Title

Black Holes: Finding Them, Building Instruments that Help You Look, and how Systemic Racism can Prevent You From Doing Any of That

Author

Abdurrahman, Fatima Naeem

Publication Date

2021

Peer reviewed|Thesis/dissertation

Black Holes: Finding Them, Building Instruments that Help You Look, and how Systemic Racism can Prevent You From Doing Any of That

by

Fatima N. Abdurrahman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Astrophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Daniel R. Weisz, Chair
Professor Gibor Basri
Professor Kris D. Gutiérrez

Spring 2021

Black Holes: Finding Them, Building Instruments that Help You Look, and how Systemic Racism can Prevent You From Doing Any of That

Abstract

Black Holes: Finding Them, Building Instruments that Help You Look, and how Systemic Racism can Prevent You From Doing Any of That

by

Fatima N. Abdurrahman

Doctor of Philosophy in Astrophysics

University of California, Berkeley

Professor Daniel R. Weisz, Chair

This work aims to connect the disparate elements of 'doing astronomy' involved in the production of knowledge, highlighting the advantages of an interdisciplinary approach and emphasizing the importance of each component to the overall success of the field. Broadly dividing the work into ideas, tools, and people, each area will be exemplified to play a critical role in the pursuit of a scientific result. The science case used to illustrate each components' contribution will be the detection of black holes in our galaxy. First, a method of black hole detection is demonstrated, relying on the theory of relativity and astronomical observations (Chapter 2). Second, a technology that makes such observations possible is described, with a novel variation making improvements on former versions (Chapter 3). Finally, social dynamics that limit the entry and progression (and by extension, the potential scientific contributions in the aforementioned areas) of minoritized groups in the field are analyzed in the context of graduate programs (Chapter 4).

Ideas: Theory and Observations

Though stellar-mass black holes (BHs) are likely abundant in the Milky Way (N= 10^8–10^9), only 20 have been detected to date, all in accreting binary systems [1]. Gravitational microlensing is a proposed technique to search for *isolated* BHs, which to date have not been detected. Two microlensing events, MACHO-1996-BLG-5 (M96-B5) and MACHO-1998-BLG-6 (M98-B6), initially observed near the lens-source minimum angular separation in 1996 and 1998, respectively, have long Einstein crossing times (> 300 days), identifying the lenses as candidate black holes. Twenty years have elapsed since the time of lens-source closest approach for each of these events, indicating that if the lens and source are both luminous, and if their relative proper motion is sufficiently large, the two components should be spatially resolvable. We attempt to eliminate the possibility of a stellar lens for these

events by: (1) using Keck near-infrared adaptive optics images to search for a potentially now-resolved, luminous lens; and (2) examining multi-band photometry of the source to search for flux contributions from a potentially unresolved, luminous lens.

We combine detection limits from NIRC2 images with light curve data to eliminate all non-BH lenses for relative lens-source proper motions above 0.81 mas/yr for M96-B5 and 2.48 mas/yr for M98-B6. Further, we use WFPC2 broadband images to eliminate the possibility of stellar lenses at any proper motion. We present the narrow range of non-BH possibilities allowed by our varied analyses. Finally, we suggest future observations that would constrain the remaining parameter space with the methods developed in this work.

Tools: Optics and Instrumentation

'Imaka is a ground layer adaptive optics (GLAO) demonstrator on the University of Hawaii 2.2m telescope with a 24×18 field-of-view, nearly an order of magnitude larger than previous AO instruments. In 15 nights of observing with natural guide star asterisms ~ 16 in diameter, we measure median AO-off and AO-on empirical full-widths at half-maximum (FWHM) of 095 and 064 in R-band, 081 and 048 in I-band, and 076 and 044 at 1 micron. This factor of 1.5-1.7 reduction in the size of the point spread function (PSF) results from correcting both the atmosphere and telescope tracking errors. The AO-on PSF is uniform out to field positions ~ 5 off-axis, with a typical standard deviation in the FWHM of 0018. Images exhibit variation in FWMM by 4.5% across the field, which has been applied as a correction to the aforementioned quantities. The AO-on PSF is also $10\times$ more stable in time compared to the AO-off PSF.

In comparing the delivered image quality to proxy measurements, we find that in both AO-off and AO-on data, delivered image quality is correlated with 'imaka's telemetry, with R-band correlation coefficients of 0.68 and 0.70, respectively. At the same wavelength, the data are correlated to DIMM and MASS seeing with coefficients of 0.45 and 0.55. Our results are an essential first step to implementing facility-class, wide-field GLAO on Maunakea telescopes, enabling new opportunities to study extended astronomical sources, such as deep galaxy fields, nearby galaxies or star clusters, at high angular resolution.

People: Society and Culture

Despite decades of interest in promoting diversity and inclusion in the field and higher education more broadly, physics and astronomy departments in American universities remain predominantly white and male. While some STEM fields have made progress in recent years towards correcting the historic overrepresentation of white men, 75% of physics and astronomy PhDs awarded in 2019 went to this demographic that constitutes only 30% of the US population, higher than chemistry, math, computer science, and engineering [2]. This

work seeks to understand the barriers to creating a more diverse field by examining how its cultural beliefs can work to maintain inequity in graduate programs.

We use the lived experiences of 12 female and gender-non conforming students of color to construct a model outlining how physics and astronomy graduate programs in American predominantly white institutions maintain equity gaps between majoritized and minoritized students. This model connects participant's observations of attitudes, policies, and behaviors that relate to two foundational cultural beliefs in Western physics and astronomy: 1) that the field functions as a meritocracy where only the fittest survive, and 2) that physicists comprise an objective, cultureless, and apolitical community, impervious to social influence. We contextualize our findings with literature to demonstrate the negative outcomes such systems create for minoritized students, including alienation from their departments, mental health problems, and lower retention in the field. In outlining this system, we emphasize the design of program structures as a critical point of intervention. To this end, we conclude the work with recommendations based on participant input.

To Kara Kundert and Jordan Fleming:

Y'all should've been here with me.

Contents

List of Figures

List of Tables

Acknowledgments

The majority of this work was completed at U.C. Berkeley, which sits on the territory of xučyun, the ancestral and unceded land of the Chochenyo speaking Ohlone people, the successors of the historic and sovereign Verona Band of Alameda County. I acknowledge that I have benefited and continue to benefit from the use and occupation of this land. I also recognize the importance of taking actions to support the rematriation of indigenous land, and pledge to take and continue action in support of American Indian and Indigenous peoples, starting with financial contributions to the Shuumi Land Tax for the duration of my graduate career.

Additionally, many of the data in this thesis were obtained through telescopes on the summit of Mauna Kea on the island of Hawaii. I wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community, as well as the complex history and conflict with the astronomical community. I have been most fortunate to have the opportunity to conduct observations from this mountain, and deeply support indigenous Hawaiian efforts towards decolonization.

For their guidance and encouragement for Chapter 2 of this work, I want to thank Calen Henderson, Gabriela Navarro Ovando, and Etienne Bachelet. For welcoming me with support and enthusiasm to the project in Chapter 3, I want to thank the ʻimaka team, led by Mark Chun. For making any of Chapter 4 possible, I want to thank Alice Olmstead, who has bookended my career by being the TA of my first astronomy class ever, getting me involved in physics education research as an undergraduate, inspiring me to pave my own way when grad school proved difficult, and finally advising me on the final chapter of my doctoral thesis.

More personally, I want to thank those who have gotten me through the past five years despite all odds. From my program, I'd like to thank Gerry Zhang, who helped me come to terms with my disappointment in the field; Wren Suess, my constant ally through all the barriers I've faced in grad school; Chris Moeckel, for making me smile when all I wanted to do was cry; and Elliana Abrahams, who's been my greatest cheerleader, advisor, and confidant since she showed up in our department three years ago. Outside of school, I want to thank my therapist Eryn Reeder, who helped me find the strength to grow through one of the most difficult times of my life; the Bay Area dance and aerial community, a much needed refuge from science in which I found friendship, solace, and joy; and my endlessly supportive friend Rosalind Goodwin, who always saw me as an astronomer, even when I couldn't see it myself.

Finally, I want to thank my dad, who showed me that 'physicist' and 'revolutionary' are not mutually exclusive terms.

Chapter 1

Introduction

1.1 A Multi-Disciplinary Approach to 'Doing Astronomy'

Observing the contradiction between the general understanding of physics as an interdisciplinary field and the growing drive towards hyper-specialization, science philosopher Terrance Quinn asks his readers, "What then, is progress in physics?". He considers the more traditionally understood areas of physics (theory, application) along with cross-scientific areas (biophysics, physical chemistry) and humanities-adjacent areas (the history and philosophy of physics, physics education) to lack unity both in goals and practice, calling for a more "functional collaboration" in the field [3]. This work exemplifies a multi-disciplinary approach to astronomical research in a similar vein, weaving together observational astronomy, instrumentation, and astronomy education to pursue a single science case. Attempting to answer Quinn's query, I demonstrate how 'progress' in the field relies on the success of all its constituent components, despite the growing tendency to focus on only a narrow subset.

The science case this work centers on is the detection of black holes in our galaxy. Chapter 2 presents an attempt at characterizing two heretofore unidentified objects as black holes through the phenomenon of gravitational microlensing, with the use of existing theory and astronomical observations. While Einstein (whose theory of relativity predicted gravitational lensing) believed that such a phenomenon would be impossible to observe, developments in astronomical imaging in the past few decades have rendered its observation trivial, with hundreds to thousands of microlensing events detected each year [4]. One such technology was used for the primary observations in this analysis: an adaptive optics (AO) instrument, which corrects the distorting effects of Earth's atmosphere on ground-based telescopes. Chapter 3 presents the results of a novel experiment using ground-layer adaptive optics, in which the author gained an expertise of AO systems that proved invaluable for the data collection described in Chapter 2. Finally, Chapter 4 is a work of socially-focused education research that outlines how cultural forces limit the entry and progression (and, therefore, the potential scientific contributions) of minoritized groups in the field. While this type of analysis is gen-

erally helpful in creating a more robust and egalitarian field, the understanding of systems it enables has also proven critical in my own navigation of graduate school, empowering me to complete the previously mentioned works despite various barriers. The following sections introduce each of these three elements in greater detail.

1.2 Searching for Black Holes with Gravitational Microlensing

Detecting isolated black holes (BHs) remains a problem of both great importance and difficulty in astrophysics. Core-collapse supernova events mark the deaths of high-mass (8) stars and are predicted to leave remnant BHs on the order of several to tens of $M_{\odot}$. An estimated $10^8 - 10^9$ stellar-mass black holes are predicted to occupy the Milky Way [5]. However, only ∼20 have been detected, all in accreting binaries [6, 7, 8]. More recently, BH binaries have been identified by gravitational waves emitted from their mergers [9]. Isolated BHs, which could comprise the majority of the BH population [10], remain elusive, with no confirmed detections to date. Detecting isolated BHs and measuring their masses would help constrain the number density of BHs as well as the initial-final mass relation–which designates which stars become BHs versus neutron stars or white dwarves–in turn informing the understanding of BH formation, supernova physics, and the equation of state for nuclear matter [11].

While an isolated BH does not produce a detectable electromagnetic radiation signature, it is in principle detectable as a lens in a gravitational microlensing event, whereby the gravitational potential of a massive lensing object refracts and focuses the light of a background source in accordance with their time-dependent angular separation on the plane of the sky [12, 13, 14]. One characterizing feature of such an event is the Einstein radius $\theta_{\rm E}$, which describes the radius of the source image given the physical alignment of an event's source, lens, and observer, and is defined as:

$$\theta_{\rm E} = \sqrt{\frac{4GM}{c^2}(d_L^{-1} - d_S^{-1})} \tag{1.1}$$

where G is the gravitational constant, M is the lens mass, c is the speed of light, d_L is the observer-lens distance and d_S is the observer-source distance. More readily recoverable from a microlensing light curve, however, is the Einstein crossing time $t_{\rm E}$, which is related to an event's Einstein radius and relative source-lens proper motion μ_{rel} by:

$$\vec{\theta_{\rm E}} = \vec{\mu_{\rm rel}} t_{\rm E} \tag{1.2}$$

While $t_{\rm E}$ encodes all of the physical parameters of an event (namely, lens mass and lens and source distances,) in the absence of additional signals such as finite source effects, astrometric shifts, or parallax, one is limited to applying a Bayesian prior from a Galactic model and estimating (rather than measuring) the lens mass. A frequently modeled second

order effect is parallax, which causes an asymmetric distortion to a point-source point-lens (PSPL) microlensing light curve. In microlensing formalism, the 'microlensing parallax', $\pi_{\rm E}$, is defined as:

$$\vec{\pi_{\rm E}} = \frac{1\,{\rm AU}(d_L^{-1} - d_S^{-1})}{\theta_{\rm E}}\hat{\theta_{\rm E}} \tag{1.3}$$

The microlensing parallax can be understood as the ratio between the Earth's orbit and the Einstein radius of the microlensing event, projected onto the observer plane.

Currently, several surveys monitor many tens of square degrees near the Galactic Bulge in search of the photometric variability characteristic of microlensing events — the fourth phase of the Optical Gravitational Lensing Experiment (OGLE-IV; [15], [16]), the Microlensing Observations in Astrophysics collaboration (MOA-II; [17], [18]), and the Korea Microlensing Telescope Network (KMTNet; [19]). Historically, the MACHO Project[1] used microlening to search specifically for MAssive Compact Halo Objects (MACHOs), a hypothesized form of dark matter in the Milky Way halo that could include BHs [20]. Previous works have considered events from the MACHO survey as well as the OGLE survey for the possibility of BH lenses [21, 22], though they have largely been limited to the examination of microlensing light curves. In Chapter 2, we will similarly examine the light curves of two candidate BH events discovered in the MACHO survey, with the addition of more recent data from several different telescopes.

1.3 Improving Image Quality through Ground Layer Adaptive Optics

Adaptive optics (AO) is a powerful tool for correcting distortions caused by Earth's atmosphere and delivering diffraction-limited images. Simply put, AO instruments measure atmospheric distortion in real time while applying a correction for these distortions to images with the use of a deformable mirror (DM). This process is shown in schematic form with more detail in Figure 1.1. However, despite the advantages they present, most AO systems have a relatively small field-of-view (FoV), which can limit the science applications. For first generation, "classical", single-conjugate adaptive optics (SCAO), which use a single guide star and a single deformable mirror (DM) to correct for turbulence at all heights in a single direction, the correction is limited to regions close to the guide star, resulting in diffraction-limited images spanning only a few arcseconds at optical wavelengths and tens of arcseconds in the near-infrared. Because of this, extended sources such as nearby galaxies or nearby star clusters cannot easily be observed in their entirety with SCAO.

The second generation of AO seeks to mitigate the limitations of SCAO systems in a variety of ways. Multi-conjugate adaptive optics (MCAO) systems use multiple DMs to

[1] A full description of the MACHO Project can be found at `http://wwwmacho.anu.edu.au/Project/Overview/status.html`

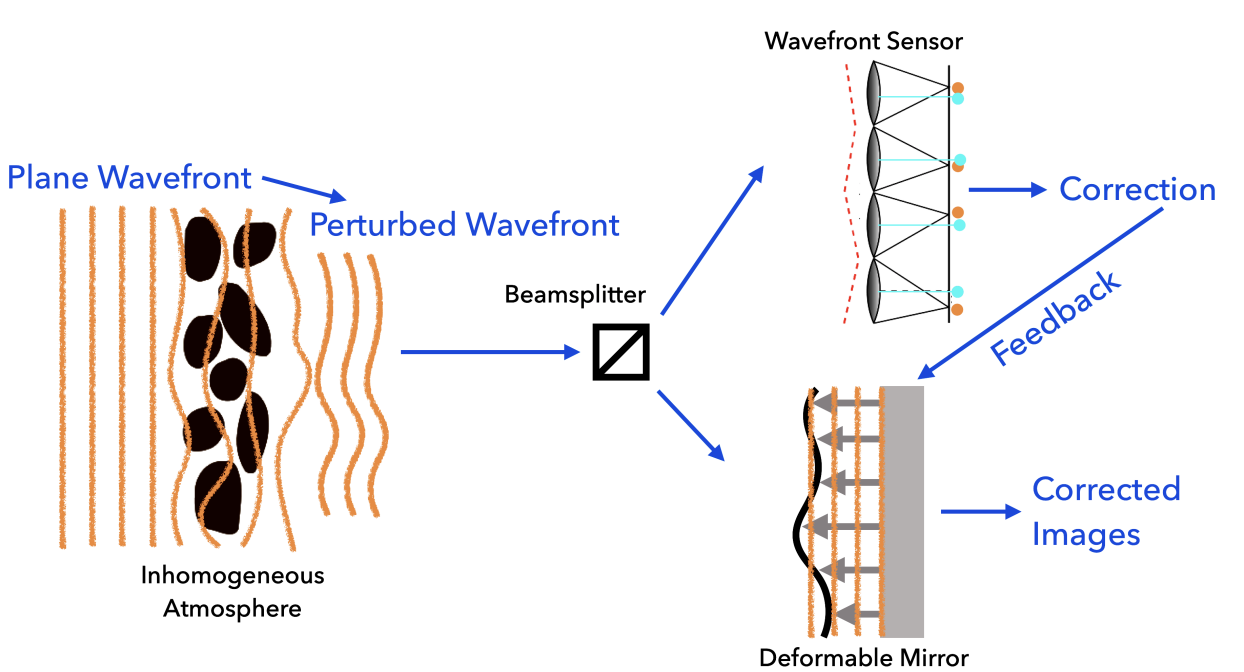


Figure 1.1: A schematic of basic AO operations: a wavefront perturbed by the Earth's atmosphere enters the instrument and is divided by a beamsplitter. One component of the beam is fed into a wavefront sensor, the simplest form of which is an array of lenslets that project and measure each wavefront component's deviation from planar. Those deviations inform the wavefront correction at that moment, which is then fed into a deformable mirror. The mirror arranges itself such that when the other beam of the perturbed wavefront is reflected off of it, the wavefront is restored to an approximately planar form. This corrected wavefront is then recorded by the camera, resulting in an atmospheric-corrected image.

correct atmospheric turbulence in more than one dimension. Correction on fields of 1'-2' with angular resolutions of $80 - 150$ milliarcseconds (mas) have been achieved by MCAO instruments such as GeMS on Gemini [23] and MAD on the VLT [24]. Multi-object adaptive optics (MOAO) systems that use tomographic reconstruction of several widely separated guide stars (3' - 5') to correct small sub-fields of a few arcseconds in size have achieved image resolutions of $\sim$150 mas. On-sky demonstrations of MOAO include Raven [25], Canary [26] and the future MOSAIC instrument, which includes a multi-object spectrograph with MOAO correction [27].

Ground layer adaptive optics (GLAO), first proposed by [28] to improve wide-field imaging for large telescopes, attempts to stretch the limits of AO even further. GLAO uses several guide stars to correct the shared ground layer turbulence of a large field of view, rather than along a single path through the atmosphere, at the expense of not correcting turbulence at

higher altitudes. Previous GLAO demonstrations have shown successful image correction for fields of 1' - 4' with resolutions of ∼300 mas, a factor of 2 improvement over the seeing [29, 30, 31, 32]. SAM on the SOAR telescope is a facility GLAO instrument, achieving free-atmosphere limited images in I-band with a resolution of ∼500 mas over a 3' FoV [33]. Similarly, ARGOS on the LBT is expected to improve spatial resolution by a factor of two across a 2' FoV [34]. ESO's Adaptive Optics Facility has implemented multiple systems with GLAO correction on VLT telescopes, including GRAAL (7'x7' in J, H, and K bands) and the wide field mode on GALACSI-MUSE (1'x1' in the optical regime) [35]. The potential for even larger fields exists for instruments such as ULTIMATE-SUBARU, which expects uniform point spread functions (PSFs) with GLAO correction across a field of 20' [36]. In Chapter 3, we present the on-sky results of a novel GLAO system with a 24'$times$18' FoV, 'imaka.

1.4 The Problem of Inequity in Astronomy Graduate Programs

After decades of interest in promoting diversity and inclusion in the field, physics and astronomy departments in American universities remain overwhelmingly white and male, with 75% of PhDs awarded in 2019 going white men, a demographic that constitutes only 30% of the national population. While STEM fields in general have faced challenges in diversifying their students and workforces, physics and astronomy have had a uniquely difficult time achieving this goal, awarding lower percentages of bachelors, masters, and PhDs to underrepresented minorities than biology, chemistry, engineering, mathematics, or computer science. Despite sometimes boasting a reputation of being 'not as bad' as physics, astronomy has the lowest percentage of underrepresented minorities at all levels of higher education of all the STEM fields. Unsurprisingly, both fields also remain far from achieving gender parity at any level [2].

As the focus on these equity gaps increase, some discourse within higher education has sought to explain away the potential effect of discriminatory pedagogy. A notable example is a Harvard University president who, in his remarks at a 2005 meeting of the National Bureau of Economic Research intended to address the underrepresentation of women and racial minorities, posited that women lack the dedication and "intrinsic aptitude" necessary of elite scientists [37]. A less cynical explanation centers inequity in K-12 and undergraduate education as cause for an unprepared pool of graduate applicants of minotized identities, as wealthy whites are more likely to attend well-resourced schools and subsequently gain material and symbolic advantages for applying to graduate school [38]. Others rely on trends in demographics' education pathways, effectively interpreting the lack of minoritized students in the field as a lack of interest on their part [39].

While these proposed explanations vary in their ability to address causes of inequity, none of them account for the low retention of minoritized students within higher education.

The percentage of underrepresented minority (UMR) students earning astronomy degrees decreases from 9% at the undergraduate level to 6% at the masters level and an abysmal 4% at the doctoral level, with a similar (though less exaggerated) trend in physics [2]. There also exist substantial inequities in graduate student well being, with racially minoritized, female, and LGBTQ graduate students showing significantly higher rates of depression and anxiety than their white, male, heterosexual and cisgender counterparts [40, 41]. Evidence suggests that these variations in outcomes do not trace variations in ability, as women tend to simultaneously achieve higher grades than men, while being more likely to leave PhD programs than men [41]. Finally, these explanations do nothing to address the ubiquity of discrimination and harassment within programs, with URM women in astronomy graduate programs being 15-20 times more likely to report such experiences than non-URM men [42].

A multitude of reports on representation in science have pointed to graduate school as a critical point of intervention for diversifying the field [43, 44, 45], recognizing the role it plays in shaping the next generation of educators and leaders. Increasingly, culture has been identified as a force maintaining homogeneity within graduate programs and the fields of physics and astronomy more generally [46], with 'culture' understood as a system of goals, values, beliefs, language, and behaviors that members of a group maintain and pass on to subsequent generations [47]. Members of a group who align with the dominant group culture can have difficulty observing it, comparable to a fish taking for granted the water it lives within. However, an inability to see the effects of culture makes it difficult to distinguish between what is necessary or beneficial to a community and what is simply tradition gone unchallenged– as Posselt puts it, "As is the case for a fish adapted to mildly toxic waters, those who stay in academia long enough may stop noticing the toxicity of the culture in which we work" [46]. To this end, Chapter 4 presents an analysis of inequity in astronomy through the lens of culture, using the lived experience of multiply minoritized graduate students to outline how cultural beliefs manifest themselves in graduate programs.

Chapter 2

On the Possibility of Stellar Lenses in Black Hole Candidate Microlensing Events

2.1 Introduction

MACHO-96-BLG-5 and MACHO-98-BLG-6 (hereafter referred to as M96-B5 and M98-B6, respectively) are microlensing events first detected and observed using long-term photometric monitoring [48]. M96-B5 and M98-B6 were two of forty-five candidates detected toward the Galactic bulge. M96-B5 was previously identified in 1996 as a BH candidate, due to its exceptionally long Einstein crossing time of ∼970 days. For this event, the mass of the lens was estimated to be $M = 6^{+10}_{-3}$, and its Heliocentric distance within the range of 0.5–2 kpc [49]. However, because these estimations were derived from fitting light curves alone–a process that necessarily exhibits degeneracies–they are heavily influenced by the imposed Galactic model prior. Another analysis of this event strongly excluded the possibility of the lens being a main-sequence star because of brightness constraints given by Hubble Space Telescope data, and their mass constraints ruled out the possibility of the lens being a neutron star [48]. The event was found by [50] to be inconsistent with a microlensing event occurring in a galaxy that did not include stellar remnant populations, based on simulations of microlensing light curves created from Galactic models that did not include neutron stars, white dwarfs, or black holes. M96-B5 was found to be measurably affected by microlensing parallax [48], and additionally exhibits minor perturbations due to xallarap [51], which refers to the accelerated motion of the source due to its binary companion, and which can complicate the determination of the source angular position at the time of the event [52]. M96-B5 was designated to be a marginal BH candidate, with a determined 37% likelihood of being a black hole [51].

The second candidate in this work, M98-B6, is a microlensing event identified in 1998 as a possible BH. The mass of the BH candidate for the event is estimated to be $M = 6^{+7}_{-3}$

Table 2.1: NIRC2 Observations

Target	RA (J2000)	Dec (J2000)	Epoch (UT)	N_{exp}	t_{int} (s)	C	N_*	Strehl	FWHM (mas)	m_{base}
M96-B5	18:05:02.5	-27:42:17	2016-07-14	9	60	1	297	0.22	70	16.864±0.035
			2017-05-21	5	5	6	130	0.21	67	16.909±0.032
M98-B6	17:57:32.8	-28:42:45	2016-07-14	22	10	1	344	0.33	58	12.895±0.033
			2017-06-08	15	2	1	74	0.19	82	12.868±0.035

Number of exposures N_{exp}, integration time t_{int}, coadds C, and number of stars N_* for each observation of both targets. Strehl and full-width half-max (FWHM) are the average values of all N_* stars over all N_{exp} individual exposures. Baseline (unlensed) magnitude m_{base} indicates the target magnitude in NIRC2's 'Kp' filter.

[49], though again, this result was dependent on Galactic models. The source star has a heliocentric radial velocity of -39 ± 20 km/s and is classified as a G5 IV spectral type [53]. [48] were not able to strongly exclude the possibility of the lens being either a main-sequence star or a neutron star from their constraints, as both possibilities still have small likelihoods due to the confidence levels used. The event is consistent with the simulations of [50], as described in the previous paragraph, which would make the possibility of a main-sequence lens seem more likely. M98-B6 is also affected by parallax, with minor perturbations from xallarap [51]. It is considered a weak BH candidate (2.2% likelihood) [51].

The chapter is structured as follows: In Section 2.2 we describe the observations and data sets that went into this study. In Section 2.3 we lay out the reduction processes used on the relevant data sets. In Section 2.4 we examine the detection limits for a luminous lens in recent, near-infrared (NIR) images. In Section 2.5 we fit microlensing light curve models to each event and use the resultant fits to separately constrain the flux contributions from the lens and source. In Section 2.6 we photometrically and astrometrically analyze the source stars in an effort to determine whether the lens is detectably luminous. In Section 3.6 we synthesize the different constraints, and then discuss the likelihood of each event being a BH in Section 4.6. Finally, in Section 3.8 we present the conclusions of our study.

2.2 Data Sets

Keck Observations

The primary data presented in this chapter are imaging observations of M96-B5 and M98-B6 taken with the NIR camera (NIRC2) on the W.M. Keck II 10m telescope behind the laser guide star adaptive optics (LGS AO) system [54]. Images were taken with the NIRC2 narrow camera in the Kp filter over a 10"×10" field with a plate scale of 9.952 mas pixel^{-1} [55]. Figure 2.1 shows the two fields, in each case with the target marked at its center.

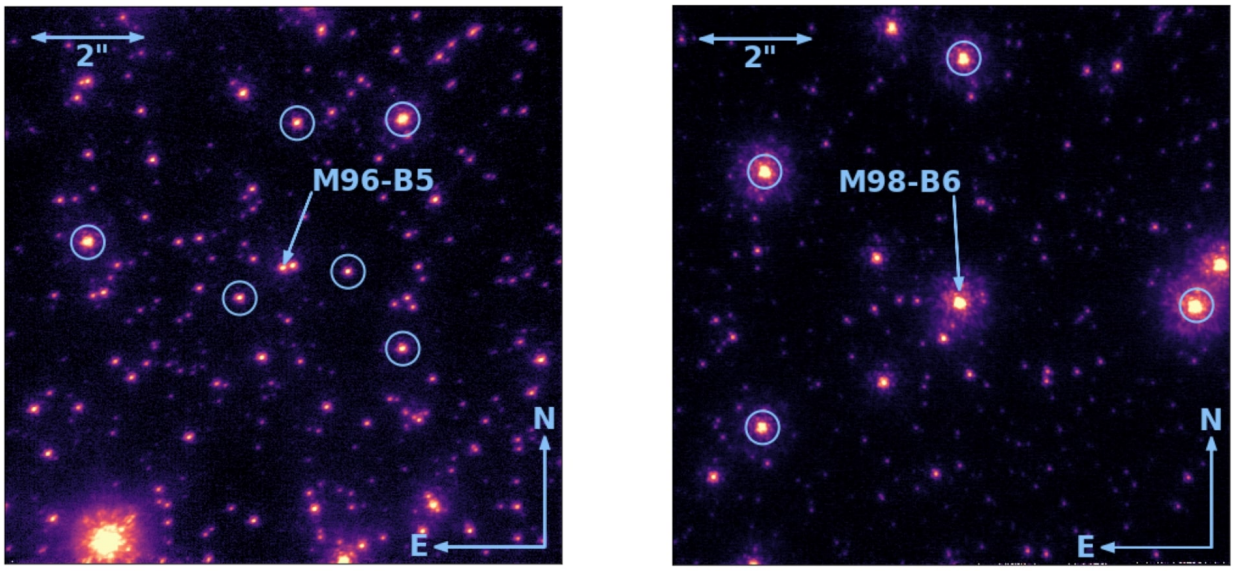


Figure 2.1: 2016 NIRC2 images of M96-B5 *(left)* and M98-B6 *(right)*. In each case, the background source star of the microlensing event is labeled with the event name. Stars used to create the mean PSF (cf. Section 2.4) are circled in light blue. PSF star selection was based on cuts in brightness, proximity to the target, and isolation. Note, the images have a logarithmic color scale.

Both targets were observed on 2016 July 14 UT, approximately 20 years after peak magnification of the corresponding microlensing events. Each target was observed several times in succession with a total integration time of 540 s for M96-B5 and 220 s for M98-B6. In order to remove detector artifacts such as bad pixels, images were taken with a random dither pattern within a 07 x 07 box.

A second epoch of NIRC2 data was taken for each of the two events in 2017 to constrain motions. Additional details of all observations are provided in Table 2.1.

MACHO Light Curves

Both events discussed in this chapter were first identified by the MACHO Project, in which 10-20 million stars in the Galactic bulge were surveyed from 1993-1999. The primary data set comes from the Mt. Stromlo 1.3m telescope in two filters: $B_{\rm MACHO}$ (∼450–630 nm) and $R_{\rm MACHO}$ (∼630–760 nm). The establishment of the MACHO Alert system made follow up observations from additional telescopes possible. As such, there are additional data for both events from the CTIO 0.9m telescope and the Mt. Stromlo 1.9m telescope [56].

In this work, we used the reduced and calibrated light curves for these targets from [48].

Table 2.2: Gaia Data for M98-B5 Position Match

Quantity	Value	Error
Source ID	4062585513471740288	
RA (deg)	269.38573681	2.4×10^{-8}
Dec (deg)	-28.71091478	1.9×10^{-8}
RA Proper Motion (mas/yr)	-2.98	0.11
Dec Proper Motion (mas/yr)	0.812	0.075
Parallax (mas)	0.150	0.096
Magnitude (g)	16.9661	0.0024
T_{eff} (K)	4465.25	+478.69/-582.60
ruwe	1.29	

Properties of the Gaia EDR3 object corresponding to the location of the event M98-B5.

Gaia Early Data Release 3

In order to define an absolute astrometric reference frame for measurements of the targets' proper motions, we use data from Gaia Early Data Release 3 (EDR3) [57, 58, 59, 60]. The EDR3 source list was queried for all sources within a 36"×36" box centered on each target. For the stars yielded by this search, we used only position information (RA and Dec). Additionally, a star in EDR3 was found to correspond to the location of the event M98-B5, the relevant characteristics of which are given below in Table 2.2. This star is likely the source in the lensing event, though it could be the lens and source together (a possibility we will further examine in Sections 2.5 and 2.6).

HST Broadband Photometry

In order to compile multi-band photometry for the microlensing event sources, we utilized archival data from the Hubble Space Telescope's WFPC2 instrument (HST-GO-8654, PI: Bennet, David P.). Observations were taken across 7 epochs between 1999 and 2003 for M96-B5, and a single epoch in 2000 for M98-B6. The data were taken in four wide filters: F439W, F555W, F675W, and F814W. In all cases, the target fell on the PC chip of the 4-chip camera. This 800×800 pixel CCD has a plate scale of 45.5 mas pixel^{-1}, corresponding to a 36"×36" field of view [61]. Additional details on the HST images used are shown in Table 2.3. The data used were downloaded in 2018 November.

Table 2.3: WFPC2 Observations

Target	Filter	Epoch	N_{exp}	t_{int} (s)	m_{base}
M96-B5	F439W	2000-06-11	6	2200	20.609±0.034
	F555W	1999-06-15	2	800	18.826±0.016
		2000-06-11	5	3240	19.027±0.027
		2001-06-03	2	800	18.877±0.095
		2001-10-01	2	800	18.884±0.047
		2002-05-25	2	800	19.021±0.009
		2002-10-02	2	800	18.995±0.041
		2003-05-27	2	800	19.042±0.009
	F675W	2000-06-11	5	1080	17.969±0.015
	F814W	1999-06-15	4	800	17.141±0.054
		2000-06-11	5	3240	17.304±0.017
		2001-06-03	4	800	17.336±0.015
		2001-10-02	4	800	17.310±0.026
		2002-05-25	4	800	17.333±0.020
		2002-10-02	4	800	17.365±0.008
		2003-05-27	4	800	17.341±0.009
M98-B6	F439W	2000-06-23	1	40	17.996±0.044
	F555W	2000-06-23	1	260	16.038±0.044
	F675W	2000-06-23	1	100	14.677±0.044
	F814W	2000-06-23	1	100	13.834±0.044

Observational parameters for archival HST data of MB96 and MB98. For a given target, filter, and epoch, there are N_{exp} frames with a combined integration time of t_{int}. The baseline (unlesnsed) magnitude m_{base} is the mean value and uncertainty in the mean of all frames in epochs with multiple observations. For epochs with a single observation, the uncertainty was set by inflating the uncertainties of multi-frame epochs with similar exposure times by the square root of the number of frames.

2.3 Raw Reduction and Catalog Creation

Keck Reduction and Star-finding with AIROPA

Initial reduction of raw data from each epoch of observation was carried out with our custom NIRC2 reduction pipeline [62, 63] and included flat field and dark calibration, sky subtraction and cosmic ray removal. The cleaned exposures were corrected for distortion and achromatic differential atmospheric refraction, shifted to a common coordinate system, and then combined, weighted by Strehl using the IRAF routine, `Drizzle` [64], as described in [65]. The final combined image was restricted to individual frames displaying core FWHM<1.25 FWHM_{min}, where FWHM_{min} is the minimum FWHM of all frames of the particular target and epoch.

On each combined map, we used the point-spread function (PSF) fitting routine, `AIROPA` [66], which is based on `StarFinder`, [67] to extract a stellar catalog of spatial coordinates and relative brightness. `AIROPA` is used in single-PSF mode, which assumes the PSF is uniform over the field of view. First, the PSFs of a subset of stars (hereafter "PSF stars") were averaged to extract a mean PSF. The PSF was then cross-correlated with the image and stars were identified as peaks with a correlation above 0.8. Additionally, we ran `StarFinder` in 'deblend' mode in order to check for a difference in sensitivity to close pairs of stars. There was no difference in the number of stars detected within the central 05 of either target (the range within which it would be reasonable to find a resolved lens). As such, `AIROPA` was run in normal ('non-deblend') mode for the remainder of this work.

The accuracy of our PSF directly impacts the contrast (i.e., our ability to detect a faint lens near the brighter source) as well as astrometric precision. Thus, we applied strict criteria in the selection of the PSF stars. Specifically, stars contributed to the mean PSF derivation if they were bright (typically Kp <18 mag), isolated, and within 4of the center of the field. The latter criterion avoids detector edge effects and ensures all PSF stars are close to the target of interest, mitigating errors due to spatial variation of the PSF caused by instrumental aberrations and atmospheric anisoplanatism.

The resulting starlists produced by `AIROPA` contain positions and fluxes for each star in detector units of pixels and counts, respectively. Residual images with found sources removed were also created for each target/epoch. Instrumental magnitudes were calibrated to Kp using J, H, and Ks magnitudes from the VVV Survey DR2 [68].

HST Reduction with `img2xymrduv`

To extract photometry from archival HST data, we used the FORTRAN program `img2xymrduv`, developed for use with WFPC2 images [69]. The code uses user-set criteria to find stars in an image before fitting them to a PSF model. It outputs instrumental magnitudes derived from fluxes in ADU, determined by the PSF fits and distortion corrected centroid positions.

To calibrate the magnitudes produced by `img2xymrduv`, we divided the observed flux in ADU by the total exposure time and applied the VEGAMAG system zero points for WFPC2

[61]. However, the zero points are defined for counts measured in a 05 radius aperture; thus we applied an additional ΔZP to calibrate the `img2xymrduv` magnitudes extracted from a smaller aperture. We determined ΔZP by performing aperture photometry on the calibrated images with a 05 radius aperture, positioned on the centroids output by `img2xymrduv`. A sigma clipped median pixel value in each image was used as the sky value to subtract off the aperture sums, and the resulting counts N_{aper} were then propagated through the equation

$$m_{aper} = -2.5 \log_{10}(\frac{N_{aper} \cdot g}{t_{exp}}) \tag{2.1}$$

along with the gain g and exposure time t_{exp} to get the magnitude, m_{aper}. The differences between m_{aper} and the `img2xymrduv` magnitudes for the brightest 10% of stars in an image was median combined to define $\Delta ZP = m_{aper} - m_{img}$. Finally, we calibrated magnitudes

$$m = m_{img} + ZP + \Delta ZP \tag{2.2}$$

where m_{img} is the `img2xymrduv` determined magnitude and ZP is the zero point from [70].

Comparison of Archival and Current Images

Equipped with starlists from both our late time (2016-2017) Keck images and early time (1999-2003), we can now determine if a new luminous lens was detected. In the case that a lens is a black hole, there should be no source nearby the late time images what wasn't present in the early time images. Searching within a radius of 3" of the source (which corresponds to a maximum proper motion of 15.2 mas/yr for M96-B5 and 16.7 mas/yr for M98-B6), no new sources appear. Visual comparisons of these images for both targets is shown in Figure 2.2. To understand the limitations of this comparison, however, we must examine our detection sensitivity.

