1 INTRODUCTION

This two-year multi-institutional project aims to capitalize on results produced during the PI's NSF-supported project "Towards a Panoramic 7-D Map of the Milky Way", which is in its final third year. The work proposed here has two main goals:

1. Combine recent observational results on mapping of the spatial, kinematic and metallicity distributions of the Milky Way stars with fast advances in cosmologically motivated N body models to critically compare data and theory: Do model galaxies look similar to the observed Milky Way? If so, can we use models to learn about quantities that are not readily observed, such as age distributions and gravitational potential? Can we use models to decipher complex correlations among various observables uncovered in surveys such as Sloan Digital Sky Survey (SDSS)?

2. Utilize a major new database of time-resolv- ed photometric observations based on LINEAR survey to data-mine and study variable objects, most notably halo RR Lyrae stars, and to produce samples of most "interesting" objects for follow up observations with small (~1m) telescopes. These time-resolved photometric observations will serve a dual purpose: they will provide more accurate characterization of various variable populations required for scientific analysis, and they will engage a large number of undergraduate students in publication-quality hands-on research.

The proposal first describes the overall motivation for studying the Milky Way and then summarizes results from prior NSF support that are relevant for this proposal, followed by a detailed description of the proposed work.

2 THE MILKY WAY AS THE ROSETTA STONE FOR GALAXY FORMATION

The current cosmological paradigm states that the Universe had its beginning in the Big Bang. Galaxies, the fundamental building blocks of the Universe, formed soon after the Big Bang. A major objective of modern astrophysics is to understand when and how galaxies formed, and how they have evolved since then. Our own galaxy, the Milky Way, provides a unique opportunity to study a galaxy in great detail by measuring and analyzing the properties of a large number of individual stars. *The proposed research will test and influence theories of galaxy formation and evolution by comparing our recently improved understanding of the distribution and kinematics of stars in in the Milky Way with the predictions of state-of-the-art N body models.*

The formation of galaxies like the Milky Way was long thought to be a steady process that created a smooth distribution of stars, with this standard view exemplified by the Bahcall & Soneira (1980) and Gilmore, Wyse & Kuijken (1989) models, and described in detail by e.g. Majewski (1993). In these smooth models, the Milky Way is usually modeled by three discrete components described by relatively simple analytic expressions: the thin disk, the thick disk, and the halo. Instead, recent discoveries of complex substructure in the distribution of the Milky Way's stars (e.g. Ivezić et al. 2000; Yanny et al. 2000; Vivas et al. 2001, 2006; Newberg et al. 2002; Majewski et al. 2003; Jurić et al. 2006; Belokurov et al. 2006; Grillmair 2006) have deeply shaken this standard view. Unlike those smooth models that involve simple components, the new data indicate much more irregular structures, such as the Sgr dwarf tidal stream in the halo and the Monoceros stream closer to the Galactic plane. These recent developments, based on accurate large-area surveys, have made it abundantly clear that the Milky Way is a complex and dynamical structure that is still being shaped by the infall (merging) of neighboring smaller galaxies.

Numerical simulations suggest that this merger process plays a crucial role in setting the structure and motions of stars within galaxies, and is a generic feature of current cosmological models. Since the individual stars that make up the stellar populations in the Milky Way can be studied in great detail, their characterization will provide *clues about galaxy merging process that cannot be extracted from observations of distant galaxies.*

Most studies of the Milky Way can be described as investigations of the stellar distribution in the seven-dimensional space spanned by the three spatial coordinates, three velocity components, and metallicity. Depending on the quality, diversity and quantity of data, such studies typically concentrate on only a limited region of this seven-dimensional (7-D) space (e.g. the nearby solar neighborhood, pencil beam surveys, kinematically biased surveys), or consider only marginal distributions (e.g. number density of stars irrespective of their metallicity or kinematics, proper motion surveys without metallicity or radial velocity information). The ongoing work, in its third final year of NSF support, has produced a series of publications which quantified this multi-dimensional distribution to unprecedented detail (described below). These breakthrough results in the understanding of the Milky Way from ours and numerous other observational studies were enabled by the advent of large digital sky surveys. At the same time, the observational progress was matched by rapid improvements in simulations - the state-of-the-art N-body codes are approaching sufficient fidelity to aid interpretation of these observations and enhance our understanding of the relevant physics. *We propose here to compare recent observational results to mock catalogs produced by N-body models, and to answer a simple but far-reaching question: Do state-of-the-art N-body codes produce spatial, kinematic and metallicity distributions in agreement with recent SDSS and other data sets?*

As summarized below, thanks to recent SDSS observations of several tens of millions of main sequence stars, we now have a significantly improved, detailed and robust, observational understanding of the Milky Way stellar distributions within about 20 kpc. However, as suggested by RR Lyrae stars from the small area covered by the so-called SDSS Stripe 82 region (300 sq.deg, where SDSS obtained time-domain data, with a distance limit of 100 kpc for RR Lyrae stars), models developed using main sequence stars within 20 kpc *cannot be extrapolated beyond about 20-30 kpc* (Sesar et al. 2009). In order to map the Galaxy beyond this distance limit with main sequence stars, a much deeper survey than SDSS is required. While there are several such surveys being built (Dark Energy Survey, Pan-STARRS, LSST), it will be years before the required data are collected. Meanwhile, the 20-40 kpc distance range can be efficiently probed with RR Lyrae stars that can be extracted from the LINEAR survey data. As we describe below, this sample includes about 4,000 RR Lyrae stars over 10,000 sq.deg. of sky, with currently unmatched combination of sky area, sample size and distance limit. With RR Lyrae stars as the leading scientific application, we propose to data-mine the LINEAR database for appropriate targets to be followed up with 1m photometric telescopes. *This effort will not only yield cutting-edge science results, but will engage a large number of undergraduate students in publication-quality hands-on research.*

We first describe the proposed comparison of the Milky Way observations with N-body models, and then discuss the proposed analysis of the LINEAR survey data, together with the photometric follow up plans. The project organization and management are described in the last section.

3 COMPARISON OF OBSERVATIONS AND N-BODY MODELS

The data obtained by the recent large-area sky surveys enable detailed unprecedented studies of the stellar distribution in the 7-D space from solar neighborhood all the way to the outer halo. However, while these results represent exciting observational breakthroughs, their interpretation is not at all simple. For example, the stellar kinematics are expected to be a strong function of the stellar age, which is a property not easily constrained by observations. Furthermore, observations reveal a number of results that are not easily understood within classical paradigms, such as the absence of a correlation between metallicity and rotational velocity for disk stars; such a correlation is predicted by the standard decomposition of disk stars into thin and thick components. Hence, to fully exploit the recent observational results, a detailed comparison with sophisticated numerical models is mandatory.

Driven by technology, the numerical N-body simulations are improving steadily. Modern simulations have sufficient resolution and physical detail to study the formation of stellar disks and spheroids over a large baseline of masses and cosmic ages. In a collaboration with the University of Washington "N-body shop", we plan to utilize state-of-the-art simulations developed by Fabio Governato, Tom Quinn and collaborators, that include in a consistent manner a simple, yet physically motivated description of energy feedback from supernovae and supermassive black holes. Initial comparison of various observed maps discussed here and model predictions is exceedingly encouraging, as detailed below. Such models will be used to study dominant physical effects that control the stellar distribution, metallicity and kinematics, and how their influence evolves with time. This direct and quantitative comparison of data and models meshes well with reseach programs led by Governato and Quinn, as testified to in the attached letters.

