Title

Constraining Dark Matter, Four Images at a Time

Author

Cohen, Jacob Sidney

Publication Date

2024

Peer reviewed|Thesis/dissertation

Constraining Dark Matter, Four Images at a Time

By

Jacob Sidney Cohen

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

Christopher Fassnacht, Chair

Tucker Jones

David Wittman

Committee in Charge

2024

Contents

List of Figures

List of Tables

Abstract of the Dissertation

Constraining Dark Matter, Four Images at a Time

Dark matter makes up roughly 85 percent of the mass budget of the Universe. Though we cannot see it, hence the name, the formation and behavior of structures from the largest cosmological scales down to sub-galactic scales betray the existence of dark matter via its gravitational interactions. The widely accepted view of cosmological history, informed by observations of hierarchical structure formation, posits non-relativistic dark matter particles that only interact gravitationally, known as cold dark matter (CDM). However, alternative theories of dark matter may describe small-scale observations as well or better. It is essential to test variations on CDM and differentiate between models that agree on larger scales. We can observe even completely invisible dark matter structures from the way they deflect and magnify light. This phenomenon, gravitational lensing, has been used to infer some of the tightest constraints on alternative dark matter models and grows stronger as telescopes improve and sample sizes increase.

This dissertation is focused on strong gravitational lensing, when multiple images of the light source appear on the sky. Particularly, I demonstrate and evaluate a method called flux-ratio analysis, which uses bright, compact sources like quasars that are lensed into four distinct images. In these systems, the main lens is an elliptical galaxy, the mass distribution of which determines the positions and relative brightnesses of the images. Low-mass dark matter structures, called halos, associated with the main lens or along the line of sight are detectable in the additional magnification or demagnification they introduce to individual images. Discrepancies between observed flux ratios and those predicted by a smooth model of the main lens indicate the presence of perturbing halos. With a sample of many lensed quasars, we can characterize the population statistics of low-mass halos. These statistics are determined by the underlying dark matter model, thus they enable us to infer constraints on it.

In Chapter 1 of this work, I provide general background information on dark matter and cosmology and detail the predictions of CDM. I then discuss the use of gravitational lensing to detect dark matter and review the history of flux-ratio analysis to put my work in the context of the field.

Chapter 2 of this work contains a detailed description of my flux-ratio analysis procedure as well as constraints on warm dark matter (WDM) inferred from a sample of 14 lensed quasars. I provide background on WDM theory and describe the process of simulating low-mass halo populations that align with theoretical predictions. Then I walk through my Bayesian inference procedure and its application to infer the strongest gravitational lensing constraints on WDM to date. Lastly, I explore the context of my analysis in relation to other works that used subsets of this lens sample. I highlight the differences in between our analysis procedures and potential sources of systematic error.

In Chapter 3, I investigate the addition of general third- and fourth-order multipoles to the mass model. I find that multipoles with realistic amplitudes, at least when compared to the isophotes of elliptical galaxies, can perturb flux ratios as significantly as low-mass halos. If their orientation angles are left to vary freely, joint third- and fourth-order multipoles are completely degenerate with CDM halo populations when modeling quadruply-imaged quasars. This work calls into question all previous dark matter constraints from flux-ratio analysis, including those presented in Chapter 2. However, those results are still critical to the field moving forward. Until this degeneracy is mitigated, the method and degree of inclusion of multipoles into an inference procedure will strongly impact the result. Chapter 2 explores one extreme end of this spectrum – the exclusion of multipoles.

In Chapter 4, I summarize and provide a framework for reevaluating past dark matter constraints from flux-ratio analysis. I explain how comparison between observations of lens galaxy isophotes and mass models from associated lensed arcs can provide more informative priors on multipoles in future lens models. Additionally, I forecast ways that high-resolution simulations can provide complementary prior information both regarding multipoles that may be present in galactic dark matter halos and their connection to the distribution of baryons in the galaxy. Finally, I give an overview of the increase in the sample size of strong lens systems that is expected from JWST and Euclid. Though data from these telescopes holds great potential to help us understand the nature of dark matter, we must first untangle the degeneracy between smooth model complexity and low-mass halos.

Acknowledgments

I'd like to acknowledge the support and influence of the people that helped me get to the point of thanking them for helping me complete a PhD in roughly chronological order. Naturally, Dave and Alexa Cohen come first. As parents, you have equipped me with belief in myself, a positive outlook and the freedom to make the choices that define me. You both have always been and continue to be role models in your approach to personal and professional challenges and your relationships with those around you. Maddie, my sister, my friend. When I started at UC Davis you were just starting your second year of college. Now, I'm 10 years deep into college and you're well on your way to an amazing career making a difference in the lives of children and their families. You're an inspiration, and I hope to be just like my little sister one day.

Alex, Skye, Jake and Glen, thank you for occasionally letting me forget that I'm in graduate school and instead pretend that I am in high school. To my graduate school cohort, thank you for building and maintaining a community that made Davis quickly feel like home and weathered a pandemic. Arsalan, Ethan, Julie, Karthik, Patrick, Pratik, Tyler, Sid and Yash, I will always think of the day we topped Tallac and packed into 2 borrowed tents as the foundational event of my graduate school career. I'd also like to thank the folks in the Davis and Sacramento running communities that shared countless hours with me chasing rabbits down by the creek. David and Brandon, thank you for pacing me. I mean that both literally and metaphorically.

Thank you to all my students over the years. From the discussion lab group that cared for me when I walked in bloodied from a bike accident, to the students that joined me on the picket line, to those that gave me thoughtful feedback on how to improve as a lecturer, it was a pleasure to work with you. Armela Keqi, thank you for your excellent guidance and partnership as an associate instructor, and thank you for your kindness, humor and friendship.

Chris, though I started working with you somewhat serendipitously, I don't think I could have possibly imagined for myself an advisor that is more patient, generous and dedicated to the success of their students. When you asked if I wanted to kick off my research by spending 6 weeks in Germany to study lensing with Simona's group, I felt like it was a ticket to Hollywood. Only a small fraction of the sparkle of work travel has worn off, and I can't thank you enough for the opportunities you have opened for me. You have always been supportive of my development as

a teacher, and it was a a pleasure to spend a couple half-summers in COSMOS with you. As I continue my career outside of academia, I am lucky to bring with me the wisdom you've imparted regarding tackling massive projects, interacting with colleagues and enjoying my work.

I am grateful to have worked with everyone in the SHARP collaboration. From biweekly telecons to pasta making in the north of Italy, it has been as fun as it has been enlightening. Simona Vegetti, thank you and your entire research group for hosting and including me back in 2020. Conor O'Riordan, thank you for challenging me to and helping me carry out higher-quality research than I had imagined I was capable.

Emily, I'll keep this brief as I don't want to use up good material for my wedding vows, but you have had the most profound impact on me over the past 3.5 years. Your constant love and encouragement made me find a new dimension of myself, and it is hard to imagine completing this PhD without having had you enter my life. I look up to you for your creativity and work ethic, and I'm eternally grateful for however much of it rubs off on me.

I also need to thank Zeek and Quinn for accompanying me through long nights of remote Keck observing and early mornings before lecturing. I thought long and hard about whether to inform the reader that they are my cats, but I have ultimately decided against it.

David Wittman and Tucker Jones, from encouraging me to dig deeper and take ownership of my research after my first attempt at a qualifying exam, to engaging in discussions about running and Dead & Company, to reading this very dissertation, you have both been supportive and kind to me throughout my tenure on the 5th floor. I thank you and the rest of the Astronomy faculty for fostering a positive environment in which I feel comfortable challenging myself and look forward to any and all gatherings.

Mom, Dad, I'm finally breaking the generational all-but-dissertation curse.

Chapter 1

An Introduction to Dark Matter and Gravitational Lensing

In this chapter, I provide a foundation for the research presented in Chapters 2 and 3. After briefly reviewing the history of dark matter and its cosmological context, I give an overview of the paradigmatic cold dark matter, its predictions and proposed alternatives. I then discuss gravitational lensing and explain how it can be applied to differentiate between and constrain competing dark matter models. This chapter ends with an introduction of other complementary astrophysical probes that have contributed to dark matter constraints and an outline of the rest of this dissertation.

1.1 The Makeup of the Universe

What makes up the Universe? The answer to this fundamental question has grown from the rough categorization of elements on Earth to our modern understanding of matter, energy and their evolution since the Big Bang. The Standard Model of particle physics describes the components and interactions of matter down to subatomic scales, and general relativity tells us how space and time behave. When we attempt to explain our observations of the cosmos with these theories alone, however, it becomes clear that this catalogue of the contents of the Universe is incomplete. Dark energy, which is the name attributed to the mysterious driver of accelerating cosmic expansion, turns out to make up ~70 percent of the Universal mass-energy budget today. Of the rest that we ascribe to matter, visible matter, the kind described by the Standard Model, can only account for ~15 percent (Planck Collaboration et al., 2020).

1.1.1 A History of Missing Matter

As far back as the 19th century, astronomers proposed that planets and stars too dim to be observed could be contributing to the dynamics of the solar system and Milky Way (see Bertone & Hooper, 2018). Early 20th century observational studies put upper limits on the local density of this so-called "dark" matter to be equal to that of visible matter (e.g. Jeans, 1922; Oort, 1932). While these pioneering studies were surely important, the contemporary field of dark matter research was jump-started by the work of Fritz Zwicky in the mid 1930s. By applying the virial theorem, which relates the total kinetic and potential energies of a system, to observations of galaxies and their apparent velocities in the Coma Cluster, Zwicky (1937) inferred that the total mass of the cluster must be many times that expected from visible matter alone. Though his estimates were later shown to be flawed, his conclusion that dark matter is not only present in the Universe, but much more abundant than visible matter has been repeatedly confirmed.

Studies in the 1970s, including the seminal work of Vera Rubin, presented optical and radio observations of galaxies and found that their rotation curves remained flat in the outer regions (e.g Rubin & Ford, 1970; Rogstad & Shostak, 1972; Roberts & Rots, 1973; Rubin et al., 1978). As the mass inferred from their light distributions would result in rotation curves that steeply drop off, this ubiquitous phenomenon further indicated the presence a substantial amount of dark matter and showed its dominance over visible matter in the outer regions of galaxies. Since then, the presence of dark matter has been verified in myriad ways, for example, via gravitational lensing (Refregier, 2003; Vegetti et al., 2010), which will be discussed in detail in this dissertation, and the formation of large-scale cosmological structure (Einasto et al., 1980; Davis et al., 1982). According to recent estimates by Planck Collaboration et al. (2020), dark matter accounts for ~85 percent of the mass budget of the universe, the rest attributed to visible, or baryonic matter. These findings indicate that dark matter does not have baryonic origins.

1.2 Cold Dark Matter

Though the precise nature of dark matter is still an open question, the answer to which is the concern of this dissertation, the phenomenological cold dark matter (CDM) theory is widely accepted as part of the "standard model" of cosmology. The CDM paradigm posits dark matter particles to be massive and non-relativistic as well as collisionless or very weakly interacting (Peebles, 1982;

Blumenthal et al., 1984). The most common CDM particle candidate is the weakly-interacting massive particle (WIMP), a class of thermal relic with mass ∼100 GeV (Jungman et al., 1996). This mass scale ensures that the time at which WIMPs fall out of thermal equilibrium in the early Universe results in a dark matter density consistent with that which is observed today. Other CDM candidates include axions and sterile neutrinos (see Feng, 2010), though the observational and experimental distinction between the various CDM candidates is outside the scope of this work. Unless otherwise specified, I will assume the WIMP formulation of CDM in this work.

Structural growth in the Universe is seeded by primordial density fluctuations that collect dark matter gravitationally. The free-streaming length, which describes the mean distance that dark matter particles travel before falling into a potential well, can be calculated as

$$\lambda_{\rm fs} = \int_0^{t_{\rm EQ}} \frac{v(t)dt}{a(t)} \quad , \tag{1.1}$$

where v is the mean particle speed, a is the scale factor and $t_{\rm EQ}$ is the time of matter-radiation equality when perturbations can start to undergo gravitational collapse (Kolb & Turner, 1990). Density perturbations below the free-streaming length scale are suppressed. CDM particles, however, are heavy and slow, which allows for perturbations down to present-day planet scales to survive. The structure formation that arises is hierarchical. Collapsed dark matter clumps, called halos, grow and merge up from planet scales ($M \gtrsim 10^{-6} M_\odot$) to the filaments and nodes of the cosmic web, the largest structural elements that we observe today. A combination of theoretical work and simulations gives us predictions for the population of halos on all scales and their internal structure.

1.2.1 Halo Populations in CDM

Early theoretical work by Press & Schechter (1974) showed that early Universe perturbations modeled as a Gaussian random field could be mapped to the resulting halo mass function (HMF), the differential number density as a function of mass and redshift, $dn(m, z)/dM$, using linear theory and a spherical collapse model. They found that the HMF follows a power law from the free-streaming scale up to a redshift-dependent cutoff which is $\sim 10^{15} M_\odot$, the scale of galaxy clusters, today. Above this cutoff, which reflects the largest structures that have had time to virialize, the mass function falls off exponentially. Sheth et al. (2001) expanded upon this work to include a

more general elliptical collapse model, finding better agreement with numerical simulations (Lacey & Cole, 1994; Sheth & Tormen, 1999) and inferring the log power law slope of the global HMF below the cutoff scale to be $\alpha \approx -1.9$.

Increased computational capacity brought about the ability to probe smaller scales with cosmological N-body simulations (e.g. Ghigna et al., 1998; Klypin et al., 1999). Though it was known that larger halos form via accumulation of smaller ones, simulations revealed that many of the small halos that are accumulated, known as subhalos, survive as identifiable structures. The associated subhalo mass function (SHMF) has a very similar shape to the global HMF with a log power law slope of $\alpha_s \approx -1.8$ below the cutoff scale and a normalization that scales with the host halo mass. The largest subhalos are generally about an order of magnitude less massive than their hosts.

1.2.2 Halo Density Profiles in CDM

Dark matter-only simulations predict that halos on all scales have roughly the same density structure (Dubinski & Carlberg, 1991; Navarro et al., 2010; Klypin et al., 2016; Ishiyama et al., 2021). Though profiles with more parameters can of course describe a broader population more accurately, the two-parameter NFW profile, proposed by and named after Navarro, Frenk, & White (1997), is the most common choice. It is a spherical density profile defined as a function of radial distance from the center, r, as

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s}\left(1 + \frac{r}{R_s}\right)^2} \quad , \tag{1.2}$$

where ρ_0 is the normalization and R_s is the scale radius. Though this expression is more directly interpretable, NFW profiles are most often parameterized instead in terms of total mass and concentration, c, which describes compactness. Though the integrated mass of the NFW profile is divergent and the total mass of a halo depends on a somewhat arbitrary choice of boundary, two common choices to which I refer in this dissertation are the virial radius, $R_{\rm vir} = cR_s$, and its enclosed mass, $M_{\rm vir}$, and the radius of maximum circular velocity, $R_{\rm max} \approx 2.16R_s$, and its enclosed mass, $M_{\rm max}$.

The mass and concentration of a halo are dependent on the history of its specific surroundings (Wechsler et al., 2002), but they are most fundamentally dependent on the time at which the halo formed and its age. The universe was denser earlier on, so the concentrations of halos that formed

earlier are, on average, higher. Since smaller halos form first in hierarchical structure formation, they are generally more concentrated. Additionally, at a fixed halo mass, halos that formed earlier are more concentrated. Though larger halos are less concentrated, they have higher densities at all physical distances from their centers than smaller halos. While the relationships between mass, concentration and age are generally predictable, there is significant scatter in these relations (Duffy et al., 2008; Moliné et al., 2023), and they are only true on average.

1.2.3 Baryon Content

Since more massive halos create larger gravitational potential wells, we expect the largest halos to trap the most baryonic matter and thus to have the largest stellar masses. This is roughly true, but the relationship between halo mass and stellar mass is far from linear. Multiple effects, such as supernova feedback and heating due to photoionization (e.g. Gnedin & Zhao, 2002), increasingly inhibit star formation in lower mass subhalos. At masses $M_{\rm halo} \lesssim 10^8 M_\odot$, halos can barely, if at all, support galaxy formation. Around $M_{\rm halo} \approx 10^{12} M_\odot$, the star forming efficiency peaks, and it continues to drop for increasingly higher mass halos which may experience significant feedback from active galactic nuclei (AGN). At Milky Way scales and larger, a halo that is an order of magnitude more massive than another will host a galaxy that is far less than an order of magnitude more massive than the other's galaxy. At galaxy cluster scales ($\gtrsim 10^{15} M_\odot$), gas heating processes exceed cooling in the host halo, and the stellar content is hosted in the galaxies that make up the cluster. At almost all mass scales, the total mass of galaxies is dominated by dark matter over baryons by more than an order of magnitude.

1.2.4 Small-Scale Challeges to CDM

While CDM is able to describe large-scale structure well, at scales smaller than $\sim 10^{11} M_\odot$, tensions exist between theoretical predictions and observations (see Bullock & Boylan-Kolchin, 2017), although these problems may mostly be solved with the addition of baryons into simulations. The mass range in which baryons are expected to most strongly affect the halo density profile roughly aligns with the mass ranges of the missing satellites, core-cusp and too-big-to-fail problems discussed below. However, their investigation has sparked inquiry about other types of dark matter particles with properties that may also solve these tensions. As the constraint of alternate dark matter models is this focus of this work, I briefly present the most significant of these challenges.

1.2.4.1 Missing Satellites

Since the HMF for CDM should be scale-invariant at small scales down to $\sim 10^{-6} M_{\odot}$, the Milky Way should host abundant substructure, including $\sim 10^3$ subhalos large enough to have supported molecular cooling and thus to host satellite galaxies. However, we observe more than an order of magnitude fewer satellite galaxies in the Milky Way. This may be solved by assuming that smaller halos are less efficient at forming stars. By comparing the observed population of galaxies to the predicted population of halos and assuming they have a one-to-one correlation (assuming the largest halos host the largest galaxies), we can establish a relationship between halo mass and expected stellar mass. This technique, extrapolated down to the scale of the smallest Milky Way satellites ($M_{vir} \approx 10^8 M_{\odot}$, $M_* \approx 10^2 M_{\odot}$), reveals that there may be no tension between predictions and observations (Garrison-Kimmel et al., 2017). The predicted halos are simply not all large enough to be luminous. More recent work that takes into account discovery of smaller dwarf galaxies, nuanced completeness corrections for larger satellites and tidal stripping as halos pass through the Milky Way's stellar disk more concretely resolve the missing satellites problem (Kim et al., 2019).

1.2.4.2 Core-Cusp Problem

Another feature of CDM-only models is the "cuspy" halo density profile, which is sharply peaked near the center and is well-described by the NFW halo profile. Close to the center, the NFW profile becomes $\rho(r \ll R_s) \propto r^{-1}$, and far from the center it drops off to $\rho(r \gg R_s) \propto r^{-3}$. However, observed rotation curves for some smaller galaxies around the size of large Milky Way satellites and smaller show that their host halos should have much flatter density profiles, or near-constant density "cores" (inner power law slope $\rho(r \ll R_s) \propto r^{-\gamma}$ with $0 < \gamma < 0.5$). Baryonic feedback is believe to be capable of creating these cores (Read et al., 2016), but the effect is sensitive to the mass of the halo and its resulting stellar mass (ineffective below $M_{vir} \approx 10^{10} M_{\odot}$), and effectiveness varies between simulations. Interaction between halos and baryons in the host galaxy or cosmic web may also give rise to lower central masses, but more observations of field dwarfs are needed to constrain this effect.

1.2.4.3 Too-Big-To-Fail

Independent of the inner slopes of halo density profiles, CDM-only simulations also predict an excess of mass in the centers of halos compared to observational inferences - a problem of normalization (see Boylan-Kolchin et al., 2011). "Central mass" and "central density" are different, yet related halo quantities. While it is possible to address these individually, effects that reduce the central densities of halos also tend to reduce their central masses. As mentioned, abundance matching between the population of expected subhalos and observed galaxies in the Milky Way and the Local Group seems to solve the issue of missing satellites. However, this method assumes that the largest subhalos host the largest satellite galaxies. Comparing the largest subhalos in simulations of Milky Way sized galaxies to the largest observed Milky Way satellites, it was found that the largest observed satellites have lower central masses than the largest subhalos. If the largest subhalos did exist with their expected mass profiles from CDM-only simulations, they would be "too-big-to-fail" at forming stars. Lower central masses as a result of baryon interactions may also solve this problem (Wetzel et al., 2016).

1.3 Alternatives to CDM

While the missing satellite, core-cusp and too-big-to-fail problems can potentially be solved within the CDM paradigm, they have also motivated investigation of other models of dark matter that may better match astrophysical observations. The solutions relying on baryonic feedback are promising, yet somewhat contentious due to the variance in results between simulations. Additionally, alternative dark matter models can vary only slightly from CDM such that the differences between the two are below our current detection thresholds. Establishing constraints on alternative models is important for testing their viability and is complementary to searches for signatures of dark matter particle interactions that are dependent on the type of particle. There are two features of halo populations that are most commonly targeted by observational dark matter studies. One is the small-scale HMF. Though CDM posits the scale-free nature to potentially continue down to planetary scales, alternate dark matter models lead to various amounts of truncation below satellite galaxy scales. The other is the range of halo density profiles and central masses, with some dark matter models allowing for the creation of cored halos and fewer star-forming satellites. I present two of the leading alternatives to CDM here. It is important to note that there is diversity in

observed inner density profiles from small subhalos to large host halo scales, and baryonic effects will occur alongside the effects of any non-CDM model. Also, these alternatives represent classes of models, and there are many proposed particle candidates for each whose properties can produce the effects described. Other solutions exist for the problems with CDM that include abandoning the idea of dark matter altogether and modifying Newtonian dynamics (see McGaugh, 2015). The focus of the work presented in this dissertation is the constraint of warm dark matter (WDM), though the methodologies are applicable to any model that predicts a clumpy distribution.

1.3.1 Warm Dark Matter

Warm dark matter (WDM) is a class of dark matter models that consist of particles, also often theorized as thermal relics, that have some effective free streaming length much larger than that of CDM particles (see Bode et al., 2001; Viel et al., 2005). This sets a minimum scale for perturbations in the early Universe that grow to become the halos we observe today, which results in suppression of the HMF at small scales. While the conversion between the scale of mass function suppression and particle mass depends on the specific WDM model, constraints are easily translated between them. In general, the lighter a WDM particle, the higher the mass scale at which suppression occurs. This suppression mass scale is often quantified by the half-mode mass, $M_{\rm hm}$. For $m \ll M_{\rm hm}$, suppression is strong, and for $m \gg M_{\rm hm}$ it is negligible. This effect can also be tuned such that the scale of suppression produces the number of observed Milky Way satellites and solves the missing satellite problem.

In addition to affecting the HMF, the suppression of small scale initial density perturbations in WDM models affects the density profiles of individual halos (Schneider et al., 2013). At mass scales much larger than $M_{\rm hm}$, the formation time and density profile of halos in CDM and WDM simulations look very similar since the density perturbation scale that produces them is not suppressed. However, at roughly an order of magnitude above $M_{\rm hm}$ and below, halos of a given mass take longer to form in WDM than in CDM. Since the central density of a halo reflects the mean density of the Universe at its redshift of formation (which increases with redshift), halos with later formation redshifts have lower central densities. The reduction in central density, and thus lower concentration, for WDM halos, however, is not limited to halos with later formation times compared to CDM. Since larger halos are formed by the merging of smaller halos, the

reduction in concentration extends to 2-3 orders of magnitude above $M_{\rm hm}$. This phenomenon, scaled appropriately, can solve both the core-cusp and too-big-to-fail problems by only reducing the concentrations of halos below or equal to the size of the smallest Milky Way satellites. This also allows for the formation of enough of the smallest inferred subhalo masses from observations.

1.3.2 Self-Interacting Dark Matter

Self-interacting dark matter (SIDM) (Spergel & Steinhardt, 2000), is another highly studied alternative to CDM that simulations show can reproduce many observations. As the name suggests, it consists of dark matter particles that can interact with each other via collisions. This creates a pressure that limits the slope of the inner density profiles of halos. Though by a different mechanism, this can be tuned to solve the core-cusp and too-big-to-fail problems by reducing the central masses of subhalos. However, at galaxy cluster scales, we do observe more cuspy density profiles that put an upper limit on the SIDM cross section inconsistent with that required to solve small scale issues (Kaplinghat et al., 2016). This can be solved with a velocity-dependent cross section that satisfies both.

1.4 Gravitational Lensing Theory

Gravitational lensing is a phenomenon caused by the deflection of light by the gravitational fields of massive objects, and the work in this dissertation is devoted to its use to constrain the properties of dark matter. While general relativity is needed to calculate the observed deflection angle of a photon passing near a massive object, Newtonian dynamics is sufficient to describe the deflection for a non-relativistic object. The result obtained when considering light differs by only a factor of 2. This factor can be conceptually explained, thus the Newtonian derivation is still instructive. The following derivation is based on one performed in Chapter 2 of Dodelson (2017) with additional information from Perlick (2012) and Wambsganss (1998).

1.4.1 Newtonian Deflection of Light by a Point Mass

Consider a point mass m traveling along the z-axis past a much more massive point mass M with impact parameter b. Assume M is sufficiently larger than m such that M is motionless. From

Newton's law of universal gravitation, the equation of motion for m is,

$$\ddot{\mathbf{R}} = -\frac{MG}{R^2}\hat{\mathbf{R}} \tag{1.3}$$

where m cancels because it appears on both sides of the equation. As a result, the Newtonian approach yields the same result for a massless photon and a massive object traveling at the same speed. Focusing on the left hand side of this equation, we first need to decompose the second time derivative of $\mathbf{R}$ into its $\hat{\mathbf{R}}$ and $\hat{\boldsymbol{\phi}}$ components. Using,

$$\dot{\hat{\boldsymbol{R}}} = -\dot{\phi}\hat{\boldsymbol{\phi}} \quad \text{and} \quad \dot{\hat{\boldsymbol{\phi}}} = \dot{\phi}\hat{\mathbf{R}} \tag{1.4}$$

we find that,

$$\ddot{\mathbf{R}} = \left(\ddot{R} - R\dot{\phi}^2\right)\hat{\mathbf{R}} + \left(\frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}t}\left(R^2\dot{\phi}\right)\right)\hat{\boldsymbol{\phi}} \quad . \tag{1.5}$$

Since there is no $\hat{\boldsymbol{\phi}}$ component of the gravitational force, we know that this component of $\ddot{\mathbf{R}}$ must vanish. Thus, we define a conserved quantity J_z,

$$J_z \equiv R^2\dot{\phi} \quad , \quad \dot{J}_z = 0 \quad \text{so} \quad -\frac{MG}{R^2} = \ddot{R} - \frac{J_z^2}{R^3} \tag{1.6}$$

which will be helpful in simplifying the rest of this solution. Another way to simplify this differential equation will be to transform the time derivatives into derivatives with respect to ϕ, which will be denoted as $dx/d\phi = x'$. This way,

$$\dot{R} = \frac{\mathrm{d}R}{\mathrm{d}\phi}\frac{\mathrm{d}\phi}{\mathrm{d}t} = R'\dot{\phi} = \frac{R'J_z}{R^2} \quad \text{and, similarly} \quad \ddot{R} = \frac{J_z^2}{R^2}\left(\frac{R''}{R^2} - \frac{2R'^2}{R^3}\right) \quad . \tag{1.7}$$

Plugging into equation 1.6 and simplifying, we get,

$$-MG = J_z^2\left(\frac{R''}{R^2} - \frac{2R'^2}{R^3} - \frac{1}{R}\right) \tag{1.8}$$

which, substituting $u = 1/R$, becomes,

$$u'' + u = \frac{MG}{J_z^2} \tag{1.9}$$

This simple second-order ordinary differential equation has the solution,

$$u = \frac{1}{R(\phi)} = A\cos(\phi - \phi_0) + \frac{MG}{J_z^2} \quad , \tag{1.10}$$

where ϕ_0 is the angle at which mass m gets closest to M (where u is maximized). This equation has two unknowns, J_z and A, that we must solve for based on the initial conditions of the problem. At very early times, when R approaches infinity and ϕ approaches 0, $\phi = \sin(b/R) \approx b/R$ using the small-angle approximation for sine. Thus, $\dot{\phi} \approx -b\dot{R}/R^2$. At this early time, projectile m, in this case a photon, has yet to be deflected significantly, so $\dot{R} \approx -c$. Since we showed that J_z is a conserved quantity,

$$J_z = R^2\dot{\phi} \approx bc \quad \text{so} \quad \dot{\phi} \approx \frac{bc}{R^2} \tag{1.11}$$

at all times. Examining the early-time state of this problem also holds the key to finding the value of A. Substituting $\phi = 0$, $R \to \infty$ and our new value for J_z into equation 1.10 yields,

$$0 \approx A\cos(\phi_0) + \frac{MG}{c^2b^2} \quad . \tag{1.12}$$

Taking the time derivative of equation 1.10, substituting our approximate values for $\dot{R}$ and $\dot{\phi}$, and using the early-time (small ϕ) approximation $\sin(\phi - \phi_0) \approx -\sin(\phi_0)$, we get,

$$\frac{c}{R^2} \approx A\sin(\phi_0)\left(\frac{bc}{R^2}\right) \quad \text{so} \quad A \approx \frac{1}{b\sin(\phi)} \tag{1.13}$$

Plugging equation 1.13 into 1.12 and setting $\sin(\phi_0) = 1$ since $\phi_0 \approx \pi/2$, we get,

$$\cos(\phi_0) \approx -\frac{MG}{c^2b} \quad . \tag{1.14}$$

Again, since $\phi_0 \approx \pi/2$, we can use the small-angle approximation to show that

$$\phi_0 \approx \frac{\pi}{2} + \frac{MG}{c^2b} \quad . \tag{1.15}$$

There are multiple ways to solve for the final deflection angle, but the most intuitive is to recognize that since ϕ_0 is the angle at which the photon is closest to the deflecting point mass, its trajectory

at this point is perpendicular to $\hat{\mathbf{R}}$. Simple geometric reasoning reveals that at ϕ_0, the deflection angle from the photon's original path is equal to the difference between ϕ_0 and $\pi/2$, the angle of minimum distance for the undeflected trajectory. Finally, since angular momentum is conserved, the deflection must be symmetric about ϕ_0. The photon undergoes an equivalent deflection to the one experienced up to ϕ_0 as it continues beyond the point mass to infinity. We thus arrive at the Newtonian approximation for the deflection angle of light deflected by a point mass,

$$\delta\theta_{Newton} = \frac{2MG}{c^2 b} \quad . \tag{1.16}$$

As mentioned previously, this result is a factor of 2 smaller than the correct answer for the deflection angle predicted by general relativity. The reason for the difference is that Newtonian dynamics fails to describe the bending of space from the deflecting mass. To describe the path of the photon in general relativity, we would use the Schwarzchild metric,

$$\mathrm{d}s^2 = -\left(1 - \frac{2MG}{c^2 R}\right) c^2 \mathrm{d}t^2 + \left(1 - \frac{2MG}{c^2 R}\right)^{-1} \mathrm{d}R^2 + R^2 \mathrm{d}\Omega^2 \tag{1.17}$$

which demonstrates the effect mass M has on both the time and space components. The Newtonian approach correctly describes the motion of a non-relativistic point mass being deflected because with a speed much smaller than c, it travels much further in time than in space. The spatial contribution to the spacetime interval is minimal. When the point mass is replaced with a photon, both the space and time curvature effects are equal in magnitude which results in twice the deflection angle of the time curvature-only approach.