2.4 Detection Limits on Luminous Lenses

Star Planting

To estimate the detection sensitivity of the images, a series of star planting simulations was conducted. Individual stars were planted in each image by taking a copy of the point spread function (PSF) that was created using the original image. In order to reduce computational time, the main image and background image were cropped from a 1064×1064 pixel array down to 200×200 pixels, and the PSF was cropped down from 200×200 pixels 30×30 pixels. Normally, the PSF is normalized such that the integral over the entire PSF is 1. However, because the cropped PSF (cPSF) was taken from an initially normalized PSF, the integral over the cropped PSF is less than 1. The cPSF was then calibrated to a chosen magnitude. Poisson noise was then added to the cPSF. This calibrated and noisy cPSF was then used as an artificial star. The artificial star was then planted in the image, with its center located

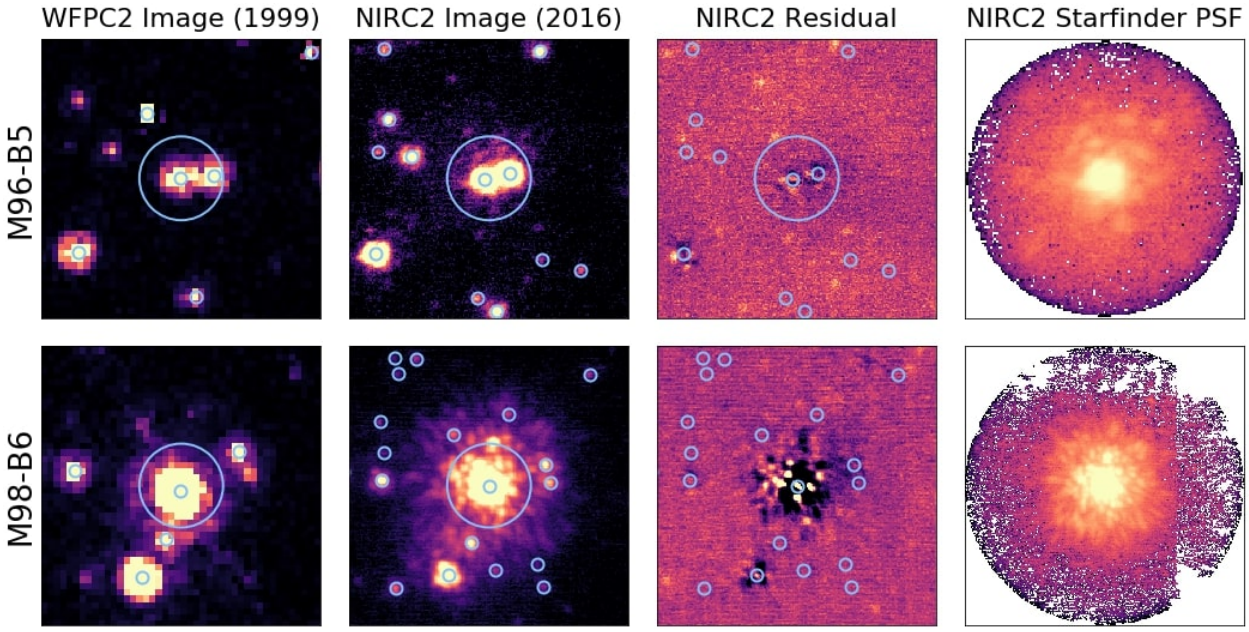


Figure 2.2: Zoomed in images of M96-B5 (top) and M98-B6 (bottom). For each row, there is an HST image right in 1999, shortly after the microlensing events (left), a NIRC2 image from 2016 (center), and the PSF of the NIRC2 image as determined by `AIROPA` (right). The sky images are each 2" across, with a large light-blue circle with radius 0.3" centered on each target. All sources found by `AIROPA` are indicated with smaller red circles. In both cases, no new sources have appeared within 0.3" of the source, which corresponds to the maximum separation of a lens moving with proper motion of 15 mas/yr.

at pixel position x_i, y_i on the image. A total of 167,445 artificial stars were planted for each event, varying in magnitude from 12.00 to 23.00, with steps of 0.25, and in location from -30 to 30 pixels in both the x- and y-directions around the target. The resulting image, with a single planted star, was analyzed with `AIROPA` in an identical manner to that used on the original image, using the exact same PSF. It was put through `StarFinder` to see if the planted star was detected. The planted star was considered detected if its position in the new starlist matched the input location to within a single pixel in both the x- and y-direction and if its magnitude agreed with the input to within 0.5 mag.

A radial completeness curve was created by binning the artificial stars into magnitude bins of 0.25 mag and radial bins of using a sliding window of 39.72 *mas* (4 *pixels*). Within each magnitude-radius bin, the completeness was determined by calculating the percentage of planted stars within the category that were detected. A completeness value of 1.0 corresponds to a detection percentage 100%, and a value of 0.0 corresponds to a detection percentage of 0%. Limiting magnitudes were determined for each radial bin by taking the faintest magnitude with a completeness value of at least 0.95. The limiting magnitudes were given

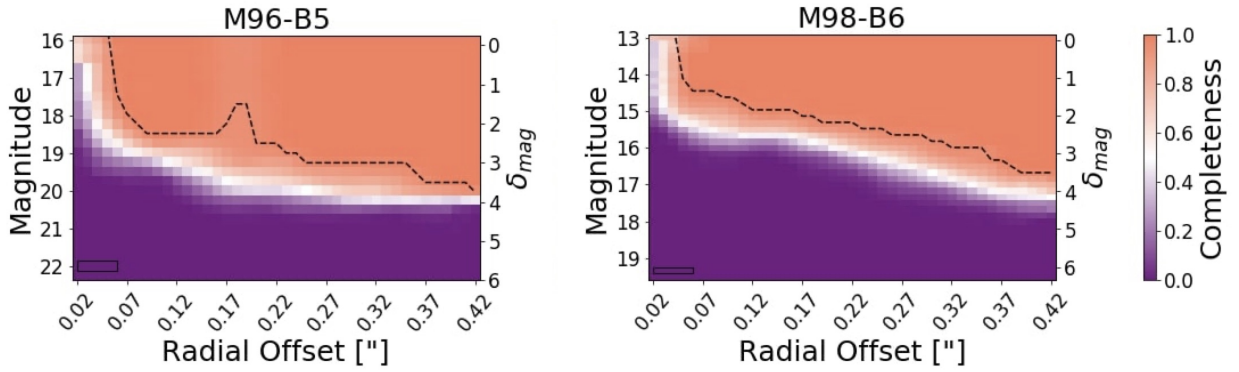


Figure 2.3: Completeness curves for MB 96-5 (top) and MB 98-6 (bottom). 95% completeness is indicated by dashed black lines. By finding where the detection completeness for a giving magnitude goes below 0.95, we determined the magnitude range of a stellar lens undetectable by `Starfinder`. These magnitude ranges gave us our upper limits to the mass of an undetectable, but luminous, stellar lens.

the average magnitude of each bin and the average radius of each window (Figure 2.3). No limiting magnitudes were used that exceeded the brightness of the respective source.

Isochrones Based on Magnitude

The limiting magnitudes were converted to stellar masses, as described below, and used to determine the allowed mass range for an undetected, luminous stellar lens. Synthetic isochrones were generated using the program `SPISEA` (Hosek et al. in prep). Isochrones were generated for a given age in $log(yr)$, distance in pc, and extinction value A_k, assuming solar metallicity. Distances ranged from 1 kpc to 10 kpc, with steps of 1 kpc. The extinction values, which are dependent on the Galactic coordinates and distance of the source, were determined using the Argonaut Skymaps [71] [72], which output an $E(B-V)$ value. That $E(B-V)$ was then used to calculate an A_V value by assuming $R_V = 3.1$. Then, the extinction value A_k was calculated using the extinction law from [73]. As a proxy for a "zero-age main-sequence", we adopted an age of 100 Myr to obtain a nearly fully main sequence while also excluding pre-main sequence stars.

Using the generated isochrones, the outputted mass values were taken for any isochrone entries that matched a limiting magnitude to within 0.2 mag. Thus, for each limiting magnitude, corresponding to a radial angular separation from the source, a mass-distance relationship was determined. We note that these mass limits are extracted under the assumption that the lens is a main-sequence star. However, pre-main-sequence or post-main-sequence stars are more luminous for a given mass; thus our use of a main-sequence mass-luminosity relationship is conservative. Finally, each radial bin was converted into a proper motion bin

using the respective time elapsed since closest approach and the distance value inputted into the synthetic isochrone. With this, our completeness in magnitude/radial seperation space has been projected to mass/lens distance/relative proper motion space. This new parameterization allows our completeness to be related to the microlensing events, the light curves of which are described by physical quantities including lens mass M, lens distance D_L, and relative proper motion μ_{rel}.

2.5 Light Curve Fitting and Analysis

Fitting Routine and Results

Microlensing models were fit to the light curves described in Section 2.2 by applying the publicly available software package `pyLIMA`, using the differential evolution fitting method [74]. Each event was fit to a standard point-source point-lens (PSPL) model with parallax. The number of free parameters in the model depends on the number of light curves of different filters fit. For any single-lens parallax event, there are initially five parameters: the time of closest approach between the source and lens (t_0); the impact parameter between the source and the lens in units of the Einstein radius (u_0), the characteristic timescale, or 'Einstein crossing time' ($t_{\rm E}$); and the North and East components of the microlensing parallax vector ($\pi_{\rm E,N}$ and $\pi_{\rm E,E}$). For each filter, there are two additional parameters: the assumed-to-be-static flux of the source in the absence of lensing, f_s and the blend flux f_b, which is the flux from companions to the lens, companions to the source, and/or ambient interloping stars that incidentally fall within the seeing disc of the microlensing target (the 'target' being the superimposed lens and source) during the event. The values obtained by the fits for this PSPL model for both events are presented in Table 2.4. Corner plots showing posteriors for all solutions are shown in Figure 12-15.

Additionally, after each initial fit was conducted, a second fit was run with the additional constraint that u_0 only be allowed to have values with the opposite sign of the initial u_0 results, in order to find degenerate solutions. For example, if u_0 was first found to equal -0.5, in the second fit, u_0 was restricted to positive values.

The two flux parameters can be recast to a more intuitive pair of parameters: un-lensed baseline magnitude of the source (m_s) and blend-source-flux-fraction (b_{sff}), which is the ratio of the flux from the source alone to the total flux in the aperture of the images from which the light curves were derived. The transformation is made with the following equations:

$$m_s = 27.4 - 2.5 \log_{10}(f_s) \tag{2.3}$$

$$b_{sff} = \frac{f_s}{f_s + f_b} \tag{2.4}$$

The resulting m_s and b_{sff} values for a subset of the data (that will be used in the following section) are presented in Table 2.5.

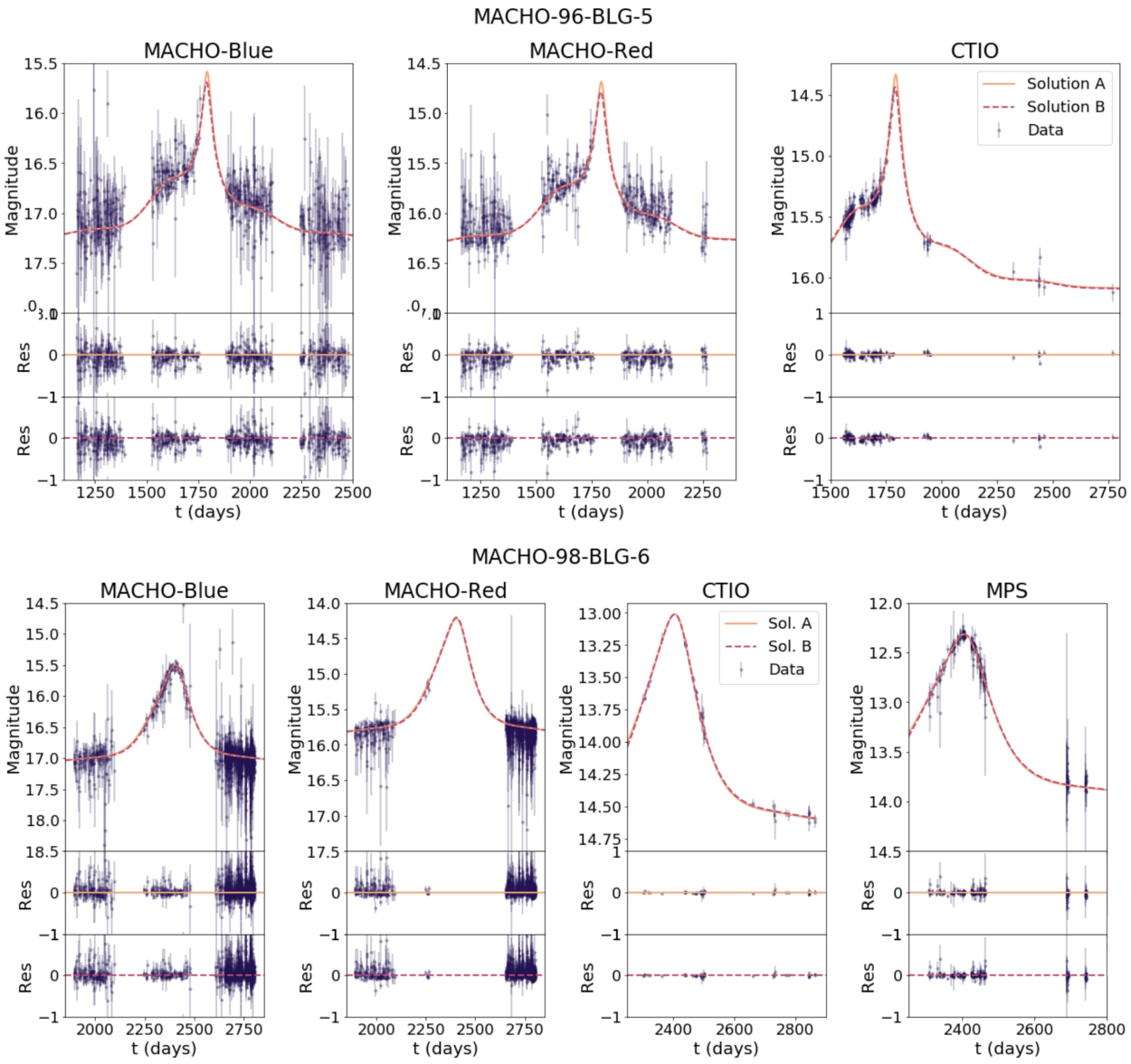


Figure 2.4: Fitted models of the light curves for M96-B5 (top) and M98-B6 (bottom). In each case, the light curve data is shown in dark purple points, the best fit model with negative u_0 in solid orange, and the best fit model with positive u_0 in dashed magenta. The residuals (data - model) for each light curve are shown immediately below in matching colors and style. Note that the visualization of the data has been truncated to only show rising/falling of the light curve, and not baseline (or approximately baseline) measurements. The date t is reported in JD - 2448623.5.

Table 2.4: Light Curve Model Fit Parameters

Parameter	M96-B5 (A)	M96-B5 (B)	M98-B6 (A)	M98-B6 (B)
t_0 [days]	$1766.2^{+1.2}_{-1.3}$	$1773.10^{+0.88}_{-0.93}$	$2413.7^{+1.4}_{-1.5}$	$2407.4^{+1.5}_{-1.6}$
u_0 [$\theta_{\rm E}$]	$-0.0296^{+0.0049}_{-0.0052}$	$0.0164^{+0.0042}_{-0.0037}$	$-0.174^{+0.019}_{-0.022}$	$0.175^{+0.023}_{-0.023}$
$t_{\rm E}$ [days]	537^{+92}_{-68}	628^{+100}_{-74}	367^{+41}_{-37}	346^{+45}_{-35}
π_{EN}	$0.0397^{+0.0074}_{-0.0070}$	$-0.0284^{+0.0050}_{-0.0051}$	$-0.087^{+0.023}_{-0.021}$	$0.0294^{+0.023}_{-0.024}$
π_{EE}	$0.076^{+0.012}_{-0.012}$	$0.0535^{+0.0073}_{-0.0077}$	$0.076^{+0.012}_{-0.013}$	$0.0873^{+0.0066}_{-0.0076}$
f_s (MACHO-B)	1900^{+330}_{-320}	1550^{+240}_{-240}	9600^{+1400}_{-1200}	9300^{+1500}_{-1500}
f_b (MACHO-B)	9110^{+250}_{-270}	9400^{+170}_{-190}	3900^{+1100}_{-1400}	4300^{+1400}_{-1500}
f_s (MACHO-R)	4320^{+750}_{-710}	3510^{+540}_{-540}	31400^{+5500}_{-4700}	31000^{+5500}_{-5000}
f_b (MACHO-R)	21820^{+540}_{-590}	22470^{+390}_{-420}	10300^{+4500}_{-5300}	11000^{+4800}_{-5300}
f_s (CTIO)	6100^{+1100}_{-990}	5000^{+770}_{-770}	96000^{+14000}_{-12000}	93000^{+15000}_{-14000}
f_b (CTIO)	26300^{+1100}_{-1200}	27200^{+1000}_{-1000}	31000^{+11000}_{-13000}	35000^{+14000}_{-15000}
f_s (MPS)	-	-	183000^{+27000}_{-23000}	177000^{+29000}_{-28000}
f_b (MPS)	-	-	56000^{+21000}_{-25000}	63000^{+27000}_{-28000}
χ^2_ν	1.57	1.56	1.17	1.18
d.o.f.	2290	2290	2302	2302

Results of fitting MACHO light curves to PSPL models. Each event has two solutions, with negative and positive u_0 due to the degeneracy in this parameter, henceforth referred to 'Solution A' and 'Solution B', respectively. Note that the date t_0 is JD-2448623.5.

The PSPL with parallax model well describes the M98-B6 light curve, fitting with a reduced chi-squared of 1.17 and 1.18 for negative and positive u_0 values, respectively. As is visible in the bottom panel of Figure 2.4, this event appears to be a smooth PSPL curve with minor perturbation due to parallax. The fit for M96-B5 is slightly poorer, with reduced chi-squared of 1.57 and 1.56; though both models appear to fit the data well, the lack of coverage for times of peak amplification introduces some uncertainty, yielding more variation in solutions than for M98-B6, which has full peak coverage in two filters.

To potentially better fit M96-B5, we attempted fitting to a point-source-binary-lens (PSBL) model, which has three additional parameters: the log of the mass ratio between the two lenses $\log q$; the log of the projected binary separation in units of the Einstein radius $\log s$; and the position angle between the binary axis and source trajectory α. However, because the PSBL model did not significantly increase the quality of the fit (χ^2_ν=1.62), we will only consider the PSPL results for the remainder of this work.

Constraining f_{lens}

In cases of a luminous lens or additional light sources within the aperture of the source star, the resultant microlensing light curve has substantial alterations to it. The additional, unlensed flux remains constant while the flux of the source varies, resulting in a smaller ap-

parent maximum magnification, as well as what appears to be a shorter $t_{\rm E}$. This information is captured in b_{sff}, which we can rewrite as as

$$b_{sff} = \frac{f_s}{f_s + f_b} = \frac{f_s}{f_s + f_l + f_n} \tag{2.5}$$

where f_l is the flux from the lens and f_n is the flux from neighboring stars. b_{sff} is fit as a free parameter when modeling the photometric light curve.

To constrain f_l, we can combine b_{sff} with two relevant quantities derived from the WFPC2 images. The first quantity is the instrumental magnitude of the target from `img2xymrduv` as described in Section 2.3. Paired with image exposure time, this magnitude is converted to a count flux which we interpret as $f_s + f_l$ (which we will call f_{targ}). As the HST images were taken only 2-3 years after the peak of each event, we assume here that the source and lens would have not had adequate time to separate appreciably. Second, by summing the total flux within the 1.2" radius circle centered on the source, imitating the observations from the original MACHO data set, we obtain $f_s + f_l + f_n$ (or f_{aper}). We can write the relationships between these quantities:

$$\left(\frac{f_s + f_l}{f_s + f_l + f_n} - \frac{f_s}{f_s + f_l + f_n}\right)(f_s + f_l + f_n) = f_l \tag{2.6}$$

$$f_l = \left(\frac{f_{targ}}{f_{aper}} - b_{sff}\right) f_{aper} \tag{2.7}$$

thereby extracting the lens flux. This quantity is then converted back into a WFPC2 instrumental magnitude and finally calibrated into Vega magnitude. The corresponding source magnitude is derived similarly. Note that this process is only possible for the components of the light curve data in the MACHO-Blue and MACHO-Red filters, as their passbands align reasonably well to that of the WFPC2 F555W and F675W filters (shown in the appendix).

This analysis was done on all HST frames in the epoch 2000-06-11 from Table 2.3. Though there are data of M96-B5 in 1999, they only exists in the F555W filter, and was thus excluded in favor of a slightly later epoch for the purpose of combining the MACHO-Blue/F555W results with the MACHO-Red/F675W results in a later section. Figure 2.5 shows a summary of these results, where source and lens magnitude are plotted as a function of b_{sff} from 0 (no source flux) to 1 (no non-source flux). As seen in this figure, the result is highly sensitive to b_{sff}. The source and lens magnitudes derived from this analysis are presented in Table 2.5.

In the case of this work's value of b_{sff}, both solutions for M96-B5 indicate a lens with non-zero flux. For M98-B6, we see that both solutions' b_{sff} indicates that there is no lens flux, which we would expect in the case of a BH lens. However, the 1-sigma error bar on b_{sff} does extend into the region of non-zero lens flux. As such, in Table 2.5 we report the lens magnitudes at the lower boundary of b_{sff} regions as limits on the lens magnitude of this event.

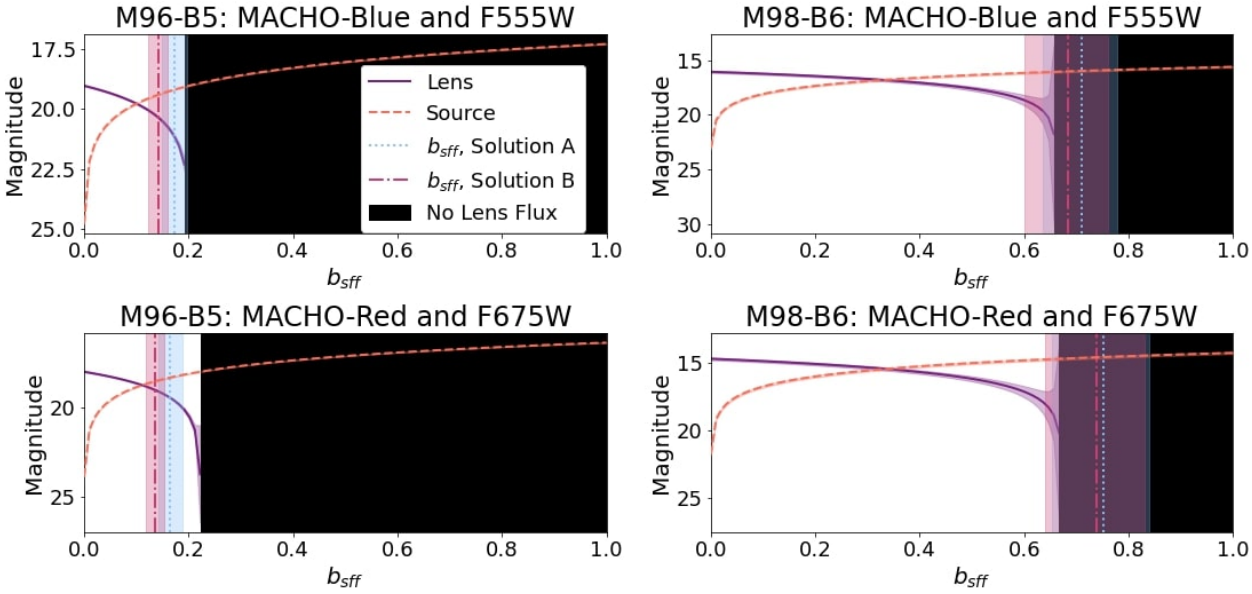


Figure 2.5: For each target and filter combination, the solid purple curve shows the magnitude of the lens for all possible values of b_{sff}. The dashed orange curve shows the corresponding source magnitude. The vertical lines show the values obtained for b_{sff} from Table 2.4, in each case with Solution A ($-u_0$) values light blue and dotted, and Solution B ($+u_0$) magenta and dot-dashed. The black regions designate the values of b_{sff} that would result in zero lens flux. The intersections between the lens curve and vertical lines are this work's estimates of the lens' magnitudes. Where the value of b_{sff} falls within a region on zero lens flux but has an error bar that extends into the region of non-zero lens flux, the intersection of the lower b_{sff} boundary and the solid purple curve are taken as lower limits on the lens magnitude.

2.6 Source Analysis

Astrometric Determination of Source Proper Motion

Before the alignment necessary for computing proper motions for the source of M96-B5, the WFPC2 starlists described in Section 2.3 were consolidated into single starlists for each epoch and filter. This was done using follow up programs to `img2xymrduv`, `xym2mat` and `xym2bar`, which match and average combine sources in a set of starlists (Anderson et al. 2008). By combining multiple observations at a given epoch, these collated starlists include uncertainties in centroid positions, which are ultimately needed for establishing the precision of the final proper motion determination.

Though the source star for M96-B5 is not in Gaia EDR3, all stars in the catalog within a 36 box centered on the source were used as an absolute reference frame to which all epochs

Table 2.5: Lens Magnitude Constraints

Event	Filter Combination	m_{targ}	Solution	b_{sff}	m_{source}	m_{lens}
M96-B5	MACHO-B/F555W	19.027± 0.027	A	$0.173^{+0.025}_{-0.024}$	$19.20^{+0.15}_{-0.16}$	$21.11^{+0.89}_{-0.93}$
			B	$0.141^{+0.019}_{-0.019}$	$19.42^{+0.15}_{-0.15}$	$20.32^{+0.33}_{-0.34}$
	MACHO-R/F675W	17.969±0.015	A	$0.165^{+0.024}_{-0.023}$	$18.30^{+0.15}_{-0.16}$	$19.42^{+0.43}_{-0.45}$
			B	$0.135^{+0.018}_{-0.018}$	$18.52^{+0.15}_{-0.15}$	$18.97^{+0.22}_{-0.23}$
M98-B6	MACHO-B/F555W	16.038±0.044	A	$0.710^{+0.067}_{-0.075}$	$15.96^{+0.15}_{-0.13}$	>18.45
			B	$0.683^{+0.078}_{-0.082}$	$16.00^{+0.16}_{-0.15}$	>18.16
	MACHO-R/F675W	14.677±0.044	A	$0.752^{+0.087}_{-0.10}$	$14.553^{+0.17}_{-0.15}$	>18.38
			B	$0.739^{+0.091}_{-0.099}$	$14.573^{+0.17}_{-0.16}$	>18.44

For a given event, light curve solution and filter combination, we present the results of using the blend fraction b_{sff} to separate the target magnitude m_{targ} into the source and lens components, m_{source} and m_{lens}, respectively.

of data were matched. This included both epochs of NIRC2 data and all epochs of WFPC2 data in the F555W and F814W filters. The starlists from these data sets were matched and transformed into the coordinate system of the EDR3 starlist with `FlyStar`, a package that performs matching and astrometric transformations on starlists. The resulting relative positions are reported in the appendix.

Once the cross-epoch alignment was complete, the positions of the source at each epoch were fit to a linear, constant proper motion model, yielding a proper motion of -2.955±0.022 mas/yr in the East direction and -0.926±0.022 mas/yr in the North direction, with a reduced chi squared of 1.38. The proper motion of this target and all Gaia EDR3 sources within a 36" box centered on the target are shown in the vector point diagram in Figure 2.6.

Unlike M96-B5, the source star of M98-B6 matches a object included in Gaia EDR3. The position, proper motion, and parallax of the source in Gaia are given in Table 2.2. There are several checks we use to confirm that this object is indeed the same as our microlensing source. First, we consider the ruwe (renormalized unit weight error) parameter, for which values below 1.4 suggest an unblended source with robust astrometric measurments [60]. As this source has a ruwe parameter of 1.29, it is unlikely that these are two distinct but spatially-unresolvable stars. Additionally, the ruwe parameter justifies our use of the measured parallax in estimating the source distance (which is $6.7^{+11.9}_{-2.6}$ kpc).

To further confirm that the Gaia object is indeed the source star and not an unrelated star at a different distance, the procedure described for determining the proper motion for M96-B5 above was repeated for M98-B6 in order to compare the results to the proper motion reported in Gaia. The result was a proper motion of -2.82±0.35 mas/yr in the East direction and 0.47±0.36 mas/yr in the North direction. Within uncertainties, this is consistent with the Gaia proper motions of -2.98±0.11 mas/yr East and 0.812±0.075 mas/yr North, further supporting the claim that these two sources are one and the same. The proper motion of this source in the context of its field is also shown in Figure 2.6.

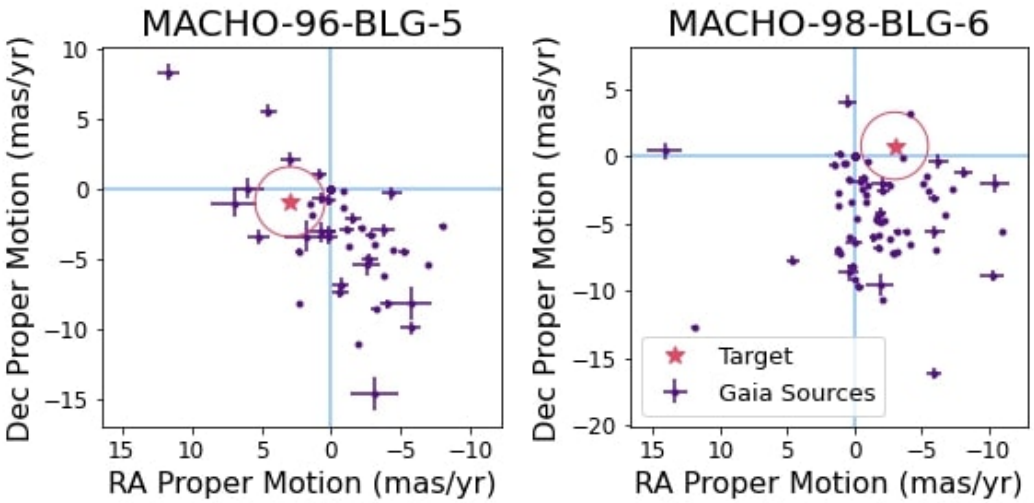


Figure 2.6: Vector point diagrams of each target (magenta), each shown with all Gaia sources within a 36" box centered on the respective target (purple). Proper motions are in the Gaia reference frame. A magenta circle centered on each target indicates the region of proper motion space in which an object would have a proper motion relative to the target of 2.5 mas/yr. An object that is within this circle and also positioned within 03 of the target would not be resolvable from the source in our 2016/2017 images.

Photometric Exploration of Source and Lens

As shown in Figure 2.3, our ability to directly observe a luminous lens becomes difficult as the radial offset from the source in the image plane decreases. If the relative proper motion between a source and lens is small, there may not be sufficient angular separation between the two objects to resolve them in our NIRC2 images. To explore a scenario in which objects appear blended even after 20 years, we use the WFPC2 photometry from the 2000-06-11 (for M96-B5) and 2000-06-23 (for M98-B6) epochs to look for indications of additional unresolved objects in/near the source. Though this concept was initially intended for the purpose of covering the low relative proper motion case, some back-of-the-envelope calculations show that it should actually hold regardless of proper motion: high lens mass (BH) microlensing events will produce Einstein radii of a few miliarcseconds, which we can use along with the light curve fit's u_0 values (see Table 2.4), which is in units of the Einstein radius, to show that at time of closest approach, the lens and source in both events should not have been separated more than $\sim$1 mas. The times between this closest approach and the observations are no more than 3 years, which, along with the aperture radius of 05 used for the photometry, means that the relative lens source proper motion would have to be $\sim$100 mas/yr in order for the lens to to have moved outside of the aperture. We can see that for both events, the source proper motion is only a few mas/yr, meaning the majority

of the required 100 mas/yr would have to be made up by the lens. Thus, we can reasonably assume that at the time of the WFPC2 observations, the lens is still within the aperture observing the source.

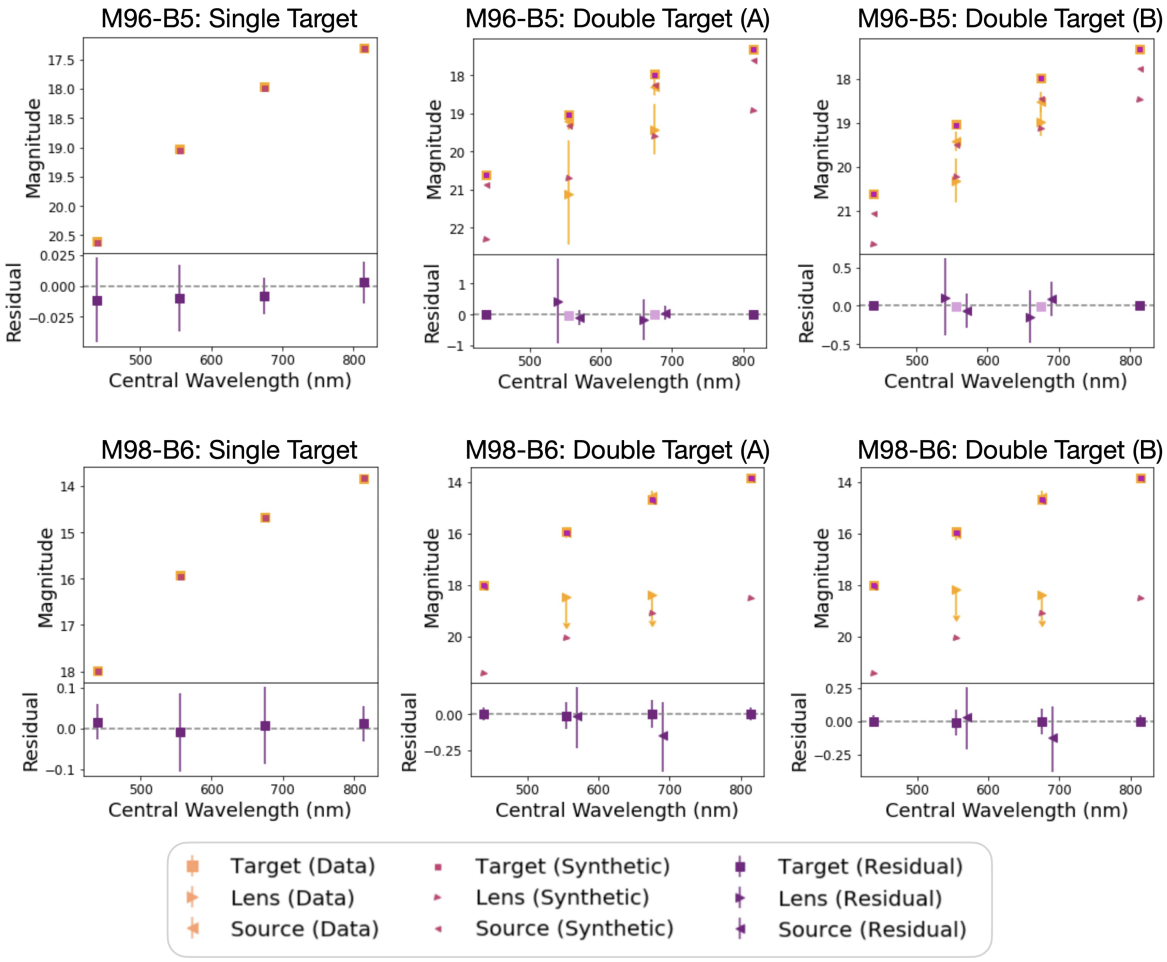
We will call the hypothetically unresolved lens and source object the 'target', for which the photometric information in four filters was given in Table 2.3. In the case of M96-B6, we also have disentangled photometric data for a lens and source in two filters (F555W and F675W), which were presented in Table 2.5. For M98-B6, the disentangled photometry consists of source magnitudes and lower limits on the lens magnitude.

In order to identify objects that our photometric data could represent, we again use `SPISEA`, the stellar population generating python package discussed in Section 2.4. `SPISEA` allows not only for the choice of many parameters for the populations it creates, but also for photometry from particular instruments and filters to be simulated. For the purposes of matching to our data, we simulated photometry for HST's WFPC2 instrument in the filters F439W, F555W, F675W and F814W.

To create a sample of stars to search in, we generated isochrones at a range of cluster distances (1 kpc to 18.5 kpc, in increments of 0.5 kpc, the upper limit set by the maximum distance allowed by M98-B6's EDR3 parallax measurement) and ages (log age in years spanned 8 to 10, in increments of 0.5). Metallicity was held constant at solar metallically, and default evolution and atmospheric models were chosen. A Cardelli extinction law with Rv=3.1 was applied along with extinction values pulled from the previously cited Argonaut Skymaps. This resulted in a library of 50 isochrones with 200 stars each. The code does not generate compact objects or brown dwarfs.

For each event, we explored three scenarios. In the first, we made the assumption that the WFPC2 photometric data from the 2000 epoch (see Table 2.3) showed only a single object and thus looked for the closest single-object match. This was done by identifying the minimum chi-squared between the photometric data and each synthetic photometry object. Though our light curve fits for M96-B5 indicated a flux contribution from both a lens and source, Figure 2.5 shows how sensitive this result is to b_{sff}, and though our result allowed for non-zero lens flux, previous works that have fit the same light curve have obtained blending parameters that, in the analysis in Section 2.5 would yield zero lens flux. For example, the fit from Bennet 2002 presented two solutions that would put b_{sff} fully in the black region of Figure 2.5). We note that though our fits yielded lower chi-squared than those of Bennet 2002, there is inherent ambiguity in this event due to the lack of coverage of the peak magnification. As such, we explore this possibility along with the scenarios more supported by this work.

The second scenario and third scenarios we explore for each event is the case of a single luminous source and single luminous lens, with magnitudes (or magnitude limits, for the lens of M98-B6) dependent on the two light curve solutions. These disentangled lens and source magnitudes come from Section 2.5, in which we derived separate lens and source photometry in the filters F555W and F675W. The least squares search for this scenario operated slightly differently than the single object case. For M96-B5, the data points matched were the F555W and F675W photometry of one object for the source and one object for the lens, in



,

Figure 2.7: Comparisons of broadband photometry of the event targets to best-matched synthetic data. For each event (M96-B5 in the top row and M98-B6 in the bottom row), three potential scenarios are considered. In the left-most panel, the target is assumed to be only the background source (meaning the lens is either non-luminous or is spatially resolved from the source). In the middle and right-most panels, the target is assumed to be an unresolved source and luminous lens. The disentanglement of the lens and source in these cases is derived from the two solutions to each light curve fit, Solution A in the center and Solution B on the right. In each figure, the total target data is represented in orange squares, and if a separate lens and source are shown, they are represented in orange triangles directed to the right and left, respectively. The synthetic data matches are symbolically represented the same, but in magenta with smaller markers. In the residual panels at the bottom of each plot, if lens and source points are present, they are slightly offset from their central wavelength to improve figure visibility. Additionally, only the dark purple points were used to calculate χ^2, as the light purple points are redundant information, but useful for visualization.

Table 2.6: Best Synthetic Photometry Matches

Event	Target	Object	$\log A_c$ (y)	D_c (kpc)	T_* (K)	$M_*/M_\odot$	$L_*/L_\odot$	$R_*/R_\odot$	χ^2/pts
M96-B5	Single	Source	9.5	8.5	4960	1.41	10.76	4.44	0.144
	Double (A)	Source	9.5	8	5043	1.41	7.39	3.56	0.079
		Lens	9	2.5	4584	0.73	0.17	0.66	
	Double (B)	Source	9.5	10	4980	1.41	9.76	4.19	0.082
		Lens	8	2	4457	0.70	0.14	0.63	
M98-B6	Single	Source	8.5	7.5	4572	3.19	297.30	27.46	0.057
	Double (A)	Source	8	16.5	4633	5.11	1466	59.36	0.056
		Lens	10	3.5	5352	0.84	0.54	0.85	
	Double (B)	Source	8	16.5	4633	5.11	1466	59.36	0.045
		Lens	10	3.5	5352	0.84	0.54	0.86	height

Closest matches from synthetic isochrones to target photometry. For each scenario, the log age and distance to the cluster(s) in which the closest match was found is given, in addition to the effective temperature, mass, luminosity, and radius of the matching object. As a metric of fit, chi-squared per point is reported. For single target scenarios, chi-squared is calculated with four data points, while double target scenarios are calculated with six.

addition to the F439W and F814W photometry for the summed source and lens combination (as there was not adequate data to separate the lens and source contribution in these two filters). Because M98-B6 has lens magnitude limits rather than specific values, the search was conducted by first finding all objects that fell above the lens limits, then summing each of these with any objects at a greater distance. For each of these combinations, chi-squared was calculated with the two source points (still F555W and F675W) and all four of the combined target points, instead of any lens points. In either case, all six points used were weighted equally in the search.

