Abstract

Luminous stars in background galaxies straddling the lensing caustic of a foreground galaxy cluster can be individually detected due to extreme magnification factors of $\sim 10^2$–$10^3$, as recently observed in deep HST images. We propose a direct method to probe the presence of dark matter subhalos in galaxy clusters by measuring the astrometric perturbation they induce on the image positions of magnified stars or bright clumps: lensing by subhalos breaks the symmetry of a smooth critical curve, traced by the midpoints of close image pairs. For the giant arc at $z = 0.725$ behind the lensing cluster Abell 370 at $z = 0.375$, a promising target for detecting image pairs of stars, we find that subhalos of masses in the range of $10^6$–$10^8 M_\odot$ with the abundance predicted in the cold dark matter theory should typically imprint astrometric distortions at the level of 20–80 mas. We estimate that $\sim 10$ hr integrations with JWST at $\sim 1$–3 $\mu$m may uncover several magnified stars whose image doublets will reveal the subhalo-induced structures of the critical curve. This method can probe a dynamic range in the subhalo-to-cluster halo mass ratio of $m/M \sim 10^{-7}$–$10^{-5}$, thereby placing new constraints on the nature of dark matter.

Key words: dark matter -- galaxies: clusters: general -- gravitational lensing: strong

1. introduction

The cold dark matter (CDM) paradigm predicts that gravitational collapse of primordial density fluctuations produces bound dark matter halos with a wide range of masses (Davis et al. 1981; Blumenthal et al. 1982, 1984; Davis et al. 1985), the smallest of which may be as tiny as $10^{-6} M_\odot$ (Diemand et al. 2005; Profumo et al. 2006). Small halos merge to assemble massive halos and create a hierarchical structure all the way up to the scale of galaxy clusters, with masses of $\sim 10^{14} M_\odot$ (Press & Schechter 1974; White & Rees 1978).

Observational probes of dark matter halos over a broad mass spectrum are crucial for testing the CDM paradigm. Several hypotheses for the nature of the dark matter predict a deficit of halos below characteristic mass scales: these include warm dark matter (Colín et al. 2000; Bode et al. 2001; Viel et al. 2005), dark matter with a macroscopic de Broglie wavelength (Turner 1983; Sin 1994; Goodman 2000; Hu et al. 2000; Peebles 2000; Hui et al. 2017), and dark matter as compact objects (Carr & Hawking 1974; Meszaros 1974; Carr 1975).

When available, gravitational lensing is the most direct probe of dark matter halos in the absence of stars. Both strong and weak lensing have been used to measure the masses of halos hosting galaxy clusters (Hoekstra et al. 2013) and galaxies (Treu 2010; Mandelbaum 2015). Measurements of the orbital motions of stars and gas have also enabled us to reconstruct the dark matter distribution in the inner parts of galaxy halos (Sofue & Rubin 2001; Iocco et al. 2015). Detecting halos on sub-galactic scales ($\lesssim 10^9 M_\odot$) has proven challenging because these typically contain little luminous matter (Okamoto et al. 2008). Stellar dynamics measurements suggest that the ultra-faint satellites of the Milky Way could be held together by dark matter halos of sub-galactic masses $10^6$–$10^8 M_\odot$ (Bullock 2010), although the initial extent and total mass of these halos is unclear. Subhalos are also expected to leave characteristic imprints in stellar streams (Ibata et al. 2002; Johnston et al. 2002; Carlb erg 2009).

Strong gravitational lensing offers powerful tools to probe the distribution of dark matter within lensing halos. A classic application is to use magnification ratios of multiple images of background quasars or galaxies to detect subhalos in galaxy halos (Mao & Schneider 1998; Metcalf & Madau 2001; Dalal & Kochanek 2002; Veggetti et al. 2010, 2012; Cyr-Racine et al. 2016; Nierenberg et al. 2017). Dark matter substructure can also be detectable through distortions in the surface brightness patterns of lensed giant arcs seen at submillimeter (Hezaveh et al. 2013, 2016; Asadi et al. 2017) or optical/infrared wavelengths (Birrer et al. 2017). Together, these methods have yielded many recent detections of subhalos with masses of $10^8$–$10^9 M_\odot$. Lowering the mass detection limit would allow comparisons with other probes of sub-galactic scale halos, which suffer from very different systematics.

This paper proposes a novel probe of subhalos in lensing galaxy clusters. Our method requires two ingredients: a background galaxy that straddles the lensing caustic of a cluster, and a number of luminous stars (or compact light clumps) detected close to the critical curve. The image pairs of these stars (or light clumps) trace the critical curve and measure its degree of irregularity as affected by the subhalos.

Highly magnified images naturally occur in pairs straddling the critical curve of the cluster. For a smooth lens mass distribution, the lensing geometry takes the form of a fold catastrophe (Schneider et al. 1999), with images lying on symmetric positions, and the radius of curvature of the critical curve is comparable to the cluster’s Einstein radius, $\theta_E \sim 20''$. When subhalos are introduced, the critical curve develops small-scale distortions, with perturbations that are dramatically enhanced in the vicinity of the critical curve even for a small subhalo abundance, analogously to the microlensing effects of intracluster stars (Vennumadhav et al. 2017).

For typical parameters of cluster lenses, stars located $\sim 10$ pc from the caustic are magnified by factors of $\sim 10^2$–$10^3$. The

---

* NASA Einstein Fellow.
most luminous stars \( L \gtrsim 10^5 L_{\odot} \) can be individually detectable at optical/IR wavelengths from cosmological redshifts \( z_s \sim 1 \) (Miralda-Escude 1991). Recently, a first detection was reported in the Hubble Space Telescope (HST) images of the strong lensing cluster MACS J1149 (Kelly et al. 2017). We show that the giant arc in Abell 370 \( (z_s = 0.725; \text{see Richard et al. 2010}) \) is a promising target for detecting the images of several stars with the upcoming James Webb Space Telescope (JWST). If the separation between image pairs can be resolved, they can be robustly identified and used as tracers of the critical curve. For the test case in Abell 370, we show that at the sensitivity of JWST, the expected angular separation between detectable image pairs of magnified stars is \( \sim 0''1 \), which is above the astrometric resolution. We propose to search for this distortion by performing astrometry on image pairs. In our test case, the typical size of the astrometric distortion is \( \sim 20-100 \text{ mas due to subhalos with masses of } \sim 10^6-10^8 M_{\odot} \).

Micro-lensing by intracluster stars causes short time variability in the magnified images of stars that allows for their identification (Diego et al. 2018; Oguri et al. 2018; Venumadhav et al. 2017; Windhorst et al. 2018). However, in this paper we are interested in using images of stars or of other bright clumps in the lensed galaxies as a probe of the shape of the critical curve, for which microlensing variability is not crucial. Accurate astrometry can be done with images of any source (stellar clusters, H II regions or irregularities in the lensed galaxy surface brightness) that is compact enough and close to the caustic.

The paper is organized as follows. In Section 2, we first review the caustic crossing phenomenon in a smooth fold model without substructure, and then discuss lensing perturbations from subhalos and the impact of micro-lensing by intracluster stars. In Section 3, we show that the giant arc in the strong lensing cluster Abell 370 is particularly promising for detecting several caustic crossing luminous stars, and present simulated lensing signatures of subhalos imprinted in the image positions within the arc. Finally, we summarize the results and discuss observational prospects for our method with optical/infrared telescopes in Section 4, highlighting the scientific potential of the forthcoming JWST. For interested readers, we collect some of the technical information in the appendices. Appendix A details our modeling of the subhalo abundance based on existing N-body simulations. Appendix B presents the analytical model used for the lensing effect of individual subhalos. Appendix C briefly explains the ray-shooting code used for the lensing simulations.

Throughout this paper, we adopt the Planck best-fit cosmological parameters of Ade et al. (2016). We work with geometrical optics, which is an excellent approximation for lensing at UV and infrared wavelengths.

2. Lensing of Caustic Crossing Stars

The method laid out in this paper relies on the behavior of lens models near singularities, the most common of which is the fold. After briefly reviewing the fold, we proceed to discuss the perturbing effect of substructure.

