

The surface density profile of a lens galaxy is 97.6 per cent elliptical power-law

Devon Powell (✉ dmpowell@mpa-garching.mpg.de)

Max Planck Institute for Astrophysics

Simona Vegetti

Max-Planck-Institut fuer Astrophysik

John McKean

Cristiana Spingola

INAF - Istituto di Radioastronomia <https://orcid.org/0000-0002-2231-6861>

Hannah Stacey

Max Planck Institute for Astrophysics

Chris Fassnacht

UC Davis

Article

Keywords:

Posted Date: March 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1335845/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

The surface density profile of a lens galaxy is 97.6 per cent elliptical power-law

The projected density profiles of gravitational lens galaxies are often assumed to be consistent with a simple elliptical power-law^{1,2}. Previous studies at low angular resolution have found this description to be satisfactory^{3,4,5}, but the limitations of this smooth model assumption have not yet been observationally tested. Here we report the analysis of a gravitational lens system observed at milli-arcsecond angular resolution with very long baseline interferometry at 1.7 GHz⁶ using a state-of-the-art Bayesian lens modelling technique for interferometric data⁷. An elliptical power-law mass distribution is found to be a remarkably good description of the data, but there remain source-plane light components that are misaligned by several milli-arcseconds. The addition of angular multipole perturbations gives a significantly better fit (Bayes factor of +4751), while the best-focused source model is recovered using pixellated corrections to the lens potential. The projected density in our best model deviates from an elliptical power-law by 2.4 per cent (RMS) inside a region within 17 mas of the lensed images. The assumption of a perfect elliptical power-law would bias the measurement of H_0 by +0.6 per cent, and the observed flux ratios by up to 10.3 per cent, for this lens. Our results demonstrate that low-level density perturbations exist in this massive elliptical galaxy. Their presence can significantly affect studies of galaxy formation via flux-ratio anomalies^{8,9}, but do not strongly affect time-delay measurements of the Hubble constant¹⁰.

MG J0751+2716 is a galaxy with powerful radio jets at redshift 3.2 that is gravitationally lensed by an intervening massive early-type galaxy at redshift 0.35^{11,12,13,14}. The jetted nature of the background radio source results in extended gravitational arcs that probe a wide range of angular and radial positions relative to the lens centre. The gravitational arcs are sufficiently bright to be imaged at 3.7 milli-arcsecond (mas) angular resolution with global Very Long Baseline Interferometry (VLBI; observing $\lambda = 18$ cm)⁶. Given the large area of the lens potential probed by the extended radio emission, coupled with the angular resolution and sensitivity of the VLBI imaging, MG J0751+2716 provides a unique laboratory to investigate the mass distribution of the foreground massive galaxy with

28 a resolution of 18 pc.

29 We have tested three parametric lens models of varying complexity. Our fiducial model is a power-
30 law ellipsoid mass distribution (PEMD; see Methods) density profile with an external shear component,
31 which we refer to as PL. Our second model, PL+FG, consists of the same PL parameterization for the
32 lens galaxy, but with five additional PEMDs to account for field galaxies within 5 arcsec of the system.
33 Finally, we consider a single PEMD with an external shear and an additional angular structure in the
34 form of multipole density perturbations up to fourth order, which we refer to as PL+M4. We quantify
35 the ability of the PL+FG and PL+M4 models to explain the observed data in terms of their Bayes
36 factors, the difference in the log-Bayesian-evidence relative to the fiducial model PL. We fit each of
37 these models to the full surface brightness distribution of the gravitational arcs using a state-of-the-art
38 grid-based lensing code that is designed to operate in the native visibility-plane of the data⁷ (see also
39 refs. ^{15,16,17,18}). We jointly infer the posterior probability distributions on the mass model parameters,
40 recover the pixellated source surface brightness distribution, and compute the marginalized Bayes
41 factor $\Delta \log \mathcal{E}$ to enable a robust model comparison. In order to recover further details in the mass
42 structure of the lens, we then augment each parametric model by iteratively computing pixellated
43 corrections to the lensing potential. We denote these potential corrections as model PX, such that, for
44 example, the PL+M4 parametric model plus potential corrections is named PL+M4+PX. As we cannot
45 efficiently marginalize over all pixels in the potential corrections (see Methods), we compare the PX
46 results separately using their maximum log-posterior values $\Delta \log P$ relative to the fiducial PL+PX
47 model. The main quantitative results are summarized in Table 1 (see also Table 2 of the Methods for
48 the individual model parameters).

49 We find that the fiducial model PL is able to fit the data remarkably well. The jet structure is
50 clearly recovered in the unlensed source surface brightness distribution, with several knots of emission
51 stretching 140 mas from northeast to southwest (components 1, 2, 3, 4 and 6 following the nomenclature
52 of ref. ¹¹; see Figure 2), which is equivalent to a projected jet length of 1.08 kpc at redshift 3.2. While
53 the recovered source is almost entirely focused by the PL lens model, it misaligns source component
54 4, which straddles the caustic, by 2 mas, and disrupts components 2 and 3. The PL+FG model, which
55 was equally favoured with a power-law model in the analysis by ref. ⁶, provides a significantly poorer

56 fit to the data, with a Bayes factor of $\Delta \log \mathcal{E} = -7060$ relative to model PL. In addition to exacerbating
57 the misalignment of source component 4 to 5 mas, the PL+FG model also produces a highly disrupted
58 source component 3 and a mis-aligned source component 6. The poor Bayes factor for model PL+FG
59 indicates that this form of parameterization for groups is still not well understood, in terms of how best
60 to account for the mass-to-light ratios of the group galaxies and/or the effects of a group halo¹³. Model
61 PL+M4 drastically improves the fit by a Bayes factor of +4751 over the PL model, properly focusing
62 source component 4. This is a strong indication that there is additional angular structure beyond the
63 simple ellipticity provided by a PEMD mass distribution. After extending each model using pixellated
64 potential corrections, we find that the source model is clearly focused in all three cases (PL+PX,
65 PL+FG+PX, PL+M4+PX). We find that PL+FG+PX is disfavored by $\Delta \log P = -1084$ relative to
66 PL+PX. PL+M4+PX gives $\Delta \log P = -21$, which is neither strongly favored nor disfavored. This
67 indicates that in the presence of pixellated potential corrections, the multipole terms in the parametric
68 model are unnecessary, as the potential corrections are able to absorb this angular structure. Due to
69 its relative simplicity and its ability to properly focus the source, we consider PL+PX to be our “best”
70 model.

71 To quantify the deviation of the PL+PX from the PL model, we determine the fractional difference in
72 the projected surface density, $\Delta\kappa$, between the two (see Figure 1; we show the results for all models in
73 Figure 3). We compute the RMS of this difference in a masked region containing all points within 17
74 mas (3 beam widths) of the lensed emission, which yields $\sigma_{\Delta\kappa} = 2.4$ per cent. Hence, on the 174-parsec
75 scales probed by the potential corrections close to the lensed images, at least 97.6 per cent of the mass
76 distribution can be described by a smooth PEMD model. In this respect, our analysis validates the use
77 of a PEMD model for those applications of gravitational lensing where only the large-scale properties
78 of the global mass model are relevant. This has important implications for modelling the large number
79 of gravitational lenses to be found with, for example, *Euclid*, where the angular resolution is relatively
80 low (> 100 mas) and a simple PEMD will most likely be assumed.

