An unexpected high concentration for the dark substructure in the gravitational lens SDSSJ0946+1006

Quinn Minor [ID],[1,2]★ Sophia Gad-Nasr [ID],[3] Manoj Kaplinghat[3] and Simona Vegetti[4]

[1]*Department of Science, Borough of Manhattan Community College, City University of New York, New York, NY 10007, USA*
[2]*Department of Astrophysics, American Museum of Natural History, New York, NY 10024, USA*
[3]*Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA*
[4]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, D-85740 Garching, Germany*

ABSTRACT

The presence of an invisible substructure has previously been detected in the gravitational lens galaxy SDSSJ0946+1006 through its perturbation of the lensed images. Using flexible models for the main halo and the subhalo perturbation, we demonstrate that the subhalo has an extraordinarily high central density and steep density slope. We robustly infer the subhalo's projected mass within 1 kpc to be $\sim$2–3.7 $\times$ 10^9 M$_\odot$ at >95 per cent CL for all our lens models, while the average log-slope of the subhalo's projected density profile over the radial range 0.75–1.25 kpc is constrained to be steeper than isothermal ($\gamma_{2D} \lesssim -1$). By modeling the subhalo light, we infer a conservative upper bound on its luminosity $L_V < 1.2 \times 10^8$ L$_\odot$ at 95 per cent CL that shows that the perturber is dark matter dominated. We analyse lensing galaxy analogues in the Illustris TNG100-1 simulation over many lines of sight, and find hundreds of subhalos that achieve a mass within 1 kpc $\gtrsim 2 \times 10^9$ M$_\odot$. However, less than 1 per cent of the mock observations yield a log-slope steep enough to be consistent with our lensing models, and they *all* have stellar masses exceeding that allowed by observations by an order of magnitude or more. We conclude that the presence of such a dark highly concentrated subhalo is unexpected in a Lambda cold dark matter universe. While it remains to be determined whether this tension can be reduced by adding more complexity to the primary lens model, it is not significantly alleviated if the perturber is assumed to be a LOS structure, rather than a subhalo.

Key words: gravitational lensing: strong – galaxies: dwarf – dark matter.

1 INTRODUCTION

A key prediction of the cold dark matter (CDM) paradigm is the presence of dark matter subhalos within larger host haloes (Klypin et al. 1999; Moore et al. 1999). This prediction is shared by closely related models like self-interacting dark matter and warm dark matter models. However, these models make distinct predictions for the density profile, concentrations, and shapes of subhalos (Vogelsberger, Zavala & Loeb 2012; Rocha et al. 2013; Lovell et al. 2014). Detecting these subhalos and inferring their properties provides an essential test of dark matter physics.

In order to test the cold dark matter paradigm, one should ideally probe the structure of dark matter haloes that are sufficiently 'dark' such that baryonic physics cannot alter its overall mass or density profile significantly. There are two principle strategies to accomplish this. In the Milky Way, dark matter haloes can be detected by their perturbations of stellar tidal streams (Carlberg 2012; Carlberg & Grillmair 2013; Erkal & Belokurov 2015). This strategy has recently led to the detection of a 10^6-10^8 M$_\odot$ subhalo in the GD1 stellar stream (Bonaca et al. 2019). In distant galaxies that act as strong gravitational lenses, dark matter subhalos [or small field haloes along the line of sight (LOS)] can be detected via their perturbations of highly magnified images (Mao & Schneider 1998; Metcalf & Madau

2001; Koopmans 2005). The presence of dark matter substructure can be established statistically (Dalal & Kochanek 2002; Kochanek & Dalal 2004; Xu et al. 2015; Hezaveh et al. 2016a; Cyr-Racine et al. 2016), or else individual perturbers can be detected for strong enough perturbations (Vegetti & Koopmans 2009); in the latter category, detections have been claimed in four lens systems thus far (Vegetti et al. 2010, 2012; Nierenberg et al. 2014; Hezaveh et al. 2016b). A great many more subhalos are expected to be detected among the avalanche of strong lenses expected from the upcoming Euclid and LSST surveys, after high-resolution follow-up imaging of these lenses.

Several papers have compared the mass function of detected substructures (as well as field haloes along the LOS; cf. Despali & Vegetti 2016; Li et al. 2017) perturbing lensed arcs to the expectation of CDM, showing broad consistency albeit with low statistical significance (Vegetti et al. 2014; Ritondale et al. 2019). More recently, the mass-concentration relation of substructure in quasar lenses has been compared to the expectation of CDM (Gilman et al. 2020). These approaches show great promise, but are still hampered by small number statistics. Additionally, the inferred dark matter halo masses themselves depend on assumptions about the density profiles and tidal radii of the subhalos, complicating the analysis (Minor, Kaplinghat & Li 2017).

Another approach, completely unexplored to date, is to constrain the density profile of individual perturbers and test their inferred properties (e.g. concentration) against the expectation from ΛCDM (Vegetti & Vogelsberger 2014). While the concentration of relatively

★ E-mail: qminor@bmcc.cuny.edu

low-mass perturbers is difficult to infer, in principle the mass and concentration of large perturbers can be constrained. Despite only four individual detections, here we benefit from the fact that the largest perturbations are the most likely to be detected, and thus our initial set of detections is likely to be biased toward relatively massive and/or concentrated perturbers. The question is then, can we constrain the individual concentrations and masses of detected subhalos and test these against the expected mass-concentration relation and its scatter in CDM simulations?

In this paper we constrain, for the first time, the concentration of a dark matter subhalo perturbing a gravitationally lensed arc, originally discovered by Vegetti et al. (2010) using *HST* observations of the lens system SDSSJ0946+1006. By comparing to analogous subhalos in the IllustrisTNG simulations, we will show that its remarkably high mass and concentration are in tension with CDM at the >99 per cent confidence level. Moreover, the lack of a distinct stellar light signal (implying $L_V \lesssim 10^8$ L$_\odot$) implies that star formation and gas physics are unlikely to resolve the discrepancy. This motivates considering modified dark matter physics as a possible explanation.

The paper is organized as follows. In Section 2 we will describe the lens modeling procedure. In Section 3 we describe the analysis carried out on SDSSJ0946+1006. The best-fitting models are compared in Section 3.1, and the subhalo mass/concentration constraints are presented and compared to CDM predictions in Section 3.2. More robust constraints on the density profile – specifically a mass scale and density log-slope – are inferred in Sections 3.3 and 3.4, respectively. In Section 3.5 we infer a conservative upper bound on the stellar luminosity of the subhalo. In Section 4 we compare these results to subhalos of analogous lens galaxies within the IllustrisTNG simulation to investigate whether CDM is capable of producing such high concentrations. The simulated sample is described in Section 4.1, while the comparison of subhalo candidates to the lensing constraints is made in Section 4.2. In Section 4.3 we estimate the likelihood of such a high concentration subhalo generating a similar-sized subhalo perturbation in CDM, and interpret the results in Section 4.4. In Section 5 we discuss whether the tension with CDM can be reduced if the perturber is a field halo along the LOS, rather than a subhalo. In Section 6 we discuss whether dark matter physics, in particular self-interacting dark matter, can explain the high concentration of the subhalo. Finally, we give our conclusions in Section 7.

2 LENS MODELING

The gravitational lens SDSSJ0946+1006 is remarkable in that it is an individual galaxy that is lensing two different source galaxies at different redshifts, $z = 0.609$ and $z \approx 2.4$ (Gavazzi et al. 2008; Sonnenfeld et al. 2012; Collett & Auger 2014), earning it the nickname 'the Jackpot'. (More recently a third faint source at $z \approx 6$ has been found by Collett & Smith 2020). The subhalo discovered in Vegetti et al. (2010) was found perturbing the images from the closest source, at $z = 0.609$. Although stronger constraints on the mass distribution of the lens (and indirectly, the subhalo) may be possible by analyzing both lensed sources; this complicates the analysis and we therefore focus only on the lower redshift source galaxy in this paper.

Our method for modeling the data uses a combination of parameterized surface brightness profiles to represent the $z = 0.609$ source galaxy. When evaluating the likelihood, image pixels are ray traced to the source plane and the following image is convolved with the point spread function (PSF; details on the ray tracing are given at the end of this section). This allows for fast likelihood evaluations and rapid exploration of the parameter space that consists of both

lens and source model parameters, thereby reducing computational cost compared to reconstructing pixellated sources. However, to use elliptical profiles would be too restrictive a prior, for two reasons: (i) actual galaxies rarely (if ever) have perfectly elliptical isophotes; (ii) even if the true source is elliptical, if one is modeling a subhalo with an incorrect mass or concentration, this systematic can often be (at least partially) absorbed into the inferred source galaxy by perturbing the isophotes. Thus, if one does not allow enough freedom in the source galaxy model, the danger arises that the constraints on concentration and mass may appear stronger than they really are.

To allow the requisite freedom in the source galaxy model, we start with a cored Sersic profile, where the core is defined by the replacement $r^2 \rightarrow r^2 + r_c^2$. We then perturb the isophotes to allow for non-elliptical profiles, as follows: first, the unperturbed isophotes are 'generalized ellipses' with a radial coordinate defined by

$$r_0(x, y) = \left(|x - x_0|^{C_0+2} + \left| \frac{y - y_0}{q} \right|^{C_0+2} \right)^{\frac{1}{C_0+2}}, \quad (1)$$

where q is the axial ratio and C_0 is the 'boxiness' parameter, such that $C_0 = 0$ corresponds to a perfectly elliptical profile. We then add Fourier mode perturbations to the isophotes, as follows:

$$r(x, y) = r_0(x, y) \left\{ 1 + \sum_{m=1}^{N} [a_m \cos(m\theta) + b_m \sin(m\theta)] \right\}. \quad (2)$$

From experimentation we find that parameter exploration can become difficult beyond 4–5 Fourier modes, and the $m = 2$ mode is quite degenerate with the axial ratio parameter q. Thus, when modeling the diffuse components of the source galaxy, we include the modes $m = 1, 3, 4, 5, 6$, with sine and cosine terms for each. For high m modes, fluctuations can easily become quite rapid, leading to noisy source solutions. To regulate this, we switch to the scaled amplitudes $\alpha_m = ma_m$, $\beta_m = mb_m$, which are the amplitudes of the azimuthal derivative $dr/d\theta$. By using these scaled amplitudes as free parameters, we can set an upper prior limit on the rate of change of the contours that applies equally to all modes. Our method for perturbing the isophotes is essentially identical to that employed by the GALFIT algorithm for fitting galaxy images (Peng et al. 2010), except that instead of including a phase angle parameter, we use the scaled amplitudes for both sine and cosine terms as free parameters; this ensures that there is no coordinate singularity in the limit of very small amplitudes.

Since the Hubble PSF is relatively undersampled compared to the pixel size, the images produced can be sensitive to how the ray tracing is done, particularly near the critical curve. To mitigate this, we split each image pixel into 2×2 subpixels and ray-trace the centre point of each subpixel to the source plane, assigning it the surface brightness given by the source profile at the position of the ray-traced point. These surface brightness values are then averaged to find the surface brightness of the image pixel. After the ray-tracing is complete, the resulting pixel values are then convolved with the PSF. We find that 2×2 splitting achieves sufficient accuracy without adding too much computational burden; a greater number of subpixels changes the surface brightness values very little while adding significant computational cost.

In a companion paper (Minor et al. 2021) we apply this method of lens modeling to a large number of mock data with a gravitational lens similar to that of SDSSJ0946+1006, but with a wide variety of subhalo perturbations. We find that for sufficiently large perturbations, the mass and concentration can be meaningfully constrained, and are recovered without bias. However, the inferred mass and concentration of the perturbing subhalo are degenerate for

Table 1. Inferred lens model parameters and their priors, and log-evidence and best-fitting χ^2 per pixel. The first 12 parameters describe the primary lens galaxy, while the following five parameters describe the subhalo. The ellipticity components are defined in terms of the axial ratio q and orientation θ (measured east of north) as shown in the first column. The final six parameters are derived parameters that relate to the subhalo, namely the projected mass enclosed within 1 kpc, the average (2D) log-slope of the projected density profile from 0.5 to 1 kpc; the total subhalo mass; the concentration c_{200}; the perturbation radius; and the mass enclosed within the perturbation radius divided by the log-slope α of the primary lens galaxy. The uncertainties are Bayesian credible intervals derived from the 2.5 per cent and 97.5 per cent percentiles of the posterior probability distributions, which are equivalent to 2σ errors if the distribution is Gaussian. Note that we omit the inferred non-linear parameters related to the source galaxy, which are described in Section 3.

	tNFW	tNFWmult	CoreCusp	CoreCuspmult	Prior
R_e (arcsec)	$1.362^{+0.002}_{-0.002}$	$1.347^{+0.005}_{-0.005}$	$1.363^{+0.001}_{-0.001}$	$1.359^{+0.002}_{-0.002}$	[1.3, 1.4]
α	$1.32^{+0.04}_{-0.04}$	$1.37^{+0.05}_{-0.05}$	$1.34^{+0.05}_{-0.05}$	$1.30^{+0.05}_{-0.05}$	[0.5, 1.5]
$e_1 \equiv (1-q)\cos(2\theta)$	$0.057^{+0.010}_{-0.009}$	$0.068^{+0.011}_{-0.011}$	$0.061^{+0.014}_{-0.013}$	$0.056^{+0.010}_{-0.010}$	[−0.15, 0.15]
$e_2 \equiv (1-q)\sin(2\theta)$	$-0.041^{+0.009}_{-0.009}$	$-0.034^{+0.010}_{-0.009}$	$-0.050^{+0.009}_{-0.009}$	$-0.032^{+0.010}_{-0.009}$	[−0.15, 0.15]
x_c (arcsec)	$0.0042^{+0.0024}_{-0.0020}$	$0.0132^{+0.0028}_{-0.0027}$	$0.0041^{+0.0017}_{-0.0019}$	$0.0102^{+0.0026}_{-0.0026}$	[−0.05, 0.05]
y_c (arcsec)	$-0.0016^{+0.0026}_{-0.0020}$	$-0.0162^{+0.0035}_{-0.0032}$	$-0.0018^{+0.0024}_{-0.0024}$	$-0.0100^{+0.0029}_{-0.0028}$	[−0.05, 0.05]
Γ_1	$0.074^{+0.003}_{-0.003}$	$0.072^{+0.003}_{-0.003}$	$0.075^{+0.004}_{-0.004}$	$0.070^{+0.003}_{-0.003}$	[−0.1, 0.1]
Γ_2	$-0.069^{+0.002}_{-0.002}$	$-0.071^{+0.003}_{-0.003}$	$-0.070^{+0.003}_{-0.003}$	$-0.070^{+0.003}_{-0.003}$	[−0.1, 0.1]
A_3	...	$0.0067^{+0.0021}_{-0.0020}$	...	$0.0067^{+0.0019}_{-0.0021}$	[−0.04, 0.04]
B_3	...	$-0.0174^{+0.0030}_{-0.0031}$	...	$-0.0146^{+0.0033}_{-0.0029}$	[−0.04, 0.04]
A_4	...	$0.0174^{+0.0027}_{-0.0026}$	...	$0.0169^{+0.0028}_{-0.0027}$	[−0.04, 0.04]
B_4	...	$-0.0228^{+0.0038}_{-0.0038}$	...	$-0.0195^{+0.0037}_{-0.0040}$	[−0.04, 0.04]
$\log(m_{200}/\mathrm{M}_\odot)$	$9.70^{+0.11}_{-0.09}$	$10.48^{+0.12}_{-0.13}$	...	...	[8, 11.95]
r_s (kpc)	$0.021^{+0.029}_{-0.011}$	$0.844^{+0.260}_{-0.230}$	...	...	[0.01, 5], log
$\log(\kappa_0)$	...	...	$0.184^{+0.113}_{-0.332}$	$0.096^{+0.149}_{-0.239}$	[−2.0, 0.3]
γ_{inner}	...	...	$1.742^{+0.302}_{-0.211}$	$0.766^{+0.379}_{-0.251}$	[0.5, 2.9]
r_t (kpc)	$24.2^{+68.5}_{-19.9}$	$38.7^{+56.8}_{-25.7}$	$0.586^{+0.197}_{-0.056}$	$1.288^{+0.351}_{-0.204}$	[0.1, 100], log
$x_{c,\,\mathrm{sub}}$ (arcsec)	$-0.652^{+0.008}_{-0.010}$	$-0.691^{+0.016}_{-0.018}$	$-0.646^{+0.011}_{-0.011}$	$-0.698^{+0.017}_{-0.020}$	[−1.7, 1.2]
$y_{c,\,\mathrm{sub}}$ (arcsec)	$0.950^{+0.011}_{-0.019}$	$1.009^{+0.026}_{-0.024}$	$0.956^{+0.019}_{-0.014}$	$1.010^{+0.022}_{-0.014}$	[−1.8, 1.5]
$M_{2\mathrm{D}}(1\,\mathrm{kpc})\,(10^9\,\mathrm{M}_\odot)$	$2.50^{+0.29}_{-0.31}$	$3.33^{+0.3}_{-0.26}$	$2.75^{+0.24}_{-0.28}$	$3.31^{+0.31}_{-0.28}$	...
$\gamma_{2\mathrm{D}}(0.75, 1.25\,\mathrm{kpc})$	$-1.98^{+0.05}_{-0.16}$	$-1.27^{+0.11}_{-0.13}$	$-3.27^{+0.27}_{-0.17}$	$-1.79^{+0.23}_{-0.25}$	...
$m_{\mathrm{sub,\,tot}}\,(10^{10}\,\mathrm{M}_\odot)$	$0.469^{+0.202}_{-0.154}$	$2.61^{+1.29}_{-1.06}$	$0.312^{+0.035}_{-0.040}$	$0.700^{+0.131}_{-0.113}$	...
c_{200}	1560^{+1440}_{-859}	$70.5^{+17.8}_{-11.9}$	...	...	...
$r_{\delta c}$ (arcsec)	$0.351^{+0.010}_{-0.013}$	$0.308^{+0.016}_{-0.014}$	$0.355^{+0.009}_{-0.011}$	$0.323^{+0.0160}_{-0.016}$	...
$\frac{M(r_{\delta c})}{\alpha}\,(10^9\,\mathrm{M}_\odot)$	$2.02^{+0.25}_{-0.26}$	$2.69^{+0.31}_{-0.25}$	$2.14^{+0.19}_{-0.22}$	$2.91^{+0.33}_{-0.30}$	...
$\log_{10}\mathcal{E}$	-1537.7	-1504.9	-1537.7	-1510.0	...
χ^2_{bf} per pixel	1.66	1.61	1.66	1.63	...

sufficiently small perturbations. A more direct constraint is provided by inferring the subhalo's projected mass within a characteristic radius and the average slope near this radius as derived parameters, which are also well recovered for perturbations in the mock data (Minor et al. 2021, fig. 6), justifying the approach used in this paper. These derived parameters will be motivated in Sections 3.3 and 3.4.