For each of the four scenarios above, a closest matched source (or sources and lens) was identified. The physical parameters of the matched sources, as well as a chi squared per point are presented in Table 2.6. Comparisons of the data to the synthetic photometry for each scenario are illustrated in Figure 2.7). The isochrones from which each match object was pulled are illustrated in Figure 2.10, in the appendix.

2.7 Results

To encapsulate the results of the three analysis sections of this chapter (4-6), we will begin by looking at Figure 4.2, which summarizes the findings of Sections 2.4 and 2.5 and illustrates the need for Sections 2.5 and 2.6.

For each event, the colorbar indicates the maximum mass of a stellar lens that could have been undetected by our process of artificial star planting and retrieval in NIRC2 images

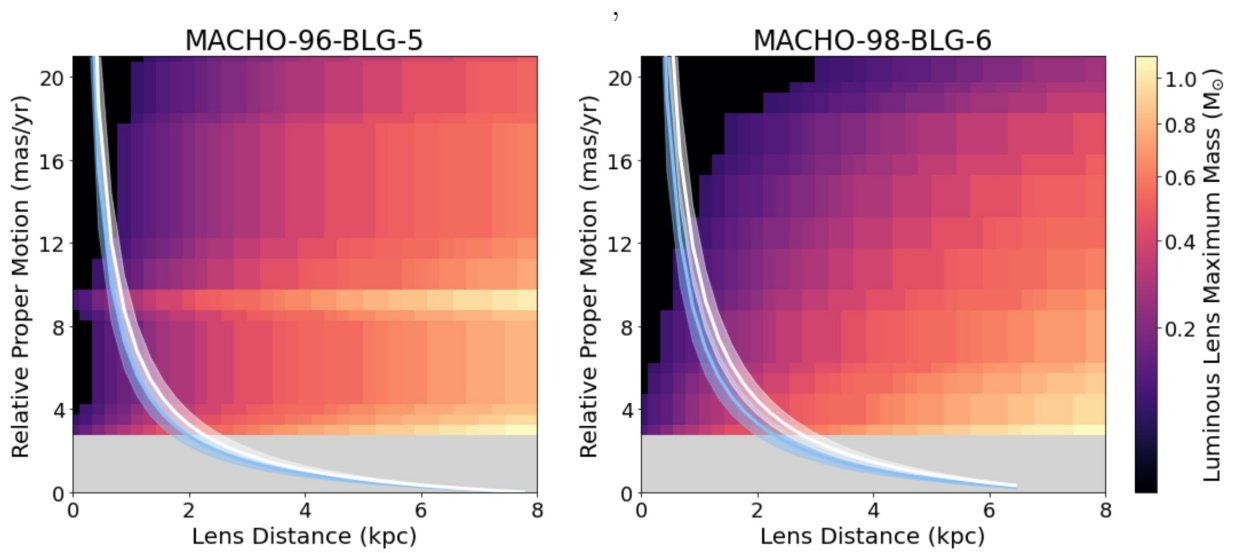


Figure 2.8: For each event, the colorbar indicates the maximum mass of a stellar lens that could have been undetected by our process of artificial star planting and retrieval in NIRC2 images described in Section 2.4 as a function of lens distance and relative proper motion, with the black area indicating parameter space for which there were no stellar matches found. The lens distance/relative proper motion relationship derived from each solution of the light curve fit is shown in cyan and white lines, with transparent surrounding regions designating plus or minus one sigma to that relationship, with Solution A in cyan and Solution B in white. The gray region represents the possibility of relative proper motions low enough (< 2.5 mas/yr) such that the lens and source are not resolvable. The only stellar lenses allowed by our model fits and not identified in our NIRC2 images must be below the masses indicated by the colorbar within the cyan and white bands.

described in Section 2.4. This mass limit is presented as a function of lens distance and relative lens-source proper motion, two microlensing parameters that cannot be constrained by a light curve alone. The black area indicates parameter space for which there were no stellar matches found. Note that despite being cast into parameters that are meaningful for microlensing events, the completeness data itself is derived solely from our 2016/17 images of the sources, and not the microlensing events.

The photometric microlensing light curve data is incorporated in the cyan and white curves (for Solution A and B, respectively), which indicate combinations of lens distance and relative lens-source proper motions allowed by the fits in Section 2.5. The surrounding shaded cyan and white regions designate plus or minus one sigma to those relationships. The curves assume a source distance: 8.5 kpc for M96-B5 and 7.5 kpc for M96-B5 (each chosen based on the the best-matched synthetic object to a single source target, reported in Table 2.6). The only stellar lenses allowed by our model fits and not identified in our NIRC2

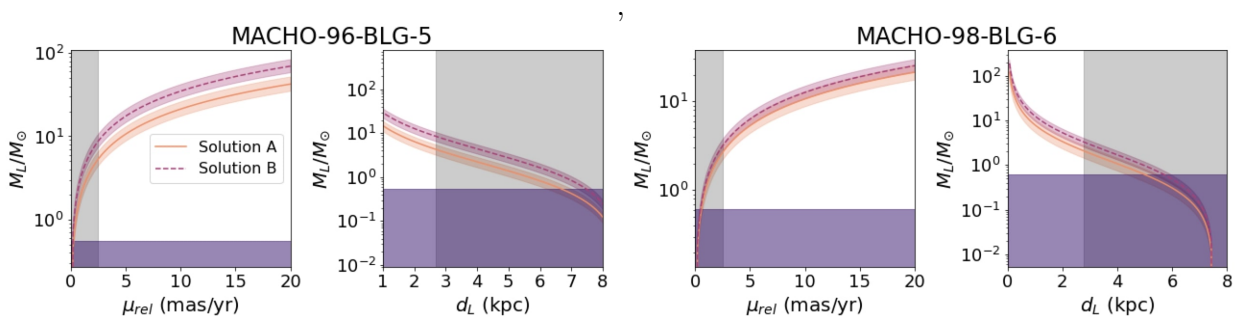


Figure 2.9: Possible lens masses as a function relative lens-source proper motion (first and third panels from the left) and as a function of lens distance (second and fourth panels from the left) for both events. In each case, the solid orange and dashed magenta lines represent the x-axis variables' relationship to lens mass as determined by Solution A and B of the light curve fit results, respectively, with shading designating plus or minus 1-sigma to that relationship. The purple regions identify all possible lens masses at which a stellar lens would have been undetected by our star finder. The gray region represents the range of relative proper motions low enough and corresponding lens distance limits such that the lens and source are not resolvable. In the regions in which we are to resolve a lens and source, undetectable masses fall below the allowed masses from the light curves in every case. Note that the curves assume a source distance; in each case, the source distance assumed is that of the best-fit photometric match to a single target, given in Table 2.6.

images must be below the masses indicated by colors within these shaded cyan and white bands.

Combining these two pieces of analysis of two different data sets, we find that the maximum possible mass of a stellar lens in each event is 0.53 $M_\odot$ and 0.55 $M_\odot$ for Solutions A and B of M96-B5, and 0.55 $M_\odot$ and 0.61 for M98-B6. We note that these mass limits are derived from simulating main-sequence stars, which are less luminous for a given mass than pre-main-sequence or post-main-sequence stars (meaning these types of stars would all be detectable as well). However, we also note that this does not account for other luminous objects, such as brown dwarfs or non-BH compact objects.

These masses, however, have little to do with the lens masses allowed by the light curve data. By combining Equations 1-3, we can derive a microlensing event's lens mass as a function of the fit parameters $t_{\rm E}$ and $\pi_{\rm E}$, and the relative proper motion μ_{rel}:

$$M = \frac{t_{\rm E}\mu_{rel}}{\kappa\pi_{\rm E}} \tag{2.8}$$

where $\kappa = \frac{4G}{1\,\mathrm{AU}\,c^2}$. Further, by assuming a source distance as described above, we can recast this relationship as lens mass as a function of lens distance using Equations 2-3:

$$M = \frac{1 \text{ AU } (d_L^{-1} - d_S^{-1})}{\kappa \pi_{\rm E}^2} \tag{2.9}$$

Both relationships for each event and solution are illustrated in Figure 2.9, in which, similar to Figure 4.2, each panel has a gray region which indicates relative proper motions too low (or corresponding lens distances too high) to resolve a lens and source (¡2.5 mas/yr). To compare to Figure 4.2, each panel of Figure 2.9 also includes a purple region along the bottom shading all masses below 1.2 $M_\odot$, the highest possible undetectable stellar mass based on our NIRC2 image completeness. In all cases outside the gray regions, the range of lens masses allowed by the light curve fit is well above the range of stellar masses undetectable in our images, eliminating the possibility of a stellar lens in this parameter space. At the boundary of this region, we find that the lens masses indicated by a relative proper motion of 2.5 mas/yr are $5.30^{+1.14}_{-0.96}$ $M_\odot$ and $8.72^{+1.70}_{-1.46}$ $M_\odot$ for solutions A and B of M96-B5, and $2.67^{+0.53}_{-0.49}$ $M_\odot$ and $3.17^{+0.52}_{-0.48}$ $M_\odot$ for solutions A and B of M98-B6. As all of these are above even a conservative upper limit of neutron star mass of 2.16 $M_\odot$ [75], this eliminates the possibility of non-BH compact objects such as neutron stars and white dwarfs, as well as low mass, luminous objects such as brown dwarfs or free floating planets.

The only remaining exception is the possibility of relative proper motions low enough such that the lens and source would not have had adequate time to separate appreciably, in which case a potentially luminous lens would not be detectable in our NIRC2 images, regardless of mass. This possibility is represented by the gray regions at the bottom of each panel in Figure 4.2 and along the sides of each panel in Figure 2.9. To address this blindspot, Sections 2.5 and 2.6 defined and explored possible scenarios with photometric and astrometric analysis of WFPC2 images shortly following the events in 2000, the results of which are in Table 2.6. We discuss the implications and limitations of these results in the following section.

2.8 Discussion

Implications of Light Curve Fit Results

From an examination of the light curve fit parameters alone (in Table 2.4), we have some clues to the nature of the lens. We first note the exceptionally long $t_{\rm E}$, particularly in the case of M96-B5. Because $t_{\rm E}$ scales with the square-root of lens mass, these long timescales (as opposed to the days-long events caused by planets and months-long events caused by stars and lower-mass compact objects) suggest very high masses. Indeed, it would be difficult for lens stars below $\sim$0.6 $M_\odot$ (the masses defined as 'allowed' in the previous section) to result in microlensing events with timescales of hundreds of days.

If we pair our events' $t_{\rm E}$ values with another output of our fits, π_E (the 'microlensing parallax', defined in Equation 1.3), we can put them in the context of the simulations constructed in [76], which emulate the results of microlensing surveys while examining the

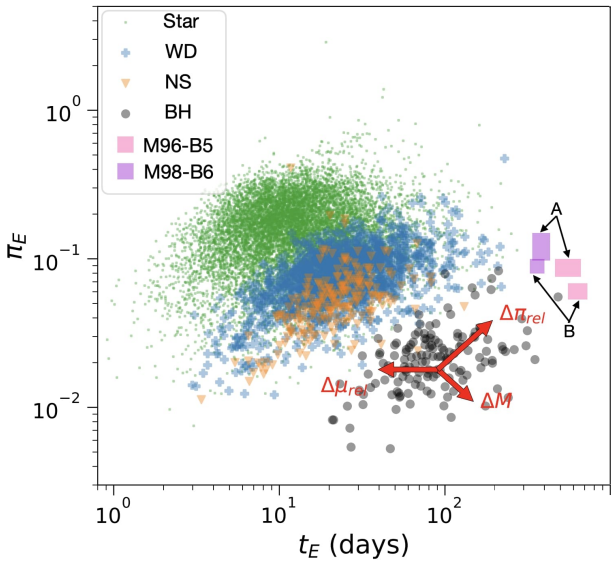


Figure 2.10: The main panel from Figure 13 in [76], showing the distributions of $t_{\rm E}$ nd π_E by lens type for a simulated microlensing survey. The one sigma ranges of these parameters from the results of this work's light curve fits are shown in magenta and purple boxes, for M96-B5 and M98-B6, respectively. The two different solutions for each event are marked with 'A' and 'B' arrows in black.

distribution of lens and source objects. One figure from this work, (replicated here as Figure 2.10), shows the distribution of star, white dwarf, neutron star, and black hole lenses as a function of $t_{\rm E}$ and π_E. The magenta and purple boxes have been added on the figure to show where our events (M96-B5 and M98-B6, respectively) fall in this distribution. This work reports that if searching for BH lenses in microlensing events, the BH detection rate will be 85% among events with $t_{\rm E} > 120$ days and $\pi_{\rm E} < 0.08$. Solutions B of M96-B5 meets this criteria fully, while solution B of M96-B5 and both solutions of M98-B6 meet the $t_{\rm E}$ cutoff, but have $\pi_{\rm E}$ values going up to 0.11.

In the context of the [76] result, the values of $\pi_{\rm E}$ and $t_{\rm E}$ for both events suggest a similarity to microlensing events simulated with compact objects lenses, though a stellar lens is not impossible. However, considering that the detection limits for this event indicate that a stellar lens would only be possible with a relatively low ($0.6 M_{\odot}$) mass, the fact that the event falls much closer to the high mass (lower right) corner of the distribution than the low mass (top right) corner of the distribution suggests that the lens is not a low mass main sequence star. Though this isn't an explicit elimination of a luminous, low mass lens in either case, it strongly suggests that the low mass stars left as potential lenses in our analysis are not likely lens candidates.

The possibility of low mass luminous sources are further diminished by the results shown

in Figure 2.9. For all relative lens-source proper motions above 2.5 mas/yr, the lens masses allowed by the light curve fit at any given proper motion or lens distance are well above the range of stellar masses undetectable in our NIRC2 images, effectively ruling out all stellar lenses in this region of parameter space. Further, we rule out any lens object in this parameter space with mass below $5.30^{+1.14}_{-0.96}$ $M_\odot$ and $2.67^{+0.53}_{-0.49}$ $M_\odot$ for M96-B5 and M98-B6 respectively, eliminating the possibility of a low-mass, non-stellar lens, including neutron stars, white dwarfs, brown dwarfs, and free-floating planets. In effect, there is no possibility for a non-BH lens in either event for relative proper motions above 2.5 mas/yr. We can reverse this calculation to limit the proper motions even further by recognizing that no luminous lens above 2.16 $M_\odot$ is possible (as the results in Table 2.6 show that an the maximum possible masses for a stellar lens don't exceed 1 $M_\odot$, and all non-stellar objects besides BHs would fall below that limit). Calculating the maximum proper motion allowed at this mass limit with Equation 2.8, we find that, using the more conservative result for each event, non-BH lenses are only possible at relative proper motions below 1.32 mas/yr and 2.48 mas/yr for M96-B5 and M98-B6, respectively. The remainder of this discussion evaluates the likelihood of a BH lens in the parameter space below these proper motion limits, in which the lens and source would be unresolved in our images.

Blend Flux Variability in Lens Flux Analysis

Light curve modeling of M98-B6 shows no indication of a luminous lens at all proper motions (though the 1-sigma error bar extends slightly into the parameter space of non-zero lens flux). This is not the case for M96-B5, which, for both solutions appear to have a non-zero amount of lens flux in our results. We note, however, that the result of this analysis is highly sensitive to the value of b_{sff} (which is calculated from the fit parameters f_s and f_b). In using this parameter as the basis of this analysis, we must consider the potential for degeneracies between this parameter, $t_{\rm E}$, and u_0. An event with long $t_{\rm E}$ but low values for u_0 and b_{sff} will be degenerate with a short $t_{\rm E}$ event with higher u_0 and b_{sff}.

We can look at previous fits in the literature to see this degeneracy manifest, such as the results for fitting the M96-B5 light curve to a PSPL model with parallax in [48]. This work reported three solutions for this event, one of which is similar to our best fit solution (long $t_{\rm E}$, small u_0 and b_{sff}), while the other two illustrated the opposite result (relatively short $t_{\rm E}$, with larger values of u_0 and b_{sff}). The values yielded in the latter two solutions for b_{sff} were 0.3 and 0.33 in MACHO-Blue and 0.28 and 0.31 in MACHO-Red. If applied to the lens flux analysis here, these results would put the lens for M96-B5 squarely in the region of no lens flux, which, paired with the long $t_{\rm E}$ and small π_E in the context of the [76] cited result above, strongly suggest a BH lens. To address this limitation of fitting (as noted previously, the lack of coverage of the peak magnification of this event introduces an inherent ambiguity), we explored a non-luminous lens scenario in addition to the two luminous lens scenarios for this target in Section 2.6, which will be further discussed below.

Similarly, by looking at the result obtained for a PSPL with parallax fit of M98-B6 in [51], we find three physical solutions for this event (excluding the fourth solution, which has

a negative, non-physical blending parameter). All three yielded lower $t_{\rm E}$ values than our reported ∼350 days, but b_{sff} values larger than ours (.76, .85, and 1). This only supports our Section 2.5 result that for this event, there is no measurable contribution of flux from a lens.

Interpreting Closest Matches in Target Photometry

Because of the degeneracy involving b_{sff} and the impact it has on our results, we took a conservative approach in using the analysis of Section 2.5 to inform the photometric examination of the source in Section 2.6 and considered both non-luminous and luminous lens scenarios for both events. Note that 'target' here refers to the unresolved object which contains at least light from the source star, and at most light from the source star and unidentified lens object.

For M96-B5, the scenario in which we consider the target as a single source (the scenario that the Bennett 2002 solution's high b_{sff} values would suggest), we find that the best match 1.41 $M_\odot$ G9 sub-giant at 8.5 kpc. This distance corresponds to a star in the galaxy's bulge, which is typically assumed for microlensing events observed in that direction. The two scenarios with a source and a luminous lens have similar objects matched to the source (both 1.41 $M_\odot$ G9 sub-dwarfs, at 8 kpc and a 10 kpc for solutions A and B), with low mass main-sequence stars matched to the lens (a K2 at 2.5 kpc and a K3 at 2 kpc, both with masses ∼0.7 $M_\odot$). Though the quality of the match for the lens+source scenarios appears to be better than the single source scenario (chi-squared values of 0.079 and 0.082 for the double objects versus 0.144 for the single object), if considered in the context of Equation 2.9, the combinations of lens distance, source distance, and lens mass of either of the double object matches is not allowed by corresponding light curve fit. As such, we report finding no combined source and luminous lens photometric matches that are consistent with all of our data for this event.

The same can be said about our results for M98-B6 (for which our analysis in Section 2.5 had already pointed to zero-lens flux, with a small possibility of a luminous lens). Though in all three scenarios, the matched source yields an object with temperature and distance consistent with the EDR3 effective temperature and parallax measurements, all lenses within our magnitude limits yielded masses too low to be allowed with the corresponding source and lens distances. Further, the single source is well described consistent with the classification of a previous work [53] which used optical spectral fitting to classify this source as a G5 sub giant.

Remaining Possibilities and Future Observations

In discussing what remains, we first examine the limitations of the synthetic photometry matching. As discussed previously, the potential matches are pulled from synthetic isochrones that include main-sequence and post-main-sequence stars. This does not account for the possibility of non-stellar lenses, including neutron stars, white dwarfs, and brown

dwarfs. Further, objects such as these may be missed by our lens flux analysis, as it was limited by observations in only optical wavelengths.

For M96-B5, this parameter space is narrowed down by several X-ray observations of the event, which found the X-ray emission to be consistent with a BH lens [77, 49]. This leaves white dwarfs, brown dwarfs, and free floating planets as possible remaining lenses for this event. By Equation 2.8, a white dwarf lens at the Chandrasekar limit of 1.4 $M_\odot$ would necessitate a relative proper motion of $0.66^{+0.15}_{-0.12}$ mas/yr for solution A and $0.40^{+0.08}_{-0.07}$ mas/yr for solution B. Without X-ray observations, our upper limit on relative proper motion is set by a conservative maximum neutron star mass of 2.16 $M_\odot$ [75], which yields relative proper motions of $2.02^{+0.46}_{-0.34}$ mas/yr and $1.71^{+0.31}_{-0.24}$ mas/yr. By take the upper limit of the more conservative solution for each event, we conclude that only a BH lens is possible for relative proper motions above 0.81 mas/yr for M96-B5 and 2.48 mas/yr for M98-B6.

The results presented here indicate that both M96-B5 and M98-B6 remain extremely good candidates for BH lenses; though we cannot make a solid confirmation, multiple methods of examination have yielded no compelling alternatives for either event. There remains, however, additional observations and analysis that could illuminate the true nature of the lenses further. Primarily, we are concerned with eliminating the possibility of non-BH compact object or brown dwarf lenses unresolved from their source stars. This could be done by spectroscopic examination of the source and searching for a second object in its spectrum, similar to this work's analysis of broad band photometry. The wavelength of observation would depend on the potential lens: near-IR for brown dwarfs, optical or UV for white dwarfs, and UV or X-ray for neutron stars. Measurement of the source spectrum would yield a secondary benefit for M96-B5: an estimation of the source distance from spectral typing (and perhaps a better constrained distance than that from the Gaia parallax for M98-B6). This would allow for better constraints on our results, which use source distances based on photometric fitting. These straightforward observations would conclusively determine if anything remains in the narrow region of parameter space in which a BH is not the only possibility.

2.9 Conclusions

Through the analysis of high-resolution images and light curve data, we have eliminated the possibility of a non-BH lens for relative lens-source proper motions above 0.81 mas/yr for M96-B5 and 2.48 mas/yr for M98-B6. To address the potential for an unresolved, luminous lens, our comparison of source images to synthetic photometry indicate that for both events, a single stellar source is a better fit than a source blended with a stellar lens, eliminating the possibility of a stellar lens at any proper motion. We discussed the unlikely, but physically possible scenario of brown dwarf or non-BH compact object lenses with extremely low relative proper motion, and described how this remaining possibility can be constrained by future observations and analyzed with the methods developed in this work.

2.10 Additional Figures and Tables

Table 2.7: M96-B5 Astrometric Measurments

Epoch	Instrument	Filter	RA Offset (mas)	Dec Offset (mas)
1999.452	WFPC2	F555W	53.1 ± 2.3	14.8 ± 0.9
1999.452	WFPC2	F814W	53.1 ± 2.0	16.3 ± 1.2
2000.444	WFPC2	F555W	49.3 ± 1.4	15.5 ± 2.3
2001.419	WFPC2	F555W	46.3 ± 1.4	16.4 ± 4.0
2001.419	WFPC2	F814W	47.5 ± 0.5	14.5 ± 0.3
2001.747	WFPC2	F555W	45.4 ± 3.9	13.8 ± 3.4
2001.750	WFPC2	F814W	47.5 ± 0.9	14.0 ± 0.9
2002.394	WFPC2	F555W	43.2 ± 2.8	16.9 ± 3.1
2002.394	WFPC2	F814W	43.5 ± 1.9	15.7 ± 1.3
2002.750	WFPC2	F555W	43.1 ± 1.8	11.8 ± 1.8
2002.750	WFPC2	F814W	43.2 ± 1.0	13.2 ± 1.0
2003.400	WFPC2	F555W	39.9 ± 1.1	15.6 ± 1.3
2003.400	WFPC2	F814W	40.3 ± 0.6	14.6 ± 0.7
2016.535	NIRC2	Kp	2.07 ± 0.10	0.49 ± 0.20
2017.387	NIRC2	Kp	0.00 ± 0.13	0.00 ± 0.21

Position of M96-B5 from 1999 to 2017. RA and Dec offset are measure in miliarcseconds, where the final epoch is placed at 0,0.

Table 2.8: M98-B6 Astrometric Measurments

Epoch	Instrument	Filter	RA Offset (mas)	Dec Offset (mas)
2000.476	WFPC2	F555W	49.3 ± 2.3	2.5 ± 6.1
2000.476	WFPC2	F675W	36.4 ± 4.1	-5.9 ± 2.0
2000.476	WFPC2	F814W	41.7 ± 1.6	-5.2 ± 2.5
2015.500	Gaia	G	0.0 ± 0.1	0.0 ± 0.1
2016.535	NIRC2	Kp	-1.9 ± 0.5	1.6 ± 0.6
2017.436	NIRC2	Kp	-4.7 ± 0.5	3.1 ± 0.8

Position of M98-B6 from 2000 to 2017. RA and Dec offset are measure in miliarcseconds, where the Gaia epoch (2015.5) is placed at 0,0.

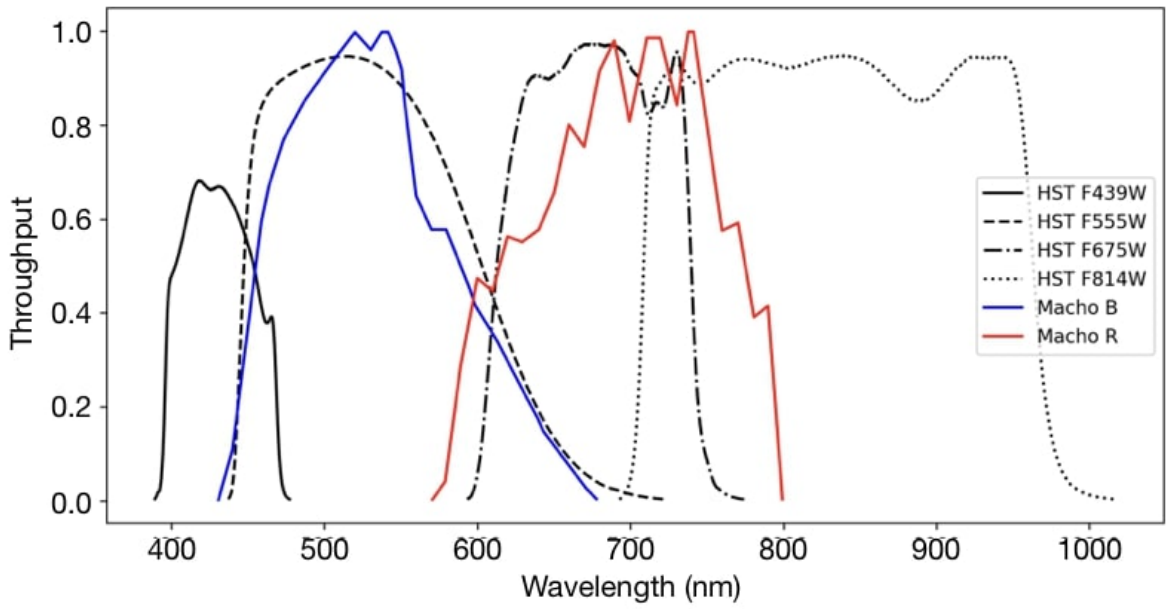


Figure 2.11: Transmission curves for the two light curve filters (MACHO-Blue and MACHO-Red) compared to the WFPC2 filters used in the analysis of f_l. We compare MACHO-Blue to F555W and MACHO-Red to F675W.

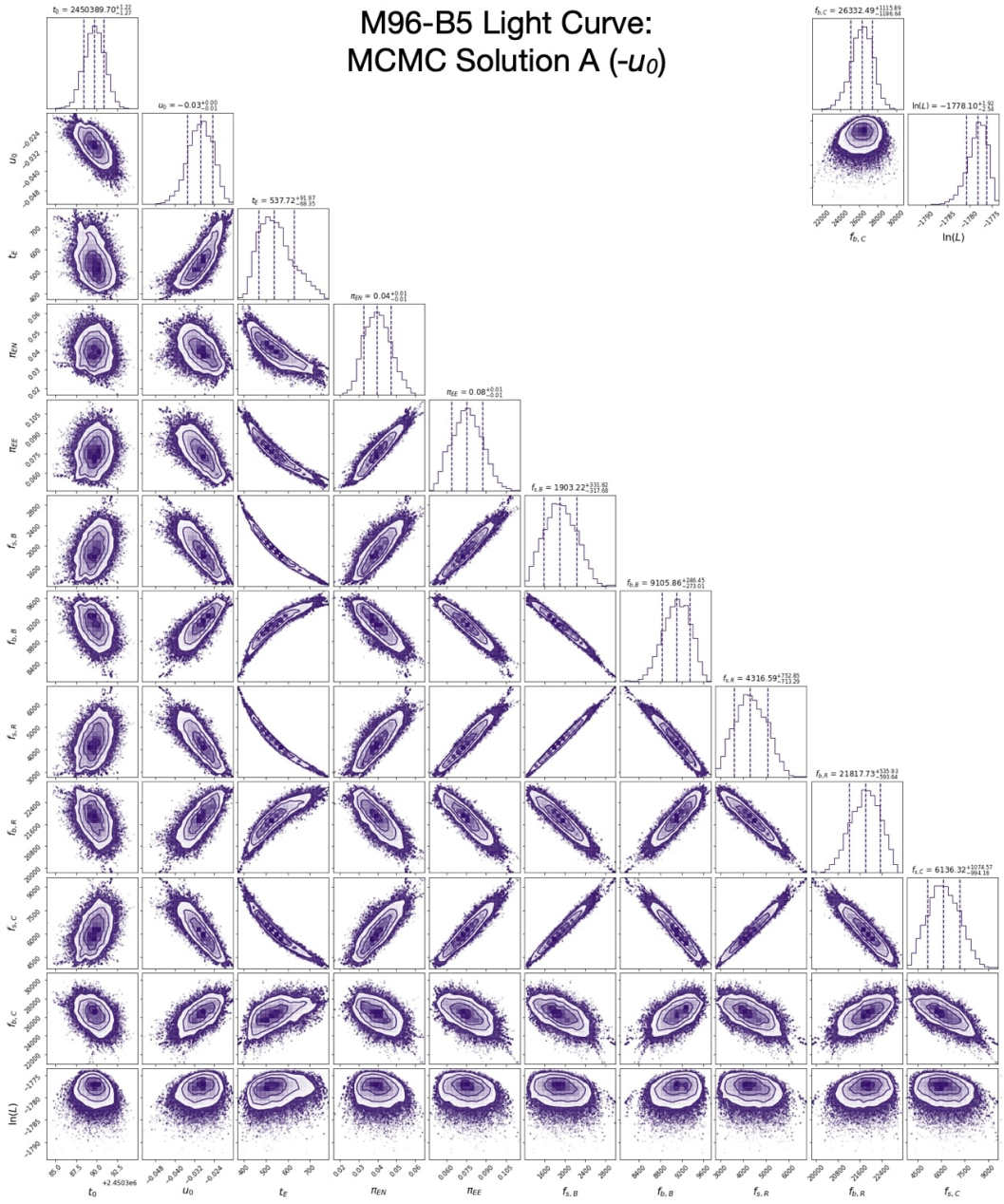


Figure 2.12: Posterior distributions from one of two MCMC solutions for the light curve of M96-B5, here with negative u_0. In addition to the 11 parameters defined for Table 2.4, there is an additional row showing the log-likelihood $\ln L$, defined here as $-0.5\chi^2$. Note that the rightmost two columns have been re-positioned to fit the figure space.

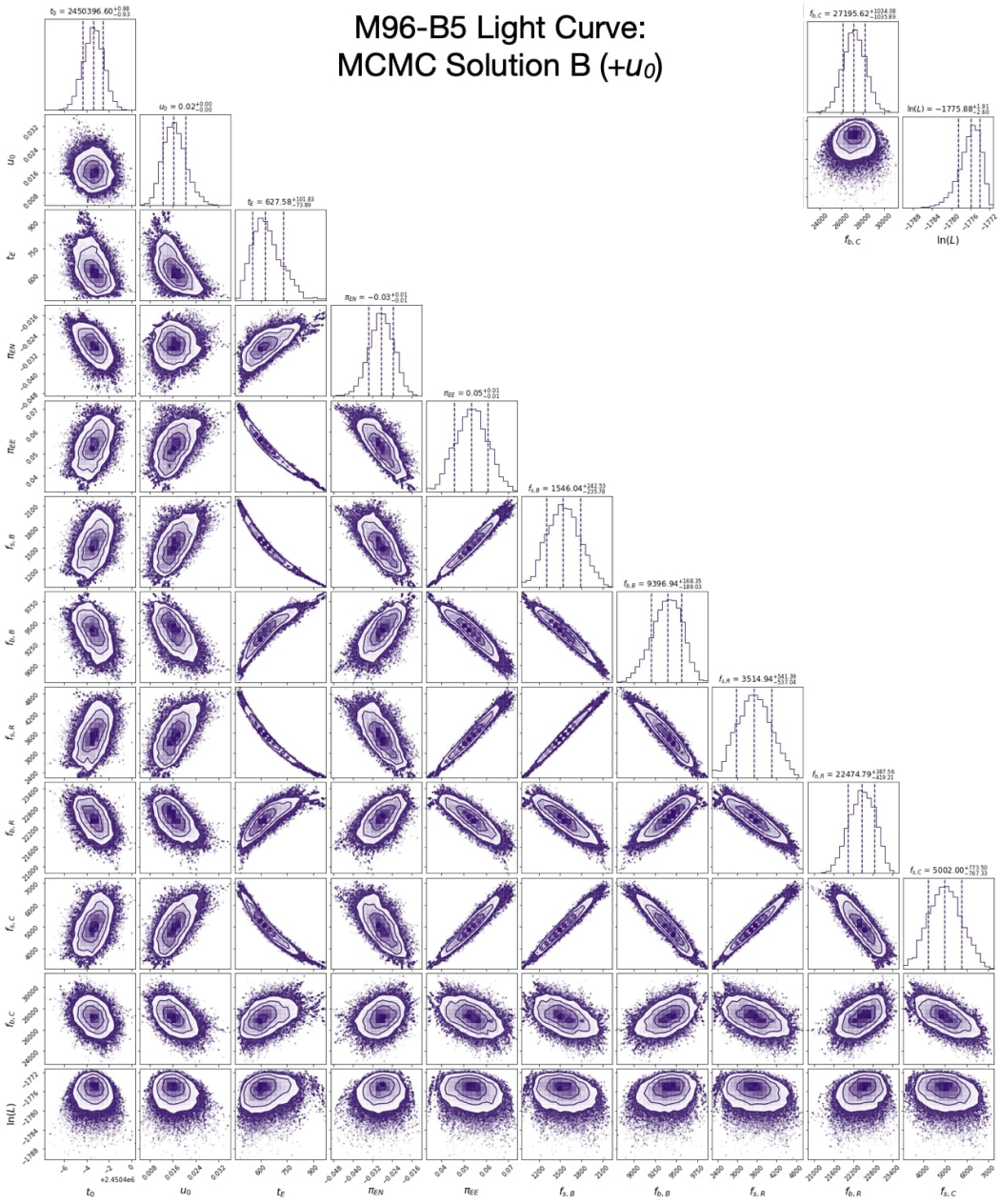


Figure 2.13: Posterior distributions from one of two MCMC solutions for the light curve of M96-B5, here with positive u_0. In addition to the 11 parameters defined for Table 2.4, there is an additional row showing the log-likelihood $\ln L$, defined here as $-0.5\chi^2$. Note that the rightmost two columns have been re-positioned to fit the figure space.

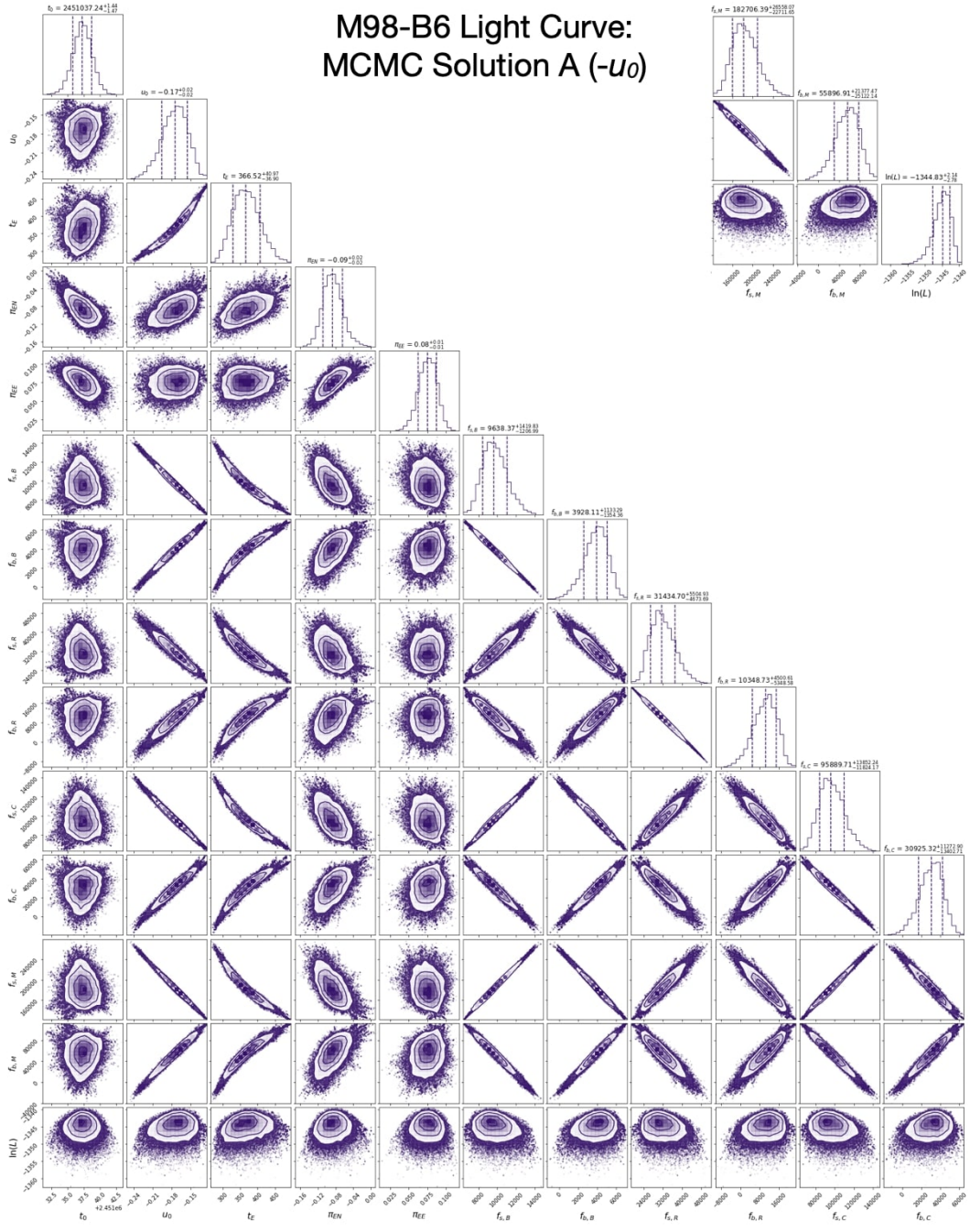


Figure 2.14: Posterior distributions from one of two MCMC solutions for the light curve of M98-B6, here with negative u_0. In addition to the 13 parameters defined for Table 2.4, there is an additional row showing the log-likelihood $\ln L$, defined here as $-0.5\chi^2$. Note that the rightmost three columns have been re-positioned to fit the figure space.

