Evidence for substructure in lens galaxies?

Shude Mao & Peter Schneider

Max-Planck-Institut für Astrophysik

Postfach 1523

D-85740 Garching, Germany

Abstract

We discuss whether one should expect that multiply imaged QSOs can be understood with 'simple' lens models which contain a handful of parameters. Whereas for many lens systems such simple mass models yield a remarkably good description of the observed properties, there are some systems which are notoriously difficult to understand quantitatively. We argue that at least in one case (B1422+231) these difficulties are not (solely) due to a 'wrong' parametrization of the lens model, but that the discrepancy between observed and model-predicted flux ratios is due to substructure in the lens. Similar to microlensing for optical fluxes, such substructure can distort also the radio flux ratios predicted by 'simple' mass models, in particular for highly magnified images, without appreciably changing image positions. Substructure also does not change the time delay significantly and therefore has little effect on the determination of the Hubble constant using time delays. We quantify these statements with several simple scenarios for substructure, and propose a strategy to model lens systems in which substructure is suspected.

1 Introduction

Multiply imaged QSOs and radio (Einstein) rings provide the most accurate mass measures of distant galaxies (Zwicky 1937; for specific examples, see e.g. Rix, Schneider & Bahcall 1992; Wallington, Kochanek & Narayan 1996) and promise to provide one of the most robust methods for measuring H_0 (Refsdal 1964; for recent examples, see Falco et

al. 1997; Courbin et al. 1997).[1] These successes and expectations are based on our ability to understand the lensing geometry in sufficient detail. It is truly remarkable that many multiple QSOs can in fact be modelled quite accurately with a simple elliptical deflection potential or an elliptical mass distribution. This has nourished the expectation that most lens systems are in fact due to fairly simple mass distributions (a well-known exception is the system MG2016+112; see Nair & Garrett 1997).

When modelling a multiple QSO lens system, one can either include or disregard the flux ratios of the images. Given that the number of observational constraints in these systems is never much larger than the number of free parameters of the lens model (and often is the same), there is of course a strong incentive to make use of the flux ratio information. As pointed out by Chang & Refsdal (1979), this may be a dangerous undertaking, given that the sizes of the optical continuum emitting regions of QSOs are expected to be of the same order as the Einstein radius of a star in the lens galaxy, so that the optical magnitudes may well be affected by gravitational microlensing (see Wambsganss 1990, and references therein), even if averaged over long periods of time. For example in the case of QSO2237+0305, the image positions can be fitted very accurately with a variety of lens models (e.g., Kent & Falco 1988, Rix et al. 1992, Wambsganss & Paczyński 1994), but the observed optical flux of image D is smaller than predicted, whereas the radio flux ratios (Falco et al. 1996) are in much better accord with the lens models. Therefore, constraints derived from optical fluxes should be used with care only, whereas radio sources are expected to be extended much beyond the Einstein radius of a stellar mass object, so that their flux should be (largely) unaffected by microlensing and thus provide useful constraints for lens modelling.

This argument would of course be weakened if lens galaxies contain mass clumps with an Einstein radius comparable to the size of the radio components – i.e., masses of the order of, or exceeding, globular clusters. In that case the same situation as microlensing

[1] For a recent overview on the determination of H_0 from lenses, see the proceedings of the Jodrell Bank meeting on *Golden Lenses*, held at Jodrell Bank on June 23–25, 1997, available at http://multivac.jb.man.ac.uk:8000/ceres/workshop1/proceedings.html

in the optical would apply. But this 'milli-lensing' could cause observable image splittings and distortions (Wambsganss & Paczyński 1992) which have not yet been detected, so that this possibility appears not very plausible.

Nevertheless, as we shall argue in Sect. 2, the difficulty in modelling some lens systems in detail, with radio flux ratios included, points towards the possibility of substructure in lens galaxies (or somewhere else along the line-of-sight to the QSO). Two kinds of substructure are briefly considered in Sect. 3, and numerical experiments to determine the probability of appreciable flux-ratio changes are presented in Sect. 4. In Sect. 5, we offer a practical way of treating (radio) flux ratios in the presence of substructure and apply it to B1422+231. We summarize our results in Sect. 6.

2 B1422+231: More than a challenge for lens modellers

The quadruply imaged QSO B1422+231 at $z_{\rm s} = 3.62$ was discovered in the course of the JVAS survey (Patnaik et al. 1992). The four images have a maximum separation of $1\farcs3$, and the lens galaxy has been accurately located (Impey et al. 1996) and its redshift ($z_{\rm d} = 0.34$) has recently been measured (Kundic et al. 1997, Tonry 1997). The flux ratios in the radio are A:B:C:D = 0.98:1:0.52:0.02, nearly independent of radio frequency. In the optical, the flux ratios of B:C:D are largely compatible with the radio values, whereas the A:B is quite different $\approx 0.8{:}1$; the optical flux ratios are somewhat dependent on the filter. Given that the optical flux may be affected by microlensing and/or dust obscuration, the radio flux ratios should be used in modelling this system.

Several serious attempts have been made to model the lens in this system (e.g., Hogg & Blandford 1994, Kormann, Schneider & Bartelmann 1994b, Keeton, Kochanek & Seljak 1997) – and they all failed! Whereas the image positions can be fitted very accurately, the radio flux ratios could not be obtained. Given that the parametrized lens models used by the different authors differ moderately, this failure is probably not due to a too restricted choice of the families of lens models. Rather it may be a generic difficulty for 'simple' (i.e., smooth) lens models, as can be seen as follows.

