Phase B WP 3.10: DESC Operations

Joe Zuntz
with James Perry

Dark Energy Science Collaboration

- DESC collects all cosmology research in LSST

- Weak & strong lensing, clustering, galaxy clusters, type IA SNe, theory.

- Associated simulations, requirements, software, infrastructure

- Mostly self-contained scientifically, though overlap with Transients and Statistics collaboration

- Not the same as DES, (Dark Energy Survey), or DESI, (Dark Energy Spectroscope Instrument)!

DESC Operations

- Branding exercise because DOE understands the concept of operations

    - But doesn't really appreciate how astro projects work, or the fuzzy boundary to software needs

DESC Operations

- Various DESC infrastructure projects need:

    - Domain-specialist expert scientists:
      Pipeline infrastructure personnel

    - HPC expertise:
      Computing infrastructure personnel

- DESC Operations Plan describes needs in full, and Operations Committee manages personnel
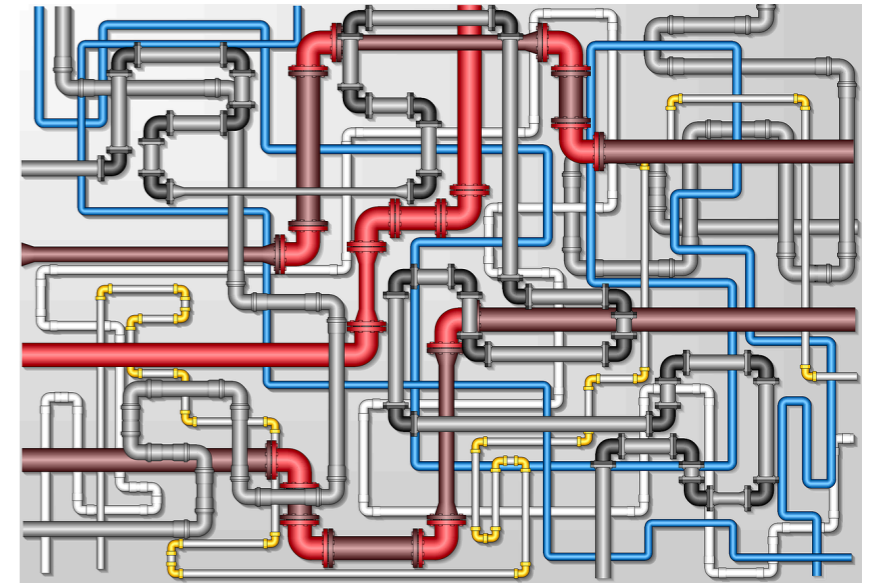
Collaboration Needs