1.4.2 Quasi-Newtonian Approximation

To simplify many gravitational lensing calculations, the result for the deflection angle of a photon by a point mass from general relativity,

$$\delta\theta_{GR} = \frac{4MG}{c^2 b} \tag{1.18}$$

can be used as a starting point. While linear superposition of deflections for a mass distribution does not hold in general relativity, it serves as a good approximation under a few assumptions that

are valid for many astrophysical applications. We first define the dimensionless quantities of the lens position ($\boldsymbol{\theta}$), source position ($\boldsymbol{\beta}$) and reduced deflection angle ($\boldsymbol{\alpha}$) as their transverse distances from the line of sight scaled by the radial distances along the line of sight of the planes in which they lie. (Note that up to this point we used $\delta\theta$ to refer to the deflection angle.) We can relate them with,

$$\boldsymbol{\beta} \approx \boldsymbol{\theta} - \boldsymbol{\alpha} \tag{1.19}$$

assuming that $\boldsymbol{\beta}, \boldsymbol{\theta}$ and $\boldsymbol{\alpha} \ll 1$ and thus are approximately equal to their tangents or sines. These are vector quantities because we are considering a mass distribution, and the angle between the differential mass elements and the photon's path will vary. For surface mass density (integrated along the line of sight to reduce the mass distribution to a single plane) at position $\boldsymbol{\xi'}$ in the lens plane, $\Sigma(\boldsymbol{\xi'})$, equation 1.18 gives us,

$$\boldsymbol{\alpha}(\boldsymbol{\xi}) \approx \frac{D_{LS}}{D_S}\frac{4G}{c^2}\int_{\mathbb{R}^2}\frac{(\boldsymbol{\xi}-\boldsymbol{\xi'})}{|\boldsymbol{\xi}-\boldsymbol{\xi'}|^2}\Sigma(\boldsymbol{\xi'})\,\mathrm{d}^2\boldsymbol{\xi'} \tag{1.20}$$

where D_{LS} is the angular diameter distance between the lens plane and the source plane, and D_S is the angular diameter distance from the observer to the source plane. Equations 1.19 and 1.20 give us a map from the lens plane to the source plane.

As mentioned, all angles involved must be small enough for the small-angle approximation to hold. In other words, the lensing effect must be relatively weak ($GM/R \ll c^2$). Additionally, the assumption of linear superposition of deflections requires a weak field approximation. Related to this limit of the strength of the field, we must assume that outside the plane of the lens there is Minkowski space. Alongside this, we must use the thin lens approximation which assumes the lens mass to be a 2-dimensional distribution in the lens plane. A final assumption that must be made is that the relative velocities of the source, lens and observer are all much less than c.

These assumptions are all valid for the strong gravitational lensing analysis techniques I use for my research. The largest deflectors involved are galaxies with coincident dark matter halos, and the extents of their mass distributions are many times smaller that the distances between the lens, source and observer. Though systems that produce multiple lensed images are referred to as "strong" lenses and thus we can use the thin-lens approximation, the deflection angles that result are on the order of milliarcseconds - small enough for the weak field approximation. While these

analyses consider deflectors all along the line of sight, their effects can be added sequentially along the line of sight, each producing a deflection in its respective plane. Though these assumptions are valid for the lens systems considered in my research, they do not apply to lensing as a result of very extended mass distributions like large scale structure. Lensing of light rays passing very close to massive objects like black holes also cannot be explained by the quasi-Newtonian approximation because the small-angle appromximation does not apply.

1.5 Strong Gravitational Lensing to Constrain Dark Matter

Strong gravitational lensing refers to lens systems in which multiple images of the source appear on the sky. It is a powerful probe of smaller dark matter halos that do not host enough luminous baryonic matter to be directly observable, as these halos will still impact the gravitational potential. Depending on the type and extent of emission that is dominant in observations of the background source, these images can be point-like and clearly separate from one another, useful for flux-ratio analysis, or extended and blended together in arcs, useful for gravitational imaging (both defined below; see Figure 1.1).

1.5.1 Flux-Ratio Analysis

When the dominant source emission is very bright and concentrated, such as an AGN, multiple point-like images result, each with a different magnification. In systems that produce 4 images, the ratios between the magnified fluxes from each image can be a marker of subhalos associated with the main lens and field halos along the line of sight halos. A smooth lens model for the main lens galaxy may accurately reproduce observed image positions, but if the flux ratios it predicts do not match observations, they are considered "anomalous". Though, the implication is that additional structures may be necessary to explain the observed flux ratios. Low-mass halos can cause local magnification perturbations that explain them, thus the frequency and magnitude of observed flux-ratio anomalies can be used to learn about clumpy dark matter. This method, flux-ratio analysis, is the subject of this dissertation.

First introduced in the late 1990s and early 2000s (eg., Mao & Schneider, 1998; Dalal & Kochanek, 2002), the power of flux-ratio-analysis was limited by the amount of usable observations. Quadruply-imaged AGN are typically first identified by continuum flux from the accretion disk around the black hole. The AGN continuum flux, as well as the broad-line emission that comes

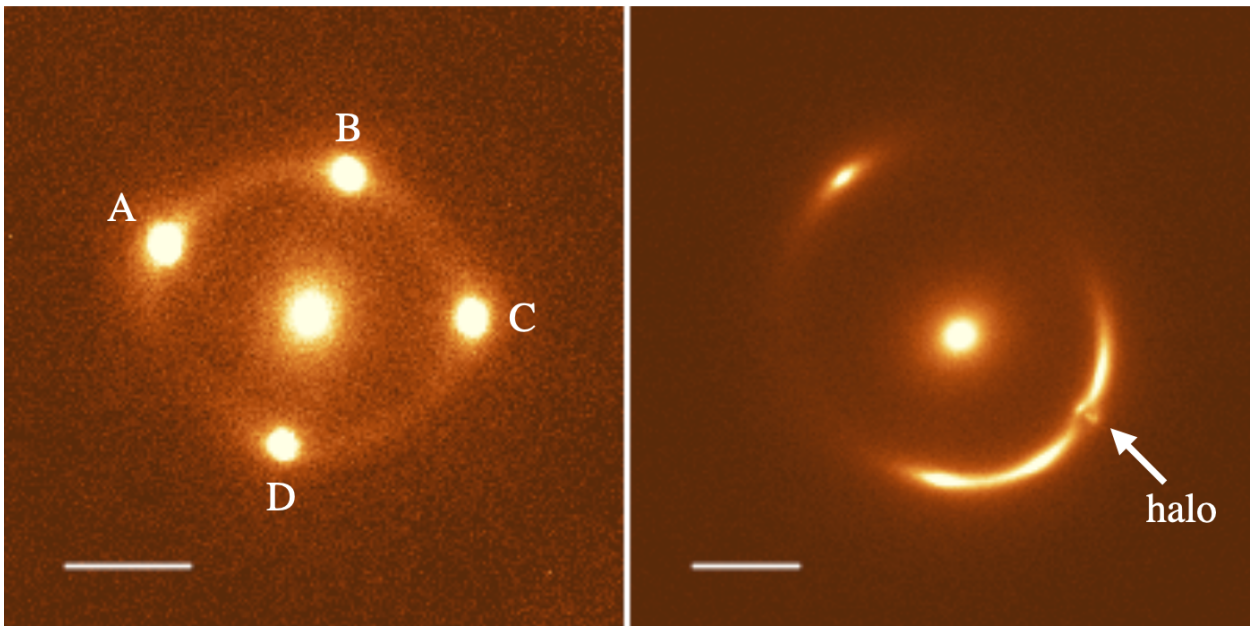


Figure 1.1 Examples of strong gravitational lensing. Each image is centered on a massive elliptical galaxy that is surrounded by multiple images and/or arcs of the background source. The scale bars indicate 1". **Left:** Quadruply-imaged quasar HE 0435–1223 from Chen et al. (2019) observed at K'-band with the Near-infrared Camera 2 (NIRC2) on the Keck II telescope. The flux ratios of the images (labeled 'A', 'B', 'C' and 'D') can reveal magnification by dark matter halos. **Right:** Another lens sytem observed using the same instrument (Shu et al., in prep). There is a clearly visible perturbation caused by a dark matter halo (labeled 'halo') in the lensed arc, though gravitational lensing is able to detect much smaller perturbations through detailed lens modeling.

from the volume surrounding the black hole, both span regions that are small enough to be subject to stellar microlensing. When the source size is smaller than the characteristic lensing scale of stars along the line of sight, magnification due to these stars has a significant effect on the observed image fluxes. Since stellar microlensing is indistinguishable from the microlensing by halos that we intend to detect, observations must be made at wavelengths at which the emission is coming from the spatially larger components of AGN. Observations of AGN radio emission, mid-infrared emission from the dust torus and more recently narrow-line region (NLR) emission have all been shown to span regions large enough to produce flux measurements unaffected by stellar microlensing (eg., Moustakas & Metcalf, 2003; Sugai et al., 2007; Patnaik et al., 1999). Each of these does present its own observational difficulty, though. Only around 10% of AGN are radio-loud, ground-based mid-infrared observations are subject to noise from thermal emission of the Earth's atmosphere and narrow-line observations require adaptive optics-fed (AO) integral field unit (IFU) instruments

or HST observations. These obstacles long prevented the accumulation of a sample size large enough to give this method the statistical power to compete with alternatives. However, recent NLR observations of 8 lens systems (Nierenberg et al., 2020a) nearly doubled the sample size, and telescopes such as the James Webb Space Telescope (JWST) and Euclid are expected to further increase it by at least an order of magnitude within the next decade.

In constraining WDM, flux-ratio analysis has most recently been applied by Hsueh et al. (2020) and Gilman et al. (2020) with independent samples of 7 and 8 lens systems, respectively. They inferred roughly the same upper bounds on the half-mode mass of $\log_{10}(M_{\rm hm}/M_\odot) < 7.8$, which translates to a lower bound on the thermal relic particle mass of $m_{\rm th} \gtrsim 5.2$keV. Considering the current expected halo detection limit of below $M = 10^6_\odot$ from this method, these findings indicate that a larger sample size should be able to improve constraints significantly (see Gilman et al., 2019). Assuming CDM is the truth and the effective half-mode mass is $\sim 10^{-6} M_\odot$, a larger sample size should bring down the upper bound asymptotically toward toward the detection limit. Below this threshold, WDM models are indistinguishable from CDM in the analysis of flux ratios.

1.5.2 Gravitational Imaging

When the dominant source emission is extended, such as a luminous galaxy, the lensed image consists of longer arcs or full rings. With high resolution observations of the arcs, detailed models can be made of the mass distribution of the host halo, subhalos around it and halos along the line of sight using a technique called gravitational imaging (Koopmans, 2005; Vegetti & Koopmans, 2009). Whereas flux-ratio analysis can only constrain parameters pertinent to the entire low-mass halo population, gravitational imaging can be used to detect the precise masses and locations of individual halos by detecting localized perturbations in extended arcs (e.g. Vegetti et al., 2010). This method has been used to detect halos down to $10^8 M_\odot$ (Vegetti et al., 2012; Vegetti & Vogelsberger, 2014), though resulting constraints on WDM (e.g. Ritondale et al., 2019) have yet to be competitive with other methods. With the milliarcsecond resolution offered by interferometry (e.g. ALMA, VLBA), gravitational imaging can theoretically probe down to $10^6 M_\odot$ (Powell et al., 2022, 2023). As more high-signal-to-noise and high-resolution data becomes available, gravitational imaging constraints should improve greatly. Additionally, lensed arcs provide much tighter constraints on the parameters of the base mass model than can be achieved from modelling only image positions

and fluxes. This information can be used to inform prior distributions for flux-ratio analysis and improve constraints from that method.

1.6 Other Astrophysical Probes of Dark Matter

It is invaluable to have independent approaches to constraining dark matter, as each entails its own assumptions, biases and sources of systematic error. While gravitational lensing avoids many astrophysical assumptions due to the well-understood nature of gravitation, it relies on assumptions about the mass distribution of the main lens that may not be accurate (see Chapter 3). Other observational phenomena that have enabled significant constraints on dark matter include Milky Way satellites and the Lyman-alpha forest.

1.6.1 Lyman-Alpha Forest

Spectroscopic observation of quasars provides a complementary method for measuring the dark matter distribution along their lines of sight (see Meiksin, 2009; McQuinn, 2016). The Lyman-alpha (Lyα) forest, a series of absorption lines blueward of the redshifted Lyα emission wavelength of a quasar, acts as a map of the neutral hydrogen along the line of sight that is causing the absorption. At redshifts $2 \lesssim z \lesssim 6$, the range in which most relevant observations are sensitive to this effect, gravitational forces dominate pressure forces on Lyα absorbing gas. As a result, the gas acts as a tracer of dark matter, the dominant gravitating entity in the Universe. For absorber redshifts $\lesssim 2$, the absorption lines exist at UV wavelengths that require satellite observations to avoid absorption by the Earth's atmosphere.

As mentioned, the Lyα forest consists of a dense group of absorption lines. The wavelength at which each line appears is related to the redshift of the absorber, and the profile of each line can be used to determine the column density. With knowledge of the background ionization rate and the average temperature at its redshift, the overdensity of the absorbing gas can then be calculated from its optical depth. As a result, each line of sight probes the overdensity of neutral hydrogen, and in turn dark matter, along it. By cross-correlating nearby lines of sight and using high resolution spectra to spatially correlate along single lines of sight, we can reveal the Lyα flux power spectrum and in turn the dark matter power spectrum that determines it. Comparison to power spectra generated from hydrodynamical simulations using various models of dark matter can then constrain the properties of dark matter candidates.

This method, like strong gravitational lensing, is sensitive to a suppression of the power spectrum at small scales. However, it allows for measurement at higher redshifts where the matter power spectrum is more linear and thus easier to reproduce with simulations. This is a unique strength of Lyα forest measurements that gravitational lensing studies cannot match. At higher redshifts, the WDM power spectrum deviates much more strongly from CDM than at lower redshifts at the same scales. Also, the decrease in the power spectrum at high wavenumbers due to free streaming of WDM is increasingly, up to at least around $z = 5$, larger that that due to Doppler broadening of absorption lines, an effect that also washes out small scale power. Thus, high redshift measurements of the Lyα forest are particularly powerful. Observations of $2 < z < 3$ quasars are also valuable since the ease of observations has allowed the accumulation of more than 50,000 quasar spectra. So far though, the uncertainty of their analysis has been unable to produce as tight constraints as higher redshift observations (e.g. Viel et al., 2005).

Though higher redshifts offer a more robust probe of the dark matter power spectrum, the Lyα forest approach to constraining dark matter requires a deeper understanding of baryonic physics. In practice, this means many assumptions are made along the way. In getting from a group of absorption lines to a map of overdensity along the line of sight, the widths of the lines are used to find the temperature of the absorbing gas. A higher temperature gas consists of particles with a wider distribution of velocities, which in turn results in a wider distribution of Lyα absorption wavelengths and a wider absorption line. Then, a redshift-dependent temperature-density relation is used to find the overdensity. The parameters of this power law relation, the mean IGM temperature and the slope, depend on the photo-heating background. The spatial fluctuations of the UV background also impact the power spectrum and are either modeled or ignored as a higher-order effect. The amplitude of the power spectrum is set by the effective optical depth along lines of sight. Additionally, the redshift of reionization impacts the level of Jeans smoothing (increased temperature/pressure inhibiting gravitational collapse) of the distribution of gas. This effect must be disentangled from the thermal broadening of absorption lines, and the power spectrum suppression of both of these must be modeled in order to discern the potential suppression from WDM. Simulations have successfully reproduced available observations of the IGM and results of reionization, but many of these variables are not well constrained and rely on marginalization over a larger range of values to produce constraints on dark matter.

Compared to strong gravitational lensing, the Lyα forest provides a view of dark matter over a larger redshift range. At redshifts comparable to those probed by strong lensing, the lower end of the mass of halos measured is of order $10^8 M_\odot$. As mentioned, at higher redshifts Lyα forest measurements get tighter constraints from larger mass scales, and the tightest constraints on WDM currently inferred from this method are similar to those recovered via flux-ratio analysis, $\log_{10}(M_{\rm hm}/M_\odot) \lesssim 8$ (Murgia et al., 2018). However, the detailed modeling of strong lensing systems allows us to resolve individual halos (gravitational imaging only) and potentially discern between the mass profiles predicted by CDM and WDM (gravitational imaging and flux-ratio analysis). The main limits of strong lensing are currently instrument resolution and sample size. On the other hand, the main limits of Lyα forest measurements are knowledge about the history of the IGM and its concurrent evolution with dark matter structures. Thus, while current bounds set by both methods are similar, their advancement will be complementary.

1.6.2 Milky Way Satellites

Dark matter constraints can also be inferred by comparing the observed population of luminous satellite galaxies in the Milky Way and other nearby galaxies to that predicted by semi-analytical models (e.g. Lovell et al., 2016) and simulations (e.g. Jethwa et al. 2018; see also Moore et al. 1999; Nierenberg et al. 2013). Born out of reasoning used to solve the missing satellite problem, this method has evolved to produce the strongest constraints on WDM to date, $\log_{10}(M_{\rm hm}/M_\odot) \lesssim 7.2$ (Nadler et al., 2021). Since our knowledge of satellite galaxies around and near the Milky Way is incomplete, the assumption of the observation selection function greatly impacts the inferred constraint. Additionally, the predicted number of luminous satellite galaxies depends on treatment of galaxy formation and the epoch of reionization in simulations (see Lacey et al., 2016).

1.7 This dissertation

The strength of flux-ratio analysis constraints on dark matter are dependent on the sample size of lens systems, and their validity requires adequate understanding and modelling of the lens galaxy mass distribution. The focus of my PhD work has been to maximize the constraining power of currently available data and to carefully investigate systematic error and biases that may affect the results. In Chapter 2, I present a joint analysis of the flux ratio samples used by Hsueh et al. (2020) and Gilman et al. (2020). I detail the creation of model CDM and WDM low-mass halo populations,

our Bayesian inference procedure and the context of this analysis within the field. We present the strongest WDM inference from strong gravitational lensing to date, $\log_{10}(M_{\rm hm}/M_{\odot}) < 7.47$, and reproduce the findings of previous works by analyzing corresponding subsamples.

In Chapter 3, I investigate the degeneracy between complex angular structure in the lens galaxy and low-mass halos when modelling quadruply-imaged quasars. We find that third- and fourth-order multipoles, which are found in observations of the isophotes of lens-like elliptical galaxies, can cause flux ratio perturbations in excess of 40 percent. Extreme simulated flux-ratio anomalies that were induced by CDM halo populations can be recovered within observational uncertainties by including realistic multipoles and no perturbing halos. Moving forward, flux ratio studies must find ways to resolve this degeneracy. Otherwise, any constraints on the nature of dark matter cannot be taken at face value.

In Chapter 4, I summarize the findings presented in this work and provide a framework for moving forward with flux-ratio analysis. This includes multiple strategies to implement multipoles within existing analysis procedures and an assessment of the viability of various ways to potentially mitigate the multipole-halo degeneracy. I conclude by commenting on the future of flux-ratio analysis as Euclid and JWST promise to dramatically increase the available sample size.

Chapter 2

Constraining Warm Dark Matter with Quadruply-Imaged Quasars

We present a Bayesian inference on thermal relic warm dark matter from joint analysis of the flux ratios of 14 gravitationally lensed quasars. At the 7:1 likelihood ratio level, we put an upper limit on the half-mode mass of $\log_{10}(M_{\rm hm}/M_\odot) < 7.47$, which corresponds to a particle mass lower limit of $m_{\rm th} > 6.63$keV. This constraint is the strongest placed on warm dark matter from gravitational lensing, however, we show that it is dependent on the precise evaluation of the likelihood function when comparing simulated realizations to observations. We also compare results derived from subsamples from the joint set to those presented in previous works and find good agreement despite differences in our analysis procedure.

2.1 Introduction

While great progress has been made in confining its behavior and distribution, the fundamental physics governing dark matter remains unknown. The cold dark matter (CDM) paradigm, which posits dark matter to consist of non-relativistic, collisionless particles, is successful in describing cosmic structure at scales larger than $\sim$ 1 Mpc (e.g. Planck Collaboration et al., 2016) and has been adopted as the standard in cosmology. However, CDM alone fails to reproduce the observed distribution of dark matter at galaxy scales and smaller (Bullock & Boylan-Kolchin, 2017). Inclusion of baryons in CDM simulations may reconcile most discrepancies (e.g. Wetzel et al., 2016; Brooks, 2014), but the potential shortcomings of CDM have motivated exploration of other dark matter particle candidates that could better match small-scale observations. Though each alter-

native model produces unique deviations from CDM, common observables used to compare their halo populations are the halo mass function (HMF) and the concentrations of halos.

The focus of this work, warm dark matter (WDM), posits dark matter particles to be more relativistic than CDM particles at the time of decoupling. This causes them to diffuse from small-density perturbations in the early universe. The result is a departure from the hierarchical structure formation predicted from CDM, a suppression of small-scale structure compared to CDM at all times, and a reduction in the central density of halos (Bode et al., 2001; Schneider et al., 2012; Lovell et al., 2014).

In addition to the gravitational lensing method discussed in this paper, other approaches to constrain alternative dark matter models have been developed over the past few decades that presently provide similar constraining power. By comparing the population of luminous satellite galaxies in the Milky Way to expectations from theory and simulations, Nadler et al. (2021) derived a lower limit of 6.50 keV on the thermal relic particle mass at the 95 per cent confidence level. This is the strongest constraint on the thermal relic WDM particle mass to date. However, this method is limited by the availability of high-resolution simulations of Milky Way-like galaxies and halos. Since the IGM acts as a tracer of the dark matter distribution in the Universe, probing it with the Lyman-α forest also provides a way to constrain dark matter candidates (Iršič et al., 2017; Murgia, 2018; Murgia et al., 2018) that has provided similar and slightly weaker constraints than lensing and Milky Way satellites depending on assumptions about the IGM. This method is highly dependent on the thermal history of the IGM and other baryonic physics (Viel et al., 2013). Disruptions of stellar streams by non-luminous halos also provide a way to probe the HMF at small scales (Banik et al., 2018, 2021). Due to a relatively small sample size, stellar stream analyses have yet to constrain WDM as strongly as other methods. Combining results from many of these complementary approaches, Enzi et al. (2021) derived a joint lower limit of 6.05 keV on the thermal relic particle mass at the 95 per cent confidence level.

Strong gravitational lensing offers a way to probe the distribution of matter in the universe without requiring the matter to be baryonic or luminous. Subhalos and line-of-sight (LOS) halos produce perturbations in the lensing potential expected from a smooth lens model. The resulting effects on lensed images are thus indicative of small-scale structure. When the dominant source emission is extended, such as a luminous galaxy with no bright active galactic nucleus (AGN),

the lensed image consists of arcs or rings. Gravitational imaging (Vegetti & Koopmans, 2009; Ritondale et al., 2019) allows for detection of perturbing halos without assuming a parametric form for the substructure.

When the dominant source emission is very bright and compact, such as an AGN, multiple point-like images result, each with its own magnification that is set by the second derivative of the lensing potential. Subhalos and field halos can further magnify or demagnify images, introducing a discrepancy between smooth models and observations. Since the population statistics and properties of these additional structures that perturb flux ratios are governed by the nature of dark matter, we can put constraints on different dark matter models based on their ability to reproduce observed flux ratios. Warmer (lower particle mass) WDM models produce fewer and less dense perturbers, so the prevalence of anomalous flux ratios provides lower bounds on the potential particle mass. On the other hand, a multitude of non-detections – lens systems with flux ratios that are well-described by smooth mass models – could similarly provide an upper bound.

Since its theoretical conception (Mao & Schneider, 1998; Metcalf & Madau, 2001) and first implementation (Dalal & Kochanek, 2002), the flux-ratio analysis method has significantly advanced. Following research demonstrating the significant impact of baryonic structures on flux ratios, we model stellar disks in galaxies in which they are observed Hsueh et al. (2016, 2017, 2018); Gilman et al. (2018). Additionally, Despali et al. (2018) and Gilman et al. (2019) showed the importance and frequent dominance of LOS halos over galactic subhalos, and simulations have illuminated more detailed ways of modelling the mass-concentration relation of subhalos and its scatter for various WDM models and at different redshifts (Duffy et al., 2008; Lovell et al., 2014).

The power of flux-ratio analyses have been limited by small sample sizes of suitable observational data. Optical emission from Type 1 quasars is usually dominated by emission from the accretion disk. However, when the size of a source is smaller than the scale of the lensing caustics of stars in the lensing galaxy, magnification due to these stars has a significant effect on the observed image fluxes. The angular size of the accretion disk is small enough to be subject to this microlensing by stars. Thus, observations must be made at different wavelengths to measure the spatially larger components of AGN. Radio, mid-infrared (MIR) and narrrow-line region (NLR) observations all fulfill this criterion, though each presents its own observational difficulties that have thus far kept growth of the sample size relatively slow. Recent NLR observations of 8 lens

systems (Nierenberg et al., 2020a), however, nearly doubled the sample size. Gilman et al. (2020, hereafter G20) analyzed 8 NLR systems, and Hsueh et al. (2020, hereafter H20) analyzed 7 Radio or MIR systems. Both works reported similar 95 per cent confidence level lower limits on the thermal relic WDM partical mass of $m_{\rm th} > 5.2$keV and $m_{\rm th} > 5.6$keV, respectively.

In this work, we perform a joint analysis on these two lens samples. With the largest flux ratio sample to date and an updated technique that accumulates developments across the field, we present the tightest constraints yet on WDM from gravitational lensing.

2.2 Methodology

In this section, we review our flux-ratio analysis framework. This framework is largely similar to that described in H20, but we have made a number of modifications. Section 2.2.1 details the lens macro-models including both the dark and baryonic components where applicable. In Sections 2.2.2 and 2.2.3, we present our calculations for the population statistics of subhalos and LOS halos, respectively. Sections 2.2.4-2.2.6 describe the combination of these components into complete model realizations. We then describe our analysis procedure in Section 2.2.7. Finally, we emphasize the differences between our analysis procedure and those used in H20 and G20 in Sections 2.2.8 and 2.2.9, respectively.

2.2.1 Macro Model

For the initial macro-modelling of the quad lens systems, we use only the observed image positions and not their fluxes. As our later analysis allows these fluxes to be different from their expected smooth model values due to mass perturbations, we do not expect macro-models to be able to reproduce them simultaneously with the observed positions. We model each main lens as a singular isothermal ellipsoid (SIE) with external shear. The SIE density as a function of position with respect to the center is defined as

$$\kappa(x, y) = \frac{b}{2\sqrt{(1-\epsilon)x^2 + (1+\epsilon)y^2}} \quad , \tag{2.1}$$

where b is the Einstein radius and the ellipticity, ϵ, is related to the axis ratio, q, by

$$q^2 = \frac{1-\epsilon}{1+\epsilon} \quad . \tag{2.2}$$

For systems with an observed satellite, we model the satellite as a singular isothermal sphere (SIE with $\epsilon = 0$) with its position fixed from observations and its Einstein radius as a free parameter. For systems with an observed stellar disk, we introduce an exponential disk component with density defined by

$$\kappa(x, y) = q^{-1}\kappa_0 \exp\left(-\frac{\xi(x, y)}{r_s}\right) \quad , \tag{2.3}$$

where q is the axis ratio, κ_0 is the central density, r_s is the scale length and

$$\xi(x, y) = \sqrt{\frac{x^2 + y^2}{q^2}} \quad , \tag{2.4}$$

where (x, y) is the position with respect to the disk center. The position of the disk's center, its axis ratio and its position angle are roughly inferred from imaging then optimized along with κ_0, r_s and the SIE parameters. For all lenses, we used the LENSMODEL feature of GRAVLENS (Keeton, 2001) to optimize the macro-model parameters.

2.2.2 Substructures

We quantify the "warmth" of WDM models with the half-mode mass, $M_{\rm hm}$, which is the mass scale at which the transfer function is suppressed by one-half compared to that of CDM. For CDM as WIMPs, $M_{\rm hm} \approx 10^{-6} M_\odot$ (Schneider et al., 2013). $M_{\rm hm}$ can be related to the thermal relic particle mass, $m_{\rm th}$, via

$$M_{\rm hm} = \frac{4\pi}{3}\bar{\rho}\left(6.97\lambda_{\rm fs}^{\rm eff}\right)^3 \quad , \tag{2.5}$$

(Schneider et al., 2012), where $\bar{\rho}$ is the background density of the universe and $\lambda_{\rm fs}^{\rm eff}$, the effective free-streaming length scale, is defined by

$$\lambda_{\rm fs}^{\rm eff} = 0.049\left(\frac{m_{\rm th}}{\rm keV}\right)^{-1.11}\left(\frac{\Omega_{\rm th}}{0.25}\right)^{0.11}\left(\frac{h}{0.7}\right)^{1.22} {\rm Mpc\ h^{-1}} \tag{2.6}$$

(Viel et al., 2005). Combining these and assuming $\Omega_{\rm th} = 0.25$ and $h = 0.68$ (Planck Collaboration et al., 2016), the relationship between the half-mode mass and the thermal relic paricle mass is

$$m_{\rm th} = 2.30\left(\frac{M_{\rm hm}}{10^9}\right)^{-0.3} {\rm keV} \quad . \tag{2.7}$$

Thus, a lower WDM particle mass results in a higher half-mode mass and a warmer model with suppression of structures up to higher mass scales.