The large flux ratio between each of the images A, B, C, and the image D (~ 50)

leaves two possibilities: (i) image D is highly demagnified, or (ii) A, B, C are highly magnified. We can exclude option (i), since the separation of D from the center of the lens galaxy – $0\farcs3$ – does not put it into a region of very high surface mass density as is necessary to obtain a strong demagnification. This leaves option (ii) only. Three highly magnified images occur generically when a source is close to, and inside a cusp (see Chap. 6 of Schneider, Ehlers & Falco 1992). In that case, there exists a universal relation between the image fluxes (see, e.g., Schneider & Weiss 1992; Mao 1992), namely that the sum of the fluxes of the outer two images (A and C) should equal the flux of the middle image (B). This relation is grossly violated in B1422+231, although the VLBI structure (Patnaik & Porcas 1997) shows a strong elongation in the direction tangent to the direction to the lens galaxy thus supporting the cusp hypothesis. Hence, although this universal relation is strictly valid only asymptotically for very high magnification, or in other words, if the source is sufficiently close to the cusp (and the theory of the lens mapping near cusps developed so far does not provide a good handle of where this 'asymptotic regime' begins), this argument easily explains why all the published lens models yield flux ratios with (A+C)/B$\lesssim$1.3 (the observed value in the radio is 1.5). A strong violation of the cusp constraint can be obtained either by putting the source further away from the cusp – that would decrease the magnification, thus the magnification ratio to D, therefore running into the same problem as option (i) above – or to impose a small-scale structure on the lens model which can locally change the magnification of individual images. Here, small scale means a scale smaller than the separation between A, B, and C, i.e., smaller than $\sim 0\farcs5$.

Although B1422+231 is the system where problems in the modelling are most easily seen, it may not be unique. Keeton et al. (1997) and Falco, Lehár & Shapiro (1997) have modelled MG0414+0534, and in both attempts the resulting χ^2 per degree of freedom is too large to view these models as satisfying. In this system, the components A1 and A2 have a separation much smaller than to the distance to the other two components. This suggests that they lie close to a critical curve passing between them. In that

case, one would expect the fluxes of A1 and A2 has to be much larger than those of B and C, and that the flux ratio A1/A2 is of order unity. This expectation is not at all satisfied with the optical data, and so again one should not be surprised to find that 'simple' models fail (for this case, microlensing by stellar mass objects offers a plausible explanation, see Witt, Mao & Schechter 1995). A similar situation occurs in PG1115+080 (Keeton & Kochanek 1996) where all the χ^2 of the best fits comes from the flux ratio of A1/A2, which is significantly different from unity (Courbin et al. 1997), even though a much larger flux uncertainty is assumed for the individual images than the measurement uncertainty in order to account for microlensing. We consider these difficulties as a hint that substructure may be present in the lens galaxies. In the next two sections, we shall show how small amplitude 'perturbations' of the mass distribution in the lens are sufficient to lead to the observed discrepancies.

3 Analytic estimates

We shall consider two particular hypotheses for the kind of substructure that may be present in a lens galaxy like that in B1422+231: globular clusters (or in lensing language: point masses of order $10^6 M_\odot$), and small amplitude fluctuations.

3.1 The 'globular cluster' picture

A point mass of $\sim 10^6 M_\odot$ yields a deflection angle at its Einstein radius of about 1 milli-arcsecond, and thus can cause multiple images with similar separation, provided that source and lens are sufficiently well aligned. Since we are interested in small perturbations of the magnification of a macroimage, the typical situation will be one where the separation of the point mass from the image is considerably larger than the Einstein radius, so that the deflection angle is also smaller, and the secondary image will be highly demagnified.

Consider a macroimage with local surface mass density κ_0 and shear γ_0, with coordinates chosen such that only the 1-component of the shear caused by the macro-model is non-zero. Let a point mass be located at position $(x\cos\varphi, x\sin\varphi)$ relative to the im-

age, where x is measured in units of the Einstein radius, $\xi_{\rm E}$, of the point mass. The unperturbed and perturbed magnifications are then, respectively,

$$\mu_0 = \frac{1}{(1-\kappa_0)^2 - \gamma_0^2} \quad ; \quad \mu = \frac{1}{(1-\kappa_0)^2 - (\gamma_0+\delta\gamma_1)^2 - \delta\gamma_2^2} \ , \tag{1}$$

where $(\delta\gamma_1, \delta\gamma_2) = (\cos 2\varphi, \sin 2\varphi)/x^2$ is the shear caused by the point mass lens. Let $R = \mu_0/\mu$ be the ratio of the unperturbed to the perturbed magnification; its dependence on x and φ can be written as

$$x^2 = \frac{\gamma_0\mu_0\cos 2\varphi}{1-R}\left(1 + \sqrt{1 + \frac{1-R}{\mu_0(\gamma_0\cos 2\varphi)^2}}\right) \approx \frac{2\gamma_0\mu_0\cos 2\varphi}{1-R} \ , \tag{2}$$

where in the second step we assumed that $|1-R| \ll 1$, since we are interested in small magnification perturbations. Thus, locations $\mathbf{x}$ of the point mass for constant flux perturbation R trace an '∞'-like curve. The area of the region in which a point mass has to be located in order to cause a flux perturbation larger (smaller) than R for $R > 1$ ($R < 1$) is

$$A(R) = 2\left|\gamma_0 \frac{\mu_0}{1-R}\right| \ , \tag{3}$$

as can be easily obtained from integrating (2) over φ. Note the area is expressed in units of $\xi_{\rm E}^2$. For the lens models applicable to B1422+231, the A image should have $\mu_0 \sim 15$, $\gamma_0 \sim 0.5$, and so the area in which a point mass lens has to be located to cause a flux perturbation of $\pm 20\%$ is $A = A(0.8) + A(1.2) \sim 150$, which corresponds to $\kappa_* = \pi/A \approx 0.02$. Thus, with a surface mass density $\kappa_* \sim 0.01$, there is a reasonable probability to find one of these point-mass lenses inside the region where it causes an appreciable magnification change.

3.2 Plane wave perturbations

As a second 'model' for lens perturbations, we consider a plane density wave. For simplicity of the argument, we again assume that the shear caused by the macromodel has only a 1-component, and that the plane wave is modulated also in the 1-direction. Then, $\delta\kappa = \delta\gamma_1$, $\delta\gamma_2 = 0$. The ratio R then becomes, up to first order in the perturbation $\delta\kappa$,

$$R = 1 - \delta\kappa \frac{2}{1-\kappa_0-\gamma_0} \ . \tag{4}$$

The denominator in (4) is one of the eigenvalues of the magnification matrix caused by the macromodel. For images near a critical curve, one of the eigenvalues becomes very small, so that R can deviate significantly from unity even for small $\delta\kappa$. For the A image of B1422+231, one expects $\kappa_0 + \gamma_0 \sim 0.9$, so that for a 20% change in magnification one needs $\delta\kappa \sim 0.01$, a truly small fluctuation.

