LSST Dark Energy Science Collaboration: Data Challenges

Eric Gawiser

Rutgers University

June 2018 TVS Collaboration Meeting

What is DESC trying to learn about the Universe?

We want to understand cosmic acceleration:

- Why is the expansion of the Universe accelerating? How does the expansion work?

What was the expansion history? How did structure grow during the expansion?

Assuming some Dark Energy drives the expansion, what is its equation of state, *w*? And has that varied in time? Is it the same everywhere?

Neutrinos will have played a role: how many species are there, and how much energy do they represent?

Are we right to assume General Relativity is correct? Can we distinguish modified gravity from Dark Energy?

How does LSST provide answers to these questions?

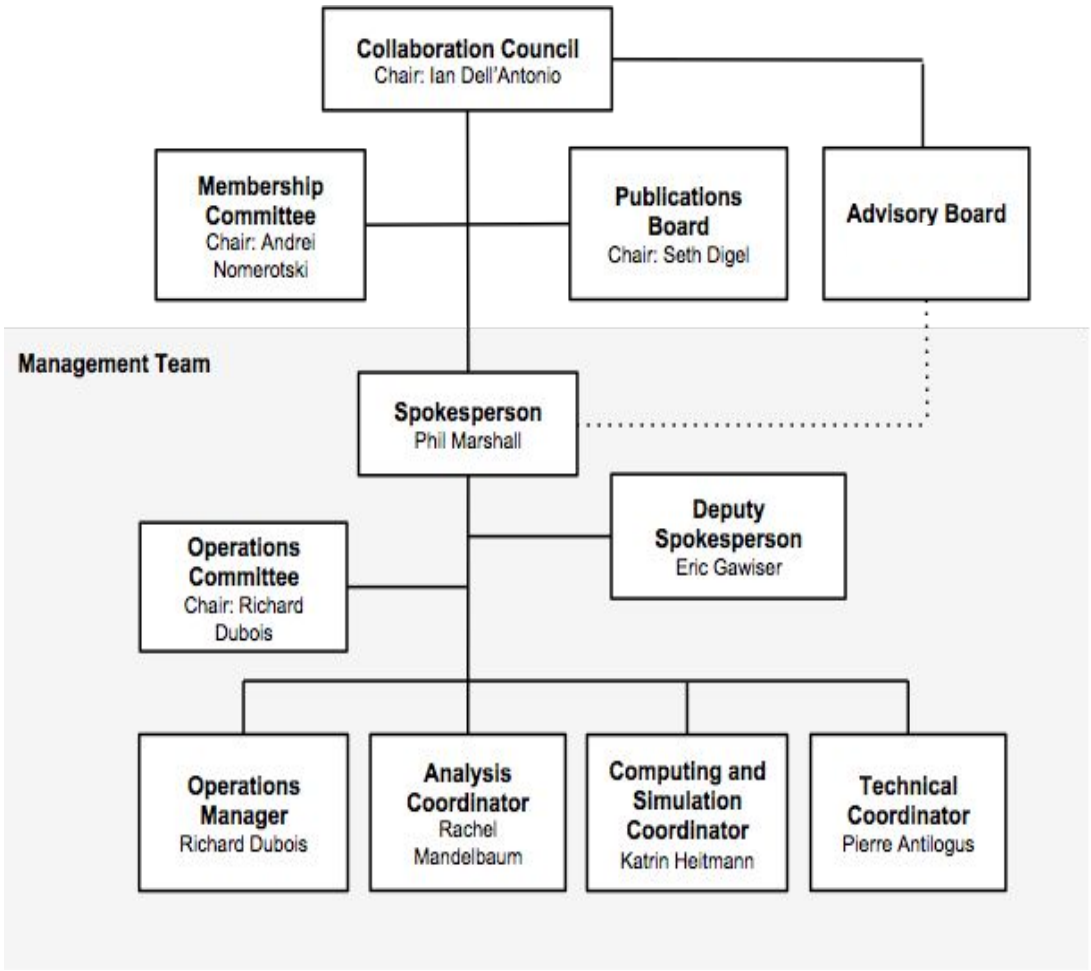
- **Weak gravitational lensing (WL)**, **galaxy clustering (LSS),** and **clusters of galaxies (CL)** all provide a direct probe of the dark matter structures on the largest scales, and how fast they grew: locate galaxies and clusters, measure galaxy shapes in images, compute correlation functions
- **Type Ia Supernovae (SN)**, **strong lens systems (SL)**, **Baryon Acoustic Oscillations** and gravitational wave sources provide standard candles, timers, rulers etc for measuring the expansion rate, independently of structure growth

Meeting LSST DESC Goals

To enable us to make cosmological measurements with LSST data, we need to:

- Develop, validate, maintain and operate a set of simulation, processing and analysis software pipelines and infrastructure
- Re-process and analyze ~petabytes of LSST images and catalogs
- Work efficiently together to perform a complex and challenging joint inference
- Feed improved processing algorithms back to the Project/Facility as needed, validated at scale