3 RESULTS OF LENSING ANALYSIS

For our primary modeling runs, the subhalo is modeled with a smoothly truncated Navarro-Frenk-White (tNFW) profile (Baltz, Marshall & Oguri 2009) with outer log-slope of −5 well beyond the truncation radius (for comparison, a different subhalo profile will also be considered in Section 3.4). All parameters of the density profile were varied freely: the mass m_{200}, scale radius r_s, and truncation radius r_t. Note that although we are describing a subhalo, one can formally define m_{200} as the virial mass such a halo would have in the

field in the absence of tidal stripping, thus providing an upper limit on the inferred subhalo mass. (The true mass of the subhalo will also depend on how it is truncated, meaning it is determined by both m_{200} and r_t; note that even for a subhalo with very little tidal stripping, r_t may indicate the virial radius at infall, which would be smaller than the virial radius at the lens redshift.) For ease of parameter exploration, for our mass parameter we use $\log_{10}(m_{200})$ and adopt a uniform prior in this parameter; we vary r_s and r_t but adopt log priors in these parameters. The prior limits on the lens parameters are listed in Table 1.

We considered two different primary models. In model tNFW, the host lens galaxy is modeled with just an elliptical power law profile; in model tNFWmult, we also use an elliptical power law, but in addition we add $m = 3$ and $m = 4$ multipole terms to the projected density of the lens, with the same power law index as the elliptical component, to capture departures from ellipticity for the primary lens. The latter model is inspired by the fact that the lens galaxy light shows evidence of perturbed isophotes, possibly as a result of a recent encounter with

a nearby galaxy (Sonnenfeld et al. 2012; Gavazzi et al. 2008). Gavazzi et al. (2008) showed that approximately half of the mass within the Einstein radius is baryonic, hence the total projected density profile is likely to show marked deviations from ellipticity. This is further supported by the fact that Vegetti et al. (2010) showed that their solution preferred to have smooth corrections to the gravitational potential of the lens (which they modeled as pixellated potential corrections) that indicated departures from ellipticity.

Although using analytic source profiles allows for very fast likelihood evaluations compared to pixellated source inversion, it has the disadvantage of being less automated in the sense that the user must choose how many profiles to employ, whether to add boxiness or Fourier modes, and what prior limits to choose. Thus, achieving a good fit requires at least a few iterations. For our initial run, we simply used one cored Sersic profile with a boxiness parameter and four Fourier modes ($m = 1, 3, 4, 5$). Remarkably, although substantial residuals remained in our best-fitting model, the lens parameters are roughly consistent with Vegetti et al. (2010) and the subhalo was still detected in approximately the same location. This suggests that the existence and location of the subhalo is not strongly dependent on the source prior, although we will see that its location does vary slightly from model to model.

Likewise, doing a pixel inversion from our best-fitting model yielded a source pixel map that is quite similar to the inferred source in Vegetti et al. (2010; Fig. 2b shows our final inversion). From the pixel map, we recognize two bright regions near the centre, as well as a possible peak with high ellipticity that intersects the lower-right part of the caustic; in the next iteration we add Sersic profiles to represent these, with the addition of two Fourier modes ($m = 1, m = 3$). In addition, there is also a small bright-spot near the centre of the caustic, so we add a spherical Sersic profile to capture this. It also became clear that the large diffuse component is not fully captured by a single profile, even with the addition of Fourier modes. In our final iteration, we thus use two cored Sersic profiles to represent the diffuse component, each with boxiness parameters and four Fourier modes ($m = 1, 3, 4, 5$). In addition, there is fainter emission near the upper cusp of the astroid caustic that is clearly perturbed by the subhalo, as well as within the additional caustic generated by the subhalo; finally, there is a large region of faint emission to the left of the large diffuse component. We add cored Sersic profiles to capture these, including two Fourier modes for the largest component. In total, then, we include eight source profiles in our fit, for a grand total of 109 parameters in the tNFW model, and 113 parameters in the tNFWmult model (with the addition of $m = 3, m = 4$ multipoles). The two profiles representing the primary diffuse component each have four Fourier modes, while the three largest additional components include only two ($m = 1, 3$) Fourier modes.

The pixel mask we choose is identical to that of Vegetti et al. (2010), while we adopt the same model PSF as in Vegetti et al. (2010) and Gavazzi et al. (2008). To handle the large number of non-linear parameters in our final run, we use the PolyChord nested sampling algorithm, which evaluates the Bayesian evidence in addition to posterior samples.

3.1 Best-fitting models

In Fig. 1, we plot the data image along with lensed images and residuals (with critical curves) from the best-fitting parameters in models tNFW and tNFWmult, using the analytic source model outlined above. The location of the subhalo is quite evident from its strong perturbation of the critical curve in the upper left. The residuals are down to noise-level in many places, with very little

correlated residuals along the location of the arcs, indicating our analytic model is capturing the primary features of the source galaxy. In Fig. 2 we show the best-fitting analytic source and caustics from the tNFW model (with tNFWmult being quite similar), and compare this to a source reconstruction obtained by a source pixel inversion from the same lens model. For the latter, we use a Cartesian grid with adaptive splitting (Dye & Warren 2005) if the pixel magnification is 4 or greater, and curvature regularization where the regularization parameter is chosen to optimize the Bayesian evidence (Suyu et al. 2006). The colour bars are chosen such that identical colours in either figure represent the same surface brightness values. The comparison shows that the essential components are clearly well reproduced by the analytic source, while fluctuations due to fitting noise are largely absent. Note that the analytic source has a bright region which peaks with a considerably higher surface brightness compared to the pixellated source; this is because such a high peak is discouraged in the pixel version due to the regularization prior that enforces smoothness even in the central regions of the source.

Close inspection of Fig. 1 shows that below the subhalo, there is some residual signal (up to $\approx 4\sigma_{\rm noise}$), which at first glance might seem to indicate a detail of the source that is not being well-captured. We found that placing a faint source blob near (0.4,-0.4) in the source plane can remove approximately half of the residual emission below the subhalo – however, this produces a partner image to the lower-right of the bottom arc, partly outside the mask, and with greater magnification. When the mask is removed, however, we found that there is no discernible signal where the partner image is supposed to be in the *HST* image. For the same reason, close inspection shows these residuals are not entirely eliminated in the pixel reconstructions in Vegetti et al. (2010). We therefore argue that the residual signal below the subhalo is likely due to unsubtracted foreground light from the primary lens galaxy. (This issue will be investigated further in Section 3.5.) The same systematic may also be responsible for other faint correlated residuals in the vicinity of the top arc. We note, however, that the presence of the subhalo is not sensitive to such faint foreground features – indeed we verified that the same subhalo solution is inferred even if the region surrounding the residual signal is masked out. We note that for the brightest pixels near the subhalo, containing the leftmost 'knob' of the top arc, the residuals are consistent with the background noise (i.e. less than $3 \times \sigma_{\rm noise} \approx 0.01$).

The best-fitting parameters and uncertainties are listed in Table 1, along with the Bayesian model evidence and best-fitting χ^2 per pixel values (the latter are calculated after doing a final optimization from the best-fitting point in the chain using Powell's method).[1] For reference, joint posteriors in the subhalo parameters for the tNFWmult model are plotted in Fig. B4 of Appendix B. The Einstein radius of the primary lens is $\approx 1.35-1.36$ arcsec, depending on the model, which is slightly higher than in Vegetti et al. (2010) but significantly lower the 1.4 arcsec value inferred by Collett & Auger (2014); likewise, the latter infer a slope that is near-isothermal ($\alpha \approx 1.03$), compared to ≈ 1.3 in our analysis. This may be explained

---

[1]The χ^2 values in the table are from the analytic source reconstruction (e.g. Fig. 2a), not from the source pixel reconstruction shown in Fig. 2(b). While the latter produces χ^2/pixel <1 due to fitting of noise and inclusion of additional fine source structure in highly magnified regions, we only quote the former since the actual fitting procedure involved the analytic reconstruction. We note, however, that the subhalo is in a relatively demagnified region within the critical curve, as evidenced by the large caustic it produces in the source plane, and hence is not sensitive to fine source structure. For this reason, we found that the subhalo constraints are essentially unchanged even if we include only the first few Sersic components.
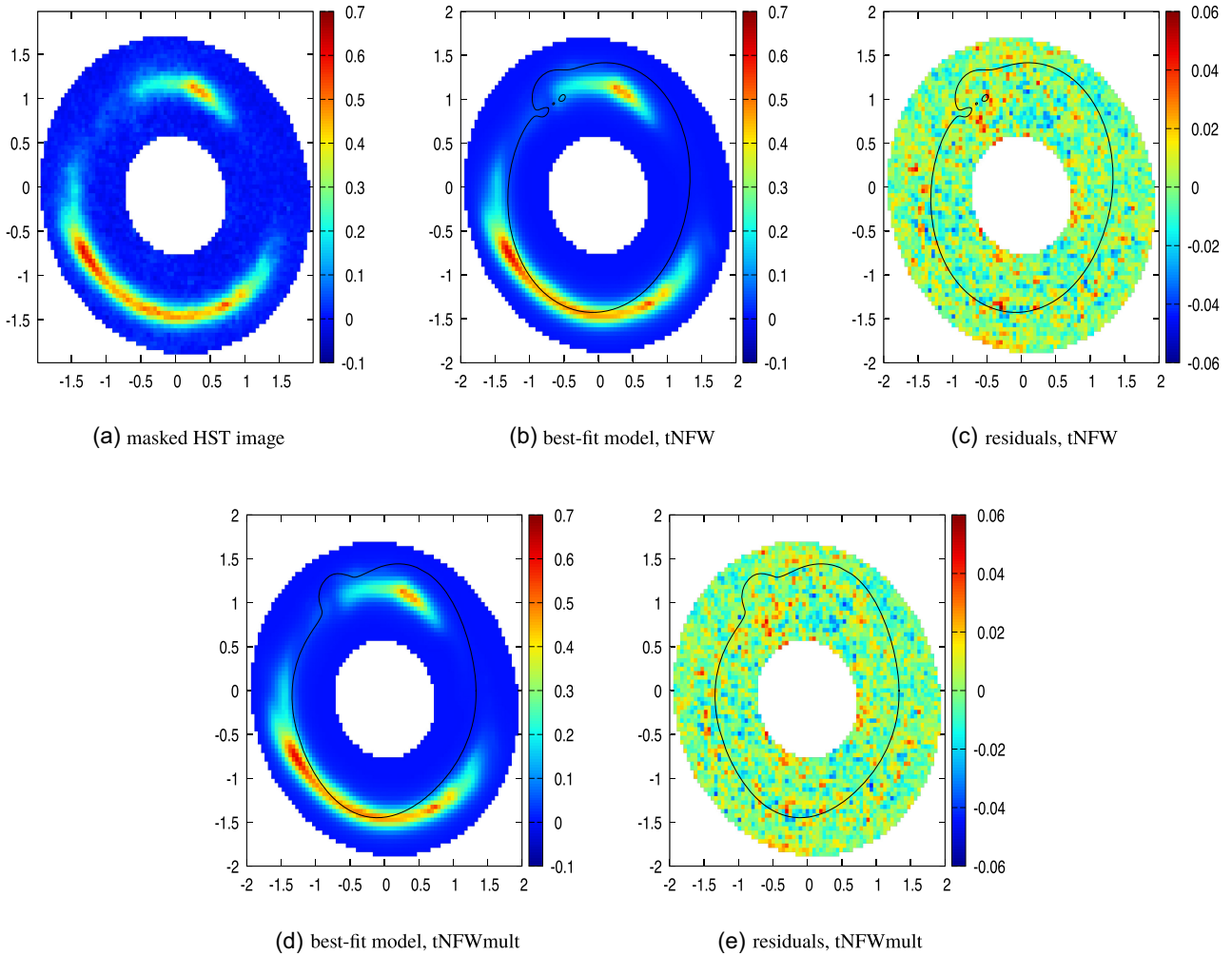
Figure 1. *HST* image along with our best-fitting lensing reconstructions and residuals using a truncated NFW (tNFW) profile for the subhalo. In (a) we show the foreground-subtracted *HST* F814W image from Gavazzi et al. (2008). The best-fitting model lensed images and residuals for the 'tNFW' model with elliptical host galaxy are shown in (b) and (c). Figures (d) and (e) show the same for our 'tNFWmult' model where multipoles are added to the host galaxy's projected density; this is the model most favoured by the Bayesian evidence. The black curves are the critical curves of the lens mapping that show the strong perturbation by the subhalo in the upper left.

by the covariance between the subhalo mass and the Einstein radius, as seen in Fig. B4: as the subhalo mass is reduced, the Einstein radius must be increased. Hence, in the limit of zero subhalo mass, a significantly larger Einstein radius would be required to fit the data. Likewise, our posteriors show that a larger Einstein radius is correlated with a shallower slope, which may explain the difference in the inferred slope parameter.

Between the different subhalo models, the Bayesian evidence strongly prefers the tNFWmult model, with an evidence ratio of nearly 30 orders of magnitude. Indeed, close inspection of Fig. 1 reveals that certain residuals are improved in the tNFWmult model; this should be taken with some caution, however, since it is possible (in principle) that lending more freedom to the source model might allow tNFW to achieve nearly as a good a fit as in tNFWmult. Nevertheless, there are certain residuals that are also present in the best-fitting models in Vegetti et al. (2010) that used pixellated sources, and at least a few of these residuals (notably above the centre region of the bottom arc) do seem to be somewhat improved by the shifting of critical curves achieved in the tNFWmult model.

To get a better sense of how the solutions differ, in Fig. 3 we show the projected density of each best-fitting solution by plotting the log(convergence) values at each pixel within the mask. Note that in tNFWmult the contours are notably boxy as a result of the high multipole amplitudes A_4, B_4, whereas the subhalo is less concentrated (although more massive, as we will see shortly). By comparison, the galaxy isophotes do show departures from ellipticity, but not nearly as boxy as the density contours in tNFWmult. It should be emphasized, however, that the data are not sensitive to the density contours everywhere; there are no images on the right-hand side of the mask, for example, and thus no direct information on the shape of the contours there. Thus, while this solution may indicate departures from ellipticity in the vicinity of the images, it does not necessarily imply the *complete* contours have the boxy shape indicated by Fig. 3. Repeating our analysis with a pixellated source may shed further light on the question of whether the data prefers such a strong departure from ellipticity. The important takeaway here is that the way we model the shape of the primary galaxy's contours can influence the properties of the inferred subhalo, a feature we will discuss further in Section 3.3.
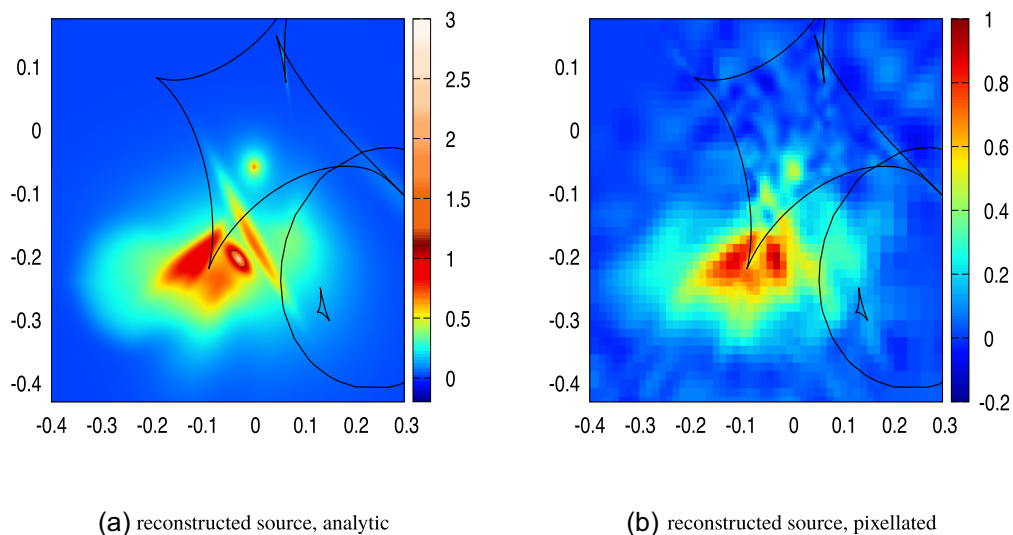
(a) reconstructed source, analytic

(b) reconstructed source, pixellated

Figure 2. Reconstructed source galaxy, analytic versus a pixel inversion from the best-fitting tNFW model with elliptical host galaxy. The analytic source contains a region with relatively high surface brightness; however, the colour scale is chosen so that the colours in each figure correspond to approximately the same surface brightness values. Black curves show the caustic curves of the lens mapping; note the large caustic generated by the subhalo in the lower left.

(a) tNFW model
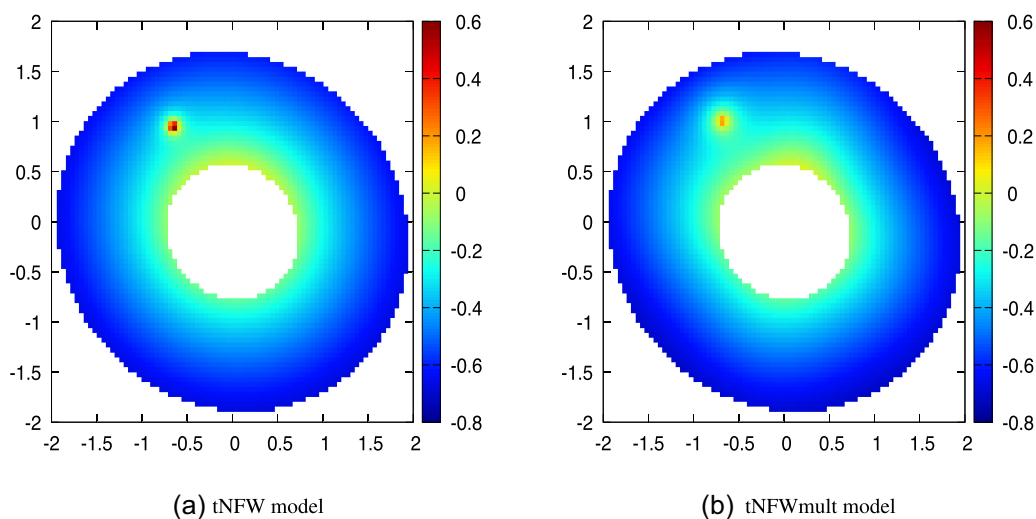
(b) tNFWmult model

Figure 3. Projected density of the lens in each model, plotted as log(convergence) values at each image pixel within the mask. Note the contours are notably boxy in the tNFWmult model, which includes multipole terms in the projected density to capture departures from ellipticity and is preferred by the Bayesian evidence.

