LSST and BigPanDA

Erin Sheldon

Brookhaven National Laboratory

March 31, 2017

Large Synoptic Survey Telescope (LSST)

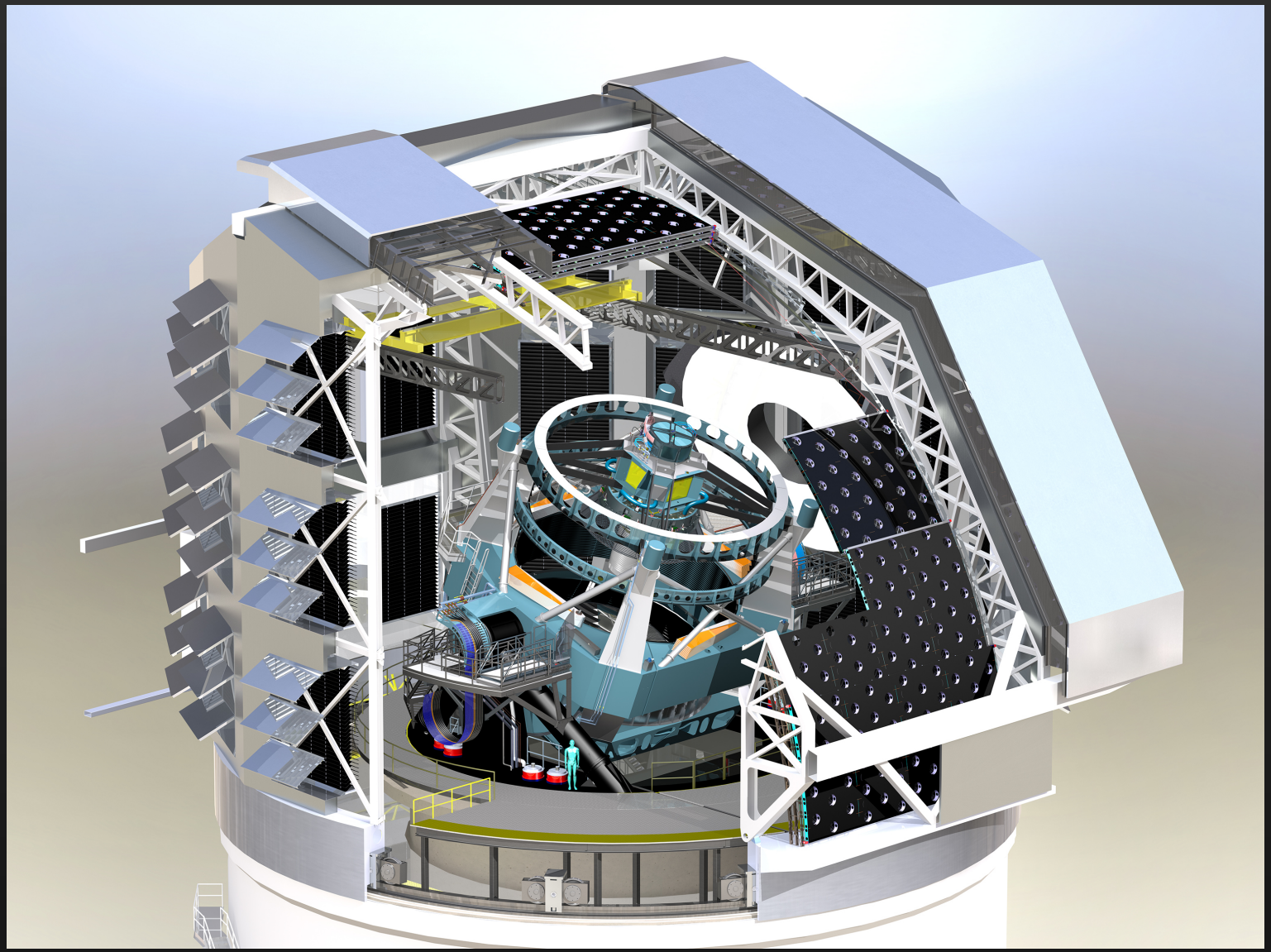
- LSST Intro
- Dark Energy Science Collaboration
- Some computing we would like to do with

