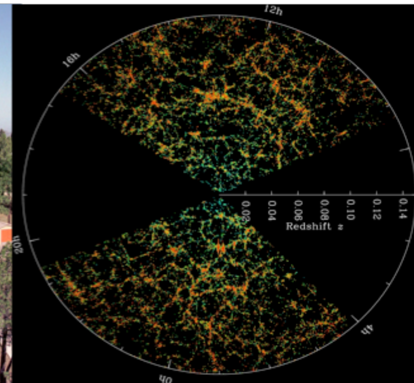
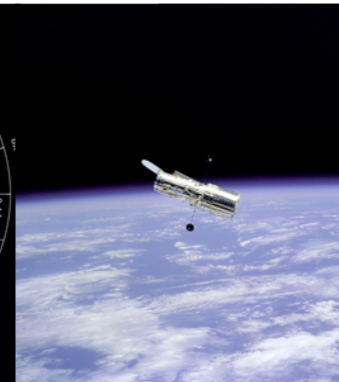
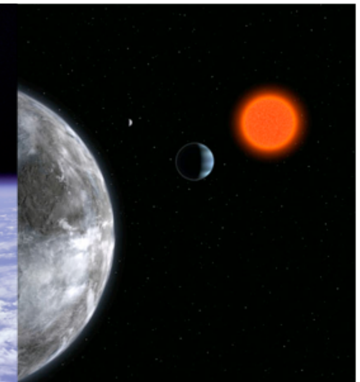
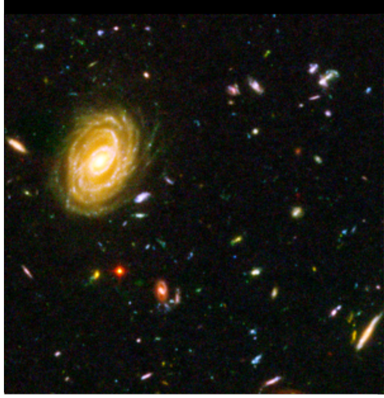
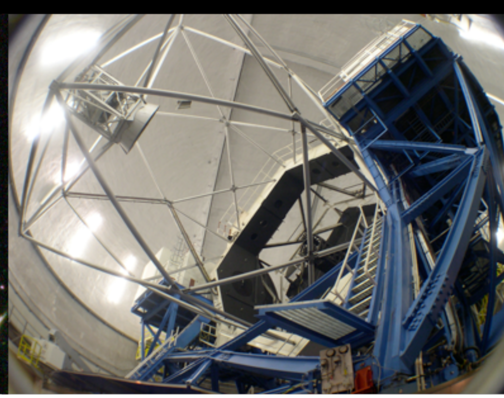




Surveys and the Time Domain

Andrew Connolly
University of Washington

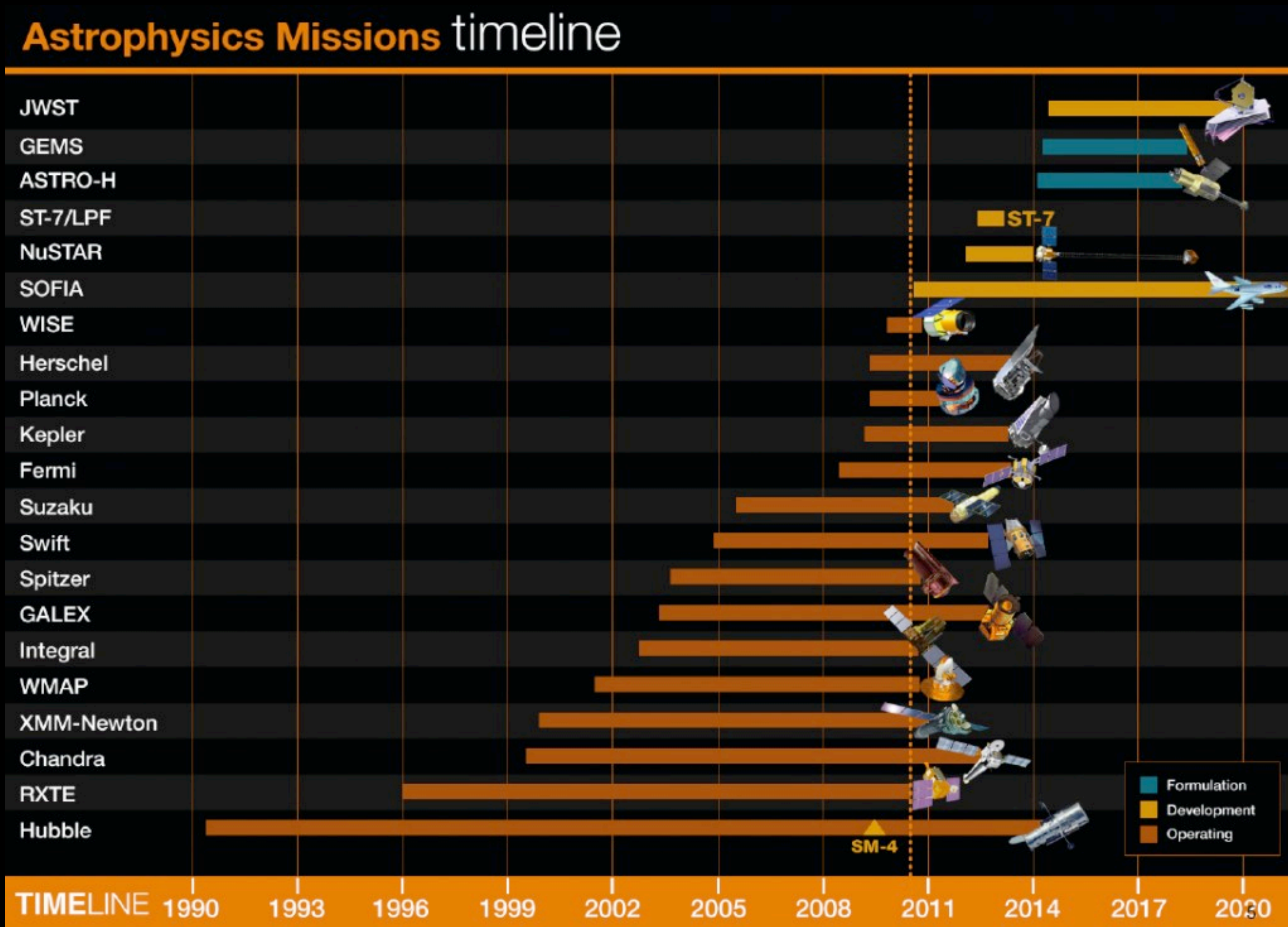
Structure of the lectures

1. LSST and the next generation of surveys
   – How will the LSST impact our view of the universe

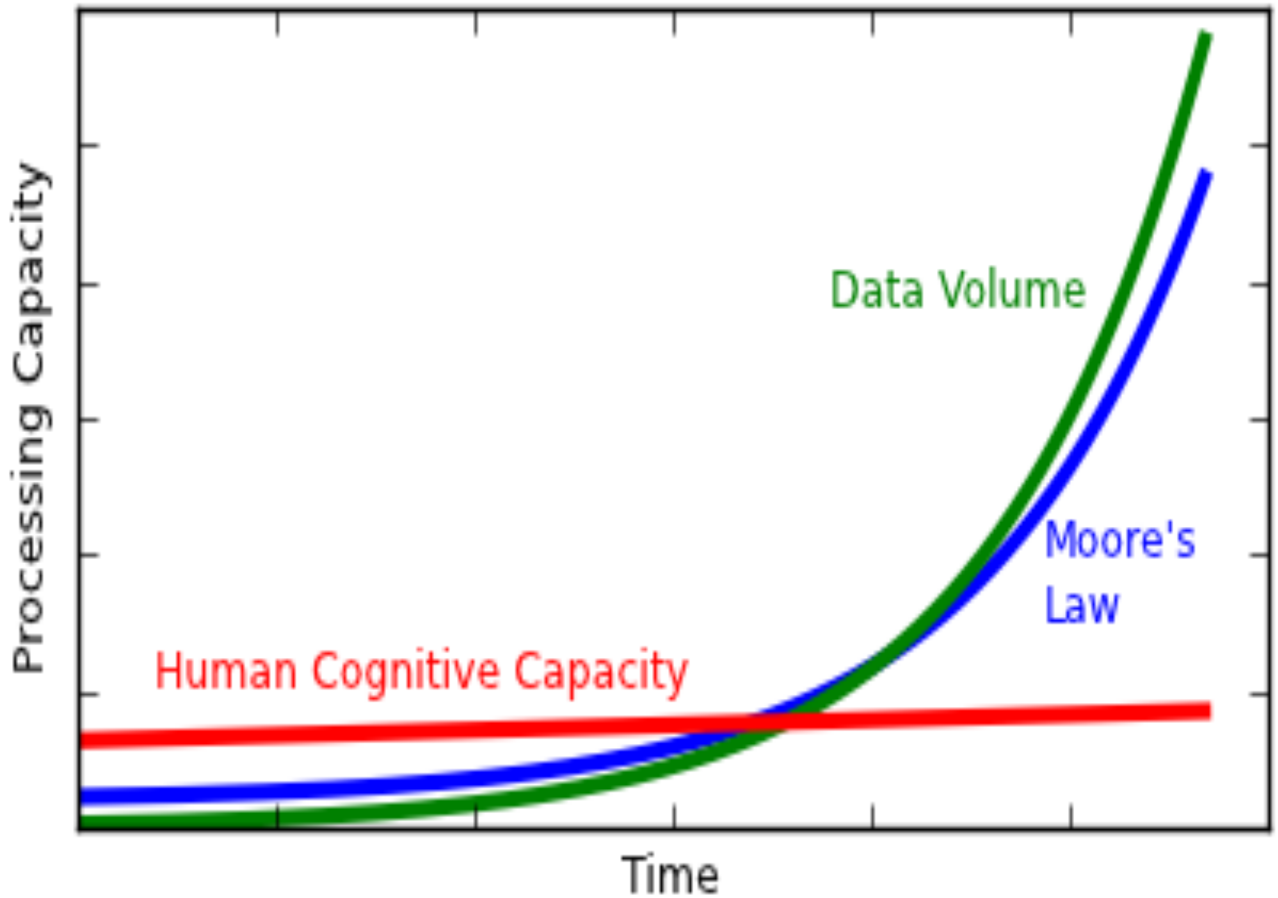
2. Working with high dimensional data
   – Techniques and tools for analyzing large complex data sets including an introduction to astroML

3. Analyzing the Time domain
   – What statistical approaches can we use to understand time domain data. Plus a shameless plug for the need for visualization

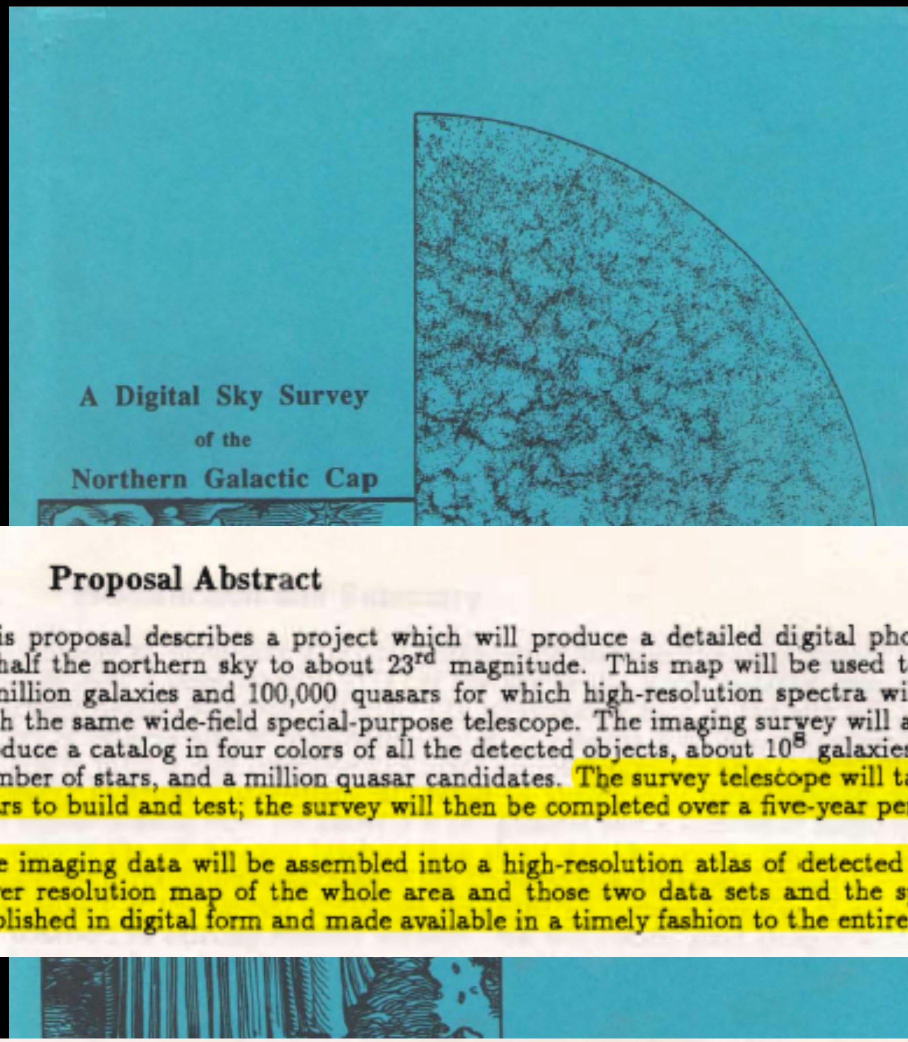
A decade of surveys



Much of the progress in astrophysics has been driven by surveys



Why do we do surveys….



i. Proposal Abstract

This proposal describes a project which will produce a detailed digital photometric map of half the northern sky to about 23rd magnitude. This map will be used to select about a million galaxies and 100,000 quasars for which high-resolution spectra will be obtained with the same wide-field special-purpose telescope. The imaging survey will also be used to produce a catalog in four colors of all the detected objects, about 10^8 galaxies and a similar number of stars, and a million quasar candidates. The survey telescope will take about four years to build and test; the survey will then be completed over a five-year period.

The imaging data will be assembled into a high-resolution atlas of detected objects and a lower resolution map of the whole area and those two data sets and the spectra will be published in digital form and made available in a timely fashion to the entire community.

The survey will cost about \$25M to complete, plus about 50 FTE-years of contributed faculty time. The funding and supporting institutions will commit \$12.7M to the project. We therefore seek outside funding for a total of \$12M, just under half the total cost for capital and operations.

