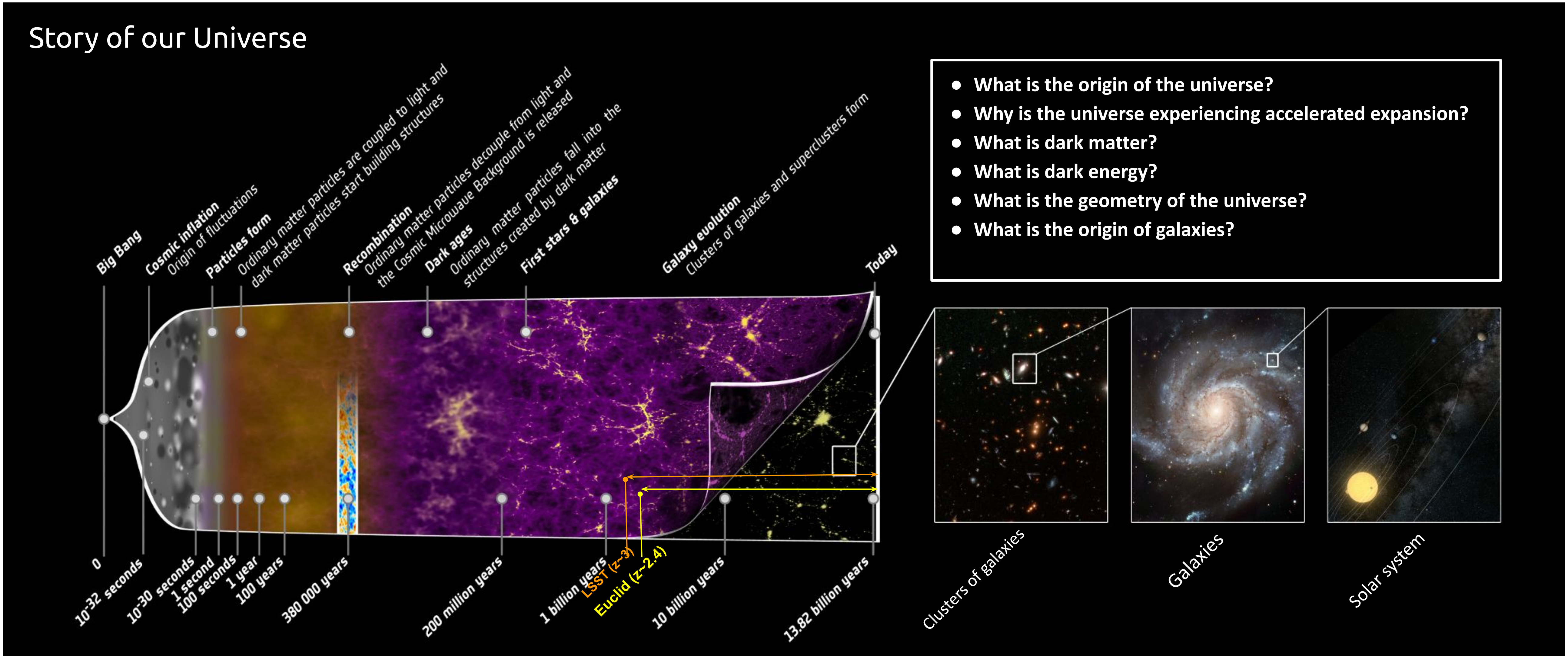
Schedule

- Introduction + Redshift Galaxy Survey
- Mock Galaxy Catalogues
- Halo Occupation Distribution model (HOD)
- Practical exercises
  - How to produce a population that follows an specific distribution
  - Estimate the halo mass function of MICECATv1.0
- (some) References

Jorge Carretero Palacios (carretero@pic.es)

Fundamental questions



Credit: Nasa web page

Perspectives

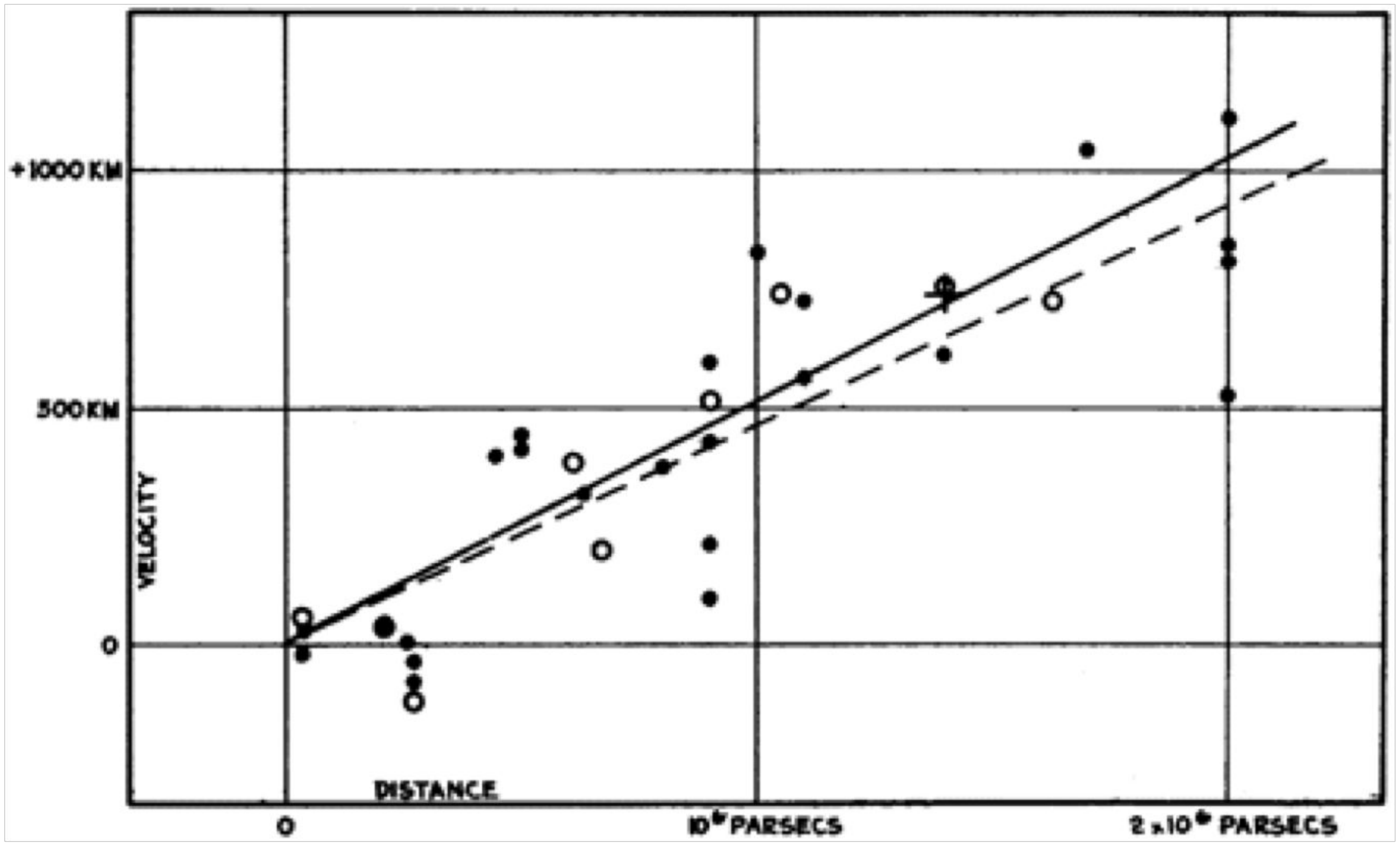
- From the Astronomy perspective:
  - The Astronet Science Vision and Infrastructure Roadmap 2022-2035
  - Pathways to Discovery in Astronomy and Astrophysics for the 2020s
- From the Astroparticles perspective:
  - European Astroparticle Physics Strategy 2017-2026
    - Mid-term update (2023) (DESI, LSST and Euclid)
  - Pathways to Innovation and Discovery in Particle Physics

Decisive moments

- **Hubble**, using cepheid stars and computing a relation between distance and redshift, realized that almost all the studied spectra of the selected **galaxies** appeared to be **moving away from us** (**1929**)

- Two independent scientific groups, the Supernova Cosmology Project and the High-z Supernova Search Team, through observations of distant supernovae Ia, discovered the **accelerating expansion of the Universe** (**1998**)

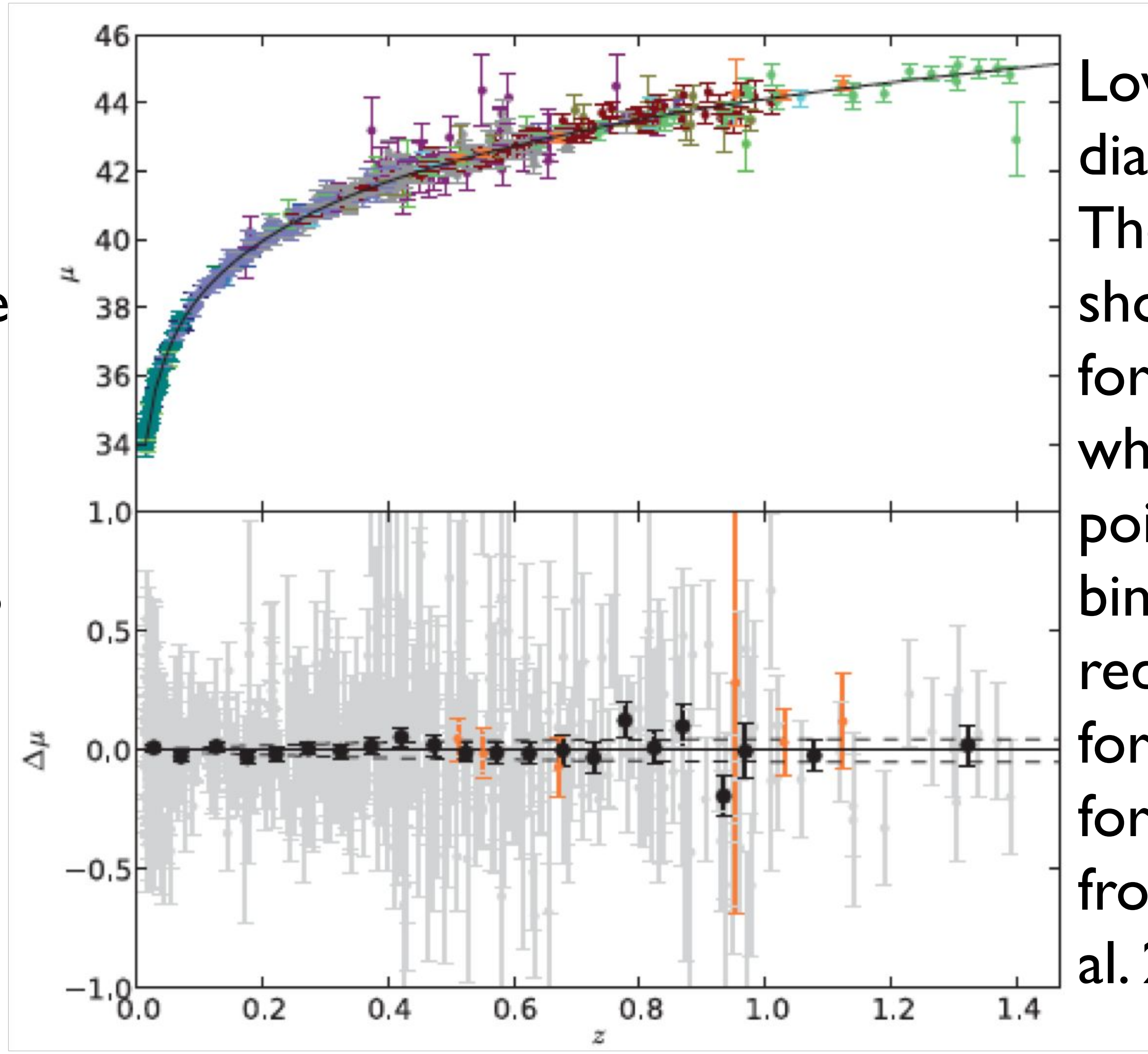
Hubble 1929



Velocity-Distance relation among Extra-Galactic Nebulae. Figure from Hubble 1929 (http://www.pnas.org/content/15/3/168.full.pdf)

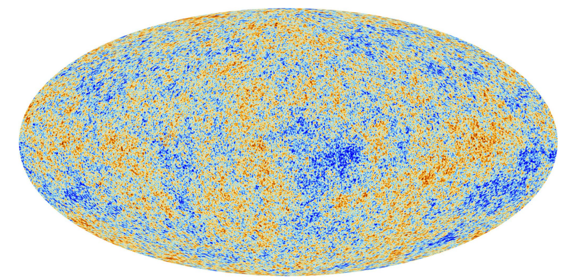
SNe Hubble diagram

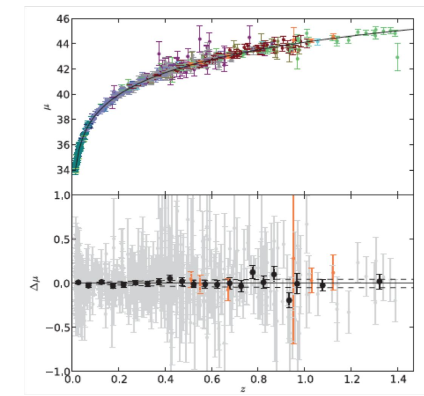
Upper panel: Hubble diagram for 557 SNe. Solid line represents the best fitted cosmology for a flat universe including the CMB and BAO constraints. Colors represent different data sets



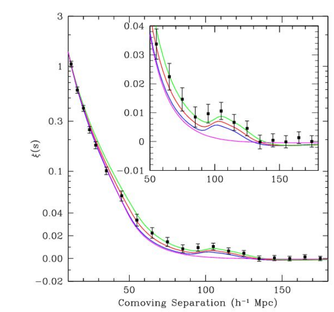
Lower panel: Hubble diagram residuals. The gray points show the residuals for individual SNe, while the black points show the binned values in redshifts bins of 0.05 for $z < 1.0$ and 0.2 for $z > 1.0$. Figure from Amanullah et al. 2010

Cosmological probes

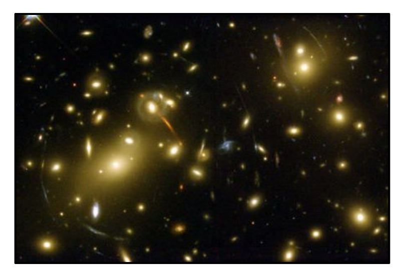


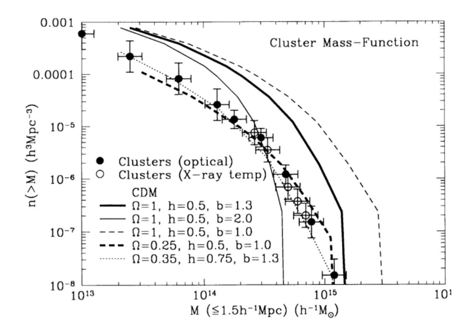


- Cosmic Microwave Background
- Supernovae Ia (standard candle)
- **Galaxy clustering**: matter power spectrum, baryon acoustic oscillations (standard ruler), redshift space distortions

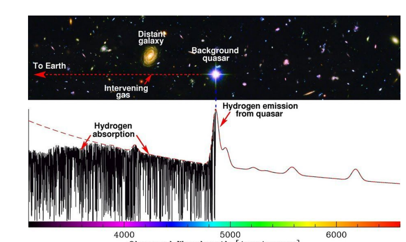


- **Weak lensing**





- Galaxy cluster abundance
- Lyman-alpha forest



- And more

The question is no longer whether the Universe is accelerating, but why (Weinberg et al. 2012)

Cosmological model

- The standard cosmological model is based on:
  - **Cosmological principle**:
    - our location is not special
    - on large enough scales the universe looks homogeneous in all direction and from every location
  - **Theory of General Relativity**

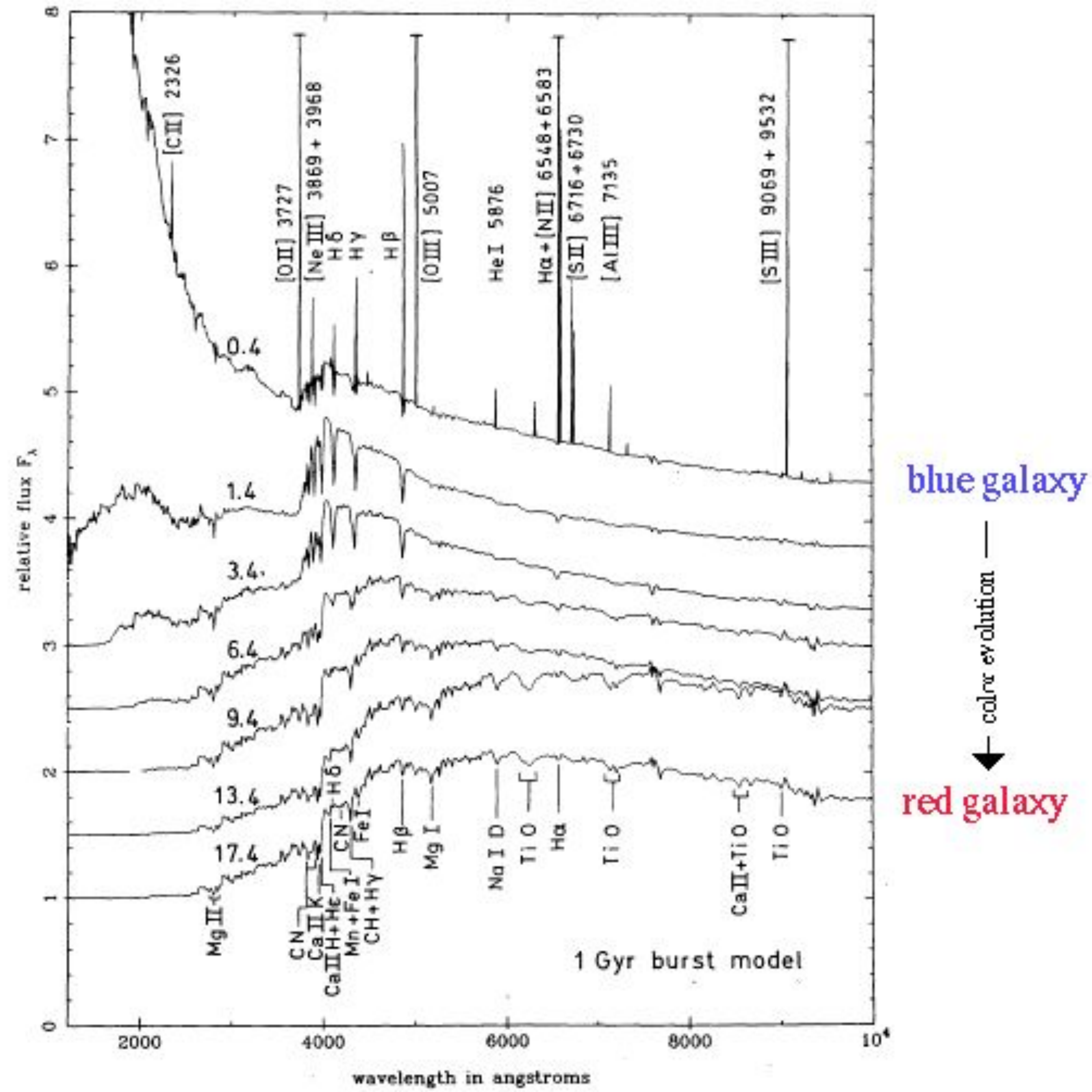
- Many different alternatives to explain the accelerating expansion. Either general relativity fails or there exists an entity (dark energy) with peculiar physical properties

Redshift galaxy surveys

- One of the most powerful tools to understand dark energy and galaxy formation
- Main goal: measure the positions of distant galaxies
- Two main possibilities to estimate the redshift of a galaxy:
  - **spectroscopy**: high precision galaxy spectra BUT more exposure time and galaxy positions known a priori
  - **photometry**: less precision than spectroscopy BUT less exposure time and no knowledge of target positions

Galaxy spectra

Single
Stellar
population
model



Galaxy spectra

- Combination of ~ billion stars + molecular clouds + star-forming regions
- We can basically distinguish 3 aspects:
  - **Strong continuum**
  - **Absorption lines**
  - **Emission lines**

Galaxy spectra

- **Continuum**:
  - combination of many Black bodies spectra spanning a range of temperatures
  - fairly flat spectrum
  - main feature: the 4000A-break caused by blanket absorption of high energy radiation from metals in the stellar atmospheres and the lack of hot blue stars. Then:
    - Ellipticals: strong break. Also characterized by strong absorption lines due to metals in the stellar atmospheres of the low luminosity stellar population
    - Spirals: weak break. Strong emission lines due to hot young stars, also absorption features due to the older underlying stellar population
    - Irregulars: no break. Strong emission lines due to hot young stars and surrounding HII regions

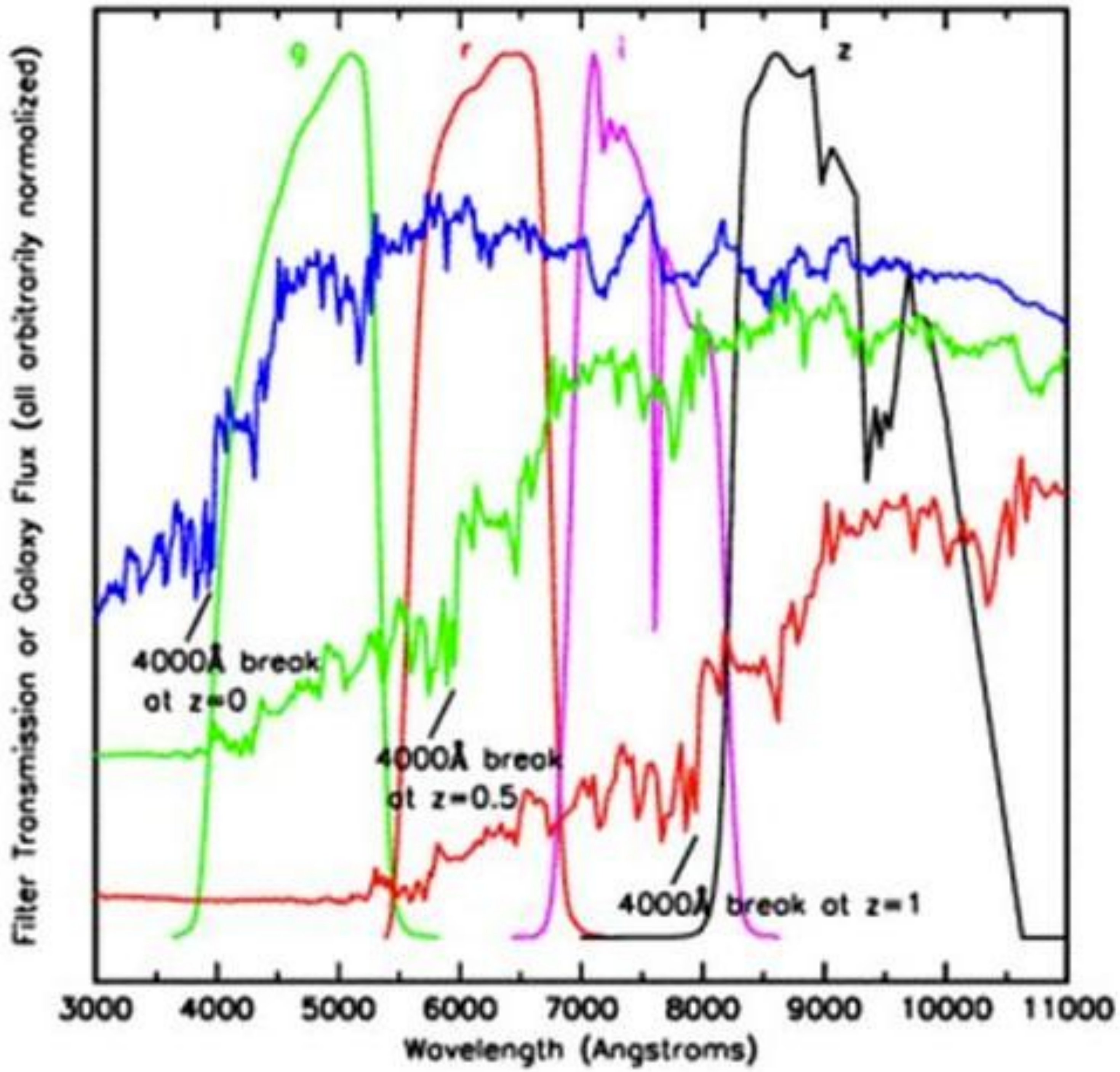
Galaxy spectra

- Absorption lines:
  - atoms / molecules in the atmosphere of stars that absorb specific wavelengths
  - cold gas in the interstellar medium that extracts energy from the passing radiation
- Emission lines caused by gas being ionized and heated and then re-radiating at a specific allowed wavelengths