3.2 Constraints on mass and concentration of the subhalo

To compare the inferred concentration of the subhalo to ΛCDM constraints, we define a derived parameter to be the halo concentration $c_{200} \equiv r_s/r_{200}$, where r_{200} is the approximate virial radius the halo would have if it were in the field where tidal stripping is absent. Joint posteriors in m_{200} and c_{200} are plotted in Fig. 4, where the red and blue shaded regions refer to model tNFW (the elliptical model) and tNFWmult (which has added multipoles), respectively. Note that in the tNFW model, a mass of $\sim 5 \times 10^9$ M$_\odot$ is preferred with a minimum concentration of several hundred, whereas in tNFWmult, a mass of $\sim 3 \times 10^{10}$ M$_\odot$ is preferred with minimum concentration of 60. For an initial comparison we plot the mass-concentration relation of field haloes from Dutton & Macciò (2014; black solid line) along with the 2σ scatter in concentration (black dashed lines). As the figure suggests, we find the inferred concentration in the

tNFWmult model exceeds the median concentration for CDM field haloes by at least $\approx 5\sigma$, where σ is the posterior uncertainty in the concentration.

However, comparing to field haloes at the lens redshift is insufficient, since it is well known that subhalos are denser than similar mass haloes in the field, for two reasons: (i) subhalos fell in earlier when the critical density of the Universe was higher; (ii) most subhalos have endured some tidal stripping that preferentially removes more mass from the outer regions and renders the subhalo denser overall compared to unstripped haloes of similar mass. To directly test against CDM predictions, we turn to cosmological simulations and examine the subhalo population in simulated galaxies analogous to the lensing galaxy in J0946+1006, to evaluate whether our subhalo is truly an outlier that cannot be explained with CDM. We will do so using IllustrisTNG in Section 4.
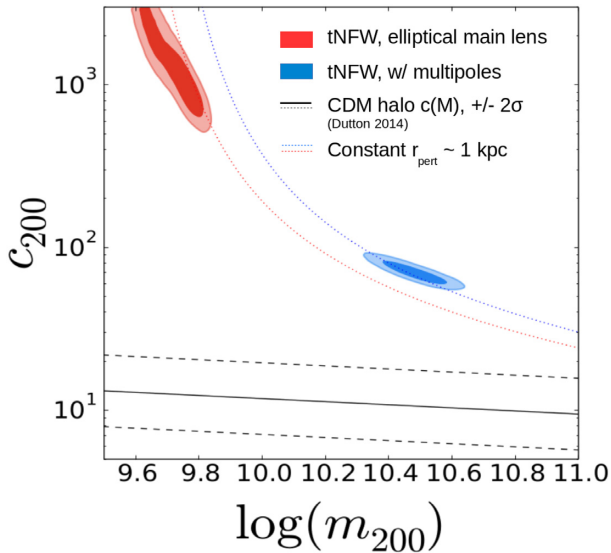
Figure 4. Posteriors in $\log(m_{200})$ versus concentration c_{200} for the two fiducial models. For each model, the dark/light shaded regions enclose 68 per cent and 95 per cent of the total probability, respectively. Black solid curve denotes the median halo concentration-mass relation, along with $\pm 2\sigma$ scatter (black dashed curves). Curves of constant perturbation radius are shown (blue/red curves), generated by starting from the best-fitting models, varying m_{200}, and finding the concentration that reproduces the mass within the perturbation radius (approximately 1 kpc).

3.3 Interpretation in terms of the subhalo's perturbation radius and estimated robust mass

It is evident from Fig. 4 that the model used for the host galaxy can influence the inferred properties of the subhalo significantly: when the primary lens galaxy is allowed to have non-elliptical contours by adding multipoles, the subhalo prefers to be more massive but less concentrated. The natural question arises: if the primary lens model is refined further − perhaps by a multicomponent model or allowing for twist of isodensity contours − might the inferred subhalo have a sufficiently low concentration to be consistent with CDM?

To address this question, it is essential to have an understanding of what is actually being constrained about the subhalo. In Minor et al. (2017) it was shown that for perturbing subhalos, one can robustly infer the subhalo's projected mass enclosed within its *perturbation radius* $r_{\delta c}$, defined as the distance from the subhalo centre to the point on the critical curve that is being perturbed the most. Provided the detection is of high significance, this mass estimate is robust to changes in the subhalo's assumed density profile and tidal radius (unlike the total subhalo mass which can be quite sensitive to these systematics). While approximate formulas can be derived for the perturbation radius $r_{\delta c}$ under the assumption of an isothermal (or truncated isothermal) profile, in the case of an NFW profile $r_{\delta c}$ must be found numerically via a root finding algorithm. For example, in model tNFW, the best-fitting model recovers $r_{\delta c} \approx 0.36$ arcsec ≈ 1.3 kpc, whereas Model tNFWmult gives $r_{\delta c} \approx 0.31$ arcsec ≈ 1.1 kpc. The primary reason for the difference is that the position of the subhalo differs slightly in each model, the difference being comparable to a single pixel length (0.05 arcsec). In order to compare different models, we will evaluate the projected mass within 1 kpc as being robust to good approximation. The inferred robust mass for tNFW is $M_{2D}(1\ \text{kpc}) = (2.5 \pm 0.3) \times 10^9\ \text{M}_\odot$, while model

tNFWmult has $M_{2D}(1\ \text{kpc}) = (3.3 \pm 0.3) \times 10^9\ \text{M}_\odot$, where the uncertainties give the 95 per cent credible intervals.[2]

To verify that the subhalo's mass enclosed within 1 kpc is being robustly inferred in our models, we start with our best-fitting models and vary the subhalo mass m_{200} over a table of values; for each m_{200}, we solve for the required concentration parameter c_{200} to keep the perturbation radius $r_{\delta c}$ (and with it, the robust mass) constant. The corresponding curves are plotted for Models 1 and 2 in Fig. 4 (red and blue curves, respectively). Note that the posteriors follow essentially the same trend as the curves do, indicating that the mass within 1 kpc is being kept approximately constant over the parameter space covered by the posterior.

For any lens model that achieves a good fit, the inferred position of the subhalo (and hence its perturbation radius) would seem unlikely to differ by more than the PSF full width at half-maximum, approximately two pixel lengths (~ 0.1 arcsec). Thus it would appear reasonable to expect that any lensing solution with an equally strong fit will produce a posterior that lies close to the band swept out by the blue/red curves in Fig. 4. If true, this implies that if a solution was found that allowed for a low enough concentration ($\lesssim 20$) to be consistent with CDM, the subhalo's m_{200} would have to be well above $10^{11}\ \text{M}_\odot$. With such a high mass, the subhalo's stellar component would likely be visible. For example, from fig. 10 of Behroozi, Wechsler & Conroy (2013) the median stellar mass in a galaxy with halo mass $2 \times 10^{11}\ \text{M}_\odot$ is estimated to be $M_* \sim 1.6 \times 10^9\ \text{M}_\odot$, reaching down to $M_* \sim 1.2 \times 10^9\ \text{M}_\odot$ for the 16th percentile of galaxies. Since the subhalo is conservatively estimated to have a luminosity no greater than $L_V \approx 1.2 \times 10^8\ \text{L}_\odot$ (see Section 3.5), this would require a V-band stellar mass-to-light ratio greater than $10\ \text{M}_\odot/\text{L}_\odot$, which would be extraordinarily high. The requirement becomes even more stringent under the assumption that the perturber is a halo along the LOS to the lens, as we will demonstrate in Section 5.

Nevertheless, although the inferred subhalo mass within 1 kpc is robust to changes in the subhalo's density profile, it is not *exactly* robust to changes in the primary lens galaxy model, as evident in the slightly different bands followed by the posteriors in Fig. 4. Whether a more sophisticated primary lens model can produce a low enough mass and concentration, sufficiently below the red band in Fig. 4, remains an open question. We will return to these arguments in Section 4.4 in the context of a more rigorous comparison to CDM subhalos using the Illustris TNG100-1 simulation.

3.4 Characterizing the constraints on the subhalo's density profile

To get a sense of how well the density profile is being constrained, we redo the analysis with a different density profile for the subhalo, this time the 'cuspy halo model' of Muñoz, Kochanek & Keeton

---

[2]More precisely, the robust quantity is $M_{2D}(r_{\delta c})/\alpha$, where α is the slope of the host halo's density profile at the position of the subhalo (as discussed in Minor et al. 2017). This implies that the inferred subhalo mass can be biased if the inferred α is biased. However, among our four lens models we fit, the inferred slope α are all consistent with each other within 90 per cent CL, suggesting the inferred slope is not sensitive to bias due the subhalo model being fit. If, however, the true density slope is shallower than our lens models are finding, the inferred subhalo mass would be reduced. For example, if the slope is close to isothermal ($\alpha \approx 1$), as found in Collett & Auger (2014), this would reduce our subhalo masses by roughly 25 per cent. Modeling both the subhalo *and* the second lensed source with a sufficiently flexible host galaxy model may shed light on whether the density slope is biased here or not.
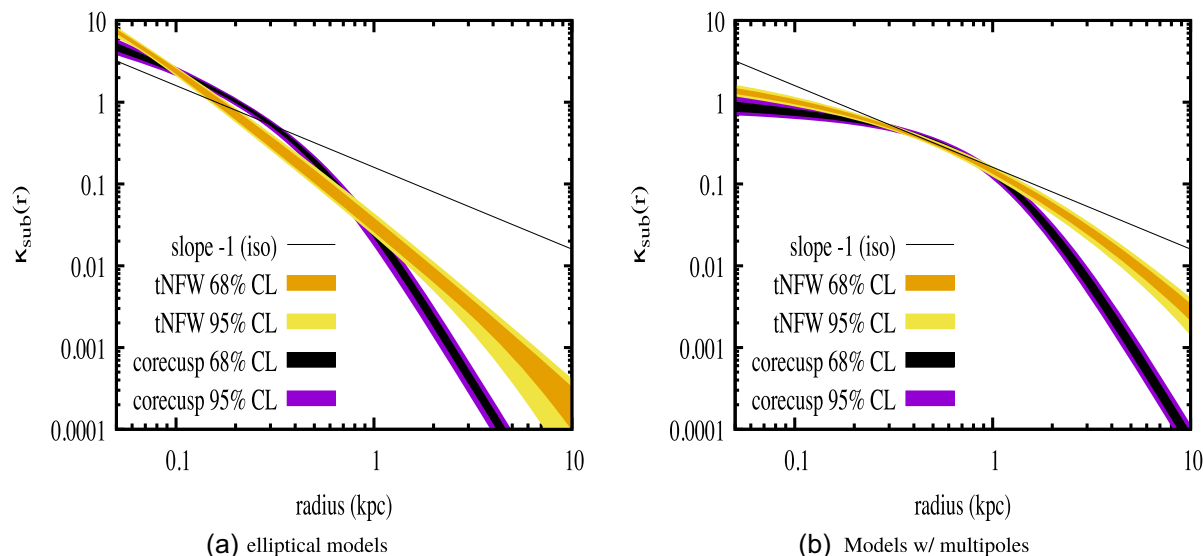
Figure 5. Projected density profiles, for two different subhalo models (tNFW and CoreCusp). Figure (a) shows the models with an elliptical primary lens, while figure (b) shows the models where multpoles are added to the primary lens galaxy to capture deviations from ellipticity. The inferred average log-slope from $r = 0.3$ kpc to $r = 1$ kpc is similar regardless of the subhalo profile used.

[2001; this is also the 'CoreCusp' model in Andrade et al. (2019) with core size set to zero]. In this model, the parameters are the inner and outer slopes and the break radius a. We fix the outer slope to -5 to be consistent with our tNFW runs, and vary the inner slope γ and break radius a freely. (Note that in the special case $\gamma = 2$, the profile is quite similar to the Pseudo-Jaffe profile used in Vegetti et al. (2010) except with outer slope -5 instead of -4.) As with the tNFW fits, we run two versions: one with an elliptical host galaxy (simply labeled 'CoreCusp'), and one with multipoles (labeled 'CoreCuspmult'). The inferred parameter values and uncertainties are given in Table 1. In Fig. 5 we plot the inferred projected density profiles by showing the 68 per cent and 95 per cent probability region of the projected density calculated over a large number of radial bins. Fig. 5(a) shows the density profiles for the CoreCusp and tNFW models with elliptical host, while Fig. 5(b) shows the corresponding profiles with multipoles added. Note that in either case, the inferred density and its log-slope within the interval between approximately 0.5 and 1 kpc are approximately consistent with each other; in the case with multipoles, the log-slopes are both approximately isothermal over this interval, although they steepen as they reach 1 kpc.

For the purpose of comparing to the IllustrisTNG results in Section 4, we define a derived parameter $\gamma_{2D, avg}$ to be the average log-slope from 0.75 to 1.25 kpc. In the tNFW model, $\gamma_{2D, avg}$ is largely determined by the scale radius or concentration parameter. Hence, we see that, to first approximation, both $M_{2D}(1 \text{ kpc})$ and $\gamma_{2D, avg}$ are determined robustly regardless of the profile, which are useful physical quantities to test against subhalos in cosmological simulations. Note that this does *not* mean that $\gamma_{2D, avg}$ is coming out exactly the same in tNFW versus tNFWmult. In Fig. 6(a) we plot the tNFW models with and without the multipoles for direct comparison. Note the slopes are significantly shallower in the models that include multipoles. This is directly related to the fact that tNFWmult allows for lower concentrations compared to tNFW. In Fig. 6(b) we plot the corresponding projected mass profiles; here it is clear that the masses are roughly similar near 1 kpc. Since the perturbation radius is approximately 1 kpc, this is consistent with the conclusion of Minor et al. (2017) that the subhalo's projected mass within this

radius is a robust measurement. (In this case the masses become equal around 0.7 kpc; this is because the position of the subhalo is not exactly the same in either case, differing by about a pixel length, which is discussed in Appendix B).

3.5 Revisiting the upper bound on the subhalo's stellar luminosity

In Vegetti et al. (2010), an upper limit on the subhalo's luminosity $L \lesssim 5 \times 10^6 \, L_\odot$ is derived based on the residuals after subtracting their best-fitting lens model. However, searching for a luminous signal in the residuals carries the risk that the lensed images from the model may have absorbed part of the subhalo's light and subtracted it out; this is especially true for pixellated sources that have a fair amount of freedom to absorb local fluctuations in surface brightness. In this section we revisit this bound and consider whether foreground light may be disguised by the lensed image near the subhalo.

One way to check whether the subhalo's foreground light may be present is to test the lens model in different bands. In Sonnenfeld et al. (2012) this lens is modeled in several different *HST* bands (without including a substructure). For comparison, in Fig. 7 we plot their image using the F336W filter with the foreground galaxy subtracted out, side-by-side with the foreground-subtracted F814W image from Gavazzi et al. (2008) used in our lens modeling. In the shorter-wavelength F336W band there is a much smaller signal from foreground light, and indeed there are clearly fewer residuals from the foreground subtraction. In addition, the bright 'knob' near the subhalo in the F814W image is barely visible in the F336W image. To test whether this feature may have a component of foreground light, we plot seven points in the vicinity of the subhalo (black points with varous markers) and show their partner images on the bottom arc using our best-fitting tNFWmult model. Note that although the surface brightness is dramatically reduced very close to the subhalo (filled circle, filled square, open circle), this is also true for the partner images in the lower right. This suggests that the variation in surface brightness is due to the source galaxy itself, rather than any disguised foreground light near the subhalo. This is a useful consistency check for our lens model, and suggests that if the subhalo's stellar light is
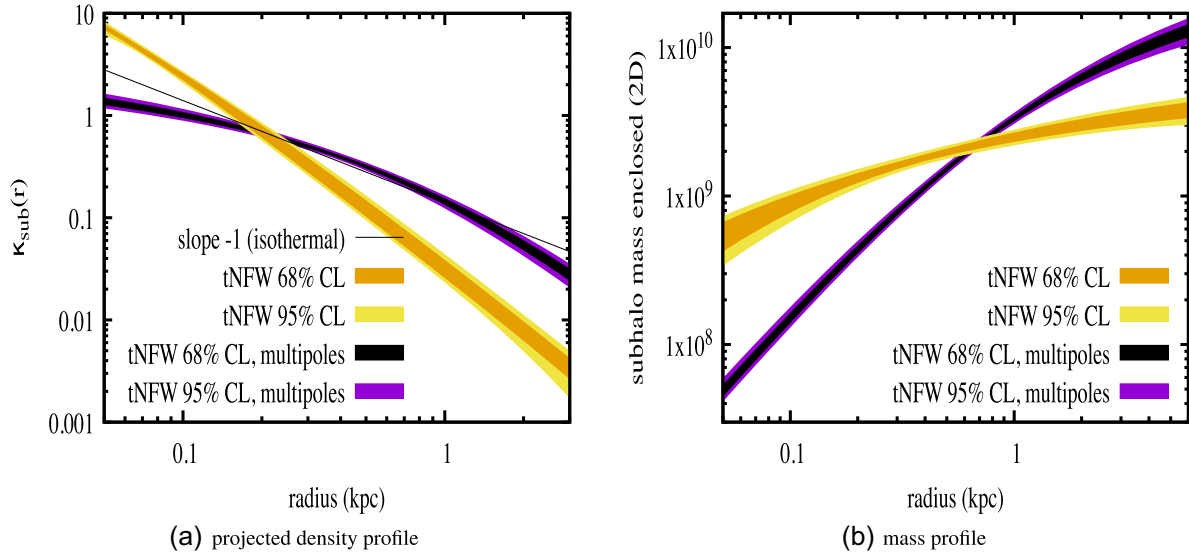
Figure 6. Projected density and mass profiles, for tNFW models with versus without multipoles added to the primary lens galaxy's projected density. Note from figure (b) that the solutions have similar projected masses around 1 kpc, even though the profile shape and the total subhalo mass is rather different in either case.
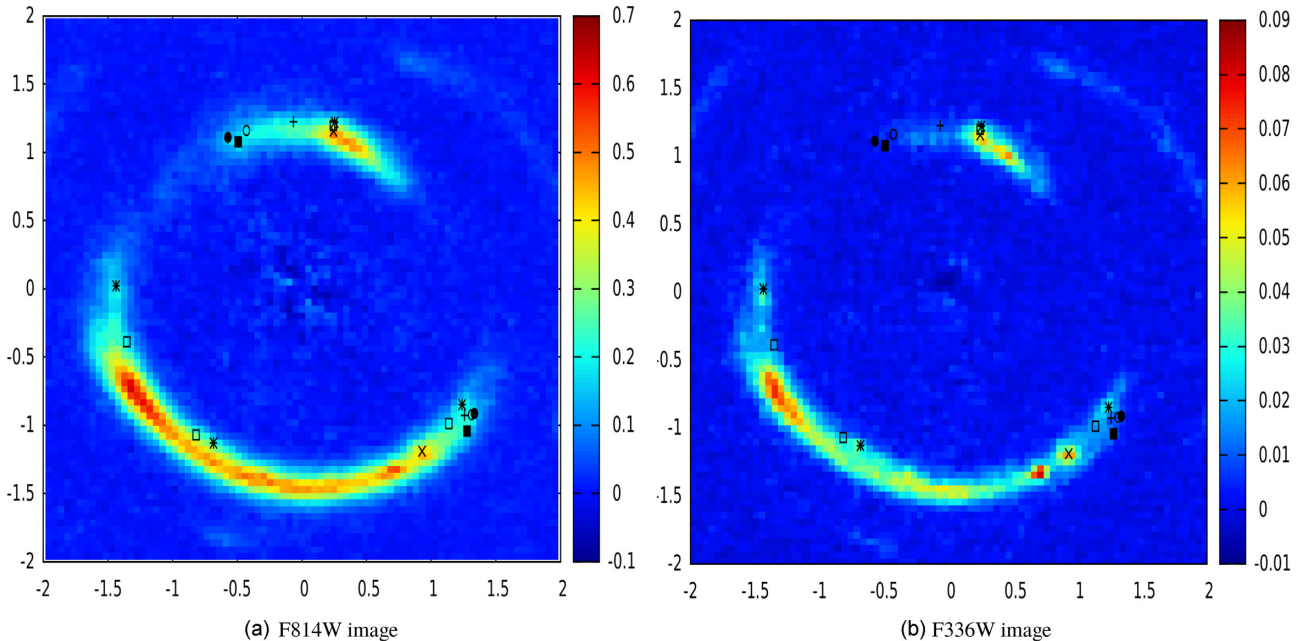