- Pre-commissioning pipeline construction

- Testing that pipeline on simulations
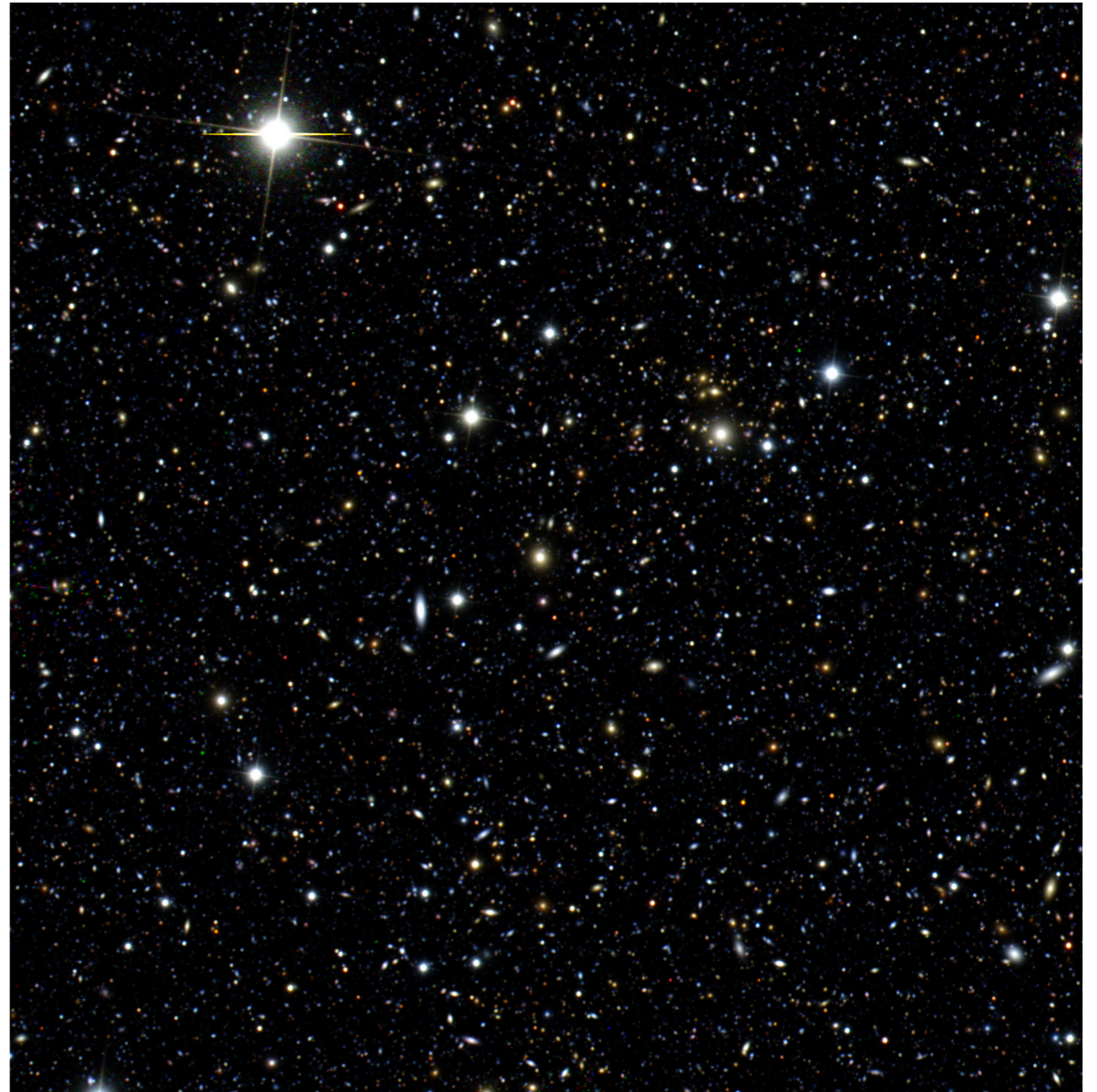
UK Incentive

- Leadership in key science areas

- Pipeline development for UK-specific science objectives

- Contribution is expected by collaboration as part of UK involvement

    - ODF??

3.10.1
# Pipeline Scientist
Joe Zuntz

Pipeline Science Contributions

- Lensing + clustering cosmology pipeline

- Infrastructure for WL catalog-to-cosmology pipeline

- Cosmological Parameter Estimation

# Lensing & Clustering Pipeline Observables

- Positions

  - Seek sample that is consistent and strong tracer of galaxy halos e.g. LRGs

- Shapes

  - Distant galaxies as backlights

  - Rays trace through gravitational fields

  - Measure two ellipticity values, e_1 and e_2
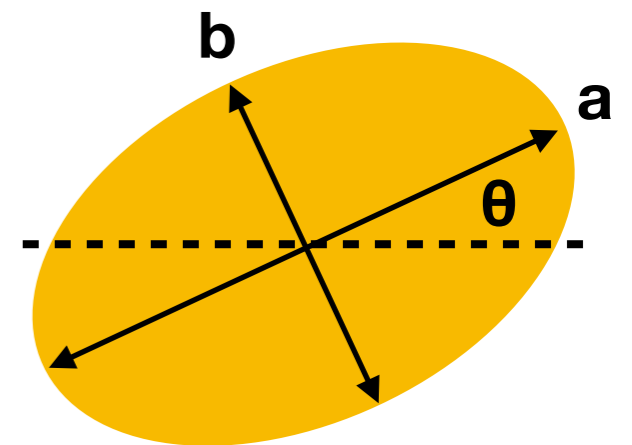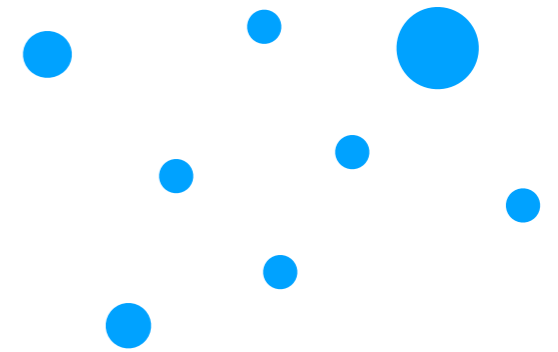
$$e_1 + ie_2 = \frac{a - b}{a + b} e^{2i\theta}$$

