



# Cosmological Simulations for Large-Scale Sky Surveys and the LSST DESC Data Challenge 2

---

Katrin Heitmann for the DESC Collaboration  
Thursday, August 2, 2018  
LineA Webinar



DC2 Motto: ***“We are slacking to hammer this out!”*** from Tom Glanzman on Github while working on DC2

# Introduction

---



# The Large Synoptic Survey Telescope



8.4m dedicated survey telescope, 9.6 sq deg field of view  
3Gpix camera, data facility and access center at NCSA

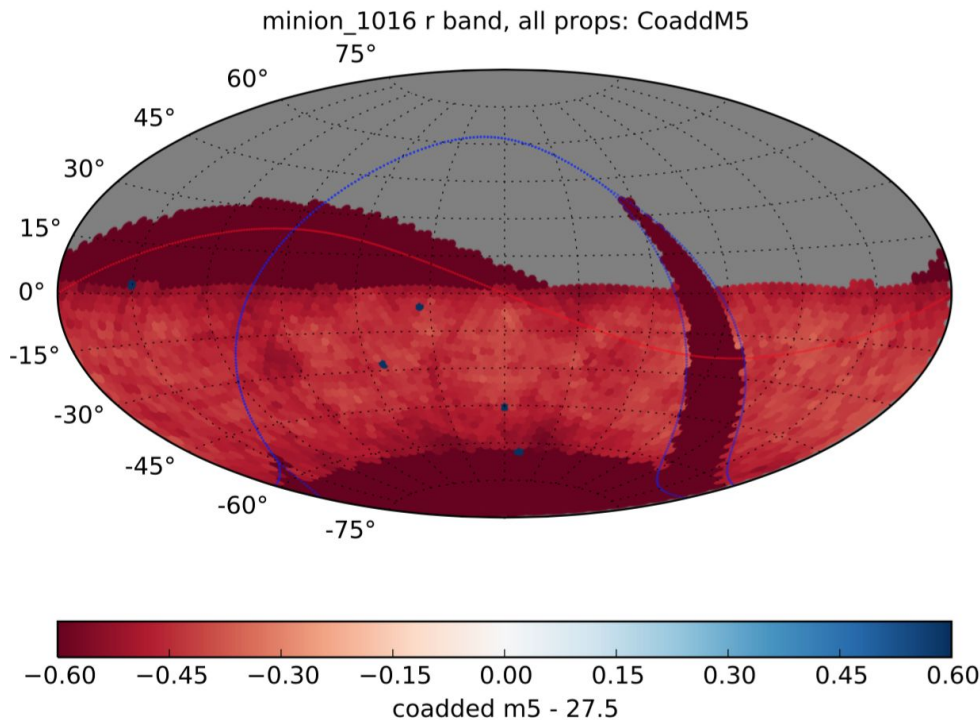


Monday May 7, 2018



<https://www.lsst.org/news/see-whats-happening-cerro-pachon>

# The LSST Survey and Data



Main survey: 18,000 sq deg of southern sky, "Deep Drilling Fields", 10 sq deg each, plus further "mini-surveys" to support special science cases

- Worldwide "Alerts" released nightly (with minimal info)
- Annual data releases: images, Object and Source tables (including galaxy model measurements and forced photometry light curves), released to LSST data rights community (all US astronomers and more)

# This is Dark Energy Science Collaboration!



Collaboration Meeting



Hack week



# What are we trying to learn about the Universe?



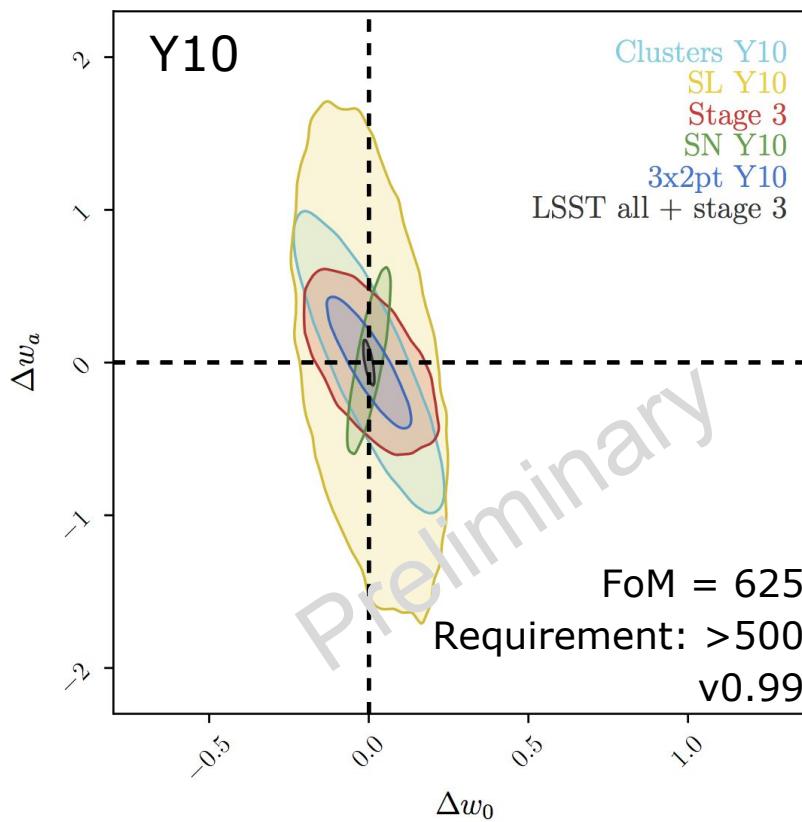
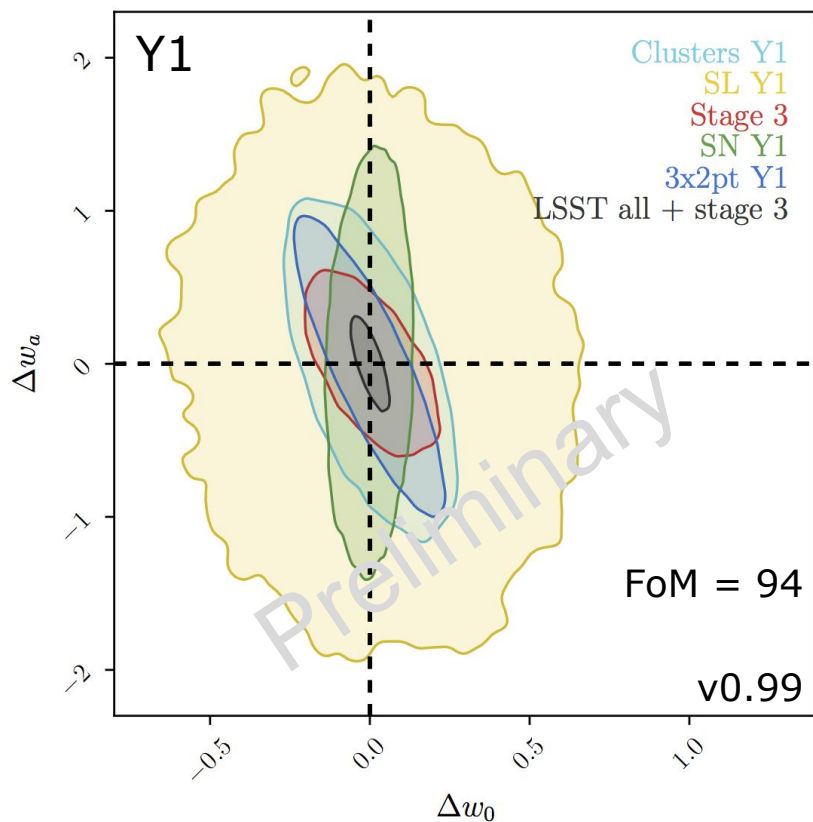
We want to understand cosmic acceleration:

- 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 do imaging surveys 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

# DESC Forecast of Dark Energy Measurement with LSST



Forecast 68% confidence level constraints on  $(w_0, w_a)$  for individual probes and their combination after 10% (Y1, left) and all (Y10, right) LSST data is analyzed.

T. Collett, T. Eifler, E. Gawiser, R. Hlozek, R. Mandelbaum, et al.



# LSST DESC Data Challenges

- DESC is carrying out large-scale data challenges (DCs) to prepare for the arrival of LSST data
- Full end-to-end simulation and processing to enable
  - Testing and developing of analysis and processing pipelines
  - Testing and developing of mitigation strategies for systematic effects (e.g., sensor defects, blending, atmospheric effects, ...)
  - Testing and developing data access strategies
- Three DCs with increasing complexity and size, currently data generation for DC2
  - DC1: Pathfinder project, 25 sq degrees, no weak lensing
  - DC2: 300 sq degrees, all working groups are engaged
  - DC3: Currently being planned

DC2 provides a complex challenge where the different working groups have to work closely together to establish a science community capable of carrying out a complex joint cosmological analysis on a scale just 10x smaller than LSST

# Focus of this talk: Data Challenge 2

- **Contributions from a LOT of people** (listed each name ones...)!
  - Design: *Humna Awan, Rahul Biswas*, Eric Gawiser, *Phil Marshall* et al.
  - Catalog production: Andrew Benson, Nick Frontiere, Salman Habib, *Andrew Hearin*, Katrin Heitmann, Joe Hollowed, Dan Korytov, Eve Kovacs, *Patricia Larsen*, Nan Li, Adrian Pope, Steve Rangel, *Jim Chiang*, Scott Daniels, Tom Glanzman, Danny Goldstein, Mike Jarvis, Bryce Kalmbach, Josh Meyers, James Perry, John Peterson, Glenn Sembrowski, Chris Walter, Dominique Boutigny, Nicolas Chortard, Dominique Fouchez, Fabio Hernandez, *Heather Kelly*, Simon Krughoff et al.
  - Validation: Anita Bahmany, Duncan Campbell, Joe deRose, Seth Diegel, Francois Lanusse, Rachel Mandelbaum, Yao Yuan Mao, Chris Morrison, Jeff Newman, Daniel Perrefort, Eli Rykoff, Javier Sanchez, Melanie Simet, Chun Hao To, Tom Uram, Vinu Vikraman, Alex Drlica Wagner, Michael Wood-Vasey, Rongpu Zhou et al.

**50+ DESC Members! Names: Contributed slides**

# The DC2 Design

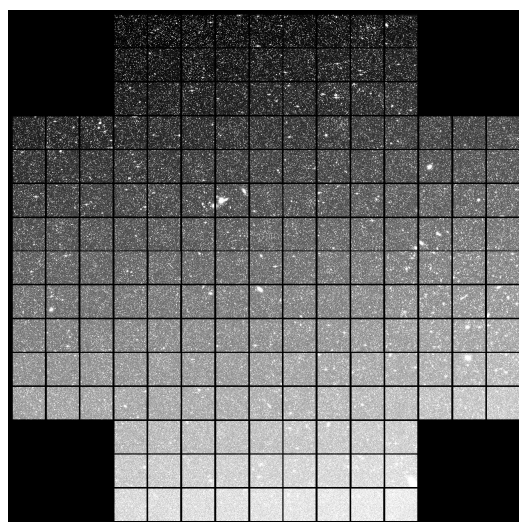
---

Humna Awan, Rahul Biswas et al.

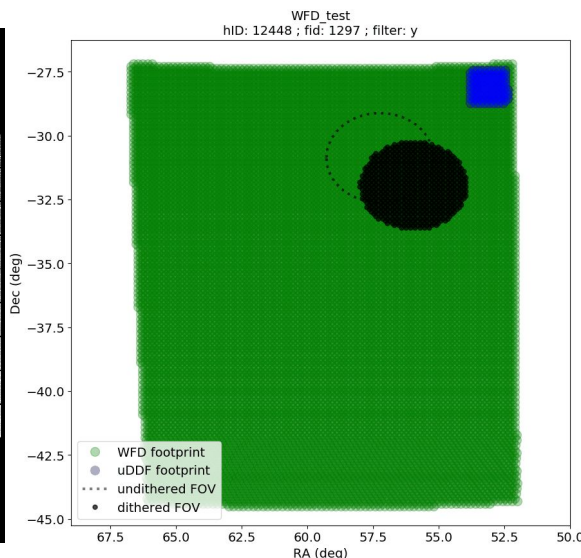


# DC2: Design

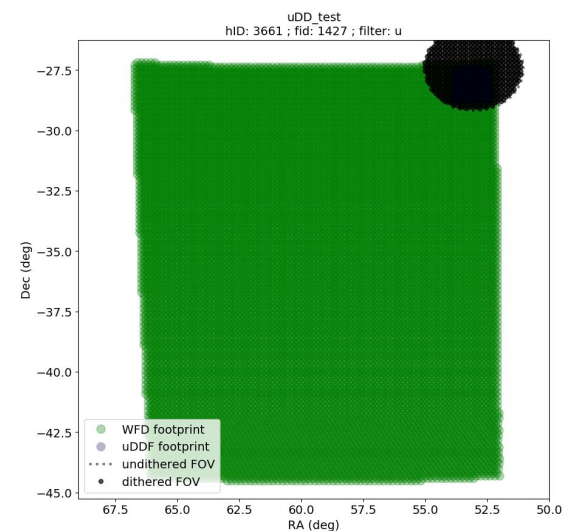
- Extragalactic sky: 5000 sq deg input catalog
- Static sky (WL, CL, LSS, PZ) with images: 300 sq deg "main survey" area, 10 years *ugrizy* Wide-Fast-Deep (WFD) cadence
- Time domain (SN, SL) analyses: 1 sq deg "ultra Deep Drilling Field (DDF)" embedded in corner of main survey, 10 years *ugrizy* WFD + DDF visits
- Work led by: [H. Awan](#), [R. Biswas](#), [E. Gawiser](#), [P. Marshall et al.](#)