Measuring redshift

- Photometric surveys:
    - Template-based methods
    - Machine learning techniques
- Spectroscopic surveys (characteristic spectral lines shifted from their usual position)

Measuring redshift



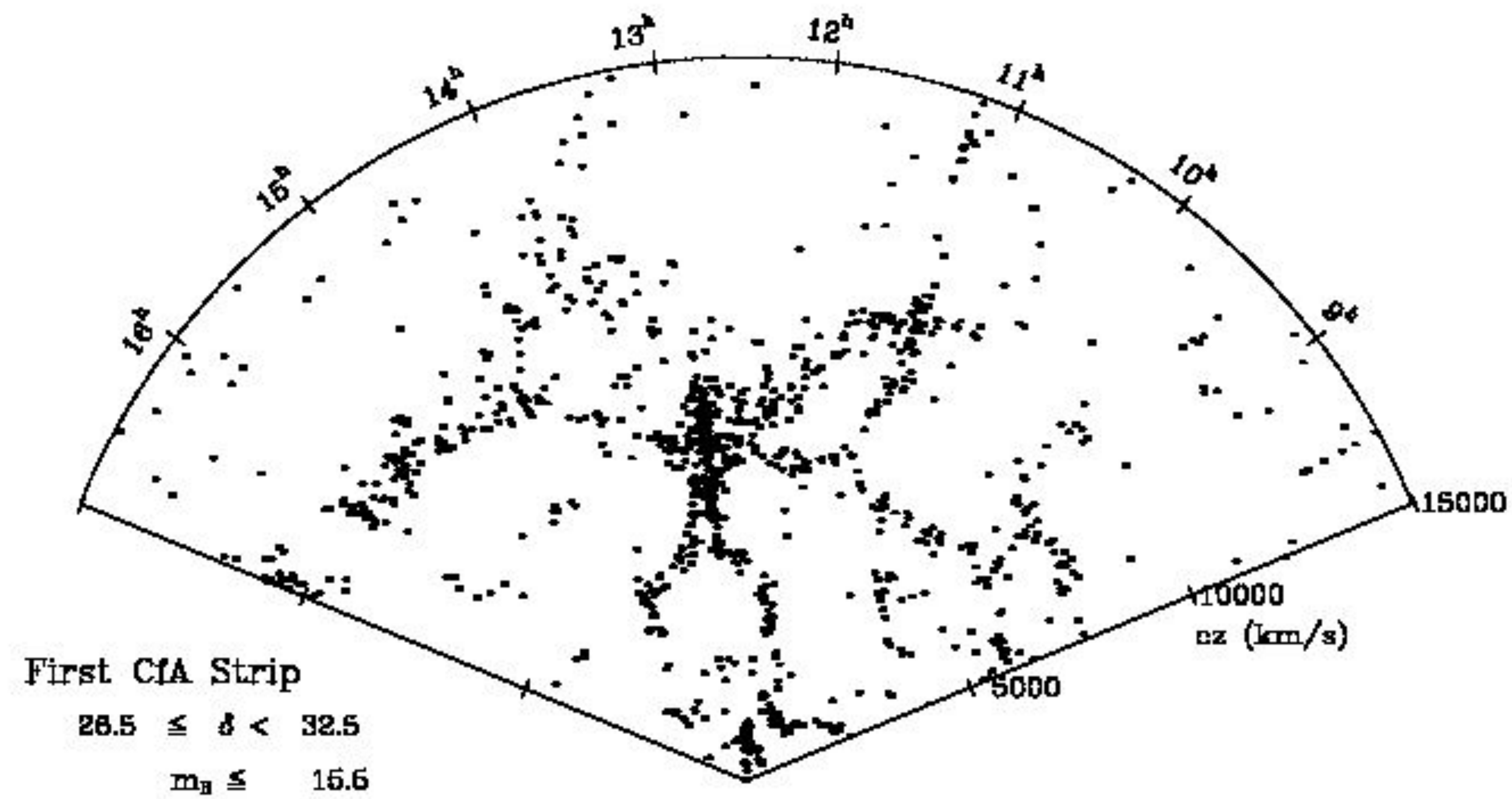
Large galaxy surveys

Completed (# objects)

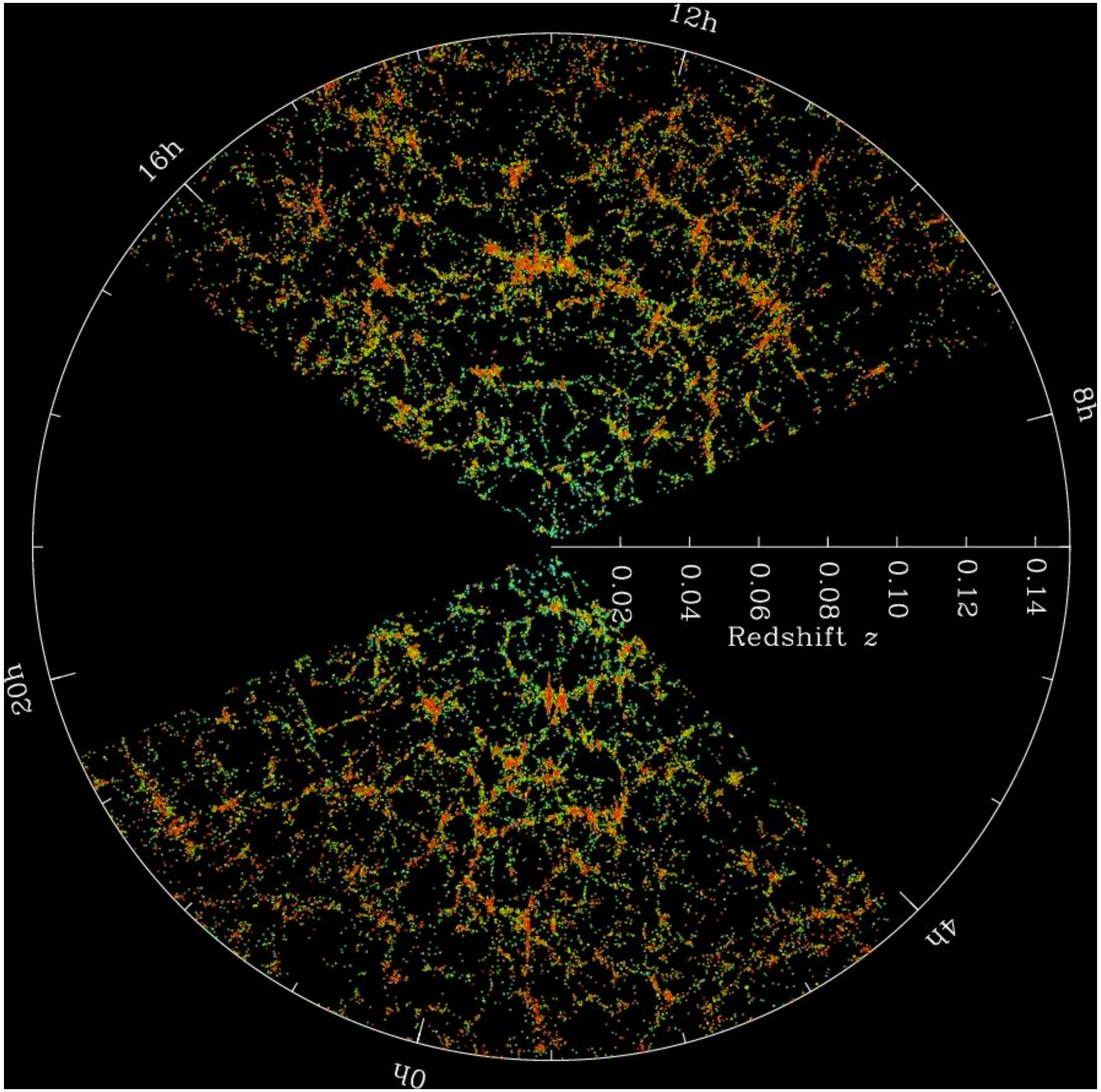
- 1977 - 1982: CfA Redshift Survey (~14K) - North
- 1985 - 1995: CfA2 (~18K) - North
- 1997 - 2002: 2dFGRS (~382K) - North
- 1997 - 2001: 2MASS (~45K) - Full
- 2000 - 2005: **SDSS**-I (700K) - North
- 2001 - 2009: 6dF Galaxy Survey (~136K) - South
- 2003 - 2012: CFHTLenS (28.7M) - North
- 2002 - 2013: DEEP2 (50K) - North
- 2008 - 2014: GAMA (~300K) - South
- 2008 - 2014: SDSS-III, BOSS (1.5M) - North
- 2010 - 2016: VIKING (100M) - South
- 2013 - 2019: KiDS (100M) - South
- 2014 - 2020: SDSS-IV, eBOSS (800K) - North
- 2015 - 2019: **PAUS** (40K) - North
- 2013 - 2019: **DES** (400M) - South

Current and future

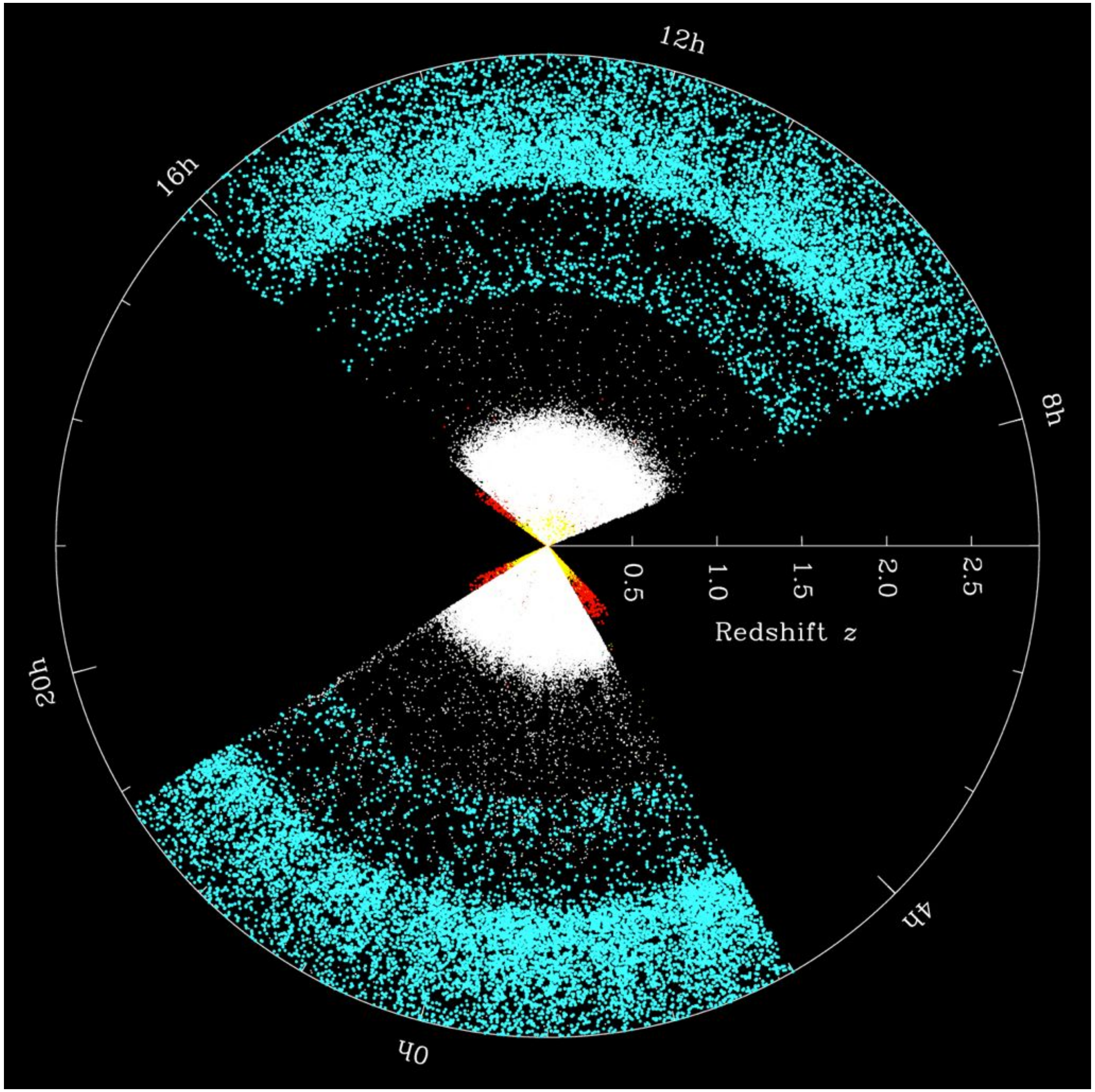
- 2022 - 2027: DESI (35M) - North
- 2023 - 2029: **Euclid** (2B) - Full
- 2024 - 2034: **LSST** (20B) - South
- 2027 - 2032: Roman - Full



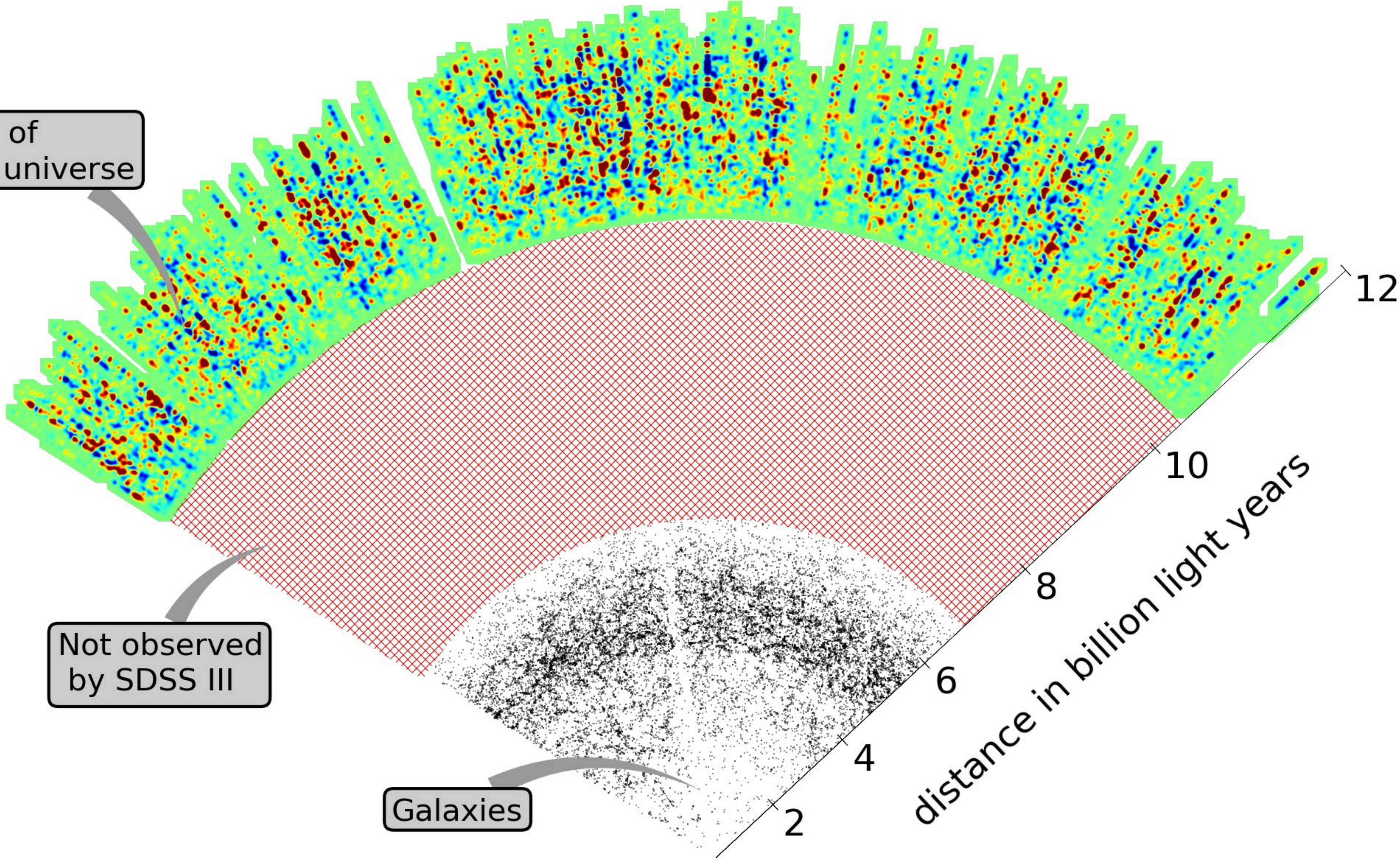
CfA survey: completed in 1982, ~14000 galaxies up to redsfhit cz~15000km/s



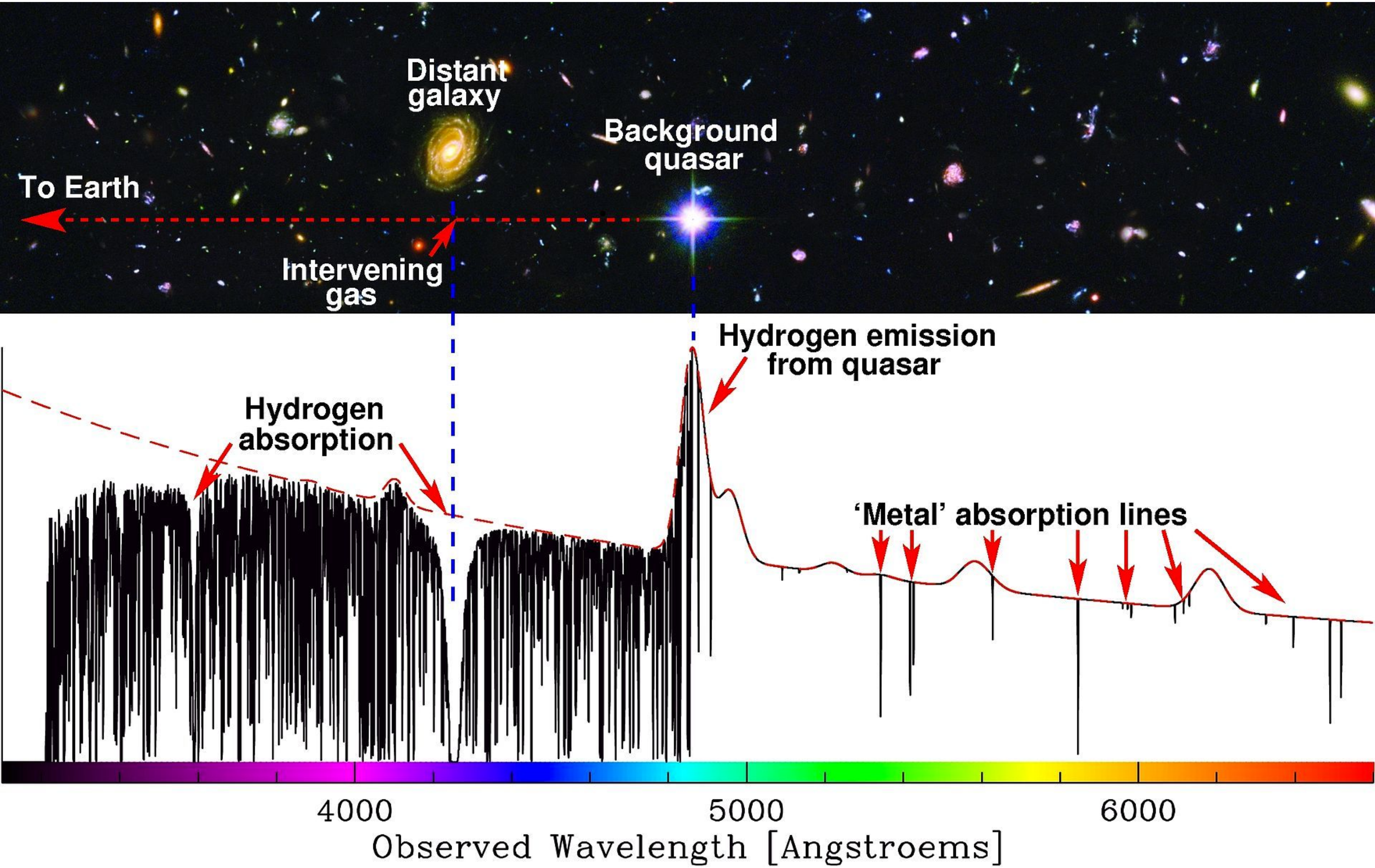
SDSS distribution of local galaxies



BOSS galaxies and quasars (in November 2011)



Largest 3D map of distant universe Lyman Alpha forest


Distant galaxy
Background quasar
To Earth
Intervening gas
Hydrogen emission from quasar
Hydrogen absorption
'Metal' absorption lines
4000
5000
6000
Observed Wavelength [Angstroems]

