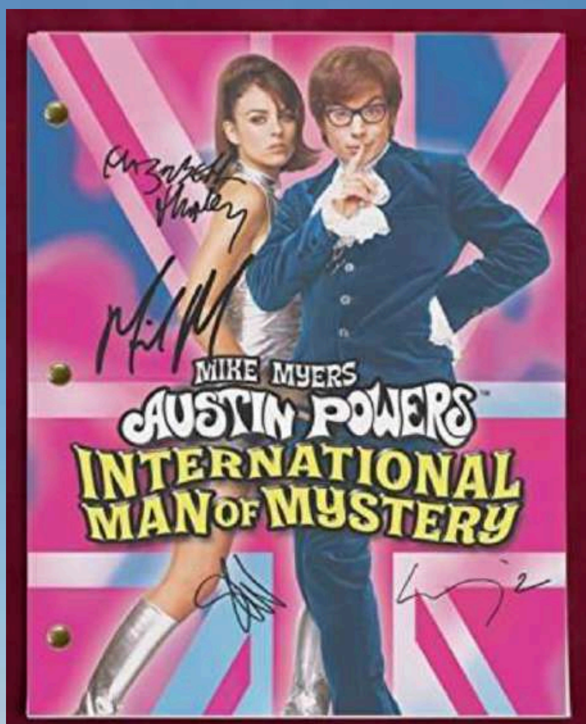
# LSST survey: millions and millions of quasars

Željko Ivezić (pronounced as Bill)
LSST/University of Washington
LSST Project Scientist and Deputy Director




FIRST LIGHT: STARS, GALAXIES AND BLACK HOLES IN THE EPOCH OF REIONIZATION
Advanced School, São Paulo, Brazil, July 28 – August 7, 2019

# OUTLINE

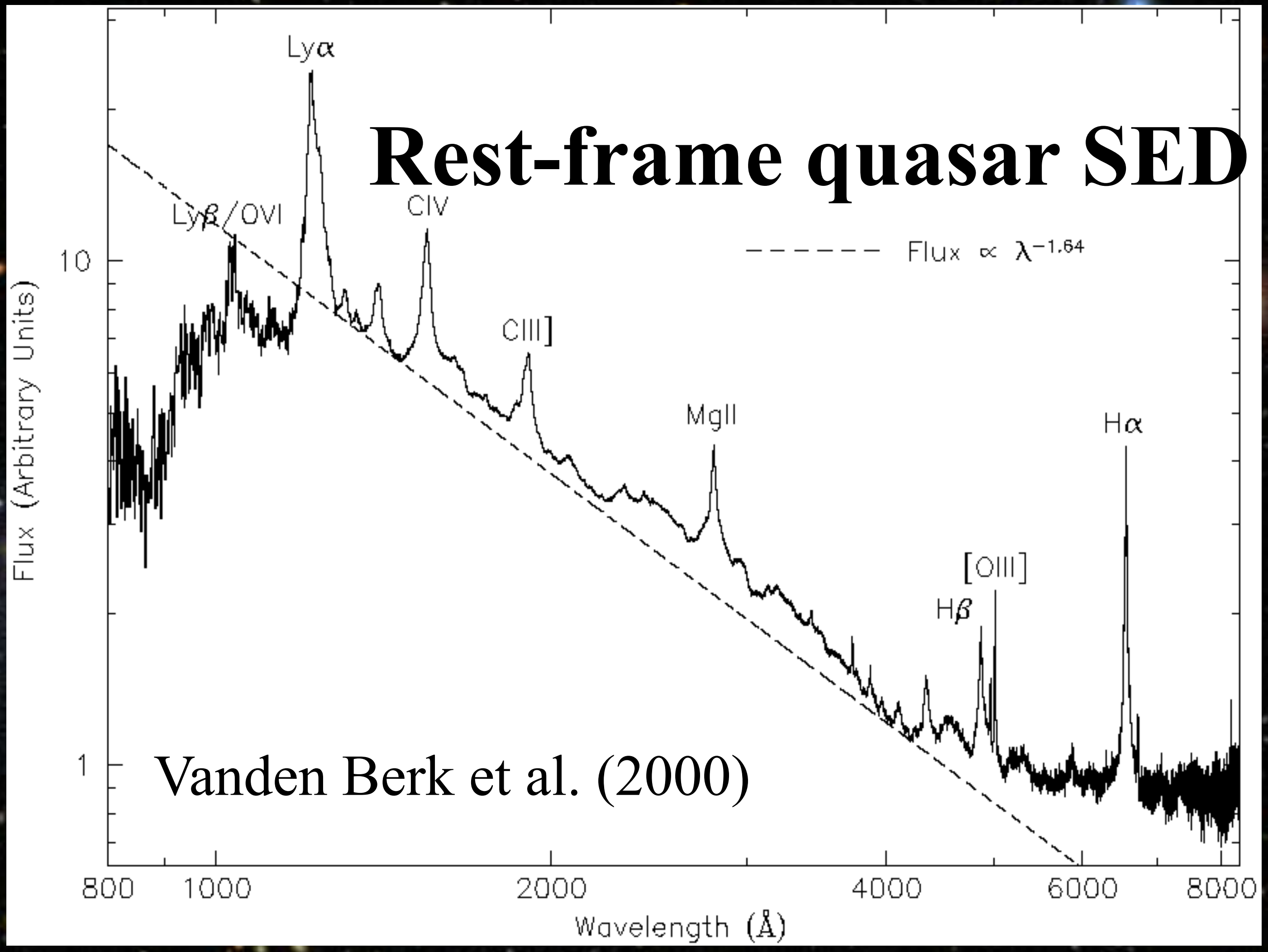
Terminology hereafter: a **quasar** is a point source in ground-based seeing, while an **AGN** is a resolved galaxy

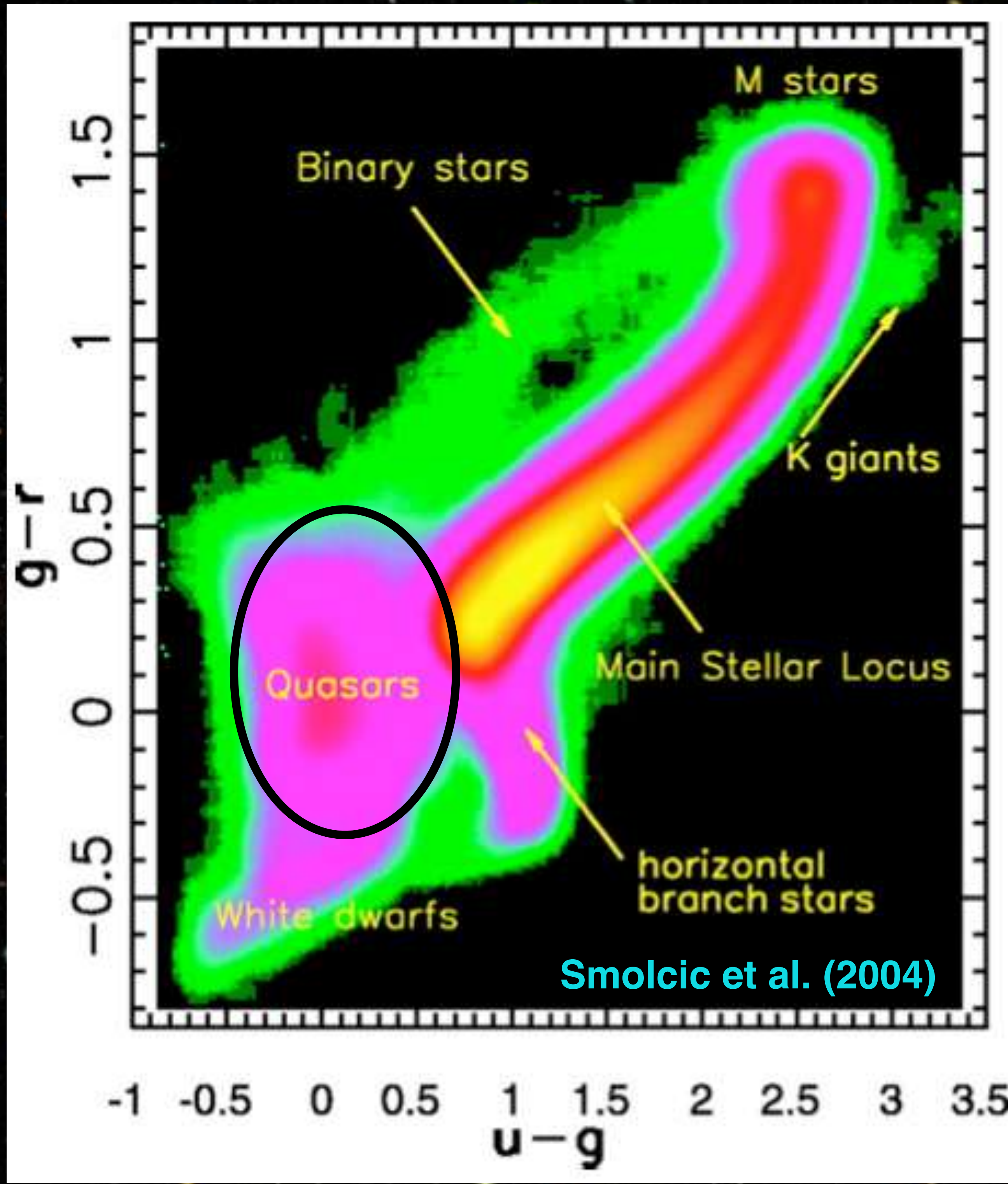
- **Finding quasars/AGNs with SDSS**
  color selection of quasars (but skipping AGN selection using galaxy spectra)
- **Brief introduction to LSST**
  science drivers, system overview, expected survey data products, status report
- **Finding quasars/AGNs with LSST**
  photometry (colors, variability) **and** astrometry (no proper motion, DCR)

# Finding Quasars…

**Quasars with z<2.2 have UV excess (e.g. U-B, u-g) compared to stars with similar blue colors (e.g. B-V, g-r):**

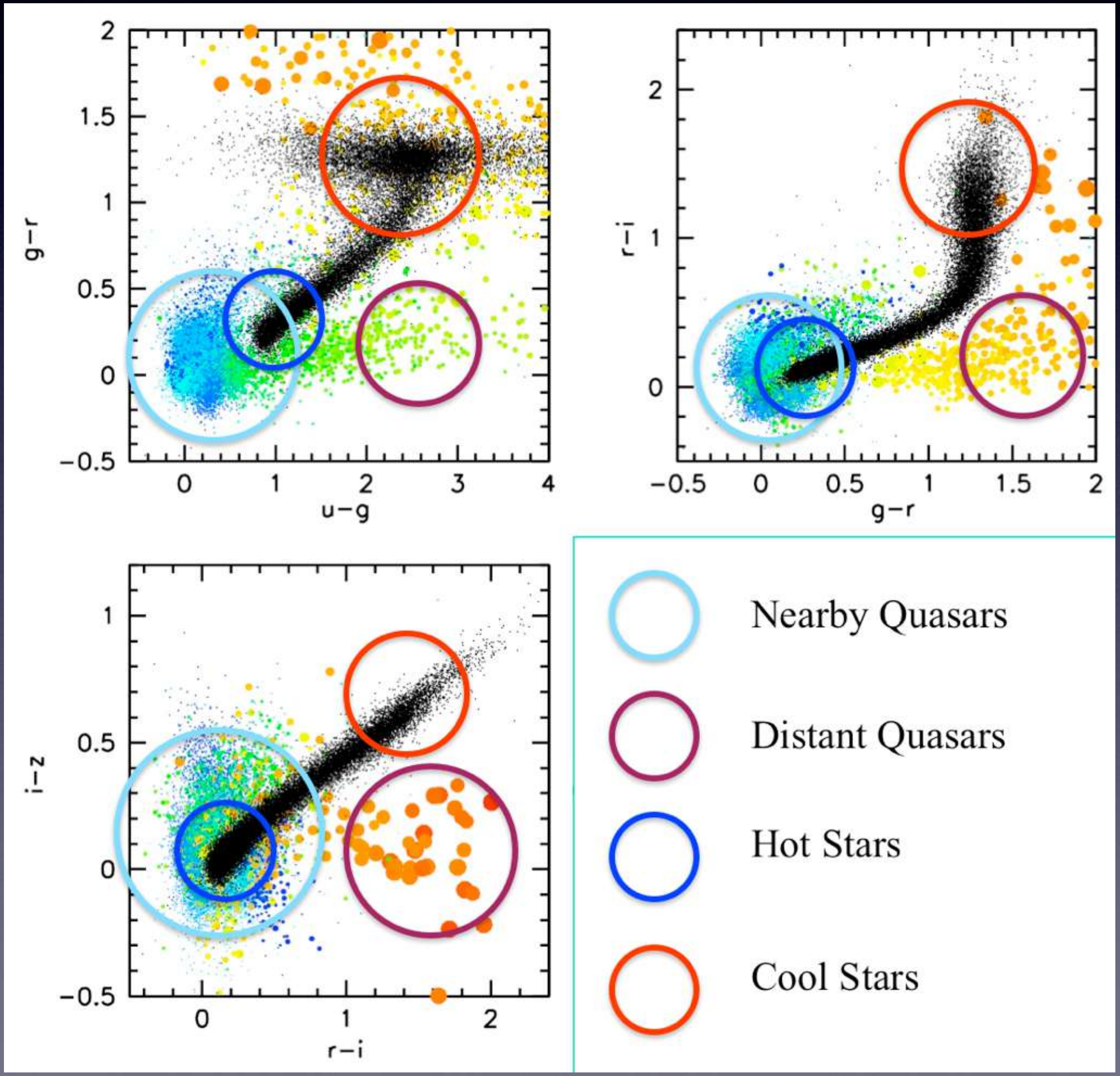
**Observed SEDs greatly differ because a given observed wavelength range samples varying rest-frame wavelength range**



# Sloan Digital Sky Survey Quasar Catalogs:

- about 1/4 of the sky, spectroscopically confirmed
- SDSS I & II: 105,783 (Schneider et al. 2010)
- SDSS III: + 78,086 (Paris et al. 2012)




- mostly color-selected, but also radio (FIRST), see Richards et al. (2002)
- over half with i<19
- 56 with z>5

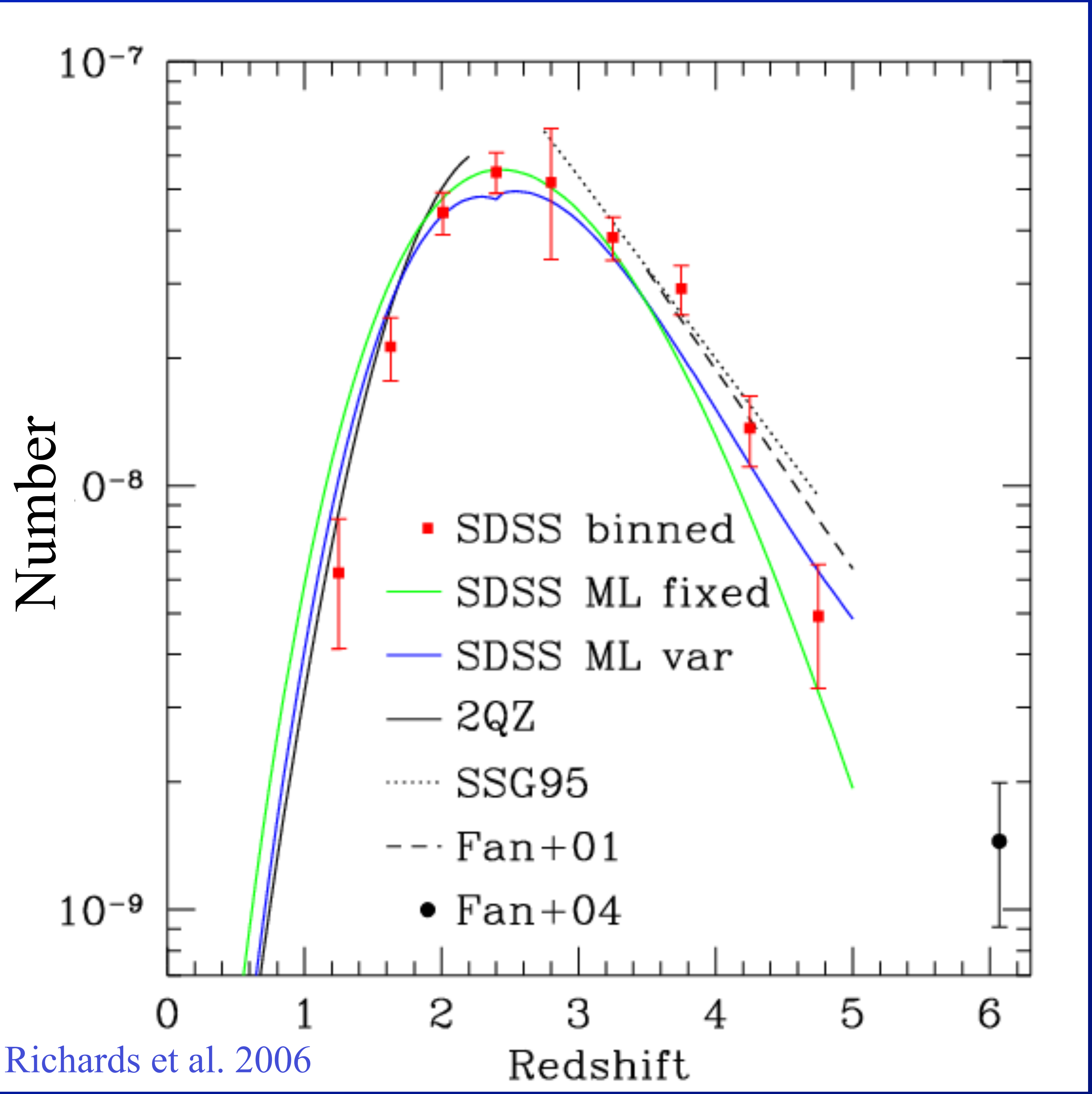
**~200,000 quasars!**

**Completeness: >90%**

**Efficiency of color selection: >65%**

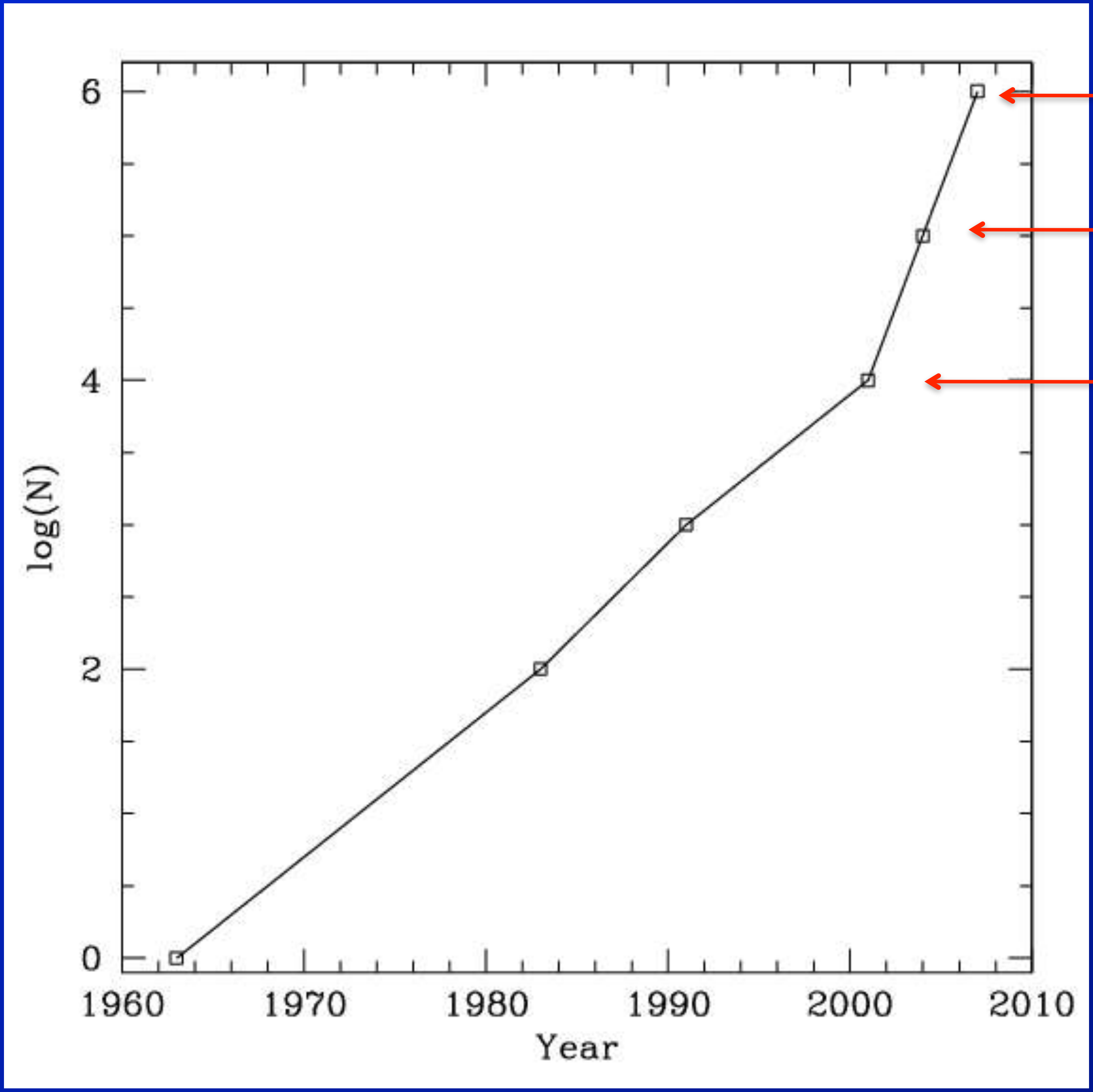
# Quasar Luminosity Function

Previous measurements were for nearby or far away quasars, but not both.  SDSS spanned the whole range.




Richards et al. 2006

# Breaking the 1,000,000 Mark




Richards et al. 2009

Richards et al. 2004

Croom et al. 2000
Schneider, Richards et al. 2002

**Can we get to 10 million?**

# Can we get to 10 million?

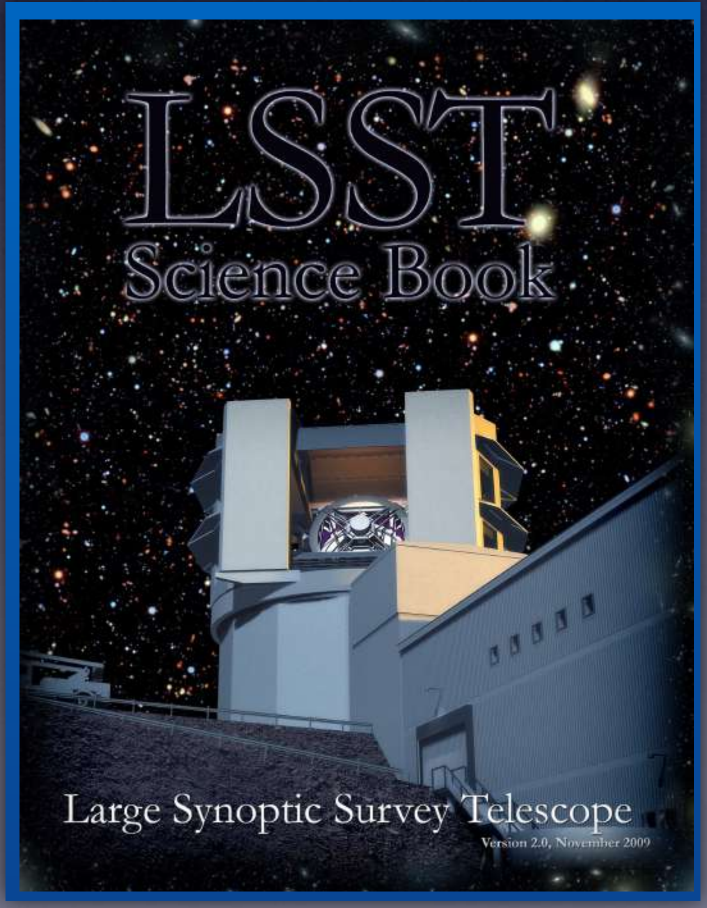
# LSST Science Themes

- 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

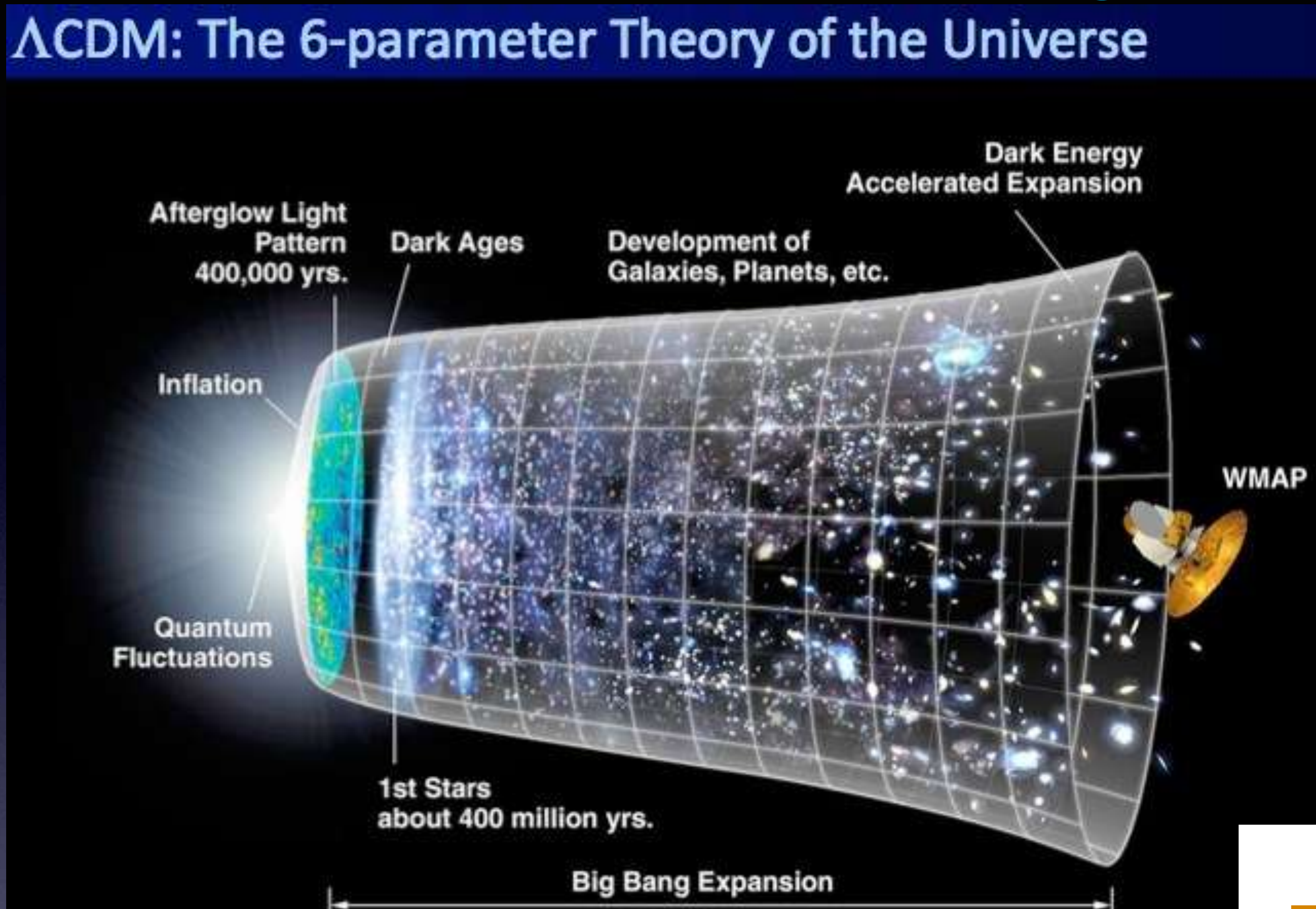
**245 authors, 15 chapters, 600 pages**

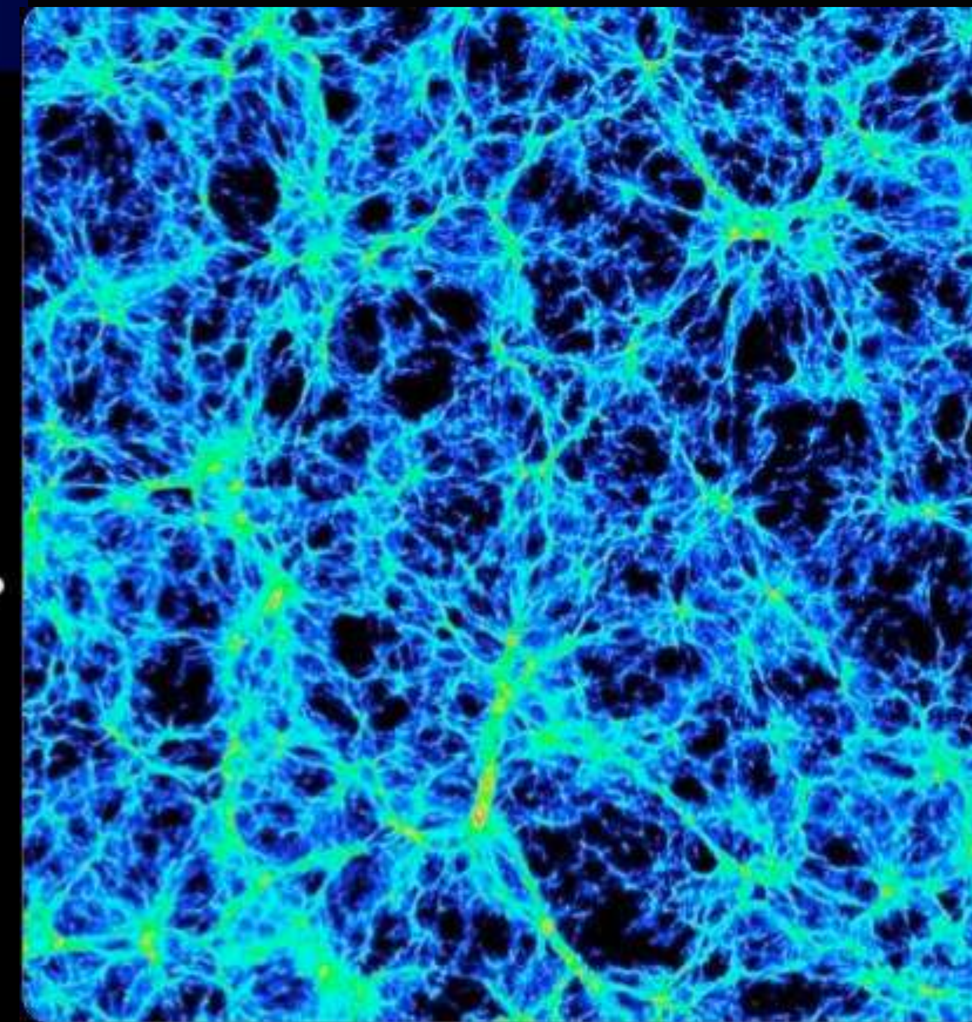


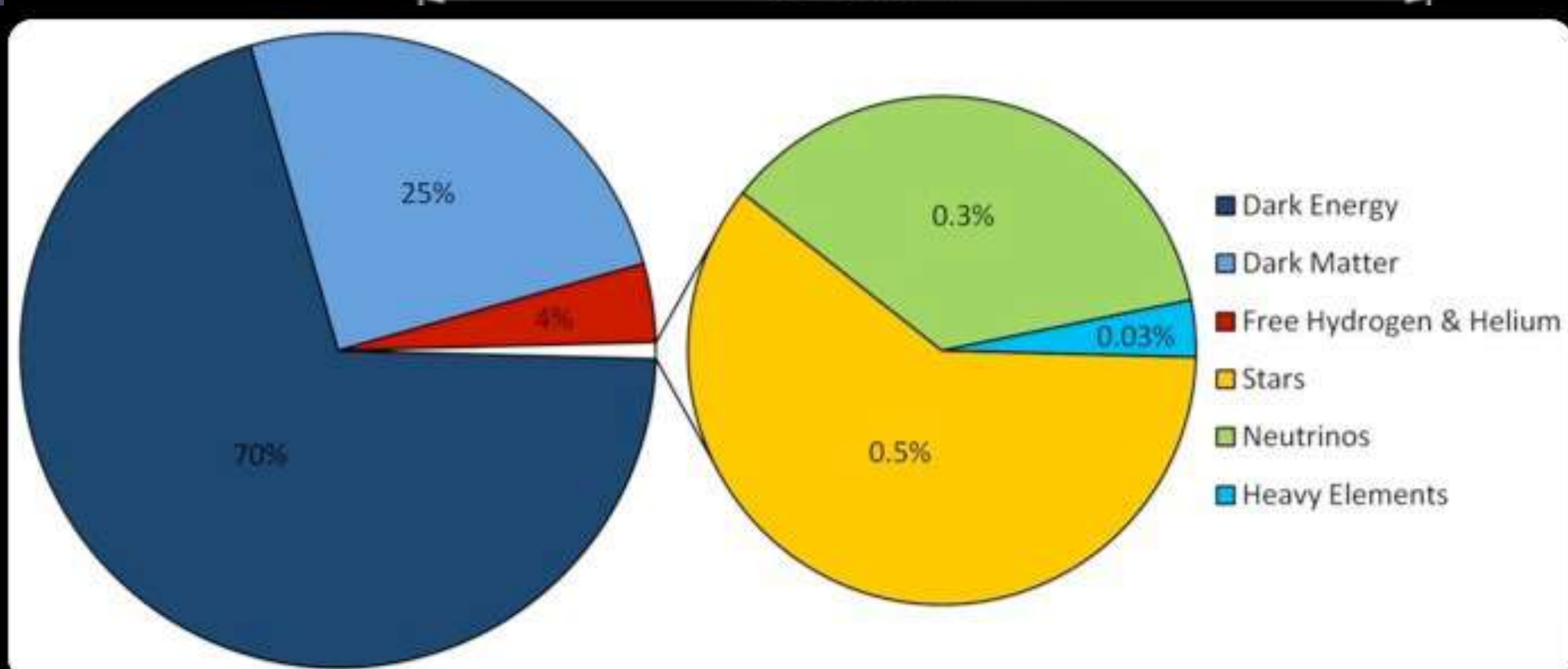
# Required system characteristics

- Large primary mirror (at least 6m) to go faint and to enable short exposures (30 s)
- Agile telescope (5 sec for slew and settle)
- Large field of view to enable fast surveying
- Impeccable image quality (weak lensing)
- Camera with 3200 Mpix
- Sophisticated software (20,000 GB/night, 20 billion objects, 20 trillion measurements)

# New Cosmological Puzzles






The modern cosmological models can explain all observations, but need to **postulate** dark matter and dark energy (though gravity model could be wrong, too)

# Modern Cosmological Probes

- Cosmic Microwave Background (the state of the Universe at the recombination epoch, at redshift ~1000)
- Weak Lensing: growth of structure
- Galaxy Clustering: growth of structure
- Baryon Acoustic Oscillations: standard ruler
- Supernovae: standard candle

**Except for CMB, measuring H(z) and growth of structure G(z)**
**$H(z) \sim d[\ln(a)]/dt$, $G(z) = a^{-1}\delta\rho_m/\rho_m$, with $a(z) = (1+z)^{-1}$**

# Cosmology with LSST: dark energy vs. modified gravity

- Even for a model with modified gravity, it is possible to assume that GR is correct and always find DE with suitable w(z) to explain data for H(z). 😱

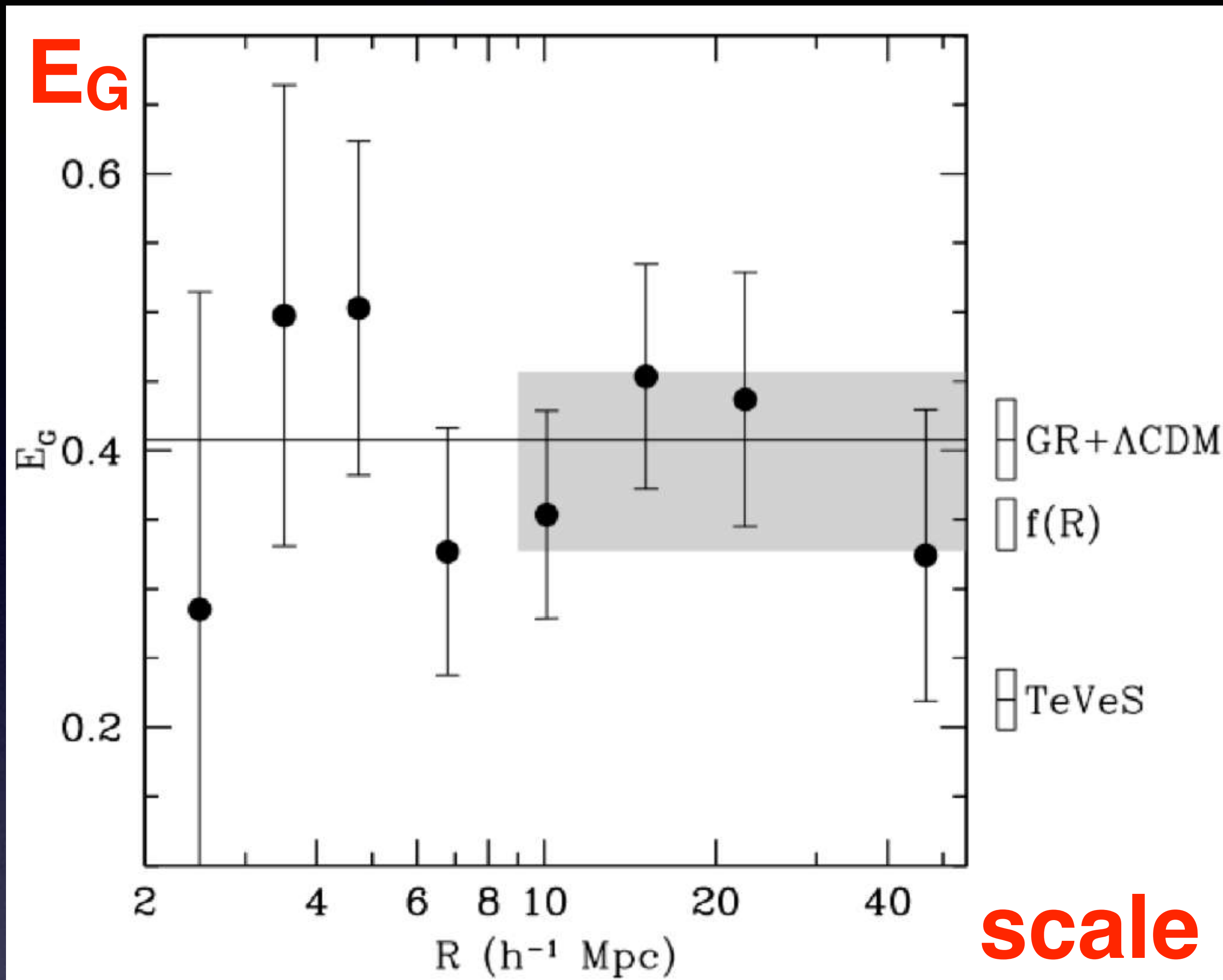
- However, the growth of structure will be different and thus when both H(z) and G(z) are measured, the degeneracy can be broken and DE vs. modified gravity models distinguished (Jain & Zhu 2008, PhysRevD 78, 063503)

$$ds^2 = -(1 + 2\psi)\ dt^2 + (1 - 2\phi)\ a^2(t)\ d\vec{x}^2$$

- φ is the curvature perturbation and ψ is the potential pertur.