2.1. Lensing without Subhalos: Fold Model

We follow the notation of Venumadhav et al. (2017) for the lens mapping \( y = y(x) \) from the image position \( x \) to the source position \( y \), in angular units. The Jacobian matrix of the map is \( A(x) = \partial y / \partial x \). In terms of this Jacobian matrix, the (signed) magnification factor \( \mu(x) \) is the inverse of the determinant, i.e.,

\[
\mu(x) = 1 / \det[A(x)],
\]

the convergence \( \kappa(x) \) is related to the trace by \( \kappa(x) = 1 - (\text{tr}[A(x)] / 2) \), and the lensing shear is the trace-free and symmetric part.

Critical curves are contours on the image plane where the magnification diverges, i.e., \( \det[A(x)] = 0 \). It is convenient to choose a coordinate system with the origin on a point of a critical curve around which we wish to examine the lens map, with its first axis \( \hat{x}_1 \) along the direction corresponding to the zero eigenvalue of the Jacobian matrix (see Figure 1). If the lens mass is smoothly distributed near the origin (i.e., for small \( x \)), we can approximate the Jacobian matrix \( A(x) \) as (Schneider et al. 1992)

\[
A(x) = \begin{bmatrix}
    x \cdot d & 0 \\
    0 & 2(1 - \kappa_0)
\end{bmatrix},
\]

which is characterized by a local convergence \( \kappa_0 \), and an eigenvalue gradient vector \( d = (d \sin \alpha, -d \cos \alpha) \). Locally, the critical curve is the straight line \( d \cdot x = 0 \), which is inclined at an angle \( \alpha \) relative to the \( \hat{x}_1 \) axis, the direction of elongation of a background galaxy image. The convergence \( \kappa_0 \) is typically of order unity, and the angular scale \( 1/d = 1/|d| \) is \( \sim \) tens of arcseconds.

Consider an image-plane trajectory \( x(\lambda) \) that intersects the critical curve and is parallel to the degenerate direction of the lens map: for the system in Equation (1), \( x(\lambda) = \lambda \hat{x}_1 \). The source location \( y(0) \) corresponding to \( \lambda = 0 \) lies on the lensing caustic. We solve for the source-plane trajectory using the definition of the Jacobian matrix as follows:

\[
y(\lambda) - y(0) = \int_0^\lambda d\lambda' A(x(\lambda')) \cdot \frac{dx(\lambda')}{d\lambda'},
\]

Figure 1. Symmetric image positions in the fold catastrophe of a smooth lens. The critical curve (solid) can be locally approximated as a straight line. The degenerate direction is along the \( x_1 \) axis. To emphasize the stretching of the images, their sizes have been exaggerated. The dashed line illustrates a critical curve with a small curvature.
solution for $\lambda$ tells us that there are two images separated by

$$2\Delta x = 2 \left( \frac{\Delta y}{|d \sin \alpha|} \right)^{1/2}$$

$$= 0''7 \times \left( \frac{\Delta y D_S}{10 \text{ pc}} \frac{1 \text{ Gpc}}{D_S} \frac{\text{arcmin}^{-1}}{|d \sin \alpha|} \right)^{1/2}. \quad (3)$$

Both images have the same magnification factor

$$|\mu| = [2 \Delta x |(1 - \kappa_0) d \sin \alpha|]^{-1}, \quad (4)$$

and are symmetrically arranged about the critical curve along the direction $\hat{y}$, as illustrated in Figure 1. Each image is itself elongated by the large magnification, although this is unresolvable for stars.

When several stars are located near a caustic, each one has a pair of images aligned along the common degenerate direction. The image separations vary, but the critical curve bisects all segments joining each pair. This can be exploited as a model-independent test of the smoothness of the lens model. While the critical curve is not directly observable, it is traced by the image positions of lensed stars.

There are two effects that can potentially change the simple symmetry of the fold: microlensing by intracluster stars and small-scale structure in the dark matter distribution that perturbs the smoothness of the fold. We will examine microlensing later in Section 2.3 and show that it is unimportant. The natural radius of curvature of the critical line is the Einstein scale of the cluster $\theta_C \approx 20''$. Across the typical width of a giant arc, $w_{\text{arc}} \approx 0''5$, this curvature causes a departure from a straight line of only $\sim w_{\text{arc}}^{2}/(2 \theta_C) \approx 10 \text{ mas}$ (see Figure 1), about one-third of the pixel size of HST. Larger curvatures can arise from the lensing influence of individual cluster galaxies if they happen to lie close in projection. We will show that dark matter subhalos produce larger curvatures and more complex distortions.

### 2.2. Lensing with Subhalos

We now consider the effect of cluster subhalos in the projected vicinity of a critical curve. Figure 2 illustrates how a single subhalo perturbs the critical curve. The perturbed critical curve has a characteristic wiggle over an angular region comparable to the separation between the subhalo center and the unperturbed critical curve, with an amplitude that grows with the subhalo mass and decreases with the separation. Owing to the quadrupolar nature of the subhalo shear, the perturbation consists of two “bulges” facing opposite directions, rather than a single “bulge.” A new closed branch of the critical curve may arise around the subhalo center if its perturbation is sufficiently strong (as is the case in Figure 2).

Dark matter subhalos are generally difficult to detect through lensing because their surface density is highly subcritical. However, near the critical line of a lensing cluster, their effect is greatly enhanced. This can be seen from the following argument, first presented in Katz et al. (1986).

In the absence of the perturber, the image position $\mathbf{x}$ satisfies the lens equation

$$\mathbf{x} - \mathbf{y} - \alpha_E(\mathbf{x}) = 0, \quad (5)$$

Figure 2. Perturbations to image locations and the critical curve due to a single subhalo centered on the black cross. Dotted and solid lines are the lensing critical curves without and with a subhalo, respectively. Arrows indicate image displacements as the subhalo mass is increased. Dashed segments indicate images that merge with their counter-image and disappear before the subhalo mass reaches its final value. The fold parameters used are typical of the critical curves of strong lensing galaxy clusters. The subhalo is modeled as a truncated Navarro–Frenk–White (NFW) sphere with bound mass $m = 2.5 \times 10^7 M_\odot$, scale radius $r_s = 300 \text{ pc}$, and concentration parameter $c_{200} = 24$ (see Section 3.3). Upper panel: subhalo located to the outside of the critical curve ($d \cdot \mathbf{x} > 0$). Images very close to the critical curve shift toward it, and some eventually merge with their counter-images and disappear. Lower panel: the same subhalo located to the inside of the critical curve ($d \cdot \mathbf{x} < 0$). Images very close to the critical curve shift apart from it, and new pairs of images emerge (not shown in this figure; they fill the vacated space on both sides of the critical curve). Here we have always assumed $\kappa_0 < 1$, which is the case for many caustic crossing arcs. The case of over-focusing $\kappa_0 > 1$ can be similarly studied.
where $\alpha_B(x)$ is the deflection angle of the smooth lens model with no subhalos. Adding a subhalo shifts the image to a new position $x' = x + \Delta x$, which satisfies the new lens equation

$$x + \Delta x - y - \alpha_B(x + \Delta x) - \alpha_{sh}(x') = 0. \quad (6)$$

Expanding to first order the smooth deflection as $\alpha_B(x + \Delta x) \approx \alpha_B(x) + \Delta x \cdot \nabla \alpha_B(x)$, the shift of the image position in Equation (6) is

$$\Delta x \approx [A_B(x)]^{-1} \cdot \alpha_{sh}(x'). \quad (7)$$

The smooth Jacobian matrix $A_B(x)$ is nearly degenerate close to a fold, as in Equation (1). Its inverse has a large eigenvalue $\mu(x)/[2(1 - \kappa_0)]$ along the degenerate direction, where $\mu(x)$ is the magnification factor in the fold model. The image displacement $\Delta x$ is of the same order as the deflection $\alpha_{sh}(x')$ along the nondegenerate direction, but is larger by a factor of $\mu$ along the degenerate direction. The astrometric perturbation of subhalos is therefore dramatically amplified near the critical curve, making highly magnified image pairs sensitive probes of substructure.

As seen in Figure 2, the image displacements caused by subhalos are predominantly along the degenerate direction. In line with Equation (7), images that are close to the smooth critical curve exhibit the largest displacements under the subhalo influence, while distant images are less affected. These displacements break the symmetry of image pairs with respect to the smooth critical curve. In the vicinity of the critical curve, a pair of images corresponding to the same source are displaced in opposite directions but by asymmetric amounts, causing their midpoint to shift.

