THE STELLAR DENSITY OF THE MAJOR SUBSTRUCTURE IN THE MILKY WAY HALO

Jake Weiss

Submitted in Partial Fullfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Approved by:
Dr. Heidi Jo Newberg, Chair
Dr. Ethan Brown
Dr. Gyorgy Korniss
Dr. Carlos Varela

Department of Physics
Rensselaer Polytechnic Institute
Troy, New York

[August 2018]
Submitted July 2018

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENT

This thesis would not have been possible without the hard work or contributions of thousands of individuals. Whether it is my family, friends, mentors, advisors, peers, colleagues, or thousands of volunteers their help was invaluable.

First, I would like to thank my mentor and advisor, Heidi Jo Newberg. Without her guidance, support, and constant drive to push me, I would never have even considered pursuing a PhD. Over the years she has taught me what it means to be a good scientist and a good person, mainly by leading through example. After countless afternoons playing soccer with her son and six years of working together, I consider her family.

Next, I would like to thank my committee members: Ethan Brown, Gyorgy Korniss, and Carlos Varela. Ethan has only been at RPI a few years, but his impact in the department can certainly be felt. Gyorgy taught the first physics class I took at RPI, and since then I have has the pleasure of taking two additional courses with him. During this time, I have gained great respect for him and think all professors should strive to teach as well as he does. Carlos is a MilkyWay@home founder and long time supporter of the project. He helped lay the foundation for the project and has helped advise several students who made the project possible. Both Gyorgy and Carlos have also been great teammates on the soccer field. They have both taught me a lot about being a goalkeeper over the years.

My colleagues and friends in the Star Lab: Jefferey Thompson, Charles Martin, Siddhartha Shelton, Paul Amy, Clayton Rayment, Keighley Rockcliffe, and countless others have been great sounding boards over the past 6 years. Whether it is helping with some rubber ducky debugging or going to trivia, these people have made working on my project much more enjoyable. I would especially like to thank Jeff for being like an older brother to me.

Thank you to my family: Melissa Weiss, Angelo Nittoli, David Weiss, and Danielle Weiss for supporting me. I feel our time apart these last several years, pursuing our own goals, has helped us grow closer together as a family. Whether its frantic late night phone calls to rant about all of the work I had to do, trips to hockey games and rugby games, or trips to New York City, you have all helped me maintain sanity through my time at RPI.

Finally, I would like to thank the more than 200,000 MilkyWay@home volunteers for

their countless hours of donated computing time, their constant community involvement with debugging and development, and their financial support over the past decade.

I would also like to thank and acknowledge contributions made by The Marvin Clan, Babette Josephs, Manit Limlamai, and the 2015 Crowd Funding Campaign to Support Milky Way Research. This publication is based on work supported by the National Science Foundation under grant No. AST 16-15688 and the NASA/NY Space Grant fellowship.

I use data from the Sloan Digital Sky Survey. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

ABSTRACT

We develop, test, and apply a statistical photometric parallax method using main sequence turn off (MSTO) stars in the Sloan Digital Sky Survey (SDSS). We show using simulated data that if our density model is similar to the actual density distribution of our data, we can reliably determine the density model parameters of at least three major substructures in the Milky Way halo, and a smooth background component, using the computational resources available on MilkyWay@home (a twenty parameter fit). As a test for the new model, we fit the stellar density in SDSS stripe 19. After confirming that the model is working as intended on both simulated and observed SDSS stripe 19 data, we moved on to fitting stripes 10 through 23 in the SDSS north Galactic cap. We found an oblate halo with an average flattening of 0.58. Seven streams were found in these fourteen stripes. The Sgr dwarf leading tidal tail detected in 8 stripes, with properties that are consistent with previous fits to the streams. The trailing tidal tail and the "bifurcated" stream were found at the previously identified distances. The Parallel Stream was traced across the sky at a distance of 15 kpc, and roughly tracks an orbit that was fit by previous authors to the Virgo Stellar Stream. A new stream, the Perpendicular Stream, was found at a distance of 15 kpc in the region of Virgo, but roughly perpendicular in orientation on the sky to the Parallel Stream. A stream possibly including the globular cluster NGC 5466 was also found at a distance between 5 and 15 kpc from the Sun.

CHAPTER 1
Introduction

1.1 Milky Way halo substructure

The distribution of stars in the Galactic halo is dominated by dwarf galaxies and tidal streams of stars that have been stripped from them [see Fig. 1 of Newberg et al. (2002), and the "Field of Streams" from (Belokurov et al., 2006)]. These substructures represent the recent minor merger history of the Milky Way (Bullock & Johnston, 2005), and contribute to the build-up of stars in the Milky Way stellar halo. Two or three dozen tidal debris streams, most of which extend tens of degrees or more across the sky, have been identified; the exact number cannot be determined due to controversy over the identity of tidal streams, particularly those discovered near the Galactic plane, and because some streams are detected at low enough significance that they are considered stream "candidates." In addition to streams, other substructures of ambiguous origin (most notably "clouds") continue to be discovered in the Galactic spheroid.

For a review of tidal streams and clouds, see Grillmair & Carlin (2016). An updated list of streams and clouds included in the 'GALSTREAMS' Python Package can be found in Table 4 of Mateu et al. (2018). More recent halo substructures are identified in Li et al. (2016b), Sohn et al. (2016), Grillmair (2017a,b), Jethwa et al. (2017), and Shipp et al. (2018). In particular, Shipp et al. (2018) identify eleven new substructures in the Milky Way halo using data from the Dark Energy Survey (DES; The Dark Energy Survey Collaboration, 2005, 2016).

Identification and measurement of tidal debris in the Milky Way halo is useful for understanding structure formation and galaxy assembly, and it has the potential to constrain the density distribution of the Milky Way's stellar halo. Methods for measuring the halo shape from tidal streams have been, and continue to be, developed (e.g. Law & Majewski, 2010; Koposov et al., 2013; Küpper et al., 2015; Bovy et al., 2016; Dierickx & Loeb, 2017a; Sanderson et al., 2017; Johnston & Carlberg, 2016, for a review). In addition, the distribution of dark subhalos can be measured by looking for stars ejected from tidal streams (Siegal-

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., Newby, M., & Desell, T. submittedb, ApJS

Gaskins & Valluri, 2008), stream heating (Johnston, 2002) or stream gaps (Carlberg, 2012). Pearson et al. (2017) show that streams can also be used to constrain the rotation rate of the Galactic bar.

These techniques to determine the dark matter distribution in the Milky Way from tidal streams rely on accurate measurements of the tidal debris itself, but as we discover that halo tidal streams are more numerous and complex than originally thought, the association of particular stars with particular tidal streams becomes more ambiguous. For example, Newberg et al. (2009) discovered that the blue horizontal branch stars (BHBs) thought to be associated with the southern portion of the Sagittarius (Sgr) dwarf tidal stream in Yanny et al. (2000) are actually part of the Cetus Polar Stream. The Sgr dwarf tidal stream, which is the most prominent tidal stream in the sky, and the so-called "bifurcated" Sgr stream that appears to split off from it, have also caused confusion; for example Newby et al. (2013) suggested that the southern Sgr stream could be associated with the "bifurcated" stream in the north, and the northern Sgr stream could be associated with the "bifurcated" stream in the south. These misidentifications and possible misidentifications of stars in the most prominent halo streams underscore the difficulties in counting and characterizing tidal streams.

1.2 Statistical photometric parallax

In this thesis we present an improved *statistical photometric parallax* (Cole et al., 2008; Newberg, 2012) method to measure the spatial density of stars in the Milky Way stellar halo, using turnoff stars from the Sloan Digital Sky Survey (SDSS; York et al., 2000; Gunn et al., 2006). Statistical photometric parallax is the use of statistical knowledge of the distribution of the absolute magnitudes of stellar populations to determine the underlying density distributions of those stars. This differs from photometric parallax in that the distance to each individual star is not determined.

The idea of using turnoff stars to trace Milky Way halo substructure was introduced by Newberg et al. (2002). They observed density substructure in the SDSS turnoff stars on the Celestial equator, and fit an absolute magnitude distribution to their blue turnoff star tracers. In Cole et al. (2008), these tracers and a simplified absolute magnitude distribution from Newberg et al. (2002) were used to build a model of the Milky Way halo and its substructure and complete preliminary fits to the stellar density of the halo and the Sagittarius dwarf

galaxy tidal stream. Several years later, Newby et al. (2013) continued working with this model and showed that the massive distributed computing network, MilkyWay@home, could be effective in constraining the parameters in the density models of tidal streams. Taking advantage of this new computational power, several optimizations were run on each stripe. Initially, the fitting algorithms were allowed great freedom in selecting the parameters. Later, the parameters were constrained based on the results from neighboring stripes.

These previous studies successfully used statistical photometric parallax to study the structure of the halo using SDSS turnoff stars. SDSS turnoff stars, detected to a limiting magnitude of $g = 22.5$, can be used to trace the structure of the Milky Way to 45 kpc from the Sun. However, the turnoff stars in a single stellar population, with the same color, can differ in absolute magnitude by two magnitudes (producing a distance error of a factor of 2.5). Photometric parallax (e.g. Jurić et al., 2008) is unusable with turnoff stars because astronomers do not have a way to determine the distance to individual stars with reasonable accuracy using photometry alone.

It has been shown that the absolute magnitude distribution of turnoff stars in halo globular clusters are surprisingly similar to each other, over a metallicity range -2.3<[Fe/H]<-1.2 dex and over ages ranging from 9 to 13.5 Gyr (Newby et al., 2011). Grabowski et al. (2013) showed that this similarity holds even for the globular cluster Whiting 1, which is only 6 Gyrs old and has a metallicity of approximately [Fe/H]$\sim$0.6 dex (Carraro et al., 2007; Valcheva et al., 2015). This surprising result, which comes about due to the age-metallicity relation for Milky Way stars, makes turnoff stars very useful for tracing the density of the stellar spheroid and outer disk.

In our work, we improve on the statistical photometric parallax method by implementing a better model for the absolute magnitude of the tracer stars and their detection efficiency, and by using better fitting methods on MilkyWay@home. In our implementation of statistical photometric parallax, we find the parameters in a density model that make the apparent magnitudes and angular positions of the observed stars most likely, using a maximum likelihood estimator (MLE; Ivezić et al., 2014, pg. 124). The statistical description of the absolute magnitudes of the stellar tracers, and of the selection effects in the data, make statistical photometric parallax somewhat complex to apply. Because we are able to take all of these effects into account, we can reliably measure density distributions in real data.

There are four parts to statistical photometric parallax: data, a density model, an

algorithm for measuring how well the model fits the data, and an algorithm for optimizing parameters. The algorithm that measures how well the model fits the data includes the MSTO absolute magnitude distribution, as well as any observational biases. It has taken us many years to perfect the algorithm that can simultaneously fit the spatial density of several tidal streams plus a smooth distribution to the SDSS MSTO stars; the smooth distribution represents the sum of: streams from small satellites, old streams that they are well mixed in density, and stars that were created during the collapse of the Milky Way, if any.

In this paper, we describe an improved algorithm for characterizing the spacial characteristics of stellar streams in the Milky Way halo using turnoff stars, and show that it is capable of simultaneously recovering the characteristics from three tidal streams plus a smooth halo component, using simulated data designed to mimic the stellar density in the actual Milky Way halo.

1.3 The big three halo substructures: the Sgr tidal stream, the "bifurcated" stream, and the Virgo Overdensity

We will present preliminary results for 15 2.5°-wide SDSS stripes (9 through 23) of data that cuts across the northern Galactic hemisphere. Figure 1.1 shows the position of Stripe 19 in the SDSS northern footprint. The results from these stripes provide measurements of the largest known substructures in the Milky Way halo: the Sgr dwarf tidal stream, the so-called "bifurcated" stream, and the Virgo Overdensity.

During the validation of our model, we will focus on Stripe 19 which crosses the Sgr dwarf tidal stream and the bifurcated stream, in a region of the sky in which they are clearly separated. Since Stripe 19 is more than 20 degrees from the densest portion of the Virgo Overdensity, it is uncertain whether a third substructure measured here is in the tails of the Virgo Overdensity, or whether it is associated with a new halo substructure. In addition to three substructures, we fit smooth Milky Way halo and thick disk distributions.

The Sgr dwarf galaxy was first discovered by Ibata et al. (1995), who found evidence of a dwarf galaxy within 16 kpc of the Galactic center, on the far side of the Milky Way, that was thought to be in the process of tidally disrupting. The tidal stream of stars stripped from this dwarf galaxy have since been found to dominate the substructure of the Galactic halo (e.g. Newberg et al., 2002; Majewski et al., 2003; Belokurov et al., 2006; Hernitschek

et al., 2017). Though the Sgr dwarf galaxy and the stream of stars that have been tidally stripped from its gravitational grasp have been studied extensively (see Law & Majewski, 2016, for a recent review); we are only starting to understand the dynamical history of this present-day merger.

It has been a challenge to find a disruption model that simultaneously fits the positions of the leading and trailing tidal streams in the sky, the line-of-sight velocities of the stream stars, and the observed extension of the trailing tidal tail to ∼100 kpc from the Galactic center (Newberg et al., 2003; Belokurov et al., 2014). Dierickx & Loeb (2017b) present a recent simulation of the tidal debris that reproduces most of the measurements of the position of the leading and trailing tidal debris, including the distant stars in the trailing tidal tail and the observed "spurs" at apogalacticon (Sesar et al., 2017), but still doesn't reproduce the line-of-sight velocities of the leading tail. A previous model by Law & Majewski (2010) was able to fit the velocities of the leading tidal tail, using a triaxial dark halo model in which the disks rotate around the intermediate axis. However, this Milky Way configuration is very unlikely (Debattista et al., 2013). Refining the spatial distribution of the Sgr dwarf tidal stream using the algorithm described in this paper will help constrain N-body simulations of the Sgr dwarf tidal disruption, and lead to a better understanding of the shape of the Milky Way's dark matter halo.

The "bifurcated" stream can be seen clearly in the "Field of Streams" as a lower surface brightness companion stream to the Sgr dwarf tidal stream (Belokurov et al., 2006). Belokurov identifies the Sgr dwarf tidal stream as "Stream A," the "bifurcated stream" as Stream B, and tentatively identifies a more distant "Stream C" behind Stream A, which is now generally associated with an extension of the Sgr trailing tidal tail (Li et al., 2016a). Koposov et al. (2012) shows the analogous bifurcated stream in the south Galactic cap. Although the origin of the second, lower surface brightness stream close to the Sgr stream is not known, a leading possibility is that it could arise from multiple wraps of the stream around the Milky Way (Fellhauer et al., 2006). Since its discovery, this stream has remained relatively unstudied compared to its sibling. Newberg et al. (2007) derive distances that are slightly farther away than Sgr for the bifurcated stream. In contrast, Niederste-Ostholt et al. (2010) says the bifurcated stream is slighly closer to the Sun than the Sgr dwarf tidal stream and Ruhland et al. (2011) finds the distances are basically the same. Slater et al. (2013) show that the southern bifurcated stream is closer to the Sun than the southern portion of

the Sgr dwarf tidal stream. Yanny et al. (2009) show that the velocities and metallicities along the bifurcated stream are similar to those in Sgr. In Koposov et al. (2013), there is evidence presented that the two streams may both pass through the progenitor, but due to the proximity of the progenitor to the Galactic bulge, it is difficult to see where exactly the two cross in reference to the progenitor. In Newby et al. (2013) it is suggested that the streams may be from two separate progenitors that accreted around the same time, but the evidence to support this is not strong. Hernitschek et al. (2017) give a possible fit to the bifurcated stream. Currently, the origin of this stream is still an open question that our results will help answer. Determining the origin of the "bifurcated" stream is critically important, as it is useful for constraining the Milky Way potential (Law & Majewski, 2010; Vera-Ciro & Helmi, 2013).

The Virgo Overdensity/Virgo Stellar Stream (Vivas et al., 2001; Jurić et al., 2008; Duffau et al., 2006; Newberg et al., 2007) is a third large halo overdensity in the northern Galactic hemisphere, at distances of $6 - 20$ kpc from the Sun. It is unclear whether this feature is a tidal stream, a "cloud," or a combination of several different pieces. Carlin et al. (2012) fit an orbit to the puffy structure, and suggest that this overdensity is the result of a recently disrupted massive ($10^9 M_\odot$) dwarf galaxy. Carlin et al. (2012) also find their orbit includes the Pisces Overdensity. Li et al. (2016b) suggest Virgo could instead be associated with the Hercules-Aquila Cloud and Eridanus-Phoenix overdensities, since they are on the same polar plane, have similar galactocentric distances (18 kpc), and are separated by 120 degrees. In Bonaca et al. (2012), it is suggested that Virgo is "cloud-like" and may have been the result of a minor merger that passed close to the Galactic center. Vivas et al. (2016) find several different, presumably unrelated, substructures of RR Lyrae stars at distances of 10-20 kpc in the Virgo region, and suggest there could be additional substructures at much larger distances. The evidence for a more distant Virgo substructure is amplified by Sesar et al. (2017), who find an outer Virgo overdensity at a distance of 80 kpc from the Sun.

The Milky Way stellar halo has traditionally been described by a smooth power-law distribution (e.g. Oort & Plaut, 1975; Preston et al., 1991). Since the discovery of significant substructure in the stellar halo (Newberg et al., 2002), researchers have had to choose whether to include or exclude these substructures when fitting the overall spheroid density. The smooth density component of the halo includes smaller or more thoroughly mixed remnants of tidal stripping, as well as any stars that were created in the initial gravitational collapse

of the Milky Way galaxy. The algorithms used in this project will fit the smooth component and streams simultaneously. Using simultaneous fitting for the background (as we will refer to the smooth component) and streams instead of subtracting a background to fit streams, we learn about the shape and density of this smooth component of the stellar halo without requiring a clean sky sample to fit it.

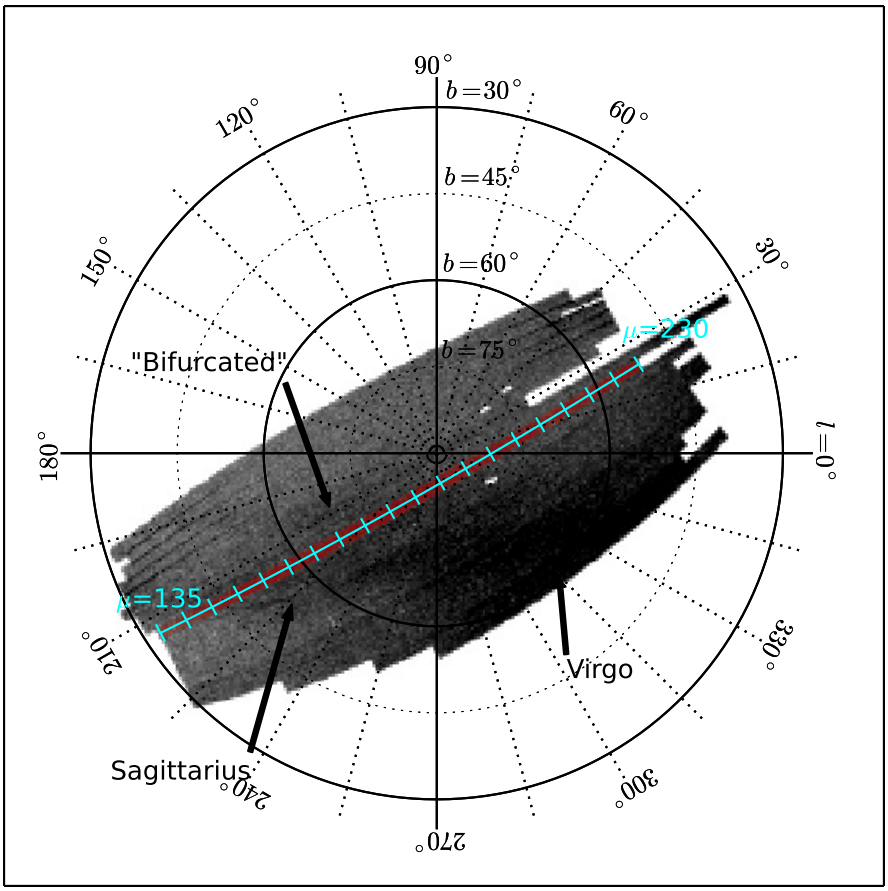


Figure 1.1: The density of SDSS turnoff stars in the north Galactic cap. We highlight stripe 19 in magenta and show μ throughout this stripe in blue. There is some overlap between the stripes, which appears in this figure as dark streaks on the borders of the stripes. The lowest latitude four stripes have a fainter magnitude limit, which explains why they appear darker. In this plot, we can see the major substructures in our data, including the Sagittarius stream, the "bifurcated" stream, and the Virgo Overdensity, with relation to the stripe we are fitting. Note that Stripe 19 passes through the Sgr dwarf tidal stream and the "bifurcated" stream, but is far from the highest density region of the Virgo Overdensity.

CHAPTER 2
Data selection

2.1 SDSS turnoff stars

Our main goal is to fit the major substructure in the Milky Way halo through the SDSS footprint using MSTO stars as tracers. Halo MSTO stars are more abundant than intrinsically brighter giant stars in the halo, and can be observed to distances of 46 kpc in SDSS photometric data. Previous studies have found that MSTO stars bluer than the thick disk turnoff are present in the Sgr dwarf tidal stream, the "bifurcated" stream, and the Virgo Overdensity. Intrinsically fainter main sequence stars are not observed at distances far enough to trace our target halo substructures.

Whereas stars like red giant stars or BHBs are often assumed to have a known absolute magnitude based on their color or spectral properties, MSTO stars of a given color and stellar population are spread over a range of absolute magnitudes. Instead of looking for a way to measure the absolute magnitude for each of these tracer stars, Newberg et al. (2002) used the average apparent magnitude of the stars in a particular substructure, compared to the average absolute magnitude of the MSTO population, to ascertain the substructure's distance. Cole et al. (2008) fit the density distribution of the Sgr dwarf tidal stream under the assumption that the absolute magnitude distribution of turnoff stars was Gaussian. Later, by studying MSTO stars in Milky Way globular clusters, Newby et al. (2011) not only fit a more accurate absolute magnitude distribution, which incorporated observational effects from SDSS and their selection efficiency, but also showed that the absolute magnitude distribution was the same in a range of globular clusters observed in the Milky Way halo.

We select our sample of turnoff stars from SDSS Data Release 7 (DR7 Abazajian et al., 2009), with the criteria: $g_0 > 16$, $0.1 < (g-r)_0 < 0.3$, $(u-g)_0 > 0.4$, and EDGE and SATURATED flags not set (Newberg & Yanny, 2006; Newberg et al., 2002). Selecting stars fainter than $g_0 = 16$ removes any saturated stars not identified with the saturated flag. The $(g-r)_0$ cut is used to pick out the blue side of the turnoff of the halo main sequence, while avoiding the redder thick disk turnoff stars. The $(u-g)_0 > 0.4$ cut is used

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., Newby, M., & Desell, T. submittedb, ApJS

to eliminate quasars. We use the subscript "0" to indicate that the magnitudes we are using are reddening corrected using the Schlegel et al. (1998) dust maps. To visualize these cuts on a color magnitude diagram, see Figure 2.1. In this figure, we plot the SDSS Data Release 14 (DR14 Abolfathi et al., 2018) stars for NGC 5272, which is a globular cluster with a population that is representative of the halo as a whole (Newby et al., 2011), and show that the color selection bin successfully captures the cluster's turnoff. To minimize disk contamination, stars with $b < 30°$ are also cut from the data (Cole et al., 2008).

An SDSS stripe or "wedge" includes a volume defined by the angular limits of a stripe and the distance (from the Sun) to the most distant object in the dataset. Each SDSS stripe is 2.5° wide, and typically 140° long. Since the density varies only a small amount in the narrow direction, we often depict the stellar density in a polar plot with the radius proportional to the distance and the angle given by μ, the angular distance along the stripe. We apply our algorithm to one wedge (stripe) of data at a time.

The footprint of the 14 SDSS stripes we fit in the SDSS north Galactic cap (10 through 23) are shown in Figure 1.1. The limits and the number of stars present for each of these stripes are listed in Table 2.1. In total, we are using 1,417,449 stars across these 14 stripes. These stripes were chosen because they trace Sagittarius and the "bifurcated" stream through the north Galactic cap.

2.2 Removing globular clusters and other stellar substructure

In some stripes, there are globular clusters or other stellar substructure such as the Monoceros ring (Newberg et al., 2002; Yanny et al., 2003) or the Galactic bulge that are not well fit by our parameterized density model. We avoid these structures by removing the area of the sky in which they are contained. We can remove a small area of the sky around globular clusters, and then also remove a section of the sky over which our model is integrated, as described in section 3.7. Low latitude substructure can be removed by removing a larger area of data near the Galactic center and anti-center where necessary, thus making the stripe shorter.

In Table 2.2, we show all of the globular clusters near the SDSS stripes we are fitting with their sky positions and distances as listed in the Harris catalog (Harris, 1996). Many of these globular clusters were removed by changing the μ boundary of the stripes to avoid the clusters. Examples of these clusters include Pal 5 and NGC 5904 which are near stripes

Table 2.1: Wedge constraints for each wedge in SDSS survey coordinates.

SDSS Stripe Constraints						
Stripe	μ_{min} (deg)	μ_{max} (deg)	g_{0min}	g_{0min}	Stars	Cuts
10	165	227	16.0	23.5	97,939	0
11	150	229	16.0	23.0	97,434	0
12	135	235	16.0	23.0	120,612	0
13	135	235	16.0	23.0	118,836	0
14	135	235	16.0	22.5	102,599	0
15	135	240	16.0	22.5	108,460	0
16	135	240	16.0	22.5	107,033	0
17	135	235	16.0	22.5	91,626	2
18	135	240	16.0	22.5	95,462	1
19	135	230	16.0	22.5	84,046	0
20	133	249	16.0	22.5	105,909	0
21	133	210	16.0	22.5	60,503	0
22	131	225	16.0	22.5	66,200	2
23	133	230	16.0	22.5	65,355	1

Table 2.2: Globular clusters near SDSS stripes 10 to 23. These values were taken from the Harris catalog (Harris, 1996).

Globular Clusters in the SDSS Footprint					
Name	l (deg)	b (deg)	R (kpc)	ra (deg)	dec (deg)
Pal 3	240.1	41.8	89.0	151.3	0.06
Pal 4	202.3	71.8	99.0	172.3	29.0
NGC 4147	252.8	77.1	18.0	182.5	18.5
NGC 5024	332.9	79.7	18.0	198.3	18.1
NGC 5053	335.6	78.9	16.0	199.1	17.7
NGC 5272	42.2	78.7	10.0	205.6	28.4
NGC 5466	42.1	73.5	16.0	211.5	28.5
Pal 5	0.8	45.8	22.0	229.0	−0.2
NGC 5904	3.8	46.8	7.0	229.6	2.1
Pal 14	28.7	42.1	72.0	242.8	14.9

10 and 11. For the globular clusters within the footprint, we use the cuts found in Table 2.3. For consistency, these cuts are the same cuts that were used in Newby et al. (2013). Newby et al. (2013) determined these cuts by querying the SDSS stars in the region of the globular cluster and finding the limits of the globular cluster by eye. In order to ensure most globular cluster stars were removed, Newby et al. (2013) slightly overestimated the bounding box for the globular clusters.

Table 2.3: Cuts in the SDSS north Galactic cap stripes. These cuts were made based on the globular clusters found in Table 2.2.

