What did we learn about the Milky Way during the last decade, and what shall we learn using Gaia and LSST? Željko lvezić, LSST Project Scientist University of Washington

# NOAC, Bejing May 29, 2013





# Outline

- Why to study the Milky Way?
  - "near-field cosmology" (galaxy formation and evolution)
  - hey, it's our Galaxy!
  - it's fun to describe galaxies with more than a few numbers
- The progress over the past decade: 3 vignettes Global stellar maps: Leaving the Solar neighborhood!! Disk: Can we map dust? Halo: Can we map dark matter?
- The promise of Gaia and LSST: 2020 vision Gaia: extraordinary astrometry will deliver precise trigonometric distances and proper motions to V<20 LSST: multi-color time-resolved faint (V<24.5) sky map: can probe all the way to the edge of the Milky Way galaxy with (numerous) main sequence stars

# Why to study the Milky Way?

- To help answer general questions about galaxies:
- How did galaxies form?
- How did galaxies evolve?
- What do they tell us about the Universe?
- Disadvantage: a single spiral galaxy
- A major advantage: no other galaxy can be studied in such an exquisite detail as our Galaxy

For individual stars in our Galaxy we can today measure:
a) 3D positions (and luminosity)
b) 3D motions (accelerations only close to the black hole)
c) chemical composition ([Fe/H], [α/Fe], individual abund.)
d) age and mass (sometimes)
Ivezić, Beers, Jurić 2012, ARA&A, 50, 251

# Rapid progress in observations and simulations 10 years ago:



# **Big questions about the Milky Way**

- Observers' view:
- Decompose Galaxy into main components
- Describe the spatial distribution of stars
- Describe stellar kinematics as a function of position
- Describe chemical composition as a function of position and kinematics (and ultimately age)
- Theoreticans'/modelers' view:
- Can we understand/model the behavior of stars in the position-kinematics-chemistry space?
- Can we use kinematics and other data to infer dynamics?
- Are the results from the Milky Way studies consistent with extragalactic constraints?

#### Classical Decomposition of the Milky Way Components



They are a product of Milky Way formation and evolution



Wide wavelength coverage, and accurate and robust photometry

#### A Primer on Dissecting the Milky Way with SDSS

- Stars on the main stellar locus are dominated (~98%) by main sequence stars (for r > 14)
  - The position of main-sequence stars on the locus is controlled by their effective temperature/luminosity/[Fe/H], and thus can be used to estimate distance: photometric parallax method for ~100 million stars (with LSST several billion!)

Accurate u-g color enables photometric metallicity estimates for 6 million SDSS F/G stars to 10 kpc; (with LSST 200 million to 100 kpc!)



#### Photometric Distance and Photometric [Fe/H]

- Determined absolute magnitude vs. color vs. metallicity relation using globular clusters observed by SDSS (blue end), and nearby stars with trigonometric parallaxes (red end)
- The g i color of a mainsequence star constrains its absolute magnitude to within 0.1-0.2 mag (0.3 mag for unresolved binaries), assuming [Fe/H] is known

This method was known half a century ago, but never before applied to tens of millions of stars because large-scale surveys did not have the required photometric accuracy



#### Photometric Distance and Photometric [Fe/H]



u-g and g-r colors provide an estimator of metallicity, [Fe/H]

because large-scale surveys did not have the required photometric accuracy



# Dissecting the Milky Way with SDSS Panoramic view of the Milky Way, akin

- to observations of external galaxies; good support for standard Galactic models (with amazing signal-to-noise!)
- Metallicity mapping supports components inferred from number counts mapping:







R [pc]

#### 0.35 < r-i < 0.40







#### R [pc]

#### Dissecting the Milky Way with SDSS

- Panoramic view of the Milky Way, akin to observations of external galaxies; good support for standard Galactic models (with amazing signal-to-noise!)
  - Removal of obvious clumps
  - Fit to least "contaminated" bins
  - Exponential disks + halo models



#### 0.35 < r - i < 0.40



#### Dissecting the Milky Way with SDSS

 Panoramic view of the Milky Way, akin to observations of external galaxies; good support for standard Galactic models (with amazing signal-to-noise!)