Goals of the SDSS (1992)

- Large Scale Structure (5 pages)
  - Topology
  - Two point correlation function
- Clusters of galaxies (7 pages)
  - Topology
  - Cluster counts
- Galaxies (25 pages)
  - Luminosity functions, Low Surface Brightness Galaxies, Morphological properties
- QSOs (11 pages)
  - QSO selection, high redshift clustering,
- Stars (5 pages)
  - Halo stars, variables, proper motions, supernovae

SDSS Impact in terms of citation

- "Cosmological parameters from SDSS and WMAP"
  Tegmark et al 2004 1769 citations

- "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies"
  Eisenstein et al 2005 1414 citations

- "The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey"
  Tegmark et al 2004 912 citations

- "The Origin of the Mass-Metallicity Relation: Insights from 53,000 Star-forming Galaxies in the Sloan Digital Sky Survey"
  Tremonti et al 2004 856 citations

- "Composite Quasar Spectra from the Sloan Digital Sky Survey"
  Vanden Berk et al 2001 619 citations

What drives survey design

- Statistical uncertainties
  - Shot noise: drives large n
  - Sample variance: drives large V

$$\frac{\Delta P(k)}{P(k)} \propto \frac{1}{\sqrt{V}}\left(1+\frac{1}{nP(k)}\right)$$

- Systematic errors
  - Absolute calibration: drives multiple visits (photometricity)
  - Relative calibration: drives offset pointings (reproducability)
- Analysis efficiency
  - Large n: drives compute resources
  - Complexity is the real challenge
- Theoretical considerations
  - Simulations exceed the volumes of galaxy surveys and have all of the same computational challenges

What drives survey design

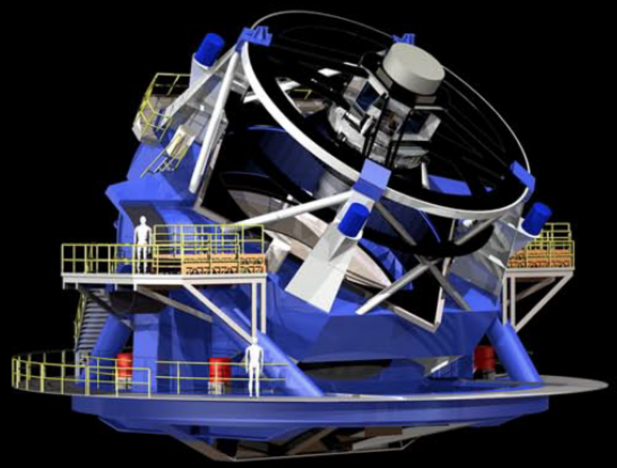
- Statistical uncertainties
  - Shot noise: drives large n
  - Sample variance: drives large V

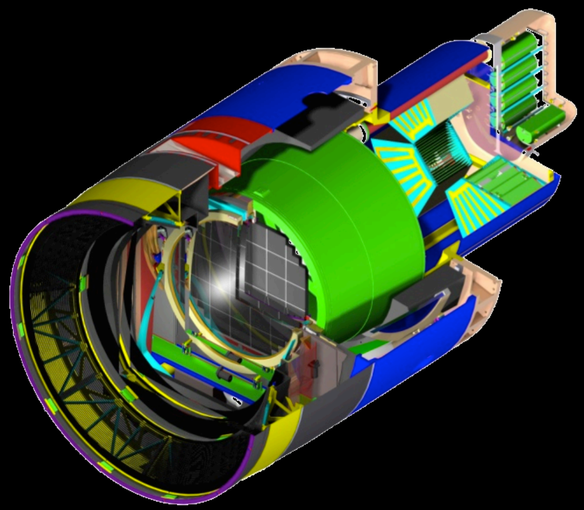
$$\frac{\Delta P(k)}{P(k)} \propto \frac{1}{\sqrt{V}}\left(1+\frac{1}{nP(k)}\right) \quad \therefore \text{ want } nP(k) > 1$$

- Systematic errors
  - Absolute calibration: drives offset pointings (reproducability)
  - Relative calibration: drives multiple visits (photometricity)
- Analysis efficiency
  - Large n: drives compute resources
  - Complexity is the real challenge
- Theoretical considerations
  - Simulations exceed the volumes of galaxy surveys and have all of the same computational challenges

Large Synoptic Survey Telescope

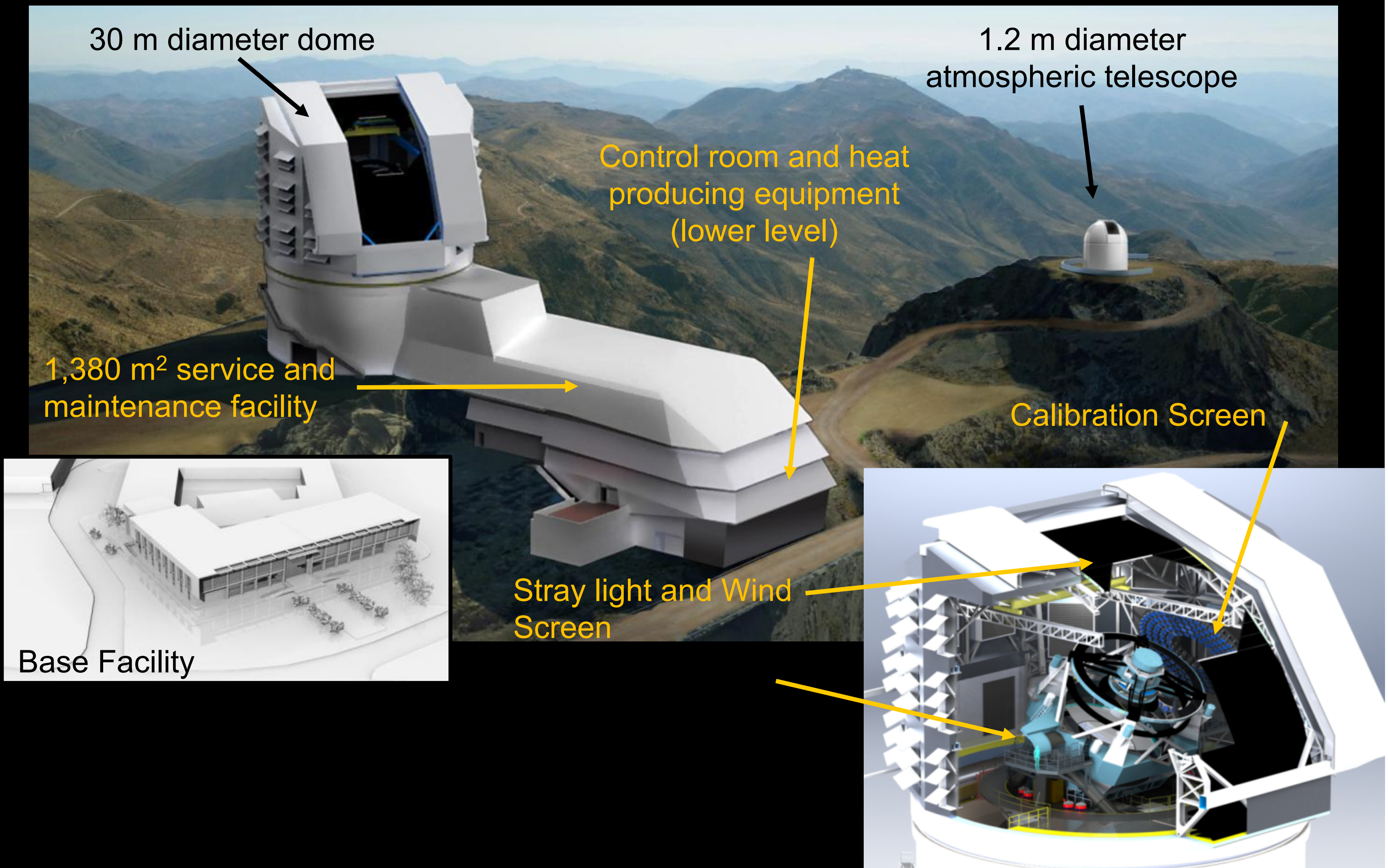
- Wide field and deep
  - 27000 sq deg (wide)
  - 100 - 200 sq deg (deep)
  - 10 years
- Broad range of science
  - Dark energy
  - Galactic structure
  - Census of the Solar system
  - Transient universe
- 3.2 Gpixel camera
  - 9.6 sq degree FOV
  - ugrizy filters



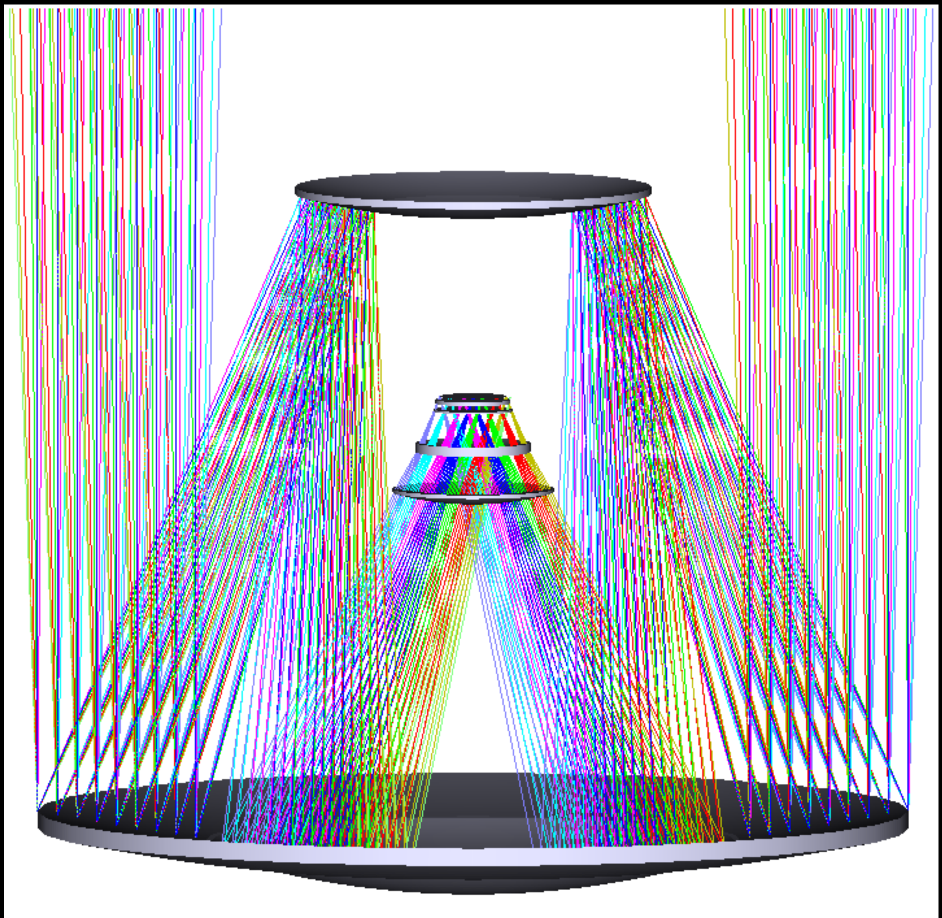


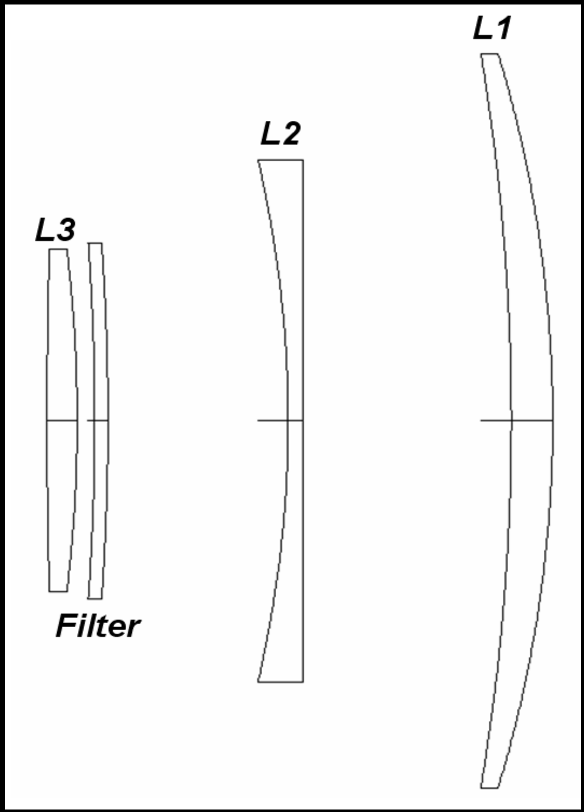
The LSST Site and Base Facilities in Chile

8.4m survey telescope and 1.2m atmospheric telescope



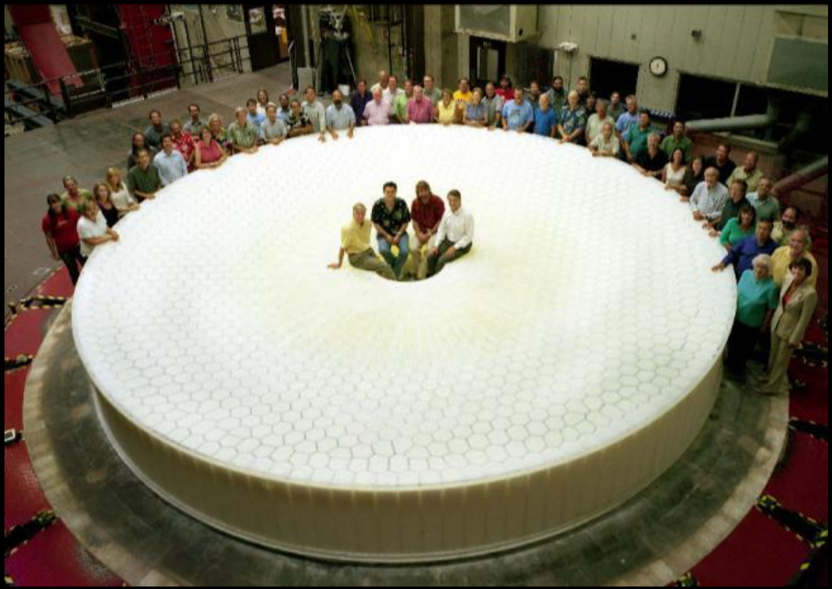
Optical design of the LSST





3 Mirror Modified Paul-Baker Design for achieving a low focal ratio over such a large field. The camera optics correct for chromatic aberration.