Cuts							
Stripe	Cluster	μ_{min} (deg)	μ_{max} (deg)	ν_{min} (deg)	ν_{max} (deg)	g_{0min}	g_{0min}
17	NGC 4147	182.4	183	0.9	1.25	16.0	22.5
17	NGC 5024/NGC 5053	197	199	0.4	1.25	16.0	22.5
18	NGC 5024	197	198	-1.25	-1.0	16.0	22.5
22	NGC 5466	207	209	0.8	1.25	16.0	22.5
22	NGC 5272	202	204	-0.5	0.8	16.0	22.5
23	NGC 5466	207	209	-1.25	-0.8	16.0	22.5

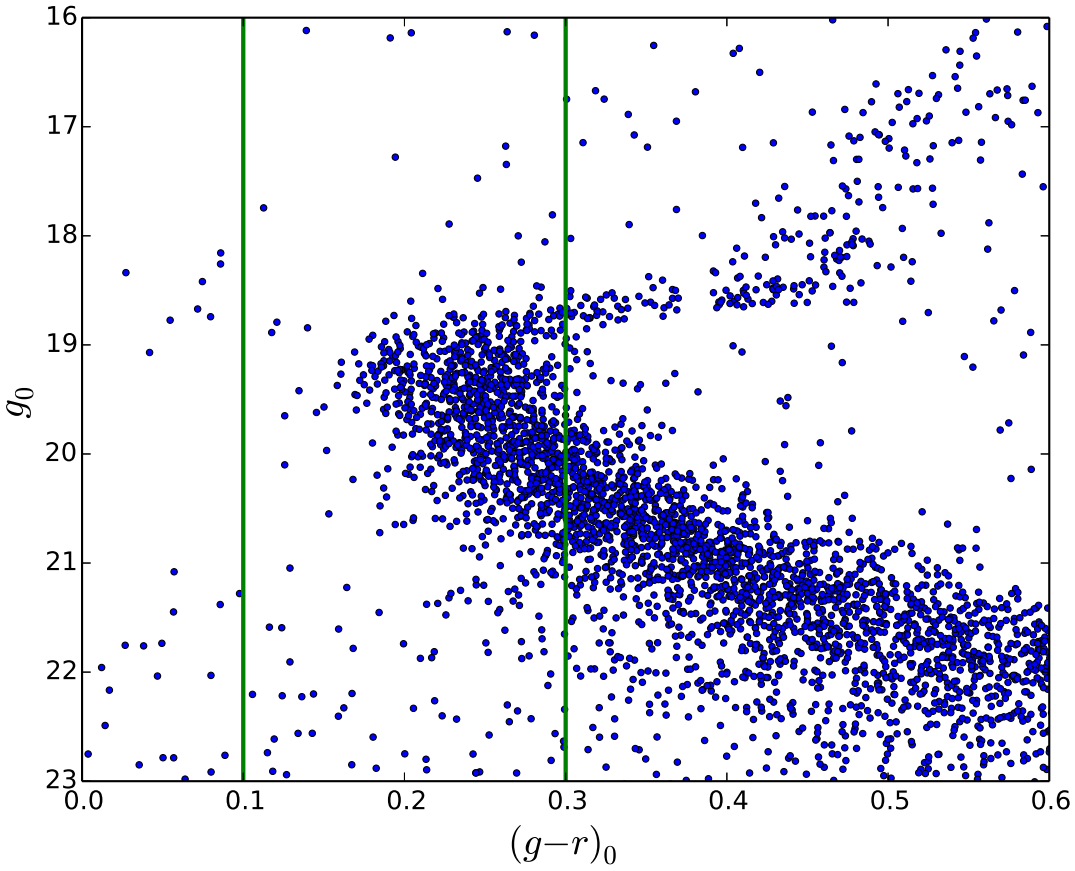


Figure 2.1: A color magnitude diagram for the globular cluster NGC 5272, which is at a heliocentric distance of 10 kpc. This figure was made using SDSS DR 14 data. Quasars are removed from the sample by selecting only stars with $(u-g)_0 > 0.4$. In this diagram we can see a main sequence, a turnoff at a $(g-r)_0$ color of 0.2, and giant branch. The green bars at $(g-r)_0 = 0.1$ and $(g-r)_0 = 0.3$ represent the color selection that we use to select MSTO stars in the halo. Since this color selection bin does a good job selecting the turnoff for this cluster, and this cluster is representative of the halo population as a whole (Newby et al., 2011), we conclude that this color selection criteria is appropriate.

CHAPTER 3
Modeling the substructure in the Milky Way halo

3.1 Parameterized Milky Way halo model

In this section we describe a stripe by stripe density model that we will test, validate and then fit to stripes 9 through 23 in the north Galactic cap. This model is adapted from a model originally used in Cole et al. (2008) and later used on the distributed computing platform MilkyWay@home by Newby et al. (2013). As color errors increase, MSTO stars are scattered outside the color selection bin and redder stars are scattered into the color selection bin. This effect is especially pronounced near the survey detection limit, where color errors are high. A major change we made to the model is the inclusion of these effects on our absolute magnitude distribution and completeness as described in Newby et al. (2011). Although the color selection range was chosen to be bluer than the turnoff of the Milky Way thick disk, we have also added a thick disk component to the smooth portion of the density profile to account for the fact that a few of these stars might have leaked into our selection.

Throughout this section we develop an MLE that measures how well a model with a particular set of parameters fits the data. This estimator is then be used to optimize the model parameters. Although we will fit three streams plus the smooth component in this thesis, the model as implemented can fit an arbitrary number of tidal streams or substructures in a given stripe. The code for this release of our model can be found at Arsenault et al. (2018).

3.2 Smooth component

We implemented two different models for the smooth component of the halo, both with two tunable parameters. One model is a Hernquist spheroid model (Hernquist, 1990; Xu et al., 2015) with a double exponential disk. The other is a broken power law (BPL) (Akhter et al., 2012) without a disk. In our model fitting, we will primarily use the Hernquist model in order to be consistent with previous versions of the model described in Cole et al. (2008) and Newby et al. (2013). First we describe the Hernquist/thick disk model, and

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., Newby, M., & Desell, T. submittedb, ApJS

then the BPL model, which we used to study the effect of an imperfectly modeled smooth component on the derived properties of the halo substructure.

3.2.1 Hernquist plus double exponential disk

The Hernquist distribution (Hernquist, 1990) is described by the equation:

$$\rho_{spheroid}(r) \propto \frac{1}{r(r+r_0)^3} \tag{3.1}$$

where $r = \sqrt{X^2 + Y^2 + \frac{Z^2}{q^2}}$; note that this is not spherical radius, but instead an ellipsoidal radius. X, Y and Z are Galactocentric Cartesian coordinates with the Sun at $(-8.5, 0, 0)$ kpc, Y in the direction of the Sun's motion, and Z in the direction of the north Galactic pole. r_0 is a scale radius and q is a flattening parameter. Using this model, there are two tunable parameters: r_0 and q . After inspecting the Hernquist distribution (Hernquist, 1990), the Einasto profile (Retana-Montenegro et al., 2012) and a broken power law distribution (Akhter et al., 2012), we found that scale radius only holds a minor effect on the overall shape of the distribution over the range of our data. In our model, we fix the scale radius at 12.0 kpc, which approximately reproduces the average of all of the distributions inspected.

The double exponential thick disk used in our model is described by the equation:

$$\rho_{disk}(R_{cyl}, Z) \propto e^{R_{cyl}/l_{disk}} e^{|Z|/h_{disk}}; \tag{3.2}$$

where $R_{cyl} = \sqrt{X^2 + Y^2}$; l_{disk} is the disk scale radius; and h_{disk} is the disk scale height. In our model, we fix the scale length and height of the disk to be the same as those found in Xu et al. (2015): $l_{disk} = 3500$ pc and $h_{disk} = 700$ pc. In Figure 5.2, we show the relative expected fractional contributions of the thick disk, thin disk and Hernquist background in the Xu et al. (2015) model at different positions in SDSS stripe 19. We do not include the thin disk in our model because our data is far enough from the Galactic plane to avoid the majority of the stars in this component, and our color cuts also help eliminate disk contamination.

Because we are using blue MSTO stars that preferentially avoid the thick disk, we cannot use a published normalization for the disk and halo, so this normalization is a fit parameter. We write the overall density of the combined stellar spheroid and disk as:

$$\rho_{smooth}(X, Y, Z) = \epsilon_{spheroid} * \rho_{spheroid}(r) + (1 - \epsilon_{spheroid}) * \rho_{thickdisk}(R_{cyl}, Z), \tag{3.3}$$

where $\epsilon_{spheroid}$ is a weighting factor that allows us to fit any mix of densities between all disk ($\epsilon_{spheroid} = 0$) and all spheroid ($\epsilon_{spheroid} = 1$). The following equation can be used to convert $\epsilon_{spheroid}$ to $f_{disk}(X, Y, Z)$, the fraction of thick disk stars at a given point:

$$f_{disk}(X, Y, Z) = \frac{(1 - \epsilon_{spheroid}) * \rho_{thickdisk}(R_{cyl}, Z)}{\rho_{smooth}(X, Y, Z)}. \tag{3.4}$$

To find the fraction of stars in the spheroid at a point, $f_{spheroid}(X, Y, Z)$, use:

$$f_{spheroid}(X, Y, Z) = 1 - f_{disk}(X, Y, Z). \tag{3.5}$$

The two parameters that we fit in the smooth Hernquist/double exponential model are q and $\epsilon_{spheroid}$.

3.2.2 Broken power law model

The BPL distribution from Akhter et al. (2012) is described by the following equation:

$$\rho_{BPL}(r) \propto (r/r_0)^n, \tag{3.6}$$

where

$$n = \begin{cases} -2.78 \text{ if } r < 45\text{kpc} \\ -5.0 \text{ if } r \geq 45\text{kpc} \end{cases}, \tag{3.7}$$

$r_0 = 8.5$ kpc is the distance from the Galactic center to the Sun, and $r = \sqrt{X^2 + Y^2 + \frac{Z^2}{q^2}}$. We use

$$q = \begin{cases} q_0 + (1 - q_0)(R_{cyl}/R_u) \text{ if } R_{cyl} < R_u \\ 1.0 \text{ if } R_{cyl} > R_u \end{cases}, \tag{3.8}$$

where $q_0 = 0.5$, $R_{cyl} = \sqrt{X^2 + Y^2}$, and $R_u = 20$ kpc as written in Keller et al. (2008).

We do not use this model for fitting the halo because the Hernquist model is more comparable to previous work. Instead, we use this model for testing our algorithm.

3.3 Stream

In each stripe, which probes a wedge-shaped volume of the Galaxy, the density of a tidal stream or cloud is fit to a cylinder with a uniform density along its length, and Gaussian

fall-off in density as a function of radius. The orientation of the stream through the wedge is described by a unit vector $\hat{a}$ with a polar angle from the Z-axis of θ, and an azimuthal angle of ϕ measured from the Galactocentric X-axis and increasing in the direction of the Galactocentric Y-axis. The position of the cylinder's axis, at the point that it passes through the center of the stripe ($\nu = 0$), is given by an angular coordinate along the stripe, μ, and a distance from the Sun, R. The (μ, ν) coordinates are SDSS great circle coordinates (York et al., 2000), which are defined separately for each stripe.

At point $\mathbf{p}$ relative to the cylinder, the stream's density ρ is described by the function:

$$\rho_{stream}(\mathbf{p}) \propto e^{\frac{-d^2}{2\sigma^2}} \tag{3.9}$$

where d is the radial distance from the stream, and σ is the standard deviation describing the stream's width (Cole et al., 2008).

The final parameter in the stream model is the stream weight, ε, which measures the fraction of stars in the wedge that belong to a stream. To get the fraction of stars in a stream or smooth component (spheroid) from the weights we use the following functions:

$$\begin{aligned} f_{stream_i} &= \frac{e^{\varepsilon_i}}{1+\sum_{j=1}^{k}[e^{\varepsilon_j}]} \\ f_{spheroid} &= \frac{1}{1+\sum_{j=1}^{k}[e^{\varepsilon_j}]}. \end{aligned} \tag{3.10}$$

Here, i and j denote the stream number for a total of k streams. The weight of a stream is used for model optimization instead of the total star counts because it is a continuous real number that is defined for all real numbers, making it a better parameter than star counts, which are discrete numbers (Cole et al., 2008).

In total, six parameters are required to fit a single stream: the 4 spatial coordinates, θ, ϕ, μ, and R, which give the stream position and orientation; the width parameter, σ; and the stream weight, ε. The model can handle an arbitrary number of streams in a single wedge.

3.4 Absolute magnitudes of MSTO stars

In our implementation of statistical photometric parallax (Newberg, 2012), we use the absolute magnitude distribution for MSTO stars described in Newby et al. (2011), which

accounts for the absolute magnitude change due to contamination from redder stars at faint magnitudes. In this implementation of statistical photometric parallax, we convolve the model density distributions (in the R direction which corresponds to our line-of-sight) with the absolute magnitude distribution. The convolution is given by:

$$\rho_{comp}^{con}(l,b,\mathcal{R}(g_0)) = \frac{1}{\mathcal{R}^3(g_0)} \int_{-\infty}^{\infty} \mathcal{R}^3(g_0-g) \cdot \rho_{comp}(l,b,\mathcal{R}(g_0-g)) \cdot N(g_0-g;g_0,u)dg \quad (3.11)$$

where $\mathcal{R}(g_0)$ is distance to a star (or position in apparent magnitude space) with an apparent magnitude of g_0 and assumed absolute magnitude of $M_{g_0} = 4.2$, ρ_{comp} is a density model, either background or stream, l and b are Galactic l and b, and $N(g_0-g;g_0,u)$ is our MSTO star absolute magnitude distribution. This integral is the combination of the MSTO star absolute magnitude distribution function with our stellar density function over all magnitude space. For an in depth explanation of this integral, see the original derivation in Cole et al. (2008).

In our model, we use the MSTO absolute magnitude distribution found in Newby et al. (2011), which is defined by two half Gaussian distributions which are normalized together:

$$N(x;g_0,u) = \frac{1}{\frac{(\sigma_l+\sigma_r(\mathcal{R}(g_0)))}{2}\sqrt{2\pi}} e^{\frac{-x^2}{2u^2}} \quad (3.12)$$

where

$$u = \begin{cases} \sigma_l = .36 \text{ if } x < 0 \\ \sigma_r(\mathcal{R}(g_0)) = \frac{\alpha}{1+e^{-(\mathcal{R}(g_0)-\beta)}} + \gamma \text{ if } x \geq 0 \end{cases}, \quad (3.13)$$

$\alpha = 0.52$, $\beta = 12.0$ and $\gamma = 0.76$. This absolute magnitude distribution accounts for an increase in the standard deviation on the faint side of the absolute magnitude distribution as a function of magnitude in the SDSS. The change in distribution is due to the change in the type of stars found in our narrow MSTO color selection criteria as increasing color errors towards the magnitude limit of the survey. Toward the survey magnitude limit, color errors increase and cause redder, lower main sequence stars to leak into our color selection as actual MSTO stars leak out. The underlying absolute magnitude distributions of all turnoff stars in the halo is assumed to be constant as found by Newby et al. (2011).

3.5 Detection and selection efficiency

To correct for the drop off in detection efficiency near the magnitude limit of the SDSS, a sigmoid curve was fit in Cole (2009) to data from Newberg et al. (2002):

$$\mathcal{E}_{sigmoid}(g_0) = \frac{s_0}{e^{s_1(g_0 - s_2)} + 1}, \tag{3.14}$$

where $(s_0, s_1, s_2) = (0.9402, 1.6171, 23.5877)$. This curve describes the survey completeness at a given magnitude. Using this curve, we can account for the decreasing completeness of the survey with increasing magnitude.

As the color errors increase, fainter main sequence stars leak into the MSTO color selection box, and MSTO stars leak out of the MSTO color selection box. The detection efficiency is adjusted using what we call the selection efficiency to account for the number of stars expected given this effect. The selection efficiency for SDSS stars in the MSTO selection bin is given by a 7th order polynomial fit by Newby et al. (2011). We reproduce it here because there was an inadvertent truncation of significant digits in the original paper that affects the result of the fit:

$$\mathcal{E}_{selction}(\mathcal{R}(g_0)) = \frac{n(\mathcal{R}(g_0))}{n_0} = \sum_{i=0}^{7} (a_{yi} + a_{ri}) \mathcal{R}(g_0)^i \tag{3.15}$$

where

$$a_y = (1.05628761, -3.14555041 \times 10^{-2}, 2.05499665 \times 10^{-4}, \\ 2.53747387 \times 10^{-6}, -2.67000303 \times 10^{-8}, 0, 0, 0) \tag{3.16}$$

and represents the remaining MSTO stars in the color selection bin and

$$a_r = (1.60879353 \times 10^{-2}, -1.97164570 \times 10^{-2}, -4.31844102 \times 10^{-4}, \\ 6.60960070 \times 10^{-3}, 1.26368065 \times 10^{-5}, -1.91560491 \times 10^{-7}, \\ 1.47140445 \times 10^{-9}, -4.53857248 \times 10^{-12}) \tag{3.17}$$

and represents the red stars that leak into the color selection bin to contaminate the sample. The percentage of stars left in the selection bin assuming 100 percent detection efficiency is

illustrated in Figure 3.1. The two detection efficiency functions are multiplied together to get the total detection efficiency for stars that fit the MSTO selection criteria.

3.6 Maximum likelihood estimator

Putting together all of the individual model components, we create an MLE which tells us how well our model, with parameters $\mathbf{Q}$, fits a set of data (N stars with measured l_i, b_i, g_i):

$$\mathcal{L}(\mathbf{Q}) = \prod_{i=1}^{N} PDF(l_i, b_i, R(g_i)|\mathbf{Q}), \tag{3.18}$$

where i is the index of each star, l and b are Galactic coordinates (Cole et al., 2008). Substituting in our density functions, correcting for imperfect knowledge of the absolute magnitudes of the stars, and taking into account the detection efficiency, our PDF is given by:

$$\begin{aligned} PDF(l, b, \mathcal{R}(g_0)|\mathbf{Q}) = \frac{1}{1+\sum_{i=1}^{n} e^{\epsilon_i}} \frac{\mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{background}(l, b, \mathcal{R}(g_0)|\mathbf{Q})}{\int \mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{background}(l, b, \mathcal{R}(g_0)|\mathbf{Q})dV} + \\ \sum_{i=1}^{n} \frac{e^{\epsilon_i}}{1+\sum_{j=1}^{n} e^{\epsilon_j}} \frac{\mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{stream_i}(l, b, \mathcal{R}(g_0)|\mathbf{Q})}{\int \mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{stream_i}(l, b, \mathcal{R}(g_0)|\mathbf{Q})dV}, \end{aligned} \tag{3.19}$$

where n is the number of streams in the model and

$$\mathcal{E}(\mathcal{R}(g_0)) = \mathcal{E}_{sigmoid}(\mathcal{R}(g_0)) \cdot \mathcal{E}_{selection}(\mathcal{R}(g_0)). \tag{3.20}$$

To avoid problems with numerical underflow due to small probabilities, we instead maximize the log likelihood:

$$\frac{1}{N}\ln(\mathcal{L}(\mathbf{Q})) = \frac{1}{N}\sum_{i=1}^{N} \ln(PDF(l_i, b_i, g_i|\mathbf{Q})). \tag{3.21}$$

3.7 Removing SDSS artifacts and globular clusters

In the SDSS, there are areas of the sky with either missing or unusable data, like the regions hidden behind a saturated bright star. There are also other sections of the sky with known structures that are not included in our model like globular clusters. We remove regions with these substructures or artifacts from our data and then integrate over the remaining usable survey volume.

The technical details of this process are as follows. We change our PDF to remove the integral of the density over the volume we are cutting from the denominator of that density's respective fraction, as follows:

$$PDF(l, b, \mathcal{R}(g_0)|\mathbf{Q}) =$$
$$\Big(\frac{1}{1+\sum_{i=1}^{n} e^{\epsilon_i}}\Big)\times$$
$$\Big(\frac{\mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{background}(l, b, \mathcal{R}(g_0)|\mathbf{Q})}{\int \mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{background}(l, b, \mathcal{R}(g_0)|\mathbf{Q})dV - \int \mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{background}(l, b, \mathcal{R}(g_0)|\mathbf{Q})dV_{cut}}\Big)+$$
$$\Big(\sum_{i=1}^{n} \frac{e^{\epsilon_i}}{1+\sum_{j=1}^{n} e^{\epsilon_j}}\Big)\times$$
$$\Big(\frac{\mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{stream_i}(l, b, \mathcal{R}(g_0)|\mathbf{Q})}{\int \mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{stream_i}(l, b, \mathcal{R}(g_0)|\mathbf{Q})dV - \int \mathcal{E}(\mathcal{R}(g_0))\rho^{con}_{stream_i}(l, b, \mathcal{R}(g_0)|\mathbf{Q})dV_{cut}}\Big), \quad (3.22)$$

where dV_{cut} represents integrating over the volume that was cut out of the data and the rest is the same as Equation 3.19.

In our data, we will either remove the stars from the globular cluster from the dataset, or we will not have data in this region due to deblending limitations in the SDSS survey.

3.8 Covariances and degeneracies

There are several degeneracies and covariances we must account for in our parametric model. The parameters which are explicitly covariant are the relative weights between the density of the streams and background. There are many other covariances that can be read from the error matrix, in Table 5.2. Matrix elements with a high value have high covariance in the corresponding parameters (Ivezić et al., 2014, pg. 128). For example, the stream parameters μ and R have high covariance. The degeneracy in stream orientation and stream selection are inherent to an optimization problem with a model of this type. The stream orientation degeneracy comes about by the cyclic nature of angles and can be mitigated by constraining the optimization to a single hemisphere. The stream selection degeneracy arises from the freedom to fit the streams in any order along the wedge. Therefore, one optimization may find stream 1 to best fit Sagittarius, while another optimization on the same data may find stream 2 to best fit Sagittarius, with the same parameters. In this instance, we do not try to eliminate this degeneracy because it does not change the result in

any substantive way, and attempting to force a particular stream order would require us to artificially constrain streams to mutually exclusive regions and could bias our optimization.

3.9 Calculating uncertainties

We calculate parameter uncertainties using the method outlined on page 128 of Ivezić et al. (2014) for calculating uncertainties for MLEs. The uncertainty is related to the Fisher information matrix which can be calculated as the negative of the Hessian matrix of the log likelihood. The equation for this matrix is:

$$H_{ij} = -\frac{d^2 ln(L)}{d\theta_i d\theta_j} \Bigg|_{\theta=\theta_0} \tag{3.23}$$

where H_{ij} is the Fischer information matrix. The variance matrix, $\mathbf{V}$, is the inverse of the normalized Fischer information matrix:

$$\mathbf{V} = \frac{1}{N} H_{ij}^{-1}, \tag{3.24}$$

where N is the number of stars. Then the standard deviation for each parameter is calculated by taking the square root of the diagonal elements of the variance matrix. As mentioned on page 128 of Ivezić et al. (2014), these are the lower bounds on the uncertainties. When we do this calculation, we use a central finite difference around the best parameter set given by our optimizations; tests have shown that the second partial derivatives are not sensitive to the (small) stepsize.

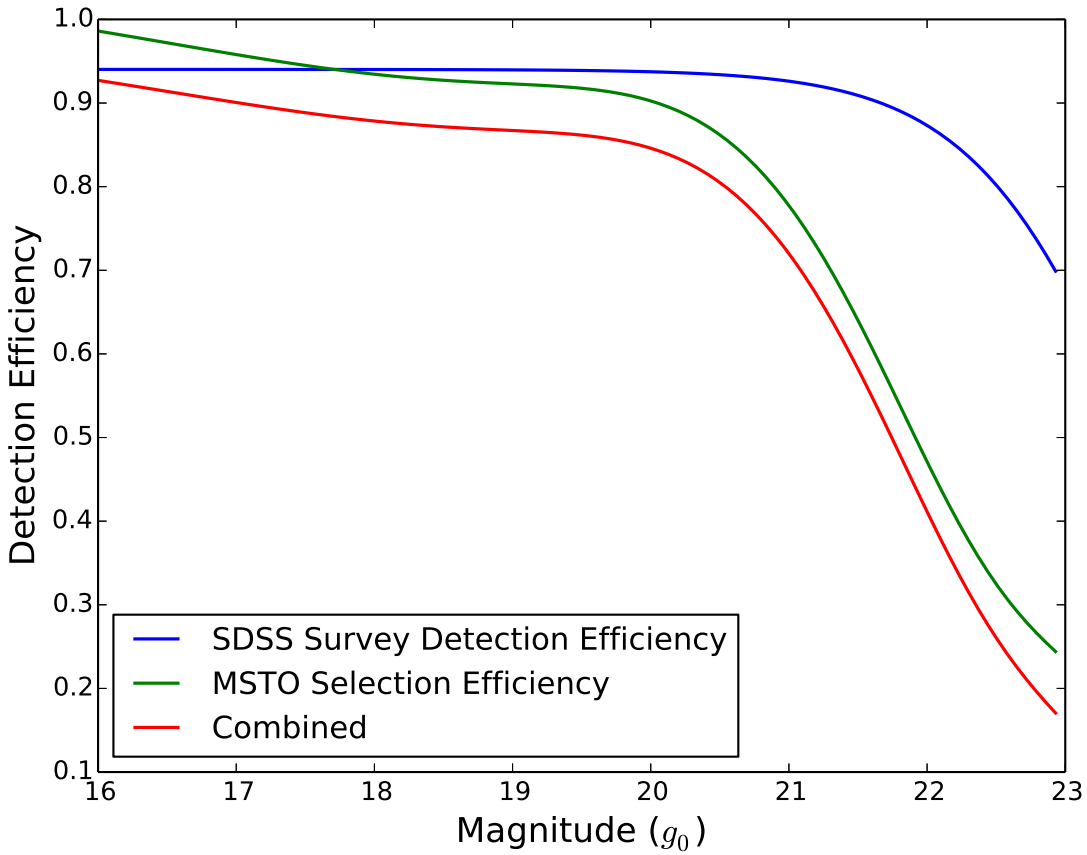


Figure 3.1: Detection and selection efficiency as a function of magnitude. The blue line shows the detection efficiency for the SDSS as fit by Newberg et al. (2002), the green line shows the detection efficiency correction to account for stars leaking in and out of the color selection bin (Newby et al., 2011) and the red line shows the overall combined detection efficiency for MSTO stars in the SDSS survey. The inclusion of the selection efficiency drastically changes the completeness for MSTO stars around 20th magnitude and dimmer.

CHAPTER 4
MilkyWay@home

4.1 Overview

MilkyWay@home is a distributed computing platform. Volunteers from around the world donate their unused CPU and GPU cycles to run "workunits," which are evaluations of the likelihood for a particular set of parameters. The MilkyWay@home server, running the Berkeley Open Infrastructure for Network Computing (BOINC) (Anderson et al., 2005), sends each volunteer an executable that will calculate the likelihood, a set of model parameters to try, and a set of Milky Way stars. The users return the likelihood, and then the MilkyWay@home server uses a differential evolution algorithm to determine the best parameters to send out in new work units, given the newly calculated likelihood and the previously returned likelihoods. For a graphical representation of this process, see Figure 4.1.

Some of the major challenges associated with volunteer distributed computing networks like MilkyWay@home are: its massively asynchronous nature, its potential for faulty results, and the need to support multiple software and hardware configurations. The project's asynchronous nature arises from the discrepancy between compute times for CPU and GPU platforms, and the episodic availability of each processor. A CPU typically takes around 30 minutes to complete a single likelihood calculation, while a GPU can take as little as 15 seconds to complete the same calculation.

This chapter will discuss how our project is set up and we solve its many challenges. We will focus our discussion on how MilkyWay@home handles the application we call, "separation," but is actually running statistical photometric parallax. Separation was the original project on MilkyWay@home. MilkyWay@home has been extended to also include an Nbody application which is still under development and testing.

4.2 Optimization with differential evolution

The differential evolution optimization algorithm used by MilkyWay@home is part of the Toolkit for Asynchronous Optimization (TAO) from Desell et al. (2007). This algorithm

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., Newby, M., & Desell, T. submittedb, ApJS

is a modified version of differential evolution for use in highly asynchronous environments such as a distributed computing platform which has heterogeneous, distributed, fault-prone hardware, and was originally designed specifically for MilkyWay@home.