- In General Relativity φ = ψ in the absence of anisotropic stresses. A metric theory of gravity relates the two potentials above to the perturbed energy-momentum tensor. 😁 φ and ψ can be constrained with astronomical observations.

# Cosmology with LSST: dark energy vs. modified gravity

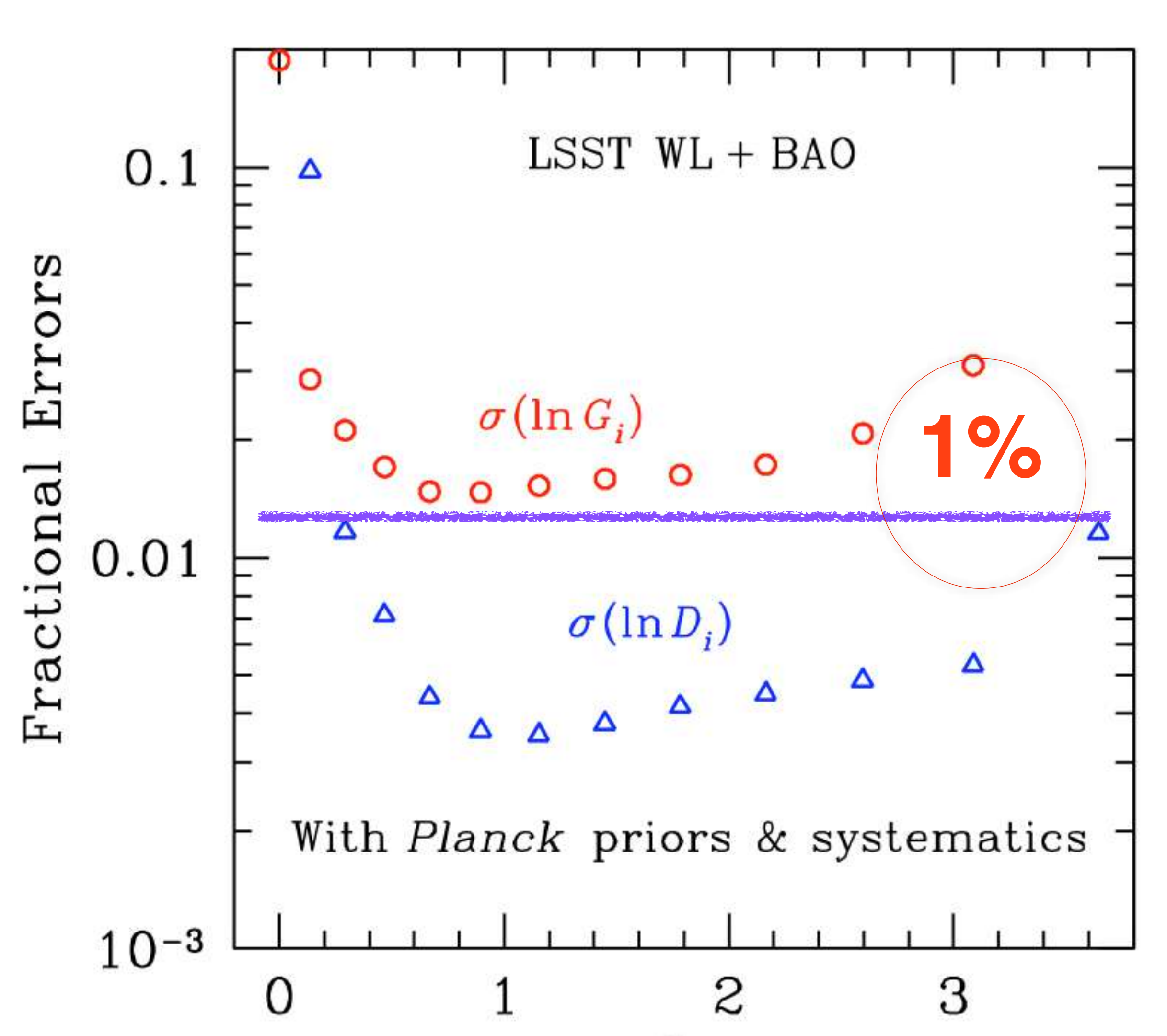



Reyes et al. (2010, Nature 464, 256)

LSST will measure $E_G$ about 10 times more precisely and will be able to rule out a large class of modified gravity theories (or GR!)

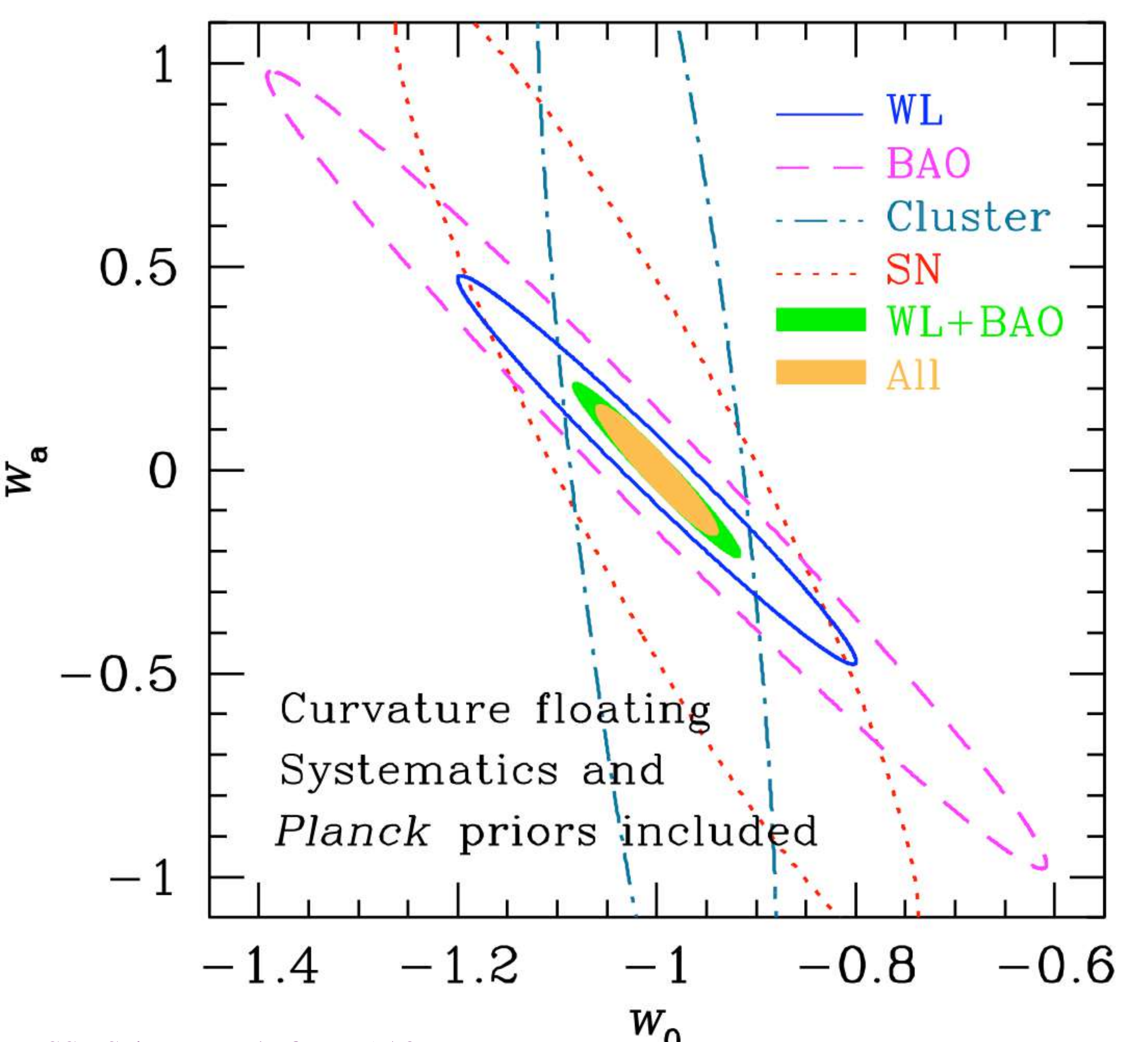
- $E_G$ combines 3 measures of large-scale structure: galaxy-galaxy lensing ($\varphi+\psi$), galaxy clustering ($\varphi$) and galaxy velocities (from galaxy redshifts; measures G(z))
- SDSS data enabled a test of GR at 15% level: it passed!
- SDSS data already excludes a model within the tensor-vector-scalar gravity theory, which modifies both Newtonian and Einstein gravity.
- Five times better precision needed to rule out f(R)

# Cosmology with LSST: high precision measurements




LSST Science Book, figure 15.2

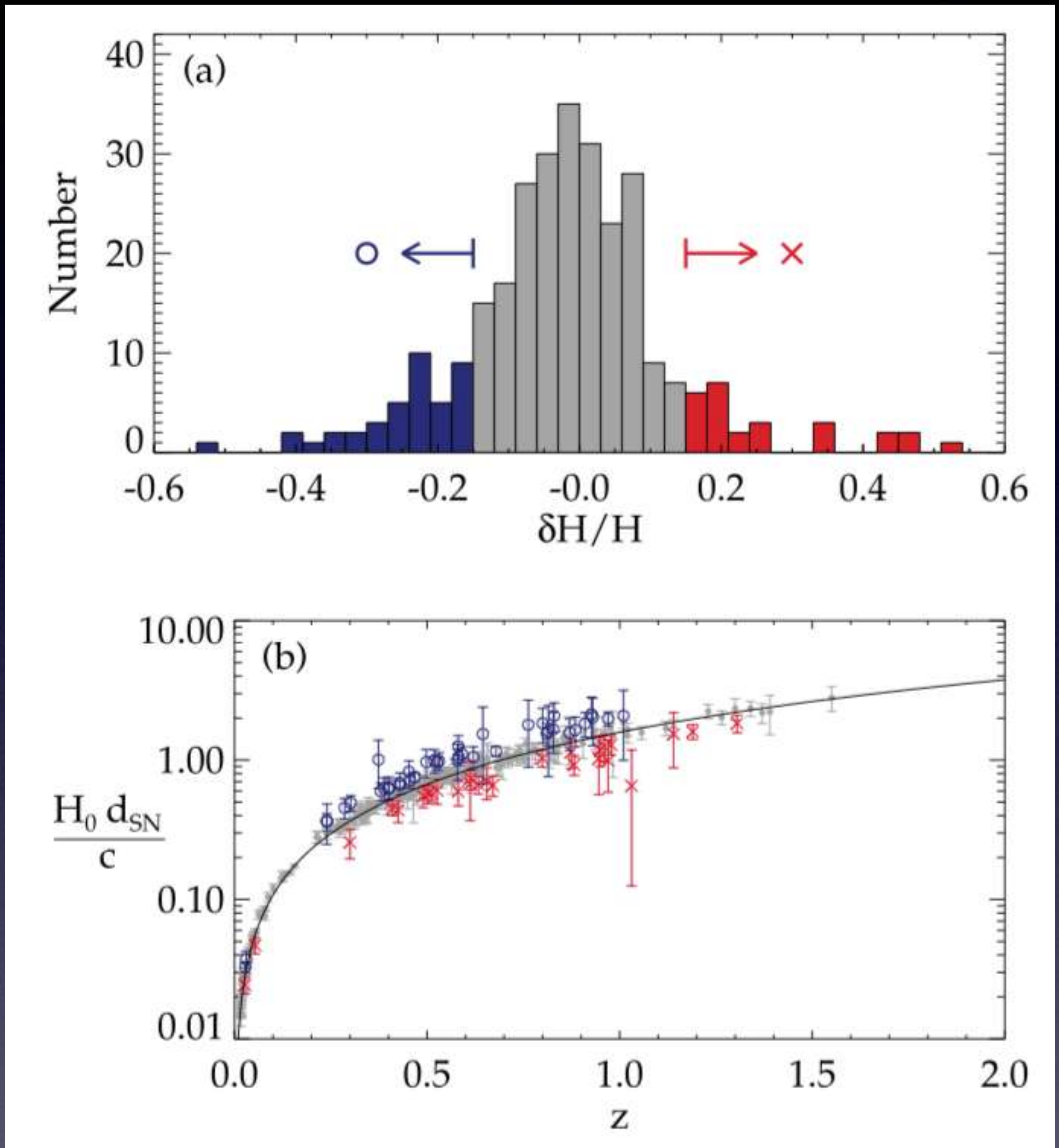
- Measuring distances, H(z), and growth of structure, G(z), with a percent accuracy for 0.5 < z < 3
- Multiple probes is the key!




LSST Science Book, figure 15.3

By simultaneously measuring growth of structure and curvature, LSST data will tell us whether the recent acceleration is due to **dark energy or modified gravity.**

# Cosmology with LSST SNe: is the cosmic acceleration the same in all directions?



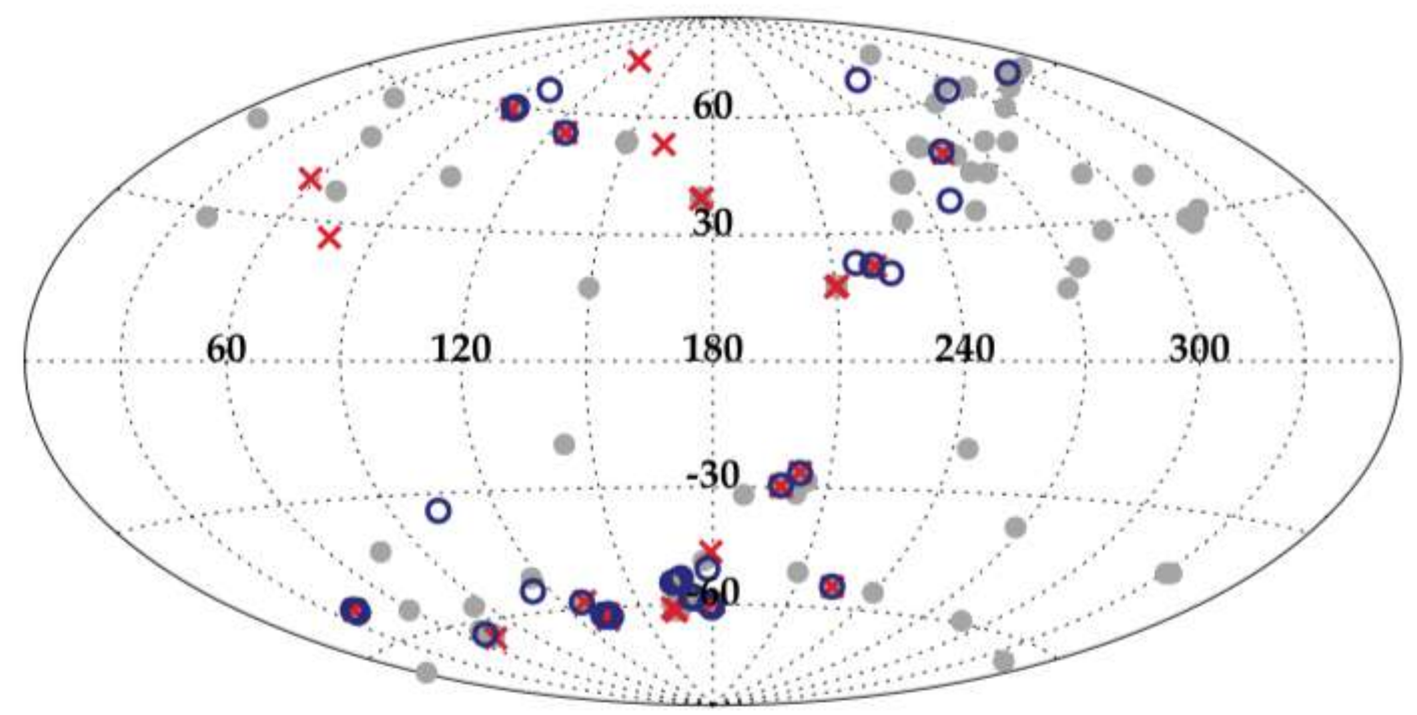




**Figure 1.** A projection of the spatial distribution of the Union SNe Ia sample in Galactic coordinates. Note the relative uniformity of the points, except around the Galactic plane. The symbols correspond to those in Fig. 2, and are explained in Section 3.1.

Cooke & Lynden-Bell (2009, MNRAS 401, 1409)

**Is there spatial structure in the SNe distance modulus residuals for the concordance model?**

- Even a single supernova represents a cosmological measurement!
- LSST will obtain light curves for several million Type Ia supernovae!

## LSST, WFIRST and Euclid are highly complementary missions.

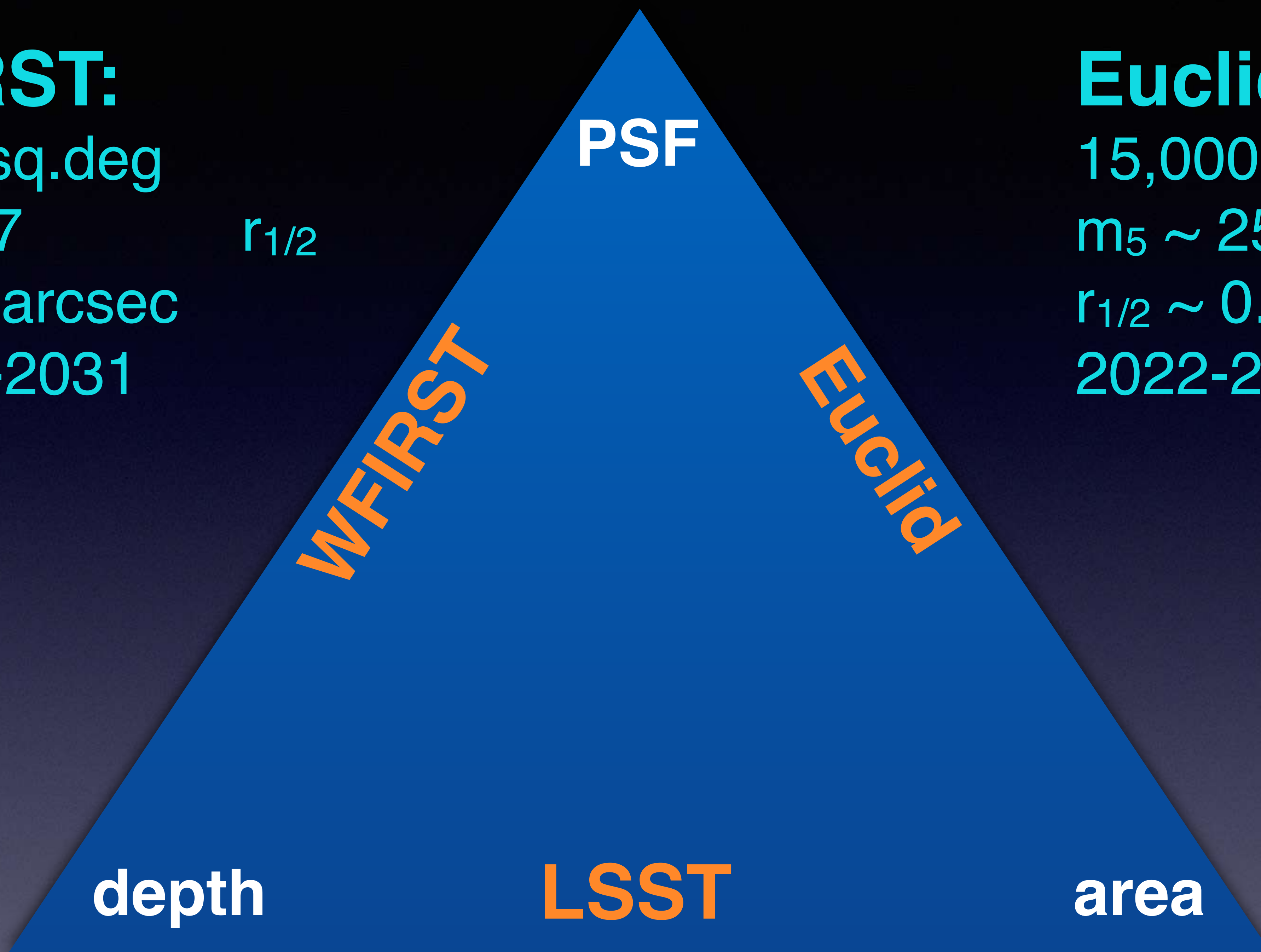
**WFIRST:**

2,200 sq.deg

$m_5 \sim 27$ $r_{1/2}$

$\sim 0.12$ arcsec

~2025-2031




**Euclid:**

15,000 sq.deg

$m_5 \sim 25$

$r_{1/2} \sim 0.13$ arcsec

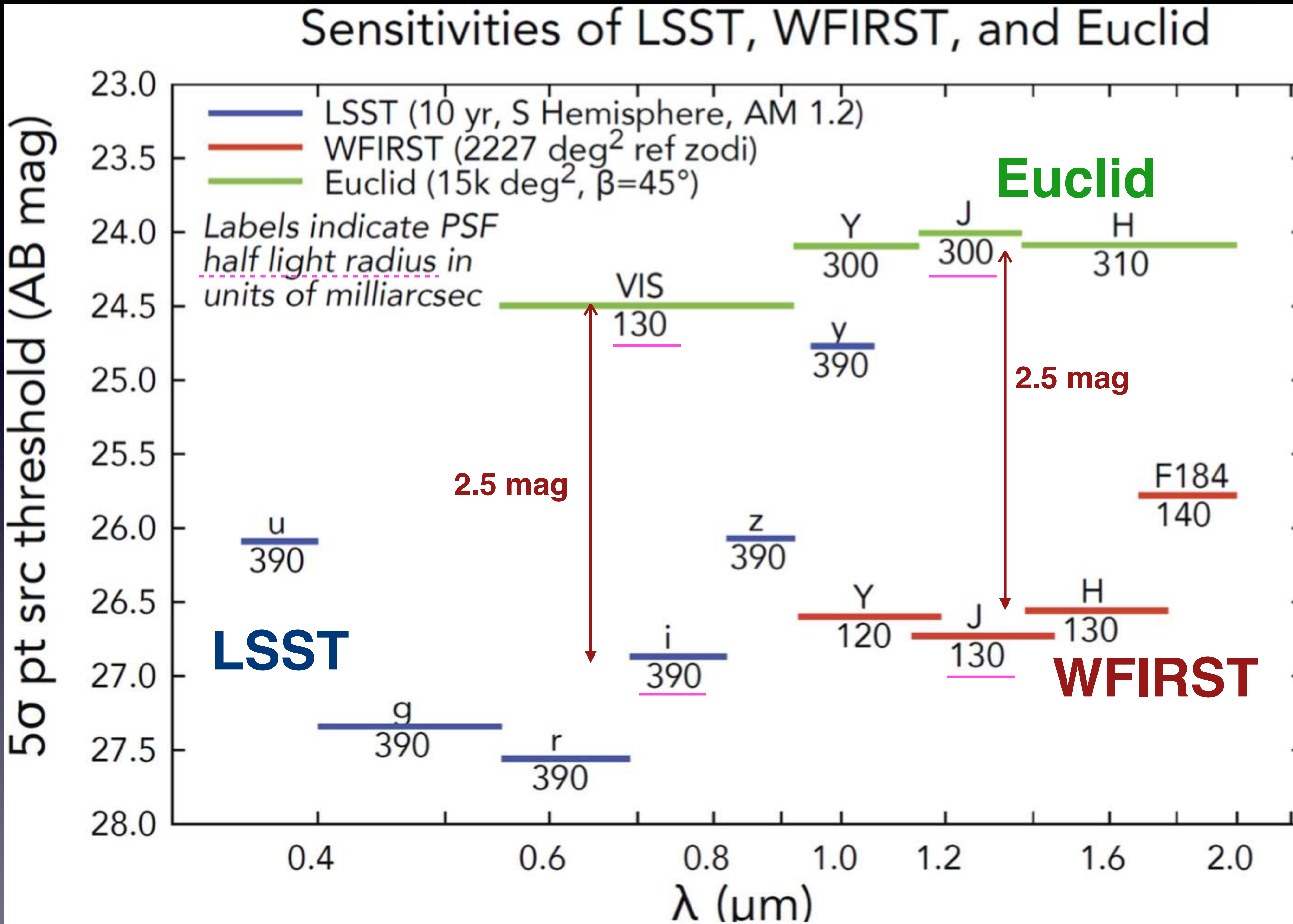
2022-2028

18,000 sq.deg

$m_5 \sim 27$

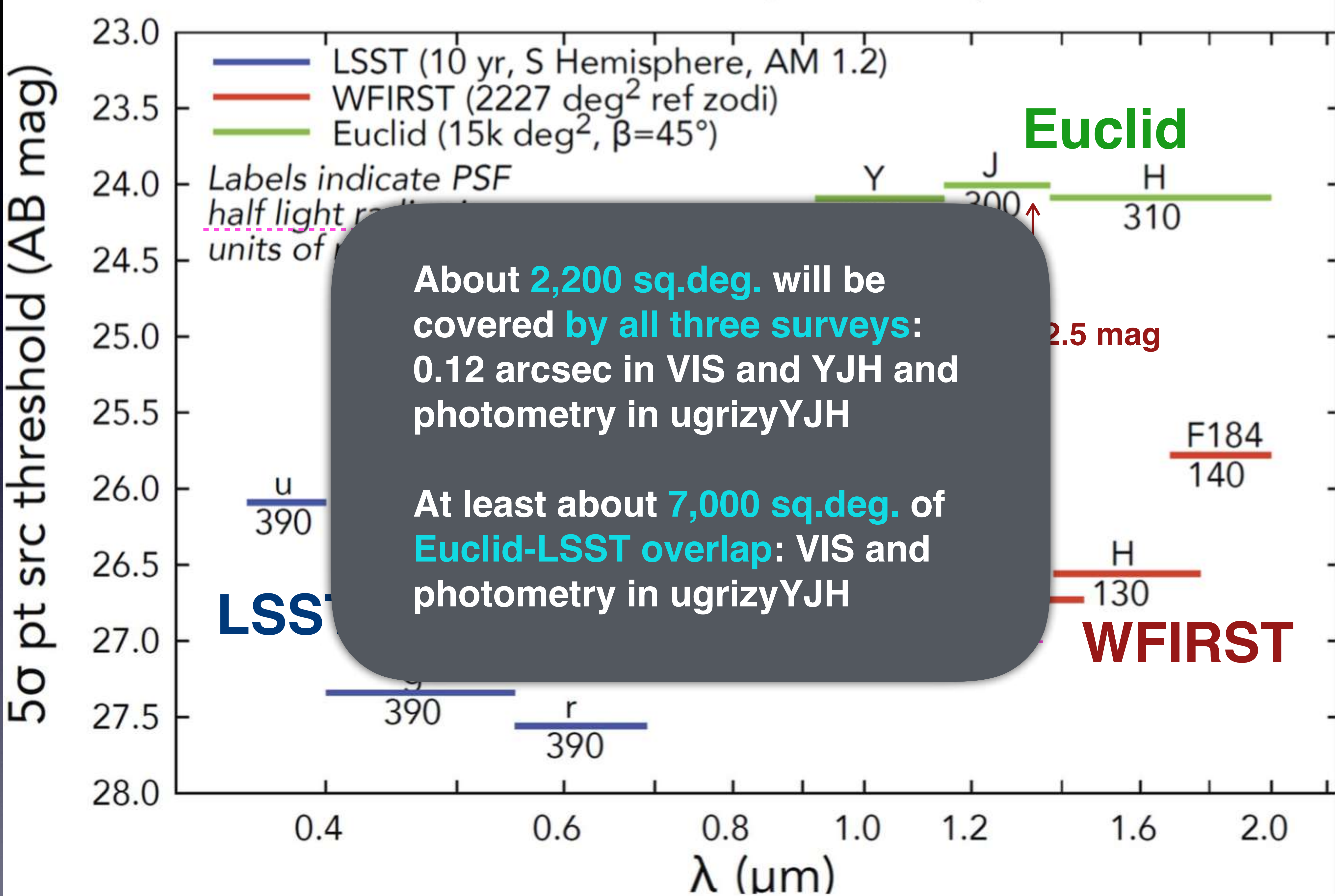
$r_{1/2} \sim 0.4$ arcsec

2022-2032



Sensitivities of LSST, WFIRST, and Euclid
LSST (10 yr, S Hemisphere, AM 1.2)
WFIRST (2227 deg² ref zodi)
Euclid (15k deg², β=45°)
Labels indicate PSF half light radius in units of milliarcsec
Euclid
LSST
WFIRST
2.5 mag
2.5 mag
VIS
130
Y
300
J
300
H
310
y
390
u
390
z
390
i
390
g
390
r
390
Y
120
J
130
H
130
F184
140
5σ pt src threshold (AB mag)
λ (μm)
23.0
23.5
24.0
24.5
25.0
25.5
26.0
26.5
27.0
27.5
28.0
0.4
0.6
0.8
1.0
1.2
1.6
2.0

# Sensitivities of LSST, WFIRST, and Euclid



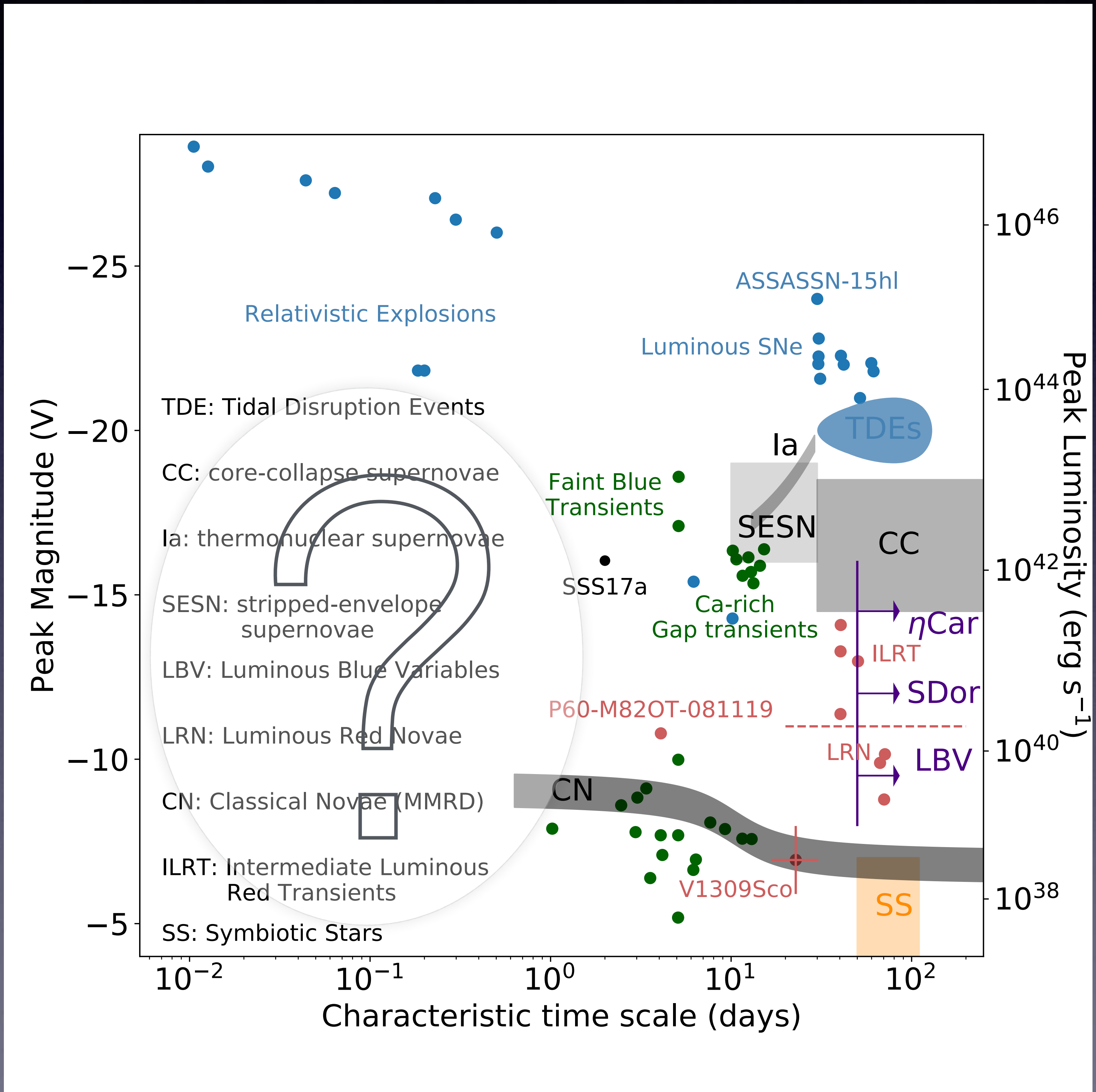

About 2,200 sq.deg. will be covered by all three surveys: 0.12 arcsec in VIS and YJH and photometry in ugrizyYJH

At least about 7,000 sq.deg. of Euclid-LSST overlap: VIS and photometry in ugrizyYJH

# Time Domain: objects changing in time

positions: asteroids and stellar proper motions

brightness: cosmic explosions and variable stars




LSST will extend time-volume space a hundred times over current surveys (new classes of object?): **multi-messenger astrophysics**

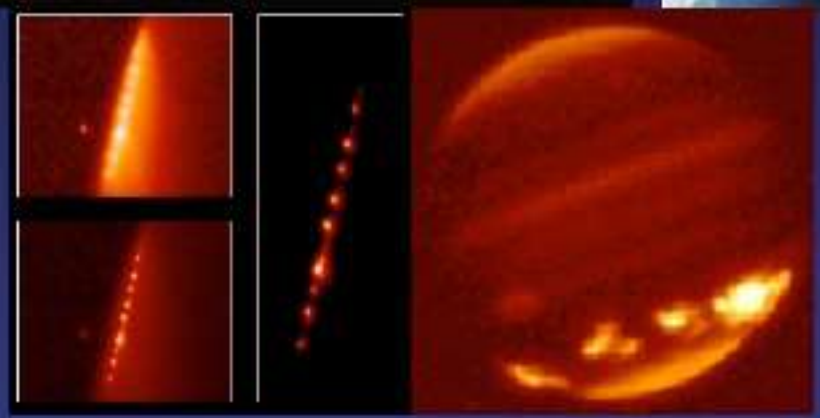
known unknowns

unknown unknowns

# **Killer asteroids:** the impact probability is not 0!

photomontage!

