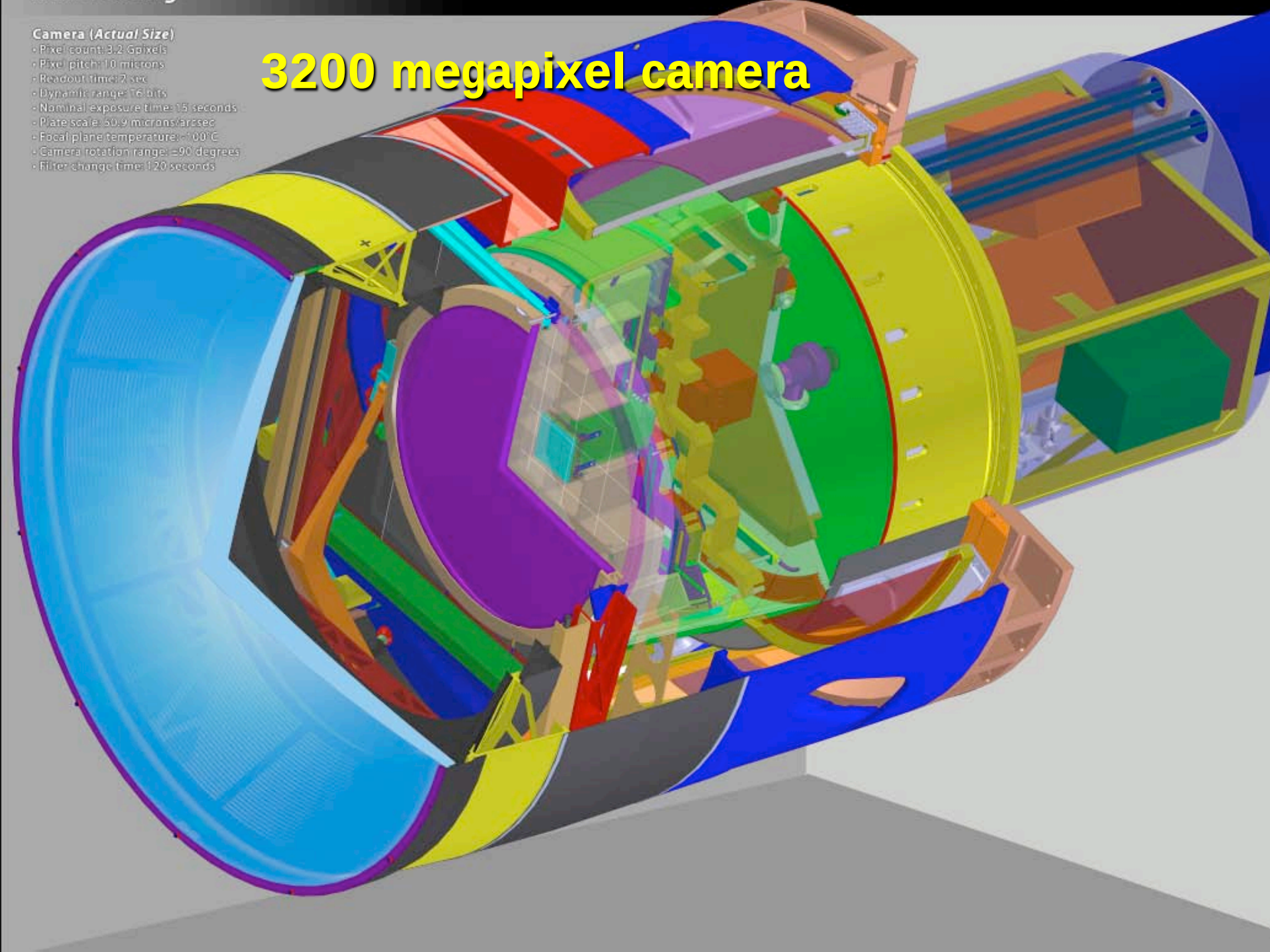
LSST
Dark Energy & Big Data

Tony Tyson

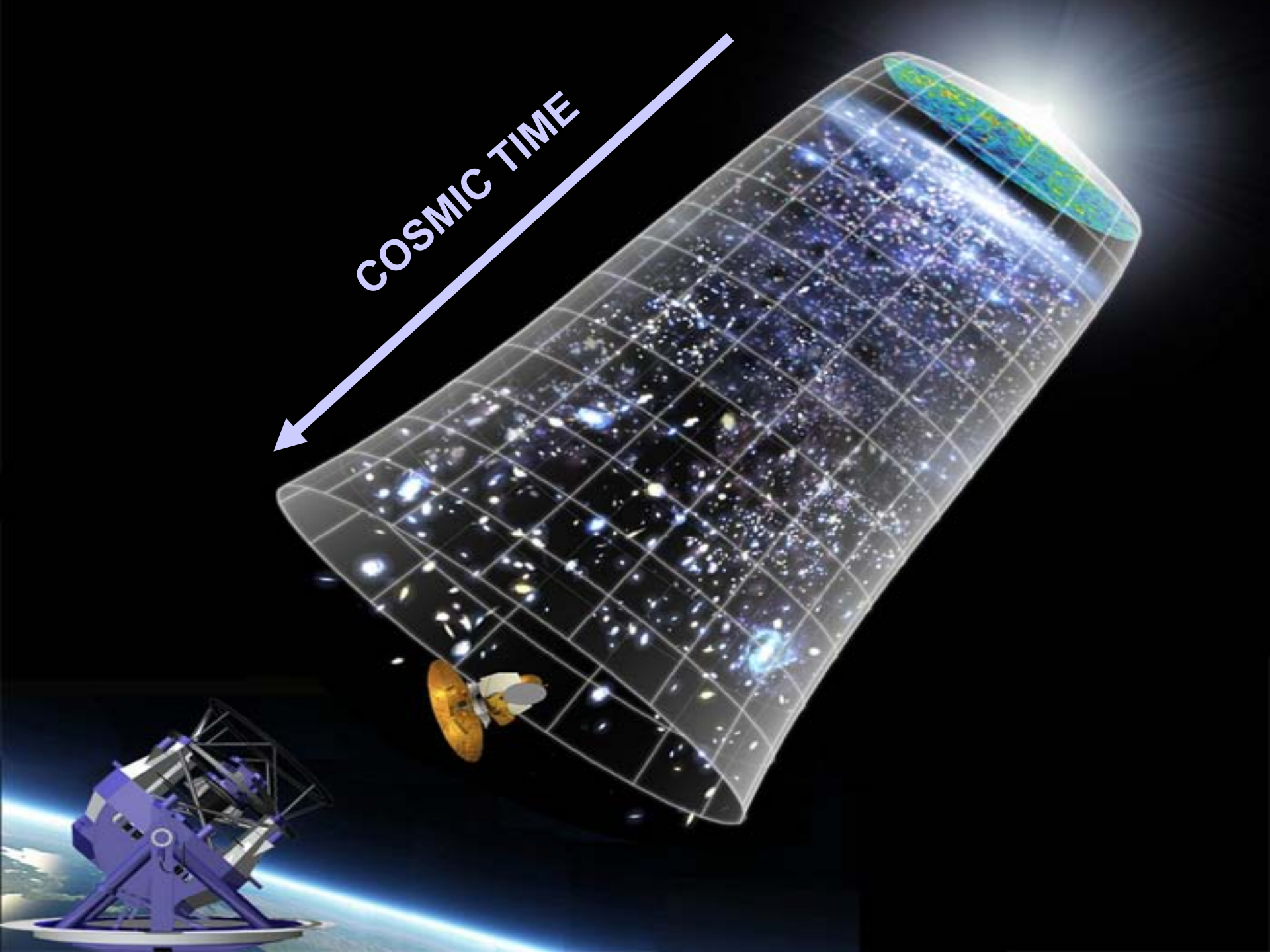
UC Davis

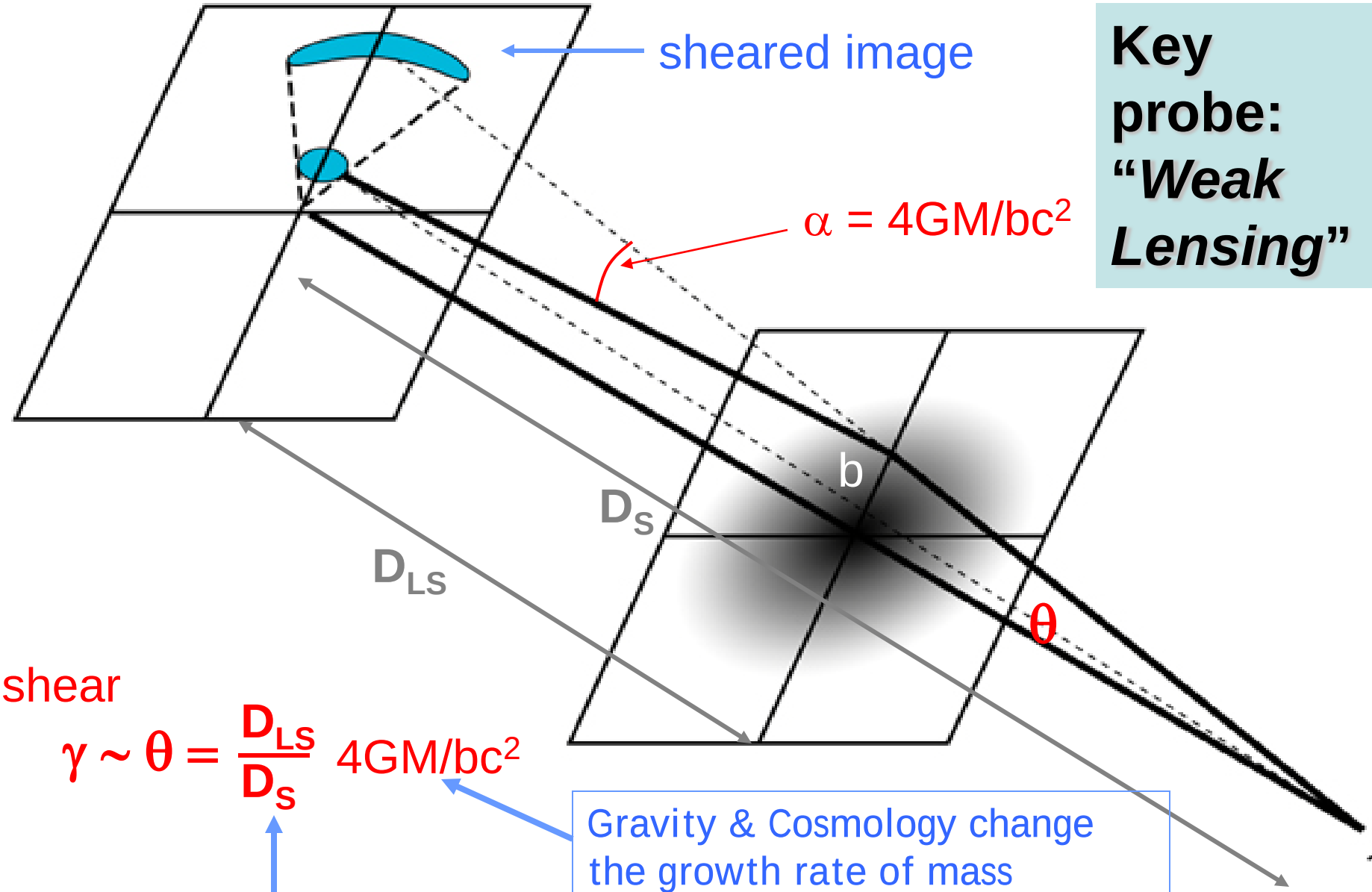
3200 megapixel camera



LSST Wide-Fast-Deep survey

- 4 billion galaxies with redshifts
- 1 million supernovae
- 1 million galaxy lenses