This version of the differential evolution algorithm begins by uniformly sampling across the parameter space within the allowed search area (as determined by the constraints on each parameter). As these random sets have their likelihoods computed, it inserts them into a population. In this context a member of a population is a complete set of model parameters (in this case 20 parameter). Once it has enough returned results for the population to be filled (for example we typically use a population size of 200, which is ten times the number of parameters being fit), returned likelihoods will then determine whether a member of the existing population should be replaced with the new parameter set.

At this point, the algorithm begins trying to replace individuals within its population. To generate new parameter sets for evaluation the algorithm uses the following method: (1) A "parent" is selected from the population using one of several methods. (2) Two population members are randomly selected, and the differential vector between them is calculated. (3) This differential vector is scaled by a fixed differential scaling factor before adding it to the parent's parameters to identify a new position in parameter space. Finally, (4) the individual potentially being replaced is combined in some way with the parameters of newly determined position in parameter space. This crossing can be done using one of several strategies, such as assigning a random chance of taking a parameter value from the parent over taking a parameter value from the population member for the new test parameter set. After this crossover, the generated set of parameters obtained is sent out as a work unit.

To make the algorithm asynchronous, TAO blurs the lines between generations of populations in this algorithm, allowing for new parameter sets to be calculated even while waiting for results for the next population (Desell et al., 2007). As results come in, the newly computed likelihood is compared with the likelihood of the population member it is attempting to replace and the parameter set with the better likelihood is kept.

We use a set of options that we have found consistently produce convergence within a reasonable time. With these options, the algorithm is: (1) Parents are selected at random from the current population, which has 200 members. (2) Two population members are selected at random, and the differences between their parameter values are calculated. (3) The parameter value differences are multiplied by 0.8, and then added to the parent parameters.

(4) This point in parameter space is then combined with the parameters in the parent using binary recombination, a crossover rate of 0.9, and a parent scaling factor of 1.0.

These parameters allow the optimizers a large number of possible guesses to reduce the chance of optimizer stall out and reduce the probability of getting stuck in a local minimum. Typically, it takes between 1.5 and 2 million returned likelihood calculations to complete a single optimization of our parameters. We often run 4 optimizations of a single parameter set to ensure agreement between the runs and help determine convergence.

We know a run is out of energy and converged when the change in likelihood between updates of the algorithm become small for a long period of time and the results from the independent runs on the same data agree. When the likelihood no longer changes in the 6th decimal place for 100,000 returned results and the independent runs agree in the 6th decimal place we assume the run is converged. We have found that once the optimizer is making improvements on this scale, the change in the parameter values for a given improvement in likelihood is much smaller than the associated error in the parameter.

4.3 BOINC

The BOINC platform facilitates communication between both the server and client applications on our distributed computing network, MilkyWay@home. Specifically, BOINC allows users to sign up for the project and manage their computational preferences and allows developers to create client-side applications, deploy a server to create and distribute workunits, and run a web interface for their project. Projects are completely configurable by the developers and any of the open source BOINC routines can be altered to suit the needs of the project as required.

To sign up for a project, users must install the BOINC manager and then sign up for projects through the client. First, users download the BOINC manager from https://boinc.berkeley.edu/. Next, they click "Add Project" on the window. This button opens a new window. On this new window, they select "Next," then find the desired project on the project list (MilkyWay@home for our project). They select the project they wish to support and then click "Next." This brings up a form to input an email and choose a password for a new account. This form is submitted by selecting "Next," and then "Finish." With that, the BOINC manager is attached to the project and ready to crunch.

BOINC and the MilkyWay@home website both offer a list of configurable settings

that allow for personalization of the running of workunits on your computer to make it less intrusive and more efficient for each user. The options on the MilkyWay@home website include project specific options such as: should CPU or GPU be used for this project, which MilkyWay@home projects should be run, and some outdated options for the Nbody screensaver. On the BOINC manager, a user can select CPU usage options, GPU usage options, resource allocation, network usage options and disk usage options. There are further options that can be configured through the use of "app_config.xml" files which allow for multiple workunits to run on a single GPU.

The provided web interface for the project includes a website, a forum, and project administration tools. BOINC provides website, configuration, and database templates. It also provides several setup scripts to help set up a server and the database. The website template includes a CSS style sheet as well as several PHP and HTML based webpages to show general project information, project server status, and user account settings or profiles. The forums are useful for helping developers interface with the community for feedback on the project and report issues. It also facilitates the community to troubleshoot frequent project issues or discuss configurations for their computers. It is extremely helpful in passing important news to the users and allows developers to push notifications directly to volunteers' BOINC managers upon posting. The administrative tools are available on the password protected "ops" portion of the website. The administrative accounts and passwords are managed using ".htaccess", a tool provided by the Apache2 server.

The BOINC administrative tools give a website interface with most of the tools necessary to do basic debugging, user management and project statistical analysis. It is possible to browse most of the database from this web interface. The provided pages allow for table querying based on "where" clauses in a MySQL statement and return most of the database columns from the query. It should be noted however that not all columns are shown here and manual database access is still necessary to see all available data for all tables. The input fields for these tables are not SQL escaped, meaning it is possible to use SQL functions in your queries and they are at risk for SQL injection attacks. With this in mind, only trusted developers should be given access to the "ops" tools on the website.

Within "ops", there is a tool to manage which applications are current and which are deprecated. All old versions of an application should be deprecated upon release of new application versions to prevent the old applications from being sent to users. If old

applications are sent to users, there can be high error rates, and incorrect results entered into the optimization populations. If a new application is determined to be unstable, it is possible to fall back to an older version of the application by deselecting the deprecated box and deprecating the newer application. This will allow the old application to be sent to users again and prevent the newer application from being sent out.

The last major tool in "ops" provides premade queries and pages to view important statistics pertaining to the overall health of applications. These pages allow error statistics over the last 24 hours or 7 days for all subprojects on MilkyWay@home. The errors can be grouped by application version, by host type, or error type. Common errors include "missing coproc errors" which occur when a user downloads a GPU application, but does not have a double precision application, or "aborted by user" which occurs when a user manually cancels a workunit on their computer. There is also a summary page which displays the number of successful, inconclusive and errored workunits in the time increment.

BOINC facilitates communication between the client and server using HTTP requests from the client to the server application. These requests are processed through the back end BOINC libraries on the client and BOINC and Apache2 server daemons. Each time a request for work is sent by the client, the server opens a database connection, selects workunits from the workunit queue, and then responds with the relevant application files, data files, and commandline from the server. Any dynamic information, such as parameters for a specific workunit, are appended to the end of the commandline from the server after a "-p" flag.

4.4 Server

4.4.1 Hardware

MilkyWay@home runs on a single Dell Poweredge server housed in the server room in the Voorhees Computing Center on the Rensselaer Polytechnic Institute campus. This server room offers a temperature controlled environment with a stable internet connection and power supply. With these resources, we are able to maintain 24 hour 7 day a week uptime for our project and website with few outages.

The Dell Poweredge server can service as many as 1000 concurrent database connections and HTTP requests. It has an 8-core Intel Xeon x5647 CPU in addition to 70 gigabytes of RAM to allow for fast processing of requests and the full database to be loaded into RAM. By loading the full database into RAM, we ensure quick database access to all of the tables. This

is important as most tables are constantly accessed to insert and query data. To maximize reliability of the server, we also run several hard drives in a hybrid RAID configuration to both increase the server's disk storage capacity and redundantly store information in case of disk failure. In addition to redundant hard drives, we also run redundant power supplies. If either were to fail, the other would immediately take over resulting in no downtime on the machine.

While this hardware is currently sufficient for our task, this server is at its upper limit for upgrading. While in a typical Dell Poweredge, there are two slots for CPUs and several additional slots for RAM, this particular server has a faulty second CPU slot. Due to this faulty slot, we are unable to add a second CPU or the corresponding memory with that CPU.

4.4.2 Toolkit for asynchronous optimization (TAO)

TAO is a software package responsible for creating and validating workunits for the MilkyWay@home server. This software is also responsible for storing and optimizing the parameters in an optimization run on MilkyWay@home. Workunit creation is done through a server daemon called the "Workunit Generator," and there are separate ones for separation and N-body. Workunit validation is done by the BOINC validator with a TAO extension. This extension is then linked into the BOINC server code, and runs in the validator daemon.

The workunit generator is responsible for taking the current population of parameters in the MilkyWay@home database and turning them into new workunits for the volunteers to compute. The generator can bundle several parameter sets together to allow multiple likelihoods to be calculated in a single workunit. We bundle workunits to reduce server load caused by workunit requests when fast single workunits are run. TAO includes implementations for several different optimization algorithms which all could, in theory, be used to generate new workunits. Currently, only differential evolution and particle swarm algorithms are available on the server code and only differential evolution is tested with workunit bundling.

The workunit generator has a pair of global parameters that can be changed if the server is struggling to keep up with workunit demand, "SLEEP_TIME" and "CUSHION." The "SLEEP_TIME" determines how much time is spent waiting between workunit generation cycles, and the "CUSHION" is the number of workunits the sever will attempt to keep in

reserve for users to request. Having the program sleep after generating new workunits is important to allow the database to propagate the inserts and ensure the number of available workunits in the database is correct next time the work generator checks. If this time is set too small, the workunit generator will generate an excessive number of workunits which is counterproductive and floods the workunit queue. It can also bog down the database as it has to handle an extremely high load from the inserts in the workunit generator. Having a "CUSHION" set too low will cause the server to run out of workunits before it generates new ones and having it set too high can cause a backlog of workunits that need to be calculated.

To use new optimization algorithms from TAO, one would have to make changes to the CMake build scripts, and the workunit generator daemon. These changes would include, but are not limited to, adding commandline parameters, adding database tables to hold relevant optimization information, and adding the desired code files into the build scripts. After compiling the new workunit generator with the added commandline parameters exposing the desired optimization algorithm and its run flags, the server binaries are updating using the update scripts on MilkyWay@home. All of this source code can be found in the "/boinc/src/" directory on the MilkyWay@home server.

The TAO validator extension helps prevent erroneous workunits or bad actors from influencing the outcome of our optimization. TAO performs adaptive cross validation on our returned likelihoods. When we receive a result, it might have to be recomputed by up to 4 more users, depending on the results of previous users. For trusted users who compute a likelihood that does not change the population, we validate a minimum of 10% of their workunits. The percentage validated changes based on the percentage success of their previous 10 workunits. All results which will be integrated into the optimizer's population must be cross validated by another volunteer, who must agree on the computed likelihood for this parameter set. If these volunteers do not agree, the work unit is sent out to additional volunteers until 2 volunteers agree on an answer. If 5 attempts to compute the likelihood do not result in 2 answers that agree, the parameter set is abandoned.

The TAO database tables include: "differential_evolution," "differential_evolution_log," "particle_swarm," "particle_swarm_log," and "tao_workunit_information." Both "differential_evolution" and "particle_swarm" hold their respective optimization flags, as well as state information about previous and current runs. "Differential_evolution_log" and "particle_swarm_log" both contain information about the previously returned workunits. In par-

ticular, these tables log the return of workunits which improved the likelihood of a given run. Finally, "tao_workunit_information" holds the overall information about a MilkyWay@home optimization run. This includes information about workunit bundling, required files (star files and parameter files for separation), and any commandline flags that need to be passed on to the client code for a specific run.

To use the TAO visualization tools, there is a public TAO page at http://milkyway.cs.rpi.edu/milkyway/tao/display_searches.php. This page allows for the selection of MilkyWay@home runs and plotting their best results over time. Suggested parameters for these plots for the separation application are "first evaluation = 100000" and "select every nth value = 100." By using both of these parameters, the likelihood results returned near the beginning of the run (which are mostly noise) will be removed. Inspecting the curve in this method can show how well converged an optimization is based on how frequently the optimization is improving and by how much it is improving. Later in a run, you may want to increase the "first evaluation" so that the y-axis of the plot has better resolution and it will be easier to determine whether the run has converged.

4.4.3 Custom server code

In addition to the TAO and BOINC code, we also have custom server code running on MilkyWay@home. This code determines the Floating Point Operations (FPOPs) per workunit, reads in parameter files, bundles multiple workunits, and alerts developers if it detects errors in the server daemons.

The code to calculate FPOPs is found in the "milkyway_server" repository, or "/boinc/src" on the server. These FPOPs are used to calculate the number of credits awarded to users for completing a workunit. To calculate the total FPOPs for a single MilkyWay@home workunit, the algorithm takes the number of stars, the step sizes in the integral, and the number of streams into account. The step size and number of streams are pulled from the parameter file provided when an optimization run is started, and the number of stars is pulled from the first line of the star file. For bundled workunits, the calculated FPOPs for a single workunit is then multiplied by the number of bundled workunits to get the total FPOPs used for credit calculations.

This code must be updated and recompiled every time a new function is added or removed from the client density model. The current method for updating the FPOPs count

is to profile the run time of the old and new client versions and multiply the FPOPs calculated by the percentage difference in run time on a CPU. More exact methods could be used to find the exact number of FPOPs required to calculate a workunit. These methods include profiling each section of the integral and likelihood code on a CPU with a known FPOPs per second to determine the number of FPOPs required throughout the application and then adding them together at the end. In general, I have found this level of accuracy is unnecessary for our application.

The parameter file interpreter reads the number of integral steps and streams from the parameter file as well as optimization flags. The main optimization flags are the "optimized" flags and constraints associated with each individual parameter in the parameter file. If a parameter is not selected as optimized, it will not be included as part of the optimization and will not be sent to the client. The constraints are used to tell the optimizer where the optimization should look for each parameter. The optimization flags work well for the N-body application, but are not currently supported by the MilkyWay@home separation client and should not be used.

The Milky Way Alerts script is a custom python script that checks the status of the MilkyWay@home daemons and memory usage for anomalies via Cronjob every 5 minutes. If an anomaly is detected, it alerts the MilkyWay@home developers by email. This code uses the login credentials of the general MilkyWay@home g-mail account and a python SMTP server to email out to developers. All active developers should have their email addresses added to the "AlertRecipients.txt" file in the "/home/boinc/Scripts/MWAlert" directory. This email is sent every 30 minutes until the script detects that errors on the server have been resolved. If an error will take a long time to fix, these notifications can be postponed by writing the number of minutes to postpone the notification by in the "SleepTimer.txt" file in the "/home/boinc/Scripts/MWAlert" directory.

4.4.4 BOINC configuration

The BOINC server configuration options are found in the config.xml file in the "/boinc /milkyway/" directory. These configuration options include: the project name, project directories, the project URL, the database login information, options to limit and change how clients interact with the server, ReCaptcha keys, and project daemon configuration. Given the sensitive nature of some of these options, this file should be kept secure and never

shared with non-project developers. In addition to the BOINC configuration files, there are configuration files for the MYSQL database and Apache2 server, these will be covered in Sections 4.4.5 and 4.4.6, respectively.

In the BOINC configuration file, there are XML tags for different directories within the project directory, such as the "apps" directory, the "download" directory, and the "log" directory. The "apps" directory is where client versions are stored on the server after they have been released. Applications in this directory should not be deleted and they should not be replaced. The server expects each application in this directory to have a corresponding signature file to ensure that that they have not been tampered with since release. That also means that if an application is released with a specific name, another application can not be released with the same name. The "download" directory is where the server will store any files, like star files or parameter files, which are downloading in conjunction with a workunit. The "log" directory is where the server stores all of the server daemon logs for the BOINC, TAO and custom server code. These logs are extremely useful for debugging server issues. On occasion it is useful to rotate these logs and delete old logs because they will grow large over time.

The configuration options for server-client interaction include the minimum time between requesting workunits, the daily result quota, maximum stored workunits, maximum work in progress, max GPU workunits, and max CPU threads. The minimum time between requests is important to reduce server load and prevent denial-of-service type attacks. It also prevents users from intentionally or inadvertently requesting and aborting workunits, which can cause other users to experience invalid result errors as the server decides to abandon workunits with too many errors. The maximum stored workunits corresponds to the most workunits a user can be assigned at any given time. If this number is set too high, users can begin requesting a large percentage of the available workunits from the workunit queue. Several users requesting at the same time can drain the queue and cause out of work errors. The goal is to balance the time between workunit requests and the maximum stored workunits to keep the fastest GPUs sated for the minimum work request time. Preferably they should have enough work to keep them crunching for several times longer than the minimum work request time to reduce network load.

We run a standard server daemon configuration with the exception of adding daemons for the workunit generators, assimilators and validators. Both N-body and separation have

their own daemons which are added on to the end of the daemon file.

4.4.5 Database

BOINC and MilkyWay@Home share a database with tables that fall into three categories: user data, work data, and TAO data. There is overlap in each of these categories; for example, user data and work data are cross correlated to determine user credit and statistics tracking, and work data and TAO data are cross correlated for optimizations. We have added columns to the default tables for things like badges, and run tracking. The database is also customized through its configuration file. These customizations allow us to optimize query speed, reliability, and uptime.

The user data tables hold information about the users, teams, website, and forums. The user information includes names, email addresses, computer hardware information, badges earned, credit earned, recent workunits completed, and team participation. Sometimes, it can be necessary for developers to search through this information during bug reports. For example, client error codes and debugging information can be found in the "stderr.txt" files returned by a user after a workunit is complete. This can be found by cross referencing the user's ID with the user ID stored in the results table and then reading the output column. In the forum tables, there is information about user posts, private user messaging, and forum thread statistics. In the team tables, the team founder and administrators are recorded along with other membership information and statistics.

All of the information for current workunits and results are stored in the work data tables, "workunit" and "result". The workunit table stores state data, validator data, assimilator data, and canonical result data. After a workunit is completed and assimilated, it is marked for deletion in the database to reduce table size. Similarly, the result table stores information about a specific workunit run's outcome after being computed by a user. Since we run workunits several times on different computers, there are often several entries in the result table per workunit. Results in the result table map to workunits in the workunit table using foreign keys stored in the result table.

The TAO tables were described previously in Section 4.4.2 and mainly hold information necessary for optimizations. In addition to the optimization data, the optimizer log tables are used for visualizing the current progress of optimizations.

The MilkyWay@home database configuration file can be found on the MilkyWay@home

server at "/etc/mysql/mysql.conf.d/mysql.cnf." This file is currently set to allow 1000 concurrent database connections and the server frequently sees in excess of 400 concurrent database connections. In order for this configuration option to work, the database daemon's systemd file, located at "/lib/systemd/system/mysql.service" must be configured to have a "LimitNOFILE" greater than the open files limit. The open files limit should be 4 or 5 times larger than the required maximum connections. Having a connection limit this high is important due to the large number of concurrent database requests for workunits and website data. Without the connection limit, the server will run out of available connections and prevent users from interacting with the website or downloading new workunits. A main cause for the large number of concurrent connections, in addition to having a large number of users, is BOINC taking too long to close or timeout connections from users. This may be something worth looking in to in the future if we start to reach the connection limit again.

To improve database performance, we attempt to have as much of the database stored in RAM, as opposed to on disk, as possible. The database is currently set to use the 46GB of RAM while the rest of the RAM is reserved for other system processes and connection overhead. If this number is configured too high then the server will begin to use swap memory which slows it down considerably. We have found once the server begins using swap space, it becomes too slow to catch up with its query queue and will stay stuck on swap memory. Currently, we have the server set to avoid using swap memory until 95% of the available RAM is used to prevent it from using swap prematurely.

The database is currently set up to run a backup on Sunday nights, once per week. These backups do a full database dump into the "/boinc/mysql/" directory and then copy the dump file to another computer on the RPI network. During these backups, database queries can be slow and the connection queue often grows large. Some symptoms of this include an increase in the workunit queue as insert queries back up from the workunit generator and a large validator queue. Currently, we do not have a solution for these symptoms and it may be possible that more database optimization should be done to reduce query and backup times in the future.

4.4.6 Website

The MilkyWay@home website is run on an Apache2 webserver configured with PHP7 and MYSQL bindings. The website is a slightly modified version of the BOINC back and

front end code which allows it to run on PHP7 instead of PHP5. The Apache2 configuration files can be found at "/boinc/milkyway/milkyway.httpd.conf" and "/etc/apache2/conf-enabled/httpd.conf." These configuration files set website aliases, permissions for the main user sections of the website, and password protect the sensitive administrative sections of the website.

The main website code exists in three different directories: "/boinc/milkyway/html /user," "/boinc/milkyway/html/ops," and "/boinc/milkyway/html/inc." The first directory holds the front end code for the website pages and forums. The second directory is password protected and contains site administration and analytics tools. The third directory contains the back end code and is shared code among the main and administrative sides of the website.

In the ops directory, there are tools for checking the error statistics for different runs, looking up hidden user host data, and for managing users and teams. The error statistics are useful during the release of a new client version or during the debugging of large scale issues to see if they are localized to specific platforms or hardware types. Sometimes users will post questions or issues to the forums, but they choose to keep their hardware hidden from other users. The user search tool can be used in conjunction with a user ID to check hidden host data, including hardware and recently returned workunits. The user and team management features are used when volunteers have decided to use the website inappropriately. Users can be banned or suspended depending on their behavior and teams can be deleted. These tools have been necessary in the past to ban users or delete teams who have posted or advertised explicit material or services on the website.

The "inc" directory houses much of the database connection and back end code used throughout the website. This directory is the most likely to have compatibility issues in any future server or code update due its technical nature. This code includes password validation, string sanitization, mailing and much more.

Due to the customized nature of the website code, it is not recommended to update it from the main BOINC Github repository.

4.4.7 Managing optimizations

The typical work flow for running a separation optimization is as follows: create star and parameter files, place star and parameter files on the server, start the optimization, monitor the optimization's status, stop the optimization, and analyze results.

Star files should be space separated coordinate files with the first line containing the number of stars in the file. Each subsequent line should list the l, b, and r of a star assuming an absolute magnitude of 4.2. Each star file should be limited so that stars only exist within the integral area defined in the corresponding parameter file. Additionally, stars should not be present within cut regions defined in the parameter file.

Parameter files should be set up following the example file in Appendix A. Since the parameter file reading in the client and server is rather primitive, the syntax and layout of the parameter file should be followed as closely as possible, limiting changes only to data values. The number of streams fit can be changed, but if it is, the corresponding number of stream sections in the parameter file should also be changed.

The data values in the parameter file are grouped into several sections: background, streams, stripe data, and cuts. The background section is broken into two different subsections: one for the background weight and one for the other background parameters. The "background_weight," as it is called in the parameter file, describes the weighting between the disk and the smooth background component and corresponds to ε_{sph} from the model. The file also defines the minimum and maximum values for this parameter during optimization. These are set by modifying the parameters on the "background_weight_min" and "background_weight_max" lines. The second subsection contains the rest of the background parameters, listed on the "background_parameters" line. The first and last parameters on the line are dummy variables. The second parameter corresponds to the halo flattening q, and the third parameter corresponds to the now deprecated halo scale length. We left the halo scale length in the parameter file so future users can reuse the variable or reimplement the scale length as needed. The minimum and maximum optimization values can be set on the "background_min" and "background_max" lines.

The "number_streams" line begins the stream section of the parameter file. This line defines the number of streams in the optimization (typically 3) and the number of non-weight parameters for each stream. Each stream is given a subsection in the stream section with a weight value, weight minimum, weight maximum, a parameter list, and lists for the minimums and maximums of the non-weight parameters. The "stream_weight" line corresponds to the ε for each stream from the model. The "stream_weight_min" and "stream_weight_max" lines correspond to the minimum and maximum values for the weight in the optimization. The stream parameters are listed on the "stream_parameters" line in the order μ, R, θ, ϕ,

and σ. The lines "stream_min" and "stream_max" are the corresponding minimums and maximums of these parameters during optimization.

The stripe data section begins with the "convolve" line. This line corresponds to the number of steps taken in the client's convolution integral. The "wedge" line corresponds to the stripe the parameter file will be used on. The "r[min,max,steps]," "mu[min,max,steps]," and "nu[min,max,steps]" lines define the stripe limits in magnitude and degrees as well as the integral steps in each direction. The choice for steps must be an even number for the integrator to work.

In the cuts section, beginning with "number_cuts," each cut is defined. The number of cuts in the stripe should be put on the "number_cuts" line. For each cut, there should be three lines defining the minimums, maximums, and steps for each cut. These should be listed next to "r_cuts[min,max,steps]," "mu_cuts[min,max,steps]," and "nu_cuts[min,max,steps]." These steps should be scaled to keep the integral steps to the original steps over the cut region. The step sizes must be even similar to the original integral.

The parameter and star files should be copied onto the MilkyWay@home server. The star file should be placed in the "/boinc/src/milkyway_separation_assimilator/star_files/" directory and the parameter file should be placed in the "/boinc/src/milkyway_separation_assimilator/parameter_files/" directory.

To start a new differential evolution run, navigate to "/boinc/milkyway/bin" and run the "/stream_fit_start_search" application. An example command line for a three stream run is:

```
sudo -u boinc ./stream_fit_start_search --command_line "-f" --bundle_size 5
    --parameters <parameter file path> --stars <star file path>
    --search_name de_<search name> --population_size 200
    --differential_scaling_factor 0.8 --crossover_rate 0.9
    --parent_selection random
```

where the sections bracketed by "<" and ">" must be filled in by the user. Note that the search name must start with "de_" to start a differential evolution run. The other parameters are "- -command_line" which passes a command line argument directly to the client. The "-f" command line tells the client to run the "modfit" version of the client which includes the new absolute magnitude distribution for MSTO stars and the selection efficiency for

MSTO stars. It is also possible to include the "-y" command line option here to run the client with the broken power law background instead of the Hernquist background. The "--bundle_size" option selects the number of workunits to bundle together. Finally, the optimization parameters of "- -population_size," "- -differential_scaling_factor," "crossover_rate," and "parent_selection" are set according to the criteria described in Section 4.2. The "bundle_size" and "population_size" should change if the number of streams is changed.

Once a run is started, its status can be checked either from the TAO visualization tool at http://milkyway.cs.rpi.edu/milkyway/tao/display_searches.php or by using the "tao_search_status" application in the "/boinc/milkyway/bin" directory. The use of the TAO visualization tools is described in 4.4.2. An example command line for the "tao_search_status" application is:

```
./tao_search_status --app milkyway --search_name <search name>
    --print_best 1
```

where "<search name>" is the name of the optimization that you are checking. The "--app" option is mandatory and can be either "milkyway" or "milkyway_nbody" depending on which application your optimization is running on. The "print_best" option is used to specify how many results to display.

When a run is complete it can be stopped using the "tao_stop_search" application in "/boinc/milkyway/bin" on the MilkyWay@home server. A typical command line for this application is:

```
./tao_stop_search --app milkyway --search_name <search name>
```

where "<search name>" is the name of the optimization that you are stopping. The "--app" can be either "milkyway" for separation or "milkyway_nbody" for the Nbody application. This application works by setting the "maximum_created" column to 1 in the "differential_evolution" table in the database. Occasionally the stop search functionality will fail and the run can manually be stopped using an update query on this database column. Additionally, a run can be restarted by setting this column back to 0.

4.5 Client

4.5.1 Bundle workunits

The current version of bundled workunits was implemented when optimizations to the client code improved its runtime by almost an order of magnitude. As users were able to drastically increase their throughput of workunits, the increased requests for new workunits from the server crippled the database. By bundling multiple workunits together, we were able to increase the runtime of the workunits on volunteers' computers without adding unnecessary calculations and in this way reduce the load on the server.

Practically, the bundled workunits are implemented on the client by calling the compute integral and likelihood functions multiple times for each set of parameters. After each likelihood computation, the results are appended to the output file and numbered for later use by the server. Additionally, the program checkpoints after each likelihood calculation to prevent loss of progress if the program is paused.

The way workunit bundling is currently written, all parameters must be present from the server. If the parameter file is set to ignore optimizing a specific parameter, this will cause an error on the client. The program will handle an arbitrary number of streams, so long as all of the parameters are present on the commandline.

