UNIVERSITY OF CALIFORNIA

Los Angeles

Investigating the Nature of Dark Matter with Strong Gravitational Lensing

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Physics

by

Daniel Alejandro Gilman

2020

ABSTRACT OF THE DISSERTATION

Investigating the Nature of Dark
Matter with Strong Gravitational Lensing

by

Daniel Alejandro Gilman
Doctor of Philosophy in Physics
University of California, Los Angeles, 2020
Professor Tommaso L. Treu, Chair

Dark matter makes up most of the mass in the Universe, and yet its particle nature remains unknown. Structure formation arguments provide a promising avenue to address this confounding mystery, as the mass and formation mechanism of the dark matter manifests in the abundance and density profiles of dark matter halos. Measurements of the halo mass function and the mass-concentration relation can therefore be cast as direct constraints on the particle nature of dark matter itself.

Strong gravitational lensing by galaxies offers a unique probe of dark matter structure across cosmological distance, circumventing the use of luminous matter to trace the underlying dark matter. Observables from strong lens systems, particularly the image magnifications in quadruply-imaged quasars, probe the halo mass function directly on sub-galactic scales, below 10^8 solar masses. In this low-mass regime, where halos become devoid of stars and gas, various dark matter models make unique predictions that lensing can constrain.

In this dissertation, I present the development and implementation of a forward modeling framework that constrains any model based on dark matter theory, provided the model predicts the form of the halo mass function, and the density profile of individual halos. Using the framework I developed, my thesis presents an unprecedented constraint on the free-streaming length of dark matter that corresponds to a lower limit of 5.2keV on the mass of a thermal relic dark matter particle. In addition, I present the first constraint on the mass-

concentration relation of Cold Dark Matter halos on sub-galactic scales across cosmological distance. The flexibility of the framework I developed broadens the scope of strong-lensing analyses to any structure formation model based on dark matter theory, underscoring the power of strong gravitational lensing as a probe of fundamental physics.

The dissertation of Daniel Alejandro Gilman is approved.

Steven R. Furlanetto

Smadar Naoz

Alexander Kusenko

Tommaso L. Treu, Committee Chair

University of California, Los Angeles

2020

To my parents, Ivelisse and David, for their unwavering love and support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Scientific publications present research as a straight-line progression from questions to conclusions. In contrast, a PhD program feels like a winding journey, full of unexpected lessons learned along the way. Throughout this interesting time, I have benefited from the guidance of several amazing scientists.

My first research experience was with Prof. Sean Scully, while I was still an undergraduate. During my senior year, Francis-Yan Cyr-Racine, Leonidas Moustakas, and Chuck Keeton brought me to do research with them at JPL, during which time I was introduced to the subject matter described in this dissertation. The opportunity to collaborate with more experienced scientists at these early stages was invaluable. I am grateful for the opportunity they gave me, and their continued support today. In particular, since day one I have collaborated closely with Anna Nierenberg. Learning from her over the last few years has made me a better scientist.

I was truly fortunate to have the chance to work closely with Simon Birrer, who started mentoring me at UCLA almost immediately after finishing his own PhD. I imagine it was a challenge to transition so quickly from student to supervisor, but it seemed natural for him. If I am ever put in a mentoring role myself and am unsure how to proceed, I'll just think to myself: What would Simon do in this situation?

It is difficult to convey just how much I admire my PhD advisor, Tommaso Treu. The astronomy community at large regards him as a creative and talented scientist. To me, he is also a role model, and a leader. I could not have asked for a better advisor.

Finally, I am thankful for the support from my amazing parents. It must have been interesting listening to me try to choose a career path: from a pilot, to joining the foreign service, and then finally setting down with astrophysics. Through it all, you challenged me and believed in me. As I get older, I realize more and more how lucky I am to have you two as my parents - I could not have done it without you!

VITA

2014 Bachelor of Science in Physics, Cum Laude, Department of Physics and Astronomy, James Madison University

2016 Master of Science in Physics, Department of Physics and Astronomy, University of California, Los Angeles

PUBLICATIONS

Gilman, D., et al. Constraints on the mass-concentration relation of cold dark matter halos with 11 strong gravitational lenses. MNRAS in press (2019)

Gilman, D., et al. Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with 8 quadruple-image strong gravitational lenses. MNRAS 491, 6077-6101 (2019)

Gilman, D., et al. Probing dark matter structure down to 10^7 solar masses: flux ratio statistics in gravitational lenses with line of sight halos. MNRAS 487, 5721-5738 (2019)

Gilman, D., et al. Probing the nature of dark matter by forward modelling flux ratios in strong gravitational lenses. MNRAS 481, 819-834 (2018)

Gilman, D., et al. Strong lensing signatures of luminous structure and substructure in early-type galaxies. MNRAS 467, 3970-3992 (2017)

CHAPTER 1

Introduction

All is not well in cosmology. Dark matter, an entity of unknown origin and particle nature, makes up approximately 80% of the mass in the Universe. Gravity mediates the only known interaction between us and this mysterious substance. Through gravity, dark matter pulls the strings of cosmic structure formation, explaining the formation and evolution of galaxies like the Milky Way. Through gravity, dark matter also bends the path of light.

Approximately 9 billion years ago, four photons were ejected from the quasar WFI 2033-4723. Left to fly freely, a distance of roughly $100,000$ light years would separate them at the present time; instead, we find the photons collected today on the primary mirror of the Hubble Space Telescope. The gravitational field of a massive galaxy and its surrounding dark matter, precisely aligned between us and WFI 2033-4723, intervened to deflect light emitted from the distant quasar. The warping of space caused by the foreground galaxy is so extreme that several different paths connect Hubble with the background source, creating multiple images of the quasar at various positions on the sky. This phenomenon is referred to as strong gravitational lensing. Figure 1.1 shows six examples of quadruple-image strong lenses, or *quads*, including the system WFI 2033-4723.

The positions and magnifications of the multiple images in quads encode information regarding the abundance and density profiles of dark matter halos between us and the lensed background source. If the reigning cosmological theory of Cold Dark Matter (CDM) is correct, countless gravitationally-bound structures, or *halos*, litter the cosmos and produce gravitational lensing effects. A positive detection of these otherwise-undetectable dark matter halos through lensing would confirm a central prediction of CDM. Alternatively, some models based on dark matter theory predict that halos do not exist below a certain mass

Figure 1.1: Six quadruple-image strong gravitational lens systems imaged by the Hubble Space Telescope (Nierenberg et al., 2020). The main lensing galaxy is visible as the faint object encircled by four highly-magnified images of a background quasar. Image credit: NASA, ESA, A. Nierenberg (JPL) and T. Treu (UCLA)

scale, which itself depends on the formation mechanism and mass of the dark matter particle(s). A non-detection of halos below a certain mass scale would therefore potentially overthrow entire classes of models, including CDM, that predict a plethora of dark matter structure in the Universe. Either result would bring us one step closer towards understanding the nature of dark matter, and resolving one of the most confounding problems facing modern cosmology.

In this dissertation, I describe research that constrains particle theories of dark matter with strong gravitational lensing. The following sections of this introduction set the stage for Chapters 2-6, which describe the development and implementation of a technique to constrain any dark matter model using the image magnifications from a sample of quads.

Section 1.1 begins with a review of how the particle nature of dark matter drives structure formation in the Universe, and what aspects of structure formation lensing can constrain. Next, I review the basic theory connecting dark matter structure to lensing observables.

1.1 Structure formation and dark matter physics

An initially diffuse field of dark matter particles will collapse into gravitationally-bound halos through a mechanism called 'violent relaxation' (Lynden-Bell, 1967). The halo mass function, or the number of halos per unit mass, encodes information about when the first dark matter halos collapsed in the early Universe. Similarly, the density profiles of individual halos as a function of mass, the mass-concentration relation, depends on the hierarchical assembly of dark matter halos through cosmic time, and the shape of the primordial matter power spectrum that seeded the growth of structure. The particle nature of dark matter affects both the initial matter power spectrum and the growth of density fluctuations initialized at early times, imprinting clues regarding the particle nature of dark matter in the large and small-scale structure of the Universe.

As a concrete example, consider two competing classes of dark matter models: Cold, and Warm Dark Matter (CDM and WDM, respectively). A quantity called the free-streaming length $\lambda_{\rm FS}$ distinguishes these two models. By definition, free-streaming effects are negligible in CDM. In WDM scenarios, diffusion of dark matter particles out of potential wells initialized in the early Universe wipes out small-scale density fluctuations. This diffusion process transforms a density field initialized with a scale-free power spectrum $P(k) \propto k^n$ into a density field with a power spectrum truncated at a scale $k_{\rm FS} = \frac{2\pi}{\lambda_{FS}}$. The characteristic length scale $\lambda_{\rm FS}$ can be approximated as the comoving distance a particle could have traveled before structure begins growing in earnest around the time of matter-radiation equality $t_{\rm EQ}$ (Schneider et al., 2012)

$$\lambda_{FS} \approx \int_0^{t_{\rm NR}} \frac{cdt}{a(t)} + \int_{t_{\rm NR}}^{t_{\rm EQ}} \frac{v(t)\,dt}{a(t)} \approx r_H(t_{\rm NR}) \left(1 + \frac{1}{2}\log\frac{t_{\rm EQ}}{t_{\rm NR}}\right), \tag{1.1}$$

where the particle has speed c before becoming non-relativistic at time t_{NR}, $r_H(t_{\rm NR})$ is the

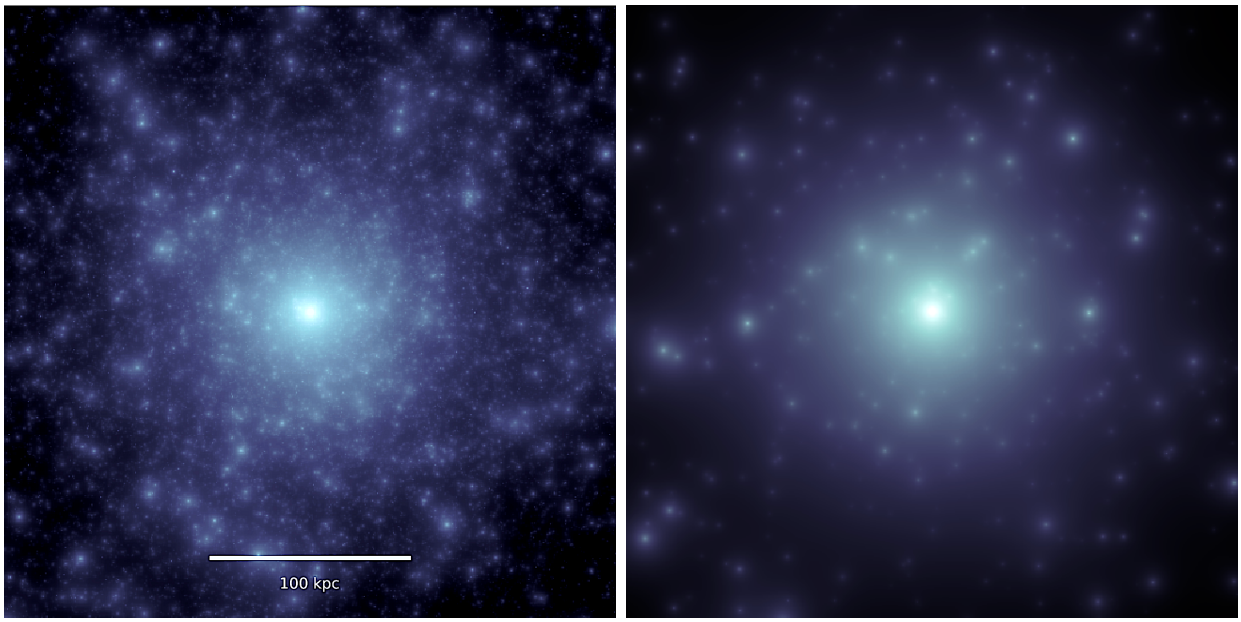

Figure 1.2: **Left:** A realization of CDM substructure, with a scale-free subhalo mass function. **Right:** A realization of WDM substructure corresponding to a 3.3keV thermal relic dark matter particle, which produces a turnover in the halo mass function around 10^8 solar masses.

comoving horizon size at $t_{\rm NR}$, $a\,(t)$ is the cosmological scale factor, and $v\,(t)$ represents the average velocity distribution of the dark matter particles[1].

The effects of free-streaming manifest in structure formation in two ways: First, erasing small-scale power at early times eliminates the small-scale density fluctuations in the primordial matter density field that would eventually collapse into the smallest dark matter halos. This suppression of small-scale power results in a turnover in the halo mass function at a certain mass scale that is proportional to the $k_{\rm FS}^{-3}$ (Avila-Reese et al., 2001; Schneider et al., 2012; Lovell et al., 2014). Second, because low-mass halos collapse first in hierarchical structure formation scenarios, eliminating the smallest halos delays the onset of structure formation. As the central density profile of a dark matter halo reflects the background density of the Universe at the time of collapse, delaying structure formation suppresses the central density of dark matter halos. Mergers between low-mass halos into larger halos propagate these effects to larger halo masses, affecting structures over an order of magnitude in mass above the scales that are directly impacted by free-streaming effects (Navarro et al., 1996; Bose et al., 2016). These structure formation arguments apply to both isolated halos in the field, and subhalos of the 10^{13} solar mass host dark matter halos that contain early-type galaxies typically acting as strong lenses (Gavazzi et al., 2007). The left and right panels of Figure 1.2 show examples of CDM and WDM subhalo populations, respectively.

The dependence of λ_{FS} on features such as t_{NR} and $v\,(t)$ links the free-streaming length of the dark matter to the formation mechanism and velocity distribution of the dark matter particle(s). As the halo mass function and mass-concentration relation depend on the free-streaming length, it follows that constraining the halo mass function and halo density profiles can be cast as a constraint on fundamental dark matter physics determining t_{NR} and $v\,(t)$. Notice, however, that none of the previous discussion depends on a particular choice dark matter particle. We can simultaneously rule out neutrinos, thermal relics with mass < 2keV, and sterile neutrinos produced via Higgs decay with a mass of 7keV (Viel et al., 2013; Abazajian & Kusenko, 2019) making up 100% of the dark matter, as the free-streaming

[1]This expression assumes that the particles become non-relativistic before $t_{\rm EQ}$, and uses the fact that $a\,(t) \propto t^{\frac{1}{2}}$ before $t_{\rm EQ}$.

lengths corresponding to each these models precludes the formation of galaxies such as the Milky Way. Properties like the free-streaming length can be computed for practically any model in the literature, illustrating the broad scope and power of structure formation arguments.

In order to employ structure formation arguments, one requires a method to detect, and measure the mass of, dark matter halos. One approach uses the fact that galaxies are believed to reside inside dark matter halos; luminous structures, such as galaxies, therefore trace the underlying dark matter. Unfortunately, the approach of using luminous matter as a proxy for invisible halos becomes increasingly difficult below 10^9 solar masses, as not every halo on these scales hosts a visible galaxy. Moreover, uncertainties that stem from astrophysics on sub-galactic scales that determine how one assigns a dark matter halo mass to an observed galaxy can sometimes be larger that the differences between predictions from the dark matter models of interest (Nierenberg et al., 2016). While recent advances in this field make considerable progress towards appropriately dealing with these complications (Nadler et al., 2019), the systematic uncertainties persist. A second technique to probe small-scale structure in the Universe relies on the flux power spectrum of the Lyman-α forest at $z \sim 5$, which under certain assumptions can be used as a proxy for the matter power spectrum (Viel et al., 2013; Iršič et al., 2017). The promise of this method must be weighed against the systematic uncertainties associated with thermodynamic processes relevant to the Lyman-α forest, which can mimic the suppression of small-scale power predicted in WDM scenarios (Garzilli et al., 2019).

Strong gravitational lensing by galaxies offers an alternative, more direct probe of dark matter structure on scales below 10^8 solar masses. Lensing couples only to gravity, and therefore circumvents the challenges associated with using baryonic matter to trace the underlying dark matter. In the next section, I review the formalism connecting lensing observables to populations of dark matter halos along the entire line of sight from the observer to the source.

1.2 Strong lensing signatures of dark matter halos

General relativity relates the deflection angle of a light ray to the mass distribution of a massive structure, dark or luminous. When the distances scales between the observer, lens, and source are much greater than the physical extent of the lensing mass distribution, the effect of a massive deflector can be approximated as a single sharp deflection in the plane of the lens, the 'thin lens' approximation. Defining $\Sigma\left(\vec{\xi}\right)$ as the projection of a deflector's three dimensional density profile onto the plane of the lens at the coordinate $\vec{\xi}$, the deflection angle is given by (Blandford & Narayan, 1986)

$$\vec{\alpha}\left(\vec{\xi}\right) = \frac{4G}{c^2}\int \frac{\left(\vec{\xi}-\vec{\xi'}\right)\Sigma\left(\vec{\xi'}\right)}{|\vec{\xi}-\vec{\xi'}|^2} d^2\xi'. \tag{1.2}$$

For multiple deflectors in a single lens plane, the cumulative effect is a linear superposition of their individual deflection angles $\vec{\alpha}$.

A strong lens system will include both subhalos associated with the host dark matter halo of the lensing galaxy, and field halos distributed along the entire line of sight. Incorporating field halos requires the multi-plane ray tracing equation, which maps an angular coordinate on the sky $\vec{\theta_1}$ to an angular coordinate on the source plane $\vec{\theta_s}$. The ray tracing equation that determines where images appear to the observer is given by (Blandford & Narayan, 1986)

$$\vec{\theta_s} = \vec{\theta_1} - \frac{1}{D_{\rm s}}\sum_{i=1}^{s-1} D_{\rm is}\vec{\alpha_{\rm i}}\left(D_{\rm i}\vec{\theta_{\rm i}}\right), \tag{1.3}$$

where the net deflection angle from all halos at the ith lens plane can be computed with Equation 1.2, $D_{\rm ij}$ is the angular diameter distance from the ith lens plane to the jth, and subscript s identifies the source plane. Equation 1.3 is a recursive equation for the position of deflected light rays at each lens plane. It describes a physical process similar to viewing an image through multiple magnifying glasses in series, coupling deflections produced by objects at different distances.

As gravitational lensing conserves surface brightness (Misner et al., 1973), the (de)magnification of the lensed images is proportional to the ratio of areas in the image and source planes. This factor is computed from the inverse determinant of the jacobian $\left(\det\frac{\partial\vec{\theta_s}}{\partial\vec{\theta_1}}\right)^{-1}$. While the full

expression for the lensing jacobian in the general multi-plane framework (see Blandford & Narayan (1986)) is long and not particularly illuminating, the key point is that the magnification of an image depends non-linearly on derivatives of the lensing deflection angle. Image magnifications are therefore highly localized probes of the mass distribution along the line of sight to strong lenses. While the exact level of perturbation to an image magnification depends on the size of the background source, the mass of the halo, and the position of the image relative to the critical curve, a dark matter halo as small as $10^7 M_\odot$ near a lensed image can induce measurable perturbations on image magnifications for background sources of size $O\,(10)$ pc.

The idea that dark matter halos frequently perturb image magnifications was first put forward in 1997 (Mao & Schneider, 1998b). Since that time, authors have attributed lensing 'flux anomalies', or the consistent failure of smoothly-parameterized mass distributions to reproduce the magnifications ratios observed in quad lens systems[2], to the presence of substructure in the lens system (Metcalf & Zhao, 2002; Dalal & Kochanek, 2002; Fadely & Keeton, 2012; Xu et al., 2012; Nierenberg et al., 2014, 2020). Early studies of strong lensing flux anomalies (e.g. Dalal & Kochanek (2002)) relied on lensed radio emission from the background quasar. This technique has drawbacks that are remedied by the advent of nuclear narrow-line emission from the background quasar as probe of substructure, a method first proposed by Moustakas & Metcalf (2003), and subsequently implemented by Sugai et al. (2007); Nierenberg et al. (2014, 2017, 2020). The use of lensed narrow-line emission has two advantages: First, the nuclear narrow-line region is spatially extended by ~ 50pc (Müller-Sánchez et al., 2011), preventing contamination from microlensing and variability in background source brightness[3], processes that considerably inflate uncertainties in radio flux ratios. Second, narrow-line emission is present in the spectrum of virtually every quasar, expanding the sample size of available lens systems. Recently, the sample size of strong lens systems with measured narrow-line flux ratios increased by nearly a factor of

[2]Since the intrinsic brightness of the source is unknown, the observable quantity is the magnification ratio, rather than the magnification itself.

[3]The light crossing time of the narrow-line region washes out the small-scale variability in the light curve of the background source.

four (Nierenberg et al., 2020).

During my PhD, I developed a flexible and powerful Bayesian inference framework that takes full advantage of flux ratios measured using nuclear narrow-line emission to constrain any dark matter model, provided the model predicts the shape of the halo mass function, and the density profile of individual halos. The methods I developed improve over previous work in two key ways: First, they account for halos along the line of sight to strong lenses, which can sometimes outnumber the subhalos associated with the main deflector. Second, the method naturally accommodates spatially extended background sources, which affect the sensitivity of lensing observables to dark matter halos. The tools I developed delivered one of the tightest constraints on the free-streaming length of dark matter to date (see Chapter 5), independent of and more stringent than those obtained from the Lyman-α forest (Viel et al., 2013), and the first observational constraint on the mass-concentration relation of CDM halos on sub-galactic scales across cosmological distance (see Chapter 6).

This dissertation is organized as follows: Chapter 2 describes a study that quantifies the intrinsic uncertainty associated with smoothly-parameterized lensing mass profiles, irrespective of the dark matter substructure content of the lens system. Chapters 3 and 4 describe the development and testing of the analysis framework I developed to combine a sample of strong lenses to constrain dark matter models. Finally, in Chapters 5 and 6 I present results obtained using methods I developed, constraining free-streaming length of dark matter, and the mass-concentration relation of CDM halos, respectively.

CHAPTER 2

Strong lensing signatures of luminous structure and substructure in early-type galaxies

This chapter was published as Gilman, D., et al. Strong lensing signatures of luminous structure and substructure in early-type galaxies. MNRAS 467, 3970-3992 (2017), and is printed here with minor formatting adjustments.

2.1 Introduction

One of the most robust predictions of cold dark matter models is that galaxy and cluster scale halos should host a large number of subhalos, described by a steep mass function of the form $dn/dM \propto M^{-1.9}$ (Klypin et al., 1999; Mao & Schneider, 1998a). Observational evidence against this prediction would force a revision of the standard model in favor of more exotic kinds of dark matter. For example, dark matter models with non-negligible free streaming lengths, such as keV scale sterile neutrinos are expected to manifest as a cutoff in the subhalo mass function (Colombi et al., 1996; Vogelsberger et al., 2016; Bose et al., 2016; Lovell et al., 2016; Menci et al., 2016).

The standard test of this prediction consists of measuring the abundance of luminous satellites around galaxies such as the Milky Way. Significant efforts over the past decades have shown that indeed the abundance of luminous satellites is lower than what is predicted for subhalos. However, the interpretation of this tension is ambiguous. Low mass subhalos might not exist in sufficient numbers, or could simply not be capable of forming stars, and thus be invisible (Nierenberg et al., 2014, 2016; Gao et al., 2011; Starkenburg et al., 2013; Wetzel et al., 2016; Sawala et al., 2016b; Despali & Vegetti, 2017).

For almost two decades it has been recognized that strong gravitational lensing offers an alternative and potentially very clear observational test of this fundamental cosmological prediction, whereby the properties of dark matter subhalos are probed directly by their impact on the arrival times, positions, and flux ratios of lensed images. A variety of techniques have been developed over the years to carry out these measurements, and applied to a variety of datasets. Broadly speaking, the measurements obtained so far are consistent with cold dark matter predictions, although their sensitivity has been limited by sample sizes and quality of the data. Fortunately, sample size and data quality are rapidly improving, and it is therefore important to explore all sources of potential systematic errors in the applications of this technique.

The goal of this paper is to study the impact of baryonic substructure on the application of the so-called lensing anomalies (in time delays, positions, and fluxes) to the study of dark matter substructure. The term anomalies arises from the standard approach in strong lensing communities where the mass distribution of a galaxy is described as the superposition of a 'smooth' mass distribution representing most of the luminous and dark matter, plus a clumpy distribution of dark substructures typically in the range $10^6 - 10^9 M_\odot$. This approach is motivated by the fact that a simple smooth component is generally sufficient to capture the main features of the lensing observables, while substructure below a certain threshold effectively behaves as smooth for given bakcground source size.

Typically, the positions and arrival time delays between lensed images are reproduced by a smooth lens model, while the ratios of the magnifications (also known as flux ratios) may or may not be recovered (Metcalf & Madau, 2001; Dalal & Kochanek, 2002; Bradač et al., 2004; Xu et al., 2009, 2015). If the observed flux ratios cannot be recovered with 'smooth' lens models, the flux ratios are deemed anomalous, and the discrepancy is attributed to the presence of a compact, massive perturbing mass near an image, such as a dark subhalo. Similarly, the inability of smooth models to reproduce image arrival times and astrometry (both for compact and extended sources) gives rise to the so-called time delay and astrometric anomalies (Chen et al., 2007; Keeton & Moustakas, 2009). Both astrometric and flux ratio anomalies have been used to characterize the distribution, abundance, mass function, and

density profile of subhalos (Metcalf & Madau, 2001; Dalal & Kochanek, 2002; Chiba, 2002; Vegetti & Koopmans, 2009; Vegetti et al., 2012; Fadely & Keeton, 2012; Vegetti et al., 2014; Nierenberg et al., 2014; Hezaveh et al., 2016b).

However, the presence of dark subhalos is not the only possible explanation for the observed anomalies. Stellar microlensing (Schechter et al., 2003) and matter along the line of sight (Metcalf, 2005; Xu et al., 2012) can give rise to anomalies in the positions and flux ratios of compact sources. The astrophysical noise from these features can be mitigated by observing sources that are sufficiently extended to smooth away microlensing, by observing at wavelengths unaffected by dust, and by carrying out multiplane lensing analysis.

In this study we focus on astrophysical noise arising from inhomogeneities in the stellar mass distribution of the lensing galaxy that may not be resolved at typical lens redshifts, and could potentially cause anomalies that could be conflated with the presence of dark subhalo. A clear and recent example is given by Hsueh et al. (2016), who show that the apparent flux ratio anomaly in the system B1555 can be readily explained by the presence of an elongated disk in the deflector, which is detected in high resolution imaging of the system.

This potential noise term was recognized early on. For example, Mao & Schneider (1998b) and Chiba (2002) calculated the impact of globular clusters based on simple analytic models. Möller et al. (2003) highlighted the importance of disk components in the statistics of flux ratios, considering their occurrence in early-type galaxies within nearby galaxy clusters. With improvements in sample size and data quality it is important to revisit theses issues and perform quantitative, systematic, and realistic calculations of the overall distribution of the anomalies induced by the stellar component on arrival times, positions, and fluxes of the multiple images. In this context, using numerical simulations, Xu et al. (2010) have shown that the density profiles in the vicinity of the Einstein radius of simulated galaxies are not as simple as those traditionally used to model galaxy-scale lenses, which could amplify the impact of the baryonic mass component of a lens.

In this work, we address this problem by using real Hubble Space Telescope (HST) observations of nearby galaxies to build mock lenses with realistic baryonic mass distributions,

and varying degrees of morphological complexity. We complement this baryonic mass component with an NFW dark matter halo, omitting dark substructure in order to isolate the effect of luminous matter. From the degree to which flux ratios from our mock lenses can be recovered with smooth lens models, we quantify the anomalies that can be attributed to the baryonic mass of a deflector (we identify stars with baryons but neglect the contribution of gas, which is assumed to be smooth on the relevant scales).

This paper is structured as follows. In Section 2.2, we detail our procedure for building mock lenses from HST images of nearby galaxies, the type of lens models considered in this work, and our fitting methodology. In Section 2.3, we present the results of our comparison between smooth models and realistic simulated lenses. In Section 2.4, we summarize the results of our analysis, and discuss the lessons learned in the context of ongoing and future strong lensing studies of dark matter. When needed to compute distances, we adopt a standard concordance cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $h = 0.7$, even though our results are independent of this choice. All of the lens simulations, ray-tracing and computation of lensing observables (positions, time-delays, magnifications) are performed using the `lensmodel` software (Keeton, 2011).

2.2 Building and fitting mock lens systems

In this Section we describe in detail our procedure to build mock lens systems and then fit them with lens models. We begin by describing our source of high resolution images about the surface brightness of early-type galaxies in Section 2.2.1. In Section 2.2.2 we summarize how we obtain the global structural parameters for the lens galaxies, either from the literature or our own fits to the light. In Section 2.2.3 we describe how we convert surface brightness into lensing potential, accounting for the dark matter halo and external shear. In Section 2.2.4 we describe the ingredients of our five different mass models used to produce mock lenses and fit them. In Section 2.2.5 we describe the process of generating data sets with our mock lenses for two of our models that are derived from the real HST images, and in Section 2.2.6 we describe the process of fitting two smooth lens models to data obtained

from the mock lenses.

2.2.1 The stellar surface brightness of early-type galaxies at high resolution

The starting point for our mocks is archival Hubble Space Telescope observations early-type galaxies from the nearby Virgo and Coma clusters (Ferrarese et al., 2006; Carter et al., 2008). In order to obtain a sample that is representative of lensing galaxies we select all the elliptical and lenticular galaxies with available HST images, central velocity dispersions between 165 and 320 km s^{-1}, and ellipticities in the range 0.05-0.43. We limit our selection to galaxies imaged with the Advanced Camera for Survey with filters F814W or F850LP, in order to minimize the effects of dust, and map the stellar light as closely as possible, while taking advantage of the wider field of view of view and finer pixel scale than the infrared channel of Wide Field Camera 3. The sample displays a variety of interesting features, including globular clusters, disks, tidal tails and shells, which we take as representative of the kind of baryonic structure and substructure that we are interested in studying. In Table 2.1, we list the galaxies used in our data set, along with their relevant physical parameters.

We avoid galaxies with prominent dust lanes, and sources of visual contamination obvious to the naked eye, as these features would be problematic in our procedure for assigning mass to light, which we discuss in the next section. There are often bright galaxies or stars in the line of sight, which we replace with a smooth interpolation of the main lens profile. We do not expect this to significantly affect our results, however, as we avoid generating lenses where an image would be located near one of these defects. When the computation of the lensing potential, according to our normalization procedure, requires information from pixels outside the ACS field of view, we extrapolate a smooth model fit to the light into these regions. After solving the lens equation, we ensure that no lensed images land in an interpolated region.

We note that real lens samples tend to be dominated by high velocity dispersion galaxies above 240 kms^{-1}(Auger et al., 2010; Sonnenfeld et al., 2013), due to their favorable lensing cross section. Surveys of high velocity dispersion galaxies (Goulding et al., 2016) show that

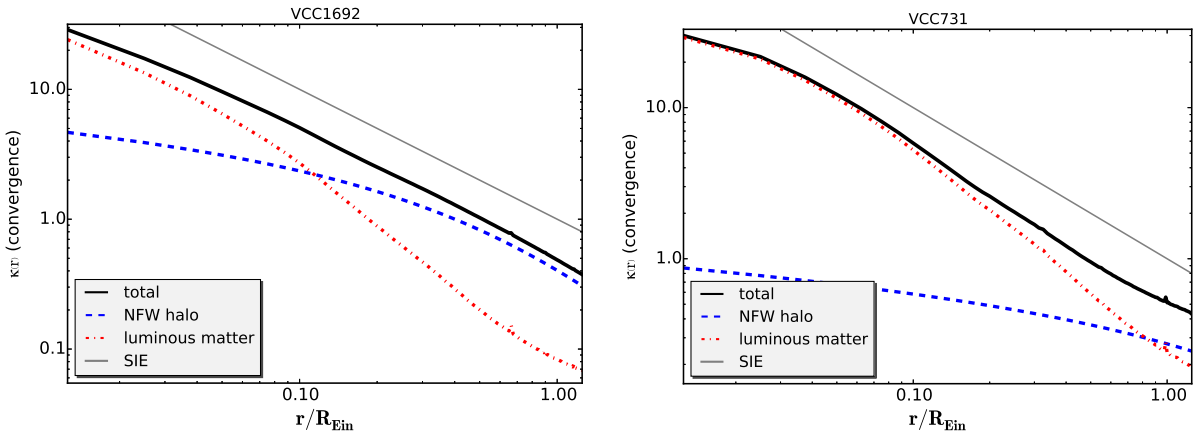

Figure 2.1: Convergence as a function of radius for a normalized map of surface mass density for the deflectors VCC1692 and VCC731. For reference, the slope of an SIE mass density profile (with arbitrary normalization) is shown in grey.

the most massive ellipticals tend to be slow rotators, while low velocity dispersion galaxies, which are more likely to be fast rotators and host disks, are over-represented in our mock sample. As such, our sample is not representative of that of typical lens galaxies, and will likely result in an over-estimate of the contribution to time delay, astrometric, and flux ratio anomalies by the baryonic mass component of a deflector. In light of this, we interpret the fraction of anomalous systems in our analysis as an upper limit to the frequency with which one expects to encounter baryon-induced anomalies in a survey of real lensed quasars.

2.2.2 Structural parameters of the sample galaxies

In order to simulate the lensing properties of the galaxies in our sample, we require a measurement of central stellar velocity dispersion σ_*, half-light radius $R_{1/2}$, ellipticity ϵ, position angle θ_ϵ, and a Sérsic index n for each host galaxy. We draw measurements of the central velocity dispersion from the HyperLeda online catalog (Makarov et al., 2014) and from (Ma et al., 2014), while measurements of the half light radii, ellipticity and position we obtain for Virgo objects from Ferrarese et al. (2006) and from HyperLeda.

When the parameters describing the host light distribution are not available in the literature, we derive them by fitting the light profiles with a single Sérsic component using

`galfit` (Peng et al., 2002) and derive the parameters ourselves, mimicking the efforts of an observer attempting to model the luminous matter of a strong lens. The parameters that we adopt for each galaxy are summarized in Table 2.1.

2.2.3 From surface brightness to surface mass density

We transform the surface brightness maps of the galaxies into maps of surface mass density (convergence) in order to determine the gravitational lensing properties. In translating between surface brightness and surface mass density, we assume that light traces luminous matter in the field of view, with a constant stellar mass-to-light ratio. This is a conservative approach as it will assign higher masses to young star populations which tend to populate disky areas, relative to the old star populations which tend to populate the smooth elliptical component. Thus, by adopting a uniform stellar mass to light ratio we tend to increase the lensing signal of disky structures, consistent with our interpretation of our results as upper limits on the perturbative effect of baryonic structure on lensing data.

For simplicity, we simulate all our systems as they would be observed for typical deflector and source redshifts $z_d = 0.5$ and $z_s = 1.5$. The smooth dark matter component of each deflector is described by a circular NFW halo, whose scale radius R_s is taken to be $5R_{1/2}$, where $R_{1/2}$ is the half-light radius of the target galaxy, in projection. We do not expect this choice for the dark matter normalization to affect our main results, as our choice for R_s simply reflects the different spatial scales over which the smooth dark matter and baryonic mass component vary. While real NFW halos are unlikely to be circular, the NFW halo in our analysis serves only to boost the convergence within the Einstein radius to that of a typical deflector. Further, ellipticity in the NFW halo is, to some extent, degenerate with external shear, which we add as a separate component.

We compute the Einstein radius of each mock lens by exploiting the observational fact (Treu et al., 2006; Koopmans et al., 2009) that in lens galaxies the stellar velocity dispersion σ_* approximates, within a few percent, the velocity dispersion $\sigma_{\rm SIE}$ of the best fitting singular

isothermal ellipsoid (SIE), for which the Einstein radius is given by

$$R_{\rm E} = 4\pi \left(\frac{\sigma_{\rm SIE}}{c}\right)^2 \frac{D_{ds}}{D_s}, \tag{2.1}$$

where D_s, and D_{ds} are the angular diameter distances to the source, and from the deflector to the source, respectively. This equation is one of the consequences of the so-called bulge-halo conspiracy (Treu & Koopmans, 2002, 2004; Koopmans et al., 2006, 2009; Dutton & Treu, 2014): the projected total mass density profile of early type galaxies is well described by a single power law with logarithmic slope -1. As a consistency check, we verify that the total convergence (after adding stellar mass to the light and a dark matter component) of our mock galaxies is well approximated by an isothermal profile, as shown in Figure 2.1. Also, we check that the stellar masses derived from our convergence maps are consistent with those reported by Gallo et al. (2008). Details of the normalization procedure, based on empirical measurements of the relative abundances of stellar mass and dark matter, are given in Appendix 5.8. In order to mimic the tidal field of the large scale structure expected at intermediate redshifts, we add, at random position angles, external shears of magnitude 0.05 or 0.08, which are typical shear magnitudes in strong lens systems (Holder & Schechter, 2003).

2.2.4 Description of the lens models

In order to carry out our quantitative analysis of the lensing effects of unresolved stellar structures, we compare the lens configurations obtained from the high resolution mass maps (the "truth"), with two models based on lower resolution data, and two simply parametrized smooth models commonly used in the literature. The two models based on a low resolution version of the "truth" are intended to simulate the best data that one could hope to extract from a distant lens using HST. The two simply parametrized lens models are meant to represent the models typically used as a reference to detect anomalies due to dark substructure.

Thus, in total, we consider five lens models, with the following characteristics:

- Model 1 (real data) - *"Truth"*. This model directly uses the image of the galaxy

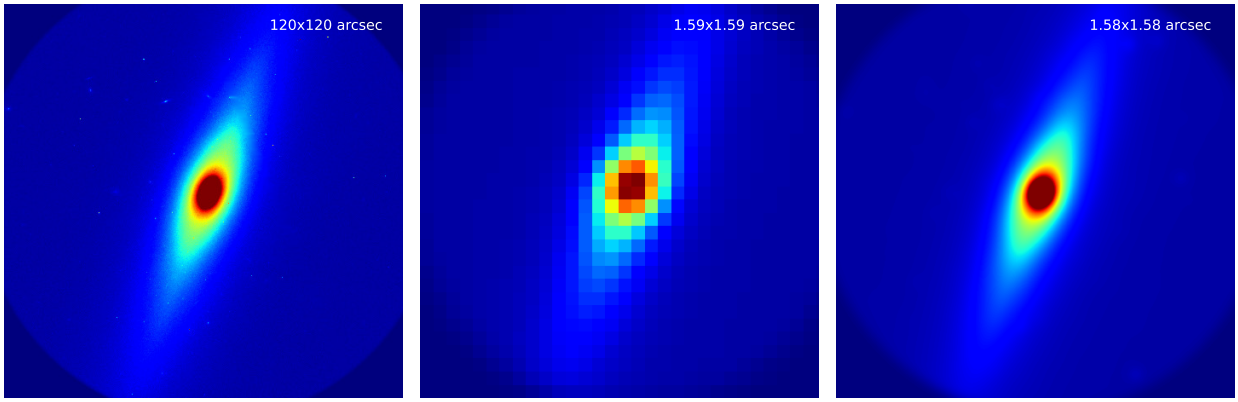

Figure 2.2: *Left:* Surface mass density of VCC1692 as it appears at a distance of 16 Mpc. *Center:* The galaxy as it appears at redshift 0.5 (1280 Mpc) after rebinning pixels 80x80 to account for a loss of spatial resolution.
Right: The galaxy after convolving with a Gaussian PSF with FWHM of 80 pixels (with pixel size the same as in far left panel, with resolution 0.05 arcsec pixel^{-1}) to simulate an observation of the galaxy where sub-pixel information has been recovered via dithering, effectively the best smooth model one could construct given HST data.

obtained by HST, converted to a convergence map as described in the previous section and Appendix 5.8.

We evaluate the following four models by their ability to reproduce the 'real' data of Model 1:

- Model 2 (relies on real data) - *"real HST"*. This is a simulated single exposure of an HST image, including the effects of a Gaussian PSF, and pixelization. First, we rebin pixels of the *Truth* model by a factor corresponding to the loss of spatial resolution going to $z_d = 0.5$ from the native redshift of the galaxy. For example, translating the angular diameter distance of the Virgo cluster ($z = 0.0038$) to $z = 0.5$ changes image resolution by a factor of 80, so the image used in the *Truth* model is rebinned 80x80. We then convolve the rebinned map with a Gaussian Point-Spread-Function (PSF) of FWHM = 2 pixels. We checked that the order of operations of rebinning and convolving does not affect the results. This model is meant to represent an attempt to

fit the stellar mass of the lens by scaling the pixel values observed by HST. An example is shown next to the *Truth* stellar mass distribution in Figure 2.2.

- Model 3 (relies on real data) - *"HST Interpolated"*. This model simulates an HST image where the sub-pixel information has been recovered via dithering, thereby representing the best possible data set obtainable for these systems at a redshift of 0.5, approximating the Hubble PSF with a Gaussian PSF. In effect, this data has been smoothed over at a scale comparable to the Hubble PSF at redshift 0.5, thereby erasing structure on scales smaller than rebinning factor at redshift 0.5, thereby erasing structures on scales < 0.32 kpc for Virgo galaxies, and < 0.51 kpc for Coma galaxies. In practice, this model represents the best description of the stellar component that one could build from HST observations, using a smooth interpolation or a fit to the pixel data. As such, the degree to which this model reproduces the flux ratios from the *Truth* model represents a noise floor for flux ratio data. An example of the stellar mass distribution corresponding to this model is shown in Figure 2.2.

The following two models are different from the previous three, as they are analytic functions fit to the data obtained from the *Truth* model.

- Model 4 (fit to *Truth* positions, time delays) - Singular isothermal ellipsoid with external shear (SIE). This model is physically motivated by the fact that the combined mass profile of baryons and a NFW halo is well approximated by an isothermal power law, as shown in Figure 2.1. We do not include information about image magnification when performing the fit, and use positional and time delay uncertainties of 0.003" and 2 days to simulate the best data currently available.

- Model 5 (fit to *Truth* positions, time delays) - Sérsic + NFW halo (SNFW). We fit an elliptical Sérsic (Sérsic, 1963) mass distribution and a NFW with external shear to image positions and time delays, with the same observational uncertainties as Model 4. The SNFW model has nearly double the number of free parameters as the SIE, which at face value suggests it would be a more adaptable functional form than the former,

and better suited to representing a possibly complex distribution of baryonic and dark matter. However, models with too many free parameters are prone to degeneracies given the limited constraints available. We will consider this point again in Section 2.6. This model is meant to represent a practical approach which might be as close as possible to the best one can do, especially in the presence of bright lensed quasar images.

We stress that because we do not explicitly add dark substructure to our mock lenses, the only source of small scale structures or non-smooth features, akin to the clumpy nature of dark matter substructure, is that of the baryons in the lensing galaxy, luminous satellites of the deflector, and background galaxies. Therefore, any discrepancy in flux ratios between the "Truth" model and models 4-5 is due entirely to a baryonic mass component that cannot be absorbed by the SIE or SNFW functions.

Similarly, with data of extraordinary quality, one could imagine using more flexible and complicated smooth lens models to describe the stellar mass component. This is captured in by the *HST Interpolated* model, which provides a reasonable upper limit on the capability of a smooth lens potential to fully account for the baryonic structure of a lensing galaxy.

2.2.5 Generating mock data sets

For each of the three lens models based on real images (*Truth*, *Real HST*, *HST Interpolated*), we manually place the source position within the astroid caustic so as to produce a cusp and a fold lens configuration. While the light traces mass hypothesis allows us to efficiently normalize and assemble realistic mock lenses, it introduces a significant complication. Shot noise in the HST images and discontinuities due to pixelization cause small scale variation in surface mass density that introduce a small scale pattern in the local magnification map. For a point source this would introduce a microlensing-like signal, which could introduce spurious scatter in the fluxes predicted by the *Truth* model. We avoid this by modeling the background quasar as an extended source 5 parsecs in diameter, a procedure we describe in detail in Appendix 5.9. For reference, this source size is roughly the size of a radio jet source

(1-10 pc), but smaller than the narrow-line region (10-100 pc) (Moustakas & Metcalf, 2003), and is large enough to avoid micro-lensing effects while preserving sensitivity to small scale structure in the image plane, and corresponds to 0.265 mas^2 in the source plane.

For the three mock deflectors (Models 1-3), we apply a Monte Carlo procedure: for each image configuration (cusp and fold), we randomly sample 250 source positions from a circular area in the source plane, centered on a reference source position guaranteed to produce a cusp or a fold lens. For each of the 250 new source positions, for each of the *Truth*, *Real HST*, and *HST Interpolated* convergence maps we directly solve the lens equation to obtain 250 new sets of positions, time delays, and flux ratios. We do not add measurement noise in this process, as we are only interested in the effects of baryonic mass on these data.
For the simply parametrized lens models (Models 4 and 5), we use the software package `lensmodel` to fit an SIE and SNFW model to each of the 250 data sets, corresponding to each of the 250 sampled source positions, constraining the models by only astrometric and time delay data and demanding that the Sérsic halo and NFW halo are centered at the same location. We introduce a χ^2 penalty to discourage `lensmodel` from adopting unphysical characteristics, such as an NFW halo with a scale radius smaller than the stellar half-light radius.

We plot the resulting data for each of our models as histograms that characterize the distributions for each lensing observable, taking into account small variations in the the unknown source position. The scatter in the distributions of the *Real HST* and *HST Interpolated* data we obtain can be attributed to variation in the source position, since the process of rebinning pixels and convolving with a PSF wipes out small scale features in the lensing potential, which could lead to flux ratio perturbations. On the other hand, the variance of the *Truth* data is affected by variations in the source position *and* perturbations from small scale features in the lensing potential, resulting in a systematically larger scatter. To account for this, we interpret significant offsets in the means of these distributions as evidence for flux ratio perturbations by luminous matter.

In Figures 3.4 and 2.5, we show distributions of flux ratios obtained for the 6 lens systems in our mock sample with the largest $R_{\rm cusp}$ or $R_{\rm fold}$ values (see Equations 2.2 - 2.3). The

frequency and magnitude of flux ratio and astrometric anomalies across our full sample of mock lenses, and the physical characteristics that give rise to these phenomena, characterize what properties of lensing galaxies are likely to perturb flux ratios and other lensing data. We will return to interpret the results of these figures in more depth in Section 2.3.

2.2.6 Fitting simply parametrized lens models to mock data

2.2.6.1 Adopted uncertainties

We assume astrometric uncertainties of 0.003 arcseconds, time delay uncertainties of 2 days, i.e. comparable to the best data currently available. For the magnification ratios we adopt uncertainties of a factor of 100 which ensures that we fit the smooth potentials only to image positions and time delays. This approach is motivated by the current standard procedure, where the flux ratios are normally not used as constraints for smooth models in order to bypass the effects of substructure and astrophysical noise arising from dust, microlensing, and variability.

2.2.6.2 Fitting procedure

When fitting the SIE, we vary the Einstein radius, position, ellipticity, shear, and the two corresponding position angles. We optimize these parameters simultaneously, first optimizing numerous random realizations of an SIE profile in the source plane, and then keeping and re-optimizing the best model in the image plane (Keeton, 2011).

In contrast, when fitting with the SNFW, we attempt to fit the lens by holding the parameters of the Sérsic profile describing each galaxy fixed while varying the properties of the NFW halo. We obtain the Sérsic parameters either from literature (see the references in Table 1), or by measuring them ourselves using `galfit`. Specifically, in the first iteration we vary only the normalization of the Sérsic profile, the normalization of the NFW halo, the scale radius of the NFW, and the external shear and position angle.

In most cases, this approach fails to fit the positions and time delays with a reduced

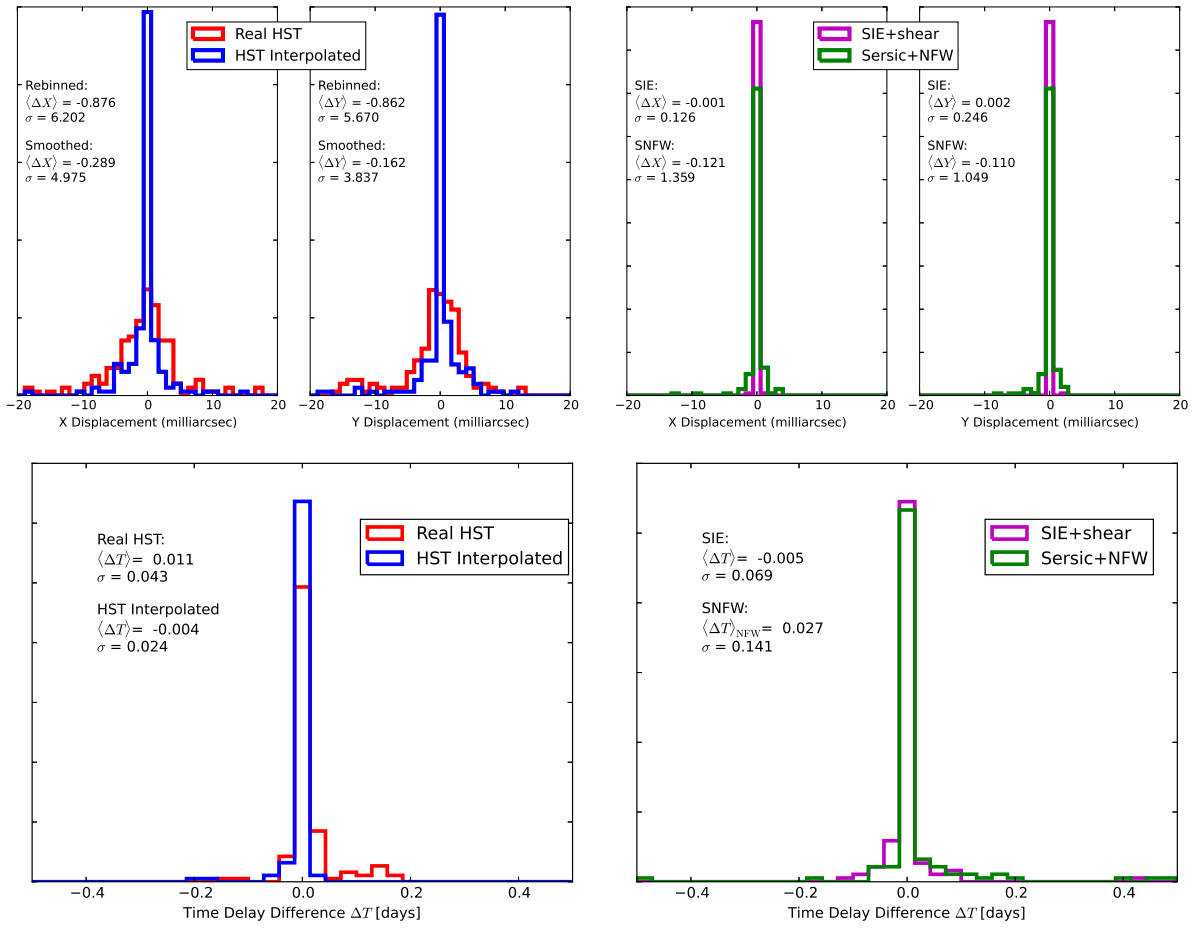

Figure 2.3: Distributions of the difference in positions (top) and times delays (bottom) from the mean of the *Truth* distributions. Standard deviation, denoted by σ is displayed for each data set. The absence of measurement noise in our mock data results in the narrow distributions, whose width is determined by specific lensing properties of each model.

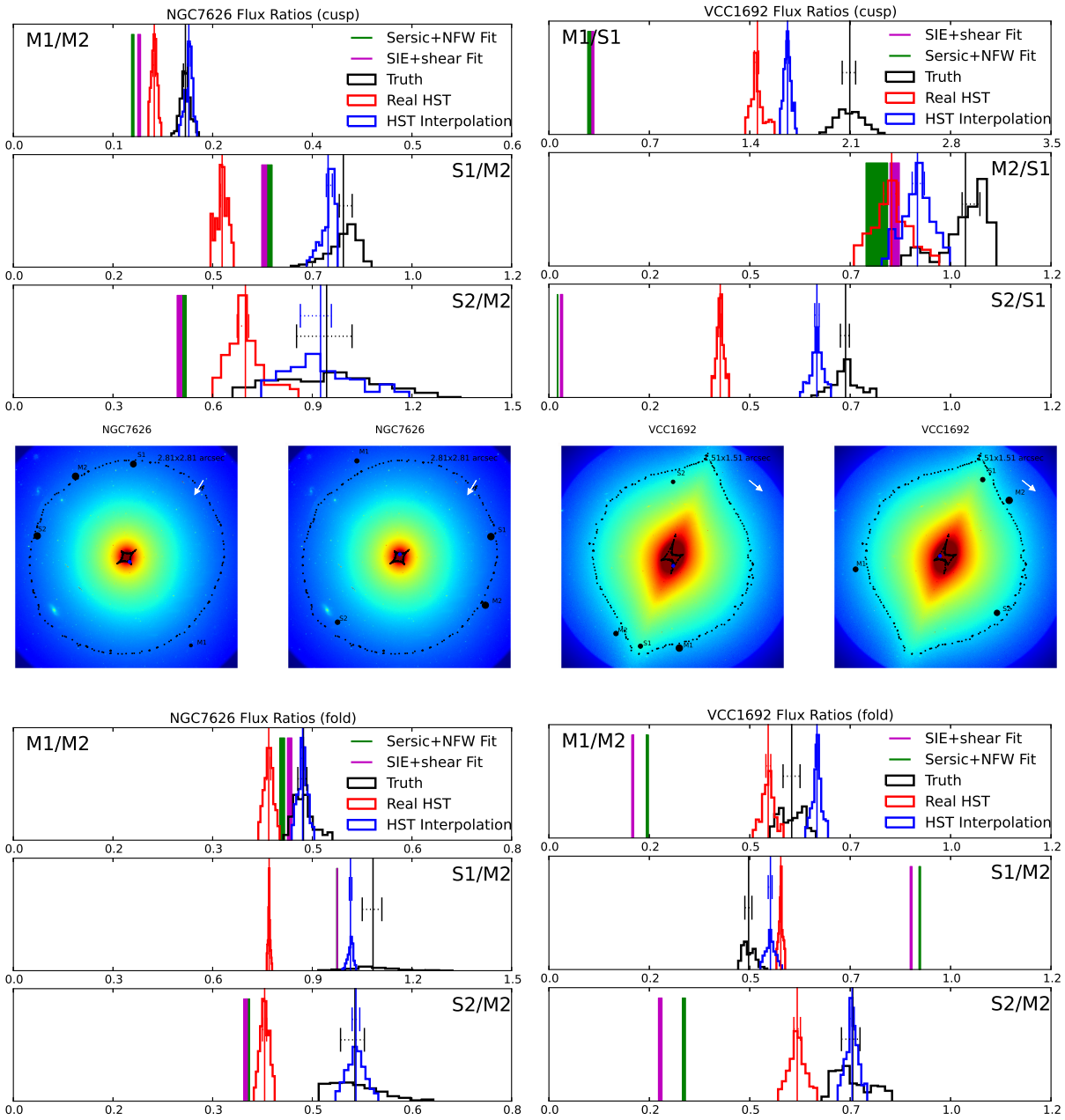

Figure 2.4: Flux ratio distributions for two anomalous systems, NGC7626 and VCC1692. Images are classified as minima (M) or saddle points (S) of the time delay surface. The outer critical curve and astroid caustic are marked by black points, while the source position is marked as a blue point. The line in the upper right corner indicates the direction of the applied external shear. *Left*: The mock lens systems created from NGC7626 *Right*: The mock lens systems created from VCC1692

.

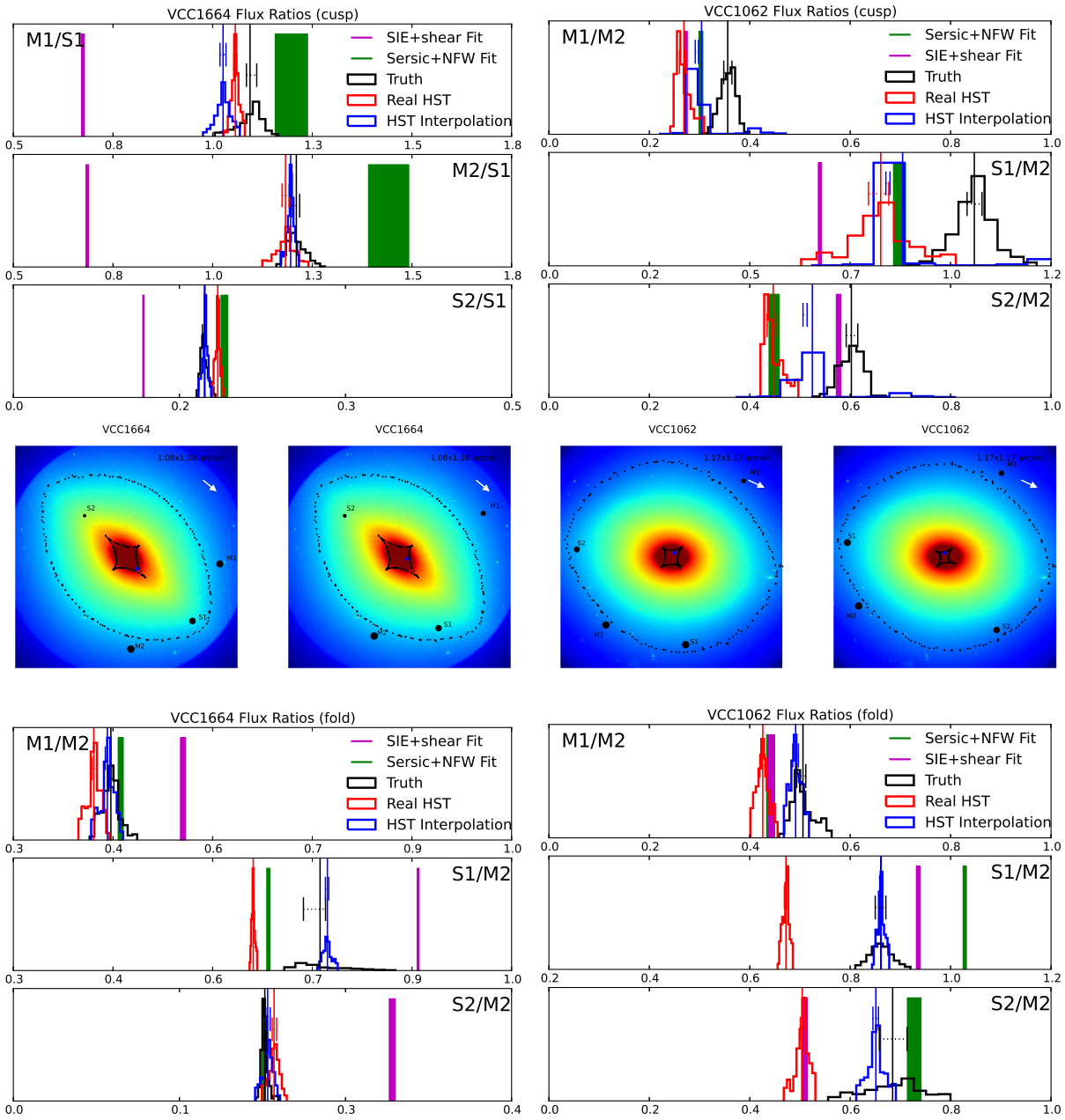

Figure 2.5: Flux ratio distributions for two anomalous systems, VCC1664 and VCC1692. Images are classified as minima (M) or saddle points (S) of the time delay surface. The outer critical curve and astroid caustic are marked by black points, while the source position is marked as a blue point. The line in the upper right corner indicates the direction of the applied external shear. *Left*: The mock lens systems created from VCC1062 *Right*: The mock lens systems created from VCC1692.

$\chi^2 < 2.5$, which we take to be the threshold acceptable χ^2 fit. We choose this χ^2 to permit individual astrometric and time delay χ^2 values greater than 1, resulting in a conservative measure of the degree to which our smooth potentials can recover image positions and time delays. As we do not add measurement noise to our data, the SIE and SNFW frequently fit the data almost exactly, resulting in reduced χ^2 much less than unity. After the first iteration of fitting, for the systems with unacceptable model fits, we allow the Sérsic index, ellipticity and position angle of the Sérsic to vary, and attempt to fit the lens again. If this approach fails, we vary the effective radius, ellipticity and position angle.

The complications we encounter trying to fit lenses with a single Sérsic model suggests that, for the purpose of lens modeling, a more complicated lens model is required to fit the luminous matter of a lens, e.g. a bulge+disk of different ellipticity or two components at different position angles. On the other hand, the success of a simple SIE model suggests that this simple model is sufficient in most cases to caputure the lensing effects of baryonic matter. Interestingly, fold configurations required more flexible models (with varying Sérsic index, ellipticity, and position angle) than cusp configurations. Overall, we find that an SIE model absorbs the combined properties of stellar mass and dark matter as well as, or better than, the SNFW model. In Table 2.2 we summarize the parameters we allow to vary when fitting with the SIE and SNFW models, and the results of the fit for each galaxy in our sample.

2.3 Results

In this section, we compare the data obtained for the *Truth*, *Real HST*, and *HST Interpolated* mock lenses, and the two analytic models, the SIE and SNFW. First, in 3.1 we investigate the extent to which positions and time delays vary between the *Truth* data set and the four comparison models. Section 3.2 reviews the $R_{\rm cusp}$ and $R_{\rm fold}$ statistics, and present the values for these statistics we obtain for the *Truth* model, side by side with observed $R_{\rm cusp}$ and $R_{\rm fold}$ statistics from real strong lenses, and characterize baryonic flux ratio anomalies by their coupling to astrometric anomalies. In 3.3, we examine in detail each anomalous system to

understand the source of flux ratio anomaly, making use of magnification maps derived from our convergence maps. Finally, in 3.4 we discuss the the degree to which the *Real HST*, *HST Interpolated*, SIE and SNFW models recover the flux ratios, $R_{\rm cusp}$ and $R_{\rm fold}$ values of the data of the *Truth* model.

2.3.1 Image positions and time delays

Since time delays and image positions depend on the total gravitational potential, and the gradient of the potential, respectively, one can expect these data to be relatively insensitive to small perturbations to the gravitational potential, whether by dark subhalos or by baryonic features. Therefore, it is expected that the rebinned and smoothed models will yield similar image positions and time delays as the *Truth* model, and that both the SIE and SNFW models we fit to the *Truth* will also accurately recover these data.

Our results are consistent with these expectations. In Figure 2.3 we plot the distributions of offsets between means of the *Truth* model, the three models based on the unfiltered data (Models 1,2,3), and the SIE and SNFW fits to the *Truth* model. The standard deviations in the distributions of image positions are comparable to the 0.003 arcsecond uncertainty we assume in our lens models, while the standard deviations in the distributions of time delays is an order of magnitude smaller than our assumed uncertainty. For the *Real HST* and *HST Interpolated* models, this indicates that the information lost in the process of rebinning pixels, or convolving with the smoothing kernal, does not, in most cases, significantly impact the predicted image positions and arrival times. For the SIE and SNFW fits, these results confirm that the image positions and time delays of real lenses are consistent with those produced by a smooth lens model, as expected.

While these results were expected a priori, this should not undermine their significance. The agreement of astrometric data between the different models we consider implies that luminous matter is highly unlikely to result in astrometric anomalies. Conversely, many real lenses with flux ratio anomalies similar to those we observe in our mocks exhibit flux ratio anomalies accompanied by astrometric anomalies, suggesting an avenue by which a baryonic

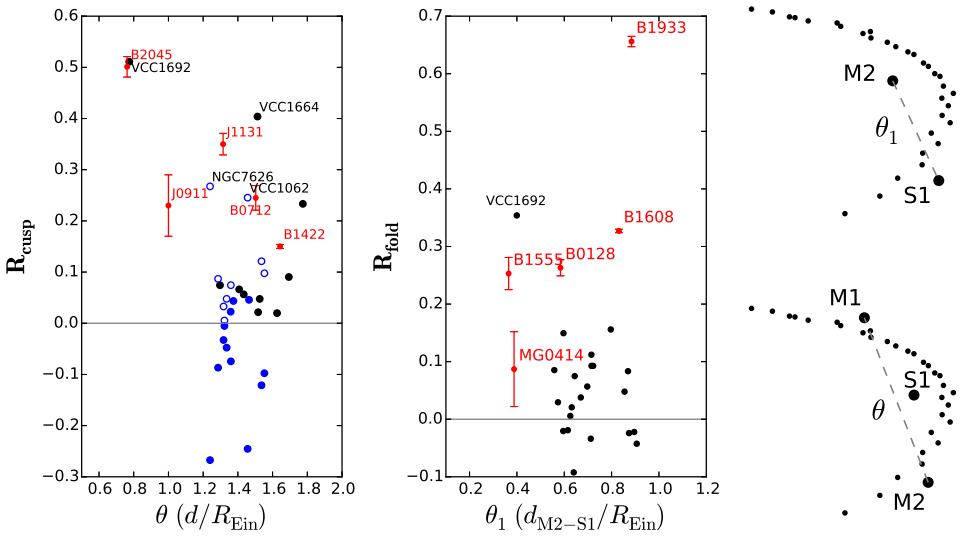

Figure 2.6: $R_{\rm cusp}$ and $R_{\rm fold}$ statistics of the *Truth* data (black), compared with a set of $R_{\rm cusp}$ and $R_{\rm fold}$ statistics of real lenses (red). An empty circle indicates the absolute value of an $R_{\rm cusp}$ that turns out to be negative. Fold (top) and cusp (bottom) image configurations, with critical curves shown as small black dots and images labeled according to their classification as minima or saddle points. The distances between images θ_1 and θ, appearing in the left panels, are shown as dashed grey lines.

lensing signal may be distinguished from other sources of anomaly. We will revisit this point in detail later in this section.

2.3.2 Flux ratios and the $R_{\rm cusp}$ and $R_{\rm fold}$ relations, *Truth* model vs. real lens systems

In order to quantify the flux ratio anomaly across the full ensemble of mock lenses, we consider the differences in the mean of each distribution (explicitly, each F_i denotes the mean of a distribution for an image generated from a lens described by model i). The mean is marked as vertical bar in Figures 3.4 and 2.5. In this work we consider the relative flux ratio anomalies δF_i, with respect to the *Truth* data:

$$\delta F_i = \frac{|F_{Truth} - F_{model}|}{F_{Truth}}$$

We calculate the commonly used $R_{\rm cusp}$ and $R_{\rm fold}$ statistics given by

$$\begin{aligned} R_{cusp} &= \frac{M_1 + M_2 - S_1}{M_1 + M_2 + S_1} \quad \text{(major axis)} \qquad (2.2) \\ &= \frac{M_2 - S_1 - S_2}{M_2 + S_1 + S_2} \quad \text{(minor axis)} \\ R_{fold} &= \frac{M_2 - S_1}{M_2 + S_1} \qquad (2.3) \end{aligned}$$

for each model. Major axis and minor axis cusps are defined by whether the central cusp image is the first saddle S1 (major) or the second minimum M2 (minor). The different configurations are shown on the right hand side of Figure 2.6. It has been shown (Schechter & Wambsganss, 2002; Keeton, 2003) that small, compact deflectors in the lensing galaxy, or local perturbations, tend to suppress the brightness of images appearing on saddle points of the time delay surface, while preferentially magnifying minima. On the other hand, perturbations to the gravitational potential on scales larger than the image separation, or global perturbations, do not discriminate between minima and saddle points. The different responses of these cusp configurations to lens structures suggest they could potentially be used to differentiate between different sources of flux ratio anomalies.

For models 2-5, we compute the offset between the unsigned $R_{\rm cusp}$ and $R_{\rm fold}$ statistics

$$\Delta R_{model} = ||R_{Truth}| - |R_{model}||$$

Smooth lens models will yield values close to zero in the limit of vanishing distance between neighboring images, with small variations depending on the image separation and properties of the main lens model (Keeton et al., 2003, 2005). Large values of $R_{\rm cusp}$ and $R_{\rm fold}$ are typically associated with perturbations to the lens potential on scales smaller than the image separation, and as such, these statistics are often used as a indicators of small scale structure near a lensed image.

2.3.2.1 Distribution of $R_{\rm cusp}$ and $R_{\rm fold}$ for target galaxies

In Figure 2.7, we plot the distributions of the $R_{\rm cusp}$ and $R_{\rm fold}$ statistics of the *Truth* data for each of our target galaxies as a function of the largest distance (normalized by $R_{\rm Ein}$) between the three merging images θ (cusp lenses) or the merging pair θ_1(fold lenses). Our sample of mock lenses contains $R_{\rm cusp}$ and $R_{\rm fold}$ statistics as high as 0.5 for cusp lenses, and as high as 0.35 for fold lenses.

