

# Introduction to the DESC data challenge 2 (DC2)

---

J. Sánchez on behalf of the DC2 team  
University of California, Irvine  
francs1@uci.edu



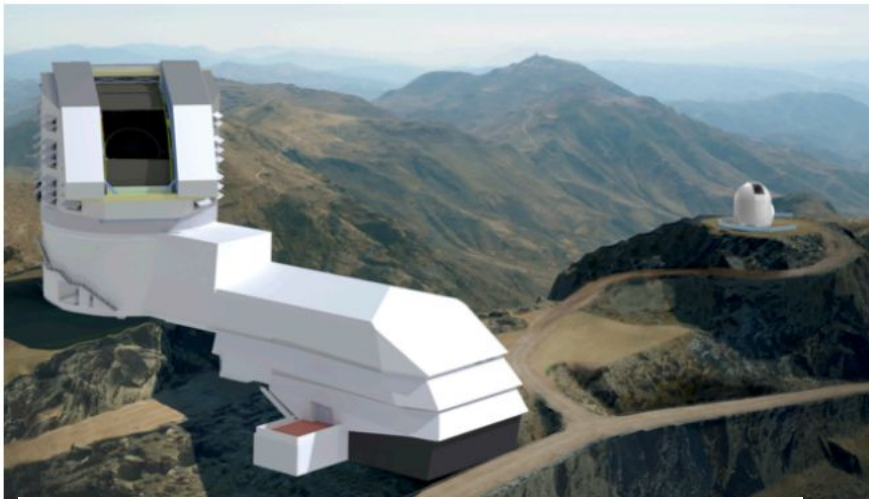
# Outline



- Motivation
- What is DC2?
- Stages
- Validation
- Goals
- Conclusions

# LSST

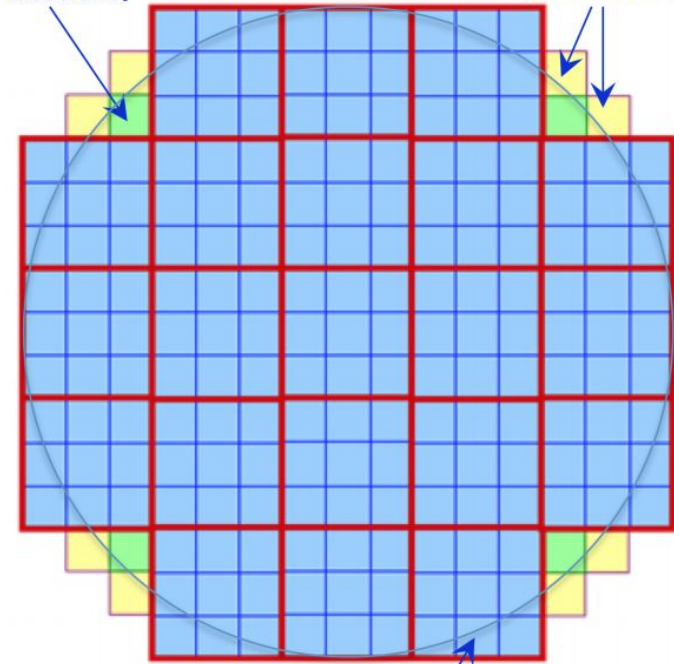
- 10 years, 6 bands,  $\sim 18,000$  sq-deg



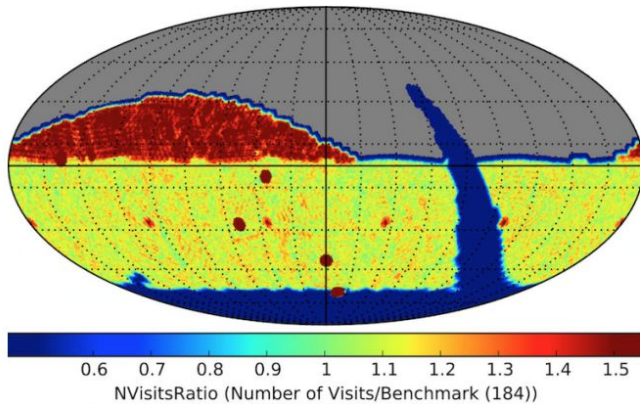
Images from Ivezić et al.

Wavefront Sensors  
(4 locations)

Guide Sensors  
(8 locations)



3.5 degree field of view,  
634mm diameter



# DESC



- “The LSST Dark Energy Science Collaboration will prepare for and carry out a variety of cosmological analyses with the LSST survey. In advance of LSST's first observations, DESC will help prepare for LSST science analysis, make synergistic connections with ongoing cosmological surveys and provide the dark energy community with state of the art analysis tools” (from [lsst-desc.org](https://lsst-desc.org)).

# Motivation

- LSST is very complex and is going to generate a very big amount of data ( $\sim 50$  PB raw/40 billion objects).
- Handling this data volume is already challenging  
-> Requires thought and training.
- We can test if the different pipelines meet the requirements of the Science Working Groups.
- Having end-to-end simulations allows to study in detail certain systematic effects and push the limits of our science.

# Data Challenges

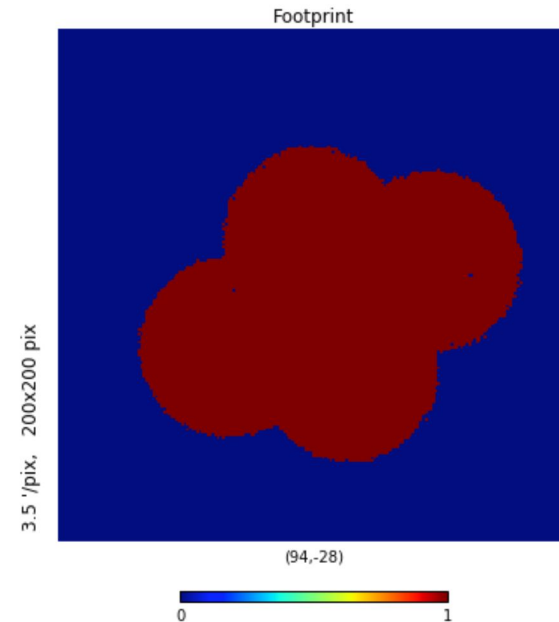
- How many data challenges is DESC going to have?
  - There are planned a total of 3 data challenges in increasing order of complexity -- DC1, DC2, and DC3.
- What is the philosophy behind these data challenges?
  - Commissioning data is expected to be large and complex, by the time of DC3 we expect to have simulations large and complex enough to be well prepared for commissioning.
  - An incremental approach makes easier to check for problems in new additions, isolate certain effects, and develop the analysis pipelines.

# Data Challenges

- Who is participating in the production of the data challenges?
  - The data challenges are coordinated by the DESC computing and simulations coordinator Katrin Heitmann.
  - The simulation and processing are responsibility of the Cosmological Simulations and Survey Simulations working groups with the assistance of Computing Infrastructure.