A problem with using the bundle workunits feature is the limited size of the commandline. It was found that when too many workunits are bundled, the commandline buffer runs out of memory due to the number of characters describing the parameters. To avoid this, we set the precision of commandline arguments to 6 decimal places and limit the number of workunits bundled. We have successfully run 5 bundled workunits with 3 stream runs and 4 bundled workunits with 4 stream runs. Changing the number of bundled workunits while changing the number of streams also has the added benefit of maintaining a more constant computation time for volunteers.

4.5.2 Adding density models

When adding density models to the MilkyWay@home code, there are several places in the code where the new model must be added. There are several versions of the code written to run on CPUs and an OpenCL version designed to run on GPUs. The CPU versions include a C code version of the model and several hand vectorized CPU intrinsic versions of the the density model. It is important to check that the same likelihood is given across all

versions of the code before release.

The C code for density models can be found in the "probabilities.c" file and several switches must be added to check which model should be used at runtime. The CPU intrinsic versions of the density models can be found in either "probabilities_sse2.c" or in "probabilities_avx.c." The CPU intrinsic versions are written for Streaming SIMD Extensions 2 (SSE2) and Advanced Vector Extensions (AVX) compatible processors. These will run on most modern processors and have vastly improved runtime over the C implementation. While we have implementations for both SSE2 and AVX, we only compile the SSE2 version and the C versions of the density model into the final release code. When determining which version of these density models to use, the client uses the functions in the "probabilities_dispatch.c" file. Any new density functions should be registered there.

The OpenCL version of the code can be found in the kernel files in the MilkyWay@home source code. This version of the code is written with a C-like syntax, and density models can be added following the format of the previously written density models. To add switches to keep track of which density model should be used, you must add new compile flags to the kernels in the "cl_compile_flags.cpp" file.

4.5.3 Changing absolute magnitude distribution

The absolute magnitude distribution code can be found in the "r_points.c" file. The desired absolute magnitude functions should be added to the "calcRConstsInt" function and "calcRConstsLik" function which calculate the star convolution for the integral and likelihood, respectively. A switch should be added to ensure only one of the absolute magnitude distributions coded is used and flags should be added to the commandline for the program to choose between them. Any physical absolute magnitude distribution should be possible in these functions, for example normalized bimodal distributions that model blue stragglers and blue horizontal branch stars or Gaussian distributions for MSTO stars.

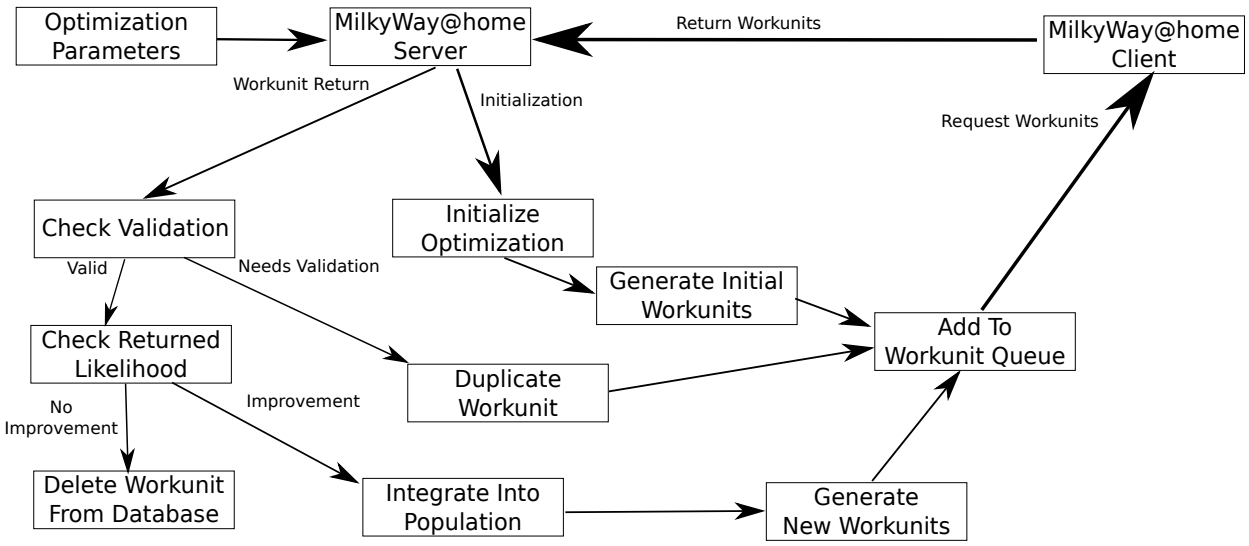


Figure 4.1: Basic flow of the main work loop on the MilkyWay@home server. An optimization is started by providing the server with a set of optimization parameters. When a new optimization is started, a short process runs to initializes a new optimization instance and insert it into the database (denoted in the flow chart by "Initialize Optimization"). With this new optimization detected in the database, the server's workunit generator will generate an initial set of workunits for the optimization and push them to the workunit queue. The main work loop takes returned workunits from the clients, checks to see if they are valid, then checks to see if they have improved the likelihood, and if it has, integrates it into the new population. If a workunit is determined to need validation, it is duplicated so it can be computed by another user and then pushed to the workunit queue. If a workunit is validated and improves the likelihood, it is integrated into the working population. The working population pushes new workunits to the workunit queue as the workunit queue gets depleted. Clients can request workunits from the workunit queue and then return their result to the server to complete the cycle.

CHAPTER 5
SDSS stripe 19

5.1 Preliminary results for SDSS stripe 19

We fit all of the turnoff stars in Stripe 19, as described in Section 2.1, with a Hernquist plus double exponential disk and three streams. We choose three streams because we expect to find, the Sgr leading tidal tail, the "bifurcated" stream, and Virgo Overdensity in our data. However, if one of these streams is absent, the algorithm has the liberty to marginalize extra streams. Although we plan to try fitting the data to a larger number of streams in the future, it was not attempted here due to the long processing time required to reach optimization convergence, and because the potential need to fit additional streams was not understood before completion of this work. The parameters for the model were determined by running four full optimizations to convergence on the SDSS data, which took approximately three weeks. The optimization with the best likelihood was selected, but the parameters and likelihoods from all four runs were similar.

The parameters found are listed in Table 5.1 along with their errors, calculated using the method described in Section 3.9. In Table 5.2, we provide the covariances found for these results as the lower half of the symmetric covariance matrix. We also provide other comparisons for our results in Table 5.1 for reference. Note that we have not used any data from neighboring stripes in creating the parameters for Stripe 19; when we do the final optimizations on this data, we might find it necessary to use constraints from neighboring stripes to ensure convergence to the best possible set of parameters and to enforce the physical constraint that streams must be continuous from stripe to stripe (as in Newby et al., 2013), but that has not been attempted here.

A more user-friendly version of our results can be found in Table 5.3. The parameters given in Table 5.3 can be used to calculate the MSTO stellar density of the background or a stream at any point in the wedge. The density of the smooth component, in MSTO stars

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., Newby, M., & Desell, T. submittedb, ApJS

per cubic kpc is given by:

$$\rho_{background} = A * \rho_{smooth}(X, Y, Z) \tag{5.1}$$

where $\rho_{smooth}(X, Y, Z)$ is from Equation 3.3 and A is provided in Table 5.3. The density of a stream, in MSTO stars per cubic kpc is:

$$\rho_{stream} = Ae^{-r^2/2\sigma^2} \tag{5.2}$$

where A and σ are provided in Table 5.3, and r is the distance from the axis of the cylinder that describes the stream. To calculate r from the values in Table 5.3, use

$$r = |(\vec{r} - \vec{r_{sc}}) \times \hat{n}| \tag{5.3}$$

where $\vec{r}$ is the position at which you want the density and $\vec{r_{sc}}$ is the position of the stream center. A conversion factor is required to convert from MSTO stellar density to the stellar density of other tracers based on the ratio of expected tracers to MSTO stars.

In Figure 5.1, we show a wedge plot of the density of SDSS turnoff stars in Stripe 19 (right panel), and the separated components (left four panels). To separate the stripe into its components, we use a probabilistic separation method. In this separation method, we assume that each star in the stripe must belong to one of the four components we fit: the background or one of the three streams. Then, using the parts of the PDF from Equation 3.19 that correspond to each component, we can calculate the relative probabilities that a star belongs in each of the components. The star is then randomly assigned to a component according to the relative probabilities. Since this is a probabilistic separation, the specific stars we select for each component do not necessarily belong to the component in which they are placed. The densities of the separated stars, however, follow the density of the components. For more information on this method, see Cole et al. (2008) for the original derivation, and Newby et al. (2013) for the derivation for multiple streams.

In this figure, you can see that the three separated streams plausibly resemble Gaussian cylinders with a slight elongation in the radial direction due to the absolute magnitude distribution, and a slight elongation due to the stream's direction of travel through the wedge; this latter elongation is not generally in the radial direction. However, the "bifurcated"

stream is less comprehensible because it is very wide and covers the whole stripe. The separated "bifurcated" stream does not look like it contains any extra substructure, which would appear as substructure in the density of stars on the third panel of Figure 5.1. The background also looks smooth, and exhibits no apparent unidentified substructure. As we will see, the smooth component looks very similar to the smooth component in a simulated stripe. Figure 5.2 compares the Xu et al. (2015) model smooth component model to our model with the parameters fit for stripe 19. We see our model finds a lower fraction of thick disk stars in the wedge, which is expected since we designed our color cuts to eliminate most of the disk stars.

To visualize the expected density distribution in this wedge, given the model and fit parameters, we used the fit parameters to simulate each component in the SDSS wedge individually, and then combined them to get a full picture of the stripe. The results are shown in the second row of Figure 5.1. The model appears to accurately fit the density of the separated components in the SDSS stripe. To get a better idea of how well it fits the SDSS data, we show a residual of the original SDSS stellar density minus the simulated density in Figure 5.3. We see only minor discrepancies in areas of particularly high density.

5.2 Validation of the results using test data

5.2.1 Generating simulated turnoff stars in SDSS stripe 19

To test that both the model and fitting algorithm work as intended, the algorithm was tested on a set of data drawn from a density model with known parameters. We wrote a completely separate program to generate the test data, so that the entire modeling and fitting procedure would be tested. Generating the test data is a multi-step process in which we simulate our background stellar distribution and streams with the correct number of stars, determine the star's absolute and apparent magnitudes using our magnitude distribution, and then apply the effects of observational bias.

The first step to generating our test data in selecting the set of parameters we wish to simulate. To do this, we use the values of the parameters from our fit to SDSS Stripe 19 from Table 5.1. This is the best set of parameters to check the accuracy of the derived Stripe 19 density model.

After we choose the parameters, we input them to our wedge simulator, which begins by calculating the number of stars in each component of our data. The program does this

Table 5.1: Constraints and preliminary results from MilkyWay@home fits for SDSS stripe 19 compared to results from Belokurov et al. (2006), Hernitschek et al. (2017), Newby et al. (2013), and Newberg et al. (2007). Our results for Sagittarius are comparable with the results found in Belokurov et al. (2006), Newby et al. (2013) and in Hernitschek et al. (2017). The slight distance and position discrepancy in the result from Hernitschek et al. (2017) could be due to the measurement being taken 1.5° outside of the plane of stripe 19. We find a much farther distance to the "bifurcated" stream than was found in Newberg et al. (2007), Belokurov et al. (2006) and Hernitschek et al. (2017). The distance discrepancy in Newberg et al. (2007) can be explained by the fact that they did not consider completeness, or the affect of larger color errors near the limit of the survey.

Preliminary Stripe 19 Results from MilkyWay@home						
Background	ε_{sph}	q				
Newby et al. (2013)	—	0.52 ± 0.12				
Fit	0.9969 ± 0.0005	0.57 ± 0.02				
Constraint	0.8 to 1.0	0.25 to 1.0				
Stream	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
Sagittarius	-1.94 ± 0.09	151.7 ± 0.4	21.1 ± 0.3	2.57 ± 0.1	2.78 ± 0.2	1.0 ± 0.2
Sagittarius Constraint	-3.5 to 0.5	135 to 172	13.88 to 32.12	1.77 to 3.03	2.37 to 3.63	0.1 to 5.0
Sagittarius (Belokurov et al., 2006)	—	152.0	20 ± 2	—	—	—
Sagittarius (Hernitschek et al., 2017)	—	168.3	25.7	—	—	5.0
Sagittarius (Newby et al., 2013)	-1.9	151.0	23.0	2.4	3.0	0.9
"bifurcated" Stream	-0.98 ± 0.11	163 ± 11	48 ± 10	1.36 ± 0.13	3.41 ± 0.09	17.6 ± 2.6
"bifurcated" Stream Constraint	-2.95 to 1.95	153 to 205	22.88 to 50.0	0.00 to 3.14	2.37 to 3.63	2.0 to 25.0
"bifurcated" Stream (Belokurov et al., 2006)	—	175.8	27 ± 1	—	—	—
"bifurcated" Stream (Hernitschek et al., 2017)	—	173	17	—	—	10.1
"bifurcated" Stream (Newberg et al., 2007)	—	179	32	—	—	—
Virgo	-0.39 ± 0.08	206.9 ± 16	6 ± 2.3	0.50 ± 0.04	3.0 ± 0.06	6.1 ± 0.3
Virgo Constraint	-3.0 to 3.0	135 to 230	6.0 to 26.24	0.00 to 3.14	0.00 to 3.14	0.5 to 11.0

Table 5.2: The full normalized variance matrix for the optimization run on the SDSS stripe 19 data. The bold diagonal elements correspond to the reported errors (σ^2 for each parameter. The off diagonal elements represent the covariances for each of the parameters. We only provide the lower half of this matrix since it is symmetric. To improve legibility, we have provided boxes which group the parameters by their corresponding streams. Although most covariances are small, for some pairs of parameters the covariances are of the order of the error.

	Error Matrix for SDSS Stripe 19													
	Background		Sagittarius						"bifurcated" stream					
	ε_{sph}	q	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
ε_{sph}	**2.3e-7**	—	—	—	—	—	—	—	—	—	—	—	—	—
q	-8.2e-6	**4.8e-4**	—	—	—	—	—	—	—	—	—	—	—	—
ε	8.6e-6	-7.2e-4	**7.7e-3**	—	—	—	—	—	—	—	—	—	—	—
μ	4.0e-5	-1.0e-3	6.5e-3	**0.17**	—	—	—	—	—	—	—	—	—	—
R	-3.2e-6	6.5e-4	-3.1e-3	0.029	**0.11**	—	—	—	—	—	—	—	—	—
θ	-1.3e-5	6.2e-4	-2.4e-3	-8.0e-3	4.2e-3	**0.013**	—	—	—	—	—	—	—	—
ϕ	1.6e-5	-7.8e-4	1.7e-3	0.018	-4.9e-3	-0.014	**0.037**	—	—	—	—	—	—	—
σ	2.6e-6	-3.3e-5	4.9e-3	-3.8e-3	3.2e-3	-4.5e-3	-0.02	**0.035**	—	—	—	—	—	—
ε	1.2e-5	-1.3e-3	3.9e-3	4.4e-3	-6.2e-3	-6.0e-4	7.1e-4	1.9e-4	**0.012**	—	—	—	—	—
μ	-6.9e-4	0.019	3.0e-3	-0.25	-0.29	0.14	-0.13	-0.084	0.015	**130**	—	—	—	—
R	8.4e-4	-0.033	0.024	0.29	0.38	-0.14	0.14	0.079	0.12	-110	**110**	—	—	—
θ	-1.6e-5	1.0e-3	-3.0e-3	-7.1e-3	-1.5e-3	1.9e-3	-1.8e-3	-1.5e-3	-9.6e-3	0.64	-0.82	**0.017**	—	—
ϕ	6.4e-6	-2.4e-4	4.9e-4	-1.0e-3	3.8e-3	-1.1e-3	1.1e-3	8.9e-4	-2.0e-3	-0.72	0.52	-6.4e-4	**7.9e-3**	—
σ	9.7e-5	-9.5e-3	-0.022	0.21	0.1	0.027	-0.028	-0.037	0.11	-5.9	11	-0.18	-0.038	**6.6**
ε	-1.3e-6	-6.6e-4	4.1e-4	-8.9e-3	3.5e-4	1.2e-4	8.8e-4	-2.2e-3	1.3e-3	0.16	-0.15	2.6e-3	6.2e-4	-0.012
μ	-9.0e-4	-0.058	0.48	-0.1	-0.38	-0.067	0.25	-0.11	0.59	73	-69	0.27	-0.26	-4.1
R	3.7e-4	-6.8e-3	-0.037	9.6e-3	-0.025	-0.014	-0.012	0.025	-0.054	-9.6	9.5	-0.048	0.035	0.71
θ	7.3e-6	-4.4e-4	-2.6e-4	-3.3e-3	-2.3e-3	-5.2e-4	5.7e-4	-3.8e-4	3.0e-4	-0.039	0.063	-3.3e-4	1.0e-4	5.3e-3
ϕ	4.9e-6	-4.4e-4	1.9e-3	-9.3e-5	-3.5e-3	-6.3e-4	5.7e-4	4.6e-4	2.2e-3	0.19	-0.15	-3.1e-4	-5.6e-4	0.011
σ	-3.3e-5	1.2e-3	-0.014	-0.024	0.014	2.1e-3	3.8e-3	-0.013	-0.016	0.26	-0.44	0.02	2.8e-3	-0.1

	Virgo					
	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
ε_{sph}	—	—	—	—	—	—
ε	—	—	—	—	—	—
μ	—	—	—	—	—	—
R	—	—	—	—	—	—
θ	—	—	—	—	—	—
ϕ	—	—	—	—	—	—
σ	—	—	—	—	—	—
ε	—	—	—	—	—	—
μ	—	—	—	—	—	—
R	—	—	—	—	—	—
θ	—	—	—	—	—	—
ϕ	—	—	—	—	—	—
σ	—	—	—	—	—	—
ε	**7.2e-3**	—	—	—	—	—
μ	0.81	**280**	—	—	—	—
R	-0.09	-36	**5.4**	—	—	—
θ	7.7e-4	-0.19	0.05	**1.3e-3**	—	—
ϕ	1.3e-3	0.71	-0.073	-1.1e-4	**3.2e-3**	—
σ	0.014	0.53	-0.12	1.2e-3	-4.3e-3	**0.087**

Table 5.3: Easy-to-use description of the density of MSTO stars in the background and streams, the preliminary results in Table 5.1. This table includes the number of stars in the wedge, converted from the weights; the stream center's Galactic l, b, and r, converted from μ, ν and r; the stream's Gaussian normalization factor A and width σ; and the stream's unit direction in Galactic (X, Y, Z). With the numbers provided in this table it is possible to find the background and stream MSTO star densities at any point the wedge using Equations 5.1 and 5.2.

Stripe 19							
Background	Stars	A	q	r_o (kpc)	ε_{sph}		
Background	38271	245099266	0.57	12.0	0.9969		
Stream	Stars	l (deg)	b (deg)	r (kpc)	A	σ (kpc)	$\hat{n}$
Sagittarius	5500	215.5	49.2	21.1	1143.4	1.00	$(-0.51, 0.19, -0.84)$
"bifurcated" Stream	14363	218.6	60.8	48.2	58.4	17.6	$(-0.94, -0.26, 0.21)$
"Virgo"	25911	12.4	74.4	6.05	360.9	6.12	$(-0.48, 0.06, 0.88)$

using Equation 3.10 with the specified weights to get the fraction of stars in each component. The fraction of stars is then multiplied by the number of MSTO stars we want to simulate (in this case 84,046, which is the number of turnoff stars observed in Stripe 19). Each component is then simulated with the required number of stars.

For each component of the density, the test data generator creates one star at a time until the number required by the weights is reached. The smooth component is generated using rejection sampling. We uniformly sample Equation 3.3 over the volume of the wedge, and we compare the density from Equation 3.3 to the maximum density possible in the wedge from Equation 3.3. Then, using a random number, we determine whether we will reject or keep the star, based on the fraction of the calculated density to the maximum density.

Stream stars are generated using active generation to create a stellar position in three dimensions. One number drawn from a uniform distribution determines the position along the cylinder's z-axis, and two numbers drawn from a normal distribution determine the position of the star in the plane perpendicular to the cylinder's z-axis. The length of the cylindrical distribution that is generated is significantly longer than the portion of the cylinder axis that is within the wedge of data in Stripe 19. The cylinder's length is long enough that all of the volume in the wedge that is within three σ of the cylinder's z-axis will be populated. The stream star is then rotated and translated into Galactic X, Y, Z based on the stream's orientation and center point.

As the program generates each star, it applies the observational biases that are corrected for in our maximum likelihood model. First, it determines the star's absolute mag-

nitude based on the absolute magnitude distribution in Equation 3.13 and then uses that to determine its apparent magnitude. At this point it checks to make sure the star lies within the wedge. Finally, it samples the combined magnitude limit from Equation 3.14 and the color selection efficiency from Equation 3.15 to determine if the star would have been observed.

5.2.2 Results from fitting test data

We generated two different sets of test data: one with a Hernquist background that exactly matches the model to which the data is fit, and one with a broken power law background. The second simulation tests the sensitivity of our measurements of the stream parameters to imperfect knowledge of the smooth component of the Milky Way spheroid. For the broken power law background, we used the parameters fit in Akhter et al. (2012): an inner halo exponent of -2.78, outer halo exponent of -5.0 and break radius of 45 kpc.

MilkyWay@home was used to find the optimum values of the parameters, fit separately for the two sets of generated test data. We ran 4 optimizations for each of the simulated data wedges, each with minimal constraints. Each set of 4 optimizations independently converged to the same results; the results are listed in Table 5.4.

Next, we determined the uncertainty in each parameter using the method outlined in Section 3.9. Since our errors are given as one standard deviation, we can expect 68% (13-14 parameters) of the results to be within uncertainty, 95% (19 parameters) within two times the uncertainty, and the rest to be within three times the uncertainty most of the time. Our results in Table 5.4 show that for the Hernquist plus disk background simulation, we had 10 parameters within 1 times the uncertainties, 16 within 2 times the uncertainties, and 18 were within 3 times the uncertainties. The result is within normal statistical variation.

In our results for the simulation with the broken power law model, ignoring the 2 background parameters, we find that 5 parameters are within 1 times the uncertainties, 9 are within 2 times the uncertainties and 10 are within 3 times the uncertainties. This shows that if our model is not a good match to the actual stellar density model in the halo, our errors, as calculated using the method in Section 3.9, could be underestimated by a factor of three or more. The fraction of stars in each component is particularly poorly measured compared to the uncertainties; the optimizations for ε differ from the simulated values by 5 to 18σ. These errors look worse than they really are because they are measured with respect

to a different smooth component. We include a column in Table 5.4 that shows the number of stars in the stripe associated with each component. The density of stars along a stream could be off by as much as 40 percent.

Nevertheless, the properties of the detected streams and the general shape of the stellar spheroid (as measured by q) are generally similar to the simulation. All of the halo substructure is immediately recognizable. The flattening of the spheroid is surprisingly accurate. The widths of the streams are approximately correct, the stream centers and distances are pretty close, and even the angles at which the steams pass through the wedge are easily matched with the correct stream.

In addition to checking the accuracy of the uncertainties, we checked the accuracy of the optimizer in finding the highest point in the likelihood surface. To do this, we ran one dimensional parameter sweeps for each of the 20 parameters in our model. The sweeps seen in Figure 5.4 show the likelihood surface for the test data with the Hernquist background. This figure shows that each model parameter peaks either at, or close to, the expected parameter value. Any small difference between the peak in the likelihood surface and the simulated values can be accounted for by the uncertainty introduced by the finite number of stars available in our data. For the broken power law parameter sweeps in Figure 5.5, there are several parameters which have extremely flat likelihood surfaces. In these cases, the optimizer still seems to find an answer close to the peak of the likelihood.

In order to visually compare the results of our simulation to the actual SDSS Stripe 19 data (Figure 5.1), we present the simulated data with a Hernquist background in Figure 5.6, and the simulated data with the broken power law background in Figure 5.7. The simulated data figures, Figure 5.6 and Figure 5.7, show three different views of the data. The first row shows the components generated by the test data simulator, and the combination of those components. The second row shows a probabilistic separation of the components by our model, given the simulation parameters. Finally, the last row shows a re-simulation of the data using the fit model parameters. By re-simulating the stripe, we can visualize the model that the optimizer found as the best fit to the data.

In each of these plots the separated streams have the appearance of streams, and the smooth component appropriately contains no substructure. Especially interesting is the ability for MilkyWay@home to determine, on its own, a set of parameters that resulted in streams that look, by eye, the same as those with the "correct" parameters for the wedge.

More evidence of this can be seen in Figures 5.8 and 5.9, which show residual densities of the simulated data and re-simulated fit data from Figures 5.6 and 5.7. The broken power law fits show that we will not easily know whether or not the real data is a poor fit to a Hernquist background. However, we will still be able to fit most stream parameters (10 of the 18 stream parameters were within 2 sigma of the simulated values), even though a few might not be correct within the calculated errors.

In summary, simulations show that in the case that our background and stream models are close to the correct answer, the resultant parameter fits can be trusted. If our background or stream models are far from the correct distribution of stars the uncertainties are underestimated, but the stream properties are generally reasonable.

Table 5.4: Results and constraints of the MilkyWay@home fits for simulated data wedges with Hernquist (Fit) and Broken Power Law (BPL) backgrounds. Here we can see that most of the simulated parameters lie within our one sigma errors from our fits, when the simulated density has the same form as the model to which it is fit (Fit). The parameter values for the BPL fits are generally similar to the simulated values, except some of the uncertainties seem to be underestimated using our uncertainty estimation method.