4. Numerical simulations

We model the primary lensing galaxy G1 as a singular isothermal ellipsoid (SIE) density distribution and take into account the effect of several nearby galaxies as an external shear (Kormann, Schneider & Bartelmann 1994b). The external shear may also include contributions from large scale structure and/or the deviation of the primary lens galaxy from the SIE model. We choose the coordinate system on the lens plane, $\mathbf{x} = (x_1, x_2)$, centered on the primary lens galaxy, with the x_1-axis along (decreasing) right ascension and the x_2-axis along (increasing) declination (cf. Fig. 3). The corresponding positions in the source plane $\mathbf{y} = (y_1, y_2)$ have the same orientations. The dimensionless coordinates $\mathbf{x}$ and $\mathbf{y}$ are the observed angles normalized to θ_0, an angular scale (Einstein radius) which characterizes the strength of the lens potential. We express the surface density in units of the critical surface density, $\Sigma_{\rm cr}$. For an Einstein-De Sitter universe, $1'' = 3.0h^{-1}$kpc at the lens plane, and the critical surface density is $\Sigma_{\rm cr} = 3650hM_\odot/{\rm pc}^2$, here h is the present-day Hubble constant in units of 100 km s^{-1} Mpc^{-1} and we take $h = 0.5$.

If the minor axis of the lens is along the x_1-axis, the SIE potential can be written as

$$\psi_{\rm g}(x,\phi) = x\frac{\sqrt{f}}{f'} \times [\sin\phi \ \arcsin(f'\sin\phi) + \cos\phi \ \mathrm{arcsinh}\ (\frac{f'}{f}\cos\phi)] \ , \qquad (5)$$

where f is the axis ratio, $f' = \sqrt{1-f^2}$, and we have used polar coordinates $\mathbf{x} = (x\cos\phi, x\sin\phi)$. The shear perturbation is given by

$$\psi_\gamma(x,\phi) = \frac{x^2}{2}\gamma\cos(2\phi) \ , \qquad (6)$$

if the shear direction is along the x_1-axis. The total potential is the sum of the SIE potential and the shear perturbation,

$$\psi_0 = \psi_{\rm g}(x, \phi - \phi_{\rm g}) + \psi_\gamma(x, \phi - \phi_{\rm s}) \ , \tag{7}$$

where we have now generalized to the case when the lens minor axis encloses an angle $\phi_{\rm g}$, and the direction of shear an angle of $\phi_{\rm s}$ with the x_1-axis. The lens equation is then simply: $\mathbf{y} = \mathbf{x} - \nabla\psi(\mathbf{x})$. The Jacobian of the lens mapping is $\mathcal{A} = \partial\mathbf{y}/\partial\mathbf{x}$, and the magnification of an image is given by $\mu = ({\rm det}\mathcal{A})^{-1}$. We refer the readers to Kormann et al. (1994a,b) for more technical details.

4.1 A macromodel

Observational data is taken from Tables 1 and 2 in Impey et al. (1996).[2] The total number of observational constraints is 12, namely we have 4×2 image positions and 4 fluxes. Our lens model is described by 8 parameters, i.e., three lens parameters, $(\phi_{\rm g}, f, \theta_0)$, two shear parameters, $(\gamma, \phi_{\rm s})$, two source coordinates, and the unlensed source flux S_0. We keep the lens galaxy position fixed. Thus, the number of degrees of freedom is 4. Nearly identical procedures to those in Kormann et al. (1994b) are used to find the best lens model. A minor difference is that we have only included the flux ratios B:C:D in the χ^2 measure, since these flux ratios are nearly the same in the radio and in the optical. The χ^2 surface is not well constrained in the $f - \gamma$ plane (cf. Figure 8 in Keeton et al. 1997). We have summarized a model that has an axis ratio close to the observed light distribution in Table 1. The model reproduces the observed positions within 1σ of the observational uncertainties. The relative magnifications are, however, unsatisfactory. The flux ratio D:B = 0.037 is much higher than the observed values in the radio (0.02) and in the optical (0.025). The predicted flux ratio A:B, 0.76, deviates significantly from the observed radio value, 0.98. Curiously, this ratio is very close to the optical value, 0.78 (this is also true for the models in Hogg & Blandford 1994; the models in Kormann et al. 1994b have A:B$\gtrsim$ 0.9; however, these previous models used less precise lens positions.) We have explored a magnification gradient as a way of changing the radio flux ratio, as seen in the double lens 0957+561 (Conner, Lehár & Burke 1992), but came to the

[2] There is a typo for the image D position in Table 1 of Impey et al. (1996). The position relative to B should be $-0\farcs957$ in RA and $+0\farcs799$ in DEC (C. Petry 1997, private communication), in good agreement with the radio positions.

conclusion that this is not an attractive explanation. The current paradigm of QSOs generally predicts that the optical continuum emission should come from a region much smaller in size than the radio emission (which in this case is at least partially resolved with VLBI). If this picture is valid, we are more inclined to regard it a coincidence that the models predict a flux ratio close to the optical value, since the flux of a 'small' source is always more sensitive to substructure than a 'big' source. For this coincidence, we first need to invoke the substructure model proposed here to change the radio flux ratios, and then microlensing to cause the the deviation of the optical flux ratios from those in the radio.