If the subhalo has a closed branch of the critical curve around its center, a source located within the corresponding closed caustic has two image pairs (instead of one), which are aligned along the common degenerate direction. If detected, this would immediately exclude a smooth lens model. However, because of the small area within closed caustics of subhalos, we expect these cases to be rare.

2.3. Microlensing by Intracluster Stars

The projected surface density of intracluster stars near critical curves, typically $\sim 50$ kpc from the cluster center, usually contributes a lensing convergence of $\kappa_s \approx 10^{-3} - 10^{-2}$ (Zwicky 1951; Lin & Mohr 2004; Zibetti et al. 2005). These stars may have been tidally stripped from cluster galaxies throughout the assembly history of the brightest central galaxy (BCG), or have formed in situ within tidally ejected intracluster clouds (Martel et al. 2012; Contini et al. 2014; Cooper et al. 2015). Despite these low values of $\kappa_s$, a similar line of reasoning as that in Equations (5)–(7) shows that the associated optical depth to microlensing is enhanced by the magnification factor of the fold model (e.g., Venumadhav et al. 2017). This is the reason for the substantial flux variations in the individual images of the caustic crossing star detected in observations of MACS J1149 (Diego et al. 2018; Kelly et al. 2017; Oguri et al. 2018).

Microlensing breaks the smooth critical curve into a band of micro-critical curves of width $\sim 2 \kappa_s/d$, where images are generally affected by more than one microlens. There are two main effects on the macroimages of background stars, i.e., their images in a model without microlenses.

First, each macroimage is broken into a series of aligned microimages along the degenerate direction, with a typical angular spread $\lesssim \theta_e \approx \theta_B/\kappa_s^2/d \sin \alpha_0^2 \approx 1$ mas, where $\theta_B$ is the Einstein radius of a microlens (Venumadhav et al. 2017). This spread is not resolvable by present telescopes, so each track of microimages appears as a single macroimage.

Second, the number of microimages and their fluxes fluctuate due to the differential proper motion of the source and lens. The net result is a stochastic variation in the flux of each macroimage, and a jitter in the position of its centroid. The latter is at the $\sim 1$ mas level. This is negligibly small compared to the astrometric signature ($\sim 10–100$ mas) of subhalos with masses of $10^{8–3} M_\odot$. Thus, we expect our astrometric method to be unaffected by microlensing.

Microlensing variations in the flux may sometimes push individual macroimages below the detectability threshold. Indeed, this occurred during the observation of the first caustic crossing star in MACS J1149 (Kelly et al. 2017). This may complicate the detection and identification of image pairs, but can be corrected through imaging at multiple epochs to average over microlensing variations. In general, our method does not have to rely on the flux to correctly identify image pairs associated with the same source star. Instead, one may use the properties of image pairs laid out in Section 2.1—macroimages of the same source star almost always align along the common degenerate direction. Colors are also helpful in identifying image pairs.

3. Case Study: Giant Arc in Abell 370

We illustrate our method via a case study of the massive galaxy cluster Abell 370 (Abell 1958) at $z_l = 0.375$ ( Mellier et al. 1988), historically one of the first studied strong lensing systems (Fort et al. 1988; Mellier et al. 1991; Kneib et al. 1993; Smail et al. 1996; Bézecourt et al. 1998, 1999). The cluster is notable for the “Dragon,” an exceptionally long and luminous arc (Lynds & Petrosian 1986; Soucail et al. 1987), which is a gravitationally lensed background galaxy at a source redshift of $z_s = 0.725$ (Soucail et al. 1988; Lynds & Petrosian 1989). Detailed lens modeling shows that the giant arc consists of five images of the background galaxy joined end-to-end at four intersections with the lensing critical curve (Richard et al. 2010; Lagattuta et al. 2017) (see upper panel of Figure 3).

In this section, we present lensing simulations of one of these intersections in the “Dragon,” including a population of subhalos in its projected vicinity, and we discuss observational prospects for the method discussed in Section 2.2. Our simulations contain the following ingredients: (1) a smooth fold in a region near our chosen intersection of the critical curve and the giant arc; (2) random realizations of subhalos causing substantial lensing perturbations in this region; and (3) a realistic, randomly generated stellar population in the corresponding region of the source galaxy. We then calculate the positions and fluxes of the star images.

We do not include microlensing by intracluster stars in our simulations. As previously argued, this does not affect the image positions, but causes additional flux variations. A full-scale simulation of the entire intersection including microlensing was not computationally feasible with our method.
1. Subhalo lensing is therefore not among other strong lensing clusters
which is parameterized by a virial mass \( M_{200} \), scale radius \( R_s \), and concentration parameter \( C_{200} \) (Navarro et al. 1996) (see Appendix A for precise definitions). Typical values for dynamically relaxed clusters are \( M_{200} \approx 2 \times 10^{15} M_\odot \) and \( C_{200} \approx 7–8 \) (Umetsu et al. 2011).

The mass modeling for Abell 370 is complicated by the dynamically unrelaxed nature of the cluster, which has two components of comparable mass that are undergoing a major merger nearly along the line of sight (Richard et al. 2010). The two components are centered near the two BCGs—a primary one near the “Dragon,” and a secondary one about 200 kpc away in projection.

We model the total mass profile by the superposition of two equal NFW halos, each with a virial mass of \( M_{200} = 1.6 \times 10^{15} M_\odot \), concentration parameter of \( C_{200} = 7 \), scale radius of \( R_s = 310 \) kpc, and virial radius of \( R_{200} = C_{200} R_s = 2.1 \) Mpc, respectively. The resulting total enclosed mass reasonably reproduces the measured value of \( M(<250 \) kpc\) = 3.8 \times 10^{14} M_\odot \) centered at the midpoint of the two BCGs (Richard et al. 2010). The concentration parameters found by Umetsu et al. (2011) are significantly higher than typical values in N-body simulations for similar halos. This is not uncommon among other strong lensing clusters (e.g., Abell 2667; Covone et al. 2006), and is likely due to a selection bias for concentrated systems that have large cross sections for strong lensing.

The “Dragon” intersects the critical curve at four locations. We focus on a small region of size 3" × 1" centered on one of these intersections, shown in the lower panel of Figure 3. The impact parameter from the primary BCG is \( \approx 10^5 \), or \( B = 0.17 R_s = 53 \) kpc in proper units. The local fold corresponding to the smooth lens model has parameters \( \kappa_0 = 0.69, |d| = 2.7 \) arcmin\(^{-1} \), and \( \alpha = 45^\circ \) (see Equation (1)). We select this fold location because of the higher magnification factors compared to the other three locations.

3.2. Subhalo Abundance

Generating random realizations of subhalos requires a model of their abundance and spatial distribution in cluster halos. Han et al. (2016) derive such a model from N-body simulations with moderate subhalo-to-host mass ratios of \( 10^{-7} < m/M_{200} < 10^{-3} \). However, subhalo masses of our interest \( m \lesssim 10^8 M_\odot \) are under-resolved even in state-of-the-art simulations of cluster-size halos. Hence, we need to extrapolate their model to lower subhalo masses.

Based on the semi-analytic model of Han et al. (2016), we use the following procedure to generate subhalos. We first generate the initial subhalo mass \( m_{\text{acc}} \) when the subhalo is incorporated into the host halo. We assume that the unevolved specific mass function, \( \frac{dn(m_{\text{acc}}, R)}{d \log m_{\text{acc}}} \), spatially traces the host halo mass density profile \( \rho(R) \),

\[
\frac{dn(m_{\text{acc}}, R)}{d \log m_{\text{acc}}} = A_{\text{acc}} \rho(R) \left( \frac{m_{\text{acc}}}{m_0} \right)^{-\alpha},
\]

where \( m_0 = 10^{10} h^{-1} M_\odot \). We adopt commonly used values for the power-law index \( \alpha = 0.9 \) (Mo et al. 2010) and normalization constant \( A_{\text{acc}} = 0.08 \) (Han et al. 2016).