# Data Challenges

- The DESC data challenge 1:
  - 43 sq-degrees
  - 2 dithering strategies
  - 1 band (r) to full depth



DC1 footprint.  
Credit: The DC1 team



# What is DC1?

SAOImage ds9 <2>

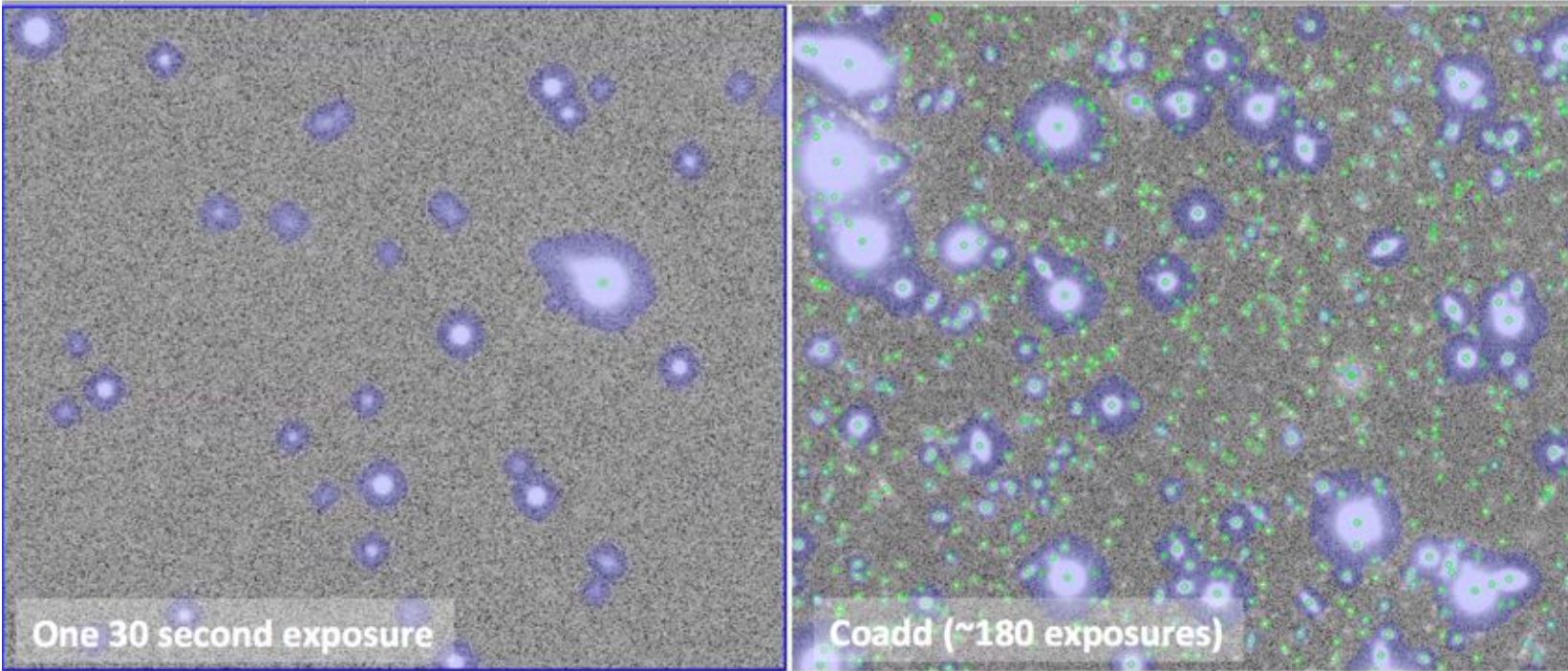
File Edit View Frame Bin Zoom Scale Color Region WCS Analysis Help

File		
Object		
Value	18.2801	
WCS a	104995	80838
Physical	X 3096.000	Y 2939.000
Image	X 3096.000	Y 2939.000
Frame 2	x 1.000	0.000 *

Courtesy of C. Walter

file edit view frame bin zoom scale color region wcs help

- + fit zoom 1/8 zoom 1/4 zoom 1/2 zoom 1 zoom 2 zoom 4 zoom 8

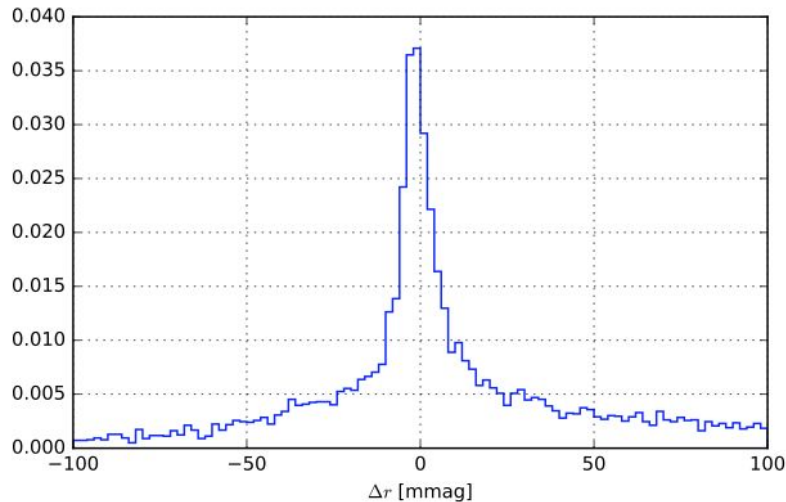


One 30 second exposure

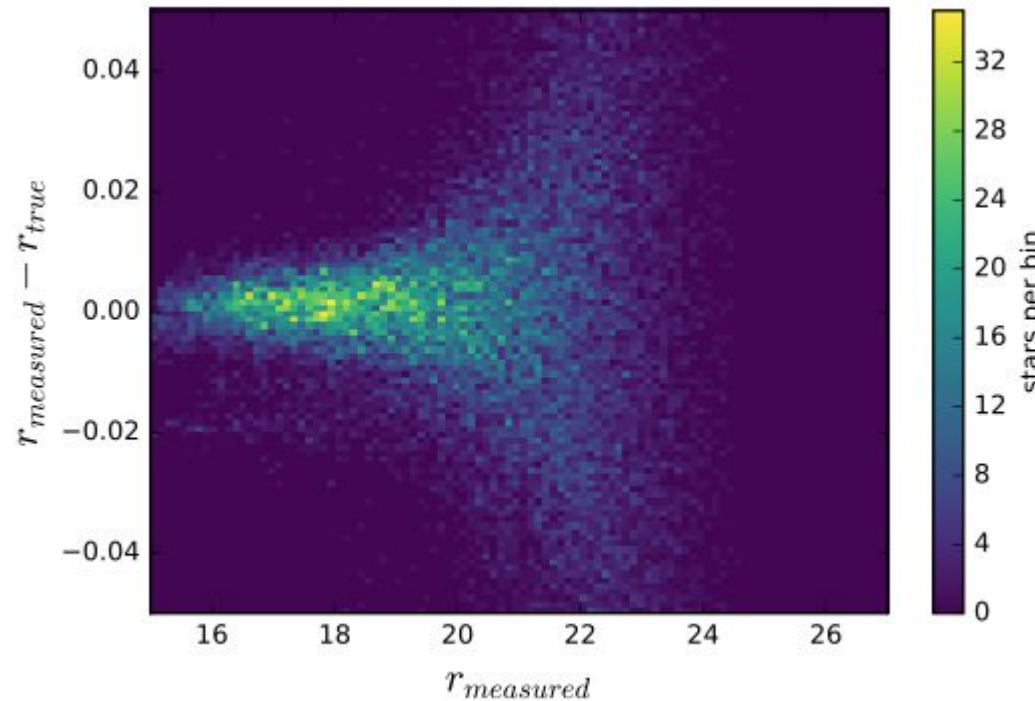
Coadd (~180 exposures)