We model all subhalos with spherical NFW profiles (Navarro et al., 1997) with concentrations determined from a mass-concentration relation (detailed in Chapter 3) derived from simulations (Ishiyama & Ando, 2020; Moliné et al., 2023). As Despali et al. (2018) have shown the effects to be small compared to the scatter on the mass-concentration relation, we do not include tidal truncation or a dependence on the distance from the main lens center of the mass-concentration relation. We parameterize the CDM subhalo mass function (SHMF, the number of subhalos within the mass range $m, m + dm$) as

$$n_{\mathrm{CDM}}(m) = \frac{dN_{\mathrm{sub}}(m)}{dm} = A_0\, m^{-1.9} \tag{2.8}$$

(Schneider et al., 2012; Lovell et al., 2014), where A_0 is a normalization factor set by the fraction of mass in substructure for CDM, f_{sub}. To calculate and define f_{sub}, we consider a region that is twice the Einstein radius (θ_{E}) of the host halo. We use this $2\theta_{\mathrm{E}}$ boundary to define the projected mass of the host halo, M_{proj}, and the total mass and number of subhalos generated using a CDM mass function, M_{sub} and N_{sub}, respectively. We then define f_{sub} as

$$f_{\mathrm{sub}} = \frac{M_{\mathrm{sub}}}{M_{\mathrm{proj}}} = \frac{\int_{M_{\mathrm{low}}}^{M_{\mathrm{high}}} n_{\mathrm{sub}}(m)\; m\; dm}{M_{\mathrm{proj}}} \quad , \tag{2.9}$$

with $(M_{\mathrm{low}}, M_{\mathrm{high}}) = (10^6, 10^9) M_\odot$. We expect subhalos below this mass range to be below our detection threshold and halos above this mass range to contain enough luminous baryonic mass to be directly observed (see e.g. Yuan et al., 2022).

For WDM models, we implement a suppression factor such that

$$\frac{n_{\mathrm{WDM}}(m)}{n_{\mathrm{CDM}}(m)} = \left(1 + \frac{M_{\mathrm{hm}}}{m}\right)^{-1.3} \tag{2.10}$$

(Schneider et al., 2012; Lovell et al., 2014). Since f_{sub} and M_{proj} are explicitly chosen for each model, we can determine A_0. Though we calculate its value to produce the desired f_{sub} assuming CDM ($M_{\mathrm{hm}} = 0$), we use the same A_0 for all WDM models with the same desired f_{sub}. This

ensures that at mass scales higher than that at which there is suppression of the WDM SHMF, the CDM and WDM SHMFs align. As a result, $f_{\rm sub}$ for WDM models is not equal to the true fraction of the mass in substructures.

2.2.3 Line of Sight

In addition to the flux perturbations from dark matter in subhalos of the main lens, significant and often dominant effects come from halos along the line of sight (Despali et al., 2018). We consider the relevant volume to be contained within two cones that share a base with radius $2\theta_{\rm E}$ in the lens plane. One cone intersects a circle with radius $\theta_{\rm E}$ in the source plane, and the other comes to a point in observer planes. This volume is centered on the lensing galaxy. We then divide this double-cone into 200 equally-spaced redshift slices. This spacing is small enough to preserve reasonable agreement between line-of-sight halo populations in different realizations using the same model parameters while ensuring computational efficiency.

For the line-of-sight HMF, we use the parameterization described by Sheth & Tormen (1999) and consider halos in the mass range $10^{5.26} M_\odot < m_{\rm vir} < 10^{10.88} M_\odot$. Specifically, we follow the implementation of Despali et al. (2016) and use their best-fitting parameters (shown to be universal when halos are identified by virial overdensity) optimized over all considered redshifts and cosmologies. We notate the LOS HMF as $n_{\rm los}(m, z_i)$, where z_i is the redshift of a particular plane along the line of sight. Using the volume of each redshift slice and the HMF evaluated at its midpoint redshift, we calculate the expected number of line-of-sight halos. We use the same suppression factor for WDM models here as we did for the subhalo mass function (see equation 2.10). For the line-of-sight mass-concentration relation we use that reported by Duffy et al. (2008) from N-body simulations using the virial radius definition of relaxed halos between redshifts 0 and 2, and we apply the associated scatter on the parameters.

2.2.4 Source

The size of the narrow-line region, the largest of the AGN components used for flux-ratio analysis, ranges from 1pc to up to 60pc (Nierenberg et al., 2017), and the effect of perturbers on the magnification of lensed images decreases with increasing source size. For increased computational efficiency, we use point-like sources in our analysis. However, the lens systems that most strongly impact our results come from the Radio/MIR sample, and are thus expected to have sources smaller

than 10pc (Burtscher et al., 2013; Kim et al., 2022). As a result, we expect this choice to have minimal impact on our findings.

2.2.5 Likelihood Calculation

We adopt a Bayesian approach to evaluate the posterior probability densities of the mass function parameters, $M_{\rm hm}$ and $f_{\rm sub}$, which we will refer to as $\vec{\theta}$. We will refer to the macro-model parameters that describe the main lens SIE and external shear (Γ), source information and any observed satellites or disks as $\vec{\theta}_M$, and we will refer to the perturber parameters that describe substructures' and LOS halos' quantity, masses, concentrations and positions as $\vec{\theta}_m$. By Bayes' theorem, the posterior probability density of $\vec{\theta}$ given the observed data $\mathbf{d}$ is,

$$P(\vec{\theta}|\mathbf{d}) \propto \int P(\mathbf{d}|\vec{\theta},\vec{\theta}_M,\vec{\theta}_m)P(\vec{\theta}_m|\vec{\theta},\vec{\theta}_M) \times P(\vec{\theta})P(\vec{\theta}_M)d\vec{\theta}_M d\vec{\theta}_m \quad . \tag{2.11}$$

$P(\vec{\theta}_m|\vec{\theta},\vec{\theta}_M)$ is detailed in Sections 2.2.2 and 2.2.3, and we discuss the choice of priors on $\vec{\theta}$ and $\vec{\theta}_M$ in Section 2.2.6. Assuming Gaussian errors on the observed fluxes, we approximate the likelihood function as,

$$P(\mathbf{d}|\vec{\theta},\vec{\theta}_M,\vec{\theta}_m) \approx \exp\left(-\frac{\chi^2_{\rm flux}+\chi^2_{\rm pos}}{2}\right) \quad . \tag{2.12}$$

2.2.6 Priors

To implement a prior on the macro-model parameters, we use Markov chain Monte Carlo (MCMC) to sample the parameter space. As with the original model, we use only the image positions and not the fluxes to evaluate χ^2 since in many cases perturbing masses are needed to reproduce observed fluxes. For each lens system, we evaluate 10^5 MCMC steps, leaving all SIE+Γ parameters free for the main halo, the Einstein radius free for any observed satellites, and the mass free for observed stellar disks. We use the EMCEE package for all MCMC (Foreman-Mackey et al., 2013). We also let the source position vary freely. For each lens, we ensure chain convergence and a densely sampled parameter space.

We set the priors on $\vec{\theta}$ to be uniform in logarithmic space and encapsulate the entire ranges that are consistent with existing constraints and our ability to detect differences between models. For $M_{\rm hm}$, we cover $10^6 < M_{\rm hm} < 10^{13} M_\odot$, and for $f_{\rm sub}$ we cover $10^{-4} < f_{\rm sub} < 10^{-1.25}$ (.25 dex spacing for both parameters). This results in a 29×11 grid of parameter combinations, each of

which describes a WDM model. For each of these model grid points, we use the same set of randomly selected macro-model realizations.

2.2.7 Analysis Scheme

For each realization, we draw a sample of perturbers according to the mass functions set by the dark matter model parameters and macro-model. For subhalos, we start by calculating the expected total number, $\mu_{\rm sub}$, from the subhalo mass function, $n_{\rm sub}(m)$ (equation 2.8). The subhalo mass function has a slope dependent on $M_{\rm hm}$ and normalization dependent on both $f_{\rm sub}$ and the mass of the host halo. For each realization, we draw the total number of subhalos, $N_{\rm sub}$, from a Poisson distribution with expectation value $\mu_{\rm sub}$. From a probability distribution created by normalizing the subhalo mass function to unity,

$$P_{\rm sub}(m) = \frac{n_{\rm sub}(m)}{\int_{M_{\rm low}}^{M_{\rm high}} n_{\rm sub}(m')dm'} \quad , \tag{2.13}$$

we randomly draw $N_{\rm sub}$ subhalo masses. The subhalos are all placed in the redshift plane of the main lens, and the angular positions of the resulting set of subhalos are assigned by drawing randomly from a uniform spatial distribution within $2\theta_{\rm E}$ of the lens center. While the spatial distribution of subhalos is not truly uniform around the center of the host halo (e.g. tidal disruption prevents more massive subhalos from surviving near the center), simulations show this to be approximately true in projection (Xu et al., 2015).

For LOS halos, generating realizations is more involved since the mass function, $n_{\rm los}(m, z_i)$, is a number density that is dependent on redshift in addition to $M_{\rm hm}$ (it is independent of $f_{\rm sub}$ and the host halo mass). To find an expectation value for the total number of LOS halos, $\mu_{\rm los}$, we calculate the expected number of halos in each redshift bin, μ_i, and take the sum as,

$$\mu_{\rm los} = \sum_i \mu(z_i) = \sum_i V_i \int_{M_{\rm low}}^{M_{\rm high}} n_{\rm los}(m, z_i)dm \tag{2.14}$$

where V_i is the volume of the ith redshift bin. For each realization we draw the total number of LOS halos, $N_{\rm los}$, from a Poisson distribution with mean $\mu_{\rm los}$. The redshifts of the LOS halos are then drawn from a probability distribution created by integrating the mass function over mass and

normalizing to unity as

$$P_{\rm los}(z_i) = \frac{\mu(z_i)}{\mu_{\rm los}} \quad . \tag{2.15}$$

The masses of these halos are then drawn from probability distributions created by normalizing to unity the mass function for each individual redshift bin as

$$P_{\rm los}(m|z_i) = \frac{n_{\rm los}(m, z_i)}{\int_{M_{\rm low}}^{M_{\rm high}} n_{\rm los}(m', z_i) dm'} \quad . \tag{2.16}$$

The angular positions are assigned randomly from a uniform spatial distribution within the boundaries defined by the double cone structure as outlined in Section 2.2.3. Finally, we remove mass from the main lens to account for that added by the perturber population.

For each ($M_{\rm hm}$, $f_{\rm sub}$) grid point, we stochastically generate 2000 model realizations consisting of a macro-model, subhalos and LOS halos. For realizations that contain at least one perturber, we then optimize the macro-model parameters to account for small shifts in image positions due to the added mass. For some lens systems, high-perturber realizations do not produce 4 images even after attempted optimization. For these realizations, we assign the likelihood to be null.

2.2.8 Changes from Hsueh et al. (2020)

Our flux-ratio analysis procedure is an extension of that used by H20. We have, however, made a number of improvements that keep this research up-to-date with WDM simulations and theory, increase computational efficiency and reduce systematic error. Whereas previous analysis by H20 used multi-dimensional Gaussian fits to the posterior to draw macro-model realizations, we draw directly from random permutations of the MCMC chains. We found this method to be computationally efficient and to produce samples that are representative of the true parameter space. For all $M_{\rm hm}$-$f_{\rm sub}$ grid points for a lens, we use the same sample of macro-model draws. Another key difference between these two approaches is the way we address the rarity of matching a macro-model to a perturber population that actually brings the flux ratios closer to the observed values. For colder models with many perturbers, most random combinations of macro-models and perturbers will result in flux ratios and image positions that are worse than the macro-model alone. H20 drew tens of thousands of realizations for each grid point for each lens (requiring upwards of 10^7 realizations per lens) to account for this. Instead, we optimize the macro-model

parameters for each realization after the perturbers are been added (see Section 2.4.1). While increasing the computing time required for each realization, this greatly reduces the number of ill-fitting realizations with insignificant likelihood contributions and increases the efficiency of our analysis.

For both LOS halos and subhalos, we also update the concentration-mass relation. H20 used the relation from Duffy et al. (2008) without the associated scatter for both. As mentioned in Section 2.2.3, we implement scatter on this relation for LOS halos. We adopt an entirely different approach dependent on both mass and redshift that is described in Chapter 3. In addition to these changes, we remedied small bugs in the code we adopted related to handling realizations with low expectation values for the number of perturbers and factors of the dimensionless Hubble constant.

2.2.9 Comparison to Gilman et al. (2020)

To aid in the comparison between this work and previous analysis of the NLR lens systems in our sample (excluding SDSS J1330+1810), we detail some of the key differences between our analysis procedure and that used by G20. We uniformly sample the parameter space of $M_{\rm hm}$ and $f_{\rm sub}$ by evaluating the likelihood over a grid of values allowed by our priors, using the same sets of MCMC-generated macro-model parameters with perturber populations generated at each grid point. G20 simultaneously sampled $M_{\rm hm}$ and $f_{\rm sub}$ along with the macro-model parameters for each lens. They also allow the power-law slope of the main deflector to vary, while we keep it fixed so that they are all isothermal. G20 also simultaneously samples the logarithmic slope of the subhalo mass function with a uniform prior between -1.95 and -1.85 whereas we leave it fixed at -1.9 (see equation 2.8). They report evaluating between 300,000 and 1,200,000 realizations per lens system, and we evaluate between 600,000 and 2,500,000.

When evaluating the likelihood contribution for each realization, G20 only include models that predict the observed image positions and use a summary statistic calculated solely using model-predicted and observed fluxes. Lens system-dependent astrometric perturbations and flux uncertainties are sampled simultaneously with other model parameters and added to the observed values before comparison to predicted values. As described in Section 2.2.5, we evaluate the likelihood contribution for each realization using the combined χ^2 from image positions and fluxes, and we use fixed values for the uncertainties on both for each lens. To allow for a readier comparison

with their work, we also include inferences calculated using only fluxes from realizations that match the observed image positions within 3σ.

2.3 Data

The 14 quadruply imaged quasar lens systems in our sample consist of 6 with radio or MIR flux measurements and 8 with NLR flux measurements. We model all lensing galaxies with singular isothermal ellipsoids with shear components. We model measured satellites with singular isothermal spheres, and we model measured stellar disks with exponential disks. The best-fit model parameters for each lens are shown in Table 3.2.

2.3.1 NLR Systems

For 7 of the 8 NLR systems, WGD J0405-3308, RX J0911+0551, SDSS J1330+1810, PS J1606-2333, WFI 2026-4536, WFI 2033-4723 and WGD J2038-4008, we collect all position and flux measurements from Nierenberg et al. (2020a). We collect position and flux measurements for lens system HE0435-1223 from Nierenberg et al. (2017). For all of these, they used the Hubble Space Telescope (HST) F105W/F140W direct imaging for image and galaxy positions, and they measured narrow line fluxes from G102/G141 spectra. Though they also report measurements for HS 0810+2554, we exclude it from our analysis. Due to the unusually high magnification of merging images A1 and A2, there is likely to be blending that renders analysis of the flux ratios inappropriate.

RX J0911+0551, PS J1606-2333 and WFI 2033-4723 all have satellite galaxies, and SDSS J1330+1810 has an associated stellar disk. All of our lens models were created independently of those included in Nierenberg et al. (2020a), but with the exception of SDSS J1330+1810 (they did not model the stellar disk) they show good agreement in both model parameters and resulting model flux ratios.

The NLR sample mostly overlaps with that used in G20 with the exceptions of our inclusion of SDSS J1330+1810 and exclusion of JVAS B1422+231. Instead, we include JVAS B1422+231 in our Radio/MIR sample since the errors on image positions and fluxes reported in Patnaik et al. (1999) are smaller than the NLR ones reported in Nierenberg et al. (2014).

Table 2.1 Macro-model parameters and image position/flux references for all lens systems. [a] Lens redshift. *b* Source redshift. [c] Einstein radius. [d] Lens right ascension with respect to observation center. [e] Lens declination with respect to observation center. [f] Ellipticity. [g] Position angle. [h] External shear. [i] Shear angle. [j] Exponential disk central density. [k] Exponential disk scale length.

Lens	z_l^a	z_s^b	$\theta_{\rm E}^c$	dRA^d	dDec^e	ϵ^f	ϕ^g	$\gamma_{\rm ext}^h$	$\phi_{\rm ext}^i$	κ_0^j	$r_{\rm s}^k$	Reference
Radio/MIR												
CLASS B0712+472	0.41	1.34	0.609	0.785	-0.158	0.124	-71.25	0.096	-34.44			Hsueh et al. (2017)
disk				0.896	-0.100	0.768	-59.65			0.294	0.389	
CLASS B1555+375*	1.0	1.5	0.197	-0.193	-0.172	0.086	-0.91					Hsueh et al. (2016)
disk				-0.159	-0.292	0.850	-6.54			0.181	0.265	
CLASS B1608+656	0.63	1.39	0.780	0.425	-1.071	0.394	76.80	0.085	6.80			Koopmans & Fassnacht (1999)
satellite			0.262	-0.291	-0.929	0.66	69.2					
CLASS B0128+437	1.15	3.12	0.212	0.210	-0.107	0.524	24.3	0.230	-43.56			Phillips et al. (2000)
PG 1115+080	0.31	1.72	1.139	0.361	-1.337	0.142	76.60	0.116	-52.77			Chiba et al. (2005)
JVAS B1422+231	0.34	3.62	0.747	0.765	-0.673	0.378	55.56	0.147	-52.49			Patnaik et al. (1999)
NLR												
SDSS J1330+1810	0.373	1.383	0.807	-0.230	0.904	0.138	-44.72					Nierenberg et al. (2020a)
disk				-0.222	1.215	0.702	-22.31			0.126	1.127	
RX J0911+0551	0.77	2.76	1.067	-0.876	0.332	0.277	33.91	0.300	-13.31			Nierenberg et al. (2020a)
satellite			0.1	-1.655	0.974							
PS J1606-2333	0.31	1.70	0.658	0.790	0.248	0.127	-47.90	0.2135	26.65			Nierenberg et al. (2020a)
satellite			0.008	0.337	-1.032							
WFI 2033-4723	0.66	1.17	1.080	-1.480	0.315	0.206	-48.77	0.097	-17.31			
satellite			0.178	-1.035	2.398							Nierenberg et al. (2020a)
WFI 2026-4536	1.04	2.23	0.652	-0.073	-0.823	0.211	73.17	0.086	-75.42			Nierenberg et al. (2020a)
WGD J0405-3308	0.29	1.71	0.696	0.381	0.560	0.142	-74.30	0.056	2.31			Nierenberg et al. (2020a)
WGD J2038-4008	0.23	0.78	1.361	-0.351	0.703	0.217	-38.37	0.051	-30.85			Nierenberg et al. (2020a)
HE0435-1223	0.45	1.69	1.120	1.297	0.035	0.108	8.50	0.060	17.14			Nierenberg et al. (2017)

2.3.2 Radio and MIR Systems

We include 5 of the same radio and MIR lens systems and references as did H20. These include CLASS B0128+437, CLASS B0712+472, PG 1115+080, JVAS B1422+231, and CLASS B1555+375. We do not include MG J0414+0534 since the source is quite extended (about 200 pc in a linear extent; see Stacey et al. (2020)) and cannot be considered a point-like source. CLASS B0712+472 and CLASS B1555+375 have observed stellar disks which we include in our models. Our lens models were created independently of those included in H20, but there is mostly good agreement between them aside from CLASS B1555+375.

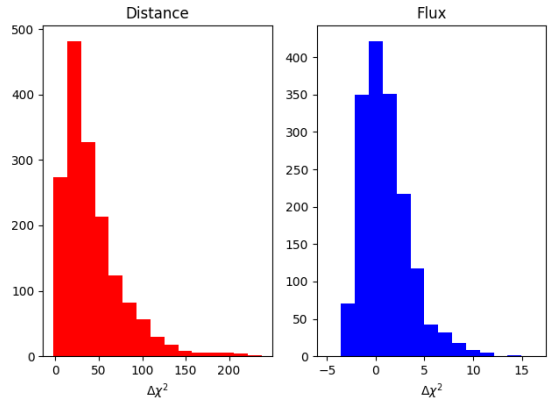
We also include CLASS B1608+656 (Myers et al., 1995; Fassnacht et al., 1999, 2002), which has two "main" lensing galaxies. For this lens, we model both lensing galaxies as SIEs with nonzero ellipticity (all other satellite galaxies are modeled as spherically symmetric). We base our initial lens model on that reported in Fassnacht et al. (2002).

In the case of CLASS B1555+375, our lens model differs significantly from that used by H20. In the methodology employed by Hsueh et al. (2016), the center position, the ellipticity and position angle of the exponential disk are roughly inferred from NICMOS F160W imaging. These four quantities are fixed in further lens modelling and analysis. They found that this SIE+exponential disk model is able to match observed flux ratios significantly better than an SIE+Γ model, but it still leaves two flux ratios with approximately 2σ discrepancies between modelled and observed. However, when we allow the center of the disk to vary with respect to the center of the SIE, we find a model that fits all image positions and flux ratios within 1σ. The resulting model still aligns qualitatively with the imaging of the disk. We leave further investigation of the comparison of these models for future work and proceed with the updated model that best fits the observed data.

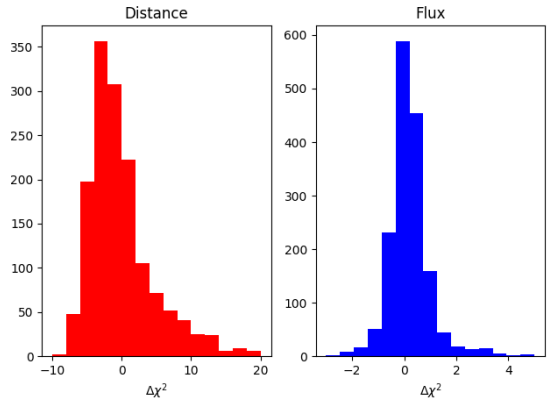
2.4 Results

2.4.1 Perturbers and Macro Model Optimization

Though perturbations to a smooth lensing potential via subhalos and LOS halos are often necessary to reproduce observed flux ratios between images, the majority of the randomly generated realizations in our analysis do not fit the observed data better than the smooth model. Optimizing the macro-model parameters after adding perturbers greatly improves the positions of most of these otherwise poor-fitting realizations. Despite this, the likelihoods calculated for each model grid



(a) Adding perturbers (subhalos and LOS halos) to macro-model without optimizing.



(b) Total change from adding perturbers and optimizing macro-model.

Figure 2.1 Histograms showing the changes in distance (left) and flux (right) χ^2 from adding perturbers and re-optimizing the macro-model parameters in 2000 realizations of PG 1115+080 with $M_{\rm hm} = 10^6 M_\odot$ and $f_{\rm sub} = 10^{-2}$. 367 of these 2000 realizations resulted in a number of images other than four, and they are not included in the histograms. In our likelihood calculations, these are assigned infinitely small likelihoods.

point are dominated by a small group of the best-fitting realizations.

Figure 2.1 shows the change in fit (measured by χ^2) of both image positions and fluxes due to the addition of perturbers to a smooth macro-model of and subsequent macro-model optimization for PG 1115+080. While the magnitudes of the changes differ between lens systems, the general behavior is similar for all lens systems with flux ratios that cannot be fit by a smooth macro-model. As shown in Figure 2.1a, the addition of perturbers, while improving some flux ratio fits and worsening others, worsens the image position fits for almost all realizations. This is expected since the macro-model parameters were chosen to fit the positions well alone. The optimization largely corrects both the positions and fluxes that were worsened by the addition of perturbers. The overall change from adding perturbers and then optimizing macro-model parameters is demonstrated in Figure 2.1b. While a large portion of realizations end up fitting worse than the smooth models alone, there is a significant group remaining with improved fluxes and/or positions. Of these 2000 realizations, 911 have overall improved fluxes, 641 have improved positions and 322 have both improved positions and fluxes.

Figure 2.2 shows the top four realizations from the same subsample of 2000 realizations for PG 1115+080 as Figure 2.1. PG 1115+080 is a lens system with fluxes that cannot be exactly

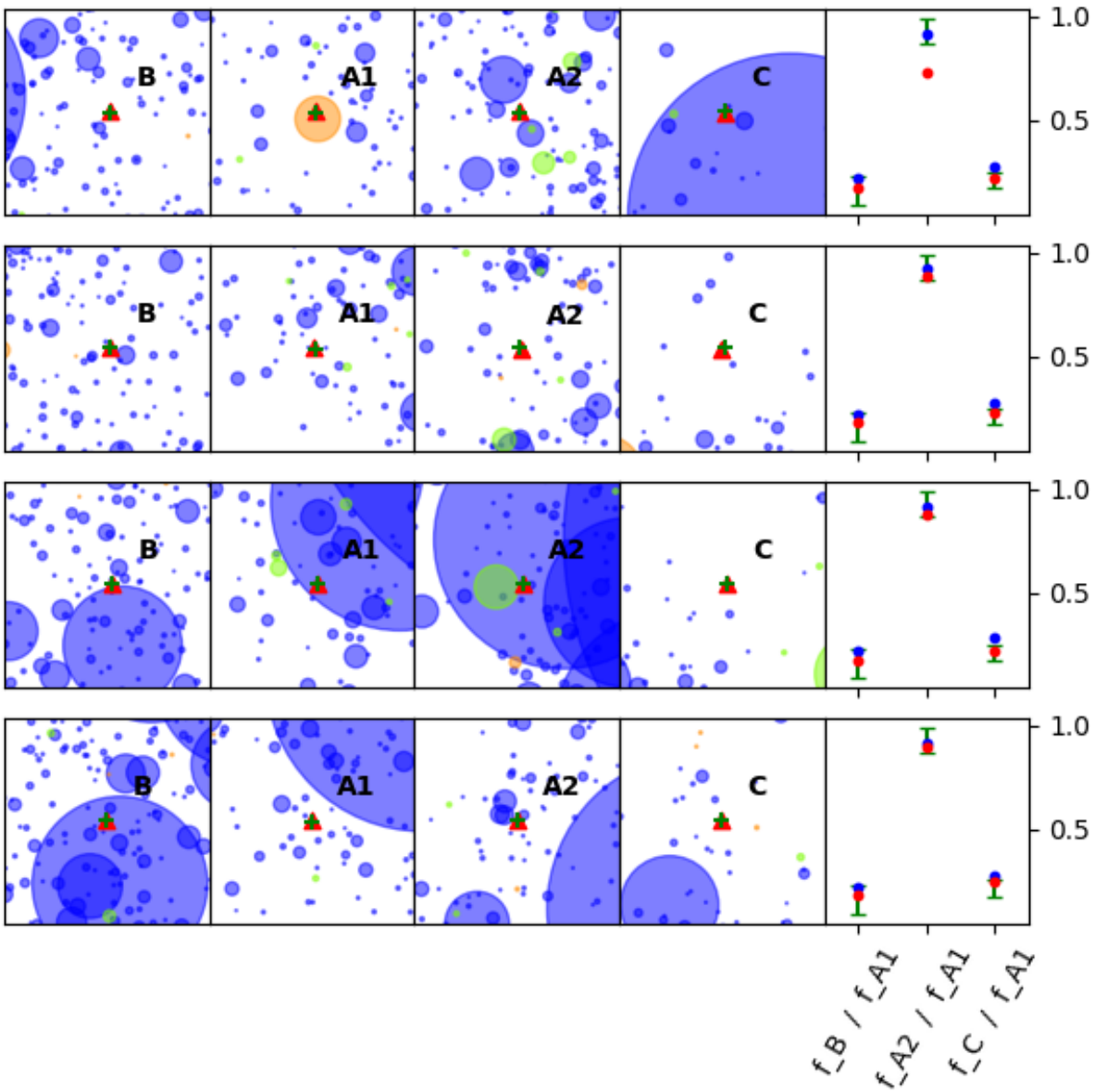


Figure 2.2 The top 4 highest likelihood (flux+pos) realizations for PG 1115+080 ($M_{\rm hm} = 10^6 M_\odot$, $f_{\rm sub} = 10^{-2}$). The first four columns from the left show 0.1" cutouts around the four images with the foreground (orange) and background (blue) LOS halos and subhalos (green) plotted as circles. The perturbers are all scaled proportionally to their Einstein radii. Observed image positions are shown by green crosses, and model-predicted (macro+perturbers+re-opt) image positions are shown by red triangles. The column on the right shows the flux ratios corresponding to each realization. The green brackets show the $\pm 1\sigma$ range of the observed flux ratios, the blue dots show the macro-model predicted flux ratios, and the red dots show the macro+perturbers+re-opt predicted flux ratios.

reproduced by a smooth lens model. Our smooth models over-predict flux ratio C/A1 and fail to produce any realizations with C/A1 within 1σ of the observed value. Smooth models do produce B/A1 flux ratios within 1σ of the observed value, but the distribution is centered above the observed value. The realizations with the best fits have numerous (or a few massive) perturbers in the vicinity of these images that are decreasing the fluxes of images B and C and/or increasing the flux of image A1. In the realization at the top of Figure 2.2, there is a foreground LOS halo magnifying A1 such that the flux ratios B/A1 and C/A1 match the predicted ones significantly better than the smooth model. Despite the fact that the A2/A1 flux ratio is made worse than the smooth model, there is still overall improvement in the fit and this realization is ranked highly. Typically, large differences between the likelihood contributions of the top few samples indicates that more realizations are necessary. Indeed, with only 2000 realizations, PG 1115+080 returns a very noisy contribution to our inference on $M_{\rm hm}$. In our complete analysis of this lens (shown in Section 2.4.2), we generate 8000 realizations for the inference to converge to a stable value.

2.4.2 Example Inferences from Individual Lens Systems

In this section, we present individual results from 3 lens systems to highlight specific features within them. We include similar plots for the remaining lens systems in Appendix A.

2.4.2.1 PG 1115+080

As shown in Figure 2.3, PG 1115+080 strongly favors colder models which have more perturbers. As shown in Section 2.4.1 and much prior analysis of this lens (Chiba et al. 2005; H20), smooth models are unable to exactly fit the observed flux ratios of this lens. Warmer models with $M_{\rm hm} > 10^{10} M_{\odot}$ generate few to no perturbers regardless of $f_{\rm sub}$ due to mass function suppression. Thus, they produce realizations with flux ratios approximately unchanged from those of the input smooth models. Increasingly colder models produce increasingly more perturbers for all $f_{\rm sub}$ values, and the resulting realizations have a wider distribution of flux ratios which include the measured values. The likelihood monotonically increases as $M_{\rm hm}$ decreases, showing that colder models more often produce realizations that match the observations.