Example Image to LSST Depths (from HSC)

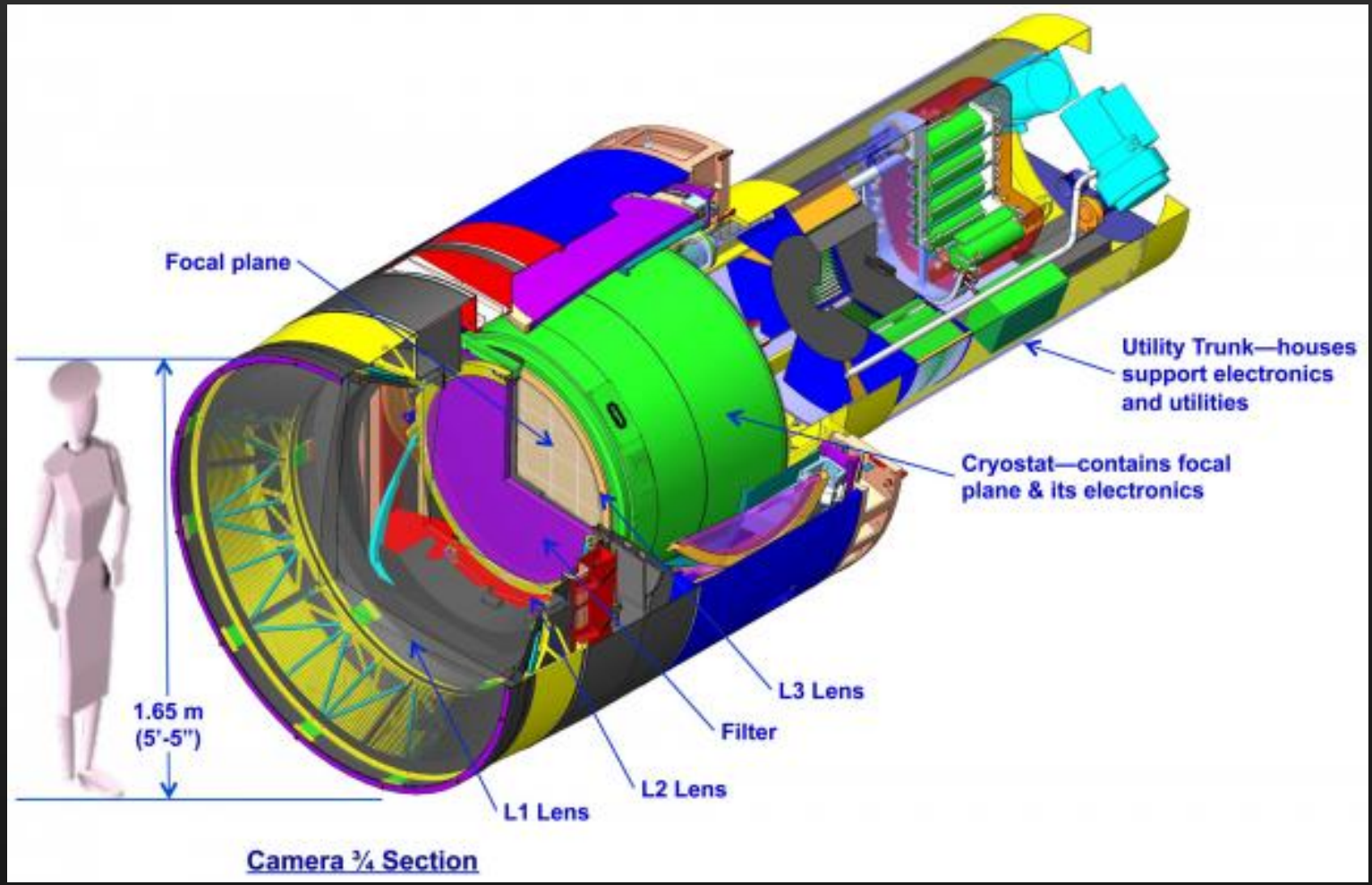
LSST

- Huge astronomical survey covering half the sky
- Diverse range of science: from cosmology, supernovae, milky way science, solar system science
- Time domain component: revisit each part of the sky of order 1000 times.
- Biggest camera ever made: 3.2 gigapixels. BNL leads the development of the CCD sensors.
- Will generate 60 PB of imaging data.