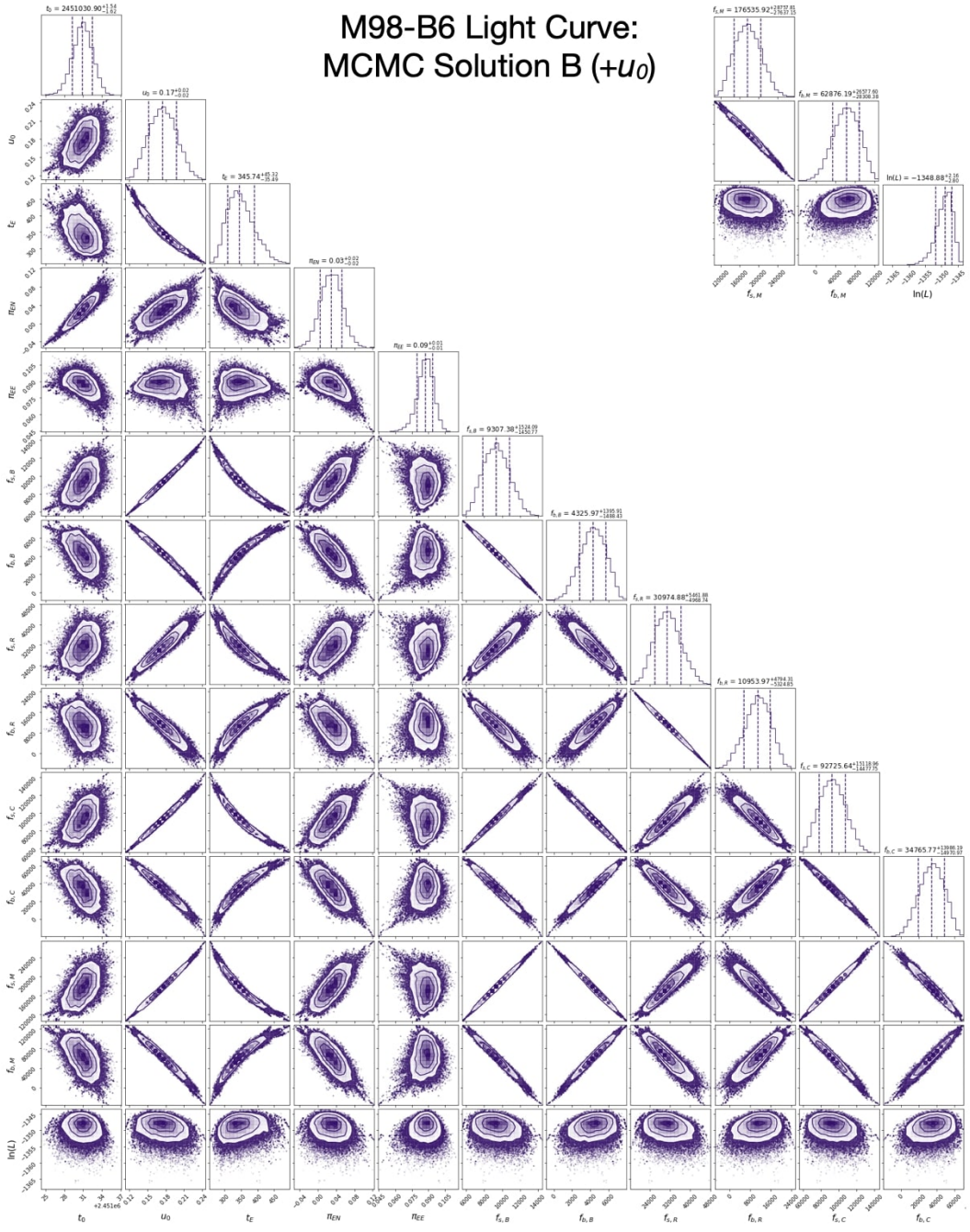


Figure 2.15: Posterior distributions from one of two MCMC solutions for the light curve of M98-B6, here with positive u_0. In addition to the 13 parameters defined for Table 2.4, there is an additional row showing the log-likelihood $\ln L$, defined here as $-0.5\chi^2$. Note that the rightmost three columns have been re-positioned to fit the figure space.

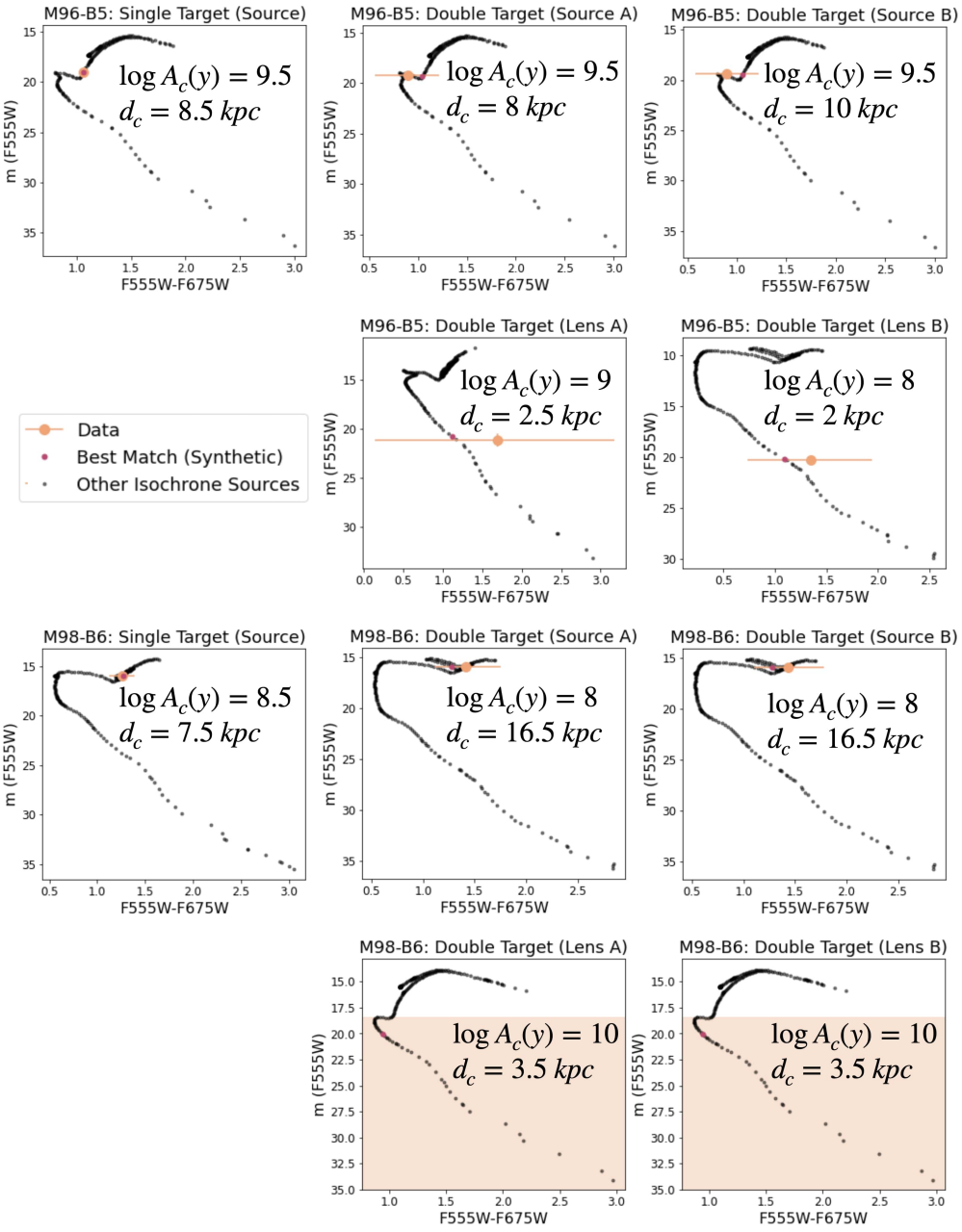


Figure 2.16: For each of the matched synthetic objects in Table 2.6, the isochrone from which that object was pulled is illustrated in a single panel here, with the matched source shown in magenta and all other sources in black. The corresponding observed data point is shown on each panel in orange with errorbars. For the two instances where we have only a limit on the object magnitude (both lenses for M98-B6), the range allowed by the data is shaded in orange. For each match, the log of the cluster age in years and the cluster distance in kpc is reported.

Chapter 3

Improved Image Quality Over 10' Fields with the 'Imaka Ground Layer Adaptive Optics Experiment

3.1 Introduction

'Imaka is a GLAO demonstrator on the University of Hawaii 2.2-meter telescope on Maunakea aimed at quantifying the potential of the site for GLAO. The full 'imaka 24'×18' field is an order of magnitude larger than its MCAO predecessors and is the largest GLAO field of view (FoV) deployed to date. Simulations for 'imaka predicted a PSF full width at half maximum (FWHM) of ∼300 mas [78] in the near-infrared, which is larger than the FWHM delivered by SCAO systems that can reach ∼30 mas at similar wavelengths [79]. As such, GLAO should be thought of not as a replacement for high-fidelity wavefront correction, small FoV AO systems, but as a distinct, wide-field, "super-seeing" system with a different set of advantages and science applications.

The science made possible with GLAO improvements are numerous. The wide field of view of 'imaka allows large, crowded stellar fields to be imaged at higher resolution than normally allowed by the natural seeing. The increased acuity over wider fields can improve astrometric measurements significantly, providing more candidates for reference stars. The observation of large resolved objects such as nearby galaxies, with stable, highly corrected PSFs will also allow more detailed study of topics such as galaxy morphology. Besides creating sharper images, the increased encircled energies obtained by GLAO correction produce a higher signal-to-noise ratio (SNR) for an otherwise identical exposure, a significant advantage for imaging as well as spectroscopy. Furthermore, 'imaka demonstrates the promising possibility of GLAO on a wide variety of telescopes of different sizes on Maunakea, including ten-meter class telescopes such as Keck as well as the forthcoming generation of extremely large telescopes.

Commissioning of 'imaka began in October 2016. Analysis of the data from the three

subsequent GLAO obsering runs in 2017 are presented in this chapter. We first present a brief description of the instrument (§2). Then we describe the different data sets used and their observations (§3), and their reduction (§4). We then define metrics used to characterize these data, including PSF modeling (§5). Analyses of the GLAO performance, including an examination of wavelength dependence and field variability, are presented (§6) before we finally discuss the implications of our findings (§7).

3.2 Instrument Description

'Imaka is a demonstrator of wide-field GLAO with its primary goal of providing an on-sky facility to gain experience using and optimizing the performance of GLAO systems. 'Imaka is mounted at the Cassegrain focus of the University of Hawaii 2.2 meter telescope on Maunakea. The instrument is built around a modified Offner optical relay (e.g., [80]) with input and output focal ratios of f/10 and f/13.25 respectively. A schematic layout of the system is shown in Figure 3.1. This optical relay provides a flat pupil position just before the second mirror (AOM2) where we place a curvature deformable mirror consisting of 36 elements. Note that the deformable mirror is conjugate to the telescope primary mirror (i.e., the ground). All powered optical elements in the system are spheres, which simplifies alignment and fabrication. An "exit" port near the relayed focal plane provides mounting points for the wavefront sensors and the science camera.

We place the wavefront sensors and the science camera at the exit port of the instrument where an oval field 24' ×18' is passed through the system. We can place up to five Shack-Hartmann wavefront sensors anywhere within this field but we generally observe fields with guide stars within the periphery of the field to avoid vignetting the science camera. Each wavefront sensor consists of its own optics and camera (Raptor Photonics Kite EMCCD cameras) configured to provide 8×8 sub-apertures with 0.4" pixel^{-1} sampling and a sampling rate up to 180 Hz. To date, the guide stars used are brighter than R $\leq$ 8.5 mag and we achieve a 0-dB rejection bandwidth of about 15 Hz.

The science focal plane is located just behind the wavefront sensors. Image quality from the optical relay was optimized within an approximately 11'×11' central field and the instrument can accommodate a large 11'×11' CCD camera or a 7'×7' near-infrared H4RG-15 camera. Both of these cameras are still in development so all images to date were taken with a Finger Lakes Instrumentation ML50100 CCD camera. The camera is mounted slightly off the center of the 'imaka science field and can be mounted in a variety of positions in order to mosaic a larger field of view. The FLI detector consists of 8176×6132, 6μm pixels with a platescale of 40 mas pixel^{-1} at the 'imaka focal plane. The field of view of an single image from the camera was limited by a band-pass filter mounted to the front of the camera to 5.4' × 4.1'.

A summary of the instrument specifications is provided in Table 3.1 and further details can be found in [78] and the references therein.

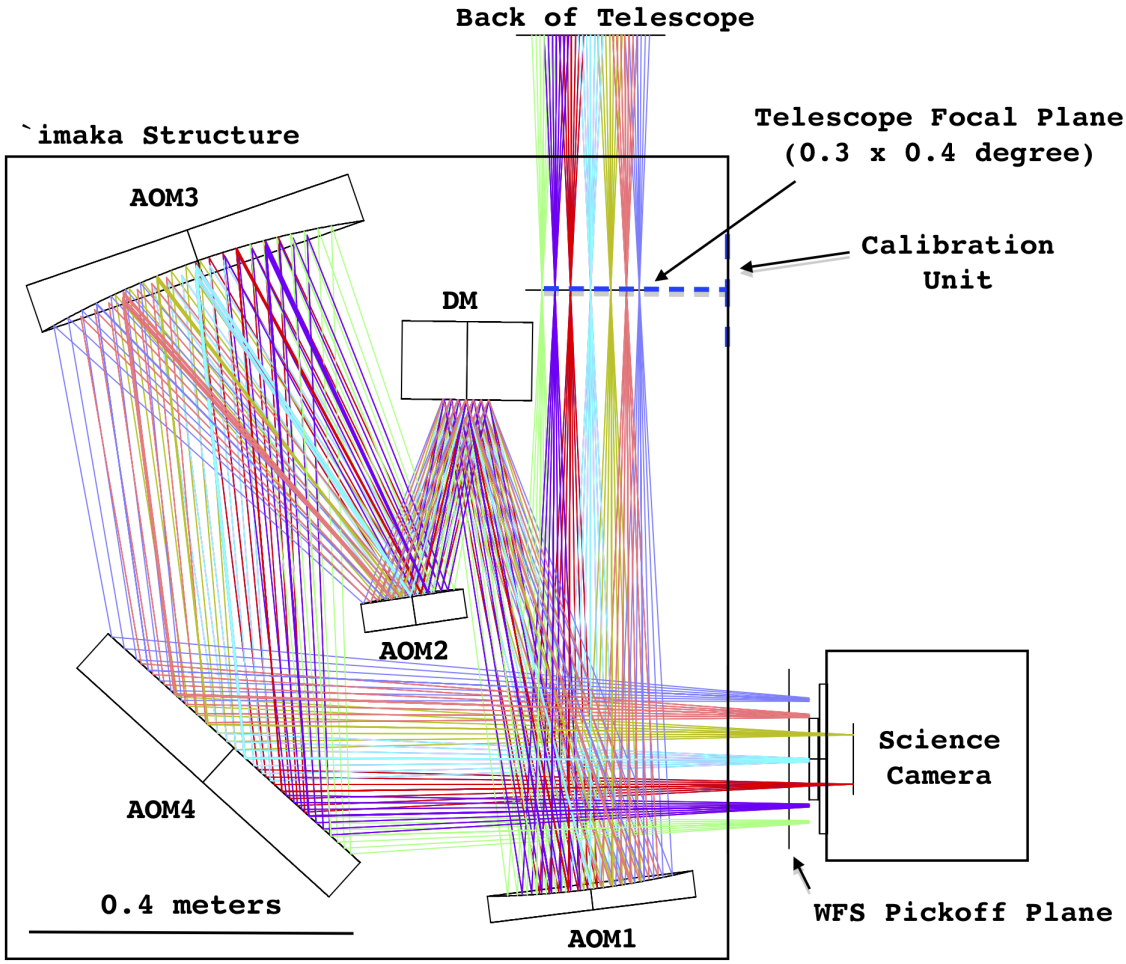


Figure 3.1: Layout of main elements of the 'imaka instrument. The light from the telescope enters from the top of the diagram. The first powered element (AOM1) creates a pupil image where the deformable mirror (DM) is placed. Following the DM are the Offner secondary (AOM2) and the final concave sphere (AOM3). A large flat (AOM4) folds the beam to a side exit port where the wavefront sensors and science camera are mounted.

Table 3.1: 'imaka AO specifications

Telescope	University of Hawaii 2.2m, Maunakea
Deformable Mirror	
Type	36-element, CILAS curvature mirror
Pupil Size	57.5 mm pupil
Wavefront Sensors	
Number	up to five Shack-Hartmann WFSs
Patrol field	24'x18'
Pupil Sampling	8x8 subaperture
Pixel size	0.4"/pixel
Subaperture Field	4.8'/subaperture
Wavefront Sensor Cameras	
Type	Raptor Photonics Kite EMCCD
Wavelengths	λ = 0.4-0.7 μm
Quantum Efficiency	peak 50%
Readout time	4.2msec
System	
Sampling Rate	up to 180Hz
Rejection Bandwidth	$f_{0dB} \sim 10 - 15$ Hz

3.3 Observations

The data represented in this chapter span three 'imaka observing runs: 2017 January 10-14 UT, 2017 February 13-17 UT, and 2017 May 17-22 UT. Data from 2017-01-13 UT were omitted in the analysis due to the MASS instrument not running that night. Our analysis combines three different types of observations, described below[1]

'Imaka Focal Plane Images

The fields' positions and guide star coordinates for each run are given in Table 3.2. The two fields used are shown in Figure 3.2. For the 2017 January and 2017 February runs, we observed the Pleiades star cluster (Field 1). This field has four guide stars, each positioned around the periphery of the science field. For the 2017 May run, we switched to a second field in the galactic plane (Field 2) in order to observe a higher density of stars. This field differed slightly in that it had its fourth guide star in the center of the field, though this star was not always used in the correction (refer to Table 3.3). For Field 2, three different guide star and WFS configurations were tested: all four (4WFS) and three outer WFS 0, 1 and 2 (3WFS). For further technical details, refer to [78].

[1]All of the data from 'imaka commissioning are available upon request.

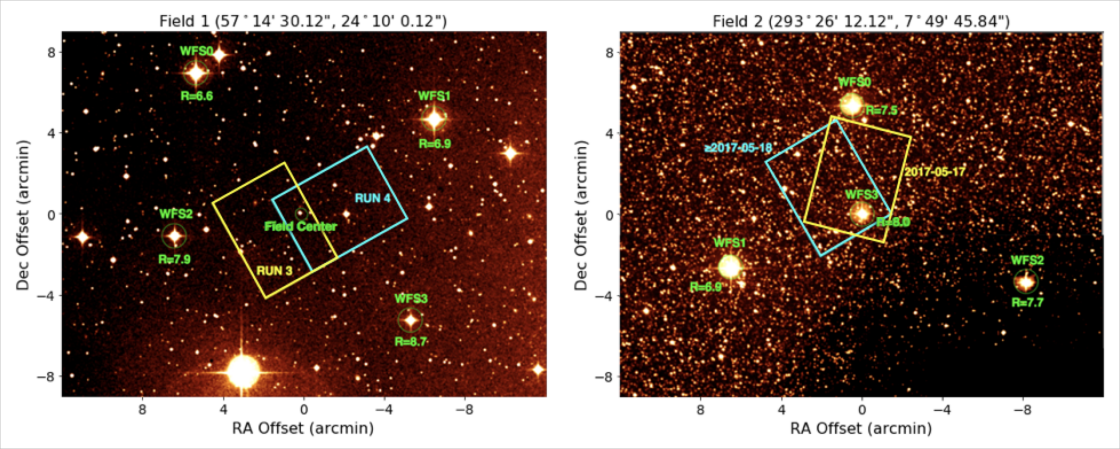


Figure 3.2: Field 1 (left; 2017 January and 2017 February) and Field 2 (right; 2017 May). In each field, all guide stars used are marked with wave front sensor number and R-band brightness. In Field 1, the east Pleiades field is shown in the yellow rectangle labeled RUN 3, while the west Pleiades field is shown in blue and labeled RUN 4. Positions 1 and 2 of Field 2 are shown in yellow and blue, respectively. In both figures, North is up and East is to the left. Background images retrieved from 2MASS [81].

In order to quantify the gains of 'imaka, we made observations both with GLAO correction (*AO-on*) and with no correction (*AO-off*). To capture both modes in a range of seeing conditions, we cycled through different modes while taking data, (i.e., one image with no correction, one image with GLAO correction, then back to no correction). Rather than either freezing or resetting when the mode switched from AO-on to AO-off, the DM was set to an average of the last 5-10 seconds at the beginning of each AO-off exposure. On average, each night produced 36 frames in each mode.

Science images were taken with exposure times of 30-60 seconds; these times were chosen in order to average over the seeing. We used I-band and R-band Johnson filters (centered at 806 nm and 658 nm, respectively), and a "long-pass" filter with an effective wavelength of approximately 1 micron. Image position and filter information is listed for each night in Table 3.3. In addition to our science images, we took dithered sky images throughout each night, as well as twilights at the beginning of each run for use as flat fields.

'Imaka Telemetry

In order to evaluate 'imaka performance, we also recorded the AO system wavefront sensor measurements and deformable mirror commands. This data set (the "telemetry") provides a means to independently estimate the uncorrected aberrations, the level of correction, and the atmospheric seeing. The telemetry data is taken concurrently with our science images;

Table 3.2: Field and Guide Star Positions

Label	Name	RA (J2000)	Dec (J2000)	R mag
Field 1 (January, February)				
Field Center	-	3h48m58.0s	24°10'00.1"	-
WFS0	HD 23873	3h49m21.7s	24°22'54.4"	6.6
WFS1	HD 23763	3h48m29.6s	24°20'48.9"	6.9
WFS2	HD 23886	3h49m25.9s	24°14'51.7"	7.9
WFS3	HD 2377	3h48m34.8s	24°10'52.3"	8.7
Field 2 (May)				
Field Center	-	19h33m44.8s	7°49'45.8"	-
WFS0	HD 184362	19h33m46.5s	7°55'05.0"	7.5
WFS1	HD 184451	19h34m11.1s	7°47'08.8"	6.9
WFS2	HD 184244	19h33m12.1s	7°46'22.0"	7.7
WFS3	HD 184336	19h33m44.8s	7°49'45.7"	8.0

however, the total recording time for the telemetry stream is limited to ∼10 seconds, while the science exposures are 3-6 times longer. The WFS telemetry data is taken without a filter and has an effective wavelength similar to a V+R-band filter.

The telemetry is used to reconstruct the seeing at a range of altitudes using SloDAR techniques [82], although turbulence profiles are not used in this chapter. Instead, the telemetry is used to simply derive the integrated, ground layer and free-atmosphere seeing independently from MASS/DIMM. Telemetry data were acquired with no AO correction and are used to compute the slope covariance matrices and maps. These are global in the sense that the covariances are computed across different wavefront sensors (the cross-covariances) as well as with themselves (the auto-covariances). A full description of the telemetry data and its analysis is described in [83]. The result of the telemetry processing is the estimated seeing for the ground-layer and the free-atmosphere for both AO-off and AO-on images.

MASS/DIMM

In addition to our own measurements, we use the seeing reported by the Mauna Kea Weather Center[2] (MKWC). The MKWC reports seeing measured by the Mauna Kea seeing monitor (MKAM) installed on a seven-meter tall tower between the Canada-France-Hawaii telescope and the Gemini-North telescope. MKAM is approximately 150 meters North of the UH88" telescope. The seeing monitor consists of a Multi-Aperture Scintillation Sensor (MASS) and a Differential Image Motion Monitor (DIMM) [84]. While the DIMM measures integrated seeing of the whole atmosphere, the MASS generates a profile of seeing at a range of altitudes (0.5, 1, 2, 4, 8, and 16 km). Note that the MKWC MASS data is processed with version v2.047 of the MASS software. However, there are improvements to the MASS data analysis

[2] `http://mkwc.ifa.hawaii.edu`

Table 3.3: Observing Runs

Date (UT)	PA	Band	N	AO-Off (mas)	AO-On (mas)	DIMM_λ (mas)	MASS_λ (mas)	DIMM_{500} (mas)	MASS_{500} (mas)
1-10	119	R	20	535 ± 17	313 ± 20	426 ± 7	164 ± 2	450 ± 8	173 ± 2
1-11	119	R	212	931 ± 20	645 ± 10	658 ± 13	643 ± 18	695 ± 14	680 ± 19
1-12	119	R	157	906 ± 13	628 ± 7	910 ± 16	701 ± 18	962 ± 17	741 ± 19
1-14	119	R	61	621 ± 15	444 ± 14	681 ± 17	560 ± 28	720 ± 18	592 ± 30
01-14	29	I	84	790 ± 13	587 ± 11	710 ± 16	414 ± 12	720 ± 18	455 ± 14
2-14	29	I	56	589 ± 17	433 ± 12	522 ± 12	220 ± 22	574 ± 13	242 ± 24
2-15	29	I	45	634 ± 42	411 ± 24	561 ± 24	184 ± 5	618 ± 27	202 ± 6
2-16	29	I	14	896 ± 28	498 ± 17	1006 ± 91	160 ± 8	1106 ± 100	176 ± 9
2-17	29	$1\mu m$	59	663 ± 52	377 ± 23	445 ± 11	297 ± 21	511 ± 12	341 ± 24
2-18	29	R	13	552 ± 22	370 ± 21	220 ± 25	112 ± 6	233 ± 27	118 ± 6
2-18	29	I	10	661 ± 66	321 ± 8	193 ± 30	110 ± 6	213 ± 33	121 ± 6
5-17	76	I	28	882 ± 14	332 ± 4	368 ± 12	112 ± 3	405 ± 13	124 ± 3
5-18	120	I	68	587 ± 9	398 ± 2	393 ± 9	208 ± 5	433 ± 9	229 ± 5
5-19	120	I	109	1078 ± 25	535 ± 2	581 ± 11	399 ± 11	639 ± 12	439 ± 12
5-20	120	I	55	712 ± 9	4290 ± 2	676 ± 16	261 ± 6	744 ± 18	288 ± 7
5-21	120	I	109	644 ± 5	368 ± 1	540 ± 12	115 ± 3	595 ± 13	127 ± 4

Observation setup, conditions, and results summarized with nightly averages. All dates are in 2017. Fields and position angle (PA; measured East of North to the positive y-axis of the detector) refer to those outlined in Figure 3.2 and Table 3.2. The number of frames is indicated by the N column. The reported FWHM is the minor FWHM of a Moffat profile. As MASS/DIMM seeing is measured at 500nm, we present both a converted DIMM_λ and MASS_λ that are scaled with equation 3.8 to match the observation wavelengths, in addition to the raw data in the DIMM_{500} and MASS_{500} columns. The nights of 01-14 and 02-18 UT each have two entries, for the two half-nights in which we observed with two different filters. In the May run, the combination of wavefront sensors changed from all four to only the outer three with different configurations. The May run is also the only data collected for Field 2; all other epochs are observing Field 1.

[85] and we have reprocessed the data using the most recent version of the software [86]. Note that the data reprocessing still uses the MKAM MASS calibrations from 2009. Throughout this work, we will refer to the MASS seeing as the integrated seeing in the entire MASS profile (e.g. from 500 m to 16 km).

Both instruments report seeing at 500 nm throughout the night. Every night of focal plane data has corresponding MASS/DIMM measurements with the exception of 2017-01-13 UT when the MASS was not working. We exclude this night from the rest of our analysis.

It should be noted that the MASS/DIMM instruments do not necessarily observe the same atmospheric profile as 'imaka. Ideally, the telemetry's integrated seeing should be comparable to the DIMM, while the free atmosphere seeing should match the MASS. However, the MASS/DIMM points close to zenith, while we observe at a range of altitudes. The differing paths through the atmosphere may result in different turbulence profiles. Additionally, the instruments are located at different places on the summit of MaunaKea and therefore experience different local seeing (e.g., seeing very close to the ground at the seeing monitor and dome seeing at the UH 2.2 m telescope).

3.4 Image Reduction

Standard image processing was applied to the science images including flat-fielding, sky subtraction, and bad-pixel and cosmic ray rejection. The flat-field image was created from several twilight exposures with different integration times (2-30 seconds) that were normalized by dividing by the σ-clipped median of all pixels in the image and then combined using a σ-clipped median. In some cases where twilight images were not taken on a given night, the flat generated from the preceding night was used. Sky images were taken periodically through the observing run each night, usually with a total of about 20 frames. Like the flats, all sky images of a given night were combined using a σ-clipping median routine that, along with a dithering procedure, removed any stars from the final combined sky image.

In addition to individual science frames, we created stacked images. For each night, one stacked image was created using the reduced images for each position, control matrix, and filter (e.g., all AO-on images on 2017-05-17 UT taken at I-band using the combination of wavefront sensors 0, 1 and 3 were combined to form one image). The frames were matched and had relative astrometric transformations applied before adding them together to create single images, without further correction (e.g., for distortion, differential refraction, etc). Because these stacked images do not capture the effects of seeing variations or zenith distance which are changing between frames, they are used only to examine the variability in AO correction across the field due to the instrument. As such, data from stacked images are only presented in §6.4 on field variability.

3.5 Image Quality Analysis

After reducing the science images, each frame was run through our star finding algorithm. Initial FWHM estimates were obtained by eye from the cleaned science images, typically 0.4" in AO-on and 1.0" in AO-off images. These initial guesses were used in DAOphot's DAOstarfinder [87], to locate stars in the images. The star finder performed two passes with this initial PSF guess, locating sources above four standard deviations of the mean pixel value of the image and updating the FWHM guess accordingly. The images of the Pleiades field yielded about 15 sources each (so few because of the sparseness of the field, the short exposure times, the rejection of saturated sources, and the high SNR needed for the subsequent fitting), while the second, much denser field of the May run had 300-500 sources each.

Once the sources were located, certain cuts were made to the sample of stars. First, stars within 20 pixels of image edges or other stars were removed. Second, saturated stars (defined as anything with a peak above 20,000 ADU) were removed. Finally, we defined a minimum flux threshold for each image. Generally, this cut saved only the brightest 80% of stars in a frame. However, for Field 1, which had few stars to begin with, no flux cut was made.

Once a clean star sample was defined, we characterized the PSF of each star in a given frame using several metrics described below. To further reduce our data, all of our statistics described below were median combined for all stars in a given image, resulting in a single measurement of a given metric for each frame. Thus, each data point in all subsequent plots represents a single frame unless otherwise stated. Note that we did not apply a correction for zenith distance of observations; Figure 3.15 shows our FWHM data as a function of airmass fit to both a constant FWHM and a 3/5 power law, and illustrates that no or very little airmass dependence is seen in our data.

Non-Parametric PSF Models

Before establishing a parametric model for the image plane PSFs, we took a numerical approach to their characterization using three different metrics, described below. Each metric quantifies the PSF shape and size slightly differently; thus the values vary for the same image and can be seen in Figure 3.3.

Empirical Full Width Half Max: Pixels were counted in the vicinity of a star's centroid with ADU counts above half of the maximum pixel value as a brute-force method of calculating a star's FWHM. Assuming this area is circular, we calculated the corresponding diameter: the empirical full width half max.

Encircled Energy Diameter: To obtain a more detailed profile of the sources' PSFs, we also calculated encircled energy (EE) diameters for our sources. To do this, we created brightness profiles of each source, counting flux contained within a series of concentric annuli centered on the source's centroid. From this profile, we measured the diameter within which 25%, 50%, and 80% of the total flux was enclosed.

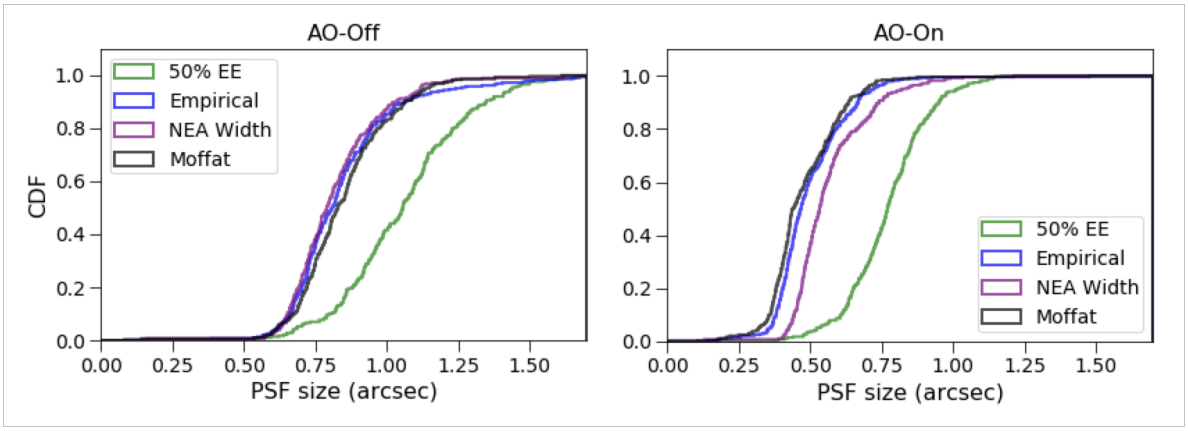


Figure 3.3: Cumulative distribution functions illustrating different PSF metrics, including the 50% encircled energy diameter, empirical FWHM, NEA width, and Moffat FWHM. Though they all approach the PSF characterization slightly differently and thus yield different results, we expect them to be somewhat consistent in the range of values measured on a given dataset. *Left:* AO-off. *Right:* AO-on. Note that only I-band data is shown here; the four metrics in other filters demonstrate similar differences.

Noise Equivalent Area: An additional way we measure the PSF is a "noise equivalent area" (NEA). A full derivation of the NEA of an arbitrary PSF can be found in [88], but simply put, it is the area of a region centered on a star's centroid in which the signal-to-noise ratio (SNR) is unity. This is calculated with a similar method to the encircled energy radius, in which concentric circular regions of increasing size around a star are isolated until reaching an SNR of 1. The NEA of a Gaussian profile, for example, is $4\pi\sigma^2$. For the sake of comparison to PSF measurements that are widths rather than areas, we define the 'NEA width' to be the diameter corresponding to a circle with area NEA.

Parametric PSF Modeling

Though the above metrics provide a good sense of image quality, they are not a complete description of stars' PSFs. These models ignore the shape of the source, which is problematic, as our sources are not radially symmetric. This is particularly apparent in AO-off images, where stars show significant elongation. As such, we parametrically modeled the sources' shape.

We tried a variety of models, including bivariate Gaussian, Lorentz, Moffat, and sinc distributions. Additionally, we tried two component models, including combinations (e.g. Gaussian + Moffat). For each model, the goodness of fit was quantified with the fraction of

variance unexplained:

$$FVU = \frac{\sum_i^{N_{pix}} [PSF_{obs,i} - PSF_{mod,i}]^2}{\sum_i^{N_{pix}} [PSF_{obs,i} - \overline{PSF_{obs}}]^2} \tag{3.1}$$

This value was calculated in a box centered on the star with a side length of about 4 times the FWHM. Using this quantity to compare different models, we found that for AO-off images, the best model was a single component elliptical Moffat profile. We had predicted that for the AO-on images, a two component elliptical Moffat would be ideal; a narrower, brighter component representing the corrected portion of the final image, and a dimmer component with a size similar to that of the AO-off image, representing the uncorrected portion of the PSF. However, as even a carefully constrained double component model showed no improvement in the FVU, we ultimately chose a single component Moffat for AO-on images as well.

The model for a two dimensional elliptical Moffat with rotation is defined as:

$$f(x,y) = a[1 + A(x-x_o)^2 + B(y-y_o)^2 + C(x-x_o)(y-y_o)]^{-\beta} \tag{3.2}$$

$$A = \left(\frac{\cos\phi}{\alpha_x}\right)^2 + \left(\frac{\sin\phi}{\alpha_y}\right)^2,\ B = \left(\frac{\sin\phi}{\alpha_x}\right)^2 + \left(\frac{\cos\phi}{\alpha_y}\right)^2 \tag{3.3}$$

$$C = 2\sin\phi\cos\phi\left(\frac{1}{\alpha_x} - \frac{1}{\alpha_y}\right) \tag{3.4}$$

where the star's centroid is given by x_o and y_o, the amplitude is a, β describes the slope of the PSF, α_x and α_y describe its width, and ϕ is the rotation of the major axis from horizontal. The data were fit with a Levenberg-Marquardt algorithm and least squares statistic, with 3σ outliers rejected iteratively over two passes. The corresponding FWHM θ is calculated as:

$$\theta = 2\alpha\sqrt{2^{1/\beta} - 1} \tag{3.5}$$

Note that unless otherwise indicated, the Moffat minor FWHM will be used to describe PSF size in further analyses in order to remove the contribution of telescope jitter apparent in the AO-off images.

The quality of the model fit is illustrated in Figure 3.4, which shows a radial profile of a AO-off and AO-on PSF and the best fit Moffat profile along the minor axis. There are small discrepancies between the observed and model PSF, particularly in the 2D residuals (Figure 3.5). The distinct high-order structure evident in the residuals is likely due to the Moffat fitting with a single exponent in the presence of jitter, since the PSF profile follows a different power law along the narrow axis than the elongated one. Static aberrations are calibrated out by optimizing the image quality on artificial sources and recording the centroid offsets on the wavefront sensors, so they are likely not the cause of this residual structure in the PSF. However, the telescope environment may introduce differential aberrations on the various wavefront sensors or some poorly controlled mirror modes could introduce such a pattern once the loop is closed and noise is injected into the system. Table **??** compares the result

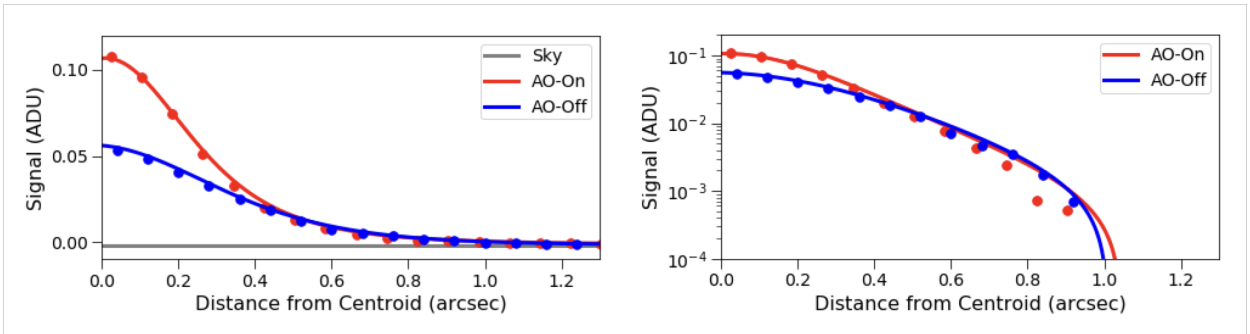


Figure 3.4: Radial profiles along the minor axis of PSFs in AO-off (blue) and AO-on (red) images, shown with linear (left) and logarithmic (right) y-axes. In each case, the image data is represented by circular points and the best-fit model is shown as a solid line. Both PSFs came from the median sized star in stacked images from 2017-05-20 UT and are representative of most of our data.

of the Moffat fit, specifically, the RMS of the minor and major FWHM to the previously described metrics.

