



Effects of Subhalos on Interpreting Highly Magnified Sources Near Lensing Caustics

Lingyuan Ji^{1,2} and Liang Dai² ¹ Berkeley Center for Cosmological Physics, University of California, Campbell Hall 341, Berkeley, CA 94720, USA² Department of Physics, University of California, 366 Physics North MC 7300, Berkeley, CA 94720, USA; liangdai@berkeley.edu

Received 2024 August 4; revised 2025 January 3; accepted 2025 January 6; published 2025 February 13

Abstract

Large magnification factors near gravitational lensing caustics of galaxy-cluster lenses allow the study of individual stars or compact stellar associations at cosmological distances. We study how the presence of subgalactic subhalos, an inevitable consequence of cold dark matter, can alter the property of caustics and hence change the interpretation of highly magnified sources that lie atop them. First, we consider a galaxy-cluster halo populated with subhalos sampled from a realistic subhalo mass function calibrated to N -body simulations. Then, we compare a semianalytical approximation and an adaptive ray-shooting method that we employ to quantify the property of the caustics. As a case study, we investigate Earendel, a $z = 6.2$ candidate of magnified single- or multiple-star system with a lone lensed image atop the critical curve in the Sunrise Arc. We find that the source size constraint ($\lesssim 0.3$ pc) previously derived from macrolens models should be relaxed by a factor of a few to 10 when subhalos are accounted for, therefore allowing the possibility of a compact star cluster. The subhalos could introduce an astrometric perturbation that is $\lesssim 0.5$, which does not contradict observation. These conclusions are largely robust to changes in the subhalo population. Subhalos therefore should be seriously accounted for when interpreting the astrophysical nature of similar highly magnified sources uncovered in recent high- z observations.

Unified Astronomy Thesaurus concepts: [Strong gravitational lensing \(1643\)](#); [Caustic curve \(2151\)](#); [Cold dark matter \(265\)](#)

1. Introduction

Gravitational magnification of light provides the best opportunities for the study of individual stars or stellar associations at cosmological distances. As first suggested by J. Miralda-Escudé (1991), individual luminous stars are detectable owing to large lensing magnification factors $\mu \sim \mathcal{O}(10^2\text{--}10^4)$ when they are fortuitously located in the vicinity of caustic curves cast by foreground massive lenses. The extreme magnification values realized there make highly magnified stars sensitive probes of dark matter (DM) substructure on scales comparable or much smaller than that of dwarf galaxies (L. Dai et al. 2018, 2020; L. Dai & J. Miralda-Escudé 2020; L. L. R. Williams et al. 2024; T. Broadhurst et al. 2024; J. M. Diego et al. 2024b) and of P compact objects, which may make up a fraction of the DM (J. M. Diego et al. 2018; M. Oguri et al. 2018; C. Vall Müller & J. Miralda-Escudé 2025). They may also offer excellent cases for studying the properties of massive stars in star formation sites at cosmological redshifts through multiband photometry (X. Han & L. Dai 2024) or spectroscopy (E. Lundqvist et al. 2024), or the accreting remnants of massive stars (R. A. Windhorst et al. 2018).

Indeed, individual stars subject to extremely large magnification factors have thus far been detected behind galaxy-cluster lenses, where it is convenient that the sightline toward the host galaxy is not severely contaminated by the diffuse starlight of cluster-member galaxies. Several earlier detections from $z \sim 0.7\text{--}1.5$ exhibited time-varying fluxes over the timescale of days to years (P. L. Kelly et al. 2018; S. A. Rodney et al. 2018; W. Chen et al. 2019; A. A. Kaurov et al. 2019), which

can be explained in terms of the macrocaustic disrupted into a network of microcaustics cast by foreground intracluster stars (T. Venumadhav et al. 2017; J. M. Diego et al. 2018; M. Oguri et al. 2018; also see L. Weisenbach et al. 2024 for a topical review). Taking advantage of microlensing-induced variability, recent studies have demonstrated that highly magnified individual stars can be efficiently detected in caustic-crossing host galaxies at these redshifts through image difference of multiepoch deep exposures, with the Hubble Space Telescope (HST; P. L. Kelly et al. 2022; A. K. Meena et al. 2023) and with the James Webb Space Telescope (H. Yan et al. 2023; Y. Fudamoto et al. 2024).

Intriguingly, recent observations have also uncovered additional lensed-star candidates in higher- z galaxies that show different characteristics. While these sources appear close to macrocaustic curves and hence must be highly magnified, they do not show a resolved pair of lensed images as expected near a fold caustic and neither does the lone image exhibit detectable flux variability. One notable example is Earendel in the $z = 6.2$ Sunrise galaxy, which has been suggested to be a single star or a system of binary or multiple stars (B. Welch et al. 2022a, 2022b). Another exceptionally bright lensed source, nicknamed Godzilla and located in the $z = 2.37$ Sunburst galaxy, appears to be a single stellar object as it also shows as a lone image without variability (E. Vanzella et al. 2020; J. M. Diego et al. 2022). S. Choe et al. (2024) propose that the source is a post-eruption luminous blue variable, while M. Pascale & L. Dai (2024) instead suggest that it is a young star cluster with a large number of hot stars powering nebular emission. L. J. Furtak et al. (2023) performed JWST/NIRSpec spectroscopy of another highly magnified compact source at $z = 4.7$ with a lone, persistent lensed image on the critical curve, and suggested that the source is likely a binary star due to a source-size constraint $\lesssim 0.5$ pc from lens modeling. If any one of the above is truly an individual stellar object, or even proven



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

to be a metal-poor or zero-metallicity star (A. T. P. Schauer et al. 2022; E. Zackrisson et al. 2023), it will be a remarkable discovery about stellar objects in the young Universe.

In B. Welch et al. (2022a), one key reasoning is that the lone, unresolved image of Earendel right on top of the theoretically predicted lensing critical curve requires that the source must be more compact than ~ 0.3 pc in radius. The authors conclude that Earendel cannot be a star cluster but must be a single star, or a small system of binary or multiple stars (B. Welch et al. 2022b). The tight size constraint derives from very small values for the third-order derivatives of the Fermat potential on the critical curve, which is a consequence of the large Einstein angular scale for the galaxy-cluster lens; in other words, the fold caustic associated with the galaxy-cluster lens has a high strength (P. Schneider et al. 1992). This small third-order derivative is independently inferred by several macrolens models constructed for Earendel’s cluster lensing field, with the uncertainty of a factor of a few between models (B. Welch et al. 2022a). While the nondetection of large temporal flux change would be natural for a lensed star cluster with many member stars (L. Dai 2021), for a single stellar source the absence could be the consequence of a hypercritical density of microcaustics due to extremely large macromagnification factors for a finite-size source (L. Dai & M. Pascale 2021; this argument is also put forth in L. J. Furtak et al. 2023).

Here we point out that the large caustic strength is inferred only from a smooth macrolens model—one that only accounts for DM subhalos massive enough to host optically visible galaxies. If the DM is cold, hierarchical structure formation should lead to a continuous spectrum of subhalo masses (W. H. Press & P. Schechter 1974) that may extend well below the galactic mass scale ($\lesssim 10^{10} M_\odot$). At cosmological distances, the subgalactic DM halos that reside within the foreground galaxy-cluster lens are not expected to host optically detectable stellar populations (T. Okamoto et al. 2008). The perturbing effects from subgalactic subhalos become prominent near a macrocritical curve (L. Dai et al. 2018, 2020; L. L. R. Williams et al. 2024). While these subgalactic subhalos make tiny corrections to ray deflection, convergence, and shear, they may increase the third-order derivatives of the Fermat potential and hence decrease the typical caustic strengths. In this paper, we will show that this effect significantly relaxes the size constraint for Earendel so that it may as well be a compact star cluster several parsecs across. In general, we suggest that the effect of subhalos should be taken into account when interpreting highly magnified sources similar to Earendel.

This paper is organized as follows. In Section 2, we start with a review of lensing in the vicinity of fold caustics and discuss the semianalytic model we adopt to model subhalos as lensing perturbations. Then in Section 3, we rigorously define observables that we can use to quantify the constraint on the size of the source as well as the position of the photometric center. In Section 4, we introduce two methods that we adopt to calculate those observables for numerically simulated lensing realizations with DM subhalos. Results will be presented and discussed in Section 5. The main body of this paper will end in Section 7 with concluding remarks.

We additionally collect technical details related to this work in two appendices. In Appendix A the readers can find details on our treatment of lensing by the macrofold caustic and by the DM subhalos. In Appendix B we present an analytic derivation

of the constraint on the source size in the vicinity of a fold caustic.

2. Lens Model

As an important case study, we will analyze Earendel as reported in B. Welch et al. (2022a). The host galaxy WHL0137-zD1 at $z_S = 6.2$ is lensed by the foreground galaxy cluster WHL0137-08 at $z_L = 0.566$, forming a $15''$ long Sunrise Arc (B. Salmon et al. 2020). Earendel is located almost exactly on the macromodel critical curve and was suspected to be highly magnified.

We approximate the lens mapping in the vicinity of Earendel by accounting for three components of lenses: (i) a smooth-fold caustic (P. Schneider et al. 1992) representing the effect of the coarse-grained mass distribution of the cluster halo as if mass locked up in subgalactic subhalos is all smoothed out; (ii) a collection of subhalos assumed to have a tidally truncated Navarro–Frenk–White (NFW) profile (J. F. Navarro et al. 1996) with random center positions within a sufficiently large circular region on the image plane centered on Earendel; and (iii) a (negative) constant surface mass density added in the same circular region to enforce mass conservation after adding subhalos.

Let $\alpha_{\text{FD}}(\mathbf{x})$, $\alpha_{\text{SH}}^{(n)}(\mathbf{x})$, and $\alpha_{\text{UD}}(\mathbf{x})$ be the ray deflection vector of the smooth local fold model, that from the n th subhalo, and that from the uniform negative-mass disk, respectively. The total deflection vector is

$$\alpha(\mathbf{x}) = \alpha_{\text{FD}}(\mathbf{x}) + \sum_{n=1}^N \alpha_{\text{SH}}^{(n)}(\mathbf{x}) + \alpha_{\text{UD}}(\mathbf{x}), \quad (1)$$

which enters the ray equation

$$\mathbf{y} = \mathbf{x} - \alpha(\mathbf{x}). \quad (2)$$

In what follows, we briefly describe the implementation of components (i) and (ii), while component (iii) is easy to treat. More detailed information for all components can be found in Appendix A.

2.1. Smooth Fold

Near the critical curve, the Cartesian components of the deflection vector caused by the smooth fold can be expressed as a power expansion

$$\alpha_{i,\text{FD}}(\mathbf{x}) = x_i - \phi_{ij}^0 x_j - \frac{1}{2} \phi_{ijk}^0 x_j x_k + O(|\mathbf{x}|^3). \quad (3)$$

Here, $\phi(\mathbf{x}, \mathbf{y}) \equiv (\mathbf{x} - \mathbf{y})^2/2 - \psi(\mathbf{x})$ is the Fermat potential with $\psi(\mathbf{x})$ being the lensing potential. We adopt the shorthand notation $\phi_{ij\dots} \equiv \partial\phi/(\partial x_i \partial x_j \dots)$ for its derivatives at various orders. The bold-faced superscript $\mathbf{0}$ indicates that the quantity is evaluated at the coordinate origin $\mathbf{x} = \mathbf{0}$, which is taken to be the location of Earendel. The first and second derivatives of deflection can be found in Appendix A.1. Since we are concerned with the vicinity of the smooth critical curve, $|\mathbf{x}|$ is small and hence from now on we keep only the leading-order term.

The smooth fold determines the main features of the criticality. If the smooth fold is the *only* component of the lens model, the Jacobian matrix $A_{ij} \equiv \partial y_i / \partial x_j = \phi_{ij}$ has two (normalized) eigenvectors ($\mathbf{e}_\parallel, \mathbf{e}_\perp$), associated with two

eigenvalues $(\lambda_{\parallel}, \lambda_{\perp})$, respectively. The eigenvectors are given by

$$\mathbf{e}_{\parallel} = \mathbf{e}_{\parallel}^0 \quad \text{and} \quad \mathbf{e}_{\perp} = \mathbf{e}_{\perp}^0, \quad (4)$$

where $\mathbf{e}_{\parallel}^0$ and $\mathbf{e}_{\perp}^0$ are the eigenvectors associated with the zero and nonzero eigenvalues of $A_{ij}^0 = \phi_{ij}^0$, respectively. The eigenvalues are given by

$$\lambda_{\parallel} = \mathbf{d}^0 \cdot \mathbf{x} \quad \text{and} \quad \lambda_{\perp} = 2(1 - \kappa^0), \quad (5)$$

where $\kappa^0 \equiv 1 - (\phi_{11}^0 + \phi_{22}^0)/2$ and $\mathbf{d}^0 \equiv \mathbf{D}^0/[2(1 - \kappa^0)]$. Here, we define the Jacobian gradient vector $\mathbf{D}^0$, which has Cartesian components

$$D_i^0 \equiv \phi_{22}^0 \phi_{11i}^0 + \phi_{11}^0 \phi_{22i}^0 - 2\phi_{12}^0 \phi_{12i}^0. \quad (6)$$

We term $\mathbf{e}_{\parallel}$ the degenerate direction (i.e., the elongation direction of the lensed arc) and $\mathbf{e}_{\perp}$ the perpendicular direction. We define

$$\mu_{\parallel} \equiv \frac{1}{\lambda_{\parallel}} = \frac{1}{\mathbf{d}^0 \cdot \mathbf{x}} \quad \text{and} \quad \mu_{\perp} \equiv \frac{1}{\lambda_{\perp}} = \frac{1}{2(1 - \kappa^0)} \quad (7)$$

to be the magnification in the degenerate direction and the perpendicular direction, respectively. Thus, $\mu_{\parallel}$ has a large absolute value that increases toward the smooth critical curve, while $\mu_{\perp}$ has a constant moderate value. The determinant of the Jacobian matrix $A_{ij}(\mathbf{x}) \equiv \phi_{ij}(\mathbf{x})$ is

$$\det A(\mathbf{x}) = \lambda_{\parallel} \lambda_{\perp} = \mathbf{D}^0 \cdot \mathbf{x}, \quad (8)$$

which implies that $\mathbf{D}^0$ is normal to the critical curve and the flux magnification is

$$\mu = \mu_{\parallel} \mu_{\perp} = \frac{1}{\mathbf{D}^0 \cdot \mathbf{x}}. \quad (9)$$

The magnitude $D^0 \equiv |\mathbf{D}^0|$ of the vector $\mathbf{D}^0$ therefore is the combination of the third-order derivatives that quantifies the strength of the caustic as it determines the magnification factor of a lensed image at a given distance to the critical curve. A stronger caustic has a smaller D^0 value, so that a lensed image at a fixed distance from the critical curve has a larger magnification factor than in the case of a weaker caustic with a larger D^0 value.

