LSST: Challenges and Opportunities in AstroInformatics

K. Simon Krughoff

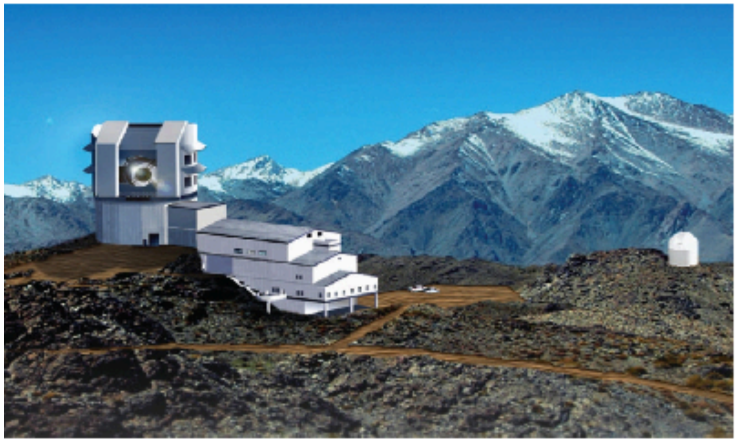
University of Washington

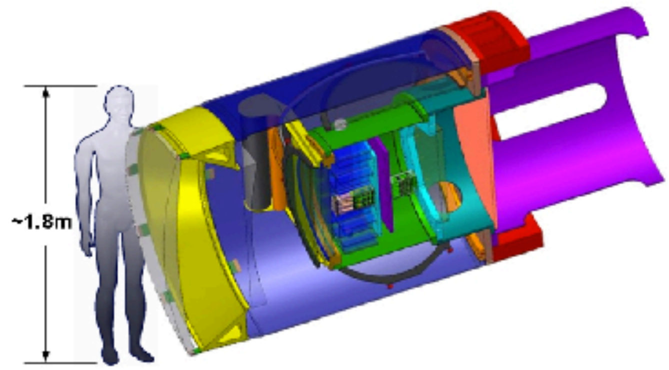
LSST Team

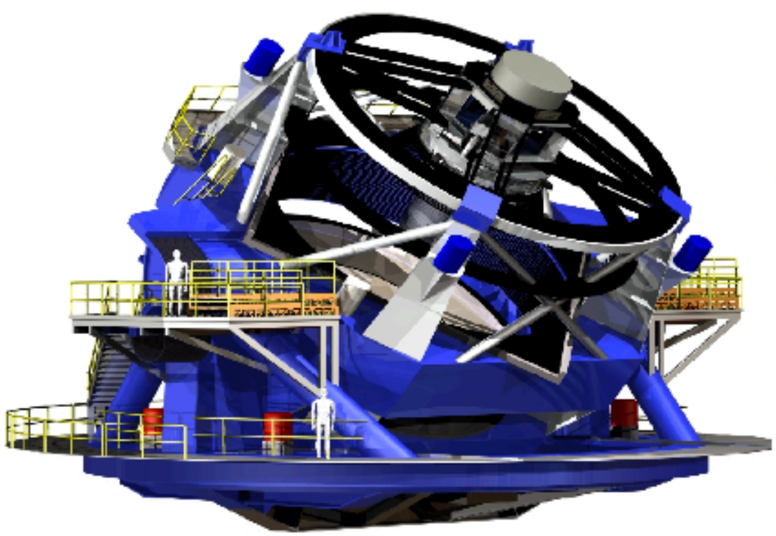
Large Synoptic Survey Telescope

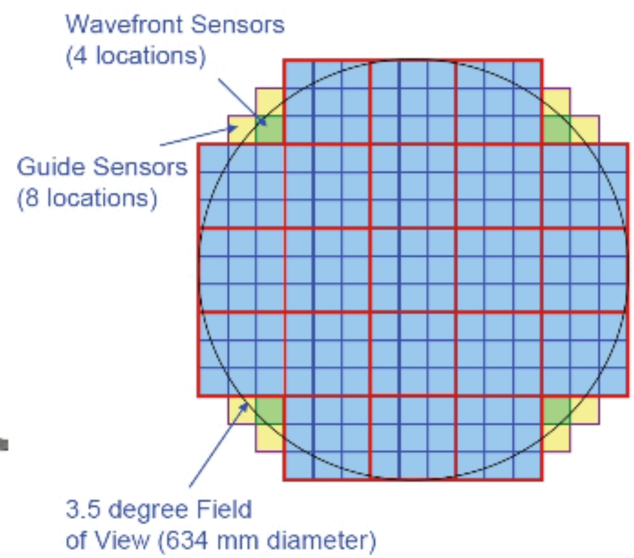
http://www.lsst.org/lsst/about/team

LSST: Overview







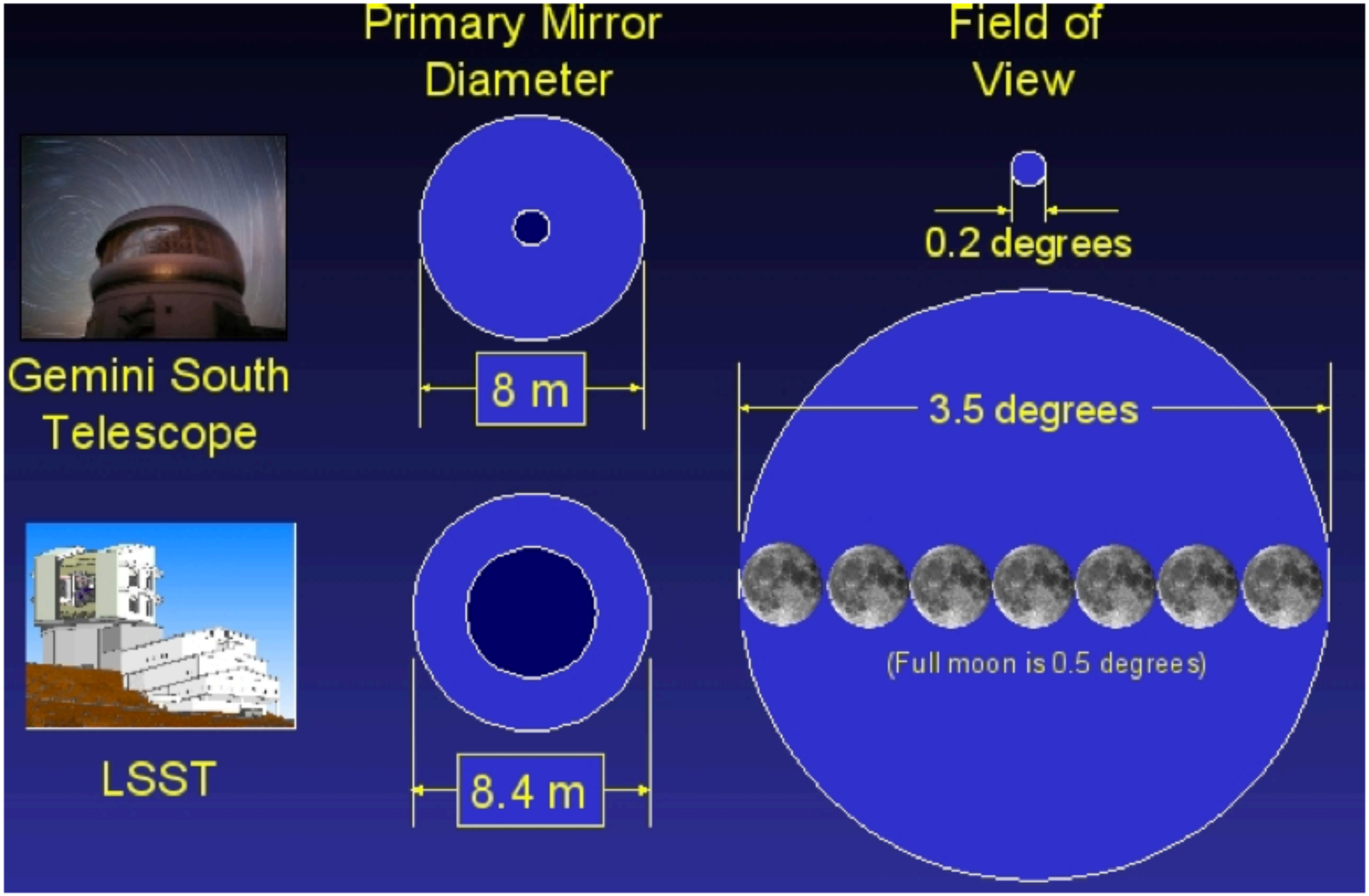