# DC1: QA

- Photometry is unbiased and consistent between exposures (for dithered and undithered)



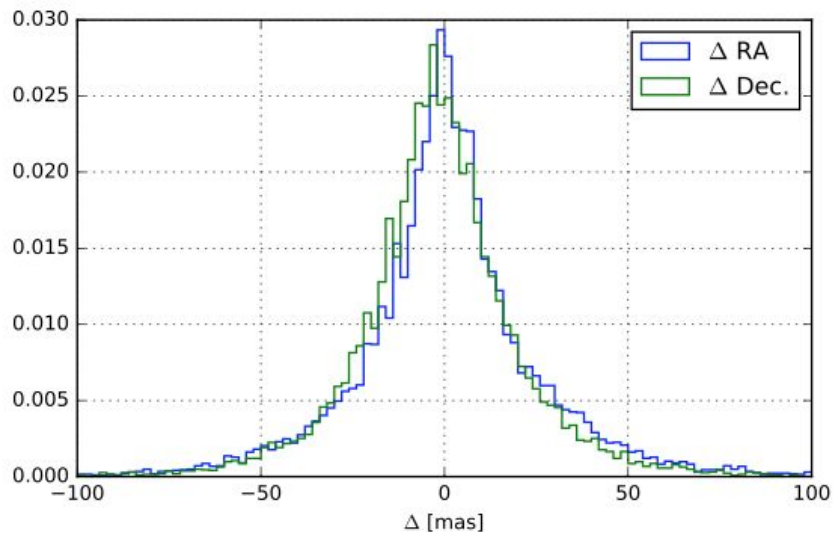
Magnitude difference of the same objects between different exposures



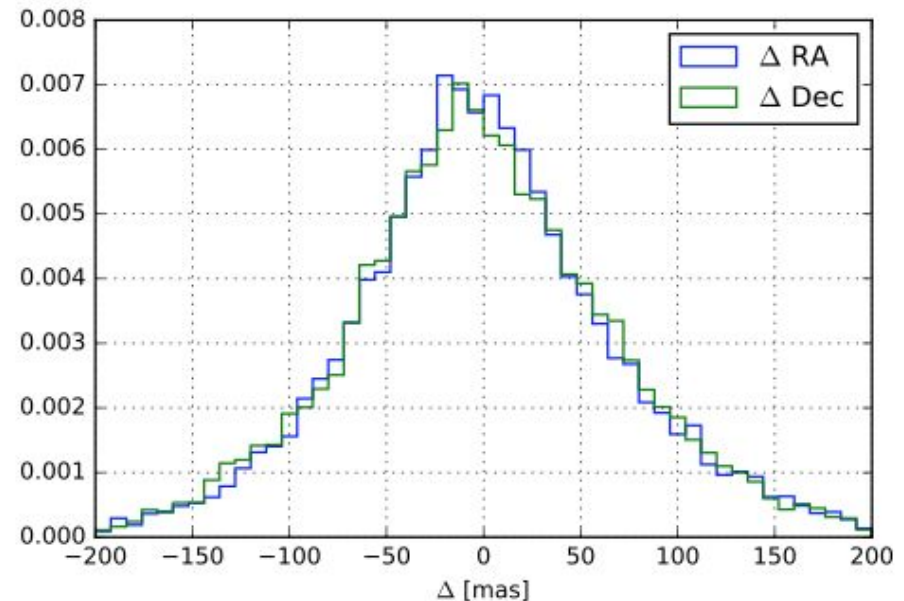
Comparison with the input catalog (stars only spatial matching)

# DC1: QA

- Astrometry is biased with respect our reference due to proper motion but is consistent between exposures (for dithered and undithered)



Position difference of the same objects between different exposures

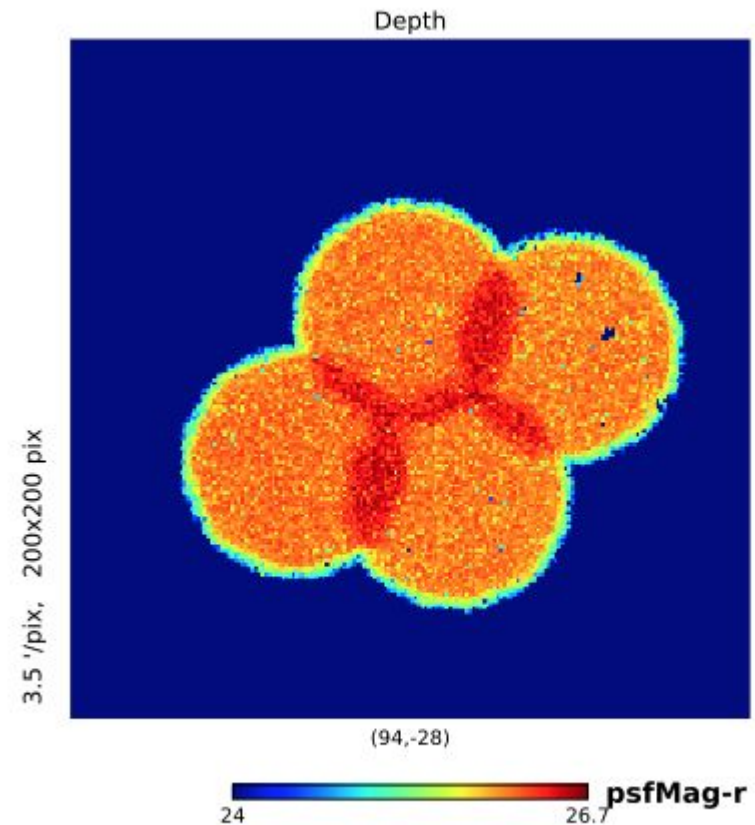
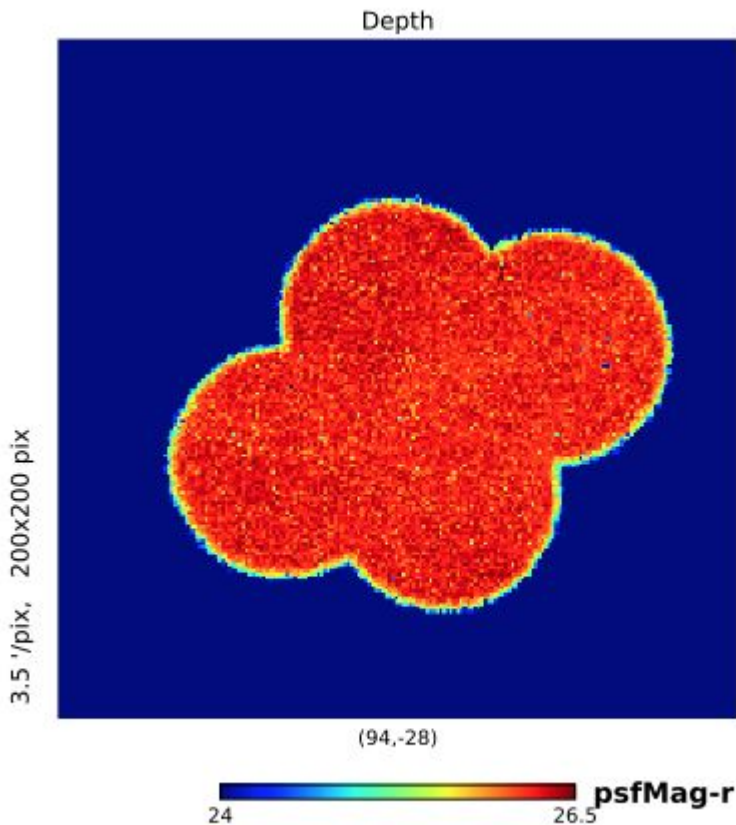