LSST is the only survey capable of delivering completeness specified in the 2005 USA Congressional NEO mandate to NASA (to find 90% NEOs larger than 140m)

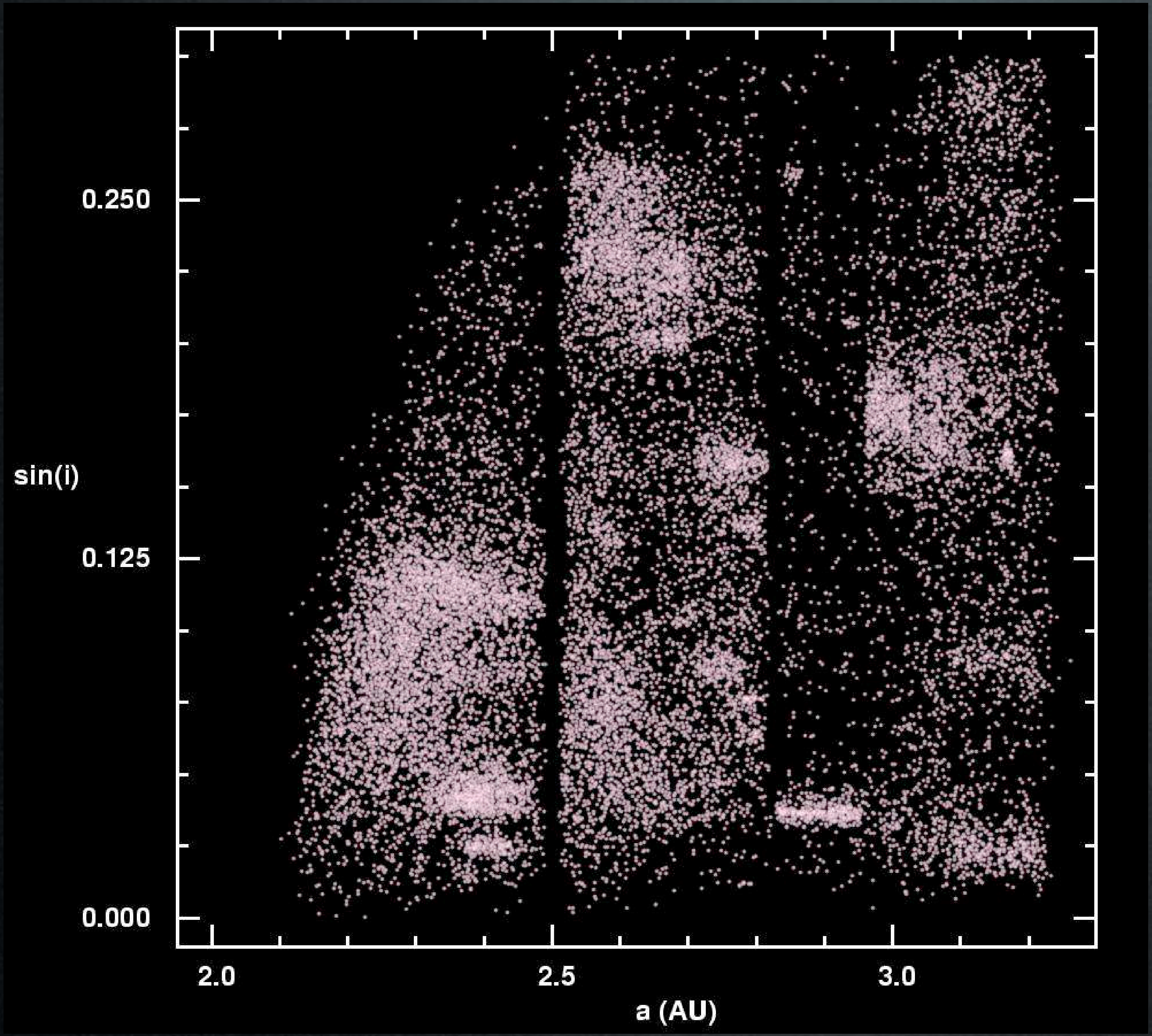


Shoemaker-Levy 9 (1994)

Tunguska (1908)


The Barringer Crater, Arizona: a 40m object 50,000 yr. ago

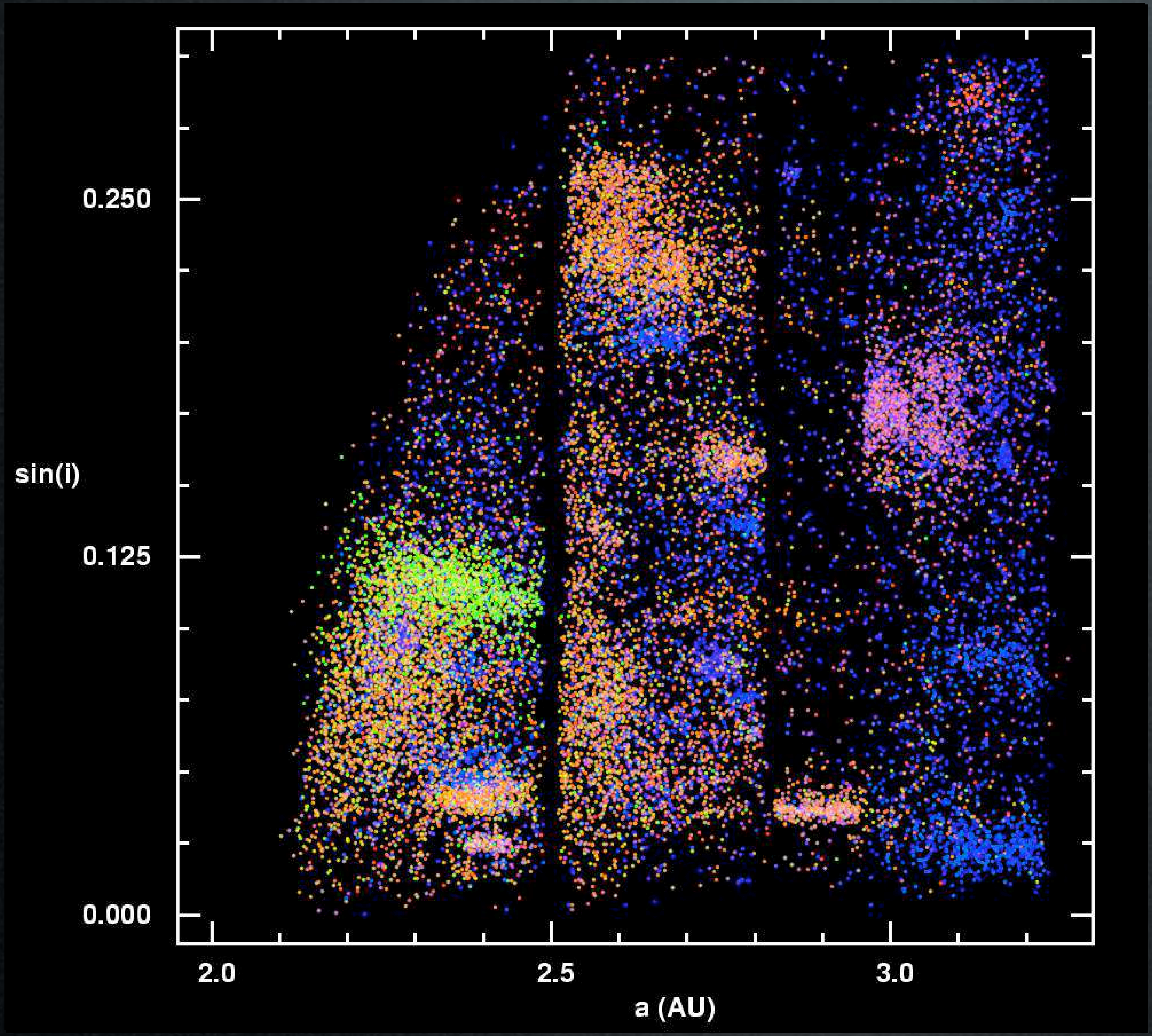
# Main-belt Inventory




30,000
Asteroids with
SDSS colors and
proper
orbital elements

(Ivezic, Juric, Lupton 2002)

# Main-belt Inventory




30,000
Asteroids with SDSS colors and proper orbital elements

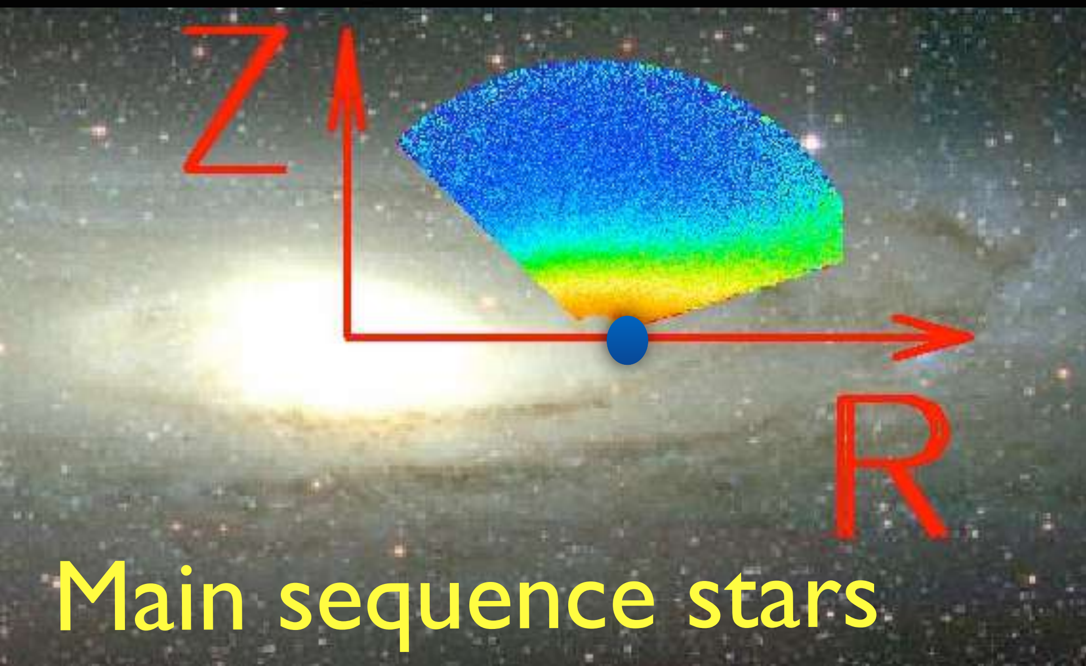
(Ivezic, Juric, Lupton 2002)

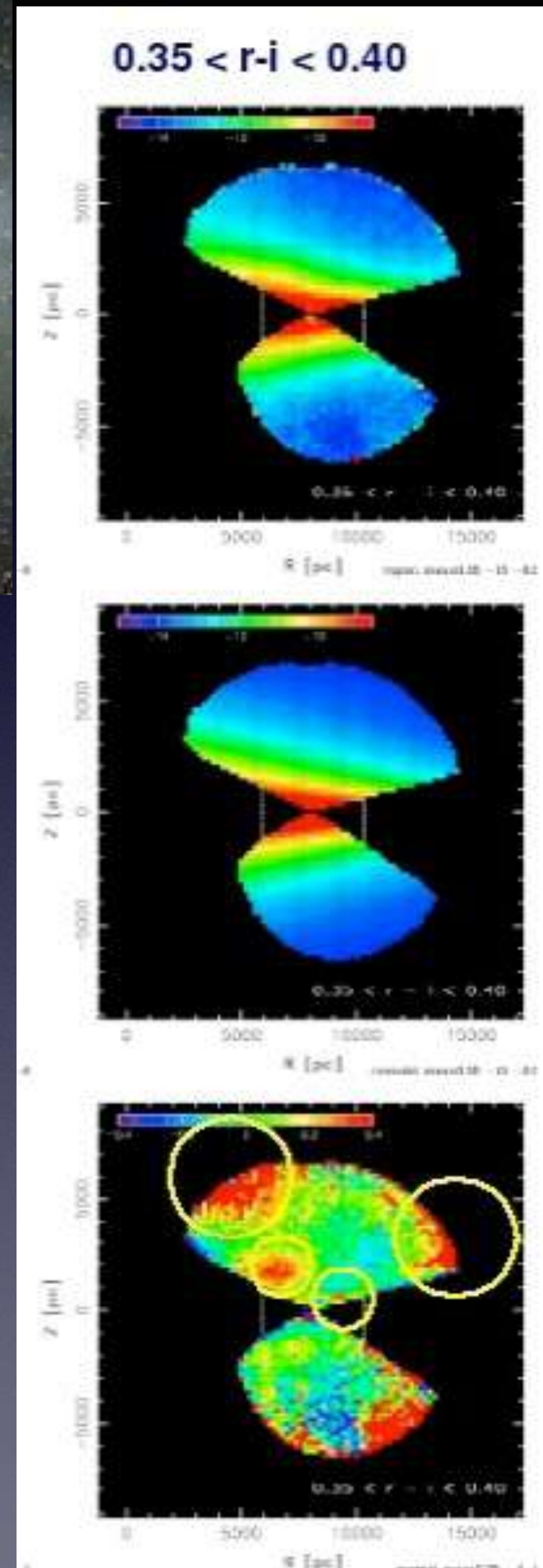
Color-coded with SDSS colors

Colors help with the definition of asteroid families.

LSST will also provide color light curves!

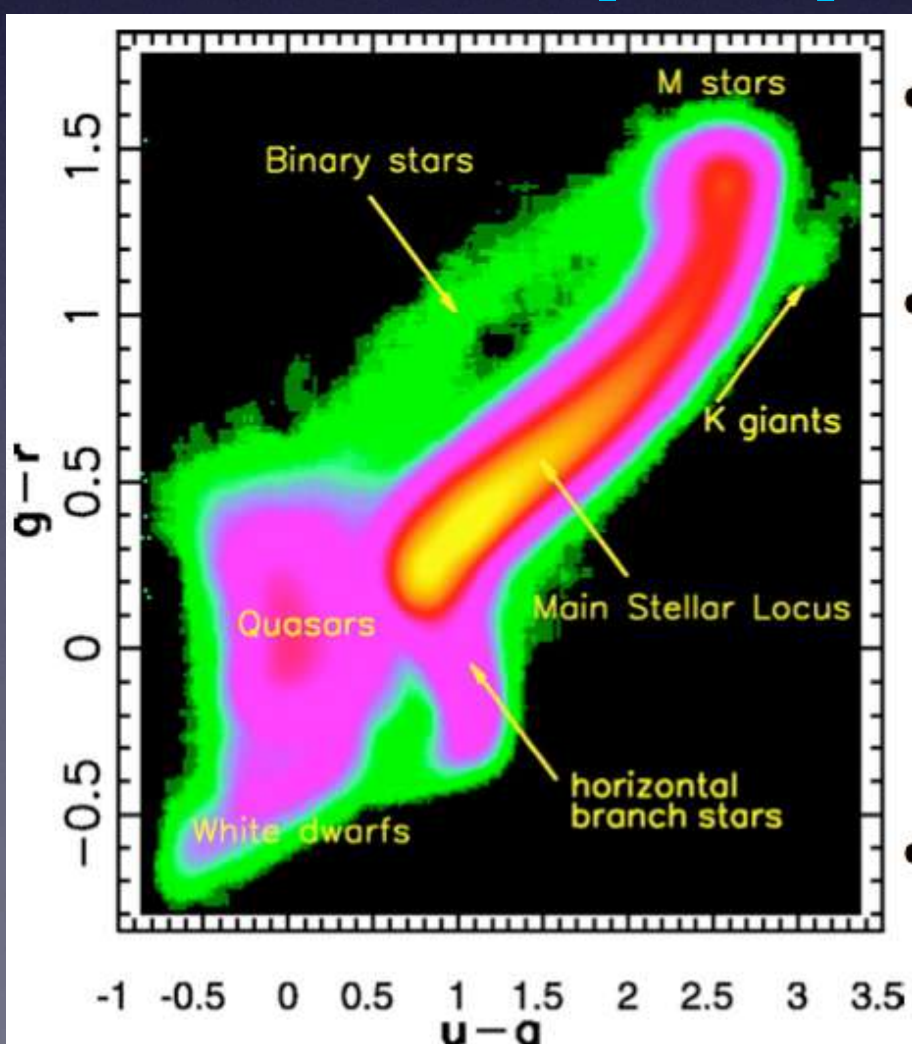
# The Milky Way structure: 20 billion stars, time domain massive statistical studies!

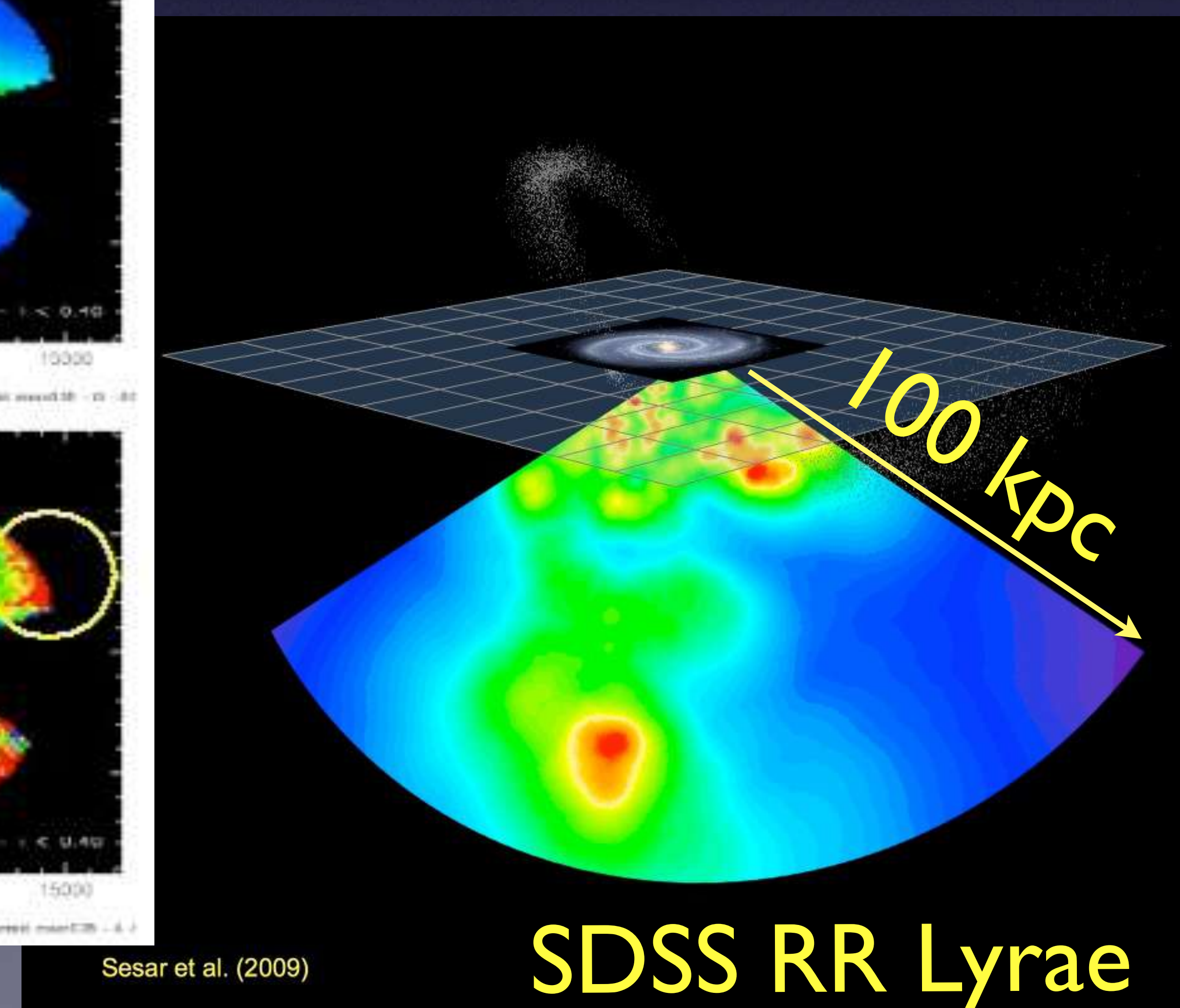


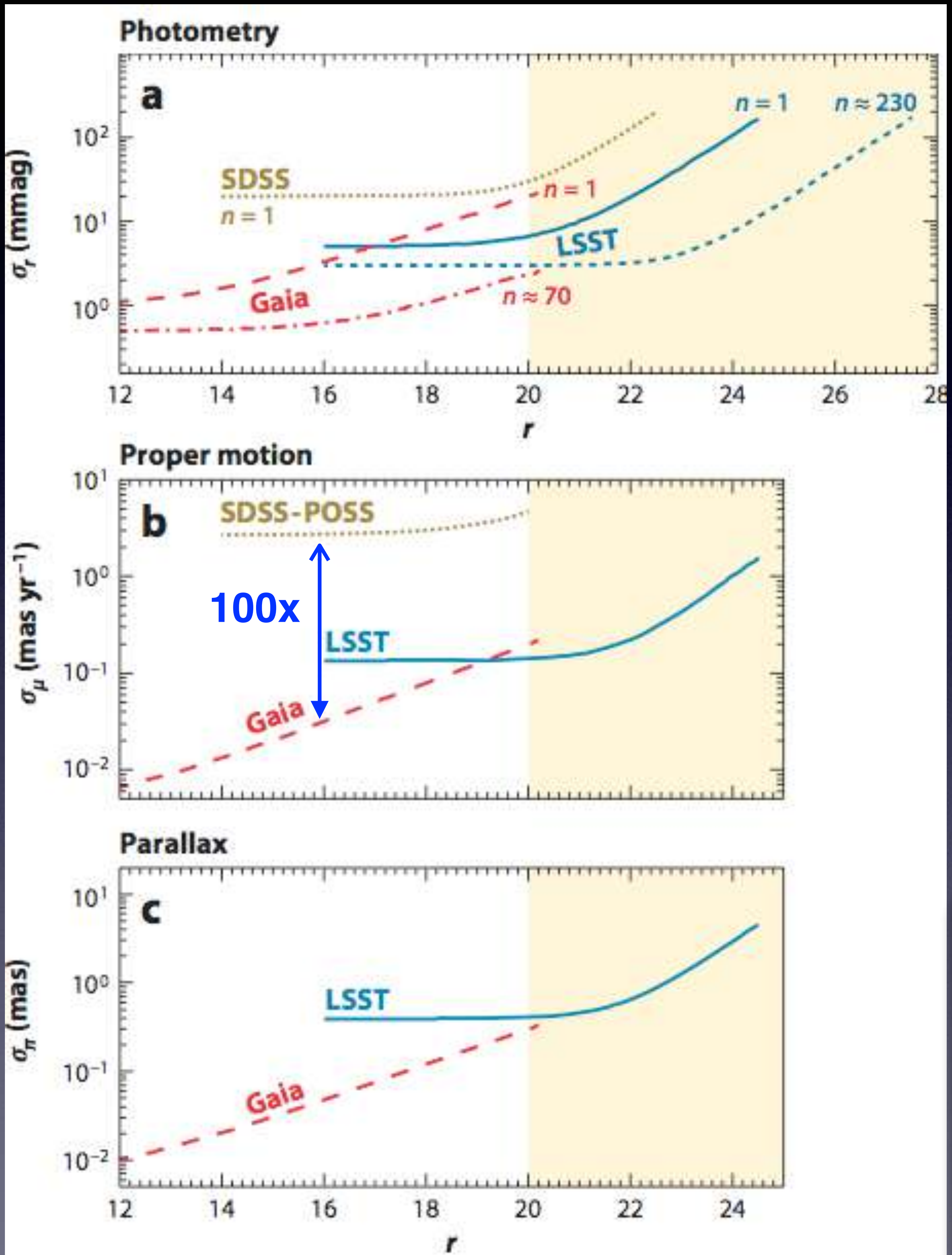

Compared to SDSS: LSST can "see" about 40 times more stars, 10 times further away and over twice as large sky area

Distance and [Fe/H]:



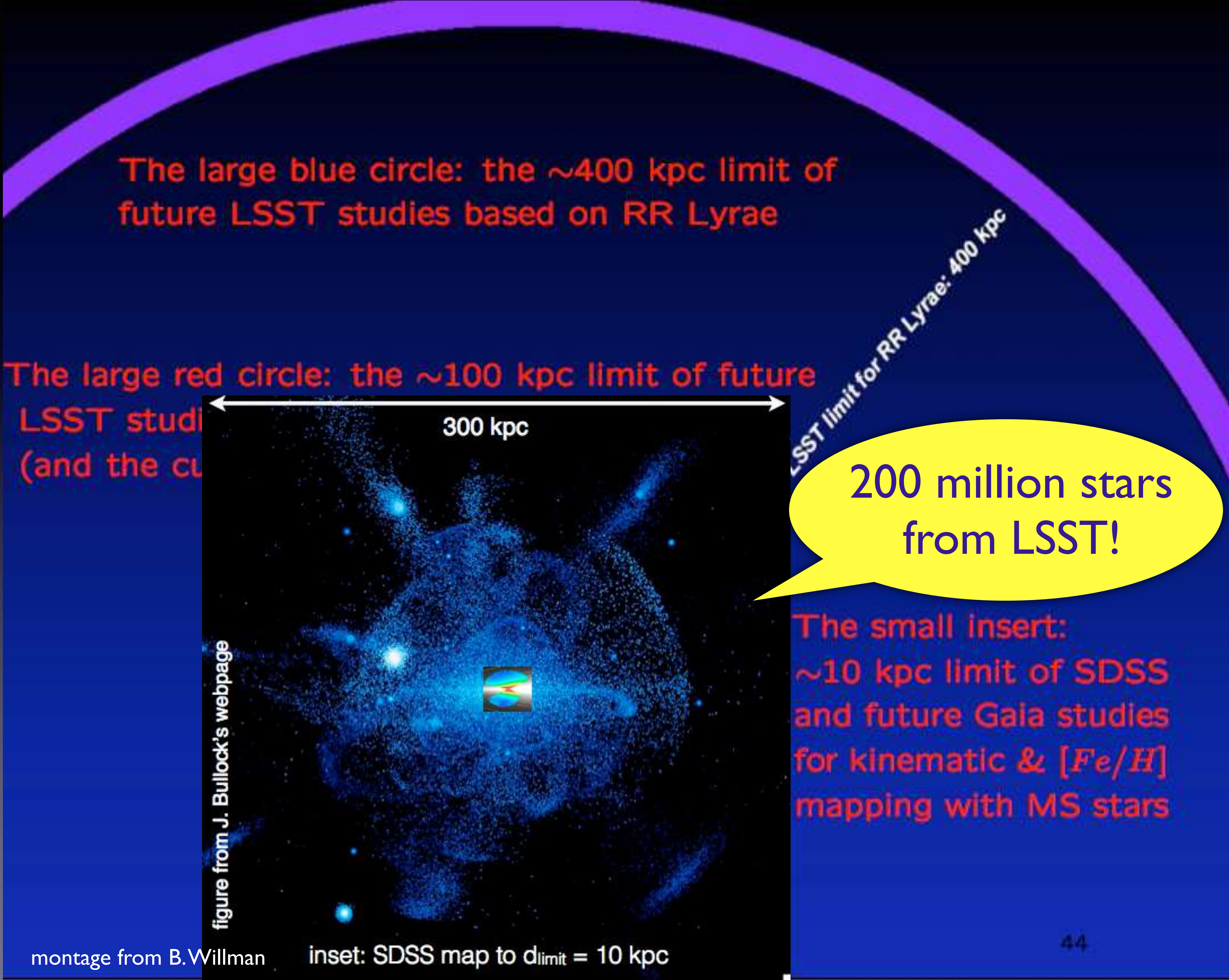
# Gaia vs. LSST comparison




Ivezić, Beers, Jurić 2012, ARA&A, 50, 251

- **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).



The large blue circle: the ~400 kpc limit of future LSST studies based on RR Lyrae
LSST limit for RR Lyrae: 400 kpc
The large red circle: the ~100 kpc limit of future
300 kpc
200 million stars from LSST!
The small insert: ~10 kpc limit of SDSS and future Gaia studies for kinematic & $[Fe/H]$ mapping with MS stars
figure from J. Bullock's webpage
inset: SDSS map to $d_{limit}$ = 10 kpc
montage from B. Willman
44

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

$3.6 \times 10^{-31}$ erg/s/cm$^2$/Hz
36 nJy
100x fainter than SDSS

**LSST in one sentence:**

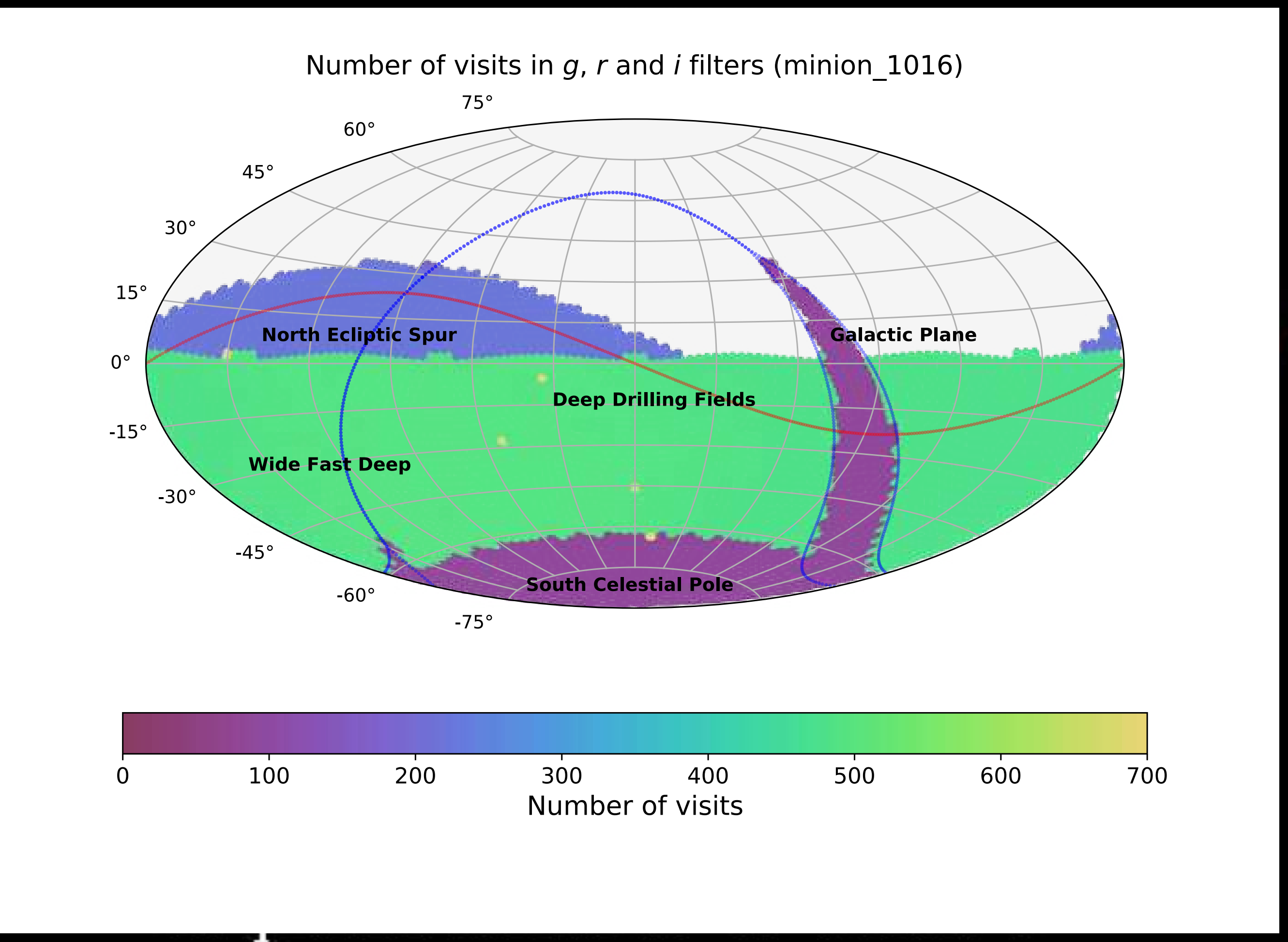
An optical/near-IR survey of half the sky in ugrizy bands to r~27.5 based on ~1000 visits over a 10-year period:

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

A catalog of 20 billion stars and 20 billion galaxies with exquisite photometry, astrometry and image quality!