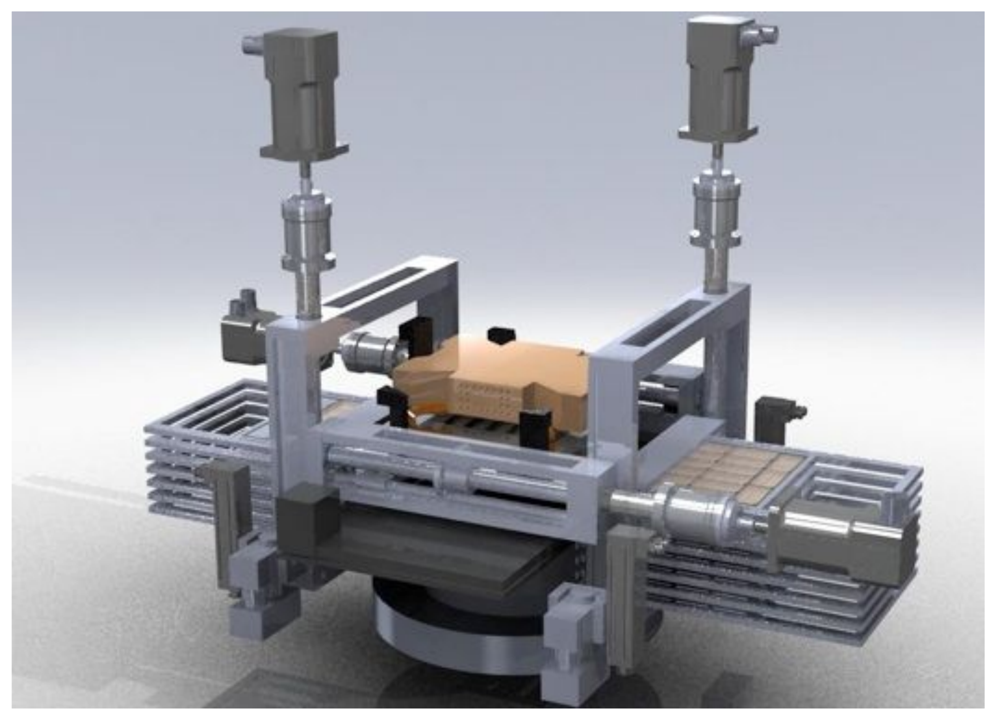
SOME COSMOLOGY PROJECTS

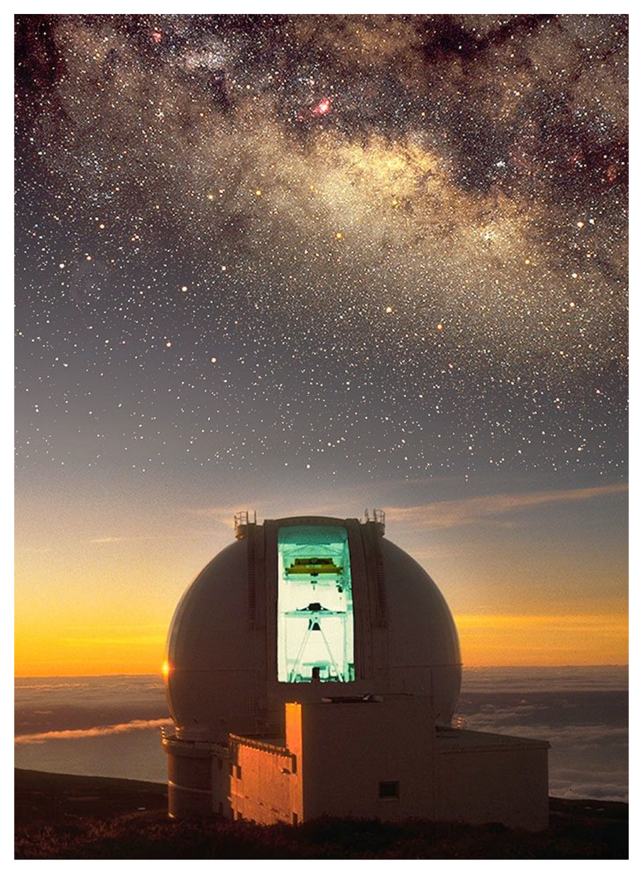
PAU

- Physics of the Accelerated Universe: photometric survey, **42 narrow band filters** (4500-8500Å) and 5 five broad band filters (3000-8575Å)
- 16 CCDs located at La Palma on the 4.2m William Herschel Telescope in the Canary islands
- Very high accuracy photometric redshifts, $\sigma_z < 0.0035(1+z)$, which corresponds to d < 10 Mpc for the whole redshift range of the survey
- Expected area 120 deg^2
- The project has been mainly conceived by people from ICE and IFAE

Cosmology Surveys

Physics of the Accelerating Universe

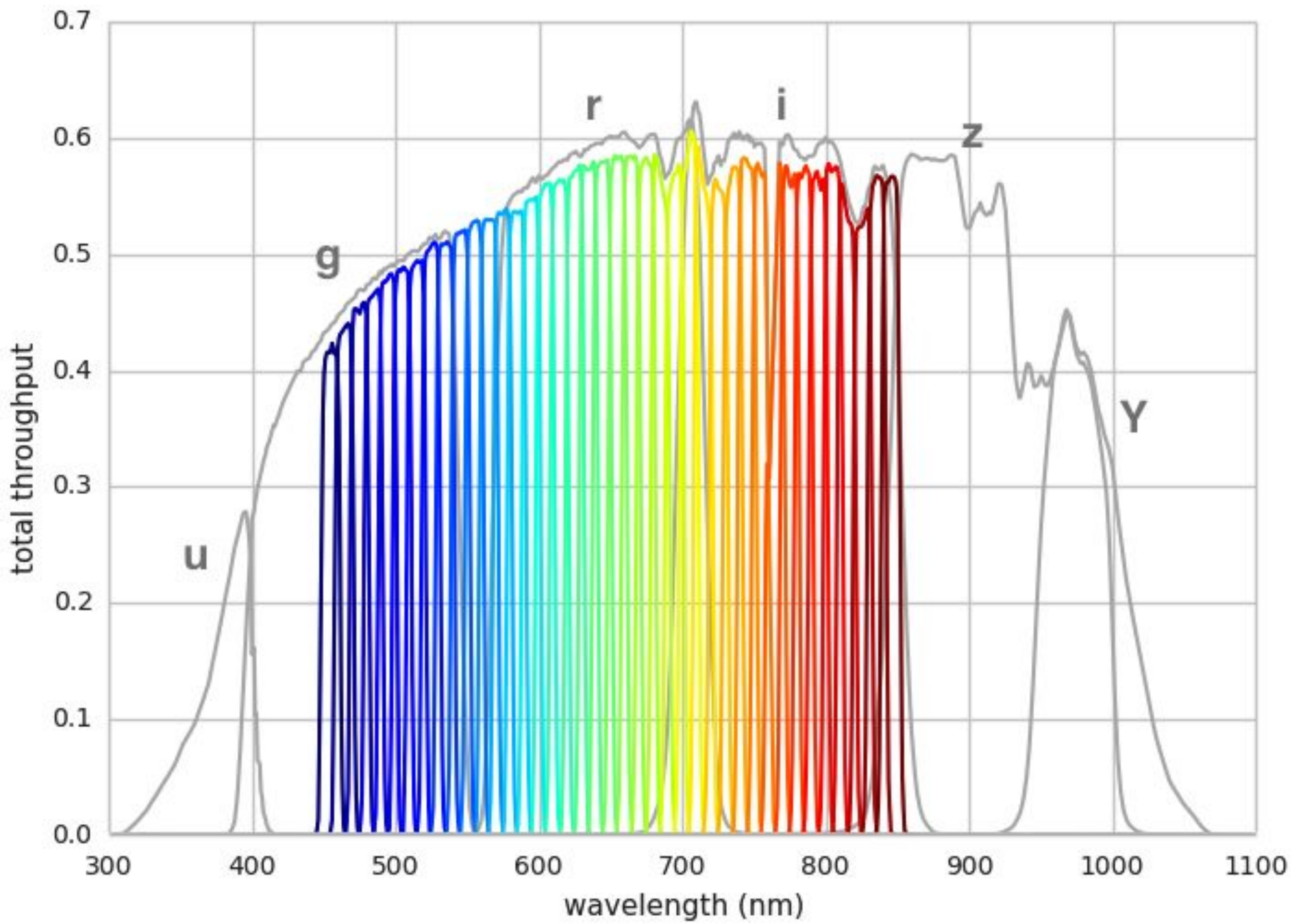




10 billion light-years

100.000.000.000.000.000.000.000.000 meters

PAU


total throughput
wavelength (nm)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
300
400
500
600
700
800
900
1000
1100
u
g
r
i
z
Y

Cosmology Surveys

Physics of the Accelerating Universe

- 42 narrow + 5 five broad filters
- Very high accuracy photometric redshifts
- 120 deg^2

10 billion light-years

100.000.000.000.000.000.000.000.000 meters

PAU

- PAU publications: https://pausurvey.org/pausurvey/publications/

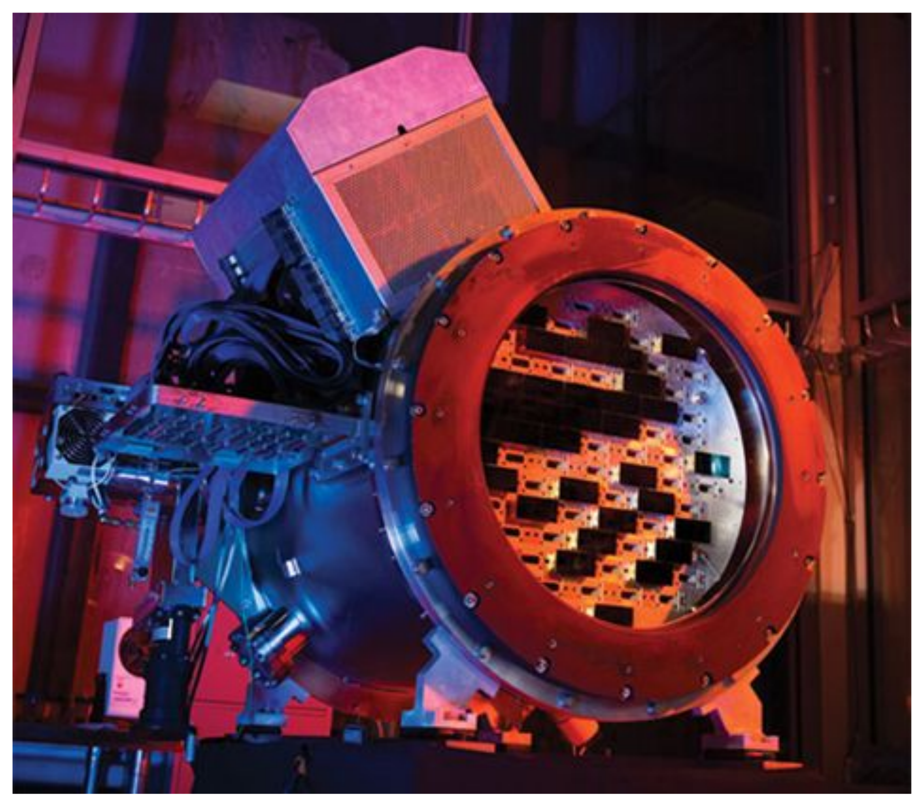
- Several different data releases (both simulated and observed)
    - some of them publicly available through: CosmoHub

DES

- Goal: probe the origin of the accelerating universe and understanding the nature of dark energy
- ~300M of galaxies, 15000 galaxy clusters up to z~1.2 and over 1000 distant type-Ia SNe
- **Photometric survey**: five filters (grizY), from the visible to the infrared, 62 CCDs located at Cerro Tololo on the Blanco Telescope in Chile
- DES combines 4 probes of dark energy: Type Ia Supernovae (**SN**), Baryon Acoustic Oscillations (**BAO**), Galaxy Clusters (**GC**) and Weak Gravitational Lensing (**WL**)
- People from ICE, IFAE and PIC are involved in photometric redshift, large scale structure and simulation science groups

Cosmology Surveys
The Dark Energy Survey

DARK ENERGY
SURVEY



10 billion light-years

26

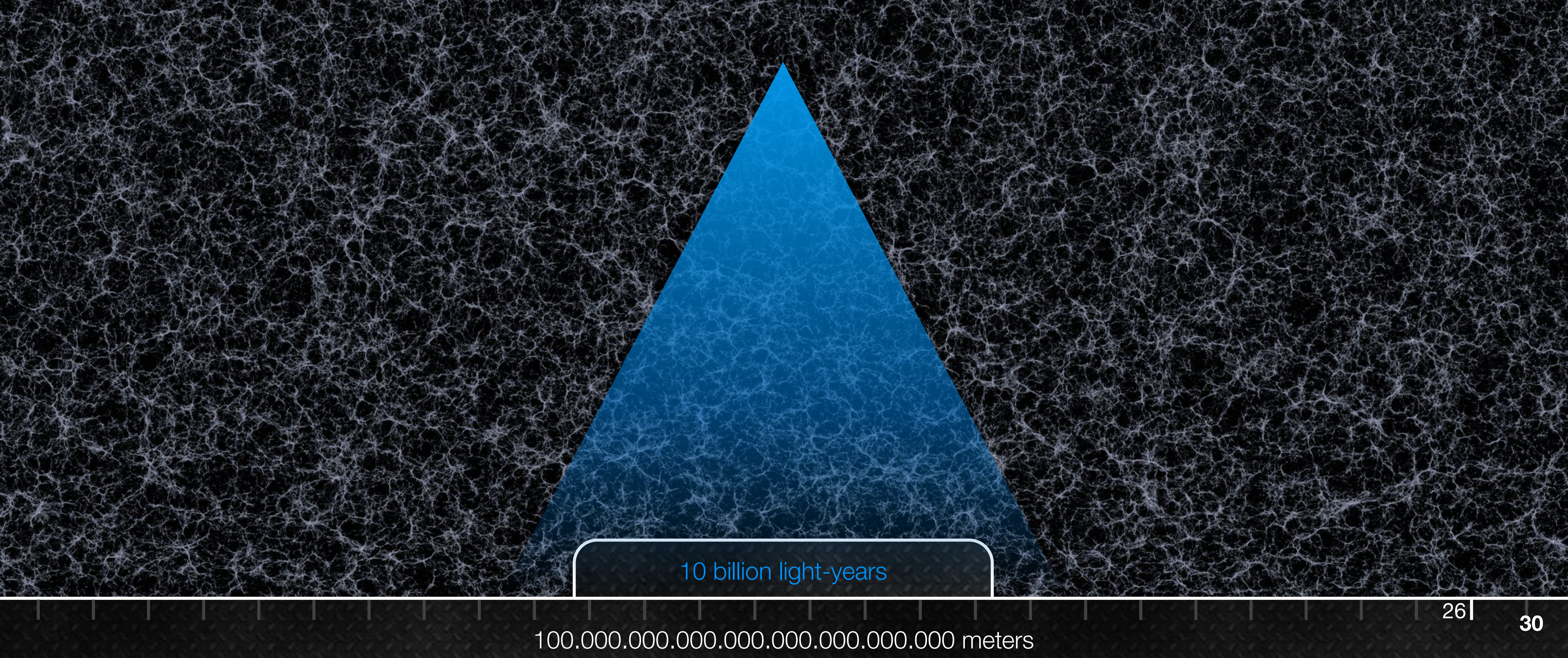
100.000.000.000.000.000.000.000.000 meters

DARK ENERGY
SURVEY

Cosmology Surveys

The Dark Energy Survey

- 300M of galaxies
- 15000 galaxy clusters up to z~1.2
- 1000 distant type-Ia SNe



2013!

The Dark Energy Survey Data Processing and Calibration System

Joseph J. Mohr[a,b], Robert Armstrong[c], Emmanuel Bertin[d], Gregory E. Daues[e], Shantanu Desai[a], Michelle Gower[e], Robert Gruendl[f], William Hanlon[g], Nikolay Kuropatkin[h], Huan Lin[h], John Marriner[h], Don Petravick[e], Ignacio Sevilla[i], Molly Swanson[j], Todd Tomashek[e], Douglas Tucker[h], and Brian Yanny[h] for the Dark Energy Survey Collaboration

[a]Ludwig-Maximilians University Department of Physics, Scheinerstr 1, Munich, Germany 81679; [b]Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, Garching, Germany 85748; [c]University of Pennsylvania Department of Physics, 203 South 33[rd] St., Philadelphia, PA, USA 19104; [d]Institut d'Astrophysique, 98bis, bd Arago, Paris, France 75014; [e]National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL, USA 61801; [f]University of Illinois Department of Astronomy, 1002 West Green St, Urbana, IL, USA 61801; [g]University of Illinois Department of Physics, 1002 West Green St, Urbana, IL, USA 61801; [h]Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL, USA 60510; [i]Centro de Investigaciones Energeticas Medioambiantales y Tocnologicas, Av. Complutense 40, Madrid, SP 28040; [j]Harvard Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, MA USA 02138;

ABSTRACT

The Dark Energy Survey (DES) is a 5000 deg^2 grizY survey reaching characteristic photometric depths of 24th magnitude (10 sigma) and enabling accurate photometry and morphology of objects ten times fainter than in SDSS. Preparations for DES have included building a dedicated 3 deg^2 CCD camera (DECam), upgrading the existing CTIO Blanco 4m telescope and developing a new high performance computing (HPC) enabled data management system (DESDM).

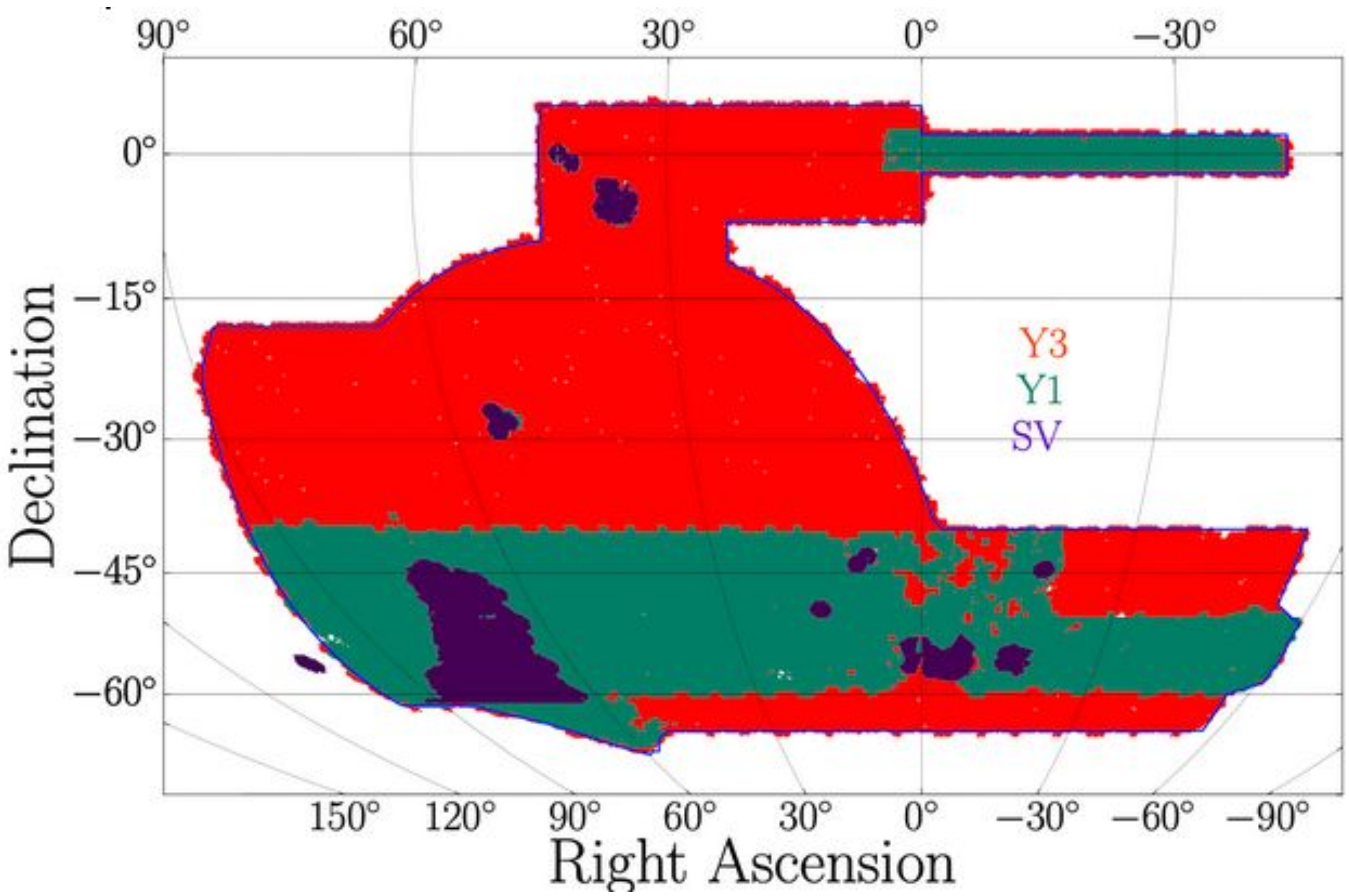
The DESDM system will be used for processing, calibrating and serving the DES data. The total data volumes are high (~2PB), and so considerable effort has gone into designing an automated processing and quality control system. Special purpose image detrending and photometric calibration codes have been developed to meet the data quality requirements, while survey astrometric calibration, coaddition and cataloging rely on new extensions of the AstrOmatic codes which now include tools for PSF modeling, PSF homogenization, PSF corrected model fitting cataloging and joint model fitting across multiple input images.

The DESDM system has been deployed on dedicated development clusters and HPC systems in the US and Germany. An extensive program of testing with small rapid turn-around and larger campaign simulated datasets has been carried out. The system has also been tested on large real datasets, including Blanco Cosmology Survey data from the Mosaic2 camera. In Fall 2012 the DESDM system will be used for DECam commissioning, and, thereafter, the system will go into full science operations.