COSMIC TIME



Key probe: "*Weak Lensing*"

shear

$$\gamma \sim \theta = \frac{D_{LS}}{D_S} \, 4GM/bc^2$$

Gravity & Cosmology change the growth rate of mass structure

Cosmology changes geometric distance factors

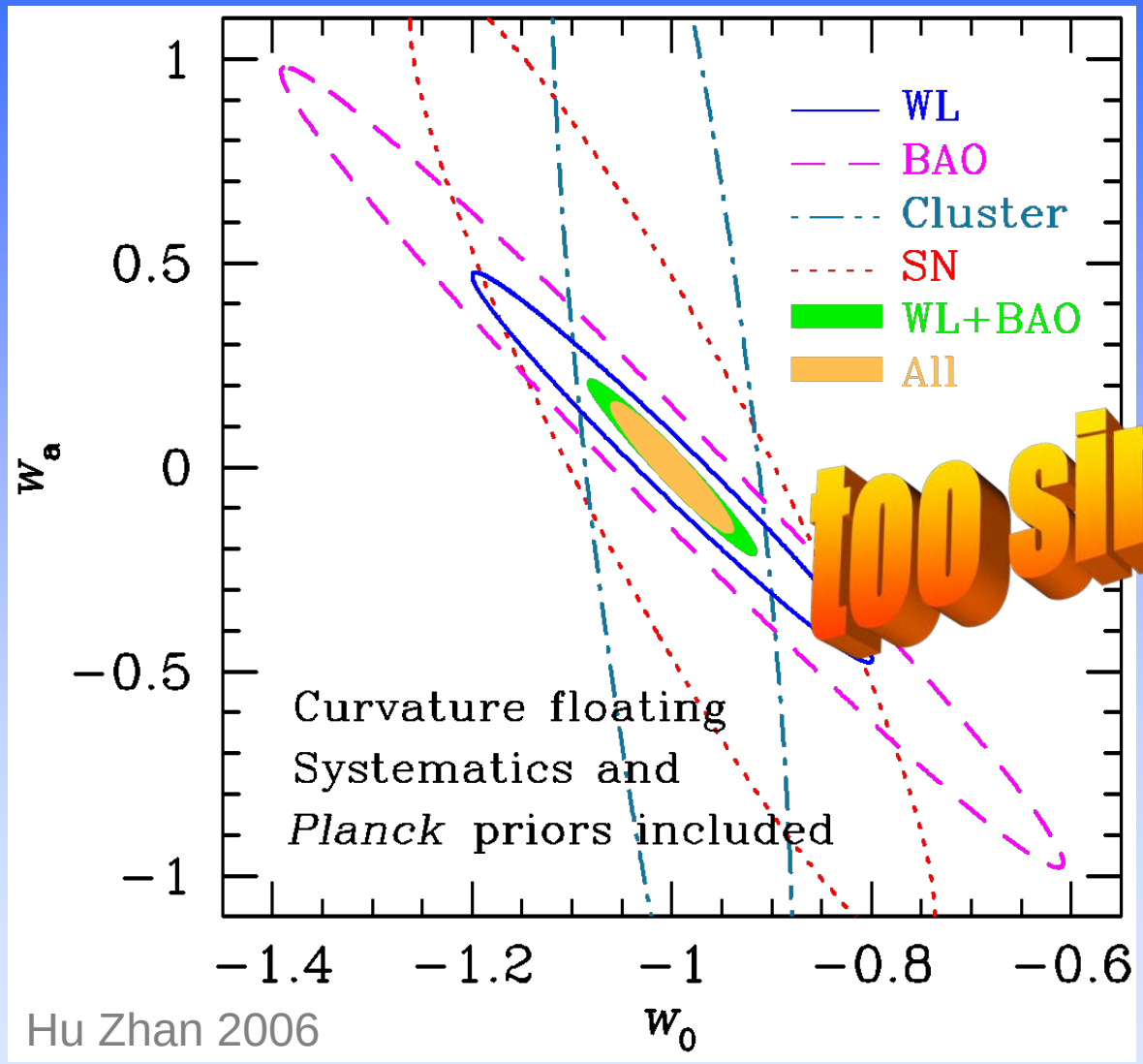
LSST Dark Energy Mission: 5 probes

Precision measurements of all dark energy signatures in a single data set. Separately measure geometry and growth of dark matter structure vs cosmic time by *combining probes*.

- Weak gravitational lensing correlations (over 50 independent lensing probes!)
- Baryon acoustic oscillations (BAO)
- Counts of dark matter clusters
- Supernovae to redshift 0.8
- Probe anisotropy

