THE SAGITTARIUS TIDAL STREAM AND THE SHAPE OF THE GALACTIC STELLAR HALO

By

Matthew T. Newby

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: PHYSICS

Approved by the
Examining Committee:

Heidi Jo Newberg, Thesis Adviser

Wayne Roberge, Member

Jim Napolitano, Member

Carlos Varela, Member

Rensselaer Polytechnic Institute
Troy, New York

July 2013
(For Graduation August 2013)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

It can be said that all of human achievement originates from two sources: trial and error, and building on previous work. Very little can be achieved through purely original thinking and individual effort, and especially in the modern era, most accomplishments are the product of a team effort. The work leading to this PhD thesis is no exception. While a great number of people have influenced my life for the better during my time in graduate school (and, thankfully, very few have made a negative impact), I am unable to thank them all here; however, I have made an effort to include here the significant contributors, of which there are still many.

First and foremost, I would like to thank my advisor, Heidi Jo Newberg. She has been a better advisor than most graduate students can ever hope to have, being kind and understanding, while at the same time not afraid to "tell it like it is." Heidi is a model scientist, and it has been an honor working with her.

I would then like to thank my committee: Professors Wayne Roberge, Jim Napolitano, and Carlos Varela. Wayne is an amazing instructor in the classroom, and his high standards for academic and professional achievement have my deepest appreciation. Jim has always been friendly and genuinely curious, and has always been great at making me think outside of my field; I only regret not having the opportunity to take one his classes. Carlos has been a major contributor to the MilkyWay@home project, and has on many occasions provided valuable advice on the computational side of the project.

Several astronomers have helped shape and refine my knowledge of physics and astronomy, many of whom have served as valuable collaborators, and to whom I owe a great quantity of thanks: Dr. Nathan Cole and Dr. Benjamin Willett ("those who came before"); Post-docs David "Bernard Black" Horne and Jeff Carlin; RPI Professors Doug Whittet and Jon Morse; collaborator Dr. Tim Beers; fellow graduate students Julie Dumas, Ray Menzel, Max Katz, and Jeff Thompson; and undergraduate collaborators ("minions") Matthew Monaco, Katie Grabowski, William Yam, and many others.

To the additional major movers and shakers on the MilkyWay@home project, without whom this thesis would not have been possible, I extend professional gratitude: RPI Computer Science Professors Malik Magdon-Ismail and Bolek Szymanski; Former RPI student and current Professor at UND, and the "father" of MilkyWay@home, Travis Desell; and former RPI student Matthew Arsenault, who kept MilkyWay@home alive.

I would like to acknowledge the NASA/NY Space Grant and the National Science Foundation (grant AST 10-09670), whose funding made this thesis research possible. I would like to acknowledge the Sloan Digital Sky Survey, from which most of the data used here was drawn (see below). I would also like to thank the many volunteers who have contributed to MilkyWay@home, who have, in their own way, enabled much of this research. A special thanks goes out to David Glogau and The Marvin Clan for their significant financial contributions to MilkyWay@home, and to the Dudley Observatory for their assistance. I would also like to acknowledge the National Astronomical Observatory of China (NAOC) for funding my trip to Beijing, and all of my collaborators in China: Licai Deng, Li Jing, Xiaoting Fu, Xu Yan, Yueyang Zhang, and Yang Fan.

I would also like to thank several of the amazing instructors that I have had over the years: RPI Professors Angel Garcia, Joel Geidt, Scott Dwyer, Peter Persans, Toh-Ming Lu, and Gwo-Ching Wang; and UWSP Professors Randy Olson, David Tamres, Greg Taft, Brad Hinaus, and many others.

I am grateful to several organizations who have invited me to present scientific lectures: The Mohawk Valley Astronomical Society; Jerry Silber and the Northeast Fans Association Star Trek Meetup; The Albany Area Amateur Astronomers (AAAA); the UWSP Physics Department; and The Dudley Observatory.

In my private life, this work would not have been possible without the support of my parents, Bob and Cathy, and my younger siblings, Sarah, Rachel and Mark, who have helped keep me in good spirits through frequent communication. I would also like to thank my local friends have also made a major (positive) impact on my sanity: Christoph Stark, Annie Roberts, Camilo Jimenez-Cruz, Pramesh Singh, Panagiotis Karampourniotis, and many others; and the Physics Graduate Student

Council, of which I was co-president, for collaborating on many social events (including the inception of the Post-Qual Pub Crawl): Nick Lanzillo, Adam Simbeck, and David Hutchinson.

With great respect, I would like to thank my local Judo instructors and teammates for giving me a physical outlet and helping me maintain my sanity. I would like to personally thank Sensei Mike D'Alessio, Sensei David Baker, and Matthew Martin, all of whom have served as mentors in the past. I would like to extend a very special thanks to (Hachidan) Sensei Kudo Nuriyasu, under whom I earned my Shodan rank, and who has served as a second advisor to me. Kudo Sensei is the most amazing person I have ever met, a man dedicated to his art, who lives like a modern samurai, and in many respects is the real-life "Mr. Miyagi." His extreme wisdom, patience, skill, and drinking prowess have made a profound impact on many lives, and it has been a great honor to study under him.

* * *

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

ABSTRACT

The stellar halo that surrounds our Galaxy contains clues to understanding galaxy formation, cosmology, stellar evolution, and the nature of dark matter. Gravitationally disrupted dwarf galaxies form tidal streams, which roughly trace orbits through the Galactic halo. The Sagittarius (Sgr) dwarf tidal debris is the most dominant of these streams, and its properties place important constraints on the distribution of mass (including dark matter) in the Galaxy. Stars not associated with substructures form the smooth component of the stellar halo, the origin of which is still under investigation. Characterizing halo substructures such as the Sgr stream and the smooth halo provides valuable information on the formation history and evolution of our galaxy, and places constraints on cosmological models.

This thesis is primarily concerned with characterizing the 3-dimensional stellar densities of the Sgr tidal debris system and the smooth stellar halo, using data from the Sloan Digital Sky Survey (SDSS). F turnoff stars are used to infer distances, as they are relatively bright, numerous, and distributed about a single intrinsic brightness (magnitude). The inherent spread in brightnesses of these stars is overcome through the use of the recently-developed technique of statistical photometric parallax, in which the bulk properties of a stellar population are used to create a probability distribution for a given star's distance. This was used to build a spatial density model for the smooth stellar halo and tidal streams. The free parameters in this model are then fit to SDSS data with a maximum likelihood technique, and the parameters are optimized by advanced computational methods. Several computing platforms are used in this study, including the RPI SUR Bluegene and the Milkyway@home volunteer computing project.

Fits to the Sgr stream in 18 SDSS data stripes were performed, and a continuous density profile is found for the major Sgr stream. The stellar halo is found to be strongly oblate (flattening parameter $q =\sim 0.53$). A catalog of stars consistent with this density profile is produced as a template for matching future disruption models. The results of this analysis favor a description of the Sgr debris system that

includes more than one dwarf galaxy progenitor, with the major streams above and below the Galactic disk being separate substructures. Preliminary results for the minor tidal stream characterizations are presented and discussed. Additionally, a more robust characterization of halo turnoff star brightnesses is performed, and it is found that increasing color errors with distance result in a previously unaccounted for incompleteness in star counts as the SDSS magnitude limit is approached. These corrections are currently in the process of being implemented on MilkyWay@home.

CHAPTER 1
Introduction and Review

1.1 Motivation

Of the many mysteries in astronomy, one of the more prominent is the assembly history of galaxies. For just over a decade, astronomers have been able to study, in depth, the stars that belong to the halo of the Milky Way. These stars are generally at large distances, old, and lacking in heavier elements (McWilliam 1997; Jofré & Weiss 2011). Because these stars may have formed in other galaxies that merged to make the Milky Way, tracing the origins of these stars will provide information on the history of our Galaxy, the formation of structure in the early universe, and the matter distribution in our Galaxy.

The density profile of the stellar halo constrains models of galaxy formation, with the two major models being the "monolithic collapse" model (Eggen et al. 1962) and the "hierarchical construction" model (Searle & Zinn 1978). In the monolithic collapse model, the galaxy formed rapidly from a single collapsing cloud of gas, leaving distinct patterns in the velocities and distributions of halo stars. In the hierarchical construction model, the Galaxy formed through repeated mergers of smaller galaxies, and evidence of these mergers should exist within the halo in the form of co-moving stars with similar elemental abundances. The discovery of the Sagittarius dwarf spheroidal galaxy (Ibata, Gilmore, & Irwin 1994) and proof of its tidal disruption (Ibata et al. 2001; Yanny et al. 2000) indicated that at least some parts of the Galaxy formed through accretion; current thinking is that galaxies form through a combination of the two processes, but more information is needed.

Through work in the 1970's, Vera Rubin showed that Galaxy rotation curves did not match theoretical curves (Rubin & Ford 1970), and that missing mass was required to explain this discrepancy; this became known as the "Galaxy rotation problem." Since then, this missing mass has been dubbed "dark matter," as it does

Portions of this chapter previously appeared as: Newby, M., Cole, N., Newberg, H. J., et al. 2013, AJ, 145, 163.

not appear to interact through the electromagnetic force, and so it is invisible to all wavelengths of light. Determining the shape of the dark matter distribution in our Galaxy will place valuable constraints on the properties on this illusive particle (or particles!). As halo stars orbit outside of the Galactic disk at large distances, they are excellent tracers of the Milky Way's gravitational potential, but these stars are distant enough that their proper motions are difficult to detect, even with current technology. However, stars that are part of tidal streams (such as the Sagittarius tidal stream) imply orbits through the Galactic halo, and therefore the past locations of these stars can be determined. Tidal streams, then, are one of the best indicators available to astronomers concerning the distribution of mass (including dark matter) in our Galaxy.

Virtual models of the Galaxy, such as the Besançon (Robin et al. 2003) and Galaxia (Sharma et al. 2011) models, allow astronomers to compare real data sets with theoretical observations. This technique provides researchers with benchmarks for discovering deviations from expected stellar densities, which can be indicative of new substructures or other dynamics. One downside of these models is that incorrect conclusions may be drawn if the models are using incorrect assumptions about stellar densities or populations. A robust characterization of the Galactic halo, then, will improve these models and the work that relies on them.

The structure of galactic halos is predicted through the use of high-resolution cosmological simulations, such as the Millennium (Springel et al. 2005), Aquarius (Springel et al. 2008), and Phoenix (Gao et al. 2012) simulations. The aim of these simulations is to connect the anisotropies observed in cosmic microwave background (Bennett et al. 2009; *Planck Collaboration* 2013), which give a "snapshot" of the early universe, with the currently observed structure of the Universe, including galaxy halos. By characterizing stellar halo densities, strong constraints are placed on these simulations, and therefore constraints are placed on cosmology itself.

The current working model of our universe is the ΛCDM model, which includes a cosmological constant (Λ) to account for the accelerating expansion of the Universe, and "cold" (non-relativistic) dark matter (CDM). For example, the ΛCDM simulation of Navarro et al. (1996) predicted that dark matter should clump

into numerous "subhalos," and that there should be a strong "cusp," or increase in dark matter density, near the center of galaxies. The previously-mentioned cosmological simulations predict similar observable structures. However, halo density searches have yet to find anywhere near as many subhalos as were predicted; either the model is in error, or the majority of subhalos are made entirely of dark matter (Bullock & Johnston 2005). Additionally, no evidence has yet been found for galaxy cusps, and they are notably absent from dwarf galaxy data. This back-and-forth between observers and modelers refines our understanding of the Universe.

Therefore, studies of stellar densities in the Galactic halo can help to answer long-standing problems in physics and astronomy, such as determining the distribution of dark matter in the galaxy, understanding the formation history of galaxies, and constraining cosmological models.

This thesis concerns the characterization of major stellar density substructures in the Galactic halo, and successfully applies these techniques to the Sagittarius (Sgr) tidal stream in the North Galactic cap (NGC). Preliminary results are presented for the characterization of other halo substructures. The tracer stars (F turnoff main-sequence stars) are rigorously analyzed in order to improve the assumed models. The results of these studies also produce some interesting information on the Galactic halo, and these are discussed in their relevant Chapters. Additionally, the powerful 0.5 PetaFLOPS MilkyWay@home volunteer computing platform, which was used for much of the work in this thesis, is discussed.

1.2 The Milky Way Galaxy

The Milky Way galaxy is our home galaxy, a collection of around 200-400 billion stars. As a spiral-type galaxy, the Milky Way is a flat disk with a radius of about 15 kpc, with a spherical central bulge. The bulge is thought to contain a prominent bar, and a central super-massive black hole with a mass of $\sim 10^6 M_\odot$. The disk can be divided into two components: the thin disk, which contains most of the Galaxy's gas and stars, and has a scale height[1] of around 350 parsecs; and the

[1]The scale height refers to the stellar density falloff factor h, where $\rho(z) \propto e^{-|z|/h}$, where z is the height above (or below) the plane of the Galactic disk.

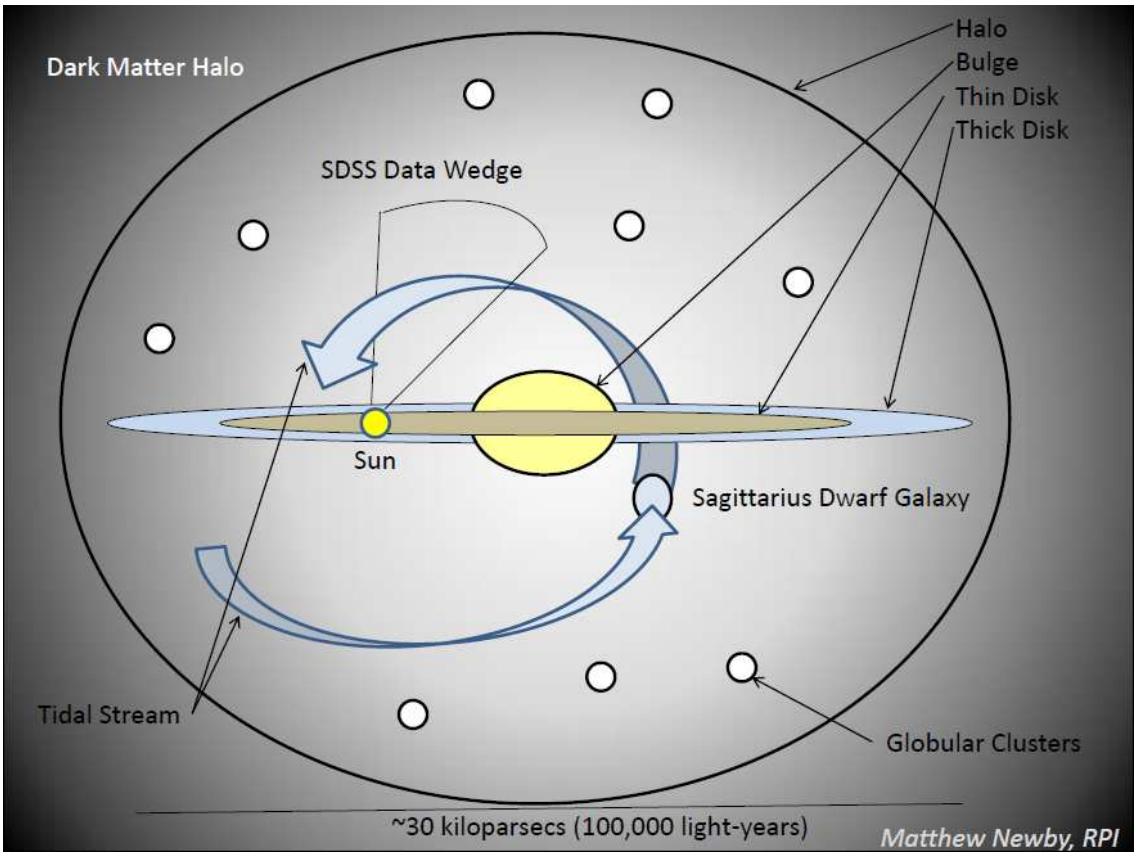


Figure 1.1: A cartoon of the Milky Way galaxy. Shown is the approximate orientation of the Sagittarius dwarf and its tidal streams, as well as an example of the volume that an example SDSS data stripe would sample, relative to the Sun and the Galaxy. Note that in reality, the Sgr dwarf is 20 kpc away and almost directly opposite the Galactic center, relative to the Sun. The major components of the Galaxy are indicated, and discussed briefly in the text.

thick disk, which contains mostly older stars, and has a scale height of around 1 kpc. The disk gas may extend beyond the stellar disk and is warped like a bent vinyl record. Together, the disks and the bulge contain 99% of the stars in the Galaxy.

Surrounding this is the Galactic halo, which consists of a sparsely populated, roughly spherical distribution of stars (the stellar spheroid), compact spherical populations of stars called globular clusters, assorted dwarf galaxies, the circum-Galactic medium (CGM), and dark matter (Majewski 1993). Dwarf galaxies and globular clusters may also have associated tidal streams. The dark matter halo surrounds the Galaxy to an unknown distance, but accounts for 80% to 90% of the total mass of the Milky Way. The constitution and distribution of the dark matter halo is currently under investigation. The CGM is a corona of hot, diffuse gas around the Galaxy, most likely kicked out of the disk, but potentially arriving from outside the Galaxy (the inter-galactic medium, or IGM). If this gas originated from the disk, it may have been ejected by disk supernovae, kinematic disk interactions, or encounters with the central black hole. The CGM is difficult to study, and so the origin and kinematics of its gas are a current active topic in astronomy research. A diagram illustrating the major components of the Galaxy can be found in Figure 1.1.

1.3 Halo Density Searches

It is difficult to study our Galaxy because we are located inside of it. The dust and gas in the Solar neighborhood blocks the visible light from most stars in the disk, and so astronomers must look above or below the plane of the disk in order to study halo stars. Ideally, a complete understanding of the halo requires knowledge of the 6-dimensional phase space coordinates (3 positional dimensions and 3 velocity dimensions) of each star. Observationally, only the angular position of a star in the sky (2 dimensions) can be well-determined (to within 1 arcsec in SDSS data). The proper motions, or velocities tangential to the line-of-sight to a star (2 dimensions), can only be reliably determined for nearby or fast-moving stars; telescope technology is slowly approaching the point at which halo star proper motions can be measured with reasonable accuracies. Spectroscopy can be used to determine the radial velocity of each star (1 dimension) and give the classification of

the star, which helps to constrain the estimated distance (1 dimension). The distance to a star can also be estimated from photometry, but this requires knowledge of the absolute magnitude distribution of the stars being studied.

Until recently, halo density searches relied on relatively few stars and/or subtended small areas of the sky (Gilmore & Reid 1983; Bahcall & Soneira 1984; Reid et al. 1996; Robin et al. 2000; Siegel et al. 2002). Early on, bright objects with known intrinsic brightnesses (such as RR Lyraes and globular clusters) were used to infer the shape of the stellar halo (Oort & Plaut 1975; de Vaucouleurs & Buta 1978; Vivas 2002). Wider-area surveys of halo stars were not very deep, and only a minority of the data was halo stars, with the majority being brighter disk stars (Gilmore & Reid 1983; Preston et al. 1991). In this era, the stellar density of the halo was routinely fit as a power law (Oort & Plaut 1975; Preston et al. 1991) or deVaucouleurs profile (de Vaucouleurs 1977; Bahcall & Soneira 1980).

The Sloan Digital Sky Survey (SDSS, York et al., 2000) was the first digital wide-area survey sensitive enough to detect halo stars with the resolution necessary to study large numbers of halo stars. Using SDSS data, Newberg et al. (2002) showed that the stellar halo contains a large number of significant substructures, and Belokurov et al. (2006b) illustrated just how dominant these structures are in the halo.

In order to effectively use photometrically detected halo stars as tracers of halo substructure, the chosen tracer population must be suited to the task. Ideally, the tracer population should meet some or all of the following criteria: bright, so that the stars are easily detected; numerous, so that they accurately trace substructures and are not susceptible to selection effects; easily segregated from other populations in color, so that there is no confusion in the types of stars selected; and tightly clustered around a single absolute magnitude, so that the inferred distances are accurate.

There are several options for choosing a reliable stellar type when tracing halo densities, and each choice has its benefits, as well as drawbacks. RR Lyrae stars make for good halo tracers (Sesar et al. 2010), but they are relatively rare, and positive identification of RR Lyraes requires multiple observations in order to confirm their

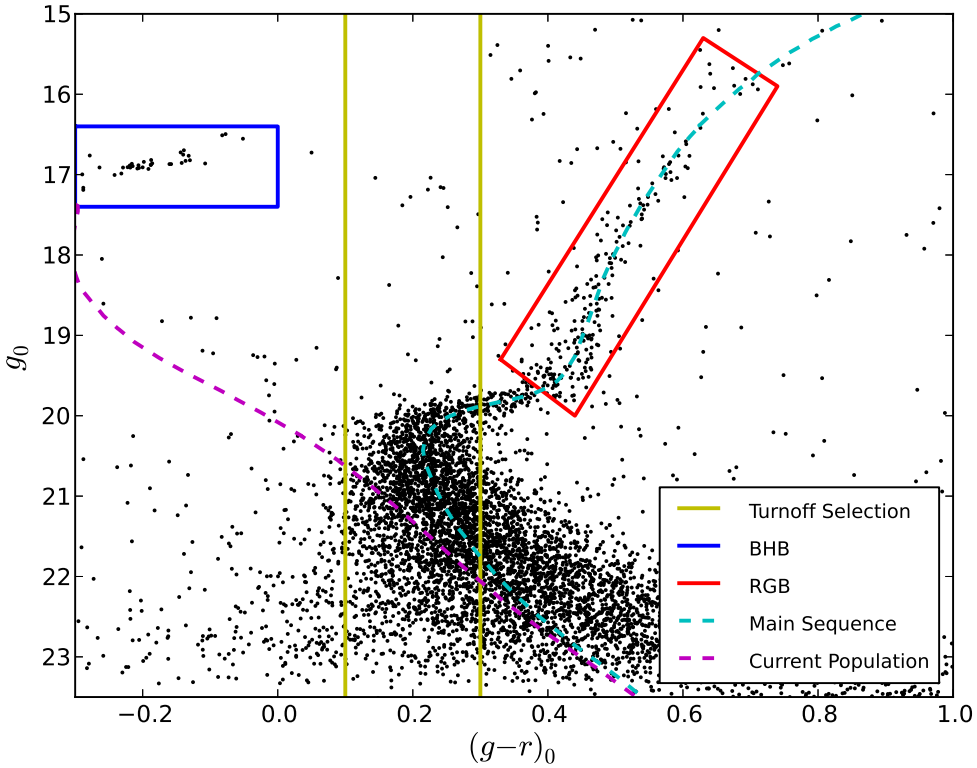


Figure 1.2: Color-magnitude diagram of globular cluster NGC 5024 in SDSS data, highlighting common halo tracers. Blue Horizontal Branch (BHB) stars are bright and tightly grouped around a single absolute magnitude, but are not present in all old stellar populations, and are fairly scarce even when present. Red Giant Branch (RGB) stars are bright, common, and fairly numerous, but have the same colors as fainter red main-sequence stars, and so knowledge of the surface gravity of these stars is required to use them as tracers; this usually requires spectroscopic data. F turnoff stars are plentiful and numerous, and are the only stellar population present in their color range, but have a large spread of absolute magnitudes, which must be taken into account when using them as tracers. An isochrone (age=1.0 Gyr) for the main sequence of this cluster is plotted as the magenta curve, while an isochrone for the current population (age=12.5 Gyr) is plotted as the cyan curve. These were generated using PARSEC isochrones (Bressan et al. 2012) and a metallicity of [Fe/H]=-1.89 (See Section 6.2).

variability. Blue Horizontal Branch (BHB) stars are very bright and tightly grouped around a single absolute magnitude, but are not present in all old halo populations, and are occasionally contaminated by blue straggler stars[2]. When present, though, BHB stars make excellent tracers (Yanny et al. 2000). Red giant branch (RGB) stars are bright and numerous, and have well defined absolute magnitudes as a function of color, but have the same colors as plentiful nearby dwarf stars. RGB stars are occasionally used as tracers, but researchers must be wary of dwarf stars that may contaminate their data and confuse their results. F turnoff stars are numerous, fairly bright, are the only star type in their color range, but have a 1-2 magnitude spread in absolute magnitudes. If this spread in magnitudes is taken into account, F turnoff stars make excellent tracers of halo substructures, and will be used in this thesis to map stellar halo densities (Chapters 3 and 5). In Chapter 6 the properties of halo F turnoff stars are studied in-depth. A color-magnitude diagram for globular cluster NGC 5024 with examples of common tracer populations highlighted can be found in Figure 1.2.

1.3.1 Sagittarius Studies

Of the spheroid substructures, the most dominant and best studied - and potentially the least understood - is the Sagittarius dwarf tidal stream. The Sgr dwarf galaxy was thought to be in the process of tidal disruption from the time of its discovery (Ibata, Gilmore, & Irwin 1994), and the tidal stream was later discovered from some dozens of carbon stars Ibata et al. (2001). At nearly the same time, overdensities of blue horizontal branch and blue straggler stars were discovered (Yanny et al. 2000), which turned out to be cross sections through the tidal stream (Ibata et al. 2001). Early calculations required the Sagittarius dwarf to have either a large dark matter component (Ibata et al. 1997), or to have been deflected into its current orbit within the last 3 Gyr (Zhao 1998). Later simulations by Helmi & White (2001) showed that the observations could be reproduced with or without a massive dark matter halo, even over 12.5 Gyr. Since its initial discovery, many groups have found Sgr stars in the halo, but two of the most striking im-

[2]Blue stragglers are stars that remain on the main sequence, even though the rest of its population has turned off and is heading towards the giant phase of stellar evolution.

ages are from the Two Micron All-Sky Survey (Majewski et al. 2003) and the SDSS "Field of Streams" (Belokurov et al. 2006b).

The Sagittarius dwarf galaxy has been the subject of much debate and confusion among astronomers. Initially, the leading tidal tail of the Sgr dwarf appeared to arc through the North Galactic Cap and rain down on the position of the Sun (Majewski et al. 2003; Law et al. 2005; Martínez-Delgado et al. 2007)[3], sparking a debate about the effects of a dark-matter dominated stream on direct detection of dark matter (Freese, Gondolo, & Newberg 2005). However, later work traced the stream with more numerous color-selected F turnoff stars (Newberg et al. 2007) clearly showed that the leading tidal tail crosses the Galactic plane well outside the solar circle, towards the Galactic anticenter. An apparent bifurcation of the Sgr leading tidal tail was discovered in Belokurov et al. (2006b). Several papers have attempted to explain the bifurcation as multiple wraps of the same stream (Fellhauer et al. 2006) or dwarf galaxy rotation (Peñarrubia et al. 2010). Spectroscopy showed that the velocities and metallicities of the stars in each part of the leading tail bifurcation were similar (Yanny et al. 2009), while follow-up observations found no significant rotation within the Sgr dwarf (Peñarrubia et al. 2011). However, Koposov et al. (2012) showed that the trailing tidal tail was also bifurcated and the photometric metallicity measurements were not the same, so they suggested that in both cases the "bifurcation" could be due to a second stream whose progenitor is not the Sagittarius dwarf spheroidal galaxy. If the stars that were previously assumed to have been stripped from the Sgr dwarf galaxy actually originate from multiple galaxies instead, then it complicates claims of debris from multiple wraps of the stream around the galaxy (Ivezić et al. 2000; Correnti et al. 2010). If true, this would not be the first time that a substructure was later discovered to be composed of more than one component. For example Grillmair (2006); Li et al. (2012) show that the "Monoceros Ring" near the galactic plane (Newberg et al. 2002; Yanny et al. 2003) in the anticenter has multiple components. Furthermore, the Virgo Overdensity (Vivas et al. 2001; Newberg et al. 2002; Jurić et al. 2008) appears to be a mix of several components (Vivas et al. 2012).

[3]Several soft journalists took the results of Majewski et al. (2003) to mean that our Sun had originated in the Sgr dwarf, with some writers extending this idea to imply that the

There is also debate over what the Sagittarius dwarf tidal tails imply about the dark matter potential. Stars that have been tidally stripped from dwarf galaxies and globular clusters are the only halo stars for which we have information about their locations in the past, since they originated from single, relatively small progenitors. These tidal streams therefore imply orbits through the stellar halo, which can be used to constrain the Milky Way's dark matter dominated gravitational potential. Semi-analytic N-body simulations, in which the Milky Way is modeled with a parameterized analytic potential and the dwarf galaxy is simulated with a set of equal-mass particles, have been compared to the locations of the Sgr tidal stream stars. Early analyses argued for a roughly spherical dark matter distribution, since the tidal debris was largely confined to an orbital plane (Ibata et al. 2001; Majewski et al. 2003). However, the measured line-of-sight velocities of the leading tidal tail favor a prolate dark matter potential (Law et al. 2004; Helmi 2004; Vivas et al. 2005); and from orbital alignment of the tails, Johnston et al. (2005) measure an oblate dark matter halo. Law et al. (2005) show that it is not possible to fit both angular positions and radial velocities of the leading Sgr dwarf tidal tail in smooth, axisymmetric models. More recently, a triaxial dark matter halo has been shown to be consistent with all of the data in the Sgr dwarf tidal stream (Law et al. 2009; Law & Majewski 2010), though this result has made some uncomfortable because the Milky Way disk in their best fit model is orbiting about the intermediate axis, and this particular triaxial halo configuration is disfavored by cold dark matter Galaxy formation models (Allgood et al. 2006). Though the triaxial model is currently the best that astronomers have, it does not explain the stream bifurcation(s), and since dark matter halos are likely more complicated than have been modeled to date for the Milky Way, it is possibly not the only model that could fit the data. From N-body cosmological simulations, for example, it seems the shape of the dark matter halo is in general complex, and is a function of radius and time (Bailin & Steinmetz 2005; Tissera et al. 2010).

Tidal streams are useful for tracing the gravitational potential of the Milky

dwarf is responsible for global warming on the Earth! A summary of these pseudoscientific ideas, and and a thorough debunking of them, can be found on Phil Plait's Bad Astronomy blog (http://blogs.discovermagazine.com/badastronomy/2007/06/27/is-the-sun-from-another-galaxy/)

Way even if the dark matter potential is complex, lumpy, or time-dependent. Several techniques for measuring the large-scale shape of the dark matter halo have been put forward (Eyre & Binney 2011; Varghese et al. 2011; Johnston et al. 1999). In addition to the Sgr dwarf tidal stream, the Milky Way halo potential has been fit using the GD-1 cold stellar stream (Willett et al. 2009; Koposov et al. 2010), the NGC 5466 globular cluster tidal stream (Lux et al. 2012), and an attempt has been made to simultaneously fit multiple streams at the same time (Willett 2010). Tidal streams are also being used to constrain the lumpiness of the Galactic halo (Johnston et al. 2002; Ibata et al. 2002; Siegal-Gaskins & Valluri 2008; Yoon et al. 2011; Carlberg 2009, 2012; Carlberg & Grillmair 2013), though tidal tail substructure may not be a good indicator of dark matter substructure (Küpper et al. 2010). The study of the dark matter potential and substructure is a young field, and competing factors such as halo shape, effect of substructure, disk shocking, and time the dwarf galaxy has been on its present orbit have not been completely disentangled. Confusion from multiple stellar streams in the same volume (and sometimes even at similar velocities) in the observational data has also hindered progress.

The Sagittarius stream and other density substructures in the Milky Way make it difficult to fit the shape of the smooth stellar halo. The shape is expected to be a combination of the stars that fell in at early times, and is spatially well mixed (Helmi et al. 2003), plus a component including dwarf galaxies and their associated tidal debris that fell in at late times or is currently infalling, as we see in for example the Sgr dwarf tidal stream. Potentially, if there was some variation of a monolithic collapse event in the history of the Milky Way, then some of the smooth halo stars may have formed during that event. One can see from Newberg et al. (2002) and Belokurov et al. (2006b) that very little of the observed halo is free from the presence of very large density perturbations from tidal streams.

To fit a shape, one either averages over the density substructures, thus making the model fit sensitive to the part of the stellar halo sampled, or uses the small region of the stellar halo that does not appear to contain a dominant substructure component to fit the shape of the stellar halo with large substructures removed. The former method is highly inaccurate, especially when sampling regions of the

halo in which the Sgr stream is present, as a large percentage of halo stars in those regions are due to the Sgr stream. The second method will be sensitive to potentially undiscovered, smaller substructures, and will miss much of the overall structure since large areas of the halo are excluded from the analysis. Due to the deficiencies of the previous two methods, the research detailed in this thesis uses a third method: simultaneously characterizing the smooth stellar halo *and* the associated substructures.

1.4 Project Overview

To characterize the 3-dimensional density structure of the stellar halo and its associated tidal streams, a maximum likelihood technique is used in which a model for stellar spheroid densities is fit to SDSS data. The maximum likelihood algorithm, its implementation, testing and validity is initially described in Cole et al. (2008), and is expanded to fit multiple tidal streams simultaneously in Cole (2009) and Newby et al. (2013). This method assumes a density model with for the smooth stellar halo, and an additional model for a single stellar stream density and orientation. A summary of the major components of this model:

- The model mimics a stripe of SDSS data (Figure 1.1), with three positional coordinates:
  - The SDSS Great Circle angle along the stripe, μ.
  - The SDSS "width" angle, ν.
  - Distance from the Sun, r.
- The smooth halo density is modeled as a Hernquist profile (Hernquist 1990):

$$\rho_{spheroid}(\vec{p}) \propto \frac{1}{r(r+r_0)^3}, \text{ where} \qquad (1.1)$$
$$r = \sqrt{X^2 + Y^2 + \frac{Z^2}{q^2}}.$$

  The Hernquist profile has 2 free parameters:
  - The Hernquist scale length, r_0.

  - The halo flattening, q.
  - The weight (ϵ, see below) is fixed at 1.0.

- The density of each tidal stream is modeled as a cylinder with a Gaussian drop-off from the center:

$$\rho_{stream}(\vec{p}) \propto e^{-\frac{d^2}{2\sigma^2}}. \tag{1.2}$$

  Each stream has 6 free parameters:

  - The weight of the stream in the data, ϵ.
  - The μ position of the stream center, simply called μ.
  - The distance (in kpc) to the stream center, R.
  - The "vertical" orientation of the stream, as the angle from the Galactic $\hat{z}$-axis, θ, with $0 < \theta < \pi$.
  - The "horizontal" orientation of the stream, as the angle from the Galactic $\hat{x}$-axis, ϕ, with $-\pi < \phi < \pi$.
  - The Gaussian width (standard deviation) of the stream, σ.
  - The ν coordinate of each stream is fixed to $\nu = 0.0$.

- The density components are convolved with the expected spread in magnitudes for a halo turnoff star:

$$\mathcal{N}(x; u) = \frac{1}{u\sqrt{2\pi}} e^{\frac{-x^2}{2u^2}}, \tag{1.3}$$

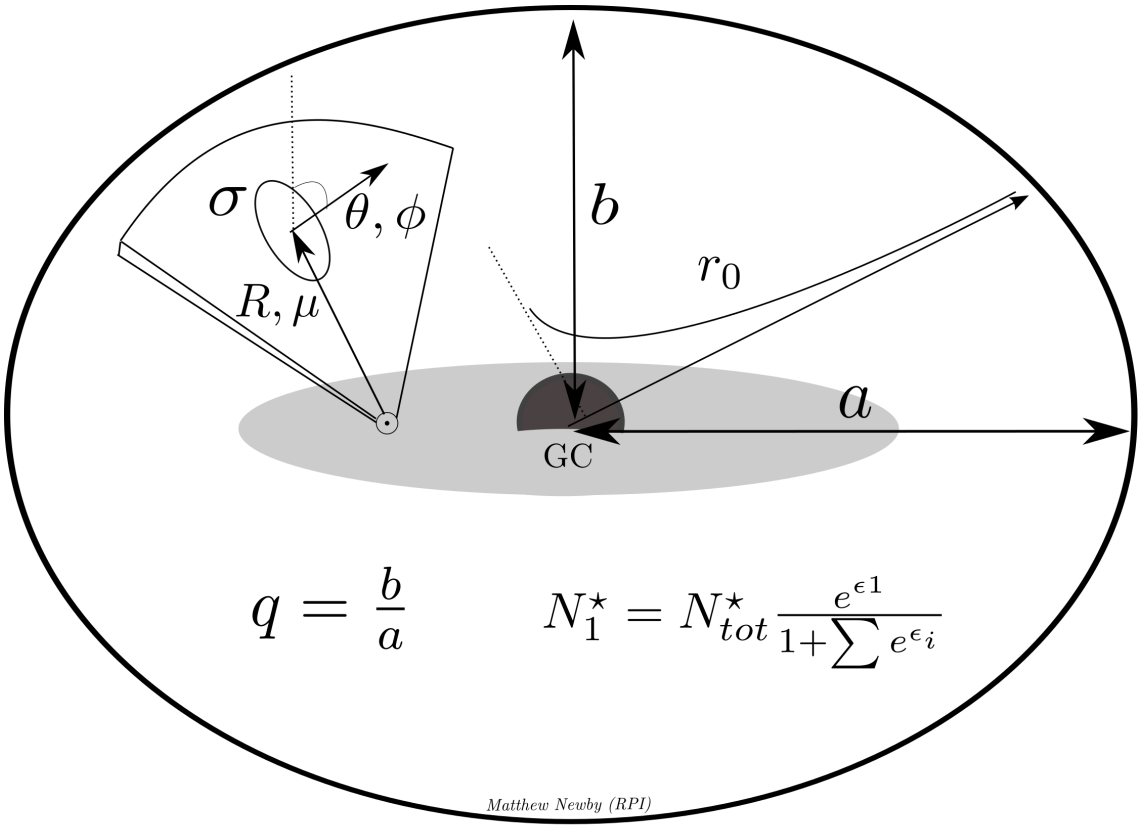
- The turnoff star absolute magnitude was assumed to have a mean of $M_{g_0} = 4.2$, and a standard deviation of $u = 0.6$. This was confirmed to be a reasonable approximation in Newby et al. (2011).

