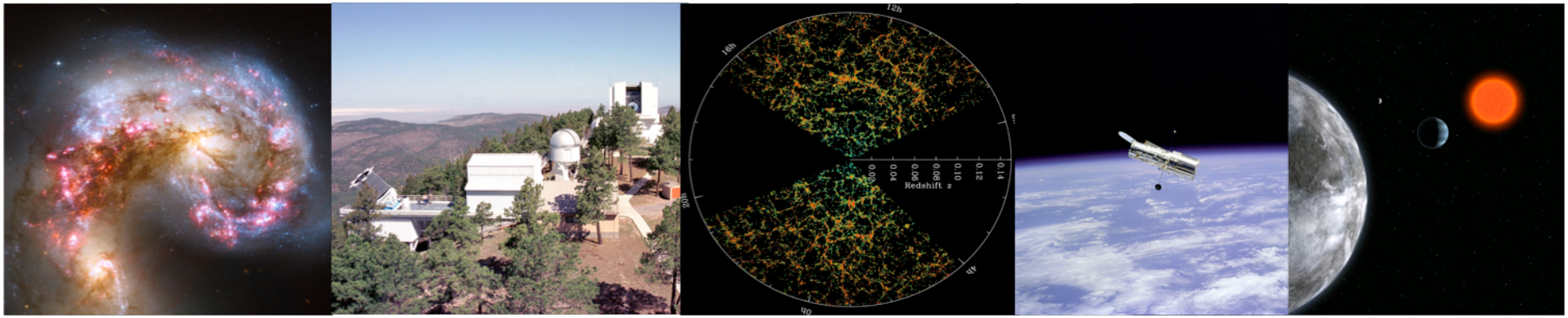
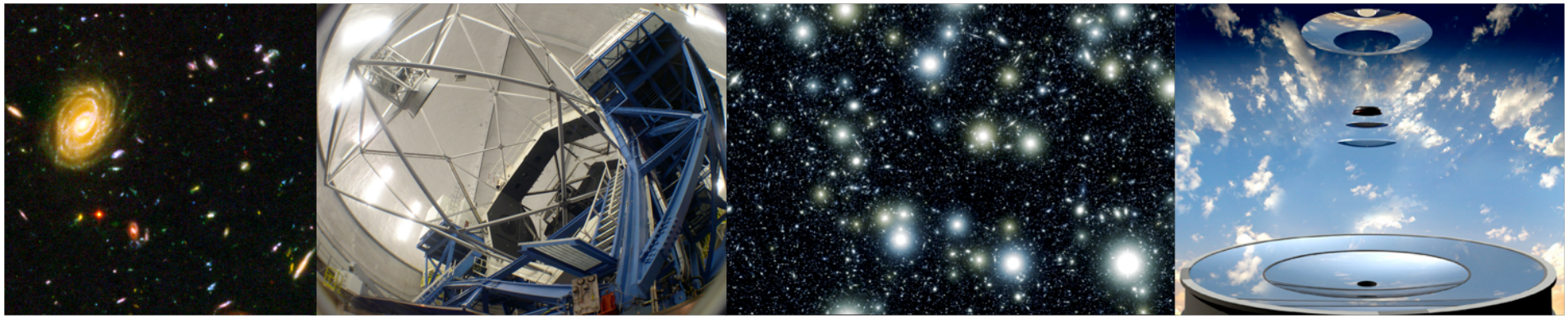


Simulations: following the photons

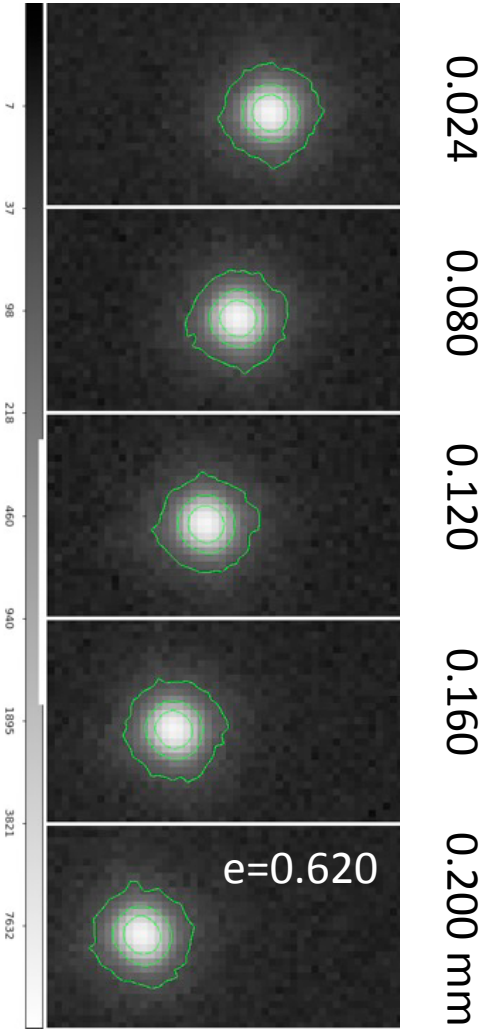
Andrew Connolly

University of Washington



Why do we need instrument simulations?

- **Requirements for level IV DE missions are substantial**
  - **1% absolute photometry, 0.5% relative photometry**
    - A factor of 2 improvement on current ground-based wide-field surveys such as the SDSS
  - **Astrometry uncertainties of 0.01 arcsec**
    - A factor of 3 improvement over constraints in current surveys such as the SDSS
  - **Correlated shape residuals of $<10^{-5}$ in the PSF**
    - In the context of weak lensing, to achieve this requires 3 orders of magnitude improvement in PSF interpolation (as a function of wavelength and position on the focal plane), and 2 orders of magnitude improvement in our ability to calibrate shear.
  - **The ability to scale production, analysis, and science**
    - LSST data volume represents a 1000-fold increase over the SDSS (30-fold over PanSTARRS/DES) – but the largest (in use) science database are around 100TB in size