Simulated and Fit Values from MilkyWay@home							
Background	N_{stars}	ε_{sph}	q				
Simulated Value	*38,270*	*0.997*	*0.565*				
Fit	$37,752 \pm 1,586$	0.997 ± 0.0004	$0.54 \pm .02$				
BPL	$29,431 \pm 853$	1.0 ± 0.0003	$0.58 \pm .02$				
Constraint	—	0.8 to 1.0	0.25 to 1.0				
Stream	N_{stars}	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
Sagittarius (Sim Value)	*5,500*	*-1.94*	*151.7*	*21.1*	*2.57*	*2.78*	*1.0*
Sagittarius (Fit)	$5,877 \pm 759$	-1.86 ± 0.08	152.0 ± 0.4	20.7 ± 0.3	2.43 ± 0.10	2.80 ± 0.14	1.2 ± 0.2
Sagittarius (BPL)	$8,776 \pm 711$	-1.21 ± 0.07	151.7 ± 0.4	21.5 ± 0.3	2.47 ± 0.10	2.43 ± 0.22	1.9 ± 0.4
Sagittarius Constraint	—	-3.5 to 0.5	135 to 172	13.88 to 32.12	1.77 to 3.03	2.37 to 3.63	0.1 to 5.0
"bifurcated" Stream (Sim Value)	*14,363*	*-0.98*	*163.2*	*48.2*	*1.36*	*3.41*	*17.6*
"bifurcated" Stream (Fit)	$18,193 \pm 1,641$	-0.73 ± 0.1	180.0 ± 5	36 ± 3	1.31 ± 0.12	3.27 ± 0.10	16.1 ± 2.1
"bifurcated" Stream (BPL)	$18,395 \pm 1,154$	-0.47 ± 0.07	204.6 ± 13	32 ± 0.8	1.31 ± 0.06	3.63 ± 0.22	15.6 ± 1.8
"bifurcated" Stream Constraint	—	-2.95 to 1.95	153 to 205	22.88 to 50.0	0.00 to 3.14	2.37 to 3.63	2.0 to 25.0
Virgo (Sim Value)	*25,911*	*-0.39*	*206.9*	*6.05*	*0.50*	*3.01*	*6.12*
Virgo (Fit)	$22,221 \pm 1,641$	-0.53 ± 0.08	214.9 ± 9	6.1 ± 0.7	0.51 ± 0.03	3.13 ± 0.04	5.5 ± 0.3
Virgo (BPL)	$27,441 \pm 1,295$	-0.07 ± 0.06	194.1 ± 4	6 ± 1	0.42 ± 0.02	3.11 ± 0.03	4.9 ± 0.2
Virgo Constraint	—	-3.0 to 3.0	135 to 230	6.0 to 26.24	0.00 to 3.14	0.00 to 3.14	0.5 to 11.0

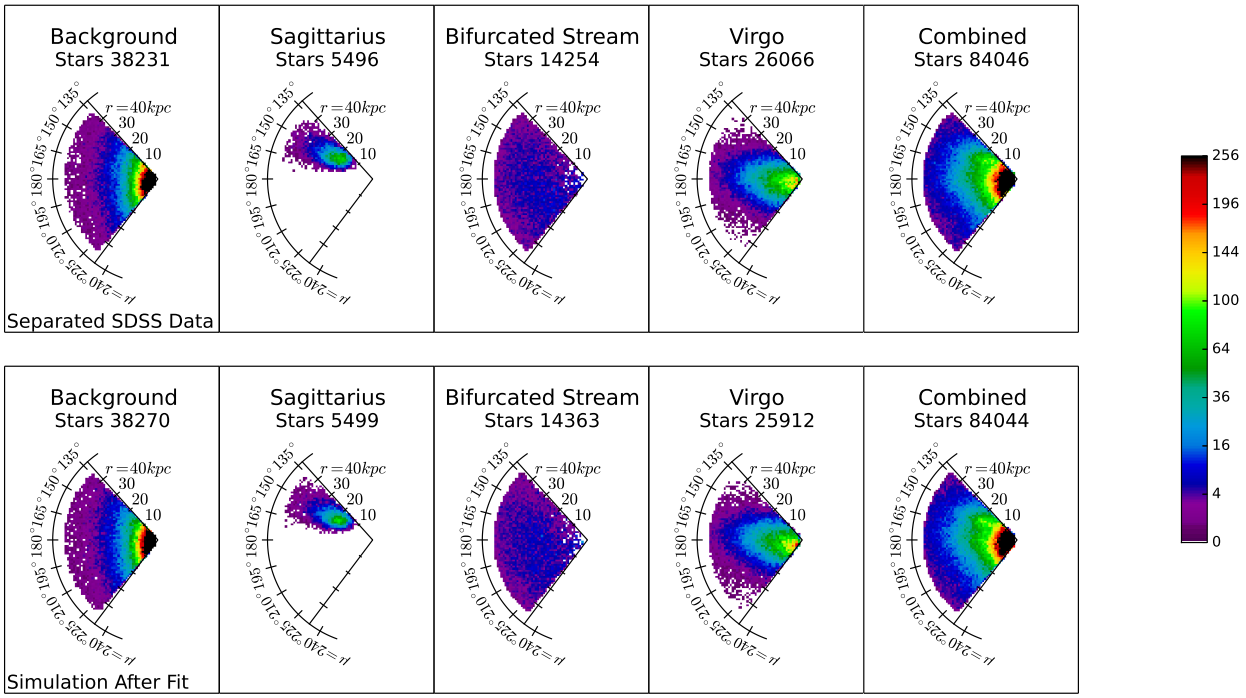


Figure 5.1: MSTO star stellar density in stars per cubic kpc in a 1 kpc by 1 kpc by 2.5 degree pixel at each point in a flattened, face on view of SDSS stripe 19 after they were fit by MilkyWay@home. The first four panels in the top row show a probabilistic separation of the SDSS stars into each different substructure fit. The last panel includes all of the stars in the original stripe of data before separation. The plots show a smoothly separated background, and three separate streams with seemingly no substructure left over in the wedge. The lower row shows the density distribution of stars in a simulated stripe, that was created with the same parameters that were fit to the actual data. Note the similarity between the simulation with three streams and a smooth background to the actual SDSS data.

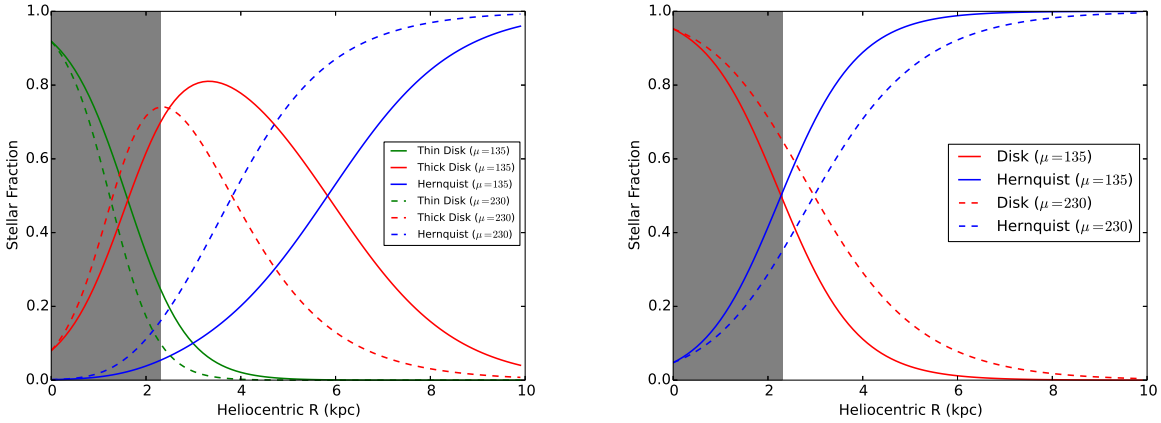


Figure 5.2: Fraction of stars from the stellar halo, the thin disk and the thick disk as a function of heliocentric distance in the wedge. The first panel shows the Xu et al. (2015) model and the panel on the right is our model after fitting stripe 19. The grey portion of the plot is outside of our data and everything to the right of it is inside of our data selection area. The green lines represent the thin disk which while present in the Xu et al. (2015) model is not present in ours since it is negligible (and not fit) in the wedge. The blue line is the Hernquist background which should dominate in our data, and the red line is the thick disk fraction which while very present in the Xu et al. (2015) model is reduced in our model since most of the disk stars are eliminated with our color cuts.

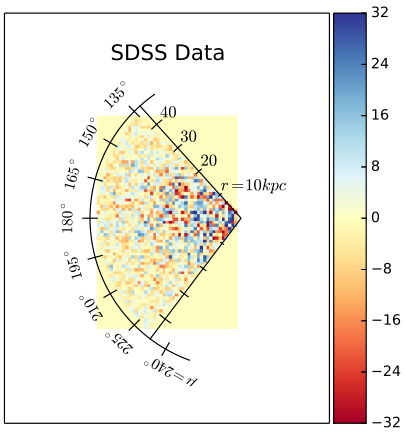


Figure 5.3: The residual between the stellar density of SDSS stars in stripe 19 and a simulation created with the fit parameters from MilkyWay@home. The color represents the residual density of MSTO stars per cubic kpc in a 1 kpc by 1 kpc by 2.5° pixel in the flattened (all stars are added together in the ν direction, so the volume of the stripe increases with radius from the Sun), face-on stripe. This residual is found by subtracting the "Simulation After Fit" data from the SDSS data from Figure 5.1. The residual shows that the model fits the data well, though there remains some structure in the highest density portion of the residual.

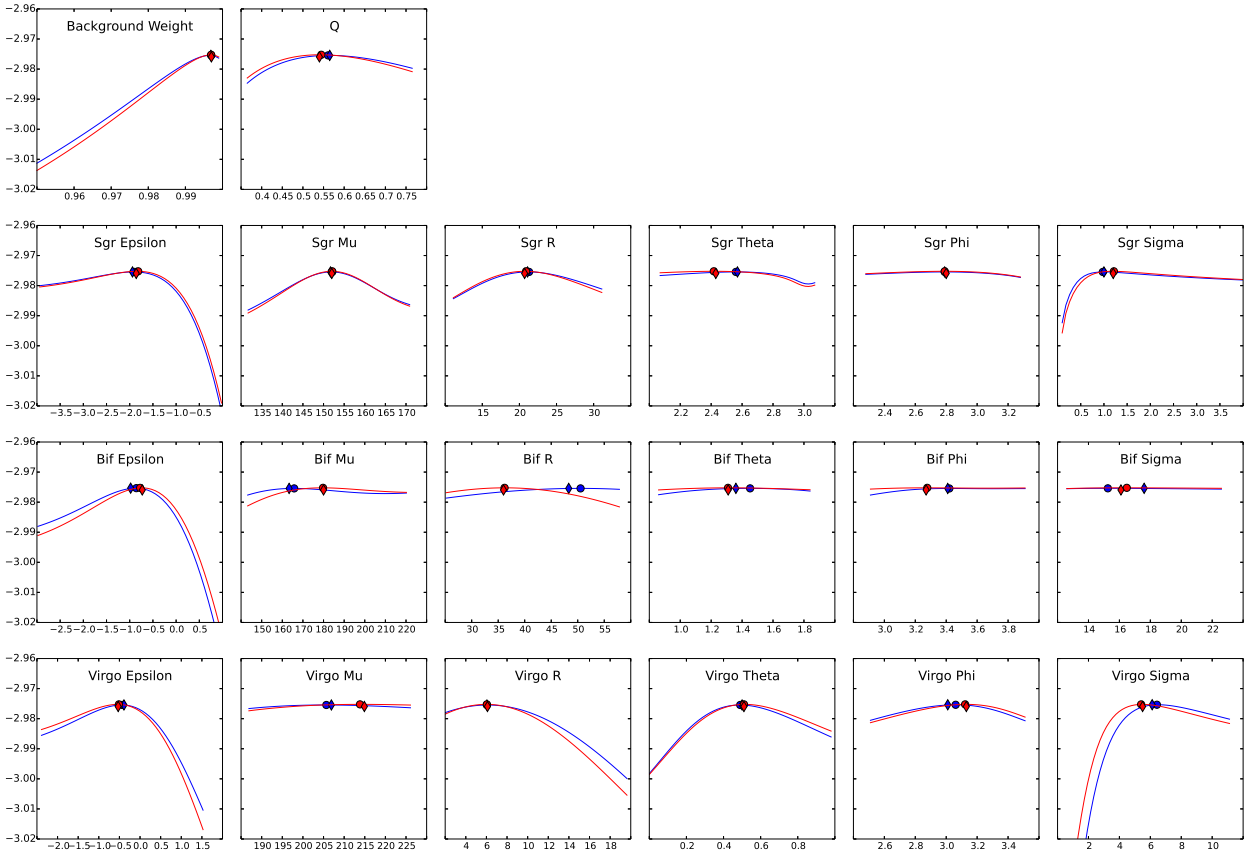


Figure 5.4: Parameter sweeps in each of the 20 parameters for the simulated SDSS stripe 19 with the Hernquist plus thick disk background. A parameter sweep gives the $log(\mathcal{L})$ as a function of one of the variables, holding the other variables constant. The blue line represents a parameter sweep around the simulated parameters from Table 5.4, and the red line represents the parameter sweep around the best results returned from MilkyWay@home, listed as "Fit" in Table 5.4. The simulated value for the parameter is shown by the blue diamond and the best likelihood in the parameter sweep around the simulated parameters is shown by the blue circle. The red diamond indicates the value of the parameter returned by MilkyWay@home and the red circle represents the best likelihood found in the parameter sweep around the result returned from MilkyWay@home. In the first row, we show parameter sweeps for the smooth background parameters. The next three rows show the parameters for, the Sagittarius stream (Sgr), the "bifurcated" stream (Bif), and Virgo, respectively. The best results will be produced if the parameter sweeps show a narrowly peaked likelihood surface. Most of the panels show well-behaved slices through the likelihood surface, which is good. All of the sweeps show that both the returned result from MilkyWay@home and the parameter sweep have the best likelihood in the same place. This means our optimizer successfully converged to a maximum in our likelihood surface.

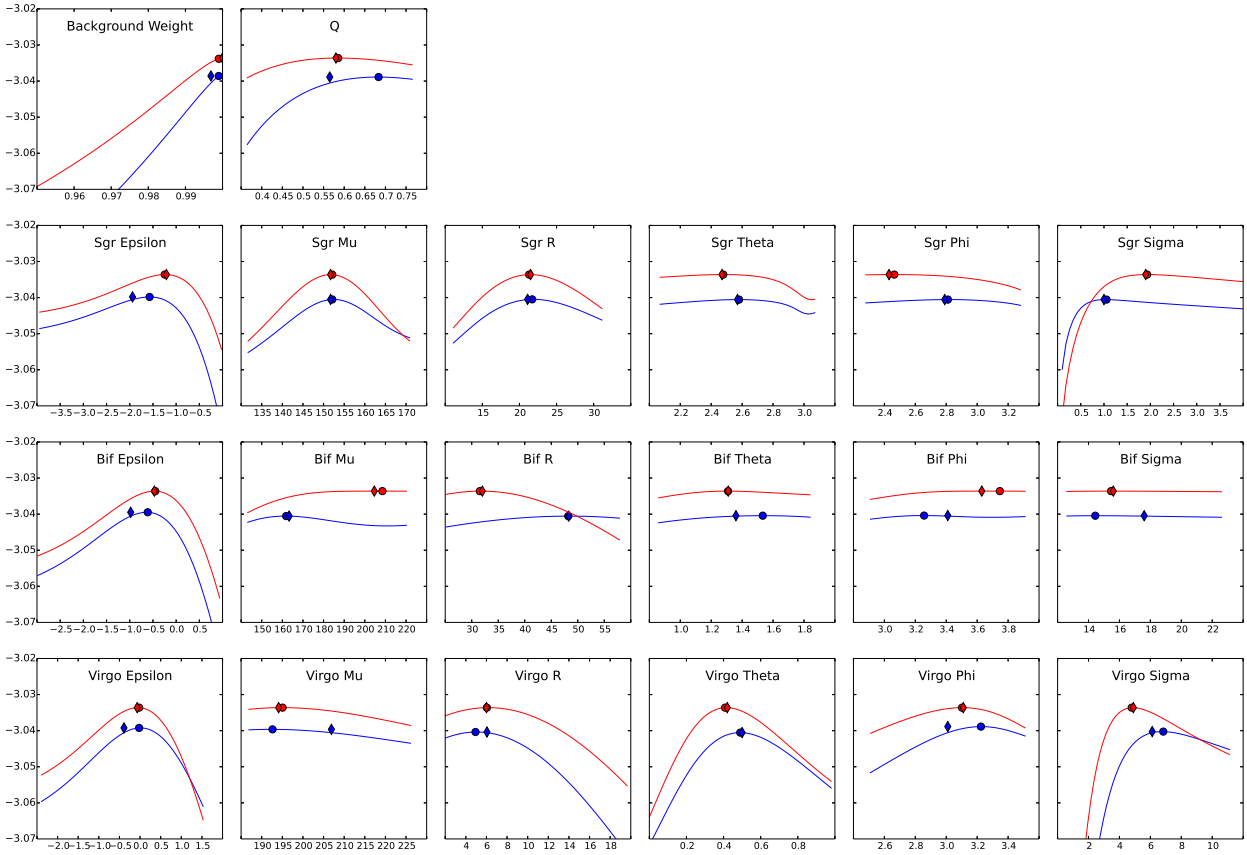


Figure 5.5: Parameter sweeps in each of the 20 parameters for the simulated SDSS stripe 19 with the broken power law background. The blue line represents a parameter sweep around the simulated parameters from Table 5.4, and the red line represents the parameter sweep around the best results returned from MilkyWay@home, listed as "BPL" in Table 5.4. The simulated value for the parameter is shown by the blue diamond and the best likelihood in the parameter sweep around the simulated parameters is shown by the blue circle. The red diamond indicates the value of the parameter returned by MilkyWay@home and the red circle represents the best likelihood found in the parameter sweep around the result returned from MilkyWay@home. In the first row, we show parameter sweeps for the smooth background parameters. The next three rows show the parameters for: the Sagittarius stream (Sgr), the "bifurcated" stream (Bif), and Virgo, respectively. Again, most of the panels show well-behaved slices through the likelihood surface. In this set of parameter sweeps, we also note the likelihood surface around several parameters in the Bif row are very flat and the optimizer was still capable of finding the peaks.

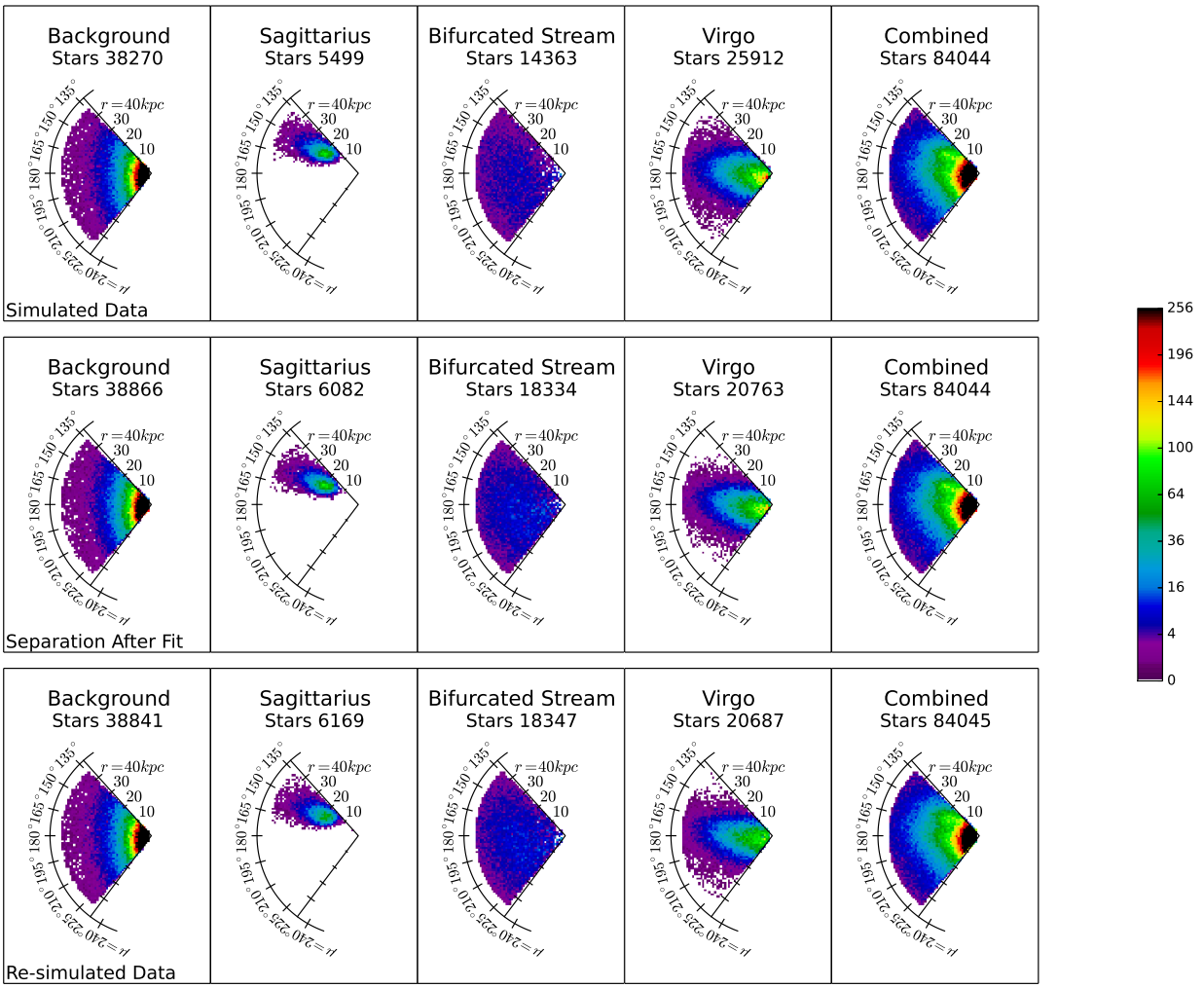


Figure 5.6: Stellar density from our simulated stripe 19 plotted in three different ways. In this figure, the color represents the density of MSTO stars per cubic kpc in a 1 kpc by 1 kpc by 2.5 degree pixel in the flattened, (all stars are added together in the ν direction, so the volume of the stripe increases with radius from the Sun) face-on stripe. Values of μ and r in each stripe are indicated. In all three rows, the last panel shows the sum of the components in the first four panels. We show the smooth component of the spheroid (background), the Sagittarius stream, the "bifurcated" stream, and the Virgo Overdensity. In the first row, the first four panels are the model components that are added together to make the simulated stripes. In the second row, the first four panels are a probabilistic separation of the stars into each of the model components using the "correct" parameters used to simulate the wedge. Finally, in the last row, we re-simulate the wedge using the parameters recovered from the MilkyWay@home results and then separate the re-simulated wedge into its model components. Using this method, we can visualize the model that the optimizer thinks is the best fit to the data. All three methods yield similar results for each stream, which is as expected if the separation algorithm is successful.

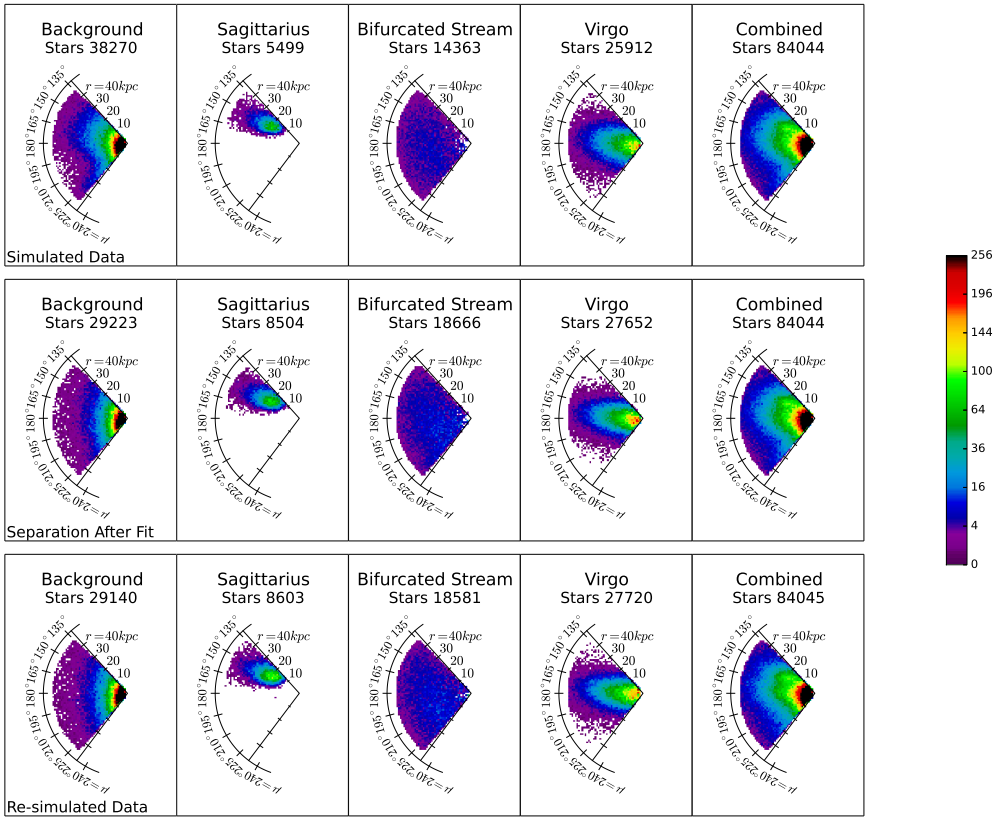


Figure 5.7: Stellar density from our simulated stripe 19 with a broken power law background. This figure is similar to Figure 5.6, except using a different simulated data set. In the first row, the first four panels are the model components, including the broken power law background, that make up the simulated stripes. In the second row, the first four panels are a probabilistic separation of the stars into each of the model components using the "correct" parameters used to simulate the wedge. This separation assumes a Hernquist background with parameters like those from our other simulated wedge. Finally, in the last row, we re-simulate the stripe using the parameters recovered by MilkyWay@home with a Hernquist background. The first four panels are the four components and the last panel is a combination of all of the stars. By re-simulating the stripe using the recovered model parameters, we can visualize the model that the optimization found as the best fit to the data. Note that MilkyWay@home is not able to faithfully reproduce the simulated background if the model that is being fit is not the same as actual stellar halo density distribution. This is evident from the fact that the shape of the smooth component in the upper left panel shows significant extra structure at $\mu = 225°$ and much less at $\mu = 210°$. Since the Hernquist profile cannot reproduce this shape, the separation in the lower left splits the difference between these two densities; the streams are shaped slightly differently to absorb or contribute stars as necessary. Note that overall, the stream densities are remarkably robust.

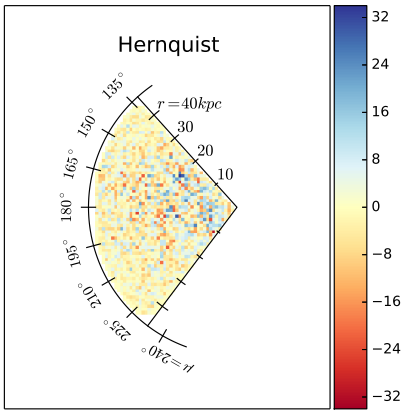


Figure 5.8: Residual density between simulated stripe 19 and the re-simulated stripe 19. The color represents the residual density of MSTO stars per cubic kpc in a 1 kpc by 1 kpc by 2.5° pixel in the flattened (all stars are added together in the ν direction, so the volume of the stripe increases with radius from the Sun), face-on stripe. This residual is found by subtracting the combined model in Row 3 of Figure 5.6 from the combined model in Row 1 of Figure 5.6. The residual shows that these models are very similar.

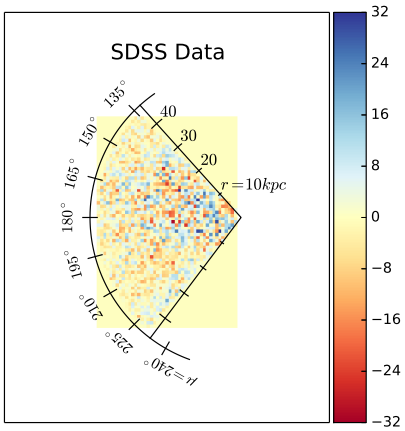


Figure 5.9: Residual density between simulated stripe 19 with BPL background and the re-simulated stripe 19 with a BPL background. The color represents the residual density of MSTO stars per cubic kpc in a 1 kpc by 1 kpc by 2.5° pixel in the flattened (all stars are added together in the ν direction, so the volume of the stripe increases with radius from the Sun), face-on stripe. This residual is found by subtracting the combined model in Row 3 of Figure 5.7 from the combined model in Row 1 of Figure 5.7. The residual shows these models mostly differ in the background. You can see the section of the plot which is between 25 kpc and 45 kpc and 180 degrees and 225 degrees which is dominated by red. This section is red dominated due to the corresponding region in the BPL model which has a large deficit of stars.

CHAPTER 6
Streams in the northern Galactic cap

6.1 Overview

We ran a single optimization on each stripe from 10 through 23 in the SDSS north Galactic cap and found stream centers consistent with 7 different streams. These runs were highly unconstrained, using the constraints found in Table 6.1 for the Halo and Table 6.2 for the streams. These constraints were chosen to prevent biasing the results of the optimizations and to allow the optimization algorithm the greatest latitude in fitting the streams in the wedge. The streams we find are: the Sgr leading tidal tail, the "bifurcated" stream, the Parallel Stream suggested in Sohn et al. (2016), the Sagittarius trailing tail in the north Galactic cap at 100 kpc (Belokurov et al., 2014), a new stream which we will refer to as the Perpendicular Stream which may be associated with Virgo and passes close to the globular cluster NGC 4147, a stream which may contain the globular cluster NGC 5466 (Belokurov et al., 2006; Grillmair & Johnson, 2006; Fellhauer et al., 2007), and one stream center at the position and distance of Stream C from Belokurov et al. (2006).