Table 1: Model Parameters

$\theta_0('')$	f	$\phi_g(^\circ)$	γ	$\phi_s(^\circ)$	μ_A	μ_B	μ_C	μ_D
0.77	0.84	-53.9	0.21	-59.8	7.49	-9.90	4.93	-0.37

4.2 The 'globular cluster' picture

The presence of massive lenses with mass $\sim 10^6 M_\odot$ provides a perturbation potential:

$$\delta\psi = \hat{m}\sum_i \frac{1}{2}\ln[(x-x_i)^2+(y-y_i)^2], \quad \hat{m} = \frac{M}{\Sigma_{\rm cr}\pi(D_{\rm d}\theta_0)^2}, \tag{8}$$

on top of ψ_0, where θ_0 is again the Einstein radius of the lens galaxy, $D_{\rm d} = 610h^{-1}$Mpc is the distance to the lens galaxy, and $M = 5\times 10^5 M_\odot$ is the mass of the lenses. For simplicity we assume that these lenses are uniformly distributed and have a surface density of κ_* in units of the critical surface density. Similar to microlensing by solar mass objects (Wambsganss 1990), we generate point lenses randomly on the lens plane with the required surface density. To quantify the deviation of magnifications from that of the unperturbed potential, we define a quantity $r = (\mu_A+\mu_B+\mu_C)/(|\mu_A|+|\mu_B|+|\mu_C|)$. If the magnifications of image A, B and C strictly obey the prediction of a cusp singularity, $r = 0$. The input value resulting from the unperturbed potential ψ_0 is $r = 0.113$. In Fig. 1, we show the probability distribution of r as obtained from Monte

Carlo simulations. For $\kappa_* = 0.005, 0.01$, the probability of having a deviation larger than the observed one is about 21% and 37%. Therefore, a surface density in 'globular clusters' $\kappa_* \approx 0.005$ may be sufficient to explain the observed deviation in flux ratios. Nearly in all the cases, the change in the image positions are a few milli-arcsecond. Such small changes are undetectable in the optical. In any case, such small changes can probably be accommodated by a small adjustment in the macromodel.

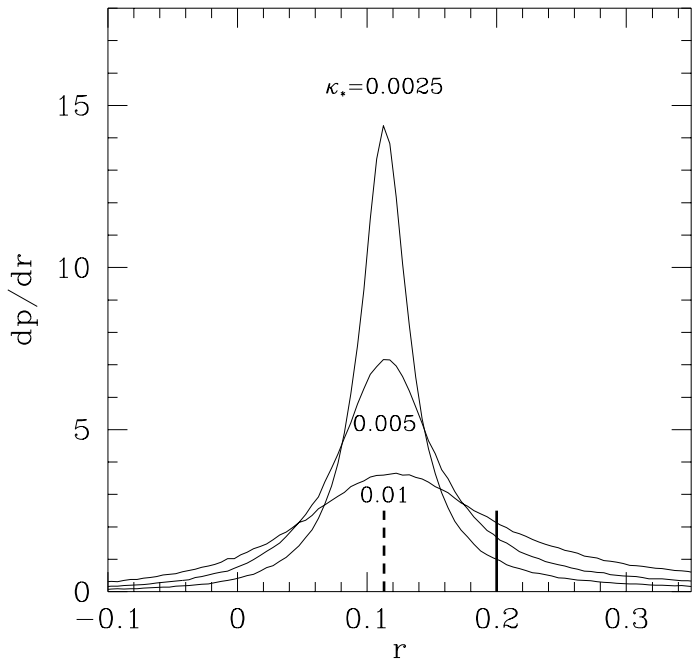

Fig. 1. Differential probability distribution for the quantity $r = (\mu_A + \mu_B + \mu_C)/(|\mu_A| + |\mu_B| + |\mu_C|)$ for a singular isothermal ellipsoid with external shear and massive point lenses. r measures the deviation of magnifications from the prediction ($r = 0$) of a cusp singularity. Equal-mass point lenses are distributed uniformly with surface density κ_*. The dashed vertical line indicates the input value for the unperturbed lens model. The solid vertical line shows the observed value of $r = 0.2$. Three curves are shown for $\kappa_* = 0.0025, 0.005$ and 0.01. The probability of having a deviation larger than the observed one is 11(21, 37)% for $\kappa_* = 0.0025$ (0.005, 0.01), respectively.

4.3 Plane wave perturbation

For each image, we consider a plane-wave fluctuation centered on the image, described by

$$\delta\psi = \frac{2\kappa_s}{|\mathbf{k}|^2}\cos(k_1\theta_0 x_1 + k_2\theta_0 x_2 + \phi_0)e^{-\frac{x^2\theta_0^2}{2\sigma_r^2}} \ , \tag{9}$$

where κ_s is the amplitude of fluctuation, (k_1, k_2) is the wave vector, ϕ_0 is a phase constant, and the last exponential term localises the perturbation within a few σ_r. Notice that in the absence of the last term, the amplitude of surface density corresponding to $\delta\psi$ is κ_s. Since the parameter space is rather large, we limit our study to one illustrative case. We fix the magnitude of the wave-vector to be $2\pi/|\mathbf{k}| = 0\farcs01$ but with random orientations. ϕ_0 is distributed uniformly between 0 to 2π, and the amplitude of fluctuations is drawn from a Gaussian distribution with zero mean and standard deviation σ_κ. We take $\sigma_r = 0\farcs03$, but the results are insensitive to the choice of σ_r. In Fig. 2, we show the probability distribution of r for $\sigma_\kappa = 0.005, 0.01, 0.02$. The probability of having deviations larger than the observed one is 1.3(5.9, 11)% for $\sigma = 0.005$ (0.01, 0.02), respectively. These values are in good agreement with the analytical estimates presented in Sect. 3.