In Figure 2.6, we plot the $R_{\rm cusp}$ values of ten real lenses, with data as reported in Keeton et al. (2003) and Xu et al. (2015). Several of these lenses, notably B2045 and B1933 have $R_{\rm cusp}$ and $R_{\rm fold}$ that are inconsistent with lensing by a smooth potential (Keeton et al., 2003). Xu et al. (2015) showed some of these anomalies could be accounted for by introducing a population of dark subhalos and multipole potential terms in smooth lens models, but noted that the observed anomalies were unlikely to caused entirely by dark subhalos. In the following paragraphs, we will focus attention on the 4 systems with large $R_{\rm cusp}$ and $R_{\rm fold}$ anomalies in order to understand their origin, and to gain insight into how a baryon-induced induced flux ratio anomaly might reveal itself in an observational scenario.While we juxtapose real lens systems with our mock lenses in Figure 2.6, we do not argue that baryons are responsible for the anomalies seen in these systems. Rather, we emphasize that features of the luminous matter in a lensing galaxy can give rise to the large values of $R_{\rm cusp}$ and $R_{\rm fold}$ typically associated with non-baryonic substructure - albeit rarely - especially in systems with a stellar disk or other irregularities. The stellar mass components and flux ratio distributions for these lenses are shown in Figures 3.4 and 2.5, while maps of the magnification surfaces are shown in Figures 2.9, 2.10, 2.11, 2.14, and 2.15. Unsurprisingly, three of the most anomalous mock lenses show evidence for disky or boxy isophotes (see Figure 2.8). The remaining system, NGC7626, has an $R_{\rm cusp}$ anomaly that can can be partially accounted for by the presence of a background galaxy and globular clusters near the images. Three out of of four anomalous mock lenses have velocity dispersion below 200 km s^{-1}, unlike the high central velocity dispersion deflectors most likely to act as strong lenses. In light of the inflated percentage of low stellar velocity dispersion targets in our lens sample, which are more likely

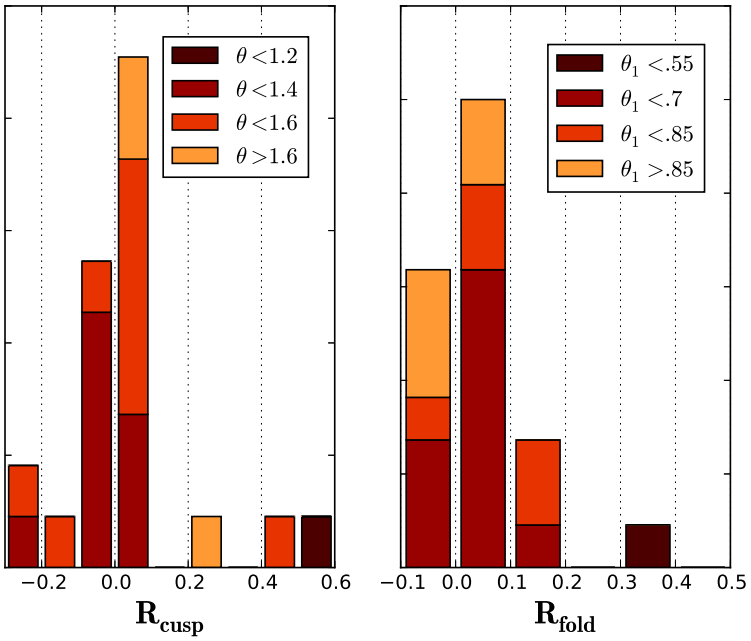

Figure 2.7: Distributions of R_{cusp} and $R_{\rm fold}$ statistics of the *Truth* data, color coded by the largest separation between the three merging images θ (for cusp configurations) or the separation between the merging image pair θ_1 (for fold configurations), normalized by the Einstein radius.

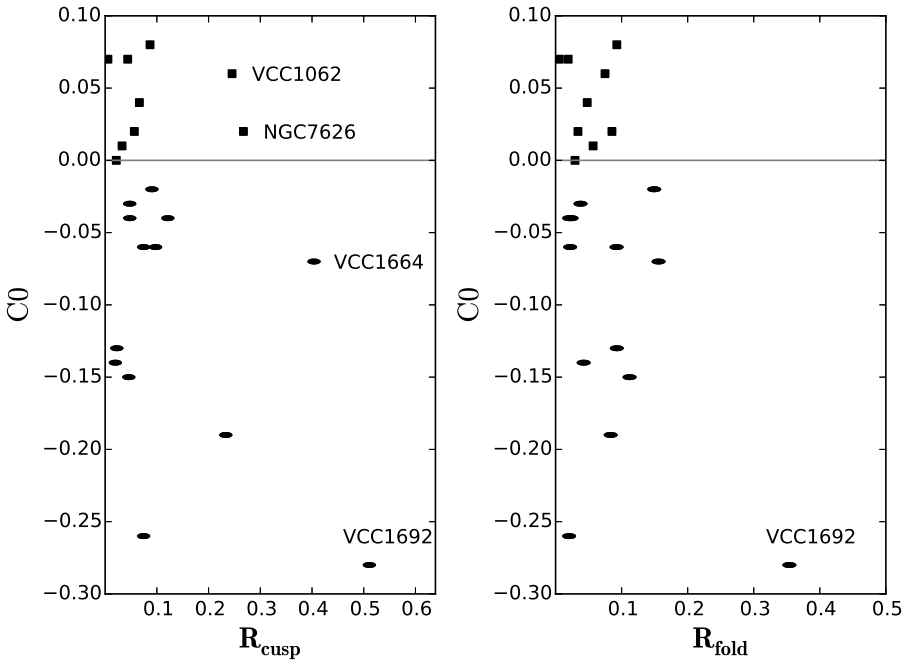

Figure 2.8: Diskyness and boxyness for each mock deflector, described by the `galfit` parameter $C0$. A positive $C0$ corresponds to boxy isophotes, while a negative value indicates diskyness. The anomalous systems shown in Figures 3.4, 2.5, and 2.6 are labeled.

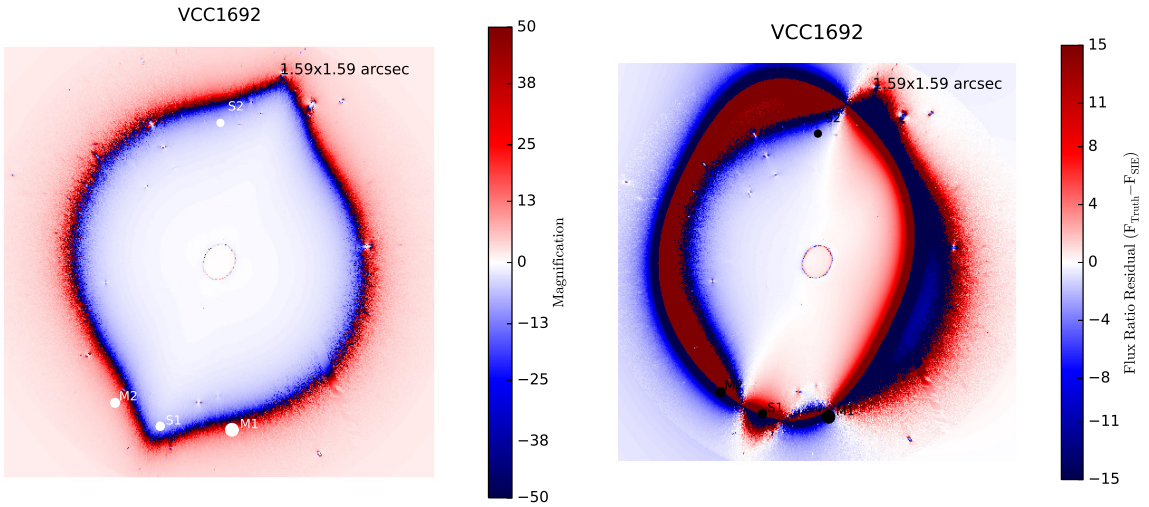

Figure 2.9: $R_{\rm cusp}$ anomaly. Left: Magnification surface derived from the convergence map. Right: Residuals after subtracting the magnification mag of the best fit SIE model. The interplay of the disk and external shear, which is nearly orthogonal to the disk position angle in this system, creates large residuals in the magnification surface that create a strong flux ratio anomaly.

to contain disks and irregular morphological features than high velocity dispersion deflectors, the non-detection of flux ratio anomalies through most of our lens sample illustrates the sub-dominant nature of flux ratio anomalies caused by luminous matter.

2.3.3 Analysis of anomalous systems

Analyzing the few mock deflectors where the luminous matter of the lensing galaxy influences the flux ratios serves to illustrate how stellar mass can affect lensing observables. We will first inspect the cusp configurations, followed by the fold configurations.

2.3.3.1 Cusps

- VCC1692: An elongated galaxy with a prominent disk, as seen in Figure 3.4, with stellar velocity dispersion $\sigma_* = 187$ km/sec. We embed this galaxy in an external shear with a 60 degree offset between the disk position angle and the position angle

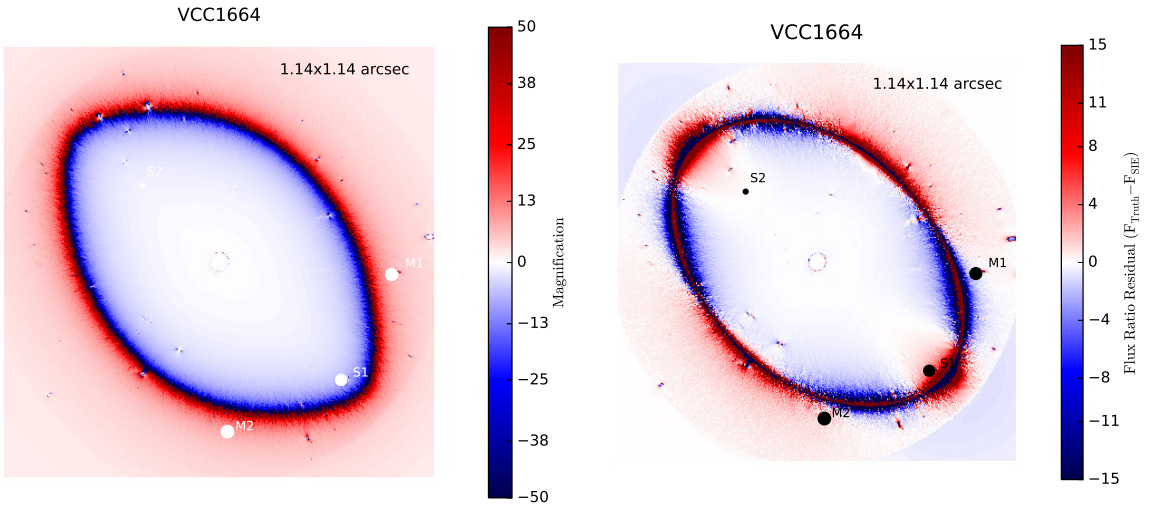

Figure 2.10: $R_{\rm cusp}$ anomaly. Left: Magnification surface derived from the convergence map. Right: Residuals after subtracting the magnification mag of the best fit SIE model. The effect of the stellar manifests itself as a multipole pattern in the residual map.

of the external shear, resulting in a warped astroid caustic. The residuals between the magnification surface of the mock lens and the best fit SIE, shown in Figure 2.9, dramatically displays this effect, with large residuals on the ends of the disk where images are located. The lens has an unusual image configuration, with the far image (S2) located off the symmetry axis of the cusp, while the three cusp images are very close to each other. Both the SNFW and SIE fit the positions reasonably well, although both models display M1/S1 flux ratio anomalies of 80%. The proximity of images M2 and S1 to the stellar disk likely significantly the perturbs the flux ratios between these images. We can compare this with the closest real analog in Figure 2.6 B2045+265. Deep imaging of the system (McKean et al., 2007) shows that the deflector galaxy is almost perfectly round ($b/a = 0.94 \pm 0.01$) when imaged with an F160W filter, while it displays irregular morphological features in F814W. (McKean et al., 2007) also investigate the possibility that a luminous satellite located between the main deflector and the three cusp images could be responsible for the observed flux ratio anomaly. The Einstein radius of 1.06 corresponds to a velocity dispersion of 278 km s^{-1}, once the correct source redshift of $z_s = 2.35$ is taken into account (Nierenberg, 2017, private

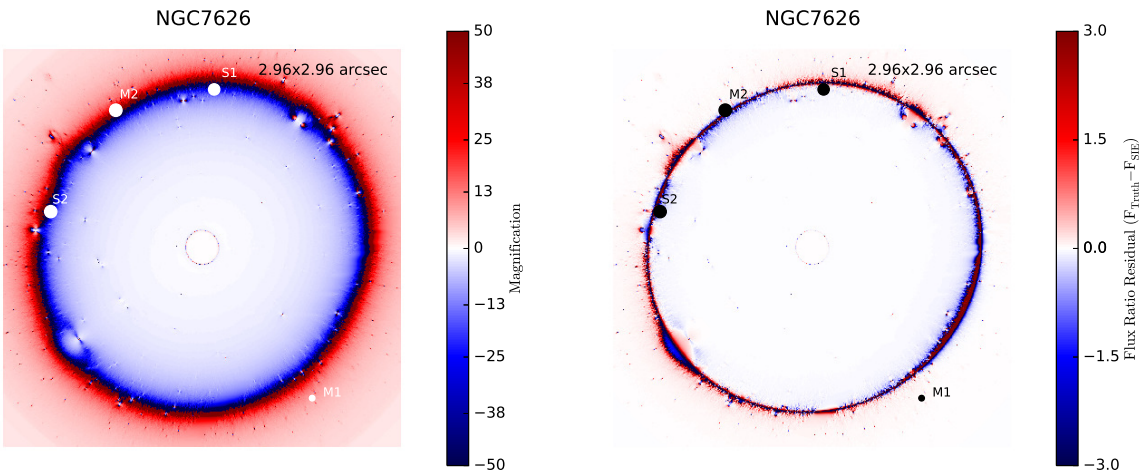

Figure 2.11: $R_{\rm cusp}$ anomaly. Left: Magnification surface derived from the convergence map. Right: Residuals after subtracting the magnification map of the best fit SIE model. This system does not exhibit any obvious large scale morphological irregularities, although the effect of a background galaxy in the lower left is clearly visible. The galaxy appears to contain many luminous substructures, some of which may be associated with a dark subhalo. However, it is possible that some of these features, such as the galaxy in the lower left, are in the background or foreground. In Section 3.3.1 we discuss the impact of luminous substructure on flux ratios and the $R_{\rm cusp}$ parameter.

communication). Thus, the deflector galaxy in B2045 is very different than the one in the mock lens discussed here, consistent with a different origin of the anomaly, even though the amplitude is the same. In our sample of mock lenses, VCC1692 is the only lens with significant astrometric anomalies. In this sense it is an outlier, as our analysis shows that lensing by luminous matter typically does not result in image positions that cannot be fit by an SIE or SNFW, while they still may result in anamalous flux ratios. The interplay between the disk and external shear is likely to blame for this unique system, resulting in the asymmetric image configuration and peculiar shape of the astroid caustic.

- VCC1664: This is a small galaxy with velocity dispersion 155 km s^{-1}. As such, it is not representative of a typical deflector of lensed quasars. It is similar to VCC1692 in that

it has a prominent disk that results in large M1/S1 and M2/S1 flux ratio anomalies that both the SIE and SNFW models fail to reproduce, as seen in the distributions of Figure 2.5. The magnification residuals between the *Truth* model and the best fit SIE result in a multipole pattern around the critical curve, seen in Figure 2.10. Since there is a significant amount of small scale structure scattered around the deflector, some of which lay close to an image, we experimented with removing these potential sources of flux ratio anomaly, but found that this did not affect the flux ratios. A lens with a similar flux ratio anomaly, RXJ1131+1231 (Sluse et al., 2003), differs in several important ways. First, Suyu et al. (2013) measured a stellar velocity dispersion in J1131 of 323 ± 23 km s^{-1}, which would likely result in images far enough from the majority of the stellar mass of the lens to be affected by morphological features of the luminous matter, especially as imaging of J1131 shows no evidence for the presence of a stellar disk or significant elongation. Second, the flux ratio anomaly in J1131 is accompanied by an astrometric anomaly; attempts to fit the lens with an single SIE with shear or a two-lens model both fail to recover the correct astrometry, whereas smooth potentials recover the image positions of VCC1664 almost perfectly. It should also be noted that because VCC1664 has a larger cusp image separation θ than J1131, the $R_{\rm cusp}$ statistic can naturally be larger without substructure.

- NGC7626: This is the most massive cusp mock lens ($\sigma_* = 274$ km sec^{-1}) with a significant $R_{\rm cusp}$ anomaly. NGC7626 is surrounded by globular clusters and luminous satellites. One background galaxy is visible to the lower left, and it induces a flux ratio anomaly in the fold configuration (see the S2/M1 ratio in Figure 3.4), although it is too far from the cusp images to be responsible for the $R_{\rm cusp}$ anomaly and does not affect the merging pair in the fold configuration. There are two structures between the M2 and S2 images, seen clearly in the convergence map (Figure 3.4) and in the map of the magnification surface (Figures 2.11 and 2.12) which visibly perturb the critical curve. The structure outside the curve resembles a background spiral galaxy, while the object just inside resembles a large globular cluster. Since NGC7626 does not possess a stellar disk or boxy isophotes that could explain the anomaly, we experimented with

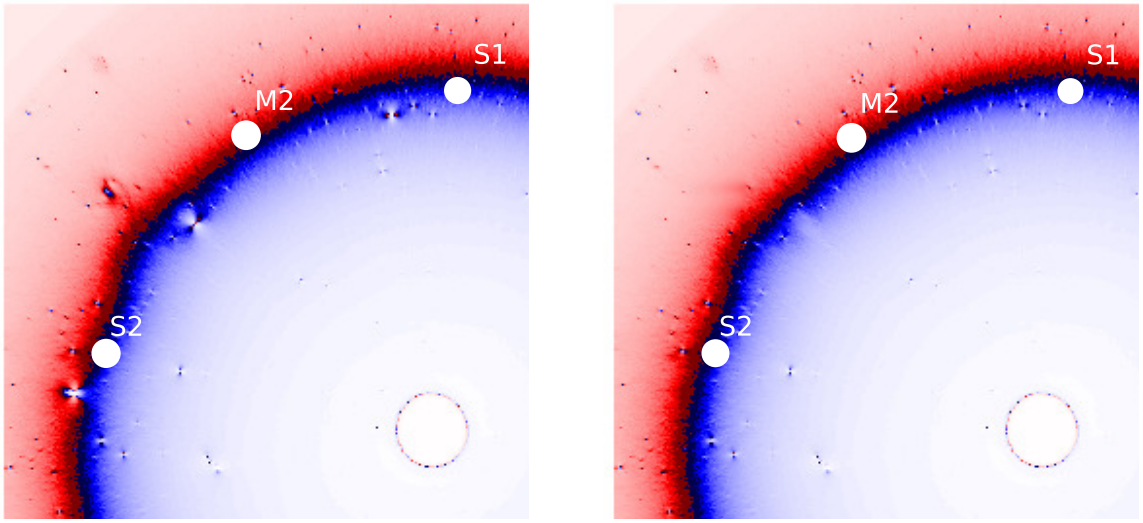

Figure 2.12: Left: Original magnification map, with all small scale structure present. Right: Magnification map with 3 globular clusters and one background or satellite galaxy removed. Before their removal, each of these features in the convergence map contributed the equivalent of $10^7 - 10^{7.5} M_\odot$.

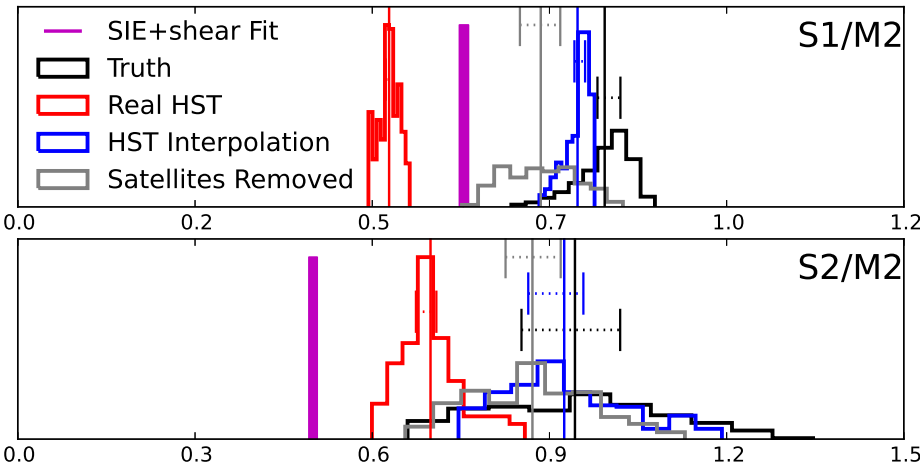

Figure 2.13: Flux ratios after the removal of small scale structure between the three cusp images in NGC7626.

removing these features individually, replacing them with smooth interpolations of the convergence map. Specifically, we removed the globular cluster and background galaxy between M2 and S2, a small globular cluster near S1, and a very small cluster near S2. The before/after magnification maps are shown in Figure 2.12, and the new flux ratios in Figure 2.13. After removing these small scale structures, which our normalization procedure assigned convergence equivalent to that produced by a $10^7 M_\odot$ perturber, we find that the $R_{\rm cusp}$ anomaly shrinks in magnitude to 0.19 from 0.26. In the context of Figure 2.16, this suggests it could be accounted for by an SIE model. We therefore conclude that the main source of anomaly in this system is due to structure in the deflector on scales smaller than the image separation. This hypothesis is supported by examining the residual map in Figure 2.11, where the alternating blue and red colors coincide with the location of the perturbing globular clusters and galaxy. NGC7626 highlights that even massive deflectors can suffer flux ratio anomalies if there is sufficient small scale structure near the critical curve, whether it is in the form of dark substructure or luminous matter. However, it is important to remember that that this is seemingly a rare occurrence, and it is possible that this signal will be overwhelmed by the lensing signatures of a full population of dark subhalos, a question we will address in a future paper.

In the θ vs. $R_{\rm cusp}$ parameter space, RXJ0911+0551 (Kneib et al., 2000) is the nearest neighbor of NGC7626. NGC7626 is a round deflector with an ellipticity of 0.17, while a best fit SIE model of J0911 (Sluse et al., 2012) favors a deflector with ellipticity 0.11. The velocity dispersion of J0911, if it is modeled as an SIE with Einstein radius 0.9 arcsec works out to $\sigma_{\rm SIE} = 239$km s^{-1} after adopting correct lens and source redshifts (Kneib et al., 2000), while NGC7626 has a velocity dispersion of 274 km s^{-1}. Neither J0911 nor NGC7626 display astrometric anomalies with respect to a smooth model when a second deflector galaxy is included in the model for J0911 (Sluse et al., 2012). NGC7626 stands out in our set of mock lenses, as it is round with high velocity dispersion, but still displays significant flux ratio anomalies, indicting that high velocity

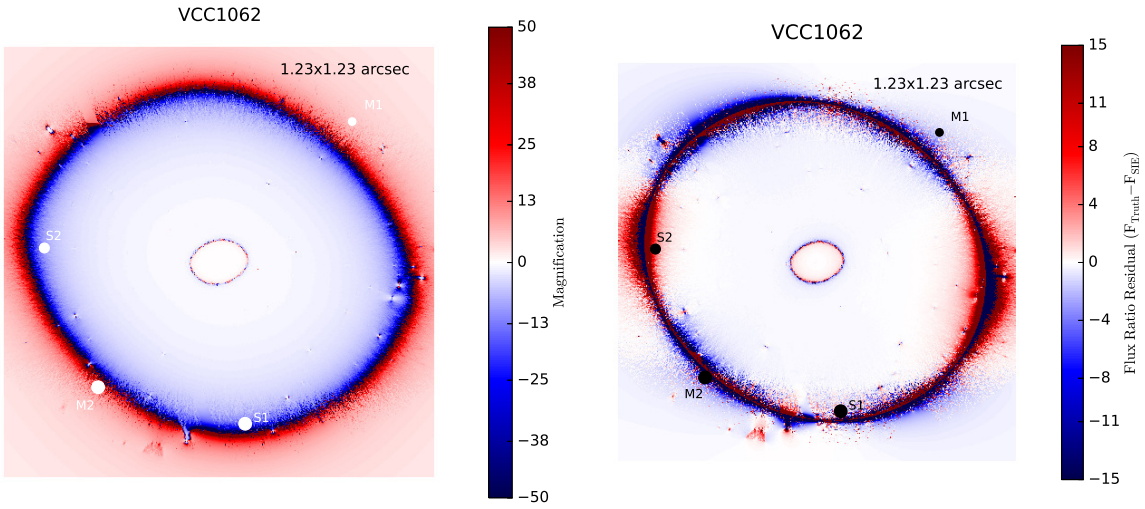

Figure 2.14: $R_{\rm cusp}$ anomaly. Left: Magnification surface derived from the convergence map. Right: Residuals after subtracting the magnification mag of the best fit SIE model. The effect of the stellar mass manifests itself as a multipole pattern in the residual map.

dispersion does not always guarantee benign flux ratios.

- VCC1062: The external shear applied in this lens forms an angle of 61 deg with the stellar quadrupole moment position angle, similar to VCC1692, that results in a cusp configuration that is not coaxial with the stellar ellipticity position angle. Coupled with the boxy isophotes (see Figure 2.8), this results in a complicated potential that the best fit SIE fails to capture, as seen in the map of flux ratio residuals in Figure 2.14.

  The small velocity dispersion ($\sigma_* = 179$ km/sec) results in a small Einstein radius,

which in turn results in images close to the center of the lens. As a result, the images S2 and M2 are located closer to the ends of the elongated baryonic mass distribution where there is more curvature in the potential, making this lens more susceptible to influence from its luminous mass component. While both the SIE and SNFW fail to recover the correct flux ratios, the anomalies are $< 40\%$. However, collectively the anomalies lead to a significant $R_{\rm cusp}$ anomaly. VCC1062 has an anomaly quite similar to that observed in B0712+472 (Jackson et al., 1998). Both deflectors have relatively

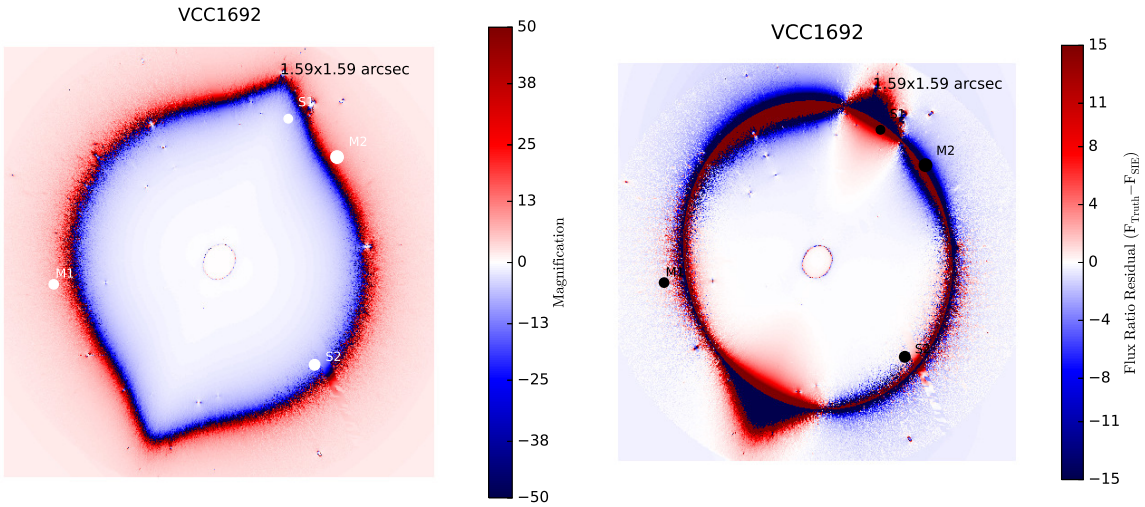

Figure 2.15: $R_{\rm fold}$ anomaly. Left: Magnification surface derived from the convergence map. Right: Residuals after subtracting the magnification mag of the best fit SIE model. The merging image pair in this fold configuration happens to land near the portion of the critical curve extended by the stellar disk, a feature the best SIE fails to capture. The resulting residual in the magnification surface gives rise to the large $R_{\rm fold}$ anomaly.

small stellar velocity dispersion (B0712 has $\sigma_{\rm SIE} = 189$ km/sec), which we estimate for B0712 by adopting the correct redshifts as cited in Sluse et al. (2012) and utilizing the relationship between image separation and velocity dispersion presented in Kochanek et al. (2000). The two lens systems both appear to have highly elliptical baryonic mass distributions, consistent with the presence of edge-on massive disk (Jackson et al., 1998). While the small mass and high ellipticity of B0712 suggests the anomaly may be influenced by baryonic matter, the astrometric anomaly noted by Kawano et al. (2004) is not a common feature among our mock lenses, and as such alternate explanations are favored, such as dark substructure. Discrepancies between the flux ratios in the optical/near IR and radio data suggest microlensing and/or dust extinction could also be present. Regardless, deeper imaging of this system could help disentangle the possible role of baryonic structure from other sources of flux ratio perturbation.

Completely different is the case of the real lens B1422+231 (Patnaik et al., 1992).

Even though the cusp flux ratio anomaly is similar to that of VCC1062, the cusp image separation θ is large enough that the measured $R_{\rm cusp}$ value (Koopmans et al., 2003) alone is not inconsistent with lensing by a smooth potential (Keeton et al., 2003), while the analysis by Nierenberg et al. (2014); Xu et al. (2015) suggests substructure in the vicinity of image A could contribute to the anomaly. Like many systems in our sample, the astrometric anomalies in B1422 are relatively tame compared to other real lens systems.

2.3.3.2 FOLDS

- VCC1692: The only significant $R_{\rm fold}$ anomaly appears in VCC1692, the lens system with the largest $R_{\rm cusp}$ anomaly. The S1/M2 flux ratio anomaly is 50%, likely because of the influence of the stellar disk. This effect is clear in Figure 2.15, where large residuals between the best fit SIE and the *Truth* model are evident. Unlike the cusp configuration, the *HST Interpolated* distribution agrees with the *Truth* flux ratios, indicating that the curvature of the gravitational potential just off the major axis of the disk is gradual enough that the convolution procedure still captures the disk's effect on the magnification surface. From a modeling standpoint, this implies that the information needed to accurately reproduce the lensing signal of a very disky deflector is lower for fold configurations than for cusp configurations, because cusps images live near the ends of the disk (for major axis cusps), where there is greater curvature. Conversely, folds will tend to straddle the sides of a disk, so one need only resolve a small portion of a relatively straight critical curve dividing the two images.

B1555+375 (Marlow et al., 1999), the closest real analogue to VCC1692 in the $R_{\rm fold}$ / θ_1 parameter space, has a nearly identical θ_1 and a large $R_{\rm fold}$. B1555 is also a very small angular separation lens, with a deflector velocity dispersion estimated from the image separation of 134 km s^{-1}, and is highly elliptical. The high ellipticity $\epsilon = 0.54$ and disk feature detected in deep AO imaging (Hsueh et al., 2016) further strengthen

the analogy between the mock and the real system. Attempts to model the lens system (Marlow et al., 1999; Miranda & Jetzer, 2007) also find that the astrometry of B1555 is consistent with an SIE model, so no extreme astrometric anomalies are present, as is the case with the majority of our mock lenses. Another clue to the nature of the anomaly in B1555 arises from the modeling of VCC1692: if flux ratios are included as constraints in the SIE fit to VCC1692, the resulting astrometric errors and flux ratios appear strikingly similar to that observed in B1555. This suggests that, while formally not a good fit to the data, a single SIE can capture the astrometry and flux ratios of a disky galaxy to within 10 mas and $\approx 70\%$, respectively. It is likely that this is not a coincidence, as Hsueh et al. (2016) show that the system can be fit to high precision by explicitly modelling the stellar disk, without the need to invoke dark subhalos. Naturally, this does not mean that dark substructure is not present, just that it is not required.

The lens system MG0414+0534 is not fit for juxtaposition with VCC1692, as the central velocity dispersion, derived from the Einstein radius of an SIE (Xu et al., 2015), is approximately 286 km s^{-1}, while lens models favor moderate SIE ellipticity of ≈ 0.2, indicating that the lensing galaxy is likely very massive and round (Hewitt et al., 1992). While the anomaly is small in magnitude, the proximity of the merging image pair makes it unlikely that a smooth potential provides an adequate description of the lens system (Minezaki et al., 2009; Keeton et al., 2005; Xu et al., 2015). However, Xu et al. (2015) point out that other sources of anomaly besides dark substructure may be needed to explain the observed anomaly. Deep imaging of this system would help rule out the possibility that baryons play a significant role, although the non-detection of baryon-induced $R_{\rm fold}$ anomalies in our sample suggests the dark matter is responsible.

B0128+437 (Phillips et al., 2000), a small deflector with $R_{\rm Ein} = 0.24$", and low sersic index, consistent with a late-type morphology (Lagattuta et al., 2010). (Biggs et al.,

2004) show that lens models favor very elliptical SIE profiles, but fail to fit the observed image positions. While the small size of the lens and elongated nature of the deflector suggest baryons may contribute a non-negligible effect to the flux ratios, the presence of astrometric anomalies suggests non-baryonic substructure may also contribute.

The system B1608+656 (Fassnacht et al., 1996) is peculiar in that it consists of two merging galaxies (Fassnacht et al., 2002), with the most massive one having a velocity dispersion of 260 ± 15 km s^{-1} (Suyu et al., 2010). The B1608 system is complex enough that a description in terms of simple anomalies is not appropriate and searches for dark matter substructure must take into account this complexity with a detailed model.

Even more anomalous than any one of our mock lenses is the system B1933+503 (Sykes et al., 1998), a well known peculiar system with a late-type deflector that contains a prominent stellar disk. Kochanek & Dalal (2004) investigate whether higher order multipole terms in the lens potential can account for the observed anomaly, and conclude that such an explanation is unlikely, which seems to favor a dark substructure as a source of flux ratio perturbation. However, it is possible that the lensing properties of galaxies with very irregular morphology, such as a prominent edge on disk, may require a more different description than can be encapsulated by adding a few higher order multipole terms.

The possibility of a large $R_{\rm cusp}$ or $R_{\rm fold}$ arising from the baryonic structure of the lens, especially in low mass ellipticals with features such as disks or boxy isophotes, behooves observers to investigate whether the lensing galaxy possesses baryonic mass distributions that require detailed modeling. Indeed, deep imaging of B1555 and B0712 shows that a disk is present, and can account for the apparent anomaly (Hsueh et al., 2016). Of the lenses in our sample with the largest $R_{\rm cusp}$ values, some have stellar disks visible even in the rebinned images. Others do not have visible disks but are significantly elongated, even in the rebinned images. These findings suggest that in most cases an elliptical galaxy at $z = 0.5$ can be imaged well enough by the HST for potential sources of baryonic anomaly to be identified and modeled, but care should be taken to account for the interplay between the

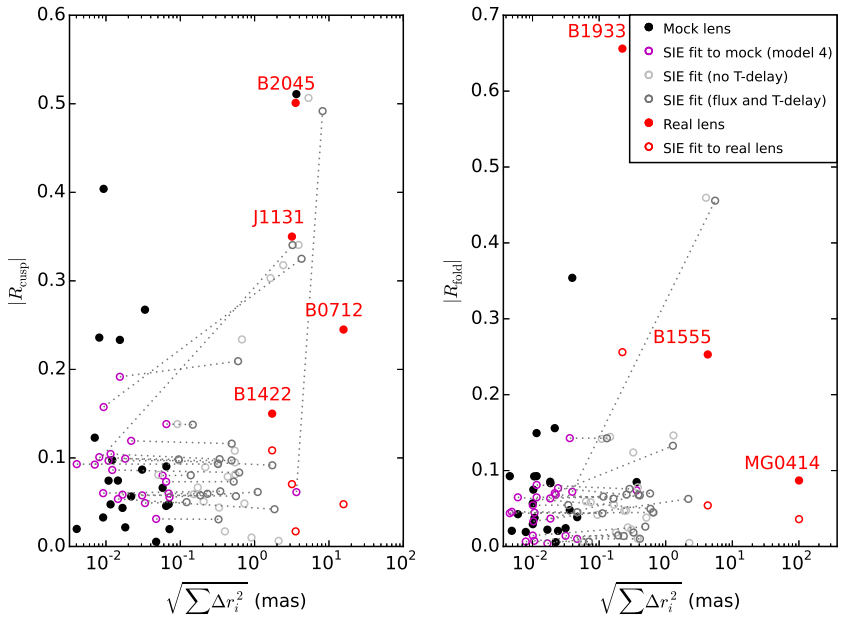

Figure 2.16: $R_{\rm cusp}$ and $R_{\rm fold}$ parameters for the *Truth* model and best fit SIE *Truth* model (black points and purple circles, respectively), together with the values observed in real lenses (red points), and SIE fits to real lenses (red circles). The real lenses and *Truth* data points are distributed along the x-axis according to the astrometric error of their best fit SIE model summed in quadrature. Counter-intuitively, as most of our mock lenses have R values larger than the best fit SIE, a positively sloped dashed line is actually an improved model for the data, although the R value increases. This flux ratio precision comes at a cost of larger astrometric and time-delay errors, irrespective of the exact uncertainty we place on the flux ratios and time delays. Further, without time delay information (light grey points), the code cannot distinguish between major and minor axis cusps, which further complicates the modeling process. In this plot, we impose flux ratio uncertainties of 10 % (grey points). When fitting the real lens systems, we adopt lens data and observational uncertainties from (Sluse et al., 2012) (MG0414, B2045), (Sluse et al., 2006) (J1131), (Jackson et al., 1998) (B0712), (Nierenberg et al., 2014) (B1422), (Cohn et al., 2001) (B1933), and (Hsueh et al., 2016) (B1555). There is a clear separation between the real lenses and the mock lenses, with real systems possessing both astrometric and flux ratio anomalies, and our set of mock lenses mostly confined to flux ratio anomalies $< 30\%$ and nearly perfect astrometric precision.

external shear and stellar ellipticity, which could result in an off-axis cusp, as in VCC1692 and VCC1062 imaged in Figures 3.4 and 2.5.

2.3.3.3 Characterizing a baryonic-lensing signal through modeling: astrometric and flux ratio anomalies

Among the properties of the baryonic mass of a deflector likely to give rise to flux ratio anomalies, stellar disks or other elogated structures, most often seen in low mass, low stellar velocity dispersion galaxies, constitute the majority of the anomalous systems in our sample. On the other hand, in round, high velocity dispersion systems such as NGC7626 where there is no obvious stellar disk or other large scale feature, anomalies could be induced by compact structures near the critical curve. Regardless of the origin, flux ratio anomalies from luminous matter may be difficult to identify solely by examining flux ratios.

In order to help distinguish a baryonic lensing signal in systems similar to NGC7626 from other sources of anomaly, we highlight a feature of our mock systems, seen even systems with significant flux ratio anomalies, that is not frequently observed in real lens systems. Our mock lenses are characterized by a conspicuous absence of astrometric anomalies, which can be present in real systems at the level of tens of mas for subhalos located near an image, in projection (Chiba, 2002; Chen et al., 2007). This suggests that a feature of perturbation by dark matter subhalos, that could be used to distinguish between baryonic and dark matter perturbations, is a flux ratio anomaly coupled to an astrometric anomaly, especially if the introduction of a dark substructure to the lens model simultaneously resolves both discrepancies.

To compare the astrometric precision of the SIE model fit to our mock lenses with that of an SIE fit to real lenses, we fit several of the real systems shown in Figure 2.6 with an SIE plus external shear, varying the Einstein radius, ellipticty, shear, position angles, and deflector centroid. For the resulting best fit model we compute the flux ratio and astrometric anomalies. We omit systems such as J0911 and B1608, which require complicated modeling involving two galaxies within the Einstein radius. We repeat the fit for each of our mock systems omitting time delays and enforcing flux ratio constraints, to see if the inclusion or

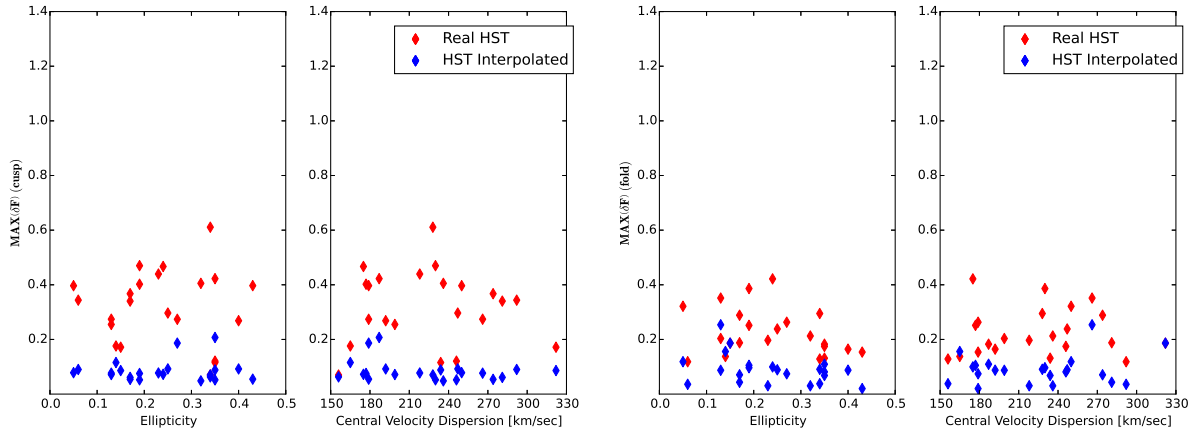

Figure 2.17: Largest flux ratio anomaly for Model 2 (*Real HST*) and Model 3 (*HST Interpolated*) in each lens as a function of ellipticity and central velocity dispersion for cusp configurations (left) and fold configurations (right).

exclusion of either these data significantly impacts the results.

In Figure 2.16, we plot the $R_{\rm cusp}$ or $R_{\rm fold}$ values of our mock lenses and some real lens systems, along with a best fit SIE to each of the real systems and mocks as a function of the total astrometric error summed in quadruature. There is a clear separation between the real lenses and the mock lenses, with real systems possessing both astrometric and flux ratio anomalies, and our set of mock lenses mostly confined to flux ratio anomalies $\leq 30\%$ and nearly perfect astrometric precision. The effect is more pronounced in cusp lenses, although both image configurations follow this general trend. Enforcing flux ratio contraints of 10 % in the SIE fit to the mock lenses sometimes corrects the flux ratio anomaly at the expensive of astrometric precision, but the resulting points still populate a different region of parameter space than the real lens systems. Additionally, while SIE fits to the mock lenses systems come close to reproducing the observed $R_{\rm cusp}$ or $R_{\rm fold}$ value, the best fit model of a real lens system differs significantly.

Our analysis aims to characterize the properties of a deflector (low velocity dispersion, high ellipticity, etc.) that may increase the likelihood of observing a baryon-induced flux ratio anomaly. While the trend in Figure 2.16 can be used to characterize a purely baryonic lensing signal by the absence of an astrometric anomaly, it should not be adopted as a criterion

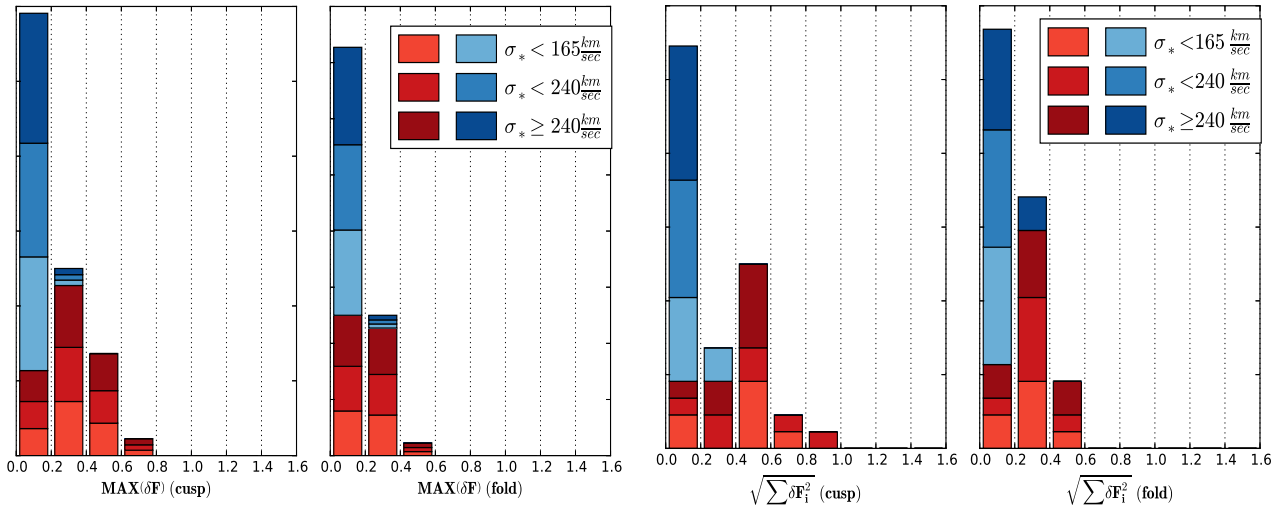

Figure 2.18: Distributions of the largest anomalies (left) and the three anomalies summed in quadrature (right) for each lens, color coded by the deflector's central velocity dispersion. The color scheme is the same as in Figure 2.17.

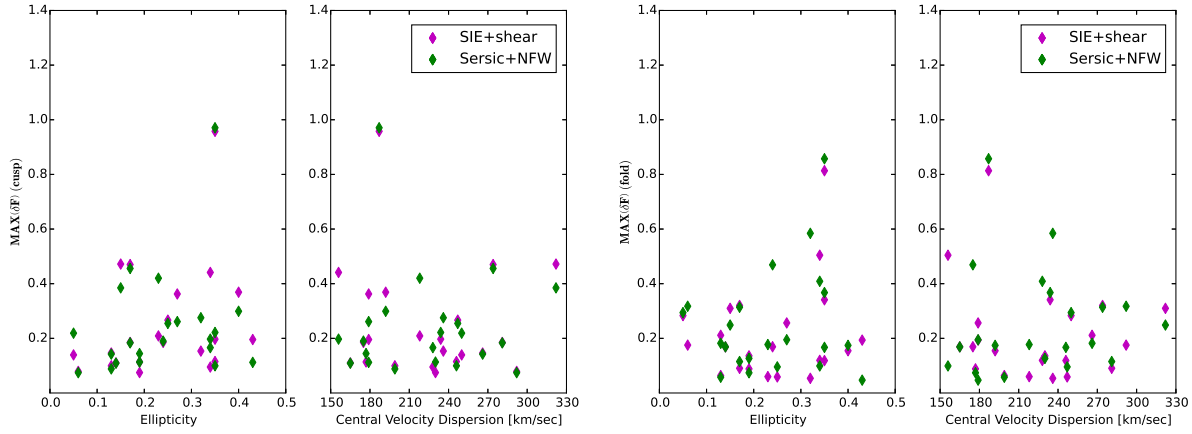

Figure 2.19: Largest flux ratio anomaly for Model 4 (purple) and Model 5 (green) in each lens as a function of ellipticity and central velocity dispersion for cusp configurations (left) and fold configurations (right).

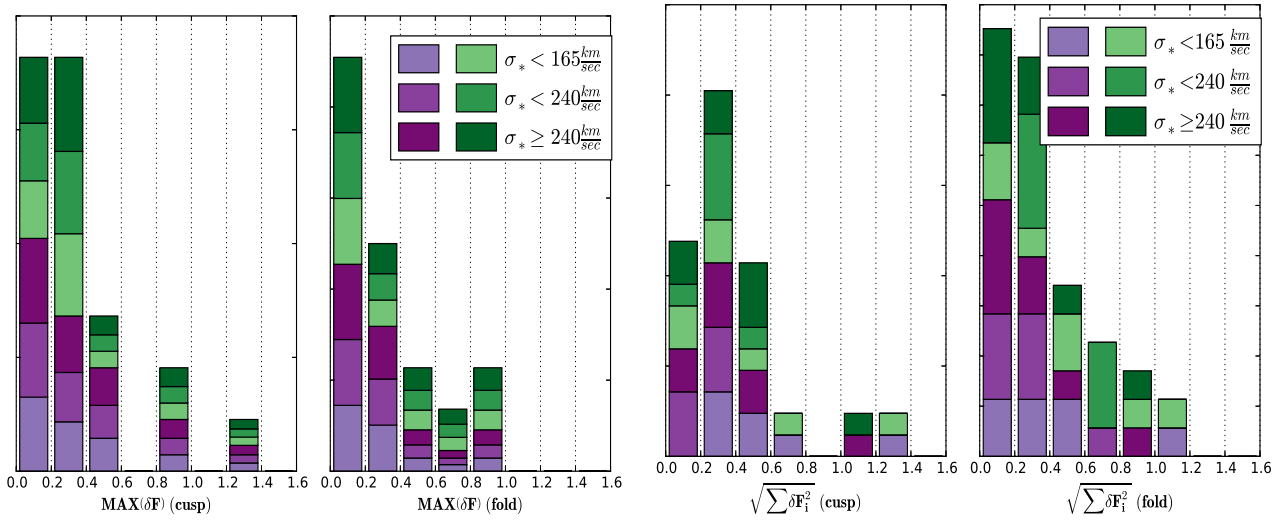

Figure 2.20: Distributions of the largest anomalies (left) and the three anomalies summed in quadrature (right) for each lens, color coded by the deflector's central velocity dispersion. The color scheme is the same as in Figure 2.19

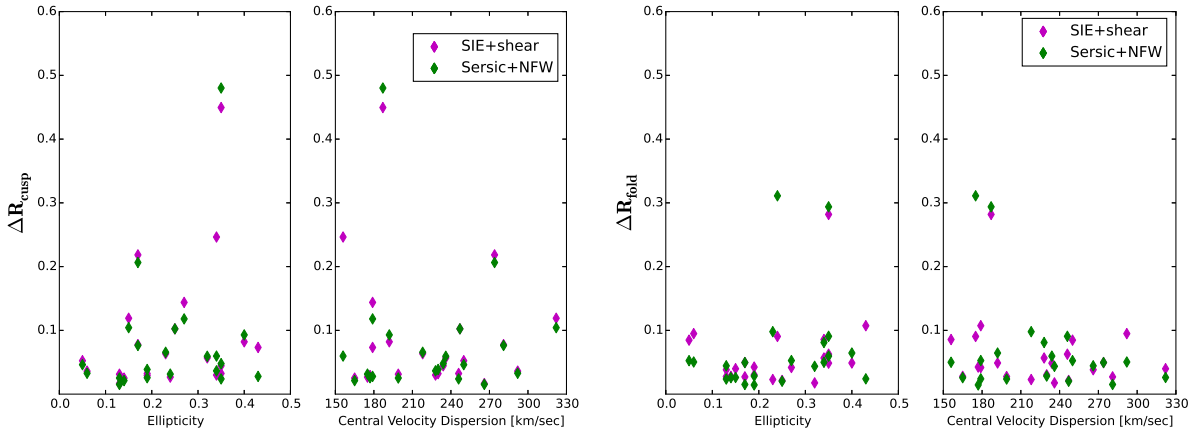

Figure 2.21: Differences in the R-cusp (left) and R-fold (right) statistics between the SIE and SNFW models, and the *Truth* data. The color scheme is the same as in Figure 2.19.

used to rule out lens systems as candidates for analysis of dark matter substructure, as our analysis does not address the question of whether a dark subhalo will necessarily result in simultaneous astrometric and flux ratio anomalies, and the relative magnitudes of these perturbations. Further, the interpretation of what constitutes an astrometric anomaly is model dependent, and depends on the precision the available data. Regardless of these nuances, given a large sample of lenses, observations of the morphological features of the lensing galaxy, together with an absence of astrometric anomalies in the presence of relatively small flux ratio anomalies, could be used to flag certain systems as more likely than others to exhibit lensing effects induced by luminous matter. This would necessitate additional observations of the lensing galaxy, and detailed modeling of its morphology.

2.3.4 $R_{\mathbf{cusp}}$, $R_{\mathbf{fold}}$, and flux ratios; Models 2-5.

2.3.4.1 Flux Ratios: Models 2 and 3 (*Real HST* and *HST Interpolated*)

The process of re-binning pixels introduces a significant source of flux ratio anomaly - relative to the *Truth* model - compared to Model 3, as seen in Figures 2.17 and 2.18. This suggests that in a real lens observed at redshift 0.5, directly using pixel values to infer properties of the lens baryonic mass distribution introduces flux ratio anomalies of about 40% for cusp configurations, and about 20% for fold configurations. There is no clear trend between ellipticity, stellar velocity dispersion, and flux ratio anomaly. Based on these findings, we conclude that the source of anomalies for the pixellated models are numerical and associated with computing lensing derivatives from coarsely sampled data. Thus, even in the presence of exquisite data, it is best to interpolate the pixellated data with smooth functions Cappellari (2002) to describe the baryonic component correctly and attribute excess anomalies to dark subhalos. As the smooth model can be interpreted as the best empirical basis one could use to model a lens, given that there exists an average variation of 9.3% and 10.6% in flux ratios between the smooth model and the *Truth* data, for fold and cusp lenses, respectively, we conclude that this is a typical perturbation induced by baryonic structure alone.

2.3.4.2 Flux Ratios: Models 4 and 5 (SIE and SNFW)

The relative flux ratio anomalies for both the SIE and SNFW models, shown in Figure 2.19 and Figure 2.20, display a clearer dependence on baryon ellipticity and the galaxy's central velocity dispersion, with the largest flux ratio anomalies present in highly elongated and low velocity dispersion galaxies. However, there is considerable scatter in the trends, as many highly elliptical and low velocity dispersion targets do not possess significant baryonic-induced anomalies. The largest anomalies are present in cusp configurations. Many of the errors are on the order of about 10%, comparable to the noise floor derived from the *HST Interpolated* model, and as such should not be formally considered 'anomalies'. There is no significant difference between the accuracy of flux ratios recovered by the SIE and SNFW models.

2.3.4.3 $R_{\rm cusp}$ and $R_{\rm fold}$: Models 4 and 5 (SIE and SNFW)

The offsets of $R_{\rm cusp}$ and $R_{\rm fold}$ statistics, shown in Figure 2.21, between the *Truth* data and the SIE and SNFW fits reveal a clear trend in anomalies for the $R_{\rm cusp}$ statistic, with highly elliptical and low velocity dispersion targets possessing the largest anomalies. However, this relationship is not deterministic, as some low velocity dispersion or significantly elongated deflectors do not display anomalies. The largest offsets in the $R_{\rm fold}$ statistic, however, do not appear to be correlated with ellipticity or velocity dispersion, which suggests that this statistic is less sensitive to the baryonic structure of the lensing galaxy. On the other hand, our results demonstrate that the $R_{\rm cusp}$ statistic is recovered almost exactly by smooth lens models in galaxies with low ellipticity. This suggests that image magnifications in the cusp configuration are more strong affected by baryon ellipticity due to the proximity of stellar mass to the lensed images in elongated deflectors. However, very massive galaxies with high velocity dispersions will result in images far enough away from the majority of the stellar mass, presumably leaving their magnifications unaffected.

2.4 Summary and conclusions

Motivated by the growing sizes of known lensed quasars samples and the interest in the lens systems as a probe of dark matter substructure, we have carried out a systematic study of "baryonic anomalies". We have used a sample of high resolution images of nearby early-type galaxies as a starting point to create mock gravitational lens systems, and then we have studied how well the arrival time, positions, and fluxes of the lensed images are reproduced by lens models based on the observed surface brightness distribution and on commonly used functional forms. Our findings can be summarized as follows:

- Arrival times and image positions are virtually unaffected by baryonic substructure and can be recovered within the uncertainties by both empirical lens models and simply parametrized models. We conclude that astrometric anomalies are unlikely to arise from baryonic lensing effects, and can therefore be used to distinguish between the lensing signal of luminous matter and a dark subhalo, which Chen et al. (2007) show can induce astrometric anomalies of order 10 mas. While the absence of astrometric anomalies is a common feature among our mock lenses, the non-detection of astrometric anomalies does not mean that dark substructure is not present. Rather, we claim that in a large sample of lenses, highly elongated deflectors with low stellar velocity dispersion and no astrometric anomalies are the most likely lens systems to posses lensing signals from baryonic structure, and warrant further study and detailed modeling.

- The baryonic structure of a lensing galaxy can introduce a source of flux ratio anomaly in strong lensing that is more pronounced in highly elongated galaxies, and galaxies with low central velocity dispersion. We interpret this as evidence that the baryonic anomalies are dominated by large-scale features such as embedded disks, or isophotal twisting. Our analysis suggests that small-scale features like globular clusters or compact dwarf satellite galaxies contribute substantially to the anomaly in NGC7626, although the non-detection of anomalies from small scale structure in the majority of our mock systems suggests this would be a sub-dominant effect in a large sample of lens systems.

- As our sample of mock lenses contains a disproportionately large number of small deflectors with low velocity dispersions, our analysis likely over-estimates the frequency and magnitude of flux ratio anomalies induced by the stellar mass of a deflector. In light of this, the fact that only 4 out of the 22 mock lenses we study display anomalies indicates that baryon induced flux ratio anomalies are a rare occurrence. Further, the magnitude of these anomalies are significantly smaller than those observed in real systems like B2045 and J1131. Together, these facts suggest that a baryonic lensing signal alone would not dominate the signal from dark substructure, although in some systems the perturbation can be non-negligible.

- By comparing the *Truth* data with the *Real HST* and *HST Interpolated* models, we show that the process of rebinning pixels introduces a large source of error in lens modeling, while convolution with a smoothing kernel produces, on average, flux ratio anomalies of 9.3% and 10.6% for cusp and fold configurations, respectively. Given that the smooth model represents the best smooth profile that may be devised to model a lens from observational data, we conclude that a relative flux ratio anomaly of roughly 10% constitutes an effective minimum flux ratio precision in strong lensing for deflectors at typical redshift 0.5, given the resolution of current optical and near infrared data. Structures that survive the smoothing procedure implemented in the *HST Interpolated* model, such as luminous satellites (likely associated with dark subhalos) and other galaxy-scale features, induce this anomaly.

- The most morphologically complex realistic lenses may have anomalous $R_{\rm cusp}$ and $R_{\rm fold}$ statistics between 0 and 0.5 purely due to luminous structure and substructure. The broad range of values makes it imperative to study in detail the distribution of luminous matter while interpreting these systems. The case of B2045 (Fassnacht et al., 1999), which posses an anomalous $R_{\rm cusp}$ value nearly identical to the mock lens VCC1692, is a good illustration. Whereas the mock lens VCC1692 is highly elongated ($\epsilon = 0.35$) and has a low stellar velocity dispersion, the deflector in B2045 is nearly round and has a much higher velocity dispersion. Thus, even though the two systems

have approximately the same R_{cusp} anomaly, the physical origin is likely different: In the case of VCC1692 it is explained by the presence of a disk, whereas this is not a viable explanation for B2045, supporting the hypothesis that a dark subhalo lurks near a lensed image. In addition, B2045 has image positions that cannot be recovered by a smooth potential, while the positions of VCC1692 can be recovered with a χ^2 value of 1.2 (see Table 2.2), illustrating how astrometric anomalies can be used to identify a lens likely susceptible to baryon-induced flux ratio anomalies, as opposed to dark substructure or line-of-sight induced anomalies.

- Real anomalous lenses and our mocks generally separate well in the space of flux-ratio and astrometric anomalies, indicating that a joint analysis of fluxes and positions is essential to disentangle galaxy-scale luminous structure from true dark matter substructure signatures.

Our results are consistent with and generalize earlier work by Mao & Schneider (1998b), Chiba (2002), and Kawano et al. (2004), who showed with simple models that small scale baryonic substructure such as globular clusters and m=3 multipole terms constitute a subdominant source of flux ratio anomalies. Similarly, our conclusion that large scale features such as disks and isophotal twisting are a non-neglibile source of uncertainties is consistent with earlier theoretical work by Möller et al. (2003), and recent work Hsueh et al. (2016), in which flux ratio anomalies of $\approx 40\%$ (with respect to a SIE model) were observed in a lens with a pronounced stellar disk. The good quality of the Hsueh et al. (2016) data made it possible to observe and model the disk, correcting the anomaly, and illustrates the importance of deep imaging of the lens galaxy. However, when deep imaging is not available, or when the lensed images are so bright to completely dominate the lens galaxy, this potential source of error can controlled and mitigated by restricting samples to deflector galaxies with high central velocity dispersions, which tend to have low ellipticity and be true elliptical galaxies with no disk (Moran et al., 2007; Cappellari, 2002).

Baryon induced anomalies are enhanced in low mass systems due to both kinematic and morphological features, and also due to the lensing geometry. Low mass systems will tend

to have smaller Einstein radii, which will result in images located closer to the center of the lens, where baryonic mass dominates. As such, the flux ratios between images in these lenses will be more strongly influenced by stellar mass.

In conclusion, in order to minimize the impact of stellar mass on flux ratios, we recommend that one restrict lensing studies to the most massive galaxies with large Einstein radii and low ellipticity, and allow for a residual noise floor to absorb both perturbations by undetected structure in the lensing galaxy, and the intrinsic uncertainties introduced by modeling lenses with smooth potentials. If possible, one should also obtain deep and high resolution images of the deflector to look for irregular morphological features, while simultaneously modeling both flux ratio and astrometric data. Furthermore, one must keep in mind the fundamental distinction between purely baryonic anomaly-inducing components, like a stellar disk or a globular cluster, and compact satellite galaxies. Whereas the former class of objects is purely noise from the point of view of dark matter studies, the latter is typically associated with the elusive subhalos. We leave to future work the analysis of how the detailed morphology of real deflectors affects the inference of the properties of a population of dark subhalos.

Acknowledgments

TT, DG, and AA acknowledge support from NSF grant AST-1450141 "Collaborative Research: Accurate cosmology with strong gravitational lens time delays". TT gratefully acknowledges support by the Packard Foundation through a Packard Research Fellowship. CRK acknowledges support from grant HST-AR-14305.002-A, which was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. We thank Chris Fassnacht, Phil Marshall and Paul Schechter, for useful comments and suggestions on this work. We acknowledge the use of the HyperLeda database (http://leda.univ-lyon1.fr).

Table 2.1: Columns 1 and 2 list galaxy name and redshift. The following columns list the stellar velocity dispersion, half light radius, stellar ellipticity, position angle, Sérsic index, external shear, and external shear position angle. Asterisks denote quantities obtained by fitting with `galfit` , while the rest are obtained from the literature. The references are as follows: 1) HyperLeda online catalog (Makarov et al., 2014) 2) (Ferrarese et al., 2006) 3) (Ma et al., 2014)

Name	Redshift	σ_*	$R_{1/2}$	ϵ	θ_ϵ	n	γ	θ_γ	References
		km/sec	arcsec		degrees			degrees	
VCC1664	0.0038	156	15.8	0.34	46.8	3.982	0.08	40	1, 2
VCC1297	0.0038	165	2.33	0.20	-32.1	2.732	0.08	25	1, 2
VCC798	0.0038	175	170.8	0.24	30.2	6.813	0.05	55	1, 2
VCC2092	0.0038	177	24.13	0.19	71.2	4.279	0.05	15	1, 2
VCC1231	0.0038	179	16.89	0.43	-86.3	2.955	0.08	35	1, 2
VCC1062	0.0038	179	16.92	0.27	86.8	3.291	0.05	25	1, 2
VCC1692	0.0038	187	9.5	0.35	-21.7	2.458	0.05	39	1, 2
VCC2000	0.0038	192	10.45	0.40	-84.9	3.977	0.05	75	1, 2
VCC355	0.0038	199	9.78	0.13	-9.4	3.725	0.05	38	1, 2
NGC4872	0.0231	218	28.25*	0.23*	104.0*	6.190*	0.05	70	1
VCC1903	0.0038	228	106.8	0.34	-16.8	6.852	0.05	17	1, 2
VCC881	0.0038	230	411.84	0.19	-57.2	7.016	0.05	85	1, 2
IC4051	0.0231	234	21.01	0.35	97.2	3.210*	0.05	71	1, 3
NGC5322	0.008	236	26.6	0.32*	97*	3.690*	0.05	57	3
NGC1132	0.0231	246	16.1	0.35*	146.0*	2.15*	0.05	135	3
VCC731	0.0050	247	115.4	0.25	43.9	5.871	0.05	70	1, 2
VCC1632	0.0038	250	85.12	0.05	-60.7	7.088	0.05	55	1, 2
NGC4874	0.0231	266	23.8	0.13	49.3	1.860*	0.05	45	1, 3
NGC7626	0.0130	274	20.1	0.17	20.5	5.240*	0.05	121	1, 3
NGC5557	0.0130	281	16.2	0.17*	97.0*	3.910*	0.05	23	3
NGC1272	0.0180	292	20.7	0.06*	76.0*	2.440*	0.05	23	3
NGC6482	0.0130	322	10.1	0.15	68.3	2.870*	0.05	68	1, 3

Table 2.2: Result of the SIE and SNFW model fits to the *Truth* data from each lens. From left to right, we display the galaxy name, lens configuration, reduced χ^2 for the fit with a SIE, Einstein radius of the SIE, ellipticity, position angle, shear, and shear angle, χ^2 of the SNFW model fit, the normalization of the Sérsic profile, ellipticty and position angle of the Sérsic profile, Sérsic index, NFW halo normalization κ_s, external shear, external shear angle, and finally the NFW scale radius. A hyphen indicates that a parameter was held fixed to a value either obtain from the literature or fit with `galfit` (see Table 1). The reduced χ^2 are very small because we do not add measurement noise to our data.

Name	type	χ^2_{SIE}	χ^2_{SIE}	$R_{\rm Ein}$	ϵ	θ_ϵ	γ	θ_γ	χ^2_{SNFW}	χ^2_{SNFW}	N	ϵ	θ_ϵ	$R_{1/2}$	n	κ_s	γ	θ_γ	R_s
		(position)	(time delay)	arcsec		degrees		degrees	(position)	(time delay)	$\Sigma/\Sigma_{\rm crit}$		degrees	arcsec		$\Sigma/\Sigma_{\rm crit}$		degrees	arcsec
VCC1664	CUSP	0.000	0.003	0.39	0.32	-40.4	0.02	6.3	0.000	0.000	6.66	0.70	-42.3	-	1.220	0.303	0.09	41.1	1.020
	FOLD	0.000	0.019	0.39	0.30	-40.7	0.03	19.0	0.000	0.001	126.48	0.63	-42.5	-	2.808	0.277	0.08	40.9	1.010
VCC1297	CUSP	0.000	0.000	0.44	0.09	27.0	0.05	21.8	0.141	0.000	193.42	-	-	-	-	2.634	0.08	23.8	0.150
	FOLD	0.000	0.000	0.44	0.09	20.6	0.05	26.6	0.291	0.000	24.93	-	-	-	-	2.804	0.08	24.2	0.140
VCC798	CUSP	0.000	0.000	0.49	0.15	34.4	0.03	-26.9	0.000	0.000	145.92	0.18	36.6	-	4.382	0.041	0.03	-33.3	10.940
	FOLD	0.000	0.003	0.49	0.18	26.6	0.04	-16.5	0.002	0.053	1.82	0.44	37.5	-	1.788	0.075	0.05	30.9	11.250
VCC2092	CUSP	0.000	0.000	0.50	0.09	-2.7	0.04	29.6	0.000	0.000	4.07	0.16	-1.8	-	1.361	0.213	0.05	22.0	1.550
	FOLD	0.000	0.002	0.50	0.09	1.3	0.04	26.9	0.000	0.000	134.53	0.28	7.7	-	3.374	0.252	0.05	19.6	1.550
VCC1231	CUSP	0.000	0.001	0.53	0.18	8.1	0.11	28.1	0.004	0.053	175.03	-	-	-	-	0.304	0.09	36.7	1.110
	FOLD	0.000	0.003	0.52	0.22	6.7	0.12	24.6	0.002	0.123	216.16	-	-	-	-	0.267	0.09	34.8	1.120
VCC1062	CUSP	0.000	0.002	0.51	0.17	-7.8	0.06	15.2	0.000	0.001	79.09	0.41	-11.3	0.50	-	0.181	0.04	23.6	2.520
	FOLD	0.000	0.016	0.50	0.27	-5.7	0.11	11.4	0.018	0.021	61.79	0.25	21.6	1.18	-	0.009	0.11	4.8	5.930
VCC1692	CUSP	0.883	0.761	0.54	0.45	8.7	0.18	19.8	1.239	0.050	37.73	0.37	17.3	0.68	-	0.001	0.13	31.6	3.410
	FOLD	0.001	0.017	0.57	0.13	2.9	0.02	-42.1	0.302	0.505	174.09	-	-	-	-	0.636	0.04	23.9	0.530
VCC2000	CUSP	0.000	0.000	0.60	0.08	-19.2	0.03	-6.2	0.725	0.819	107.05	-	-	-	-	0.925	0.05	78.3	0.520
	FOLD	0.000	0.002	0.60	0.10	6.1	0.04	-25.3	0.000	0.004	54.16	0.18	-34.9	-	1.422	0.170	0.07	-15.1	0.670
VCC355	CUSP	0.000	0.000	0.65	0.01	-12.1	0.05	37.7	0.443	0.008	106.28	-	-	-	-	0.903	0.05	35.4	0.570
	FOLD	0.000	0.000	0.65	0.02	-12.4	0.05	38.4	1.011	0.008	42.98	-	-	-	-	0.887	0.05	35.0	0.590
NGC4872	CUSP	0.000	0.003	0.77	0.24	8.4	0.04	36.8	0.000	0.010	91.82	0.58	15.8	0.20	-	0.355	0.06	-18.9	1.220
	FOLD	0.000	0.004	0.76	0.25	9.3	0.05	35.5	0.000	0.044	24.29	0.42	4.0	-	3.229	0.101	0.03	40.4	10.560
VCC1903	CUSP	0.000	0.000	0.83	0.26	-17.3	0.05	-44.9	0.001	0.007	7.26	0.27	40.1	-	1.976	0.060	0.04	-16.7	6.870
	FOLD	0.000	0.017	0.83	0.26	-18.8	0.06	43.6	0.001	0.061	3.41	0.22	-15.8	-	1.346	0.006	0.05	-44.2	6.890
VCC881	CUSP	0.000	0.001	0.84	0.26	35.1	0.07	-32.1	0.001	0.001	25.80	0.24	30.3	-	3.404	0.010	0.05	-34.0	26.820
	FOLD	0.000	0.001	0.84	0.24	33.0	0.06	-31.3	0.000	0.004	14.26	0.22	28.9	-	2.996	0.001	0.05	-38.1	26.450
IC4051	CUSP	0.000	0.024	0.88	0.26	15.8	0.06	37.3	0.001	0.036	26.95	0.23	14.8	-	3.359	0.000	0.06	37.4	26.260
	FOLD	0.000	0.003	0.88	0.27	18.1	0.07	38.4	0.000	0.000	79.90	0.26	9.8	1.30	-	0.048	0.06	-6.6	6.570
NGC5322	CUSP	0.000	0.001	0.89	0.20	6.0	0.06	35.7	0.054	0.221	329.38	-	-	-	-	0.139	0.06	50.3	3.790
	FOLD	0.000	0.000	0.89	0.18	3.6	0.05	38.1	0.000	0.001	38.14	0.30	-0.4	-	2.692	0.181	0.05	-34.0	3.590

Table 2.3: Table 2.2 cont.