Several macrolens models have been constructed for the Sunrise Arc without including the numerous subhalos that are not expected to host visible galaxies. These models make moderately different predictions for the criticality near Earendel (B. Welch et al. 2022a). At the location of Earendel, lens-model predictions for the macroconvergence fall in the range $\kappa^0 = 0.5$ – 0.8 , and we shall use a fiducial value $\kappa^0 = 0.5$. The degenerate direction and the critical direction (i.e., the tangential direction of the critical curve at the location of Earendel) form an angle α whose value is in the range 22° – 41° , and we shall use a fiducial value $\alpha = 30^\circ$. The Jacobian gradient $\mathbf{D}^0$ of the smooth fold is perpendicular to the critical curve with a model-predicted magnitude in the range $D^0 = 1/(113'')$ – $1/(18'')$. We will numerically study the case $D^0 = 1/(18'')$. There are two arguments for this choice. First, as we will show in Section 3.1, the largest D^0 implies the weakest caustic and thus the most conservative size constraint for Earendel. This most conservative case is crucial for B. Welch et al. (2022a) to rule out the star-cluster scenario. Second, the weakest caustic is also the one that is most immune to

perturbations from subhalos, due to their already large third-order derivatives. As a result, subhalos will be important to all other possible D^0 values in the range $D^0 = 1/(113'')$ – $1/(18'')$ if it is important for the case where $D^0 = 1/(18'')$.

2.2. Subhalo Abundance

Direct N -body simulations of a galaxy-cluster halo that resolve subhalos with subgalactic masses $m = 10^6$ – $10^{10} M_{\odot}$ are computationally demanding. Instead, we adopt a semianalytical model devised by J. Han et al. (2016) that is calibrated to and verified by simulations for the more massive subhalos and then extrapolated to subgalactic subhalo masses. This detailed model accounts for the effects of subhalo infall and tidal stripping. The essential parameters of the model are the host-halo density profile $\rho(R)$ as a function of the host halo–centric three-dimensional radius R and the lower (upper) limit of the subhalo infall mass $m_{\min}$ ($m_{\max}$). The model outputs a joint distribution $d^2N/(d \ln m d^3R)$ as a function of the subhalo mass m and the location R . The explicit expression for the distribution, as well as other secondary parameters, are provided in Appendix A.2.

We then map the three-dimensional spatial distribution to the projected two-dimensional distribution needed for studying the lensing effect, through a change of variable

$$\frac{dN(m, Z, B)}{d \ln m dZ d^2B} = \frac{dN(m, \sqrt{Z^2 + B^2})}{d \ln m d^3R}, \quad (10)$$

where B is the host halo–centric impact parameter and Z is the longitudinal coordinate, related by $R^2 = Z^2 + B^2$. In the vicinity of Earendel, the variation in B is negligible. Thus, for a given impact parameter B and within a field of view of area ΔS , we draw subhalos from the distribution

$$\left. \frac{dN(m, Z)}{d \ln m dZ} \right|_B = \frac{dN(m, Z, B)}{d \ln m dZ d^2B} \times \Delta S, \quad (11)$$

and randomly distribute them within the field of view. It is important that we sample the joint distribution of (m, Z) , as the subhalo parameters (c_{200}, r_t) at given m depends on the three-dimensional radial distance $R = (Z^2 + B^2)^{1/2}$.

The host halo of the galaxy cluster WHL0137-08 has an estimated mass $M_{500} = 9 \times 10^{14} M_{\odot}$ with uncertainty in the concentration parameter C_{500} (B. Welch et al. 2022a). We assume that the host halo can be described by the NFW model and follow B. Diemer & M. Joyce (2019) to obtain $C_{500} = 2.6$. We choose a mass $M_{200} = 1.3 \times 10^{15} M_{\odot}$ and a concentration $C_{200} = 4.0$ for the host halo at the density contrast 200. These numbers decently reproduce the NFW parameter values at the density contrast 500. The line of sight toward Earendel is roughly $40''$ away from the brightest cluster galaxy, which we take to be the center of the host halo. The impact parameter with respect to the host-halo center is $B = 268$ kpc. Since we mainly concern the effect of the more numerous subgalactic subhalos, we shall fix $m_{\max} = 10^9 M_{\odot}$, but explore different $m_{\min}$ values.

2.3. Subhalo Density Structure

Following L. Dai et al. (2018), we model the subhalos using the truncated NFW profile (E. A. Baltz et al. 2009;

F.-Y. Cyr-Racine et al. 2016; L. Dai et al. 2018)

$$\rho_{\text{SH}}(r) = \frac{m_{200}}{4\pi r_s^3 f(c_{200})} \frac{1}{(r/r_s)(1+r/r_s)^2} \frac{1}{1+r^2/r_t^2}, \quad (12)$$

where $f(x) \equiv \ln(1+x) - x/(1+x)$. This model is parameterized by three independent parameters: the underlying NFW mass m_{200} , the underlying NFW concentration parameter c_{200} , and a tidal truncation radius r_t . The underlying NFW scale radius is then given in terms of m_{200} and c_{200}

$$r_s \equiv \frac{1}{c_{200}} \left(\frac{3}{4\pi} \frac{m_{200}}{200 \rho_{\text{crit}}} \right)^{1/3}. \quad (13)$$

We note that our subhalo profile has a central cusp, unlike the cored density profile considered in L. L. R. Williams et al. (2024).

For a given subhalo with mass m and location R , we find the parameters (m_{200}, c_{200}, r_t) in the following way: we first identify the tidal truncation radius with the instantaneous tidal radius,

$$r_t(m, R) = R \left(\frac{m}{M_<(R)} \right)^{1/3} \left(3 - \frac{d \ln M_<(R)}{d \ln R} \right)^{-1/3}, \quad (14)$$

where $M_<(R)$ is the host-halo mass enclosed within radius R . We then determine the subhalo's underlying NFW concentration c_{200} using the phenomenological relation (J. S. Bullock et al. 2001; J. Diemand et al. 2007; R. Bartels & S. Ando 2015; Á. Moliné et al. 2017)

$$c_{200}(m, R) = \bar{c}_{200}(m) \left\{ 1 + \frac{1}{15} \left[\frac{1.5^2}{0.1^2 + (R/R_{200})^2} \right]^{1/2} \right\}, \quad (15)$$

where R_{200} is the host-halo radius and $\bar{c}_{200}(m)$ is the mean concentration for field halos as described in A. D. Ludlow et al. (2016). Finally, we solve for an underlying NFW mass m_{200} that reproduces the correct total mass $m = \int_0^\infty 4\pi r^2 dr \rho_{\text{SH}}(r)$.

The deflection caused by a subhalo with parameters (m_{200}, c_{200}, r_t) located at $\mathbf{x}_0$ is given by

$$\alpha_{\text{SH}}(\mathbf{x}) = \frac{4 G m_{200}}{c^2 f(c_{200}) r_s} \frac{D_{\text{LS}}}{D_s} \frac{\mathbf{r}}{|\mathbf{r}|} S \left(\frac{D_L}{r_s} |\mathbf{r}|, \frac{r_t}{r_s} \right), \quad (16)$$

where $\mathbf{r} \equiv \mathbf{x} - \mathbf{x}_0$ and $S(\xi, \tau)$ is a dimensionless function, whose explicit definition and method of evaluation are detailed in Appendix A.3. In the above, D_L , D_s , and D_{LS} are the angular diameter distances to the lens, to the source, and from the lens to the source, respectively. We also provide analytic expressions for the first and second derivatives of the deflection angle in Appendix A.3.

3. Observables

Through the ray equation, Equation (2), subhalos perturb the smooth-fold caustic and, in the most dramatic cases, break it into a corrugated collection of smaller fold caustics. We shall refer to those as the “subfolds.” The fact that Earendel does not show as a resolved pair of images suggests that it is nearly exactly on one of the subfolds, if the original smooth macrofold caustic is disrupted. However, the subfolds, on average, have weaker caustic strengths than the smooth fold. This biases the

size constraint and hence affects the interpretation of Earendel's astrophysical nature. In this section, we define several important observables of lensing.

3.1. Size Constraint

Earendel appears as a single, unresolved source. In the analysis of B. Welch et al. (2022a), where only the smooth fold is accounted for, one expects that the two lensed images of Earendel straddle the smooth critical curve at $\mathbf{x}_{1,2} = \pm \Delta \mathbf{x} \mathbf{e}_\parallel$. We shall adopt the agnostic assumption that the intrinsic surface-brightness profile of Earendel is a two-dimensional isotropic Gaussian function with an angular standard deviation σ_s . Then each image will be an elongated two-dimensional Gaussian profile with size $|\mu_\parallel| \sigma_s$ and $|\mu_\perp| \sigma_s$ along the degenerate and the perpendicular direction, respectively. Since $|\mu_\parallel| \gg |\mu_\perp|$ near the critical curve, the lensed images are flattened to one-dimensional Gaussians with an angular standard deviation $|\mu_\parallel| \sigma_s$.

As neither the image-pair separation nor the image size is astrometrically resolved, it is required that

$$\Delta x \lesssim \mathcal{R} \quad \text{and} \quad |\mu_\parallel| \sigma_s \lesssim \mathcal{R}, \quad (17)$$

where $\mathcal{R}$ represents the angular resolution ($\simeq 0.055$ for HST as tested by B. Welch et al. 2022a) of the observation, as quantified by the size of the point-spread function (PSF). Using Equation (7), it follows from Equation (17) that the proper size $r_s \equiv \sigma_s D_s$ of Earendel is constrained to be

$$r_s \lesssim \mathcal{R}^2 D_s |\mathbf{d}^0 \cdot \mathbf{e}_\parallel^0|. \quad (18)$$

Beyond this crude analysis, it is nontrivial to determine if an arbitrary set of images will be exactly unresolved (i.e., consistent with a point source) under a given PSF. To facilitate better comparison between different methods that we will use to investigate this observable in Section 4, we shall use the photometric moments for quantification.

Specifically, the source surface-brightness profile is

$$b(s) = \frac{1}{2\pi \sigma_s^2} \exp \left(-\frac{s^2}{2 \sigma_s^2} \right), \quad (19)$$

where $\mathbf{s} \equiv \mathbf{y} - \mathbf{y}_s$ with $s \equiv |\mathbf{s}|$ and $\mathbf{y}_s$ is the center of the source. The surface brightness observed at image position $\mathbf{x}$ is then $b(\mathbf{y}(\mathbf{x}) - \mathbf{y}_s)$.

Along the direction of elongation, we define the photometric moment (a two-by-two matrix)

$$\Sigma_x \equiv \frac{\int (\mathbf{x} - \mathbf{C}_x)(\mathbf{x} - \mathbf{C}_x) b[\mathbf{y}(\mathbf{x}) - \mathbf{y}_s] d^2x}{\int b[\mathbf{y}(\mathbf{x}) - \mathbf{y}_s] d^2x}, \quad (20)$$

where $\mathbf{C}_x$ is the photometric center defined as

$$\mathbf{C}_x \equiv \frac{\int \mathbf{x} b[\mathbf{y}(\mathbf{x}) - \mathbf{y}_s] d^2x}{\int b[\mathbf{y}(\mathbf{x}) - \mathbf{y}_s] d^2x}. \quad (21)$$

The size σ_x of a set of images is then defined using the photometric moment Σ_x as

$$\sigma_x \equiv [\varrho(\Sigma_x)]^{1/2}, \quad (22)$$

where ϱ is the spectral-radius operator that returns the maximum of the absolute values of the matrix eigenvalues.

We define that the set of images is unresolved if $\sigma_x < \mathcal{R}$. Under this definition, the size constraint $\bar{\sigma}_S$ is defined such that all sources with size $\sigma_S > \bar{\sigma}_S$, for any center position $\mathbf{y}_S$, would have been resolved. Let $\sigma_x(\mathbf{y}_S, \sigma_S)$ be the size defined in Equation (22) as a function of $\mathbf{y}_S$ and σ_S . Then $\bar{\sigma}_S$ can be expressed as

$$\bar{\sigma}_S \equiv \min\{\sigma_S | \forall \mathbf{y}_S, \sigma_x(\mathbf{y}_S, \sigma_S) \geq \mathcal{R}\}. \quad (23)$$

However, there is one subtlety that compels us to reconsider the definition presented above. There is a half-plane region in the source plane such that, for any $\mathbf{y}$, the ray equation, Equation (2), admits no solution $\mathbf{x}$ (P. Schneider et al. 1992). This means that no ray originating from this region will arrive at the observer (other modestly magnified lensed images may still form, with the corresponding rays passing the lens plane somewhere else on the lensed arc and far away from the particular critical curve we are concerned with here). We term this region the external side of the fold caustic. Because of this, it is in fact possible to have an arbitrarily large source, with nearly all of it hiding on the external side. In this case, only a tiny portion of the source is visible, which can be unresolved. This was noted as a difficult corner case in B. Welch et al. (2022a) but was not further considered due to its small probability. This ambiguity can be removed if we instead constrain the size of the *visible* portion of the source along the degenerate direction $\mathbf{e}_\parallel^0$. We define the source-plane photometric moment for the visible portion of the source as

$$\Sigma_{V,y} \equiv \frac{\int_V (\mathbf{y} - \mathbf{C}_{V,y})(\mathbf{y} - \mathbf{C}_{V,y})b(\mathbf{y} - \mathbf{y}_S) d^2y}{\int_V b(\mathbf{y} - \mathbf{y}_S) d^2y}, \quad (24)$$

where the subscripts V indicate that the integral is performed only over the visible portion of the source. The photometric center of the visible portion of the source is defined as

$$\mathbf{C}_{V,y} \equiv \frac{\int_V \mathbf{y} b(\mathbf{y} - \mathbf{y}_S) d^2y}{\int_V b(\mathbf{y} - \mathbf{y}_S) d^2y}. \quad (25)$$

We now define the one-dimensional size for the visible portion of the source as

$$\sigma_V \equiv [\mathbf{e}_\parallel^0 \cdot \Sigma_{V,y} \cdot \mathbf{e}_\parallel^0]^{1/2}. \quad (26)$$

This is measured along the direction of elongation $\mathbf{e}_\parallel^0$ because astrometric constraint is only possible in this direction, along which the source is hugely magnified.

