


























# STRIDES: a 3.9 per cent measurement of the Hubble constant from the strong lens system DES J0408–5354

A. J. Shajib <sup>1</sup>★ S. Birrer,<sup>1,2</sup> T. Treu,<sup>1†</sup> A. Agnello,<sup>3</sup> E. J. Buckley-Geer,<sup>4</sup> J. H. H. Chan <sup>5</sup>, L. Christensen,<sup>3</sup> C. Lemon <sup>5</sup>, H. Lin,<sup>4</sup> M. Millon,<sup>5</sup> J. Poh,<sup>6,7</sup> C. E. Rusu <sup>8</sup>, D. Sluse,<sup>9</sup> C. Spiniello <sup>10,11</sup>, G. C.-F. Chen <sup>12</sup>, T. Collett <sup>13</sup>, F. Courbin,<sup>5</sup> C. D. Fassnacht,<sup>12</sup> J. Frieman,<sup>4,7</sup> A. Galan,<sup>5</sup> D. Gilman,<sup>1</sup> A. More <sup>14</sup>, T. Anguita <sup>15</sup>, M. W. Auger,<sup>16,17</sup> V. Bonvin,<sup>5</sup> R. McMahon,<sup>16,17</sup> G. Meylan,<sup>5</sup> K. C. Wong <sup>18</sup>, T. M. C. Abbott,<sup>19</sup> J. Annis,<sup>4</sup> S. Avila <sup>20</sup>, K. Bechtol,<sup>21,22</sup> D. Brooks,<sup>23</sup> D. Brout,<sup>24</sup> D. L. Burke,<sup>25,26</sup> A. Carnero Rosell <sup>27,28</sup>, M. Carrasco Kind <sup>29,30</sup>, J. Carretero,<sup>31</sup> F. J. Castander,<sup>32,33</sup> M. Costanzi,<sup>34,35</sup> L. N. da Costa,<sup>28,36</sup> J. De Vicente <sup>27</sup>, S. Desai,<sup>37</sup> J. P. Dietrich,<sup>38,39</sup> P. Doel,<sup>23</sup> A. Drlica-Wagner,<sup>4,7</sup> A. E. Evrard <sup>40,41</sup>, D. A. Finley,<sup>4</sup> B. Flaugher,<sup>4</sup> P. Fosalba,<sup>32,33</sup> J. García-Bellido,<sup>42</sup> D. W. Gerdes,<sup>40,41</sup> D. Gruen <sup>25,26,43</sup>, R. A. Gruendl,<sup>29,30</sup> J. Gschwend,<sup>28,36</sup> G. Gutierrez,<sup>4</sup> D. L. Hollowood,<sup>44</sup> K. Honscheid,<sup>45,46</sup> D. Huterer,<sup>41</sup> D. J. James,<sup>47</sup> T. Jeltema,<sup>44</sup> E. Krause,<sup>48</sup> N. Kuropatkin,<sup>4</sup> T. S. Li <sup>4,7,49,50</sup>, M. Lima,<sup>28,51</sup> N. MacCrann <sup>45,46</sup>, M. A. G. Maia,<sup>28,36</sup> J. L. Marshall,<sup>52</sup> P. Melchior <sup>49</sup>, R. Miquel,<sup>31,53</sup> R. L. C. Ogando,<sup>28,36</sup> A. Palmese <sup>4</sup>, F. Paz-Chinchón,<sup>29,30</sup> A. A. Plazas <sup>49</sup>, A. K. Romer,<sup>54</sup> A. Roodman,<sup>25,26</sup> M. Sako,<sup>24</sup> E. Sanchez,<sup>27</sup> B. Santiago,<sup>28,55</sup> V. Scarpine,<sup>4</sup> M. Schubnell,<sup>41</sup> D. Scolnic,<sup>56</sup> S. Serrano,<sup>33,57</sup> I. Sevilla-Noarbe,<sup>27</sup> M. Smith <sup>58</sup>, M. Soares-Santos <sup>59</sup>, E. Suchyta <sup>60</sup>, G. Tarle,<sup>41</sup> D. Thomas <sup>13</sup>, A. R. Walker<sup>19</sup> and Y. Zhang<sup>4</sup>

Affiliations are listed at the end of the paper

Accepted 2020 March 18. Received 2020 February 24; in original form 2019 October 14

## ABSTRACT

We present a blind time-delay cosmographic analysis for the lens system DES J0408–5354. This system is extraordinary for the presence of two sets of multiple images at different redshifts, which provide the opportunity to obtain more information at the cost of increased modelling complexity with respect to previously analysed systems. We perform detailed modelling of the mass distribution for this lens system using three band *Hubble Space Telescope* imaging. We combine the measured time delays, line-of-sight central velocity dispersion of the deflector, and statistically constrained external convergence with our lens models to estimate two cosmological distances. We measure the ‘effective’ time-delay distance corresponding to the redshifts of the deflector and the lensed quasar  $D_{\Delta t}^{\text{eff}} = 3382_{-115}^{+146}$  Mpc and the angular diameter distance to the deflector  $D_d = 1711_{-280}^{+376}$  Mpc, with covariance between the two distances. From these constraints on the cosmological distances, we infer the Hubble constant  $H_0 = 74.2_{-3.0}^{+2.7}$  km s<sup>-1</sup> Mpc<sup>-1</sup> assuming a flat  $\Lambda$ CDM cosmology and a uniform prior for  $\Omega_m$  as  $\Omega_m \sim \mathcal{U}(0.05, 0.5)$ . This measurement gives the most precise constraint on  $H_0$  to date from a single lens. Our measurement is consistent with that obtained from the previous sample of six

\* E-mail: [ajshajib@astro.ucla.edu](mailto:ajshajib@astro.ucla.edu)

† Packard Fellow

lenses analysed by the  $H_0$  Lenses in COSMOGRAIL’s Wellspring (H0LiCOW) collaboration. It is also consistent with measurements of  $H_0$  based on the local distance ladder, reinforcing the tension with the inference from early Universe probes, for example, with  $2.2\sigma$  discrepancy from the cosmic microwave background measurement.

**Key words:** gravitational lensing; strong–cosmological parameters–distance scale–cosmology; observations.

## 1 INTRODUCTION

The concordance  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmology explains the accelerated expansion of the Universe by incorporating the cosmological constant  $\Lambda$  (Riess et al. 1998; Perlmutter et al. 1999). The  $\Lambda$ CDM model is very successful in predicting observations covering a large range of physical scales – from the scale of sound horizon at the recombination epoch, down to the structure formation at the megaparsec scale (e.g. Alam et al. 2017; Abbott et al. 2018; Planck Collaboration VI 2018). The Hubble constant,  $H_0$ , plays a central role in cosmology, including in the  $\Lambda$ CDM model. The Hubble constant is not only crucial to determine the age of the Universe, it also normalizes the distances to distant galaxies. As a result, a precise understanding of the galaxy formation and evolution, and the Universe as a whole, closely depends on the precise knowledge of the Hubble constant.

Recently, a significant tension has been reported between the measurements of the Hubble constant using early-Universe and late-Universe probes (e.g. Planck Collaboration VI 2018; Riess et al. 2019; Wong et al. 2020). Among others, the most precise constraints on the Hubble constant come from extrapolating the cosmic microwave background (CMB) observation at the early-Universe, and from the measurement based on the cosmic distance ladder calibrated with parallax distances, Cepheids, and type Ia supernovae (SNIae). Assuming a  $\Lambda$ CDM cosmology, the *Planck* measurement gives  $H_0 = 67.4 \pm 0.5 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (Planck Collaboration VI 2018). The Supernovae,  $H_0$ , for the Equation of State of dark energy (SH0ES) team measures  $H_0 = 74.03 \pm 1.42 \text{ km s}^{-1} \text{ Mpc}^{-1}$  by calibrating the SNIa distance ladder using Cepheids and parallax distances (Riess et al. 2019). These two measurements are at  $4.4\sigma$  tension. A cosmic distance ladder measurement from the Carnegie–Chicago Hubble project calibrated by the tip of the red giant branch (TRGB) stars reports  $H_0 = 69.8 \pm 1.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , consistent with both of the above values on opposite sides (Freedman et al. 2019). However, the SH0ES team finds  $H_0 = 72.4 \pm 1.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$  using the TRGB stars to calibrate the SNIae distance ladder (Yuan et al. 2019). Additional probes, all consistent with the tension at varying degrees of significance are summarized by Verde, Treu & Riess (2019). This tension between the early-Universe and the late-Universe probes can be due to unknown systematics in any or all of the probes. However, if systematics can be ruled out as the source of this tension, then this tension would require extension of the  $\Lambda$ CDM model. In order to reach a conclusion on the tension and whether new physics is needed, it is paramount to have multiple independent measurements of the Hubble constant, each with sufficient precision on its own to resolve the discrepancy ( $<2$  per cent). In parallel it is also crucial to investigate in detail all possible sources of systematic uncertainties in each method.

Time-delay cosmography measures  $H_0$  and other cosmological parameters independently of both the CMB or other high-redshift observations and the local probes such as the ones using the cosmic distance ladder (Refsdal 1964). The time-delay between

the arrival time of photons at multiple images of a strong-lensing system (hereafter, lens) depends on the three angular diameter distances – between the observer and the deflector, between the deflector and the source, and between the observer and the source. A combination of these three angular diameter distances gives the so-called ‘time-delay distance’ (Suyu et al. 2010). This time-delay distance is inversely proportional to  $H_0$  and thus measuring this distance directly constrains  $H_0$ .

To measure the time delay between the arrivals of photon at different lensed images that were emitted at the same time, we require a time-variable source. Although Refsdal (1964) originally proposed using strongly lensed supernovae as a time-variable source to measure the time-delay, only a few such supernovae have been discovered so far (e.g. Kelly et al. 2015; Goobar et al. 2017; Grillo et al. 2018). Even though the number of lensed supernova is still too small to be a competitive cosmological probe, the re-appearance of supernova Refsdal as predicted provides an important validation of the method (Treu et al. 2016).

Strongly lensed quasars have provided time-variable sources in larger numbers. As a result, these objects have been predominantly used to measure  $H_0$  from their time-delays (e.g. Schechter et al. 1997; Treu & Koopmans 2002; Suyu et al. 2010). Although some of the early measurements had shortcomings in the data quality or the analysis technique, both of these aspects have tremendously improved over the past decade (for a review with historical perspective, see Treu & Marshall 2016). The key breakthroughs in the past two decades have been: (i) high cadence monitoring to determine the time delays (e.g. Fassnacht et al. 2002; Tewes et al. 2013), (ii) high resolution images of the lensed arcs from the quasar host galaxy and pixel-based lens modelling to constrain the lens mass distribution (Suyu et al. 2010), (iii) adding stellar kinematics of the deflector (Treu & Koopmans 2002), and (iv) statistical analysis of the line of sight to constrain the external convergence (Suyu et al. 2010; Greene et al. 2013; Rusu et al. 2017). Implementing these improvements, the  $H_0$  Lenses in COSMOGRAIL’s Wellspring (H0LiCOW) collaboration measure  $H_0 = 73.3_{-1.8}^{+1.7} \text{ km s}^{-1} \text{ Mpc}^{-1}$  using six lens systems (Suyu et al. 2010, 2013, 2014; Wong et al. 2017; Bonvin et al. 2017; Birrer et al. 2019; Chen et al. 2019; Rusu et al. 2020; Wong et al. 2020).

To reach 1 per cent precision in the Hubble constant with time-delay cosmography, a sample of  $\sim 40$  lenses is necessary (Shajib, Treu & Agnello 2018). To have such a large sample of strongly lensed quasars available in the first place, the STRong-lensing Insights into Dark Energy Survey collaboration (STRIDES; Treu et al. 2018) has discovered numerous new lenses from the Dark Energy Survey (DES) footprint, in cases combining data from other large-area sky surveys (e.g. Agnello et al. 2015b; Nord et al. 2016; Ostrovski et al. 2017; Agnello et al. 2018b; Anguita et al. 2018; Lemon et al. 2020). The STRIDES is an external collaboration of the DES. The DES data are particularly useful in discovering new lenses due to its combination of uniform depth and coverage of area in the Southern hemisphere that is not covered by the Sloan Digital Sky Survey (SDSS). Additionally, thanks to new

data mining and machine learning based techniques, new lenses have been discovered also from other photometric surveys – such as the VLT Survey Telescope-ATLAS (VST-ATLAS), Kilo-Degree Survey Strongly lensed Quasar Detection project (KiDS-SQuAD) (e.g. Agnello et al. 2015a, 2018a; Spiniello et al. 2018).

In this paper, we present a blind analysis of the lens system DES J0408–5354 and infer  $H_0$  from its time delays (Courbin et al. 2018). This lens was discovered in the DES footprint (Diehl et al. 2017; Lin et al. 2017). This paper sets down two goals underlying our analysis. First, we aim to increase the statistical precision of the  $H_0$  determination by presenting results from the analysis of a new lens system. Secondly, this system is being analysed independently and in parallel by two teams using two different codes in order to estimate potential systematics arising from modelling choices and software. This paper presents the first of these two independent and blind analyses for DES J0408–5354. To facilitate meaningful comparison between independent modelling teams, the participating teams agreed beforehand on a set of baseline models with minimal but sufficient specifications. The teams are free to extend on the baseline models for exploring different sources of systematics as they see fit. This additional exploration by a team proceeds independently while keeping the cosmographic inferences blind. In this way, we aim to check on systematics that can potentially arise from different codes through comparison of the baseline models from different teams, and also from different model choices within one team’s analysis. In this paper, we infer  $H_0$  using the lens modelling software LENSTRONOMY, which is publicly available online at GitHub<sup>1</sup> (Birrer, Amara & Refregier 2015; Birrer & Amara 2018). A second independent team uses the software GLEE to analyse the same lens system (Suyu & Halkola 2010). In a future paper, cosmographic inference based on this second analysis and a comparison between the two analyses will be presented (Yıldırım et al., in preparation). Both of the independent modelling works use the results from a companion paper, which analyses the lens environment to detect galaxy groups and estimate the external convergence using the DES data, and measure the stellar kinematics of the central deflector galaxy from spectroscopic observations (Buckley-Geer et al. 2020).

Our concerted effort to analyse a system independently but based on the same data, and with some overlap in modelling choices, is an important step forward in estimating the modelling errors with respect to previous works. Previous efforts by the H0LiCOW and Strong-lensing High Angular Resolution Programme (SHARP) collaborations took some step in this direction by assigning different lead investigators and softwares to the analysis of the six lenses (Lagattuta et al. 2012; Suyu et al. 2017). The lens systems B1608+656, RXJ1131–1231, HE 0435–1223, WFI2033–4723, and PG 1115+080 were analysed using the lens modelling software GLEE, whereas the systems RXJ1131–1231 and SDSS 1206+4332 were analysed using the software LENSTRONOMY (Suyu et al. 2010, 2013; Birrer, Amara & Refregier 2016; Wong et al. 2017; Birrer et al. 2019; Chen et al. 2019; Rusu et al. 2020). In total, four different lead investigators modelled these six lenses, even though there was overlap between the team members. The two softwares used in the modelling differ in various aspects. For example, LENSTRONOMY performs source-reconstruction using a basis set of shapelets, whereas GLEE performs a pixelized source-reconstruction with regularization. LENSTRONOMY is a publicly available open-source software, whereas GLEE is not.

In order to preserve the blindness of the analysis, this paper and the companion describing the analysis of the environment and line of sight used to compute the external convergence were internally reviewed by the STRIDES collaboration and the DES strong-lensing working group before unblinding. Once both the analyses and manuscripts met the approval of the internal reviewers and co-authors, unblinding happened on 2019 September 25. After unblinding, the only changes to the manuscript were the addition of the unblinded measurements, discussion on the unblinded results in Section 7, minor editing for clarity, grammar, and typos after the DES collaboration-wide review, and the addition of the plot showing the galaxy group’s convergence described in Appendix C.

This paper is organized as follows. In Section 2, we lay out the necessary formalism and describe the analysis framework. We present the data sets used in our analysis in Section 3. Next in Section 4, we describe the different mass and light profiles that are used in the lens modelling. We present the various lens model choices in Section 5. We report the results from the lens modelling and the cosmographic inference in Section 6. Finally, we discuss the results and summarize the paper in Section 7. We provide summaries of the uncertainty budget in our inferred  $H_0$ , systematic checks, adopted models, and parameter priors in Appendices F and G. The reported uncertainties in this paper are computed from 16th and 84th percentiles of the posterior probability distribution.

## 2 FRAMEWORK OF THE COSMOGRAPHIC ANALYSIS

In this section, we outline our cosmographic analysis using strong-lensing time delays. We briefly lay out the strong-lensing time-delay formalism in Section 2.1, discuss the lensing degeneracies in Section 2.2, present an overview of the kinematic analysis in Section 2.3, describe the cosmological analysis in Section 2.4, and formalize the underlying Bayesian inference framework of our analysis in Section 2.5.

### 2.1 Strong-lensing time delay

The framework described in this subsection was developed in previous studies – e.g. see Schneider, Ehlers & Falco (1992), Blandford & Narayan (1992) – and was applied in previous studies to the measure  $H_0$  from time delays (e.g. Suyu et al. 2010; Wong et al. 2017; Birrer et al. 2019).

The time delay  $\Delta t_{XY}$  between arrival of photons at two images, indexed with X and Y, of a multiply-imaged quasar by a single deflector is given by

$$\Delta t_{XY} = \frac{1+z_d}{c} \frac{D_d D_s}{D_{ds}} \left[ \frac{(\theta_X - \beta)^2}{2} - \frac{(\theta_Y - \beta)^2}{2} - \psi(\theta_X) + \psi(\theta_Y) \right]. \quad (1)$$

Here, the three angular diameter distances are  $D_d$ : between the observer and the deflector,  $D_s$ : between the observer and the source, and  $D_{ds}$ : between the deflector and the source. Additionally,  $z_d$  is the deflector redshift,  $c$  is the speed of light,  $\theta$  is the image position,  $\beta$  is the source position, and  $\psi$  is the deflection potential. The deflection potential is defined such that its gradient gives the deflection field  $\alpha \equiv \nabla \psi$ . Then, the deflection potential relates to the convergence  $\kappa$  as  $\kappa \equiv \nabla^2 \psi / 2$ . We define the Fermat potential  $\phi$  as

$$\phi(\theta) \equiv \frac{(\theta - \beta)^2}{2} - \psi(\theta), \quad (2)$$

<sup>1</sup><https://github.com/sibirrer/lenstronomy>

and the time-delay distance as

$$D_{\Delta t} \equiv (1 + z_d) \frac{D_d D_s}{D_{ds}}. \quad (3)$$

Then, we can express equation (1) in a more compact form as

$$\Delta t_{XY} = \frac{D_{\Delta t}}{c} (\phi(\theta_X) - \phi(\theta_Y)) \equiv \frac{D_{\Delta t}}{c} \Delta \phi_{XY}. \quad (4)$$

If multiple deflectors are present at close angular proximity at different redshifts, then we need to use the multilens-plane formalism to describe the lensing effect with sufficient accuracy. The time delay between two images for the case of lensing with  $P$  lens planes can be obtained by tracing the lensed light-ray backward from the image plane to the source plane as

$$\Delta t_{XY} = \sum_{i=1}^P \frac{D_{\Delta t, i, i+1}}{c} \left[ \frac{(\theta_{X, i} - \theta_{X, i+1})^2}{2} - \frac{(\theta_{Y, i} - \theta_{Y, i+1})^2}{2} - \zeta_{i, i+1} \{ \psi_i(\theta_{X, i}) - \psi_i(\theta_{Y, i}) \} \right] \quad (5)$$

(cf. equation 9.17 of Schneider et al. 1992). Here, the first lens plane is the nearest to the observer and the  $(P + 1)$ -th plane refers to the source plane. The time-delay distance  $D_{\Delta t, i, j}$  between a pair of planes is defined as

$$D_{\Delta t, i, j} \equiv \frac{1 + z_i}{c} \frac{D_i D_j}{D_{ij}}, \quad i < j, \quad (6)$$

where  $D_i$  is the angular diameter distance from the observer to the  $i$ th plane and  $D_{ij}$  is the angular diameter distance between the  $i$ th and  $j$ th planes. The rescaling factor  $\zeta_{i, j}$  is defined as

$$\zeta_{i, j} \equiv \frac{D_{ij} D_s}{D_j D_{is}}, \quad i < j. \quad (7)$$

In this multilens-plane case, we can define the time-delay distance between the central deflector plane and the source plane as the effective time-delay distance  $D_{\Delta t}^{\text{eff}} \equiv D_{\Delta t, d, s}$  that normalizes the multilens-plane time delay as

$$\begin{aligned} \Delta t_{XY} &= \frac{D_{\Delta t}^{\text{eff}}}{c} \sum_{i=1}^P \frac{1 + z_i}{1 + z_d} \frac{D_i D_{i+1} D_{ds}}{D_d D_s D_{i+1}} \left[ \frac{(\theta_{X, i} - \theta_{X, i+1})^2}{2} \right. \\ &\quad \left. - \frac{(\theta_{Y, i} - \theta_{Y, i+1})^2}{2} - \zeta_{i, i+1} \{ \psi_i(\theta_{X, i}) - \psi_i(\theta_{Y, i}) \} \right] \\ &= \frac{D_{\Delta t}^{\text{eff}}}{c} \Delta \phi_{XY}^{\text{eff}}. \end{aligned} \quad (8)$$

Here, we defined the effective Fermat potential for the multilens-plane case as

$$\phi^{\text{eff}}(\theta) \equiv \sum_{i=1}^P \frac{1 + z_i}{1 + z_d} \frac{D_i D_{i+1} D_{ds}}{D_d D_s D_{i+1}} \left[ \frac{(\theta_i - \theta_{i+1})^2}{2} - \zeta_{i, i+1} \psi_i(\theta_i) \right]. \quad (9)$$

In equation (8), the effective Fermat potential difference  $\Delta \phi_{XY}^{\text{eff}}$  only contains the distance ratios. Thus, this term does not depend on  $H_0$ . However, the distance ratios weakly depend on the relative expansion history, thus on the density parameter  $\Omega$  in the context of  $\Lambda$ CDM. Only the effective time-delay distance  $D_{\Delta t}^{\text{eff}}$  depends on  $H_0$  in equation (8). For the single lens plane case with  $P = 1$ , the effective Fermat potential  $\phi^{\text{eff}}$  and the effective time-delay distance  $D_{\Delta t}^{\text{eff}}$  naturally take the form of their single-lens-plane equivalents  $\phi$  and  $D_{\Delta t}$  from equations (2) and (3), respectively.

## 2.2 Mass-sheet degeneracy

For lensing, the imaging observables such as the flux ratios and the relative astrometry are invariant with respect to the mass-sheet transformation (MST; Falco, Gorenstein & Shapiro 1985). If we transform the convergence and the source plane as

$$\begin{aligned} \kappa(\theta) &\rightarrow \kappa_{\vartheta}(\theta) = \vartheta \kappa(\theta) + 1 - \vartheta, \\ \beta &\rightarrow \beta' = \vartheta \beta, \end{aligned} \quad (10)$$

then the lensing observables except the time delay remain invariant. This invariance under the MST is called the mass-sheet degeneracy (MSD). Notably, the MST also rescales the magnification, thus the MSD can be broken with standard candles (Bertin & Lombardi 2006).