Name	type	χ^2_{SIE} (position)	χ^2_{SIE} (time delay)	$R_{\rm Ein}$ arcsec	ϵ	θ_ϵ degrees	γ	θ_γ degrees	χ^2_{SNFW} (position)	χ^2_{SNFW} (time delay)	N $\Sigma/\Sigma_{\rm crit}$	ϵ	θ_ϵ degrees	$R_{1/2}$ arcsec	n	κ_s $\Sigma/\Sigma_{\rm crit}$	γ	θ_γ degrees	R_s arcsec
NGC1132	CUSP	0.000	0.010	0.97	0.23	-37.3	0.02	22.3	0.494	0.311	8.72	-	-	-	-	0.150	0.03	-52.4	5.090
	FOLD	0.000	0.019	0.97	0.23	-37.0	0.02	20.2	0.021	0.413	6.85	-	-	-	-	4.038	0.06	-46.8	0.180
VCC731	CUSP	0.000	0.002	1.00	0.20	36.2	0.05	2.4	0.000	0.000	27.85	0.19	40.2	-	3.384	0.014	0.04	1.6	26.400
	FOLD	0.000	0.011	1.00	0.17	42.0	0.04	-1.2	0.000	0.016	4.72	0.16	-44.2	-	1.662	0.001	0.03	1.1	9.880
VCC1632	CUSP	0.000	0.000	1.01	0.07	-32.5	0.07	-32.6	0.056	1.625	943.00	-	-	-	-	0.314	0.03	57.3	3.730
	FOLD	0.000	0.002	1.01	0.07	-21.3	0.07	-29.3	0.000	0.006	18.04	0.26	-42.7	-	3.116	0.227	0.03	-31.0	5.260
NGC4874	CUSP	0.001	0.020	1.18	0.02	34.6	0.06	43.9	0.103	0.012	3.32	-	-	-	-	0.130	0.03	42.9	9.010
	FOLD	0.000	0.006	1.18	0.03	26.4	0.06	42.2	0.337	0.034	3.40	-	-	-	-	0.127	0.03	42.2	9.070
NGC7626	CUSP	0.004	0.006	1.23	0.09	35.6	0.07	35.4	0.852	0.936	39.21	-	-	-	-	0.369	0.04	-54.7	3.490
	FOLD	0.000	0.085	1.23	0.09	-40.1	0.07	41.8	0.000	0.059	16.12	0.13	-43.0	-	1.933	0.102	0.07	40.1	4.290
NGC5557	CUSP	0.000	0.005	1.28	0.09	9.1	0.07	19.8	0.181	0.152	564.82	-	-	-	-	0.231	0.06	23.3	3.500
	FOLD	0.000	0.000	1.28	0.10	9.5	0.07	19.7	0.000	0.020	14.67	0.28	-21.0	-	2.600	0.347	0.04	23.1	3.360
NGC1272	CUSP	0.000	0.029	1.41	0.07	15.9	0.07	19.4	0.428	0.337	6.60	-	-	-	-	0.291	0.04	19.4	4.470
	FOLD	0.000	0.019	1.40	0.07	40.4	0.07	27.9	0.000	0.010	3.29	0.64	4.9	-	2.160	0.299	0.04	26.8	5.140
NGC6482	CUSP	0.002	0.164	1.69	0.17	-0.8	0.05	34.4	0.001	0.188	29.40	0.25	-7.0	-	3.343	0.073	0.03	36.6	26.420
	FOLD	0.017	2.410	1.70	0.16	-13.7	0.03	40.9	0.015	1.672	39.70	0.27	-14.8	1.75	-	0.146	0.04	8.4	8.800

2.5 Appendix A: From surface brightness to surface mass density

The dimensionless surface mass densities of an NFW halo and an image of a galaxy are given by:

$$\kappa_D(r) = 2\kappa_s \frac{1 - F(r/r_s)}{(r/r_s)^2 - 1} = \kappa_s g(r) \qquad\qquad \kappa_B = \lambda\, c(r)$$

where

$$F(x) = \begin{cases} \frac{\tanh^{-1}\left[\sqrt{1-x^2}\,\right]}{\sqrt{1-x^2}}; & x \le 1 \\ 1 & x = 1 \\ \frac{\tan^{-1}\left[\sqrt{x^2-1}\,\right]}{\sqrt{x^2-1}}; & x \ge 1 \end{cases}$$

with $x \equiv r/r_s$, and r_s the scale radius of the NFW halo. λ is a normalization factor we apply to the images obtained from the HST with units of [convergence]/[pixel count], and c(r) is a 2 dimensional image with pixel values corresponding to photon counts. There are two free parameters (κ_s, λ), and two constraints on the surface mass densities of dark matter and baryons (κ_D, κ_B):

- the average convergence (baryons plus NFW) within the Einstein radius $R_{\rm Ein}$ is 1, a standard result for lenses with circular symmetry:

$$\bar{\kappa}_s = \frac{1}{\pi R^2_{\rm Ein}} \int dA_{R_{\rm Ein}} \kappa(x) = \frac{1}{\pi R^2_{\rm Ein}} \int (\kappa_B + \kappa_D)\, dA_{R_{\rm Ein}} = 1$$

- The contribution to the total convergence within $R_0 = \frac{R_{1/2}}{2}$ from the NFW halo is some fraction f:

$$\frac{\int dA_{R_0}\ \kappa_D}{\int dA_{R_0}\left(\kappa_B + \kappa_D\right)} = f$$

Inserting the expressions for κ_D and κ_B into (1) and (2) yields two equations in two unknowns:

$$\frac{1}{\pi R_{\rm Ein}^2}\int dA_{R_{\rm Ein}}\Big[\kappa_s g\left(r\right) + \lambda\ c\left(r\right)\Big] = 1$$
$$\left[\int dA_{R_0}\kappa_s g\left(r\right)\right]\left[\int dA_{R_0}\left[\kappa_s g\left(r\right) + \lambda c\left(r\right)\right]\right]^{-1} = f$$

It is useful to introduce the notation:

$$\int dA_R\ g\left(r\right) = 2\pi R_s^2\ G\left(n\right);\ G\left(n\right) \equiv \log\left(\frac{n^2}{4}\right) + \frac{2\tanh^{-1}\left(\sqrt{1-n^2}\right)}{\sqrt{1-n^2}};\ n \equiv \frac{R}{R_S}$$
$$\int dA_R\ c\left(r\right) = \pi R^2\left(\frac{1}{N}\sum_{i,j=0}^{N} D_{ij}\right) = \pi R^2\langle C\left(r\right)\rangle\ \ ;\ [r <\ R]$$

Where the integral over the count map is expressed as a discrete sum over pixels within the radial limit of integration.

Solving for κ_s and λ:

$$\kappa_s = \frac{R_{\rm Ein}^2}{2R_s^2}\left[G\left(n_1\right) + G\left(n_2\right)\left(\frac{1-f}{f}\right)\left(\frac{R_{\rm Ein}}{R_0}\right)^2\frac{\langle C\rangle_{R_{\rm Ein}}}{\langle C\rangle_{R_0}}\right]^{-1}$$
$$\lambda = \left(\frac{1-f}{f}\right)\frac{R_{\rm Ein}^2}{R_0^2}\left[\frac{G\left(n_1\right)}{G\left(n_2\right)}\langle C\rangle_{R_0} + \left(\frac{1-f}{f}\right)\left(\frac{R_{\rm Ein}}{R_0}\right)^2\langle C\rangle_{R_{\rm Ein}}\right]^{-1}$$

where $\left[n_1 \equiv \frac{R_{\rm Ein}}{R_s}, n_2 \equiv \frac{R_0}{R_s}\right]$. The scale radius R_s is taken to be $5R_{1/2}$, while the number f is taken to be $\frac{1}{3}$, consistent with the results of Auger et al. (2010). With these choices, the only free parameters are the Einstein radius $R_{\rm Ein}$, which we obtain via the measured central velocity dispersion of the galaxy, and the half-light radius for each galaxy found in the literature. To check the validity of this approach, we calculate the stellar mass within the

Table 2.4: Stellar mass estimates derived via our normalization procedure and by Gallo et al. (2008).

Galaxy	$Log_{10}M_{\odot}$ (from convergence map)	$Log_{10}M_{\odot}$ (from Gallo, Treu et al 2008)
VCC1664	10.8	10.6
VCC1692	10.8	10.6
VCC2000	10.6	10.4
VCC881	11.5	11.9
VCC731	11.3	11.7
VCC1903	11.3	11.3
VCC1231	10.9	10.8
VCC355	10.6	10.3
VCC1062	10.6	10.7

Einstein radius of galaxies in our sample, and compare the results to stellar mass estimates in the literature. The results in Table 1 confirm our normalization procedure does not imply an unrealistic mass to light ratio.

2.6 Appendix B: Simulating an extended source

To get around the issue of background noise in the optical images that introduces an artificial micro-lensing signal, which results in a large scatter in the distribution of image magnifications, we model the source as an extended object of diameter 5 parsecs in the source plane.

To model an extended source, we take the original source position as the center of a 2-d Gaussian, characterized by a covariance matrix

$$\Sigma_0 = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_0^2 \end{bmatrix}$$

where we take $\sigma_0 = 2.5$ pc. We take the area of the source to be

$$A_{src} = \pi\sigma_0^2$$

which corresponds to a circle in source plane of diameter 5 parsecs. We draw 100 random source positions from this distribution and solve the lens equation with gravlens for each one, yielding 4 images positions for each source position. This process results in 4 clusters of 100 points each, with each cluster described by its own covariance matrix describing an ellipse in the image plane. The area of this ellipse is given by

$$A_{img} = \pi\sqrt{\lambda_1\lambda_2}$$

where (λ_1, λ_2) are the eigenvalues of the covariance matrix describing the 100 (x, y) coordinates for each of the four images. The magnification for each image is then given by the ratio of A_{img} to the area of the source:

$$M_i = \frac{A_{img}}{A_{src}}$$

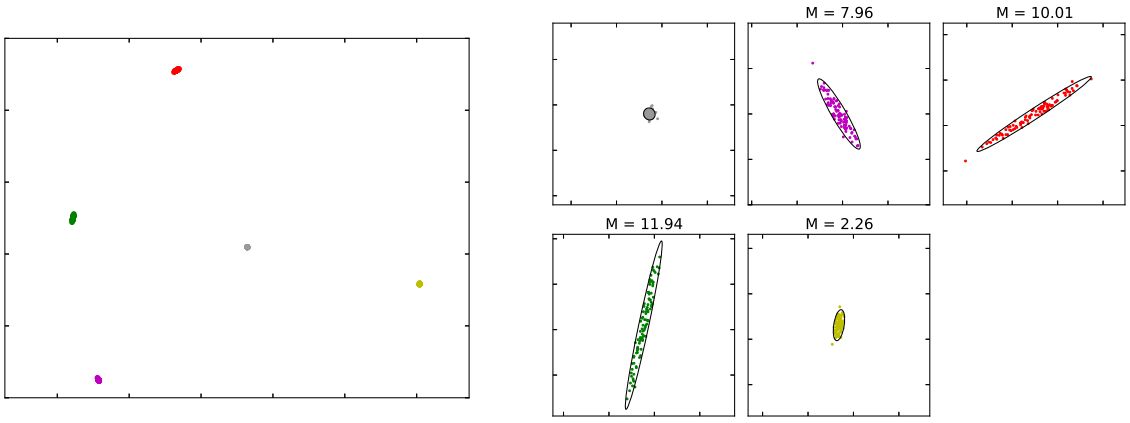

Figure 2.22: The individual panels on the right hand side show zoomed-in images of the clusters of points in the left panel. The grey points are drawn from a circular Gaussian distribution, centered at a reference source position, simulating an extended background source of diameter 5 pc. For each of these source points, we use *gravlens* to solve the lens equation, resulting in 4 additional points, each representing an image produced by the lens system. After repeating this procedure 100 times, the area of the ellipse describing the covariance matrix for each set of points is used to compute the compute the magnification. This procedure is repeated for each of the 250 randomly sampled reference source positions. The ellipses in the right-hand panels of correspond to 90% confidence intervals.

CHAPTER 3

Probing the nature of dark matter by forward modelling flux ratios in strong gravitational lenses

This chapter was published as Gilman, D., et al. Probing the nature of dark matter by forward modelling flux ratios in strong gravitational lenses. MNRAS 481, 819-834 (2018), and is printed here with minor formatting adjustments.

3.1 Introduction

Dark matter models make testable predictions regarding the abundance and mass profiles of substructure in galactic dark matter halos. In the standard ΛCDM picture, structure grows bottom-up at practically all length scales (Schneider et al., 2013), resulting in a scale-free mass function for subhalos (Springel et al., 2008; Gao et al., 2011; Fiacconi et al., 2016), and density profiles fit by the Navarro Frenk and White (NFW) profile (Navarro et al., 1996). In contrast, in warm dark matter (WDM) models, free streaming washes out small density perturbations, resulting in a paucity of structure below a certain scale, which depends on the free streaming scale of the dark matter particle(s) (Schneider et al., 2012; Pullen et al., 2014; Viel et al., 2013; Lovell et al., 2016; Bose et al., 2016; Menci et al., 2016). Thermal relics and sterile neutrinos represent two WDM candidates (Kusenko, 2009; Abazajian, 2017), and under certain assumptions produce similar mass functions. Self-interacting dark matter alters the density profiles of individual subhalos, producing cores rather than cusps in the center of halos (Schneider et al., 2017; Vogelsberger et al., 2016; Kamada et al., 2017).

These alternative models to ΛCDM have gained traction, motivated by apparent failures of the ΛCDM picture on small scales (Bullock & Boylan-Kolchin, 2017). For example, by

invoking WDM one explains the non-detection of low-mass satellites in the Milky Way, the "missing satellites problem" (Klypin et al., 1999), by reducing the expected number of subhalos. In another example, rotation curves of satellite galaxies imply shallower-than-isothermal inner density profiles, characteristic of the mass profiles associated with self interacting dark matter.

These apparent anomalies are based on observations of luminous structure and rely on assumptions about the connection between baryons and dark matter halos. Under different combinations of models for star formation and kinematic data in satellites, some small-scale challenges to the ΛCDM picture can be resolved (Kim et al., 2017). Unfortunately, it is difficult to measure kinematics of low mass galaxies owing to the small number of stars that can be used for this purpose. This leads to large uncertainties in the inferred halo masses and densities. Different models for tidal disruption and baryonic feedback (Despali & Vegetti, 2017; Garrison-Kimmel et al., 2017), differing luminosity functions for dark subhalos (Nierenberg et al., 2016), and large scatter in stellar-mass-halo-mass relations below $10^8 M_\odot$ (Munshi et al., 2017) complicate constraints on the nature of dark matter. An independent and direct probe, which does not rely on assumptions regarding the physics of star formation in low mass galaxies, is needed to disentangle the physics of baryons and the nature of dark matter.

Gravitational lensing offers a direct probe of dark substructure below halo masses of $10^8 M_\odot$. Lensing relates a set of three observables - time delays, positions, and magnifications - to the gravitational potential along a path traversed by light emitted by a background source. As the observables depend only on the gravitational potential of the deflector and the potential along the line of sight, lensing offers a tool to study dark matter substructure directly, without relying on baryons as tracers.

Various techniques employ strong lensing as a probe of dark matter structure. When the light from a spatially extended background source is warped by a foreground deflector into a highly magnified arc, substructure can distort the arc. By simultaneously modeling the mass distribution in the lens plane and reconstructing the luminosity distribution of the background source, one can infer the mass of a perturber and constrain the subhalo mass

function (Koopmans, 2005; Vegetti & Koopmans, 2009; Vegetti et al., 2012, 2014; Li et al., 2016; Hezaveh et al., 2016b; Birrer et al., 2017b; Vegetti et al., 2018).

Flux ratios in lensed quasars offer an alternative probe of dark substructure (Mao & Schneider, 1998b; Metcalf & Madau, 2001; Dalal & Kochanek, 2002; Chiba, 2002). The sensitivity of this observable derives from the compact size of a quasar and the fact that lensing magnifications can be perturbed by a subhalo whose deflection angle is comparable to or larger than the source size (Dobler & Keeton, 2006). From a theoretical standpoint, several forecasting studies (Xu et al., 2012; Graus et al., 2018) use N-body simulations to anticipate the lensing signal from CDM substructures. From an observational and lens modeling point of view, given observed flux ratios, one can add a subhalo to the lens model and vary its properties to infer the mass (Fadely & Keeton, 2012; MacLeod et al., 2013; Nierenberg et al., 2014) or rule out the presence of substructure near the images (Nierenberg et al., 2017). This lens modeling technique, and the direct detection of subhalos via gravitational imaging, comprise a class of observations that yield constraints on individual substructures.

Recently, authors have called attention to a potential bias present in flux ratios, wherein the morphology of deflector in the lens plane produces flux ratio anomalies reminiscent of those induced by substructure. Gilman et al. (2017) showed that realistic deflectors with a luminous mass component drawn from HST images of nearby galaxies occasionally produce flux ratio anomalies with respect to a simplified smooth lens model. Hsueh et al. (2018) performed a similar analysis with galaxies produced in the Illustris simulation (Vogelsberger et al., 2014), and reached similar conclusions. In two observed lenses, Hsueh et al. (2016, 2017) argued that a disk component in the deflector can explain observed flux ratio anomalies. These effects may contribute to frequency of observed flux ratio anomalies in strong lenses, which occur more frequently than one would expect in a CDM scenario, as pointed out by Xu et al. (2015). Among non-dark matter sources of flux ratio anomaly, microlensing can also induce drastic fluctuations in image magnifications, although this effect can be mitigated by using flux ratios measured from the more spatially extended narrow-line region of a background quasar (Moustakas & Metcalf, 2003; Nierenberg et al., 2014, 2016).

In the context of dark matter, it is important to note that models predict large ensembles

of dark subhalos, which could act together to affect a lensed image. In contrast to single subhalo models, other methods attempt to probe the collective impact of numerous perturbers whose individual effects are not statistically significant, but which together produce a measurable signal (Dalal & Kochanek, 2002; Fadely & Keeton, 2012; Hezaveh et al., 2016a; Cyr-Racine et al., 2016; Daylan et al., 2018; Birrer et al., 2017b). Since these methods do not require high significance detections of individual perturbers, they extract information from a larger area around Einstein rings.

In an example, Birrer et al. (2017b) quantify substructure in the lens RX J1131−1231 by modeling surface brightness anomalies detected in HST imaging data. Through a forward modeling approach that relies on generating an extensive suite of realistic simulations, they are able to constrain models of dark matter statistically. They could rule out WDM mass functions with thermal relics below the mass of 2 keV at 2σ.

In this work, we present a statistical method that utilizes the flux ratios from an ensemble of multiply imaged quasars to distinguish between dark matter models. Our technique takes as input data from a sample of strong lens systems and a prescription for rendering substructure realizations for a dark matter theory, and returns posterior probability distributions for the parameters describing the substructure population. We use the technique of Approximate Bayesian Computing (ABC; Rubin, 1984), also applied in Birrer et al. (2017b), which enables an application of Bayesian statistics to the problem of substructure lensing. In our framework, we are able to efficiently explore the parameter space spanned by dark matter models with different predictions regarding the nature of substructure without explicitly computing a likelihood function, which in substructure lensing is a computationally and analytically daunting task. ABC permits one to circumvent calculation of formidable likelihoods through the use of summary statistics, which quantify agreement between an observation and data computed in a forward model. Our method also naturally accommodates joint inference from multiple strong lens systems. The method can be applied to any parameterization of dark substructure, provided one specifies the mass function, spatial distribution, and density profile of individual subhalos. Since the method relies on flux ratio statistics rendered in a forward model, accurate lens models and control over systematic

errors in flux ratios are crucial for attaining robust constraints. Finally, by omitting line of sight substructure that is expected to contribute significantly to flux ratio anomalies (Chen et al., 2003; Inoue & Takahashi, 2012; Despali et al., 2018b), we do not capture the full information content of each lens, so our results can be interpreted as conservative theoretical bounds.

We present the formalism of our method and illustrate its general capabilities via a case study in which we distinguish between two simplified dark matter models. We consider a subhalo mass function with variable normalization and damping below a free-streaming scale, and provide forecasts for the constraints afforded under different flux precisions with up to 180 systems, which is a sample size attainable with future surveys such as Euclid, LSST and WFIRST (Oguri & Marshall, 2010). With our forward modeling framework, we forecast the possible constraints on a WDM subhalo mass function using flux ratios, and quantify how these constraints scale with the number of lenses, the uncertainty in fluxes, and the overall normalization of the mass function.

This paper is structured as follows. Section 3.2 poses the problem of substructure lensing in a Bayesian framework and reviews the basics of Approximate Bayesian Computing. In section 3.3, we describe our parameterization for the subhalo mass function, our method for creating mock data sets, and the procedure to compute posterior distributions from the forward model. In Section 6.4, we examine how the signal from different substructure models appears surfaces as flux ratio anomalies, and provide forecasts for constraints on the half-mode mass for under various levels of precision in image flux measurements. We use a cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $h = 0.7$. We use the software package `lensmodel` to solve the lens equation and fit smooth models to lensing observables (Keeton, 2011).

3.2 Bayesian inference on the subhalo mass function

In this section we describe how we infer the parameters describing subhalo populations within a Bayesian framework, and propose an Approximate Bayesian Computing (ABC) algorithm to contend with the highly stochastic nature of substructure lensing. Section 3.2.1 derives

Table 3.1: Definitions and descriptions for parameters relevant to Equation 3.7

parameter	definition	description
$\mathbf{d_n}$	data from the nth lens	positions, time delays, flux ratios
$\mathbf{q}_{\rm sub}$	vector of hyper-parameters describing the global subhalo mass function	$(A_0, m_{\rm hm})$, spatial distribution
A_0	normalization of subhalo mass function	1% substructure mass fraction at $R_{\rm Ein} \approx A_0 = 2 \times 10^8 {\rm M}_\odot{}^{-1}$ mass range $10^6 \leq M_{200} \leq 10^{10}\,[M_\odot]$
$m_{\rm hm}$	half-mode mass	number of subhalos below $m_{\rm hm}$ is strongly suppressed
$\mathbf{m}_{\rm sub}$	defines parameters for an individual substructure realization	subhalo positions, masses, density profiles
$\mathbf{q}^\star_{\rm mac(n)}$	maximum-likelihood macromodel for nth lens	fits nth positions, time delays in presence of substructure
$\mathbf{f_n}^\star$	flux ratios computed in the forward model for nth lens	computed with $\mathbf{q}^\star_{\rm mac(n)}$ as opposed to true mass distribution

the expression for the posterior distribution of dark matter model parameters given a set of lensing observables. In Section 3.2.2, we briefly review the technique of Approximate Bayesian Computing.

3.2.1 Connecting dark matter model parameters to lensing observables

The observables from a strong lens system are image time delays $\mathbf{t}$, positions $\mathbf{x}$, and fluxes $\mathbf{f}$.[1] The data vector for the nth lens can be written $\mathbf{d}_{\rm n} = \{\mathbf{x_n}, \mathbf{t_n}, \mathbf{f_n}\}$, and we represent the dataset for a sample of N lenses as set $\mathbf{D} = \{\mathbf{d_1}, \mathbf{d_2}, ..., \mathbf{d_N}\}$.

Given a dark matter model with global properties described by a set of hyper-parameters $\mathbf{q}_{\rm sub}$, the desired posterior distribution for the parameters $\mathbf{q}_{\rm sub}$ is given by

$$p\left(\mathbf{q}_{\rm sub}|\mathbf{D}\right) \propto \mathcal{L}\left(\mathbf{D} \mid \mathbf{q}_{\rm sub}\right) \pi\left(\mathbf{q}_{\rm sub}\right) \tag{3.1}$$

[1] To impose constraints, we actually use flux ratios in order to divide out the unknown source flux.

where $\pi\left(\mathbf{q}_{\rm sub}\right)$ is the prior probability for the parameters. In practice, $\mathbf{q}_{\rm sub}$ describes the shape and normalization of the subhalo mass function, the spatial distribution of subhalos, the density profile of subhalos, etc.

Since the data from each lens is independent, the joint likelihood in Equation 3.1 can be written as a product of the likelihoods for each lens

$$\mathcal{L}\left(\mathbf{D} \mid \mathbf{q}_{\rm sub}\right) = \prod_{n=1}^{N} \mathcal{L}\left(\mathbf{d}_{\rm n} \mid \mathbf{q}_{\rm sub}\right). \tag{3.2}$$

We specify the model for each lens as a combination of two mass components. The first is a macromodel, which accounts for most of the mass of the deflector and its environment. For the nth lens, we denote this component $\mathbf{q}_{\rm mac(n)}$. The second component is the substructure population described by by $\mathbf{q}_{\rm sub}$. With these definitions, the components of $\mathbf{q}_{\rm mac(n)}$ are nuisance parameters which are marginalized out of the posterior

$$\mathcal{L}\left(\mathbf{d}_{\rm n} \mid \mathbf{q}_{\rm sub}\right) \propto \int p\left(\mathbf{d}_{\rm n} \mid \mathbf{q}_{\rm mac(n)}, \mathbf{q}_{\rm sub}\right) \pi\left(\mathbf{q}_{\rm mac(n)}\right) d\mathbf{q}_{\rm mac(n)}, \tag{3.3}$$

where we have assumed that the macromodel $\mathbf{q}_{\rm mac(n)}$ is independent of the dark matter parameters $\mathbf{q}_{\rm sub}$, and introduce the prior $\pi\left(\mathbf{q}_{\rm mac(n)}\right)$. The assumption that $\mathbf{q}_{\rm mac}$ and $\mathbf{q}_{\rm sub}$ are independent is not formally correct, as parameters such as the Einstein radius may be informative of the total halo mass and the normalization of subhalo mass function. Working with real datasets, this information would need to be incorporated in the analysis. For the purpose of forecasting the possible constraints on $\mathbf{q}_{\rm sub}$, we examine the case of two fixed normalizations that span the expected range of substructure abundance for the halo masses implied by the distribution of Einstein radii in our mock data (see Section 3.3). As we will demonstrate in Section 6.4, the information content of each lens scales with the overall normalization, such that the bounds on a sample of lenses with diverse halos masses will be bound by the two limiting cases of the overall normalization we analyze.

Dark matter models do not directly map $\mathbf{q}_{\rm sub}$ to a set of lensing observables. Rather, $\mathbf{q}_{\rm sub}$ specifies statistical distributions for the masses, positions, density profiles, etc. of the subhalos. Defining a vector $\mathbf{m}_{\rm sub}$ that specifies a specific substructure realization, the likelihood becomes

$$\mathcal{L}\left(\mathbf{d}_{\rm n} \mid \mathbf{q}_{\rm sub}\right) \propto \int \mathcal{L}\left(\mathbf{d}_{\rm n} \mid \mathbf{m}_{\rm sub}, \mathbf{q}_{\rm mac(n)}\right) p\left(\mathbf{m}_{\rm sub}|\mathbf{q}_{\rm sub}\right) \times \pi\left(\mathbf{q}_{\rm mac(n)}\right) d\mathbf{m}_{\rm sub} d\mathbf{q}_{\rm mac(n)} \tag{3.4}$$

To make progress in evaluating Equation 3.4, we note that the astrometric and time delay perturbations from substructure are generally small and can be partially absorbed by small adjustments in $\mathbf{q}_{\rm mac(n)}$.[2] The flux ratios, on the other hand, are determined by the second derivative of the gravitational potential near an image, thus the effects of substructure are difficult to reproduce by adjustments in $\mathbf{q}_{\rm mac(n)}$. With this in mind, we write image positions and time delays separately from the flux ratios, writing $\mathbf{d}_{\rm n} = \left\{\mathbf{d_{tx(n)}}, \mathbf{f_n}\right\}$, where $\mathbf{d_{tx(n)}}$ denotes the positions and time delays $\{\mathbf{t}, \mathbf{x}\}$. To relate $\mathbf{q}_{\rm sub}$ to $\mathbf{d_{tx(n)}}$ and $\mathbf{f_n}$, our strategy is to forward model simulated data sets of image positions, time delays, and fluxes $\left\{\mathbf{d'_{tx(n)}}\left(\mathbf{q}_{\rm mac(n)}, \mathbf{m}_{\rm sub}\right), \mathbf{f_n}'\left(\mathbf{q}_{\rm mac(n)}, \mathbf{m}_{\rm sub}\right)\right\}$, which depend on $\mathbf{q}_{\rm sub}$ through the realizations of substructure $\mathbf{m}_{\rm sub}$. The likelihood of observing $\mathbf{d}_{\rm n}$ is therefore

$$\mathcal{L}\left(\mathbf{d}_{\rm n} \mid \mathbf{m}_{\rm sub}, \mathbf{q}_{\rm mac(n)}\right) = \mathcal{L}\left(\mathbf{d_{tx(n)}} \mid \mathbf{d'_{tx(n)}}\right) \mathcal{L}\left(\mathbf{f_n} \mid \mathbf{f_n}'\right). \tag{3.5}$$

Next, we note that most choices of $\mathbf{q}_{\rm mac(n)}$, with a wide prior distribution, yield the incorrect positions and time delays, and therefore do not contribute substantially to the integral in Equation 3.4. We therefore approximate the marginalization over the macromodel parameters by fixing the macromodel in a certain configuration $\mathbf{q}^{\star}_{\rm mac(n)}$, which fits the image positions and time delays. This step avoids sampling the potentially vast parameter space of $\mathbf{q}_{\rm mac}$. Explicitly, $\mathbf{q}^{\star}_{\rm mac(n)}$ is defined by the relation

$$\mathbf{d_{tx(n)}} = \mathbf{d'_{tx(n)}}\left(\mathbf{q}^{\star}_{\rm mac(n)}, \mathbf{m}_{\rm sub}\right). \tag{3.6}$$

The procedure of re-optimizing the macromodel was also employed in Dalal & Kochanek (2002).

[2]Positions and time delays have some ability to probe substructure (Chen et al., 2007; Keeton & Moustakas, 2009), but flux ratios experience stronger perturbations that are our focus here.

By evaluating the flux ratios only with respect to $\mathbf{q}^{\star}_{\mathrm{mac(n)}}$, we effectively take a derivative, while formally an integral is required to marginalize over $\mathbf{q}_{\mathrm{mac}}$. The procedure of optimizing, rather than marginalizing, the macromodel will yield a good approximation to the true likelihood as long as the image fluxes do not vary significantly over the range of macromodel parameters space for which the image positions and time delays are fit. We verify that the variation in image fluxes for different macromodel configurations that fit the observed image positions is smaller that the typical $5-8\%$ variations derived empirically in (Gilman et al., 2017) by fitting smooth lens models to realistic deflectors. [3] In macromodel parameterizations more complicated than power-law ellipsoids that possess additional degrees of freedom, the fluxes may vary substantially even for configurations of $\mathbf{q}_{\mathrm{mac}}$ with fixed image positions, and the macromodel may be better able to absorb flux perturbations from substructure. Stellar disks fall into this category (Hsueh et al., 2016; Gilman et al., 2017), as do models with extreme angular structures (Congdon & Keeton, 2005), but the former are unlikely to be present in a sample of massive elliptical deflectors, and the latter are unphysical. If external sources of error were reduced such that the dominant source of flux uncertainty stemmed from marginalizing the macromodel, one would need to explicitly do the marginalization. The procedure outlined here should be amended to sample prior distributions of $\mathbf{q}_{\mathrm{sub}}$ and $\mathbf{q}_{\mathrm{mac}}$ constructed on a lens-by-lens basis when working with real data, but for the purpose of computational expediency and making approximate forecast statements we leave this level of detail for future work.

After replacing the integral over $\mathbf{q}_{\mathrm{mac}}$ with $\mathbf{q}^{\star}_{\mathrm{mac(n)}}$, Equation 3.4 becomes

$$\mathcal{L}\left(\mathbf{d}_{\mathrm{n}}|\mathbf{q}_{\mathrm{sub}}\right) \propto \int \mathcal{L}\left(\mathbf{f_n}\ |\mathbf{f_n}^{\star}\right) p\left(\mathbf{m}_{\mathrm{sub}}, \mathbf{q}_{\mathrm{sub}}\right) d\mathbf{m}_{\mathrm{sub}} \tag{3.7}$$

where we introduce the notation $\mathbf{f_n}^{\star} = \mathbf{f_n}'\left(\mathbf{q}^{\star}_{\mathrm{mac(n)}}, \mathbf{m}_{\mathrm{sub}}\right)$. The parameters relevant to Equation 3.7 are summarized in Table 3.1.

[3] We find that constraining the magnitude of the external shear, external shear angle, axis ratio and position angle to $0.01, 5^{\circ}, 0.05, 5^{\circ}$, as is possible with detailed modeling (e.g. Wong et al., 2017), is sufficient to ensure the flux variations associated with marginalizing the macromodel is below 4% per image. We also verify that uncertainty associated with the power-law slope of the main deflector does not incur a serious bias in our forecasts (see Appendix 3.10).

At this step, a Markov Chain Monte Carlo integration scheme would be inefficient, as the flux ratios corresponding to the overwhelming majority of realizations $\mathbf{m}_{\rm sub}$ would not match those observed in the data. Rather than computing Equation 3.7 directly, we employ a computational method that allows us to efficiently explore the parameter space spanned by $\mathbf{q}_{\rm sub}$.

3.2.2 Approximate Bayesian Computing

Approximate Bayesian Computing (ABC) is a computational algorithm rooted in Bayesian statistics that circumvents the direct calculation of intractable likelihoods, and enables inferences from simulated data sets computed in a forward model. For details in addition to those presented in this section, see e.g. Csilléry et al. (2010); Lintusaari et al. (2017). In recent years, ABC has seen applications across a wide range of problems in cosmology and astrophysics (Weyant et al., 2013; Hahn et al., 2017; Birrer et al., 2017b; Herbel et al., 2017).

In an implementation of ABC, one draws samples from a prior probability distribution, creates a forward model of simulated data from the samples, compresses the data sets into summary statistics (optional, but often necessary to keep computation costs low), and accepts or rejects the samples based on the similarity of the simulated to the observed data. An implementation of the algorithm therefore proceeds as follows:

1. Sample from a set of model parameters θ.

2. From the samples θ, forward-model a set of simulated data $\mathbf{d}'$.

3. The data vector $\mathbf{d}'$ is often multi-dimensional, but in many cases the relevant information that will discriminate between different parameters θ is contained in only a subset of the data. To reduce the dimensions of the problem, introduce the summary statistics $S(\mathbf{d}')$ and $S(\mathbf{d})$, which compress the relevant information contained in $\mathbf{d}'$ and $\mathbf{d}$.

4. Introduce a distance metric $R\left(S(\mathbf{d}), \mathbf{S}(\mathbf{d}')\right)$ - for instance, the Euclidean distance between the summary statistics in N-dimensions - and accept the proposition θ under

the requirement

$$R\left(S(\mathbf{d}'), S(\mathbf{d})\right) \leq \epsilon \tag{3.8}$$

for some tolerance ϵ.

5. Repeat steps [1-4] until the distribution of accepted samples is stable to changes in the total number of samples computed, and the total number of samples retained in the posterior.

Formally, when implementing ABC one obtains samples from the posterior density

$$p\left(\theta|\mathbf{R}\left(\mathbf{d}', \mathbf{d}\right) \leq \epsilon\right), \tag{3.9}$$

with the property

$$p\left(\theta|\mathbf{d}\right) = \lim_{\epsilon \to\ 0} \left[p\left(\theta|\mathbf{R}\left(\mathbf{d}', \mathbf{d}\right) \leq \epsilon\right)\right]. \tag{3.10}$$

Thus, assuming the summary statistic contains the information necessary to distinguish between different models, the distribution of accepted samples from θ converges to the true posterior as ϵ tends to zero. Put another way, the relative number of accepted samples between multiple competing models reflects the relative probabilities of these models as $\epsilon \to 0$. In practice, one must compromise between an ϵ large enough to ensure timely convergence of the ABC procedure, and a value stringent enough to ensure the distribution of accepted samples is representative of the true posterior.

Crucial steps in the implementation of ABC include the choice of summary statistic $S(\mathbf{d})$, and the acceptance criterion ϵ. A summary statistic which erodes the discriminating information contained in the data will not converge to the true posterior. In a similar vein, an acceptance threshold too lax will result in a posterior distribution too broad, with the extreme limit of accepting all samples from θ and returning the prior. For this reason, assuming the algorithm has converged, the joint posterior distribution for the model parameters approximated in ABC will always be conservative, in the sense that it contains more volume than the true posterior.

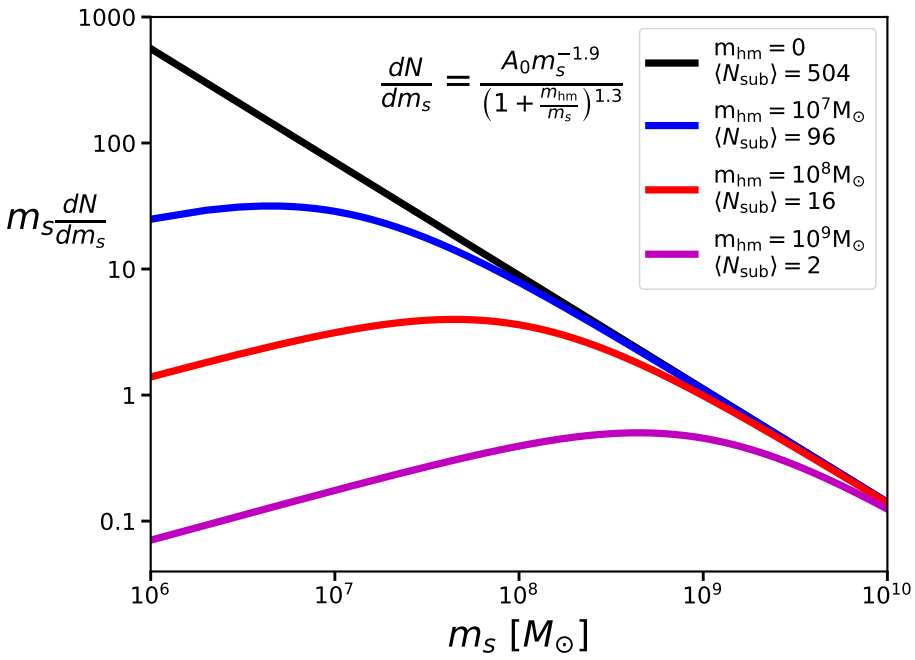

Figure 3.1: The subhalo mass function of Equation 3.13 which we use to generate substructure realizations. In the figure we vary the half-mode mass $m_{\rm hm}$ with fixed normalization $A_0 = 1.2 \times 10^8 {\rm M}_\odot{}^{-1}$. We generate subhalo populations in such a way that the amplitude of a CDM-like and a WDM-like mass functions are identical for masses $m \gg m_{\rm hm}$, rendering subhalos in projection to a radius of 18.6 kpc.

In the context of substructure lensing, we compute a summary statistic for each realization based on the observed flux ratios, or the fluxes of three images normalized by the flux of the fourth. The summary statistic we use based on the observed flux ratios $\mathbf{f_n}$ and the forward model flux ratios $\mathbf{f_n}^\star$ is given by

$$S\left(\mathbf{f_n}, \mathbf{f_n}^\star\right) = \sqrt{\sum_{i=1}^{3} \left(f_{n(i)} - f^\star_{n(i)}\right)^2}, \tag{3.11}$$

where the summation runs over the three flux ratios of the nth lens, making use of the full information contained in these data. For an example of results using a different summary statistic, see Appendix 5.10.

3.3 Simulation Setup

In this section, we describe our lensing simulations, including the ingredients necessary to render substructure realizations and how we implement ABC to constrain the subhalo mass function. Section 3.3.1 describes our prescription for modeling substructure populations for both cold and warm dark matter scenarios. Section 3.3.2 describes the mock data sets we use in our simulations, including lens and source redshift configurations. In section 3.3.3, we explain how we use the information contained in the forward model within the ABC framework to make inferences on $\mathbf{q}_{\rm sub}$.

3.3.1 Subhalo density profile, mass definition, and mass function

When quoting subhalo masses, we refer to the mass inside a sphere of radius r_{200}, M_{200}. We model subhalos as tidally truncated NFW profiles (Baltz et al., 2009)

$$\rho\left(r\right) = \frac{\rho_0}{x\left(1+x\right)^2}\frac{\tau^2}{x^2+\tau^2} \tag{3.12}$$

where $x = \frac{r}{r_s}$, $\tau = \frac{r_t}{r_s}$, r_t is the truncation radius and r_s is a scale radius. [4] The finite, truncated mass can differ from M_{200} to a varying degree depending on the truncation radius and the concentration, but the effect on image flux ratios is primarily determined by the central density, with the truncation playing a sub-dominant role provided $r_t > r_s$.

We render subhalos with M_{200} (denoted m_s) between $10^6 \leq \frac{m_s}{M_\odot} \leq 10^{10}$, drawing from a subhalo mass function written as a broken power law,

$$\begin{aligned}\frac{dN_{\rm wdm}}{dm_s} &= \frac{dN_{\rm cdm}}{dm_s}\left(1+\frac{m_{\rm hm}}{m_{\rm s}}\right)^{-1.3} \\ &= A_0\left(\frac{m_{\rm s}}{\rm M_\odot}\right)^{-1.9}\left(1+\frac{m_{\rm hm}}{m_{\rm s}}\right)^{-1.3},\end{aligned} \tag{3.13}$$

a functional form which resembles the subhalo mass function of a WDM particle (Schneider et al., 2012; Lovell et al., 2014). We restrict our analysis to this specific parameterization,

[4] While experimenting with Pseudo-Jaffe profiles, we find that our results depend sensitively on how convergence is partitioned between the truncation radius and the central density. Since the central density dominates the flux ratio signal, different choices for the truncation and central density normalization yield significantly different results.

and do not attempt to constrain the exponent -1.3 appearing in Equation 3.13 or the -1.9 slope of the CDM function. More complex scenarios, such as multi-component dark matter in which only a fraction of the dark matter is warm, will require a more careful treatment of the parameterization (see Vegetti et al., 2018). For reference, $A_0 \approx 2 \times 10^8 \mathrm{M}_\odot{}^{-1}$ corresponds to a convergence in substructure of 0.005, or a mass fraction of 1% at the Einstein radius (for more details see Appendix 5.8).

The parameter $m_{\rm hm}$, the half-mode mass, denotes the mass scale at which the WDM power spectrum is damped with respect the CDM case by one-half. Assuming a thermal relic particle of mass m comprises the dark matter, one can establish scaling relation $m_{\rm hm} \propto m^{-3.33}$ (Schneider et al., 2012). We normalize this relation to the $2 \times 10^8 \mathrm{M}_\odot h^{-1} \sim 3.3$keV result of Viel et al. (2013), and translate between the two parameters as

$$m_{\rm hm}\,(m) \;\; = \;\; 10^{10} \left(\frac{m}{1\mathrm{keV}}\right)^{-3.33} \mathrm{M}_\odot h^{-1}. \tag{3.14}$$

In our simulations, we convert to physical masses in the lens plane using $h = 0.7$. With this metric, the 2 keV result from Birrer et al. (2017b) corresponds to $m_{\rm hm} = 10^{8.8}\mathrm{M}_\odot$. As shown in Figure 3.1, we normalize the mass function such that $m_{\rm hm}$ does not affect the amplitude at scales $m_s \gg m_{hm}$, yielding the same numbers of very massive subhalos, on average.

We combine this mass definition with a form for the mass-concentration relation for warm dark matter halos presented by Bose et al. (2016) (see also Macciò et al., 2008; Ludlow et al., 2016)

$$c\,(m_s) = 6 \left(\frac{m_{\rm s}}{10^{12}\mathrm{M}_\odot}\right)^{-0.098} \left(1 + 60\frac{m_{\rm hm}}{m_{\rm s}}\right)^{-0.17}. \tag{3.15}$$

which results in lower central densities at a given mass for warm dark matter models. The relation between concentration and $m_{\rm hm}$ reflects the later collapse epoch of small WDM subhalos, which prevents them from building up their concentrations over time. [5]

Given an M_{200} drawn from the mass function in Equation 3.13, and a concentration from Equation 5.12, we compute the normalization ρ_0 and the scale radius r_s. To obtain the

[5]We do not model the scatter in the mass-concentration relation. In a careful measurement, this feature should be included.

truncation radius, and the lensing properties associated with the mass profile in Equation 5.5, we generate subhalos in a 3-D sphere of radius 250 kpc (see Appendex 5.8 for details regarding the spatial distribution). Given r_{3d}, we compute the truncation radius (Cyr-Racine et al., 2016)

$$r_t = \left(\frac{m_s r_{\rm 3d}^2}{2\Sigma_{\rm crit} R_{\rm Ein}} \right)^{\frac{1}{3}}, \tag{3.16}$$

where $\Sigma_{\rm crit}$ is the critical density and where $R_{\rm Ein} \approx 1''$ is a typical Einstein radius.

3.3.2 Mock Data Sets

We consider three subhalo mass functions: a WDM mass function with $m_{\rm hm}$of $10^8 {\rm M}_\odot$ and normalization of $10^8 {\rm M}_\odot{}^{-1}$, and two CDM mass functions with normalizations of $8 \times 10^7 {\rm M}_\odot{}^{-1}$ and $40 \times 10^7 {\rm M}_\odot{}^{-1}$. The two normalizations in the CDM simulations correspond to projected mass fractions at the Einstein radius of 0.4% and 2%, respectively and bracket a plausible range than spans the theoretical uncertainties associated with the connection to halo mass (Jiang & van den Bosch, 2017) and the tidal destruction of subhalos by the host galaxy (Despali & Vegetti, 2017; Garrison-Kimmel et al., 2017). The low normalization case corresponds to the scenario in which all subhalos inside the halo NFW scale radius are destroyed, while the latter corresponds to no destruction (see the discussion in Appendix 5.8 for details on obtaining these numbers).

When rendering subhalos to create mock data sets, we solve the lens equation and ray trace with every subhalo between 10^6 and $10^{10} {\rm M}_\odot$ included in the computation. We distribute substructures over an SIE+shear macromodel with randomly oriented shear and ellipticity. Ellipticity (shear) is sampled from a Gaussian with mean 0.2 (0.05) and standard deviation 0.05 (0.01), and Einstein radii sampled from a Guassian with mean 1" and variance 0.2". We randomly sample source positions to produce equal numbers of cusp, fold and cross configurations. In Appendix 5.9, we compare the sensitivities of the different image configurations by using each on separately to infer $\mathbf{q}_{\rm sub}$. We take the deflector to lie at a typical redshift $z_d = 0.5$, and put the source at $z_s = 1.5$. The source in both the data and

forward model is parameterized by a circular Gaussian with width of 10 pc, or 1.2 m.a.s. To check to what degree this source size plays a role in our analysis, using source sizes as large as 30 pc we generate distributions of thousands of flux ratios from identical mass functions, and verify the distributions are nearly identical.

Even in the presence of identical subhalo populations, the observed and simulated flux ratios will differ due to the underlying macromodel and measurement errors. Examining mock lens systems with luminous mass components from galaxies in the Virgo and Coma clusters, in Gilman et al. (2017) we found flux ratio anomalies in mock deflectors composed of a NFW halo with a galaxy at its center, with respect to an SIE+shear model. Ray tracing through galaxies formed in the Illustris simulations, the authors of Hsueh et al. (2017) also conclude that the incorrect macromodel can contribute to a measured flux ratio anomaly.

To ensure that modest deviations away from an isothermal-ellipsoid macromodel parameterization does not bias the precision of our inference on $\mathbf{q}_{\rm sub}$, we have simulated mock lenses with with power law slopes drawn from a distribution offset from isothermal and modeled them with SIE profiles in the forward model. Encouragingly, the precision of our forecast is not degraded when subjected to this source of error. See Appendix 3.10 for more details regarding this test. In practice, systematic error associated with the macromodel can be dealt with by sampling additional parameters in the forward model. For instance, macromodel deviations around an isothermal profile can be handled by simultaneously sampling the power law slope and $\mathbf{q}_{\rm sub}$. [6]

Even for samples of morphologically simple deflectors (i.e. no disks or other prominent morphological features), we expect deviations in flux ratios at the percent level from measurement errors, and residual uncertainties in the image fluxes caused by the parameterization of the macromodel. To incorporate these uncertainties in our simulations, we add perturbations to the fluxes in our mock data sets. We model flux anomalies as Gaussian, and perturb

[6]In principle, this approach could be extended in the forward modeling framework to more complex morphological features on a lens by lens basis.

each flux F as

$$\begin{aligned} F &\rightarrow F + \delta F; \\ \delta F &= \mathcal{N}\left(\prime, \delta \times \mathcal{F}\right), \end{aligned} \tag{3.17}$$

examining specific cases of $\delta = 0.02$ $\delta = 0.04$ and $\delta = 0.08$, which correspond to flux errors of $2, 4$ and 8%. Conceptually, one can interpret this operation as erasing information at the δ level, which enables one to track the sensitivity of lensing constraints on small sources of flux perturbations. These perturbations lump together all deviations in image fluxes away from those of an SIE+shear model fit to the data that are not caused by dark substructure, including measurement errors and the baryonic structure of the deflector, and bracket the range of errors one excepts for morphologically simple deflectors. These perturbations are empirically motivated by the flux residuals we encounter in (Gilman et al., 2017) fitting smooth lens models to lenses build from high resolution images of galaxies in the nearby Virgo cluster. We reiterate that extreme morphological features like stellar disks may produce larger systematic perturbations than those we are mimicking with the δ perturbations, but these prominent features are unlikely to be present in massive ellipticals (Auger et al., 2010; Sonnenfeld et al., 2013). [7] Furthermore, deflectors likely to contain disks can be readily identified based on velocity dispersion and stellar mass, in addition to high resolution imaging. Since we add these flux errors independent of perturbations to the image positions and time delays, we assume that the macromodel and measurement-error induced flux ratio anomalies are independent from astrometric and time delay anomalies, a conservative choice as correlations provide additional information that can be used to identify these features in the data.

For reference, current techniques using measurements of narrow-line fluxes achieve precision of roughly $4 - 6\%$ (Nierenberg et al., 2014, 2017). The 8% errors can therefore be interpreted as pessimistic case, with errors induced by the use of a simplified macromodel compounding measurement errors, while 2% simulates an optimistic future precision. The

[7] In fact, studies of massive elliptical lensing galaxies find their mass profiles are well modeled by nearly isothermal power law ellipsoids (Shankar et al., 2017; Gilman et al., 2017).

4% curve serves to illustrate how the bounds evolve between these two extremes. Finally, we add measurement errors of 3 m.a.s. to the mock image positions, typical of astrometric uncertainties with current instruments.

3.3.3 Constraining the subhalo mass function

In our simulations, the subhalo mass function is defined by the free parameters A_0 and $m_{\rm hm}$; our goal is to relate these parameters to the observed data from N simulated lenses, $\mathbf{D}$. To do so, for each A_0 and $m_{\rm hm}$, we render ≈ 2000 substructure realizations $\mathbf{m}_{\rm sub}$ per proposed set of parameters $\mathbf{q}_{\rm sub}$, sampling the prior uniformly in A_0 and uniformly in $\log_{10}(m_{\rm hm})$, yielding in $\approx 10^6$ realizations per lens. In Appendix 4.10, we perform tests to verify convergence with this number of samples. For each realization, we use `lensmodel` to re-optimize the macromodel parameters $\mathbf{q}_{\rm mac(n)}$ that satisfy Equation 3.6, and ray-trace through a grid sampled at 0.4 m.a.s. per pixel to obtain image magnifications and flux ratios with an extended background source modeled as a circular Gaussian with a full-width at half-maximum of 10 pc. When re-optimizing the macromodel, we assume uncertainties of 3 m.a.s. and 2 days on positions and time delays. We use flux constraints weak enough that they do not impact the re-optimization of the macromodel, allowing us to impose different flux perturbations in our mock data sets after running the simulation.

To deal with the flux errors we add to our mock data sets, we add perturbations of the same form as Equation 3.17 to the forward model fluxes. With the perturbed fluxes in hand, we evaluate the statistic in Equation 5.4. After computing the summary statistics, each proposal of $\mathbf{q}_{\rm sub}$ in each of the N lenses has a set metric distances associated with it. At this juncture, we apply a rejection criterion to select the most probable models, including the 1,800 samples with smallest corresponding summary statistics. In principle, because the systems are statistically independent, we could apply this criteria to each one individually and multiply the resulting distributions. Practically, however, multiplying a large number of probability densities together computed on a discretely sampled space is numerically unstable. To handle these numerical issues, we first reduce the dimensionality adding the

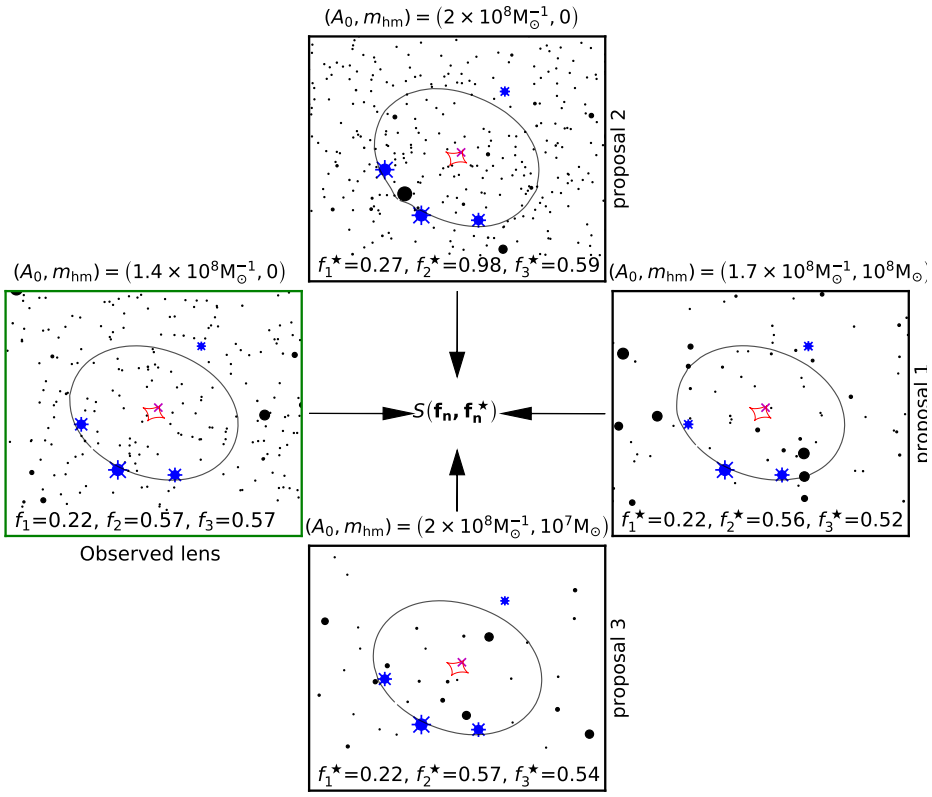

Figure 3.2: This figure presents an illustration of the forward modeling analysis. We use one realization of substructure to create mock data, as shown in the panel outlined in green (left). Lensed images (blue) have sizes corresponding to their flux, while subhalos (black) have sizes corresponding to M_{200}. We compute flux ratios with respect to the middle image in the triplet; f_1 denotes the image in the upper right, while f_2 and f_3 denote images in the left and right of the triplet, respectively. We then discriminate between different parameters describing the subhalo mass function by drawing many substructure realizations from proposal mass functions, three examples of which are shown here (panels outlined in black). For each realization we re-optimize the macromodel and compute the summary statistic $S\left(\mathbf{f_n}, \mathbf{f_n}^\star\right)$ from Equation 5.4, which we use to accept or reject the realization and the parameters describing the mass function it came from. The procedure visualized here is repeatedly applied in the full analysis shown in Figure 3.3.

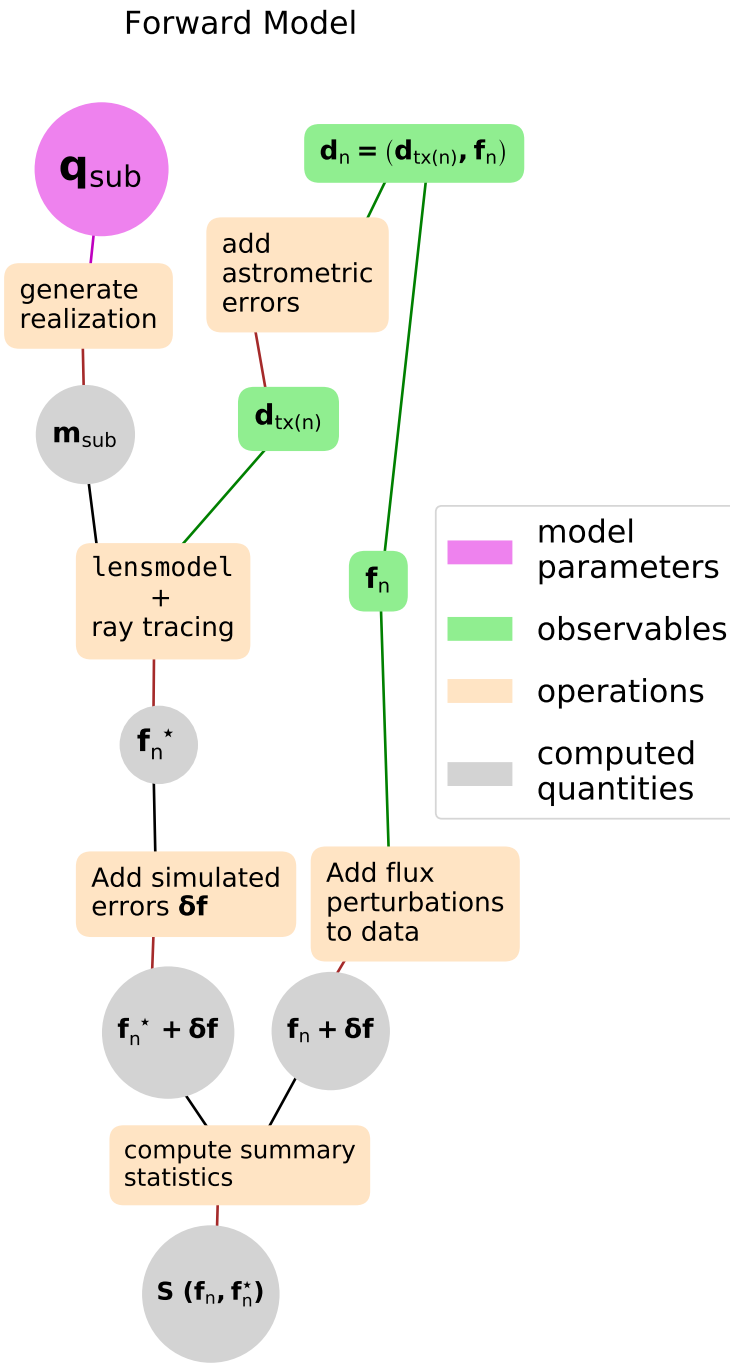

Figure 3.3: A schematic overview of the forward model used to compare flux ratios in the simulated data sets to flux ratios derived from substructure realizations drawn from $\mathbf{q}_{\rm sub}$.

summary statistics for pairs of lenses. In the limit of an infinite number of realizations and infinitely stringent acceptance criteria, this is equivalent to multiplying the likelihoods. To each of the resulting $\frac{N}{2}$ probability densities, we apply a Gaussian Kernel Density Estimator (KDE), and multiply them. We verify that, with these choices, our algorithm satisfies a rudimentary test of convergence (see Appendix 4.10).

In summary: the forward model $\mathbf{f_n}^\star$ contains the information needed to discriminate between different dark matter models through the realizations $\mathbf{m}_{\rm sub}$, drawn from the hyper-parameters $\mathbf{q}_{\rm sub}$. For any parameterization of $\mathbf{q}_{\rm sub}$, Equation 3.7 relates the flux ratios $\mathbf{f_n}'\left(\mathbf{q}^\star_{\rm mac(n)}, \mathbf{m}_{\rm sub}\right)$ to the observed data by evaluating the flux ratios at fixed image positions, enforced by first re-optimizing the macromodel by fitting a smooth lens model to the data $\mathbf{d_{tx(n)}}$, and then computing the flux ratios $\mathbf{f_n}^\star$ for this lens model in the presence of the substructure realization $\mathbf{m}_{\rm sub}$.

We account for flux ratio anomalies caused by measurement errors, and by the imperfect SIE+shear macromodel fit to realistic deflectors, by adding random Gaussian perturbations to the forward model fluxes. We then compute a summary statistic which reflects the degree to which the observed flux ratios match those computed in the forward model. The final posterior probability distribution is composed of the set samples from the prior $\pi\left(\mathbf{q}_{\rm sub}\right)$ whose corresponding realizations $\mathbf{m}_{\rm sub}$ yield summary statistics closest to those computed from the flux ratios in the observed data. To keep computational costs low, in the forward model we only render substructures below masses of $10^{7.5}\mathrm{M}_\odot$ if they lie within 0.5 arcseconds of an image. All higher mass subhalos are included regardless of position.

3.4 Results

Based on the method we outlined in the previous two sections, we are able to quantify the effect of substructure on image flux ratios, and forecast the constraining power of this method. We analyze three scenarios: a CDM mass function with a mass fraction in substructure of 0.4% substructure at the Einstein radius $A_0 = 8 \times 10^7 M_\odot^{-1}$), a CDM mass function with a mass fraction in substructure of 2% at the Einstein radius ($A_0 = 4 \times 10^8 M_\odot^{-1}$), and a WDM

mass function with a mass fraction in substructure of 0.5% substructure at the Einstein radius ($A_0 = 10^8 M_\odot^{-1}$) and a half-mode mass of $10^8 M_\odot$ corresponding to a 3.6 keV thermal relic. [8] We begin in Section 3.4.1 by discussing how variations in the normalization and half-mode mass impact distributions of flux ratios. In 3.4.2, we show the results of simulations in which the mock data sets are free from errors; these simulations serve to determine the constraints achievable with the best possible data. In 3.4.3, we add measurement errors and uncertainties to the fluxes in the mock data sets, account for them in the forward model, and quantify their effect on ones inference of $\mathbf{q}_{\rm sub}$.

3.4.1 Flux ratio signal from structure in the lens plane

In Figure 3.4, we plot the cumulative distribution of flux ratio anomalies for $10,000$ substructure realizations. Several trends emerge which help to understand the signal coming from substructure in image flux ratios.

If we consider modest anomalies whose strength is $< 60\%$ (summed in quadrature), we see that the normalization and half-mode mass both affect the frequency of anomalies. This suggests a degeneracy between the two mass function parameters, which indeed surfaces in a joint inference. The tails of the distributions, shown in the lower panel of Figure 3.4, tell a different story: the curves behave similarly for anomalies whose strength is $> 140\%$ (summed in quadrature), except for the most extreme WDM case (shown in magenta). Together, these results suggest that the most massive substructures, which survive the free-streaming cutoff and are present in the black, blue, grey, and red curves, are responsible for the largest flux ratio anomalies. The frequency of flux ratio anomalies from the model with $m_{\rm hm} = 10^9 {\rm M}_\odot$ suggests that substructures with masses between $10^7 {\rm M}_\odot$ and $10^9 {\rm M}_\odot$ dominate the lensing cross section for the largest flux ratio anomalies, while the cross section for small flux ratio anomalies is dominated by subhalos of mass $\lesssim 10^7 {\rm M}_\odot$.

[8] Due to the demanding computational resources required per lens in the forward modeling procedure, we limit our present analysis to these cases.

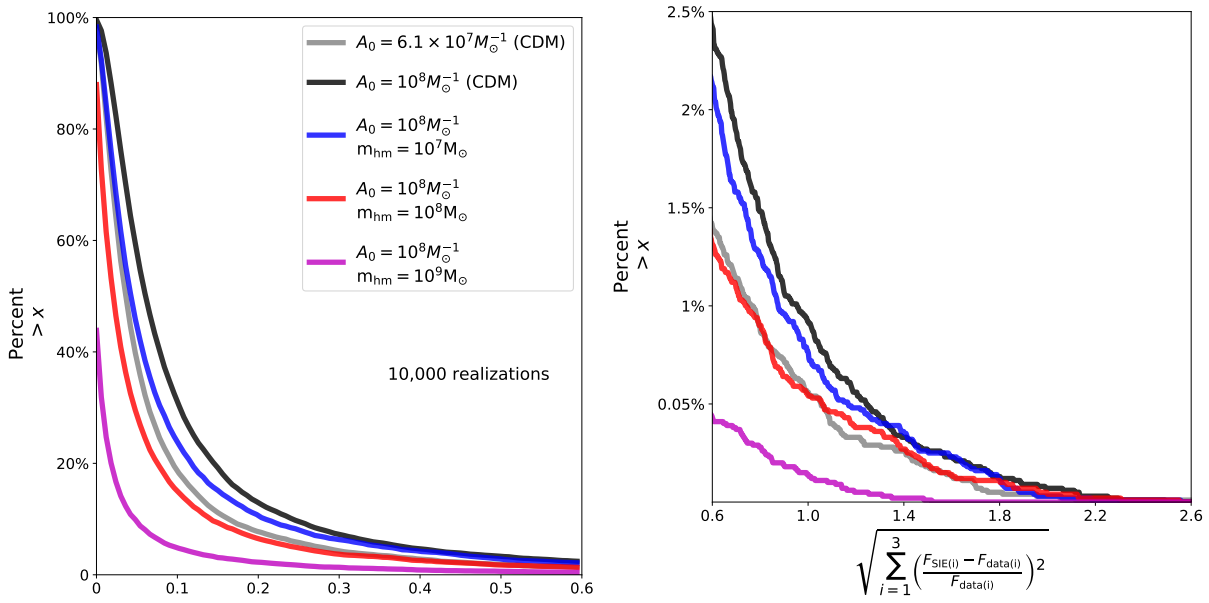

Figure 3.4: Cumulative distributions of flux ratio anomalies summed in quadrature for different subhalo mass functions with varying normalization (grey, black), and identical normalizations but varying $m_{\rm hm}$ (black, blue, red, magenta). The lower panel shows a zoom-in of the long, low probability tail of the distributions. Models with higher normalization (black vs. grey) produce more frequent flux ratio anomalies. Models with high $m_{\rm hm}$ produce less frequent anomalies than the black curve with $m_{\rm hm} = 0$ for flux ratio anomalies < 1.4.

3.4.2 Inference on subhalo mass function with idealized data sets

Before adding simulated errors to the measured flux ratios, as will be present in a real sample of lenses, we first perform the inference on data sets where the flux ratios in the data and the forward model are un-perturbed. Effectively, in these simulations, the only unknowns are the properties of the underlying subhalo mass function, as the macromodel in both the data and the forward model is the same. They demonstrate the best one could hope to do by modeling subhalos only in the lens plane.

First, Figure 3.5 shows the joint posterior distribution for a simulated data set of 30 lenses with $\mathbf{q}_{\rm sub} = (A_0, m_{\rm hm}) = \left(10.4 \times 10^7 {\rm M}_\odot{}^{-1}, 10^8 {\rm M}_\odot\right)$. This idealized calculation provides a useful limit to the sensitivity of the flux ratio anomaly method. The closing of the 2σ contour around $10^{6.5}{\rm M}_\odot$ demonstrates that the signal in flux ratios is sensitive to subhalos of this mass. On their own, these objects produce a very weak lensing signal, but they create collective effects that should make it possible to distinguish between a CDM scenario in which they are abundant and a WDM scenario in which their numbers are depleted. In thie case shown here, the 2σ bounds on the half-mode mass are are $10^7{\rm M}_\odot < m_{\rm hm} < 10^{8.7}{\rm M}_\odot$, which correspond to bounds on the WDM particle mass of 7.3 and 2.3 keV, respectively.

In a second simulation, we use a data set composed of 180 systems with an input mass function with $(A_0, m_{\rm hm}) = \left(8.2 \times 10^7 {\rm M}_\odot{}^{-1}, 0\right)$. Figure 3.6 shows the joint posterior distribution on this data set as the number of lenses is increased. We interpret the 2σ bound of $10^{6.4}{\rm M}_\odot$ as the best one could do with 180 systems. [9] As we show in the next section, these bounds weaken significantly under flux ratio errors of 2%, 4% and 8%, which mimic the signal in flux ratios produced by the smallest subhalos, or by subhalos far from an image.