**Simulated LSST focal plane**



**WFD dithers, Run 2.0**



**DDF dithers, Run 2.0**

# The DC2 Workflow

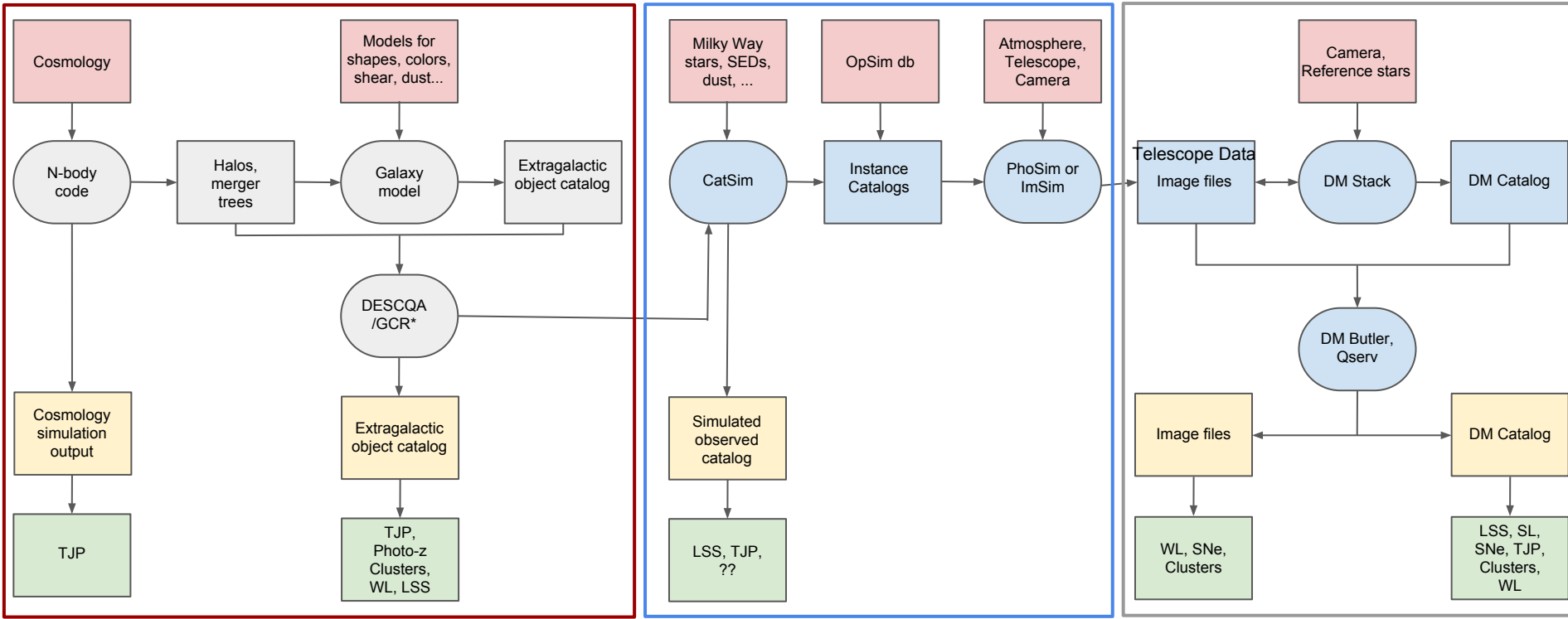
---

# DESC End-to-End Simulation Workflow



- Responsibility of Cosmological Simulations Working Group
- Responsibility of Survey Simulation Working Group

- Input
- Output delivered to collaboration
- Users



Extra-galactic catalog generation

Image simulations

DM processing

Codes

Data

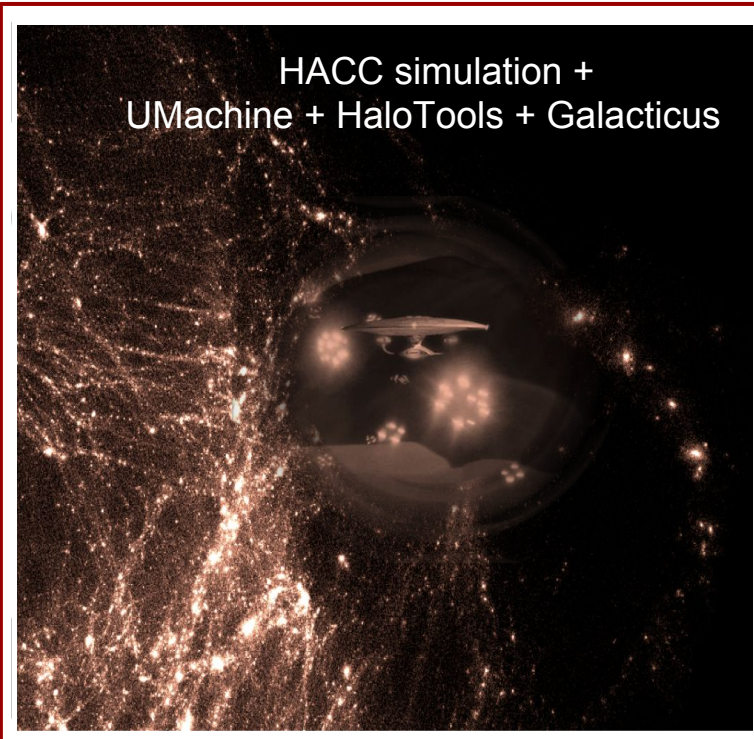
# DESC End-to-End Simulation Workflow



- Responsibility of Cosmological Simulations Working Group
- Responsibility of Survey Simulation Working Group

- Input
- Output delivered to collaboration

● Users



Extra-galactic catalog generation

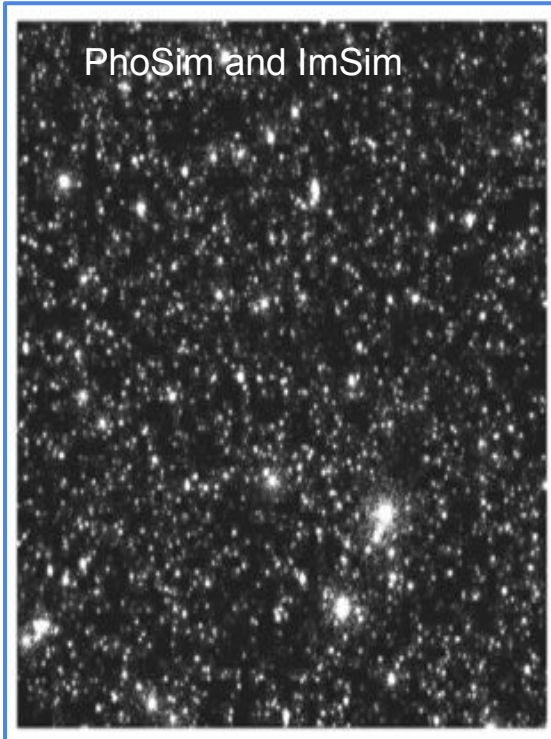
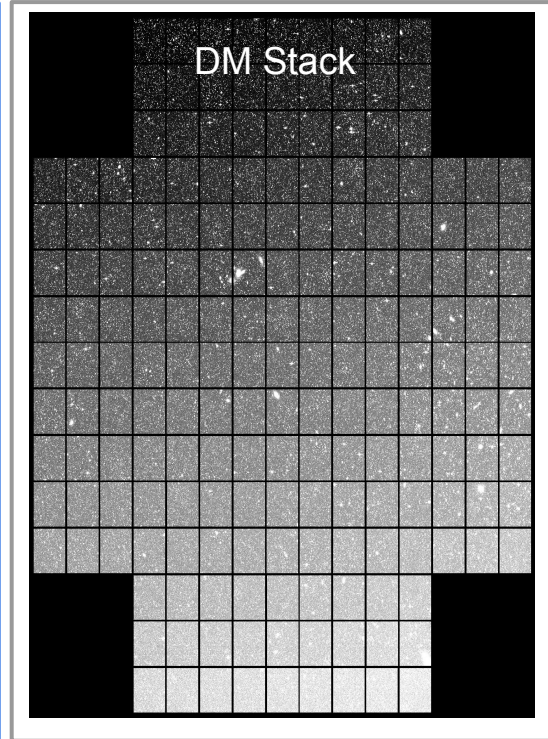


Image simulations



DM processing

# The DC2 Extragalactic Catalog

---

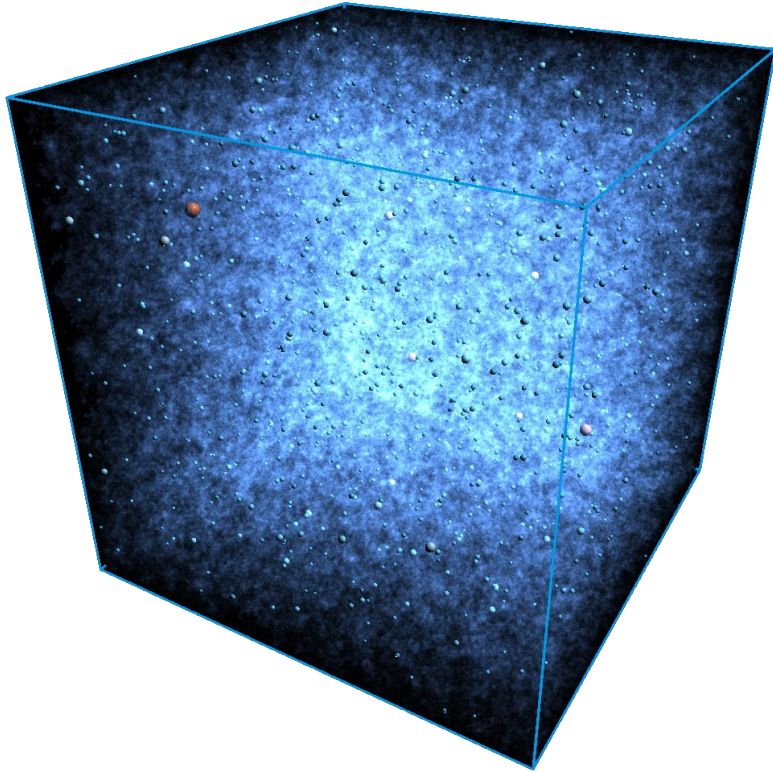
Andrew Hearin, Patricia Larsen

(in collab. with Dan Korytov, Eve Kovacs, J. Hollowed, S. Rangel et al.)





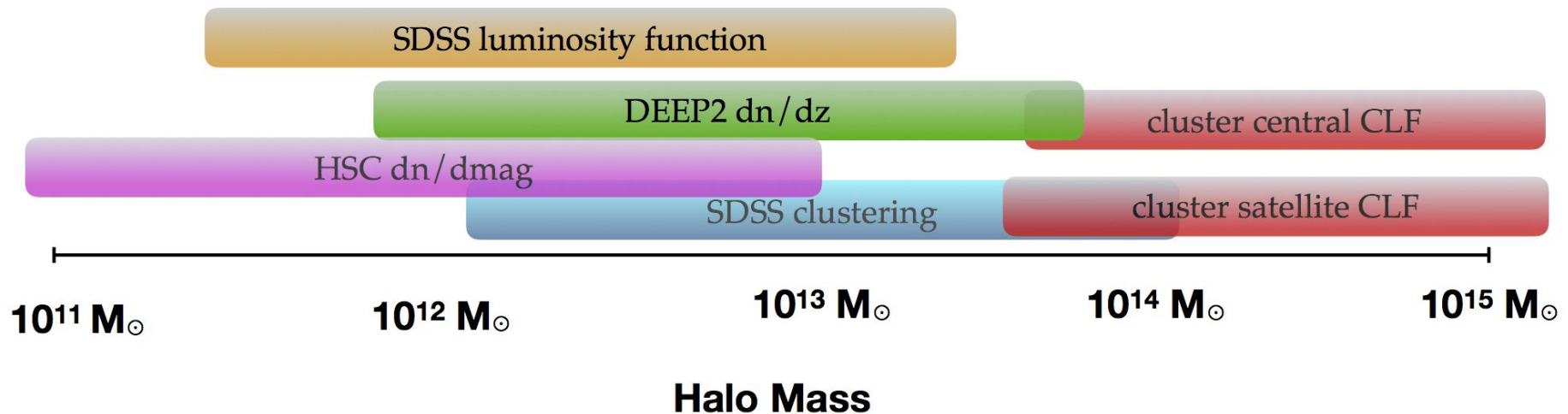
# DC2: Input Extragalactic Catalog



- HACC “Outer Rim” N-body simulation (more than *one trillion particles*, 3Gpc/h) provides matter distribution
- Empirical approach used to populate halos with galaxies to get correct clustering statistics and colors; galaxies then matched with galaxies from semi-analytic model to enrich their properties
- Range of observational data sets is used to tune the model to satisfy diverse DESC science needs at various mass/length scales
- Close collaboration between analysis and computing groups

# Extragalactic Catalog Validation

- Developed “DESCQA”, a webportal located at NERSC to run validation tests and inspect catalog quality in (almost) automatic fashion (Mao et al. ApJS 2018)
- Analysis working groups provided wide range of validation tests and validation requirements for the extragalactic catalog
- Currently 16 carefully chosen tests that cover a range of mass scales
- Leverages collective DESC expertise: Full validation suite beyond the grasp of any individual!