Owing to tidal stripping, the bound mass \( m \) of a subhalo decreases with time from its initial value \( m_{\text{acc}} \). In simulations, the ratio \( m/m_{\text{acc}} \) is highly stochastic, since tidal stripping depends on the orbital history of the subhalo. We present our procedure for generating the distribution of \( m/m_{\text{acc}} \) in detail in Appendix A. In brief, and following Han et al. (2016), we draw the ratio \( m/m_{\text{acc}} \) from the log-normal distribution of Equation (13), which depends on the halo-centric radius \( R \) and includes a probability for complete tidal disruption that is set to 56%.

The left panel of Figure 4 shows the projected subhalo abundance as a function of the projected distance \( B \) to the primary BCG. The surface number density increases with decreasing distance \( B \) down to the scale radius \( R_s \) of the host, and plateaus thereafter. Note that the caustic crossing location on the giant arc has an impact parameter \( B = 53 \) kpc to the nearby BCG, which is much less than the scale radius \( R_s \). Since the projected separation between the two-component halos is also less than \( R_s = 310 \) kpc, the subhalo contributions from both components are nearly equal for equal component masses. Given the mass-modeling uncertainty for the non-relaxed cluster, we generate subhalos within the NFW halo associated with the nearby BCG using Equation (8), and simply multiply the abundance by a factor of two. Including both components, we find on average \( \approx 1 \) subhalo more massive than \( 10^7 M_\odot \) within a 0\" × 0\" 2/patch. Subhalo lensing is therefore not rare, and we have checked that the result does not substantially change even if we set the concentration parameter to its cosmic mean of \( C_{200} = 3.5 \) for cluster-sized halos at low redshifts (Ludlow et al. 2016).

The right panel of Figure 4 shows that even for low impact parameters, the subhalos in the projected vicinity of the line of sight are most likely to have halocentric distance \( R \approx R_s \). Therefore, although the impact parameter is typically \( B < R_s \), most detectable subhalos that our method is sensitive to are

![Figure 3. Upper panel: false color image of the “Dragon,” a giant arc in Abell 370, plotted using the archival HST data from f450w + f606w (blue), f814w (green), and f105w + f125w (red) filters. The white line is the critical curve at the arc redshift \( z_\text{a} = 0.725 \). Lower panel: zoomed-in region of size 3\" × 1\"2 of the most promising crossing location of the critical curve. The lens model is adopted from Richard et al. (2010) and the critical curve was calculated using Lenstool (Jullo & Kneib 2009).](image-url)
likely far in three-dimensional distance from the central region near the BCG.

For the lensing simulations, we generate subhalos within a cylinder of radius \( R = 1'' \) and line-of-sight depth 100 \( \rho_\text{s} \), centered on the simulated 1'' × 1'' region. We conserve the mass within the cylinder by removing an appropriate uniform surface-mass density. We focus on subhalo masses \( m \in [10^4, 10^9] M_\odot \), since more massive subhalos are rare, while subhalos with lower masses do not have a resolvable astrometric signature.

### 3.3. Internal Structure of Subhalos

The lensing effects of subhalos depend on their detailed mass profiles. Our method probes subhalos with mass lower than that of the host cluster halo by factors of \( \sim 10^{7-9} \). Such a large dynamic range has never been achieved even in the highest resolution dark matter only simulations run so far (Diemand et al. 2008). Recent work has shown that simulations are prone to artificial numerical relaxation effects at the edge of their mass resolution (van den Bosch et al. 2018); this precludes a direct use of numerically obtained subhalo mass profiles near the resolution limit. We will extrapolate simple fits to the profiles of more massive subhalos and neglect complications such as triaxiality.

We parameterize subhalo density profiles as smoothly truncated NFW (TNFW) profiles (Baltz et al. 2009; Cyr-Racine et al. 2016), with the functional form

\[
\rho_{\text{TNFW}}(r) = \frac{m_{200}}{4\pi r_s^3 f(c_{200})} \frac{1}{(r/r_s)(1 + r/r_s^2)^2} \frac{1}{1 + r^2/r_s^2},
\]

where \( r \) is the distance to the subhalo center, and the finite total mass \( m \) differs from \( m_{200} \). The parameters \( m_{200}, r_s, \) and \( c_{200} \) are the usual parameters of (untruncated) NFW halos. The instantaneous tidal radius \( r_i \) is defined in terms of the halo mass

\[
distance within the host halo, \( R \), as (Binney & Tremaine 1987)

\[
r_i = \left( \frac{m}{M_\odot(R)} \right)^{2} \left( 1 - \frac{d \ln M_\odot(R)}{d \ln R} \right) \frac{1}{R},
\]

where \( M_\odot(R) \) is the mass enclosed within \( R \). At fixed \( m \), this always underestimates the true compactness of the subhalo, which is generally truncated at a radius smaller than \( r_i \).

We use a concentration parameter that increases toward the inner part of the host halo according to the fit (Bullock et al. 2001; Diemand et al. 2007; Bartels & Ando 2015; Moliné et al. 2017)

\[
c_{200}(m, R, z) = C_{200}(m, z) \left[ 1 + \frac{1}{15} \left( \frac{(1.5 R_{200})^2}{R^2 + (0.1 R_{200})^2} \right)^2 \right],
\]

where \( C_{200}(m, z) \) is the concentration for field halos (Ludlow et al. 2016). We collect some analytical formulae for lensing by TNFW density profiles in Appendix B.

### 3.4. Source Stellar Population

The final ingredient in our simulations is the magnitude distribution of stars in the vicinity of the caustic. We model the source stellar population based on archival HST images of Abell 370 in seven filters: f435w, f606w, f814w, f105w, f125w, f140w, and f160w.

We model the data within a 1'' × 1'' square patch that is centered on the midpoint of the segment of the lensing critical curve shown in the zoomed-in panel of Figure 3, and has its first coordinate axis aligned with the direction of arc elongation. Based on the surface brightness and color inferred from the multi-band HST data, we divide the patch into five regions, and separately model the stellar population in each region using the following procedure.

We consider six values of stellar metallicity: \( \log(Z/Z_\odot) = 0.2, 0.0, -0.3, -0.5, -1.0, -1.5 \). For each metallicity, we use the population synthesis code FSPS (Conroy et al. 2009;
Conroy & Gunn 2010) to construct simple stellar populations (SSPs) in 14 age bins: \( t_{\text{age}} = 300 \) kyr, 500 kyr, 1 Myr, 2 Myr, 5 Myr, 10 Myr, 25 Myr, 50 Myr, 100 Myr, 250 Myr, 500 Myr, 1 Gyr, 3 Gyr, and 7 Gyr (the age of the universe at \( z_s = 0.725 \) is about 7 Gyr). We then find the restricted linear combination of age bins (with non-negative coefficients) that best reproduces the measured fluxes in the seven HST filters. When constructing the SSP templates, we correct for dust attenuation in the source galaxy using the two-component power-law prescription of Charlot & Fall (2000). We also allow for an additional component to account for the contamination from a nearby cluster member galaxy (only \( \sim 3^{\prime\prime} \) away in projection), and from the intracluster light surrounding the primary BCG. We restrict this component to have the same color as that of the cluster member galaxies. Finally, we choose the value of metallicity \( \log(Z/Z_\odot) \) with the smallest best-fit residual. Although the number of linear components invoked exceeds the number of HST filters used, our approach does not have an over-fitting problem. We found that the physical requirement that the coefficients of the linear components be non-negative is sufficiently restrictive so as to prevent over-fitting.

### 3.5. Simulating Caustic Crossing Stars

Our fits favor \( -0.5 \lesssim \log(Z/Z_\odot) \lesssim -0.3 \) in regions of high surface brightness, which is consistent with the mean stellar metallicity found across the entire source galaxy in Patricio et al. (2018). We identify SSP components with ages in two widely separated ranges, \( t_{\text{age}} = 2–10 \) Myr and \( t_{\text{age}} = 1–7 \) Gyr, with the former suggesting recent star formation. This is in line with the appearance of the galaxy in the reconstructed source-plane images based on multi-band HST images, with a red core, and blue star-forming clumps in the outskirts.