Decentering of LSST M2

Learning from past surveys

- **Photometric calibration**
  - **SDSS: photometric survey of ~10,000 sq degrees**
  - **Need for accurate PSF characterization recognized 2 years after first light, scattered light was found to impact the flat fields 4 years after first light, photometric calibration completed 5 years after first light, background subtraction issues 7 years after first light**
- **Shape characterization**
  - **CFHTLS: weak lensing survey of 200 sq degrees**
  - **Simplified data analysis showed coaddition of the image stacks was appropriate for weak lensing. Systematics showed up in more detailed analyses (B-modes in good seeing data)**
  - **Papers appearing 3 years after completion of CFHTLS, 9 years after start**
- **Simulations provide a limited view of the truth**

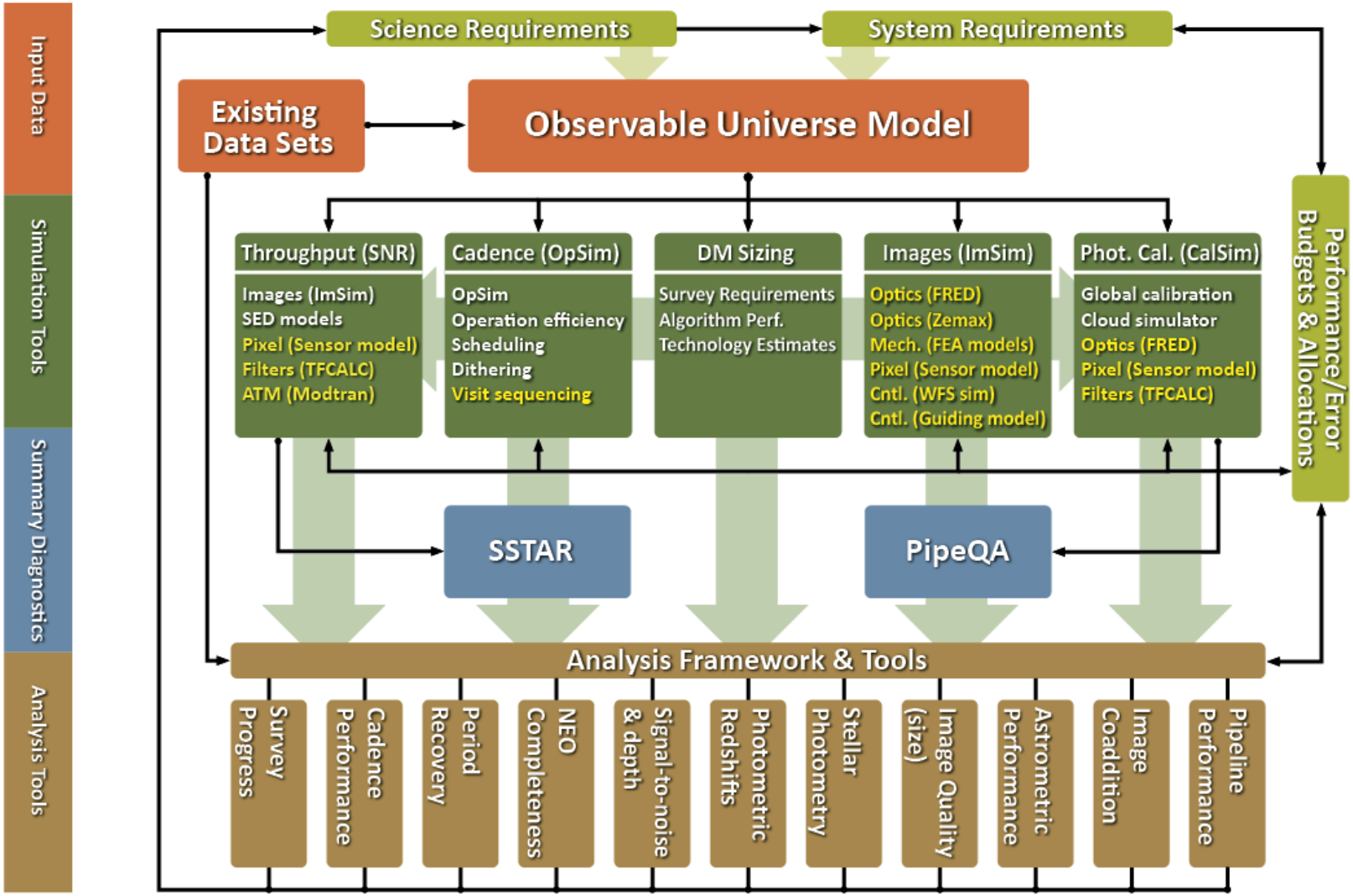
Delays are expensive

Comparison of Astronomical Simulations

Project	Role	Cosmology	Galactic Structure	Galaxy Model	Atmosphere	Telescope	Camera
LSST	End-to-end	Millenium	Juric et al	Sersic	Raytrace	Raytrace	Raytrace
DES	End-to-end	Las Damas	Giradi et al	Shapelets	Analytic	Analytic	Analytic
PanSTARRS	Instrument	N/A	N/A	N/A	Wavefront	Wavefront	Analytic
GAIA	End-to-end	N/A	Mansana et al	Skymaker	N/A	Analytic	Analytic includes CTE
ALMA	Instrument	N/A	N/A	Real images	Analytic	Analytic	
Zemax	Engineering	N/A	N/A	N/A	Geometric Wavefront	Geometric Wavefront	N/A