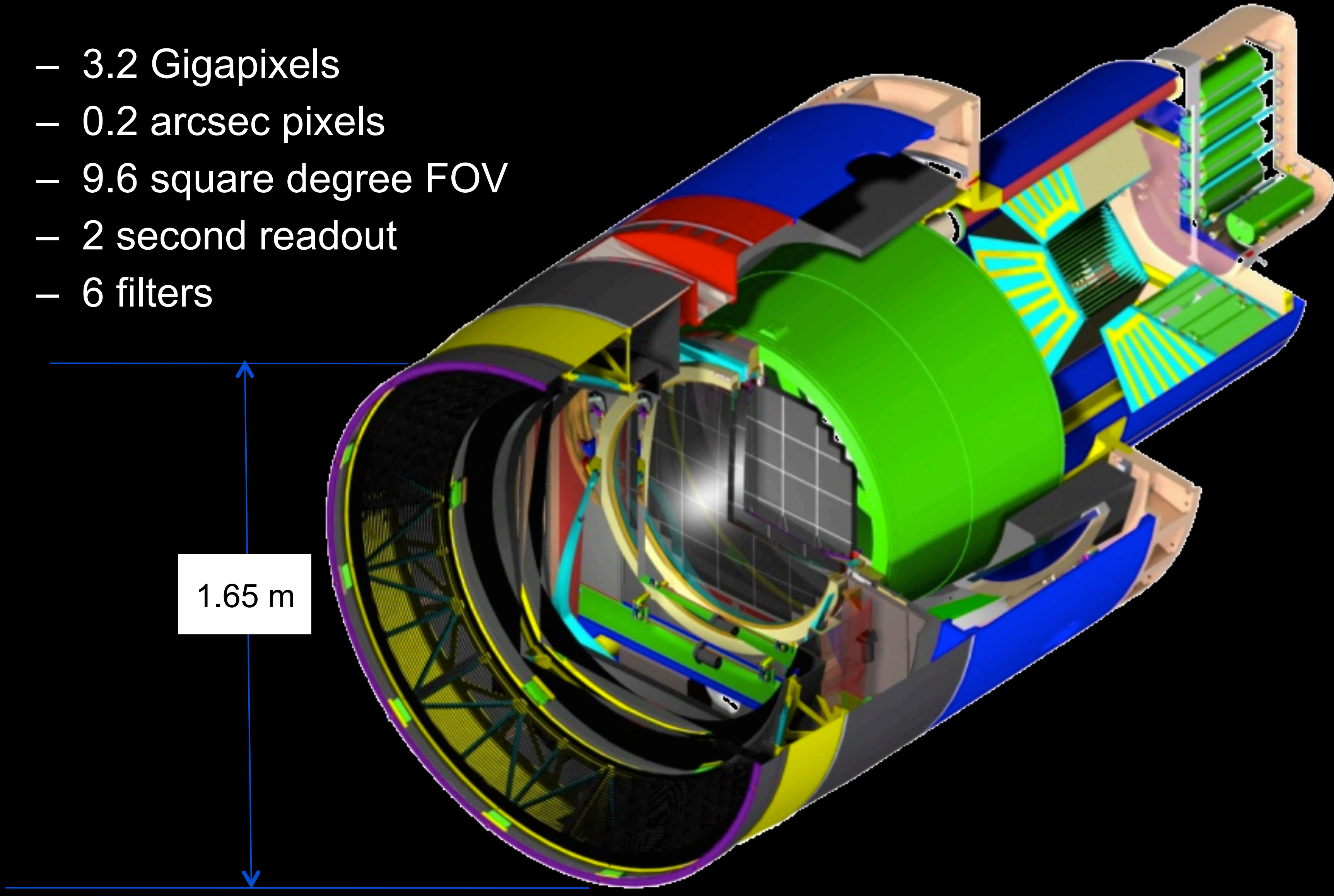
Primary/Tertiary cast from a single borosilicate blank.



- Primary-Tertiary was cast in the spring of 2008.
- Secondary fabricated by Corning in 2009.

A 3.2 Gigapixel camera

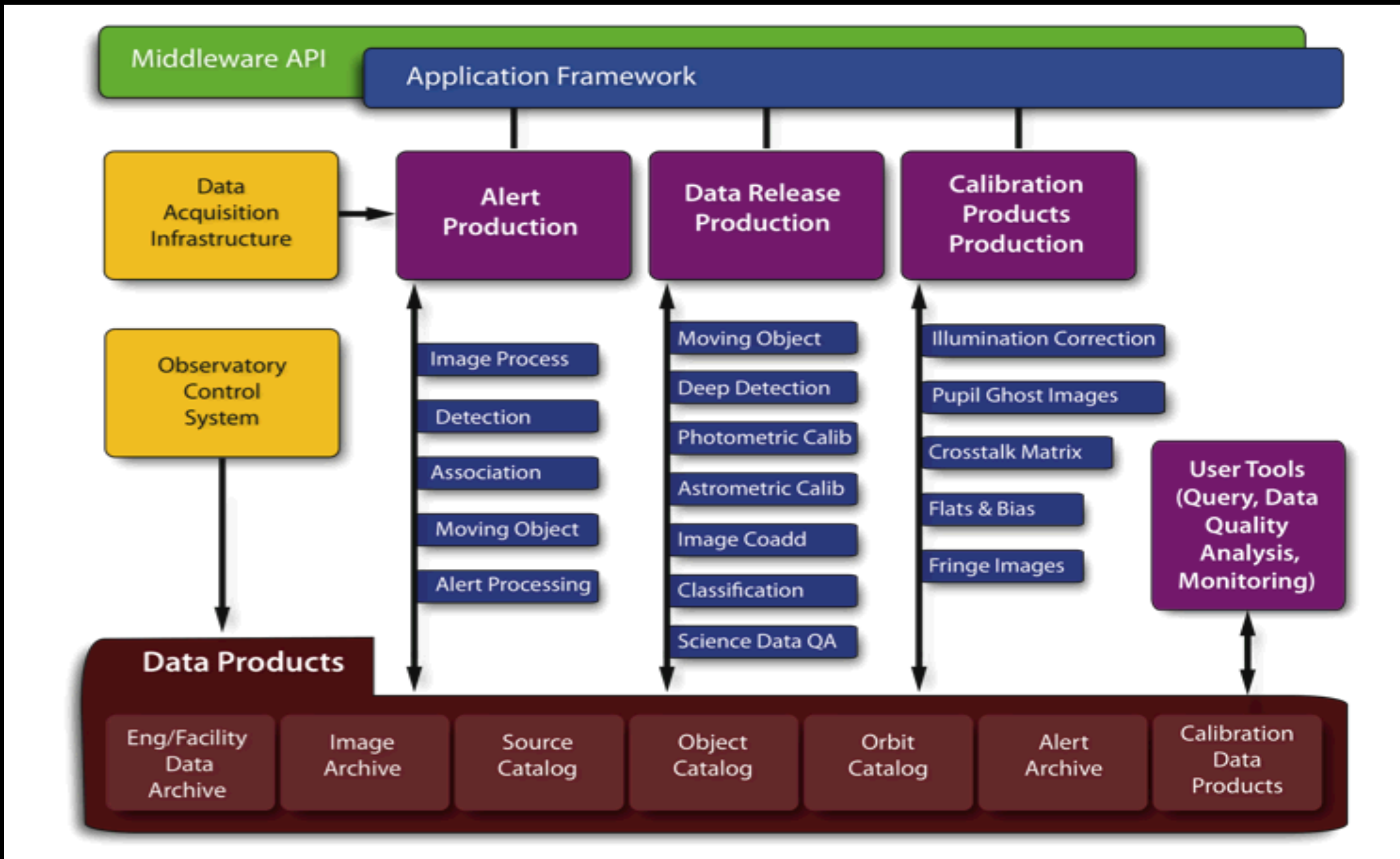
- 3.2 Gigapixels
- 0.2 arcsec pixels
- 9.6 square degree FOV
- 2 second readout
- 6 filters



Processing the data flow from the LSST

- Each “Visit” comprises a pair of back-to-back exposures
  - 2x15 sec exposure; duration = 34 seconds with readout
- The data volume associated with this cadence is unprecedented
  - one 6-gigabyte image every 17 seconds
  - 15 terabytes of raw scientific image data / night
  - 100-petabyte final image data archive
  - 20-petabyte final database catalog
  - 2 million real time events per night every night for 10 years

LSST data processing pipelines are designed, prototyped and tested in data challenges



The Science Book

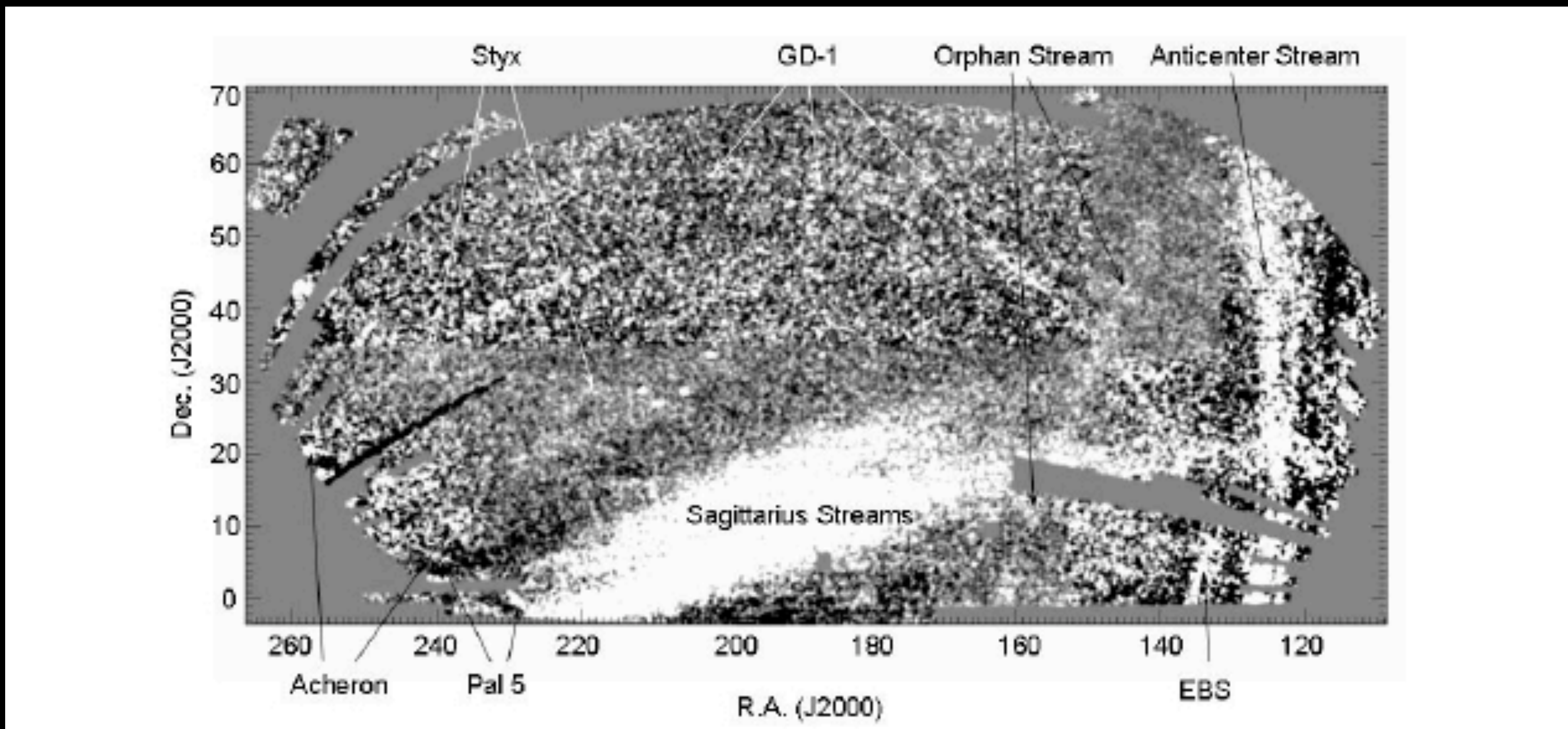
- **Contents:**
  - Introduction
  - LSST System Design
  - System Performance
  - Education and Public Outreach
  - The Solar System
  - Stellar Populations
  - Milky Way and Local Volume Structure
  - The Transient and Variable Universe
  - Galaxies
  - Active Galactic Nuclei
  - Supernovae
  - Strong Lenses
  - Large-Scale Structure
  - Weak Lensing
  - Cosmological Physics

http://www.lsst.org/lsst/scibook

Stellar Populations

- LSST will individually resolve and detect billions of stars in the Milky Way and neighboring Local Group
- Studies of field stars and stellar associations can address a multitude of astrophysical issues associated with star formation, the assembly of the MW galaxy, and the origin of the chemical elements.
- Key techniques for these investigations include:
  - Construction of color magnitude diagrams
  - Trigonometric parallaxes to establish absolute distances
  - Stellar proper motions to separate associations from background stars and from one another
  - Using RR Lyrae and other variables as “standard candles”
  - Using eclipsing binaries to measure stellar masses

Cuts in the CMD and in Proper Motion Allow the Detection of Faint Tidal Streams

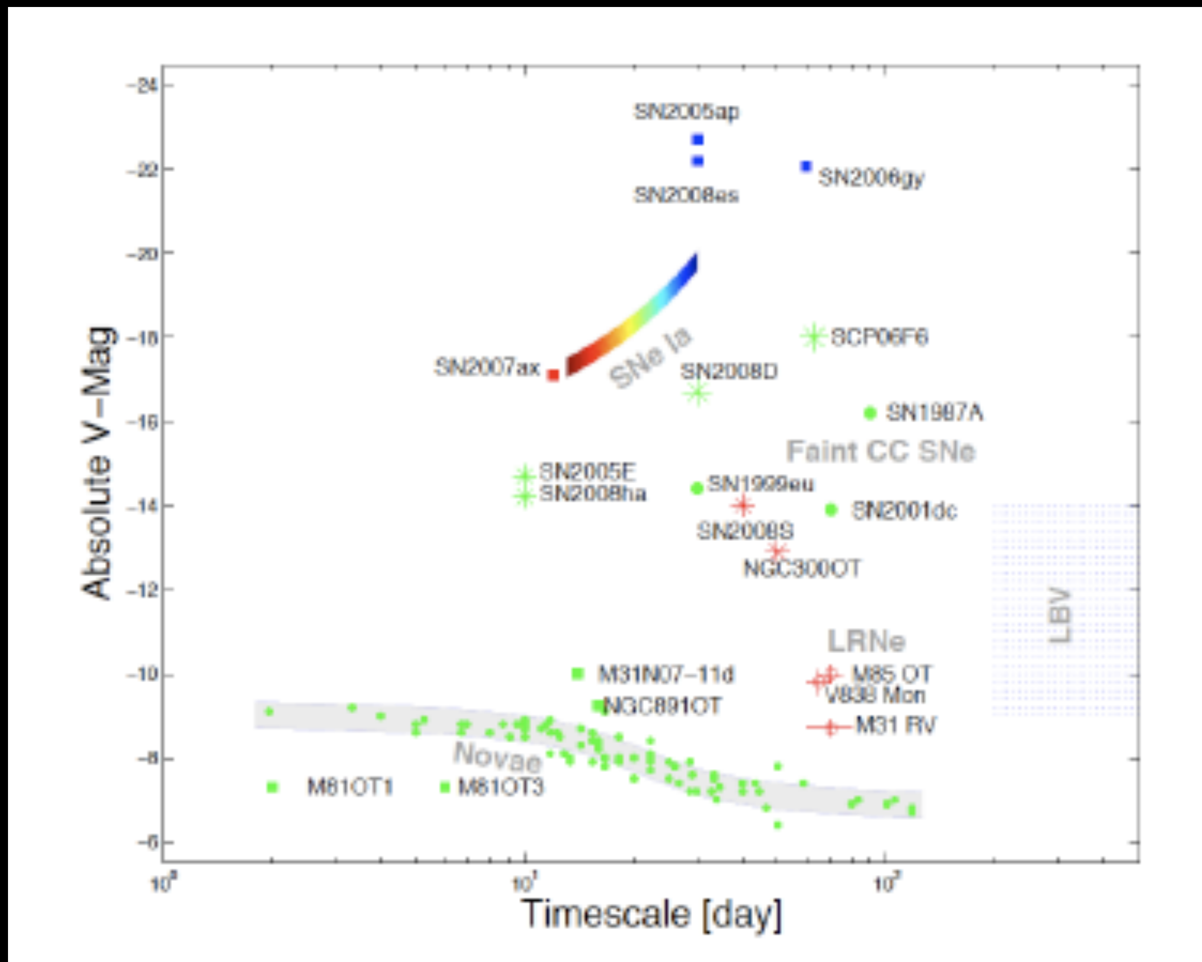


Transients and Variable Stars

- LSST's unique time sampling allows the detection of stellar variability on timescales from seconds to years.