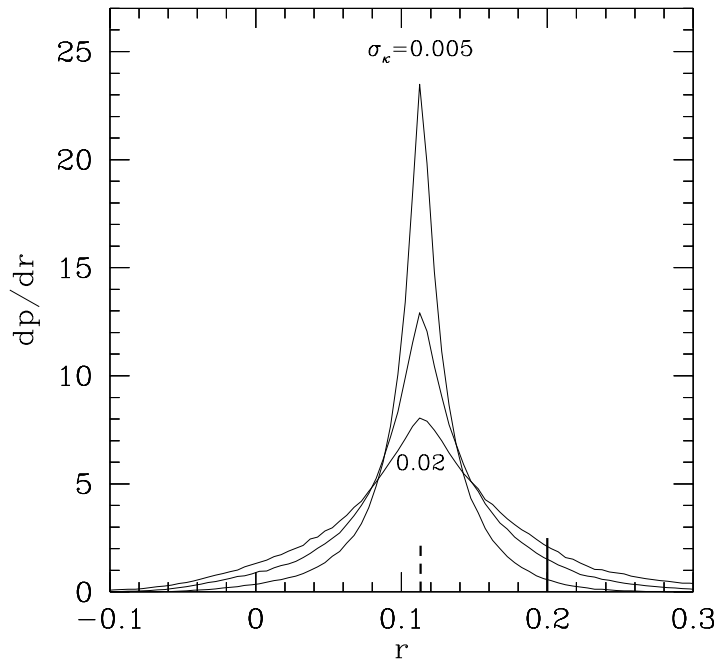

Fig. 2. Differential probability distribution for the quantity $r = (\mu_A + \mu_B + \mu_C)/(|\mu_A| + |\mu_B| + |\mu_C|)$ for a singular isothermal ellipsoid with external shear and localised plane-wave fluctuations (cf. eq. [9]). The amplitude of fluctuations is drawn from a Gaussian distribution with standard deviation σ_κ. Three curves are shown with $\sigma_\kappa = 0.005, 0.01, 0.02$. The probability of having deviations larger than the observed one is 1.3(5.9, 11)% for $\sigma_\kappa = 0.005$ (0.01, 0.02), respectively. The dashed and solid lines have the same meaning as those in Fig. 1.

4.4 Size of fluctuation

Let us estimate how small (in size) the smooth fluctuations have to be in order to explain the observed deviation in B1422+231. We consider a similar perturbation potential as that in Eq. (9), but without the localisation exponential term. For simplicity, we take Fourier modes that are stationary in a square with a side length, L, centered on the primary lens galaxy, i.e., the perturbation potential is

$$\delta\psi = \sum_{i,j=1}^{N} \sum_{l,m=0}^{1} \frac{2\kappa_{ijlm}}{k_{ij}^2} \cos(i\frac{\pi}{L}\theta_0 x_1 - \frac{l\pi}{2}) \cos(j\frac{\pi}{L}\theta_0 x_2 - \frac{m\pi}{2}), \quad k_{ij}^2 = \frac{\pi^2}{L^2}\left(i^2+j^2\right) \quad (10)$$

where l and m are either 0 or 1 for cosine and sine waves, i and j specify the order of Fourier modes, N controls the number of Fourier modes used, and κ_{ijlm} are the surface density of fluctuations for the mode specified by i, j, l, and m. Smaller and smaller scale fluctuations are used when N increases. We take the side length of the box to be $4''$. This size is chosen such that the periodic boundary condition does not influence the central region of the deflection potential. Notice that the number of extra parameters is $4N^2$, therefore for $N = 1$, the number of degrees of freedom is already zero; for $N \geq 2$, the system has more free parameters than constraints. The best fit parameters are again found by minimizing a χ^2 measure, which now utilizes all the flux ratios (cf. Kormann et al. 1994b). To avoid excessively large fluctuations, we include a term $\delta\chi^2 = \sum_{i,j}\sum_{l,m}(\kappa_{ijlm}/0.05)^2$. For the model presented in Table 1, we find $\chi^2 = 100, 13.8, 5.7, 1.8$ for $N = 0, 1, 2, 3$, while for $N \geq 4$, χ^2 has essentially dropped to zero. For $N = 0, 1, 2$, most χ^2 comes from modified discrepancies with observed flux ratios. It is interesting that for $N = 1, 2$, when the number of parameters is equal to or larger than the number of constraints, the χ^2 remains high. This shows that large scale fluctuations are ineffective in reducing the χ^2, again highlighting the need for small-scale structures. The fluctuation wavelengths corresponding to $N = 4, 5$ are $2'', 1\farcs6$. Therefore fluctuations with wavelength of $\lesssim 1\farcs5$ are sufficient to reproduce the observed flux ratios. The resulting surface density distribution for $N = 5$ is shown in Fig. 3 as the solid contours. It is obvious that the contours close to the brighter images are somewhat

twisted. In particular, the surface density at the positions of image A and C are increased while that at the position of image B is decreased. The magnifications are 17.1, −17.5, 9.1 for A, B and C, respectively while that for image D is essentially unchanged (−0.36). However, notice in Fig. 3, the outer density contours are moderately distorted. Such distortions can be avoided by fluctuations with smaller wavelengths. Therefore smaller scale fluctuations are preferred.

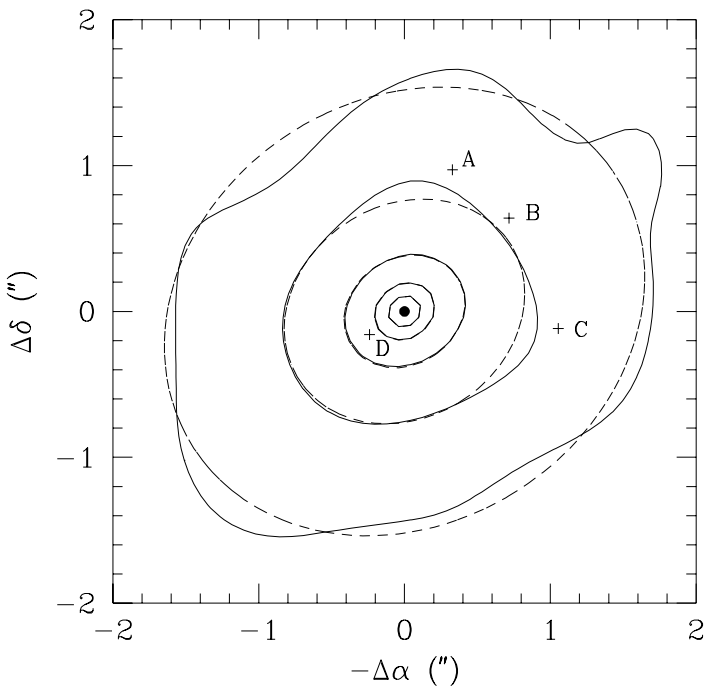

Fig. 3. The contours of constant surface density for a singular isothermal ellipsoid with external shear and plane wave fluctuations described by eq. (10), with $N = 5$ and $L = 4''$. The contour levels are 4, 2, 1, 0.5 and 0.25 of the critical surface density, respectively. The thin lines show the corresponding contours for the underlying singular isothermal ellipsoid. The four crosses show the image positions and the central dot shows the position of the primary lens galaxy, G1.