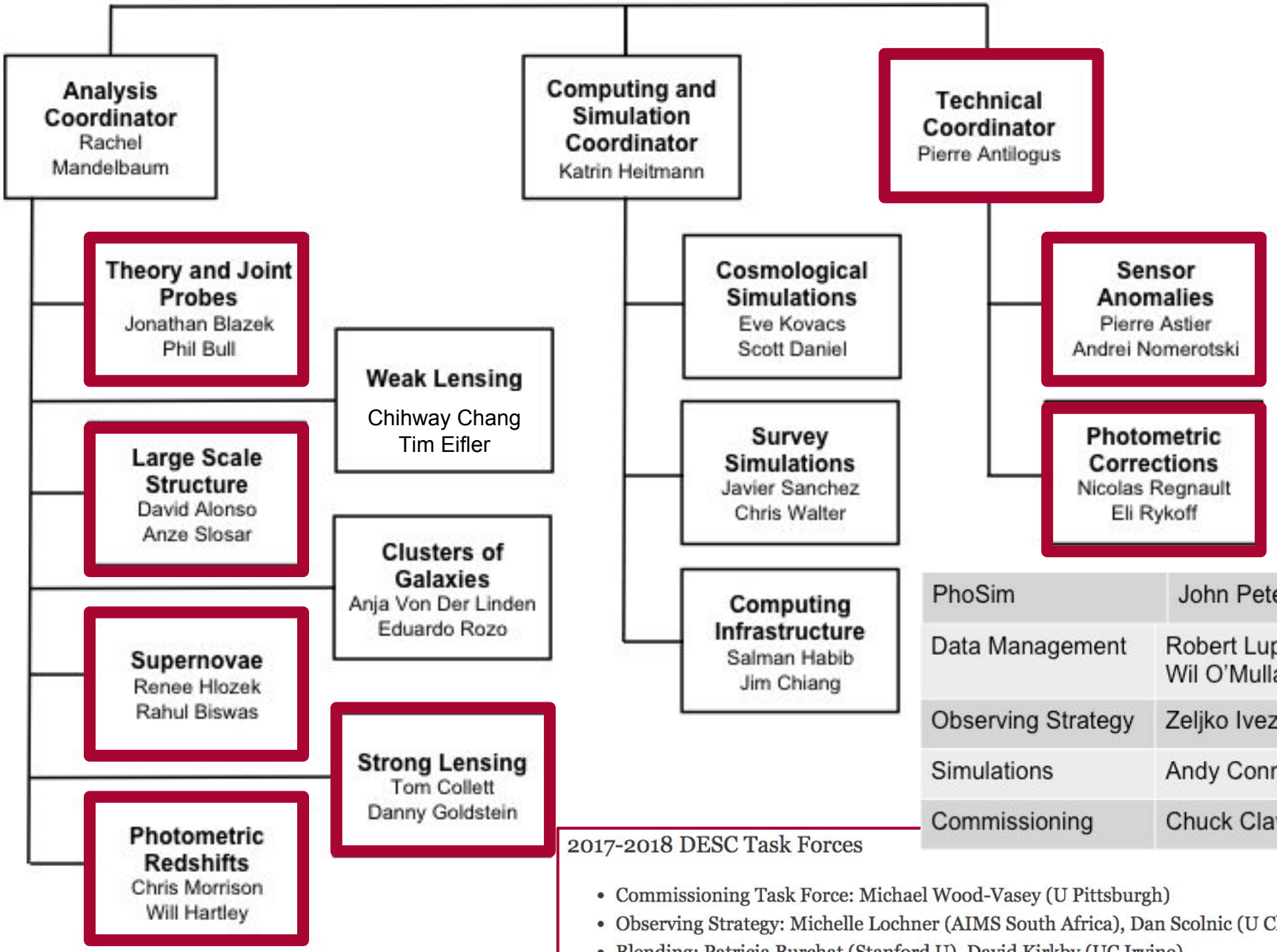
LSST DESC organization



The LSST DESC has a Spokesperson + Council democratic governance model, with biennial Spokesperson elections and annual CC elections.

The Mgmt Team is appointed by the Spokesperson and confirmed by the CC.

Full Members (who have committed a significant fraction of their research time to DESC) can vote. JuDO for junior members; emphasis on diversity in governance; professional conduct is demanded



PhoSim	John Peterson
Data Management	Robert Lupton Wil O'Mullane
Observing Strategy	Zeljko Ivezic
Simulations	Andy Connolly
Commissioning	Chuck Claver

2017-2018 DESC Task Forces

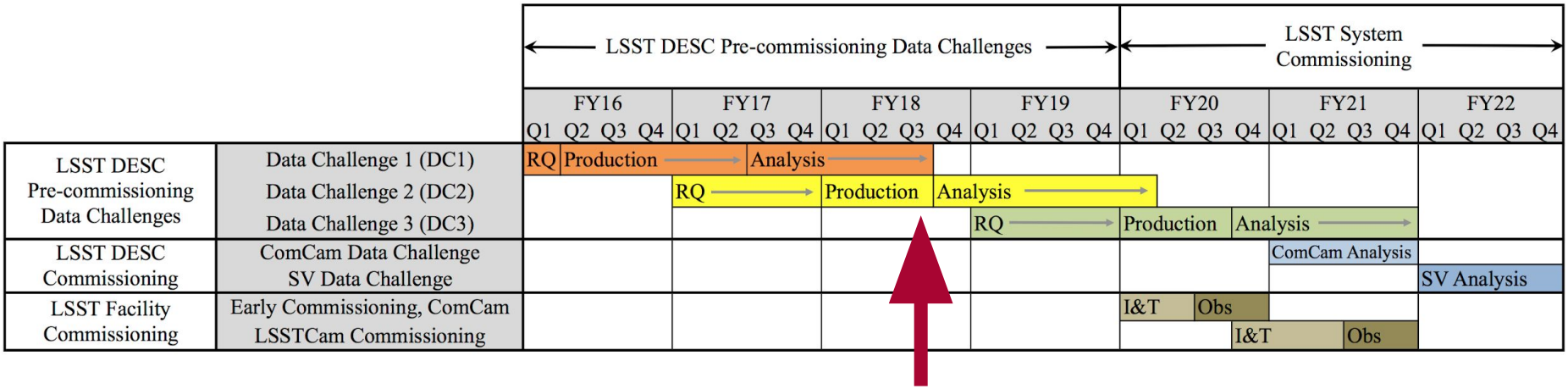
- Commissioning Task Force: Michael Wood-Vasey (U Pittsburgh)
- Observing Strategy: Michelle Lochner (AIMS South Africa), Dan Scolnic (U Chicago)
- Blending: Patricia Burchat (Stanford U), David Kirkby (UC Irvine)

DESC Data Challenges

- Created so the collaboration will be ready for the first LSST data
- 3 sets of Data Challenges scheduled
  - Increase in size and complexity in each iteration
- Data Challenge 1 completed, included 10 years of r-band visits with random dithers covering 40 square degrees

  (60M galaxies, 2M stars)
- Currently in Data Challenge 2 era
  - 300 sq degrees, 10 years, ugrizy
  - Includes variable sources in both WFD and “uDDF” regions
- Data Challenge 3 being defined; will mix analysis of precursor surveys, ComCam and simulated LSST data

