Tri-Agency Effort Regarding Joint Pixel Processing of LSST, WFIRST, and Euclid

Will Dawson (LLNL) and Peter Melchior (Princeton)

Tri-Agency Group Task Overview

Combining surveys at catalog level ***is*** powerful for cosmology

Key question: Do we need pixel-level analysis and, if so, why?

NASA, NSF, DoE formed group to investigate and make recommendations

Tri-Agency Group Task Overview

Phase I (previous effort): “Multi-Survey Processing: Science Data Handling” report submitted to DOE, NASA, and NSF on 2016 February 16.

Phase II (current effort):

> Scoping activities including requirements definition, improved scoping of algorithms, and design and architecture work.
>
> March 2019 final report due

Phase 3 (future effort): Developing Software and Systems

> Approx. start by Sep. 2019 to allow 3 year development for LSST & Euclid start

3 Primary Efforts

a) Developing requirements for joint processing
b) Scoping effort required for developing and optimizing algorithms
c) Designing and optimizing computing architecture

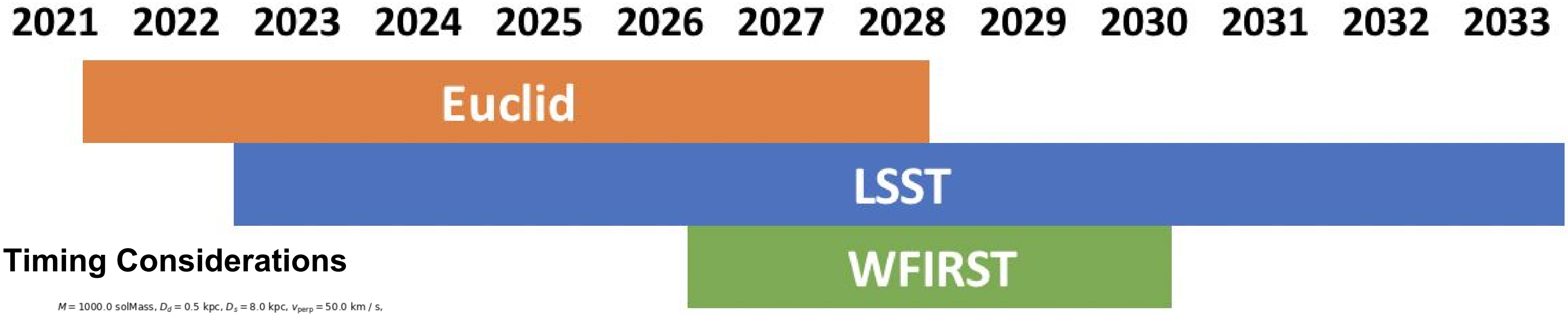
Task	a.i	a.ii	a.iii	a.iv	b.i	b.ii	b.iii	c.i	c.ii	c.iii	c.iv	c.v
Topics	Best phot-z for 3D WL, Breaking z-degeneracies	Time domain astronomy/SN /AGN transients/rapid response	Galactic, Solar System, Stellar Streams, Reionization	Microlensing, time baselines for SSOs	LSST Deconfusion, optimal photometry for phot-z, cross-mission color selection, dust corrections			Infrastructure, networking, hardware, interface				
Lead(s)	Newman (Upitt.)	Chary (IPAC)	Paladini (IPAC)	Helou (IPAC), Dawson (LLNL)	Lupton (Princeton) & Ferguson (STScI)	Lupton & Ferguson	Melchior (Princeton)	Dawson & A. Smith (STScI)	Dawson & A. Smith	Rusholme (IPAC)	Appleton (IPAC)	Teplitz (IPAC)
Members	Momcheva, Ferguson, Schneider, Prakash, Chary, Capak	Ferguson, Juric, Momcheva, Prakash, Capak, Armus, Wood-vasey	Ferguson, Momcheva, Wachter, Kirkpatrick, Chary, Grillmair	van der Marel, Carey, Grillmair	Dawson, Melchior, Schneider, Schulz, B. Lee, Appleton	Dawson, Melchior, Ferguson, Schneider, B. Lee, Grillmair, Armus	Dawson, Ferguson, Koekemoer, Lupton, Schulz	Schneider, Fox, Groom, Ebert	Fox, Flynn, Ebert	Smith, Fox, Groom, Berriman	Smith, Fox, Wachter, B. Lee, Rusholme, Berriman	Smith, Fox, Wachter, Groom, Rusholme, Berriman

Developing Requirements

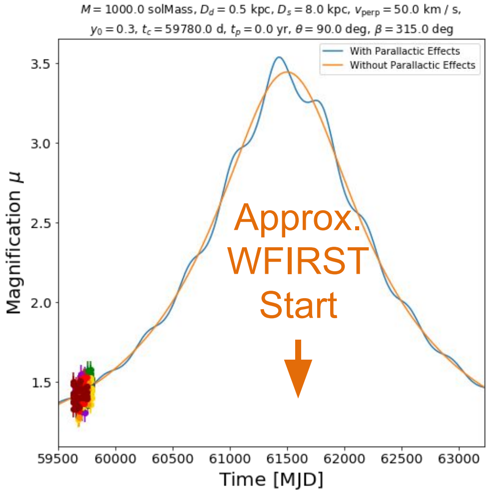
Task	a.i	a.ii	a.iii	a.iv
Topics	Best phot-z for 3D WL, Breaking z-degeneracies	Time domain astronomy/SN /AGN transients/rapid response	Galactic, Solar System, Stellar Streams, Reionization	Microlensing, time baselines for SSOs
Lead(s)	Newman (Upitt.)	Chary (IPAC)	Paladini (IPAC)	Helou (IPAC), Dawson (LLNL)
Members	Momcheva, Ferguson, Schneider, Prakash, Chary, Capak	Ferguson, Juric, Momcheva, Prakash, Capak, Armus, Wood-vasey	Ferguson, Momcheva, Wachter, Kirkpatrick, Chary, Grillmair	van der Marel, Carey, Grillmair

I. Fully Characterized, most accurate solutions
II. Rapid, agile exploration of alternative hypotheses
III. Use cases and science/data requirements to meet needs of broader community
IV. Implications of various timelines of data acquisition in different surveys