It should be noted that the model presented in Fig. 3 is not at all unique. The 'best fit model' depends certainly on the macromodel from which the χ^2-minimization was started, and even for a fixed starting model, there are many different models for which χ^2 is nearly zero for $N \geq 4$. Therefore, the model in Fig. 3 should not be viewed as a realistic lens model, but the exercise should demonstrate the largest possible scale of

substructure which can bring the flux ratios within the observed range.

5 Practical modelling of lenses with substructure

If we accept the view that the surface mass density of lens galaxies is not as smooth as described by 'simple' lens models, using flux ratios for modelling may yield misleading conclusions. On the other hand, discarding them altogether would result in a loss of information; in particular, as long as the perturbations are small, the magnification is only weakly affected for images which are not highly magnified. Therefore, a statistical treatment may be appropriate. In this section we propose a simple method to perform lens modelling in the presence of small-scale structure in the lens.

Let $\hat{p}_\mu(|\mu|\,;\kappa_0,|\gamma_0|)$ be the probability density for a magnification $|\mu|$, given that the smooth model has surface mass density κ_0 and (complex) shear γ_0. If the unlensed flux of the source is S_0, the probability density for the magnified flux $S=|\mu|\,S_0$ becomes

$$p_S(S;\kappa_0,|\gamma_0|)=\frac{1}{S_0}\,\hat{p}_\mu\left(\frac{S}{S_0};\kappa_0,|\gamma_0|\right)\ . \tag{11}$$

Furthermore, let the observed flux be $S^{(\mathrm{obs})}=xS$, where the deviation of x from unity accounts for measurement uncertainties. Then, the likelihood to obtain a flux measurement $S^{(\mathrm{obs})}$ becomes

$$\hat{p}_S(S^{(\mathrm{obs})};\kappa_0,|\gamma_0|)=\int_0^\infty\frac{\mathrm{d}S}{S}\,p_S(S;\kappa_0,|\gamma_0|)\,p_x\left(\frac{S^{(\mathrm{obs})}}{S}\right)\ , \tag{12}$$

where $p_x(x)$ is the probability density of the 'measurement error' x. Therefore, for N images with observed positions $\boldsymbol{\theta}_i^{(\mathrm{obs})}$, positional uncertainty σ_i, and observed fluxes $S_i^{(\mathrm{obs})}$, the likelihood function that can be maximized for obtaining model fits is

$$\mathcal{L}=\prod_{i=1}^{N}\frac{1}{\pi\sigma_i^2}\exp\left(-\frac{\left|\boldsymbol{\theta}_i-\boldsymbol{\theta}_i^{(\mathrm{obs})}\right|}{\sigma_i^2}\right)\,\hat{p}_S\left(S_i^{(\mathrm{obs})};\kappa_i,\gamma_i\right)\ , \tag{13}$$

where $\boldsymbol{\theta}_i$ are the predicted image positions, and $\kappa_i=\kappa_0(\boldsymbol{\theta}_i)$ and $\gamma_i=|\gamma_0(\boldsymbol{\theta}_i)|$ are the surface mass density and shear at these positions.

For concreteness, we shall assume that the determinant $D \equiv \det(\mathcal{A}) = (1-\kappa_0-\delta\kappa)^2 - |\gamma_0+\delta\gamma|^2$ is distributed like a Gaussian with mean $\langle D\rangle$ and dispersion σ_D. If, for example, the perturbation quantities $\delta\kappa$ and $\delta\gamma$ were Gaussian,

$$p(\delta\kappa,\delta\gamma) = \frac{1}{\sqrt{2}\pi^{3/2}\sigma_\kappa\sigma_\gamma^2}\exp\left(-\frac{(\delta\kappa)^2}{2\sigma_\kappa^2} - \frac{|\delta\gamma|^2}{\sigma_\gamma^2}\right) , \tag{14}$$

then D follows very closely a Gaussian distribution, with

$$\langle D\rangle = D_0 + \sigma_\kappa^2 - \sigma_\gamma^2 \quad ; \quad \sigma_D^2 = 4(1-\kappa_0)^2\sigma_\kappa^2 + 2|\gamma_0|^2\sigma_\gamma^2 + \sigma_\gamma^4 + 2\sigma_\kappa^4 , \tag{15}$$

where $D_0 = (1-\kappa_0)^2 - |\gamma_0|^2$ is the determinant for the unperturbed lens. Since $\mu = 1/D$, the probability density for the magnification is $p_\mu(\mu) = p_D(1/\mu)/\mu^2$, and so the probability density for the absolute value of μ becomes

$$\hat{p}_\mu(|\mu|) = \frac{1}{|\mu|^2}\left[p_D\left(\frac{1}{|\mu|}\right) + p_D\left(\frac{-1}{|\mu|}\right)\right] ; \tag{16}$$

for notational simplicity we currently drop the arguments κ_0 and γ_0. Assuming a Gaussian distribution for x, with mean 1 and dispersion σ_x, we obtain from the foregoing expressions that

$$\hat{p}_S(S^{(\rm obs)}) = \frac{1}{2\pi S_0\sigma_D\sigma_x\, a}\left(\mathrm{e}^{-b} + \frac{\sqrt{\pi}}{2}\left[g\left(\frac{c_1+c_2}{2\sqrt{a}}\right) + g\left(\frac{c_1-c_2}{2\sqrt{a}}\right)\right]\right) , \tag{17}$$

and we introduced the abbreviations

$$\begin{aligned} a &= \frac{1}{2\sigma_D^2} + \frac{(S^{(\rm obs)}/S_0)^2}{2\sigma_x^2} \quad , \quad b = \frac{\langle D\rangle^2}{2\sigma_D^2} + \frac{1}{2\sigma_x^2} , \\ c_1 &= \frac{S^{(\rm obs)}}{S_0\,\sigma_x^2} \quad , \quad c_2 = \frac{\langle D\rangle}{\sigma_D^2} , \\ g(z) &= z\,\mathrm{e}^{z^2-b}\,\mathrm{erfc}(-z) , \end{aligned} \tag{18}$$

where $\mathrm{erfc}(z)$ is the complimentary error function. This expression for $\hat{p}_S$ is easy to calculate, e.g., by using the routine `erfcc` of Press et al. (1992).[3] In Fig. 4, we have plotted the function $\hat{p}_S(S^{(\rm obs)})$ for several combinations of κ_0 and γ_0. As can be readily seen, the probability distribution broadens considerably for larger values of the magnification of the

[3] A FORTRAN77 routine which calculates $\hat{p}_S(S^{(\rm obs)})$ can be obtained from the authors on request.