We first summarize the main observational results that models will be compared to, and then show a few results obtained as part of the ongoing research that demonstrate feasibility of the proposed analysis.

3.1 Stellar Number Density Maps

A common methodology employed in a series of recent papers that resulted from the ongoing grant is the use of photometric parallax and photometric metallicity relations. While such approach has been attempted before, the SDSS produced the first dataset which extends over a large fraction of the sky (one quarter), with sufficiently accurate photometry to obtain distance accuracy of 10-20% and metallicity accuracy of 0.1-0.2 dex (enabled by the u band data) for millions of stars.

Using photometric data for 50 million stars from SDSS Data Release 4, Jurić et al. 2008, hereafter J08) have constructed 3-dimensional maps (data cubes) of the stellar number density distribution for 19 narrow color bins that span spectral types from mid-F to early M. As the bin color is varied from the reddest to the bluest, the subsamples cover distances ranging from 100 pc to 20 kpc. Distance to each star is estimated using a maximum likelihood implementation of photometric parallax method, and stars are binned and counted in small 3-dimensional pixels whose size depends on dynamical range provided by each color bin and Poisson noise limits (typically there are 250,000 pixels per map). An example of a 2-dimensional projection of the resulting maps is shown in Fig. 1 (left panel).

These maps are a powerful tool for studying the Milky Way's stellar number density distribution. Traditional methods for modeling stellar counts in the magnitude-color space need to adopt a large number of poorly known functions such as the initial mass function, the mass-luminosity relationship, the luminosity function, and geometric description of the postulated components such as disks, bulge and halo. Instead, with these number density maps the Milky Way's structure can be studied without any a priori assumptions about its components. *With these maps analysis of the Milky Way's structure is now akin to studies of external galaxies.*

The description of these maps is not a trivial task because of the rich substructure. While halo substructure has been known for a while (e.g. Ivezić et al. 2000, Yanny et al. 2000, Vivas et al. 2001, Majewski et al. 2003, and references therein), these new maps demonstrate that the thin and thick disk substructure is equally complex. Nevertheless, the gross behavior can be captured by assuming "standard" Galaxy models based on two exponential disks and an oblate power-law halo. J08 determined the best-fit parameter values for full two-dimensional smooth models and further refined them using residual minimization algorithms. The complex Galaxy substructure becomes readily discernible in residual maps obtained by subtracting the best-fit smooth Galaxy models from the data, as shown in Fig. 1 (middle and right panels). When using different color bins, the clumps discernible in residual maps are detected in different apparent magnitude ranges, but at the *same geometric positions.* This consistency strongly argues that the overdensities are not artefacts of the adopted photometric parallax relation.

In summary, J08 analysis produced the following key findings:

1. The data confirmed prior evidence for a Galaxy consisting of a halo and an exponential disk component; however, these smooth compo-

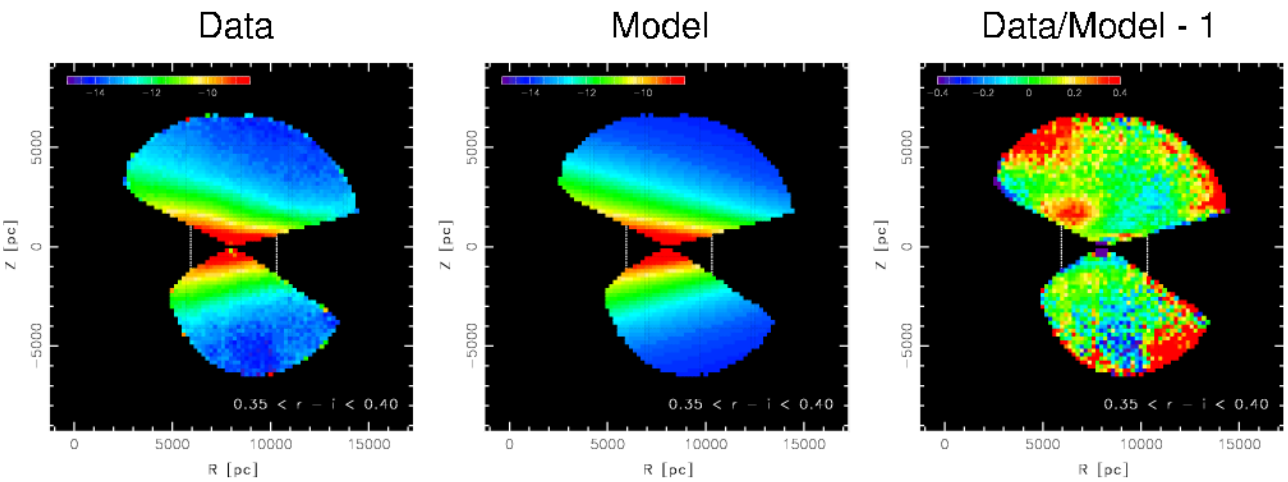

Figure 1: Local overdensities in the number of main sequence stars revealed after subtracting large-scale background models. The left panel shows the measured stellar number density as a function of Galactic cylindrical coordinates for stars with 0.35< $r-i$ <0.40. The middle panel shows the best-fit smooth model taken from J08, and the right panel the normalized (data-model) difference. The large 10-30% overdensities visible in the right panel are hard to identify in the observed maps (left panel), but are easy to discern in difference maps.

nents are also punctuated by numerous localized overdensities (the substructure).

2. The substructure, by now well known in the halo, permeates the thick disk as well. While consistent with merger remnants, the origin of the identified substructure could not be determined with certainty.

3. All parameters (scales, normalizations, flattening, and the power law index) of the Galactic disk and halo are determined using a dataset 400 times larger than in any previously published analysis, and an iterative fitting method that accounts for substructure. The large sky area breaks all degeneracies inherent in the multi-parameter model, and the distributions of disk and halo stars are now known better than ever.

4. A very large, diffuse, substructure in the halo, covering over 1000 deg^2 of the sky at distances of 6-15 kpc is seen towards the Virgo constellation and dubbed the Virgo overdensity.

3.2 Stellar Metallicity Distribution and Kinematics

Encouraged by the demonstrated feasibility of direct mapping as the method of choice for analysis of large-scale photometric studies, (Ivezić et al. (2008, hereafter I08) used it to obtain an unbiased, three-dimensional, volume-complete metallicity distribution of ∼2.5 million F/G stars at heliocentric distances of up to ∼8 kpc. SDSS spectroscopic metallicity was used to calibrate a photometric metallicity indicator based on the $u-g$ and $g-r$ colors, and an explicit metallicity dependence term was added to the photometric parallax relation. This study also had a kinematic component, with velocities deduced from SDSS-POSS proper motion (Munn et al. 2004). The main conclusions from I08 study are

1. The metallicity distribution function (MDF) of the halo and its kinematics are clearly distinct and separate from those of the disk (see Figure 2).