In the left panel of Figure 5, we show the Hertzsprung–Russell diagram for the best-fit stellar population. In general, stars with very high bolometric luminosities \( L_{\text{bol}} \gtrsim 10^3 L_\odot \) are promising candidates for individual detection as caustic crossing stars. Specific to \( z_s = 0.725 \), the brightest stars in the near-IR bands (1–4 \( \mu m \)) are typically red supergiants of spectral type K. The brightest stars in optical bands (0.4–0.8 \( \mu m \)) are white and blue supergiants of spectral type A or B and with surface temperatures \( T_{\text{eff}} = 7500–20,000 \) K. Main-sequence stars of spectral type ranging from B–O can have higher temperatures \( T_{\text{eff}} \gtrsim 20,000 \) K, but their bolometric luminosities are only comparable to those of the supergiants. This is because their typical stellar radii \( \sim 10 R_\odot \) are much smaller than those of the supergiants \( \sim 10^2–10^3 R_\odot \). Since most of their energy output is in the UV (\( \lambda_{\text{peak}} \approx 0.15 \mu m \) in the rest frame), hot main-sequence stars are not ideal candidates for detection in optical/IR bands (Kelly et al. 2017).

The right panel of Figure 5 shows distributions of the absolute magnitude in a few selected filters. The distributions have power-law tails at the bright end, \( dN/dL \propto L^{-2} \). This slope compares well to the values measured in actively star-forming systems, such as 30 Doradus (Kelly et al. 2017). Caustic crossing arcs that lack recent star formation are unlikely to host a substantial number of supergiants. Thus, giant arcs that host star-forming structures are more promising targets for detecting caustic crossing stars.

To predict the magnitudes of observable highly magnified stars, we randomly draw stars from the best-fit stellar population and distribute them uniformly on the source plane (in reality young stars cluster; we recall that image pairs of bright clumps such as open clusters or H II regions can be equally used to trace the critical curve). We read off the stellar radius \( R_\star \) and luminosity of each star in various photometric bands (redshifted to \( z_s = 0.725 \)) from the outputs of FSPS. We do not simulate microlensing, and hence resolve only the stellar sizes of giant stars with \( R_\star \sim 10^2–10^3 R_\odot \) that approach the macrocaustic. We normalize the overall number of source stars to produce the measured surface brightness of the giant arc (which is preserved by lensing).

We populate stars within an elongated \( \sim 7 \) kpc \( \times \) 0.2 kpc region within the source galaxy (that the cluster caustic passes through); out of \( \sim 10^9 \) stars, a few thousand have \( L_{\text{bol}} > 10^5 L_\odot \). Individually detectable stars mainly belong to this ultra-luminous population. We compute the magnitudes of the (lensed) macroimages for a smooth fold model, i.e., a model without dark matter substructure and without microlensing by intracluster stars. The upper panel of Figure 6 shows the number of image pairs with apparent magnitudes brighter than a given value, and that lie within the 1″ \( \times \) 1″ region at the..
Figure 6. Expected number of highly magnified pairs of star images for Abell 370 assuming a simple fold model. Upper panel: cumulative count of image pairs vs. the maximum apparent magnitude. Results for a few selected HST and JWST wide filters are shown. Middle panel: cumulative count of image pairs vs. the minimum signal-to-noise ratio and JWST exposure time of 6 ks with JWST, for four selected NIRCam wide filters. The mean metallicity is \( \log(Z/Z_\odot) = -0.4 \) according to our modeling. Lower panel: same as middle panel but assuming \( \log(Z/Z_\odot) = 0.2 \). In all panels, horizontal bars show the mean and shaded bands show the sample variance.
able to simulate realizations of highly magnified stars in the vicinity of the critical curve, incorporating the effect of subhalo lensing. For each realization, we pixelize a 1′′ × 1′′ region centered at the smooth critical curve at a resolution of 8 × 8 mas. Detectable magnified stars have high luminosities, so it suffices to individually keep track of stars with $L_{\text{crit}} > 10^4 L_\odot$ and locate the pixels containing their macro-images. See Appendix C for details on how we numerically simulate lensing with substructure.

### 3.6. Astrometric Sensitivity to Subhalo Lensing

We now discuss lensing simulations computed for our fiducial parameters of the giant arc in Abell 370.

Based on a typical realization of subhalos and source stars in the giant arc, Figure 7 exemplifies the subhalo lensing effect on the shape of the critical curves and the image positions of magnified stars. We choose the JWST ultra-wide filter f150w2 as an example. We select stars with both macroimages brighter than a threshold magnitude of $m_{\text{AB}} = 31$ (without microlensing), corresponding to a point-source detection $S/N \approx 3$ (5) with $\sim$10 hr of integration.

Figure 7 shows that subhalos strongly distort the smooth critical curve, create wiggles, and break it up into loops. To study the effects of subhalos of different masses, we set the maximum subhalo mass to $m = 10^9 M_\odot$ and successively include subhalos of smaller masses. Large subhalos $m \geq 10^9 M_\odot$ (left panel) create features on large angular scales. Smaller subhalos ($10^6 M_\odot \lesssim m \lesssim 10^8 M_\odot$; middle panel), with a higher surface density, cause ubiquitous distortions on angular scales $\lesssim 100$ mas. Very small subhalos $10^4 M_\odot \lesssim m \lesssim 10^6 M_\odot$ imprint distortions on very fine angular scales of $\lesssim 10$ mas, which are difficult to resolve with current or forthcoming instruments.

As seen in Figure 7, most detectable image pairs of magnified stars are located within $\sim$0′′1 of the cluster critical curve. Because of the perturbations by subhalos, the astrometric midpoints of the image pairs do not align along a smooth curve. To test the null hypothesis that the simple fold model can account for the image positions, we fit the midpoints to a straight line using orthogonal regression (Deming 1943).

Under the simplifying assumption that all image pairs have an equal, uncorrelated astrometric precision, we quantify the statistical significance of the departure from the null hypothesis with the quantity,

$$S = \frac{2}{\sigma_\theta} \sqrt{\frac{\sum_{i=1}^{N_p} s_i^2}{N_p - p}} - \left(\frac{N_p}{p} - 1\right)^{1/2}.$$

Here, $N_p$ is the total number of image pairs, and $s_i$ is the perpendicular residual from the best-fit critical curve to each midpoint. For the most simple approximation to the critical curve, a straight line, the number of free parameters is $p = 2$. We use $\sigma_\theta$ to denote the standard deviation of astrometric errors of image positions along each Cartesian axis. The astrometric standard deviation of the midpoint of two images is then $\sigma_\theta / \sqrt{2}$. For the realization shown in Figure 7, an astrometric precision of $\sigma_\theta \approx 60$–80 mas is required to reject the null hypothesis at 2σ($S = 2$).

Despite their low number density, subhalos of larger masses $m \gtrsim 10^{10} M_\odot$ (which may be hosting dwarf or large galaxies) can accidentally lie close to the critical curves. Alternatively, a group of less massive subhalos clustered together by chance can collectively generate a significant lensing perturbation. Acting at intermediate distances (say, $\gtrsim 1'$), these effects can induce a local curvature in the smooth critical curve. If the optical counterpart of a large subhalo is measured, its contribution to lensing can be modeled and corrected, but this will generally still leave uncorrected perturbations from relatively massive subhalos. In order to relax the null hypothesis to account for this possibility, we modify our model fitting the midpoints to a partial circle, for which $p = 3$ in Equation (12). However, a very small best-fit radius of curvature suggests the presence of dark subhalos (if not
accompanied by a dwarf galaxy in the lensing cluster) in the mass range we are interested in detecting. For this reason, we impose a minimum radius of curvature 300 mas to prevent falsely excluding evidence for substructure.

Our analysis assumes that image pairs can be correctly identified and that astrometric accuracy is not worsened by blending and crowding. In practice, the identification of image pairs may not be easy, but is aided by the known and fixed direction of elongation along which all pairs should be split.

To statistically quantify the required astrometric precision for rejecting the null hypothesis, we simulate a large number of random realizations of subhalos and source stars, and perform both the line fit and the circle fit for each realization as in the case of Figure 7. Figure 8 shows that the cumulative distribution of the required $\sigma_\theta$ has a median at \( \sim 30-80 \) mas. Larger values of $\sigma_\theta$ are required when subhalos smaller than \( 10^6 M_\odot \) are added, but this saturates for \( m_{\text{min}} \lesssim 10^8 M_\odot \).