#### 0.35 < r-i < 0.40







#### # [sc]

## Dissecting the Milky Way with SDSS

- Panoramic view of the Milky Way, akin to observations of external galaxies; good support for standard Galactic models (with amazing signal-to-noise!)
  - Removal of obvious clumps
  - Fit to least "contaminated" bins
  - Exponential disks + halo models

$$\rho(R,Z) = \rho_{thin} e^{-\frac{R-R_{e}}{l_{thin}}\frac{|Z+Z_{0}|}{h_{thin}}} + \rho_{thick} e^{-\frac{R-R_{e}}{l_{thick}}\frac{|Z+Z_{0}|}{h_{thick}}} + \rho_{halo} \left(\frac{R_{GC}}{\sqrt{R^{2} + (z+z_{0})^{2}/q^{2}}}\right)^{n}$$

 The merger history of the Milky Way can be deciphered by mapping the substructure (spatially, and in velocity space, as a function of chemical composition [metallicity])



#### Deviations from the Smooth Model

- Monoceros stream was discovered using stellar counts (Newberg et al. 2002)
- It is also identified as a substructure in metallicity space... LEFT
- And kinematics, too: it rotates faster than LSR by ~50 km/s
- More details: Ivezić et al. 2008 (ApJ 684, 287)
- Search for substructure using radial velocity: more effective at large distances (Schlaufman et al. 2009, Klement et al. 2009)

Li+2012 (ApJ 757, 151) Monoceros (or Lynx? at I=180,b=30) 21

# SDSS view along the Milky Way Disk



# But there is dust there! (not to scale)

22 5 5 5 5

Schlegel, Finkbeiner, Davis (1998) Dust Map

## **ISM dust in the MW thin disk:**

SFD maps **cannot** be used to correct low-latitude optical photometry for extinction: a major obstacle for mapping stellar distribution

• At high Galactic latitude (|b|>30°), stars are mostly behind dust

• At low Galactic latitude (|b|<30°), stars are **embedded** in dust

## **Major Challenge**

There is a degeneracy between dust reddening and intrinsic stellar colors!

With high-precision photometry from SDSS and 2MASS, this degeneracy can be lifted (for most stars)

Stellar SEDs are parametrized at high latitudes as a (nearly) one-parameter family; dust extinction is parametrized as a two-parameter family Berry et al. (2012, ApJ, 757, 166)



Determine intrinsic stellar SED and the dust extinction along the line of sight by fitting seven SDSS-2MASS colors with 2-3 free parameters

Since distance to main sequence stars can be estimated, this yields 3D dust map

# Why does this work? Near-IR colors break degeneracy between the stellar color and dust reddening:



But wide wavelength baseline (UV to near-IR) and precise photometry (errors ~0.02 mag or smaller) is required (i.e. SDSS & 2MASS)

# **Distance Bracketing**

Distance bins are (from left to right, in kpc): 0.1-0.5, 0.5-0.7, 0.7-0.9, 0.9-1.1

Note the abrupt jump in A<sub>r</sub> at b~2° for stars with D>0.9 kpc

- Lower limit for the distance to this dust cloud
- Implies molecular clouds have D>1kpc



# Spatial Distribution of Dust

## Difference in median A<sub>r</sub> along the line of sight for distances: 1kpc, 1.5kpc, 2kpc, 2.5kpc The dust structures at $b \sim 2^{\circ}$ and b~13° are confined to 1-1.5kpc The substructure at $-3^{\circ} < b < 0^{\circ}$ is at D~2.5kpc

**3D dust tomography!** All data (fits) are public:



http://www.astro.washington.edu/users/ivezic/r\_datadepot.html

# The Milky Way halo: not smooth, not simple



The halo structure is interesting because:
I) dynamical times are long (merger history), and
2) gravitational potential becomes dominated by the dark matter halo

#### 0.35 < r-i < 0.40

![](_page_22_Figure_1.jpeg)

![](_page_22_Figure_2.jpeg)

R (kpc)

#### Dissecting the Milky Way with SDSS

- Panoramic view of the Milky Way, akin to observations of external galaxies; good support for standard Galactic models (with amazing signal-to-noise!)
- Metallicity mapping supports components inferred from number counts mapping
- Kinematics correlated with metallicity: high-metallicity (disk) stars rotate, lowmetallicity (halo) stars on random highly eccentric orbits