In the rest of the paper, we shall use σ_V in place of σ_S . We follow the same procedure in obtaining the constraints $\bar{\sigma}_V$ on σ_V as

$$\bar{\sigma}_V \equiv \min\{\sigma_V | \forall \mathbf{y}_S, \sigma_x(\mathbf{y}_S, \sigma_V) \geq \mathcal{R}\}, \quad (27)$$

where $\sigma_x = \sigma_x(\mathbf{y}_S, \sigma_V)$ is now a function of $\mathbf{y}_S$ and σ_V , not σ_S . According to Appendix B, the constraint on the size of the visible portion of the source $r_V \equiv D_S \sigma_V$ is

$$r_V < \bar{r}_V \equiv D_S \bar{\sigma}_V = \mathcal{R}^2 D_S |\mathbf{d}^0 \cdot \mathbf{e}_\parallel^0|. \quad (28)$$

This result agrees with that of the crude analysis in Equation (18). Nevertheless, we emphasize that Equation (28) is the exact size constraint for the visible portion of the source.

Using numerical values from B. Welch et al. (2022a), we derive

$$\bar{r}_V \simeq 0.48 \text{ pc} \left(\frac{\mathcal{R}}{0.''055} \right)^2 \left(\frac{D_S}{1182 \text{ Mpc}} \right) \left(\frac{|\mathbf{d}^0 \cdot \mathbf{e}_\parallel^0|}{\frac{1}{18''} \cdot \cos \frac{\pi}{3}} \right) \quad (29)$$

for the conservative choice $D^0 = 1/(18'')$. The size constraint tightens to $\bar{r}_V = 0.08 \text{ pc}$ if $D^0 = 1/(113'')$. This is the key argument in B. Welch et al. (2022a) that disfavors the star-cluster scenario. For a comparison, one of the most compact star clusters found in our cosmic backyard, the R136 cluster in the Large Magellanic Cloud, has a half-light radius of $\sim 1.7 \text{ pc}$ (D. A. Hunter et al. 1995). In the Milky Way and M31, globular clusters all have half-light radii larger than 1 pc (P. Barmby et al. 2007; H. Baumgardt & M. Hilker 2018).

We will show that the argument presented in B. Welch et al. (2022a) needs to be revised if the smooth-fold caustic is disrupted into subfolds due to perturbations from subhalos. Furthermore, given the possibility that more than two images could form with subhalos, we must extend the definition “unresolved” to the following: we first find all unresolved clusters of lensed images (hereafter image clusters) at an angular resolution $\lesssim \mathcal{R}$; we then require that Earendel must correspond to the brightest image cluster, while the second-brightest image cluster, if in existence, must be at least $r_{\text{thres}} = 10$ times fainter and hence remains undetected. We have chosen this value for the minimum magnification ratio based on the point-source detection limit of the JWST/NIRCam images. We will derive a more reliable size constraint following this requirement.

3.2. Astrometry of Photometric Center

Subhalos can significantly perturb the position of the lensed images near a macrocritical curve (L. Dai et al. 2018). Judging from Earendel's location relative to the lensed image pair of a compact source on the same arc, whose separation must be bisected by the *smooth* critical curve (each image is nearly $1''$ from Earendel), Earendel shows almost exactly on top of the *smooth* critical curve. This implies that the typical astrometric departure of the perturbed critical curve from the smooth one at Earendel's location cannot be too large, which may constrain the abundance of subhalos. We shall define an astrometric observable to quantify this effect.

Let $\mathbf{c}$ be the observed photometric center of Earendel. For a smooth-fold caustic we would expect a photometric center $\mathbf{c}'$ that is along the degenerate direction and exactly on top of the smooth critical curve:

$$(\mathbf{c} - \mathbf{c}') \cdot \mathbf{e}_\perp^0 = 0 \quad \text{and} \quad \mathbf{D}^0 \cdot \mathbf{c}' = 0, \quad (30)$$

respectively. Solving for $\mathbf{c}'$ gives

$$\mathbf{c}' = (\mathbf{c} \cdot \mathbf{e}_\perp^0) \left[\mathbf{e}_\perp^0 - \left(\frac{\mathbf{D}^0 \cdot \mathbf{e}_\perp^0}{\mathbf{D}^0 \cdot \mathbf{e}_\parallel^0} \right) \mathbf{e}_\parallel^0 \right]. \quad (31)$$

We then define the difference $\Delta \mathbf{c} \equiv (\mathbf{c} - \mathbf{c}') \cdot \mathbf{e}_\parallel^0$ of these two predictions along the degenerate direction $\mathbf{e}_\parallel^0$ to be the

astrometric shift. Using Equation (31), we obtain

$$\Delta c = c \cdot \left[e_{\parallel}^0 + \left(\frac{D^0 \cdot e_{\perp}^0}{D^0 \cdot e_{\parallel}^0} \right) e_{\perp}^0 \right]. \quad (32)$$

We emphasize that this vector is along the degenerate direction, not perpendicular to the smooth critical curve. Thus it is suitable in determining, for instance, how far Earendel is from the midpoint of other lensed image pairs along the arc.

4. Methods

We employ two methods, one semianalytic and another numerical, to derive the key quantitative results of this work. In the first method, we assume that Earendel is located close enough to a subfold, so that the local lensing map can still be approximated by a simple fold, with the caustic strength describing the subfold rather than the smooth macrofold. We refer to this method as the “subfold approximation.”

In the second method, we exactly compute all lensed images of a finite-sized source via the inverse ray-shooting method augmented with adaptive mesh refinement. We refer to this method as the “exact numerical” method. In what follows, we first introduce both methods, and then describe how the observables defined in Section 3 can be extracted.

Both methods start with shooting rays according to Equation (2) on an uniform $N_1 \times N_2$ parallelogram mesh

$$\mathbf{x}_{i_1 i_2} \equiv l_{1,i_1} \mathbf{n}_1 + l_{2,i_2} \mathbf{n}_2. \quad (33)$$

Here, $i_1 = 1, 2, \dots, N_1$ and $i_2 = 1, 2, \dots, N_2$ are the indices for the mesh basis $\mathbf{n}_1$ and $\mathbf{n}_2$, respectively. And l_{1,i_1} and l_{2,i_2} are uniform nodes on $[-L_1/2, +L_1/2]$ and $[-L_2/2, +L_2/2]$ (including endpoints), respectively. We choose $\mathbf{n}_1$ to be along the degenerate direction, i.e., $\mathbf{n}_1 = e_{\parallel}^0$, and $\mathbf{n}_2$ to satisfy $\mathbf{n}_2 \cdot D^0 = 0$ so that it is along the smooth critical curve. This choice mitigates the loss of one of the lensed image pair beyond the ray-shooting domain. The area of the domain $\Delta\Omega$ is hence the area of the parallelogram,

$$\Delta\Omega \equiv L_1 L_2 \sqrt{1 - (\mathbf{n}_1 \cdot \mathbf{n}_2)^2}. \quad (34)$$

We note that $\Delta\Omega$ is different from the field ΔS of subhalo sampling used in Equation (11). We ensure that the field ΔS always encloses the field $\Delta\Omega$. For each ray $\mathbf{x}_{i_1 i_2}$ shot, we compute and store the deflection α , the first-order derivatives $\partial\alpha_i/\partial x_j$, and the second-order derivatives $\partial^2\alpha_i/\partial x_j \partial x_k$.

4.1. Subfold Approximation

Following ray shooting, we use a contour-finding algorithm to trace out the critical curve $\mathbf{x}(t)$ where $\det A(\mathbf{x})$ vanishes. Here, t is the parameter along the critical curve, and we define the subfold $\mathbf{y}(t) \equiv \mathbf{x}(t) - \alpha[\mathbf{x}(t)]$. At $\mathbf{y}(t)$, similar to Section 2.1, we define the subfold parameters $(\kappa^t, \mathbf{d}^t, e_{\parallel}^t, e_{\perp}^t)$, now with the superscript t emphasizing the location of evaluation. We now assume that the source is very close to the subfold and only one pair of highly magnified images are relevant, with the midpoint $\mathbf{x}(t)$ on the critical curve. For each source, we study the following three properties:

Consistency with Earendel. We compute the photometric size σ_x of the source as defined in Equation (22) using the analytic expression Equation (B4) provided in Appendix B. If

$\sigma_x < \mathcal{R}$, we deem the source as consistent with Earendel, and otherwise not.

Visible Proper Size. We compute the visible proper size σ_v of the source as defined in Equation (26) using the analytic expression Equation (B7) provided in Appendix B, which is appropriate for the fold caustic lens mapping.

Astrometry of the Photometric Center. Under the subfold approximation, the two images formed only differ in parity. Thus the photometric center will lie exactly on the critical curve, i.e., $\mathbf{c} = \mathbf{x}(t)$, giving

$$\Delta c = \mathbf{x}(t) \cdot \left[e_{\parallel}^0 + \left(\frac{D^0 \cdot e_{\perp}^0}{D^0 \cdot e_{\parallel}^0} \right) e_{\perp}^0 \right]. \quad (35)$$

Note that now σ_x , σ_v , and Δc are all stochastic variables dependent on the subhalo realization drawn from the population model in Section 2.2. In practice, we approximate the ensemble statistics as statistics of the parameter t weighted by the subfold length $|dy/dt| dt$.

4.2. Exact Numerical

The subfold approximation is straightforward but only accounts for one pair of lensed images. This approximation may also be inaccurate if the source's distance to the subfold is greater than the source-plane length scale over which the simple fold is a good approximation. The latter problem becomes particularly severe when lower-mass subhalos induce increasingly smaller and numerous wiggles along the critical curve, rendering the critical curve essentially a fractal.

To improve upon the subhalo approximation, we develop an image-finding algorithm using adaptive mesh refinement, which follows the algorithm described in R. B. Metcalf & M. Petkova (2014). We term this algorithm the “telescoping” algorithm. When the ray equation, Equation (2), is given, the algorithm finds all images with magnification $|\mu| \gtrsim \mu_{\min}$ for a finite-sized source.

Let $b(s)$ be the surface-brightness profile of the source, where $s \equiv \mathbf{y} - \mathbf{y}_s$ is the displacement from the source center $\mathbf{y}_s$. We start by choosing a circular disk centered at $\mathbf{y}_s$ with a large radius Δy_s that circumscribes the source. We then consider a sequence of concentric disks with decreased radii $\Delta y_n = \Delta y_0/3^n$ starting with $\Delta y_0 = \Delta x_0/\mu_{\min}^{1/2}$, where Δx_0 is the spatial resolution of the initial ray-shooting mesh (see Equation (33)). For the n th step, rays that hit the circular disk with radius Δy_n are kept. The parallelogram cells that are occupied by these rays are individually divided into 3×3 congruent parallelogram subcells, and new rays are shot at the center of each subcell. The new rays, together with the old rays that hit, are then tested against a circular disk with radius Δy_{n+1} in the $(n+1)$ th step. This process is iterated until the disk becomes smaller than Δy_s , in which case we instead set the disk radius to be Δy_s and collect all rays $\mathbf{x}_i$ that hit the disk. We then compute the surface brightness $\{b_i\}$ associated with rays $\{\mathbf{x}_i\}$ with $b_i \equiv b(\mathbf{y}_i - \mathbf{y}_s)$, where $\mathbf{y}_i \equiv \mathbf{x}_i - \alpha(\mathbf{x}_i)$.

For the two-dimensional Gaussian source profile (Equation (19)), the total unlensed flux is then

$$I_{\text{UL}} = 1 - \exp\left(-\frac{\Delta y_s^2}{2\sigma_s^2}\right). \quad (36)$$

We choose the bounding circle radius $\Delta y_S = 3\sigma_S$ so that we retain 99% of the flux. We also set a safe magnification floor $\mu_{\min} = 1.0$.

Our method is a simplified version of the one described in R. B. Metcalf & M. Petkova (2014) in that we do not refine the image further after the source size Δy_S is reached. This is because the Gaussian surface-brightness profile we use is smooth, and sufficient accuracy for the magnification factor is achievable without further refining.

We now determine whether multiple images of a given source are unresolved using the condition defined at the end of Section 3.1. First, we sort groups of adjacent rays into images. Since the rays $\{\mathbf{x}_i\}$ are on a parallelogram mesh, albeit at a finer resolution than that defined in Equation (33), we define adjacency using the refined mesh. A depth-first search is carried out to find the N_J adjacent groups of rays $\{\mathbf{x}_i | i \in J_j\}$, where J_j with $j = 1, 2, \dots, N_J$ are sets of ray indices.