3 x 2pt Correlations

- w(θ)
  Galaxy density (position) correlation function

- $\xi_+(\theta)$, $\xi_-(\theta)$
  Shear (shape) correlation functions

- $\gamma_t(\theta)$
  Shear around lens galaxies (galaxy-galaxy lensing)
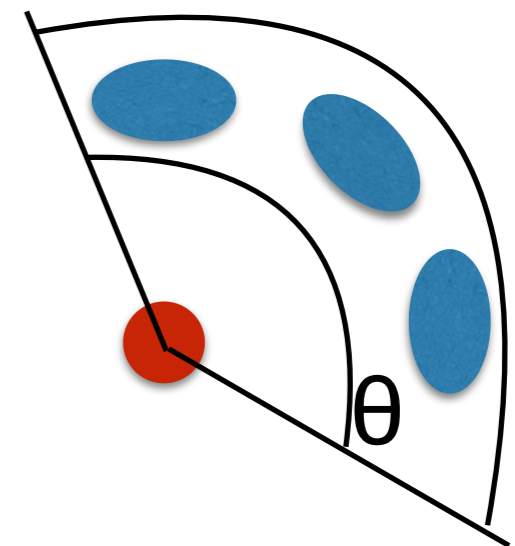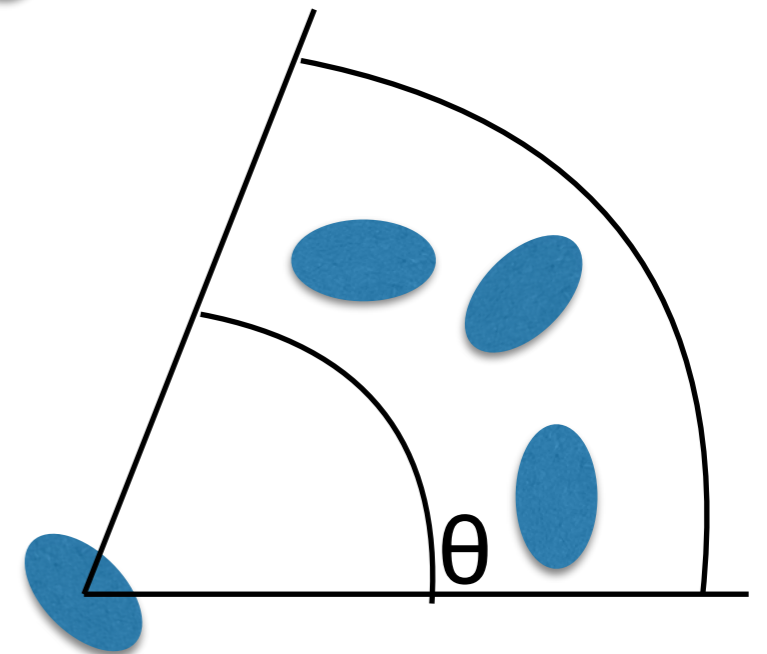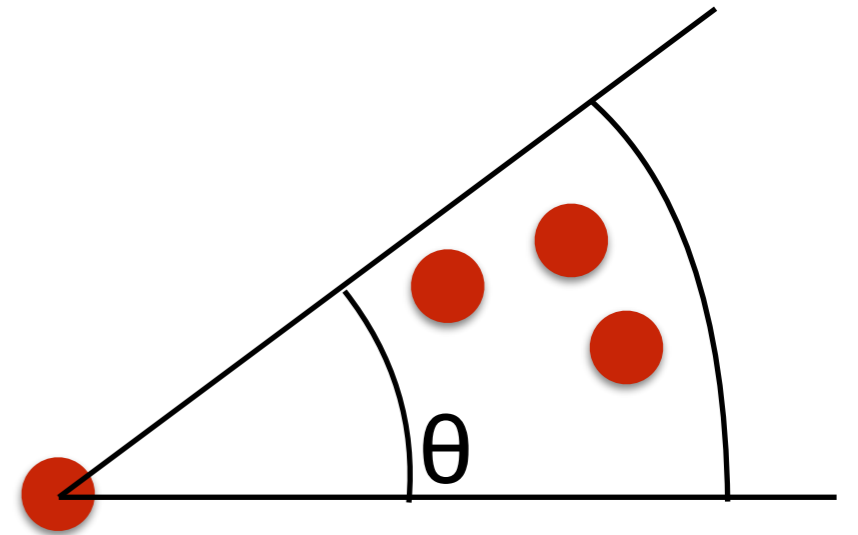
3 x 2pt Pipelines