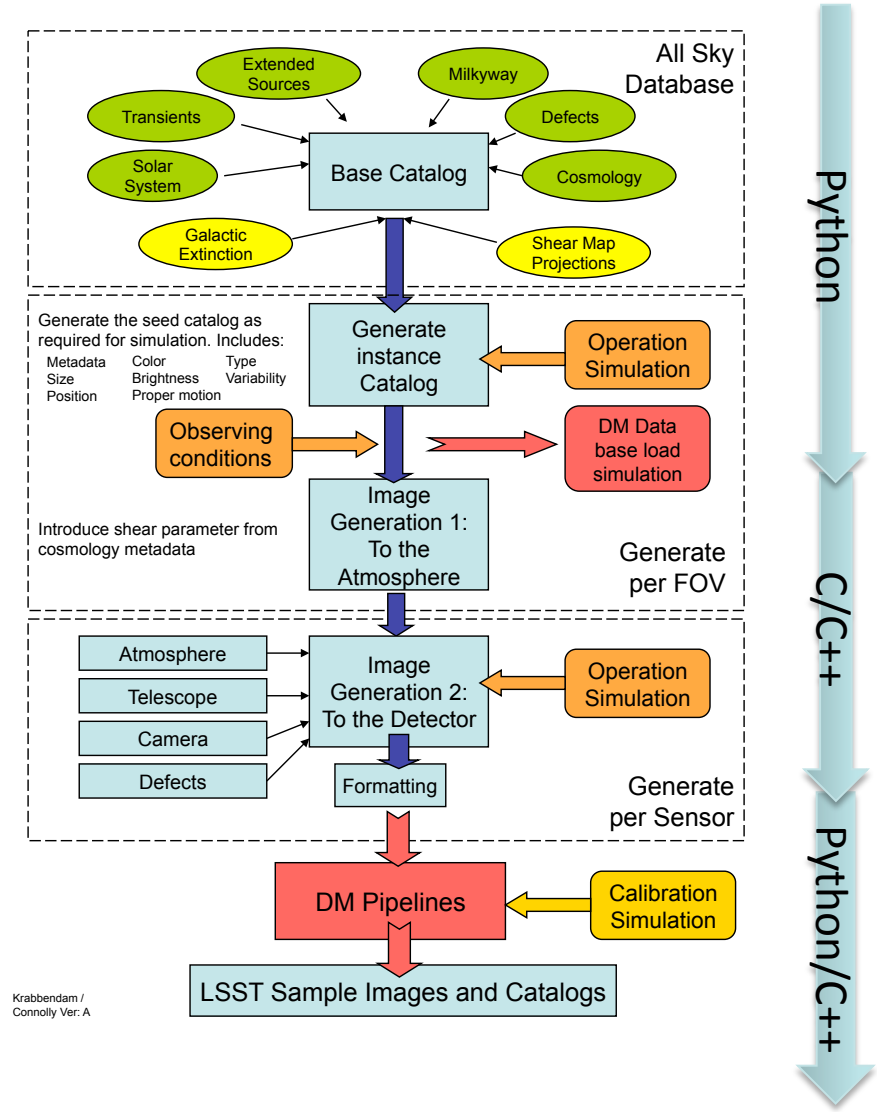
Simulation frameworks to date have focused on simplified data flows (e.g. assuming an analytic form for the point-spread function) that can be used in testing the end-to-end performance (cosmology to images) or engineering simulations that simulate the complete wavefront but are too slow for large scale simulation tests

What do we need?



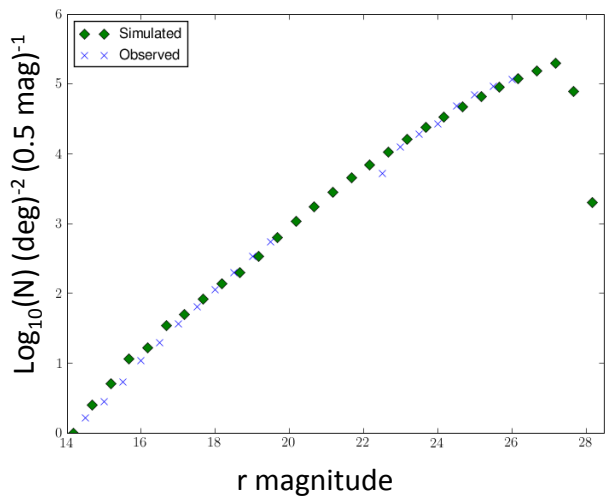
How might we address this

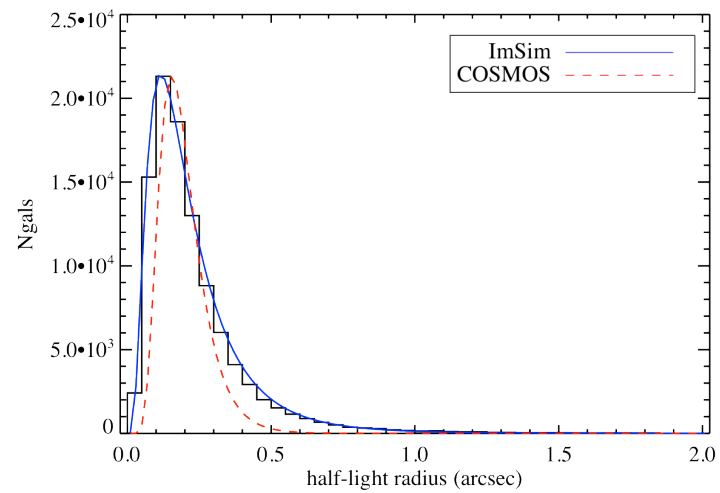
- A representative view of the universe and its properties
- The implementation of the camera, telescope, and site reflects the current LSST designs
- The framework scales from the individual images to Terabytes
- Fidelity of the simulations is designed to match the current needs of the project
- The system is flexible to respond to as-delivered components
- Images complement the capabilities of the analytic models, catalog simulations and real data



Data driven models of the universe

A universal model of the sky





- **Source counts are based on simulations of the universe matched to observed densities and color of sources. For example, galaxy catalogs are based on the Millennium survey (de Lucia et al 2006) but modified to reproduce the observed number counts, size distributions, and redshifts.**
- **Simulations complement the observed data, providing a simplified view of the sky which can be used as a reference to evaluate the performance of the LSST system**
- **More than just galaxies**