Figure 7. Image of SDSSJ0946+1006 in two different *HST* filters: F814W on the left (from Gavazzi et al. 2008), which is used for the primary lensing fits in this paper; and F336W on the right, from Sonnenfeld et al. (2012). Both images have the foreground galaxy subtracted out. To illustrate the lens mapping near the subhalo, seven points in the top arc are shown (black points with various markers) along with their partner images on the bottom arc according to our best-fitting lens model (tNFWmult). Note that in F336W, the surface brightness is dramatically reduced near the subhalo (filled circle, filled square, and open circle), but this is also true for the partner images in the lower right, indicating the lens mapping is consistent in both bands.

visible in the image, it is clearly subdominant compared to the lensed source.

To place more rigorous constraints on the subhalo's luminosity, the most direct approach is to include a light profile for the subhalo in our model whose parameters are varied simultaneously with the lens/source parameters. We carry out this strategy by modeling the subhalo's light with a Gaussian profile, varying the total luminosity LV and half-light radius $r_{1/2}$ as free parameters. The surface brightness from the subhalo is added to the light from the lensed sources before convolving with the PSF. We allow the half-light radius to vary from

0.005 arcsec up to 0.3 arcsec (or from $\approx$18 pc up to $\approx$1.1 kpc). In all other respects, our model is identical to the tNFWmult model.

Since the light profile approaches a point source at the low end of the prior range in $r_{1/2}$, numerical error can arise since the extent of the profile is small compared to the size of an image pixel. To solve this issue, we implement a 'zoom' feature where any pixels whose borders lie close enough to the subhalo light (we choose the distance criterion to be within twice the half-light radius) are split into subpixels, such that the subpixels have lengths no longer than $\frac{1}{4}r_{1/2}$ in either direction; if a pixel already satisfies this criterion,

no splitting occurs. The surface brightness is then evaluated at each subpixel, and these values are averaged to find the surface brightness of each parent pixel. With our choice of subpixel size, the surface brightness is quite accurate and does not significantly improve with further splittings. (Finally, we note that while we are using the zoom feature for a foreground light profile here, it can also be used for compact lensed sources as well.)

The best-fitting half-light radius of the subhalo is $r_{1/2} = 0.74^{+0.15}_{-0.10}$ kpc, while the stellar luminosity constraint is surprisingly stringent, giving $L_V = (1.1 \pm 0.1) \times 10^8 \, L_\odot$ (the uncertainties here give the 95 per cent credible interval). In addition, the subhalo's dark matter constraints are slightly different than in tNFWmult: the best-fitting $M_{2D}(1 \, \text{kpc})$ is $\approx 2 \times 10^9 \, M_\odot$, which is significantly smaller, while the scale radius has shrunk considerably to ≈ 0.1 kpc. The question remains, however; can the apparently large subhalo luminosity be trusted?

Comparing the residuals to those of our fiducial models (Fig. 1b) show that the residuals below the subhalo are much reduced by the inclusion of the stellar light profile. However, as discussed in Section 3, these residuals – which are located a few pixel-lengths away from the inferred subhalo centre – are likely to be due to unsubtracted foreground light from the primary lens galaxy. This can be seen in fig. 2 of Sonnenfeld et al. (2012), where they subtracted the foreground galaxy using an elliptical Sersic model: the residuals in question (the F814W image is plotted in row 2 of that figure) form part of a continuum that go all the way to the central region where the foreground galaxy is brightest. (The foreground subtraction in Gavazzi et al. 2008, used in this paper, employed a more sophisticated B-spline technique to remove the residuals, making this continuum less apparent.) As a result, the true luminosity of the subhalo is likely to be lower than our solution suggests. We thus regard $L_V < 1.2 \times 10^8 \, L_\odot$ to be a conservative upper bound on the subhalo's stellar luminosity. The true luminosity is likely to be lower than this, but need not be as low as the bound derived in Vegetti et al. (2010).

Finally, we note that Collett & Smith (2020) have recently discovered a third lensed source at redshift $z = 5.975$ using the MUSE spectrograph, one of whose images is very close to the location of the perturbing subhalo. However, the partner image (the larger of the two) from this source is observed to the lower right of where the large arcs are, and there is no discernible signal in that region in the F814 image. Therefore, we do not expect this high-redshift source to affect our upper bound on the subhalo luminosity derived here.

4 COMPARISON TO SUBHALO POPULATIONS WITHIN ANALOGUE LENSING GALAXIES IN THE ILLUSTRIS TNG100-1 SIMULATION

In the preceding sections, we have identified three properties of the subhalo in J0946+1006 for which we have constraints in order to fit the lensing data well. Conservatively, the following bounds are required: the projected mass within 1 kpc, $M_{2D}(1 \, \text{kpc}) \gtrsim 2 \times 10^9 \, M_\odot$; the average log-slope of the 2D density profile from 0.75 to 1.25 kpc $\gamma_{2D} < -1$; and the V-band luminosity $L_V < 1.2 \times 10^8 \, L_\odot$. To address the question whether such a highly concentrated subhalo is consistent with CDM, we now explore whether subhalos that satisfy these properties arise in cosmological simulations.

To find analogues of the main lensing galaxy, we require a cosmological simulation that includes the effects of baryons on the mass distribution of galaxies. This is partly because the radiative cooling of baryons and subsequent adiabatic contraction (AC) of the dark matter halo results in a denser galaxy, with larger Einstein radius, compared to the same galaxy in a dark-matter-only (DMO)

simulation (Blumenthal et al. 1986; Gnedin et al. 2004). The baryonic physics, in particular adiabatic contraction and active galactic nucleus (AGN) feedback, also affects the radial distribution and mass function of the subhalos (Despali & Vegetti 2016), which is critical for the comparison we are making here. We choose IllustrisTNG for this purpose, which is a suite of both hydrodynamical and DMO N-body simulations. The hydrodynamical suite of simulations includes the effect of gas cooling, as well as a prescription for subgrid physics including star formation/evolution, black hole seeding and evolution, and stellar and AGN feedback, while the DMO runs are strictly N-body simulations based on the CDM model (Nelson et al. 2017; Pillepich et al. 2017; Springel et al. 2017; Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2019a).

4.1 Sample selection

Although our primary results will be obtained using a full hydrodynamical simulation in IllustrisTNG, we will also analyse and compare to its DMO counterpart to gauge the effects of baryons on the satellite populations. We choose in particular the TNG100-1 and TNG100-1-DARK simulations, both of which have a box length of about 100 Mpc and the smallest dark matter particle mass of all the publicly available simulations in the TNG suite, with $m_{DM} = 7.5 \times 10^6 \, M_\odot$ for the hydrodynamical run, and $m_{DM} = 8.9 \times 10^6 \, M_\odot$ for the DMO run (TNG50 will have smaller mass particles but is not yet publicly available; Nelson et al. 2019b). Note that with this resolution, subhalos with masses less than $7.5 \times 10^9 \, M_\odot$ will have fewer than 1000 particles, and hence might be vulnerable to artificial disruption due to being unresolved (Carlsten et al. 2020). For this reason, we restrict our search to subhalos with total dark matter masses greater than $5 \times 10^9 \, M_\odot$ so as to exclude subhalos with less than ~ 700 particles, to ensure all subhalos have ~ 1000 particles or more. We do the same for the DMO run despite the slightly lower resolution (as we will see, however, among lower mass subhalos only a tiny fraction are expected to produce a large enough $M_{2D}(1 \, \text{kpc})$, eliminating the subhalos with lower particle counts).

We find analogue lensing galaxies in a manner similar to Despali & Vegetti (2016), by selecting TNG100-1 hosts and subhalos at a redshift z = 0.2 with masses close to those measured for J0946+1006 (Auger et al. 2010). We accomplish this by searching the TNG100-1 API for host haloes with total masses $1 \times 10^{13} - 6 \times 10^{13} \, M_\odot$ and stellar masses between 8×10^{10} and $8 \times 10^{12} \, M_\odot$ to match the host (for the DMO run we only use the total mass criteria). From the hosts that match our search criteria, we select subhalos with total masses $8 \times 10^9 - 4 \times 10^{11} \, M_\odot$, and dark matter masses $5 \times 10^9 - 4 \times 10^{11} \, M_\odot$ to address the resolution issues aforementioned. In the DMO run, we only use the latter range.

Although our expectation that a strong candidate would have undergone significant tidal stripping suggests restricting our selection of subhalos to those closest to their host haloes, we cannot predict where in the subhalo's orbit it is or when it passed pericentre as we restrict our analysis to one snapshot (or redshift). We therefore select all subhalos within r_{vir} of their hosts, as they may have been stripped at some earlier redshift when at pericentre of their orbits. Additionally, we expect the subhalo to be close to its host in projection. This can be true for any subhalo depending upon the LOS, further motivating relaxing the condition of distance from host.

This yielded a total of 3056 subhalos from 167 different hosts for the hydrodynamical run, and 4909 subhalos from 188 hosts in the DMO run. For the hydrodynamical run, we further discarded subhalos that have less than 30 star particles, none of which had a projected mass within 1 kpc of at least $10^9 \, M_\odot$; likewise, we
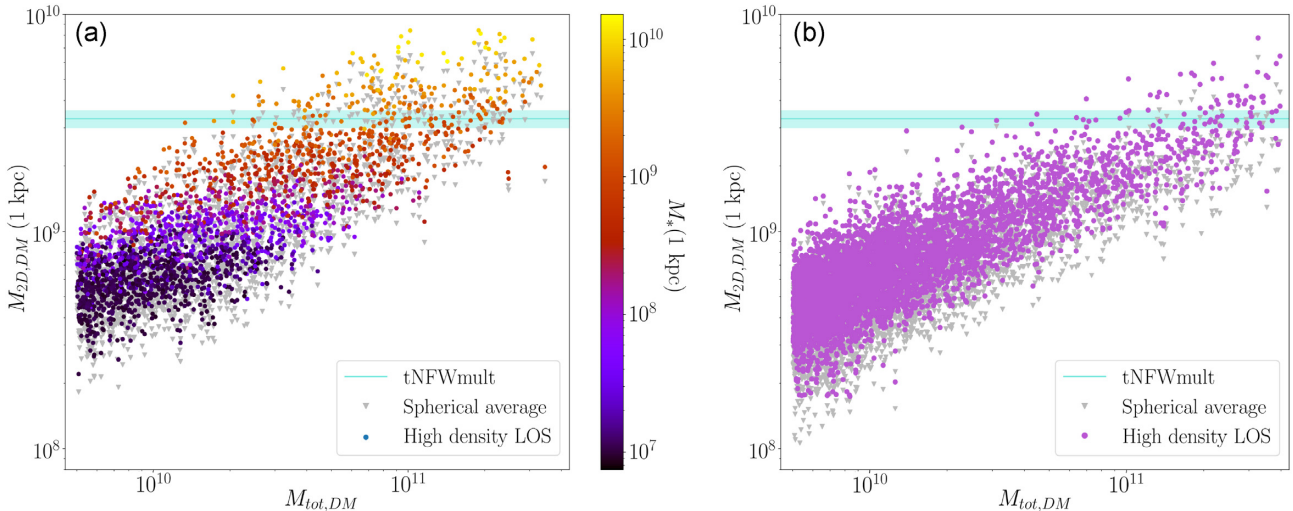
Figure 8. Projected mass within radius 1 kpc from subhalo centre, plotted with respect to total mass for simulated subhalos in the hydrodynamical TNG100-1 run [left-hand panel (a)] and the DMO TNG100-1-DARK run [right-hand panel (b)]. In both panels the grey triangle points are from the full spherical average over 1000 LOS, while the filled coloured circles are the same subhalos averaged over the top 10 per cent of highest density in 1 kpc LOS. In panel (a), the circles are colour coded by $M_*(1\,\mathrm{kpc})$, while in panel (b) they are solid violet. The 2σ error bands for $M_{\mathrm{2D,\,DM}}(1\,\mathrm{kpc})$ are also included for the 'tNFWmult' lensing model, where the subhalo is modeled with a tNFW profile and our model allows for departures from ellipticity in the host galaxy. All the plotted subhalos lie within r_{vir} from the host galaxy centre.

discarded subhalos with less than $8.5 \times 10^8\,\mathrm{M}_\odot$ projected mass in 1 kpc, which is below the minimum requirement of $10^9\,\mathrm{M}_\odot$. This leaves us with a final sample of 2205 subhalos for the hydrodynamical run.

For each subhalo, a spherically averaged density profile is constructed by binning the dark matter particles in radial mass bins, and 2D mass and surface density profiles are constructed similarly by binning in cylindrical shells and averaging over 1000 random LOS. Previous studies have pointed out that due to the triaxiality of dark matter haloes, LOS along the long axis of a halo can result in a significantly higher projected density and thus a stronger lensing signal due to selection bias (Sereno et al. 2015). In view of this, the strongest subhalo perturbations may be generated by subhalos whose long axes are oriented nearly along our LOS. In the following analysis, we explore this possibility by selecting the top 10 per cent of LOS which return the highest projected density within 1 kpc and produce profiles by averaging over these LOS for each subhalo, and contrasting their properties to the sample averaged over all LOS (https://www.overleaf.com/project/5e31dc3c2192ad00016cf4bb). Finally, to obtain γ_{2D} and circumvent the issue of noise near the inner regions of the profiles of these subhalos, we fit the profile with a power law from 0.75 to 1.25 kpc, and determine the slope for both the full spherical average, as well as the high density LOS. It should be borne in mind, however, that we are probing a regime in which the resolution is not sufficient to produce smooth density profiles for many subhalos; although our averaging procedure mitigates the noise, this may still lead to some artificial scatter in the inferred subhalo slopes.[3]

4.2 Subhalo candidates and comparison to lensing constraints

In Fig. 8, we plot the projected DM mass of each subhalo within 1 kpc, $M_{\mathrm{2D,\,DM}}(1\,\mathrm{kpc})$, versus the total subhalo mass for both the hydrodynamical [panel (a)] and DMO [panel (b)] simulation runs. The grey triangle points represent the full spherical average over 1000 random LOS for each subhalo, while the circles colour coded by $M_*(1\,\mathrm{kpc})$ represent the average over the top 10 per cent of the highest density LOS for each subhalo. Note that the required $M_{\mathrm{2D,\,DM}}(1\,\mathrm{kpc})$ is achieved only for the subhalos of total mass $\sim 10^{10}\,\mathrm{M}_\odot$ or greater. The remaining subhalos with $M_{\mathrm{2D,\,DM}}(1\,\mathrm{kpc}) < 10^9\,\mathrm{M}_\odot$ are eliminated from the proceeding analysis. This cut also eliminates most of the subhalos with low stellar content below $10^8\,\mathrm{M}_\odot$ within 1 kpc. The colour coding for the circle points reveals a trend in which the stellar mass in 1 kpc increases for larger projected halo mass within 1 kpc. Consistency with the lensing results (given by the horizontal shaded bar) therefore requires at least $\sim 2 \times 10^9\,\mathrm{M}_\odot$ of stellar mass within 1 kpc, despite the observed subhalo luminosity implying a much smaller stellar mass, a point to which we will return in Section 4.4.

In panel (b) of Fig. 8, we plot all 4909 subhalos found in TNG100-1-DARK. The trend of the points is similar to panel (a); however, with a shallower slope. Additionally, less of the subhalo population falls in the 2σ band for $M_{\mathrm{2D}}(1\,\mathrm{kpc})$ than the hydro run in panel (a), and those that fall in, do as a result of selection bias, which is visible in a comparison between the violet circle points and the full spherically averaged grey triangle points: the high density averaged ones (violet) obtain a large boost in projected mass within 1 kpc compared to the grey triangles. The large boost implies selection bias is amplified in DMO subhalos, which are more spherical in hydrodynamical simulations as a result of the added central potential from the baryons (Dubinski 1994; Kazantzidis et al. 2004). Additionally, the trend is less steep in panel (b) than in panel (a), indicating subhalos are denser in the hydrodynamic run, particularly for higher subhalo masses.