We can express the physically existent ‘true’ mass distribution as

$$\kappa_{\text{true}} = \kappa_{\text{lens}} + \kappa_{\text{ext}}, \quad (11)$$

where,  $\kappa_{\text{lens}}$  is the convergence of the central deflector including nearby satellites and perturbers, and  $\kappa_{\text{ext}}$  is the convergence from projecting all the line-of-sight inhomogeneities on to the lens plane. If we impose the condition  $\lim_{\theta \rightarrow \infty} \kappa_{\text{lens}} = 0$ , then we have  $\lim_{\theta \rightarrow \infty} \kappa_{\text{true}} = \kappa_{\text{ext}}$ . As a result, we can interpret the external convergence  $\kappa_{\text{ext}}$  as the convergence far from or ‘external’ to the central deflector. However, as we cannot constrain  $\kappa_{\text{ext}}$  only from the lensing observables due to the MSD, we aim to constrain a model  $\kappa'_{\text{model}}$  that captures all the lensing effects of  $\kappa_{\text{true}}$ . By taking  $\vartheta = 1/(1 - \kappa_{\text{ext}})$  in equation (10), we can obtain an MST of  $\kappa_{\text{true}}$  as

$$\kappa_{\vartheta} = \frac{1}{1 - \kappa_{\text{ext}}} (\kappa_{\text{lens}} + \kappa_{\text{ext}}) - \frac{\kappa_{\text{ext}}}{1 - \kappa_{\text{ext}}} = \frac{\kappa_{\text{lens}}}{1 - \kappa_{\text{ext}}} = \kappa'_{\text{model}}. \quad (12)$$

Here, we name this  $\kappa_{\vartheta}$  as  $\kappa'_{\text{model}}$  because it captures all the lensing effect of  $\kappa_{\text{true}}$  by the virtue of MST. If we can constrain  $\kappa'_{\text{model}}$ , then we can obtain  $\kappa_{\text{true}}$  simply through an MST with  $\vartheta = 1 - \kappa_{\text{ext}}$  where  $\kappa_{\text{ext}}$  is separately constrained by studying the lens environment. However, the lens model  $\kappa_{\text{model}}$  that we actually constrain can potentially be an internal MST of  $\kappa'_{\text{model}}$  given by

$$\kappa'_{\text{model}} = \vartheta_{\text{int}} \kappa_{\text{model}} + 1 - \vartheta_{\text{int}}. \quad (13)$$

The internal MST factor  $\vartheta_{\text{int}}$  only changes the shape of the mass profile, but it does not add any physical mass to the model within the Einstein radius. Note that both  $\kappa_{\text{model}}$  and  $\kappa'_{\text{model}}$  can satisfy  $\lim_{\theta \rightarrow \infty} \kappa = 0$  by construction. In that case,  $\vartheta_{\text{int}}$  is not a constant over the whole plane and we have the condition  $\lim_{\theta \rightarrow \infty} \vartheta_{\text{int}} = 1$  (Schneider & Sluse 2014). This condition implies that  $\vartheta_{\text{int}}$  does not physically add an infinite background-mass-sheet. With such a  $\vartheta_{\text{int}}$ , both models  $\kappa_{\text{model}}$  and  $\kappa'_{\text{model}}$  can reproduce the lensing observables that are indistinguishable within the noise level in the data. Finally, combining equations (11), (12), and (13), we write the relation between the ‘true’ convergence  $\kappa_{\text{true}}$  and the modelled convergence  $\kappa_{\text{model}}$  as

$$\kappa_{\text{true}} = (1 - \kappa_{\text{ext}}) [\vartheta_{\text{int}} \kappa_{\text{model}} + 1 - \vartheta_{\text{int}}] + \kappa_{\text{ext}}. \quad (14)$$

Using different but equally plausible model parametrizations – e.g. power-law profile, composite profile – we explore different model families related by equation (13). To alleviate the MSD within a model family by constraining  $\vartheta_{\text{int}}$ , we utilize non-lensing observables, e.g. kinematics of the deflector galaxy. Kinematics probes the 3D deprojection of  $\kappa_{\text{lens}}$  for a given combination of  $\kappa_{\text{model}}$  and  $\kappa_{\text{ext}}$ . Moreover, the addition of the kinematic information also constrains the angular diameter distance to the deflector  $D_d$



(Paraficz & Hjorth 2009; Jee, Komatsu & Suyu 2015). As a result, the uncertainty on the estimated  $H_0$  is improved by kinematics in two ways:

- (i) by alleviating the MSD, and
- (ii) by adding extra constraint on cosmology through  $D_d$

(Birrer et al. 2016; Jee et al. 2016; Shajib et al. 2018). In the next subsection, we outline the kinematic analysis framework.

### 2.3 Kinematic analysis

The kinematic observable is the luminosity-weighted line-of-sight stellar velocity dispersion  $\sigma_{\text{los}}$ . To model the 3D mass distribution consistent with the observed velocity dispersion, we adopt the spherical solution of the Jeans equations. Although the true mass distribution is non-spherical, the assumption of spherical symmetry is sufficient given the 10–25 per cent uncertainty in our kinematic data (Section 3.4; Sonnenfeld et al. 2012). We can express the spherical Jeans equation as

$$\frac{d(l(r)\sigma_r^2)}{dr} + \frac{2\beta_{\text{ani}}l(r)\sigma_r^2}{r} = -l(r)\frac{d\Phi}{dr}. \quad (15)$$

Here,  $l(r)$  is the 3D luminosity density of the stars,  $\sigma_r$  is the intrinsic radial velocity dispersion, and  $\beta_{\text{ani}}(r)$  is the anisotropy parameter relating  $\sigma_r$  with the tangential velocity dispersion  $\sigma_t$  as

$$\beta(r) \equiv 1 - \frac{\sigma_t^2}{\sigma_r^2}. \quad (16)$$

By solving equation (15), we can obtain the luminosity-weighted, line-of-sight velocity dispersion as

$$\sigma_{\text{los}}^2(R) = \frac{2G}{I(R)} \int_R^\infty \mathcal{K}_\beta \left( \frac{r}{R} \right) \frac{l(r)M(r)}{r} dr, \quad (17)$$

where  $M(r)$  is the enclosed mass within radius  $r$  (equation (A15)–(A16) of Mamon & Łokas 2005). Here, the function  $\mathcal{K}_\beta(\varrho)$  depends on the parametrization of  $\beta(r)$ . We adopt the Osipkov–Merritt parametrization of the anisotropy parameter given by

$$\beta_{\text{ani}}(r) = \frac{r^2}{r^2 + r_{\text{ani}}^2}, \quad (18)$$

where  $r_{\text{ani}}$  is the anisotropy scale radius (Osipkov 1979; Merritt 1985a, b). For this parametrization, the function  $\mathcal{K}_\beta$  takes the form

$$\mathcal{K}_\beta(u) = \frac{u_{\text{ani}}^2 + 1/2}{(u_{\text{ani}} + 1)^{3/2}} \left( \frac{u^2 + u_{\text{ani}}^2}{u} \right) \tan^{-1} \left( \sqrt{\frac{u^2 - 1}{u_{\text{ani}}^2 + 1}} \right) - \frac{1/2}{u_{\text{ani}}^2 + 1} \sqrt{1 - \frac{1}{u^2}}, \quad (19)$$

where  $u_{\text{ani}} = r_{\text{ani}}/R$  (Mamon & Łokas 2005).

The enclosed mass  $M(r)$  is computed from the 3D mass profile. For the convergence and surface brightness profiles that cannot be straightforwardly deprojected into 3D, we decompose them into concentric Gaussian components (Bendinelli 1991; Emsellem et al. 1994; Cappellari 2002; Shajib 2019). We then deproject the Gaussian components into 3D Gaussians to compute the enclosed mass  $M(r)$  and 3D light density profile  $l(r)$ .

### 2.4 Cosmological distances

In this section, we effectively follow Birrer et al. (2016, 2019) to jointly infer  $D_{\Delta t}$  and  $D_d$ . From the modelled convergence profile  $\kappa'_{\text{model}}$  of the deflector, we derive the time-delay distance  $D'_{\Delta t}$

particular to the deflector’s line of sight. We need to correct  $D'_{\Delta t}$  for the external convergence  $\kappa_{\text{ext}}$  to obtain the true time-delay distance  $D_{\Delta t}$ . From equations (1) and (12), we can express the true time-delay distance  $D_{\Delta t}$  as

$$D_{\Delta t} = \frac{D'_{\Delta t}}{1 - \kappa_{\text{ext}}}. \quad (20)$$

We can express  $\sigma_{\text{los}}$  in terms of parameters characterizing the 2D mass and light distributions and relevant angular diameter distances as

$$\sigma_{\text{los}}^2 = \frac{D_s}{D_{\text{ds}}} c^2 J(\xi_{\text{lens}}, \xi_{\text{light}}, \beta_{\text{ani}}), \quad (21)$$

where  $\xi_{\text{lens}}$  is the set of mass parameters,  $\xi_{\text{light}}$  is the set of light distribution parameters,  $c$  is the speed of light, and the function  $J$  captures all the dependencies from the mass profile, the light profile, and the orbital anisotropy (Birrer et al. 2016). The parameters in the argument of the function  $J$  are expressed in angular units, thus they do not depend on the cosmology. Then from equation (1), we have

$$\frac{D_d D_s}{D_{\text{ds}}} = \frac{c \Delta t_{\text{XY}}}{(1 + z_d) \Delta \phi_{\text{XY}}^{\text{eff}}(\xi_{\text{lens}})}. \quad (22)$$

Combining this equation with equation (21), we can write

$$D_d = \frac{c^3 \Delta t_{\text{XY}} J(\xi_{\text{lens}}, \xi_{\text{light}}, \beta_{\text{ani}})}{(1 + z_d) \sigma_{\text{los}}^2 \Delta \phi_{\text{XY}}^{\text{eff}}(\xi_{\text{lens}})} \quad (23)$$

(Birrer et al. 2016). As a result, we can estimate the angular diameter distance  $D_d$  to the deflector by combining the kinematics with the lensing observables. Therefore, we can infer two cosmological distances,  $D_{\Delta t}$  and  $D_d$ , at specific redshifts relevant to the lens system. Thus, we can constrain the Hubble constant and other cosmological parameters from the distance–redshift relation for a given cosmology. In the next section, we describe the combined Bayesian framework to infer the Hubble constant from the observables.

### 2.5 Bayesian inference framework

We tabulate the notations used in this section in Table 1 as a quick reference for the readers. At the top level, the two cosmological distances  $D_{\Delta t}$  and  $D_d$  contain all the cosmographic information. We express the set of cosmological distances using the notation  $D$ , which is a function  $D(\omega; C)$  of the set of cosmological parameters  $\omega$  for a given cosmology  $C$ . We denote the set containing all the observables as  $O \equiv \{O_{\text{img}}, O_{\Delta t}, O_{\text{kin}}, O_{\text{env}}\}$ , where  $O_{\text{img}}$  contains the imaging data of the lens system,  $O_{\Delta t}$  contains the observed time delays,  $O_{\text{kin}}$  contains the spectra of the deflector to estimate the kinematics, and  $O_{\text{env}}$  contains photometric and spectroscopic survey data of the lens environment to estimate the external convergence. Then from Bayes’ theorem, we can write

$$p(\omega | O, C) \propto p(O | \omega, C) p(\omega | C) = p(O | D(\omega; C)) p(\omega | C), \quad (24)$$

where the probability density  $p(\omega|O, C)$  is called the posterior of  $\omega$ , the probability density  $p(O|\omega, C)$  is called the likelihood of  $O$  given  $\{\omega, C\}$ , and the probability density  $p(\omega|C)$  is called the prior for  $\omega$ . In the last line of the above equation, we have changed  $\{\omega, C\}$  into  $D(\omega; C)$  in the likelihood term, as it allows us to break down the computation of the likelihood into two steps. First, we compute the likelihood  $p(O|D)$  of the observed data for given cosmological distances marginalizing over various model choices and their respective parameters. Then, we can fold in the prior of

**Table 1.** List of notations used in Section 2.5.

Notation	Definition
$D$	$\equiv \{D_{D_{\Delta t}, D_d}\}$ , set of cosmological distances
$C$	a given cosmological model
$\omega$	set of cosmological parameters in $C$
$O_{\text{img}}$	imaging data of the lens systems
$O_{\Delta t}$	observed time delay
$O_{\text{kin}}$	spectra of the deflector to measure the stellar kinematics
$O_{\text{env}}$	data to estimate the external convergence
$O$	$\equiv \{O_{\text{img}}, O_{\Delta t}, O_{\text{kin}}, O_{\text{env}}\}$
$M$	mass and light model for the central deflector
$\xi_{\text{lens}}$	mass model parameters in $M$
$\xi_{\text{light}}$	light model parameters in $M$
$S$	model specifying the source light and the perturber mass
$\xi_{\text{source}}$	source light parameters in $S$
$\xi_{\text{pert}}$	perturber mass parameters in $S$
$\kappa_{\text{ext}}$	external convergence
$\beta_{\text{ani}}$	anisotropy parameter from equation (18)
$\xi_{\beta}$	model parameters characterizing $\beta_{\text{ani}}$
$\sigma_{\text{los}}$	line-of-sight velocity dispersion
$\Xi$	$\equiv \{\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{source}}, \xi_{\text{pert}}\}$
$\Delta\phi_{XY}^{\text{eff}}$	effective Fermat potential difference from equation (4)
$\mathcal{Z}$	Bayesian model evidence
$\nu$	non-linear parameters in $\Xi$
$\lambda$	vector containing linear parameters in $\Xi$
$\hat{\lambda}$	maximum likelihood estimator of $P(O_{\text{img}} \nu, \lambda, M, S)$
$\mathbf{K}_{\lambda\lambda}$	covariance matrix of $\lambda$
$w$	uniform prior width of $\lambda \sim \mathcal{U}(-w/2, w/2)$
$\xi_{\text{micro}}$	parameters relevant to the microlensing time-delay effect

the cosmological parameters  $p(\omega|C)$  to obtain the posterior  $p(\omega|O, C)$ . As the different pieces of the data in  $O$  are independent, we can breakup the likelihood into likelihoods of each observable type as

$$p(O|D) = p(O_{\text{img}}|D) p(O_{\Delta t}|D) p(O_{\text{kin}}|D) p(O_{\text{env}}|D). \quad (25)$$

When computing these likelihood functions, we adopt a combination of model choices. We denote the model choice containing the mass model parameters  $\xi_{\text{lens}}$  and deflector light model parameters  $\xi_{\text{light}}$  as  $M$ . In addition, we have to make specific choices for the parametrization  $\xi_{\text{source}}$  of the source light distribution and the parametrization  $\xi_{\text{pert}}$  of the mass profiles of the line-of-sight perturbers. We denote the model choice encompassing  $\xi_{\text{source}}$  and  $\xi_{\text{pert}}$  as  $S$ . We also marginalize over the external convergence  $\kappa_{\text{ext}}$  and the parameters  $\xi_{\beta}$  characterizing  $\beta_{\text{ani}}$ . Adding it all together, we can marginalize all the specific model parameters to express the total likelihood given the distances as

$$p(O|D, M) = \int p(O_{\text{img}}|\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{source}}, \xi_{\text{pert}}, M, S) \times p(O_{\Delta t}|\Delta t(\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{pert}}, \kappa_{\text{ext}}; M, S)) \times p(\xi_{\text{source}}, \xi_{\text{pert}}|S) p(S) \times p(O_{\text{kin}}|\sigma_{\text{los}}(\xi_{\text{lens}}, \xi_{\text{light}}, \kappa_{\text{ext}}, \xi_{\beta}; M)) \times p(\xi_{\beta}|\xi_{\text{lens}}, \xi_{\text{light}}, M) \times p(O_{\text{env}}|\kappa_{\text{ext}}) p(\kappa_{\text{ext}}) \times p(\xi_{\text{lens}}, \xi_{\text{light}}|M) \times d\xi_{\text{source}} d\xi_{\text{pert}} dS d\xi_{\beta} d\kappa_{\text{ext}} d\xi_{\text{lens}} d\xi_{\text{light}}. \quad (26)$$

Here, we omitted some model parameters and model specifications in the conditional statements of the likelihoods where the corre-

sponding likelihood does not depend on them. Breaking up the likelihood as above allows us to partially separate the computation of the likelihoods for different observable types before marginalizing over the model parameters. We first describe the imaging likelihood and marginalization over relevant models and model parameters in Section 2.5.1, then we explain the derivation of the joint posterior combining time delay and kinematics likelihoods with the lens model posterior in Section 2.5.2.

### 2.5.1 Lens model posterior and evidence from imaging likelihood

We can first obtain the posterior of the lens model parameters  $\Xi \equiv \{\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{source}}, \xi_{\text{pert}}\}$  as

$$p(\Xi|O_{\text{img}}, M, S) = \frac{p(O_{\text{img}}|\Xi, M, S) p(\Xi|M, S)}{p(O_{\text{img}}|M, S)}. \quad (27)$$

Here, the term in the denominator  $\mathcal{Z} \equiv p(O_{\text{img}}|M, S)$  is the evidence for the imaging data  $O_{\text{img}}$  given the model  $\{M, S\}$ . We first change the variables  $\{\xi_{\text{lens}}, \xi_{\text{pert}}\} \rightarrow \{\xi_{\text{lens}}, \Delta\phi_{XY}^{\text{eff}}\}$  in equation (27) to be able to marginalize over parameters related to the line-of-sight galaxies while retaining their effect on the Fermat potential difference  $\Delta\phi_{XY}$ . As the Jacobian determinant  $|d\{\xi_{\text{lens}}, \Delta\phi_{XY}^{\text{eff}}\}/d\{\xi_{\text{lens}}, \xi_{\text{pert}}\}|$  cancels out from both sides, we have

$$p(\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{source}}, \Delta\phi_{XY}^{\text{eff}}|O_{\text{img}}, M, S) = p(O_{\text{img}}|\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{source}}, \Delta\phi_{XY}^{\text{eff}}, M, S) \times \frac{p(\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{source}}, \Delta\phi_{XY}^{\text{eff}}|M, S)}{p(O_{\text{img}}|M, S)}. \quad (28)$$

We can marginalize this posterior over  $\xi_{\text{source}}$  and  $S$  as

$$p(\xi_{\text{lens}}, \xi_{\text{light}}, \Delta\phi_{XY}^{\text{eff}}|O_{\text{img}}, M) = \int p(\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{source}}, \Delta\phi_{XY}^{\text{eff}}|O_{\text{img}}, S, M) \times p(S) d\xi_{\text{source}} dS. \quad (29)$$

Since the term inside the integral contains the evidence term  $\mathcal{Z}$ , the integral over the model space  $S$  automatically weights the models  $\{S\}$  according to their evidence ratios. As we can only discretely sample models  $\{S_n\}$  from the model space  $S$ , the integral in equation (27) becomes a discrete sum as

$$p(\xi_{\text{lens}}, \xi_{\text{light}}, \Delta\phi_{XY}^{\text{eff}}|O_{\text{img}}, M) = \sum_n \Delta S_n \int p(\xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{source}}, \Delta\phi_{XY}^{\text{eff}}|O_{\text{img}}, S_n, M) \times p(S_n) d\xi_{\text{source}}. \quad (30)$$

Here, the term  $\Delta S_n$  can be interpreted as the model space volume represented by the model  $S_n$ , thus it can account for sparse sampling from the model space.

In our model, we have both linear and non-linear parameters. The linear parameters are the amplitudes of the surface brightness profiles that we treat with a basis set in our model (Birrer et al. 2015). We denote the linear parameters using the vector  $\lambda$  and non-linear parameters using the set  $\nu$ . Hence, the lens model parameters can alternatively be expressed as  $\Xi \equiv \{\nu, \lambda\}$ . We can write the evidence integral as

$$\mathcal{Z} = \int p(O_{\text{img}}|\nu, \lambda, M, S) p(\lambda|M, S) p(\nu|M, S) d\lambda d\nu. \quad (31)$$

We can first marginalize over the linear parameters to get the likelihood  $P(O_{\text{img}}|\nu, M, S)$  in terms of only the non-linear

parameters as

$$p(O_{\text{img}} | \nu, M, S) = \int p(O_{\text{img}} | \nu, \lambda, M, S) p(\lambda | M, S) d\lambda. \quad (32)$$

If  $\hat{\lambda}$  is the maximum-likelihood estimator of  $P(O_{\text{img}} | \nu, \lambda, M, S)$  for given  $\{\nu, M, S\}$ , then we can approximate the likelihood using up to the second order terms in the Taylor expansion in the vicinity of  $\hat{\lambda}$  as

$$p(O_{\text{img}} | \nu, \lambda, M, S) \approx p(O_{\text{img}} | \nu, \hat{\lambda}, M, S) \times \exp \left[ -\frac{1}{2} \Delta\lambda^T \mathbf{K}_{\lambda\lambda}^{-1} \Delta\lambda \right], \quad (33)$$

where  $\Delta\lambda = \lambda - \hat{\lambda}$ , and  $\mathbf{K}_{\lambda\lambda}$  is the covariance matrix of  $\lambda$  (equation [12] of Birrer et al. 2015). We can directly obtain  $\hat{\lambda}$  given the set of non-linear parameters  $\nu$  by solving a set of linear equations. If we take a uniform prior  $\mathcal{U}(-\mathbf{w}/2, \mathbf{w}/2)$  for  $\lambda$ , then from equation (32) we have

$$p(O_{\text{img}} | \nu, M, S) \approx p(O_{\text{img}} | \nu, \hat{\lambda}, M, S) \frac{[(2\pi)^n \det(\mathbf{K}_{\lambda\lambda})]^{1/2}}{\prod_i^n w_i}, \quad (34)$$

where  $n = \dim(\lambda)$  is the number of linear parameters. Then, we can express the evidence as

$$\mathcal{Z} \approx \frac{[(2\pi)^n \det(\mathbf{K}_{\lambda\lambda})]^{1/2}}{\prod_i^n w_i} \int p(O_{\text{img}} | \nu, \hat{\lambda}, M, S) p(\nu | M, S) d\nu. \quad (35)$$

### 2.5.2 Joint posterior combining time delay and kinematics likelihoods

Next, we can fold in the time-delay likelihood to update the posterior and marginalize over the Fermat potential  $\Delta\phi_{XY}^{\text{eff}}$  as

$$\begin{aligned} & p(D, \xi_{\text{lens}}, \xi_{\text{light}}, \kappa_{\text{ext}} | O_{\text{img}}, O_{\Delta t}, O_{\text{env}}, M) \\ & \propto \int p(O_{\Delta t} | \Delta t(D, \xi_{\text{lens}}, \xi_{\text{light}}, \Delta\phi_{XY}^{\text{eff}}, \kappa_{\text{ext}}; O_{\text{img}}, O_{\text{env}}, M)) \\ & \quad \times p(\xi_{\text{lens}}, \xi_{\text{light}}, \Delta\phi_{XY}^{\text{eff}} | O_{\text{img}}, M) \\ & \quad \times p(\kappa_{\text{ext}} | O_{\text{env}}) p(D) d\Delta\phi_{XY}^{\text{eff}}. \end{aligned} \quad (36)$$

Tie & Kochanek (2018) introduce a possible microlensing time-delay effect due to the asymmetric magnification of a quasar accretion disc – assuming the lamppost model (Shakura & Sunyaev 1973) – due to microlensing by the foreground stars in the deflector galaxy. Note, the time-delay measurement from the quasar light curves accounts for the long-term variation in the microlensing magnification pattern. Tie & Kochanek (2018)'s microlensing time-delay effect is due to the non-uniform weighting of the quasar accretion disc brightness by the microlensing magnification pattern, thus this effect depends on the gradient of the magnification across the accretion disc. The long-term change in the magnification pattern is not necessarily correlated with the gradient of the magnification across the accretion disc. For the case that marginalizes over this microlensing time-delay effect, the above equation becomes

$$\begin{aligned} & p(D, \xi_{\text{lens}}, \xi_{\text{light}}, \kappa_{\text{ext}} | O_{\text{img}}, O_{\Delta t}, O_{\text{env}}, M) \\ & \propto \int p(O_{\Delta t} | \Delta t(D, \xi_{\text{lens}}, \xi_{\text{light}}, \xi_{\text{micro}}, \kappa_{\text{ext}}, \\ & \quad \times \Delta\phi_{XY}^{\text{eff}}; O_{\text{img}}, O_{\text{env}}, M)) \\ & \quad \times p(\xi_{\text{lens}}, \xi_{\text{light}}, \Delta\phi_{XY}^{\text{eff}} | O_{\text{img}}, M) \\ & \quad \times p(\kappa_{\text{ext}} | O_{\text{env}}) p(\xi_{\text{micro}}) p(D) d\Delta\phi_{XY}^{\text{eff}} d\xi_{\text{micro}}. \end{aligned} \quad (37)$$

where  $\xi_{\text{micro}}$  is the set of parameters relevant to the microlensing time-delay effect, e.g. parameters related to the properties of the black hole and the accretion disc (Chen et al. 2018, Section 6.2.3).

Then, we can update the posterior once again by folding in the kinematic likelihood as

$$\begin{aligned} & p(D, \xi_{\text{lens}}, \xi_{\text{light}}, \kappa_{\text{ext}}, \xi_{\beta} | O_{\text{img}}, O_{\Delta t}, O_{\text{kin}}, O_{\text{env}}, M) \\ & = p(O_{\text{kin}} | \sigma_{\text{los}}(D, \xi_{\text{lens}}, \xi_{\text{light}}, \kappa_{\text{ext}}, \xi_{\beta}; M)) \\ & \quad \times p(\xi_{\beta} | \xi_{\text{lens}}, \xi_{\text{light}}, M) \\ & \quad \times p(D, \xi_{\text{lens}}, \xi_{\text{light}}, \kappa_{\text{ext}} | O_{\text{img}}, O_{\Delta t}, M). \end{aligned} \quad (38)$$

Now, we can marginalize over the model parameters to obtain the posterior of the cosmological distances  $D$  as

$$\begin{aligned} & p(D | O, M) = p(D | O_{\text{img}}, O_{\Delta t}, O_{\text{kin}}, O_{\text{env}}, M) \\ & = \int p(D, \xi_{\text{lens}}, \xi_{\text{light}}, \kappa_{\text{ext}}, \xi_{\beta} | O_{\text{img}}, O_{\Delta t}, O_{\text{kin}}, O_{\text{env}}, M) \\ & \quad \times d\xi_{\beta} d\kappa_{\text{ext}} d\xi_{\text{lens}} d\xi_{\text{light}}. \end{aligned} \quad (39)$$

Finally, we can marginalize over the deflector mass model choices as

$$p(D | O) = \sum_M p(D | O, M) p(M | O). \quad (40)$$

A particular choice of mass model  $M$  breaks the MSD (Schneider & Sluse 2014). However, we cannot ascertain that our adopted mass model choice represents the true mass distribution. As a result, we cannot weigh different mass models according to their evidence ratios as a higher evidence value may just be a fluke from breaking the MSD near a better fit of the data. Therefore, we take  $p(M|O) = 1$  to equally weight different deflector mass model choices.

As the likelihood  $p(O|D)$  follows the proportionality relation

$$p(O | D) \propto \frac{p(D | O)}{p(D)}, \quad (41)$$

we can then use the distance posterior  $p(D|O)$  to obtain the posterior of the cosmological parameters  $p(\omega|O, C)$  from equation (24).

## 3 THE LENS SYSTEM AND DATA SETS

In this paper, we perform cosmographic analyses of the lens systems DES J0408–5354. This lens was discovered and confirmed by Lin et al. (2017) from a large sample of potential galaxy–galaxy lenses in the DES footprint (Diehl et al. 2017). Agnello et al. (2017) acquired follow-up data and modelled the system presenting evidence for a faint perturber G2 near one of the quasar images, which was later confirmed by the deeper and higher resolution imaging from the *Hubble Space Telescope* (*HST*; Shajib et al. 2019).