Second, we find the photometric center

$$\mathbf{c}_{J_j} \equiv \langle \mathbf{x}_i \rangle_{J_j} \quad (37)$$

of the image J_j and perform a single-linkage clustering with threshold $\mathcal{R}$ on the centers $\{\mathbf{c}_{J_j}\}$ to find N_K image clusters $\{\{\mathbf{x}_i | i \in J_j\} | j \in K_k\}$, where K_k with $k = 1, 2, \dots, N_K$ are sets of images. Here we have defined the shorthand notation

$$\langle f_i \rangle_{J_j} \equiv \frac{\sum_{i \in J_j} b_i f_i}{\sum_{i \in J_j} b_i}. \quad (38)$$

Finally, the magnification can be computed from

$$\mu_{K_k} = \frac{1}{I_{\text{UL}}} \sum_{j \in K_k} \sum_{i \in J_j} b_i S_{\text{cell}} \quad (39)$$

of image cluster K_k , where S_{cell} is the area of the refined cell. We sort the image cluster so that $\mu_{K_1} \geq \mu_{K_2} \geq \dots$ when $N_K \geq 2$. We also compute

$$l_{K_k} \equiv [\varrho(\Sigma_{K_k})]^{1/2} \quad (40)$$

as an estimation for the size of image cluster K_k , where ϱ is the spectral-radius operator that returns the maximum of the absolute values of the eigenvalues of the photometric-moment matrix

$$\Sigma_{K_k} \equiv \langle (\mathbf{x}_i - \mathbf{c}_{K_k})(\mathbf{x}_i - \mathbf{c}_{K_k}) \rangle_{K_k}. \quad (41)$$

This estimator is designed such that it yields the correct value Δx when there are only two rays, with the same surface brightness, separated by Δx . Here we have introduced the shorthand notation

$$\langle f_i \rangle_{K_k} \equiv \frac{\sum_{j \in K_k} \sum_{i \in J_j} b_i f_i}{\sum_{j \in K_k} \sum_{i \in J_j} b_i} \quad (42)$$

and defined the photometric center

$$\mathbf{c}_{K_k} \equiv \langle \mathbf{x}_i \rangle_{K_k}. \quad (43)$$

Following these calculations, we discard the source if the photometric center $\mathbf{c}_{K_1}$ of the brightest image cluster is in the outer 30% margin of the ray-shooting mesh. This is to avoid the situation in which some lensed images are clipped off the simulation domain. After this selection, the following three properties are studied for each allowed source:

Consistency with Earendel. The source is deemed consistent with Earendel if either (i) $N_K = 1$ and $l_{K_1} < \mathcal{R}$ or (ii) $N_K \geq 2$ but $\mu_{K_1}/\mu_{K_2} > r_{\text{thres}}$ and $l_{K_1} < \mathcal{R}$. Here $r_{\text{thres}} = 10$ is the minimum magnification ratio we choose to rule out a second detectable image cluster elsewhere.

Visible Proper Size. We compute the visible proper size σ_V via the definition in Equation (26) by using the rays $\{\mathbf{y}_i\}$ as samples to estimate the integral. Since the rays $\{\mathbf{x}_i\}$ are uniformly sampled in the image plane, each ray $\mathbf{y}_i$ should be downweighted by the magnification $\mu(\mathbf{x}_i)$.

Astrometry of the Photometric Center. For the unresolved sources, we use the photometric center of the brightest image cluster to compute the astrometric deflection,

$$\Delta c = \mathbf{c}_{K_1} \cdot \left[\mathbf{e}_{\parallel}^0 + \left(\frac{\mathbf{D}^0 \cdot \mathbf{e}_{\perp}^0}{\mathbf{D}^0 \cdot \mathbf{e}_{\parallel}^0} \right) \mathbf{e}_{\perp}^0 \right]. \quad (44)$$

5. Results

We now apply the methods introduced in Sections 4.1 and 4.2 to study the effect of subhalos on the interpretation of Earendel.

First, we consider three subhalo population scenarios with the minimum infall masses $m_{\min} = \{10^6, 10^7, 10^8\} M_{\odot}$ but a common maximum infall mass $m_{\max} = 10^9 M_{\odot}$. Subhalos with $m_{\min} > 10^9 M_{\odot}$ have a low surface number density on the lens plane such that possibly Earendel is not affected by any such massive subhalos. Smaller subhalos are expected to be sufficiently numerous that the macrocritical curve is perturbed by them everywhere (L. Dai et al. 2018; L. L. R. Williams et al. 2024). Including subhalos with $m_{\min} < 10^6 M_{\odot}$ would be computationally expensive, but they may not imprint significant observable effects in lensing. These subhalos are orders of magnitude less massive than the $m_{\min} = 10^6 - 10^8 M_{\odot}$ ones but not orders of magnitude smaller in size according to the mass-concentration relation predicted by the halo collapse theory. Because of their small convergence and shear, these subhalos are inefficient in creating additional large caustics, and may have limited effects in relaxing the source-size constraint for Earendel. Neglecting those very small subhalos in our numerical calculations is therefore a conservative choice. For the three minimum infall masses $m_{\min} = \{10^6, 10^7, 10^8\} M_{\odot}$ that we consider, the subhalo model produces the mean subhalo convergences $\kappa_{\text{SH}} = \{11.6, 8.6, 4.8\} \times 10^{-3}$, respectively.

The image-plane magnification map for random realizations of these scenarios are shown in Figure 1 on a parallelogram domain with the size $(L_1, L_2) = (2'', 2'')$. We sample the source center by randomly selecting a point on the caustic and randomly move the point along the degenerate direction $\mathbf{e}_{\parallel}$ with an amount uniformly distributed in $(-10\sigma_S, +10\sigma_S)$. We perform ray shooting to find the lensed images and determine if it is consistent with the observation of Earendel. We also compute the size for the visible portion of the source r_V (described in Section 3.1) and the astrometric effect Δc (described in Section 3.2). We study four different source sizes $r_S = \{0.1, 0.3, 1.0, 3.0\}$ pc, and set an astrometric resolution $\mathcal{R} = 0.''055$ (B. Welch et al. 2022a) and a magnification-ratio cut $r_{\text{thres}} = 10$.

Figure 2 shows a comparison between the subfold approximation and the exact numerical methods. In the reference case with only a smooth fold, the two methods yield very close results. When numerous low-mass subhalos are included, with

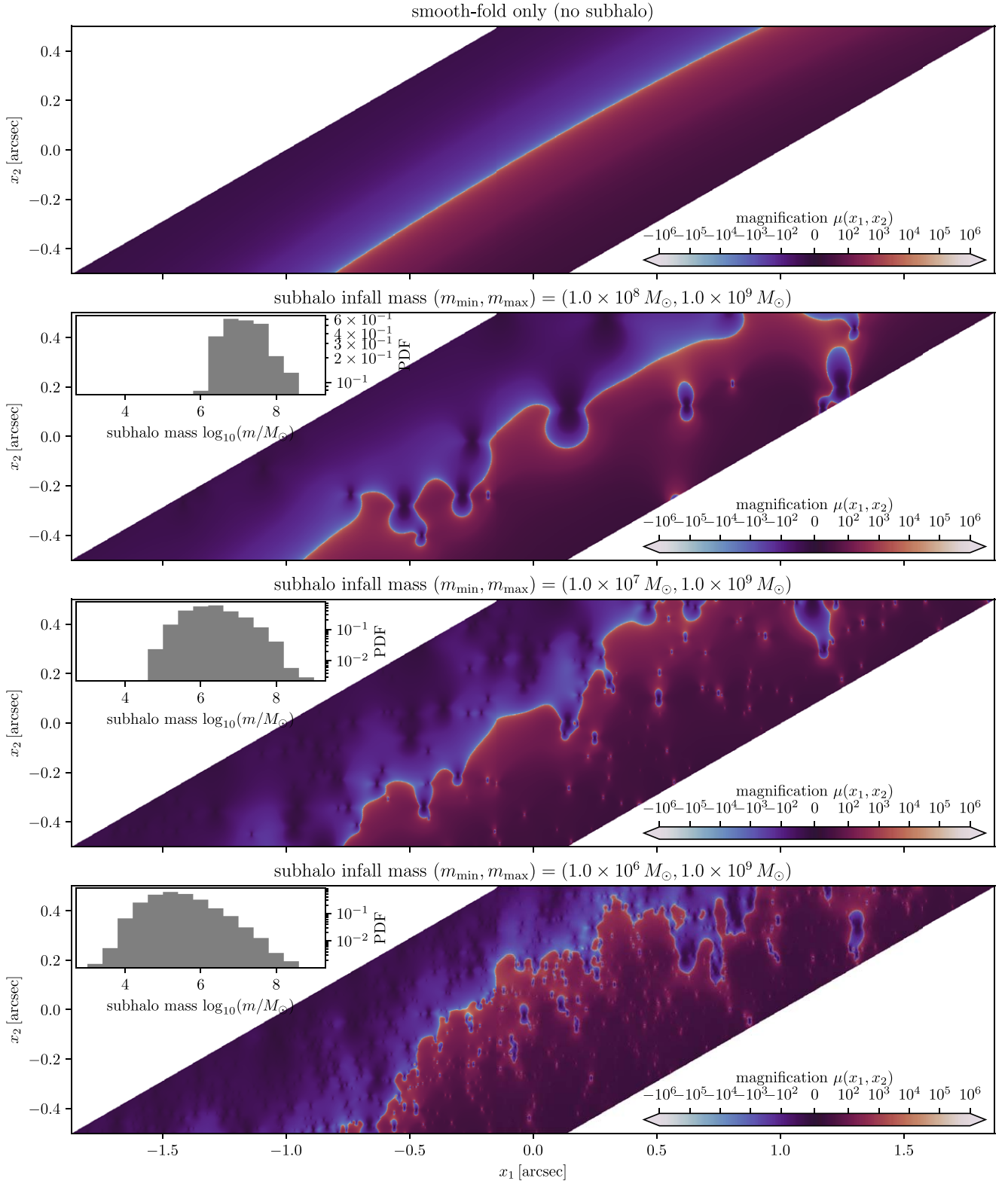


Figure 1. Magnification map for random realizations of subhalos under different subhalo population scenarios in the vicinity of Earendel where the macrocritical curve intersects the Sunrise Arc. The top panel shows the case where there is only a single smooth fold, as assumed by B. Welch et al. (2022a). The rest, from top to bottom, show the cases where the subhalo *infall* masses are between $m_{\min} = \{10^8, 10^7, 10^6\} M_{\odot}$ and $m_{\max} = 10^9 M_{\odot}$. Within each panel, the inset histogram shows the (*final*, after tidal disruption) subhalo-mass distribution. Subhalos disturb the shape of the smooth critical curve, and change the local strength of the fold.

$m_{\min} = 10^6 M_{\odot}$, the exact numerical method finds less sources to have a single unresolved image. This is because the exact numerical method uses the full lens mapping and a more realistic procedure in determining whether the set of images

will appear as a single unresolved image consistent with the observation of Earendel.

Since the exact numerical method is a more accurate treatment for the lensing effect of subhalos, next we only

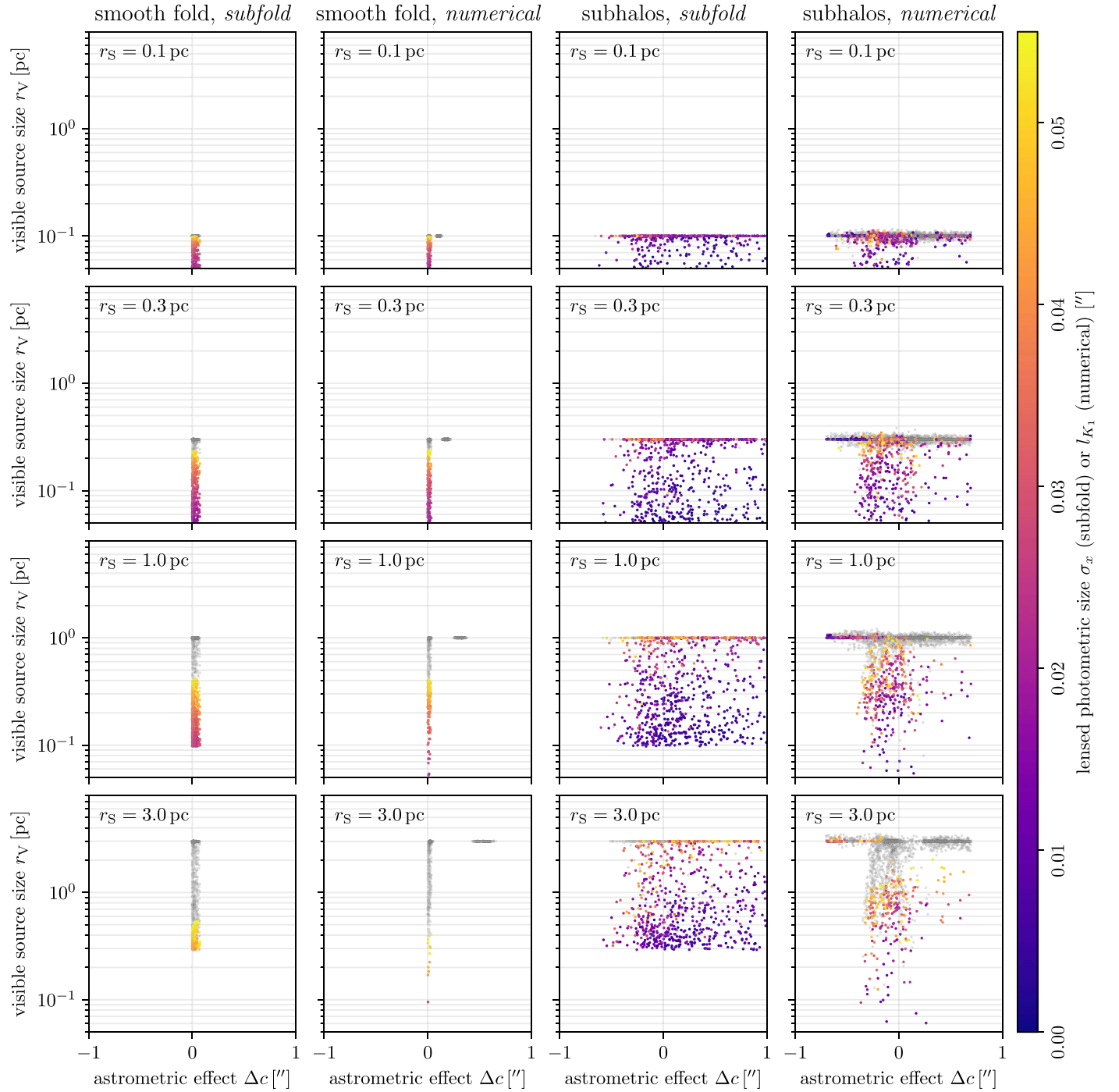


Figure 2. Comparison between the subfold approximation and the exact numerical method on the possible source size of Earendel. The left and the right two columns are for the case with and without the presence of the subhalos, respectively. In the subhalo case, the infall mass is chosen in the range $(m_{\min}, m_{\max}) = (10^6, 10^9) M_{\odot}$. In each case, the first and the second columns are obtained via the subfold approximation and the exact numerical method, respectively. Each panel shows, with color coding, the visible source size r_v (defined in Equation (26)) vs. the astrometric effect Δc (defined in Equations (35) and (44) respectively for the two methods). We choose sources with sizes $\sigma_s = \{0.1, 0.3, 1.0, 3.0\}$ pc in each row. Sources compatible with Earendel are color coded by their lensed photometric size σ_x (defined in Equation (22)) or l_{K1} (defined in Equation (40)) respectively for the two methods. Sources incompatible with Earendel are colored in gray. In the presence of subhalos, the exact numerical method is needed to obtain accurate results, while the subfold method provides a fair approximation. The presence of subhalos relaxes the constraint on the visible size r_v by a factor of a few to 10.

discuss results derived from the exact numerical method. Figure 3 compares results derived from the exact numerical method for three different values of the minimum subhalo infall masses $m_{\min} = \{10^6, 10^7, 10^8\} M_{\odot}$. Figure 4 is the same as Figure 3, but with the subhalo abundance reduced to 25% of the fiducial value. For these scenarios, the mean subhalo convergence is then 25% of that of the fiducial model.