Though this lens does not provide nearly as strong a constraint on $f_{\rm sub}$ as it does on $M_{\rm hm}$, the marginalized likelihood for lower values of $f_{\rm sub}$ is slightly higher than it is for higher values (above 10^{-3}). At low values of $M_{\rm hm}$ and high values of $f_{\rm sub}$, excessive perturbation often results in

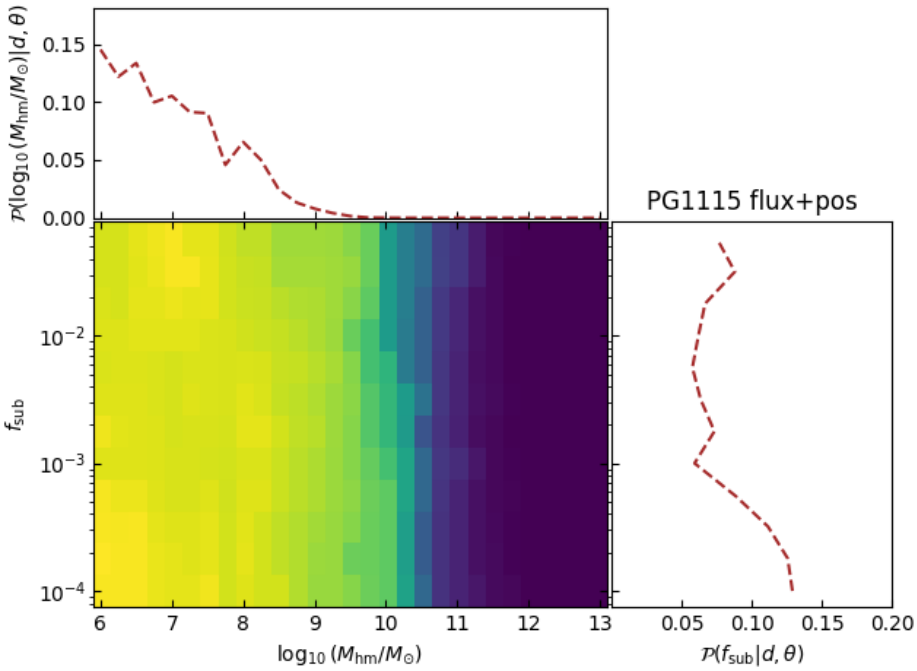


Figure 2.3 The grid on the bottom left shows the two-dimensional posterior probability distribution of $M_{\rm hm}$ and $f_{\rm sub}$ recovered from analysis of PG 1115+080. The likelihood for each grid point is calculated using both the image positions and fluxes. The top panel shows the result of marginalizing over $f_{\rm sub}$ to get the posterior of $M_{\rm hm}$, and the right panel shows the result of marginalizing over $M_{\rm hm}$ to get the posterior of $f_{\rm sub}$.

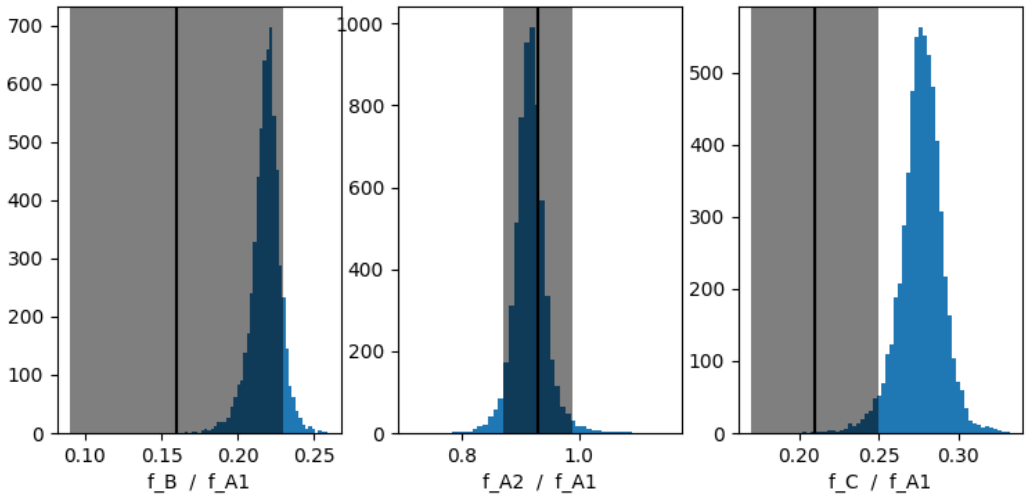


Figure 2.4 Flux ratio histograms for PG 1115+080 with $M_{\rm hm} = 10^6 M_\odot$ and $f_{\rm sub} = 10^{-2}$. The black lines and surrounding gray regions show the observed values and their corresponding measurement errors, and the blue histograms show the results from full model raytracing. 6642/8000 realizations produced 4 images and are included in this plot.

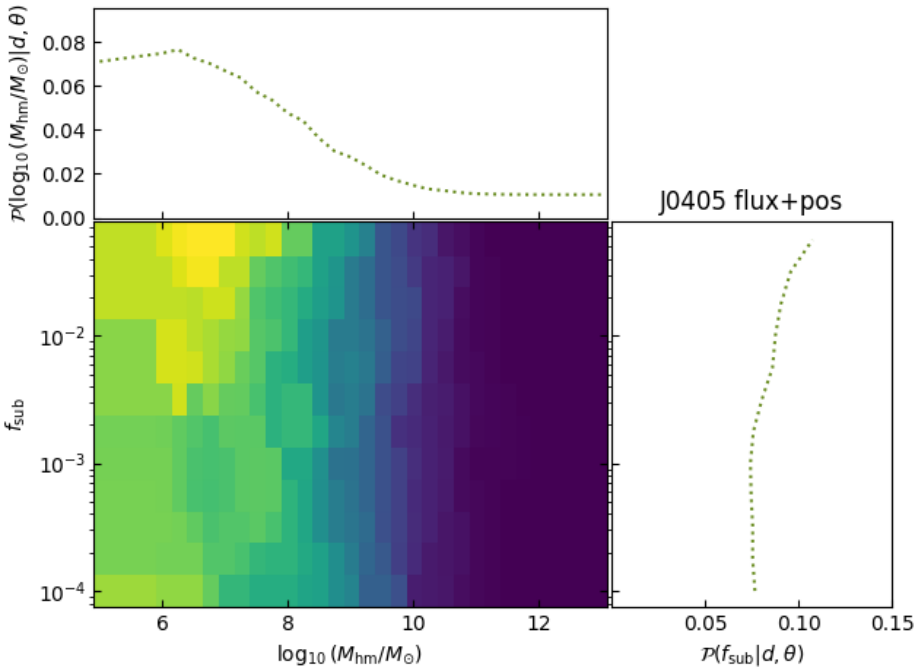


Figure 2.5 Same as Figure 2.3, except for WGD J0405-3308.

lens configurations that do not produce four images even after attempted macro-model parameter optimization. For $M_{\rm hm} = 10^6 M_\odot$, lower $f_{\rm sub}$ values (between 10^{-4} and 10^{-3}) result in more than 95% of realizations returning four images. Each successively higher $f_{\rm sub}$ value results in fewer four-image realizations, and by the maximum $f_{\rm sub} = 10^{-1.25}$, only about 40% of realizations produce four images. Since realizations that do not produce four images do not contribute to the likelihood for a grid point, higher $f_{\rm sub}$ values end up with slightly lower marginalized likelihoods.

2.4.2.2 WGD J0405-3308

As shown in Figure 2.5, WGD J0405-3308 favors colder models which have more perturbers. Like PG 1115+080, smooth models are unable to exactly fit the observed flux ratios of this lens. The degree to which this lens favors colder models is lesser than it is for PG 1115+080. The reason may be that there is only disagreement between macro-model predictions and one of the observed flux ratios for WGD J0405-3308 (as opposed to two for PG 1115+080), and a smaller percentage of realizations are able to improve the discrepant flux ratio. Higher values of $f_{\rm sub}$ are slightly favored. While cold, high-$f_{\rm sub}$ models result in some realizations returning non-four-image realizations, the frequency of this is much smaller than for PG 1115+080.

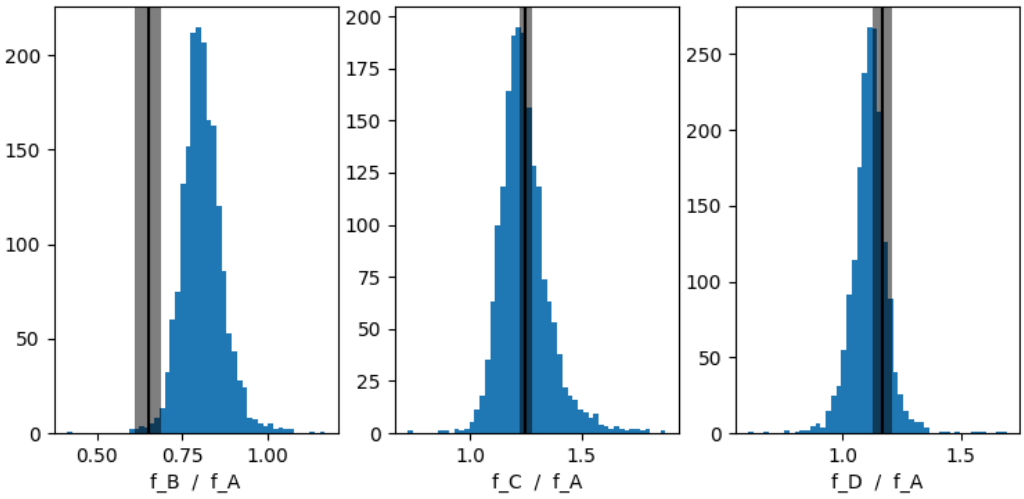


Figure 2.6 Same as Figure 2.1, except for WGD J0405-3308. 1890/2000 realizations produced 4 images and are included in this plot.

2.4.2.3 WFI 2033-4723

Like PG 1115+080 and WGD J0405-3308, WFI 2033-4723 has flux ratios that cannot be matched by smooth models alone. As shown in Figure 2.7, it favors increasingly colder models down to $M_{\rm hm} \approx 10^7 M_\odot$. Below this, however, the posterior probability distribution begins to drop again. To verify this trend and allow more complete comparison to the results for this lens from G20, we evaluate $M_{\rm hm} = 10^5 M_\odot$ grid points and interpolate. The likelihood does continue to drop off, though at a declining rate, and for high $f_{\rm sub}$, $M_{\rm hm} = 10^5 M_\odot$ grid points, almost no realizations return four images. This lens does not provide any constraint on $f_{\rm sub}$.

These findings do not agree with a similar analysis of the lens by G20. They found that the likelihood continues to rise monotonically as $M_{\rm hm}$ decreases. Additionally, their highest-likelihood realizations were able to exactly reproduce the observed flux ratios, whereas ours still have $< 1\sigma$ discrepancies. While the differences between our overall analysis procedures (see Section 2.2.9) may contribute this difference, G20 added an additional component to their macro-model to describe a nearby galaxy that we exclude. We suspect this added freedom in their model to contribute to the difference between our results for this lens.

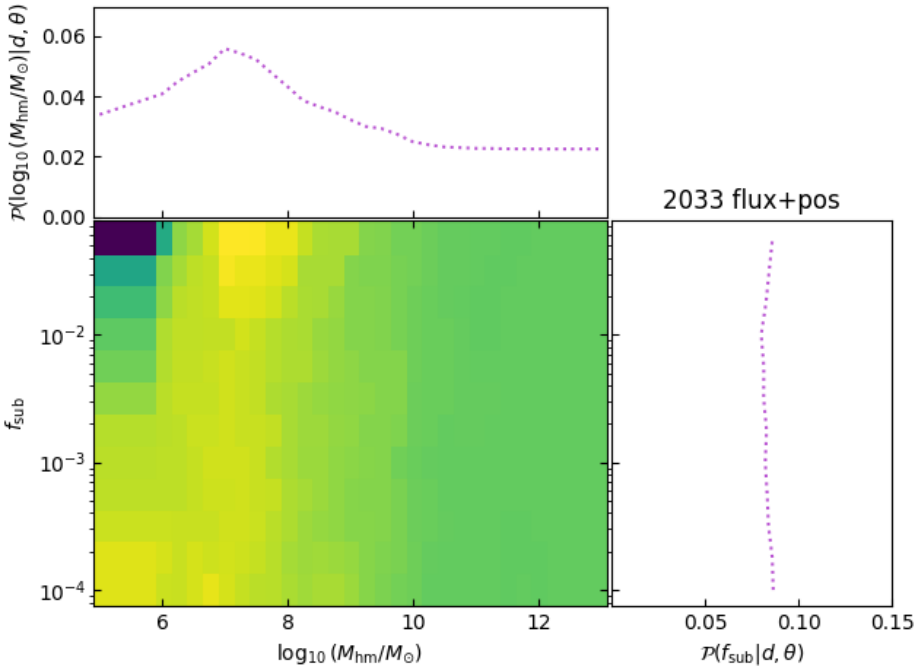


Figure 2.7 Same as Figure 2.3, except for WFI 2033-4723.

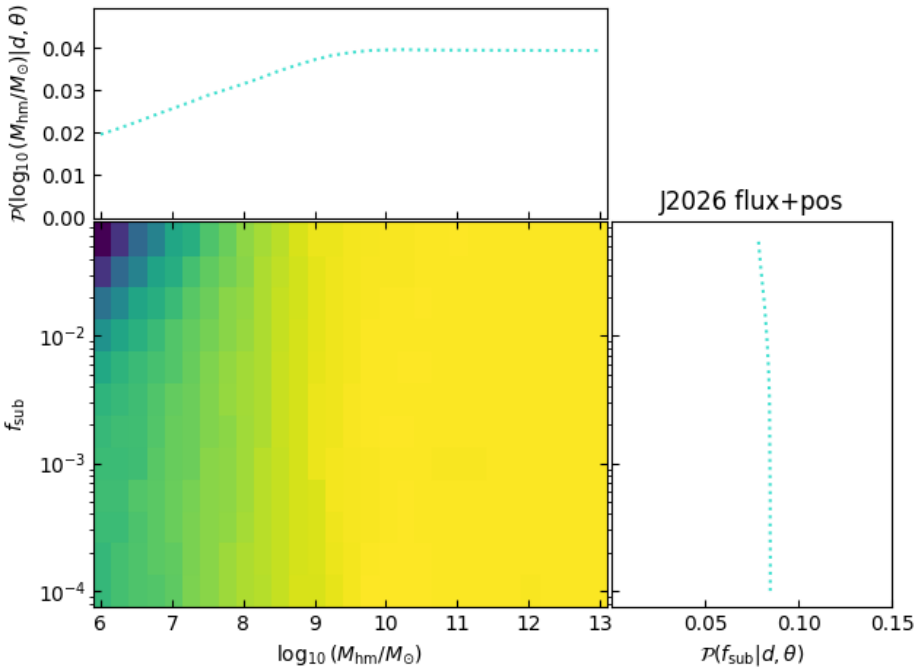


Figure 2.8 Same as Figure 2.3, except for WFI 2026-4536.

2.4.2.4 WFI 2026-4536

As shown in Figure 2.8, WFI 2026-4536 favors warmer models that have fewer perturbers. This lens system has image positions and fluxes that can be perfectly described by a smooth macro-model. Warmer models with few to no perturbers can thus produce realizations that fit very well. Since there is no room for improvement of image positions and fluxes, colder models with more perturbers can only fit as well or worse. Many perturber arrangements leave images relatively unaffected. In rare cases, perturber arrangements may improve macro-models that do not fit well on their own. However, a much more common scenario is for perturbations to worsen the fit. As a result, models with lower $M_{\rm hm}$ values have lower marginalized likelihoods for this lens. This lens system provides almost no information about $f_{\rm sub}$ since the marginalized $f_{\rm sub}$ likelihoods are dominated by warm realizations with few to no subhalos.

Note that this lens favors higher $M_{\rm hm}$ values much less strongly than PG 1115+080 favors lower $M_{\rm hm}$ values. In general, lens systems with 'anomalous' flux ratios that can be resolved with perturbers (such as PG 1115+080) favor colder models more strongly than lens systems with flux ratios that can be reproduced with smooth models (such as WFI 2026-4536) favor warmer models. The former is a 'detection' of perturbation by low-mass dark matter halos that provides strong evidence against a model with significant mass function suppression at that mass scale. The latter is a 'non-detection' and requires many others like it to provide similarly strong evidence for a model with significant mass function suppression at that mass scale.

2.4.3 Inference on Thermal Relic WDM Particle Mass

In order to obtain an an inference on $M_{\rm hm}$ from the 2D likelihoods shown in Figure 2.9, we marginalize over $f_{\rm sub}$. Figure 2.10 presents the normalized posterior probability distributions that result. Table 2.2 shows the resulting 7:1, 20:1 and 30:1 likelihood ratio as well as 95% confidence level upper limits on the half-mode mass, $M_{\rm hm}$, and corresponding lower limits on particle mass, $m_{\rm th}$, assuming thermal relic WDM. We choose these constraint criteria in order to allow for comparison to constraints published in previous work. We report the values derived from the joint sample, the NLR and Radio/MIR subsets, and a subset that matches the lens sample used by G20, which we refer to as NLR+B1422 or G20-like.

Though it is a commonly used statistic in the literature, the 95% (2σ) bound in this case is

highly dependent on the choice of a prior on $M_{\rm hm}$. H20 use $-6.0 < \log_{10}(M_{\rm hm}/M_\odot) < 9.0$, G20 use $4.8 < \log_{10}(M_{\rm hm}/M_\odot) < 10.0$, and in this work we use $6.0 < \log_{10}(M_{\rm hm}/M_\odot) < 13.0$. Models with $\log_{10}(M_{\rm hm}/M_\odot) < 6.0$ produce similar populations of perturbers above our detection threshold. Thus, we do not expect to find a significant difference between the likelihood inferred for CDM and a WDM model with $\log_{10}(M_{\rm hm}/M_\odot) < 6.0$. Thus, we expect the posterior probabilities for all models below this point to be roughly equal. Lowering the extent of the prior on $M_{\rm hm}$ can then significantly change the constraining power of the 95% (2σ) statistic. The lenses which we have explored down to $\log_{10}(M_{\rm hm}/M_\odot) = 5.0$ indicate that we are approaching our detection threshold, but we will check this in more detail in future work. Likelihood ratio constraints avoid these concerns related to the choice of prior and are only dependent on the maximum of the distribution. If the likelihood function flattens below our lower bound on $M_{\rm hm}$, extending the prior lower will not change their strength. The 7:1 likelihood ratio roughly corresponds to the ratio between the maximum height of a Gaussian distribution and its height at 2σ, and we mainly rely on this statistic in the following discussion for its robustness.

2.4.3.1 Joint Sample

As expected, our strongest constraints come from the joint sample of all 14 lenses studied in this work. Using only the fluxes from realizations with image positions that fit within 3σ to calculate the likelihood, we constrain thermal relic WDM to models with $\log_{10}(M_{\rm hm}/M_\odot) < 7.47$ ($m_{\rm th} > 6.63$keV) with a 7:1 likelihood ratio. This is the strongest constraint on WDM to be set from a gravitational lensing-only analysis to date. This inference weakens to $\log_{10}(M_{\rm hm}/M_\odot) < 7.63$ ($m_{\rm th} > 5.93$keV) if we instead use both image fluxes and positions to calculate the likelihood. In either case, the inference is most strongly driven by two lenses from our Radio/MIR sample, PG 1115+080 and JVAS B1422+231, which have flux ratios that cannot be matched with a smooth macro-model and can be matched with the addition of perturbers (see figures 2.9 and 2.10).

2.4.3.2 NLR

Our NLR sample, which includes SDSS J1330+1810 and excludes JVAS B1422+231, provides relatively weak constraints on its own of $\log_{10}(M_{\rm hm}/M_\odot) < 9.15$ ($m_{\rm th} > 2.08$keV) with a 7:1 likelihood ratio using fluxes and positions. The lens that contributes most strongly to this inference is WGD J0405-3308. Though they favor warmer models less strongly than other lenses favor colder models, HE0435-1223 and WFI 2026-4536, the lenses which most strongly favor warmer models,

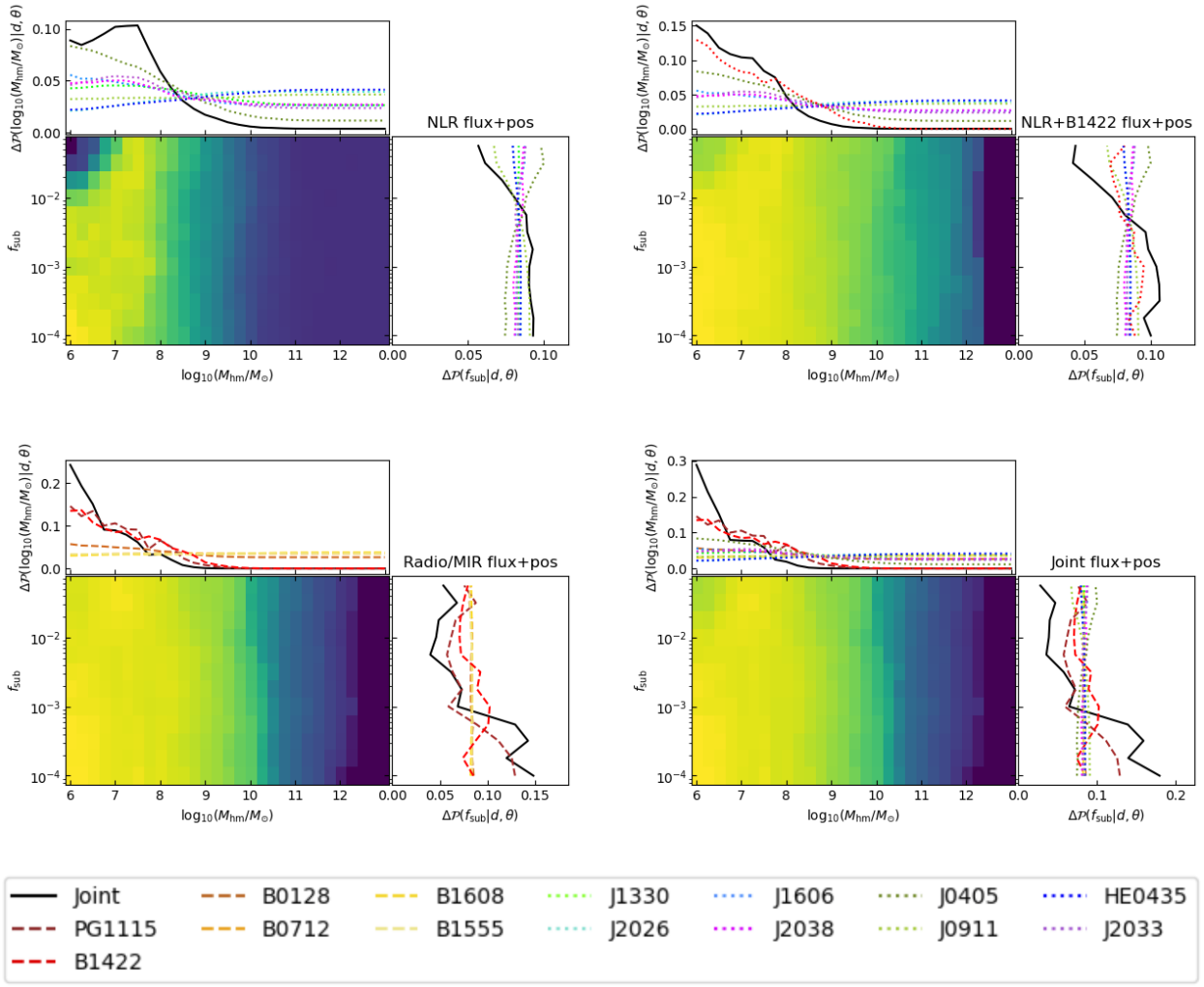


Figure 2.9 Plots similar to Figure 2.3, except for combined NLR (top left), NLR+B1422 (top right), Radio/MIR (bottom left) and full Joint (bottom right) samples. Distributions from individual lenses are shown in dashed (NLR) and dotted (Radio/MIR) colorful lines, and joint distributions for each sample are shown in solid black.

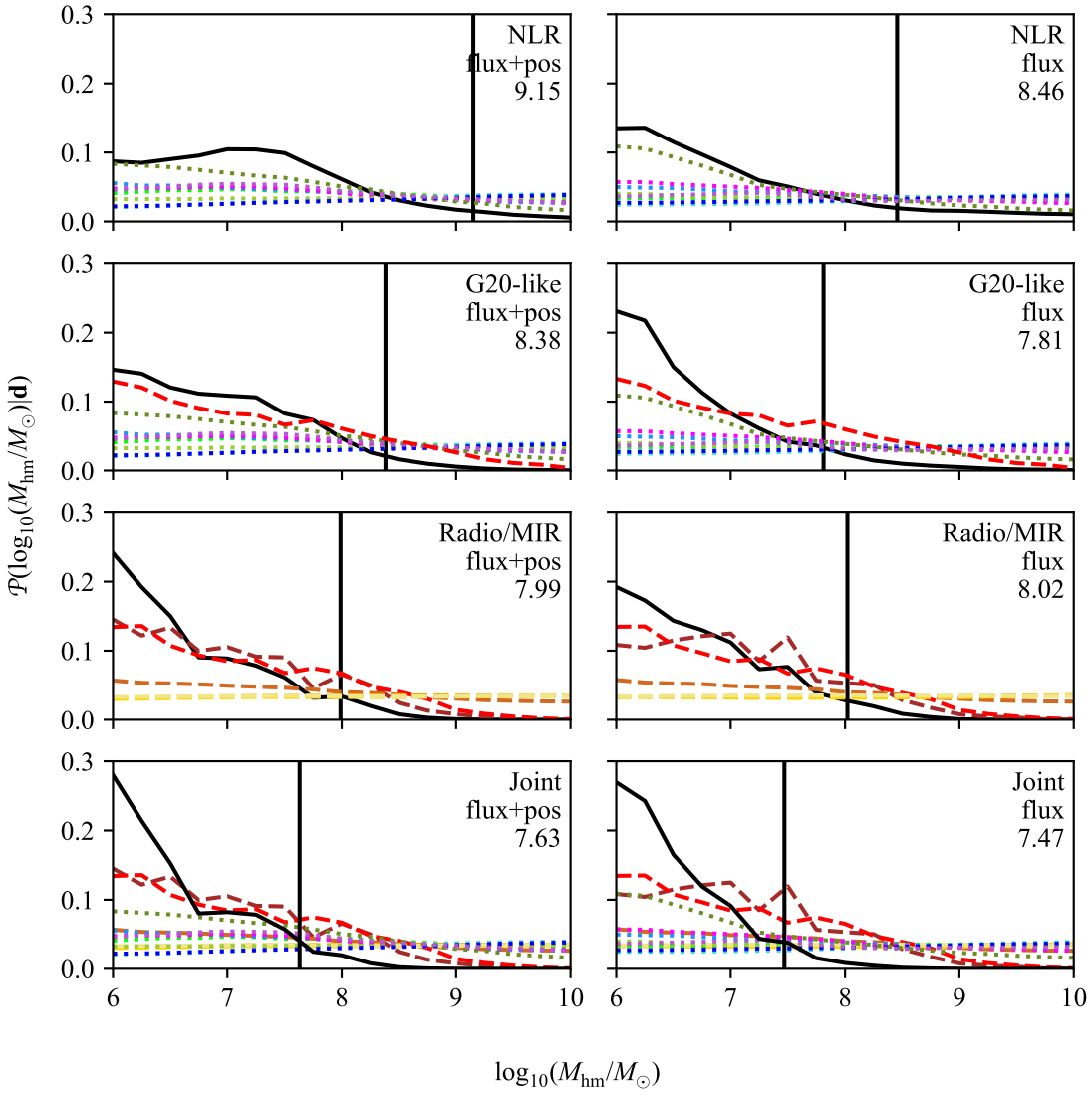


Figure 2.10 Normalized posterior probability distributions of $\log_{10}(M_{\rm hm})$ from the (top to bottom) NLR, NLR+B1422, Radio/MIR and joint samples. Distributions from individual lenses are shown in dashed (NLR) and dotted (Radio/MIR) lines, and joint distributions for each sample are shown in solid black. Colors for individual lenses are the same as those described in the legend of Figure 2.9. The left hand column is calculated using both the image positions and fluxes, and the right hand column is calculated using only the fluxes from realizations with all image positions within 3σ of observed positions. Vertical black lines show joint 7:1 t likelihood ratios relative to the peaks of the joint distributions.

Table 2.2 Inferences on the upper limit of $M_{\rm hm}$ from all sample sets and various levels of constraint. Shown are upper limits on $\log_{10}(M_{\rm hm}/M_\odot)$ with corresponding lower limits on $m_{\rm th}$/keV shown in parentheses (see equations 2.5 and 2.6). We calculate 7:1, 20:1 and 30:1 likelihood ratios with respect to the maximum likelihoods. The last column shows the 95% confidence level limits. Note that we were unable to place 20:1 and 30:1 likelihood ratio constraints with the flux-only method for the NLR sample that excludes B1422.

	$\log_{10}(M_{\rm hm}/M_\odot)$ Upper Bound				$m_{\rm th}$/keV Lower Bound			
Sample	7:1	20:1	30:1	95%	7:1	20:1	30:1	95%
NLR								
flux+pos	9.15	9.88	10.25	9.93	2.08	1.25	0.97	1.21
flux	8.46	-	-	12.0	3.35	-	-	0.29
NLR+B1422*								
flux+pos	8.38	8.88	8.98	8.27	3.53	2.50	2.34	3.81
flux	7.81	8.40	8.47	8.13	5.23	3.49	3.31	4.20
Radio/MIR								
flux+pos	7.99	8.48	8.62	8.01	4.63	3.29	2.99	4.56
flux	8.02	8.63	8.70	8.00	4.53	2.98	2.84	4.59
Joint								
flux+pos	7.63	8.49	8.31	7.80	5.93	3.28	3.72	5.28
flux	7.47	7.91	7.96	7.56	6.63	4.88	4.72	6.23

* In order to replicate the sample used by G20, the NLR+B1422 includes lens system B1422+231 with larger measurement uncertainties on image positions and fluxes that correspond to the NLR measurements of this system and excludes SDSS J1330+1810.

are also included in this set.

2.4.3.3 NLR+B1422

This group of lenses, which is the same as the NLR sample with the inclusion of JVAS B1422+231 and exclusion of SDSS J1330+1810, exactly matches the set used by G20. From it, we draw constraints of $\log_{10}(M_{\rm hm}/M_\odot) < 7.81$ ($m_{\rm th} > 5.23$keV) with a 7:1 likelihood ratio using fluxes only and $\log_{10}(M_{\rm hm}/M_\odot) < 8.38$ ($m_{\rm th} > 3.53$keV) using fluxes and positions. These 7:1 flux-only constraints are in excellent agreement those made by G20 at 2σ, $\log_{10}(M_{\rm hm}/M_\odot) < 7.8$ ($m_{\rm th} > 5.20$keV). Our 2σ limits are significantly weaker than theirs, which is expected due to their lower-extending prior. The 7:1 and 30:1 likelihood ratio constraints that they report, however, fall between our flux-only and flux and position values.