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**

Fine

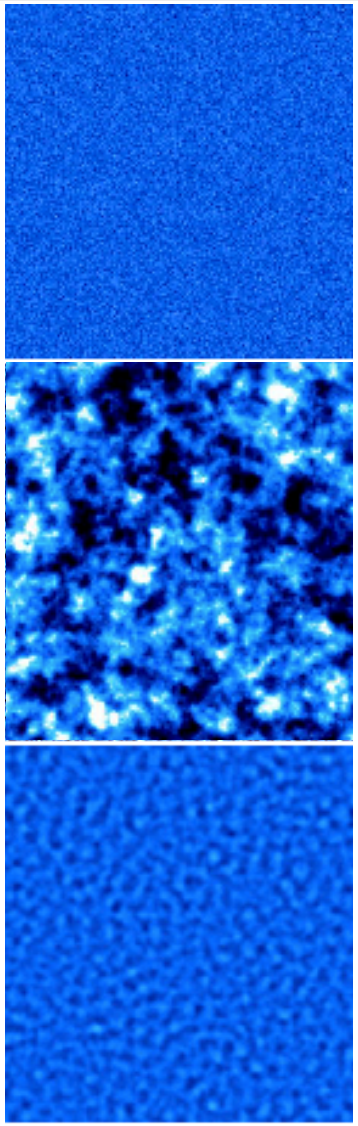
1pix=1cm

Medium

1pix=4cm

Coarse

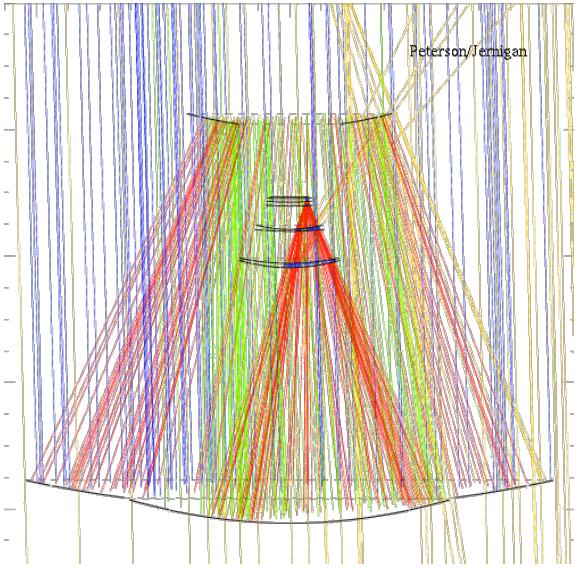
1pix=16cm

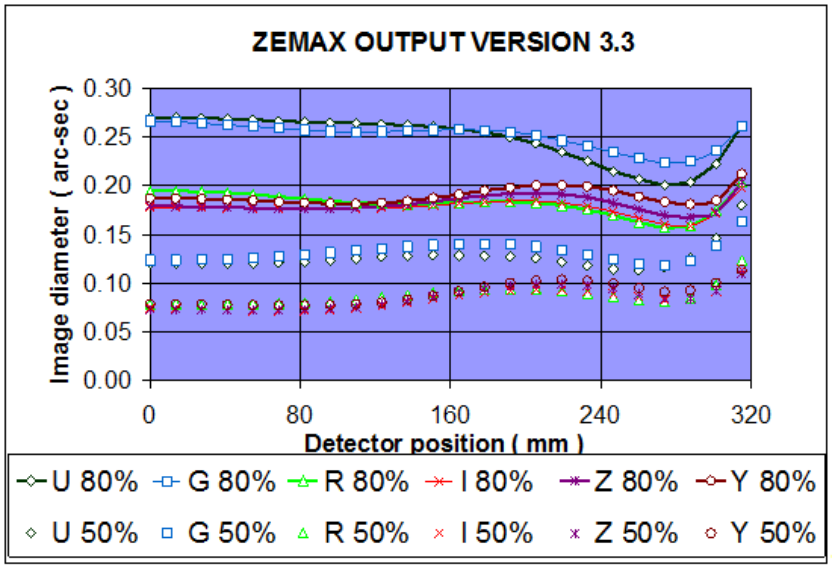


Vorontsov et al. 2008

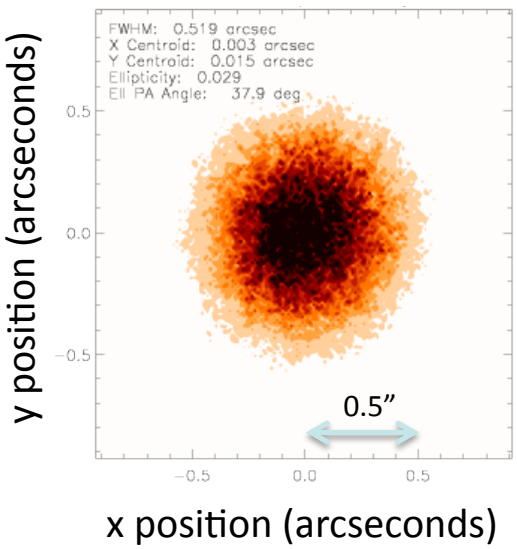
Telescope and camera

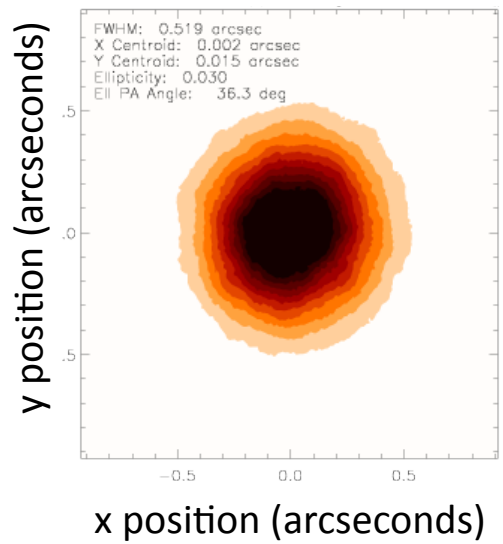
- **Telescope modeling**
  - Implemented from the LSST optical design
  - Currently incorporates a simplified perturbation model
- **Camera modeling**
  - Full focal plane model including wavefront sensors
  - Throughputs matched to baseline design
  - A simplified model of the electronics and perturbations within the camera