Keywords: Data Management, High Performance Computing, Optical Astronomy

[redacted]

DES



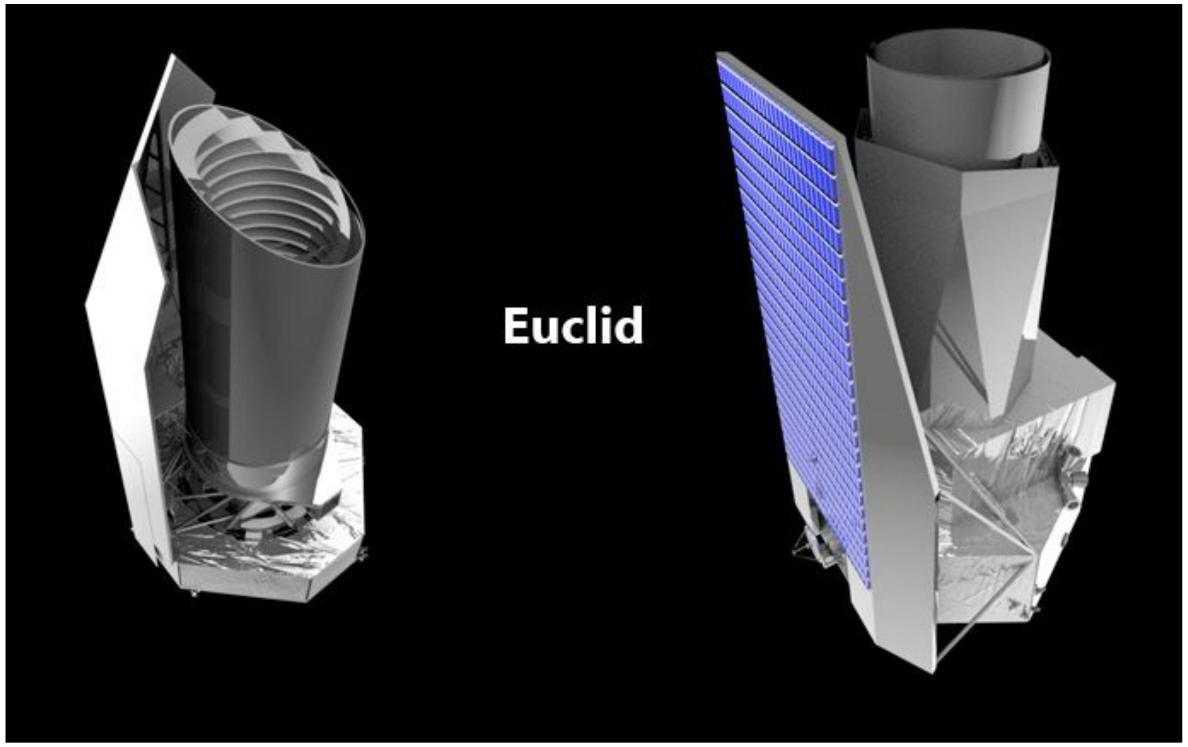
- DES Year 3 Cosmology Results papers: "Our year 3 cosmology analysis of galaxy clustering and gravitational lensing is a massive effort from more than a hundred scientists"
- DES Year 6 Cosmology Results papers: "Our year 6 cosmology analysis of galaxy clustering and gravitational lensing is a massive effort from more than two hundred scientists"
    - Y6 Gold: https://arxiv.org/abs/2501.05739

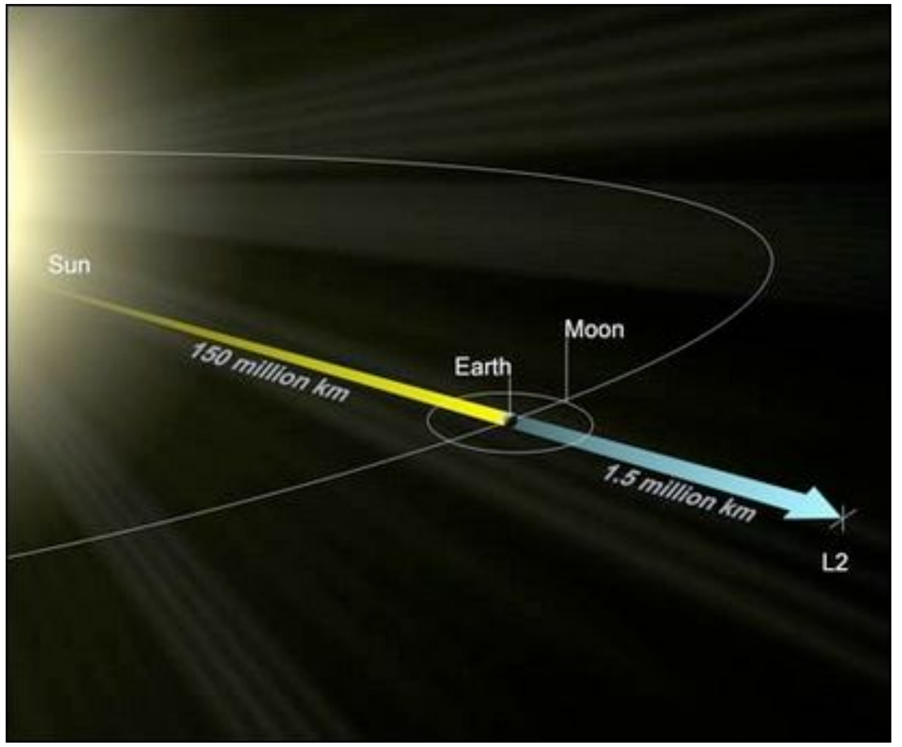
Euclid

- M-class ESA space mission
  - **First space wide field galaxy survey**
  - Nature of dark energy and dark matter
- > 2000 members from 300 laboratories
  - 13 European countries, USA, Canada and Japan
- Combine independent methods
  - **Weak lensing and BAO**
- **15000 deg^2** (and a deep survey of 40 deg^2)
  - Billions of galaxies
- 2 instruments - 3 modes
  - Visual imager (shear measurements)
  - Near Infrared Spectrometer and Photometer
    - Near Infrared imager
    - Near infrared Slitless Spectrometer
- 1.2m telescope will be launched in 2020 (date planned in 2016!) (the end of May 2023 in january 2021!) and will cover 15000 deg^2 (and a deep survey of 40 deg^2)
  - **Finally launched in july 2023** in the Falcon 9 Block 5 (SpaceX) from Cape Canaveral
- People at ICE, IFAE and PIC involved in the photometric redshift, simulations and data management science groups, and also responsible of building the filter wheel in the NIR

Cosmology Surveys
Euclid Mission

Euclid Consortium





10 billion light-years

100.000.000.000.000.000.000.000.000 meters

26

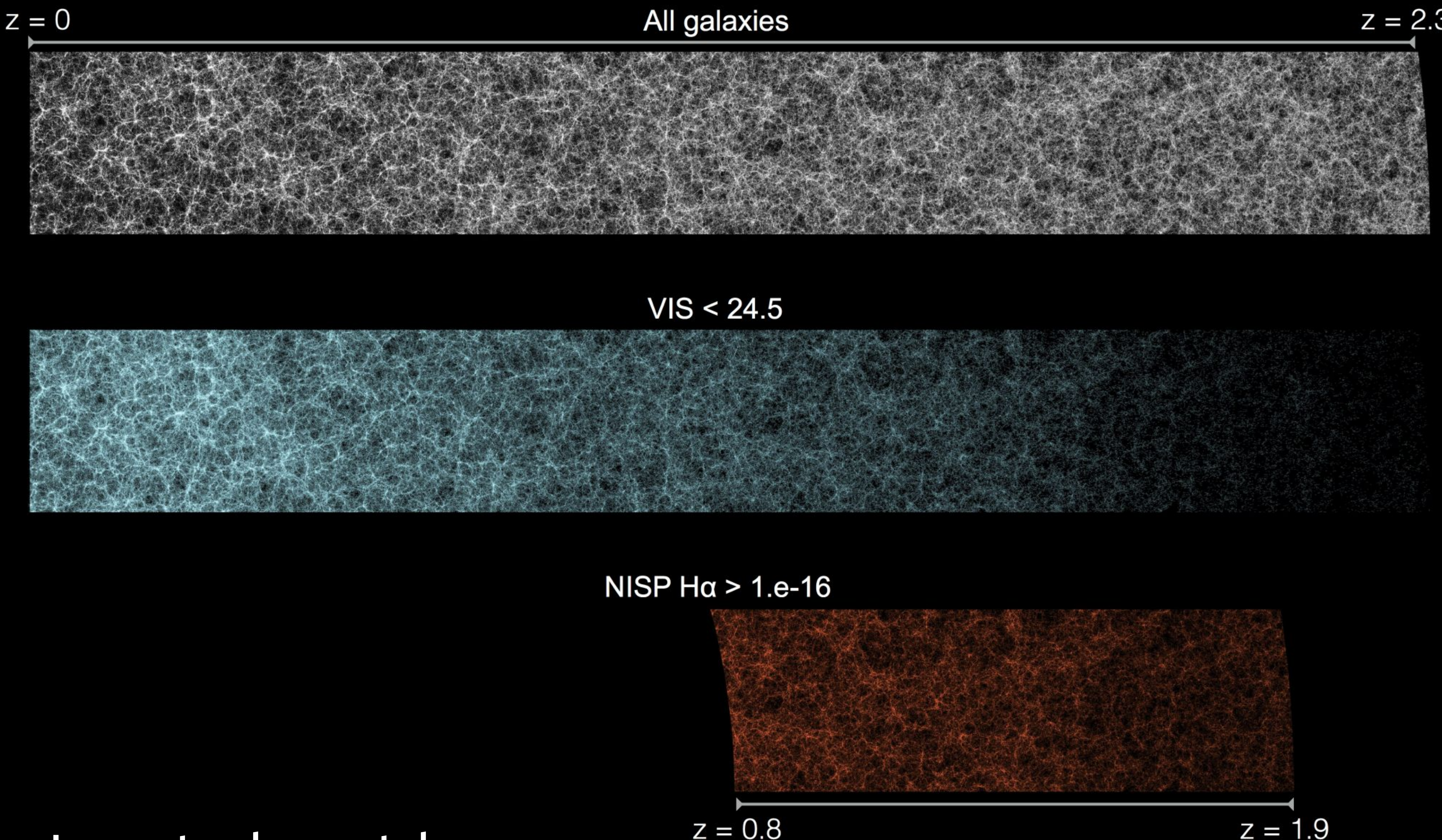
Consortium

Cosmology Surveys
Euclid Mission

- 10 billion astronomical sources
- spectroscopic redshifts for 50 million objects
- 15,000 deg2

10 billion light-years

100.000.000.000.000.000.000.000.000 meters



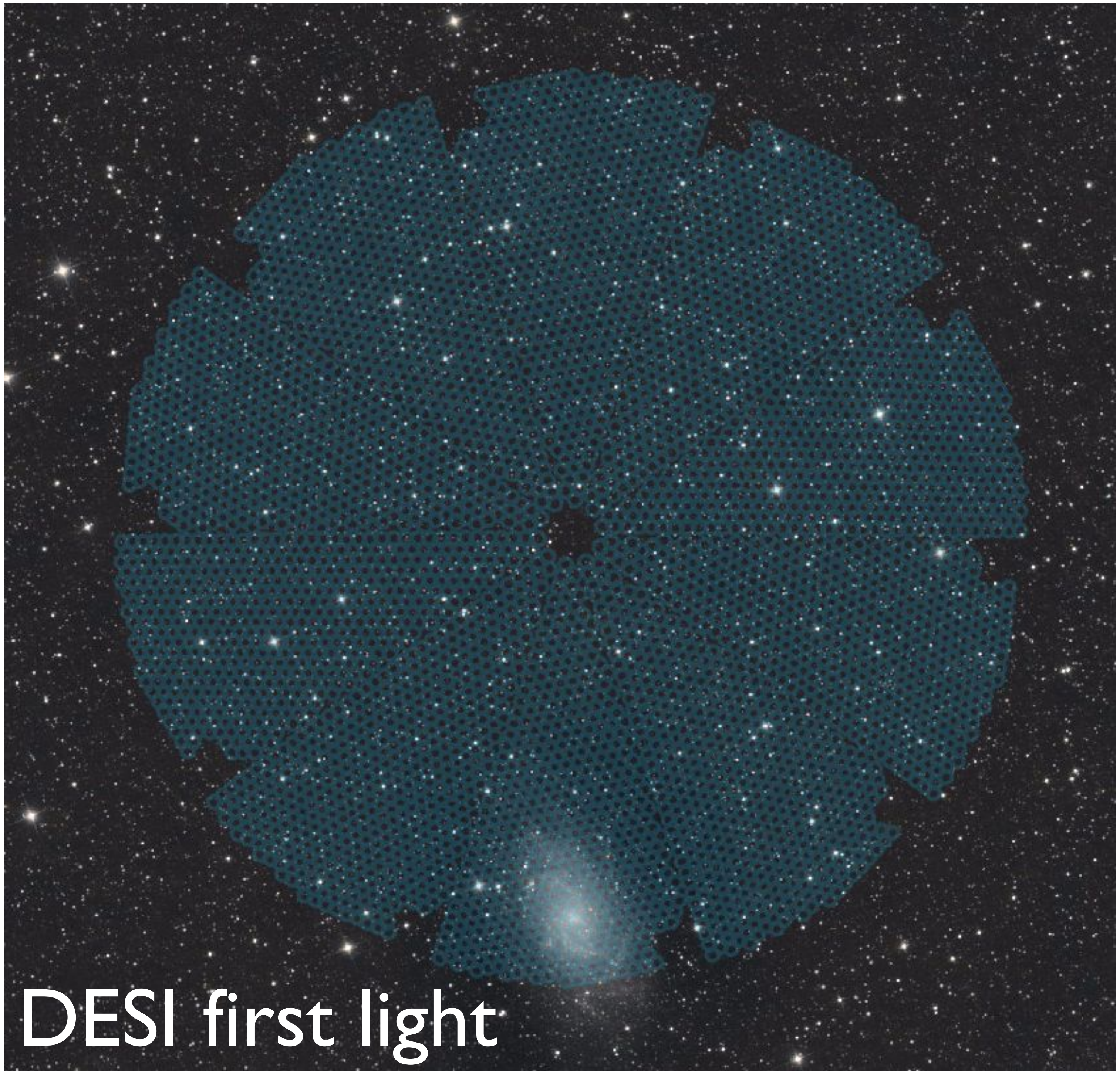
Largest galaxy catalog ever built!

Euclid STAR TEAM prize (2018)

Dark Energy Spectroscopic Instrument (DESI)

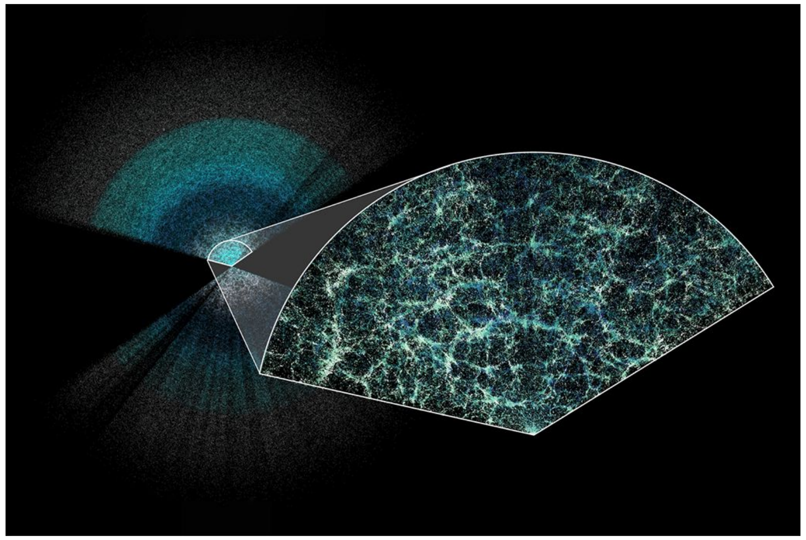
- Goal: probe diverse aspects of cosmology: from DE to alternatives to GR to neutrino masses
- **Spectroscopic astronomical survey**, located at the Kitt Peak Observatory in Arizona (United-States), with a focal plane containing 5K fiber-positioning robots (currently in the commissioning and science validation process)
- Position of 35M galaxies and 2.4M QSOs during 5 years
- Barcelona-Madrid group (IFAE, ICE-CSIC, IEEC, CIEMAT and IFT-UAM) contributions: design, development and installation of the guiding, focusing and alignment of the telescope

Dark Energy Spectroscopic Instrument (DESI)



DESI results

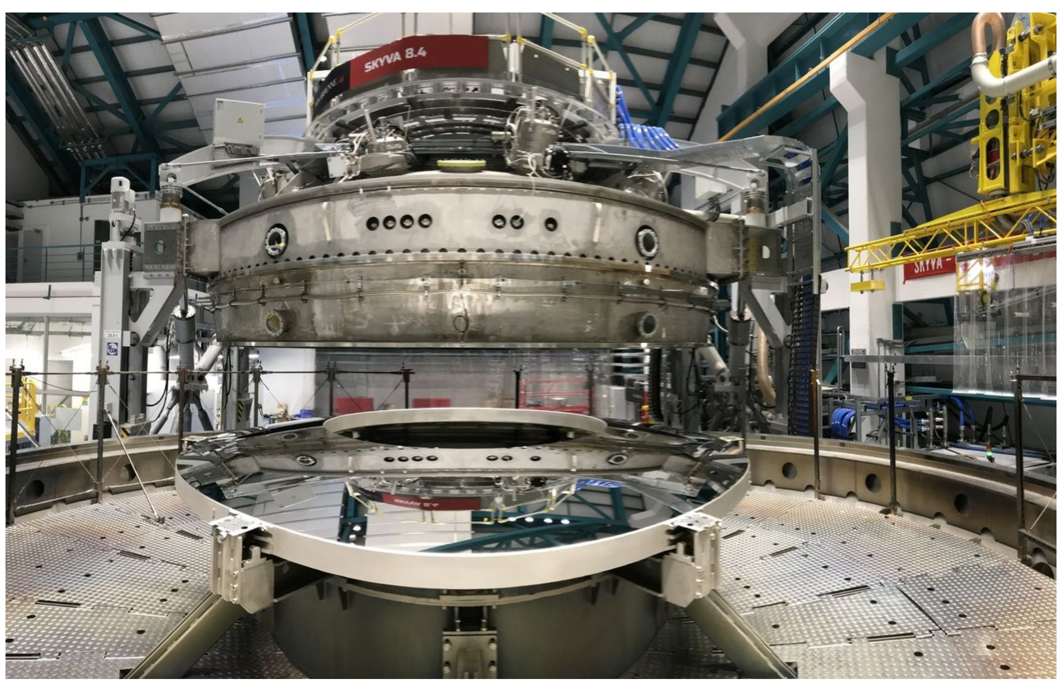
- Most Precise Measurement of Our Expanding Universe (April 2024)
  - Possible hint of dark energy evolving with time (Hubble tension):
    - "Early-universe observations, such as those from the Cosmic Microwave Background (CMB), suggest a lower H_0 (~67.4 km/s/Mpc), while late-universe or local measurements indicate a higher value (~73 km/s/Mpc). This inconsistency challenges our current understanding of cosmology"
- DESI key papers:
  - https://data.desi.lbl.gov/doc/papers/
- [5000 Eyes - Mapping the Universe with DESI]()

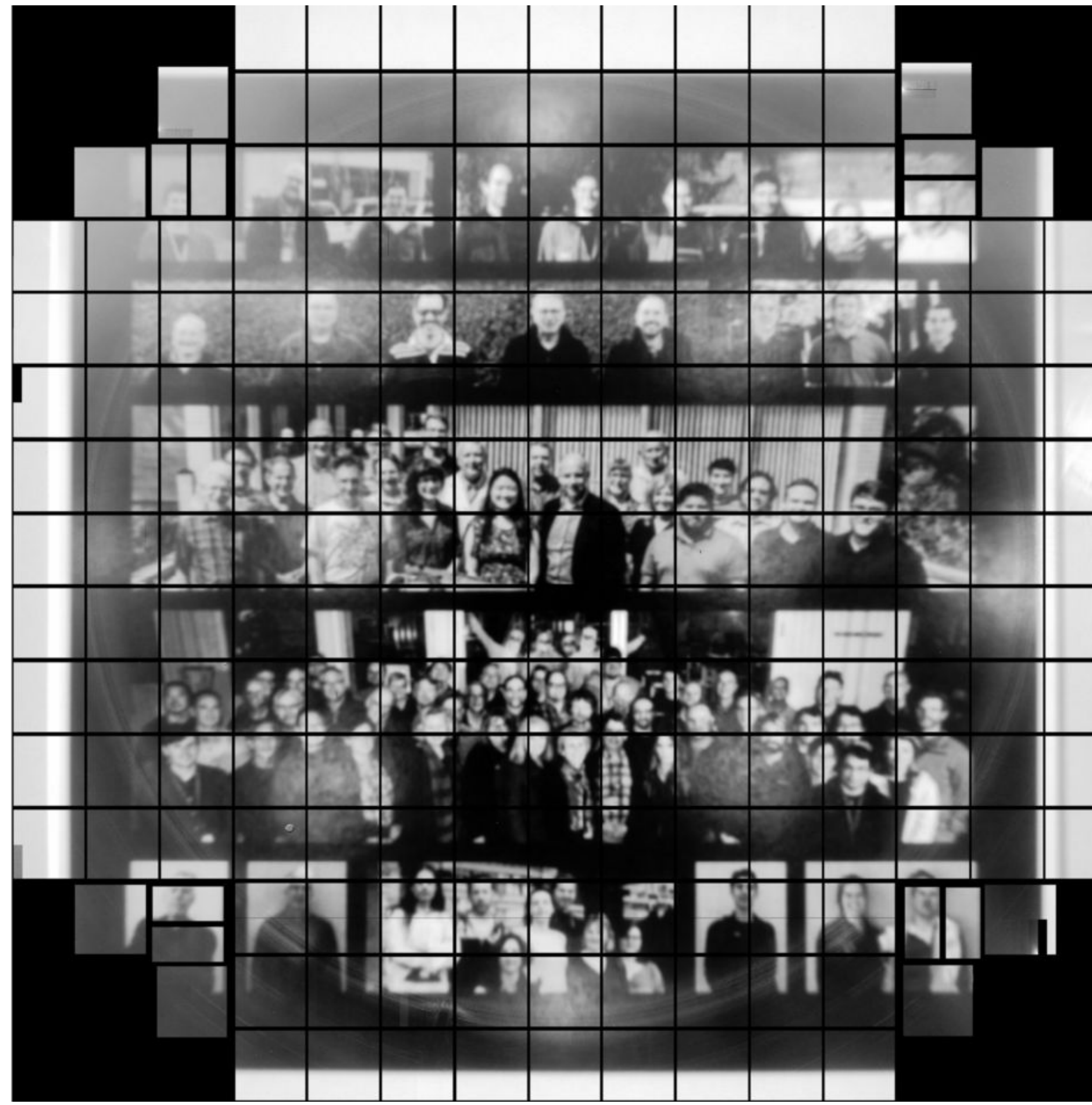


Legacy Survey of Space and Time (LSST)

- Conducted by the Vera C. Rubin Observatory
- Designed to address four science areas:
  - Probing dark energy and dark matter
  - Taking an inventory of the solar system
  - Exploring the transient optical sky (10M alerts/night)
  - Mapping the Milky Way
- Camera with ~ **3 billion pixels** of solid state detectors with 5 filters (**photometric survey**)
- 8.4-meter telescope can survey the full sky in only 3 nights. Science verification (2022), operations (2023)
- Data management: software is one of the most challenging aspects (>20 TiB of data processed and stored / night)
- People at ICE, IFAE and PIC will be involved in several areas.