A summary of the density fitting model can be found in Figure 1.3.

The stellar density models are then used to construct a probability density function (PDF), which provides the likelihood that the model describes an SDSS data stripe, for given a set of input parameters. By intelligently iterating through test parameter sets, the maximum value of the PDF - the maximum likelihood - is found. The set of parameters that produces the maximum likelihood is considered to

be the "best fit" parameters, that is, the parameters for which the model most closely matches the data. To search through the possible parameter values, a combination of Conjugate Gradient Descent (CGD) and Line Search techniques are used, and these are described in the appendices of Cole et al. (2008). The model errors in the final parameters were determined through a Hessian matrix technique, as detailed in Newby et al. (2013). More recently, global search techniques such as Particle Swarm and Differential Evolution were used on the asynchronous MilkyWay@home volunteer computing platform, which is the subject of Chapter 4. Additional physical constraints can be applied, such as enforcing density continuity from stripe to stripe, and are discussed later in this text.

With a complete density model and search algorithm, the densities of the stellar spheroid and associated substructures can now characterized. The main focus of this thesis is to characterize the density of Sgr tidal stream and stellar halo; however, the other components of the halo will also be discussed. Chapter 2 discusses the data used in this study. Chapter 3 discusses the density characterization of the main Sgr stream in the North Galactic cap, and the associated density fit to the smooth halo. Chapter 4 describes the MilkyWay@home computing platform, and the updated algorithms that it uses. Chapter 5 discusses the preliminary results of characterizing the density of the secondary, "bifurcated" piece of the Sgr stream. Chapter 6 focuses on the study of halo turnoff star properties, especially their absolute magnitude distribution. Chapter 7 summarizes this thesis, and describes a plan for future work.



Halo:
q (unitless) — $\hat{z}$ spheroid flattening, ratio of axes: b/a
r_0 (kpc) — Spheroid scale length (Hernquist Model)
Stream:
ϵ (unitless) — Stream weight, see equation for $N^\star$ (above)
μ (degrees) — Angle along SDSS stripe (wedge)
R (kpc) — Distance to stream center from Sun
θ (rad) — Stream orientation relative to Galactic $\hat{z}$
ϕ (rad) — Stream orientation relative to Galactic $\hat{x}$
σ (kpc) — Gaussian width of stream
(Note: ν for stream position is always locked to zero)
Integration Volume:
g (mag) — Written as r in param files, but actually g magnitudes!
Steps: 175 or 1400
μ (deg) — μ range for integration. Steps: 400 or 1600
ν (deg) — ν range for integration. Steps: 20 or 640
—*Starfiles are in Galactic l, b and r(kpc), # of stars is first line*

Figure 1.3: A reference sheet for the maximum likelihood algorithm, with the free model parameters placed in physically relevant locations with respect to the Galaxy.

CHAPTER 2
Data Selection

2.1 The Sloan Digital Sky Survey

The main source of data for this work is the Sloan Digital Sky Survey (SDSS), which is a large-field five-filter (*ugriz*) photometric sky survey, and includes the SEGUE-I and SEGUE-II spectroscopic surveys[4]. The photometric survey was carried out with a dedicated 2.5-meter telescope located at Apache Point Observatory, which utilized a 120-megapixel CCD camera capable of imaging over a square degree of sky at a time, using a drift-scan technique. Data from the SDSS was made available to the public, in large blocks called Data Releases (DR). The main goals of SDSS were met with SDSS-II, which included DR1 through DR7, and contained over 230 million photometric objects.

SDSS was extended by the SDSS-III program, including the BOSS, APOGEE, and MARVELS spectroscopic surveys[5], and will continue releasing new data until 2014. Photometric data in all Data Releases was taken primarily away from the Galactic disk in order to avoid the dense dust and gas of the thin disk, and corrected for extinction via the Schlegel maps (Schlegel, Finkbeiner, & Davis 1998). With the most current data release (DR9), the whole of the SDSS images over 14,500 square degrees, or roughly a third of the entire sky. In total, the database now contains almost 0.5 *billion* uniquely detected photometric objects, and over 1.5 *million* unique, moderate- to high-quality spectroscopic objects.

The five *ugriz* flux-calibrated magnitudes for each photometric object are available through the SDSS database, along with additional information describing the type and quality of the data. Included in this data are the magnitude errors ("asinh"

Portions of this chapter previously appeared as: Newby, M., Cole, N., Newberg, H. J., et al. 2013, AJ, 145, 163, and Newby, M., Newberg, H. J., Simones, J., Cole, N., & Monaco, M. 2011, ApJ, 743, 187.

[4]SEGUE: Sloan Extension for Galactic Understanding and Exploration.

[5]BOSS: Baryon Oscillation Spectroscopic Survey; APOGEE: Apache Point Observatory Galactic Evolution Experiment; MARVELS: Multi-object APO Radial Velocity Exoplanet Large-area Survey.

errors, Lupton et al., 1999) and the Schlegel, Finkbeiner, & Davis (1998) reddening values used to correct each magnitude. The SDSS pipeline also computes the probability that an object is a point-source (star-like) or extended (galaxy-like), then assigns the object a "type" based on this information. Flags[6] that describe the quality of the measurement are included for each object; for example, a "SATURATED" flag indicates that one or more of the CCD's pixels were saturated (maximum incoming flux) when measuring that object. All of this data is available through the appropriate Casjobs[7] web interface using SQL-type database queries.

2.2 SDSS Stripe Turnoff Data

The data for the stream-fitting study was queried from the SDSS photometric catalog, with data for the North Galactic cap taken from the DR 7 catalog (Abazajian et al. 2009). Preliminary data for the SGC study was taken from the DR 6 catalog (Adelman-McCarthy et al. 2007), with additional data queried from Data Release 8 (Aihara et al. 2011) when it became available.

The data selection for the stream-fitting algorithm is detailed in Section 4.1 of Cole (2009). The following is a summary of these data selection criteria:

- Magnitude cuts:[8] $16.0 < g_0$, and $g_0 < 22.5$ to $g_0 < 23.5$ (stripe-dependent)
- Color cuts: $0.1 < (g - r)_0 < 0.3$ and $(u - g)_0 > 0.4$
- Galactic latitude cuts: $|b| > 30°$
- Stars flagged as EDGE or SATURATED were rejected.

A complete list of individual stripe bounds can be found in Table 1 of Newby et al. (2013), while a list of data cuts and their associated properties can be found in Table 2 of the same paper.

The resulting dataset includes 1.7 million turnoff stars (selected to be bluer than the turnoff of the Milky Way's thick disk) from eighteen SDSS stripes - fifteen

[6]See: http://www.sdss3.org/dr8/algorithms/photo_flags_recommend.php
[7]http://skyservice.pha.jhu.edu/casjobs/
[8]This corresponds to a distance range of $2.3 < R < 45.7$ kpc for a turnoff star with $M_g = 4.2$.

of which are in the North Galactic cap. The DR 8 data for the Galactic South stripes numbered 78 to 86 include 0.5 million additional turnoff stars.

2.2.1 Casjobs Query

Data from SDSS is acquired from the Casjobs web-based data access application, scripted through the use of modified SQL queries. This section will describe the queries used to acquire the SDSS turnoff star wedges used with the maximum-likelihood algorithm, which were taken from SDSS DR6, using the following query for each stripe:

```
SELECT
   s.run,s.rerun,s.camcol,s.field,s.obj,s.ra,s.dec,
   s.psfmag_g-s.extinction_g as g0,
   s.psfmag_g-s.extinction_g-s.psfmag_r+s.extinction_r as gmr0,
   s.psfmag_u-s.extinction_u-s.psfmag_g+s.extinction_g as umg0,
   seg.stripe
FROM photoobjall s, segment seg, field f
WHERE s.fieldid = f.fieldid
    and f.segmentid = seg.segmentid and stripe = 10
    and type = 6 and (s.status & 512) != 0
```

with "stripe = 10" being replaced by the respective target stripe number, using stripe numbers 9-23 for the NGC, and 79, 82, and 86 for the SGC. The "type = 6" criteria selects only objects that are star-like, as determined by the SDSS photometric pipeline. The "(s.status & 512) != 0" clause selects against stars that have bad photometry (flagged as "SATURATED" or "EDGE"). Further cuts were made to the data in order to use only F turnoff stars, de-select possible quasars, and fine-tune the wedge volume (see previous section). The l, b coordinates for each star are derived subsequently from the queried RA and Dec values.

For newer data from SDSS-III (DR8 and up), different queries were used to reflect some of the changes and features that had been added to the system. It is important to note that the SDSS-III database is independent from the the previous SDSS programs; however, all previous data is accessible through the SDSS-III

Casjobs portal (but not vice-versa). The following query was used in order to acquire the F turnoff stars located in the SGC:

```
SELECT l,b,
  dered_u, dered_g, dered_r, dered_i, dered_z,
  psfMag_u, psfMag_g, psfMag_r, psfMag_i, psfMag_z,
  extinction_u, extinction_g, extinction_r,
  extinction_i, extinction_z
FROM star
WHERE (dered_g between 16.0 and 23.0) and
  (b < -20.0) and
  (dered_g - dered_r between 0.1 and 0.3) and
  (dered_u - dered_g > 0.4) and
  (flags & dbo.fPhotoFlags('SATURATED')) = 0 and
  (flags & dbo.fPhotoFlags('EDGE')) = 0
```

This query includes the magnitude limits, F turnoff color selection, quasar removal clause, and a Galactic disk avoidance criteria, but does not separate the stars into numbered wedges. Numbered wedges were created from the queried data by asking, for each SGC stripe number, which stars were within 1.25° of that stripe's great circle ($\nu = 0.0°$). Stars meeting this criteria were placed in the wedge data subset for that stripe number. Note that since SDSS stripes overlap near the the ends of each stripe, many stars appear in more than one wedge.

In order to be consistent with the data previously queried from DR6 and DR7, only "EDGE" and "SATURATED" stars are removed from the data, and PSF magnitudes (to have their respective extinctions subtracted from them during processing) are used. SDSS-III adds a "clean" flag, which when selected for ("clean=1"), removes all objects with non-ideal photometry; however, this criteria was considered too broad for our needs and would create datasets that are not comparable to previously queried NGC wedges. Note that the "from Star" clause queries a different table (in this case a SQL database "view") than before, as the "Star" view only contains primary and type=6 (star) objects from the PhotoObjAll table.

The entirety of the North Galactic cap F turnoff data can be retrieved in this fashion as well, but since there is much more data, the queries must be broken up into smaller chunks in order to produce manageable file sizes. It was found that the following four b ranges produced reasonably-size data files: b between 20.0 and 30.0; b between 30.0 and 45.0; b between 45.0 and 60.0; and $b > 60.0$.

2.3 Globular Cluster Data Selection

SDSS data for several globular clusters was used in the halo turnoff star magnitude distribution analysis performed in Chapter 6. The Harris globular cluster catalog (Harris 1996) was consulted to determine which globular clusters were located within the SDSS imaging footprint. Clusters were rejected even if they were within the survey footprint if they had a distance of 60.0 kpc or more, as this is the maximum distance at which a cluster's turnoff star distribution could be effectively studied in SDSS data. This is because the 50% detection efficiency for SDSS occurs at a g magnitude of 23.5, so that an average turnoff star with $M_g \sim 4.2$ would not be detected effectively at a distance of 72.4 kpc. Since an effective characterization of turnoff stars requires stars that are fainter than the average by at least a magnitude, any clusters greater than 60.0 kpc away ($g = 23.1$) were rejected.

2.3.1 Casjobs Query

Globular clusters are fairly small objects within the SDSS footprint, and so a more constrained data selection method was used when querying the database. First, the list of Globular clusters located in the SDSS footprint was created by referencing the Harris catalog. To determine the limits in right ascension and declination for selecting stars for each field, the SDSS SkyServer's Navigate tool was used[9]. The image field of view was expanded so that the cluster was clearly visible, then expanded further until a sizable background distribution was also contained within this view. In general, the entire field of view spanned a rectangle with sides of lengths between 6 and 8 times the apparent radius of the cluster. The bounds of this rectangle were then used as the right ascension and declination limits in the

[9]Located at http://cas.sdss.org/dr7/en/tools/chart/navi.asp

Casjobs data query.

Within these limits, the data selection included all objects classified as STAR, and that had $(u-g)_0 > 0.4$. The latter requirement was designed to avoid possible contamination from quasars (Yanny et al. 2000). Selecting from the database of "STAR"s ensured that only one instance of each object was obtained. Also extracted were the extinction-corrected (denoted by the subscript '0') point-spread function (PSF) apparent magnitudes with errors. The radius within which the majority of the stars belong to the globular cluster, r_{clus}, and the radius outside of which there is little contamination from cluster stars, r_{cut}, was determined by visual inspection of the data. Stars with $r_{clus} < r < r_{cut}$ were removed from the data set. The r_{clus} and r_{cut} values used for each cluster are included in Table 6.1. An example data set is shown in Figure 2.1.

The following is the globular cluster data query for NGC 4147:

```
# --- NGC_4147

SELECT ra, dec, dered_g, dered_r, dered_u, flags
FROM Star
WHERE (dered_g - dered_r BETWEEN -0.30 AND 0.60) AND
  (dered_u - dered_g > 0.4) AND
  (ra BETWEEN 182.34 AND 182.70) AND
  (dec BETWEEN 18.38 AND 18.69)
```

The Right Ascension (R.A.) and Declination (Dec.) limits would change depending on the target cluster.

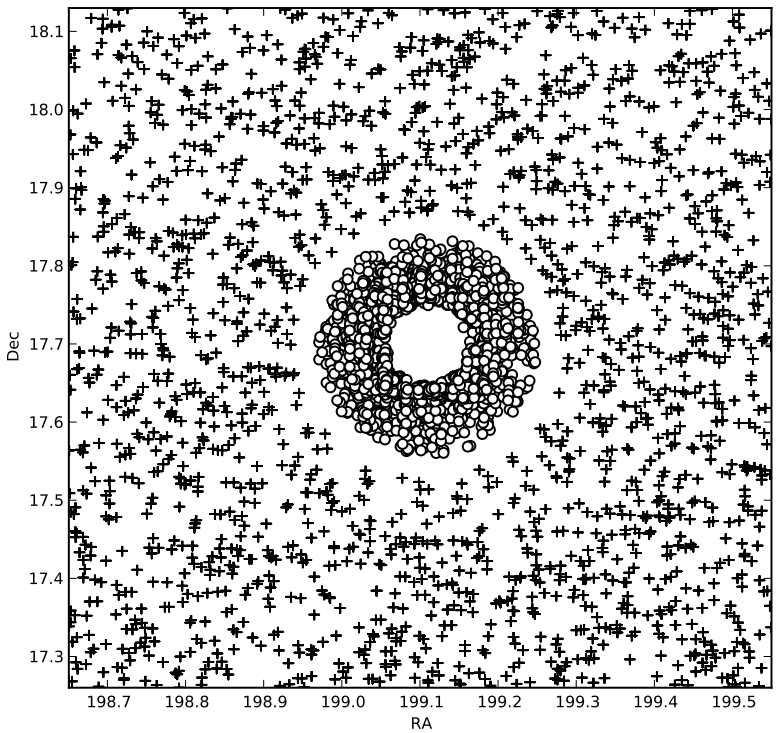


Figure 2.1: Positions of stars near globular cluster NGC 5053 in right ascension and declination. Stars assigned to the cluster are marked with circles, while stars assigned to the background are marked with crosses. At the edge of the cluster, it is difficult to separate cluster stars from background stars, so a ring of stars at the interface has been removed, leaving a gap on the plot. The empty area in the center of the plot is where the SDSS photometric pipeline was unable to separate individual stars, and the data was not included in the database. This missing area has been subtracted from the cluster area determinations in Chapter 6.

CHAPTER 3
The Sagittarius Tidal Debris

The major substructures present in the Galactic halo affect density studies of the smooth stellar component, and if they are not accounted for, incorrect conclusions may be drawn about the stellar halo. This problem has been encountered by several authors, most notably in early pencil-beam studies which could return a large number of different halo density profiles. More recent studies (such as Jurić et al. (2008)) have also found halo substructures to be a hindrance to studying the smooth stellar halo. Sagittarius has proven to be a dominant influence on the results of halo substructure studies, even with modern techniques, as Sgr stars are well-mixed with the smooth halo stars, and have similar colors and metallicities. It is necessary to characterize Sgr - and potentially other halo substructures, as well - in order to accurately study the smooth component of the spheroid.

Using the method of statistical-photometric parallax (Newberg 2013a) with halo turnoff stars, the overdensities in the halo can be characterized. A Gaussian probability distribution for turnoff star absolute magnitudes in the SDSS g band is assumed, with a mean of $\mu = 4.2$ and a standard deviation of $\sigma = 0.6$ magnitudes (Newberg & Yanny 2006). This was included in the probability density of the maximum likelihood algorithm, as described in Cole (2009) and Newby et al. (2013), and confirmed to be a reasonable approximation in Newby et al. (2011). The bright Sgr tidal stream in the North Galactic cap can then be characterized. Once this dominant substructure is accounted for, the less prominent substructures (Virgo and the faint Sgr stream) can be characterized, and the the true density distribution of the stellar halo can be determined.

Portions of this chapter previously appeared as: Newby, M., Cole, N., Newberg, H. J., et al. 2013, AJ, 145, 163.

3.1 Fitting the Sagittarius Tidal Stream

While the Sgr tidal stream is the dominant over-density in the Galactic halo, it is not the only significant structure in the SDSS data, as is commonly illustrated by the "Field of Streams" image (Belokurov et al. 2006b). In addition to the main Sgr stream, a secondary, fainter (less-dense), roughly parallel structure can be seen; this is has been dubbed the "bifurcated" piece of Sgr (Belokurov et al. 2006b). This fainter stream is a source of confusion for the maximum-likelihood algorithm, especially when it coincides with the angular coordinates of the Sgr stream. Also located in SDSS, and close to Sgr, is the Virgo stellar stream (Vivas et al. 2001). Virgo is at lower Galactic latitude than Sgr, and represents a significant over-density in stripes 9-12. At low Galactic latitudes, in the opposite direction of the Galactic center (the anti-center) lies the Monoceros over-density, as well as other, fainter structures (Li et al. 2012). These low-latitude structures are easily avoided by ignoring data at low Galactic latitude ($|b| < 30°$). In the South Galactic cap, the Hercules-Aquila cloud (Belokurov et al. 2007) represents the only (known) significant structure in our data. This structure is easily avoided, as it is well-separated from the Sgr detections in the Southern stripes. Also, there may exist other substructure, data artifacts, or non-uniformities in the data. A complete list of the bounds used in each stripe can be found in Table 1 of Newby et al. (2013), and the excluded volumes are tabulated in Table 2 of that paper.

Although the volume cuts removed several potential sources of confusion from the data, there are still several over-densities present, in addition to the main Sgr stream. In order to prevent these objects from contaminating the fits to the Sgr stream, 3 streams were fit to stripes located in the North Galactic cap: the first stream fits Sgr, the second stream fits either the bifurcated piece or Virgo, and the third stream is available to pick up any remaining non-Sgr over-density (the "garbage collection" stream). In the event that too many streams were being fit to the data, testing showed that the algorithm would reduce the ϵ of the extraneous streams to a large negative number, effectively placing zero stars in that stream (Cole 2009). Therefore, allowing the algorithm to fit one or two additional streams would not affect the accuracy of the method, but does increase the time the algorithm

Table 3.1: Sagittarius Stream Fit Parameters

Stripe	ϵ	$\mu(^\circ)$	R (kpc)	$\theta(rad)$	$\phi(rad)$	σ	# Stars
9	-0.4 ± 0.6	220.0 ± 3.0	43.0 ± 3.0	1.4 ± 0.4	-1.8 ± 0.4	4.8 ± 0.8	19,386
10	-1.4 ± 0.4	210.0 ± 20.0	44.0 ± 5.0	1.1 ± 0.5	3.0 ± 2.0	6.0 ± 0.5	18,734
11	-1.3 ± 0.3	206.0 ± 1.0	41.0 ± 13.0	1.2 ± 1.3	-2.9 ± 0.5	3.8 ± 0.8	11,095
12	-1.3 ± 0.4	202.0 ± 5.0	41.0 ± 6.0	1.3 ± 0.5	3.1 ± 0.3	5.8 ± 1.9	17,044
13	-1.0 ± 0.4	200.0 ± 12.0	36.0 ± 5.0	1.5 ± 0.2	-3.0 ± 1.0	3.8 ± 0.9	18,736
14	-1.4 ± 0.6	184.0 ± 7.0	28.0 ± 9.0	1.8 ± 0.8	2.9 ± 0.3	4.0 ± 1.0	15,409
15	-1.7 ± 0.6	180.0 ± 10.0	27.0 ± 8.0	1.9 ± 0.7	3.0 ± 0.6	3.0 ± 2.0	12,519
16	-1.5 ± 0.6	169.0 ± 7.0	28.0 ± 6.0	2.3 ± 0.7	2.9 ± 0.5	2.4 ± 0.6	12,248
17	-1.6 ± 0.7	162.0 ± 7.0	26.0 ± 6.0	2.0 ± 0.8	2.9 ± 0.3	2.5 ± 0.8	8,853
18	-2.0 ± 0.8	157.0 ± 7.0	25.0 ± 7.0	2.1 ± 0.7	3.0 ± 1.0	2.2 ± 0.5	7,328
19	-1.9 ± 0.4	151.0 ± 7.0	23.0 ± 7.0	2.4 ± 1.6	3.0 ± 2.0	0.9 ± 3.0	5,479
20	-2.8 ± 0.5	148.0 ± 7.0	23.0 ± 6.0	2.3 ± 0.3	3.0 ± 0.7	1.1 ± 0.3	4,450
21	-2.1 ± 0.3	140.0 ± 15.0	21.0 ± 8.0	2.6 ± 0.9	3.0 ± 2.0	1.1 ± 0.4	3,486
22	-2.0 ± 1.0	140.0 ± 14.0	18.0 ± 8.0	2.4 ± 1.0	1.9 ± 2.5	2.0 ± 1.0	2,425
23	-4.0 ± 1.0	130.0 ± 16.0	17.0 ± 18.0	2.7 ± 1.6	1.4 ± 1.6	1.0 ± 2.0	971
79	-2.4 ± 0.4	38.0 ± 5.0	30.0 ± 5.0	2.2 ± 0.5	0.3 ± 0.6	2.8 ± 0.1	9,511
82	-1.8 ± 0.3	31.0 ± 3.0	29.0 ± 3.0	1.7 ± 0.5	0.0 ± 1.0	3.0 ± 1.0	16,119
86	-1.7 ± 0.4	16.0 ± 5.0	26.0 ± 3.0	1.4 ± 0.1	0.1 ± 0.2	2.4 ± 0.5	16,603

takes to converge. The non-Sgr components were not subject to stripe-to-stripe continuity restrictions to allow them to adapt to the potentially different conditions in each stripe. Since the main concern was fitting the Sgr stream, these additional components are not analyzed here and will be discussed only briefly in this chapter; the characterization of these secondary substructures is the subject of Chapter 5.

The data in the South Galactic Cap appeared to be less cluttered than in the North, so only one stream was fit to these stripes during this study. Noticeable additional structures in the South (such as Hercules-Aquila) lie far enough from the Sgr stream that we could reduce the angular extent of the data in order to exclude them. However, the fainter, bifurcated piece in the South was not accounted for in the fits. It is possible that this bifurcated piece was included in the main Sgr stream by the algorithm; if so, it may have skewed the fit to have a larger width or gave it a slightly different direction. It is also possible that the bifurcated piece in the South is too faint for our algorithm to detect. In any case, the separation plots (described in the next section) for the Southern stripes appeared to cleanly remove all over-density, so it is believed that the fits are reasonable. The Southern stripes are currently in the process of rigorous re-analyzation, as these methods are being applied to SDSS DR8 (Thompson et al., *in prep.*).

Table 3.2: Sagittarius Stream Centers

Stripe	$X_\odot(kpc)$	$Y_\odot(kpc)$	$Z_\odot(kpc)$	l	b	$\Lambda_\odot$	$B_\odot$	Λ_{new}[1]	B_{new}[1]	$R_\odot(kpc)$
9	18.3	-5.3	33.8	348.8	51.0	-66.6	-1.3	-74.6	-0.5	43.4
10	14.4	-7.9	37.3	340.9	57.0	-74.0	0.6	-82.2	-1.4	44.5
11	8.7	-8.8	36.3	332.9	62.0	-80.3	1.6	-88.6	-1.4	41.1
12	5.2	-9.3	37.1	325.9	66.0	-85.4	1.7	-93.6	-0.7	40.7
13	-0.2	-9.2	34.0	312.1	70.0	-91.9	2.4	-100.1	-0.4	36.2
14	-7.7	-9.3	26.9	275.1	70.9	-103.7	5.6	-112.3	-1.9	28.5
15	-9.9	-8.8	25.8	260.8	71.0	-108.3	5.1	-116.8	-0.8	27.3
16	-15.2	-10.3	24.9	236.9	63.7	-119.7	7.8	-128.3	-2.1	27.8
17	-17.7	-9.9	22.0	226.9	58.5	-126.8	8.6	-135.4	-2.1	25.8
18	-19.7	-9.6	20.6	220.5	54.3	-132.3	8.6	-140.8	-1.6	25.4
19	-21.0	-8.9	17.8	215.5	49.3	-138.3	9.1	-146.7	-1.6	23.5
20	-21.9	-8.4	16.3	212.0	45.9	-142.5	8.8	-150.9	-1.0	22.7
21	-22.7	-8.0	13.5	209.2	39.7	-149.0	10.0	-157.4	-1.8	21.2
22	-21.9	-6.9	10.1	207.2	33.9	-155.0	11.1	-163.3	-2.6	18.1
23	-21.4	-6.2	8.5	205.7	30.7	-158.5	11.1	-166.8	-2.6	16.7
79	-27.7	5.8	-22.6	163.3	-48.4	114.4	-2.9	105.8	5.0	30.2
82	-23.2	5.6	-24.7	159.2	-57.6	105.1	-1.2	96.8	1.8	29.2
86	-14.1	5.6	-24.8	134.8	-72.3	87.4	-0.0	79.5	-2.1	26.1

[1]Defined in the Appendix of Newby et al. (2013)

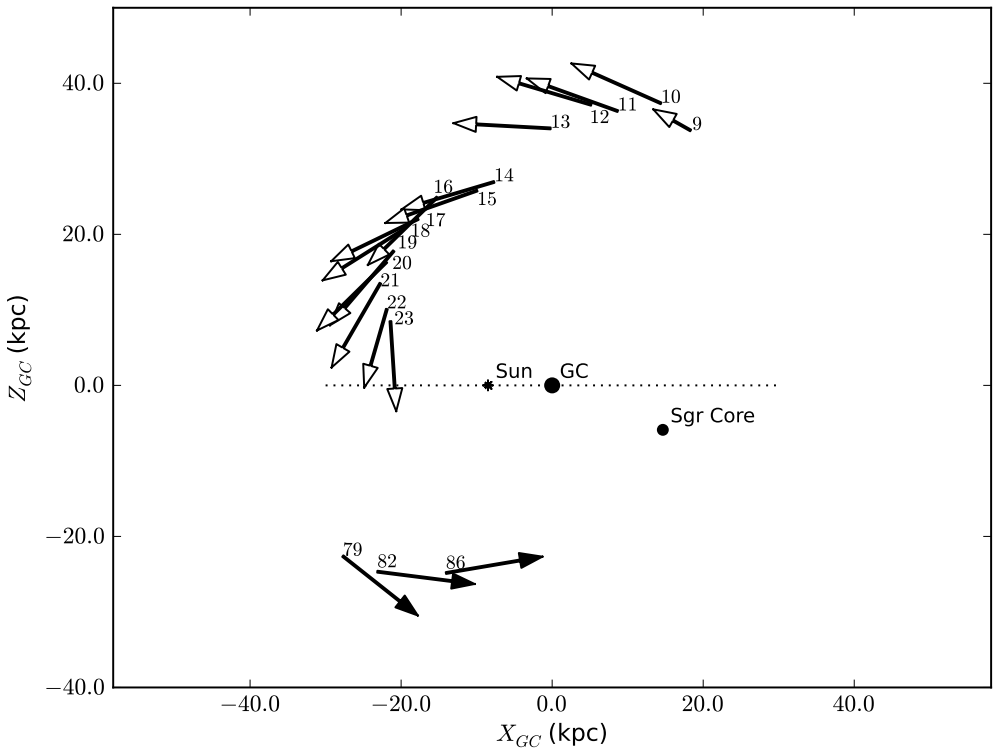


Figure 3.1: Sagittarius stream fit directional vectors plotted in Galactic X, Z. The base of each arrow is labeled with its respective SDSS stripe number. White arrows represent stream detections located above the disk of the galaxy, while black arrows represent stream detections located below the disk. The base of each arrow represents the location of the stream detection, while the head of the arrow illustrates the direction of the axis ($\hat{a}$) of the cylindrical shape that is fit to the stream in that stripe. The length of the arrows is set by the projection of the unit vector $\hat{a}$ in the X_{GC}, Z_{GC} plane, each scaled by the same arbitrary multiple to make the relative lengths more apparent. The dotted line shows the plane of the Milky Way.

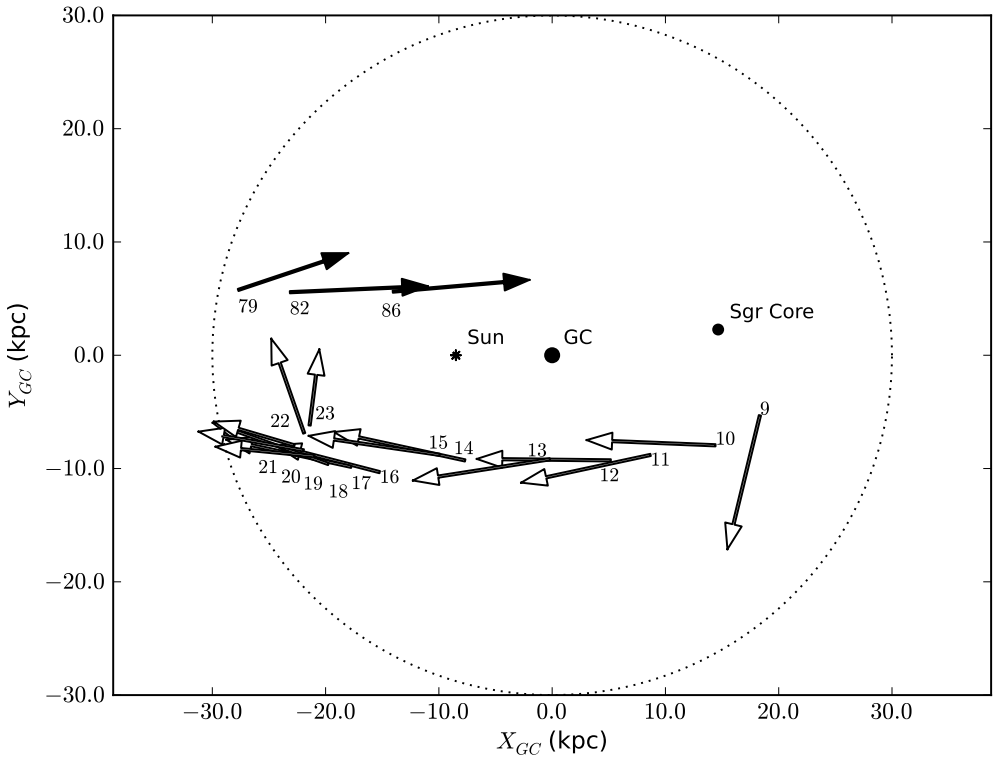


Figure 3.2: Sagittarius stream fit directional vectors plotted in Galactic X, Y. The base of each arrow represents the location of the stream detection, while the head of the arrow illustrates the direction of the stream. The length of the arrows is the same arbitrary multiple of the unit vector $\hat{a}$ projections used in Figure 3.1. The dotted line represents a 30 kpc ring centered on the Galactic center. Note that the direction vectors are less accurate at the edges of the Northern data (SDSS stripes 9-23), where no adjacent stripes are available to guide the fit.

Using the maximum-likelihood algorithm, the best-fit Sgr stream parameters were found, and are are tabulated in Table 3.1. The μ,R parameters define the spatial location of the Sgr stream center, and the θ,ϕ parameters define the stream direction. The ϵ parameter describes the weight of the stream relative to the background, and σ gives the standard deviation of the stream width. "# Stars" is the number of stars that are consistent with the Sgr stream in a given stripe. The fit values are depicted in Figures 3.1 and 3.2, which show the position and orientation of the Sgr stream in each stripe of SDSS data analyzed here. Our fits describe a continuous, consistent, circum-Galactic orbit. When compared with Table 1 of Newberg et al. (2007), where a similar study of halo turnoff stars was performed, the optimized longitudes and latitudes are similar, but the distances for stripes 15-22 are on average 2 kpc larger than those produced here by the maximum likelihood method. It is possible that since the bifurcated Sgr stream was not accounted for in their study, it may have skewed their distance determinations.

The stream centers have been converted into common coordinate systems in Table 3.2: Galactic Cartesian XYZ, centered on the Sun with the Galactic center in the positive X direction; Sun-centered Galactic longitude and latitude, l, b; the Sun-centered Sagittarius $\Lambda_\odot, B_\odot$, coordinate system first described in Majewski et al. (2003), using the C++ code made publicly available by David Law[10]; and a new Sun-centered Sagittarius $\Lambda_{\rm new}, B_{\rm new}$ that reflects the new stream center fits, as defined in Appendix A.2 of Newby et al. (2013).

3.2 Separating Sgr Tidal Debris from other Halo Stars

Using the separation method described in Cole (2009), sets of stars were extracted that are consistent with the density profile of the maximum likelihood fits to the Sgr tidal stream. For each stripe, the original dataset of F Turnoff stars in μ, ν, g_0 would be taken, as illustrated using stripe 18, in Figure 3.3. Using the separation method, the original data could be separated into subsets - one subset for each fit component: one subset for the background, and one more subset for each stream. The "garbage collection" stream would always move out to outer r limit

[10]http://www.astro.virginia.edu/~srm4n/Sgr/code.html

of the data, and grow to be very wide. It is possible that this indicates that the Hernquist model does not sufficiently describe the smooth component of the stellar halo, and so the "garbage collection" stream compensated for the Hernquist model's shortcomings.

Another possibility is that the "garbage collection" is accounting for inaccuracies in the density model. SDSS photometric color errors increase with distance, and it will be shown in Chapter 6 that these errors cause the magnitude distribution of F turnoff halo stars to change with distance, especially near the detection limit of the survey. The "garbage collection" stream may, then, be fitting the extra density that is created by photometric errors, and is not accounted for in the model used here.

Figure 3.4 illustrates the separation process, by simultaneously plotting the μ, g_0 subsets of stripe 18. Here, μ is the angle along the stream in SDSS Great Circle coordinates. The subset corresponding to the Sgr stream can be seen as the top-right plot, while the smooth Hernquist background can be seen as the subset plotted in the top-left. The diffuse stream-like structure, at bottom-left, is the rough fit to substructure secondary to Sgr; in stripe 18, this is most likely to be the so-called bifurcated piece. The very large, distant "garbage collection" stream is shown as the bottom-right dataset. The smooth appearance of the remaining background is confirmation that the three streams have completely fit any non-smooth structure.

Presented in Figure 3.5 is the entirety of the North Galactic cap data, in an l, b polar plot. The North Galactic pole is located at the center of the plot, with Galactic latitude (b) decreasing radially. This plot is comparable to the "Field of Streams" plot, with the Sgr stream, its bifurcated piece, and the Virgo stream all clearly visible.

After extracting the stars consistent with the Sgr tidal stream from each SDSS data stripe, the results were combined to produce the complete Northern profile of the Sgr tidal stream, as seen in SDSS data, in Figure 3.6. The stream is moving away from the Galactic center ($l = 0°$), and therefore the Sgr core, and becomes thinner and less dense the further it is traced from the core. The width appears constant along the stream; this is an illusion caused by the stream being more distant closer

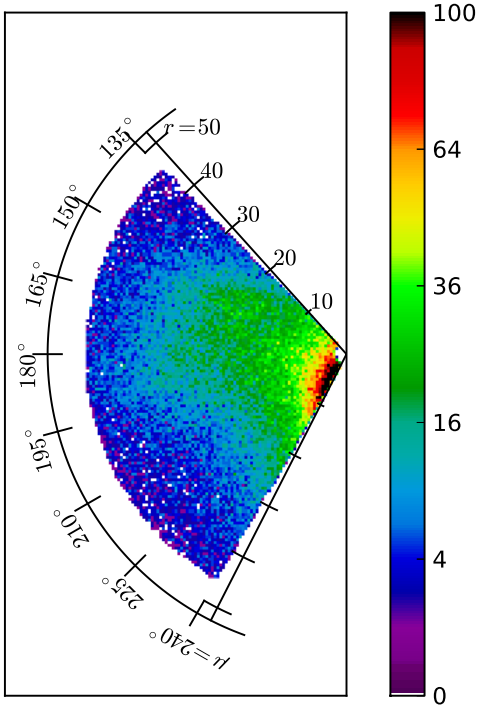


Figure 3.3: Face-on wedge plot of SDSS Stripe 18 stellar density, in μ (angle along stripe as marked on the outside circle) and distance r (kpc, radial from center). The radial position assumes each star has an absolute magnitude of $M_g = 4.2$. The number of stars contained in each pixel is given by the pixel's color. Tidal debris can be seen as a dark density 'cloud' in the region bounded approximately by $15 < R < 35$, $150° < \mu < 190°$. This stellar density wedge plot is typical of the SDSS data stripes analyzed in this thesis.