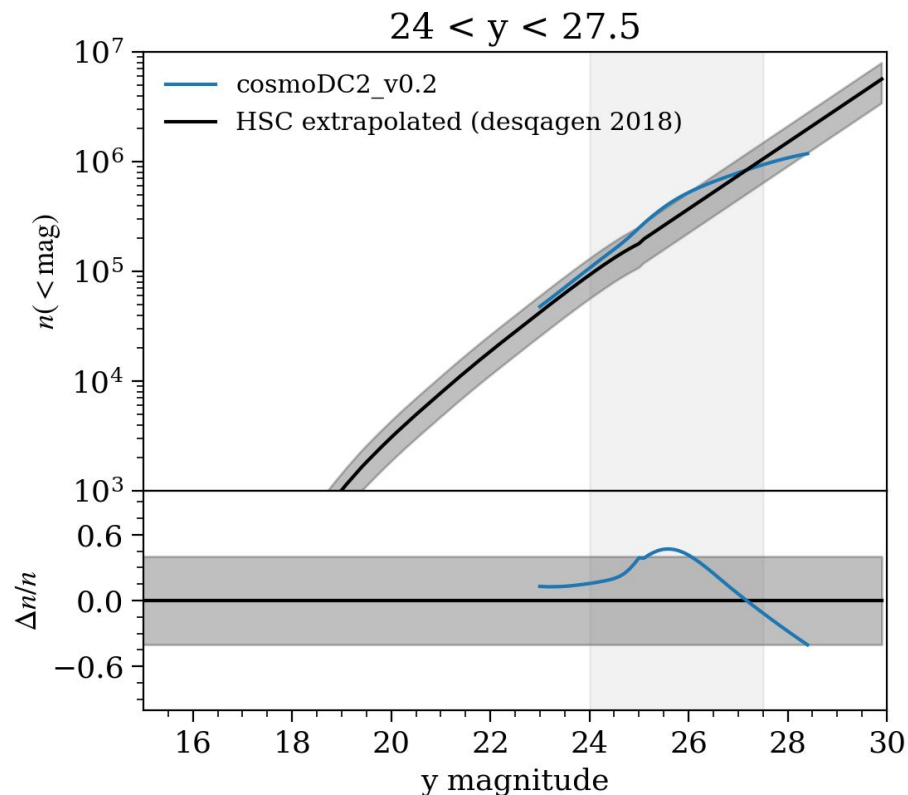
# DESCQA Validation of $dn/dmag$



## Near power-law scaling to $ugrizy < 28$

Developed new techniques to push below simulation resolution limit for this specific purpose

## Enables wide range of deblending science applications

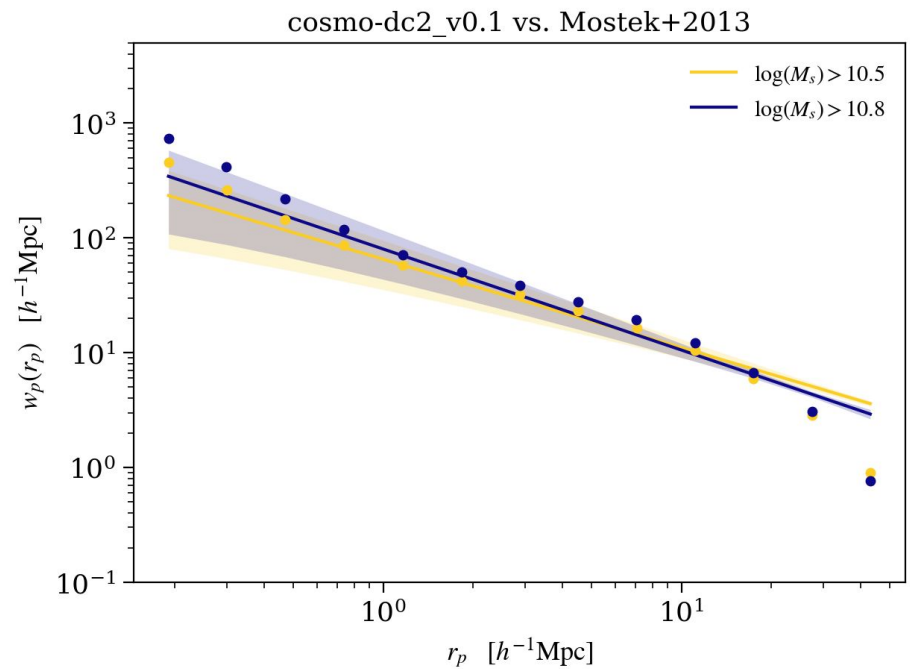
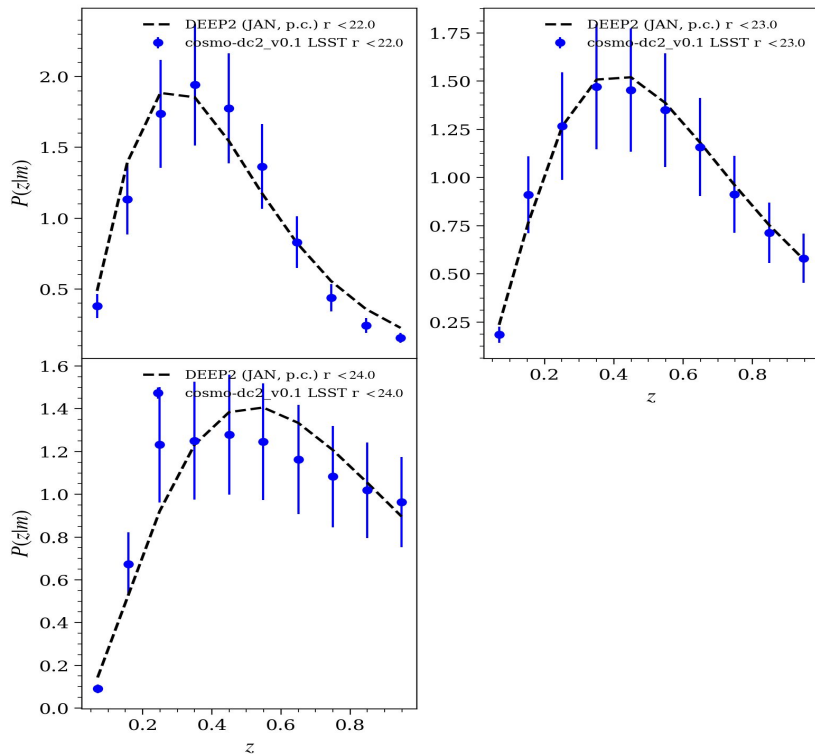


# DESCQA Validation of high-z LSS



**Good agreement with DEEP2  $dn/dz(m_r)$  &  $w_p(r_p | M_*)$**

**Ensures realistic galaxy distribution at  $z \sim 1$**



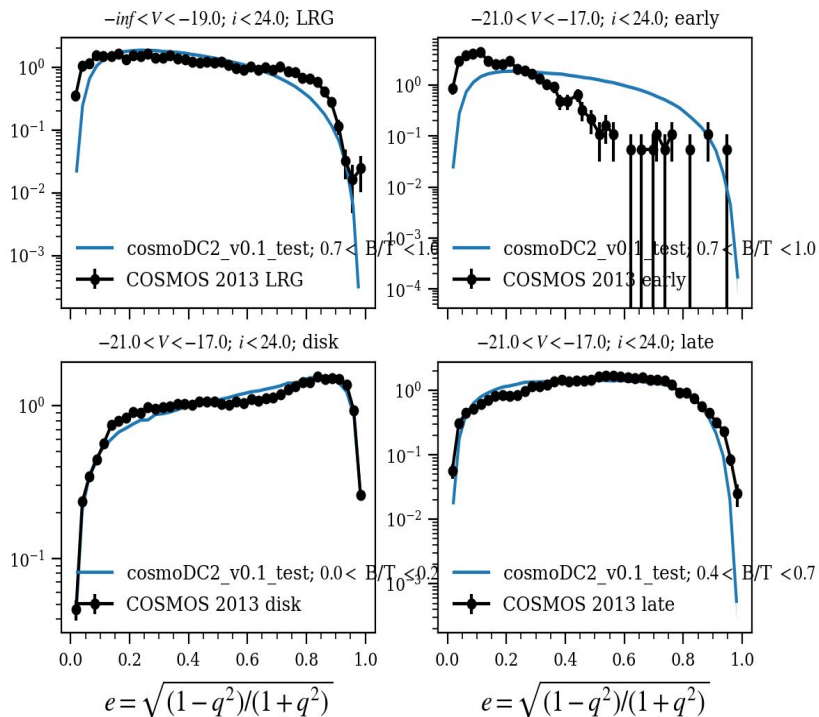
# DESCQA validation of shape & size



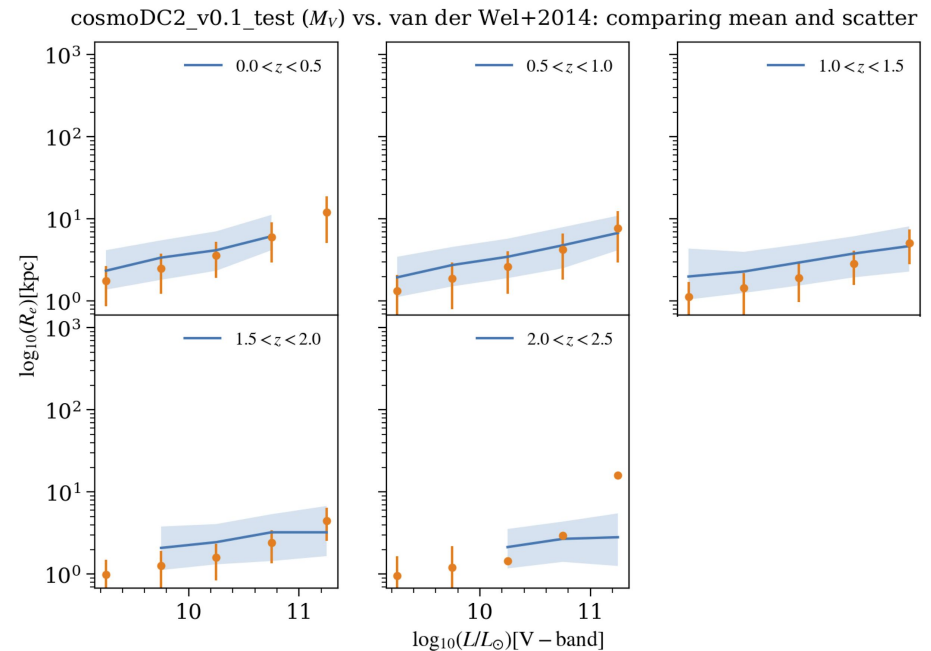
## Comparison to COSMOS

Accurate shape & size model with simultaneous dependence upon brightness, color, bulge/disk decomposition