81 We now discuss the observational consequences of this 2.4 per cent deviation from a globally smooth
82 mass distribution. Measurements of the Hubble constant (H_0) inferred with time-delay cosmography
83 are known to be biased if the mass profile of the lens is not sufficiently well-known^{19,20,21}. We find

that the models with the largest bias are PL+FG and PL+FG+PX, which would underestimate H_0 by 22.4 and 24.8 per cent, respectively, when compared to the PL+PX model; this can be explained by the additional group of galaxies that creates an effective mass sheet around the lensing galaxy. Also, for the PL+M4 model, the presence of additional degrees of freedom in the mass profile could change the measured value of H_0 by a significant amount; in this case it would be lowered by 6.2 per cent. Our best parametric model, PL, gives a bias in H_0 of +0.6 per cent. Most of the relative changes in H_0 are due to the underlying parametric lens model, rather than the potential corrections themselves, in line with the findings of ref.¹⁰. The fact that the PL+PX model is able to focus the source with minimal departure from the smooth PL model, in terms of density and the inferred Hubble constant, indicates that multipole terms, while attractive in their parametric simplicity, can bias inferred global properties of the lens while attempting to fit local image positions and magnifications^{20,22}.

We also assess the impact of the assumed lens profile on the measured flux-ratios of the lensed images. We compute the observed magnifications at the position of the brightest source component (component 1; see Figures 2 and 3 of the Methods) for each of the lens models. We find that the maximum change in magnification for any of these images (relative to PL+PX, with the mean magnification removed) is between 10.3 and 14.9 per cent for the different parametric models. Such large changes are comparable to the scatter in the measured flux ratios from gravitational lens systems, which are typically attributed to unconstrained mass structure in the lens^{8,9}. Models that include potential corrections find a maximum change in magnification of up to 1.6 per cent. This demonstrates that density structure in the lens beyond a PEMD is a plausible source of this scatter, and that the density profile must depart from a PEMD by only a few per cent locally in order to produce flux-ratio anomalies.

To illustrate the importance of high resolution imaging in differentiating lens model parameterizations, we repeated our analysis using a 2.12 μm observation of MG J0751+2716 taken using the adaptive optics (AO) system on the W. M. Keck-II Telescope, which provides an angular resolution of ~ 65 mas (point spread function FWHM). We find that models PL+FG and PL+M4 give a Bayes factor of 3.4 and 6.7, respectively, three orders of magnitude weaker than those obtained from the global VLBI observation. Therefore, the ability to resolve source structure in the lensed images is

112 of major consequence for sensitivity to the mass structure in lens galaxies. Our analysis shows the
113 computational feasibility of modeling large, high-resolution interferometric data with 2.5×10^8 visi-
114 bilities. The numerical techniques from⁷, along with the fast χ^2 computation (see Methods), allows
115 the efficient evaluation of the posterior samples, independent of the number of visibilities. This new
116 ability to rapidly model high-resolution interferometric data opens a new window into the complex
117 physical processes that determine the small-scale mass structure of elliptical galaxies.

H	Description	$\Delta \log \mathcal{E}_H$	$\Delta \log P_H$	$\sigma_{\Delta\kappa}$ (%)	f_{H_0} (%)	$ \Delta\mu _{\max}$ (%)
PL	Power-law ellipsoid (PEMD)	$\equiv 0$	-	2.4	+0.6	10.3
PL+FG	PEMD plus field galaxies	-7060	-	6.3	-22.4	14.9
PL+M4	PEMD plus multipoles $m \leq 4$	+4751	-	1.8	-6.2	12.6
PL+PX	PL plus potential corrections	-	$\equiv 0$	$\equiv 0$	$\equiv 0$	$\equiv 0$
PL+FG+PX	PL+FG plus potential corrections	-	-1084	4.2	-24.8	1.2
PL+M4+PX	PL+M4 plus potential corrections	-	-21	1.3	-6.6	1.6

Table 1: **Summary of the main quantitative results for each model parameterization.** We present the Bayes factor $\Delta \log \mathcal{E}_H$ relative to the fiducial model PL. For models that include pixellated potential corrections (PX), we present the maximum log-posterior $\Delta \log P_H$ relative to PL+PX. The RMS fractional difference in convergence, $\sigma_{\Delta\kappa}$, relative to the best model PL+PX, is measured inside a masked region within 17 mas (3 beam widths) of the lensed images (see Figure 1). The fractional difference f_{H_0} in the measurement of H_0 inferred using time-delay cosmography for each model is stated relative to model PL+PX. The maximum change in flux for the brightest part of the source (source component “1”; see Figures 1, 2, and 3) in each model, $|\Delta\mu|_{\max}$, is given relative to the flux-weighted mean magnification. Although the projected surface mass densities depart from the best PL+PX profile by only a few per cent (RMS) within the mask, the effect on inferences made using time-delays or flux ratios can be substantial.