LSST: Overview

Large Synoptic Survey Telescope

- Telescope
  - Modified Paul-Baker
    - 8.4m (6.5 effective) Primary
    - Three mirror system
  - 9.6□°fov
  - $35m^2$ effective collecting area

LSST: Overview



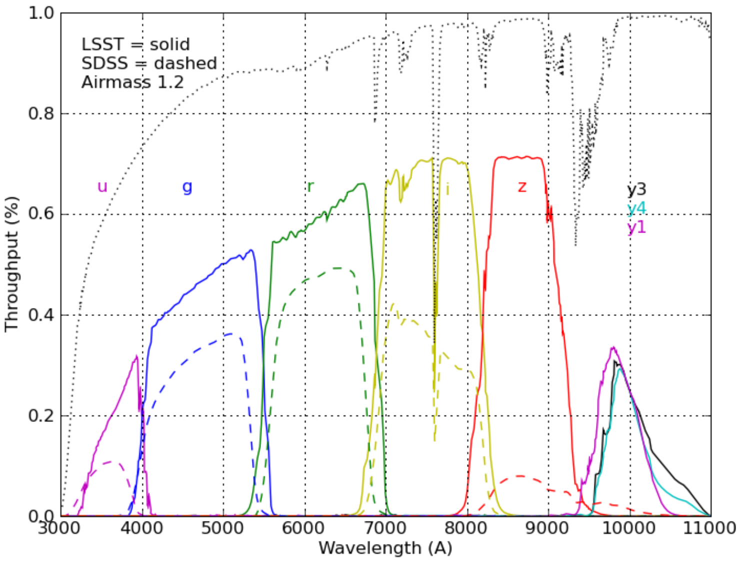
LSST: Overview

- Telescope
  - Modified Paul-Baker
    - 8.4m (6.5 effective) Primary
    - Three mirror system
  - 9.6□°fov
  - 35m2 effective collecting area