The seeing-limited AO-off images are well described by a Moffat profile with a median $\beta \simeq 4.8$, which is slightly above the β of roughly 3-4 typically measured in astronomical images [89]. Additionally, they show relatively little dependence on conditions or FWHM, as expected (Figure 3.6). For the AO-on images, there is a correlation between β and FWHM.

Wavelength Conversion

As previously stated in section 3.3, the data comprising our observations were taken through three different filters (details in Table 3.3). For ease of comparison, the MASS/DIMM data (which are observed at 500 nm) are converted to the wavelength of the corresponding 'imaka science images. The telemetry data have been similarly converted. The data are converted in the following way:

For a given wavelength λ, the Fried parameter describing the amount of wavefront distortion due to atmospheric turbulence scales as:

$$r_o \propto \lambda^{6/5} \tag{3.6}$$

To scale a given seeing measurement θ_{obs} observed at a wavelength, λ_{obs}, we assume an infinite outer scale and that the seeing measurement is proportional to the ratio of the imaging wavelength and the Fried parameter, which gives

$$\theta \propto \frac{\lambda}{r_o} \propto \lambda^{-1/5} \tag{3.7}$$

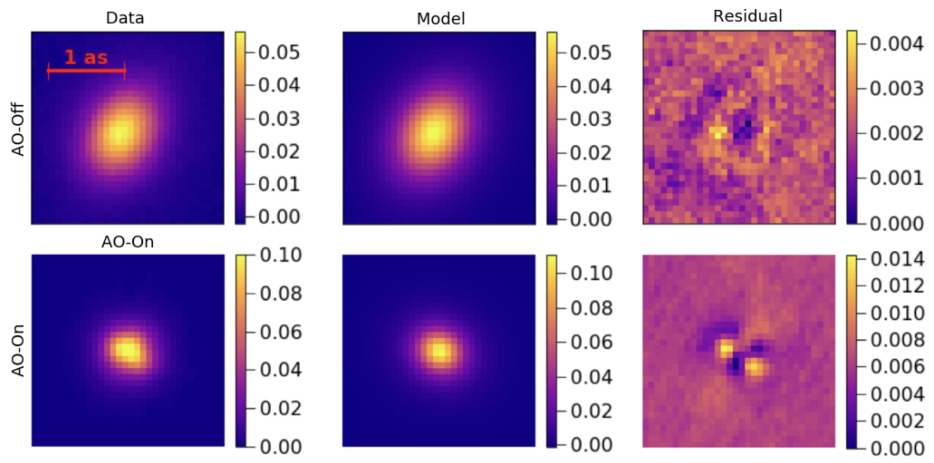


Figure 3.5: Best-fit PSF model composed of a single component, elliptical Moffat for the AO-off (*top row*) and AO-on (*bottom row*) data. The original image (*left*), the best fit model (*middle*), and the residuals (*right)* are shown over the fitting box of 2.56" by 2.56". Colorbar units are in ADU. Image samples were taken from stacked images on 2017-05-20 UT, a night with typical seeing conditions and performance. Note that the color scale is different between AO-off and AO-on to highlight PSF structure, so the differences in SNR between AO-off and AO-on are not represented here.

$$\theta_{conv} = \theta_{\rm obs} \cdot \left(\frac{\lambda_{\rm obs}}{\lambda_{conv}} \right)^{1/5} \tag{3.8}$$

where θ_{conv} is the seeing scaled to a wavelength λ_{conv}.

3.6 Results

GLAO Performance

The simplest analysis that can be done to characterize the GLAO performance is to compare the FWHM in AO-off images to that in AO-on images, as shown in Figure 3.7. The FWHM along the minor axis decreases from 0.76" in the AO-off case to 0.45" with AO correction in I-band, a change of 51%. Changes in all other PSF size metrics at all observed wavelengths are given in Table 3.4, and show roughly consistent improvements.

Though already showing a fair amount of correction, this comparison does not capture the full impact of the instrument. As was visible in Figure 3.5, a change occurs not only in the PSF size, but in its shape as well. Static and dynamic aberrations within the telescope and instrument are also corrected by the GLAO system. As noted above the static aberrations are removed in the AO-off images by applying average voltages to the deformable mirror during

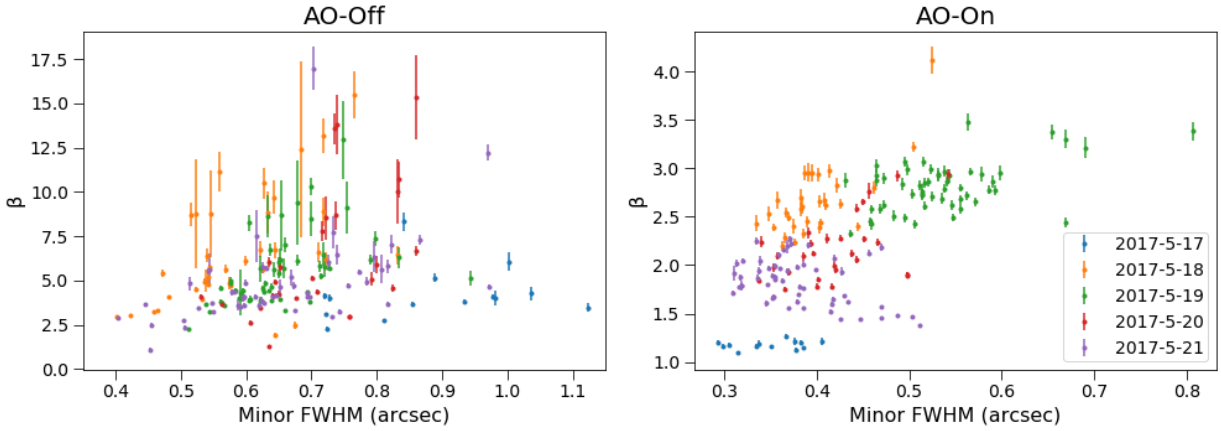


Figure 3.6: Comparisons of the slope parameter β in the Moffat profile to the minor FWHM of the same source. The data show each night of Run 3 in a different color, with each data point representing the median value PSF for a single frame. Points without error bars have uncertainties smaller than the marker size. *Left:* AO-off (median β=4.82). *Right:* AO-on (median β=2.27).

AO-off images. The correction of the PSF elongation, caused predominately by telescope jitter, is readily seen in the images in Figure 3.5. AO-off images show elongation in one direction by a factor of ∼1.3. We attribute this elongation to telescope jitter, as it is generally oriented along either the East-West direction (i.e. RA drive of the telescope). The elongation could be caused by a combination of astigmatism and focus if the deformable mirror voltages used for AO-off images were not averaged for a sufficiently long period. However, this effect would give rise to a random direction and amplitude to the image elongation. Since the direction of the PSF elongation is constant over many hours, its unlikely that a static aberration from the DM shape during AO-off images is not the dominant source of elongation. Regardless, elongation is well corrected by the AO system and is nearly nonexistent in the AO-on images. This stark difference can be seen in Figure 3.8, which shows the elongation parameter:

$$E = \frac{\theta_{maj}}{\theta_{min}} \tag{3.9}$$

The median value of the elongation decreases from 1.34 to 1.08. This residual elongation of 8% is consistent with the plate scale variation in the 'imaka design, which is roughly 8% in the East-West direction. Even more, the AO-on elongation values demonstrate substantially less spread than the AO-off values, with standard deviations of 0.28 and 0.35, respectively.

In addition to changes in FWHM, we can use our other PSF metrics to quantify 'imaka's gains. As seen in Table 3.4, the AO-on/AO-off change in the 50% EE diameter is greater than the same change in the 80% diameter (e.g., in R-band, the improvements in 50% EE

Table 3.4: PSF Size Improvement

Metric	AO-Off (arcsec)	AO-On (arcsec)	Change
R-Band (658 nm)			
Moffat FWHM (min)	0.857 ± 0.014	0.589 ± 0.012	37%
Moffat FWHM (maj)	1.097 ± 0.016	0.639 ± 0.013	53%
Empirical FWHM	1.008 ± 0.015	0.666 ± 0.013	41%
NEA Width	1.000 ± 0.014	0.736 ± 0.015	30%
50% EE Diameter	1.221 ± 0.030	0.899 ± 0.031	30%
80% EE Diameter	2.101 ± 0.026	1.790 ± 0.032	16%
I-Band (806 nm)			
Moffat FWHM (min)	0.761 ± 0.059	0.450 ± 0.007	51%
Moffat FWHM (maj)	1.089 ± 0.070	0.499 ± 0.008	74%
Empirical FWHM	0.845 ± 0.012	0.498 ± 0.008	52%
NEA Width	0.822 ± 0.011	0.565 ± 0.008	37%
50% EE Diameter	1.255 ± 0.020	0.985 ± 0.023	24%
80% EE Diameter	1.799 ± 0.020	1.435 ± 0.014	22%
1000 nm			
Moffat FWHM (min)	0.663 ± 0.037	0.377 ± 0.025	55%
Moffat FWHM (maj)	0.868 ± 0.047	0.416 ± 0.023	70%
Empirical FWHM	0.796 ± 0.036	0.456 ± 0.023	54%
NEA Width	0.715 ± 0.038	0.516 ± 0.028	32%
50% EE Diameter	0.849 ± 0.020	0.682 ± 0.027	22%
80% EE Diameter	1.468 ± 0.031	1.304 ± 0.035	12%

Mean PSF sizes by several different metrics, in the AO-off and AO-on. All values are give in arcseconds. Reported errors are the uncertainty in the mean. The data have been divided between observation filter.

and 80% EE are respectively 24% and 16%). The relatively low change in the 80% diameter is indicative of the less substantial correction the system makes to the extended halo. The radius in which any correction of the PSF is possible for a given AO system is set by the number of actuators in the DM. 'imaka's current DM has 36 actuators; increasing this number could potentially the radius of its AO-correction, and subsequently increase this amount of correction for the higher encircled energies.

The improvement in our final metric, noise-equivalent area (NEA), is shown in Figure 3.9. Note that unlike other metrics and the representation of NEA previously, this figure reports NEA as an area; this representation is useful in addition to the NEA width because photometric precision scales with NEA, while astrometric precision scales with $\mathrm{NEA}^{1/2}$. The NEA decreases from 1.96 square arcseconds to 0.89 in I-band, a change of over a factor of two.

Besides improvements in individual PSFs, AO-on images demonstrate higher stability throughout each night of observing. The spread of minor FWHM for each night in both AO-off and AO-on images can be seen in Figure 3.10. Regardless of seeing conditions for a given night, there is noticeably more variation in the size of AO-off PSFs than those of the AO-on images. This effect is particularly apparent when the free-atmosphere is weak. This is well illustrated in the May run, where the night-averaged standard deviation of FWHM decreases from 0.59" to 0.05", a change of more than an order of magnitude. In contrast, when the total seeing is dominated by poor free-atmosphere seeing (e.g. Run 1), the PSF variability even in GLAO mode, is more pronounced. This trend can also be seen with more detail in Figure 3.16, where the same data are shown with MASS/DIMM data over time for each night. Generally, even when the ground layer seeing has more variation over the course of the night than the free atmosphere, the AO-on FWHM remains more stable than in the AO-off case.

Comparisons to Seeing

The comparison between AO-off and AO-on images already begins to characterize the gains from GLAO. However, a more objective indicator of success would be to compare the GLAO PSFs to the seeing above the ground layer using MASS/DIMM measurements described in §3.3. First, we compare the FWHM of the focal plane images to the contemporaneous MASS/DIMM seeing in Figure 3.11, where the top and bottom panels show the correlations between AO-off images and DIMM, and AO-on images and MASS, respectively.

Assuming accurate measurements by the MASS/DIMM instruments and no error in our

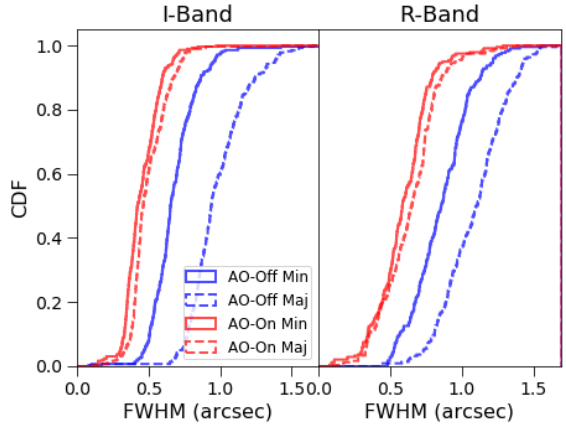


Figure 3.7: A comparison of the FWHM in AO-off (*blue*) and AO-on (*red*) images. Minor FWHM is shown in *solid* lines, and major FWHM in *dashed* lines. Data are divided between *I*-band *left* and *R*-band *right*.

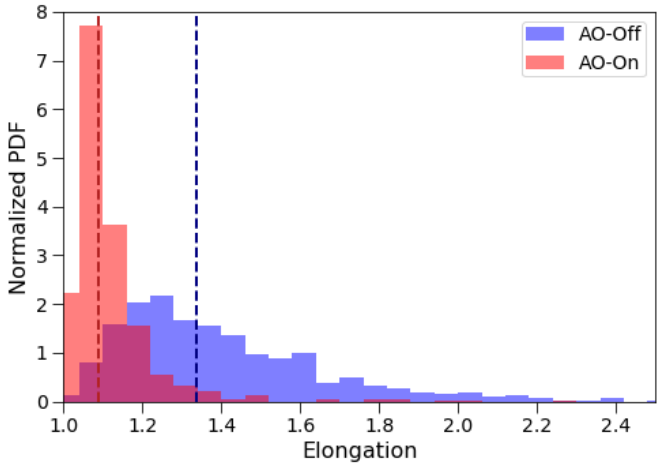


Figure 3.8: A comparison of the elongation (derived from the Moffat fit and defined by Equation 3.9) of PSFs in AO-off (*blue*) and AO-on (*red*) images. The median value of each distribution is indicated by the vertical *dashed line* in the corresponding color: 1.34 for AO-off and 1.08 in AO-on. The elongation in the closed-loop images is consistent with the 8% plate scale variation in the optical design.

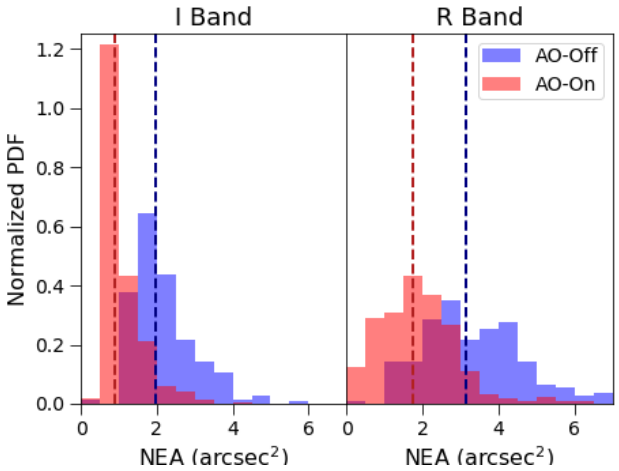


Figure 3.9: The change in noise equivalent area between AO-off (blue) and AO-on (red) images across all nights, shown as normalized probability distribution functions. Median values of both distribution are designated by vertical dashed lines. The data are divided into I-band *left* and R-band *right*.

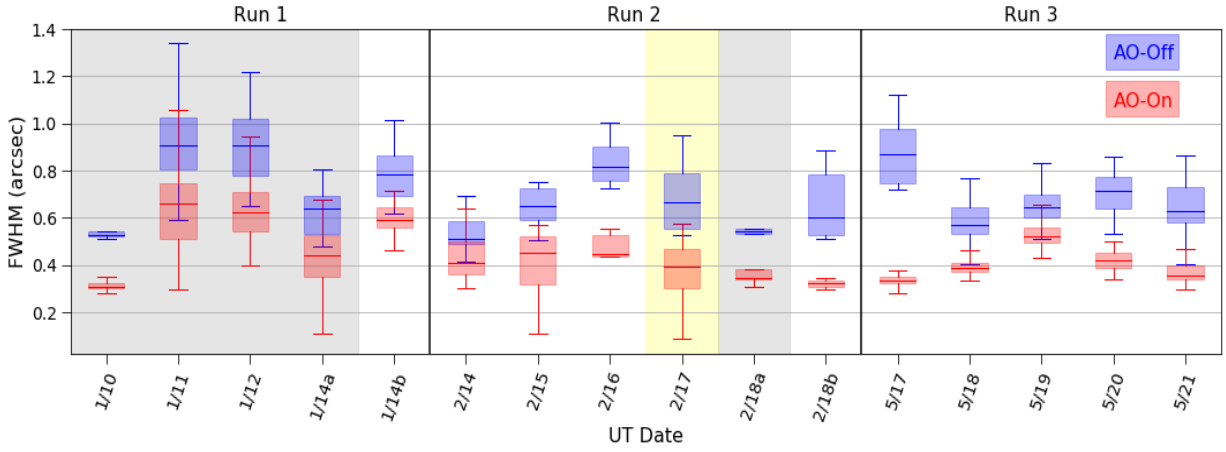


Figure 3.10: A comparison of the PSF stability between AO-off (*blue*) and AO-on (*red*) across all nights. For each data point, the median value of the empirical FWHM is represented by the center line, with a box surrounding it and spanning the second and third quartiles of the data. The full range of data is shown by the extended lines. Shaded regions refer to the observation wavelength: R-band in *gray*, 1000 nm in *yellow*, and I-band is unshaded.

correction, the DIMM should show some correspondence with the AO-off image FWHM and similarly the MASS should correspond to the AO-on data. To examine this relationship, we looked at the correlation between the focal plane data, telemetry, and MASS/DIMM seeing. Across all nights, we calculated the correlation coefficient for six combinations: three in the total atmosphere regime (AO-off vs. free atmosphere telemetry; AO-off vs. DIMM; integrated seeing telemetry vs. DIMM) and the free atmosphere regime (AO-on vs. integrated seeing telemetry; AO-on vs. MASS; free atmosphere seeing telemetry vs. MASS). In the case of AO-on images, we see good correlation with the MASS, with a correlation coefficient of 0.46 in I-band and 0.52 in R-band. These correlations are shown in Figure 3.11. The image data is similarly well correlated with 'imaka's telemetry in I and R band, though some discrepancies occur at 1000nm. We note that the amount of data taken in this filter is small, so the lack of correlations here may not be significant. A full list of correlation coefficients is presented in Table 3.6.

Although the AO-on images are correlated with the MASS seeing, there is a distinct offset between their cumulative distributions with the MASS seeing being smaller than the AO-on FWHM. One limiting factor in the agreement is the apparent floor in the AO-on images at approximately 0.3". Across all analyses, our GLAO PSF sizes generally don't go below this minimum even when the seeing predicted by the MASS does drop below 0.3". This limitation is likely due to systematic effects (e.g., static aberrations, internal seeing), that the low order system (due to a relative lack of DM actuators and the low bandwidth of our wavefront sensor cameras) cannot correct.

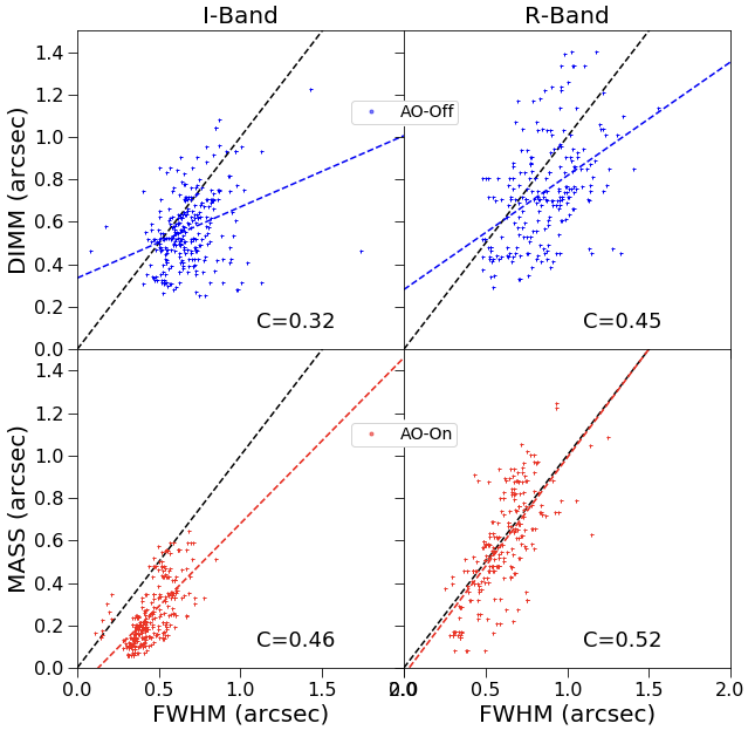


Figure 3.11: Correlations between focal plane data and MASS/DIMM. *Top:* correlation between AO-off images and DIMM seeing, matched by observation time. *Bottom:* AO-on images and MASS. In each panel, data points are shown with a line of best fit in the same color, with the corresponding correlation coefficient C reported. A dashed black line in each panel shows a 1:1 correlation. The AO-on/AO-off data are separated by filter, with I-band on the left and R-band on the right. The MASS/DIMM are converted to match the wavelength of the 'imaka image data using equation 3.8. Note that the I-band and R-band data sets were taken on different sets of nights with different seeing conditions.

As mentioned in §3.3, comparisons to MASS/DIMM are also limited by the fact that they are not observing along the same lines of sight through the atmosphere. To compensate for this, we compare the image quality to our estimates of the integrated and free-atmosphere seeing derived from our telemetry measurements. These have the advantage that these data are synchronized with the images. Figure 3.12 compares the science image PSF FWHM to both the telemetry measured by 'imaka and to the seeing measured by the MASS/DIMM. Median values and standard deviations are given for all these data sets in Table 3.5.

Figure 3.12 shows the cumulative distributions for the image FWHM and the seeing from the MASS/DIMM and the imaka telemetry. There is a mix of agreement and disagreement between the distributions of seeing and image FWHM. During the nights when we observed in I-band, both the AO-on and AO-off images FWHM distributions were shifted to values

Table 3.5: Comparison of Image Quality with Seeing Estimates

	Total Atm. or AO-Off	Free Atm. or AO-on	Change
	(arcsec)	(arcsec)	
	***R*-Band (658nm)**		
'Imaka Images	0.857 ± 0.014	0.589 ± 0.012	37%
'Imaka Telemetry	0.854 ± 0.014	0.665 ± 0.012	25%
MASS/DIMM	0.728 ± 0.011	0.616 ± 0.013	17%
	***I*-Band (806nm)**		
'Imaka Images	0.761 ± 0.059	0.45 ± 0.007	51%
'Imaka Telemetry	0.572 ± 0.012	0.401 ± 0.009	35%
MASS/DIMM	0.567 ± 0.007	0.253 ± 0.006	76%
	1000nm		
'Imaka Images	0.663 ± 0.037	0.377 ± 0.025	55%
'Imaka Telemetry	0.563 ± 0.023	0.453 ± 0.023	22%
MASS/DIMM	0.445 ± 0.011	0.297 ± 0.021	40%

Mean PSF sizes by three different measurements, observing the total integrated atmosphere (from top to bottom: AO-off images, AO-off telemetry, MASS integrals) and the free atmosphere (AO-on images, AO-on telemetry, DIMM integrals). All values are give in arcseconds. Reported uncertainties are the error on the mean. The data are separated by observation wavelength of the 'imaka images, but all data are shown at their original wavelengths.

larger than the I-band converted MASS/DIMM seeing and telemetry derived seeing. In these cases the distribution of MASS seeing values is significantly lower than either the telemetry derived free-atmosphere seeing or the AO-on image FWHM. On the other hand, on nights when we observed in R-band, the AO-off image FWHMs are smaller than the telemetry seeing but worse than the MASS/DIMM seeing but the AO-on image FWHMs are comparable or better than the MASS or telemetry derived free-atmosphere seeing.

Some of this difference could come from the fact that the I-band and R-band nights had quite different seeing conditions. During most of the nights when we observed in I-band, the free-atmosphere seeing was weak and both the MASS and DIMM seeing are better than their corresponding image FWHMs. During the the R-band nights the total seeing was dominated by large free-atmosphere seeing. Here, the MASS seeing CDF and AO-on image CDF are similar for the better free-atmosphere seeing but skewed to values larger than the AO-on image FWHM CDF in the poorer free-atmosphere seeing conditions. Together, the I-band and R-band AO-on data suggest that we have an instrumental floor which prevents GLAO images of better than about 0.3 arcseconds. Alternatively, the data may be indicating that the MASS underestimates the free-atmosphere seeing under good seeing conditions, as seen in [85].

For the telemetry data the shapes of the distribution functions are similar in all cases to

the distributions for the image FWHM, but with fixed offsets between them. One possibility is that the effective wavelength of the wavefront sensors changes between the observed fields. All of the R-band data were taken on the Pleiades field where the guide stars are generally bluer than the guide stars in the other fields. Since we assume a wavelength of 0.7 microns for the WFSs, we may be overestimating our seeing estimates in Pleiades field (e.g. R-band data). Alternative explanations could be an effect that happens on longer time scales, such that it is seen in the 30-45 second science camera exposures but not the 15-second telemetry exposures. Additionally, if the dome seeing consists of mostly high frequencies (which the AO system can not correct for), these residual wavefront aberrations could contribute to a broadening of the wings of the PSF, in turn slightly enlarging the FWHM.

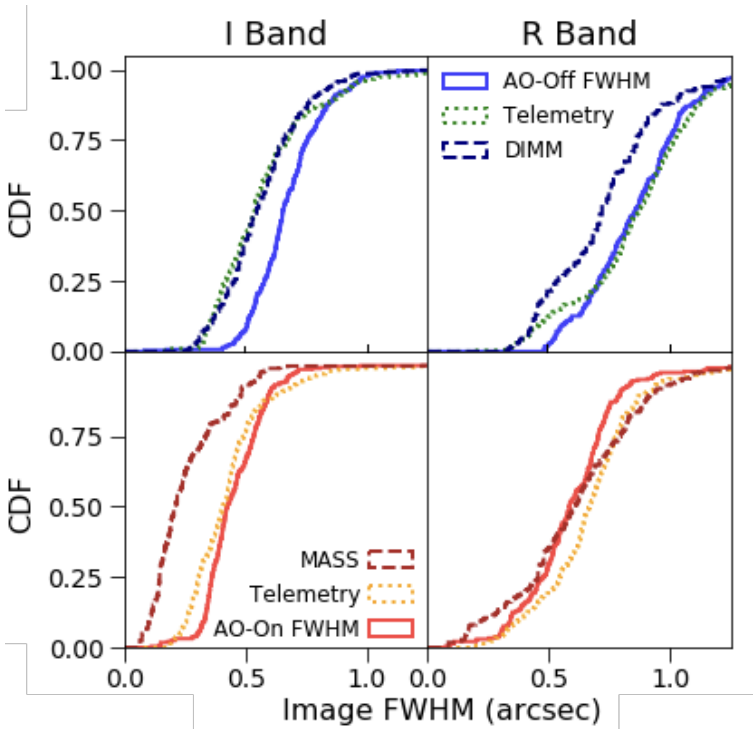


Figure 3.12: Cumulative distribution functions comparing image plane, telemetry, and MASS/DIMM data. The top panels show values associated with the total atmosphere: telemetry and DIMM seeing, and are compared to the AO-off image plane PSF size. The bottom panels show the free atmosphere case: the telemetry case is for all turbulence above $\sim 600m$, along with MASS and AO-on image plane PSF size. The AO-on/AO-off data are separated by filter, with I-band on the left and R-band on the right. The MASS/DIMM and telemetry, taken from nights that corresponded to each set of observation data, are converted to match the wavelength of the 'imaka image data using equation 3.8. Note that the I-band and R-band data sets were taken on different sets of nights with different seeing conditions.

Table 3.6: Correlations Coefficients between Data and Seeing

	Total Atmosphere (AO-off/DIMM)	Free Atmosphere (AO-on/MASS)
	***R*-Band (658nm)**	
Focal Plane vs. Telemetry	0.68	0.70
Focal Plane vs. MASS/DIMM	0.45	0.52
Telemetry vs. MASS/DIMM	0.35	0.47
	***I*-Band (806nm)**	
Focal Plane vs. Telemetry	0.33	0.36
Focal Plane vs. MASS/DIMM	0.32	0.46
Telemetry vs. MASS/DIMM	-0.10	0.23
	1000nm	
Focal Plane vs. Telemetry	0.15	0.43
Focal Plane vs. MASS/DIMM	-0.02	0.14
Telemetry vs. MASS/DIMM	-0.23	0.27

Correlation coefficients for 'imaka image data, 'imaka telemetry, and MASS/DIMM seeing in both the total and free atmosphere cases. Telemetry and MASS/DIMM data have been converted to the matching observation wavelength with equation 3.8.

Wavelength Dependence

We can examine the wavelength dependence of GLAO correction by plotting the raw FWHM values at different wavelengths, as shown in Figure 3.13. Ideally, we would simply look at how FWHM varies with wavelength. However, because the observations in the different filters were done over different nights and over a wide range of observing conditions, direct comparisons cannot be made from night to night. For this reason, we present each data set with the MASS/DIMM seeing measurements for the corresponding times, converted from 500 nm to the relevant wavelength. As in previous figures, we see an offset between AO-off and DIMM, as well as between AO-on and MASS. Despite this discrepancy, at each wavelength there is a clear improvement between AO-off and AO-on, sometimes larger than the difference between DIMM and MASS. Though we do not yet have enough data in consistent seeing conditions to conclude anything about the specific dependence of GLAO correction on wavelength, we can already see that 'imaka achieves substantial correction at all currently observed wavelengths.

Field Variability

From simulations the GLAO PSF is predicted to be quite uniform across the field. However, as GLAO is deployed over wider and wider fields, it is important to ensure that the instrumental PSF quality does not suffer with increasing field size. The left panel of Figure

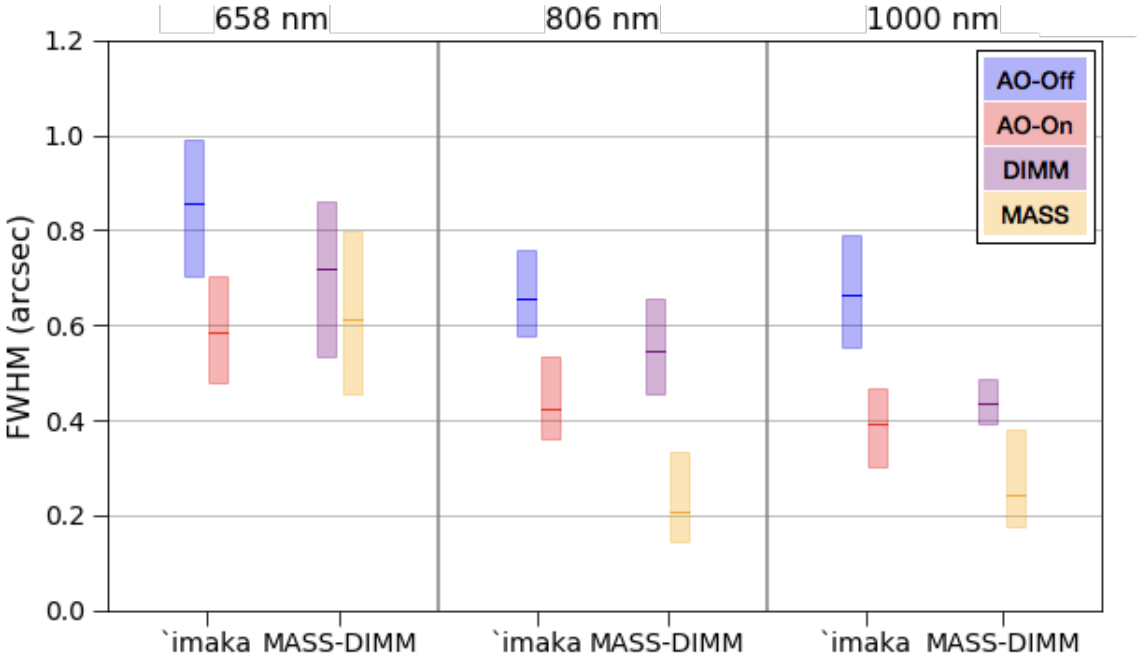


Figure 3.13: Wavelength dependence of PSF size is shown here as minor Moffat FWHM for AO-off (*blue*) and AO-on (*red*) images for all nights, with no wavelength conversion applied. Each set is shown next to the MASS (*orange*) and DIMM (*purple*) seeing measurements for the same observation times. As the MASS/DIMM data is initially measured at 500 nm, their values are presented here scaled to the wavelength of the corresponding 'imaka data points. Each box represents the mean value with a solid line and the second and third quartiles with a shaded region. We note that the R-band and I-band points are from an average of $\geq$5 nights, while the 1 μm point is from only one night (refer to Table 3.3 for observation details).

3.14 shows the FWHM of each star as a function of position in the stacked AO-on image from 2017-05-19 UT (chosen because its seeing conditions were typical of the run), with the location of guide stars marked in green. To better view any global trends across the field, these data had outliers removed through an iterative sigma clipping routine, rejecting points more discrepant than 3σ from the median, with 5 iterations.

The dominant field variation is a gradient in FWHM, with PSFs increasing in size towards the East side of the field. In single guide star AO, the expectation is for the PSF quality to improve closer to the guide star. The trend visible here is not well correlated to the distance to the field center or to any single guide star. The trend is repeated over multiple nights and appears fixed to the sky, as we describe in more detail below.

We used several methods of quantifying the global variation. First, the data shown in the

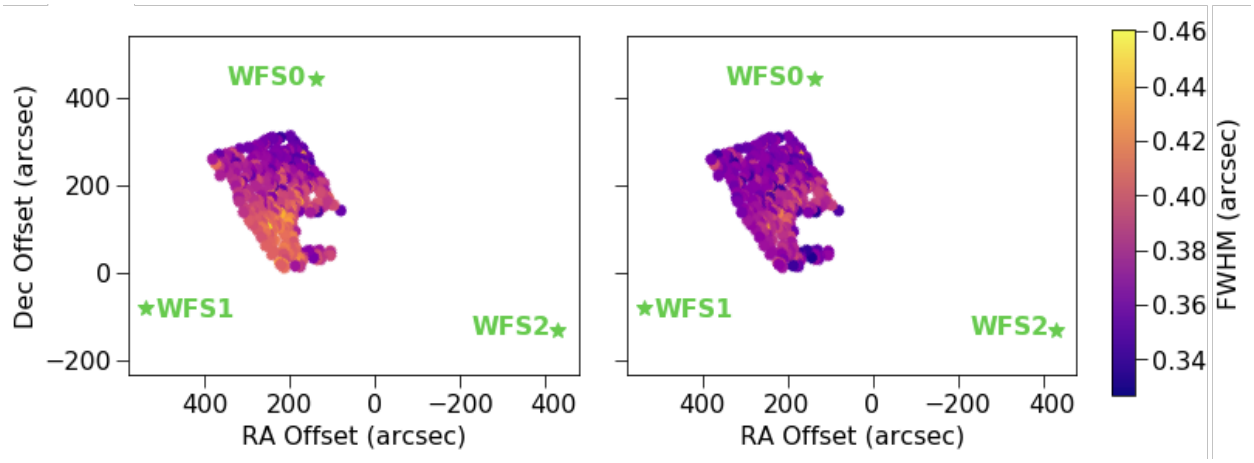


Figure 3.14: The FWHM of individual stars in a stacked image of all AO-on frames from 2017-05-18 UT. *Left:* The original data, showing the variability of the PSF across the science field. *Right:* The same data after subtraction of the best fit plane, with the median data point from the left (0.40") added to all points for comparison purposes. Wavefront sensors' positions relative to the science frame are marked in green and WFS1 is the brightest guide star. The data were taken at I-band with no wavelength conversion applied. By visual inspection, the field on the left shows more structure in its variability, with a strong gradient in FWHM increasing to the bottom left. This global pattern in the variability is largely removed in the right panel. Between the two figures, the range of FWHM decreases from 0.201" on the left to 0.110" on the right.

Table 3.7: Nightly Variability Plane Fit for Run 3

Date (UT)	$\Delta FWHM_i$ (")	$\Delta FWHM_f$ (")	Direction ϕ (°)	Gradient (arcsec/arcmin)
2017-5-17	0.128	0.032	153.4 ± 6.0	$(16.6 \pm 0.2) \times 10^{-3}$
2017-5-18	0.201	0.110	97.0 ± 9.7	$(8.8 \pm 0.1) \times 10^{-3}$
2017-5-19	0.104	0.091	95.8 ± 9.7	$(8.2 \pm 0.1) \times 10^{-3}$
2017-5-20	0.115	0.086	86.0 ± 14.5	$(5.0 \pm 0.1) \times 10^{-3}$
2017-5-21	0.178	0.092	92.3 ± 31.3	$(4.2 \pm 0.1) \times 10^{-3}$

For each night of the May run, the mean range of FWHM values of all stars in AO-on images is given for both before ($\Delta FWHM_i$) and after ($\Delta FWHM_f$) the removal of a best fit plane. The value ϕ represents the direction of the best fit plane (in degrees, East of North with respect to the detector) while the gradient describes the variation in arcseconds per arcminute of the plane.

left panel of Figure 3.14 was fit in three dimensions (x, y, and FWHM) to a plane. The best-fit plane is defined by two quantities. The first, ϕ, is the position angle of the plane's normal vector projected onto the focal plane and measured East of North in degrees. The second is the gradient, which measures the spatial change of FWHM in arcseconds per arcminute. The right panel of Figure 3.14 shows the residual FWHM after subtraction of the best-fit plane for the stacked image on 2017-05-19 UT, with an offset of 0.44" added for the purpose of comparison. In addition to these parameters, we calculated the range of FWHM values before and after the plane removal ($\Delta FWHM_i$ and $\Delta FWHM_f$).

As seen in Figure 3.14, though some faint, high-order structure appears to remain, the variation decreases from the left to the right panels substantially. Averaged over the five nights of the May run, the range decreases from 0.145" to 0.082". Finally, the average variation of FWHM across the best-fit planes in the direction of greatest change is 0.0086" per arcminute of the field.

To justify the removal of the fitted plane, the variability would have to be attributed to an instrumental effect, such as a tilt in the CCD relative to the incoming beam. To check whether the planar variability could be due to an atmospheric effect, similar plane fits were generated for each individual frame on the same night. On a frame-by-frame basis for all images taken 2017-05-18 UT, neither the direction nor the gradient showed significant correlation to time, ground layer seeing, wind speed or direction, or zenith angle of the telescope (a potential cause of flexure in the instrument). Based on the lack of any trends in these analyses, we conclude that the field variability of the FWHM is dependent on an instrumental effect such as misalignment, which will be addressed in the future.

Further evidence for an instrumental origin of the PSF field-variability comes from the stability of the plane over multiple nights. The stacked AO-on images from each night of the 2017 May run were fit with a plane in the same manner as above and results are listed in Table 3.7. The plane parameters ϕ and gradient are given, along with $\Delta FWHM_i$ and $\Delta FWHM_f$. The uncertainties in angles were generated using a full sample bootstrap with replacement routine in the fit. In order to generalize this effect, we report 4.5% as the typical variation in FWHM across the field, calculated as the mean of each night's percent change from the median to the minimum point. As such, we can tentatively apply a correction of 4.5% to any focal plane PSF size measurements due to field variability.