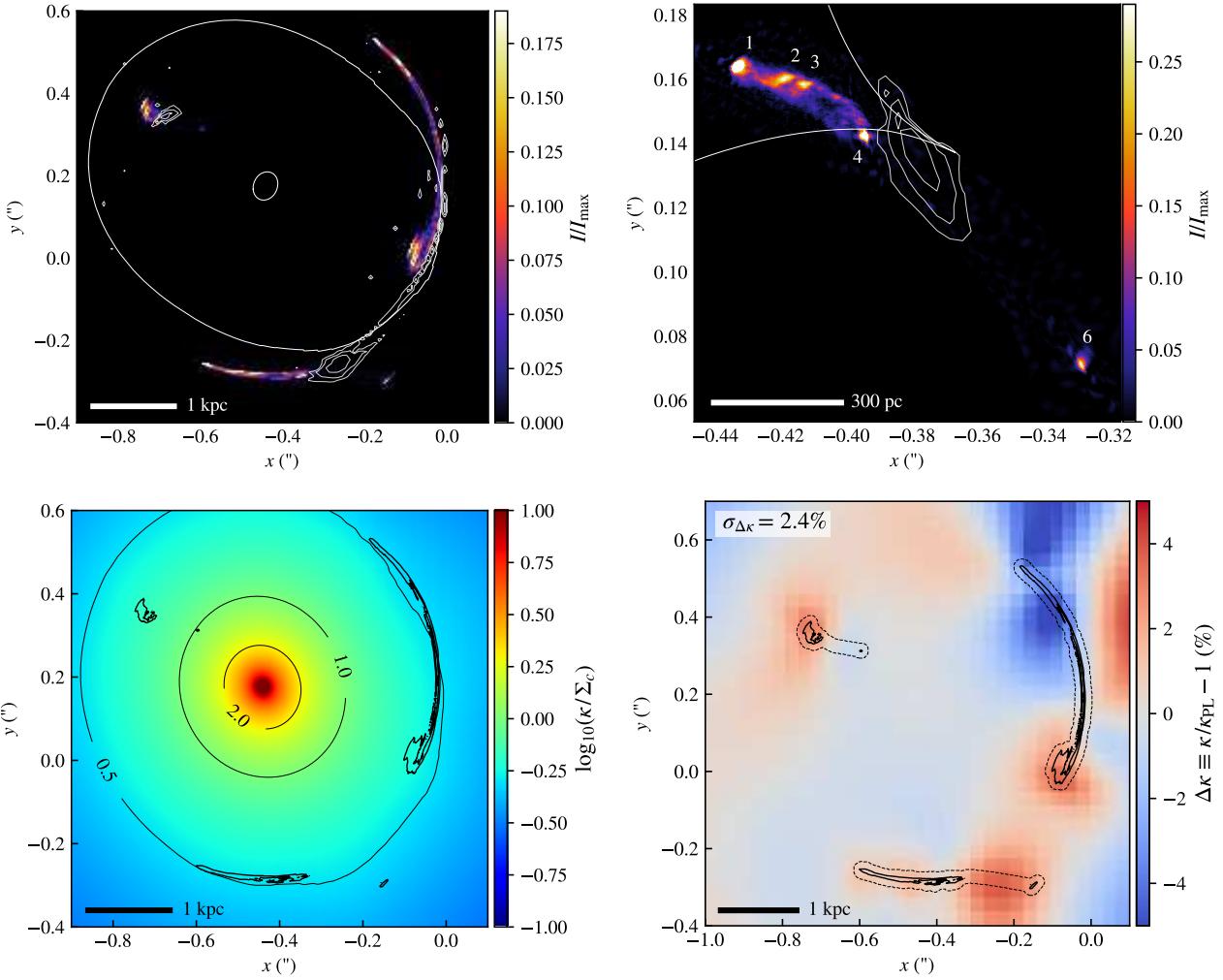


Figure 1: The best-fit lens mass and source surface brightness models. These were obtained using a power-law ellipsoid (PEMD) plus pixellated potential corrections (PL+PX; see Table 1). **Top-left:** The sky model, with critical curves shown in white. **Top-right:** The source surface brightness distribution, with the caustic curve in white, and the brightest source light components labeled following^{6,11,13}. The grey contours in the top panels show the (rest-frame) optical emission from the Keck AO observation of this lens. **Bottom-left:** The total projected surface mass density. The black density isocontour levels are in units of critical density Σ_c , while the colour map is in logarithmic units. **Bottom-right:** The fractional difference between the projected surface mass density and that of the simple PEMD model (PL). The RMS fractional deviation from an elliptical power-law model, computed within the mask (black dashed outlines), is $\sigma_{\Delta\kappa} = 2.4\%$.

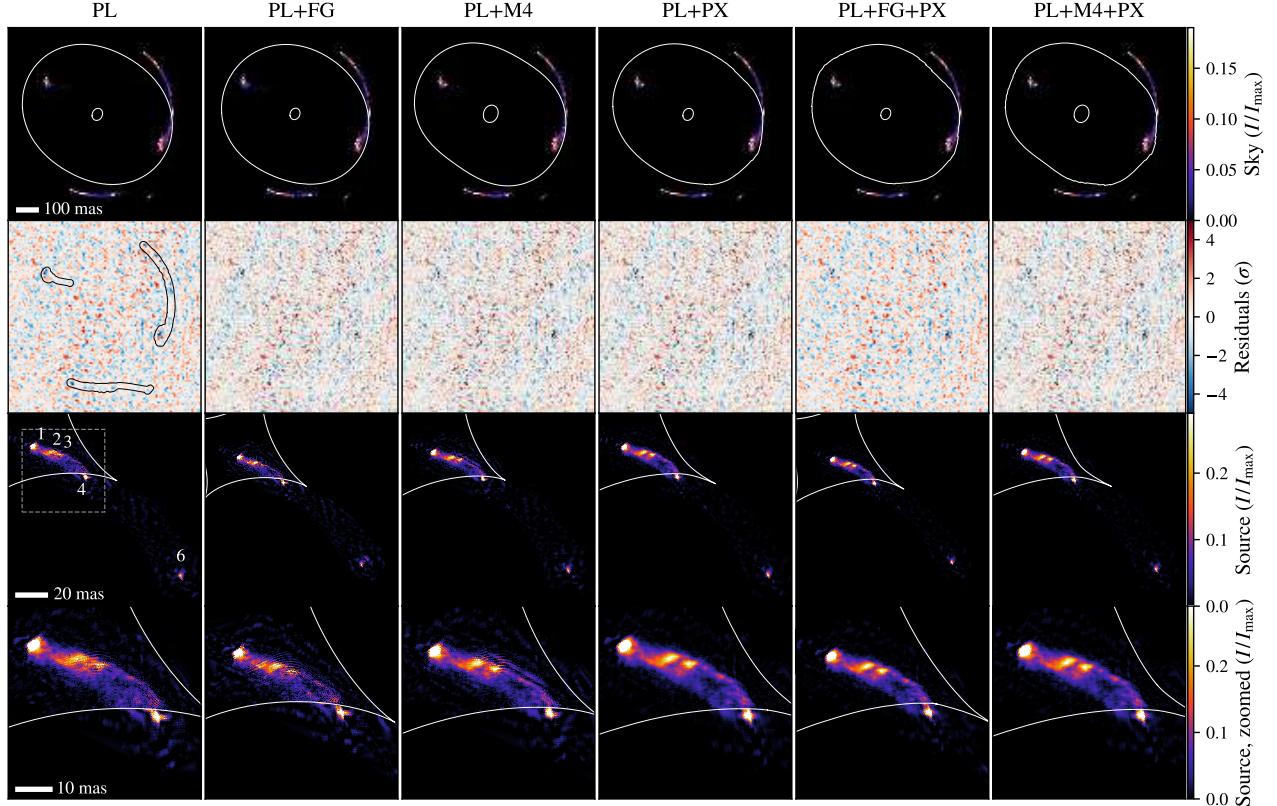


Figure 2: Comparison of maximum a posteriori (MP) source and sky models for each lens mass profile. Colour scales and physical sizes are consistent across each row. **Top row:** The lens-plane sky model, with critical curves in white. **Second row:** An image-plane representation of the residuals, which have been normalized to the noise in the visibility space and then Fourier transformed for visualization purposes. The mask is shown as a thin black outline in the left panel. The top two rows use the same physical extent and scale for each panel. **Third row:** The source surface brightness for each model, which are translated relative to one another depending on the mean deflection of the lens model. In the leftmost panel of the third row, we label the brightest light components of the source following refs.^{6,11,13}. **Bottom row:** A zoomed view of the main source features (with an extent shown by the dashed square in the left-hand panel of the third row) in order to illustrate the improvement in the focusing of the source when using pixellated potential corrections (PX).