# 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 40 billion objects**




**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 10-year 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)

SDSS
gri
3.5'x3.5'
r~22.5

3 arcmin is 1/10 of the full Moon's diameter

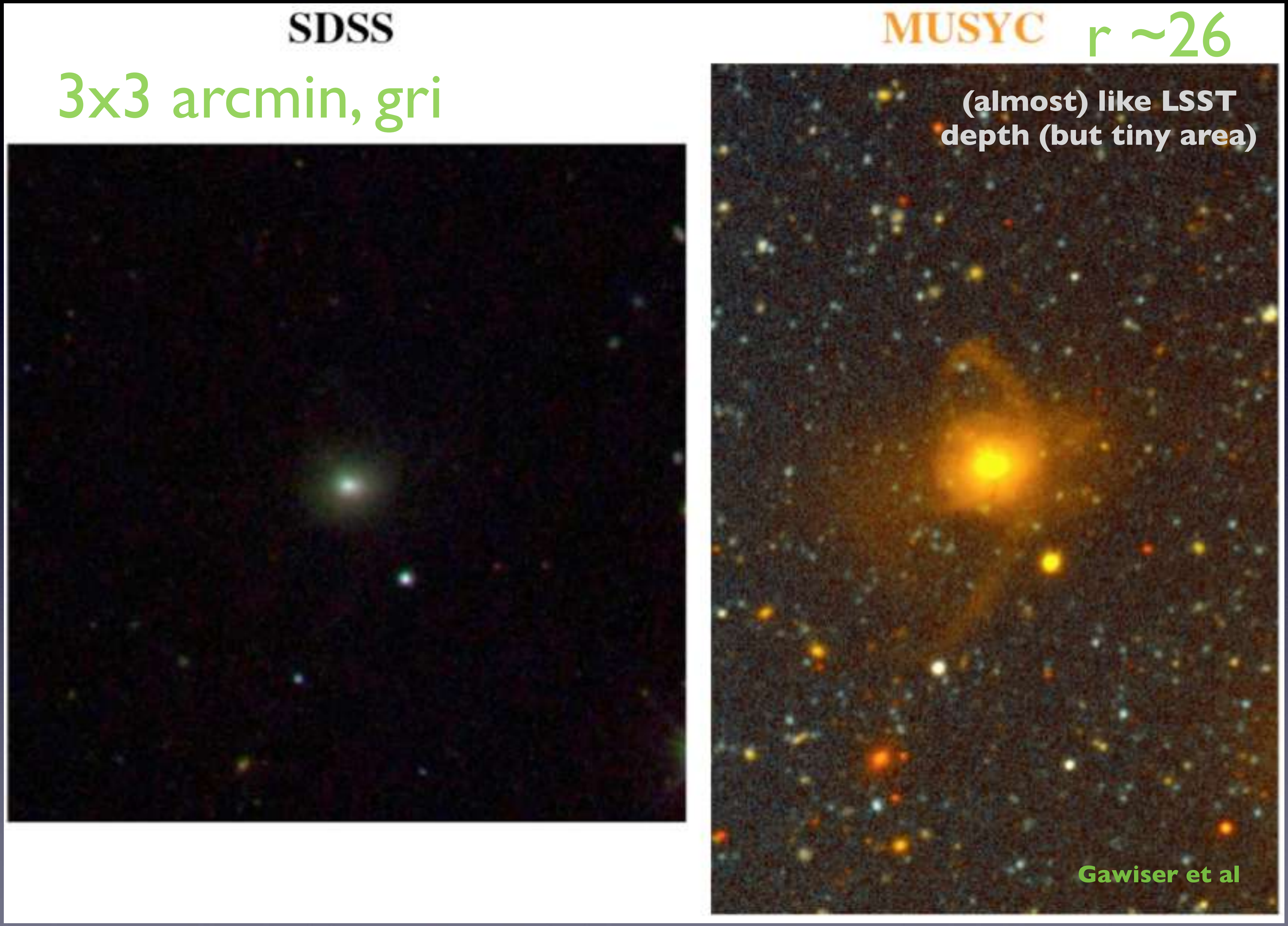
HSC
gri
3.5'x3.5'
r~27

3 arcmin is 1/10 of the full Moon's diameter

like LSST depth (but tiny area)

LSST will deliver 5 million such images

# Extragalactic astronomy: faint surface brightness limit



# Filter complement

- **Photometric redshifts for galaxies:** random errors smaller than 0.02, bias below 0.003, fewer than 10% >3σ outliers
- These photo-z requirements are one of the primary drivers for the photometric depth and accuracy of the main LSST survey (and the definition of filter complement)

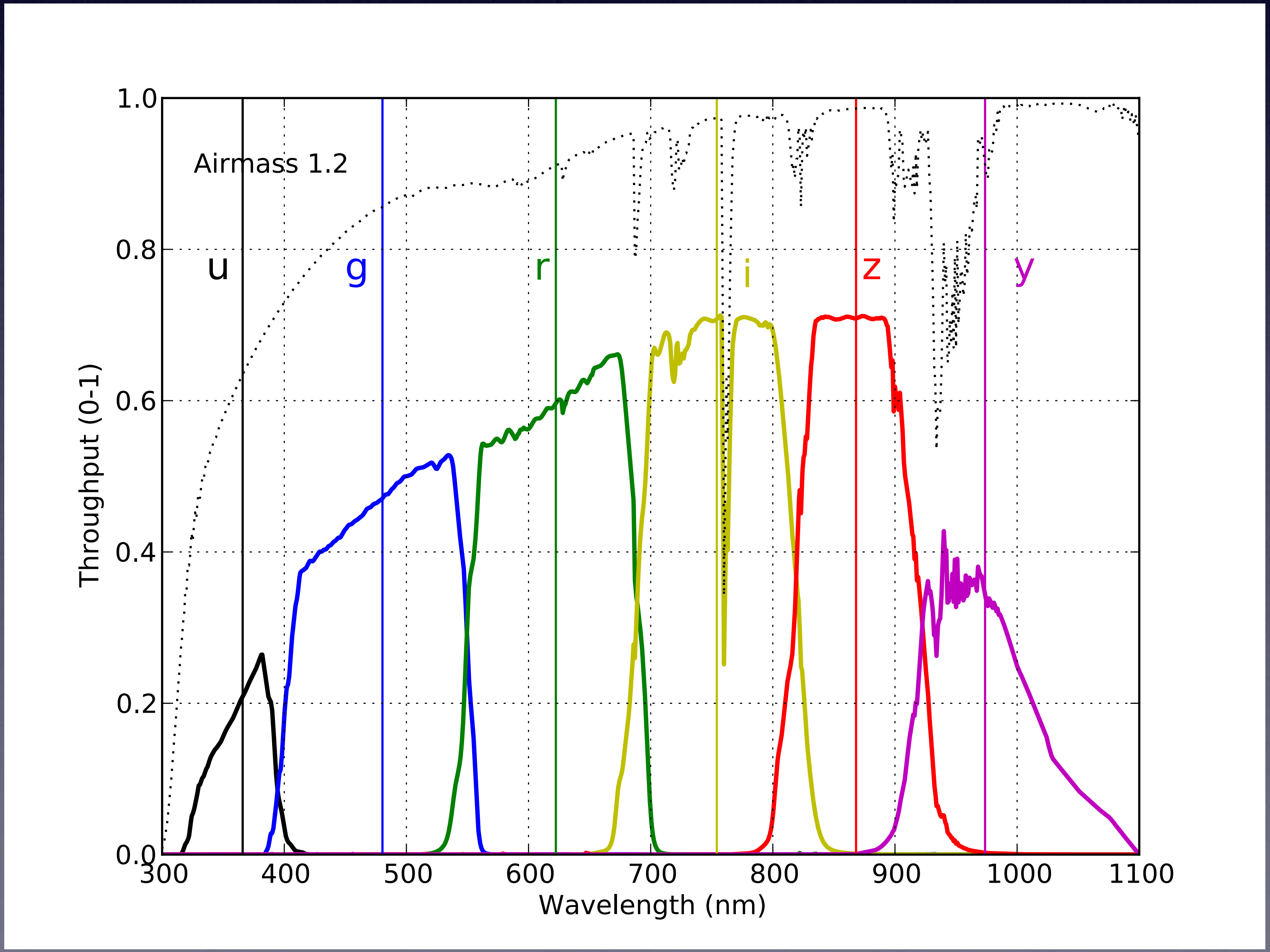



**Photo-z requirements correspond to r~27.5**

with the following per band time allocations:

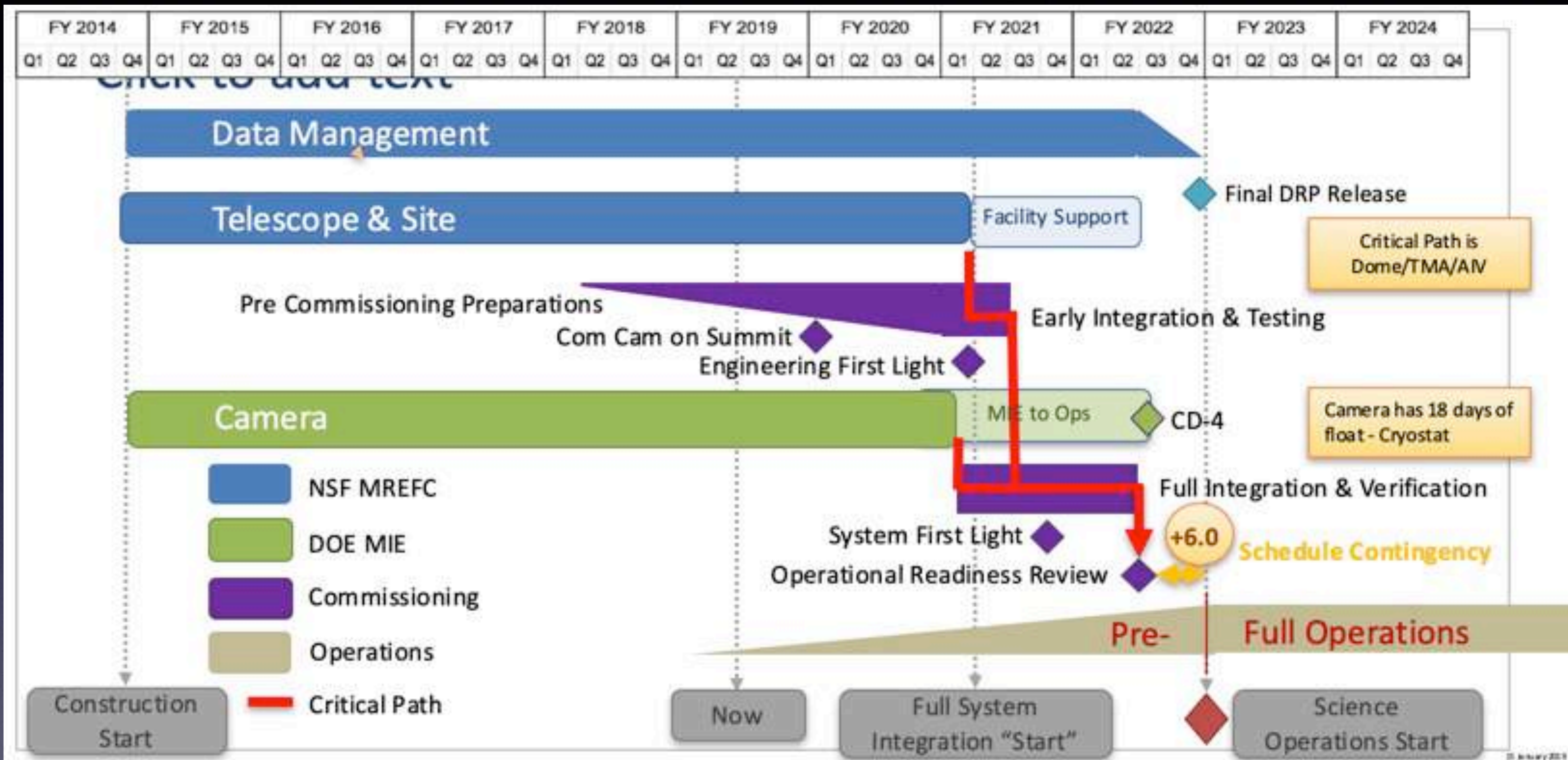
u: 8%; g: 10%

r: 22%; i: 22%

z: 19%; y: 19%

**Consistent with other science themes (stars)**

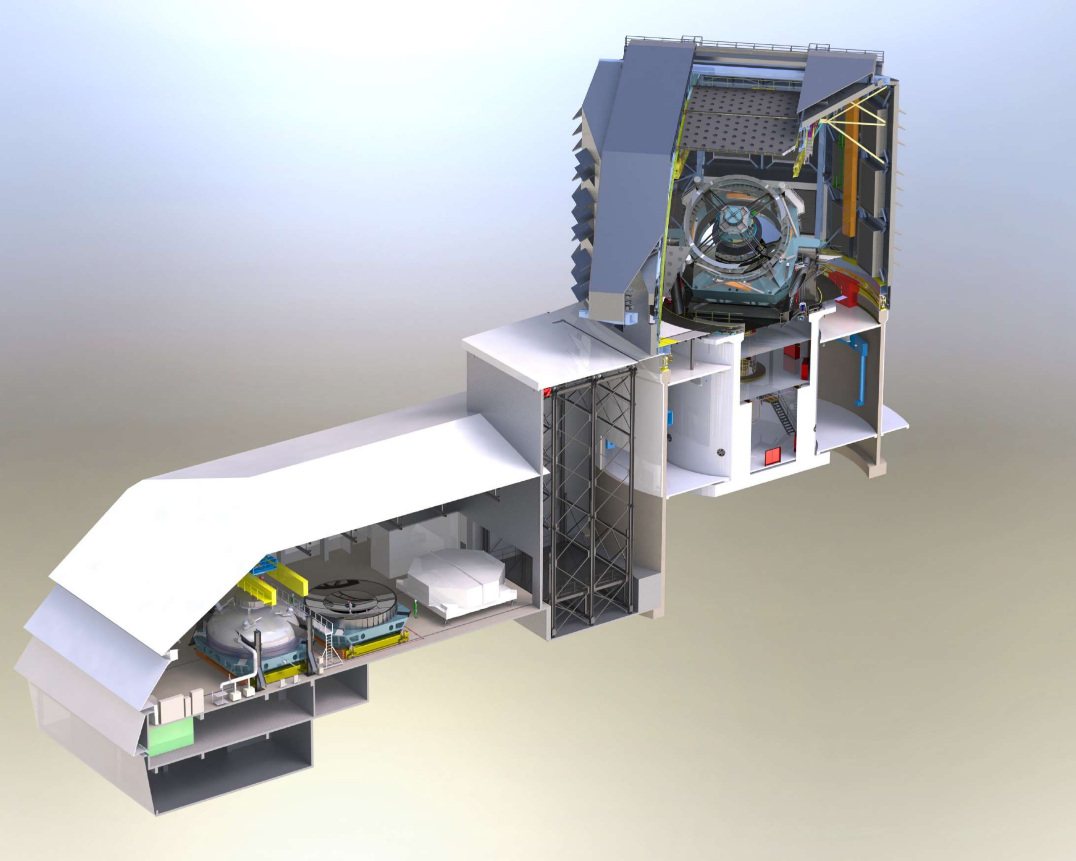
# Project Schedule as of April 28, 2019




**Start of Operations: Oct 1, 2022**

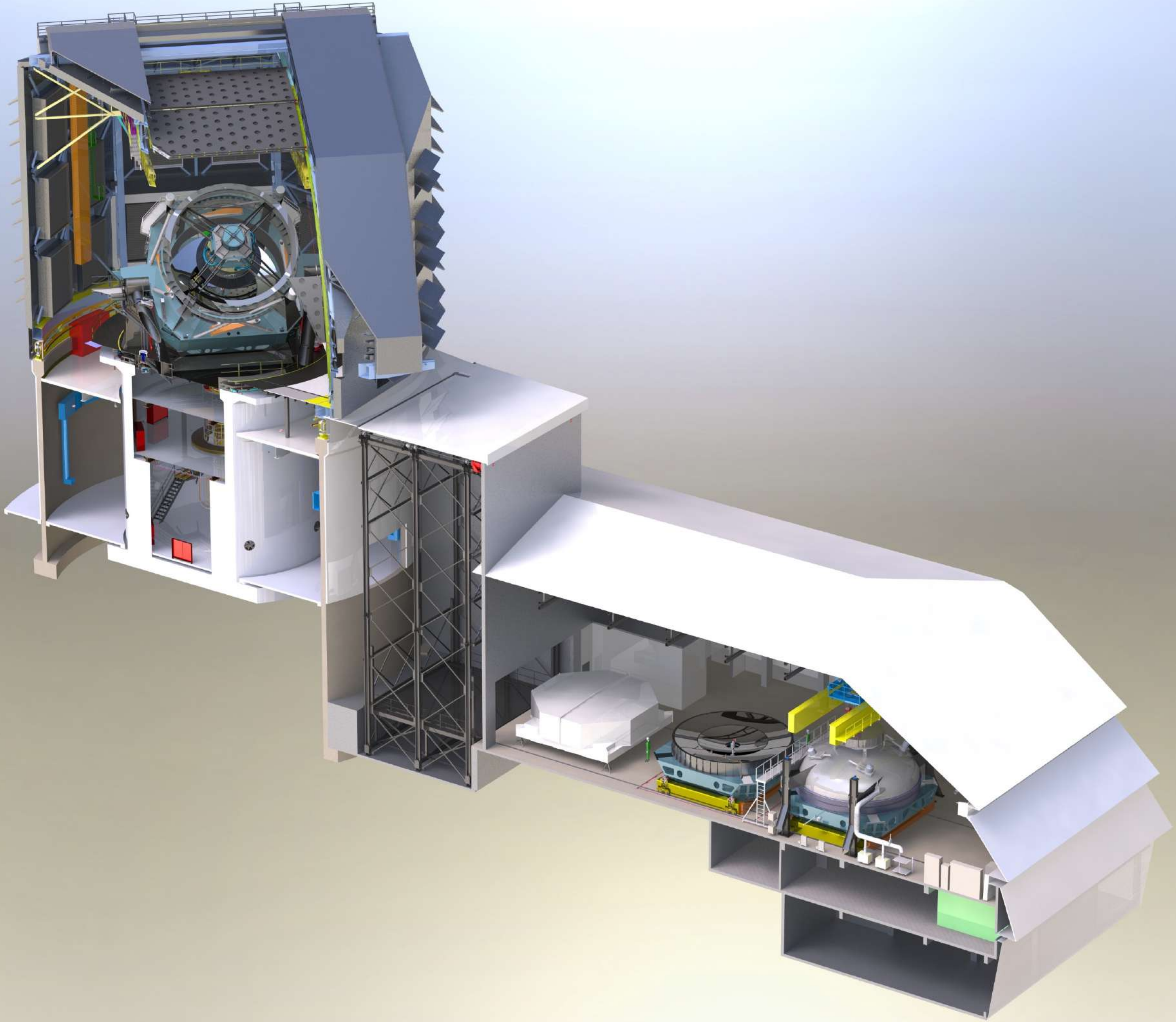
April 14, 2015


February 15, 2016
material staging
Provisional plywood walkways
Excavation for telescope pier foundation - rebar placed for pour Feb 22 to 24
Excavation for lower enclosure foundation
Excavation for platform lift
Service building concrete structure in progress
Elec. & tank area
Formwork for beams to support level 3 floor & mirror cart rails