The variation in the data of Table 3.7 may be explained by adjustments made to the instrument between nights. For example, the science camera was removed and rotated by 45^o between 2017-05-17 UT and 2017-05-18 UT, corresponding to the large change in the direction of variation. Also, between 2017-19-20 and 2017-5-20, the camera was taken off and returned to the same position. The constant direction with different inclination suggests that after remounting the camera, we modified the camera tilt slightly. Though not all the changes between nights can be explained with camera changes, these data do present us with potential experiments we can conduct in the future in order to understand the source of this aberration.

It should be noted that the entirety of this variability analysis comprises of only AO-on images. The gradient apparent in these images is not seen in AO-off frames, as the image

quality is too poor to detect the small change in the FWHM over the field of view. Any attempts at fitting the variability had uncertainties too large to provide valuable insight, so they were omitted from this analysis.

3.7 Discussion

The results of our three commissioning runs already show the value of 'imaka and GLAO and the science cases it makes possible. From the size of the field alone, the potential to obtain sharp images of extended sources or large, crowded fields is apparent. Across all nights, our science images display an improvement in FWHM of up to a factor of 1.8 , depending on the metric used (summarized in Table 3.4). The median (I-band) minor FWHM across all nights of 0.45"$\pm$0.03" is consistent with Monte-Carlo simulations for 'imaka, which predicted FWHM in the same band of approximately 0.45"-0.50" within the central 6.5' [90]. The range in improvement comes from the difference in how each metric quantifies the shape of the PSF, and can be most simply summarized by looking at the change in minor and major FWHM of the Moffat profile (0.72" to 0.49" and 1.00" to 0.52", respectively). The difference along the two axes in AO-off images are indicative of significant elongation, likely due to telescope jitter, which is almost entirely removed in AO-on images, where the median ratio of major to minor FWHM decreases from 1.34 to 1.08. This highlights the value of a GLAO system in cleaning up telescope and instrumental artifacts that would otherwise reduce image quality in "seeing-limited" data.

As an alternative metric, the radius of 50% encircled energy decreased from 1.26" to 0.99" in I-band. The decrease of 50% EE radius by a factor of 1.3 corresponds to increasing the depth of observable stars by approximately 0.9 magnitudes at a given exposure time, yielding larger samples of stars in a given field. This is particularly advantageous for high-precision astrometry and photometry, where the precision of a measurement scales with the square root of the number of reference stars used.

Besides the increased number of observable objects, the improvement in individual PSFs allows for other opportunities. Photometric precision, for example, scales with NEA, which 'imaka has improved from 1.96 to 0.89 square arcseconds (I-band), a change of more than a factor of 2. Astrometric precision is also improved by a more concentrated PSF, scaling with the square root of NEA (or what we term the NEA width).

In addition to improving individual PSFs, AO-on images from 'imaka demonstrate improved stability in the FWHM over time and through varying seeing conditions. This effect is most clearly seen in the five nights of the May run, where the nightly standard deviation of FWHM went down from 0.59" in AO-off images to 0.05" in AO-on images, a change of nearly an order of magnitude. These improvements greatly aid spectroscopic observation. Besides the fact that increased encircled energy yields a higher SNR for a given exposure time, the stability of the PSF size allows for optimized slit width selection. This optimization in turn allows for maximum spectral resolution in an observing run: Assume, for example, the ratio of maximum FWHM in a night in with AO-off and AO-on is 'imaka's average of

1.6, as previously reported. A decrease in the slit width by this amount would improve the spectral resolution by nearly a factor of 1.6. In addition, since the GLAO PSF is more stable spatially and temporally, there is the possibility to decrease the slit width further.

Furthermore, 'imaka has demonstrated an amount of correction that could be the difference between an unusable night of observing and a usable night. This is a major benefit of GLAO observing; Andersen et al. [91] demonstrated in simulations that at the Cerro Pachon site in Chile, a GLAO system would improve an observatory's efficiency by up to 40% by increasing the percentage of the 'best seeing-limited image quality nights' from 20% to about 70%. As is visible in Figure 3.16, sporadic changes in seeing are minimized by 'imaka, in addition to the seeing being significantly decreased. Eliminating nights unusable because of bad seeing would increase the science output of any facility, an advantage for all types of observing.

Considering the fact that 'imaka has thus far only been run in natural guide star (NGS) mode, its gains are already comparable to GLAO systems using laser guide stars (LGS). For example, the previously mentioned ARGOS instrument and the GLAO system on the MMT telescope both achieved improvement in FWHM of roughly a factor of two in LGS mode [31, 30]. We have come close to this improvement in larger fields while still using natural guide stars, which, for GLAO, inject noise from atmospheric profiles unrelated to the corrected field. The initial success of 'imaka compared to such instruments makes the prospect of adding laser guide star capability in the future particularly promising.

In that vein, 'imaka was constructed on a low budget, as it is meant primarily as a demonstrator for GLAO on Maunakea. As such, it exhibits limitations that the high-quality optics and system engineering commonly used in large-telescope instrumentation projects could easily address.

Despite the clear gains 'imaka has achieved, there are distinct differences between the distributions of FWHM we observe in our science images and the seeing (both integrated and free-atmosphere) estimated from the 'imaka wavefront sensors and Maunakea MASS/DIMM. Though the median MASS measurement of free atmosphere seeing was 0.25" across all nights of observing in I-band, the image quality of AO-on images those nights was reduced to only 0.45", while 'imaka's telemetry measured 0.40" seeing for what is ostensibly the same atmospheric profile experienced by the science camera. Despite this offset, the focal plane images are typically correlated to the telemetry and MASS/DIMM. The offset in the cumulative probability distribution (Figure 3.12) between the telemetry and the focal plane images indicates an instrumental error; it is nearly constant with image FWHM, suggesting that it is not a fixed non-common path aberration since its contribution to the image degradation would decrease as the seeing degrades. There is some indication that the tomographic reconstruction error (e.g. our ability to separate ground-layer turbulence from high-altitude turbulence) may be larger than expected but this is left to a later work to quantify. There could also be a floor in the 'imaka image quality of $FWHM_{min} \sim 0.3$" that the AO-on images cannot go below. Potential causes for this limitation will also be explored in relation to 'imaka's error budget in a later work.

The initial characterization of AO-on images in relation to wavelength and field position

in this chapter reveals that we have not yet reached the atmospheric limits with 'imaka and we may be able to improve instrument calibration and performance in the future. Though very preliminary, we can already see a weak trend in FWHM as a function of observation wavelength. Because of the limited amount of multi-wavelength data and constraints on seeing conditions, it is currently difficult to conclude more about the wavelength dependence. However, we plan to extend this analysis in the future by observing both at shorter and longer wavelengths over a wider range of seeing conditions. As AO correction is substantially more difficult to achieve with shorter wavelengths, we will test how far this limitation can be pushed.

We will also test larger fields in order to better understand the variability of the GLAO correction across images. Initial analysis of the five nights of the May run show an average range of 0.145". This variation is reduced to a range of 0.082" when the planar structure is removed. The plane itself varies on average by 0.0082 arcseconds per arcminute across the field, which corresponds to a typical variation (percent change between minimum and median) of 4.5%. We will continue to examine this effect in future runs, both with larger fields and different asterisms. The data presented in this chapter represent one configuration of wavefront sensors and control matrices per night on relatively small science fields; in reality, we collected data with a range of asterisms, which we are currently analyzing in more detail.

This chapter has only looked at a narrow range of commissioning results. It serves as a complement to (Chun, in prep), the main results paper of 'imaka's commissioning and will be followed by additional papers specifically on 'imaka's PSF and 'imaka's photometric and astrometric precision. We will also be continuing regular 'imaka runs with new experiments. In order to better address some of the unanswered questions in performance thus far, one new component of the coming runs will be the use of larger science cameras, with plans for an 11' optical camera and a 7' infrared camera. This will allow us to further examine the GLAO correction over large areas and at longer wavelengths.

Collectively, the results of 'imaka demonstrate the potential for a GLAO system on telescopes on Maunakea. The improvements we see in image quality across large fields could yield a significant boost in sensitivity for multi-object spectrographs, potential for high precision astrometry, and high spatial resolution of large sources, all of which can be valuable to a wide variety of science cases. Though the delivered image quality of 'imaka on the UH 2.2m may not be identical to expectations for larger telescopes, our image quality improvement by a factor of 1.4-1.9 is comparable to the gains made by GLAO systems deployed on larger telescopes, such as the factor of 2 improvement in FWHM achieved by ARGOS on each of LBT's 8.4 m telescopes [31] and the improvement of 1.5-2 in FWHM on the VLT's 8.2 m UT4 telescope [32]. Furthermore, many instruments on Keck, such as LRIS and DEIMOS, are impacted by effects (e.g. focus changes, telescope jitter) that we have shown 'imaka can correct [92, 93].

3.8 Conclusion

We have demonstrated the potential of GLAO with guide stars distributed over fields of about 16 arcminutes over 15 nights of observing. These initial findings are showing promising results:

Between AO-off and AO-on modes, 'imaka's focal plane images demonstrate a decrease in PSF size by up to a factor of 1.8, depending on the metric used. The AO-on PSF also shows an elongation of 1:1.1, which is consistent no elongation given the design variation of the plate scale of 8% in the North-South and East-West directions. This represents a substantial reduction from the AO-off value of 1:1.4. The median minor FWHM derived from a Moffat profile for AO-on images was 0.56" in R-band and 0.43" in I-band, and 0.82" for AO-off images in R-band and 0.72" in I-band (including a 4.5% correction for field variability). These quantities are typically correlated with 'imaka's telemetry measurements of the free-atmosphere and integrated seeing, with correlation coefficients in R-band of 0.68 and 0.70 respectively. Despite variation in the seeing over the course of a night, AO-on images demonstrate significantly more temporal stability than AO-off images, with the standard deviation of minor FWHM decreasing from 0.59" to 0.05", averaged across the five nights of the final run. The variation in FWHM across the field is also minimized in AO-on images, with FWHM varying on average by 0.0082" per arcminute and an average percent variation of 10.3%. Overall, these early results from 'imaka suggest that a facility-class ground-layer adaptive optics system on Maunakea would deliver valuable gains over seeing-limited images and spectroscopy.

3.9 Additional Figures: Nightly Performance

Figure 3.15: PSF FWHM versus airmass for each night. Each data point (black) represents the median value of all sources in a frame (AO-on minor FWHM). The data were fit to two models: a constant FWHM (red solid line) and a 3/5 power law (blue dashed line). The corresponding reduced χ^2 for both fits are reported in matching colors.

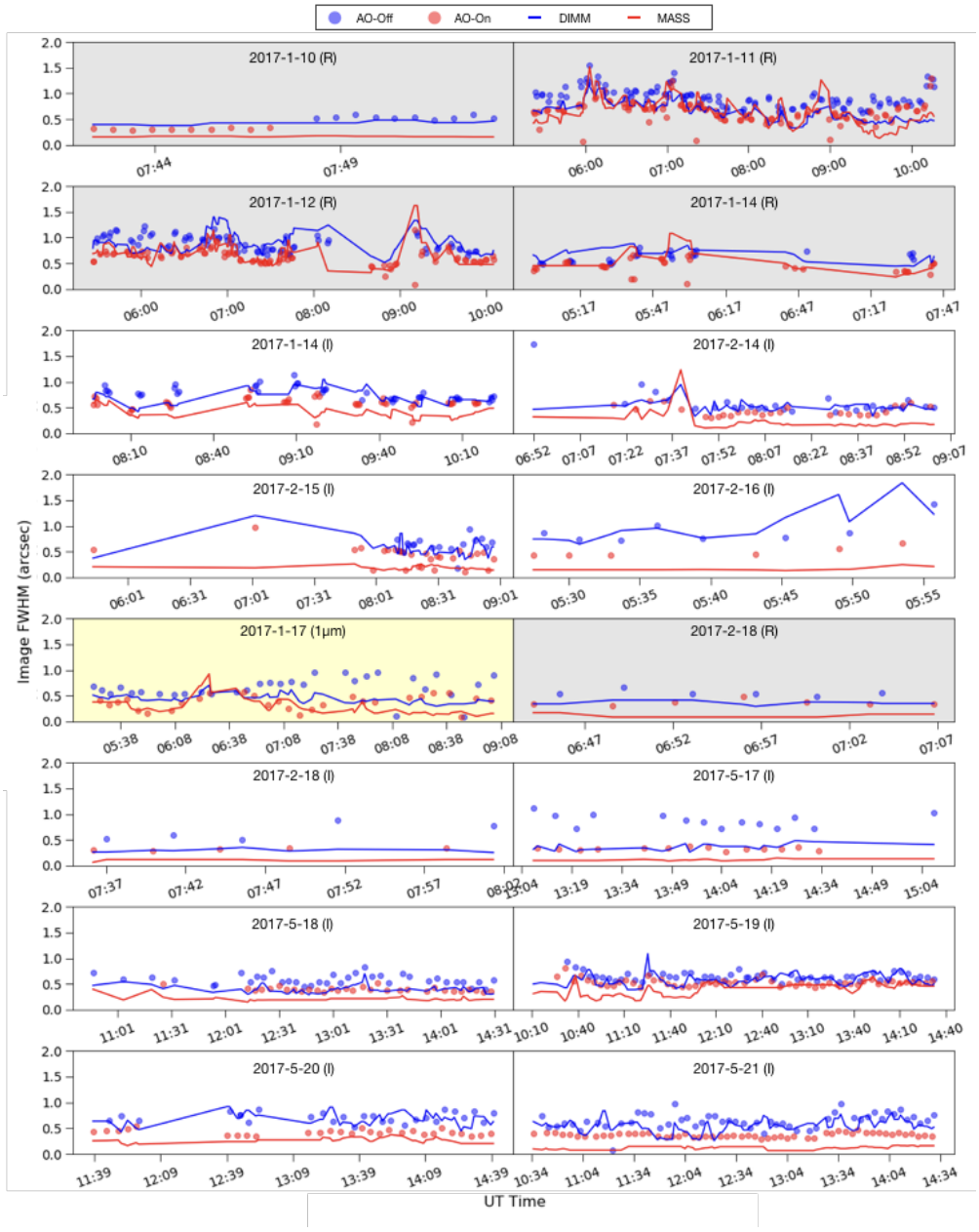


Figure 3.16: Nightly performance summaries. In each plot, AO-off and AO-on image FWHM (blue and red dots, respectively) are compared over time with the MASS/DIMM seeing (solid lines, red and blue respectively). The MASS/DIMM values are converted to the corresponding observation wavelength with equation 3.8. The observation wavelength is reported with the date and indicated by each figure's background color: *gray* for R-band, *yellow* for 1000 nm, and no shading for I-band. The image FWHM data points each represent the median value of all sources in a single frame. Each panel's caption gives the UT date of observation.

Chapter 4

It's the System, Man: How Cultural Beliefs Maintain Inequity in Astronomy Graduate Programs

4.1 Introduction

State of the Field

Despite accounting for only 30% of the US population, 75% of US PhDs in physics and astronomy in 2019 were awarded to white men. In contrast, women of underrepresented racial minorities (URM) earned less than 3% of the same PhDs while comprising 15% of the US population, illustrating a glaringly wide gap in educational outcomes between racial and gender groups [2]. The field exists in this grossly unrepresentative state decades into growing interest in supporting women and underrepresented minorities, with research, educational initiatives, and entire careers being dedicated to this goal [94]. While adjacent fields such as chemistry and biology have made significant strides towards achieving equity goals in this time, physics maintains one of the lowest percentages of bachelors, masters, and doctoral degrees awarded to underrepresented minorities in the STEM fields, second only to astronomy. Similarly, both fields remain far from achieving gender parity at any level of higher education [46].

While much of the discussion of achieving equity goals focuses on the presumed deficits of minoritized students or demographics and metrics of representation, growing evidence indicates the manifestation of systems of oppression (e.g., racism and sexism, among others) in institutions of higher education as being a barier to said equity goals, as well as the institutional actors that maintain them [95]. Notions of the abilities of minoritized students have been demonstrated to be present and negatively influential in the admissions, teaching, and advising of students [96, 97, 98]. Furthermore, female URM graduate students in astronomy are 15-20 times more likely to report discrimination or harassment than non-URM males [42]. It comes as no surprise, then, that both gender and race influence the persistence of

graduate students in astronomy, a fact which points to not only gate keeping, but the internal operations of graduate programs as reducing the presence of minoritized scientists in the field [42]. As such, this work aims to understand how these internal operations may create barriers for minoritized students while perusing their graduate studies.

Research Question

This work initially sought to understand how the environments in physics and astronomy departments lead to negative outcomes for non-majoritized students (e.g., higher rates of discrimination and harassment and lower retention) by examining how multiply minoritized graduate students in predominantly white and male departments relate to, observe, and interact with department culture. As the data collection and analysis progressed, our scope narrowed to a specific question we attempt to answer with this work: **How can the cultural beliefs of academic astronomy and physics contribute to inequity in graduate programs?**

4.2 Background

Culture and Cultural Beliefs

Before attempting to characterize the culture of academic physics and astronomy, we must define what we mean by 'culture'. Broadly speaking, 'culture' is a system of inherited values, goals, knowledge, beliefs, and norms that provide members with a shared sense of who they are and a common purpose for action [47]. Simply put, this can be understood as what is done, how it is done, and who is doing it. Culture is enabling, by serving as a resource for social inheritance, but it can also be constraining as people participate in cultural practices that were developed by previous generations. Furthermore, culture is dynamic, and is always being reconstituted by people as they participate in it; while it is never passed down or taken up uniformly, there are stable elements that are appropriated by new members, even as there is change [99]. There are many ways in which a culture can demonstrate its shared values, including how it defines goals, how it evaluates and socializes people, how decisions are prioritized, how conflicts are managed, and what expectations group members, especially leaders, have [46].

In this work, we are primarily concerned with cultural beliefs, which we define as the related patterns of attitudes and behaviors maintained by a group and passed down to new group members. [100]. In the context of this work, we can think of the cultural beliefs of physics and astronomy as the stories told in this community about how physics is done, who physicists are, and what makes a great physicist. Implicit in these stories are how physics isn't done, who physicists aren't, and what makes a bad physicist.

In the following section, we will describe two dominant cultural beliefs that are relevant to this study: the belief that that the field is a meritocratic system in which only the top

talent survive, and the belief that, as a community, physicists are immune to cultural and social influences, that is, they are 'cultureless'.

Culture in Academic Physics and Astronomy

To set up our examination of how these cultural beliefs influence academic programs, we must first establish what the cultural beliefs of academic physics and astronomy are. Though most of the literature discussed below focuses on physics (and science/higher education more broadly), in the absence of more information, we assume that modern astronomy and astrophysics have inherited similar cultural practices and beliefs as physics in general. Additionally, we recognize that the culture of academia in these specific fields in the US is inextricably linked to the broader culture of contemporary Western universities, particularly predominantly white institutions (PWIs). Thus, the descriptions of cultural beliefs that follow attempts to describe the communities that sit at the intersection of physics culture and university culture in a Western context.

'Survival of the Fittest' Meritocracy

Like most of contemporary Western science, physics and astronomy operate as highly individualistic and competitive endeavors, where a select few are exceptionally 'great' enough to prove their worthiness against the rigors of the field. To see how this belief can manifest in education, one need look no further studies of physics textbooks, which can simultaneously paint the history of the field as "the product of individual great men" while urging students "to assume that they are not going to be an Einstein or Dirac" [101].This characterization is heavily gendered, positioning science as "a form of competitive and aggressive activity, a contest of man against man that provides knowledge as a side-product" [102]. The origins of this persistent belief trace back to the late nineteenth century Cavendish Laboratory at Cambridge, where physics students were expected to compete in the Mathematics Tripos, a test combining physical athleticism with mental agility as a demonstration of "manliness" [103]. Interestingly, the inextricable nature of physics and competition is distinctly American, but the gendering of physics as masculine is not. In a comparative anthropological study of high-energy physics laboratories in the United States and Japan, Traweek finds that communities in both nations value masculine traits in physicists, but while American masculine traits include competition, individualism, and arrogance, culturally masculine traits in Japan include humility, cooperation and interdependence [101].

A justifying belief for this epistemic approach is meritocracy, or the belief that opportunities are distributed on the basis of individual effort, talent, and achievement rather than by heredity or privilege [104]. However, belief in meritocracy largely functions to obscure the causes of variation between minoritized and majoritized students' educational outcomes. Besides ignoring variation in initial conditions that can lead to varied outcomes, a belief in meritocracy has been demonstrated to result in biased decision making. When empirically tested, managers that explicitly valued meritocracy favored and rewarded male employees

over female employees with equal qualifications, a bias that effectively disappeared when the emphasis on meritocracy was absent [105]. This phenomenon can be seen in a study of physics graduate admissions which found that "some programs seek to admit students who are judged to have innate talent, primarily measured by grades and standardized test scores (especially the physics GRE)" [106]. However, because of racial and gender biases present in both standardized metrics and individual attitudes of faculty assessing students, this pseudo-meritocratic approach conceals the fact that it negatively impacts the admission of minoritized students. The belief in meritocracy also serves to protect the belief that negative outcomes for minoritized students are the fault of that student; an ethnographic study of physics (among other) programs found that the physics faculty interpreted high failure rates as their "simply dismissing the students who lacked the skills and commitment to study physics" and proudly considered such high rates to be a "natural feature of physics programs" [107].

It is important in discussing the belief in meritocracy to identify what is defined as 'merit', as the arbitrary designation of behaviors or attitudes as 'meritful' is itself an opportunity to introduce bias. Two traits stand out as being worthy of merit and relevant to this study in the literature examining the culture of physics: a single-mindedness for physics above all else, and relatedly, an enthusiasm to commit as much time as possible to that endeavor [101]. The former has been discussed previously in studies on the experiences of science-major students of color, which find that "survival [in programs] required a single-minded focus on individual [scientific] goals, which presented a conflict to some of the Black, Latino, and American Indian students in their study, who felt communication obligations to serve their communities and be a role model" [108]. Similarly, the latter has also come up in the literature–this time in women's experience in physics doctoral programs–where women were found to struggle with the time commitment asked of them, finding difficulty balancing graduate school with the rest of their life and feeling guilt for wanting to spend time on anything outside of graduate school [109]. While we will focus on these two traits in this work, it is worth noting that there are other gendered and racialized traits that typically constitute 'physics merit', such as secularism, stoicism, and masculine appearance [101].

Objective 'Culturelessness' and 'Apoliticism'

An ironic feature of the culture of physics is the prevalent idea that, as a community, physicists are somehow immune to the influences of culture [101]. This paradox is explained in a survey of feminist science critiques: "[The physicists'] nonsocial subject matter and the paradigmatic status of their methods appear to preclude critical reflection on social influences on their conceptual systems; indeed, prevalent dogma holds that it is the virtue of modern science to make such reflection unnecessary" [110]. However, as Hyater-Adams points out in her work understanding how Black folks identify with physicists, the belief in the independence of physics from the broader culture does harm to Black (and otherwise minoritized) physicists, who understand that, rather than disappearing in physics spaces, racism is in fact amplified by the collective denial stemming from the belief that physics is cultureless.

Similar to how meritocracy places the blame for negative student outcomes on the students, Hyater-Adams shows how in a 'cultureless' community of physicists, the blame for racism denied by the community ultimately falls back on to the person who experienced it [111], amplifying its impact rather than mitigating it. This mimics the function of 'colorblind' racial ideology in contemporary American society, which purports to support diversity while failing to address the existence of racism [112]. Thus, putative 'culturelessness' in the field serves to obfuscate the very real presence of systems of oppression inherited from broader American culture.

A foundational belief that holds up this image of 'culturelessness' is the physicist's dedication to the ideal of objectivity, by which impartiality and a focus on evidence take precedence over personal preference or feeling. Of course, human beings are cognitively incapable of true objectivity (think back to the supposedly meritocratic managers and admissions officers), and the perseverance of the ideal is simply another way in which men are prioritized; historical studies of science have indicated that the establishment of modern universities intentionally associated itself with masculine qualities of intellectual rigor and objectivity, while distancing itself from the feminine qualities of emotion and subjectivity [113]. The deeply social nature of what is considered to be 'objective' is described By Harell, stating "Claims that are held up as 'Objective' are actually 'socially negotiated'...and 'partial and stationed knowledge'...that emerges only through the interaction of particular individuals with particular cultural, social, and cognitive values" [114].

Closely related to this objective 'culturelessness' is an emerging veneer of 'apoliticism' in physics communities, particularly in the context of diversity. This aversion to the 'political' is present broadly in higher education, as shown by a study that found that white admissions counselors were more likely to respond to deracialized and apolitical black students than black students who demonstrate a commitment to racial justice [115]. As this example shows, 'political' here is not limited to electoral or federal politics, but the broader discussion of "who gets what, when, and how", which includes the sociopolitical aspects of diversity work [116]. This broad characterization of 'politics' as something to avoid has been found to be present in how diversity work is currently approached in physics education, with some physicists feeling that they should not be involved in political discussion, while others recognize the necessity of addressing broader social issues in the work of addressing diversity in STEM [117]. As an example, Cochran finds that "some individuals engaged in diversity work [in physics] are comfortable saying that Black Lives Matter and others are not", a 'political' tension that ultimately makes addressing anti-blackness in the field a matter of personal opinion and debate [117]. We draw a parallel here between 'culturelessness' and 'apoliticism' because they both seek to present their holders as neutral parties while acting as a veneer, preventing the underlying beliefs of the field (which are influenced by culture and inherently political) from being challenged.

4.3 Theoretical Framework

This work is guided by two analytical frames: equity as a cognitive frame, and critical race theory.

Equity as a Cognitive Frame

The first analytical frame is intended to guide where we look to identify causes of inequity. We adopt 'equity' as a cognitive frame, which, along with 'deficit' and 'diversity', comprise the three cognitive frames outlined by Bensimon in her analysis of inequity in higher education through a lens of organizational learning theory [95]. She distinguishes between the three frames in their orientation, discourse, and strategies. For example, while a person operating within the 'diversity' cognitive frame focuses on representation in demographics, discusses diversity and its value, and works to that end through methods like trainings and workshops; someone using the 'deficit' cognitive frame focuses on "stereotypical characteristics associated with the culture of disadvantage and poverty", discusses the shortcomings of minoritized students, and attempts to correct them through remediation based courses and programs.

In this work, we adopt the third frame, the 'equity cognitive frame', which focuses on "institutional practices and the production of unequal educational outcomes for minority group students", discusses the manifestation of systems of oppression in institutions and society more broadly, and approaches these problems through the development of institutions increasing accountability of equitable education. Bensimon describes how discourses that result from deficit or diversity cognitive frames are more likely to be heard when institutions of higher education discuss equity, despite their tendency to serve as cognitive 'blinders' to critical aspects of such discussions. To avoid the blind-spots of these frames, we will operate with the assumptions of the 'equity cognitive frame', namely, that it is not the shortcomings of students nor the implicit bias of faculty members that perpetuate unequal outcomes, but rather the "institutional actors", including faculty, administrators, and program directors, that maintain inequitable institutions as a result of their beliefs, values, practices, and expectations. As a result, this work aims to examine features of graduate programs and how Bensimon's "institutional actors" operate within them, rather than students themselves.

Critical Race Theory

Critical race theory (CRT) emerged from the work of American civil-rights legal scholars and activists in the mid-1970s to counter the mainstream understandings of racism and civil rights, which at the time were dominated by the principle of non-discrimination and a belief in an increasingly post-racial society [118]. Instead, CRT scholars aimed to fortify and further the gains of the Civil Rights movement, using law and academic scholarship to argue that, despite gains made against segregation and employment discrimination, there remain putatively neutral laws which continue to uphold white supremacy. CRT is founded on the premises that white supremacy exists and is upheld through laws and institutions,

and that achieving racial emancipation is possible through the tranformation of such laws and institutions. To this end, CRT is comprised of a set of tenets, three of which we will use in this work: the ordinariness of racism, the value of counter-storytelling, and intersectional analysis.

The Ordinariness of Racism

This premise stands in opposition to the majority culture ideals of meritocracy and color-blindness, asserting that such ideals not only obfuscate how people of color are marginalized in American society, but actively support that marginalization [119]. Stefancic and Delgado describe the way racism is the norm, rather the exception, in American society, saying it is "an ingrained feature of our landscape" which "looks ordinary and natural to persons in the culture". Racism is understood here to be the collection of policies and ideas that produce and normalize racial inequities which, in the United states, operate in favor of white people and at the expense of people of color [112]. Assuming this prevalence of racism in both systems and individuals within the US, this work aims to understand graduate programs by examining both how racist ideologies can manifest in program structures, as well as the extent to which such structures prevent or empower individuals to act on racial biases.

Standpoint Theory and Counter-Storytelling

Another central tenet of CRT is the use of counter-storytelling as a method of meaning making and challenging dominant narratives [120]. Counter-storytelling centrally positions the experiential knowledge of people of color and other minoritized groups in understanding and combating racism, as opposed to majoritarian stories which prioritize the social locations of Whites, men, and otherwise privileged groups as normative reference points. This tenet derives from standpoint theory, founded on the Marxist idea that members of an oppressed group have unique knowledge of their oppression that is not understood by members of the oppressing group [118]. As such, this work will center the stories of minoritized students in our understanding of how departments maintain inequity: the primary data set consists of such student's stories about their experiences in graduate school.

Intersectional Theory

Intersectionality combines the examination of race with other axes of identity (e.g., gender, class, or sexual orientation) in order to understand how their combinations result in unique forms of oppression. The term was coined by black feminist scholar and legal scholar Kimberlé Williams Crenshaw to challenge the dominant paradigms of discrimination, in which sexism is understood through the experiences of white women while racism is understood through the experiences of "the most privileged Blacks" [121]. The result of such division of analysis is an understanding that erases the ways in which multiply minoritized people can experience a form of discrimination that is more than the sum of its parts. Therefore, to obtain a more complete understanding of how various systems of oppression (racism, sexism,

etc) operate simultaneously in a system, this work seeks to understand the experiences of those who are multiply minoritized in that system- the this case, students who are neither cisgender-men nor white.

4.4 Methods

Sample

The primary criteria for participants in this work were that they have both minoritized gender and racial identities (i.e., not white and not cisgender men); that they are pursuing (or already earned) a PhD in astronomy, astrophysics, or physics; and that their program is/was in a predominantly white institution in the United States. We limited participants who had already completed their PhDs to those who had graduated within the last three years. We recruited participants by circulating the above criteria and our research goals to current astronomy and physics graduate students and postdocs via department chairs of the first 100 programs on ranked lists of programs[1], excluding minority serving institutions. Additional recruitment was done through advertisement via professional societies (e.g., Society for the Advancement of Chicanos/Hispanics and Native Americans in Science) or social media hashtags (e.g, #BlackInAstro). All interested parties completed an initial screening survey soliciting demographic information as well as basic program information (e.g., program length, and candidacy status).

The initial screening survey totaled 145 respondents. This sample was narrowed by removing respondents that did not fulfil the preliminary demographic criteria (leaving 87), eliminating current students who had completed less than three years of their program (leaving 36). From this group, we retained all nine URM-identifying students (note that URM is defined here as African American/Black, Hispanic/Latinx, Native American/Alaskan Native, Native Hawaiian/Other Pacific Islander, Filipino, Hmong, Vietnamese, or two or more races, with at least one from the preceding list). The sample was then reduced to 14 participants (set by available funds to compensate participants) by prioritizing respondents with multiply minoritized identities along other axes. Ultimately, two participants were removed from the data set, due to complicating variables brought up in their interviews (e.g., completing most of their degree outside of their official institution), leaving a final sample size of 12 participants. Due to the identifiable nature of multiply minoritized identities in a highly homogeneous field, we will not give individual demographic information for the participants. Instead, list the participant pseudonyms with career stage in Table 4.1, and describe the demographics of the sample as a whole.

All participants in the final sample focused their research in astronomy, whether in an astronomy department, in an astronomy and physics department, or in an astronomy research group within a physics department. Two graduated within the past year, while the remaining 10 are currently enrolled and have passed their qualifying examinations. All participants'

[1] US graduate programs for 'physics' and 'space science' from www.usnews.com

Table 4.1: Final Participant Sample

Pseudonym	Current Position	Year
Muna	Graduate Student	4
Nur	Graduate Student	4
Khadija	Graduate Student	4
Ferdous	Graduate Student	4
Mariam	Graduate Student	5
Anwar	Graduate Student	5
Sarah	Graduate Student	5
Marwa	Graduate Student	5
Sawsan	Graduate Student	5
Amina	Graduate Student	6
Mabrooka	Postdoc	-
Najla	Professor	-

Pseudonyms of the 12 participants used in the final analysis of this work, along with each participant's position at the time of interview. For graduate students, the year in program is also reported.

programs end in a PhD, with normative times to completion from 5-7 years. The participants ranged in age from 26-34. Eight participants have URM identities (3 Black/African American, 3 Hispanic/Latinx, 1 American Indian/Alaska Native, and 2 Filipino). Nine participants identified as cisgender women, while 2 participants identified as genderqueer and 1 reported their gender identity as 'female/questioning'. The sample contained 2 international students, 3 participants who practice a non-dominant (i.e., non-Christian) religion, 6 participants with non-heterosexual orientations, and 5 participants from low-income backgrounds.

Data Collection

The data presented in this work were collected in individual interviews conducted and recorded by the first author over Zoom. Interviews ranged from 1 to 2 hours, with two participants, Nur and Mariam, having a second interview within a few weeks of the first. Interviews were conducted in a semi-structured format, using an interview protocol of 30 open-ended questions for all participants, but refining questions along with the flow of the conversation. Questions focused on topics including department culture, senses of othering and belonging, and academic experiences.

In addition to open ended questions, the interview included the Trait Ranking Exercise (TRE), an activity designed to outline what participants see value in scientists, as well as what they perceive their departments to value in its graduate students. The TRE was conducted on a collaborative digital whiteboard, where participants could click and drag items to rank them. Participants were given 22 'trait cards' to position on arrows from

'most positive' to 'most negative'. The traits were selected based on the the literature review on physics culture, with about half as 'consistent' traits (e.g., 'competitive'), about half as 'inconsistent' traits ('nurturing') and a few neutral traits ('creative'). This set of traits was ranked twice, first with the prompt 'Traits you believe are valuable for a physicist/astronomer to have' and second with the prompt 'Traits you believe your department values in its graduate students'. Participants were prompted to talk through their thought processes while sorting. Following the sorting, participants were given the opportunity to add traits they considered important for either ranking set, so long as they add that trait to both sets. Additionally, five questions of the interview protocol were focused on the rankings that participants generated. Both the participants' dialogue regarding the TRE and the sorted traits are analysed as data in this work. The dialogue was analysed along with the rest of the interview, while the positions of traits on the arrows in the TRE were converted to a point scale, from -3 to +3. For each trait, the mean and uncertainty in the mean of the distribution was obtained and are reported in Section 4.5 to contextualize interview data. The aggregate results of the TRE are illustrated in Figure 4.1, with violin plots showing the distributions of value rankings for each trait, with the participants' opinion in green and their understanding of their departments' value in orange.

Data Analysis

Interview recordings were transcribed and coded for emergent themes using the qualitative data analysis software `MAXQDA`. The analysis process was highly iterative, with our code system and research question evolving with each pass through the data set. As described in Section 4.1, the initial goal was to understand how our participants observe, relate to, and interact with the culture of their department.

The zeroth order coding pass happened concurrent to transcription, initially as rough in vivo code generation, using participants' words to generate a list of potential codes relating to experiences regarding department culture, as well as effects, coping strategies, and potential solutions. Once the full data set were transcribed, these initial code ideas were organized into ten broad categories, such as 'department community', 'DEI work', and current events', each with 3-10 subcodes. Using `MAXQDA`, all transcripts were then coded line by line with this coding scheme, marking pieces of text with the appropriate subcode, or if necessary, creating a new subcode in the appropriate category.

Following the initial pass at coding transcripts, the relative frequency of different codes was used to re-organize the coding system by identifying four dominant themes: the creation of researchers, conflict and diversity, equity, and inclusion (DEI) work, gendered and racialized expectations, and community and socialization. Again, the transcripts were re-coded using this new scheme. The coding scheme went through five iterations, with the final version focusing on commonalities between participants' programs' academic structures and approaches to conflict management and DEI work.

As the coding scheme developed to better organize the data, the research questions also evolved, moving from the general (how do students perceive and interact with their

departments' culture) to the specific (how do these students observe the cultural beliefs of the field to affect their programs' structure and their experience within those structures) as the authors' understanding of the data evolved. The data coded to answer the developed question were grouped into participants' descriptions of attitudes, policies, and behaviors in their department that relate to the two cultural beliefs outlined in Section 4.2.

4.5 Findings

In this section, we present interview and TRE data that indicate a pattern of response which related to program structures and faculty practices that correlate to the two cultural beliefs discussed in Section 4.2, as well as the student experiences that result from these structures and practices. Quotations and summary descriptions of interviews will be presented, along with value rankings from the TRE on a scale from -3 to +3. Each cultural belief is connected to a different component of graduate programs: survival of the fittest meritocracy with academic structures, and objective 'culturelessness' and apoliticalism to conflict management.

Academic Structures

A foundational component of culture to examine is what Traweek calls "the developmental cycle, or how the group transmits to novices the skills, values, and knowledge that constitute a sensible, competent person" [101]. In the context of graduate programs, we can think of this as the process by which a student with an undergraduate education is taught the skills to conduct original research. Importantly, we want here to understand not just how grads are trained, but in what ways their access to adequate training may be barred. To this end, we will start by examining the relevant formative belief.

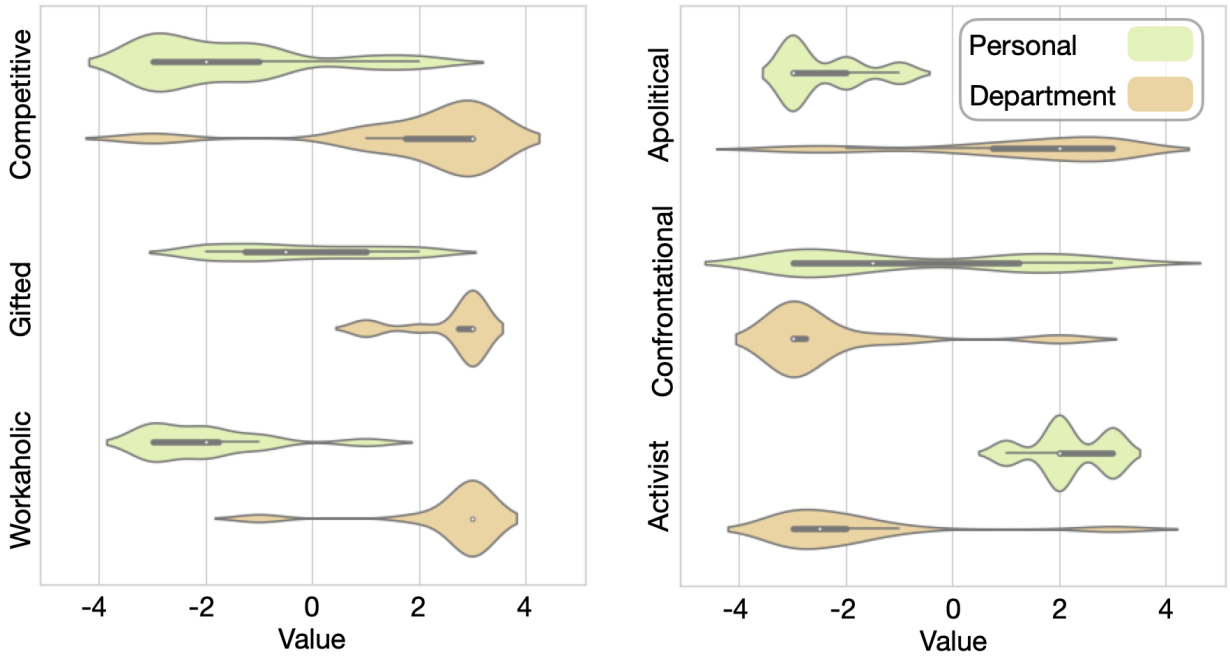
Cultural Belief: 'Survival of the Fittest' Meritocracy

Four students explicitly described an attitude of 'survival of the fittest' present in their departments, by which students who are judged to be 'good' are believed to deserve their place, and the 'failure' of those who don't measure up is natural and not worth attention. Najla illustrates this attitude in the context of their department's comprehensive exam:

> A lot of the attitudes around the comprehensive exam are, either you got it or you don't.... I see this all the time where people say, "Oh, you're good at physics, so we're gonna work with you more" or "You're bad at physics, so instead of trying to train you to become better, we're just going to say no, you're bad at this."