2. The median metallicity of the disk exhibits a clear vertical (with respect to the Galactic plane; Z) gradient. The MDF of the disk at Z > 0.5 kpc is consistent with no gradient in the radial direction (6 < R/kpc < 10). The spatial variation of the median metallicity does not follow the distribution of stellar number density (see Figure 3).

3. Disk stars show a rotational velocity gradient with distance Z from the Galactic plane. However, there is no correlation between the

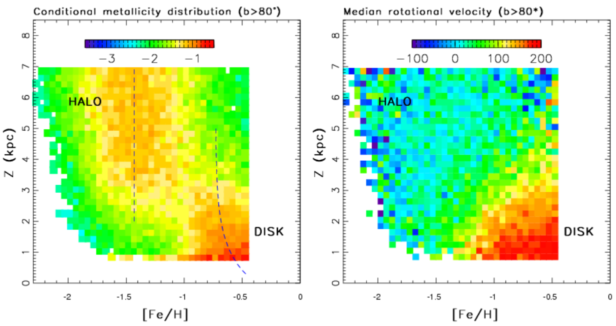

Figure 2: The left panel shows the full metallicity distribution, as a function of the distance from the Galactic plane, for about 60,000 stars within 10 degrees from the North Galactic Pole (I08). The distribution is displayed on a logarithmic scale and color-coded as shown in the inset. Two distinct Galaxy components, halo and disk, are evident. High-metallicity disk stars dominate close to the plane, while low-metallicity halo stars dominate beyond 3 kpc from the plane. The dashed lines mark the median metallicity for each component. The median metallicity for disk stars shows a gradient, while halo stars have spatially invariant metallicity distribution. These two components with distinct metallicity distributions also have different kinematics: the right panel shows the median rotational velocity component for the same stars as in the left panel. The velocity is determined from displacements of stars on the sky over half a century that lapsed between the Palomar Observatory Sky Survey (POSS) in the 1950s and SDSS. The high-metallicity disk stars have large rotational velocity (about 200 km/s, see the inset), while the low-metallicity halo stars display behavior consistent with no net rotation. The rotational velocity for disk stars decreases with the distance from the Galactic plane, while it is constant for halo stars, similarly to the behavior of their metallicity distributions.

metallicity and rotational velocity of stars at $Z > 0.4$kpc, in conflict with traditional thin/thick disk decomposition.

4. The Monoceros stream is seen as a structure with different metallicity distribution and kinematics than its surroundings.

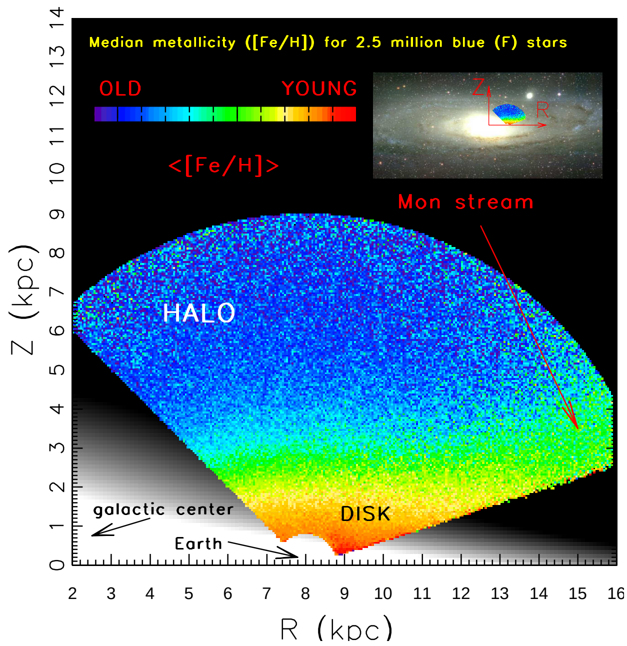

Figure 3: The median metallicity map for 2.5 million main-sequence F-type stars within 10 kpc from the Sun (I08). The metallicity is estimated using $u-g$ and $g-r$ colors measured by SDSS, and linearly color-coded according to the legend from $[Fe/H] = -1.5$ (blue) to $[Fe/H] = -0.5$ (red). The position and size of the mapped region, relative to the rest of the Milky Way, is illustrated in the top right corner, where the same map is scaled and overlaid on an image of the Andromeda galaxy. The gray scale background shows the stellar number density generated using the J08 best-fit model. The gradient of the median metallicity is essentially parallel to the Z axis, except in the Monoceros stream region, as marked.

These results are encapsulated in a new Galaxy modeling code *Galfast* developed in collaboration with Mario Jurić as a part of the ongoing grant, and made available to the community at www.mwscience.net/galfast. The mock catalogs produced by *Galfast* can be now compared to mock catalogs produced by N-body codes using *identical processing and visualization algorithms*. The next section describes some of the preliminary analysis which demonstrates that N-body models show behavior that is very similar to the observed one.