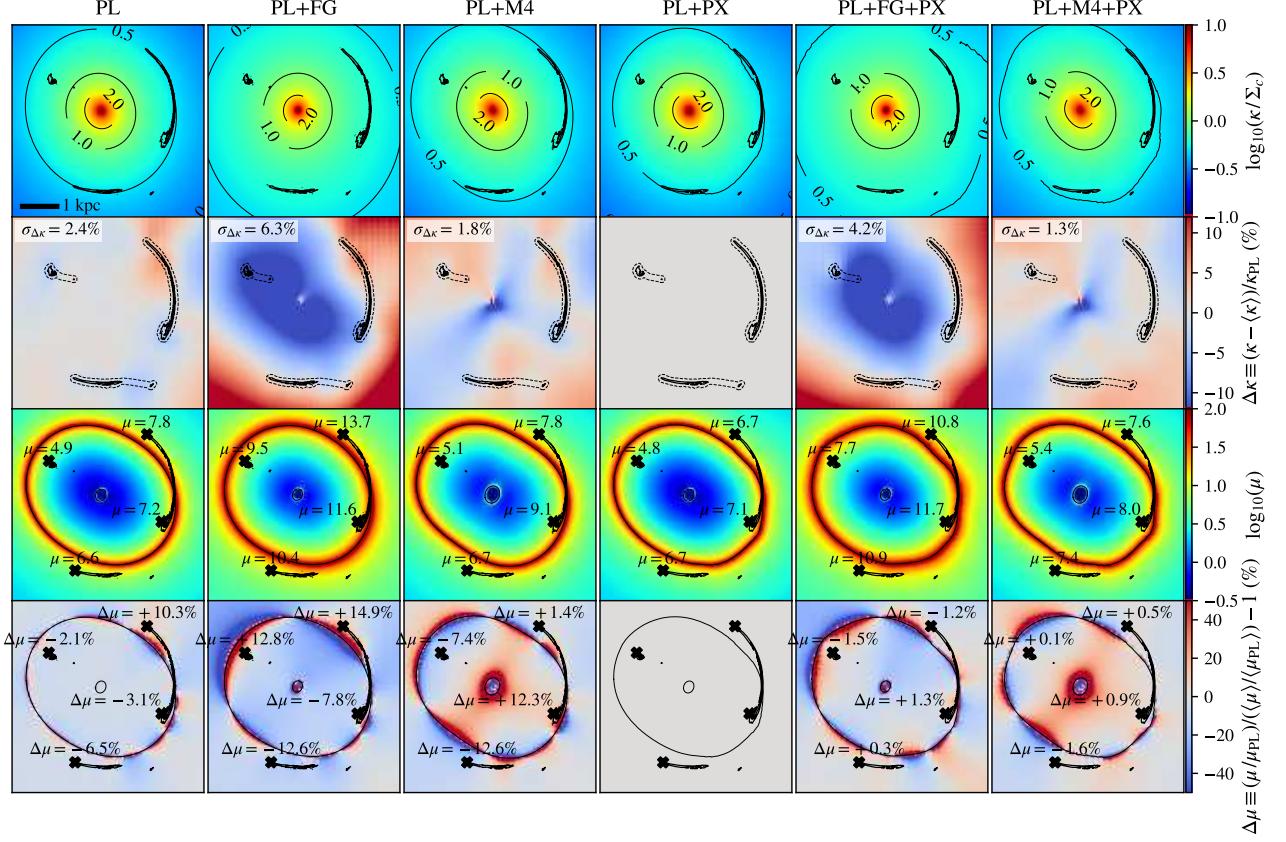


Figure 3: Comparison of convergence and magnification maps for the lens mass profiles. The physical extent is the same for all panels; a scale bar is shown in the top-left panel. The colour scales are consistent within each row. **Top row:** Total convergence maps, in units of critical density Σ_c . **Second row:** Differences from the best model PL+PX as a fraction of the total convergence. **Third row:** Magnification maps, with magnifications of the bright source component 1 (see Figure 2) labeled for each of the four images. **Bottom row:** Differences in magnification relative to model PL+PX. We additionally plot the change in magnification from PL+PX at the location of the brightest image corresponding to source component 1. Relative magnifications are standardized to the flux-weighted mean magnification for each model, in order to remove the mass-sheet degeneracy from the comparison.

¹¹⁸ **Methods**

¹¹⁹ **Radio interferometric data**

¹²⁰ MG J0751+2716 is a strongly-lensed quasar initially observed in the MG-VLA survey²³. Its
¹²¹ discovery was presented along with follow-up observations at systematically higher angular
¹²² resolution with the VLA and MERLIN, as well as a preliminary lens model by ref.¹¹. Ref.¹³
¹²³ proposed an improved lens model incorporating the mass of additional nearby galaxies, which were
¹²⁴ spectroscopically confirmed by ref.¹² to be members of a common group with the lens galaxy at
¹²⁵ $z_l = 0.35$, as well as the source redshift of $z_s = 3.2$, measured by ref.²⁴. Recently, ref.⁶ presented an
¹²⁶ analysis of a global VLBI observation of MG J0751+2716. This is the highest angular resolution
¹²⁷ observation of any gravitational lens system containing extended arcs to date, with sharply resolved
¹²⁸ arcs and images localized to within a few milli-arcseconds.

¹²⁹ **Measurement sets and flagging**

¹³⁰ The observation of MG J0751+2716 was carried out on 21 October 2012 using the global VLBI
¹³¹ array composed of 24 antennas from the European VLBI Network (EVN) and the Very Long
¹³² Baseline Array (VLBA) (project GM070; PI: McKean). The total time on-source was 18.5 hours,
¹³³ with a visibility integration time of 2 s. The total bandwidth is 64 MHz, centred around 1.65 GHz.
¹³⁴ This bandwidth is divided into 256 frequency channels (32 channels in each of 8 spectral windows).
¹³⁵ For detailed information about the calibration and data reduction procedures, we refer to ref.⁶.
¹³⁶ In addition to ref.⁶, we estimate the noise from the data using the procedure described in the next
¹³⁷ section. We then flag all visibilities with noise greater than 1 Jy. Finally, we flagged the
¹³⁸ Effelsberg-Jodrell Bank baseline so as not to allow our inference to be dominated by this single, very
¹³⁹ sensitive baseline. The final calibrated and edited observation used in this work contains 2.5×10^8
¹⁴⁰ unflagged visibilities.

¹⁴¹ **Noise estimation**

¹⁴² The noise column provided in a CASA measurement set is computed from the radiometer equation,
¹⁴³ $\sigma \propto (\Delta\nu\Delta t)^{-\frac{1}{2}}$, which depends on the channel bandwidth $\Delta\nu$ and the integration time Δt . However,

¹⁴⁴ this is a simple theoretical estimate which may not capture other instrumental and atmospheric effects
¹⁴⁵ that vary on timescales shorter than the full observation. We instead measure the noise empirically
¹⁴⁶ from the data as follows.

¹⁴⁷ We first partition the data by baseline, observation epoch, spectral window, and polarization. We
¹⁴⁸ further divide these data into 15-minute blocks, giving ~ 250 visibilities per block. We then subtract
¹⁴⁹ time-adjacent visibilities from one another. Under the assumption that the sweep of each baseline
¹⁵⁰ across the uv plane is sufficiently small between integrations, this difference between neighboring
¹⁵¹ visibilities cancels the sky signal and provides a sample of the noise. We then take the RMS of these
¹⁵² time-differenced samples, corrected by $\sqrt{2}$ to account for the subtraction, to obtain our estimate of the
¹⁵³ noise. Using this differencing scheme, we attempt to utilize as much information from the data as
¹⁵⁴ possible by computing a detailed noise estimate for each visibility.

¹⁵⁵ **Image plane**

¹⁵⁶ The image plane pixel scale and dimensions are chosen to meet two criteria. First, it must be large
¹⁵⁷ enough to contain all of the lensed light that we wish to model; we use a $1.2 \text{ arcsec} \times 1.2 \text{ arcsec}$ field
¹⁵⁸ of view. Second, the pixel scale must be small enough that the dirty beam is properly Nyquist
¹⁵⁹ sampled, which is determined by the uv coverage of the observation. We accomplish this by choosing
¹⁶⁰ an image-plane resolution of $N_{\text{pix}} = 1024^2$ and a pixel size of $1.17 \text{ mas} \times 1.17 \text{ mas}$.

¹⁶¹ In order to aid the inference process, we mask the image plane (see ref.²⁵). This serves both to reduce
¹⁶² the dimensionality of the reconstructed source and to constrain the region of the image plane that is
¹⁶³ allowed to contain emission. It is desirable to make the mask as tight as possible without excluding
¹⁶⁴ any real emission. This can be interpreted as a prior on the model; we refer the reader to ref.⁷ for a
¹⁶⁵ more detailed discussion.

¹⁶⁶ To generate the mask, we use the CLEANed image of MG J0751+2716 by ref.⁶. We first threshold
¹⁶⁷ the image at $5\sigma_{\text{RMS}}$, where $\sigma_{\text{RMS}} = 41 \mu\text{Jy}/\text{beam}$ is the residual RMS off-source noise of the
¹⁶⁸ self-calibrated image. We then pad the resulting region of the image by three beams ($3 \text{ mas} \times 9.8$
¹⁶⁹ mas) in all directions. Because the emission comes from multiple disjoint components, we lastly
¹⁷⁰ connect these components along a path defined by the locations determined by ref.⁶. The resulting
¹⁷¹ mask is shown in the lower-right panel of Figure 1. The image plane (and hence the triangulated

¹⁷² source grid) contains $N_{\text{src}} = 4.5 \times 10^4$ unmasked pixels.

¹⁷³ Keck adaptive optics data

¹⁷⁴ For comparison purposes, we also make use of an infrared ($2.12 \mu\text{m}$) observation of MG J0751+2716
¹⁷⁵ taken with the W. M. Keck-II Telescope as part of the Strong-lensing at High Angular Resolution
¹⁷⁶ Programme (SHARP; reg.²⁶, e.g.). The adaptive optics (AO) system on Keck provides a PSF width
¹⁷⁷ of 65 mas (FWHM) at this wavelength. See ref.¹⁴ for a detailed description of this observation and
¹⁷⁸ the data reduction process. Contours of the image- and source-plane emission are shown in Figure 1
¹⁷⁹ of this work.