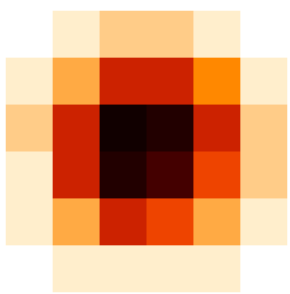


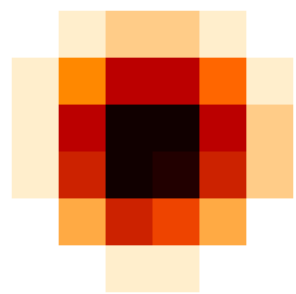


The speed and fidelity trade off





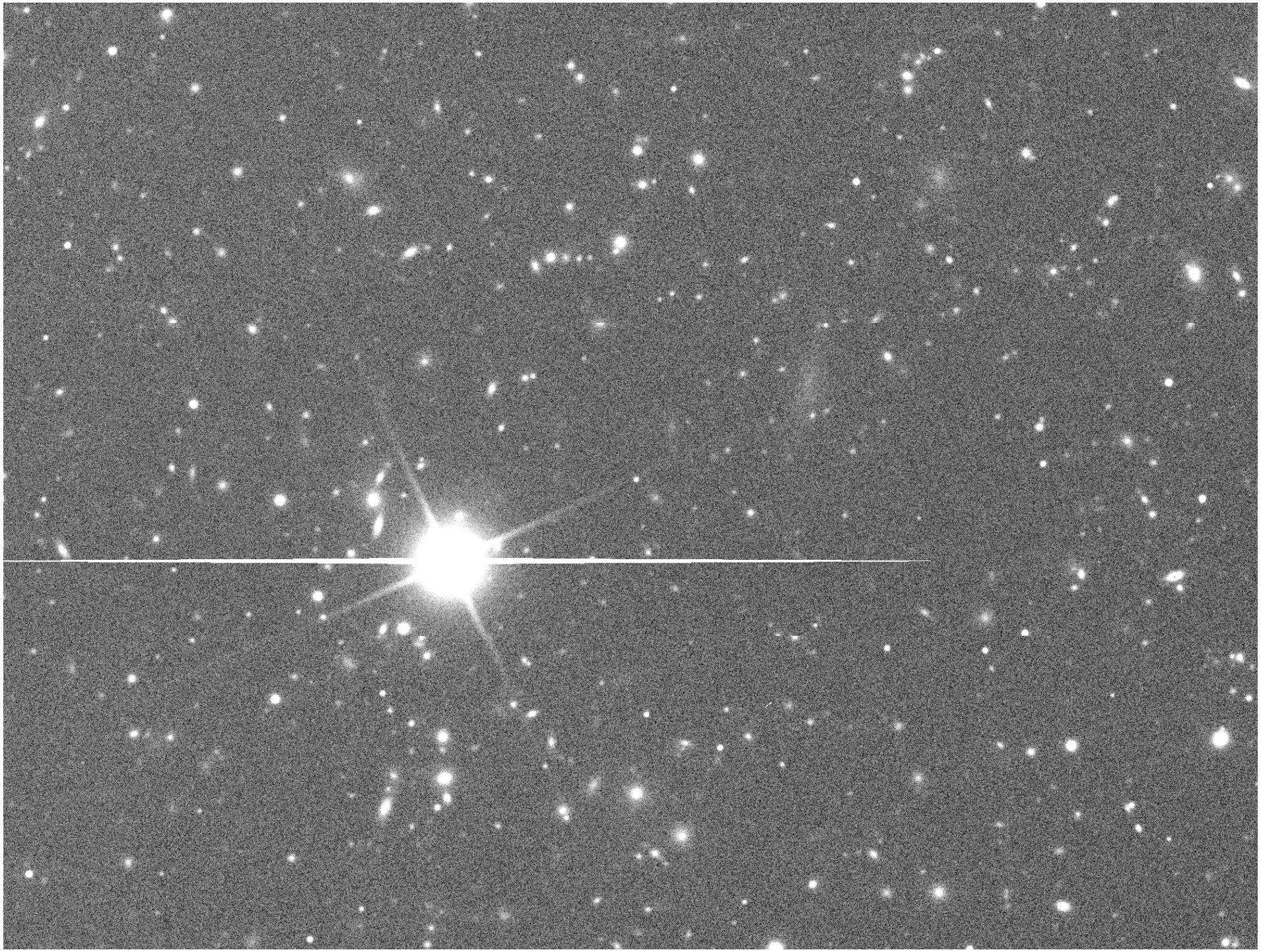




Peterson et al 2013

GalSim/Phosim

GalSim/Phosim



3 Gigapixels

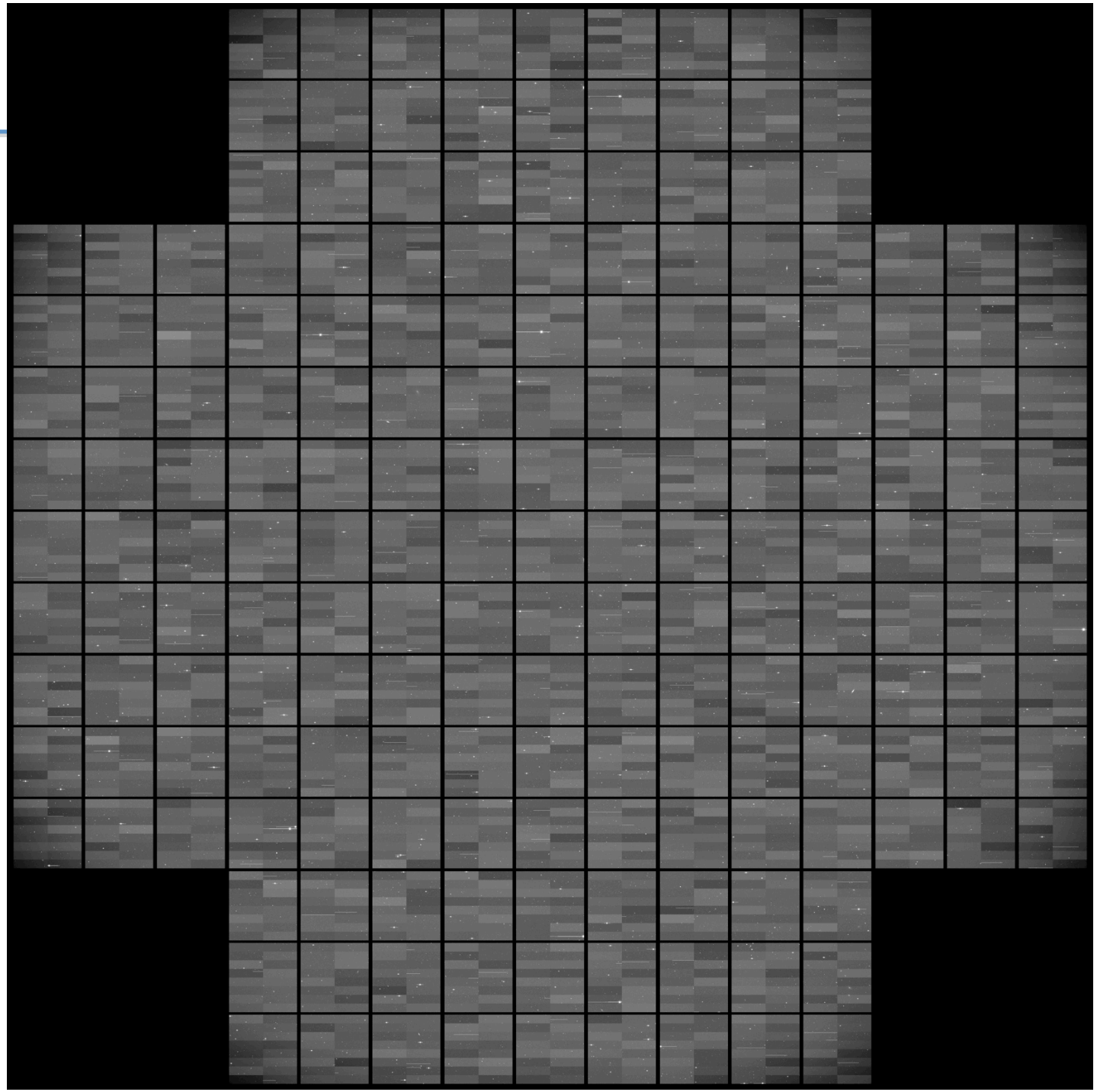
10 sq. degrees

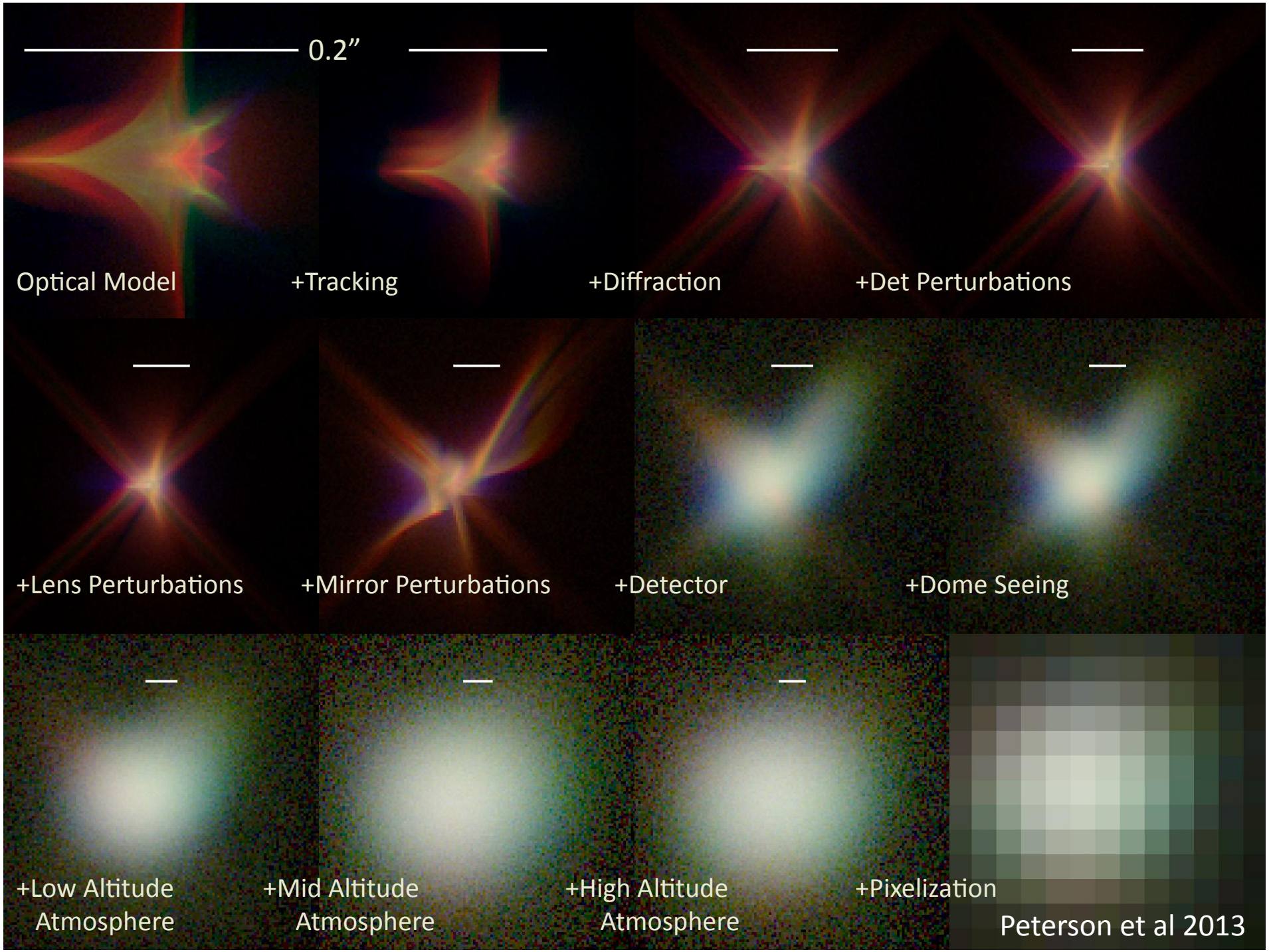
20 million sources

10^{10} photons

11 Gbytes

1000 CPU hours




0.2”
Optical Model
+Tracking
+Diffraction
+Det Perturbations
+Lens Perturbations
+Mirror Perturbations
+Detector
+Dome Seeing
+Low Altitude Atmosphere
+Mid Altitude Atmosphere
+High Altitude Atmosphere
+Pixelization
Peterson et al 2013

Public simulation frameworks

- **PhoSim**
  - https://dev.lsstcorp.org/cgit/LSST/sims/phosim.git/refs
    - Python and C++ code (v3.2.5) and walkthroughs
    - Continuous development

- **GalSIm**
  - **https://github.com/GalSim-developers/GalSim**
    - Note: first fully-supported public release in mid-2013
    - Already has 95% of necessary functionality

- **Public simulations enable testing, validation, and extension to different applications**

Importance of data challenges

GREAT3

The third GRavitational lEnsing Accuracy Testing challenge

http://great3challenge.info

Supported in part by:

1. NASA via the Strategic University Research Partnership (SURP) Program of the Jet Propulsion Laboratory
2. The IST Programme of the European Community, under the PASCAL2 Network of Excellence, IST-2007-216886

PASCAL2 Pattern Analysis, Statistical Modelling and Computational Learning

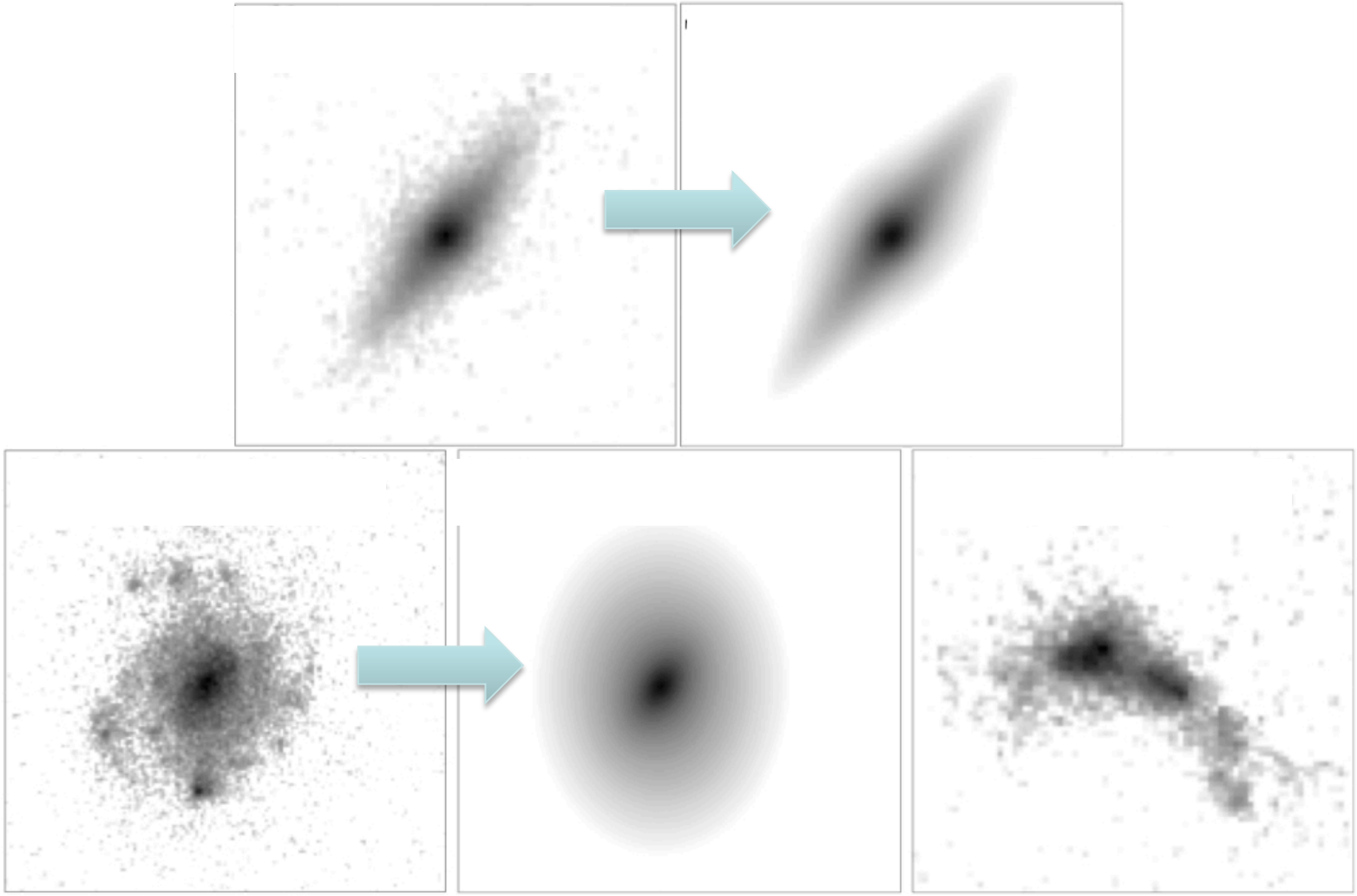
Progressive development of challenges

- **Breaking down development into incremental challenges**
- **GREAT10 star challenge: first systematic test of PSF estimation, interpolation – crucial for identifying promising methods! (Kitching et al. 2012)**
  - **Did not propagate through to inferred shears**
  - **Effect on shear depends not just on magnitude of PSF estimation/ interpolation error, but its form (Size? Ellipticity? Other moments?) and spatial correlation of those errors**
- **GREAT3 includes the first systematic test that goes from PSF estimation/ interpolation to the effect on shear**

What are the challenges ahead?

- **Engineering simulations**
  - Tools are in good shape (though slow)
  - Evaluating the sensor characteristics becomes one of the primary challenges (PSF shapes, CTE, non-linear effects)
- **Environment**
  - Structure function of the cloud adds zeropoint uncertainties
  - Variation in the sky absorption and emission limits calibration
  - Extinction model for our Galaxy could have systematic
  - Variations in seeing, sky, airmass etc couple with the astrophysics
- **Astrophysics**
  - Galaxies are messy
  - Properties of galaxies at high redshift and faint magnitudes are not well constrained
- **Building the End-to-End system**
  - Testing more than just individual modes (coupling to the cosmology)