Combining four LSST probes

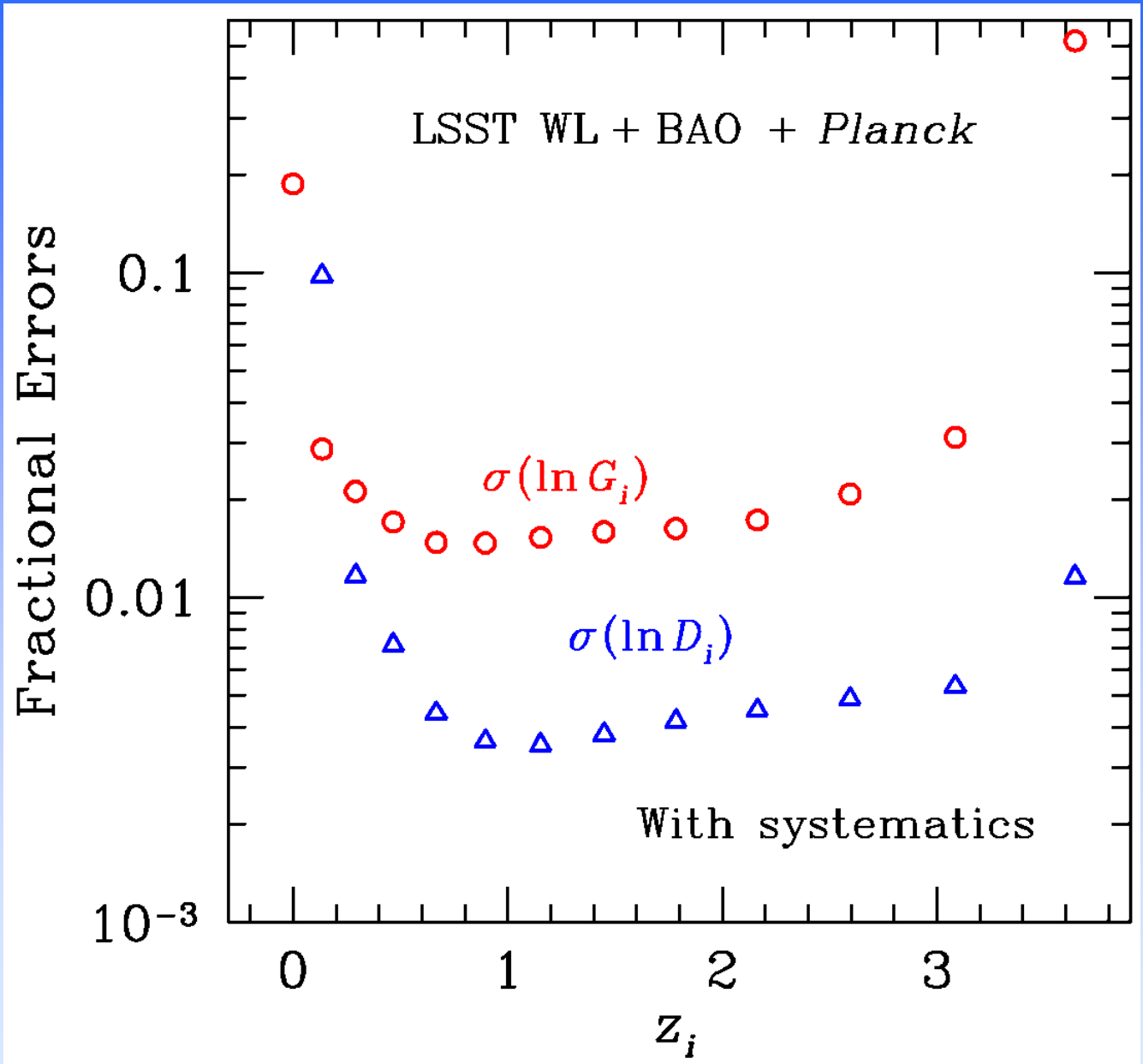
$p/\rho = w_0 + w_a (1-a)$



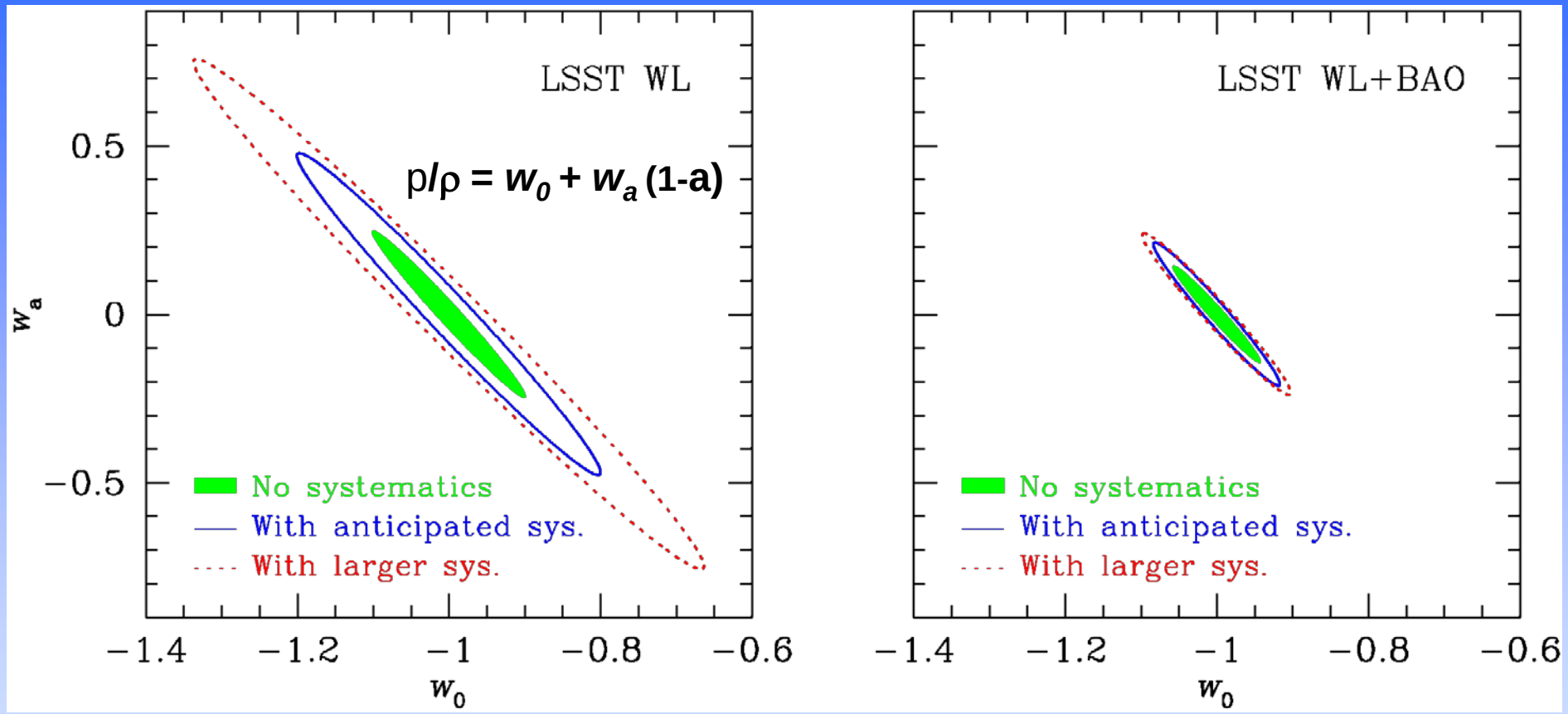
too simple

Hu Zhan 2006

Testing general models of dark energy



sensitivity of WL+BAO to systematic error



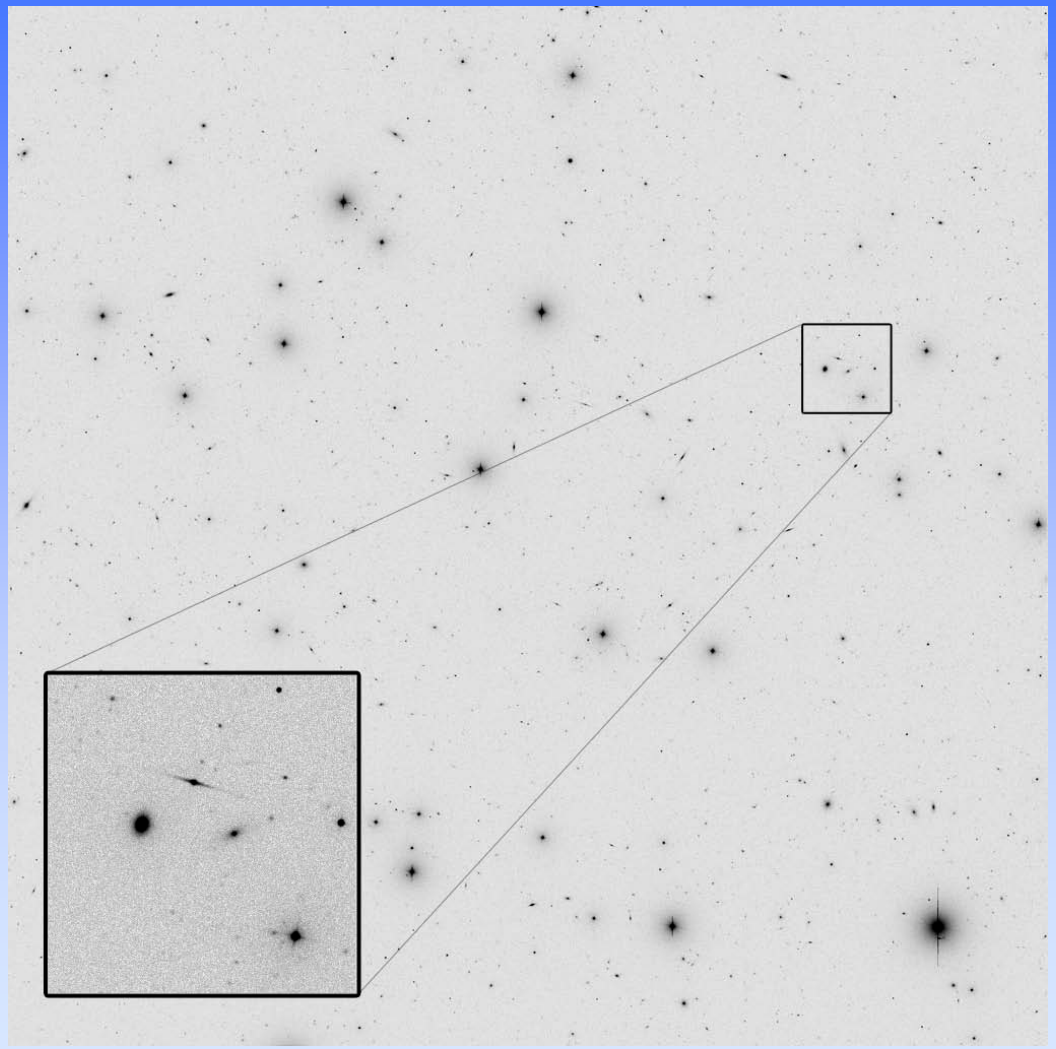
Combining WL and BAO breaks degeneracies.

Joint analysis of WL & BAO is far less affected by systematics.

LSST Survey

- One 6-Gigabyte image every 17 seconds
- 30 Terabytes every night for 10 years
- 200-Petabyte final image data archive anticipated
- 20-Petabyte final database catalog anticipated
- 20 Trillion photometric measurements of 20 billion objects
- Repeat images of the entire night sky every 3 nights
- *~1000 deep images of each of 10 billion galaxies*