¹⁸⁰ Lens mass models

¹⁸¹ The lens mass profile of MG J0751+2716 has been studied in detail by several authors to date. Ref.¹¹
¹⁸² used VLA and MERLIN data (with a maximum resolution of 50 mas) to fit a lens model containing
¹⁸³ ellipsoidal power-law potentials for the main lens galaxy, plus four additional group galaxies. Ref.¹³
¹⁸⁴ use the same data, along with improved optical HST observations, to model the lens and group using
¹⁸⁵ elliptical power-law density profiles. In addition, ref.¹³ include a common group dark matter halo,
¹⁸⁶ claiming an improved fit. Ref.⁶ then provided improved lens modeling for this system using several
¹⁸⁷ bright features in the same global VLBI observation used in this work. They propose both a single
¹⁸⁸ PEMD model, as well as a model that includes additional PEMDs for five nearby group galaxies.

¹⁸⁹ Both of these models were only able to account for the observed image positions to within ~ 5 mas,
¹⁹⁰ which is quite significant considering the astrometric precision of the observation. All of these
¹⁹¹ aforementioned studies fit lens models by first identifying the image positions obtained from a
¹⁹² separate imaging step, then fitting the lens parameters that best reproduce those positions.

¹⁹³ In this section, we enumerate an extended set of lens mass model parameterizations with which we
¹⁹⁴ attempt to improve upon previous modeling attempts using the method of ref.⁷, using the full surface
¹⁹⁵ brightness of the arc and fitting to the native visibility data. We describe them in terms of their
¹⁹⁶ projected surface mass density κ . The models as defined here lie their local reference frame, centred
¹⁹⁷ at the origin and aligned with the x and y axes. In practice, we translate and rotate them to the lens
¹⁹⁸ centre (x_0, y_0) and position angle θ_q . Additionally, we include for all models an external shear

199 defined by its strength Γ and direction θ_Γ .

200 *Power-law ellipsoid*

201 The power-law ellipsoid mass distribution (PEMD, see e.g. ref.²⁷) is a ubiquitous lens mass profile
202 due to its simplicity and ability to fit a wide range of observed lens systems. As this is the simplest
203 model, with only 8 free parameters, we use the PEMD as our fiducial density profile, which we label
204 “PL”. The PEMD has a normalized projected mass density

$$\kappa(x, y) = \frac{\kappa_0 (2 - \frac{\gamma}{2}) q^{\gamma - \frac{3}{2}}}{2 [q^2 x^2 + y^2]^{\frac{\gamma-1}{2}}}, \quad (1)$$

205 where κ_0 is the mass normalization, q is the elliptical axis ratio, and γ is the power-law slope (with
206 $\gamma = 2$ corresponding to an isothermal power-law). In practice we use the FASTELL library²⁸ to
207 compute the deflection angles.

208 *Field galaxies*

209 Ref.²⁹ identify a total of 13 galaxies that are spectroscopically confirmed to be members of the same
210 compact group along with the main lens galaxy. Both ref.¹¹ and ref.¹³ include additional mass
211 components for these galaxies in their lens models. Most recently, ref.⁶ also modeled this system
212 with group galaxy properties inferred as follows: positions, ellipticities, and position angles were
213 measured using archival optical HST data (GO-7495; PI: Falco). Mass normalizations were set
214 relative to the main lens galaxy using their optical magnitudes and scaling relations appropriate to
215 their Hubble types (refs.^{30,31}, e.g.). However, they also find that the inclusion of a global dark matter
216 halo as in ref.¹³ does not afford a well-constrained position or mass, so they do not include it in the
217 model.

218 In order to test whether external differential shear due to tidal forces from neighboring galaxies can
219 improve upon the fiducial PL model, we also test a mass model that includes these field galaxies. We
220 use the same positions, masses, and ellipticities as found by⁶. We plot the total convergence of the
221 group in a 15 arcsec \times 15 arcsec field of view in Figure 5, along with labels for the galaxies
222 consistent with the aforementioned previous works. The free parameters in this mass distribution are

223 those of the main lens galaxy, G3, which is a PEMD. We hereafter label this model as “PL+FG”.

224 *Multipole perturbations*

225 As an extension to the PEMD profile PL, we include multipole-like terms describing angular
226 variation in the mass distribution. This model is meant to encompass generic smooth deviations from
227 the PEMD, which may arise from mergers, tidal forces, and/or baryonic processes. The functional
228 form of the convergence is

$$\kappa_m(r, \theta) = r^{-(\gamma-1)} [a_m \sin(m\theta) + b_m \cos(m\theta)]. \quad (2)$$

229 Here, we express the convergence more naturally in polar coordinates, with r in arcseconds. a_m and
230 b_m together describe the strength and orientation of the multipole perturbation. These coefficients
231 give the deviation from the underlying PEMD in units of the critical density Σ_c at a radius of 1 arcsec
232 from the lens centre. The slope γ is fixed to the slope of the PEMD (Equation 1). The potential and
233 deflection angles are obtained trivially via the Poisson equation.

234 We impose a flat prior of width 0.02 on a_m and b_m . Our choice of prior is motivated by ref.³², who
235 note a typical amplitude of $\kappa_0 \sqrt{a_4^2 + b_4^2} \sim 0.005$ from numerical simulations. The conversion to our
236 units is not exact, as they assume an isothermal ($\gamma = 2$) density slope, but it is sufficient for our
237 purposes. We see in Table 2 that this choice of prior is indeed able to accommodate the multipole
238 amplitudes favored by the data. We include multipole perturbations up to order $m = 4$, labelling this
239 model “M4”. We do note consider multipole components with $m > 4$, as they can correspond to the
240 presence of subhaloes or other fine structure in the lens, which we describe instead with pixellated
241 potential corrections.

242 **Hierarchical Bayesian inference**

243 In radio interferometry, the data take the form of visibilities, which sample Fourier modes of the sky.
244 Hence, our data \mathbf{d} is a vector of N_{vis} complex numbers. The source \mathbf{s} is a vector of length N_{src} , which
245 we represent on an adaptive Delaunay-tessellated grid as in ref.²⁵. The source light is mapped from
246 the source plane to the image plane by the lens operator $\mathbf{L}(\boldsymbol{\eta}_H)$, which has dimensions $N_{\text{src}} \times N_{\text{pix}}$.
247 The image dimension N_{pix} is set by the field of view and the angular resolution of the instrument,

248 while N_{src} is determined by a light mask in the image plane. η_H is the set of parameters describing
 249 the lens mass distribution used to generate \mathbf{L} , where H denotes model parameterization. The
 250 instrumental response is \mathbf{D} , the Fourier transform corresponding to the uv coverage of the
 251 interferometer. Written as a matrix, \mathbf{D} is dense with dimensions $N_{\text{vis}} \times N_{\text{pix}}$. We assume additive
 252 Gaussian noise \mathbf{n} with covariance \mathbf{C}^{-1} .