Comparison with the input catalog (stars only spatial matching)



# DC1: QA

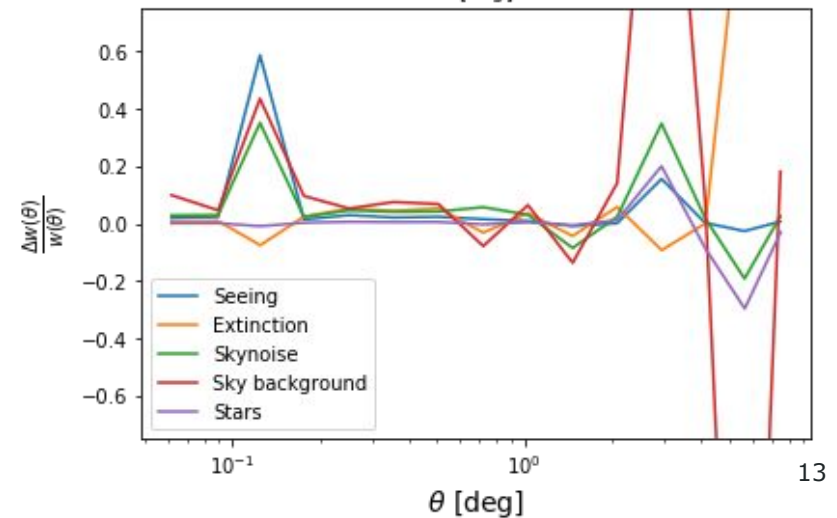
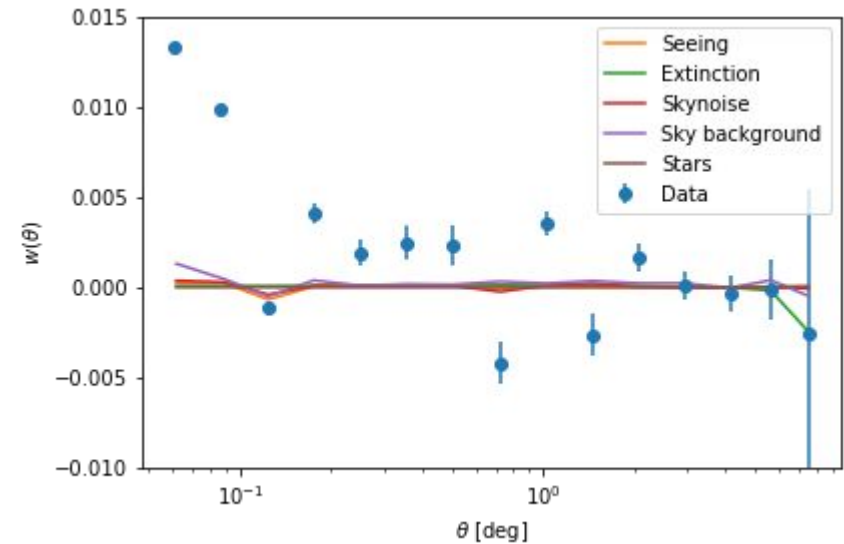
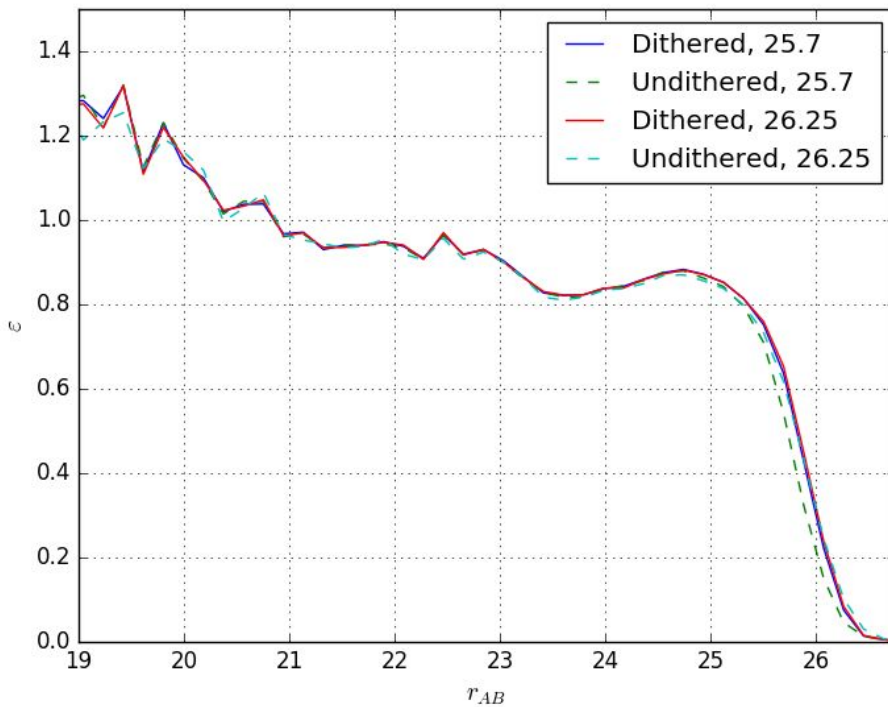
- Good uniformity for dithered (not so much for undithered)



5-sigma depth dithered (left) and undithered (right)