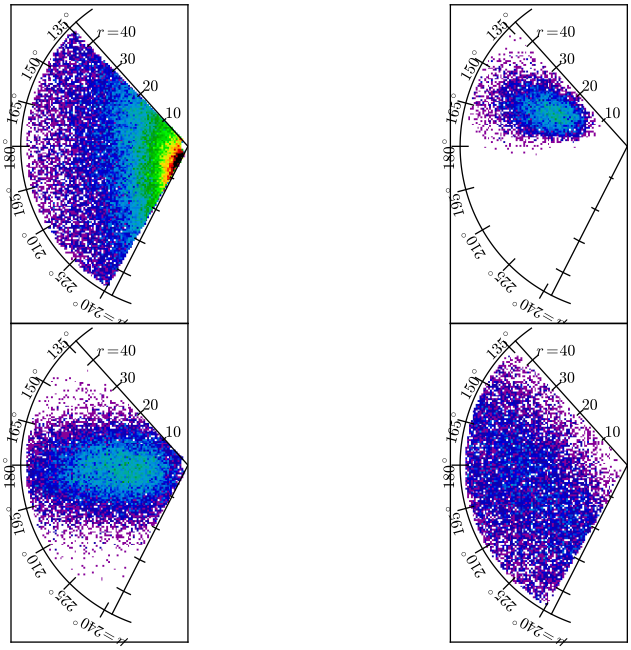


Figure 3.4: Wedge plots of the probabilistically separated SDSS Stripe 18, with r plotted radially, and the angle along the stripe, μ, plotted in angle around the center in each panel. Top left: remaining smooth component, after fit streams have been removed. Top-right: Separated Sgr Stream, as fit within SDSS Stripe 18 data. Bottom-left: Secondary stream fit, most likely the bifurcated piece. Bottom-right: "Garbage collection" stream. Substructure appears to have been completely removed, leaving only the smooth component behind. The segment of the Sagittarius tidal stream that passes through Stripe 18 also appears to have been cleanly separated. The other two streams were not subjected to the stripe-to-stripe continuity requirements that was enforced for Sgr, and so were free to fit any density in the data not described by Sgr or the background Hernquist profile.

to the Galactic center, and becoming closer as it travels towards the edge of the galaxy. A catalog of stars consistent with the density profile of Sgr can be found in Table 5 of Newby et al. (2013).

The stars remaining after the removal of the Sgr leading tidal tail can be seen in Figure 3.7. Note that Sgr is cleanly removed (compare with Figure 3.5), leaving a smooth background in its place. The bifurcated piece can still be found close to the center of the plot, and Virgo can be seen near the bottom of the data, near $l = 300°$. The smooth density of stars increases toward the Galactic center, below $b = 45°$. These components are not analyzed further in this paper, as they were not subject to the rigorous fitting procedure used to derive the model fits to the Sgr debris, and will be the subject of Chapter 5.

3.3 Plane Fits to the Sgr Tidal Tails

From the data in Figures 3.1 and 3.2 and Table 3.2, it was found that the leading and trailing tidal tails trace independent orbital planes (Johnston et al. 2005; Law et al. 2005; Koposov et al. 2012). The original orbital plane of Majewski et al. (2003) does not coincide with our stream centers, as the Majewski et al. (2003) fit has the equation:[11]

$$-0.064X + 0.970Y + 0.233Z + 0.232 = 0. \tag{3.1}$$

A new set of orbital planes is then defined: one for the leading tail and one for the trailing tail. Using all fifteen detections of the leading tail, a gradient descent method was used to fit a plane to the points along the leading tail (See Appendix A.1 of Newby et al. (2013) for details, including error analysis. A new set of Sun-centered in-plane Λ, B coordinates is also defined in that section.) The resulting plane has the equation

$$-0.199X + 0.935Y + 0.293Z - 0.988 = 0 \tag{3.2}$$

[11]Note that this equation is presented here in right-handed X, Y, Z coordinates, giving it a slightly different appearance from the left-handed Majewski et al. (2003) presentation.

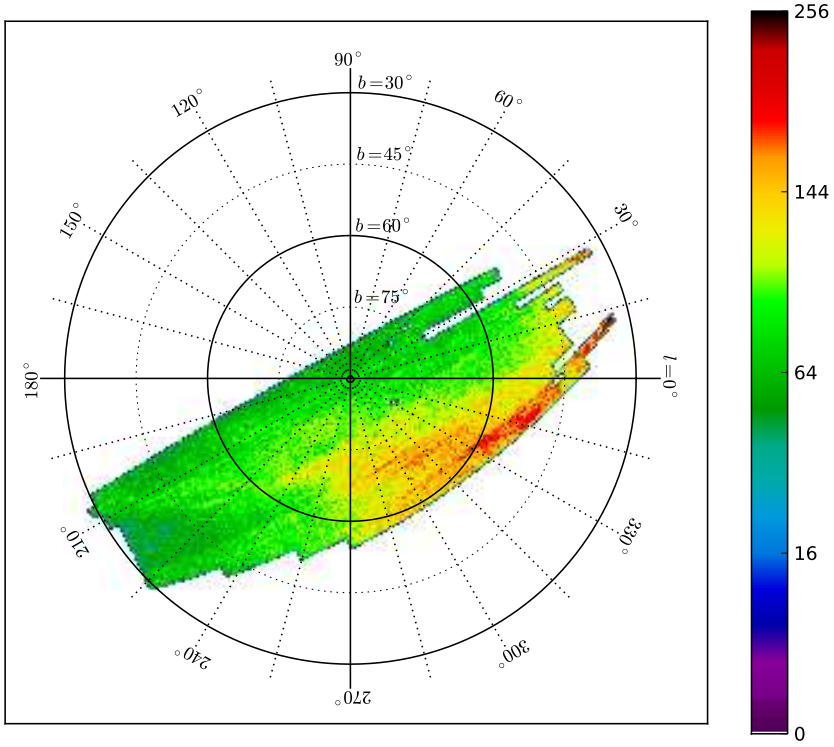


Figure 3.5: Polar plot of F turnoff stars in the Northern Galactic Cap in Galactic l, b coordinates. Each colored pixel is 0.5 degree by 0.5 degrees in size, with color indicating the total number of turnoff stars in that area of the sky. The Sagittarius stream is the overdensity near the center of the plot, spanning from $l, b \approx (210°, 40°)$ to $l, b \approx (345°, 55°)$. The bifurcated piece can be seen as a faint secondary stream just above Sagittarius (closer to $b = 90°$). The Virgo over-density is lower in the plot, and is centered (in the data) near $(l, b) \approx (315°, 60°)$.

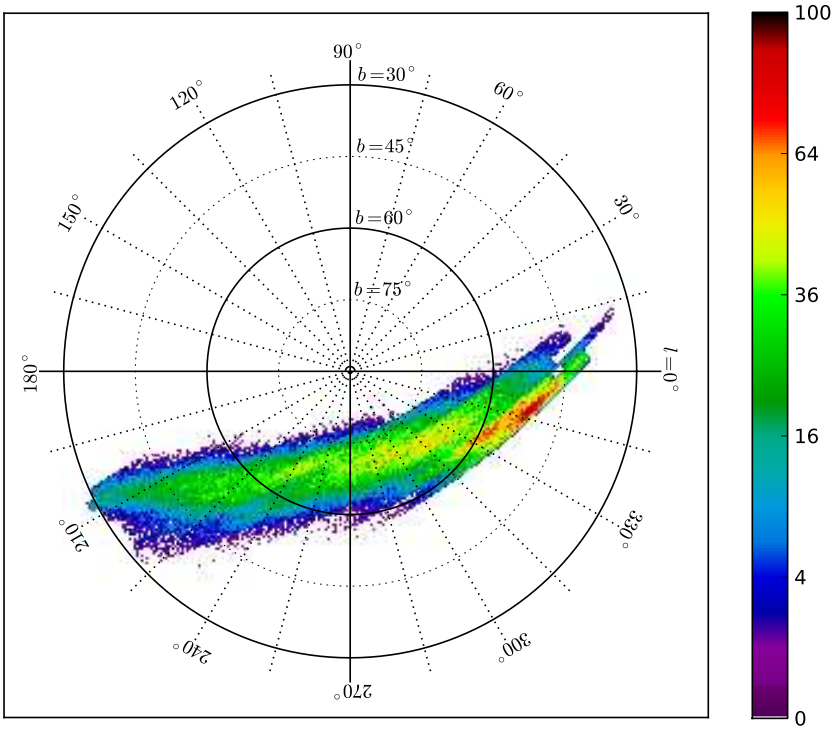


Figure 3.6: Polar plot, similar to Figure 3.5, but with only the separated Sgr Stream shown. The separated stream has a smooth, continuous distribution, despite having been fit to 15 separate SDSS stripes.

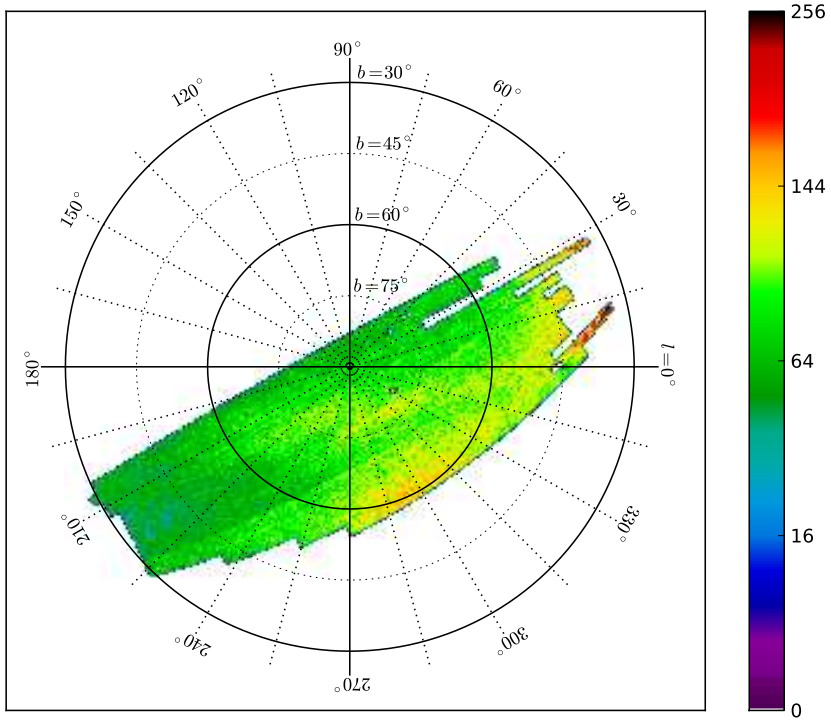


Figure 3.7: Polar plot, similar to Figure 3.5, but with the Sagittarius Stream removed. The bifurcated piece is clearly visible near the center of the plot, and the Virgo over-density is visible near the bottom of the data. The Sgr Stream has cleanly been removed, leaving only Virgo, the Sgr bifurcated piece, and a smooth background.

with a correlation value of 0.019.

A separate plane was fit to the three trailing stream centers, which resulted in a best-fitting plane to Sgr South with the following equation:

$$0.032X + 0.987Y + 0.158Z - 1.073 = 0 \tag{3.3}$$

with correlation value 0.011.

A convention was chosen such that the Z-component of the plane normals (which are given by the coefficients of X, Y and Z in the above equations) is positive. The normals to the two planes of equations 3.2 and 3.3 are approximately 15.6° (±0.1°) apart. The leading and trailing tails do not share the same orbital plane, and the Sgr core is found -3.5 (±0.7) kpc away from the leading orbital plane (that is, opposite the plane normal) and 0.7 (±1.0) kpc from the trailing orbital plane. The Galactic center lies -1.0 (±0.3) kpc from the leading plane, and -1.1 (±0.5) kpc from the trailing plane. The Sgr core is in better agreement with the trailing tail orbital plane than the leading tail orbital plane, which is expected, as the Sgr core and trailing tail are both located below the disk, and the trailing tail debris is closer to the dwarf and therefore has not yet evolved in the Galactic potential for as long as the leading tail. Both orbital planes pass within the same distance of the Galactic Center (∼1 kpc), but do not pass through it.

The fit planes could have been forced to pass through the Sgr dwarf core or the Galactic center, but the aim was to characterize the Sgr tidal debris; forcing the planes to intersect either of these points requires presumptions about the halo potential and stream associations. By ignoring the Sgr core during the plane fitting process, only the planes implied by the analysis are examined, and the association with the Sgr core is discussed below. Also, a study by Koposov et al. (2012) indicates that one or more of the tails and associated bifurcations may not belong to Sgr; if this is the case, then including the core in the plane fits would not be appropriate. If the fits were required to pass through the Galactic center, then a spherical Galactic halo would be presumed. Given that the current best fit dark matter halo is triaxial (Law & Majewski 2010), a symmetric orbit seems unlikely. In the past, the deviations of the leading and trailing tidal streams from a common

Table 3.3: Plane Fit Parameters

Plane	a	b	c	d_{GC} (kpc)	d_{sgr} (kpc)
Sgr North	-0.20±0.02	0.935±0.008	0.29±0.01	-1.0±0.3	-3.5±0.7
Sgr South	0.03±0.05	0.987±0.009	0.16±0.05	-1.1±0.5	0.7±1.0
Bif North (N07)	-0.13±0.05	0.980±0.002	0.15±0.04	-0.9±0.6	-1.5±1.1
Sgr South (K12)	-0.06±0.07	0.97±0.02	0.24±0.04	-1.0±0.7	-1.1±2.0
Bif South (K12)	-0.19±0.06	0.98±0.02	0.10±0.04	-1.0±0.6	-2.2±1.9
Majewski (2003)	-0.064±0.002	0.970±0.008	0.233±0.002	0.23±0.04	0.12±0.07

plane were used as evidence for a non-spherical Galactic potential.

After finding plane fits for Sgr in the North and South Galactic Caps, it is instructive to compare the angles between planes with fits to the bifurcated Sgr pieces. Results from Koposov et al. (2012) were used to get an alternate fit to Sgr South and a fit to Sgr South bifurcation, using sky positions from their Table 1 and distances (corrected with recent Errata) from Table 2 of that paper. Distances were extrapolated using their magnitude gradient (see Figure 5 of that paper), where necessary. Also included in our analysis is the plane fit from Majewski et al. (2003). The best-fit plane parameters are given in Table 3.3, and the angles between respective plane pairs are presented in Table 3.4.

From Table 3.4 it can be seen that, of all the North-South plane pairs, the Sgr North and Sgr South planes from this study are the farthest apart (15.6±0.1), while the two bifurcated pieces (Newberg 2007, Koposov 2012) are the most similar (3.9±2.2). Note that the main Sgr South orbital plane for Koposov et al. (2012) is nearly identical to that of Majewski et al. (2003).

There are three possibilities here: (1) Sgr and the bifurcated pieces are from one dwarf galaxy, being described by multiple wraps, internal dynamics of the Sgr dwarf (rotation, etc.), or significant lumps in the halo; (2) The remains of two dwarfs are present, with some combination of leading-trailing streams describing both; or (3) tidal debris from more than two dwarf galaxies are present.

If (1) is true, and the two bifurcated pieces are from the same orbital wrap, then the low angle between their respective orbital planes is indicative of a symmetric, well-behaved halo potential. However, the large angle between the main Sgr stream orbital planes then implies a strongly asymmetric halo potential; both of these

Table 3.4: Angles Between Planes

Plane	Sgr North °	Sgr South °	Bif North (N07) °	Sgr South (K12) °	Bif South (K12) °	Majewski (2003) °
Sgr North	···	15.6±0.10	9.4±1.7	8.8±3.8	10.9±2.2	8.7±4.3
Sgr South	15.6±0.10	···	9.6±0.2	7.0±4.6	13.0±0.7	7.0±3.1
Bif North (N07)	9.4±1.7	9.6±0.2	···	6.5±3.0	3.9±2.2	6.3±7.6
Sgr South (K12)	8.8±3.8	7.0±4.6	6.5±3.0	···	10.4±0.3	0.2±1.3
Bif South (K12)	10.9±2.2	13.0±0.7	3.9±2.2	10.4±0.3	···	10.2±3.2
Majewski (2003)	8.7±4.3	7.0±3.1	6.3±7.6	0.2±1.3	10.2±3.2	···

cannot be true. Dynamics internal to the dwarf are still a possibility, but would have to describe a bifurcated piece that appears on the same side of the main stream both above and below the Galactic plane. Several authors (Fellhauer et al. 2006; Law et al. 2009; Peñarrubia et al. 2010; Law & Majewski 2010) have proposed mechanisms that would allow the Sgr streams to be a single continuous object, but no single theory has been able to completely explain all of the data.

If there are two dwarf progenitors (case 2), then minimizing the difference between two orbital plane angles is not desirable, but instead the two North-South orbital plane pairs that have the most consistent (similar) angle differences and distances from the Sgr core are important. Through inspection of Sgr core distances from Table 3.3 and plane-plane angles from Table 3.4, the best way to pair the streams, assuming two dwarfs, is (a) Sgr-North and Bifurcated South; and (b) Sgr-South and Bifurcated North. The angles between these stripes pairs are 10.9° and 9.6°, respectively. Interestingly, these angles match the the difference in North/South orbital poles in 2MASS data (10.4°$\pm$ 2.6°) (Johnston et al. 2005). The planes for Sgr North/Bifurcated South both pass over 2 kpc away from the Sgr core (-3.5 and -2.2 kpc, respectively) while Sgr South/Bifurcated North intersect (within errors) the Sgr core itself. This seems to imply that the two more prominent tails, that have traditionally been assumed to be Sgr dwarf tidal debris, may in actuality be from two separate streams.

If there are more than two dwarfs (case 3), then it is difficult to say anything about the system without having a complete understanding of the stream density distribution. Currently, this information is unavailable, as the plane of the disk (in the direction of the Galactic center) blocks our line of sight to where these dwarfs would be discernible as separate entities. This description seems to add more complexity than is needed to describe the problem; however, case 1 either does not contain enough details to describe the current situation, or adds theoretical complexities (in the form of internal dynamics or dark matter distribution) that have yet to be solved satisfactorily. From the distances to the dwarf in Table 3.3, it seems more likely that the Sgr dwarf is associated with the Sgr South stream and the bifurcated piece of Sgr North.

In summary, our maximum-likelihood optimizations, applied to SDSS turnoff star data, serve as strong constraints for any future models of the Sgr tidal debris system. We provide the best-fit stream parameters found by our algorithm, and a catalog of SDSS turnoff stars consistent with the density distribution of the main Sgr stream in the Galactic North, which will serve as a useful tool for future authors when studying the Sgr debris and models of the stellar halo.

3.4 Properties of Sgr as a Function of Angle Along the Sgr Stream

The tidal streams have a width and density as a function of angle along the progenitor, which can be derived from our stream fitting results . The angle along the stream is given in the Sun-centered $\Lambda_\odot, B_\odot$ coordinate system used in Law & Majewski (2010), in a right-handed (XYZ) Galactic coordinate system. The $\Lambda_\odot = 0°$ reference is the Sgr Dwarf galaxy, and increases towards the Galactic South (towards the trailing tail, or opposite the orbital direction of Sgr). Therefore $\Lambda_\odot = 180°$ is in the opposite direction from the Sgr dSph as viewed from the Sun. The leading (North) stream is traveling from 360° towards 180°, while the trailing (South) stream is moving towards 0°.

The stellar density along the stream is defined as the number of Sgr tidal stream stars per degree of $\Lambda_\odot$ in the sky. To get this number, the Northern Sgr stream was separated from the SDSS data (Figure 3.6), then this subset was sliced into 15 (the number of SDSS data stripes used) equal-width wedges, in $\Lambda_\odot$, $7.65\bar{3}$ degrees wide. The number of stars in each wedge were then counted, and divided by the degree width to get counts per degree.

However, these counts do not represent the true number of F turnoff stars, as stars are missing from the data due to the SDSS detection efficiency (Newberg et al. 2002) and in places where the Sgr tidal stream is not fully contained within the volume of data. To compensate for this, a Monte-Carlo simulation method was used to infer the number of missing stars in the stripe. Valid stream stars were randomly generated given the best-fit parameters for that stripe, as per Section 3 of Cole et al. (2008), and then tested to see if the star would have been detected

by SDSS, given the F turnoff detection efficiency and stripe boundaries. If the star would have been detected by SDSS, it was flagged as "detected" and added to the dataset; otherwise, the star would be added to the dataset and not flagged. Any star outside of the ν bounds ($-1.25° < \nu < 1.25°$) was rejected outright to prevent overlap with adjacent stripes. When the number of "detected" stars was equal to the number of stars belonging to Sgr in that stripe (Table 3.1), the process was stopped. This process resulted in a catalog of stars with a density consistent with that of the Sgr tidal stream in a given stripe, but with additional stars correcting for losses due to the stripe bounds and the detection efficiency. After the stream in each stripe was corrected in this way, the simulated data was merged into one dataset and then sliced into $\Lambda_\odot$ bins and divided into wedges, as described above.

Since the great circle of $B_\odot = 0°$ (that is, the Sgr orbital plane given by Law & Majewski, 2010) does not intersect the SDSS data stripes at a 90° angle, there will be large volumes of data missing from the $\Lambda_\odot$ bins located at the edge of the data. The $\Lambda_\odot$ stripes that are suspected to contain this bias were flagged, and their densities plotted using open points.

The SDSS South stripes used in this stage of the analysis were not contiguous, and so their densities were approximated by dividing the star counts by the length of the Sgr stream in that stripe. The missing stars were then reconstructed, as above, and divided by the same length to get the expected density. If the density of the stream is not varying rapidly near the location of the SDSS detection, then it is expected that this method will provide a fairly accurate description of the stream density in that data stripe.

Although it is questionable to compare the density of stars along the stream with an N-body simulation, presumed to represent primarily dark matter particles, it is nonetheless interesting to compare the stream fitting results to the densities expected from four Sgr dSph tidal disruption models created by Law & Majewski (2010); Law et al. (2005). These models are N-body simulations that best describe the Sgr tidal stream data using different gravitational potentials for the Galactic halo: tri-axial, non-axisymmetric (Law & Majewski 2010), spherical ($q = 1$), oblate ($q < 1$), and prolate($q > 1$) (Law et al. 2005). Masses were selected from each model

by "Pcol" (which indicates the orbit on which a given particle became unbound from the simulated satellite) and "Lmflag" (which is positive for leading tail debris, negative for trailing debris). Only stars with Pcol< 3 (corresponding to the first wrap only) and $160.0 < \Lambda_\odot < 185.0$ were selected. The latter condition is because, in these models, the leading and trailing tails each wrap more than 180° around the galaxy, and so each tail will extend into the opposite respective Galactic hemisphere, creating a apparent overlap in angular density near the Galactic anti-center. Since only the regions of the streams closest to the Sgr core would have been characterized by the stream fitting analysis, it is necessary that only stream stars located in their respective Galactic cap be selected. For the tri-axial model, stars were selected with "Lmflag"> 0 and $185.0 < \Lambda_\odot < 360.0$ to isolate the leading (Northern) tail, and "Lmflag"< 0, $0.0 < \Lambda_\odot < 160.0$ to select only the trailing (Southern) tail. For the axisymmetric models, cuts were made in $\Lambda_\odot, W$ (z-axis velocity) in order to separate the primary leading and trailing tails into their respective hemispheres.

Using the data from these models, the simulated streams were sliced into 1.0° $\Lambda_\odot$ bins and the number of stars in each bin were counted, thereby obtaining the stellar density along the stream. These densities are plotted as lines in Figure 3.8 (top pane), multiplied by an arbitrary factor to bring them in line with the observed densities (since the number of bodies in the simulations does not match the number of stars in the actual stream).

It can be seen from Figure 3.8 that none of the models are a good match to the data. This is not too surprising, since the N-body models trace the path of all mass originally associated with the the Sagittarius dwarf, including dark matter, while this analysis only traces the luminous matter (specifically, F turnoff stars, which are plentiful and can be assumed to be well-mixed with other stars in the dwarf). The results from the leading tidal tail, then, suggest that mass does not follow light exactly within the disrupted dwarf system. However, it is worth noting that the density of stars along the smaller Orphan Stream can be well-fit with N-body models (Newberg et al. 2010). If one were interested in developing a function for the number of stars in the tidal stream for a given total mass per $\Lambda_\odot$ (so as to make the N-body simulations directly comparable to the data), the tri-axial and prolate

model densities would require the least amount of correction.

The densities in the trailing tail tell an entirely different story: only one of the three points matches any of the four models, and the remaining two points show a completely separate trend. Interestingly, while this analysis shows the Sgr South stream to be the most consistent with the Sgr dwarf, the density data from the South is not consistent with the best Sgr dwarf disruption models. There are three possibilities for this discrepancy: (1) The N-body models are somehow in error; (2) The above approximation for single-stripe density is incorrect; or (3) the dominant stream in the South Galactic Cap is not from Sagittarius. Note that possibility (3) is the opposite of the conclusion reached from the plane fits of the tidal tails. The plane fits favor an association between the Sgr dwarf and the dominant trailing tidal tail, as is seen in 2MASS data (Majewski et al. 2003). In order to rule out possibility (2), the SDSS South data will need to be analyzed in more detail. This is possible with the addition of SDSS Data Release 8 (DR8), which fills in large parts of the Southern Galactic cap, and will be the topic of a future paper. If in fact the leading tidal tail is not composed of stars from the Sgr dwarf galaxy, then it would be natural for the N-body simulations to be in error, since they were optimized using the leading tidal tail stars.

The bottom pane of Figure 3.8 displays the stream FWHM (full-width at half-maximum) as a function of $\Lambda_\odot$, as well as the stream widths from Law & Majewski (2010); Law et al. (2005). The FWHMs and errorbars are derived by multiplying the σ stream width values and errors (Table 3.1) by 2.35. The N-body model stars were sliced into wedges of size $\Lambda_\odot = 7.65\bar{3}$ (as above). In each $\Lambda_\odot$ wedge, it was assumed that each star had a $\Lambda_\odot$ equal to the middle value of the wedge, thereby creating a 2D plane of data in $R, B_\odot$. A 2D Cartesian system is then defined in this plane, with $\hat{x}$ being along the line-of-sight, and $\hat{y}$ being the perpendicular direction, such that $x = R\cos B_\odot$ and $y = R\sin B_\odot$. The x and y standard deviations were then determined for each wedge, and the geometric mean of these values then gives the cylindrical stream width: $\sigma_{xy} = \sqrt[2]{\sigma_x \cdot \sigma_y}$. Note that the N-body stream was not cylindrical - the x (or radial) standard deviations were almost always larger than the y standard deviations. The stream profile, then, is more elliptical in the simulations

than the assumed circular profile. Even if this is the case in the actual Sgr tidal stream, the assumed-cylindrical σ fit from the algorithm should be analogous to σ_{xy}, and therefore it is valid to compare the two.

It is seen from Figure 3.8 that all four of the N-body models follow roughly the same width trend in the leading tidal tail as the best-fit widths. The stream widths in the South are also consistent with the model, although the actual data appears to be a bit flatter than the model. The stream widths are raw parameters from the data fits, so these are expected to be accurate. Again, more information about the South is required before an accurate comparison can be made, but at this point, the widths appear to be in good agreement with both the Law & Majewski (2010) model and the spherical potential model from Law et al. (2005), for both the dominant leading and dominant trailing tidal tails. The oblate and prolate potentials produce width profiles that are a worse fit to the data, but still within the given errors.

3.5 Fit Parameters for the Spheroid Star Density

Since spheroid parameters were measured for each SDSS stripe individually, there are now 18 measurements of these parameters. If the Hernquist profile is a good fit to the smooth stellar density in the halo (which may not be true; see below), then the parameters should be similar for each of the stripes.

The Hernquist q, r_0 parameter fits to the SDSS stripe data are presented in Table 3.5. It is found that the stellar spheroid is clearly oblate, as the q value for every stripe is less than 1.0 (spherical). The Northern Galactic cap has very consistent q values, with a mean value of 0.53 and a standard deviation in the fit values of 0.06. There is no apparent trend in q with stripe number in the Northern cap, and the errors in the fits are larger than the average difference in q fits, and so the halo flattening has been well-fit in the North.

The South Galactic Cap exhibits a broader range of q fit values, despite consisting of only three stripes. The Southern cap contains an apparent decrease in q with increasing angle away from the Sgr core, although the errors in these values place them within 2σ of the Northern values, so this may not be significant. The average and standard deviation of the Southern q values are 0.48 and 0.12, respec-

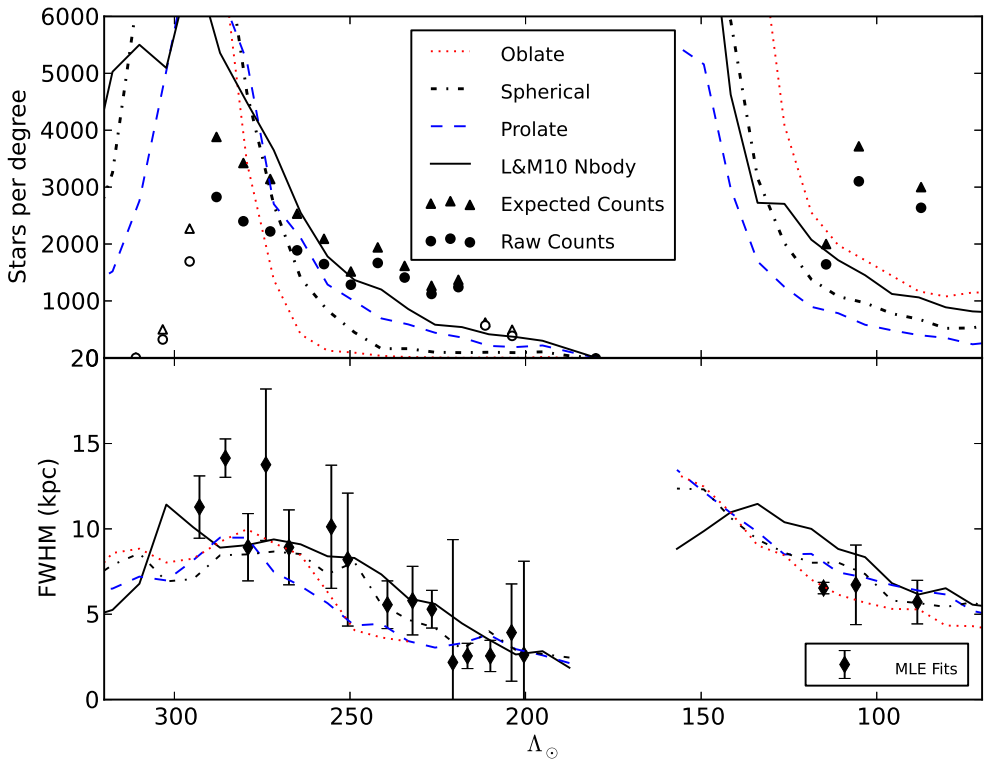


Figure 3.8: Sagittarius stream density (top) and width (bottom) versus $\Lambda_\odot$. In the Sgr stream density plot (top), square points represent the density as given by the number of F turnoff stars observed per $\Lambda_\odot$ degree in the SDSS data, while triangles show the corrected density, accounting for the SDSS detection efficiency and edge effects. Open symbols represent points that may be missing counts due to edge effects (see text). Typical errors are on the order of the point sizes, while maximum errors are roughly twice the point sizes. The over-plotted lines represent the star densities from best-fitting N-body simulations in various halo potentials: Solid black line: triaxial halo (Law & Majewski 2010); Red dotted line: axisymmetric oblate halo; Black dotted line: spherical halo; Blue dotted line: axisymmetric prolate halo. The latter three models are from Law et al. (2005). All N-body densities are scaled by an arbitrary factor (1.8, 1.8, 2.0, and 2.4, respectively). It is clear the the Southern data does not match any of the theoretical models. The lower plot shows the width (FWHM) of the Sgr stream versus $\Lambda_\odot$. The diamond symbols and associated error bars are the stream widths and errors as given by the maximum likelihood algorithm in this paper. The solid line represents widths derived from the Law & Majewski (2010) triaxial N-body model, using "Pcol"< 3 and "Lmflag" > 0 for leading (Northern) tail, or "Lmflag" < 0 for the trailing (Southern) tail (see text). Note that while one could imagine a reasonable variation in the mass-to-light ratio along the stream that would match the data to the model for $\Lambda_\odot > 200$ (the North Galactic Cap), the data in the South ($\Lambda_\odot \sim 100$) do not fit the model.

tively. If the Southern and Northern cap q's are included together, the average value of q becomes 0.52, with a standard deviation in the fit values of 0.08.

The Hernquist profile scale length, r_0, varies widely in the parameter fits, varying between $1.84 < r_0 < 11.55$ kpc in the Northern cap, and between $1.84 < r_0 < 25.95$ kpc over all stripes. The model errors in the r_0 fits are also large (on the order of ∼50% in the North). This is partially due to fact that the algorithm is not as sensitive to changes in r_0 as it is for the other parameters (Cole et al. 2008). There is no consistent trend with angle along the stream in the North Galactic Cap, and so the variations are most likely due to random errors in the data, or the lack of sensitivity of the algorithm. Within errors, the Northern r_0 values are consistent, with an average value of 6.73 kpc and a standard deviation of 2.51 kpc.

The three Southern stripes appear to have an increasing trend in r_0 with angle away from the Sgr core. However, the large errors on these values imply that this correlation is a weak one. That the Southern values for r_0 are significantly larger than the Northern values (even after accounting for errors) implies that there is a difference in the stellar spheroid between the North and the South. This could be due to the presence of the Hercules-Aquila cloud, which is present in SDSS South but not accounted for in this analysis, or this may be additional evidence that the smooth stellar spheroid is non-uniform, or this could indicate that a Hernquist profile is not a good fit to the smooth component. The mean and standard deviation of r_0 in the Southern stripes is 20.71 and 3.89 kpc, respectively. If the Northern and Southern stripes are taken together, the average of r_0 becomes 9.06 kpc, and the standard deviation rises to 5.91 kpc.

It is quite possible that the Hernquist profile is not a good descriptor of the Galactic halo stellar density. The insensitivity of the algorithm to the fit parameter r_0 is possible evidence of this. More problematic, however, is the fact that the "garbage collection" stream (Section 3.2) always converged to a set of parameters that would collect stars at the edge of each data set, implying that the Hernquist profile by itself did not adequately describe the smooth stellar halo. Note that the algorithm did not have problems when fitting the Hernquist profile to simulated data (Cole et al. 2008), which strongly implies that the real data is not well-fit by

Table 3.5: Spheroid Parameters

Stripe	q	r_0 (kpc)
9	0.55 ± 0.17	1.84 ± 2.97
10	0.49 ± 0.13	9.82 ± 5.04
11	0.56 ± 0.13	4.65 ± 2.15
12	0.55 ± 0.10	6.94 ± 2.50
13	0.52 ± 0.11	6.08 ± 4.75
14	0.57 ± 0.11	7.15 ± 4.09
15	0.56 ± 0.09	8.59 ± 4.40
16	0.58 ± 0.13	6.22 ± 4.29
17	0.59 ± 0.13	5.37 ± 5.80
18	0.57 ± 0.12	7.32 ± 4.30
19	0.52 ± 0.12	5.99 ± 5.59
20	0.54 ± 0.11	10.14 ± 4.82
21	0.53 ± 0.14	6.45 ± 3.60
22	0.31 ± 0.16	2.88 ± 4.88
23	0.55 ± 0.07	11.55 ± 3.68
79	0.34 ± 0.08	25.95 ± 9.20
82	0.46 ± 0.10	19.53 ± 7.01
86	0.63 ± 0.14	16.66 ± 5.42

the Hernquist profile. Since no "garbage collection" stream was fit to stripes in the South, it is possible that the discrepant q and r_0 values found there are due to the smooth component having to compensate for the density that would have been accounted for by that stream. The additional unknown, here, is whether the selected F turnoff stars, particularly those near the magnitude limit of the survey, are good tracers of the spheroid density. In the near future, the data will be re-analyzed after being corrected for selection effects, due to increasing color errors near the survey limit. The selection effects and proposed corrections to the algorithm are described in Section 6.7.

Note that Cole et al. (2008) showed that the stream parameters were robust to improperly specified smooth stellar halo component models, so even if the Hernquist profile is eventually found to be a bad choice, the stream parameters should still be reliable.

Several similarities can be found in these halo fits with the findings of Jurić et al. (2008). In that paper, the authors used an earlier release of SDSS and photometric parallax methods to fit disk and power-law halo profiles to Milky Way star densities within 30 kpc of the Sun. Although the power-law and Hernquist profiles are not directly comparable, the Jurić et al. (2008) analysis also found a clearly oblate stel-

lar halo, with flattening parameter $q_h = 0.64$, but consistent with $0.5 < q_h < 0.8$, which is in good agreement with this study's flattening value of $q = 0.53$. Jurić et al. (2008) also found that their density profile had difficulty matching the densities at the edge of their data, a problem which manifested itself in the stream fitting analysis through the "garbage collection" stream. This study agrees with the Jurić et al. (2008) analysis that a single density function may not be adequate for describing the Milky Way stellar halo. In a future paper, multiple halo density profiles (including double-halo profiles) will be explored as possible better fits to spheroid data.