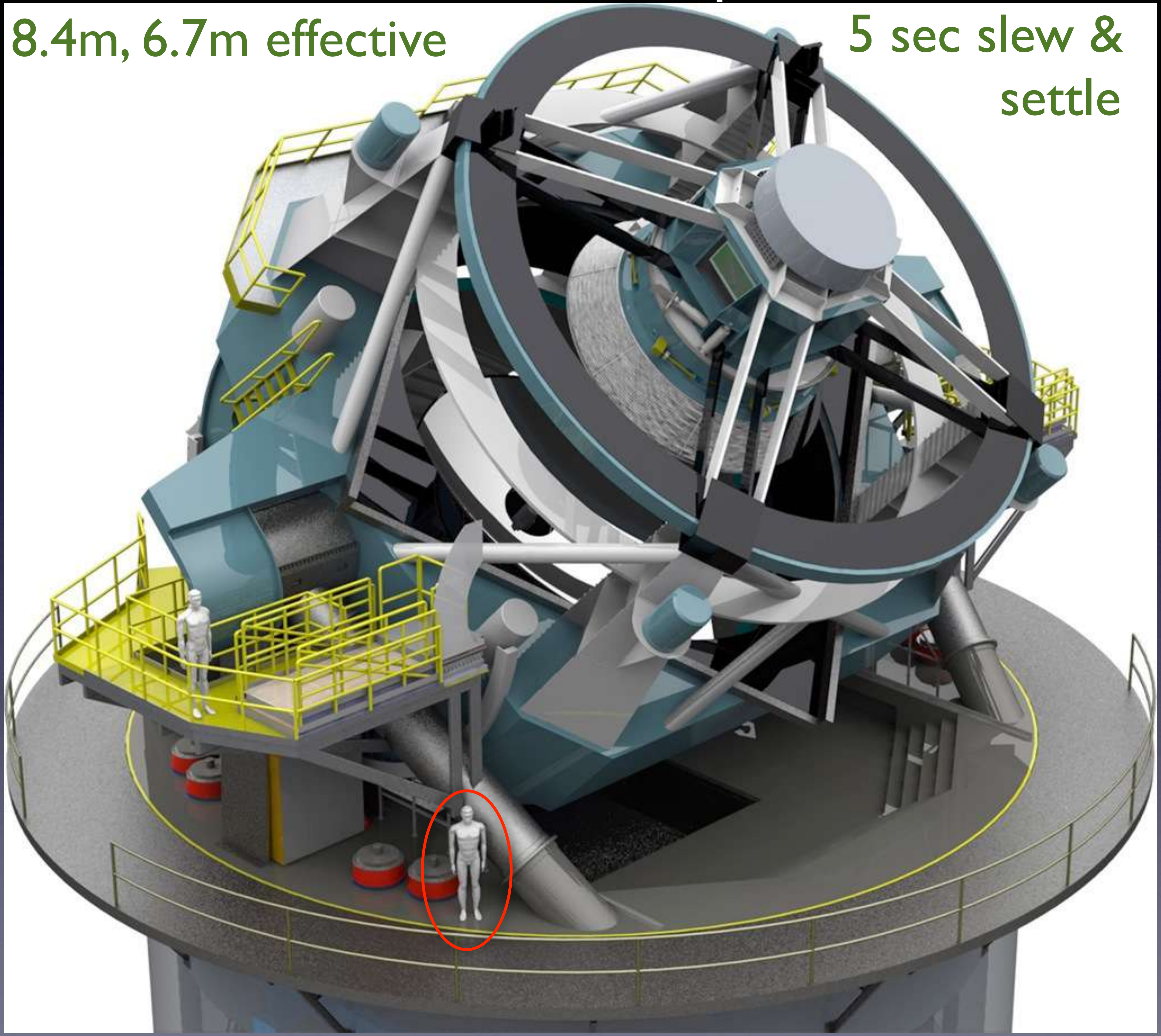


Oct 17, 2017


# First light: 2021

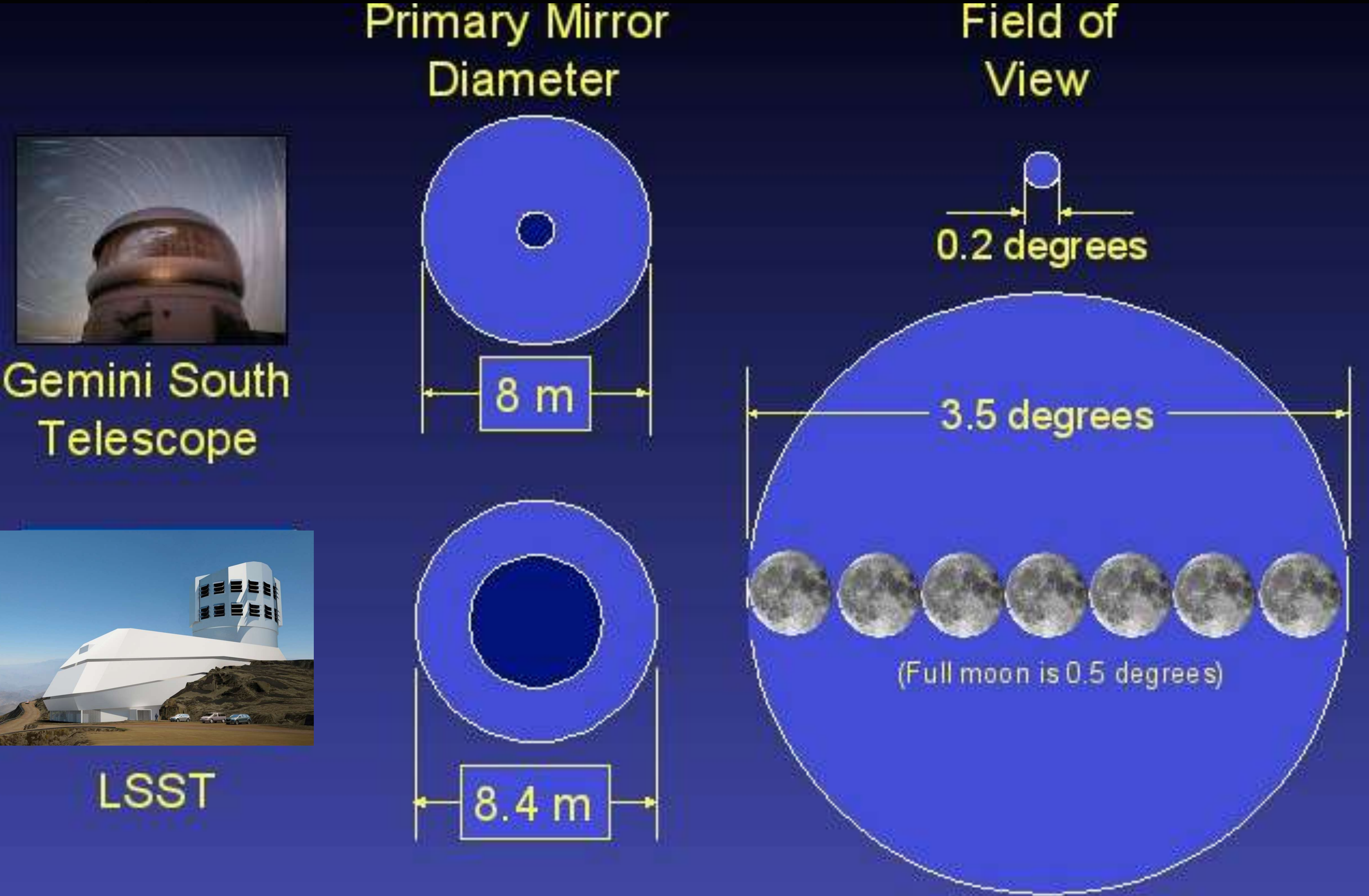


# LSST Telescope

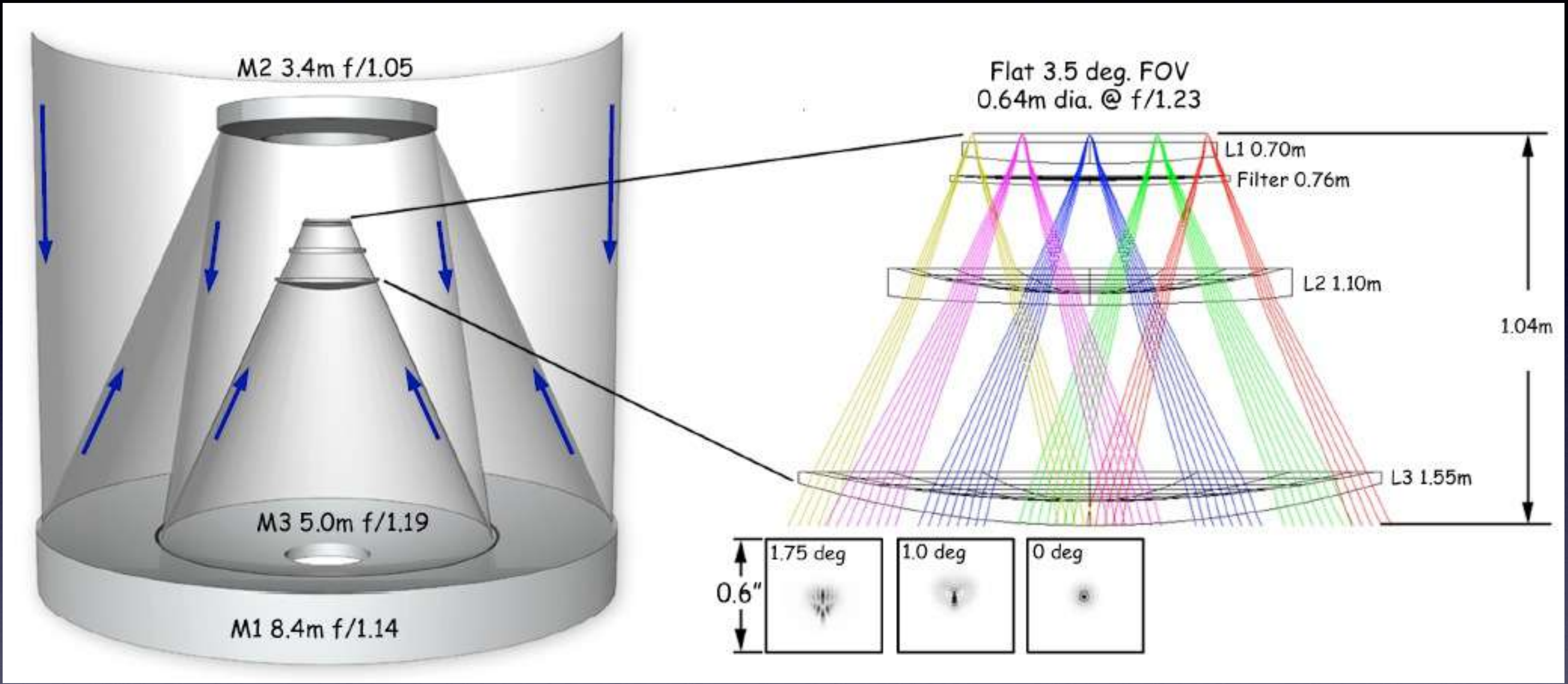



**Telescope Mount Assembly is on its way from Spain to Chile**

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

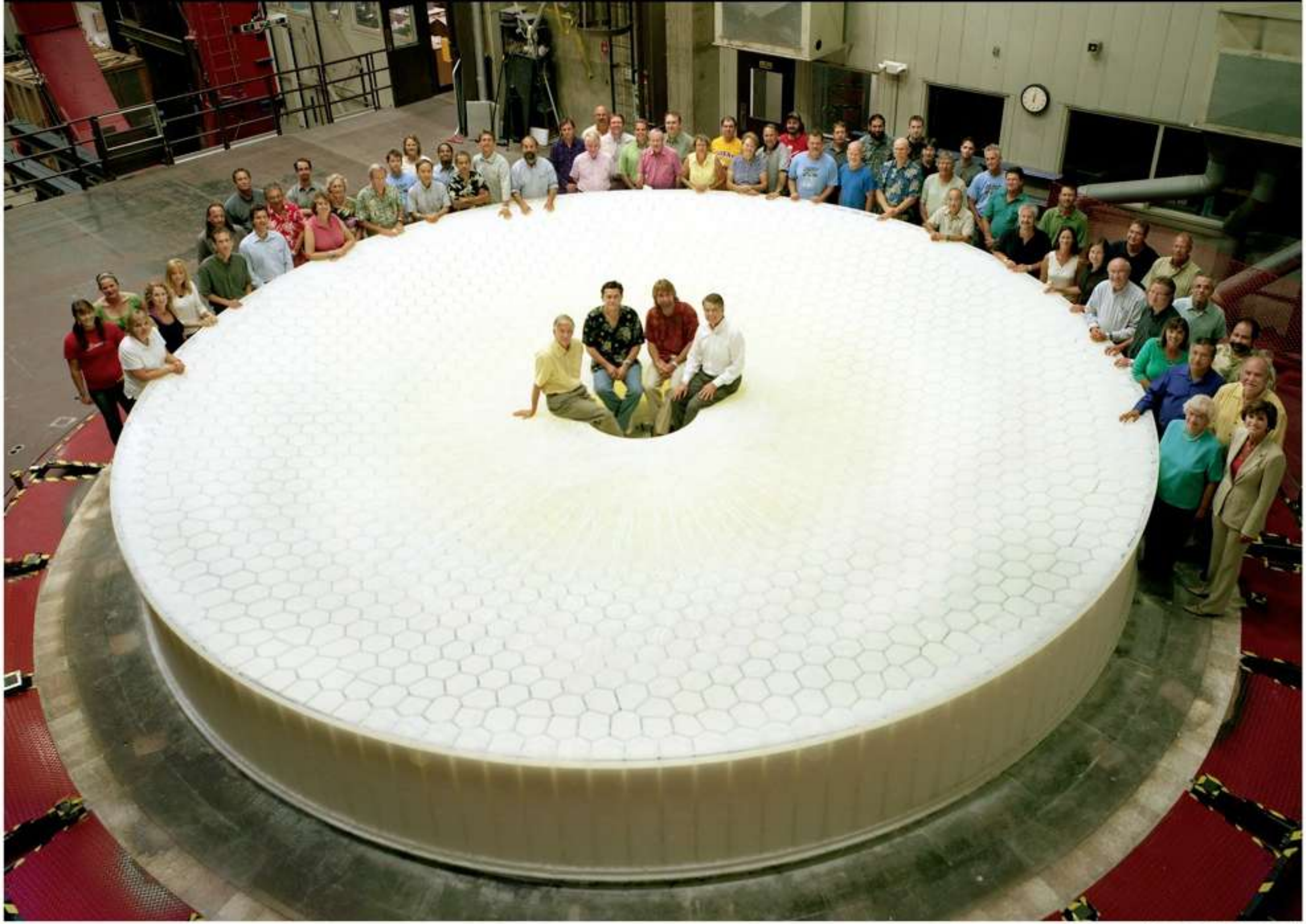


# Optical Design for LSST




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

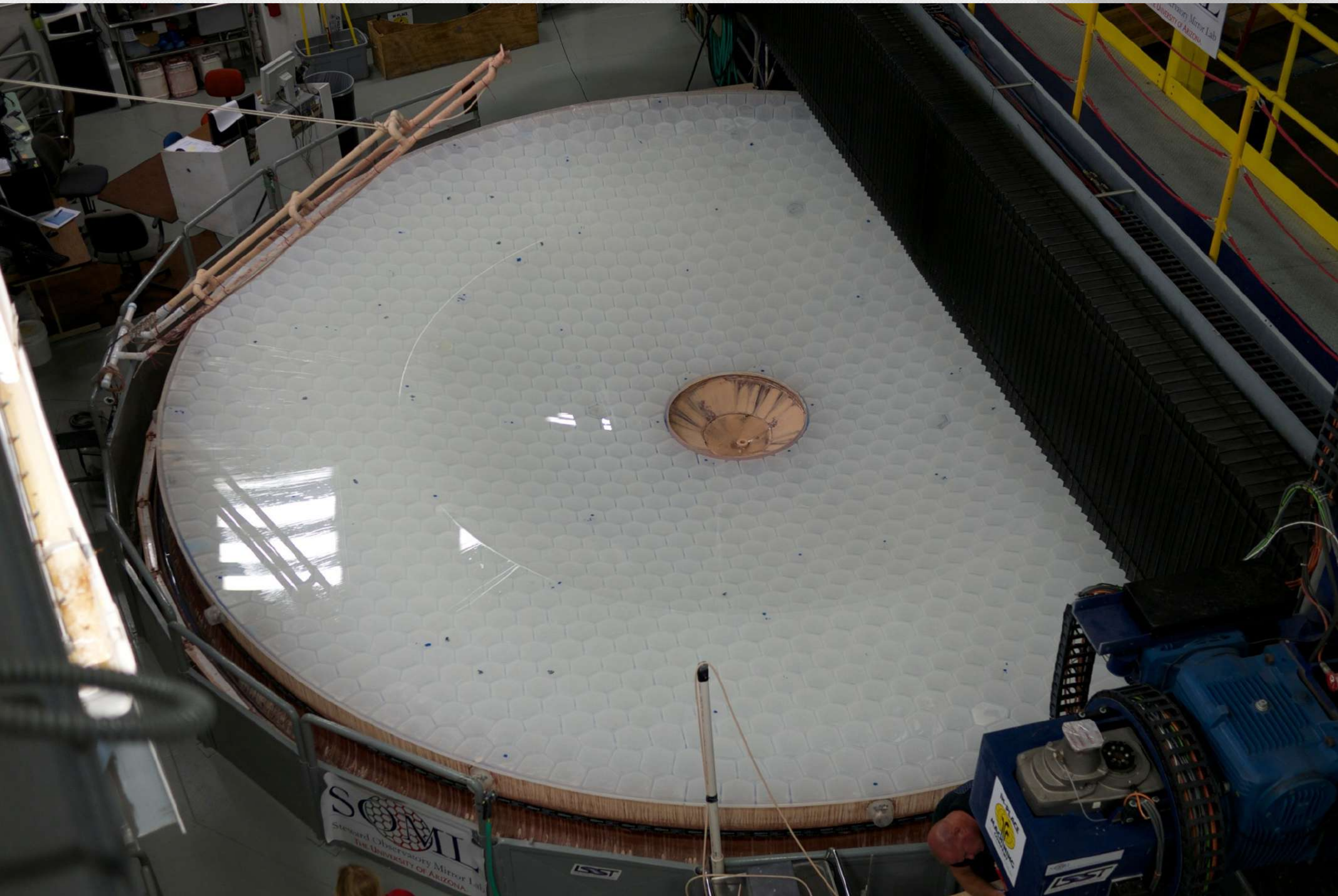
# Large Synoptic Survey Telescope



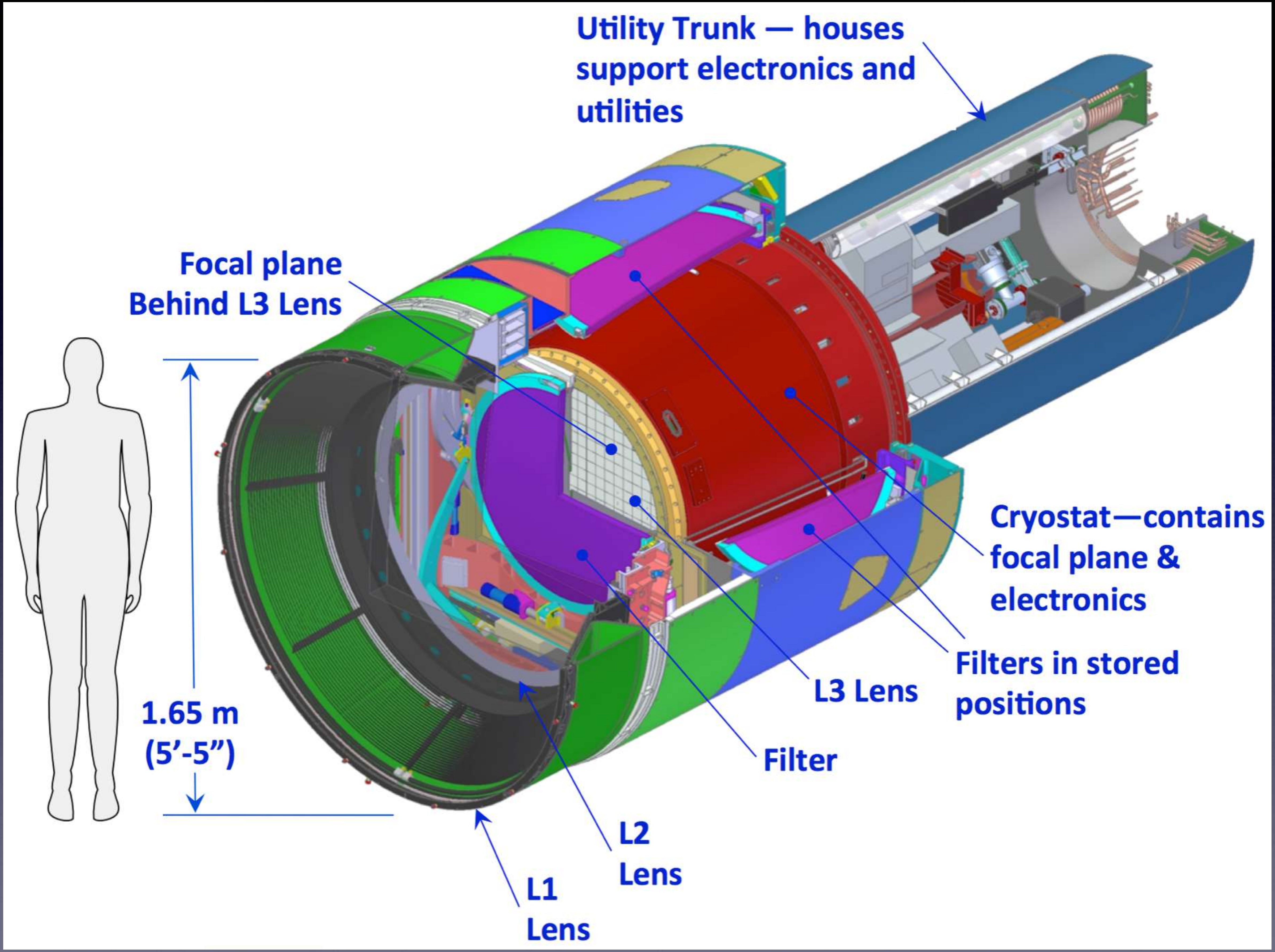
LSST Primary/Tertiary Mirror Blank
August 11, 2008, Steward Observatory Mirror Lab, Tucson, Arizona

The University of Arizona, Tucson Arizona — SOML — LSST Large Synoptic Survey Telescope

# Large Synoptic Survey Telescope

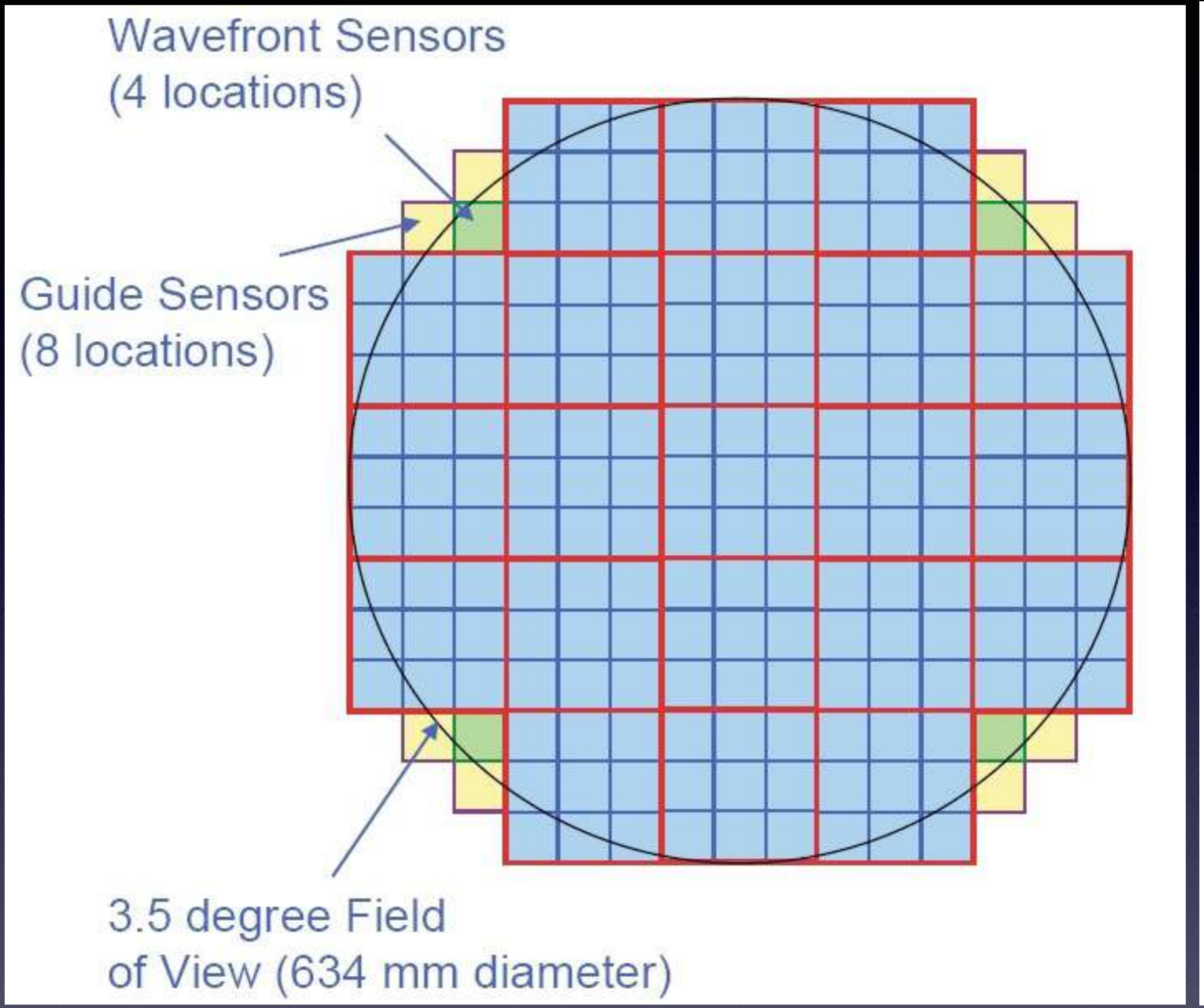

Steward Observatory Mirror Lab
THE UNIVERSITY OF ARIZONA

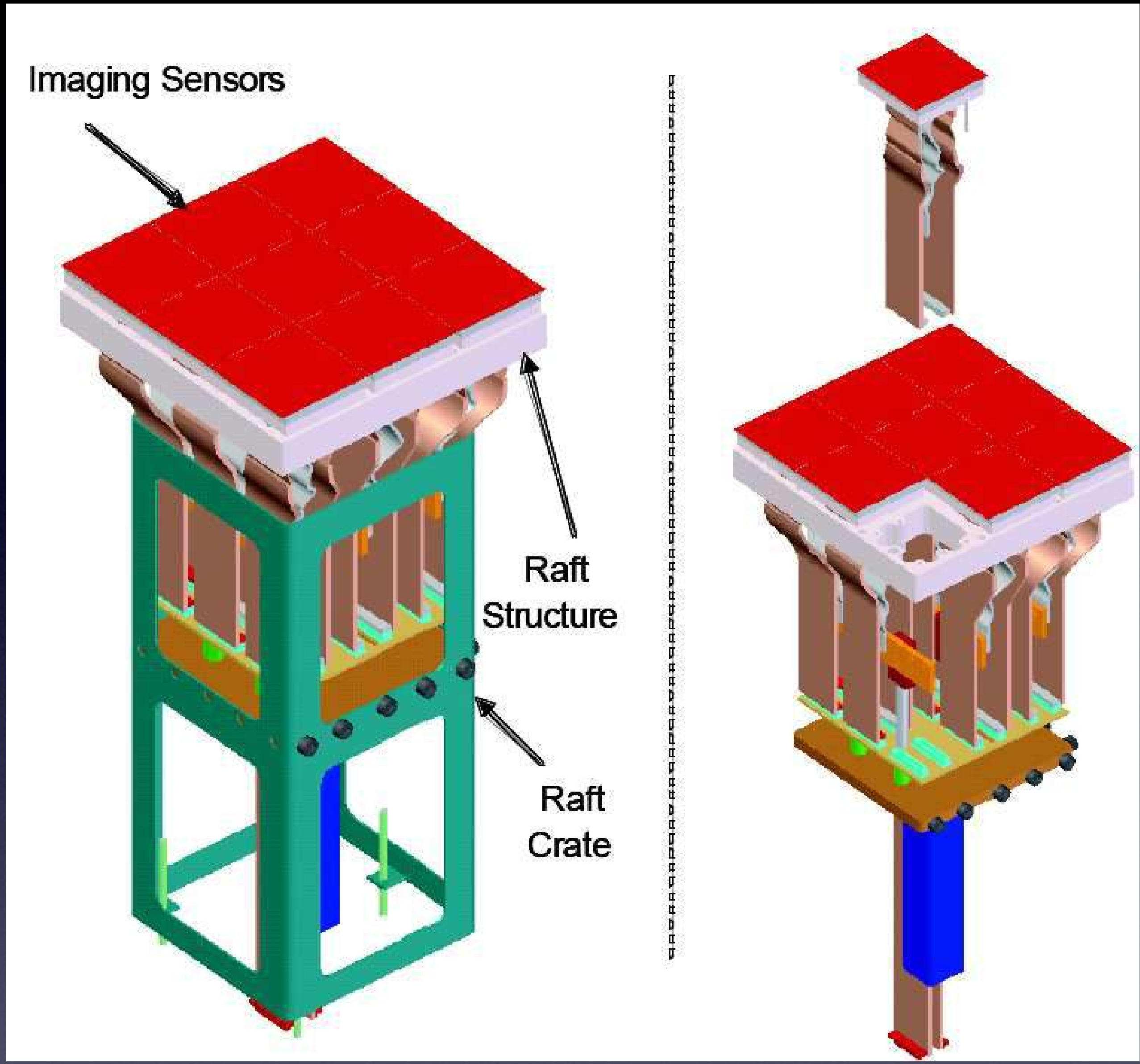
# LSST camera




The largest astronomical camera: 2800 kg, 3.2 Gpix

# LSST camera




Modular design: 3200 Megapix = 189 x16 Megapix CCD

9 CCDs share electronics: raft (=camera)

Problematic rafts can be replaced relatively easily

**LSST Science Sensor procurement is complete!**

# Start of LSST Operations: Oct 3, 2022

## Get ready!

# At the highest level, LSST objectives are:

1) Obtain about 5.5 million images, with 189 CCDs (4k x 4k) in the focal plane; this is about **a billion 16 Megapixel images of the sky**

2) Calibrate these images (and provide other metadata)

3) Produce catalogs ("model parameters") of detected objects (37 billion)

4) **Serve** images, catalogs and all other metadata, that is,
**LSST data products to LSST users**

**The ultimate deliverable of LSST is** not just the telescope, nor the camera, but **the fully reduced science-ready data** as well. **Software!**

- 20 TB of data to process every day (~one SDSS/day)
- Existing tools and methods do not scale up to LSST data volume and rate (100 PB!)
- About 5-10 million lines of new code (C++/python)

## LSST data products are organized into 3 categories:

### Prompt Data Products

**Real Time Difference Image Analysis (DIA)**

- A stream of ~10 million time-domain events per night (Alerts), transmitted to event distribution networks within 60s of camera readout.
- Images, Object and Source catalogs derived from DIA, and an orbit catalog for ~6 million Solar System bodies within 24h.
- Enables discovery and rapid follow-up of time domain events

### Data Release Data Products

**Reduced single-epoch & deep co-added images, catalogs, reprocessed DIA products**

- Catalogs of ~37 billion objects (20 billion galaxies, 17 billion stars), ~7 trillion sources and ~30 trillion forced source measurements.
- 11 Data Releases, produced ~annually over 10 years of operation
- Accessible via the LSST Science Platform & LSST Data Access Centers.

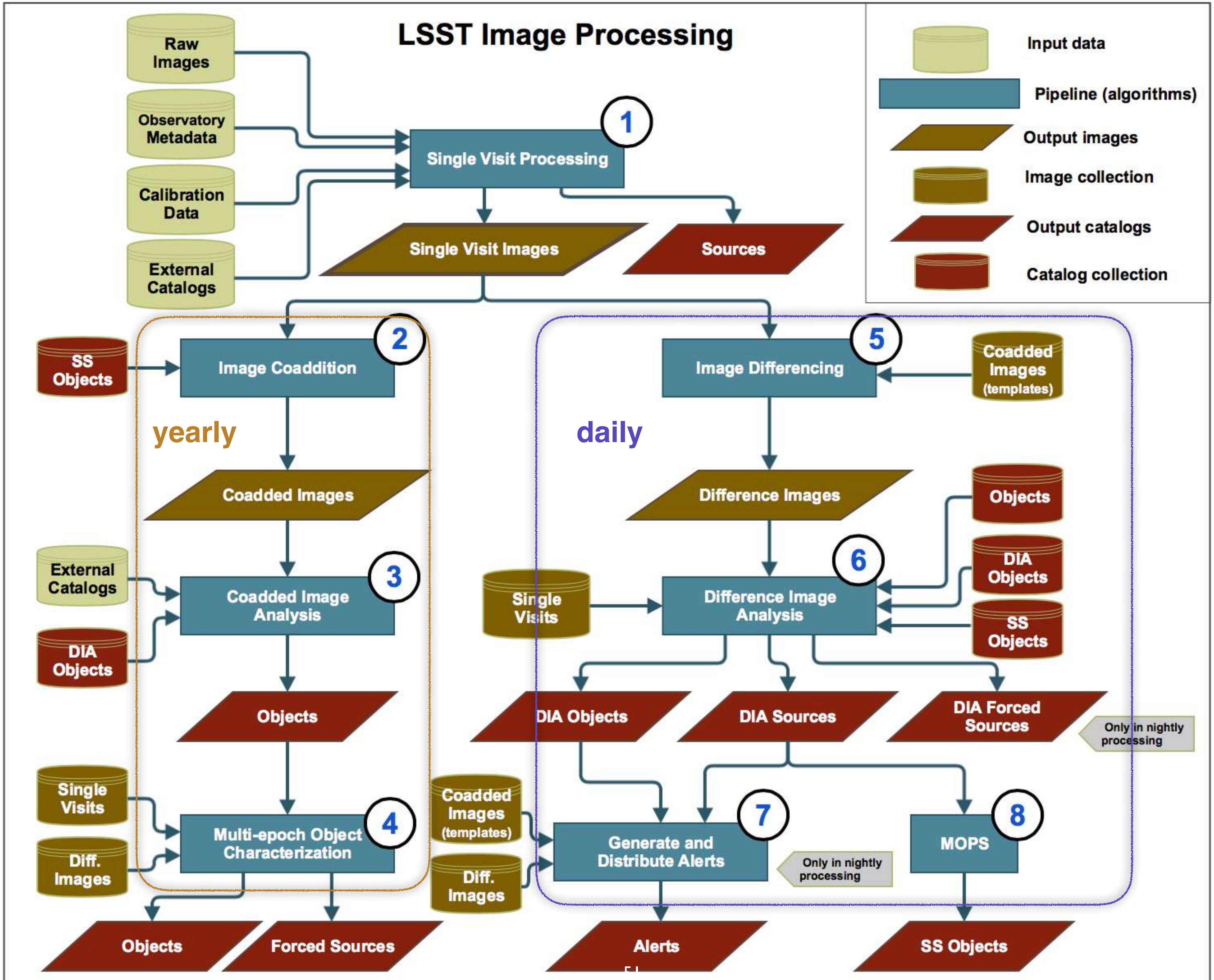
### User Generated Data Products

**User-produced derived, added-value data products**

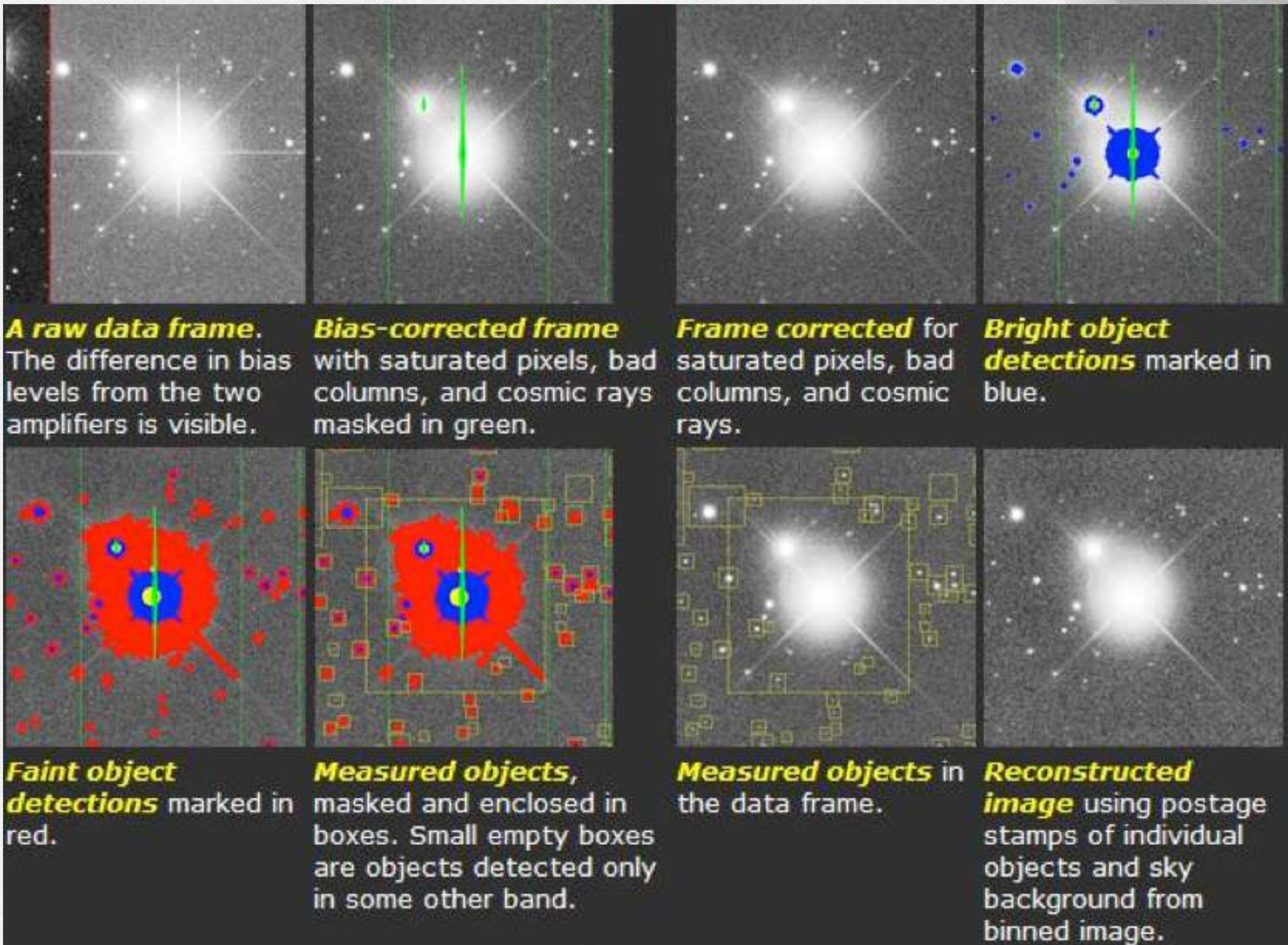
- Deep KBO/NEO, variable star classifications, shear maps, etc ...
- y services & computing resources at the LSST DACs and via the LSST Science Platform (LSP).
- 10% of LSST computing resources will be allocated for User Generated data product storage & processing.

Product Categories & DM Data Products & LSST Key Numbers

**LSST Data Products: see http://ls.st/dpdd**


LSST Image Processing
Input data
Pipeline (algorithms)
Output images
Image collection
Output catalogs
Catalog collection
Raw Images
Observatory Metadata
Calibration Data
External Catalogs
1
Single Visit Processing
Single Visit Images
Sources
2
Image Coaddition
SS Objects
yearly
Coadded Images
3
Coadded Image Analysis
External Catalogs
DIA Objects
Objects
4
Multi-epoch Object Characterization
Single Visits
Diff. Images
Objects
Forced Sources
5
Image Differencing
Coadded Images (templates)
daily
Difference Images
6
Difference Image Analysis
Single Visits
Objects
DIA Objects
SS Objects
DIA Objects
DIA Sources
DIA Forced Sources
Only in nightly processing
Coadded Images (templates)
Diff. Images
7
Generate and Distribute Alerts
Only in nightly processing
8
MOPS
Alerts
SS Objects

# Basic steps in astronomical image processing



***A raw data frame***. The difference in bias levels from the two amplifiers is visible.

***Bias-corrected frame*** with saturated pixels, bad columns, and cosmic rays masked in green.

***Frame corrected*** for saturated pixels, bad columns, and cosmic rays.

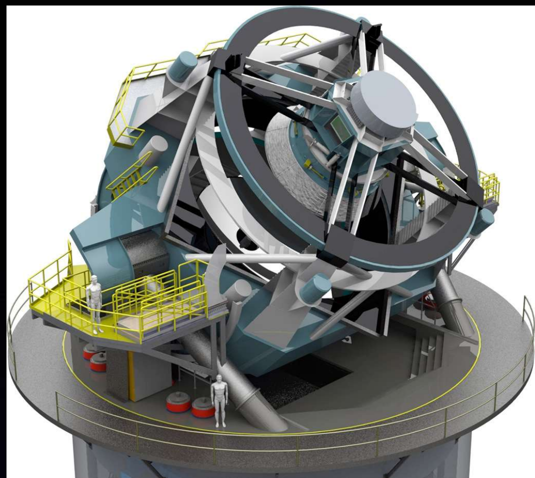
***Bright object detections*** marked in blue.

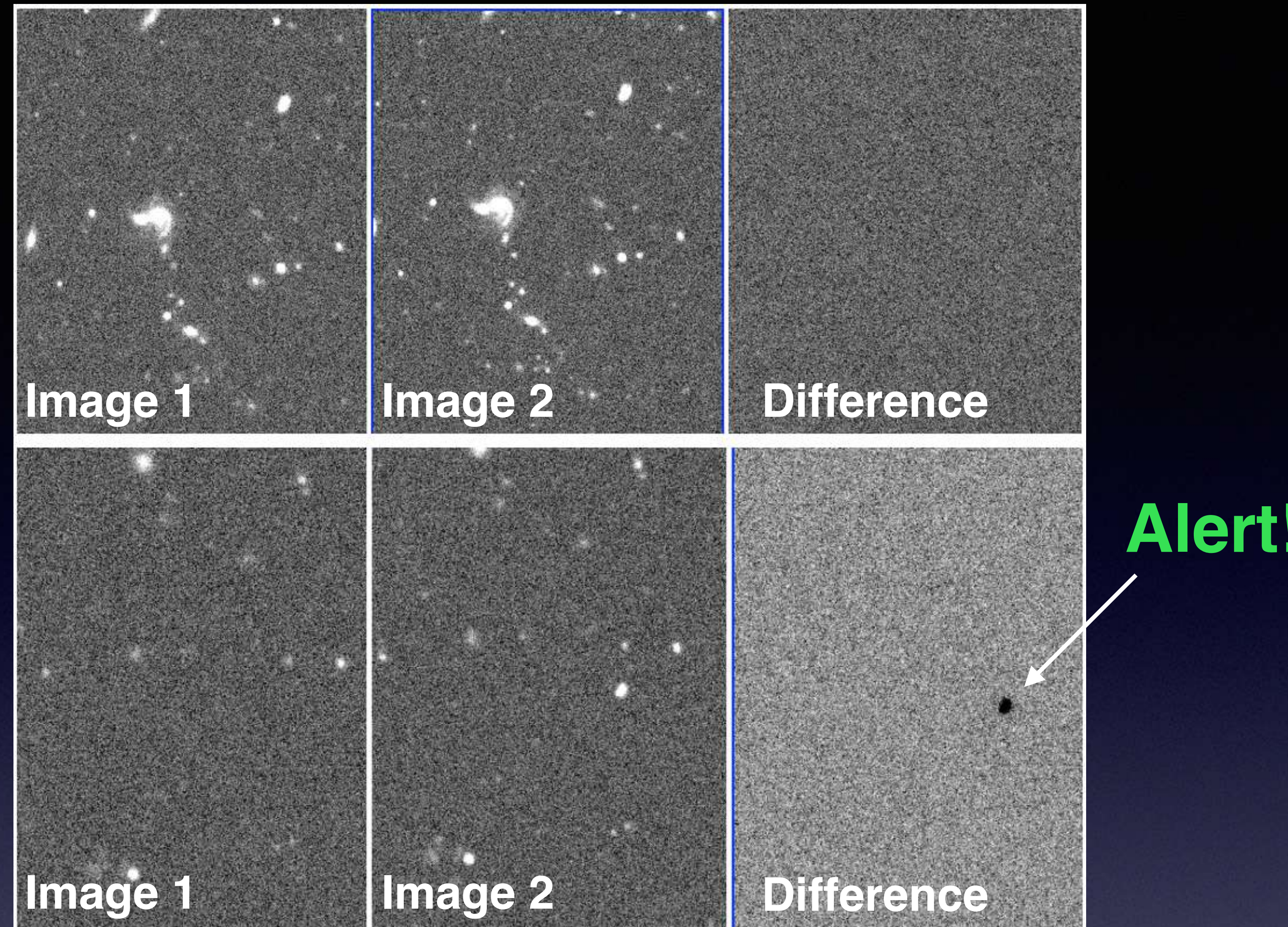
***Faint object detections*** marked in red.

***Measured objects***, masked and enclosed in boxes. Small empty boxes are objects detected only in some other band.

***Measured objects*** in the data frame.

***Reconstructed image*** using postage stamps of individual objects and sky background from binned image.






**Additional "followup" data obtained to:**

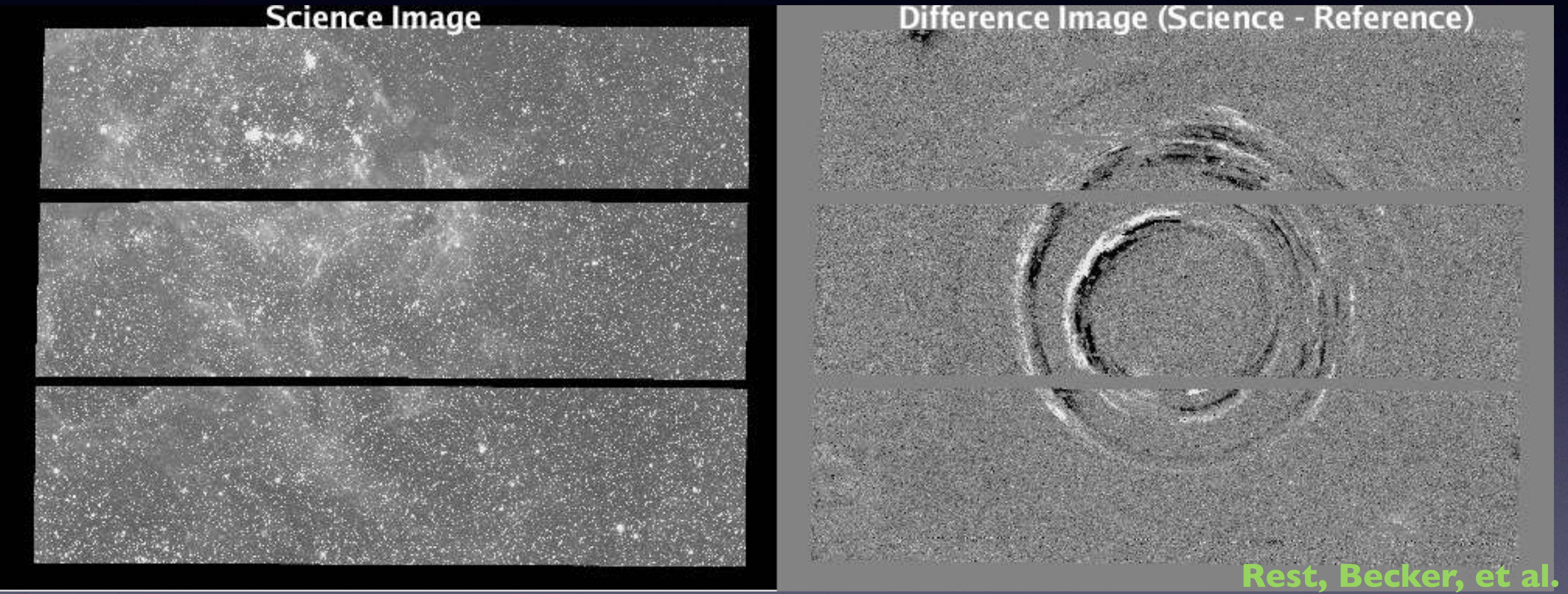
- confirmation and classification
- provide better temporal resolution
- use different filters/wavelengths
- obtain spectra (distance!)
- other measurements (e.g. polarimetry)

**Alerts can trigger "Followup" observations:**

# **Time Domain:** objects changing in time

positions: asteroids and stellar proper motions
brightness: cosmic explosions and variable stars

Not only point sources - echo of a supernova explosion:




As many variable stars from LSST, as all stars from SDSS

**Web stream with data for transients within 60 seconds.**

**Real time alerts!**

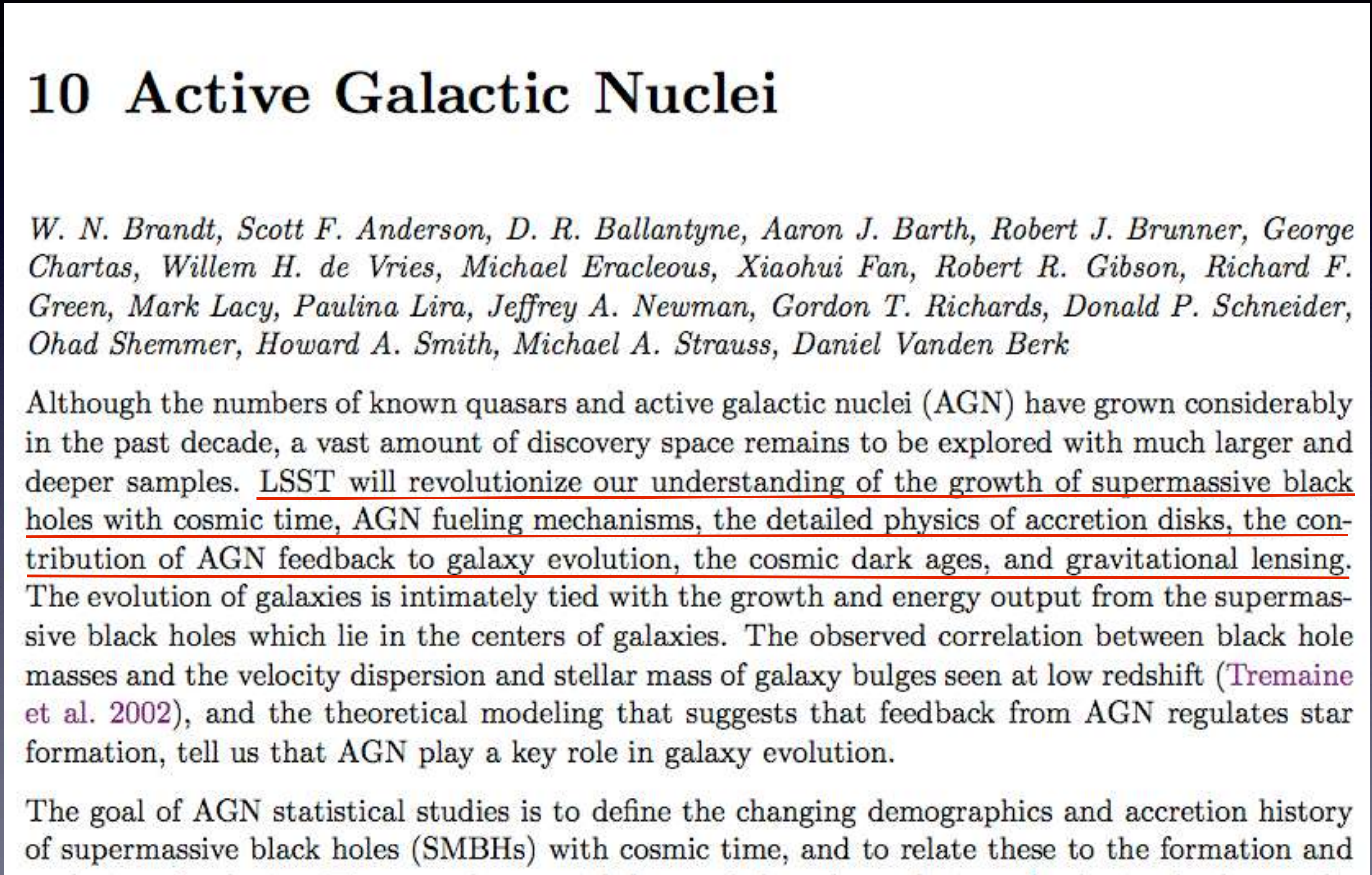
"Ask Not What Data You Need To Do Your Science, Ask What Science You Can Do With Your Data."