Image simulation: validate system performance

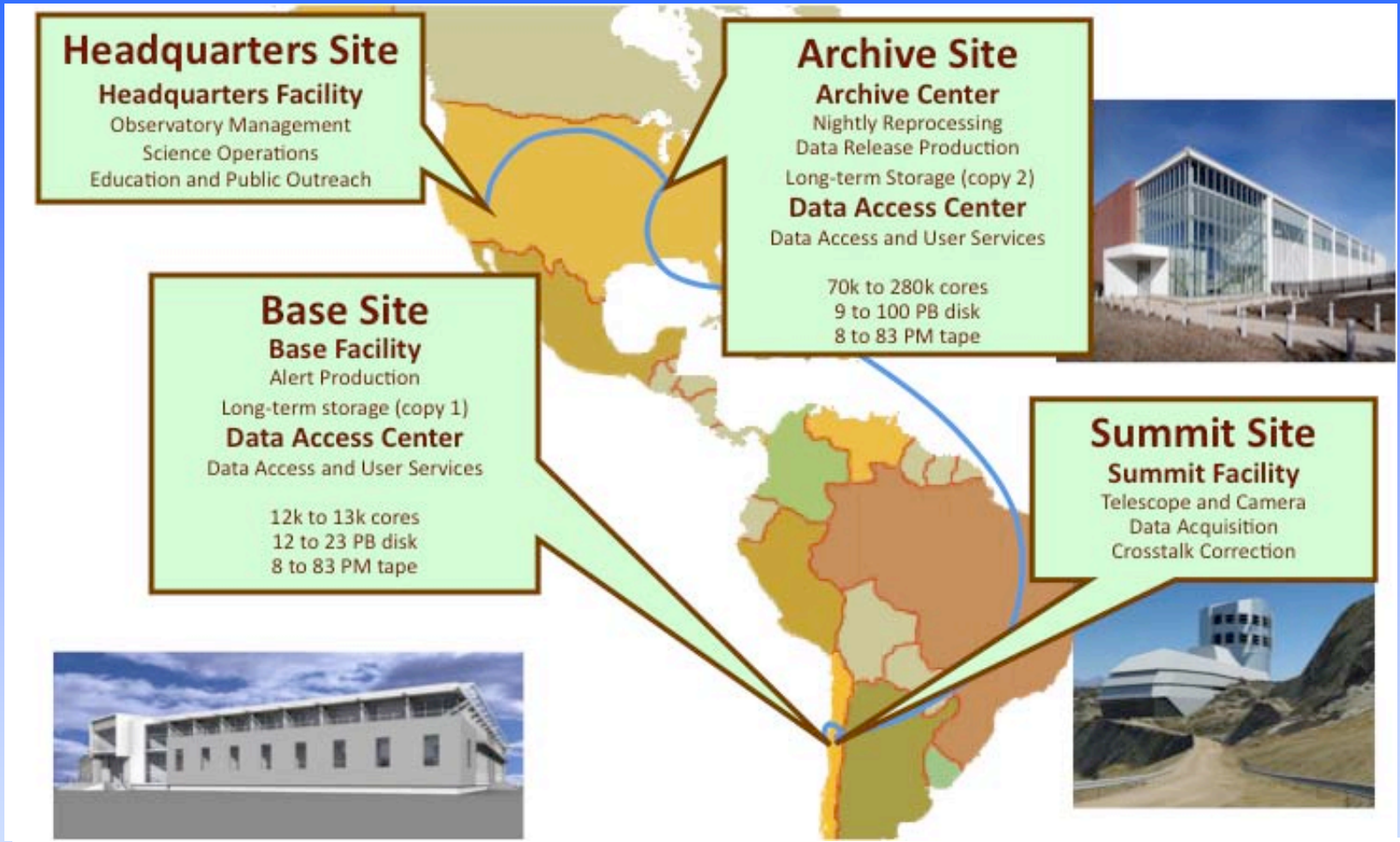


LSST is simulating image performance from end-to-end, including:

- **Cosmology + Milky Way + Solar System**
- **Atmosphere: Turbulence, Refraction, Motion**
- **Optics: Telescope + Camera**
- **Focal Plane: Detectors + Electronics**

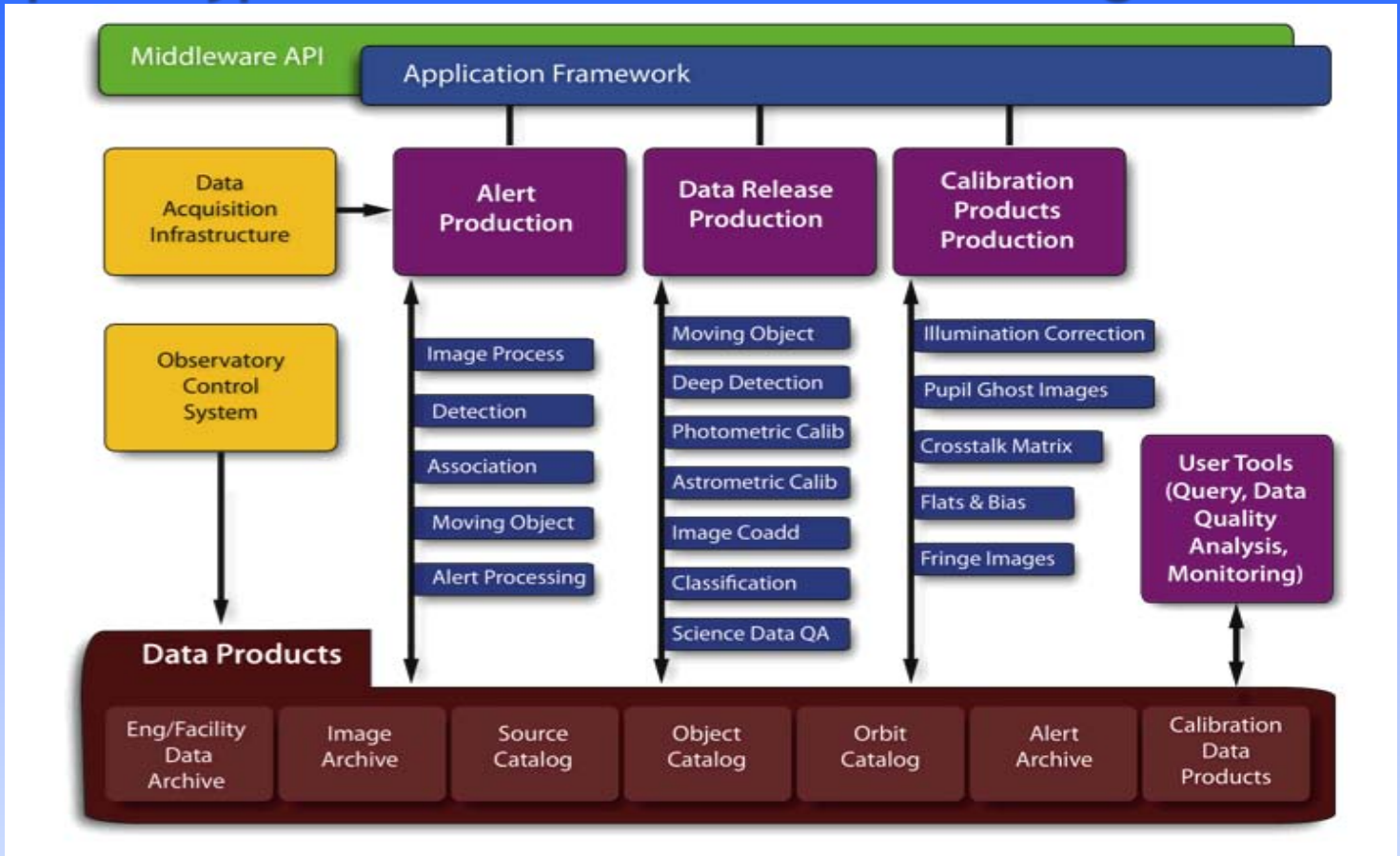
Image simulations are used during all phases of the project

Cyber infrastructure defined and capacity identified to handle data volume



- Summit-Base network will be installed by the project.
- Working with network consortia on capacity.
- International protected network identified and quoted.

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



- **Ongoing Data Challenges.**
- **Unique database design to address multiple trillion row data sets.**

Drivers for Computational Challenges

- **The cross-match problem (multi-λ, time domain)**
- **The distance problem, photometric redshifts**
- **Spatial correlations (auto, cross, higher order)**
- **Outlier detection in many dimensions**
- **Statistical errors vs systematics**
- ***Inverse problem: galaxy shapes***
- **Comparing observations to models**
- **Repeated end-to-end simulations**
- **Scalability!!!**

Enabling Data – to - Science

- 10^3 observations each of $2x10^{10}$ objects
- Each object has ~100 attributes
- Frequent reprocessing
- Processing speed: 1000 TF, sustained
- Search for correlations: many dimensions
- Automated quality assessment
- *Novel database SciDB*

Image Analysis Processing for Dark Energy

Inverse problem:

Use all exposures on a galaxy to measure its shape.