The necessary data sets and ancillary measurements for cosmographic analysis are

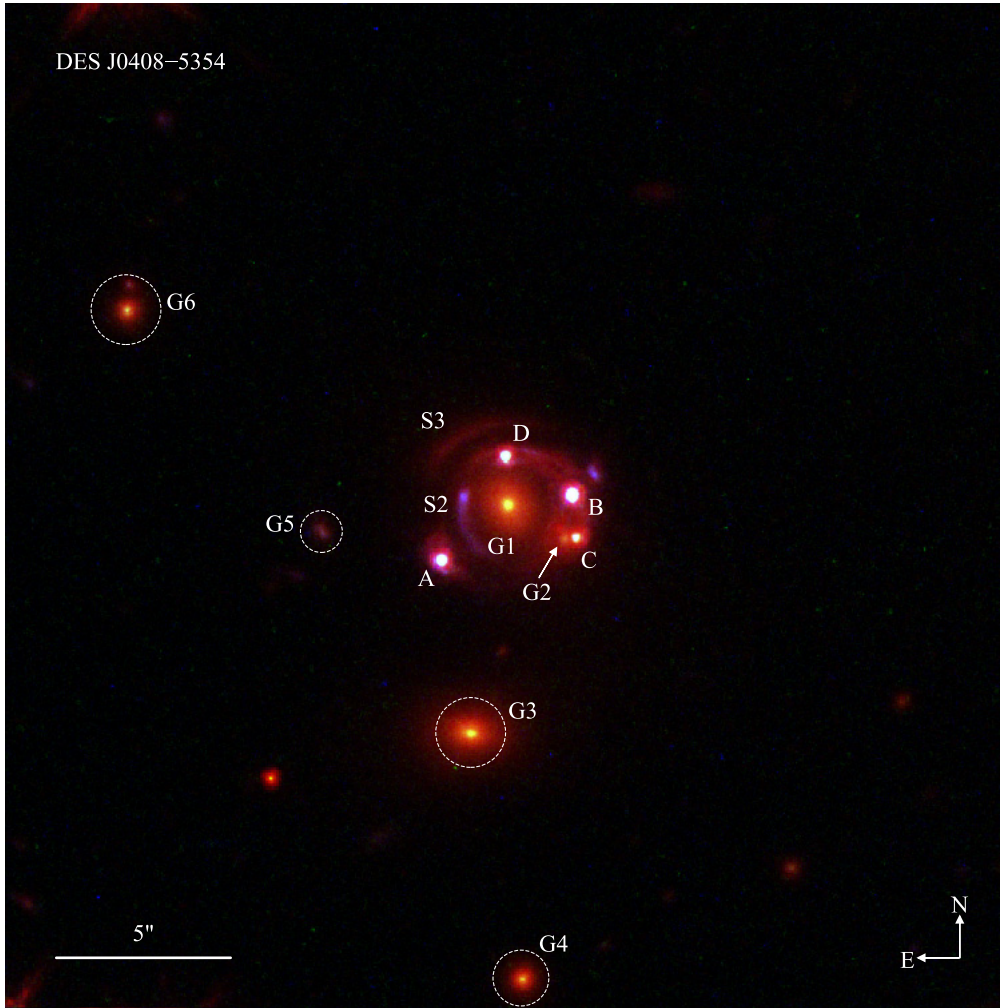
- (i) high-resolution imaging of the lens system,
- (ii) spectroscopy of the lens components to measure redshifts,
- (iii) measured time-delays between the images,
- (iv) LOS velocity dispersion of the central deflector galaxy, and
- (v) estimate of the external convergence.

Each type of data set or ancillary measurement is described in the following subsections.

### 3.1 *HST* imaging of the lens system

*HST* Wide-Field Camera 3 (WFC3) imaging was obtained under the program GO-15320 (PI: Treu; Shajib et al. 2019). The images were taken in three filters: F160W in infrared (IR), F814W





**Figure 1.** RGB colour composite of the lens systems DES J0408–5354. The three *HST* filters used to create the RGB image are F160W (red), F814W (green), and F475X (blue). The relative amplitudes between the three filters are adjusted in this figure for better visualization by achieving a higher contrast. We label different components of the lens system. G1 is the main deflector galaxy and G2 is its satellite galaxy. In addition to the lensed arcs from the extended quasar host galaxy, this lens system has extra source components S2 and S3. The source component S2 is doubly imaged and forms an extended arc inside the Einstein radius. S3 forms another fainter extended arc on the North-East outside the Einstein radius without a noticeable counter-image. Four nearby perturbers G3–G6 along the line of sight are marked with the dashed, white circles.

and F475X in ultraviolet–visual (*UVIS*). For each filter, four exposures – two short and two long – were taken to cover the large dynamic range in brightness encompassing the bright quasar images and the fainter extended host galaxy. For the IR band, we chose a 4-point dither pattern and STEP100 readout sequence for the MULTIACCUM mode. For the *UVIS* bands, we adopted a 2-point dither pattern. The total exposure times for the three filters are 2196.9 s in F160W, 1428 s in F814W, and 1348 s in F475X.

The data in each band were reduced with the standard ASTRODRIZZLE package (Avila et al. 2015). The final pixel scale after drizzling is 0.08 arcsec in the IR band, and 0.04 arcsec in the *UVIS* band. We estimate the background level in the reduced image from each band using SEXTRACTOR and subtract it from the reduced image (Bertin & Arnouts 1996).

Fig. 1 shows the colour-composite image for the lens system and its surrounding. The central deflector galaxy G1 has a visible satellite galaxy G2. The four prominent nearby galaxies along the line of sight are marked with G3, G4, G5, and G6. Note that the

naming convention of these galaxies is different in Lin et al. (2017) and Agnello et al. (2017).

The lens has multiple lensed arcs from additional source components, S2 and S3. The lensed arc S2 lies inside the Einstein radius and it has a noticeable counterimage on the North–West of image B. Another faint lensed arc S3 lies on the East of image D. We could not identify the counterimage of S3 from visual inspection.

### 3.2 Spectroscopic observations of the lens components

The central deflector G1 sits at the redshift  $z_d = 0.597$  and the quasar sits at redshift  $z_{QSO} = 2.375$  (Lin et al. 2017). Buckley-Geer et al. (2020) measure redshifts for the nearby line-of-sight galaxies G3–G6 from spectroscopic observations using the Magellan and the Gemini telescopes obtaining  $z_{G3} = 0.769$ ,  $z_{G4} = 0.771$ ,  $z_{G5} = 1.032$ , and  $z_{G6} = 0.594$ . The redshifts are precise up to the specified decimal point.

We measure the redshift of S2  $z_{S2} = 2.228$  from the integral-field spectroscopy observations of DES J0408–5354 with the Multi-Unit



**Table 2.** Redshift and stellar velocity dispersion for the line-of-sight galaxies G3–G6 and S2. The relative offsets of the observed centroids for G3–G6 are computed from the coordinate RA 04:08:21.71 and Dec  $-53:53:59.34$ . The tabulated uncertainties for the velocity dispersions are statistical uncertainties. However, we adopt a  $20 \text{ km s}^{-1}$  uncertainty for each measurement to account for the typical systematic uncertainty for kinematics obtained from MUSE spectra (Guérou et al. 2017). The redshifts are precise up to the specified decimal point.

Galaxy	$\Delta\text{RA}$ (arcsec)	$\Delta\text{Dec}$ (arcsec)	Redshift	Stellar velocity dispersion ( $\text{km s}^{-1}$ )
G3	1.08	−6.52	0.769	$226 \pm 7$
G4	−0.40	−13.58	0.771	$153 \pm 10$
G5	5.34	−0.78	1.032	$56 \pm 2$
G6	10.90	5.53	0.594	$63 \pm 7$
S2	–	–	2.228	$46 \pm 9$

Spectroscopic Explorer (MUSE, on the ESO VLT UT4). The MUSE observations of the lens and its immediate neighbourhood, within approximately 45 arcsec, were carried out in Period 102 during two nights on 2019 January 11 and 13 [run 0102.A-0600(E), PI Agnello]. The observations were executed in wide-field mode with adaptive-optics (AO) corrections, so that the multiple images and galaxies in this lens could be properly deblended. The AO wide field mode of MUSE results in a wavelength coverage from 4700 to 5803 Å; and 5966–9350 Å; at a spectral resolution of  $R \sim 1700$ –3400. Each observation block contains four exposures, with the main target placed in four different quadrants of the instrument’s field of view. An approximately  $15 \text{ arcsec} \times 15 \text{ arcsec}$  region centred on the lens was exposed for 4 h, with a dither-and-rotation pattern that minimized artefacts due to the multiple instrument slicers and channels. We reduced the data cubes using the standard ESOREX pipeline recipes and flux calibrated them using observations of standard stars obtained on the two nights. Offsets between 20 individual exposures were determined from cross-correlations of white light images created from individual data cubes. We cleaned strong sky-line residuals from the final combined data cube using ZAP (Soto et al. 2016). The setup results in a final data cube with a full field of view of  $92 \text{ arcsec} \times 95 \text{ arcsec}$ . For this work, we analysed a  $8 \text{ arcsec} \times 8 \text{ arcsec}$  ‘mini-cube’ centred around the lens. We use three stars in the field as reference point source function (PSF) cubes. We decompose the ‘mini-cube’ as a superposition of four Moffat profiles for the quasar images, and a convolved de Vaucouleurs profile for the deflector light distribution. By means of this procedure, all component spectra could be reliably separated and the quasar shot noise on the deflector spectra was minimized. We use Mg II emission lines to measure S2’s redshift and velocity dispersion. As S2 and the quasar are at different redshift, the quasar’s Mg II contamination does not overlap with S2’s Mg II lines. Also, given the large systematic uncertainty on the velocity dispersion

described in the next paragraph, residual AGN contamination is not a dominant source of bias.

We also measure the line-of-sight velocity dispersions of G3–G6 and S2 from the MUSE spectra (Table 2). We adopt an uncertainty of  $20 \text{ km s}^{-1}$  on the measured velocity dispersion to account for the typical systematic uncertainty for kinematics extracted from MUSE spectra (Guérou et al. 2017). The estimated PSF from the stars in the MUSE observation can be different than the PSF of the quasar due to different SED within a filter. However, the impact in the estimated velocity dispersion from this potential PSF mismatch is subdominant to this conservative estimate of the systematic uncertainty.

### 3.3 Time delays

Courbin et al. (2018) present the measured time delays between the images of DES J0408–5354. This system was monitored to obtain light-curves of the lensed images using the MPIA 2.2 m telescope at La Silla observatory between 2016 October 1 and 2017 April 8. The system was observed almost daily except for 14 consecutive nights between 2016 December 10 and 2016 December 24, and for one week in 2017 January due to bad weather and technical problems. Additional monitoring was carried out using the 1.2 m the *Leonhard Euler* 1.2 m Swiss Telescope (*Euler*) between 2016 July and 2017 April. The mean observation cadence with *Euler* is 5 d. From these light-curves of the lensed images, the measured time delays are  $\Delta t_{\text{AB}} = -112.1 \pm 2.1 \text{ d}$ ,  $\Delta t_{\text{AD}} = -155.5 \pm 12.8 \text{ d}$ , and  $\Delta t_{\text{BD}} = -42.4 \pm 17.6 \text{ d}$  (see Fig. 1 for the naming of the images). The time delays relative to image C could not be measured due the close proximity of a satellite galaxy as it is difficult to deblend the quasar flux from the satellite’s in the ground-based monitoring data.

### 3.4 Velocity dispersion of the central deflector

Buckley-Geer et al. (2020) measure the velocity dispersion of G1. The velocity dispersion is measured with four different observing setups: two mask setups with the Magellan telescope, one with the Gemini telescope, and one with the MUSE spectra. The specifics and the measured values from these four setups are tabulated in Table 3. We estimate the systematic uncertainty  $\sigma_{\text{los}}^{\text{sys}}$  in the measured velocity dispersion to add the reported statistical uncertainty  $\sigma_{\text{los}}^{\text{stat}}$ . We infer a systematic uncertainty of  $17 \text{ km s}^{-1}$  from the variance in the estimated velocity dispersions when different settings – e.g. the stellar population library, the stellar templates, the wavelength region – are varied in the kinematic fitting. We form a covariance matrix for the velocity dispersion measurements with  $(\sigma_{\text{los}}^{\text{sys}})^2 + (\sigma_{\text{los}}^{\text{stat}})^2$  for the diagonal terms and  $(\sigma_{\text{los}}^{\text{sys}})^2$  for off-diagonal terms, as the source of the systematic in the kinematic fitting is common between all the measurements.

**Table 3.** Measurements of velocity dispersion from three different setups from Buckley-Geer et al. (2020). The quoted uncertainties are only statistical, see Section 3.4 for the estimated systematic uncertainty.

Instrument and setup	Aperture dimension (arcsec $\times$ arcsec)	Aperture rotation (deg E of N)	Seeing (arcsec)	Moffat PSF exponent	Velocity dispersion ( $\text{km s}^{-1}$ )
Magellan mask A	$1 \times 1$	99	0.68	−2.97	$230 \pm 37$
Magellan mask B	$1 \times 1$	99	0.76	−3.20	$236 \pm 42$
Gemini mask A2	$0.75 \times 1$	0	0.52	−3.06	$220 \pm 21$
MUSE	$1 \times 1$	0	0.61	−1.55	$227 \pm 9$

### 3.5 Estimate of the external convergence

Buckley-Geer et al. (2020) present the distribution of the external convergence  $\kappa_{\text{ext}}$  for DES J0408–5354. This analysis is based on the weighted galaxy number counts approach of Greene et al. (2013), which was further developed by Rusu et al. (2017, 2020) and Birrer et al. (2019). In brief, weighted number counts are computed in 45 arcsec- and 120 arcsec-radii apertures centred on the lensing system, up to a depth of  $I = 22.5$  mag, using simple physical weights robust to measurement errors, such as the inverse of the distance to the lens and photometric/spectroscopic redshifts. Analogous number counts are computed in a large number of same-size apertures and depth in a cosmological survey, in this case DES, so as to measure the over/underdensity of the DES J0408–5354 line of sight relative to the median line of sight through the Universe, in terms of weighted number count ratios. In this case, the line of sight was found to be underdense, and a combination of weighted number count ratios was used as constraint to select statistically similar lines of sight from the Millennium Simulation (Springel et al. 2005). Using the external convergence maps from Hilbert et al. (2009) corresponding to each Millennium Simulation line of sight, we construct a probability distribution function of  $\kappa_{\text{ext}}$ . This probability distribution function of  $\kappa_{\text{ext}}$  is provided in Section 6.2.2 (specifically in Fig. 7).

## 4 LENS MODEL INGREDIENTS

In this section, we describe the mass and light profiles used to construct the lens model in our analysis.

### 4.1 Central deflector’s mass profiles

To model the main deflector’s mass distribution, we adopt two sets of profiles: (i) power-law, and (ii) composite mass profile.

#### 4.1.1 Power-law mass profile

We adopt the power-law elliptical mass distribution (Barkana 1998). This profile is described by

$$\kappa_{\text{PL}}(\theta_1, \theta_2) = \frac{3 - \gamma}{2} \left[ \frac{\theta_E}{\sqrt{q_m \theta_1^2 + \theta_2^2 / q_m}} \right]^{\gamma - 1}, \quad (42)$$

where  $\gamma$  is the power-law slope,  $\theta_E$  is the Einstein radius, and  $q_m$  is the axial ratio. The coordinates  $(\theta_1, \theta_2)$  are in the frame that is aligned with the major and minor axes. This frame is rotated by a position angle  $\varphi_m$  from the frame of on-sky coordinates.

#### 4.1.2 Composite mass profile

In the composite mass profile, we adopt separate mass profiles for the baryonic and the dark components of the mass distribution.

For the dark component, we choose a Navarro–Frenk–White (NFW) profile with ellipticity defined in the potential. The spherical NFW profile in 3D is given by

$$\rho_{\text{NFW}}(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}, \quad (43)$$

where  $r_s$  is the scale radius, and  $\rho_s$  is the normalization (Navarro, Frenk & White 1997).

For the baryonic mass profile, we adopt the Chameleon convergence profile. The Chameleon profile approximates the Sérsic profile within a few per cent in the range  $0.5\text{--}3R_{\text{eff}}$ , where  $R_{\text{eff}}$

is the effective or half-light radius of the Sérsic profile. The Chameleon profile is the difference between two non-singular isothermal ellipsoids given by

$$\kappa_{\text{Chm}}(\theta_1, \theta_2) = \frac{\kappa_0}{1 + q_m} \left[ \frac{1}{\sqrt{\theta_1^2 + \theta_2^2/q_m^2 + 4w_c^2/(1 + q_m^2)}} - \frac{1}{\sqrt{\theta_1^2 + \theta_2^2/q_m^2 + 4w_l^2/(1 + q_m^2)}} \right] \quad (44)$$

(Dutton et al. 2011; Suyu et al. 2014). This profile is convenient to compute lensing properties using closed-form expressions.

With each of these models, we include an external shear profile parametrized with the shear magnitude  $\gamma_{\text{ext}}$  and shear angle  $\varphi_{\text{ext}}$ .

### 4.2 Central deflector’s light profile

We use the Sérsic profile and the Chameleon profile to model different light components of the lens system.

#### 4.2.1 Sérsic profile

The Sérsic profile is given by

$$I_{\text{Sérsic}}(\theta_1, \theta_2) = I_{\text{eff}} \exp \left[ -b_n \left\{ \left( \frac{\sqrt{\theta_1^2 + \theta_2^2/q_m^2}}{\theta_{\text{eff}}} \right)^{1/n_s} - 1 \right\} \right], \quad (45)$$

where  $\theta_{\text{eff}}$  is the effective radius,  $I_{\text{eff}}$  is the amplitude at  $\theta_{\text{eff}}$ , and  $n_s$  is the Sérsic index (Sérsic 1968). The factor  $b_n$  normalizes the profile such that half of the total luminosity is contained within  $\theta_{\text{eff}}$ .

#### 4.2.2 Chameleon profile

We use the same Chameleon profile from equation (44) to fit the central deflector’s light profile by replacing the convergence amplitude  $\kappa_0$  with flux amplitude  $I_0$ .

### 4.3 Quasar host galaxy’s light profile

We choose an elliptical Sérsic profile to model the smooth component of the quasar host galaxy’s light distribution. Additionally, we use a basis set of shapelets to reconstruct the non-smooth features in the extended source light distribution (Refregier 2003; Birrer et al. 2015). The set of shapelets is characterized with a scale size  $\zeta$  and maximum polynomial order  $n_{\text{max}}$ . The order  $n_{\text{max}}$  determines the total number of shapelet components  $n_{\text{shapelet}} = (n_{\text{max}} + 1)(n_{\text{max}} + 2)/2$ .

We model the quasar images as point-sources on the image plane convolved with the reconstructed point spread function (PSF). We let the amplitudes of each quasar image free.

## 5 LENS MODEL SETUPS

In this section, we present the specific model choices for DES J0408–5354. Extending on the baseline models, we choose different options – that we consider equally viable – for some particular components of the model. A combination of these options then make up our model settings  $S$  for each mass profile family  $M$ . To be specific, the model settings  $S$  include the model components describing the source and the line-of-sight galaxies, and the model settings  $M$  include the mass and light profiles of the central deflector

galaxy G1. We marginalize over these model settings  $S$  to account for any possible source of systematics that may be introduced from adopting only one specific choice. We first state the baseline models in Section 5.1. Then, we elaborate on the different additional model choices in Sections 5.2–5.9, and we summarize the set of model settings  $S$  that we marginalize over in Section 5.11. A summary of the adopted models and the parameter priors are tabulated in Appendix G.

### 5.1 Baseline models

The specifics of the baseline models agreed by the participating independent modelling teams are:

(i) Central deflector G1’s mass profile: power-law profile and composite profile (elliptical NFW potential for the dark component, double Chameleon convergence for luminous component),

(ii) Central deflectors G1’s light profile:

(a) *For models with power-law mass profile:* double Sérsic profile in all three bands,

(b) *For models with composite mass profile:* double Chameleon light profile in the F160W band linked with the double Chameleon mass profile, double Sérsic profiles in *UVIS* bands,

(iii) Satellite G2’s mass profile: singular isothermal sphere (SIS) placed on G1’s lens plane,

(iv) External shear,

(v) Explicit modelling of the line-of-sight galaxies G3–G6, multilens-plane treatment for G3,

(vi) Multisource-plane treatment for quasar host S1 and additional source component S2.

In the next sections, we explain these model settings and further extend on some of these settings as we see fit.

### 5.2 Central deflector G1’s mass and light profiles

We choose two sets of mass profile for the central deflector G1: power-law mass profile and composite mass profile.

For the corresponding light profile distribution of G1 with the power-law mass profile, we adopt a double Sérsic profile in the IR band, and a single Sérsic profile for each of the *UVIS* bands. Here, we deviated from the baseline model of double Sérsic profile for the *UVIS* bands, as we find a single Sérsic profile for each of the *UVIS* bands is sufficient and the posteriors of the lens model parameters are almost identical between the double Sérsic and single Sérsic profiles for the *UVIS* bands. Therefore, we adopt the single Sérsic profile for the *UVIS* bands to increase numerical efficiency by simplifying our model. However, we still use the double Sérsic profile in the IR band where the signal-to-noise ratio of the galaxy light is higher and thus more flexibility is needed to render it within the noise. The centroids are joint for all the Sérsic profiles between the bands. The axial ratio  $q_L$ , position angle  $\phi_L$ , and the Sérsic index  $n_s$  are also joint between the *UVIS* bands. We let effective radius  $\theta_{\text{eff}}$  and amplitude  $I_{\text{eff}}$  as free parameters independently for all bands to allow a colour gradient.

For the composite mass profile, we model the dark matter distribution with a NFW profile with ellipticity parametrized in the potential. For the baryonic matter distribution, we adopt two concentric Chameleon profiles to model both the luminous mass distribution and the light distribution in the F160W band. We join the scaling and ellipticity parameters of each pairing of the Chameleon profiles between the baryonic mass distribution and the F160W

light distribution. We do not fix the amplitude ratio between the two Chameleon profiles and this ratio is sampled as a non-linear parameter in our model. For each of the two *UVIS* bands, we adopt a single Sérsic profile. Similar to power-law profile, the Sérsic profile parameters except  $\theta_{\text{eff}}$  and  $I_{\text{eff}}$  are joint between the *UVIS* bands and the centroids of the all the deflector light profiles are joint together. The amplitudes of the mass and light profiles are independent of each other, thus we allow the mass-to-light ratio ( $M/L$ ) to be free. We adopt a Gaussian prior equivalent to  $12.74 \pm 1.71$  arcsec for the NFW scale radius  $r_s$  based on the results of Gavazzi et al. (2007) for the Sloan Lens ACS (SLACS) survey lenses (Bolton et al. 2006). G1’s velocity dispersion and redshift are within the range of those from the SLACS lenses, thus SLACS is a representative sample of elliptical galaxies such as G1 (Treu et al. 2006). Similar priors were adopted in previous H0LiCOW analyses of time-delay lenses (e.g. Wong et al. 2017; Rusu et al. 2020).

We find the half-light radius  $\theta_{\text{eff}}$  of the Sérsic profiles to be degenerate with the Sérsic index  $n_s$  in our models and the models tend to optimize towards large values of  $\theta_{\text{eff}}$  that is inconsistent with our observational prior. To prevent  $\theta_{\text{eff}}$  from converging towards abnormally large values, we impose an empirical prior on  $\theta_{\text{eff}}$ . We derive a scaling relation from the distribution of the central velocity dispersion  $\sigma_{e/2}$  measured within half of effective radius and  $R_{\text{eff}}$  in physical unit for the lenses in the SLACS sample (Auger et al. 2010). We account for intrinsic scatter in the derived scaling relation as we are ignoring the average surface brightness  $\bar{I}(R_{\text{eff}})$  in the relation between the three quantities along the Fundamental Plane. Then, we derive a distribution for  $R_{\text{eff}}$  for DES J0408–5354 for the given central velocity dispersion measurements from Table 3. In practice, we simultaneously sample  $R_{\text{eff}}^{\text{J0408}}$  and the parameters  $\{m, b, S\}$  – for the scaling relation

$$\log_{10}(\sigma_{\text{los}}/\text{km s}^{-1}) = m \log_{10}(R_{\text{eff}}/\text{kpc}) + b \quad (46)$$

with scatter  $S$  – from a joint likelihood for the SLACS sample data and the measured velocity dispersions of DES J0408–5354. For each sampled  $R_{\text{eff}}^{\text{J0408}}$ , we transform the measured central velocity dispersions within each aperture into  $\sigma_{e/2}$  using the aperture correction formulae given by Jorgensen, Franx & Kjaergaard (1995). We include the intrinsic scatter in the likelihood term for DES J0408–5354’s velocity dispersions, thus the scatter in the scaling relation propagates into the  $R_{\text{eff}}^{\text{J0408}}$  distribution. We estimate the scaling relation parameters as  $m = 0.18_{-0.04}^{+0.05}$ ,  $b = 2.2 \pm 0.04$ ,  $S = 1.53 \pm 0.05$ . We convert the  $R_{\text{eff}}^{\text{J0408}}$  distribution in physical unit into  $\theta_{\text{eff}}$  distribution in angular unit using the angular diameter distance to DES J0408–5354 for our fiducial cosmology, however we add 10 per cent uncertainty to the distribution to remove any strong dependence on the choice of cosmology. We take a Gaussian prior with the same mean and standard deviation of the resultant  $\theta_{\text{eff}}$  distribution from the SLACS lenses (Table G2). We adopt this prior only to prevent  $\theta_{\text{eff}}$  from veering off to very large values. The adopted prior is broad enough not to bias the  $\theta_{\text{eff}}$  posterior within the plausible range of values, including for the double Sérsic profile.