Vera C. Rubin Observatory

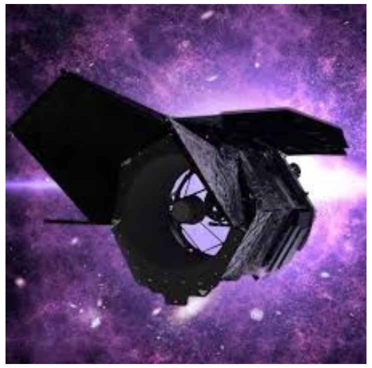




Credits: Rubin Observatory Gallery

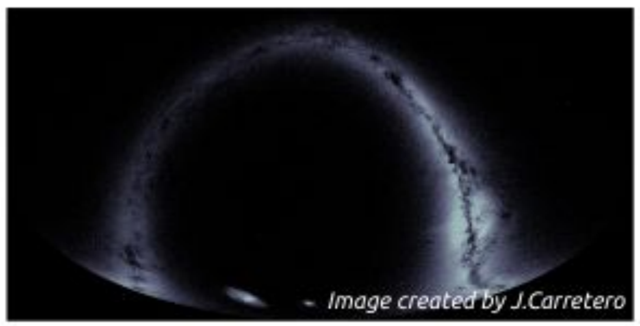
Nancy Grace Roman Space Telescope

- NASA infrared space telescope
  - Scheduled to be launched on a Falcon Heavy rocket by October 2027
  - Sun-Earth L2 orbit
- Designed to address four science areas
  - Dark energy and the expansion of the Universe
  - Extrasolar planets (gravitational microlensing)
- 2.4m telescope (wide field of view) with 2 instruments
  - Wide-Field Instrument (300.8-megapixel) multi-band visible and near-infrared camera (higher resolution than Euclid; deeper surveys)
  - Coronagraphic Instrument (CGI) (small field of view) camera and spectrometer covering visible and near-infrared wavelengths
- ICE involved in SN (as far as I know)



Other large surveys and detectors

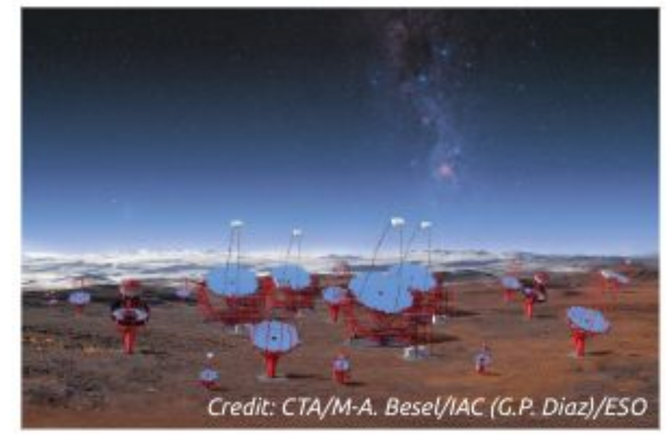
Gaia 3D map of the MW (2013-2025)



James Webb Space Telescope (2022-)

LIGO / Virgo collaborations

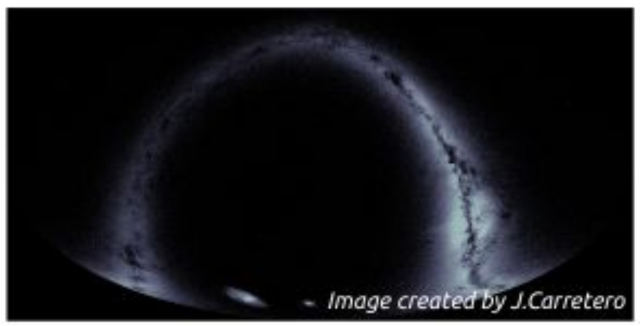
Cherenkov Telescope Array (2026 -)



Square Kilometer Array (mid-2020s -)

Other large surveys and detectors

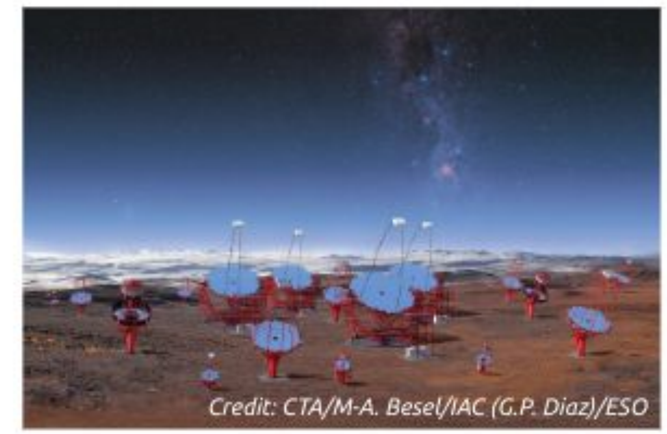
Gaia 3D map of the MW (2013-2025)



James Webb Space Telescope (2022-)

LIGO / Virgo collaborations

Cherenkov Telescope Array (2026 -)



Square Kilometer Array (mid-2020s -)

Multimessenger Astronomy is now a reality

(new windows to probe the cosmos

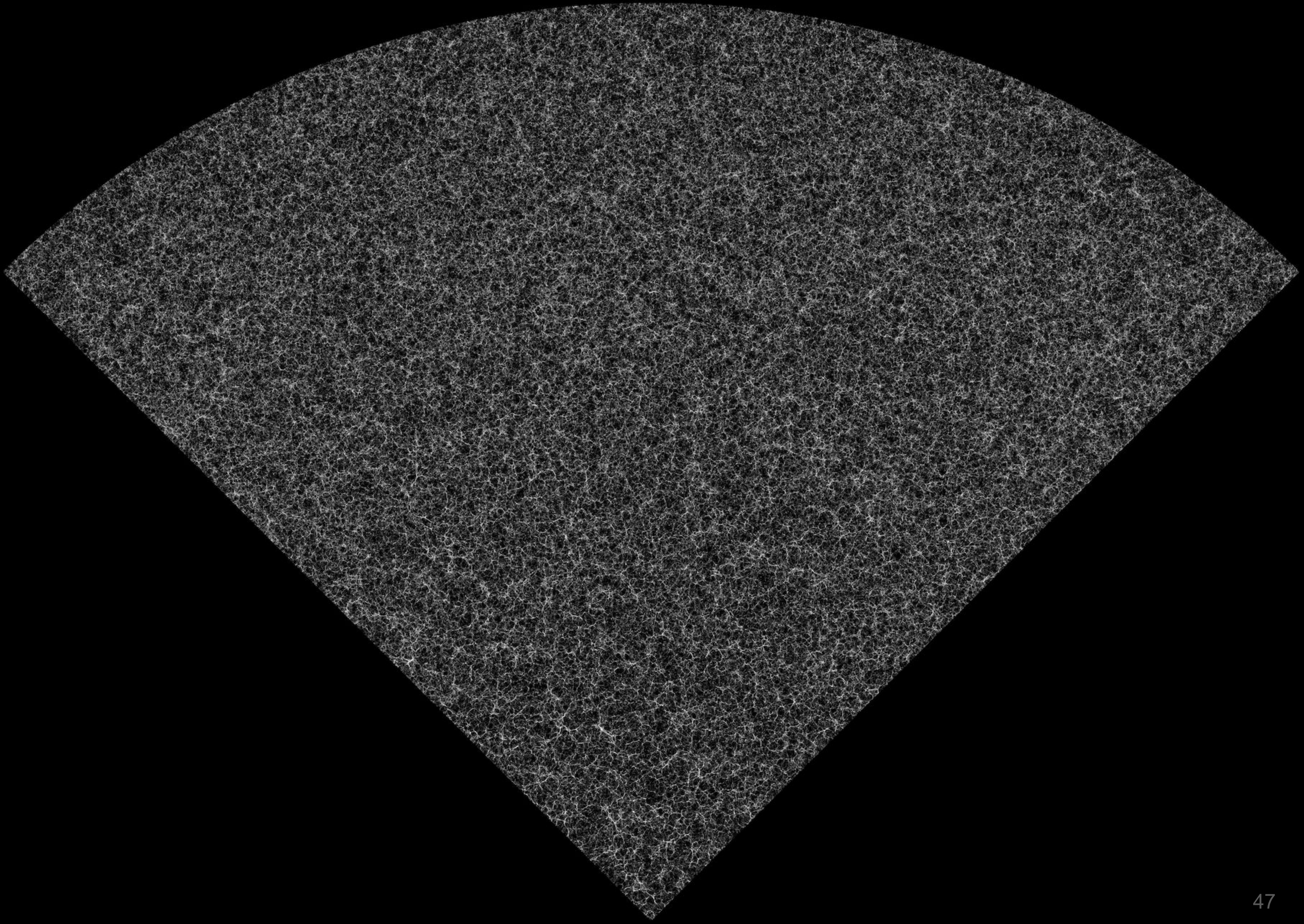
Galaxy formation

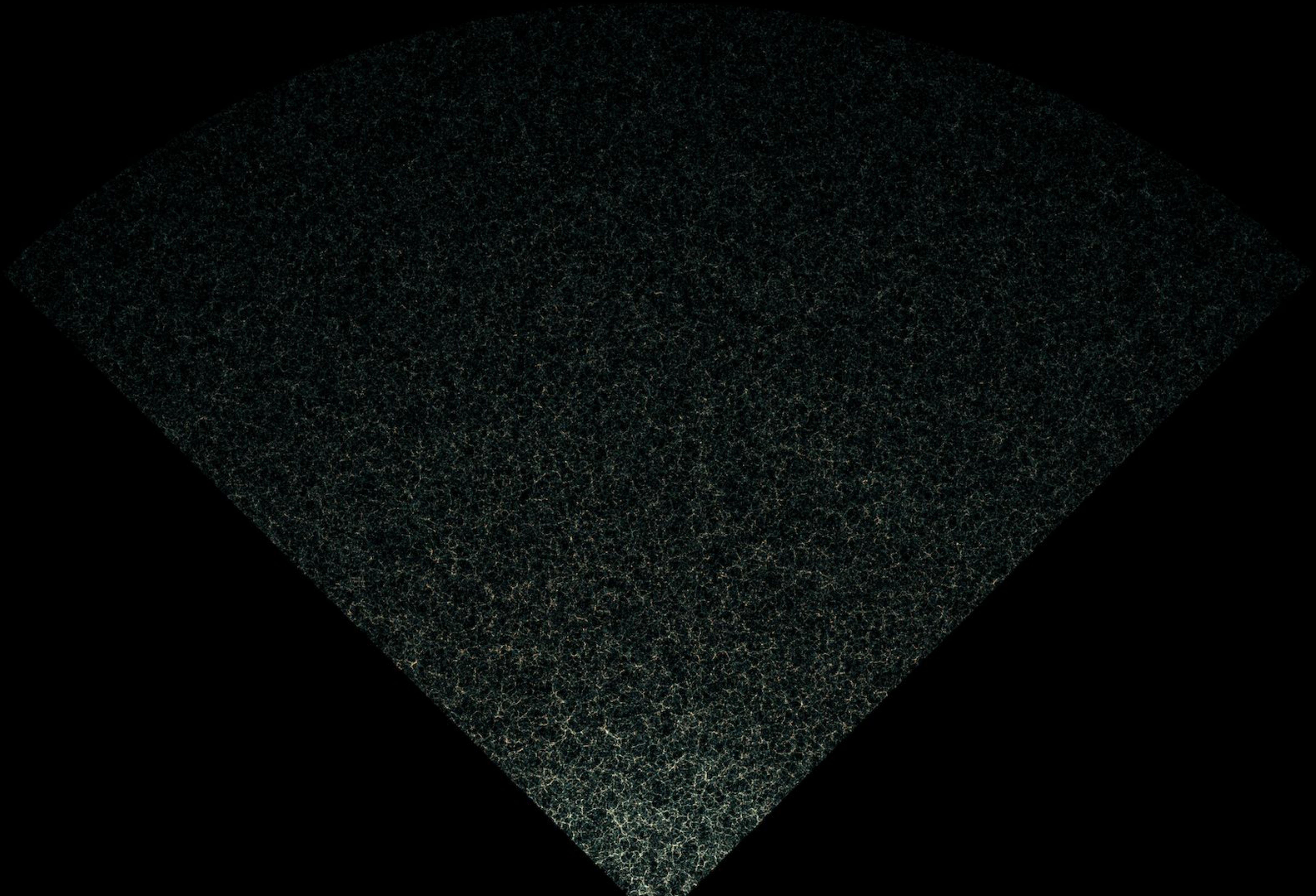
- Galaxy surveys are a **powerful tool** to solve two of the most challenging problems in Astrophysics and Cosmology: understand the **nature of dark energy** and solving the problem of **galaxy formation**
- Current paradigm of structure formation: galaxies are formed and reside in overdensities of the underlying dark matter field (these overdensities are called dark matter haloes)
- The link between the properties of the halo population and the galaxy population is a key ingredient in order to understand galaxy formation

N-body simulations and analytical models give a detailed picture of the underlying dark matter field

Galaxy formation

- Brief summary: galaxies form by the cooling and condensation of gas inside the dark matter haloes. Besides cooling, complicated physical properties such as star formation, merging, tidal interactions and several feedbacks processes, determine galaxy formation and its evolution
- To simulate these processes and reproduce galaxy formation there exist several methods, which by themselves, and together with other ingredients, are used to build mock galaxy catalogs

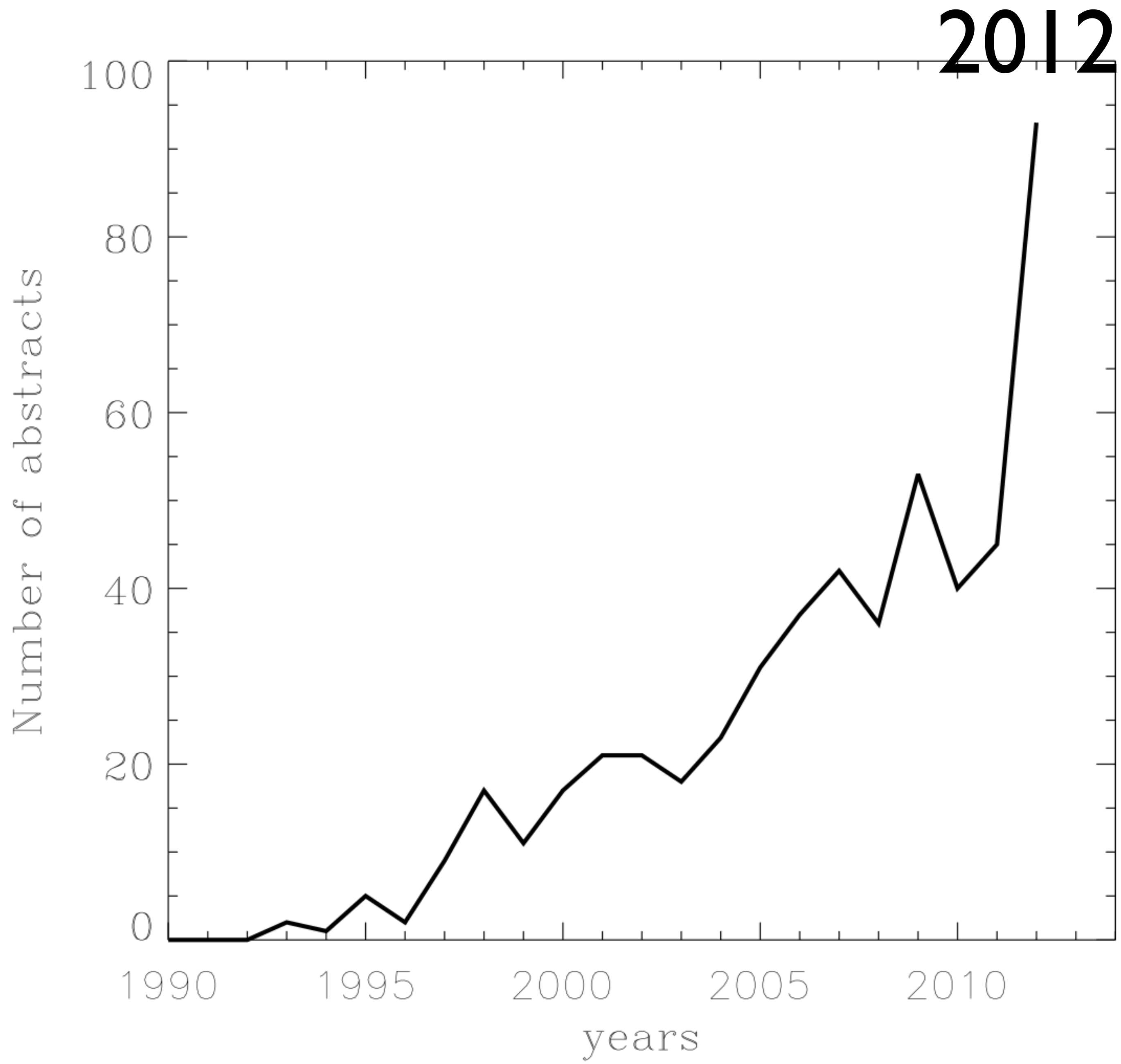




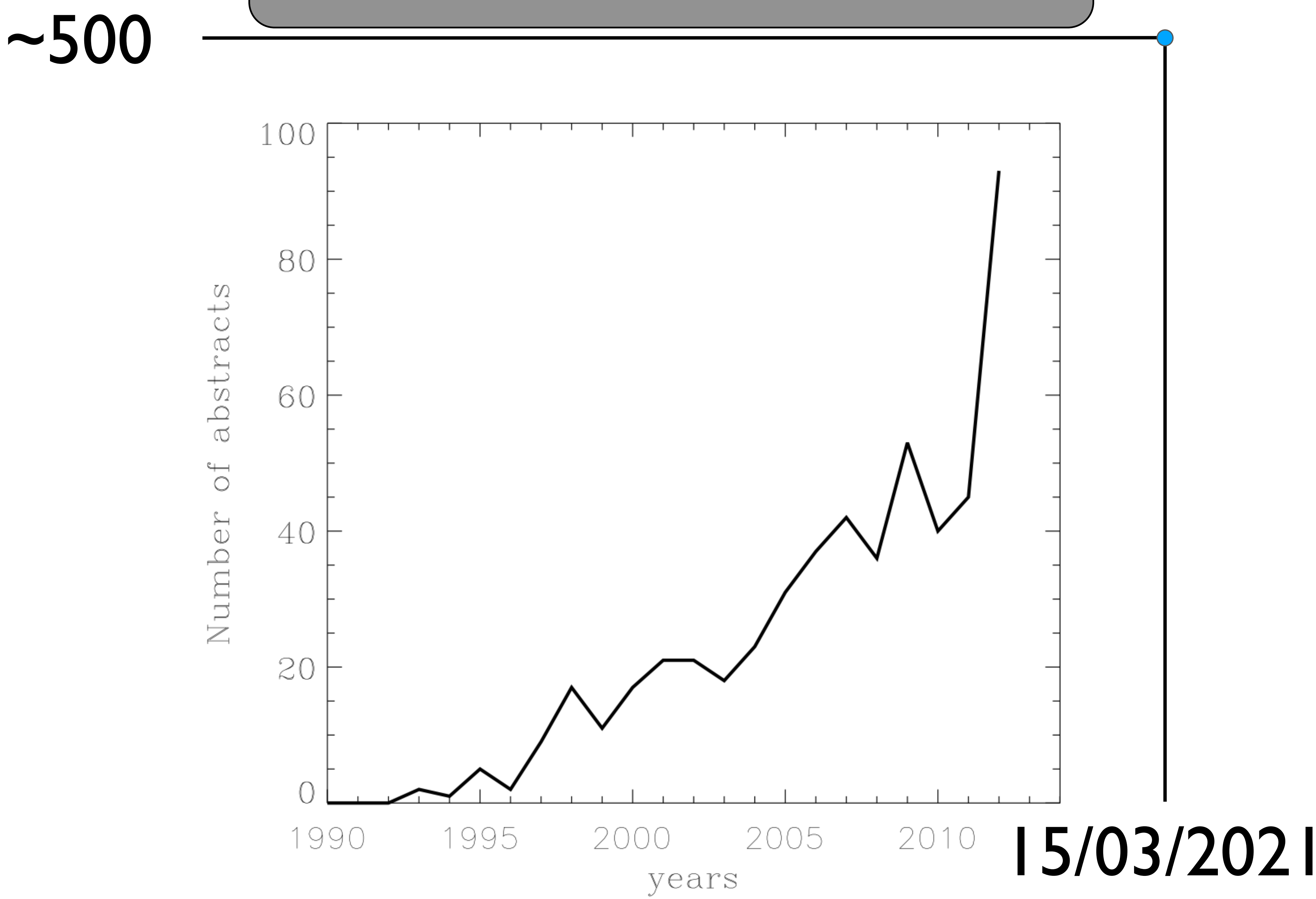
Why mock galaxy catalogs?

- Design and calibrate galaxy surveys (e.g. DES, PAU or Euclid):
  - study selection effects (magnitude, colour, photometric redshift, etc.)
  - explore systematic effects
  - calibrate errors
  - test new techniques to measure cosmological parameters
  - calibrate cluster finder codes, photometric redshift code s,etc…
- Understand and interpret observations
- Essential tool to develop the data processing

Articles with mocks



Articles with mocks



Methods

- We can separate the methods into the ones involving N-body simulations and the ones that do not use them
- The former are more common mainly because one obtains much more information about the underlying dark matter distribution.
- The later avoid the problem of managing and solving the equations of motion of a huge number of particles, which is computationally expensive, and usually take advantage from results of N-body simulations (e.g. density profile to place satellite galaxies)
- For both methods, in most cases, the essential ingredients are a correct number density and spatial distribution of dark matter haloes

Methods

- We can separate the methods into the ones inv… simulations and the ones that do not use th…
- The former are more common mai… obtains much more information about the underl… distribution.
- The later avoid the pro… …ng and solving the equations of motion of a huge … …cles, which is computationally expensive, a… …advantage from results of N-body simulati… …y profile to place satellite galaxies)
- F… …s, in most cases, the essential ingredients are a c… …ber density and spatial distribution of dark matter haloes

Recent overview (2020): Cosmological simulations of Galaxy Formation

Without N-body

- Scoccimarro & Sheth 2002 cite in their conclusions some methods of making mock galaxy catalogues without using as input N-body simulations, e.g:
    - Soneira & Peebles 1978: algorithm to reproduce the observed two-point statistics at small-scales. By adjusting the parameters of the model they produce a galaxy map that visually appears as a first approximation of the Lick survey
    - Coles & Jones 1991: model for the distribution of matter using a lognormal random field
    - Sheth & Saslaw 1994: technique to distribute cluster centers using a Poisson distribution and place galaxies around them following a specific recipe for the shape, the radius, the number of particles and the radial density profile of the cluster

Without N-body / SAMs

- Another approach to model galaxy formation, and to build mock galaxy catalogs, are Semi-Analytic Models. They need a halo merger history tree and, in principle, they do not necessary use N-body simulations
- In its origin the halo merger history tree was obtained using an analytical prescriptions (the extended Press-Schechter prescription), but more recently halo merger history trees are obtained by using N-body simulations

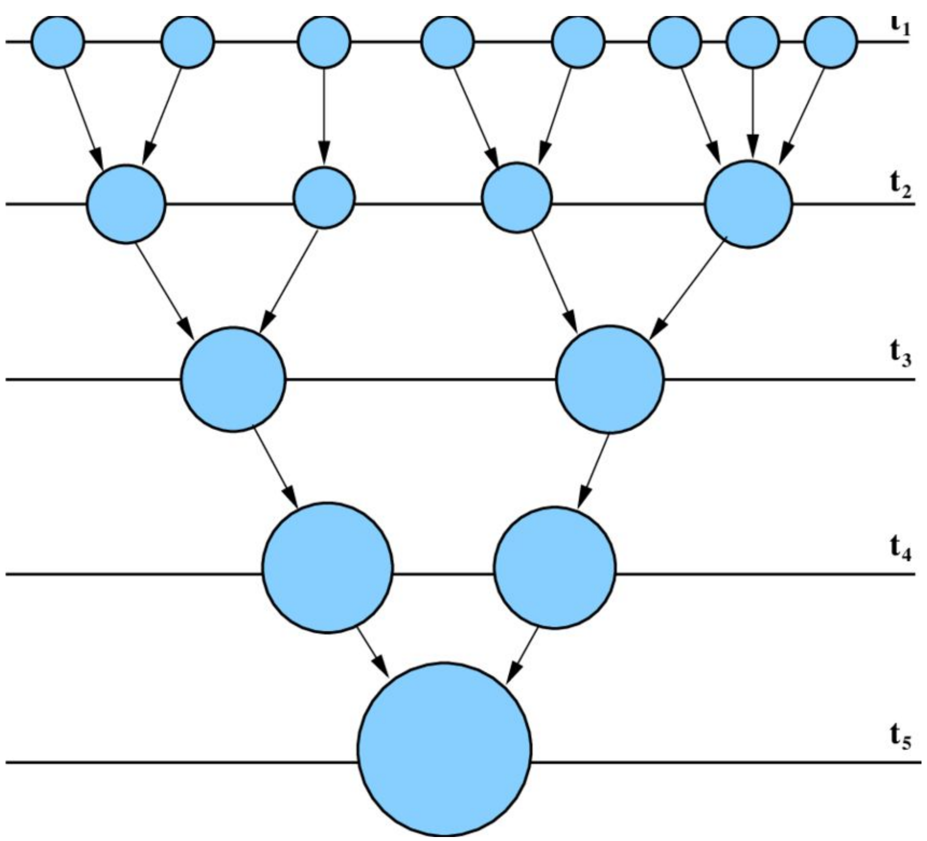
Without N-body / SAMs

- Another approach to model galaxy formation, a ... k galaxy catalogs, are Semi-Analytic Models ... o merger history tree and, in principle, they do ... se N-body simulations

- In its origin the halo mer... was obtained using an analytical prescripti... ded Press-Schechter prescription), but more rece... r history trees are obtained by using N-body s...