- A wide range of phenomena can be studied with such a rich dataset:
  - Explosive events (supernovae, novae, gamma-ray bursts)
  - Periodic variability associated with binarity
  - Intrinsic stellar variables like Cepheids, RR Lyrae, Miras, which are important for distance measurements
  - Geometrical effects such as gravitational microlensing
  - Dimming of stars as they are occulted by transiting planets

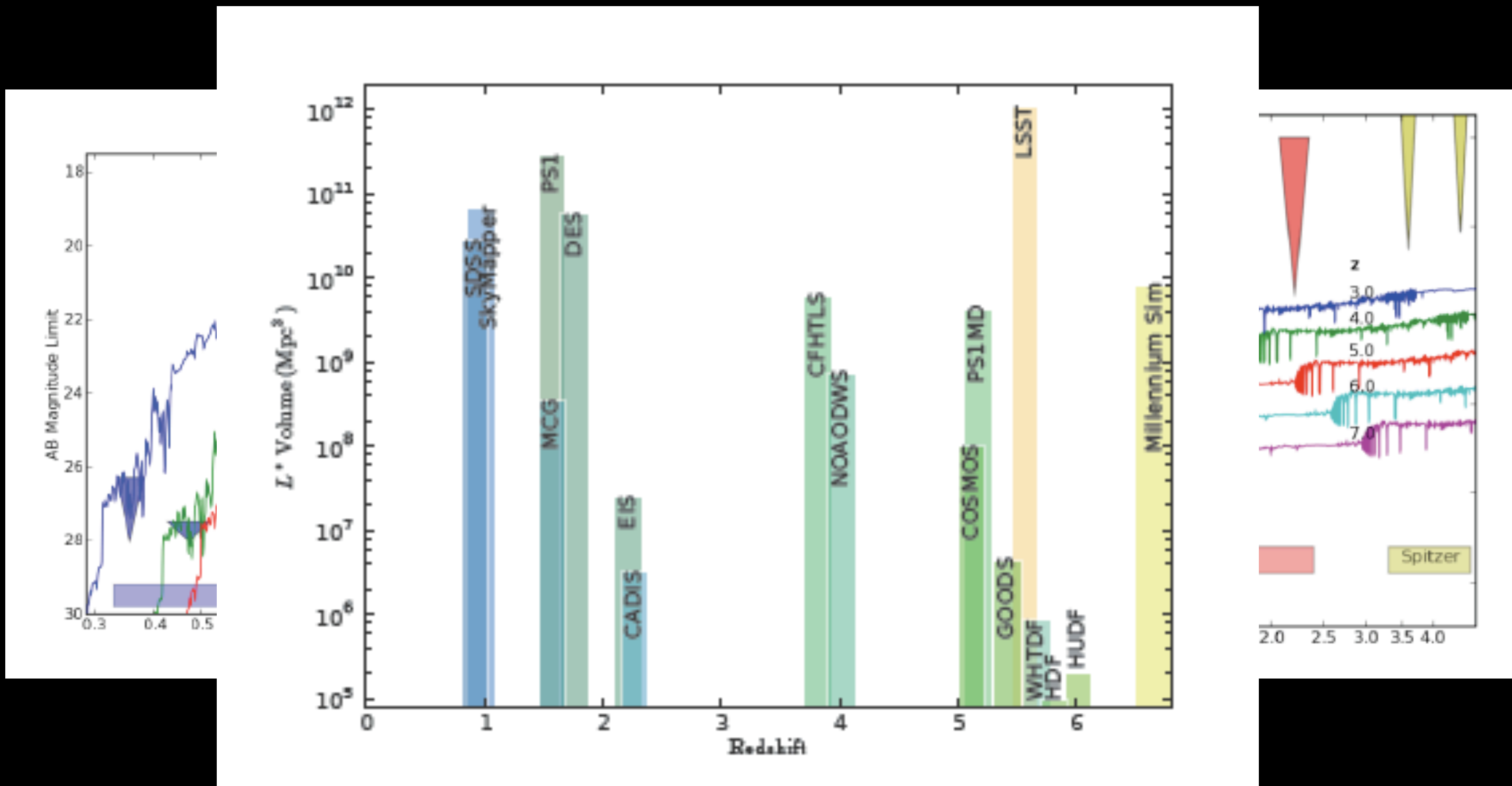
Our Knowledge of Explosive Transients is Limited
Much of the Phase Space Has Been Unexplored



Galaxies

- **LSST will be a unique tool for studies of galaxy formation and galaxy properties.**
- **The database will include photometry for 10^{10} galaxies from the Local Group to $z > 6$.**
- **6-band photometry for $4x10^9$ galaxies.**
- **Key diagnostic tools will include:**
  - Luminosity functions
  - Color-luminosity relations
  - Size-luminosity relations
  - Quantitative morphological classifications
  - Dependence on environment

Galaxies: Expected Sensitivity Leads to Near Complete Samples Out to High Redshifts



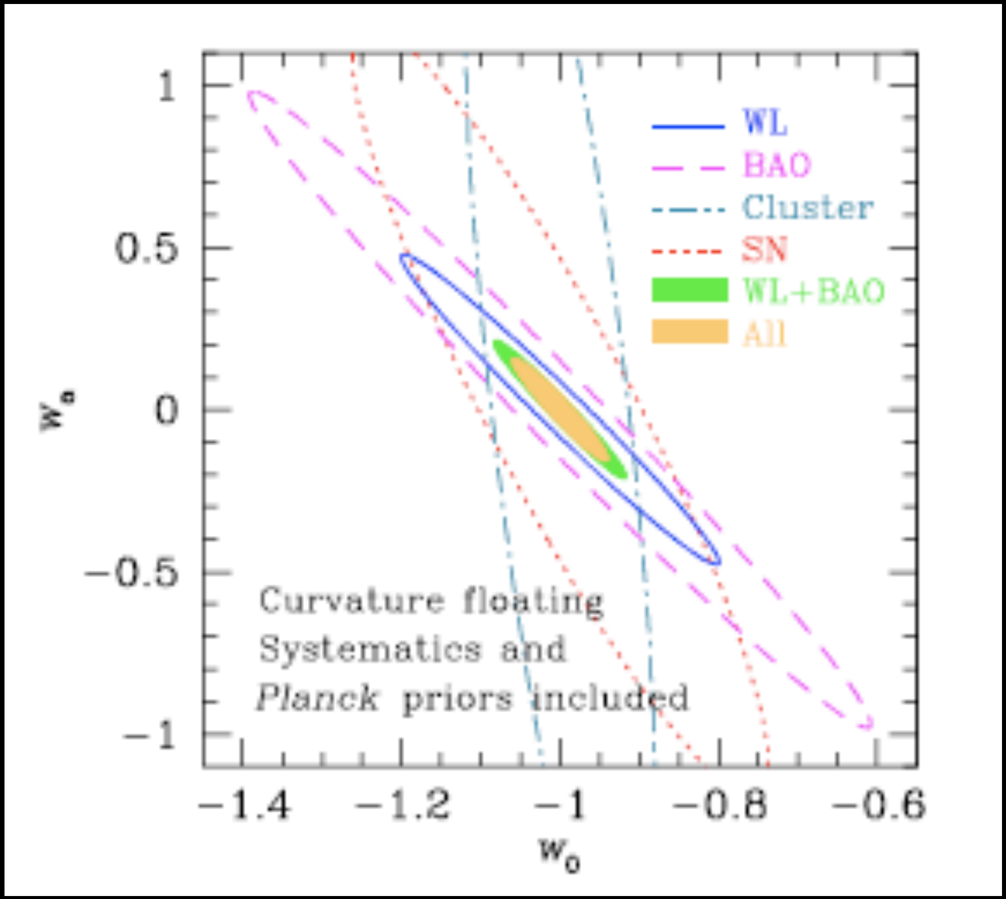
Precision Cosmology: Constraints on Dark Energy

- LSST will probe the nature of Dark Energy via a distinct set of complementary probes:
  - SNe Ia's as "standard candles"
  - Baryon acoustic oscillations as a "standard rulers"
  - Studies of growth of structure via weak gravitational lensing
  - Studies of growth of structure via clusters of galaxies

- In conjunction with one another, this rich spectrum of tests is crucial for reduction of systematics and dependence on nuisance parameters.

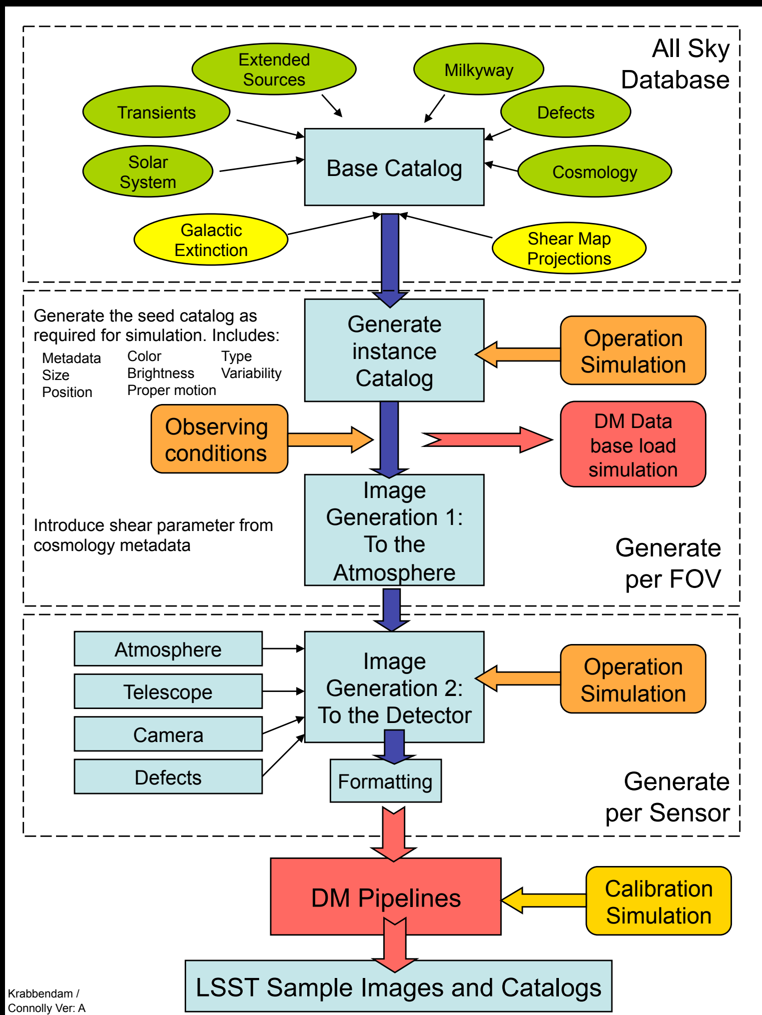
- These tests also provide interesting constraints on other topics in fundamental physics: the nature of inflation, modifications to GR, the masses of neutrinos.

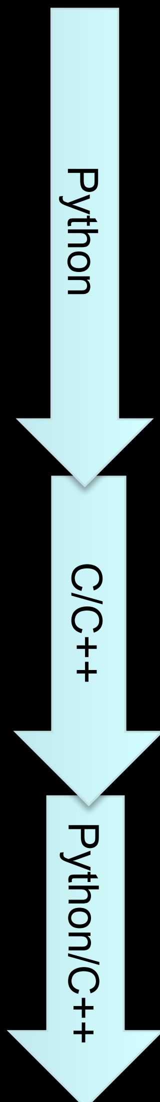
Separate and Joint Constraints on the Dark Energy Equation of State



$w(a) = w_0 + w_a(1+a)$

Observing the LSST before 2020





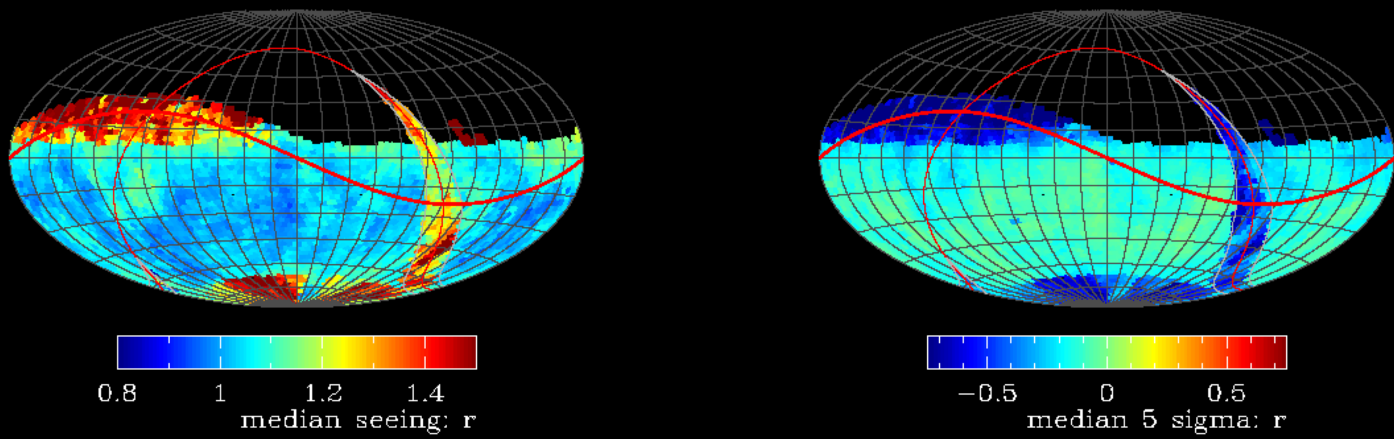
16TB of catalog data on a parallel database system with Python interfaces. Defined **schema** for sources, extinction and lensing

Flexible python framework integrated through sqlAlchemy with databases and opSim. Extensible to new catalogs and capabilities (calibration and lensing)

Modular simulation codes separating atmosphere, telescope, camera, electronics, sky background.