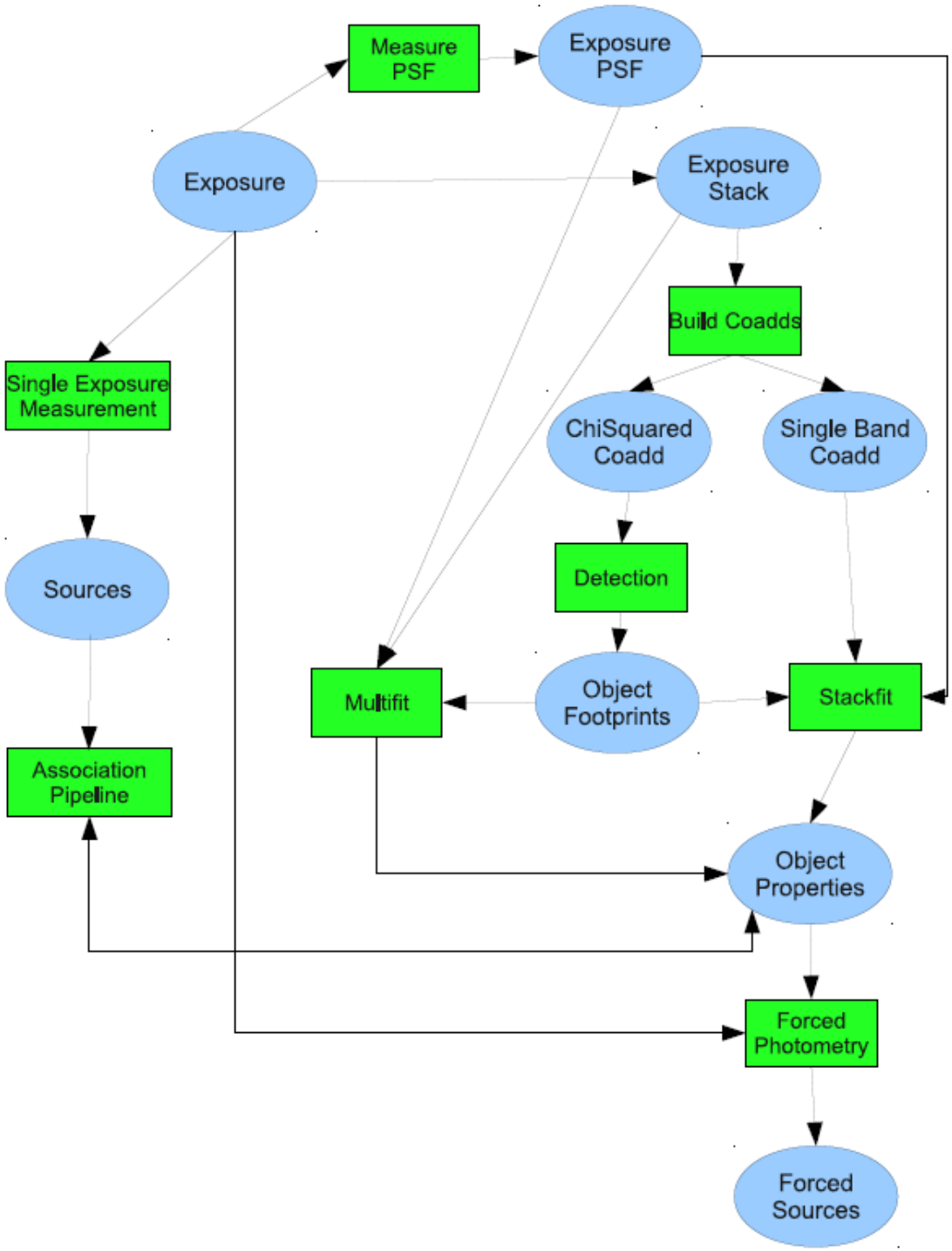
Estimate requirements for analysis of 4 billion galaxies.