3.4.3 Inference on subhalo mass function with simulated uncertainties

In this Subsection we demonstrate the effect of flux errors on our inference.

We start by using the the same data drawn from the low subhalo mass normalization

[9] In each of the simulations with mock data sets containing CDM mass functions, we only quote the upper bound because the 2σ lower bound is set by the limits of the prior assigned to $m_{\rm hm}$.

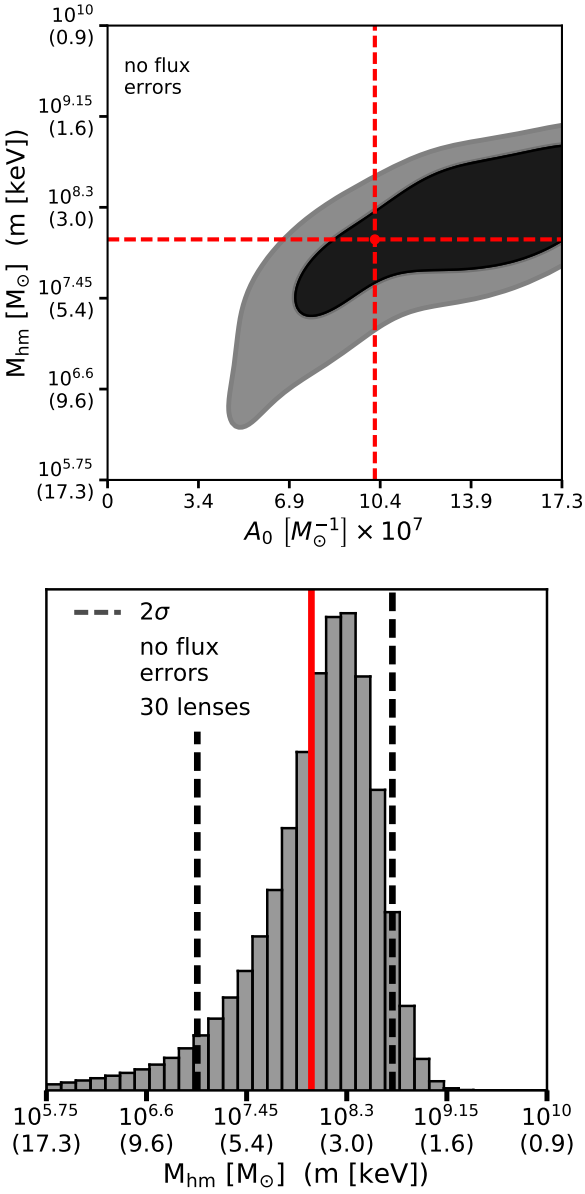

Figure 3.5: (Top) Joint posterior distribution for data with input $(A_0, m_{\rm hm}) = \left(1.04 \times 10^8 {\rm M}_\odot{}^{-1}, 10^8 {\rm M}_\odot\right)$, marked by red lines. The closing of the 1σ contour from the bottom, ruling out CDM mass functions at 2σ, demonstrates the sensitivity of the flux ratios, and this method, for probing substructure on scales $M_{200} \approx 10^{6.5} {\rm M}_\odot$. (Bottom) Marginalized constraints on $m_{\rm hm}$. The 2σ bounds correspond to $10^7 {\rm M}_\odot < m_{\rm hm} < 10^{8.7} {\rm M}_\odot$.

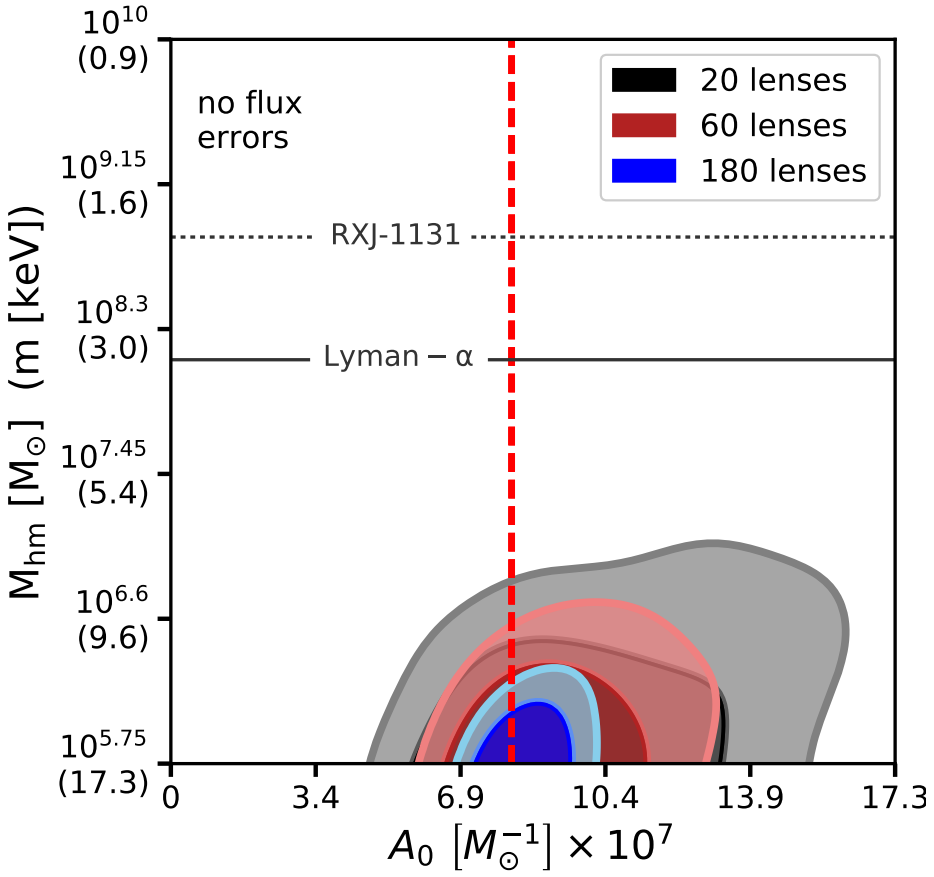

Figure 3.6: Posterior distribution for a CDM-like mass function with input mass function parameters $(A_0, m_{\rm hm}) = \left(8.2 \times 10^7 {\rm M}_\odot{}^{-1}, 0\right)$. In this idealized simulation, we do not add measurement errors or any other perturbations to fluxes in the simulated data sets. Effectively, the only unknown variables are A_0 and $m_{\rm hm}$, which describe the subhalo mass function. In such an idealized case, flux ratios probe scales below $10^{6.5}{\rm M}_\odot$, ruling out WDM models with $m_{\rm hm} > 10^{6.1}{\rm M}_\odot$ and $m_{\rm hm} > 10^{6.4}{\rm M}_\odot$ at 1 and 2σ, respectively. For reference, grey solid and dashed lines show the 2σ bounds on a WDM particle mass from Viel et al. (2013) and Birrer et al. (2017b), respectively.

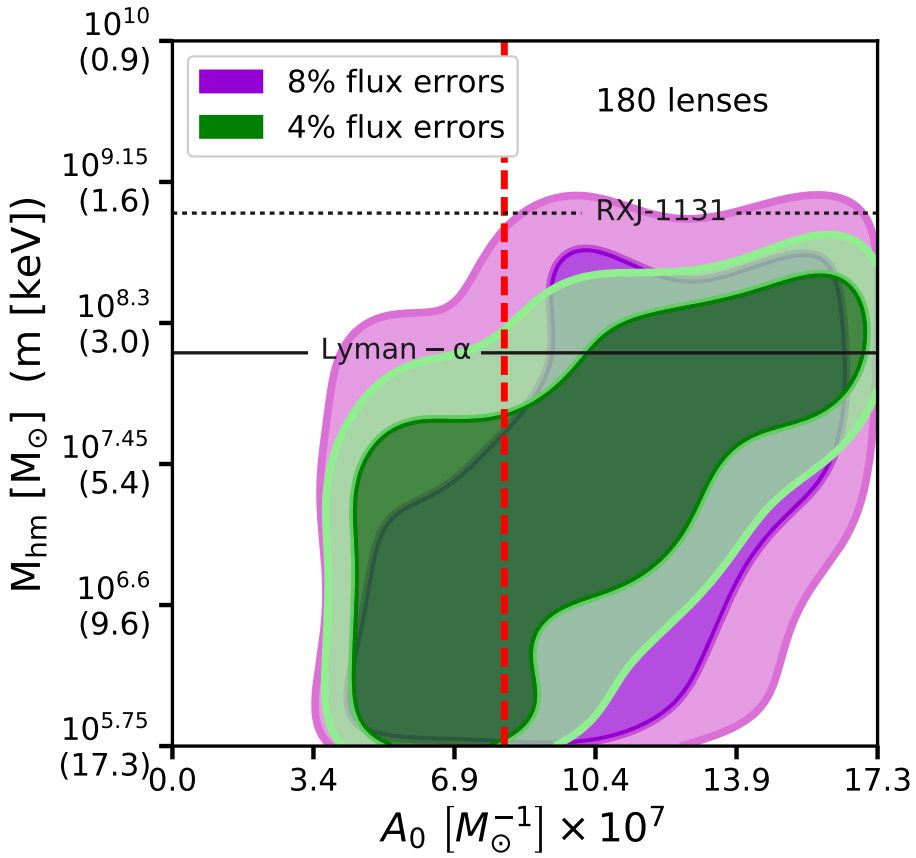

Figure 3.7: Similar to Figure 3.6, but in this case the fluxes in the mock data sets receive 4% and 8% uncorrelated Gaussian errors. On their own, these errors look like substructure, biasing the inference to models with more subhalos. However, we are able to account for this uncertainty by introducing perturbations to the fluxes in the forward model that match those applied to the mock data. Adding noise to the fluxes washes out the signal from the smallest subhalos, and the 2σ constraining power on $m_{\rm hm}$ is diminished by over an order of magnitude. This figure shows the result of a single draw of Gaussian errors in the fluxes. In Figure 3.9, we compute 2σ bounds on $m_{\rm hm}$ averaged over many draws of these errors.

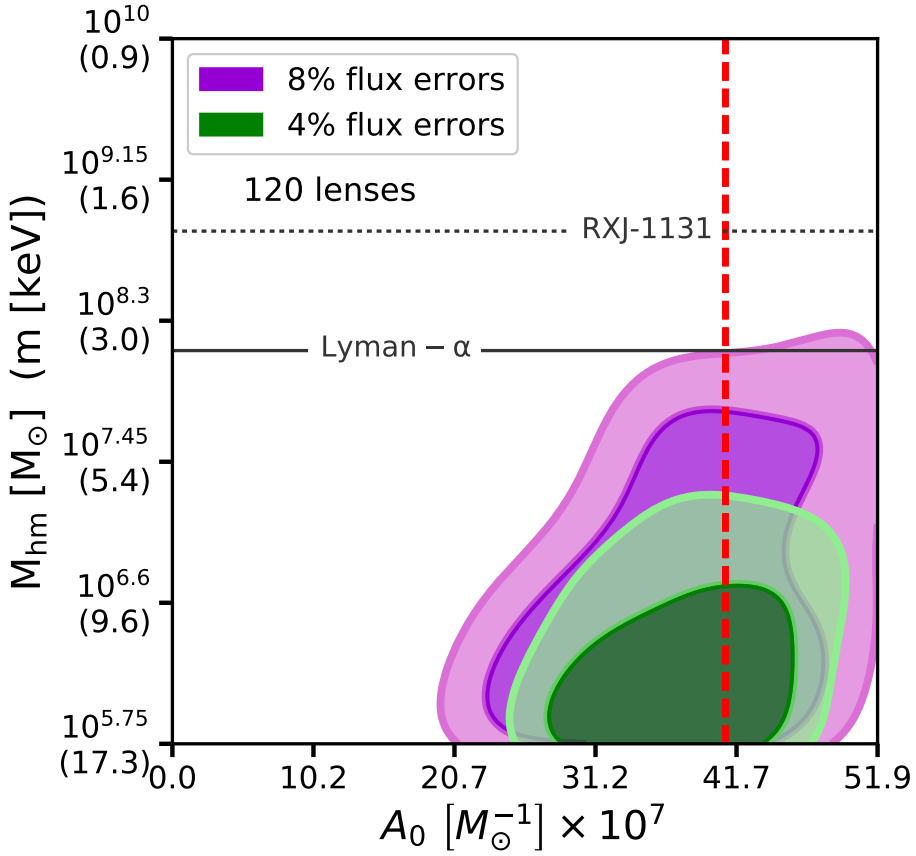

Figure 3.8: Similar to Figure 3.7, but an inference performed on a dataset with a subhalo mass function normalization corresponding to the expected mass fraction in substructure if there is no tidal disruption of subhalos within the host scale radius (see Appendix A). The higher abundance of substructures in this scenario results in a higher probability of observing flux perturbation larger than the effective detection threshold, which is determined by the precision of the image fluxes. Neither the high nor low normalization scenarios includes the expected boost from line of sight structure which we will consider in a future work.

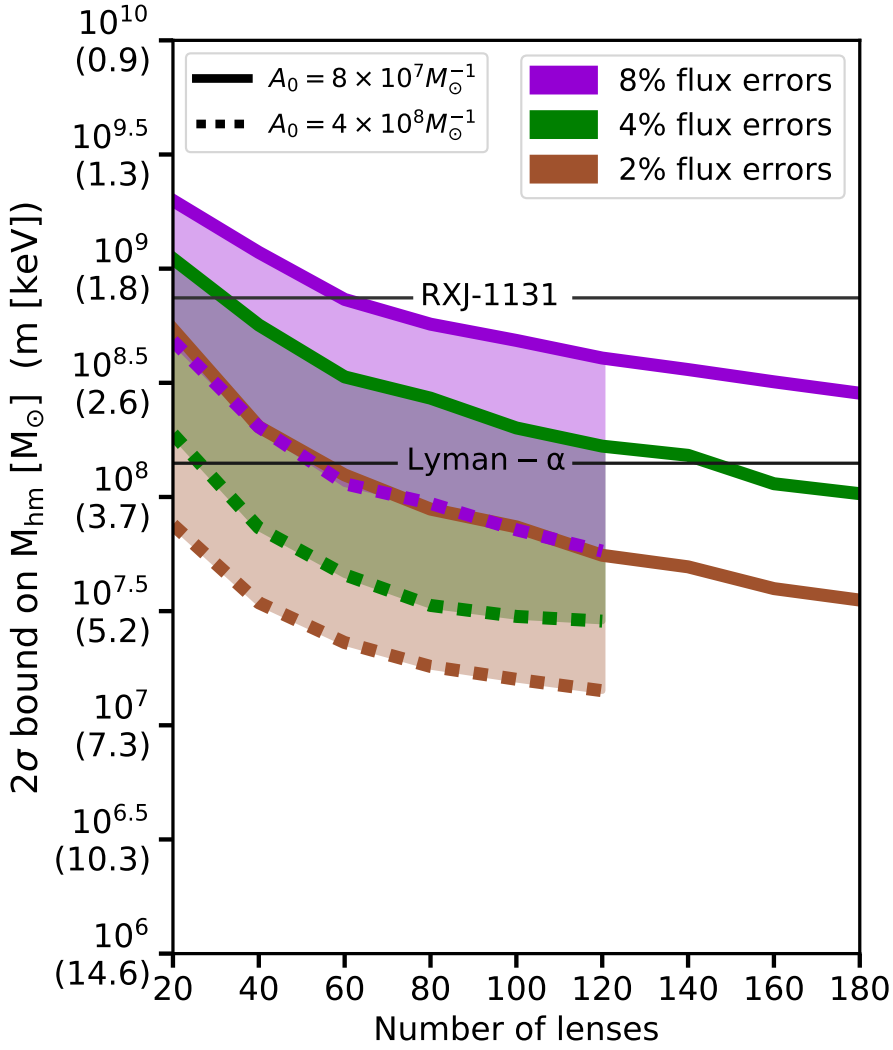

Figure 3.9: Curves show the dependence of the 2σ bounds on $m_{\rm hm}$, and the mass m of a corresponding thermal relic dark matter particle, as a function of the number of lenses. For each plotted point, we randomly sample different combinations of N lenses, each with a random draw of flux errors of 2, 4, and 8 percent. We iterate this procedure 200 times, and compute the mean of the two sigma bounds over the 200 iterations . With 180 lenses, the 2σ bound on $m_{\rm hm}$ is $10^{6.4}{\rm M}_\odot$, $10^{7.5}{\rm M}_\odot$, $10^{8}{\rm M}_\odot$, $10^{8.4}{\rm M}_\odot$ for fluxes with errors controlled at 0%, 2%, 4%, and 8%, respectively. For the higher normalization case, (dotted curves) the signal to noise ratio in the data is higher, allowing tighter constraints for a fixed sample size. For reference, horizontal lines show bounds on the mass of a thermal from the Lyman-α forest (Viel et al., 2013) and an analysis of the strong lens RXJ-1131 (Birrer et al., 2017b).

scenario as is plotted in Figure 3.6. We add flux ratio errors of 4 and 8 percent to the data, and add errors of the same form to the forward model. Random flux errors applied to the data and forward model weaken the constraint on A_0 by a factor of ~ 2 and on $m_{\rm hm}$ over an order of magnitude, as shown in Figure 3.7. This is in part due to the loss of signal from subhalos below masses $10^8 M_\odot$. The 2σ posterior probability contours slope upwards, mirroring the degeneracy seen in Figure 3.5. We also explore the effect of flux errors in the case of a higher normalization (Figure 3.8). A higher normalization results in more frequent high-significance flux perturbations from substructure, which translates to stronger constraints on $m_{\rm hm}$ for a fixed number of lenses, and for fixed flux uncertainties.

In the case of low normalization, the inference with 4% errors rules out models with no substructure and very warm mass functions, but with this specific realization of flux errors applied to the data does not quite surpass the bounds set by Viel et al. (2013) with the Lyman-α forest. If more subhalos survive in the lens plane than the most extreme models for tidal stripping suggest, the constraints on the free streaming length improve to ≈ 5.5 keV, as shown in Figure 3.8. Of course neither of these scenarios incorporate the boost to the overall normalization which will comes from line of sight structure, so the low normalization case may be seen as a lower limit on the signal we expect. In a future paper we will consider the effects of this additional signal.

We note that our method can robustly distinguish between the two normalization scenarios even in the case of 8% flux errors. This is important as it indicates that we are able to measure the difference between WDM and a high normalization and CDM and a low normalization. This can potentially provide useful input to simulations of tidal disruption in massive halos, although the signal from the lens plane will be somewhat diminished by the boost from line-of-sight structure.

Figures 3.7 and 3.8 each show one realization of flux perturbations applied to the data. In Figure 3.9, we average over many realizations of flux perturbations to properly account for the constraints possible from N lenses. Stepping through samples of N lenses in increments of 20, we compute 200 bootstraps for each sample, and repeat this procedure for errors of 2%, 4%, and 8%. The resulting curves in reflect a compromise between more precise flux

measurements and increasing the lens sample size. Both options improve the constraints afforded by image flux ratios on the subhalo mass function, but increasing the flux precision by better handling systematic errors in the lens modeling and by more accurate measurements results yields better marginal gains. The curves do not appear to level off, suggesting that these bounds may improve by including more lenses in the inference, but without simulating more systems it is difficult to make definitive statements for the regime $N > 180$. The normalization is also observed to play a key role, as it effectively boosts the signal to noise ratio in the data and enables measurements more robust to deviations in the fluxes at the percent level.

Conceptually, one can interpret the dependence of the 2σ bounds on different flux errors δF as tracers of the probability distribution $p\left(\delta F|\mathbf{m}_{\rm sub}\right)$. Supposing that $m_{\rm s}*$ defines the subhalo mass scale that dominates this probability density, it follows that adding random noise at the δF level will erode the sensitivity to mass scales $\approx m_{\rm s}*$, so the fast deterioration of constraining power on $m_{\rm hm}$ tracks the loss of sensitivity to mass scales $\approx m_{\rm s}*$. In this context, a higher normalization increases the probability at each scale $m_{\rm s}*$ of observing a flux ratio anomaly δF, counteracting the loss of signal induced by adding flux errors at the δF level.

3.5 Discussion and Conclusions

We have introduced a new method to infer the nature of dark matter from observations of flux ratios in quadruply lensed quasars. The method uses an Approximate Bayesian Computing algorithm to statistically infer the input parameters describing the subhalo mass function without directly computing a likelihood function. We have illustrated the method by performing simulations of strong lenses systems with substructure populations of NFW subhalos rendered in the lens plane, in the case of cold and warm dark matter, and for two different normalizations of the subhalo mass function.

While a real sample of lenses will be diverse in both redshift distribution and host halo mass, the lens and source redshift will primarily impact the contribution from the line of

sight, while the connection between the normalization and host halo mass, and the effect on lensing observables, is subject to considerable theoretical uncertainty (Despali & Vegetti, 2017; Garrison-Kimmel et al., 2017). We handle this uncertainty by considering two limiting cases of the normalization, corresponding to scenarios with and without complete subhalo disruption within the scale radius of the parent halo. As the normalization of each lens effectively weights the information content available, the limiting cases we analyze bound the constraints from a sample of lenses with diverse halo masses.

Our main results can be summarized as follows:

- In an idealized scenario, where the macromodel is known to high precision and other sources of flux ratio errors are mitigated, the only source of flux ratio perturbation comes from dark substructure and flux ratios probe the mass function at scales below $10^7 \mathrm{M}_\odot$. With flux uncertainty at the level 2%, 4% and 8%, the bounds on the half-mode mass [$M_\odot$] (thermal relic mass [keV]) are $m_{\rm hm} = 10^{7.5}$ (5.0), $m_{\rm hm} = 10^{8}$ (3.6) and $m_{\rm hm} = 10^{8.4}$ (2.7) with 180 systems. For the higher normalization case, the improvement in the signal to noise ratio in the data yields constraints of $m_{\rm hm} = 10^{7.2}$ (6.6), $m_{\rm hm} = 10^{7.5}$ (5.3) and $m_{\rm hm} = 10^{7.8}$ (4.3) with just 120 lenses. In a WDM scenario, we find with no uncertainty in flux ratios that we can measure the position of the free-streaming cutoff in the subhalo mass function with just 30 lenses, constraining it to between $10^7 M_\odot$ and $10^{8.7} M_\odot$ at 2σ. With less control over systematics and degraded measurement precision, more than 30 lenses would be required to achieve these constraints, but our simulations suggest that this method can, in principle, measure the warmth of dark matter should CDM be the incorrect model. Provided one controls for systematic errors in flux ratios associated with incorrect macromodels, these constraints will likely improve after adding the contribution from line of sight structure, which contributes substantial additional signal.

- The 2σ bound on $m_{\rm hm}$ improves rapidly with increasing flux uncertainties, and falls slowly after $N \approx 100$ lenses. This reflects the sensitivity of flux ratios to low mass subhalos, which impart deviations at the level of a few percent. In terms of overall

strategy for the study of strong lens systems, this establishes the necessity of measuring fluxes precisely and controlling for systematic errors arising from the parameterization of the macromodel. The simple SIE+shear parameterization implemented in this work may not be sufficient for systems in which additional mass components are present in the main deflector, such as stellar disks. In practice, identifying morphological complexity in the main deflector can be achieved by deep imaging of the lensing galaxy to identify luminous mass components, and preferentially analyzing slow-rotators with high central velocity dispersions.

- The frequency and magnitude of flux ratio anomalies differentiates between different dark matter models. In this work, we have explored the effect of a varying normalization and half-mode mass. While the half-mode mass scale carries information regarding the nature of dark matter, baryonic effects can impact the normalization, which plays a crucial role as it effectively determines the signal to noise ratio in the data. This translates into better constraints on the shape of the mass function, making it easier to distinguish WDM from CDM. There is some degeneracy between the normalization of the mass function and the half-mode mass, but flux ratios are sensitive enough to break this degeneracy and probe mass scales below $10^8 M_\odot$, where CDM and WDM subhalo abundance differs significantly.

Recent analysis has shown the contribution from line-of-sight subhalos is substantial (Despali et al., 2018b). Since inclusion of the line-of-sight structure will likely improve our projected constraints, we interpret our results as understated limits on the power of substructure lensing. We leave the extension of our method to include line-of-sight structure to future work.

Comparing the posterior probability distributions in Figures 3.7 and 3.8, although there is some degeneracy between the normalization and the half-mode mass these parameters can be constrained simultaneously with ≈ 100 lenses. [10] The normalization has important

[10] In fact, the differences between the low normalization and high normalization scenarios become apparent with fewer than 100 lenses, but the extent to which one can differentiate the two depends on the degree to which one controls for systematic errors in flux ratios, and the difference between the two normalizations.

implications for dark matter and baryonic physics through its connection to total halo mass and tidal stripping of subhalos, respectively, and thus potentially will provide an important constraint for theoretical models.

Our method hinges on accurate measurement of image flux ratios and controlling for systematic errors in their modeling. To quantify the impact of small uncertainties in these observables, we simulate observations with different errors applied to the image fluxes to erase information at the 2%, 4%, and 8% level, and find the projected constraints are extremely sensitive to loss of information at this level. Case in point: the difference between perfect models and perfect measurements, and an observational scenario with 2% uncertainties in image fluxes is an order of magnitude in $m_{\rm hm}$. To achieve the required level of measurement precision, we will need flux ratios computed from the narrow-line emission of the background quasar (Nierenberg et al., 2014, 2017), which yield measurements accurate to $4-6\%$ in flux, and are resilient to microlensing. The presence of systematic errors in the modeling can also be mitigated by restricting analysis to deflectors with high velocity dispersions and no complicated morphological features like stellar disks, assuring that dark substructure dominates as the source of flux ratio anomaly computed with respect to simple SIE macromodels. Alternatively, deep imaging of the deflector and its stellar mass distribution may enable the construction of lens models that map the luminous structure of the deflector in detail, as was done in Hsueh et al. (2016, 2017). For these more complicated systems, additional observable information in the form of deep imaging is required to constrain the mass distribution of the deflector.

Finally, we note that our method accommodates any arbitrary dark matter model, provided it specifies the form of the subhalo mass function and the density profiles of individual substructures. Possible extensions of our method can explore subhalo populations from mixed or self-interacting dark matter (Rocha et al., 2013), and models in which a fraction of the dark matter is composed of primordial black holes (Cotner & Kusenko, 2017).

Acknowledgments

We thank Adriano Agnello, Andrew Benson, Michael Boylan-Kolchin, Francis-Yan Cyr-Racine, Chris Fassnacht, Stacy Kim, Alex Kusenko, Phil Marshall, Leonidas Moustakas, Annika Peter, Simona Vegetti, and Dandan Xu for helpful suggestions and interesting discussions throughout the course of this project. We also thank the anonymous referee for feedback that improved the quality of the paper.

DG, TT, and SB acknowledge support by the US National Science Foundation through grant AST-1714953. CRK acknowledges support by the NSF through grant AST-1716585. This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by the UCLA Institute for Digital Research and Education's Research Technology Group.

3.6 Appendix A: More on generating substructure realizations

3.6.1 Spatial distribution and truncation

As discussed in Section 3.3, we tidally truncate the subhalos according to their 3-d position in the halo $r_{3\mathrm{d}}$ according to

$$r_t = \left(\frac{m_s r_{3\mathrm{d}}^2}{2\Sigma_{\mathrm{crit}} R_{\mathrm{Ein}}} \right)^{\frac{1}{3}} \tag{3.18}$$

(see Cyr-Racine et al., 2016), where m_s refers to M_{200} at redshift $z_d = 0.5$. Although strong lensing quantities typically live in projection in the lens plane (omitting the line of sight), the third spatial dimension enters through the truncation, via the 3-d position $r_{3d} = \sqrt{z^2 + r_{2d}^2}$.

To include this effect in our simulations, we begin by noting that in projection, in the inner portions of a galactic halo where strong lensing takes place, subhalos appear distributed uniformly in two dimensions (Xu et al., 2015). We therefore assign each subhalo a projected position r_{2d} with a spatially uniform probability density out to $R_{\mathrm{max}} = 18.6$kpc, or 3 arcseconds at the lens redshift.

To obtain the 3-dimensional z coordinate for a subhalo, we start with a two dimensional

distribution that is uniform (to a very good approximation) with in 18.6 kpc

$$p\left(r_{2d}|r_c\right) \propto \left(1+\frac{r_{2d}^2}{r_c^2}\right)^{-1} \tag{3.19}$$

with $r_c = 75$kpc. We then de-project this 2-d density into the third dimension to obtain a density for the z coordinate, out to a maximum 3-d radius $R = 250$kpc. The corresponding distribution for the z coordinate, given a 2-d position, becomes

$$p\left(z|r_{2d}, r_c\right) \propto \left(1+\frac{z^2+r_{2d}^2}{r_c^2}\right)^{-1.5}. \tag{3.20}$$

We note that this has the same asymptotic form $\propto r^{-3}$ as an NFW profile for large z. The z coordinate affects the lensing only indirectly through the truncation radius. For an NFW profile, this does not significantly impact the image magnification, as this observable is principally determined by the central density which is unchanged. We verify that our truncation scheme consistently yields $r_t > r_s$. Typical values for $\frac{r_t}{r_s}$ range between 5-30, depending on the concentration of the subhalo.

3.6.2 Mass Function

Focusing first on the form of the mass function $\frac{dN}{dm}$, numerical simulations of cold dark matter halos (Springel et al., 2008; Gao et al., 2011; Fiacconi et al., 2016) suggest a scale free mass function $\frac{dN}{dm_s} \propto m_s^{-1.9}$ for $10^6 \leq \frac{m_s}{M_\odot} \leq 10^{10}$. We acknowledge that tidal disruption may alter this prediction significantly in the inner portions of galactic halos, but we do not address this concern here, as our principle aim is to demonstrate the method rather than focus on the most realist mass function.

When normalizing the subhalo mass function, we wish to compare CDM realizations with WDM realizations with the same amplitude at mass scales far above the half-mode mass $m_{\rm hm}$. To accomplish this, we start with a scale free CDM mass function

$$\frac{dN_{\rm cdm}}{dm_s} = A_0 \left(\frac{m_s}{\rm M_\odot}\right)^{-\alpha} \tag{3.21}$$

taking $\alpha = 1.9$. In the regions of dark matter halos, the spatial distribution of substructure is approximately uniform in projection (Xu et al., 2015). Uniformly distributing subhalos

in a plane with radius $R_{\rm max}$, we relate the substructure convergence at the Einstein radius, $\kappa_{\rm sub}$, to the mass in substructure between m_L and $m_{\rm s} \leq m_H$

$$\Sigma_{\rm crit}\kappa_{\rm sub}\pi R^2_{\rm max} = \int_{m_L}^{m_H} m_s \frac{dN_{\rm cdm}}{dm_s} dm_s. \tag{3.22}$$

This yields the normalization A_0 in terms of $\kappa_{\rm sub}$, and the mean number of subhalos

$$\begin{aligned} A_0 &= \frac{(2-\alpha)\,\pi R^2_{\rm max}\Sigma_{\rm crit}\ \kappa_{\rm sub}}{M^{\alpha}_{\rm M_\odot}\left(m_H^{2-\alpha} - m_L^{2-\alpha}\right)} \\ \langle N_s \rangle &= \frac{A_0}{1-\alpha}\left(m_H^{1-\alpha} - m_L^{1-\alpha}\right) {\rm M_\odot}^{\alpha}. \end{aligned} \tag{3.23}$$

We then draw N_s subhalos from a Poisson distribution with average value $\langle N_s \rangle$.

3.6.3 Normalization

The normalization of the subhalo mass function depends both on the accretion history and evolution of subhalos in the lens halo, and the effects of baryonic physics in the central regions of the halo. The former effect, the accretion history of dark matter halos as a function of halo mass, has been well studied, and here we adopt the result of (Han et al., 2016) and assume a total surviving halo mass fraction of $f_{sub,halo} \sim 6\%$. There is scatter in the predictions depending on halo accretion history and redshift which can raise this value by as much as a factor of 2 (Fiacconi et al., 2016; Jiang & van den Bosch, 2017), however we conservatively adopt the lower normalization for this estimate, and lower the overall subhalo mass normalization for both the extreme and no tidal disruption cases by 30%, bringing the assumed total halo mass fraction in substructure to $\sim 4\%$.

The effect of baryonic physics is more uncertain. Recent state of the art hydrodynamic simulations indicate that in a Milky Way mass host halo, the central disk may destroy all subhalos within the central 20 kpc of the host, in addition to reducing the total number of subhalos by $\sim 30\%$ compared to a dark matter only run (Garrison-Kimmel et al., 2017). In these simulations, the destruction appears independent of subhalo mass.

In order to bracket the range of possibilities and to demonstrate how tidal disruption in the central region would affect our inference, we consider two scenarios. In the first, we

assume that dark matter subhalos follow the NFW density profile of the host (Han et al., 2016), which is seen in dark matter only simulations. In the second case we mimic the effects of central tidal disruption by assuming that the destruction radius scales with the host scale radius. Dedicated simulations of lens mass halos will be necessary to calibrate this effect in more detail.

For a lower bound on the normalization, we assume that all tidal disruption destroys all subhalos within the scale radius of the host (for a typical lens halo of mass $\sim 10^{13} M_{\odot}$ this radius is $\sim$ 150 kpc). For an upper bound, we assume that the subhalo number density follows the density profile of the host. In each case, we compute the projected mass density in substructure along the longitudinal virial radius of the host inside a projected cylinder of 18 kpc ($\sim$ 3 arcseconds at $z = 0.5$), and obtain $(\kappa_{\rm sub}) = 0.002$ (0.01) in the case of extreme (minimal) tidal stripping. These values correspond to values of $A_0 = 8$ (40) $\times 10^7 {\rm M}_{\odot}{}^{-1}$.

3.6.4 Extension to WDM

The half mode mass $m_{\rm hm}$ corresponds to a characteristic length scale at which the linear matter power spectrum with pure WDM is damped with respect to that of CDM by one-half. For details, see e.g. Viel et al. (2005); Schneider et al. (2012). We wish to compare a range of WDM subhalo populations with varying $m_{\rm hm}$ according to the mass function

$$\frac{dN_{\rm wdm}}{dm_s} = \frac{dN_{\rm cdm}}{dm_s} \left(1 + \frac{m_{\rm hm}}{m_s}\right)^{-1.3} \tag{3.24}$$

while preserving the amplitude of the mass function for masses high above $m_{\rm hm}$ to isolate the effect of $m_{\rm hm}$ from the normalization A_0. To do this, we first generate subhalos according to Equation 3.22. This results in scale free mass function, which we deplete by removing subhalos probabilistically with probability

$$P \propto \left(1 + \frac{m_{\rm hm}}{m_{\rm s}}\right)^{-1.3}. \tag{3.25}$$

This yields substructure populations obeying the curves plotted in Figure 3.1.

3.7 Appendix B: Sensitivity of cusp, fold, and cross configurations

The response of an image magnification to small scale structure is heightened if it lies close to a critical curve. Similarly, images close to one another may be affected by the same substructures, introducing a correlation between flux anomalies in different images.

Cusp (fold) configurations are characterized by three (two) images straddling the critical curve, and by the three (two) images close in proximity to each other. One therefore expects a ranking in sensitivity to substructure of cusp, fold, cross, in descending order. In Figure 3.10, we show that this is indeed the case. Interestingly, the degeneracy between a warm mass function with a high normalization, and a cold mass function with low normalization is reduced in cusp and fold configurations compared to crosses.

To identify image configurations, we adopt the following classification scheme based on the Einstein radius $R_{\rm Ein}$ and the image separations. If the smallest image separation is greater than $0.7R_{\rm Ein}$, the lens is immediately classified as a cross. If the second largest separation is $< 1.2R_{\rm Ein}$, we classify it as a cusp, and otherwise it is a fold.

3.8 Appendix C: Use of other summary statistics

The summary statistic in Equation 5.4 is closely the likelihood, or a χ^2 value. It penalizes models which do not reproduce the anomalies observed in the data in the correct images. There is, however, no 'correct' choice of summary statistic, only ones that perform better than others.

We experimented with using other statistics, including $R_{\rm cusp}$ and $R_{\rm fold}$ (see Keeton et al., 2003, 2005). An advantageous feature of these parameters is that they can be computed directly from the observed image fluxes, and do not rely on a lens model to identify anomalies. After experimenting with other summary statistics, however, we find the summary statistic in Equation 5.4 yields the strongest constraints, because it does not discard information by adding and subtracting fluxes from different images.

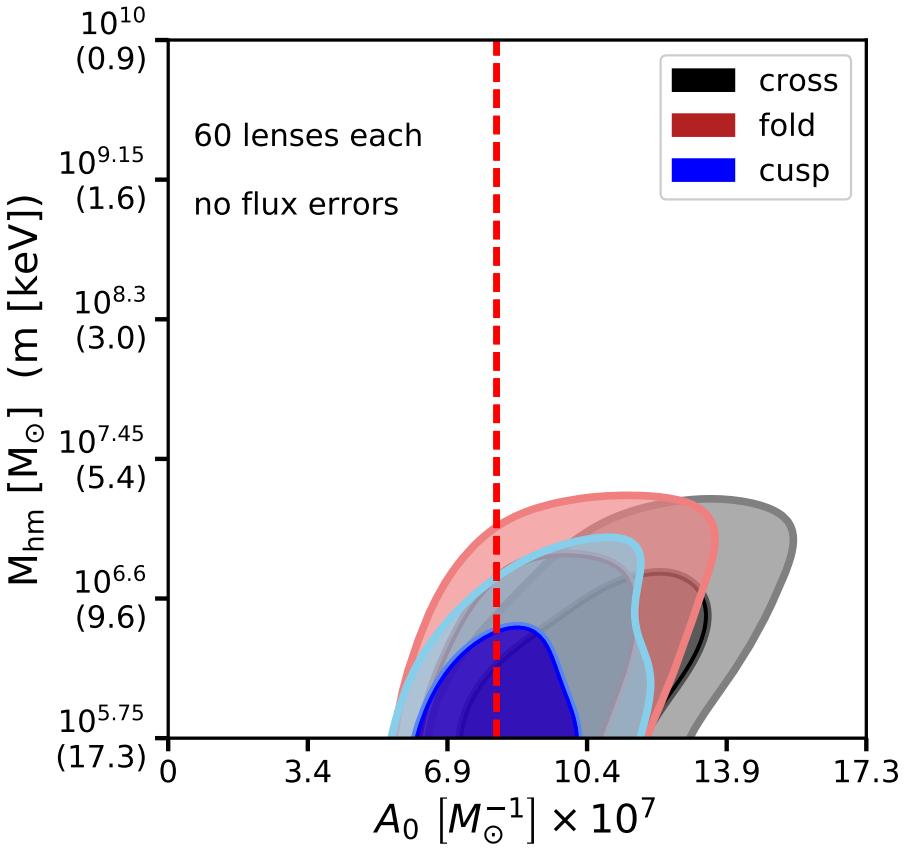

Figure 3.10: Response of cusp, fold and cross image configurations to small scale structure. For a fixed number of lenses, cusps yield the strongest constraints, followed by folds and crosses.

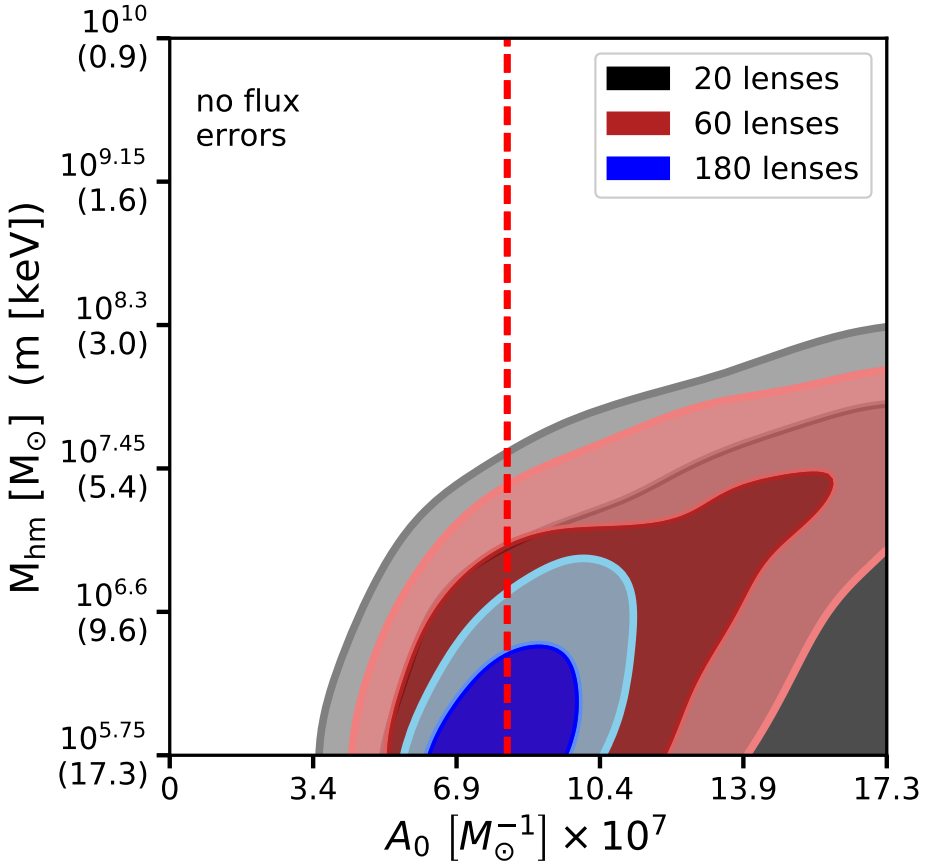

Figure 3.11: Inference made using the statistic in defined in Equation 3.26, rather than Equation 5.4.

As an example, consider the summary statistic

$$S\left(\mathbf{f_n}, \mathbf{f_n}^\star\right) = \left| \sum_{i=1}^{3} \left(f_{n(i)} - f^\star_{n(i)} \right) \right|. \tag{3.26}$$

Using this equation instead of Equation 5.4, we perform an inference on the same simulation as in Figure 3.6. The results are shown in Figure 3.11. Equation 5.4 performs better.

3.9 Appendix D: Convergence of ABC simulations and posteriors

To assess convergence of the ABC algorithm, we compare two inferences made on the same data set in which one has half the number of realizations as the other. To produce the black distribution, we retain the draws from $\mathbf{q}_{\rm sub}$ associated with the lowest 1,800 summary

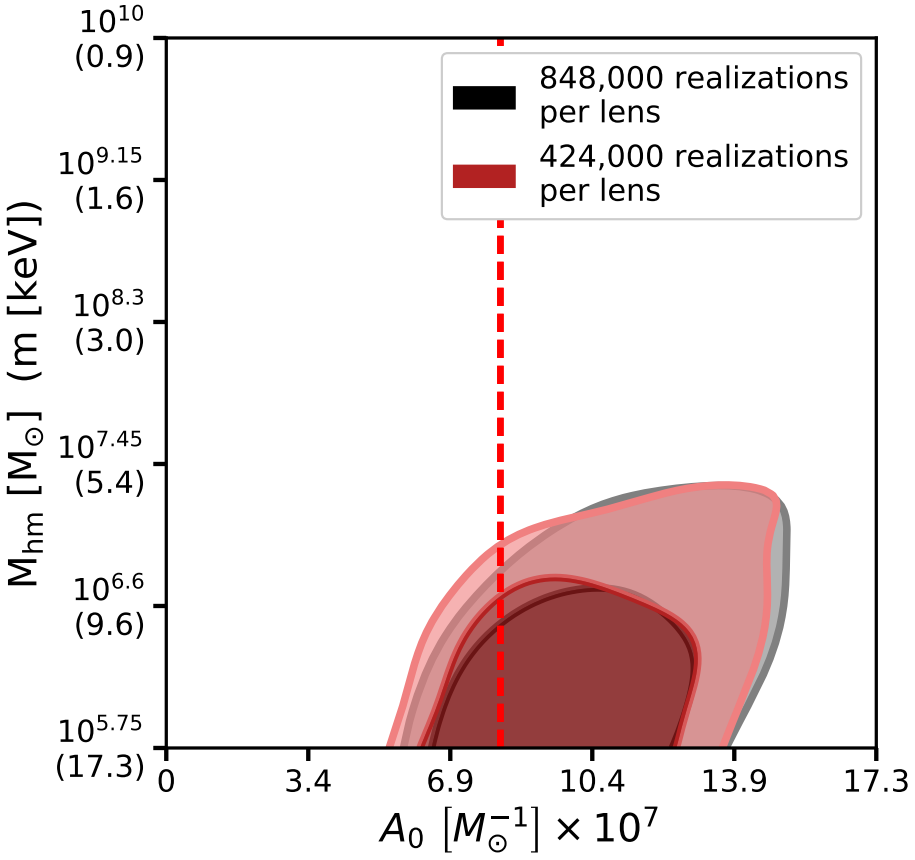

Figure 3.12: A convergence test in which we discard half of the realizations before applying the acceptance criterion. Each distribution is composed of the samples from $\mathbf{q}_{\rm sub}$ associated with the lowest 1,800 summary statistics.

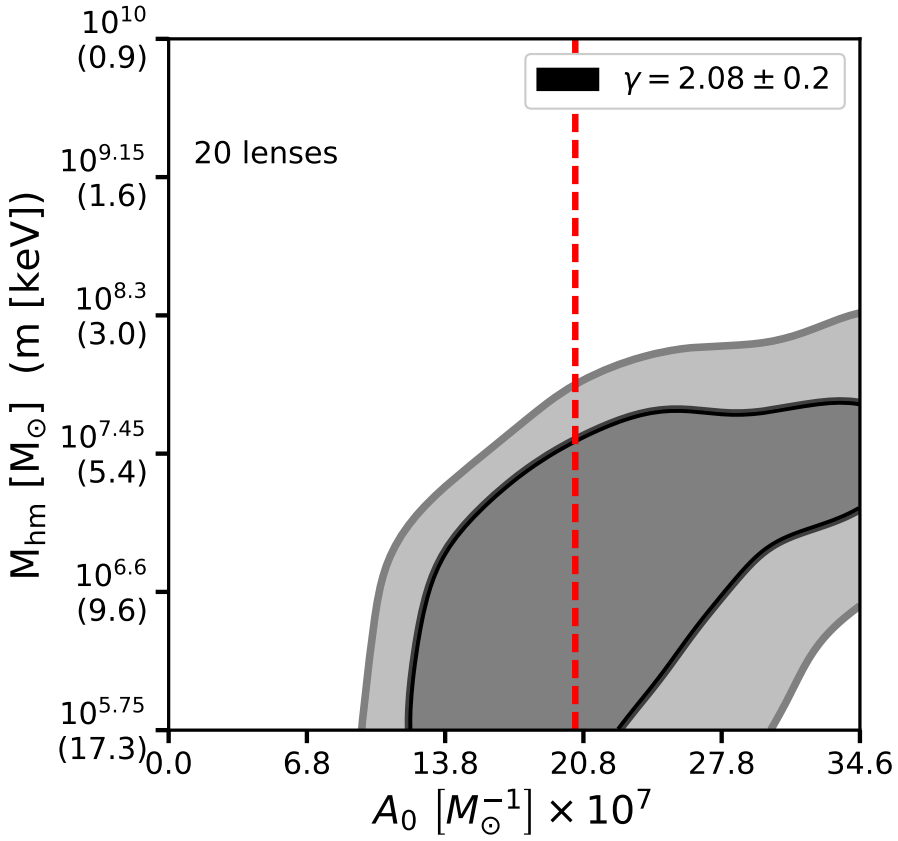

Figure 3.13: We test the effect of an incorrectly parameterizing the macromodel by generating mock data with non-isothermal ellipsoids, and modeling the data with an SIE in the forward model. The power law slope γ is sampled from a Guassian with mean 2.08 and variance 0.2.

statistics. We then discard half of the realizations, and reject all but the lowest 1,800 summary statistics from the depleted simulation. Under-sampling by a factor of two, we recover the same bounds on A_0 and $m_{\rm hm}$ to a high degree of precision, indicating the inference made with ≈ 2000 realizations per proposal $\mathbf{q}_{\rm sub}$ has converged.

3.10 Appendix E: Fitting SIE macromodels to datasets built with non-isothermal power laws

We test the effect of incorrectly parameterizing the macromodel by created datasets with non-isothermal power law ellipsoids, and fitting them with SIEs in the forward model. When generating mock lenses, we sample power law slopes from a Guassian distribution offset from isothermal at $\gamma = 2.08 \pm 0.2$, consistent with the findings of Shankar et al. (2017), who find the power law slopes for massive ellipticals are consistently steeper than isothermal. The results of this exercise are shown in Figure 3.13. The effect of varying the power law slope in the data is seen to not significantly degrade the inference on A_0 and $m_{\rm hm}$, and therefore does not affect the precision of our forecast statements.

By completely neglecting the presence of this systematic, we exaggerate its potential bias in the inference. In practice, this systematic should be properly dealt with by sampling different power laws slopes in the forward model. A prior γ may be constructed on a lens by lens basis based on measurements of the central velocity dispersion, if available.

CHAPTER 4

Probing dark matter structure down to 10^7 solar masses: flux ratio statistics in gravitational lenses with line of sight halos

This chapter was published as Gilman, D., et al. Probing dark matter structure down to 10^7 *solar masses: flux ratio statistics in gravitational lenses with line of sight halos. MNRAS 487, 5721-5738 (2019), and is printed here with minor formatting adjustments.*

4.1 Introduction

Theories of particle dark matter predict that the enigmatic particle(s) collect in gravitationally bound halos. The mass function and density profiles of these objects depend on the particle nature of dark matter itself. For example, theories with cold dark matter predict an abundance of low mass halos, and cuspy r^{-1} central density profiles (Moore et al., 1999; Springel et al., 2008; Fiacconi et al., 2016). In warm dark matter (WDM) models, diffusion of dark matter particles in the early universe wipes out density fluctuations below a characteristic scale that depends on the production mechanism of the WDM particle candidate (Kusenko, 2009; Shoemaker & Kusenko, 2009; Abazajian, 2017). Suppression of small-scale power in WDM models results in a turnover in the halo mass function and a dearth of small-scale structure at later times (Bode et al., 2001; Schneider et al., 2012; Lovell et al., 2014). In self-interacting dark matter (SIDM) theories, scattering between dark matter particles produces cored density profiles in individual halos (Spergel & Steinhardt, 2000; Rocha et al., 2013; Vogelsberger et al., 2016; Kamada et al., 2017; Tulin & Yu, 2018). Finally, in 'fuzzy'

dark matter scenarios the kpc-scale de Broglie wavelength of ultra-light dark matter particles results in quantum mechanical phenomena on galactic scales, which produces large soliton cores (Hui et al., 2017; Robles et al., 2019). To date, the strongest constraints on WDM come from the Lyman-α forest (Viel et al., 2013; Iršič et al., 2017), while cosmological probes on large scales (Cyr-Racine et al., 2014; Bringmann et al., 2017) and in galaxy clusters (Kim et al., 2017; Andrade et al., 2019) constrain the interaction cross section in SIDM models.

Two challenges to CDM in particular spur interest in alternative theories. First, natural CDM particle candidates have not yet been detected, despite decades of experimental searches (Aprile et al., 2018). Second, the suppression of small scale structure in WDM, and cored density profiles associated with SIDM, possibly alleviate tension between observations and the predictions on sub-galactic scales, dubbed the 'Small-Scale Crisis' of CDM (see Bullock & Boylan-Kolchin, 2017, and references therein).

Traditional astrophysical challenges to the CDM model, however, are predicated on assumptions related to baryonic physics. This has the undesirable consequence of propagating uncertainties from sub-galactic astrophysics onto inferences of dark matter properties, and results in covariance between baryonic astrophysics and dark matter physics. Supernova and stellar feedback inside halos, for instance, and the tidal destruction of subhalos by their host galaxy, mimic the observable signatures of SIDM and WDM models, respectively (Tollet et al., 2016; Read et al., 2018; Despali & Vegetti, 2017; Garrison-Kimmel et al., 2017; Kim et al., 2018; Despali et al., 2018a). Moreover, in some cases, the uncertainties related to baryonic astrophysical processes can be larger than the differences between CDM, WDM, and SIDM (e.g. Nierenberg et al., 2016). To isolate dark matter physics from sub-galactic astrophysics, and to differentiate between CDM, WDM, and SIDM, one must look to masses below $10^8 M_\odot$, where subhalos are expected to be devoid of stars and completely dark in the case of CDM, or absent in the case of WDM.

Gravitational lensing offers a direct probe of this elusive, low-mass regime. It circumvents the complications associated with using luminous matter to trace the dark matter by enabling the direct measurement of the distribution of matter across cosmological distance, and is sensitive to mass scales where astrophysical effects are thought to be too weak to significantly

alter the structure of halos. It also compliments other probes of dark matter, such as the Lyman-α forrest, since lensing depends on different systematics and measures the halo mass function directly.

Ultimately, analysis of strong lenses hinges on separating mass distributions that vary on large scales (the lensing galaxy and its parent dark matter halo) from small scale structure in the main lens plane and along the line of sight. In strong lens systems with luminous arcs, the analysis consists of iteratively fitting a smooth model to the flux in pixels of an image while simultaneously reconstructing the background source. This process can reveal the presence of small scale structure in the arcs (see e.g. Vegetti et al., 2014; Hezaveh et al., 2016b; Vegetti et al., 2018; Ritondale et al., 2018). Birrer et al. (2017b) performed this analysis, and placed constraints on the free streaming length of dark matter. Recently, several authors have proposed using the surface brightness residuals from lens models fit to luminous arcs and to infer the power spectrum of dark matter in strong lenses (Hezaveh et al., 2016a; Díaz Rivero et al., 2018; Cyr-Racine et al., 2019), and Bayer et al. (2018) applied this method to a strong lens system.

In addition to extended arcs, some strong gravitational lenses produce four images (quads) of an unresolved background source, such as a quasar. The magnification ratios (flux ratios) between multiple images of unresolved sources have long been recognized as powerful probes of small scale structure near lensed images (Mao & Schneider, 1998a; Metcalf & Madau, 2001), and have been used to test the predictions of CDM and to detect structure near individual objects (Dalal & Kochanek, 2002; Amara et al., 2006; Xu et al., 2012; Fadely & Keeton, 2012; MacLeod et al., 2013; Nierenberg et al., 2014; Xu et al., 2015; Nierenberg et al., 2017). Recently, Nierenberg et al. (2014, 2017) used image flux ratios measured from narrow line emission, a method first proposed by (Moustakas & Metcalf, 2003), to study substructure in strong lenses. The significance of this advance derives from the fact that the magnification of a lensed image is a function of background source diameter (Dobler & Keeton, 2006); the narrow-line region, which typically subtends angles on scales of a few tens of milliarcseconds, is resilient to contaminating effects of microlensing by stars, while still being sensitive to the milliarcsecond perturbations sourced by dark matter halos above

$10^6 M_\odot$ with current astrometric precision of a few m.a.s. (Nierenberg et al., 2017).

In this work, we extend the formalism presented by (Gilman et al., 2018) to include the contribution from dark matter halos along the line of sight. Since field halos do not orbit in a steep galactic potential with star formation, stellar feedback, and other complications, they constitute an ideal laboratory for studying the intrinsic structure of dark matter halos. Several studies investigate the role of the line of sight halos on flux ratio perturbations in strong lenses (Chen et al., 2003; Metcalf, 2005; Miranda & Macciò, 2007; Xu et al., 2012; Inoue & Takahashi, 2012), and Despali et al. (2018b) address the line of sight contribution in the context of gravitational imaging with luminous arcs. The consensus from these works is that the line of sight halos affect lensing observables, possibly becoming the dominant source of perturbation to smooth lens models for lenses at high redshift.

The analysis presented here builds on previous analysis of multiple image lenses in several ways. First, we quantify the signal from non-linear multi-plane lensing effects on flux ratios with finite-size background sources, and combine this multi-plane lensing machinery with a forward-generative model to measure the shape and amplitude of the halo mass function by combining flux ratio statistics from a sample of lenses. We also marginalize over parameters such as the background source size and the power law profile of the main deflector, both of which can affect the flux ratios between images. We demonstrate how well this method constrains the free-streaming length of dark matter in the presence of uncertainties associated with measurements and lens modeling, and apply the machinery to a set of 50 simulated quads. The number 50 is chosen since it is roughly the size of the current sample of known quads (Shajib et al., 2019, HST GO-15652) with a similar distribution of lens and source redshifts.

This paper is organized as follows: First in Section 4.2, we describe our prescription for modeling the line of sight halo mass function, and the subhalo mass function in the main lens plane. In Section 4.3, we discuss the impact of line of sight halos on flux ratio observables. In Section 4.4, we describe the forward modeling procedure implemented in the simulations, and the process for creating mock datasets. Finally, in Section 6.4 we present the results of simulations run with a mock data set in which we infer dark matter

and lens model parameters with a Bayesian framework. Finally, Section 4.6 summarizes our main results. All lensing computations performed in this work utilize the open-source gravitational lensing software LENSTRONOMY (Birrer et al., 2015; Birrer & Amara, 2018). Cosmological calculations, in particular the line of sight halo mass function and two-halo term, are computed with the software package COLOSSUS (Diemer, 2018). We assume a flat cosmology with parameters from WMAP9 (Hinshaw et al., 2013): $\sigma_8 = 0.82$, matter density $\Omega_m = 0.28$ and hubble constant $h = 0.7$. When quoting halo masses, we refer to M_{200} computed with respect to the critical density of the universe at z=0.

4.2 Modeling the line of sight and subhalo mass functions

This section describes the parameterization of the subhalo mass function in the main deflector, and the halo mass function along the line of sight, as well as the density profile for individual halos. We then describe our parameterization of free-streaming effects in WDM models, both on the mass functions and the mass-concentration relation. The forward model, described in Section 4.4, will use these parameterizations to render realistic populations of dark matter structure for lensing computations.

4.2.1 Mass profile of individual halos

We model the density profiles of dark matter halos using truncated NFW profiles (Baltz et al., 2009)

$$\rho\left(r, r_s, r_t\right) = \frac{\rho_0}{x\left(1+x\right)^2} \frac{\tau^2}{x^2+\tau^2} \tag{4.1}$$

where $\tau = \frac{r_t}{r_s}$ and $x = \frac{r}{r_s}$.

In the main lens plane, we tidally truncate subhalos through a Roche limit approximation, assuming a roughly isothermal mass profile for the main lens halo mass distribution. This truncation corresponds to a scaling $r_t \propto (M_{200} r_{\rm 3D}^2)^{\frac{1}{3}}$ (Tormen et al., 1998; Cyr-Racine et al., 2016). We truncate according to this scaling using the expression

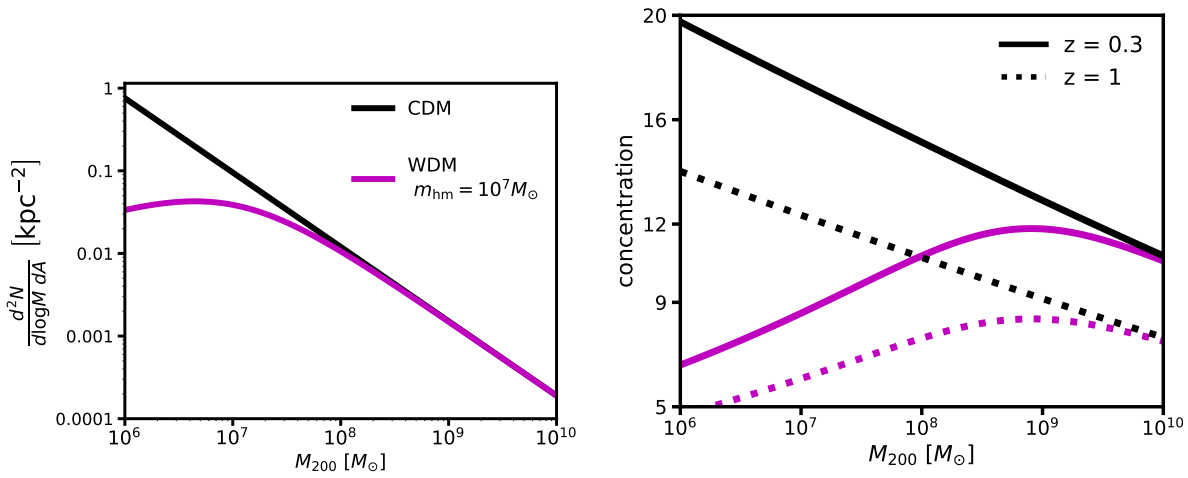

Figure 4.1: **Top:** The subhalo mass function for CDM and a WDM scenario with a half-mode mass of $10^7 \mathrm{M}_\odot$. The line of sight halo mass function looks qualitatively similar, but evolves slightly with redshift. **Bottom:** The mass concentration relation for the same CDM and WDM models as in the upper panel. The effects of free-streaming on the mass-concentration relation alter the properties of halos two orders of magnitude above the half-mode mass.

$$r_t = 0.68 \left(\frac{M_{200}}{10^6 \mathrm{M}_\odot}\right)^{\frac{1}{3}} \left(\frac{r_{3\mathrm{D}}}{100\mathrm{kpc}}\right)^{\frac{2}{3}} [\mathrm{kpc}]\,. \tag{4.2}$$

This results in a skewed distribution of τ with mean $\langle\tau\rangle \sim 6$, and a tail extending to $\tau \sim 20$.

We truncate line of sight halos at r_{50}, or the radius where the mean enclosed density is $50\rho_{\rm crit}$ [1]. Finally, we adopt the mass-concentration-redshift relation for CDM halos presented by (Diemer & Joyce, 2019) with a scatter of 0.13 dex (Dutton & Macciò, 2014). We render halos and subhalos in the range $10^6 - 10^{10} \mathrm{M}_\odot$, which captures perturbations from the smallest subhalos that affect image magnifications, given the source sizes we model. We discuss the rationale for using this mass range in Section 4.4.1.

[1] We introduce this truncation to keep the total mass per unit volume along the line of sight finite, since the mass of an NFW profile diverges. Since r_{50} is much larger than the scale radius of an NFW halo, this truncation negligibly impacts observables.

4.2.2 The line of sight halo mass function

We model line of sight structure using the mass function of Sheth and Tormen (Sheth et al., 2001), plus a boost from the 2-halo term at a distance r from the main deflector $\xi_{\rm 2halo}\left(r, M, z\right)$, where M denotes the halo mass of the parent dark matter halo. The two-halo term accounts for the correlated structure in the vicinity of the main lens halo. To leading order, this term rescales the background density and the amplitude of the halo mass function. The inclusion of $\xi_{\rm 2halo}$ results in a roughly $5-15\%$ boost in the number of halos located at approximately the main lens redshift, depending on the normalization of the subhalo mass function and the lens redshift. We review the form of the two-halo term and its implementation in lensing simulations in Appendix 5.8.

We introduce a rescaling factor $\delta_{\rm los}$ to account for theoretical uncertainty regarding the amplitude of the halo mass function. This term also accounts for statistical fluctuations around the mean density of the universe, which may lead to modestly over-dense or under-dense lines of sight to individual lenses. We note, however, that due to the vast cosmological distances probed by strong lenses (order Gpc, versus kpc-scale dark matter halos and filament diameters) the dark matter structure in these volumes should be well represented by the average halo mass function in the universe which corresponds to $\delta_{\rm los} = 1$, modulo uncertainties in parameters such as σ_8 and Ω_m.

With these modifications, the line of sight halo mass function takes the form

$$\frac{d^2N_{\rm los}}{dmdV} = \delta_{\rm los}\left(1+\xi_{\rm 2halo}\right)\frac{d^2N}{dmdV}\Big|_{\rm ShethTormen}. \tag{4.3}$$

This mass function yields accurate counts of isolated halos over a wide mass range. We do not model the subhalos of these objects along the line of sight, subsuming the possible effects of these small perturbers into the marginalization over $\delta_{\rm los}$. Line of sight halos are distributed in a double-cone geometry with opening angle $3R_{\rm Ein}$, where $R_{\rm Ein}$ is the Einstein radius of a given lens, and a closing angle behind the main lens plane such that the cone closes at the source redshift.

The addition of halos along the line of sight and specifying a flat cosmology introduces an artificial focusing of light rays. To counteract this effect, we add negative convergence

sheets along the line of sight computed with respect to the mean mass in dark matter halos we render (see Birrer et al., 2017a).

4.2.3 The subhalo mass function of the main deflector

We parameterize the subhalo mass function in terms of a projected number density per unit area $\Sigma_{\rm sub}$. In principle, the abundance and spatial distribution of substructure depends on the total mass of the parent dark matter halo and redshift (Gao et al., 2011; Han et al., 2016), and tidal stripping, which can dramatically reduce the subhalo content of galactic halos (Despali & Vegetti, 2017; Han et al., 2016; Garrison-Kimmel et al., 2017; Jiang & van den Bosch, 2017; Richings et al., 2018). We may therefore write the subhalo mass function as

$$\frac{d^2N}{dmdA} = \frac{\Sigma_{\rm sub}}{m_0}\left(\frac{m}{m_0}\right)^{-\alpha} F\left(M\right) H\left(z\right) \tag{4.4}$$

where F and H encode dependence on the parent halo mass M and redshift, respectively.

We render subhalos out to a maximum projected radius of $R_{\rm max} = 3R_{\rm Ein}$, and render the subhalo z-coordinates in three dimensions out to the virial radius of the parent halo. In the semi-cylindrical volume defined by the viral radius and $R_{\rm max}$, we assume the spatial distribution of subhalos follows the mass profile of the parent NFW halo outside $r_{\rm 3D} = 0.5R_{\rm s}$, where R_s is the scale radius of the parent halo, and assume the spatial distribution (per unit volume) is constant inside $0.5R_{\rm s}$. This reflects the impact of tidal stripping, which tends to preferentially destroy subhalos in the central regions of halos (Jiang & van den Bosch, 2017). This procedure sets the distribution of subhalo z-coordinates, which affects the truncation of subhalos through Equation 5.6. When we render halo populations from this mass function and the line of sight halo mass function, we draw from a Poisson distribution with mean $\langle N \rangle$ obtained by normalizing and integrating Equation 4.4 (see Section 4.4).

4.2.4 Modeling free-streaming effects in WDM

Diffusion of dark matter particles in the early universe suppresses small scale power in the matter power spectrum below a characteristic 'free-streaming length' that depends on the

WDM particle mass and formation mechanism. For a more detailed review of WDM theory, see Benson et al. (2013); Schneider et al. (2013).

We parameterize free-streaming effects on the mass function through the half-mode mass $m_{\rm hm}$, defined with respect to the length scale where the WDM transfer function is damped with respect to the CDM transfer function by one-half. In WDM models, the number of halos below $m_{\rm hm}$ is strongly suppressed with respect to CDM. We adopt the functional form for this effect given by Lovell et al. (2014)

$$\frac{dN_{\rm wdm}}{dm} = \frac{dN_{\rm cdm}}{dm}\left(1 + \frac{m_{\rm hm}}{m}\right)^{-1.3}. \tag{4.5}$$

We note that other parameterizations for the turnover in the mass function differ slightly from Equation 4.5 (see Schneider et al., 2012; Benson et al., 2013). For instance, the WDM mass functions by Benson et al. (2013) exhibit a harder turnover than the parameterization in Equation 4.5 due to physical effects, namely, the presence of dark matter velocity dispersion at early times. Other (non-physical) variables, including the different algorithms for identifying and assigning mass to halos, and the choice of window function used to compute the matter power spectrum, can yield different mass functions for the same dark matter model. We do not explicitly address these complications in this work. Finally, we note that the effects of dark matter free-streaming may be enhanced at high redshift, suppressing halo counts relative to CDM more than that predicted by Equation 4.5. This would increase the disparity between CDM and WDM on small scales, which would result in stronger constraints on $m_{\rm hm}$ than those we project in this work. However, lacking a clear prediction for the redshift evolution of the WDM mass function, we do not model the effect in our forecasts.