Theoretically speaking, too many low-mass subhalos can limit the maximum magnification factor and therefore suppress the overall number of detectable bright images. Our simulation results show this effect when the minimum subhalo mass is decreased from \( 10^5-10^6 M_\odot \), but we find no evidence that this trend continues to \( m \lesssim 10^6 M_\odot \). We therefore conclude that the astrometric test with magnified stars is most sensitive to subhalos in the mass range of \( 10^6 M_\odot \lesssim m \lesssim 10^8 M_\odot \). More than six orders of magnitude below the mass scale of the host halo, these are beyond the reach of the state-of-the-art N-body simulations of halo formation. Since star formation should be quenched due to the hot intracluster environment (Taranu et al. 2014), we expect these subhalos to be non-luminous systems. Constraints on their abundance should be of great interest as a probe to the dark matter.

The left panel of Figure 8 shows that allowing the smooth critical curve to have a small curvature (radius of curvature \( > 10^3 \text{mas} \)) can reduce but not fully eliminate the astrometric residuals. As demonstrated by Figure 7, perturbations by subhalos shift the midpoints of image pairs incoherently, in a way that cannot be described by a slight curvature of the critical curve.

The middle panel of Figure 8 shows that detecting more caustic crossing stars enhances the astrometric sensitivity to subhalo lensing, because an increased number of tracer stars allows for a denser sampling of the critical curve across the width of the giant arc.

In the right panel of Figure 8, we vary the normalization $A_{\text{acc}}$ of the subhalo mass function to assess the sensitivity of the subhalo perturbations to their abundance. We find that if the subhalo surface number density is reduced by a factor of four compared to our fiducial model, there is still a large probability that the lensing imprints of subhalos can be revealed along the giant arc with an astrometric precision of \( \gtrsim 30 \) mas.

Despite the large uncertainty in the subhalo abundance after an extrapolation down to \( m \simeq 10^4-10^6 M_\odot \), the results suggest that the phenomenon we study here should be observationally detectable in standard CDM. These observations can constrain several properties of the subhalos like concentration and tidal radius, which will need more detailed studies.

4. Discussion and Conclusions

We have proposed a novel method to probe the subhalo contents of cluster dark matter halos in the low-mass regime of \( 10^8-10^9 M_\odot \). Although substructure on these mass scales has been numerically studied for galaxy-sized halos (Diemand et al. 2008), state-of-the-art N-body simulations of individual galaxy clusters have not been able to resolve down to that level, corresponding to an impressive subhalo-to-host mass ratio of \( \sim 10^{-7}-10^{-8} \). Probing this wide dynamic range tests structure formation on sub-galactic scales and explores modifications of dark matter from the CDM paradigm.

The proposed method uses astrometry of highly magnified image pairs of stars in caustic-straddling giant arcs to seek subhalo-induced departures from the symmetric positions expected in a smooth fold. The method is different from previously proposed methods to detect substructure induced astrometric anomalies in lensed quasars (see e.g., Metcalf 2002). In particular, we take the advantage that these astrometric perturbations are strongly amplified when subhalos lie close in projection to a critical curve. The identification of image pairs of source stars is aided by their property of aligning along a unique direction of elongation. Microlensing by intracluster stars is not expected to appreciably shift image positions but causes flux changes.

Figure 8. Cumulative distribution functions for the required astrometric precision $\sigma_\theta$ to rule out the fold model at $2\sigma(S = 2)$ for the giant arc in Abell 370. We select image pairs brighter than a threshold magnitude in the JWST’s f150w2 filter. Different curve colors correspond to minimum subhalo masses $\log(m_{\text{min}}/M_\odot) = 4, 6, 8$. Each curve is computed based on 200 random realizations. Left: comparison between fitting the midpoints of the image pairs to a straight line and fitting them to a circle. Middle: comparison between different magnitude limits $m_{\text{AB}} < 31, 32$ for image detection. Right: effect of varying the normalization of the subhalo abundance $A_{\text{acc}}$ (Equation (8)) from the fiducial value $A_{\text{acc}} = 0.08$. (Due to pixelization at 8'' x 8'' resolution in our simulation, $\sigma_\theta$ has an artificial minimum at the level of 8 mas; this is however unimportant compared to a physical signature $\sigma_\theta \sim \text{tens of mas}$.)
variability. This information can help identify image pairs, but one must properly account for the additional flux fluctuation due to microlensing of unresolved stars within a single pixel (see, e.g., Tuntsov et al. 2004).

Cluster member galaxies in subhalos of a mass of \( m \gtrsim 10^{10} M_\odot \) can also cause curvature on the critical curve. However, distortions from these subhalos should be coherent over patches around the critical curve as large as \(~1''\), and are distinguishable from the signatures of smaller subhalos. The lensing influence of nearby galaxies can be modeled and corrected if their optical counterparts are detected.

As a specific example, we studied the giant arc of Abell 370 in the vicinity of one caustic crossing location, by simulating random realizations of subhalos based on a realistic model of halo substructure. We found that with images of \(~5–10\) stars with astrometric precisions of \( \sigma_\theta \approx 20–100\) mas, the astrometric perturbations by subhalos of masses \( m \approx 10^6–10^8 M_\odot \) can most often be detected.

The forthcoming JWST offers improved prospects to detect enough lensed stars with the required astrometry to probe subhalos compared to HST, owing to increased number of bright stars, particularly supergiants, that can be detected at \( 1–4\) \( \mu\)m in typical star-forming galaxies at \( z \approx 1\). For the giant arc of Abell 370, we forecast that \(~5–10\) stars down to \( m_{AB} < 31\) can be measured with \(~10\) hr integrations with the widest JWST filters. Current and forthcoming ground-based giant telescopes can reach higher angular resolution with adaptive optics in the near-infrared and may also be promising for this purpose.

Finding new caustic-straddling galaxies that can probe dark matter substructure would be highly beneficial. The best targets are low-redshift, star-forming galaxies (to maximize the number of detectable stars) with a local convergence \( \kappa_0 \) close to unity and a small shear gradient \( |\partial| \) (to maximize magnification of the stellar flux and the substructure perturbations), and a large arc width that allows sampling a larger portion of the critical curve.

Our study has considered only subhalos orbiting within the lensing cluster, ignoring intervening field halos that are randomly projected along the line of sight. The projected abundance of intervening halos, which have better chances of surviving tidal disruption than subhalos inside clusters, can dominate over that of cluster subhalos by a factor of few in the mass range we are interested in (Despali et al. 2018), and have been considered important in applications of strong lensing reconstruction (Birrer et al. 2017; McCully et al. 2017) and quasar flux ratio tests (Metcalf 2005a, 2005b; Xu et al. 2012, 2015). Simulating lensing by intervening halos requires ray tracing with multiple lens planes, which is beyond the scope of this work. Intervening halos should be included in future work and are expected to enhance and make more easily detectable the effect of substructure we have discussed.

Intracluster globular clusters (GCs) (White 1987; West et al. 1995) observed in nearby clusters (Lee et al. 2010; Peng et al. 2011; West et al. 2011; Alamo-Martinez et al. 2013; D’Abrusco et al. 2016; Lee & Jang 2016) may also generate astrometric perturbations through lensing. Having typical masses of \( 10^4–10^6 M_\odot \), these are on average less massive than halos at the lower end (\(~10^4 M_\odot\)) of the interesting subhalo mass range, but should be more efficient lenses due to their high concentrations. However, their surface density of \(~\text{few} \times 10^{-2} \text{kpc}^{-2} \) at the typical impact parameter of the critical curve (Ramos-Almendares et al. 2018) is two orders of magnitude smaller than for subhalos of \( m \gtrsim 10^6 M_\odot \) (see Figure 4).

Therefore, intracluster GCs are unlikely to be a major source of confusion for detecting dark matter subhalos.