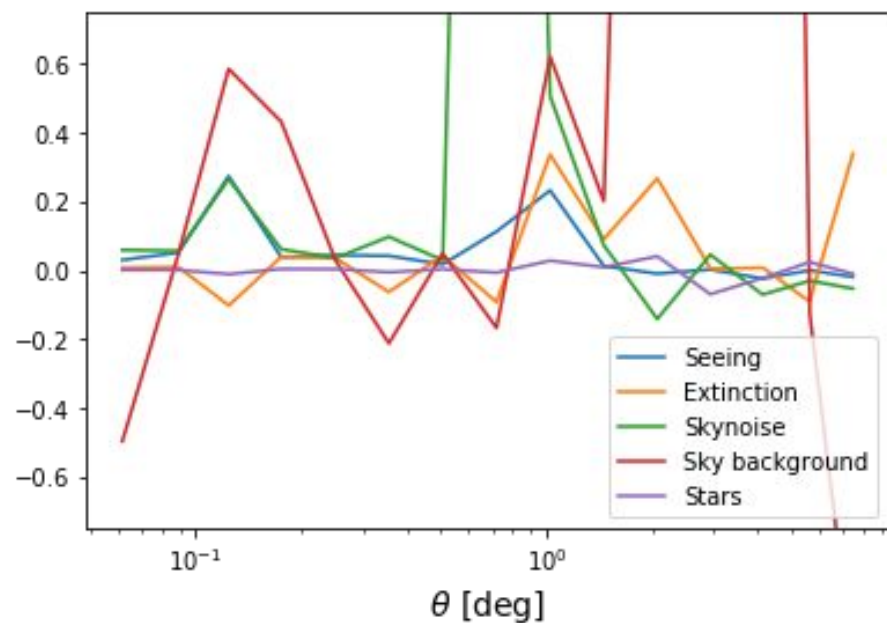
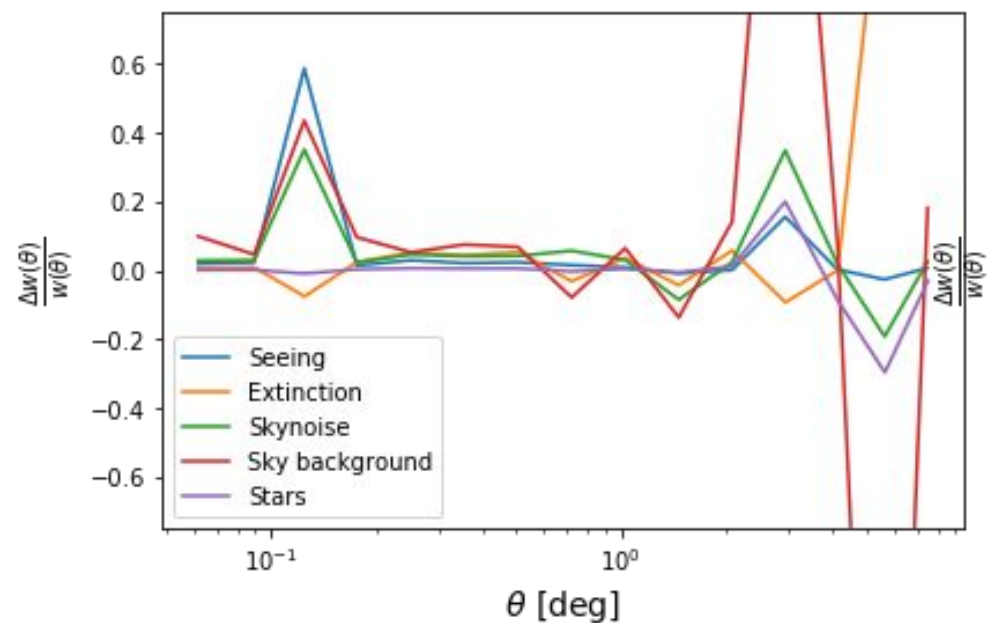
# DC1: Results

- High completeness up to 25.3 and low impact from systematics

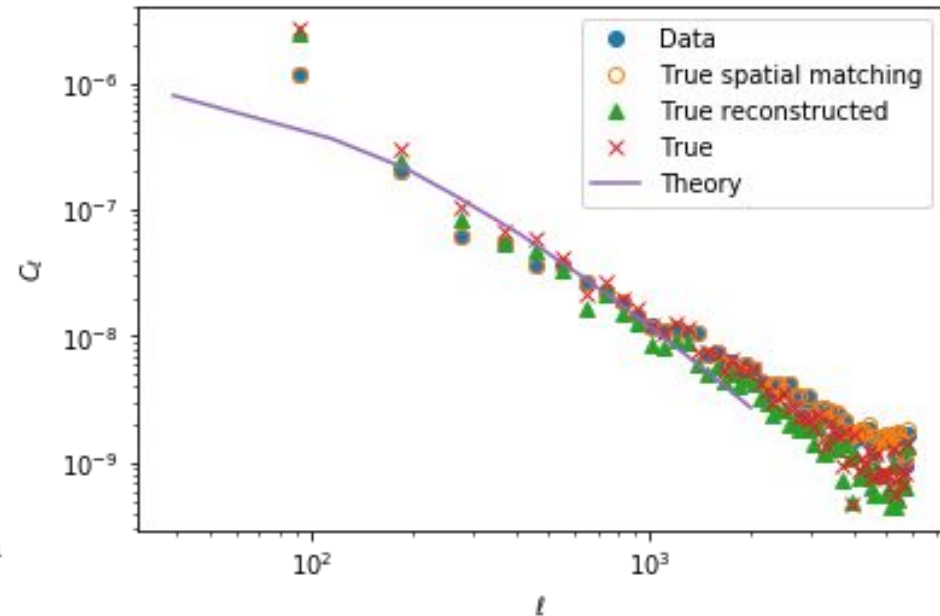
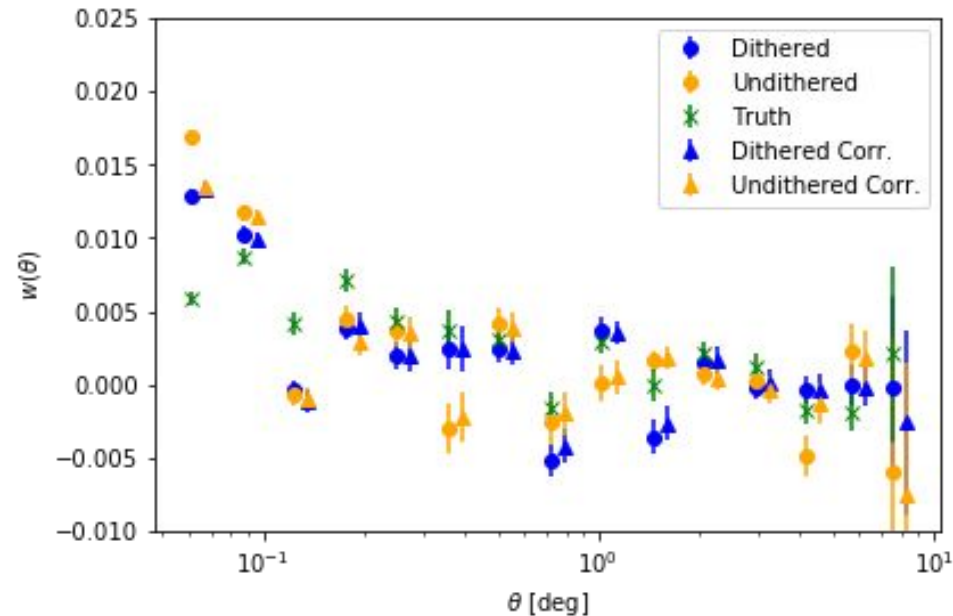


# DC1: Results

- The dithering strategy seems to diminish the systematics considered



# Results



- Reconstructing the selection function is challenging.
- One band only (no photo-z) and small area make difficult to make good predictions for our sensitivity and characterize main systematic (photo-z)..
- Some potential sources of systematic uncertainty not included.
- Need for a new and more complex data challenge!