253 With this notation, our model for the data \mathbf{d} is

$$\mathbf{d} = \mathbf{DL}(\eta_H)\mathbf{s} + \mathbf{n}. \quad (3)$$

254 We jointly infer \mathbf{s} and η_H as follows. We first compute the maximum a posteriori source \mathbf{s}_{MP} for a
 255 given set of lens parameters η_H and source regularization strength λ_s . The regularized least-squares
 256 equation for \mathbf{s}_{MP} is

$$\mathbf{A}\mathbf{s}_{\text{MP}} = (\mathbf{DL})^T \mathbf{C}^{-1} \mathbf{d}, \quad (4)$$

257 where

$$\mathbf{A} \equiv [(\mathbf{DL})^T \mathbf{C}^{-1} \mathbf{DL} + \lambda_s \mathbf{R}_s^T \mathbf{R}_s]. \quad (5)$$

258 The operator \mathbf{R}_s is a discrete gradient operator defined on the Delaunay mesh. The Gaussian source
 259 prior has covariance $\lambda_s \mathbf{R}_s^T \mathbf{R}_s$, which penalizes large gradients in the reconstructed source. We solve
 260 Equation (4) using a preconditioned conjugate gradient solver, where the Fourier operator \mathbf{D} is
 261 implemented using a nonuniform fast Fourier transform (NUFFT). We refer the reader to ref.⁷ for
 262 further details on the method. Marginalizing over all possible source configurations for a given η_H
 263 and λ_s gives the likelihood

$$2 \log P(\mathbf{d} | \eta_H, \lambda_s) = -\chi^2 - \lambda_s \mathbf{s}_{\text{MP}}^T \mathbf{R}_s^T \mathbf{R}_s \mathbf{s}_{\text{MP}} - \log \det \mathbf{A} + \log \det(\lambda_s \mathbf{R}_s^T \mathbf{R}_s) + \log \det(2\pi \mathbf{C}), \quad (6)$$

264 which follows from the Gaussianity of both the noise and source prior. As \mathbf{R}_s and \mathbf{C} are sparse, the
 265 terms containing them are easy to evaluate. Computing $\log \det \mathbf{A}$ is non-trivial; we approximate it
 266 using the preconditioner from the inference on \mathbf{s}_{MP} as described in ref.⁷.

267 The χ^2 term, needed for the likelihood in Equation (6), is

$$\chi^2 = (\mathbf{D}\mathbf{L}\mathbf{s}_{\text{MP}} - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{D}\mathbf{L}\mathbf{s}_{\text{MP}} - \mathbf{d}). \quad (7)$$

We speed its evaluation as follows. We first expand the quadratic form into its individual terms,

$$\chi^2 = \mathbf{s}_{\text{MP}}^T \mathbf{L}^T \mathbf{D}^T \mathbf{C}^{-1} \mathbf{D} \mathbf{L} \mathbf{s}_{\text{MP}} - 2 \mathbf{s}_{\text{MP}}^T \mathbf{L}^T \mathbf{D}^T \mathbf{C}^{-1} \mathbf{d} + \mathbf{d}^T \mathbf{C}^{-1} \mathbf{d}.$$

268 We next observe that $\mathbf{D}^T \mathbf{C}^{-1} \mathbf{d}$ is the naturally-weighted dirty image, which we denote \mathbf{d}_x . Similarly,
269 $\mathbf{D}^T \mathbf{C}^{-1} \mathbf{D}$ performs a convolution with the naturally-weighted dirty beam, which we carry out
270 efficiently using an FFT⁷. We define $\tilde{\mathbf{C}}_x^{-1} \equiv \mathbf{D}^T \mathbf{C}^{-1} \mathbf{D}$, where the tilde indicates that $\tilde{\mathbf{C}}_x^{-1}$ is
271 implemented as a function rather than an explicit dense matrix. The last term, $\mathbf{d}^T \mathbf{C}^{-1} \mathbf{d}$, is a constant
272 that need only be evaluated once. Making these substitutions yields

$$\chi^2 = \mathbf{s}_{\text{MP}}^T \mathbf{L}^T \tilde{\mathbf{C}}_x^{-1} \mathbf{L} \mathbf{s}_{\text{MP}} - 2 \mathbf{s}_{\text{MP}}^T \mathbf{L}^T \mathbf{d}_x + \mathbf{d}^T \mathbf{C}^{-1} \mathbf{d}. \quad (8)$$

273 We have shown here that the χ^2 can be computed entirely in the image and gridded uv -plane bases,
274 without the need to explicitly enter the (extremely high-dimensional) visibility space. No information
275 is lost between the visibility space and the image plane, given that the image plane is sub-Nyquist
276 sampled. After precomputing and storing the dirty image and beam, each evaluation consists of just
277 one forward/backward FFT pair and a few sparse matrix multiplications. This fast method for
278 evaluating the χ^2 is crucial to the feasibility of our analysis, which would otherwise require an
279 expensive de-gridding operation at every likelihood evaluation.

280 The final step of inference is to compare the relative probability of each lens mass parameterization,
281 \mathbf{H} , given the observed data. This is done using the Bayesian evidence, which is computed by
282 marginalizing over the entire parameter space of $\boldsymbol{\eta}_{\mathbf{H}}$, λ_s , and \mathbf{s} :

$$P(\mathbf{d} | \mathbf{H}) = \int P(\mathbf{d} | \boldsymbol{\eta}_{\mathbf{H}}, \lambda_s) P(\boldsymbol{\eta}_{\mathbf{H}}) P(\lambda_s) d\lambda_s d\boldsymbol{\eta}. \quad (9)$$

283 We use a uniform prior $P(\boldsymbol{\eta}_{\mathbf{H}})$ and log-uniform $P(\lambda_s)$. Note that the marginalization over \mathbf{s} has

284 already taken place in Equation 6.

285 This integral has no closed-form solution, so it must be computed numerically. We accomplish this
286 using the MultiNest algorithm³³, which samples the full posterior distribution in the parameter
287 space, while also computing the total evidence. For practical purposes, we express the evidence in
288 logarithmic units using the notation

$$\log \mathcal{E}_H \equiv \log P(\mathbf{d} | H). \quad (10)$$

289 We can then compare models using the difference in log-evidence, $\Delta \log \mathcal{E}$, between the two. In this
290 context, by “different models” we mean different parameterizations of the lens mass distribution. The
291 Bayesian evidence provides us with an objective means to compare the ability of different models to
292 explain the data, while automatically penalizing unnecessarily complex models.

293 **Pixellated potential corrections**

294 In addition to the three fully parametric mass profiles, we also fit models that augment the parametric
295 profiles using pixellated corrections to the lensing potential. We denote these potential corrections by
296 $\delta\psi$, which is a vector that inhabits a 41×41 grid with an extent of 1.4×1.4 parsec, corresponding to
297 a pixel scale of 34 mas, or 174 parsecs when projected to the lens redshift $z_l = 0.35$. This pixel scale
298 was chosen to be fine enough to focus the source, but coarse enough to avoid introducing too many
299 additional degrees of freedom to the model. Our results show that the potential corrections in the
300 preferred model PL+PX vary on scales several times larger than one pixel (Figure 1), even with this
301 relatively coarse grid.

302 We regularize the potential corrections using a Gaussian prior on $\delta\psi$,

$$2 \log P(\delta\psi | \lambda_{\delta\psi}) = -\lambda_{\delta\psi} \delta\psi^T \mathbf{R}_{\delta\psi}^T \mathbf{R}_{\delta\psi} \delta\psi + \log \det(\lambda_{\delta\psi} \mathbf{R}_{\delta\psi}^T \mathbf{R}_{\delta\psi}), \quad (11)$$

303 where $\mathbf{R}_{\delta\psi}$ minimizes gradients in the convergence (surface mass density) of $\delta\psi$. The prior modifies
304 the usual source-marginalized likelihood (Equation 6) as

$$\log P(\mathbf{d} | \boldsymbol{\eta}_H, \lambda_s, \lambda_{\delta\psi}) = \log P(\mathbf{d} | \boldsymbol{\eta}_H, \delta\psi, \lambda_s) + \log P(\delta\psi | \lambda_{\delta\psi}). \quad (12)$$

305 $\log P(\mathbf{d} \mid \boldsymbol{\eta}_H, \delta\psi, \lambda_s)$ is computed via the usual source inversion described above, where $\delta\psi$ is
306 included implicitly via the lens operator \mathbf{L} .