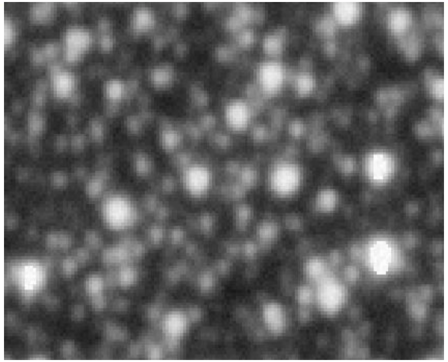
The Microlensing Science Case



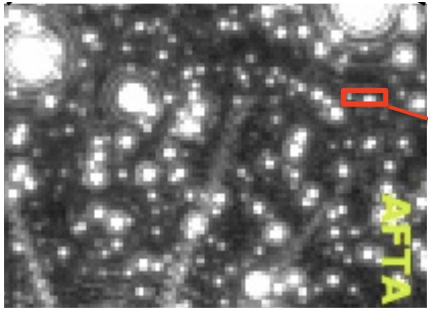
Timing Considerations



Spatial and Resolution Considerations



DECam image of the Bulge

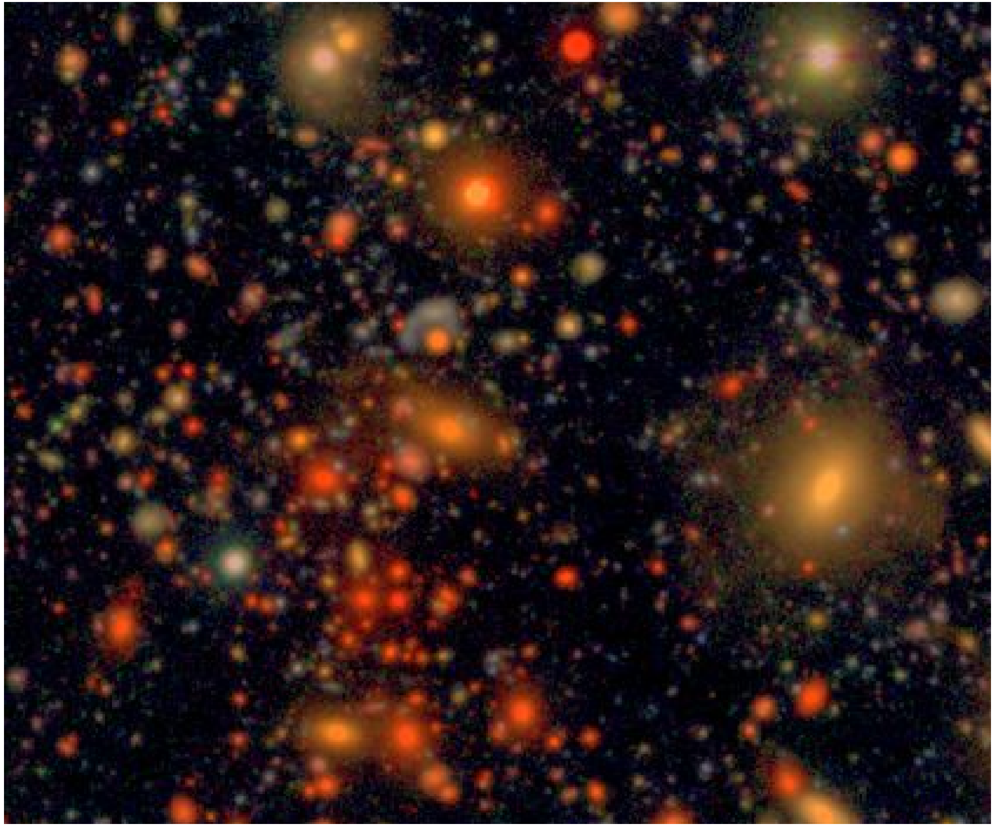


WFIRST image of the Bulge
Credit: M. Penny

Algorithmic Challenges

b.i	b.ii	b.iii
LSST Deconfusion, optimal photometry for phot-z, cross-mission color selection, dust corrections		
Lupton (Princeton) & Ferguson (STScI)	Lupton & Ferguson	Melchior (Princeton)
Dawson, Melchior, Schneider, Schulz, B. Lee, Appleton	Dawson, Melchior, Ferguson, Schneider, B. Lee, Grillmair, Armus	Dawson, Ferguson, Koekemoer, Lupton, Schulz

HSC 5-band, comparable to LSST



Algorithmic Challenges

b.i	b.ii	b.iii
LSST Deconfusion, optimal photometry for phot-z, cross-mission color selection, dust corrections		
Lupton (Princeton) & Ferguson (STScI)	Lupton & Ferguson	Melchior (Princeton)
Dawson, Melchior, Schneider, Schulz, B. Lee, Appleton	Dawson, Melchior, Ferguson, Schneider, B. Lee, Grillmair, Armus	Dawson, Ferguson, Koekemoer, Lupton, Schulz

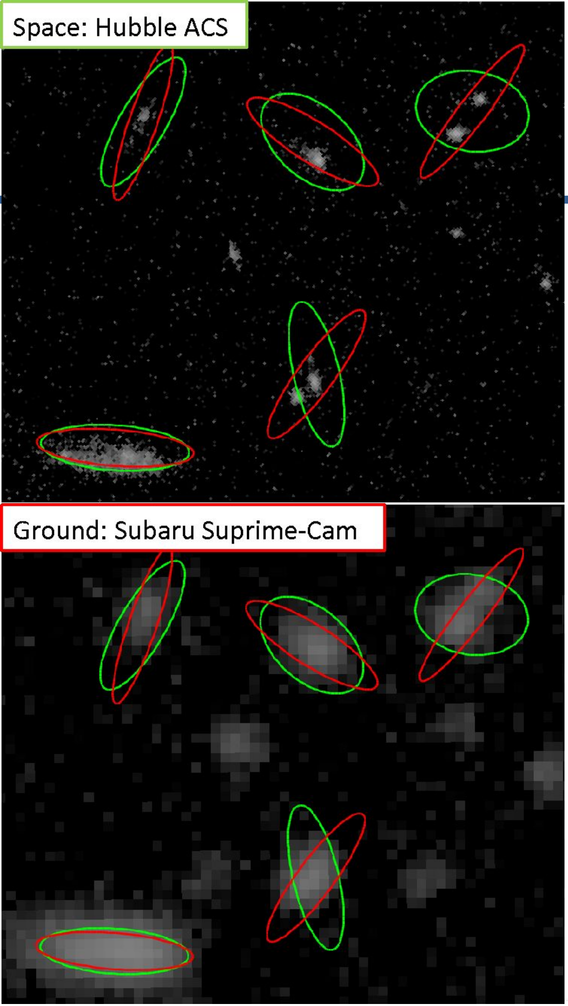
Association of pixels to objects ambiguous

Catalog-level combinations inferior

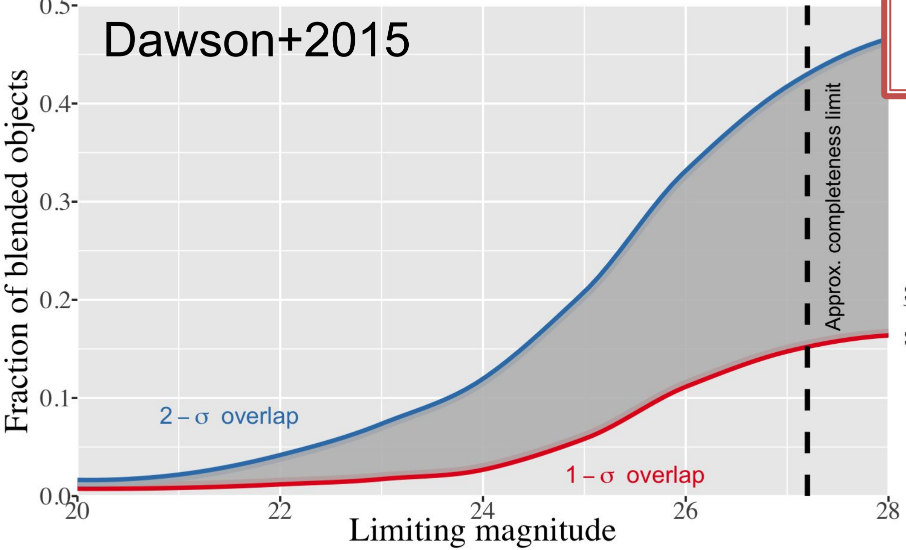
- for blended objects
- for transient/moving objects
- compared to forced/matched photometry

Challenge: accounting & utilizing different wavelengths, resolutions, cadences

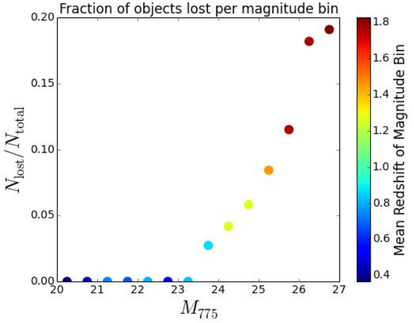
Catalog cross-matching between space and ground is confused by significant object blending as seen by LSST