2.4.3.4 Radio/MIR

Our Radio/MIR sample contains the two most constraining (favoring colder models) lenses, JVAS B1422+231 and PG 1115+080. From it, we draw constraints of $\log_{10}(M_{\rm hm}/M_\odot) < 8.02$ ($m_{\rm th} > 4.53$keV) with a 7:1 likelihood ratio using fluxes only and $\log_{10}(M_{\rm hm}/M_\odot) < 7.99$ ($m_{\rm th} > 4.63$keV) using fluxes and positions. The difference between the two likelihood calculation methods is much smaller than it is for any of the other sample sets. This is weaker than the $\log_{10}(M_{\rm hm}/M_\odot) < 7.80$ ($m_{\rm th} > 5.58$keV) constraints set by H20, but their constraint includes two additional lens systems, MG J0414+0534 and CLASS B2045+265. Due to uncertainties surrounding the analyses of these two systems, they also report constraints excluding them of $\log_{10}(M_{\rm hm}/M_\odot) < 8.03$ ($m_{\rm th} > 4.77$keV at the 95% confidence level that align well with our 7:1 constraints. Our Radio/MIR sample includes CLASS B1608+656 whereas theirs does not, however this lens lacks an anomalous flux ratio (unlike MG J0414+0534 and CLASS B2045+265) and only works to weaken our bounds on WDM.

2.4.4 Inference on the CDM Subhalo Mass Function

Figure 2.11 shows the posterior probability distribution that we recover for $f_{\rm sub}$ using the joint 14 lens sample, and also the result from excluding PG 1115+080. This value determines the normalization of the subhalo mass function (see Section 2.2.2) assuming CDM. PG 1115+080 has a much stronger influence than any of the other lens systems we include, and it drives the joint inference to favor $f_{\rm sub}$ values below 0.001. This is in disagreement with existing constraints that all place $f_{\rm sub}$ above 0.001 (Dalal & Kochanek 2002; Xu et al. 2015; H20; G20). When we remove PG 1115+080 and recalculate, this preference weakens significantly such that $f_{\rm sub}$ is unconstrained. As expected, there is degeneracy between $f_{\rm sub}$ and $M_{\rm hm}$ as both alter the amount of perturbation expected. The top-left panel of Figure 2.9 illustrates this most clearly.

2.5 Discussion

2.5.1 Comparison to Predictions

Gilman et al. (2019) used a sample of 50 mock lensed quasars in order to predict the expected constraining power from increasingly large data sets. They generated subhalos and LOS halos using a CDM mass function with various SHMF normalizations and multiple WDM mass functions with

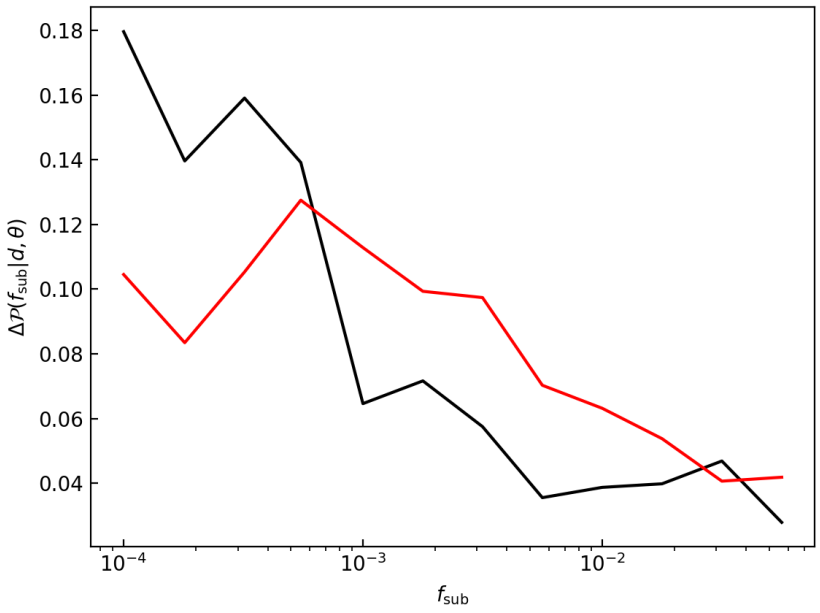


Figure 2.11 Posterior probability distributions of the normalization of the SHMF, $f_{\rm sub}$, for the joint sample of 14 lenses (black) and all lenses except PG 1115+080 (red), which dominates the inference otherwise. For CDM, this represents the true substructure fraction, the total mass in substructure divided by the mass of the host halo. For WDM models, the suppression of the low-mass end of the SHMF is implemented after the normalization. As a result, there are less subhalos than in colder models with the same $f_{\rm sub}$ values and the true fraction of mass in substructure is lower than $f_{\rm sub}$.

the same SMHF normalization. Assuming CDM with $f_{\rm sub} \approx 0.015$ and 6% flux uncertainties, they expected a sample of 10 lensed quasars to be able to set a 2σ upper bound of $\log_{10}(M_{\rm hm}/M_\odot) < 8.6$. However, using the same analysis method on a sample of 8 real lens systems with similar uncertainties in a follow up paper, G20 were able to set an upper bound of $\log_{10}(M_{\rm hm}/M_\odot) < 7.8$. They note that this discrepancy is likely due to the higher mean value of $f_{\rm sub} = 0.035$ recovered from the real data. The 7:1 likelihood ratio constraints we recover from the same sample of 8 lens systems using only fluxes to calculate the likelihood, $\log_{10}(M_{\rm hm}/M_\odot) < 7.81$, are likewise consistent with their predictions based on a higher $f_{\rm sub}$ value that is in agreement with our findings. Using fluxes and positions, our constraint of $\log_{10}(M_{\rm hm}/M_\odot) < 8.38$ agrees with their predictions for a smaller $f_{\rm sub}$ value.

Though Gilman et al. (2019) do not make predictions for samples with fewer than 10 lens systems, they predict that doubling the sample size from 10 to 20 lens systems would lower the

upper bound on $\log_{10}(M_{\rm hm}/M_\odot)$ by roughly 0.4 dex or more, depending on flux uncertainty and SHMF normalization. Our sample of 14 lensed quasars is double the size of that used by H20 (7) and nearly double the size of that used by G20 (8). Depending on the subsample comparison, likelihood calculation method and the constraint statistic, we find improvement of between 0.3 and 1.0 dex, which is slightly larger than their expectations. However, considering the smaller SHMF normalizations used for their predictions, our results are broadly consistent with them. Additionally, Gilman et al. (2019) created their mock lens sample with equal numbers of cross, fold and cusp configurations. Gilman et al. (2018) found that fold and cusp configurations tend to produce stronger constraints on WDM, and our sample consists of mostly folds and cusps. This may also contribute to a greater-than-expected improvement from the increased sample size.

2.5.2 Systematic Uncertainties

Though we have introduced multiple methods to reduce systematic uncertainty in our analysis, we have yet to closely investigate their efficacy and effects on our final inferences. Additionally, there are effects that we assume to be second-order that may be modeled to reduce systematic uncertainty in future work.

2.5.2.1 Macro Models

The foundational concept of our analysis is that we can constrain dark matter properties by trying to reproduce observed flux ratios by adding perturbing dark matter halos as substructure to a macro-model and along the line of sight. Dark matter models that produce populations of halos that are more likely to reproduce observed flux ratios are thus favored. Thus, to test the viability of a perturber population we need to ensure that the underlying macro-model is correct. Many of the lens systems in our sample have lensed arcs from extended emission in addition to the discreet images used for flux ratios. Modelling these arcs provides much tighter constraints on the macro-model than can be achieved with only image positions and fluxes. In future work, we plan to incorporate the constraints from modelling arcs into our flux-ratio analysis procedure to greatly reduce this source of uncertainty.

Optimizing the macro-model parameters after the addition of perturbers allows us to greatly improve the pairing of perturber populations and macro-models. However, especially for realizations with the most perturbers (low $M_{\rm hm}$, high $f_{\rm sub}$), this optimization step is often unsuccessful when the

initial macro-model/perturber combination does not produce four images. The initial macro-model parameters are drawn from a distribution created by fitting only to the image positions and with only the smooth model. We have shown that the addition of perturbers can have a significant effect on the image positions of some realizations, and it is possible that macro-model parameter values that are outside our sampled distribution may combine with these perturber populations to produce well-fitting, four-image realizations. Assuming CDM, this may be systematically reducing the posterior probability of high-perturber realizations, making our constraints on WDM weaker.

There is also systematic uncertainty introduced in lens systems with stellar disks included in the macro-model, CLASS B0712+472, CLASS B1555+375 and SDSS J1330+1810. As shown in Hsueh et al. (2018), stellar structures can increase flux ratio anomalies by up to 10%. From the imaging available, we have weak constraints on the elliptical disk parameters including the centroid positions and position angles. The model parameters we report and with which we initialize our MCMC analysis are the best-fit values from initial smooth modeling. For these lenses, we fit to images positions and fluxes with the assumption that the disk is contributing to the flux ratio anomalies. We found that different disk models that fit the image positions equally well can result in different flux ratios and thus different final inferences. If there are undetected stellar structures in other lens systems, this uncertainty may be even more pervasive. In future work, we intend to quantify the degree to which reasonable modification of the disk parameters can affect a lens system's constribution to an inference on $M_{\rm hm}$.

2.5.2.2 Source Size

In future work, we can improve our analysis by modelling sources as extended Gaussian brightness distributions, as in G20. However, we do not currently have the observational ability to determine the approximate source size for individual lenses. Instead, source sizes must be randomly sampled. Future work may investigate the affect on the inference on $M_{\rm hm}$ from changing the source size and predict expected improvement from measuring the source structure for individual lenses.

2.5.2.3 Dark Matter Halo Population

In modelling the population of dark matter halos as substructure and along the line of sight, we make a number of choices with the intention of balancing the inclusion of up-to-date findings and computational efficiency. In future work, we intend to explore potential biases and uncertainties resulting from these choices. For both subhalos and LOS halos, we adopt stochastic mass-

concentration relations based on simulation results (see Section 2.2.2). As the lensing signal of a halo is highly dependent on its central density, changes in the mass-concentration relation may have large impacts on our final inference. Assuming CDM, producing halos with higher concentrations would increase the lensing signal would weaken our ability to constrain WDM.

Our parameterizations of the HMFs also introduce systematic uncertainty. For both the subhalo and LOS mass functions, we use the the same suppression factor for WDM models in relation to CDM (see equation 2.10). Recent work based on simulations suggests that a slightly more complex parameterization that differs between subhalos and LOS halos may be more appropriate (Lovell, 2020). Including this distinction may help reduce the degeneracy between $f_{\rm sub}$ and $M_{\rm hm}$. We expect the effects of these choices to be of secondary importance, but future exploration of their relative effects on our final inference on $M_{\rm hm}$ will help inform best practices for future analyses.

2.6 Conclusions

Using a sample of 14 four-image lensed quasars, we place the strongest constraints to date on the properties of thermal relic dark matter using gravitational lensing. We place an upper bound of $\log_{10}(M_{\rm hm}/M_{\odot}) < 7.47$ ($m_{\rm th} > 6.63$keV) with a 7:1 likelihood ratio. Our findings for subsets of this sample analyzed in previous work are consistent with their results (H20; G20). We find that our constraints are most strongly driven by two lens systems with anomalous measured flux ratios, JVAS B1422+231 and PG 1115+080. To further strengthen constrains, it is crucial to increase the sample size. Upcoming observations from the James Webb Space Telescope (JWST) should increase the sample size of gravitationally lensed quasars by more than an order of magnitude. This should enable a joint analysis to reach the theoretical potential of flux ratios to constrain WDM. Thus, it is crucial to understand the systematics involved in this method. This will be the subject of our future work.

Acknowledgements

This work used the following software packages: AstroPy (The Astropy Collaboration et al., 2018), Matplotlib (Hunter, 2007), NumPy (van der Walt et al., 2011), GRAVLENS (Keeton, 2011), pylens (provided by M. Auger) and emcee (Foreman-Mackey et al., 2013).

Chapter 3

Forgotten multipoles and their implication for dark matter inference

This chapter has been submitted to *MNRAS* as of March, 2024 and is currently under review.

The flux ratios of strongly lensed quasars have previously been used to infer the properties of dark matter. In these analyses it is crucial to separate the effect of the main lensing galaxy and the low-mass dark matter halo population. In this work, we investigate flux-ratio perturbations resulting from general third- and fourth-order multipole perturbations to the main lensing galaxy's mass profile. We simulate four lens systems, each with a different lensing configuration, without multipoles. The simulated flux ratios are perturbed by 10-40 per cent by a population of low-mass haloes consistent with CDM and, in one case, also a satellite galaxy. This level of perturbation is comparable to the magnitude of flux-ratio anomalies in real data that has been previously analyzed. We then attempt to fit the simulated systems using multipoles instead of low-mass haloes. We find that multipoles with amplitudes of 0.01 or less can produce flux-ratio perturbations in excess of 40 per cent. In all cases, third- or fourth-order multipoles can individually reduce the magnitude of, if not eliminate, flux-ratio anomalies. When both multipole orders are jointly included, all simulated flux ratios can be fit to within the observational uncertainty. Our results indicate that low-mass haloes and multipoles are highly degenerate when modelling quadruply-imaged quasars based just on image positions and flux ratios. In the presence of this degeneracy, flux-ratio anomalies in lensed quasars alone cannot be used to place strong constraints on the properties of dark matter without additional information that can inform our priors.

3.1 Introduction

The cold dark matter (CDM) paradigm, which posits that dark matter consists of non-relativistic, collisionless particles (e.g. Planck Collaboration et al., 2016), is successful at describing cosmic structure at scales larger than $\sim$ 1 Mpc (Springel et al., 2005; Planck Collaboration et al., 2020) and has been adopted as the standard in cosmology. CDM predicts a clumpy distribution of dark matter on sub-galactic scales and the existence of a large population of low-mass haloes (e.g., Vogelsberger et al., 2014; Schaye et al., 2015). On the other hand, warm dark matter models (WDM) predict a smaller amount of such objects with a less concentrated mass density profile (e.g. Bode et al., 2001; Viel et al., 2005; Lovell et al., 2014; Lovell, 2020). The difference between CDM and still-viable WDM models is strongest at halo masses lower than $10^9 M_\odot$ (Hsueh et al., 2020), where most of these objects are predicted to be faint or even completely dark. Strong gravitational lensing allows us to detect them via their gravitational effect on the strongly lensed images.

In this paper, we focus on galaxy-scale strong lensing of unresolved sources, specifically quadruply-imaged quasars (quads). The image configuration is determined by the mass distribution of the lens and the position of the source. Low-mass haloes associated with the lens galaxy, called subhaloes, and haloes along the line of sight, called field haloes, can produce measurable changes in the relative fluxes of the lensed images due to the dependence of the image magnifications on the second derivative of the lensing potential.

This method of investigating dark matter, known as flux-ratio analysis, cannot be used to precisely determine the masses and positions of individual haloes that cause the perturbations. Rather, the viability of a dark matter model is assessed based on the probability that its halo population could have produced the observed flux ratios, marginalised over the possible individual halo configurations. Dark matter models that produce larger numbers of low-mass haloes will lead to a higher incidence of lens systems that show so-called flux-ratio anomalies, in which a standard unperturbed smooth mass distribution cannot reproduce the observed flux ratios. In contrast, dark matter models that suppress the formation of low-mass haloes will lead to fewer and less significant flux-ratio anomalies in samples of lensed quasar systems. Single lens systems provide only a weak inference on dark matter models, since even in the presence of a large number of associated low-mass perturbers, these may be spatially distributed in a way such that no flux-ratio anomaly is produced. Thus, large samples of lens systems are needed.

Flux-ratio analysis was first proposed by Mao & Schneider (1998) and Metcalf & Madau (2001), and it was originally limited to subhaloes. Soon after, Dalal & Kochanek (2002) applied it to a sample of seven quads and reported results that were consistent with CDM simulations at 90 per cent confidence.

Follow-up studies argued for the inclusion of additional components that could also influence flux ratios, such as field haloes and stellar disks (Möller et al., 2003; Inoue & Takahashi, 2012; Metcalf, 2005; Despali et al., 2018). The contribution from field haloes is especially important given that they are often more numerous than substructures (Despali et al., 2018). Furthermore, field haloes should provide a cleaner test of dark matter models because their properties are not influenced by the tidal effects that can be so important for subhaloes. Baryonic components such as stellar disks have been discovered in real lens systems and, when included in the lens model, have successfully been able to reproduce the flux-ratio anomalies in those systems without having to resort to low-mass dark matter haloes (Hsueh et al., 2016, 2017). Similarly, more general explorations have shown that baryonic structures in lensing galaxies can mimic perturbations by low-mass haloes if not properly accounted for in the lens model (Gilman et al., 2017, 2018; Hsueh et al., 2018). Sensitive high-resolution imaging can be used to estimate the contribution of baryonic structure in lensing galaxies, but other complexities may remain as confounding factors.

In this paper, we will investigate an important form of additional complexity for lens mass models, namely, the angular structure in the lensing galaxy, parameterised here as multipole perturbations. The most common mass profile used to model lens galaxies is the elliptical power law (EPL) with external shear (e.g., Tessore & Metcalf, 2015). We will hereafter refer to this type of base model as the EPLγ model. The multipoles that we consider add Fourier-type perturbations to the angular part of the density profile, leaving the radial part unchanged. The use of multipoles is motivated by optical and infrared observations of elliptical galaxies, which show that the isophotes of many of them deviate from perfect ellipticity (e.g., Bender et al., 1988, 1989; Cappellari, 2016). These deviations can be modeled by simply-parameterized multipole components. While many treatments of elliptical galaxy isophotes focus only on fourth-order multipoles, Hao et al. (2006) present an extensive investigation of the surface brightness distribution in elliptical galaxies in which they fit both third- and fourth-order multipoles with a variety of orientation angles.

It is thus natural to consider multipole components in the mass distributions of galaxies as

well. Demonstrating the impact of angular complexity in the lens galaxy on flux ratios, Evans & Witt (2003) and Congdon & Keeton (2005) showed that the joint inclusion of third- and fourth-order multipoles with unrestricted orientation angles can reproduce many anomalies that had been observed at the time. We focus specifically on third- and fourth-order multipoles because they are expected to cause flux-ratio effects degenerate with those of perturbing haloes. Lower-order multipoles have effects that are analogous to changes in the macro-model parameters, and higher-order multipoles may introduce deviations from ellipticity that either produce greater than four images or are unphysical (Evans & Witt, 2003; Congdon & Keeton, 2005). Despite these findings, flux-ratio analyses since then have neglected to implement them completely. Hsueh et al. (2020) study some of the same lenses as Evans & Witt (2003) and Congdon & Keeton (2005), including angular structure in one lens in the form of an exponential disk, but they modelled all other lenses with only an EPL and shear. Recent flux-ratio analyses by Gilman et al. (2021, 2022, 2023) have included multipoles but in a specific and restrictive way. In those analyses, the orientation angle of the fourth order multipole is fixed to align with the EPL. Third order multipoles are not included. Gilman et al. (2024) inferred constraints on WDM from simulated lens systems including both third- and fourth-order multipoles, but the fourth order was fixed to align with the EPL. The effect of multipoles has also been considered in the context of the extended emission of lensed galaxies, where O'Riordan & Vegetti (2023) found that the inclusion of third- and fourth-order multipoles with unrestricted orientation angles could produce false substructure detections.

In this paper, we extend earlier work on lensed quasar flux-ratios (Evans & Witt, 2003; Congdon & Keeton, 2005) in several important ways. First, while those papers modeled real lenses with multipole components, our investigation uses simulated lenses so that we can directly compare the perturbative effects of low-mass haloes with those of multipoles. In addition, we consider the effects of third- and fourth-order multipoles separately as well as jointly, and have more generality in our base models by allowing the power-law index to be different from the isothermal value. Our particular focus is an investigation of the potential for general third- and fourth-order multipoles to perturb the flux ratios of quadruply-imaged quasars in a way that is degenerate with perturbations from low-mass dark matter haloes. In Section 3.2, we describe our procedure for obtaining EPLγ base models for a sample of real lens systems. In Section 3.3, we detail the creation of four simulated lens systems from the combination of EPLγ base models and CDM low-mass halo

populations plus, for one of the systems, a satellite galaxy. Section 3.4 describes how we model the simulated lenses in our sample using just an EPLγ model plus multipole components, and Section 3.5 presents the results. We discuss the implications of our results and future work in Section 3.6.

3.2 Modelling real data: EPLγ

To quantify the effect of multipoles in realistic scenarios, we create simulated strong gravitational lens systems using image configurations from real lens systems taken from recent flux-ratio analysis studies (Hsueh et al., 2020; Gilman et al., 2020). To ensure applicability of our results across different image configurations, we select lens systems that fall into one of each of the general categories (see Figure 3.1 for visualizations): cross (WGD J0405-3308), fold (WFI 2026-4536) and cusp (B1422+231). We also select the PS J1606-2333 system, which has a luminous satellite associated with the main lensing galaxy, as the basis for a fourth simulated strong lens system. The satellite will allow us to investigate the degeneracy between multipoles and haloes beyond the low-mass range we otherwise consider (see Section 3.3.2).

For the mass model, we use an EPLγ. The corresponding dimensionless surface mass density (convergence) is given by

$$\kappa(R) = \frac{2-t}{2}\left(\frac{b}{R}\right)^t \quad , \tag{3.1}$$

where R is an ellitpical radius such that $R^2 = (qx)^2 + y^2$. The model parameters are the power-law slope, t, axis ratio, q, and scale length, $b = R_E\sqrt{q}$, where R_E is the Einstein radius. We define external shear with amplitude, $\gamma_{\rm ext}$, and orientation angle, $\phi_{\rm ext}$. The satellite in J1606 is modelled as a singular isothermal sphere (SIS), an EPL profile with $t = 1$ and $q = 1$.

Fitting only the observed image positions, we perform Markov chain Monte Carlo (MCMC) sampling to approximate posterior distributions of the mass model parameters for each lens system. The location of the luminous satellite in J1606 is fixed to the observed position, but we allow all other parameters, including the source position, to vary freely. For convenience, we will hereafter refer to the distributions generated in this step, including the system with the satellite, as EPLγ distributions. These sets of parameters describe the base models that will be perturbed either by low-mass haloes (EPLγ+CDM; Section 3.3) or by multipoles (EPLγ+MP; Section 3.4).

3.3 Simulated data: EPLγ+CDM

We create our four simulated lens systems by adding populations of CDM subhaloes and field haloes to base models drawn from the EPLγ distributions described in Section 3.2. The results are mock quads with image positions that match those of a real lens system and flux ratios that are perturbed only by low-mass haloes. These lens systems do not contain any multipole components.

3.3.1 Background source

Typically, flux-ratio investigations focus on emission from regions of the background objects that are large enough to avoid being affected by microlensing by stars in the primary lensing galaxy. These include mid-infrared emission from dust surrounding quasar accretion disks, which are typically smaller than 10 pc (Burtscher et al., 2013); emission from the narrow-line regions surrounding a quasar, which can extend up to 60 pc (Müller-Sánchez et al., 2011; Nierenberg et al., 2017); or radio emitting regions, for which individual observations give estimates of sizes smaller than 10 pc (Lee et al., 2017; Kim et al., 2022). Generally, as the size of the background source increases, it becomes less susceptible to flux perturbations from low-mass haloes (Dobler & Keeton, 2006). In this paper, we want to quantify the degeneracy between low-mass haloes and multipoles in the scenario in which the effect of the former is maximal, hence the background sources in our mock observations and models are point-like. Their location in each realization is drawn from the MCMC chains associated with the modelling of the real data. In Section 3.6, we further investigate the effect of the source size and its implication for the degeneracy under study.

3.3.2 Low-mass halo population

To generate the CDM halo populations that we add to the EPLγ models, we largely follow the process described in Hsueh et al. (2020) with updated treatments of the mass-concentration relations for subhaloes and field haloes. All low-mass haloes are modelled as NFW profiles (**?**; however, see Heinze et al., 2024).

For the field halo mass-concentration relation, we use that reported in Table 1 of Duffy et al. (2008) from N-body simulations. We use values derived using the virial radius definition of relaxed haloes between redshifts 0 and 2. Unlike Hsueh et al. (2020), we apply the associated scatter on the parameters. We follow the implementation of Despali et al. (2016), which is based on the approach introduced by Sheth & Tormen (1999), for the field halo mass function. We use their best-fitting

Table 3.1 Unitless constants used in equations 3.2 and 3.3.

$R^{\rm CDM}_{\rm max,mean}$	
a	0.24986592
b	1.55822031
c	-0.01885084
d	0.38482671
σ	
p	1.28099
m	0.21388
n	0.46263
k	0.01501

parameters optimized over all considered redshifts and cosmologies.

We determine subhalo concentrations from a redshift-dependent mass-concentration relation extracted from the ShinUchuu N-body simulation (Ishiyama & Ando, 2020; Moliné et al., 2023, also see O'Riordan et al., 2023 for more details). This relation is derived in terms of $R_{\rm max}$ and $m_{\rm max}$, the radius of maximum tangential velocity and mass enclosed within it, as these more accurately describe the characteristics of haloes in simulations than the usual virial quantities. Our choice of mass-concentration relation results in more concentrated subhaloes than does the typical one from Duffy et al. (2008). We use a subhalo mass function that comes from fitting to the data in Lovell (2020) and has been reparameterised in terms of $m_{\rm max}$. After drawing the subhalo mass, $m_{\rm max}$, from the mass function, we draw the corresponding $R_{\rm max}$ value from a log-normal distribution with mean

$$R^{\rm CDM}_{\rm max,mean} = \left(\frac{A^{2/B} G m_{\rm max} M_\odot}{100\ ({\rm km/s})^{-2}\ {\rm kpc}\ M_\odot}\right)^{B/(B+2)} {\rm kpc} \quad , \tag{3.2}$$

and standard deviation

$$\sigma = \exp\left[(p + mz) + (n - kz)\ln\left(\frac{m_{\rm max}}{10^{10} M_\odot}\right)\right] \quad . \tag{3.3}$$

Here, $A = az + d$ and $B = cz + b$. Values for a, b, c, d, k, m, n, and p are listed in Table 3.1, and z is the redshift. The NFW profile for a subhalo then has normalization

$$\rho_0 [M_\odot {\rm kpc}^{-3}] = \frac{m_{\rm max}(1+C)^2}{4\pi C^2 R_{\rm s}^3} \quad , \tag{3.4}$$

where R_s is the scale radius and $C = R_{max}/R_s = 2.16$ (Bullock et al., 2001). We do not include tidal truncation or a dependence on the distance from the main lens centre of the mass-concentration relation because Despali et al. (2018) have shown the effects to be small compared to the scatter on the mass-concentration relation.

We generate populations consistent with predicted CDM subhalo and field halo mass functions down to a halo mass of $10^5 M_\odot$, and we assume the total mass in substructure in the region of the lensed images to be ~ 2 per cent of the total mass of the main lens in that region, which is roughly consistent with observational constraints (Dalal & Kochanek, 2002; Hsueh et al., 2020; Gilman et al., 2020). This substructure fraction is higher than simulation predictions by Xu et al. (2015), but since we are testing the ability of multipoles to mimic low-mass haloes that strongly perturb the flux ratios, a bias towards models that have more perturbing haloes is a conservative choice.

3.3.3 Selecting realizations for simulated lens systems

Because we will proceed to stress test multipoles as they try to reproduce the flux ratios resulting from these EPLγ+CDM models, we generate 2000 realizations for each of the four main types of simulated lens in our sample (cross, fold, cusp, or satellite) and then select for each type the realization that produces the most extreme flux-ratio anomalies. These four EPLγ+CDM models should thus present the flux ratios that are most difficult for the EPLγ+MP models to reproduce. If models with multipoles but without low-mass haloes can fit perturbations produced by the most extreme halo populations, they should be able to do so in nearly all cases. We stress that even though strong the perturbations in our simulated lens systems are of comparable magnitude to flux-ratio anomalies in real observations (e.g. Nierenberg et al., 2020b). Each simulated lens system contains total flux-ratio perturbations in excess of 10 per cent, and some images are perturbed beyond 20 per cent. Table 3.2 lists the macro-model parameters used in each of the four simulated systems, and Table 3.3 presents the image positions and flux ratios. We add to each flux in the models uncertainties that are based on the observations of the real lenses on which they were based.

Table 3.2 Macro-model parameters used to create mock observations. The columns are z_l, the lens redshift, z_s, the source redshift, $\theta_{\rm E}$, the Einstein radius, dRA, the lens right ascension with respect to the observation center, dDec, the lens declination with respect to the observation center, t, the power law slope, ϵ, the ellipticity, ϕ, the position angle, $\gamma_{\rm ext}$, the external shear strength, $\phi_{\rm ext}$, the shear angle, $\mathrm{dRA_{src}}$, the source right ascension with respect to the observation center, and $\mathrm{dDec_{src}}$, the source declination with respect to the observation center.

Lens	z_l	z_s	$\theta_{\rm E}$	dRA	dDec	t	ϵ	ϕ	$\gamma_{\rm ext}$	$\phi_{\rm ext}$	$\mathrm{dRA_{src}}$	$\mathrm{dDec_{src}}$
Cross	0.29	1.713	0.702	0.380	0.552	1.013	0.076	-62.66	0.031	-2.89	0.385	0.560
Fold	1.04	2.23	0.655	-0.003	0.000	1.00	0.100	-90.00	0.100	90.00	0.030	0.079
Cusp	0.34	3.62	0.765	0.699	-0.629	1.00	0.176	60.90	0.193	52.96	0.388	-0.414
Satellite (main)	0.31	1.70	0.653	0.794	0.248	1.02	0.140	-54.51	0.216	26.81	0.798	0.177
Satellite (satellite)			0.101	0.337	-1.032	1.00	0.0					

3.4 Modelling of simulated data: EPLγ+MP

In accordance with previous work examining the lensing effects of complex angular structure, we describe the convergence of multipoles in polar form

$$\kappa(R, \phi) = R^{-t}[a_m \cos(m\phi) + b_m \sin(m\phi)] \quad . \tag{3.5}$$

Here, a_m and b_m are the standard multipole sine and cosine amplitudes, and m is the multipole order. For ease of interpretation, we also describe multipoles in terms of their overall amplitudes,

$$\eta_m = \sqrt{a_m^2 + b_m^2} \quad , \tag{3.6}$$

and orientation angles,

$$\phi_m = \frac{1}{m} \arctan\left(\frac{b_m}{a_m}\right) \in \frac{1}{m}[0, 2\pi) \quad . \tag{3.7}$$

We focus on third- and fourth-order multipoles ($m = 3, 4$), as these are expected to cause flux-ratio effects degenerate with those of perturbing haloes (Evans & Witt, 2003; Congdon & Keeton, 2005). Lower-order multipoles have effects that are analogous to changes in the macro-model parameters, and higher-order multipoles may introduce deviations from ellipticity that either produce greater than four images or are unphysical (Evans & Witt, 2003; Congdon & Keeton, 2005). There is also ample evidence for third- and fourth-order multipoles in the observed isophotes of lens-like elliptical galaxies (Hao et al., 2006).