These results show that the constraints to the visible size r_v of Earendel is relaxed by a factor of a few to 10 when the effect of subhalos is accounted for. This conclusion is robust when the infall mass cutoff $m_{\min}$ or the overall subhalo abundance are varied. In particular, the visible portion of Earendel can be larger than ~ 1 pc, thus allowing the interpretation of Earendel being a compact star cluster. For example, there are many star

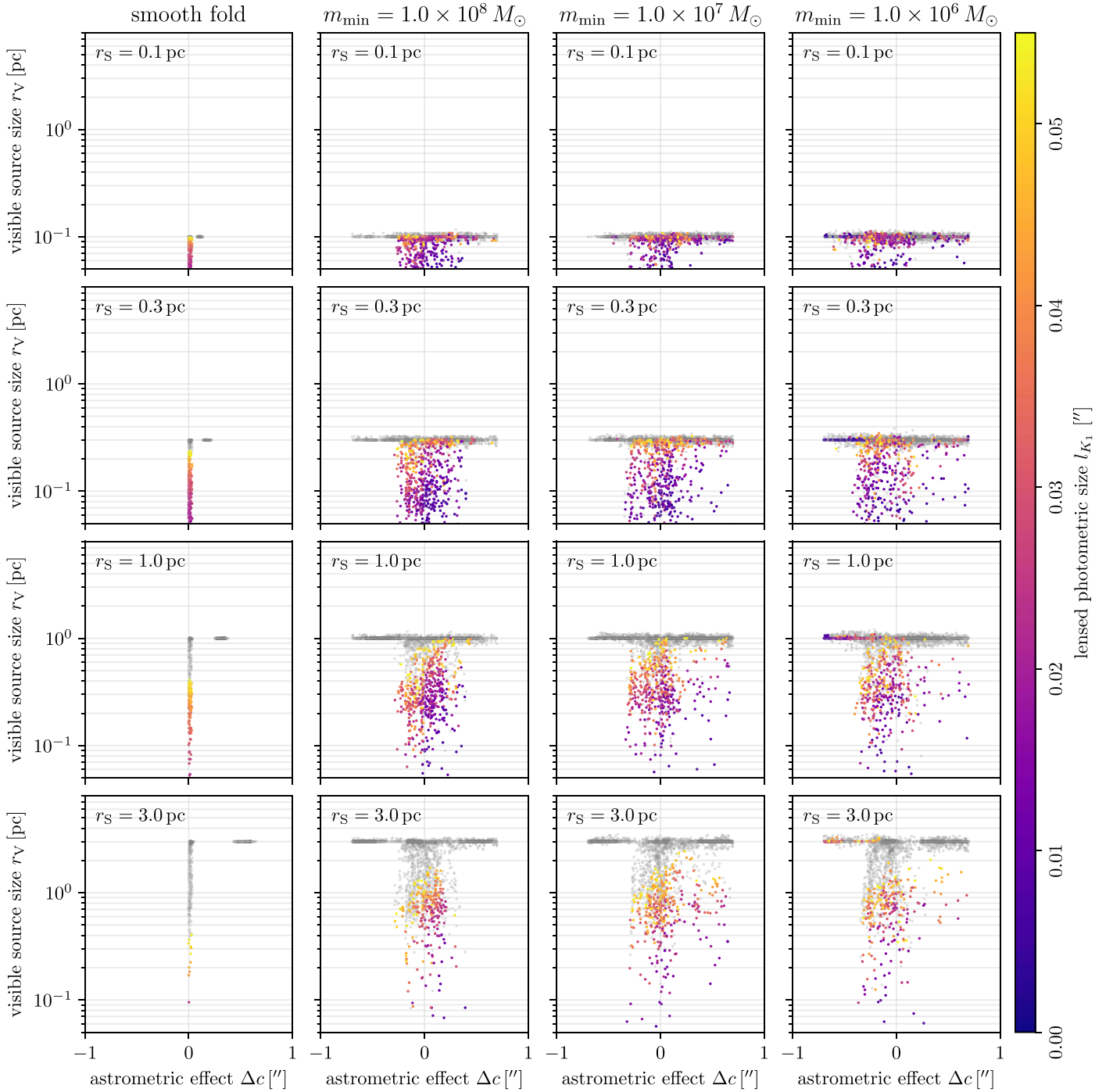


Figure 3. Results derived using the exact numerical method for different subhalo population scenarios. In the second to the fourth columns are the subhalo scenarios with minimum infall masses $m_{\min} = \{10^6, 10^7, 10^8\} M_\odot$ and a maximum infall mass $m_{\max} = 10^9 M_\odot$. In the first column, the smooth-fold-only scenario is plotted for reference. All other elements of this figure are the same as those in Figure 2. The constraint on the visible size r_V of Earendel is not very sensitive to the minimum cutoff $m_{\min}$ of the infall mass.

clusters in local spiral galaxies with a half-light radius $\sim 1\text{--}2$ pc (Figure 9 of M. R. Krumholz et al. 2019).

We also found that subhalos can perturb the astrometry of the lone image so that it does not sit exactly atop the smooth macrocritical curve. However, we find that this astrometric shift is typically $\lesssim 0.5''$ and has a distribution that peaks at $\Delta c = 0''$. Since it is highly expected from cold DM (CDM) that the population of subgalactic subhalos exists, we conclude that the source-size constraint alone cannot rule out star cluster as the possible underlying nature of Earendel.

6. Discussion

So far we studied how DM subhalos, a prediction of the vanilla CDM model, will relax the constraints to the physical size of Earendel. However, other particle physics models that produce small-scale DM clustering will have similar effects.

A notable model of this kind is ultralight boson DM. With a suitable boson mass m_a , it can produce de Broglie interference patterns on a scale similar to the subhalos considered here. The

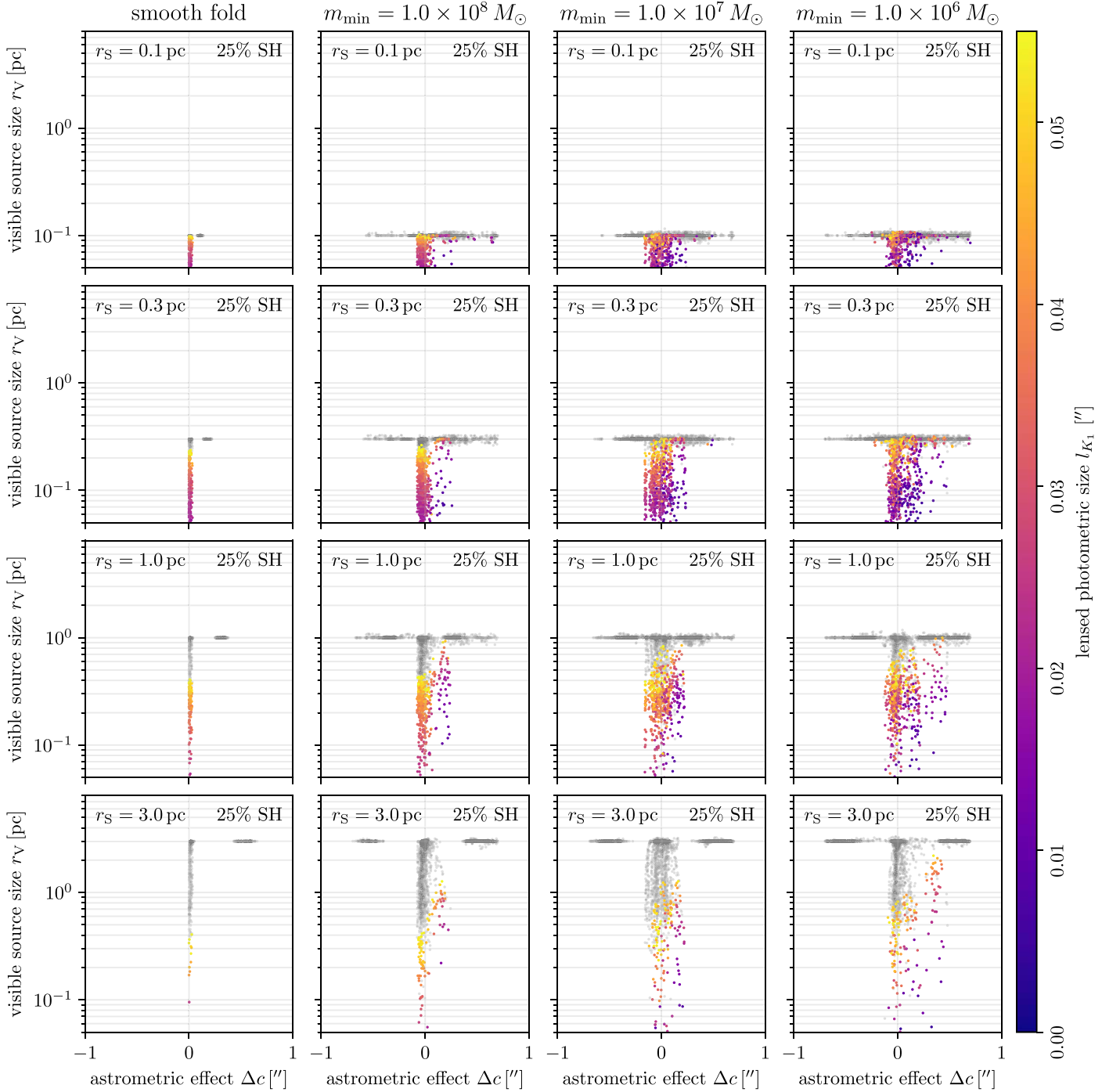


Figure 4. Same as Figure 3, but with the subhalo abundance reduced to 25% of the fiducial model. The constraint on the visible size r_V of Earendel does not significantly tighten with this reduction of the subhalo abundance.

de Broglie wavelength of the DM particle is

$$\lambda_{\text{dB}} = \frac{h}{mv} = 12 \text{ pc} \left(\frac{10^{-21} \text{ eV}}{m} \right) \left(\frac{1000 \text{ km s}^{-1}}{v} \right), \quad (45)$$

where h is the Planck constant and v is the typical virial velocity of the DM particles inside the galaxy-cluster halo. The $10^5 M_\odot$ subhalo that we consider has a scale radius $r_S \sim 80$ pc, which is several times larger than λ_{dB} for a boson with mass $m_a \simeq 10^{-21}$ eV. For similar boson masses, the effect of interference on lensing in the vicinity of macrocritical curves

has been previously investigated (A. Amruth et al. 2023; T. Broadhurst et al. 2024; J. M. Diego et al. 2024b).

However, if the boson mass is much larger, the length scale of the interference pattern is small, and a large number of small wave-interference structures integrated along the line of sight will reduce both the magnitude and coherence length of the lens surface density fluctuations. Recently, tight constraints on ultralight boson mass $m_a \gtrsim 10^{-19}$ eV have been derived from the study of stellar dynamics in ultrafaint dwarf galaxies (D. J. E. Marsh & J. C. Niemeyer 2019; N. Dalal & A. Kravtsov 2022). Since each λ_{dB} patch has order-unity

density fluctuation, the convergence fluctuation $\delta\kappa$ over distance L is roughly $\delta\kappa \sim (\lambda_{\text{dB}}/L)^{1/2}$, giving

$$\delta\kappa \sim 10^{-3} \left(\frac{m}{10^{-19} \text{ eV}} \right)^{-\frac{1}{2}} \left(\frac{v}{10^3 \text{ km s}^{-1}} \right)^{-\frac{1}{2}} \left(\frac{L}{100 \text{ kpc}} \right)^{-\frac{1}{2}} \quad (46)$$

for a line-of-sight projection across $L \sim 100 \text{ kpc}$ of the dense part of the galaxy-cluster DM halo. If the boson mass bounds quoted above are valid, ultralight DM may be relevant on tiny angular scales analogous to and subdominant to intracluster microlensing, but is not expected to generate the same larger-scale, persistent effect from CDM subhalos as we have studied here.

7. Conclusion

In this paper, we studied the effect of DM subhalos on interpreting highly magnified sources on lensing caustics. This is a subgalactic population with subhalo masses in the range 10^6 – $10^9 M_\odot$ that are expected to arise from hierarchical structure formation in the cold DM model but are thought to lack the stellar component. We consider the particular example of the highly magnified source Earendel, discovered in a lensed galaxy at $z_s = 6.2$ as a case study, as it has been suspected to be an extremely magnified single or multiple star system but a crucial constraint on the underlying source size was previously derived based on lensing analysis neglecting the presence of subgalactic subhalos.

We have recreated the smooth lens model in the vicinity of Earendel as reported in B. Welch et al. (2022a) and have reproduced the previously reported tight size constraint. We have clarified that the size constraint is actually on the visible portion of the source when the source overlaps the caustic, which is not an improbable corner case for large sources. We then embedded in the smooth lens model randomly generated subhalos according to calibrated semianalytic models of the subhalo internal structure and population. Such lensing situations with random realizations of subhalos are then numerically examined under the semianalytical subfold approximation as well as with the exact numerical ray-shooting method.