# The era of surveys...

- Standard: "What data do I have to collect to (dis)prove a hypothesis"?
- Data-driven: "What theories can I test given the data I already have?"

# Extragalactic astronomy: AGNs

## From LSST Science Book (arXiv:0912.0201):

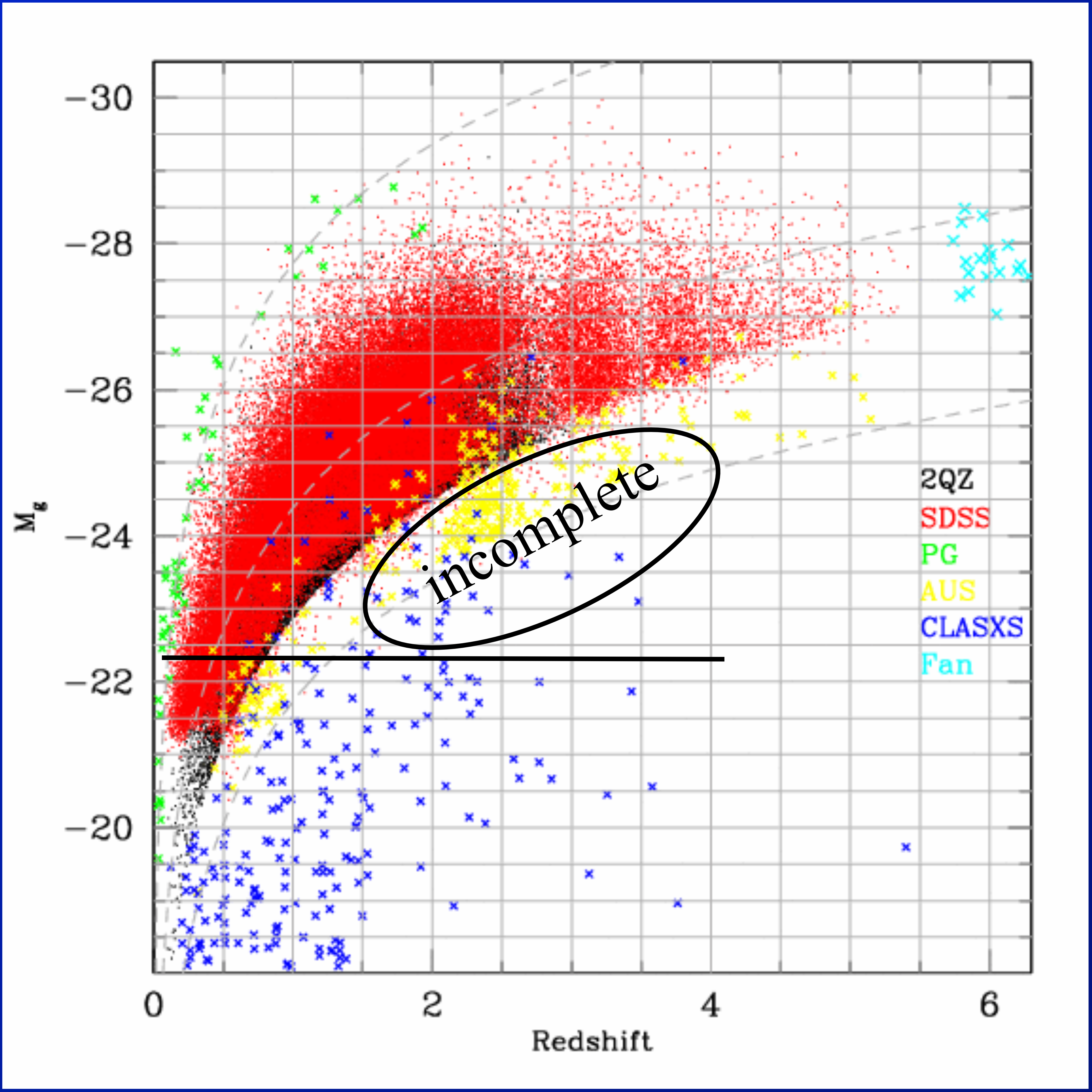

# 10 Active Galactic Nuclei

*W. N. Brandt, Scott F. Anderson, D. R. Ballantyne, Aaron J. Barth, Robert J. Brunner, George Chartas, Willem H. de Vries, Michael Eracleous, Xiaohui Fan, Robert R. Gibson, Richard F. Green, Mark Lacy, Paulina Lira, Jeffrey A. Newman, Gordon T. Richards, Donald P. Schneider, Ohad Shemmer, Howard A. Smith, Michael A. Strauss, Daniel Vanden Berk*

Although the numbers of known quasars and active galactic nuclei (AGN) have grown considerably in the past decade, a vast amount of discovery space remains to be explored with much larger and deeper samples. LSST will revolutionize our understanding of the growth of supermassive black holes with cosmic time, AGN fueling mechanisms, the detailed physics of accretion disks, the contribution of AGN feedback to galaxy evolution, the cosmic dark ages, and gravitational lensing. The evolution of galaxies is intimately tied with the growth and energy output from the supermassive black holes which lie in the centers of galaxies. The observed correlation between black hole masses and the velocity dispersion and stellar mass of galaxy bulges seen at low redshift (Tremaine et al. 2002), and the theoretical modeling that suggests that feedback from AGN regulates star formation, tell us that AGN play a key role in galaxy evolution.

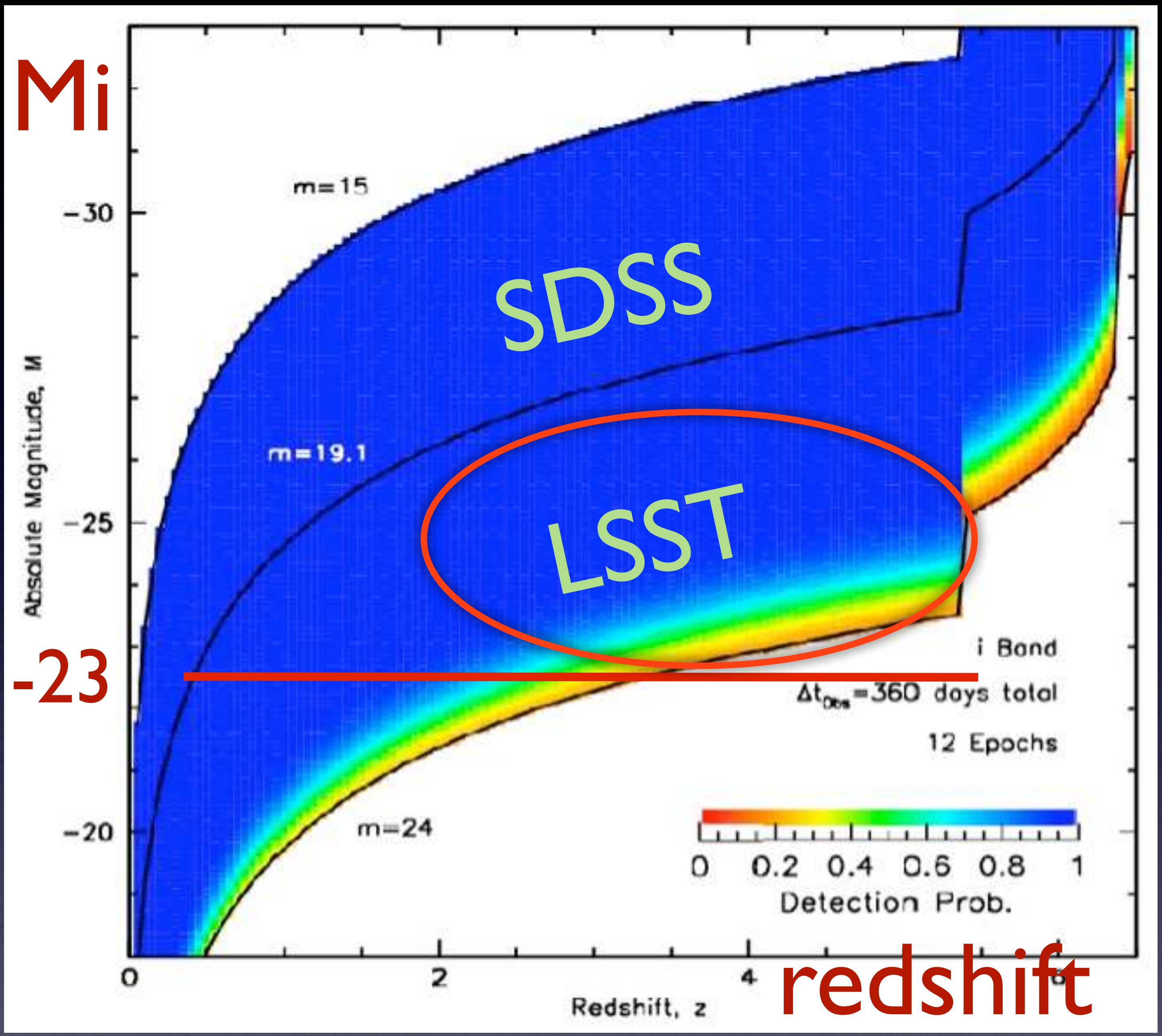
The goal of AGN statistical studies is to define the changing demographics and accretion history of supermassive black holes (SMBHs) with cosmic time, and to relate these to the formation and

# Quasar Surveys Status




**Great progress over the last decade, but: still incomplete for low-L objects already at z~1!**

# Extragalactic astronomy with LSST: quasars



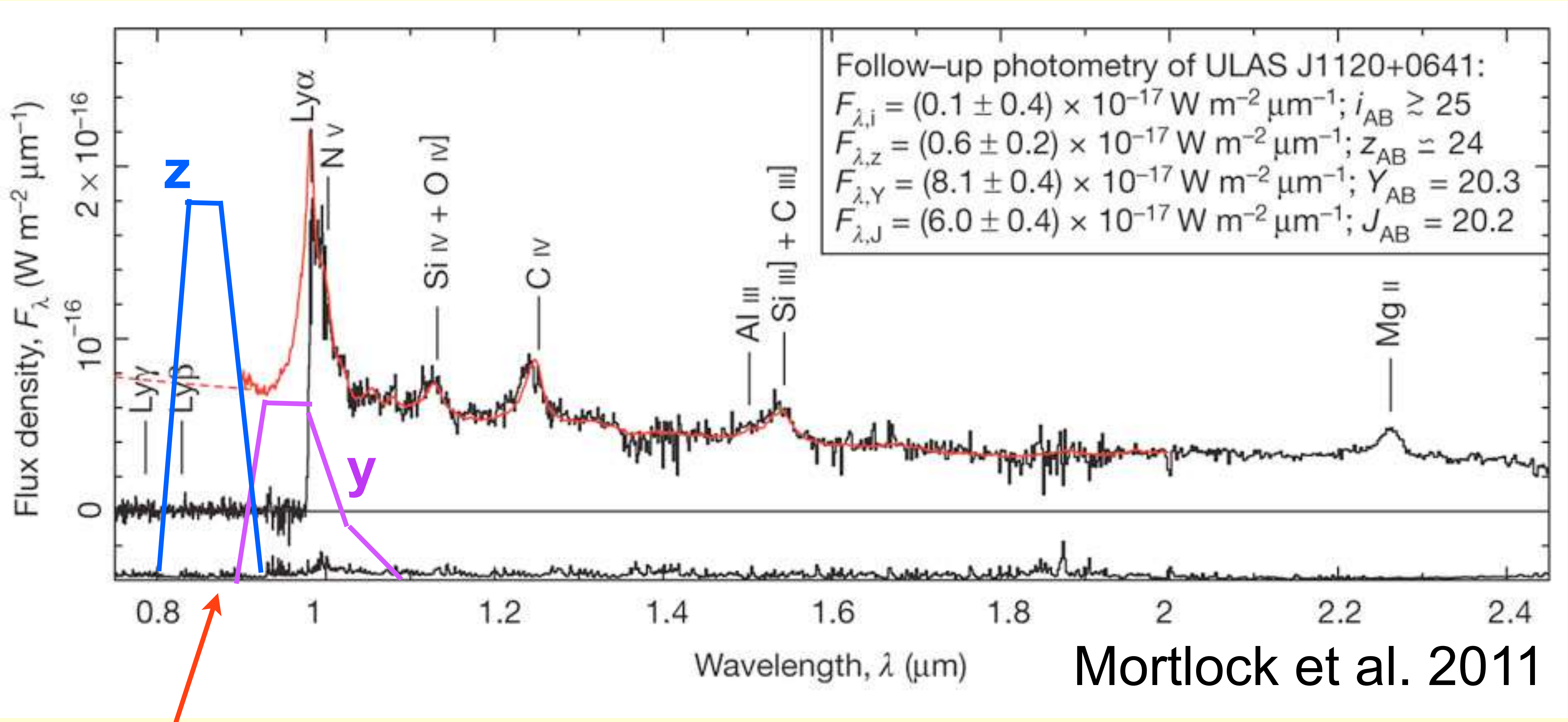

**Top:** absolute magnitude vs. redshift diagram for quasars

- About 10 million quasars will be discovered using variability, colors, and the lack of proper motions Really?? SDSS: yes!
- The sample will include Mi=-23 objects even at redshifts beyond 3
- Quasar variability studies will be based on millions of light curves with 1000 observations over 10 yrs

Today: <100 quasars with 6<z<7.5

**LSST will detect ~10,000 quasars with 6<z<7.5!**

# A High-Redshift Quasar at z=7.085 from UKIDSS




**Such a quasar would be detected by LSST as a z-band dropout**

(multi-epoch data will greatly help with false positives)

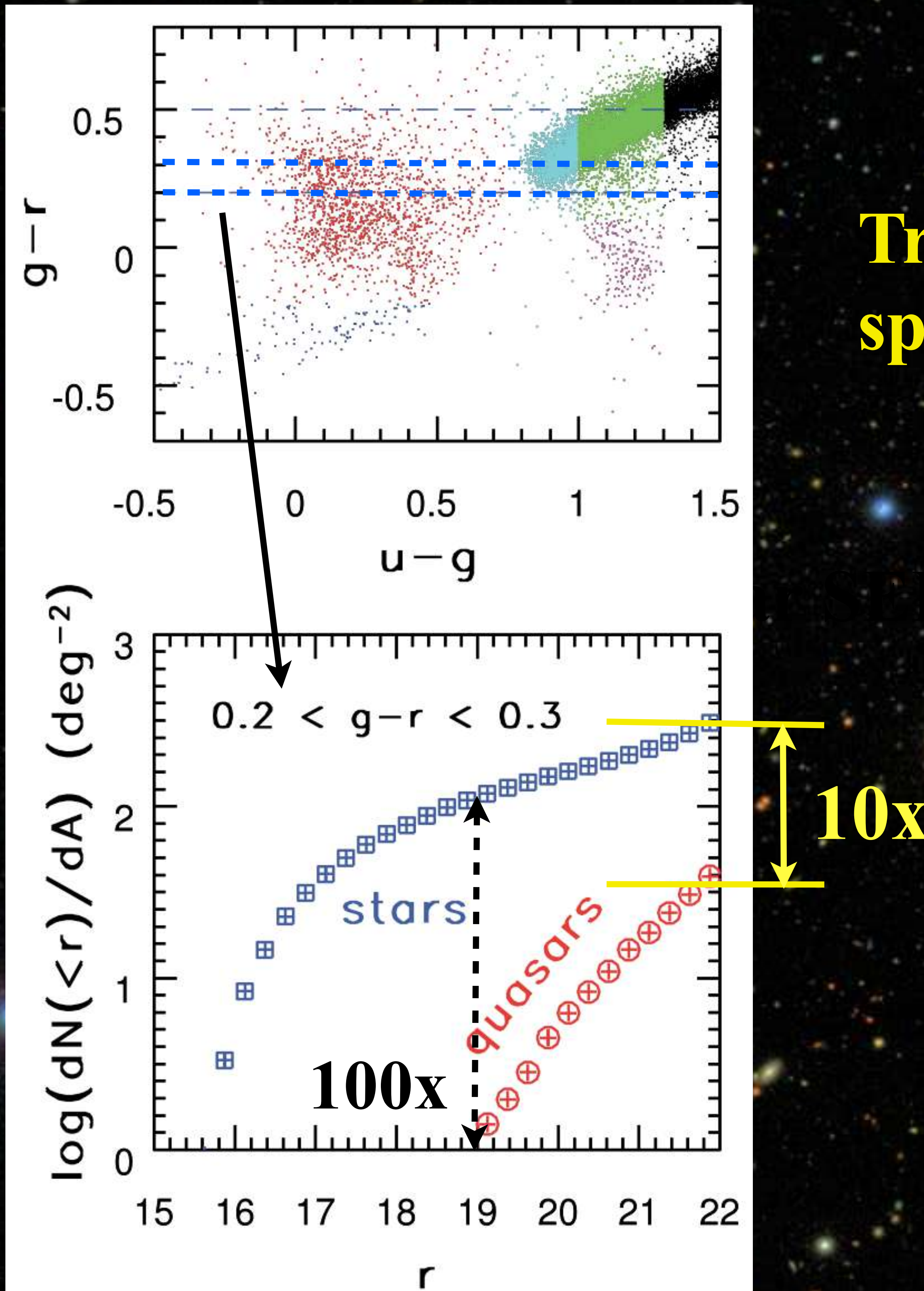
**LSST will discover about 1,000 quasars with z>7**

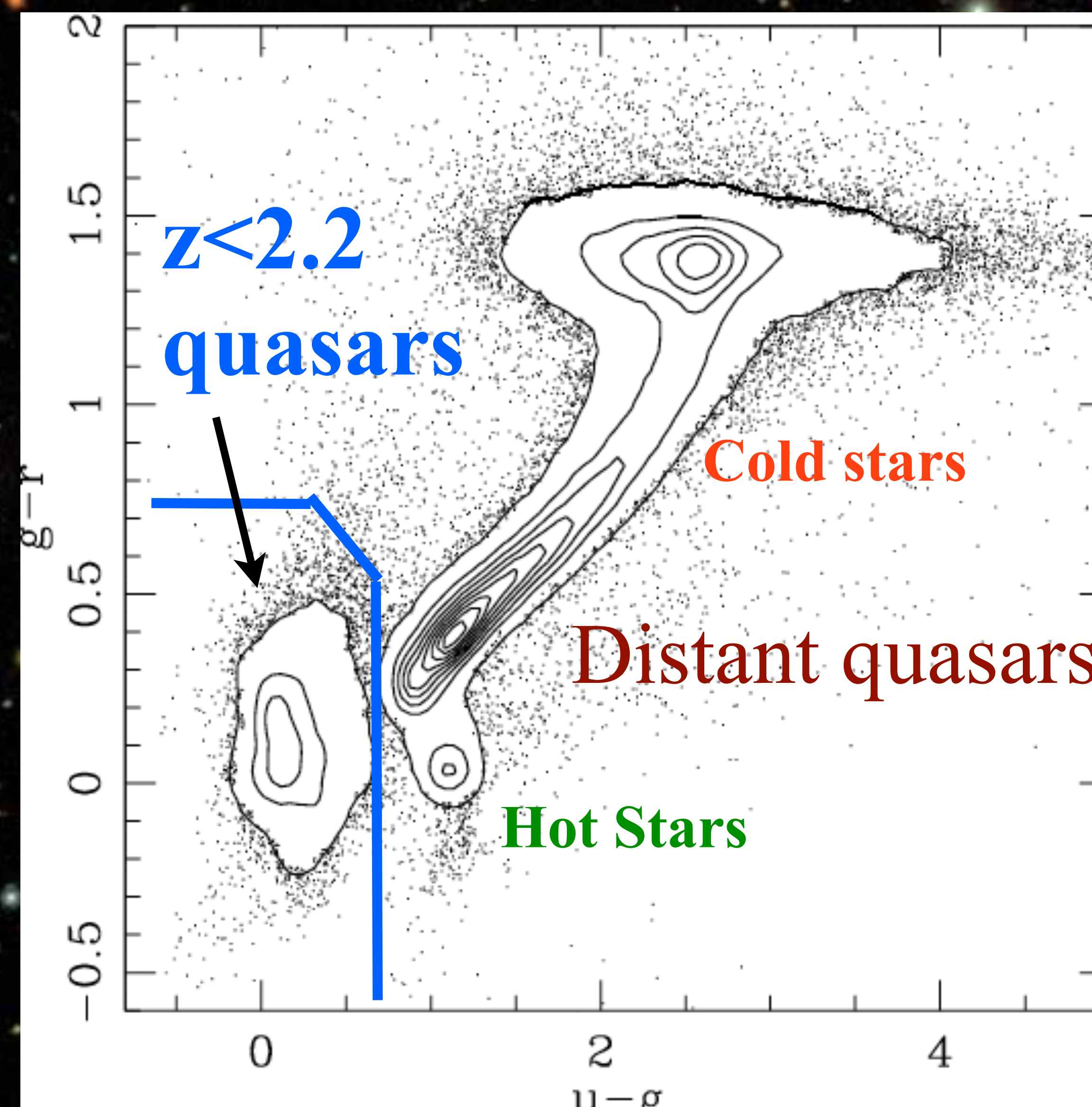
Today: <10 quasars with z>7

# Finding Quasars…

Static sky!
More information in time domain...

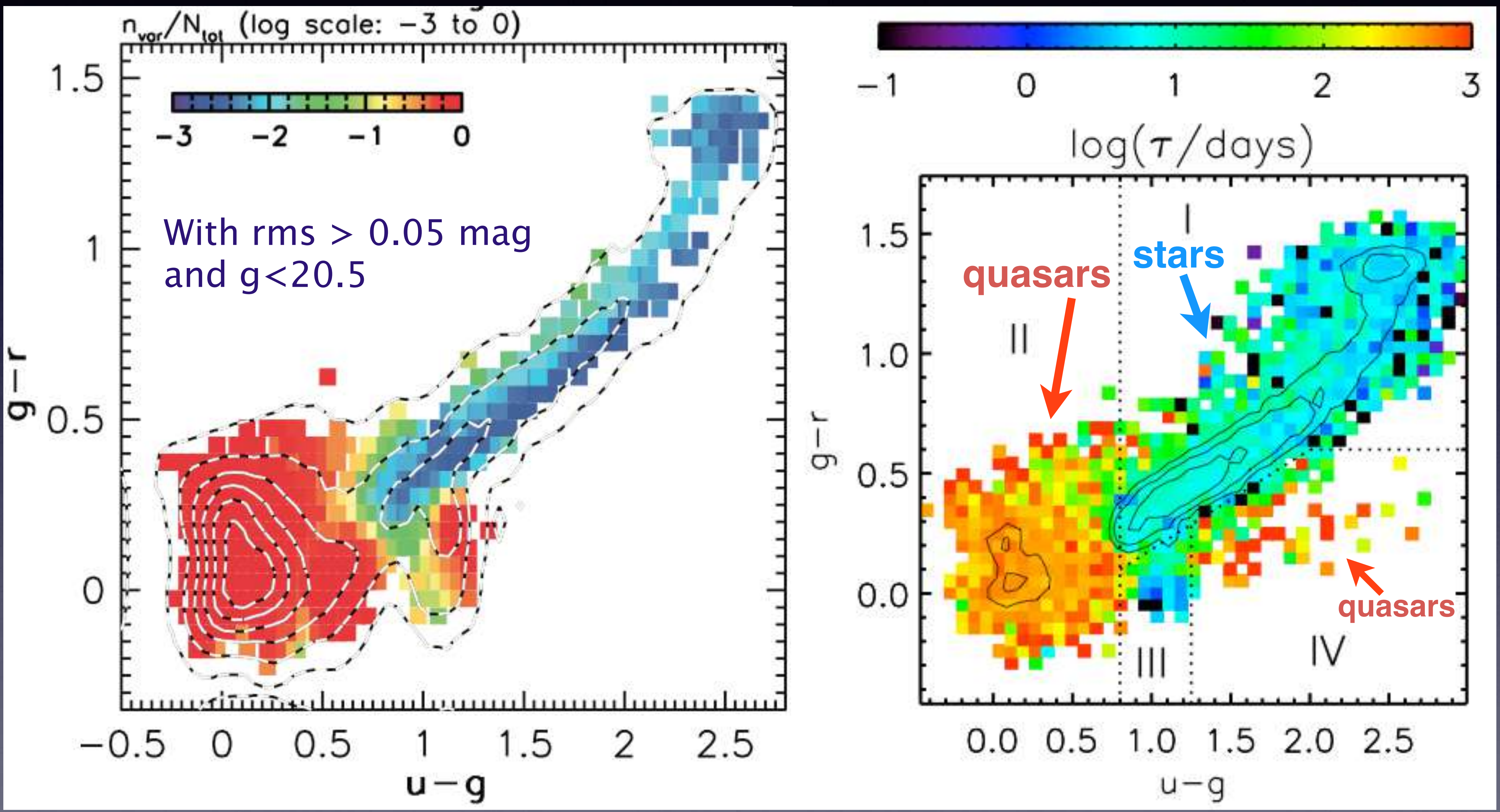
**Traditionally need spectroscopic confirmation**



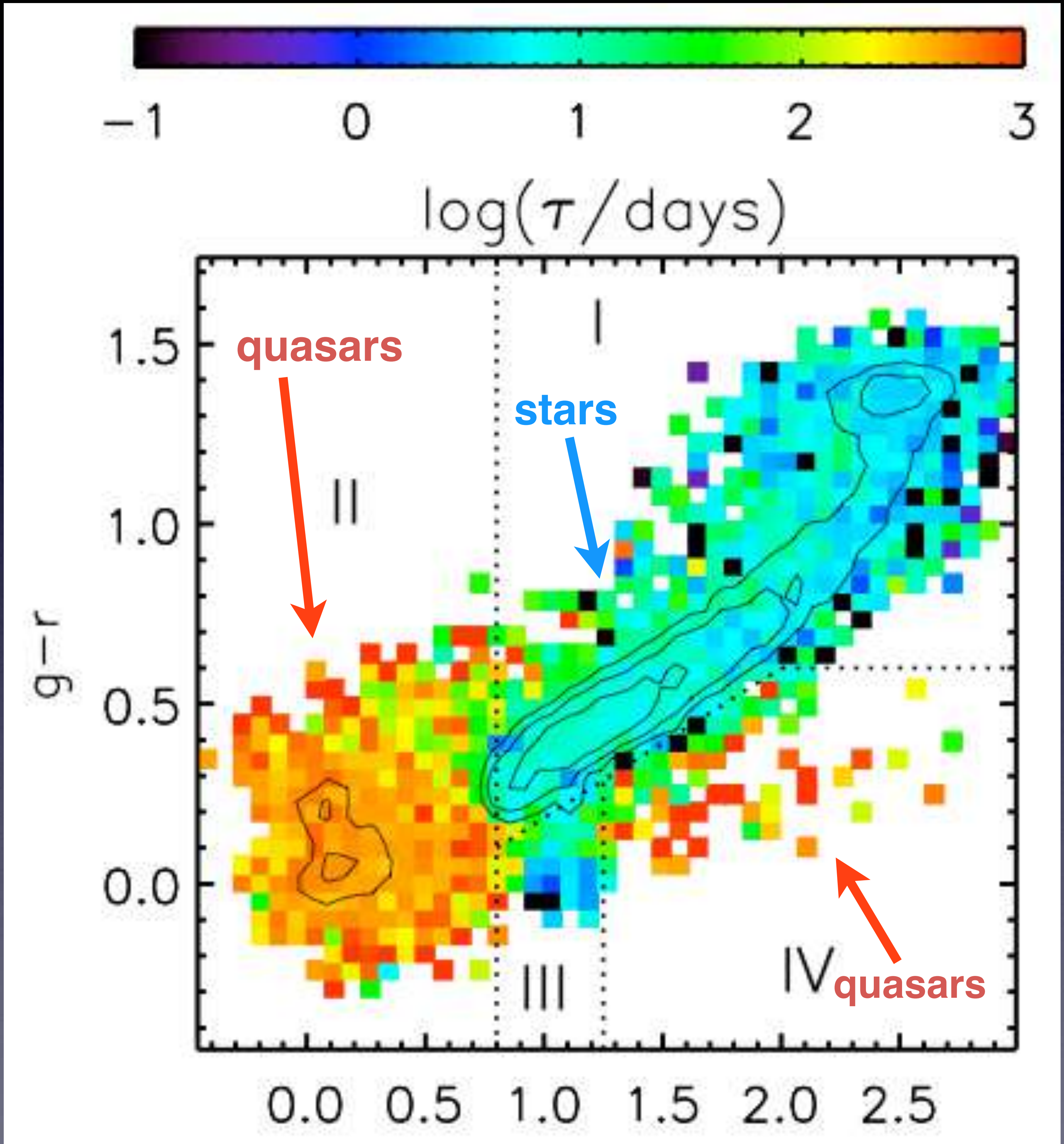



# Practically all quasars are variable!

The fraction of variable objects in SDSS Stripe 82:



# The variability time scales




McLeod et al. 2011

Time scale τ is defined via **damped random walk** (because not all variable sources are periodic)

Quasars are easily distinguished from stars by their long time scales.

**Variability is even better than color selection!**

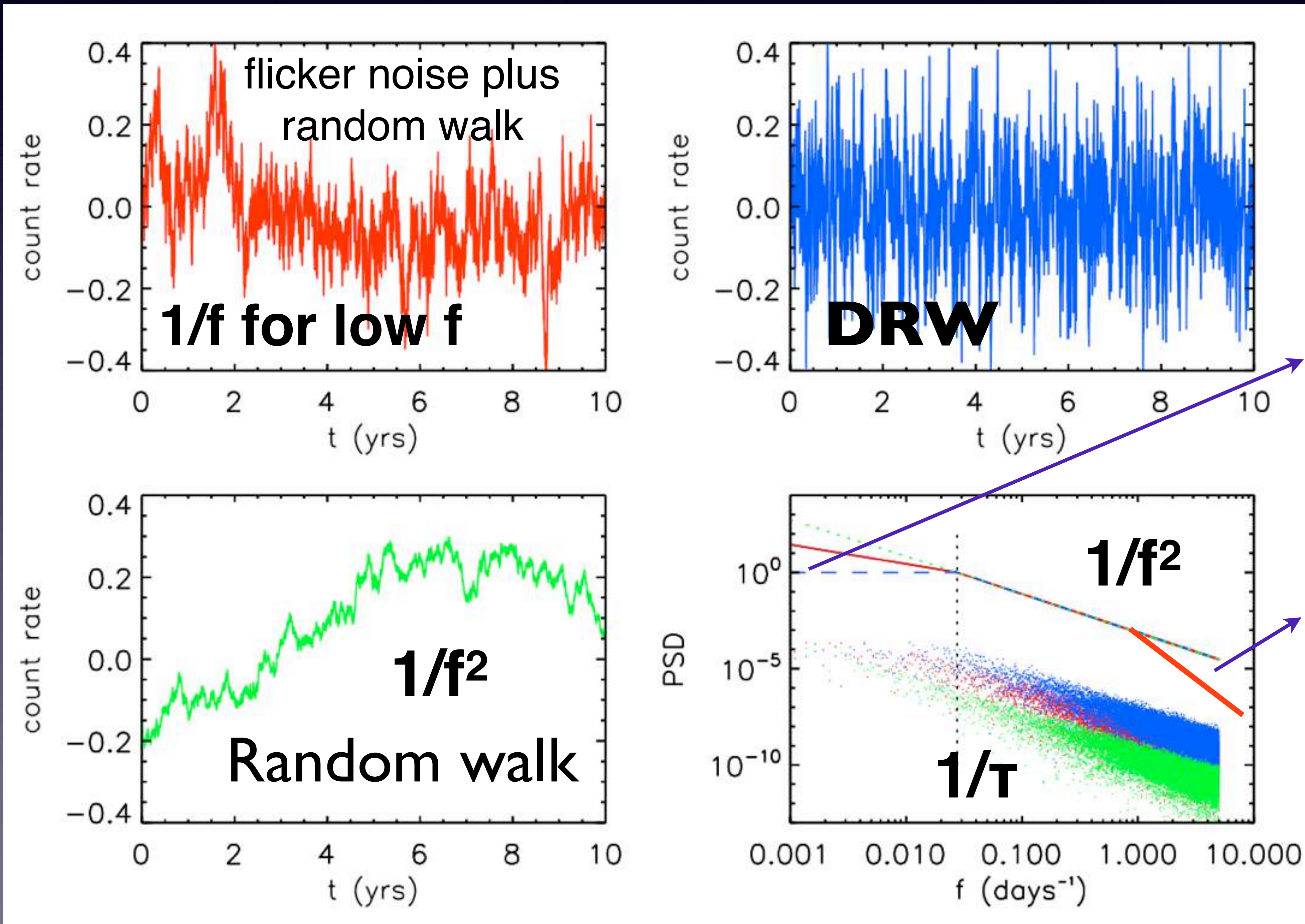
**Case study:** light curve data and proper motion data for over 1 million sources from SDSS Stripe 82 (all are publicly available)

# Damped random walk

also known as Ornstein-Uhlenbeck process and as CAR(1) process; it has exponentially decaying ACF and it is a

Stochastic process with

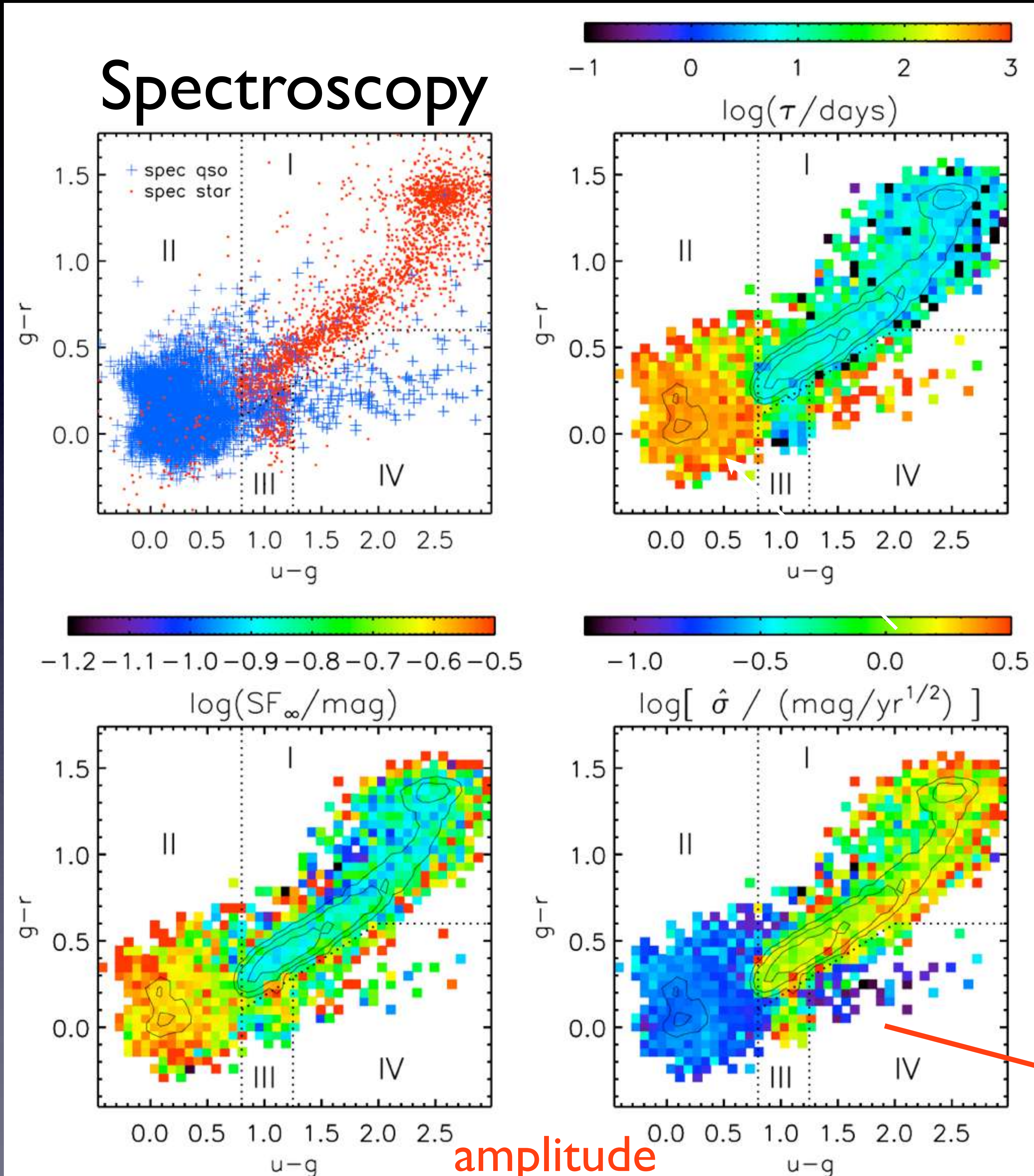
**PSD(f) = $1/f^2$ for f > 1/τ, and PSD(f)=const. for f < 1/τ**




**RDW cannot be rejected (MacLeod+201x) (Zu+2012)**

**Kepler data: $1/f^3$ for high frequencies i.e. short time scales (Mushotzky+2011)**

# Damped random walk fits to SDSS Stripe 82




3-parameter fits: DRW time scale, amplitude, and mean magnitude

Using variability, one gets the same morphology in the g-r vs. u-g diagram as when using spectroscopy!