307 Starting from the maximum a posteriori (MP) parametric model, we iteratively compute $\delta\psi$
308 following the procedure of ref.³⁴. To ensure convergence while avoiding local minima, we first
309 perform 100 iterations with an over-regularized source model, where λ_s is set artificially high by a
310 factor of 100. We then do 100 iterations with λ_s high by a factor of 10. Finally, we do 100 iterations
311 with λ_s at its best value obtained from the parametric modeling. This totals 300 iterations, with λ_s
312 slowly cooling towards its optimal value. We perform this procedure for a range of $\lambda_{\delta\psi}$, choosing for
313 our best model the $\lambda_{\delta\psi}$ that maximizes Equation 12.

314 The posterior (Equation 12) is now non-Gaussian in $\delta\psi$ in addition to $\boldsymbol{\eta}_H$, totaling ~ 1700 free
315 parameters, which cannot be efficiently marginalized, either analytically or with `MULTINEST`. This
316 prevents us from computing a formal log-evidence as in Equation 9. In order to compare models that
317 include potential corrections, we instead adopt the maximum log-posterior $\log P$ as our figure of
318 merit, following ref.¹⁰.

319 Comparison of H_0 in time-delay cosmography

320 We compute the fractional bias in H_0 that would be inferred assuming each lens model considered
321 using the relation given by ref.²⁰. This describes how the inferred H_0 relates to the convergence at the
322 Einstein radius of the lens:

$$f_{H_0} = \frac{H_{0,\text{true}}}{H_{0,\text{model}}} - 1 = \frac{1 - \kappa_{E,\text{true}}}{1 - \kappa_{E,\text{model}}} - 1, \quad (13)$$

323 where $\kappa_E \equiv \kappa(R_E)$, the convergence at the Einstein radius of the lens. We use our best model PL+PX
324 as the “truth” for this comparison. We compute R_E and κ_E numerically from the total convergence
325 maps.

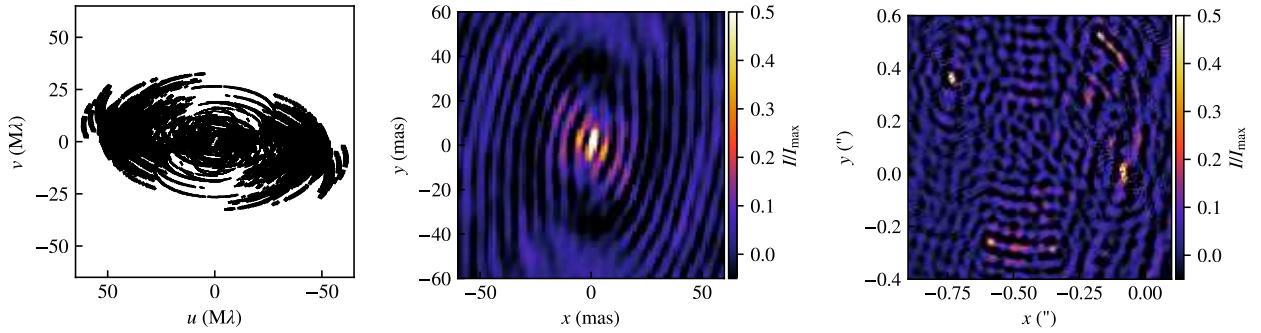


Figure 4: *uv*-coverage (left), naturally-weighted dirty beam (center), and dirty image (right) of the global VLBI observation of MG J0751+2716. The main lobe of the dirty beam is $5.5 \text{ mas} \times 1.8 \text{ mas}$ (FWHM) with a position angle of 9.8 degrees. The (x, y) coordinates of the dirty image are given in arcseconds relative to the phase center.

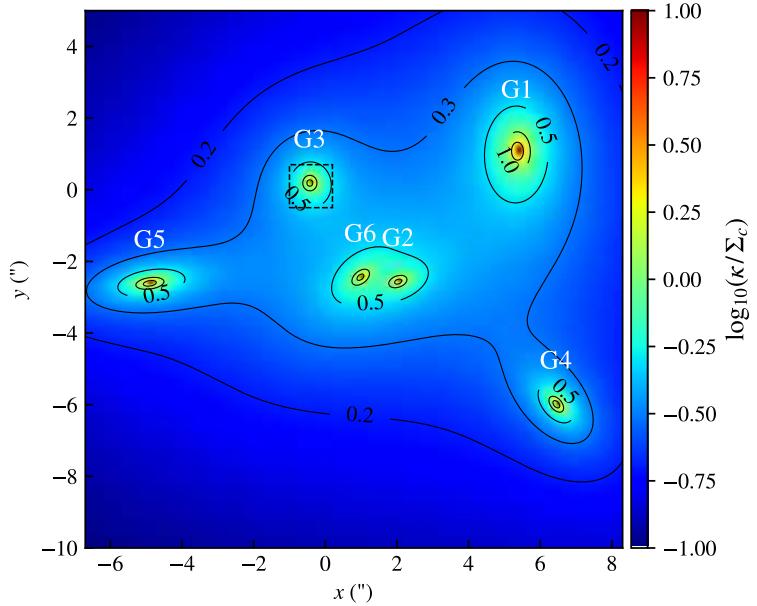


Figure 5: Convergence map showing the projected mass density for model PL+FG. This model includes five additional group galaxies near the lens, which we label following refs.^{6,11,13}. The main lens galaxy, G3, is shown enclosed in the black dashed square, which corresponds to the $1.2 \text{ arcsec} \times 1.2 \text{ arcsec}$ field used in our modelling procedure (see e.g. Figure 2). The units are in critical density (Σ_c) and the field of view is $15 \text{ arcsec} \times 15 \text{ arcsec}$.

Par.	PL		PL+FG		PL+M4	
	Mean $\pm 2\sigma$ (95% CI)	MP	Mean $\pm 2\sigma$ (95% CI)	MP	Mean $\pm 2\sigma$ (95% CI)	MP
κ_0	0.4627 ± 0.0028	0.4624	0.37062 ± 0.00056	0.37059	0.49531 ± 0.00048	0.49535
θ_q	19.17 ± 0.12	19.16	-1.180 ± 0.083	-1.187	26.87 ± 0.15	26.86
q	0.89926 ± 0.00035	0.89924	0.90964 ± 0.00011	0.90961	0.84563 ± 0.00042	0.84567
x_0	-0.44153 ± 0.00003	-0.44152	-0.43739 ± 0.00001	-0.43739	-0.44408 ± 0.00008	-0.44408
y_0	0.17507 ± 0.00002	0.17507	0.17805 ± 0.00002	0.17805	0.17852 ± 0.00006	0.17852
γ	1.8977 ± 0.0047	1.8982	1.9033 ± 0.0012	1.9034	1.84119 ± 0.00081	1.84113
Γ	0.09247 ± 0.00039	0.09252	0.03042 ± 0.00008	0.03043	0.08961 ± 0.00012	0.08962
θ_Γ	73.703 ± 0.050	73.699	57.304 ± 0.032	57.309	81.531 ± 0.052	81.525
a_3	-		-		-0.00082 ± 0.00002	-0.00083
b_3	-		-		0.00166 ± 0.00006	0.00166
a_4	-		-		0.00229 ± 0.00003	0.00229
b_4	-		-		0.00167 ± 0.00005	0.00167

Table 2: **Comparison of lens parameters for the parameteric models.** We quote the mean and 95 per cent confidence interval, as well as the MP parameters.

³²⁶ **Acknowledgements**

³²⁷ This research used SciPy, NumPy and Matplotlib packages for Python^{35,36,37}.

³²⁸ **Data Availability**

³²⁹ The data sets generated during the current study are available from the corresponding author on
³³⁰ reasonable request.