### 5.3 Satellite G2’s mass and light profile

In addition to the power-law or composite mass profile for the central deflector, we add a singular isothermal sphere (SIS) profile for G2’s mass distribution and a circular Sérsic profile for its light distribution. The Sérsic profile parameters except  $\theta_{\text{eff}}$  are joint between all bands. We join the centroid between the SIS and Sérsic profiles. Although a deviation from the isothermal profile in G2’s mass can potentially change the deflection potential at image C,

such a change will be negligible in our inference of  $H_0$  as time-delays with respect to image C are not used in our inference. The SIS profile is sufficient to capture the astrometric position of the image C in our modelling.

### 5.4 Nearby line-of-sight galaxies

We explicitly model the mass distributions of line-of-sight galaxies G3–6 to fully capture their higher than second-order lensing effects that cannot be accounted for by the external convergence and the external shear profiles. First in Section 5.4.1, we describe our selection criterion for the line-of-sight galaxies to explicitly include in our lens model. Then in Section 5.4.2, we explain the multilens-plane treatment of the line-of-sight galaxies. Lastly in Sections 5.4.3 and 5.4.4, we describe the mass profiles we adopt to model these line-of-sight galaxies.

#### 5.4.1 Selection criterion of the line-of-sight galaxies for explicit modelling

To select the line-of-sight galaxies for explicit modelling, we first estimate the contribution in time-delays between the images from higher than second-order derivatives of the deflection potential of these galaxies. To quantify this effect, we set an SIS profile for each perturber with its Einstein radius corresponding to the estimated central velocity dispersion. We infer the velocity dispersion for all the line-of-sight galaxies from their stellar masses using two scaling relations – one from Auger et al. (2010) and the other from Zahid et al. (2016). To be conservative, we choose the upper limit of the  $1\sigma$  confidence interval of the estimated stellar mass and choose the larger value of the velocity dispersions estimated from the two scaling relations (Buckley-Geer et al. 2020). We select the line-of-sight galaxies that may cause more than 1 per cent shift in the measured Hubble constant if higher than second-order derivatives of their deflection potential are ignored. The shift in the Hubble constant can be related to the relative astrometric shift  $\delta\theta_{AB}$  between image A and B as

$$\frac{\delta H_0}{H_0} \lesssim \frac{D_{\Delta t}}{c \Delta t_{AB}} (\theta_A - \theta_B) \cdot \delta\theta_{AB} \quad (47)$$

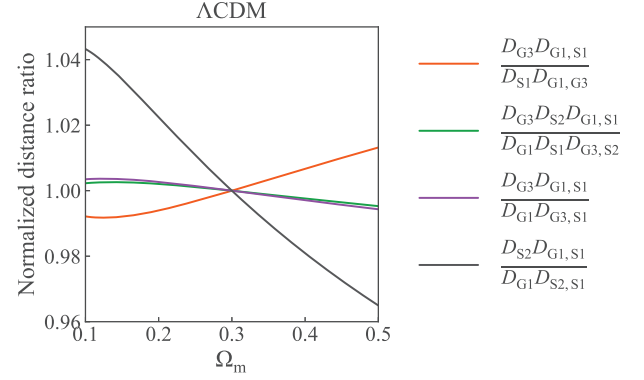
(Birrer & Treu 2019). We take the relative astrometric shift  $\delta\theta_{AB} = \alpha_A^{(3)} - \alpha_B^{(3)}$ , where the term on the right-hand side is the relative deflection angle for third and higher order lensing effects from the SIS profile corresponding to each line-of-sight galaxy. Thus, we set the selection criterion

$$\frac{D_{\Delta t}^{\text{fiducial}}}{c \Delta t_{AB}} (\theta_A - \theta_B) \cdot (\alpha_A^{(3)} - \alpha_B^{(3)}) \geq 0.01. \quad (48)$$

This criterion selects G3–G6 for explicit modelling. Note that the perturber selection criterion based on the ‘flexion shift’  $\Delta_{3,x} > 10^{-4}$  also selects G3–G6 for explicit modelling (McCully et al. 2017; Sluse et al. 2019, Buckley-Geer et al. 2020).

#### 5.4.2 Multilens-plane modelling of the line-of-sight galaxies

We model this lens system with a multilens-plane treatment by setting G3’s lens plane at its own redshift  $z_{G3} = 0.769$ , as G3 is close enough to G0 to cause more than 1 per cent deviation in the computed time-delays if we place it on G0’s lens plane. We place G4–G6 on G0’s lens plane as we assume that the deviation in computed time-delays due to this assumption is negligible given the combinations of their stellar masses and distances from G0.



**Figure 2.** Impact of varying  $\Omega_m$  in the  $\Lambda$ CDM cosmology on the angular diameter distance ratios between the lens and source planes. All the distance ratios except for the black line changes less than 1 per cent for a wide range of  $\Omega_m$ . The black line corresponds to the distance ratio involving S2’s lens plane. As the S2’s Einstein radius is small ( $\sim 0.002$  arcsec), the change in the black line only has a small effect on the effective Fermat potential [cf. equation (49)]. Therefore, fixing the distance ratios for the fiducial cosmology with  $\Omega_m = 0.3$  is not a strong assumption in our analysis. See Appendix A for tests validating this point.

Additionally, we model the mass profile of S2 at its redshift  $z_{S2} = 2.228$ . Therefore, we have three lens-planes in our model. We can express the effective Fermat potential for the triple-lens-plane case from equation (9) as

$$\begin{aligned} \phi^{\text{eff}}(\theta) = & \left[ \frac{D_{G3} D_{G1,S1}}{D_{S1} D_{G1,G3}} \frac{(\theta_{G1} - \theta_{G3})^2}{2} - \psi_{G1}(\theta_{G1}) \right] \\ & + \frac{1 + z_{G3}}{1 + z_{G1}} \left[ \frac{D_{G3} D_{S2} D_{G1,S1}}{D_{G1} D_{S1} D_{G3,S2}} \frac{(\theta_{G3} - \theta_{S2})^2}{2} \right. \\ & \left. - \frac{D_{G3} D_{G1,S1}}{D_{G1} D_{G3,S1}} \psi_{G3}(\theta_{G3}) \right] \\ & + \frac{1 + z_{S2}}{1 + z_{G1}} \frac{D_{S2} D_{G1,S1}}{D_{G1} D_{S2,S1}} \left[ \frac{(\theta_{S2} - \beta)^2}{2} - \psi_{S2}(\theta_{S2}) \right]. \quad (49) \end{aligned}$$

Here,  $\theta_G$  is the quasar’s image position on G’s plane with  $G \in \{G1, G3, S2\}$ ,  $\psi_G$  is the deflection potential of G, and  $\beta$  is the quasar’s position on the source-plane. We fix the distance ratios in the above equation in our modelling. We adopt the  $\Lambda$ CDM cosmology with the cosmological density parameters  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$  to obtain these distance ratios. The relevant distance ratios change by less than 1 per cent within  $0.25 \lesssim \Omega_m \lesssim 0.35$  and  $-1.1 \lesssim w \lesssim -0.9$  (Fig. 2). Therefore, adopting this fiducial cosmology is only a weak assumption in our analysis. Fixing these distance ratios does not linearly affect our inference of  $H_0$ , as the ratios do not depend on  $H_0$ . However, there can potentially be a small non-linear shift in the inferred  $H_0$  from our analysis had we adopted a different set of values for  $\Omega_m$  and  $\Omega_\Lambda$ . The non-linear effect on  $H_0$  from fixing density parameters in the multilens-plane treatment was demonstrated to be less than 1 per cent for two previously analysed lens systems, HE 0435–1223 and WFI 2033–4723 (Wong et al. 2017; Rusu et al. 2020). In Appendix A, we show that  $H_0$  shifts by less than 1 per cent if we change the matter density parameter to  $\Omega_m = 0.1$  and to  $\Omega = 0.45$  within the  $\Lambda$ CDM cosmology. This range in  $\Omega_m$  covers nearly the full range of our prior  $\Omega_m \in [0.05, 0.5]$  for inferring  $H_0$  for the  $\Lambda$ CDM cosmology. As a shift less than 1 per cent in  $H_0$  is much smaller than the typical precision on  $H_0$  ( $\sim 5$ –8 per cent) allowed by the current data quality, we consider



that the impact of fixing the distance ratios using a fiducial  $\Lambda$ CDM cosmology has negligible impact in our analysis. However, we find that our inference of  $H_0$  is sensitive to the dark energy equation of state parameter  $w$  in the  $w$ CDM cosmology. As we adopt a double source plane model – as described in Section 5.6 – the distance ratios or the  $\zeta$  terms in equation (8) become sensitive to  $w$  (Gavazzi et al. 2008; Collett et al. 2012; Collett & Auger 2014). Therefore, the distance posteriors from this analysis should not be used to infer  $H_0$  in extended cosmologies other than the  $\Lambda$ CDM model. We postpone the derivation of a distance posterior in more general cosmologies to future work.

We model G3 and S2 with SIS profiles. We place G3 at its ‘true’ position on its own lens plane by tracing back from its observed position accounting for the foreground deflectors. As we also model the flux distribution from S2, we join the centroid of S2’s mass profile with its light centroid on its plane.

For G4–G6, we fix their centroids at their observed position on the lens plane of G1. For the mass profiles of G4–G6, we adopt two choices: the SIS profile and the spherical NFW profile. We choose the additional NFW model for G4–G6 as the NFW scale radius estimated from each of their stellar masses is smaller than the distance between the galaxy and G0 (Fig. 3, Section 5.4.4). Thus, their mass profile slopes can potentially be different from the isothermal profile at the centre of G0. In Sections 5.4.3 and 5.4.4, we describe the priors for the SIS and NFW profile parameters, respectively, of the line-of-sight galaxies.

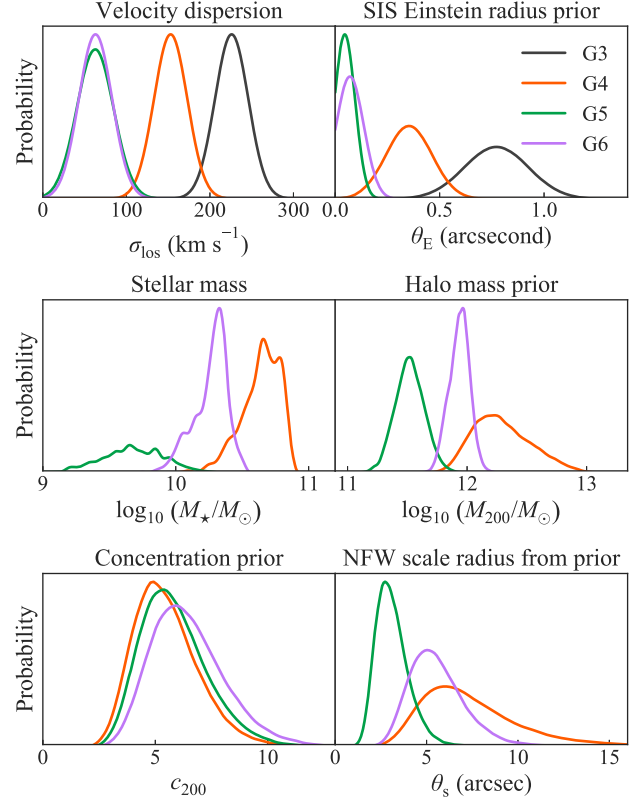
#### 5.4.3 SIS profile for the line-of-sight galaxies

We estimate the SIS Einstein radius distribution from each galaxy’s SIS velocity dispersion  $\sigma_{\text{SIS}}$  using the relation

$$\theta_{\text{E,SIS}} = 4\pi \left( \frac{\sigma_{\text{SIS}}}{c} \right)^2 \frac{D_{\text{G,S1}}}{D_{\text{S1}}}, \quad (50)$$

where  $\theta_{\text{E,SIS}}$  is the Einstein radius for an SIS profile, and  $D_{\text{G,S1}}$  is the angular diameter distance between a line-of-sight galaxy  $G \in \{G3, G4, G5, G6, S2\}$  and S1. We calculate the distance ratio in the above equation using our fiducial cosmology. We do not need to add uncertainty to the fiducial cosmology used here as the distance ratios are independent of  $H_0$  and a large shift (e.g. by 0.1) in  $\Omega_m$  changes them by negligible amount relative to the  $20 \text{ km s}^{-1}$  uncertainty we adopted for the velocity dispersions.

G4 and G6’s observed morphologies indicate that they are elliptical galaxies. Therefore, we take their observed stellar velocity dispersions  $\sigma_{\text{ap}}$  as  $\sigma_{\text{SIS}}$  in equation (50) to obtain these galaxies’ Einstein radius prior distributions (Treu et al. 2006; Auger et al. 2010). In contrast, G5’s spectra contains bright [O II] emission lines indicative of a star-forming galaxy. We also take S2 as a star-forming galaxy due to its blue colour in the *HST* three-band imaging (Fig. 1). Therefore, we estimate the rotational velocities  $v_c$  of G5 and S2 from their observed ‘aperture-averaged’ velocity dispersions  $\sigma_{\text{ap}}$  using the scaling relation between  $v_c^2/\sigma_{\text{ap}}^2$  and Sérsic index  $n_s$  from Agnello, Evans & Romanowsky (2014). We obtain the Sérsic index of G5  $n_s = 4$  by fitting a Sérsic profile to its light distribution in the F814W band. From a preliminary lens model, we adopt S2’s Sérsic index as  $n_s = 1.5$ . For these Sérsic indices, the  $v_c^2/\sigma_{\text{ap}}^2$  ratios are approximately 2.5 and 2.2, respectively, for G5 and S2. We adopt a Gaussian uncertainty with standard deviation 0.2 for these ratios to account for the scatter observed in the  $v_c^2/\sigma_{\text{ap}}^2 - n_s$  distribution (cf. fig. 6 of Agnello et al. 2014). Then, to convert the estimated rotational velocity  $v_c$  into the corresponding SIS velocity dispersion  $\sigma_{\text{SIS}}$ , we



**Figure 3.** Observed and estimated properties of the line-of-sight galaxies G3–G6. Top left: velocity dispersions derived from the MUSE integral field spectra. Top right: SIS Einstein radius distributions obtained from the observed velocity dispersions. Middle left: estimated stellar masses from Buckley-Geer et al. (2020). Middle right: halo mass  $M_{200}$  inferred from the estimated stellar mass using the stellar mass–halo mass relation from Behroozi et al. (2019). Bottom left: halo concentration parameter  $c_{200}$  obtained using a halo mass–concentration relation for our fiducial cosmology (Diemer & Joyce 2019). Bottom right: scale radius of the NFW profile in angular unit for our fiducial cosmology from the  $M_{200}$  and  $c_{200}$  priors. The intrinsic scatter and uncertainties of the adopted scaling relations are accounted for at each conversion step. We use the SIS Einstein radius distributions as priors for the SIS model and the  $M_{200}$  and  $c_{200}$  distributions as priors for the NFW model for G4–G6.

use the relation

$$\sigma_{\text{SIS}}^2 = \frac{GM(R)}{2R} = \frac{v_c(R)^2}{2}, \quad (51)$$

where  $M(R)$  is the enclosed 3D mass within a radius  $R$ . The estimated SIS velocity dispersions are  $\sigma_{\text{SIS}}^{\text{G5}} = 62 \pm 22 \text{ km s}^{-1}$  and  $\sigma_{\text{SIS}}^{\text{S2}} = 48 \pm 11 \text{ km s}^{-1}$ . We parametrize the SIS Einstein radius distributions derived from the velocity dispersions as Gaussian priors for the SIS mass profiles of G3–G6 and S2.

#### 5.4.4 NFW profile for the line-of-sight galaxies

We parametrize the NFW profiles for G4–G6 with the halo mass  $M_{200}$  and concentration  $c_{200}$ . We obtain the priors on the NFW profile parameters from the estimated stellar masses of G4–G6 (Buckley-Geer et al. 2020). We derive the halo mass distribution from the stellar mass distribution using the stellar mass–halo mass relation from Behroozi et al. (2019) for the respective redshift of the line-of-sight galaxy. We weight the halo-mass distribution with the halo

**Table 4.** Evidence for different lens model setups. The evidence ratio  $\Delta \log \mathcal{Z}$  is calculated only within the particular mass profile family – power law or composite. The model setups are ordered from higher to lower evidence within each mass profile family. The relative weights for each model are obtained from the evidence ratios adjusted for sparse sampling from the model space as described in Section 6.1.

Mass profile	Source $n_{\max}$	$z_{S3}$	G4–G6 mass profile	$\log \mathcal{Z}$ ( $\pm 24$ )	$\Delta \log \mathcal{Z}$ ( $\pm 34$ )	Relative weight, $W_{\text{rel}}$
Power law	{8, 5, 3}	2.375	SIS	–25087	0	1.00
Power law	{6, 3, 2}	2.375	SIS	–25215	128	0.77
Power law	{8, 5, 3}	2.228	SIS	–25232	145	0.74
Power law	{7, 4, 2}	2.375	SIS	–25317	230	0.60
Power law	{7, 4, 2}	2.228	SIS	–25421	333	0.45
Power law	{6, 3, 2}	2.228	SIS	–25450	363	0.41
Power law	{8, 5, 3}	2.375	NFW	–25578	490	0.26
Power law	{7, 4, 2}	2.228	NFW	–25624	537	0.22
Power law	{6, 3, 2}	2.228	NFW	–25656	569	0.19
Power law	{6, 3, 2}	2.375	NFW	–26432	1345	0.00
Power law	{8, 5, 3}	2.228	NFW	–26469	1382	0.00
Power law	{7, 4, 2}	2.375	NFW	–26551	1464	0.00
Composite	{7, 4, 2}	2.228	SIS	–25055	0	1.00
Composite	{8, 5, 3}	2.228	SIS	–25121	66	0.96
Composite	{8, 5, 3}	2.375	SIS	–25147	92	0.94
Composite	{6, 3, 2}	2.228	SIS	–25155	100	0.94
Composite	{7, 4, 2}	2.375	SIS	–25155	100	0.93
Composite	{6, 3, 2}	2.375	NFW	–25252	197	0.87
Composite	{6, 3, 2}	2.375	SIS	–25292	237	0.85
Composite	{7, 4, 2}	2.375	NFW	–25482	427	0.72
Composite	{6, 3, 2}	2.228	NFW	–25985	930	0.44
Composite	{8, 5, 3}	2.375	NFW	–26541	1486	0.22
Composite	{8, 5, 3}	2.228	NFW	–27073	2018	0.09
Composite	{7, 4, 2}	2.228	NFW	–28979	3924	0.00

mass function for our fiducial cosmology and the relevant redshift from Tinker et al. (2008). We obtain the concentration distribution from the halo mass distribution for our fiducial cosmology using the  $M$ – $c$  relation from Diemer & Joyce (2019). We propagate the uncertainties and scatters in these relations when deriving one quantity from another. The  $M_{200}$  priors and  $c_{200}$  priors for G4–G6 are shown in Fig. 3. We can also derive the NFW scale radius  $r_s = R_{200}/c_{200}$  in physical unit, and convert it to the scale radius  $\theta_s$  in angular unit given our fiducial cosmology (Fig. 3). We do not use these scale radii as prior, we only show the distributions to motivate our choice of the NFW profile for the galaxies G4–G6.

### 5.5 Galaxy group containing G1

We do not explicitly model the galaxy group that contains the central deflector G1 [Group 5 in Buckley-Geer et al. (2020)]. The estimated flexion shift  $\log_{10} \Delta x_3 = -3.86^{+0.97}_{-0.72}$  for this group is marginally above the conservative threshold  $\Delta x_3 > 10^{-4}$  (Buckley-Geer et al. 2020). However, the larger end of the flexion shift is provided by the case where the group is centred near to the central deflector. In that case, the group’s halo coincides with the deflector’s halo, which is already accounted for in our lens models. However, if the group’s centroid is offset from the central deflector, then the flexion shift becomes smaller. Then, the group’s contribution can be considered only in the approximated convergence, as the external shear profile already captures the shear contribution from the group. In Appendix C, we show that the impact of the group’s convergence, if explicitly accounted for, only shifts  $H_0$  by 0.4 per cent. This small shift justifies our choice of not including the group in our lens model.

### 5.6 Source component light profiles

We use an elliptical Sérsic light profile and a set of shapelets to reconstruct the quasar host S1’s light profile. We join all the Sérsic profile parameters across the three bands. We join the shapelet scale size  $\zeta$  across the *UVIS* bands and leave  $\zeta$  in the IR band as a free parameter.

To reconstruct S2’s light profile, we take a basis set of one elliptical Sérsic profile and multiple shapelets. We join the Sérsic profile parameters and the shapelet scale size  $\zeta$  across bands. For S3’s light profile, we adopt only an elliptical Sérsic profile. All the profile parameters for this profile except the amplitude  $I_{\text{eff}}$  are joint across the three bands.

For each model setup, we choose three fixed values of  $n_{\max}$ : S1’s  $n_{\max}$  in the IR band, S1’s  $n_{\max}$  in the *UVIS* band, and S2’s  $n_{\max}$  common across the three bands. We adopt three different sets of  $\{n_{\max}^{\text{S1, IR}}, n_{\max}^{\text{S1, UVIS}}, n_{\max}^{\text{S2}}\}$ : {6, 3, 2}, {7, 4, 2}, {8, 5, 3}. A minimum number of shapelets is necessary to sufficiently capture the complex structures in the lensed arcs, however we would start to fit the noise in the data by adopting too many more shapelets than necessary. We choose these values for  $n_{\max}$  so that we hit a balance between these two scenarios. As we show in Table 4, the model evidence peaks around these values leading to our choice of these  $n_{\max}$  values. We check if the inferred  $H_0$  value from our analysis depends on the particular range of  $n_{\max}$  values adopted above. We find that a set of larger  $n_{\max}$  values {12, 9, 9, 4} and a set of smaller  $n_{\max}$  values {2, 2, 2, 2} both infer  $H_0$  within the range spanned by the models with our adopted  $n_{\max}$  values. Thus, our inference of  $H_0$  is robust against the particular range of adopted  $n_{\max}$  values.

We place the additional source component S2 at the source plane with redshift  $z_{S2} = 2.228$ . As we do not know the redshift of S3, we adopt two choices for its redshift:  $z_{QSO}$  and  $z_{S2}$ .

For the model where we place S3 on S2's plane, we ignore the mass distribution of S3. From our lens model, we find that S3 is approximately twice further away from the quasar position on S2's plane than S2. We run a lens model ignoring S2's mass profile as well and find that the time-delay distance shifts by 0.94 per cent. The total flux from the reconstructed source light distribution of S2 and S3 are comparable after accounting for lensing magnification. If S2 and S3 are at a similar redshift, then they have similar total mass. If we assume SIS profile for S2 and S3, then the convergence of S3 at the quasar position would be approximately half of that from S2. We estimate that the time-delay distance will shift by  $\lesssim 0.5$  per cent due to ignoring S3's mass distribution, if it indeed lies at a similar redshift of S2. This shift is negligible compared to the typical uncertainty (5–8 per cent) on the estimated time-delay distance given the quality of the current data. Therefore, we do not include its mass distribution in our model as we do not know the true redshift of S3.

### 5.7 Potential additional image C2 split from the image C

A faint blob is noticeable on a few pixels toward North-East from the position of G2 in the F160W band. This blob can potentially be another quasar image C2 split off from the image C by the nearby satellite G2. This potential additional image is not noticeable in the *UVIS* bands, but this non-detection in the *UVIS* bands can be caused by differential extinction through G2. If such an additional image is predicted by the model, we allow the model to assign point-source-like flux at the position of the predicted additional image. Note that we do not impose the existence of this additional image in the model.

### 5.8 *HST* image region for likelihood computation

To compute the imaging likelihood, we choose a large enough circular region of the *HST* image centred on the deflector galaxy in each band so that it contains most of the flux from the lens system in that particular band. The radii of these regions are 4.3, 3.3, and 3.3 arcsec in the F160W, F814W, and F475X bands, respectively. We mask out some of the pixels around the faint blob visible between G1 and G3 to block its light that would otherwise be within the chosen apertures. See the 'Normalized Residual' plots of Figs 4 or 5 for the shape of the likelihood computation regions. In Appendix B, we show that this particular choice of likelihood computation regions is not a source of bias in our analysis.