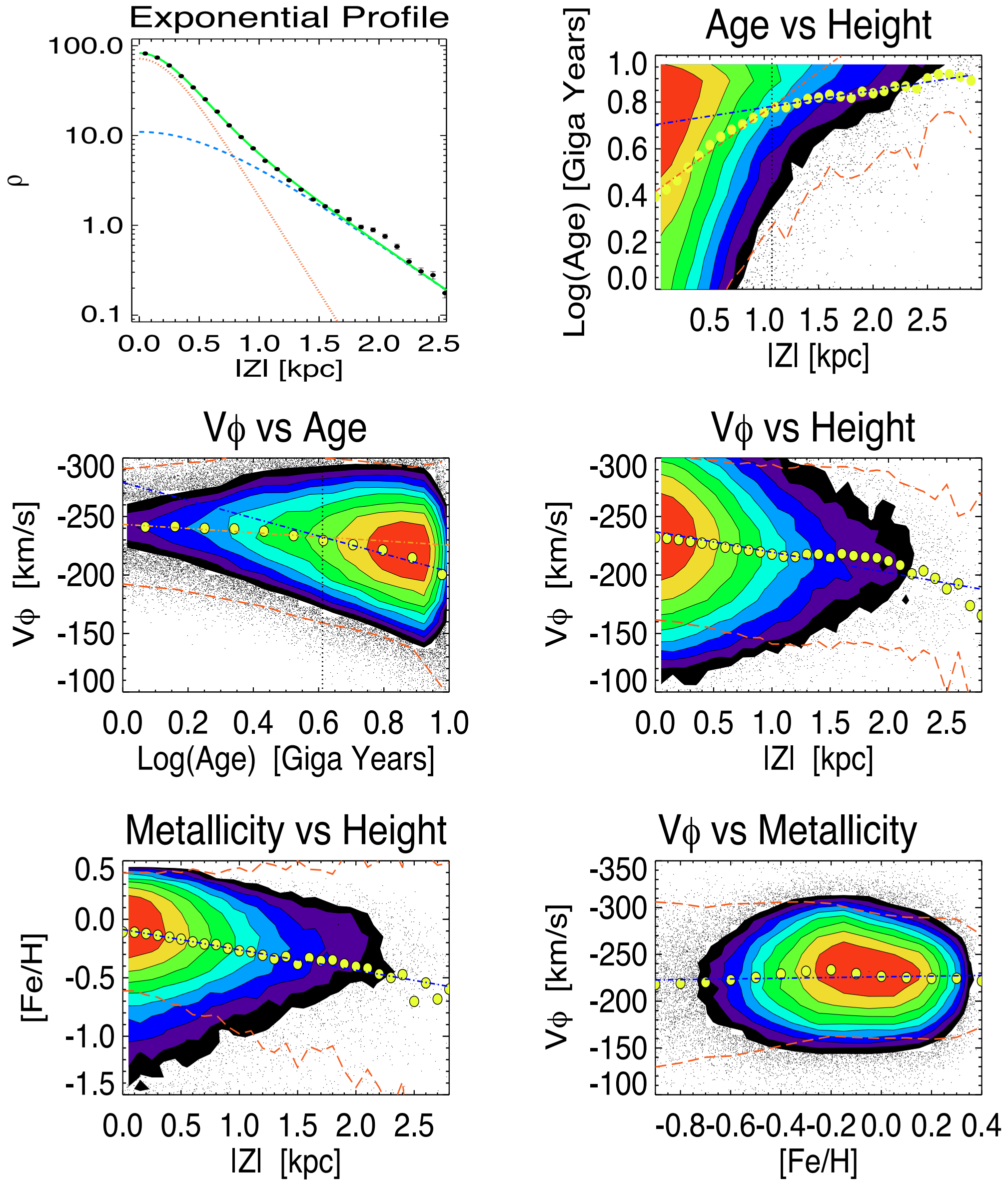

Figure 4: Analysis of an N-body model (Roškar et al. 2008) using the same methodology as applied to SDSS data, with additional analysis of stellar age distribution which is inaccessible to SDSS observations. Model particles are selected from a galactocentric cylindrical annulus with 7 kpc $< R <$ 9 kpc. The histogram in the top left panel showsw counts as a function of distance from the Galactic plane, Z. The two solid straight lines show the best-fit thin and thick exponential density profiles, and the vertical dashed line marks the transition. The age distribution as a function of Z is shown in the top right panel as color-coded contours (low to medium to high: black to green to red) in the regions of high density of points, and as individual points otherwise. The large symbols show the median values in Z bins, and the dashed lines show a 2σ envelope around the medians. The dot-dashed line shows the best linear fit to these medians. The remaining four panels are analogous, except that they show the rotational velocity vs. age (middle left), the rotational velocity vs. Z (middle right), the metallicity vs. Z (bottom left) and the rotational velocity vs. metallicity (bottom right) diagrams. In the last panel, only a subset of data from a thin slice in Z centered at $\sim$1 kpc is shown. Note the absence of a correlation between the velocity and metallicity, in agreement with observations.

3.3 A Preliminary Comparison of Observations and N-body Models

In a study supported by the ongoing grant, Loebman et al. (2008, hereafter L08) have compared SDSS results to an N–body + Smooth Particle Hydrodynamics (SPH) simulation designed to mimic the quiescent formation and evolution of a Milky Way–type galactic disk following the last major merger (Roškar et al. 2008, hereafter R08). The system consists of a rotating, pressure–supported gas halo embedded in an NFW dark matter halo. As the simulation proceeds, the gas cools and collapses to the center of the halo, forming a thin disk from the inside–out. When the gas reaches densities and temperatures conducive to star formation, the sub–grid star formation and stellar feedback recipes are initiated. The stellar feedback prescriptions include SN II, SN Ia and AGB metal production, as well as injection of supernova energy which impacts the hydrodynamic properties of the disk ISM. Importantly, no *a priori* assumptions about the disk's structure are made – its growth and the subsequent evolution of its stellar populations are completely spontaneous and governed only by hydrodynamics, stellar feedback, and gravity. For example, the stellar radial migration resulting from the interactions of stars with transient spiral arms, such as recently explicitly modeled by Schoenrich & Binney (2008), is naturally incorporated in the model.

Loebman et al. model analysis is illustrated in Fig. 4. In all the important features, models at least qualitatively, and often quantitatively, agree with the observational results. For example, the model distribution of stars as a function of Z (top left panel) resembles a sum of two exponential profiles, with the "break" height of $Z \sim 1$ kpc, in agreement with SDSS results (J08). Since models now nothing about "thick" disk hypothesis, this agreement is remarkable and suggests that "thick" disk component can be produced without major mergers. Equally encouraging is the lack of correlation between velocity and metallicity at $Z \sim 1$ kpc (bottom right panel), as found in the SDSS data, but in conflict with simplistic models for "thin" and "thick" disk components. On the other hand, and again in agreement with observations, both velocity and metallicity show a correlation with Z (middle right and bottom left panels). Our initial analysis indicates that the naively expected correlation between these two quantities gets erased by radial migration of stars and evolution of kinematic behavior with age.

In summary, the models of R08 show remarkable similarity to SDSS observations of the distribution of Milky Way stars in the position–velocity–metallicity space. In addition, these models provide a quantity that SDSS observations cannot – the stellar age. We propose to

1. Extend this analysis from a limited volume analyzed in L08 to the full Galaxy volume explored by SDSS.

2. In addition to the low-order moments (median and dispersion), compare the full probability distributions for metallicity and velocity components.

3. Further explore the distribution of stellar age and its correlations to various observables.

4. Repeat the entire analysis for models that have different prescriptions for stellar feedback, and have different initial conditions.

Most of the proposed analysis is by now fully automated and new models can be expeditiously processed.

3.4 Determining Gravitation Potential Via Jeans Equations

The second subproject based on comparison of observations and N-body models will investigate how well gravitational potential can be determined from the observed kinematics and spatial distribution of halo and disk stars. Together with the first subproject described above, it will form the bulk of PhD thesis for the supported graduate student.

If we could measure acceleration of stars, we could easily map gravitational potential via Newton's second law. Since we still cannot measure acceleration except for a few stars around the black hole in the center of the Galaxy, we need to use statistical approach and velocity measurements for a large number of stars. The mathematics of this method are given by Jeans equations which imply that when the spatial distribution of a relaxed population of stars and the spatial variation of their velocity components are known, one can deduce the spatial variation of the gravitational potential in which they move. This is an exciting possibility because the resulting potential would include the contribution from the dark matter, in turn giving us *the most detailed measurement of a dark halo structure.*

Until recently this approach was limited to local (solar neighborhood) datasets due to observational constraints. Its success was limited because the local nature of the data required more or less ad hoc assumptions about the spatial gradients of kinematic quantities. To reiterate, with the recent SDSS datasets these gradients are now measured out to distances of 10 kpc, and separately for disk and halo stars. Given the vastly different spatial and kinematic distributions of these two populations (which are easily separated by metallicity, see the left panel in Fig. 2), an important and timely question is whether they imply mutually consistent gravitational potentials. However, for the interpretation of the likely observational difference in the implied potentials, systematic errors due to sampling effects and various assumptions underlying the use of Jeans' equations need to be quantified with the aid of simulations. Using the same well-understood simulations from the first project, we will subject N-body mock catalogs to exactly the same processing pipeline as the SDSS catalogs and obtain the first robust measurement of the gravitational potential via Jeans method, together with estimates for both random and systematic errors.