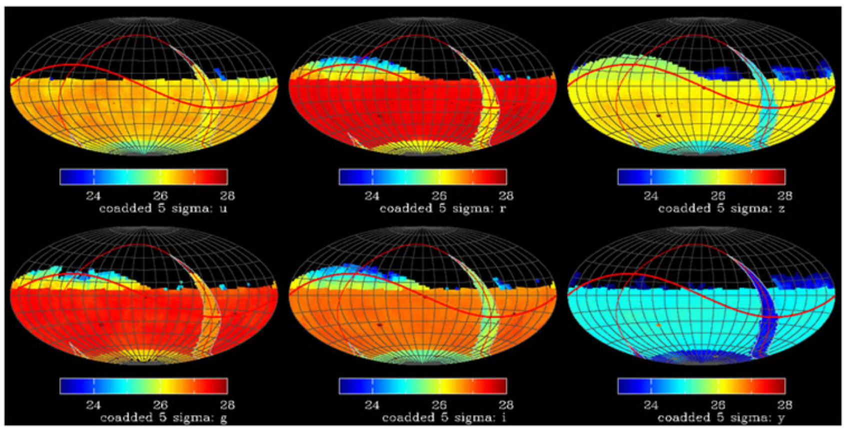
- Camera and Filters
  - 6 filters
  - 2 min filter change
  - Prime focus
  - 3.2 Gigapixels
  - 1 sec shutter overhead
  - 2 sec readout