In summary, the smooth component fits to the stellar halo are very consistent in the North Galactic Cap, using a Hernquist profile with $q = 0.53$, and $r_0 = 6.73$ kpc. The data in the South is not as consistent, however, due to either additional over-densities (such as Hercules-Aquila), deficiencies in the Hernquist profile, or the fitting technique.

CHAPTER 4
Milkyway@home

4.1 Computational Resources

The maximum-likelihood algorithm requires a large amount of computing power to search through between 8 and 20 dimensions of parameter-space, and so several computing platforms were used in order to produce our results. The foremost of these is the MilkyWay@home[12] BOINC (Berkeley Open Infrastructure for Network Computing, Anderson, Korpela & Walton , 2005) volunteer computing project, which is the subject of this chapter. While MilkyWay@home was still under development, RPI's SUR BlueGene/L was used to compute best likelihoods, and is still used to compute error estimates through the Hessian method. Its 1024 dual-core nodes produced about 6 TeraFLOPS (Trillion Floating Point Operations Per Second) of computing power; however, this system is shared with other researchers, and so its full power was never brought to bear on the problem. For comparison, the MilkyWay@home platform averages $\sim$ 500 TeraFLOPS (or 0.5 PetaFLOPS) of computing power. In addition to these platforms, a trio of in-house servers provides platforms for code testing, development, and small-scale crunching: *M31*, a Dell PowerEdge 2600 with 4x 3.0GHz CPUs; *Fornax*, a Dell PowerEdge 2900, with 8x 3.0 GHz CPUs; and *LMC*, a machine custom-built by Matthew Arsenault, with two 4-core CPUs (8 cores in total) and support for up to 4 GPUS.

Because the 2- and 3-stream fits have very complicated likelihood surfaces, the local, synchronous conjugate gradient descent (CGD) searches would commonly converge to local maxima, or get close to the "right" answer and then become lost in a flat region near the true best parameters. Since a single CGD run could not be guaranteed to find the global maximum, multiple searches for a single stripe, started from different random initial parameter sets, were run simultaneously on the SUR BlueGene/L. Additional runs were then started around the parameter set

Portions of this chapter previously appeared as: Newby, M., Cole, N., Newberg, H. J., et al. 2013, AJ, 145, 163.

[12] http://milkyway.cs.rpi.edu/milkyway/

that produced the best likelihood. This process was repeated until the returned likelihoods stopped improving. To enforce continuity between adjacent stripe fits, the initial starting parameters for each run were set to the average of the most recent results for the adjacent data stripes. This was a time-consuming process, not only due to the computing power necessary (and time spent in the process queue) on the SUR BlueGene/L, but also because the researchers spend time studying the previous results, generating unbiased starting parameter sets, applying continuity requirements, and preparing the next set of runs.

It became clear that more global search techniques would be required to improve the efficiency and accuracy of the maximum likelihood algorithm. MilkyWay@home's global, asynchronous search techniques converge, repeatably, to the best-fit parameters (the global maximum) for a given problem, even though it must navigate the same complicated likelihood surface. MilkyWay@home, then, provides a platform capable of effectively searching complicated likelihood surfaces, and the background and capabilities of the MilkyWay@home project are the subject of this chapter.

4.2 The MilkyWay@home BOINC project

By necessity and design, the MilkyWay@home BOINC project is a highly asynchronous platform running modern numerical optimization techniques on users spare CPU and GPU computer processing power. As of May 2013, MilkyWay@home has around 28,000 active volunteers in over 150 countries donating computer time, for an average aggregate processing power equal to ~ 0.50 petaFLOPS (10^{15} floating operations per second). For comparison, this is about 6 times the power of state-of-the-art BlueGene/Q supercomputers, and places MilkyWay@home on par with the $\sim 50th$ supercomputer on the TOP500 list[13]. MilkyWay@home is one of the most popular and successful projects on BOINC, consistently ranking in the top 5 most popular projects and responsible for 11% of all credits distributed through BOINC[14], where credit is theoretically proportional to the amount of numerical

[13]http://www.top500.org/lists/, June, 2013
[14]http://boincstats.com/

operations performed for a project[15].

The *modus operandi* of MilkyWay@home is to send a "work unit" to a volunteer (client), who processes that work unit then returns the result to the server, where it is integrated into the current likelihood search. When a client requests work from the server, the server will send that client a work unit from its current stack of available work units, known as the "feeder." The server will also send the relevant star file to the user, if it is not already stored on their computer. The client then crunches the work unit and returns the likelihood value of the given parameters (or the result) to the server. If this result would improve the current search, the server then validates (through the "validator") this work unit by matching it to another identical work unit crunched by a different user. Additionally, 10% of other work units are randomly selected for validation, and each time a computer fails validation, its frequency of being selected for validation increases by 10%. If a computer has a validation rate greater than 10% and returns a valid work unit, its validation frequency is reduced (currently, it is multiplied by 0.95). By validating work units in this fashion, we prevent erroneous or fabricated results from contaminating the search, while not wasting computing time in validating every result (Desell et al. 2011). Once validated, the result is assimilated (by the "assimilator") by the current search method, which will influence the next set of generated work units.

This process is automated, and repeats until an admin manually stops the search; a search is considered to have converged when the likelihoods stop improving within a certain threshold (Figure 4.1). For the stream fitting problem, this threshold is usually $\Delta\mathcal{L} \leq 10^{-6}$. A complete summary of the MilkyWay@home server structure can be found in Desell (2009). Detailed BOINC server documentation can be found on the BOINC wiki[16].

Each work unit uses a single set of parameters to integrate the probability density function (Equation 2.31 of Cole, 2009) over a target SDSS stripe. The work unit then returns to the server the likelihood that the data is drawn from the

[15] "Theoretically" because the actual number of flops performed varies by application, and cross-BOINC project standards are difficult to enforce.

[16] http://boinc.berkeley.edu/trac/wiki/ProjectMain. Wikipedia also contains a good summary of the BOINC wiki (accessed July 2, 2013).

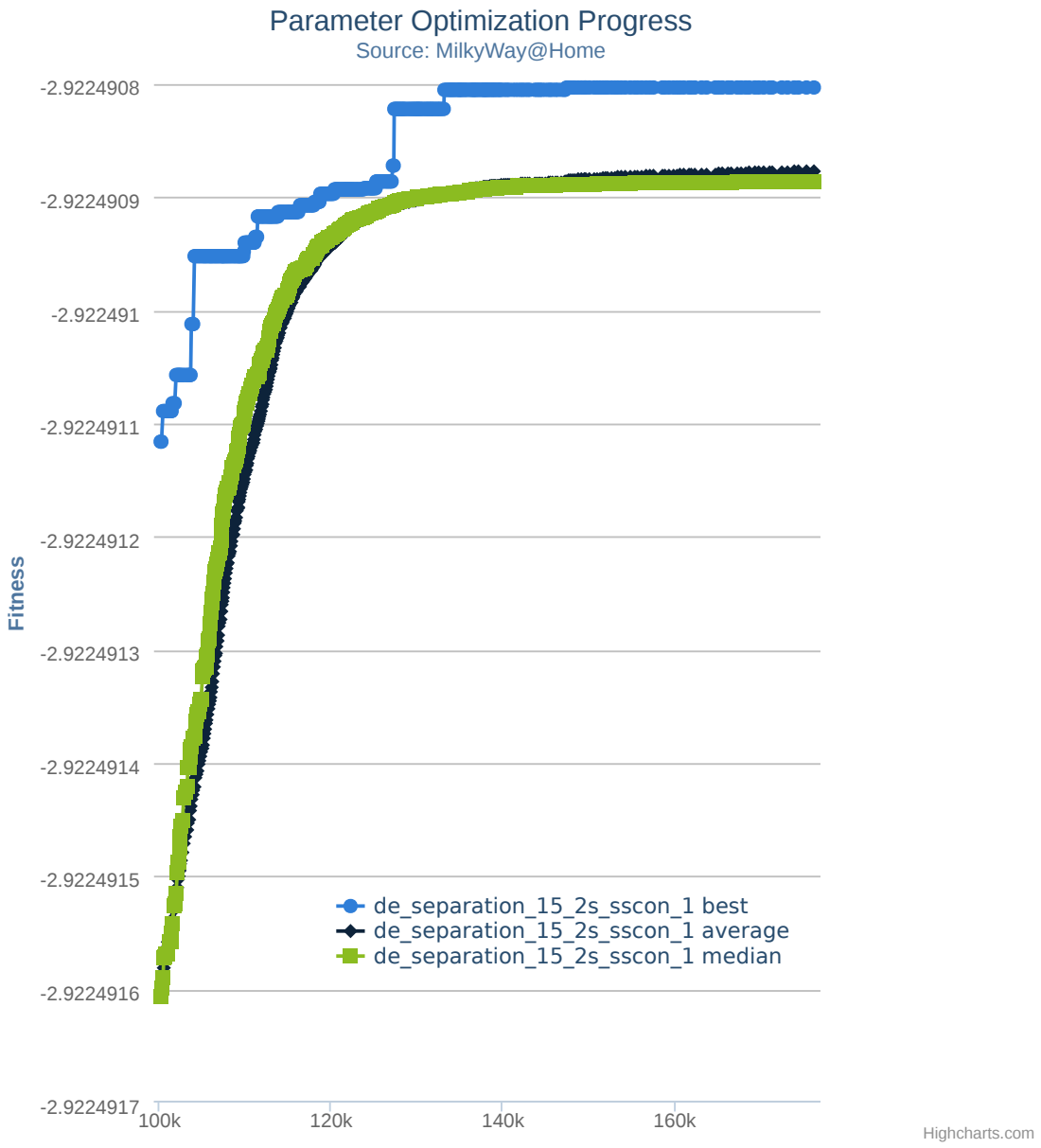


Figure 4.1: Plot of MilkyWay@home likelihood progress for a 2-stream differential evolution search, using data from SDSS stripe 15. Along the x-axis is the number of valid particles (unique parameter sets) that have been returned to the server, with particle likelihood on the y-axis. This search's unique identifier (or "'name") is "de_separation_15_2s_sscon_1," which describes the properties of the search; this search is a differential evolution search, using the separation method on SDSS stripe 15, using two streams in the model fit, and the remaining information is a unique tag that differentiates similar searches. The "best" (blue), "average" (dark blue) and "median" (green) curves reference the likelihood statistics of the current population, after the newest particle has been assimilated. The best likelihood here has not improved to within 10^{-8} for the last 20,000 particles, and so this search has clearly converged.

model, given the input parameters. Each user receives an estimated completion time for each downloaded work unit, calculated from the total number of floating point operations required to complete a work unit, multiplied by a hardware specific factor (derived from hardware benchmarks). The amount of credit given for a completed work unit is proportional to the number of floating point operations required to complete a work unit, and is directly related to the number of volume elements in the PDF integral and the number of points used for the convolution:[17]

$$\text{time} \propto \text{FLOPs} \propto R_{\text{steps}} * \mu_{\text{steps}} * \nu_{\text{steps}} * N_{\text{convolve}} \tag{4.1}$$

Where the subscript "steps" implies the number of steps that coordinate will be subdivided into by the the integral calculation, and N_{convolve} is the number of points used to calculate the convolution.

The integral steps have an important impact on the likelihood searches, with small numbers of steps producing less accurate integrals, but in a shorter amount of time. Also, if the number of total steps in the integral is too small, users with high-end GPUs will crunch work units quickly enough that their systems are left idle while downloading new work units. Even though these idle periods are usually short, they represent time that could be better spent crunching data and earning BOINC credits; serious users will then disconnect from the Milkway@home project and join other GPU-enabled BOINC projects that provide a more steady source of work and credits. Short work units may also cause too many results to be returned at once and overload the server; however, this has not been an issue since the server hardware upgrade. On the other hand, if there are too many steps in the integral, slower computers will not be able to finish a work unit before it expires. This effect is significant only for very large numbers of integral steps, and experience has shown that it is better to err on the side of larger work units and more accurate integrals.

In May of 2013, it was discovered that larger integral numbers of integral steps increase the number of work units that return errors. By experimenting with different integral sizes and watching error rates and keeping track of the hardware which produced errors, it seems that these large work units may be overflowing the

[17] "FLOPs" here refers floating point operations, *not* floating point operations *per second*.

GPU memory on older computers. The precise source of these errors is currently under investigation.

A standard set of integral step sizes has been found to be ideal for 3-stream searches on SDSS data stripes with $\sim 100,000$ stars. These searches appear to be most efficient when 1400 μ steps, 1600 r steps, 320 ν steps, and 120 convolution points are used, regardless of a stripe's physical extent. These step numbers correspond to a number of floating-point operations on the order of 10^9 per likelihood calculation, which grants the users ~ 80 BOINC credits per work unit completed. Since each stream represents an additional set of density calculations for the algorithm, searches with fewer than 3 streams should have one of their integral step numbers increased by 33% to 50% per omitted stream. This will produce work units of approximately the same size as the standard 3 stream searches.

The optimization methods currently implemented on Milkway@home are the particle swarm and differential evolution searches. These searches are well suited to our problem, as they are capable of escaping local maxima and eventually working their way to the true global maximum likelihood, and they are effective in an asynchronous, heterogeneous computing environment. By the virtue of being asynchronous, new parameter sets are not strictly dependent on previously sampled parameters, and so work units do not need to be returned in the order that they were sent out. Returned work units do influence the choice of the next generation of test parameters, and so every work unit, even those crunched on slower, older computers, will improve the search. By supporting a large number of different operating systems and hardware configurations, and by using standard, statically-linked libraries within the client-side application, the MilkyWay@home project will return consistent results across all supported computers. This means that if the same work unit is sent to two drastically different computers, both systems will return the same likelihood to within at least 10^{-8}. A thorough review of the MilkyWay@home server's BOINC setup, optimization methods, and validation techniques can be found in Desell (2009).

4.2.1 Project History

MilkyWay@home launched in 2007 as a collaboration between the Physics and Computer Science departments at RPI, with Heidi Newberg, Malik Magdon-Ismail, Bolek Symanski, Carlos Varela, Travis Desell, and Nathan Cole as the major developers, adopting the maximum likelihood algorithm that had begun development under Heidi Newberg, Malik Magdon-Ismail, Joe Doran, Jonathon Purnell, and Nathan Cole (Cole et al. 2008).

MilkyWay@home was one of the first BOINC projects to implement a GPU version of its application, and the likelihood integral is well-suited to the highly-parallel architecture of GPUs. GPUs provide significant performance improvements over CPUs (Desell et al. 2009); CPUs are capable of 0.25 to 7.5 GigaFLOPS, while cutting-edge GPUs are capable of over 2 TeraFLOPS of processing power when applied to parallelized problems. This is roughly a factor of 100 increase in performance. Actual runs on MilkyWay@home confirm this; a CPU may take over 5 hours to complete an single likelihood calculation, while a top-of-the-line GPU may take less than a minute to complete the same calculation.

The GPU code for the stream fitting problem was originally developed by a volunteer in Germany whose username was "Cluster Physik," and was released on MilkyWay@home by computer science graduate student Anthony Waters. The source code for this version of the code was lost when both of these developers moved on to other projects, but new code was subsequently written by Matthew Arsenault. Matthew Arsenault later worked with Ben Willett to produce the prototype for a new N-body application that would run side-by-side with the the integral calculation code on MilkyWay@home; the integral calculation code became the "separation"[18]code, since the final goal is to separate the streams from the smooth spheroid, and the new code became known as the "N-body" code. A series of students continue to improve the N-body codebase (Table 4.1), which is still under development at of the writing of this thesis.

[18]The separation process actually occurs outside of MilkyWay@home, and so this term is technically incorrect, however, the name has stuck and is now the standard. *Although* the term "stream fit" appears in some places; this is colloquial with "separation." This non-rigorous naming scheme is one of the "tribal" features of a large project with many developers.

Table 4.1: Contributors to MilkyWay@home

Name	Affiliation	Position	Contribution
Heidi Jo Newberg	RPI[1], Physics	Professor	PI
Malik Magdon-Ismail	RPI, CS[2]	Professor	PI
Carlos Varela	RPI, CS	Professor	PI
Boleslaw Szymanski	RPI, CS	Professor	PI
Travis Desell	RPI; UND[3], CS	Grad Student/Post-Doc;	Server, Search Methods,
...	...	Professor	Separation, N-body
Matthew Arsenault	RPI, Physics/CS	Undergrad,	Server, GPU, Separation,
...	...	Grad Student	N-body, Visualization
Jonathon Purnell	RPI, CS	Grad Student	Separation
Joe Doran	RPI, CS	Grad Student	Separation
Nathan Cole	RPI, CS/Physics	REU[4]/Grad Student	Separation
Anthony Waters	RPI, CS	Grad Student	Separation, GPU
Cluster Physik[5]	Independent	Volunteer	Separation, GPU
Matthew Newby	RPI, Physics	Grad Student	Separation, Server
Benjamin Willett	RPI, Physics	Grad Student	N-body
Jeffery Thompson	RPI, Physics	Grad Student	Separation, N-body
Torrin Bechtel	RPI, Physics	Undergraduate	Separation
John Vickers	RPI, Physics	Undergraduate	Separation
Jake Weiss	RPI, Physics	Undergraduate	Separation
Dave Przybylo	RPI, CS	Undergraduate	Separation
Eric Wyler	RPI, CS	Undergraduate	N-body
Colin Rice	RPI, Physics/CS	Undergraduate	N-body, Server
Steve Ulin	RPI, Physics	Undergraduate	N-body
Kegham Khosdegian	RPI, CS	Undergraduate	N-body
Jake Bauer	RPI, Physics	Undergraduate	N-body
Roland Judd	RPI, Physics	Undergraduate	N-body, GPU
Adam Susser	RPI, Physics	Undergraduate	Visualization
Liam Moynihan	RPI, Physics	Undergraduate	Visualization
Shane Reilly	RPI, CS	Undergraduate	Visualization
Brian Chitester	RPI, CS	Undergraduate	Web Design

[1]Rensselaer Polytechnic Institute

[2]Computer Science

[3]University of North Dakota

[4]Research Experience for Undergraduates student

[5]MilkyWay@home volunteer, Cluster Physik is a username.

Each stream-fitting search is given a unique and meaningful identifier, in order to prevent confusion with similar runs. A typical name will be of the form "(search method)_(search type)_(stripe number)_(number of streams)_(identifier)_(iteration)". For example, "de_separation_15_2s_sscon_1" (Figure 4.1): the "de" indicates that this is a differential evolution search; "separation" implies the use of the separation algorithm; "15" states that SDSS stripe 15 is being used; "2s" indicates that this is a two-stream search; "sscon" is a tag that indicates other information about the run (in this case, a sans-Sgr data set is in use, with some constraints applied to the parameter file); and the "1" denotes that this is the first search of this type. In theory,

only one run should be needed per data stripe, but in practice, code or user errors sometimes require a search to be restarted. An N-body naming scheme has not yet been standardized, but are generally of the form "(search method)_nbody_(other information)".

4.2.2 Hardware

The MilkyWay@home server is currently a Dell PowerEdge R710 blade server, located in the secure server room beneath the Voorhees Computing Center (VCC) at RPI. The operating system is the long-term support (LTS) version of Ubuntu server 11.10 (GNU/Linux 3.0.0-21-server x86_64). This machine has a Intel Xeon X5647, 2.93 GHz, quad-core CPU, 64GB of Memory (RAM), and four 500GB hard drives. This setup allows for the entire ∼60GB MilkyWay@home project database to be loaded directly into memory, resulting in very smooth and fast data turnaround to the volunteers.

This is the second server to house MilkyWay@home; the previous server was not able to keep up with the increased load placed on it, especially once the GPU code became popular. The old server had only 8GB of memory, and so would have to read/write to the hard drives frequently. This meant that if a large number of users returned results at the same time, the server would be to busy processing them to create new work, and would occasionally put such a load on the server that it would crash. Thankfully, the current server does not suffer from these shortcomings.

The hard drives on the current server are placed in a mirror-writing RAID-array, so that if one drive fails, another drive with an identical copy of the data will exist. In the event of a failure, the server will seamlessly switch to the good drive, and the failed drive can be replaced without requiring a server shutdown. This was also an improvement over the old server, which was backed up to magnetic tape; when a drive failed, the system had to be taken down while a new one was inserted, then had to be rebuilt from the backup, which meant the server had to be down for several hours.

4.2.3 The MilkyWay@home Codebase

The MilkyWay@home codebase is divided into two main parts: the server code and the client code. The server code contains all of the functions and processes that run on the MilkyWay@home server, while the client code contains the applications that run on the volunteer computers. The separation and N-body codes are grouped together inside of the client code, since they both call many of the same BOINC client functions. All of the code is open-source, version-controlled using Git, and stored on Github[19]. The server code can be found on Github under the project name "milkyway_server"[20], and the client code can be found under "milkywayathome_client"[21].

The server code is dependent on Travis Desell's Toolkit for Asynchronous Optimization (TAO), which is a collection of methods for solving maximum likelihood problems (Desell et al. 2013). MilkyWay@home uses the asynchronous particle swarm (PS) and differential evolution (DE) methods, which are detailed in Chapter 5 of Desell (2009). These two techniques will return the same parameters for the best likelihood, but occasionally one method will converge more quickly than the other for a given data stripe. Also, with two separate search methods available, we can double-check results to ensure that a given result is not due to a technicality with a search method. The TAO source code can be found on Github under the project name "travisdesell/tao"[22].

The majority of the code is written in the C programming language, and is built using the CMake cross-platform build system. The GPU code is an OpenCL (Open Computing Language) wrapper that modifies the standard C code to run on GPU hardware. Currently, only the separation code has a production level GPU version, while a GPU-capable N-body application is currently under development.

Readying the project code for release to the volunteers requires building, testing, signing, and release. The code must be built separately for each platform that is supported: 32-bit versions of Windows and Linux binaries, and 64-bit versions of

[19] www.github.com
[20] https://github.com/Milkyway-at-home/milkyway_server
[21] https://github.com/Milkyway-at-home/milkywayathome_client
[22] https://github.com/travisdesell/tao

Windows, MacOS, and Linux binaries, for a total of 5 applications for a single CPU release. A GPU release requires the same 5 operating system releases per GPU manufacturer, and since Nvidia and AMD cards are both supported, this means that 15 applications must be compiled for each update to the separation code (5 for CPU + 10 for GPU)[23]. After the binaries are successfully linked to BOINC and built, they must be tested. Currently, virtual machines are installed on LMC that simulate each platform, and dummy jobs are sent to them to check that the compiled code is operating properly[24]. Once this is done, the applications must be cryptographically signed in order to make them secure, then they can be released (through a process on the MilkyWay@home server) as an update.

4.2.4 The MilkyWay@home Community

The users of MilkyWay@home are key to the current and future success of the project. Many users are drawn to MilkyWay@home because of their interest in astronomy or science, but a large number of the volunteers are computer hardware enthusiasts looking for beefy computational tasks with which they can benchmark their machines. These users delight in building top-of-the-line, multi-GPU, multi-CPU computers, then optimizing them for computational speed. MilkyWay@home gives them an opportunity to benchmark these machines in fair arena; they can easily compare the performance of their machines with others' by looking at the earned credit statistics. These "hot-rodders" represent a significant fraction of the project's computing power.

Volunteers who sign up because they are interested in the science involved not only contribute valuable computational time to the project, but also provide the MilkyWay@home team with the opportunity to increase public understanding of astronomy research. One of the many difficulties with advanced research is that it is often difficult to communicate scientific findings to the general public. With MilkyWay@home, we are able to keep our users informed through news posts on the

[23]The current build system documentation can be found here (requires log-on): http://milkyway.cs.rpi.edu/milkyway_ops/MilkyWayBuildEnvironment.html

[24]The current test environment documentation can be found here (requires log-on): http://milkyway.cs.rpi.edu/milkyway_ops/MilkyWayTestEnvironment.html

project webpage, and a science summary page[25]. Additionally, users are encouraged to ask science questions and discuss scientific topics in our community forums.

The community forums also serve as a vital diagnostic and feedback tool for the project. Users who are having trouble with specific platforms or work units will post about their issue, which not only brings the issue to the attention of the team but also provides a starting point for troubleshooting. Several bugs in early releases of the code were eliminated thanks to user communication. Also, since the team is unable to test every new release on *every* possible combination of hardware and software, the users will use the forums to discuss workarounds and resolve issues in special cases. Indeed, many of the computer-savvy users on the project will resolve issues faster than the science team is able to respond - in a way, the users are their own tech support group.

The users will also use the forums to police each other, pointing out other users who are exhibiting unusual behavior. This was especially useful when MilkyWay@home's validation system was not as stringent, and several users took to spoofing results to get "free" credits. Any user proven to not be "playing fair" is banned from the project, and their credit totals deleted. Users have also been useful in pointing out hosts that produce only errors, often due to out-of-date software or obsolete hardware.

MilkyWay@home has a large number of users located across the entire world (Newberg 2013b). Most of these users are concentrated in industrialized and English-speaking nations (Figure 4.2), which makes sense, as by definition our users require access to a computer, and most of MilkyWay@home's documentation is in English. Even so, several countries (notably Germany, France, Poland, Czech Republic, and Russia) have large, robust virtual communities which allow even non-native English speakers to enjoy participating in the project.

4.3 Using MilkyWay@home To Verify Sgr Stream Fits

Once MilkyWay@home became mature, it was applied to the maximum-likelihood stream-fitting algorithm, using the 15 SDSS stripes studied in Chapter 3. The

[25]http://milkyway.cs.rpi.edu/milkyway/science.php

Rank	Country name	Total credit	Credit /day	Credit /week	Credit /month	Average credit
1	United States	51,810,620,201	33,022,589	200,622,579	811,922,941	28,466,211
2	Germany	17,758,438,407	8,217,511	57,790,612	222,212,204	7,926,613
3	United Kingdom	9,805,327,315	6,614,409	44,172,165	181,062,651	6,383,875
4	France	9,229,098,529	3,107,190	20,237,003	76,636,036	2,845,811
5	Canada	7,519,437,097	3,373,790	22,882,383	155,389,780	4,239,384
6	International	7,292,426,944	1,573,751	18,199,976	79,768,907	2,798,292
7	Poland	7,120,045,624	2,834,302	18,191,028	74,939,622	2,615,080
8	Japan	7,118,753,347	4,041,077	29,165,066	105,904,971	3,868,174
9	Australia	6,395,804,846	4,811,643	28,550,211	99,015,171	3,956,531
10	Russian Federation	5,544,294,659	4,222,526	25,070,789	112,655,613	3,861,230
11	Norway	4,740,620,332	1,545,809	9,318,995	49,157,136	1,570,030
12	Czech Republic	4,630,314,775	2,561,669	17,159,675	75,642,417	2,548,404
13	Finland	3,911,341,261	2,891,950	16,574,251	69,826,131	2,434,827
14	Sweden	2,474,276,896	2,187,244	14,528,450	54,397,941	1,982,264
15	Netherlands	2,355,877,918	1,260,904	8,319,962	31,397,684	1,182,852
16	Austria	2,353,048,606	1,092,915	7,642,431	30,685,413	1,112,215
17	Spain	1,756,944,009	275,623	1,824,447	7,096,582	268,931
18	Italy	1,754,817,790	557,795	3,386,712	14,060,014	507,609
19	Switzerland	1,709,906,635	906,714	4,831,964	15,718,069	606,131
20	New Zealand	1,413,495,538	645,962	4,514,104	19,459,547	667,082
21	Belgium	1,099,480,546	638,077	3,593,731	16,983,248	548,009
22	Taiwan	1,063,249,044	319,525	2,869,620	11,169,709	397,814
23	China	1,026,880,076	480,458	3,011,236	11,682,400	410,327
24	South Africa	976,311,738	789,761	4,979,661	24,205,196	783,361
25	Portugal	931,028,991	223,220	1,712,448	4,699,503	215,100

Figure 4.2: The top 25 countries by credits earned through MilkyWay@home, where credits are proportional to the total number of cycles donated by each country. This figure was modified from www.boincstats.com, which uses data from the MilkyWay@home project (Accessed July 3rd, 2013).

goal was to test the platform by giving it previously-characterized data, and determine if the previous results could be recovered. Additionally, these runs on MilkyWay@home allowed us to benchmark the performance of the platform, especially with respect to the RPI SUR Bluegene/L used previously.

With MilkyWay@home, accurate results for each stripe were obtained in 1-2 weeks, with 4-6 stripes running simultaneously. For comparison, with relevant factors taken into account, the BlueGene would take 2-4 weeks per single stripe. Since the numerical optimization methods used by MilkyWay@home were global optimization techniques (that is, they used randomized evolutionary algorithms which are much less likely to be trapped in a local optimum) they did not require "by eye" starting points and repeated runs. However, stripe-to-stripe continuity conditions were difficult to enforce on MilkyWay@home. While the local search method used on the BlueGene/L (gradient descent) allowed the algorithm to start from an initial set of parameters that are continuous with neighboring stripes and find the nearest minimum, the global optimization methods explore the entire parameter space to find the parameter set with the best likelihood, even if it does not make physical sense in light of additional information from adjacent stripes. To compensate for this shortcoming, bounds were enforced on the fit parameters that were shrunk with each successive fit. Initially, all fits were unbound, and in the final iteration all fit parameters were bound at $\pm 20\%$ of the average parameters of the adjacent stripes. This constraint forced the searches to look for parameters that were consistent with adjacent stripes, but was still allowed to search a range of parameters in detail.

The results of the MilkyWay@home fits to stripes 10-23 are presented in Table 4.2. Additionally, the "N_σ" values are presented; these are the number of standard deviations (from Table 3.1) that MilkWay@home differs from the BlueGene results. The two sets of best-fit parameters are very much in agreement with each other; most of the differences are less then 1-σ apart. The only stripes with $N_\sigma > 2$ are stripes 11 and 23. For these stripes, the only outlying parameter is r_0, to which the algorithm is not very sensitive (Cole et al. 2008). This shows that the technique of constraining parameter ranges in the conjugate gradient descent searches produced the global maximum likelihoods, and that the searches were not "coerced" to

their final values.

It has now been shown that MilkyWay@home is a fast, powerful and accurate platform for analyzing astronomical data (Desell et al. 2010a,b). Additionally, MilkyWay@home succeeded in producing the same or better results than the BlueGene searches, with virtually no artificial constraints. MilkyWay@home is currently active crunching data for 4 projects: extending this analysis of the Galactic halo to the Galactic South, using data from SDSS DR8 (Thompson et al., *in prep.*); fitting the additional structure in the Northern Cap, such as the Sgr Bifurcation and Virgo (see Chapter 5); implementing the the results of the F turnoff star analysis (Chapter 6) to the stream-fitting algorithm; and developing the N-body application (Desell et al. 2011), so that future analysis can include orbit fits and probe the distribution of the Dark Matter halo.

Table 4.2: MilkWay@home Stream Fits and N_σ

Stripe	q	N_σ	r_0 (kpc)	N_σ	ϵ	N_σ	$\mu(^\circ)$	N_σ	R (kpc)	N_σ	$\theta(rad)$	N_σ	$\phi(rad)$	N_σ	σ	N_σ
10	0.53	0.34	8.0	0.36	-1.26	0.31	230.0	1.1	36.4	1.69	1.50	0.80	3.08	0.04	5.0	2.0
11	0.50	0.44	15.4	5.00	-1.64	1.10	205.2	0.8	42.0	0.07	1.53	0.25	3.09	0.60	5.1	1.5
12	0.55	0.02	6.77	0.07	-1.34	0.16	201.6	0.1	40.6	0.01	1.31	0.02	3.13	0.10	5.5	0.2
13	0.50	0.10	6.52	0.09	-1.10	0.12	197.3	0.1	40.0	0.72	1.70	1.00	-2.99	0.01	4.2	0.4
14	0.57	0.01	7.19	0.01	-1.55	0.30	183.6	0.1	28.2	0.04	1.91	0.14	2.92	0.07	3.9	0.3
15	0.56	0.00	8.64	0.01	-1.80	0.17	180.1	0.01	27.6	0.03	1.91	0.01	3.00	0.00	3.3	0.1
16	0.59	0.09	5.93	0.07	-1.47	0.01	169.0	0.04	28.2	0.07	2.33	0.04	2.87	0.06	2.5	0.2
17	0.61	0.11	9.67	0.74	-1.61	0.02	165.3	0.4	27.8	0.33	1.92	0.80	2.92	0.07	2.9	0.5
18	0.56	0.08	7.32	0.00	-1.97	0.00	157.3	0.03	25.2	0.02	1.97	0.18	2.82	0.18	2.3	0.2
19	0.52	0.02	5.97	0.00	-1.93	0.03	151.5	0.00	23.6	0.02	2.40	0.00	3.04	0.02	1.0	0.02
20	0.53	0.08	9.77	0.08	-2.76	0.11	147.9	0.01	22.8	0.02	2.36	0.20	2.95	0.07	1.1	0.03
21	0.53	0.03	6.65	0.06	-2.01	0.29	142.1	0.03	21.4	0.02	2.62	0.02	2.84	0.08	1.1	0.1
22	0.35	0.25	3.45	0.12	-2.00	0.00	135.8	0.3	17.9	0.01	2.54	0.14	1.94	0.02	1.4	0.6
23	0.47	1.13	1.0	2.87	-1.86	1.85	138.2	0.3	14.1	0.14	2.20	0.30	1.04	0.22	0.9	0.1

CHAPTER 5
Additional Substructure in the North Galactic Cap

After the removal of the major Sgr tidal stream (Section 3.2), there remain two noticeable substructures in the North Galactic cap (Figure 3.7): the so-called "bifurcated" piece of Sgr, and the Virgo overdensity. The density signal from the main Sgr stream was so large that the algorithm had trouble fitting these secondary structures alongside it, and so were ignored during the previous study (Chapter 3). As no single model of the Sgr dwarf debris system has been able to account for all of its observed properties, especially the bifurcated pieces, a characterization of the bifurcated piece of Sgr will be useful to future N-body modelers. By providing strong constraints on the bifurcation's position and density, we hope to give modelers the ability to better match the actual properties of the dwarf, and thereby probe the physics of the stellar halo. This chapter highlights the preliminary results obtained for these substructures, including current progress with searches on MilkWay@home, and discusses the current difficulties in this process.

5.1 SDSS Without Sagittarius

Using the separation method outlined in Cole (2009), the main Sgr tidal stream was removed from the SDSS data (see Figure 3.6). With the "elephant in the halo" removed from the data, the remaining substructures become the dominant densities in the data, and so the algorithm should characterize them accurately. The new, "sans-Sgr" data wedges were then used to start global likelihood searches on MilkyWay@home, searching with two streams where only the bifurcated piece is expected (Stripes 14 through 21), and searching with 3 streams where both the bifurcated piece and Virgo may be present (Stripes 9 through 13). Only stripes 9 through 21 were used, as the bifurcated piece is only strongly apparent in stripes 13 through 21, and Virgo only spans stripes 9 through 13.

5.1.1 Initial, Unconstrained Parameter Fits

The initial searches run on the sans-Sgr data were unable to satisfactorily converge to reasonable parameters for the secondary halo substructures. No constraints were placed on the fit parameters in these searches (although spatial parameters were not allowed to leave the volume of the data), and so the global optimization methods were allowed to search a large volume of the parameter-space, including parameter sets that might be unphysical or discontinuous across the data. The separation method could be used to isolate each fit component and evaluate the validity of the fits. From the appearance of the smooth Hernquist spheroid fits shown in Figure 5.1, it is clear that the best-fit parameters in each stripe do not result in a continuous distribution, and so continuity constraints, like those applied to the main Sgr stream (Section 3.1), may have to be used here.

Figure 5.2 displays the best stream fits from the initial likelihood searches for each stripe. While the spheroid in Figure 5.1 is clearly not continuous, and therefore not well-fit, a few of the stream fits did in fact find reasonable fits to the bifurcated piece and Virgo. This is an indication that the remaining sub-structures in the North Galactic cap can be accurately fit, and that constraints may be necessary in order to enforce stripe-to-stripe continuity and prevent the additional model streams from over-fitting the spheroid.

5.1.2 Constrained Parameter Stream Fits

A second set of likelihood searches were run on MilkyWay@home, this time with parameter constraints motivated by the previous, unconstrained runs. The constraints restricted Virgo to a range of spatial (μ, r_0, σ) parameters extending to the visible edges of overdensities in the face-on wedgeplots. The bifurcated piece was allowed to explore a range of parameters consistent with where previous authors (Belokurov et al. 2006b; Newberg & Yanny 2006; Koposov et al. 2012) detected the overdensity. An extra stream was fit in each stripe (as in Chapter 3), but this time limited to $\epsilon \leq 0.0$, preventing it from becoming more dominant than the smooth component.