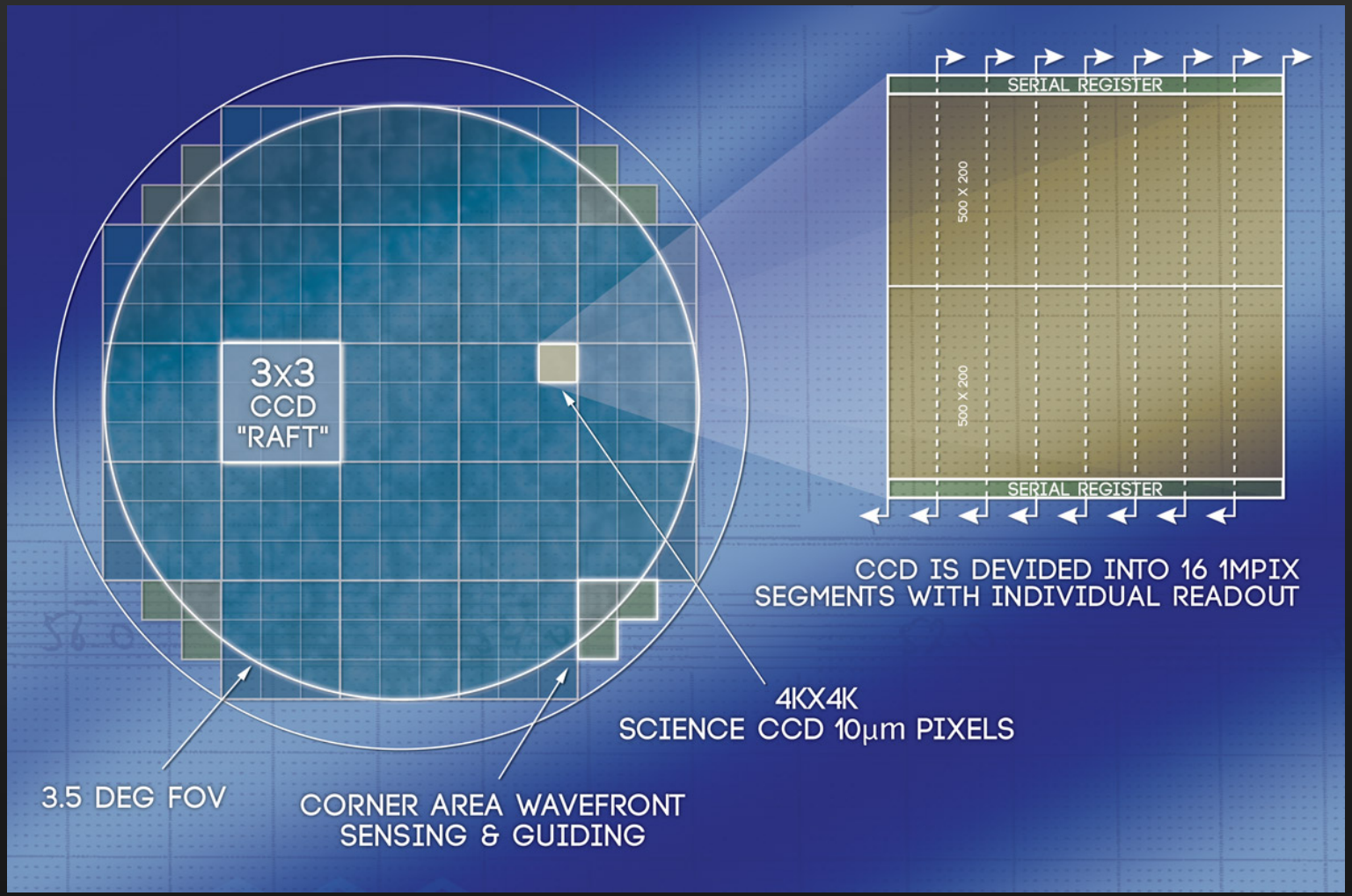
LSST Camera



LSST Camera


Focal plane
Utility Trunk—houses support electronics and utilities
Cryostat—contains focal plane & its electronics
L3 Lens
Filter
L2 Lens
L1 Lens
1.65 m
(5'-5")
Camera ¾ Section

LSST Camera


3x3
CCD
"RAFT"