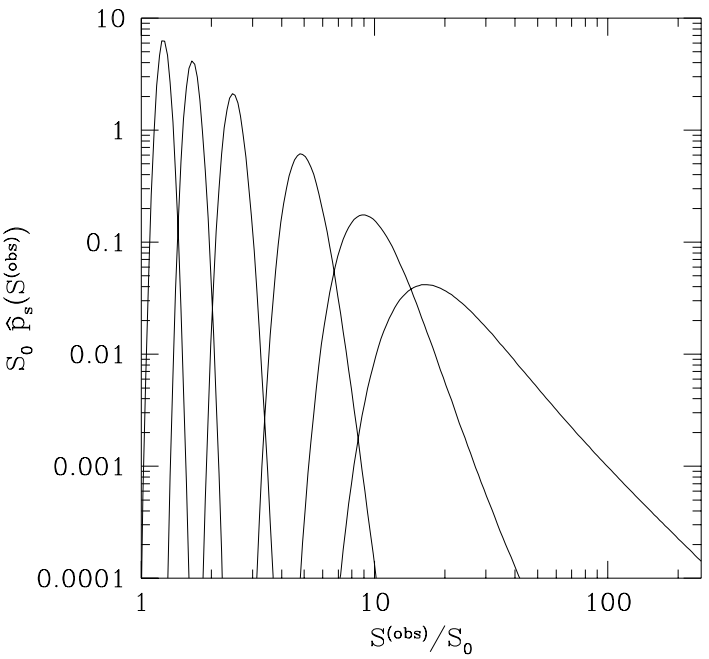

Fig. 4. The probability density $\hat{p}_S(S^{(\rm obs)})$, according to eq. (17) of the text. The parameters of the macromodel was chosen to be $\kappa_0 = \gamma_0$, with $\kappa_0 = 0.1$ (left-most curve), 0.2, 0.3, 0.4, 0.45, 0.48. The probability density for the perturbations $\delta\kappa$ and $\delta\gamma$ have been assumed to follow a Gaussian, with width $\sigma_\kappa = \sigma_\gamma = 0.02$. Furthermore, the measurement errors were characterized by a Gaussian distribution in x with width $\sigma_x = 0.02$. It is easily seen that images with higher magnifications attain a much broader flux distribution than those where the magnification is of order unity.

macromodel. Hence, the fluxes of weakly magnified images provide strong constraints, whereas highly magnified images yield weaker constraints on the model.

We now apply this formalism to B1422+231. We start with the macromodel in Table 1, and examine the best likelihood, $\mathcal{L}$, as a function of σ_κ (we take $\sigma_\gamma = \sigma_\kappa$). The goodness-of-fit is measured by the difference, $\Delta\ell$, between $\ell = \ln\mathcal{L}$ and the maximum $\ell_{\rm max} = \ln\mathcal{L}_{\rm max}$. $\ell_{\rm max}$ can be obtained from equation (13) by setting the exponential term to 1 (perfect fit to the positions) and finding the maximum for the function of $\hat{p}_s$ for a given set of parameters (cf. Figure 4). For a good fit, we expect $2\Delta\ell$ to be close to the number of degrees of freedom, 4 (Lupton 1993). Figure 5 shows the numerical result. $2\Delta\ell$ decreases dramatically when σ_κ increases from 0 to 0.01, and then stays nearly flat for $\sigma_\kappa \gtrsim 0.01$. This shows that a perturbation of $\sigma_\kappa \approx 0.01$ is efficient in increasing

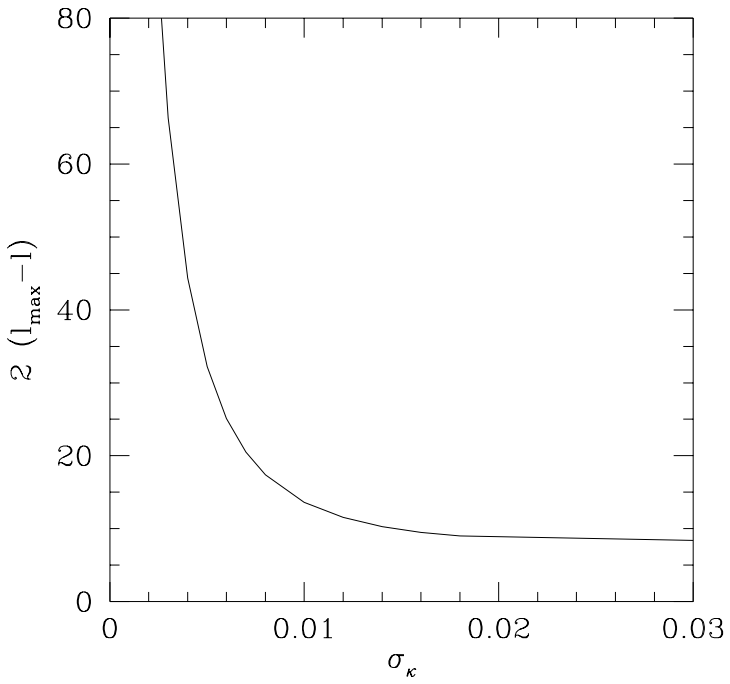

Fig. 5. The 'goodness-of-fit', $2\Delta\ell = 2(l_{\rm max} - l)$, as a function of σ_κ (see text). A good fit is expected to have a value close to 4 for B1422+231.

the likelihood by relaxing the constraint on the flux ratios. Unfortunately, $\Delta\ell$ remains somewhat high, about a factor of 2 above the expected value. This is primarily because our macromodel predicts too bright an image D, and substructures with amplitude $\sigma_\kappa \approx 0.01$ do not reduce the magnification of image D sufficiently to be consistent with the observed value. A possible cure is to increase the local surface density at image D, by, e.g., making the density slope at small radius steeper. It is beyond the scope of this paper to explore this issue in detail. We plan return to this issue and perform a more extensive study of the parameter space.