Thermally produced dark matter particles (thermal relics), assuming they comprise the entirety of the dark matter, admit a one-to-one mapping between the half-mode mass $m_{\rm hm}$ and the mass of the dark matter particle $m_{\rm DM}$. To translate between these two quantities, we use the scaling $m_{\rm hm} \sim m_{\rm DM}^{-3.33}$ (Schneider et al., 2012), and normalize this relation using the $2 \times 10^8 {\rm M}_\odot h^{-1} \sim 3.3$ keV constraint from the Lyman-α forrest (Viel et al., 2013). This yields

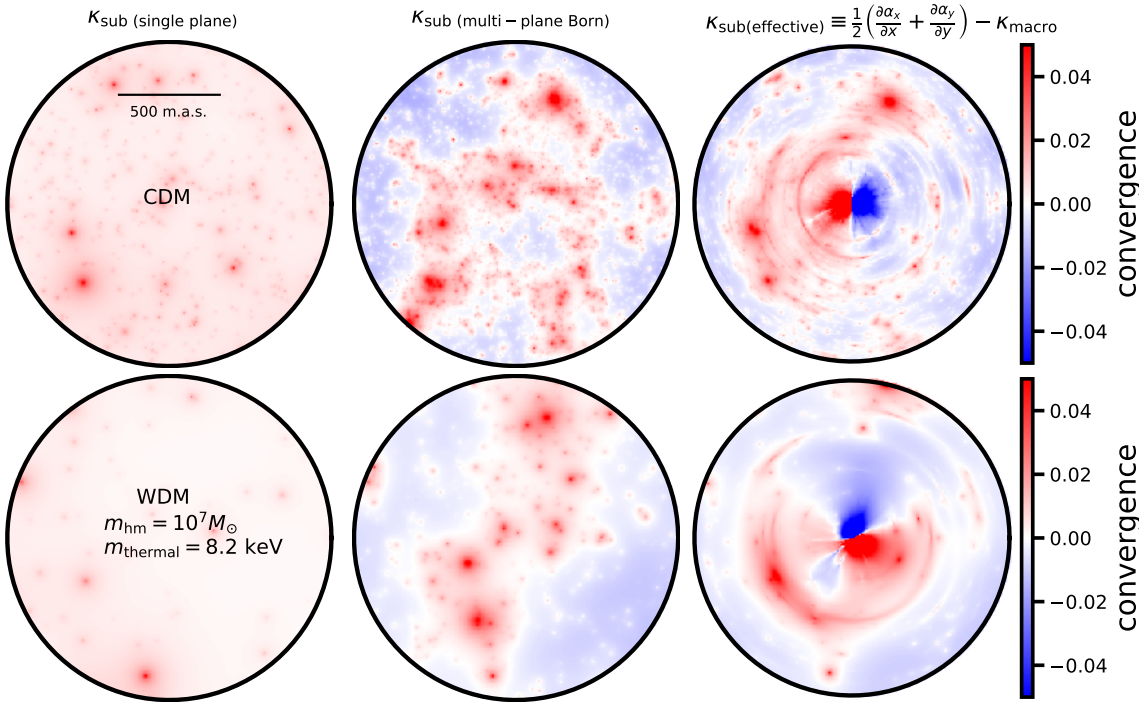

Figure 4.2: A visualization of the mass distributions that affect observables in single-plane and multi-plane lensing. The top and bottom rows show a single realization of CDM and WDM structure, respectively. **Left:** The convergence map from subhalos of the main deflector only, with $\Sigma_{\rm sub} = 0.012 {\rm kpc}^{-2}$, which corresponds to a projected mass fraction in substructure at the Einstein radius of 1% at $z = 0.5$. **Center:** The full line of sight realization viewed in projection. Computing deflection angles with respect to these mass distribution effectively employs the Born approximation, in which the deflection angles from halos at different redshifts are computed by assuming light travels along an unperturbed path. There are blue regions with negative mass due to the inclusion of negative convergence sheets at each lens plane (see discussion in Section 2.2). **Right:** The *effective multi-plane convergence* for these realizations. The deflection angles corresponding to these convergence maps, after subtracting off the convergence from the main deflector, include the non-linear effects present in multi-plane lensing not captured by the Born approximation (see Appendix 5.8).

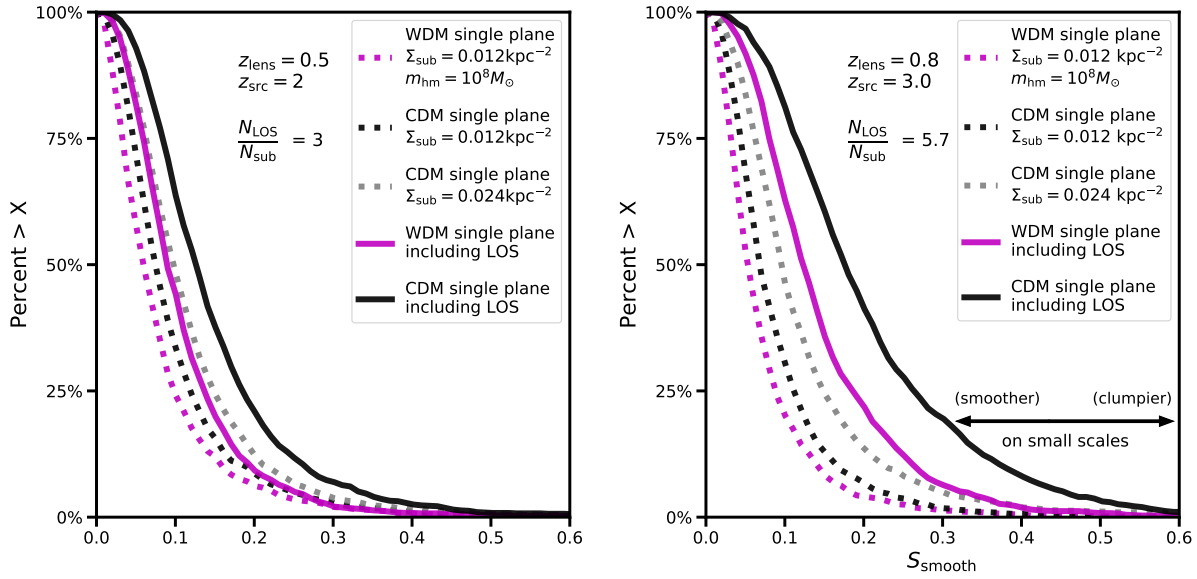

Figure 4.3: Distributions of the summary statistic in Equation 4.10 for different dark matter mass functions, and lens and source redshifts. Dotted curves represent realizations of main deflector subhalos only, while the solid curves include both subhalos and line of sight halos. Black and grey curves denote CDM mass functions with normalizations $\Sigma_{\rm sub} = 0.012{\rm kpc}^{-2}$ and $0.024{\rm kpc}^{-2}$, respectively, while magenta curves correspond to WDM mass functions with $\Sigma_{\rm sub} = 0.012{\rm kpc}^{-2}$ and $m_{\rm hm} = 10^8{\rm M}_\odot$. Mass functions with more small scale structure produce more frequent flux ratio anomalies with respect to smooth lens models, which results in longer tails in the cumulative distribution of these statistics. The boost in the frequency and magnitude of flux ratio anomalies is much stronger for configurations with higher lens/source redshifts.

$$m_{\rm hm}\left(m\right) = 10^{10}\left(\frac{m_{\rm DM}}{1{\rm keV}}\right)^{-3.33}{\rm M}_{\odot}h^{-1}. \tag{4.6}$$

In addition to a suppressed mass function below the free streaming scale, free streaming alters the concentration-mass relation of WDM halos (Schneider et al., 2012; Macciò et al., 2013; Bose et al., 2016; Ludlow et al., 2016). We model this suppression using the parameterization given by (Bose et al., 2016)

$$\frac{c_{\rm wdm}\left(m,z\right)}{c_{\rm cdm}\left(m,z\right)} = \left(1+z\right)^{\beta(z)}\left(1+60\frac{m_{\rm hm}}{m}\right)^{-0.17}. \tag{4.7}$$

with $\beta\left(z\right) = 0.026z - 0.04$.[2] We plot the subhalo mass function and the mass concentration relation in Figure 5.3. Due to the factor of 60 in Equation 5.12, the effect on halo concentrations affects the central densities of objects with masses significantly above the half-mode mass.

4.3 Effect of line of sight structure on image flux ratios

In order to constrain different dark matter models, we must accurately predict image flux ratios in the presence of perturbing dark matter halos in the main lens plane and along the line of sight. To this end, in this exploratory section we investigate the effect of halos at multiple redshifts on flux ratio observables. First, we present visualizations of the non-linear effects present in multi-plane lensing by defining an effective single plane mass distribution for a multi-plane lens system. We then quantify the signal in flux ratios from line of sight structures using a summary statistic, and compare the contributions from subhalos in the main deflector to the signal from line of sight objects for lenses at different redshifts.

4.3.1 Multi-plane lensing

As photons traverse the cosmos from a background source to an observer, they experience numerous deflections by dark matter halos along the line of sight. One formulation of

[2]We remind the reader that we use the $c_{\rm cdm}\left(m,z\right)$ relation presented in Diemer & Joyce (2019).

the equation describing these deflections and the path of deflected light rays is given by (Schneider, 1997)

$$\boldsymbol{\beta_S} = \boldsymbol{\theta} - \frac{1}{D_s} \sum_{n=1}^{S-1} D_{ns} \boldsymbol{\alpha_n} \left(D_n \boldsymbol{\beta_n} \right). \tag{4.8}$$

where $\boldsymbol{\beta_S}$ and $\boldsymbol{\theta}$ denote angular coordinates in the source plane and on the sky, respectively, and where D_n and D_{ns} denote angular diameter distances to the nth lens plane, and between the nth lens plane and the source plane.

In the case of a single lens plane, the deflection field from multiple halos is a linear superposition of the deflections from each individual halo. In the case of multiple lens planes, however, Equation 5.3 becomes a recursive equation for the $\boldsymbol{\beta_n}$, coupling the deflections from halos at different redshifts. Equation 5.3 describes a physical process akin to looking through a magnifying glass through the lens of another magnifying glass (or in the case of substructure lensing, through thousands of other magnifying glasses). For additional details on multi-plane lensing, see Schneider et al. (1992).

The number of halos along the line of sight often outnumber main lens plane subhalos, to a degree that depends on the lens and source redshifts, and the normalization of the subhalo mass function. However, number counts do not accurately reflect the effects of these line of sight objects on lensing observables. First, the geometry defined by the lens and source redshifts results in different lensing efficiencies for halos at different redshifts. Second, the coupling between deflections by halos at different redshifts results in non-linear effects that impact the deflection angles.

To glean some physical intuition of the lensing effects at play in a multi-plane system, we adopt a definition of the lensing surface mass density for multi-plane systems that encodes redshift-dependent lensing efficiency, and non-linear coupling between different lens planes. We define $\kappa_{\text{effective}}$, the *effective multi-plane convergence*, as

$$\kappa_{(\text{effective})} \equiv \frac{1}{2} \boldsymbol{\nabla} \cdot \boldsymbol{\alpha} \tag{4.9}$$

where $\boldsymbol{\alpha}$ is the deflection field of the lens system, or the mapping from a coordinate on the sky to a position in source plane through multi-plane ray-tracing.

This definition expresses the convergence of a multi-plane realization in terms of deflections angles (α_x, α_y) rather than a lensing potential, but is equivalent to the usual definition of convergence in the case of a single lens plane. [3] We compute these deflection angles by ray-tracing through the line of sight according to Equation 5.3. To obtain an effective substructure convergence $\kappa_{\rm sub(effective)}$, we simply subtract the convergence profile of the main deflector $\kappa_{\rm macro}$ (the macromodel), from the full $\kappa_{\rm (effective)}$.

The definition of κ in Equation 4.9 permits a comparison between single plane and multi-plane 'convergence' maps. For illustrative purposes, in Figure 4.2, we render a full multi-plane realization of NFW halos between $10^{5.7}$ and $10^{10}{\rm M}_{\odot}$, for a CDM and WDM scenario. The far left panels show only the single-plane realizations of the subhalo mass function, as would be present in a typical strong lens halo. The central panels show the single plane realizations plus the a full line of sight realization viewed in projection, with coupling between the multiple lens planes turned off. The lensing properties of this convergence map correspond to adopting the Born approximation in lensing, in which lensing quantities are computed by assuming the light rays follow unperturbed paths through the lens planes in front of and behind the main deflector. The far right panels show the *effective multi-plane convergence* for these realizations. In Appendix 5.9, we compare flux ratios computed with the Born approximation to those computed with full ray-tracing, and find the two approaches yield significantly different observables.

Comparing the mass distribution in the far left panels with those on the far right suggests the inclusion of line of sight objects will dramatically affect the statistics of flux ratio distributions in strong lenses caused by small scale density fluctuations in the projected mass density. In the following sections, we will show that this is indeed the case.

[3] Convergence is equivalent to the projected surface mass density in units of the critical density for lensing $\Sigma_{\rm crit} = \frac{c^2}{4\pi G}\frac{D_s}{D_{ds}D_d}$ in single plane lensing, where subscripts d and s denote the lens and source redshifts. For multiple lens planes, we express κ as a vector-field derived quantity.

4.3.2 Flux ratio statistics with line of sight halos

We perform a simple experiment to build intuition for the impact of line of sight halos on flux ratio observables. First, we compute a set of image positions $\boldsymbol{x}$ and flux ratios $\boldsymbol{f_{\text{reference}}}$ for a smooth lens mass distribution, which for simplicity we model as en elliptical isothermal-ellipsoid with external shear (SIE+Shear). Next, given a dark matter model with fixed Σ_{sub} and m_{hm} (with $\delta_{\text{los}} = 1$ and a background source size of 40 pc FWHM), we render 1,000 realizations of halos this model from Equations 5.9 and 4.4. For each of these realizations, we optimize a smooth model to fit the image positions, and compute the model flux ratios $\boldsymbol{f'}$ with respect to this optimized lens model. We then compute the summary statistic[4]

$$S_{\text{smooth}}\left(\boldsymbol{f'}, \boldsymbol{f_{\text{reference}}}\right) \equiv \sqrt{\sum_{i=1}^{3} \left(f'_i - f_{\text{reference(i)}}\right)^2}. \tag{4.10}$$

The statistic S_{smooth} encodes the amount of flux ratio anomaly with respect to a smooth lens model induced by the presence of dark matter halos. In principle, the distributions of this statistic depend on the reference smooth lens model used to compute $\boldsymbol{f_{\text{reference}}}$, but since we construct these distributions merely for visualization purposes the choice of smooth model is not crucial. These complications notwithstanding, we note that the SIE+Shear profile used to compute S_{smooth} reasonably describes the large-scale mass profile of a typical deflector (Auger et al., 2010; Gilman et al., 2017).

Figure 4.3 shows distributions of S_{smooth} for different lens (source) redshifts of 0.5 (2) and 0.8 (3) with different dark matter models. The addition of line of sight halos increases the frequency of a flux ratio anomaly with respect to a smooth lens model, and the boost is substantially higher for configurations with higher lens and source redshifts. The inclusion of line of sight structure also increases the difference in relative amplitudes between the CDM and WDM (solid black and magenta curves) relative to models with lens plane subhalos only. Finally, the distribution of summary statistics for a CDM mass function with a high normalization (grey dotted curve) resembles the statistics produced in a WDM model with

[4]The summation i runs over the three flux ratios derived from the four image fluxes.

a lower value of $\Sigma_{\rm sub}$. This reflects a degeneracy between the amplitude of the subhalo mass function in the main lens plane, and the turnover scale in the mass function.

In the next Section, we amend the definition of the summary statistic in Equation 4.10 slightly, replacing $\boldsymbol{f_{\rm reference}}$ with a set of observed fluxes from a strong lens $\boldsymbol{f_{\rm obs}}$. We write this new statistic $S_{\rm lens}$ as

$$S_{\rm lens}\left(\boldsymbol{f'}, \boldsymbol{f_{\rm obs}}\right) \equiv \sqrt{\sum_{i=1}^{3}\left(f'_i - f_{\rm obs(i)}\right)^2}. \tag{4.11}$$

Through the forward model, we will attempt to minimize this statistic by computing flux ratios $\boldsymbol{f'}$ with different dark matter mass functions. The model flux ratios that minimize this statistic match the observed flux ratios at the particular image positions, and as such the model flux ratios minimizing the statistic satisfy the same correlations as those present in the data. In Appendix 5.10, we describe the implementation of a fast algorithm for lens model optimizations with many line of sight halos, which we use to compute the statistic in Equation 5.4.

4.4 Simulations of substructure lensing: setup and methodology

In this section, we describe the setup of simulations designed to project the constraining power of flux ratios on a WDM mass function. We first outline the physical assumptions imposed in the simulations, and the priors on the parameters sampled in the forward model. Next, we walk through the forward modeling procedure. The subsequent section describes our implementation of flux uncertainties, both from measurement errors and lens modeling. We then describe how, after accounting for uncertainty in the image fluxes, we construct posterior distributions for the model parameters. Finally, we describe the procedure for creating simulated datasets we will use to test this machinery and make forecasts.

4.4.1 Physical assumptions and priors

The methodology we present is flexible, and accommodates any parameterization for the quantities such as the subhalo mass function, line of sight halo mass function, main deflector mass profile, etc. However, for the purpose of making forecast statements and presenting the methodology, we make several simplifying assumptions regarding the implementation of dark matter physics, mass models, and lensing quantities.

4.4.1.1 The subhalo mass function

First, we do not marginalize over the mass, concentration, or ellipticity of the host dark matter halo. We assume a halo mass of $10^{13}\mathrm{M}_{\odot}$, which is typical for a lensing galaxy (Gavazzi et al., 2007), when distributing halos spatially and evaluating the two-halo term in Equation 5.9. We do not expect the ellipticity of the parent dark matter halo to affect the lens model predictions for image fluxes, since the ellipticity of the lensing galaxy and external shear dominate the quadrupole moment of the mass distribution (Keeton et al., 1997). We also ignore any redshift dependence in the subhalo mass function, although we evolve the line of sight halo mass function evolve with redshift. With these simplifications, the subhalo mass function in Equation 4.4 takes the form

$$\frac{d^2N_{(13)}}{dmdA} = \frac{\Sigma_{\mathrm{sub}}}{m_0}\left(\frac{m}{m_0}\right)^{-\alpha} \tag{4.12}$$

where the subscript (13) refers to the assumed halo mass of $10^{13}\mathrm{M}_{\odot}$. We assume $\alpha = 1.9$ (Springel et al., 2008; Fiacconi et al., 2016).

We derive a projected mass density in subhalos by integrating Equation 4.12 over mass, and find values of $\Sigma_{\mathrm{sub}} \sim 0.01-0.02$ kpc^{-2} yield surface mass densities in substructure similar to those derived in simulations of early-type galaxy halos of $10^7\mathrm{M}_{\odot}\mathrm{kpc}^{-2}$ with a pivot mass of $m_0 = 10^8\mathrm{M}_{\odot}$ (Fiacconi et al., 2016). This normalization in principle depends on the severity of tidal stripping, the host halo mass, the halo redshift, and the halo formation time. Rather than modeling all of these effects from first principles, we subsume them in the normalization Σ_{sub}, and impose a wide (flat) prior on this parameter between $0-0.045\mathrm{kpc}^{-2}$.

Table 4.1: Parameters sampled in the forward model

parameter	definition	prior
$\Sigma_{\rm sub}\,[{\rm kpc}^{-2}]$	normalization of subhalo mass function (Equation 4.12) (rendered between $10^6 - 10^{10}{\rm M}_\odot$)	uniform: $[0, 0.045]$
$m_{\rm hm}\,[M_\odot]$	half-mode mass (Equations 4.5 and 5.12) $\propto$ to free streaming length and thermal relic mass $m_{\rm DM}$	log-uniform: $[4.8, 10]$
$\delta_{\rm los}$	rescaling factor for the line of sight Sheth-Tormen mass function (Equation 5.9, rendered between $10^6 - 10^{10}{\rm M}_\odot$)	uniform: $[0.7, 1.3]$
$\sigma_{\rm src}\,[{\rm pc}]$	source size parameterized as FWHM of a Gaussian	uniform: $[25, 50]$
$\gamma_{\rm macro}$	logarithmic slope of main deflector mass model	uniform: $[2, 2.2]$
$\delta_{xy}\,[{\rm m.a.s.}]$	image position uncertainties	$\mathcal{N}(0, 3)$

Gilman et al. (2018) demonstrate that the mean normalization in the lens sample effectively scales the information content available per lens; we perform the same analysis in this work, examining how the constraints on dark matter respond to different values of $\Sigma_{\rm sub}$.

Given a detailed model for the redshift evolution and halo mass depedence of the normalization, as well as the effects of tidal stripping, a non-flat, more informative prior could be used. Since we lack this information, and since we subsume the halo mass dependence and redshift evolution into $\Sigma_{\rm sub}$, we assume minimum information and use a flat prior.

4.4.1.2 Free streaming in WDM

Regarding the implementation of WDM mass functions, we assume that the parameterization of the mass function turnover near $m_{\rm hm}$ (Equation 4.5) applies to both halos along the line of sight, and for subhalos in the main lens halo. As we vary the half-mode mass $m_{\rm hm}$

between $10^{4.8} - 10^{10}\mathrm{M}_{\odot}$, none of the models considered are truly 'cold' in the sense of GeV-scale WIMPS with free-streaming masses of order an Earth mass. However, provided $m_{\mathrm{hm}} << m_{\mathrm{low}} = 10^{6}\mathrm{M}_{\odot}$, the halo populations rendered result in the same observables as those in a CDM universe. [5] We therefore interpret inferences that favor models with $m_{\mathrm{hm}} < 10^{6}\mathrm{M}_{\odot}$ as consistent with CDM, even though the true half-mode mass may be in fact be much lower than the value we recover. Finally, while we implement scatter and redshift dependence in the mass concentration relation in Equation 5.12, we do not marginalize over the parameters describing the turnover for WDM models.

4.4.1.3 Halo and subhalo mass range

We render subhalos and line of sight halos in the mass range $10^{6} - 10^{10}\mathrm{M}_{\odot}$. We choose the lower bound by reducing the smallest rendered halo mass until the distributions of S_{smooth} (like those in Figure 4.3) become insensitive to lower masses (see also footnote 6). On the other hand, halos more massive than the upper bound of $10^{10}\mathrm{M}_{\odot}$ would likely host stars and be visible, allowing them to be directly included in the main lens model (e.g. Birrer et al., 2019).

4.4.1.4 Scaling of the LOS halo mass function

We vary the rescaling parameter for the line of sight halo mass function between 0.7 and 1.3. This accounts for theoretical uncertainties in the prediction of the halo mass function, which is typically at the $10 - 30\%$ level (Despali et al., 2016). This term also accounts for variance in the average density along the line of sight to strong lenses. This parameter is not meant to account for correlated structure near the main lens plane, which we model through the two-halo term ξ_{2halo}.

[5] This is only true if the signal in flux ratio saturates at m_{low}, otherwise we would miss part of the signal from halos with mass $< m_{\mathrm{low}}$. We verify that halos of mass below $10^{6}\mathrm{M}_{\odot}$ do not significantly affect the flux ratio signal for the background source sizes 25-50 pc.

4.4.1.5 The background source size

The background source size enters the forward model because the perturbation to image magnifications depends on the source size relative to the deflection angle of a perturber (Dobler & Keeton, 2006). Upper limits on the size of the narrow-line region from (Nierenberg et al., 2017) correspond to physical sizes of $\sim$ 50pc, which agrees with the surface brightness profiles seen in low redshift AGN (Müller-Sánchez et al., 2011). We therefore allow the source size to vary between 25 and 50 pc. While in this work we forward model source sizes appropriate for narrow-line emission, the method we present can accommodate flux ratios measured from any band provided it is free from contamination from micro-lensing, including mid-infrared bands (Minezaki et al., 2009; MacLeod et al., 2013).

4.4.1.6 The main deflector

We model the main deflector as a power-law ellipsoid plus external shear. This is a generalization of the widely applied, physically motivated (e.g. Treu et al., 2006) singular isothermal sphere (SIE) profile used to model lensing galaxies. Studies of early-type deflectors find mass profiles $\rho(r) \sim r^{-\gamma_{\rm macro}}$ modestly steeper than r^{-2} (Treu et al., 2009; Auger et al., 2010; Shankar et al., 2017), so we allow the power-law profile $\gamma_{\rm macro}$ to vary between 2 and 2.2. We assume deflectors with complex morphologies, including features like stellar disks, have been identified and removed from our sample, and describe residual baryonic effects by adding perturbations to the forward model image fluxes, a process we describe in Section 4.4.3. We marginalize over uncertainties in image positions by rendering Gaussian astrometric uncertainties of ± 3 m.a.s. in the forward model.

4.4.1.7 Summary

We point out that many of the simplifying assumptions we impose in our forecasts effectively ignore relevant information that could be used to inform a prior. For example, the velocity dispersion of the lensing galaxy could inform a prior on the halo mass and the normalization $\Sigma_{\rm sub}$, and possibly the macromodel profile $\gamma_{\rm macro}$. Since $\Sigma_{\rm sub}$ is somewhat correlated with

$m_{\rm hm}$ (see Section 6.4), this could improve constraints on the free-streaming length of the dark matter. Similarly, modeling redshift dependence in the normalization of the subhalo mass function could break the covariance between $\Sigma_{\rm sub}$ and $\delta_{\rm los}$ (see Section 6.4). This information would therefore improve the precision on the inferred dark matter properties, and it is possible that we overestimate uncertainties by omitting it.

4.4.2 Forward modeling procedure

To constrain the halo mass function, we adopt a forward modeling approach. This consists of generating mock data sets by simulating the physical processes that affect lensing observables, including the size of the background source, dark matter halos in the main lens halo and along the line of sight, the mass profile of the main deflector, and statistical measurement errors. This approach handles complicated degeneracies between model parameters - for example, between halo redshift and halo mass (Despali et al., 2018b) - by building these features directly into the forward-generated data sets. In effect, we exchange the task of computing a complicated likelihood function with the challenge of simulating the relevant physics in strong lensing.

This first step in the forward model is to sample all parameters from their respective prior probability densities, summarized in Table 5.1. For convenience, for the ith realization, we denote the collection of the model parameters $\boldsymbol{M_i}$. Using the parameters describing the dark matter $(\Sigma_{\rm sub}, \delta_{\rm los}, m_{\rm hm})$, we render a the full population of line of sight halos and lens plane subhalos, as described in Section 4.2.

Next, using the observed image positions [6] and fluxes from a strong lens, we optimize a power-law plus external shear lens model with power law slope $\gamma_{\rm macro}$ to fit the observed image positions in the presence of the full population of dark matter halos, and ray-trace to compute the flux ratios with background source modeled as a Gaussian with a FWHM of $\sigma_{\rm src}$. While optimizing the macromodel to fit image positions, we allow the lens Einstein radius, centroid, ellipticity, ellipticity angle, shear, and shear angle to vary, while keeping

[6] We add random statistical measurement errors of ± 3 m.a.s. to the image positions for each realization.

the power-law slope $\gamma_{\rm macro}$ fixed for each optimization. If necessary, we may extend the forward modeling of $\gamma_{\rm macro}$ to additional mass profile parameters to add complexity in the lens macromodel.

At this stage, we have a set of observed flux ratios and a set of flux ratios simulated in the forward model. We use the model-predicted flux ratios $\boldsymbol{f'}$ with the observed flux ratios $\boldsymbol{f}_{\rm obs}$ to compute the summary statistic in Equation 5.4, which we then assign to the set of parameters $\boldsymbol{M_i}$. We repeat this entire procedure 600,000 times for each lens (see the convergence test in Appendix 4.10).

4.4.3 Accounting for uncertainty in image fluxes

We introduce uncertainties in the image fluxes by adding perturbations to the fluxes in the mock data, and by rendering these perturbations in the model fluxes. Explicitly, we modify each model-predicted image flux f_i as

$$f_i \rightarrow f_i + \mathcal{N}\left(\prime, \delta\right). \tag{4.13}$$

The most straightforward interpretation of this procedure is the incorporation of statistical measurement errors. For reference, current measurements of narrow-line fluxes achieve precision of $3-6\%$ (Nierenberg et al., 2014, 2017). These perturbations also simulate the role of unknown sources of uncertainty, or simply those we do not explicitly model. For example, in cases where a more complex macromodel is required, the additional degrees of freedom that must be marginalized over result in a larger variation in image fluxes at fixed image positions, which effectively introduces an additional source of flux uncertainty.

We will explicitly consider flux perturbations of 2%, 4%, 6%, and 8%. The intermediate values of 4% and 6% represent current measurement precision (Nierenberg et al., 2017) and modeling uncertainties (Gilman et al., 2017). The 2% value represents a best-case scenario with precise measurements — perhaps with observations from future telescopes such as JWST — and a sample of morphologically simple deflectors that do not require complex macromodels. The 8% value corresponds to a scenario where the majority of the systems in the lens sample require marginalization over complex macromodels.

4.4.4 Bayesian Inference

To construct posterior probability densities for the parameters $\boldsymbol{M}$ listed in Table 5.1, we rank the 600,000 $\boldsymbol{M_i}$ by their summary statistics, with those that minimize the statistic ranked highest. A subset of these models (we use the top 1,500) form a probability density $p'(\boldsymbol{M}|\text{data})$, which becomes an increasingly good approximation of the true posterior distribution $p(\boldsymbol{M}|\text{data})$ as the number of forward model samples increases. This procedure falls in the category of Approximate Bayesian Computing methods (for a review, see (Lintusaari et al., 2017)), and is widely applied to problems with intractable likelihood functions (Akeret et al., 2015; Hahn et al., 2017; Birrer et al., 2017b; Davies et al., 2018). We apply a kernel density estimator to the 1,500 sample that form $p'(\boldsymbol{M}|\text{data})$, and multiply the resulting probability densities to obtain the final posterior. We test for convergence in this algorithm in Appendix 4.10.

We acknowledge that, formally, a marginalization of the macromodel, rather than an optimization of the macromodel, yields the desired posterior distribution of dark matter parameters. We avoid this computationally prohibitive step [7] with two justifications: First, the volume of macromodel parameter space is typically tightly constrained by the requirement that the macromodel fit the image positions. For macromodels parameterized as power-law ellipsoids, the image fluxes do not vary significantly over this volume, and the variation in image fluxes induced by marginalizing over the macromodel is negligible compared to other sources of uncertainty [8] Second, we note that for each of the 600,000 realizations rendered in the forward model, each macromodel re-optimization is independent. Thus, over the course of many realizations, covariance between macromodel parameters and the parameters describing the dark matter content is reflected in the summary statistics.

[7] This is computationally prohibitive because the vast majority of macromodel parameter configurations do not fit the image positions, and therefore consume computation time without contributing to the desired posterior distribution.

[8] We test this by re-sampling a once-optimized macromodel around the peak of the likelihood, and computing the variation in image fluxes.

4.4.5 Creating simulated data sets

To create mock data sets, we parameterize the lens macromodel as a power-law ellipsoid, and generate mock lenses by sampling the Einstein radii, ellipticity, and external shears, as well as lens and source redshifts, from the distributions of these quantities used by Oguri & Marshall (2010). We plot the lens and source redshifts of the 50 quads in our mock lens sample in Figure 4.4. We sample power law slopes drawn from a distribution centered at 2.05 ± 0.04, consistent with the morphological properties of the early-type galaxies that dominate the strong lensing cross section (Auger et al., 2010; Shankar et al., 2017). The background source is parameterized by a circular Gaussian with a FWHM, which we specify within the range $25 - 50$ pc, consistent with the upper limits on the size inferred by Nierenberg et al. (2017), and comparable to the luminous extent of the narrow line region of quasars (Müller-Sánchez et al., 2011).

We choose background source positions to produce roughly equal numbers of cross, fold, and cusp image configurations. Cusp and fold configurations generally yield the strongest constraints on WDM properties (see Appendix B in Gilman et al. (2018)), and since the images in these types of quads have higher magnifications they may be more easily discovered. It is therefore possible that a real sample of quads would consist of more cusp and fold configurations than crosses, in which case the resulting constraints on WDM would be stronger than those obtained in this work.

When generating the mock data sets, we add measurement errors to the image positions of 3 m.a.s., and model statistical measurement errors by adding perturbations to the image fluxes, as described in Section 4.4.3.

4.5 Simulations of substructure lensing: Results

This section presents the results of our analysis, in which we test the forward modeling machinery described in the previous section to constrain dark matter properties. We discuss how measurement and modeling uncertainties affect the precision of constraints on both

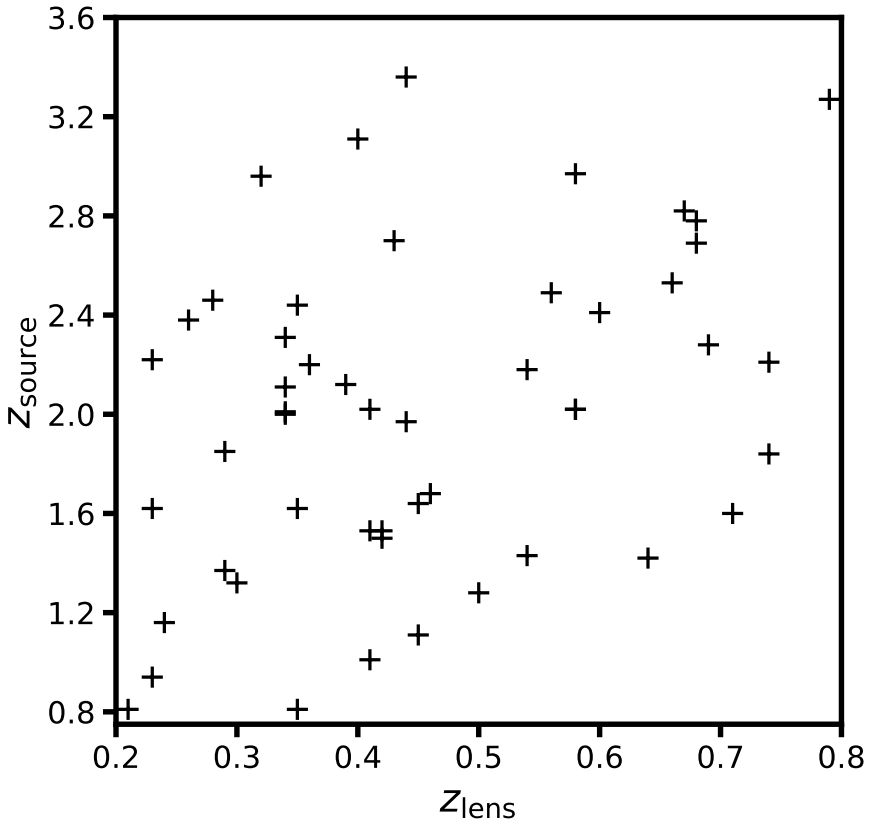

Figure 4.4: The lens and source redshifts for the 50 quads in our mock lens sample. We draw these parameters, along with the lens velocity dispersion, ellipticity, and shear from the distributions used by Oguri & Marshall (2010).

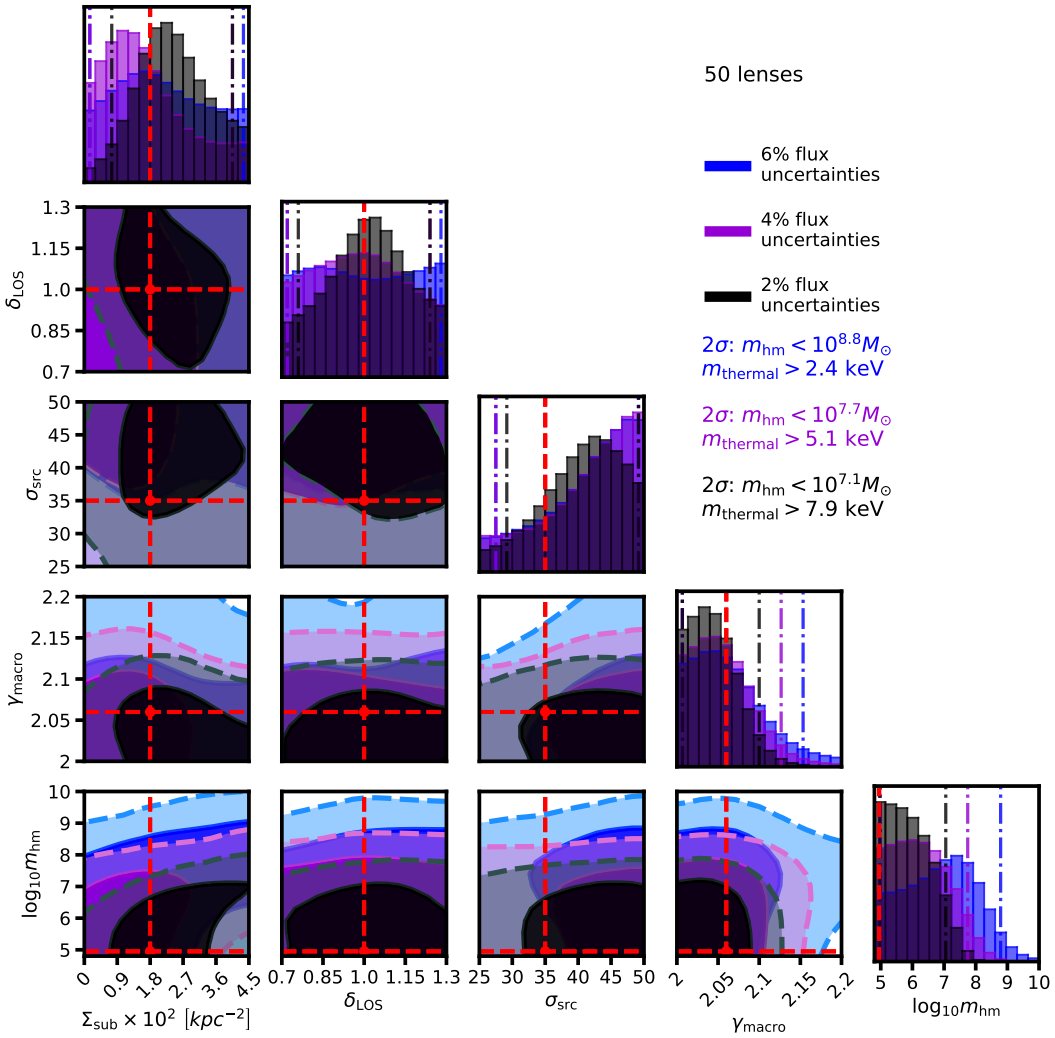

Figure 4.5: The posterior distributions resulting from the forward modeling analysis of a sample of 50 lenses, with flux uncertainties stemming from measurement errors and lens modeling controlled at the 2%, 4%, and 6% level. Vertical bars in the marginal distribution indicate 2σ confidence interval, while dashed (solid) lines in the panels denote 2σ (1σ) contours. The marginalized constraints on $m_{\rm hm}$ range from $10^{7.2}{\rm M}_\odot$ for the case of 2% flux uncertainties, to $10^{8.8}{\rm M}_\odot$ for uncertainties of 6%.

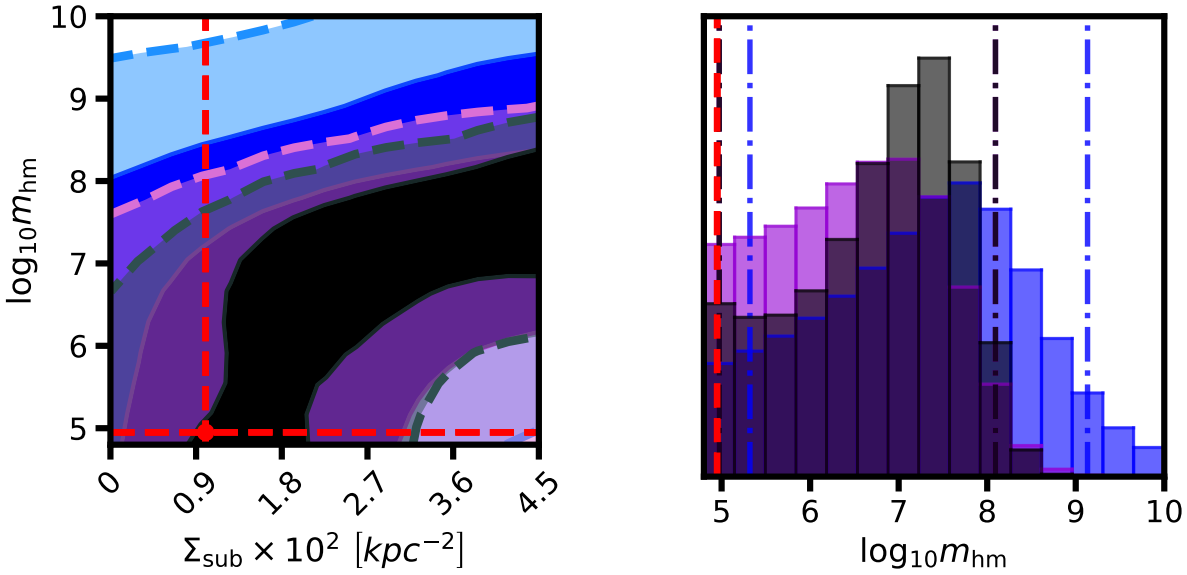

Figure 4.6: Inference on a CDM mass function with a normalization of the subhalo mass function $\Sigma_{\rm sub} = 0.01$, roughly half the value of the normalization assumed in Figure 4.5. The color scheme is the same as in Figure 4.5, with black, magenta, and blue representing flux uncertainties of 2%, 4%, and 6%, respectively. In this case, the marginalized constraints on $m_{\rm hm}$ are $10^{9.1}{\rm M}_\odot$, $10^{8.1}{\rm M}_\odot$, and $10^{8.1}{\rm M}_\odot$ for flux uncertainties of 6%, 4%, and 2% (for the 4% and 2% flux uncertainties, the 2σ confidence interval happen to be the same). These constraints are weaker by roughly an order of magnitude in mass over the bounds quoted in Figure 4.5, which illustrates the role of the normalization of the subhalo mass function on the possible constrains on $m_{\rm hm}$.

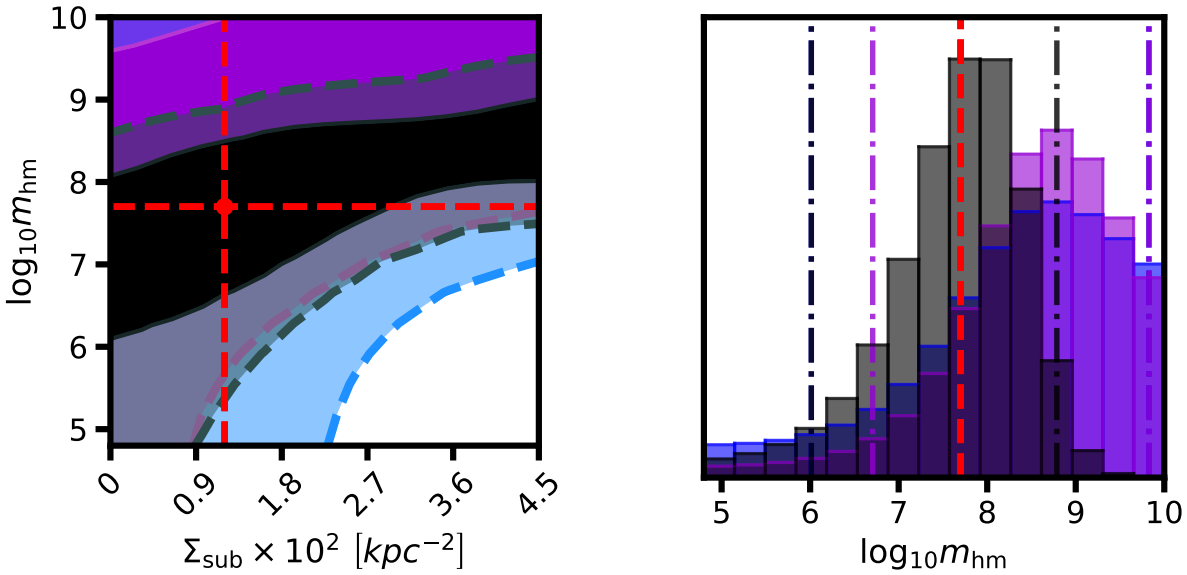

Figure 4.7: Inference on a WDM mass function with a half-mode mass of $10^{7.7}\mathrm{M}_\odot$ ($m_{\rm thermal} = 5.4$ keV), with the same color scheme as Figure 4.5. As in Figure 4.5, we marginalize over the parameters listed in 5.1 and over various degrees of flux uncertainty, and the color scheme is the same as in Figures 4.5 and 4.6. For flux uncertainties of 2%, 4%, and 6%, we favor WDM with $m_{\rm hm} > 10^{7.7}\mathrm{M}_\odot$ over CDM with likelihood ratios of 22, 30, and 8, respectively.

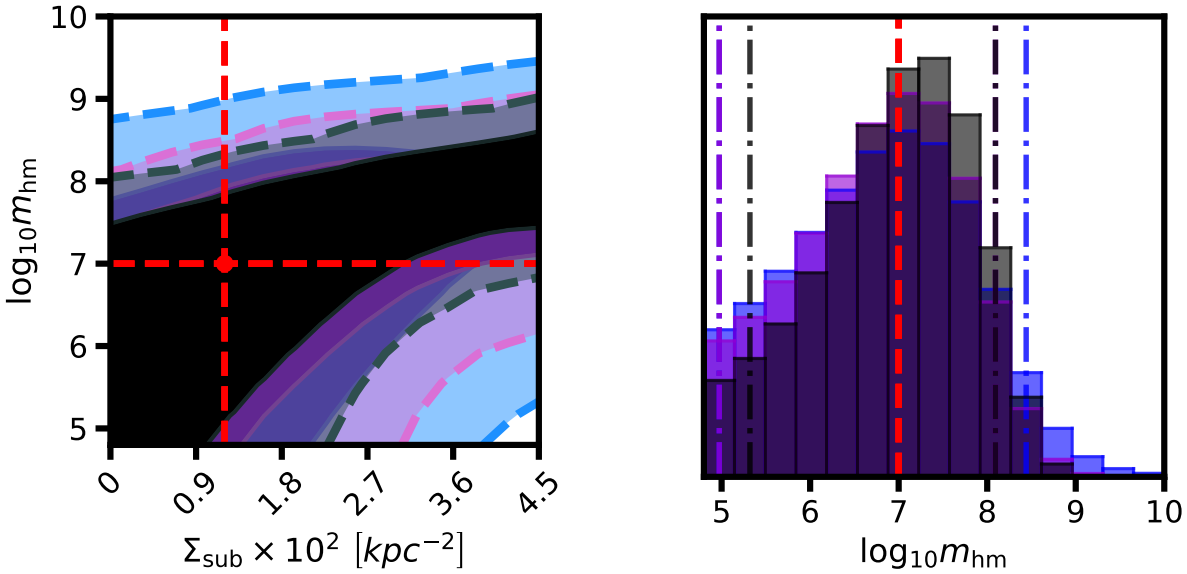

Figure 4.8: Inference on a WDM mass function with a half-mode mass of $10^7 M_\odot$ ($m_{\rm thermal} = 8.2$ keV)., marginalized over the parameters listed in Table 5.1, with the same color scheme as Figure 4.5. For each degree of uncertainty in image fluxes, the peak of the posterior coincides with the location of the the turnover at $10^7 M_\odot$, but the width of the distributions increases. With uncertainties of 2%, 4%, and 6% we favor WDM mass functions with $m_{\rm hm} > 10^7 M_\odot$ over CDM with likelihood ratios of 4:1, 3:1, and 2:1, respectively.

CDM and WDM mass functions, and make projections for the constraints on the half-mode mass. We explicitly consider 4 models: Two CDM cases with a different normalization of the subhalo mass function $\Sigma_{\rm sub}$, and two WDM cases with half-mode masses of $10^{7.7}\mathrm{M}_\odot$, and $10^{7}\mathrm{M}_\odot$.

4.5.1 Joint inference on model parameters

Beginning with the CDM mass functions, in Figure 4.5 we show posterior distributions for all the parameters sampled in the forward model for a CDM mass function with a normalization of $\Sigma_{\rm sub} = 0.018$ kpc^{-2}. As described in Section 4.4, we add flux perturbations of 2%, 4%, and 6% the mock data and model fluxes to simulate measurement errors, and additional sources of flux uncertainty that stem from lens modeling. We marginalize over ten realizations of these flux perturbations to reduce shot noise in the posterior distributions.

The boost in signal from the line of sight halos permits 2σ bounds on the half-mode mass that range between $m_{\rm hm} < 10^{7.1}\mathrm{M}_\odot$, or a 7.9 keV thermal relic particle, to $m_{\rm hm} < 10^{8.8}\mathrm{M}_\odot$ (2.4 keV) as statistical measurement errors and modeling uncertainties in image fluxes increase from 2% to 6%. This rapid erosion of constraining power underscores the necessity of accurately measuring image fluxes, and accurate lens model predictions for these observables.

The most visibly striking covariance in Figure 4.5 exists between $\Sigma_{\rm sub}$ and $m_{\rm hm}$ (see also Figure 4.6). Physically, this feature corresponds to adding more substructure by increasingly the normalization, and subsequently removing some of the subhalos by raising the half-mode mass such that the total amount of flux perturbation remains relatively constant. Thus, above a sensitivity threshold of roughly $10^{6}\mathrm{M}_\odot$, $\Sigma_{\rm sub}$ and $m_{\rm hm}$ are positivity correlated. The opposite is true for $\Sigma_{\rm sub}$ and $\delta_{\rm los}$: the additional source of flux perturbation from extra line of sight structure is partially offset by reducing the number of lens plane subhalos, and these parameters are anti-correlated. Finally, there is weak evidence (notice the slightly tilted 2σ contours) for a positive correlation between the power-law slope of the macromodel $\gamma_{\rm macro}$ and the source size $\sigma_{\rm src}$. Without a priori knowledge of the true source size, the focusing

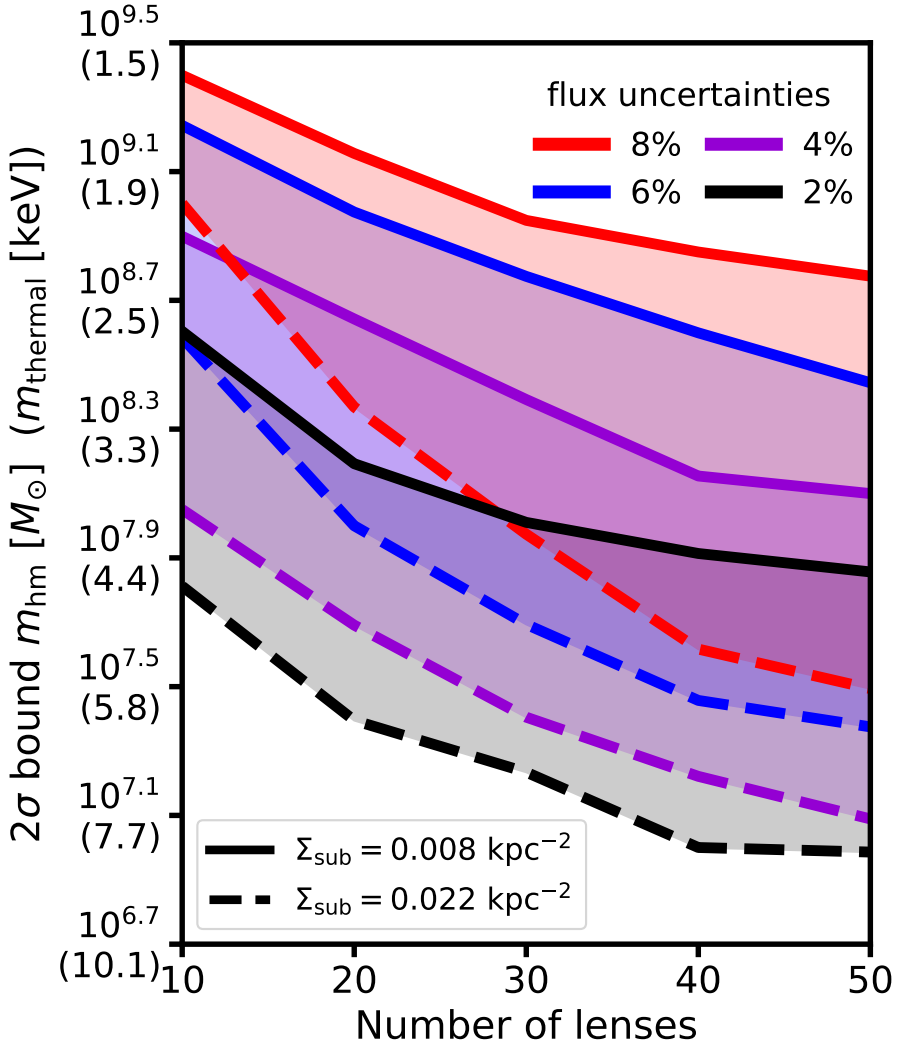

Figure 4.9: Forecasts for the constraints on the half-mode mass as a function of the number of lenses, including line of sight halos and subhalos of the main deflector. Black, purple, blue, and red colors denote flux uncertainties of 2%, 4%, 6%, and 8%. The solid line corresponds to a normalization $\Sigma_{\rm sub} = 0.008{\rm kpc}^{-2}$, while dashed lines correspond to $\Sigma_{\rm sub} = 0.022{\rm kpc}^{-2}$. The y-axis labels represent the 2σ bound on $m_{\rm hm}$, with the mass of the corresponding thermal relic dark matter particle in parentheses. Models with more subhalos (dashed lines), and hence more signal, are more resilient to flux uncertainties than models with fewer lens planes subhalos (solid lines) and produce stronger constraints on $m_{\rm hm}$.

power of a lens with a steeper mass profile makes larger background sources look smaller. Thus, a more extended background source is focused to the same size image by a steeper mass profile and these parameters are positively correlated. We emphasize that despite the covariance between parameters such as $\Sigma_{\rm sub}$ and $m_{\rm hm}$, the data still constrains these parameters independently. The covariance affects the precision of the inference, but it does not result in completely unconstrained posterior distributions.

The normalization of the subhalo mass function $\Sigma_{\rm sub}$ plays an important role in the constraints on WDM and CDM models. Systems with more substructure are effectively weighted more than systems with fewer subhalos, and the strength of the constraints on $m_{\rm hm}$ reflect this weighting. We illustrate this effect in Figure 4.6, through comparison with Figure 4.5. The former has $\Sigma_{\rm sub} = 0.01$ kpc^{-2}, while the latter has nearly twice as many lens plane subhalos with $\Sigma_{\rm sub} = 0.018$ kpc^{-2}. The constraints on $m_{\rm hm}$ are weaker for the simulation with less substructure, because the data contains less signal. Due to the covariance between $\Sigma_{\rm sub}$ and $m_{\rm hm}$, a significant portion of the volume of the posterior lies in high $\Sigma_{\rm sub}$, high $m_{\rm hm}$ parameter space, which results in a peak in the marginalized constraint on $m_{\rm hm}$. Stronger theoretical priors on $\Sigma_{\rm sub}$, which take into account the role of halo mass, redshift, and tidal stripping, may improve constraints on $m_{\rm hm}$ by breaking this covariance.

It is possible that by extending the range of the prior on $\Sigma_{\rm sub}$ to higher values, the covariance between $m_{\rm hm}$ and $\Sigma_{\rm sub}$ would result in weaker constraints on the half-mode mass. However, extending the prior in this manner would imply a degree of ignorance surrounding the parameter $\Sigma_{\rm sub}$ that would likely be exaggerated given the current state of numerical simulations of dark matter halos and their substructure (Benson, 2012; Wheeler et al., 2018; Bozek et al., 2019; Lovell et al., 2018). Keeping the width of the prior fixed, we implicitly assume that one may predict $\Sigma_{\rm sub}$ for each lens halo to within the width a factor of 4.5, or the width of the prior on $\Sigma_{\rm sub}$.

In Figures 4.7 and 4.8, we show the constraints on WDM mass functions with $m_{\rm hm}$ of $10^{7.7}$M$_\odot$ and 10^{7}M$_\odot$, which correspond to thermal relic dark matter particles of 5.1 and 8.2 keV, respectively. Both datasets have $\Sigma_{\rm sub} = 0.012$ kpc^{-2}. As in Figures 4.5 and 4.6, we marginalize over every parameter listed in Table 5.1, but focus only on the joint distribution

of $\Sigma_{\rm sub}$ and $m_{\rm hm}$. We see evidence for a turnover in the mass function, even though it lies below $10^8 {\rm M}_\odot$. When interpreting the marginalized posteriors for $m_{\rm hm}$ in cases where there is a clear peak in WDM territory, we use the relative likelihood between the lowest $m_{\rm hm}$ bin (at $10^{4.8} {\rm M}_\odot$) and the peak of the posterior as a summary statistic, since the statement regarding the 2σ confidence interval depends on the width of the prior. [9]

In the case of $m_{\rm hm} = 10^{7.7} {\rm M}_\odot$, with flux uncertainties of 2%, 4%, and 6%, we favor WDM mass functions with $m_{\rm hm} > 10^{7.7} {\rm M}_\odot$ over CDM with relative likelihoods of 22:1, 30:1, and 8:1, respectively [10]. With uncertainties of 4% and 6%, the posterior distributions of $m_{\rm hm}$ shift towards higher masses, and the posteriors no longer resolve the position of the turnover in the mass function and mass-concentration relation. The shift to higher values of $m_{\rm hm}$ is a consequence of the weak signal produced by very warm mass functions with a paucity of small-scale structure. Increased flux uncertainties wash out the information from the 'weak signal' regime of parameter space with $m_{\rm hm} > 10^{7.7} {\rm M}_\odot$, and the constraints on this region of parameter space deteriorate because the data itself lies in this 'weak signal' regime. This reasoning is similar to the interpretation of $\Sigma_{\rm sub}$ as an information scaling parameter for CDM mass function: like a CDM mass function with a high normalization, a 'colder' WDM mass function produces more significant flux perturbation events, and is more resilient to increased uncertainties in image fluxes. If this reasoning is correct, we should expect the posteriors on $m_{\rm hm}$ for 'colder' WDM mass function to remain relatively stationary, modulo an increased variance, after adding perturbations to the image fluxes.

This effect is seen in Figure 4.8, which has $m_{\rm hm} = 10^7 {\rm M}_\odot$. The shift of the posterior distributions towards higher masses as flux uncertainties increase does not happen in this case because the WDM mass function with $m_{\rm hm} = 10^7 {\rm M}_\odot$ produces stronger perturbations in the data than the warmer, 'weak signal' model with $m_{\rm hm} = 10^{7.7} {\rm M}_\odot$. This is because the halos are both more numerous and more concentrated that the WDM model with $m_{\rm hm} = 10^{7.7} {\rm M}_\odot$.

[9] Sometimes, inference on CDM mass functions results in a posterior distribution peaked around some value of $m_{\rm hm}$, due to the covariance between $m_{\rm hm}$ and other parameters. This effect is visible in Figure 4.6. In the case of Figure 4.6, the maximum likelihood ratio between WDM and CDM with uncertainties of 2% equals two.

[10] The increase from 22 to 30 is likely due to shot noise.

In turn, the stronger signal survives additional flux uncertainties, and is sufficient to constrain very warm mass functions. The locations of the peaks of the posteriors coincide with the true value of $m_{\rm hm}$, but the width of the distributions widen. In this case, we favor WDM mass functions with $m_{\rm hm} > 10^7 {\rm M}_\odot$ over CDM mass functions with relative likelihoods of 4:1, 3:1, and 2:1 with flux uncertainties of 2%, 4%, and 6%, respectively. The fact that we statistically favor WDM models over CDM models suggests that we could infer a turnover in the mass function at $m_{\rm hm} = 10^7 {\rm M}_\odot$ (or an 8.2 keV WDM particle) at higher significance with a larger sample of quads.

4.5.2 Marginalized constraints on the free-streaming length

The posterior distributions in Figures 4.5 and 4.6 give a sense for how the constraints on the half-mode mass in WDM models depends on the precision with which one measures image fluxes and predicts them with lens models, and on parameters such as the normalization of the subhalo mass function. To take into account sample variance, in Figure 4.9 we plot the marginalized constraints on the half-mode mass as a function of the number of lenses, $\Sigma_{\rm sub}$, and flux measurement uncertainties of 2%, 4%, 6%, and 8%. We plot the bounds on $m_{\rm hm}$ for both a high ($\Sigma_{\rm sub} = 0.02 {\rm kpc}^{-2}$) and low ($\Sigma_{\rm sub} = 0.008 {\rm kpc}^{-2}$) normalization of the subhalo mass function. To produce these curves, we compute 200 bootstraps of 50 lenses, and average over many realizations of flux uncertainties.

With a sample of 50 lenses it will be possible to probe below $10^8 {\rm M}_\odot$ in the halo mass function, to a degree that depends on the amount of substructure in the main deflector, measurement precision of image fluxes, and precise lens model predictions for this observable. With control over image fluxes at the level for 4%, routinely achieved at present (Nierenberg et al., 2014, 2017), the bounds on $m_{\rm hm}$ with 50 quads range between $10^{7.1} - 10^{8.1} {\rm M}_\odot$ for values of $\Sigma_{\rm sub}$ of 0.01 and 0.022 kpc^{-2}, respectively. With more precise predictions of $\Sigma_{\rm sub}$ made on a lens-by-lens basis, these bounds may improve. We also note that future surveys, such as LSST, WFIRST, and Euclid, will discover hundreds of quads (Oguri & Marshall, 2010), so the sample of available quads will eventually be much larger than 50.

4.6 Summary and conclusions

We have presented a method to perform Bayesian inference on the halo mass function through a forward modeling analysis of image flux ratios in quadruply imaged quasars. We model the contribution from line of sight halos, which boost the signal per lens and permit stronger constraints on the properties of dark matter with fewer systems. We demonstrate the method with a sample of 50 quads, comparable in number to the currently observed sample size, and project the constraints on the free streaming length of a WDM particle under different degrees of flux measurement and lens modeling uncertainties, while marginalizing over parameters describing the size of the background source, the lens macromodel, and the amplitude of the line of sight halo mass function.

Our key results can be summarized as follows:

- With a sample of 50 quads, we are able to constrain the free streaming length of dark matter on scales below $10^8 \mathrm{M}_\odot$. Assuming CDM, with mean subhalo mass function normalizations $\Sigma_{\rm sub} = 0.022 \mathrm{kpc}^{-2}$ $(0.008 \mathrm{kpc}^{-2})$ we forecast bounds on the half-mode mass of 10^7 $(10^{7.9})\,\mathrm{M}_\odot$, $10^{7.1}$ $(10^{8.1})\,\mathrm{M}_\odot$, $10^{7.4}\,(10^{8.4})\,\mathrm{M}_\odot$, $10^{7.5}$ $(10^{8.8})\,\mathrm{M}_\odot$ for flux uncertainties of 2%, 4%, 6%, and 8%, respectively. These $m_{\rm hm}$ limits translate to bounds on the mass of thermal relic particles of 8.2 (4.4), 7.7, (3.8), 6.2 (3.1), 5.8 (2.4) keV.

- Line of sight halos contribute substantially to the signal in flux ratios, even dominating the signal in lens systems with higher lens and source redshifts. However, the normalization of the subhalo mass function still plays a key role in scaling the information content per lens, with higher values of this parameter translating into tighter constraints on the mass function. The half-mode mass is also covariant with the normalization, which affects the marginalized constraints on this parameter. These features underscore the importance of theoretical work to predict the projected surface mass density of substructure inside galactic halos with accurate models of baryonic feedback and tidal stripping.

- In the case that dark matter is warm, we are able to infer the location of the turnover

in the mass function with 50 quads, even if it lies below $10^8 M_\odot$. With a half-mode mass of $10^{7.7} M_\odot$, which corresponds to a 5.1 keV thermal relic particle, we favor WDM mass functions with $m_{\rm hm} > 10^{7.7} M_\odot$ over CDM with relative likelihoods of 22:1, 30:1 and 8:1 for flux uncertainties of 2%, 4%, and 6%, respectively. With the same set of flux uncertainties and a half-mode mass of $10^7 M_\odot$, we favor WDM with $m_{\rm hm} > 10^7 M_\odot$ over CDM with relative likelihoods of 4:1, 3:1, and 2:1. These constraints will likely improve with additional lenses, which suggests that a future large sample of quads could be used to infer a turnover in the halo mass function at $10^7 M_\odot$ at high statistical significance.

Our work is broadly consistent with other studies of the line of sight contribution in substructure lensing. For instance, by ray tracing through N-body simulations, (Xu et al., 2012) compare the frequency of flux anomalies induced by line of sight versus main lens halos, and reach the conclusion that line of sight halos contribute at the same level as subhalos. More recently, (Despali et al., 2018b) analyze the role of line of sight halos in the context of gravitational imaging. This method differs somewhat from this analysis in that it aims to detect individual halos along the line of sight, and in the main lens plane, but the authors reach a similar conclusion: the line of sight contribution substantially boosts the signal per lens. In terms of relative numbers, line of sight halos can outnumber lens plane subhalos by a factor of 2-25, depending on the normalization of the subhalo mass function, and the lens and source redshifts. However, the most robust metric of the influence of line of sight halos comes from the resulting constraints on the half-mode mass. Differences in the treatment of the subhalo mass function, background source size, lens macromodel, and the lens redshift distribution complicate a simple comparison between this work and the results obtained in Gilman et al. (2018) by modeling only subhalos of the main deflector. With that said, the constraints obtained in this work by including line of sight halos, at the level of $10^7 M_\odot$, are stronger by half an order of magnitude to one full order of magnitude over those obtained by Gilman et al. (2018).

The strength of the constraints on WDM models depend sensitively on the normalization of the subhalo mass function. This is partly due to the interpretation of the normalization

as scaling the information content per lens, and also due to the covariance between the normalization and the half-mode mass, although we stress that despite this covariance both parameters can be constrained independently. This highlights the importance of refining theoretical predictions for the value of the normalization, accounting for halo mass, redshift, and the destruction of subhalos by tidal stripping. To this end, observables from each lens system, such as the central velocity dispersion, half-light radius, redshift, etc. may be used to inform the prior on the normalization and thus further improve the inferred posterior with actual data.

The macromodel used to describe the mass profile of the main deflector plays a key role in this analysis. Several studies demonstrate that simple parameterizations sometimes fail to fit the flux ratios of substructure-less mass profiles, leading to 'artificial' flux ratio anomalies in the sense that they do not derive from dark matter substructure (Gilman et al., 2017; Hsueh et al., 2018). However, we note that these cases are dominated by the presence of undetected stellar disks, which are rare in the early-type galaxies that dominate the lensing cross section (Auger et al., 2010; Shankar et al., 2017). Also, we point out that identifying morphologically complexity in the main deflector and modeling it can remove these 'artificial' anomalies (Hsueh et al., 2016). While we do not explicitly account morphologically complex deflectors in this work, we do allow some freedom in the macromodel by marginalizing over the power law slope, and account for additional variations in the image fluxes as high as 8% that would result from marginalizing over additional macromodel parameters in the forward model.

Finally, we note that the formalism we present naturally accommodates other parameterizations of the halo mass function, and density profile for individual objects.

Acknowledgments

We thank Francis-Yan Cyr-Racine, Giulia Despali, Chuck Keeton, Stacy Kim, Alex Kusenko, Leonidas Moustakas, Annika Peter, and Simona Vegetti for helpful suggestions and interesting discussions throughout the course of this project. We also thank the anonymous referee

for comments that improved the quality of this work.

DG, TT, and SB acknowledge support by the US National Science Foundation through grant AST-1714953. DG, TT, SB and AN acknowledge support from HST-GO-15177. Support for Program number GO-15177 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. AN acknowledges support from the NASA Postdoctoral Program Fellowship.

This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by the UCLA Institute for Digital Research and Education's Research Technology Group. This work also used computational and storage services associated with the Aurora and Halo super computers. These resources were provided by funding from the JPL Office of the Chief Information Officer.

4.7 Appendix A: Implementing the two-halo term

The two-halo term describes an excess of matter (relative to the mean density of the universe) near a large halo, or a peak in the density field. It is evaluated using the software package COLOSSUS (Diemer, 2018), and takes the form

$$\xi_{\rm 2halo}\left(r, M, z\right) = b\left(M, z\right)\xi_{\rm lin}\left(r, z\right) \tag{4.14}$$

where $b\left(M, z\right)$ is the halo bias around a mass M, computed with the model presented by Tinker et al. (2010), and

$$\xi_{\rm lin}\left(r, z\right) = \frac{1}{2\pi^2}\int_0^\infty k^2 P(k, z)\frac{\sin(kr)}{kr}dk \tag{4.15}$$

is the linear matter-matter correlation function at a distance r. While in principle WDM free-streaming should affect the linear power spectrum $P\left(k, z\right)$, we do not model this effect.

We define a boost parameter β in terms of $\xi_{\rm 2halo}$ as

$$\beta\left(M, z\right) = \frac{2}{r_{\rm max} - r_{\rm min}}\int_{r_{\rm min}}^{r_{\rm max}} \xi_{\rm 2halo}\left(r', M, z\right)dr' \tag{4.16}$$

where M denotes the parent halo mass, and the factor of 2 accounts for symmetry around the parent halo. We choose $r_{\rm min} = 0.5$Mpc and $r_{\rm max} = 10$Mpc, which captures most of the contribution from the correlation function while omitting the contribution from regions inside the virial radius of the parent halo. Defining $A_0\,(z)$ as the normalization of the halo mass function in the lens plane closest to the main lens halo, we incorporate the two halo term by taking $A_0\,(z) \rightarrow (1+\beta) \times A_0\,(z)$, and add these halos at the main lens redshift.

In Figure 4.10 we plot the distribution of summary statistics $S_{\rm smooth}$ for a CDM mass function that includes the boost from the two-halo term, and one that does not. In both cases, we set to $\Sigma_{\rm sub} = 0$ to isolate the impact of the two-halo term. The lens and source redshifts are set at 0.6 and 2, respectively. The largest differences between the curves occurs at $S_{\rm smooth} \sim 0.2$, and is equal to 4%. We conclude that the contribution from $\xi_{\rm 2halo}$ is at most at the level of a few percent, although this may increase if a larger halo mass than $10^{13}{\rm M}_{\odot}$ is used to evaluate Equation 4.16.

4.8 Appendix B:The Born approximation in substructure lensing

The Born approximation computes the deflection at each subsequent plane along an unperturbed path. This speeds up lensing computations since a full backwards ray-tracing routine is not required. In Figure 4.11, we compare the distribution of flux ratio anomalies computed with respect to a smooth lens model (see the discussion in Section 4.3.2) using the Born approximation, and through full multi-plane ray-tracing. The difference between the solid and dotted curves in the figure, which represent flux ratios computed with and without the Born approximation, respectively, is comparable to the difference of WDM and CDM mass functions in Figure 4.3. Thus, we conclude that full multi-plane ray-tracing approach is required to accurately predict image flux ratios and probe dark matter on small scales.