To assess the possible degeneracy between the lensing effects of multipoles and haloes above our standard mass range, we do not include the SIS satellite galaxy in any of our models of that lens system.

3.5 Results

In this section, we attempt to reproduce the simulated lens systems with multipoles in two distinct ways. First, we apply third- and fourth-order multipoles separately (Section 3.5.1). We explicitly step through the parameter space of strengths and orientation angles for each to explore the possible range of flux-ratio perturbations induced by the individual multipole orders. Then, we use MCMC to fit each simulated system with both multipole orders simultaneously (Section 3.5.2).

Table 3.3 Image positions and flux ratios for each of the simulated lens systems along with the fractional degree of flux perturbation by low-mass haloes compared to the best-fitting EPL and shear-only fiducial model. All position units are in arcsec. These are generated using an EPL+shear macro-model in addition to a population of perturbing CDM subhaloes and field haloes. No multipoles are present in any of the simulated systems. Positions are given with respect to the observation centre, and all dRA and dDec uncertainties are 0.005. Flux ratio uncertainties are listed in parentheses next to each flux ratio.

Lens	Image	Position		Flux Ratio
		dRA	dDec	
Cross	A	1.0656	0.3204	1.000 (0.030)
	B	0.0026	-0.0017	0.508 (0.001)
	C	0.7222	1.1589	0.920 (0.030)
	D	-0.1562	1.0206	0.658 (0.030)
Fold	A1	-0.4985	-0.2207	0.288 (0.012)
	A2	0.2364	-0.6048	1.000 (0.040)
	B	0.4897	-0.3895	0.893 (0.036)
	C	0.0725	0.8233	0.299 (0.015)
Cusp	A	0.3908	0.3213	1.000 (0.010)
	B	0.0003	0.0003	1.149 (0.010)
	C	-0.3330	-0.7463	0.537 (0.010)
	D	0.9511	-0.8018	0.045 (0.010)
Satellite	A	1.6217	0.5890	0.867 (0.030)
	B	-0.0005	0.0003	1.000 (0.010)
	C	0.8328	-0.3170	0.670 (0.030)
	D	0.4948	0.7377	0.694 (0.030)

3.5.1 Independent investigation of third- and fourth-order multipoles

At this stage we are not trying to fit to the flux ratios in our simulated sample, but rather to explore the dependence of the flux-ratio perturbations on the multipole amplitudes (η_m, where m is either 3 or 4) and position angles (ϕ_m). We do this by generating a grid of (η_m, ϕ_m) pairs and, for each grid point, adding multipole components with those parameters to the 200 base EPLγ models. We do this exercise for $m = 3$ and $m = 4$ separately. While, judging from isophotes, multipole amplitudes are not expected to be much larger than $\eta_m \approx 10^{-2}$ in real galaxies (see Hao et al., 2006), we explore 11 multipole strengths ranging from $10^{-3} - 10^{-1}$ with equal logarithmic spacing. For each multipole strength, we examine an evenly spaced set of 10 orientation angles encompassing the full range over which they are unique, i.e., from −60 to +60 degrees for the third-order multipoles and from −45 to +45 degrees for the fourth-order multipoles. To correct for

any astrometric perturbations introduced by the addition of the multipole component, we optimize the macro-model parameters to fit to the simulated image positions after adding the multipole.

We show the results of this exploration in Figures 3.1 (third-order) and 3.2 (fourth-order). In each of the panels in the first two columns, the horizontal lines show the flux ratios of our simulated lenses while the points show the flux ratios produced by the EPLγ+MP models. In the left-hand columns we show how the perturbations change with multipole amplitude, showing only the points for values of ϕ_m corresponding to the orientations of the highest-likelihood realizations. We calculate the likelihood of a realization from

$$\chi^2 = \sum_i \frac{\left\|\mathbf{x}_i^{\mathrm{m}} - \mathbf{x}_i^{\mathrm{d}}\right\|^2}{\delta x_i^2} + \sum_i \frac{\left(f_i^{\mathrm{m}} - f_i^{\mathrm{d}}\right)^2}{\delta f_i^2} \quad , \tag{3.8}$$

where $\mathbf{x}_i$ and f_i are the image positions and flux ratios of a model realization (denoted ^{m}) and simulated data (denoted ^{d}). δx_i and δf_i are the simulated uncertainties. As expected, flux-ratio perturbations get larger with increasing multipole amplitude. However, the size of this effect is dependent on the configuration and the particular image in question.

The centre columns of Figures 3.1 and 3.2 show perturbations due to multipoles over the full range of orientation angles with η_m values fixed to those of the highest-likelihood realizations. All four lens systems show clear periodic behavior as the orientation angle changes. In all cases, some realizations bring the flux ratios closer to the simulated values than the macro-model alone. The simulated flux ratios in the fold and cusp lens systems can be reproduced within observational uncertainty by either third- or fourth-order multipoles, though the highest-likelihood realizations for the cusp system both have potentially unrealistic[1] amplitudes of $\eta_m = 0.1$. Though no third- or fourth-order multipole perturbations can reproduce the simulated flux ratios in the cross and satellite systems, the magnitude of the discrepancy between model-predicted and simulated flux ratios can be significantly reduced by either order with reasonable amplitudes.

The flux-ratio perturbations induced by the SIS in conjunction with low-mass haloes in our simulated satellite system are comparable to those induced by low-mass haloes alone in other mocks. While satellite galaxies may be directly observable from their light, there may be a

[1]What we mean here by unrealistic is large compared to what is on average observed in the isophotes of elliptical galaxies. However, as we discuss in Section 3.6, the amplitude of multipole components in the light and mass distribution do not necessarily have to agree.

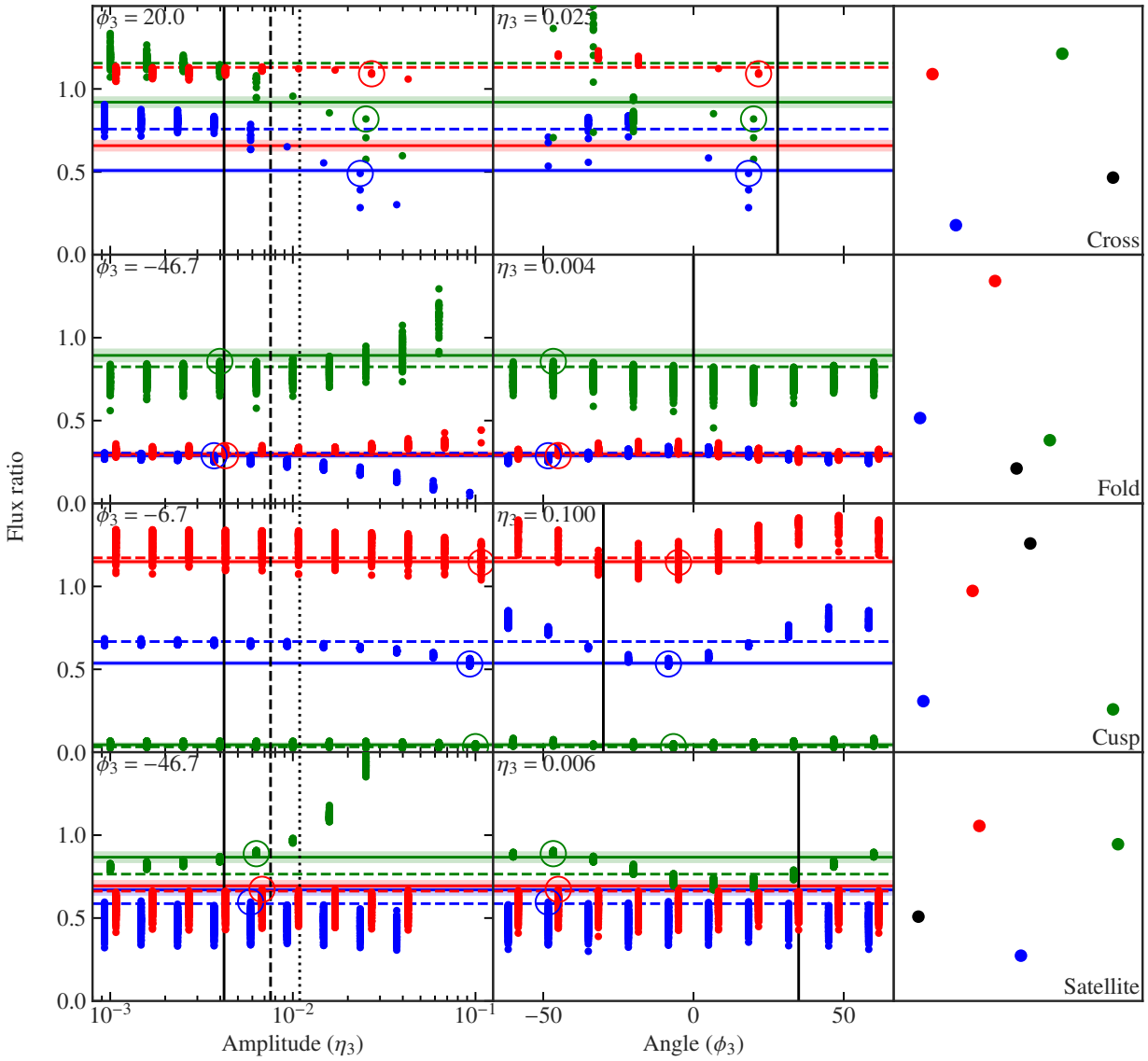


Figure 3.1 The range of flux-ratios as a function of third-order multipole parameters η_3 (left) and ϕ_3 (middle). Only flux ratios from models with position error less than 1σ for each image are included. The horizontal lines and shaded regions represent the flux-ratios and uncertainties in the simulated data, which is generated by an EPLγ+CDM model. The filled dots show the flux-ratios of each image for each multipole parameter value. Open circles represent the flux-ratios from the EPLγ+MP model that most closely matches the simulated flux ratios. The solid, dashed and dotted vertical black lines in the left column indicate the mean, 1σ and 2σ values, respectively, of the third-order multipole amplitude distributions from elliptical galaxy isophotes (Hao et al., 2006). The solid black line in the middle column indicates the orientation angle of the EPL major axis in each simulated lens system. The rightmost column shows the image configuration for each lens system, and the color of each image corresponds to data for its respective flux ratio in the other columns. The black point marks the image used in the denominator of the flux ratios.

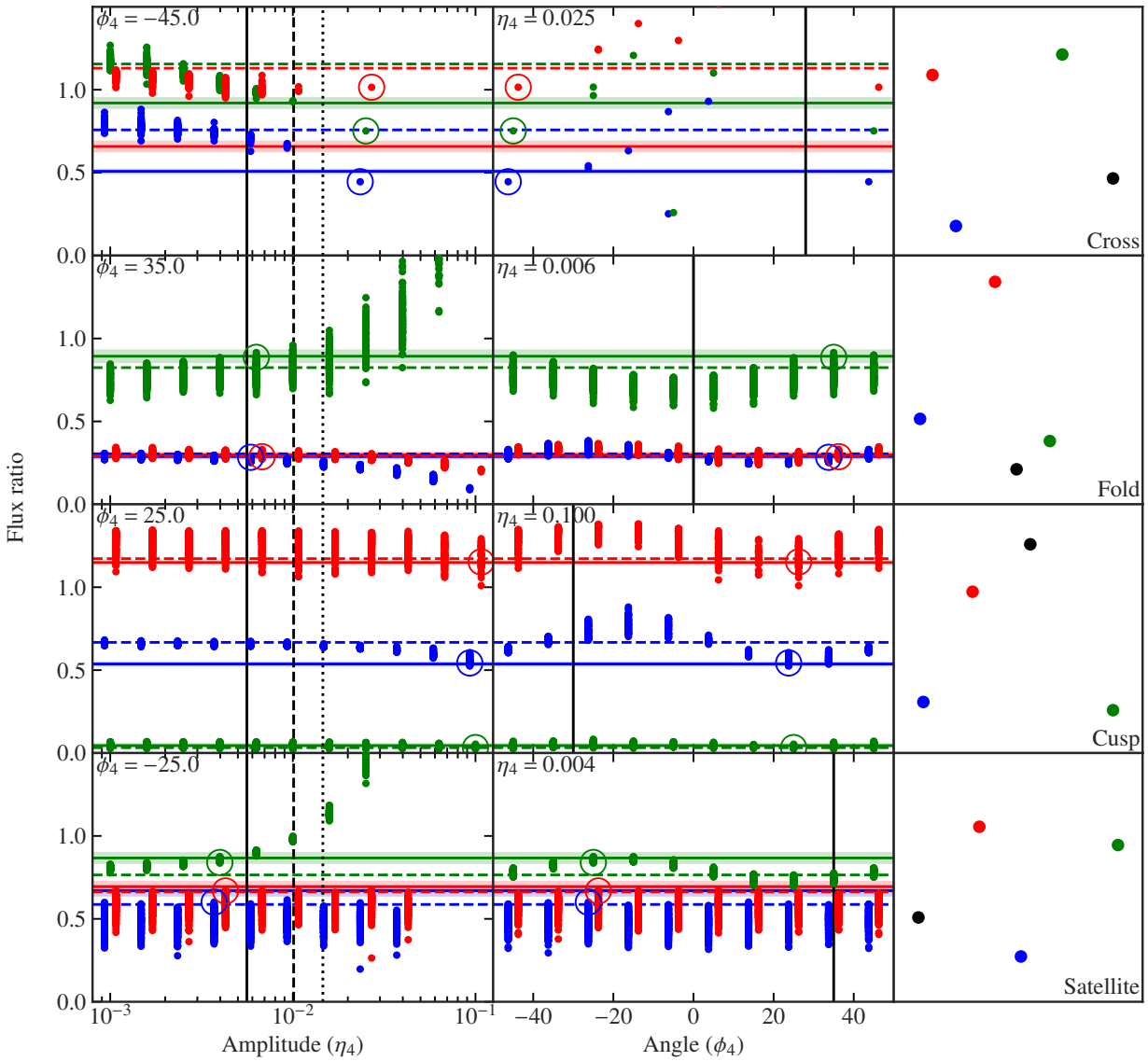


Figure 3.2 Same description as Figure 3.1, except with fourth-order multipoles.

degeneracy between their inferred properties and multipole amplitudes. We leave the investigation of this potential degeneracy to a future work.

3.5.2 Joint investigation of third- and fourth-order multipoles

We now fit the mass model and multipole parameters simultaneously to both the flux ratios and image positions of our simulated lens systems. These data are the elements of a vector $\mathbf{d}$. Similarly, the macro model parameters are the elements of a vector $\theta_{\rm MM} = \{\theta_E, ...\}$, and the multipole amplitudes form a vector $\theta_m = \{a_3, b_3, a_4, b_4\}$. From Bayes' theorem, the posterior distribution of the parameters given the data is

$$\Pr(\theta_m, \theta_{\rm MM}|\mathbf{d}) = \frac{\Pr(\mathbf{d}|\theta_m, \theta_{\rm MM}) \Pr(\theta_m, \theta_{\rm MM})}{\Pr(\mathbf{d})}. \tag{3.9}$$

The first term in the numerator $\Pr(\mathbf{d}|\theta_m, \theta_{\rm MM})$ depends only on χ^2 defined previously (see Equation 3.8). The prior probability $\Pr(\theta_m, \theta_{\rm MM}) = \Pr(\theta_m) \Pr(\theta_{\rm MM})$ is the probability of a given parameter value before the data is observed, based on other information. The normalisation of the posterior, or the evidence, $\Pr(\mathbf{d})$ can be ignored in this case as we only consider one model. The posterior we use in practice is then

$$\Pr(\theta_m, \theta_{\rm MM}|\mathbf{d}) \propto \Pr(\theta_m, \theta_{\rm MM}) \exp\left(-\frac{1}{2}\chi^2\right). \tag{3.10}$$

The calculation of this many-dimensional posterior is intractable, so we use MCMC to obtain samples of θ_m and $\theta_{\rm MM}$. The density of these samples represents the posterior probability distribution. To conduct the MCMC sampling, we use the EMCEE[2] ensemble sampler with 120 walkers and 10,000 burn-in steps that are discarded, followed by 20,000 recorded steps. We use broad uniform priors on the mass model parameters $\theta_{\rm MM}$ and use a normally distributed prior with mean zero and standard deviation of one per cent on the multipole amplitudes.

With both third- and fourth-order multipole parameters free, all flux ratios in the simulated lens systems can be fit within 1σ. Figure 3.3 displays flux ratio histograms from the joint MCMC sampling of both multipole orders and macro-model parameters (EPLγ+MP). For reference, we also include the flux ratio histograms that result from the MCMC sampling with only an EPLγ.

[2]Foreman-Mackey et al. (2013)

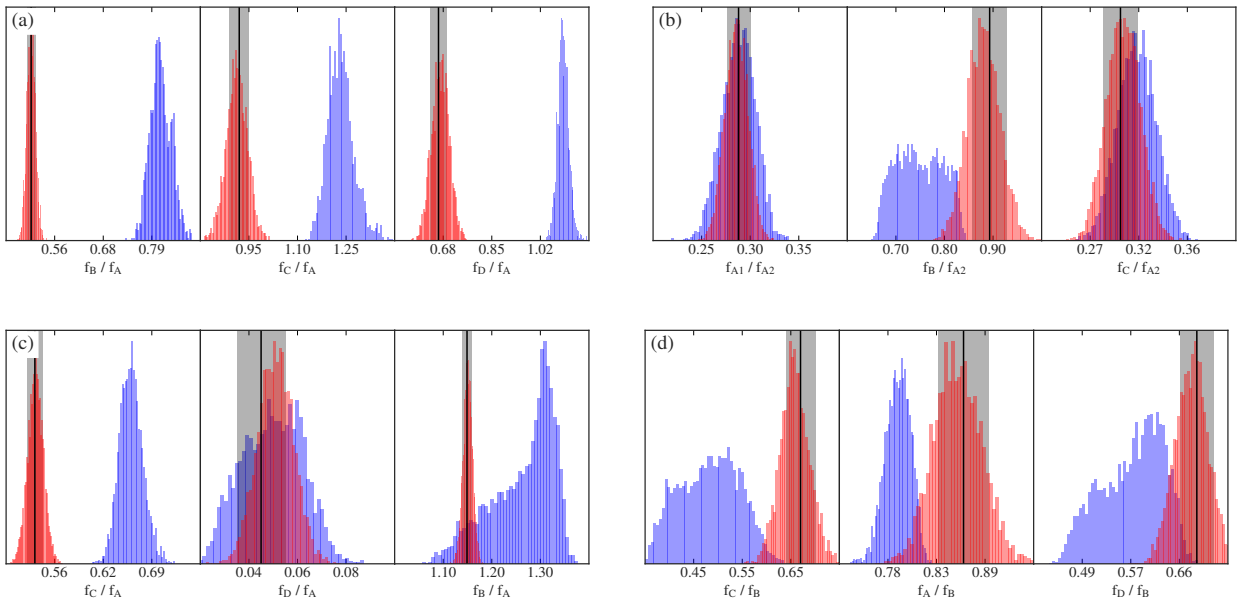


Figure 3.3 Flux ratio histograms created using random samples of 10,000 models from MCMC chains for all four simulated lens systems: cross (a), fold (b), cusp (c) and satellite (d). For comparison, we show distributions resulting from using EPLγ models (blue) and EPLγ+MP models with third- and fourth-order multipoles included simultaneously (red). Vertical black lines show flux ratios from the simulated data, and the surrounding grey regions show the associated uncertainties.

These EPLγ-only models are unable to reproduce the simulated flux ratios within observational uncertainty in every case. Generally, the addition of third- and fourth-order multipoles to the model makes the model flux ratio distributions narrower and roughly centered around the simulated values. In other words, when multipoles with amplitudes that are consistent with the amplitudes of observed isophotes are included in the model, the simulated flux ratios do not appear anomalous despite the fact that they were perturbed by a population of CDM low-mass haloes, and, in one case, also a satellite galaxy.

Figures 3.4, 3.5, 3.6 and 3.7 show the resulting posterior distributions for the third- and fourth-order multipole parameters for all four lens systems. Non-zero third- and fourth-order multipole amplitudes are preferred in every case, and most orientation angles are constrained within 20 degrees.

Isodensity contours for the best-fitting models for each system and their underlying macro-models are shown in Figure 3.8. The shapes of the contours for the cross and fold systems generally become 'boxier'. However, offsets from perfect alignment between the fourth-order

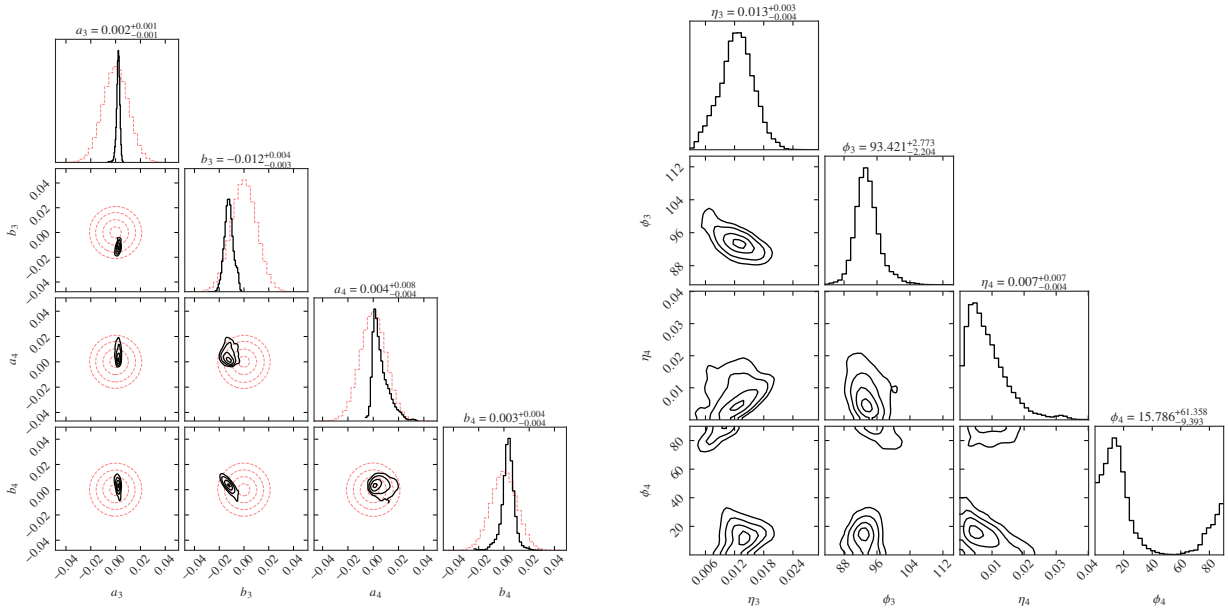


Figure 3.4 Corner plots from simultaneous MCMC sampling of EPL, shear and both multipole order parameters on the cross lens system. MCMC was conducted in the sine/cosine amplitude basis (left), and we also present the same data transformed into the overall amplitudes and angles (right) for ease of interpretation. Only multipole parameters are displayed, but all EPL and shear parameters, along with the source position, were allowed to vary freely.

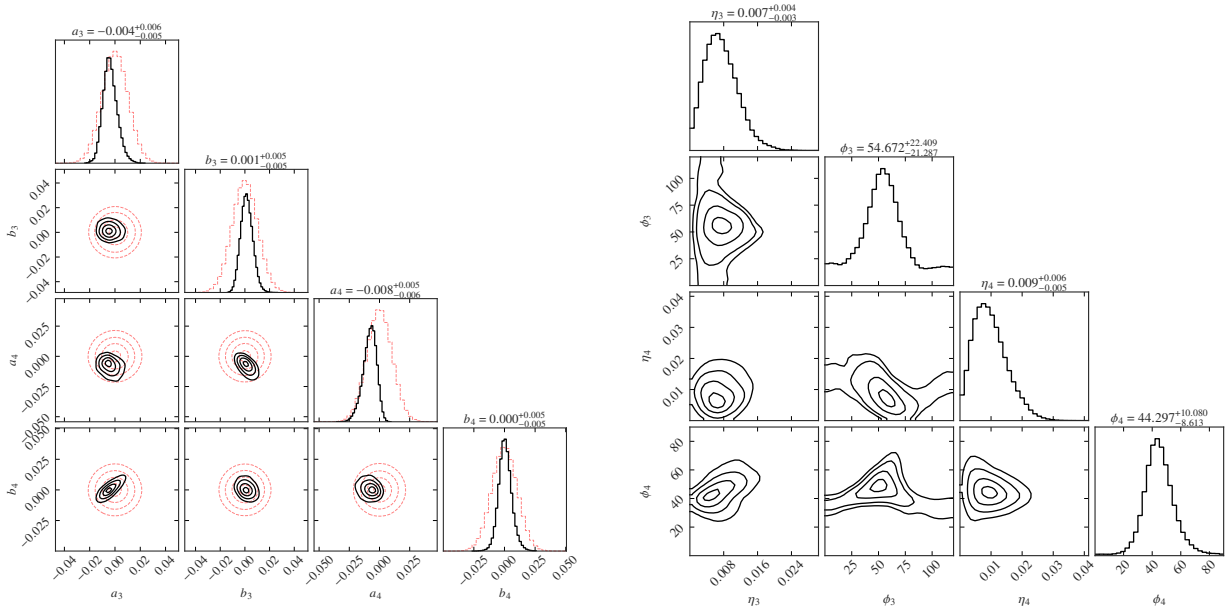


Figure 3.5 Same description as Figure 3.4, except for the fold system.

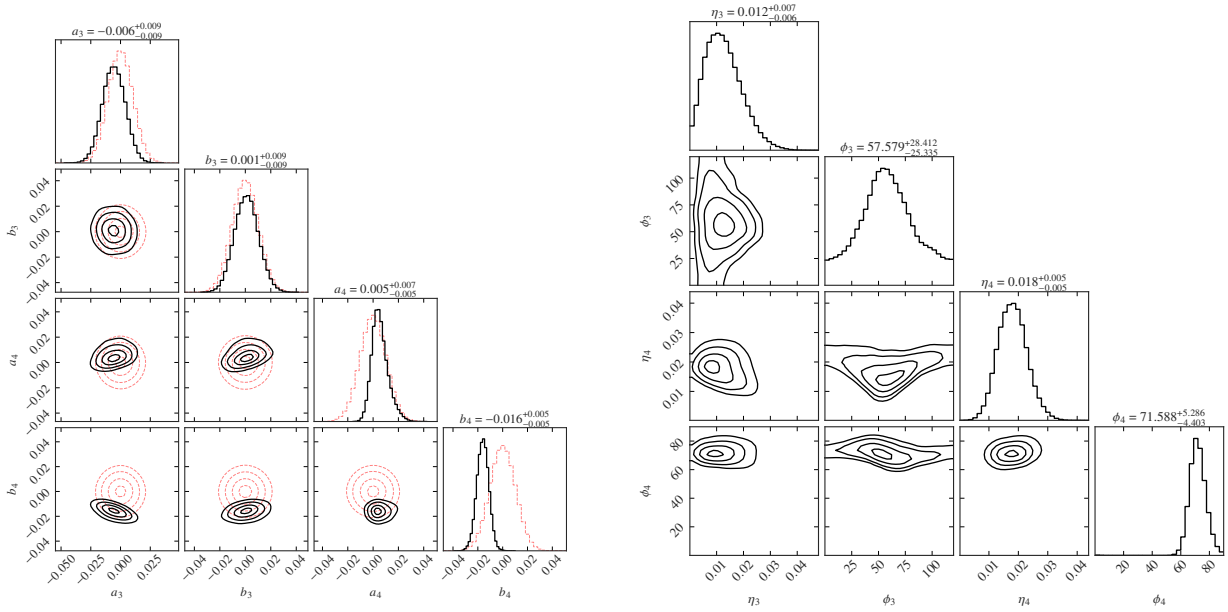


Figure 3.6 Same description as Figure 3.4, except for the cusp system.

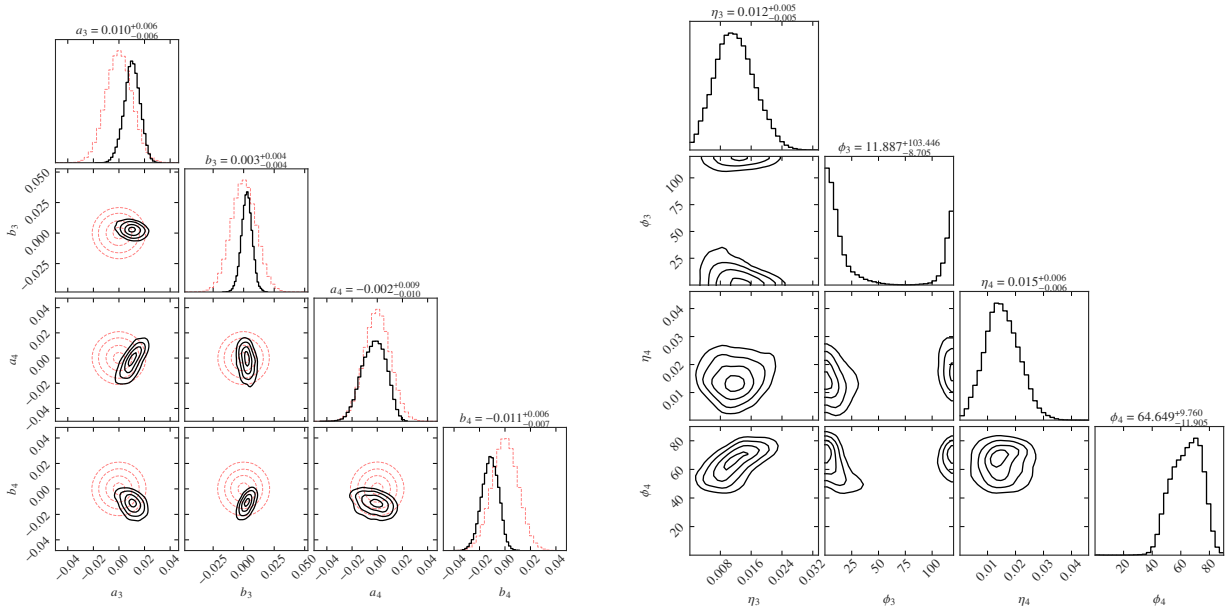


Figure 3.7 Same description as Figure 3.4, except for the satellite system.