### Ellipticity distributions



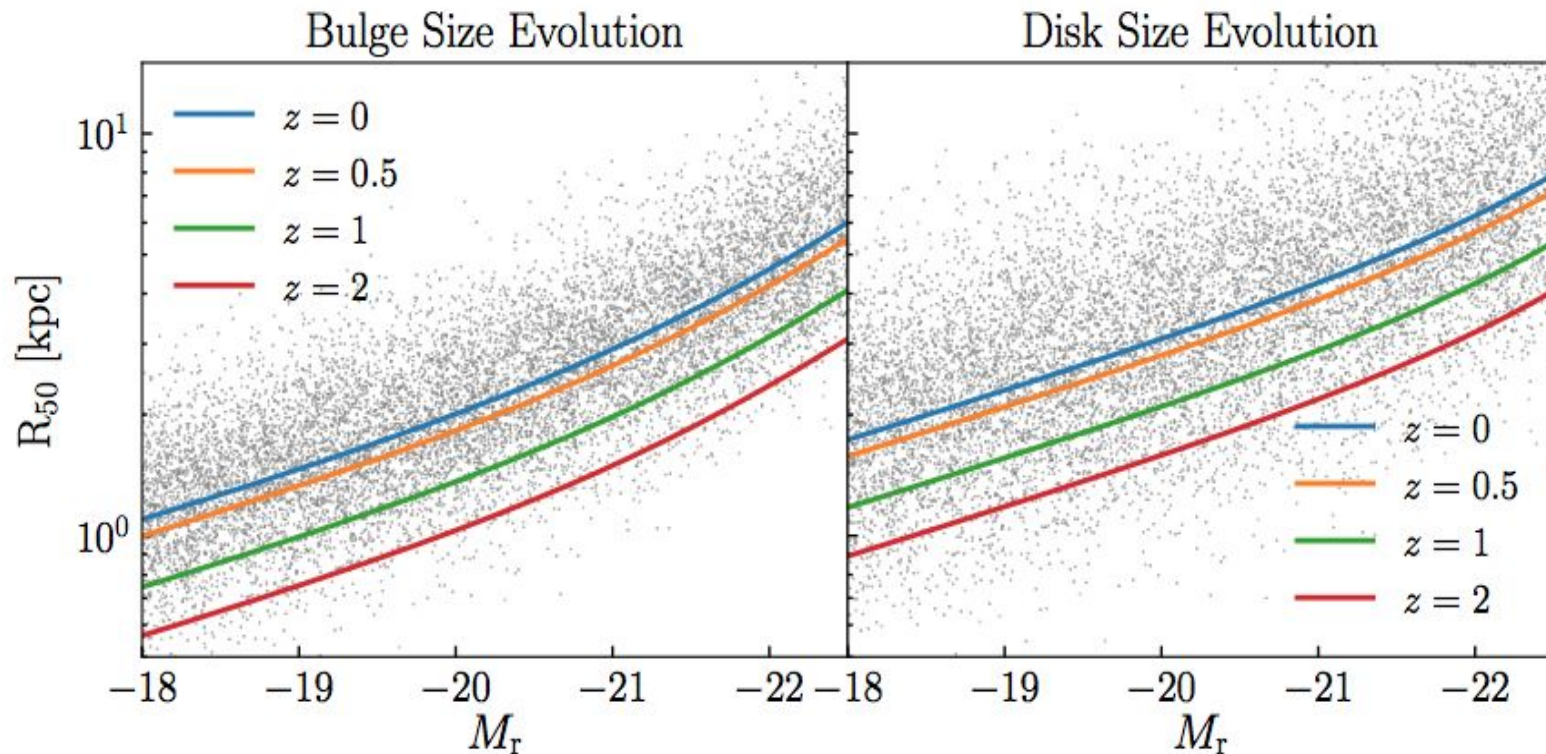
### Size distributions



# Realistic Galaxy Formation Physics

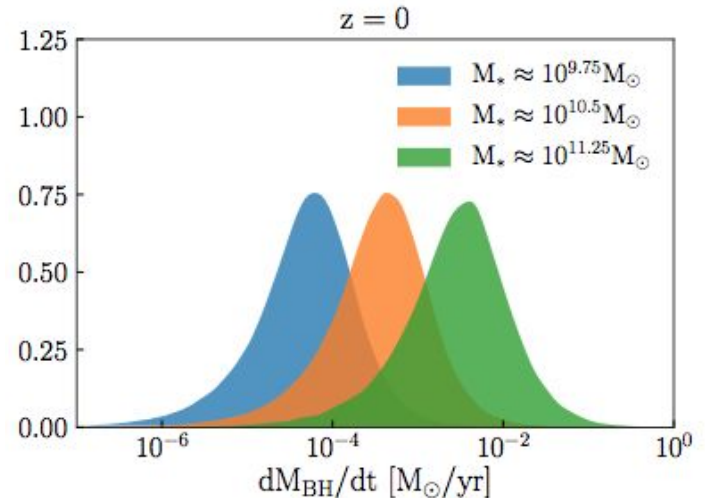
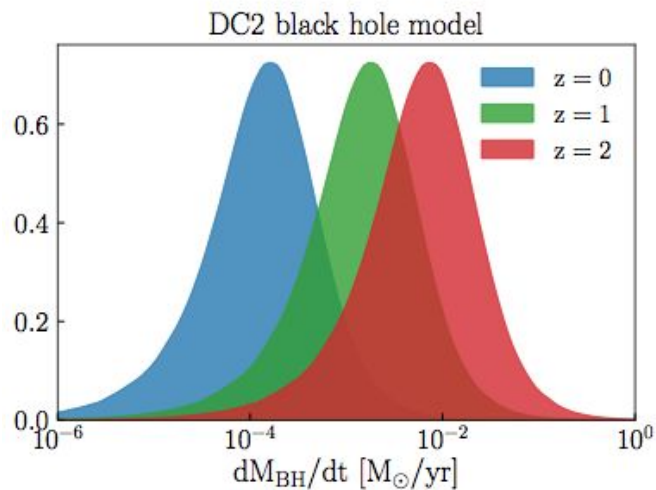
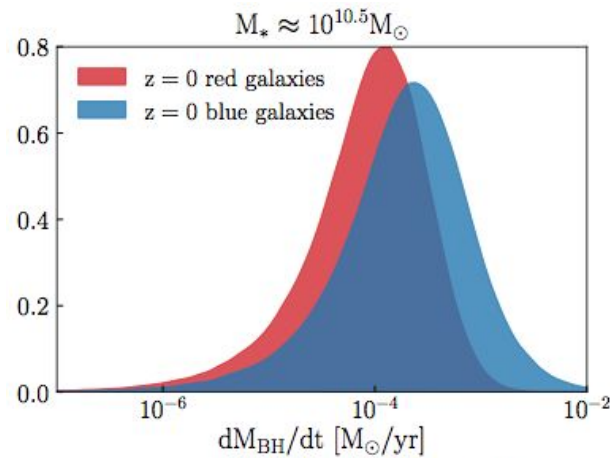


**Every single galaxy represents Galacticus solution to complex ODE system regulating evolution**  
**(way) beyond simple empirical modeling!**

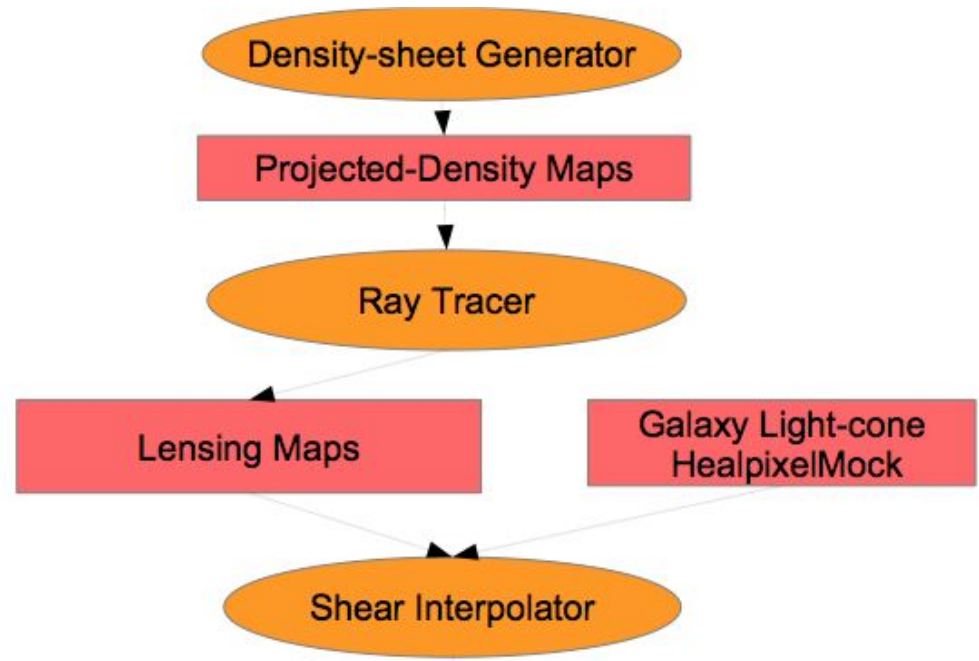
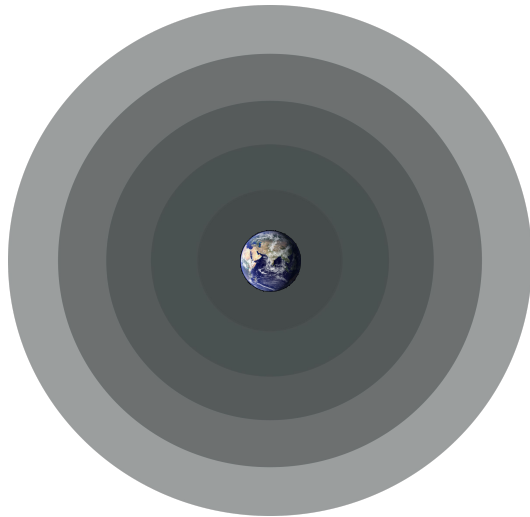


# Realistic Galaxy Formation Physics

## Black hole model with multi-dimensional correlations enhances time-domain AGN studies

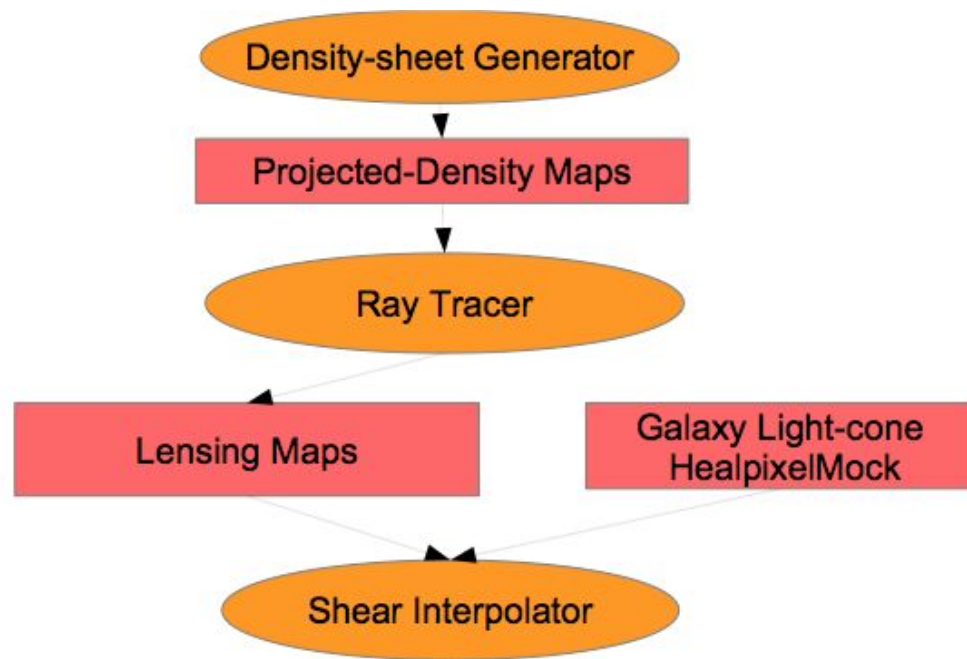
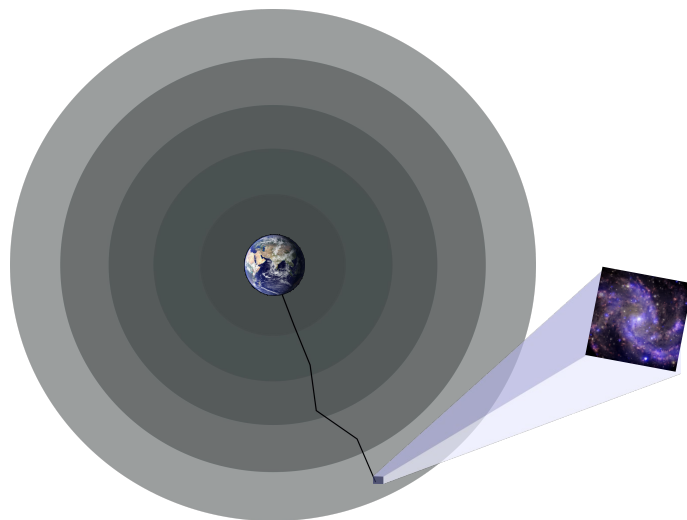