Due to the chosen constraints, there is a degeneracy in the stream fitting order for each stripe. As a result, the streams fit in each stripe can be assigned to any of the three fit streams. We report the results for each stream in each stripe as they were fit in Table 6.3 and we use that table as a the basis for our later stream analysis. Easier-to-use versions of these parameters are provided in Table 6.4. These parameters, in conjunction with Equation 5.2, can be used to determine the MSTO stellar density at any point in a stripe. The results from Table 6.3 are plotted in a right ascension (RA) and declination (dec) plot in Figure 6.1 and an RA and distance plot in Figure 6.2. Globular clusters from table 2.2 have been added to these figures for reference.

After fitting the whole sky and plotting all of the results in Figures 6.1 and 6.2, we attempted to determine which stream centers were likely to belong to which streams. This was done by cross-correlating our RA and dec plots, our RA and distance plots, and previous stream results. We found 7 different stream candidates which we separated into three tables:

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., & Desell, T. submitteda, ApJ. https://arxiv.org/abs/1807.03754

Tables 6.5, 6.7 and 6.6. These tables report the candidate stream centers for Sgr, the "bifurcated" stream, and other substructure, respectively. We plotted these results in different colors on Figures 6.3 and 6.4 to show where these stream centers can be found on the sky and in distance. One stream of particular note is the stream we refer to as the Perpendicular Stream. This stream runs roughly perpendicular to both Sgr and the "bifurcated" stream at the distance of Virgo and may be a new stream.

Currently, periodic angle parameters θ and ϕ are not allowed to wrap from 3.14 to 0 or from 0 to 3.14 making it possible for optimizations to get stuck on a boundary of these parameters even though the best value lies within the constraints. If the best value for one of these parameters is on the opposite side of the periodic boundary from the current search neighborhood, the optimizer will continue approaching the boundary causing it to get stuck. Any stripe with streams stuck in this way should be rerun at a later date allowing these parameters to have periodic boundary conditions. This will help the optimizations find the best possible set of parameters within the parameter space.

To determine what the density model found by MilkyWay@home looks like, we simulated the entire sky using our test wedge generator, which was developed for testing our model with stripe 19 data in Section 5.2.1. The result of these simulations can be found in Figures 6.5 and 6.6. In Figure 6.5, we show the stellar density of the SDSS data in the top panel compared to the stellar density of the simulated sky from our model in the bottom panel. Our density model reproduces most of the visible halo structure in the data, but is missing some substructure such as the bifurcated stream, especially in stripes above stripe 17.

We also provide wedge plots of the SDSS data, the simulated SDSS data from our model, and a residual between these two data sets in Figure 6.6. The residual plot is found by subtracting the simulated SDSS data from the observed SDSS data. Each residual is on a different scale to better reveal the substructure. Some stripes, such as stripe 20, have a large residual and monochromatic patching. This patchiness suggest that there are areas which have unfit substructure in the data. In the future, we will attempt to fit the remaining substructure by adding additional streams to our model.

Table 6.1: The constraints for fitting the halo in the SDSS stripes 10 through 23 in the north Galactic cap. These constraints were chosen to allow the optimizations to decide the shape of the halo and disk fraction while avoiding unphysical or undefined parameter values.

Halo Constraints for MilkyWay@home		
Stripe	ε_{sph}	q
10	0.8 to 1.0	0.2 to 1.75
11	0.8 to 1.0	0.2 to 1.75
12	0.8 to 1.0	0.2 to 1.75
13	0.8 to 1.0	0.2 to 1.75
14	0.8 to 1.0	0.2 to 1.75
15	0.8 to 1.0	0.2 to 1.75
16	0.8 to 1.0	0.2 to 1.75
17	0.8 to 1.0	0.2 to 1.75
18	0.8 to 1.0	0.2 to 1.75
19	0.8 to 1.0	0.2 to 1.75
20	0.8 to 1.0	0.2 to 1.75
21	0.8 to 1.0	0.2 to 1.75
22	0.8 to 1.0	0.2 to 1.75
23	0.8 to 1.0	0.2 to 1.75

6.2 Halo shape and disk fraction

In stripes 10 through 23, we find similar results for all of the disk weights and the halo shapes. These results can be found in Table 6.8. We also provide easier-to-use parameters for our halo model in Table 6.9. These parameters can be used with Equation 5.1 to derive the MSTO density in the background component at any point in the associated stripe.

In our results, we find a oblate halo with an average flattening of 0.58 with a dispersion in the measurements of 0.04. This results is consistent with results previously reported by Newby et al. (2013) even though our background model is slightly different than theirs since we fix the Hernquist scale length to 12 kpc. The dispersion in measurements is somewhat larger than the statistical error in each measurement, which suggests systematic error due to an imperfect density model.

6.3 Sagittarius leading tidal tail

Sgr leading tidal tail candidates were only fit in 8 of the 14 stripes on which our optimizations were run. Since we allowed the optimizations to fit the streams in any order,

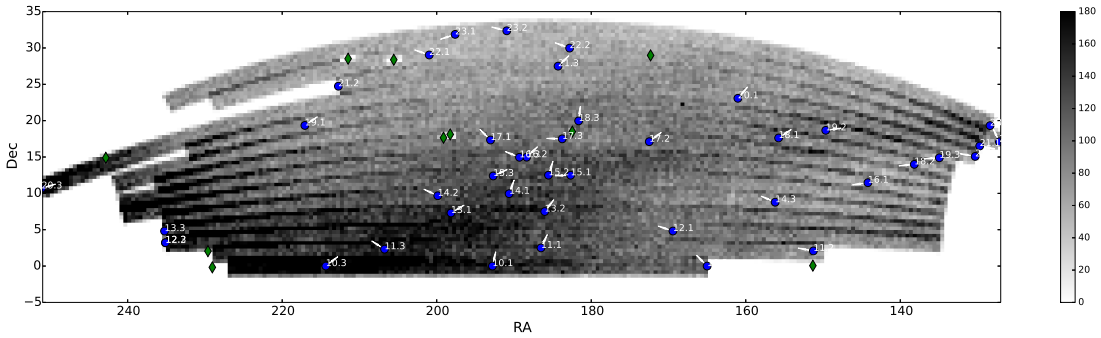


Figure 6.1: RA and dec of the stream centers from Table 6.3. These stream centers are overplotted on selected SDSS turnoff stars in the north Galactic cap for reference. Each pixel is $0.5°$ by $0.5°$. There is some overlap between the stripes, which appears in this figure as dark streaks on the borders of the stripes. The lowest latitude four stripes have a fainter magnitude limit, which explains why they appear darker. In this plot, we can see the major substructure in our data (including the Sagittarius leading tidal tail, the "bifurcated" stream, and the Virgo Overdensity) with relation to the stripe we are fitting. The blue points correspond to centers from our stream fits. There is a large cluster of points at (RA, dec) of (190, 12) near the Sagittarius stream and corresponds to the overlap of Sagittarius, the "bifurcated" stream and Virgo.

we selected the stream fits corresponding to the Sgr leading tidal tail based on sky positions and distances compared to values for the Sgr leading tidal tail found in the literature. These candidates can be found in Table 6.5. The stripe column in this table corresponds to a value on Figures 6.1 and 6.2.

Our Sgr leading tidal tail results are consistent with previous values of the distance, sky position, and width from Belokurov et al. (2006), Newberg et al. (2007), Newby et al. (2013), Belokurov et al. (2014), and Hernitschek et al. (2017). We find an (RA, dec, distance) of $(218°, -2.1°, 41.7$ kpc) to $(128°, 19.3°, 14.2$ kpc). We show that our results are consistent with the Belokurov et al. (2006) distances in Figure 6.2 where we indicate our stream centers as blue circles and the Belokurov et al. (2006) results as red stars. Additionally, the RA and dec positions of the stream centers are consistent with the overdensity associated with Sagittarius in 6.1.

The model seems to struggle in stripes 14 through 17 where several streams all converge on the sky. Since we are missing Sgr leading tidal tail centers in stripes 12, 14, 15, 16, 17, and 20, and we find find other structures in most of these stripes, we are likely not fitting enough streams to our data. Specifically, our model finds two streams at approximately 15

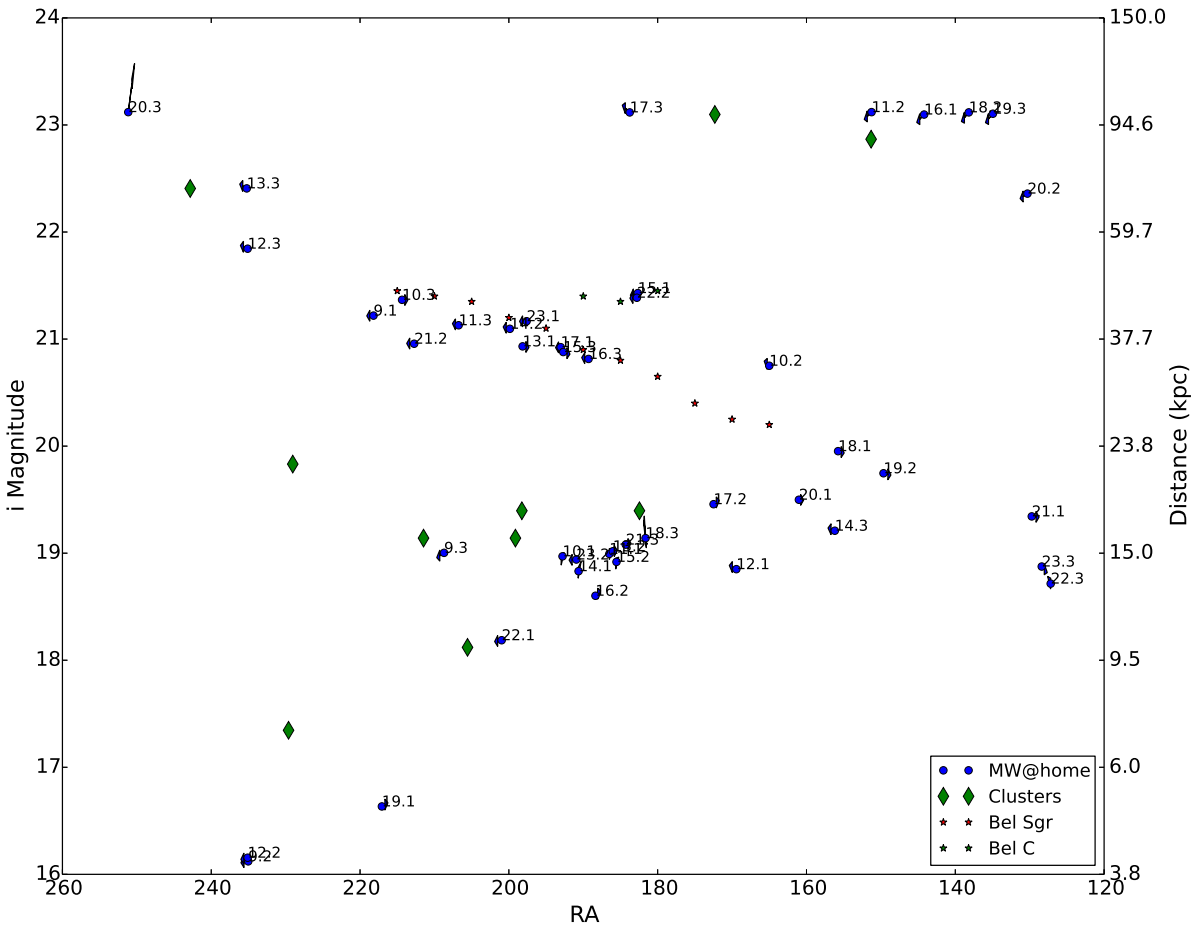


Figure 6.2: Results from MilkyWay@home in RA and distance. MilkyWay@home results are blue circles, the Sgr leading tail from Belokurov et al. (2006) are red stars, and the Belokurov et al. (2006) Stream C results as green stars. We plot the effective i band magnitude against the right ascension in degrees for each stream center fit by MilkyWay@home assuming an average absolute magnitude of 3.1. The number before the decimal point next to each point indicates the stripe to which the stream center belongs, and the number after the decimal point represents the stream number in Tables 6.3 and 6.4. This figure shows that MilkyWay@home struggled to fit Sagittarius between RA of 190° and RA of 160°. The cluster of stream centers at 19^{th} magnitude (or 15 kpc) and RA of 190° are likely making it difficult to fit Sagittarius in that region.

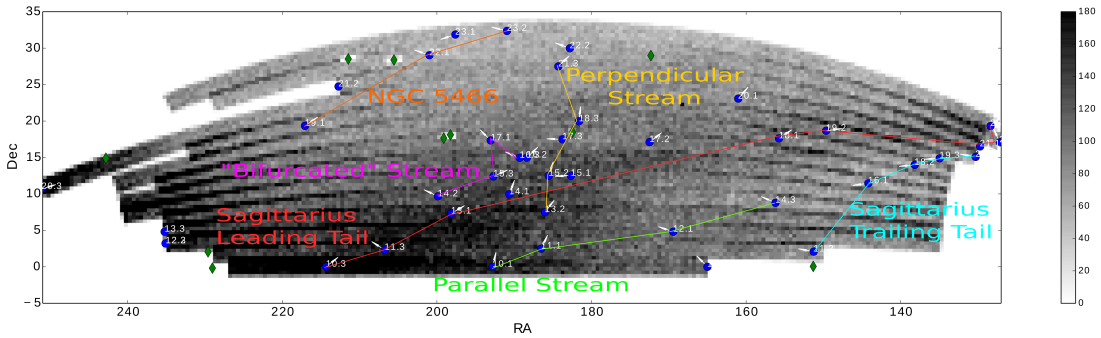


Figure 6.3: Data from Figure 6.1 along with highlighted stream candidates from Tables 6.5, 6.7, and 6.6. The lines are highlighted in the following way: Sgr leading tidal tail (red), "bifurcated" stream (magenta), NGC 5466 stream (orange), Perpendicular Stream (gold), 100 kpc stream (cyan), and Parallel Stream (green). These colors correspond to those used in Figure 6.4.

kpc in that section of the sky which could be be associated with Virgo. These streams are listed as the Perpendicular Stream and the Parallel Stream in Table 6.6.

6.4 "Bifurcated" stream

The fainter, "bifurcated" stream was not fit well by MilkyWay@home. Candidates for this stream were found in 5 of the 14 stripes and can be found in Table 6.7. The major reason these stream centers were determined to be "bifurcated" stream candidates instead of the Sgr leading tidal tail candidates is their sky positions. These stream centers are between the Sgr leading tidal tail and the "bifurcated" stream which distinguishes them from the stream centers we chose for the Sgr leading tidal tail. The Sgr leading tidal tail centers were very close to the stream center of Sgr visually.

"Bifurcated" stream candidates were fit from stripe 14 to stripe 17, the stripes where Sagittarius was not fit. The results are consistent in distance with those found in Belokurov et al. (2006), Newberg et al. (2007), and Belokurov et al. (2014), but their sky positions are inconsistent. Since these candidates lie between the two stripes, it is possible that these are misidentified or are the optimizer's attempt to fit both the Sgr leading tidal tail and the "bifurcated" stream in a single stream. We plan to fit more streams in the future which could improve our "bifurcated" stream fits.

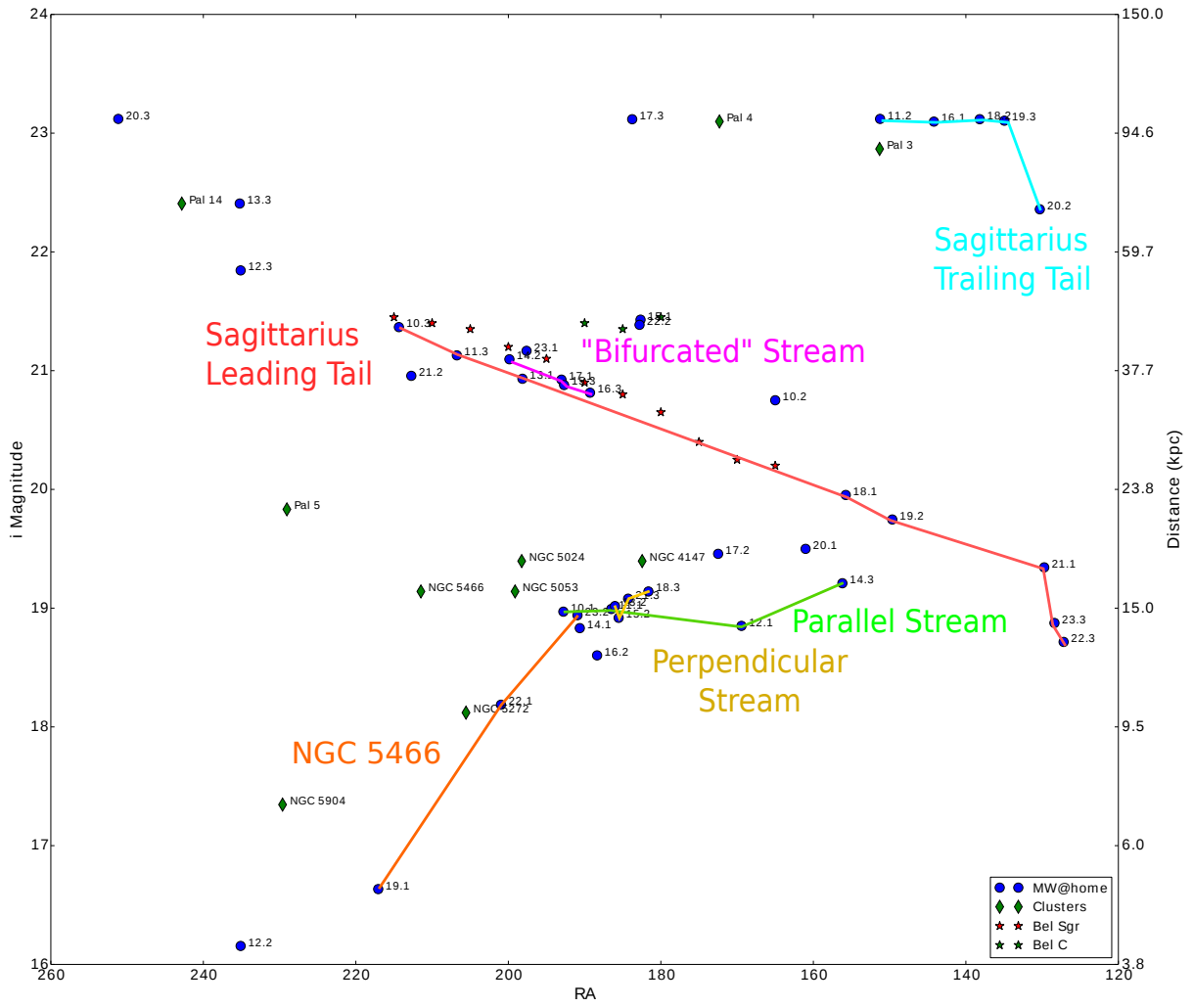


Figure 6.4: Recreate Figure 6.2 with highlighted streams from Tables 6.5, 6.7, and 6.6. The lines are highlighted in the following way: red the Sgr stream, fuchsia for the "bifurcated" stream, orange for the NGC 5466 stream, gold for the Perpendicular Stream, cyan for the 100 kpc stream, and green for the Parallel Stream. These colors correspond to their respective stream in Figure 6.3 for comparison.

6.5 Other substructure

Inspecting Figures 6.1 and 6.2, five other streams seem apparent in our results. We present these other streams and their candidate stream centers in Table 6.6. The substructure we present in this table are: the Sagittarius trailing tail in the north Galactic cap at 100 kpc (Belokurov et al., 2014), the Parallel Stream suggested in Sohn et al. (2016), a new stream which we will refer to as the Perpendicular Stream which may be associated with

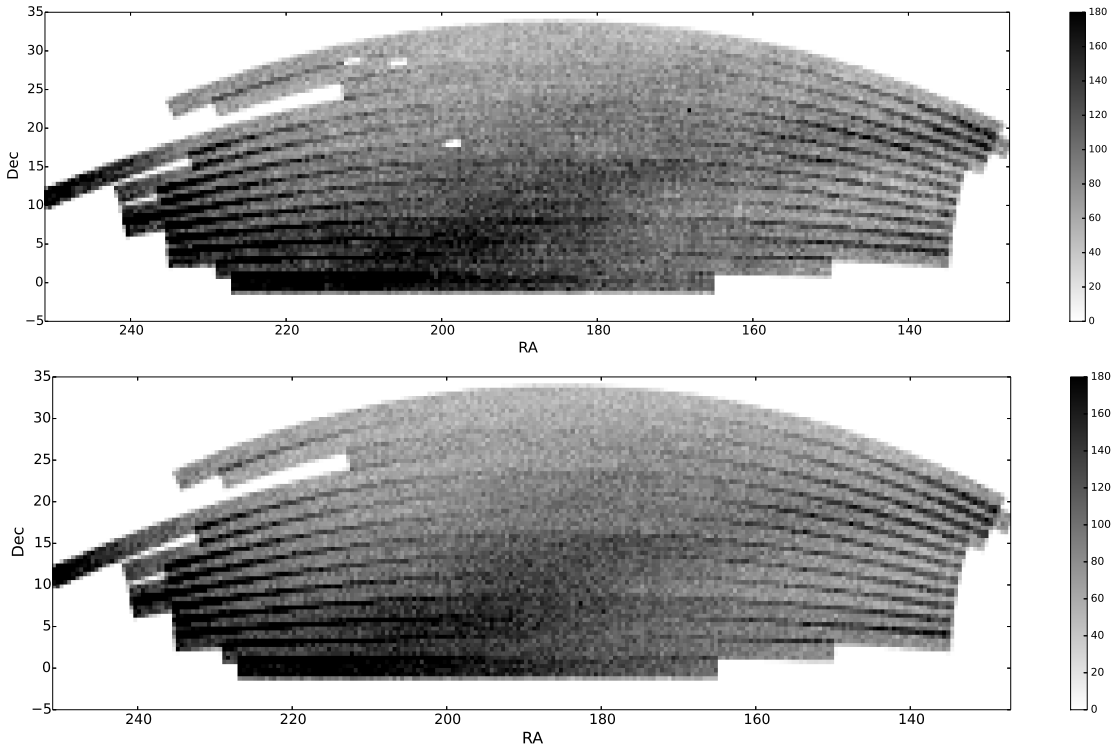


Figure 6.5: Comparison between SDSS north Galactic cap data and fit Milky-Way@home model for the north Galactic cap in RA and dec. We recreate Figure 6.1 here without stream centers in the top panel, and in the bottom panel, we simulate the entire sky using the parameters fit in Tables 6.8 and 6.3. Each stripe was simulated using the test wedge generator following the description from Section 5.2.1 and then plotted in RA and dec. We can see that our model reproduces much of the density substructure in the sky when compared to Figure 6.1. The major substructure missing in the model is the "bifurcated" stream. Maybe fitting the sky with additional streams will improve our model's ability to fit this substructure.

the overdensity in Virgo and passes close to the globular cluster NGC 4147, a stream near the globular cluster NGC 5466 (Belokurov et al., 2006; Grillmair & Johnson, 2006; Fellhauer et al., 2007), and one stream center at the position and distance of Stream C from Belokurov et al. (2006).

We find a pile up of stream centers at 100 kpc in stripes 11, 16, 18, 19, and 20 which may be the trailing tail of Sagittarius in the north Galactic cap. We believe this stream is the Sgr trailing tail in the north Galactic cap. We find that our stream centers are not in the same area of the sky as the Sgr trailing tail in Belokurov et al. (2014). In their work, they fit the trailing tail using a selection of data between $-10 < B < 10$ in Sagittarius coordinates.

On an (RA, dec) plot, this would be above the leading tail of Sgr in declination. Instead, we find the trailing tail is at a lower declination than the leading tidal tail of Sgr, and outside the region selected by Belokurov et al. The Sgr trailing tidal tail is expected to be between 70 and 100 kpc, which is consistent with our result. It should be noted that our distance and μ uncertainties are large for these stream centers. Our distances are also constrained to 100 kpc or less, so if we had more complete data for faint magnitudes it is possible that we might find a result more consistent with Belokurov et al. (2014). It is also possible that Belokurov et al. are fitting the fringes of this wide substructure which should be present in their data selection area.

Using Hubble Space Telescope proper motions, Sohn et al. (2016) presented a new stream that they suggested runs parallel to Sagittarius in velocity. The only detection of this stream is in a pencil beam at (RA, dec) of $(158°, 7°)$. This pencil beam lies on the line between our stream centers for stripes 12 and 14 for the Parallel Stream. The three stars they suggest are part of this stream are at a distance of 17 kpc, which is consistent with our estimate of 13 to 16 kpc for the Parallel Stream. We trace this stream from (RA, dec) of $(209°, -2.3°)$ to $(156°, 8.8°)$, which stream passes right through Virgo. We find it is consistent with the orbit fits to the Virgo Stellar Stream from Carlin et al. (2012). These stream centers suggest that the Parallel Stream might be the same as the Virgo Stellar Stream.

The Perpendicular Stream is a stream that runs almost perpendicular to the Sgr leading tidal tail, the "bifurcated" stream, and the Parallel Stream from (RA, dec) of $(186°, 7.5°)$ to $(184°, 27.5°)$, at a distance of 15 kpc. We find that this stream has a consistent distance to the Virgo overdensity, and if extrapolated 5 degrees in declination, it a crosses the Parallel Stream in the region of the Virgo Overdensity.

We find a stellar stream near the globular cluster NGC 5466 that runs from (RA, dec) of $(217°, 19°)$ to $(191°, 32.3°)$, at a distance of 5 to 15 kpc. The stream we found consists of 3 stream centers across stripes 19, 22, and 23. Our stream is estimated to have an average $\sigma = 5.4$ kpc or 12.6 kpc full width at half maximum (FWHM). We find this stream is very wide and low density. This could explain why it is not readily apparent in density in Figure 6.3. Given this width, the globular cluster NGC 5466 falls within the extent of this stream.

The stream center associated with stream C from Belokurov et al. (2006) is found at an RA and dec of $(182°, 12.5°)$ with a distance of 45.9 kpc. This stream was only fit by

a single stream center with an extremely large width. Our stream is consistent with their measurement of RA, dec and distance even though we only fit it with a single stream center. This stream was fit with a width of 25 kpc as a standard deviation and almost 60 kpc FWHM. It is likely that due to its position in a highly congested portion of the sky, this stream was trying to fit multiple streams in the area, inflating our width measurements.

Table 6.2: Constraints used in the optimizations of SDSS stripes 10 through 23 in the north Galactic cap.

Sagittarius Stream Fits from MilkyWay@home						
Stream 1						
Stripe	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
10	−20 to 20	165 to 227	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
11	−20 to 20	150 to 229	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
12	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
13	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
14	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
15	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
16	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
17	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
18	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
19	−20 to 20	135 to 230	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
20	−20 to 20	133 to 249	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
21	−20 to 20	133 to 210	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
22	−20 to 20	131 to 225	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
23	−20 to 20	133 to 230	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
Stream 2						
10	−20 to 20	165 to 227	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
11	−20 to 20	150 to 229	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
12	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
13	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
14	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
15	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
16	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
17	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
18	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
19	−20 to 20	135 to 230	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
20	−20 to 20	133 to 249	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
21	−20 to 20	133 to 210	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
22	−20 to 20	131 to 225	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
23	−20 to 20	133 to 230	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
Stream 3						
10	−20 to 20	165 to 227	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
11	−20 to 20	150 to 229	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
12	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
13	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
14	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
15	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
16	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
17	−20 to 20	135 to 235	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
18	−20 to 20	135 to 240	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
19	−20 to 20	135 to 230	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
20	−20 to 20	133 to 249	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
21	−20 to 20	133 to 210	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
22	−20 to 20	131 to 225	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0
23	−20 to 20	133 to 230	2.0 to 100.0	0.0 to 3.14	0.0 to 3.14	0.1 to 25.0

Table 6.3: Preliminary results from fitting SDSS stripes 10 through 23 in the north Galactic cap. Streams which are stuck on a constraint boundary are marked with an asterisk.