Building and developing within a single framework enables consistent checks between multiple probes

Operations Simulator: characterizing the cadence of the LSST

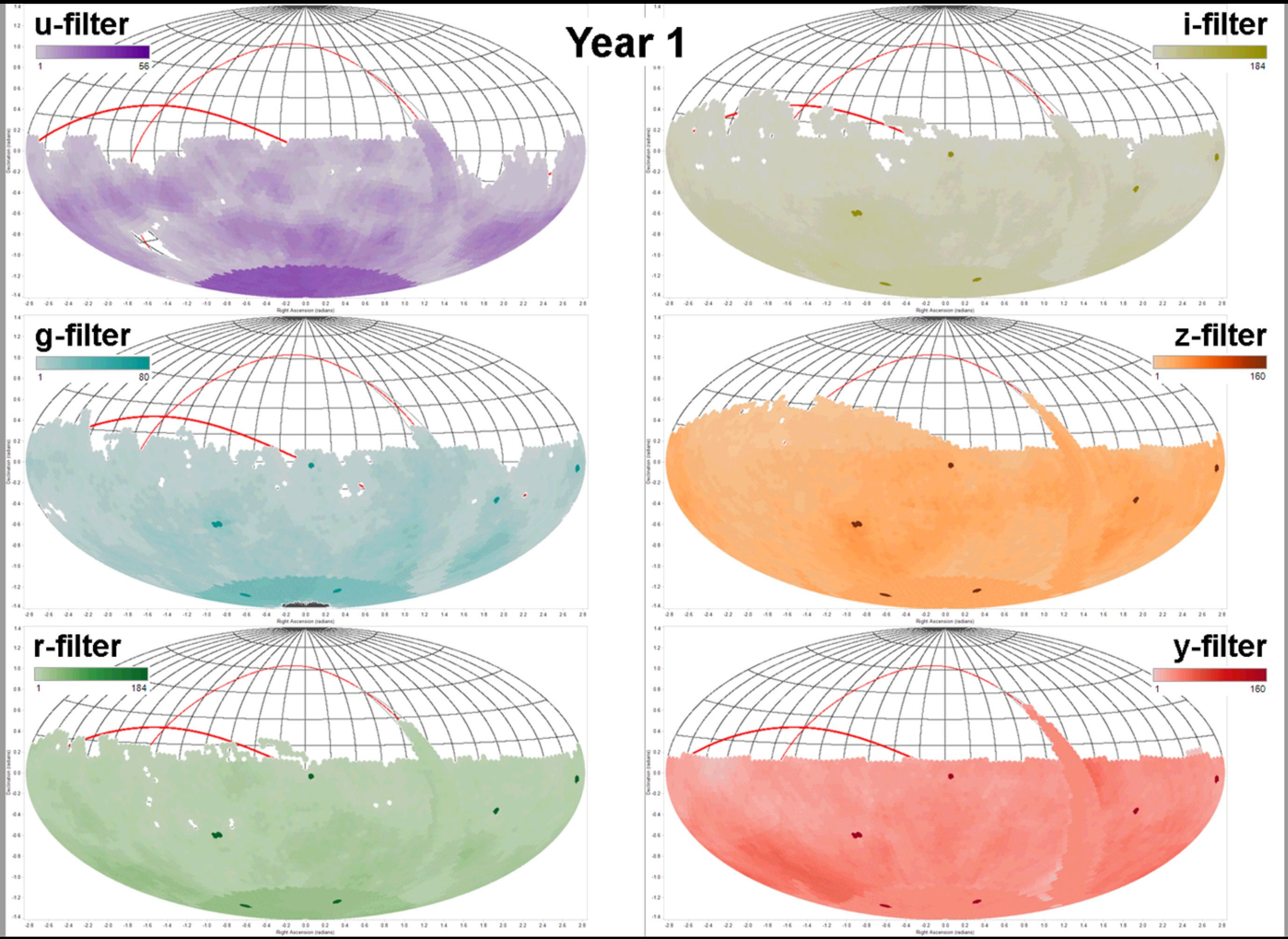


System inputs: FOV, QE, sky brightness, weather readout, slew and settle times science requirements on cadence and depth, seeing distributions

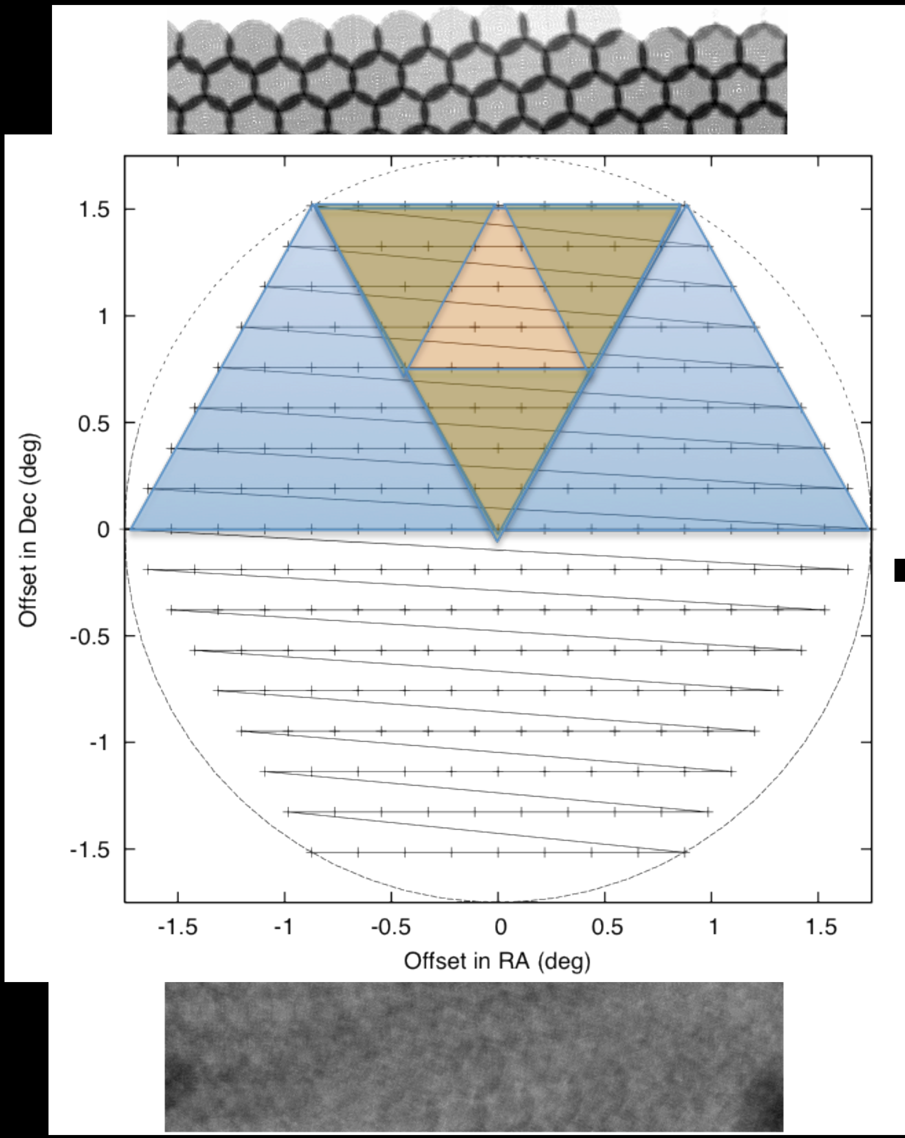
System outputs: sequences of observations and summaries of the properties of a 10 year survey

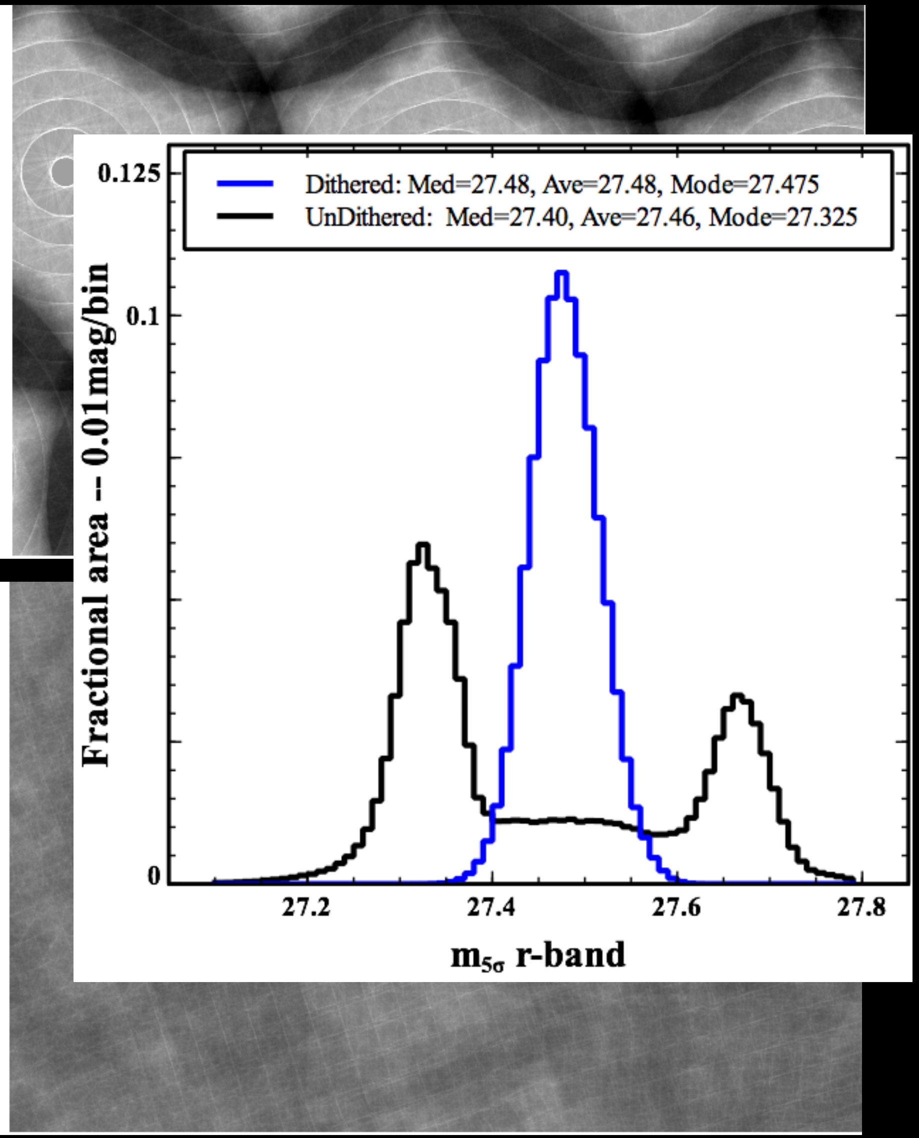
Many Monte-Carlo realizations of an LSST survey can be generated and evaluated for how much the science requirements are impacted by the system design