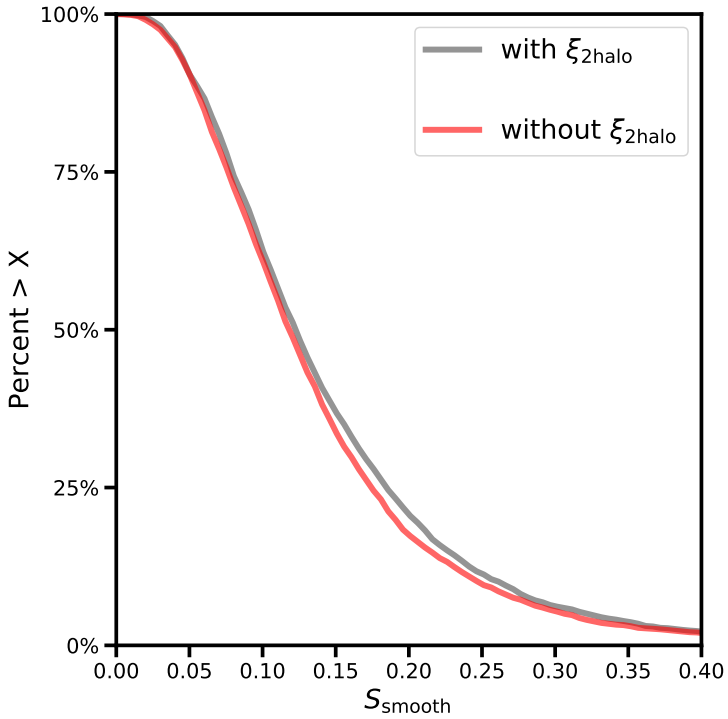

Figure 4.10: Distributions of the summary statistic $S_{\rm smooth}$, which represents the amount of flux ratio anomaly with respect to a smooth lens model (see the discussion in Section 4.3). The grey curve is computed with the two-halo contribution, and the red curve is computed without it. Both models include only line of sight halos to isolate the contribution from $\xi_{\rm 2halo}$. The largest difference between the curves, an offset of 4%, lies at $S_{\rm smooth} \sim 0.2$.

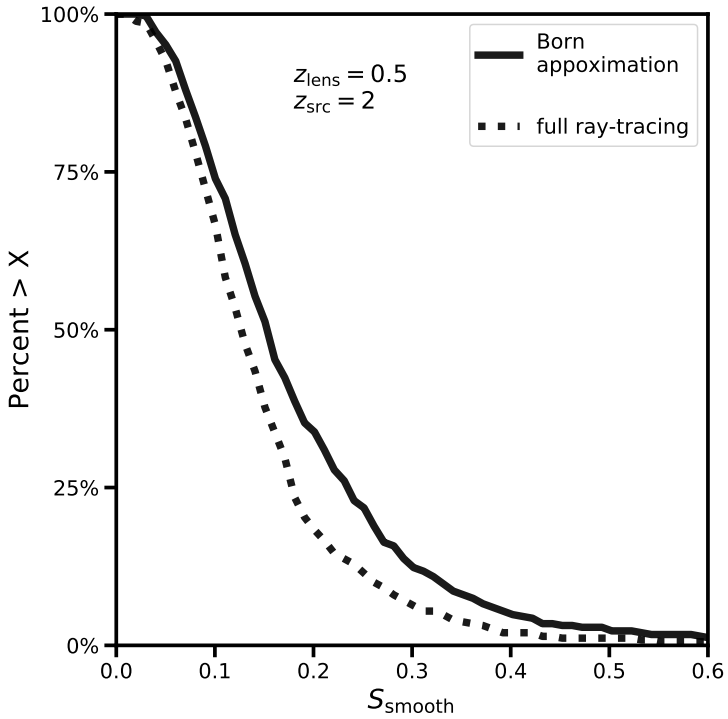

Figure 4.11: The two curves show distributions of summary statistics computed with respect to a smooth lens model. The curves are computed for the same CDM mass function, with and without the use of the Born approximation. The disagreement between the two curves suggests that the Born approximation does not predict image flux ratios accurately enough to differentiate between dark matter models.

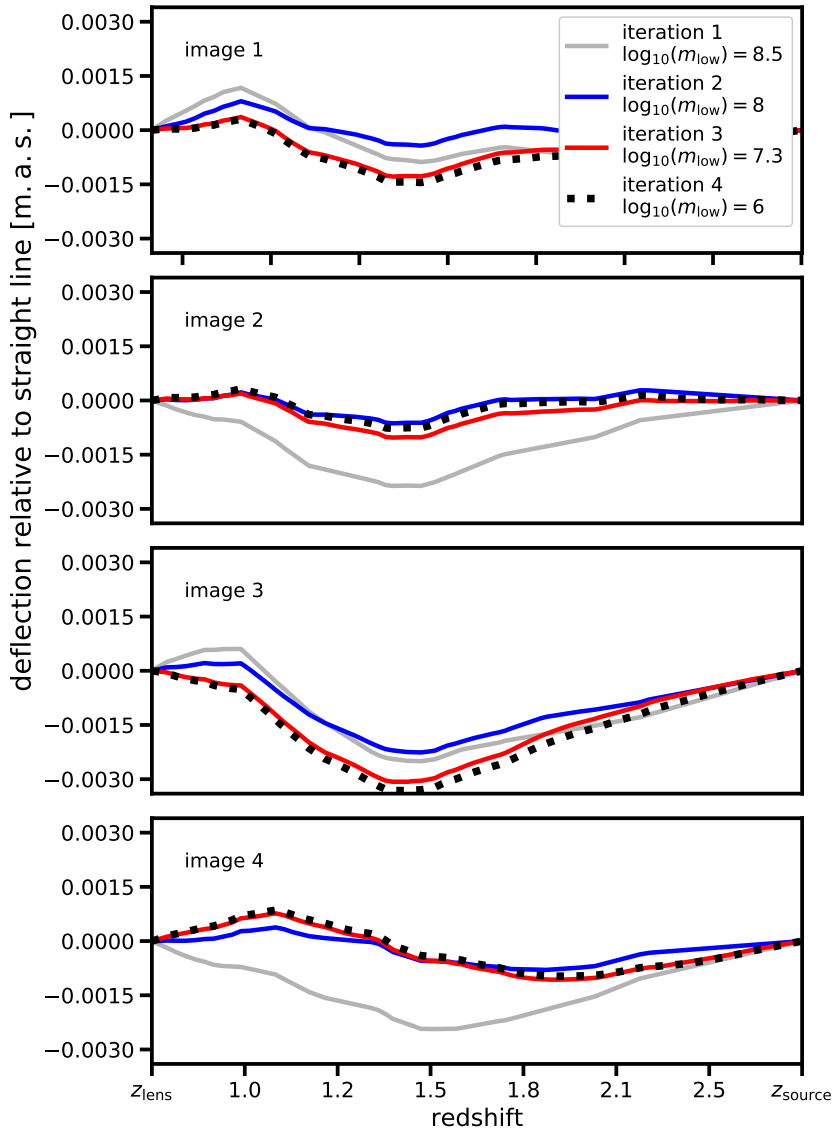

Figure 4.12: A visualization of the perturbative ray tracing algorithm we use to optimize lens models with potentially thousands of line of sight halos. The panels show the path through the background field relative to a straight line for multiple iterations of the algorithm, in which progressively smaller halos are rendered in progressively smaller apertures around the path of the light rays. This procedure speeds up optimizations of lens models with line of sight halos by at least an order of magnitude.

4.9 Appendix C: A fast algorithm for multi-plane lensing computations

For each observed lens, our forward modeling approach requires finding a set of macromodel parameters that cast the four light rays in a quadrupole image system to the same location in the source plane. For a single realization, this typically requires hundreds to thousands of backwards ray-tracing computations.

This task is computationally light for models with halos only in front of and at the same redshift as the main deflector because the path through the foreground field of halos is not coupled to the deflections produced in the main lens plane (owing to the recursive nature of Equation 5.3). Put differently, as soon as one specifies image positions on the sky and draws a realization of dark matter halos, the path through the foreground field is fully determined. In contrast, the path through the field of background halos is coupled to the deflections produced by the macromodel. The path through the background field therefore changes for each new proposal of macromodel parameters. This necessitates repeated computations of the potentially thousands of deflection angles of halos behind the main lens plane, which requires hundreds to thousands as many function evaluations as those needed in single plane lensing computations.

We address this computational challenge by implementing a perturbative approach to lens model optimizations. First, we optimize the macromodel to fit image positions with only foreground halos and main deflector subhalos present. We denote this optimized lens model $\vec{m}*$. This proceeds quickly, since the macromodel deflection angles are not coupled to those from foreground and main lens plane halos. Next, we add the largest background halos with $m > 10^8 \mathrm{M}_\odot$, and *re-optimize* $\vec{m}*$. Even though the deflections from these massive halos are coupled to those of the macromodel and need to be continuously re-evaluated during the optimization, since there are relatively few of them this proceeds fairly quickly. Next, we add halos in the range $10^{7.5} - 10^8 \mathrm{M}_\odot$, but only in 300 m.a.s. apertures around the path of the rays computed with respect to $\vec{m}*$. Since the area in which we render these smaller halos is relatively small, and since the macromodel solution $\vec{m}*$ is already close to the true

solution, this optimization also proceeds quickly. We iterate this process for progressively smaller halos until we reach $10^6 M_\odot$.

A visual representation of this process is presented in Figure 4.12, where we plot the path through the background halos relative to a straight line for subsequent iterations of the perturbative approach. After adding the $10^8 M_\odot$ background halos, the path through the background lens planes changes only slightly, which reflects the fact that these massive objects dominate the deflection field.

This procedure accomplishes the optimization of a macromodel with background halos 10-50 times faster than a naive optimization with all background halos included simultaneously. We test that the flux ratio statistics are identical to those obtained by ray tracing through full realizations without the perturbative approach implemented. We note that this algorithm is reminiscent of the Born approximation in that it initially neglects the presence of small deflections from subhalos along the line of sight, but differs fundamentally from the Born approximation in that the full non-linear coupling between every subhalo is eventually accounted for.

4.10 Appendix D: Convergence of posterior distributions

We approximate the true posterior distributions for model parameters by retaining the top 1,500 samples (ranked by their summary statistics) out of the 600,000 realizations computed per lens. To test whether this procedure yields an accurate approximation to the true posterior distribution, we appeal to a certain feature of Approximate Bayesian Computing algorithms, namely, that the approximation to the true posterior distribution converges as the number of samples increases. We can therefore test for convergence by applying the same cut on the top 1,500 samples to an 'under-sampled' model with only 400,000 realizations per lens, and check that the posterior distribution stays approximately fixed in place. We generate the sample of 400,000 by drawing the realizations randomly from the computed set of 600,000.

We perform this test and plot the results in Figure 4.13. While there is some movement

in the 1σ contours, the 2σ contours trace each other closely. Importantly the constraints on the half-mode mass are the same between the two inferences, which is the most important criterion for our purpose of forecasting bounds on dark matter warmth. Finally, we note that ABC routines tend to yield conservative approximations to the true posterior distributions, in the sense that with more samples the volume of the resulting posterior distribution shrinks. This explains why black contours (400,000 samples) tend to cover more area than the red contours (600,000 samples). As additional forward model samples improve the precision of the inference, the constraints we present would only improve by computing additional realizations.

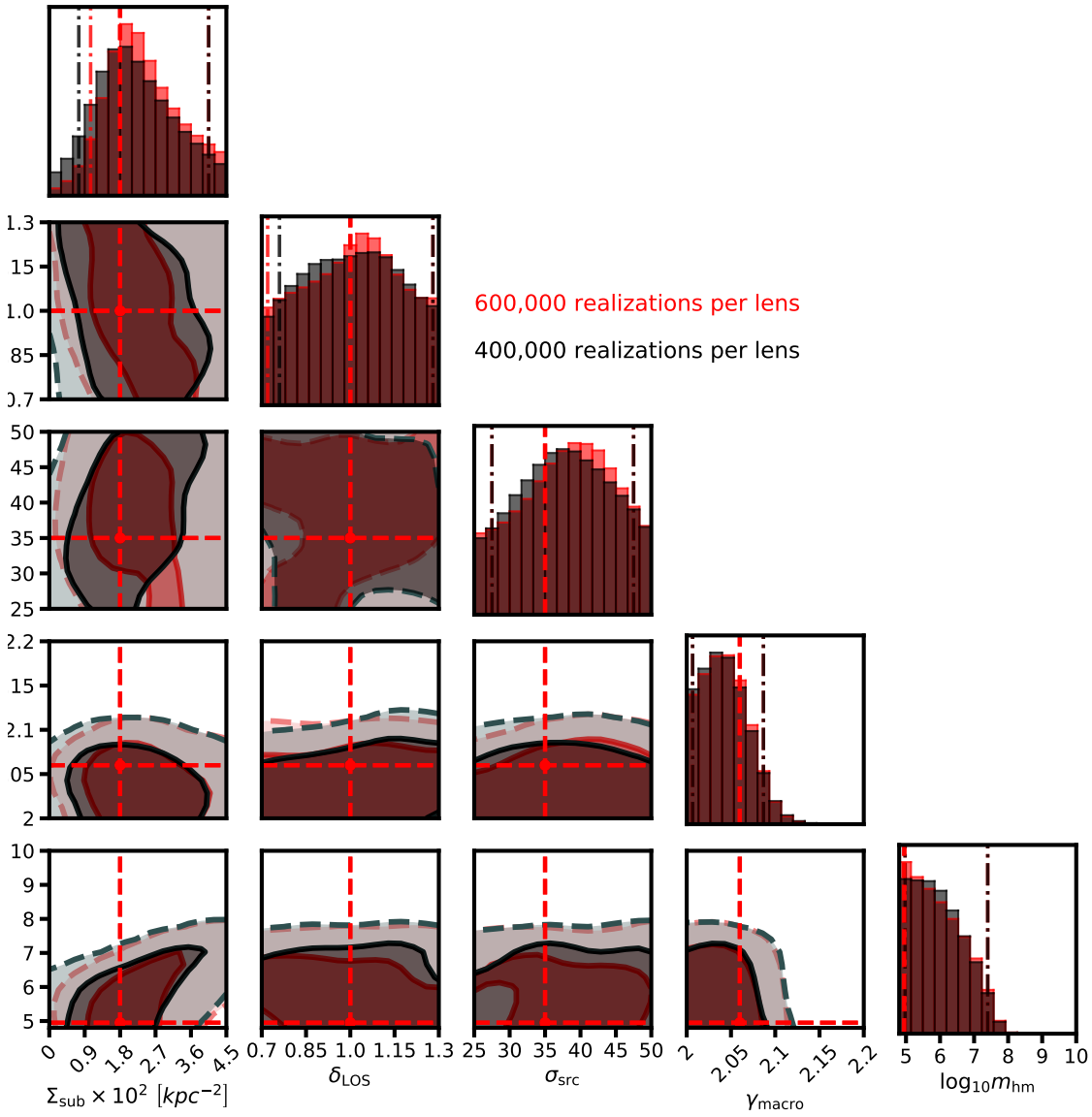

Figure 4.13: A convergence test for the forward model simulations. The overall agreement between the black and red distributions indicates that the posteriors we derive, and the numerical operations involved to produce them including the kernel density estimation, are robust to changes in the number of forward model samples per lens.

CHAPTER 5

Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with 8 quadruple-image strong gravitational lenses

This chapter was published as Gilman, D., et al. Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with 8 quadruple-image strong gravitational lenses. MNRAS 491, 6077-6101 (2020), and is printed here with minor formatting adjustments.

5.1 Introduction

The theory of cold dark matter (CDM) has withstood numerous tests on scales spanning individual galaxies to the large scale structure of the Universe and the cosmic microwave background (Tegmark et al., 2004; de Blok et al., 2008; Hinshaw et al., 2013). The next frontier for this highly successful theory lies on sub-galactic scales, where CDM makes two distinct predictions: First, CDM predicts a scale-free halo mass function, possibly down to halo masses comparable to that of a planet (Hofmann et al., 2001; Angulo et al., 2017). Second, in CDM models halo concentrations decrease monotonically with halo mass, a result of hierarchical structure formation (Moore et al., 1999; Avila-Reese et al., 2001; Zhao et al., 2003; Diemer & Joyce, 2019). A confirmation of these predictions through a measurement of the mass function and halo concentrations on mass scales below $10^9 M_\odot$ would at once constitute a resounding success for CDM and rule out entire classes of alternative dark matter

theories.

The abundance of small-scale dark matter depends on the matter power spectrum at early times. If the velocity distribution of the dark matter particles causes them to diffuse out of small peaks in the density field, this will prevent the direct collapse of over-densities below a characteristic scale referred to as the free-streaming length (Benson et al., 2013; Schneider et al., 2013). The delay in structure formation in these scenarios also suppresses the central densities of the smallest collapsed halos, changing the mass-concentration relation for low-mass objects (Avila-Reese et al., 2001; Schneider et al., 2012; Macciò et al., 2013; Bose et al., 2016; Ludlow et al., 2016). By definition, free-streaming effects are negligible in CDM, while models with cosmologically relevant free-streaming lengths are collectively referred to as warm dark matter (WDM). As the free-streaming length depends on the dark matter particle(s) mass and formation mechanism, an inference on the small-scale structure of dark matter on mass scales where some halos are expected to be completely dark directly constrains fundamental dark matter physics and the viability of specific WDM particle candidates, including sterile neutrinos (Dodelson & Widrow, 1994; Shi & Fuller, 1999; Abazajian & Kusenko, 2019) and keV-mass thermal relics.

Interest in alternatives to the canonical CDM paradigm, such as WDM, were motivated in part by apparent failures of the CDM model on small scales (see Bullock & Boylan-Kolchin, 2017, and references therein). Two challenges in particular dominate scientific discourse, and provide illustrative examples of the complexity associated with testing CDM's predictions on sub-galactic scales. The 'missing satellites problem' (MSP), first pointed out by Moore et al. (1999), refers to the paucity of observed satellite galaxies around the Milky Way, in stark contrast to dark-matter-only N-body simulations that predict hundreds of dark matter subhalos hosting a luminous satellite galaxy. Invoking free-streaming effects in WDM to remove these small subhalos would resolve the problem, and hence WDM models gained traction. A second challenge to the CDM picture emerged with the 'too big to fail' (TBTF) problem (Boylan-Kolchin et al., 2011), which points out that the subhalos housing the largest Milky Way satellites are either under-dense or too small. Self-interacting dark matter, which results in lower central densities in dark matter subhalos (see Tulin & Yu, 2018, and references

therein), gained traction in part as a resolution to the TBTF problem.

Today, new astrophysical solutions to the MSP and TBTF problems diminish the immediate threat to CDM, but the resolutions to these issues are riddled with assumptions regarding complicated physical processes on sub-galactic scales. The inclusion of baryonic feedback and tidal stripping in N-body simulations results in the destruction of subhalos, pushing the surviving number down to observed levels (Kim et al., 2017), although recently it has been suggested that the role of tidal stripping in N-body simulations is artificially exaggerated by resolution effects (van den Bosch et al., 2018; Errani & Peñarrubia, 2019). The continuous discovery of new dwarf galaxies seems to resolve the MSP, and might even suggest a 'too-many-satellites problem' (Kim et al., 2018; Homma et al., 2019), but the number of expected satellite galaxies in CDM itself rests on assumptions regarding the process of star formation in low mass halos, which can introduce uncertainties larger than the differences between CDM and WDM on these scales (Nierenberg et al., 2016; Dooley et al., 2017; Newton et al., 2018).The inclusion of baryonic feedback from star formation processes and supernova in low-mass halos can reduce halo central densities, and at least partially alleviates the issues associated with the TBTF problem (Tollet et al., 2016). However, the degree to which baryonic feedback resolves the problem depends on the manner in which this feedback is implemented in simulations.

Regarding constraints on WDM models, analysis of the Lyman-α forest (Viel et al., 2013; Iršič et al., 2017) and the luminosity function of distant galaxies (Menci et al., 2016; Castellano et al., 2019), while robust to the systematics associated with examining Milky Way satellites, to some degree rely on luminous matter to trace dark matter structure. Constraints from the Lyman-α forest also invokes certain assumptions for the relevant thermodynamics. The common theme is that disentangling the role of baryons and dark matter physics on sub-galactic scales is difficult and fraught with uncertainty. It would be ideal to test the predictions of matter theories irrespective of baryonic physics.

Strong gravitational lensing by galaxies provides a means of testing the predictions of dark matter theories directly, without relying on baryons to trace the dark matter. As photons emitted from distant background sources traverse the cosmos, they are subject to

deflections by the gravitational potential of dark matter halos along the entire line of sight and by subhalos around the a main lensing galaxy. Each warped image produced by a strong lens contains a wealth of information regarding the dark matter structure in the Universe. The aim of this work is to extract that information.

When the lensed background source is spatially extended – for example, a galaxy – the lensed image becomes an arc that partially encircles the main deflector. Dark matter halos near the arc produce small surface brightness distortions, which allows for the localization of the perturbing halo and enables constraints on its mass down to scales somewhere between $10^8 - 10^9 M_\odot$ (Vegetti et al., 2014; Hezaveh et al., 2016b). Analysis of the surface brightness fluctuations over the entirety of the arc can also constrain the abundance of small halos too diminutive to be detected individually, and results in a 2 keV lower bound on the mass of thermal relic WDM (Birrer et al., 2017b). A joint analysis of individual detections and non-detections in a sample of arc-lenses can constrain certain models of dark matter and test the predictions of CDM (Vegetti et al., 2018; Ritondale et al., 2018). Recently, several works have proposed measuring the substructure convergence power spectrum in by analyzing surface brightness fluctuations in extended arcs (Hezaveh et al., 2016a; Cyr-Racine et al., 2019; Díaz Rivero et al., 2018; Brennan et al., 2018), and Bayer et al. (2018) applied this method to a strong lens system.

We focus on a second kind of lens system, quadruply imaged quasars (quads). Rather than extended arcs, the observables in quads are four image positions and three magnification ratios, or flux ratios (the observable is the flux ratio, not the intrinsic flux, because the intrinsic source brightness is unknown) with unresolved sources. Flux ratios depend on non-linear combinations of second derivatives of the lensing potential near an image, providing localized probes of small-scale structure down to scales of $10^7 M_\odot$. These systems have been used in the past to constrain the presence of dark matter halos near lensed images (Metcalf & Madau, 2001; Metcalf & Zhao, 2002; Amara et al., 2006; Nierenberg et al., 2014, 2017) and measure the subhalo mass function (Dalal & Kochanek, 2002). Recently, Hsueh et al. (2019) improved on previous analyses of quadruply imaged quasars by including halos along the line of sight, which can contribute a significant signal in flux ratio perturbations (Xu

et al., 2012; Gilman et al., 2018). They found results consistent with CDM, ruling out WDM models to a degree comparable to that of the Lyman-α forest (Viel et al., 2013; Iršič et al., 2017).

In the case of quadruple-image lenses, the luminous source is often a compact background object, such as the ionized medium around a background quasar. Broad-line emission from the accretion disk is subject to microlensing by stars, whereas light that scatters off of the more spatially extended narrow-line region is immune to microlensing while retaining sensitivity to the milli-arcsecond scale deflection angles produced by dark matter halos in the range $10^7 - 10^{10} M_\odot$ (Moustakas & Metcalf, 2003; Sugai et al., 2007; Nierenberg et al., 2014, 2017). Likewise, radio emission from the background quasar, while generally expected to be more compact than the narrow-line emission based on certain quasar models (Elitzur & Shlosman, 2006; Combes et al., 2019), is extended enough to absorb micro-lensing effects.

We carry out an analysis of eight quads using a forward modeling approach we have tested and verified with mock data sets (Gilman et al., 2018, 2019). The sample of lenses we consider contains six systems with flux ratios measured with narrow-line emission presented in Nierenberg et al. (2020), and two others with data from Nierenberg et al. (2014) and Nierenberg et al. (2017). We expect the sample is robust to microlensing effects and yield reliable data with which to constrain dark matter models. None of the quads show evidence for morphological complexity in the form of stellar disks, which require more detailed lens modeling (Hsueh et al., 2016; Gilman et al., 2017; Hsueh et al., 2017).

This paper is organized as follows: In Section 5.2 we describe our forward modeling analysis method and our implementation of a rejection algorithm in Approximate Bayesian Computing. Section 5.3 describes our parameterizations for the dark matter structure in the main lens plane and along the line of sight, and our modeling of free-streaming effects in WDM. Section 6.3 contains a brief description of the data used in our analysis and the relevant references for each system. In Section 5.5 we describe in detail each physical assumption we make and the modeling choices and prior probabilities attached to these assumptions. In Section 6.4, we present our inferences on the free-streaming length of dark matter and the amount of lens plane substructure. We discuss the implications of our results

and our general conclusions in Section 5.7.

All lensing computations are performed using `lenstronomy`[1] (Birrer & Amara, 2018). Cosmological computations involving the halo mass function and the matter power spectrum are performed with `colossus` (Diemer, 2018). We assume a standard cosmology using the parameters from WMAP9 (Hinshaw et al., 2013) ($\Omega_m = 0.28, \sigma_8 = 0.82, h = 0.7$).

5.2 Bayesian inference in substructure lensing

In this section we frame the substructure lensing problem in a Bayesian context, and describe our analysis method which relies on a forward-generative model to sample the target posterior distribution through an implementation of Approximate Bayesian Computing. We have tested this analysis method using simulated data (Gilman et al., 2018, 2019). The full forward modeling procedure we describe in this section is illustrated in Figure 5.1, and the relevant parameters are summarized in Table 5.1.

5.2.1 The Bayesian inference problem

Our goal is to obtain samples from the posterior distribution

$$p\left(\mathbf{q}_{\rm s}|\boldsymbol{D}\right) \propto \pi\left(\mathbf{q}_{\rm s}\right) \prod_{n=1}^{N} \mathcal{L}\left(\mathbf{d}_{\rm n}|\mathbf{q}_{\rm s}\right) \tag{5.1}$$

where $\mathbf{q}_{\rm s}$ is a set of hyper-parameters describing the subhalo and line of sight halo mass functions, $\boldsymbol{D}$ denotes the set of positions and flux ratios from a set of N lenses with the data from each lens denoted by $\mathbf{d}_{\rm n}$, and where π represents the prior on $\mathbf{q}_{\rm s}$.

A certain dark matter model makes predictions for the parameters in $\mathbf{q}_{\rm s}$, which includes quantities such as the normalization of the subhalo mass function, the logarithmic slope of the mass function, a free streaming cutoff, etc. For a given $\mathbf{q}_{\rm s}$, we may generate specific realizations of line of sight halos and main deflector subhalos (including the halo/subhalo masses, positions, concentrations, etc.), that affect lensing observables. We refer to a specific

[1] https://github.com/sibirrer/lenstronomy

realization of dark matter structure corresponding to a model specified by $\mathbf{q}_{\rm s}$ as $\mathbf{m}_{\rm sub}$. In addition to generating the realizations $\mathbf{m}_{\rm sub}$, computing the likelihood function $\mathcal{L}\left(\mathbf{d}_{\rm n}|\mathbf{q}_{\rm s}\right)$ in Equation 5.1 requires marginalizing over nuisance parameters $\mathbf{M}$, which include the background source size $\sigma_{\rm src}$, and the lens model that describes the main lensing galaxy (hereafter the macromodel). Integrating over the macromodel and the space of possible dark matter realizations $\mathbf{m}_{\rm sub}$, the likelihood is given by

$$\mathcal{L}\left(\mathbf{d}_{\rm n}|\mathbf{q}_{\rm s}\right) = \int p\left(\mathbf{d}_{\rm n}|\mathbf{m}_{\rm sub}, \mathbf{M}\right) p\left(\mathbf{m}_{\rm sub}, \mathbf{M}|\ \mathbf{q}_{\rm s}\right) d\mathbf{m}_{\rm sub}\ d\mathbf{M}. \tag{5.2}$$

Note that we write the joint distribution $p\left(\mathbf{m}_{\rm sub}, \mathbf{M}|\mathbf{q}_{\rm s}\right)$, and do not assume the parameters in $\mathbf{M}$ and $\mathbf{q}_{\rm s}$ are independent.

Evaluating Equation 5.2 is a daunting task. We highlight two main reasons:

- Exploring the parameter space spanned by $\mathbf{q}_{\rm s}$ and $\mathbf{M}$ through traditional MCMC methods is extremely inefficient. $\mathbf{M}$ is a high-dimensional space, where the overwhelming majority of volume does not result in model-predicted observables that resemble the data, and in particular does not predict the correct image positions. Thus the overwhelming majority of samples drawn from $\mathbf{M}$, and the corresponding samples $\mathbf{q}_{\rm s}$ (even if they described the 'true' nature of dark matter) would not contribute to the integral.

- The parameters $\mathbf{M}$ describing the lens macromodel may depend indirectly on the dark matter parameters $\mathbf{q}_{\rm s}$ through the realizations $\mathbf{m}_{\rm sub}$ generated from the model specified by $\mathbf{q}_{\rm s}$. This necessitates the simultaneous sampling of $\mathbf{q}_{\rm s}$ and $\mathbf{M}$ in the inference. However, it is difficult to impose an informative prior on $\mathbf{M}$ since the 'true' parameters in $\mathbf{q}_{\rm s}$ are unknown. Recognizing this and using a very uninformative prior on $\mathbf{M}$, most samples will be rejected since they do not resemble the data, which alludes back to the issue of dimensionality described in the first bullet point.

To address these challenges, we use a statistical method that bypasses the direct computation of the integral in Equation 5.2.

5.2.2 Forward modeling the data

Rather than compute the likelihood function, we recognize that by creating simulated observables $\mathbf{d}'_{\rm n} = \mathbf{d}'_{\rm n}(\mathbf{m}_{\rm sub}, \mathbf{M})$ from the model $\mathbf{q}_{\rm s}$, and accepting the proposed $\mathbf{q}_{\rm s}$ if they satisfy $\mathbf{d}'_{\rm n} = \mathbf{d}_{\rm n}$, the accepted $\mathbf{q}_{\rm s}$ samples will be direct draws from the posterior distribution in Equation 5.1 (Rubin, 1984). In this forward-generative framework, simulating the relevant physics in substructure lensing replaces the task of evaluating the likelihood function in Equation 5.2. We propagate photons from a finite-size background source through lines of sight populated by dark matter halos, a lensing galaxy and its subhalos, and finally into a simulated observation with statistical measurement errors added. Provided the forward model contains all of the relevant physics, the simulated data $\mathbf{d}'_{\rm n}$ will express the same potentially complex covariances present in the observed data.

The 'curse of dimensionality' that prohibits direct evaluation of Equation 5.2 also afflicts the criterion of exact matching between $\mathbf{d}_{\rm n}$ and $\mathbf{d}'_{\rm n}$. In particular, most draws of macromodel parameters $\mathbf{M}$ will not yield the observed image positions, and would therefore be rejected from the posterior. To deal with this, our strategy will be to ensure that the macromodel and other nuisance parameters sampled in the forward model, when combined with the full line of sight and subhalo populations specified by $\mathbf{m}_{\rm sub}$, yield a lens model that predicts the same image positions as observed in the data.

Obtaining a lens model that returns the observed image positions amounts to demanding that the the four images seen by the observer on the sky at positions $\boldsymbol{\theta}$ map to the same position on the source plane $\boldsymbol{\beta_K}$. This requires the use of the full multi-plane lens equation describing the path of deflected light rays (e.g. Schneider, 1997)

$$\boldsymbol{\beta_K} = \boldsymbol{\theta} - \frac{1}{D_{\rm s}} \sum_{k=1}^{K-1} D_{\rm ks} \boldsymbol{\alpha}_{\mathbf{k}} \left(D_{\rm k} \boldsymbol{\beta}_{\mathbf{k}} \right), \tag{5.3}$$

where the quantities $D_{\rm s}$, $D_{\rm k}$ and $D_{\rm ks}$ denote angular diameter distances to the source plane, to the kth lens plane, and from the kth lens plane to the source plane, respectively. Equation 5.3 is a recursive equation for the $\boldsymbol{\beta}_{\mathbf{k}}$ that couples deflection angles from objects at different redshifts, similar to looking through potentially thousands of magnifying glasses in series.

Throughout this process, we account for uncertainties in the measured image positions by sampling astrometric perturbations δ_{xy}, and applying them to the observed image positions during the forward modeling.

To solve for macromodel parameters $\mathbf{M}$, for each realization $\mathbf{m}_{\rm sub}$ we sample the power-law slope of the main deflector mass profile $\gamma_{\rm macro}$ and the external shear strength $\gamma_{\rm ext}$. If the lens system in question has satellite galaxies or nearby deflectors, we sample priors for their masses and positions. The remaining parameters describing the lens macromodel[2] are allowed to vary freely until a lens model that fits the image positions is found[3].

The approach of simultaneously sampling $\mathbf{M}$ and $\mathbf{q}_{\rm s}$ does not involve lens model optimizations with respect to the observed image fluxes, because the information from the observed fluxes is not used at this stage of the analysis. This method therefore avoids potential biases incurred by optimizing the macromodel with respect to the observed fluxes, rather than marginalizing over these parameters. As we will show in Section 5.6.1, by sampling $\mathbf{M}$ and $\mathbf{q}_{\rm s}$ simultaneously we obtain joint posterior distributions that account for potential covariance between these quantities, recognizing that the addition of substructure may affect the distributions for the macromodel parameters in $\mathbf{M}$.

With a lens model that fits the image positions in hand, we draw a background source size and ray-trace on a finely sampled grid around each image position using Equation 5.3 to compute the image fluxes $\boldsymbol{f'}$. To incorporate statistical measurement errors in image fluxes, we sample flux uncertainties $\boldsymbol{\delta f}$, and render these perturbations onto the model-predicted fluxes $\boldsymbol{f'} \rightarrow \boldsymbol{f'} + \boldsymbol{\delta f}$ prior to computing the flux ratios.

[2]The full set of macromodel parameters for a power-law ellipsoid are the overall normalization $b_{\rm macro}$, the mass centroid g_x and g_y, the ellipticity and ellipticity position angle ϵ and θ_ϵ, the external shear and shear angle $\gamma_{\rm ext}$ and $\theta_{\rm ext}$, and the power-law slope $\gamma_{\rm macro}$. Nearby galaxies are modeled as Singular Isothermal Spheres.

[3]The four image positions provide $4 \times 2 = 8$ constraints, and the macromodel parameters that are allowed to vary freely, plus the source position, give 8 degrees of freedom.

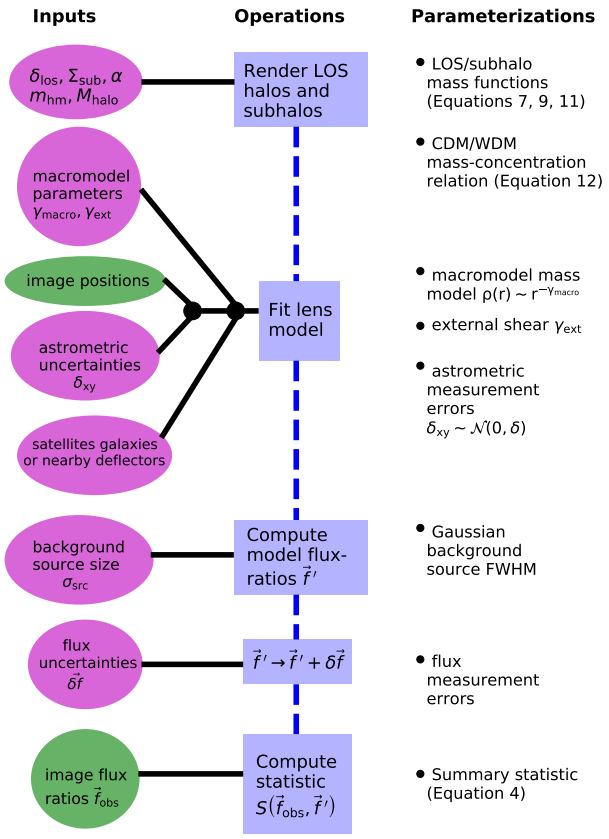

Figure 5.1: A graphical representation of the forward modeling procedure. Purple colors correspond to the action of sampling from a prior, blue represents an operation performed using the parameters sampled from a prior, and green colors indicate the use of observed information from the lenses. The arrow of time points from top to bottom: The first step is the rendering of dark matter structure, while the use of the information from observed flux ratios happens only at the very end.

5.2.3 Deriving posteriors from the forward model samples

For each realization, we compute a summary statistic between the three observed flux ratios $\boldsymbol{f_{\rm obs}}$ and those computed in the forward model

$$S_{\rm lens}\left(\boldsymbol{f'}, \boldsymbol{f_{\rm obs}}\right) \equiv \sqrt{\sum_{i=1}^{3}\left(f'_i - f_{\rm obs(i)}\right)^2}, \tag{5.4}$$

and assign this statistic to the draw of $\mathbf{q_s}$. This summary statistic contains the full information content of the data, as the simultaneous matching of the three ratios requires that the forward model samples that minimize this statistic contain the same correlations present in the data. We repeat this procedure between 300,000 and 1,200,000 times for each quad, depending on with frequency with which the realizations, with the statistical flux uncertainties added, match the observed fluxes to within 1%.

We select the $\mathbf{q_s}$ parameters corresponding to the 800 lowest summary statistics $S_{\rm lens}$. The exact matching criterion $\mathbf{d_n} = \mathbf{d'_n}$, which guarantees that the accepted samples $\mathbf{q_s}$ form the desired posterior, is replaced by selecting the realizations that look most like the data through the summary statistic $S_{\rm lens}$. The resulting distribution of $\mathbf{q_s}$ is therefore an approximation to the posterior distribution for each lens, with the approximation converging to the true posterior as the number of forward model samples increases while keeping the number of accepted samples fixed. The quality of the approximation can be quantified through a convergence test, in which we verify that the posteriors are unchanged as one removes realizations from the forward-modeled data while keeping the same number of accepted samples (see Appendix 5.8). This method is an implementation of a rejection algorithm in Approximate Bayesian Computing (Rubin, 1984; Marin et al., 2011; Lintusaari et al., 2017), a technique applied to problems where it is possible to generate simulated data from the model, but difficult to compute the likelihood (see also Beaumont et al., 2002; Akeret et al., 2015; Birrer et al., 2017b; Hahn et al., 2017).

To obtain the final posterior distribution $p\left(\mathbf{q_s}|\boldsymbol{D}\right)$ (Equation 5.1), we multiply together the likelihoods obtained for each lens [4]. This procedure is only possible when using uniform

[4]Before taking the product, we use a Gaussian kernel density estimator (KDE) with a first order boundary

priors in the forward model sampling, as the use of non-uniform priors would effectively move $\pi(\mathbf{q}_s)$ inside the product in Equation 5.1 and over-use this information. We may, however, impose any prior we wish a-posteriori by re-weighting the forward model samples accordingly.

5.3 The subhalo and line of sight halo populations

In this section, we describe the models we implement for the line of sight and subhalo mass functions in cold and warm dark matter that we sample in the forward model. We also describe the density profiles for individual halos, including their truncation radii and their distribution both along the line of sight and in the main lens plane. We begin with the parameterizations used for the halo and subhalo density profiles and the spatial distribution of subhalos in Section 5.3.1. In Sections 5.3.2 and 5.3.3 we describe the parameterizations of the subhalo and line of sight halo functions, respectively, and in Section 5.3.4 describe how we model WDM free-streaming effects.

5.3.1 Subhalo density profiles and spatial distribution

We model subhalos as tidally truncated NFW profiles (Baltz et al., 2009)

$$\rho(r) = \frac{\rho_s}{x(1+x)^2}\frac{\tau^2}{x^2+\tau^2} \tag{5.5}$$

where $x = \frac{r}{r_s}$, $\tau = \frac{r_t}{r_s}$, and r_t is a truncation radius and r_s is the NFW profile scale radius. We use the mass definition of M_{200} computed with respect to the critical density at $z = 0$, and a concentration mass relation that accounts for free-streaming effects in WDM as is specifically designed to accurately predict the concentrations of low-mass halos (see Section 5.3.4).

In the main lens plane, we truncate halos according to their three-dimensional position inside the host halo r_{3D} through a Roche-limit approximation that assumes a roughly

correction (e.g. Lewis, 2019) to obtain a continuous approximation of the likelihood for each lens. We compute the bandwidth according to Scott's factor (Scott, 1992), but caution that care should be taken with the choice of bandwidth to avoid over or under smoothing the likelihood.

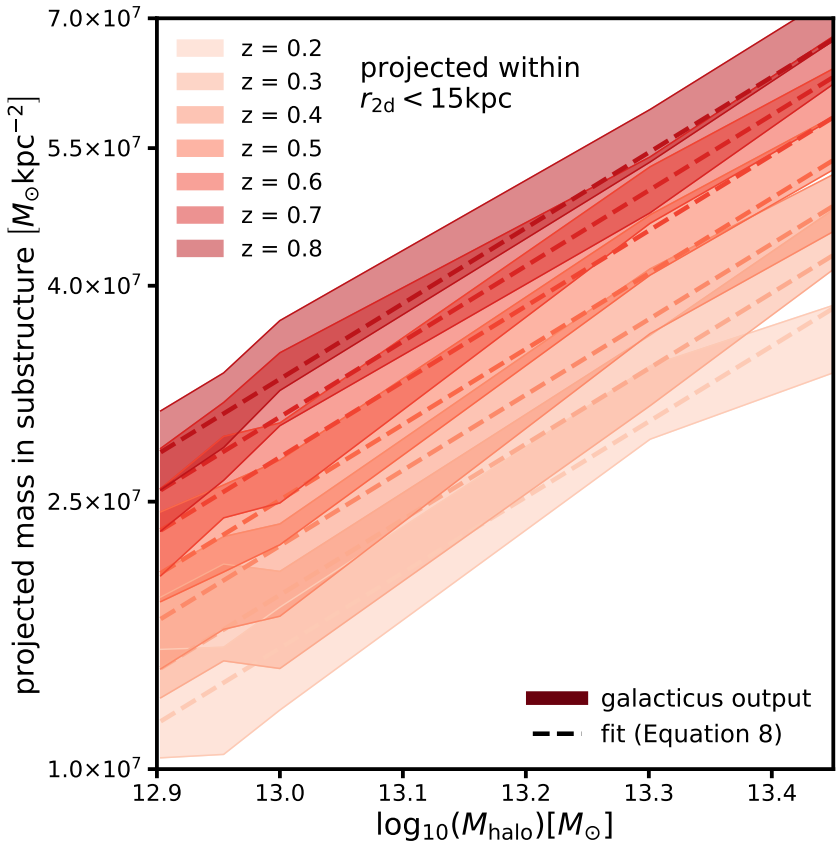

Figure 5.2: Output from the `galacticus` semi-analytic simulations of substructure within halos used to calibrate the evolution of the subhalo mass function with halo mass and redshift. While on the y-axis we plot the actual projected surface mass density in substructure output by `galacticus`, we only use the scaling with halo mass in redshift in our modeling, treating the overall normalization of the subhalo mass function as a free parameter. The projected mass density in substructure on the y-axis corresponds to a mass range $10^6 - 10^{10} M_\odot$, where we have extrapolated the mass function from the smallest resolved subhalo ($10^8 M_\odot$) to $10^6 M_\odot$ to compute the projected mass.

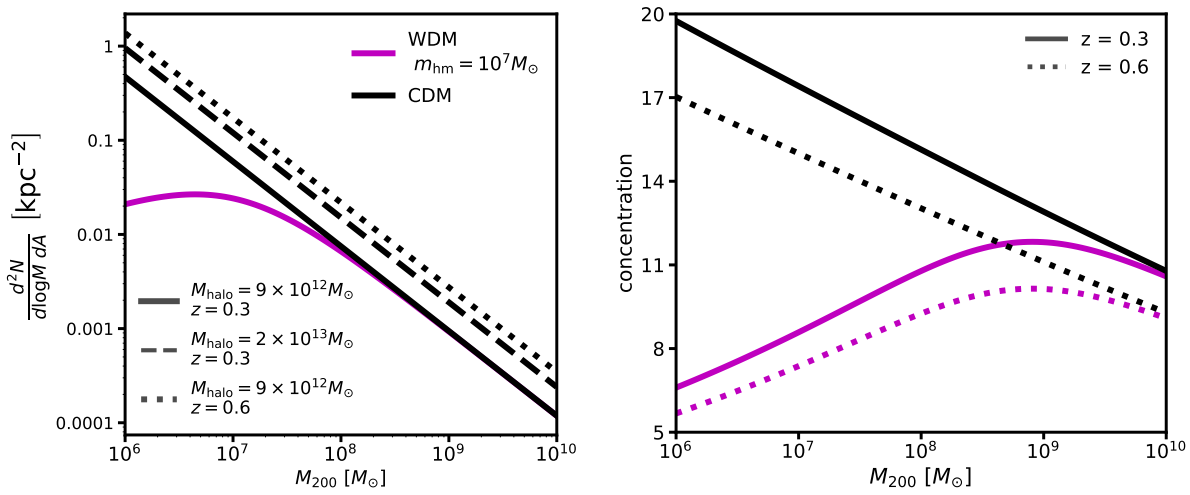

Figure 5.3: **Top:** The subhalo mass function as a function of halo mass, redshift, and the half-mode mass $m_{\rm hm} = 10^7 M_\odot$ with $\Sigma_{\rm sub} = 0.012 {\rm kpc}^{-2}$. The line of sight halo mass function looks similar, but evolves differently with redshift. **Bottom:** The mass-concentration-relation for CDM and the same WDM model with $m_{\rm hm} = 10^7 M_\odot$. Free-streaming affects the concentration of halos over one order of magnitude above $m_{\rm hm}$.

isothermal global mass profile. The relevant scaling is $r_t \propto (M_{200} r_{\rm 3D}^2)^{\frac{1}{3}}$ (Tormen et al., 1998; Cyr-Racine et al., 2016), which we implement as

$$r_t = 1.4 \left(\frac{M_{200}}{10^7 M_\odot} \right)^{\frac{1}{3}} \left(\frac{r_{\rm 3D}}{50 {\rm kpc}} \right)^{\frac{2}{3}} [{\rm kpc}] \,. \tag{5.6}$$

This results in truncation radii of $\sim 4 - 10 r_s$. We note that the truncation radius depends implicitly on the host halo mass $M_{\rm halo}$ through $r_{\rm 3D}$, which depends on the scale radius and the virial radius of the host halo at the lens redshift (see Figure 5.4). We note that the definition of r_t in Equation 5.6 does not depend on the structural parameters of the subhalo, which are altered in WDM models (see Section 5.3.4). Incorporating these modeling details requires prescriptions for the tidal evolution of subhalos in the host halo as a function of the physical properties of the subhalo at infall (e.g. Green & van den Bosch, 2019).

We render subhalos out to a maximum projected radius $3 R_{\rm Ein}$ and assign a three-dimensional z-coordinate between $-r_{200}$ and r_{200}, where r_{200} is the virial radius of the host. Inside this volume, we distribute the subhalos assuming the spatial distribution follows the

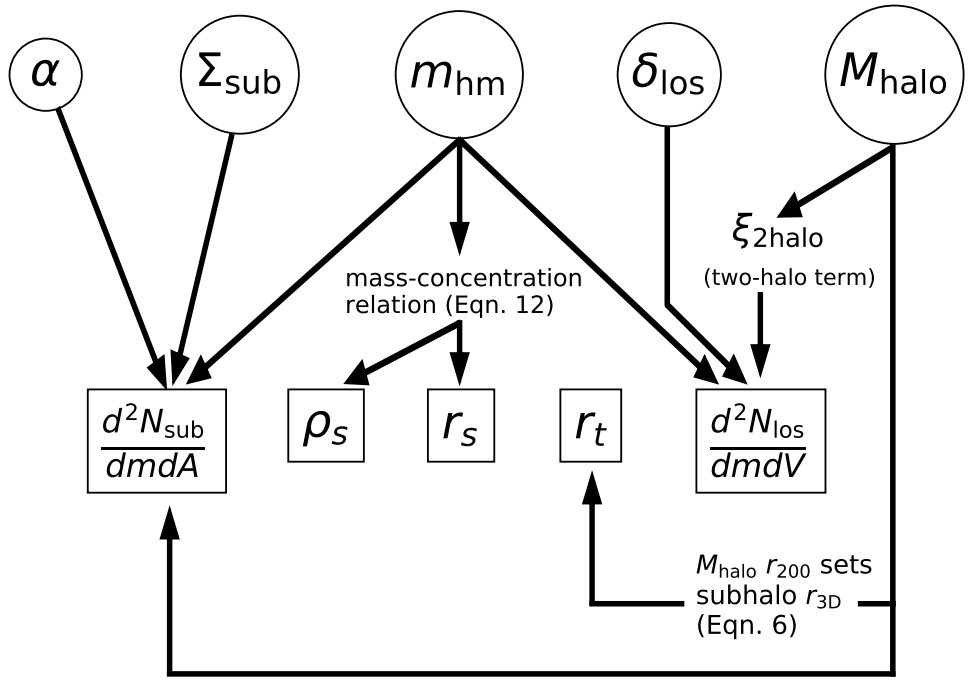

Figure 5.4: Summary of model parameter inter-dependency

A graphical representation of the dark matter parameters in $\mathbf{q}_s$: α, the logarithmic slope of the subhalo mass function, $\Sigma_{\rm sub}$, the overall scaling of the subhalo mass function, $m_{\rm hm}$, the WDM half-mode mass, $\delta_{\rm los}$, the overall factor for the line of sight halo mass function, and $M_{\rm halo}$, the main deflector's parent halo mass. $\xi_{\rm 2halo}$ is implemented through Equation 5.9 (see Section 5.3.3). These parameters are linked to the physical dark matter quantities they affect. From left to right: the subhalo mass function $\frac{d^2N}{dmdA}$, the normalization ρ_s, scale radius r_s, and truncation radius r_t of individual halos (see Equation 5.5), and the line of sight halo mass function $\frac{d^2N}{dmdV}$. The priors for each of these parameters are summarized in Table 5.2, and discussed at length in Section 5.5.

Table 5.1: Free parameters sampled in the forward model. Notation $\mathcal{N}(\mu, \sigma)$ indicates a Gaussian prior with mean μ and variance σ, and $\mathcal{U}(u_1, u_2)$ indicates a uniform prior between u_1 and u_2. Lens-specific priors are summarized in Table 5.2.

parameter	definition	prior
$\log_{10}(M_{\rm halo})\,[M_\odot]$	main lens parent halo mass	(lens specific)
$\Sigma_{\rm sub}\,[{\rm kpc}^{-2}]$	normalization of subhalo mass function (Equation 6.5)	$\mathcal{U}$ (0, 0.1)
	(rendered between $10^6 - 10^{10} M_\odot$)	
α	logarithmic slope of the subhalo mass function	$\mathcal{U}$ (-1.95, -1.85)
$\log_{10}(m_{\rm hm})\,[M_\odot]$	half-mode mass (Equations 5.11 and 5.12)	$\mathcal{U}$ (4.8, 10)
	$\propto$ to free streaming length and thermal relic mass $m_{\rm DM}$	
$\delta_{\rm los}$	rescaling factor for the line of sight Sheth-Tormen	$\mathcal{U}$(0.8, 1.2)
	mass function (Equation 5.9, rendered between $10^6 - 10^{10} M_\odot$)	
$\sigma_{\rm src}\,[{\rm pc}]$	source size	$\mathcal{U}$ (25, 60)
	parameterized as FWHM of a Gaussian	
$\gamma_{\rm macro}$	logarithmic slope of main deflector mass model	$\mathcal{U}$ (1.95, 2.2)
$\gamma_{\rm ext}$	external shear in the main lens plane	(lens specific)
$\delta_{xy}\,[{\rm m.a.s.}]$	image position uncertainties	(lens specific)
δf	image flux uncertainties	(lens specific)

mass profile of the host dark matter halo outside an inner tidal radius, which we fix to half the scale radius of the host. Inside this radius, we distribute subhalos with a uniform distribution in three dimensions. This choice is motivated by simulations that predict tidal disruption of subhalos near the lensing galaxy, resulting in an approximately uniform number of subhalos per unit volume in the inner regions of the halo (Jiang & van den Bosch, 2017). The spatial distribution of subhalos that results from this procedure is approximately uniform in projection, which agrees with the predictions from N-body simulations (Xu et al., 2015).

5.3.2 The CDM subhalo mass function

In principle, the projected mass in subhalos near the Einstein radius can depend on the host halo mass, redshift, and the severity of tidal stripping by the main lensing galaxy. We will ultimately combine the inferences from multiple lenses at different redshifts and with different host halo masses, so we parameterize the subhalo mass function in such a way that a single parameter $\Sigma_{\rm sub}$ can be used to simultaneously describe the projected mass density in substructure for each quad, regardless of halo mass or redshift.

We use the functional form for the subhalo mass function

$$\frac{d^2N_{\rm sub}}{dmdA} = \frac{\Sigma_{\rm sub}}{m_0}\left(\frac{m}{m_0}\right)^{\alpha}\mathcal{F}\left(M_{\rm halo}, z\right) \tag{5.7}$$

where scaling function $\mathcal{F}\left(M_{\rm halo}, z\right)$ encodes the differential evolution of the projected number density with redshift and host halo mass, such that $\Sigma_{\rm sub}$ can be interpreted as a common parameter for all the lenses. We choose the normalization such that $\mathcal{F}\left(\mathcal{M}_{\rm halo} = \infty\prime^{\infty\ni}\mathcal{M}_{\odot}, \ddagger = \prime.\bigtriangledown\right) = \infty$, anchoring $\Sigma_{\rm sub}$ at $z = 0.5$ with a halo mass of $10^{13}M_{\odot}$. We use a pivot mass $m_0 = 10^8 M_{\odot}$. We will marginalize over $\Sigma_{\rm sub}$ and α when quoting constraints on dark matter warmth to account for tidal stripping of subhalos and halo-to-halo scatter.

To determine the scaling function $\mathcal{F}\left(\mathcal{M}_{\rm halo}, \ddagger\right)$, we run a suite of simulations using the semi-analytic modeling code `galacticus`[5] (Benson, 2012; Pullen et al., 2014), simulating host halos and their substructure in the redshift range $0.2 < z < 0.8$ and mass range $0.8 - 3 \times 10^{13}M_{\odot}$, with a subhalo mass resolution of $10^8 M_{\odot}$. In each redshift and mass bin we simulate 24 halos, resulting in 840 halos with $M_{\rm halo} \sim 10^{13}M_{\odot}$ in total[6]. We average over the projected number densities along each principle axis inside a 15 kpc aperture to obtain trends in the projected number density with host halo mass and redshift in the vicinity of the Einstein radius, where lensed images appear. The `galacticus` simulations include tidal destruction of subhalos by the global dark matter mass profile, which affects the evolution of the projected mass density with host halo redshift: at early times, subhalos are more

[5]Code version 7175:2bd6b8d84a39

[6]The entire simulation suite using `galacticus` completed in 1,000 CPU hours.

concentrated in the host, while at later times tidal stripping from the host depletes the population of subhalos at small radii and the projected number density near the Einstein radius decreases. In addition, the physical size of the host halo at higher redshifts is smaller by a factor of $(1+z)^{-1}$, so the number of subhalos per square physical kpc is higher. We also note that early-type galaxy host halos simulated by Fiacconi et al. (2016) also show significant evolution with redshift in the projected number density of subhalos by about a factor of two, very similar to the `galacticus` predictions.

We fit the evolution with halo mass and redshift predicted by `galacticus` with the relation

$$\log_{10}(\mathcal{F}) = k_1 \log_{10}\left(\frac{M_{\rm halo}}{10^{13} M_\odot}\right) + k_2 \log_{10}(z+0.5) \tag{5.8}$$

with $k_1 = 0.88$ and $k_2 = 1.7$. The `galacticus` output and the fit from Equation 6.6 are shown in Figure 5.2. We only extract information regarding the scaling of projected mass density with halo mass and redshift from the `galacticus` simulations, and treat the overall normalization of the number density as a free-parameter that absorbs the effects of tidal destruction of subhalos by the main lens galaxy. We discuss our modeling assumptions in more detail in Section 5.5.4.

5.3.3 The line of sight halo mass function

We model line of sight structure by drawing halo masses from the Sheth-Tormen halo mass function (Sheth et al., 2001), with two modifications. First, we introduce an overall rescaling factor $\delta_{\rm los}$ which accounts for theoretical uncertainty in the predicted amplitude of the halo mass function (see e.g. Despali et al., 2016). The factor $\delta_{\rm los}$ accounts for the possibility of a selection bias in the quads towards systematically over or under-dense lines of sight. The second modification we add is a contribution from the two-halo term $\xi_{\rm 2halo}(M_{\rm halo}, z)$, which accounts for the presence of correlated structure in the vicinity of main deflector parent dark

matter halo [7]. With these modifications the line of sight halo mass function takes the form

$$\frac{d^2N_{\rm los}}{dmdV} = \delta_{\rm los}\big(1 + \xi_{\rm 2halo}\left(M_{\rm halo}, z\right)\big)\frac{d^2N}{dmdV}\Big|_{\rm ShethTormen}. \tag{5.9}$$

Halos along the line of sight are rendered in a double-cone geometry with opening angle $3R_{\rm Ein}$, where $R_{\rm Ein}$ is the Einstein radius of the main deflector, and a closing angle behind the main deflector such that the cone closes at the source redshift. Finally, we add negative convergence sheets to subtract the mean expected convergence from line of sight halos at each line of sight plane. Without this numerical procedure, lines of sight are systematically overdense relative to the expected matter density of the Universe, akin to lensing in a universe with positive curvature (Birrer et al., 2017a). This may bias results as the macromodel will attempt to compensate for the artificial focusing of light rays in this scenario.

5.3.4 Modeling free-streaming effects in WDM

Free-streaming refers to the diffusion of dark matter particles out of small peaks in the matter density field in the early Universe. This has the effect of erasing structure on scales below a characteristic free-streaming length which depends on the velocity distribution of the dark matter particles, and hence on their mass and formation mechanism. For a more in-depth discussion, see Schneider et al. (2013).

It is convenient to express free-streaming effects in terms of the half-mode mass $m_{\rm hm}$, which is defined in terms of the length scale where the transfer function between the CDM and WDM power spectra drops to one-half. In the specific case that all of the dark matter exists in the form of thermal relics, a one-to-one mapping between the half-mode mass and the mass of the candidate particle $m_{\rm DM}$ exists, and has the scaling $m_{\rm hm} \propto m_{\rm DM}^{-3.33}$ (Schneider et al., 2012) (see also Blandford & Narayan (1986))

$$m_{\rm hm}\left(m_{\rm DM}\right) = 3 \times 10^8 \left(\frac{m_{\rm DM}}{3.3{\rm keV}}\right)^{-3.33} M_\odot. \tag{5.10}$$

[7] In Appendix A of Gilman et al. (2019), we describe how this effect is implemented and show that this term contributes a $\sim 4\%$ increase in the frequency of flux ratio perturbations induced by objects outside the virial radius of the main deflector.

We have run `galacticus` models (Benson et al., 2013) with WDM mass functions corresponding to 3.3 and 5 keV thermal relics to investigate the effects of free-streaming on the trends with host halo mass and redshift of the projected mass in substructure near the Einstein radius, and determine that the fit in Equation 6.6 is common to both CDM and WDM. We therefore use the same scaling function $\mathcal{F}\left(\mathcal{M}_{\rm halo}, \ddagger\right)$ for WDM subhalo mass functions, and model the effects of free streaming using the fitting formula from (Lovell et al., 2014)

$$\frac{dN_{\rm WDM}}{dm} = \frac{dN_{\rm CDM}}{dm}\left(1 + \frac{m_{\rm hm}}{m}\right)^{-1.3}. \tag{5.11}$$

Since the parameter $m_{\rm hm}$ is related to the WDM transfer function, it should affect the subhalo and field halo mass functions in a similar manner. We therefore apply the same suppression factor in Equation 5.11 to both the subhalo mass function and the line of sight halo mass function in Equations 6.5 and 5.9, respectively. Lacking a theoretical prediction for the evolution of the turnover with redshift, we do not evolve the shape or position of the free-streaming cutoff in the mass function at higher redshifts.

In WDM scenarios, the delayed onset of structure formation affects the assembly history of dark matter halos and suppresses their concentrations $c \equiv \frac{r_{\rm vir}}{r_s}$[8] on mass scales that extend above $m_{\rm hm}$ (Schneider et al., 2012; Bose et al., 2016). We use the functional form proposed by (Bose et al., 2016), and write the WDM concentration-mass relation as

$$\frac{c_{\rm WDM}\left(m,z\right)}{c_{\rm CDM}\left(m,z\right)} = \left(1+z\right)^{\beta(z)}\left(1 + 60\frac{m_{\rm hm}}{m}\right)^{-0.17} \tag{5.12}$$

with $\beta\left(z\right) = 0.026z - 0.04$, using the CDM mass-concentration model of Diemer & Joyce (2019) and a scatter of 0.1 dex (Dutton & Macciò, 2014). The WDM suppression factor for the mass-concentration relation we use was calibrated for halos on mass scales below $M_{200} \sim 10^9 M_\odot$, and is accurate in the redshift range $z = 0 - 3$. We note that since flux ratios are particularly sensitive to the central density of perturbing halos, the suppression of halo concentrations far above $m_{\rm hm}$ (because of the factor of 60 in Equation 5.12) is possibly

[8]We define $r_{\rm vir}$ with respect to the matter density contrast $200\rho_{\rm crit}$.

the dominant effect of dark matter free-streaming on lensing observables. We plot the subhalo mass function and the halo mass-concentration-redshift relation in Figure 5.3.

5.4 The data

We apply the forward-modeling methodology outlined in Section 5.2 using the physical model described in Section 5.3 to eight quadruply imaged quasars. In this Section, we describe the sample selection, and how the data for these eight systems was obtained. In Table 5.3 in Appendix 5.10, we summarize the data used in the analysis and provide the relevant references.

5.4.1 The narrow-line systems

The quads in our sample have image fluxes measured using the narrow-line emission from the background quasar. Six of these (WGD 2038, WFI 2033, RX J0911, PS J1606, WGD J0405, and WFI 2026) have flux and astrometry presented by Nierenberg et al. (2020), while the data for B1422 and HE0435 are taken from Nierenberg et al. (2014) and Nierenberg et al. (2017), respectively. The flux uncertainties for the narrow-line lenses are estimated from the forward-modeling method used to fit the narrow-line spectra. For additional details regarding the measurement methodology for the narrow-line flux ratios, we refer to Nierenberg et al. (2017, 2020).

Shajib et al. (2019) analyzed several systems in our sample. They measured satellite galaxy location and provided the photometric information for the systems J1606 and WGD J0405, which we used to obtain photometric redshifts (see Appendix 5.9).

5.4.2 Lenses omitted from our sample

We apply our analysis to a sample of eight quads, although additional systems exist in the literature with measured flux ratios. We choose only a subset of the total number of possible lenses since the remaining systems either do not have reliable flux measurements, or

have complicated deflector morphology that introduces significant uncertainties in the lens modeling. We do not include lenses with fluxes measured using radio emission from the background quasar. Some of these systems may be analyzed in a future work upon revision of our modeling strategy and new flux measurements.

Specifically, we do not include quads with main lensing galaxies that contain stellar disks, since accurate lens models for these systems require explicit modeling of the disk. This excludes the system J1330 presented by Nierenberg et al. (2020). We also exclude HS 0810, a system with narrow-line flux measurements presented by Nierenberg et al. (2020) because the flux from the merging images becomes blended together for source sizes larger than 20 pc. This complicates our analysis, as our method for computing image fluxes with extended background sources cannot be applied to merging pairs when the images blur together.

5.5 Physical assumptions and priors

The parameterizations we introduce in Section 5.3 and the priors use in the forward model reflect certain physical assumptions. In this section we describe these assumptions, and the prior probabilities attached to each parameter in the forward model for our sample of quads.

5.5.1 The extended background source

The effect of a dark matter halo of a given mass on the magnification of a lensed image is a function of the background source size (Dobler & Keeton, 2006), see also Figure 14 in Amara et al. (2006) and Figure 8 in Xu et al. (2012). In general, more extended background sources are less sensitive to dark matter halos (in terms of the image magnifications) on the mass scales relevant for substructure lensing, and the minimum sensitvity threshold for a halo of a given max to produce a measurable flux perturbation is determined by the background source size.

The lenses in our sample have fluxes measured using emission from the narrow-line region of the background quasar (Nierenberg et al., 2017, 2020). The narrow-line region is expected

Table 5.2: A summary of deflector z_d and source z_s redshifts, and satellite galaxies included in the lens model for the quads in our sample. Galaxy positions prior marked by * denote observed locations, which may differ from the true physical location due to foreground lensing effects from the lens macromodel. We correct for foreground lensing effects in our inference pipeline (see Section 5.5.8). Satellite galaxy locations are quoted with respect to the light centroid of the main deflector (see Table 5.3). All priors on the satellite mass $G2_{\theta_E}$ are positive definite. The raised and lowered numbers around the deflector redshifts for PS J1606, WGD J0405, and WFI 2026 are the 68% confidence intervals on the estimated lens redshifts (see Appendix 5.9), which we marginalize over.

lens	z_d	z_s	$\log_{10} M_{\rm halo}$	$\gamma_{\rm ext}$	$\mathrm{G2_x}$	$\mathrm{G2_y}$	$\mathrm{G2_z}$	$\mathrm{G2}_{\theta_\mathrm{E}}$
WGD J0405-3308	$0.29^{0.32}_{0.25}$	1.71	$\mathcal{N}(13.3, 0.3)$	$\mathcal{U}(\prime.\prime\in, \prime.\infty)$	-	-	-	-
HE0435-1223	0.45	1.69	$\mathcal{N}(13.2, 0.3)$	$\mathcal{U}(0.02, 0.13)$	${}^*\mathcal{N}(2.585, 0.05)^*$	${}^*\mathcal{N}(-3.637, 0.05)^*$	$z_d + 0.33$	$\mathcal{N}(0.37, 0.03)$
RX J0911+0551	0.77	2.76	$\mathcal{N}(13.1, 0.3)$	see Section 5.5.9	$\mathcal{N}(-0.767, 0.05)$	$\mathcal{N}(0.657, 0.05)$	z_d	$\mathcal{N}(0.2, 0.2)$
B1422+231	0.36	3.67	$\mathcal{N}(13.3, 0.3)$	$\mathcal{U}(0.12, 0.35)$	-	-	-	-
PS J1606-2333	$0.31^{0.36}_{0.26}$	1.70	$\mathcal{N}(13.3, 0.3)$	$\mathcal{U}(0.1, 0.28)$	$\mathcal{N}(-0.307, 0.05)$	$\mathcal{N}(-1.153, 0.05)$	z_d	$\mathcal{N}(0.27, 0.05)$
WFI 2026-4536	$1.04^{1.12}_{0.9}$	2.2	$\mathcal{N}(13.3, 0.3)$	$\mathcal{U}(0.03, 0.16)$	-	-	-	-
WFI 2033-4723	0.66	1.66	$\mathcal{N}(13.4, 0.3)$	$\mathcal{U}(0.13, 0.32)$	$\mathcal{N}(0.245, 0.025)$	$\mathcal{N}(2.037, 0.025)$	z_d	$\mathcal{N}(0.02, 0.005)$
					${}^*\mathcal{N}(-3.965, 0.025)^*$	${}^*\mathcal{N}(-0.025, 0.025)^*$	$z_d + 0.085$	$\mathcal{N}(0.93, 0.05)$
WGD 2038-4008	0.23	0.78	$\mathcal{N}(13.4, 0.3)$	$\mathcal{U}(0.04, 0.12)$	-	-	-	-

to subtend angular scales larger than a micro-arcsecond, corresponding to physical scales larger than $\sim$ 1pc, such that it is immune to microlensing by stars. This physical extent also corresponds to a light-crossing time greater than the typical time delay between lensed images, such that variability in the background quasar should be washed out of the light curves if the source size is indeed large enough to avoid microlensing.

The size of the narrow-line region typically spans up to $\sim$ 60pc (Müller-Sánchez et al., 2011) defined as the full-width at half maximum (FWHM) of the radially averaged luminosity profile. Upper limits of 50-60 pc may also be obtained by forward modeling the spectrum of the lensed images themselves (Nierenberg et al., 2017). We therefore model the background source as a circular Gaussian and impose a uniform prior on the FWHM between $25 - 60$pc.

5.5.2 Halo and subhalo mass ranges

We render halos for both the line of sight and subhalo mass functions in the range $10^6 - 10^{10} M_\odot$. Halos with masses below $10^6 M_\odot$ do not leave imprints on lensing observables for the extended source sizes we consider, which we verify by comparing distributions of image flux ratios with different minimum subhalo masses. The smallest halo masses flux ratios are sensitive to depends on the background source size and the concentration of the halo, but we estimate through ray-tracing simulations that the lower limit lies somewhere between $10^6 - 10^7 M_\odot$ for the smallest source sizes we model. We include the rare objects more massive than $10^{10} M_\odot$ by explicitly including them in the lens model, assuming that they host a luminous galaxy, in which case they are detected in the observations of the lenses themselves. This assumption is consistent with current abundance matching techniques (Kim et al., 2018; Nadler et al., 2019).

5.5.3 The line of sight halo mass function

We use the Sheth-Tormen (Sheth et al., 2001) halo mass function to model structure along the line of sight, with two modifications: First, we introduce a rescaling term $\delta_{\rm los}$ to account for a systematic shift in the predicted mean amplitude of the mass function. Second, we

include a term $\xi_{\rm 2halo}\left(M_{\rm halo}, z\right)$ that rescales the amplitude of the mass function near the main deflector to account for the presence of correlated structure in the density field near the parent dark matter halo. This results in a $5-10\%$ increase in the number halos near the main deflector.