![](_page_22_Figure_7.jpeg)

$$a_{R} = \sigma_{RR}^{2} \times \frac{\partial(\ln \nu)}{\partial R} + \frac{\partial \sigma_{RR}^{2}}{\partial R} + \sigma_{RZ}^{2} \times \frac{\partial(\ln \nu)}{\partial Z} + \frac{\partial \sigma_{RZ}^{2}}{\partial Z} + \frac{\sigma_{RR}^{2}}{R} - \frac{\sigma_{\Phi\Phi}^{2}}{R} - \frac{\overline{\nu_{\Phi}}^{2}}{R},$$

$$a_{Z} = \sigma_{RZ}^{2} \times \frac{\partial(\ln \nu)}{\partial R} + \frac{\partial \sigma_{RZ}^{2}}{\partial R} + \sigma_{ZZ}^{2} \times \frac{\partial(\ln \nu)}{\partial Z} + \frac{\partial \sigma_{ZZ}^{2}}{\partial Z} + \frac{\sigma_{RZ}^{2}}{R}.$$

$$M_{1}^{2} = \sigma_{RZ}^{2} \times \frac{\partial(\ln \nu)}{\partial R} + \frac{\partial \sigma_{RZ}^{2}}{\partial R} + \sigma_{ZZ}^{2} \times \frac{\partial(\ln \nu)}{\partial Z} + \frac{\partial \sigma_{ZZ}^{2}}{\partial Z} + \frac{\sigma_{RZ}^{2}}{R}.$$

$$M_{1}^{2} = \sigma_{RZ}^{2} \times \frac{\partial(\ln \nu)}{\partial R} + \frac{\partial \sigma_{RZ}^{2}}{\partial R} + \sigma_{ZZ}^{2} \times \frac{\partial(\ln \nu)}{\partial Z} + \frac{\partial \sigma_{ZZ}^{2}}{\partial Z} + \frac{\sigma_{RZ}^{2}}{\partial Z} + \frac{\partial(\ln \nu)}{\partial R} + \frac{\partial \sigma_{RZ}^{2}}{\partial Z} + \frac{\sigma_{RZ}^{2}}{\partial Z} + \frac{\partial(\ln \nu)}{\partial Z} + \frac{\partial($$

Kinematics of halo stars based on SDSS-POSS proper motions: velocity ellipsoid is nearly invariant in spherical coordinate system

Given stellar distribution from Juric+2008 and stellar kinematics from Bond+2010, we can apply **Jeans equations** and infer the gravitational potential, and ultimately the distribution of dark matter! Can we really?

![](_page_24_Figure_1.jpeg)

Kinematics of halo stars based on SDSS-POSS proper motions: velocity ellipsoid is nearly invariant in spherical coordinate system

![](_page_24_Figure_3.jpeg)

![](_page_25_Figure_1.jpeg)

SDSS aZ normalized by model aZ Top: model with DM; bottom: no DM Kinematics of halo stars based on SDSS-POSS proper motions: velocity ellipsoid is nearly invariant in spherical coordinate system

![](_page_25_Figure_4.jpeg)

![](_page_26_Figure_1.jpeg)

acc.

ratio

Obs/

Mod

**Figure 5.** Pixel-by-pixel comparison of the acceleration values implied by the SDSS data and the two model predictions that include (black lines) and do not include (red lines) contributions from dark matter. The top panel corresponds to the  $a_Z$  acceleration maps shown in Figure 4 and the bottom panel to the  $a_R$  acceleration maps. The model-based acceleration maps that include a dark matter contribution provide a significantly better description of the acceleration maps derived from the SDSS data.

Kinematics of halo stars based on SDSS-POSS proper motions: velocity ellipsoid is nearly invariant in spherical coordinate system, which implies that the DM halo is oblate! q=0.5±0.1 (Loebman et al. 2012; ApJ, 758, L23)

![](_page_26_Figure_4.jpeg)

## Halo density profiles out to 35kpc

![](_page_27_Figure_1.jpeg)

Lecture II: Discovering The Galaxy With Large Photometric Surveys

Mario Juric <mjuric@cfa.harvard.edu>, Tuesday, August 3rd, 2010. XV IAG-USP Advanced School on Astrophysics, Campos do Jordão, Brazil

![](_page_28_Picture_0.jpeg)