6 Discussion

We have considered the possibility that the flux ratios in lens systems may be affected by small scale structure in the lens galaxy. Although we discussed only the SIE model, we have verified that it still seems difficult to fit the observed flux ratios using other power-law models of the deflection potential. A general analytic argument was presented

why 'simple' lens models are unable to account for the radio flux ratios in B1422+231. We considered two potential pictures for substructure, namely 'globular clusters' with $\sim 10^6 M_\odot$ and smooth perturbations, modelled as plane density waves. We have studied both pictures analytically, as well as through numerical simulations. Owing to the large magnifications of images A, B, C in B1422+231, only small perturbations (in mass) are needed to change magnifications, and thus flux ratios appreciably.

The required perturbation is of the order of $\sim 1\%$ of the critical surface density, or about $(40 h M_\odot \mathrm{pc}^{-2})$ in physical units for B1422+231. There are about 200 globular clusters in the Galaxy with a total mass of $\approx 10^8 M_\odot$. These globular clusters form a quasi-spherical distribution concentrated toward the galactic center. The average surface density of these objects within 5kpc is $\sim 1 M_\odot \mathrm{pc}^{-2}$. This is $2.5 \times 10^{-4} h^{-1}$ of the critical surface density, somewhat too low to produce a significant change in magnification (cf. Fig. 1). The 'globular cluster' picture should apply equally well to the hypothetical massive black holes in galactic halos (e.g., Lacey & Ostriker 1985). They can produce much more visible deviations if just a few percent of the halo mass are in these objects. Our second model of substructure addresses smooth fluctuations. An obvious example of smooth fluctuations is the spiral arms observed in disk galaxies. Smooth inhomogeneities are also quite naturally expected in hierarchical structure formation models as a result of continuous merging and accretion of sub-clumps.

An important question is whether the added perturbation affects the time delay significantly: we found from models like those shown in Figs. 2 and 3 that the perturbation changes the time delay at most by a few percent. The physical explanation is clear: the time delay is $\propto \psi$. Since our perturbation only changes the potential by $\delta\psi/\psi \propto \delta\kappa/\kappa \sim 1\%$, therefore the time delay is only affected at a few percent level. We conclude that substructure can have only minor effects on the time delay – gravitational lensing is still potentially a golden tool for determining the Hubble constant.

We thank Matthias Bartelmann, Alok Patnaik, Cathy Petry, Simon White and Hongsheng Zhao for helpful discussions and comments on the paper. In particular, we

are grateful to the referee Chuck Keeton for his most constructive comments which helped to improve the paper. This work was supported by the "Sonderforschungsbereich 375-95 für Astro–Teilchenphysik" der Deutschen Forschungsgemeinschaft.

References

Chang K., Refsdal S., 1979, Nature, 282, 561

Conner S. R., Lehár J., Burke B. 1992, ApJ, 387, L61

Courbin F., Magain P., Keeton C.R., Kochanek C.S., Vanderriest C., Jaunsen A.O., Hjorth J. 1997, preprint (astro-ph/9705093)

Falco E.E., Lehár J., Perley R.A., Wambsganss J., Gorenstein M.V., 1996, AJ, 112, 897

Falco E.E., Lehár J., Shapiro I.I., 1997, AJ, 113, 540

Falco E.E., Shapiro I.I., Moustakas L.A., Davis M., 1997, ApJ, 484, 70

Hogg D.W., Blandford R.D., 1994, MNRAS, 268, 889

Impey C.D., Foltz C.B., Petry C.E., Browne I.W.A., Patnaik A.R., 1996, ApJ, 462, L53

Keeton II C.R., Kochanek C.S., 1996, preprint (astro-ph/9611216)

Keeton II C.R., Kochanek C.S., Seljak U., 1997, ApJ, 482, 604

Kent S.M., Falco E.E., 1988, AJ, 96, 1570

Kochanek C.S., 1991, ApJ, 373, 354

Kormann R., Schneider P., Bartelmann M., 1994a, A&A, 284, 285

Kormann R., Schneider P., Bartelmann M., 1994b, A&A, 286, 357

Kundic T., Hogg D.W., Blandford R.D., Cohen J.G., Lubin L.M., Larkin J.E., 1997, preprint (astro-ph/9706169)

Lacey C. G., Ostriker J. P., 1985, 290, 154

Lupton R. 1993, *Statistics in theory and practice* (Princeton University Press: Princeton)

Mao S., 1992, ApJ, 389, 63

Nair S., Garrett M.A., 1997, MNRAS 284, 58.

Patnaik A.R. et al., 1992, MNRAS, 259, $1_{\rm P}$

Patnaik A.R., Procas R.W., 1997, preprint

Refsdal S., 1964, MNRAS, 128, 307

Rix H.-W., Schneider D.P., Bahcall J.N., 1992, AJ, 104, 959

Schneider P., Ehlers J., Falco E. E., 1992, *Gravitational Lenses* (Springer-Verlag: New York)

Schneider P., Weiss A., 1992, A&A, 260, 1

Tonry J. L. 1997, preprint (astro-ph/9706199)

Wallington S., Kochanek C.S., Narayan R., 1996, ApJ, 465, 64

Wambsganss J., 1990, report MPA 550, Garching

Wambsganss J., Paczyński B., 1992, ApJ, 397, L1

Wambsganss J., Paczyński B., 1994, AJ, 108, 1156

Witt H.J., Mao S., Schechter P., 1995, ApJ, 443, 18

Zwicky F., 1937, Phys. Rev., 51, 290