SERIAL REGISTER
500 X 200
500 X 200
SERIAL REGISTER
CCD IS DEVIDED INTO 16 1MPIX
SEGMENTS WITH INDIVIDUAL READOUT
4KX4K
SCIENCE CCD 10μm PIXELS
3.5 DEG FOV
CORNER AREA WAVEFRONT
SENSING & GUIDING

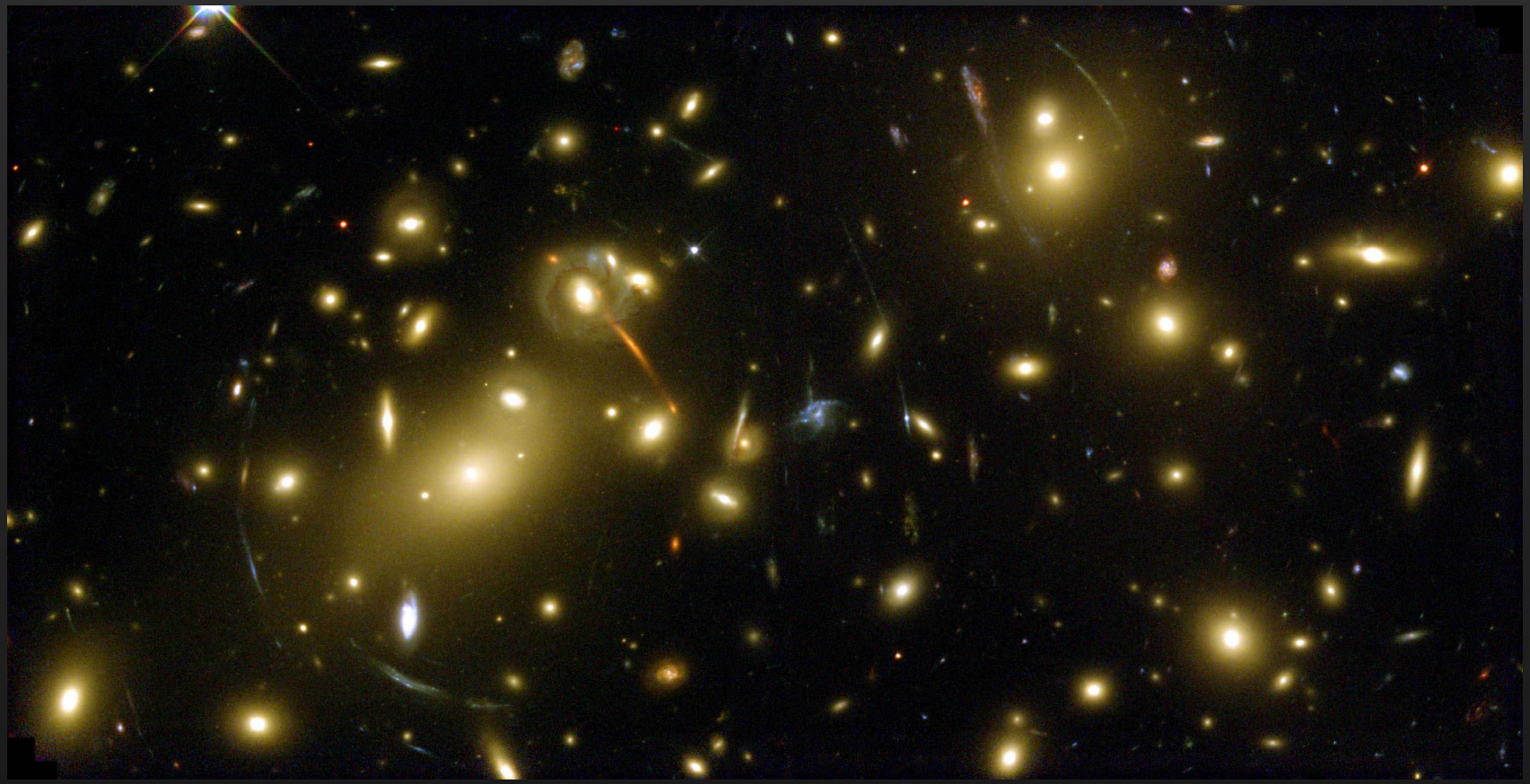
Dark Energy Science Collaboration (DESC)