- Many different analysis stages between catalog and cosmology

- Each stage is research problem in itself

- Most are also HPC problems

- Collecting and combining them into a coherent pipeline is a key infrastructure challenge

    - Traditional approach is somewhat incoherent
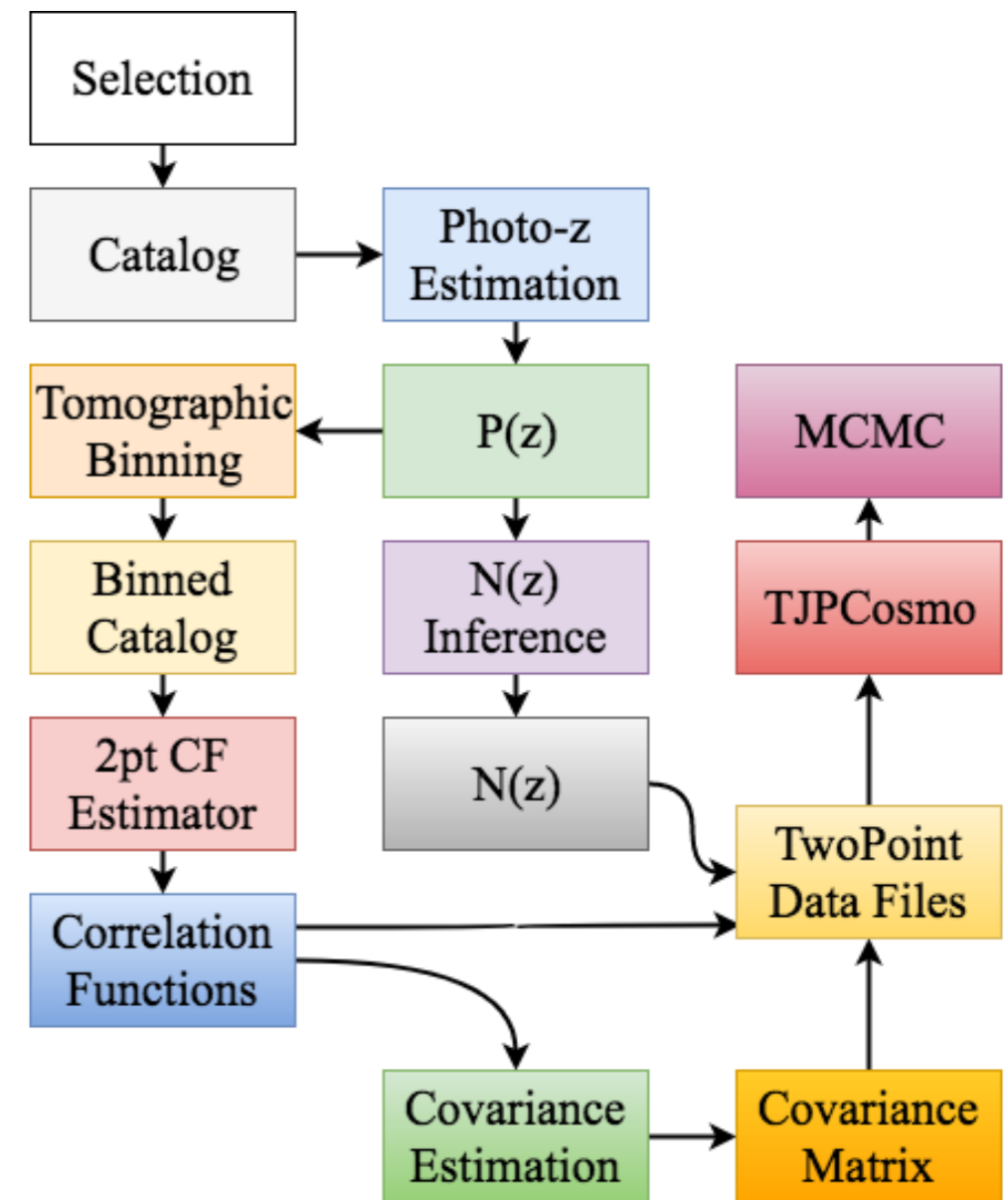
Pipeline Infrastructure Tools