Apart from uncertainty in the overall amplitude $\delta_{\rm los}$, we assume the halo mass function in the lens cone volume is well-described by the mean halo mass function in the Universe. This is a reasonable approximation as lensing volumes span several Gpc, and we expect fluctuations in the dark matter density along the line of sight should average out over large distances. We note, however, that there is some scatter among the predictions from different parameterizations of the halo mass function below $10^{10} M_{\odot}$ (e.g. Despali et al., 2016) and cosmological model uncertainties, for instance associated with σ_8 and Ω_m. It is also possible that lenses are selected preferentially in over or under-dense lines of sight. We use a flat prior on $\delta_{\rm los}$ between 0.8 and 1.2 to account for these uncertainties.

5.5.4 The subhalo mass function

Our parameterization of the subhalo mass function is an improvement over previous modeling efforts in predicting strong lensing observables since it explicitly accounts for the evolution of the subhalo mass function with redshift and halo mass, and accounts for the tidal stripping of subhalos by the host dark matter halo. However, since the `galacticus` runs do not include a central galaxy[9] we cannot predict the effects of tidal stripping on the projected mass in substructure near the Einstein radius, or the possible redshift and halo mass dependence of this effect. Since tidal destruction of substructures appears to be independent of subhalo mass (Garrison-Kimmel et al., 2017; Graus et al., 2018), we absorb the effects of tidal stripping into the normalization parameter $\Sigma_{\rm sub}$ in Equation 6.5. Finally, we note that the prescription for rendering halos outlined in Section 5.3 does not couple parameters such as the truncation radius to the concentration of subhalos at infall, and does not model the tidal evolution of subhalos from the time of infall until the time of lensing. These additional

[9] `galacticus` is capable of including the tidal stripping effects from a central galaxy, but we did not include them to minimize computation costs.

degrees of modeling complexity will be implemented in a future analysis that uses a larger sample size of lenses.

To determine reasonable bounds on $\Sigma_{\rm sub}$, we compare the predicted surface density in substructure obtained by integrating Equation 6.5 over mass with the output from N-body simulations, and from the `galacticus` runs. At $z \sim 0.7$, the $\sim 10^{13} M_{\odot}$ halos in Fiacconi et al. (2016) have projected substructure mass densities of $10^7 M_{\odot}{\rm kpc}^{-2}$ at $0.02 R_{\rm vir}$. Fiacconi et al. (2016) show that this value increases when accounting for baryonic contraction of the halo. The `galacticus` halos contain more substructure at the same redshift without accounting for baryonic contraction, corresponding to projected mass densities between $2.5 \times 10^7 M_{\odot}{\rm kpc}^{-2}$ and $6 \times 10^7 M_{\odot}{\rm kpc}^{-2}$. Both of these projected mass densities would likely decrease when accounting for tidal stripping. We note, however, that recent works call attention to possible numerical issues that can lead to the artificial fragmentation of subhalos in N-body simulations (van den Bosch et al., 2018; Errani & Peñarrubia, 2019). For reference, $\Sigma_{\rm sub} = 0.012{\rm kpc}^{-2}$ corresponds to a projected mass density of $10^7 M_{\odot}{\rm kpc}^{-2}$ at $z = 0.5$ in a $10^{13} M_{\odot}$ halo, using Equation 6.5.

With these considerations in mind, we use a wide, flat prior on $\Sigma_{\rm sub}$ between 0 and 0.1 ${\rm kpc}^{-2}$ that should encompass the theoretical uncertainties present in the literature. We reiterate that by factoring out the evolution with halo mass and redshift, we intend for the parameter $\Sigma_{\rm sub}$ to be common for all the lenses in our sample with scatter from different tidal stripping scenarios and halo-to-halo variance.

The power-law slope α of the subhalo mass function predicted by N-body simulations is consistently in the range -1.95 to -1.85 (Springel et al., 2008; Fiacconi et al., 2016), and because tidal stripping appears independent of mass the presence of a central galaxy should not cause significant deviations from this prediction. We therefore impose a flat prior on α between -1.95 and -1.85.

5.5.5 Free-streaming in WDM

The prior on $m_{\rm hm}$ needs to be chosen with care since statements using confidence intervals depend on the choice of prior. We specify the lower bound on the prior for $m_{\rm hm}$ with the WDM mass-concentration relation (Equation 5.12) in mind, since the factor of 60 in the denominator of Equation 5.12 results in suppressed halo concentrations nearly two orders of magnitude above the location of the turnover in the mass function (see Figure 5.3). We choose a lower bound for $m_{\rm hm}$ at $10^{4.8} M_{\odot}$ that preserves the CDM-predicted halo concentrations down to $10^{7} M_{\odot}$. At $10^{6} M_{\odot}$, even the coldest mass function we model with $m_{\rm hm} = 10^{4.8} M_{\odot}$ result in halo concentrations for $10^{6} M_{\odot}$ objects 25% lower than the CDM prediction, but we expect the signal from these very low-mass halos will be sub-dominant given that we model extended background sources which decrease sensitivity to low-mass halos.

5.5.6 The parent dark matter halo mass

We use information about the mean population of early-type galaxy lenses, as well as empirical relations between stellar mass, halo mass, and observable quantities such as the image separations and lens/source redshifts, to construct priors for the halo mass of each system.

First, we estimate the 'lensing' velocity dispersion from the Einstein radius and lens/source redshifts using the empirical relation between the stellar mass and velocity dispersion derived by Auger et al. (2010) for a sample of strong lens galaxies. We account for the scatter between spectroscopic velocity dispersion and the 'lensing' velocity dispersion (Treu et al., 2006), and uncertainties in the fit by Auger et al. (2010), and convert the estimated stellar mass into a halo mass using the halo-to-stellar mass ratio $\frac{M_{\rm halo}}{M_{*}} = 75^{+36}_{-27}$ inferred by Lagattuta et al. (2010). The typical uncertainty in the resulting prior for the halo mass is 0.3 dex.

We use this procedure to construct a prior for the halo mass of each quad, with the exceptions of B1422, PS J1606, and WGD J0405. The stellar velocity dispersions implied by the Einstein radii of these systems is significantly lower than the stellar velocity dispersion in the sample of quads used to calibrate the halo-to-stellar mass ratio in Lagattuta et al. (2010), and as such the estimate of the halo mass using the above procedure may not be

accurate for these systems. For B1422, PS J1606, and WGD J0405, we therefore assume the population mean of $10^{13.3\pm0.3}M_{\odot}$ inferred by Lagattuta et al. (2010). We also assume the population mean halo mass for WFI 2026 since the lens redshift used to estimate the central velocity dispersion is very uncertain.

The system RX J0911 is known to reside near a cluster of galaxies, and thus convergence from the cluster halo contributes to the mass within the Einstein radius. We approximate the contribution from the cluster convergence by noting that it should be approximately equal to the mean external shear we infer of 0.3. We then rescale the Einstein radius by $\sqrt{0.7}$, since the stellar mass scales as $R^2_{\rm Ein}$ and where we have used the fact that the mean convergence inside the Einstein radius is approximately equal to one for an isothermal deflector. The priors for the parent halo mass used for each quad are listed in Table 5.2.

Since we explicitly model the evolution with halo mass, we vary $\Sigma_{\rm sub}$ and $M_{\rm halo}$ independently. We note however, that $M_{\rm halo}$ and $\Sigma_{\rm sub}$ are not completely degenerate in our analysis. While the number of lens plane subhalos depends on both parameters, the truncation radius of the subhalos depends on $M_{\rm halo}$ through the distribution of subhalo z-coordinates, which in turn depends on the virial radius of the parent halo (see Equation 5.6), and the 2-halo term appearing in Equation 5.9 depends on the halo mass as a larger halo will have more correlated structure around it. Figure 5.4 provides a visual representation of the link between $M_{\rm halo}$, $\Sigma_{\rm sub}$, α, $\delta_{\rm los}$, and $m_{\rm hm}$.

5.5.7 The main deflector lens model

The galaxies that dominate the lensing cross-section are typically massive early-types with stellar velocity dispersions $\sigma > 200$ km sec^{-1} (Gavazzi et al., 2007; Auger et al., 2010; Lagattuta et al., 2010). The mass profiles of these systems are typically inferred to be isothermal, or close to isothermal (Treu et al., 2006, 2009; Auger et al., 2010; Shankar et al., 2017). These observations motivate a simple parameterization for the main deflector lens model, the singular isothermal ellipsoid (SIE) plus external shear. We generalize this model to a power-law ellipsoid with a variable logarithmic slope $\gamma_{\rm macro}$ to account for uncertainties

associated with the mass profile of the lensing galaxy, and the model-predicted flux ratios. We assume a flat prior on the power-law slope $\gamma_{\rm macro}$ between 1.95 and 2.2 for each deflector (Auger et al., 2010).

In addition to the logarithmic slope of the main deflector mass profile, we sample values for the external shear strength $\gamma_{\rm ext}$. The prior for $\gamma_{\rm ext}$ is chosen on a lens-by-lens basis by first sampling the macromodel parameter space without subhalos to determine a reasonable starting range for $\gamma_{\rm ext}$. The width and center of the prior is adjusted after adding substructure such that the posterior distribution of $\gamma_{\rm ext}$ obtained for each lens is contained well within the bounds of the prior. The specific priors used for each system are summarized in Table 5.2. Finally, we use a Gaussian prior for the mass centroid of each quad centered on the main deflector light with a variance of 0.05 arcseconds, a typical modeling uncertainty for quadruple-image systems (Shajib et al., 2019; Nierenberg et al., 2020).

Several studies (Evans & Witt, 2003; Hsueh et al., 2016; Gilman et al., 2017; Hsueh et al., 2017, 2018) explore the role of complicated main deflector morphologies on the model predicted flux ratios. As image magnifications are local probes of the gravitational potential, if there are fluctuations in the surface mass profile on scales comparable to the image separation these structures can affect the image magnifications. In particular, stellar disks, if they go unnoticed, can result in systematically inaccurate lens models. With deep Hubble Space Telescope (HST) images of the narrow-line quads in our sample, we can confirm that they do not contain disks, and indeed are representative of the massive elliptical galaxies with roughly isothermal mass profiles that typically act as strong lenses (Auger et al., 2010; Shankar et al., 2017). Gilman et al. (2017) and Hsueh et al. (2018) quantified the systematic uncertainties introduced by modeling early-type galaxy lenses as isothermal ellipsoids with fixed logarithmic slopes $\gamma = 2$. These works found that the resulting systematic uncertainties on image magnifications are typically less than 10%. This degree of uncertainty is comparable to the variance in model-predicted image magnifications resulting from marginalizing over a power-law ellipsoid mass model with additional degrees of freedom implemented through a variable logarithmic slope γ (Nierenberg et al., 2020). Based on these considerations, we use a power-law ellipsoid with variable logarithmic slopes γ to model the main deflector mass

profile.

Three quads in our sample do not have measured spectroscopic redshifts. For two of these, we use photometry from Shajib et al. (2019) to compute photometric redshifts probability distributions with the software `eazy` (Brammer et al., 2008), and sample the deflector redshift from these distributions in the forward model. For the third system (WFI 2026), which does not have multi-band photometry from Shajib et al. (2019), we assume a typical velocity dispersion for a massive elliptical galaxy, and derive a probability distribution for the lens redshift from measured quantities such as the source redshift and measured image separation. We give more details regarding this procedure in Appendix 5.9.

5.5.8 Satellite galaxies and nearby deflectors

We model satellite galaxies and other deflectors near the main lens as Singular Isothermal Spheres (SIS), and assume they lie at the lens redshift unless they have measured redshifts that place them elsewhere. We marginalize over the position and Einstein radius of these objects using Gaussian priors on the positions centered on the light centroid with a variance of 0.05 arcseconds. We use a Gaussian prior on the Einstein radius which is estimated from lens model fitting, or in some cases by direct measurements on the central velocity dispersion (e.g. Wong et al., 2017; Rusu et al., 2019).

In the cases of HE0435 and WFI 2033, the nearby galaxy lies at a higher redshift than the main lens plane. The light from the galaxy is therefore subject to lensing by the main deflector, and its true physical location differs from its observed position. We estimate the true physical locations of these objects by sampling the macromodel parameter space using the image positions as constraints, and read out the physical position of background satellite given its observed (lensed) position. We then place the satellite at this derived physical location in the forward model sampling with uncertainties of 0.05 arcseconds. This process significantly speeds up the lensing computations since it does not require the continuous reevaluation of the physical satellite location given its observed position during each lens

model computation[10]. The boost in speed comes at the cost of decoupling the satellite galaxy position from the dark matter parameters $\mathbf{q}_s$ in the inference, but we expect the covariance between these quantities will be negligible because the satellite galaxies, even when their locations are corrected for foreground lensing effects, are relatively far from the images, introducing convergence at the main deflector light centroid of < 0.1 in both cases.[11]

In the case of HE0435, we estimate the angular location without foreground lensing of the satellite to be $(-2.37, 2.08)$, while for WFI 2033 we obtain $(-3.63, -0.08)$, for observed (lensed) locations of $(-2.911, 2.339)$ and $(-3.965, -0.022)$, respectively. These coordinates are with respect to the galaxy light centroid (see Table 5.3). The angular locations of the lensed background satellites are closer to the mass centroid of the main deflector, just as the physical location of the lensed background quasar is concentric with the mass centroid.

The lens-specific priors on satellite galaxies are summarized in Table 5.2.

5.5.9 Lens-specific modeling for RX J0911+0551 and WGD 2038-4008

For system RX J0911, we alter the modeling strategy slightly to increase computational efficiency by allowing the external shear strength $\gamma_{\rm ext}$ to vary freely while solving for macromodel parameters that fit the observed image positions. For the system WGD 2038, we widen the prior on the power-law slope of the macromodel as the posterior using the default range for $\gamma_{\rm macro}$ between 1.95 - 2.2 is biased towards higher values of $\gamma_{\rm macro}$. For WGD 2038, the posterior peaks at $\gamma_{\rm macro} \sim 2.25$.

[10]The physical location of the nearby galaxy needs to be continuously reevaluated because it's observed location depends on the foreground lensing effects from the macromodel, and the parameters describing the macromodel are continuously changing while finding a solution to the lens equation (Equation 5.3).

[11]The default convention in `lenstronomy` is to place deflectors at their observed angular locations in the Universe, but it is now possible (in code versions 0.8.0+) to specify which objects should be treated using the observed (lensed) position instead. We note that the default convention in `lensmodel` (Keeton et al., 1997) is to place objects at their observed (lensed) locations during multi-plane ray-tracing.

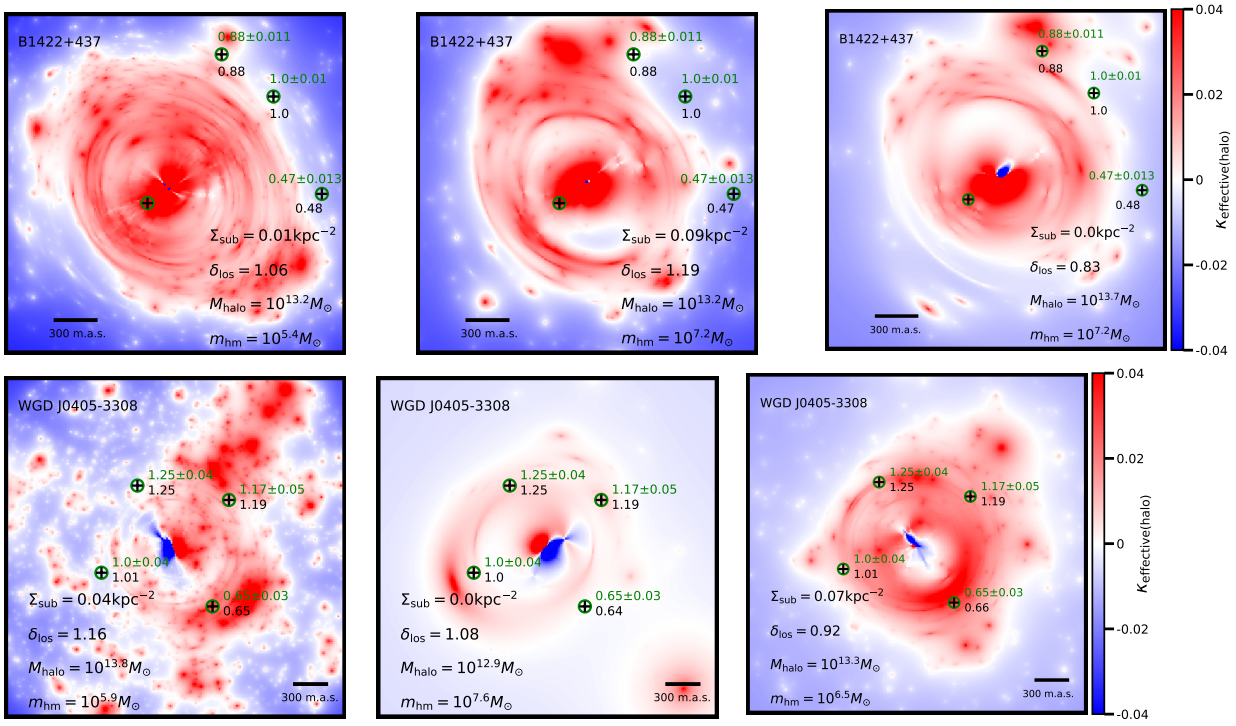

Figure 5.5: Dark matter halo *effective multi-plane convergence* maps for some of the highest-ranked realizations for the quads B1422 and WGD J0405, each of which has flux ratios inconsistent with smooth lens models. The defintion of the *effective multi-plane convergence* takes into account the non-linear effects present in multi-plane lensing, and is defined with respect to the mean dark matter density in the Universe such that some regions are underdense (blue), while other regions (specifically, dark matter halos) are over-dense (red). The sub-halo mass function normalization, line of sight normalization, halo mass and half-mode mass are displayed for each realization. Green text/circles denote observed image positions and fluxes, while black text/crosses denote the model positions and fluxes. The forward-model data sets fit the image positions and fluxes to within the measurement uncertainties.

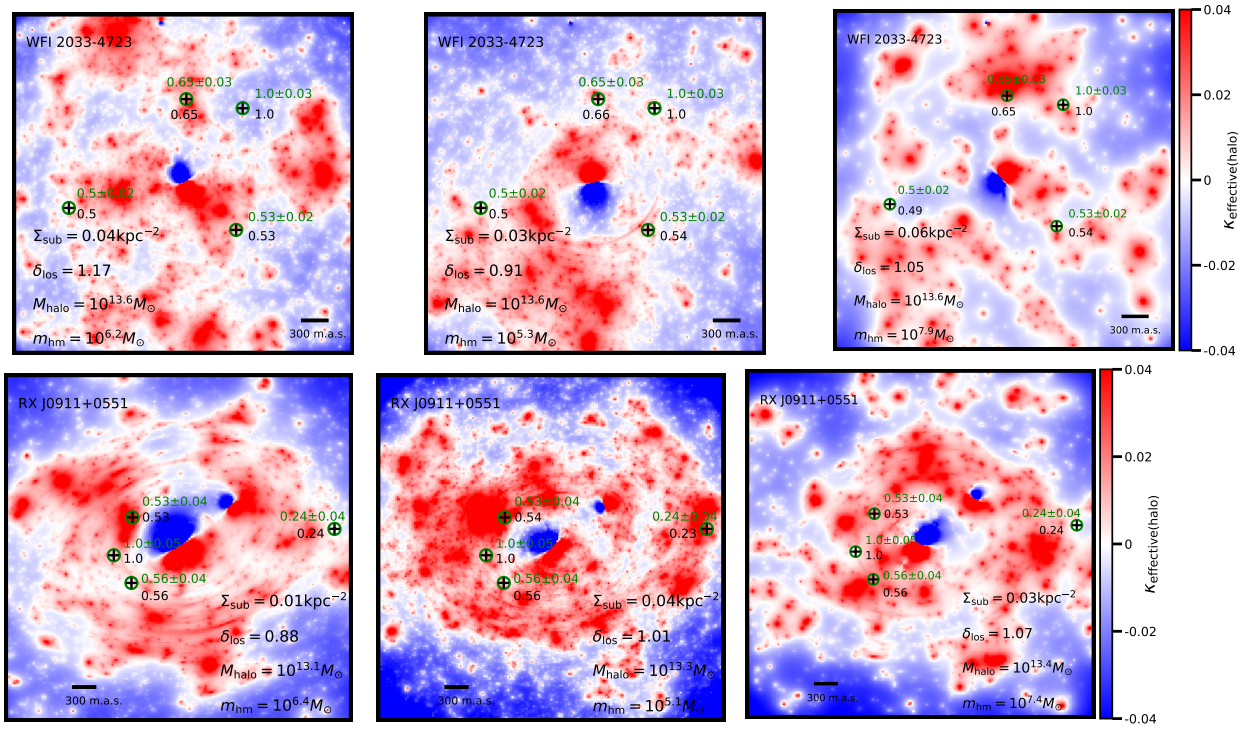

Figure 5.6: Same as Figure 5.5 for the systems WFI 2033-4723 and RX J0911+0551.

5.6 Results

In this section we present the results of our analysis. We begin in Section 5.6.1 by showing dark matter halo convergence maps for some of the top-ranked realizations drawn in the forward model. We then display the posterior distributions for a few individual lenses, showing the simultaneous inference of parameters describing the macro lens model and the dark matter hyper-parameters. In Section 5.6.2 we present the constraints on the abundance of substructure and dark matter warmth for the full sample of 11 quads.

5.6.1 Top-ranked realizations and posteriors for individual lenses

Minimizing the summary statistic in Equation 5.4 selects realizations that resemble the observed data as closely as possible. This guarantees that the set of accepted dark matter hyper-parameters $\mathbf{q}_s$ yield an accurate approximation of the true posterior distribution for each individual lens with data $\mathbf{d}_n$: $p(\mathbf{q}_s|\mathbf{d}_n)$. For visualization purposes, and to reinforce

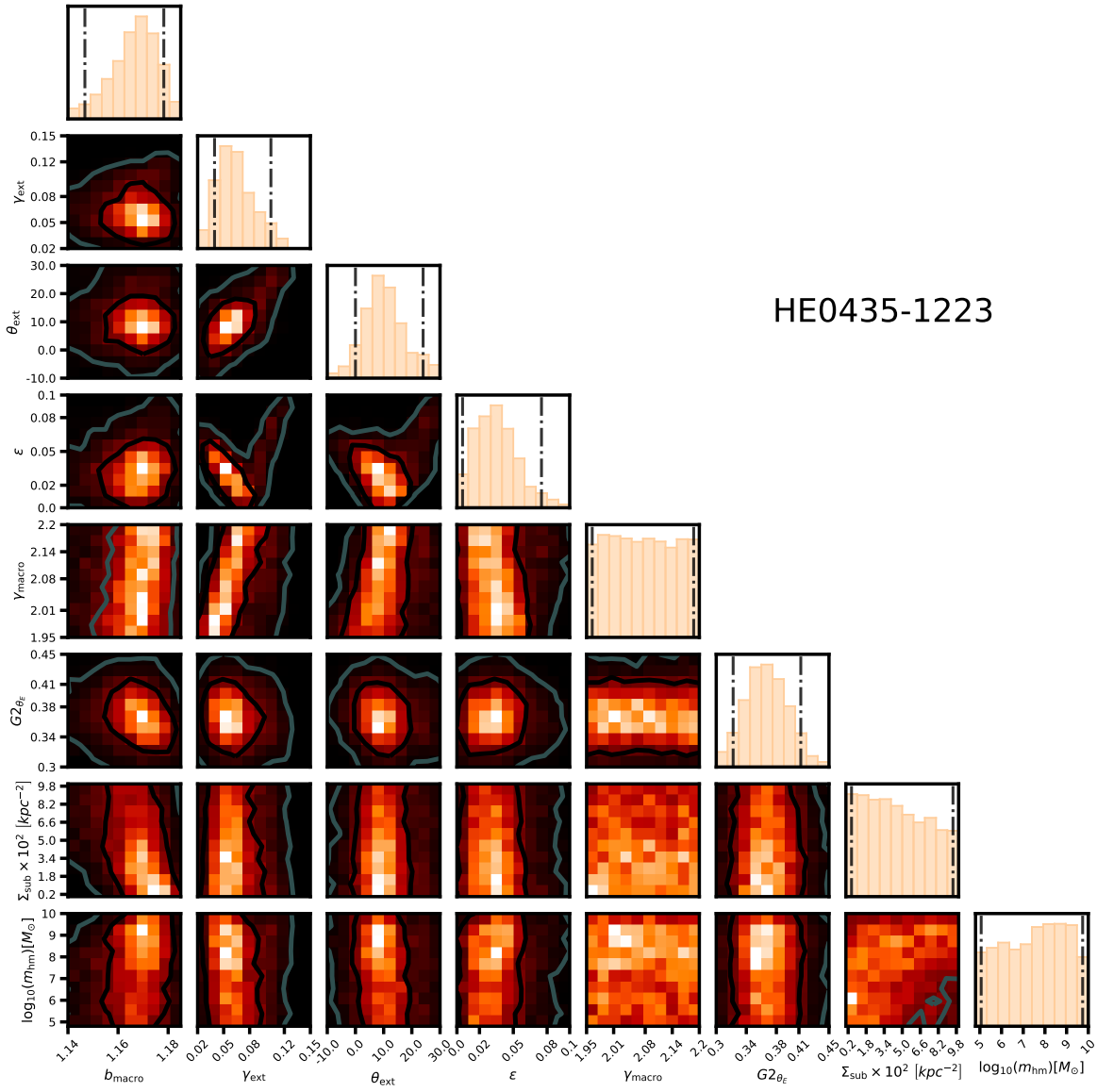

Figure 5.7: Joint posterior distribution for a subset of $\mathbf{M}$ and $\mathbf{q}_{\rm s}$ parameters for the system HE0435. We display the normalization of the main deflector lens model $b_{\rm macro}$, the external shear strength and position angle $\gamma_{\rm ext}$ and $\theta_{\rm ext}$, the deflector ellipticity ϵ, the power-law slope of the main deflector mass profile $\gamma_{\rm macro}$, the Einstein radius of the satellite galaxy $G2_{\theta_E}$, the normalization of the subhalo mass function $\Sigma_{\rm sub}$, and the half-mode mass $m_{\rm hm}$. We simultaneously sample the distributions of these parameters to account for covariance between the macromodel and the dark matter hyper-parameters $\mathbf{q}_{\rm s}$. Vertical lines denote 95% confidence intervals.

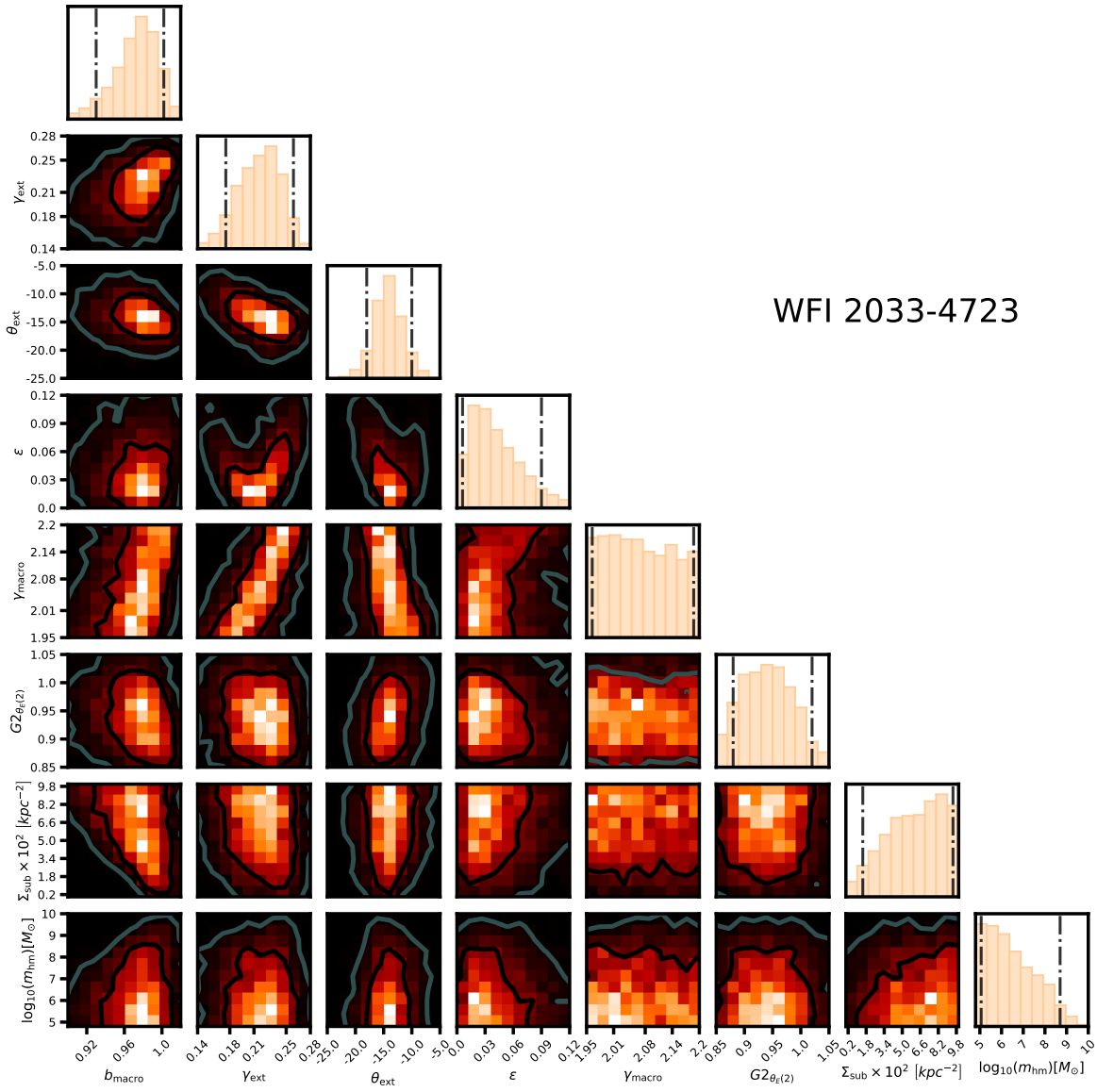

Figure 5.8: Joint posterior distribution for a subset of $\mathbf{M}$ and $\mathbf{q_s}$ parameters for the system WFI 2033. The parameters are the same as in Figure 5.7. In addition to the main deflector we model two additional nearby galaxies, with Einstein radii $G2_{\theta_E(1)}$ and $G2_{\theta_E(2)}$. We show the distributions of the Einstein radius for the larger nearby galaxy ($G2_{\theta_E(2)}$), whose position we correct for foreground lensing effects (see Section 5.5.8).

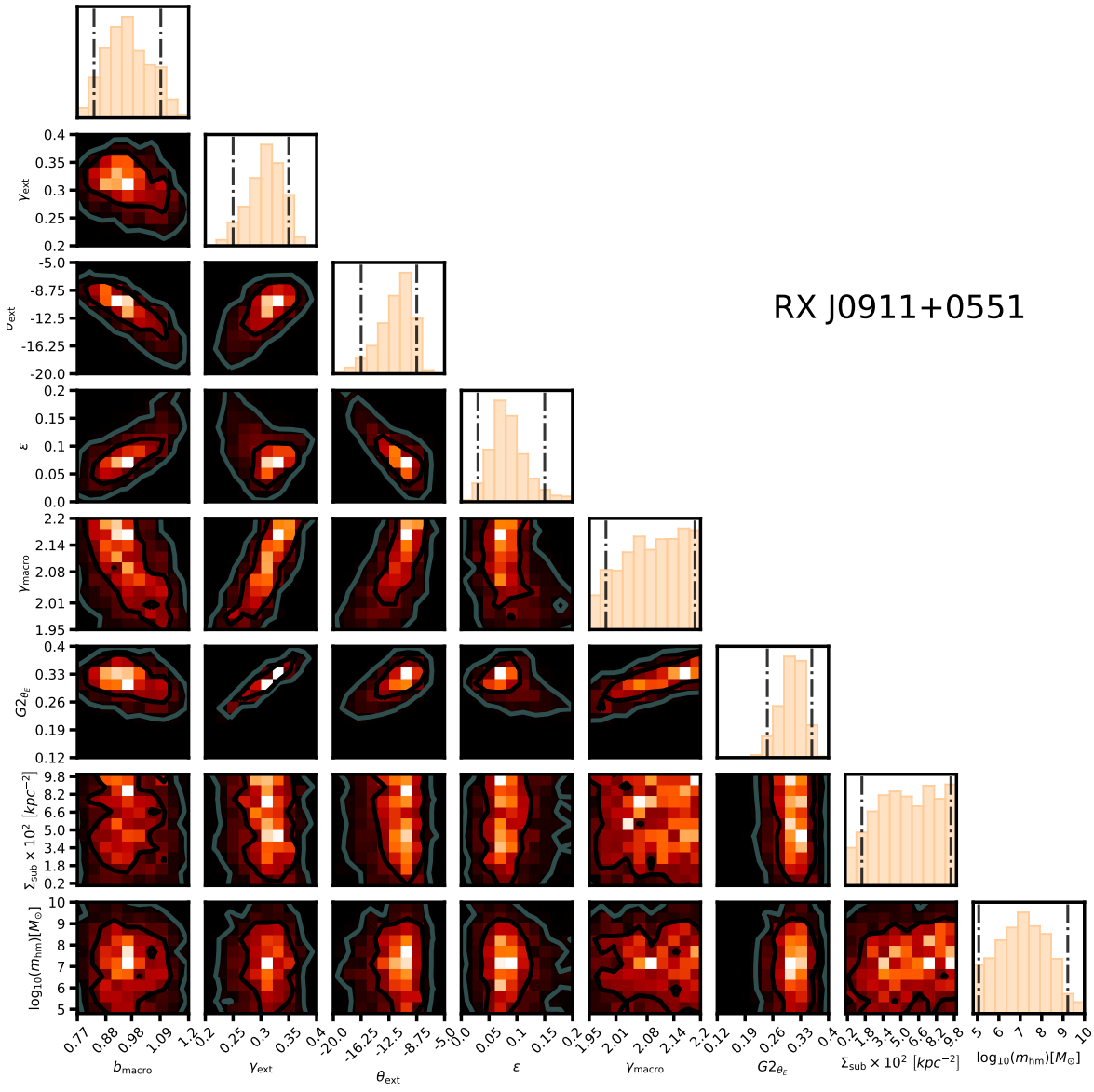

Figure 5.9: Joint posterior distribution for a subset of $\mathbf{M}$ and $\mathbf{q}_{\rm s}$ parameters for the system RX J0911. The parameters are the same as in Figure 5.7.

the fact that the top-ranked realizations look like the data and satisfy $S_{\rm lens} \approx 0$ (Equation 5.4), in Figure 5.5 we display the dark matter halo *effective multi-plane convergence* maps for some of the top ranked realizations for a subset of quads in our sample. The *effective multi-plane convergence* is defined as half the divergence of the full deflection field $\boldsymbol{\alpha}$

$$\kappa_{\rm effective} \equiv \frac{1}{2}\, \boldsymbol{\nabla} \cdot \boldsymbol{\alpha}. \tag{5.13}$$

This definition of the multi-plane convergence accounts for the non-linear effects present in multi-plane lensing, and satisfies the single-plane definition of convergence as second derivatives of a lensing potential in the absence of multiple lens planes.

To visualize individual realizations of dark matter structure, we define $\kappa_{\rm effective(halo)} \equiv \kappa_{\rm effective} - \kappa_{\rm macro}$, where $\kappa_{\rm macro}$ is the convergence from the lens macromodel, including satellite galaxies and nearby deflectors. In the resulting convergence maps, halos located behind the main lens plane appear sheared tangentially around the Einstein radius due to coupling to the large deflections produced by the macromodel.

In Figure 5.5, we show $\kappa_{\rm effective(halo)}$ maps of randomly selected realizations of dark matter structure whose corresponding $\mathbf{q}_{\rm s}$ parameters were accepted in the final posterior on the basis of their summary statistic $S_{\rm lens}$. The specific realizations and the corresponding dark matter parameters $\mathbf{q}_{\rm s}$ correspond to a diverse set of substructure populations, warm and cold, which yield similarly good fits to the observed flux ratios satisfying $S_{\rm lens} \sim 0$. Some models, however, predict flux ratios that match the observed flux ratios more frequently than others. In terms of the Approximate Bayesian Computing algorithm described in Section 5.2, the frequency with which one dark matter model relative to another predicts observables that resemble the data is a surrogate for the relative likelihood of the models. The probability of accepting a proposed $\mathbf{q}_{\rm s}$ based on the summary statistic in Equation 5.4 is therefore equal to the likelihood $p\left(\mathbf{d}_{\rm n}|\mathbf{q}_{\rm s}\right)$ (Equation 5.2), even though the form of this function is unknown and it is never directly evaluated.

The top-ranked realizations for B1422 shown in Figure 5.5 each have a relatively massive dark matter halo, or several smaller ones, located near the top left merging triplet image with (normalized) flux 0.88. This is in agreement with the analysis by Nierenberg et al.

(2014), who find that a blob of dark matter near this image brings the model-predicted flux ratios into agreement with a smooth lens model.

Although not obvious from examining Figure 5.5, the underlying macromodels for each accepted realization are unique, with different external shears, power-law slopes, lens ellipticity, etc. We marginalize over different macromodel configurations by simultaneously sampling the macromodel parameters and the dark matter hyper-parameters in the forward model. To illustrate, in Figures 5.7, 5.8, and 5.9 we show the posterior distributions for several parameters in the lens macromodel, along with the dark matter hyper-parameters $\Sigma_{\rm sub}$ and $m_{\rm hm}$ for HE0435, WFI 2033, and RX J0911. The system HE0435 generally favors models with low subhalo mass function normalizations (low $\Sigma_{\rm sub}$), or a turnover the mass function with higher $\Sigma_{\rm sub}$. The system WFI 2033 is the opposite, with a posterior favoring CDM-like mass functions with many lens plane subhalos. The system RX J0911 lies somewhere in between, with a peak in the posterior distribution of $m_{\rm hm}$ near $10^7 M_\odot$.

For each of these systems, in particular WFI 2033, there is a visibly obvious covariance between the overall normalization of the main deflector mass profile $b_{\rm macro}$[12], and the parameters $\Sigma_{\rm sub}$ and $m_{\rm hm}$. This covariance is readily understood: To reproduce the observed image positions, the macromodel responds to the addition of mass in the form of subhalos in main lens plane by decreasing the overall normalization of the main deflector mass profile, and hence these quantities are anti-correlated. Similarly, WDM models correspond to macromodels with larger $b_{\rm macro}$ because WDM realizations contain fewer subhalos. Interestingly, there is some structure in the posterior distribution for the lens ellipticity ϵ in WFI 2033, and both $m_{\rm hm}$ and $\Sigma_{\rm sub}$.

By simultaneously sampling the lens macromodel and dark matter hyper-parameters, we obtain posterior distributions that account for covariance between $\mathbf{M}$ and $\mathbf{q_s}$. We do not use lens model priors from more sophisticated lens modeling efforts (e.g. Wong et al. (2017); Shajib et al. (2019)) because these analyses did not include substructure in the lens models and therefore do not account for covariances between the macromodel parameters

[12] $b_{\rm macro}$ has units of convergence, or projected mass density divided by the critical surface mass density for lensing.

and the dark matter parameters of interest. For the same reason, we do not decouple the lens macromodel parameters from the dark matter hyper-parameters by first sampling the macromodel parameter space that fits the image positions, and using these distributions as priors in the forward modeling.

5.6.2 Constraints on the free-streaming length of dark matter

For each quad, we obtain a joint likelihood between the macromodel parameters $\mathbf{M}$ and the dark matter-hyper parameters $\mathbf{q}_{\rm s}$. We marginalize over the parameters in this 20+ dimensional space to obtain the four-dimensional space of $\mathbf{q}_{\rm s}$ parameters that includes logarithmic slope of the subhalo mass function α, the scaling of the line of sight halo mass function $\delta_{\rm los}$, the overall scaling of the subhalo mass function $\Sigma_{\rm sub}$, and the half-mode mass $m_{\rm hm}$. We reiterate that these four parameters describe universal properties of dark matter and should therefore be common to all the lenses, while the parameters $\mathbf{M}$ and the halo mass $M_{\rm halo}$ are lens-specific. After marginalizing, we compute the product of the resulting likelihoods and obtain the desired posterior distribution in Equation 5.1, which we display in Figure 5.10.

The marginalized constraints on $m_{\rm hm}$ rule out $m_{\rm hm} > 10^{7.8} M_{\odot}$ at 2σ, corresponding to thermal relic particle mass of < 5.2 keV. It is apparent from Figure 5.10 that $m_{\rm hm}$ and $\Sigma_{\rm sub}$ are correlated, since halos added by increasing the normalization can be subsequently removed by increasing $m_{\rm hm}$ such that the total amount of lensing substructure remains relatively constant. As a result, the marginalized distribution for the normalization $\Sigma_{\rm sub}$ appears unconstrained from above, as the normalization can be significantly higher in WDM scenarios. With only eight quads we cannot simultaneously measure $m_{\rm hm}$ and $\Sigma_{\rm sub}$, although our previous forecasts indicate this is possible with more lenses (Gilman et al., 2018).

The constraints on dark matter warmth in terms of confidence intervals depend on the range of allowed values specified by the prior on $\Sigma_{\rm sub}$. Similarly, the confidence interval on $m_{\rm hm}$ depends on the lower bound of this parameter that is set by the prior on $m_{\rm hm}$. As discussed in Section 5.5.5, we have chosen the prior on $m_{\rm hm}$ to encompass the region of parameter space where the data can constrain $m_{\rm hm}$, keeping in mind that the WDM

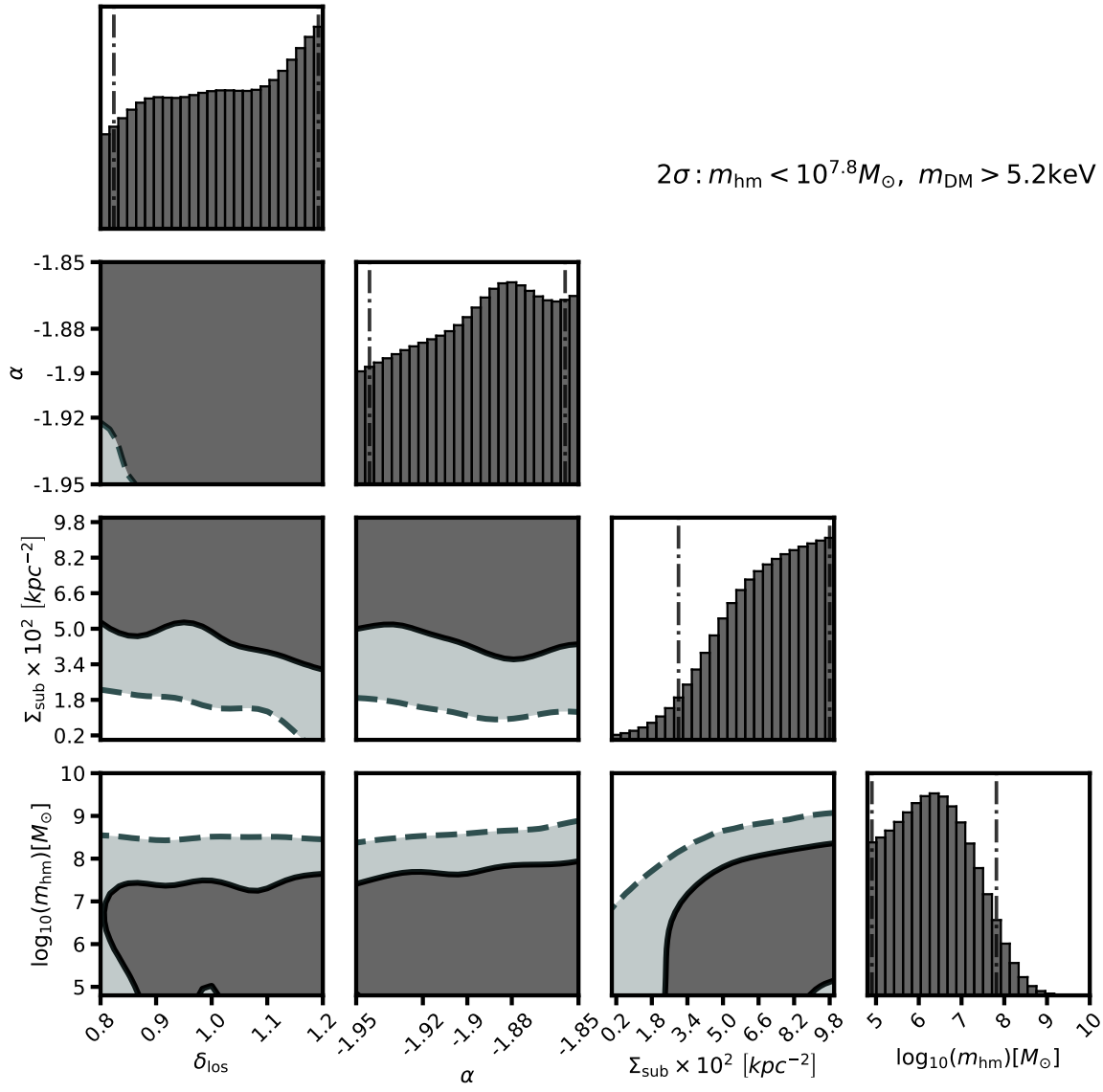

Figure 5.10: Marginal and joint posterior distributions for the dark matter hyper-parameters $\delta_{\rm los}$, α, $\Sigma_{\rm sub}$, and $m_{\rm hm}$, which represent the overall scaling of the line of sight halo mass function, the logarithmic slope of the subhalo mass function, the global normalization of the subhalo mass function that accounts for evolution with halo mass and redshfit (see Equation 6.5), and the half-mode mass $m_{\rm hm}$ relevant to WDM models. Contours show 68% and 95% confidence intervals, while the dot-dashed lines on the marginal distributions show the 95% confidence intervals.

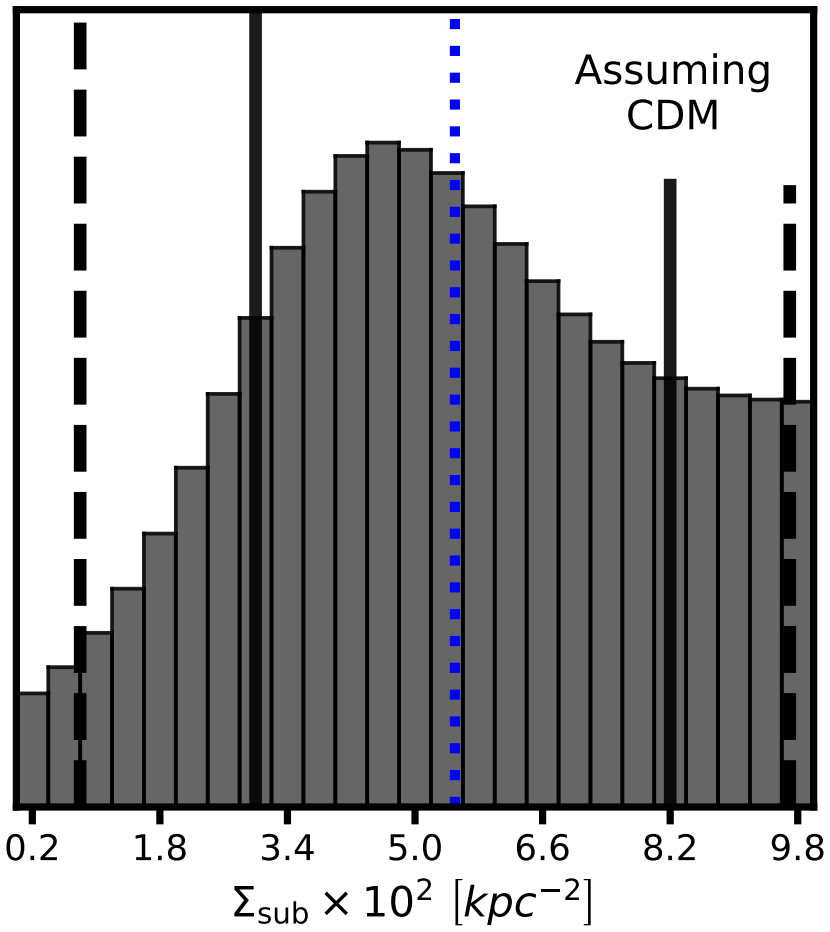

Figure 5.11: Inference on the global normalization of the subhalo mass function $\Sigma_{\rm sub}$ assuming CDM, marginalized over the logarithmic slope α and uncertainty in the overall amplitude of the line of sight halo mass function $\delta_{\rm los}$. The blue dashed lines shows the mean of the marginal distribution, while black solid (dashed) lines represent 68% and 95% confidence intervals. The contours in the joint distribution also represent 68% and 95% confidence intervals.

mass concentration relation affects the central densities of subhalos 60 times above $m_{\rm hm}$ (Equation 5.12), and the upper bound of $\Sigma_{\rm sub} = 0.1{\rm kpc}^{-2}$ is a conservative choice as most N-body simulations and the `galacticus` runs predict values below $0.05{\rm kpc}^{-2}$. In light of these complications, we also quote likelihood ratios which do not depend on the choice of prior. Relative to the peak of the $m_{\rm hm}$ posterior, we obtain likelihood ratios for WDM with $m_{\rm hm} = 10^{8.2} M_\odot$ ($m_{\rm hm} = 10^{8.6} M_\odot$) of 7:1 (30:1)[13].

The posterior for $\delta_{\rm los}$ indicates the data favors more line of sight structure, but the preference is not statistically significant. The parameters $\delta_{\rm los}$ and $\Sigma_{\rm sub}$ are anti-correlated, as one would expect as one can, to a certain degree, remove lens plane subhalos and replace them with line of sight halos while keeping the total amount of flux perturbation constant. This is not a perfect degeneracy, however, since lensing efficiency and the relative number of subhalos and line of sight halos changes with redshift. Thus, a larger sample of quads at different redshifts could break the covariance between $\Sigma_{\rm sub}$ and $\delta_{\rm los}$.

5.6.3 Constraints on the subhalo mass function assuming CDM

We perform a suite of CDM simulations using the same priors listed in Table 5.2, minus the WDM parameter $m_{\rm hm}$, with the aim of inferring $\Sigma_{\rm sub}$. We marginalize over $\delta_{\rm los}$, and over a theoretical-motivated prior on α (between -1.95 and -1.85) based on predictions from N-body simulations (Springel et al., 2008; Fiacconi et al., 2016).

The inference on $\Sigma_{\rm sub}$ is shown in Figure 5.11. We infer $\Sigma_{\rm sub} = 0.055{\rm kpc}^{-2}$, with a 1σ confidence interval $0.029 < \Sigma_{\rm sub} < 0.083\ {\rm kpc}^{-2}$. At the 2σ level we obtain $\Sigma_{\rm sub} > 0.008{\rm kpc}^{-2}$. We do not quote an upper 2σ bound on $\Sigma_{\rm sub}$ as it is prior dominated. To put these numbers in physical units, the mean value of $\Sigma_{\rm sub}$ corresponds to a mean projected mass in substructure for the lenses in our sample between $10^6 - 10^9 M_\odot$ of $4.0 \times 10^7 M_\odot {\rm kpc}^{-2}$, and the 1σ confidence interval corresponds to $2.0 - 6.1 \times 10^7 M_\odot {\rm kpc}^{-2}$. At 2σ, the projected mass constraint is $\Sigma_{\rm sub} > 0.6 \times 10^7 M_\odot {\rm kpc}^{-2}$. To convert into the average projected mass, we have computed

[13]We remind the reader that the relative heights of the peaks in the posterior somewhat depend on the binning method, or in this case the bandwidth estimator of the KDE. In this work, we have applied a KDE with a first order boundary correction and a bandwidth selected according to Scott's factor (Scott, 1992).

the average of the projected masses for each of the eight lenses in our sample, using the scaling of the halo mass function with redshift in Equation 6.6 while assuming a halo mass of $10^{13} M_{\odot}$.

5.7 Discussion and conclusions

In this section we review the main results of this work and discuss the implications for cold and warm dark matter. In Section 5.7.1 we summarize our main results, and in Section 5.7.2 we compare our results with those obtained in previous works. In Section 5.7.3 we discuss the sources of systematic uncertainty in our analysis, and we conclude in Section 5.7.4 by discussing the implications of our result for cold and warm dark matter.

5.7.1 Summary of the analysis and main results

We have carried out a measurement of the free-streaming length of dark matter and the subhalo mass function using a sample of eight quadruply-imaged quasars. The methodology we use to constrain the dark matter parameters of interest has been tested and verified with simulated data (Gilman et al., 2019). Lenses that show evidence for morphological complexity in the form of stellar disks are excluded from our analysis. We model halos both in the main deflector and along the line of sight, including correlated structure around the main deflector through the two-halo term, and account for evolution of the projected subhalo mass function with redshift and halo mass using a suite of simulations using the semi-analytic modeling code `galacticus`. We compute image flux ratios by ray-tracing to finite-size background sources, which correctly accounts for the sensitivity of image flux ratios to perturbing halos. We also marginalize over the macromodel parameters for each system, including the power-law slope of the main deflector, and simultaneously constrain the lens macromodel and dark matter hyper-parameters to account for covariance between these quantities. In addition to the turnover in the halo mass function, we model WDM free-streaming effects on the mass-concentration relation, accounting for the effect of reduced central densities of WDM halos on lensing observables.

The main results of this analysis are summarized as follows:

- We constrain the half-mode mass $m_{\rm hm}$ (thermal relic dark matter particle mass) to $m_{\rm hm} < 10^{7.8} M_\odot$ ($m_{\rm DM} > 5.2$keV) at 2σ. Since the confidence intervals depend on the prior used for both $m_{\rm hm}$ and $\Sigma_{\rm sub}$, we also quote likelihood ratios relative to the peak of the posterior distribution for $m_{\rm hm}$: we disfavor $m_{\rm hm} = 10^{8.2} M_\odot$ ($m_{\rm DM} = 4$keV) with a likelihood ratio of 7:1, and with $m_{\rm hm} = 10^{8.6} M_\odot$ ($m_{\rm DM} = 3.0$keV) the relative likelihood is 30:1. These bounds are marginalized over the amplitude of the subhalo mass function, the amplitude of the line of sight halo mass function, the power-law slope of the subhalo mass function, the parent halo mass, the background source size, and the parameters describing the main deflector mass profile.

- Assuming cold dark matter, we infer a value of the global amplitude of the subhalo mass function $\Sigma_{\rm sub} = 0.055^{0.032}_{-0.027}{\rm kpc}^{-2}$ at 1σ, and $\Sigma_{\rm sub} > 0.008{\rm kpc}^{-2}$ at 2σ. In our lens sample, these values correspond to an average projected mass density in substructure between $10^6 - 10^9 M_\odot$ of $4.0^{+2.1}_{-2.0} \times 10^7 M_\odot{\rm kpc}^{-2}$ and a lower bound of $0.6 \times 10^7 M_\odot{\rm kpc}^{-2}$, respectively. At fixed redshifts, for a $10^{13} M_\odot$ halo at $z = 0.2$ ($z = 0.6$) the 1σ constraint corresponds to a projected mass in substructure of $1.9^{+0.9}_{-0.9} \times 10^7 M_\odot{\rm kpc}^{-2}$ ($4.1^{+2.0}_{-2.0} \times 10^7 M_\odot{\rm kpc}^{-2}$) in the subhalo mass range $10^6 - 10^9 M_\odot$. The 2σ constraint corresponds to a projected mass in substructure of greater than $0.3 \times 10^7 M_\odot{\rm kpc}^{-2}$ ($0.6 \times 10^7 M_\odot{\rm kpc}^{-2}$) in the same mass range.

5.7.2 Discussion and comparison with previous work

5.7.2.1 Constraints on dark matter warmth and the amplitude of the CDM subhalo mass function

The first comprehensive analysis of multiply-imaged quasars was carried out by Dalal & Kochanek (2002) (hereafter DK2), who inferred a projected mass fraction in substructure $\bar{f}_{\rm sub}$ [14] between $0.006 < \bar{f}_{\rm sub} < 0.07$ at 2σ modeling only lens-plane substructure, and assuming

[14] Throughout this section, we will use $\bar{f}_{\rm sub}$ to refer to the average mass fraction in substructure inferred from a sample of multiple lenses in halos of different masses at different redshifts, and $f_{\rm sub}$ to refer to the

CDM. Recently, Hsueh et al. (2019) (hereafter H19) improved on the analysis of DK2 by including the effects of line of sight halos, measuring $0.006 < \bar{f}_{\rm sub} < 0.018$ at 1σ with a mean of 0.011 assuming CDM, and also constrained the free-streaming length of dark matter to $m_{\rm hm} < 10^{8.4}$ ($m_{\rm DM} > 3.8$keV).

The 2σ bound from H19 of $m_{\rm hm} < 10^{8.4} M_\odot$ is weaker than the constraint from this work $m_{\rm hm} < 10^{7.8} M_\odot$. One possible reason for this difference is that unlike previous work (Birrer et al., 2017b; Gilman et al., 2018, 2019) H19 did not model the suppression of the mass-concentration relation in warm dark matter scenarios, which suppresses the lensing signal more than one order of magnitude above the position of the turnover in the mass function. This is of particular relevance for flux ratio studies because the effect of a perturbing dark matter halo depends on its central density profile. Free-streaming effects on the mass-concentration relation therefore increase the relative differences between CDM and WDM on the scales relevant for substructure lensing, which leads to greater constraining power over WDM models. Finally, we note that in a future analysis modeling the tidal evolution of substructures from the time of infall to the time of lensing may introduce additional constraining power over WDM models by coupling the structural parameters of subhalos at the time of lensing to their structural properties, such as concentration, at the time of infall.

To facilitate direct comparison between this analysis and that of DK2 and H19 regarding the constraints on the subhalo mass function assuming CDM, we convert our $\Sigma_{\rm sub}$ values into estimates of $\bar{f}_{\rm sub}$ by computing the projected mass density Σ, and then using the fact that $\frac{\Sigma}{\Sigma_{\rm crit}} = 0.5$ near the Einstein radius, where $\Sigma_{\rm crit}$ is the critical surface mass density for lensing. In these conversions, we also assume a halo mass of $10^{13} M_\odot$, and take care to compute $\bar{f}_{\rm sub}$ using the same mass range $10^6 - 10^9 M_\odot$ used by H19. Our 2σ bounds on $\Sigma_{\rm sub}$ correspond to an average mass fraction in substructure $\bar{f}_{\rm sub} > 0.005$ with a mean of $\bar{f}_{\rm sub} = 0.035$. At 1σ $0.018 < \bar{f}_{\rm sub} < 0.056$. This result is statistically consistent with the constraints from H19, and also with those of DK2.

There are several key differences between our analysis and those of H19 and DK2 that pull

mass fraction in substructure implied by a certain $\Sigma_{\rm sub}$ value at a specific redshift and halo mass.

in opposite directions in terms of constraining power over dark matter models. As mentioned previously we model free-streaming effects on the mass-concentration relation, and include the contribution from the two-halo term to account for correlated structure near the main deflector. These pieces of additional physics add information and increase our constraining power over WDM models. On the other hand, accounting for finite-size background sources decreases the expected magnification signal caused by dark matter halos and subhalos, and we expect to infer a higher normalization of the subhalo mass function in our analysis as more substructure is needed to produce the same degree of flux perturbation. Explicitly, by ray-tracing to finite-size background sources we find that the peak of the magnification cross section for a $5 \times 10^7 M_\odot$ halo is reduced by a factor of two for a 15pc background source relative to a 5pc background source, and by a factor of three for a 40pc source. The simplifying assumption of point-sources for the background quasar invoked by H19 and DK2 introduces signal from low-mass halos whose effects would otherwise be washed out by an extended source.

The tidal truncation of lens plane subhalos that we model may also reduce the overall impact of subhalos on lensing observables. We also marginalize over the power-law slope of the main deflector and simultaneously sample the macromodel parameters and the dark matter hyper-parameters. These processes introduce additional covariances in the posterior distributions, and should lead to weaker constraints on $\Sigma_{\rm sub}$ and $m_{\rm hm}$.

Other lensing studies, primarily those using the technique of gravitational imaging, have also sought to measure the subhalo mass function. Vegetti et al. (2014) inferred $\bar{f}_{\rm sub} = 0.0064^{0.0080}_{-0.0042}$ at 1σ in the mass range $4 \times 10^6 - 4 \times 10^9 M_\odot$ assuming a prior on the slope of the subhalo mass function centered on $\alpha = -1.9$, while Hezaveh et al. (2016b) constrained the normalization of subhalo mass function assuming $\alpha = -1.9$, inferring $\bar{f}_{\rm fsub}$ values comparable to the median $\bar{f}_{\rm sub} = 0.02$ result from DK2 (and our constraint), but with larger uncertainties.

To compare with the analysis of Vegetti et al. (2014), we assume a halo mass of $10^{13} M_\odot$ at a lens redshift $z_d = 0.25$ and a source at $z_{\rm src} = 0.7$, characteristic values for the lens sample analyzed by Vegetti et al. (2014). Using these values with our expression for the subhalo mass function in Equation 6.5, we obtain $f_{\rm sub} = 0.014^{+0.008}_{-0.007}$ between 4×10^6 and $6 \times 10^9 M_\odot$ at

1σ, in the same mass range used by Vegetti et al. (2014). This result is consistent with that of Vegetti et al. (2014)[15]. We quote constraints on $f_{\rm sub}$ to make comparisons with previous work, but we caution that the conclusions derived from inferences of $f_{\rm sub}$ should be interpreted with care. The physical meaning of this parameter depends on specific assumptions regarding the subhalo mass range and the contribution from dark substructure to the convergence near the Einstein radius, which may change with halo mass and redshift.

Comparing our results with semi-analytic simulations of massive $10^{13}M_{\odot}$ hosts, our results in terms of the projected mass in substructure is consistent with the `galacticus` simulations used to calibrate the evolution of the subhalo mass function with halo mass and redshift. We stress that our model was not tuned to match the normalization predicted by `galacticus`, it only made use of the trends of projected substructure mass density with host halo mass and redshift.

Our results are also consistent with N-body simulations of $10^{13}M_{\odot}$ halos by Fiacconi et al. (2016), who predict projected substructure mass densities of $2.0 - 2.8 \times 10^{7} M_{\odot}{\rm kpc}^{-2}$ after accounting for baryonic contraction of the halo. We infer roughly triple the predicted mass in substructure than the amount predicted by Xu et al. (2015), who simulated $10^{13}M_{\odot}$ halos by rescaling Milky Way size and cluster size hosts to halo masses of $\sim 10^{13}M_{\odot}$. Finally, we note that our results arrive on the heels of several works that examine numerical features of N-body simulations that may result in the artificial fragmentation of subhalos (van den Bosch et al., 2018; Errani & Peñarrubia, 2019). Taken at face value, these results suggest that N-body simulations may underpredict substructure abundance in dark matter halos.

We may also compare our constraints with the projections from Gilman et al. (2019). With a sample of ten quads, they projected a 2σ bound on $m_{\rm hm}$ with $\Sigma_{\rm sub} = 0.022{\rm kpc}^{-2}$ of $10^{7.7}M_{\odot}$ with 2% uncertainties in image fluxes, and $10^{8.6}M_{\odot}$ with 6% uncertainties. Our constraint of $m_{\rm hm} < 10^{7.8}M_{\odot}$ is broadly consistent with these predictions[16], given the higher

[15]Although Vegetti et al. (2014) did not model line of sight halos, the low lens/source redshifts their sample lessen the impact of line of sight halos on the inferred subhalo mass fraction such that we may compare our results, which include line of sight halos, with theirs.

[16]The conversion between the half-mode mass and the mass of the corresponding thermal relic dark matter particle used by Gilman et al. (2019) is off by a factor of h=0.7, but the comparison between the half-mode

mean $\Sigma_{\rm sub}$ value of $0.055{\rm kpc}^{-2}$ we infer in this analysis, and the flux uncertainties in the lens sample which are $\sim 6\%$ on average.

The overall scaling of the line of sight halo mass function $\delta_{\rm los}$ is unconstrained with our sample size and choice of prior. This is likely because the prior on $\delta_{\rm los}$ spans a relatively limited range of $\pm 20\%$ around the Sheth-Tormen mass function prediction, and with the current sample size of only eight quads we cannot constrain departures from the Sheth-Tormen prediction at the level of $10 - 20\%$.

5.7.3 Sources of systematic uncertainties

5.7.3.1 The lens macromodels

Several works (Gilman et al., 2017; Hsueh et al., 2018) have investigated the ability of smooth isothermal mass models plus external shear to fit the smooth mass component of galaxy scale strong lenses. These works reach similar conclusions, determining that isothermal models predict image flux ratios to better than 10% unless a stellar disk is present, in which case explicit modeling of the disk is required (e.g. Hsueh et al., 2017, 2018). Each of these analysis restricted the smooth lens models to exactly isothermal mass density profiles.

The deflectors in our sample show no evidence for morphological complexity that would require explicit modeling beyond a power-law ellipsoid model. Specifically, we exclude all lens systems with known stellar disks to avoid any bias they may introduce. To account for remaining uncertainties associated with the lens macromodel, we highlight two features of our lens modeling implemented in an effort to mitigate this source of systematic uncertainty. First, we note that flux ratios are highly localized probes of the surface mass density in the immediate vicinity of the lensed images, and therefore the main requirement for this work is to accurately predict the mass profile in these four small isolated regions. By relaxing the strictly isothermal mass profile assumption and marginalizing over the logarithmic slope of the main deflector mass profile, we allow for the local mass profile in the vicinity of the

masses is robust.

lensed images to vary. The additional degree of freedom added in the lens macromodel increases our uncertainties, but accounts for deviations from power-law ellipsoids limited to exact $\rho(r) \propto r^{-2}$ mass profiles.

Second, we note that smooth power-law models predict a distribution of flux ratios, rather than single values (for example, see Figures A1-A8 in Nierenberg et al. (2020)). Following common practice, Gilman et al. (2017) and Hsueh et al. (2018) identified flux ratio 'anomalies' with respect to a single smooth model fit to lensed images, a procedure that does not account for the distribution of flux ratios predicted by smooth lens models that is marginalized over in the full forward modeling analysis we perform. In this work, we also take care to explore the macromodel parameter space and the dark matter hyper-parameter space simultaneously, which accounts for additional covariances that contribute to the model-predicted flux uncertainties.

5.7.3.2 Modeling of the dark matter content

We assume specific functional forms for the halo and subhalo mass functions (Equations 6.5 and 5.9), and the mass-concentration-redshift relation (Equation 5.12). We acknowledge that there are other parameterizations in the literature for both of these quantities (e.g. Schneider et al., 2012; Benson et al., 2013), but in this work we implement only one parameterization of WDM effects on the mass function (Equation 5.11) and halo concentrations (Equation 5.12), which corresponds to one specific WDM model. We note that additional physics, such as the velocity dispersion of dark matter particles in the early Universe, can alter the shape of the mass function, but with the current sample size of lenses it is unlikely we have enough information to constrain these additional features if they were included in the model.

It is possible that free-streaming effects on the halo mass function near the half-mode mass scale may become more pronounced at high redshifts. This could affect both the location and shape of the turnover in the mass function. However, in the absence of a specific prediction for the evolution of the turnover with redshift, we apply the parameterization in Equation 5.11 through the relevant redshift range $z = 0 - 3.5$. We note that since

the lensing efficiency of halos decreases approaching source redshift, systematic errors from possible redshift evolution of the WDM turnover will be correspondingly down-weighted. We note that the mass-concentration-redshift relation for WDM calibrated by Bose et al. (2016) that we implement does evolve with redshift, as does the CDM mass-concentration relation from Diemer & Joyce (2019).

5.7.4 Implications for WDM models

Galaxy-galaxy strong lensing provides a useful compliment to the strongest existing probe of the free-streaming length of dark matter from the Lyman-α forest (Viel et al., 2013; Iršič et al., 2017). Our 2σ bound on the thermal relic mass of $m_{\rm DM} > 5.2$keV surpasses than the 3.3 keV constraint from Viel et al. (2013) and matches the 5.3keV constraint from Iršič et al. (2017), who invoked additional assumptions regarding the relevant thermodynamics. The key point of this comparison, however, is not so much which method achieves the most precision, but the fact that both methods provide stringent limits and that they are completely independent of each other in observational data and astrophysical assumptions. Independently and in combination, the results from lensing and the Lyman-α forest support the following statement: the halo mass function extends down in a scale-free manner to mass scales of $\sim 10^8 M_\odot$, where halos are mostly, if not completely, dark. There appears to be little room left for a viable warm dark matter solution to the small-scale issues of cold dark matter.

5.8 Appendix A: Convergence of the posterior distributions

The approximation of the true posterior obtained in Approximate Bayesian Computing (ABC) algorithms converges to the true posterior distribution as the acceptance criterion becomes increasingly more stringent. In our framework, changing the acceptance criterion is equivalent to reducing the number of forward model samples while keeping the number of total accepted realizations fixed. We exploit this property to test for convergence of the posteriors.