---

[3]The Illustris TNG50 simulation will solve this issue, at the cost of a smaller sample of subhalos due to smaller box size. The higher resolution of TNG50 will be able to better resolve the stellar profiles, as well as better resolution in mass by an order of magnitude and improved spatial resolution by a factor of ~ 2 compared to TNG100-1 (Pillepich et al. 2019). A reanalysis combining the results here of TNG100-1 with those from TNG50 would be useful for more robust measurements of the inner slope of the density profile, while

TNG100-1 gives us access to a larger population of subhalos in this mass range (Nelson et al. 2019b).
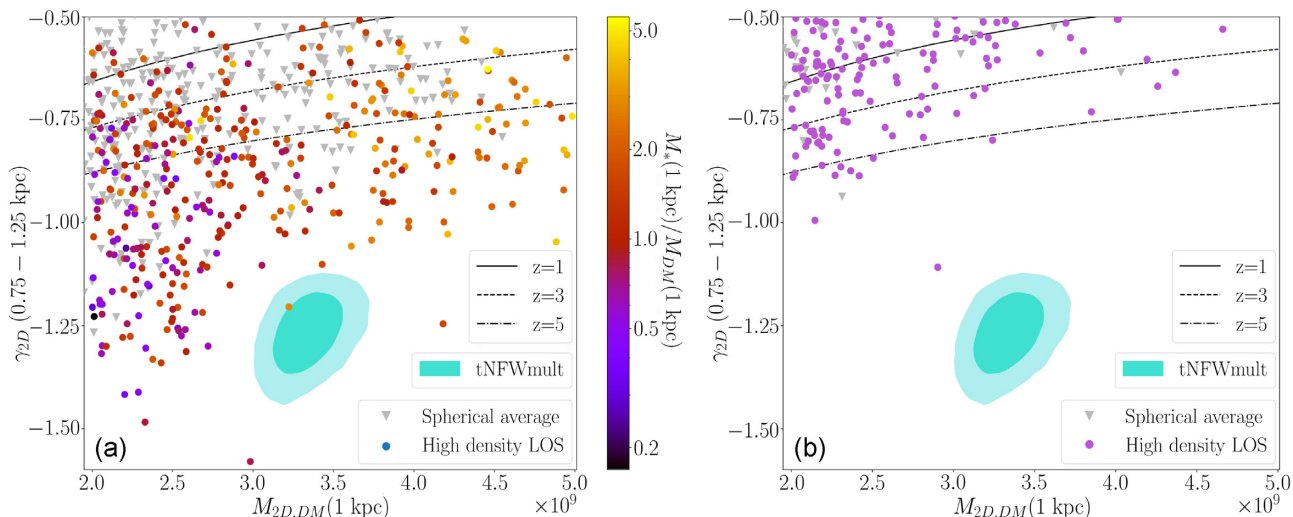
Figure 9. Slope of the surface density profile versus projected mass in 1 kpc for the TNG subhalos in the range $M(1\,\text{kpc}) = 2 - 5 \times 10^9\,\text{M}_\odot$ and $\gamma_{2D} \leq -0.5$ for both the hydrodynamical [panel (a)] and DMO [panel (b)] runs. Plotted are the full spherical average over 1000 LOS (grey triangles) as well as the average over the top 10 per cent highest density LOS in 1 kpc, colour coded by $M_*(1\,\text{kpc})/M_{DM}(1\,\text{kpc})$ in panel (a), and solid violet for panel (b). Both panels include the 68 per cent and 95 per cent contours from lensing constraints for our most conservative model, in which the subhalo is modeled with a tNFW profile and our model allows for departures from ellipticity in the host galaxy (which we term the 'tNFWmult' model), and the cosmological relation for dark matter for z = 1, 3, 5 at 3σ.

Next, we examine the density slope constraints. In Section 3.4, we showed that the log-slope of the subhalo's projected density profile is well constrained from the lensing signal over the approximate range from 0.1 up to 1 kpc. However, since the surface density profiles of the subhalos in TNG100-1 are often noisy near the central regions within 1 kpc due to poor resolution (see Appendix), we pick the upper end of this range and fit a line to the binned profiles between 0.75 and 1.25 kpc to obtain the average log-slope γ_{2D} for each subhalo.

In Fig. 9 we plot the average log-slope of the projected density profiles, γ_{2D}, with respect to $M_{2D}(1\,\text{kpc})$, for all subhalos with $2 \times 10^9 < M_{2D,\,DM}(1\,\text{kpc}) < 5 \times 10^9\,\text{M}_\odot$ and $\gamma_{2D} < -0.5$. Here again, panel (a) corresponds to the hydrodynamical run, while panel (b) is the DMO run, and the grey triangles again represent the full spherical average over 1000 LOS, whereas the coloured circles are the average over the top 10 per cent of the highest density LOS for each subhalo. In Fig. 9(a) the circles are colour coded by the ratio of stellar mass over dark matter mass within 1 kpc of the subhalo's centre. As in Fig. 8, the LOS is a significant factor, with the grey triangle markers being significantly shifted towards shallower slopes and smaller $M_{2D,\,DM}(1\,\text{kpc})$ from the colour-coded points. This effect is more significant for the DMO run, which is again consistent with a greater triaxiality expected for the DMO haloes (Dubinski 1994; Kazantzidis et al. 2004). The corresponding lensing constraints for the slope and the projected mass within 1 kpc are also plotted as contours at 68 per cent and 95 per cent CL. We show only our most conservative model 'tNFWmult', which modeled the subhalo with a tNFW profile and had the highest Bayesian evidence among the lens models we fit. A similar comparison that includes the other lens models, which have even steeper slopes compared to the model shown, is shown in Appendix B.

We also include the cosmological relation for dark matter-only field haloes with concentrations 3σ above the median for redshifts z = 1, 3, 5 from Dutton & Macciò (2014). Note that even at the 3σ level, typical CDM field haloes with masses within this range generally produce slopes shallower than -1; this can be seen in Fig. 9(b), where aside from some scatter, subhalos generally reside in the region where slopes are shallower than -1, in contrast with

Fig. 9(a) which shows the hydrodynamical subhalo population having steeper slopes. This suggests that for many of the subhalos that achieve $M_{2D}(1\,\text{kpc})$ in the hydrodynamical run, tidal stripping, and adiabatic contraction has steepened the profiles by contracting more dark matter into 1 kpc. However, in the vast majority of cases this does not result in a steep enough slope to fit the lensing constraints. Roughly ∼60 of the hydrodynamical subhalos are steep enough to be consistent with our inferred slope constraints when averaged over the top 10 per cent of high density LOS (coloured circles), while only six subhalos are steep enough when measured from the full spherical average (grey triangles). In addition, none of the subhalos are consistent with both the inferred slope and $M_{2D}(1\,\text{kpc})$ when averaged over all LOS. The difference is starker in the DMO simulation [panel (b)], where none of the subhalos produce a steep enough slope even for the highest density LOS.

4.3 Likelihood of a perturbing subhalo being as concentrated as the subhalo in SDSSJ0946+1006

Since the projected slope and mass can depend significantly on the LOS, we can ask the following question: supposing one makes a large number of mock 'observations' over many LOS for each subhalo in our TNG sample, what fraction of such observations would involve $M_{2D}(1\,\text{kpc})$ and $\gamma_{2D}(0.75-1.25\,\text{kpc})$ values consistent with our lensing constraint? This gives an idea of the likelihood of seeing a lensing perturbation by a subhalo as concentrated as the one in SDSSJ0946+1006. To address this question, it would be a mistake to use all 1000 individual LOS: the noisiness of the profiles affects the slope calculation, as discussed in Section 4.1. However, we find that averaging the profiles over as few as 20 nearby LOS reduces the noise dramatically (see Appendix B for comparison of different LOS averaging). Thus, we generate mock 'observations' by averaging 20 nearby LOS. The procedure is carried out as follows: for each subhalo, we calculate $M_{2D}(1\,\text{kpc})$; any LOS whose $M_{2D}(1\,\text{kpc})$ lies outside the range $3-4 \times 10^9\,\text{M}_\odot$ favoured by our fiducial lens model is discarded. The remaining LOS are sorted by their M_{2D} values and divided into groups of 20 LOS (with a few groups
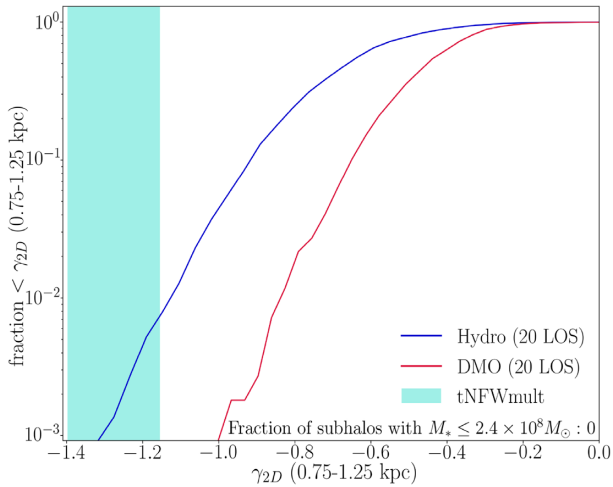
Figure 10. Fraction of mock 'observations' in our TNG100 sample for which a subhalo's projected density log-slope is steeper than a given threshold $\gamma_{2D}(0.75–1.25\,\text{kpc})$. Here we include only the observations that produce a projected mass within 1 kpc $M_{2D}(1\,\text{kpc})$ in the range $3–4 \times 10^9\,\text{M}_\odot$, for comparison to our fiducial lensing model ('tNFWmult'); the 95 per cent credible interval in γ_{2D} from our lens model is shown as the vertical shaded bar. (Joint constraints in $M_{2D}(1\,\text{kpc})$ and γ_{2D} are shown in Fig. 9). For each subhalo, 1000 LOS were generated, and each mock 'observation' consists of 20 similar LOS over which the profiles are averaged to reduce noise. Note than less than 1 per cent of the mock observations involve slopes that satisfy the lensing constraint. The fraction is even smaller for our other lens models (shown in Appendix B.).

possibly obtaining one extra LOS); each group constitutes a mock 'observation', over which the profiles are averaged and the resulting slope $\gamma_{2D}(0.75–1.25\,\text{kpc})$ is calculated. This process is repeated for each subhalo in our TNG100-1 sample, both for the hydrodynamical and DMO simulations. Note that in referring to each 'observation', we are not simulating the lensing perturbation of the subhalos here, but rather just calculating the projected mass and (smoothed) density slope of each simulated subhalo.

In Fig. 10 we plot the fraction of mock observations for which the projected (smoothed) density slopes are steeper than a given threshold γ_{2D}. The blue curve corresponds to subhalos in the TNG100-1 hydrodynamical simulation, whereas the red curve gives the DMO simulation. The vertical shaded bar gives the 95 per cent credible interval from the lensing constraint for our fiducial model (defined by the 2.5 per cent and 97.5 per cent percentiles of the marginalized posterior in γ_{2D}). Note that less than 1 per cent of the observations have projected density slopes consistent with the lensing constraint; in the DMO simulation, *none* of the observations produce steep enough density slopes. The situation is even starker for our other lens models, for which not a single mock observation produced a steep enough slope to match the lensing constraints even in the hydrodynamical simulation (see Appendix B for the corresponding plots with all lens models).

This result can be interpreted as follows. The angular size of a lensing perturbation is determined primarily by the subhalo's projected mass and proximity to the lens's critical curve (Minor et al. 2017), whereas to first approximation, the density slope affects the perturbation's shape, but not its size (Minor et al. 2021). This means that any of the mock observations in Fig. 10 would produce a lensing perturbation of similar angular size if the simulated subhalo were placed at the position of the subhalo in SDSSJ0946, and thus the majority of them are likely to be detectable. However, fewer than

1 per cent of these observations would produce a projected density slope as steep as what the lensing constraints are telling us. This suggests that the subhalo in SDSSJ0946+1006 is truly an outlier in CDM (at the 99 per cent confidence level) in terms of its central density slope, and hence concentration, given the size of the observed lensing perturbation.

Note that this argument can be made independently of the stellar mass of the subhalo. However, as we discuss in the following section, even those subhalos that do achieve densities and slopes that match our lensing constraints have stellar masses that far exceed the conservative upper bound on the subhalo's stellar mass, significantly worsening the tension with CDM.

4.4 Discussion

Although the subhalo in SDSSJ0946+1006 is clearly an outlier in our subhalo sample in TNG100-1, as Figs 9 and 10 show, we do find a few subhalos whose projected density profiles are consistent with the lensing constraints within 99 per cent confidence level (CL) for specific LOS. How are such high densities and density slopes achieved?

To understand this, we first consider the simulated subhalos' stellar mass. Among the sample of subhalos with $M_{2D,\,\text{DM}}(1\,\text{kpc}) \geq 2 \times 10^9\,\text{M}_\odot$, Fig. 8 shows that nearly all of them have total stellar masses greater than $10^9\,\text{M}_\odot$. Given the stellar luminosity constraint $L_V < 1.2 \times 10^8\,\text{L}_\odot$ found in Section 3.5, achieving a stellar mass this high would require the actual subhalo to have a V-band mass-to-light ratio greater than $8\,\text{M}_\odot/\text{L}_\odot$, which would be extraordinarily high and would imply a very steep stellar initial mass function (IMF). In old stellar populations typical for dwarf satellites, a Salpeter-like IMF generally produces $M_*/L_V \approx 2\,\text{M}_\odot/\text{L}_\odot$ (Martin, de Jong & Rix 2008). Direct measurements of the stellar IMF in Milky Way satellite galaxies infer a stellar IMF that is shallower than Salpeter, and hence imply even lower M_*/L_V (Geha et al. 2013; Gennaro et al. 2018); this is also consistent with the expectation from scaling relations in Local Group dwarfs (Woo, Courteau & Dekel 2008) and the low measured M_*/L_V in low surface brightness dwarfs (Du et al. 2020). While super-Salpeter IMF's have been deduced in high-redshift giant elliptical galaxies via spectroscopy and/or lensing (Conroy & van Dokkum 2012; Newman et al. 2013), an IMF extreme enough to produce $M_*/L_V > 8\,\text{M}_\odot/\text{L}_\odot$ has only been observed in the centres of massive galaxies (Conroy, van Dokkum & Villaume 2017), likely as a result of extreme starburst conditions (Chabrier, Hennebelle & Charlot 2014) unlikely to occur in a dwarf galaxy. If we assume a value that is likely a typical upper limit for dwarf galaxies, $M_*/L_V \approx 2\,\text{M}_\odot/\text{L}_\odot$, our lensing constraint would imply $M_* \lesssim 2.4 \times 10^8\,\text{M}_\odot$.

Among the simulated subhalos in our sample that achieve $M_{2D}(1\,\text{kpc}) \geq 2 \times 10^9\,\text{M}_\odot$ along some LOS, *none* have stellar masses equal to or smaller than our conservative upper bound of $2.4 \times 10^8\,\text{M}_\odot$. Since TNG100-1 has been shown to produce stellar masses roughly consistent with observations of dwarf galaxies in the Local Volume down to at least $M_{\text{halo}} \sim 10^{11}\,\text{M}_\odot$ (Carlsten et al. 2021), this suggests that subhalos with such high projected masses must necessarily be accompanied by a stellar mass that exceeds that of the subhalo in SDSSJ0946+1006. However, the strong upper bound on the subhalo's stellar mass does not by itself disqualify the simulated subhalos in our sample, since we do not necessarily expect the simulation to reproduce the correct stellar mass within 1 kpc. The stellar softening length ≈ 0.7 kpc implies the stellar density within 1 kpc must be taken with a grain of salt (Pillepich et al. 2017). Nevertheless, all the subhalo candidates with steep enough slopes (the < 1 per cent of observations in Fig. 10) have at least

$3 \times 10^9\,M_\odot$ in stellar mass within 1 kpc, an order of magnitude higher than our upper bound on the observed subhalo. Most also contain more stellar mass than dark matter mass within 1 kpc. This strongly suggests that the dark matter mass within 1 kpc has been boosted significantly by adiabatic contraction, an effect that may also be steepening the slope within 1 kpc (Gnedin et al. 2004). This is supported by the fact that none of the mock 'observations' produced a steep enough density slope in the DMO simulation (Fig. 10). Indeed, previous work suggests that IllustrisTNG has over-contracted dark matter haloes at higher galaxy masses, producing high central dark matter fractions compared to observational constraints (Lovell et al. 2018). (See Appendix A for a direct comparison of the stellar and dark matter density profiles for 21 of the closest candidate subhalos.) If the high concentration of these dark matter subhalos was achieved primarily through AC, then they are no longer consistent with the lensing results since the true stellar mass is likely to be below 3×10^8 $M_\odot$, not enough for AC to be a significant factor in reality – again, unless the stellar mass-to-light ratio takes an unprecedented value $M_*/L_V \gtrsim 8\,M_\odot/L_\odot$.

We conclude that the subhalo in the lens SDSSJ0946+1006 appears to be a remarkable outlier in ΛCDM, in two respects. Its projected mass within 1 kpc, constrained by the lensing, is an outlier given the upper bound on its luminosity, unless its stellar mass-to-light ratio takes an extraordinarily high value (making it an outlier in stellar mass). In addition, its density slope is an outlier (at >99 per cent confidence, by Fig. 10), unless a lens model could be found that produces a strong fit and infers a shallower density slope; but even if such a lensing solution could be found, this would imply a large halo mass and hence would again require a very high stellar mass-to-light ratio.

4.5 Can a more sophisticated primary lens model reconcile the inferred subhalo properties with ΛCDM?

Perhaps the most plausible scenario for reconciling the subhalo with ΛCDM is if a good-fitting lensing solution could be found that produces a projected mass within 1 kpc $\lesssim 2 \times 10^9\,M_\odot$. As Fig. 8 shows, subhalos with $M_{2D} \approx 1.5 \times 10^9\,M_\odot$ can have stellar masses reaching down to $10^8\,M_\odot$, which would make it consistent with our upper bound on the stellar mass. Since $M_{2D}(1\,\mathrm{kpc})$ is robust to changes in the subhalo profile, however, this would likely require some change in the host galaxy model that would allow the subhalo to be placed at a slightly different position with respect to the critical curve.