Annual review (2015): Physical Models of Galaxy Formation in a Cosmological Framework

Merger trees



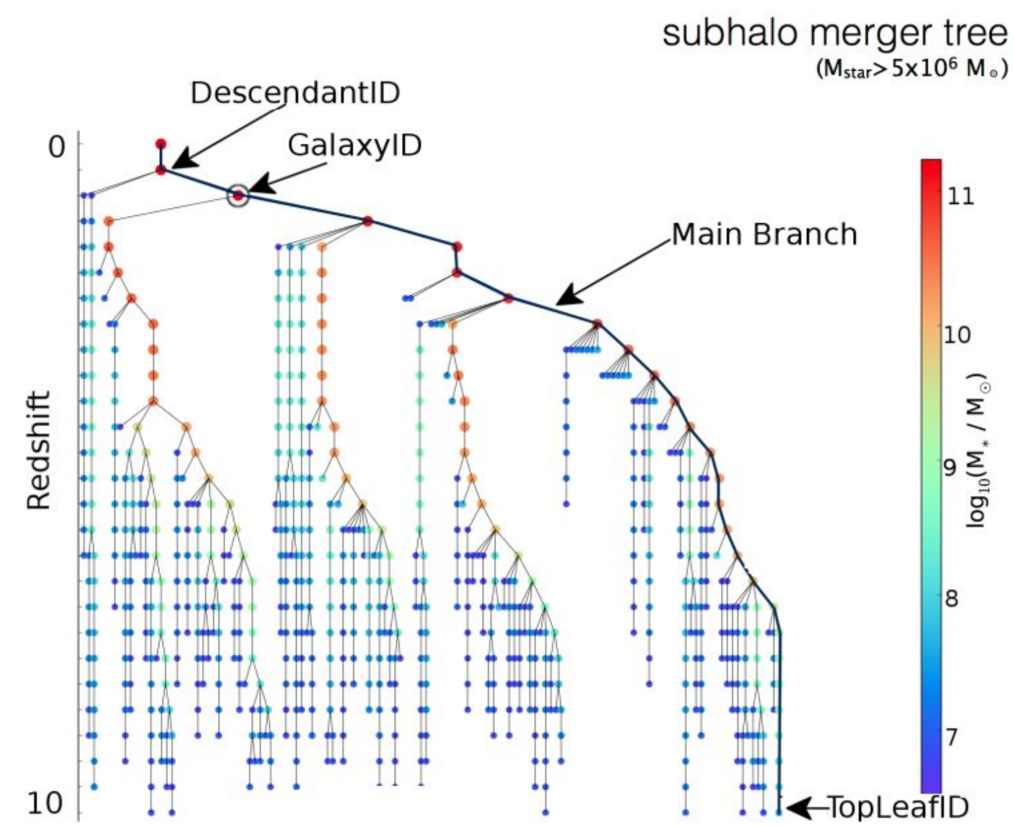


Figure 2: Merger history of a galaxy with a $z = 0.18$ stellar mass $M_* \sim 10^{10}\,M_\odot$ indicated by the circled dot. Symbol colours and sizes are scaled with the logarithm of the stellar mass. The `GalaxyID` of this galaxy points towards it, as indicated by the arrow. The main progenitor branch is indicated with a thick black line, all other branches with a thin line. The `TopLeafID` gives the `GalaxyID` of the highest redshift galaxy on the main progenitor branch whilst the `LastProgID` (not shown) gives the maximum `GalaxyID` of all the progenitors of the galaxy considered. Querying all galaxies with an ID between `GalaxyID` and `LastProgID` will return all the progenitor galaxies in the tree.

Eagle simulation of galaxy formation

Fast approximated methods

- From nIFTy Cosmology paper: "Many new tools (see Table 1) have been developed for reconstructing in an approximate way the **large-scale structures down to the mildly nonlinear scales**, allowing a fast generation of simulated volumes of the Universe. In this way, a direct computation of the covariance matrices by means of large numbers of realizations is possible"

Methodology	reference
Log-Normal	Coles & Jones 1991
PTHalos	Manera et al. 2012, 2015
PINOCCHIO (PINpointing Orbit-Crossing Collapsed Hierarchical Objects)	Monaco et al. 2002, 2013
COLA (COmoving Lagrangian Acceleration simulation)	Tassev et al. 2013
PATCHY (PerturbAtion Theory Catalog generator of Halo and galaxY distributions)	Kitaura et al. 2014a,b
QPM (quick particle mesh)	White et al. 2013
EZmock (Effective Zel'dovich approximation mock catalogue)	Chuang et al. 2015
HALOgen	Avila et al. 2014

Table 1. The methodologies of generating mock halo/galaxy catalogues developed in the last years. The methodologies included in this study are highlighted using bold font.

PTHALOS

- New method orders of magnitudes faster than N-body simulations.
- They argue that since the behavior of baryons inside dark-matter haloes is not completely known it is possible to obtain similar galaxy distributions without solving the equation of motion for the dark matter due to gravitational clustering.
- **The idea is to obtain a correct number and space distribution of virialized dark-matter haloes without using the computationally expensive method used in N-body simulations**.
- Combine perturbation theory with halo models of the non-linear density and velocity fields.

With N-body

- Some mock galaxy catalogs use N-body simulations.
- Most of them use a halo catalogue as input.
- **To reproduce the correct clustering the essential point is to have a halo population with the correct number density and the correct clustering**
- There are galaxy catalogues built by directly populating dark-matter particles

DM halo - galaxy relation

- 3 common approaches to relate DM haloes with galaxies:
  - Hydrodynamical simulations
  - Semi-analytic models
  - Extensions of the Halo model
    - Conditional Luminosity Function (CLF)
    - Halo Occupation Distribution (HOD)
    - Halo abundance matching technique (HAM)

Hydrodinamical sims

- Another type of numerical simulations in a cosmological context, study galaxy formation and evolution
- Most explicit way to model galaxy formation is using numerical hydrodynamic techniques, in which the equations of gravity, hydrodynamics, and thermodynamics are concurrently solved for particles and/or grid cells representing dark matter, gas, and stars (different particle types rather than pure dark matter N-body simulations)
- Unlike gravity, which is a long range force, hydrodynamic mechanisms are important on small scales (<1Mpc/h), e.g. in the formation of galaxies and in linking the matter distribution of the universe to the observed light distribution
- The main limitation of these techniques is that computational exigencies restrict the dynamic range that can be explicitly simulated

Hydrodinamical sims

- Another type of numerical simulations in a cos[...], study galaxy formation and evolution
- Most explicit way to model galaxy f[...] numerical hydrodynamic techniques, in wh[...] of gravity, hydrodynamics, and therm[...] ncurrently solved for particles and/or grid ce[...] dark matter, gas, and stars (different particle [...] pure dark matter N-body simulations)
- Unlike gr[...] long range force, hydrodynamic mechanisms are im[...] scales (<1Mpc/h), e.g. in the formation of [...] king the matter distribution of the universe to the [...] distribution
- Th[...] limitation of these techniques is that computational exigencies restrict the dynamic range that can be explicitly simulated

Recent overview (2020): Cosmological simulations of Galaxy Formation

Some examples (very outdated!)

- Cole et al. 1998 (2dF and SDSS surveys; N-body)
- Coil et al. 2001 (DEEP survey; GIF Virgo Consortium simulations with SAMS)
- Yang et al. 2004 (2dFGRS; N-body with CLF)
- Millenium simulation project 2005 (N-body with SAMs)
- Jouvel et al. 2009 (realistic galaxy spectro-photometric catalogs without N-body sims)
- Cabré & Gaztañaga 2011(BAO catalogs directly populating dark-matter particles)
- Las Damas project (set of N-body sims using HOD)
- Manera et al. 2012 (600 catalogs using PTHALOS)

Galaxy and some other Mock Catalogues available online

https://w.astro.berkeley.edu/~jcohn/mocklist.html

Halo model / HOD

- Born to explain the observed power spectrum and correlation function of galaxies
- To estimate the statistics of the large scale structure of the density and velocity fields as a consequence of the non-linear gravitational clustering ($\delta >> 1$)
- Not easy to put the limit between the Halo model and the Halo Occupation distribution model
- HOD is a logical extension of the halo model to relate dark matter and baryonic matter

Halo model

- Main assumption: **all matter is contained in dark matter haloes**
- Can describe many different observational probes:
  - the large and small scale distribution of galaxies (which constrain the dark matter distribution)
  - the weak gravitational lensing (which provides a direct detection of the dark matter density field)
  - the distribution of the pressure on large scales given by the Sunyaev Zel'dovich effect
  - the signature on small angular fluctuations on the CMB of the density, velocity, momentum and pressure fields of the dark or/and baryonic matter

Halo model

- Review "Halo models of large scale structure" by Cooray & Sheth 2002
- The halo model is a formalism to describe the **non-linear gravitational clustering (δ >> 1)**
- The distribution of matter and the distribution of velocities inside the haloes with the number and spatial distribution of these dark-matter haloes are used to estimate the statistics of the large scale structure of the density and velocity fields as a consequence of the non-linear gravitational clustering

Halo model

HOD

- Review "Halo models of large scale structur… …neth 2002
- The halo model is a formalism to de… …inear **gravitational clustering**
- The distribution of matt… …bution of velocities inside the haloes with the nu… …distribution of these dark-matter haloes are us… …e statistics of the large scale structure of the de… …y fields as a consequence of the non-linear gra… …ing

Recent overview (2023): The halo model for cosmology: a pedagogical review

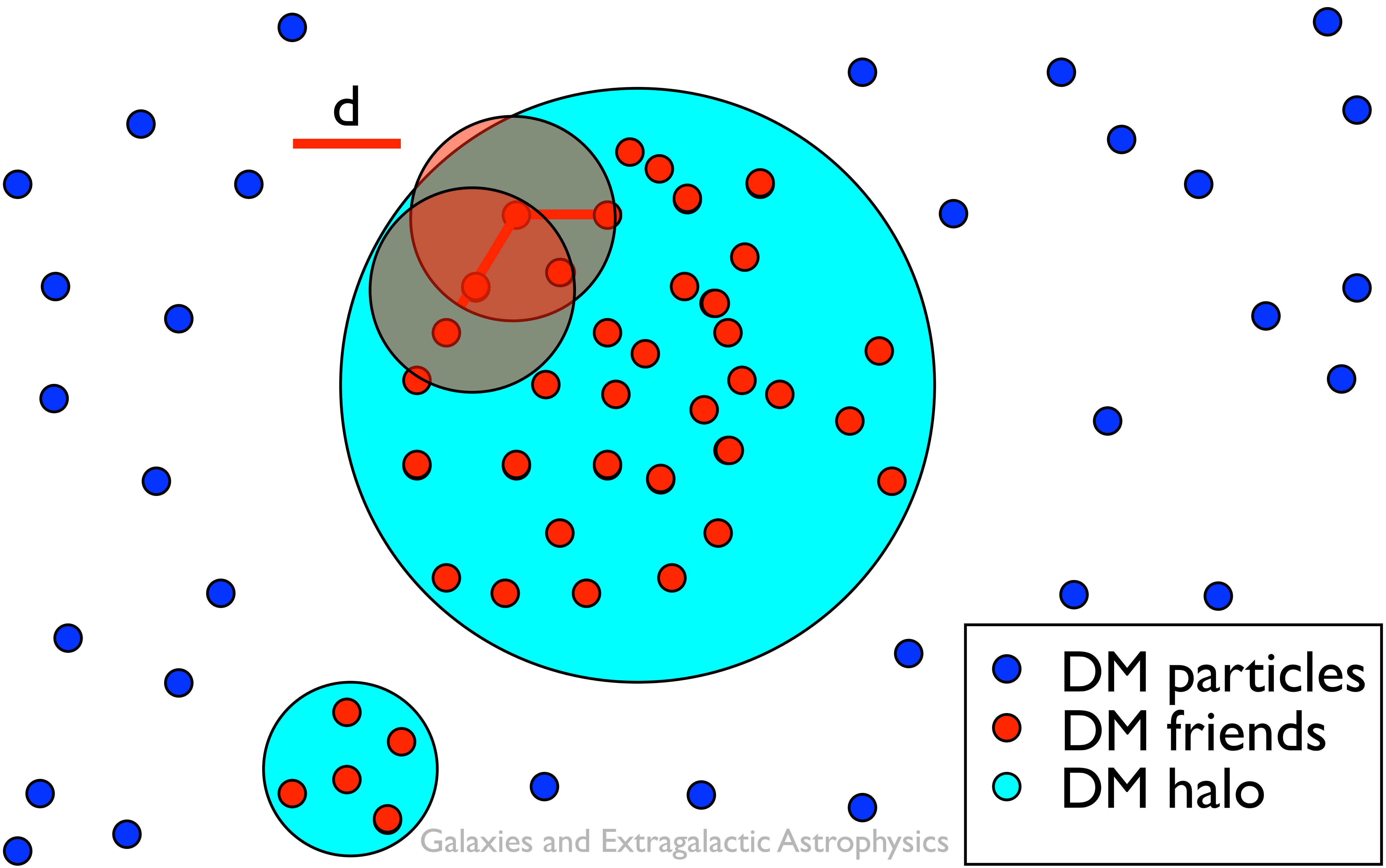
Halo model

- To describe the DM clustering (following Martinez & Saar 2002):
  1. Population of virialized DM haloes sampling the large-scale linear density field (position and velocity)
  2. Mass density profile of the haloes
  3. Concentration parameter
  4. Halo mass function
  5. Halo bias function

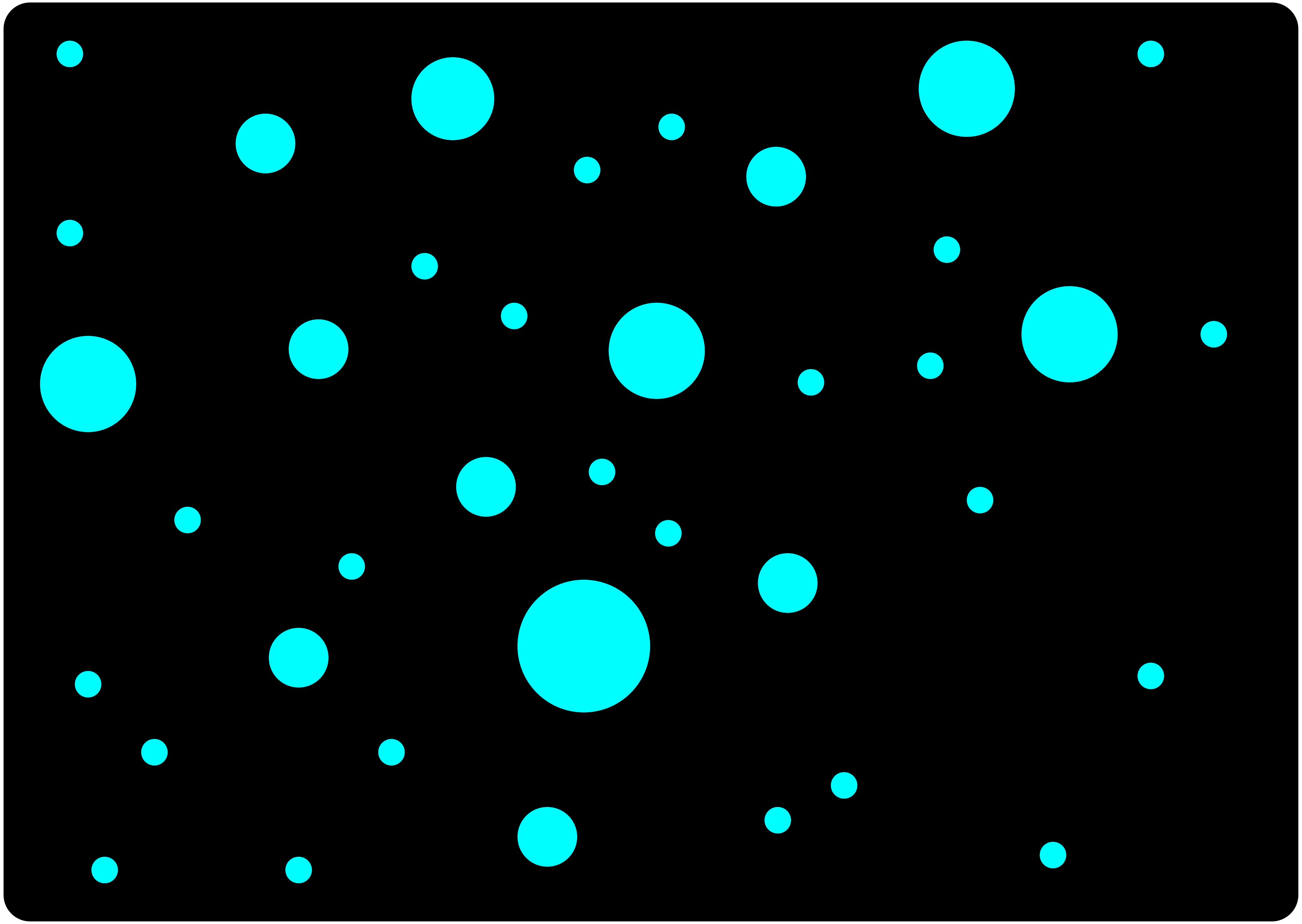
Dark matter haloes

- What defines a DM halo? Two main different methods:
  - Friends-of-Friends algorithm (FoF, e.g. Davis et al. 1985a)
  - Spherical overdensity method (SO, e.g. Warren et al. 1992)
- If dark matter haloes can be distinguished it is reasonable to assume that they are small in comparison with the common separation between them
- Two different regimes depending on the scale (even two different physics)

Friends of friends (FoF)