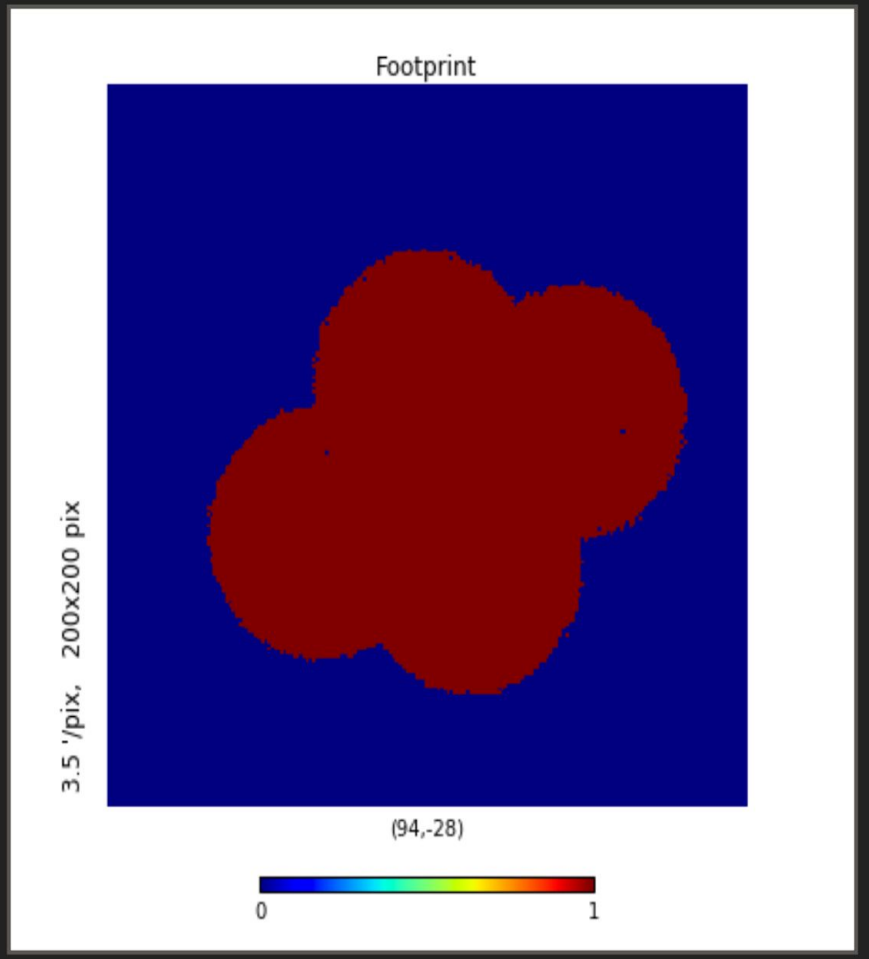
DESC Data Challenges: Schedule



- We are in DC2 production in FY2018 Q3, and the DC1 analysis phase is drawing to a close.
- The DC2 analysis phase will extend through FY19, and *drive development of the prototype analysis pipelines*
- DC3 will be designed during its year-long requirements (RQ) phase, based on a) the DC2 experience and results and b) *the needs of the analysis pipeline development.* The aim is to use this challenge to ready ourselves for the LSST SV commissioning data (which sets an overall scale of ~100k reprocessed visit images in DC3)

DC1 OVERVIEW: INPUT OBSERVING STRATEGY

- OpSim: minion_1016_new_dithers.
- Dithering strategy by Humna and Eric. Trimmed chips out of the field.
- 10 year visits. r-band only.
- The total area of the field is ~40 sq-deg.
- ~ 180 visits per coadd