LSST: Overview

Large Synoptic Survey Telescope

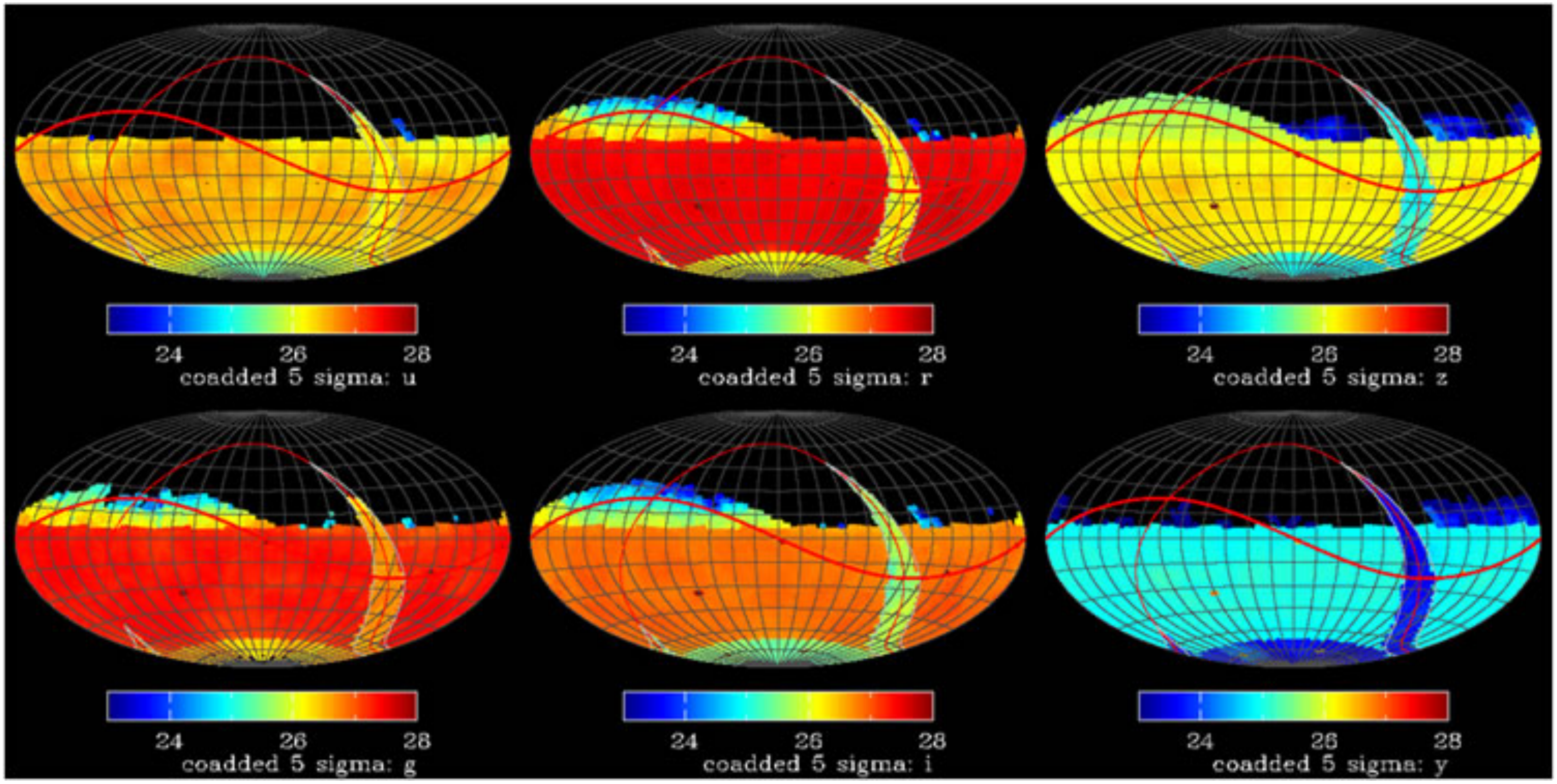


LSST: Overview

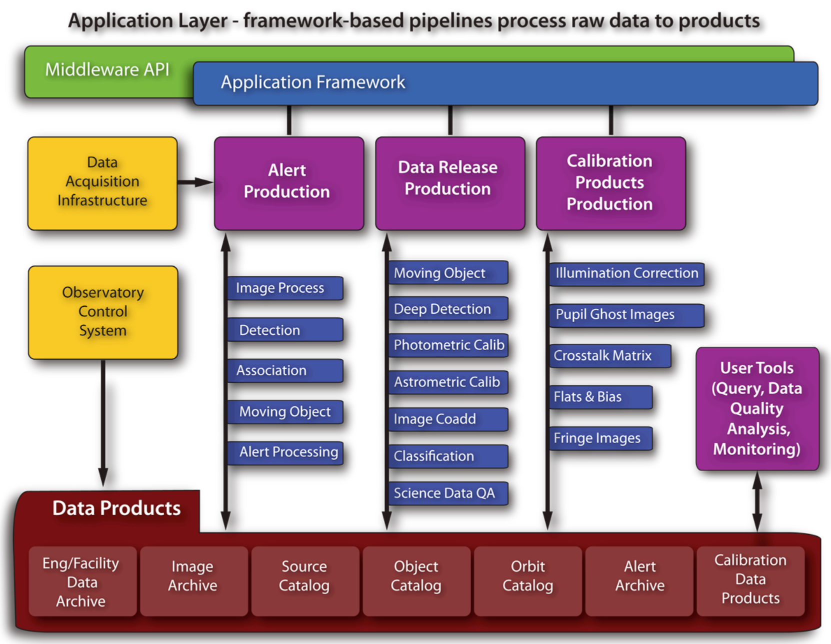


- Cadence and Survey
  - 20,000 □°
  - ~3M visits
    - 2 15 sec exposures
  - 6-8 Sub surveys
    - Solar system, deep drilling, galactic, main
  - ~1000 visits of entire southern sky over 10 years
  - Single image: ~24.5 in r
    Co-added: ~27.5 in r

LSST: Overview



LSST: Overview



- Data
  - Per Night
    - 15 TB of images
    - 1 TB catalogs
    - 60 sec alerts for 10^5-10^6 objects
  - Per Year
    - 6.5 PB per year of images and catalogs
  - Lifetime
    - 10B Stars and 10B Galaxies
    - 60-70PB of images