Halo population



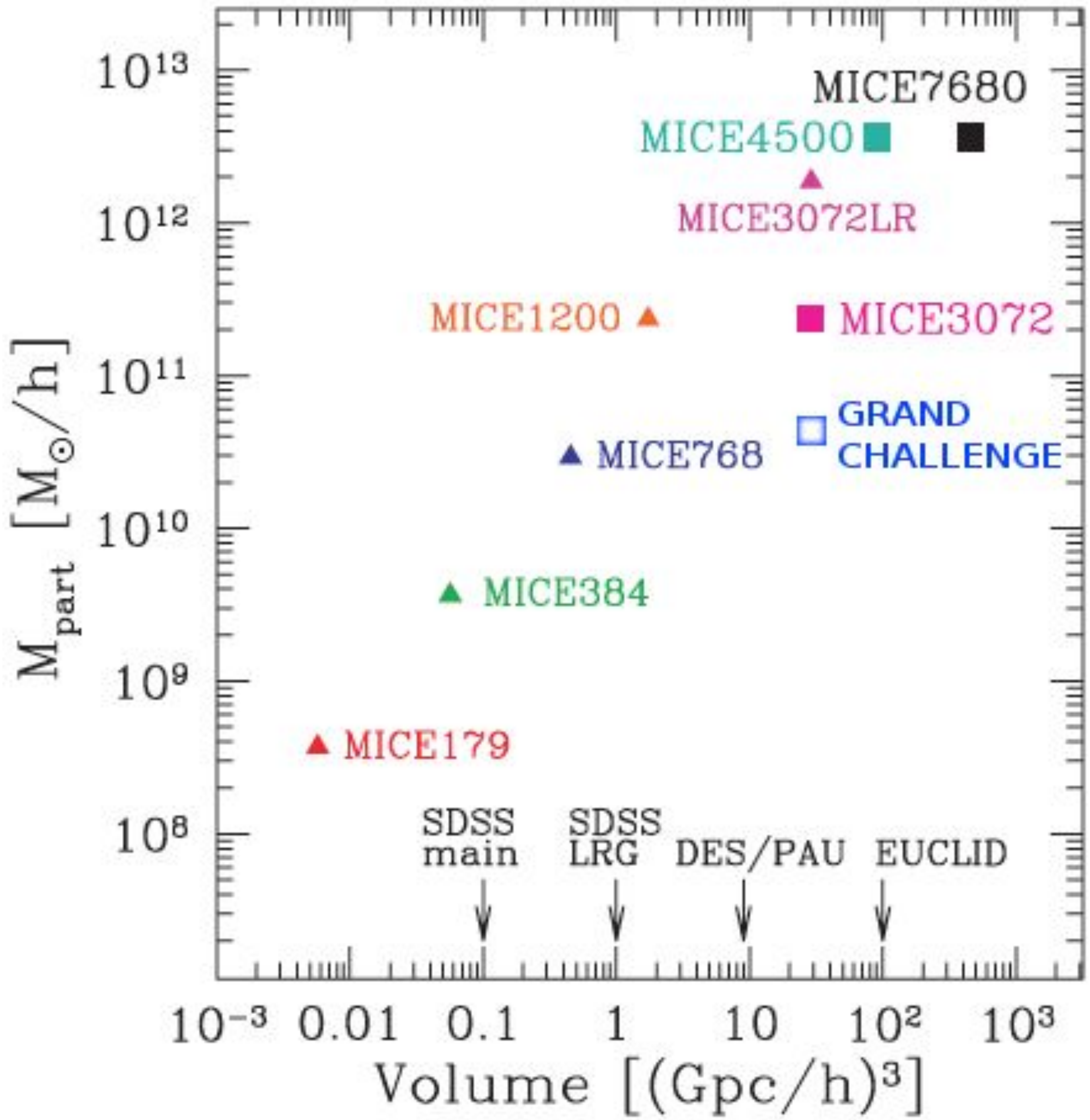
MICE simulations

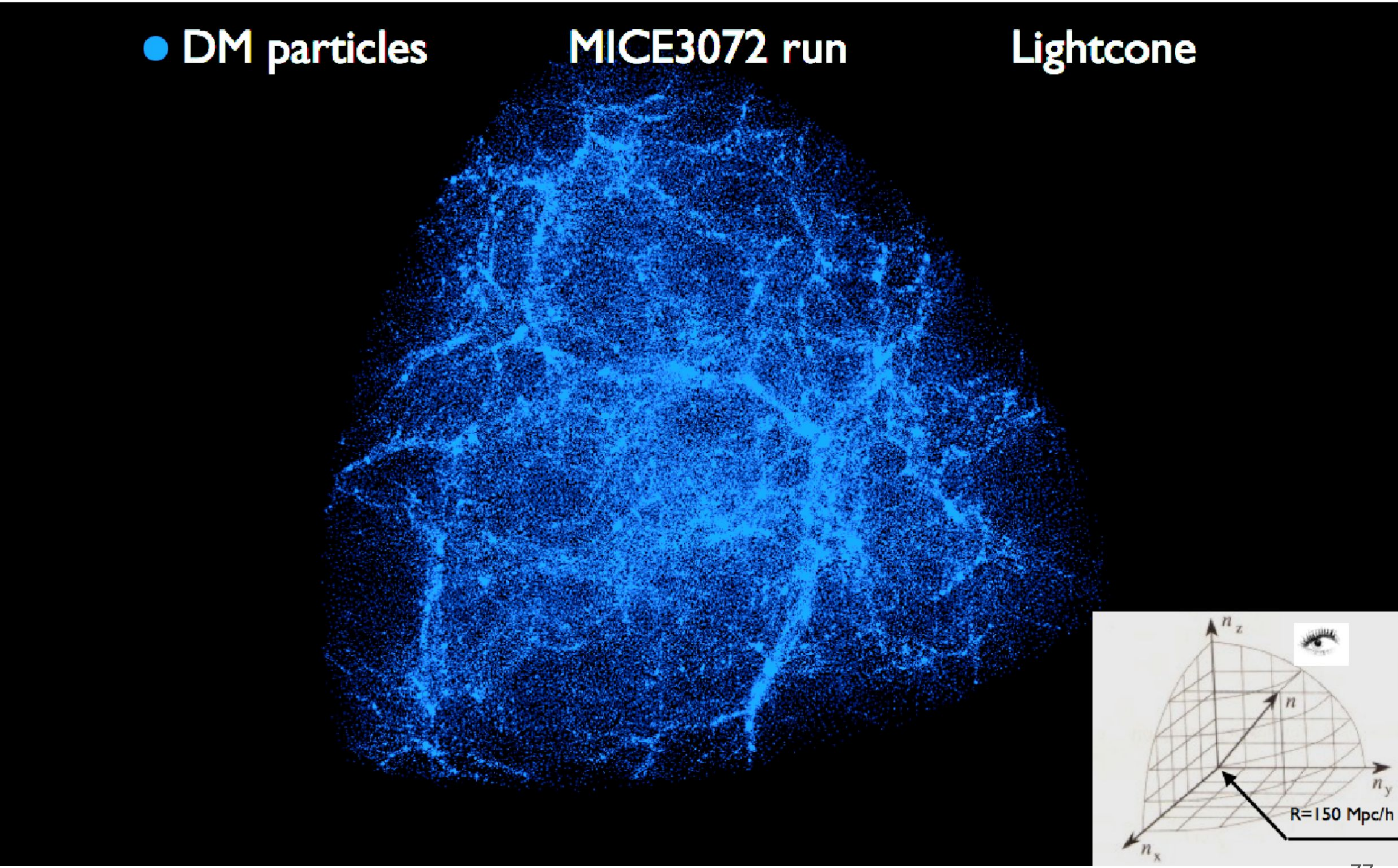
- Set of different DM N-body simulations specially suited to study large scale structure of the Universe
- Cosmological paramters: Ω_m=0.25, Ω_Λ=0.75, Ω_k=0, σ_8=0.8, n_s=0.95, h=0.7

MICE simulations

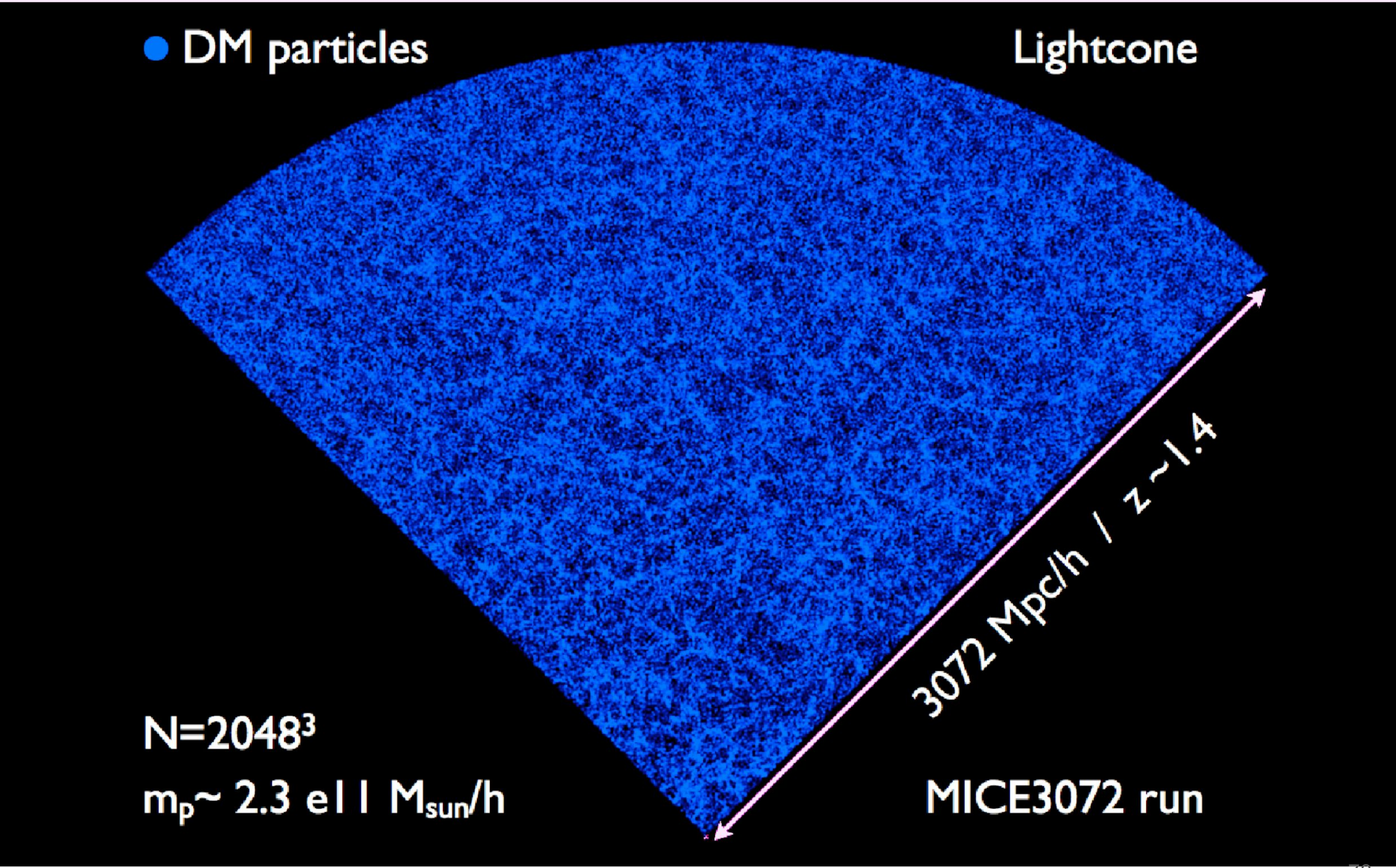
Run	$L_{box}/h^{-1}Mpc$	N_{part}	$m_p/h^{-1}M_{\odot}$	$l_{soft}/h^{-1}Kpc$	z_i
GRAND CHALLENGE	3072	4096^3	2.93×10^{10}	50	100
MICE7680	7680	2048^3	3.66×10^{12}	50	150
MICE3072	3072	2048^3	2.34×10^{11}	50	50
MICE4500	4500	1200^3	3.66×10^{12}	100	50
MICE3072LR*	3072	1024^3	1.87×10^{12}	50	50
MICE768*	768	1024^3	2.93×10^{10}	50	50
MICE384*	384	1024^3	3.66×10^{9}	50	50
MICE179*	179	1024^3	3.70×10^{8}	50	50
MICE1200* (x20)	1200	800^3	2.34×10^{11}	50	50

MICE simulations



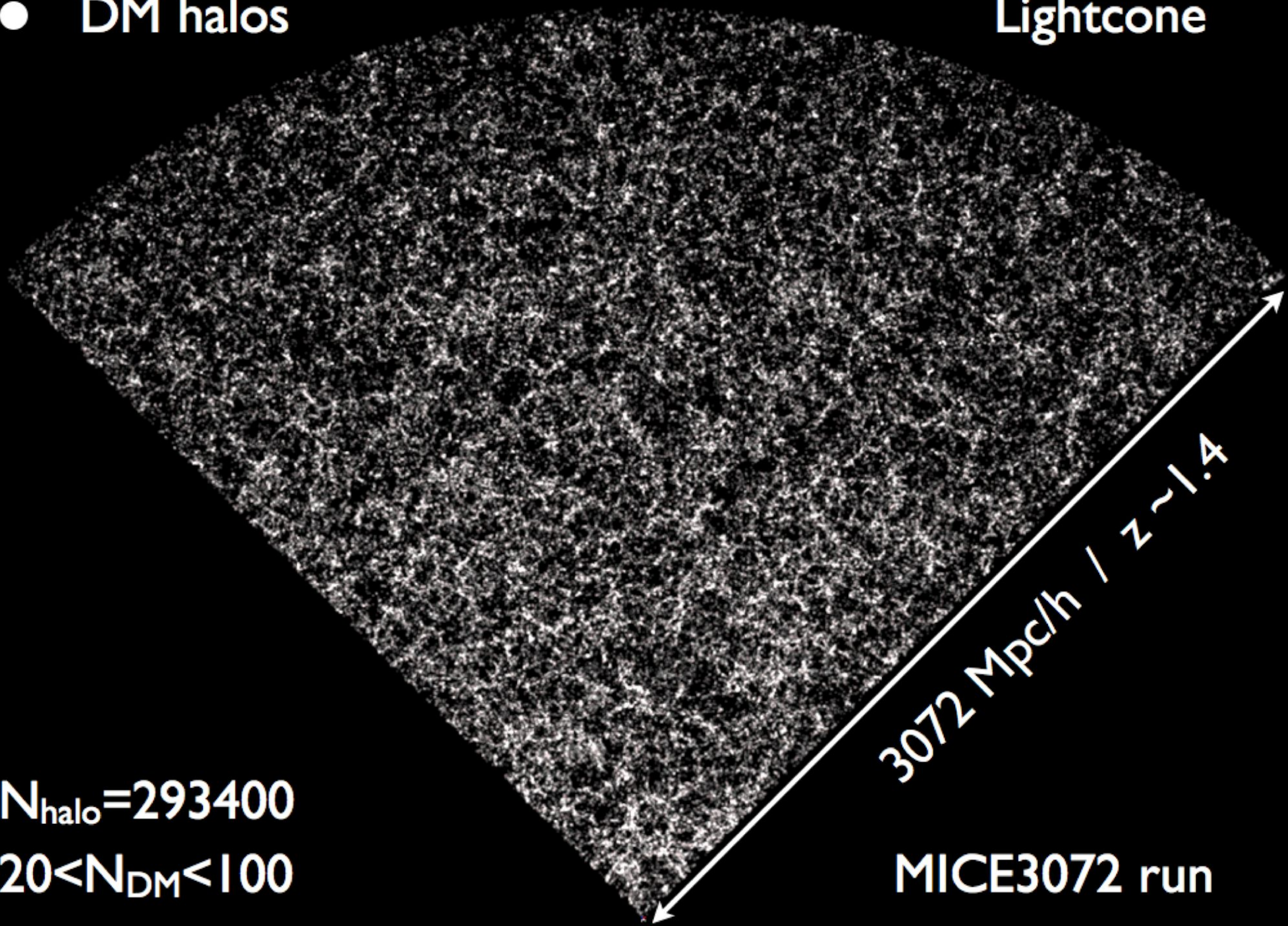


● DM particles
MICE3072 run
Lightcone
n_z
n
n_y
n_x
R=150 Mpc/h


DM particles
Lightcone
3072 Mpc/h / z ~1.4
N=2048³
mp~ 2.3 e11 Msun/h
MICE3072 run

- DM halos

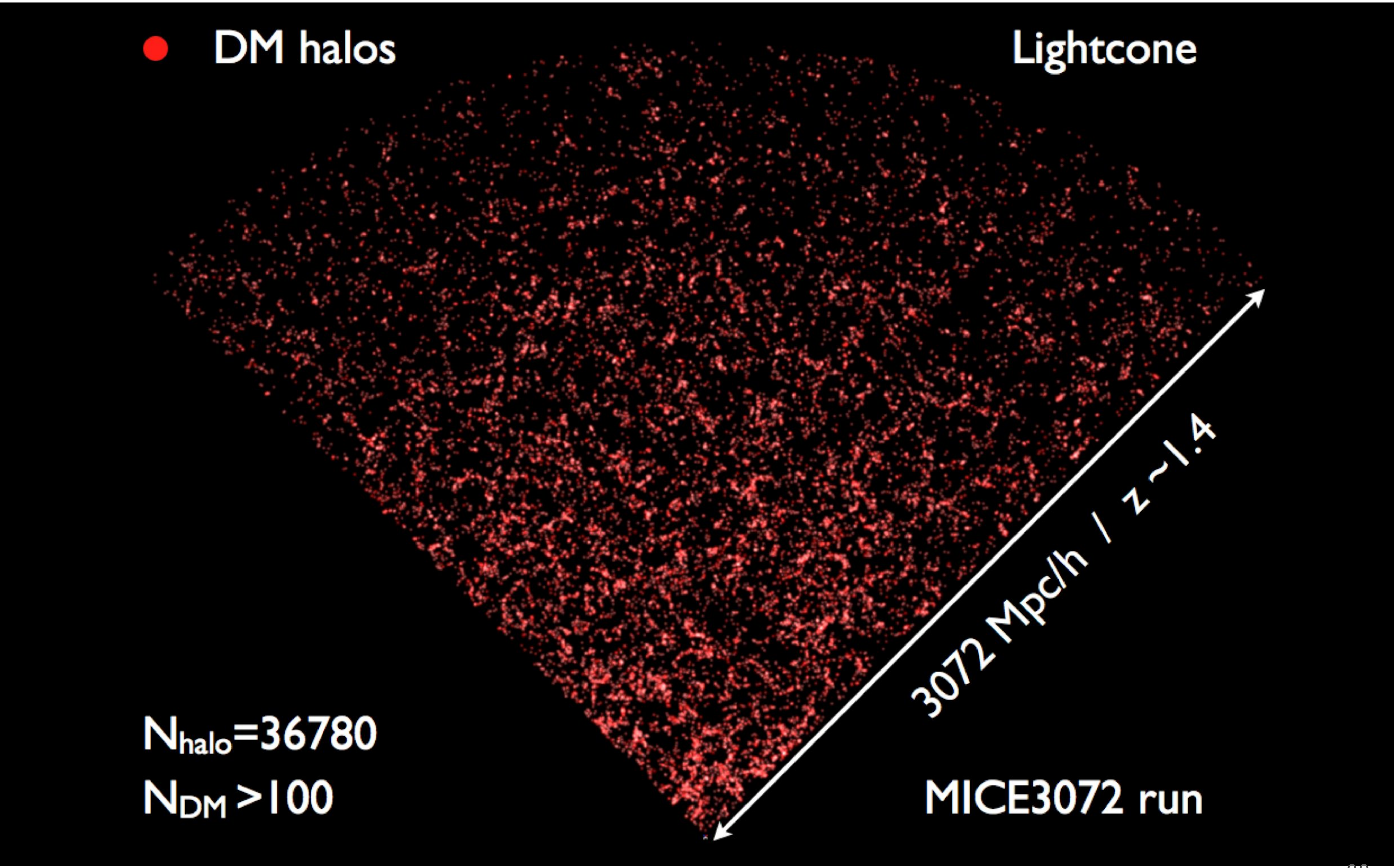
Lightcone

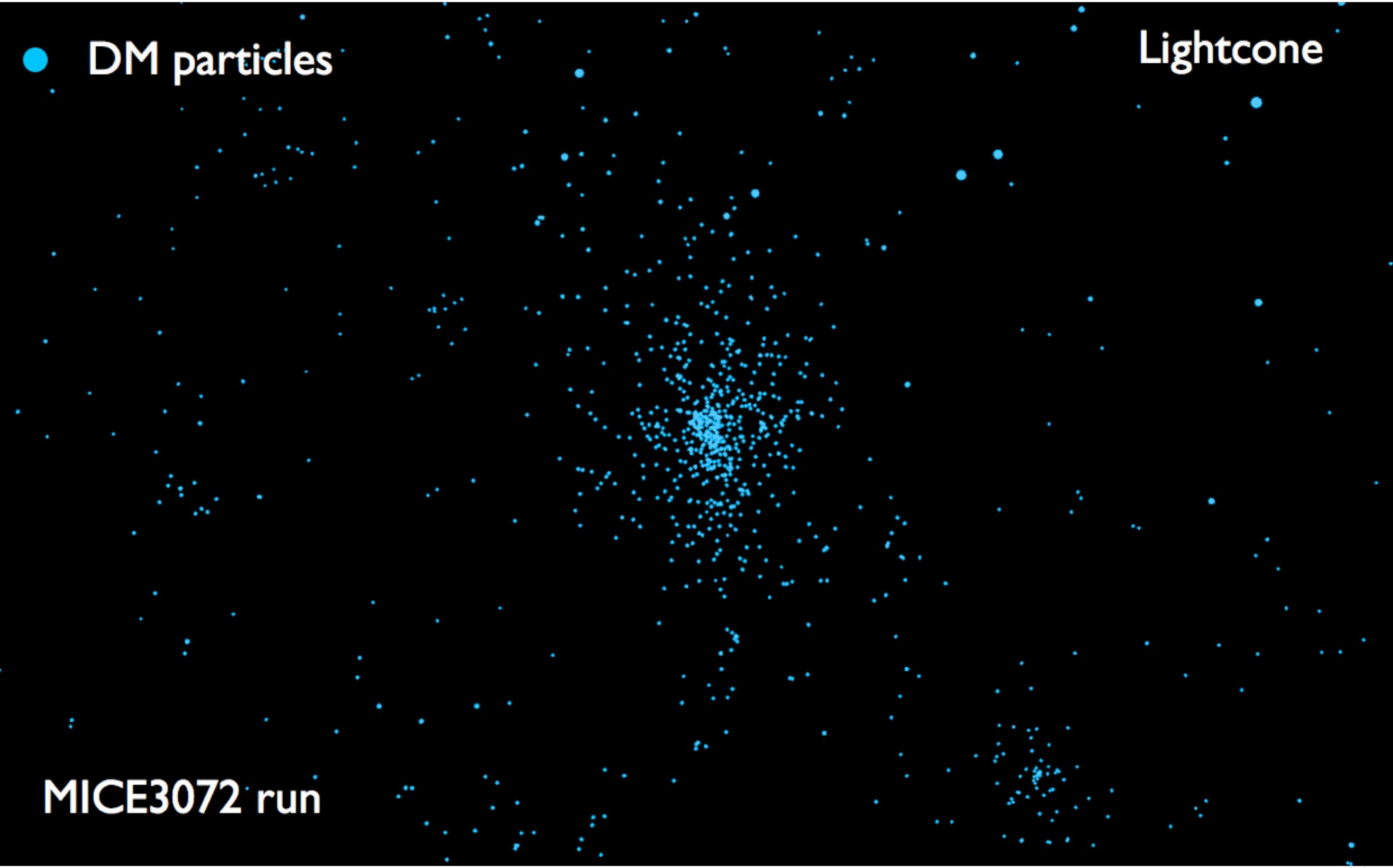


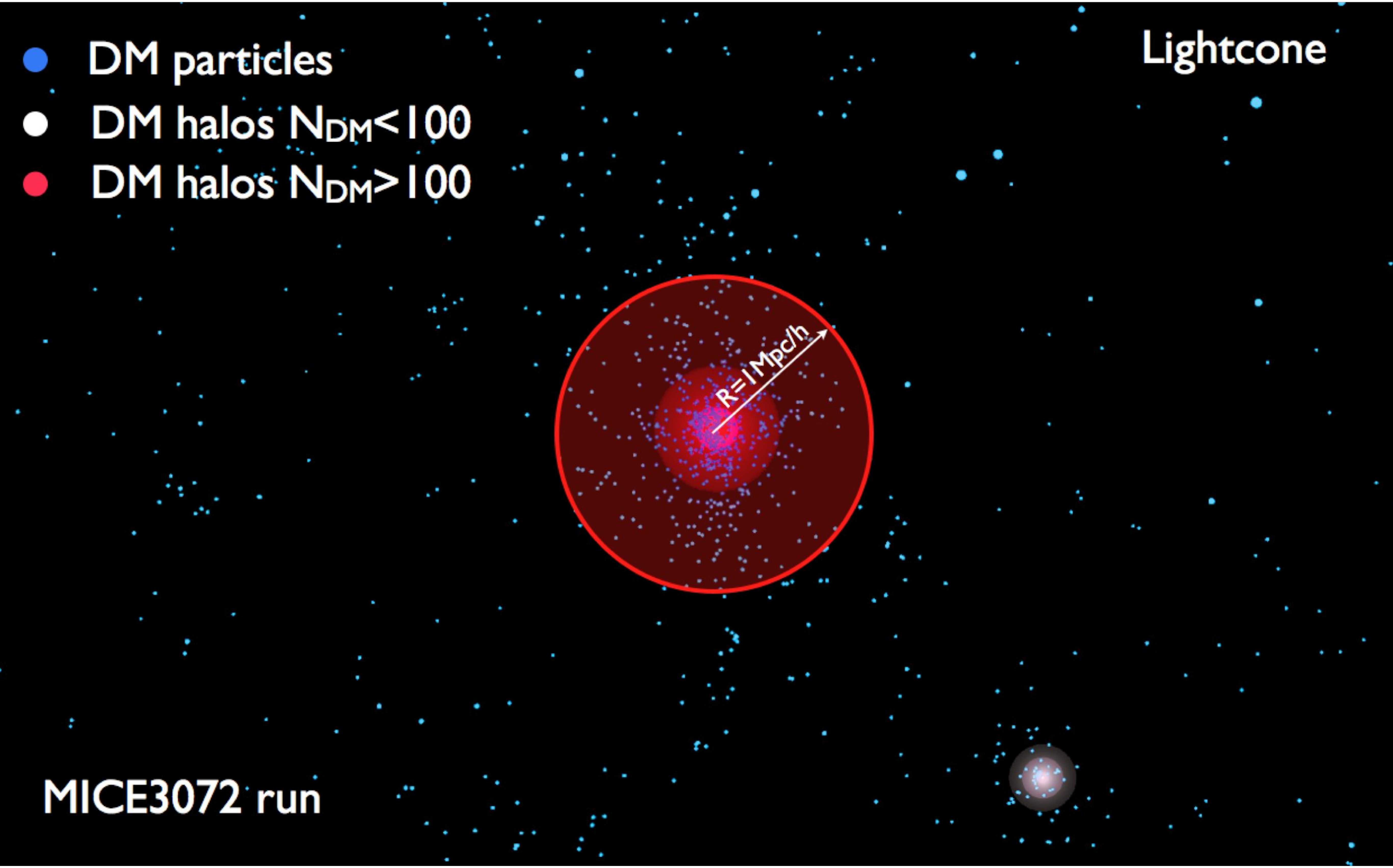
N_{halo}=293400

20<N_{DM}<100

MICE3072 run


DM halos
Lightcone
3072 Mpc/h / z ~1.4
Nhalo=36780
NDM >100
MICE3072 run


DM particles
Lightcone
MICE3072 run


DM particles
DM halos $N_{DM}<100$
DM halos $N_{DM}>100$
Lightcone
R=1Mpc/h
MICE3072 run