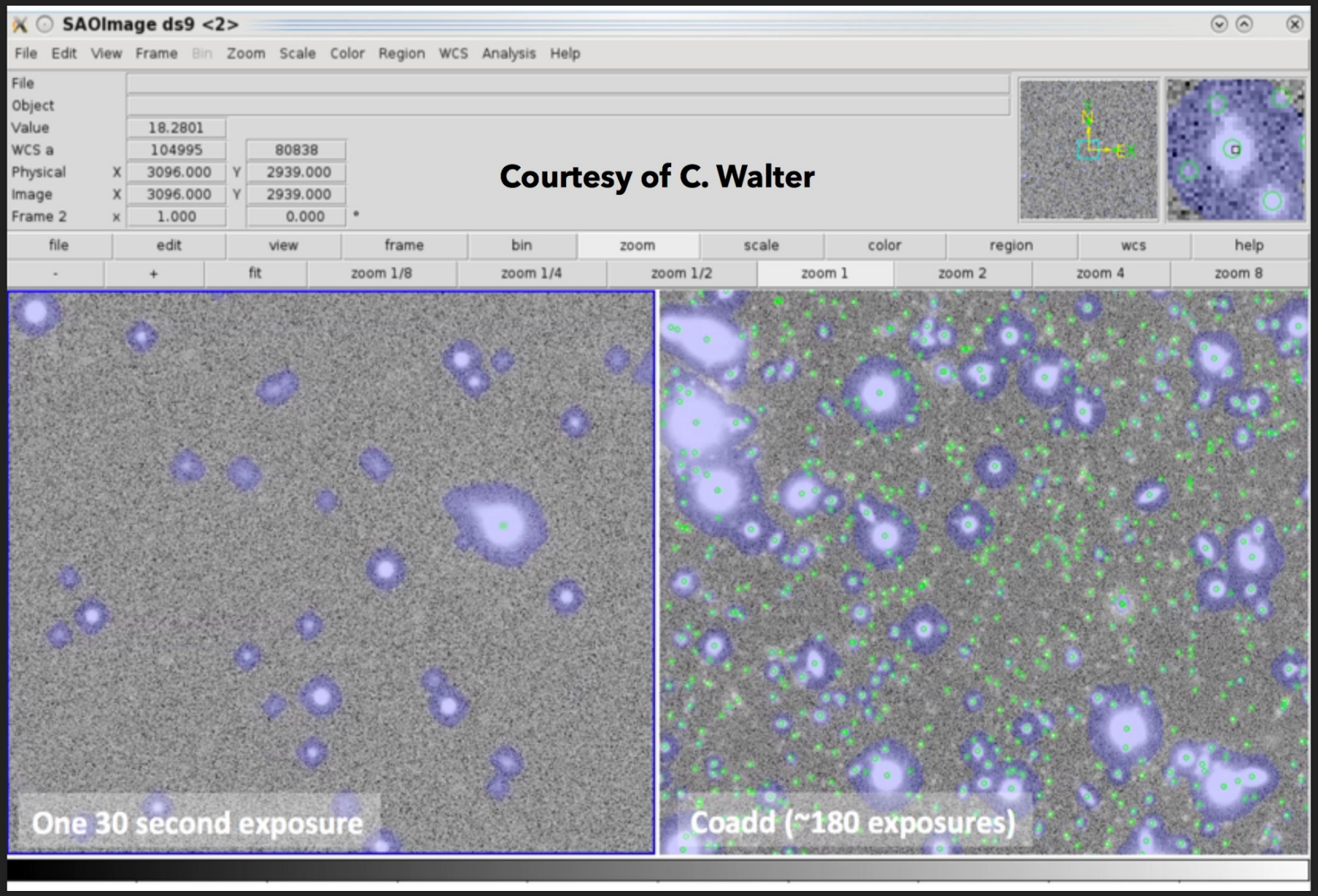
Slide from Javier Sanchez

DC1 OVERVIEW: SIMULATION RUNS AND REDUCTION

- imSim + dithering
- imSim + no dithering (undithered)
- PhoSim + dithering (~55 day run/ 24 M CPU hours 189,354 images)
- Reduction by DM stack: reduced single exposures, co-adds and catalogs generated
- Each dataset ~75 TB
- Data available at NERSC