True even for short observation spans

# Wonderful, but can we do variability selection with extended sources? YES!

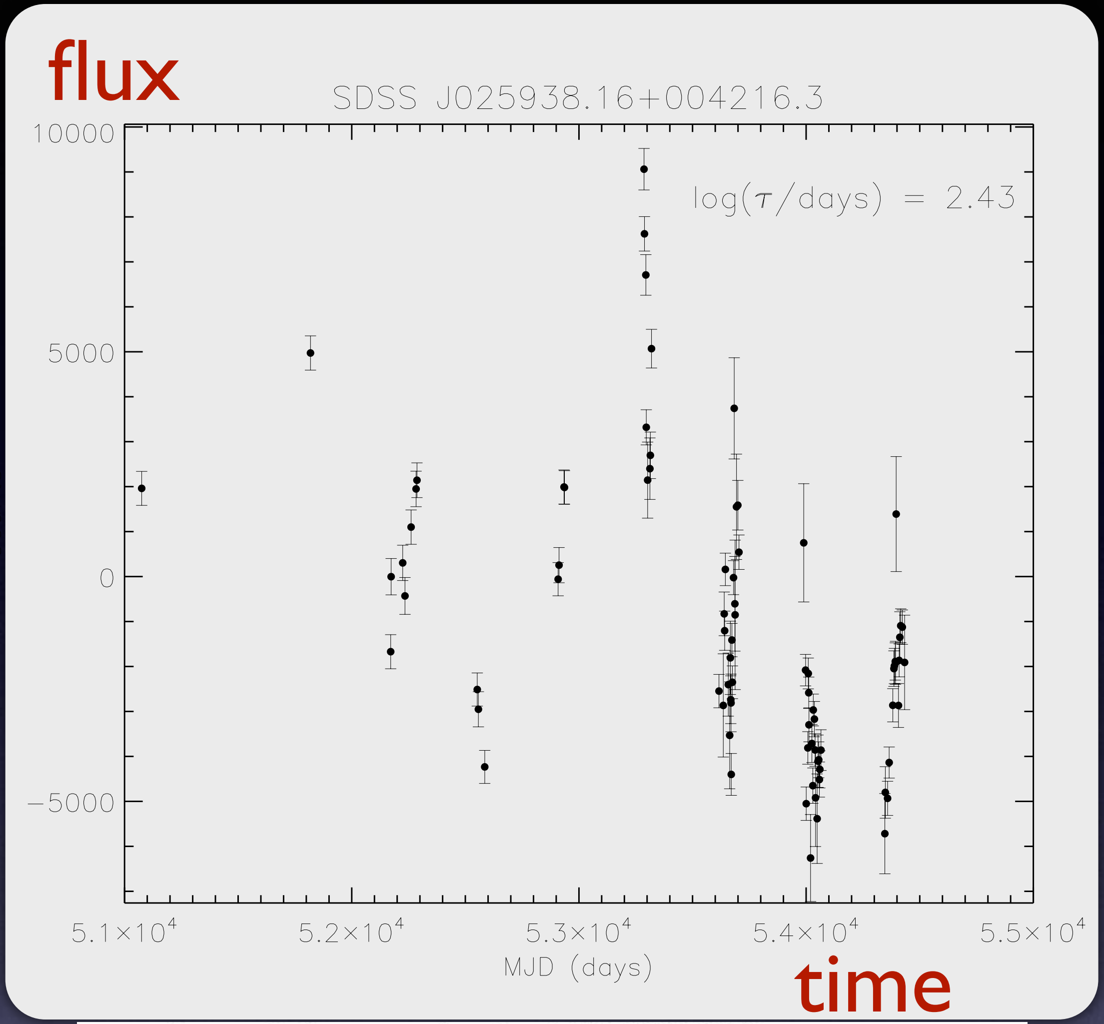
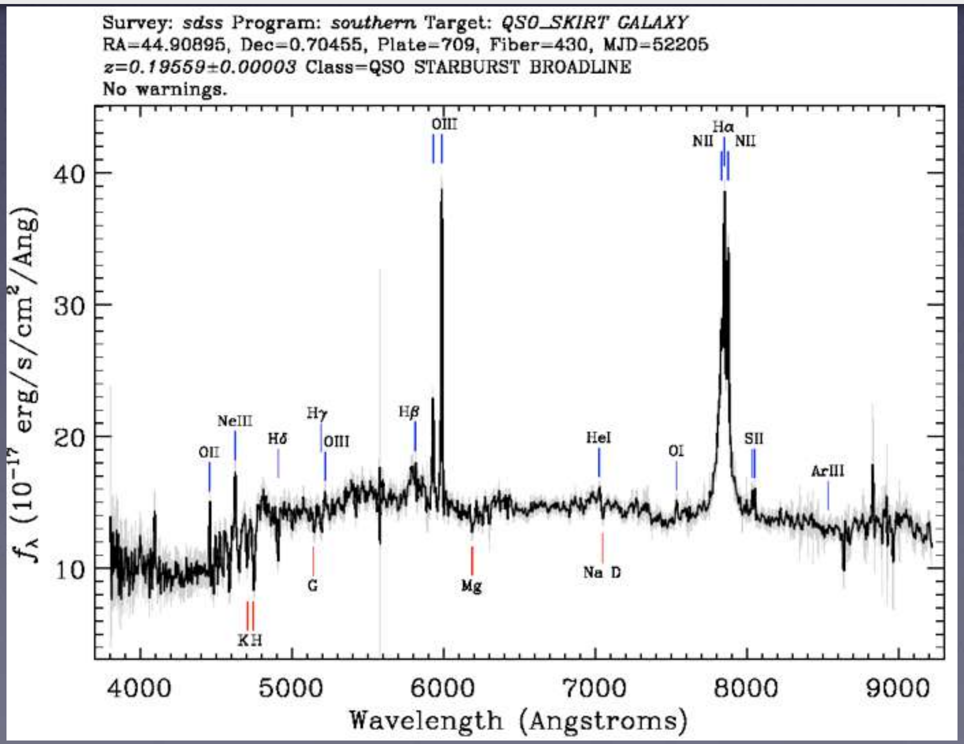
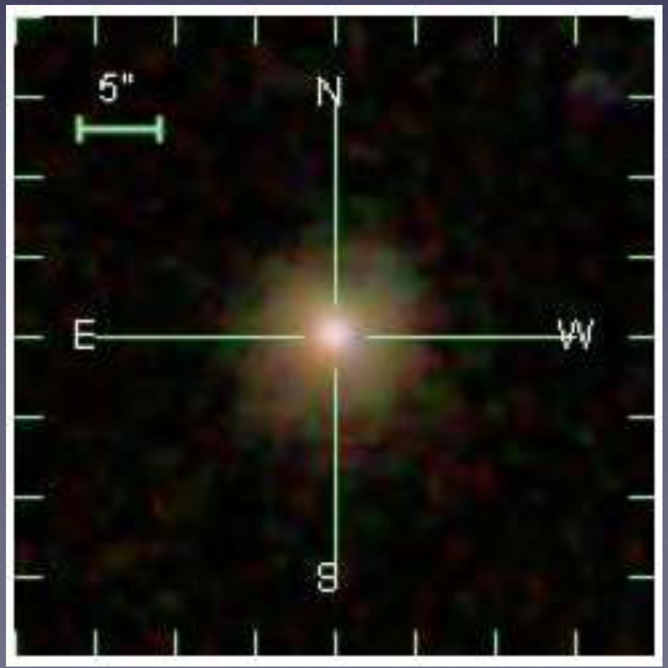



**Image differencing** can be used to extract light curves for extended sources with nearly the same SNR as for unresolved sources

**Case study:** image differencing using SDSS Stripe 82 (Choi+2014, ApJ 782, 37): light curves that have time scales as long as AGNs are independently confirmed as AGNs using X-ray data and optical emission lines (BPT diagram)!

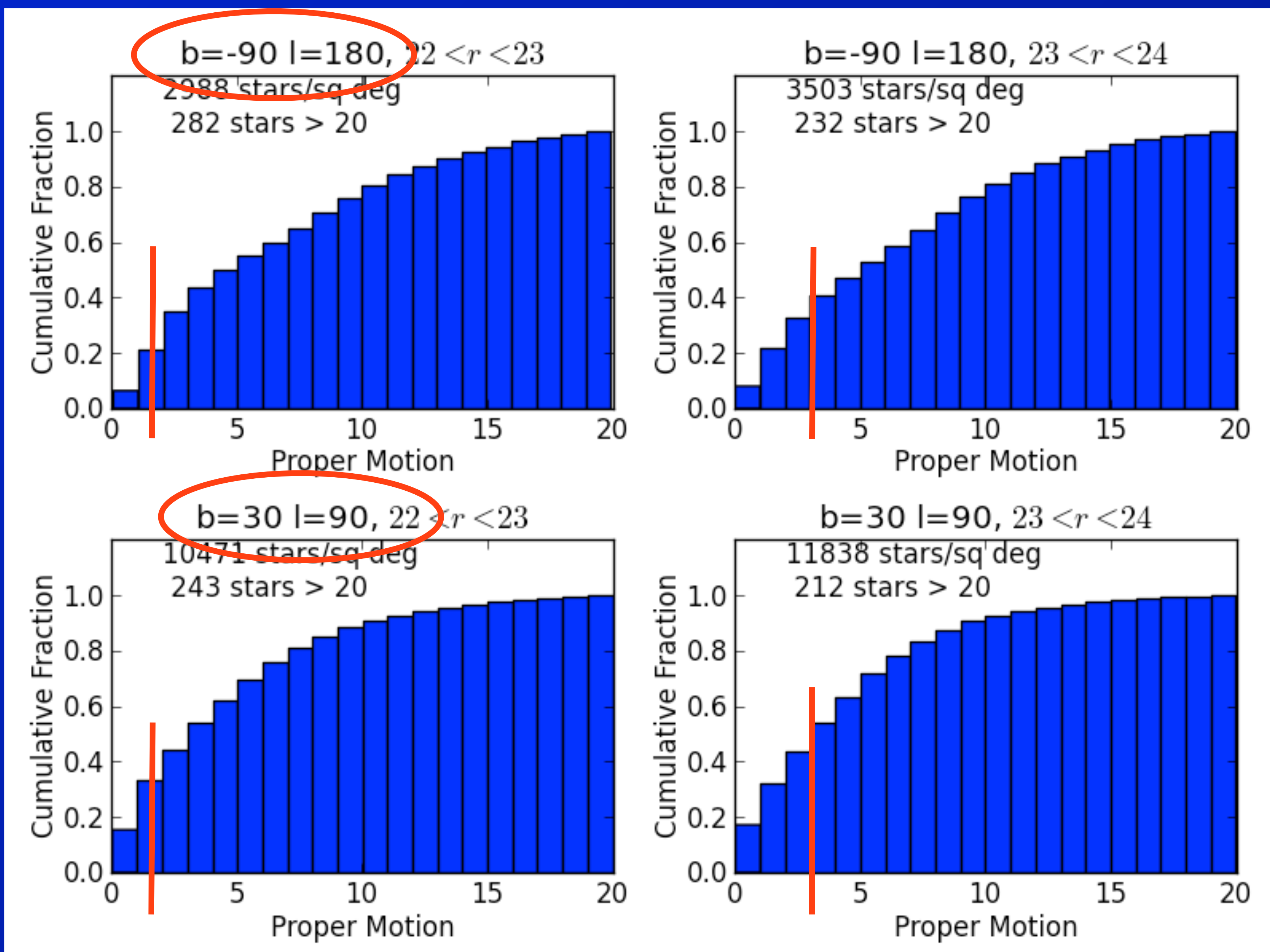


# How will LSST proper motion measurements help with the selection of faint quasars?

**LSST proper motion errors: 0.5 mas/yr for r=23 and 1.0 mas/yr for r=24.**

**22 < r < 23** **23 < r < 24**




**By adopting a 3sigma rejection cut:**

**About 2/3 of faint stars rejected due to proper motions without any selection by color or photometric variability, even at the faint end!**

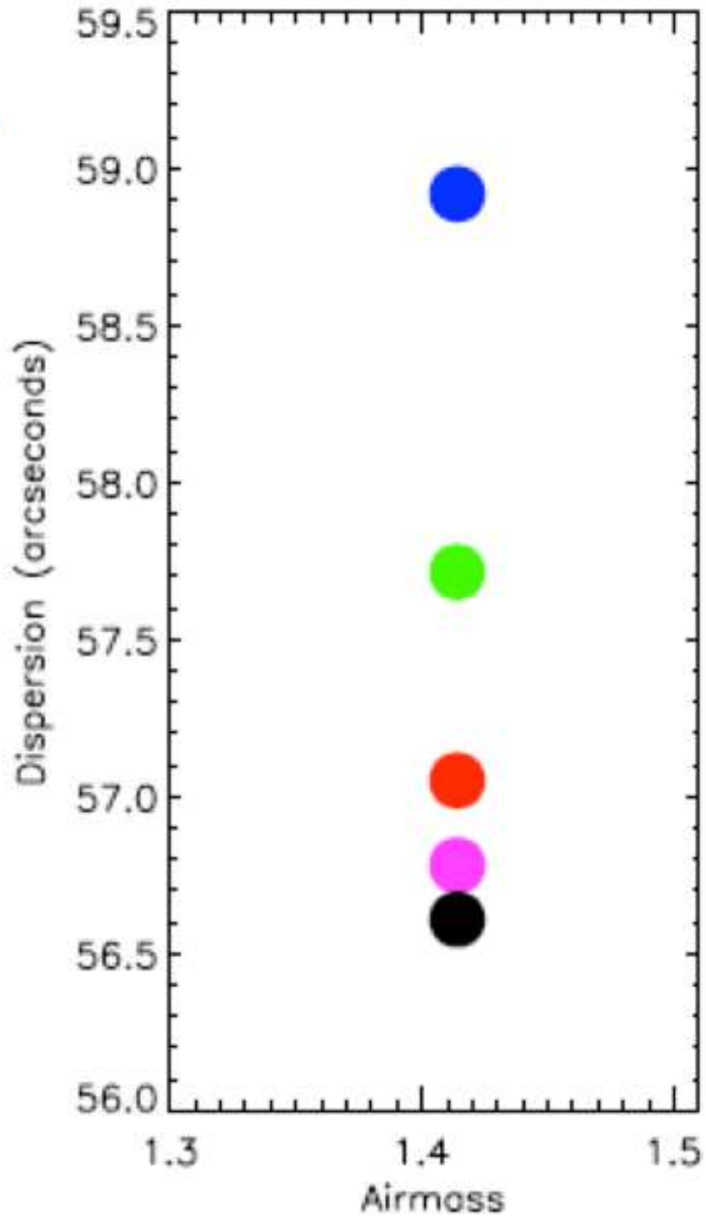
# Astrometric Classification Kaczmarczik et al. (2009)

1) Atmospheric refraction depends on object's SED (within a passband)
2) Astrometric solution is derived using stars (with different SEDs than quasars')
3) Quasar's calibrated position will change with airmass of observation:



$$R \simeq R_0 \tan(Z),$$

$$R_0 = \frac{n^2 - 1}{2n^2}$$




$$[n(\lambda) - 1]10^6 = 64.328 + \frac{29498.1}{146 - (1/\lambda)^2} + \frac{255.4}{41 - (1/\lambda)^2}$$

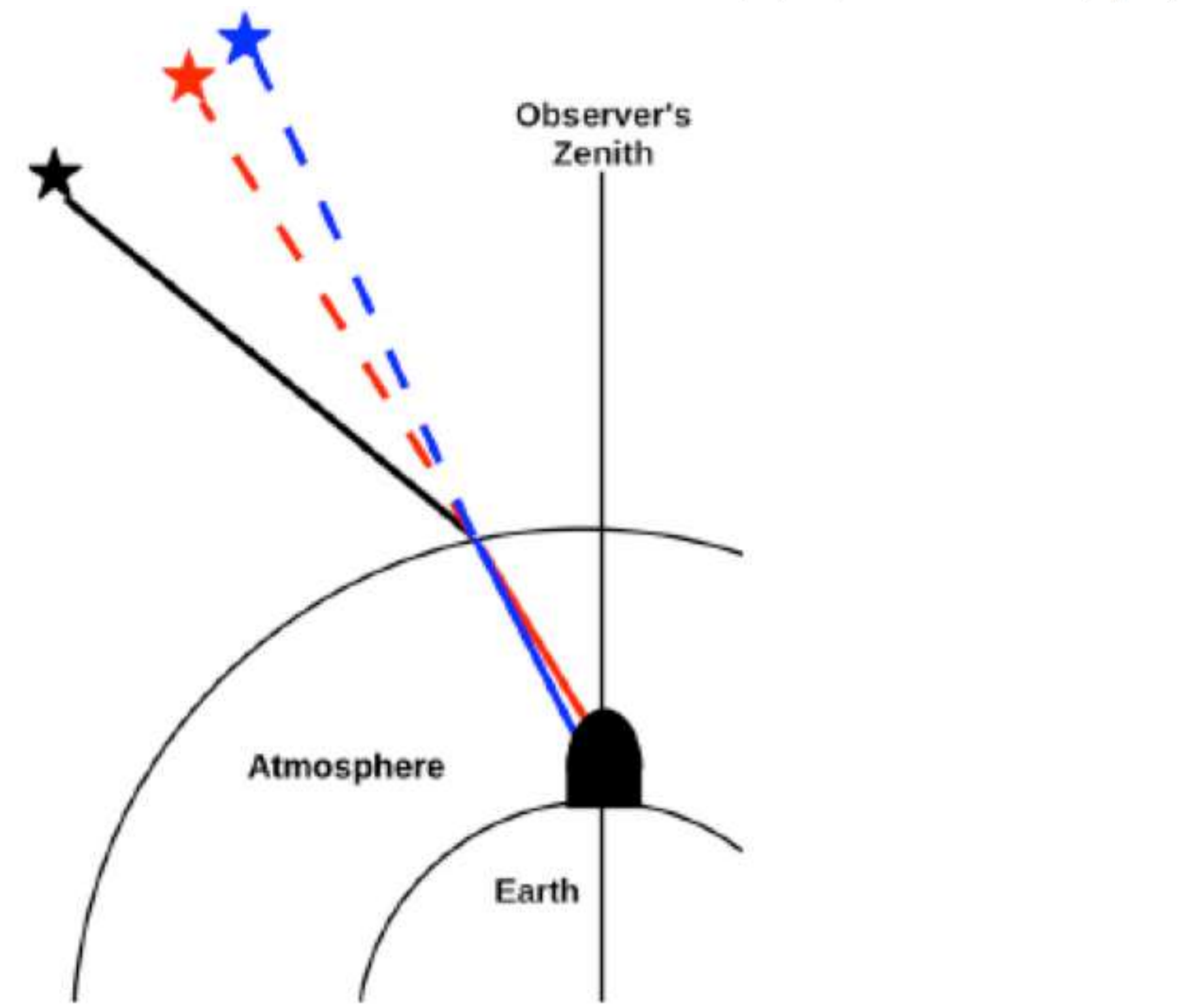


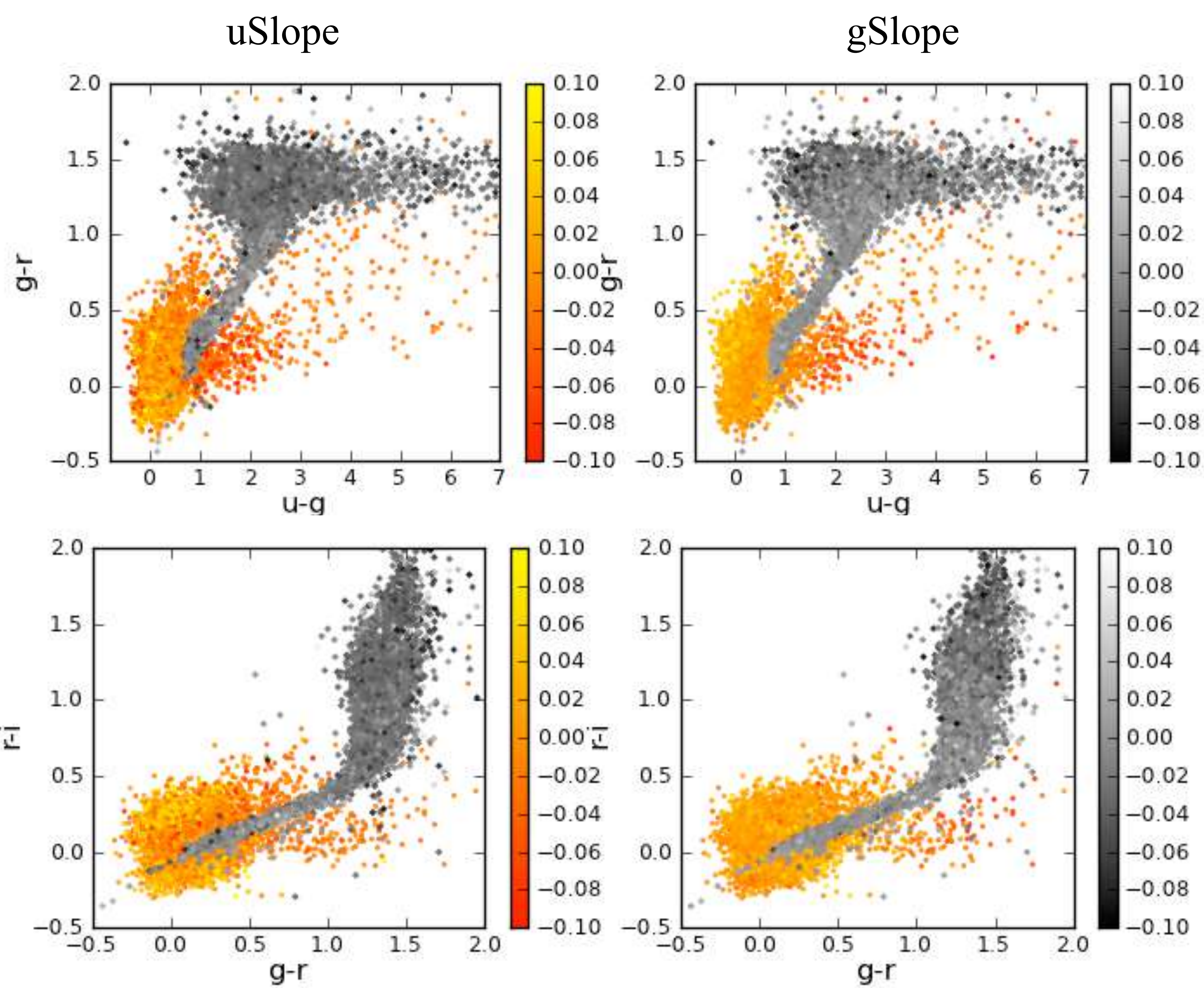

**Figure 1.** Left: DCR for a flat-spectrum object observed in the SDSS photometric system at a zenith angle of 45° (AM = 1.414). The color coding is $u$ = blue, $g$ = green, $r$ = red, $i$ = magenta, and $z$ = black. Objects appear higher in the sky when observed in blue bandpasses than in red bandpasses. Right: DCR schematic example. The solid black line indicates the incoming multi-chromatic light rays. The solid red and blue lines indicate the DCR of the incoming beam, with blue light rays being bent more than red. The dashed blue and red lines indicate the apparent location on the sky of the object as seen by the blue and red filters.

1) Atmospheric refraction depends on object's SED (within a passband)

2) Astrometric solution is derived using stars (with different SEDs than quasars')

**3) Quasar's calibrated position will change with airmass of observation:**

The slope of the change of the object's position with the airmass of observation clearly differentiates quasars and stars:




**Additional quasar selection method:**

**DCR**

**the variation of position with airmass (relative to the reference frame set by stars)**

**Kaczmarczik et al.: it is sufficient to sample airmass <1.4 (which is consistent with the baseline LSST survey cadence)**

# Plausible LSST Yields

Variability Selected Quasar Predictions from Palanque-Delabrouille et al. (2013)

**Table 8.** Predicted number of quasars over 15.5 $< g <$ 25 and 0 $< z <$ 6 for a survey covering 10 000 deg$^2$, based on our best-fit luminosity function.

| $g/z$ | 0.5 | 1.5 | 2.5 | 3.5 | 4.5 | 5.5 | Total |
|---|---|---|---|---|---|---|---|
| 15.75 | 76 | 15 | 0 | 0 | 0 | 0 | 92 |
| 16.25 | 174 | 55 | 11 | 0 | 0 | 0 | 239 |
| 16.75 | 402 | 172 | 61 | 0 | 0 | 0 | 635 |
| 17.25 | 939 | 535 | 180 | 6 | 0 | 0 | 1661 |
| 17.75 | 2163 | 1630 | 508 | 21 | 1 | 0 | 4323 |
| 18.25 | 4740 | 4720 | 1409 | 57 | 2 | 0 | 10 928 |
| 18.75 | 9456 | 12 380 | 3784 | 156 | 5 | 0 | 25 781 |
| 19.25 | 16 612 | 27 796 | 9409 | 422 | 14 | 0 | 54 255 |
| 19.75 | 25 537 | 51 561 | 20 579 | 1128 | 39 | 1 | 98 846 |
| 20.25 | 35 185 | 80 209 | 38 096 | 2923 | 107 | 4 | 156 523 |
| 20.75 | 45 008 | 110 341 | 59 939 | 7085 | 289 | 10 | 222 671 |
| 21.25 | 54 980 | 141 918 | 82 650 | 15 386 | 779 | 27 | 295 740 |
| 21.75 | 64 988 | 176 959 | 103 733 | 28 916 | 2036 | 74 | 376 706 |
| 22.25 | 74 189 | 217 815 | 122 861 | 46 636 | 5064 | 201 | 466 766 |
| 22.75 | 80 370 | 266 716 | 141 310 | 65 652 | 11 408 | 545 | 566 001 |
| 23.25 | 79 024 | 325 945 | 160 621 | 82 972 | 22 419 | 1436 | 672 417 |
| 23.75 | 61 347 | 398 006 | 182 048 | 97 320 | 37 756 | 3632 | 780 110 |
| 24.25 | 15 976 | 480 676 | 206 510 | 109 295 | 55 090 | 8401 | 875 949 |
| 24.75 | 0 | 492 283 | 234 874 | 120 118 | 71 481 | 17 111 | 935 866 |
| Total | 571 169 | 2 789 734 | 1 368 583 | 578 092 | 206 489 | 31 444 | 5 545 510 |

**Notes.** Bins are centered on the indicated magnitude and redshift values. The ranges in each bin are $\Delta g = 0.5$ and $\Delta z = 1$.

**LSST: 10 million!**

where we call "quasar" an object with a luminosity $M_i[z = 2] < -20.5$ and either displaying at least one emission line with FWHM greater than 500 km s$^{-1}$ or, if not, having interesting/complex absorption features.

**The connections between optical and radio regimes:**

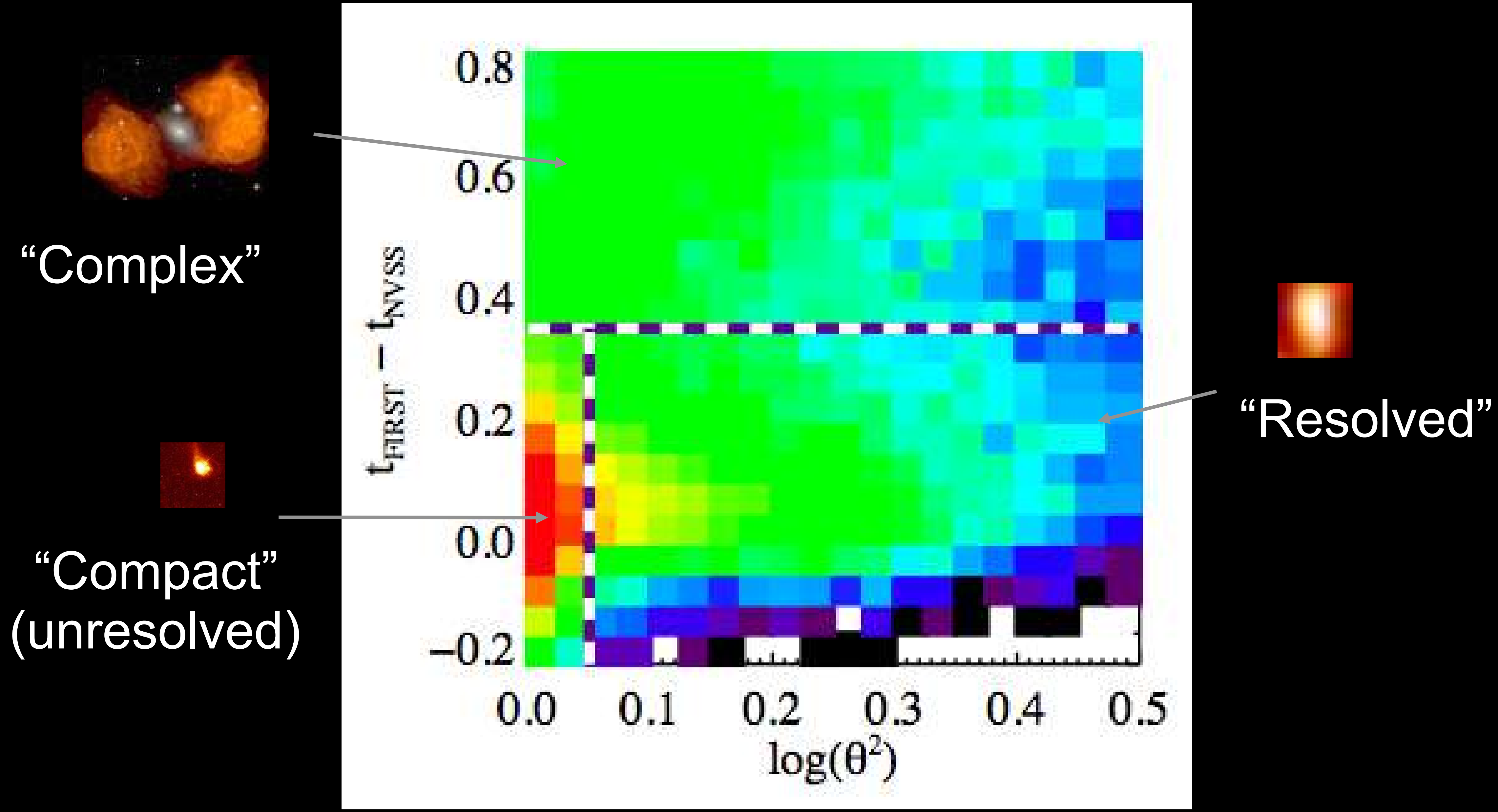
1) Science Results (asking similar and often same questions; e.g. stellar and galaxy formation and evolution, dark energy)

2) Tools and Methods (e.g. massive databases)

3) Supplemental data (identification, physical processes, HI)

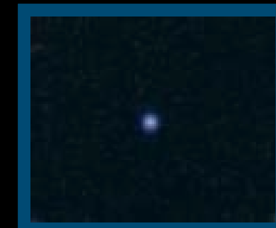
# AUTOMATED radio morphology classification for over 100,000 radio sources