![](_page_28_Figure_1.jpeg)

#### Outer halo studies: RR Lyrae from SDSS Stripe 82

- Top left: the disk structure (artist's conception based on the Spitzer and other surveys of the Galactic plane)
- Bottom left: the halo density (multiplied by  $R^3$ ; yellow and red are overdensities relative to mean  $\rho(R) \propto R^{-3}$  density) as traced by RR Lyrae from SDSS Stripe 82 (Sesar et al. 2009), compared in scale to the top panel
- Conclusions: the spatial distribution of halo stars is highly inhomogeneous (clumpy); when averaged, the stellar volume density decreases as  $\rho(R) \propto R^{-3}$ . Limited by data!

# The Milky Way is complex and big!

![](_page_29_Figure_1.jpeg)

# And interesting and informative!

![](_page_30_Picture_0.jpeg)

# Only 2.5 deg wide: <1% halo volume! **"We need more data"**

## U.S. Decadal Survey 2010

# Priorities:

# • Spaced-based:

Wide-Field Infrared Survey Telescope WFIRST
 The Explorer Program "rapid response"
 Laser Interferometer Space Antenna LISA
 International X-ray Observatory IXO

# Ground-based:

Large Synoptic Survey Telescope LSST
 Mid-scale Innovations Program "rapid response"
 Giant Segmented Mirror Telescope (30m) GSMT
 Atmospheric Čerenkov Telescope Array (Y) ACTA
 Cerro Chajnantor Atacama Telescope (submm) CCAT

![](_page_33_Picture_0.jpeg)

EE

EE

1

An optical/near-IR survey of half the sky in ugrizy bands to r~27.5 based on ~1000 visits over a 10-year period: More information at www.lsst.org and arXiv:0805.2366

LSST: a digital color movie of the Universe...

A catalog of 10 billion stars and 10 billion galaxies with exquisite photometry, astrometry and image quality!

# **LSST Science Themes**

- Dark matter, dark energy, cosmology (spatial distribution of galaxies, gravitational lensing, supernovae, quasars)
- Time domain (cosmic explosions, variable stars)
- The Solar System structure (asteroids)
- The Milky Way structure (stars)

LSST Science Book: arXiv:0912.0201 Summarizes LSST hardware, software, and observing plans, science enabled by LSST, and educational and outreach opportunities

245 authors, 15 chapters, 600 pages

![](_page_34_Picture_7.jpeg)

# **LSST Science Themes**

- Dark matter, dark energy, cosmology (spatial distribution of galaxies, gravitational lensing, supernovae, quasars)
- Time domain (cosmic explosions, variable stars)
- The Solar System structure (asteroids)
- The Milky Way structure (stars)

These drivers not only require similar hardware and software systems, but also motivate a universal cadence: **about 90% of time will be spent on a uniform survey** 

# Basic idea behind LSST: a uniform sky survey

- 90% of time will be spent on a uniform survey: every 3-4 nights, the whole observable sky will be scanned twice per night
- after 10 years, half of the sky will be imaged about 1000 times (in 6 bandpasses, ugrizy): a digital color movie of the sky
- ~100 PB of data: about a billion 16 Mpix images, enabling measurements for 20 billion objects!

![](_page_36_Figure_4.jpeg)

## LSST in one sentence:

An optical/near-IR survey of half the sky in ugrizy bands to r~27.5 (36 nJy) based on 825 visits over a 10year period: deep wide fast.

Left: a 10-year simulation of LSST survey: the number of visits in the r band (Aitoff projection of eq. coordinates)

Cadence details still open to optimization (e.g. see Gould, arXiv:1304.3455)

# SDSS-LSST comparison: LSST=d(SDSS)/dt, LSST=SuperSDSS 7x7 arcmin, gri

![](_page_37_Figure_1.jpeg)

![](_page_37_Picture_2.jpeg)

# SDSS-LSST comparison: LSST=d(SDSS)/dt, LSST=SuperSDSS 7x7 arcmin, gri

![](_page_38_Picture_1.jpeg)

# SDSS: one US Library of Congress worth of data LSST: one SDSS per night, or all the words ever printed! (Deep Lens Survey)

![](_page_38_Picture_3.jpeg)

LSST system Telescope Camera Software

and the second second

# LSST Telescope

![](_page_40_Picture_1.jpeg)