# What is DC2?

- What is DC2?
  - DC2 is the second data challenge of LSST DESC.
- Who has access to it?
  - DESC collaborators will be able to access the data.
- What does DC2 entail?
  - DC2 entails the production, validation, and analysis of:
    - 5000 sq-deg mock catalog (extragalactic catalog)
    - 300 sq-deg end-to-end simulation: mock + image simulation + image processing (100 M CPU hours)



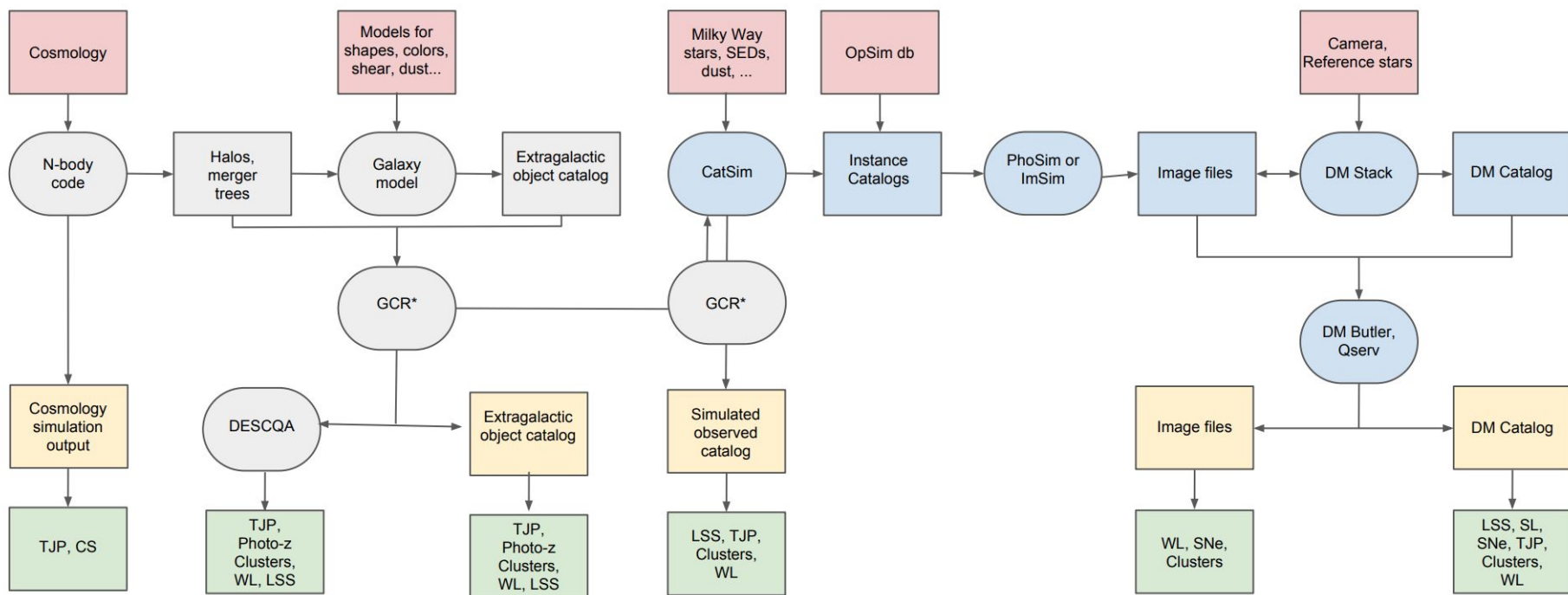
# What is DC2?

- Who is participating in DC2?
  - All Science Working groups have plans to analyze DC2 data.
  - We receive assistance from the technical working groups and the project (data management -- DM).
  - All DESC members are encouraged to participate!
- We hope to make DC2 available to a larger community.

# What is DC2?

## Workflow for Generating Simulated Data v3

- Responsibility of Cosmological Simulations Working Group
- Responsibility of Survey Simulation Working Group
- Input
- Output delivered to collaboration
- Users



\*GCR = Generic catalog reader

Courtesy of K. Heitmann

# DC2: Extragalactic catalogs

- Input (extragalactic) catalog: CosmoDC2, N-body from Argonne (Outer Rim),  $(4.225 \text{ Gpc})^3$ , 1 trillion particles.
- Mass resolution of  $2.6 \times 10^9$  solar masses.
- WMAP-7 cosmology.
- 4 PB!
- 99 snapshots from  $z=0$  to  $z=10$ .
- 1% dilution used to compute intrinsic alignments.
- More details in [\(Korytov et al. 2018 in prep\)](#)\*.

\*DESC Confluence credentials  
required to see

## DC2: Extragalactic catalogs

- Including galaxies in haloes with [Galacticus](#) ([Benson et al. 2010](#)).
- Semianalytic model: Complex galaxy-halo connection.
- Monte Carlo approach to speed up/scale up the inpainting in 5000 sq-degrees using Galsampler (Hearin et al. in prep).
- Available information: Position, size, shape (disk+bulge), stellar mass, redshift (including RSD), shear, magnification, convergence, SDSS and LSST magnitudes, star formation rate, stellar metallicity, BH mass and accretion rate.

## DC2: Image simulations

- A subset of the extragalactic catalog ( $\sim 300$  sq-deg) will be used to create image simulations.
- 2 different approaches:
  - PhoSim (Peterson et al. 2015) -> Monte Carlo photon shooting, you can think of this as GEANT4 -> Computationally expensive (3 to 10+ hours per sensor/visit depending on the configuration)
  - imSim -> Uses GalSim (Rowe et al. 2015) to render the images passing specific LSST information (connections to OpSim, CatSim, sensor effects) -> Alternative approach

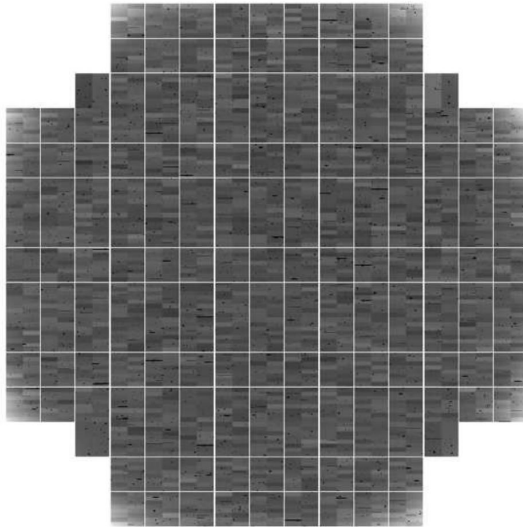
# PhoSim



- Ray traced atmosphere and optics -> Realistic PSF
- Weather and clouds: TBD.
- Moonlight included
- Includes brighter-fatter, tree rings, cosmic rays, bleeding, cross-talk, and readout noise.
- Differential chromatic refraction included.
- Standalone package, self-consistent physics.

24

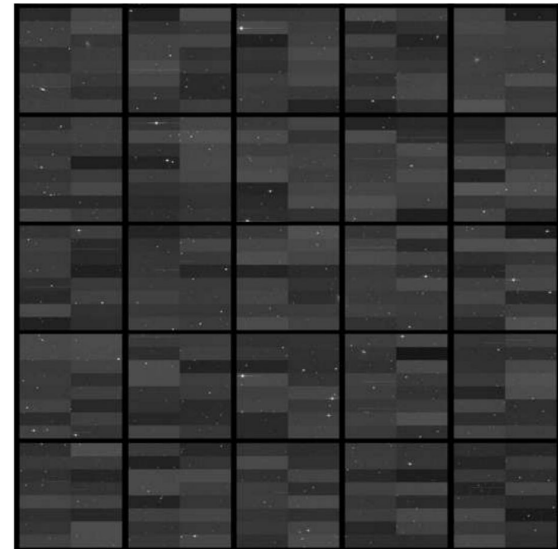
PETERSON ET AL.



Images from  
[Peterson et al  
2015](#)

**Figure 5.** A simulation of the entire field (10 sq. degrees) of the LSST field of view. 10 million stars and galaxies are in the simulation with over 1 trillion photons. This simulation was executed using CONDOR grid computing for about 1000 CPU hours in which each individual CCD was simulated in parallel. The image has over 3 billion pixel, so the full detail cannot be observed. The variation in bias levels of the individual amplifier is visible, as well the vignetting of the background near the corners of the field.