A ten year observing plan



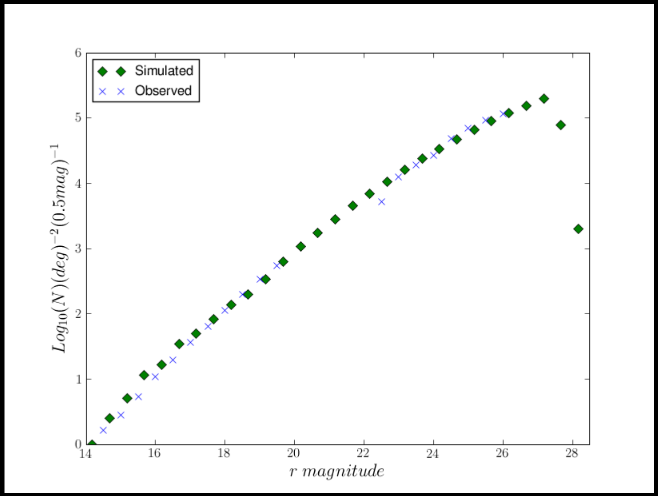
Dithering the pointings

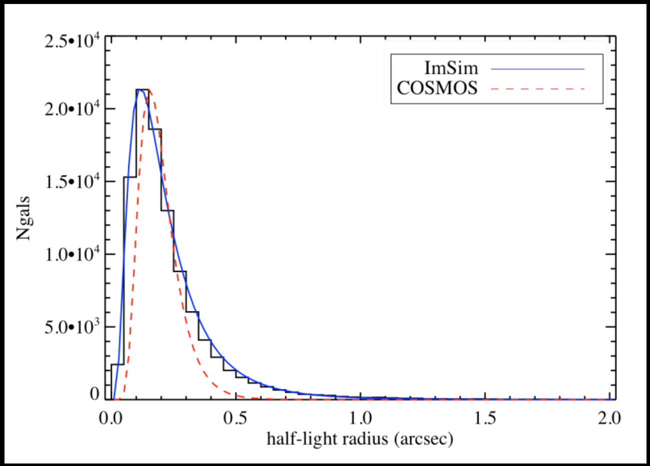




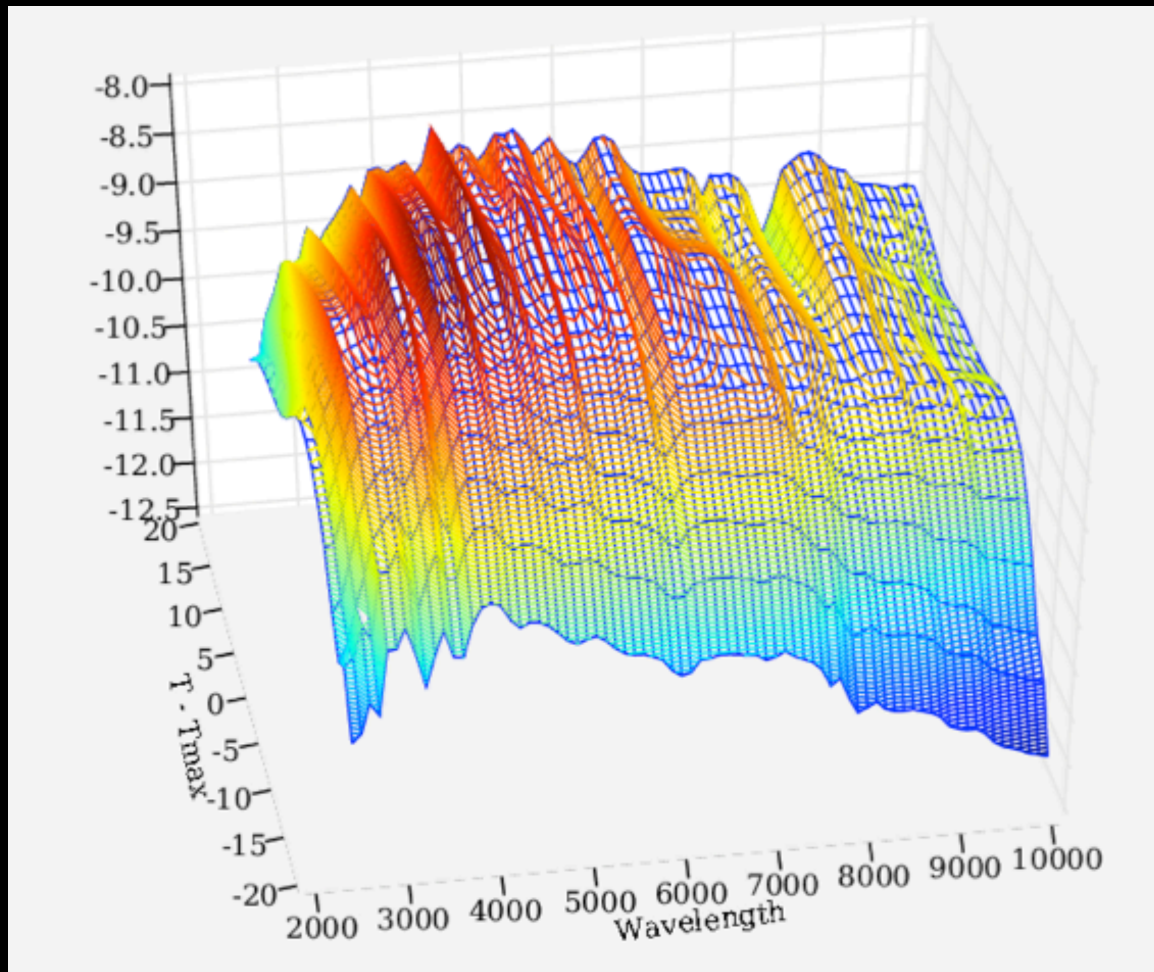
Extensible base catalogs

- Extragalactic (de Lucia et al 2006)
  - Cosmological models cloned to reproduce the observed size, redshift and magnitude distributions (r<28)
  - Sersic morphologies (with AGNs) based on local distributions of sources
  - Prototype lensing signatures
- Galactic (Juric et al 2008)
  - F-T main sequence stars
  - Giants and white dwarfs
  - Flexible variability model for stars (including cepheids, flares, micro-lensing)
  - Proper motion, parallax, 3D extinction
- Solar System sources (Grav et al 2007)
  - 11 million asteroids (MB, Trojans, TNOs, KBOs)
  - Predicted orbital positions (<5 mas), velocities and magnitudes recovered in 10s



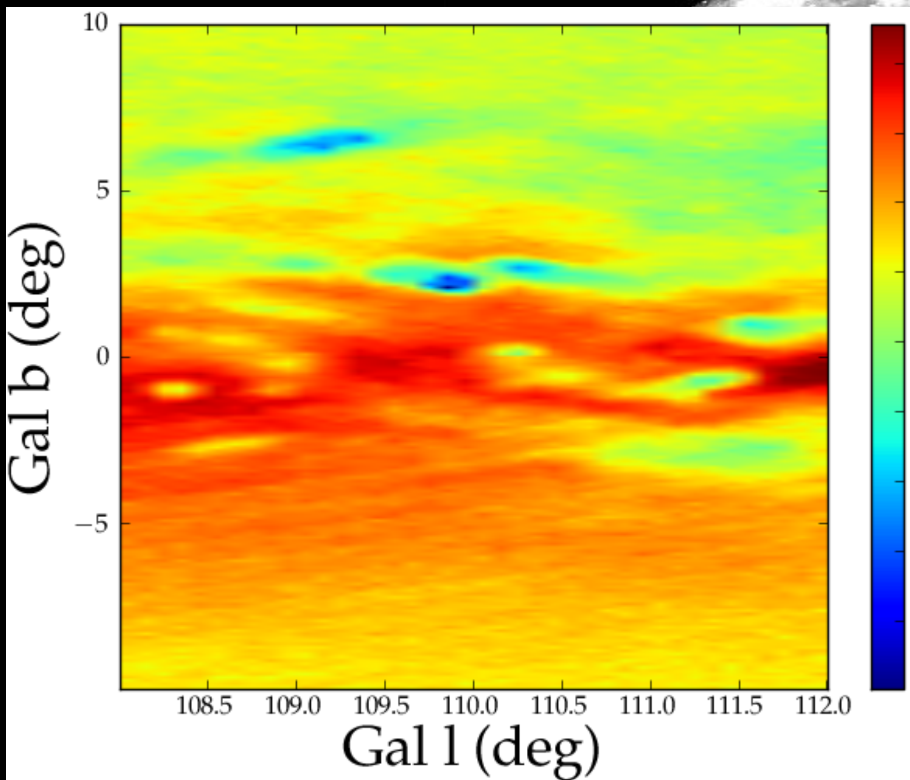


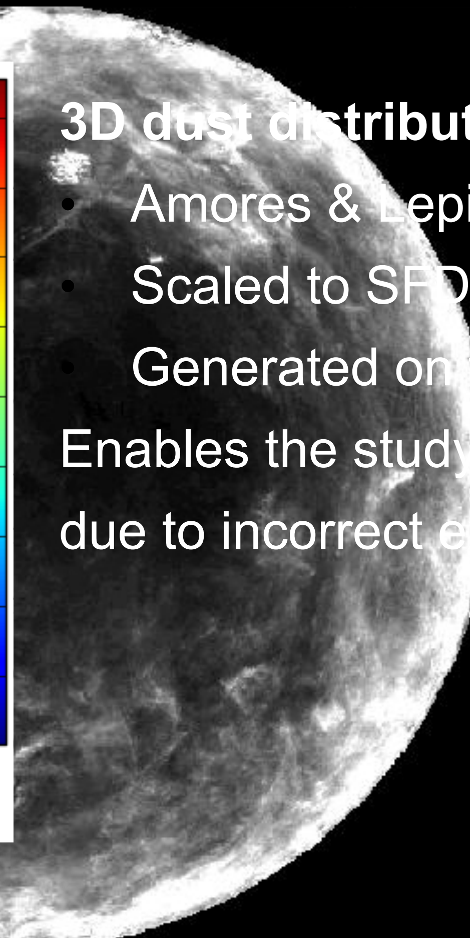
Variability in the universe



The time dependent evolution of the spectral energy distribution of a Type Ia supernova as modeled within the simulation database

Modeling the distribution of dust





3D dust distributions

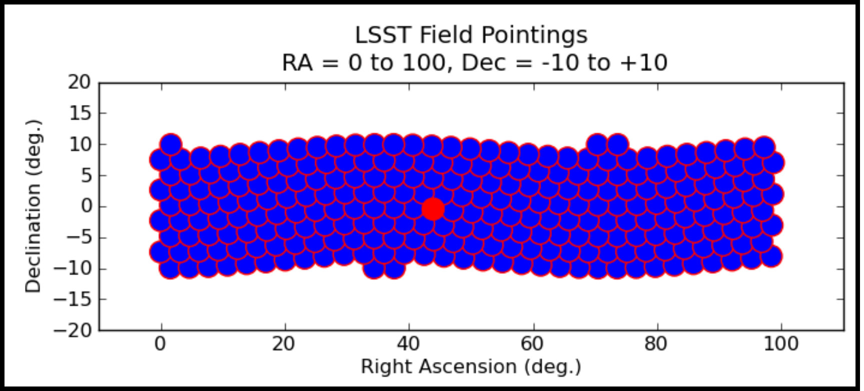
- Amores & Lepine (2005)
- Scaled to SFD at 100 kpc
- Generated on the fly

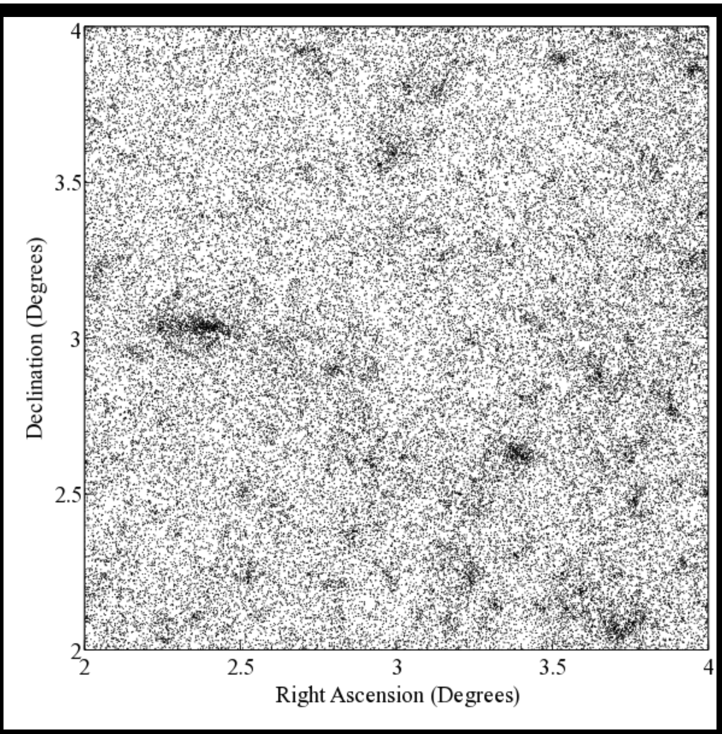
Enables the study of induced clusteri
due to incorrect extinction correction

Distance is 1 kpc, compare to SFD98 maps

Flexible source generation system

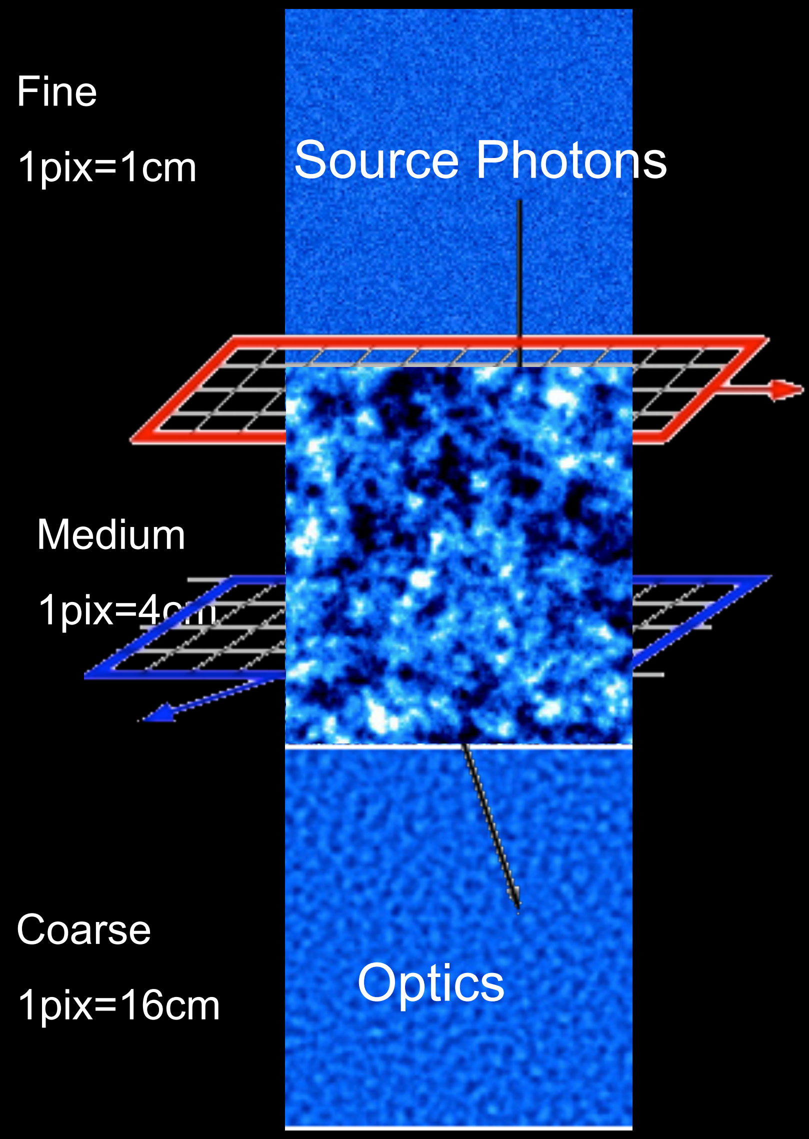
- **Framework generation**
  - Native python queries interact with SQL databases through a database independent layer
  - Fully integrated with OpSim pointings (including dithering and observing conditions)
- **Catalog generation**
  - Framework for specifying format of source catalogs (e.g. input to image generation, reference catalogs, calibration catalogs)
  - Customizable inputs and outputs (specified in a configuration file)
  - Extensible functionality within the Python framework