We expect these projects to become templates for future investigations of both new survey data and new N-body models.

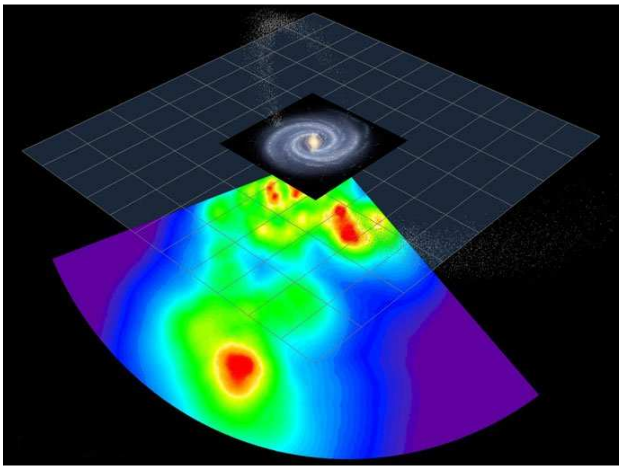

Figure 5: A visualization of the halo substructure traced by 500 RR Lyrae stars discovered using the SDSS Stripe 82 data (300 sq.deg. with about 50 repeated scans of the sky). The local number density is multiplied by the cube of the galactocentric radius and shown on a linear scale (increasing from magenta to blue to green to yellow to red). If the halo were smooth and following $1/r^3$ density law, the entire map would be green. The large red clump in the bottom left corner is the Pisces overdensity at a galactocentric distance of 80 kpc. The red elongated clump in the middle is the intersection with the Sgr tidal stream at 30 kpc. An artist conception of the Galactic plane (based on the Spitzer data) is shown for scale. The barely visible white dots show a model for the Sagittarius dSph and its tidal streams by Law, Johnston & Majewski (2005).

4 RR LYRAE STARS AS A TOOL FOR EXPLORING HALO STRUCTURE

RR Lyrae stars are more luminous than main sequence stars and thus enable exploration of the Milky Way structure to large distances; with SDSS data RR Lyrae are detected to about 100 kpc, compared to ∼20 kpc for main sequence stars. While red giants can also be traced to large distances (to almost 100 kpc with 2MASS data, e.g. Majewski et al. 200X), RR Lyrae stars are more numerous (per unit volume) by almost an order of magnitude and have more accurate distances. In a recent study, we have discovered complex halo substructure traced by ∼500 RR Lyrae stars from a 300 sq.deg. large area covered by the so-called SDSS Stripe 82 region (Sesar et al. 2009). This sample is currently the deepest available and, within its 5–100 kpc distance range, there are numerous well-defined clumps, including interesection with the Sgr tidal stream and a new, as of yet not fully understood, substructure dubbed the Pisces overdensity at 80 kpc (see Fig. 5). The Pisces clump has been recently shown to be kinematically cold thereby proving that it is a real coherent structure (Kollmeier et al. 2009). Equally puzzling is the evidence for a change in the slope of the power law that describes the distribution of halo RR Lyrae stars around 20–30 kpc from the Galactic center. This result is supported by the coadded stripe 82 observations of main sequence stars (distance limit of ∼30 kpc) and implies that the best-fit model for halo stellar number density distribution obtained by J08 (using main sequence stars within a distance limit of ∼20 kpc) cannot be extrapolated beyond 30 kpc. This conclusion agrees with recent work based on entirely different observations (e.g., a kinematic study by Carollo et al. 2007). Such a change of halo density profile is strong evidence in favor of "dual halo" hypothesis advocated by Carollo et al. and its robust confirmation would have a fundamental impact on our models for the Milky Way formation and evolution. Therefore, there is strong observational motivation to explore the Galaxy to distances of at least ∼30 kpc and over a large sky area – much larger than the 300 sq.deg. area covered by stripe 82 to minimize the impact of local substructure on the measurement of overall halo structure. Such data are not available yet for main sequence stars (a multi-band survey to $V \sim 23$ over several thousand sq.deg.).

The second best probe are RR Lyrae stars, which could be discovered using a time-domain survey with a limiting depth of $V \sim 18$, compared to $V \sim 23$ for main sequence stars. Fortunately, such a time-domain survey exists, although it was developed and executed for entirely different reasons.

4.1 Exploration of the Halo Substructure with RR Lyrae Stars from the LINEAR Survey

Studies of the halo structure with RR Lyrae stars can be simultaneously extended to a larger area and to a larger distance limit, with a larger sample than in any published work, with about 4,000 RR Lyrae stars which we discovered using LINEAR survey.

LINEAR has been operated by the MIT Lincoln Laboratory since 1998 to search for asteroids. The focus of the program is to discover and track near-Earth asteroids ("killer asteroids) larger than 1 km diameter. Its archive now contains approximately 6 million images of the sky most of which are 5 megapixel images covering 2 sq.deg. The LINEAR image archive contains a unique combination of coverage depth ($V \sim 18$), coverage area (about 10,000 sq. deg. of overlap with the SDSS survey area), and cadence (several hundred observations repeated every few days) that is potentially useful in exploring a number of time-varying astronomical phenomena.

A shortcoming of LINEAR data is that its photometric calibration is fairly inaccurate because the effort was focused on astrometric observations of asteroids, and because its photometric bandpass is very broad (essentially white light) and requires additional color information to accurately transform to a standard bandpass. Motivated by our desire to map RR Lyrae stars over a large sky area and to a faint distance limit, the University of Washington group, in collaboration with Scott Stuart from the LINEAR project, has recalibrated LINEAR photometry using the Sloan Digital Sky Survey data (in essence, most SDSS stars brighter than $V = 18$, or $r \sim 18$, act as photometric calibration stars). Scott Stuart, a member of LINEAR team, employed SExtractor (Bertin & Arnouts 1996) to extract and measure objects, and then the UW group reused tools developed for recalibrating Palomar Observatory Sky Survey data as described in Sesar et al. (2006). We obtained photometric repeatability of 0.03 mag (root-mean-square scatter) for LINEAR sources not limited by photon statistics, with fairly Gaussian behavior.