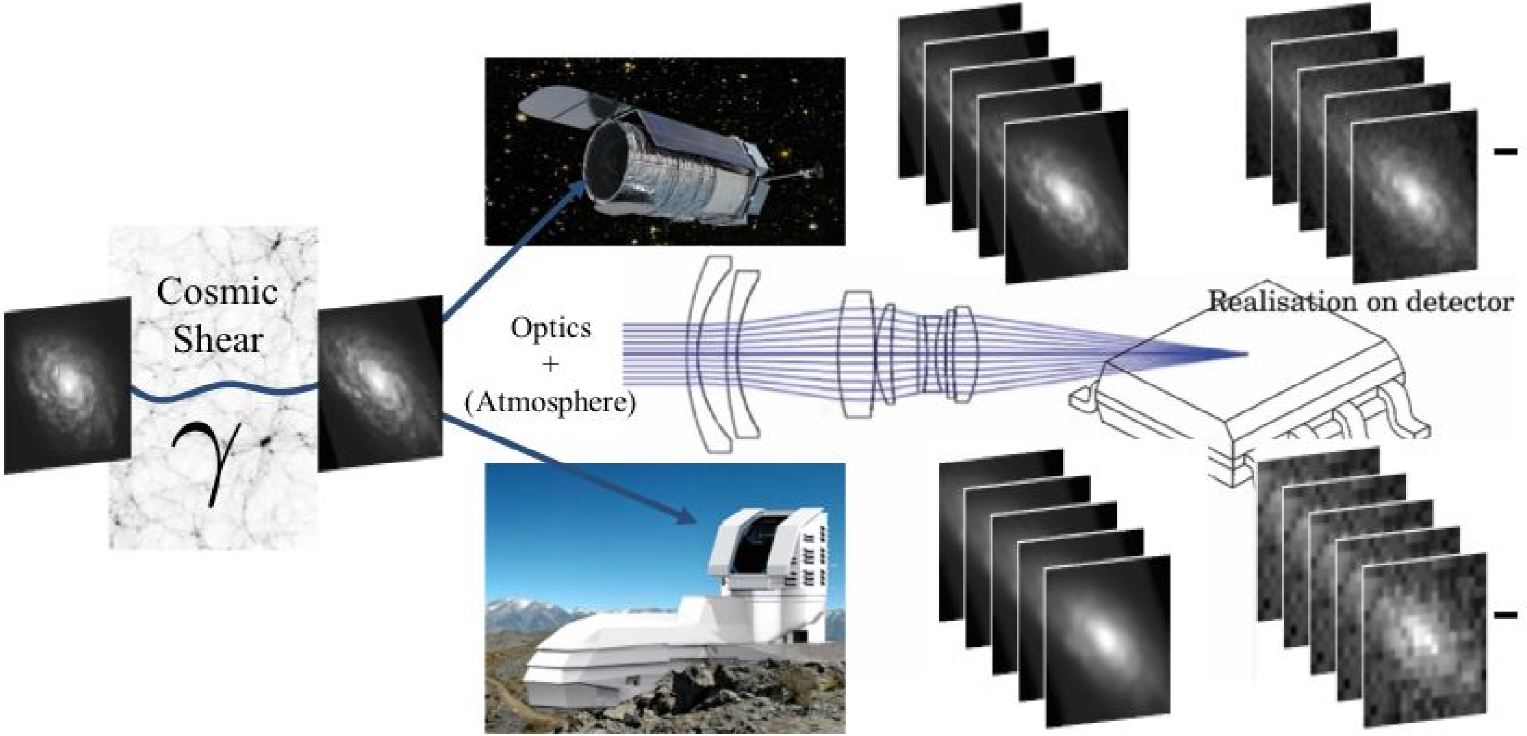
LSST blend fractions estimated from Subaru & HST overlapping imaging



Catalog cross-matched studies are limited largely to population studies.



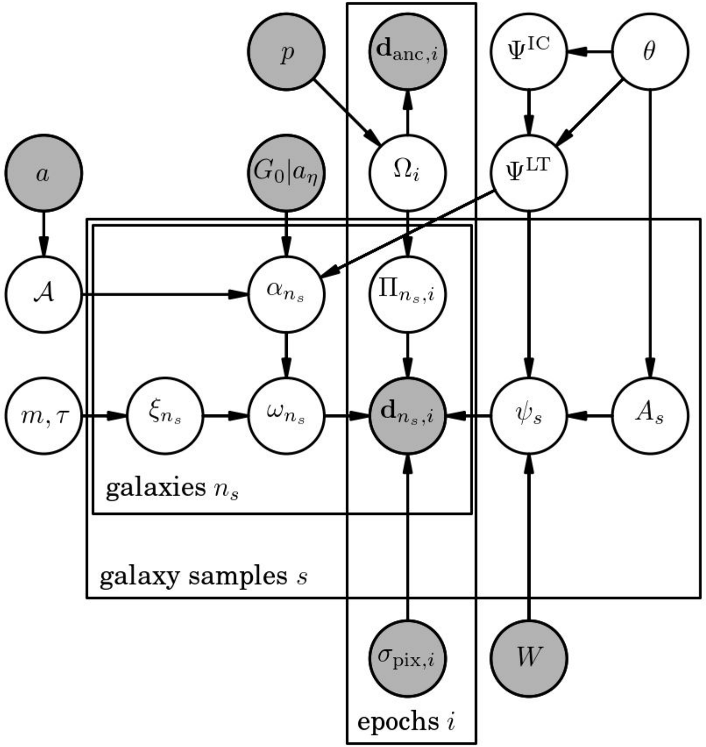
To go beyond population statistics we really need to go back to the pixels.



The complete statistical model for cosmic shear

arXiv:1411.2608

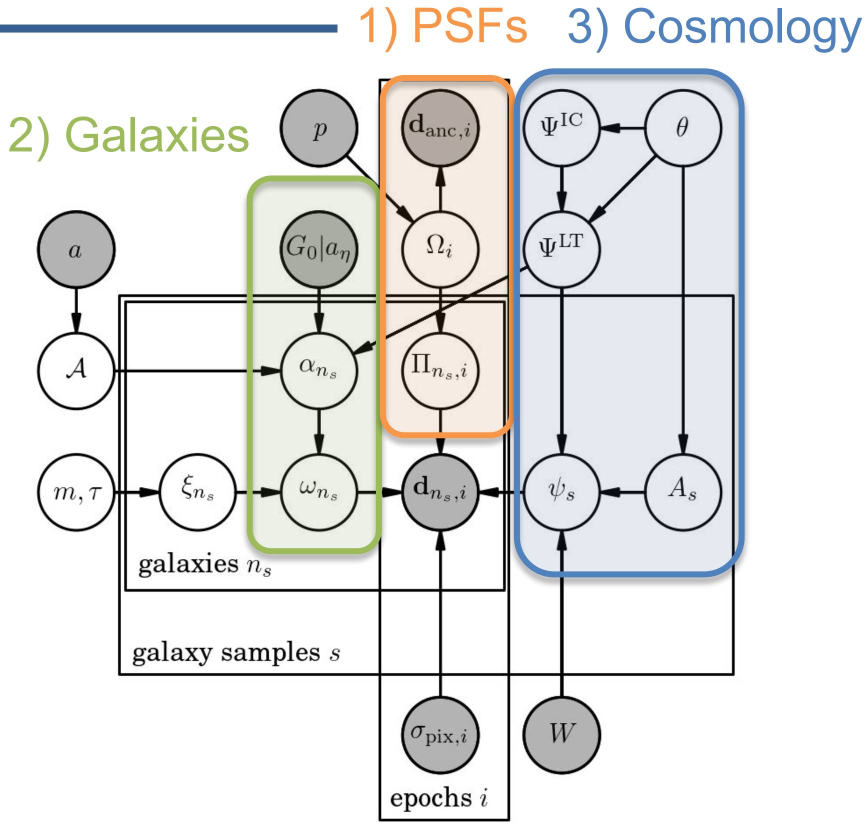
Parameter	Description
$\boldsymbol{\theta}$	Cosmological parameters
Ψ^{IC}	Initial conditons for the 3D gravitational potential
Ψ^{LT}	Late-time 3D gravitational potential
ψ_s	2D lens potential (given source photo-z bin s)
A_s	Parameters for the line-of-sight source distribution
$\Pi_{n_s,i}$	PSF for galaxy n_s observed in epoch i
Ω_i	Observing conditions in epoch i
$\{\omega_{n_s}\}$	Galaxy model parameters; $n_s = 1, \ldots, n_{\mathrm{gal},s}$
$\{\alpha_{n_s}\}$	Parameters for the distribution of $\{\omega_{n_s}\}$
$\{\xi_{n_s}\}$	Scaling parameters for $\{\omega_{n_s}\}$
m, τ	Hyperprior parameters for $\{\xi_{n_s}\}$
$\mathcal{A}$	Hyperparameter for $\{\alpha_{n_s}\}$ classifications
$\{\mathbf{d}_{n_s,i}\}$	Pixel data for galaxies $n_s = 1, \ldots, n_{\mathrm{gal},s}$ in epoch i
$G_0 \vert a_\eta$	Prior specification for $\{\alpha_{n_s}\}$
s	Source sample (e.g., photo-z bin)
W	Survey window function
$\mathbf{d}_{\mathrm{anc},i}$	Ancillary data for PSF in epoch i
p	Prior params. for observing conditions
a	Prior params. for $\mathcal{A}$
$\sigma_{\mathrm{pix},i}$	Pixel noise r.m.s. in epoch i
I	Model selection assumptions