- **MULTI-FIT:**

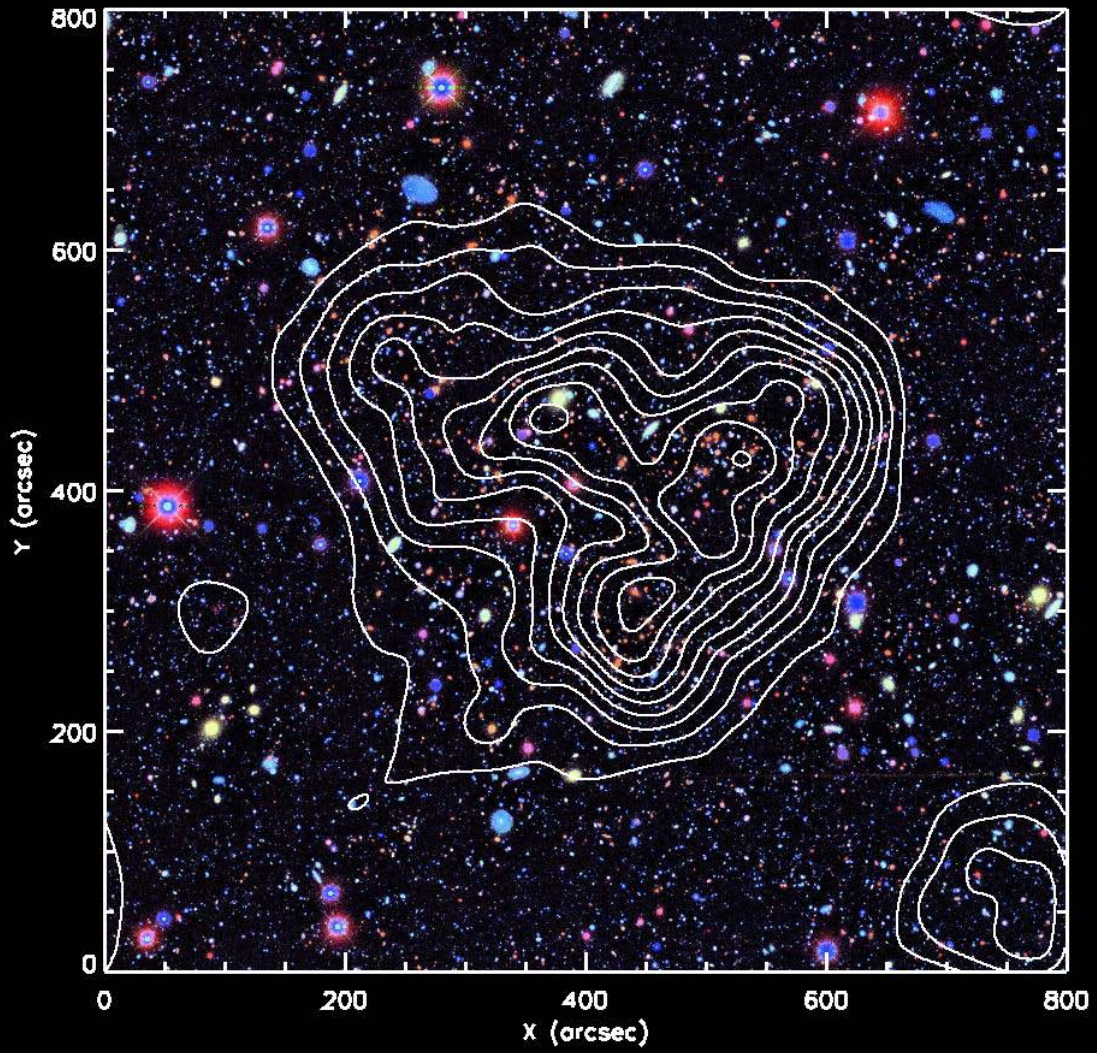
~10^22 Flops / run

- **STACK-FIT:**

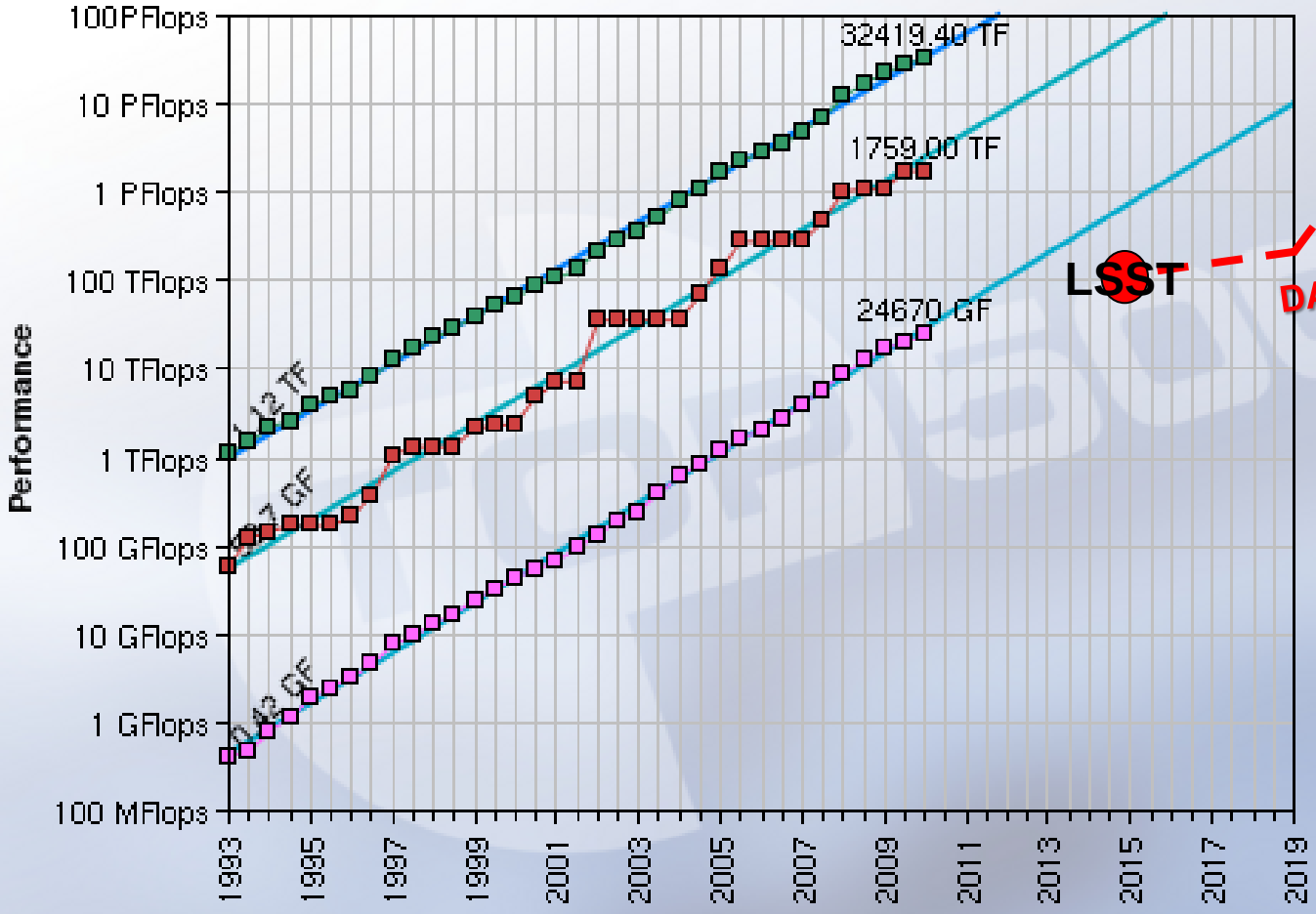
~10^19 Flops / run



Gravitational lensing dark matter reconstruction



Projected Performance Development



LSST Cosmology Computation

- Production of Data Products during operations phase will not be challenging in CPU, but disk I/O could be an issue for some deep co-add calculations.

- Automated data quality assessment and validation is key. Spinoff to other fields of science and engineering.

- Mining science from the database (if further pixel math is required) will be challenging, especially for Dark Energy. Even high dimensional statistical correlations over the database alone can be expensive in CPU and I/O.

- An LSST Dark Energy Experiment Collaboration is being set up. Computational challenges in comparing with theory must be addressed.

Science at the Limit

- Much of the breakthrough science using surveys occurs at the *limits* of the surveys

  - Noise, Sample incompleteness
  - Subtle systematic errors

Data exploration and statistical studies must correct for these errors