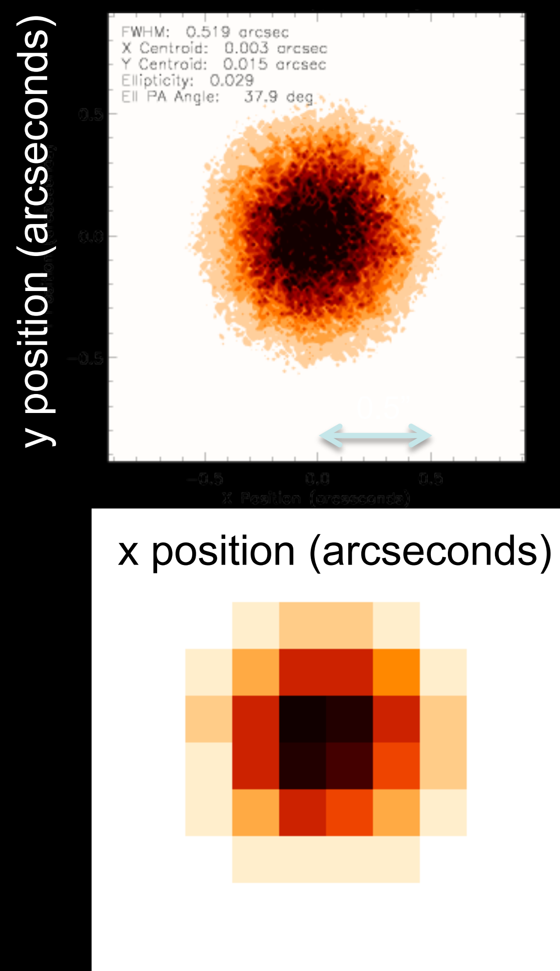


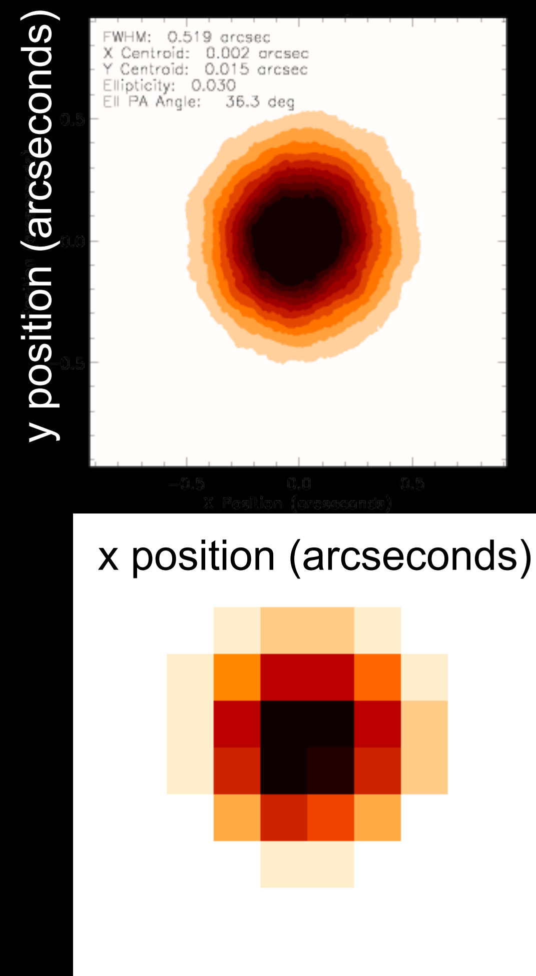
From points to pixels (and back)

- Generating images
  - From a parameterized view above the atmosphere to the distortions from the atmosphere, telescope and camera
- Turbulent atmospheres
  - Modeled as a series of frozen screens moving at 20m/s, refracting
    light based on the wavelength, cloud
    scattering and atmosphere.
  - Fidelity is optimized for the volume
    of data required for data challenges

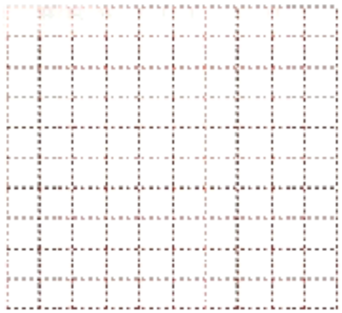


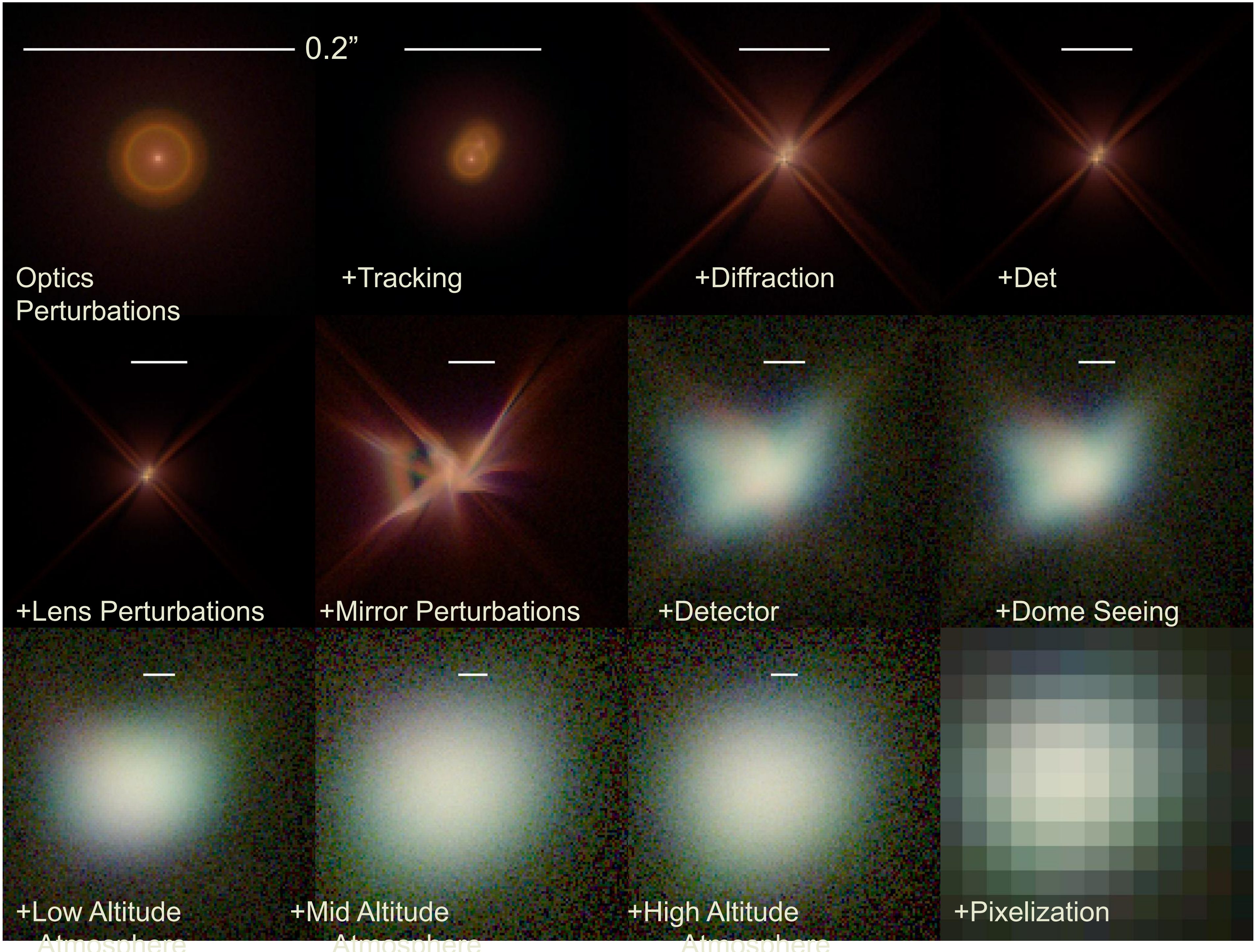
The speed and fidelity trade off





Following the photon flow…





0.2"
Optics Perturbations
+Tracking
+Diffraction
+Det
+Lens Perturbations
+Mirror Perturbations
+Detector
+Dome Seeing
+Low Altitude Atmosphere
+Mid Altitude Atmosphere
+High Altitude Atmosphere
+Pixelization