We find that the presence of subhalos relaxes the constraint on the size of the visible portion of Earendel by a factor of a few to 10 compared to that derived for the smooth lens model. This aligns with the intuition that small-scale perturber lenses weaken and disrupt lensing caustics despite their small contributions to convergence and shear. Our results imply that the physical size of Earendel may be larger than $\sim 1 \text{ pc}$ without contradicting observation, making it a viable possibility that Earendel is a compact star cluster with a large number of stars. We point out that subhalos can also astrometrically perturb the lensed image of Earendel so that it may not necessarily sit exactly atop the smooth critical curve. However, we find that the deviation is typically smaller than $\sim 0.5''$ and having nearly zero astrometric departure is a probable realization. Our findings are robust to moderate changes in the assumed subhalo population, such as the abundance of subhalos or the lower cutoff on the subhalo mass.

Based on JWST NIRCам photometry, B. Welch et al. (2022b) found that if Earendel is visually dominated by a single star a cool supergiant with $T_{\text{eff}} = 15,000 \text{ K}$ would be the best

fit. However, the authors noted that the single-star SED fit had significant discrepancy with data and explored a scenario that Earendel is a binary consisting of one cool supergiant and one O star. The challenge with the binary interpretation is that the hot O star is required to have a higher bolometric luminosity, which is at odds with the expectation that the more massive binary component evolves off the main sequence faster (see A. Nabizadeh et al. 2025 for possible solutions to this problem). This difficulty, together with the nondetection of microlensing-induced time variability, may imply instead that Earendel's light is dominated by many stars, but we note that SED fitting to star-cluster models has not been presented in the literature and would be a useful investigation toward unveiling Earendel's true nature.

The idea and methodology of this work can be generalized to a few other recently reported highly magnified lone images atop critical curves in other lensed galaxies (J. M. Diego et al. 2022; L. J. Furtak et al. 2023). We expect that such analysis will impact the interpretation of the nature of these sources as source-size constraints may have to be revised. Analyses of more similar highly magnified sources in the future should account for the possible effect of DM subhalos when applying the smooth lens model to deducing de-lensed properties. The time-varying microlensing effect of the intracluster stars, which is inevitable and can have an observable effect on both clusters of stars or individual stars, is a complementary approach to constraining de-lensed source properties.

Acknowledgments

The authors would like to thank Shude Mao, Christopher McKee, Massimo Pascale, and Wenrui Xu for useful discussions. L.D. acknowledges research grant support from the Alfred P. Sloan Foundation (Award Number FG-2021-16495), and support of Frank and Karen Dabby STEM Fund in the Society of Hellman Fellows. L.D. is also supported by the Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Award Number DE-SC-0025293.

Appendix A Details of the Lens Models

We are in the regime of a single lens plane where all deflections of light rays happens in a region whose thickness is negligible compared to the angular diameter distances D_s , D_L , and D_{LS} involved. In this case, the total deflection $\alpha(\mathbf{x})$ in the ray equation $\mathbf{y}(\mathbf{x}) = \mathbf{x} - \alpha(\mathbf{x})$ is simply the sum of deflections caused by individual lenses,

$$\alpha(\mathbf{x}) = \alpha_{\text{FD}}(\mathbf{x}) + \sum_{n=1}^N \alpha_{\text{SH}}^{(n)}(\mathbf{x}) + \alpha_{\text{UD}}(\mathbf{x}), \quad (\text{A1})$$

where $\alpha_{\text{FD}}(\mathbf{x})$, $\alpha_{\text{SH}}^{(n)}(\mathbf{x})$, and $\alpha_{\text{UD}}(\mathbf{x})$ are the deflections from the smooth-fold catastrophe, from the n th subhalo, and from the compensating uniform disk, respectively. To evaluate the Jacobian $\partial y_i / \partial x_j(\mathbf{x}) = \delta_{ij} - \partial \alpha_i / \partial x_j(\mathbf{x})$, we need the first-order derivatives of the individual deflections (which are the second-order derivatives of the Fermat potential),

$$\frac{\partial \alpha_i}{\partial x_j}(\mathbf{x}) = \frac{\partial \alpha_{i,\text{FD}}}{\partial x_j}(\mathbf{x}) + \sum_{n=1}^N \frac{\partial \alpha_{i,\text{SH}}^{(n)}}{\partial x_j}(\mathbf{x}) + \frac{\partial \alpha_{i,\text{UD}}}{\partial x_j}(\mathbf{x}), \quad (\text{A2})$$

and the second-order derivatives as well (which are the third-order derivatives of the Fermat potential),

$$\frac{\partial^2 \alpha_i}{\partial x_j \partial x_k}(\mathbf{x}) = \frac{\partial^2 \alpha_{i,\text{FD}}}{\partial x_j \partial x_k}(\mathbf{x}) + \sum_{n=1}^N \frac{\partial^2 \alpha_{i,\text{SH}}^{(n)}}{\partial x_j \partial x_k}(\mathbf{x}) + \frac{\partial^2 \alpha_{i,\text{UD}}}{\partial x_j \partial x_k}(\mathbf{x}). \quad (\text{A3})$$

$$A_{ij}(\mathbf{x}) \equiv \frac{\partial y_i}{\partial x_j}(\mathbf{x}) = \phi_{ij}^0 + \phi_{ijk}^0 x_k + \mathcal{O}(|\mathbf{x}|^2) = \begin{pmatrix} \phi_{11}^0 + \phi_{111}^0 x_1 + \phi_{112}^0 x_2 & \phi_{12}^0 + \phi_{121}^0 x_1 + \phi_{122}^0 x_2 \\ \phi_{21}^0 + \phi_{211}^0 x_1 + \phi_{212}^0 x_2 & \phi_{22}^0 + \phi_{221}^0 x_1 + \phi_{222}^0 x_2 \end{pmatrix} + \mathcal{O}(|\mathbf{x}|^2), \quad (\text{A10})$$

In what follows, we collect the formulas we use for each lens model.

A.1. Fold Catastrophe

We start by expanding the lensing potential $\psi(\mathbf{x})$ to cubic order at the image-plane origin $\mathbf{x} = \mathbf{0}$, which is on the critical curve:

$$\psi(\mathbf{x}) = \psi^0 + \psi_i^0 x_i + \frac{1}{2!} \psi_{ij}^0 x_i x_j + \frac{1}{3!} \psi_{ijk}^0 x_i x_j x_k + \mathcal{O}(|\mathbf{x}|^4). \quad (\text{A4})$$

Here, we have adopted the Einstein summation convention and the shorthand notation for derivatives $\psi_{ij\dots} \equiv \partial\psi/(\partial x_i \partial x_j \dots)$. The superscript $\mathbf{0}$ indicates that the quantities are evaluated at the origin $\mathbf{x} = \mathbf{0}$. The resulting deflection is

$$\alpha_{i,\text{FD}}(\mathbf{x}) = \psi_i(\mathbf{x}) = \psi_i^0 + \psi_{ij}^0 x_j + \frac{1}{2!} \psi_{ijk}^0 x_j x_k + \mathcal{O}(|\mathbf{x}|^3). \quad (\text{A5})$$

Without loss of generality, we require that the lens-plane origin $\mathbf{x} = \mathbf{0}$ is mapped to the source-plane origin $\mathbf{y} = \mathbf{0}$, i.e., $\alpha(\mathbf{0}) = \mathbf{0}$, for which we set $\psi_i^0 = 0$. We define the Fermat potential $\phi(\mathbf{x}, \mathbf{y}) \equiv (\mathbf{x} - \mathbf{y})^2/2 - \psi(\mathbf{x})$, and switch to using derivatives of the Fermat potential $\phi_{ij}^0 = \delta_{ij} - \psi_{ij}^0$ and $\phi_{ijk}^0 = -\psi_{ijk}^0$, instead of those of the lensing potential, to characterize the fold. Doing so, we have

$$\alpha_{i,\text{FD}}(\mathbf{x}) = x_i - \phi_{ij}^0 x_j - \frac{1}{2!} \phi_{ijk}^0 x_j x_k + \mathcal{O}(|\mathbf{x}|^3). \quad (\text{A6})$$

The requirement that the origin $\mathbf{x} = \mathbf{0}$ is on the fold implies that $\det A(\mathbf{0}) = 0$, thus $\phi_{11}^0 \phi_{22}^0 - (\phi_{12}^0)^2 = 0$. The Jacobian matrix is

$$\frac{\partial \alpha_{i,\text{FD}}}{\partial x_j}(\mathbf{x}) = \delta_{ij} - \phi_{ij}^0 - \phi_{ijk}^0 x_k + \mathcal{O}(|\mathbf{x}|^2). \quad (\text{A7})$$

The second derivative of the deflection is

$$\frac{\partial^2 \alpha_{i,\text{FD}}}{\partial x_j \partial x_k}(\mathbf{x}) = -\phi_{ijk}^0 + \mathcal{O}(|\mathbf{x}|). \quad (\text{A8})$$

A single fold has the lens map

$$y_i(\mathbf{x}) = x_i - \alpha_{i,\text{FD}}(\mathbf{x}) = \phi_{ij}^0 x_j + \frac{1}{2!} \phi_{ijk}^0 x_j x_k + \mathcal{O}(|\mathbf{x}|^3). \quad (\text{A9})$$

The Jacobian matrix is

with the determinant (accurate to the leading order in $\mathbf{x}$)

$$\det A(\mathbf{x}) = \mathbf{x} \cdot \mathbf{D}^0 + \mathcal{O}(|\mathbf{x}|^2), \quad \text{where} \\ \times D_i^0 = \phi_{22}^0 \phi_{11i}^0 + \phi_{11}^0 \phi_{22i}^0 - 2 \phi_{12}^0 \phi_{12i}^0. \quad (\text{A11})$$

A fold is therefore characterized by the independent components of ϕ_{ij}^0 (i.e., ϕ_{11}^0 , ϕ_{12}^0 , and ϕ_{22}^0) and ϕ_{ijk}^0 (i.e., ϕ_{111}^0 , ϕ_{112}^0 , ϕ_{122}^0 , ϕ_{222}^0), a total of seven numbers, minus the singularity constraint $\phi_{11}^0 \phi_{22}^0 - (\phi_{12}^0)^2 = 0$. The magnification $\mu(\mathbf{x}) \equiv 1/\det A(\mathbf{x})$ is

$$\mu(\mathbf{x}) = \frac{1}{\mathbf{x} \cdot \mathbf{D}^0} + \mathcal{O}\left(\frac{1}{|\mathbf{x}|^2}\right). \quad (\text{A12})$$

A fold is defined by a few parameters that depend only on the local property of the lens (P. Schneider et al. 1992; J. Wagner 2019; J. Lin et al. 2022; J. Wagner 2022).

A.2. Abundance of the Subhalos

N -body simulations suggest that the mass and spatial distributions of subhalos follow simple universal laws independent of the host-halo mass. Here, we follow the semianalytical model developed in J. Han et al. (2016), which is based on three simple assumptions about subhalo formation. First, the infall mass m_{acc} of a subhalo, before any tidal stripping, follows a power-law distribution, $dN(m_{\text{acc}})/d \ln m_{\text{acc}} \propto m_{\text{acc}}^{-\alpha}$. Second, the spatial distribution of subhalos at a given infall mass m_{acc} follows the density profile $\rho(R)$ of the host halo, $dN(R|m_{\text{acc}})/d^3R \propto \rho(R)$, where R is the host halo-centric distance and $d^3R = 4\pi R^2 dR$ is the volume element. Third, tidal forces gradually strip a subhalo of infall mass m_{acc} either (1) completely apart with a probability $1 - f_s$, or (2) to a reduced final mass $m \propto R^\beta m_{\text{acc}}$ (with a normal scatter σ in $\ln m$) with probability f_s . These three assumptions are individually tested in J. Han et al. (2016) for subhalos with mass $m \gtrsim 10^8 M_\odot$. In this work, we assume that these results can be extrapolated to lower subhalo masses.

Together, the joint distribution of subhalo mass and position is (Equation (10) of J. Han et al. 2016, but a minus sign is missing in front of the power-law index α in that equation)

$$\frac{dN(m, R)}{d \ln m d^3R} = A_{\text{acc}} f_s \frac{\rho(R)}{m_0} \left[\frac{m}{m_0 \bar{\mu}(R)} \right]^{-\alpha} \frac{e^{\sigma^2 \alpha^2/2}}{2} \times \left\{ \text{erf} \left[\frac{\ln \frac{\bar{\mu}(R)}{\mu_{\min}(m)} + \alpha \sigma^2}{\sqrt{2} \sigma} \right] - \text{erf} \left[\frac{\ln \frac{\bar{\mu}(R)}{\mu_{\max}(m)} + \alpha \sigma^2}{\sqrt{2} \sigma} \right] \right\}, \quad (\text{A13})$$

where A_{acc} is a dimensionless normalization constant and $m_0 \equiv 10^{10} M_{\odot}/h$ is the mass scale of normalization. Here $\mu \equiv m/m_{\text{acc}}$ is the fractional mass that remains in the subhalo after tidal stripping, so that the third assumption takes the form $\bar{\mu}(R) \equiv \mu_{\star} (R/R_{200})^{\beta}$ with a normalization constant $\mu_{\star}$. Similarly, $\mu_{\text{min}}(m) \equiv m/m_{\text{max}}$ and $\mu_{\text{max}}(m) \equiv m/m_{\text{min}}$, where m_{min} and m_{max} are the minimum and maximum cutoff in the power-law distribution of the infall mass m_{acc} , respectively. In this work, for a galaxy-cluster-sized host halo, we adopt the fiducial values $A_{\text{acc}} = 0.08$, $\alpha = 0.9$, $f_s = 0.56$, $\mu_{\star} = 0.34$, $\beta = 1.0$, and $\sigma = 1.1$ (J. Han et al. 2016). We choose to fix $m_{\text{max}} = 10^9 M_{\odot}$, and explore the effect of varying m_{min} .