The complete statistical model for cosmic shear

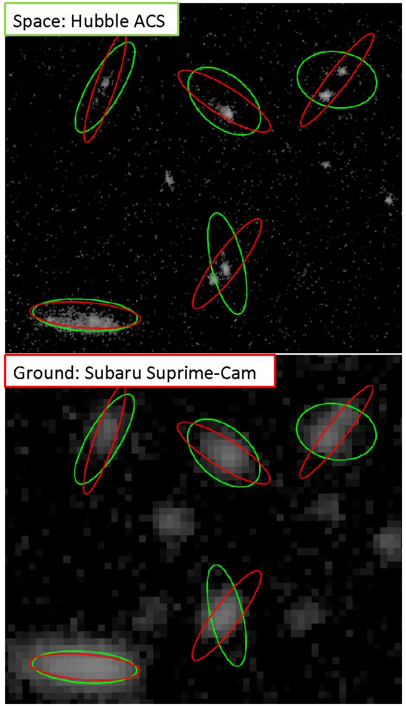
arXiv:1411.2608

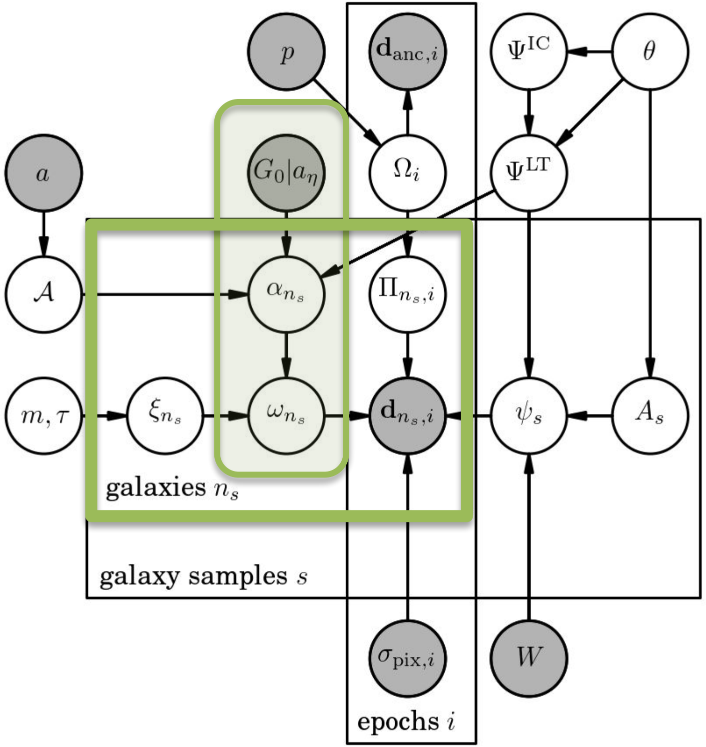
Parameter	Description
$\boldsymbol{\theta}$	Cosmological parameters
Ψ^{IC}	Initial conditons for the 3D gravitational potential
Ψ^{LT}	Late-time 3D gravitational potential
ψ_s	2D lens potential (given source photo-z bin s)
A_s	Parameters for the line-of-sight source distribution
$\Pi_{n_s,i}$	PSF for galaxy n_s observed in epoch i
Ω_i	Observing conditions in epoch i
$\{\omega_{n_s}\}$	Galaxy model parameters; $n_s = 1, \ldots, n_{\mathrm{gal},s}$
$\{\alpha_{n_s}\}$	Parameters for the distribution of $\{\omega_{n_s}\}$
$\{\xi_{n_s}\}$	Scaling parameters for $\{\omega_{n_s}\}$
m, τ	Hyperprior parameters for $\{\xi_{n_s}\}$
$\mathcal{A}$	Hyperparameter for $\{\alpha_{n_s}\}$ classifications
$\{\mathbf{d}_{n_s,i}\}$	Pixel data for galaxies $n_s = 1, \ldots, n_{\mathrm{gal},s}$ in epoch i
$G_0\|a_\eta$	Prior specification for $\{\alpha_{n_s}\}$
s	Source sample (e.g., photo-z bin)
W	Survey window function
$\mathbf{d}_{\mathrm{anc},i}$	Ancillary data for PSF in epoch i
p	Prior params. for observing conditions
a	Prior params. for $\mathcal{A}$
$\sigma_{\mathrm{pix},i}$	Pixel noise r.m.s. in epoch i
I	Model selection assumptions



Naïve approach is intractable

1. Galaxy models be joint fitted to all available epochs i

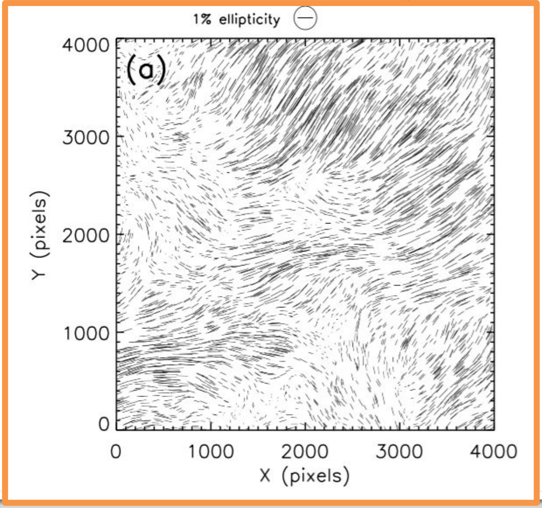


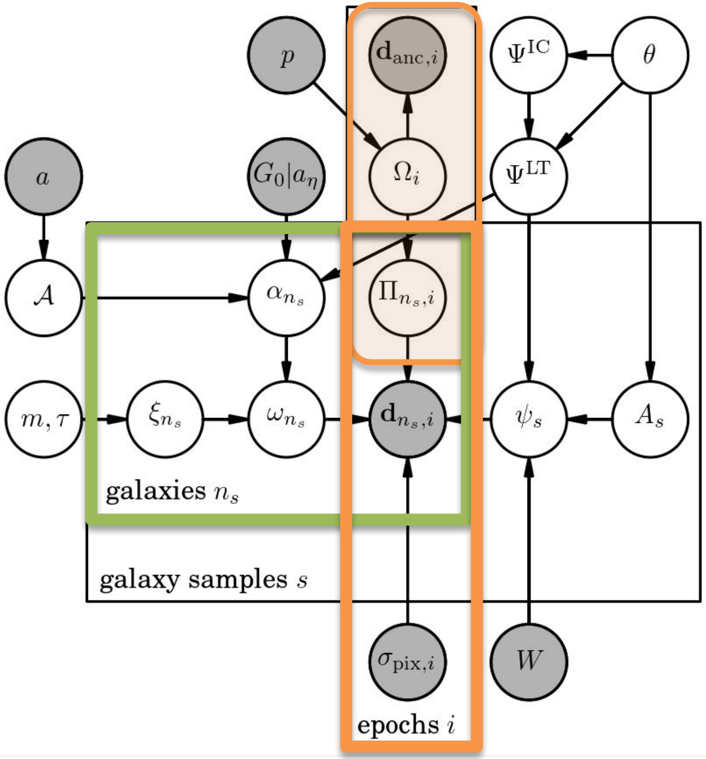


Naïve approach is intractable

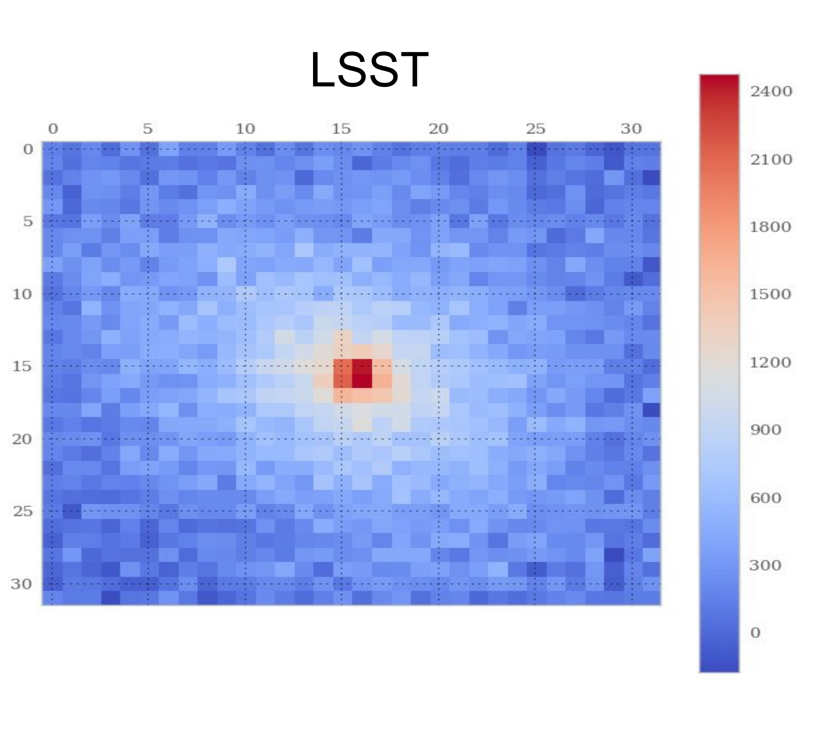
1. Galaxy models be joint fitted to all available epochs i