In general, the bifurcation is located at greater r than the bright stream, and

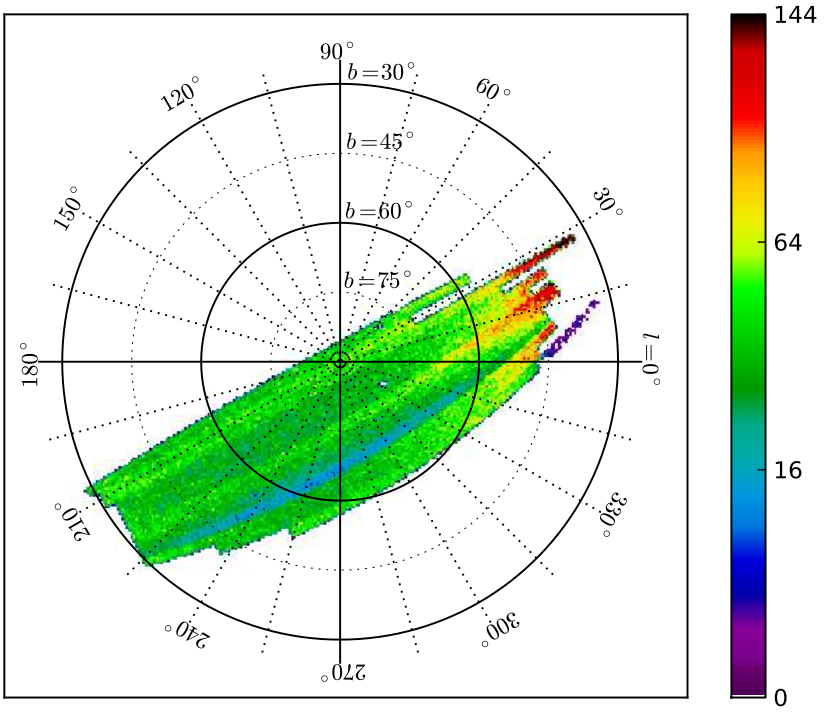


Figure 5.1: Polar plot of the "sans-Sgr" data (Figure 3.7) with the secondary stream fits removed. The stripe-to-stripe discontinuities make it clear that the initial, unconstrained likelihood searches were unable to characterize the sans-Sgr spheroid satisfactorily.

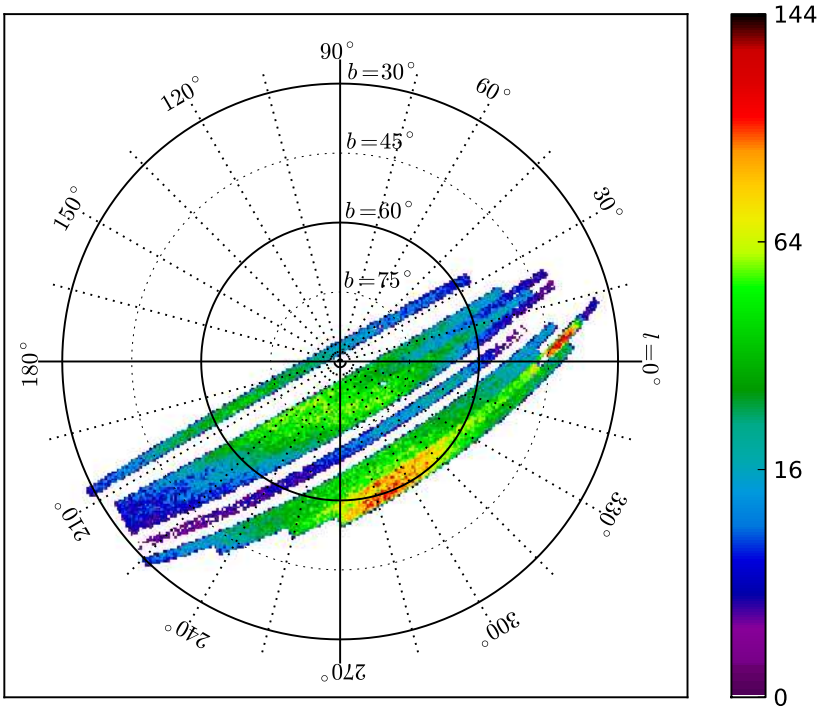


Figure 5.2: Polar plot, similar to Figure 5.1, but showing only the streams fit during the first set of likelihood searches on the sans-Sgr data. Data stripes in which the stream fits were deemed unphysical (overly-wide or located on the physical edge of the data volume) were omitted from this plot. It appears as though parts of the bifurcated piece and Virgo are visible, but not yet continuous from stripe to stripe.

so the minimum distance allowed was $r = 15.0$ kpc, while the maximum was left at the distance limit of the data stripe. The μ limits were a 20° to 40° range, chosen to include the best fit stream from previous runs, as well as regions cited by previous authors. Since the weights of the main Sgr stream never rose above -1.0 in the region of interest (Table 3.1), and was much more dominant in the halo, the bifurcated piece is expected to have a lower weight and was therefore constrained to $\epsilon \leq 0.0$. It is possible that the removal of the bright stream took enough stars out of the data that the bifurcated piece weights (ϵ) could exceed 0.0 (that is, there are more stars in the bifucation than in the smooth background), but this is not expected. If this is the case, then the stream weights will run into the bound at $\epsilon = 0.0$, and new runs will need to be started with a larger maximum allowed value.

The model streams meant to fit the Virgo overdensity were allowed a large range of μ values, but a constrained set of distance values. For the stripes that contain Virgo, the only disallowed μ values were those within 5° of the stripe edge, forcing it to look within the stripe, where it is visibly located. The distances were constrained to be within ± 4.0 kpc of the best-fit distance from the unconstrained round of fits. Since Virgo is a large, dominant substructure, the stream weights (ϵ) for Virgo were unconstrained.

The results from this round of likelihood searches are presented in Tables 5.1, 5.2, and 5.3. The first stream in each stripe was constrained as the bifurcated stream, the second stream was constrained as Virgo for stripes 9-13, and the remaining stream was treated as the extra "garbage collection" stream. Several values have run into the constraints, as indicated by whole-integer values (no decimal point) or values of $\pm 2\pi = \pm 6.28$ radians for θ and ϕ. Noticeably, several fits to stream 1 (Table 5.1) run into the inner and outer r_1 bounds (15.0 and 45.6 kpc, respectively), and the remaining stripes span a large range of distances. In stripes 9 through 11 the bifurcated piece is close to or within the increasing densities near the edge of the data, which may be a source of confusion for the algorithm. The fits to other stripes may be confusing the stream with a far-edge (45.6 kpc) overdensity bias (see the SDSS photometric color error analysis in the next chapter), or other substructures (such as Virgo). Many of the stream 1 fits run into the maximum stream width

Table 5.1: New Stream 1 Fits

Stripe #	ϵ_1	μ_1	r_1	θ_1	ϕ_1	σ_1
9	-0.64	186.52	18.57	2.69	0.82	1.77
10	-0.45	220.39	25.27	-5.45	3.22	20
11	-1.75	229	51.2	1.12	-2.51	3.63
12	-2.03	233.09	46.8	-4.07	-5.66	5
13	0	213.6	10.52	-0.93	-6.28	5
14	-1.37	193.12	15	1.85	-2.22	2.85
15	-1.93	191.26	15	-4.58	-6.28	2.92
16	-1.14	190.58	15	-3.43	4.96	5
17	-2.14	170.79	45.6	1.54	3.59	5
18	-2.19	173.11	45.6	-1.81	3.62	5
19	-0.43	190	15	-3.95	-0.32	6.74
20	-1.72	179.22	18.71	-4.22	6.05	5
21	-1.45	176.01	15	-3.77	-5.0	5

(σ) of 5 kpc, possibly because the algorithm is trying to simultaneously fit the bifurcated piece and smooth component that deviates from a Hernquist density profile. Alternatively, the stream may actually be much wider than the main Sgr stream. Therefore, either the imposed bounds were too restrictive, or the the "free" (extra) stream found a better fit to the bifurcated stream and left stream 1 to attempt to fit potential deviations of the Hernquist model. In either case, it is clear that at least one more round of likelihood searches will be necessary.

Few of the fits for streams 2 and 3 run into spatial constraints, but many of these fits are stuck on the 2π angular limits. Due to the cyclical nature of angles, it may be that the actual best-fit value is greater than 2π (or 0 radians), and the searches became stuck at this bound since the parameter fits do not "wrap-around" back to 0 at the bounds. Otherwise, these streams are doing a good job of fitting the local density substructures that are not associated with a Hernquist density profile.

To see if these fits provide a reasonable description of the data, all of the fit streams were removed from the background using the separation method, and plotted together in Figure 5.3. Excepting the stripes that contain the bifurcated stream and Virgo together, the aggregate sum of the separated streams does not look terrible, with the bifurcated piece readily apparent near the center, and Virgo near the bottom of the plot. These fits are a definite improvement over the previous set, and are a clear sign that these substructures can eventually be characterized with the algorithm. However, the streams appear very wide, much wider than was

Table 5.2: New Stream 2 Fits

Stripe #	ϵ_2	μ_2	r_2	θ_2	ϕ_2	σ_2
9	0.02	210	26.66	0.58	-2.27	4.91
10	-1.69	170.93	21.98	1.61	6.28	3.28
11	-0.31	189.2	14.89	1.68	1.83	7.2
12	-1.15	170	20	5.55	6.28	20
13	-1.96	147.61	19.19	-1.96	-6.28	2.06
14[1]	-1.29	200.79	42.82	-1.73	-6.28	10
15	-1.29	192.4	34.21	1.1	1.32	10
16	-1.16	226.82	17.7	-0.49	-6.07	18.18
17	-0.55	195.85	12.22	-2.4	5.1	8.26
18	-0.8	186.78	17.47	-0.75	-4.51	7.2
19	-1.21	191.43	35.01	4.35	-3.06	14.64
20	-1.12	249	21.96	-1.02	-5.32	11.72
21	-2.99	133	2.4	-4.91	-5.79	1.79

[1]Stripes 14 through 21 were only fit with two streams, with the second stream (shown here) as the "free" stream.

Table 5.3: New Stream 3 Fits

Stripe #	ϵ_3	μ_3	r_3	θ_3	ϕ_3	σ_3
9	0	240	2.4	0.17	0.29	3.63
10	-0.72	199.97	7.64	-6.28	6.28	4.22
11	-1.94	150	42.43	3.14	0.42	20
12	-0.54	228.95	4.29	5.92	-6.28	7.05
13	0	210.73	45.6	1.85	0.67	20

found for the main Sgr stream (Chapter 3). Since the bifurcation is known to be less dominant than the bright stream, it does not seem likely that this stream's width would be the wider of the two. It will require at least another round of fits to determine if current fits are in error, but if the final fits also find large widths, then there are three possibilities: (1) This is an artifact of the faintness of the stream, and the algorithm is unable to differentiate the Hernquist background from the faint edges of the stream; (2) the bifurcated piece is actually much wider than the bright stream; or (3) the bifurcated piece is not well-described by our assumed stream density profile. If case (2) is true, than this might imply that the bifurcated piece is more disrupted than the bright stream, or that it was created by a more diffuse object than that of the bright stream's progenitor. This will require further investigation, including using N-body simulations to match the observed densities.

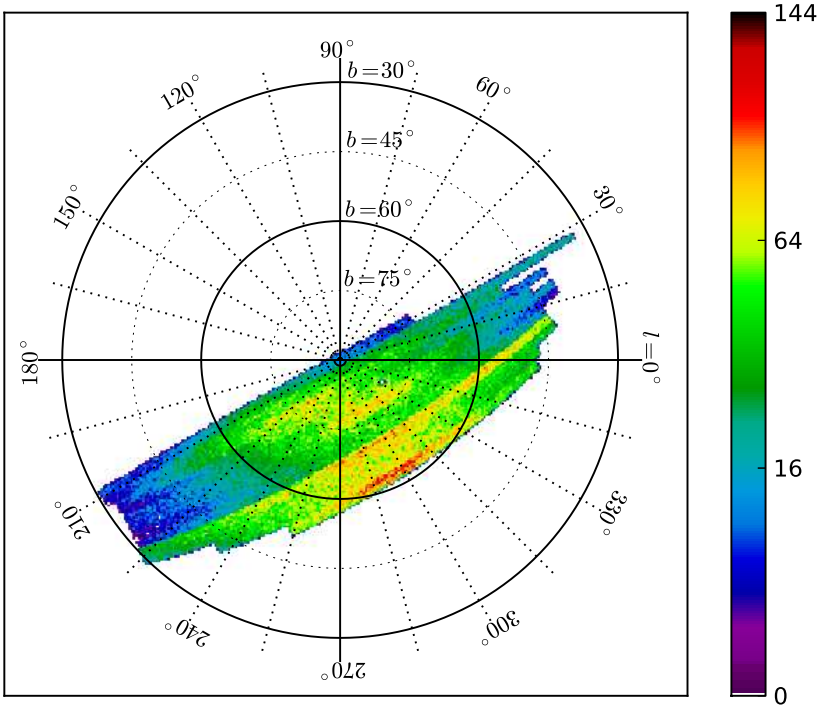


Figure 5.3: Polar plot, similar to Figure 5.2, but with the streams separated using the first round of constrained fits. Virgo and the bifurcated piece are distinct and relatively smooth here, except in the transition stripes between the two.

Table 5.4: New Spheroid Fits

Stripe #	$\mathcal{L}$	q	r_0
9	-2.78	0.58	1.69
10	-2.98	0.58	4.59
11	-2.99	0.56	5.57
12	-3.09	0.56	6.07
13	-3.0	0.43	2.68
14	-2.88	0.56	6.13
15	-2.92	0.56	8.53
16	-2.94	0.58	6.38
17	-2.94	0.59	6.52
18	-2.96	0.58	8.4
19	-2.95	0.53	6.91
20	-2.97	0.52	8.89
21	-2.88	0.79	15.59

5.1.3 Constrained Stream Spheroid Fits

The smooth spheroid parameters (q, r_0) were allowed to explore the full range of physically allowed values during the constrained stream likelihood searches, with the results presented in Table 5.4. With the exception of stripe 21, the fits find similar values for the flattening parameter q across all stripes (excepting 21), averaging to 0.57 with a standard deviation of 0.07.

The Hernquist scale length, r_0, is also very similar across all stripes, but with some variation, which is acceptable because r_0 is the least sensitive parameter in the algorithm (Cole et al. 2008). The fits to all 12 stripes find a average $r_0 = 6.76$, with a standard deviation of 3.26. This is consistent with the results of the bright stream fits and their large errors, as discussed in Section 3.1.

After extracting the fit streams from this data (Figure 5.3), the smooth Hernquist background remains, and is plotted in Figure 5.4. This figure shows that this round of fits produces a smooth, continuous density profile, except for stripes 13 and 21, which appear to be under-dense and over-dense, respectively. Most likely, the algorithm was over-zealous in fitting the streams in these stripes to local over-densities, since the bifurcated stream is expected to be very faint in stripe 21, and stripe 13 appears to be generally confusing to the algorithm.

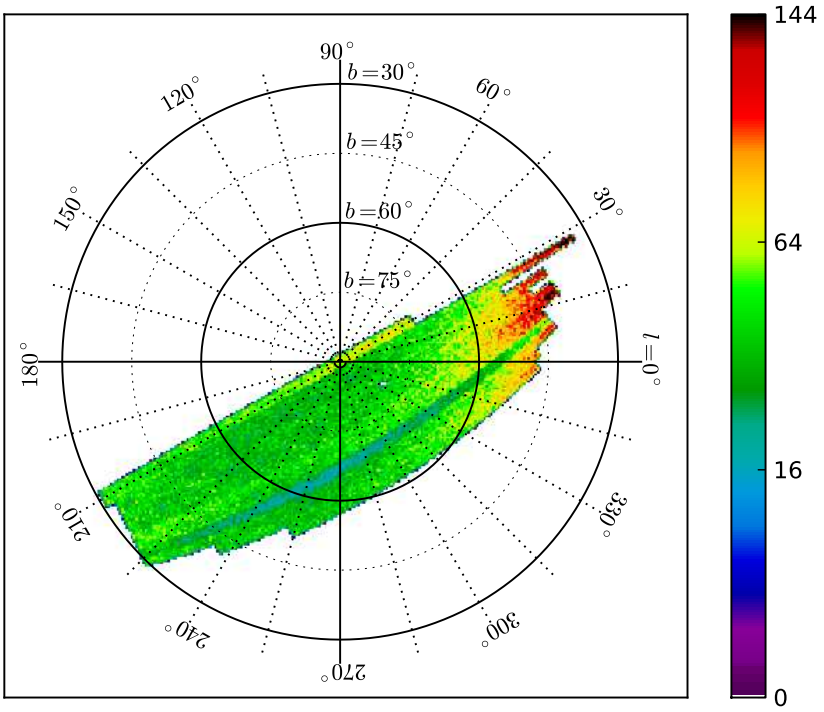


Figure 5.4: Polar plot, similar to Figure 5.1, but showing only the the spheroid fits to the "sans-Sgr" data, fit with constraints placed on the stream parameters. It appears that the algorithm finds a smooth Herquist density profile across all stripes, except for stripes 13 and 21, where the overdensities may be overly faint or confusing.

5.2 Discussion

The fits to the sans-Sgr data will require additional rounds of likelihood fits before a physical fit is achieved, but the second round of fits is a noticeable improvement over the first. The minimum distance (r) constraints for the bifurcated stream will have to be re-analyzed due to several of the stream 1 fits ending on the 15 kpc bound, which, assuming the absolute magnitude of an average turnoff star is $M_g = 4.2$ (Chapter 6), corresponds to a distance of 20 kpc. This bound may need to be increased; Belokurov et al. (2006b) shows that the brightest part of the bifurcation is at SDSS r magnitude of greater than 20.7, which is roughly equivalent to a g magnitude of 20.9, or a distance of greater than 21.9 kpc. Additionally, when Newberg et al. (2007) analyzed both Sgr streams, they found the nearest distance to the bifurcated stream to be 26 kpc. It seems safe, then, for future stream 1 fits to have a minimum r bound of 21 or 22 kpc.

The current μ constraints were well-motivated by the previous fits, the visual location of the bifurcated piece, and previously published work. They should not need to be modified for future likelihood searches.

The θ and ϕ angle parameters present a problem due to their cyclical nature. Traditionally, the bounds for these parameters have been placed at $\pm 2\pi$, granting these parameters a 4π ($720°$) range in which the likelihood searches can explore[26]. These bounds were chosen to give the original gradient descent maximum likelihood algorithm a "cushion" of values around the initial parameter choices, since the initial parameters could be started from values within the meaningful range, and then would explore in either direction according to the shape of the local likelihood surface. This would, in theory, prevent the gradient descent from reaching either of the bounds, since equivalent best solutions would be found regardless of the direction that the search initially traveled. However, the global searches on MilkyWay@home do not need this cushion, since a large portion of the parameter space is searched by the particle swarm and differential evolution methods. The multiple equivalent

[26]Many combinations of these angles are equivalent, and the unique *vector* solutions are physically bounded by the ranges $0 < \theta < \pi$ and $0 < \phi < 2\pi$. Since the combination of θ and ϕ give the inclination of a stream center within a stripe, a vector pointing at one θ, ϕ value is equivalent to one pointing in the opposite direction, $(\pi - \theta), (\phi + \pi)$.

minimums may actually serve to confuse these global search techniques, since they require multiple independent particles (parameter sets) to converge to a single value.

If a number of particles were to sample the parameter space near one of the angle bounds, and the "true" solution is on the other side of the bound, then it is possible that the search will incorrectly converge on that boundary value. For example, if the "true" solution is at $\frac{\pi}{10}$ and a large number of particles get close to, but below, the 2π boundary, then direction towards the best likelihood will be towards the boundary, to which the particles will travel and become stuck. If enough particles do this, then the remaining particles may also be pulled away from the global maximum. Therefore fitting angles using global likelihood searches is a double-edged sword: If one makes the allowed parameter-space small, then there will be only one "true" solution, but it will be easier for the particles to become stuck on the boundaries; on the other hand, if the allowed parameter-space is large, then there will be a large number of equivalent solutions which may confuse the algorithm, while it becomes more difficult for the parameters to run into any boundary.

A solution to this problem was implemented within TAO (by Travis Desell) by allowing angle parameters to "wrap-around" at the boundaries. If an angle parameter tries to step over a bound, TAO will move that particle to the opposite boundary instead of stopping it at the boundary. For example, if a particle were to to try to to move from $0.9(2\pi)$ to $1.1(2\pi)$, with boundaries at 0 and 2π, TAO will move that particle to $0.1(2\pi)$. This fix has just been implemented as of the writing of this thesis, and results are pending.

If modifying the parameter bounds and implementing the angle wrap-around fails to produce reasonable fits to the bifurcated stream, then there are several other courses of action available. To increase the strength of the bifurcated stream's density signal, two adjacent stripes of data can be combined to produce a single stripe with a ν thickness of $5.0°$. Combining two stripes in this way will double the length of the stream in the data, thereby roughly doubling the number of stream stars present and the length of the stream, which should make it easier for the algorithm to find a reasonable fit. This would be relatively easy to implement, as the current software is able to handle this procedure with no modification. The two

adjacent stripes do have some spatial overlap near the μ edges of the data, and so a matching technique would have to be used to identify and remove duplicate stars, then the two resulting data sets could be combined and run on MilkyWay@home.

Another option would be to attempt to fit each stripe with only a single stream, which would force the algorithm to fit that stream to the dominant over-density in the stripe, which should be Virgo or the bifurcation. A single stream fit would have the advantage of finding only the dominant structure in a stripe, but would not have the luxury of an extra stream with which deficiencies in the assumed smooth spheroid model or turnoff star completeness function could be fit.

In summary, more work needs to be done in fitting the secondary, fainter stream found in the North Galactic cap in SDSS turnoff star data. The current round of constrained fits appears to be a step in the right direction, as it produces reasonable-looking stream profiles; however, several of the fits run into imposed constraints, and therefore are not necessarily the global best fits. At the current rate, the secondary stream fits are on-schedule to be finished by during the Fall of 2013.

CHAPTER 6
The Properties of Halo Turnoff Stars

The use of statistical photometric parallax requires that the absolute magnitude distribution of the target population be well-known, and so an understanding of halo turnoff star properties is required in order to use them as valid density tracers. Globular clusters were chosen as adequate analogs for stellar halo stars, as they are of similar ages and may even be the original hosts of many of these stars. In previous work (Cole 2009; Newby et al. 2011, 2013) a Gaussian with $\mu_g = 4.2, \sigma_g = 0.6$ was used to describe the absolute magnitude distribution of turnoff stars, as motivated by the stellar density searches performed by Newberg & Yanny (2006). In theory, the turnoff magnitude distributions could vary by location within the stellar halo, or even along tidal streams, and could be a function of age or metallicity.

This study used halo globular clusters from SDSS DR7 data and found them to have intrinsically similar turnoff properties. This analysis also produced an updated distance-dependent absolute magnitude distribution for color-selected turnoff stars in SDSS data. This distance dependence appears to be entirely due to photometric errors that increase dramatically near the faint edge of the data, such that a different population of stars falls into the color-selection bin near the survey limit. Two additional globular clusters were studied in Grabowski et al. (2013), using data from SDSS DR8, and they were found to be consistent with the results from this study.

6.1 Globular Clusters as Proxies

In this chapter, eleven Milky Way halo globular clusters are analyzed. We find that the peak of the distribution of absolute magnitudes of F turnoff stars is typically $M_g = 4.18$, and that the asymmetric distribution can be approximated by a half Gaussian on the bright side with a width of $\sigma = 0.36$, and a half Gaussian on the faint side with a width of $\sigma = 0.76$. This distribution is surprisingly similar for

Portions of this chapter previously appeared as: Newby, M., Newberg, H. J., Simones, J., Cole, N., & Monaco, M. 2011, ApJ, 743, 187.

all of the globular clusters studied; these clusters range in age from 9.5 to 13.5 Gyr and from [Fe/H]= -1.17 to [Fe/H]= -2.30, which are the typical values for halo stars. This surprising result is due to the age-metallicity relationship for Galactic stars: Older clusters should have fainter, redder turnoffs; however, older clusters also contain fewer metals, which pushes the turnoff brighter and bluer. Although the color of the turnoff varies slightly from cluster to cluster, the absolute magnitude of the turnoff only shifts about 0.1 magnitudes from the mean.

Although the turnoff magnitude was found to be similar for the clusters studied, observational effects may considerably change the properties of turnoff star distributions. As stars are sampled to the limit of the survey, the photometric errors increase. Although these photometric errors are small compared to the uncertainty with which the absolute magnitude of a single F turnoff star is measured, they can be large compared to the width of the turnoff color selection box. For bright magnitudes, the actual colors of most of the stars selected are within the color range selected. For magnitudes near the survey limit, some stars that should be selected are randomly measured with colors that are too red or too blue, and a larger number of stars that are too red are randomly scattered into the selection range. The largest flux into and out of the color selection range is on the red side of the range, and primarily broadens the measured width on the faint side of the peak as the survey magnitude limit is approached.

The data acquisition for the eleven globular clusters is described in Section 2.3. From this data, g_0 vs $(g-r)_0$ color-magnitude diagrams were made, fiducial sequences were constructed, and then isochrones were fit in order to obtain the cluster distances and ages; cluster metallicities were taken from previous literature. Once distances were determined and cluster-specific biases removed, the absolute magnitude distribution of each cluster was fit with an asymmetric Gaussian. The observational biases due to increasing distances were characterized by convolving nearby clusters with the photometric errors that cluster would have if it were observed at a greater distance. Finally, the absolute magnitude distribution of F turnoff halo stars was characterized as a function of distance, and the results discussed in the context of the Milky Way's age-metallicity relationship.

6.1.1 The Globular Cluster Sample

In all, the turnoff star project used photometric data from eleven globular clusters, taken from the SDSS database's seventh data release (DR7; Abazajian et al. (2009)). There are a total of seventeen globular clusters within the SDSS DR7 footprint, but five clusters (NGC 2419, NGC 7006, Pal3, Pal 4, and Pal 14) were eliminated because they are too distant, and one (Pal 1) was eliminated because there were too few stars to obtain accurate measures of the F turnoff star distribution using SDSS data. A list of all clusters studied in this investigation can be found in Table 6.1.

The following clusters are listed in Table 6.1, but were excluded in some of the analyses:

NGC 5053 has a very low Zinn & West metallicity ($[Fe/H]_{ZW84}$ ∼-2.58) which falls outside the range of the Carretta & Gratton (1997) conversion scale, and is also below the minimum metallicity value for which Padova isochrones can be generated. In the newer work of Carretta et al. (2009), an $[Fe/H]$ of -2.30 is found, which is within the range of the Padova isochrones (Girardi et al. 2000; Marigo et al. 2008), and therefore this metallicity value is used in the analysis. To indicate status as a potential outlier, results for NGC 5053 are plotted using a red-dotted series in figures showing turnoff stars properties.

M15 (NGC 7078) is a 'core-collapsed' cluster, and so has a very compact core. It may also have different dynamics and stellar distributions than standard globular clusters (Haurberg et al. 2010). The SDSS photometric pipeline does not attempt to deblend crowded star fields, and so information on M15 in SDSS is highly biased towards stars found on the edges of the cluster. This cluster is not necessarily expected be consistent with the other clusters, but it is included in the analysis anyway. M15's status as potential outlier is indicated in figures showing turnoff star properties by plotting it with a red dotted series.

NGC 4147 contains only slightly more stars than M92, (583 versus 334) but they are more concentrated around the turnoff due to a lack of stars below $M_g = 6.5$. This lack of stars is due to SDSS crowded-field photometry detection efficiency problems at fainter magnitudes (see Section 6.3). NGC 4147 is analyzed alongside

Table 6.1: Results of modified Padova isochrone fits

NGC Number	Messier Number	l °	b °	Metallicity CG97 [Fe/H]	Fit Distance (kpc)	Fit Age ± Error (Gyr)	r_{clus} °	r_{cut} °	# Stars in Cluster[1]
NGC 4147	···	252.85	77.19	-1.58	19.3	11.2 ± 0.8	0.07	0.08	583
NGC 5024	M53	332.96	79.76	-1.89	18.7	12.5 ± 1.3	0.25	0.30	5871
NGC 5053	···	335.70	78.95	-2.30[2]	18.5	11.5 ± 1.1	0.14	0.17	2280
NGC 5272	M3	42.22	78.71	-1.43	10.4	12.2 ± 1.1	0.30	0.40	4395
NGC 5466	···	42.15	73.59	-2.14	15.6	13.4 ± 0.9	0.18	0.25	2029
NGC 5904	M5	3.86	46.80	-1.17	8.0	9.4 ± 0.1	0.21	0.28	3488
NGC 6205	M13	59.01	40.91	-1.42	7.7	12.7 ± 0.2	0.24	0.33	2192
NGC 6341	M92	68.34	34.86	-2.17	8.7	13.4 ± 0.9	0.135	0.20	334
NGC 7078	M15	65.01	-27.31	-2.04	11.0	11.2 ± 0.9	0.19	0.25	2939
NGC 7089	M2	53.37	-35.77	-1.38	11.5	12.6 ± 0.3	0.14	0.18	1451
Pal 5	···	0.85	45.86	-1.24	21.0	11.5 ± 0.7	0.087	0.12	1478

[1]The number of stars represents the number of stars contained within r_{clus}, and therefore the number of stars used to build the cluster dataset.

[2]The Zinn & West metallicity of NGC 5053 (-2.58) was outside the effective conversion range of the CG97 metallicity scale, as well as the range of available Padova isochrones, so the Carretta et al. (2009) value of -2.30 was used instead

the other clusters, but errors are expected to be large due to the low number of data points available. In figures showing turnoff star properties, a blue dotted series is used for NGC 4147 to indicate it's low star counts and high expected errors.

Cluster M92 (NGC 6341) fell on the edge of the SDSS DR7 footprint, and relatively few stars were observed in the cluster. The number of cluster stars was large enough to produce a fiducial fit with large bins in magnitude (0.5 magnitudes). Therefore we were able to fit a modified Padova isochrone to this cluster. However, as there are so few M92 stars in the data, especially close to the turnoff, that M92 was omitted from the turnoff analysis.

6.2 Isochrone Fitting to Determine Cluster Distances

In order to convert the observed apparent magnitudes to absolute magnitudes, the measured distances to each globular cluster are needed. To study the effects of age and metallicity on the absolute magnitudes of the turnoff stars, measurements of these quantities that are as accurate and as uniform as possible for the sample of globular clusters are preferred. In this section, the spectroscopically determined metallicities ($[Fe/H]$) are assembled from a single group of authors; measurements were obtained from Zinn & West (1984) and then converted to more modern Carretta & Gratton scale using the conversion provided in (Carretta & Gratton 1997). Using these metallicities, ages and distances are then fit to the clusters in a consistent fashion using Padova isochrones.

The Padova theoretical isochrones were fit to fiducial sequences determined from data for eleven clusters found in SDSS. Stars in each cluster were separated into g_0 bins, then the mean and standard deviation (σ_{g-r}) of the $(g-r)_0$ distribution in each bin was found. Any stars in the g_0 bin with a $(g-r)_0$ value beyond $2\sigma_{g-r}$ from the mean were rejected, and then the $(g-r)_0$ mean and σ_{g-r} of the remaining population was determined. The $(g-r)_0$ mean and average M_g value were accepted as a point on the fiducial sequence once the entire bin was within $2\sigma_{g-r}$ of the current mean. Isochrones were then fit to the fiducial sequences using distance and age as free parameters, while metallicity was held constant at the spectroscopically determined value.

In the initial attempts to use this technique to determine cluster properties, it was found that Padova isochrones that are good fits to both the main sequence and the subgiant branch require unreasonably high ages (>15 Gyr) for most clusters [27] The lack of agreement between theoretical isochrones and cluster data has been explored by previous authors. Using eclipsing binary stars in the Hyades open cluster, Pinsonneault et al. (2003) showed that theoretical isochrones do not match true star populations; there are discrepancies in mass, luminosity, temperature, and radius. In the second paper in the series, Pinsonneault et al. (2004) uses *Hipparcos* parallax data for the Hyades cluster to further calibrate theoretical isochrones, finding that offsets in color indexes are sufficient to bring a theoretical isochrone in line with real main sequence data. In the fourth and final paper of the series An et al. (2007) fits isochrones to Galactic open clusters using color corrections in $(B-V)_0$ as a function of M_v, while using Cepheid variables as calibration points. In An et al. (2009), updated Yale Rotating Evolutionary Code with MARCS model atmospheres were used to produce *ugriz* isochrones which were fit to main sequences of five globular clusters, producing ages and distances to these clusters.

Using similar techniques, Padova *ugriz* isochrones were calibrated to the An et al. (2009) results. Comparing a Padova isochrone generated from the An et al. (2009) derived age and distance for globular cluster NGC 6205 with the derived fiducial fit, it was found that the difference between the theoretical isochrone and the data is very nearly linear along the main sequence (Figure 6.1). Therefore, linear color correction function was applied in $(g-r)_0$, holding M_g as the independent variable, to the Padova isochrone to bring it into agreement with the fiducial fit. Only the main sequence and subgiant branch were fit, and the giant branch ($\sim M_g > 3.5$) was ignored, as the linear correction is not valid for these stars. The lower main sequence was also not fit, as the data does not extend to fainter absolute magnitudes. The best-fit color-correction functional fit is:

$$\Delta(g-r) = -0.015 * M_g + 0.089 \tag{6.1}$$

[27] These ages would be in significant conflict with the recent WMAP and Plank ages of 13.8 Gyr, derived from measurements of the cosmic microwave background (Bennett et al. 2009; *Planck Collaboration* 2013)

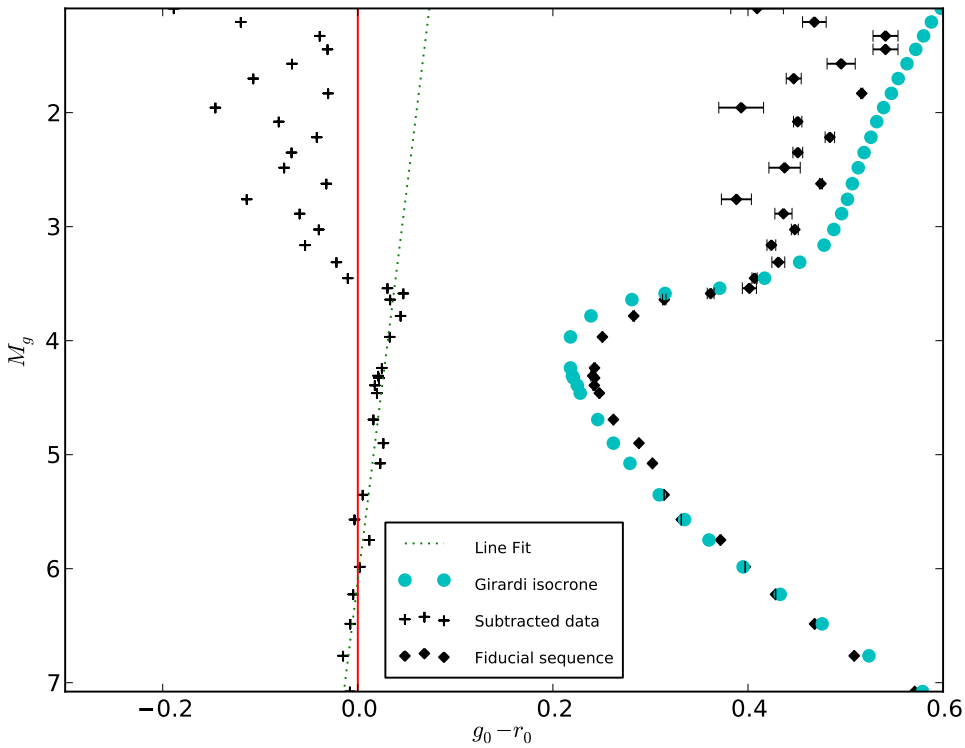


Figure 6.1: Diagram of Padova Isochrone fit for cluster NGC 6205. The values of the raw Padova Isochrone, (blue circles) generated from the An et al. (2009) results, were subtracted from the values of the cluster fiducial fit (black diamonds) to produce the offset (crosses). A red vertical line indicates zero offset, or a perfect match between isochrone and fiducial fit. A linear fit (dotted line) to the offset was made for values below the giant branch ($M_g > 3.5$): $\Delta(g-r)_0 = -0.015 * M_g + 0.089$. This correction allows Padova isochrones to fit the turnoff on color-magnitude diagrams. We then used this fit to match isochrones to our cluster sample (see text).

This function was applied to all of the colors in the model isochrones, and then these models were fit to the fiducial sequences derived from the data. A Padova isochrone was chosen for each cluster that, when modified by Equation 6.1, fit the fiducial sequence well. Additional isochrones were then chosen, identical except for age, which was varied about the "by eye" best fit in increments of 0.2 Gyrs. A Gaussian function was fit to the residuals of these isochrone fits, and the mean of this Gaussian was taken as the best fit age.

Errors in the fit ages were determined through a Hessian matrix method, by comparing the residuals in the isochrone fit. In the limit of a single fit parameter (age), the Hessian error method reduces to:

$$\sigma = \left(\frac{R^2(t+2h) + R^2(t-2h) - 2R^2(t)}{8h^2} \right)^{-\frac{1}{2}} \tag{6.2}$$

Where $R^2(t)$ is the residual of the isochrone fit at age t, and $h = 0.2$ is the step size in the age determination method. Using Equation 6.2, the age fit errors for each cluster were determined through the use of three isochrones: the isochrone of best fit age, and two isochrones generated at the best fit age ± 0.4 Gyr.

Padova isochrones fit using the above correction function produce a consistent set of metallicities, ages, and distances to the globular cluster sample, as presented in Table 6.1. Cluster color-magnitude diagrams, fiducial fits, and modified isochrone fits are shown in Figure 6.2.