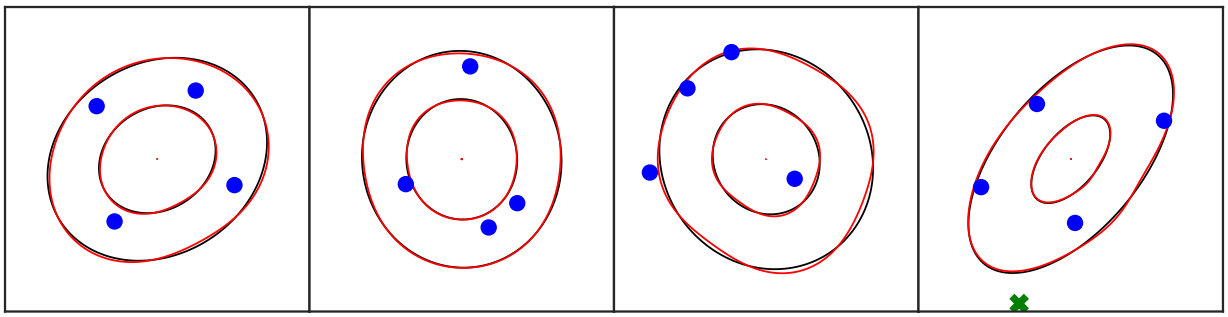


Figure 3.8 Isodensity contours of the best-fitting models from the MCMC analysis of the cross, fold, cusp and satellite systems (left to right). The red curves show contours for the total convergence, including the EPL, shear and third- and fourth-order multipoles. The underlying black curves show contours of the EPL and shear only. Multipole parameter values: **cross**: $\eta_3 = 0.0109$, $\phi_3 = 93.91$, $\eta_4 = 0.0023$, $\phi_4 = 20.73$; **fold**: $\eta_3 = 0.0068$, $\phi_3 = 62.21$, $\eta_4 = 0.0052$, $\phi_4 = 44.86$; **cusp**: $\eta_3 = 0.0073$, $\phi_3 = 51.36$, $\eta_4 = 0.0151$, $\phi_4 = 73.19$; **satellite**: $\eta_3 = 0.0088$, $\phi_3 = 32.15$, $\eta_4 = 0.0037$, $\phi_4 = 54.78$.

multipole and the EPL major axes along with the inclusion of third-order multipoles, makes this effect more irregular than pure 'diskiness' or 'boxiness'. Deviation between the perturbed and unperturbed contours for the satellite system occurs roughly in alignment with the location of the missing satellite galaxy. It is worth noting again that the simulated data were generated without any multipoles.

3.6 Discussion and conclusions

We find that there is a substantial degeneracy between the perturbative lensing effects of low-mass dark matter haloes and multipoles when modelling quadruply-imaged quasars. Individually, third- and fourth-order multipoles with amplitudes around or below 0.01 can produce flux-ratio perturbations over 40 per cent. This is sufficient to reproduce perturbations induced not only by a population of low-mass CDM haloes, but also a satellite galaxy. Our results agree well with early investigations of this topic by Evans & Witt (2003) and Congdon & Keeton (2005). From the analysis of real data, they find that all but the most extreme flux-ratio anomalies, such as those in the cusp systems B2045+265 and B1422+231, which differ from EPLγ predictions in excess of ~50 per cent, can be fit using multipoles with amplitudes smaller than 0.01.

3.6.1 A more general treatment of multipoles

In our analysis of the ability of multipole components to reproduce flux ratio anomalies, we utilized two important general treatments of the multipoles: we included third-order multipoles in addition to the more standard fourth-order one, and we did not fix their position angle to the major axis of the underlying EPL model.

The freedom allowed to the position angle plays a key role in the degeneracy between multipoles and low-mass haloes. As we have shown, the orientations of multipoles that most closely reproduce the perturbed flux ratios in our simulated data do not always align with the EPL. Evans & Witt (2003), Kochanek & Dalal (2004) and Congdon & Keeton (2005) also found that a larger portion of observed flux-ratio anomalies can be reduced or eliminated when multipole angles are allowed to vary freely as opposed to being fixed.

Recent flux-ratio analyses, where fourth-order multipoles have been included (e.g. Gilman et al., 2021, 2022, 2023, 2024), have fixed the orientation angles to align with the EPL major axis. However, neither the total mass density nor the light distribution of real galaxies seem to suggest this assumption to be generally valid. Using high-angular resolution observations of real gravitational lens systems with an extended source, Powell et al. (2022) and Stacey (2024) have found that the multipoles can be misaligned with respect to the underlying EPL. Additionally, tilted dark matter haloes are commonly observed in numerical simulations (e.g Han et al., 2023). At the same time, the observed isophotes of elliptical galaxies only have order $m = 4$ perturbations aligned with the EPL for larger values of a_4, in excess of one per cent, i.e. for truly boxy isophotes (Hao et al., 2006, their fig. 4). The fourth-order multipole in the best-fitting model for our cusp system has an amplitude of 1.5 percent, and it does not align with the EPL. The best-fitting models for our other systems all have fourth-order multipoles with amplitudes $\lesssim 0.5$ per cent.

3.6.2 Isophotes and iso-density contours

An important comparison to our work is the study of isophotes in elliptical galaxies, since multipole patterns in the former are often used as motivation for including multipole components in the mass distribution of lens galaxies.

The amplitudes of the multipoles that we find to be highly degenerate with low-mass haloes are consistent with amplitudes of the isophotes observed in elliptical galaxies. Indeed, our best-

fitting values fall within the central 1-2 σ of the distributions found by Hao et al. (2006), even though we use Gaussian priors that allow for more extreme values. Interestingly, despite isophote observations indicating that third-order multipoles may on average be weaker than fourth-order (Hao et al., 2006), our findings indicate that they can still have significant impacts on flux ratios. We repeated our analysis with much tighter priors on the multipole amplitudes, Gaussians with widths 0.35 per cent instead of 1 per cent, and found that the cross and fold simulated lens systems can still be fit with $\chi^2 < 2$. The best-fitting realizations for the cusp and satellite systems both had $\chi^2 < 4$, much better than the best-fitting EPLγ models. Even with priors on the multipole amplitudes that are comparable to, and tighter in the case of a_4, those suggested from isophotes, multipoles can have a significant impact on the flux ratios.

Still, it is currently unclear how closely galaxy mass distributions resemble their observed isophotes. Recently, Stacey (2024) analysed a sample of three strong gravitational lens systems and inferred multipole amplitudes in the total mass density distribution of up to ~0.01 for both third and fourth orders. They found that in some cases the multipole amplitudes in the isophotes of the lens galaxies differ from those inferred in the total mass density distribution by 0.01 or more. Modelling very long baseline interferometric (VLBI) observations of a lensed radio jet, Powell et al. (2022) found the third-order multipole to have an amplitude that is roughly twice as large as that of the fourth order, in disagreement with expectations from observations of isophotes in elliptical galaxies. Whether this is related to a genuine discrepancy between the light and mass distributions or to the fact that third-order multipoles in the mass density are more degenerate with low-mass haloes, at least for resolved lensing observations (O'Riordan & Vegetti, 2023), remains to be determined. As already pointed out by Stacey (2024), the environment within which the lens galaxies reside may also play a role in observed differences between the isophotes and the iso-density contours. In general, our lack of knowledge about how well the projected mass density profile at the Einstein radius is traced by light limits our capability to use isophote distributions as a strong prior for the properties of multipoles in the mass distribution of gravitational lens galaxies.

3.6.3 Considerations on the source size

To understand the role played by the source on the relative effect of low-mass haloes and multipoles, we quantify the level of flux-ratio anomaly induced by the two contributions, i.e., low-mass haloes

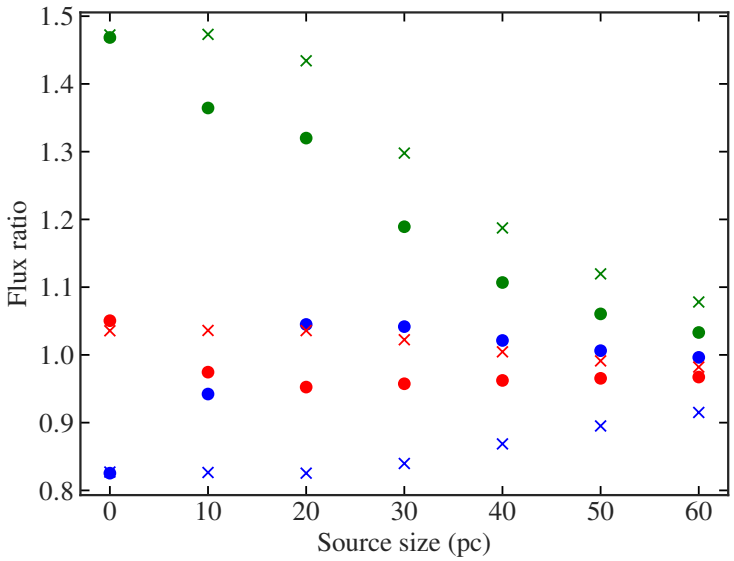


Figure 3.9 Flux ratios as a function of source size for fixed EPLγ+CDM (circles) and EPLγ+M34 (crosses) models that predict the same image positions and point-source flux ratios. Source size denotes the standard deviation of a Gaussian brightness distribution. Note that the physical extents of the radio-emitting, mid-infrared-emitting and warm dust regions of quasars are typically expected to be $\lesssim$ 10pc.

and multipoles, independently as a function of the source size (see Figure 3.9). We select two models, one perturbed by low-mass haloes and the other by multipoles, which result in the same image positions and flux ratios from a lensed point source. As we increase the source size from a point to 40 pc, we find that the flux ratios perturbed by low-mass haloes approach their unperturbed values (i.e. the values expected from an EPLγ-only mass distribution). Even when the source is only 10 pc in size, there is a $\sim$10 percent change of two of the flux ratios from their point-source values. On the other hand, the flux ratios perturbed by multipoles are hardly affected by increasing the source size up to 20 pc (i.e. 0-3 per cent). While in both cases larger source sizes produce smaller flux ratio anomalies, the effect is much smaller for the multipoles than for low-mass haloes. As the multipoles are less sensitive to changes in the source size, the more extended the source, the lower the multipole amplitudes that are needed to reproduce the effect of a given halo population. We can conclude, therefore, that our assumption of a point source gives a conservative quantification of the degeneracy between low-mass haloes and multipoles.

3.6.4 Implication for dark matter inferences using flux ratios

Traditionally, flux-ratio constraints have been based on the concept that any anomaly that is seen with respect to the flux ratios predicted by the smooth EPLγ model (perhaps with the addition of a stellar disc; Hsueh et al., 2016, 2017, 2018), is due to perturbations by clumpy dark matter. Thus, the underlying assumption is that the prevalence of observed flux-ratio anomalies can be used to infer the mass-function of low-mass dark matter haloes and, in turn, the properties of dark matter. If, instead, some unknown fraction of those anomalies are caused by multipole structure in the primary lensing galaxy, the constraining power of this method is diminished.

In this paper we have shown that, when presented with a simple set of observables, namely four image positions and four fluxes, it is generally possible to reproduce those observations with a model consisting of an EPLγ base model plus either (1) low-mass CDM haloes but no multipoles or (2) multipoles but no low-mass haloes. The fluxes in real lens systems are likely to be affected by a combination of these two effects. Because we cannot identify from just the lensed image positions and fluxes whether observed anomalies are due to low-mass haloes or multipoles, any dark matter inference from real flux-ratio data will be affected by what weight is given to each of these two components, i.e., by the priors on the relative contributions of the perturbing haloes and multipole components. For example, neglecting the contribution of the multipole component or reducing its complexity, e.g., by not including third-order multipoles or by fixing the multipole position angle to match the major axis of the EPL, will lead to a bias in favour of colder dark matter models. Most extremely, an inference could be made under the assumption that all perturbations were due to low-mass haloes, as was done in Hsueh et al. (2020), in which case a CDM-like model, which predicts large numbers of low-mass haloes, is likely to be preferred. Similarly, an inference could be made assuming the maximum multipole contribution, in which case models that predict fewer low-mass haloes, such as WDM, are likely to be preferred.

All but the most extreme flux-ratio anomalies seen in real lensed quasar systems are comparable to or smaller than the anomalies considered in this paper, indicating that our ability to fit the anomalies with multipole components is not due to selectively choosing easy systems to fit. Thus, one conclusion to be drawn from our results is that, in the absence of other data that can inform our priors, lensed quasar flux ratios alone are ineffective at placing constraints on dark matter models.

In a follow-up paper, we will quantify how the inclusion of multipole components in the lens

mass distribution affect existing constraints on dark matter from the two recent investigations by Hsueh et al. (2020) and Gilman et al. (2020). Despite their different approach to multipoles, both works found nearly identical constraints on the WDM particle mass. However, neither work allowed for a general treatment of multipoles as presented in this paper. We expect that allowing more freedom in the angular structure of the lens galaxies will reduce the contribution from low-mass haloes to the observed flux ratios, and potentially weaken the constraints on dark matter.

3.6.5 Future prospects

As the prospects to derive meaningful constraints on the properties of dark matter with flux ratio only observations strongly depends on independent knowledge about the general mass density distribution of lens galaxies, it is fair to wonder where the required information could be obtained.

One option is deep observations that could allow one to identify and quantify the prevalence of multipoles in the light distribution of lens galaxies. However, it is currently unknown how well the total mass distribution follows that of the light. Though the best-fit multipole amplitudes that we have found in this paper are not in tension with those observed in galaxy isophotes by Hao et al. (2006), the current number of quadruply-imaged quasar observations is not sufficient to precisely quantify the distributions of inferred multipole parameters across all elliptical galaxies from lens modelling.

Another option is modelling high-resolution images of strongly lensed extended emission, which is expected to allow for the constraint of more complex macro model features including multipoles (e.g. Powell et al., 2022; Stacey, 2024). In principle, this information could be applied to the analysis of flux ratios when available (Gilman et al., 2024). However, high-resolution imaging of lensed extended emission is at the moment not available for the majority of the ~ 15 lens systems in the current flux ratio sample. In addition, O'Riordan & Vegetti (2023) have shown that the degeneracy between multipoles and low-mass haloes persists in the modelling of lensed extended emission, leaving as reliably detectable only low-mass haloes which are located in projection close to the lensed images. Furthermore, Herle et al. (2023) have shown that automatic lens finding techniques may lead to the selection of gravitational lens galaxies that have different mass distributions when the sources are extended instead of unresolved. If confirmed, their result may place some doubts on the possibility of transferring knowledge from one type of lens system

to another.

Galan et al. (2022) have recently shown that it may be possible to distinguish between the lensing effects of multipoles and low-mass haloes using extended sources and a non-analytical description of the lensing potential, but this method has yet to be demonstrated on real data. If proven successful, in combination with the large number of gravitational lens systems expected from the James Webb Space Telescope, Euclid and the SKA (Nierenberg et al., 2023; Collett, 2015; McKean et al., 2015), it may represent a path forward to obtain the necessary information on the mass density distribution of lens galaxies. Similarly, numerical simulations could provide useful insights on the mass and light distribution of (lens) galaxies (e.g. Gao & White, 2006; Lokas, 2022; Despali et al., 2022; Han et al., 2023).

To conclude, our results as well as previous similar works demonstrate that the analysis of flux-ratio anomalies, in its current form, cannot directly constrain the properties of the low-mass dark matter halo population. This follows from the main issue identified in this work, that the priors imposed on multipole parameters play a dominant role in the inference process. Though our focus is on CDM, we expect our results to be applicable to other dark matter models, but potentially with some differences. We expect, for example, the degeneracy here identified to be stronger for warm (WDM) and weaker for self-interacting (SIDM) dark matter. In the former case, the amplitude of the multipoles should not be significantly different than in CDM, but as low-mass haloes are less numerous and concentrated they result in weaker flux ratio anomalies. In SIDM models, galaxies are predicted to be rounder (e.g. Brinckmann et al., 2018; however, see Despali et al., 2022), while low-mass haloes, which experience core-collapse, have a stronger lensing effect (Gilman et al., 2023). We will quantify the degeneracy between multipoles and low-mass haloes in different dark-matter models in a follow-up publication.

Acknowledgements

This work used the following software packages: AstroPy (The Astropy Collaboration et al., 2018), Matplotlib (Hunter, 2007), NumPy (van der Walt et al., 2011), GRAVLENS (Keeton, 2011), pylens (provided by M. Auger) and emcee (Foreman-Mackey et al., 2013).

Chapter 4

Summary and Conclusions

In this dissertation, I have detailed my development and use of strong gravitational lensing techniques as a means for learning about the dark matter distribution on sub-galactic scales. Because the flux ratios of quadruply-imaged quasars are sensitive to magnification by low-mass halos, they offer a probe of the populations along their lines of sight. The information gained about the population of subhalos, field halos and their internal structure allows us to test the predictions of CDM and constrain alternative models. Discovery of frequent and extreme flux ratio anomalies would provide evidence for dark matter models that predict ample and concentrated halo populations, like CDM. On the other hand, a large sample of flux ratios that lack evidence of perturbation could indicate a suppression of the number of halos or their central densities, as in WDM.

A growing sample of observations that are eligible for flux-ratio analysis should lead to tighter constraints on the properties of dark matter. Constraints from lensing analyses are complementary to other astrophysical probes such as Milky Way satellites and the Lyα forest, as each method entails a different set of assumptions and biases. However, a thorough understanding of the main lens mass distribution, comparison to which defines anomalous flux ratios, is essential for gravitational lensing constraints. In Chapter 2, I perform flux-ratio analysis on 14 lens systems, the largest sample to date. In addition to introducing the connection between the WDM particle mass and its effect on halos, I detail each of the base mass model components, our procedure for simulating CDM halo populations and our model of truncation of the mass functions and concentrations under WDM. I then present constraints on WDM of $\log_{10}(M_{\rm hm}/M_{\odot}) < 7.81$ ($m_{\rm th} > 5.23$keV), which have only been surpassed by analysis of Milky Way satellites. Finally, I note potential sources of systematic

error and the effect that different ways of interpreting the results have on the constraining power. Though in Chapter 3 I argue that this analysis cannot be taken at face value due to its exclusion of multipoles, the findings are still relevant as the proper treatment of multipoles is ambiguous and it is important to compare their implementation in various ways. The constraints inferred in Chapter 2 represent the limiting case in which third-order and higher multipoles have zero amplitude.

Next, in Chapter 3 I investigate the degree to which third- and fourth-order multipoles can perturb flux ratios. When applied independently, both multipole orders can cause perturbations of up to 40 percent. Using a diverse set of mocks that have been strongly perturbed by CDM halo populations, I show that they can be modeled within the observational uncertainties using smooth mass models with the joint inclusion of realistic third- and fourth-order multipoles and, importantly, without low-mass halos. This indicates, at least when analyzing only flux ratios and image positions from quads, that there is a complete degeneracy between low-mass halos and general third- and fourth-order multipoles. Powell et al. (2022) and O'Riordan & Vegetti (2023) find that this degeneracy extends to modeling lensed extended emission. This calls for a better understanding of multipoles in galaxy mass distributions, reevaluation of previous constraints from flux-ratio analysis that did not sufficiently include them, and rigorous exploration of the ways in which they can be included in an inference procedure.

4.1 Moving Forward with Multipoles

This work has uncovered an important source of systematic uncertainty. Future work will need to address this in order to fully realize the potential of lensing to constrain the nature of dark matter. There are two avenues of research that can be pursued to more definitively address the problem of multipoles. The reason for the ambiguity in implementing multipoles is that we do not have a good understanding of the frequency and degree to which they exist in the mass distributions of lens-like elliptical galaxies. Observations of both lens galaxy isophotes and lensed arcs around them as well as simulations will better inform multipole priors that can have a profound impact on dark matter inferences. As mentioned, there is also a wide range of ways in which multipoles can be included in mass models. It is important to experiment with them and see how, if at all, the resulting inferences vary.

Most gravitational lensing work that has included multipoles, including my own, uses the

distribution of multipoles found in the isophotes of lens-like galaxies as a reference for realistic multipole perturbations. Hao et al. (2006) analyzed 847 ellipticals from the Sloan Digital Sky Survey (Blanton et al., 2017) and derived distributions of third- and fourth-order multipole amplitudes that very roughly follow Gaussian distributions centered around 0 with widths of 0.0035. They also found that multipoles occur in all orientations. However, it is not known whether multipoles in the mass distribution of a galaxy match those of the light distribution. In fact, comparing multipoles found from modelling arcs in high-resolution ALMA observations to those found in isophotes of the lens galaxies Stacey (2024) find disagreement in two of the three systems investigated. While this tension already indicates that the multipoles in the mass and light distribution may differ, a larger sample of lens systems with observations of both lens galaxy isophotes observations and the arcs will allow us to quantify the discrepancy.

Even if a relationship, or a lack thereof, between multipoles in galaxy isophotes and lens models is concretely established, we still cannot be sure of the origin of inferred multipoles in the mass distribution. My work in Chapter 3 and that of O'Riordan & Vegetti (2023) indicate that there is a degeneracy between low-mass halos and multipoles that may lead to false detections of either. In this regard, simulations will be an invaluable complementary tool for searching for the emergence of angular complexity in galaxies through mergers and interactions with baryons (Prada et al., 2019; Chua et al., 2019). Galaxies from simulations have been found to display offsets between the halo and baryon centers (Liao et al., 2017) and radially dependent twists (Emami et al., 2021). With the next generation of simulations, it may be possible to resolve multipoles in the dark matter and baryon distributions. This would be doubly helpful, as dark matter multipole distributions alone could be used to set priors in lensing analysis, and their comparison to baryons would inform the proper use of isophotes for this purpose.

Given the current uncertainty regarding appropriate priors for multipoles in flux-ratio analysis, exploration of various choices is warranted. My work in Chapter 2 presents the resulting inference on WDM from currently available data if multipoles are left out. Other recent works have partially implemented multipoles, excluding the third order altogether and restricting the fourth order to align with the major axis of the EPL (Gilman et al., 2021, 2022, 2023). As expected given my findings in Chapter 3, this limited inclusion does not have a measurable impact on constraining power. It will be extremely important to repeat the analysis presented in Chapter 2 with general third-

and fourth-order multipoles free to vary in orientation and with varying priors on the amplitudes. This will allow us to find the degree of limitation of multipole amplitudes at which constraining power starts to return. As I have shown, I expect that any prior which allows for both amplitudes to be around or above ~0.01 will eliminate all constraining power. Beyond the types of priors on multipoles, other decisions about their place in the analysis procedure remain. For instance, should multipoles parameters be optimized after the addition of a low-mass halo population as are the macro-model parameters? Also, pending further knowledge from simulations about their relationship to dark matter models, can inferred multipoles actually contribute to dark matter constraints? Regardless of the results of the proposed inquiries via observation and simulation, multipoles must be implemented using a careful Bayesian approach in order for dark matter constraints from flux-ratio analysis to be trusted.

4.2 Future of the Field

Despite the challenges presented by multipoles, the enormous increase in the sample size of lens systems expected in the coming decades from JWST (see Nierenberg et al., 2023), Euclid (see O'Riordan & Vegetti, 2023) and the application of machine learning to efficiently identify lens candidates (see Herle et al., 2023) promises to accelerate progress in constraining dark matter using strong gravitational lensing. Flux-ratio analysis on its own may provide weaker constraints than originally predicted (e.g. Gilman et al., 2019). With the aid of simulation results, detailed mass models from lensed arcs and lens light observations, however, it has the potential to provide great value. Comparing the magnification of lensed quasar images allows us to probe down to smaller halo mass scales than any other method at redshifts $0 \lesssim z \lesssim 3$, thus it provides a unique insight into the distribution of dark matter. If the next generation of flux-ratio analysis reveals fewer anomalies that can be attributed to low-mass halos, it could indicate tension with longstanding CDM predictions. Otherwise, the space of viable alternatives would be further narrowed. The results of this thesis serve to highlight important next steps to mitigate systematic uncertainty, which will be necessary for any major future discoveries. It's a case of scientific progress, addressing successive systematic errors, which occasionally leads to remarkable discoveries.

References

Banik, N., Bertone, G., Bovy, J., & Bozorgnia, N. 2018, Journal of Cosmology and Astroparticle Physics, 2018, 061

Banik, N., Bovy, J., Bertone, G., Erkal, D., & de Boer, T. 2021, Journal of Cosmology and Astroparticle Physics, 2021, 043

Bender, R., Doebereiner, S., & Moellenhoff, C. 1988, A&AS, 74, 385

Bender, R., Surma, P., Doebereiner, S., Moellenhoff, C., & Madejsky, R. 1989, A&A, 217, 35

Bertone, G., & Hooper, D. 2018, Reviews of Modern Physics, 90, 045002, 1605.04909

Blanton, M. R. et al. 2017, AJ, 154, 28, 1703.00052

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

Bode, P., Ostriker, J. P., & Turok, N. 2001, ApJ, 556, 93, astro-ph/0010389

Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, MNRAS, 415, L40, 1103.0007

Brinckmann, T., Zavala, J., Rapetti, D., Hansen, S. H., & Vogelsberger, M. 2018, MNRAS, 474, 746, 1705.00623

Brooks, A. 2014, Annalen Phys., 526, 294, 1407.7544

Bullock, J. S., & Boylan-Kolchin, M. 2017, Annual Review of Astronomy and Astrophysics, 55, 343, https://doi.org/10.1146/annurev-astro-091916-055313

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, astro-ph/9908159

Burtscher, L. et al. 2013, A&A, 558, A149, 1307.2068

Cappellari, M. 2016, ARA&A, 54, 597, 1602.04267

Chen, G. C. F. et al. 2019, MNRAS, 490, 1743, 1907.02533

Chiba, M., Minezaki, T., Kashikawa, N., Kataza, H., & Inoue, K. T. 2005, ApJ, 627, 53, astro-ph/0503487

Chua, K. T. E., Pillepich, A., Vogelsberger, M., & Hernquist, L. 2019, MNRAS, 484, 476, 1809.07255

Collett, T. E. 2015, ApJ, 811, 20, 1507.02657

Congdon, A. B., & Keeton, C. R. 2005, MNRAS, 364, 1459, astro-ph/0510232

Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25, astro-ph/0111456

Davis, M., Huchra, J., Latham, D. W., & Tonry, J. 1982, ApJ, 253, 423

Despali, G., Giocoli, C., Angulo, R. E., Tormen, G., Sheth, R. K., Baso, G., & Moscardini, L. 2016, MNRAS, 456, 2486, 1507.05627

Despali, G., Vegetti, S., White, S. D. M., Giocoli, C., & van den Bosch, F. C. 2018, MNRAS, 475, 5424, 1710.05029

Despali, G., Walls, L. G., Vegetti, S., Sparre, M., Vogelsberger, M., & Zavala, J. 2022, MNRAS, 516, 4543, 2204.12502

Dobler, G., & Keeton, C. R. 2006, ApJ, 653, 1391, astro-ph/0608391

Dodelson, S. 2017, Gravitational Lensing (Cambridge University Press)

Dubinski, J., & Carlberg, R. G. 1991, ApJ, 378, 496

Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64, 0804.2486

Einasto, J., Joeveer, M., & Saar, E. 1980, MNRAS, 193, 353

Emami, R. et al. 2021, ApJ, 918, 7, 2012.12284

Enzi, W. et al. 2021, MNRAS, 506, 5848, 2010.13802

Evans, N. W., & Witt, H. J. 2003, MNRAS, 345, 1351, astro-ph/0212013

Fassnacht, C. D., Pearson, T. J., Readhead, A. C. S., Browne, I. W. A., Koopmans, L. V. E., Myers, S. T., & Wilkinson, P. N. 1999, ApJ, 527, 498, astro-ph/9907257

Fassnacht, C. D., Xanthopoulos, E., Koopmans, L. V. E., & Rusin, D. 2002, ApJ, 581, 823, astro-ph/0208420

Feng, J. L. 2010, ARA&A, 48, 495, 1003.0904

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, 1202.3665

Galan, A., Vernardos, G., Peel, A., Courbin, F., & Starck, J. L. 2022, A&A, 668, A155, 2207.05763

Gao, L., & White, S. D. M. 2006, MNRAS, 373, 65, astro-ph/0605687

Garrison-Kimmel, S. et al. 2017, MNRAS, 471, 1709, 1701.03792

Ghigna, S., Moore, B., Governato, F., Lake, G., Quinn, T., & Stadel, J. 1998, MNRAS, 300, 146, astro-ph/9801192

Gilman, D., Agnello, A., Treu, T., Keeton, C. R., & Nierenberg, A. M. 2017, MNRAS, 467, 3970, 1610.08525

Gilman, D., Benson, A., Bovy, J., Birrer, S., Treu, T., & Nierenberg, A. 2022, MNRAS, 512, 3163, 2112.03293

Gilman, D., Birrer, S., Nierenberg, A., & Oh, M. S. H. 2024, arXiv e-prints, arXiv:2403.03253, 2403.03253

Gilman, D., Birrer, S., Nierenberg, A., Treu, T., Du, X., & Benson, A. 2020, MNRAS, 491, 6077, 1908.06983

Gilman, D., Birrer, S., Treu, T., Keeton, C. R., & Nierenberg, A. 2018, MNRAS, 481, 819, 1712.04945

Gilman, D., Birrer, S., Treu, T., Nierenberg, A., & Benson, A. 2019, MNRAS, 487, 5721, 1901.11031

Gilman, D., Bovy, J., Treu, T., Nierenberg, A., Birrer, S., Benson, A., & Sameie, O. 2021, MNRAS, 507, 2432, 2105.05259

Gilman, D., Zhong, Y.-M., & Bovy, J. 2023, Phys. Rev. D, 107, 103008, 2207.13111

Gnedin, O. Y., & Zhao, H. 2002, MNRAS, 333, 299, astro-ph/0108108

Han, J. J., Semenov, V., Conroy, C., & Hernquist, L. 2023, The Astrophysical Journal Letters, 957, L24

Hao, C. N., Mao, S., Deng, Z. G., Xia, X. Y., & Wu, H. 2006, MNRAS, 370, 1339, astro-ph/0605319

Heinze, F. M., Despali, G., & Klessen, R. S. 2024, MNRAS, 527, 11996, 2311.13639

Herle, A., O'Riordan, C. M., & Vegetti, S. 2023, arXiv e-prints, arXiv:2307.10355, 2307.10355

Hsueh, J.-W., Despali, G., Vegetti, S., Xu, D., Fassnacht, C. D., & Metcalf, R. B. 2018, MNRAS, 475, 2438, 1707.07680