LSST: Overview

Large Synoptic Survey Telescope

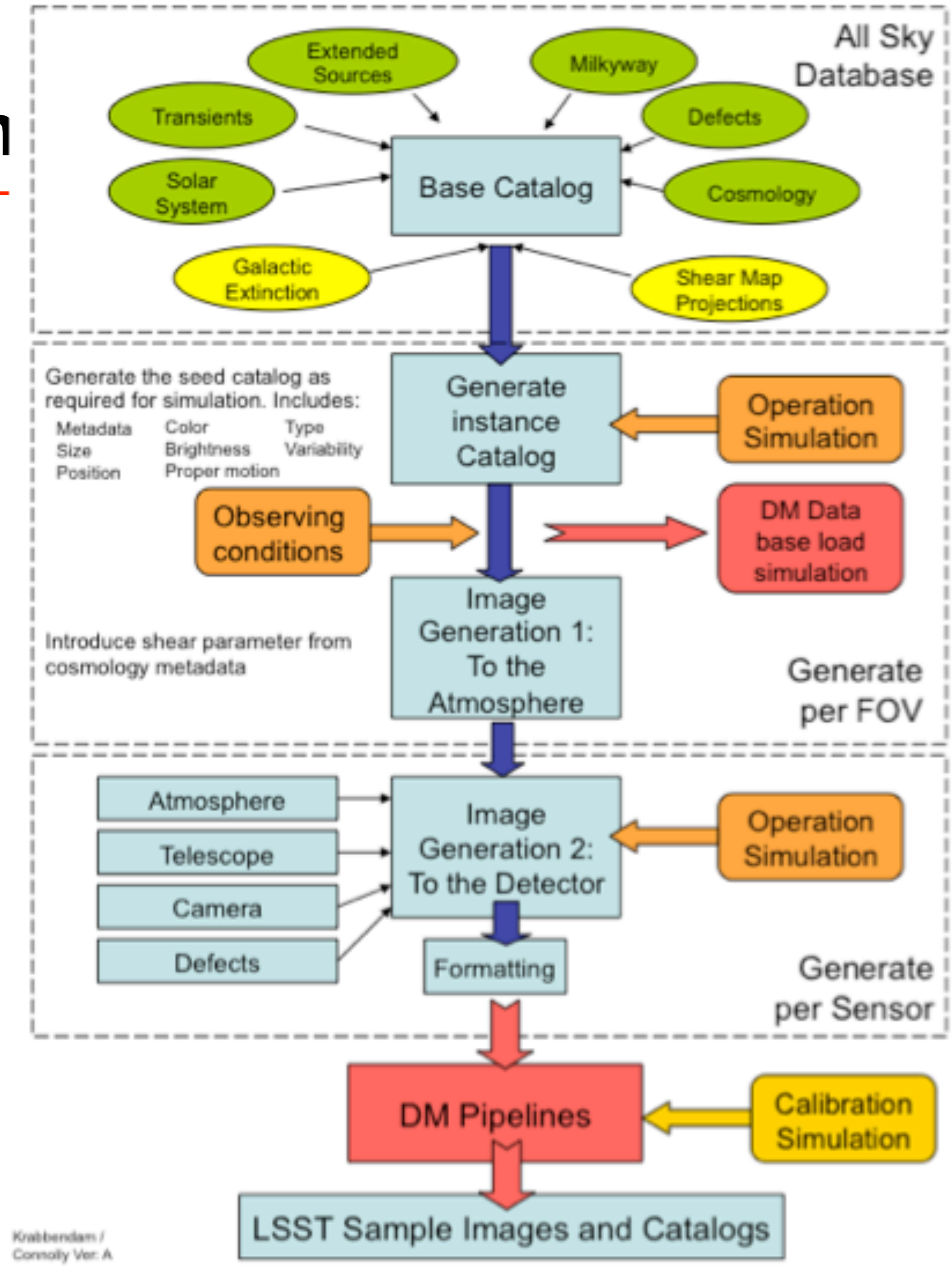
- LSST products
  - The Best Sky Image Ever: 60 PB of astronomical imagery
  - The Greatest Movie of All Time: The entire observable southern sky every three nights for 10 years
  - The Largest Astronomical Catalog: 20 Billion objects (for the first time more objects than people)
- The BIG challenges
  - Rapid and robust classification of sources detected in difference images.
  - Simultaneous treatment of diverse astrometric and photometric time series measurements for an unprecedented number of objects.
    - Classification of 20 billion objects in 3000 dimensional space.