# CosmoDC2 Weak Lensing Pipeline

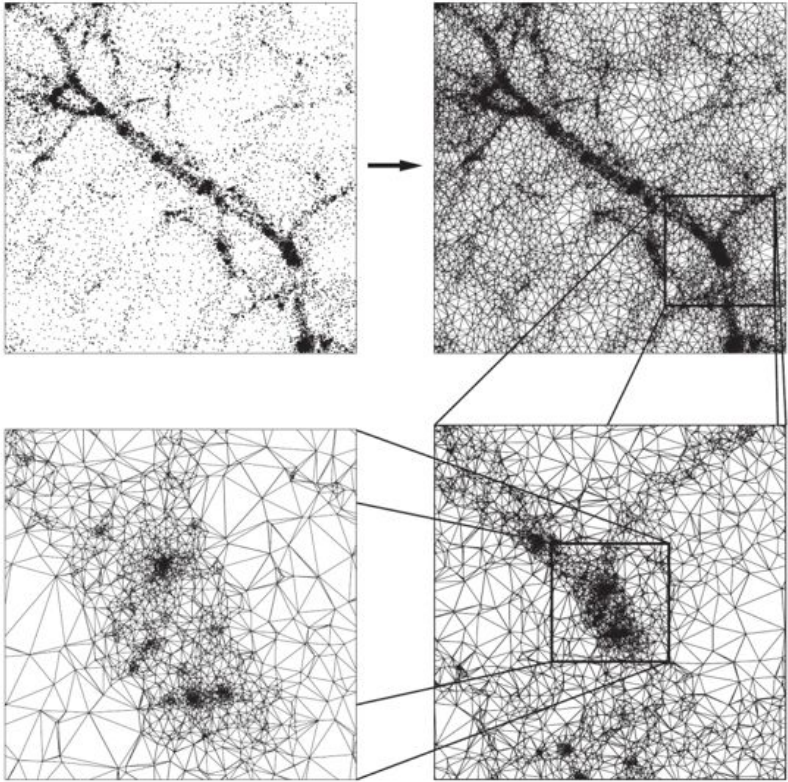
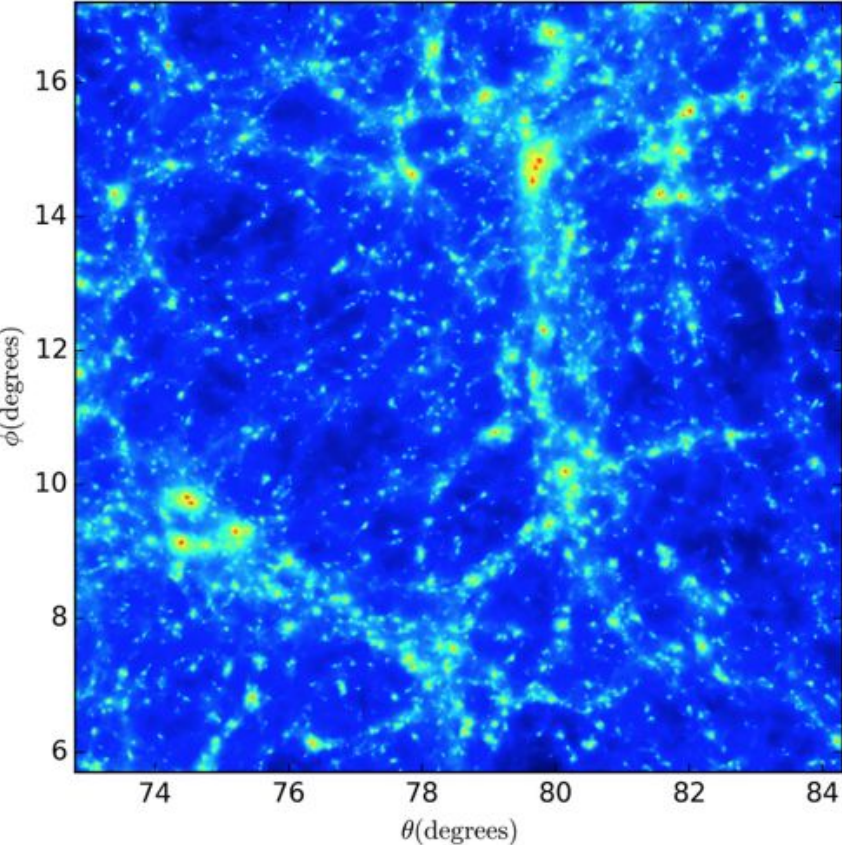




# CosmoDC2 Weak Lensing Pipeline



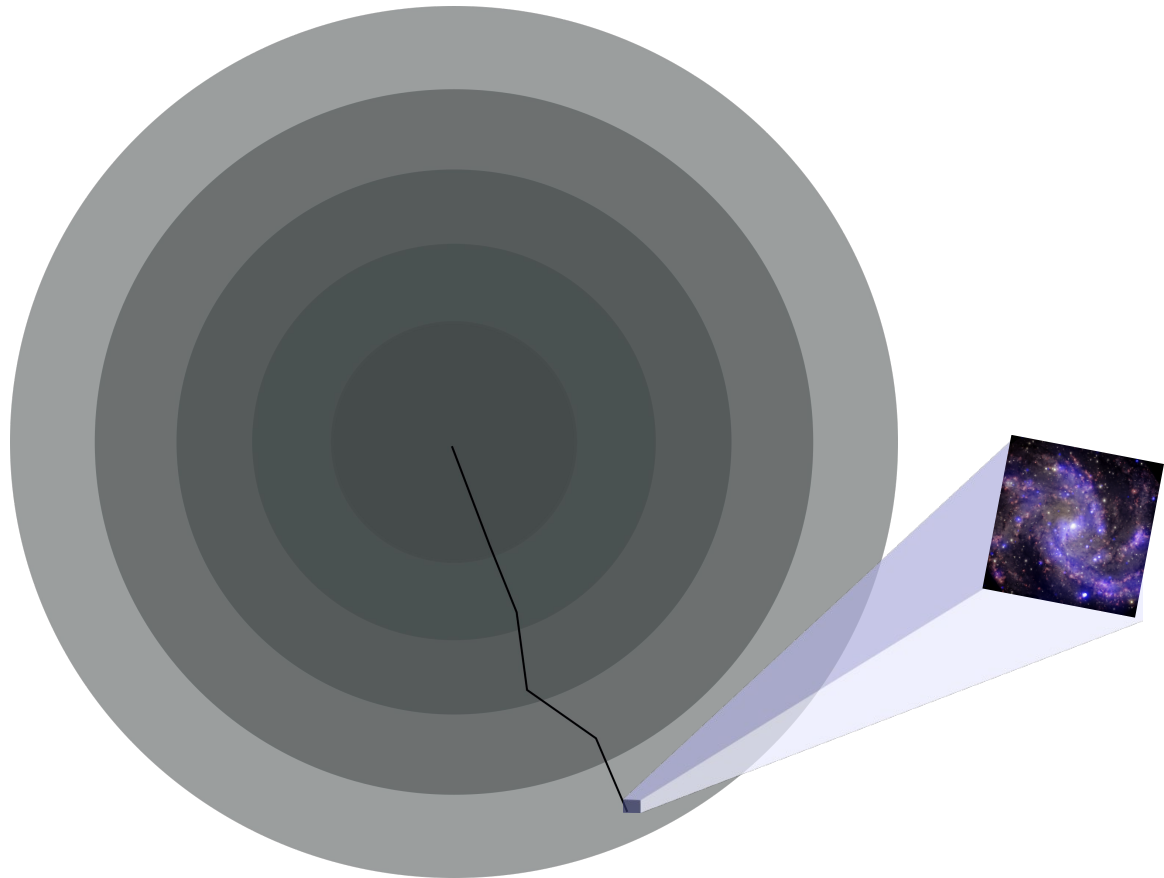
# Density Sheet Generation



Credit: Cautun et. al 2011

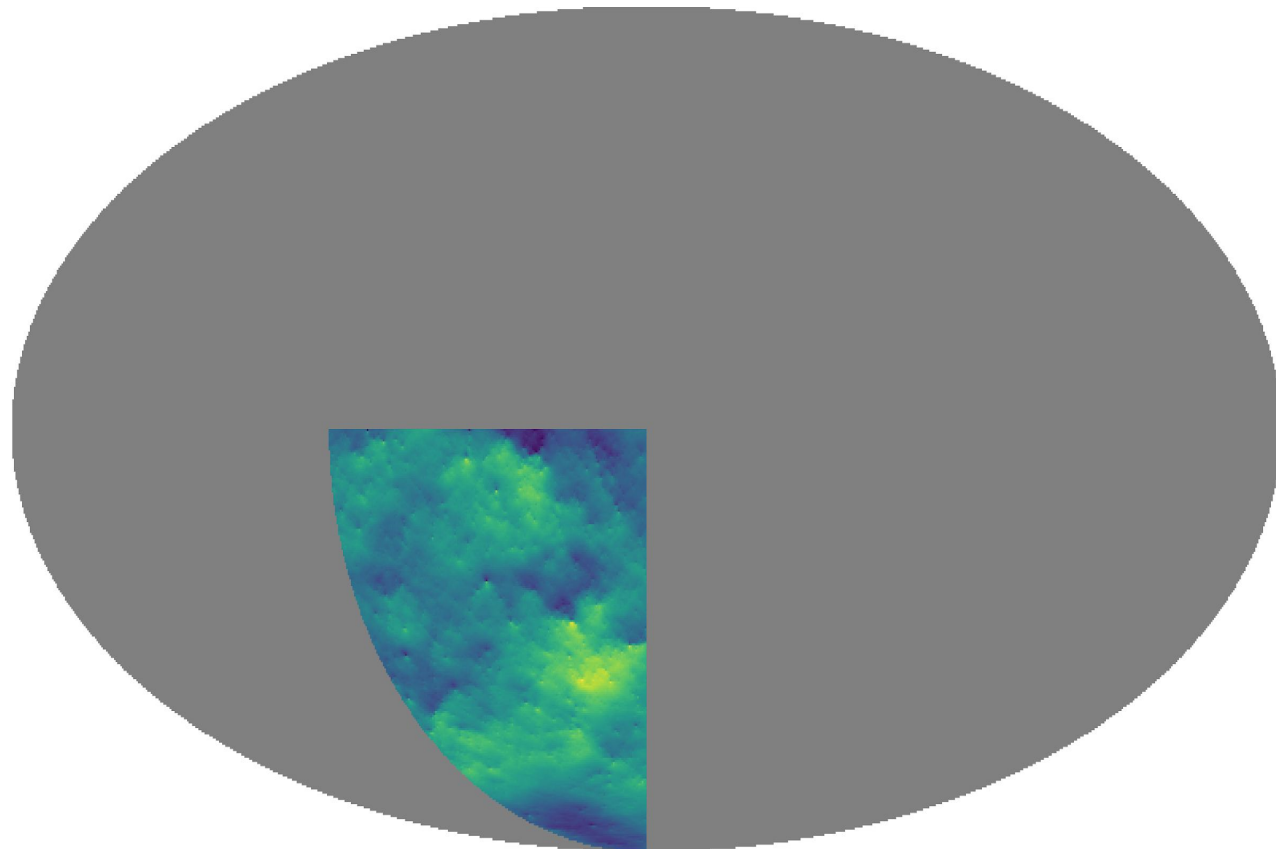
# Ray-tracing

- Lensing quantities computed using iterative ray-tracing through density shells
- Uses parallelized Lenspix routines, performing deflections on high resolution sub-grids

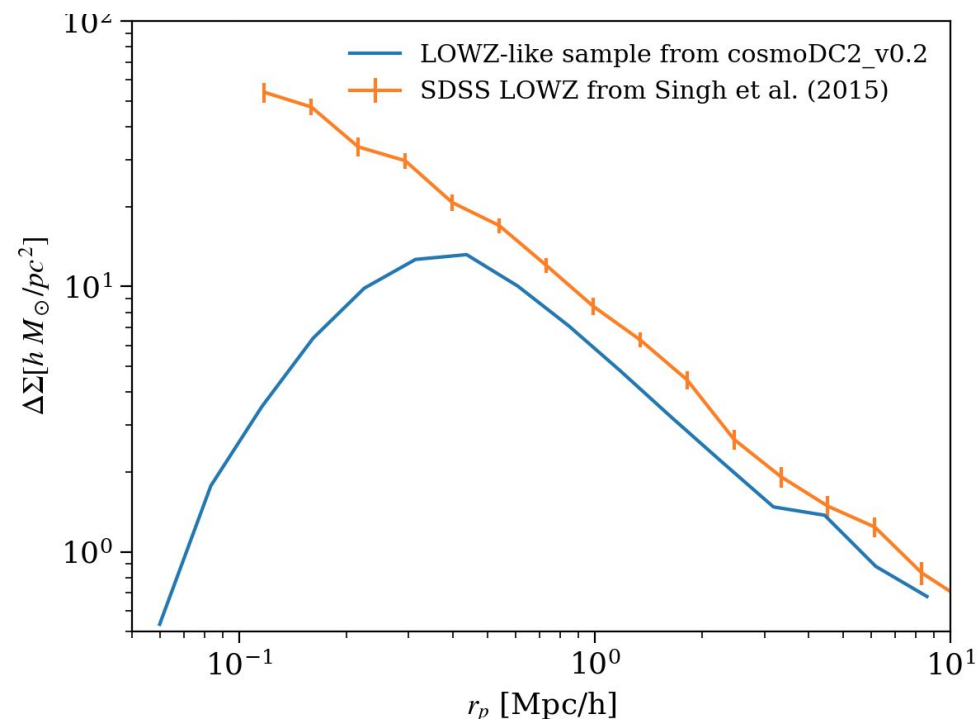


# Weak lensing maps

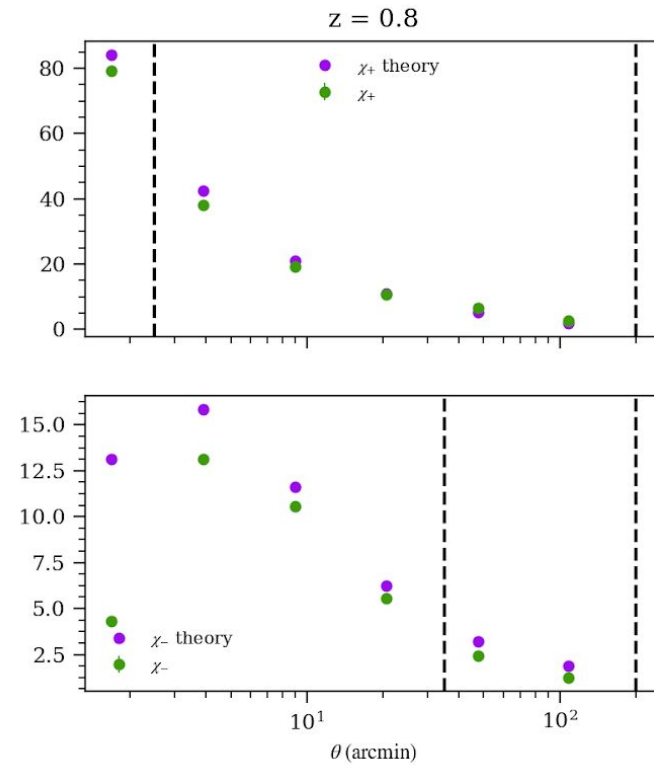
- Deflection field for source plane at  $z \sim 0.2$
- Roughly 0.85 arcminute resolution
- Interpolated to galaxy positions for catalog



# Weak lensing validation tests



F. Lanusse + DESCQA team



P. Larsen + DESCQA team

# Transients in DC2

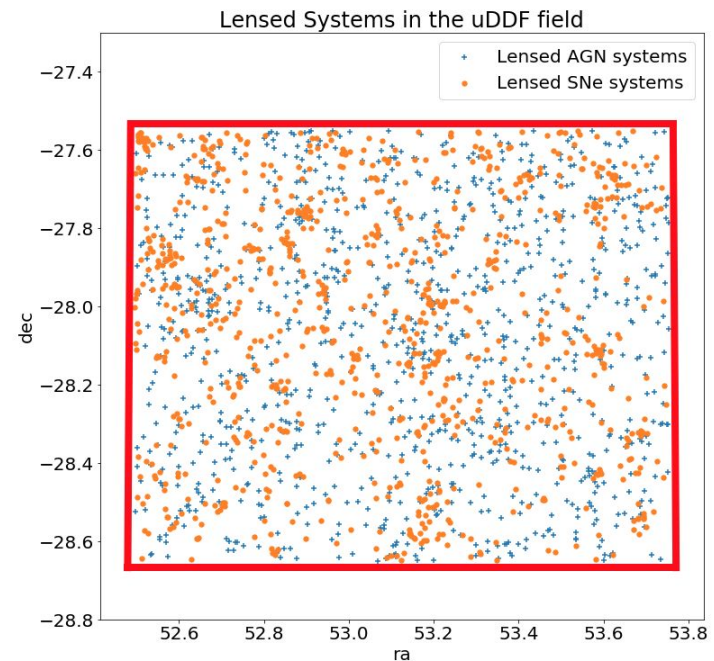
---

Bryce Kalmbach et al.