Hsueh, J. W., Enzi, W., Vegetti, S., Auger, M. W., Fassnacht, C. D., Despali, G., Koopmans, L. V. E., & McKean, J. P. 2020, MNRAS, 492, 3047, 1905.04182

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, 1601.01671

Hsueh, J. W. et al. 2017, MNRAS, 469, 3713, 1701.06575

Hunter, J. D. 2007, Computing in Science Engineering, 9, 90

Inoue, K. T., & Takahashi, R. 2012, MNRAS, 426, 2978, 1207.2139

Iršič, V. et al. 2017, Phys. Rev. D, 96, 023522

Ishiyama, T., & Ando, S. 2020, MNRAS, 492, 3662, 1907.03642

Ishiyama, T. et al. 2021, MNRAS, 506, 4210, 2007.14720

Jeans, J. H. 1922, MNRAS, 82, 122

Jethwa, P., Erkal, D., & Belokurov, V. 2018, MNRAS, 473, 2060, 1612.07834

Jungman, G., Kamionkowski, M., & Griest, K. 1996, Phys. Rep., 267, 195, hep-ph/9506380

Kaplinghat, M., Tulin, S., & Yu, H.-B. 2016, Phys. Rev. Lett., 116, 041302, 1508.03339

Keeton, C. R. 2001, arXiv e-prints, astro, astro-ph/0102340

——. 2011, GRAVLENS: Computational Methods for Gravitational Lensing, Astrophysics Source Code Library, record ascl:1102.003

Kim, S., Peter, A., & Hargis, J. 2019, in American Astronomical Society Meeting Abstracts, Vol. 233, American Astronomical Society Meeting Abstracts #233, 439.02

Kim, S.-H. et al. 2022, MNRAS, 510, 815, 2111.14025

Klypin, A., Gottlöber, S., Kravtsov, A. V., & Khokhlov, A. M. 1999, ApJ, 516, 530, astro-ph/9708191

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340, 1411.4001

Kochanek, C. S., & Dalal, N. 2004, ApJ, 610, 69, astro-ph/0302036

Kolb, E. W., & Turner, M. S. 1990, The early universe, Vol. 69

Koopmans, L. V. E. 2005, MNRAS, 363, 1136, astro-ph/0501324

Koopmans, L. V. E., & Fassnacht, C. D. 1999, ApJ, 527, 513, astro-ph/9907258

Lacey, C., & Cole, S. 1994, MNRAS, 271, 676, astro-ph/9402069

Lacey, C. G. et al. 2016, MNRAS, 462, 3854, 1509.08473

Lee, J. W., Lee, S.-S., Hodgson, J. A., Kim, D.-W., Algaba, J.-C., Kang, S., Kang, J., & Kim, S. S. 2017, ApJ, 841, 119, 1705.06835

Liao, S., Gao, L., Frenk, C. S., Guo, Q., & Wang, J. 2017, MNRAS, 470, 2262, 1610.07592

Lokas, E. L. 2022, A&A, 662, A53, 2204.01456

Lovell, M. R. 2020, ApJ, 897, 147, 2003.01125

Lovell, M. R. et al. 2016, MNRAS, 461, 60, 1511.04078

Lovell, M. R., Frenk, C. S., Eke, V. R., Jenkins, A., Gao, L., & Theuns, T. 2014, MNRAS, 439, 300, 1308.1399

Mao, S., & Schneider, P. 1998, MNRAS, 295, 587, astro-ph/9707187

McGaugh, S. S. 2015, Canadian Journal of Physics, 93, 250, 1404.7525

McKean, J. et al. 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14), 84, 1502.03362

McQuinn, M. 2016, ARA&A, 54, 313, 1512.00086

Meiksin, A. A. 2009, Rev. Mod. Phys., 81, 1943

Metcalf, R. B. 2005, ApJ, 622, 72, astro-ph/0407298

Metcalf, R. B., & Madau, P. 2001, ApJ, 563, 9, astro-ph/0108224

Moliné, Á. et al. 2023, MNRAS, 518, 157, 2110.02097

Möller, O., Hewett, P., & Blain, A. 2003, in The Mass of Galaxies at Low and High Redshift, ed. R. Bender & A. Renzini, 196

Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., & Tozzi, P. 1999, ApJ, 524, L19, astro-ph/9907411

Moustakas, L. A., & Metcalf, R. B. 2003, MNRAS, 339, 607, astro-ph/0206176

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, 1107.3140

Murgia, R. 2018, in Journal of Physics Conference Series, Vol. 956, Journal of Physics Conference Series, 012005, 1712.04810

Murgia, R., Iršič, V., & Viel, M. 2018, Phys. Rev. D, 98, 083540, 1806.08371

Myers, S. T. et al. 1995, ApJ, 447, L5

Nadler, E. O., et al. 2021, Phys. Rev. Lett., 126, 091101, 2008.00022

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493, astro-ph/9611107

Navarro, J. F. et al. 2010, MNRAS, 402, 21, 0810.1522

Nierenberg, A. M. et al. 2020a, MNRAS, 492, 5314, 1908.06344

——. 2020b, MNRAS, 492, 5314, 1908.06344

——. 2023, arXiv e-prints, arXiv:2309.10101, 2309.10101

——. 2017, MNRAS, 471, 2224, 1701.05188

Nierenberg, A. M., Treu, T., Menci, N., Lu, Y., & Wang, W. 2013, ApJ, 772, 146, 1302.3243

Nierenberg, A. M., Treu, T., Wright, S. A., Fassnacht, C. D., & Auger, M. W. 2014, MNRAS, 442, 2434, 1402.1496

Oort, J. H. 1932, Bull. Astron. Inst. Netherlands, 6, 249

O'Riordan, C. M., Despali, G., Vegetti, S., Lovell, M. R., & Moliné, Á. 2023, MNRAS, 521, 2342, 2211.15679

O'Riordan, C. M., & Vegetti, S. 2023, arXiv e-prints, arXiv:2310.10714, 2310.10714

Patnaik, A. R., Kemball, A. J., Porcas, R. W., & Garrett, M. A. 1999, MNRAS, 307, L1, astro-ph/9905311

Peebles, P. J. E. 1982, ApJ, 263, L1

Perlick, V. 2012, Gravitational Lensing (ZARM University of Bremen)

Phillips, P. M. et al. 2000, MNRAS, 319, L7, astro-ph/0009334

Planck Collaboration et al. 2016, A&A, 594, A13, 1502.01589

——. 2020, A&A, 641, A6, 1807.06209

Powell, D. M., Vegetti, S., McKean, J. P., Spingola, C., Stacey, H. R., & Fassnacht, C. D. 2022, MNRAS, 516, 1808, 2207.03375

Powell, D. M., Vegetti, S., McKean, J. P., White, S. D. M., Ferreira, E. G. M., May, S., & Spingola, C. 2023, MNRAS, 524, L84, 2302.10941

Prada, J., Forero-Romero, J. E., Grand, R. J. J., Pakmor, R., & Springel, V. 2019, MNRAS, 490, 4877, 1910.04045

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Read, J. I., Agertz, O., & Collins, M. L. M. 2016, MNRAS, 459, 2573, 1508.04143

Refregier, A. 2003, ARA&A, 41, 645, astro-ph/0307212

Ritondale, E., Vegetti, S., Despali, G., Auger, M. W., Koopmans, L. V. E., & McKean, J. P. 2019, MNRAS, 485, 2179, 1811.03627

Roberts, M. S., & Rots, A. H. 1973, A&A, 26, 483

Rogstad, D. H., & Shostak, G. S. 1972, ApJ, 176, 315

Rubin, V. C., Ford, W. K., J., & Thonnard, N. 1978, ApJ, 225, L107

Rubin, V. C., & Ford, W. Kent, J. 1970, ApJ, 159, 379

Schaye, J. et al. 2015, MNRAS, 446, 521, 1407.7040

Schneider, A., Smith, R. E., Macciò, A. V., & Moore, B. 2012, MNRAS, 424, 684, 1112.0330

Schneider, A., Smith, R. E., & Reed, D. 2013, MNRAS, 433, 1573, 1303.0839

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1, astro-ph/9907024

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119, astro-ph/9901122

Spergel, D. N., & Steinhardt, P. J. 2000, Phys. Rev. Lett., 84, 3760, astro-ph/9909386

Springel, V. et al. 2005, Nature, 435, 629, astro-ph/0504097

Stacey, H. 2024

Stacey, H. R., Lafontaine, A., & McKean, J. P. 2020, MNRAS, 493, 5290, 2002.07186

Sugai, H., Kawai, A., Shimono, A., Hattori, T., Kosugi, G., Kashikawa, N., Inoue, K. T., & Chiba, M. 2007, ApJ, 660, 1016, astro-ph/0702392

Tessore, N., & Metcalf, R. B. 2015, A&A, 580, A79, 1507.01819

The Astropy Collaboration et al. 2018, AJ, 156, 123

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science Engineering, 13, 22

Vegetti, S., & Koopmans, L. V. E. 2009, MNRAS, 392, 945, 0805.0201

Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, MNRAS, 408, 1969, 0910.0760

Vegetti, S., Lagattuta, D. J., McKean, J. P., Auger, M. W., Fassnacht, C. D., & Koopmans, L. V. E. 2012, Nature, 481, 341, 1201.3643

Vegetti, S., & Vogelsberger, M. 2014, MNRAS, 442, 3598, 1406.1170

Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, Phys. Rev. D, 88, 043502

Viel, M., Lesgourgues, J., Haehnelt, M. G., Matarrese, S., & Riotto, A. 2005, Phys. Rev. D, 71, 063534, astro-ph/0501562

Vogelsberger, M. et al. 2014, MNRAS, 444, 1518, 1405.2921

Wambsganss, J. 1998, Living Reviews in Relativity, 1, 12

Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V., & Dekel, A. 2002, ApJ, 568, 52, astro-ph/0108151

Wetzel, A. R., Hopkins, P. F., Kim, J.-h., Faucher-Giguère, C.-A., Kereš, D., & Quataert, E. 2016, ApJ, 827, L23, 1602.05957

Xu, D., Sluse, D., Gao, L., Wang, J., Frenk, C., Mao, S., Schneider, P., & Springel, V. 2015, MNRAS, 447, 3189, 1410.3282

Yuan, S., Hadzhiyska, B., Bose, S., & Eisenstein, D. J. 2022, MNRAS, 512, 5793, 2202.12911

Zwicky, F. 1937, ApJ, 86, 217

Appendix A

Individual Lens Inferences

The grid on the bottom left shows the two-dimensional posterior probability distribution of $M_{\rm hm}$ and $f_{\rm sub}$ recovered from each of the individual lenses in our sample that were not discussed in Section 2.4.2. The likelihood for each grid point is calculated using both the image positions and fluxes. The top panel shows the result of marginalizing over $f_{\rm sub}$ to get the posterior of $M_{\rm hm}$, and the right panel shows the result of marginalizing over $M_{\rm hm}$ to get the posterior of $f_{\rm sub}$.

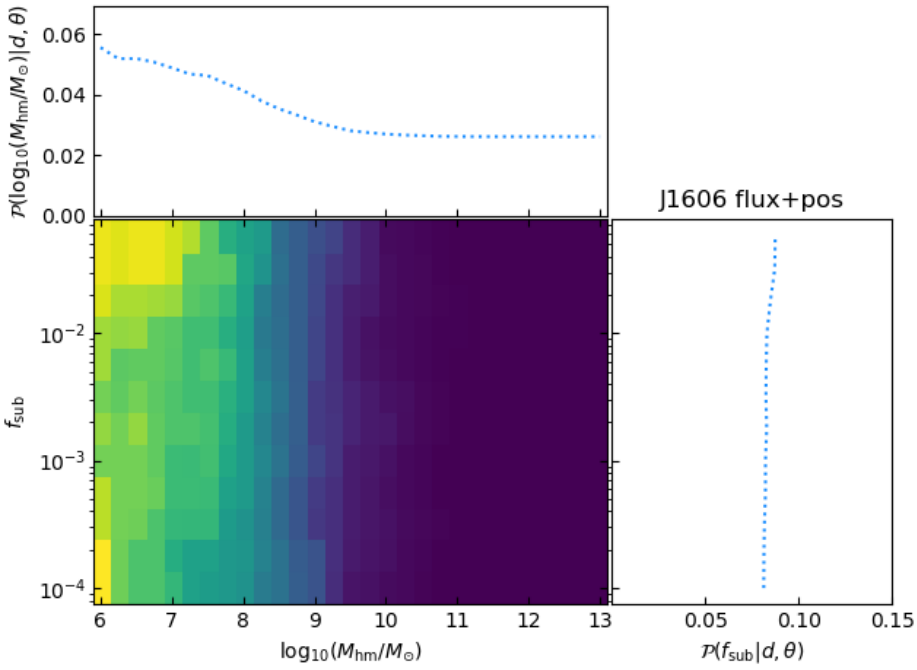


Figure A.1 PS J1606-2333 (NLR)

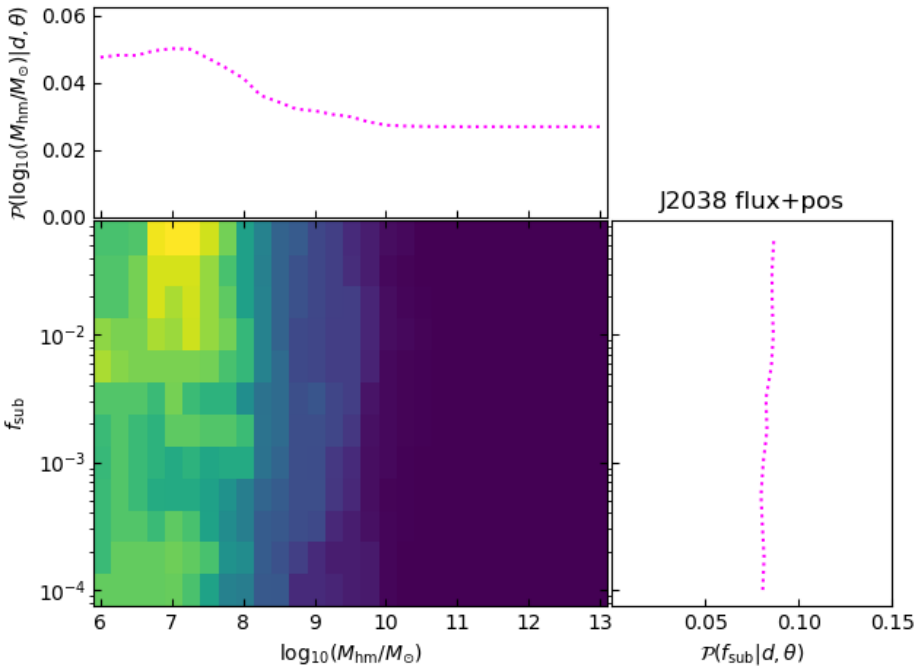


Figure A.2 WGD J2038-4008 (NLR)

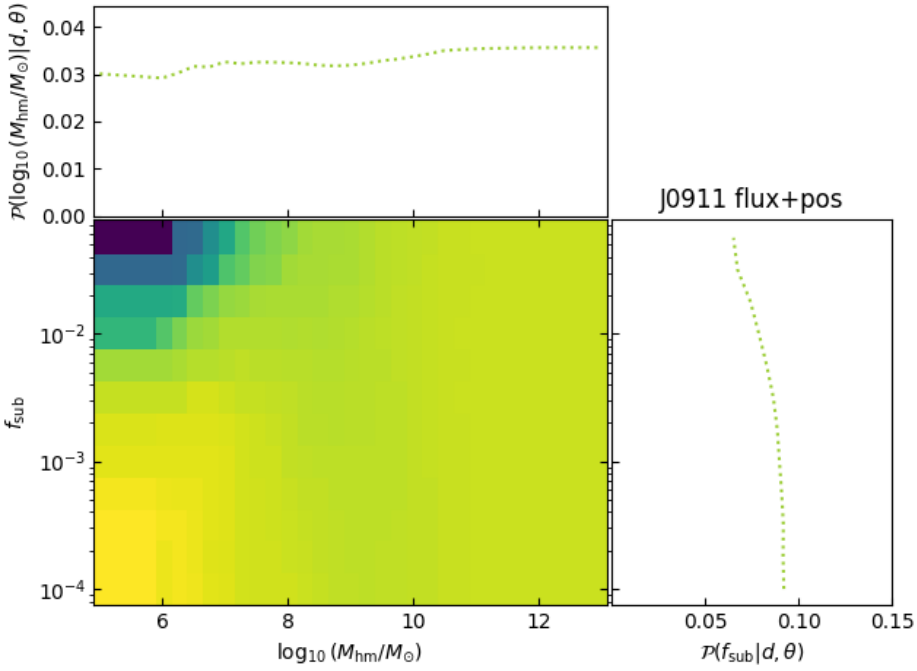


Figure A.3 RX J0911+0551 (NLR)

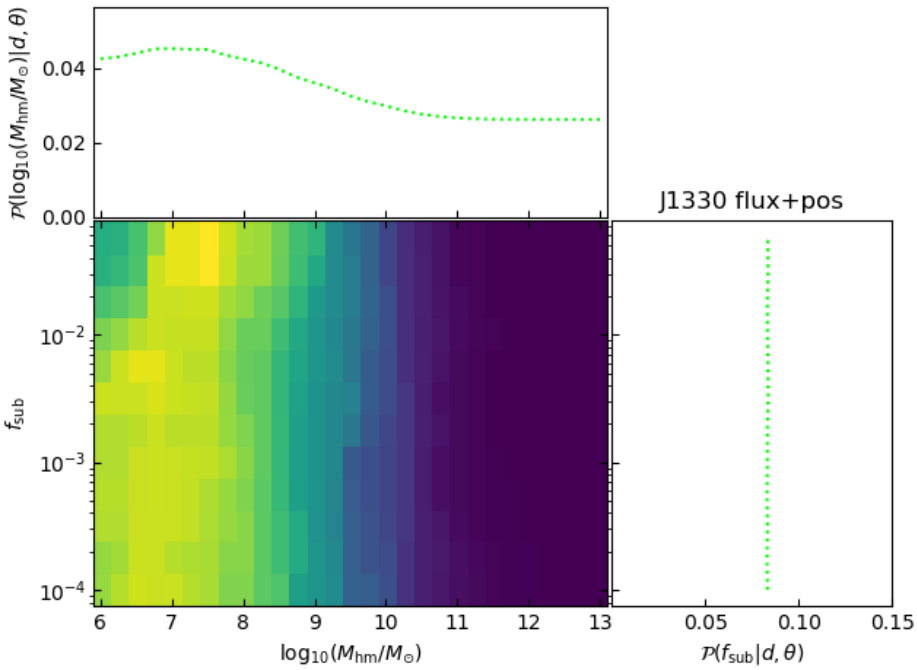


Figure A.4 SDSS J1330+1810 (NLR)

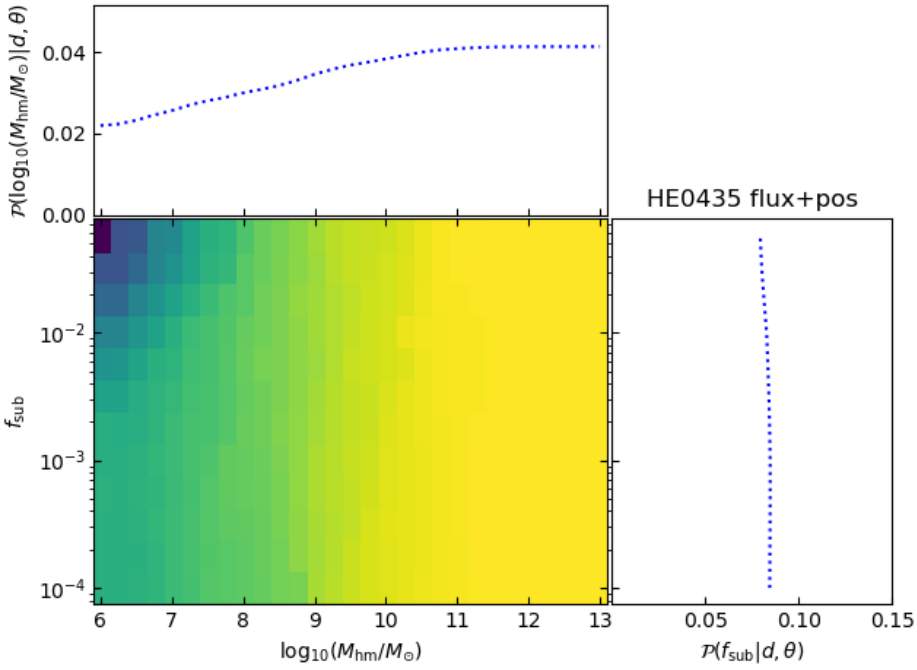


Figure A.5 HE0435-1223 (NLR)

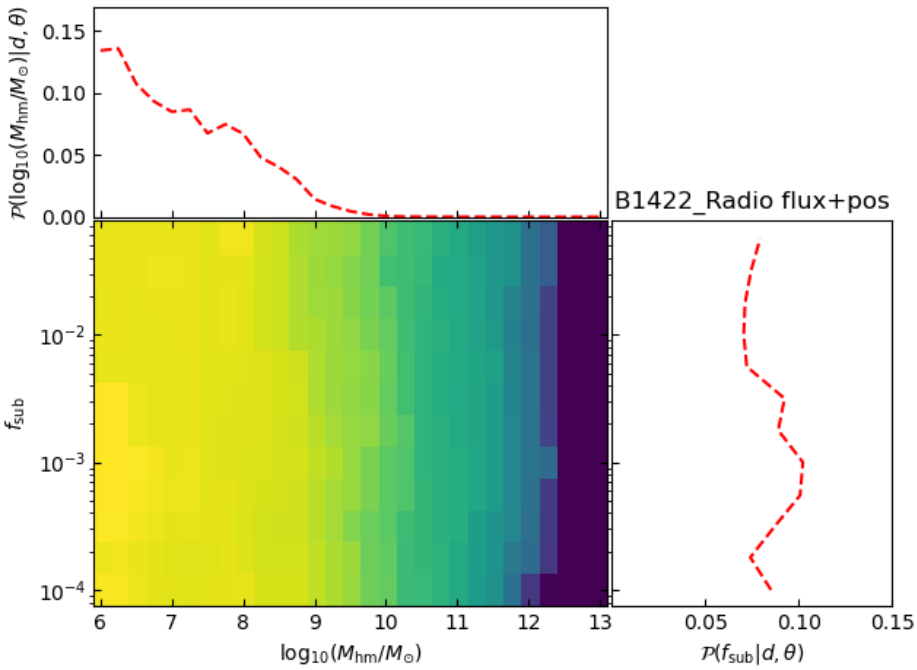


Figure A.6 JVAS B1422+231 (Radio/MIR)

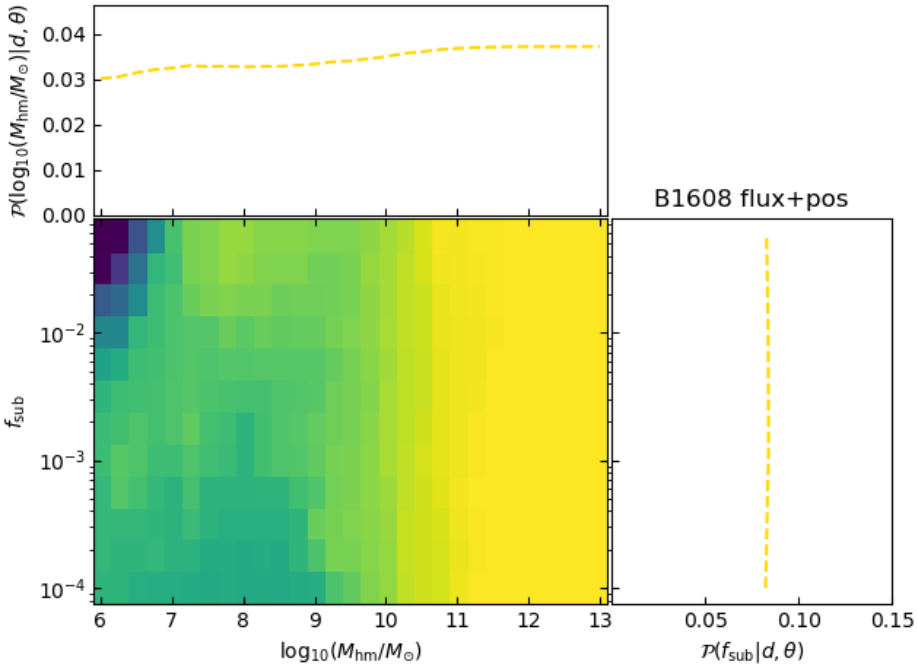


Figure A.7 CLASS B0128+437 (Radio/MIR)

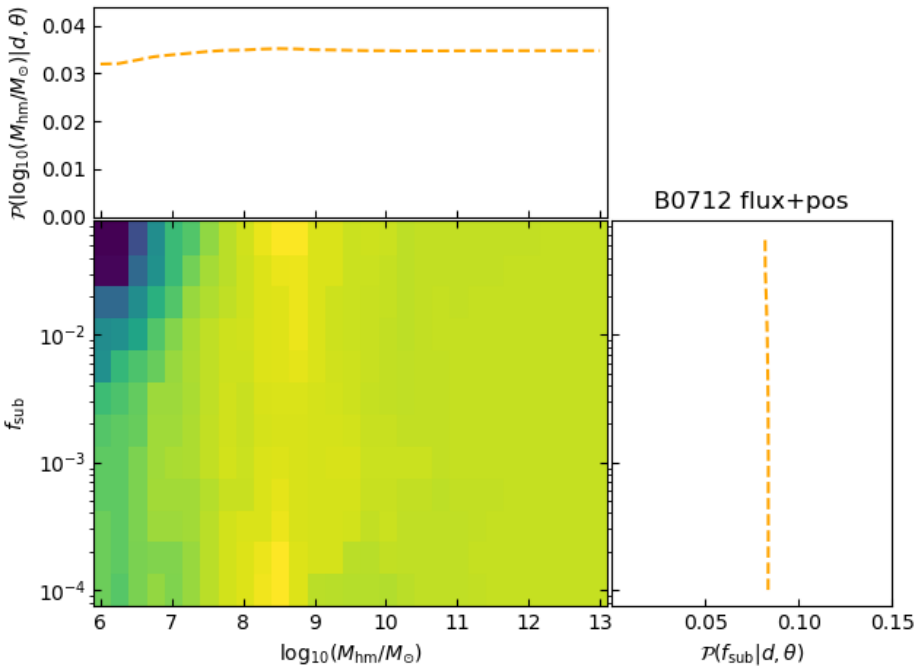


Figure A.8 CLASS B0712+472 (Radio/MIR)

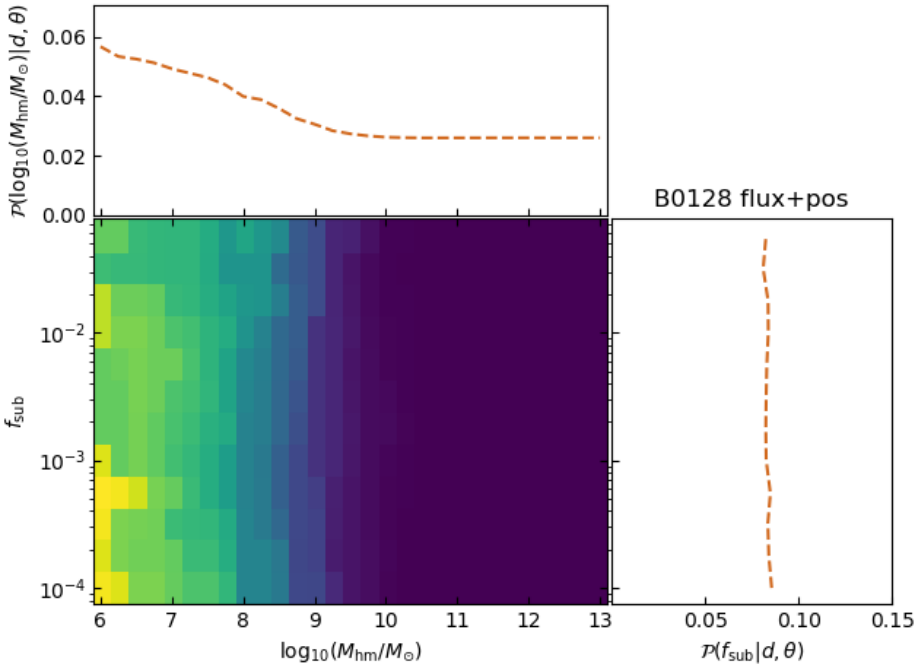


Figure A.9 CLASS B0128+437 (Radio/MIR)

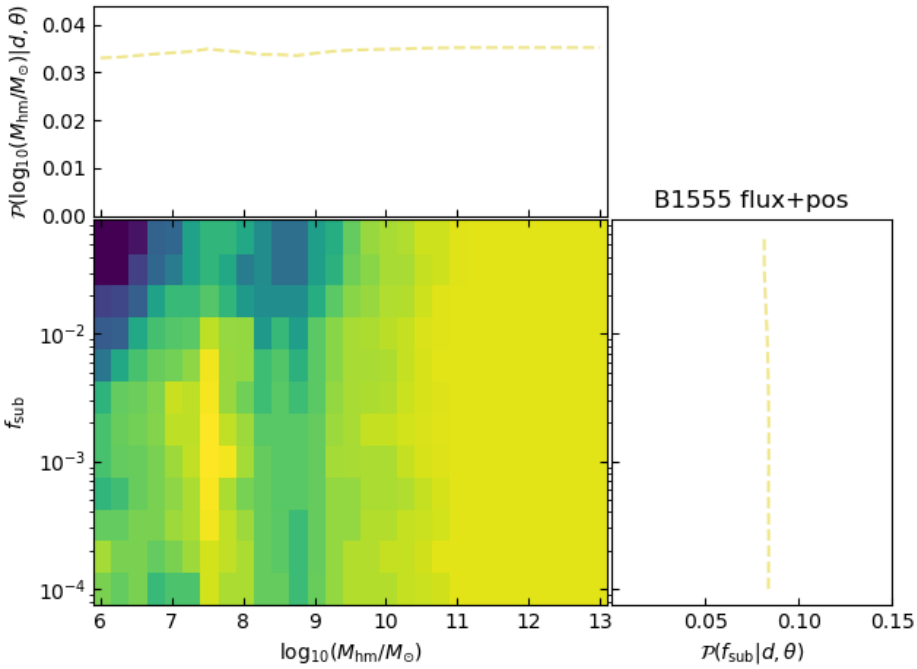


Figure A.10 CLASS B1555+375 (Radio/MIR)