### 5.9 Dust extinction by the satellite G2

The satellite G2 may cause differential dust extinction to the lensed light distribution near image C. Ignoring this differential extinction may produce poor fitting around image C in the modelling. To account for this effect, we multiply a differential extinction factor  $\exp[-\tau_\lambda(\theta_1, \theta_2)]$  to the lensed light distribution from the quasar host galaxy in all three bands. Here,  $\tau_\lambda$  is equivalent to an optical depth parameter. We set the differential extinction profile proportional to G2's IR surface brightness with a wavelength-dependent normalization. Therefore, we take  $\tau_\lambda(\theta_1, \theta_2) = \tau_\lambda^0 I_{G2}(x, y)$ , where  $I_{G2}(\theta_1, \theta_1)$  is G2's light distribution parametrized with a Sérsic profile as described in Section 5.3. Thus, we are only modelling the differential extinction effect by G2 and this extinction goes to zero far away from G2. We do not model the differential extinction effect for the central deflector G1 as elliptical galaxies like G1 are typically dust-poor. We connect the proportionality constant  $\tau_\lambda^0$  for

each band using the differential extinction law of Cardelli, Clayton & Mathis (1989) with  $R_V = 3.1$ . As a result, we only have  $\tau_{F814W}^0$  as a non-linear parameter in our model. As a check, we run a lens model with the proportionality constant  $\tau_\lambda^0$  in each band independent of each other and we find that the three constrained  $\tau_\lambda^0$  parameters follow the extinction law from Cardelli et al. (1989) for  $R_V \sim 3-5$ . The amplitudes of the quasar images are free parameters, therefore any possible differential extinction effect in the quasar image flux is already accounted for.

### 5.10 Requirement for astrometric precision

For the lens system DES J0408–5354, a precision of 6 mas is required in the estimated source position to match the precision of the most precise time delay,  $\Delta t_{AB}$  (Birrer & Treu 2019). Given the magnification and the multiplicity of the images, this precision in the source position translates to an astrometric precision of approximately 40 mas for each image position on the image plane under a fixed lens model. As we can constrain the image positions in our models within 10 mas, we meet the requirement for astrometric precision. We expect any non-accounted astrometric uncertainty on the level of 10 mas or below to be subdominant in the error budget and the systematic impact.

### 5.11 Model choice combinations

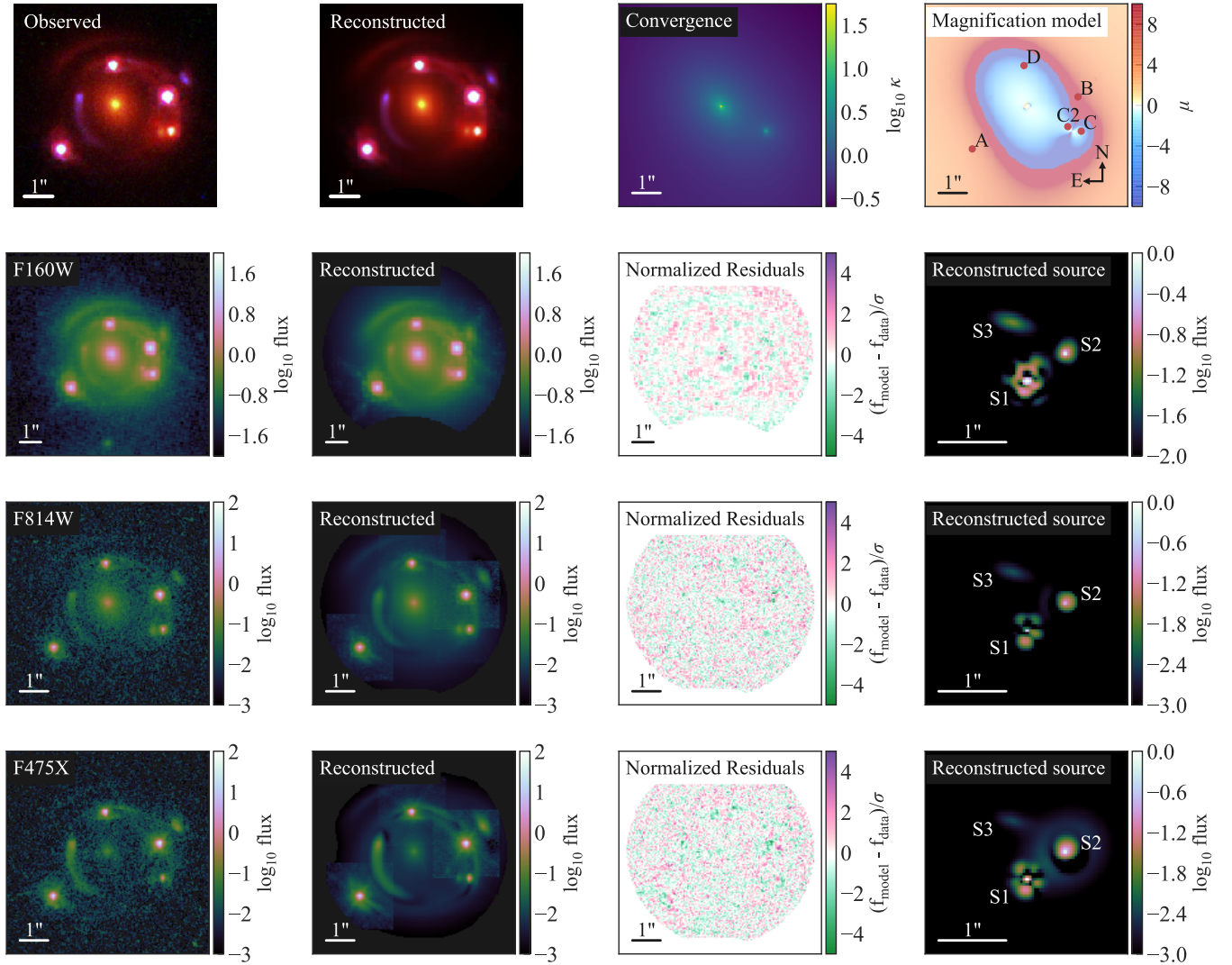
Assembling the different choices described above for various components in our models, we have the following options that we vary:

- (i) Central deflector G1's mass profile: power-law, composite,
- (ii) Source  $n_{\max}$ : {6, 3, 2}, {7, 4, 2}, {8, 5, 3},
- (iii) S3 redshift:  $z_{QSO} = 2.375$ ,  $z_{S2} = 2.28$ , and
- (iv) G4–G6 mass profile: SIS, NFW.

Taking all possible combinations of these choices, we have 24 different models in total – 12 for the power-law and 12 for the composite mass profiles. All the light profiles for lens and source light distribution form a linear basis set, thus all the amplitude parameters are linear (Birrer et al. 2015; Birrer & Amara 2018). We have 85–137 linear parameters and 57–62 non-linear parameters in the 24 model setups with either power-law or composite mass profiles.

We can compare the number of chosen models in this study with the 128 model runs performed in the cosmographic analysis of SDSS 1206+4332 (Birrer et al. 2019). Since Birrer et al. (2019) performed two separate sampling runs for the same model, these authors adopt 64 different models in practice combining the power-law and composite mass profiles. As Birrer et al. (2019) find that explicitly accounting for the non-linear components of the foreground shear has negligible impact in the cosmographic analysis, we choose not to include it in our analysis. Note, the linear components of the foreground shear is already accounted by the adopted external shear profile. Moreover, Birrer et al. (2019) incorporate two different likelihood-computation region sizes in their model choices, whereas we do not vary it in our analysis as we show that our analysis is stable against different choices of the likelihood-computation region size (Appendix B). As a result, the comparable number of models in Birrer et al. (2019) is 16 to contrast with our adopted model number of 24. These numbers, although not identical, are comparable and difference between the exact number of chosen models to check systematics can arise naturally due to different complexity in different lens systems.





**Figure 4.** The most likely lens model and reconstructed image of DES J0408–5354 using the power-law model. The top row shows the observed RGB image, reconstructed RGB image, the convergence profile, and the magnification model in order from the left-hand side to the right-hand side. The next three rows show the observed image, the reconstructed image, the residual, the reconstructed source in order from the left-hand side to the right-hand side for each of the *HST* filters. The three filters are F160W (second row), F814W (third row), and F475X (fourth row). All the scale bars in each plot correspond to 1 arcsec. The patchy or ring-like artefacts in the source reconstruction translate to lensed features below the noise level in the image, thus they do not affect our lens model.

## 6 LENS MODELLING AND COSMOGRAPHIC INFERENCE

In this section, we first present the lens modelling results (Section 6.1), combine the time-delay and kinematics likelihoods with the lens model posterior to produce the cosmological distance posterior (Section 6.2), and infer  $H_0$  from the distance posterior (Section 6.3).

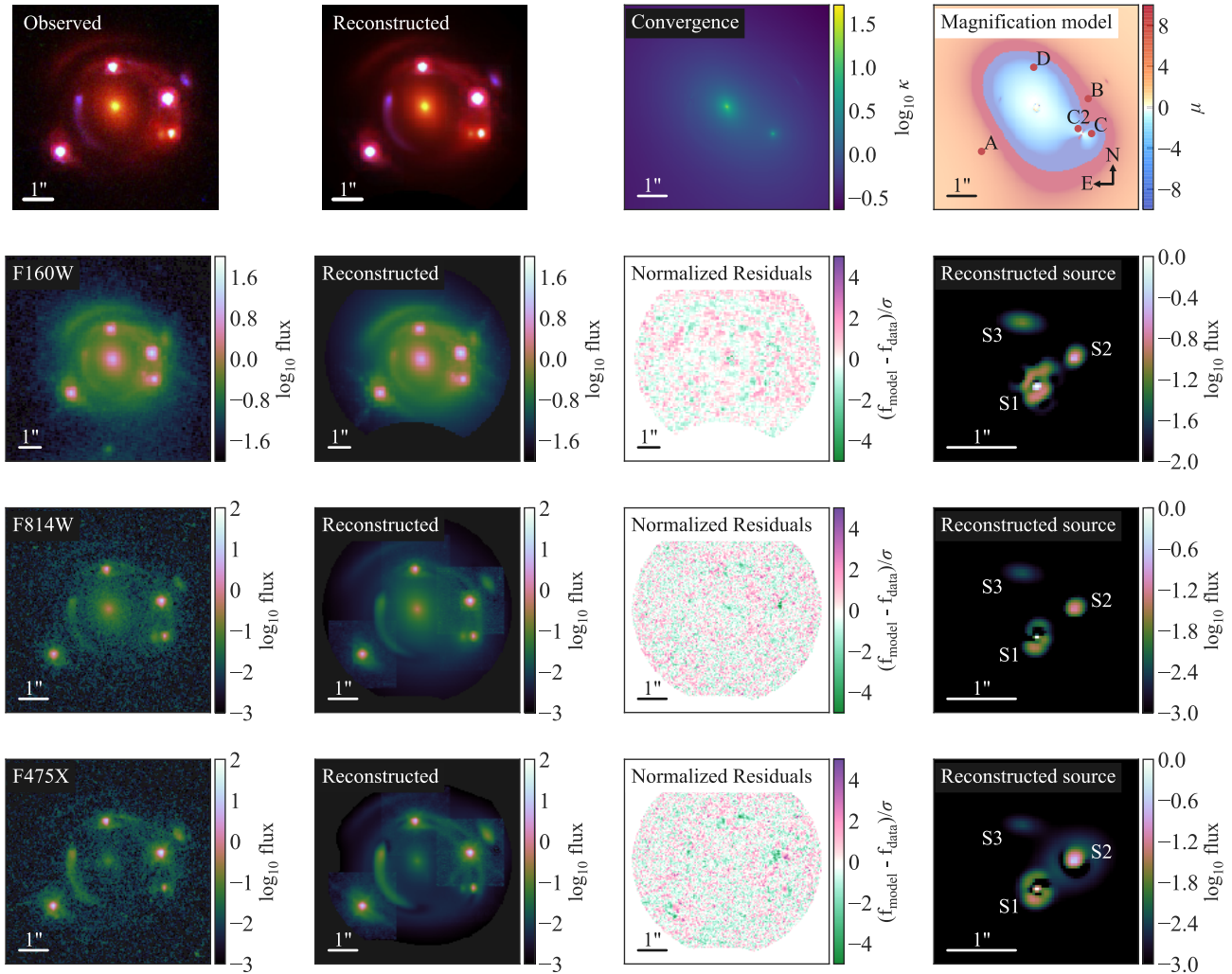
### 6.1 Modelling workflow and results

We simultaneously model the images from all three *HST* bands. For each model choice from Section 5.11, we reconstruct the PSF for each *HST* band. Thus, a set of three reconstructed PSFs is part of the model choice  $S$  that we marginalize (cf. equation 27). To initiate the PSF reconstruction, we take an initial PSF estimate by taking the median of a few ( $\sim 4$ – $6$ ) stars from each *HST* image and then re-centring the median PSF. At each iteration

of the PSF reconstruction process, we first realign the IR band’s coordinate system with the *UVIS* bands’ coordinate system using the quasar image positions (Shajib et al. 2019). Then, we optimize the lens model given the PSF from the initial estimate or the previous iteration of PSF reconstruction. Finally, we subtract the extended host-galaxy and the lens-galaxy light from the image and optimize the PSF using the residual quasar images (see for details Birrer et al. 2019, and for similar procedure Chen et al. 2016). We use LENSTRONOMY for lens modelling and PSF reconstruction and the particle swarm optimization (PSO) routine of COSMOHAMMER for optimizing the model (Kennedy & Eberhart 1995; Akeret et al. 2013; Birrer & Amara 2018). We repeat the set of the following three steps five times in total to reconstruct the PSF:

1. IR band image re-alignment,
2. lens model optimization using PSO,
3. PSF reconstruction.





**Figure 5.** The most likely lens model and reconstructed image of DES J0408–5354 using the composite model. The top row shows the observed RGB image, reconstructed RGB image, the convergence profile, and the magnification model in order from the left-hand side to the right-hand side. The next three rows show the observed image, the reconstructed image, the residual, the reconstructed source in order from the left-hand side to the right-hand side for each of the *HST* filters. The three filters are F160W (second row), F814W (third row), and F475X (fourth row). All the scale bars in each plot correspond to 1 arcsec. The patchy or ring-like artefacts in the source reconstruction translate to lensed features below the noise level in the image, thus they do not affect our lens model.

We check that the reconstructed PSF stabilizes after five such iterations as the PSFs from additional iterations do not lead to higher imaging likelihood.

After the PSF reconstruction, we simultaneously sample from the lens model posterior and compute the model evidence  $\mathcal{Z}$  using the dynamic nested sampling algorithm (Skilling 2004; Higson et al. 2018). We use the nested sampling software DYPOLYCHORD (Handley, Hobson & Lasenby 2015; Higson et al. 2019). In Appendix E, we describe the sampler settings, assess the numerical performance, and conclude that the chosen settings allow for sufficient exploration of the posterior space.

We perform our analysis while blinding  $H_0$  and other model parameters and lensing quantities directly related to  $H_0$ , i.e. the model-predicted time delays. We also blind the mass profile slope  $\gamma$  of the power-law model after the initial exploration stage to find a stable preliminary lens model. In practice, the mean of the blinded quantities are always subtracted from the distribution within the analysis software, so that the printed values or plotted distributions

are centred at zero. After all the co-authors had agreed during a teleconference on 2019 September 25 that sufficient amount of checks for modelling systematics were carried out, the analysis was frozen and the actual posterior distribution of  $H_0$  was revealed for the first time. We report this  $H_0$  posterior in this paper without any further alteration.

Figs 4 and 5 display the most likely models for the power-law and composite profiles, respectively. In addition to the lensed complex structures in the Einstein ring from the extended quasar host galaxy, the lensed arcs S2 and S3 are also reproduced very accurately. Moreover, the models reproduce the additional split image C2 on the other side of G2 from image C.

Table 4 tabulates the evidences for different model choices. We combine the model posteriors weighted by the evidence ratios within each lens model family – power-law or composite – to marginalize over the model choices. Previous studies – e.g. Birrer et al. (2019), Chen et al. (2019), Rusu et al. (2020) – use the Bayesian information criterion (BIC) as an estimate of the evidence for Bayesian model

averaging (BMA; e.g. Madigan & Raftery 1994; Hoeting et al. 1999). Whereas BIC estimates the model evidence based on the maxima of the likelihood function under certain assumptions, nested sampling directly computes the model evidence by integrating over the whole prior space. Hence, the evidence obtained from nested sampling is more robust.

We account for sparse sampling from the model space by down-weighting the evidence ratios between the models. Effectively, we want to estimate the factor  $\Delta S_n$  in equation (30) to account for sparse sampling. We estimate the sparsity of the sampled models by taking the variance of  $\Delta \log \mathcal{Z}$  between ‘neighbouring’ model pairs that differ in only one model setting. In this way, we are being conservative by accepting more variance in our lens model posterior to avoid any bias due to sparse sampling from the model space. For 12 models within each mass profile family, we then have 20 such ‘neighbouring’ models. We obtain  $\sigma_{\Delta \log \mathcal{Z}}^{\text{model}} = 436$  for the power-law models and  $\sigma_{\Delta \log \mathcal{Z}}^{\text{model}} = 1210$  for the composite models. We follow Birrer et al. (2019) to adjust the relative weights of the model by convolving the evidence ratios with a Gaussian kernel with standard deviation  $\sigma_{\Delta \log \mathcal{Z}} = (\sigma_{\Delta \log \mathcal{Z}}^{\text{model}^2} + \sigma_{\Delta \log \mathcal{Z}}^{\text{numeric}^2})^{1/2}$ . Here, we take  $\sigma_{\Delta \log \mathcal{Z}}^{\text{numeric}} = 34$  as explained in Appendix E. Following Birrer et al. (2019), we first calculate the absolute weight  $W_{n, \text{abs}}$  of the  $n$ th model by convolving the evidence uncertainty with the evidence ratio function  $f(x)$  as

$$W_{n, \text{abs}} = \frac{1}{\sqrt{2\pi}\sigma_{\Delta \log \mathcal{Z}}} \int_{-\infty}^{\infty} f(x) \exp\left[-\frac{(\log \mathcal{Z}_n - x)^2}{2\sigma_{\Delta \log \mathcal{Z}}^2}\right] dx, \quad (52)$$

where we define the evidence ratio function  $f(x)$  as

$$f(x) \equiv \begin{cases} 1 & x \geq \log \mathcal{Z}_{\text{max}}, \\ \exp(x - \log \mathcal{Z}_{\text{max}}) & x < \log \mathcal{Z}_{\text{max}}. \end{cases} \quad (53)$$

We then obtain the relative weight  $W_{n, \text{rel}}$  by simply normalizing the absolute weights by the maximum absolute weight as

$$W_{n, \text{rel}} = \frac{W_{n, \text{abs}}}{\max\{W_{n, \text{abs}}\}}. \quad (54)$$

In the limit of  $n \rightarrow \infty$ , we would have a perfect sampling of models from the model space. In that case, we have  $\sigma_{\Delta \log \mathcal{Z}}^{\text{model}} \rightarrow 0$  and  $\sigma_{\Delta \log \mathcal{Z}} \rightarrow \sigma_{\Delta \log \mathcal{Z}}^{\text{numeric}}$ . Furthermore, if the evidence value is perfectly computed with  $\sigma_{\Delta \log \mathcal{Z}}^{\text{numeric}} = 0$ , then the exponential function inside equation (52) becomes a Dirac delta function. In that limit, the relative weight  $W_{n, \text{rel}}$  approaches the evidence ratio as

$$\lim_{\sigma_{\Delta \log \mathcal{Z}} \rightarrow 0} W_{n, \text{rel}} = \frac{\mathcal{Z}_n}{\mathcal{Z}_{\text{max}}}. \quad (55)$$

In Fig. 6, we compare the posteriors of important lens parameters between power-law and composite mass profiles after marginalizing over the model space using the adjusted evidence ratios as described above.

## 6.2 Combining the time delays, kinematics, and external convergence information

To combine the time-delay likelihood with the lens imaging likelihood, we importance sample from the lens model posterior weighted by the joint time-delay and kinematics likelihood (Lewis & Bridle 2002). In Section 6.2.1, we fold in the time-delay and kinematic likelihoods into the lens model posterior. Then in Section 6.2.2, we add the external convergence distribution to the cosmological distance posteriors. Finally in Section 6.2.3, we check the impact of

microlensing time-delay on our inference of the effective time-delay distance.

### 6.2.1 Combining time-delay and kinematics likelihoods

The posterior samples from nested sampling carry weights proportional to their contribution to the posterior mass. We obtain 10 000 equally weighted posterior samples through weighted random sampling from the nested sampling chain for each lens model setup. To combine distance posteriors from different lens model setups, we randomly sample a number of points from the lens model posterior for each setup within a mass-model family, where the sampled number is proportional to the relative weight computed from the adjusted evidence ratio (Table 4). We then uniformly sample 4000 points of  $(D_{\Delta t}^{\text{eff}}, D_s/D_{\text{ds}})$  for each lens model sample from  $[0, 2.15]D_{\Delta t}^{\text{eff, fiducial}} \times [0.35, 1.35](D_s/D_{\text{ds}})^{\text{fiducial}}$ . The chosen boundaries fully contain ( $>5\sigma$ ) the distance posteriors and they also encompass the range allowed by the priors  $H_0 \in [0, 150] \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $\Omega_m \in [0.05, 0.5]$ , given our fiducial cosmology. This procedure effectively gives us  $4000 \times N_{\text{sample}}$  points from the joint space combining the lens model parameters and  $(D_{\Delta t}^{\text{eff}}, D_s/D_{\text{ds}})$ , where  $N_{\text{sample}}$  is the number of lens model samples. We then importance sample from these  $4000 \times N_{\text{sample}}$  points weighted by the joint time-delay and kinematic likelihood to obtain the marginalized posterior distribution of  $(D_{\Delta t}^{\text{eff}}, D_s/D_{\text{ds}})$ . We only consider  $\Delta t_{\text{AB}}$  and  $\Delta t_{\text{AD}}$  in the time-delay likelihood as  $\Delta t_{\text{BD}}$  is not independent of the others. We then transform the  $(D_{\Delta t}^{\text{eff}}, D_s/D_{\text{ds}})$  distribution into the  $(D_{\Delta t}^{\text{eff}}, D_d)$  distribution.

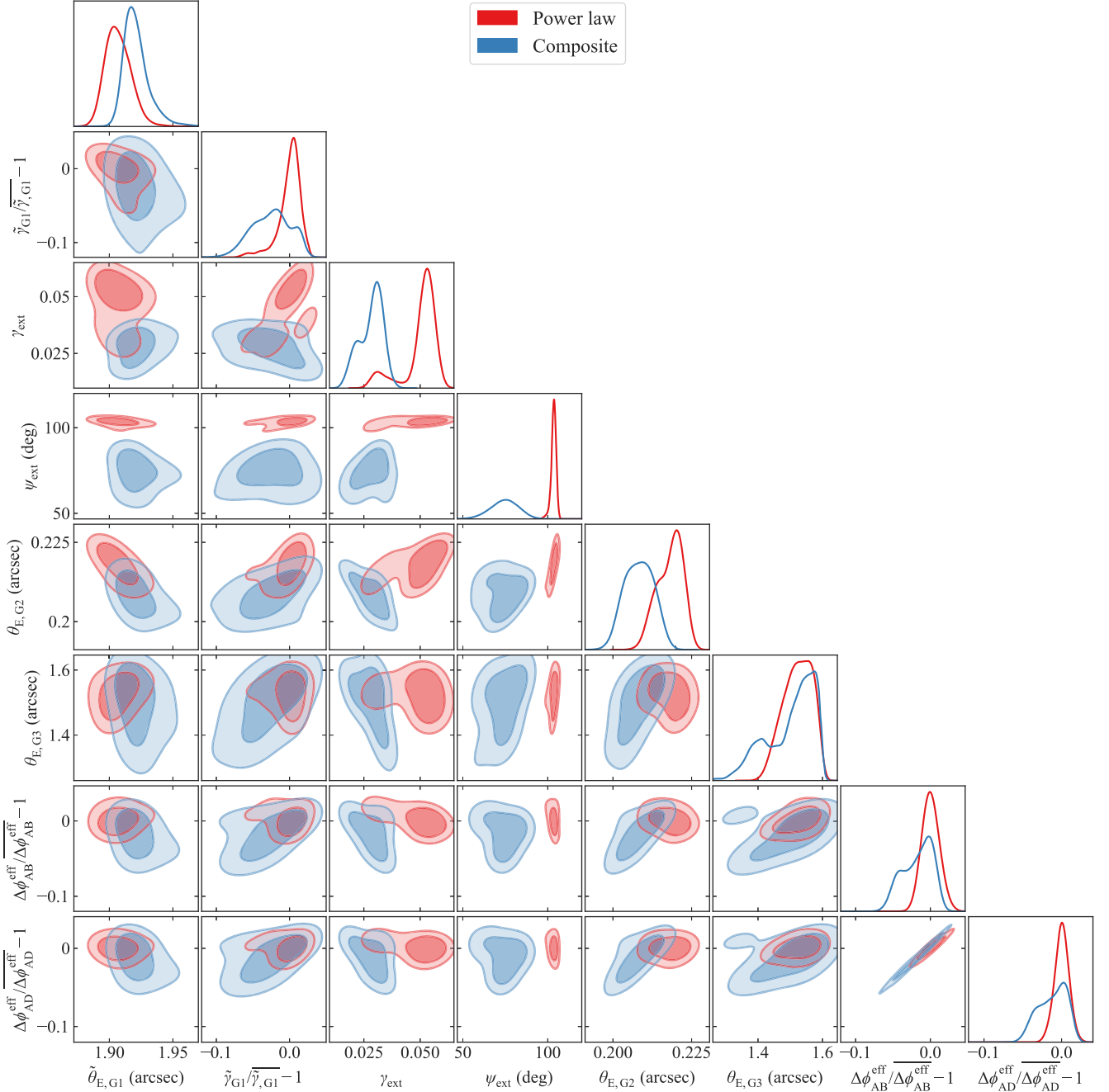
Since there are four observing setup for G1’s central line-of-sight velocity dispersion, we compute four line-of-sight velocity dispersions for each sample from the lens model posterior. We account for covariance between these four measurements in the kinematic likelihood (see Section 3.4 for covariance matrix definition). We choose a uniform prior for the anisotropy scale radius as  $r_{\text{ani}} \sim \mathcal{U}(0.5\theta_{\text{hl}}, 5\theta_{\text{hl}})$ , where  $\theta_{\text{hl}}$  is the half-light radius in the F160W band. As we model the deflector light distribution using a double Sérsic profile, we numerically compute the radius of the circular aperture that contains half of the total flux from the double Sérsic profile.