Challenges and Opportunities

- Simulation
  - Of entire system from end to end.
- Survey Design
  - Dithering and cadence
  - Metrics for building the optimal survey
- Analysis
  - Custom analysis on PB datasets.
  - How to keep the development time reasonable.

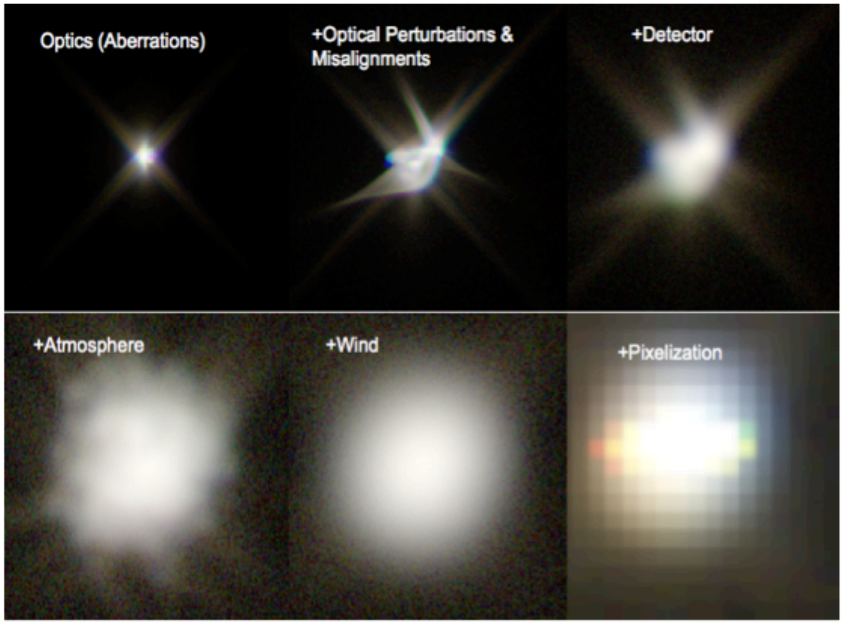
Simulation

Simulation



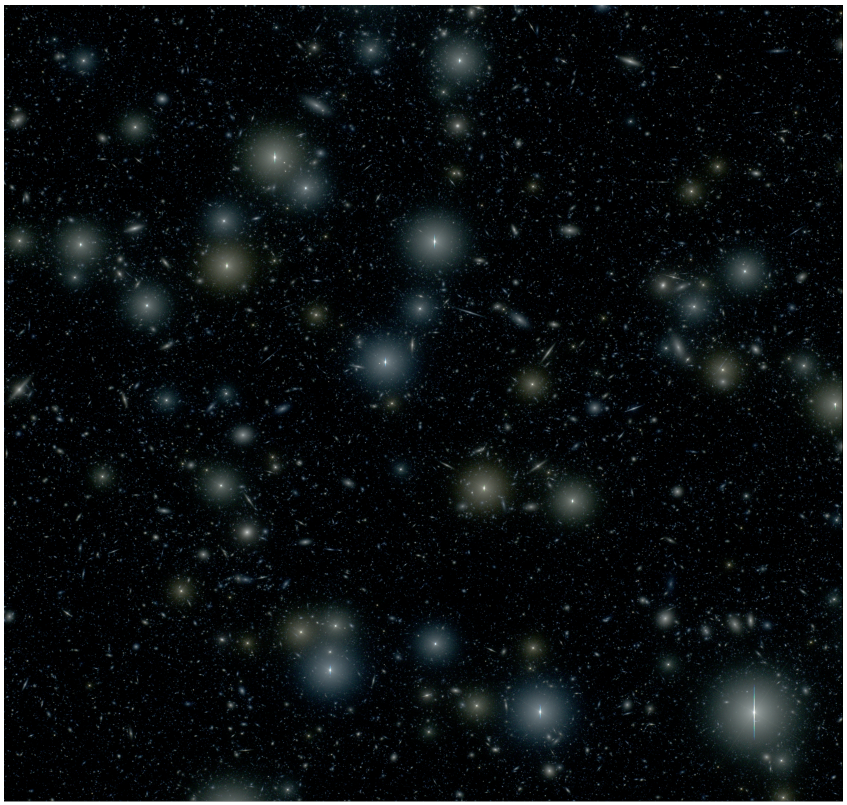
All Sky Database
Extended Sources
Milkyway
Transients
Defects
Base Catalog
Solar System
Cosmology
Galactic Extinction
Shear Map Projections
Generate the seed catalog as required for simulation. Includes:
Metadata Size Position
Color Brightness Proper motion
Type Variability
Generate instance Catalog
Operation Simulation
Observing conditions
DM Data base load simulation
Image Generation 1: To the Atmosphere
Introduce shear parameter from cosmology metadata
Generate per FOV
Atmosphere
Telescope
Camera
Defects
Image Generation 2: To the Detector
Operation Simulation
Formatting
Generate per Sensor
DM Pipelines
Calibration Simulation
LSST Sample Images and Catalogs
Krabbendam / Connolly Ver. A

Simulation



- Raytrace
  - High fidelity
  - Realistic atmosphere, clouds, optics

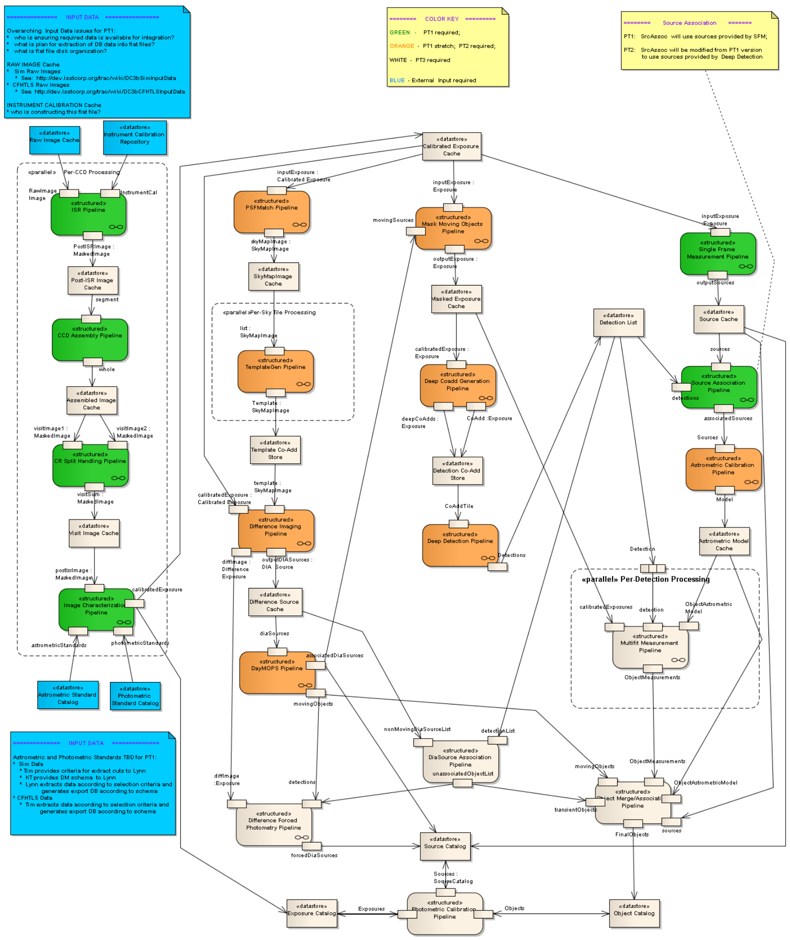
Simulation



- Raytrace
  - High fidelity
  - Realistic atmosphere, clouds, optics
- Catalogs
  - Galactic, solar system, extragalactic
  - Transient and variable
    - SNe, flares, GRBs, AGN, RR Lyrae, etc