# Transients and Variables in DC2

- Main Survey
  - Galactic
    - M-dwarf flares
    - RR-Lyrae
    - Phenomenological variability models from Kepler light curves
  - Extragalactic
    - AGN
    - SNe Ia
- uDDF
  - Strongly Lensed AGN and SNe



# Main Survey Transients and Variables



- Milky Way Sources
  - Milky Way catalog comes from output of the Galfast simulation of Juric et al. (2008)
    - Stored in a database at UW
- Galactic Variable Sources
  - M-dwarf flares based upon Davenport et al. (2014)
  - RR Lyrae from Sesar et al. (2009)
  - All other stars were matched to Kepler light curves
    - Based upon color-magnitude diagram position ( $r$  vs  $g-r$ )



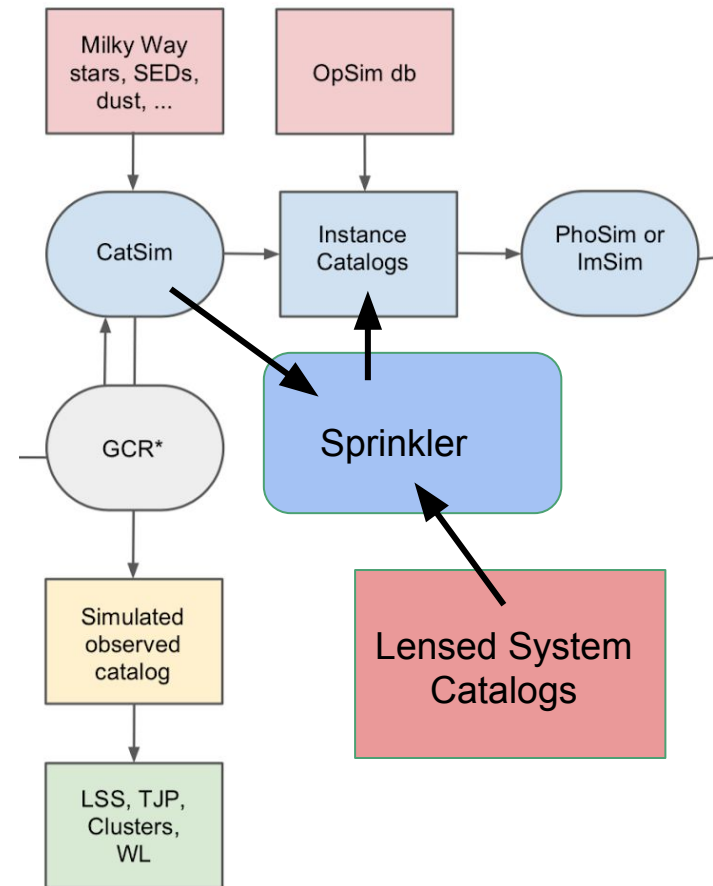
# Main Survey Transients and Variables



- AGN
  - Matched to galaxies in the extragalactic catalogs provided by cosmological simulations group
    - Variability modeled as a damped random walk based upon MacLeod et al. (2010)
- SNe
  - All type Ia
  - Added to extragalactic catalog hosts based upon total stellar mass and light distribution
  - Numbers boosted in uDDF region

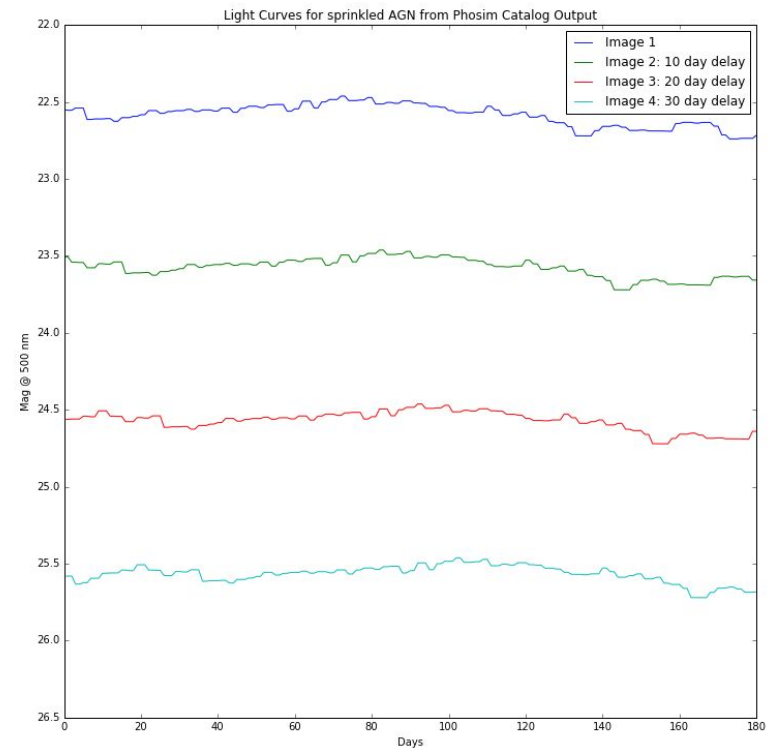
# DESC Sprinkler

- Strongly Lensed Systems are inserted using the DESC Sprinkler code
  - Matches to existing objects in CatSim catalog and replaces them with lensed systems



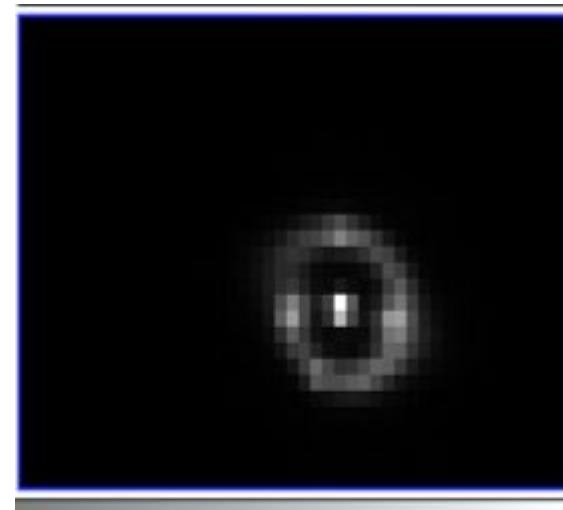
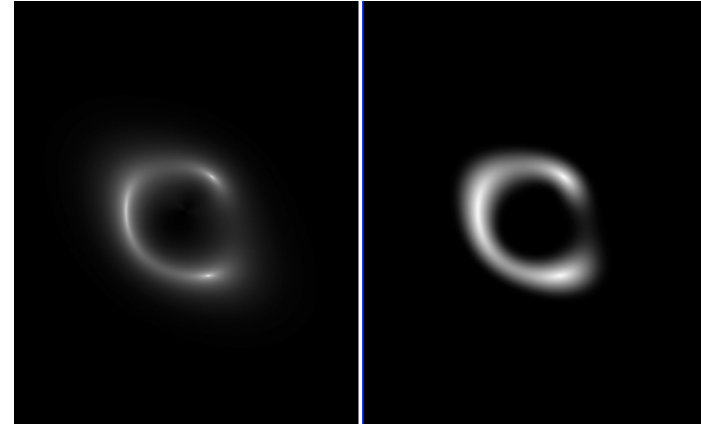
# uDDF Strongly Lensed Objects

- Strongly Lensed objects appear only in uDDF
  - AGN
    - From Oguri and Marshall (2010) catalog
  - SNe
    - From Goldstein et al. (2017) catalog
- ~1000 of each type of system added
- Time delays included



# Adding Host Galaxies

- Working on including host galaxies in lensed images
  - Work by Nan Li and Matt Wiesner
  - Use lens system properties to properly lens the host galaxies
  - Finishing up to get this into DC2



# The First Image Simulation Runs

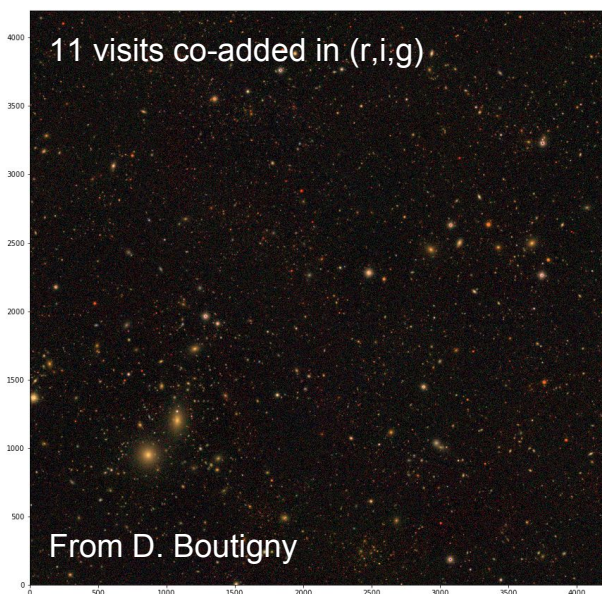
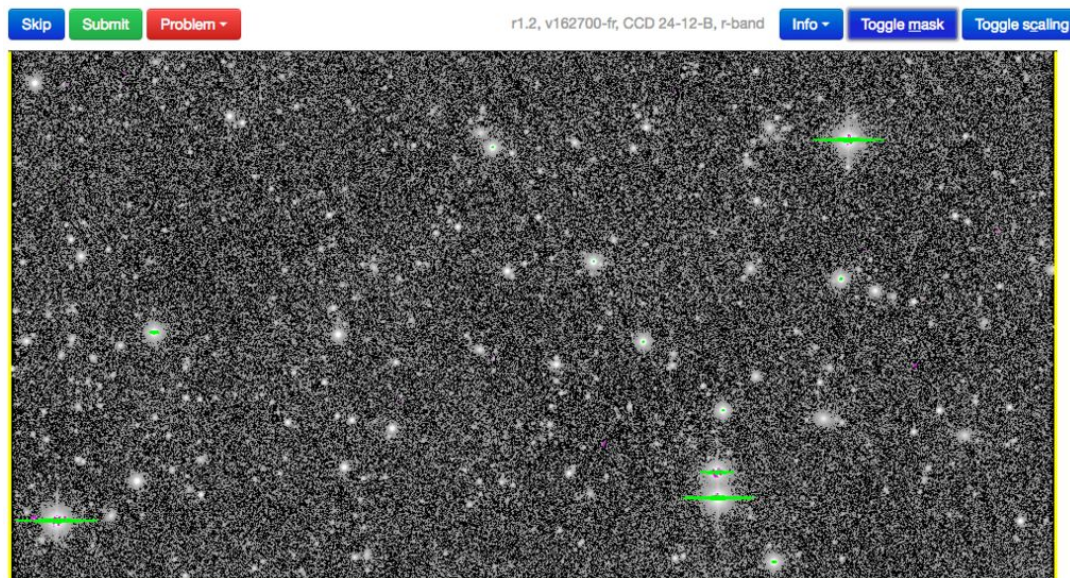
---

Jim Chiang et al.