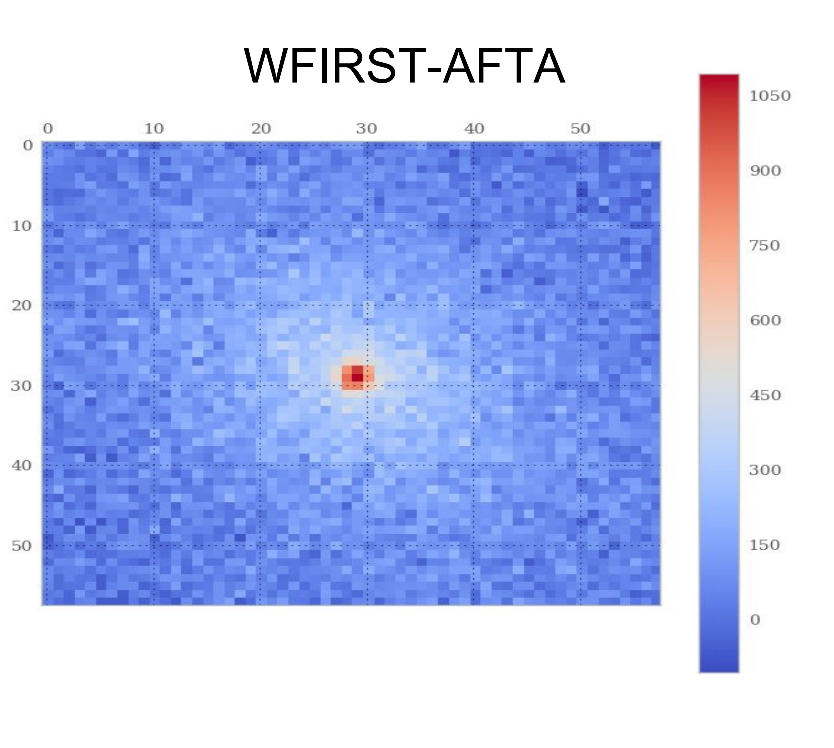
2. PSF models must be joint fitted to all galaxies in an exposure n_s



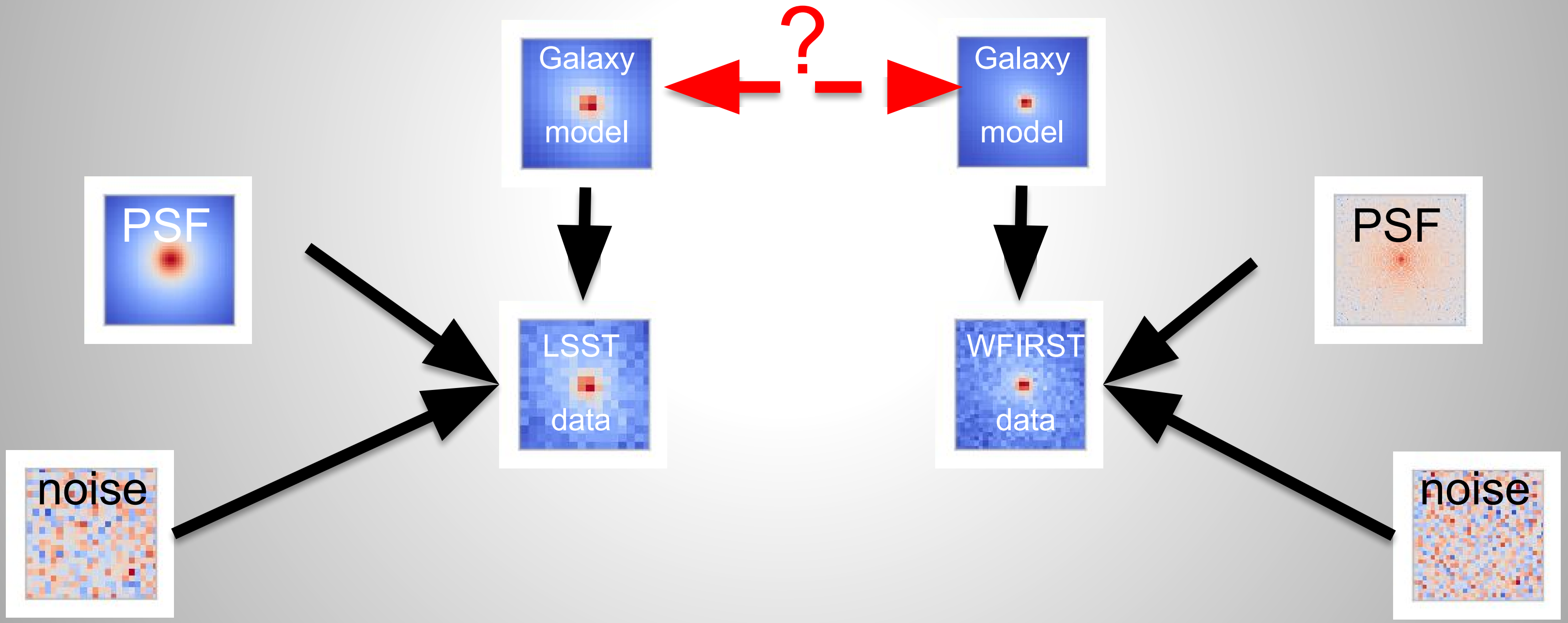


How do ellipticity measurements compare for an isolated, high-SNR galaxy?