As a result of this effort, we now have a database with about 5 billion LINEAR photometric measurements for over 20 million objects (mostly stars), with the mean number of observations per object of about 250. This database is the largest such resource currently available, and will remain so for at least the next few years, if not for a decade (especially when considering

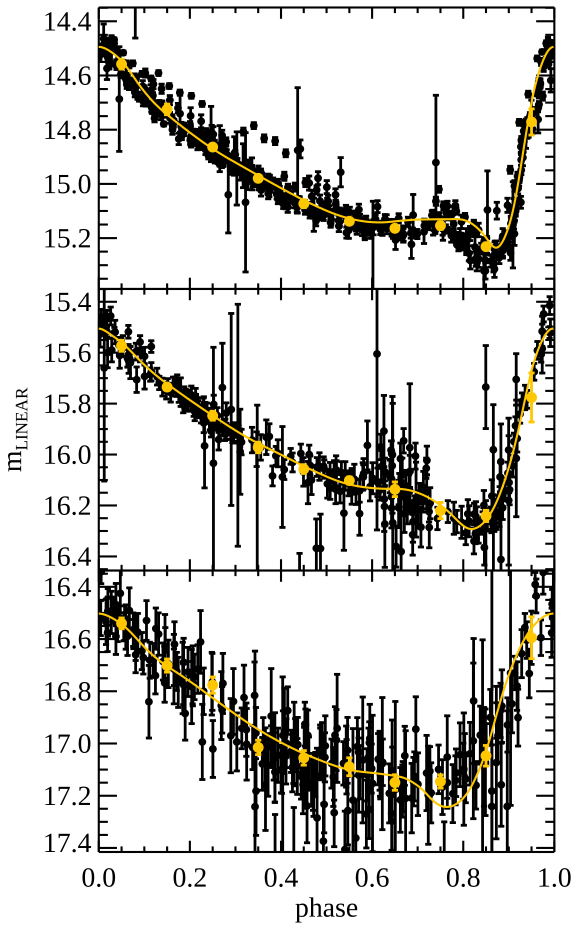

Figure 6: Examples of LINEAR data for RR Lyrae stars (small dots are individual observations and large symbols are medians for phase bins) and best-fit templates in three characteristic magnitude ranges.

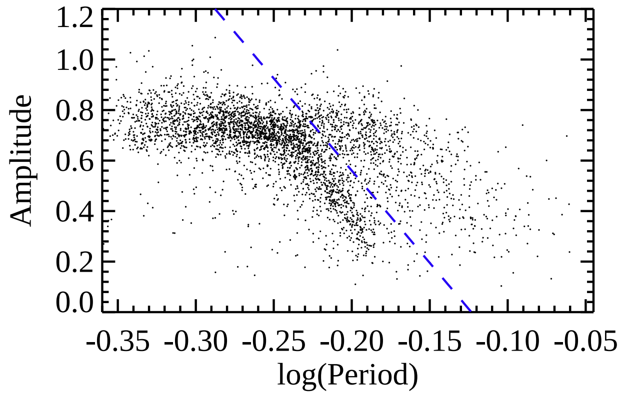

Figure 7: The period-amplitude (Bailey) diagram for about 3,000 ab type RR Lyrae stars selected from the LINEAR database. The main locus of points corresponds to Oosterhof type I stars, and the cloud seen to the right from the dashed lineare Oosterhof type II stars. The two types have different spatial distributions, and probably different metallicity distributions (Miceli et al. 2008; Sesar et al. 2009). The clear separation of the two Oosterhof classes in this diagram demonstrates the accuracy of periods and amplitudes based on template fits to LINEAR light curves.

the number of observations per source). Its main value is the potential for discovery and characterization of several hundred thousand new variable objects (based on variable object statistics from Sesar et al. 2007). The probed magnitude range ($14 < V < 18$) meshes well with another large-area survey, the Northern Sky Variability Survey, ($V < 15$), developed by collaborator Przemek Wozniak.

As a pilot project to investigate the quality of recalibrated LINEAR database, we have employed variability-based selection criteria and light curve template fitting tools from Sesar et al. (2009, see Fig. 6) and selected about 4,000 RR Lyrae stars over 10,000 sq.deg. of sky. This sample has an unmatched combination of sky area, sample size and distance limit ($\sim$ 30 kpc). The template fitting method performs robustly to $V \sim 18$ as testified by easily discernible Oosterhof dichotomy seen in the Bailey diagram (Fig. 7). In addition to providing a statistically significant sample to study overall halo structure, several localized overdensities are detected (Fig. 8). Encouragingly, the closest clumps of LINEAR RR Lyrae correspond to known substructures traced by SDSS main sequence stars (e.g., the Virgo overdensity).

While detailed analysis of this sample is still in progress, two important conclusions are already apparent. First, this sample of 4,000 RR Lyrae stars is only a small fraction of several hundred thousand variable stars present in the database. Therefore, the utility of this resource extends far beyond our immediate scientific program and its public release would have a strong impact on the community. Second, while LINEAR data are

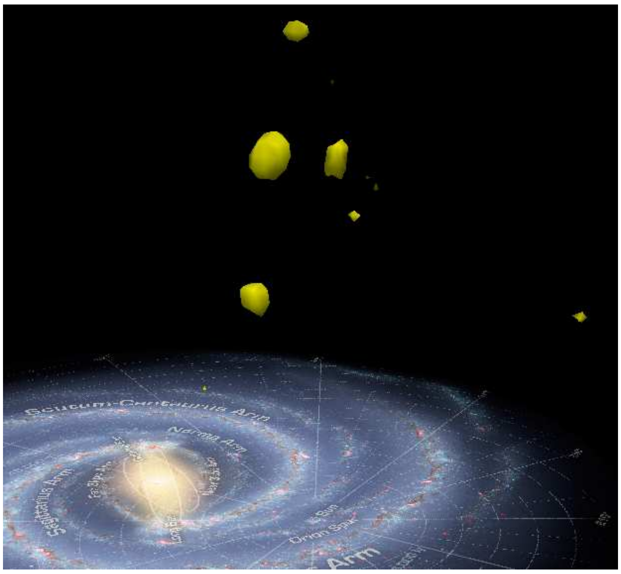

Figure 8: A three-dimensional rendering of halo overdensities within 25 kpc from the Sun traced by RR Lyrae stars discovered using the LINEAR survey. An artist conception of the Milky Way disk (motivated by the Spitzer data) is added for scale. The two closest clumps can be associated with known structures traced by SDSS main sequence stars.

great for robustly selecting RR Lyrae stars with large amplitudes (ab type), there is some contamination of the low-amplitude c type subset by eclipsing binary stars (Fig. 9). This contamination could be minimized with additional time-resolved color observations. Similarly, photometric metallicity estimates could be derived from light curves for ab type stars associated with localized substructures with followup photometric data obtained around minimum light. We describe our proposals motivated by these conclusions in the next two sections.

4.2 A Proposal for Public LINEAR Photometric Database

A photometric database with about 5 billion photometric measurements for over 20 million objects would be a major resource for studying faint optical sky. Even if the upcoming new surveys started tomorrow, it would take them years to collect 250 observations of each source (the mean

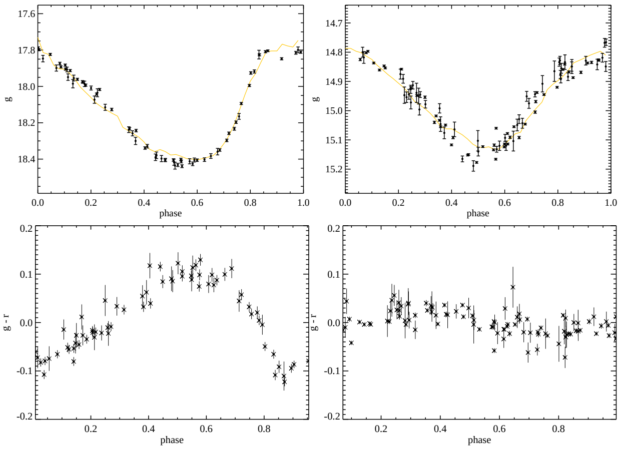

Figure 9: Example of a case where photometric follow up can improve the classification of a variable source selected from the LINEAR database. The g (*top*) and $g-r$ (*bottom*) show SDSS light curves of a RRc star (*left*) and an eclipsing binary (*right*). Even though their g band light curves are quite similar, and their periods are in the $0.2 < P < 0.43$ days range typical for RRc stars, their $g-r$ light curves are quite different. Since the LINEAR has only single-band data, a few additional color measurements would reliably distinguish between these two cases.

number for LINEAR survey). Since the sample fully overlaps the SDSS survey area, *the LINEAR photometric database provides time-domain information for all SDSS sources with* $14 < r < 18$, or roughly the brightest half of the SDSS magnitude range.