# The field-of-view comparison: Gemini vs. LSST

![](_page_41_Picture_1.jpeg)

# **Optical Design for LSST**

![](_page_42_Figure_1.jpeg)

Three-mirror design (Paul-Baker system) enables large field of view with excellent image quality: delivered image quality is dominated by atmospheric seeing

# Mirror fabrication is advanced - Private funding enabled early start of both reflective optics

![](_page_43_Picture_1.jpeg)

![](_page_43_Picture_2.jpeg)

![](_page_43_Picture_3.jpeg)

- Primary-Tertiary was cast in 2008
- Fabrication completed in 2013
- Secondary substrate fabricated by Corning in 2009.

## The largest astronomical camera: 2800 kg, 3.2 Gpix

![](_page_44_Figure_1.jpeg)

# LSST camera

![](_page_45_Figure_1.jpeg)

Modular design: 3200 Megapix = 189 x16 Megapix CCD 9 CCDs share electronics: raft (=camera) Problematic rafts can be replaced relatively easily

# LSST Software

![](_page_46_Figure_1.jpeg)

# Software: the subsystem with the highest risk

- 20 TB of data to process every day (~one SDSS/day)
- 1000 measurements for 20 billion objects during 10 years
- Existing tools and methods (e.g. SDSS) do not scale up to LSST data volume and rate (100 PB!)

![](_page_47_Picture_4.jpeg)

# LSST Timeline

![](_page_48_Figure_1.jpeg)

#### **Estimate:** survey operations begin in 2019 (if MREFC in FY2014)

- Primary/Tertiary Mirror being polished, have secondary mirror blank
- Sensor development program delivered first prototype sensors
- Processing pipelines under construction, hand-in-hand with simulations of Operations, Images, Catalogs
- **Cost:** ~\$850M in \$2011

contributions from NSF, DOE and private gifts

# Extragalactic astronomy: 10 billion galaxies

## SDSS MUSYC

![](_page_49_Picture_2.jpeg)

![](_page_49_Picture_3.jpeg)

## The Milky Way structure: 10 billion stars, time domain massive statistical studies!

## Main sequence stars Distance and [Fe/H]:

![](_page_50_Figure_2.jpeg)

# 0.35 < r-i < 0.40

![](_page_50_Figure_4.jpeg)

![](_page_50_Figure_5.jpeg)

Compared to SDSS: LSST can "see" about 40 times more stars, 10 times further away and over twice as large sky area

![](_page_50_Picture_7.jpeg)

Sesar et al. (2009

# Gaia vs. LSST comparison

![](_page_51_Figure_1.jpeg)

**Gaia:** excellent astrometry (and photometry), but only to r < 20

LSST: photometry to r < 27.5 and time resolved measurements to r < 24.5

Complementarity of the two surveys: photometric, proper motion and trigonometric parallax errors are similar around r=20

The Milky Way disk "belongs" to Gaia, and the halo to LSST (plus very faint and/or very red sources, such as white dwarfs and LT(Y) dwarfs).

## Comparison of SDSS, Gaia and LSST for main sequence stars:

![](_page_52_Figure_1.jpeg)

# **Dwarfs in LSST**

White dwarfs: LF is age probe

~400,000 halo white dwarfs from LSST (10 million total):

![](_page_53_Figure_3.jpeg)

L / T dwarfs: L dwarfs are dime a dozen: 200,000 in LSST with proper motion and trigonometric parallax measurements

Simulations predict 2400 T dwarfs with >50 proper motion and parallax measurements

Compared to UKIDSS, 5 times larger sample of T dwarfs, with parallaxes and 10-20 times more accurate proper motions

(~100 Y dwarfs [model based])

The large blue circle: the  $\sim$ 400 kpc limit of future LSST studies based on RR Lyrae

The large red circle: the ~100 kpc limit of future LSST studies based on main-sequence stars (and the current limit for RR Lyrae studies)

![](_page_54_Figure_2.jpeg)

LSST Innitor PRIVIDE 6D information from LSST: 3D spatial, 2 velocities, [Fe/H] The small insert: ~10 kpc limit of SDSS and future Gaia studies for kinematic & [Fe/H]mapping with MS stars