### 6.2.2 Adding the external convergence distribution into the cosmological distance posterior

We apply a selection criterion on the  $P(\kappa_{\text{ext}})$  estimated in Buckley-Geer et al. (2020) by requiring that the selected lines of sight also correspond to the external shear values predicted by our lens models. In Fig. 7, we show the probability distribution of  $\kappa_{\text{ext}}$  for the fiducial choice of constraints explored in Buckley-Geer et al. (2020), and the two  $\kappa_{\text{ext}}$  distributions consistent with the external shear values for the power-law and composite mass profiles [see Buckley-Geer et al. (2020) for further details].

We simple-sample from the external convergence distribution corresponding to each mass profile. We correct the distance posterior using the sampled external convergence according to equation (20).

Fig. 8 shows the comparison of distance posteriors between the mass model families and between different settings within a mass model family. The distance posteriors are consistent between different model setups. From the combined distance posterior from all the models, we obtain the 1D marginalized constraints for the effective time delay distance  $D_{\Delta t}^{\text{eff}} = 3382_{-115}^{+146} \text{ Mpc}$  and the angular diameter distance  $D_d = 1711_{-280}^{+376} \text{ Mpc}$  (Fig. 9).



**Figure 6.** Comparison of the lensing properties between the power-law and composite mass models. The posteriors are weighted combinations of 12 models for each mass model family. Here,  $\hat{\theta}_E$  is the Einstein radius defined to contain mean convergence of 1, and the profile slope  $\tilde{\gamma}$  is defined as the derivative of the convergence profile at  $\hat{\theta}_E$ . We blind the profile-slope and the effective Fermat potential differences by subtracting the mean and then normalize it with the mean to show relative offsets in percentage.

### 6.2.3 Microlensing time-delay effect

We check the impact of microlensing by the deflector galaxy’s stars on the measured time-delays in our analysis (Tie & Kochanek 2018). Note, this microlensing time-delay from Tie & Kochanek (2018) is based on the lamp-post model for the AGN accretion disc (Shakura & Sunyaev 1973). This effect will not necessarily exist for other disc models.

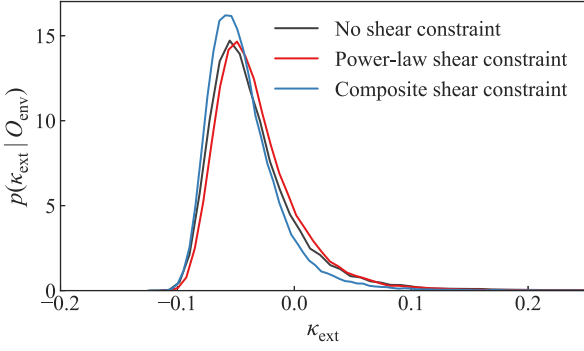
We generate the microlensing time-delay maps following Bonvin et al. (2018) and Chen et al. (2018). The estimated microlensing time-delay depends on the magnification of the lens model, on the

stellar contribution at the image position, and on the properties of the black hole’s accretion disc

We estimate the black hole mass using the scaling relation

$$\log_{10} \left( \frac{M_{\text{BH, Mg II}}}{M_{\odot}} \right) = b + m \log_{10} \left( \frac{\text{EW}_{\text{Mg II}}}{\text{\AA}} \right) + 2 \log_{10} \left( \frac{\text{FWHM}_{\text{Mg II}}}{\text{km s}^{-1}} \right) \quad (56)$$

between the black hole mass  $M_{\text{BH, vir}}$ , and the rest-frame full width at half-maximum (FWHM) and equivalent width (EW) of the Mg



**Figure 7.** Distribution of external convergence  $\kappa_{\text{ext}}$ . The black line shows the  $\kappa_{\text{ext}}$  distribution from Buckley-Geer et al. (2020) without imposing any shear constraint. The red and blue lines show the distributions with shear constraints from the power-law and the composite mass models, respectively.

II broad line. This equation is equivalent to the  $M_{\text{BH}} \propto R_{\text{BLR}} \sigma_{\text{BLR}}^2$ , where  $R_{\text{BLR}}$  is the radius of the broad-line region and  $\sigma_{\text{BLR}}$  is the velocity dispersion of the broad-line region. Here, we used the EW as a proxy for  $R_{\text{BLR}}$  and the FWHM as a proxy for  $\sigma_{\text{BLR}}$  [cf. equation (2) of Shen et al. (2011)]. We estimate the parameters of this scaling relation using the SDSS quasars from the catalogue provided by Shen et al. (2011) as  $b = 2.71$  and  $m = -0.61$ . We only take the quasars with non-zero entries for  $M_{\text{BH, vir}}$ ,  $\text{FWHM}_{\text{Mg II}}$ , and  $\text{EW}_{\text{Mg II}}$ . Moreover, we only select the quasars within  $1300 \text{ km s}^{-1} <$

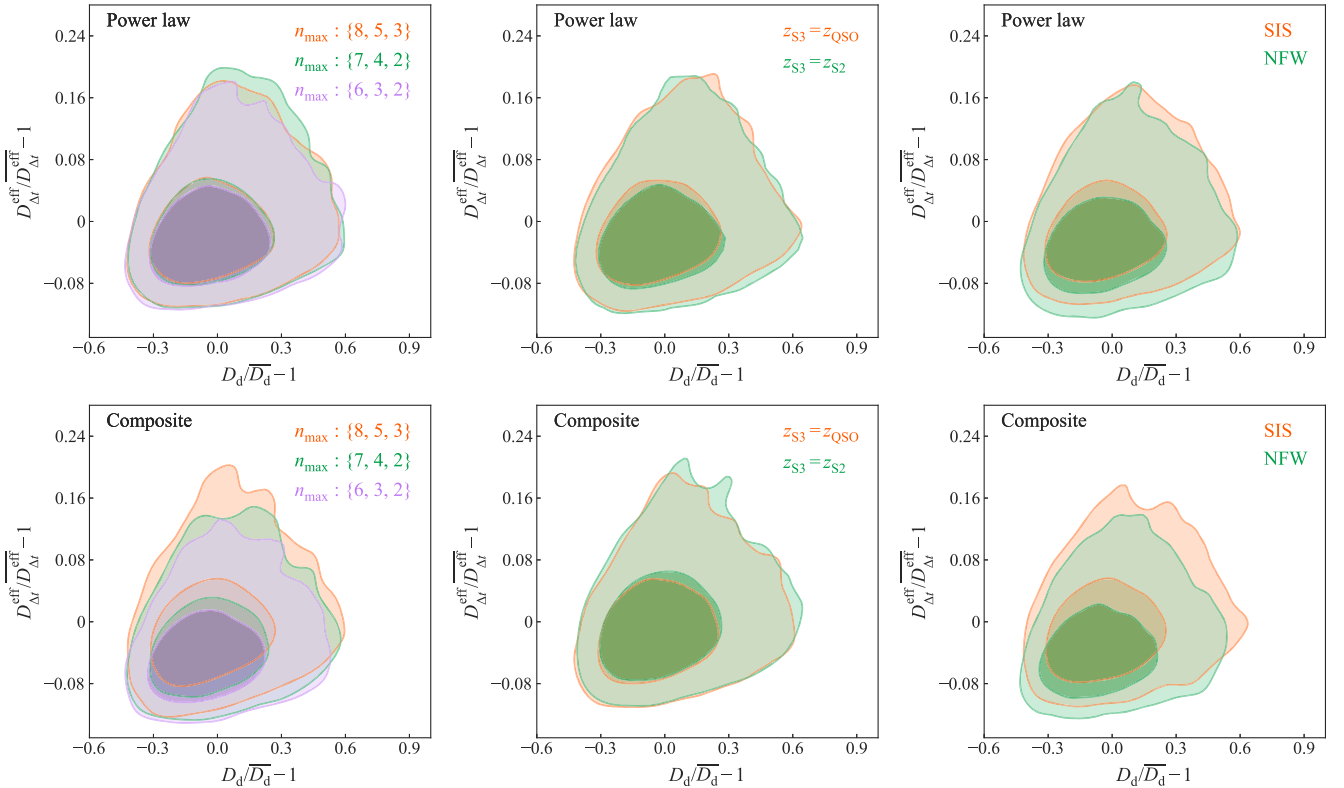
$\text{FWHM}_{\text{Mg II}} < 30\,000 \text{ km s}^{-1}$  to remove the quasars creating stripe-like features at the boundaries of the  $M_{\text{BH, vir}}\text{--FWHM}_{\text{Mg II}}$  scatter plot. As a result, we have 85 038 selected quasars to fit the above scaling relation. We obtain the rest-frame FWHM and EW of the Mg II line from the quasar spectra at image B and image D from Agnello et al. (2017) as  $\text{FWHM}_{\text{Mg II}}^{\text{B}} = 3413 \text{ km s}^{-1}$ ,  $\text{EW}_{\text{Mg II}}^{\text{D}} = 37.3 \text{ \AA}$ ,  $\text{FWHM}_{\text{Mg II}}^{\text{D}} = 2952 \text{ km s}^{-1}$ ,  $\text{EW}_{\text{Mg II}}^{\text{B}} = 30.5 \text{ \AA}$ . We apply a magnification correction to the estimated black hole mass from each image as

$$\log_{10} \left( \frac{M_{\text{BH}}}{M_{\odot}} \right) = \log_{10} \left( \frac{M_{\text{BH, Mg II}}}{M_{\odot}} \right) - g \log_{10} \mu, \quad (57)$$

where we take the calibration factor  $g = 0.5$  (Vestergaard & Peterson 2006). We also add 0.25 dex uncertainty to the estimated black hole mass to account for the limitation of using Mg II to measure it (Woo et al. 2018). Averaging over the estimates from image B and D, we obtain the black hole mass of the quasar as  $\log_{10} (M_{\text{BH, vir}}^{10408}/M_{\odot}) = 8.41 \pm 0.27$ . We also estimate the Eddington ratio using the scaling relation

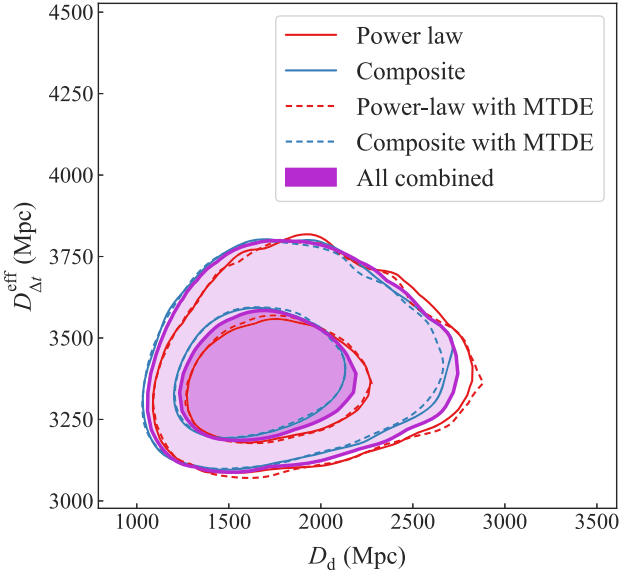
$$\log_{10} \left( \frac{L'_{\text{bol}}}{L'_{\text{Edd}}} \right) = b + m \log_{10} \left( \frac{M_{\text{BH, Mg II}}}{M_{\odot}} \right). \quad (58)$$

We estimate  $m = -0.33$ ,  $b = 2.2$  with an intrinsic scatter of 0.64 dex by fitting the relation to the same objects selected from Shen et al. (2011)’s catalogue. We also apply a magnification



**Figure 8.** Comparison of the distance posteriors between choices of the lens model setups for the power-law models (top row) and the composite models (bottom row). The distance posteriors are weighted combinations of different runs with one common model setting as specified. One of the posteriors mean is subtracted from all them and then they are normalized by the mean to get the relative shifts in percentage. Left-hand panel: the distance posteriors for different settings of the source components’  $n_{\text{max}}$ . Centre: the distance posteriors for different redshifts of S3. Right-hand panel: the distance posteriors for SIS and NFW mass profiles for the line-of-sight galaxies. The distance posteriors from different models are consistent.





**Figure 9.** Cosmological distance posteriors for power-law (red) and composite mass-profile families (blue), and for all models combined (purple). All the models are combined within each categories weighted by their adjusted evidence ratios (Table 4). The solid contours for the power-law and composite mass profiles are without accounting for the microlensing time-delay effect (MTDE), and the corresponding dashed contours show the ones with the MTDE. The MTDE is negligible in our analysis and we do not incorporate this effect in the final distance posterior (purple contours) combining all the models.

**Table 5.** Lensing quantities at the image positions used to create the microlensing time-delay map. The total convergence  $\kappa$ , the stellar convergence  $\kappa_*$ , the shear  $\gamma_{\text{shear}}$ , and the magnification  $\mu$  are obtained from the best-fitting composite model. As the local slope of the composite model at the image positions can deviate from the 3D slope  $\gamma = 2$  of an isothermal profile,  $\kappa$  and  $\gamma_{\text{shear}}$  are not necessarily identical.

TImage	$\kappa$	$\kappa_*$	$\gamma_{\text{shear}}$	$\mu$
A	0.46	0.03	0.19	3.9
B	0.59	0.06	0.32	15.5
D	0.70	0.13	0.69	-2.6

correction on the Eddington ratio obtained for each image as

$$\log_{10} \left( \frac{L_{\text{bol}}}{L_{\text{Edd}}} \right) = \log_{10} \left( \frac{L'_{\text{bol}}}{L'_{\text{Edd}}} \right) + (g - 1) \log_{10} \mu \quad (59)$$

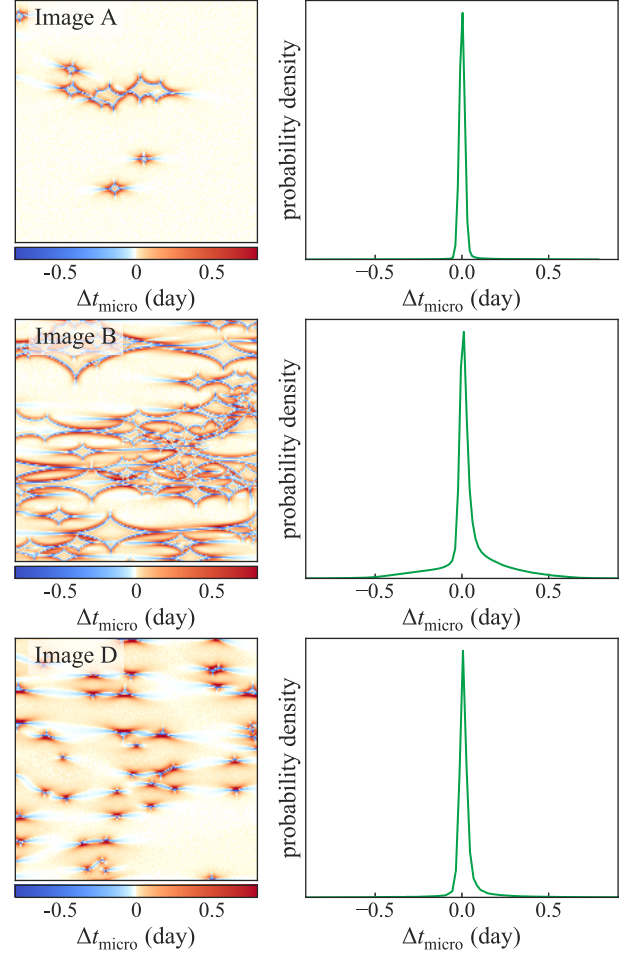
(Birrer et al. 2019). As a result, we obtain  $\log_{10}(L_{\text{bol}}/L_{\text{Edd}}) = -1.48 \pm 0.27$  after averaging over the estimates from images B and D. The accretion disc size  $R_0$  is determined assuming a standard accretion disc model (Shakura & Sunyaev 1973). In Tables 5 and 6, we tabulate the values used to create the microlensing time delay maps shown in Fig. 10. We assumed Salpeter initial mass function (IMF) and ignored the uncertainty on the convergence and shear parameters. Shifting the stellar convergence  $\kappa_*$  by the typical uncertainty of 10 per cent or changing the IMF has negligible impact on the estimated microlensing time-delay distribution in Fig. 10.

We account for the microlensing time-delay effect in the measured time delay by sampling from the microlensing time-delay distribution and adjusting the measured time delay as

$$\Delta t_{XY, \text{adjusted}} = \Delta t_{XY, \text{measured}} + t_{X, \text{micro}} - t_{Y, \text{micro}}. \quad (60)$$

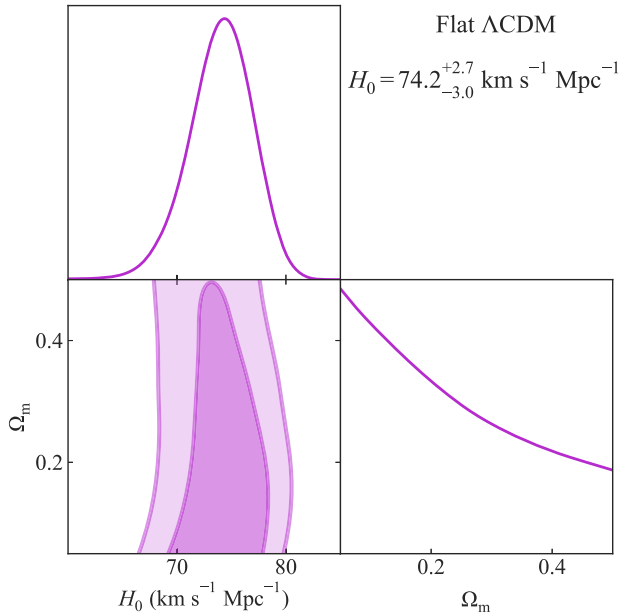
**Table 6.** Properties of the quasar accretion disc used to compute the microlensing time-delay maps.

Quantity	Value
Black hole mass, $\log_{10}(M_{\text{BH}}/M_{\odot})$	$8.41 \pm 0.27$
Eddington ratio, $\log_{10}(L_{\text{bol}}/L_{\text{Edd}})$	$-1.48 \pm 0.27$
Accretion disc size, $R_0$ (cm)	$3.125 \times 10^{14}$
Accretion efficiency, $\eta$	0.1
Central wavelength for light curve observation, $\lambda$ ( $\mu\text{m}$ )	0.668
Average foreground stellar mass, $\langle M_{\star}/M_{\odot} \rangle$	0.3



**Figure 10.** Microlensing time-delay maps and the probability of microlensing time-delays for images A, B, and D. Each microlensing time-delay map on the left-hand column is created from the relevant image magnification, stellar contribution to the convergence at the image position, and the accretion disc properties of the black hole. The probability density function of the microlensing time-delay at each image position is shown in the right-hand column. The expected fluctuation in the measured time-delays is small relative to the measurement uncertainty.

The microlensing time-delay effects is small compared to the uncertainty on the measured time delays. Thus, accounting for this microlensing time-delay effect does not shift the effective time-delay distance by more than 0.1 per cent (Fig. 9). We only perform this step as a check and we do not include this effect in our inference of  $H_0$ .



**Figure 11.** Posterior probability distribution functions of  $H_0$  and  $\Omega_m$  in the  $\Lambda$ CDM cosmology for the 2D distance posterior in Fig. 9. The inferred Hubble constant is  $H_0 = 74.2^{+2.7}_{-3.0} \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

### 6.3 Inference of $H_0$

The cosmological distance posterior contains all the cosmographic information. We infer  $H_0$  from this distance posterior for a flat  $\Lambda$ CDM cosmology with priors  $H_0 \in [0, 150] \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $\Omega_m \in [0.05, 0.5]$ . We take these priors for consistency with previous H0LiCOW analyses (Birrer et al. 2019; Chen et al. 2019; Rusu et al. 2020). The  $\Omega_m$  prior is based on our knowledge from various observations that the Universe is neither empty nor closed. We take a kernel density estimate of the distance posterior as the likelihood function for cosmological parameters to retain the full covariance between  $D_{\Delta t}^{\text{eff}}$  and  $D_d$ . Similar to Birrer et al. (2019), we take the bandwidth for the kernel density estimation to be sufficiently narrow so as to not affect the resultant posteriors of the cosmological parameters. We infer  $H_0 = 74.2^{+2.7}_{-3.0} \text{ km s}^{-1} \text{ Mpc}^{-1}$  in the  $\Lambda$ CDM cosmology, which is a 3.9 per cent measurement (Fig. 11). In this 3.9 per cent uncertainty, we estimate that the time-delay measurement contributes 1.8 per cent, the external convergence contributes 3.3 per cent, and the lens modelling and other sources contribute the remaining 1 per cent uncertainty. As a systematic check on our model weighting scheme, we infer  $H_0$  by combining all the models with equal weight as done in the first few analyses of the H0LiCOW lenses (e.g. Wong et al. 2017). For this most conservative weighting scheme, we find  $H_0 = 74.8^{+2.7}_{-3.0} \text{ km s}^{-1} \text{ Mpc}^{-1}$ , which is a 0.8 per cent deviation from our quoted  $H_0$  above. We summarize the various systematic checks performed in this paper and their corresponding impacts on the inferred  $H_0$  in Appendix F.

## 7 DISCUSSION AND SUMMARY

In this paper, we analyse the lens system DES J0408–5354 to blindly infer the effective time-delay distance from the observed time delays. We model the mass profile of the lens system using high-resolution *HST* imaging from three bands. We combine the time-delay and kinematic likelihoods with the lens model posterior,

and factor in the statistically inferred external convergence to obtain the cosmological distance posteriors in the  $D_{\Delta t}^{\text{eff}}-D_d$  plane. We perform a thorough check for systematic effects arising from model choices, and we marginalize over them to account for this source of systematic uncertainties in our analysis. As a result, we constrain the 2D joint posterior of the effective time delay distance  $D_{\Delta t}^{\text{eff}}$  and the angular diameter distance to the deflector  $D_d$  that fully incorporates their covariance. The marginalized estimates for these two distances are  $D_{\Delta t}^{\text{eff}} = 3382^{+146}_{-115}$  with 3.9 per cent uncertainty, and  $D_d = 1711^{+376}_{-280}$  with 19.2 per cent uncertainty. These constraints translate into  $H_0 = 74.2^{+2.7}_{-3.0} \text{ km s}^{-1} \text{ Mpc}^{-1}$  with a precision of 3.9 per cent. This estimated value of  $H_0$  is consistent with from the previously analysed sample of six lenses by the H0LiCOW collaboration,  $H_0 = 73.3^{+1.7}_{-1.8} \text{ km s}^{-1} \text{ Mpc}^{-1}$  (Wong et al. 2020). It is also consistent with measurements of  $H_0$  based on the local distance ladder (Freedman et al. 2019; Riess et al. 2019), reinforcing the tension (Verde et al. 2019) with the inference from early Universe probes (Abbott et al. 2018; Planck Collaboration VI 2018).

The one presented in this paper is the first of two independent cosmographic analysis of the lens system DES J0408–5354, which is based on the lens-modelling software LENSTRONOMY. A second independent and blind analysis using the lens-modelling software GLEE will be presented in a future work (Yıldırım et al., in preparation). In this future paper, we will compare the results from the two modelling efforts and quantitatively evaluate the systematic uncertainty that may arise due to using different softwares and adopting different modelling choices by different investigators. The posterior probability distribution function of  $H_0$  from DES J0408–5354 will be combined with previous measurements by the H0LiCOW team after the second analysis is complete, so as to include this modelling systematic uncertainty in the combination.

The analysis presented in this paper can be improved in the future. Due to a multitude complexities required in the lens model of DES J0408–5354, we fix the distance ratios between the multiple lens and source planes in our analysis using a fiducial  $\Lambda$ CDM cosmology to make our analysis computationally feasible. We show that the choice of fiducial cosmology has negligible impact on our inference within the  $\Lambda$ CDM model. However, it would be ideal to treat the distance ratios as independent non-linear parameters in the model. We leave this improvement to be implemented in future works, where more general cosmological models will be considered. Furthermore, the precision on the inferred  $H_0$  can be improved in the future with the help of spatially resolved kinematics (Shajib et al. 2018; Yıldırım, Suyu & Halkola 2019).

In the modelling, more general mass profiles can be used in the composite model – e.g. the generalized NFW profile, or a mass-to-light ratio gradient in the stellar component (e.g. Zhao 1996; Sonnenfeld et al. 2018). However, these types of generalized mass profile had been computationally intractable for lens modelling in the elliptical case until only recently (Shajib 2019). Given the computational cost of this study ( $\sim 10^6$  CPU h) already pushing far beyond the typical case of modelling endeavours for a single lens system, we leave the explorations of more general mass models and estimation of the corresponding impact in the inference of  $H_0$  for future studies.

Improving the precision on  $H_0$  measurement from each single lens system, increasing the number of systems to  $\sim 40$ , and investigating the presence of yet unknown systematic errors are all necessary steps towards reaching 1 per cent precision from time-delay cosmography (Shajib et al. 2018). The analysis presented in this paper took one step in each direction.