- Parsl: parsl.readthedocs.io

  - External, authors working closely with DESC

  - Workflow management: launches jobs, checks inputs/ outputs, logging,

- Ceci: ceci.readthedocs.io

  - Developed for DESC

  - Connects DESC tasks as python classes with pre-defined structure to Parsl / other workflow tools
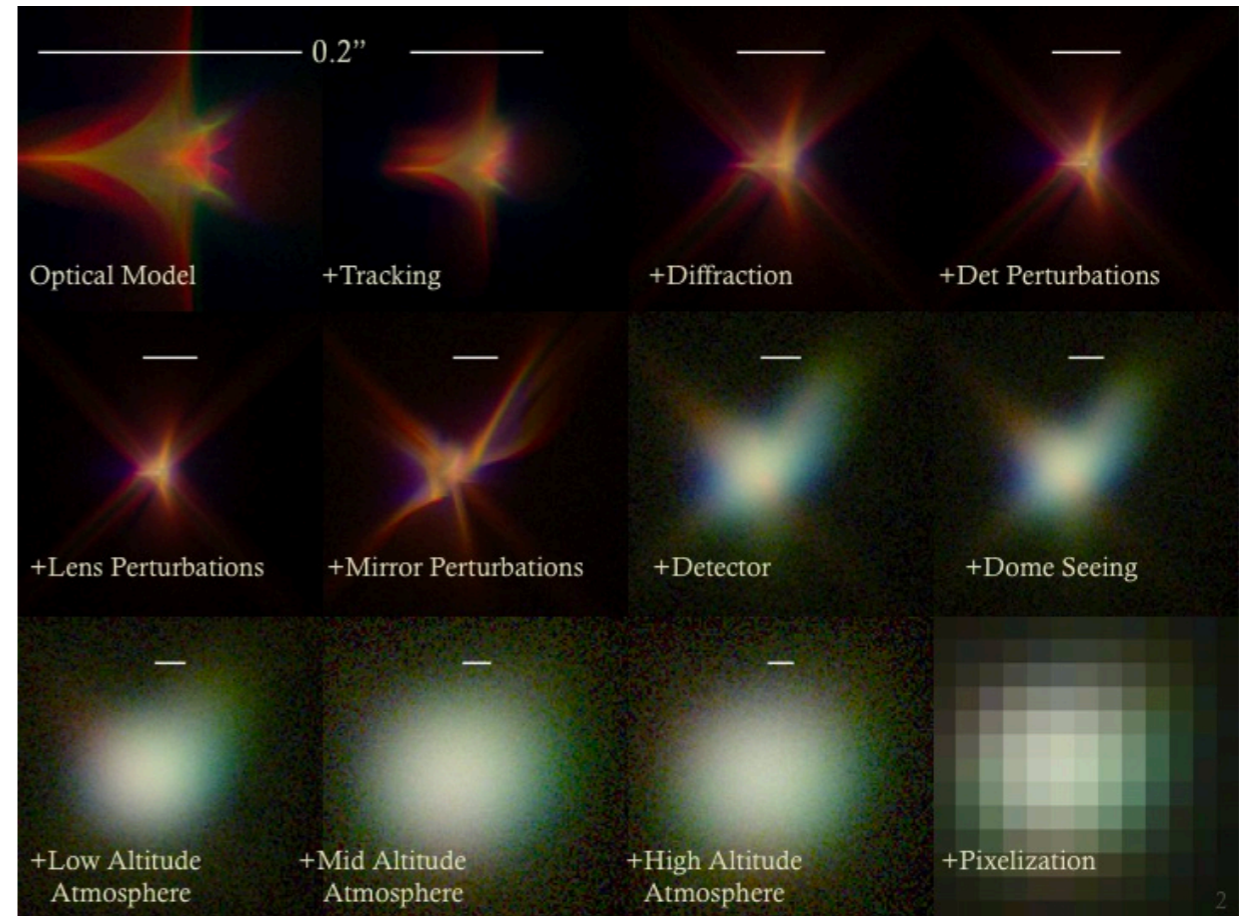
3.10.2
Data & Sim Wrangling
James Perry