# DC2: Simulated LSST Images

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

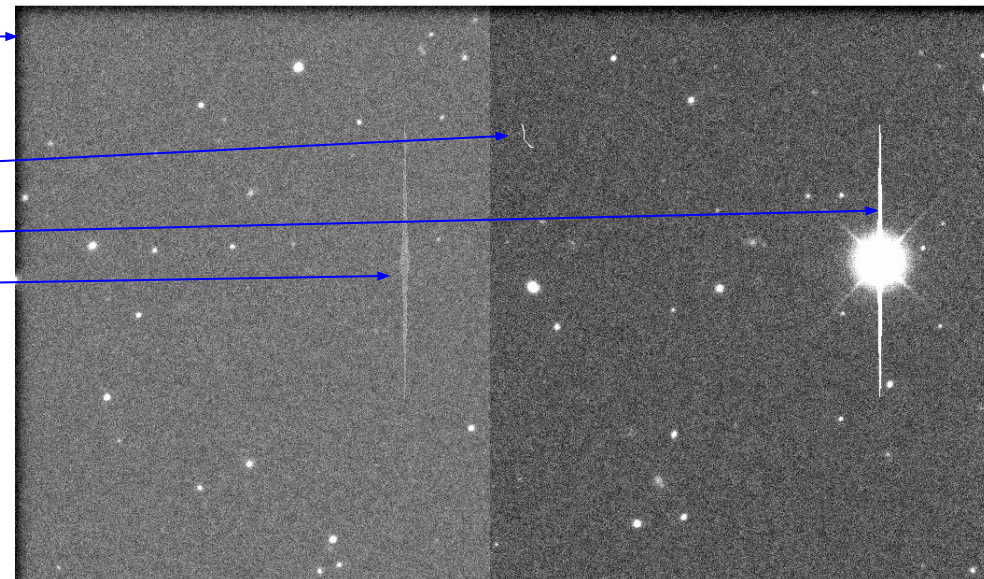
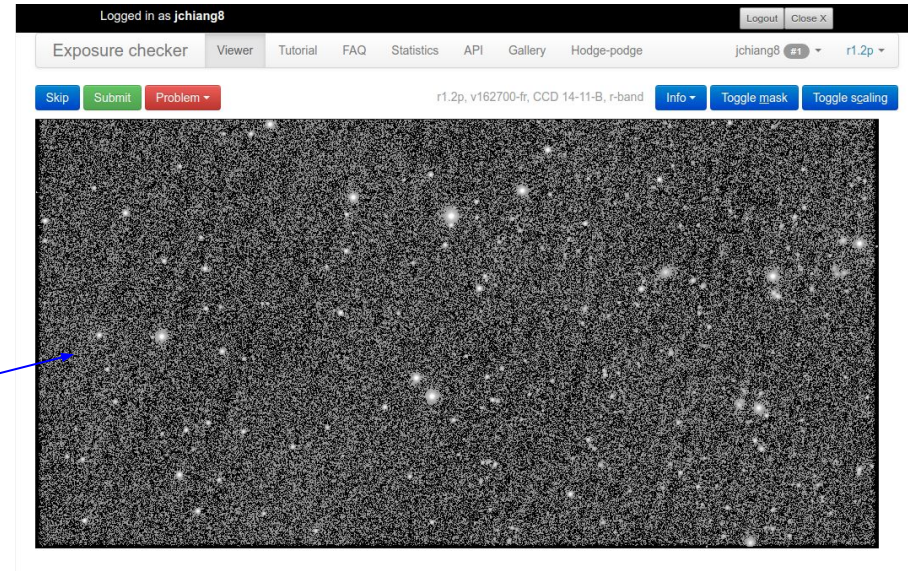


Second image simulation tool; ImSim uses GalSim to produce LSST-specific simulations with tunable levels of complexity.

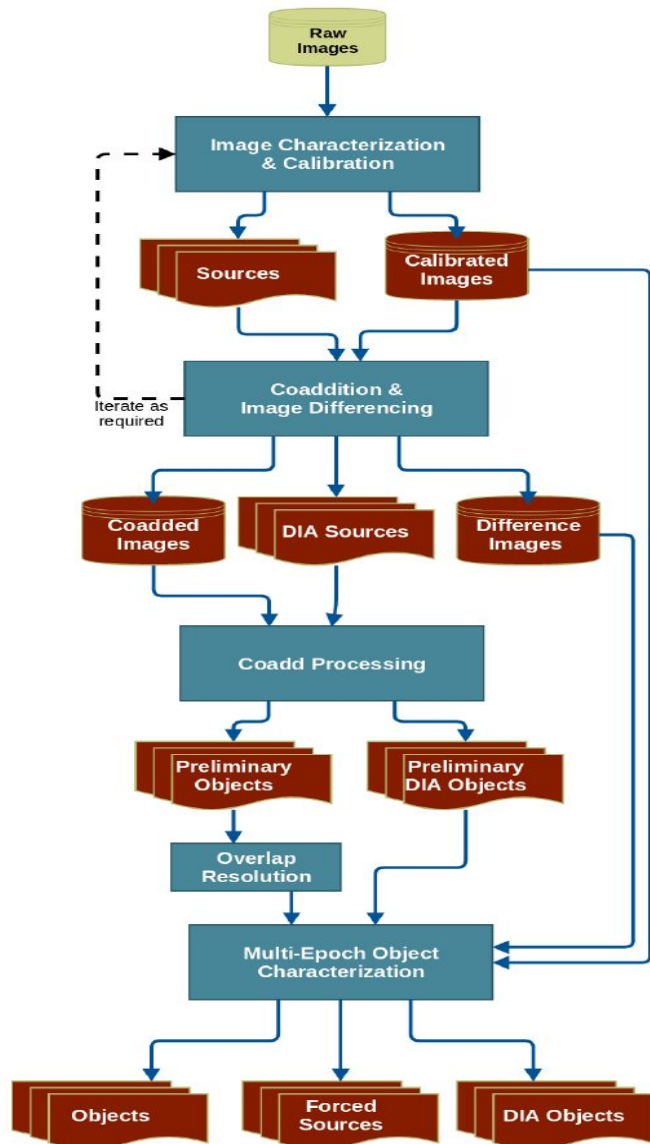
We use both codes, with a plausible variety of astrophysical and observational effects, to stress-test our analyses

# Image Simulation Properties

- Objects
  - DC2 galaxies (clustering, WL)
  - AGNs
  - Stars from CatSim
  - Sprinkled time-varying objects (lensed AGNs, SNe)
- PSF modeled with atmospheric effects and optical distortions
- Sensor effects:
  - Treerings
  - Brighter-fatter
  - Edge rolloff
- Instrument signature features
  - Vignetting
  - Cosmic Rays
  - Saturation and bleed trails
  - Crosstalk
  - Bright/dark defects
  - Charge transfer inefficiency



# 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

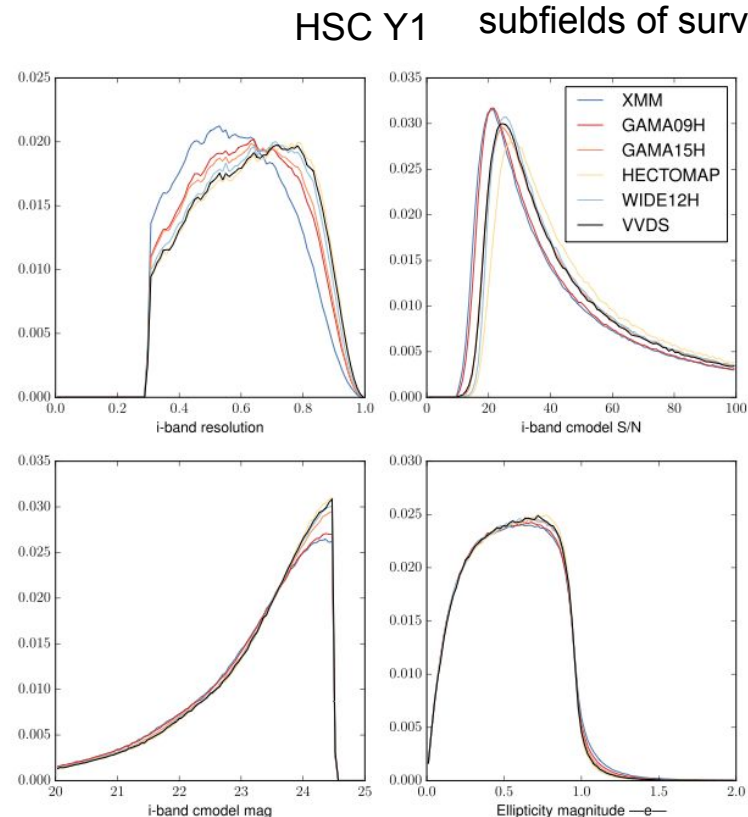
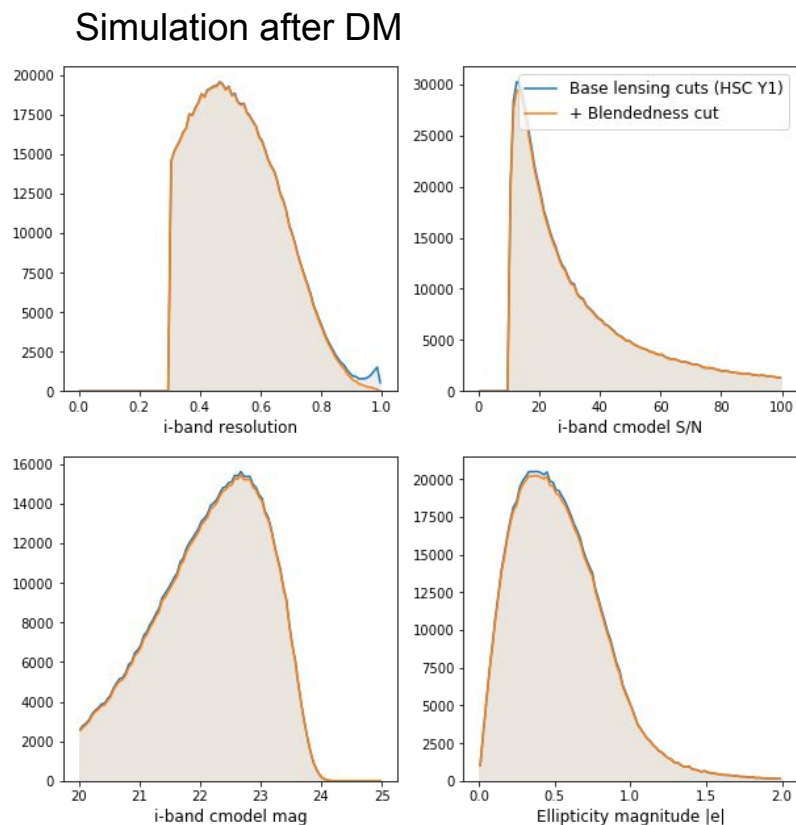


# Data Challenge 2: Validation Tests



Work by Francois Lanusse (CMU)

Colors: Different subfields of survey



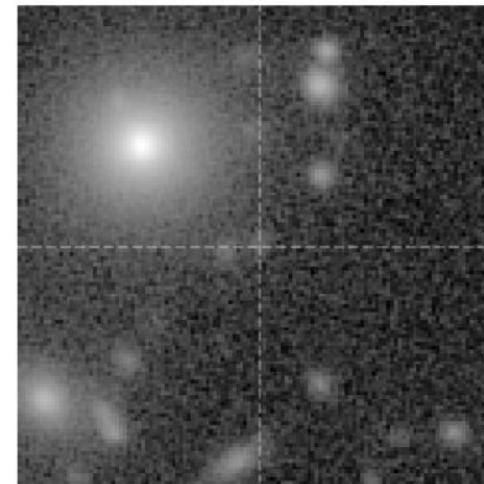
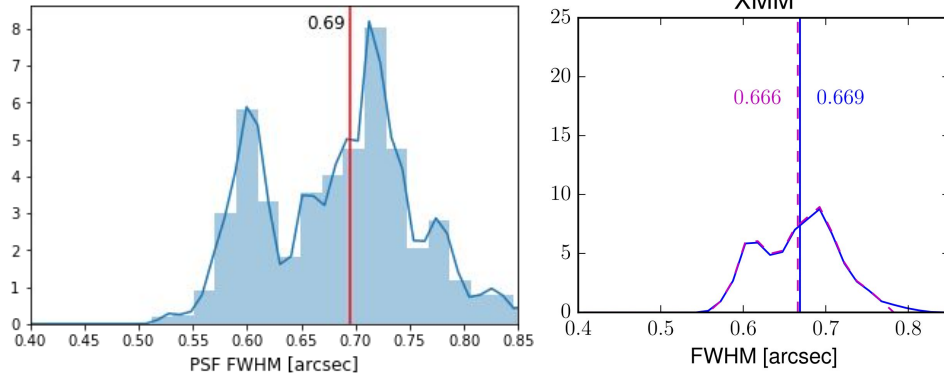
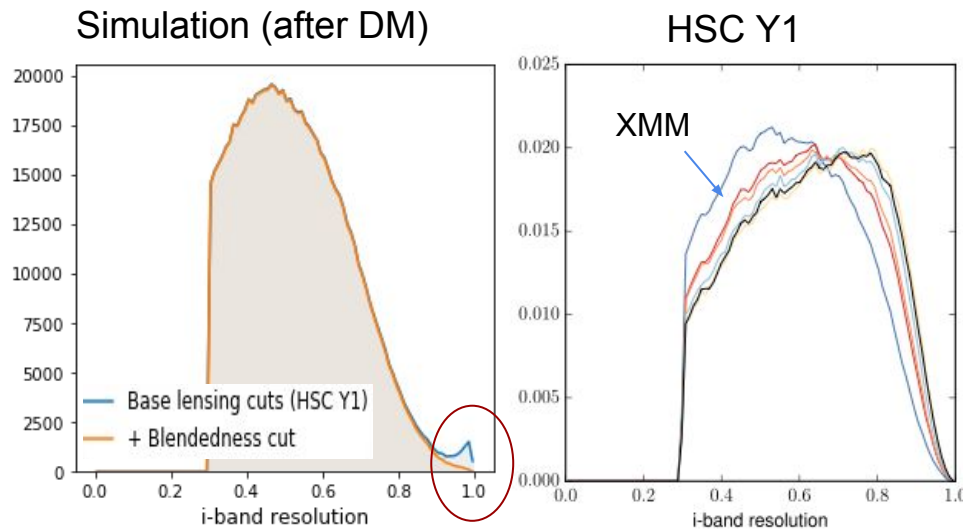
Comparison to HSC Y1 shape sample (Mandelbaum et al. 2017)

- Resolution: Indicator of galaxy size compared to size of PSF:  $R = 1 - \text{PSF\_size} / \text{gal\_size}$
- Galaxy properties look very reasonable!