Our distance determinations are compared to the distances in three other sources (De Angeli et al. 2005; Harris et al. 1997; Dotter et al. 2010), and compared with other isochrone-derived ages (De Angeli et al. 2005; Marín-Franch et al. 2009; Dotter et al. 2010), in Figure 6.3. The fit distances appear to be in excellent agreement with other sources. The fit ages agree to within the formal errors for each cluster, but appear to have a small linear systematic offset. The derived ages are a very close match around 13 Gyr, but are a Gyr or two higher for ages of 10 Gyr, so the scale of the fit ages is slightly more compressed than the ages in the comparison sources.

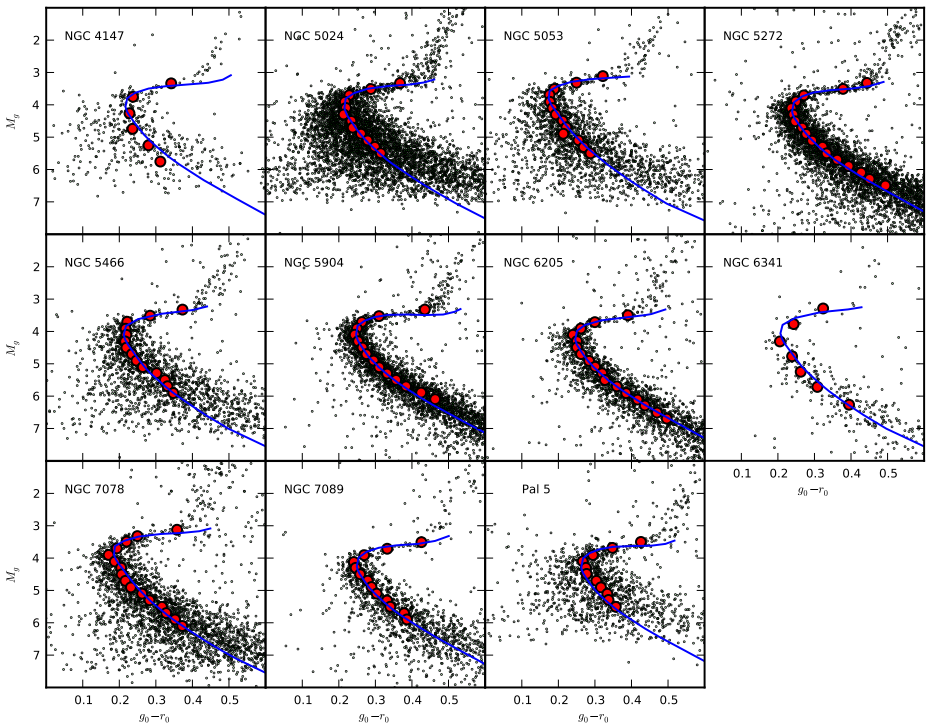


Figure 6.2: Color-magnitude diagrams, fiducial fits, and isochrone fits for all eleven globular clusters used in this study. Black dots represent the positions of individual stars in M_g, $(g-r)_0$; red circles represent 2-σ rejection fiducial fits to the stars in M_g strips; and blue lines represent Padova isochrones modified by our correction function. The ages, metallicities, and distances for the isochrones are given in Table 6.1. Note the absence of faint stars in some clusters. This is due to the poor performance of the SDSS photometric pipeline in crowded fields. The shape of this incompleteness is described in the text. Also note the sparse data in NGC 6341; it will not be used for F turnoff analysis.

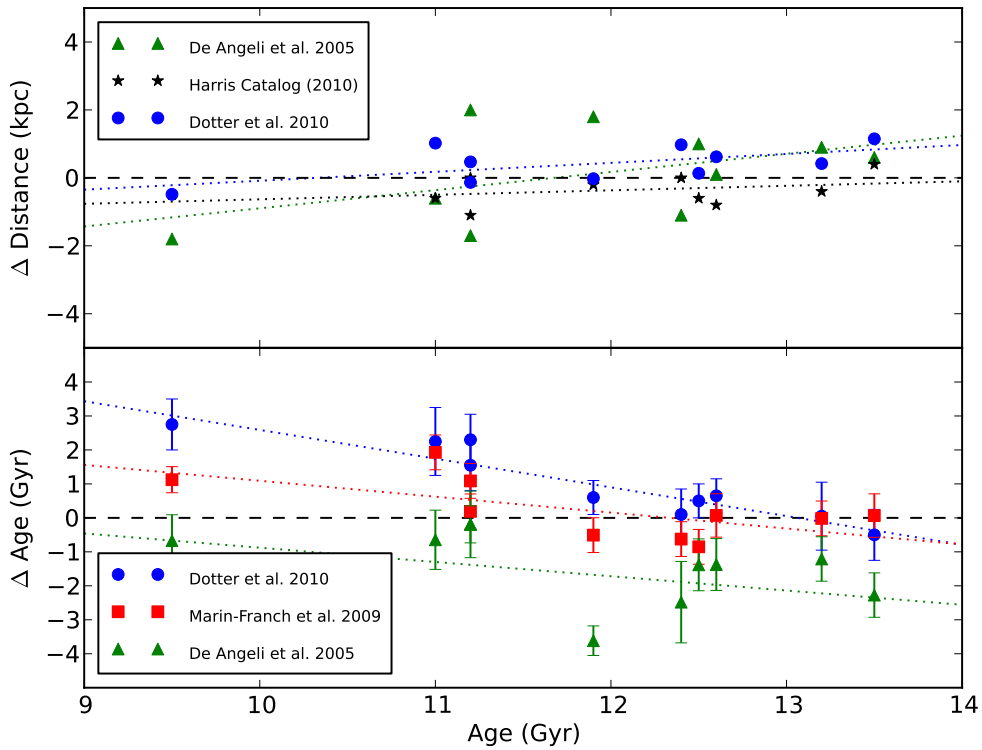


Figure 6.3: Comparison of the distance and age values from this paper with values from other sources. The top panel plots the distances from this paper, subtracted from the respective external values, plotted vs. our fit ages. Dotted trend lines are linear fits to the points. Our distances are in good agreement with the previous literature. The lower panel is the same as the top panel, but with the age differences as the y-axis. Our ages agree with the previous literature to within the formal errors, but have a small linear offset. Note that Palomar 5 is absent from this plot, as the other authors did not study this cluster.

6.3 Detection Efficiency for Stars in SDSS Globular Clusters

It is clear from Figure 6.2 that the cluster data is incomplete at fainter absolute magnitudes, especially amongst the farther clusters: Pal 5, and NGC 4147, NGC 5024 and NGC 5053. This incompleteness begins at a brighter magnitude than is expected from the SDSS detection efficiency for stars (Newberg et al. 2002). The poorer detection efficiency in globular clusters is due to difficulty in detecting faint sources in highly crowded star fields, and in particular the poor performance of the SDSS photometric pipeline in this regime (Adelman-McCarthy et al. 2008). For sufficiently crowded fields, the code cannot deblend and resolve faint stars, since they are washed out by much brighter stars.

To quantify the cluster detection efficiency, nearby clusters with relatively complete CMDs were examined and compared to the farther, incomplete CMDs under the assumption that there are similar absolute magnitude distributions for the stars in nearby and distant clusters. Stars were selected from the CMDs in a rectangular box with bounds $3.567 < M_{g_0} < 7.567$, $0.0 < (g-r)_0 < 0.8$, and an additional color cut of $(u-g)_0 < 0.4$. Three nearby clusters - NGC 6205, NGC 5904, and NGC 5272 - were chosen, and it was found that their normalized M_{g_0} histograms in the box were, in fact, similar. An average of these histograms was used to create a reference absolute magnitude distribution for cluster stars. Assuming that this histogram represents the true magnitude distribution of globular clusters, the normalized histograms of the four most distant clusters (NGC 4147, NGC 5024, NGC 5053, and Pal 5) were compared to this reference. To make the normalizations comparable across the incompleteness of the farther clusters, the average difference of the first five bins (which showed a similar rising trend in all of our clusters) was taken, and then the entire histogram scaled by this value, thereby scaling the clusters by matching the initial trends.

To quantify the detection efficiency of the globular clusters, the ratios between the reference histogram and the distant cluster histograms were fit using a parabolic function. These globular clusters were expected to have the same intrinsic absolute magnitude distribution as the bright clusters, but are missing faint stars that were

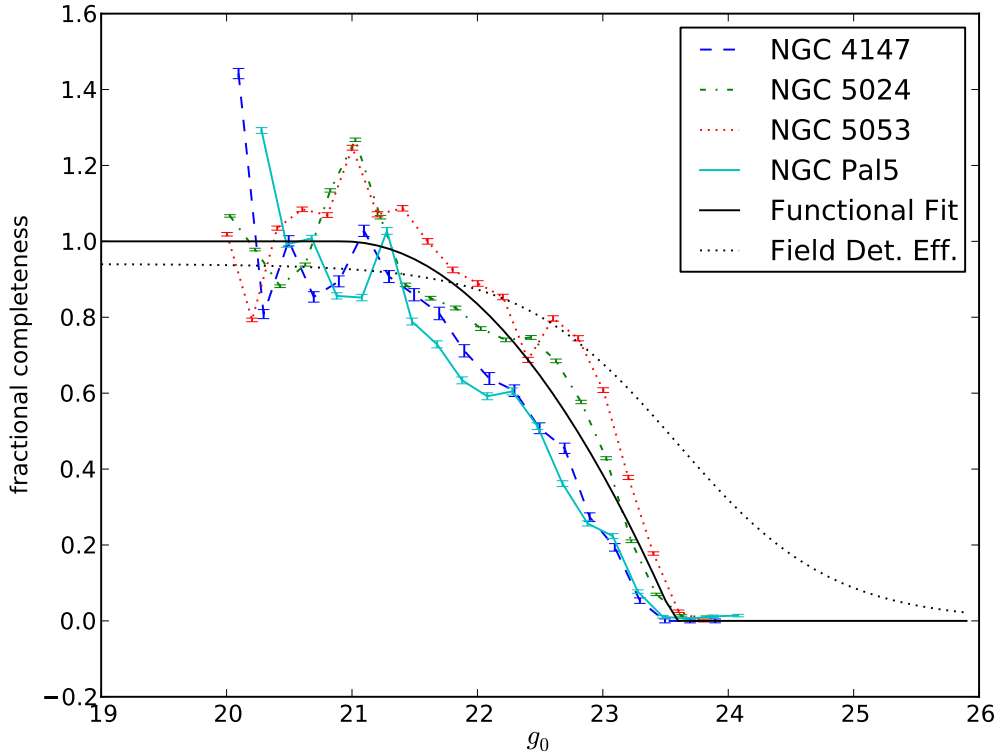


Figure 6.4: The SDSS crowded-field photometry detection efficiency. Due to the poor performance of the SDSS photometric pipeline in crowded fields, more distant clusters have a deficit of stars at fainter magnitudes relative to brighter clusters. The ratio between the number of stars observed in distant clusters and the reference histogram are plotted as colored lines. The reference histogram was created from the average histogram of M_g for the nearby clusters NGC 6205, NGC 5904, and NGC 5272, and then shifted to the distance of each of the distant clusters NGC 4147, NGC 5024, NGC 5053, and Pal 5. The parabolic functional fit ($CDE(g_0) = 1.0 - 0.14 \cdot (g_0 - 20.92)^2$, for $20.92 < g_0 < 23.57$) is plotted as the solid black line. The stellar detection efficiency from Newberg et al. (2002) is plotted as a dotted series for reference.

lost to the crowded-field photometry. The histogram residuals and functional fit can be seen in Figure 6.4, and the resulting function is given empirically as:

$$CDE(g_0) = \begin{cases} 1.0 \text{ if } g_0 < 20.92 \\ 0.0 \text{ if } g_0 > 23.57 \\ 1.0 - 0.14 \cdot (g_0 - 20.92)^2 \text{ otherwise} \end{cases} \tag{6.3}$$

This function can be used to reconstruct incomplete cluster star distributions, and to model the effect of SDSS crowded-field photometry at farther distances. Errors in the parameter fits in Equation 6.3 are 0.14 ± 0.001, and 20.92 ± 0.007.

6.4 Absolute Magnitude Distribution of F Turnoff Stars

6.4.1 Fitting the Turnoff

The absolute magnitude distribution of F turnoff stars in old stellar populations can now be characterized. Stars were selected from the F turnoff region ($0.1 < (g-r)_0 < 0.3$), which is the color range used in the photometric F turnoff density searches in Cole et al. (2008), and the range originally chosen in Newberg et al. (2002) to include stars redder than most blue horizontal branch stars, but bluer than the turnoff of Milky Way disk stars. A histogram is then built in M_g from the cluster data, using a bin size of 0.2 magnitudes over the range $2.0 < M_g < 8.0$, which minimizes potential contamination from non-cluster stars. Each bin is then divided by the cluster detection efficiency (Equation 6.3) for the g_0 that corresponds to the bin's M_g, thereby rebuilding the component of each cluster lost due to the detection efficiency.

The expected number of field stars in each magnitude bin was subtracted from this histogram, as determined from the sky area around each cluster, scaled to the sky area of cluster data. If the cluster lies on the edge of the SDSS footprint, or where sections of the clusters were unresolved by SDSS, the areas with no data were excluded from the area used in scaling the background bins. The magnitude histogram for the background was divided by the field stellar detection efficiency (Newberg et al. 2002) as a function of apparent magnitude.

The top panel in Figure 6.5 shows the distribution of turnoff star M_g absolute

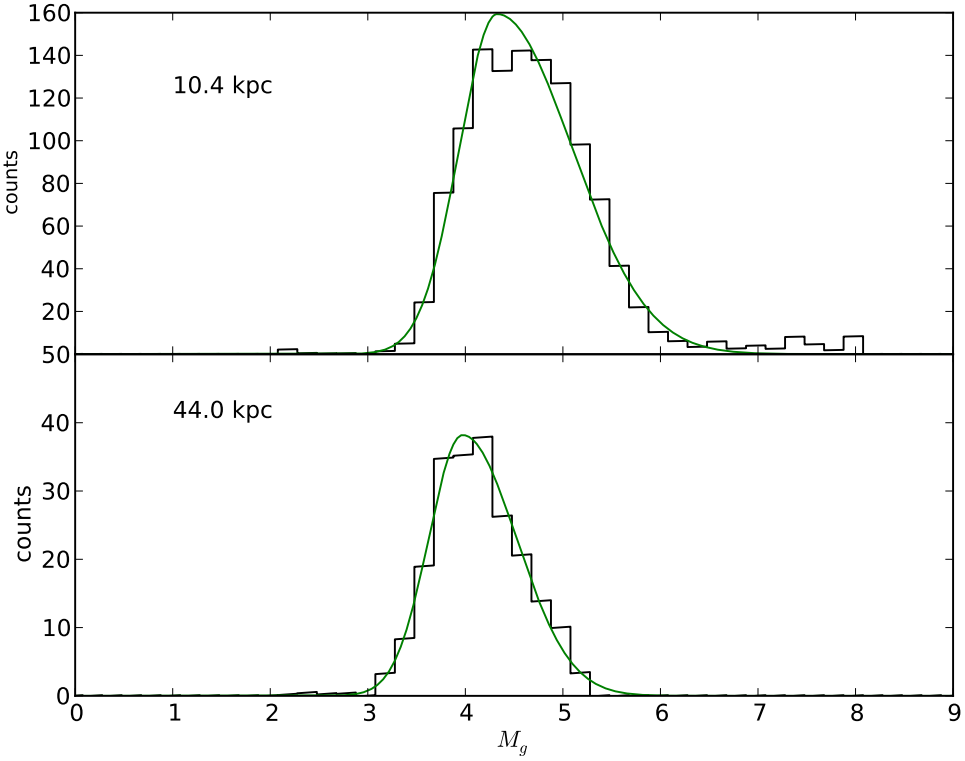


Figure 6.5: Histogram fits for NGC 5272 original data (top) and convolved to to an effective distance of 44.0 kpc (bottom) with the cluster detection efficiency applied. Black lines outline the bins, to which a double-sided Gaussian (green line) was fit. Fit values are (μ,σ_l,σ_r) = (4.33, 0.39, 0.77) for the NGC 5272 raw data, and (3.97, 0.35, 0.55) if the cluster was instead observed at a distance of 44.0 kpc. Note the difference in total counts between the two plots; this is caused by losses due to color errors with effective distance.

magnitudes in globular cluster NGC 5272. Each SDSS cluster histogram was fit with a 'double-sided' Gaussian distribution, where the standard deviation is different on each side of the mean. This choice provides a good fit to the data using a simple, well-known function. There is no theoretical motivation behind this choice of fitting function; however, this function appears to effectively match the form of the data without over-determining the system. The form of the fit function is given by:

$$\mathcal{G}(x;\mu,\sigma_l,\sigma_r,A) = A \cdot exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_i^2}\right] \tag{6.4}$$

where:

$$\sigma_i = \begin{cases} \sigma_l & if M_g \leq \mu \\ \sigma_r & if M_g > \mu \end{cases}$$

When normalized:

$$A = \frac{1}{\sqrt{2\pi}(\frac{\sigma_l+\sigma_r}{2})} \tag{6.5}$$

The fit parameters A, μ, σ_l, and σ_r are the amplitude, mean, left-side standard deviation, and right-side standard deviations, respectively. All bins outside of the range $2.0 < M_g < 8.0$ were set to zero.

Poisson counting errors were assumed when fitting this function to the data histograms. Typically, the approximation of "square-root n" is used to simulate true Poisson errors, but it was found that this over-emphasized low-count bins at the expense of the overall fit. Instead, the Poisson errors were modeled in the fashion of Equation 10 from Gehrels (1986). At the standard 1-σ confidence level, this equation simplifies to $\delta n = 1 + \sqrt{n+1}$.

A two stage process was used in fitting the parameters of the double-sided Gaussian function to the M_g histograms. First a Markov-Chain Monte Carlo technique was used to sweep the parameter space, then the best fit point was fed into a gradient descent algorithm. This two-stage fit avoided local minima and found the true best fit. A sample histogram and functional fit are provided for M3 (NGC 5272) on the top plot of Figure 6.5. Results from the fits to ten clusters can be

Table 6.2: Fit values and errors to cluster data

NGC ID	Messier ID	FWHM	μ	μ error	σ_l	σ_l error	σ_r	σ_r error	A	A error	$(g-r)_{TO}$	$(g-r)_{TO}$ error
NGC 4147	···	1.70	4.42	0.33	0.50	0.22	0.98	0.29	26.6	3.34	0.223	0.022
NGC 5024	M53	1.69	4.49	0.02	0.50	0.02	0.94	0.03	271.8	3.53	0.240	0.010
NGC 5053	···	1.92	4.68	0.07	0.70	0.06	0.93	0.07	105.3	3.00	0.219	0.011
NGC 5272	M3	1.37	4.33	0.05	0.39	0.04	0.77	0.04	154.9	3.71	0.251	0.009
NGC 5466	···	1.92	4.28	0.15	0.40	0.09	1.22	0.13	78.7	3.17	0.217	0.010
NGC 5904	M5	1.23	4.13	0.28	0.32	0.18	0.72	0.19	134.7	4.03	0.253	0.006
NGC 6205	M13	1.30	4.31	0.13	0.31	0.09	0.79	0.10	69.0	4.07	0.251	0.008
NGC 7078	M15	2.18	4.48	0.11	0.62	0.08	1.23	0.11	88.7	3.26	0.219	0.011
NGC 7089	M2	1.27	4.33	0.10	0.31	0.09	0.76	0.12	46.0	4.07	0.258	0.010
Pal 5	···	1.56	4.21	0.14	0.33	0.10	0.99	0.12	54.1	3.55	0.284	0.011
Average	···	1.613	4.367	0.138	0.440	0.097	0.930	0.122	103.0	3.573	0.242	0.011
Standard	Deviation	0.309	0.149	0.092	0.129	0.059	0.173	0.073	67.5	0.367	0.021	0.003

found in Table 6.2. A Hessian matrix, calculated at the best fit parameters, was used to determine the model errors in the parameter fit values. This matrix was multiplied by 2 to make it equivalent to the curvature matrix, then inverted, so that the diagonal elements are equivalent to the squared parameter variances.

To determine the turnoff color, $(g-r)_{TO}$, for a cluster, a Gaussian profile was fit to the $(g-r)_0$ histogram of all stars within 0.5 magnitudes of the μ parameter determined by the turnoff fits. The mean value from this Gaussian is taken as the turnoff value of the cluster. This definition of turnoff color is then consistent with the turnoff color of theoretical isochrones (the bluest point of the isochrone turnoff). Note that "bluest" (minimum $(g-r)_0$ value) point is not necessarily the the turnoff color of the data, since that definition ignores the intrinsic color spread of the data, and could result in a bluer measurement of the turnoff with distance as color errors become larger.

6.4.2 F Turnoff Distribution Results

The F turnoff distribution fit parameters are now examined as a function of the ages, metallicities, and distances to the clusters in the sample. Throughout this section, the turnoff magnitude is describing the SDSS g filter, and the turnoff color is $(g-r)_0$.

In Figure 6.6, no significant correlation is found with cluster age and the fit parameters μ, σ_l, σ_r, and $(g-r)_{TO}$. Intuitively, as a cluster ages μ should move to higher magnitudes (fainter) and the turnoff color should move to smaller (redder) values of $(g-r)_0$. The invariance of these parameters with age is evidence for a fairly uniform globular cluster sample and for the the Age-Metallicity Conspiracy discussed in Section 6.6.

When these fit parameters are plotted vs. metallicity, possible relationships emerge (Figure 6.7). In general, when the metallicity of a theoretical isochrone is increased while holding other variables constant, the turnoff becomes more "pinched" and dimmer - that is, the main sequence and subgiant branch move closer together in magnitude while the turnoff point becomes fainter. The decrease of the σ_l and σ_r parameters with increasing $[Fe/H]$ indicates that "pinching" is occurring, however,

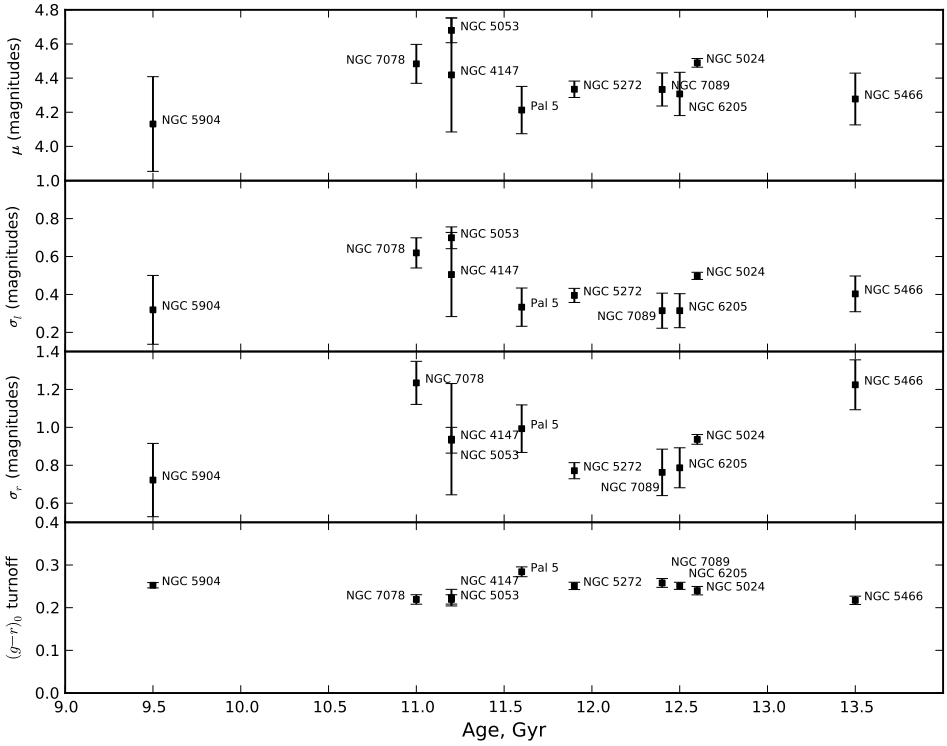


Figure 6.6: Plot of globular cluster μ, σ_l, σ_r and turnoff color values vs. isochrone fit ages. There are no statistically significant trends in our parameters as a function of cluster age. It is expected that the turnoff should shift to redder color and fainter magnitude as it ages, but since these clusters fall along the Age-Metallicity Relationship, the change in metallicity counterbalances the effects of the change in age.

the μ fits *decrease*, showing the opposite behavior from what is expected. This is evidence that metallicity is not the dominant factor in determining turnoff brightness, and will be further explored in Section 6.6. A reddening (increasing) of turnoff color is expected with increasing $[Fe/H]$, and is observed in the fits. Therefore, metallicity is dominant over age in determining the turnoff color. It is important to note that, for the cluster sample, the absolute magnitude distribution has little dependence on metallicity.

Having explored the F turnoff distribution as a function of age and metallicity, associations with distance will now be explored (Figure 6.8). Position in the Galactic halo is not expected to significantly change the structure of a cluster, and so it is expected that all of the turnoff fit parameters will be intrinsically independent of the observed distance for similar clusters. Any parameter dependence on distance is therefore actually a result of observational errors. From Figure 6.8 it is seen that μ and σ_l stay roughly constant with distance, and σ_r rises first, then decreases. The turnoff color ($(g-r)_{TO}$) appears to be invariant with distance. In the next section, the observational errors in the SDSS database are studied, and it is shown that these errors explain the observed changes in the distribution of F turnoff stars as a function of distance.

6.5 Observational Effects of SDSS Errors

6.5.1 Photometric Errors in SDSS Star Colors

Photometric errors widen the observed magnitude and color distributions of turnoff stars in globular clusters as plotted in color-magnitude diagrams, by an increasing amount with fainter magnitudes. The photometric errors in the SDSS database are now examined, and the observational biases that result from these errors are extrapolated. SDSS forgoes the conventional "Pogson" logarithmic definition of magnitude in favor of the "asinh magnitude" system described in Lupton et al. (1999). These systems are virtually identical at high signal-to-noise, but in the low signal-to-noise regime asinh magnitudes are well-behaved and have non-infinite errors. The two magnitude systems do not differ noticeably in magnitude or magnitude error until fainter than 24th magnitude (signal-to-noise $\sim$2.0),

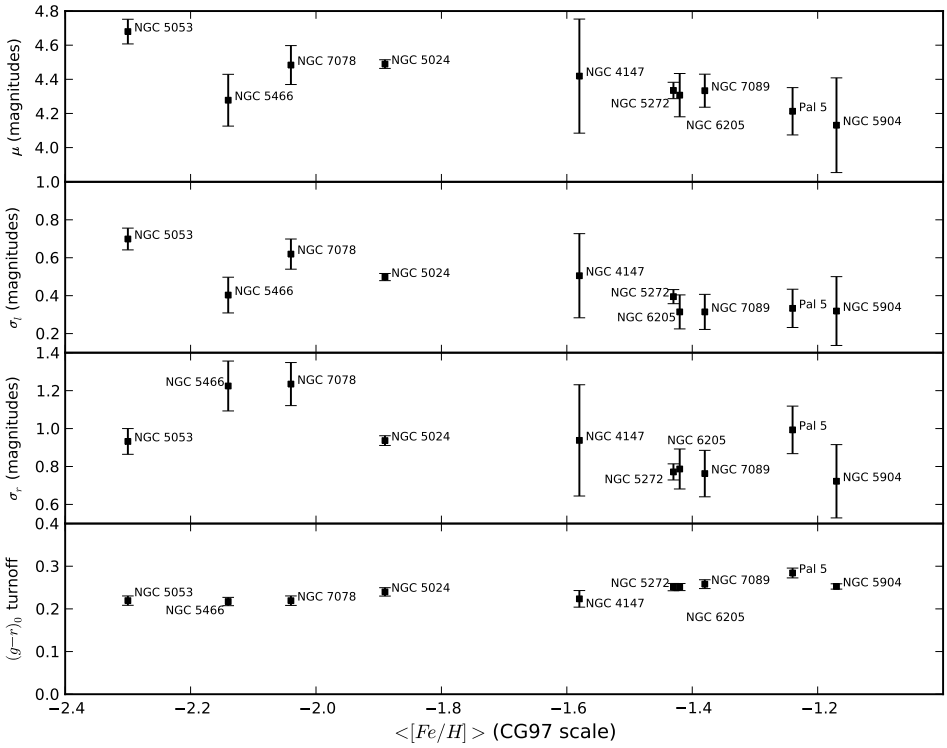


Figure 6.7: Plot of globular cluster μ, σ_l, σ_r and turnoff color values vs. CG97 scale metallicities. The fit parameters μ and σ_l appear to decrease with metallicity, indicating that isochrone fits should have slightly brighter and sharply curved turnoffs as $[Fe/H]$ increases. The fit parameter σ_r does not appear to be correlated with metallicity; We will show that changes in σ_r are primarily due to distance (therefore, error) effects. One would expect an increase in metallicity to produce a fainter turnoff, however, since the more metal-rich clusters are also younger, the turnoff is actually slightly brighter for a higher metallicity star. A slight increase in the turnoff color with metallicity is observed.

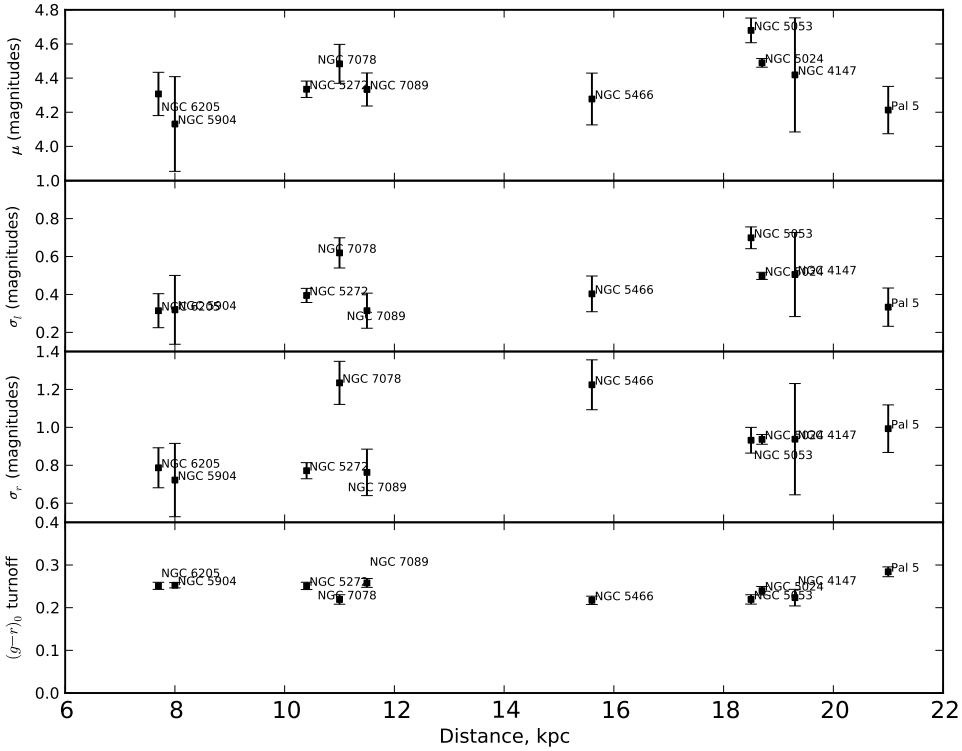


Figure 6.8: Plot of globular cluster μ, σ_l, σ_r and turnoff color values vs. cluster distances. We see that μ appears to increase (become fainter) slightly with distance, σ_l stays roughly constant, and σ_r first increases, then decreases. The turnoff color ($(g-r)_{TO}$) appears to be invariant with distance. Since cluster properties should not be dependent on a cluster's distance from the Sun, any trends in this figure are a result of the increasing magnitude errors with distance.

so this analysis will be valid in both systems. Figure 6.9 shows the relationship between apparent magnitude and magnitude error in the u_0, g_0 and r_0 passbands for stars in the Palomar 5 selection field. An exponential function with an argument that is linear in apparent magnitude is fit to the error vs. magnitude data for each relevant color passband:

$$\epsilon(u_0) = 0.0027 + e^{(0.80u_0 - 19.2)} \tag{6.6}$$

$$\epsilon(g_0) = 0.00031 + e^{(0.79g_0 - 20.0)} \tag{6.7}$$

$$\epsilon(r_0) = -0.000026 + e^{(0.80r_0 - 19.7)} \tag{6.8}$$

It is the negative constant term in the exponential that determines the magnitude at which the errors start to become significant. It is evident from the functional fits and Figure 6.9 that, of the three studied passbands, errors in u_0 grow most quickly, while errors in g_0 are the smallest throughout. Most of the good data is brighter than 23rd magnitude, where g_0 and r_0 magnitude errors are less than 0.3. This will have little effect on the overall appearance of the H-R diagrams, however, magnitude errors of this degree will have a noticeable affect on the observed colors, where the faint magnitude errors are on the order of the observed $(g - r)_0$ values.

6.5.2 F Turnoff Contamination

Since F turnoff stars are selected through color cuts, SDSS color errors are expected to cause the turnoff to be contaminated by misidentified non-turnoff stars. Cluster NGC 6205 was taken as a reference data set, as the relatively low distance (7.7 kpc) implies minimal photometric magnitude errors near the turnoff. Cluster NGC 6205 color errors do not become noticeable until $M_g = 7.0$ ($g_0 = 21.43$); at the absolute magnitude limit of $M_g = 8.0$ ($g_0 = 22.43$), the magnitude errors are close to 0.1 in g_0 and 0.15 in r_0, resulting in a maximum color error at $M_g = 8.0$ equal to 0.18. From Figure 6.2, it is seen that NGC 6205 stars between $7.0 < M_g < 8.0$ are sufficiently red that even for the maximum color error, these stars are statistically

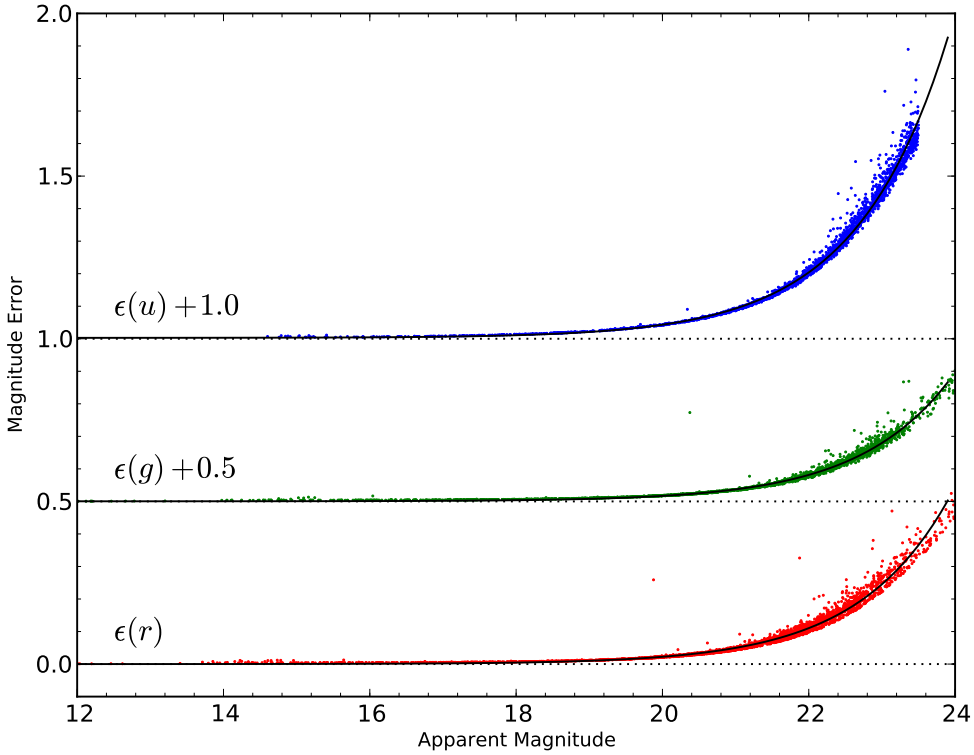


Figure 6.9: Plot of SDSS color magnitudes versus respective color errors. g_0 and and u_0 errors are offset by constants of 0.5 and 1.0, respectively, for illustration. Errors and dereddened photometric values are from the Sloan Digital Sky Survey DR7 database using stars near the globular cluster Palomar 5. It is clear from the plots that the errors in u_0 rise the fastest, followed by the errors in r_0. u_0 data was cut for $u_0 > 23.5$, where the u_0 errors became unreliable. Black lines are functional fits to the data (colored points): $\epsilon(u_0) = 2.71e-03 + e^{0.80u_0-19.2}$; $\epsilon(g_0) = 3.11e-04 + e^{0.79g_0-20.0}$; $\epsilon(r_0) = -2.63e-05 + e^{0.80r_0-19.8}$.

unlikely to to be detected in the turnoff color cut. Therefore, it can be assumed that NGC 6205 will be illustrative as an example of a cluster with an uncontaminated turnoff.

In order to understand how the errors affect a cluster with increasing distance, a process must first be defined that will view a cluster as though it were observed by SDSS at a farther distance, taking into account the increasing magnitude errors as apparent magnitudes increase. The cluster is to be observed at a new chosen "effective distance" ($d_{\rm eff}$), and then a "distance shift" is performed on each star for each of u_0, g_0 and r_0; the magnitude of each star is increased by an amount equivalent to observing the cluster at $d_{\rm eff}$ instead of the original distance (d_0). The magnitude error associated with the new magnitude is then derived from the appropriate error equation (Equations 6.6, 6.7 and 6.8). The original magnitude error is then subtracted in quadrature from the new magnitude error to produce the relative increase in error. The new magnitude value is then modified by the relative increase in error: a random value is selected from a Gaussian distribution with a mean equivalent to the new magnitude, and with a standard deviation equal to the relative increase in error. This random value produces the new shifted magnitude, equivalent to observing the star at $d_{\rm eff}$, including the effects of SDSS magnitude errors.