Slide from Javier Sanchez

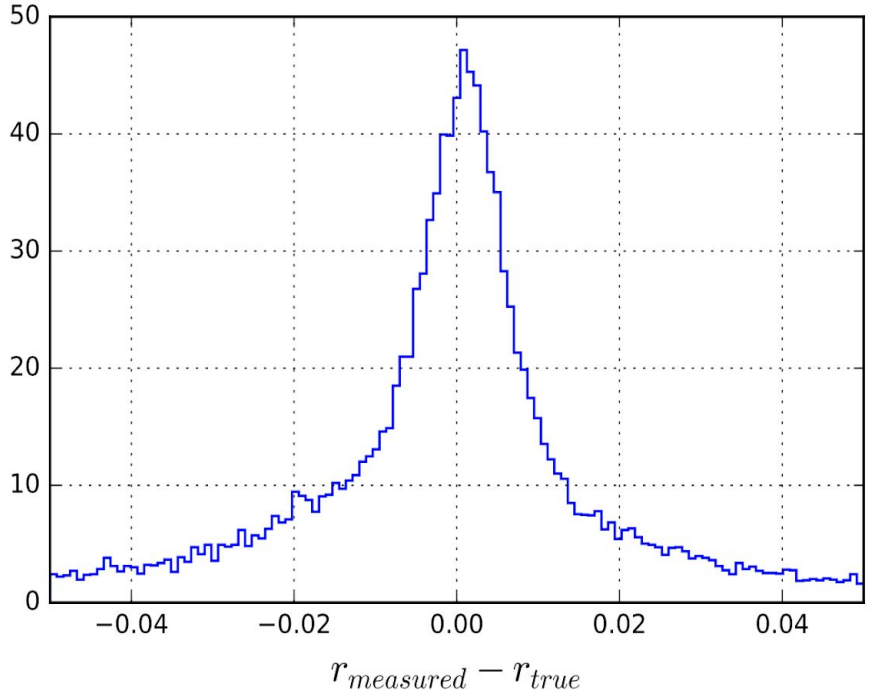
DC1: THE PRODUCTS

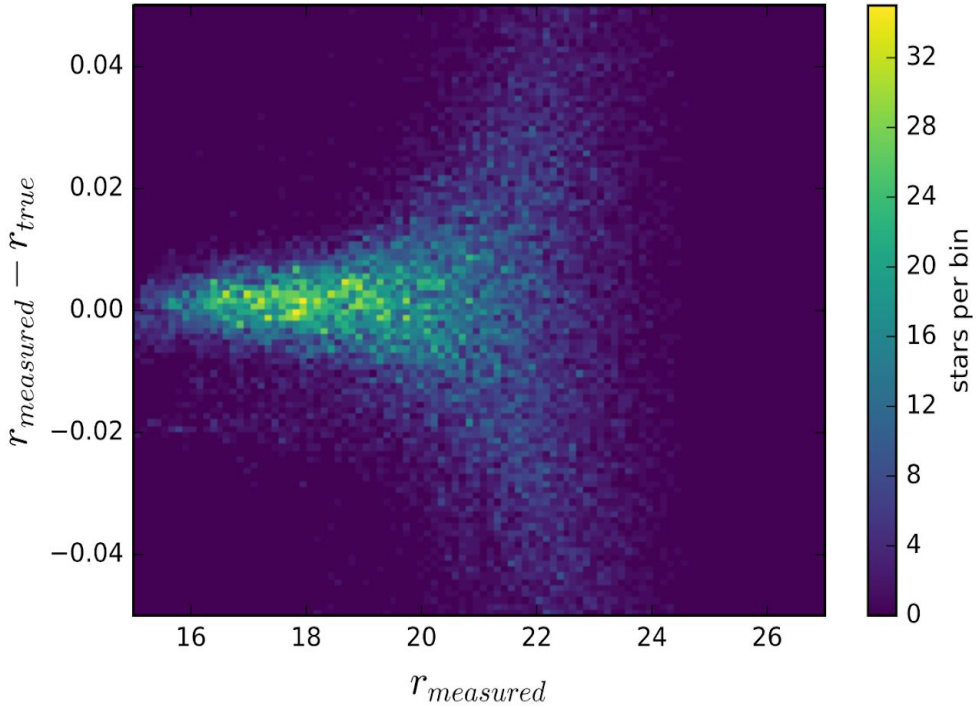


Slide from Javier Sanchez

DC1 QA: PHOTOMETRY CHECKS

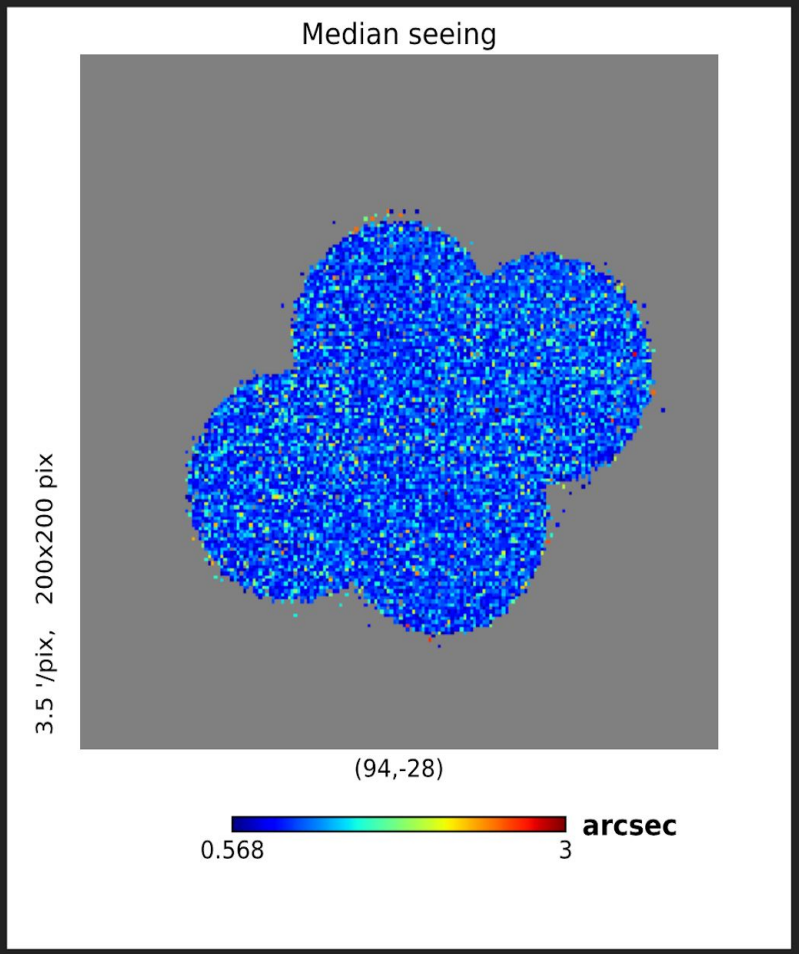
- External tests: Compare the input (true) flux with the measured flux