Large Synoptic Survey Telescope

Simulation

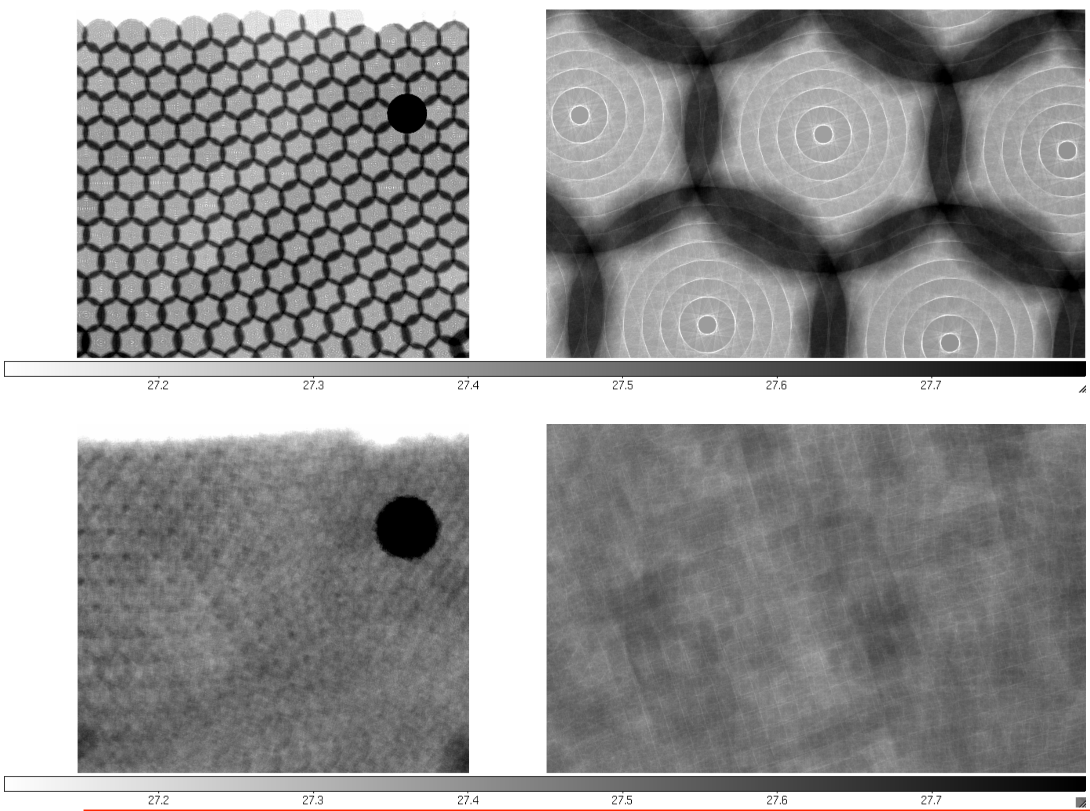


- **Raytrace**
  - High fidelity
  - Realistic atmosphere, clouds, optics
- **Catalogs**
  - Galactic, solar system, extragalactic
  - Transient and variable
    - SNe, flares, GRBs, AGN, RR Lyrae, etc.
- **Data flow**
  - Realistic conditions and rate