To make such a fairly large dataset publicly available, together with appropriate search tools and supporting documentation, is not trivial. Fortunately, a similar distribution system, developed for the NSVS, already exists and is used by a large number of astronomers (e.g., Kinemuchi et al. 2006, more refs from Przemek?). The principal developer of the NSVS portal, Przemek Wozniak, is excited about supplementing the NSVS database with the LINEAR database and is a collaborator on this proposal. If provided input files, he can reuse all the existing tools and extend the NSVS portal with relatively minor effort (compared to starting from scratch). One of the main components of this proposal is to enable the preparation of UW LINEAR database files for ingestion by the NSVS portal by an UW postdoc, and the execution of ingestion step by P. Wozniak. The UW postdoc will also prepare all the supporting documentation about the LINEAR survey and its SDSS-based photometric recalibration (together 0.5 FTE years of postdoc effort).

Users will be enabled to search the database by position to extract light curves of their favorite objects, as well as to select a subset of sources and download LINEAR light curves for all of them. All LINEAR sources will also have linked information from SDSS and 2MASS surveys (e.g., the latter is invaluable when selecting and studying long-period variables). As a component of the quality testing for the final product, we will run our analysis tools developed for Stripe 82 data and publish a summary paper similar to Sesar et al. (2006) with the basic statistical properties of the LINEAR variable sample (e.g., the distribution of variable sources, as a function of amplitude and low-order light curve moments in the SDSS color-color and color-magnitude diagrams). This data-mining based analysis will provide users with a useful data summary that will help them quickly start their own projects, such as designing selection criteria for their own science targets.

4.3 A Photometric Follow-up of LINEAR Variables

With its several hundred thousand new variable objects, the LINEAR photometric database will have a major impact on our understanding of faint variable populations, including the studies of the Milky Way structure. Although LINEAR data is sufficient by itself to address many scientific questions, for our particular case of RR Lyrae stars follow up photometric observations can enhance the science outcome. This fact meshes well with our desire to enhance undergraduate education at the University of Washington and the Florida Institute of Technology by providing first-hand data mining and observing experience, as described below.

There are two main drivers for photometric follow up of LINEAR RR Lyrae stars. First, color observations of the low-amplitude c type subset can minimize the contamination by eclipsing binary stars. About three to four $B-V$ or similar measurements at well placed light curve phases (LINEAR light curves will enable us to fine tune the observed phases) can robustly distinguish c type RR Lyrae from eclipsing binary stars (Fig. 9). At the same time, the newly identified eclipsing binary stars will provide material for a stand-alone publication by undergraduate students. Second, metallicity distribution, both for newly discovered clumps and for Oosterhof type I vs. type II stars is an important quantity which is hard to obtain spectroscopically for a large number of stars. Fortunately, metallicity can be estimated from RR Lyrae light curves with reasonable accuracy if the phase of minimum light is accurately measured (Jurcsik & Kovacs 1996; Szczygiel et al. 2009; Sesar et al. 2009). To deliver metallicity errors of 0.1-0.2 dex for ab type stars, the phase of the minimum light needs to be measured with a precision of about 1% (about 15 minutes for a typical ab star). For the majority of LINEAR RR Lyrae stars, this seems to be the case, but we do not have yet an independent estimate of measurement errors. In summary, we need to get follow up photometric data for about a hundred RR Lyrae stars to i) quantify the contamination of the c type subset, and ii) to quantify errors in the phase of minimum light obtained from LINEAR light curves.

With about 15 minutes per star (10 minutes for open-shutter time, and about 5 minutes for acquiring the next target), 4 stars can be observed per hour. With about 4 hour long observing sequence per star, about 8 stars can be observed per night for a total of about 12 observing nights. With a 50% contigency, the required number of nights is about 18. Given the relatively bright magnitude limit ($V < 18$) and assumed exposure time (10 min), these observations can be performed with modest 1m class telescopes. We have access to several 1m class telescopes.

The University of Washington operates the Manastash Ridge Observatory (MRO) equipped with a photometric 0.8m telescope. MRO is located 100 miles East of Seattle and it is used primarily for undergraduate research projects and classes at UW. The MRO director, Chris Lows, is a collaborator on this proposal and will support observing trips by undergraduate students. Several MRO observing campaigns over the last two years have targeted 40 SDSS Stripe 82 RR Lyrae stars and produced light curves that were presented by undergraduate students at the 211 AAS meeting (Fraser et al. 2007, which was a collaboration of 9 undergraduate and 4 graduate students, supervised by Lows and Ivezić), and incorporated in journal publication (Sesar et al. 2007, 2009). Based on this experience, we are both confident that the data will have required quality, and that it will be easy to recruit undergraduate observers. The program includes four five-day long MRO observing trips, with 3-5 students per observing run.

The Florida Institute of Technology (FIT) serves as the lead institution for the 10 member universities that comprise the Southeastern Association for Research in Astronomy (SARA). The SARA institutions run a summer REU program that supports mentor-student teams to participate in observations at the SARA Observatory. We will have access to the 10% FIT's share of SARA telescope time (0.9m SARA North at Kitt Peak, and 0.6m SARA South at Cerro Tololo), and to 100% of time on FIT's 0.8m Ortega telescope sited in Florida (which is also equipped with a spectrograph). The program includes two five-day long observing trips to Kitt Peak (two students per trip) and two trips to Cerro Tololo (one student per trip), all led by Hakeem Oluseyi.

Given that good weather is more frequent at MRO during summer months, and during winter months at the Ortega telescope, together with SARA North and South facilities, we will have a full year coverage needed to optimize the observing program.

In summary, with RR Lyrae stars as the leading scientific application, we propose to datamine the LINEAR database for appropriate targets to be followed up with our 1m class photometric telescopes. *This effort will not only yield cutting-edge science results, but will engage a large number of undergraduate students in publication-quality hands-on research.*

5 EDUCATION & OUTREACH: Exploring Astronomical Frontiers with Computer Galaxies, Digital Sky Surveys and Small Telescopes

The research proposed in this work is heavily based on a new reality in observational astronomy: the availability of science-ready large complex astronomical databases. Such readily available resources are having a strong impact on areas ranging from high school education to graduate training of future astronomers. Data mining of these new datasets is particularly powerful when combined with follow up observations and analysis of sophisticated numerical models. At the same time, there is strong evidence that research experiences for undergraduates have exceedingly strong impact on students' decision to stay in the sciences (Rothman & Narum 1999). Taken together, these facts motivate our proposal to combine our scientific goals with an effort to provide meaningful research experiences for undergraduate students at the University of Washington and the Florida Institute of Technology.