This fixed-mindset approach to evaluating grads is echoed by how participants' understood the traits 'gifted' and 'apt' (which are meant to represent fixed and growth mindset

Figure 4.1: Violin plots illustrating the aggregate results of the 12 participants' responses in the Trait Ranking Exercise (TRE) for six traits discussed in this work. For each trait (designated to the left of the data), the distribution of participants' personal value is shown in green, while the participants' observations of their department value is shown in orange. The inner component (gray line, gray bar, and white dot) of each violin function as a box and whisker plot, showing the full range of the data, the central two quartiles, and the median (respectively). The outer 'violin' component of each plot shows a kernel density estimation of the underlying distribution, with the vertical 'thickness' indicating frequency of response at that value. The six traits shown here are only the subset of traits in the TRE that were relevant for our analysis. The results for the remainder of the traits from this exercise are presented in Figure 4.3.



thinking, respectively) to be valued by their departments, with 10 of the 12 participants ranking their department as valuing being 'gifted' above being 'apt'. Moreover, one of highest ranked traits for departments after 'gifted'(at 2.58±0.75) was 'competitive' (2.0±0.58).

Program Structure: Unclear Academic Expectations

All but three participants expressed frustration due to what they believe was their departments inadequate communication of expectations on grads. For example, they noted that confusion about typical benchmarks in their program, what an acceptable amount of research is, expectations on teaching and service, how to organize and interact with thesis committees, how success is defined for exams, and requirements on advancement to PhD candidacy and theses completion, among others. Khadija's experience illustrates the confusion and possibility for exploitation that arise when such fundamental program expectations are not clearly defined:

> Your advisor would tell you something, but the head of the department would tell you something else because they don't actually know. And they don't communicate. And so one of the professors expects all of his students to TA at least one semester because he thinks that's a very important thing. And so he tells everybody he meets 'everyone in this department TAs at least one semester'. I never TA-ed, because my advisor didn't care about that. But this guy's going around telling everybody this... It's kinda like, whoever you heard it from, you have to fact check with 10 other sources to know what the actual truth is.

This is a critical example of the difficulty unstructured programs can cause in determining expectations where none are given, as it shows not only hassle of finding expectations when none are given, but also how faculty can leverage that ambiguity to their advantage and potentially, the student's detriment: if there is not a clear policy in place, a faculty member has the freedom to answer student questions as they see appropriate, and this freedom can be used 'define' policies that are beneficial to them. On the other side of this coin, four participants reported issues stemming from unclear expectations on the part of faculty–that is, not knowing departmental advising norms. A notable instance of this was shared by Sawsan, who was not aware that one-on-one between grads and their advisors was a typical expectation:

> I didn't have individual meetings with my advisor for two or three years. I just thought that's how things were.... I just didn't know that's something that I could have requested...It was never communicated to me that we should have one."

Some participants observed a relationship between this lack of clear academic expectations and faculty attitudes on who is a 'good' graduate student, with the idea that a 'good'

enough student simply does not require a thoroughly outlined academic path. Ferdous comments on this relationship, saying:

> "There's a lack of expectations and there's a lack of explicit communication... And that kind of comes back to what the department's ideal graduate student is. The ideal graduate student doesn't need that kind of communication. They don't need to be taught how to do things."

This student's point begs the question: in the absence of a clearly defined process to or definition of success, how do faculty approach the education of grads?

Faculty Practice: Selecting 'Pre-Trained' Researchers

The first practice we can examine is participants' accounts of who is favored by faculty in their departments. On the most basic level, 6 participants describe their department's ideal grad as one that is 'pre-trained'– that is, they enter the program already having all the necessary skills and knowledge to conduct original research, are self-sufficient, and require little to no guidance, education, or support. Ferdous describes this expectation as they understand it:

> Their ideal grad student is basically one that is already a self-sufficient scientist. Doesn't need to be taught anything about being a researcher, being a scientist. Doesn't need any sort of non-academic support. They are just able to function immediately once they get in the door.

One of the most brought up characteristics of this 'pre-trained' student is a narrow focus on research. While research is of course a critical goal of a graduate program, 10 participants experienced an emphasis on research that they found excessive. Six participants described a prioritization of research above all that they found excessive. Muna gave an example of a teacher training being deprioritized in favor of focusing on research:

> They would much rather you focus on your science...I've seen a lot of faculty members look down on people that want to do outreach or want to do things that involve education and stuff... We have this program, it's a certificate in teaching science and math at the college level... I've seen faculty members look down on those grad students just be like, 'Oh, why do you want to teach? What's the point of teaching? I don't get it. You're just going to be a researcher.'

In addition to the pedagogical training referred to in this example, other participants describe being discouraged from participating in professional development, equity work, science communication, outreach, and identity-based conferences on the basis that they are less important than research or that these practices aren't perceived to align as well with a research trajectory. Ferdous explains how this narrow research focus is valued by their

department, saying "the ideal grad student is a researching machine...You are here as a scientist and nothing else."

Faculty Practice: Unlimited Work Expectations

Another common faculty practice was an expectation for students to spend "as much time as possible" on research. This can be seen in the trait rankings, where 'workaholic' was tied with 'gifted' as the highest ranked value for departments (2.58 ± 0.75). Eight participants describe faculty highly valuing students who are willing to work extra hours, often at the expense of other areas of life. This was generally communicated by faculty either directly stating it (Muna was explicitly told "If you're not working 80 hours a week, then you're not doing a good job") or by praising grads who voluntarily commit themselves to extra work. Mariam observed a conflation of this voluntary and excessive work commitment, and perceived talent, saying:

> There's always going to be students who are workaholics, and the faculty who say, yeah, we have another paper. It's great. He's so smart and really talented. That's the word they use when they're workaholics. Are they talented or are they just working 24/7? Maybe it's both. I don't know. I could probably also be talented if I worked 20 hours a day.

Several students report this attitude in their advisors, despite being unionized (and therefore having legal caps on the amount of work asked from them). Najla discussed advisor expectations that were inconsistent with their union rules, saying:

> I was explicitly told... 'for seven days, 24 seven, you were available because we had this deadline coming up... And I expect that from you all the time'.... We are a part of a union. I get to do 20 hours a week of teaching no more.

Something to note is that, while almost all participants discussed the perceived importance faculty put on working as much as possible, no participants brought up any evaluation metric for research besides quantity of papers published, which is generally how they observed faculty to identify 'good research'.

Faculty Practice: Discouraging Questions

This high prioritization of 'pre-trained' students who commit themselves solely and completely to research is juxtaposed by the experience of grads who, though by no means are unprepared for graduate school, require some education before becoming fully formed scientists. One example of this difference is shown by some faculty's responses to grads asking questions. Marwa gives a salient example of this, from a time when she and the other two women that made up her cohort went to a professor's office hours for help with a difficult problem set:

> We thought maybe it would be easier for us if we went as a group and just explained that we're having difficulty because each of these problems are taking us upwards of 10 hours to do... [We asked], 'Are there other resources that we can have access to, like another book that you might recommend or something?', And the professor just looked at us and flipped through the problems and said, "You shouldn't be taking that long..." And he just continued to look at us like we were idiots and was like, "I don't know, I feel if you can't do this, then maybe..." And it was very clear that he was thinking: if you can't do this, then you shouldn't be here.

Six participants total gave examples of scientific questions (as opposed to bureaucratic or programmatic questions) being discouraged by professors in similar ways, both in one-on-one situations and in front of classrooms of peers. Mabrooka described asking professors to clarify concepts she already understood because she recognized it was being taught poorly, as she had completed a masters degree in astronomy before enrolling in a PhD program. In response, she felt that her peers thought she didn't know anything because she is "always asking questions".

The tendency to treat asking for help as incompetence occurred in a more general sense, as well. Three participants describe the way that in their departments, any difficulty grads may experience (be it academically or otherwise) is framed as a failing on their part; and subsequently, two of these participants recount hiding any struggles they may have had for fear of being labeled a problem student. Sawsan explains this conflict, saying:

> I feel like I have to pretend that I'm confident, that I know what I'm talking about... I can't tell them that I'm frustrated or I have problems... And whenever I do mention, I'm willing, I'm choosing to be vulnerable, the response I often get is "Oh, I'm sure you'll get better." It seems it's my problem that I'm individually having, that no one else is having problems.

This cutthroat learning environment where questions are treated as an indication of inability rather than an opportunity to learn may be related to most participants ranking 'nurturing' as being negative or neutral values for their department (-0.5 ± 0.14).

Student Experiences: Inconsistent and Unaccountable Education

Though only two participants explicitly made this observation, the ubiquity of problems caused by unclear program expectations in our sample suggests that participants' overall graduate school experience is largely a function of who their advisor is– for better or for worse. Amina framed their impression of graduate school through this observation, saying:

> The grad school experience can be very different for different people, and it is mostly because the programs are designed in a way that they are very dependent on your advisor...So your advisor is going to determine how your experience

> would be...Some people need to spend most of their graduate school time just doing research. Some other people can have a balance between life and grad school and research. And all of them are depending from my perspective on the advisor."

There are a few salient outcomes enabled by these circumstances, the first being the possibility for grads to be effectively unadvised by any department faculty. This was percieved to be a persitent or recurring issue in programs, discussed by three participants. In most cases, issues between the grad and their advisor led to the grad seeking advising outside of their department, keeping their department 'advisor' on in a nominal role. In a different case, Mabrooka had to 'change advisors', only to find that no faculty would then take her on and that the department had no contingency for such a scenario:

> I asked people, would they advise me? They all said no. So then I was like, okay, what am I supposed to do?... I was in this department, watching all of these other grad students meet with their advisors, like weekly going to group meetings with each other. And I was just sitting there in my office . . . [The faculty] didn't justify [their refusal]. And then within a year they took on new students. I knew it was just that they didn't think that I was going to be strong enough to work with them, that it was going to be too much work on their part to bring me up to speed.

This tenuous assurance about support from an advisor takes on another dimension in the case of international students, where their visa status may be dependent on academic appointments. Amina, who is an international student from a country on President Trump's 2017 travel ban, discussed their worry about the precarity of being international and having conflicts with an advisor, saying "[faculty] don't have any thought about our visa situation, how fragile it is."

Short of academic abandonment and the looming threat of potential deportation, participants experienced a myriad of other issues with faculty resulting from the lack of clear expectations in the student-advisor relationship. These issues included faculty encouraging students to leave the program prematurely, having grads compete for projects or funding, avoiding funding students entirely, and taking advantage of student's relative lack of information to assign them more work.

Conflict Management and DEI Structures

In addition to robust processes to prepare novice scientists, graduate programs need structures in place that are readily available to address issues of conflict and to promote diversity, equity, and inclusion (DEI). In this section, we will focus on how conflicts arising from harassment or discrimination get addressed, particularly as they related to students with

minoritized identities. This is in line with the aim of promoting DEI (that is, creating outcomes for all students, regardless of minoritization). Again, we begin by examining beliefs that undergird and help shape these structures.

Cultural Belief: Objective 'Culturelessness' and 'Apoliticism'

Seven participants discussed their departments' value of being 'apolitical', with three of them reporting that they were explicitly told at some point to "not be political". Describing their interaction with a faculty member who is responsible for generally advising graduate students, Sarah says:

> She has told a number of students, including myself, when [we ask]... 'What sorts of things do you have on campus that will support students of color, or women of color?' And one of her comments is, 'Oh, don't get political. Let's not get political. Just focus on your research. Don't think about that at all.'

It is worth noting that generally, the topics that participants understood as being labeled 'political' were largely issues relating to minoritized identities, such as anti-blackness in departments follwing the #BlackLivesMatter protests of Summer 2020. Nur elaborates on these specific and subjective definition of 'political' and 'apolitical', saying:

> Usually it depends who I'm talking to what 'apolitical' even means. If I'm talking to a white, wealthy, male faculty, 'political' can mean something as trivial as talking about trans rights or talking about having a gender neutral restroom. That, to them, can seem very political.

This emphasis on being 'apolitical' can also be seen in the results of the ranking activity, where participants rated it as one of the most highly and consistently valued traits in their departments (1.25 ± 0.36).

Program Structure: Ineffective Systems of Accountability

The most concrete structure to demonstrate the influence of this aversion to 'politics' in our data set was department codes of conduct. Five participants reported that, though their department may have a code of conduct that prohibits harassment and other problematic behaviors, they are largely ineffectual, with one participant saying their department does not even have a code of conduct. Mariam shared a representative story in which students reported a faculty member for making racist and transphobic remarks in classes and colloquia, and despite the department having "a robust code of conduct", no action was taken aside from "limiting their intersection with students". They described the frustration and false sense of security from this lack of accountability:

> [We don't have] clear pipelines for repercussions, for breaking this supposed code of conduct that's supposed to help us... There's absolutely nothing on there about what happens if you do those things... There were several grad students, especially more junior ones, who would come in, see the code of conduct, feel protected by it, then watch someone break it, and try to do something about it. And nothing happens.

Participants who spoke to this point found that, in the absence of clear systems of accountability, faculty (particularly those with tenure) were effectively unrestricted in their behavior. Khadija spoke about their frustration with being unable to expect "basic, what-you-learned-in-kindergarten respect", saying:

> [For faculty,] you can say whatever you want, you can do whatever you want. Especially if you're the head of the department, you can do whatever you want, nothing bad is ever going to happen to you. And so that kind of gives you more of a reason to say whatever you want and it gives other people less of a reason to try and stop you.

Faculty Practice: Avoiding DEI Issues

In the absence of clearly defined procedures for dealing with harassment and discrimination, participants reported a range of problematic faculty approaches to DEI conflict management, the most frequent of which was simply to not acknowledge the conflict. Seven participants described faculty members who avoid discussing problems that exist in their departments. Ferdous described how this avoidance (in this case, justified by the departments high ranking) made it difficult to improve the department climate:

> It's very difficult to point out the flaws that are in the department when the people who are in leadership... don't want to rock the boat. Don't want to acknowledge that there might be issues. Don't want to acknowledge that the way they've been doing things can be improved. Especially when that department [is highly ranked], they think that they're doing something right.

Besides not acknowledging conflict themselves, 4 participants described how their department's faculty value students who "don't rock the boat". In describing their department's ideal graduate student, Sarah said:

> Students that are less interested in political aspects of the human experience, anything to do with diversity or inclusion...students that just purely focus on putting their nose down and doing good research. I think that those are the students that are most rewarded in my department: the students that don't push back against the established system.

Other ways in which participants were implicitly encouraged to not criticize their departments included feeling "tone policed" or "gaslit" when discussing racism or other systems of oppression (4 participants) and witnessing retaliation to those who call out problematic behavior (3 participants). These observations are consistent with the ranking of the trait 'confrontational', which participants ranked as the most negatively valued trait by their departments (-2.33±0.67), with the majority of participants noting that they interpreted 'confrontational' roughly as discussing existing problems.

Faculty Practice: Discouraging DEI Work

In addition to ignoring DEI problems, seven participants discussed faculty members actively discouraging DEI work. For three participants, this discouragement came in the form of the previously mentioned 'research comes first' attitude; Ferdous put this discouragement in the context of research prioritization, saying:

> It comes back to, they don't want us to have any other stake in anything else [besides research]. . . . [Faculty ask,] 'Why are you spending all this time doing non-research things? Going to conferences, trying to recruit or talk to other minority students about your experience? And why are you doing these things that are not science related?'

In some cases, observed attitudes towards DEI were not simply deprioritization, but outright antagonism. Marwa recounts seeing this attitude in faculty both in the assessment of program applicants and current grads, saying:

> We had some [applicants] explicitly include equity and inclusion related stuff in their statements of purpose and [faculty] would be like, 'I don't know, I feel like grad students are entitled enough already.'

This emphasis on discouraging DEI work is consistent with participants' ranking of the trait 'activist', which they ranked as highly negative in their department's values (-2.0±0.58), second only to 'confrontational'.

Faculty Practice: Performatively Promoting DEI

Despite these modes of de-emphasis on DEI work, all participants mentioned some DEI efforts existing in their departments. However, nine participants found these efforts–at least, in part–performative. This performative approach to DEI work was observed in many forms, the most frequent broadly being the performance of symbolic efforts that lacked follow through, for example, publishing a statement in reaction to #BlackLivesMatter affirming a commitment to combating anti-racism in a department, but failing to fulfill those commitments. This specific example came up several times, with Mariam's case taking on an extra dimension of questionable authorship:

> A bunch of grad students and postdocs wrote [a statement]. And then [the faculty] took it, took out the things that were too extreme for them, and then published it as their own... And not only that, they obviously haven't done anything that they said they would do on that piece of paper that they watered down to begin with.

Participants who discussed these 'broken promises' of performative DEI efforts report it happening both as department-wide projects (like the #BlackLivesMatter statements) and in one-on-one interactions with faculty (for example, agreeing to help a student in some way and not following through). In the case of individuals, four participants encountered faculty members who simultaneously presented themselves as 'allies' in public while actiong differently with minoritized students in private. Muna observed this performance play out on Twitter, saying:

> One of the professors, they're just super, super performative on Twitter...."I'm so diverse, I'm so progressive, look at me" and things like that. But behind closed doors, he would just be the worst person ever, saying the worst things about students...about people of color.

Other things participants brought up as demonstrations of performative DEI work include 'bait-and-switch' recruitment (which three participants used to describe their own recruitment), tokenizing the minoritized identities of students while not supporting them adequately, directing all DEI efforts as outreach rather than internal department work, and only engaging in DEI work in response to crises (such as a department scandal or national tragedy).

One complexity of this feature is the existence of DEI efforts uninformed by an intersectional lens, or in other words, efforts that do not recognize the multiple ways people can be simultaneously marginalized. For example, four participants sought out venues for supporting women in their work, only to find that these venues of primarily white women were uninterested in supporting the unique positions of women of color, if not outright hostile to them. Muna describes this mixed support and antagonism for different axes of their identity in the context of an individual faculty 'ally':

> This is the first time I had run into the situation where someone who I saw as a champion of feminism and black liberation and the rights of people of color also believes some pretty transphobic things... It's pretty awkward to continue working with someone who's advocating for you in some ways, but also advocating against you in other ways.

Student Experiences: "Suffering in Silence"

The most immediate outcome of the practices described above was an educational environment in which students with minoritized identities were effectively unprotected from discrimination and harassment, whether from faculty, peers, or other department members. This is

echoed by all but one participant reporting at least one instance (though usually multiple instances) of racial, gendered, or otherwise related to minoritized identity microaggressions from department members. These microaggressions included racist, sexist, or otherwise discriminatory remarks from faculty and exclusion from academic and social activities by fellow grads, among others.

There was additionally a general sense that, because of the lack of channels for reporting such harassment and overall lack of interest in rectifying such problems, there was no recourse aside from simply dealing with it. Mabrooka describes the pressure to accept inequitable treatment, which she observed acting both on herself and some peers:

> If you never point out a flaw in someone or ask for something to be different, it might help you to succeed...Any sort of feedback that calls into question [faculty's] perfect advising or their perfect teaching or whatever, it's viewed as a negative on the person who's [giving feedback]... [Students] just suffer in silence, terrified to say anything.

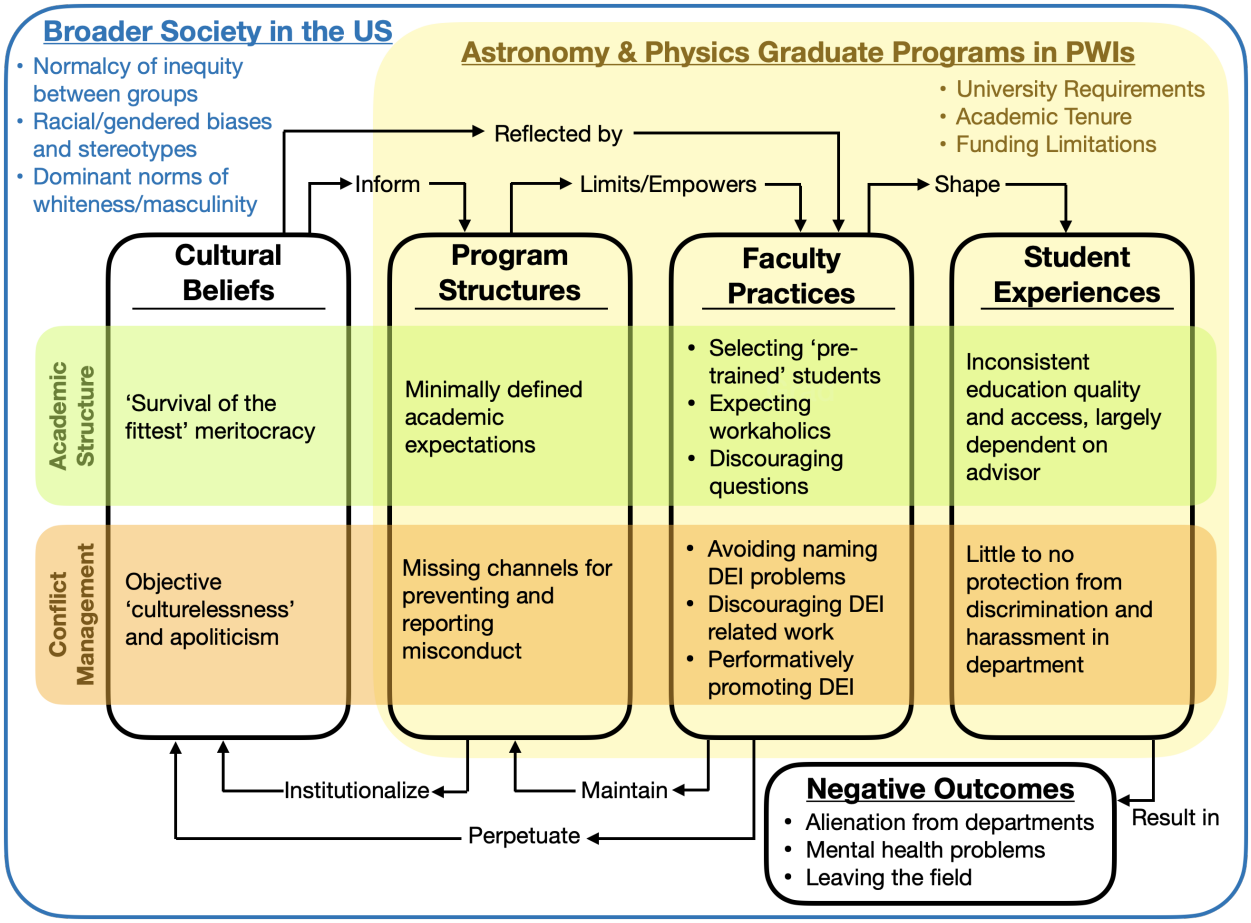
This circumstance of "suffering in silence" was common, with five students describing similar situations of struggling with interpersonal problems and feeling unable to seek solutions in their departments.

Student Outcomes

The student experiences laid out above (inconsistent education dependent on advisors and lack of protection from misconduct) affected the participants of this study in a myriad of ways. To begin, 8 participants reported struggling with research as a result of their experiences in their departments (for example, problematic interpersonal interactions). Five participants reported interactions with department members leading to feelings of dissociation or apathy, while another five felt alienated enough from such interactions to avoid coming to their departments whenever possible. Participants also reported suffering from a variety of mental health problems following experiences in their departments, including reaching a point of burnout (3 participants), dealing with anxiety and/or depression (6 participants), and being driven to substance use as a coping mechanism (2 participants). In a similar vein, seven participants felt the need to seek therapy or counseling due to their experiences in their departments.

Another important outcome to consider is participants' career trajectories, which varied in several ways. While nearly all participants entered graduate school with the intention of ultimately becoming faculty at a research university, at time of interview only four confidently felt that this was still their goal. Five participants describe their career plans changing in non-academic research directions (e.g., industry or observatory work) while three participants report their goals oscillating between their initial plans for academia and leaving the field.

Figure 4.2: A model system outlining how cultural beliefs, program structures, and faculty practices work in tandem to create student experiences that result in negative outcomes. The functioning of this system is illustrated with two regimes: that of academic structures (the green band) and conflict management (the orange band). The context of this system is outlined by the yellow region (which represents an individual department, limited by the broader university) and the all-encompassing blue rectangle (which represents American society at large).



4.6 Discussion

It's the System, Man

Following two threads, our findings in Section 4.5 illustrate attitudes in department members that align with the cultural beliefs of physics identified in Section 4.2, aspects of graduate program's design related to these attitudes, and common faculty member practices that operate within that context. Further, we have outlined the experiences of multiply minoritized grads along these threads, as well as their educational outcomes. Figure 4.2 presents a summary of these findings (grouped as 'cultural beliefs', 'program design', 'faculty practices', 'grad experiences', and 'student outcomes'), how they relate to each other, and how they are positioned in the broader contexts of universities and American society at large. We discuss these findings in two parts: first, we outline the systemic interaction between beliefs, design, and practices, and second, utilize literature to understand the impact of the individual practices on students.

Stepping through the feedback loop connecting 'cultural beliefs', 'program design', and 'faculty practices' in Figure 4.2, the upper set of arrows illustrates how ideas shape the physical world. The most persistent cultural beliefs of a field inform how that field operates, including how academic programs are set up; we can see this in the preservation of certain traditions in physics, such as the modern competitive environment of physics and astronomy programs hearkening back to the Mathematics Tripos contests of 1800's Cambridge physics labs [103]. These beliefs are also reflected by faculty members in their practices and attitudes: this same emphasis on competition as the ideal mode of knowledge production, for example, is apparent in the tendency for physics faculty to accept high failure rates as a necessary and natural part of the process, rather than a shortcoming of the system [107]. The ability of faculty members to act on such beliefs is, to an extent, determined by the program design, which can function to limit the freedom of faculty members in some areas while empowering them to behave as they like in others.

In the opposite direction, we can see how the choices of Bensimon's 'institutional actors' solidify the ideas of the culture by following the leftward arrows along the bottom of Figure 4.2 [95]. Faculty members maintain their department graduate programs, and can choose to adopt modern, research-based pedagogical methods, or elect to preserve the way things are traditionally done. While some of the design of graduate programs is dictated by the larger institution (for example, university-wide requirements for benchmark exams or campus policy on sexual harassment) many of the details of a program can be determined within a department, such as the design and evaluation of benchmark exams or how conflicts between students and advisors are handled. However a program is designed, it has the opportunity to institutionalize the values and beliefs of the designers, and by influencing this institution, faculty members are able to perpetuate the cultural beliefs, completing the feedback loop.

This context of these system components also affect how they interact. As mentioned, the design of programs is limited by broader university requirements; similarly, faculty members and students are also limited (or empowered) by university policies, the existence of

academic tenure, and access to funding, among other factors. Zooming out even farther, the university and broader physics community exist within American society at large, which, as established in our CRT framework, is understood to maintain white supremacy (and similarly, patriarchy) by default (Delgado, 1995). As such, we should expect those socialized in the United States (i.e., American professors, and to an extent, foreign-born professors who have studied or worked in the US for some time) to have absorbed some of the ideas that maintain its systems of oppression, such as racial and gendered stereotypes, or norms of masculinity and whiteness [112]. Because it is faculty members who have the most immediate influence over the experience of grads, it is critical to understand how the design of programs can prevent or allow faculty members to project these absorbed stereotypes and norms onto their students.

The Green Band: Academic Structures

To put the faculty practices presenting in Section 4.5 into context, we begin with those relating to academics, following the green upper band in Figure 4.2. The first practice, discussed in Section 4.5, is the preference towards working with who they perceive to be gifted– namely, the 'pre-trained' students, who are described as entering programs with most of the knowledge and skills required to independently conduct research. There are two problematic aspects to note in this practice: the first is the well-documented inability of physics professors to evaluate the quality of students without racial and gender bias. This can be seen at the graduate level both in professors' preferential assessment of white and male applicants to graduate programs [106] and their higher ratings of identical post-doctoral candidate applications with white, Asian, and male sounding names over those with female, Black, and Latinx sounding names [122]. As such, we can expect professors' assessments of who the 'gifted' students in a program are to be similarly biased, effectively resulting in a preference for white, Asian, and male students justified by prejudiced notions of intelligence.

Even in the absence of assessment affected by bias, this preference for pre-trained students has the potential to perpetuate inequity, essentially propagating inequity from upstream in the educational pipeline. Far more literature has been dedicated to understanding the nature and sources of inequity at the K-12 and undergraduate levels than at the graduate level, indicating variation between racial and gender groups' academic preparedness at each subsequent stage [94]. Rather than investing in the creation of excellence, our findings suggest that many faculty select students on the basis of already existing excellence, tacitly accepting and perpetuating the inequity of preceding educational structures. This practice is consistent with a 'fixed-mindset' approach to learning, which presupposes that intelligence and ability are static traits, rather than things that can be developed over time (as a 'growth mindset' approach would assume). Interestingly, research would suggest that this belief not only explains away equity gaps, but reproduces them as well: a study of over 150 STEM professors found that racial achievement gaps were twice as large in courses taught by 'fixed mindset' faculty compared to those taught by 'growth mindset' faculty' [97].

Section 4.5 outlined another preference faculty demonstrated: students who work without

limit. In the absence of standardized work expectations for students, participants indicated that advisors often had expectations that their students work beyond typical work hours. While this attitude has been pervasive in physics for some time [101], the impact it has on minoritized groups in the field has been studied at length, with women in math and physical sciences reporting the difficulty of balancing demanding work hours with family responsibilities to be the most common barrier to success in their fields [123]. In general, achieving work-life balance is significantly correlated with positive mental health outcomes, with one study finding graduate students who report poor work-life balance being more than twice as likely to suffer from anxiety or depression [40]. The same study found higher rates of depression and anxiety in cisgender-female graduate students (and yet higher rates in transgender graduate students), a trend consistent with our six participants who reported struggling with such things all indicating this expectation of excessive work in their department's faculty.

While these two practices work to elevate dominant-group students, the third practice, discussed in Section 4.5 is directed the opposite way: responding to student questions in discouraging ways. This dynamic has been studied in male-dominated undergraduate courses, with various works finding that asking questions reduced women's perceived intelligence in the eyes of their classmates [124] and women of color (particularly those with darker skin) generally avoid asking questions in large lecture classes due to that perception [108]. It comes as no surprise, then, that so many participants described questions being met with negative judgments of their ability or intelligence. Discouragement from educators has been demonstrated to have a range of negative effects on students, a notable one in the context of higher education being the gradual diminishing of their degree aspirations [125]. This practice, then, functions to burden minoritized students in two ways: stigmatizing them to the point of not feeling able to ask questions, and enhancing the effect of stereotype threat if they choose to do so anyways.

The cumulative effect of these practices in the context of an academic program with little structure or oversight results in an extremely variable academic experience, influenced by student identity and ultimately shaped by faculty members' choices and actions. Through the praise, recognition, and opportunities participants describe faculty members granting to students they perceive to be gifted and hard-working, majority-identifying students are more likely to have positive experiences in their programs, while minoritized students are more likely to be undervalued, discouraged, or overlooked entirely. Of course, some participants had positive experiences with their advisors; about half described having a good scientific and social relationship, in most cases citing their advisor as the only or one of few 'non-problematic' faculty members in their department. However, there are two things worth noting about this point: first, that many of the behaviors described in Section 4.5 manifest outside of the student-advisor relationship, for example, through professors of courses, or faculty speaking at colloquia or town halls. Second, due to the prominence of advisors that hold the potentially harmful biases described above, the success of the student relationship is effectively a coin toss, increasingly weighted against the student as a function of their minoritization. What's more, the presence of students who are not advised by faculty in their own department (whether effectively or officially) is a prime example of how, in the

absence of standardized academic opportunities, the cumulative biases of individual faculty members can maintain a deeply inequitable program, without the blame falling on any one actor.

The Orange Band: Conflict Management

While the green 'Academic Structure' thread in Figure 4.2 outlines attitudes and behaviors that can contribute to the marginalization of racial and gender minorities, the orange 'Conflict Management' thread follows the ways in which this marginalization is handled– or in many cases, ignored. Interestingly, while the competition-based ethos present in Western physics for centuries was likely an absorption of gender roles which existed at the time, the incomparable value of objectivity was likely borne from the field itself. Sandra Harding posits that the physical sciences are in fact the origin of the "positivist, excessively empiricist philosophy" that pervades both physics and American culture more broadly (Harding, 1986). This Enlightenment ideal of the disembodied, transcendent scientist observing nature without bias or emotion has become the essential characteristic of the archetypal 'physicist', positioning it as the purest form of knowledge acquisition and sharing its aura of rationality with those who are perceived as similar to the scientists of yore [126]. It comes as no surprise then, that, in a country where the field remains dominated by the same demographic that produced such ideals (namely, white men), modern physicists extend their faith in the objectivity of their science to their understanding of the people who do science: "indeed, prevalent dogma holds that it is the virtue of modern science to make [critical reflection on social influences] unnecessary" [110].

The 'social influences' we have examined in this work are largely the manifestation of systems of oppression in educational settings, which, as illustrated in Section 4.5, students are encouraged not to think or talk about. By naming discussions of subjects like affinity-based support groups or bathroom access for gender-noncomforming folks as 'political' (along with the implication that such 'political' discussion is unprofessional and inappropriate for the workplace), faculty members create boundaries defining what is and isn't relevant to doing science, positioning the concerns of marginalized groups as irrelevant to their scientific bottom line. It follows that, if concerns of socially-influenced misconduct (e.g., microaggressions or discrimination) are irrelevant to doing science, creating robust structures for accountability and conflict resolution are similarly unimportant. The lack of effective codes of conduct in nearly half of our sample's departments (Section 4.5) is consistent with this thinking, and creates an environment where students see faculty members as autocratic actors, unfettered and unaccountable. What's more, as outlined in Section 4.5, this 'apolitical' environment is maintained by faculty ignoring critique and encouraging students to do the same, even in the presence of blatant harassment.

Most participants in this work found it difficult to buy into false images of inclusive departments, and often reported taking it upon themselves to try improving circumstances (for example, by creating peer-mentorship groups or serving on DEI committees). However, these efforts were often blocked on two ends; in Section 4.5, we described how participants

were discouraged from engaging in DEI oriented work both by faculty implying that there is no time for it, and instead their time should be spent on research, or by explicitly admonishing such efforts, like the faculty member who referred to applicants interested in DEI as "entitled". For participants who experienced discrimination or harassment in their departments, this created an impossible bind: a toxic work environment makes producing research difficult, the student attempts to improve their environment, but their advisor urges them to focus on research instead. While faculty members were observed to treat DEI work as an extra, possibly unnecessary part of the job, participants who endured misconduct due to their identities saw that work as a foundational piece of being able to do the job at all. This bind becomes especially worrying considering the near-ubiquity of such misconduct, with URM women being 15-20 times more likely then non-URM men to encounter discrimination or harassment in astronomy graduate programs [42].

These points paint a seemingly contradicting picture: though every participant's departments communicate in some form they are inclusive and value diversity in their students, all participants had experiences that were inconsistent with this stated value to some extent. Parallels to this paradox exist throughout the literature: Cochran and Boveda describe a similar tension between the desire to affirm diversity in physics spaces and the physicists doing said diversity work that are not "comfortable saying that Black Lives Matter" [117], while Thornhill found that white admissions counselors seek to increase diversity while favoring 'apolitical' Black students over 'political' ones [115]. Several participants accounts of their departments were strikingly similar to how Carbado and Gulati describe employers desiring aesthetic 'diversity' without attending to the complexities of a diverse population, identifying what they call "good Blacks": "'Good blacks' will think of themselves as people first and black people second (or third or fourth); they will neither 'play the race card' nor generate racial antagonism or tensions in the workplace; they will not let white people feel guilty about being white; and they will work hard to assimilate themselves" [127]. This practice of selecting assimilated 'diversity' (or expecting the existing 'diversity' to assimilate themselves) allows departments to satisfy the social and legal requirements of including minoritized people, without having to do the difficult work of creating an equitable environment, all the while accruing what Carbado and Gulati called the "diversity benefit".

The desire for the aesthetic of 'diversity' without meaningfully diversifying the culture of a department is also demonstrated by the ways in which participants observed faculty members to perform the motions of DEI work without meaningfully changing the conditions or treatment of minoritized groups, as described in Section 4.5. In her work analyzing the function of "diversity documents" in UK organizations, Sara Ahmed explains how documents stating commitments to diversity, anti-racism or similar values are often used as symbolic gestures of inclusion that get counted as 'having done' the work of creating inclusive work environments [128]. Her description of diversity work as being "about changing perceptions of whiteness rather than changing the whiteness of organizations" perfectly parallels most participants' observation of DEI work in their departments: performatively valuing and seeking diverse students, without meaningfully changing the cultural norms of the department, as would necessarily happen in a truly diverse population.

While Ahmed's analysis primarily focuses on institutional efforts (comparable to several participants accounts of their departments' #BlackLivesMatter statements), the performative dynamic that she outlines can be seen just as clearly in the behaviors of individual faculty members as well. At a time when 'diversity statements' or 'broader impacts' are a required component of many job applications and grant proposals in the field, there is strong incentive for faculty members to cultivate an image of supporting and contributing to DEI work. However, as our participants explain in Section 4.5, it is not unheard of for a faculty member who publicly presents an inclusive image (for example, via social media) while treating minoritized students poorly behind the scenes. Regardless of intention, this double-faced approach to DEI work has the potential to be damaging; as Ahmed said, "Self-perception of being good at communication and care blocks the recognition of racism" [128].

Predictably, participants who spoke of performative DEI work generally characterized that work as lacking an intersectional frame, that is, being unable to identify and treat the unique circumstances of multiply minoritized groups (see Section 4.5). The necessity of this type of intersectional analysis was initially discussed by Kimberle Crenshaw, who identified how being simultaneously black and female in certain workplaces allowed for a type of discrimination that didn't affect black men or non-black women [121]. A similar study in astronomy and planetary science observed Crenshaw's "double-jeopardy" to exist within these fields as well, finding that women of color are significantly more likely than white women, white men, or men of color to experience harassment or assault in the workplace on the basis of their gender or race [129]. Though one would think this warrants extra attention for the needs of multiply minoritized students, participants found that extra attention lacking, and felt alienated by groups and people meaning to support them along one axis of identity while being largely apathetic towards other components of their identity (a common example being organizations for women in the field that were largely composed of and catering to white women). Like the performative writing of statements, this indicates a 'checklist' attitude to doing DEI work: if white women and men of color are present and supported, then racism and sexism have been ostensibly defeated.

When Academics and Conflict Collide

All these factors work together to create an environment where the harassment and discrimination that have been demonstrated time and again to exist in the field go unreported and unprevented. For most participants of our study, this translated to enduring misconduct and alienation from advisors, other faculty members, and peers, while being devalued if they attempted to name the problem or improve their situation. What's more, departments can create a false sense of security and inclusivity (for example, to show to prospective students) by performatively engaging in DEI work. While this chain of events is outlined in the orange 'Conflict Management' band of Figure 4.2, we note here that it is very possible for these circumstances to overlap with the green 'Academic Structures' band.