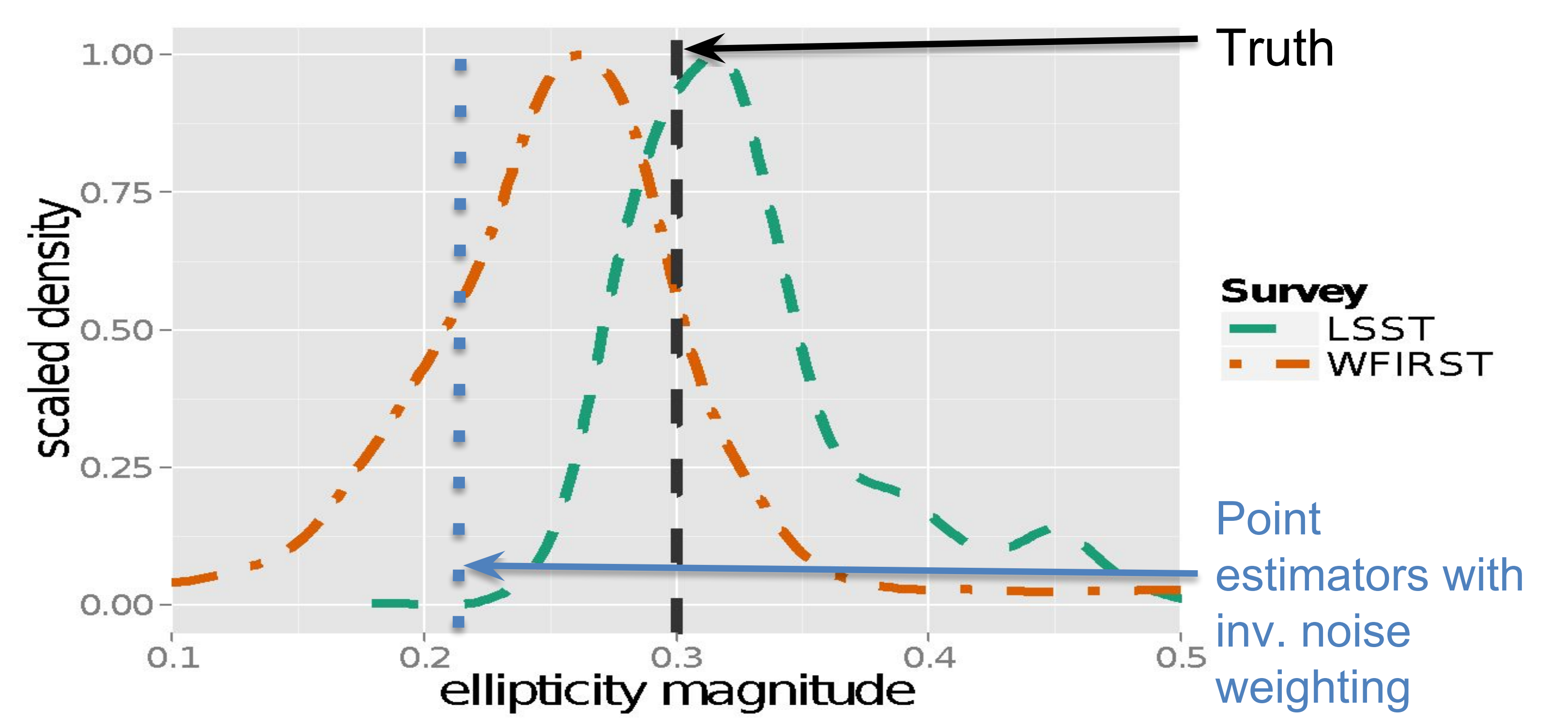




Measuring galaxy ellipticity (and other properties) from different surveys

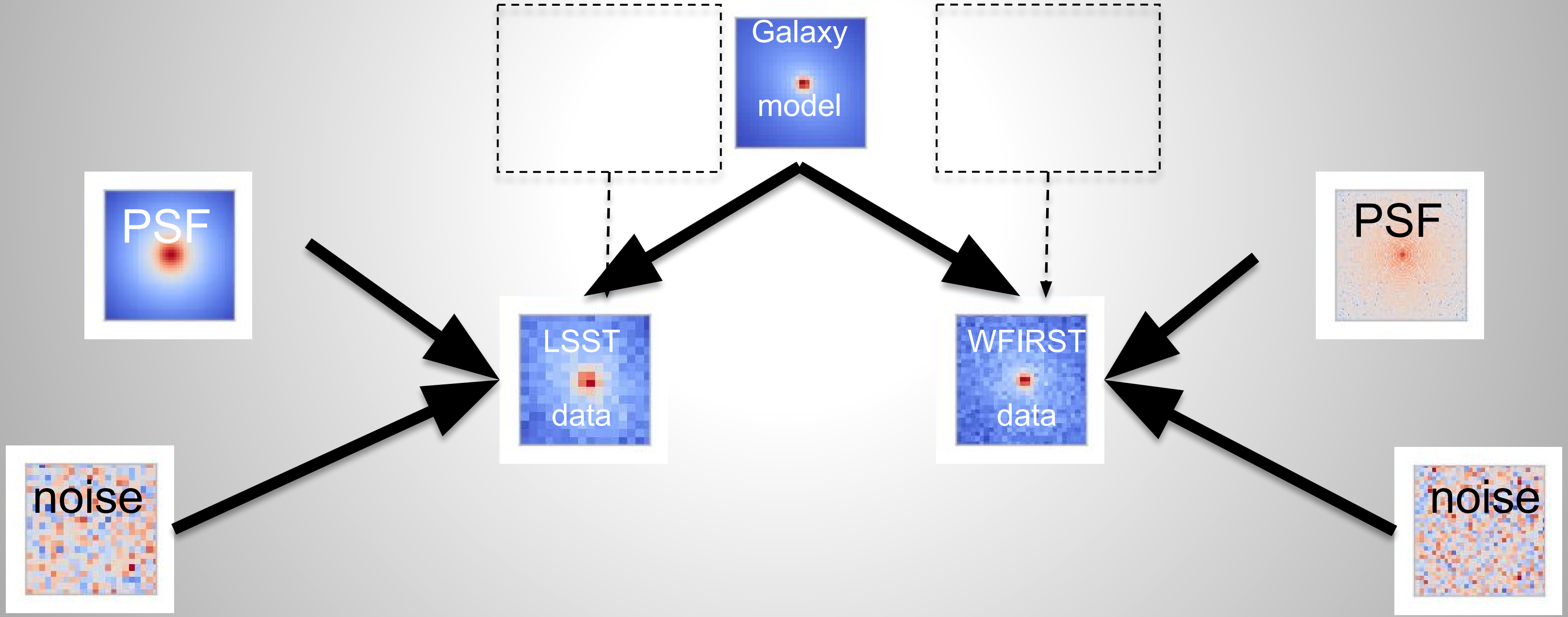


Measuring galaxy ellipticity (and other properties) from different surveys

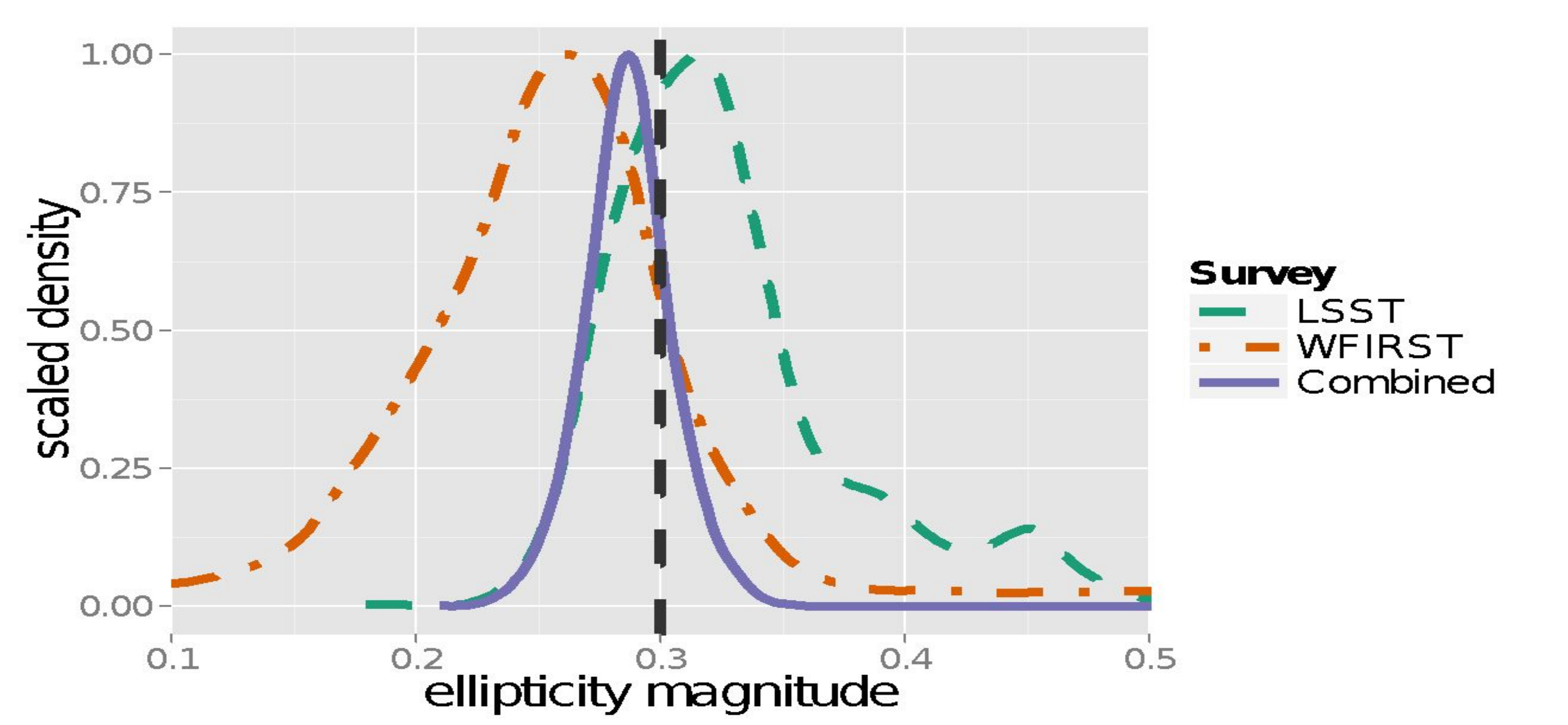


Measuring galaxy ellipticity (and other properties) from combined survey data

Analyze pixels from both surveys assuming same galaxy model.
Analogous to forced photometry.