A.3. Structure of the Subhalos

We adopt the smoothly truncated NFW density profile for subhalos

$$\rho_{\text{SH}}(r) = \frac{m_{200}}{4\pi r_s^3 f(c_{200})} \frac{1}{(r/r_s)(1+r/r_s)^2} \frac{1}{1+r^2/r_t^2}, \quad (\text{A14})$$

parameterized by the mass m_{200} , the concentration parameter c_{200} of the underlying untruncated NFW halo, and the

$$H_1(t, \tau) \equiv \frac{\tau^2}{2(1+\tau^2)^2} \times \left[-\frac{2t}{1+t}(1+\tau^2) + 4\tau \arctan \frac{t}{\tau} + 2(\tau^2-1) \ln[(1+t)\tau] - (\tau^2-1) \ln(t^2+\tau^2) \right]. \quad (\text{A19})$$

truncation radius r_t . Here, $f(x) \equiv \ln(1+x) - x/(1+x)$, and the scale radius r_s of the untruncated NFW halo is given in terms of m_{200} and c_{200} as

$$r_s = \frac{1}{c_{200}} \left(\frac{3}{4\pi} \frac{m_{200}}{200\rho_{\text{crit}}} \right)^{1/3}. \quad (\text{A15})$$

Note that the total mass m of the halo differs from m_{200} . It is given by

$$m \equiv \int_0^{+\infty} 4\pi r^2 dr \rho_{\text{SH}}(r) = \frac{m_{200}}{f(c_{200})} \frac{\tau^2}{(\tau^2+1)^2} [(\tau^2-1)\ln\tau + (\pi-\tau)\tau - 1], \quad (\text{A16})$$

where $\tau \equiv r_t/r_s$. A subhalo centered at $\mathbf{x}=\mathbf{x}_0$ causes a ray deflection

$$\alpha_{\text{SH}}(\mathbf{x}) = \frac{4Gm_{200}}{c^2 f(c_{200})} \frac{D_{LS}}{r_s} \frac{\mathbf{r}}{D_S |\mathbf{r}|} S\left(\frac{D_L}{r_s} |\mathbf{r}|, \frac{r_t}{r_s}\right), \quad (\text{A17})$$

where we have defined $\mathbf{r} \equiv \mathbf{x} - \mathbf{x}_0$ and the dimensionless function

$$S(\xi, \tau) \equiv \int_0^{+\infty} \frac{\xi d\eta}{(\eta^2 + \xi^2)^{3/2}} H_1(\sqrt{\eta^2 + \xi^2}, \tau) \quad (\text{A18})$$

with a closed-form expression for $H_1(t, \tau)$

The Jacobian matrix is

$$\frac{\partial \alpha_{i,\text{SH}}}{\partial x_j}(\mathbf{x}) = \frac{4Gm_{200}}{c^2 f(c_{200})} \frac{D_{LS}}{r_s} \frac{D_S}{r} \left[(\delta_{ij} - \hat{r}_i \hat{r}_j) \frac{1}{r} S\left(\frac{D_L}{r_s} r, \frac{r_t}{r_s}\right) + \hat{r}_i \hat{r}_j \frac{D_L}{r_s} S_1\left(\frac{D_L}{r_s} r, \frac{r_t}{r_s}\right) \right], \quad (\text{A20})$$

where we define $S_1(\xi, \tau) \equiv \partial S(\xi, \tau)/\partial \xi$, $r \equiv |\mathbf{r}|$, and $\hat{r} \equiv \mathbf{r}/r$. The second-order derivative of the deflection is

$$\begin{aligned} \frac{\partial^2 \alpha_{i,\text{SH}}}{\partial x_j \partial x_k}(\mathbf{x}) &= \frac{4Gm_{200}}{c^2 f(c_{200})} \frac{D_{LS}}{r_s} \frac{D_S}{r} \\ &\times \left\{ [3\hat{r}_i \hat{r}_j \hat{r}_k - (\delta_{ij} \hat{r}_k + \delta_{jk} \hat{r}_i + \delta_{ki} \hat{r}_j)] \frac{1}{r^2} S\left(\frac{D_L}{r_s} r, \frac{r_t}{r_s}\right) \right. \\ &+ [(\delta_{ij} \hat{r}_k + \delta_{jk} \hat{r}_i + \delta_{ki} \hat{r}_j) - 3\hat{r}_i \hat{r}_j \hat{r}_k] \frac{1}{r} \frac{D_L}{r_s} S_1\left(\frac{D_L}{r_s} r, \frac{r_t}{r_s}\right) \\ &\left. + \hat{r}_i \hat{r}_j \hat{r}_k \left(\frac{D_L}{r_s}\right)^2 S_2\left(\frac{D_L}{r_s} r, \frac{r_t}{r_s}\right) \right\}, \quad (\text{A21}) \end{aligned}$$

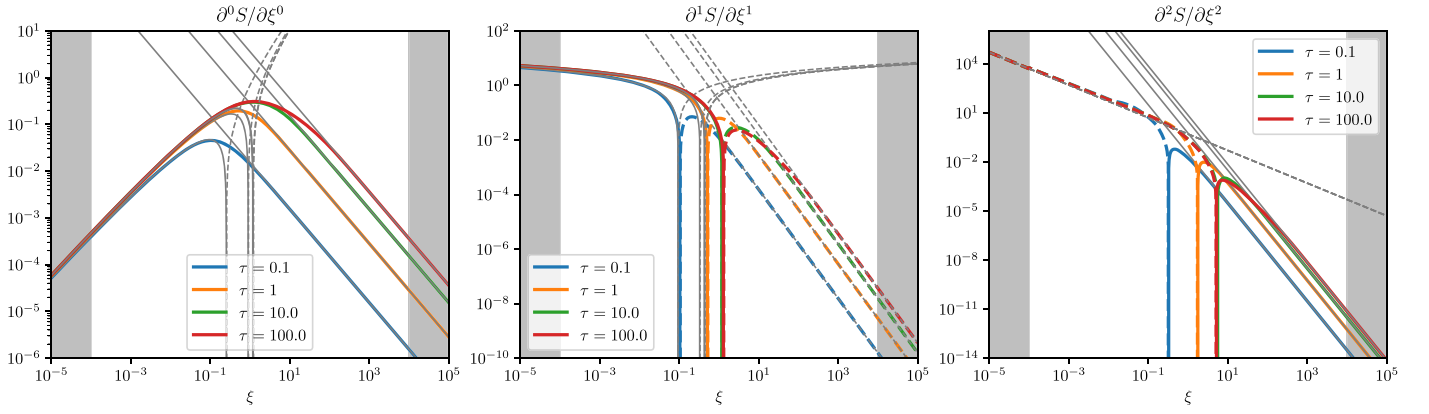


Figure 5. Functions $S(\xi, \tau)$, $S_1(\xi, \tau) \equiv \partial S(\xi, \tau)/\partial \xi$, and $S_2(\xi, \tau) \equiv \partial^2 S(\xi, \tau)/\partial \xi^2$. The colored lines are for different τ values and the gray lines are for the corresponding asymptotic formulas for both $\xi \rightarrow 0$ and $\xi \rightarrow +\infty$. The shaded regions in ξ show where the appropriate asymptotic formulas are used. In between the gray regions, uniform rectangular bivariate spline interpolation in logarithmic space is used. Solid (dashed) lines indicate positive (negative) values.

where we define $S_2(\xi, \tau) \equiv \partial^2 S(\xi, \tau)/\partial \xi^2$.

Although we found analytical expressions for $S(\xi, \tau)$, $S_1(\xi, \tau)$, and $S_2(\xi, \tau)$, they are too lengthy to be computationally efficient. Hence, we proceed with a hybrid scheme of using interpolation and asymptotic formulas. For $(\xi, \tau) \in [10^{-4}, 10^{+4}] \times [10^{-2}, 10^{+3}]$, we tabulate $S(\xi, \tau)$ on a uniform rectangular grid and use bivariate spline interpolation, all in logarithmic space. For values of ξ outside of this range, we directly evaluate the appropriate asymptotic formulas for $\xi \rightarrow 0$ or $\xi \rightarrow +\infty$. In summary, we have

$$S(\xi, \tau) = \begin{cases} \left\{ \frac{1}{2} \ln \frac{2}{\xi} - \frac{1}{4(1+\tau^2)^2} [\tau^4 + 2(\tau^2 - 1) \ln \tau + 2\pi\tau - 1] \right\} \xi, & \text{for } \xi < 10^{-4}, \\ \frac{\tau^2}{(\tau^2 + 1)^2} [(\tau^2 - 1) \ln \tau + (\pi - \tau)\tau - 1] \xi^{-1}, & \text{for } \xi > 10^{+4}, \\ \text{bivariate spline interpolation of } S(\xi, \tau) & \text{otherwise.} \end{cases} \quad (\text{A22})$$

For the derivatives $S_1(\xi, \tau)$ and $S_2(\xi, \tau)$, we use the spline derivatives for $\xi \in [10^{-4}, 10^{+4}]$ and the derivatives of the

The first-order and second-order derivatives of the deflection are

$$\frac{\partial \alpha_{i,\text{UD}}}{\partial x_j}(\mathbf{x}) = \begin{cases} \kappa_0 \delta_{ij} & (x \leq a), \\ \kappa_0 a^2 \frac{1}{x^2} \left(\delta_{ij} - 2 \frac{x_i x_j}{x^2} \right) & (x > a) \end{cases} \quad (\text{A24})$$

and

$$\frac{\partial \alpha_{i,\text{UD}}}{\partial^2 x_j \partial x_k}(\mathbf{x}) = \begin{cases} 0 & (x \leq a), \\ \kappa_0 a^2 \left[\frac{8}{x^6} x_i x_j x_k - \frac{2}{x^4} (\delta_{ij} x_k + \delta_{jk} x_i + \delta_{ki} x_j) \right] & (x > a) \end{cases} \quad (\text{A25})$$

appropriate asymptotic forms outside that range. The accuracy of this method is demonstrated in Figure 5.

A.4. Uniform Circular Disk Lens

For a uniform circular disk with angular radius a and dimensionless surface density κ_0 located at the origin $\mathbf{x} = \mathbf{0}$, the ray deflection is

$$\alpha_{\text{UD}}(\mathbf{x}) = \frac{\mathbf{x}}{x} \times \begin{cases} \kappa_0 \mathbf{x} & (x \leq a), \\ \kappa_0 a^2 / \mathbf{x} & (x > a). \end{cases} \quad (\text{A23})$$

respectively.

Appendix B Detailed Size Constraints

In this appendix, we derive in detail the size constraint given in Section 3.1. First, we explicitly compute the photometric center $\mathbf{C}_x$, the photometric moment Σ_x , and the associated image size σ_x (defined in Equations (21), (20), and (22), respectively); then, we explicitly compute the unlensed photometric center $\mathbf{C}_{V,y}$, the unlensed photometric moment $\Sigma_{V,y}$, and the associated unlensed size σ_v , all for the visible portion of the source (defined in Equations (25), (24), and (26), respectively); finally, we derive the size constraint.

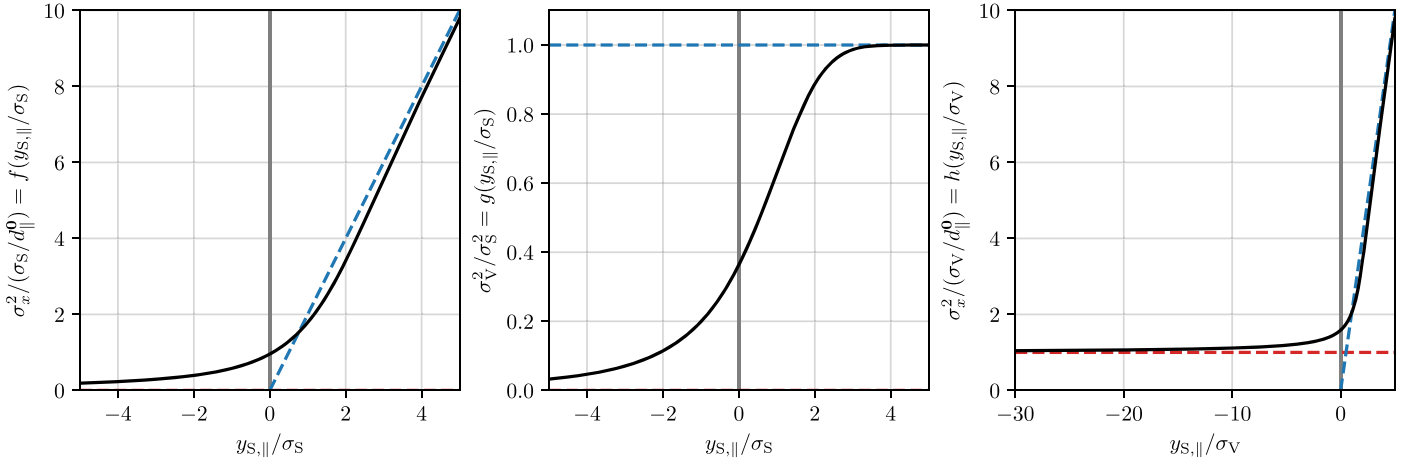


Figure 6. Functions $f(\xi)$, $g(\xi)$, and $h(\eta)$ (from left to right) as defined in Equations (B4), (B7), and (B9), respectively. The left and right asymptotes of each function are shown in red and blue dashed lines, respectively.

Without loss of generality, we choose a coordinate system with basis $(\mathbf{e}_\parallel^0, \mathbf{e}_\perp^0)$. In these coordinates, the lensing Jacobian has a simple form,

$$A(\mathbf{x}) = \begin{bmatrix} \mathbf{d}^0 \cdot \mathbf{x} & 0 \\ 0 & 2(1 - \kappa^0) \end{bmatrix}. \quad (\text{B1})$$

This indicates that the images will be extremely stretched in the $\mathbf{e}_\parallel$ direction, which implies that a two-dimensional Gaussian source will be stretched to effectively a one-dimensional Gaussian. We now focus on the $\mathbf{e}_\parallel$ direction. The lens mapping along the $\mathbf{e}_\parallel^0$ direction is obtained by the line integral

$$\mathbf{y}(\lambda) - \mathbf{y}(0) = \int_0^\lambda A[\mathbf{x}(\lambda')] \frac{d\mathbf{x}(\lambda')}{d\lambda'} d\lambda' \quad (\text{B2})$$

along the image-plane trajectory $\mathbf{x}(\lambda) \equiv \lambda \mathbf{e}_\parallel^0$. With the convention that $\mathbf{y}(0) = \mathbf{0}$, we have the one-dimensional lens mapping along the $\mathbf{e}_\parallel$ direction,

$$y_\parallel = \frac{1}{2} d_\parallel^0 x_\parallel^2, \quad (\text{B3})$$

where $d_\parallel^0 \equiv \mathbf{d}^0 \cdot \mathbf{e}_\parallel^0$. In what follows, we consider the case where $d_\parallel^0 > 0$. The case of $d_\parallel^0 < 0$ is essentially the same by flipping the coordinate basis.