## FIRST vs. NVSS flux, and FIRST peak vs. integrated flux:



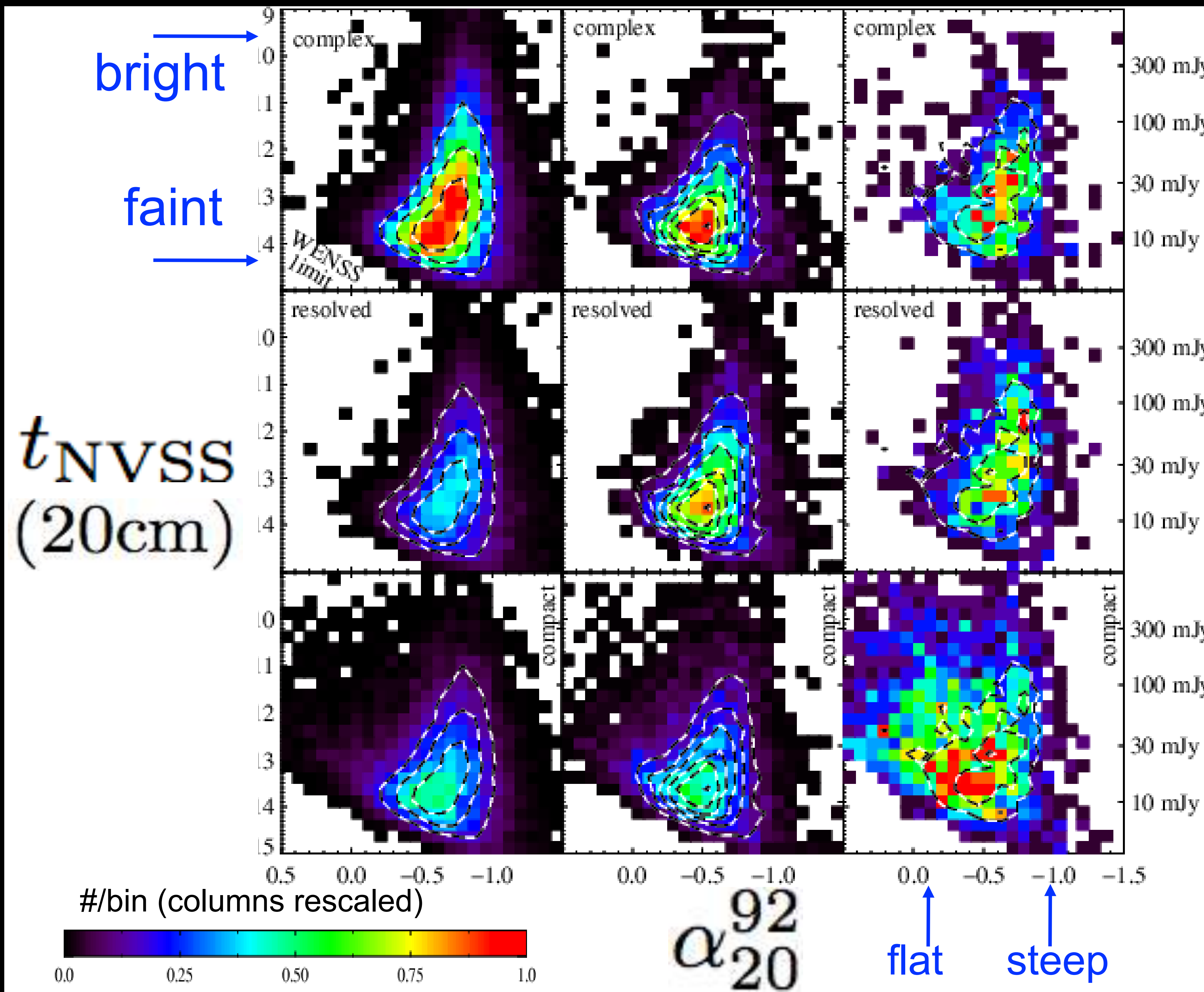

Kimball & Ivezić 2008

# Radio “color”-mag. diagrams

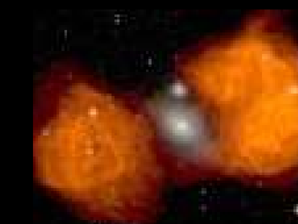


71% 22% 7%

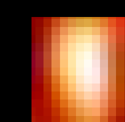
Optical: No SDSS | Galaxies | Point sources




Complex 45%



Resolved 25%



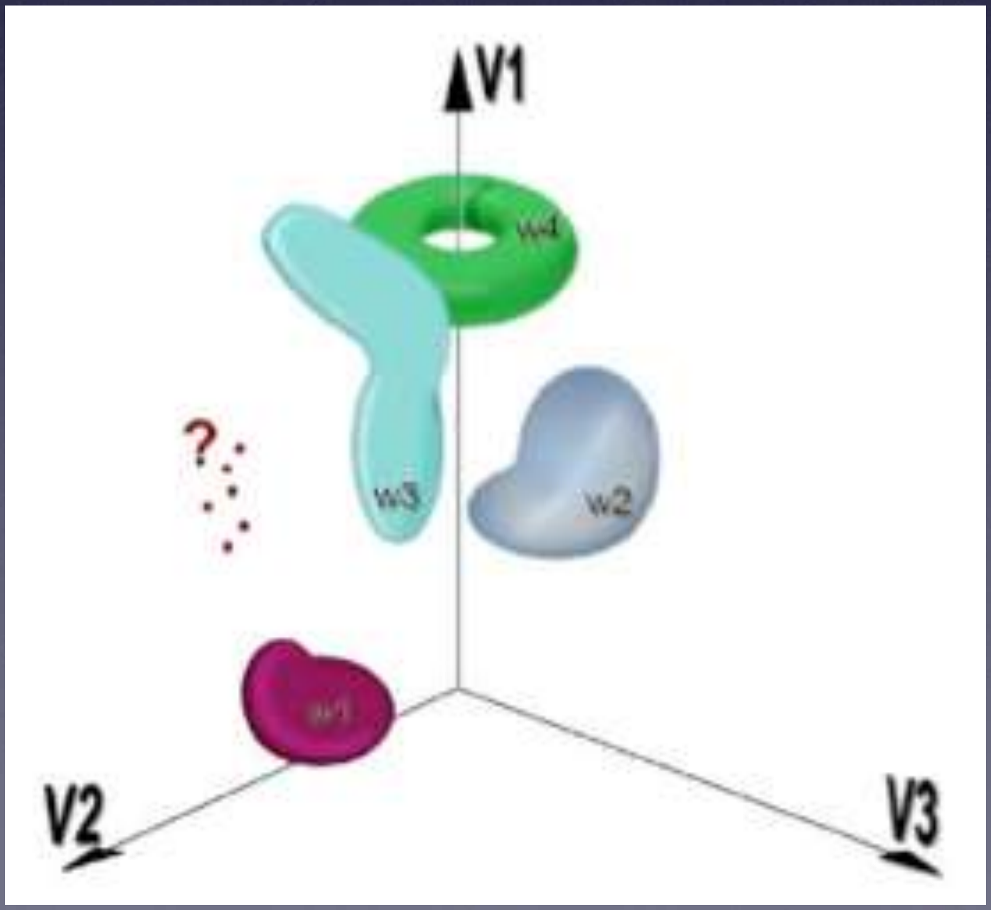
Compact 30%

**massive statistical studies!**

# Statistical analysis of a massive LSST dataset

- A large (100 PB) database and sophisticated analysis tools: for each of 40 billion objects there will be about 1000 measurements (each with a few dozen measured parameters)

## Data mining and knowledge discovery




- 10,000-D space with 40 billion points
- Characterization of known objects
- Classification of new populations
- Discoveries of unusual objects

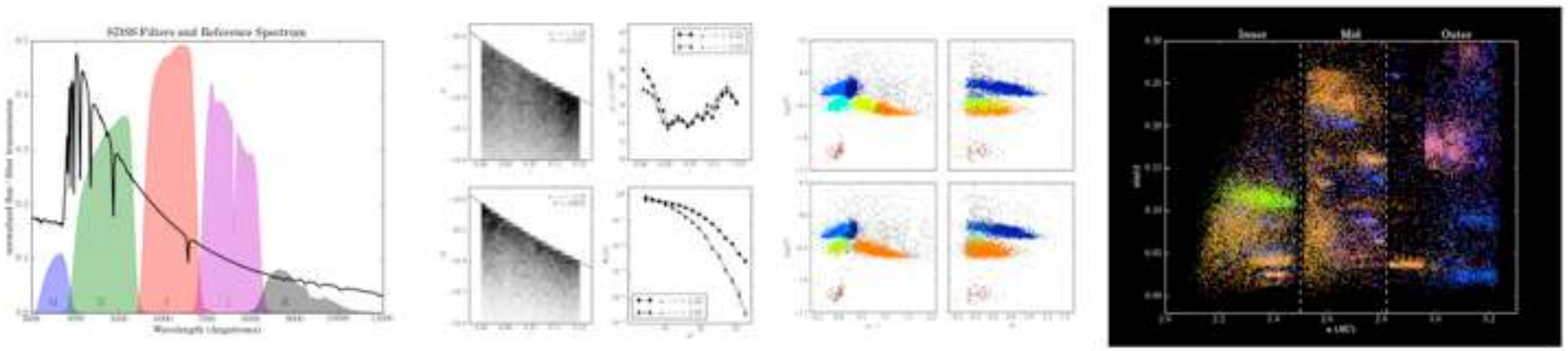
Clustering, classification, outliers

astroML

Home | User Guide | Book Figures | Examples Plots

Google™ Custom Search

# AstroML: Machine Learning and Data Mining for Astronomy



## 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

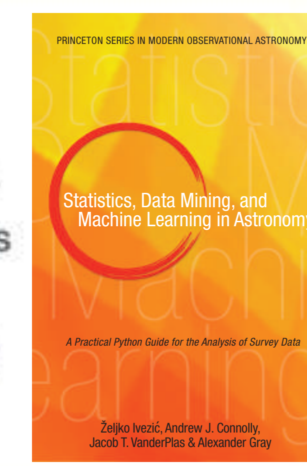
If you use the software, please consider citing astroML.

AstroML 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 easy-to-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)`.




## User Guide

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

### 1. Introduction

- 1.1. Philosophy

# Textbook Figures

This section makes available the source code used to generate every figure in the book *Statistics, Data Mining, and Machine Learning in Astronomy*. Many of the figures are fairly self-explanatory, though some will be less so without the book as a reference. The table of contents of the book can be seen **here(pdf)**.

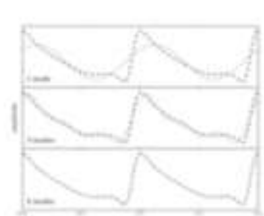
## Figure Contents

Each chapter links to a page with thumbnails of the figures from the chapter.

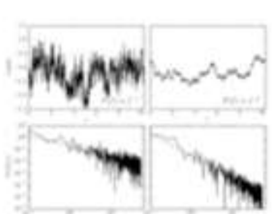
- Chapter 1: Introduction
- Chapter 2: Fast Computation and Massive Datasets
- Chapter 3: Probability and Statistical Distributions
- Chapter 4: Classical Statistical Inference
- Chapter 5: Bayesian Statistical Inference
- Chapter 6: Searching for Structure in Point Data
- Chapter 7: Dimensionality and its Reduction
- Chapter 8: Regression and Model Fitting
- Chapter 9: Classification
- Chapter 10: Time Series Analysis
- Appendix

# Chapter 10: Time Series Analysis

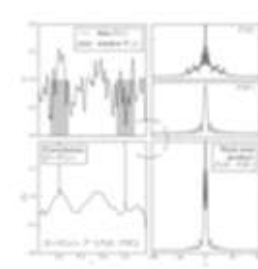
This chapter covers the analysis of both periodic and non-periodic time series, for both regularly and irregularly spaced data.



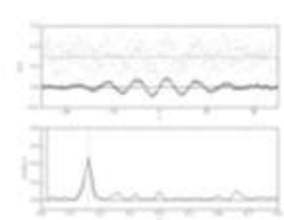
*Fourier Reconstruction of RR-Lyrae Templates*



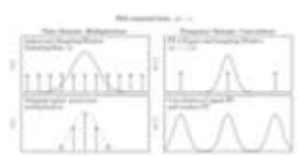
*Generating Power-law Light Curves*



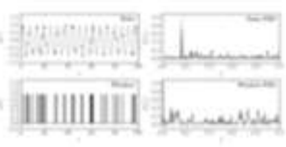
*Plot a Diagram explaining a Convolution*



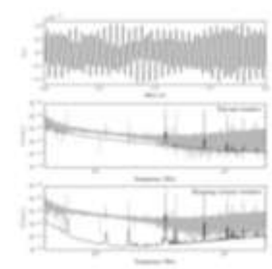
*Fast Fourier Transform Example*



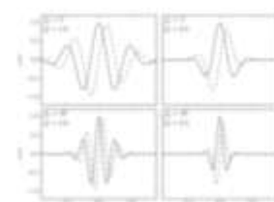
*The effect of Sampling*



*The effect of Sampling*



*Plot the power spectrum of the LIGO big dog event*



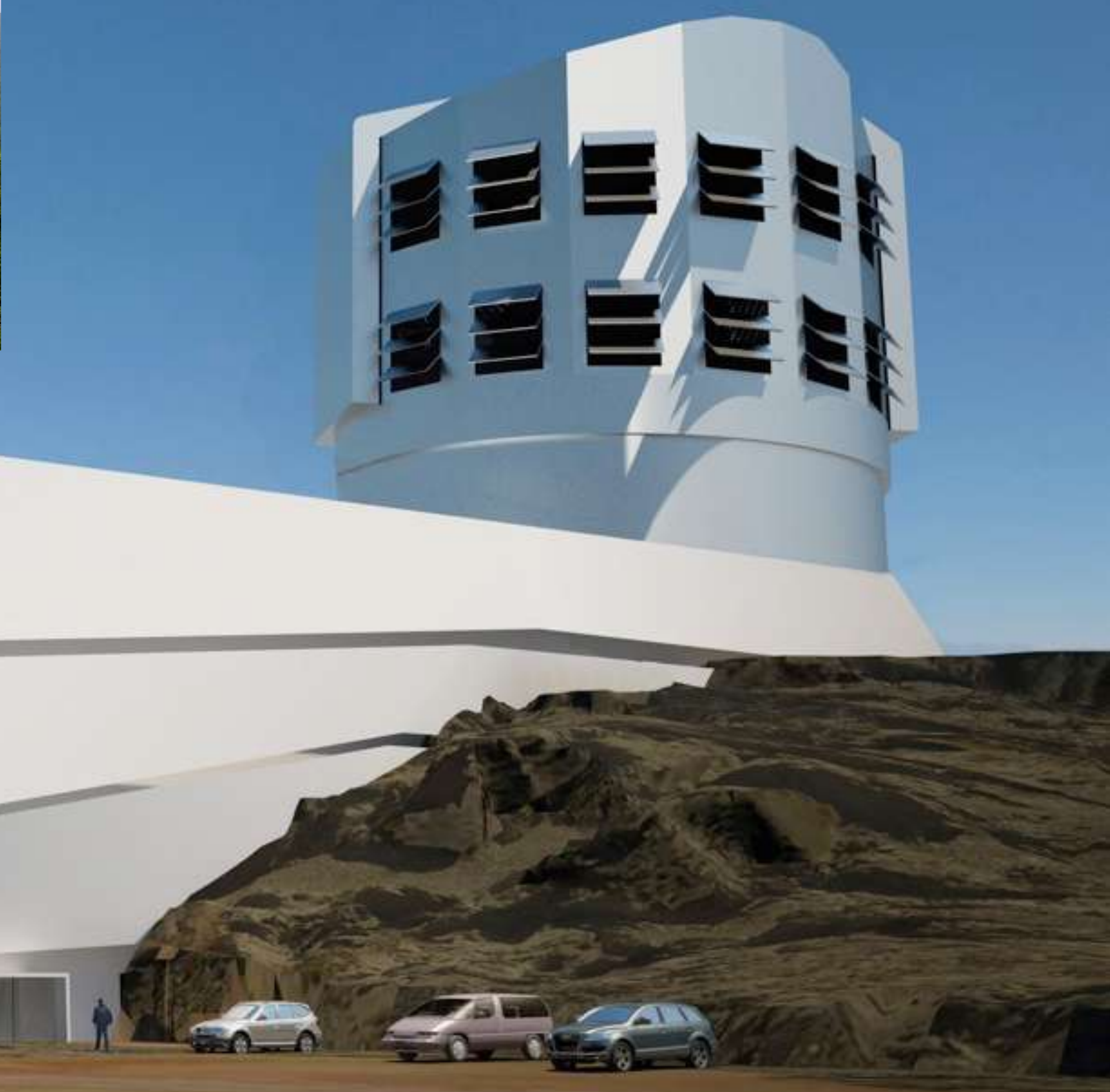
*Examples of Wavelets*

# SUMMARY

- Finding quasars/AGNs with SDSS
  color selection of quasars produced samples with ~200,000 spectroscopically confirmed objects, and ~1 million quasar candidates!
- Finding quasars/AGNs with LSST
  a combination of photometry (colors and variability) and astrometry (no proper motion and DCR) will yield a highly clean and complete sample of 10 million objects, including ~10,000 quasars at redshifts exceeding ~6 (and 1,000 with $z>7$)!

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!"



Quasar variability is stochastic - how can we describe it mathematically?

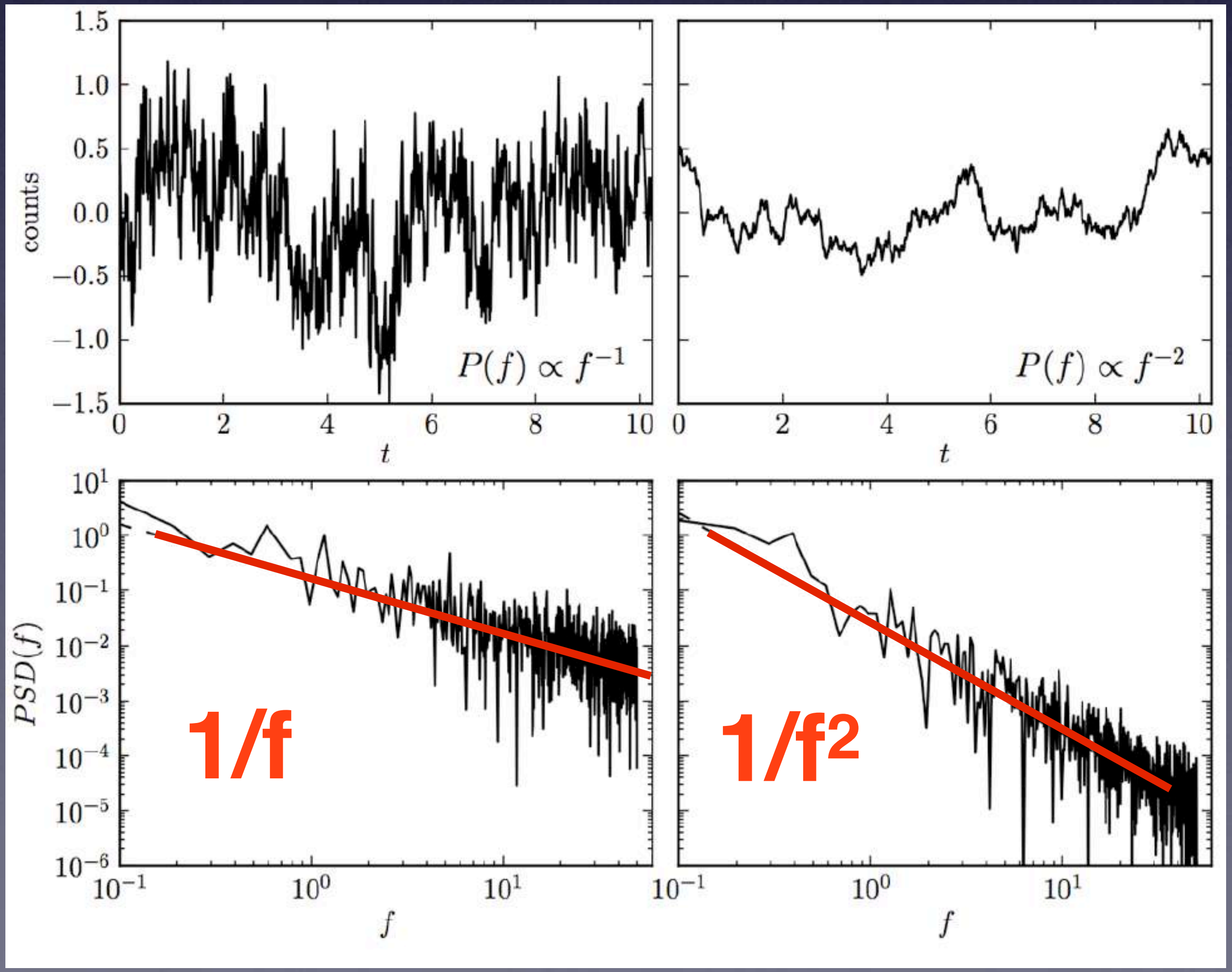
# Damped random walk

also known as Ornstein-Uhlenbeck process and as CAR(1) process; it has exponentially decaying ACF and it is a

Stochastic process with

**PSD(f) = $1/f^2$ for f > 1/τ, and PSD(f)=const. for f < 1/τ**

So, like ordinary random walk at high frequencies, but less power at low frequencies: the process tends towards its mean value and doesn't drift away

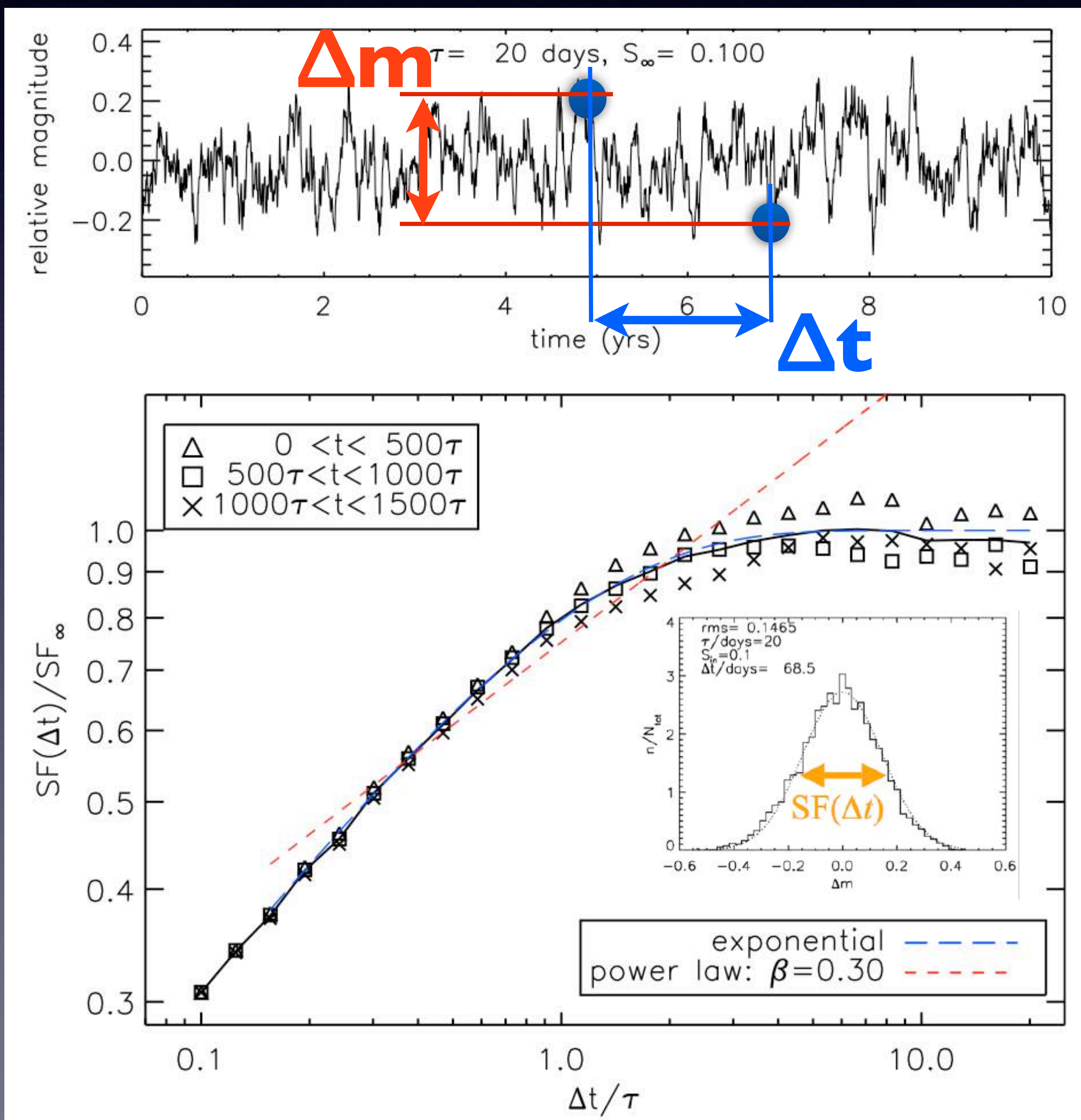



**LEFT:**

Both PSDs have equal power at low frequencies; the 1/f on the left has more power at high frequencies and thus the light curve in the top left panel appears more "noisy"

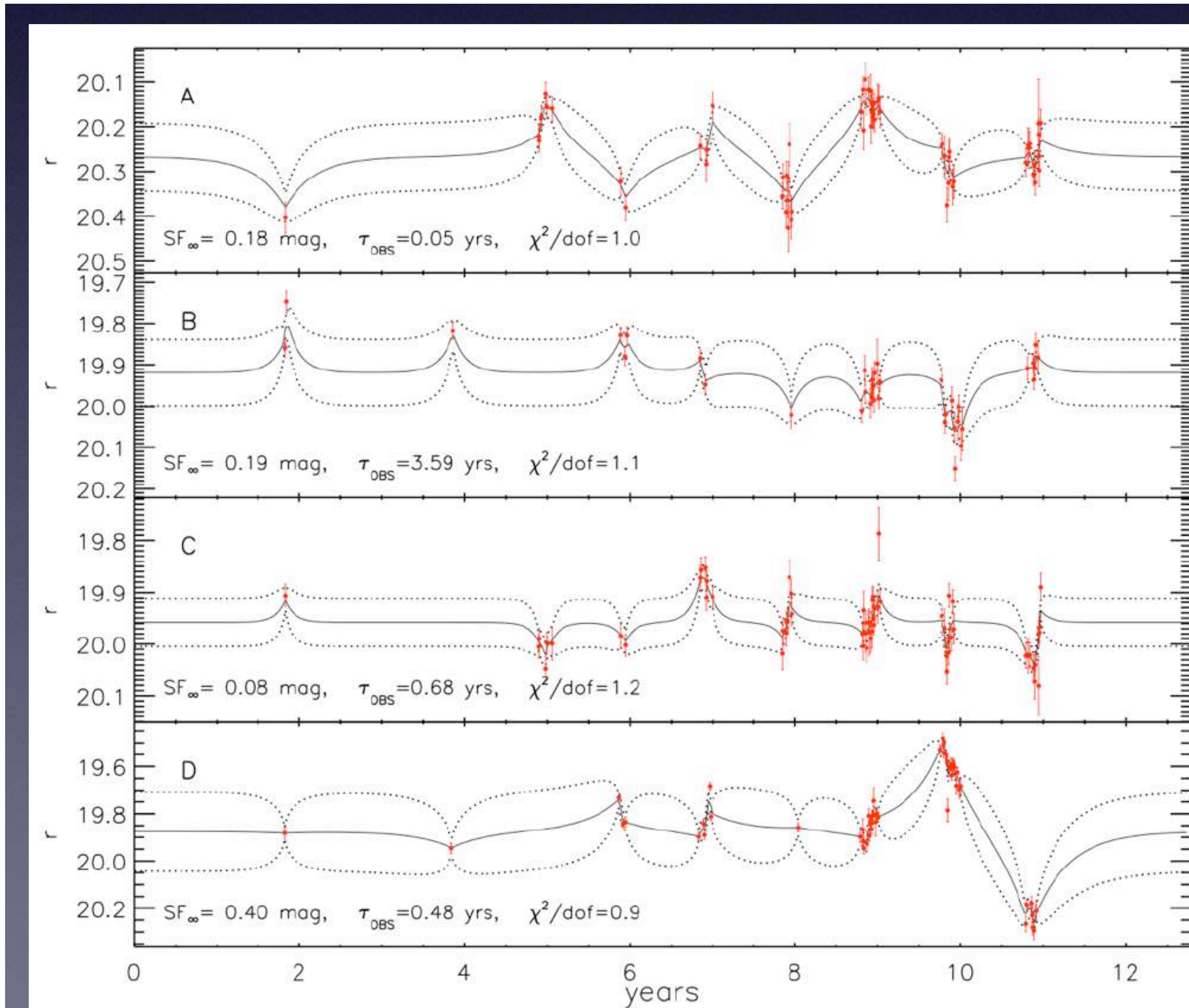
# Damped random walk

For irregularly sampled data, best analyzed using the structure function, or alternatively by fitting individual light curves for the best-fit time scale τ and variability long-term variance (e.g. see "Gaussian Processes" in Numerical Recipes, or Kozłowski, S., et al. 2010, ApJ, 708, 927)



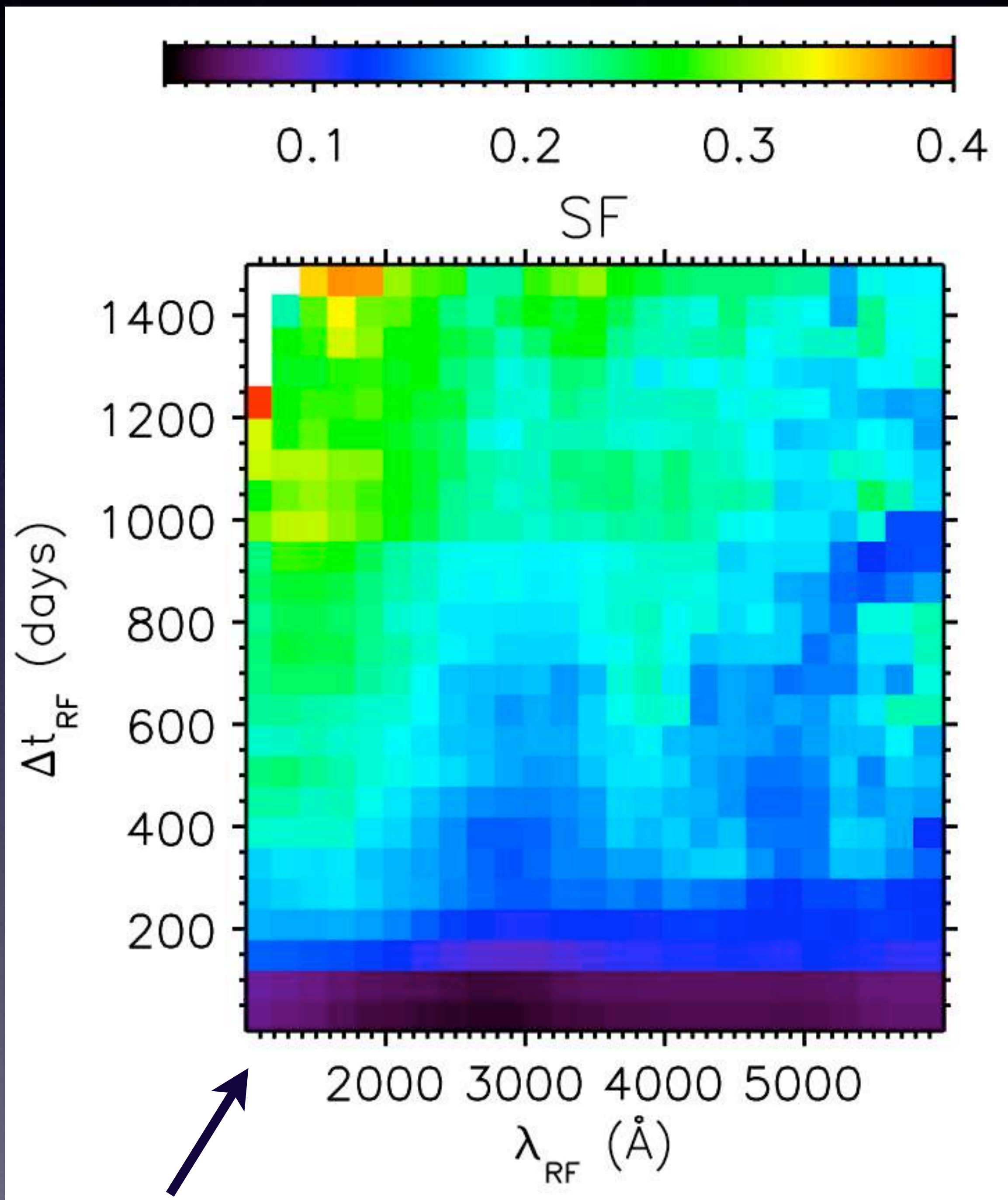

auto-correlation function

$$SF(\Delta t) = SF_\infty (1 - e^{-|\Delta t|/\tau})^{1/2}$$

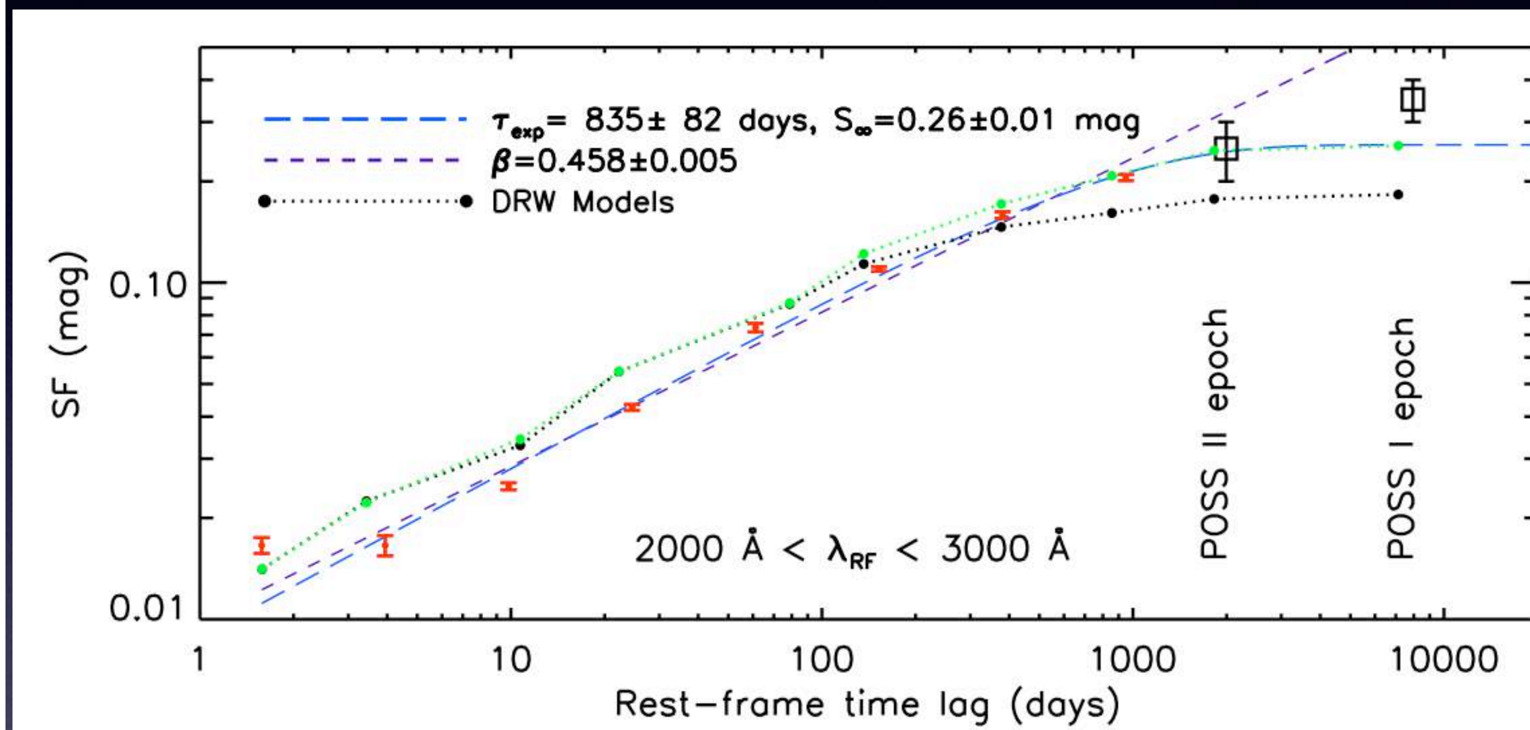


# Damped random walk

For irregularly sampled data, statistical samples are best analyzed using the (model-independent) structure function




$$SF(\Delta t) = SF_\infty (1 - e^{-|\Delta t|/\tau})^{1/2}$$




Observing baseline of 10 years (SDSS) is sufficient to constrain variability time scale for the majority of quasars

Variability rms decreases with wavelength and increase with time

McLeod et al. 2012