SIMULATION OF OPTICAL SURVEY TELESCOPES



**Figure 6.** The central 5x5 chips covering about 1 sq. degree. Individual bright stars are visible as well as hot columns.

# imSim



- ESO sky model (by LSST project)
- Aberrated phase screen optical model. Raytraced atmosphere.
- Moonlight included
- Includes brighter-fatter, tree rings, cosmic rays, bleeding, cross-talk, and readout noise.
- Differential chromatic refraction included.
- Uses GalSim to generate images and specific models from LSST project

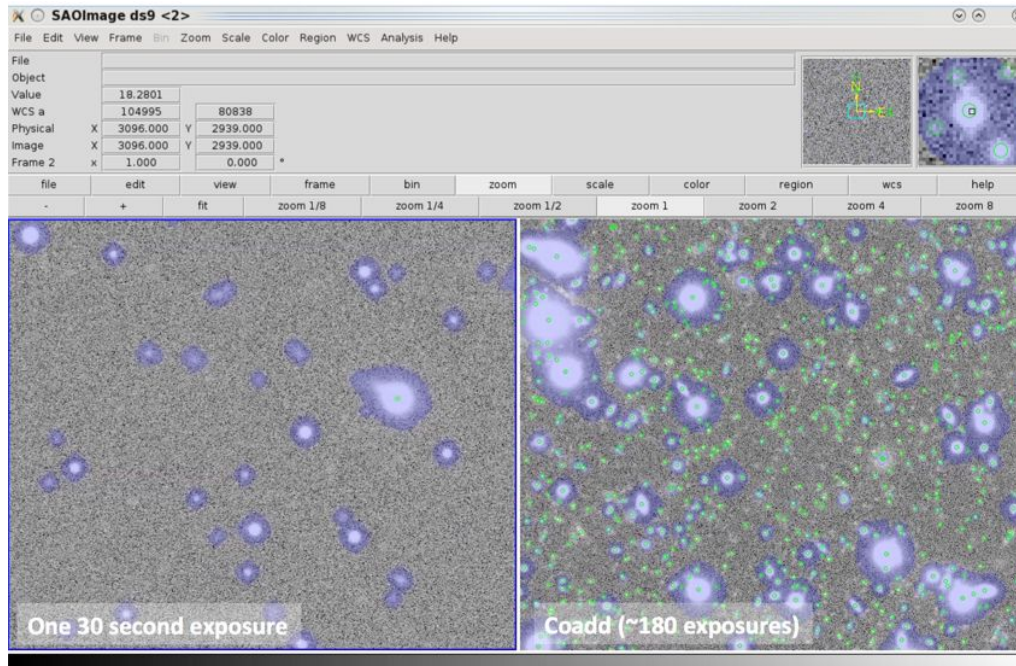


Image from imSim DC1  
Courtesy of C. Walter

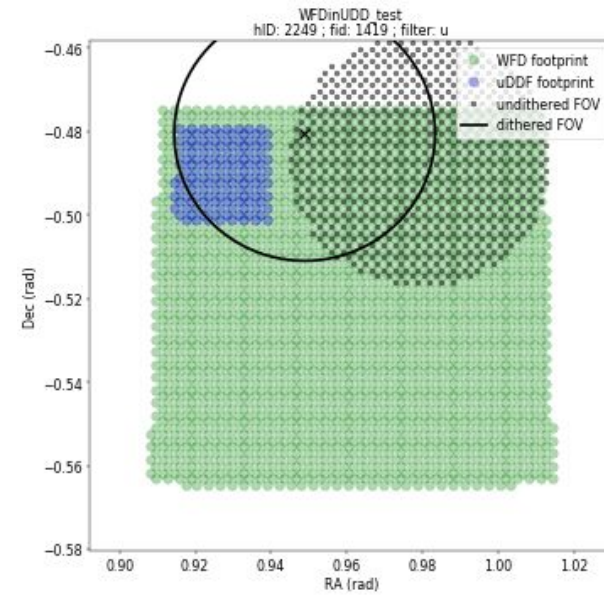
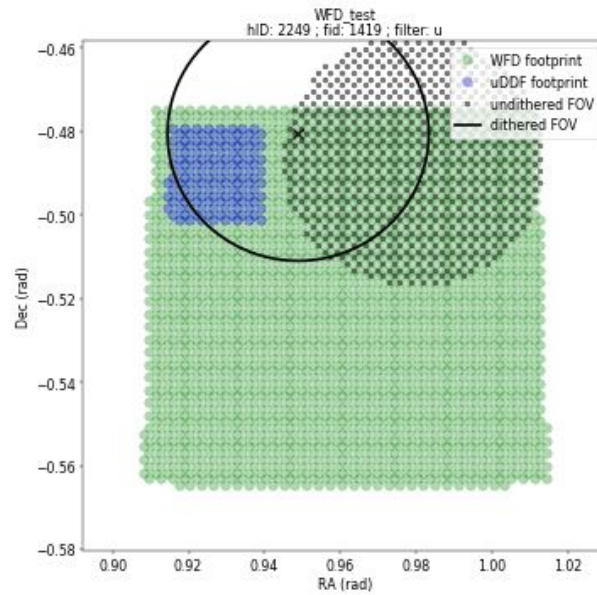
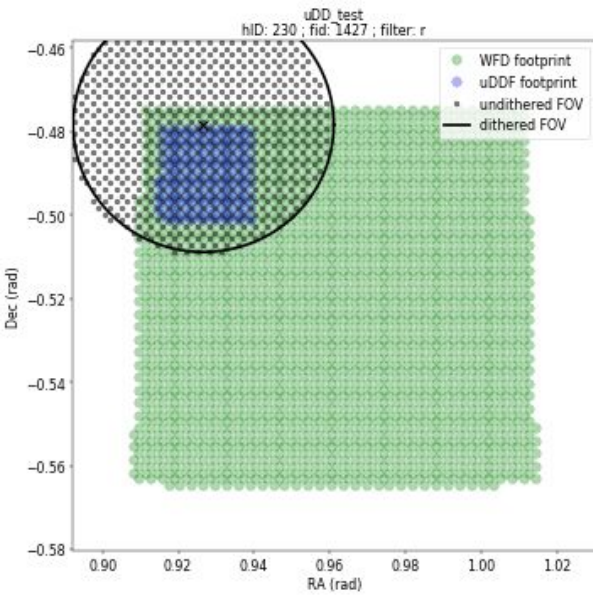
## DC2: Image simulations

- We will have 2 sets of simulated images for 300 sq-deg in u,g,r,i,z,Y LSST bands up to 10-year depth with the baseline cadence --  $\sim 27,000$  visits.
- We will also have a deeper field with transients of 1.25 sq-deg ( $\sim 20,000$  visits!!) -> ultra deep drilling field (uDDF).
- Over-abundance of AGNs, SN Ia and core collapse in this field.
- Realistic dithers included (rotation + translation of the FOV).
- Parallel generation (1 job/sensor/visit).



# DC2: Image simulations

## uDDF



Animation credit: Humna Awan and the DC2 team

# Image processing

- These images will be processed using the LSST DM stack (configuration similar to HSC).
- CModel fluxes (like SDSS), HSM (re-Gaussianization) shapes included.
- The science working groups may create additional products.
- Transient pipeline included.
- Access to coadds and individual calibrated exposures (calexp) if needed.
- DC2-DM task-force planning the configuration and execution.