# Data Challenge 2: Validation Tests

Work by Francois Lanusse (CMU)

- PSF size distribution closest to the HSC XMM field, resolution for that field is quite close to Run1.1p
- Effect of blending (and deblending failures) visible in size distribution



Example of galaxy with  $R=0.99$  and blendedness = 0.97

Removed by blendedness cut

Simulation!

## Resolution distribution

# DC2 status



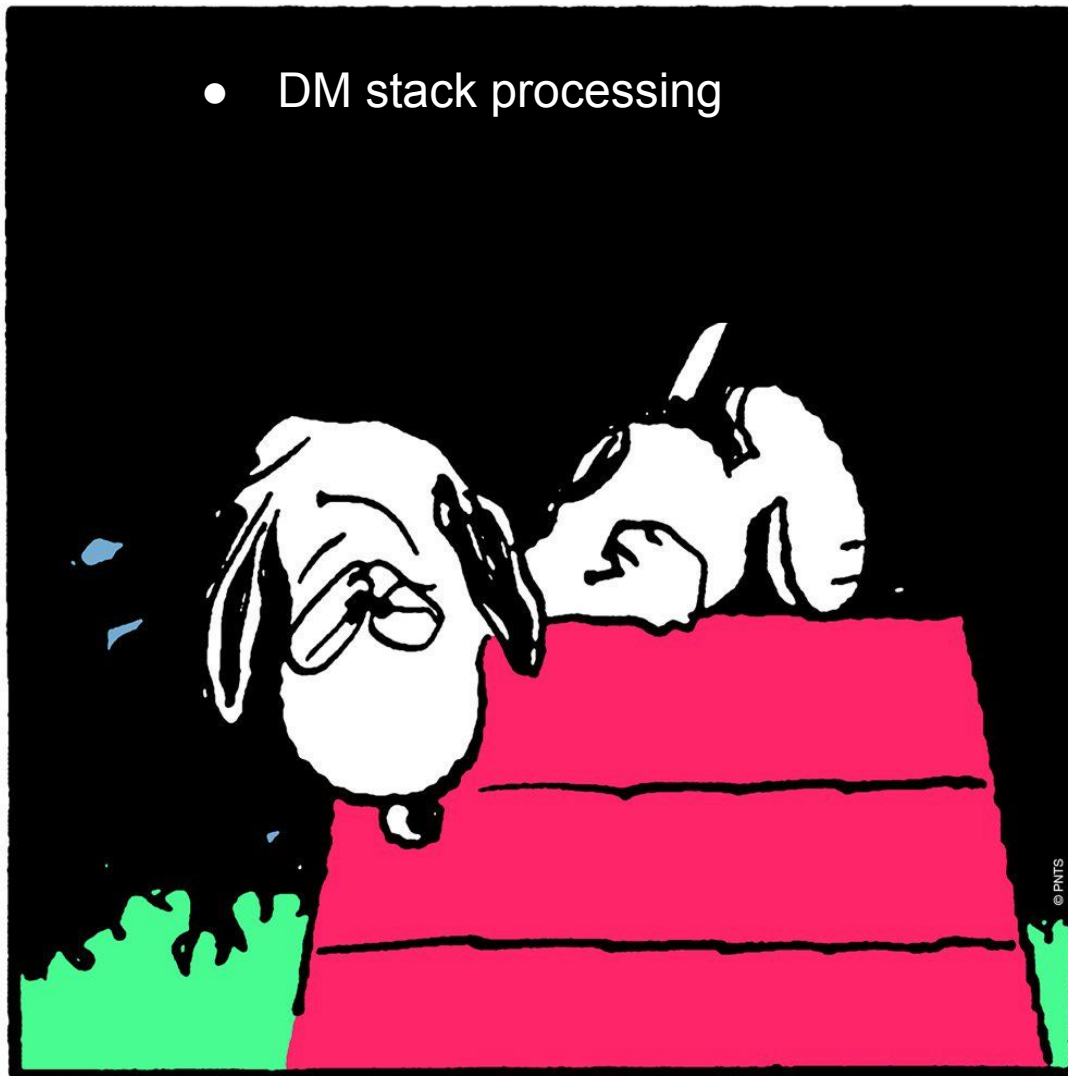
- We had an *initial internal release* of processed images and DM catalogs at the last DESC collaboration meeting (last week) at CMU
- Release included range of notebooks and tutorials to teach the collaboration how to access LSST data products
- First version of the “CosmoDC2” extragalactic catalog was released
- The DC2 data will be used to drive early development of the DESC end-to-end analyses, inferring cosmological parameters from LSST data
- 20+ DC2 projects have been announced already

# What keeps us up at night ...



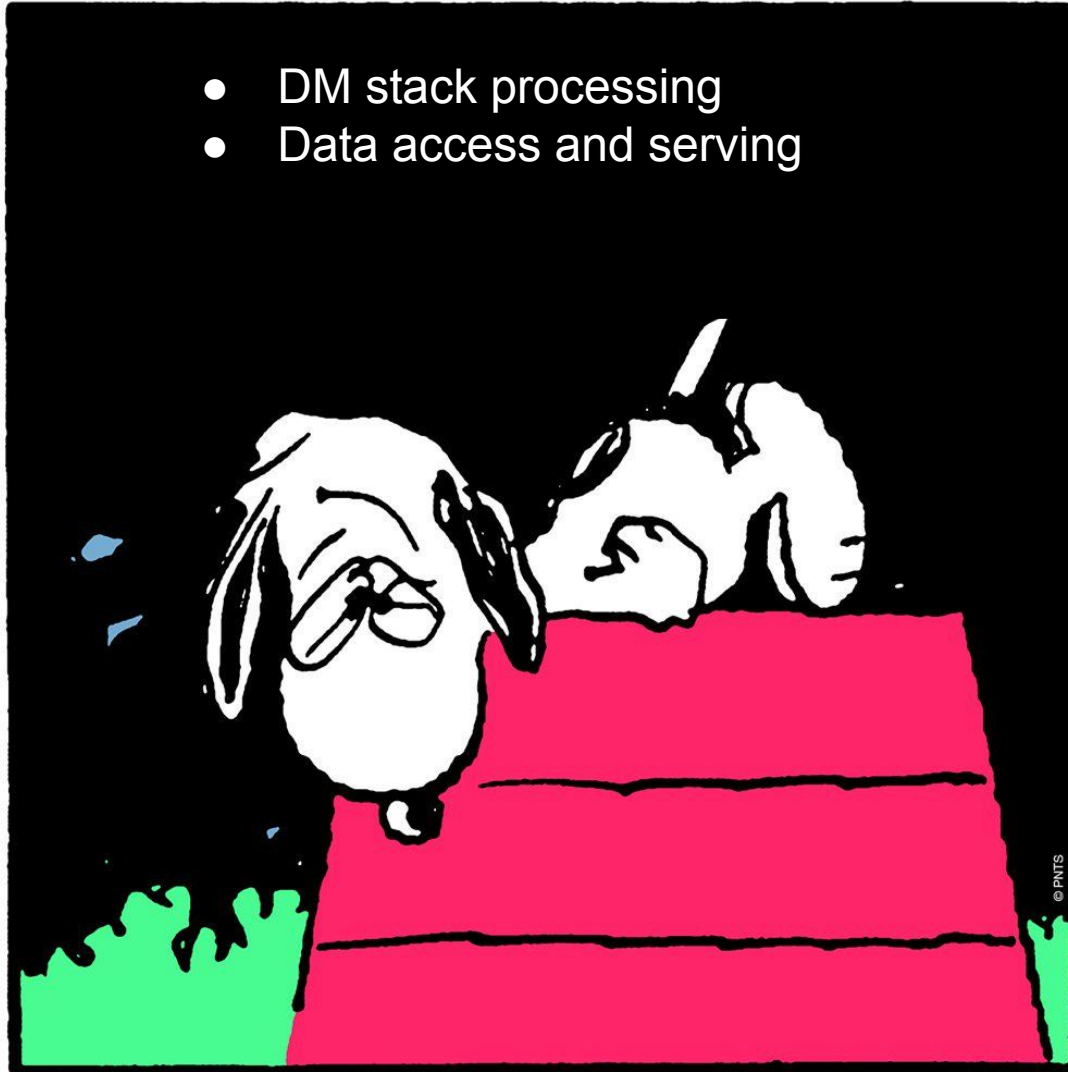
# What keeps us up at night ...

- DM stack processing



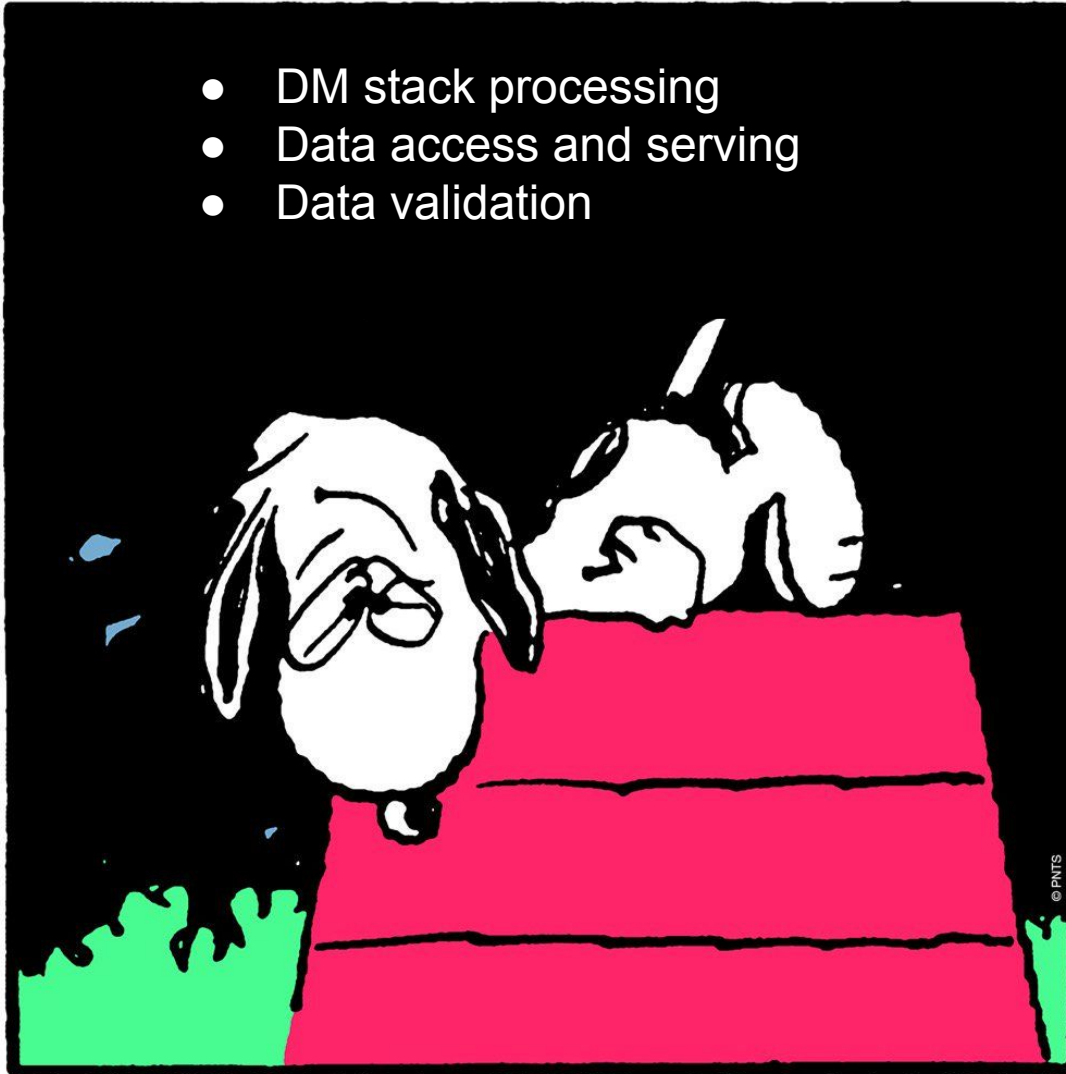
# What keeps us up at night ...

- DM stack processing
- Data access and serving



# What keeps us up at night ...

- DM stack processing
- Data access and serving
- Data validation



# DESC welcomes new members!





# Back-up Slides

---



# Total Planned Visits



WFD: minion\_1016 baseline cadence for WFD visits

	u	g	r	i	z	y	Total
Run 1.1 WFD	572	822	1825	1855	1669	1646	8389
Run 2 WFD	3468	4873	11096	11133	9911	9934	50415
uDDF	4963	1911	3796	3818	4930	3742	23160

500% more WFD visits in Run 2!

More about the uDDF visits now.