Data & Simulations Wrangling

- Major series of simulations being generated by DESC: Data Challenges 1, 2, 3

- Loads of infrastructure and operations work needed:

- Writing and optimizing sim s/w

  - ImSim / PhoSim

- Running production at scale

- Storing, moving, & documenting generated data

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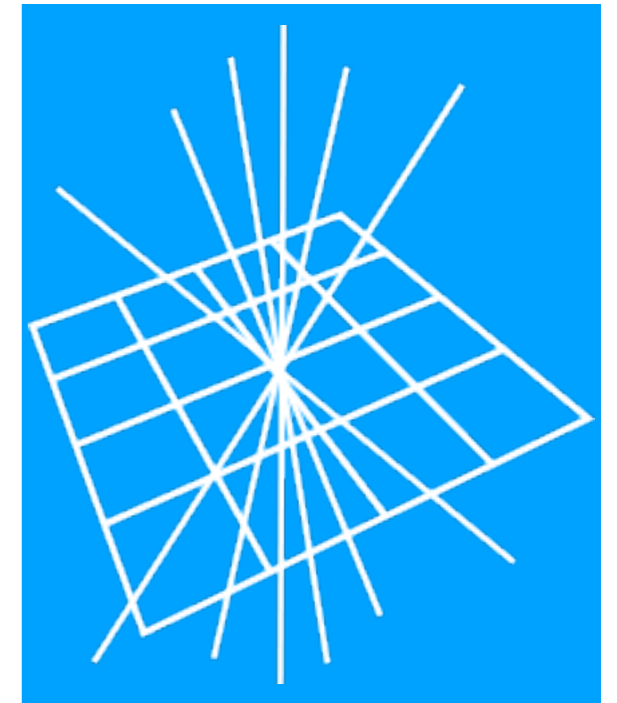
0.2"

Data & Simulations Wrangling

- Also need to run analysis pipeline on these simulations

- Data Management Stack

  - Behemoth low-level processing pipeline

  - Significant expertise needed!

  - Data butler

  - Crucial for UK projects needing specialist reduction / analyses

- Higher level processing

GridPP

- As discussed yesterday, we are using the grid via the UK *GridPP* organisation

  - Great computing power available, but not straightforward to access

  - Access, data & code management, submission, and monitoring all difficult

  - Being worked on by James - potentially useful to other LSST:UK people

Other Inter-Collaboration Possibilities

- Simulations / software useful elsewhere?

  - Mixed single-exposure and catalog-only sims

  - 500 sq deg of 10-year images

  - 5000 sq deg of catalogs

- Tools useful elsewhere?

  - Pipeline management for large jobs