3 gigapixels

10 sq. degrees

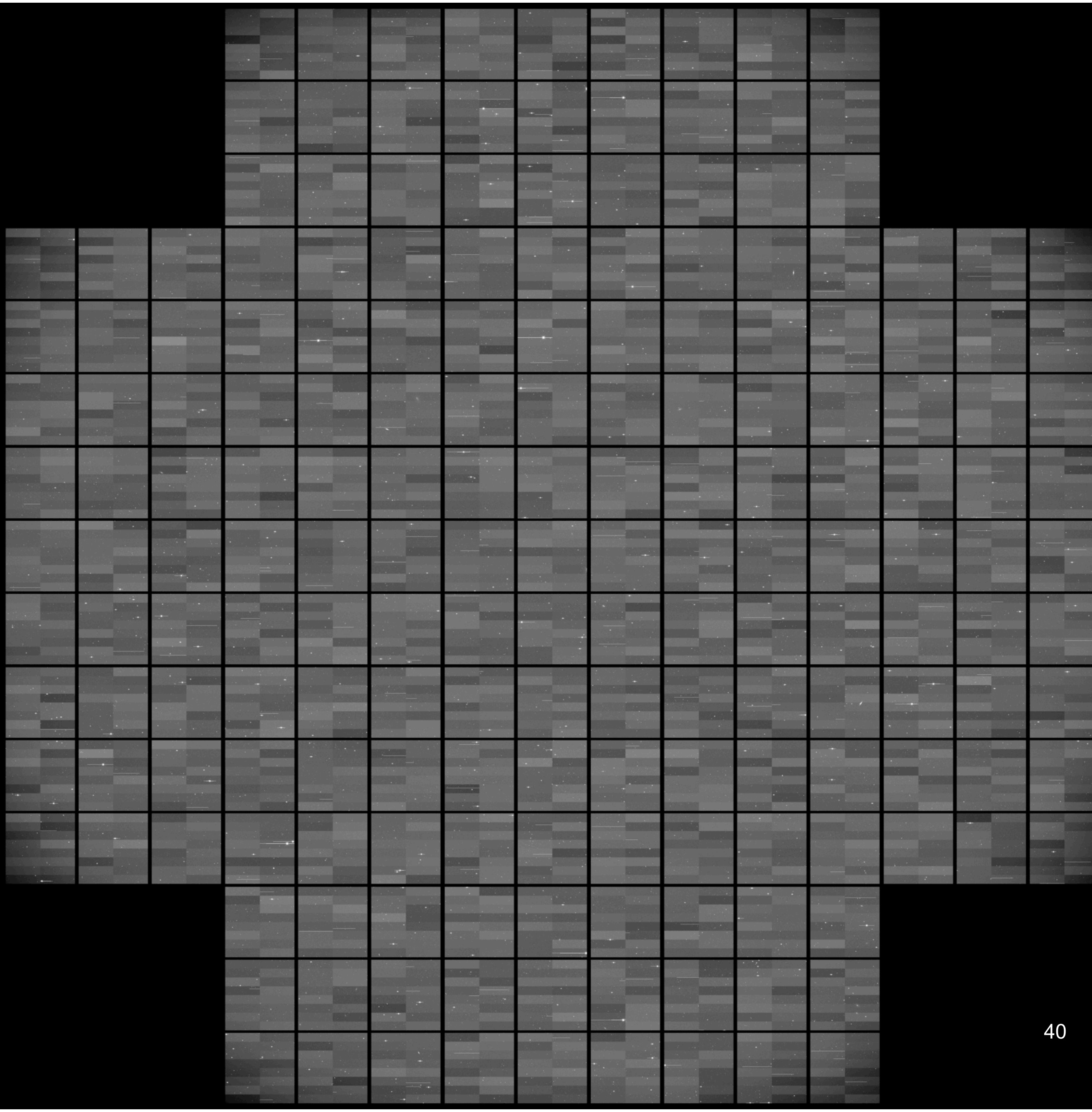
1 trillion photons

11 Gbytes

1000 CPU hours

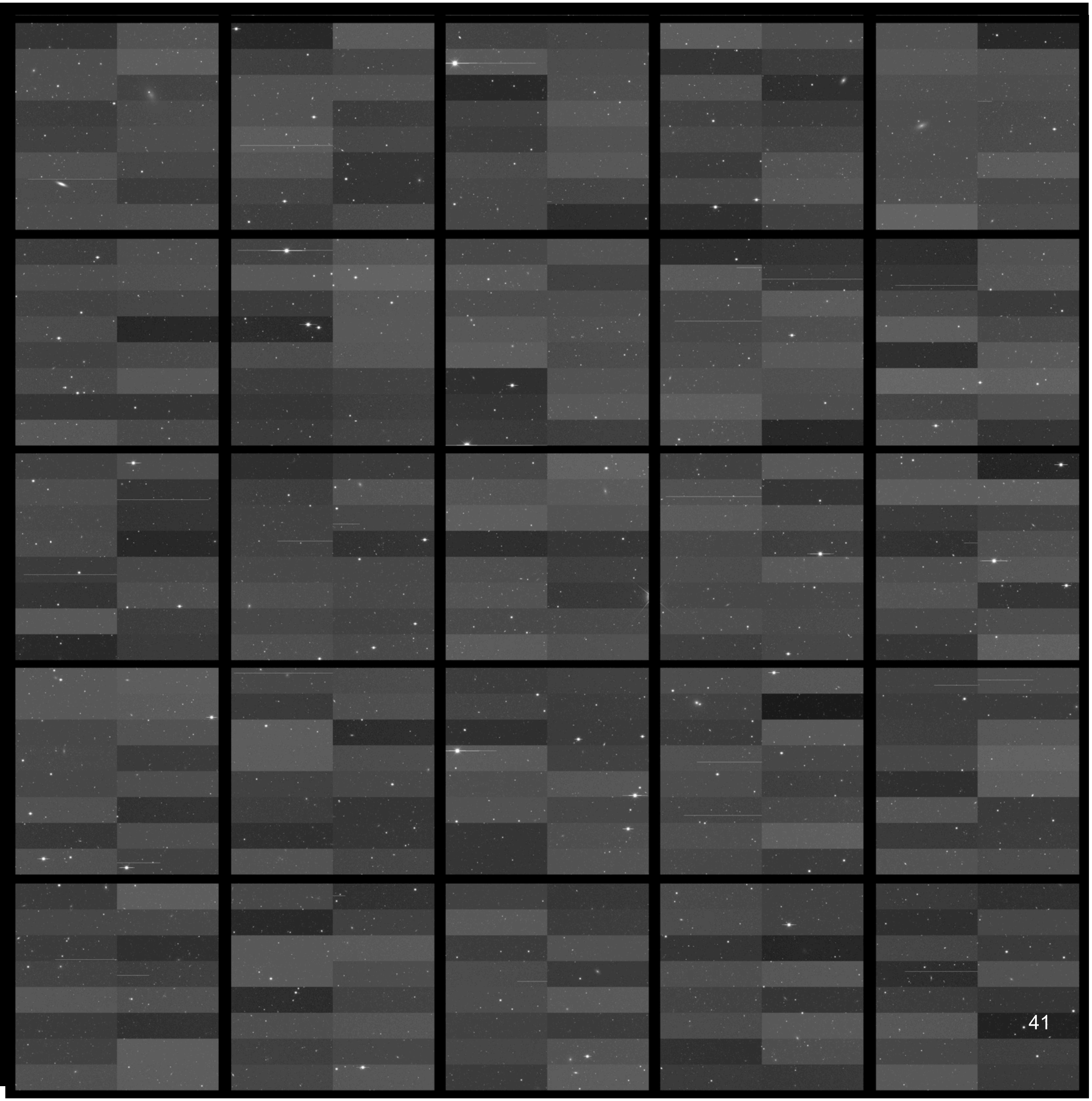
10 million objects

Now routine



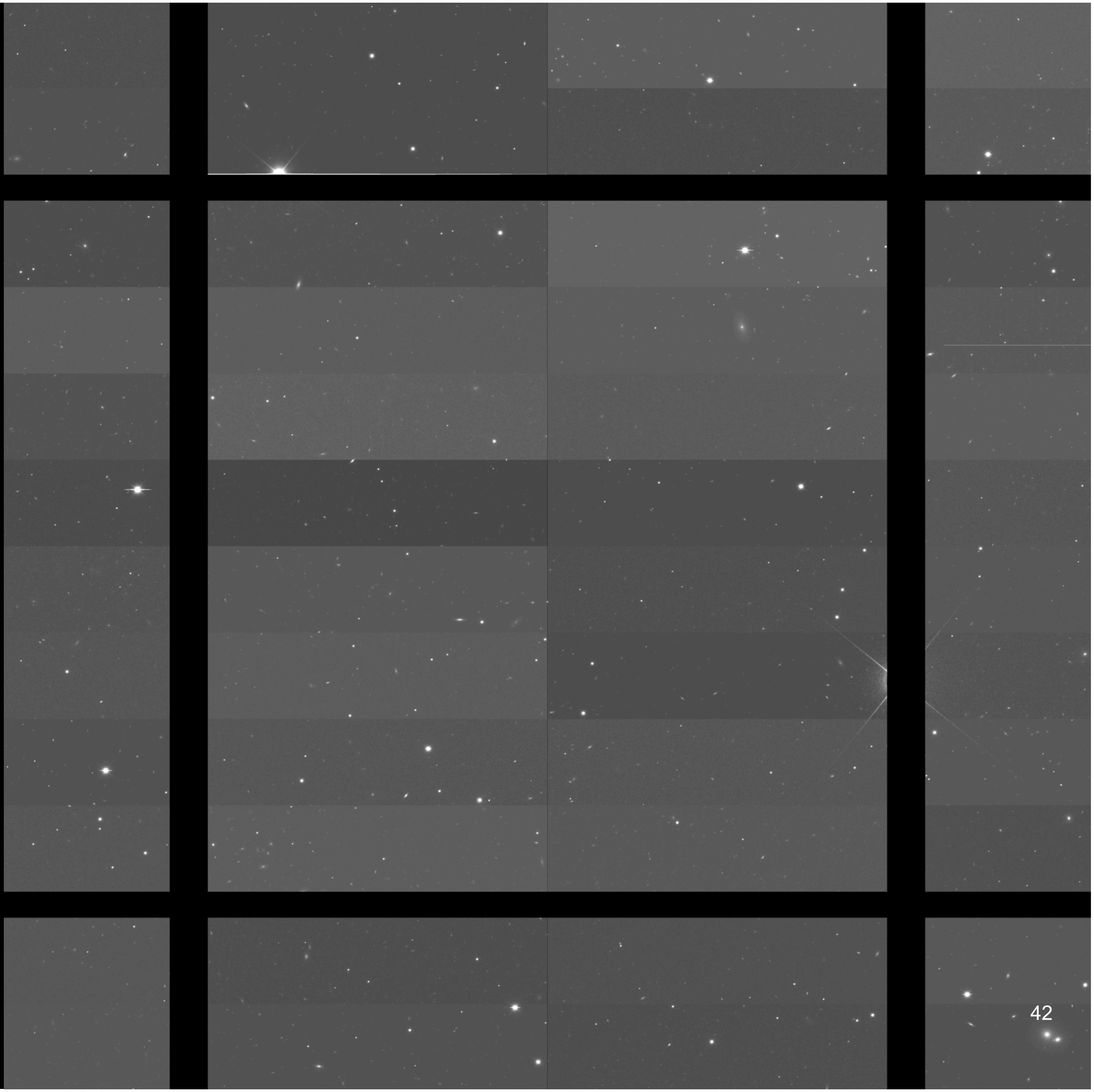
Central
5 x 5 chips

~1 sq. degree

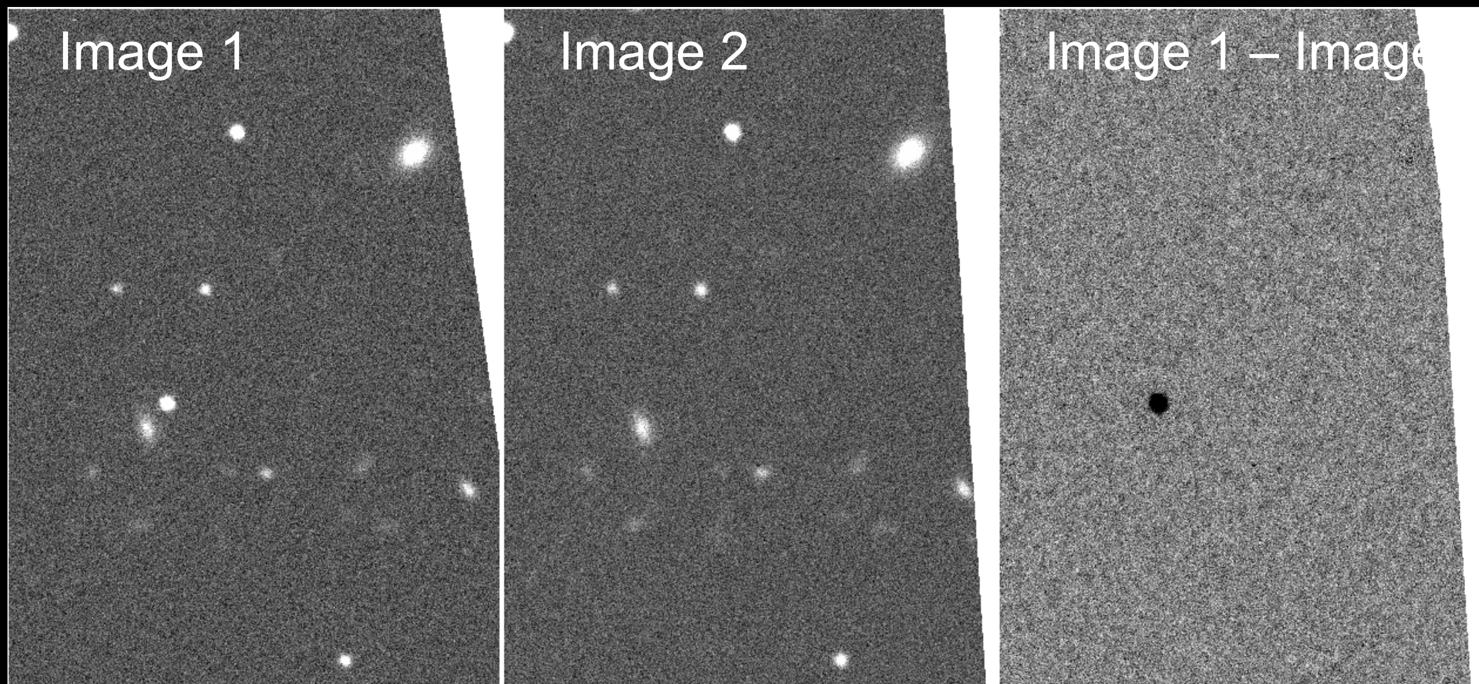


Central chip

15’ x 15’



Detecting variability



Describing an image as a model, a convolution and noise

$$S(x,y) = (PSF \otimes R)(x,y) + \varepsilon(x,y)$$

Difference imaging requires solving for the mapping kernel

$$\chi^2 = \sum_{x,y} \frac{\left[K(x,y) \otimes S_2(x,y) - S_1(x,y)\right]^2}{\sigma_{x,y}^2}$$

Tracking variability in incomplete data

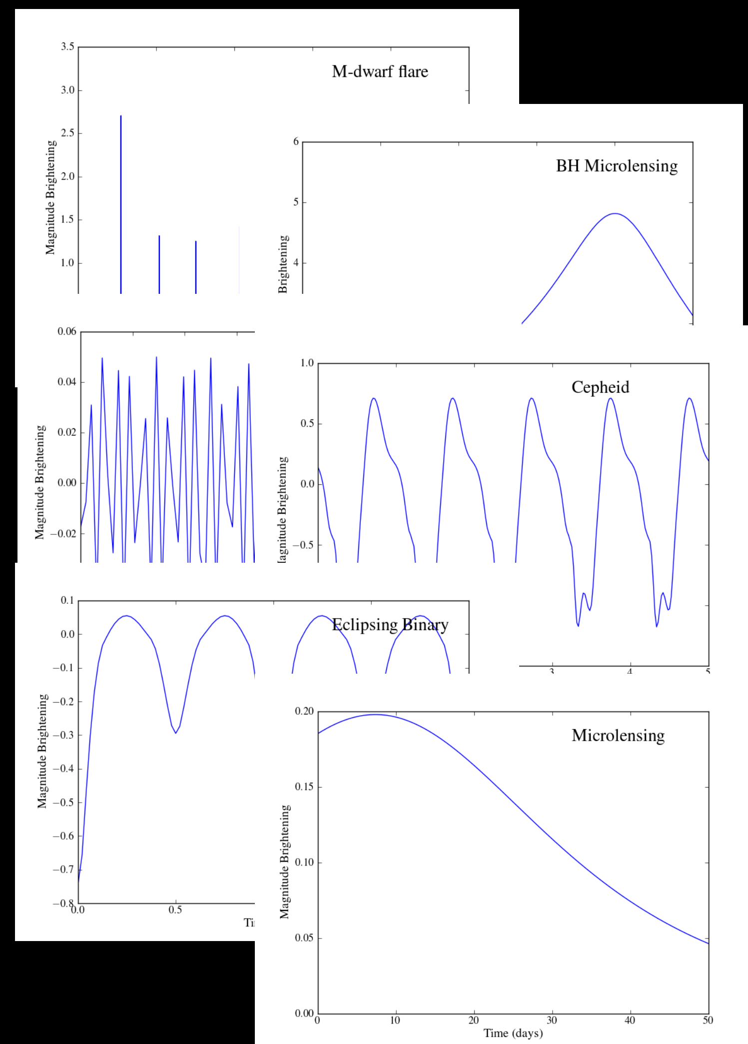
Periodicity from light curves

$$y_i = S\left(wt_i\right) + \varepsilon_i$$

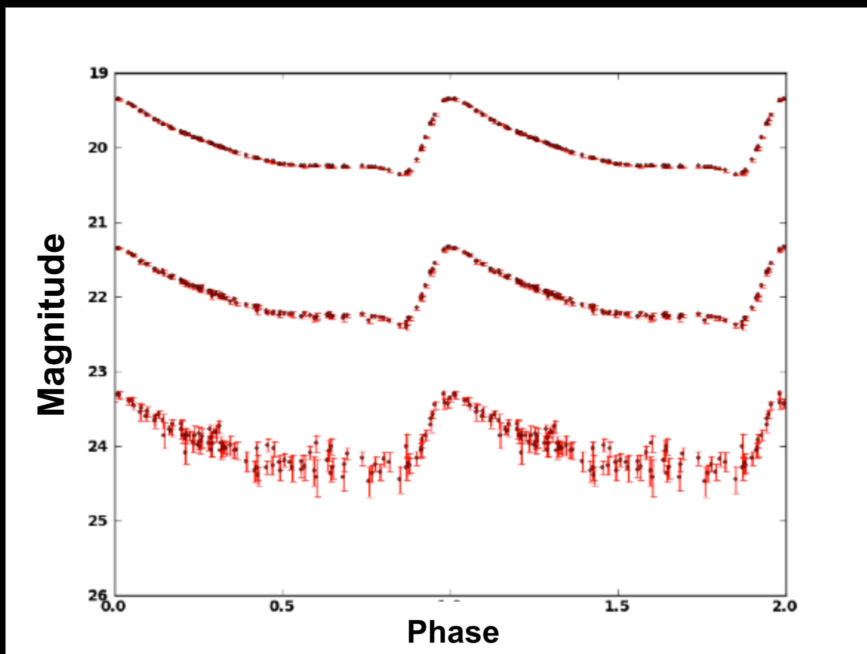
Generalize a source, S, into a set of harmonic functions or search for smooth representations of the data

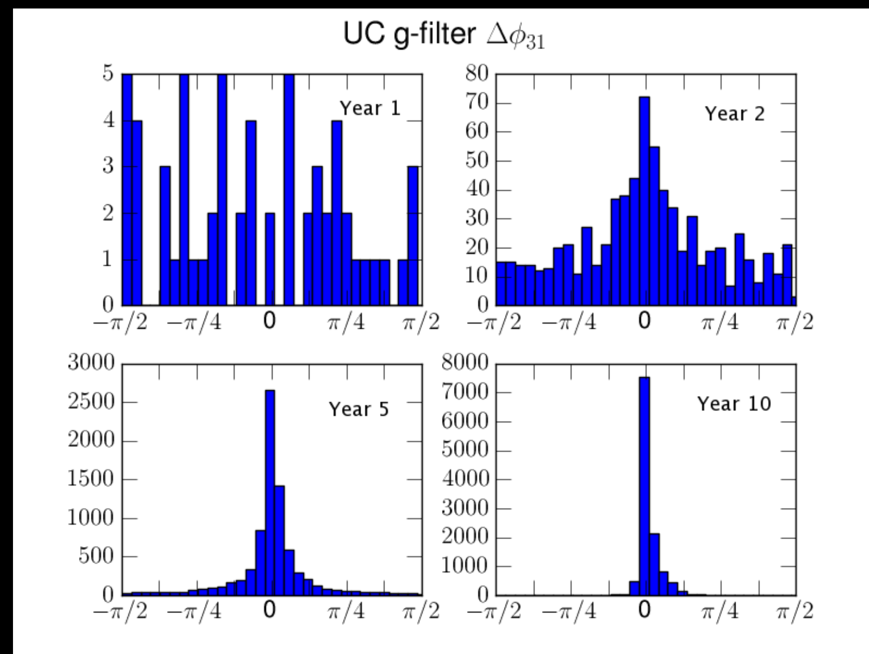
$$\mathrm{SAR}_{\mathrm{n}}(\mathrm{v}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{1}{\sigma_i} \left| y_i - \hat{y}_{i,v} \right|$$

Supersmoother (Friedmann 1984, Reimann 1994): a variable span smoother. Span of a local linear smooths chosen by cross-validation.



Building a lightcurve (a day at a time)





Folding over 10 years of data (inhomogeneous sampling)

Lightcurve shape as a proxy for metalicity (phases in a fourier series). Noise in period determination (sparse sampling) reflected in the metalicity accuracy)

Combining all wavelengths would produce year 10 results in year 2

How do we work with large complex data sets?

Lets start by looking at AstroML

Statistics, Data Mining and Machine Learning in Astronomy:
A Practical Guide for the Analysis of Large Survey Data

Ivezic, Connolly, vanderPlas, Gray (2013)

The AstroML website