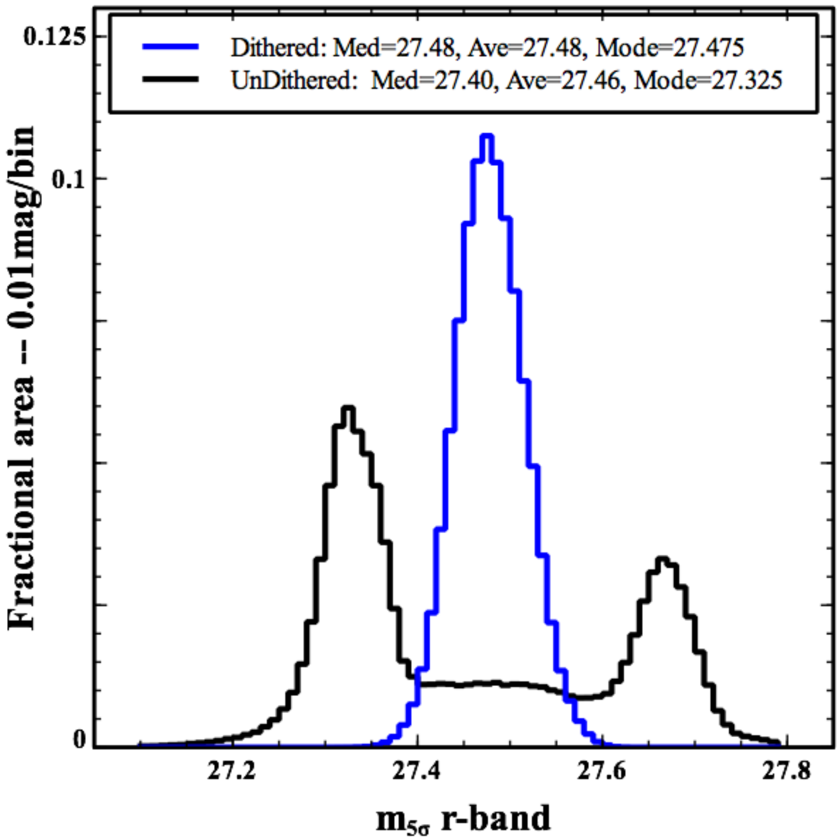
Survey Design

Survey Design

- To dither or not
  - Smooth the instrument signature
  - Gain 0.1 mag in depth

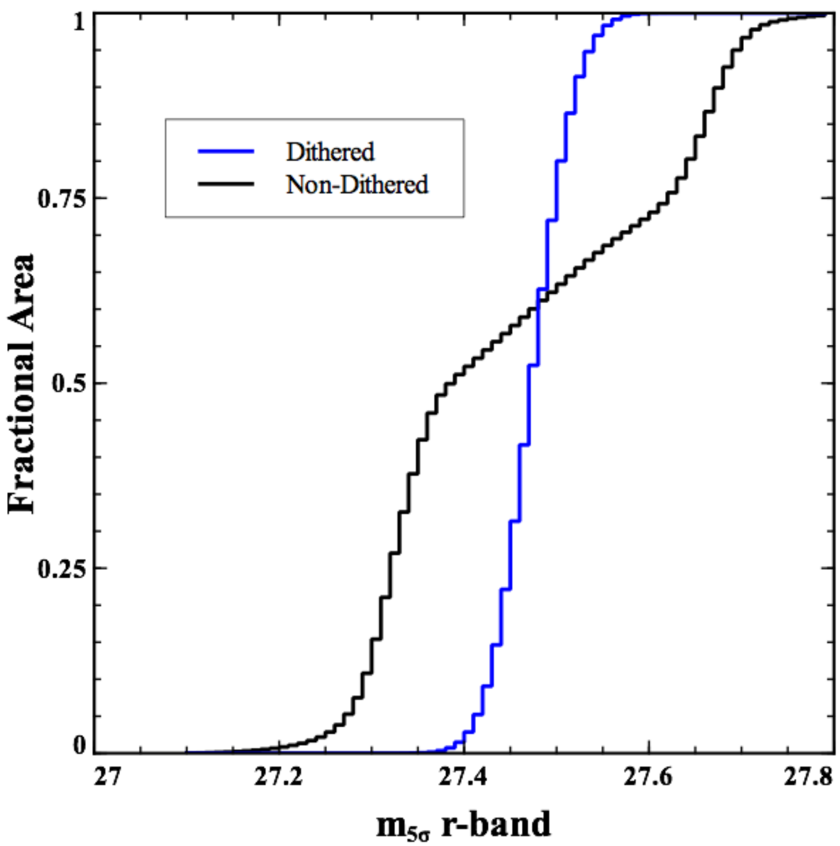


Survey Design



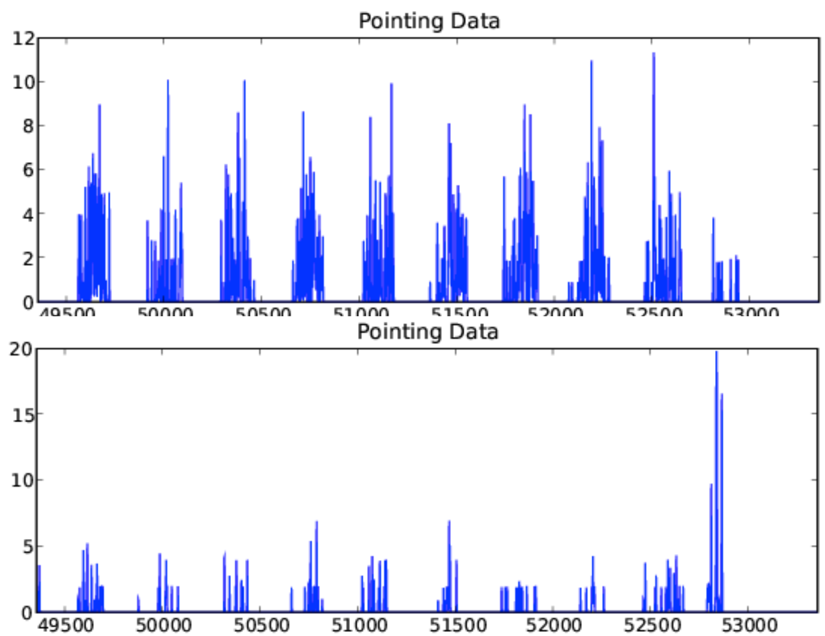
- To dither or not
  - Smooth the instrument signature
  - Gain 0.1 mag in depth

Survey Design



- To dither or not
  - Smooth the instrument signature
  - Gain 0.1 mag in depth

Survey Design



- To dither or not
  - Smooth the instrument signature
  - Gain 0.1 mag in depth
- Cadence
  - Revisit time can have large impact on detection efficiency
  - How does one measure this?
- How to divide survey time between major projects

Analysis

Analysis

- The right tool for the job
  - How can typical astronomers execute custom analysis on PB data sets?
  - Can data providers be analysis engines too?
  - Does this require a new language?
  - Can industry standards for large scale commodity computation aid in this task?

Analysis

- The right tool for the job
  - LSST data centers will touch every pixel every time there is a reduction made.

Analysis

- The right tool for the job
  - How can typical astronomers execute custom analysis on PB data sets?
  - Can data providers be analysis engines too?
  - Does this require a new language?
  - Can industry standards for large scale commodity computation aid in this task?

Analysis

Large Synoptic Survey Telescope

- The right tool for the job
  - In order to keep development costs down, interaction with the data will require a new way of programming.
  - Examples:
    - PIG – Uses user defined functions to build workflows that are translated to map reduce
    - SciDB – Understands interacting with array/pixel based data in a relational DBMS language.

Analysis

- The right tool for the job
  - How can typical astronomers execute custom analysis on PB data sets?
  - Can data providers be analysis engines too?
  - Does this require a new language?
  - Can industry standards for large scale commodity computation aid in this task?

Analysis

- The right tool for the job
  - Current “cloudy” architectures don’t work well with astronomical data.
  - Distributed file systems dislike lots of relatively small files.
  - Scheduling overhead in open source MapReduce implementations require individual work units to run for ~minutes.

Thank you!

Questions?