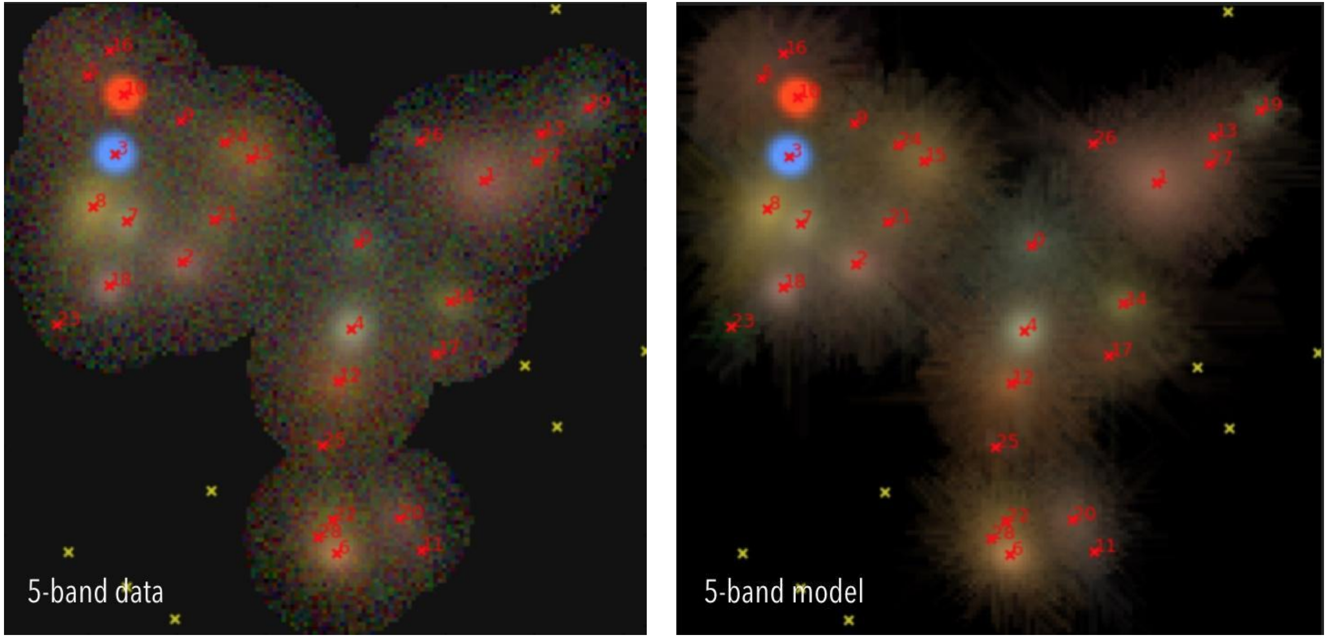
Measuring galaxy ellipticity (and other properties) from combined survey data



Algorithmic Challenges: Deblending

LSST & Euclid/WFIRST: resolution critical for detection and overlap regions

Multi-band coverage: utilizes color variation between components



Melchior+ in prep.

Algorithmic Challenges: Deblending

Component separation by modeling spectrum $\otimes$ shape

Minimizing $||Y - A\,S||_2^2$ under (non-differentiable) constraints (arXiv:1708.09066)

SCARLET: Fast, flexible, extendable, open source (Melchior+, in prep.)

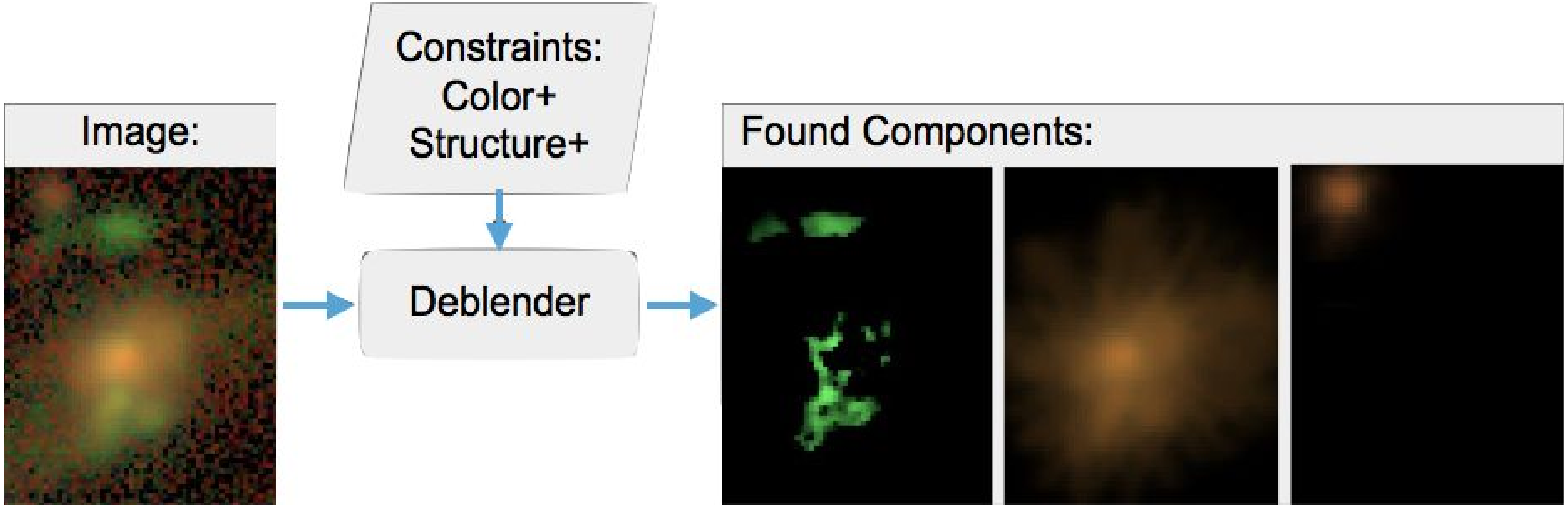
For LSST & Euclid & WFIRST:

- PSF-matched, forced photometry across all bands
- Multiple components per galaxy possible
- Can deal with transients if told

Algorithmic Challenges: Detection

Resolution vs Sensitivity vs Variability

Solution: Detection in multi-band, multi-resolution data frames



Melchior+ in prep.

Algorithmic Challenges: Data compression

LSST has ~200 of exposures in each band, a lot more than Euclid, WFIRST

Which tasks can be done on coadds (what do I mean by “coadd”)?

- Detection for anything that isn’t moving
- Photometry for anything that isn’t variable
- Shapes?

Collapsing the time axis

- sufficient for many users
- allows for self-calibration (e.g. Balrog, metacalibration),
- allows more complex algorithms to be run

Defining Architecture

c.i	c.ii	c.iii	c.iv	c.v
Infrastructure, networking, hardware, interface				
Dawson & A. Smith (STScI)	Dawson & A. Smith	Rusholme (IPAC)	Appleton (IPAC)	Teplitz (IPAC)
Schneider, Fox, Groom, Ebert	Fox, Flynn, Ebert	Smith, Fox, Groom, Berriman	Smith, Fox, Wachter, B. Lee, Rusholme, Berriman	Smith, Fox, Wachter, Groom, Rusholme, Berriman

I. Point design of computing architecture for most accurate solutions
II. Point design of computing architecture for exploration of alternative hypotheses
III. Optimized design of computing architecture
IV. Architecture for use by broader community of users/science applications
V. Definition of science/data services to meet needs of broader community

We won't just have more data with 'Stage IV' surveys.
- We're in an era with qualitatively new computing capabilities

<u>Dijkstra's Law</u>

A quantitative difference is also a qualitative difference if the quantitative difference is greater than an order of magnitude.