On the image plane, according to Equations (21) and (B3), from symmetry, $\mathbf{C}_x = \mathbf{0}$. Then it follows from Equations (20)–(22) that

$$\sigma_x^2 = \frac{\sigma_s}{d_\parallel^0} f\left(\frac{y_{s,\parallel}}{\sigma_s}\right), \quad (\text{B4})$$

where $f(\xi)$ has the explicit form

with $I_\nu(z)$ and $K_\nu(z)$ being the modified Bessel function of the first and the second kind, respectively. On the source plane, the visible side of the fold is the side in which $y_\parallel \geq 0$. So according to Equation (25), we have

$$\mathbf{C}_{\mathbf{v},\mathbf{y}} = \left(y_{s,\parallel} + \sigma_s \sqrt{\frac{2}{\pi}} \frac{\exp[-y_{s,\parallel}^2 / (2\sigma_s^2)]}{1 + \text{erf}[y_{s,\parallel} / (\sqrt{2}\sigma_s)]} \right) \mathbf{e}_\parallel^0. \quad (\text{B6})$$

Then it follows from Equations (24) and (26) that

$$\sigma_v^2 = \sigma_s^2 g\left(\frac{y_{s,\parallel}}{\sigma_s}\right), \quad (\text{B7})$$

where $g(\xi)$ has the explicit form

$$g(\xi) \equiv 1 - \xi \sqrt{\frac{2}{\pi}} \frac{\exp(-\xi^2/2)}{1 + \text{erf}(\xi/\sqrt{2})} - \frac{2}{\pi} \left[\frac{\exp(-\xi^2/2)}{1 + \text{erf}(\xi/\sqrt{2})} \right]^2. \quad (\text{B8})$$

We now express σ_x in terms of σ_v by combining Equations (B4) and (B7) to eliminate σ_s . Through dimensional analysis, the result must take the form

$$\sigma_x^2 = \frac{\sigma_v}{d_\parallel^0} h\left(\frac{y_{s,\parallel}}{\sigma_v}\right), \quad (\text{B9})$$

where $h(\eta)$ is the dimensionless function to be determined. By using Equation (B7), (B9) can be cast as

$$f(\xi) \equiv \begin{cases} \frac{\xi^2 L_{-1/4}(\xi^2/4) + (\xi^2 + 2) I_{1/4}(\xi^2/4) + \xi^2 [I_{3/4}(\xi^2/4) + I_{5/4}(\xi^2/4)]}{\xi [L_{-1/4}(\xi^2/4) + I_{1/4}(\xi^2/4)]} & (\xi \geq 0), \\ \frac{\pi}{\sqrt{2}} \cdot \frac{\xi^2 L_{-1/4}(\xi^2/4) - (\xi^2 + 2) I_{1/4}(\xi^2/4) + \xi^2 [I_{3/4}(\xi^2/4) - I_{5/4}(\xi^2/4)]}{\xi K_{1/4}(\xi^2/4)} & (\xi < 0), \end{cases} \quad (\text{B5})$$

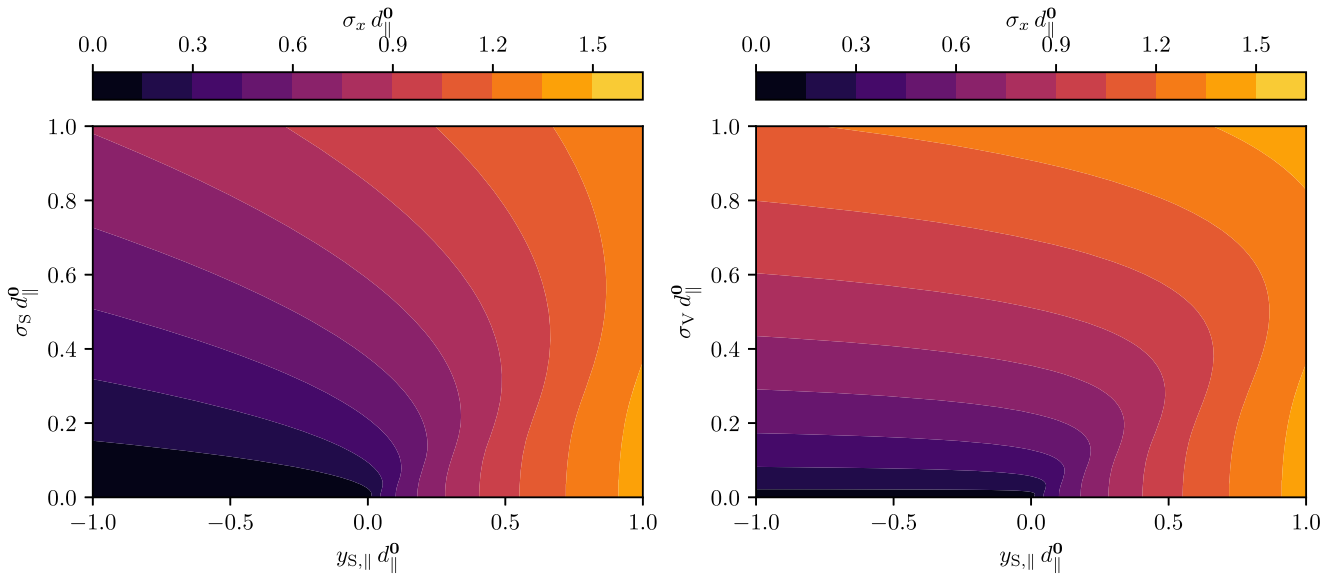


Figure 7. σ_x as a function of $y_{S,||}$ and σ_S (left) or σ_V (right). All quantities are in units of $1/d_{||}^0$. For a given resolution $\mathcal{R}$ (i.e., a given contour line in the plot), when σ_S is used to quote the size constraints, a source, no matter how big, can always appear unresolved by having a very negative $y_{S,||}$. This is not the case for σ_V as the contours flatten out for very negative $y_{S,||}$. So a valid size constraint can be derived using σ_V .

$$\sigma_x^2 = \frac{\sigma_V}{d_{||}^0} h\left(\frac{y_{S,||}}{\sigma_V}\right) = \frac{\sigma_S}{d_{||}^0} \frac{\sigma_V}{\sigma_S} h\left(\frac{y_{S,||} \sigma_S}{\sigma_S \sigma_V}\right) = \frac{\sigma_S}{d_{||}^0} \left[g\left(\frac{y_{S,||}}{\sigma_S}\right)\right]^{\frac{1}{2}} h\left(\frac{y_{S,||}}{\sigma_S} \left[g\left(\frac{y_{S,||}}{\sigma_S}\right)\right]^{-\frac{1}{2}}\right). \quad (\text{B10})$$

We compare this with Equation (B4), finding that the function $h(\eta)$ is implicitly, but uniquely, defined by

$$h\left(\frac{\xi}{[g(\xi)]^{1/2}}\right) = \frac{f(\xi)}{[g(\xi)]^{1/2}}. \quad (\text{B11})$$

We are not able to find an explicit formula for $h(\eta)$. The functions $f(\xi)$, $g(\xi)$, and $h(\eta)$ are plotted in Figure 6. The fact that $h(\eta)$ has range $(1, +\infty)$, together with Equations (B9) and (27), implies that $\bar{\sigma}_V = \mathcal{R}^2 d_{||}^0$, which then implies Equation (28). The reason why using σ_V in place of σ_S will remove the corner case mentioned in Section 3.1 is illustrated in Figure 7.

ORCID iDs

Lingyuan Ji <https://orcid.org/0000-0002-3947-7362>

Liang Dai <https://orcid.org/0000-0003-2091-8946>

References

- Amruth, A., Broadhurst, T., Lim, J., et al. 2023, *NatAs*, **7**, 736
 Baltz, E. A., Marshall, P., & Oguri, M. 2009, *JCAP*, **2009**, 015
 Barmby, P., McLaughlin, D. E., Harris, W. E., Harris, G. L. H., & Forbes, D. A. 2007, *AJ*, **133**, 2764
 Bartels, R., & Ando, S. 2015, *PhRvD*, **92**, 123508
 Baumgardt, H., & Hilker, M. 2018, *MNRAS*, **478**, 1520
 Broadhurst, T., Li, S. K., Alfred, A., et al. 2024, *ApJL*, **978**, L5
 Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, *MNRAS*, **321**, 559
 Chen, W., Kelly, P. L., Diego, J. M., et al. 2019, *ApJ*, **881**, 8
 Choe, S., Rivera-Thorsen, T. E., Dahle, H., et al. 2024, arXiv:2405.06953
 Cyr-Racine, F.-Y., Moustakas, L. A., Keeton, C. R., Sigurdson, K., & Gilman, D. A. 2016, *PhRvD*, **94**, 043505
 Dai, L. 2021, *MNRAS*, **501**, 5538

- Dai, L., Kaurov, A. A., Sharon, K., et al. 2020, *MNRAS*, **495**, 3192
 Dai, L., & Miralda-Escudé, J. 2020, *AJ*, **159**, 49
 Dai, L., & Pascale, M. 2021, arXiv:2104.12009
 Dai, L., Venumadhav, T., Kaurov, A. A., & Miralda-Escudé, J. 2018, *ApJ*, **867**, 24
 Dalal, N., & Kravtsov, A. 2022, *PhRvD*, **106**, 063517
 Diego, J. M., Amruth, A., Palencia, J. M., et al. 2024a, *A&A*, **690**, A359
 Diego, J. M., Kaiser, N., Broadhurst, T., et al. 2018, *ApJ*, **857**, 25
 Diego, J. M., Li, S. K., Amruth, A., et al. 2024b, *A&A*, **689**, A167
 Diego, J. M., Pascale, M., Kavanagh, B. J., et al. 2022, *A&A*, **665**, A134
 Diemand, J., Kuhlen, M., & Madau, P. 2007, *ApJ*, **667**, 859
 Diemer, B., & Joyce, M. 2019, *ApJ*, **871**, 168
 Fudamoto, Y., Sun, F., Diego, J. M., et al. 2024, arXiv:2404.08045
 Furtak, L. J., Meena, A. K., Zackrisson, E., et al. 2023, *MNRAS*, **527**, L7
 Han, J., Cole, S., Frenk, C. S., & Jing, Y. 2016, *MNRAS*, **457**, 1208
 Han, X., & Dai, L. 2024, *ApJ*, **964**, 160
 Hunter, D. A., Shaya, E. J., Holtzman, J. A., et al. 1995, *ApJ*, **448**, 179
 Kaurov, A. A., Dai, L., Venumadhav, T., Miralda-Escudé, J., & Frye, B. 2019, *ApJ*, **880**, 58
 Kelly, P. L., Chen, W., Alfred, A., et al. 2022, arXiv:2211.02670
 Kelly, P. L., Diego, J. M., Rodney, S., et al. 2018, *NatAs*, **2**, 334
 Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, *ARA&A*, **57**, 227
 Lin, J., Wagner, J., & Griffiths, R. E. 2022, *MNRAS*, **517**, 1821
 Ludlow, A. D., Bose, S., Angulo, R. E., et al. 2016, *MNRAS*, **460**, 1214
 Lundqvist, E., Zackrisson, E., Hawcroft, C., Amarsi, A. M., & Welch, B. 2024, arXiv:2404.10817
 Marsh, D. J. E., & Niemeyer, J. C. 2019, *PhRvL*, **123**, 051103
 Meena, A. K., Chen, W., Zittrn, A., et al. 2023, *MNRAS*, **521**, 5224
 Metcalf, R. B., & Petkova, M. 2014, *MNRAS*, **445**, 1942
 Miralda-Escudé, J. 1991, *ApJ*, **379**, 94
 Moliné, Á., Sánchez-Conde, M. A., Palomares-Ruiz, S., & Prada, F. 2017, *MNRAS*, **466**, 4974
 Nabizadeh, A., Zackrisson, E., Lundqvist, E., et al. 2025, *A&A*, **693**, A116
 Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, **462**, 563
 Oguri, M., Diego, J. M., Kaiser, N., Kelly, P. L., & Broadhurst, T. 2018, *PhRvD*, **97**, 023518
 Okamoto, T., Gao, L., & Theuns, T. 2008, *MNRAS*, **390**, 920

- Pascale, M., & Dai, L. 2024, [ApJ](#), **976**, 166
- Press, W. H., & Schechter, P. 1974, [ApJ](#), **187**, 425
- Rodney, S. A., Balestra, I., Bradac, M., et al. 2018, [NatAs](#), **2**, 324
- Salmon, B., Coe, D., Bradley, L., et al. 2020, [ApJ](#), **889**, 189
- Schauer, A. T. P., Bromm, V., Drory, N., & Boylan-Kolchin, M. 2022, [ApJL](#), **934**, L6
- Schneider, P., Ehlers, J., & Falco, E. E. 1992, *Gravitational Lenses* (Berlin: Springer)
- Vall Müller, C., & Miralda-Escudé, J. 2025, [MNRAS](#), **536**, 1579
- Vanzella, E., Meneghetti, M., Pastorello, A., et al. 2020, [MNRAS](#), **499**, L67
- Venumadhav, T., Dai, L., & Miralda-Escudé, J. 2017, [ApJ](#), **850**, 49
- Wagner, J. 2019, [Univ](#), **5**, 177
- Wagner, J. 2022, [A&A](#), **663**, A157
- Weisenbach, L., Anguita, T., Miralda-Escudé, J., et al. 2024, [arXiv:2404.08094](#)
- Welch, B., Coe, D., Diego, J. M., et al. 2022a, [Natur](#), **603**, 815
- Welch, B., Coe, D., Zackrisson, E., et al. 2022b, [ApJL](#), **940**, L1
- Williams, L. L. R., Kelly, P. L., Treu, T., et al. 2024, [AJ](#), **961**, 200
- Windhorst, R. A., Wyithe, S., Alpaslan, M., et al. 2018, *AAS Meeting*, 232, 325.09
- Yan, H., Ma, Z., Sun, B., et al. 2023, [ApJS](#), **269**, 43
- Zackrisson, E., Hultquist, A., Kordt, A., et al. 2023, [MNRAS](#), **533**, 2727