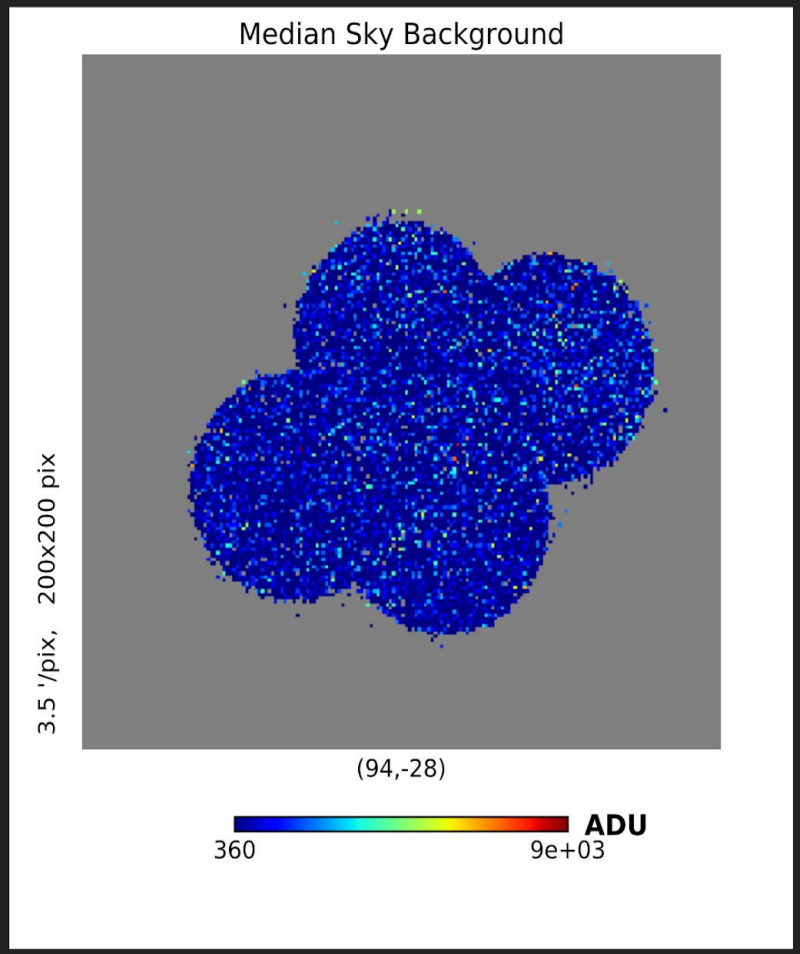




Slide from Javier Sanchez

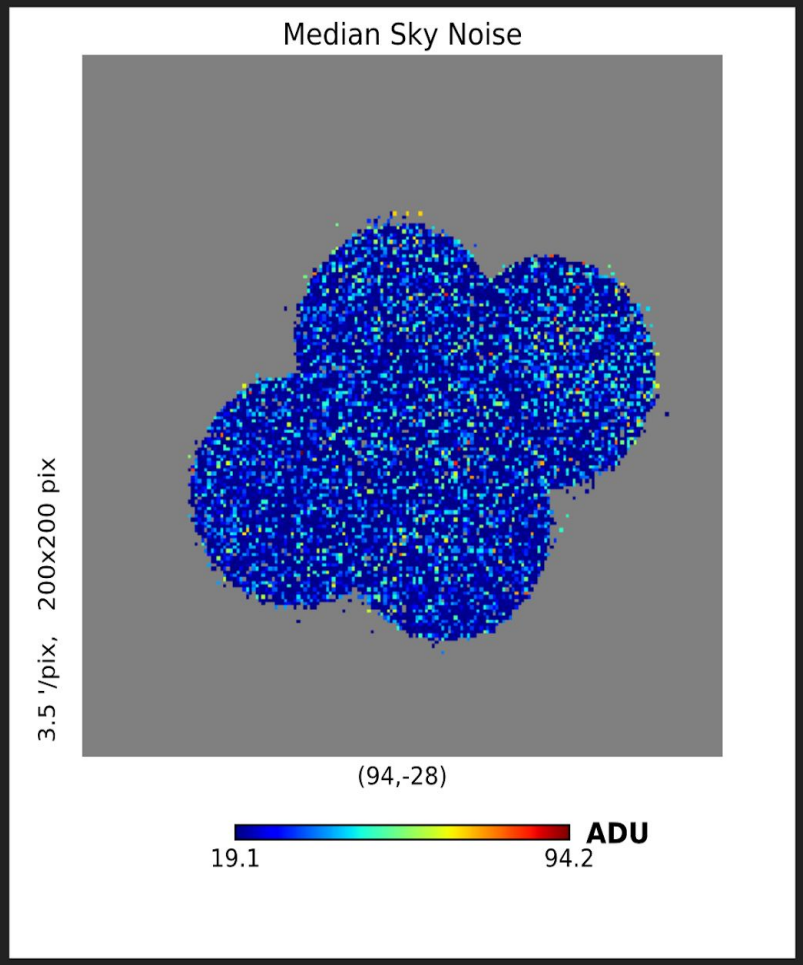
SYSTEMATICS

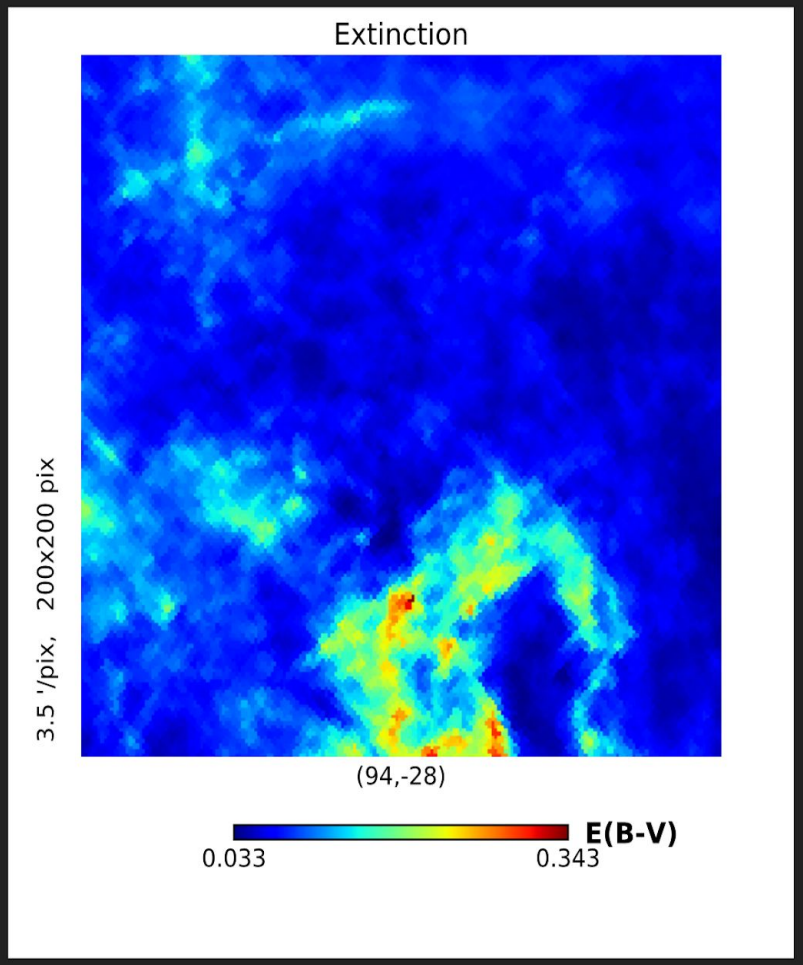




Slide from Javier Sanchez

SYSTEMATICS





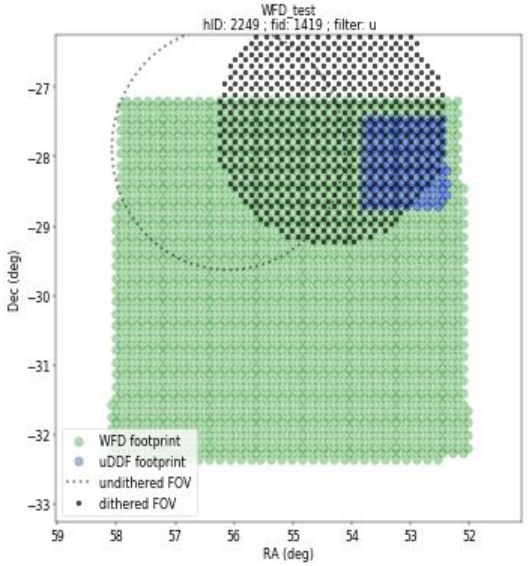
Slide from Javier Sanchez

DESC Data Challenges

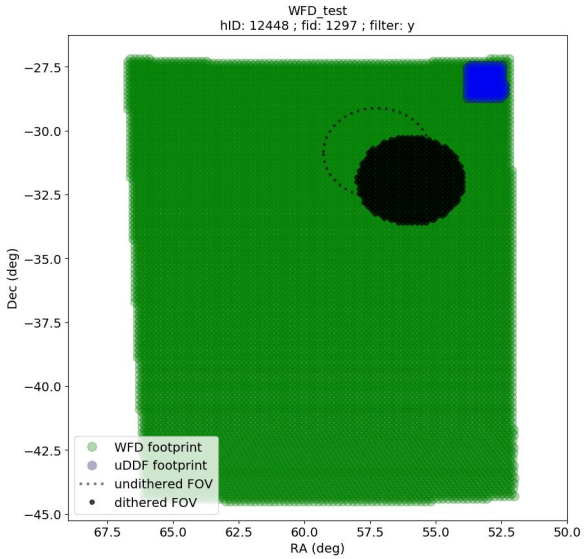
- Created so the collaboration will be ready for the first LSST data
- 3 sets of Data Challenges scheduled
  - Increase in size and complexity in each iteration
- Data Challenge 1 completed, included 10 years of visits covering 40 square degrees in r-band with large random dithers (recommended by Awan et al. 2016)
- Currently in Data Challenge 2 era
  - 300 sq degrees, 10 years, ugrizy
  - Includes variable sources in both WFD and “uDDF” regions
- Data Challenge 3 being defined; will mix analysis of precursor surveys, ComCam and simulated LSST data