Deep images at high resolution are likely to reveal star clusters and compact star-forming associates in the source galaxy. Near a caustic, their physical size \( l \) limits the maximum magnification factor \( \mu \approx [2(\gamma/D_{\gamma})/d \sin \alpha]^{1/2} /\sqrt{2(1 - \kappa_0)} \) (Venumadhav et al. 2017), or \( \mu \approx 600 \text{(}/\text{pc})^{1/2} \) in the case of Abell 370, at which the elongated image has a length of \(~50\) mas(\(1/\text{pc})^{1/2} \). Depending on the compactness, these structures may be under-resolved or marginally resolved. They may be more easily detectable as they are typically brighter than individual stars, and they are not susceptible to microlensing. These sources may be used as substitutes for individual stars in the astrometric test.

Finally, surface brightness irregularities exist in the source galaxy, which include H II regions, open clusters, features of spiral arms, or dust lanes. Those can cause complications in the detection of individual stars. In multi-epoch observations, those should be distinguishable from individual stars based on flux variabilities, since extended source sizes should quench microlensing. On the other hand, surface brightness irregularities themselves can be used to probe subhalos, in which case a variety of techniques developed for other applications (e.g., strong lensing of sub-mm galaxies) are applicable. Further studies are warranted to assess the impact of surface brightness irregularities near caustics.

Given the above considerations, it would be valuable to have mock telescope images created for ultrabright stars on top of a realistic source galaxy surface brightness profile and with the effect of confusion accounted for. More sophisticated treatment at the level of realistic data analysis is warranted for future work.

To conclude, observing the effect of mass clumps near cluster critical curves is a powerful probe to constrain the physical nature of the dark matter. It is a promising method to robustly test theories of warm dark matter or ultralight boson dark matter that make different predictions for the substructure inside dark matter halos.

The authors thank Hsiao-Wen Chen, Brenda Frye, Simon White, Rogier Windhorst, Barak Zackay, Matias Zaldarriaga for inspiring discussions. J.M. thanks the Institute for Advanced Study for their support during visits.

We made use of the lens model for Abell 370 implemented in Lenstool. We modeled the giant arc based on archival HST images. We computed point-source detection S/Ns for caustic crossing stars using the JWST Exposure Time Calculator.

L.D. is supported by the Institute for Advanced Study by NASA through Einstein Postdoctoral Fellowship grant number PF5-160135 awarded by the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for NASA under contract no. NAS8-03060. T.V. and A.A.K. acknowledge support from the Schmidt Fellowship. T.V. is also supported by the W.M. Keck Foundation Fund. J.M. was partially supported by Spanish MINECO grants AYA2015-71091-P and MDM-2014-0369.
Appendix A
Subhalo Abundance

In this appendix, we detail how we modeled the subhalo content of the host galaxy cluster expected in the standard ΛCDM cosmology.

N-body simulations suggest that dark matter halos are well described by the NFW density profile (Navarro et al. 1996). In galaxy clusters, the NFW profile appears to fit the sum of the dark matter and the baryonic matter (Newman et al. 2013). As a function of the halocentric distance $R$, the density is given by $\rho(R) = \rho_{\text{crit}} \delta_c / (R/R_s)^2 (1 + R/R_s)^2$. Here $R_s$ is the scale radius, $\rho_{\text{crit}} (z) = 3H^2(z)/(8\pi G)$ is the critical density, and the characteristic overdensity $\delta_c$ depends on the concentration parameter $C_{200} = R_{200}/R_s$ through $\delta_c = (200/3)(C_{200}/f(C_{200}))$, where $f(x) \equiv \ln(1 + x) - x/(1 + x)$, and $R_{200}$ is the radius within which the mean density is 200 times the cosmic mean. A commonly adopted characteristic mass is the total mass enclosed within $R_{200}$. For field halos, we use the mean concentration-mass-redshift relation $C_{200} = C_{200}(M_{200}, z)$ found by Ludlow et al. (2016).

In hierarchical structure formation, a cluster-sized halo acquires substructure by accreting smaller halos (Kravtsov et al. 2004). Each subhalo has an initial mass $m_{\text{acc}}$ at the time of accretion and subsequently loses mass to tidal stripping. While the evolved subhalo mass function depends substantially on the host mass, the unevolved mass function in terms of $m_{\text{acc}}$ appears universal (Giovi et al. 2008). This is the picture of “unbiased accretion,” in which the unevolved subhalo specific mass function spatially traces the host density profile (Han et al. 2016) and is parametrized as Equation (8), which becomes inaccurate only in the inner part $R/R_{200} \lesssim 0.1$ (Wetzel 2011; Jiang et al. 2015).

The tidal stripping ratio $m/m_{\text{acc}}$ is subject to large scatter because it is sensitive to the complete subhalo orbital history. In particular, subhalos on eccentric orbits suffer from strong tidal stripping and tidal shocking during pericenter passages (Hayashi et al. 2003). On average, a strong scaling of the tidal stripping ratio with the halocentric distance of $m/m_{\text{acc}} \propto R^3$ with $\beta \sim 1$ is seen. Following Han et al. (2016), we assign a probability for bound mass $m$ given an infall mass $m_{\text{acc}}$.

$$dP(m|m_{\text{acc}}, R) = (1 - f_s) \delta_P(m) dm + f_s \mathcal{N}(\ln \frac{m}{m_{\text{acc}}}, \ln \mu_{\text{acc}}(R), \sigma_{\text{acc}}) d m.$$  \hspace{1cm} (13)

The first term accounts for a finite fraction $(1 - f_s)$ of complete tidal disruption, while the second term describes a random draw for $m/m_{\text{acc}}$ from a log-normal distribution with a mean of $\mu_{\text{acc}}(R) = \mu_0(R/R_{200})^3$ and a constant variance of $\sigma_{\text{acc}}$. To satisfy physical constraints, the log-normal distribution is truncated to $m/m_{\text{acc}} < 1$. For a cluster-sized host halo, we use fiducial values $f_s = 0.56$, $\mu_0 = 0.34$, $\beta = 1.0$, and $\sigma_{\text{acc}} = 1.1$ as suggested by simulations (Han et al. 2016).

Appendix B
Subhalo Lens Model

Many studies have found that substructure halos have systematically higher concentrations than comparable field halos and are subject to larger scatter (Bullock et al. 2001; Diemand et al. 2007; Bartels & Ando 2015). This may be explained by tidal stripping, or may be related to biased halo formation within the proto-cluster. Moliné et al. (2017) quantified the enhancement of subhalo concentration in simulated Milky Way-sized halos, which increases as $\propto R^{-1/2}$ toward smaller halocentric distances. Guided by these findings, we match Moliné et al. (2017) with the prescription for subhalo concentration given in Equation (11). We introduce this prescription as a simple, reasonable model, with the caveat that it is unclear how the results of Moliné et al. (2017) may be extrapolated to cluster-sized host halos and our knowledge about subhalo internal structure is in general poor.

With our fiducial model setup, subhalos are generated in the following way: for a given infall mass $m_{\text{acc}}$ at a given halocentric distance $R$, we draw the bound mass $m$ according to Equation (13), compute $c_{200}$ from Equation (11), compute $r_s$ from Equation (10), and finally tune $m_{\text{acc}}$ and $r_s$ such that the total mass equals $m$.

We now present the analytical results for the lensing effect of a subhalo with our chosen density profile, the truncated NFW profile in Equation (9). The gravitational potential gradient is

$$\frac{d\Phi_{sb}}{dr} = \frac{Gm_{200}}{f(c_{200})r^2} H_t \left( \frac{r}{r_s}, \tau \right).$$  \hspace{1cm} (14)

where $\tau = r/r_s$, and we have defined the auxiliary function

$$H_t(t, \tau) \equiv \frac{\tau^2}{2(1 + t)(1 + \tau^2)^2} [ - 2(1 + \tau^2) + 4 \tau (1 + t) \tan^{-1} \frac{t}{\tau} + 2(1 + t)(\tau^2 - 1) \ln[(1 + t)\tau] - (1 + t)(\tau^2 - 1) \ln(\tau^2 + \tau^2)] > 0. \hspace{1cm} (15)$$