# Catalog distribution

- Database + pandas dataframes (HDF5) + original files (FITS).
- The working groups will apply their analysis on their favorite catalog format.
- **Y1, Y3 and Y10 releases** are being considered.
- Having the different data releases allows us to check the data complexity and our statistical power at different stages.
- We can learn and improve the observing strategy.

# Stages



- Catalog production: protoDC2 finished in its 4th version. CosmoDC2 on the way -> Extensive validation (DESCQA)
- Run 1.1p: Test on a smaller area ( $\sim 40$  sq-deg) with PhoSim on 6 bands to full depth + deep field.
- Run 1.2p: Same characteristics as 1.1p but different configuration for the background. Running now!
- Run 1.1p and 1.2p validation: Happening now!
- Run 1.1i: On the way. Same as 1.1p but with imSim.
- **Run 2.0: Final run of DC2.** (This will be the actual DC2). We expect this to be done by the next DESC collaboration meeting (July at CMU).
- Publications about the simulation generation process to be ready by the end of the year.



# Validation

## Preliminary Exposure Checker



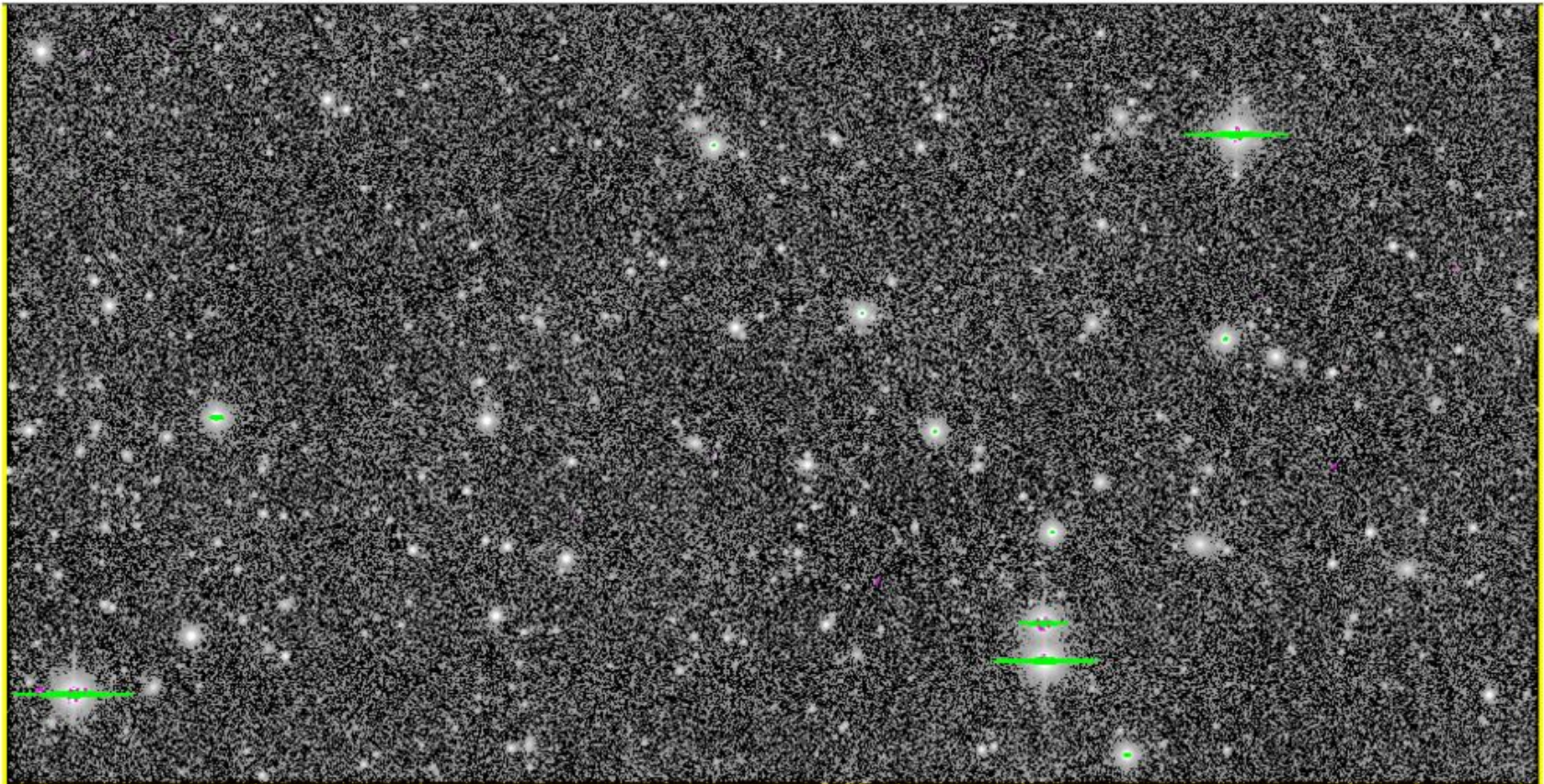
Skip Submit Problem ▾

r1.2, v162700-fr, CCD 24-12-B, r-band

Info ▾

Toggle mask

Toggle scaling



Credit: Alex Drlica-Wagner, Simon Krughoff and the DC2 team

# Validation



- Catalog validation: [DESCQA \(Y-Y Mao, et al. 2018\)](#). See related DESC seminar.
  - We are using this framework to test the extragalactic catalog and the inputs for PhoSim and imSim (instance catalogs).
- Image validation:
  - DESC exposure checker (based on DES exposure checker) provides a quick look to the images so we can identify problems early.
  - Quick checks in images\*: Background mean/median/power-spectra, comparison with expected background.
  - Run through DM stack.
- Final catalog validation\*: DM QA tools (astrometry, photometry, etc), DESCQA, analysis by working groups.

\* help needed

# How the working groups will use DC2?



- The different working groups presented a summary of the studies that they want to perform in DC2:
  - Galaxy clusters:
    - Impact of observational systematics in cluster finding algorithms.
    - Impact of blending and shear biases in mass reconstruction.
    - Understand systematics impact on cluster lensing profiles.
  - Large-scale structure:
    - Investigate galaxy selection.
    - Test analysis pipeline.
    - Test systematics in 2-point statistics.
    - Measure magnification bias in presence of systematics.
- And many more! More details [here](#).

# Where can DESC members contribute?



- We need people to test and analyze these catalogs.
- We need to test the catalogs themselves, the data distribution, whether they include all the information needed or not.
- We need to test the processing pipelines. Do they meet your requirements? DM wants to know about this.
- Are there effects missing that you want to test? Let us know!
- Do you have suggestions for the data formats and distribution?
- Do you have any QA resources/ are you interested in developing QA resources or testing/improving the ones available, e.g. DM-QA?
- We need to learn how to use the tools developed by the project and give feedback.



# Conclusions

- The DESC DC2 is an underway complex effort to produce, validate and analyze LSST 10 year-like data.
- 300 sq-deg/6 bands with 1.25 sq-deg uDDF
- 2 different independent simulations (imSim and PhoSim).
- Releases for Y1, Y3 and Y10 depth are being considered.
- New cosmological extragalactic 5000 sq-deg catalog with complex galaxy-halo connection.
- Interesting analyses will be performed.
- All DESC members are encouraged to contribute!
- We hope for these data to be available for a wider community.