In Vegetti et al. (2010), the authors modeled this lens with the addition of pixellated corrections to the gravitational potential and showed that the data prefer a localized perturbation in the location of the inferred subhalo, regardless of the overall structure of the smooth lens model. Nevertheless, the question remains whether a more limited degeneracy might allow for the subhalo to be somewhat less dense than our inferences allow. Prior work has suggested that the properties of individual substructures can indeed be degenerate with the structure of the primary lens galaxy due to fundamental lensing degeneracies (Wagner 2018a), such as the source position transformation (Schneider & Sluse 2014; Wagner 2018b). While Enzi et al. (2020) have shown that complexity in the lensing potential is generally not degenerate with the inferred amount of low-mass substructures in gravitational lenses; there may be significant degeneracy with the detailed properties of high-mass substructures. Such degeneracies, when accounted for, may plausibly weaken constraints on the properties of substructure or time delay cosmography (Enzi et al. 2020; Denzel et al. 2021).

In this work, we have taken a step toward investigating this effect by introducing multipole terms to allow the lens model to deviate from a purely elliptical density profile. The trend we observe is that although the elliptical solution (without multipoles) infers a smaller $M_{2D}(1\,\mathrm{kpc})$, it requires a steeper slope to fit the lensing data, presumably to produce a strong enough perturbation. For example, our tNFW model with a perfectly elliptical host galaxy finds $M_{2D}(1\,\mathrm{kpc}) \approx 2.5 \times 10^9\,M_\odot$, but requires a significantly steeper slope $\gamma_{2D} \approx -2$, making it more difficult to reconcile with substructures in IllustrisTNG (see Fig. B1 in Appendix B, purple contour). As we shall see in Section 5, a similar trend is evident when the perturber is assumed to be a LOS halo, rather than a subhalo: solutions with smaller M_{2D} require a steeper slope. Thus it remains to be seen whether a more sophisticated model for the lens galaxy might allow for a good solution with $M_{2D}(1\,\mathrm{kpc}) \lesssim 2 \times 10^9\,M_\odot$ and a slope shallow enough to be consistent with the subhalos in Illustris TNG. Fitting both of the lensed sources (and perhaps even a third source recently detected by Collett & Smith 2020) with a flexible host galaxy and subhalo model, perhaps in combination with spatially resolved kinematics, may shed light on this by better constraining the host galaxy's density slope and contour shape in the vicinity of the subhalo.

5 WHAT IF THE PERTURBER IS A LOS HALO RATHER THAN A SUBHALO?

Several studies have suggested that a significant fraction of lensing perturbations are caused by dark matter haloes lying along the LOS to the lensed source, instead of being subhalos of the lensing galaxy (Xu et al. 2012; Li et al. 2017; Despali et al. 2018). Unfortunately in our fits, the tidal radius of the perturber is largely prior dominated and hence cannot be used to determine *a priori* whether the perturber is indeed a subhalo or not. This raises the question, can the inferred density of the perturber in SDSSJ0946+1006 be reduced if it lies along the LOS, such that the tension with CDM is mitigated? For example, if the perturber lies behind the lensing galaxy at z = 0.4, its angular diameter distance is nearly 50 per cent higher compared to the lens redshift (z = 0.222). Assuming the perturbation radius would be constrained to have a similar angular size (~0.3 arcsec), then the same mass within this perturbation radius would now be spread over roughly 1.5 kpc in radius, perhaps implying a lower central density. On the other hand, if the perturber lies in front of the lens at lower redshift, its Einstein radius becomes larger, perhaps indicating the perturber does not need to be so massive. Either effect could possibly reduce the mass of the perturber contained within 1 kpc. In this section, we investigate whether the assumption of a LOS perturber can lessen the tension with CDM.

The complicating factor in these scenarios is the role played by recursive lensing: instead of one lens plane, there are now two, and the lensing proceeds in two stages as the lensed light from the more distant galaxy is in turned lensed by the closer one (Schneider, Ehlers & Falco 1992). This recursive lensing has observable effects that can alter the quality of the fit, for better or worse. This is most obviously true when the perturber is behind the primary lens at high redshift, so that the convergence from the perturber is essentially lensed by the primary lens galaxy. This has the effect of smearing its convergence into an arc, or in extreme cases even leading to multiple 'images' of the same perturber (i.e. perturbations in two different locations in the lens plane).

We now investigate whether changing the redshift of the perturber can lead to a less dense halo without significantly degrading the fit, taking into account the full effect of recursive lensing by solving the recursive lensing equation (Schneider et al. 1992). In principle one could redo the analyses of Section 3, but varying the perturber's
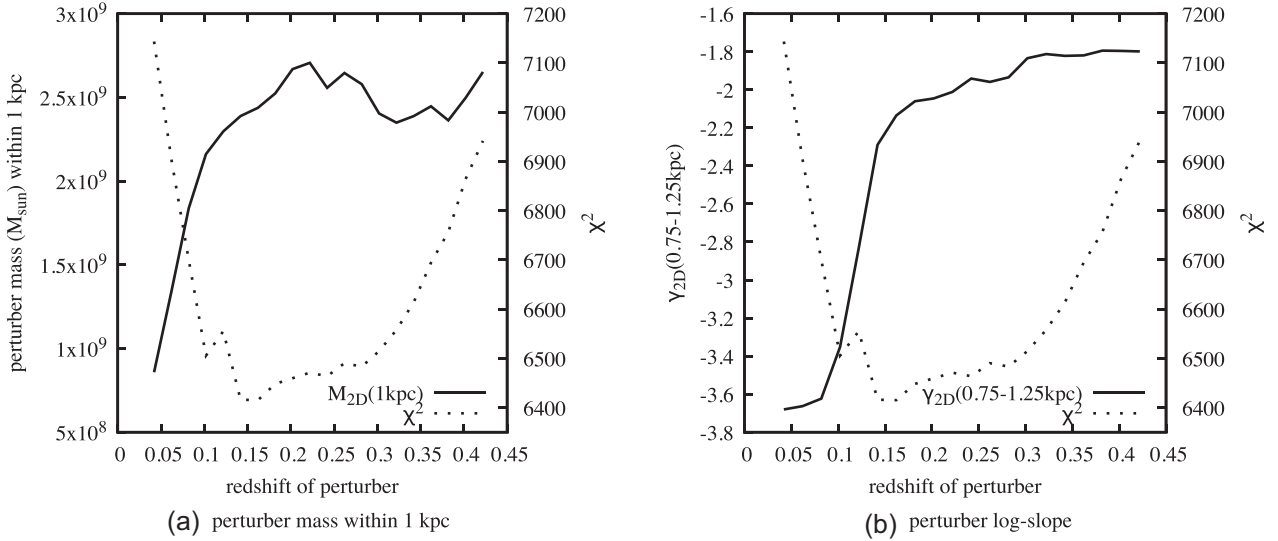
Figure 11. Best-fitting perturber's projected mass within 1 kpc and average log-slope $\gamma_{2D}(0.75{-}1.25\ \text{kpc})$ as the perturber's redshift is varied. In (a) the $M_{2D}(1\ \text{kpc})$ is given by the solid curve, while the χ^2 of the fit is given by the dashed curve with corresponding values given by the axis on the right. In (b) the log-slope is given by the (solid curve), again with the χ^2 given by the dashed curve for comparison. These plots were generated by taking the best-fitting model with tNFW profile, gradually varying the redshift in either direction with steps of $\Delta z = 0.02$ and reoptimizing the χ^2 function.

redshift as an additional model parameter. However, the degeneracies between the subhalo parameters (m_{200}, r_s, r_t, z_{pert}) would require running the nested sampling algorithm for a much longer time to achieve convergence, straining our computational resources. We thus opt for a simpler investigative approach: starting from the best-fitting model for the subhalo, we gradually change the redshift of the perturber, by steps $\Delta z = 0.02$, reoptimizing the chi-square at every step using Powell's method to find a new set of best-fitting lens and source parameters. First, we reduce the redshift down to $z = 0.05$, bringing the perturber in front of the lens plane; the procedure is then repeated going in the other direction up to $z = 0.45$. Provided that the redshift is changed slowly enough, one is still tracing out the approximate best-fitting model as a function of redshift.

The results are plotted in Fig. 11 for the tNFW model (the corresponding plots for the tNFWmult model whose behaviour is qualitatively similar, are given in Appendix C, Fig. C1). In Fig. 11(a) we plot the perturber's projected mass within 1 kpc (solid line, left axis) along with the best-fitting χ^2 value (dashed line, right axis) as a function of the perturber's redshift. Surprisingly, Fig. 11 shows the the variation in goodness-of-fit is equally apparent at both high and low redshifts. The recursive lensing effect is entirely responsible for this variation in the χ^2: if the recursive lensing effect is 'turned off', so that the convergences are simply added together, one can always achieve an equally good chi-square by rescaling the masses and scale lengths regardless of redshift. Hence, the fact that the χ^2 increases at low redshifts shows that the recursive lensing effect is not only important behind the lens plane, but also in front as the redshift gets lower. This is because the perturber is strongly lensing by itself (with Einstein radius of ~ 0.1 arcsec) and its Einstein radius increases as it comes closer, for fixed m_{200} and r_s values.[4] Hence, to achieve a similar perturbation radius (in angular size), the mass of the perturber must be reduced, driving the mass within 1 kpc to be lower for $z =$

0.05, but by this point, the fit is degraded significantly, with the χ^2 increasing by ~ 500. (See Appendix C for a discussion of how the perturbation size varies with redshift.)

Although one achieves a similar (or perhaps slightly better) fit down to $z_{\text{pert}} = 0.15$, the mass scale $M_{2D}(1\ \text{kpc})$ is only slightly reduced. To get down to 1 $M_\odot$ within 1 kpc, the perturber must be brought down to $z = 0.05$, and the fit is quite poor by this time with significant residuals in the vicinity of the subhalo. In addition, Fig. 11(b) shows that the log-slope $\gamma_{2D}(1\ \text{kpc})$ becomes extremely steep in this limit, diving below -3.6 at $z = 0.05$, due to the best-fitting tidal radius becoming very small. This indicates an extremely high concentration (indeed, approaching the limit of a point mass) and the slope is much steeper than that of field haloes in CDM, as Fig. 9 shows (black curves give the log-slope for concentrations 3σ above the median value at different redshifts). Therefore, the limit of low redshift, while leading to a lower $M_{2D}(1\ \text{kpc})$, also exacerbates the concentration issue and leads to a significantly poorer fit.[5]

We conclude that the assumption of a LOS perturber does not significantly lessen the tension with CDM; indeed, it likely worsens the tension, since field haloes are in general less dense than subhalos of similar mass. In addition, even if a slightly better fit might be achieved down to $z_{\text{pert}} = 0.15$, the upper bound on the stellar luminosity of the perturber becomes stricter: the luminosity distance at $z = 0.15$ is less than 70 per cent its value at the primary lens redshift, which leads to the upper bound on luminosity being less than half as bright, making it harder to explain with a very massive halo. Thus, although both scenarios are in tension with CDM, we

---

[4]For completeness' sake, we note that if you bring the perturber much closer than z = 0.05, its Einstein radius will finally decrease again as the mass 'spreads out' over the lens plane, driving the surface density per arcsec to be low. But the perturber in this case is so dense that this effect has not kicked in yet by z = 0.05.

[5]A related question is whether the perturber could be a black hole within the Milky Way. We did a PolyChord run with a point mass perturber (and no multipoles) at the redshift of the lens and found a best-fittingting solution with χ^2/pixel ≈ 1.71, only modestly worse than the tNFW or CoreCusp models listed in Table 1. Following the above procedure, if we gradually reduce the perturber redshift down to $z = 10^{-6}$ (about 4 kpc away), the pertuber has a best-fittingting mass $\approx 1.1 \times 10^4$ $M_\odot$, consistent with a large intermediate-mass black hole. However, there are strong residuals in the vicinity of the perturber and we have χ^2/pixel ≈ 1.95, indicating the fit is severely degraded in the low-redshift limit, consistent with the results using tNFW in this section.

consider the scenario of a perturbing subhalo to be more plausible than that of a LOS halo.

6 DISCUSSION: CAN DARK MATTER PHYSICS EXPLAIN THE HIGH CONCENTRATION OF THE SUBHALO IN LENS SDSSJ0946+1006?

Having established that the subhalo in SDSSJ0946+1006 is an outlier in CDM based on the lensing constraints, we ask if a modification of the dark matter particle physics can explain its high central density and concentration? Alternatives to cold, collisionless CDM (such as warm dark matter or self-interacting dark matter) are typically invoked to mitigate the cusp-core or 'too big to fail' problems by *reducing* the central densities of small dark matter haloes (Vogelsberger et al. 2012; Lovell et al. 2014), which would appear to exacerbate the problem here.

However, one intriguing possibility is the phenomenon of core collapse due to dark matter self-interactions (Ahn & Shapiro 2005; Elbert et al. 2014). In the self-interacting dark matter model, the self-interactions between dark matter particles thermalize the central regions of dark matter haloes, producing a constant density core (Dave et al. 2001). Over long enough time-scales, however, the central core collapses as heat is transferred outward (the 'gravothermal catastrophe' familiar in globular clusters; Quinlan 1996), resulting in high central densities. For subhalos, the collapse time-scale can be significantly shorter than for field haloes (Nishikawa, Boddy & Kaplinghat 2020). It has been argued that this phenomena could be playing a role in setting the observed diversity of dark matter densities measured in the Milky Way satellite galaxies (Kahlhoefer et al. 2019; Zavala et al. 2019; Kaplinghat, Ren & Yu 2020; Sameie et al. 2020; Turner et al. 2021). Could this core collapse phenomenon also be responsible for the observed properties of the substructure in SDSSJ0946+1006? It is not possible to answer this question at present. The interplay between the velocity dependence of the cross-section, the orbit of the subhalo, the concentration at infall, and the tidal field of the host halo allow for a rich variety of evolution in the subhalo density (Kahlhoefer et al. 2019), which needs to be investigated in detail. We highlight in passing another intriguing possibility that core collapse could also explain the apparent high concentration of the dark matter substructure discovered in the GD-1 stellar stream by Bonaca et al. (2019).

7 CONCLUSIONS

We have fit several lens models to the gravitationally lensed images in SDSSJ0946+1006 and confirm the discovery of a dark and remarkably concentrated subhalo made by Vegetti et al. (2010). Our most robust measurement is the subhalo's projected mass within 1 kpc of its centre of $(2-3.7) \times 10^9$ M$_\odot$, with our highest evidence model inferring $M_{2D}(1 \text{ kpc}) = (3.3 \pm 0.3) \times 10^9$ M$_\odot$. This mass inference is robust to changes in the subhalo's assumed density profile or tidal radius, as 1 kpc is quite close to the subhalo's lensing perturbation radius defined in Minor et al. (2017). This is an extraordinarily high central density given our conservative upper bound on its stellar luminosity $L_V \approx 1.2 \times 10^8$ L$_\odot$. In addition to the mass, we find the average log-slope of the subhalo's projected density profile over the range 0.75–1.25 kpc to be steeper than isothermal for all models, with the highest-evidence model giving $\gamma_{2D} = -1.27^{+0.11}_{-0.13}$. Our reconstructions of the lensed images and source galaxy when the subhalo is modeled with a tNFW profile are shown in Figs 1 and 2, and the inferred lens parameters for all models are given in Table 1.

To test whether such a dense subhalo is expected to occur in ΛCDM, we compared our inferred subhalo mass and density slope to those of subhalos within 167 analogue lensing galaxies in the Illustris TNG100-1 hydrodynamical simulation, and 188 analogue lensing galaxies in the Illustris TNG100-1-DARK DMO simulation. The results are encapsulated in Figs 9 and 10. We conclude with the points listed below.

(i) By generating mock 'observations' along many LOS for each subhalo in our Illustris TNG100 sample, we find that while many subhalos have projected masses within 1 kpc that satisfy the lensing constraint, fewer than 1 per cent of such observations produce projected density slopes consistent with the lensing constraints at the 95 per cent confidence level (Fig. 10). This implies that if a CDM subhalo produces a lensing perturbation of similar angular size to what is observed in SDSSJ0946+1006, the likelihood of the subhalo having a density slope as steep as our lensing constraints is < 1 per cent. This conclusion can be made independently of the subhalo's stellar mass.

(ii) Among the simulated subhalos that have a projected mass within 1 kpc of $>2 \times 10^9$ M$_\odot$ along at least one LOS, as required by the lensing constraints, *all* of them have stellar masses exceeding our conservative upper bound for the observed subhalo ($M_* < 2.4 \times 10^8$ M$_\odot$) for which we assume a stellar mass-to-light ratio $M_*/L_V = 2$ M$_\odot$/L$_\odot$, consistent with a Salpeter-like IMF. This is the case regardless of the subhalo's density slope (i.e. concentration). For the actual subhalo to have a stellar mass consistent with the TNG100 candidates would require at least $M_*/L_V \gtrsim 8$ M$_\odot$/L$_\odot$, which would be unprecedentedly high in a dwarf galaxy.

(iii) The small subset of simulated subhalos that are consistent with the lensing constraints (both projected mass and density slope) within 99 per cent CL for specific LOS all have very high stellar masses within 1 kpc ($\gtrsim 3 \times 10^9$ M$_\odot$, higher than the dark matter mass). These subhalos are likely to have had their central dark matter fractions boosted by adiabatic contraction, a view that is supported by the fact that *none* of the mock observations in the DMO simulation achieve a steep enough slope (red curve in Fig. 10). In view of our upper bound on the subhalo's stellar mass (2.4×10^8 M$_\odot$), adiabatic contraction is unlikely to have been a significant factor in the observed subhalo.

(iv) We find that the tension with CDM cannot be mitigated if the perturber is a field halo along the LOS, rather than a subhalo of the lens galaxy: although the perturber's best-fitting projected mass within 1 kpc drops below 2×10^9 M$_\odot$ for a low-redshift perturber, the goodness-of-fit plummets along with it, and the best-fitting density slope becomes much steeper (Fig. 11). In addition, the constraint on the perturber's luminosity becomes more stringent at low redshift, requiring an even smaller stellar mass, while the lack of tidal stripping in the field makes the density slope harder to explain.

Thus, the central dark matter density, the concentration, and the apparent darkness of the observed subhalo in SDSSJ0946+1006 make it an extreme outlier in the ΛCDM model.