The gravitational potential is given by

$$\Phi_{sb}(r) = - \frac{Gm_{200}}{f(c_{200})r_s} H_2 \left( \frac{r}{r_s}, \tau \right),$$  \hspace{1cm} (16)

where $H_2(t, \tau)$ is a second auxiliary function,

$$H_2(t, \tau) \equiv \frac{\tau}{2t(1 + \tau^2)^2} [ - 2(2\tau^2 + t(\tau^2 - 1)) \times \tan^{-1} \frac{t}{\tau} + \tau (2\pi - 2t(1 + \tau^2)) \ln \left( 1 + \frac{1}{t} \right) - 2(1 + t)(\tau^2 - 1) \ln \left( \frac{1 + t}{1 + \tau} \right) + (1 + 2t - \tau^2) \ln \left( 1 + \frac{\tau^2}{\tau^2} \right) ] > 0. \hspace{1cm} (17)$$

The lensing potential is given by

$$\psi_{sb}(\theta) = - \frac{D_{LS}}{D_{L} D_{S}} \left( \frac{4 \pi G m_{200}}{c^2 f(c_{200})} \right) \int_0^{+\infty} d\eta \ H_2 \left( \sqrt{\eta^2 + \left( \frac{b}{r_s} \right)^2}, \tau \right),$$ \hspace{1cm} (18)

where $c$ is the speed of light in vacuum, and $D_{LS}$, $D_{L}$, and $D_{S}$ are the angular diameter distances to the lens, to the source, and from the lens to the source, respectively. The impact parameter $b$ is the related to the angular impact parameter through...
\[ b = D_L \frac{\alpha_{sh}(x)}{|x|}, \] where \( x \) is the angular displacement vector on the lens plane.

The lensing deflection at an angle \( x \) away from the center of the subhalo is given by

\[
\alpha_{sh}(x) = \frac{x}{|x|} \frac{D_{LS}}{D_S} \frac{4Gm_{200}}{c^2 f(c_{200}) r_s} S \left( \frac{D_L |x|}{r_s}, \tau \right),
\]

where we introduce a dimensionless function

\[
S(\xi, \tau) \equiv \int_0^{+\infty} d\eta H_1(\sqrt{\eta^2 + \xi^2}, \tau) \eta \xi / (\eta^2 + \xi^2)^{3/2}.
\]

Similarly, the lensing convergence is given by

\[
\kappa_{sh}(x) = \frac{1}{2} \frac{D_{LS} D_L}{D_S} \frac{4Gm_{200}}{c^2 f(c_{200}) r_s^2} K \left( \frac{D_L |x|}{r_s}, \tau \right).
\]

There, a second dimensionless function reads

\[
K(\xi, \tau) = \frac{\partial S(\xi, \tau)}{\partial \xi} + \frac{S(\xi, \tau)}{\xi} = \int_0^{+\infty} \frac{\xi^2 d\eta}{(\eta^2 + \xi^2)^2} H_1(\sqrt{\eta^2 + \xi^2}, \tau) + \left( \frac{2 \eta^2}{\xi^2} - 1 \right) H_1(\sqrt{\eta^2 + \xi^2}, \tau),
\]

with our notation for partial derivatives

\[
\left[ \begin{array}{l} \gamma_{sh,1}(x) \\ \gamma_{sh,2}(x) \end{array} \right] = \left[ \begin{array}{cc} -1 / 2 & D_{LS} D_L / D_S \frac{4Gm_{200}}{c^2 f(c_{200}) r_s^2} G \left( \frac{D_L |x|}{r_s}, \tau \right) \cos 2\phi \\ 1 & \sin 2\phi \end{array} \right].
\]

Here \( \phi \) is the orientation angle of the vector \( x \) on the lens plane, and we have introduced another dimensionless function,

\[
G(\xi, \tau) \equiv \frac{S(\xi, \tau) - \partial S(\xi, \tau)}{\xi} = \int_0^{+\infty} \frac{\xi^2 d\eta}{(\eta^2 + \xi^2)^2} \times \left[ \frac{3 H_1(\sqrt{\eta^2 + \xi^2}, \tau)}{\sqrt{\eta^2 + \xi^2}} - H_{1,1}(\sqrt{\eta^2 + \xi^2}, \tau) \right].
\]

### Appendix C

**Simulating Lensing**

In this appendix, we briefly discuss how we simulate the lensing of stars in the vicinity of a cluster lensing caustic. Each 1°024 \times 1°024 field of view centered on the critical curve is divided into \( N_{\text{side}} \times N_{\text{side}} \) square pixels with \( N_{\text{side}} = 128 \). Each pixel has a size 8 mas \times 8 mas.

Each pixel is sampled with the technique of inverse ray shooting (Kayser et al. 1986; Schneider & Weiss 1987; Wambsganss 1990, 1999) on an adaptively refined grid, which is designed to ensure good accuracy near critical curves.

Starting with the entire pixel, we shoot rays at the four vertices and at the center, and compute deflection, the lensing Jacobian matrix, and the magnification factor. If the magnification varies substantially (by more than 20%) across the pixel, or if it changes sign (suggesting crossing of a critical curve), the pixel is further divided into four equal sub-pixels, and rays are shot at their vertices and centers. The algorithm proceeds recursively until either the magnification becomes sufficiently uniform within a sub-pixel, or a maximum number of recursions \( N_{\text{rec}} = 8 \) have been performed.

The source plane is sampled using a technique similar to the polygon mapping method (Mediavilla et al. 2006, 2011). Rays shot on the adaptively refined grid on the image plane are mapped onto the source plane, where they define the vertices of polygons that cover the source plane. Special sub-pixels that straddle critical curves on the image plane are further divided into parts on either side of the critical curve and are then mapped to polygons on the source plane. The magnification factor is then interpolated from the values at the vertices. For a source of finite size, such as a stellar disk, we calculate the fraction contained within each polygon (either the source is contained within a single polygon or it straddles several neighboring polygons) and compute the flux using the interpolated magnification. For each polygon on the source plane, we efficiently search through all source stars by organizing them into a hierarchical tree according to their positions (Wambsganss 1999). This method enables to correctly compute the magnifications of small but finite sources by adaptively shooting a realistic number of rays.

**ORCID iDs**

Liang Dai \( \text{https://orcid.org/0000-0003-2091-8946} \)

Tejaswi Venumadhav \( \text{https://orcid.org/0000-0002-1661-2138} \)

Alexander A. Kaurov \( \text{https://orcid.org/0000-0003-0255-1204} \)

**References**

Abell, G. O. 1958, ApJS, 3, 211

Ade, P. A., Aghanim, N., Arnaud, M., et al. 2016, A&A, 594, A13

Alamo-Martinez, K. A., Blakeslee, J. P., Jee, M. J., et al. 2013, ApJ, 775, 20

Asadi, S., Zackrisson, E., & Freeland, E. 2017, MNRAS, 472, 129

Baltz, E. A., Marshall, P., & Oguri, M. 2009, JCAP, 0901, 015

Bartels, R., & Ando, S. 2015, PhRvD, 92, 123508

Bézecourt, J., Ellis, R. S., Soucail, G., & Kneib, J. 1999, A&A, 351, 433

Bézecourt, J., Kneib, J., Soucail, G., & Ebbels, T. 1998, arXiv:astro-ph/9810199

Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton, NJ: Princeton Univ. Press)

Birrer, S., Amara, A., & Refregier, A. 2017, JCAP, 5, 037

Birrer, S., Welschen, C., Amara, A., & Refregier, A. 2017, JCAP, 1704, 049

Blumenthal, G. G., Faber, S., Primack, J. R., & Rees, M. J. 1984, Natur, 311, 517

Blumenthal, G. R.,Pagels, H., & Primack, J. R. 1982, Natur, 299, 37

Bode, P., Ostriker, J. P., & Turok, N. 2001, ApJ, 556, 93

Bullock, J. S. 2010, arXiv:1009.4505

Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, MNRAS, 321, 559

Carlberg, R. 2009, ApJL, 705, L223

Carr, B. J. 1975, ApJ, 201, 1

Carr, B. J., & Hawking, S. W. 1974, MNRAS, 168, 399

Colin, P., Avila-Reese, V., & Valenzuela, O. 2000, ApJ, 542, 622

Conroy, C., Gunn, J. E. 2010, ApJ, 712, 833

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Contini, E., De Lucia, G., Villalobos, A., & Borgani, S. 2014, MNRAS, 437, 3787

Cooper, A. P., Gao, L., Guo, Q., et al. 2015, MNRAS, 451, 2703

Covone, G., Kneib, J-P., Soucail, G., et al. 2006, A&A, 456, 409