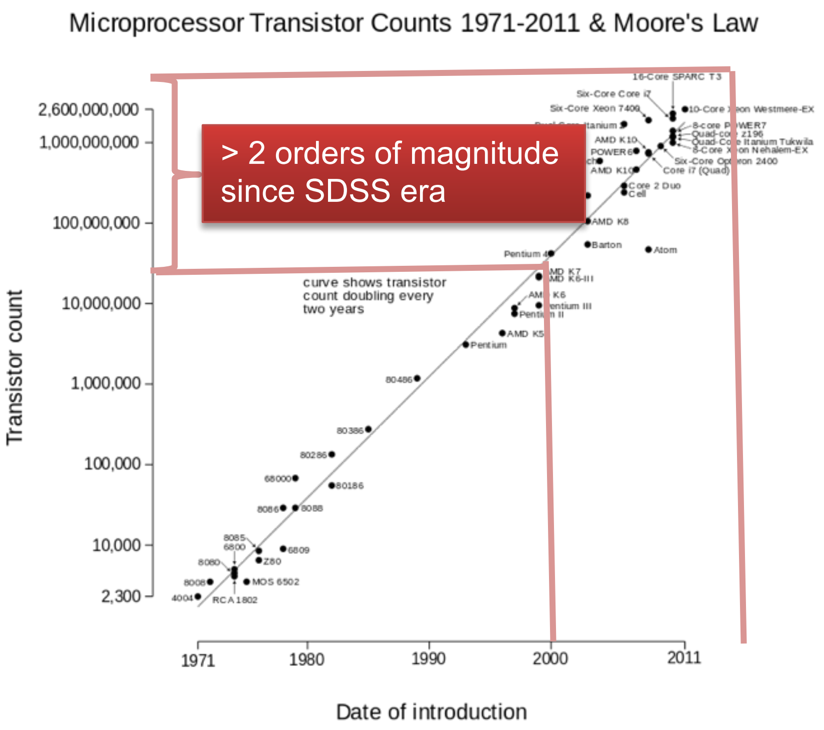


Figure credit: Wikipedia

Qualitative changes in computing enable new scientific methods

“…predictive simulation has brought together theory and experiment in such a compelling way that it’s fundamentally extended the scientific method for the first time since Galileo Galilei invented the telescope in 1609…”

- Mark Seager, CTO for the HPC Ecosystem at Intel

(interview in Inside HPC on June 6, 2016)

Data + Compute convergence in cosmology – DOE ASCR initiative, April 2016

- We're facing **systematics-limited** measurements
  - End-to-end simulations of the experiment are the best approach to improve accuracy & precision
  - Ties data and simulation more intricately than in past cosmology pipelines
- Image and catalog **summary statistics** are no longer good enough to meet next generation science requirements
  - Probabilistic hierarchical models and related machine-learning approaches show promise but are much more computationally intensive
  - Potential changes to the traditional 'facility' / 'user' separate analysis stages

Removing the line between 'analysis' and 'simulation'.

How will this effort enhance our current knowledge of dark energy?

In overlap area(s)

- Systematics control (important since systematics limited)
- Improved detection catalog
- Robust photometry for LSST at full-depth
- Cleaner photo-z training samples

Outside of overlap

Priors/calibrations for detection completeness, photo-z/shape contaminations

Recent review

Rhodes+2017 “Scientific Synergy Between LSST And Euclid”

How does the idea complement other efforts?

Unclear which other efforts qualify

Cost estimate

TBD, key aspect of Tri-Agency Group report (due 2019)

Timeline

Survey starts: LSST 2022, Euclid 2021+, WFIRST 2025+

LSST 1-year depth well matched to Euclid, but wide-first vs deep-first

Final overlap 7000 sq. deg (or more with LSST extensions in North or South)

LSST 10-year depth well matched to WFIRST HLS, overlap complete

	u	g	r	i	z	*Note*
LSST	23.7	24.9	24.4	24.0	23.5	Depth reached per single visit
Euclid	25.4	25.6	25.3	25.3	24.9	Average requirement for ground based complement
LSST visits	23	4	5	11	13	Number of visits to reach the *Euclid* depths
LSST exp. (mn)	11.5	2	2.5	5.5	6.5	Total integration needed (open shutter time)
LSST time (mn)	15	2.6	3.3	7.2	8.5	Total budget (integration plus overheads)

$\longmapsto$ *We quote the total magnitude for a point source at 5-σ with a PSF fitting photometry*

Rhodes+ 2017

Key Technical and Logistical Obstacles

Technical

R&D needed for forecasts, algorithmic, and computing

Logistical

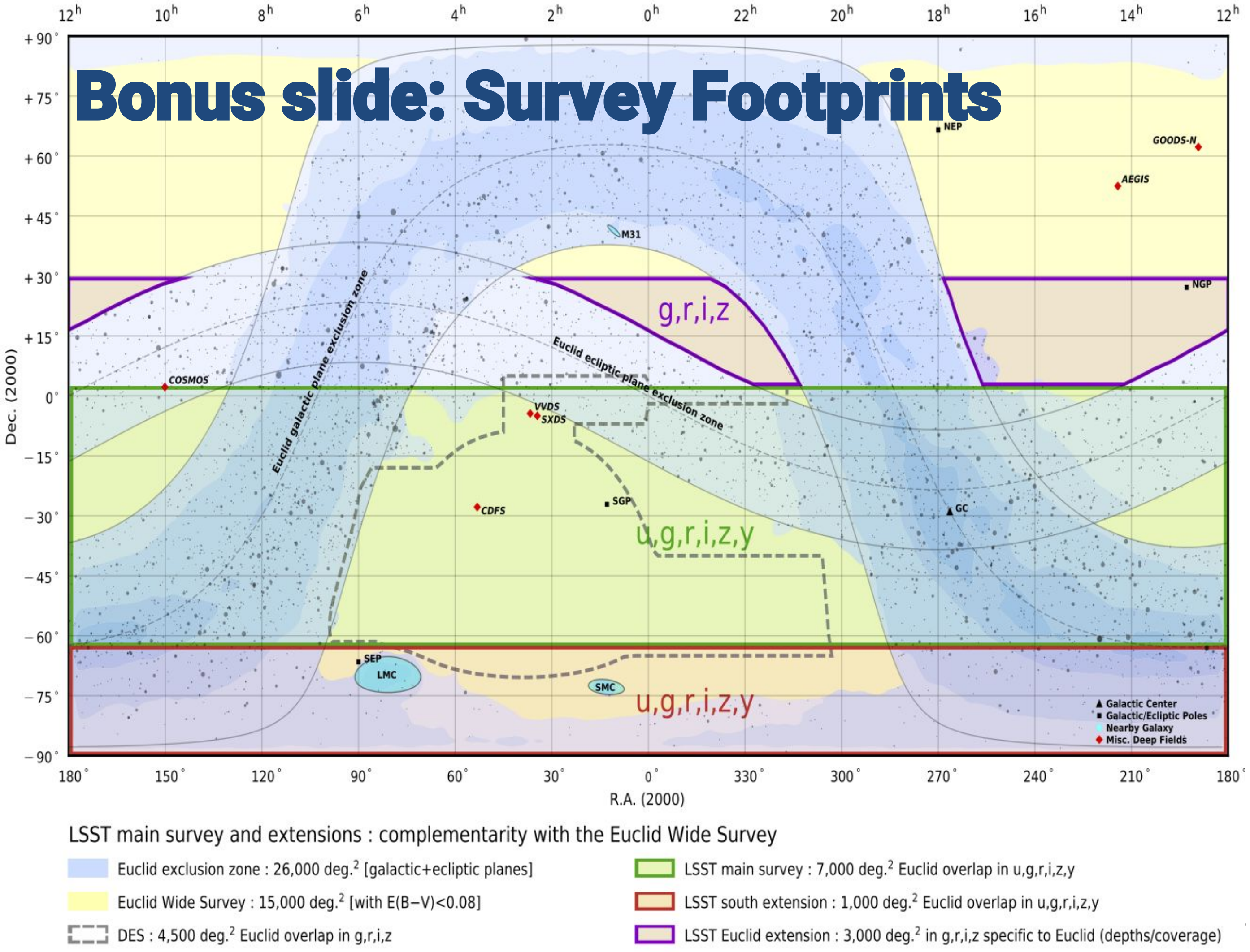
Data formats/procedures for survey-specific processing

Data sharing policies of three surveys (projects and science users)

Big Plus

Tri-Agency Group represents/coordinates shareholders from every survey

Bonus slide: Survey Footprints



Rhodes+ 2017