331 References

- 332 [1] Koopmans, L. V. E. & Treu, T. The Structure and Dynamics of Luminous and Dark Matter in
333 the Early-Type Lens Galaxy of 0047-281 at $z = 0.485$. *ApJ* **583**, 606–615 (2003).
334 [astro-ph/0205281](#).
- 335 [2] Sonnenfeld, A. *et al.* Evidence for Dark Matter Contraction and a Salpeter Initial Mass
336 Function in a Massive Early-type Galaxy. *ApJ* **752**, 163 (2012). [1111.4215](#).
- 337 [3] Gavazzi, R. *et al.* The Sloan Lens ACS Survey. IV. The Mass Density Profile of Early-Type
338 Galaxies out to 100 Effective Radii. *ApJ* **667**, 176–190 (2007). [astro-ph/0701589](#).
- 339 [4] Koopmans, L. V. E. *et al.* The Structure and Dynamics of Massive Early-Type Galaxies: On
340 Homology, Isothermality, and Isotropy Inside One Effective Radius. *ApJL* **703**, L51–L54
341 (2009). [0906.1349](#).
- 342 [5] Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T. & Bolton, A. S. Two-dimensional
343 kinematics of SLACS lenses - III. Mass structure and dynamics of early-type lens galaxies
344 beyond $z \simeq 0.1$. *MNRAS* **415**, 2215–2232 (2011). [1102.2261](#).
- 345 [6] Spingola, C. *et al.* SHARP - V. Modelling gravitationally lensed radio arcs imaged with global
346 VLBI observations. *MNRAS* **478**, 4816–4829 (2018). [1807.05566](#).
- 347 [7] Powell, D. *et al.* A novel approach to visibility-space modelling of interferometric gravitational
348 lens observations at high angular resolution. *MNRAS* **501**, 515–530 (2021). [2005.03609](#).
- 349 [8] Hsueh, J. W. *et al.* SHARP - II. Mass structure in strong lenses is not necessarily dark matter
350 substructure: a flux ratio anomaly from an edge-on disc in B1555+375. *MNRAS* **463**, L51–L55
351 (2016). [1601.01671](#).
- 352 [9] Gilman, D., Agnello, A., Treu, T., Keeton, C. R. & Nierenberg, A. M. Strong lensing signatures
353 of luminous structure and substructure in early-type galaxies. *MNRAS* **467**, 3970–3992 (2017).
354 [1610.08525](#).

- 355 [10] Suyu, S. H. *et al.* Dissecting the Gravitational Lens B1608+656. I. Lens Potential
356 Reconstruction. *ApJ* **691**, 277–298 (2009). [0804.2827](#).
- 357 [11] Lehar, J. *et al.* The Gravitationally Lensed Radio Source MG 0751+2716. *AJ* **114**, 48–53
358 (1997). [astro-ph/9702191](#).
- 359 [12] Tonry, J. L. & Kochanek, C. S. Redshifts of the Gravitational Lenses MG 0414+0534 and MG
360 0751+2716. *AJ* **117**, 2034–2038 (1999). [astro-ph/9809063](#).
- 361 [13] Alloin, D., Kneib, J. P., Guilloteau, S. & Bremer, M. Dust and molecular content of the lensed
362 quasar, MG0751+2716, at $z = 3.2$. *A&A* **470**, 53–60 (2007).
- 363 [14] Spingola, C. *et al.* SHARP - VI. Evidence for CO (1-0) molecular gas extended on kpc-scales in
364 AGN star-forming galaxies at high redshift. *MNRAS* **495**, 2387–2407 (2020). [1905.06363](#).
- 365 [15] Vegetti, S. & Koopmans, L. V. E. Bayesian strong gravitational-lens modelling on adaptive
366 grids: objective detection of mass substructure in Galaxies. *MNRAS* **392**, 945–963 (2009).
367 [0805.0201](#).
- 368 [16] Rybak, M., McKean, J. P., Vegetti, S., Andreani, P. & White, S. D. M. ALMA imaging of
369 SDP.81 - I. A pixelated reconstruction of the far-infrared continuum emission. *MNRAS* **451**,
370 L40–L44 (2015). [1503.02025](#).
- 371 [17] Rizzo, F., Vegetti, S., Fraternali, F. & Di Teodoro, E. A novel 3D technique to study the
372 kinematics of lensed galaxies. *MNRAS* **481**, 5606–5629 (2018). [1809.07340](#).
- 373 [18] Ritondale, E. *et al.* Low-mass halo perturbations in strong gravitational lenses at redshift $z \approx 0.5$
374 are consistent with CDM. *MNRAS* **485**, 2179–2193 (2019). [1811.03627](#).
- 375 [19] Xu, D. *et al.* The inner structure of early-type galaxies in the Illustris simulation. *MNRAS* **469**,
376 1824–1848 (2017). [1610.07605](#).
- 377 [20] Kochanek, C. S. Overconstrained gravitational lens models and the Hubble constant. *MNRAS*
378 **493**, 1725–1735 (2020). [1911.05083](#).

- 379 [21] Enzi, W., Vegetti, S., Despali, G., Hsueh, J.-W. & Metcalf, R. B. Systematic errors in strong
380 gravitational lensing reconstructions, a numerical simulation perspective. *MNRAS* **496**,
381 1718–1729 (2020). [1911.02581](#).
- 382 [22] Van de Vyvere, L. *et al.* TDCOSMO. VII. Boxyness/discyness in lensing galaxies :
383 Detectability and impact on H_0 . *arXiv e-prints* arXiv:2112.03932 (2021). [2112.03932](#).
- 384 [23] Lawrence, C. R. *et al.* 5 GHz Radio Structure and Optical Identifications of Sources from the
385 MG Survey. II. Maps and Finding Charts. *ApJS* **61**, 105 (1986).
- 386 [24] Barvainis, R. & Ivison, R. A Submillimeter Survey of Gravitationally Lensed Quasars. *ApJ*
387 **571**, 712–720 (2002). [astro-ph/0201424](#).
- 388 [25] Vegetti, S. & Koopmans, L. V. E. Bayesian strong gravitational-lens modelling on adaptive
389 grids: objective detection of mass substructure in Galaxies. *MNRAS* **392**, 945–963 (2009).
390 [0805.0201](#).
- 391 [26] Lagattuta, D. J. *et al.* SHARP - I. A high-resolution multiband view of the infrared Einstein ring
392 of JVAS B1938+666. *MNRAS* **424**, 2800–2810 (2012). [1206.1681](#).
- 393 [27] Keeton, C. R. A Catalog of Mass Models for Gravitational Lensing. *arXiv e-prints*
394 astro-ph/0102341 (2001). [astro-ph/0102341](#).
- 395 [28] Barkana, R. FASTELL: Fast calculation of a family of elliptical mass gravitational lens models
396 (1999). [9910.003](#).
- 397 [29] Momcheva, I., Williams, K., Keeton, C. & Zabludoff, A. A Spectroscopic Study of the
398 Environments of Gravitational Lens Galaxies. *ApJ* **641**, 169–189 (2006). [astro-ph/0511594](#).
- 399 [30] McKean, J. P. *et al.* CLASS B2108+213: a new wide-separation gravitational lens system.
400 *MNRAS* **356**, 1009–1016 (2005). [astro-ph/0410554](#).
- 401 [31] More, A. *et al.* Probing a massive radio galaxy with gravitational lensing. *MNRAS* **384**,
402 1701–1710 (2008). [0712.1145](#).

- 403 [32] Kochanek, C. S. & Dalal, N. Tests for Substructure in Gravitational Lenses. *ApJ* **610**, 69–79
404 (2004). [astro-ph/0302036](#).
- 405 [33] Feroz, F., Hobson, M. P. & Bridges, M. MULTINEST: an efficient and robust Bayesian inference
406 tool for cosmology and particle physics. *MNRAS* **398**, 1601–1614 (2009). [0809.3437](#).
- 407 [34] Koopmans, L. V. E. Gravitational imaging of cold dark matter substructures. *MNRAS* **363**,
408 1136–1144 (2005). [astro-ph/0501324](#).
- 409 [35] Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.
410 *Nature Methods* **17**, 261–272 (2020).
- 411 [36] Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
- 412 [37] Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* **9**,
413 90–95 (2007).