DC2: Design

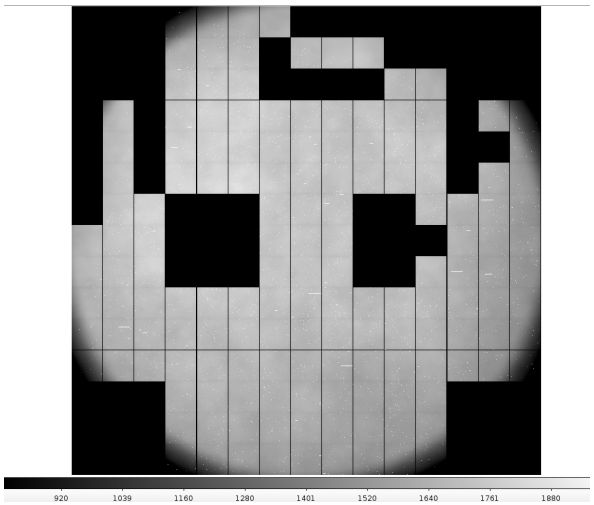
- Static sky (WL, CL, LSS, PZ) with images: 300 sq deg "main survey" area, 10 years *ugrizy* Wide-Fast-Deep (WFD) cadence
- Time domain (SN, SL) analyses: 1 sq deg "ultra Deep Drilling Field (DDF)" embedded in corner of main survey, 10 years *ugrizy* WFD + DDF visits