**ACKNOWLEDGEMENTS**

We thank the anonymous referee for helpful comments that improved this manuscript. AJS, SB, TT, GCFC, and CDF were supported by the National Aeronautics and Space Administration (NASA) through the Space Telescope Science Institute (STScI) grant *HST-GO-15320*. AJS was also supported by the Dissertation Year Fellowship from the University of California, Los Angeles (UCLA) graduate division. TT acknowledges support by the Packard Foundation through a Packard Research fellowship and by the National Science Foundation through NSF grants AST-1714953 and AST-1906976. This project is partly funded by the Danish council for independent research under the project ‘Fundamentals of Dark Matter Structures’, DFF - 6108-00470. AA was supported by a grant from VILLUM FONDEN (project number 16599). JHHC, MM, CL, DS, FC, and AG acknowledge support from the Swiss National Science Foundation (SNSF) and by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (COSMICLENS: grant agreement No 787866). CS has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie actions grant agreement No 664931. GCFC acknowledges support from the Ministry of Education in Taiwan via Government Scholarship to Study Abroad (GSSA). CDF and GCFC acknowledge support for this work from the National Science Foundation under Grant No AST-1715611. TA acknowledges support from Proyecto FONDECYT N° 1190335. KCW was supported by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan.

This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Institute for Digital Research and Education’s Research Technology Group. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) through allocation TG-AST190038, which is supported by National Science Foundation grant number ACI-1548562 (Towns et al. 2014). Specifically, this work used the Comet and Oasis system at the San Diego Supercomputer Center (SDSC), and the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC, Nystrom et al. 2015). AJS thanks Smadar Naoz for providing access to additional computing nodes on the Hoffman2 Shared Cluster and Khalid Jawed for assisting with additional computational resources from the Structures–Computer Interaction Laboratory at UCLA.

AJS acknowledges the hospitality of the Aspen Center of Physics (ACP) and the Munich Institute for Astro- and Particle Physics (MIAPP) of the Excellence Cluster ‘Universe’, where part of this research was completed. ACP is supported by National Science Foundation grant PHY-1607611.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro,

Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020, and the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) e-Universe (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

This research made use of LENSTRONOMY (Birrer et al. 2015; Birrer & Amara 2018), DYPOLYCHORD (Higson et al. 2019), POLYCHORD (Handley et al. 2015), COSMOHAMMER (Akeret et al. 2013), FASTELL (Barkana 1999), NUMPY (Oliphant 2015), SCIPY (Jones et al. 2001), ASTROPY (Astropy Collaboration 2013, 2018), JUPYTER (Kluyver et al. 2016), MATPLOTLIB (Hunter 2007), SEABORN (Waskom et al. 2014), NESTCHECK (Higson 2018), SETRACTOR (Bertin & Arnouts 1996), EMCÉE (Foreman-Mackey et al. 2013), COLOSSUS (Diemer 2018), and GETDIST (<https://github.com/cmban/getdist>).



REFERENCES

Abbott T. M. C. et al., 2018, *Phys. Rev. D*, 98, 043526  
 Agnello A., Evans N. W., Romanowsky A. J., 2014, *MNRAS*, 442, 3284  
 Agnello A., Kelly B. C., Treu T., Marshall P. J., 2015a, *MNRAS*, 448, 1446  
 Agnello A. et al., 2015b, *MNRAS*, 454, 1260  
 Agnello A. et al., 2017, *MNRAS*, 472, 4038  
 Agnello A. et al., 2018a, *MNRAS*, 475, 2086  
 Agnello A. et al., 2018b, *MNRAS*, 479, 4345  
 Akeret J., Seehars S., Amara A., Refregier A., Csillaghy A., 2013, *Astron. Comput.*, 2, 27  
 Alam S. et al., 2017, *MNRAS*, 470, 2617  
 Anguita T. et al., 2018, *MNRAS*, 480, 5017  
 Astropy Collaboration, 2013, *A&A*, 558, A33  
 Astropy Collaboration, 2018, *AJ*, 156, 123  
 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  
 Avila R. J., Hack W., Cara M., Borncamp D., Mack J., Smith L., Ubeda L., 2015, in Taylor A. R., Rosolowsky E., eds, ASP Conf. Ser. Vol. 495. Astronomical Data Analysis Software and Systems XXIV (ADASS XXIV). Astron. Soc. Pac., San Francisco, p. 281  
 Barkana R., 1998, *ApJ*, 502, 531  
 Barkana R., 1999, FASTELL: Fast calculation of a family of elliptical mass gravitational lens models, Astrophysics Source Code Library, record ascl:9910.003  
 Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, 488, 3143  
 Bendinelli O., 1991, *ApJ*, 366, 599  
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393  
 Bertin G., Lombardi M., 2006, *ApJ*, 648, L17  
 Birrer S., Amara A., 2018, *Phys. Dark Universe*, 22, 189  
 Birrer S., Amara A., Refregier A., 2015, *ApJ*, 813, 102  
 Birrer S., Amara A., Refregier A., 2016, *J. Cosmol. Astropart. Phys.*, 8, 020  
 Birrer S., Treu T., 2019, *MNRAS*, 489, 2097  
 Birrer S. et al., 2019, *MNRAS*, 484, 4726  
 Blandford R. D., Narayan R., 1992, *ARA&A*, 30, 311  
 Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Moustakas L. A., 2006, *ApJ*, 638, 703  
 Bonvin V. et al., 2017, *MNRAS*, 465, 4914  
 Bonvin V. et al., 2018, *A&A*, 616, A183  
 Buckley-Geer E. J. et al., 2020, preprint (arXiv:2003.12117)  
 Cappellari M., 2002, *MNRAS*, 333, 400  
 Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245  
 Chen G. C.-F. et al., 2016, *MNRAS*, 462, 3457  
 Chen G. C.-F. et al., 2018, *MNRAS*, 481, 1115  
 Chen G. C.-F. et al., 2019, *MNRAS*, 490, 1743  
 Collett T. E., Auger M. W., 2014, *MNRAS*, 443, 969  
 Collett T. E., Auger M. W., Belokurov V., Marshall P. J., Hall A. C., 2012, *MNRAS*, 424, 2864  
 Courbin F. et al., 2018, *A&A*, 609, A71  
 Diehl H. T. et al., 2017, *ApJS*, 232, 15  
 Diemer B., 2018, *ApJS*, 239, 35  
 Diemer B., Joyce M., 2019, *ApJ*, 871, 168  
 Dutton A. A. et al., 2011, *MNRAS*, 417, 1621  
 Emsellem E., Monnet G., Bacon R., Nieto J.-L., 1994, *A&A*, 285, 739  
 Falco E. E., Gorenstein M. V., Shapiro I. I., 1985, *ApJ*, 289, L1  
 Fassnacht C. D., Xanthopoulos E., Koopmans L. V. E., Rusin D., 2002, *ApJ*, 581, 823  
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
 Freedman W. L. et al., 2019, *ApJ*, 882, 34  
 Gavazzi R., Treu T., Koopmans L. V. E., Bolton A. S., Moustakas L. A., Burles S., Marshall P. J., 2008, *ApJ*, 677, 1046  
 Gavazzi R., Treu T., Rhodes J. D., Koopmans L. V. E., Bolton A. S., Burles S., Massey R. J., Moustakas L. A., 2007, *ApJ*, 667, 176  
 Goobar A. et al., 2017, *Science*, 356, 291  
 Greene Z. S. et al., 2013, *ApJ*, 768, 39  
 Grillo C. et al., 2018, *ApJ*, 860, 94  
 Guérou A. et al., 2017, *A&A*, 608, A5  
 Handley W. J., Hobson M. P., Lasenby A. N., 2015, *MNRAS*, 453, 4384  
 Higson E., 2018, *J. Open Source Softw.*, 3, 916  
 Higson E., Handley W., Hobson M., Lasenby A., 2018, *Stat. Comput.*, 29, 891  
 Higson E., Handley W., Hobson M., Lasenby A., 2019, dyPolyChord: Super fast dynamic nested sampling with PolyChord, Astrophysics Source Code Library, record ascl:1902.010  
 Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, *A&A*, 499, 31  
 Hoeting J. A., Madigan D., Raftery A. E., Volinsky C. T., 1999, *Stat. Sci.*, 14, 382  
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90  
 Jee I., Komatsu E., Suyu S. H., 2015, *J. Cosmol. Astropart. Phys.*, 2015, 033  
 Jee I., Komatsu E., Suyu S. H., Huterer D., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 031  
 Jones E., Oliphant T., Peterson P., Others, 2001, SciPy: Open source scientific tools for Python. Available at: <http://www.scipy.org/>  
 Jorgensen I., Franx M., Kjaergaard P., 1995, *MNRAS*, 276, 1341  
 Kelly P. L. et al., 2015, *Science*, 347, 1123  
 Kennedy J., Eberhart R., 1995, Proceedings of ICNN'95 - International Conference on Neural Networks. IEEE, Piscataway, NJ  
 Kluyver T. et al., 2016, in Loizides F., Schmidt B., eds, Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press BV, Amsterdam, Netherlands, p. 87  
 Lagattuta D. J., Vegetti S., Fassnacht C. D., Auger M. W., Koopmans L. V. E., McKean J. P., 2012, *MNRAS*, 424, 2800  
 Lemon C. et al., 2020, *MNRAS*, preprint (arXiv:1912.09133)  
 Lewis A., Bridle S., 2002, *Phys. Rev. D*, 66, 103511  
 Lin H. et al., 2017, *ApJ*, 838, L15  
 Madigan D., Raftery A. E., 1994, *J. Am. Stat. Assoc.*, 89, 1535  
 Mamon G. A., Łokas E. L., 2005, *MNRAS*, 363, 705  
 McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2017, *ApJ*, 836, 141  
 Merritt D., 1985a, *AJ*, 90, 1027  
 Merritt D., 1985b, *MNRAS*, 214, 25P  
 Munari E., Biviano A., Borgani S., Murante G., Fabjan D., 2013, *MNRAS*, 430, 2638  
 Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493  
 Nord B. et al., 2016, *ApJ*, 827, 51  
 Nystrom N. A., Levine M. J., Roskies R. Z., Scott J. R., 2015, in Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure. XSEDE '15. ACM, New York, NY, USA, p. 30:1  
 Oliphant T. E., 2015, Guide to NumPy, 2nd edn. CreateSpace Independent Publishing Platform, USA  
 Osipkov L. P., 1979, *Pisma Astron. Z.*, 5, 77  
 Ostrovski F. et al., 2017, *MNRAS*, 465, 4325  
 Paraficz D., Hjorth J., 2009, *A&A*, 507, L49  
 Perlmutter S. et al., 1999, *ApJ*, 517, 565  
 Planck Collaboration VI, 2018, preprint (arXiv:1807.06209)  
 Refregier A., 2003, *MNRAS*, 338, 35  
 Refsdal S., 1964, *MNRAS*, 128, 307  
 Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, *ApJ*, 876, 85  
 Riess A. G. et al., 1998, *AJ*, 116, 1009  
 Rusu C. E. et al., 2017, *MNRAS*, 467, 4220  
 Rusu C. E. et al., 2020, *MNRAS*, preprint (arXiv:1905.09338)  
 Schechter P. L. et al., 1997, *ApJ*, 475, L85  
 Schneider P., Ehlers J., Falco E. E., 1992, *Gravitational Lenses*. Springer, Berlin  
 Schneider P., Sluse D., 2014, *A&A*, 564, A103  
 Shajib A. J., 2019, *MNRAS*, 488, 1387  
 Shajib A. J., Treu T., Agnello A., 2018, *MNRAS*, 473, 210  
 Shajib A. J. et al., 2019, *MNRAS*, 483, 5649  
 Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337  
 Shen Y. et al., 2011, *ApJS*, 194, 45  
 Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, AIP Conf. Proc. Vol. 735, Nested Sampling. Am. Inst. Phys., New York, p. 395

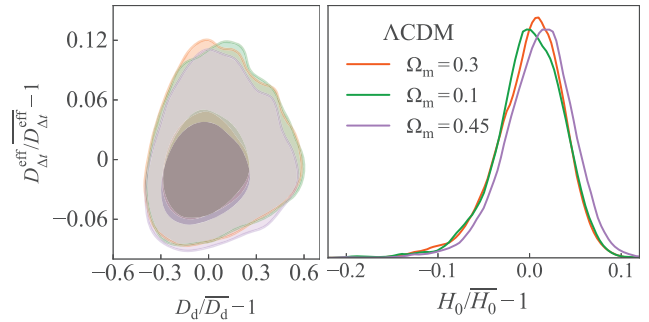


- Sluse D., Hutsemékers D., Courbin F., Meylan G., Wambsganss J., 2012, *A&A*, 544, A62
- Sluse D. et al., 2019, *MNRAS*, 490, 613
- Sonnenfeld A., Leauthaud A., Auger M. W., Gavazzi R., Treu T., More S., Komiyama Y., 2018, *MNRAS*, 481, 164
- Sonnenfeld A., Treu T., Gavazzi R., Marshall P. J., Auger M. W., Suyu S. H., Koopmans L. V. E., Bolton A. S., 2012, *ApJ*, 752, 163
- Soto K. T., Lilly S. J., Bacon R., Richard J., Conseil S., 2016, *MNRAS*, 458, 3210
- Spiniello C. et al., 2018, *MNRAS*, 480, 1163
- Springel V. et al., 2005, *Nature*, 435, 629
- Suyu S. H., Halkola A., 2010, *A&A*, 524, A94
- Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, *ApJ*, 711, 201
- Suyu S. H. et al., 2013, *ApJ*, 766, 70
- Suyu S. H. et al., 2014, *ApJ*, 788, L35
- Suyu S. H. et al., 2017, *MNRAS*, 468, 2590
- Sérsic J. L., 1968, Atlas de Galaxias Australes. Available at: <http://adsabs.harvard.edu/abs/1968adga.book....S>
- Tewes M. et al., 2013, *A&A*, 556, A22
- Tie S. S., Kochanek C. S., 2018, *MNRAS*, 473, 80
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, 688, 709
- Towns J. et al., 2014, *Comput. Sci. Eng.*, 16, 62
- Treu T., Koopmans L. V., Bolton A. S., Burles S., Moustakas L. A., 2006, *ApJ*, 640, 662
- Treu T., Koopmans L. V. E., 2002, *MNRAS*, 337, L6
- Treu T., Marshall P. J., 2016, *A&AR*, 24, 11
- Treu T. et al., 2016, *ApJ*, 817, 60
- Treu T. et al., 2018, *MNRAS*, 481, 1041
- Verde L., Treu T., Riess A. G., 2019, *Nat. Astron.*, 3, 891
- Vestergaard M., Peterson B. M., 2006, *ApJ*, 641, 689
- Waskom M. et al., 2014, seaborn: v0.5.0 (November 2014)
- Wong K. C. et al., 2017, *MNRAS*, 465, 4895
- Wong K. C. et al., 2020, *MNRAS*, preprint ([arXiv:1907.04869](https://arxiv.org/abs/1907.04869))
- Woo J.-H., Le H. A. N., Karouzos M., Park D., Park D., Malkan M. A., Treu T., Bennert V. N., 2018, *ApJ*, 859, 138
- Yuan W., Riess A. G., Macri L. M., Casertano S., Scolnic D. M., 2019, *ApJ*, 886, 61
- Yıldırım A., Suyu S. H., Halkola A., 2020, *MNRAS*, 493, 4783
- Zahid H. J., Geller M. J., Fabricant D. G., Hwang H. S., 2016, *ApJ*, 832, 203
- Zhao H., 1996, *MNRAS*, 278, 488

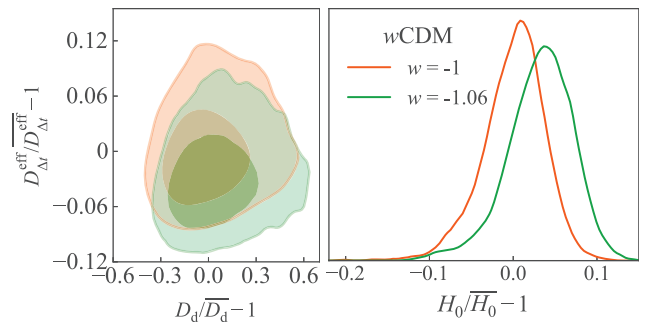
## APPENDIX A: IMPACT OF FIDUCIAL COSMOLOGY IN THE LENS MODELLING

We check the impact of fixing the distance ratios between the lens and the source planes with a fiducial  $\Lambda$ CDM cosmology with density parameters  $\Omega_m = 0.3$  and  $\Omega_\Lambda = 0.7$ . For this purpose, we run two separate lens models with the same model setup for the power-law mass profile, but with the cosmological parameters ( $\Omega_m = 0.1$ ,  $\Omega_\Lambda = 0.9$ ) and ( $\Omega_m = 0.45$ ,  $\Omega_\Lambda = 0.55$ ) to fix the distance ratios. Within this wide-range of  $\Omega_m$  values within the flat  $\Lambda$ CDM cosmology,  $H_0$  only shifts by  $\lesssim 1$  per cent (Fig. A1). As this shift is much smaller than the precision of the measured Hubble constant allowed by the quality of our data, we conclude that our inferred cosmological distance posterior on the  $D_{\Delta l}-D_d$  plane is effectively independent of the choice of cosmological parameters within the flat  $\Lambda$ CDM cosmology.

However, we find that the inferred distance posteriors depend on the fiducial cosmology within the  $w$ CDM model. If we adopt the fiducial  $w$ CDM cosmology with  $w = -1.06$ ,  $\Omega_m = 0.3$ ,  $\Omega_{de} = 0.7$ , then the inferred  $H_0$  shifts by approximately 3 per cent from that inferred from the cosmology with  $w = -1$  (Fig. A2). We adopt the

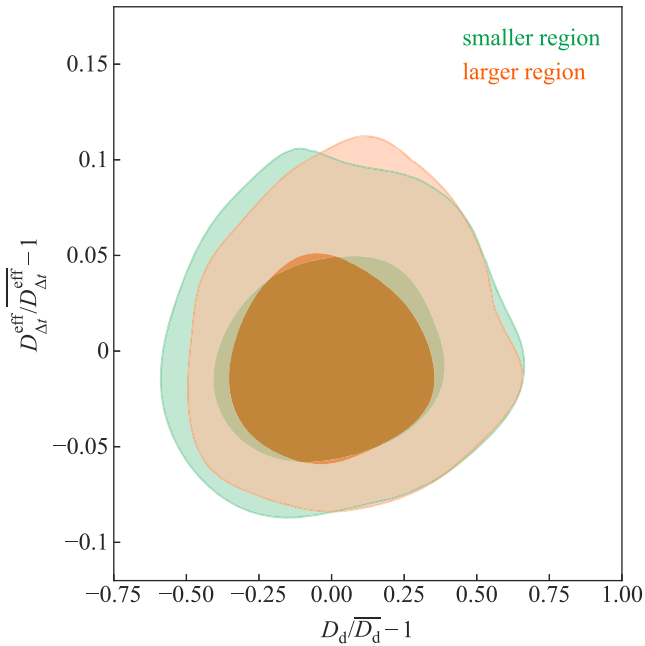


**Figure A1.** Comparison of the distance posteriors and inferred  $H_0$  for different fiducial cosmologies within the flat  $\Lambda$ CDM model. We compare between three set of cosmological parameters:  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$  (orange),  $\Omega_m = 0.1$ ,  $\Omega_\Lambda = 0.9$  (green), and  $\Omega_m = 0.45$ ,  $\Omega_\Lambda = 0.55$  (purple). The distance posteriors are from identical lens model setups with the power-law mass profile except for the fiducial cosmology.  $H_0$  shifts by less than 1 per cent within these wide range of  $\Omega_m$  values. This shift is much smaller than the precision on  $H_0$  allowed by our current data quality. As a result, we can treat the distance posteriors inferred from our analysis to be independent of cosmological assumptions within the flat  $\Lambda$ CDM cosmology.



**Figure A2.** Comparison of the distance posteriors and inferred  $H_0$  for different fiducial cosmologies within the  $w$ CDM model. We compare between three set of cosmological parameters:  $w = -1$ ,  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$  (orange), and  $w = -1.06$ ,  $\Omega_m = 0.1$ ,  $\Omega_\Lambda = 0.9$  (green). The distance posteriors are from identical lens model setups with the power-law mass profile except for the fiducial cosmology.  $H_0$  shifts by  $\sim 3$  per cent for a shift  $\Delta w = 0.06$ , which is approximately the joint precision on  $w$  from the *Planck* with CMB lensing, SNIae, and the baryon acoustic oscillation measurements (Planck Collaboration VI 2018). This shift in  $H_0$  shows that the double source plane treatment in our analysis is sensitive to the dark energy equation of state parameter  $w$  (Gavazzi et al. 2008; Collett & Auger 2014). As a result, our distance posterior should not be used to constrain parameters in cosmologies that extend the flat  $\Lambda$ CDM model.

shift  $\Delta w = 0.06$  for comparison, as this range is in  $w$  is the joint precision from the *Planck* with CMB lensing, SNIae, and baryon acoustic oscillation measurements (Planck Collaboration VI 2018). This significant shift in  $H_0$  demonstrate that our double source plane treatment is sensitive to the dark energy equation of state parameter  $w$  (e.g. Gavazzi et al. 2008; Collett et al. 2012; Collett & Auger 2014). Therefore, our distance posteriors should not be used to constrain cosmological parameters in extended cosmologies other than the flat  $\Lambda$ CDM model. We leave the computation of a posterior distribution function valid in more general cosmologies for future work.



**Figure B1.** Comparison of the distance posteriors from the power-law lens model for two likelihood computation region sizes. The smaller regions have radii 4.3, 3.3, and 3.3 arcsec in the F160W, F814W, F475X bands, respectively. The larger region sizes are 4.5, 3.5, and 3.5 arcsec in the same order. All the other model setups are same between the two runs. The distance posteriors are almost identical. Therefore, the choice of likelihood computation region has negligible impact in our analysis.

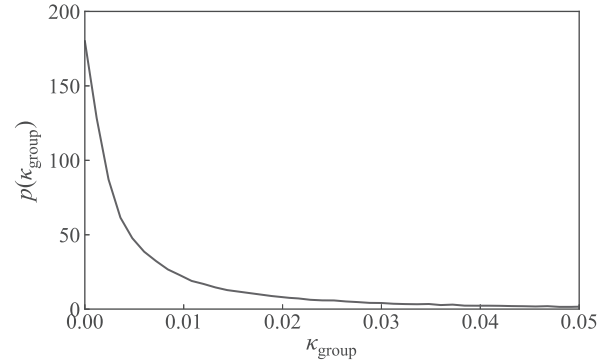
### APPENDIX B: IMPACT OF LIKELIHOOD COMPUTATION REGION CHOICE

We check if our adopted region for imaging likelihood computation can be a source of systematic bias in the lens modelling. We perform the modelling procedure for two different region sizes keeping every other settings in the model the same for a power-law model. The regular region sizes are 4.3, 3.3, and 3.3 arcsec in radius for the F160W, F814W, and F475X bands, respectively. The larger region sizes are larger by 0.2 arcsec in each band. The median of the effective time-delay distance shifts by less than 0.1 per cent and the median of the angular diameter distance shifts by less than 2 per cent between these two choices of the likelihood computation region (Fig. B1).

### APPENDIX C: IMPACT OF THE CONVERGENCE FROM THE GROUP CONTAINING G1

We estimate the convergence at G1’s centre from the galaxy group containing G1 [group ID 5 in Buckley-Geer et al. (2020)]. We randomly sample haloes from the centroid and velocity dispersion distributions estimated as described in Buckley-Geer et al. (2020). However, we use a uniform prior to obtain the velocity dispersion for the halo, whereas Buckley-Geer et al. (2020) adopt the Jeffrey’s prior. We convert the group’s velocity dispersion into halo mass using the scaling relation

$$\log_{10} [h(z)M_{200}] = 13.98 + 2.75 \log_{10} \left( \frac{\sigma_{\text{group}}}{500 \text{ km s}^{-1}} \right) \quad (\text{C1})$$



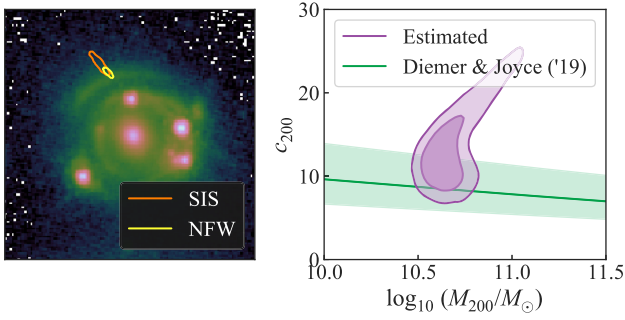
**Figure C1.** Distribution of the convergence at G1’s centre from the galaxy group’s halo containing G1. The distribution  $p(\kappa_{\text{group}})$  is estimated for the estimated centroid and velocity dispersion of this group in Buckley-Geer et al. (2020).