Perhaps the most plausible way to reconcile the subhalo with CDM is if a lensing solution could be found with projected mass $M_{2D}(1 \text{ kpc}) \lesssim 2 \times 10^9$ M$_\odot$ within 1 kpc, since subhalos in this range can have a stellar mass consistent with our upper bound ($M_* < 2.4 \times 10^8$ M$_\odot$). Since $M_{2D}(1 \text{ kpc})$ is robust to changes in the subhalo model, such a solution would likely require a more sophisticated model for the host galaxy, which may be degenerate with the inferred subhalo properties (see Section 4.5). However, our lensing solutions that achieve smaller $M_{2D}(1 \text{ kpc})$ require significantly steeper slopes (Fig. B1), worsening the tension with CDM, so it is unclear whether such a solution would

be viable. Fitting both of the lensed sources in SDSSJ0946+1006 with a more flexible host galaxy model (along with the subhalo) may provide clarity on this question, perhaps in combination with spatially resolved kinematic data for the lens galaxy. In addition, repeating our analysis in Illustris TNG50 would also provide a more robust measurement of the subhalos' dark matter slopes along individual LOS, particularly at lower subhalo masses.

If the unexpected properties of the perturbing subhalo in SDSSJ0946+1006 are due to the particle physics of dark matter, such as dark matter self-interactions, one may expect that many more perturbers of its kind will be detected among the thousands of strong lenses expected to be discovered by the Euclid and LSST surveys. Such highly concentrated subhalos may also be detected by their perturbations of tidal streams; indeed, such a detection may have already occurred (Bonaca et al. 2019). If so, we may be catching our first glimpses of a population of low-mass concentrated dark matter subhalos whose properties would allow us to directly constrain the particle physics of dark matter in the coming years.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

No new data were generated or analysed in support of this research.

REFERENCES

Ahn K., Shapiro P. R., 2005, MNRAS, 363, 1092
Andrade K. E., Minor Q., Nierenberg A., Kaplinghat M., 2019, MNRAS, 487, 1905
Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, ApJ, 724, 511
Baltz E. A., Marshall P., Oguri M., 2009, J. Cosmol. Astropart. Phys., 1, 015
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986, ApJ, 301, 27
Bonaca A., Hogg D. W., Price-Whelan A. M., Conroy C., 2019, ApJ, 880, 38
Carlberg R. G., 2012, ApJ, 748, 20
Carlberg R. G., Grillmair C. J., 2013, ApJ, 768, 171
Carlsten S. G., Greene J. E., Peter A. H. G., Greco J. P., Beaton R. L., 2020, ApJ, 902, 124
Carlsten S. G., Greene J. E., Peter A. H. G., Beaton R. L., Greco J. P., 2021, ApJ, 908, 109
Chabrier G., Hennebelle P., Charlot S., 2014, ApJ, 796, 75
Collett T. E., Auger M. W., 2014, MNRAS, 443, 969
Collett T. E., Smith R. J., 2020, MNRAS, 497, 1654
Conroy C., van Dokkum P. G., 2012, ApJ, 760, 71
Conroy C., van Dokkum P. G., Villaume A., 2017, ApJ, 837, 166
Cyr-Racine F.-Y., Moustakas L. A., Keeton C. R., Sigurdson K., Gilman D. A., 2016, Phys. Rev. D, 94, 043505

Dalal N., Kochanek C. S., 2002, ApJ, 572, 25
Dave R., Spergel D. N., Steinhardt P. J., Wandelt B. D., 2001, ApJ, 547, 574
Denzel P., Coles J. P., Saha P., Williams L. L. R., 2021, MNRAS, 501, 784
Despali G., Vegetti S., 2016, MNRAS, 469, 1997
Despali G., Vegetti S., White S. D. M., Giocoli C., van den Bosch F. C., 2018, MNRAS, 475, 5424
Du W., Cheng C., Zheng Z., Wu H., 2020, AJ, 159, 138
Dubinski J., 1994, ApJ, 431, 617
Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359
Dye S., Warren S. J., 2005, ApJ, 623, 31
Elbert O. D., Bullock J. S., Garrison-Kimmel S., Rocha M., Oñorbe J., Peter A. H. G., 2014, MNRAS, 453, 29
Enzi W., Vegetti S., Despali G., Hsueh J.-W., Metcalf R. B., 2020, MNRAS, 496, 1718
Erkal D., Belokurov V., 2015, MNRAS, 454, 3542
Gavazzi R., Treu T., Koopmans L. V. E., Bolton A. S., Moustakas L. A., Burles S., Marshall P. J., 2008, ApJ, 677, 1046
Geha M. et al., 2013, ApJ, 771, 29
Gennaro M. et al., 2018, ApJ, 863, 38
Gilman D., Du X., Benson A., Birrer S., Nierenberg A., Treu T., 2020, MNRAS, 492, L12
Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, ApJ, 616, 16
Hezaveh Y., Dalal N., Holder G., Kisner T., Kuhlen M., Perreault Levasseur L., 2016a, J. Cosmology Astropart. Phys., 2016, 048
Hezaveh Y. D. et al., 2016b, ApJ, 823, 37
Kahlhoefer F., Kaplinghat M., Slatyer T. R., Wu C.-L., 2019, J. Cosmol. Astropart. Phys., 2019, 010
Kaplinghat M., Ren T., Yu H.-B., 2020, J. Cosmol. Astropart. Phys., 2020, 027
Kazantzidis S., Kravtsov A. V., Zentner A. R., Allgood B., Nagai D., Moore B., 2004, ApJ, 611, L73
Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82
Kochanek C. S., Dalal N., 2004, ApJ, 610, 69
Koopmans L. V. E., 2005, MNRAS, 363, 1136
Li R., Frenk C. S., Cole S., Wang Q., Gao L., 2017, MNRAS, 468, 1426
Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, MNRAS, 439, 300
Lovell M. R. et al., 2018, MNRAS, 481, 1950
Mao S., Schneider P., 1998, MNRAS, 295, 587
Marinacci F. et al., 2018, MNRAS, 480, 5113
Martin N. F., de Jong J. T. A., Rix H.-W., 2008, ApJ, 684, 1075
Metcalf R. B., Madau P., 2001, ApJ, 563, 9
Minor Q. E., Kaplinghat M., Li N., 2017, ApJ, 845, 118
Minor Q. E., Kaplinghat M., Chan T. H., Simon E., 2021, MNRAS, in press
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Muñoz J. A., Kochanek C. S., Keeton C. R., 2001, ApJ, 558, 657
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Nelson D. et al., 2017, MNRAS, 475, 624
Nelson D. et al., 2019a, Comput. Astrophys. Cosmol., 6, 2
Nelson D. et al., 2019b, MNRAS, 490, 3234
Newman A. B., Treu T., Ellis R. S., Sand D. J., 2013, ApJ, 765, 25
Nierenberg A. M., Treu T., Wright S. A., Fassnacht C. D., Auger M. W., 2014, MNRAS, 442, 2434
Nishikawa H., Boddy K. K., Kaplinghat M., 2020, Phys. Rev. D, 101, 063009
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010, AJ, 139, 2097
Pillepich A. et al., 2017, MNRAS, 475, 648
Pillepich A. et al., 2019, MNRAS, 490, 3196
Quinlan G. D., 1996, New Astron., 1, 255
Ritondale E., Vegetti S., Despali G., Auger M. W., Koopmans L. V. E., McKean J. P., 2019, MNRAS, 485, 2179
Rocha M., Peter A. H. G., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Oñorbe J., Moustakas L. A., 2013, MNRAS, 430, 81
Sameie O., Yu H.-B., Sales L. V., Vogelsberger M., Zavala J., 2020, Phys. Rev. Lett., 124, 141102
Schneider P., Sluse D., 2014, A&A, 564, A103
Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer, Berlin/Heidelberg

Sereno M., Giocoli C., Ettori S., Moscardini L., 2015, MNRAS, 449, 2024

Sonnenfeld A., Treu T., Gavazzi R., Marshall P. J., Auger M. W., Suyu S. H., Koopmans L. V. E., Bolton A. S., 2012, ApJ, 752, 163

Springel V. et al., 2017, MNRAS, 475, 676

Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983

Turner H. C., Lovell M. R., Zavala J., Vogelsberger M., 2021, MNRAS, 505, 5327

Vegetti S., Koopmans L. V. E., 2009, MNRAS, 392, 945

Vegetti S., Vogelsberger M., 2014, MNRAS, 442, 3598

Vegetti S., Koopmans L. V. E., Bolton A., Treu T., Gavazzi R., 2010, MNRAS, 408, 1969

Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., 2012, Nature, 481, 341

Vegetti S., Koopmans L. V. E., Auger M. W., Treu T., Bolton A. S., 2014, MNRAS, 442, 2017

Vogelsberger M., Zavala J., Loeb A., 2012, MNRAS, 423, 3740

Wagner J., 2018a, A&A, 615, A102

Wagner J., 2018b, A&A, 620, A86

Woo J., Courteau S., Dekel A., 2008, MNRAS, 390, 1453

Xu D. D., Mao S., Cooper A. P., Gao L., Frenk C. S., Angulo R. E., Helly J., 2012, MNRAS, 421, 2553

Xu D., Sluse D., Gao L., Wang J., Frenk C., Mao S., Schneider P., Springel V., 2015, MNRAS, 447, 3189

Zavala J., Lovell M. R., Vogelsberger M., Burger J. D., 2019, Phys. Rev. D, 100, 063007

APPENDIX A: CANDIDATE SUBHALOS FROM ILLUSTRIS TNG100-1

To take a closer look at the closest subhalo candidates in TNG100-1, we further investigated 21 subhalos that had the closest match with the lensing constraints by selecting subhalos with 2.7×10^9 $M_\odot < M_{2D,DM}(1 \text{ kpc}) < 4.2 \times 10^9 M_\odot$ and $\gamma_{2D} < -1$. These subhalos are shown in panel 1 of Fig. A1. Notice here that only two of these subhalos make the cut when full spherical averaging was used. For these subhalos, we plot and show the 3D stellar and dark matter density profiles, grouping them by increasing projected mass within 1 kpc [panels (b), (c), and (d) in Fig. A1]. In all the panels, the stellar densities are either higher than the dark matter densities or are comparable near the central region of the subhalo. The subhalos with the highest projected mass in 1 kpc [panel (d) in Fig. A1] have stellar densities that are higher than their respective dark matter densities by as much as an order of magnitude. These high stellar densities strongly indicate that adiabatic contraction may have modified the dark matter halo
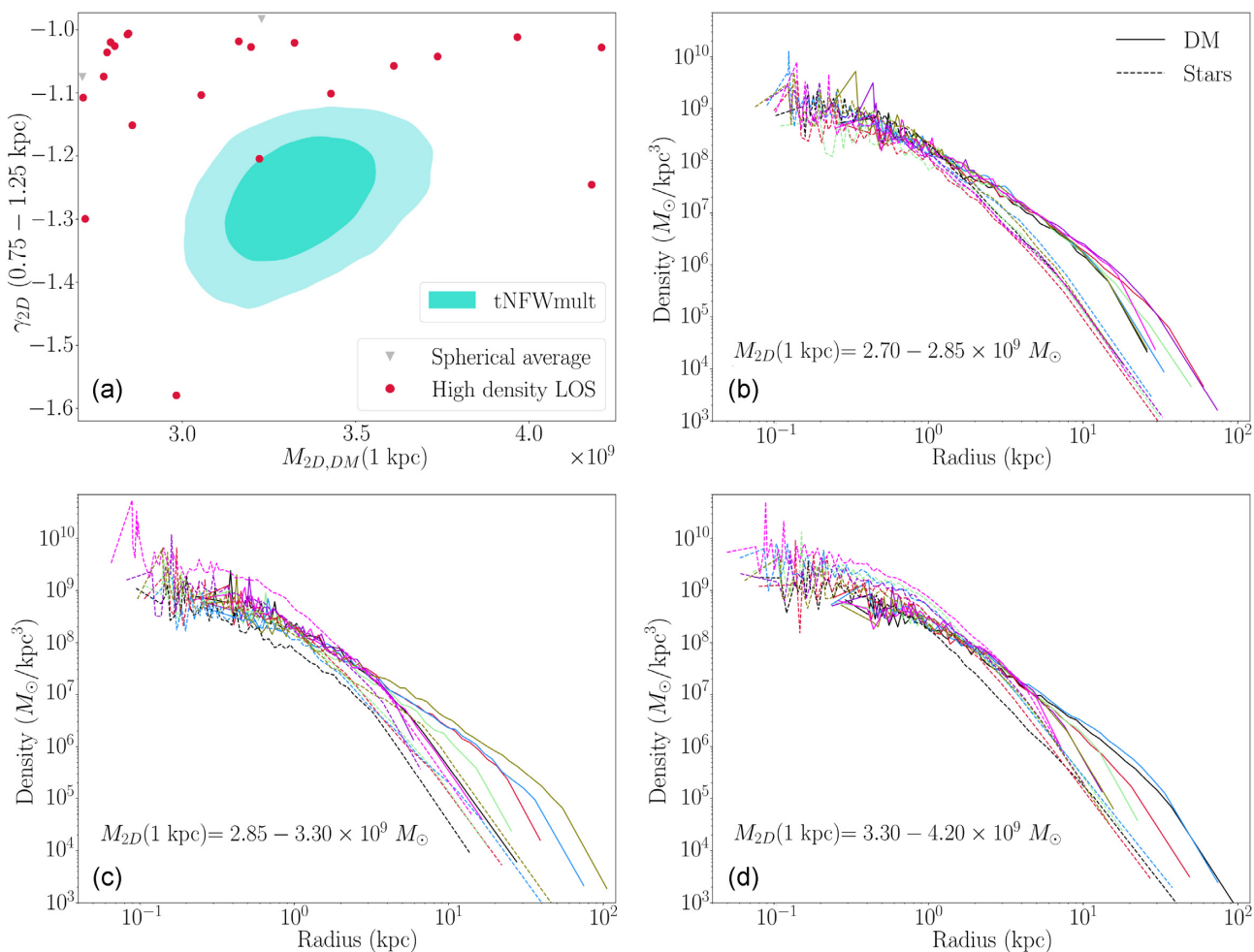
Figure A1. The 21 TNG100-1 subhalo candidates in closest range to the lensing constraints from the tNFW model with multinest [panel (a)]. 3D dark matter (solid line) and stellar (dashed line) density profiles corresponding to the subhalo candidates in panel (a), arranged by increasing projected mass in 1 kpc [panels (b), (c), (d)].

structures of these subhalos, steepening them further than the NFW expectation. We thus conclude that the subhalo analogues found in TNG100-1 are analogues because AC steepens their slopes and/or pushes enough mass into 1 kpc, combined with tidal truncation during the subhalo's pass through pericentre along its orbit in its host halo.

APPENDIX B: COMPARISON OF ALL FOUR LENSING MODELS TO ILLUSTRIS TNG100-1 SAMPLE

Here we make a comparison of all four of our lensing models (whose inferred parameters are given in Table 1) to the subhalo sample in the IllustrisTNG simulation presented in Section 4.

In Fig. B1 we make a plot that is identical to Fig. 9, except that all four of our lens models are shown. Note that the models with elliptical host galaxy (tNFW and CoreCusp) both have consistent subhalo masses within 1 kpc, while the models that include multipoles (tNFWmult and CoreCuspmult) have nearly identical inferred masses within 1 kpc. This is a consequence of the fact that the subhalo's perturbation radius (defined by the distance of the subhalo's centre to the point of maximum warping of the lens' critical curve), is nearly 1 kpc, and hence the projected mass within 1 kpc is independent of the subhalo's density profile (Minor et al. 2017). However, the tNFW, CoreCuspmult, and CoreCusp models all infer significantly steeper slopes $\gamma_{2D}(0.75-1.25$ kpc) compared to the tNFWmult model, and hence are even more difficult to explain in IllustrisTNG-100.

Next, we examine the likelihood of the simulated subhalos having a projected mass and density slope consistent with the model. Using the same procedure for generating mock 'observations' as in Section 4.3, we plot the cumulative distributions in γ_{2D}. The results are shown in Fig. B2. We now examine two different ranges in $M_{2D}(1$ kpc): $2-3 \times 10^9\,M_\odot$, favoured by the tNFW and CoreCusp models, and $3-4 \times 10^9\,M_\odot$, favoured by the models with multipoles (tNFWmult, CoreCuspmult). Vertical bars show the 95 per cent credible interval (defined by the 2.5 per cent and 97.5 per cent percentiles of the

marginalized posterior in γ_{2D}) for each model. To examine whether our results are dependent on how we average the profiles over nearby LOS, we plot the distributions for 20 LOS averaging (solid curve) and 50 LOS averaging (dashed curve) for both the hydro (blue) and DMO (dashed) simulations. Note the resulting curves are relatively unchanged whether we average over 20 or 50 LOS, although the 50 LOS averaging gives a less smooth distribution due to having fewer observations overall. Remarkably, in the tNFW, CoreCusp, and CoreCuspmult models, none of the observations give a steep enough slope to be consistent with the lensing constraints. For our most conservative model, tNFWmult (which is the one shown in Fig. 10), the fraction is < 1 per cent. The fractions become negligible in the DMO sample for all models. We conclude from this that the actual subhalo is an outlier in its density slope (and hence concentration) in CDM for all models. In addition, as discussed in Section 4.4, the few subhalos that do achieve a steep enough slope for some LOS all have $>10^9\,M_\odot$ within 1 kpc, far exceeding our conservative bound on the subhalo's stellar mass.

From the lens modeling point of view, one might hold out hope that a more sophisticated model could yield a solution that is more in accord with ΛCDM predictions – either by reducing $M_{2D}(1$ kpc), or by yielding a shallower slope. However, as discussed in Section 3.3, there may be relatively little wiggle room for this (a third possibility, that the perturber is a LOS halo rather than a subhalo, was discussed in Section 5). The most robust inference that can be made from the subhalo is its perturbation radius and the subhalo's projected mass enclosed therein. The four models we fit to the subhalo yielded perturbation radii whose values (roughly 1 kpc) differ by as much as a pixel length, as one might expect; it is this difference that results in different inferred masses within 1 kpc, spanning a range $2-3.5 \times 10^9\,M_\odot$.