The main features of the proposed program are

Engage students in observations: using target lists that in many cases will be designed by students themeselves, about a dozen undergraduate students will participate in follow up photometric observations at the MRO and SARA observatories. They will be trained to perform image reduction and photometric calibration.

Expose students to data-mining approach to astronomy: we will use SDSS SkyServer exercises and the LINEAR database to train students in statistical analysis of large datasets. For example, students will be required to download a number of lightcurves, compute amplitudes, and correlate them with SDSS and 2MASS colors. They will also participate in the selection of observing targets.

Educate students about state-of-the-art numerical models of the Universe: we will place their observing experience in a broader context by discussing the predictions of N body galaxy models and how their observations could lead to support for, or rejection of, such models.

Expand students' experience: all students will participate in the full publication process, starting with discussions of rough ideas, early data reduction and paper drafts, all the way to writing the last reply to the referee. Their activities will not cease after returning from the observing run.

Although the proposed work is limited to two institutions, it is likely that the public release of LINEAR database, and examples that will be set by our work, will motivate other teams to engage their students in similar projects.

6 PROJECT ORGANIZATION

6.1 Key Project Aims

This project will **i)** deliver a quantitative answer to the broad and far-reaching question, "Do cosmologically motivated state-of-the-art N body models reproduce the observed behavior of spatial, kinematic and metallicity distributions of the Milky Way stars'?', which will represent a PhD thesis work, **ii)** enhance observational constraints on halo structure constraints by extending distance limit of current maps using an unprecedented sample of 4,000 RR Lyrae stars from LINEAR survey, **iii)** make the LINEAR photometric database publicly available, and **iv)** engage about a dozen undergraduate students in publication-quality hands-on research.

These deliverables will have an impact on the community that is much broader than the focus of this proposal. In particular, the expected several hundred thousand variable sources present in LINEAR dataset will enhance our understanding of the faint variable sky, and the educational observing program will serve as a template for similar follow up efforts with small telescopes. The educational impact of undergraduate training, with students drawn from minority populations, cannot be overemphasized.

6.2 Responsibilities and Schedule

The PI, Željko Ivezić, will be responsible for the overall success of the project, and will serve as the advisor for PhD thesis work (comparison of models with observations). The thesis work will be performed by the UW graduate student, Sarah Loebman, who has already made significant progress (section 3.3). She is also an expert for large databases, has extensive experience with undergraduate recruitment and mentoring, and will work with undergraduate students on designing the target lists for observations. In collaboration with Chris Lows, the director of the MRO, Sarah will supervise observing trips, and will coordinate the publication of observing results (with help from PI and the co-PIs).

The Co-Pi Andrew Becker has been instrumental in the design and construction of the UW LINEAR database. Together with Scott Stuart, Branimir Sesar and the PI, he has developed quality tests for the recalibrated photometry. He will be responsible for the database transfer to a public site, including the preparation of supporting documentation about the LINEAR survey, and will lead a publication summarizing the basic statistical properties of the LINEAR variable sample.

The Co-PI Przemek Wozniak (Los Alamos National Laboratory) will facilitate the inclusion of the LINEAR photometric database within his distribution framework developed for the Northern Sky Variability Survey. His time commitment will be 0.1 FTE during the first year and will be used to make adjustments to database schema design, modifications to database software and web interface, to load the data, and to test and deploy the public website.

The Co-PI Hakeem Oluseyi (Florida Institute of Technology, FIT) will lead the follow up effort by SARA telescopes, and will coordinate interaction between the FIT undergraduate students and the rest of the team. Dr. Oluseyi has been a mentor of student researchers for several years and is dedicated to alleviating systematic underrepresentation of minorities and women in STEM areas. Since returning to academia in 2002, he has supervised the research activities of 9 graduate students, 29 undergraduates and 2 high school students. Of these 40 students, 21 were from underrepresented groups including 15 females, 13 minorities, and 7 who were minority females. All of the students have successfully continued their career progress in STEM-related fields. Dr. Oluseyi has also served as a co-organizer, panelist, and presenter for multiple education-related conferences for organizations such as the AAPT, the National Conference of Black Physics Students, the Southern Growth Policy Board, the National Society of Black Physicists (NSBP) and Global Hands-On Universe. Therefore, we do not anticipate any difficulties with student recruitment at FIT.

To summarize, the project schedule is as follows:

Year 1: Loebman: Complete the first part of N-body vs. data comparison, including a publication which will represent the first milestone. Begin work on constraining gravitational potentials. **Wozniak and Becker:** make LINEAR database publicly available via the NSVS site, including documentation, which will represent the second milestone. **Ivezić, Oluseyi, Loebman, Becker**: prepare target lists for follow up observations, with Becker leading a paper on statistical analysis of LINEAR variables. **Oluseyi, Lows:** begin with observing runs, and supervise data analysis.

Year 2: Loebman: Complete work on constraining gravitational potentials, including publication. **Ivezić, Oluseyi, Loebman**: continue with observing runs, and supervise preparation of publications.

6.3 Results from Prior NSF Support

The work proposed here by and large derives from the PI's existing NSF grant AST-0707901 "Towards a Panoramic 7-D Map of the Milky Way". During the first two years of this grant, a series of papers have been published that essentially quantified the distribution of the Milky Way stars, as observed by SDSS, in the position-velocity-metallicity space (J08, I08, Loebman et al. 2008, Sesar, Ivezić & Jurić 2008, Sesar et al. 2009, Bond et al. 2009). Together, these papers have already collected over 200 citations. In addition to four graduate students participating in this research program, 11 undergraduate students were engaged in analysis and they are coauthors on these publications, and have presented four posters at AAS meetings. Thanks to these results, the PI has delivered over a dozen invited talks at universities and national and international meetings, and has incorporated these results in an undergraduate and a graduate class on galaxy formation and evolution (Astr323, Astr598). These results form the basis for simulations of the stellar content of the LSST survey, and the simulation code, *Galfast*, is publicly available at www.mwscience.net/galfast.

The PI for this proposal is also PI for the NSF award AST-0507529 "Interpretation of Modern Radio Surveys: Test of the Unification Paradigm". The key project aims are unification of several modern radio catalogs into a single database containing several million sources, morphological classification of the matched sources, and analysis of the properties of the resulting classes with the aid of state-of-the-art models for extragalactic radio source evolution and expansion. This project has produced six journal publications, two PhD thesis, and has engaged six undergraduate students in data analysis and publication process. This grant is expiring this year.

The PI for this proposal is also PI for the NSF award AST-0807500 "Statistical Description and Modeling of the Variability of Optical Continuum Emission from Quasars", which is beginning its second year. This project aims to use time-domain data for the exploration of quasars. The characteristics of the quasar optical continuum variability are studied as functions of luminosity, wavelength, time scale, black hole mass, and the presence of X-ray and radio emission. The results of these studies will be used to constrain the origin of quasar emission, and to inform variability-based surveys of quasars. This project will result in a PhD thesis and during its first year public catalogs with quasar light curves have been realized and two journal publications, led by a graduate student and by an undergraduate student (both female) are in preparation. Additional two undergraduate students are engaged in this research.