Having defined the distance shift process, the NGC 6205 cluster data is then separated into three color bins and bin cross-contamination is examined as distance increases. The three color bins are: the primary 'yellow' turnoff bin $0.1 < (g-r)_0 < 0.3$, the 'red' star bin ($(g-r)_0 > 0.3$) and the 'blue' star bin ($(g-r)_0 < 0.1$). Each of these bins were treated as separate data sets. Distance shifts were then performed on each bin at 1.0 kpc steps, up to a maximum $d_{\rm eff}$ of 80.0 kpc. At each step the $(u-g)_0$ color cut was enforced, then the number of stars that remained in their original bin were counted, as were as the number that had leaked into other color bins due to color errors. This process was performed 100 times with the NGC 6205 cluster data, and the results averaged to smooth out potential random errors. This process was then repeated, but this time the field stellar detection efficiency was included in the calculations in order to represent the actual observed turnoff population.

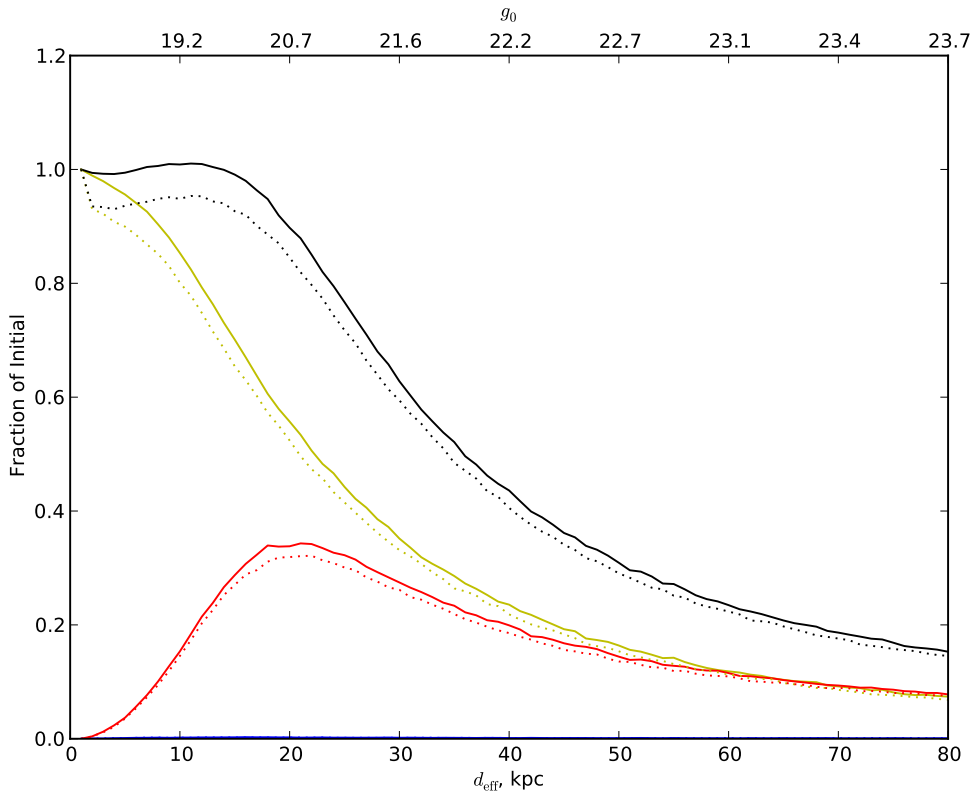


Figure 6.10: Percentage of stars by type detected as F turnoff stars due to SDSS color errors as a function of distance/mean apparent magnitude. The yellow (middle) curve indicates the percentage of 'true,' (original) F turnoff stars detected as F turnoff stars, while the red (lower) curve indicates the percentage of 'red' stars ($(g-r)_0 > 0.3$) incorrectly identified as F turnoff stars due to color errors. The blue curve, which is nearly coincident with the x-axis, represents the same trend but for 'blue' stars ($(g-r)_0 < 0.1$). The black (top) curve is the sum of all lower curves, and illustrates the total percentage of stars detected as turnoff stars relative to the true turnoff distribution. The dotted curves indicate the effect of the field star detection efficiency on star counts. It is clear that the relatively higher color errors at fainter magnitudes, which largely consist of 'red' stars, initially causes a large number of red stars to contaminate the turnoff, which then tapers off as the errors increase. True turnoff stars quickly and continually leak out of the turnoff bin. By as near as 25.0 kpc from the Sun, half of all detected turnoff stars are actually redder star contaminants.

Figure 6.10 illustrates the composition of the selected turnoff stars ($0.1 < (g-r)_0 < 0.3$) as a fraction of the original 'true' turnoff count, and as a function of distance. This calculation assumed 100% detection efficiency for the shifted stars. The results of applying the field stellar detection efficiency is shown by the dotted lines. The trends in this figure are color-coded by the bin of origin; the black trend is a sum of all lower curves, and represents the total number of stars detected as turnoff stars for a given distance. Readily apparent from Figure 6.10 is the quick influx of 'red' stars into the turnoff between distances of (10.0-20.0 kpc), and the constant loss of true turnoff stars with distance.

Red G stars that lie just below the turnoff are on the densely populated main sequence, and have $(g-r)_0$ values just above the turnoff color cut of 0.3. Even slight color error perturbations will tend to shift some these stars into the turnoff color cut, resulting in significant 'red' star contamination at relatively low halo distances. It is clear from Figure 6.10 that this is a rapid effect that occurs at fairly low distances (10.0-20.0 kpc), but as the errors continue to increase it becomes just as likely for a red G star to jump over the turnoff selection box as it is to land in it, thereby causing the red star contamination to stop increasing around 20.0 kpc.

It is interesting to note that the $(u-g)_0$ color cut causes some suppression of the red contamination effect. Since red stars are fainter and farther down the main sequence, they have higher measured magnitudes and magnitude errors than true turnoff stars. As the u_0 passband has the highest associated magnitude errors, faint stars with large errors are perturbed more in $(u-g)_0$ than in $(g-r)_0$. Therefore, faint red stars with large color errors may be perturbed beyond the $(u-g)_0$ cut and subsequently removed from the data set, even if errors would place that star in the $(g-r)_0$ cut for the turnoff. The $(u-g)_0$ cut then serves to remove some of the red contamination stars from the turnoff selection.

As the errors increase, true F turnoff stars can only leak out of the turnoff selection box. Since there are very few bluer ($(g-r)_0 < 0.1$) A-type stars or redder subgiants in globular clusters, and these in any event are bright, few A stars leak into the F turnoff star selection due to errors in color. Around the distance that the red star contamination stops ($\sim$25.0 kpc), the number of stars in the turnoff

selection box is comprised of roughly 60% true F turnoff stars and 40% redder G star contamination. As distances increase, the fraction of true turnoff stars in the color selection range decreases to 50%, and the total number of stars selected as turnoff stars decreases. This is a significant effect that has never been accounted for in previous research papers.

So that future authors may compensate for these effects, we provide analytical functions for the F turnoff dissipation and the red star contamination. These fits are to the 100% detection efficiency case (solid curves in Figure 6.10), representing the effect caused by color errors only. The F turnoff dissipation is represented with a 4th-order polynomial function in $d_{\rm eff}$, and the red contamination is fit with a similar function of 7th-order, with the coefficients given by $\mathbf{a_y}$ and $\mathbf{a_r}$ (where $\mathbf{a} = (a_0, a_1, a_2, ...)$, subscripts corresponding to order of term), respectively[28]:

$$\begin{aligned} \mathbf{a_y} = \quad (1.05628761, \quad -3.14555041e-02, \quad 2.05499665e-04, \\ 2.53747387e-06, \quad -2.67000303e-08) \end{aligned} \tag{6.9}$$

$$\begin{aligned} \mathbf{a_r} = \quad (1.60879353e-02, \quad -1.97164570e-02, \quad 6.60960070e-03, \\ -4.31844102e-04, \quad 1.26368065e-05, \quad -1.91560491e-07, \\ 1.47140445e-09, \quad -4.53857248e-12) \end{aligned} \tag{6.10}$$

These functions are valid in the range of $0.0 < d_{\rm eff} < 80.0$, that is, the range of the distance shifts used in the above analysis. Note that more digits are included in these coefficients than are scientifically significant; this is because polynomial fit coefficients are highly sensitive to rounding, such that the output values are radically different between functions with rounded coefficients and those with all trailing digits.

[28]Here "e" indicates that the scientific exponent follows, such that $1.0e-02 \Rightarrow 1.0 \times 10^{-02}$.

6.5.3 The Effects of Magnitude Errors on the Distribution of M_g

In order to study the effects of SDSS errors on the measured fit parameters, distance shifts (see previous section) were performed on the u_0, g_0 and r_0 magnitudes of each cluster at varying $d_{\rm eff}$ steps, to a maximum of 44.0 kpc. At every $d_{\rm eff}$, the $(u-g)_0 < 0.4$ color cut was applied in order to remove stars that would have been removed as if this data had been selected from the SDSS database.

At each distance shift the background subtraction and detection losses were taken into account. Since distance shifts must be performed on a dataset of individual stars, while the correction functions must be applied to histogrammed data, distance shift was performed before the background subtraction and detection efficiency correction. In order to keep the background subtraction consistent with a new, effectively more distant cluster data set, applied an equivalent distance shift was applied to the background prior to binning and subtraction. This reproduces the effect of subtracting the background prior to the shift. The resulting histogram was then divided by the cluster detection efficiency, but with the parabola center shifted with the cluster to the new magnitude. If the detection efficiency function was not shifted prior to dividing, it would be applied to the wrong portion of the cluster histogram, since the cluster has been shifted to higher magnitudes.

Before performing the functional fit to our shifted data, one of three observational biases were applied to the M_g histogram: The cluster detection efficiency detailed in Section 6.3, the stellar detection efficiency described in Newberg et al. (2002), and 100% detection efficiency (in which no correction is applied). The first bias system reveals the evolution with increasing distance of globular cluster turnoff distributions as observed in SDSS data. The second bias system will produce the evolution of non-cluster turnoff distributions as a function of distance. The final system reveals the evolution of turnoff distributions if no detection bias is applied, that is, if all of the stars originally detected in a cluster continue to be detected as the distances increase.

All ten clusters fit in Section 6.4 were distance shifted to increasingly greater $d_{\rm eff}$ values, using the methods outlined above. At each $d_{\rm eff}$, a new set of fit parameters for the observed absolute magnitude distribution were evaluated using the methods

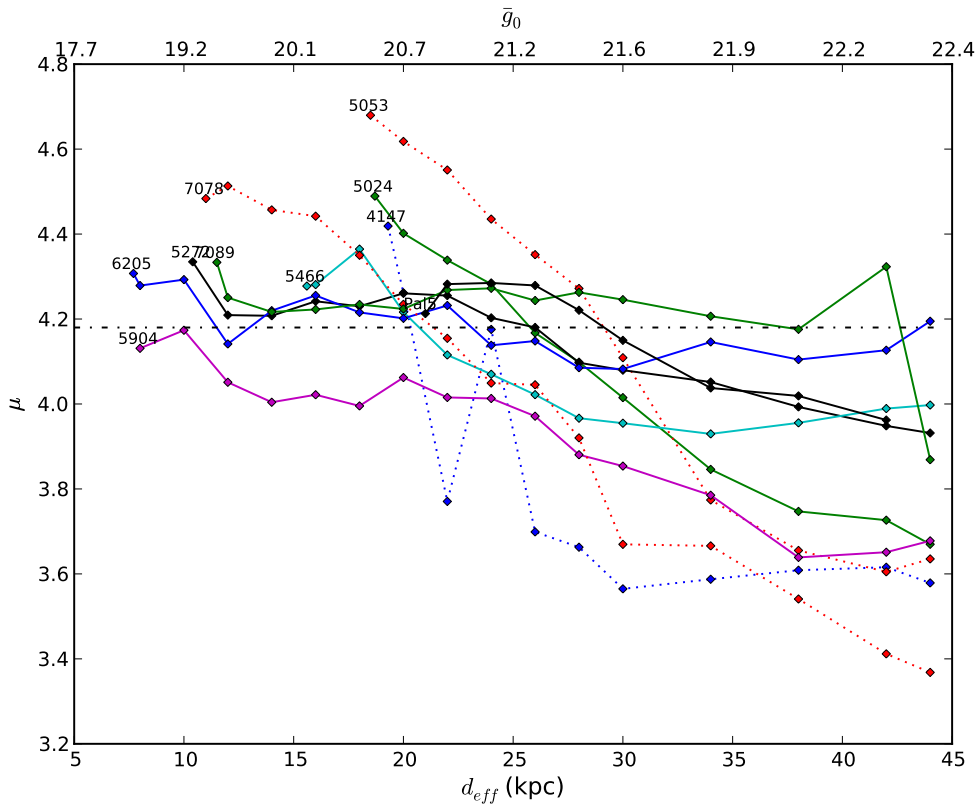


Figure 6.11: μ series fits to ten globular clusters, convolved to errors consistent with observing them at larger distances ($d_{\rm eff}$). The cluster detection efficiency (Equation 6.3) was applied during the distance shifts. Although the fit μ values decrease with distance, the errors increase as well. To within errors, the μ values are consistent with a constant value. The red dotted series have been rejected as outliers. The blue dotted series for NGC 4147 indicates large expected errors due to low star counts. The error-weighted average of initial μ points for the remaining globular clusters is 4.18($\pm$ 0.008), plotted as the black dot-dash line.

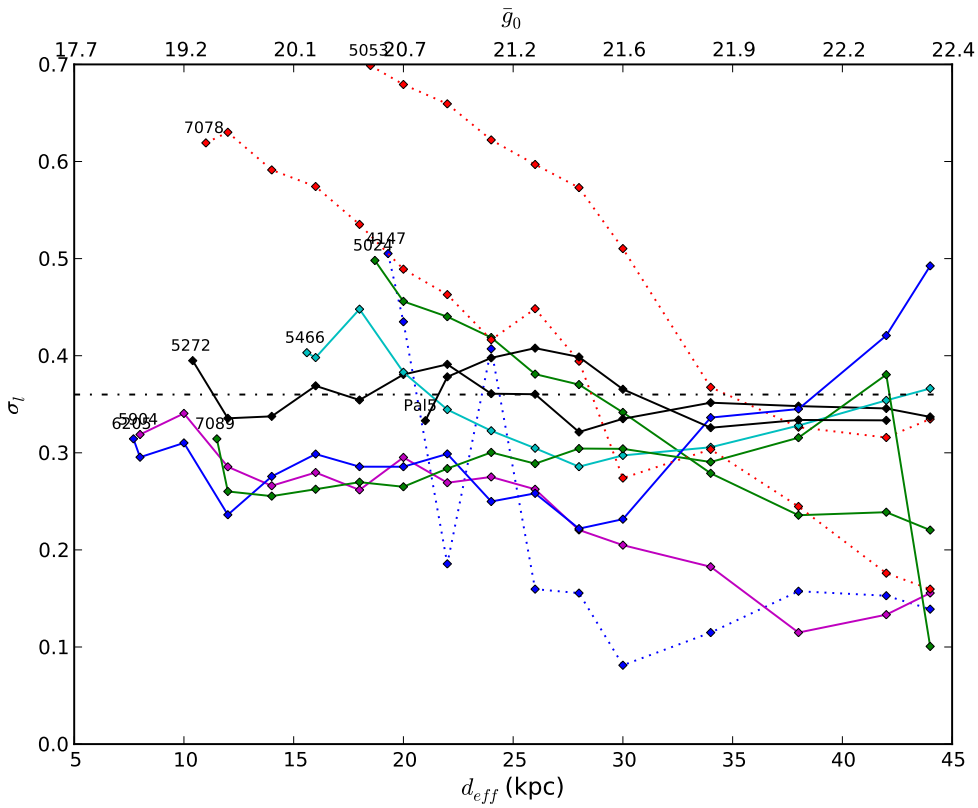


Figure 6.12: σ_l series fits to ten globular clusters, convolved to errors consistent with observing them at larger distances ($d_{\rm eff}$). The cluster detection efficiency (Equation 6.3) was applied during the distance shifts. To within errors, the σ_l values stay constant. The red dotted series have been rejected as outliers. The blue dotted series for NGC 4147 indicates that large errors are expected due to low star counts. The error-weighted average (ignoring outliers) of initial σ_l points for the remaining clusters is 0.36($\pm$ 0.006), plotted as the black dot-dash line.

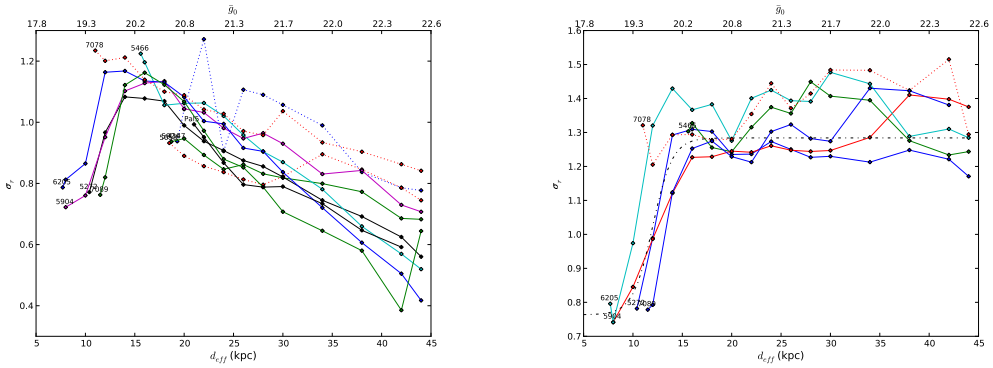


Figure 6.13: σ_r series fits to ten globular clusters, convolved to errors consistent with observing them at larger distances ($d_{\rm eff}$). The left plot represents clusters convolved while applying the SDSS crowded-field photometry detection efficiency function (Section 4). When this detection efficiency is applied, all clusters in our analysis follow the same trend (reproducing the observed dependence of σ_r with distance) as they are convolved to farther $d_{\rm eff}$. The right plot contains only clusters not initially affected by the detection efficiency, and did not have it applied during the convolutions, since it was not possible to remove the crowding effect from the more distance cluster data. This represents the ideal observation, in which no SDSS detection losses occur beyond the initial status of the cluster, but with only photometric errors increasing with distance. In both cases a quick rise in σ_r with distance is observed, due to the large influx of redder main-sequence stars into the F turnoff cut, as described in Figure 6.10. In the left plot, the detection efficiency continually cuts into the fainter stars, causing a constant fall in σ_r after the quick rise. When there is no loss due to detection efficiency (right), σ_r remains constant after the initial rise. To this trend we fit a sigmoid function (dashed line, see text): α, β, and γ values of 0.52($\pm$ 0.04), 12.0($\pm$ 0.31), and 0.76($\pm$ 0.04), respectively. Dotted lines indicate clusters that have been treated as outliers.

described above. An example of the histogram and fit of cluster M3 (NGC 5272), shifted to a $d_{\rm eff}$ of 44.0 kpc, is presented in the lower plot of Figure 6.5.

The results of the distance shifted fits for the parameters μ, σ_l, and σ_r in are presented in Figures 6.11, 6.12, and 6.13 respectively. NGC 5053 and M15 (NGC 7078) are plotted with a red dotted series to indicate their status as expected outliers, while NGC 4147 is plotted with a blue dotted series to indicate that it contains few stars and therefore the fits contain large errors.

6.5.4 Observed vs. Intrinsic M_g Distribution of F Turnoff Stars

It is important to note that all of the clusters studied, including suspected outliers, have similar μ and σ_l values despite differences in distance, age and metallicity. Although there are differences in σ_r, it has been shown these are due to photometric errors, and not differences in the absolute magnitude distribution of turnoff stars in globular clusters. This implies that the halo cluster population is intrinsically similar throughout. In the next section we will show that this similarity is related to the Age-Metallicity Relationship.

In Figure 6.14 we show a series of fits to the turnoff star magnitude distribution for nearby cluster NGC 6205 at increasing effective distances, including the effects of the cluster detection efficiency. From this plot, the most obvious effect is the loss of stars with distance; however, one can see that these losses balance to cause σ_l to stay constant throughout, and μ shifts slightly to the left (brighter magnitudes) as the detection efficiency cuts in with distance.

μ is found to be approximately constant with distance, regardless of the applied detection efficiency bias. From the plot of μ vs $d_{\rm eff}$ (Figure 6.11), it is seen that μ values decrease slowly with increasing distance; however, μ fit errors increase quickly with distance due to the loss of turnoff stars. Within the fit errors, μ is adequately described by a constant value. The error-weighted constant-value fit to μ gives $\mu = 4.18(\pm\ 0.008)$, with a cluster dispersion of 0.073. Also provided is a linear fit to μ with distance, and a slope of zero (which would imply a constant value) is within the one-σ errors:

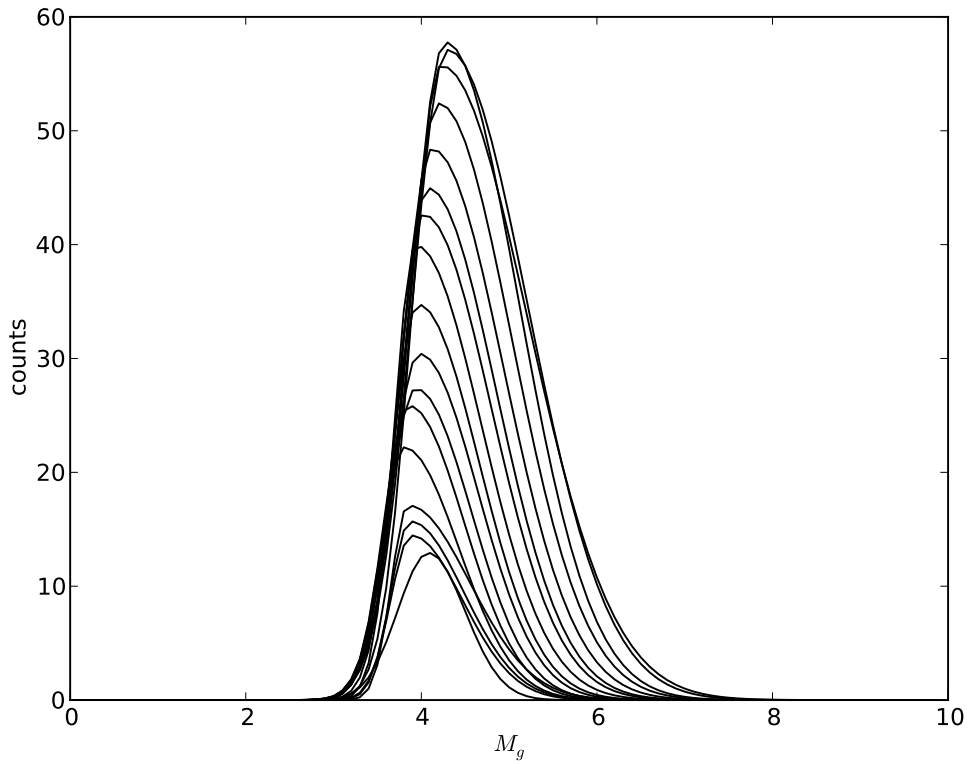


Figure 6.14: Double-sided Gaussian fits to NGC 6205 turnoff ($0.1 < (g-r)_0 < 0.3$) histograms with increasing effective distance ($d_{\rm eff}$). The third tallest Gaussian is at the original distance of 7.7 kpc, while the remaining Gaussians are fits to the data shifted to $d_{\rm eff}$s of 8.0 kpc through 28.0 kpc, in 2.0 kpc increments. Initially, the Gaussian fits become taller and σ_r becomes larger as redder contamination stars enter the data set. Then the peak of the histogram becomes smaller as the cluster moves to larger $d_{\rm eff}$ values, where increasing color errors cause a bleed-off of turnoff stars. Note that as the cluster is shifted beyond 20.0 kpc (fourth curve from the bottom), the σ_r value eventually begins to shrink as the cluster detection efficiency removes fainter stars from the data.

$$\mu(d_{\rm eff}) = -0.011(\pm 0.02)d_{\rm eff} + 4.39(\pm 0.57) \tag{6.11}$$

The σ_l values also stay constant with distance, regardless of the applied detection efficiency, to within calculated errors (Figure 6.12). It is also apparent that the two expected outlier clusters NGC 5053 and NGC 7078 have σ_l behaviors that differ from the other clusters. An error-weighted average to the σ_l values, excluding the two outliers, give $\sigma_l = 0.36(\pm\ 0.006)$, with a cluster dispersion of 0.18.

The values of σ_r do not stay constant with the distance shifts for any of the three detection efficiency cases. Because of increasing color errors with distance, all but the nearest turnoff star populations are contaminated by redder main-sequence stars, as described in Section 6.5.2. These redder stars enter the turnoff histograms on the fainter ($> \mu$) side, thereby widening the overall distribution and increasing σ_r while leaving the other two parameters unchanged. The nearest clusters (NGC 5272, NGC 5904, NGC 6205, NGC 7089; excluding the core-collapsed NGC 7078) are near enough that they do not exhibit significant red main-sequence contamination, and are consistent with each other. These nearby, uncontaminated clusters are representative of the "true" globular cluster distribution, with a σ_r fit of $0.76(\pm\ 0.04)$, equivalent to the fit value of γ, below.

When the SDSS detection efficiency is applied to globular clusters, all of the clusters show the same behavior as a function of distance (Figure 6.13, left). The initial, quick rise in σ_r with distance is due to a large influx of red main-sequence stars due to color errors. If the cluster is observed at even farther distances, σ_r is reduced due to the cluster detection efficiency removing increasingly larger portions of the faint edge of the turnoff. This consistent series in σ_r is evidence that the observed spread in initial cluster σ_r values is not a real feature of the clusters, but is instead an observational bias due to the incompleteness of SDSS crowded-field photometry at faint magnitudes. Figure 6.15 illustrates a fit to the variation of σ_r with distance.

In order to see what happens to the distribution of turnoff stars as a function of distance for SDSS field stars, the SDSS stellar detection efficiency is also applied when distance-shifting the cluster. The nearby clusters (NGC 5272, NGC 5904,

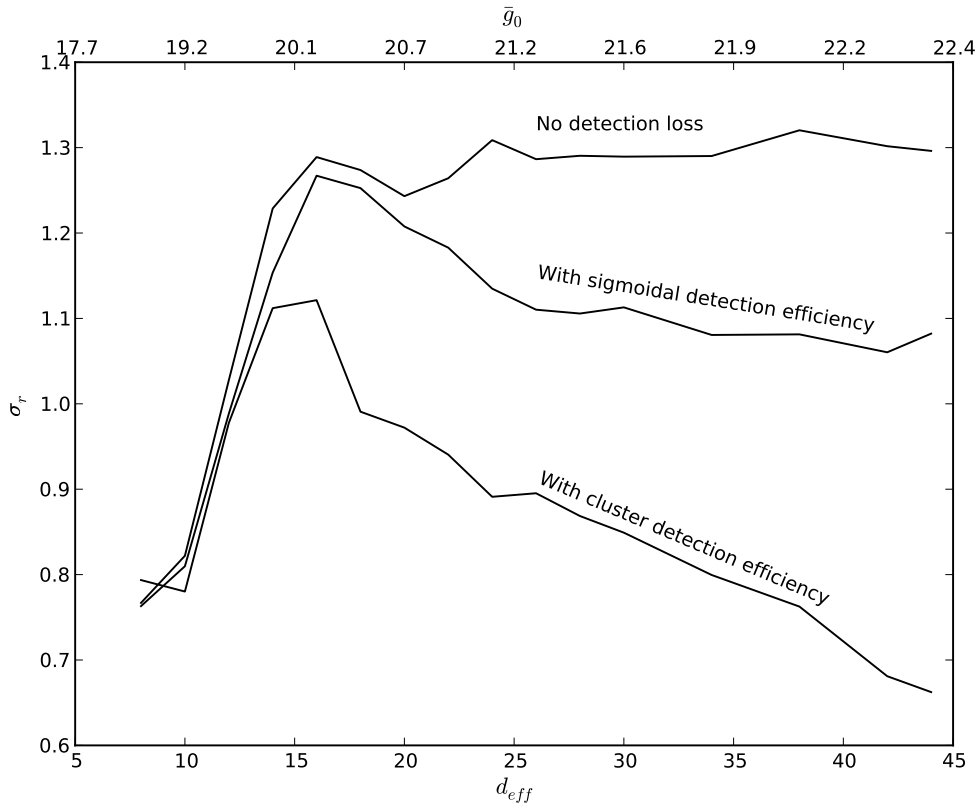


Figure 6.15: Weighted averages of σ_r series fits with effective distance ($d_{\rm eff}$) for three different detection efficiency functions: no detection efficiency applied, (100% detection, upper curve) the sigmoidal SDSS detection efficiency from Newberg et al. (2002), (middle curve) and the SDSS crowded-field photometry parabolic cluster detection efficiency discussed in Section 6.3 (lower curve).

NGC 6205, NGC 7089, and NGC 5466) are found to follow a similar pattern as in the cluster detection efficiency system, while the initially more distant clusters (NGC 4147, NGC 5024, NGC 5053, Pal 5) maintain a relatively constant value for σ_r. The initially more distant clusters are already incomplete due to the cluster detection efficiency, which has modified their apparent σ_r fit. For the nearby clusters, the initial rapid rise in σ_r with distance peaks at a higher value of σ_r, and the subsequent gradual decline is less severe. This is due to the wider and fainter drop off for the stellar detection efficiency as compared to the cluster detection efficiency; that is, the stellar detection efficiency starts removing fainter stars at greater distances, and to a less severe degree. As the stellar field detection efficiency is the dominant observational bias in the SDSS, a 4th order polynomial fit to σ_r versus distance was performed, with the coefficients given by $\mathbf{a}_{\rm sdss}$:

$$\begin{aligned}\mathbf{a}_{\rm sdss} = \quad (-1.74163321, \quad & 0.457079636, \quad -0.0250001309, \\ & 5.72077958e-04, \quad -4.70331735e-06)\end{aligned} \tag{6.12}$$

If one assumes the 100% detection efficiency at all magnitudes, it is found that nearby clusters see a quick rise in σ_r as in the previous two cases, but then level out at a constant value (Figure 6.13, right; Figure 6.15), while farther clusters remain constant (as in the previous case). As discussed in Section 6.5.2, the ratio of "true" turnoff stars to red contaminants remains constant after the initial influx of red stars into the turnoff color cut. Since this ratio remains constant, there is no appreciable change to the observed turnoff distribution after the initial rise. Figure 6.15 shows the difference in σ_r evolution for different detection efficiency cases.

The sudden inflow of red turnoff contaminants between 10.0 and 15.0 kpc is responsible for the rapid rise in σ_r fits for all three detection efficiency systems. The nearby clusters (NGC 5272, NGC 5904, NGC 6205, and NGC 7089) are close enough ($\leq$ 11.5 kpc) to not contain significant turnoff contamination; they are therefore representative of the intrinsic σ_r value for globular clusters. When the 100% detection efficiency is assumed at all magnitudes, the deviation of σ_r fits from this intrinsic value is then purely an effect of the color errors due to magnitude,

and are not influenced by pre-existing red star contamination or detection efficiency losses. These σ_r trends then serve as a basis for understanding the effects of color errors on the observed distribution of turnoff stars, and so a sigmoid functional fit to the 100% detection efficiency case is provided:

$$\sigma_r = \frac{\alpha}{1 + e^{-(d_{\rm eff} - \beta)}} + \gamma \tag{6.13}$$

Where $d_{\rm eff}$ is the effective distance of the shifted cluster. The best fit values for the sigmoid functional fit to the nearby clusters are $\alpha = 0.52(\pm\ 0.04)$, $\beta = 12.0(\pm\ 0.31)$, and $\gamma = 0.76(\pm\ 0.04)$. The three different trends for σ_r with distance, produced by the three separate detection efficiency systems, are compared in Figure 6.15.

It has been shown that the intrinsic turnoff distribution, as quantified by the fit parameters μ, σ_l, and σ_r, is similar for all observed clusters, but that the measured value of σ_r depends on the distance to the cluster due to photometric errors that increase if the cluster is farther away. Stars that affect σ_l are bright, and therefore have smaller color errors than stars that affect σ_r, and will therefore not be shifted into or out of the turnoff color cut as easily. Also, there are few stars adjacent to the brighter region of the turnoff, so potential contamination is low. Finally, detection efficiency biases must effect the fainter turnoff stars before they can affect the brighter stars. Therefore, bright turnoff stars are well insulated from contamination and detection bias. The dominant process affecting σ_l, then, is the loss of turnoff stars due to color errors. Turnoff star losses also affect μ, but at large distances μ will also be modified by detection biases.

6.6 The Age-Metallicity Conspiracy

The age-metallicity relationship (AMR) describes the observed relationship between a cluster's age and its average metallicity. Recent observational work (De Angeli et al. 2005; Marín-Franch et al. 2009; Dotter et al. 2011) indicates that there are two such AMRs present in the Milky Way: one in which metallicity increases with age, attributed to clusters and dwarf galaxies that were that were gradually captured by the Milky Way over time (considered by these authors to

be "outer halo" clusters); and a set of clusters with age 13 Gyr that spans a very large range of metallicities, believed to be due to old clusters that formed rapidly alongside the Milky Way during the initial formation event (considered to be "inner halo" clusters).

In Figure 6.16, it is shown that the clusters in this study follow the trend of metallicity decreasing with age, which we will henceforth refer to as *the* AMR, as presented in Muratov & Gnedin (2010). Muratov & Gnedin provide a galaxy-independent model of galaxy formation history using semi-analytic models that take into account cosmological simulations. Note that the only two clusters in the sample that are not good matches to Muratov & Gnedin have already been identified as outliers in previous discussions. Since all of this study's cluster sample are located at high Galactic latitudes, and are likely members of the Galactic halo (and thought to have been accreted during galaxy assembly), it is not surprising that these clusters are similar to the Muratov & Gnedin (2010) relations. Figure 6.16 shows that the Muratov & Gnedin AMR model does not reproduce the constant (old, rapid co-forming clusters) metallicity trend described by De Angeli et al. (2005), Marín-Franch et al. (2009), Dotter et al. (2011)[29], but that it is consistent with the AMR (young, late accretion) cluster values from this study.

Using the Muratov & Gnedin AMR and the modified Padova isochrones, it is shown that they predict the absolute magnitudes of turnoff stars will be similar for old stellar populations that follow the AMR. Taking the approximate mean metallicity at range of ages over 8 Gyr from Figure 8 of Muratov & Gnedin (2010) (plotted in Figure 6.16 as the solid line), a set of metallicity versus age points are produced that are representative of the AMR. The Padova isochrones (modified by the linear correction function in Equation 6.1) are shown in Figure 6.17. The blue turnoff point can be found between $0.20 < g - r < 0.23$, the turnoff magnitude lies between $3.77 < M_g < 4.16$, and the subgiant branch is constrained between magnitudes of $3.4 < M_g < 3.6$. Also shown in Figure 6.17 are two isochrones generated for age-metallicity combinations far from the AMR. These illustrate that the tight grouping of the series is a property of clusters on AMR, and is not true

[29]Dotter et al. (2010) used Zinn & West metallicities; these were converted to the Carretta & Gratton scale in Figure 6.16 to be consistent.

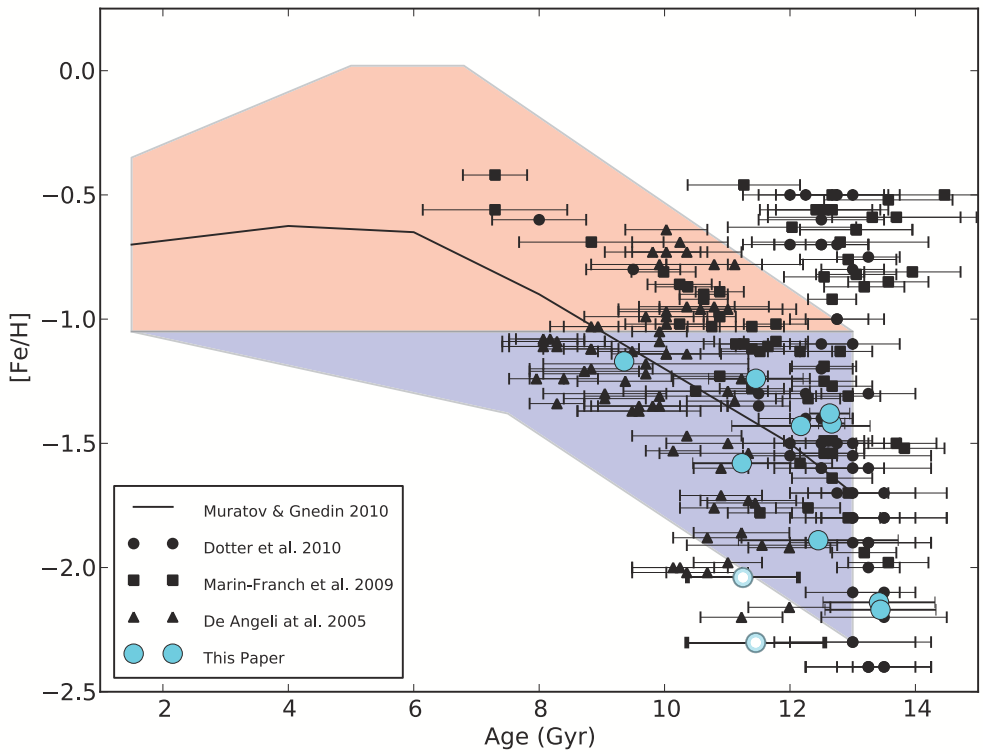


Figure 6.16: Plot of globular cluster ages versus metallicity, from four different sources: this paper (blue circles), Dotter et al. (2010); De Angeli et al. (2005); Marín-Franch et al. (2009). Two clusters, NGC 5053 and NGC 7078, are shown as blue rings to indicate that they are outliers in our analysis. These are overlaid on a theoretical age-metallicity relationship from Muratov & Gnedin (2010) (blue- and red-shaded areas). Note that most of the clusters, including all of the clusters in this study besides noted outliers NGC 5053 and NGC 7078, are consistent with the theoretical age-metallicity relationship. Note that there are some high age, high metallicity clusters in the Marín-Franch et al. (2009) and Dotter et al. (2010) data that they attribute to a constant age with metallicity relationship at around 13 Gyr in the inner halo.