WFD dithers, Run 1.2p

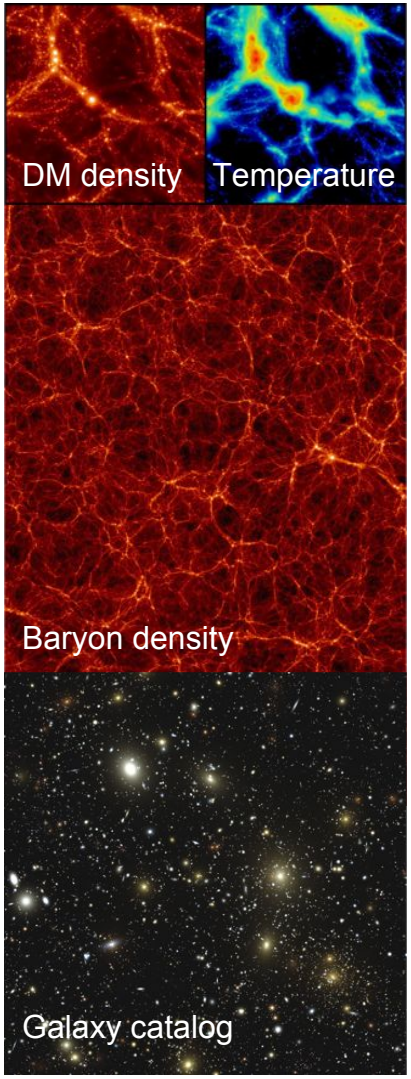


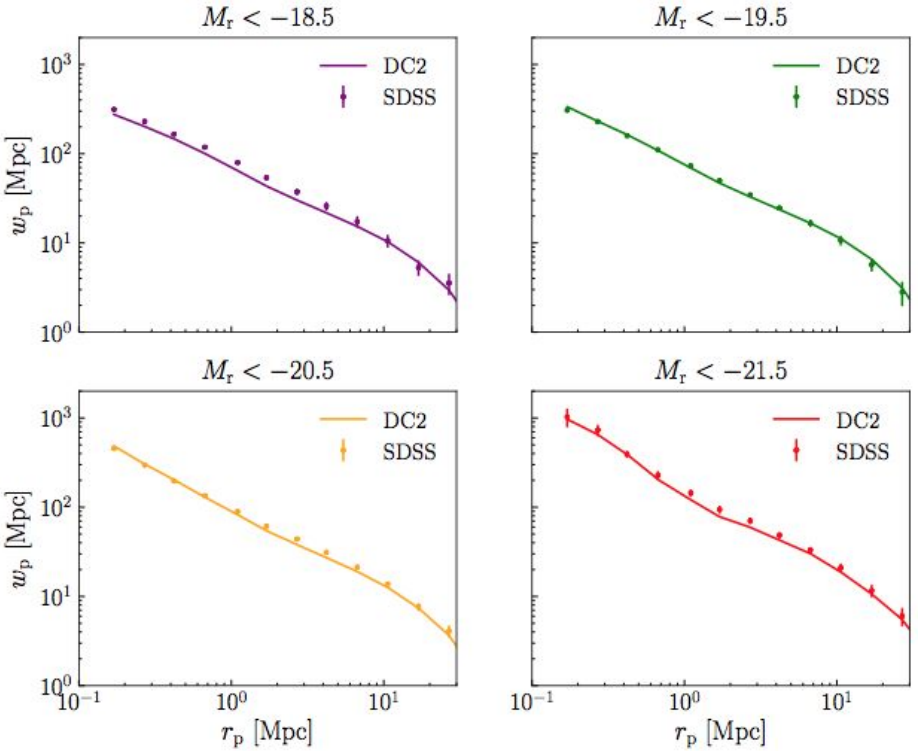
WFD dithers, Run 2.0



Partially-simulated LSST focal plane

DC2: Input Extragalactic Catalog



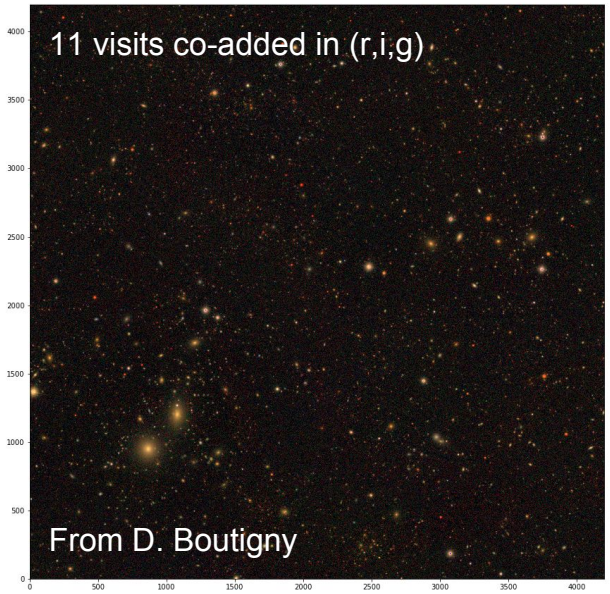


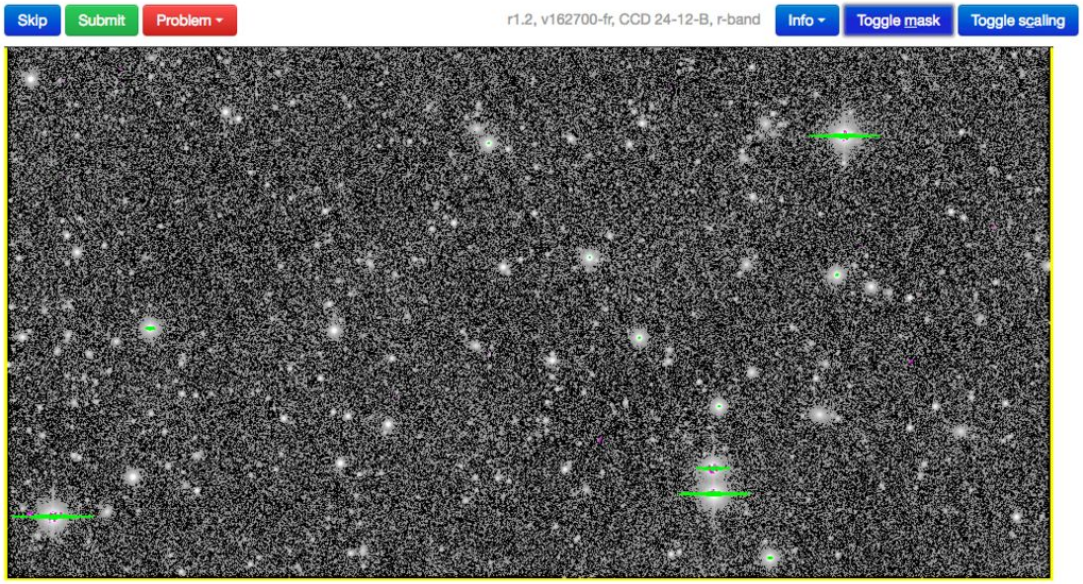
Hearin, Koryatov, Larsen et al

- HACC "Outer Rim" N-body simulation provides 5000 sq degree halo catalog
- Empirical population of halos with galaxies to get correct clustering statistics and colors; galaxies then matched with galaxies from semi-analytic model to enrich their properties

DC2: Simulated LSST Images

PhoSim on NERSC KNL: 1000 nodes allows us to produce 1 full focal plane visit image every 30 seconds - as fast as LSST

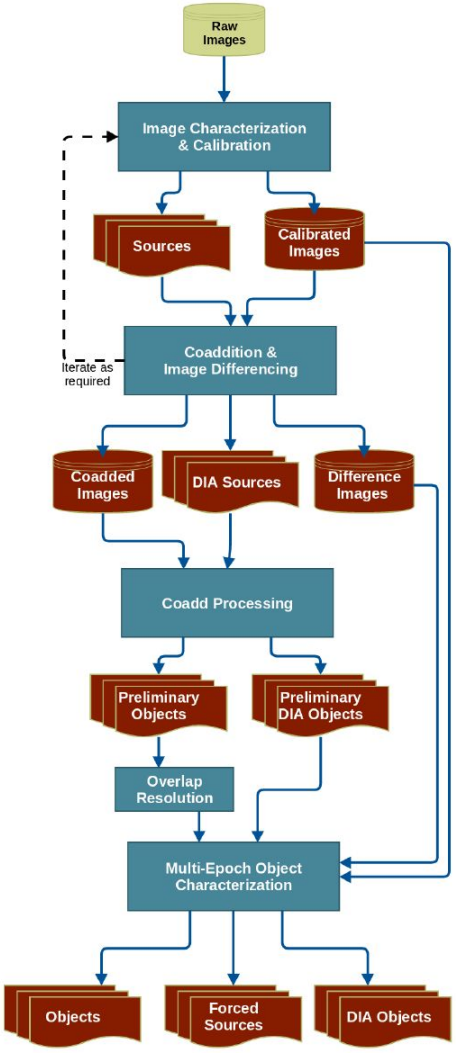




DESC-hosted, collaboratively-coded experimenter platform ImSim uses GalSim to produce LSST-specific simulations with tunable levels of complexity.

We use both codes, with a plausible variety of astrophysical and observational effects, to stress-test our analyses

DC2: LSST catalog data, from DM "data release processing" of the images



- Image processing, object detection and measurement, catalog generation will be done by the LSST Facility
- DESC needs its own mirror of the "data release processing" (DRP) pipeline, to reprocess 10% of the LSST images to probe for systematics
- DM-DC2 Task Force is assembling our prototype DRP pipeline, to run at CC-IN2P3 and NERSC
- (Data release includes "reprocessed" difference image analysis to make eg SN light curves)

Public “data” release?

DESC plans a public release of images and catalogs from at least some part of Data Challenge 2, circa January 2020

Details TBD but we are (always) open to hearing your suggestions!