The large blue circle: the  $\sim$ 400 kpc limit of future LSST studies based on RR Lyrae

300 kpc

The large red circle: the  $\sim$ 100 kpc limit of future

LSST studies (and the curr

figure from J. Bullock's webpa

200 million stars from LSST!

he small insert: 10 kpc limit of SDSS nd future Gaia studies or kinematic & [Fe/H] apping with MS stars

montage from B.Willman

inset: SDSS map to  $d_{limit} = 10$  kpc

The large blue circle: the  $\sim$ 400 kpc limit of future LSST studies based on RR Lyrae

300 kpc

The large red circle: the ~100 kpc limit of future

LSST studies (and the curi

# ~500 RR Lyrae from SDSS

igure from J. Bu

#### montage from B.Willman

inset: SDSS map to d<sub>limit</sub> = 10 kpc

# 200 million stars from LSST!

he small insert: 10 kpc limit of SDSS nd future Gaia studies or kinematic & [Fe/H] apping with MS stars Data analysis challenges in the era of LSST

- 1) Large data Volume
- 2) Large number of objects
- 3) Highly multi-dimensional space
- 4) Unknown statistical distributions
- 5) Time-series data
- 6) Truncated, censored and missing data

7) Unreliable quantities (e.g. unknown systematics and random errors)

# Statistics, Data Mining and Machine Learning in Astronomy

Željko Ivezić, Andrew Connolly, Jacob Vanderplas, Alex Gray

**Princeton University Press, 2013** 

Statistics, Data Mining, and Machine Learning in Astronomy Zelleo Iveric, Andrew Concolly, Just Waderster, Alex Gen

![](_page_58_Picture_4.jpeg)

- Complete *Practical* guide to statistical analysis, data exploration, and machine learning
- Example-driven approach, using real data (SDSS, LIGO, LINEAR, WMAP, and others)
- All book figures and examples generated in python (matplotlib), with code available online – for free!
- Makes use of numpy, scipy, matplotlib, scikit-learn, pymc, healpy, and others
- Supporting python package: astroML

![](_page_58_Picture_10.jpeg)

astroML

#### News

October 2012: astroML 0.1 has been released! Get the source on Github

Our Introduction to astroML paper received the CIDU 2012 best paper award.

#### Links

astroML Mailing List GitHub Issue Tracker

#### Videos

Scipy 2012 (15 minute talk)

#### Citing

If you use the software, please consider citing astroML.

## AstroML: Machine Learning and Data Mining for Astronomy

![](_page_59_Figure_14.jpeg)

AstroML is a Python module for machine learning and data mining built on numpy, scipy, scikit-learn, and matplotlib, and distributed under the 3-clause BSD license. It contains a growing library of statistical and machine learning routines for analyzing astronomical data in python, loaders for several open astronomical datasets, and a large suite of examples of analyzing and visualizing astronomical datasets.

#### **Downloads**

- Released Versions: Python Package Index
- Bleeding-edge Source: github

The goal of astroML is to provide a community repository for fast Python implementations of common tools and routines used for statistical data analysis in astronomy and astrophysics, to provide a uniform and easyto-use interface to freely available astronomical datasets. We hope this package will be useful to researchers and students of astronomy. The astroML project was started in 2012 to accompany the book **Statistics**, **Data Mining, and Machine Learning in Astronomy** by Zeljko Ivezic, Andrew Connolly, Jacob VanderPlas, and Alex Gray, to be published in late 2013. The table of contents is available here: here(pdf).

![](_page_59_Picture_20.jpeg)

#### User Guide

#### 1. Introduction

1.1. Philosophy

## **Open source!** www.astroML.org

![](_page_60_Figure_0.jpeg)

# The Excitement of LSST

EE

EE

More information at www.lsst.org and arXiv:0805.2366

- The Best Sky Image Ever: 60 petabytes of astronomical image data (resolution equal to 3 million HDTV sets)
- The Greatest Movie of All Time: digital images of the entire observable sky every three nights, night after night, for 10 years (11 months to "view" it)
- The Largest Astronomical Catalog: 20 billion sources (for the first time in history more than living people)

# SDSS: a digital color **map** of the night sky LSST: a digital color **movie** of the sky

"If You Liked SDSS, You will Love LSST!"

![](_page_61_Picture_2.jpeg)