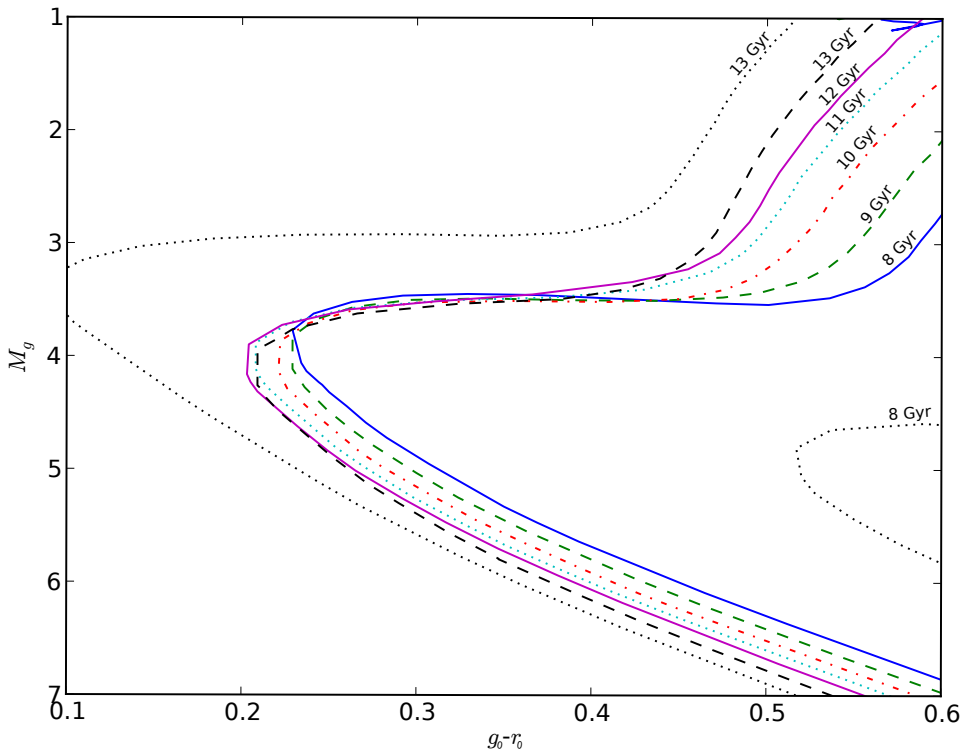


Figure 6.17: Plot of Padova isochrones, modified by a correctional fit (Equation 6.1), from the Age-Metallicity Relationship (AMR) presented in Muratov & Gnedin (2010). Two isochrones not on the AMR are also plotted as black dotted series, at 8 Gyr and 13 Gyr, in order to show the behavior of extreme outliers. All of the isochrones on the AMR have turnoff values that are similar to each other, with a turnoff color between between $0.2 < (g-r)_0 < 0.23$ and a turnoff magnitude of $3.77 < M_g < 4.16$. The subgiant branch is also well constrained in magnitude: $3.4 < M_g < 3.6$. These constraints will be useful in determining distances to old stellar populations. The following Age (Gyr) and [Fe/H] (dex) value sets were used to produce the AMR isochrones: 13.0, -1.7; 12.0, -1.5; 11.0, -1.35; 10.0, -1.20; 9.0, -1.05; 8.0, -0.90. The age, [Fe/H] values for outlying isochrones are: 13.0, 0.0; 8.0, -2.0.

for arbitrary combinations of age and metallicity.

The turnoff parameter fits to the unmodified cluster data in Table 6.2 are consistent with the trends seen in Figure 6.17. Along the AMR, as age decreases and metallicity increases, the turnoff moves to slightly brighter magnitudes (lower μ) and slightly redder colors (higher $(g-r)_0$). This shows that the AMR produces clusters with similar isochrones, but the age of a cluster slightly dominates over the metallicity in determining the g_0 turnoff brightness, while the metallicity slightly dominates over the age in determining the $(g-r)_0$ turnoff color.

This finding implies that age and metallicity values along the AMR "conspire" to produce a similar population distribution for all older (8.0+ Gyr) clusters. As a cluster ages, it's turnoff becomes redder since it has had more time for stars to evolve off of the main-sequence. However, an increase in metal content also corresponds to a redder value for the entire isochrone. The plot of isochrones along the theoretical AMR in Figure 6.17 implies that these two effects almost exactly cancel each other out, and therefore old stars formed in accordance with the AMR should fall in a very narrow range on a color-magnitude diagram. Provided one does not select a cluster from the old ("inner halo") population, this finding simplifies distance measurements for the predominantly old stellar population in the Galactic halo, but complicates the age determination process, which depends on the uniqueness of isochrones. We removed this complication by using only metallicities determined from spectra; only age and distance were used free parameters in the isochrone fits.

6.7 Application to SDSS Data

This chapter has thusfar described SDSS photometric errors and the resulting effects on observed F turnoff distributions ($0.1 < (g-r)_0 < 0.3$). In this section a method will be outlined by which SDSS observations can be corrected for these effects. If one sums the polynomials whose coefficients are presented in Equation 6.9 and Equation 6.10, it will produce a function that describes the ratio of stars selected as turnoff stars to the number of actual turnoff stars present, as a function of distance

$$\varepsilon(r) = \frac{n(r)}{n_0} = \sum_{i=0}^{7} (\mathbf{a_{yi}}(r) + \mathbf{a_{ri}}(r)) r^i \tag{6.14}$$

where r is the distance to the stellar population. Note that this equation assumes a $(u-g)_0$ cut to the data; if this cut is not performed, then additional stars will be included in the data set and this equation will not be valid. Instead, the following equation should be used:

$$\varepsilon(r) = \frac{n(r)}{n_0} = \sum_{i=0}^{7} (\mathbf{b_{yi}} + \mathbf{b_{ri}}) r^i \tag{6.15}$$

$$\begin{aligned} \mathbf{b_{yi}} = \quad (1.02299501, \quad -1.11841006e-02, \quad -4.25644389e-04, \\ 9.71742045e-06, \quad -5.49938115e-08) \end{aligned} \tag{6.16}$$

$$\begin{aligned} \mathbf{b_{ri}} = \quad (3.40566382e-02, \quad -3.95498538e-02, \quad 1.15149959e-02, \\ -7.12231940e-04, \quad 1.99285419e-05, \quad -2.89890541e-07, \\ 2.14094767e-09, \quad -6.35861985e-12) \end{aligned} \tag{6.17}$$

For simple density searches that ignore F turnoff distribution statistics, dividing the observed density at a given distance by $\varepsilon(r)$ is sufficient to correct data for missing turnoff stars due to magnitude errors. For example, if one measures the number of background-corrected F turnoff stars in a section of the Sagittarius tidal debris stream at a distance of 45 kpc, that number should be corrected for completeness (if necessary) and then divided by ε(45 kpc) in order to produce the true number of turnoff stars present in that field.

For statistical models that seek to quantify F turnoff star densities in the Galactic halo, a more complicated correction method is required. In Cole et al. (2008), the authors sought to map the Sagittarius tidal debris stream by performing a maximum likelihood fit of a statistical model to F turnoff stars taken from the SDSS data. They include the effect of the turnoff star distribution by convolving a normalized Gaussian, representing the turnoff star distribution (mean=4.2, σ =0.6), with a halo density model in which all of the turnoff stars have the same absolute magnitude (M_g =4.2). A correction for the stellar detection efficiency is subsequently applied.

The methods of Cole et al. (2008) are now modified by incorporating these results into their analysis. In order to include the turnoff star distribution pre-

sented in this paper, the turnoff distribution in the density convolution must be changed from a symmetrical Gaussian ($\mathcal{N}$ in that paper) to the double-sided Gaussian, $\mathcal{G}(g_0; \mu, \sigma_l, \sigma_r, A)$ presented in Equations 6.4 and 6.5, with a σ_r given by Equation 6.13. Note that the 100% detection efficiency function is used for σ_r, as the stellar detection efficiency is applied subsequent to the turnoff convolution. The normalization for $\mathcal{G}$ must be adjusted for the fact that turnoff stars are lost (and a smaller number of G stars are gained) at farther distances. To account for the turnoff dissipation and contamination, the distribution $\mathcal{G}$ should be multiplied by $\varepsilon(g)$. Putting this together, a convolution recipe is produced which provides a statistical description of the Galactic halo turnoff population, including the effects of the double-sided Gaussian magnitude distribution and the effects of dissipation and contamination with distance. Equation 14 in Cole et al. (2008) is then replaced by:

$$\rho(g_0, \Omega) = \int_{-\infty}^{\infty} dg \; \varepsilon\left(r(g, \mu)\right) \rho_\mu(g, \Omega) \mathcal{G}\left(g_0 - g; \mu, \sigma_l, \sigma_r(r(g, \mu))\right) \tag{6.18}$$

where ρ represents a density function of magnitude and solid angle (Ω), $\rho_\mu(g, \Omega)$ is the model turnoff density function, in which all stars are assumed to be at absolute magnitude μ (4.18), ε is Equation 6.14, and $\mathcal{G}$ is Equation 6.4 with Equation 6.13 as the input for σ_r. ε and σ_r are both functions of r, which is equal to the distance ($r = 10^{\frac{g_0 - \mu - 10}{5}}$).

Using these examples as templates, and with the information contained in the rest of this document, future authors should be able to adapt these results to correct their models for observational biases due to turnoff magnitude errors. Future authors must be prepared to compensate for three mechanisms that cause turnoff star incompleteness and misidentification: The combined effects of turnoff bleed-out and red star contamination, given by Equation 6.14; the SDSS stellar field detection efficiency, given in Newberg et al. (2002); and the intrinsic magnitude distribution of turnoff stars within the Galactic halo, as discussed in Section 6.5 and previously in this section.

6.8 Discussion

6.8.1 Effects of Abundance Variation

Recent observational studies (Pritzl et al. 2005; Roederer 2009) (see also reviews by Gratton et al., 2004 and McWilliam, 1997) indicate that both outer halo globular clusters and halo field stars have similar alpha-element abundances, consistent with an average [α/Fe] value of 0.3. This is strong evidence that halo clusters and field stars share a similar formation history, and it is therefore reasonable to assume that they share similar population distributions and photometric properties. These studies also find that halo clusters and stars are chemically distinct from dwarf spheroidal galaxies (Venn et al. 2004) and the galactic disk, including stars in the thick disk. While thick disk stars can be removed from SDSS photometric data through magnitude and color cuts, dwarf spheroidal stars may be present in the halo due to tidal disruptions of infalling dwarfs. If significantly different [α/Fe] affect the photometric properties of a population, disrupted dwarf spheroidal galaxies may pose a problem for this technique. However, the success of Cole et al. (2008) in mapping segments the Sagittarius tidal stream with a similar technique is evidence that that dwarf galaxy turnoff star distributions are not significantly different from those in the globular clusters in this study.

The set of globular clusters in this study are limited to old, metal poor Milky Way halo clusters, with a maximum metallicity of [Fe/H]= -1.17 and a minimum age of 9.5 Gyr (NGC 5904). Figure 6.17 indicates that these results can be extended to an age of at least 8.0 Gyr. From Figure 6.16 it can be seen that the entire cluster sample falls within the "metal poor" (blue) region of the Muratov & Gnedin AMR. While these results are well-suited to the goal of describing the old, metal-poor halo of the Milky Way, this work did not test how far it can be extrapolated to stellar populations less than 8 Gyr old or more metal-rich than NGC 5904.

However, recent work by Grabowski et al. (2013) extended this analysis to two additional clusters, Pal 13 and Whiting 1, with data taken from SDSS DR8 in the South Galactic Cap. Both clusters were found to be consistent with the Milky Way AMR, with Pal 13 having [Fe/H]=-1.7 and and age of 12 Gyr, and Whiting 1 having [Fe/H]=-0.65 and an age of 6.5 Gyr. Despite being consistent with the

AMR, Pal 13 was found to be only marginally consistent with the results of this study. It was determined that the the Pal 13 data contained large errors (due to the low number of stars detected), especially on the faint side of the turnoff, and so no definite conclusions could be reached regarding this cluster. Whiting 1 was found to be consistent with the previously derived absolute magnitude distribution for halo stars. This allows the results to be extended to clusters on the AMR that are as young as 6.5 Gyr (or as metal-rich) as Whiting 1.

The results of this study may not apply to globular clusters with age and metallicity values that fall far from the Milky Way AMR. These results may not apply to stellar populations in other galaxies, which have differing assembly histories and potentially different age-metallicity relationships. Indeed, the two unusual globular clusters that fall outside of the AMR in Figure 6.16, are the two clusters that differ from each other in turnoff magnitude (Figure 6.10).

Recently, multiple star formation periods have been detected in globular clusters (Milone et al. 2008; Bedin et al. 2004; Piotto et al. 2007), (see review by Piotto, 2009). Multiple stellar populations are not expected to significantly affect these results, however, certain clusters with several separate, strongly visible main sequences (such as M54) may be poorly described by these results.

Above average helium enrichment has been proposed as a potential mechanism for producing separations in multiple-population cluster main sequences (Piotto et al. 2007; Milone et al. 2008; Pasquini et al. 2011). Due to the difficulty of measuring He enrichment, the complete effects of enrichment are still under investigation (see preliminary work by Valcarce et al., 2011, and Milone & Merlo, 2008), but it is thought that it will result in slightly brighter subgiant branches and slightly bluer turnoffs, since He is less opaque than H. If the He abundances have only a small effects on photometry, they will not significantly influence these results.

6.8.2 Potential Influence of Binaries

Binary stars present in globular clusters have the potential to bias color-magnitude diagrams towards brighter, redder values (Romani & Weinberg 1991). The maximum increase in brightness occurs when the binaries are of the same spec-

troscopic type, $2.5*log(2) \approx 0.75$ magnitude, and the maximum increase in "reddening," which serves to widen the color-magnitude diagram towards redder values, can shift colors by as much as 0.05 magnitudes. Depending on the binary fraction (f), this effect could introduce significant biases when determining population statistics of a globular cluster.

The binary fraction in old globular clusters is generally low, around a few percent to 20%, with the binaries being concentrated near the cluster's center (Sollima et al. 2007; Fregeau et al. 2009), although a select few globular clusters have f values as high as 50%. The primary result of Monte-Carlo simulations in Fregeau et al. (2009) is that "True" binary fractions are likely to remain constant with age, implying that one can also rule out age-dependent biases on cluster f values. Carney et al. (2003) studied the binary fractions of metal-poor ([Fe/H] $\leq$ -1.4) field red giants and dwarfs, and found them to be 16%±4% and 17%±2%, respectively, which is on the high end of a typical globular cluster f. Globular clusters also feature apparent binaries, which are stars that appear as binaries due to crowding, which become more likely as one looks closer to the densely-packed cluster cores.

Due to the poor performance of the SDSS photometric pipeline in crowded fields, the centers of globular clusters are absent from the database, and therefore are not present in this study. If binaries are concentrated towards a cluster's center as in Sollima et al. (2007); Fregeau et al. (2009), then this "disadvantage" serves to reduce the potential bias that binaries would have on the results, and will also greatly reduce the number of apparent binaries in our data.

The effect of a binary population is to shift the observed turnoff to brighter magnitudes, and to increase the spread in magnitudes. An experiment was run in which 1000 stars were generated with magnitudes given by a double-sided Gaussian with $\mu = 4.2$, $\sigma_l = 0.36$ and $\sigma_r = 0.76$, and with colors given by a standard Gaussian with a $\mu_{(g-r)} = 0.25$ and $\sigma_{(g-r)} = 0.025$. Then 20% of these stars were randomly selected and shifted brighter by 0.75 magnitudes and redder by 0.05 magnitudes, which should produce a larger effect than would be expected from the binary fraction of a typical halo globular cluster. The 20% binary population was found to have a new magnitude fit of $\mu = 4.22$, $\sigma_l = 0.55$ and $\sigma_r = 0.75$, and a color fit of

$\mu_{(g-r)} = 0.26$ and $\sigma_{(g-r)} = 0.03$. The variations in these values are all within the errors of the fitting process, except for σ_l in the magnitude fit. It is possible that the quantity σ_l is sensitive to to the binary fractions typically found in halo globular clusters, in the sense that a large binary fraction could produce a larger σ_l. The spread in σ_l values in the accepted sample is low (Figure 6.12), so it is not expected that the effect of binaries would change the results significantly. A future study could potentially use a more rigorous study of binary fraction effects to see if the binary fraction is directly correlated to the value of σ_l.

The results of this study are expected to provide a close approximation of the F turnoff absolute magnitude distribution in old, metal poor stars in the Milky Way halo, as observed by SDSS. These results should also be a reasonable approximation of small dwarf galaxy turnoff distributions. These results are not expected to be applicable to very young, relatively metal-rich stellar populations, to populations that are outliers on the AMR, or to populations that are outside of the Milky Way.

CHAPTER 7
Summary and Conclusions

7.1 Summary

This thesis describes the effort to characterize the major components of the Milky Way stellar halo, using data from the Sloan Digital Sky Survey and the method of statistical photometric parallax. With an approximation of the magnitude distribution of F turnoff stars, large numbers of these stars may be used to determine the statistical density of a given region, and thereby pick out major substructures. Previous authors developed a model that would describe the Galactic halo in a piecewise fashion, using the stripe format of SDSS data. Software was then developed to determine, given a set of model parameters, the likelihood that the model was a good match to a given data stripe. A maximum-likelihood algorithm was then implemented to intelligently search for the best-fitting model parameters for each SDSS stripe. Subsequently, the maximum-likelihood algorithm would be modified to run with the highly asynchronous, global searches on the MilkyWay@home volunteer computing platform, making use of idle CPU and GPU cycles on the computers of citizen scientists around the world. Using this system, the main ("bright") Sagittarius tidal stream has been characterized in the North Galactic cap.

A catalog of stars was produced that reflects the density profile of the main Sgr stream, and an early fit to the smooth component of the Galactic halo was found. Additionally, the intrinsic absolute magnitude distribution of old halo F turnoff stars was found, and a degeneracy discovered in isochrone fits to old halo clusters that lie along the Milky Way age-metallicity relationship. Preliminary results for the characterization of the secondary North streams are also presented.

The characterization of the Virgo and Sgr bifurcation in the North Galactic cap (NGC) is currently underway, using SDSS data with the main Sgr stream removed. Constraints were placed on the allowed parameter ranges in order to focus the search algorithms on the areas where the streams are expected to be found, but

these constraints need to be re-evaluated. The current fits to the secondary NGC streams are not yet satisfactory, possibly due to a weak bifurcated stream density signal, an inadequate model for the smooth spheroid, or the as yet unaccounted-for increasing color errors with distance.

By studying globular clusters in the Milky Way halo, we found that the turnoff populations of old, metal-poor halo stars have similar intrinsic brightnesses, and use the globular clusters as proxies for all halo stars. This assumption was found to be valid for any stellar halo population that is consistent with the age-metallicity relationship of the Milky Way halo, and so is a good approximation for halo turnoff stars. It was also found that as the SDSS photometric errors increased with distance, turnoff stars would bleed out of the color selection bin and redder stars would leak in and contaminate the data, resulting in an unaccounted-for incompleteness in the number of turnoff stars detected with distance. This effect has been characterized, and the modified fitting algorithm has recently been released on MilkyWay@home. The preliminary results from these searches seem to indicate that the modification greatly improves the fits, but a proper comparison is still underway.

The analysis of the Southern Galactic cap streams has also begun, with data from SDSS data release 8. The results from the Southern stripe analyses will help to answer several questions that the NGC stream characterizations have created, especially in determining if the stellar densities along the entire stream match those of recent N-body simulations.

Ultimately, we find that the Sagittarius tidal stream debris system, while still consistent with recent N-body simulations, may deviate enough from previous models that a new interpretation of Sgr tidal debris may be necessary. The results seem to indicate that the bifurcated ("faint") streams may not have originated with the Sgr dwarf itself, and may be evidence that more than one dwarf galaxy is present in this system. Further study is needed, but it is clear that the Sgr dwarf debris system is still not well understood.

7.2 Conclusions

The major conclusions of this thesis work are as follows:

1. Using the technique of statistical photometric parallax (Newberg 2013a) and the maximum-likelihood evaluator developed in Cole et al. (2008) and Desell (2009), the main Sgr dwarf tidal debris was successfully characterized in 15 NGC stripes of SDSS DR7 data, and 3 SGC stripes of DR6 data.

2. A catalog of turnoff stars consistent with the density profile of the characterized Sgr tidal stream was developed, and is available from the online edition of the associated journal article (Newby et al. 2013).

3. Plane fits to the components of the Sgr tidal streams, in light of previous analysis, indicate that a more complicated, multi-dwarf hypothesis is more favorable than the assumption that only a single dwarf is being disrupted. Additionally, the stream densities in the SGC may indicate that current N-body models are in error, but more analysis is needed.

4. The Hernquist profile fits to smooth component of the stellar halo are strongly oblate, with a flattening parameter of $q = 0.53$ and a scale length of $r_0 = 6.73$ kpc.

5. It was found that a Hernquist profile is probably not an adequate descriptor of the smooth component of the stellar halo, and that other models will have to be tested in the future.

6. It has been shown that the MilkyWay@home volunteer computing platform provides robust, accurate results, and is capable of producing useful scientific results. This platform is a complicated but powerful system for solving computationally intensive, asynchronous problems, with a diverse and computationally savvy user-base.

7. In order to characterize the fainter streams, the density profile of the main Sgr tidal stream was removed from SDSS stripes 9-23, and the remaining stars used as inputs for stream searches on MilkyWay@home. Preliminary fits show that MilkyWay@home will be able to fit the secondary streams, but at least one more iteration of searches is required.

8. Eleven globular clusters in SDSS data were analyzed, and it was found that the absolute magnitude distributions of turnoff stars in these clusters is intrinsically similar. The mean absolute magnitude and bright-side standard deviation are therefore approximately constant across the studied range of clusters, with $\mu = M_g = 4.18 \pm 0.008$ and $\sigma_l = 0.36 \pm 0.006$. It was shown that these values will be a good approximation to all halo turnoff stars that formed in a population that lies along the Milky Way's age-metallicity relationship.

9. The intrinsic faint-side standard deviation for turnoff star magnitudes was found to be $\sigma_r = 0.76 \pm 0.04$, but is greatly effected by photometric errors as distances increase. The increasing color errors with distance cause turnoff stars to bleed out of their color selection range, and redder stars to move in and contaminate the selection. Techniques to correct for these effects have been presented, such that the faint-side distribution (σ_r), changes with distance, and a function to correct for the changing completeness with distance for color-selected turnoff stars is also provided. Correction functions characterizing both of these effects, in the context of statistical photometric parallax, are provided.

10. Halo star populations have similar turnoff properties due to the age-metallicity relationship. Halo populations are generally old and metal-poor, with younger populations having higher metallicities. As a stellar population ages, its turnoff becomes redder, however, a lower metal content will result in a bluer turnoff. It is found that these two effects almost cancel out entirely for old halo populations that lie on the Milky Way's age-metallicity relationship, and so they will have similar intrinsic magnitude and color properties.

With the work described in this thesis, we are now in a position to develop a complete description of stellar densities in the Galactic halo. This study is being completed through a combination of the stellar halo density likelihood model, the characterized magnitude distribution for old halo turnoff stars, and the powerful MilkyWay@home computing platform. Currently underway are density studies of the main Sgr stream in the SGC and the secondary substructures in the NGC; a

study of secondary SGC substructures is planned for the near future. Once these studies are complete the significant halo substructures found in SDSS data will have been accounted for, and then the smooth stellar halo can be accurately characterized. Since the Hernquist may not be the best descriptor of the smooth spheroid, it will be necessary to fit other density models and compare the results. Once the best functional form has been determined, it will be fit to the entirety of SDSS data (with the substructures removed) at once, producing the most accurate description of the stellar halo to date. This thesis, then, represents a major step towards completing a full, accurate, self-consistent density model of the stellar halo and major substructures, and towards a better understanding of the Galactic halo as a whole.

BIBLIOGRAPHY

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2007, ApJS, 172, 634

Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297

Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29

Allgood, B., Flores, R. A., Primack, J. R., et al. 2006, MNRAS, 367, 1781

An, D., Terndrup, D. M., & Pinsonneault, M. H. 2007, ApJ, 671, 1640

An, D., Pinsonneault, M. H., Masseron, T., et al. 2009, ApJ, 700, 523

Anderson, D. P., Korpela, E., & Walton, R. 2005, The First International Conference on e-Science and Grid Techniques (Melbourne, Australia: e-Science 2005), 196

Bahcall, J. N., & Soneira, R. M. 1980, ApJS, 44, 73

Bahcall, J. N., & Soneira, R. M. 1984, ApJS, 55, 67

Bailin, J., & Steinmetz, M. 2005, ApJ, 627, 647

Bedin, L. R., Piotto, G., Anderson, J., et al. 2004, ApJ, 605, L125

Belokurov, V., Zucker, D. B., Evans, N.W., et al. 2006b, ApJ, 642, L137

Belokurov, V., Evans, N. W., Bell, E. F., et al. 2007, ApJ, 657, L89

Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, arXiv:1212.5225v3.

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127

Bullock, J. S., & Johnston, K. V. 2005, ApJ, 635, 931

Carlberg, R. G. 2009, ApJ, 705, L223

Carlberg, R. G. 2012, ApJ, 748, 20

Carlberg, R. G., & Grillmair, C. J. 2013, ApJ, 768, 171

Carney, B. W., Latham, D. W., Stefanik, R. P., Laird, J. B., & Morse, J. A. 2003, AJ, 125, 293

Carretta, E., & Gratton, R. G. 1997, A&AS, 121, 95

Carretta, E., Bragaglia, A., Gratton, R., D'Orazi, V., & Lucatello, S. 2009, A&A, 508, 695

Cole, N., et al. 2008, ApJ, 683, 750

Cole, N. 2009, Ph.D. Thesis,

Correnti, M., Bellazzini, M., Ibata, R. A., Ferraro, F. R., & Varghese, A. 2010, ApJ, 721, 329

De Angeli, F., Piotto, G., Cassisi, S., et al. 2005, AJ, 130, 116

de Vaucouleurs, G. 1977, AJ, 82, 456

de Vaucouleurs, G., & Buta, R. 1978, AJ, 83, 1383

Desell, T. 2009, PhD Thesis, RPI

Desell, T., Anderson, D., Magdom-Ismail, M., et al. 2010a, in The 2010 IEEE Congress on Evolutionary Computation, ed. C.-K. Ting et al. (IEEE CEC; Los Alimitos, CA: IEEE)

Desell, T., Anderson, D., Magdon-Ismail, M., et al. 2010b, in Distributed Applications and Interoperable Systems, ed. F. Eliassen & R. Kapitza (Lecture Notes in Computer Science 6115; Berlin: Springer), 29

Desell, T., Waters, A., Magdon-Ismail, M., et al. 2009, in Parallel Processing and Applied Mathematics Pt I, ed. R. Wyrzykowski et al. (Lecture Notes in Computer Science 6067; Berlin: Springer), 276

Desell, T., Willet, B. A., Arsenault, M., et al. 2011, in Proc. IPDPS Workshops, ed. A. Sussman et al. (Los Alimitos, CA: IEEE), 1888

Desell, T., Holman M., Noghanian, S., et al. 2013, in The 2013 IEEE Congress on Evolutionary Computation (Cancun, Mexico: IEEE), *submitted*

Dotter, A., Sarajedini, A., Anderson, J., et al. 2010, ApJ, 708, 698

Dotter, A., Sarajedini, A., & Anderson, J. 2011, ApJ, 738, 74

Eggen, O. J., Lynden-Bell, D., & Sandage, A. R. 1962, ApJ, 136, 748

Eyre, A., & Binney, J. 2011, MNRAS, 413, 1852

Fellhauer, M., Belokurov, V., Evans, W. et al. 2006, ApJ, 651, 167

Freese, K., Gondolo, P., Newberg, H. J. 2005, Physical Review D, 71, 4

Fregeau, J. M., Ivanova, N., & Rasio, F. A. 2009, ApJ, 707, 1533

Gao, L., Navarro, J. F., Frenk, C. S., et al. 2012, MNRAS, 425, 2169

Gehrels, N. 1986, ApJ, 303, 336

Gilmore, G., & Reid, N. 1983, MNRAS, 202, 1025

Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, VizieR Online Data Catalog, 414, 10371

Grabowski, K., Newby, M., & Newberg, H. J. 2013, JURP, *submitted*

Gratton, R., Sneden, C., & Carretta, E. 2004, ARA&A, 42, 385

Grillmair, C. J. 2006, ApJ, 645, L37

Grillmair, C. J. 2009, ApJ, 693, 1118

Harris, W. E. 1996, AJ, 112, 1487

Harris, W. E., Bell, R. A., Vandenberg, D. A., et al. 1997, AJ, 114, 1030

Haurberg, N. C., Lubell, G. M. G., Cohn, H. N., et al. 2010, ApJ, 722, 158

Helmi, A., & White, S. D. M. 2001, MNRAS, 323, 529

Helmi, A., White, S. D. M., & Springel, V. 2003, MNRAS, 339, 834

Helmi, A., 2004, ApJ, 610, L97

Hernquist, L. 1990, ApJ, 356, 359

Ibata, R. A., Gilmore, G., & Irwin, M. J. 1994, Nature, 370, 194

Ibata, R. A., Wyse, R. F. G., Gilmore, G., Irwin, M. J., & Suntzeff, N. B. 1997, AJ, 113, 634

Ibata, R., Irwin, M., Lewis, G. F., & Stolte, A. 2001, ApJ, 547, L133

Ibata, R., Lewis, G. F., Irwin, M., Totten, E., & Quinn, T. 2001, ApJ, 551, 294

Ibata, R. A., Lewis, G. F., Irwin, M. J., & Cambrésy, L. 2002, MNRAS, 332, 921

Ivezić, Ž., Goldston, J., Finlator, K., et al. 2000, AJ, 120, 963

Johnston, K. V., Majewski, S. R., Siegel, M. H., Reid, I. N., & Kunkel, W. E. 1999, AJ, 118, 1719

Johnston, K. V., Spergel, D. N., Hadyn, C., 2002, ApJ, 570, 656

Johnston, K. V., Law, D. R., Majewski, S. R. 2005, ApJ, 619, 800

Jofré, P., & Weiss, A. 2011, A&A, 533, A59

Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864

Koposov, S. E., Rix, H.-W., & Hogg, D. W. 2010, ApJ, 712, 260

Koposov, S. E., Belokurov, V., Evans, N. W., et al. 2012, ApJ, 750, 80

Küpper, A. H. W., Kroupa, P., Baumgardt, H., & Heggie, D. C. 2010, MNRAS, 401, 105

Law, D. R., Majewski, S. R., Skrutskie, M. F., & Johnston, K. V. 2004, Satellites and Tidal Streams, 327, 23

Law, D. R., Johnston, K. V., & Majewski, S. R. 2005, ApJ, 619, 807

Law, D. R., Majewski, S. R., & Johnston, K. V. 2009, ApJ, 703, L67

Law, D. R., & Majewski, S. R. 2010, ApJ, 714, 229

Li, J., Newberg, H. J., Carlin, J. L., et al. 2012, ApJ, 757, 151

Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, AJ, 118, 1406

Lux, H., Read, J. I., Lake, G., & Johnston, K. V. 2012, MNRAS, 424, L16

Majewski, S. R. 1993, ARA&A, 31, 575

Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, ApJ, 599, 1082

Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883

Marín-Franch, A., Aparicio, A., Piotto, G., et al. 2009, ApJ, 694, 1498

Martínez-Delgado, D., Peñarrubia, J., Jurić, M., Alfaro, E. J., & Ivezić, Z. 2007, ApJ, 660, 1264

McWilliam, A. 1997, ARA&A, 35, 503

Milone, A. P., Bedin, L. R., Piotto, G., et al. 2008, ApJ, 673, 241

Milone, L. A., & Merlo, D. C. 2008, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 51, 121

Muratov, A. L., & Gnedin, O. Y. 2010, ApJ, 718, 1266

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, ApJ, 569, 245

Newberg, H. J., & Yanny, B. 2006, Journal of Physics Conference Series, 47, 195

Newberg, H. J., Yanny, B., Cole, N., et al. 2007, ApJ, 668, 221

Newberg, H. J., Willett, B. A., Yanny, B., & Xu, Y. 2010, ApJ, 711, 32

Newberg, H. J. 2013, IAU Symposium, 289, 74

Newberg, H. J. 2013, IAU Symposium, *submitted*

Newby, M., Newberg, H. J., Simones, J., Cole, N., & Monaco, M. 2011, ApJ, 743, 187

Newby, M., Cole, N., Newberg, H. J., et al. 2013, AJ, 145, 163

Oort, J. H., & Plaut, L. 1975, A&A, 41, 71

Pasquini, L., Mauas, P., Käufl, H. U., & Cacciari, C. 2011, A&A, 531, A35

Peñarrubia, J., Belokurov, V., Evans, N. W., et al. 2010, MNRAS, 408, L26

Peñarrubia, J., Zucker, D. B., Irwin, M. J., et al. 2011, ApJ, 727, L2

Pinsonneault, M. H., Terndrup, D. M., Hanson, R. B., & Stauffer, J. R. 2003, ApJ, 598, 588

Pinsonneault, M. H., Terndrup, D. M., Hanson, R. B., & Stauffer, J. R. 2004, ApJ, 600, 946

Piotto, G., Bedin, L. R., Anderson, J., et al. 2007, ApJ, 661, L53

Piotto, G. 2009, arXiv:0902.1422

Planck Collaboration, 276 authors, 2013, arXiv:1303.5062.

Preston, G. W., Shectman, S. A., & Beers, T. C. 1991, ApJ, 375, 121

Pritzl, B. J., Venn, K. A., & Irwin, M. 2005, AJ, 130, 2140

Reid, I. N., Yan, L., Majewski, S., Thompson, I., & Smail, I. 1996, AJ, 112, 1472

Robin, A. C., Reylé, C., & Crézé, M. 2000, A&A, 359, 103

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

Roederer, I. U. 2009, AJ, 137, 272

Romani, R. W., & Weinberg, M. D. 1991, ApJ, 372, 487

Rubin, V. C., & Ford, W. K., Jr. 1970, ApJ, 159, 379

Schlegel, D.J., Finkbeiner, D.P., & Davis, M. 1998, ApJ, 500, 525

Searle, L., & Zinn, R. 1978, ApJ, 225, 357

Sesar, B., Ivezić, Ž., Grammer, S. H., et al. 2010, ApJ, 708, 717

Sharma, S., Bland-Hawthorn, J., Johnston, K. V., & Binney, J. 2011, ApJ, 730, 3

Siegel, M. H., Majewski, S. R., Reid, I. N., & Thompson, I. B. 2002, ApJ, 578, 151

Siegal-Gaskins, J. M., & Valluri, M. 2008, ApJ, 681, 40

Sollima, A., Beccari, G., Ferraro, F. R., Fusi Pecci, F., & Sarajedini, A. 2007, MNRAS, 380, 781

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685

Tissera, P. B., White, S. D. M., Pedrosa, S., & Scannapieco, C. 2010, MNRAS, 406, 922

Valcarce, A. A. R., Catelan, M., & Sweigart, A. V. 2011, Revista Mexicana de Astronomia y Astrofisica Conference Series, 40, 257

Varghese, A., Ibata, R., & Lewis, G. F. 2011, MNRAS, 417, 198

Venn, K. A., Irwin, M., Shetrone, M. D., et al. 2004, AJ, 128, 1177

Vivas, A. K., Zinn, R., Andrews, P., et al. 2001, ApJ, 554, L33

Vivas, A. K. 2002, Ph.D. Thesis

Vivas, A. K., Zinn, R., & Gallart, C. 2005, AJ, 129, 189

Vivas, A. K., Zinn, R., Duffau, S., & Jaffé, Y. 2012, European Physical Journal Web of Conferences, 19, 2007

Willett, B. A., Newberg, H. J., Zhang, H., Yanny, B., & Beers, T. C. 2009, ApJ, 697, 207

Willett, B. A. 2010, Ph.D. Thesis,

Yanny, B., Newberg, H. J., Kent, S., et al. 2000, ApJ, 540, 825

Yanny, B., Newberg, H. J., Grebel, E. K., et al. 2003, ApJ, 588, 824

Yanny, B., Newberg, H. J., Johnson, J. A., et al. 2009, ApJ, 700, 1282

Yoon, J. H., Johnston, K. V., & Hogg, D. W. 2011, ApJ, 731, 58

York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579

Zhao, H. 1998, ApJ, 500, L149

Zinn, R., & West, M. J. 1984, ApJS, 55, 45