In Figures 5.12 and 5.13, we compare posteriors constructed from the full set of forward model samples with others derived from a depleted set of forward model samples, where we have discarded one-third of the realizations and accepted the same rejection criterion (accept the realizations corresponding to the 800 lowest values of $S_{\rm lens}$) to those that remain. The mass of the posterior distributions remains relatively unchanged, and the 1σ and 2σ contours are nearly identical. We conclude we have generated enough realizations of dark matter structure to reliably construct posterior distributions using the ABC rejection algorithm described in Section 5.2.

5.9 Appendix B: Obtaining deflector redshifts

The quads PS J1606 and WGD J0405 do not have measured spectroscopic redshifts, so we use photometry from Shajib et al. (2019) to obtain photometric redshift estimates. The photometry from Shajib et al. (2019) comes in three bands: F160W, F814W, and F475X with magnitude uncertainties of $0.1 - 0.3$ dex. We use the software package `eazy` (Brammer et al., 2008), and restrict the templates to only consider the SEDs for early-type galaxies, which are 90% of galaxies acting as strong lenses. We verify this procedure is accurate by applying it to other deflectors in sample analyzed by Shajib et al. (2019) that have measured spectroscopic redshifts, and then proceed to derive PDFs for deflector redshifts in the systems PS J1606 and WGD J0405.

The results are shown in Figure 5.14. The top row shows four quads from the sample analyzed in (Shajib et al., 2019) with measured spectroscopic redshifts, and the bottom row shows the pdfs output by EAZY for the systems PS J1606 and WGD J0405.

The system WFI 2026 does not have a photometric redshift, and the photometry available in the literature comes in only one or two bands with larger uncertainties. For this system, we use the equation for isothermal mass profiles relating the Einstein radius $R_{\rm Ein}$, source redshift z_s, lens redshift z_d, velocity dispersion σ and speed of light c

$$R_{\rm Ein} = 4\pi \left(\frac{\sigma}{c}\right)^2 \frac{D_{\rm ds}\left(z_d, z_s\right)}{D_{\rm s}\left(z_s\right)} \tag{5.14}$$

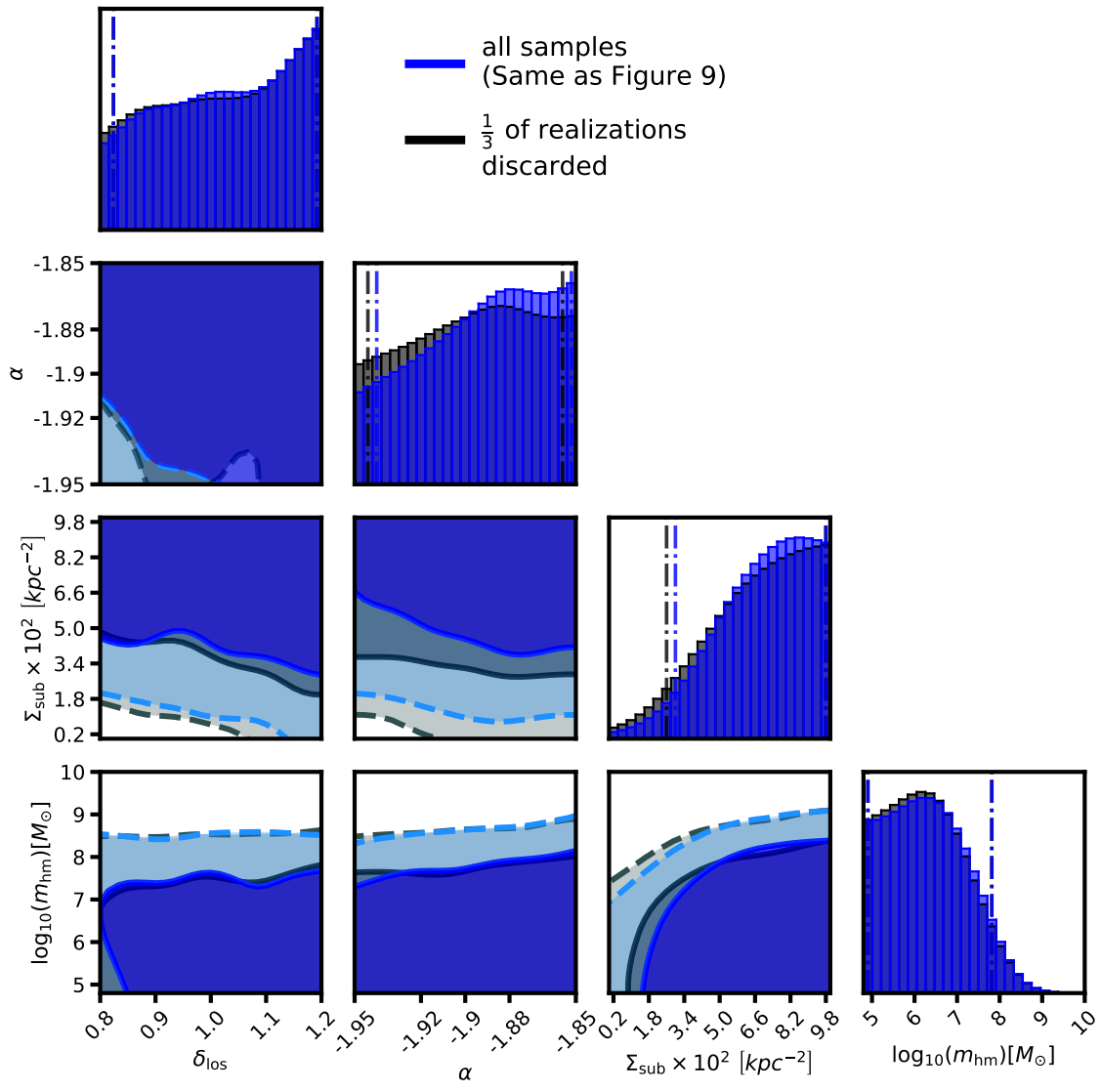

Figure 5.12: A convergence test of the posterior distributions. By discarding one-third of the forward model samples and applying the same rejection criterion to those that remain, we verify the inference obtained through the ABC rejection algorithm is robust.

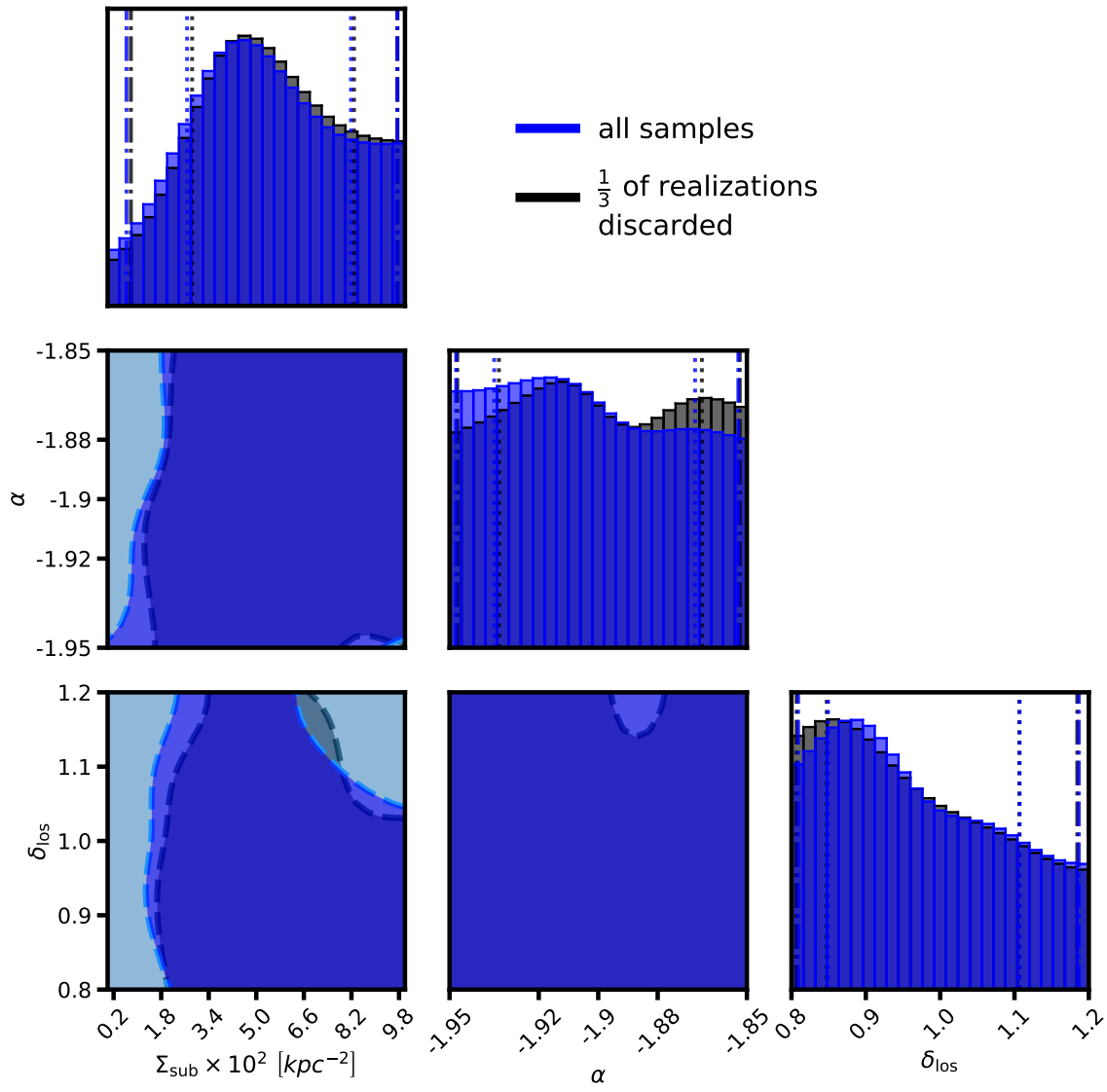

Figure 5.13: A convergence test of the posterior distributions assuming CDM. Like Figure 5.12, one-third of the samples are discarded and the same number of realizations are accepted into the posterior.

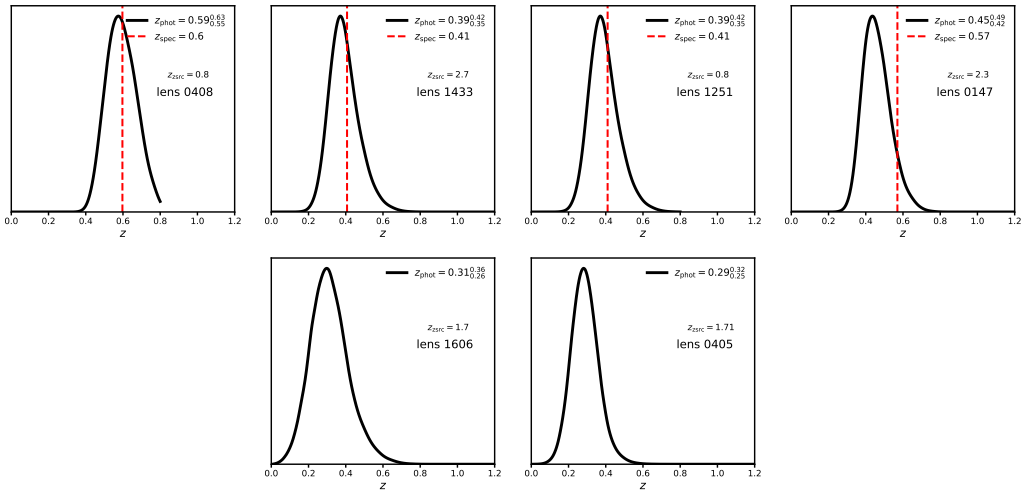

Figure 5.14: PDFs for main deflector redshifts computed with the software EAZY and photometry from Shajib et al. (2019), restricting the photometry templates to those of early-type galaxies. Top rows show four applications of this procedure to quads with measured spectroscopic redshifts (red dotted lines). The bottom row shows the results of this procedure, using the same photometry and template assumptions, applied to the quads PS J1606 and WGD J0405, which do not have spectroscopic redshift measurements.

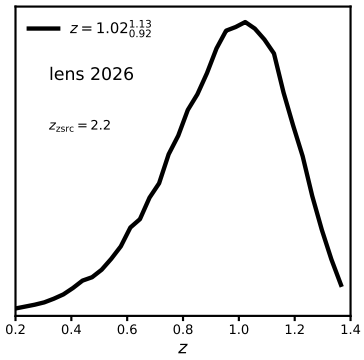

Figure 5.15: The PDF for the deflector redshift of WFI 2026 obtained by assuming a velocity dispersion of 240 ± 30 km s^{-1} and a roughly isothermal mass profile.

where $D_{\rm ds}$ and $D_{\rm s}$ are angular diameter distances between the lens and the source, and the observer and the source, respectively.

We sample a Gaussian distribution of velocity dispersions typical of early-type galaxies $240 \pm 30 {\rm kms}^{-1}$, evaluate the right hand side of Equation 5.14, and numerically solve for the lens redshift that yields the resulting angular diameter distance. The resulting PDF shown in the bottom right panel of Figure 5.15 peaks around $z_d = 1$, for the measured values $R_{\rm Ein} = 0.67$", $z_s = 2.2$. We have experimented with placing WFI 2026 at various specific redshifts, but find the posteriors for $\Sigma_{\rm sub}$, $\delta_{\rm los}$, α, and $m_{\rm hm}$ are unchanged within the uncertainties.

5.10 Appendix C: Data

We summarize the data used in this analysis, and the references for the astrometry, fluxes or flux ratios, and the corresponding uncertainties, and satellite galaxies or nearby nearby deflectors in Table 5.3.

Table 5.3: The data used in this analysis. Letters A-D correspond to the lensed images, while G is the galaxy light centroid. The priors sampled for the satellite galaxies or nearby deflectors are quoted in Table 5.2. Discovery papers are marked with a †.

Lens	Image	dRA	dDec	NL flux
WGD J0405-3308	A	1.066 ± 0.003	0.323 ± 0.003	1.00 ± 0.04
Nierenberg et al. (2020)	B	0 ± 0.003	0 ± 0.003	0.65 ± 0.04
†Anguita et al. (2018)	C	0.721 ± 0.003	1.159 ± 0.003	1.25 ± 0.03
	D	-0.157 ± 0.003	1.021 ± 0.003	1.17 ± 0.04
	G	0.358 ± 0.05	0.567 ± 0.05	-
HE0435-1223	A	2.424 ± 0.008	0.792 ± 0.008	0.97 ± 0.05
Nierenberg et al. (2017)	B	1.458 ± 0.008	-0.456 ± 0.008	0.98 ± 0.049
Wong et al. (2017)	C	0 ± 0.008	0 ± 0.008	1 ± 0.048
†Wisotzki et al. (2002)	D	0.768 ± 0.008	1.662 ± 0.008	0.54 ± 0.056
	G	1.152 ± 0.05	0.636 ± 0.05	-
RX J0911+0551	A	0 ± 0.003	0 ± 0.003	0.56 ± 0.04
Nierenberg et al. (2020)	B	0.258 ± 0.003	0.405 ± 0.003	1.00 ± 0.05
†(Bade et al., 1997)	C	-0.016 ± 0.003	0.959 ± 0.003	0.53 ± 0.04
Blackburne et al. (2011)	D	-2.971 ± 0.003	0.791 ± 0.003	0.24 ± 0.04
	G	-0.688 ± 0.05	0.517 ± 0.05	-
B1422+231	A	0.387 ± 0.005	0.315 ± 0.005	0.88 ± 0.01
Nierenberg et al. (2014)	B	0 ± 0.005	0 ± 0.005	1.00 ± 0.01
†Patnaik et al. (1992)	C	-0.362 ± 0.005	-0.728 ± 0.005	0.474 ± 0.006
	D	0.941 ± 0.01	-0.797 ± 0.01	-
	G	0.734 ± 0.01	-0.649 ± 0.01	-

Table 5.4: Table 5.3 cont.

Lens	Image	dRA	dDec	NL flux
PS J1606-2333	A	1.622 ± 0.003	0.589 ± 0.003	1.00 ± 0.03
Nierenberg et al. (2020)	B	0 ± 0.003	0 ± 0.003	1.00 ± 0.03
Shajib et al. (2019)	C	0.832 ± 0.003	-0.316 ± 0.003	0.59 ± 0.02
†Lemon et al. (2018)	D	0.495 ± 0.003	0.739 ± 0.003	0.79 ± 0.02
	G	0.784 ± 0.05	0.211 ± 0.05	-
WFI 2026-4536	A	0.164 ± 0.003	-1.428 ± 0.003	1.00 ± 0.02
Nierenberg et al. (2020)	B	0.417 ± 0.003	-1.213 ± 0.003	0.75 ± 0.02
†Morgan et al. (2004)	C	0 ± 0.003	0 ± 0.003	0.31 ± 0.02
	D	-0.571 ± 0.003	-1.044 ± 0.003	0.28 ± 0.01
	G	-0.023 ± 0.05	-0.865 ± 0.05	-
WFI 2033-4723	A	-2.196 ± 0.003	1.260 ± 0.003	1.00 ± 0.03
Nierenberg et al. (2020)	B	-1.484 ± 0.003	1.375 ± 0.003	0.65 ± 0.03
Vuissoz et al. (2008)	C	0 ± 0.003	0 ± 0.003	0.50 ± 0.02
†Morgan et al. (2004)	D	-2.113 ± 0.003	-0.278 ± 0.003	0.53 ± 0.02
	G	-1.445 ± 0.05	2.344 ± 0.05	-
WGD 2038-4008	A	-2.306 ± 0.003	1.708 ± 0.003	1.00 ± 0.01
Nierenberg et al. (2020)	B	0 ± 0.003	0 ± 0.003	1.16 ± 0.02
†Agnello et al. (2018)	C	-1.518 ± 0.003	0.029 ± 0.003	0.92 ± 0.02
	D	-0.126 ± 0.003	2.089 ± 0.003	0.46 ± 0.01
	G	-0.832 ± 0.05	1.220 ± 0.05	-

CHAPTER 6

Constraints on the mass-concentration relation of cold dark matter halos with 11 strong gravitational lenses

This chapter was published as Gilman, D., et al. Constraints on the mass-concentration relation of cold dark matter halos with 11 strong gravitational lenses. MNRAS 492, L12-L16 (2020), and is printed here with minor formatting adjustments.

6.1 Introduction

Dark matter structure formation in cold dark matter (CDM) cosmologies proceeds hierarchically. Small peaks in the density field collapse first, followed by the collapse of over-densities on larger scales and mergers between collapsed halos (Navarro et al., 1997; Moore et al., 1999). The scale-free nature of structure formation in CDM scenarios results in self-similar density profiles for individual dark matter halos, first pointed out by Navarro et al. (1996) (hereafter NFW). The concentration parameter, defined as the ratio of the virial radius of the halo to its scale radius $c \equiv \frac{r_{\rm vir}}{r_s}$, determines the density profile of NFW halos $\rho(r)$

$$\frac{\rho(r)}{\rho_{\rm crit}} = \frac{200}{3} \frac{1}{x(1+x)^2} \frac{c^3}{\ln(1+c) - \frac{c}{1+c}} \tag{6.1}$$

where $\rho_{\rm crit}$ is the critical density of the Universe today, $x = \frac{r}{r_s}$, and where we define the boundary of the halo as the virial radius r_{200} enclosing a mean density $200\rho_{\rm crit}$. The function $c(M, z)$ relates the concentration of a halo to its mass and redshift, and is known as the mass-concentration relation.

Navarro et al. (1997) argued that the anti-correlation between halo mass and concentration seen in N-body simulations reflects the collapse epoch of the halo, with the high concentrations of low-mass objects reflective of the higher background density of the Universe at early times when the majority of these small over-densities collapsed. The logarithmic slope of the matter power-spectrum $P(k) \propto k^n$ is also understood to affect halo concentrations, with larger n resulting in more centrally concentrated low-mass halos (Eke et al., 2001). These realizations provided a starting point for attempts at predicting the mass-concentration relation of cold dark matter halos (Bullock et al., 2001; Wechsler et al., 2002; Prada et al., 2012; van den Bosch et al., 2014; Ludlow et al., 2014; Diemer & Joyce, 2019). As the various models for halo concentrations depend on the mass accretion history of dark matter halos and the matter-power spectrum, the mass-concentration relation encodes information regarding the process of dark matter structure formation in the Universe. Halo concentrations also play a central role in determining the dark matter annihilation signals from dwarf galaxies (Strigari et al., 2007).

Despite progress over the past two decades in identifying the astrophysical process shaping the mass-concentration relation, a first-principles derivation does not exist and the form of the mass-concentration relation on mass scales below $10^9 M_\odot$ remains unconstrained by observations. We remedy this situation using the flux ratios and images positions from 11 quadruple-image strong gravitational lenses to constrain the mass-concentration relation on scales $\leq 10^8 M_\odot$, where halos are expected to be mostly devoid of stars and gas (Sawala et al., 2016a). Strong lensing is a powerful tool for constraining the abundance and structure of dark matter halos as it measures dark matter structures directly, without relying on luminous matter to trace the dark matter (Dalal & Kochanek, 2002; Vegetti et al., 2014; Nierenberg et al., 2014; Birrer et al., 2017b; Hsueh et al., 2019; Gilman et al., 2020). In this letter, we deploy the statistical machinery developed and tested by Gilman et al. (2019) to constrain the mass-concentration relation of CDM halos on scales below $10^8 M_\odot$ at cosmological distance.

This letter is organized as follows. In Section 6.2, we review the parameterizations of the subhalo and halo mass functions, and describe the parameterization of the mass-concentration relation we constrain in this work. We also briefly discuss the observable

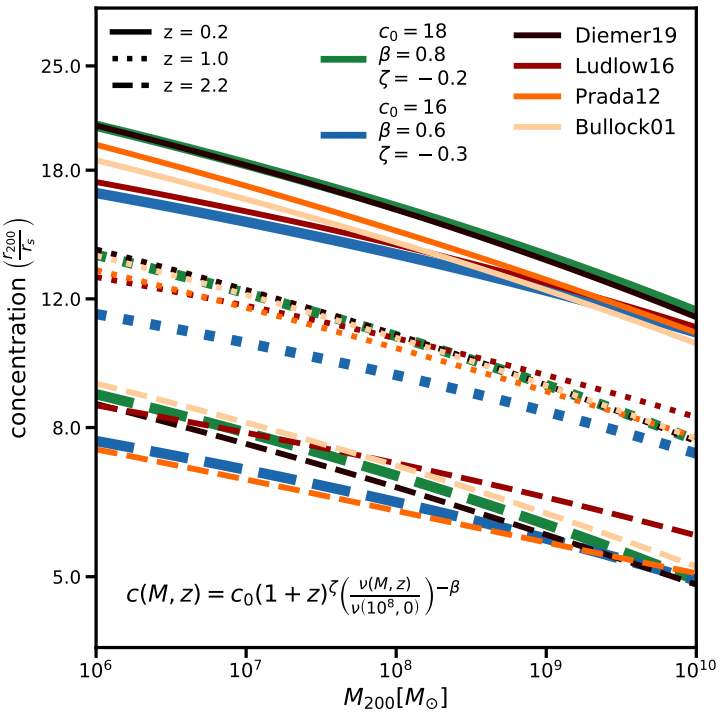

Figure 6.1: Mass-concentration relations from the literature as a function of halo mass and redshift, compared with the functional form for the relation in Equation 6.3. The parameterization of the mass-concentration relation we constrain in this work has a variable normalization c_0 and logarithmic slope β, with a redshift evolution modified by an empirical factor ζ. Models plotted from the literature are valid in the mass range shown in the figure.

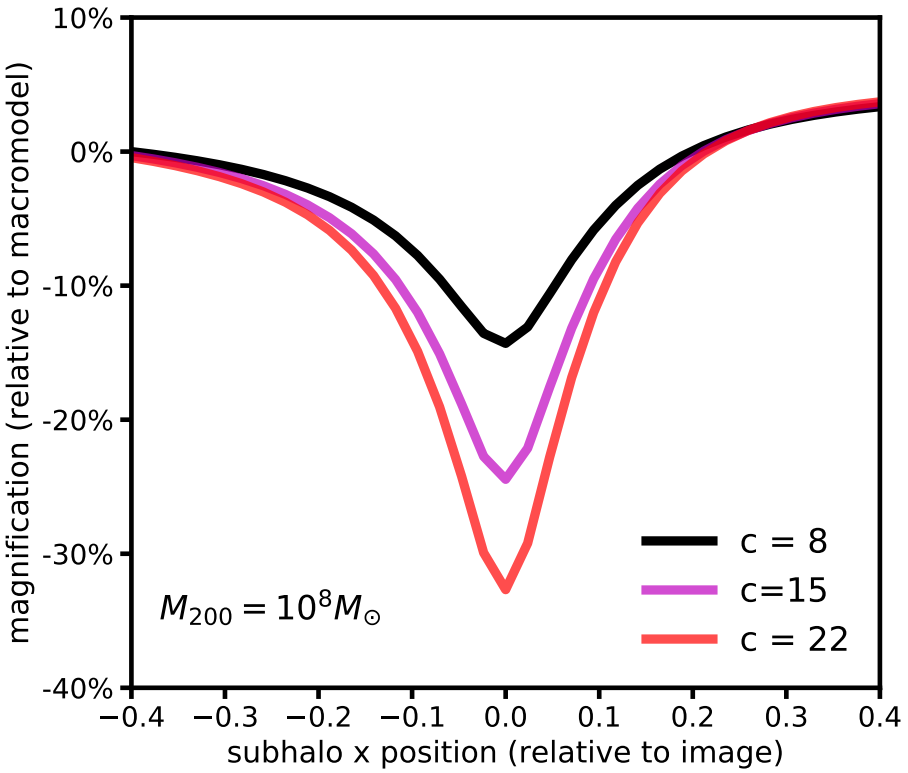

Figure 6.2: Magnification perturbation cross section of a $10^8 M_\odot$ halo at $z = 0.5$ with various concentrations for a 30 pc background source at $z = 1.5$. More concentrated halos are more efficient lenses, resulting in stronger flux perturbations.

signatures of halo concentrations on image flux ratios, and briefly review our Bayesian inference methodology. For more detailed discussion of the mass functions and inference method, we defer to the text of Gilman et al. (2020). Section 6.3 discusses the data used in this analysis. In Section 6.4, we present our main results, and we provide concluding remarks in Section 6.5.

Lensing computations are carried out using `lenstronomy`[1] (Birrer & Amara, 2018). Computations involving the halo mass function and the matter power spectrum are performed with `colossus` (Diemer, 2018). We assume a standard cosmology using the parameters from WMAP9 (Hinshaw et al., 2013) ($\Omega_m = 0.28, \sigma_8 = 0.82, h = 0.7$).

6.2 Modeling strategy and inference method

In this section we describe the modeling of the CDM subhalo and line of sight halo mass functions, and a model for the mass-concentration relation of CDM halos expressed in terms of the peak height ν. In Sections 2 and 3 of (Gilman et al., 2020) we provide additional detail regarding the Bayesian inference methodology and the mass function parameterizations.

6.2.1 A model for the CDM mass-concentration relation for field halos

The mass function for field halos[2] is parameterized in terms of the Sheth-Tormen (ST) halo mass function (Sheth et al., 2001) $\frac{d^2N}{dmdV}\big|_{\rm ShethTormen}$

$$\frac{d^2N_{\rm los}}{dmdV} = \delta_{\rm los}\big(1 + \xi_{\rm 2halo}\left(M_{\rm halo}, z\right)\big)\frac{d^2N}{dmdV}\big|_{\rm ShethTormen}. \tag{6.2}$$

The parameter $\delta_{\rm los}$ is an overall scaling term that accounts for a systematic shift the mean number of halos predicted by the ST mass function, and $\xi_{\rm 2halo}$ is the two-halo term that introduces additional correlated structure around the host dark matter halo (Gilman et al., 2020).

[1] https://github.com/sibirrer/lenstronomy

[2] In lensing nomenclature these are also referred to as 'line of sight' halos.

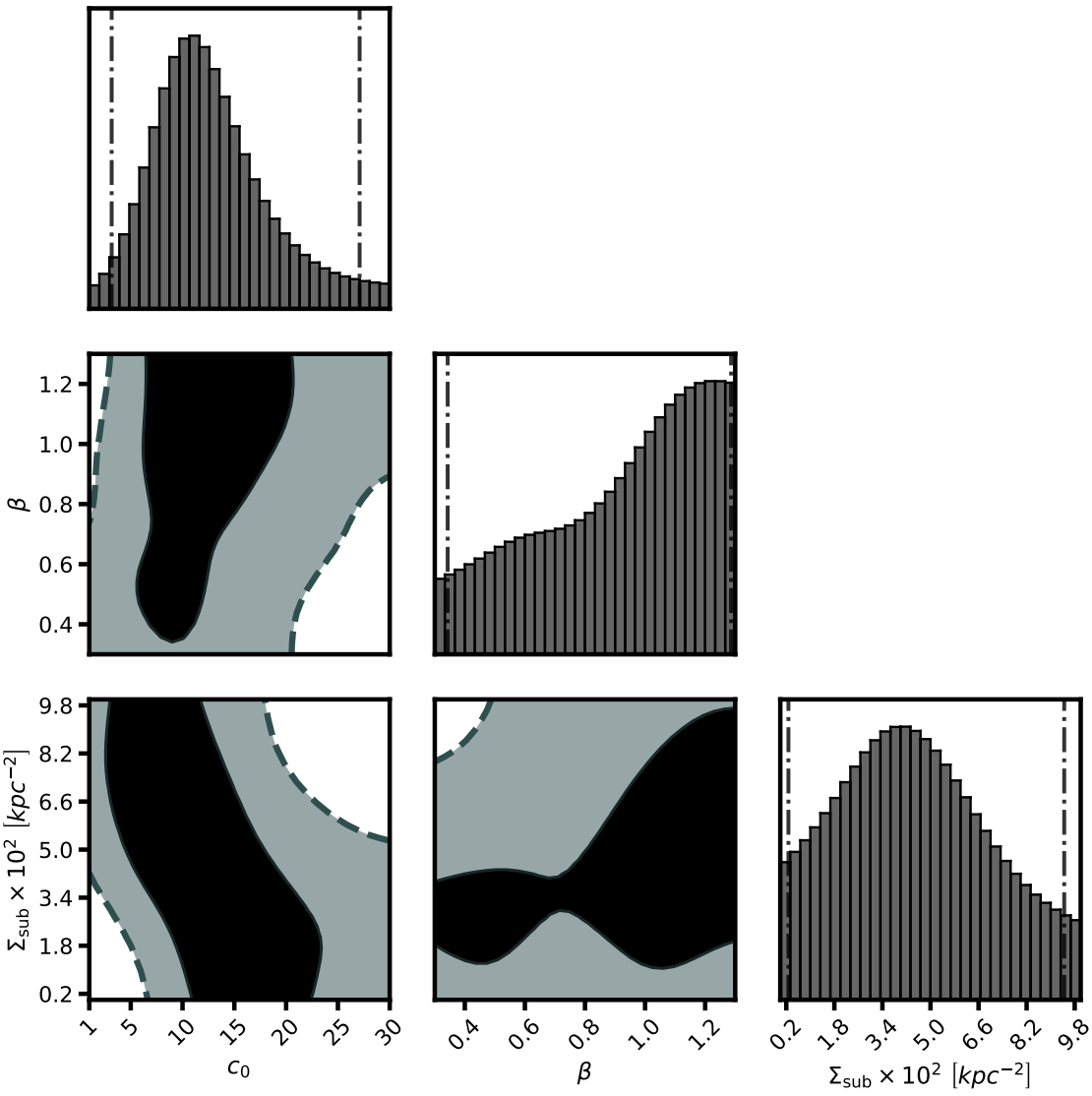

Figure 6.3: Constraints on the normalization c_0 and the logarithmic slope β of the mass-concentration relation in Equation 6.3. We include the constraints on the normalization of the subhalo mass function $\Sigma_{\rm sub}$, as it is covariant with both c_0 and β. Contours show 68% and 95% confidence intervals. The parameter ζ, for which we use a Gaussian prior $\mathcal{N}(-\prime.\in\bigtriangledown, \prime.\prime\bigtriangledown)$, is unconstrained. Vertical bars represent 95% confidence intervals.

Halos in the field by definition follow the median mass-concentration relation, which we parameterize as

$$c\left(M,z\right)=c_{0}\left(1+z\right)^{\zeta}\left(\frac{\nu\left(M,z\right)}{\nu\left(10^{8},0\right)}\right)^{-\beta} \tag{6.3}$$

with a scatter of 0.1 dex (Dutton & Macciò, 2014). The above relation is expressed as a power-law with slope β in terms of the peak height ν at a particular length scale $R=\left(\frac{3M}{4\pi\rho_{\mathrm{m},0}}\right)^{\frac{1}{3}}$

$$\nu\left(R,z\right)=\frac{\delta_{c}}{\sigma\left(R,z\right)}$$

where $\delta_{c}=1.686$ is the threshold for spherical collapse in an Einstein de-Sitter universe, $\rho_{\mathrm{m},0}$ is the component of the critical density of the Universe in matter at $z=0$, and $\sigma\left(R,z\right)$ is the variance of the matter density field on the scale R. The variance depends on the linear matter power spectrum $P\left(k,z\right)$ through

$$\sigma^{2}\left(R,z\right)=\frac{1}{2\pi^{2}}\int_{0}^{\infty}k^{2}P\left(k,z\right)\left|\tilde{W}\left(kR\right)\right|^{2}dk \tag{6.4}$$

where $\tilde{W}\left(kR\right)$ is the Fourier transform of the spherical top-hat window function. The parameter c_{0} anchors the normalization to that of a $10^{8}M_{\odot}$ halo at $z=0$. We introduce the factor $\left(1+z\right)^{\zeta}$ to account for additional redshift evolution, similar to the empirical approach employed by Prada et al. (2012). We marginalize over a Gaussian prior on ζ with mean -0.25 and variance 0.05, which tracks the redshift evolution of the theoretical mass-concentration relations.

In Figure 6.1 we show the mass-concentration relation in Equation 6.3 as blue and green curves alongside several models from the literature (Bullock et al., 2001; Prada et al., 2012; Ludlow et al., 2016; Diemer & Joyce, 2019).

6.2.2 Mass-concentration relation for subhalos and the subhalo mass function

When a field halo is accreted into a more massive host, it becomes a subhalo and ceases to evolve through 'pseudo-evolution', which refers to changing halo concentrations due to the evolving background density of the Universe while the density normalization and scale radius

remain fixed in physical coordinates (Diemer et al., 2013). For subhalos, the concentration defined in terms of r_{200} becomes an ill-defined concept, and the structure of the subhalo evolves through tidal stripping effects that alter the density profile (Errani et al., 2017). A complete prescription for subhalo density profiles requires a model for how tidal stripping evolves a subhalo whose physical parameters are determined at the time of infall (Green & van den Bosch, 2019). Without a detailed prescription for this effect implemented at the present time, and given the need for the number of free parameters to match the statistical constraining power of the current sample size of lenses, we simply evaluate subhalo concentrations at the time of infall when the mass-concentration relation in Equation 6.3 is valid. To this end, we sample a probability density for the infall redshift as a function of halo mass and the main deflector redshift using the semi-analytic modeling code `galacticus` (Benson, 2012).

We render subhalos from a mass function parameterized as

$$\frac{d^2 N_{\rm sub}}{dm dA} = \frac{\Sigma_{\rm sub}}{m_0} \left(\frac{m}{m_0} \right)^{\alpha} \mathcal{F}\left(M_{\rm halo}, z \right), \tag{6.5}$$

where the scaling function $\mathcal{F}\left(\mathcal{M}_{\rm halo}, \ddagger \right)$

$$\log_{10}\left(\mathcal{F} \right) = k_1 \log_{10} \left(\frac{M_{\rm halo}}{10^{13} M_{\odot}} \right) + k_2 \log_{10}\left(z + 0.5 \right) \tag{6.6}$$

accounts for evolution of the differential projected number density of subhalos with host halo mass $M_{\rm halo}$ and redshift. The fit $k_1 = 0.88$ and $k_2 = 1.7$ is determined from a suite of simulated host halos and their substructure generated with `galacticus`. We add a tidal truncation radius to subhalo density profiles that depends on the mass of the subhalo and its position inside the host halo (Gilman et al., 2020).

6.2.3 Where does the lensing signal come from?

We show the magnification cross section for a $10^8 M_{\odot}$ halo as a function of its concentration in Figure 6.2. The maximum magnification perturbation from a halo with $c = 8$ is 10%, while the perturbation from a halo with $c = 22$ reaches 30%. More concentrated halos will increase the frequency of flux-ratio perturbations relative to a population of low-concentration halos, as more concentrated halos are more efficient lenses.

6.2.4 Forward modeling methodology

The most important conceptual feature of our Bayesian inference technique is the recognition that we may obtain posterior distributions of model parameters from simulated datasets generated with a forward model, circumventing the direct computation of an intractable likelihood function. The forward modeling technique detailed by Gilman et al. (2020) simultaneously samples the dark matter quantities of interest and nuisance parameters such as the logarithmic slope of the main deflector mass profile $\gamma_{\rm macro}$ and the extent of the lensed background source $\sigma_{\rm src}$. Comparisons between the forward model output and the observed data are performed through the use of a summary statistic, which is used to estimate the likelihoods for each lens and compute the posterior.

We use a uniform prior on c_0 and β between $1-30$ and $0.3-1.3$, respectively, a Gaussian prior on ζ with mean -0.25 and variance 0.05. Comparing the theoretical predictions for the mass-concentration relation shown in Figure 1 with the blue and green curves from the parametric model in Equation 6.3, the differences between the various theoretical predictions in the literature for the mass-concentration relation are smaller than our choice of prior on c_0 and β by factors of about 10 and 5, respectively. Though our choice of prior on these quantities, we therefore implicitly assume that theoretical predictions are accurate to this degree of precision.

Based on predictions from N-body simulations (Springel et al., 2008; Fiacconi et al., 2016), we use a Gaussian prior on the slope of the subhalo mass function α with mean (variance) -1.9 (0.025), and a flat prior on the overall normalization of the subhalo mass function $\Sigma_{\rm sub}$ between $0-0.1{\rm kpc}^{-2}$ (see also Figure 2 in Gilman et al. (2020) for predictions from `galacticus`). For reference, N-body simulations and `galacticus` predict values of $\Sigma_{\rm sub}$ in the range 0.01 - $0.04{\rm kpc}^{-2}$. We defer to Gilman et al. (2020) for further details regarding the choice of priors.

6.3 Data

We use the image positions and flux ratios from eleven quadruply-imaged quasars. Eight of these systems have measured narrow-line emission (Nierenberg et al., 2014, 2017, 2020), while three (B0128+437, MG0414+0543, and PG 1115+080) are measured in radio wavelengths (Koopmans et al., 2003; Katz et al., 1997; Chiba et al., 2005). Both the narrow-line systems and radio lenses have background source sizes large enough ($\sim 1 - 60$pc) to avoid contaminating effects from micro-lensing, while retaining sensitivity to dark matter halos in the mass range $10^6 - 10^9 M_\odot$.

We assume the mean host halo mass inferred in strong lens systems $\log_{10}(M_{\rm halo}) = 13.3$ derived by (Lagattuta et al., 2010) for B0128+437, and $\log_{10}(M_{\rm halo}) = 13.0$ and 13.5 for MG0414+0543 and PG 1115+080, respectively, each with variance 0.3 dex. We defer to Gilman et al. (2020) for details regarding the derivation of these priors. We model the luminous satellite galaxy visible near MG0414+0543 (Ros et al., 2000) with a Gaussian prior the Einstein radius $\mathcal{N}(\text{ı.}\in, \text{ı.ı}_\bigtriangledown)$ and the mass centroid, with astrometric uncertainties of 50 m.a.s. We use uniform priors on the background source size of $1 - 25$pc ($25 - 60$pc) for the radio (narrow-line) lenses.

6.4 Results

Figure 6.3 shows posterior distributions for c_0, β, and $\Sigma_{\rm sub}$. The parameters c_0 and $\Sigma_{\rm sub}$ are covariant, as more concentrated halos and more numerous halos both act to increase the clumpiness of dark matter structure on small scales, and hence the lensing signal. Similarly, steeper logarithmic slopes β lower the concentrations of halos with mass above $10^8 M_\odot$ above the point point in Equation 6.3, which can be partially accounted for by increasing the overall normalization c_0.

We sample the posterior distribution of hyper-parameters in Figure 6.3 to obtain constraints on the halo concentrations as a function of mass, and show the results in Figure 6.4. Our constraints are consistent with the subset of models from the literature applicable in the

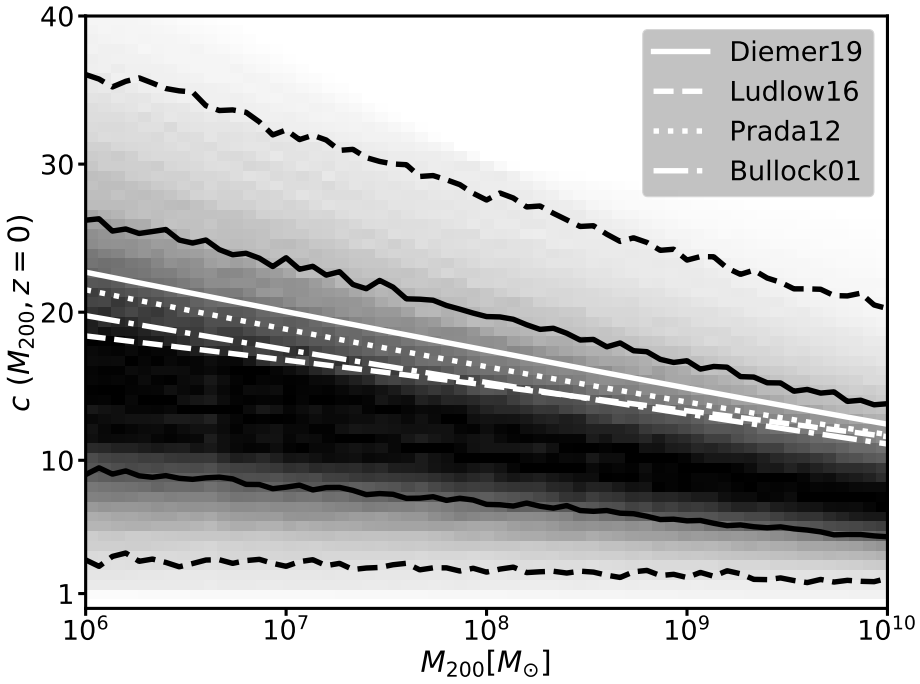

Figure 6.4: Constraints on the concentration-mass relation of CDM halos derived from the posterior distribution of hyper-parameters shown in Figure 6.3, computed with eleven strong gravitational lenses. Black solid (dashed) lines contain 68% (95%) confidence intervals at fixed halo mass. White curves show several concentration-mass relations from the literature also plotted in Figure 6.1.

halo mass range $10^6 - 10^{10} M_\odot$ relevant for our analysis. The uncertainties on concentrations of low-mass halos are larger than those of high mass halos, a result of the correlation between c_0 and β in the posterior distribution in Figure 6.3. The inference on the parameters describing the mass-concentration relation are marginalized over the normalization of the subhalo mass function, the amplitude of the line of sight halo mass function, the slope of the subhalo mass function, and nuisance parameters describing the main deflector lens models and the background source size, given our choice of priors and their motivation detailed by Gilman et al. (2020).

6.5 Discussion and conclusions

We have extended the Bayesian inference framework detailed by Gilman et al. (2020) to accommodate a variable mass-concentration relation assuming a CDM mass function, and constrain the parameters describing this relation on sub-galactic scales using 11 quadruple-image strong gravitational lenses. We have used priors on the parameters describing the mass-concentration relation assuming theoretical predictions from the literature are accurate to within a factor of ten in the overall normalization c_0, and to within a factor of five in the logarithmic slope β. Given these assumptions, our main results are summarized as follows:

- We constrain the normalization of the mass concentration relation c_0, defined as the concentration of a $10^8 M_\odot$ halo at $z = 0$. At 68% CI, $c = 12^{+6}_{-5}$, and at 95% CI $c = 12^{+15}_{-9}$.
- We convert the constraints on the hyper-parameters describing the mass-concentration relation into physical halo concentrations as a function of halo mass. At 68% (95%), the concentration of a $10^7 M_\odot$ halo is $c = 15^{+9}_{-8} (c = 15^{+18}_{-11})$, while for a $10^9 M_\odot$ halo $c = 10^{+7}_{-4} (c = 10^{+14}_{-7})$.

The results of this paper conclusively establish strong gravitational lensing by galaxies as perhaps the only probe of the mass-concentration relation of dark matter halos across cosmological distance on mass scales where they are expected to be completely or mostly dark. Dedicated studies from N-body simulations and semi-analytic models will be necessary

to refine theoretical predictions for the complicated processes that can shape the density profiles of dark matter subhalos in tidal fields. The sample size of strong lens suitable for the kind of study carried out in this work will increase by order of magnitude (Oguri & Marshall, 2010; Treu et al., 2018) in the coming decade, allowing strong lensing to constrain these processes.

Acknowledgments

We thank James Bullock and Benedikt Diemer for interesting discussions, and the referee for comments that improved the presentation of our results.

DG, TT, and SB acknowledge support by the US National Science Foundation through grant AST-1714953. DG, TT, SB and AN acknowledge support from HST-GO-15177. AJB and XD acknowledge support from NASA ATP grant 17-ATP17-0120. Support for Program number GO-15177 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. TT and AN acknowledge support from HST-GO-13732. AN acknowledges support from the NASA Postdoctoral Program Fellowship, the UC Irvine Chancellor's Fellowship, and the Center for Cosmology and Astroparticle Physics Fellowship.

CHAPTER 7

Conclusions

In my thesis, I developed techniques that synthesize data from quadruply-imaged quasars to place constraints on any dark matter model by constraining the shape of the halo mass function, and the density profiles of individual dark matter halos. The multi-year effort of developing the necessary analysis tools culminated in two main results. First, we placed a an upper bound on the free-streaming length of 11kpc, assuming a mass function based a thermal relic dark matter particle (Schneider et al., 2012). The corresponding constraint on the particle mass sets a lower bound of 5.2keV, one of the strongest constraints on Warm Dark Matter to date. Second, we placed the first observational constraint on the mass-concentration relation of Cold Dark Matter halos on sub-galactic scales across cosmological distance. Neither of these results would have been possible without pioneering work to measure narrow-line flux ratios, presented in Nierenberg et al. (2020).

The constraints on the free-streaming length and the mass-concentration arrived quickly on each other's heels, published only a few months apart. Once the analysis framework, which took years to develop, was in place, it was rapidly deployed to test two quite different parameterizations of the mass function and halo density profiles. This illustrates both the power of strong lensing as a probe of fundamental physics, and the flexibility of the tools I developed to efficiently exploit this power.

The results I have presented in this dissertation suggest new research avenues to pursue. The flexibility of my method accommodates any parameterization of halo density profiles, including the cored profiles characteristic of self-interacting and ultra-light 'fuzzy' dark matter, as well as the intrinsically point-like mass profiles of primordial black holes. In coming years, the constraining power of strong lensing analyses over fundamental dark matter physics will

increase in tandem with the continuously-growing sample size of strongly-lensed quasars with reliable, high-quality data. This enhanced constraining power could enable precise measurements of the amplitude of subhalo mass function, constraining tidal stripping processes that determine the observed number counts of dwarf galaxies. Finally, although strong lensing is often presented as direct competitor with 'rival' techniques that rely on the Lyman-α forest and stellar streams, the strongest bounds will result from analyses that combine each of these independent probes to place joint constraints on the nature of dark matter through structure formation arguments.

Bibliography

Abazajian K. N., 2017, Phys. Rept., 711, 1

Abazajian K. N., Kusenko A., 2019, arXiv e-prints, p. arXiv:1907.11696

Agnello A., et al., 2018, MNRAS, 479, 4345

Akeret J., Refregier A., Amara A., Seehars S., Hasner C., 2015, JCAP, 8, 043

Amara A., Metcalf R. B., Cox T. J., Ostriker J. P., 2006, MNRAS, 367, 1367

Andrade K. E., Minor Q., Nierenberg A., Kaplinghat M., 2019, arXiv e-prints, p. arXiv:1901.00507

Anguita T., et al., 2018, MNRAS, 480, 5017

Angulo R. E., Hahn O., Ludlow A. D., Bonoli S., 2017, MNRAS, 471, 4687

Aprile E., et al., 2018, Phys. Rev. Lett., 121, 111302

Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, ApJ, 724, 511

Avila-Reese V., Colín P., Valenzuela O., D'Onghia E., Firmani C., 2001, ApJ, 559, 516

Bade N., Siebert J., Lopez S., Voges W., Reimers D., 1997, AA, 317, L13

Baltz E. A., Marshall P., Oguri M., 2009, JCAP, 1, 015

Bayer D., Chatterjee S., Koopmans L. V. E., Vegetti S., McKean J. P., Treu T., Fassnacht C. D., 2018, arXiv e-prints, p. arXiv:1803.05952

Beaumont M. A., Zhang W., Balding D. J., 2002, Genetics, 162, 2025

Benson A. J., 2012, NewA, 17, 175

Benson A. J., et al., 2013, MNRAS, 428, 1774

Biggs A. D., Browne I. W. A., Jackson N. J., York T., Norbury M. A., McKean J. P., Phillips P. M., 2004, MNRAS, 350, 949

Birrer S., Amara A., 2018, Physics of the Dark Universe, 22, 189

Birrer S., Amara A., Refregier A., 2015, ApJ, 813, 102

Birrer S., Welschen C., Amara A., Refregier A., 2017a, JCAP, 4, 049

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

Birrer S., et al., 2019, MNRAS, 484, 4726

Blackburne J. A., Pooley D., Rappaport S., Schechter P. L., 2011, ApJ, 729, 34

Blandford R., Narayan R., 1986, ApJ, 310, 568

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

Bose S., Hellwing W. A., Frenk C. S., Jenkins A., Lovell M. R., Helly J. C., Li B., 2016, MNRAS, 455, 318

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40

Bozek B., et al., 2019, MNRAS, 483, 4086

Bradač M., Schneider P., Lombardi M., Steinmetz M., Koopmans L. V. E., Navarro J. F., 2004, AA, 423, 797

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Brennan S., Benson A. J., Cyr-Racine F.-Y., Keeton C. R., Moustakas L. A., Pullen A. R., 2018, arXiv e-prints, p. arXiv:1808.03501

Bringmann T., Kahlhoefer F., Schmidt-Hoberg K., Walia P., 2017, Physical Review Letters, 118, 141802

Bullock J. S., Boylan-Kolchin M., 2017, Annual Review of Astronomy and Astrophysics, 55, 343

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559

Cappellari M., 2002, MNRAS, 333, 400

Carter D., et al., 2008, ApJS, 176, 424

Castellano M., Menci N., Grazian A., Merle A., Sanchez N. G., Schneider A., Totzauer M., 2019, arXiv e-prints, p. arXiv:1903.12580

Chen J., Kravtsov A. V., Keeton C. R., 2003, ApJ, 592, 24

Chen J., Rozo E., Dalal N., Taylor J. E., 2007, ApJ, 659, 52

Chiba M., 2002, ApJ, 565, 17

Chiba M., Minezaki T., Kashikawa N., Kataza H., Inoue K. T., 2005, ApJ, 627, 53

Cohn J. D., Kochanek C. S., McLeod B. A., Keeton C. R., 2001, ApJ, 554, 1216

Colombi S., Dodelson S., Widrow L. M., 1996, ApJ, 458, 1

Combes F., et al., 2019, AA, 623, A79

Congdon A. B., Keeton C. R., 2005, MNRAS, 364, 1459

Cotner E., Kusenko A., 2017, Phys. Rev. Lett., 119, 031103

Csilléry K., Blum M. G. B., Gaggiotti O. E., François O., 2010, Trends in ecology & evolution, 25, 410—418

Cyr-Racine F.-Y., de Putter R., Raccanelli A., Sigurdson K., 2014, PhysRevD, 89, 063517

Cyr-Racine F.-Y., Moustakas L. A., Keeton C. R., Sigurdson K., Gilman D. A., 2016, PhysRevD, 94, 043505

Cyr-Racine F.-Y., Keeton C. R., Moustakas L. A., 2019, PhysRevD, 100, 023013

Dalal N., Kochanek C. S., 2002, ApJ, 572, 25

Davies F. B., Hennawi J. F., Eilers A.-C., Lukić Z., 2018, ApJ, 855, 106

Daylan T., Cyr-Racine F.-Y., Diaz Rivero A., Dvorkin C., Finkbeiner D. P., 2018, ApJ, 854, 141

Despali G., Vegetti S., 2017, MNRAS, 469, 1997

Despali G., Giocoli C., Angulo R. E., Tormen G., Sheth R. K., Baso G., Moscardini L., 2016, MNRAS, 456, 2486

Despali G., Sparre M., Vegetti S., Vogelsberger M., Zavala J., Marinacci F., 2018a, arXiv e-prints,

Despali G., Vegetti S., White S. D. M., Giocoli C., van den Bosch F. C., 2018b, MNRAS, 475, 5424

Díaz Rivero A., Dvorkin C., Cyr-Racine F.-Y., Zavala J., Vogelsberger M., 2018, PhysRevD, 98, 103517

Diemer B., 2018, ApJS, 239, 35

Diemer B., Joyce M., 2019, ApJ, 871, 168

Diemer B., More S., Kravtsov A. V., 2013, ApJ, 766, 25

Dobler G., Keeton C. R., 2006, MNRAS, 365, 1243

Dodelson S., Widrow L. M., 1994, Phys. Rev. Lett., 72, 17

Dooley G. A., Peter A. H. G., Carlin J. L., Frebel A., Bechtol K., Willman B., 2017, MNRAS, 472, 1060

Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359

Dutton A. A., Treu T., 2014, MNRAS, 438, 3594

Eke V. R., Navarro J. F., Steinmetz M., 2001, ApJ, 554, 114

Elitzur M., Shlosman I., 2006, ApJ, 648, L101

Errani R., Peñarrubia J., 2019, arXiv e-prints, p. arXiv:1906.01642

Errani R., Peñarrubia J., Laporte C. F. P., Gómez F. A., 2017, MNRAS, 465, L59

Evans N. W., Witt H. J., 2003, MNRAS, 345, 1351

Fadely R., Keeton C. R., 2012, MNRAS, 419, 936

Fassnacht C. D., Womble D. S., Neugebauer G., Browne I. W. A., Readhead A. C. S., Matthews K., Pearson T. J., 1996, ApJL, 460, L103

Fassnacht C. D., et al., 1999, AJ, 117, 658

Fassnacht C. D., Xanthopoulos E., Koopmans L. V. E., Rusin D., 2002, ApJ, 581, 823

Ferrarese L., et al., 2006, ApJS, 164, 334

Fiacconi D., Madau P., Potter D., Stadel J., 2016, ApJ, 824, 144

Gallo E., Treu T., Jacob J., Woo J.-H., Marshall P. J., Antonucci R., 2008, ApJ, 680, 154

Gao L., Frenk C. S., Boylan-Kolchin M., Jenkins A., Springel V., White S. D. M., 2011, MNRAS, 410, 2309

Garrison-Kimmel S., et al., 2017, MNRAS, 471, 1709

Garzilli A., Ruchayskiy O., Magalich A., Boyarsky A., 2019, arXiv e-prints, p. arXiv:1912.09397

Gavazzi R., Treu T., Rhodes J. D., Koopmans L. V. E., Bolton A. S., Burles S., Massey R. J., Moustakas L. A., 2007, ApJ, 667, 176

Gilman D., Agnello A., Treu T., Keeton C. R., Nierenberg A. M., 2017, MNRAS, 467, 3970

Gilman D., Birrer S., Treu T., Keeton C. R., Nierenberg A., 2018, MNRAS, 481, 819

Gilman D., Birrer S., Treu T., Nierenberg A., Benson A., 2019, MNRAS, p. 1618

Gilman D., Birrer S., Nierenberg A., Treu T., Du X., Benson A., 2020, MNRAS, 491, 6077

Goulding A. D., et al., 2016, ApJ, 826, 167

Graus A. S., Bullock J. S., Boylan-Kolchin M., Nierenberg A. M., 2018, MNRAS, 480, 1322

Green S. B., van den Bosch F. C., 2019, MNRAS, 490, 2091

Hahn C., Vakili M., Walsh K., Hearin A. P., Hogg D. W., Campbell D., 2017, MNRAS, 469, 2791

Han J., Cole S., Frenk C. S., Jing Y., 2016, MNRAS, 457, 1208

Herbel J., Kacprzak T., Amara A., Refregier A., Bruderer C., Nicola A., 2017, JCAP, 2017, 035

Hewitt J. N., Turner E. L., Lawrence C. R., Schneider D. P., Brody J. P., 1992, AJ, 104, 968

Hezaveh Y., Dalal N., Holder G., Kisner T., Kuhlen M., Perreault Levasseur L., 2016a, JCAP, 11, 048

Hezaveh Y. D., et al., 2016b, ApJ, 823, 37

Hinshaw G., et al., 2013, The Astrophysical Journal Supplement Series, 208, 19

Hofmann S., Schwarz D. J., Stöcker H., 2001, PhysRevD, 64, 083507

Holder G. P., Schechter P. L., 2003, ApJ, 589, 688

Homma D., et al., 2019, arXiv e-prints,

Hsueh J.-W., Fassnacht C. D., Vegetti S., McKean J. P., Spingola C., Auger M. W., Koopmans L. V. E., Lagattuta D. J., 2016, MNRAS, 463, L51

Hsueh J. W., et al., 2017, MNRAS, 469, 3713

Hsueh J.-W., Despali G., Vegetti S., Xu D., Fassnacht C. D., Metcalf R. B., 2018, MNRAS, 475, 2438

Hsueh J.-W., Enzi W., Vegetti S., Auger M., Fassnacht C. D., Despali G., Koopmans L. V. E., McKean J. P., 2019, arXiv e-prints, p. arXiv:1905.04182

Hui L., Ostriker J. P., Tremaine S., Witten E., 2017, PhysRevD, 95, 043541

Inoue K. T., Takahashi R., 2012, MNRAS, 426, 2978

Iršič V., et al., 2017, PhysRevD, 96, 023522

Jackson N., et al., 1998, MNRAS, 296, 483

Jiang F., van den Bosch F. C., 2017, MNRAS, 472, 657

Kamada A., Kaplinghat M., Pace A. B., Yu H.-B., 2017, Phys. Rev. Lett., 119, 111102

Katz C. A., Moore C. B., Hewitt J. N., 1997, ApJ, 475, 512

Kawano Y., Oguri M., Matsubara T., Ikeuchi S., 2004, PASJ, 56, 253

Keeton C. R., 2003, ApJ, 584, 664

Keeton C. R., 2011, GRAVLENS: Computational Methods for Gravitational Lensing (ascl:1102.003)

Keeton C. R., Moustakas L. A., 2009, ApJ, 699, 1720

Keeton C. R., Kochanek C. S., Seljak U., 1997, ApJ, 482, 604

Keeton C. R., Gaudi B. S., Petters A. O., 2003, ApJ, 598, 138

Keeton C. R., Gaudi B. S., Petters A. O., 2005, ApJ, 635, 35

Kim S. Y., Peter A. H. G., Wittman D., 2017, MNRAS, 469, 1414

Kim S. Y., Peter A. H. G., Hargis J. R., 2018, Phys. Rev. Lett., 121, 211302

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82

Kneib J.-P., Cohen J. G., Hjorth J., 2000, ApJL, 544, L35

Kochanek C. S., Dalal N., 2004, ApJ, 610, 69

Kochanek C. S., et al., 2000, ApJ, 543, 131

Koopmans L. V. E., 2005, MNRAS, 363, 1136

Koopmans L. V. E., et al., 2003, ApJ, 595, 712

Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 649, 599

Koopmans L. V. E., et al., 2009, ApJL, 703, L51

Kusenko A., 2009, Phys. Rept., 481, 1

Lagattuta D. J., et al., 2010, ApJ, 716, 1579

Lemon C. A., Auger M. W., McMahon R. G., Ostrovski F., 2018, MNRAS, 479, 5060

Lewis A., 2019, arXiv e-prints, p. arXiv:1910.13970

Li R., Frenk C. S., Cole S., Gao L., Bose S., Hellwing W. A., 2016, MNRAS, 460, 363

Lintusaari J., Gutmann M. U., Dutta R., Kaski S., Corander J., 2017, Systematic Biology, 66, e66

Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, MNRAS, 439, 300

Lovell M. R., et al., 2016, MNRAS, 461, 60

Lovell M. R., et al., 2018, MNRAS, 481, 1950

Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, MNRAS, 441, 378

Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, MNRAS, 460, 1214

Lynden-Bell D., 1967, MNRAS, 136, 101

Ma C.-P., Greene J. E., McConnell N., Janish R., Blakeslee J. P., Thomas J., Murphy J. D., 2014, ApJ, 795, 158

MacLeod C. L., Jones R., Agol E., Kochanek C. S., 2013, ApJ, 773, 35

Macciò A. V., Dutton A. A., van den Bosch F. C., 2008, MNRAS, 391, 1940

Macciò A. V., Ruchayskiy O., Boyarsky A., Muñoz-Cuartas J. C., 2013, MNRAS, 428, 882

Makarov D., Prugniel P., Terekhova N., Courtois H., Vauglin I., 2014, AA, 570, A13

Mao S., Schneider P., 1998a, MNRAS, 295, 587

Mao S., Schneider P., 1998b, MNRAS, 295, 587

Marin J.-M., Pudlo P., Robert C. P., Ryder R., 2011, arXiv e-prints, p. arXiv:1101.0955

Marlow D. R., et al., 1999, AJ, 118, 654

McKean J. P., et al., 2007, MNRAS, 378, 109

Menci N., Sanchez N. G., Castellano M., Grazian A., 2016, ApJ, 818, 90

Metcalf R. B., 2005, ApJ, 622, 72

Metcalf R. B., Madau P., 2001, ApJ, 563, 9

Metcalf R. B., Zhao H., 2002, ApJ, 567, L5

Minezaki T., Chiba M., Kashikawa N., Inoue K. T., Kataza H., 2009, ApJ, 697, 610

Miranda M., Jetzer P., 2007, APSS, 312, 203

Miranda M., Macciò A. V., 2007, MNRAS, 382, 1225

Misner C. W., Thorne K. S., Wheeler J. A., 1973, Gravitation

Möller O., Hewett P., Blain A. W., 2003, MNRAS, 345, 1

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJL, 524, L19

Moran S. M., Loh B. L., Ellis R. S., Treu T., Bundy K., MacArthur L. A., 2007, ApJ, 665, 1067

Morgan N. D., Caldwell J. A. R., Schechter P. L., Dressler A., Egami E., Rix H.-W., 2004, AJ, 127, 2617

Moustakas L. A., Metcalf R. B., 2003, MNRAS, 339, 607

Müller-Sánchez F., Prieto M. A., Hicks E. K. S., Vives-Arias H., Davies R. I., Malkan M., Tacconi L. J., Genzel R., 2011, ApJ, 739, 69

Munshi F., Brooks A. M., Applebaum E., Weisz D. R., Governato F., Quinn T. R., 2017, arXiv e-prints, p. arXiv:1705.06286

Nadler E. O., Mao Y.-Y., Green G. M., Wechsler R. H., 2019, ApJ, 873, 34

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Newton O., Cautun M., Jenkins A., Frenk C. S., Helly J. C., 2018, MNRAS, 479, 2853

Nierenberg A. M., Treu T., Wright S. A., Fassnacht C. D., Auger M. W., 2014, MNRAS, 442, 2434

Nierenberg A. M., Treu T., Menci N., Lu Y., Torrey P., Vogelsberger M., 2016, MNRAS, 462, 4473

Nierenberg A. M., et al., 2017, MNRAS, 471, 2224

Nierenberg A. M., et al., 2020, MNRAS, 492, 5314

Oguri M., Marshall P. J., 2010, MNRAS, 405, 2579

Patnaik A. R., Browne I. W. A., Walsh D., Chaffee F. H., Foltz C. B., 1992, MNRAS, 259, 1P

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

Phillips P. M., et al., 2000, MNRAS, 319, L7

Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

Pullen A. R., Benson A. J., Moustakas L. A., 2014, ApJ, 792, 24

Read J. I., Walker M. G., Steger P., 2018, preprint, (arXiv:1808.06634)

Richings J., Frenk C., Jenkins A., Robertson A., 2018, arXiv e-prints,

Ritondale E., Vegetti S., Despali G., Auger M. W., Koopmans L. V. E., McKean J. P., 2018, arXiv e-prints, p. arXiv:1811.03627

Robles V. H., Bullock J. S., Boylan-Kolchin M., 2019, MNRAS, 483, 289

Rocha M., Peter A. H. G., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Oñorbe J., Moustakas L. A., 2013, MNRAS, 430, 81

Ros E., Guirado J. C., Marcaide J. M., Pérez-Torres M. A., Falco E. E., Muñoz J. A., Alberdi A., Lara L., 2000, AA, 362, 845

Rubin D. B., 1984, The Annals of Statistics, 12, 1151

Rusu C. E., et al., 2019, arXiv e-prints, p. arXiv:1905.09338

Sawala T., et al., 2016a, MNRAS, 456, 85

Sawala T., et al., 2016b, MNRAS, 457, 1931

Schechter P. L., Wambsganss J., 2002, ApJ, 580, 685

Schechter P. L., et al., 2003, ApJ, 584, 657

Schneider P., 1997, MNRAS, 292, 673

Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses, doi:10.1007/978-3-662-03758-4.

Schneider A., Smith R. E., Macciò A. V., Moore B., 2012, MNRAS, 424, 684

Schneider A., Smith R. E., Reed D., 2013, MNRAS, 433, 1573

Schneider A., Trujillo-Gomez S., Papastergis E., Reed D. S., Lake G., 2017, MNRAS, 470, 1542

Scott D. W., 1992, Multivariate Density Estimation

Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Shajib A. J., et al., 2019, MNRAS, 483, 5649

Shankar F., et al., 2017, ApJ, 840, 34

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Shi X., Fuller G. M., 1999, Phys. Rev. Lett., 82, 2832

Shoemaker I. M., Kusenko A., 2009, PhysRevD, 80, 075021

Sluse D., et al., 2003, AA, 406, L43

Sluse D., et al., 2006, AA, 449, 539

Sluse D., Chantry V., Magain P., Courbin F., Meylan G., 2012, AA, 538, A99

Sonnenfeld A., Treu T., Gavazzi R., Suyu S. H., Marshall P. J., Auger M. W., Nipoti C., 2013, ApJ, 777, 98

Spergel D. N., Steinhardt P. J., 2000, Physical Review Letters, 84, 3760

Springel V., et al., 2008, MNRAS, 391, 1685

Starkenburg E., et al., 2013, MNRAS, 429, 725

Strigari L. E., Koushiappas S. M., Bullock J. S., Kaplinghat M., 2007, PhysRevD, 75, 083526

Sugai H., Kawai A., Shimono A., Hattori T., Kosugi G., Kashikawa N., Inoue K. T., Chiba M., 2007, ApJ, 660, 1016

Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, ApJ, 711, 201

Suyu S. H., et al., 2013, ApJ, 766, 70

Sykes C. M., et al., 1998, MNRAS, 301, 310

Tegmark M., et al., 2004, PhysRevD, 69, 103501

Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878

Tollet E., et al., 2016, MNRAS, 456, 3542

Tormen G., Diaferio A., Syer D., 1998, MNRAS, 299, 728

Treu T., Koopmans L. V. E., 2002, ApJ, 575, 87

Treu T., Koopmans L. V. E., 2004, ApJ, 611, 739

Treu T., Koopmans L. V., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 640, 662

Treu T., Gavazzi R., Gorecki A., Marshall P. J., Koopmans L. V. E., Bolton A. S., Moustakas L. A., Burles S., 2009, ApJ, 690, 670

Treu T., et al., 2018, MNRAS, 481, 1041

Tulin S., Yu H.-B., 2018, Phys. Rept., 730, 1

Vegetti S., Koopmans L. V. E., 2009, MNRAS, 400, 1583

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

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

Vegetti S., Despali G., Lovell M. R., Enzi W., 2018, preprint, (arXiv:1801.01505)

Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, PhysRevD, 71, 063534

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, PhysRevD, 88, 043502

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Vogelsberger M., Zavala J., Cyr-Racine F.-Y., Pfrommer C., Bringmann T., Sigurdson K., 2016, MNRAS, 460, 1399

Vuissoz C., et al., 2008, AA, 488, 481

Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, ApJ, 568, 52

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJL, 827, L23

Weyant A., Schafer C., Wood-Vasey W. M., 2013, ApJ, 764, 116

Wheeler C., et al., 2018, arXiv e-prints,

Wisotzki L., Schechter P. L., Bradt H. V., Heinmüller J., Reimers D., 2002, AA, 395, 17

Wong K. C., et al., 2017, MNRAS, 465, 4895

Xu D. D., et al., 2009, MNRAS, 398, 1235

Xu D. D., Mao S., Cooper A. P., Wang J., Gao L., Frenk C. S., Springel V., 2010, MNRAS, 408, 1721

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

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

Zhao D. H., Mo H. J., Jing Y. P., Börner G., 2003, MNRAS, 339, 12

de Blok W. J. G., Walter F., Brinks E., Trachternach C., Oh S. H., Kennicutt R. C. J., 2008, AJ, 136, 2648

van den Bosch F. C., Jiang F., Hearin A., Campbell D., Watson D., Padmanabhan N., 2014, MNRAS, 445, 1713

van den Bosch F. C., Ogiya G., Hahn O., Burkert A., 2018, MNRAS, 474, 3043