Challenges ahead: galaxies are messy



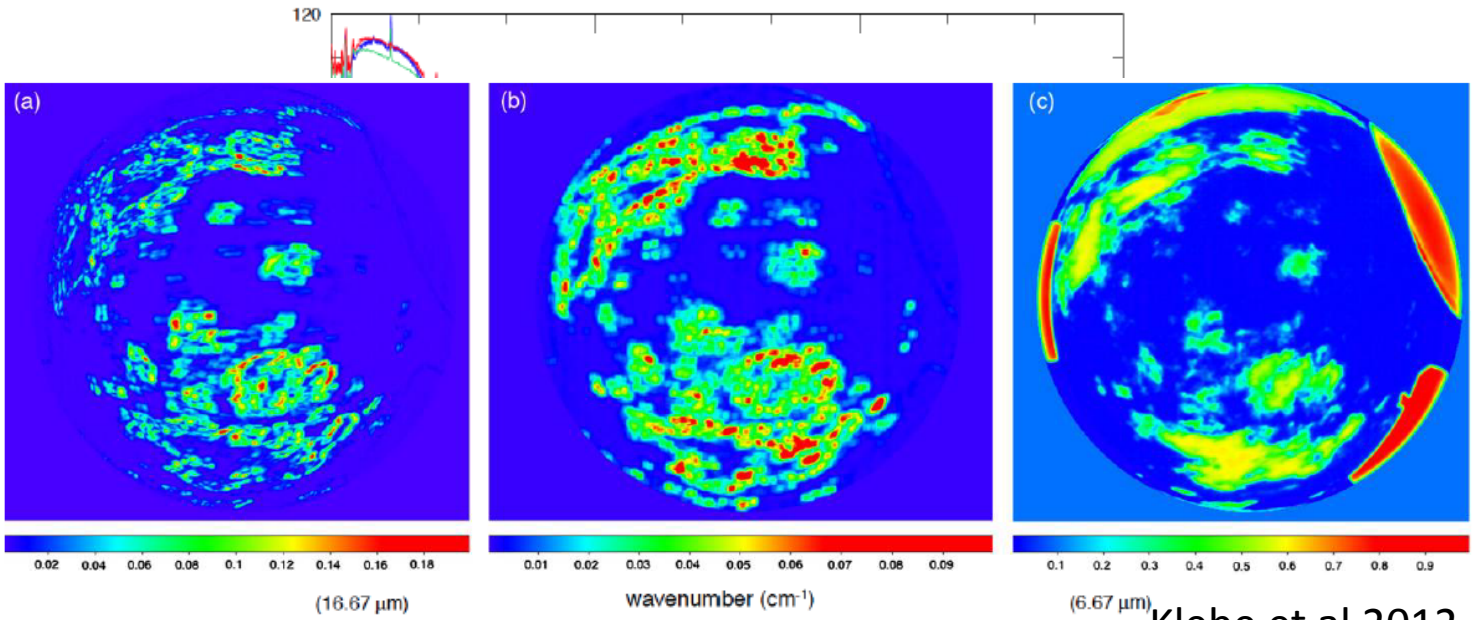
R. Mandelbaum

Challenges for realistic training dataset

- **Knowing the population of galaxies**
  - **Size and shape distributions as a function of redshifts**
  - **No simple method based on 2nd moments is independent of galaxy population**
  - **Second moments can couple to higher order moments**
- **Creating realistic training sets**
  - **Noise in the images gets sheared, convolved, resampled: need to understand this OR get really deep data**
  - **If we want to simulate multi-wavelength data, we need lots of high-res, high-S/N data in many filters to create wavelength dependent data**
  - **Can only properly do this calculation if simulated data has lower-resolution than the original.**

R. Mandelbaum

Challenges ahead: the environment is messy



Klebe et al 2012

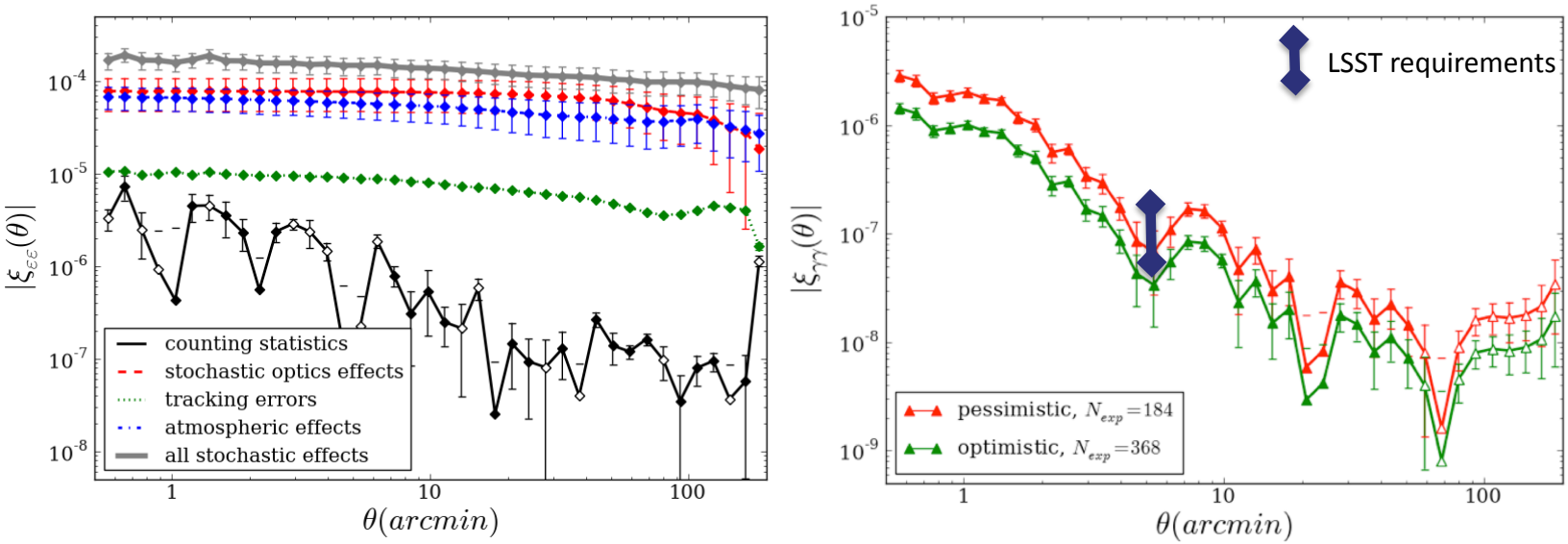
Shaw et al 2005

Models for the distribution of cloud (and its temporal and spatial variations are not well defined (best structure functions are based on 55s SDSS Observations). Requires details site observations.

Steps ahead

- **Deep high resolution multiband imaging surveys**
- **Detailed cosmological surveys (validated against data)**
- **Instrumentation of the site (seeing, transparency, extinction…)**
- **Coupling cosmology to simulations to analysis and back**

Why we want to undertake this effort



Chang et al 2012

- **The contribution to the ellipticity as a function of component (tracking errors, atmosphere, shot noise) and the ellipticity residuals can be characterized for a range of scenarios**