In Fig. B3 we plot the critical curves from all four models, along with the subhalo position in each model; note that all four models have a critical curve perturbation of similar size, but differ slightly in the subhalo position (resulting in a slightly different $r_{\delta c}$). The covariance between subhalo position and perturbation radius is evident in the posteriors. For example, in Fig. B4 we plot joint
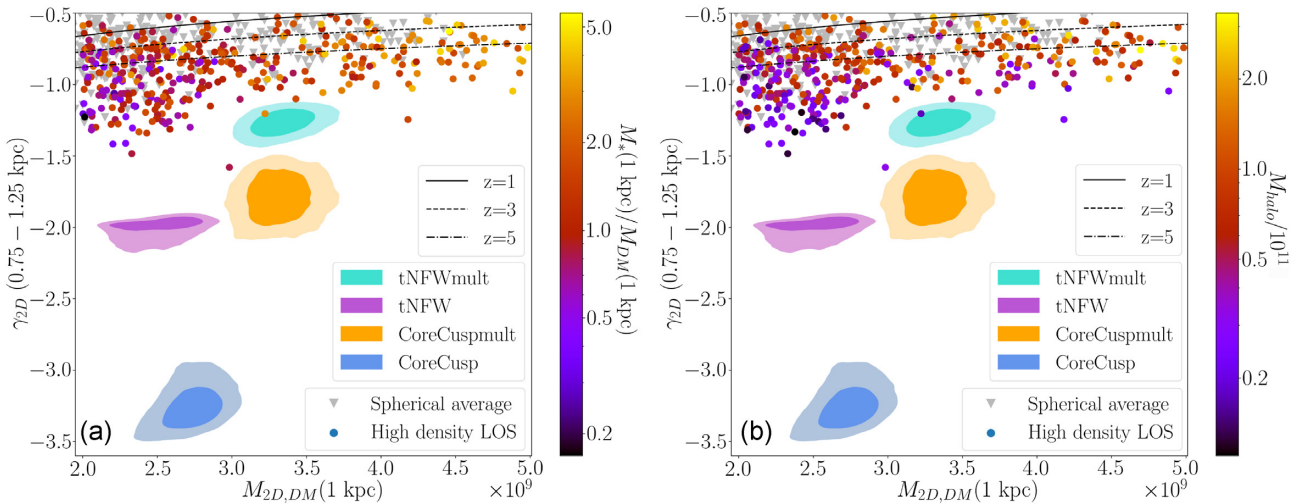

Figure B1. The same as Fig. 9, except here we plot the contours for all four models: tNFW and CoreCusp, both the elliptical case as well as with multipoles. Here the points from the high density LOS average are colour coded by $M_*(1\text{ kpc})/M_{\text{DM}}(1\text{ kpc})$ in panel (a), and by total halo mass in panel (b).
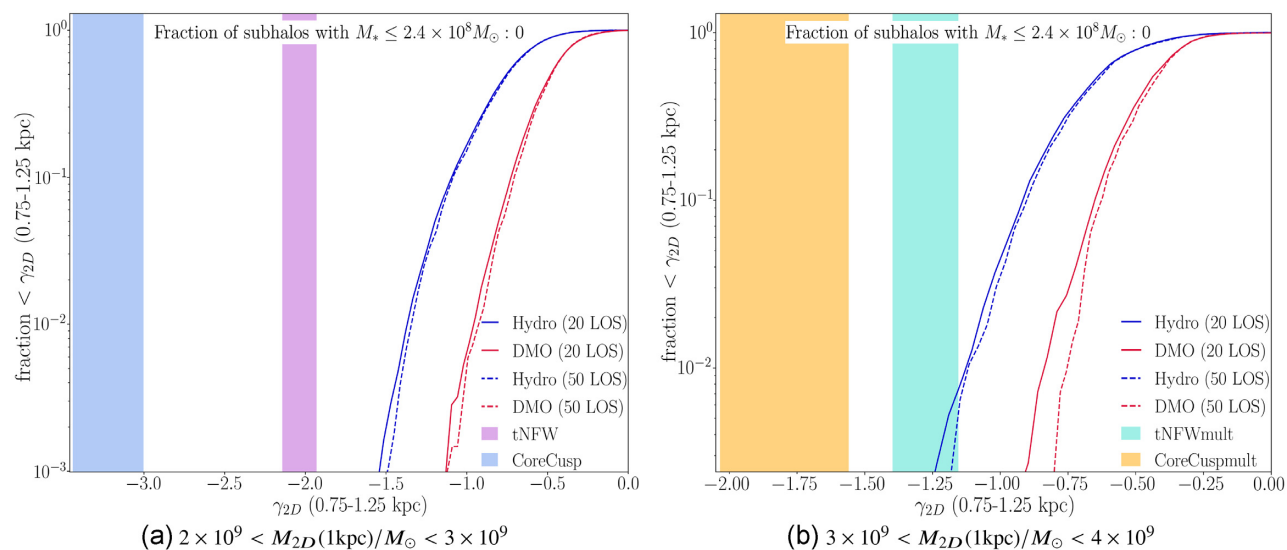
Figure B2. The same as Fig. 10, except we plot two different ranges in $M_{2D}(1\,\mathrm{kpc})$: figure (a) plots the $2–3 \times 10^9\,\mathrm{M}_\odot$ range favoured by the elliptical tNFW and CoreCusp models, while figure (b) plots the $3–4 \times 10^9\,\mathrm{M}_\odot$ range favoured by the models that include multipoles (tNFWmult and CoreCuspmult). Vertical bars representing the 95 per cent credible interval for each lens model are overlaid. We plot the cumulative distribution in γ_{2D} for two different levels of averaging: for the solid curves, each mock 'observation' is produced by averaging the profiles for 20 nearby LOS (out of 1000 total for each subhalo), while for the dashed curves, 50 LOS are averaged. Note that for all but one of the lens models, less than 0.1 per cent of observations produce density slopes consistent with the lensing constraints; the only exception is tNFWmult, our most conservative model, for which the fraction is < 1 per cent.

posteriors in the subhalo parameters for the tNFWmult model. The joint posteriors in $r_{\delta c}$ versus $x_{c,\,sub}$ and $y_{c,\,sub}$ show that if the subhalo is moved toward higher x and/or lower y, $r_{\delta c}$ increases; the same relationship is evident in all the models we fit. A similar covariance with the centre coordinates is seen with $M_{2D}(1\,\mathrm{kpc})$ as well. However, as all models agree on the size of the critical curve perturbation itself (Fig. B3), we consider it unlikely that a model could produce a much smaller $r_{\delta c}$ without degrading the fit. This is supported by the results of Section 5: when the perturber is modeled as a field halo along the LOS instead of a subhalo, for sufficiently low redshift ($z \lesssim 0.1$) the goodness-of-fit plummets as the best-fitting $M_{2D}(1\,\mathrm{kpc})$ drops below $2 \times 10^9\,\mathrm{M}_\odot$ (see Fig. 11).
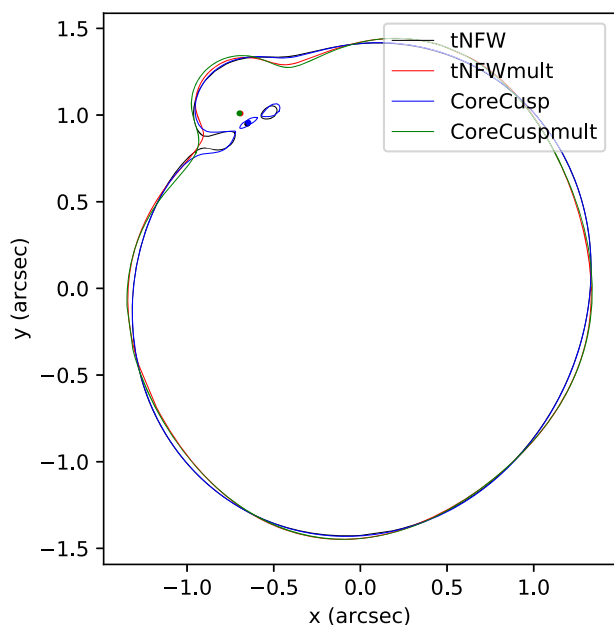
Figure B3. Critical curves for all four lens models we fit to SDSSJ0946+1006, whose inferred parameters are listed in Table 1 The best-fitting position of the subhalo's centre for each model is shown by the filled circles, whose colours match the corresponding critical curve colours. Note the size of the subhalo's perturbation to the critical curve (in the upper left) is fairly consistent in all four models. However, the position of the subhalo differs noticeably in the models that include multipoles compared to those that do not; as a result, the inferred perturbation radius differs by up to roughly 0.05 arcsec, the width of an image pixel.
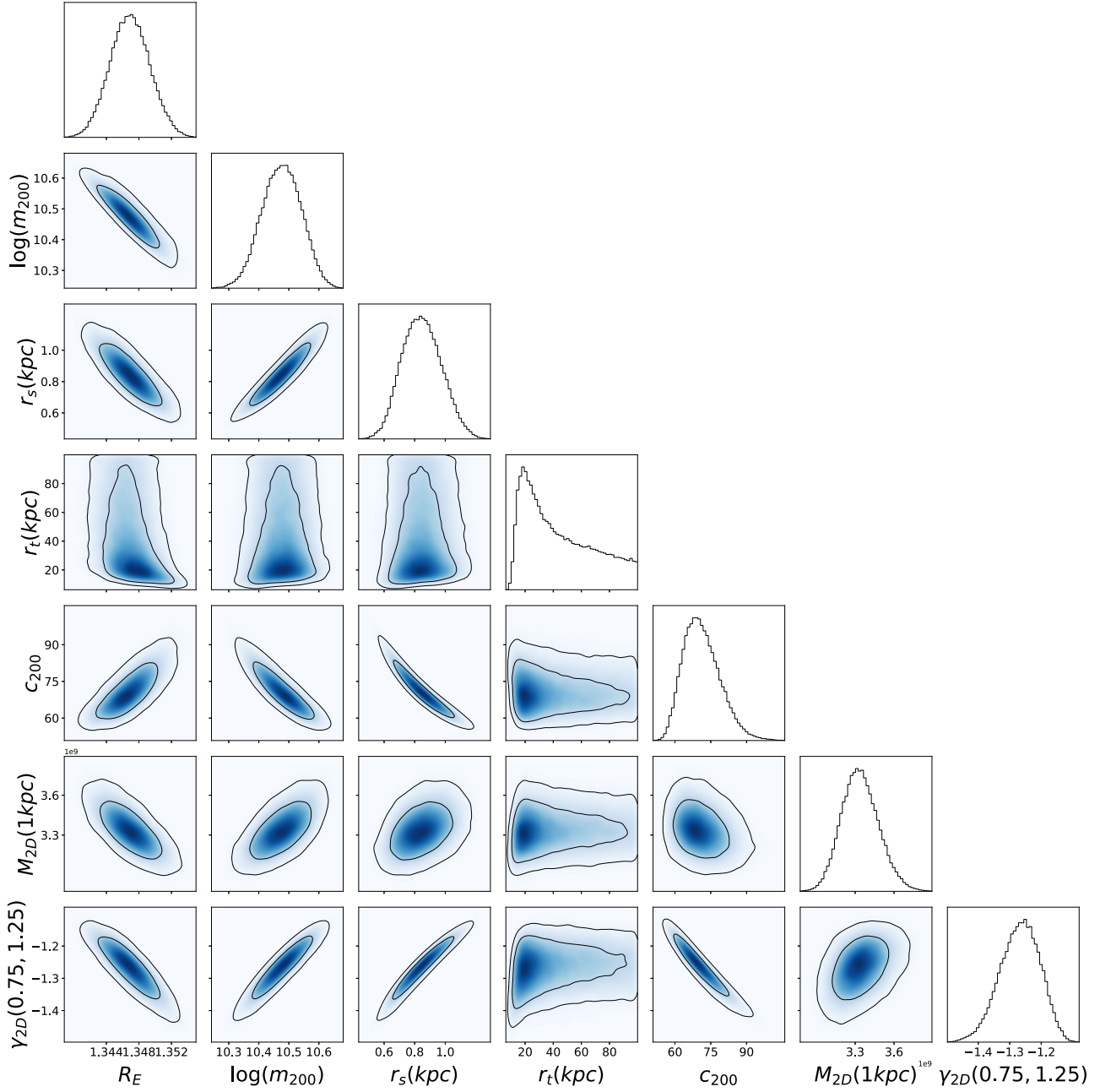
Figure B4. Posteriors in the subhalo lensing parameters for model tNFWmult, whose uncertainties are tabulated in Table 1. The host galaxy parameters are omitted here except for R_E, the Einstein radius parameter; we have also omitted the subhalo position parameters, which have very little degeneracy with the other parameters, to avoid clutter. The structural parameters of the subhalo r_s, r_t are given in kpc. Note that c_{200}, $M_{2D}(1\,\text{kpc})$, and $\gamma_{2D}(0.75, 1.25)$ are both derived parameters.

APPENDIX C: VARIATION OF BEST-FITTING SUBHALO MASS AND PERTURBATION RADIUS WITH REDSHIFT FOR TNFW, TNFWMULT MODELS

In Fig. C1 we plot the $M_{2D}(1\,\text{kpc})$ and χ^2 as the perturber redshift is varied, starting from the tNFWmult model. Note that, as in Fig. 11, the fit becomes significantly degraded as the perturber mass drops down to $2 \times 10^9\,\text{M}_\odot$ for low redshift. Likewise, the required slope becomes very steep at low redshift.

It is interesting to consider how the perturbation radius of the subhalo in SDSSJ0946+1006 changes as a function of redshift, since this has some bearing on whether $M_{2D}(1\,\text{kpc})$ can be considered a robust quantity. In Fig. C2 we plot the perturbation radius in units of kpc (solid line, left axis) and arcsec (dashed line, right axis). Note that at low redshift, the angular size of the perturbation remains approximately the same, while the physical size becomes smaller. This is easily understood, since the angular diameter distance becomes smaller as the redshift is reduced. Naively one might expect the mass within the (angular) perturbation radius is preserved, and hence there would be a greater overall mass within 1 kpc since this corresponds to a bigger angular size at low redshift. However, as discussed above, the lensing efficiency of the perturber increases at lower redshift, requiring a smaller mass to keep the same perturbation scale, and this is the dominant effect, as Fig. C2 shows.

The change in the perturbation at high redshift is more subtle: the size of the perturbation decreases as the perturber is moved behind the lens, in both physical and angular units. The decreased size in kpc, in particular, defies the naive expectation that the larger angular diameter distance leads to a greater physical size for the perturbation radius. This is a result of the recursive lensing that gradually smears the perturber's convergence into an arc, and requires the perturber to be closer to the critical curve for an optimal fit. (A secondary effect is that the dashed line is showing the unlensed angular size of the
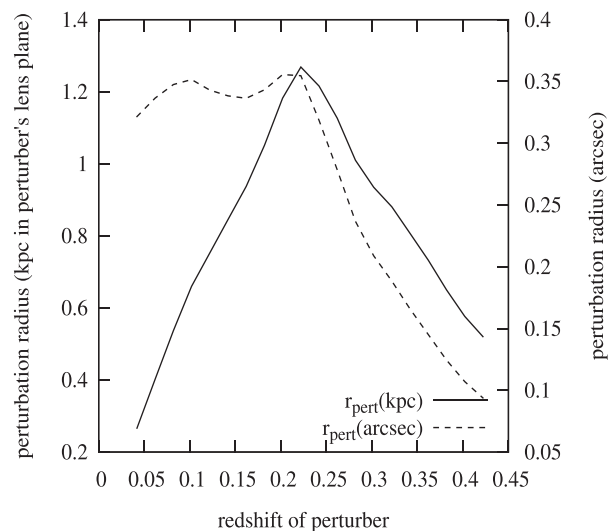

Figure C2. The perturber's perturbation radius (defined in Section 3.3) as a function of redshift. The perturbation radius is given in kpc (solid curve) and arcsec (dashed curve, axis on right). These plots were generated by taking the best-fitting model with tNFW profile, gradually varying the redshift in either direction with steps of $\Delta z = 0.02$ and reoptimizing the χ^2 function.

perturbation radius, i.e. the angular size it maps to in the perturber's lens plane, which is slightly smaller than the observed angular size after being lensed by the primary lens galaxy.) Note that outside the redshift range from 0.15 to 0.35, the perturbation radius dips below 0.8 kpc, hence the projected mass within 1 kpc may be less robustly determined; on the other hand, we have seen in Fig. 11(a) that the fit becomes severely degraded outside this range, so this is less of a concern. Over this redshift range, the projected mass within 1 kpc is approximately constant and greater than $2 \times 10^9\,\text{M}_\odot$.

(a) perturber mass within 1 kpc
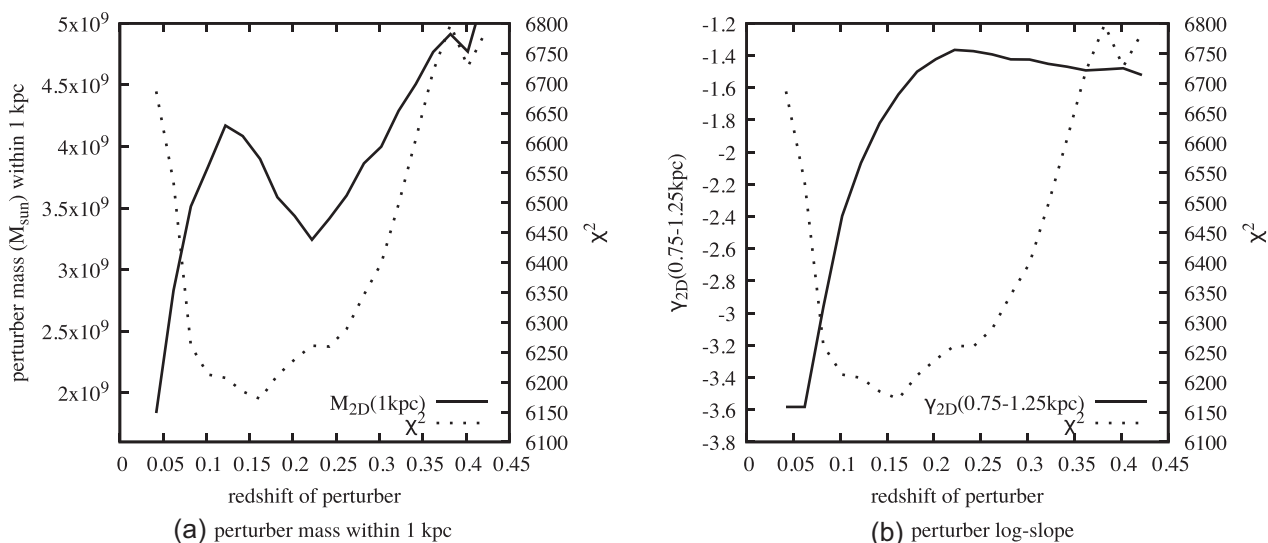

(b) perturber log-slope

Figure C1. Same as Fig. 11, except using the tNFWmult model instead of tNFW. Note that as with the tNFW model, the log-slope becomes very steep in the limit of low redshift. Likewise, the perturber mass drops below $2 \times 10^9\,\text{M}_\odot$ for low redshift, but the fit becomes significantly degraded in this limit.

This paper has been typeset from a TEX/LATEX file prepared by the author.