- Collaboration focused on Cosmology and Dark Energy from LSST
- Why is the expansion of the universe accelerating?
- Primary probes
  - Weak Gravitational Lensing and Large Scale Structure
  - Supernovae
  - Clusters of galaxies
- The main challenges involve image processing and simulations

Weak Gravitational Lensing

- Light is deflected as it passes massive objects, such as galaxies.
- The deflection distorts the images of galaxies behind the lens in a coherent way
- Can be used to measure the mass of objects in the universe, including Dark Matter
- Can be used to measure the effect of Dark Energy

Strong Lensing



Weak lensing you can't see by eye, but it can be measured anywhere.

Computing Challenges for Weak Gravitational Lensing

- The lensing effect changes the ellipticities of galaxies in a coherent way. Basic task is to measure ellipticities.
- There are many things that can bias the measurement: smearing by the atmosphere, optical distortions, noise, overlap of light from nearby objects.
- These are image processing challenges.
- We don't know how to do all these yet at the level required (0.1%!). This is still in research!
- Yet the success of LSST is depends on accurately measuring weak lensing.

Computational Challenges

- Once algorithms are finalized, run on 60PB of images.
- But the algorithms are still in development.
- Testing the algorithms requires a huge amount of computing, because most of the tests need to be done on the scale of the real survey.
- Testing also means running on the real data multiple times until systematic errors are reduced to required level
- Key Point: We need a computing environment that supports development with quick feedback loops, as well as production runs, all with high throughput (basically the ATLAS/RHIC system at BNL is a perfect example).

Computational Challenges

- The DOE currently wants us to utilize NERSC
- NERSC is optimized for HPC: Massive production runs using proven codes. Grab 100,000 cores in parallel.
- NERSC is a poor environment for our type of work and development. Even getting codes compiled requires expertise. Getting running efficiently is a signifcant challenge even for experts
- You need to anticipate your work long in advance, submit job for huge number of cores, wait a week for the job to start, then once the job starts run your own queue on top of that job to spinoff each of the 1-core jobs.
- We need an automated system to make it easy to do development and production work without all this mental overhead.
- And remember we are a very diverse collaboration with many different code bases, science goals, and little HPC computing expertise.

Current Status with PanDA

- Heyun Park and I have begun working with Sergey and Pavlo at BNL.
- I have been able to run simple simulation jobs using PanDA.
- Would like to get running on OSG since those are standard redhat systems. Will take some learning curve to get started, but then developement work should be straightforward.
- Begin with simulations and processing data from pre-cursor survey DES (Dark Energy Survey).
- Eventually we may need work at NERSC through PanDA. Still need to compile codes on that system, and deal with non-standard environement, but hope PanDA can simplify the work flow.

Time Scale

- We will have “first light” engineering data in 2019, science data in 2021.
- We would like to have a standard system in place well before that, a system that makes it transparent to work on a variety of systems: OSG, NERSC, whatever is available.
- My goal would be to move all my processing to PanDA in a year. Would take longer to get others on board.
- Short term, process data from current surveys through the system. Process LSST image simulations.
- Long term, process real LSST data.