Stream Fits from MilkyWay@home						
			Stream 1			
Stripe	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
10	-0.89 ± 0.17	192.8 ± 1.4	14.8 ± 0.9	0.39 ± 0.12	2.57 ± 0.26	4.5 ± 0.4
11	-0.79 ± 0.09	186.5 ± 1.1	14.9 ± 0.6	0.82 ± 0.15	2.74 ± 0.16	5.2 ± 0.3
12	-0.69 ± 0.06	169.5 ± 7.1	14.0 ± 1.8	0.43 ± 0.02	0.28 ± 0.04	6.8 ± 0.3
*13	-1.45 ± 0.10	198.0 ± 1.4	36.5 ± 1.1	1.59 ± 0.06	3.14 ± 0.23	4.9 ± 1.7
*14	-1.30 ± 0.08	190.5 ± 1.3	13.9 ± 1.3	0.20 ± 0.04	3.14 ± 0.22	3.7 ± 0.2
15	-1.42 ± 0.09	182.7 ± 14.2	45.9 ± 7.3	1.91 ± 0.13	0.60 ± 0.20	25.0 ± 4.9
16	-1.39 ± 0.1	145.2 ± 4.4	99.0 ± 29.2	1.9 ± 0.14	0.46 ± 0.09	21.0 ± 4.5
17	-1.19 ± 0.14	192.7 ± 1.5	36.4 ± 1.2	1.56 ± 0.13	1.43 ± 0.19	10.1 ± 1.0
18	-2.04 ± 0.15	157.2 ± 0.6	23.3 ± 0.4	2.23 ± 0.13	2.44 ± 0.26	2.5 ± 0.6
19	-0.36 ± 0.07	215.1 ± 12.4	5.0 ± 0.4	0.49 ± 0.04	3.02 ± 0.03	6.2 ± 0.3
20	-0.98 ± 0.05	163.0 ± 1.5	18.9 ± 0.7	1.87 ± 0.22	2.43 ± 0.21	8.2 ± 0.7
*21	-1.51 ± 0.10	133.0 ± 9.1	17.6 ± 0.9	2.78 ± 0.05	3.14 ± 0.12	2.2 ± 0.5
22	-0.23 ± 0.07	198.9 ± 3.4	10.3 ± 0.5	2.09 ± 0.04	1.08 ± 0.04	6.2 ± 0.4
23	-1.08 ± 0.14	195.7 ± 20.4	40.7 ± 5.0	1.67 ± 0.20	0.22 ± 0.15	25.0 ± 6.7
			Stream 2			
*10	-1.17 ± 0.24	165.0 ± 5.9	33.6 ± 8.8	0.06 ± 0.17	1.2 ± 1.31	14.5 ± 1.3
*11	-0.83 ± 0.09	151.3 ± 10.3	100.0 ± 27.6	1.85 ± 0.09	0.8 ± 0.08	25.0 ± 4.6
*12	-2.05 ± 0.11	235.0 ± 23.1	4.0 ± 0.8	2.87 ± 0.05	0.76 ± 0.11	2.3 ± 0.2
13	-1.12 ± 0.13	186.0 ± 0.9	15.1 ± 0.5	1.18 ± 0.16	2.82 ± 0.22	4.8 ± 0.8
14	-1.01 ± 0.06	199.6 ± 1.2	39.4 ± 0.9	1.12 ± 0.04	0.93 ± 0.04	8.9 ± 0.6
15	-1.03 ± 0.10	185.5 ± 1.1	14.4 ± 1.0	0.55 ± 0.14	2.68 ± 0.14	5.2 ± 0.2
*16	-0.62 ± 0.08	188.2 ± 3.5	12.5 ± 1.6	0.61 ± 0.06	3.14 ± 0.07	5.9 ± 0.2
*17	-0.90 ± 0.10	173.0 ± 1.6	18.5 ± 1.2	0.98 ± 0.05	3.13 ± 0.07	5.2 ± 0.4
18	-1.40 ± 0.09	140.0 ± 4.9	99.9 ± 14.4	1.86 ± 0.08	0.42 ± 0.03	16.1 ± 2.2
19	-1.93 ± 0.08	151.8 ± 0.4	21.1 ± 0.3	2.57 ± 0.10	2.79 ± 0.18	1.0 ± 0.2
*20	-2.35 ± 0.39	133.0 ± 6.3	70.4 ± 3.2	1.46 ± 0.12	0.42 ± 0.02	7.3 ± 2.1
*21	-1.10 ± 0.18	210.0 ± 27.2	36.9 ± 9.3	1.8 ± 0.34	0.83 ± 0.18	25.0 ± 8.3
*22	-0.88 ± 0.09	183.1 ± 0.7	45.0 ± 8.4	1.58 ± 0.10	0.91 ± 0.11	25.0 ± 8.0
23	-1.17 ± 0.16	190.0 ± 5.7	14.6 ± 0.7	1.77 ± 0.06	0.86 ± 0.05	3.7 ± 0.5
			Stream 3			
*10	-0.53 ± 0.07	214.4 ± 0.4	44.6 ± 0.3	1.1 ± 0.05	3.14 ± 0.07	7.1 ± 0.5
11	-1.46 ± 0.09	206.7 ± 0.5	40.0 ± 0.6	1.17 ± 0.09	1.03 ± 0.08	4.7 ± 0.6
*12	-0.77 ± 0.05	235.0 ± 5.2	55.6 ± 4.0	1.68 ± 0.03	0.45 ± 0.02	16.4 ± 1.2
*13	-0.88 ± 0.06	235.0 ± 13.6	72.0 ± 21.6	1.55 ± 0.03	0.37 ± 0.03	19.9 ± 3.2
14	-3.07 ± 0.18	156.6 ± 2.7	16.5 ± 0.7	0.33 ± 0.14	0.59 ± 0.46	1.4 ± 0.7
*15	-1.73 ± 0.11	192.5 ± 1.8	35.6 ± 1.1	1.94 ± 0.05	3.14 ± 0.09	4.5 ± 0.8
16	-1.77 ± 0.16	189.1 ± 1.9	34.6 ± 1.0	1.12 ± 0.08	0.9 ± 0.09	3.9 ± 0.8
*17	-4.37 ± 0.23	183.8 ± 10.4	99.9 ± 26.9	0.38 ± 0.13	0.13 ± 0.15	2.1 ± 2.3
18	-0.93 ± 0.07	181.8 ± 1.6	16.0 ± 2.1	2.61 ± 0.05	1.83 ± 0.07	5.8 ± 0.3
*19	-1.00 ± 0.10	137.2 ± 3.6	99.3 ± 17.7	1.81 ± 0.12	0.39 ± 0.04	18.4 ± 3.1
*20	-2.13 ± 0.25	249.0 ± 14.5	100.0 ± 6.3	0.92 ± 0.36	0.48 ± 0.06	11.9 ± 4.2
21	-1.71 ± 0.20	184.4 ± 1.4	15.6 ± 0.5	1.34 ± 0.24	2.8 ± 0.18	3.0 ± 0.6
*22	-2.38 ± 0.14	131.0 ± 4.8	13.2 ± 2.1	0.75 ± 0.11	3.07 ± 0.21	1.6 ± 0.3
*23	-3.00 ± 0.14	133.0 ± 2.4	14.2 ± 1.1	2.25 ± 0.13	1.11 ± 0.31	1.7 ± 0.4

Table 6.4: Easy-to-use preliminary results from fitting SDSS stripes 10 through 23 in the north Galactic cap.

Easy-to-use Stream Results							
Stream 1							
Stripe	Stars	l (deg)	b (deg)	R (kpc)	A	σ (kpc)	$\hat{n}$
10	17417	302.8	62.9	14.8	476.5	4.5	$(-0.32, 0.21, 0.92)$
11	20835	288.0	64.6	14.9	463.4	5.2	$(-0.67, 0.29, 0.68)$
12	28916	253.7	58.4	14.0	196.5	6.8	$(0.4, 0.12, 0.91)$
13	14118	318.1	69.6	36.5	234.3	4.9	$(-1.0, 0.0, -0.02)$
14	16643	295.5	72.7	13.9	374.8	3.7	$(-0.2, 0.0, 0.98)$
15	14751	267.8	72.5	45.9	46.7	25.0	$(0.78, 0.53, -0.33)$
16	13613	222.1	41.8	99.0	58.2	21.0	$(0.85, 0.42, -0.32)$
17	16140	303.9	80.2	36.4	140.0	10.1	$(0.15, 0.99, 0.01)$
18	6987	220.6	54.5	23.3	445.6	2.5	$(-0.6, 0.51, -0.62)$
19	26538	18.8	66.5	5.0	376.0	6.2	$(-0.47, 0.06, 0.88)$
20	25073	213.8	61.0	18.9	260.4	8.2	$(-0.72, 0.63, -0.29)$
21	7670	209.2	31.0	17.6	351.3	2.2	$(-0.35, 0.0, -0.94)$
22	22863	49.7	82.6	10.3	336.8	6.2	$(0.41, 0.77, -0.5)$
23	13064	83.0	83.7	40.7	44.3	25.0	$(0.97, 0.22, -0.1)$
Stream 2							
10	13154	253.7	51.9	33.6	71.2	14.5	$(0.02, 0.06, 1.0)$
11	20072	237.9	43.0	100.0	64.1	25.0	$(0.67, 0.69, -0.28)$
12	7432	9.8	43.1	4.0	829.1	2.3	$(0.2, 0.19, -0.96)$
13	19569	283.4	69.3	15.1	454.0	4.8	$(-0.88, 0.29, 0.38)$
14	22219	325.0	71.3	39.4	178.1	8.9	$(0.54, 0.72, 0.44)$
15	21784	276.4	73.8	14.4	391.1	5.2	$(-0.46, 0.23, 0.85)$
16	29400	283.0	77.1	12.5	371.7	5.9	$(-0.58, 0.0, 0.82)$
17	21687	236.7	68.7	18.5	296.4	5.2	$(-0.83, 0.01, 0.56)$
18	13289	215.9	37.6	99.9	64.9	16.1	$(0.87, 0.39, -0.29)$
19	5512	215.6	49.5	21.1	1132.0	1.0	$(-0.51, 0.18, -0.84)$
20	6374	211.0	31.0	70.4	79.6	7.3	$(0.91, 0.4, 0.11)$
21	11659	29.8	72.0	36.9	49.5	25.0	$(0.66, 0.72, -0.23)$
22	11959	192.7	80.7	45.0	45.7	25.0	$(0.61, 0.79, -0.01)$
23	11904	140.2	84.5	14.6	239.9	3.7	$(0.64, 0.74, -0.2)$
Stream 3							
10	24985	343.7	55.9	44.6	364.3	7.1	$(-0.89, 0.0, 0.45)$
11	10634	333.5	61.9	40.0	239.3	4.7	$(0.48, 0.79, 0.39)$
12	26650	9.8	43.1	55.6	80.6	16.4	$(0.89, 0.43, -0.11)$
13	25009	11.7	43.9	72.0	78.2	19.9	$(0.93, 0.36, 0.02)$
14	2827	233.9	50.8	16.5	287.8	1.4	$(0.27, 0.18, 0.95)$
15	10790	302.3	75.3	35.6	203.6	4.5	$(-0.93, 0.0, -0.36)$
16	9280	286.9	77.4	34.6	224.6	3.9	$(0.56, 0.71, 0.43)$
17	673	259.9	77.2	99.9	149.3	2.1	$(0.37, 0.05, 0.93)$
18	21202	245.4	77.5	16.0	232.6	5.8	$(-0.13, 0.49, -0.86)$
19	14031	213.3	35.0	99.3	64.8	18.4	$(0.9, 0.37, -0.24)$
20	7936	28.1	33.0	100.0	83.0	11.9	$(0.71, 0.37, 0.6)$
21	6304	208.2	82.4	15.6	341.6	3.0	$(-0.92, 0.32, 0.23)$
22	2660	207.5	29.0	13.2	452.1	1.6	$(-0.68, 0.05, 0.73)$
23	1909	205.6	30.8	14.2	449.4	1.7	$(0.35, 0.7, -0.63)$

Table 6.5: Candidates for the Sgr leading tidal tail. We did not find stream centers consistent with Sgr in all stripes.

Sagittarius Candidates						
Stripe	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
10.3	-0.53 ± 0.07	214.4 ± 0.4	44.6 ± 0.3	1.1 ± 0.05	3.14 ± 0.07	7.1 ± 0.5
11.3	-1.46 ± 0.09	206.7 ± 0.5	40.0 ± 0.6	1.17 ± 0.09	1.03 ± 0.08	4.7 ± 0.6
13.1	-1.45 ± 0.10	198.0 ± 1.4	36.5 ± 1.1	1.59 ± 0.06	3.14 ± 0.23	4.9 ± 1.7
18.1	-2.04 ± 0.15	157.2 ± 0.6	23.3 ± 0.4	2.23 ± 0.13	2.44 ± 0.26	2.5 ± 0.6
19.2	-1.93 ± 0.08	151.8 ± 0.4	21.1 ± 0.3	2.57 ± 0.10	2.79 ± 0.18	1.0 ± 0.2
21.1	-1.51 ± 0.10	133.0 ± 9.1	17.6 ± 0.9	2.78 ± 0.05	3.14 ± 0.12	2.2 ± 0.5
22.3	-2.38 ± 0.14	131.0 ± 4.8	13.2 ± 2.1	0.75 ± 0.11	3.07 ± 0.21	1.6 ± 0.3
23.3	-3.00 ± 0.14	133.0 ± 2.4	14.2 ± 1.1	2.25 ± 0.13	1.11 ± 0.31	1.7 ± 0.4

Table 6.6: The remaining substructure candidates in our results. These streams were found by looking at the correlation between stream centers in RA and dec as well as RA and distance. For most of these streams, we were able to find a corresponding, previously discovered stream in the neighborhood of the stream centers we fit.

Candidates for Other Stellar Streams						
Parallel Stream (Sohn et al., 2016)						
Stripe	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
10.1	-0.89 ± 0.17	192.8 ± 1.4	14.8 ± 0.9	0.39 ± 0.12	2.57 ± 0.26	4.5 ± 0.4
11.1	-0.79 ± 0.09	186.5 ± 1.1	14.9 ± 0.6	0.82 ± 0.15	2.74 ± 0.16	5.2 ± 0.3
12.1	-0.69 ± 0.06	169.5 ± 7.1	14.0 ± 1.8	0.43 ± 0.02	0.28 ± 0.04	6.8 ± 0.3
14.3	-3.07 ± 0.18	156.6 ± 2.7	16.5 ± 0.7	0.33 ± 0.14	0.59 ± 0.46	1.4 ± 0.7
100 kpc Stream (Sagittarius Trailing Tail)						
11.2	-0.83 ± 0.09	151.3 ± 10.3	100.0 ± 27.6	1.85 ± 0.09	0.8 ± 0.08	25.0 ± 4.6
16.1	-1.39 ± 0.1	145.2 ± 4.4	99.0 ± 29.2	1.9 ± 0.14	0.46 ± 0.09	21.0 ± 4.5
18.2	-1.40 ± 0.09	140.0 ± 4.9	99.9 ± 14.4	1.86 ± 0.08	0.42 ± 0.03	16.1 ± 2.2
19.3	-1.00 ± 0.10	137.2 ± 3.6	99.3 ± 17.7	1.81 ± 0.12	0.39 ± 0.04	18.4 ± 3.1
20.2	-2.35 ± 0.39	133.0 ± 6.3	70.4 ± 3.2	1.46 ± 0.12	0.42 ± 0.02	7.3 ± 2.1
Perpendicular Stream						
13.2	-1.12 ± 0.13	186.0 ± 0.9	15.1 ± 0.5	1.18 ± 0.16	2.82 ± 0.22	4.8 ± 0.8
15.2	-1.03 ± 0.10	185.5 ± 1.1	14.4 ± 1.0	0.55 ± 0.14	2.68 ± 0.14	5.2 ± 0.2
18.3	-0.93 ± 0.07	181.8 ± 1.6	16.0 ± 2.1	2.61 ± 0.05	1.83 ± 0.07	5.8 ± 0.3
21.3	-1.71 ± 0.20	184.4 ± 1.4	15.6 ± 0.5	1.34 ± 0.24	2.8 ± 0.18	3.0 ± 0.6
NGC 5466 Stream						
19.1	-0.36 ± 0.07	215.1 ± 12.4	5.0 ± 0.4	0.49 ± 0.04	3.02 ± 0.03	6.2 ± 0.3
22.1	-0.23 ± 0.07	198.9 ± 3.4	10.3 ± 0.5	2.09 ± 0.04	1.08 ± 0.04	6.2 ± 0.4
23.2	-1.17 ± 0.16	190.0 ± 5.7	14.6 ± 0.7	1.77 ± 0.06	0.86 ± 0.05	3.7 ± 0.5
Stream C (Belokurov et al., 2006)						
15.1	-1.42 ± 0.09	182.7 ± 14.2	45.9 ± 7.3	1.91 ± 0.13	0.60 ± 0.20	25.0 ± 4.9

Table 6.7: Candidates for the "bifurcated" stream. These stream centers resemble the "bifurcated" stream most closely, however these stream centers fall between Sgr and the "bifurcated" stream. This makes it possible that the optimization tried to fit both streams with a single stream in this region and more streams are required to be fit in these stripes to properly disentangle the two.

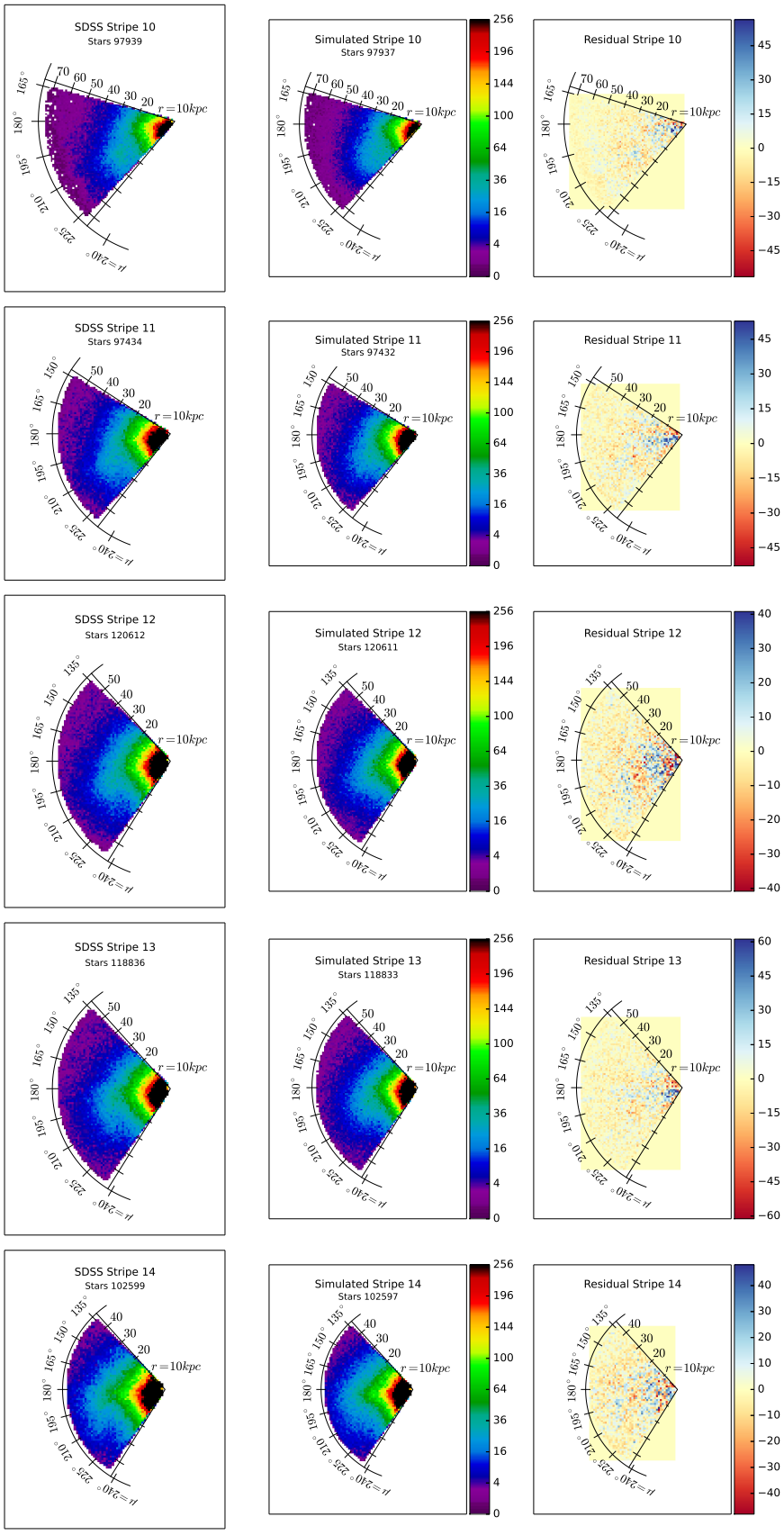
"Bifurcated" Stream Candidates						
Stripe	ε	μ (deg)	R (kpc)	θ (rad)	ϕ (rad)	σ (kpc)
14.2	-1.01 ± 0.06	199.6 ± 1.2	39.4 ± 0.9	1.12 ± 0.04	0.93 ± 0.04	8.9 ± 0.6
15.3	-1.73 ± 0.11	192.5 ± 1.8	35.6 ± 1.1	1.94 ± 0.05	3.14 ± 0.09	4.5 ± 0.8
16.3	-1.77 ± 0.16	189.1 ± 1.9	34.6 ± 1.0	1.12 ± 0.08	0.9 ± 0.09	3.9 ± 0.8
17.1	-1.19 ± 0.14	192.7 ± 1.5	36.4 ± 1.2	1.56 ± 0.13	1.43 ± 0.19	10.1 ± 1.0

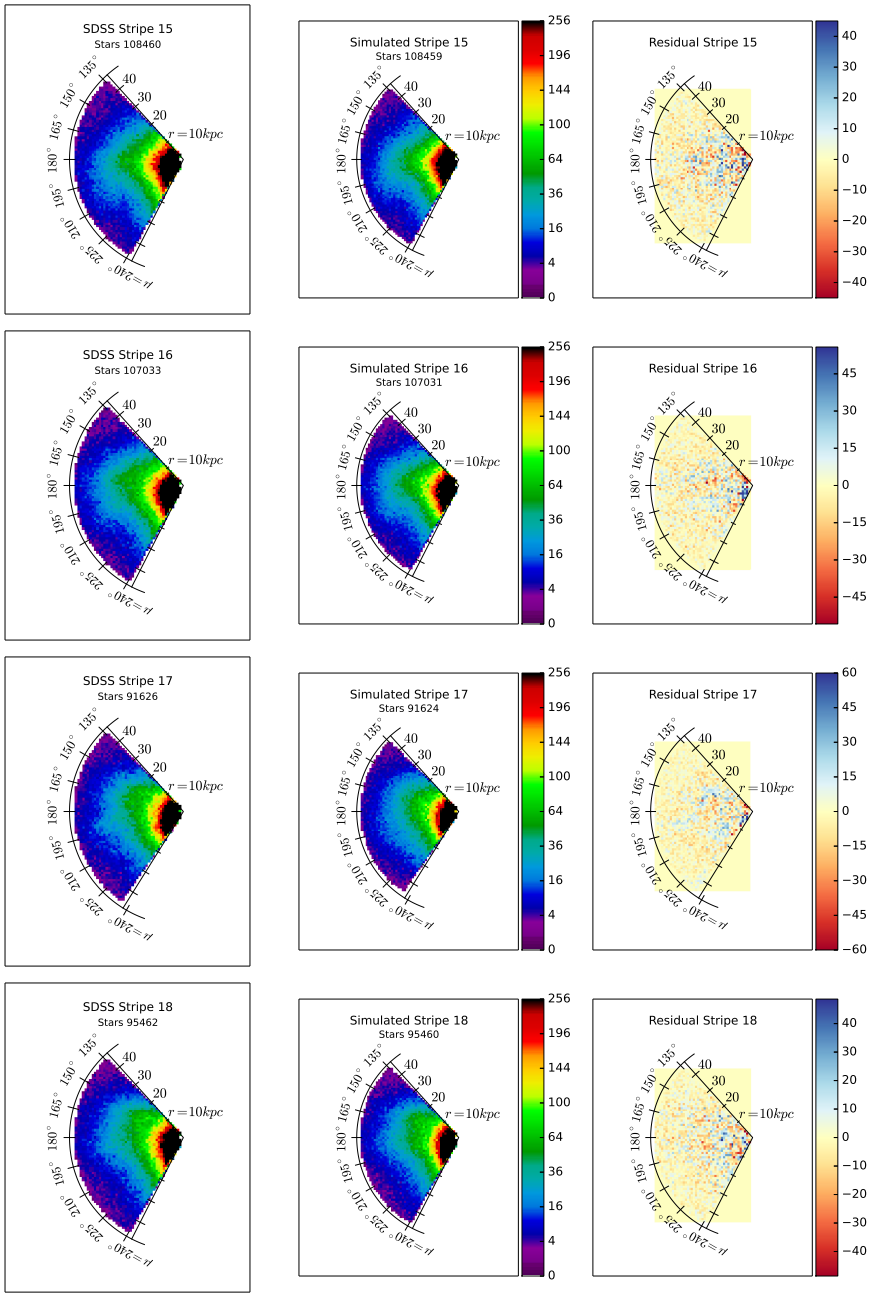
Table 6.8: Preliminary results from fitting the SDSS stripes 10 through 23 in the north Galactic cap. These results are all consistent showing a similarly low disk weight and a prolate halo.