(Munari et al. 2013). We weight this halo mass distribution using the halo mass function from Tinker et al. (2008) corresponding to our fiducial cosmology and the lens redshift. We obtain the concentration parameter distribution using the theoretical  $M_{200}-c$  relation from Diemer & Joyce (2019) with 0.16 dex uncertainty. We also apply 10 per cent uncertainty on  $M_{200}$  to remove any strong dependency on  $H_0$  through the fiducial cosmology. We compute the convergence distribution at G1’s centre due to this distribution of the halo masses and we apply a cut in the group’s shear distribution  $\gamma_{\text{group}} < 0.1$  to remove haloes that are inconsistent with the model predicted shear (Fig. C1). The median of the group’s convergence distribution is 0.004. As we are explicitly accounting for the group’s convergence here, we re-estimate  $\kappa_{\text{ext}}$  after removing the galaxies in this group from the number count statistic of Buckley-Geer et al. (2020). The re-estimated  $\kappa_{\text{ext}}$  decreases by 0.005 for the power-law mass models and by 0.008 for the composite mass models. As a result, explicitly accounting for the group’s convergence decreases  $H_0$  by approximately 0.4 per cent. This shift is negligible compared to the 3.9 per cent uncertainty in our estimated value of  $H_0$ .

### APPENDIX D: CHECKING FOR THE EXISTENCE AND IMPACT OF A DARK SUBSTRUCTURE

Agnello et al. (2017) propose a possible dark subhalo near image D toward the North–East direction. We check the impact of such a dark substructure in our analysis by including a mass profile for the substructure in our lens model. We check with both SIS and spherical NFW profile for the substructure. We take a broad uniform prior of  $0.8 \text{ arcsec} \times 0.8 \text{ arcsec}$  for the centroid of the mass profile to encompass the possible position of the substructure given in fig. 9 of Agnello et al. (2017).

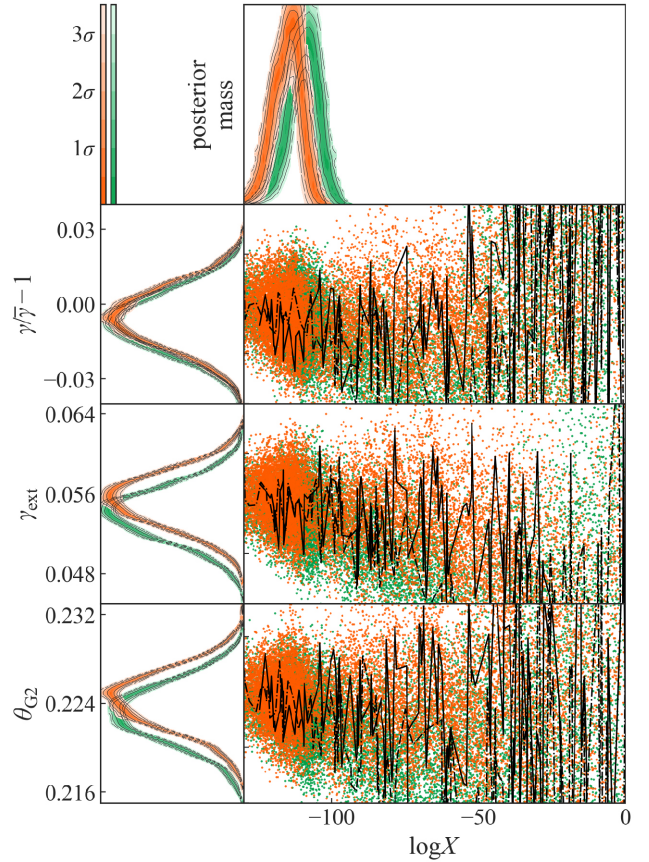
For both of the SIS and NFW profiles, our lens model constrains the possible position of the mass profile for the potential substructure (Fig. D1). Interestingly, the constrained position is consistent with the proposed position by Agnello et al. (2017), although the model had the freedom to offset the position by  $\sim 0.4 \text{ arcsec}$  from the constrained position. We estimate the SIS velocity dispersion of this possible dark substructure to be  $\sigma_{\text{SIS}} = 33.7^{+1.9}_{-1.3} \text{ km s}^{-1}$ . From the model with the NFW profile for the dark substructure, we estimate the halo mass  $\log_{10}(M_{200}/M_{\odot}) = 10.65^{+0.10}_{-0.06}$ , halo radius  $r_{200} = 45.7^{+3.5}_{-2.1} \text{ kpc}$ , and concentration  $c_{200} = 12.2^{+4.1}_{-2.2}$ .



**Figure D1.** Constraints from our lens model for the position and NFW halo properties of the dark substructure proposed by Agnello et al. (2017). Left-hand panel:  $2\sigma$  credible region for the position of the dark substructure in our lens model assuming the SIS profile (the orange contour) and the NFW profile (the yellow contour) for the substructure. Right-hand panel: The constraint on the  $M_{200}$ – $c_{200}$  plane assuming the NFW profile for the dark substructure. The purple contours show the  $1\sigma$  and  $2\sigma$  credible regions. We show a comparison with the theoretical  $c_{200}$ – $M_{200}$  relation for our fiducial cosmology from Diemer & Joyce (2019).

However, we do not add the potential substructure in our final lens model as the addition of the dark substructure shifts the estimated  $H_0$  by less than 1 per cent. Moreover, it is not clear if the constrained parameters for the additional mass profile to account for the dark substructure indeed reflect the existence of the substructure. A similar effect can also arise if the additional source component S2 lies at a redshift between the quasar and the central deflector G1. The proximity of the constrained position of the dark substructure and the lensed position of S2 hints this scenario to be a possibility.

Agnello et al. (2017) use the dust-corrected and delay-corrected flux ratios observed in the DES data as a constraint for the lens model and propose that the existence of a dark substructure fits the data better. We check if microlensing can be a possible source for the deviation of the flux ratios from a smooth model observed by Agnello et al. (2017). We derive the amplitude of microlensing in images A, B, and D by comparing their MUSE spectra. Microlensing is stronger in the continuum than in the broad emission lines. Therefore, we can isolate the microlensed fraction of the spectra if we assume that microlensing is more important in one of the lensed images under scrutiny and derive a lower limit on the amplitude of microlensing effect in the continuum emission (e.g. Sluse et al. 2012). This procedure reveals substantial differential microlensing between the continuum and the broad lines when we consider image pairs A–D and B–D, but not A–B. The data are compatible with a microlensing demagnification of image D by at least a factor of 2. This demagnification translates into a microlensing corrected flux ratio  $\Delta m_{AD} = 0.25$  mag. This estimate, however, may be affected by systematic errors caused by intrinsic variability. From the past light-curves of this system, we estimate that over the period corresponding to the time delay between images A and D, this systematic error could reach up to 0.2 mag. Therefore, we cannot definitively attribute the observed ‘flux-ratio anomaly’ to microlensing. In summary, whereas we cannot find strong evidence for the existence of the potential substructure, we also cannot rule out its existence. Since the presence or the impact of the dark substructure is not significant in our analysis, we omit it in our lens models for simplicity.



**Figure E1.** NESTCHECK diagnostic plot showing samples’ distributions in two different runs with the same lens model for our chosen sampling settings. The top right-hand panel shows the relative posterior mass as a function of  $\log X$ , where  $X$  is the prior volume fraction. The next three panels show the progressions of sampling (right to left) for three of the main lens mass parameters. The black solid and dashed lines show the evolution of one random thread of the dynamic nested sampler. The left column shows the posterior distributions of each parameter. The coloured contours represent the iso-probability credible regions of the probability density functions. The posteriors of the parameters are consistent within  $1\sigma$ , therefore we accept that the chosen settings allow for sufficient exploration of the prior volume.

## APPENDIX E: NESTED SAMPLING SETTINGS

In this Appendix, we provide our adopted settings for the nested sampling software DYPOLYCHORD and validate that the numerical requirements for our analysis are met.

We choose the DYPOLYCHORD settings `ninit = 100`, `nrepeats = 30`, `nconst_live = 140`, `dynamic_goal = 0.9`, `precision_criterion = 0.001` [see Higson et al. (2019) for explanation of these settings]. To check the appropriateness of these settings, we run two sampling runs with the same lens model and sampler settings (Fig. E1). We find that the posteriors PDFs of the parameters are consistent within  $1\sigma$  between the two runs, therefore we accept the chosen settings to be appropriate for sufficient exploration of the prior space. However, we find the estimated evidence values to differ by more than the estimated statistical uncertainty. This difference indicates that there is a systematic scatter in the computed evidence value. To estimate this scatter, we run a second set of nested sampling runs for 17 different lens models with `precision_criterion = 0.01`. We choose a lower



precision\_criterion for this second set of runs to make the sampling run terminate faster. A lower precision\_criterion does not largely impact the evidence values, although it may affect the posterior estimation (Higson et al. 2018). As we are only interested to obtain a conservative estimate of the scatter present in the computed  $\log \mathcal{Z}$  values, this lower precision\_criterion is sufficient for this purpose. By taking the mean of the evidence difference between runs from the two sets with the same lens model, we estimate the scatter in the evidence value as 24. Therefore, we take  $\sigma_{\log \mathcal{Z}}^{\text{numeric}} = 24$  as the numerical error in the computed evidence values. Albeit, if we increase nlive\_const or nrepeats, we can decrease the error in the computed  $\log \mathcal{Z}$  values in the exchange of a higher computational cost. However, as we down-weight the relative evidence ratios to account for sparse sampling of our models from the model space, this numerical error in  $\log \mathcal{Z}$  is a subdominant factor (Section 6.1).

**APPENDIX F: SUMMARY OF UNCERTAINTY BUDGET AND SYSTEMATIC CHECKS**

In this appendix, we summarize the uncertainty budgets from different parts of the cosmographic analysis (Table F1) and the systematic checks performed in our analysis with their impacts on the inferred  $H_0$  in Table F2.

**Table F1.** Uncertainty contributions from different parts of the cosmographic analysis.

Analysis component	Uncertainty budget (per cent)
Time-delay measurement	1.8
$\kappa_{\text{ext}}$ estimation	3.3
Lens modelling and other sources	1.0
<b>Total uncertainty</b>	<b>3.9</b>

**Table F2.** Summary of systematic checks and their impacts on the inferred  $H_0$ . The relevant section in the paper is referenced for each systematic check.

Systematic check	$\Delta H_0$ (per cent)
Fixing $\Omega_m = 0.10$ in fiducial cosmology (Appendix A)	−0.06
Fixing $\Omega_m = 0.45$ in fiducial cosmology (Appendix A)	+0.94
Choosing a larger likelihood computation region (Appendix B)	+0.10
Explicitly accounting for the galaxy group (Appendix C)	−0.40
Accounting for a possible dark subhalo near S3 (Appendix D)	−0.46
Not accounting for the deflection by S2’s mass (Note, we included S2’s mass in our models, thus already marginalizing in our quoted posterior; Section 5.6)	+0.95
Accounting for Tie & Kochanek (2018)’s microlensing time-delay (Section 6.2.3)	−0.10
<b>Total systematic shift</b>	<b>−0.92–0.08</b>

**APPENDIX G: MODEL SUMMARY AND PARAMETER PRIORS**

In this appendix, we summarize the adopted models (Table G1) and provide the priors for the parameters in our lens models (Table G2).

To make the nested sampler efficiently explore and integrate over the high-dimensional (~60D) prior volume in our models, we narrow down the width of the uniform priors for some of the parameters more than that would be known purely *a priori*. We choose these prior bounds by looking at the posterior PDFs for the lens models from the initial exploratory phase of this study and we further adjust these prior widths through trial and error. We check that the posterior PDFs of the parameters are fully contained within the chosen bounds for our final lens models, unless we specifically set the bound using an empirical or physical prior. Below we explain some of the parameters from Table G2 that were not introduced within the main body of this paper.

The amplitude of the Chameleon convergence profile is parametrized with the deflection angle at 1 arcsec,  $\alpha_1^{\text{Chm}}$ . The ellipticity parameters  $e_1$  and  $e_2$  in the relevant profiles are related to the axial ratio  $q$  and the position angle  $\phi$  as

$$q = \frac{1 - \sqrt{e_1^2 + e_2^2}}{1 + \sqrt{e_1^2 + e_2^2}}, \tag{G1}$$

$$\tan 2\phi = \frac{e_2}{e_1}.$$

Parametrizing the ellipticity with  $e_1$  and  $e_2$  avoids the periodicity in the polar coordinate  $\phi$  and makes the sampling more efficient. The centroids ( $\theta_1^c, \theta_2^c$ ) of the relevant profiles are relative to the coordinate RA 04:08:21.71 and Dec −53:53:59.34. We take a uniform prior  $\mathcal{U}(-5 \times 10^4, 5 \times 10^4)$  for the amplitudes of the



**Table G1.** Summary of model components. Alternative choices for some components are shown with bullet point lists.

Component	Model description
G1	(i) Elliptical power-law convergence, double elliptical Sérsic light, external shear (ii) Elliptical NFW potential for dark matter, double elliptical Chameleon for luminous matter and light, external shear
G2	SIS mass, Sérsic light
G3	SIS mass
G4, G5, G6	(i) SIS mass (ii) NFW mass
S1	Elliptical Sérsic light and shapelets on source plane
S2	Elliptical Sérsic light and shapelets on source plane
S3	Elliptical Sérsic light on source plane
Quasar images	Point sources on image plane
Cosmology	$\Lambda$ CDM, $H_0 \sim \mathcal{U}(0, 150) \text{ km s}^{-1} \text{ Mpc}^{-1}$ , $\Omega_m \sim \mathcal{U}(0.05, 0.5)$

**Table G2.** Prior for the model parameters.

Model component	Parameter	Prior
G1 mass, power law	$\gamma$	$\mathcal{U}(1.80, 2.15)$
G1 mass, composite	$r_s^{\text{NFW}}$ (arcsec)	$\mathcal{N}(12.42, 2.94)$ with bound [5.6, 19.3] (Section 5.2)
G1 mass, composite	$\alpha_1^{\text{Chm}}$ (arcsec)	$\mathcal{U}(0.2, 1.5)$
External shear	$\gamma_{\text{ext}}$	$\mathcal{U}(0.01, 0.09)$
External shear	$\psi_{\text{ext}}$ (rad)	$\mathcal{U}(0, \pi)$
G2 mass, SIS	$\theta_E$ (arcsec)	$\mathcal{U}(0.19, 0.27)$
G3 mass, SIS	$\theta_E$ (arcsec)	$\mathcal{N}(0.772, 0.024)$ with bound [0.3, 1.6] (Section 5.4.3, Fig. 3)
G4 mass, SIS	$\theta_E$ (arcsec)	$\mathcal{N}(0.353, 0.012)$ with bound [0.0, 1.0] (Section 5.4.3, Fig. 3)
G4 mass, NFW	$\log_{10}(M_{200}/M_{\odot})$	Empirical prior with bound [11.3, 13.4] (Section 5.4.4, Fig. 3)
G4 mass, NFW	$c_{200}$	Empirical prior with bound [0.0, 16.0] (Section 5.4.4, Fig. 3)
G5 mass, SIS	$\theta_E$ (arcsec)	$\mathcal{N}(0.046, 0.002)$ with bound [0.0, 0.2] (Section 5.4.3, Fig. 3)
G5 mass, NFW	$\log_{10}(M_{200}/M_{\odot})$	Empirical prior with bound [10.8, 12.3] (Section 5.4.4, Fig. 3)
G5 mass, NFW	$c_{200}$	Empirical prior with bound [0.0, 16.5] (Section 5.4.4, Fig. 3)
G6 mass, SIS	$\theta_E$ (arcsec)	$\mathcal{N}(0.070, 0.004)$ with bound [0.0, 0.3] (Section 5.4.3, Fig. 3)
G6 mass, NFW	$\log_{10}(M_{200}/M_{\odot})$	Empirical prior with bound [11.4, 12.5] (Section 5.4.4, Fig. 3)
G6 mass, NFW	$c_{200}$	Empirical prior with bound [0.0, 20.0] (Section 5.4.4, Fig. 3)
S2 mass, SIS	$\theta_E$ (arcsec)	$\mathcal{N}(0.0022, 9.98 \times 10^{-6})$ with bound [0.000, 0.022] (Section 5.4.3)
S1 light, Sérsic	$\theta_{\text{eff}}$ (arcsec)	$\mathcal{U}(0.04, 0.15)$
S1 light, Sérsic	$n_s$	$\mathcal{U}(0.6, 5.0)$
S1 light, Sérsic	$e_1$	$\mathcal{U}(-0.05, 0.35)$
S1 light, Sérsic	$e_2$	$\mathcal{U}(-0.16, 0.20)$
S1 light, F814W+F475X, shapelets	$\zeta$ (arcsec)	$\mathcal{U}(0.06, 0.11)$
S1 light, F160W, shapelets	$\zeta$ (arcsec)	$\mathcal{U}(0.08, 0.15)$
S2 light, Sérsic	$\theta_{\text{eff}}$ (arcsec)	$\mathcal{U}(0.08, 0.40)$
S2 light, Sérsic	$n_s$	$\mathcal{U}(1.0, 5.0)$
S2 light, Sérsic	$e_1$	$\mathcal{U}(0.04, 0.37)$
S2 light, Sérsic	$e_2$	$\mathcal{U}(-0.20, 0.00)$
S2 light	$\theta_1^{\zeta}$ (arcsec)	$\mathcal{U}(-0.47, 0.30)$
S2 light	$\theta_2^{\zeta}$ (arcsec)	$\mathcal{U}(-2.48, -1.48)$
S2 light, shapelets	$\zeta$ (arcsec)	$\mathcal{U}(0.06, 0.12)$
S3 light, Sérsic	$\theta_{\text{eff}}$ (arcsec)	$\mathcal{U}(0.18, 0.90)$
S3 light, Sérsic	$n_s$	$\mathcal{U}(0.6, 2.5)$
S3 light, Sérsic	$e_1$	$\mathcal{U}(0.20, 0.42)$
S3 light, Sérsic	$e_2$	$\mathcal{U}(0.00, 0.35)$
S3 light	$\theta_1^{\zeta}$ (arcsec)	$\mathcal{U}(0.75, 1.80)$
S3 light	$\theta_2^{\zeta}$ (arcsec)	$\mathcal{U}(1.70, 2.30)$
G1 light, F814W	$\theta_{\text{eff}}$ (arcsec)	$\mathcal{N}(0.61, 0.27)$ with bound [0.1, 2.7] (Section 5.2)
G1 light, F814W	$n_s$	$\mathcal{U}(2.0, 8.0)$
G1 light, F814W+F475X	$e_1$	$\mathcal{U}(-0.20, 0.00)$
G1 light, F814W+F475X	$e_2$	$\mathcal{U}(-0.05, 0.13)$
G1 light	$\theta_1^{\zeta}$ (arcsec)	$\mathcal{U}(0.023, 0.035)$
G1 light	$\theta_2^{\zeta}$ (arcsec)	$\mathcal{U}(-0.010, 0.003)$
G1 light, F475X	$\theta_{\text{eff}}$ (arcsec)	$\mathcal{N}(0.61, 0.27)$ with bound [0.1, 2.7] (Section 5.2)

**Table G2** – *continued*

Model component	Parameter	Prior
G1 light, F475X	$n_s$	$\mathcal{U}(1.0, 5.0)$
G1 light, F160W, Sérsic 1	$\theta_{\text{eff}}$ (arcsec)	$\mathcal{N}(0.61, 0.27)$ with bound [0.1, 2.7] (Section 5.2)
G1 light, F160W, Sérsic 1	$n_s$	$\mathcal{U}(1.0, 5.0)$
G1 light, F160W, Sérsic 1	$e_1$	$\mathcal{U}(-0.25, 0.00)$
G1 light, F160W, Sérsic 1	$e_2$	$\mathcal{U}(-0.10, 0.12)$
G1 light, F160W, Sérsic 2	$\theta_{\text{eff}}$ (arcsec)	$\mathcal{N}(0.61, 0.27)$ with bound [0.1, 2.7] (Section 5.2)
G1 light, F160W, Sérsic 2	$n_s$	$\mathcal{U}(0.6, 6.0)$
G1 light, F160W, Sérsic 2	$e_1$	$\mathcal{U}(-0.05, 0.10)$
G1 light, F160W, Sérsic 2	$e_2$	$\mathcal{U}(0.07, 0.25)$
G1 light, double Chameleon	$I_0, \text{Chm1}/I_0, \text{Chm2}$	$\mathcal{U}(0.2, 9.5)$
G1 light, Chameleon 1	$w_c$ (arcsec)	$\mathcal{U}(0.00, 0.10)$
G1 light, Chameleon 1	$w_t$ (arcsec)	$\mathcal{U}(0.20, 1.00)$
G1 light, Chameleon 1	$e_1$	$\mathcal{U}(-0.25, 0.25)$
G1 light, Chameleon 1	$e_2$	$\mathcal{U}(-0.25, 0.25)$
G1 light, Chameleon 2	$w_c$ (arcsec)	$\mathcal{U}(0.01, 1.50)$
G1 light, Chameleon 2	$w_t$ (arcsec)	$\mathcal{U}(2.50, 9.00)$
G1 light, Chameleon 2	$e_1$	$\mathcal{U}(-0.20, 0.20)$
G1 light, Chameleon 2	$e_2$	$\mathcal{U}(-0.20, 0.20)$
G2 light	$\theta_{\text{eff}}$ (arcsec)	$\mathcal{U}(0.25, 1.10)$
G2 light	$n_s$	$\mathcal{U}(2.0, 6.0)$
G2 light	$\theta_1^c$ (arcsec)	$\mathcal{U}(-1.60, -1.56)$
G2 light	$\theta_2^c$ (arcsec)	$\mathcal{U}(-0.97, -0.93)$
Image A	$\Delta\alpha$ (arcsec)	$\mathcal{U}(1.940, 1.948)$
Image A	$\Delta\delta$ (arcsec)	$\mathcal{U}(-1.576, -1.568)$
Image B	$\Delta\alpha$ (arcsec)	$\mathcal{U}(-1.819, -1.809)$
Image B	$\Delta\delta$ (arcsec)	$\mathcal{U}(0.263, 0.290)$
Image C	$\Delta\alpha$ (arcsec)	$\mathcal{U}(-1.935, -1.926)$
Image C	$\Delta\delta$ (arcsec)	$\mathcal{U}(-0.954, -0.940)$
Image D	$\Delta\alpha$ (arcsec)	$\mathcal{U}(0.096, 0.110)$
Image D	$\Delta\delta$ (arcsec)	$\mathcal{U}(1.385, 1.392)$
Differential dust extinction	$\tau_0^{\text{F814W}}$	$\mathcal{U}(0.1, 2.0)$

shapelet components – which are linear parameters – to compute the evidence using equation (35).

<sup>1</sup>Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA

<sup>2</sup>Kavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, Stanford, CA 94305, USA

<sup>3</sup>DARK, Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen, Denmark

<sup>4</sup>Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA

<sup>5</sup>Institute of Physics, Laboratoire d’Astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland

<sup>6</sup>Department of Astronomy & Astrophysics, University of Chicago, Chicago, IL 60637, USA

<sup>7</sup>Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

<sup>8</sup>National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

<sup>9</sup>STAR Institute, Quartier Agora - Allée du six Août, 19c B-4000 Liège, Belgium

<sup>10</sup>INAF - Osservatorio Astronomico di Capodimonte, Salita Moiariello, 16, I-80131 Napoli, Italy

<sup>11</sup>European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany

<sup>12</sup>Physics Department, UC Davis, 1 Shields Ave., Davis, CA 95616, USA

<sup>13</sup>Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

<sup>14</sup>The Inter-University Center for Astronomy and Astrophysics, Post bag 4, Ganeshkhind, Pune 411007, India

<sup>15</sup>Departamento de Ciencias Físicas, Universidad Andres Bello, Fernandez Concha 700, Las Condes, Santiago, Chile

<sup>16</sup>Institute of Astronomy, Madingley Rd, Cambridge CB3 0HA, UK

<sup>17</sup>Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

<sup>18</sup>Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

<sup>19</sup>Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile

<sup>20</sup>Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

<sup>21</sup>LSST, 933 North Cherry Avenue, Tucson, AZ 85721, USA

<sup>22</sup>Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390, USA

<sup>23</sup>Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

<sup>24</sup>Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>25</sup>Kavli Institute for Particle Astrophysics and Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA

<sup>26</sup>SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

<sup>27</sup>Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

<sup>28</sup>Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil

<sup>29</sup>Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA

- <sup>30</sup>*National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*
- <sup>31</sup>*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona) Spain*
- <sup>32</sup>*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
- <sup>33</sup>*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
- <sup>34</sup>*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
- <sup>35</sup>*Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*
- <sup>36</sup>*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil*
- <sup>37</sup>*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
- <sup>38</sup>*Excellence Cluster Origins, Boltzmannstr. 2, D-85748 Garching, Germany*
- <sup>39</sup>*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
- <sup>40</sup>*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
- <sup>41</sup>*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
- <sup>42</sup>*Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*
- <sup>43</sup>*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
- <sup>44</sup>*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
- <sup>45</sup>*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
- <sup>46</sup>*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
- <sup>47</sup>*Center for Astrophysics, Harvard and Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
- <sup>48</sup>*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
- <sup>49</sup>*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
- <sup>50</sup>*Observatories of the Carnegie Institution for Science, 813 Santa Barbara St., Pasadena, CA 91101, USA*
- <sup>51</sup>*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP 05314-970, Brazil*
- <sup>52</sup>*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*
- <sup>53</sup>*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
- <sup>54</sup>*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK*
- <sup>55</sup>*Instituto de Física, UFRGS, Caixa Postal 15051, Porto Alegre, RS - 91501-970, Brazil*
- <sup>56</sup>*Department of Physics, Duke University Durham, NC 27708, USA*
- <sup>57</sup>*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
- <sup>58</sup>*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
- <sup>59</sup>*Physics Department, Brandeis University, 415 South Street, Waltham, MA 02453, USA*
- <sup>60</sup>*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.