We can understand this potential overlap by stepping a hypothetical example through our model: if a student is being discriminated against by their advisor (perhaps by holding

them to different expectations than their non-minoritized students, a fair possibility in the absence of standardized academic expectations), that student may be blocked from seeking recourse, or marked as 'political' if they attempt to do so. The advisor may be able to further obfuscate the nature of the relationship by performatively contributing to DEI efforts, effectively absolving them of the possibility of misconduct. If the student-advisor relationship ends, that student may then seem undesirable to other potential advisors, who might consider the student's 'failure' with their past advisor or 'confrontational' attempts to report their past advisor as reason enough to not take them on. If enough faculty members buy into this thinking, as Mabrooka's experience demonstrates, it is possible in some departments for a student to be left completely unadvised. In this way, the system outlined in Figure 4.2 not only allows for the marginalization of minoritized students, but compounds that marginalization if students resist it.

After demonstrating how the system in Figure 4.2 functions, we would like to identify how to interrupt it. While many efforts to improve department climate emphasize faculty training on topics like microaggressions or implicit bias, we believe that the best point of intervention at the department level is in the structure of programs. In both of the regimes we have discussed in this work, it is largely the lack of standardized and accountable structures that create the potential for individual faculty members' bias to have as much impact as it does. Rather than encourage their faculty to act in less equitable ways, we urge departments to intentionally and rigorously design their programs with equity as a foundational value, institutionalizing structures designed to address inequity rather than simply encouraging equitable behavior. The effectiveness of such an approach can be seen by the University of Michigan applied physics program, which was "designed for inclusion from the inside out": the result was a program that graduated about 10% of Black physics PhDs in the country for the previous decade, with 33% of the program's PhDs in 2011 awarded to URM students and 33% to women (as opposed to the national averages of 5% and 16%, respectively) [130]. In a case study of this program, the themes identified to explain these successes include a reconceptualization of the 'ideal student' and corresponding reformation of admissions, as well as an effort to create a "familylike climtate" by building department community. To this end, we offer suggested practices for departments in Section 4.7 that were compiled from our participants' recommendations.

Limitations

The primary limitation of this work is the relatively small sample size of 12 participants, limited both by funds available to compensate participants and the workload of conducting, transcribing, and coding a large collection of interviews. However, we argue that it is not unreasonable to extrapolate the more frequently occurring findings of this work more broadly, considering how small the population we are trying to represent is. From 2009-2019, the total number of physics and astronomy PhDs earned by Black/African American, Latinx/Hispanic, and American Indian/Alaskan Native women ranged from about 10 to 20 per year, suggesting less than 100 total URM women enrolled in such programs at a given

time (NSF, 2019). Thus, our 8 URM participants constitute nearly a tenth of the larger population, suggesting that our sample is highly representative despite the small size.

Due to this dearth of racially minoritized women (as well as gender nonconforming folks, who are not accounted for in the NSF statistics), many participants indicated concern for anonymity, feeling that basic demographic information could render them identifiable in the field. This concern is emphasized by the sensitive subject nature of the interviews, which, if identified, could create problems for participants in their institutions. As such, we recognize the possibility of our findings being limited by what participants felt safe disclosing. However, we assume that such unsafe information would be 'worse' than what participants did choose to share, making our findings something of a lower-limit description of negative graduate school outcomes.

Additionally, significant efforts were made to protect participants' identities and ensure them of this protection (for example, only describing demographics of the sample as whole, rather than sharing the various identities of each participant). Further, the interviews and all communication with participants was done by the first author, a current graduate student in astronomy who is a woman of color. This positionality was leveraged to generate a comfortable rapport with participants, which can be difficult when soliciting such sensitive information from a stranger. Several participants mentioned afterwards that this helped them to feel at ease during the interview, hopefully supporting the veracity of our findings.

The positionality of the first author was also an influence in the production of this work. A research collaboratory served as a check on the rigor of the first author's data collection and analysis, allowing us to take advantage of the benefits that this positionality allows for. Besides the peer-like interview rapport described above, the first author's lived experiences function as a foundational counter-story in this work, as it was the source of inspiration for the research question and helped inform how data could best be collected.

Finally, we recognize that the organization of academic departments is partially dictated by policies and norms of the larger university. While we attempted to contextualize our results with this fact, we had insufficient information to fully disentangle it's impact from that of the specific participant department. Participants brought up various university-level practices, including academic tenure and minimal incentive for DEI work relative to research, and while we did not discuss them in depth in this work, we note that the influence of such factors on inequity at any stage of higher education may be a worthwhile area of study.

4.7 Conclusions

This work used the lived experiences of 12 female and gender-non conforming students of color to outline how astronomy and physics graduate programs in American PWIs maintain inequity between majoritized and minoritized students through the interplay of cultural beliefs, program structures, and faculty practices. We present a model of this system as exemplified by this sample, constructed by identifying attitudes, policies, and behaviors that produced negative outcomes for participants and related to two cultural beliefs of Western

physics: 1) that the field is a meritocracy where the most brilliant survive, and 2) that physicists comprise an objective, cultureless, and apolitical community. These findings are supported by the literature that has documented the negative outcomes such systems create for minoritized students, including alienation from their departments, mental health problems, and lower retention in the field. In outlining this system, we emphasize the design of program structures as a critical point of intervention, as they can serve to limit faculty members from acting on biases, and can largely be determined within a department (though some requirements are generally set by the larger university). We refer to a case study of a department "designed for inclusion from the inside out" as an example of the success of such methods at achieving equity goals.

Recommendations

We have compiled a list of recommendations based on participant input for departments to consider in their efforts to attain equitable graduate programs. As the final interview question, participants were asked what they would do to achieve this goal if they were "chair for a day". The following list presents recurring answers in aggregate, from most frequently suggested to least.

- **Transparency and communication:** Explicitly and officially defining the details of academic programs could go a long way in remedying the problematic structures described in this work. Graduate handbooks, for example, can be written up to clearly communicate to students what they can expect from their department, advisor, and professors, as well as what they are expected to do in return. This can also be done on an individual level by faculty members by creating 'lab handbooks', defining student and advisor expectations in a group. Guides like these help by putting "into writing things that are usually transmitted informally" [131]. Wherever standards and clear policies can be set, they should be easily accessible to current and prospective students, for example via department website, with changes communicated as they are made.

- **Orienting hiring and admissions towards building a 'critical mass':** An effective way to change an organization's culture is to change its composition, which department's have the ability to do through the admission of new students and hiring of faculty and administrators. Utilizing equitable practices in recruitment and assessment of candidates can drastically change the demographics of accepted additions, without the use of affirmative action or quotas [46].

- **Trainings:** Though the effectiveness of implicit bias training is not certain [132], participants generally believed faculty practices could be improved through trainings or workshops. Rather than only addressing implicit bias, trainings might focus on mentoring, leadership, or 'soft skills', such as communication and interpersonal skills, especially through a frame of equity.

- **Robust codes of conduct:** While some departments may already have a code of conduct, work can likely be done to improve the communication of acceptable conduct as well as the defined procedures in the event of misconduct for all department members. Importantly, the code of conduct must be shown to be enforceable for it to have a meaningfully positive impact on department climate [133]. Additionally, thought should be put into ensuring safe channels for reporting misconduct in order to prevent intimidation or retaliation, and the safety and existence of these channels should be clearly communicated.

- **Community building:** In the aforementioned University of Michigan program which was redesigned "for inclusion from the inside out", one of the central elements they attribute their success to is the family-like community developed among department members [130]. Many participants in this study felt they could have benefited from similarly warm environments, recommending departments invest in community building for students, faculty, and staff.

- **Decentralized advising:** As many problems occurred for participants related to their dependence on a single advisor, we encourage exploring different models for training graduate students. Research on holistic advising, where students are advised by a full team of faculty, have shown positive results on student retention [134]. Other models to consider include advisor and project rotations, or advising committees. Additionally, procedures for changing advisors and ensuring all students are adequately advised can alleviate the pressure of depending on a single advisor.

- **Mentoring structures:** Mentoring, both between peers and across professional levels, have the potential to provide students with academic and social/emotional support in graduate programs. Building structures to ensure that all students have at least one mentor they can reliably and comfortably talk to can go far in improving student outcomes.

- **Recognizing and rewarding service work:** In whatever ways departments recognize or reward work in areas of research, there should be comparable recognition and reward for teaching, service, and DEI work. Recognition has been demonstrated to be a critical component of developing a science identity [135], so recognition provides the double function of validating minoritized department members who are more likely to take on this work, and incentives others to value and contribute to this work.

- **Service requirements:** Similar to teaching requirements for both students and faculty, departments can implement service requirements, which may include outreach, department community building, and DEI work. By standardizing the expectation, service works becomes an institutionally valued activity, and spreads the burden of such work between all members.

- **Utilizing research based pedagogy:** Some participants expressed frustration that their departments continued practices that have been rigorously demonstrated to be biased, such as the use of the Physics GRE in graduate admissions [96]. The literature of physics education research (PER) can be a source of invaluable information for the development of more equitable programs, such as best admissions/hiring practices, effective mentoring, inclusive teaching, and more. This could be implemented, for example, with a PER journal club.

Future Research

There were several interesting themes in our data set that we did not explore in this work that could be the basis of future research. One theme was that of inter-grad dynamics, specifically, the ways in which students marginalized other students: some participants described harassment and ostracization from other grads, and our data suggest that this behavior may be incentivized by the tone and values set by faculty members. Another theme was a bias against religiousness (particularly, non-christian religiousness), which some religious-minority identifying participants reported as being alienating and discouraging. Finally, while we discussed the emergence of performative DEI efforts throughout this work, we have only introduced the idea; much more work is needed to identify and understand the potential harm of performative DEI work, especially as it becomes an increasingly necessary part of job applications and grant proposals. Some participants offered interesting thoughts on the question of differentiating between genuine and performative DEI work, a distinction that could be invaluable in the evaluation of potential faculty members.

4.8 Additional Figures: Complete TRE Results

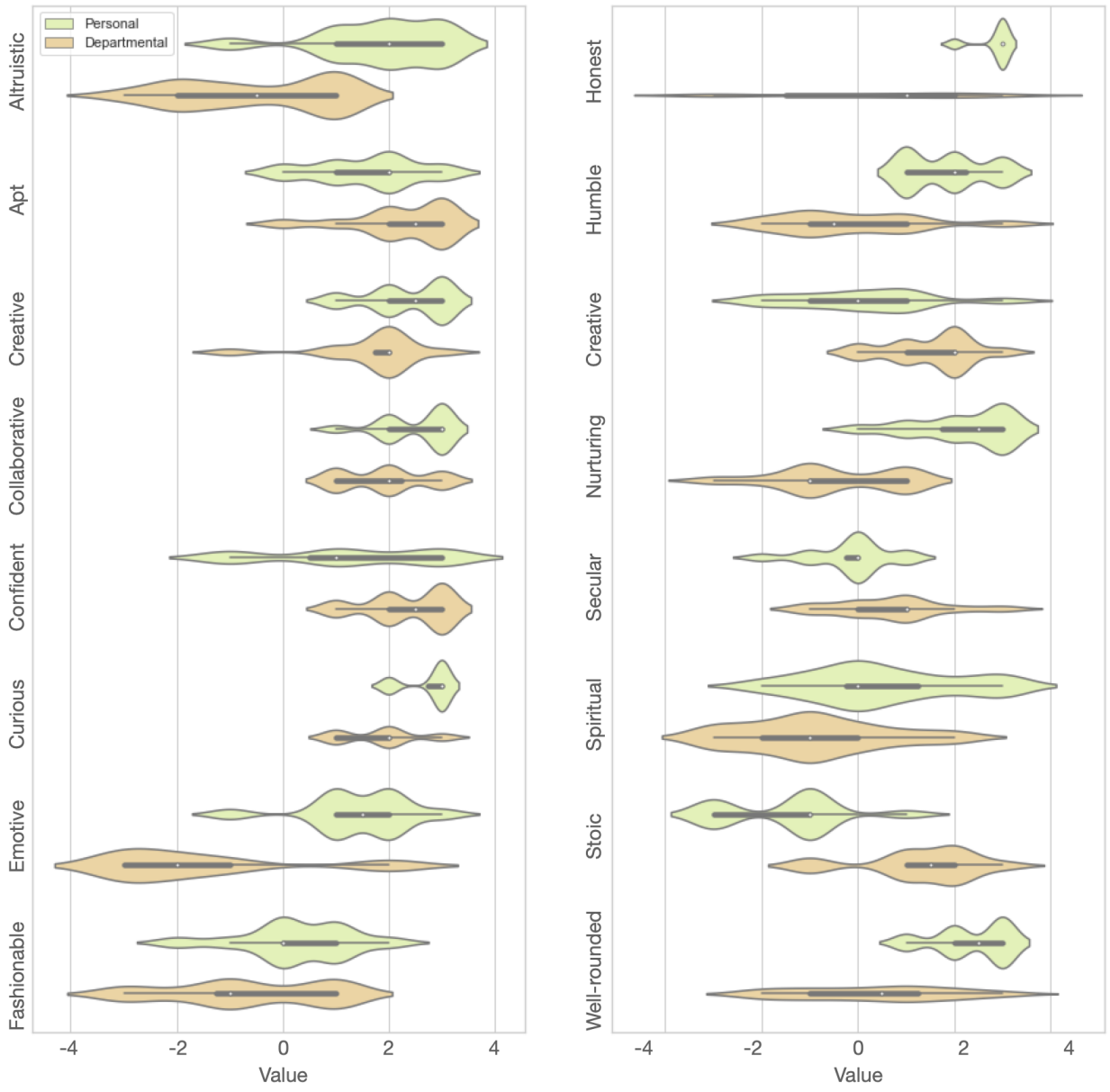


Figure 4.3: This figure is an extension of Figure 4.1 which presents the remaining TRE data for the 16 traits not included in the main text.

Bibliography

[1] Jorge Casares. "Observational evidence for stellar-mass black holes". In: *Proceedings of the International Astronomical Union* 2.S238 (2006), pp. 3–12.

[2] National Center for Science and National Science Foundation Engineering Statistics. *Doctorate Recipients from U.S. Universities: 2019*. Alexandria, VA, 2019.

[3] Terrance J Quinn. *The (pre-) dawning of Functional Specialization in Physics*. World Scientific, 2017.

[4] Albert Einstein. "Lens-like action of a star by the deviation of light in the gravitational field". In: *Science* 84.2188 (1936), pp. 506–507.

[5] E. Agol et al. "Finding Black Holes with Microlensing". In: 576 (Sept. 2002), pp. L131–L135. DOI: 10.1086/343758. eprint: astro-ph/0203257.

[6] J. Casares. "Observational evidence for stellar-mass black holes". In: *Black Holes from Stars to Galaxies – Across the Range of Masses*. Ed. by V. Karas and G. Matt. Vol. 238. IAU Symposium. Apr. 2007, pp. 3–12. DOI: 10.1017/S1743921307004590. eprint: astro-ph/0612312.

[7] J. Casares and P. G. Jonker. "Mass Measurements of Stellar and Intermediate-Mass Black Holes". In: 183 (Sept. 2014), pp. 223–252. DOI: 10.1007/s11214-013-0030-6. arXiv: 1311.5118 [astro-ph.HE].

[8] M. T. Reynolds and J. M. Miller. "A Swift Survey of Accretion onto Stellar-mass Black Holes". In: 769, 16 (May 2013), p. 16. DOI: 10.1088/0004-637X/769/1/16. arXiv: 1112.2249 [astro-ph.HE].

[9] B. P. et al. Abbott. "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Phys. Rev. Lett.* 116 (6 Feb. 2016), p. 061102. DOI: 10.1103/PhysRevLett.116.061102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.061102.

[10] Grzegorz Wiktorowicz et al. "Populations of Stellar-mass Black Holes from Binary Systems". In: *The Astrophysical Journal* 885.1 (2019), p. 1.

[11] J. R. Lu et al. "A Search For Stellar-mass Black Holes Via Astrometric Microlensing". In: 830, 41 (Oct. 2016), p. 41. DOI: 10.3847/0004-637X/830/1/41. arXiv: 1607.08284 [astro-ph.SR].

[12] B. Paczynski. "Gravitational microlensing at large optical depth". In: 301 (Feb. 1986), pp. 503–516. DOI: 10.1086/163919.

[13] B. Paczynski and J. Wambsganss. "Gravitational microlensing." In: *Physics World* 6 (May 1993), pp. 26–30.

[14] Shude Mao. "Astrophysical applications of gravitational microlensing". In: *Research in Astronomy and Astrophysics* 12.8 (2012), p. 947. URL: http://stacks.iop.org/1674-4527/12/i=8/a=005.

[15] A Udalski et al. "The optical gravitational lensing experiment. Discovery of the first candidate microlensing event in the direction of the Galactic Bulge". In: *Acta astronomica* 43 (1993), pp. 289–294.

[16] A Udalski, MK Szymański, and G Szymański. "OGLE-IV: fourth phase of the Optical Gravitational Lensing Experiment". In: *arXiv preprint arXiv:1504.05966* (2015).

[17] IA Bond et al. "Real-time difference imaging analysis of MOA Galactic bulge observations during 2000". In: *Monthly Notices of the Royal Astronomical Society* 327.3 (2001), pp. 868–880.

[18] T Sako et al. "MOA-cam3: a wide-field mosaic CCD camera for a gravitational microlensing survey in New Zealand". In: *Experimental Astronomy* 22.1-2 (2008), pp. 51–66.

[19] Seung-Lee Kim et al. "KMTNET: a network of 1.6 m wide-field optical telescopes installed at three southern observatories". In: *Journal of Korean Astronomical Society* 49 (2016), pp. 37–44.

[20] C. Alcock et al. "Possible gravitational microlensing of a star in the Large Magellanic Cloud". In: 365 (Oct. 1993), pp. 621–623. DOI: 10.1038/365621a0.

[21] Ł Wyrzykowski et al. "Black hole, neutron star and white dwarf candidates from microlensing with OGLE-III". In: *Monthly Notices of the Royal Astronomical Society* 458.3 (2016), pp. 3012–3026.

[22] Shude Mao et al. "Optical Gravitational Lensing Experiment OGLE 1999 BUL 32: the longest ever microlensing event–evidence for a stellar mass black hole?" In: *Monthly Notices of the Royal Astronomical Society* 329.2 (2002), pp. 349–354.

[23] François Rigaut et al. "Gemini multiconjugate adaptive optics system review – I. Design, trade-offs and integration". In: *Monthly Notices of the Royal Astronomical Society* 437.3 (2014), pp. 2361–2375. DOI: 10.1093/mnras/stt2054. eprint: /oup/backfile/content_public/journal/mnras/437/3/10.1093_mnras_stt2054/1/stt2054.pdf. URL: +%20http://dx.doi.org/10.1093/mnras/stt2054.

[24] Enrico Marchetti et al. "On-sky testing of the multi-conjugate adaptive optics demonstrator". In: *The Messenger* 129.8 (2007).

[25] David R Andersen et al. "Status of the Raven MOAO science demonstrator". In: *Proc. SPIE*. Vol. 8447. 2012, 84473F.

[26] E Gendron et al. "MOAO first on-sky demonstration with CANARY". In: *Astronomy & Astrophysics* 529 (2011), p. L2.

[27] Ray M Sharples, Richard M Myers, and Nicholas A Walton. "MOSAIC: a multi-object spectrograph with adaptive image correction". In: *Optical and IR Telescope Instrumentation and Detectors*. Vol. 4008. International Society for Optics and Photonics. 2000, pp. 228–237.

[28] F. Rigaut. "Ground Conjugate Wide Field Adaptive Optics for the ELTs". In: *European Southern Observatory Conference and Workshop Proceedings*. Ed. by E. Vernet et al. Vol. 58. European Southern Observatory Conference and Workshop Proceedings. 2002, p. 11.

[29] C. Baranec et al. "On-Sky Wide-Field Adaptive Optics Correction Using Multiple Laser Guide Stars at the MMT". In: 693 (Mar. 2009), pp. 1814–1820. DOI: 10.1088/0004-637X/693/2/1814. arXiv: 0812.0352.

[30] M. Hart et al. "A ground-layer adaptive optics system with multiple laser guide stars". In: 466 (Aug. 2010), pp. 727–729. DOI: 10.1038/nature09311.

[31] G. Orban de Xivry et al. "First on-sky results with ARGOS at LBT". In: *Adaptive Optics Systems V*. Vol. 9909. July 2016, p. 990936. DOI: 10.1117/12.2240094.

[32] R. Arsenault et al. "The Adaptive Optics Facility: Commissioning Progress and Results". In: *The Messenger* 168 (June 2017), pp. 8–14. DOI: 10.18727/0722-6691/5019.

[33] A. Tokovinin et al. "SOAR Adaptive Module (SAM): Seeing Improvement with a UV Laser". In: 128.12 (Dec. 2016), p. 125003. DOI: 10.1088/1538-3873/128/970/125003. arXiv: 1608.05593 [astro-ph.IM].

[34] Wolfgang Gässler et al. "Status of the ARGOS ground layer adaptive optics system". In: *Adaptive Optics Systems III*. Vol. 8447. International Society for Optics and Photonics. 2012, p. 844702.

[35] Harald Kuntschner et al. "Operational concept of the VLT's adaptive optics facility and its instruments". In: *Proc. SPIE*. Vol. 8448. 2012, p. 07.

[36] Yutaka Hayano et al. "ULTIMATE-SUBARU: project status". In: *SPIE Astronomical Telescopes+ Instrumentation*. International Society for Optics and Photonics. 2014, 91482S–91482S.

[37] Lawrence H Summers. "Remarks at NBER conference on diversifying the science & engineering workforce". In: *The Office of the President, Harvard University* (2005).

[38] Mitchell L Stevens. *Creating a class*. Harvard University Press, 2020.

[39] William G Bowen and Neil L Rudenstine. *In pursuit of the PhD*. Vol. 75. Princeton University Press, 2014.

[40] Teresa M Evans et al. "Evidence for a mental health crisis in graduate education". In: *Nature biotechnology* 36.3 (2018), pp. 282–284.

[41] Julie Posselt. "Discrimination, competitiveness, and support in US graduate student mental health". In: *Studies in Graduate and Postdoctoral Education* (2021).

[42] Rachel Ivie. "Results from the Longitudinal Study of Astronomy Graduate Students". In: *American Astronomical Society Meeting Abstracts# 223*. Vol. 223. 2014, pp. 304–03.

[43] Tom Arrison, Steve Olson, et al. *Rising above the gathering storm: Developing regional innovation environments: A workshop summary*. National Academies Press, 2012.

[44] Expanding Underrepresented Minority Participation. "America's Science and Technology Talent at the Crossroads". In: *National Academy of Sciences, National Academy of Engineering and Institute of Medicine* (2011).

[45] Engineering National Academies of Sciences, Medicine, et al. *Graduate STEM education for the 21st century*. National Academies Press, 2018.

[46] Julie R Posselt. *Equity in Science: Representation, Culture, and the Dynamics of Change in Graduate Education*. Stanford University Press, 2020.

[47] Edward Tylor. *Primitive Culture [1871], New York, JP Putnam's Sons*. 1920.

[48] D. P. Bennett et al. "Gravitational Microlensing Events Due to Stellar-Mass Black Holes". In: 579 (Nov. 2002), pp. 639–659. DOI: 10.1086/342225. eprint: astro-ph/0109467.

[49] A. A. Nucita et al. "An XMM-Newton Search for X-Ray Emission from the Microlensing Event MACHO-96-BLG-5". In: 651 (Nov. 2006), pp. 1092–1097. DOI: 10.1086/507784.

[50] M. C. Smith et al. "The nature of parallax microlensing events towards the Galactic bulge". In: 361 (July 2005), pp. 128–140. DOI: 10.1111/j.1365-2966.2005.09147.x. eprint: astro-ph/0504536.

[51] S. Poindexter et al. "Systematic Analysis of 22 Microlensing Parallax Candidates". In: 633 (Nov. 2005), pp. 914–930. DOI: 10.1086/468182. eprint: astro-ph/0506183.

[52] Andrew Gould and Jennifer C. Yee. "Microlens Masses from Astrometry and Parallax in Space-based Surveys: From Planets to Black Holes". In: *The Astrophysical Journal* 784.1 (2014), p. 64. URL: http://stacks.iop.org/0004-637X/784/i=1/a=64.

[53] M. Soto, D. Minniti, and M. Rejkuba. "Spectroscopic and light curve characterization of bulge microlensing events". In: 466 (Apr. 2007), pp. 157–164. DOI: 10.1051/0004-6361:20066211.

[54] Peter L. Wizinowich et al. "The W. M. Keck Observatory Laser Guide Star Adaptive Optics System: Overview". In: *Publications of the Astronomical Society of the Pacific* 118 (Feb. 2006), pp. 297–309. DOI: 10.1086/499290.

[55] M. Service et al. "A New Distortion Solution for NIRC2 on the Keck II Telescope". In: *Publications of the Astronomical Society of the Pacific* 128 (Sept. 2016), p. 095004. DOI: 10.1088/1538-3873/128/967/095004.

[56] Charles Alcock et al. "The MACHO project: microlensing results from 5.7 years of Large Magellanic Cloud observations". In: *The Astrophysical Journal* 542.1 (2000), p. 281.

[57] Anthony GA Brown et al. "Gaia Early Data Release 3: Summary of the contents and survey properties". In: *arXiv preprint arXiv:2012.01533* (2020).

[58] L Lindegren et al. "Gaia Early Data Release 3: The astrometric solution". In: *arXiv preprint arXiv:2012.03380* (2020).

[59] Marco Riello et al. "Gaia Early Data Release 3: Photometric content and validation". In: *arXiv preprint arXiv:2012.01916* (2020).

[60] Claus Fabricius et al. "Gaia Early Data Release 3–Catalogue validation". In: *arXiv preprint arXiv:2012.06242* (2020).

[61] S. Gonzaga and J. Biretta. *HST WFPC2 Data Handbook v5.0*. 2010.

[62] A. Stolte et al. "The Proper Motion of the Arches Cluster with Keck Laser-Guide Star Adaptive Optics". In: 675, 1278-1292 (Mar. 2008), pp. 1278–1292. DOI: 10.1086/527027. arXiv: 0706.4133.

[63] J. R. Lu et al. "A Disk of Young Stars at the Galactic Center as Determined by Individual Stellar Orbits". In: 690 (Jan. 2009), pp. 1463–1487. DOI: 10.1088/0004-637X/690/2/1463. arXiv: 0808.3818.

[64] A. S. Fruchter and R. N. Hook. "Drizzle: A Method for the Linear Reconstruction of Undersampled Images". In: 114 (Feb. 2002), pp. 144–152. DOI: 10.1086/338393. eprint: astro-ph/9808087.

[65] S. Yelda et al. "Improving Galactic Center Astrometry by Reducing the Effects of Geometric Distortion". In: 725, 331-352 (Dec. 2010), pp. 331–352. DOI: 10.1088/0004-637X/725/1/331. arXiv: 1010.0064.

[66] Gunther Witzel et al. "The AIROPA software package: milestones for testing general relativity in the strong gravity regime with AO". In: *Adaptive Optics Systems V*. Vol. 9909. International Society for Optics and Photonics. 2016, 99091O.

[67] E. Diolaiti et al. "StarFinder: a code to analyse isoplanatic high-resolution stellar fields." In: *The Messenger* 100 (June 2000), pp. 23–27.

[68] D. Minniti, P. Lucas, and VVV Team. "VizieR Online Data Catalog: VISTA Variable in the Via Lactea Survey DR2 (Minniti+, 2017)". In: *VizieR Online Data Catalog* 2348 (Apr. 2017).

[69] Jay Anderson and Ivan R King. "PSFs, Photometry, and Astrometry for the ACS/WFC". In: *ACS Instrument Science Report* 1 (2006).

[70] Jon A Holtzman et al. "The photometric performance and calibration of WFPC2". In: *Publications of the Astronomical Society of the Pacific* 107.717 (1995), p. 1065.

[71] G. M. Green et al. "A Three-dimensional Map of Milky Way Dust". In: 810, 25 (Sept. 2015), p. 25. DOI: 10.1088/0004-637X/810/1/25. arXiv: 1507.01005.

[72] G. M. Green et al. "Galactic Reddening in 3D from Stellar Photometry - An Improved Map". In: *ArXiv e-prints* (Jan. 2018). arXiv: 1801.03555.

[73] E. F. Schlafly and D. P. Finkbeiner. "Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD". In: 737, 103 (Aug. 2011), p. 103. DOI: 10.1088/0004-637X/737/2/103. arXiv: 1012.4804 [astro-ph.GA].

[74] E Bachelet et al. "pyLIMA: An Open-source Package for Microlensing Modeling. I. Presentation of the Software and Analysis of Single-lens Models". In: *The Astronomical Journal* 154.5 (2017), p. 203.

[75] Luciano Rezzolla, Elias R Most, and Lukas R Weih. "Using gravitational-wave observations and quasi-universal relations to constrain the maximum mass of neutron stars". In: *The Astrophysical Journal Letters* 852.2 (2018), p. L25.

[76] Casey Y Lam et al. "PopSyCLE: A New Population Synthesis Code for Compact Object Microlensing Events". In: *The Astrophysical Journal* 889.1 (2020), p. 31.

[77] Yoshitomo Maeda et al. "A search for X-rays from the long-duration microlensing event MACHO-96-BLG-5". In: *The Astrophysical Journal Letters* 631.1 (2005), p. L65.

[78] M. R. Chun et al. "Imaka: a ground-layer adaptive optics system on Maunakea". In: *Adaptive Optics Systems V*. Vol. 9909. Proc. SPIE. July 2016, p. 990902. DOI: 10.1117/12.2233208. arXiv: 1608.01804 [astro-ph.IM].

[79] P Wizinowich et al. "First light adaptive optics images from the Keck II telescope: a new era of high angular resolution imagery". In: *Publications of the Astronomical Society of the Pacific* 112.769 (2000), p. 315.

[80] C. Baranec, M. Lloyd-Hart, and M. Meyer. "Loki: a ground-layer adaptive optics high-resolution near-infrared survey camera". In: *Astronomical Adaptive Optics Systems and Applications III*. Vol. 6691. Sept. 2007, p. 669106. DOI: 10.1117/12.733144.

[81] MF Skrutskie et al. "The two micron all sky survey (2MASS)". In: *The Astronomical Journal* 131.2 (2006), p. 1163.

[82] RW Wilson. "SLODAR: measuring optical turbulence altitude with a Shack–Hartmann wavefront sensor". In: *Monthly Notices of the Royal Astronomical Society* 337.1 (2002), pp. 103–108.

[83] O Lai et al. "Deconstructing Turbulence and optimizing GLAO using imaka telemetry". In: *Proc. SPIE*. Vol. 10703. 2018, pp. 10703–229.

[84] Andrei Tokovinin. "From differential image motion to seeing". In: *Publications of the Astronomical Society of the Pacific* 114.800 (2002), p. 1156.

[85] G Lombardi and M Sarazin. "Using MASS for AO simulations: a note on the comparison between MASS and Generalized SCIDAR techniques". In: *Monthly Notices of the Royal Astronomical Society* 455.3 (2015), pp. 2377–2386.

[86] Matwey V. Kornilov. *atmos 3.0*. 2017. URL: `http://curl.sai.msu.ru/~matwey/atmos/`.

[87] P. B. Stetson. "DAOPHOT - A computer program for crowded-field stellar photometry". In: 99 (Mar. 1987), pp. 191–222. DOI: `10.1086/131977`.

[88] Ivan R King. "ACCURACY OF MEASUREMENT OF STAR IMAGES ON A PIXEL ARRAY." In: *Publications of the Astronomical Society of the Pacific* 95.564 (1983), p. 163.

[89] R Saglia et al. "THE EFFECTS OF SEEING ON THE PHOTOMETRIC PROPERTIES OF ELLIPTIC GALAXIES". In: *MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY* 264.4 (1993), pp. 961–974.

[90] M. R. Chun et al. "'imaka: a path-finder ground-layer adaptive optics system for the University of Hawaii 2.2-meter telescope on Maunakea". In: *Adaptive Optics Systems IV*. Vol. 9148. Aug. 2014, 91481K. DOI: `10.1117/12.2056917`.

[91] David R Andersen et al. "Performance Modeling of a Wide-Field Ground-Layer Adaptive Optics System". In: *Publications of the Astronomical Society of the Pacific* 118.849 (2006), p. 1574.

[92] Sandra M Faber et al. "The DEIMOS spectrograph for the Keck II Telescope: integration and testing". In: *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. Vol. 4841. International Society for Optics and Photonics. 2003, pp. 1657–1670.

[93] C Rockosi et al. "The low-resolution imaging spectrograph red channel CCD upgrade: fully depleted, high-resistivity CCDs for Keck". In: *Ground-based and Airborne Instrumentation for Astronomy III*. Vol. 7735. International Society for Optics and Photonics. 2010, 77350R.

[94] Robert Beichner. "An Introduction to Physics Education Research". In: *Getting Started in PER*. Vol. 2. Apr. 2009.

[95] Estela Mara Bensimon. "Closing the achievement gap in higher education: An organizational learning perspective". In: *New directions for higher education* 2005.131 (2005), pp. 99–111.

[96] Casey W Miller et al. "Typical physics Ph. D. admissions criteria limit access to underrepresented groups but fail to predict doctoral completion". In: *Science Advances* 5.1 (2019), eaat7550.

[97] Elizabeth A Canning et al. "STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes". In: *Science advances* 5.2 (2019), eaau4734.

[98] Kerrie G Wilkins-Yel et al. "Understanding the impact of personal challenges and advisor support on stem persistence among graduate women of color." In: *Journal of Diversity in Higher Education* (2021).

[99] Kris D Gutiérrez and Barbara Rogoff. "Cultural ways of learning: Individual traits or repertoires of practice". In: *Educational researcher* 32.5 (2003), pp. 19–25.

[100] Ernieda Hatah et al. "The influence of cultural and religious orientations on social support and its potential impact on medication adherence". In: *Patient preference and adherence* 9 (2015), p. 589.

[101] Sharon Traweek. *Beamtimes and lifetimes*. Harvard University Press, 2009.

[102] Richard C Lewontin. "Sociobiology: Another biological determinism". In: *International Journal of Health Services* 10.3 (1980), pp. 347–363.

[103] Paula Gould. "Women and the culture of university physics in late nineteenth-century Cambridge". In: *The British Journal for the History of Science* 30.2 (1997), pp. 127–149.

[104] Amy Liu. "Unraveling the myth of meritocracy within the context of US higher education". In: *Higher education* 62.4 (2011), pp. 383–397.

[105] Emilio J Castilla and Stephen Benard. "The paradox of meritocracy in organizations". In: *Administrative Science Quarterly* 55.4 (2010), pp. 543–676.

[106] Rachel E Scherr et al. "Fixed and growth mindsets in physics graduate admissions". In: *Physical Review Physics Education Research* 13.2 (2017), p. 020133.

[107] Jan Nespor. *Knowledge in motion: Space, time and curriculum in undergraduate physics and management*. Routledge, 2014.

[108] Angela C Johnson. "Unintended consequences: How science professors discourage women of color". In: *Science Education* 91.5 (2007), pp. 805–821.

[109] Katherine P Dabney and Robert H Tai. "Female physicist doctoral experiences". In: *Physical Review Special Topics-Physics Education Research* 9.1 (2013), p. 010115.

[110] Sandra G Harding. *The science question in feminism*. Cornell University Press, 1986.

[111] Simone Hyater-Adams et al. "Deconstructing Black physics identity: Linking individual and social constructs using the critical physics identity framework". In: *Physical Review Physics Education Research* 15.2 (2019), p. 020115.

[112] Ibram X Kendi. *How to be an antiracist*. One world, 2019.

[113] Maria Ong. "Body projects of young women of color in physics: Intersections of gender, race, and science". In: *Social problems* 52.4 (2005), pp. 593–617.

[114] Allison Harell, Stuart Soroka, and Shanto Iyengar. "Race, prejudice and attitudes toward redistribution: A comparative experimental approach". In: *European Journal of Political Research* 55.4 (2016), pp. 723–744.

[115] Ted Thornhill. "We want black students, just not you: How white admissions counselors screen black prospective students". In: *Sociology of Race and Ethnicity* 5.4 (2019), pp. 456–470.

[116] Harold D Lasswell. *Politics: Who gets what, when, how*. Pickle Partners Publishing, 2018.

[117] Geraldine L Cochran and Mildred Boveda. "A framework for improving diversity work in physics". In: ().

[118] Richard T Schaefer. *Encyclopedia of race, ethnicity, and society*. Vol. 1. Sage, 2008.

[119] Jean Stefancic and Richard Delgado. *Critical race theory: The cutting edge*. Temple University Press Harvard, 2000.

[120] L Hunn, Talmadge C Guy, and Elaine Manglitz. "Who can speak for whom? Using counterstorytelling to challenge racial hegemony". In: *annual conference of the Adult Education Research Conference, University of Minnesota, Minneapolis*. 2006.

[121] Kimberlé Crenshaw. "Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics". In: *u. Chi. Legal f.* (1989), p. 139.

[122] Asia A Eaton et al. "How gender and race stereotypes impact the advancement of scholars in STEM: Professors' biased evaluations of physics and biology post-doctoral candidates". In: *Sex Roles* 82.3 (2020), pp. 127–141.

[123] Sue V Rosser and Eliesh O'Neil Lane. "Key barriers for academic institutions seeking to retain female scientists and engineers: Family-unfriendly policies. Low Numbers, stereotypes, and harassment". In: *Journal of Women and Minorities in Science and Engineering* 8.2 (2002).

[124] Elaine Seymour and Nancy M Hewitt. *Talking about leaving*. Westview Press, Boulder, CO, 1997.

[125] Burton R Clark. "The" Cooling-Out" Function in Higher Education." In: (1963).

[126] Donna Haraway. "Situated knowledges: The science question in feminism and the privilege of partial perspective". In: *Feminist studies* 14.3 (1988), pp. 575–599.

[127] Devon W Carbado and Mitu Gulati. *Acting white?: Rethinking race in post-racial America*. Oxford University Press, 2013.

[128] S. Ahmed. In: *Ethnic and Racial Studies* 30.4 (2007), pp. 590–609.

[129] K. Clancy. In: *Journal of Geophysical Research: Planets* 122.7 (2017), pp. 1610–1623.

[130] Julie Posselt et al. "Equity efforts as boundary work: How symbolic and social boundaries shape access and inclusion in graduate education". In: *Teachers College Record* 119.10 (2017), pp. 1–38.

[131] Mariam Aly. "The key to a happy lab life is in the manual". In: *Nature* 561.7721 (2018), pp. 7–8.

[132] Doyin Atewologun, Tinu Cornish, and Fatima Tresh. "Unconscious bias training: An assessment of the evidence for effectiveness". In: *Equality and Human Rights Commission Research Report Series* (2018).

[133] Zabihollah Rezaee, Robert C Elmore, and Joseph Z Szendi. "Ethical behavior in higher educational institutions: The role of the code of conduct". In: *Journal of business ethics* 30.2 (2001), pp. 171–183.

[134] Barnard A Polnarieve et al. "Nourishing STEM Student Success via a TEAM-Based Advisement Model." In: *International Journal of Higher Education* 6.6 (2017), pp. 31–43.

[135] Heidi B Carlone and Angela Johnson. "Understanding the science experiences of successful women of color: Science identity as an analytic lens". In: *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 44.8 (2007), pp. 1187–1218.