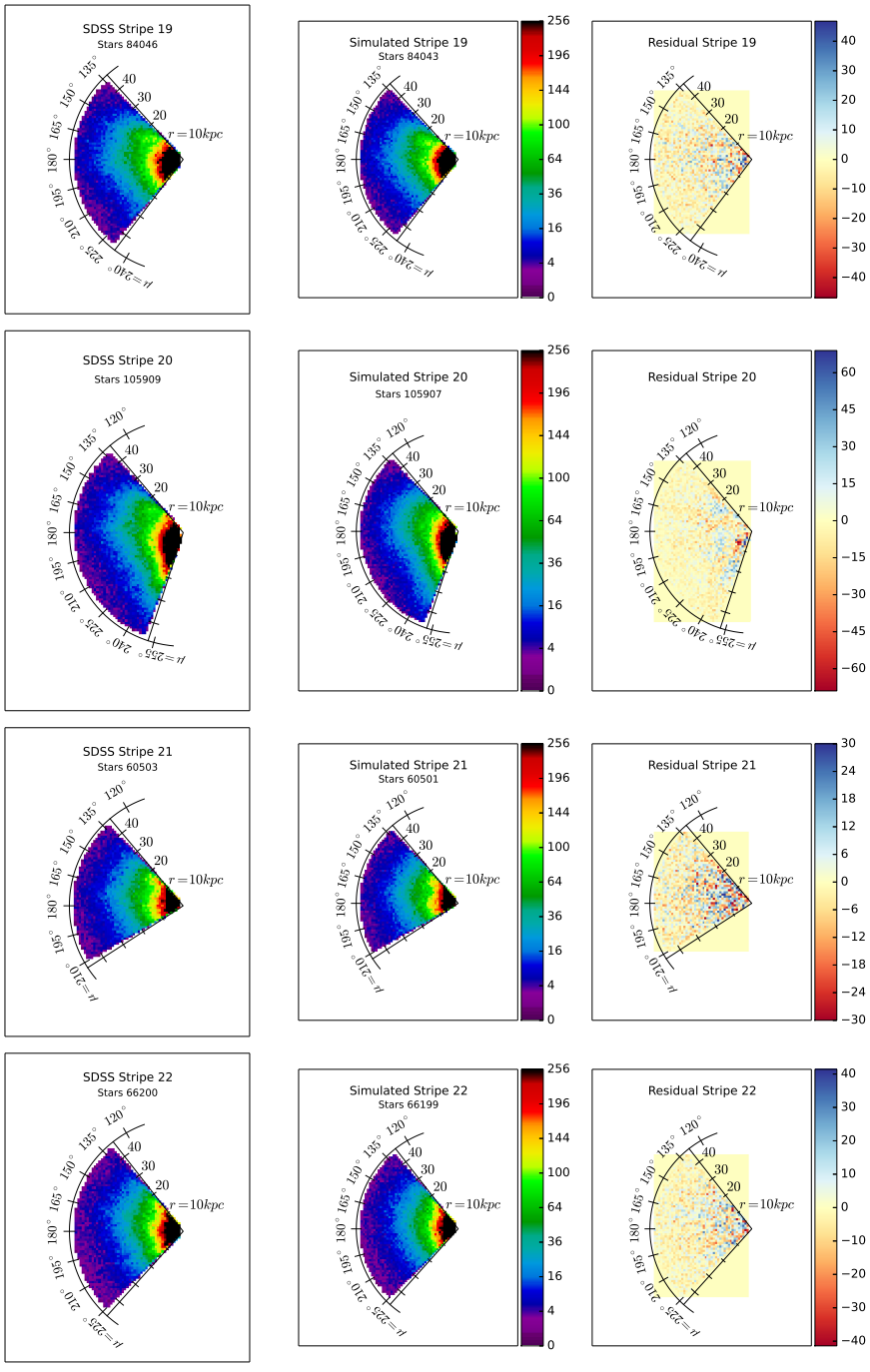
Halo Fits from MilkyWay@home		
Stripe	ε_{sph}	q
10	0.9979 ± 0.0004	0.54 ± 0.02
11	0.9980 ± 0.0003	0.54 ± 0.02
12	0.9972 ± 0.0003	0.61 ± 0.01
13	0.9982 ± 0.0003	0.54 ± 0.01
14	0.9972 ± 0.0003	0.62 ± 0.01
15	0.9976 ± 0.0003	0.57 ± 0.01
16	0.9970 ± 0.0004	0.58 ± 0.02
17	0.9968 ± 0.0003	0.62 ± 0.02
18	0.9969 ± 0.0004	0.62 ± 0.01
19	0.9968 ± 0.0005	0.57 ± 0.02
20	0.9967 ± 0.0003	0.64 ± 0.02
21	0.9977 ± 0.0004	0.62 ± 0.03
22	0.9970 ± 0.0007	0.51 ± 0.03
23	0.9979 ± 0.0004	0.55 ± 0.03

Table 6.9: Easy-to-use preliminary results from fitting the SDSS stripes 10 through 23 in the north Galactic cap.

Easy-to-Use Halo Results					
Stripe	Stars	A	q	r_0 (kpc)	ε_{sph}
10	42383	423311785	0.54	12.0	0.9979
11	45893	392253082	0.54	12.0	0.9980
12	57614	291929528	0.61	12.0	0.9972
13	60141	414265102	0.54	12.0	0.9982
14	60909	334772542	0.62	12.0	0.9972
15	61134	369434130	0.57	12.0	0.9976
16	54740	326199868	0.58	12.0	0.9970
17	53126	314525188	0.62	12.0	0.9968
18	53984	295907096	0.62	12.0	0.9969
19	37966	315552526	0.57	12.0	0.9968
20	66526	278955944	0.64	12.0	0.9967
21	34870	348488676	0.62	12.0	0.9977
22	28718	317605856	0.51	12.0	0.9970
23	38477	355214806	0.55	12.0	0.9979


SDSS Stripe 10
Stars 97939
Simulated Stripe 10
Stars 97937
Residual Stripe 10
SDSS Stripe 11
Stars 97434
Simulated Stripe 11
Stars 97432
Residual Stripe 11
SDSS Stripe 12
Stars 120612
Simulated Stripe 12
Stars 120611
Residual Stripe 12
SDSS Stripe 13
Stars 118836
Simulated Stripe 13
Stars 118833
Residual Stripe 13
SDSS Stripe 14
Stars 102599
Simulated Stripe 14
Stars 102597
Residual Stripe 14
r = 10kpc


SDSS Stripe 15
Stars 108460
Simulated Stripe 15
Stars 108459
Residual Stripe 15
SDSS Stripe 16
Stars 107033
Simulated Stripe 16
Stars 107031
Residual Stripe 16
SDSS Stripe 17
Stars 91626
Simulated Stripe 17
Stars 91624
Residual Stripe 17
SDSS Stripe 18
Stars 95462
Simulated Stripe 18
Stars 95460
Residual Stripe 18
r =10kpc
μ =240°


SDSS Stripe 19
Stars 84046
Simulated Stripe 19
Stars 84043
Residual Stripe 19
SDSS Stripe 20
Stars 105909
Simulated Stripe 20
Stars 105907
Residual Stripe 20
SDSS Stripe 21
Stars 60503
Simulated Stripe 21
Stars 60501
Residual Stripe 21
SDSS Stripe 22
Stars 66200
Simulated Stripe 22
Stars 66199
Residual Stripe 22
r = 10kpc

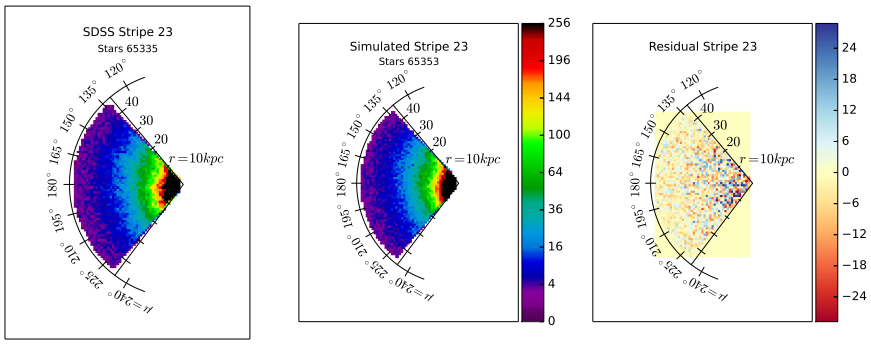


Figure 6.6: Wedge plots and residuals for all fit stripes. We created a 2D wedge plot for each SDSS stripe from 10 to 23 using SDSS data, simulated SDSS data and the residual between the two. Each pixel in these plots shows the MSTO density in a 1 kpc by 1 kpc by 2.5° pixel. We find that most stripes are not reproduced exactly in our simulated stripes. This suggests that there are likely unfit substructures or there are improvements we can make to our density model.

CHAPTER 7
Discussion and future work

7.1 Discussion

Of the streams that we found in the north Galactic cap, we believe the 100 kpc stream, the Parallel Stream, the Perpendicular Stream, and the stream near NGC 5466 are the most exciting. The 100 kpc stream fits could have an implication for the shape and structure of the Milky Way dark matter halo. The Parallel and Perpendicular streams could explain previous results in the Virgo Overdensity. Finally, the stream near NGC 5466 may be a new stream unassociated with the globular cluster or it may contain the globular cluster and its stream.

If the 100 kpc stream we found is the trailing tail of Sgr in the north Galactic cap, then our width measurements could have implications for the shape and properties of the Milky Way dark matter halo (Ibata et al., 2001; Siegal-Gaskins & Valluri, 2008; Ngan et al., 2016; Sandford et al., 2017). Ibata et al. (2001) find that a wide tail for a stellar stream is indicative of a flattened or oblate dark matter halo. Siegal-Gaskins & Valluri (2008), Ngan et al. (2016), and Sandford et al. (2017), present methods to determine the number of dark matter subhalos and constraints on their mass by looking at the width and gaps present in the tidal streams. In general, more gaps and wider streams are correlated with a higher number of subhalos in the dark matter halo as well as more massive subhalos. Our measurements suggest a stream that is a 47° FWHM across, which corresponds to a width 17.6 kpc for the trailing tail of Sgr. This wide stream could suggest an oblate or a lumpy dark matter halo.

The distances to both the Parallel Stream and the Perpendicular Stream are consistent with the distance of 6-20 kpc found in Jurić et al. (2008) for the Virgo Overdensity. Jurić et al. (2008) also suggest that Virgo extends over 1000 square degrees on the sky, which is consistent with our model of two perpendicular, close, and wide streams with distance around 15 kpc and widths around 4.3 and 4.7 kpc. A sigma of 4.5 kpc corresponds to a

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., Newby, M., & Desell, T. submittedb, ApJS

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., & Desell, T. submitteda, ApJ. https://arxiv.org/abs/1807.03754

10.5 kpc FWHM, which at 15 kpc away is 38.5° across. A circular region of the sky with this width is 1200 sq. deg. Duffau et al. (2006) also suggest the stream is wide, taking up over 100 square degrees on the sky; they find the structure at 18 kpc which is consistent with our result. Carlin et al. (2012) fit an orbit to the Virgo Stellar Stream and present an N-body simulation. The Parallel Stream lies directly on the orbit in both RA and dec, as well as distance. This strongly suggests the Parallel Stream is in fact the Virgo Stellar Stream. This also suggests the Perpendicular Stream is a previously unknown component to the Virgo Overdensity.

Having two crossing streams within the Virgo Overdensity can also explain the multiple velocity measurements for moving groups in the region of Virgo. Duffau et al. (2014) and Newberg et al. (2007) find a group of stars with a positive line-of-sight velocity around 130 km s^{-1}. Duffau et al. find two other moving groups with line-of-sight velocities of -94 km s^{-1} and 32 km s^{-1}. Newberg et al. find two group with line-of-sight velocities of -168 km s^{-1} and -76 km s^{-1}. It is reasonable to expect that these two streams could account for two of the moving groups found by these papers.

The Perpendicular Stream passes directly through the NGC 4147 globular cluster which has an RA, dec and distance of $(183°, 18.5°, 18\text{kpc})$ in the Harris catalog (Harris, 1996). Considering the proximity of this globular cluster to the stream, this globular cluster is a candidate for belonging to this stream.

We found a stream near the globular cluster NGC 5466, but we believe it is a different stream than the one previously found with the globular cluster as its progenitor. To avoid confusion, I will refer to the streams by the name NGC 5466 Stream W and the previously found stream by the name NGC 5466 Stream B. The NGC 5466 Stream B has a heliocentric distance between 15 and 20 kpc at an (RA, dec) near $(211°, 28.5°)$ (Belokurov et al., 2006; Grillmair & Johnson, 2006; Fellhauer et al., 2007). We measure NGC 5466 Stream W comes within 6 kpc from the sun which is closer than Fellhauer et al. (2007) predict in their orbit fit to NGC 5466 Stream B. NGC 5466 Stream W is also too wide 12.6 kpc FWHM to be the same stream as NGC 5466 Stream B. In Grillmair & Johnson (2006) they find the width of NGC 5466 Stream B to be between 1 and 2 degrees, which at 15 kpc away gives the stream a width of ~ 250 pc. This is significantly smaller than NGC 5466 Stream W. However, it is reasonable given the proximity and similar orientation of the two streams that the NGC 5466 Stream B belongs to NGC 5466 Stream W.

7.2 Future work

In the future, we plan to improve our model, fit the SDSS south, fit data sets from other surveys, and we will consider other applications for statistical photometric parallax, such as fitting the Milky Way disk. There are several ways to improve our model in the future: more streams will be fit in some stripes, more stellar background model will be explored, and different stellar tracers can be used.

The most important improvement required for this research is to increase the number of streams fit in a given wedge of data. In this thesis, we made an educated guess of how many streams we expect there to be, but our work has shown that three streams is insufficient for at least some of the stripes. There are some technical or statistical limitations to fitting large numbers of streams. Examples of technical limitations are: the size of the command line when entering stream parameters to the modeling program (which currently limits us to four streams fit per wedge), the increased computational time required for each likelihood evaluation, and the increased number of likelihood evaluations required to reach convergence as the number of parameters increases. The number of stars in the dataset, the degree of stream overlap, and the similarity of the stream density profile to the model for each stream will also limit the number of streams that can be fit.

In the future, we plan to use different background models to try to determine which gives the best fit to the data. While we explored the Einasto profile, the broken power law, and the Hernquist profile, we only tried fitting the Hernquist profile with a fixed scale radius. Attempting to fit these other profiles may yield better results. If we are successful in deriving the smooth background component from statistical photometric parallax or if our model can be informed from other analyses, then we could fix the smooth component and fit the stream parameters.

We will fit the SDSS data in the south Galactic cap with our model. Sgr and the "bifurcated" stream are also seen in the south and we plan to trace their density there. It is also possible we will find new streams in the south, just as we did in the north.

Our model will also be applied to stars from surveys other than the SDSS such as Pan-STARRS (Chambers et al., 2016) and DES (The Dark Energy Survey Collaboration, 2005, 2016). These surveys will require a different absolute magnitude distribution, detection efficiency model, and selection efficiency model since ours was determined for SDSS data only. These models will change based on the detection efficiency and color errors of the survey

being used.

Different tracer populations can be used in conjunction with the statistical photometric parallax method we described. Any stellar population can be used with this model as long as the absolute magnitude distribution of that population can be determined. An example of this is main sequence stars where there are an even larger number of stars than there are in the MSTO population, but they are fainter so they probe a shorter distance.

We will develop a statistical photometric parallax method for fitting the disk using MSTO stars. Recent work by Xu et al. (2015) has shown that there are wiggles present in the disk density. Our method will measure the wiggles out to a distance of ~ 45 kpc from the Sun, depending on extinction due to dust. When implementing this new model, we plan to use Pan-STARRS for stellar data since the SDSS only has a few stripes that extend to low Galactic latitudes where disk structure can be studied.

CHAPTER 8
Conclusions

In this thesis, we show the result of an updated version of the maximum likelihood technique developed by Cole et al. (2008) and Newby et al. (2013), using updated MSTO absolute magnitude distributions from Newby et al. (2011) and then fit both simulated data and real data from the SDSS. The major conclusions are:

1. Our new model returns parameters that are consistent with those used to create simulations, and is ready to use on real SDSS data. Clean separations of all of the simulated substructure in a simulated SDSS stripe 19 can be seen in Figures 5.1 and 5.6, and the comparisons of simulated to optimized parameter values can be found in Table 5.4. During our model validation, and with the inclusion of the absolute magnitude distribution and selection efficiency from Newby et al. (2011) we fit the Sgr leading tidal tail, the trailing tail of Sgr in the north, and a third stream which is consistent with at least some measurements of the Virgo Overdensity in stripe 19. The fit density parameters for these structures can be found in Table 5.1; the parameters from which the densities can be more easily derived using Equations 5.1 and 5.2 are listed in Table 5.3.

2. We find an oblate halo in all 14 stripes we fit in the north Galactic cap with an average halo flattening parameter of 0.58. We also find that our bluer MSTO color selection effectively removes most of the disk contamination in our stripes.

3. When fitting the entire sky with minimal constraints, our model was able to fit the Sgr leading tidal tail in 8 stripes with results consistent with previous values found in Belokurov et al. (2006), Newberg et al. (2007), Newby et al. (2013), Belokurov et al. (2014), and Hernitschek et al. (2017).

4. We find stream centers consistent with the Sgr trailing tidal tail in 5 stripes. These

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., Newby, M., & Desell, T. submittedb, ApJS

Portions of this chapter have been submitted to: Weiss, J., Newberg, H. J., & Desell, T. submitteda, ApJ. https://arxiv.org/abs/1807.03754

stream centers are found between 70 and 100 kpc from the Sun and centered several degrees from the selection area of Belokurov et al. (2014).

5. We traced the Parallel Stream from the position it was detected in Sohn et al. (2016) to Virgo. The path of the stream matches the orbit fit to the Virgo Stellar Stream in Carlin et al. (2012) suggesting that the Virgo Stellar stream and the Parallel Stream are the same structure.

6. We found a new stream which we refer to as the Perpendicular Stream, which runs roughly perpendicular to Sgr, the "bifurcated" stream, and the Parallel Stream. It has a heliocentric distance of 15 kpc, the same distance as the Virgo Overdensity. This stream may contain the globular cluster NGC 4147. The Perpendicular and Parallel Streams cross at the same distance and position as the most dense region of the Virgo Overdensity.

7. We found a new stream near NGC 5466. With an average radius of 5.4 kpc this stream is too wide to have been created by the globular cluster, but it appears to be in a similar sky position and oriented in a similar direction to the narrow tidal tails previously detected around NGC 5466.

8. There was a single, 25 kpc wide stream in our model that was fit to Stream C from Belokurov et al. (2006). This stream is in a congested part of the sky, making it likely to overestimate the stream's width as it tries to compensate for not having enough streams to fit all of the substructure in the area.

9. Our results strongly indicate that in the future we need to fit more than three tidal streams to each stripe of SDSS data.

In summary, we have developed and tested an improved algorithm for fitting stellar substructure in the Milky Way halo and demonstrated it on the data available from SDSS stripe 19 as well as on simulated test data. We provide preliminary density information for the substructure fit in SDSS stripe 19. We demonstrate our ability to correct for biases introduced when streams are near the limiting magnitude of the survey and our ability to separate overlapping substructures from each other.

We identify 7 streams in the SDSS north Galactic cap in stripes 10 through 23 with MilkyWay@home. We report their model parameters and compare them with relevant literature. We believe 2 of these 7 streams may be newly discovered streams, the Parallel Stream and the NGC 5466 Stream W. The discovery of these streams was made possible by the powerful modeling and optimization capabilities of MilkyWay@home and its ability to simultaneously fit multiple substructures.

Looking forward, this algorithm will be run on all of the available SDSS data in the north and south Galactic caps with a halo model that includes additional streams. By fitting additional streams, we hope to improve our fits of the previously found substructure and find new substructure to expand our understanding of tidal streams in the Milky Way.

BIBLIOGRAPHY

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

Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42

Akhter, S., Costa, G. S. D., Keller, S. C., & Schmidt, B. P. 2012, ApJ, 756, 23

Anderson, D. P., Korpela, E., & Walton, R. 2005, in First International Conference on e-Science and Grid Computing (e-Science'05) (Melbourne, Vic., Australia: IEEE), 8

Arsenault, M., Shelton, S., Desell, T., et al. 2018, Milkyway-at-home/milkywayathome_client: Separation, v1.46, Zenodo, doi: 10.5281/zenodo.1257427

Belokurov, V., Evans, N. W., Irwin, M. J., Hewett, P. C., & Wilkinson, M. I. 2006, ApJ, 637, L29

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

Belokurov, V., Koposov, S. E., Evans, N. W., et al. 2014, MNRAS, 437, 116

Bonaca, A., Jurić, M., Željko Ivezić, et al. 2012, AJ, 143, 105

Bovy, J., Bahmanyar, A., Fritz, T. K., & Kallivayalil, N. 2016, ApJ, 833, 31

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

Carlberg, R. G. 2012, in ASPCS, Vol. 458, Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way, ed. W. Aoki, M. Ishigaki, T. Suda, T. Tsujimoto, & N. Arimoto (San Francisco, CA: ASP), 379

Carlin, J. L., Yam, W., Casetti-Dinescu, D. I., et al. 2012, ApJ, 753, 145

Carraro, G., Zinn, R., & Moni Bidin, C. 2007, A&A, 466, 181

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560

Cole, N. 2009, PhD thesis, Rensselaer Polytechnic Institute

Cole, N., Newberg, H. J., Magdon-Ismail, M., et al. 2008, ApJ, 683, 750

Debattista, V. P., Roškar, R., Valluri, M., et al. 2013, MNRAS, 434, 2971

Desell, T., Cole, N., Magdon-Ismail, M., et al. 2007, in IEEE International Conference on e-Science and Grid Computing (Bangalore, India: IEEE), 337–344

Dierickx, M. I. P., & Loeb, A. 2017a, ApJ, 847, 42

—. 2017b, ApJ, 836, 92

Duffau, S., Vivas, A. K., Zinn, R., Méndez, R. A., & Ruiz, M. T. 2014, A&A, 566, A118

Duffau, S., Zinn, R., Vivas, A. K., et al. 2006, ApJ, 636, L97

Fellhauer, M., Evans, N. W., Belokurov, V., Wilkinson, M. I., & Gilmore, G. 2007, MNRAS, 380, 749

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

Grabowski, K., Newby, M., & Newberg, H. J. 2013, Journal of Undergraduate Research in Physics, 26

Grillmair, C. J. 2017a, ApJ, 847, 119

—. 2017b, ApJ, 834, 98

Grillmair, C. J., & Carlin, J. L. 2016, in Astrophysics and Space Science Library, Vol. 420, Tidal Streams in the Local Group and Beyond, ed. H. J. Newberg & J. L. Carlin (Cambridge, MA: Springer International Publishing), 87

Grillmair, C. J., & Johnson, R. 2006, ApJ, 639, L17

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332

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

Hernitschek, N., Sesar, B., Rix, H.-W., et al. 2017, ApJ, 850, 96

Hernquist, L. 1990, ApJ, 356, 359

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

Ibata, R. A., Gilmore, G., & Irwin, M. J. 1995, MNRAS, 277, 781

Ivezić, v., Connolly, A. J., VanderPlas, J. T., & Gray, A. 2014, Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (Princeton, NJ, USA: Princeton University Press)

Jethwa, P., Torrealba, G., Navarrete, C., et al. 2017, arXiv:1711.09103

Johnston, K. V. 2002, in The Shapes of Galaxies and their Dark Halos, ed. P. Natarajan (Hackensack, NJ: World Scientific Publishing Co. Pte. Ltd.), 194–201

Johnston, K. V., & Carlberg, R. G. 2016, in Astrophysics and Space Science Library, Vol. 420, Tidal Streams in the Local Group and Beyond, ed. H. J. Newberg & J. L. Carlin (Cambridge, MA: Springer International Publishing), 169

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

Keller, S. C., Murphy, S., Prior, S., DaCosta, G., & Schmidt, B. 2008, ApJ, 678, 851

Koposov, S. E., Belokurov, V., & Evans, N. W. 2013, ApJ, 766, 79

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

Küpper, A. H. W., Zonoozi, A. H., Haghi, H., et al. 2015, arXiv:1512.05967

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

Law, D. R., & Majewski, S. R. 2016, in Astrophysics and Space Science Library, Vol. 420, Tidal Streams in the Local Group and Beyond, ed. H. J. Newberg & J. L. Carlin (Cambridge, MA: Springer International Publishing), 31

Li, J., Smith, M. C., Zhong, J., et al. 2016a, ApJ, 823, 59

Li, T. S., Balbinot, E., Mondrik, N., et al. 2016b, ApJ, 817, 135

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

Mateu, C., Read, J. I., & Kawata, D. 2018, MNRAS, 474, 4112

Newberg, H. J. 2012, Proceedings of the International Astronomical Union, 8, 74

Newberg, H. J., & Yanny, B. 2006, JPhCS, 47, 195

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

Newberg, H. J., Yanny, B., & Willett, B. A. 2009, ApJ, 700, L61

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

Newberg, H. J., Yanny, B., Grebel, E. K., et al. 2003, ApJ, 596, L191

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

Ngan, W., Carlberg, R. G., Bozek, B., et al. 2016, ApJ, 818, 194

Niederste-Ostholt, M., Belokurov, V., Evans, N. W., & Peñarrubia, J. 2010, ApJ, 712, 516

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

Pearson, S., Price-Whelan, A. M., & Johnston, K. V. 2017, NatAs, 1, 633

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

Retana-Montenegro, E., van Hese, E., Gentile, G., Baes, M., & Frutos-Alfaro, F. 2012, A&A, 540, A70

Ruhland, C., Bell, E. F., Rix, H.-W., & Xue, X.-X. 2011, ApJ, 731, 119

Sanderson, R. E., Hartke, J., & Helmi, A. 2017, ApJ, 836, 234

Sandford, E., Küpper, A. H. W., Johnston, K. V., & Diemand, J. 2017, MNRAS, 470, 522

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

Sesar, B., Hernitschek, N., Dierickx, M. I. P., Fardal, M. A., & Rix, H.-W. 2017, ApJ, 844, L4

Shipp, N., Drlica-Wagner, A., Balbinot, E., et al. 2018, arXiv:1801.03097

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

Slater, C. T., Bell, E. F., Schlafly, E. F., et al. 2013, ApJ, 762, 6

Sohn, S. T., van der Marel, R. P., Kallivayalil, N., et al. 2016, ApJ, 833, 235

The Dark Energy Survey Collaboration. 2005, arXiv:astro-ph/0510346

—. 2016, MNRAS, 460, 1270

Valcheva, A. T., Ovcharov, E. P., Lalova, A. D., et al. 2015, MNRAS, 446, 730

Vera-Ciro, C., & Helmi, A. 2013, ApJ, 773, L4

Vivas, A. K., Zinn, R., Farmer, J., Duffau, S., & Ping, Y. 2016, ApJ, 831, 165

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

Weiss, J., Newberg, H. J., & Desell, T. submitteda, ApJ. https://arxiv.org/abs/1807.03754

Weiss, J., Newberg, H. J., Newby, M., & Desell, T. submittedb, ApJS

Xu, Y., Newberg, H. J., Carlin, J. L., et al. 2015, ApJ, 801, 105

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

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

APPENDIX A
Example parameter file: stripe 19

```
number_parameters: 4
background_weight: 0.99
background_weight_step: 0.000001
background_weight_min: 0.8
background_weight_max: 1.0
optimize_background_weight: 1
background_parameters[4]: 1.0, 0.52, 5.99, 1.0
background_step[4]: 0.02, 0.000004, 0.00008, 0.02
background_min[4]: 0.0, 0.25, 2.00, 0.1
background_max[4]: 3.0, 1.75, 20.24, 3.0
optimize_parameter[4]: 0, 1, 0, 0
number_streams: 3, 5
stream_weight: -1.9
stream_weight_step: 0.000001
stream_weight_min: -20.0
stream_weight_max: 20.0
optimize_weight: 1
stream_parameters[5]: 151.0, 23.0, 2.4, 3.0, 0.9
stream_step[5]: 0.00003, 0.00004, 0.00006, 0.00004, 0.000004
stream_min[5]: 135.0, 2.0, 0.0, 0.0, 0.1
stream_max[5]: 240.0, 100.0, 3.14, 3.14, 25.0
optimize_parameter[5]: 1, 1, 1, 1, 1
stream_weight: -0.95
stream_weight_step: 0.000001
stream_weight_min: -20.0
stream_weight_max: 20.0
optimize_weight: 1
stream_parameters[5]: 179.0, 32.0, 2.4, 3.0, 10.0
```

```
stream_step[5]: 0.00003, 0.00004, 0.00006, 0.00004, 0.000004
stream_min[5]: 135.0, 2.0, 0.00, 0.0, 0.1
stream_max[5]: 240.0, 100.0, 3.14, 3.14, 25.0
optimize_parameter[5]: 1, 1, 1, 1, 1
stream_weight: -3
stream_weight_step: 0.000001
stream_weight_min: -20.0
stream_weight_max: 20.0
optimize_weight: 1
stream_parameters[5]: 190.0, 27.0, 1.9, 3.0, 3.0
stream_step[5]: 0.00003, 0.00004, 0.00006, 0.00004, 0.000004
stream_min[5]: 135.0, 2.0, 0.0, 0.0, 0.1
stream_max[5]: 240.0, 100.0, 3.14, 3.14, 25.0
optimize_parameter[5]: 1, 1, 1, 1, 1
convolve: 120
sgr_coordinates: 0
wedge: 18
r[min,max,steps]: 16.0, 22.5, 700
mu[min,max,steps]: 135, 240, 800
nu[min,max,steps]: -1.25, 1.25, 320
number_cuts: 1
r_cut[min,max,steps][3]: 16.0, 22.5, 700
mu_cut[min,max,steps][3]: 197.0, 198.0, 8
nu_cut[min,max,steps][3]: -1.25, -1.0, 288
```