Statistics

- Two different regimes depending on the scale
- Statistics of the mass density field on small scales are mainly determined by the distribution of matter inside the haloes
- Halo mass density profiles are not relevant at scales larger than the size of a typical halo

$$\xi(r) \equiv < \delta(\mathbf{x})\delta(\mathbf{x}+r) > \qquad \delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}$$

$$\xi(r) = \xi^{1h}(r) + \xi^{2h}(r)$$

Statistics

- Two different regimes depending on the scale
- Statistics of the mass density field on sm[...]ainly determined by the distribution of [...] haloes
- Halo mass density profiles a[...] scales larger than the size of a typical halo

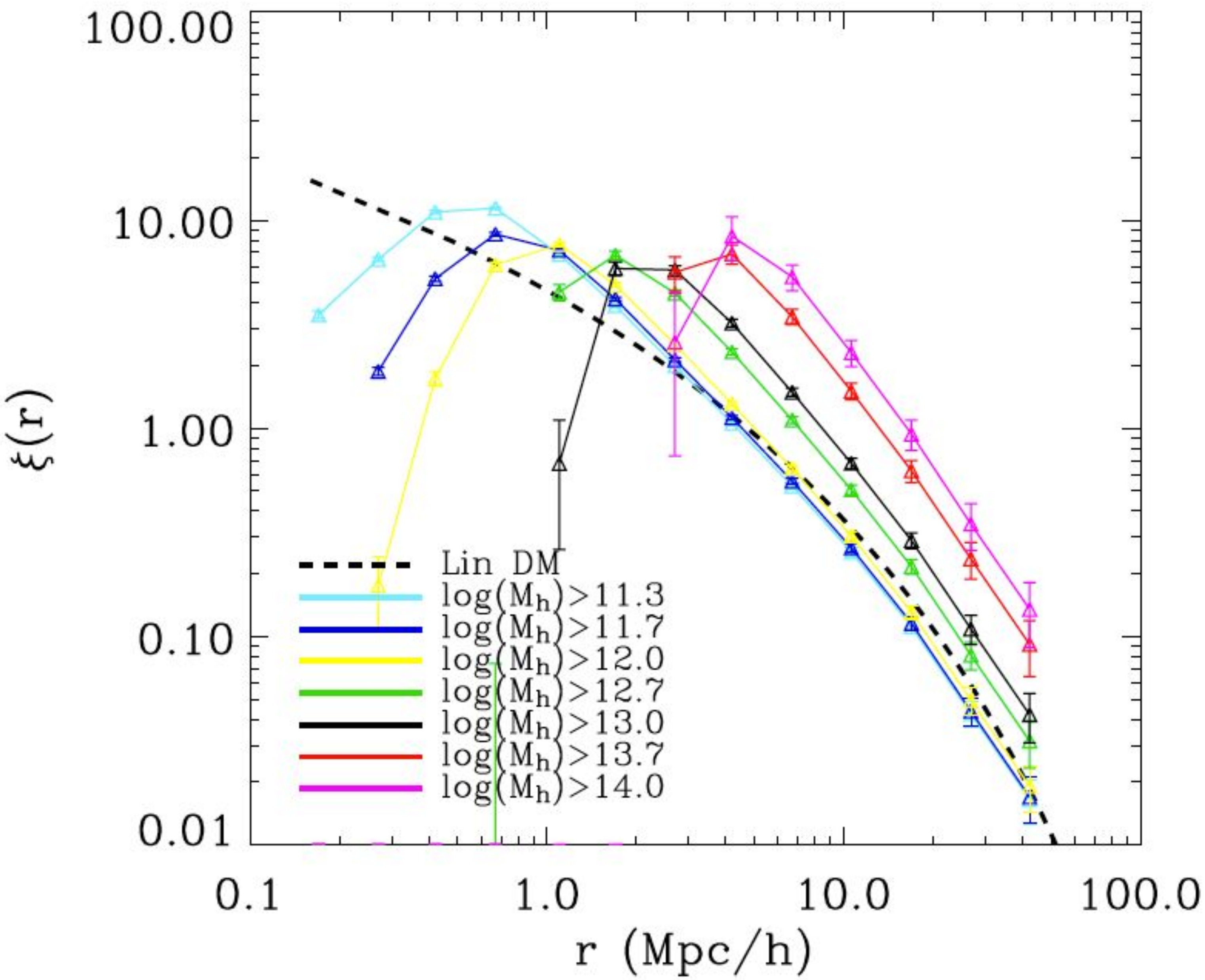
A comparison of estimators for the two-point correlation function (Kerscher, Szapudi & Szalay; 2000)

$$\ldots(\mathbf{x}+r)>$$

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}$$

$$\xi(r) = \xi^{1h}(r) + \xi^{2h}(r)$$

Halo spatial distribution



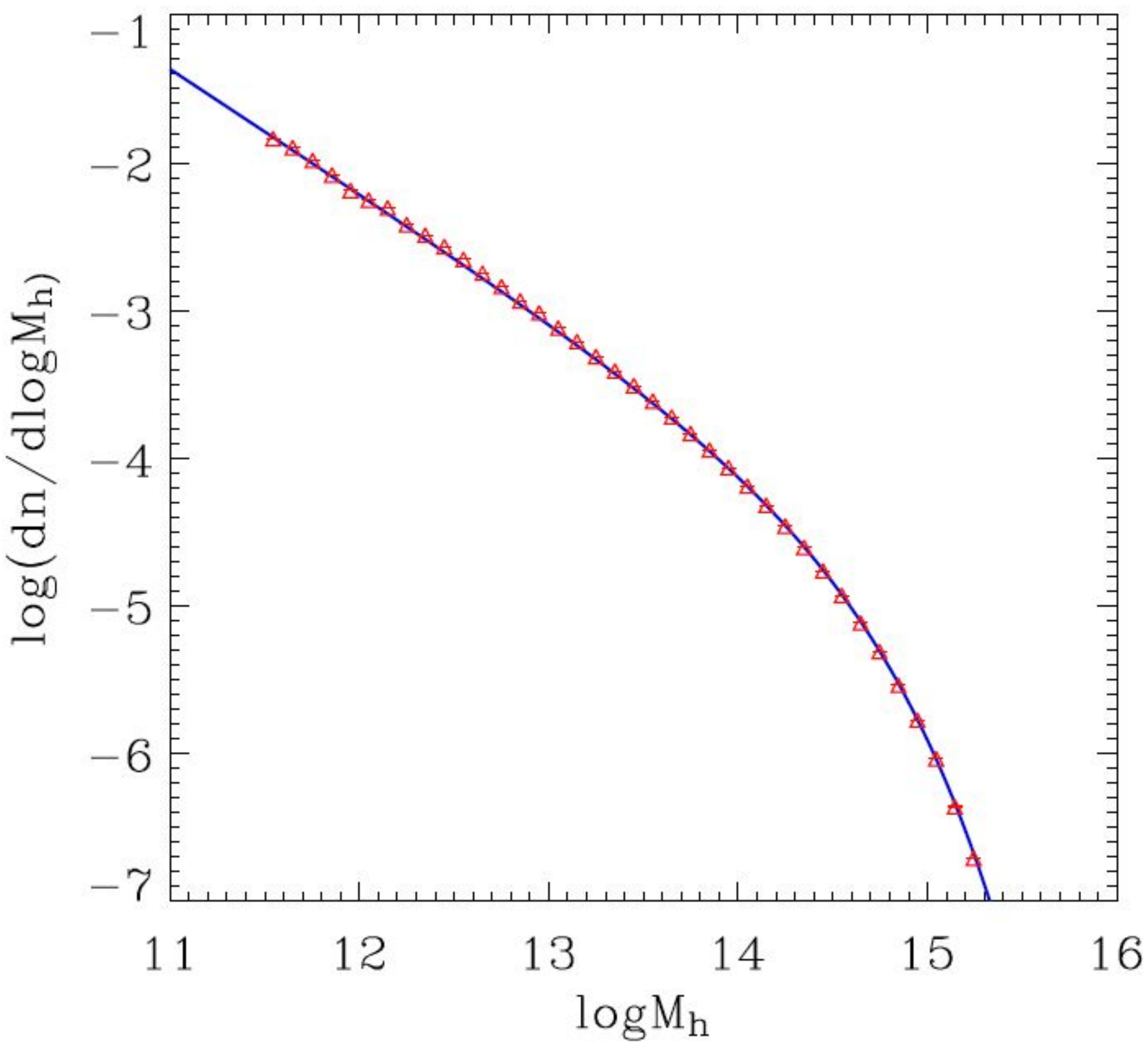
$$\xi_h(r) = \frac{DD(r)}{RR(r)} - 1$$

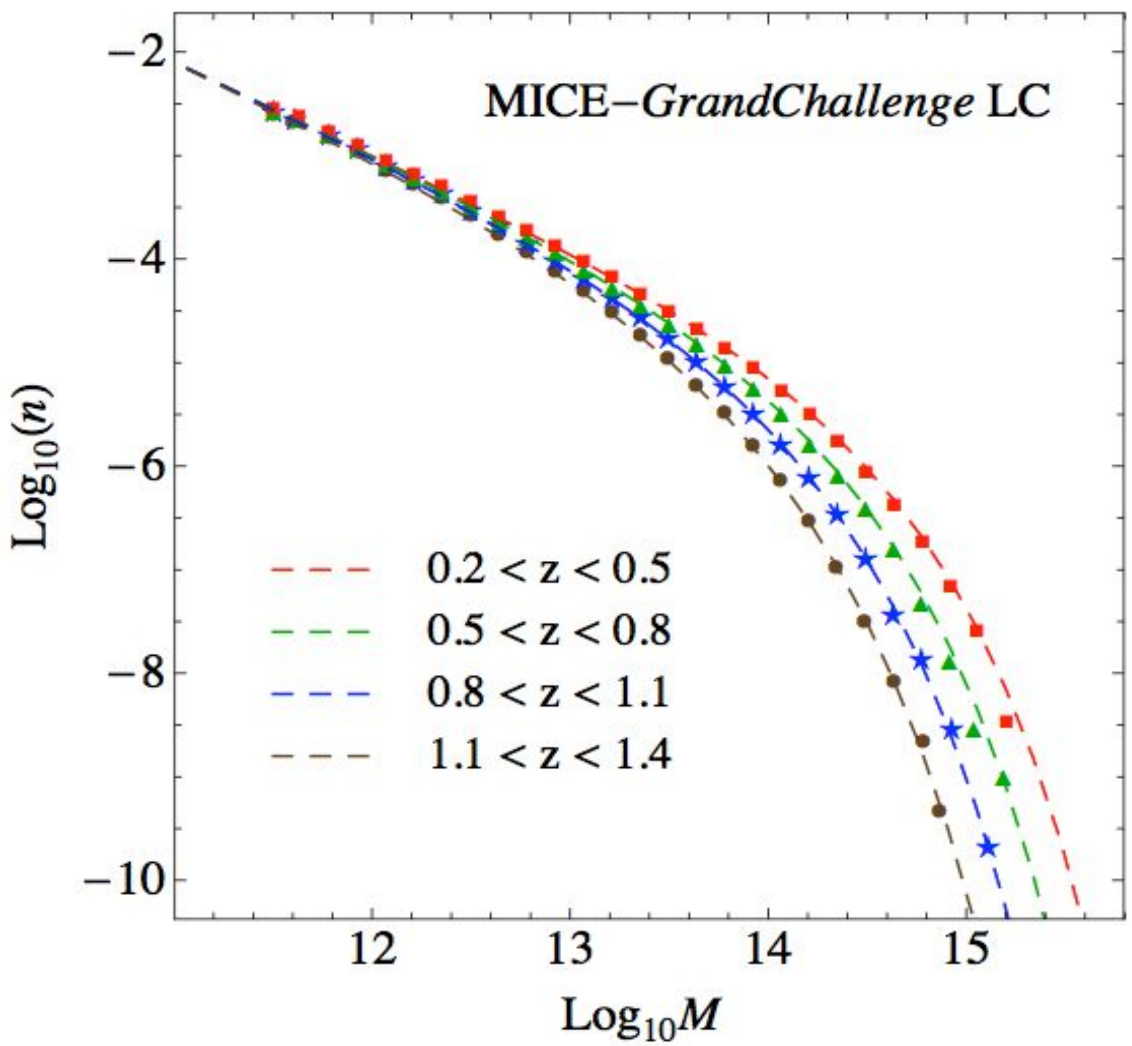
Halo abundance

- The halo mass function gives the mean number density of haloes as a function of their mass
- Several analytical derivations and fits have been provided in the literature, e.g:
    - Press-Schechter formalism (1974)
    - Sheth & Tormen (1999), Jenkins et al. (2001)

$$\boxed{f_{ST}(\nu) = A\sqrt{\frac{\alpha\nu^2}{2\pi}}\left[1 + \frac{1}{(\alpha\nu^2)^p}\right]\exp\left(\frac{-\alpha\nu^2}{2}\right)}$$

Halo Mass Function



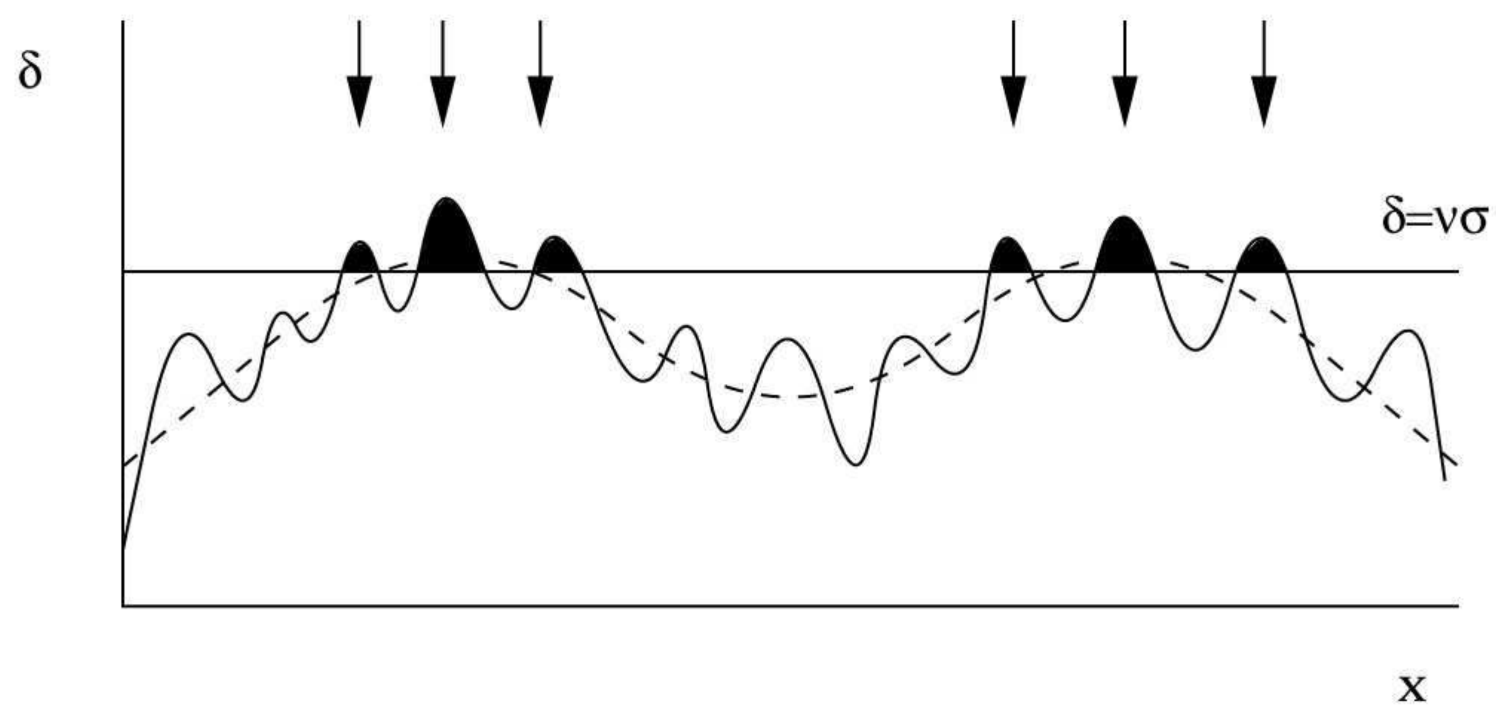


$$f(\sigma, z) = A(z)\left[\sigma^{-a(z)} + b(z)\right]\exp\left[-\frac{c(z)}{\sigma^2}\right]$$

ST-like MF

Halo bias

- **DM haloes are bias tracers of the underlying DM distribution**
- The halo spatial distribution and the halo abundance are not independent (peak-background split)



Halo bias

- Several analytical derivations or fits:
  - Mo & White 1996 (following the spherical collapse model):

$$\boxed{b(\nu(M_h)) = 1 + \frac{\nu^2 - 1}{\delta_c}}$$

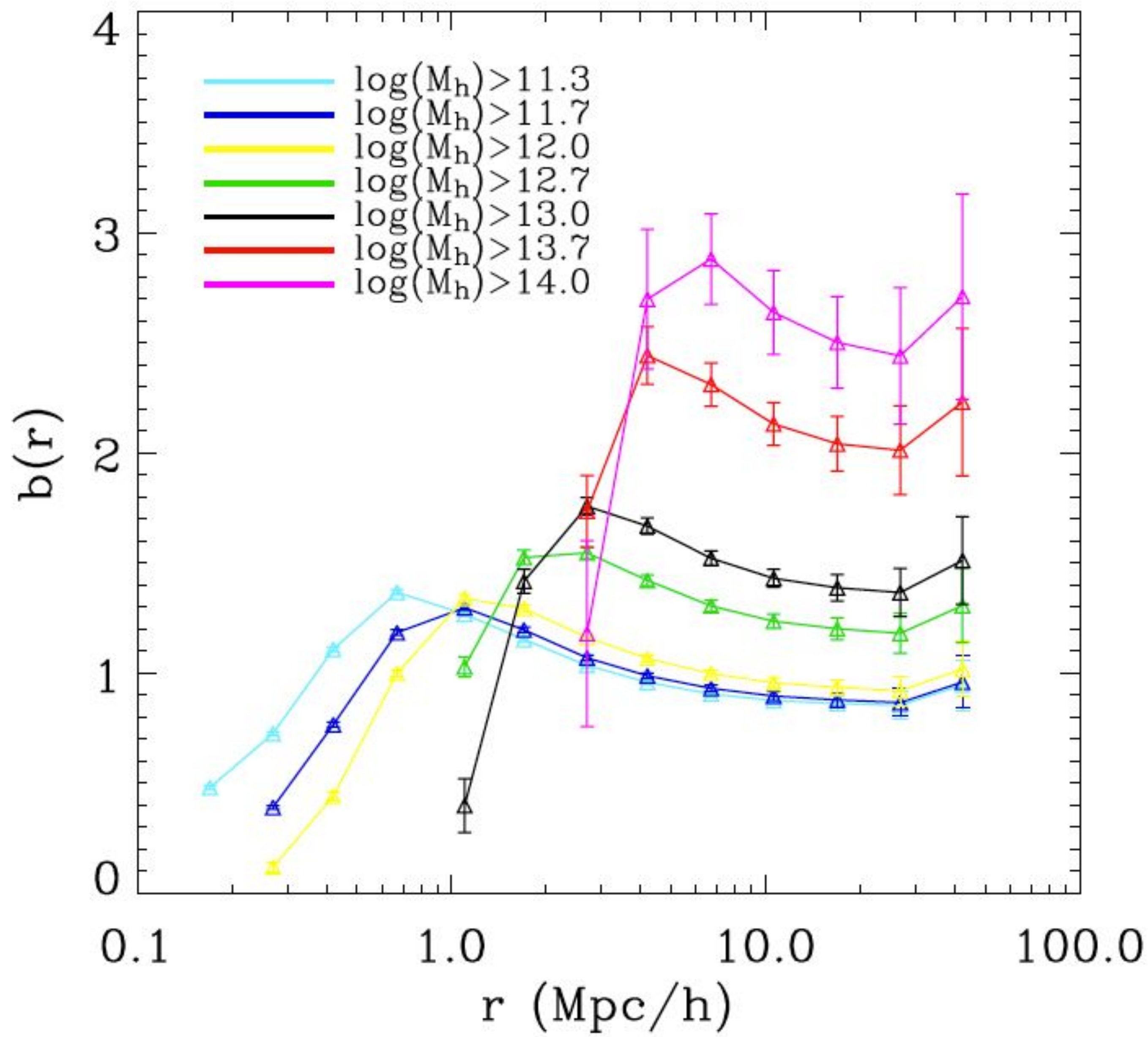
  - Sheth & Tormen 1999 (more general form):

$$\boxed{b(\nu(M_h)) = 1 + \frac{a\nu^2 - 1}{\delta_c} + \frac{2p}{\delta_c[1 + (a\nu^2)^p]}}$$

  - Manera et al. 2010 (using a Warren-form MF)

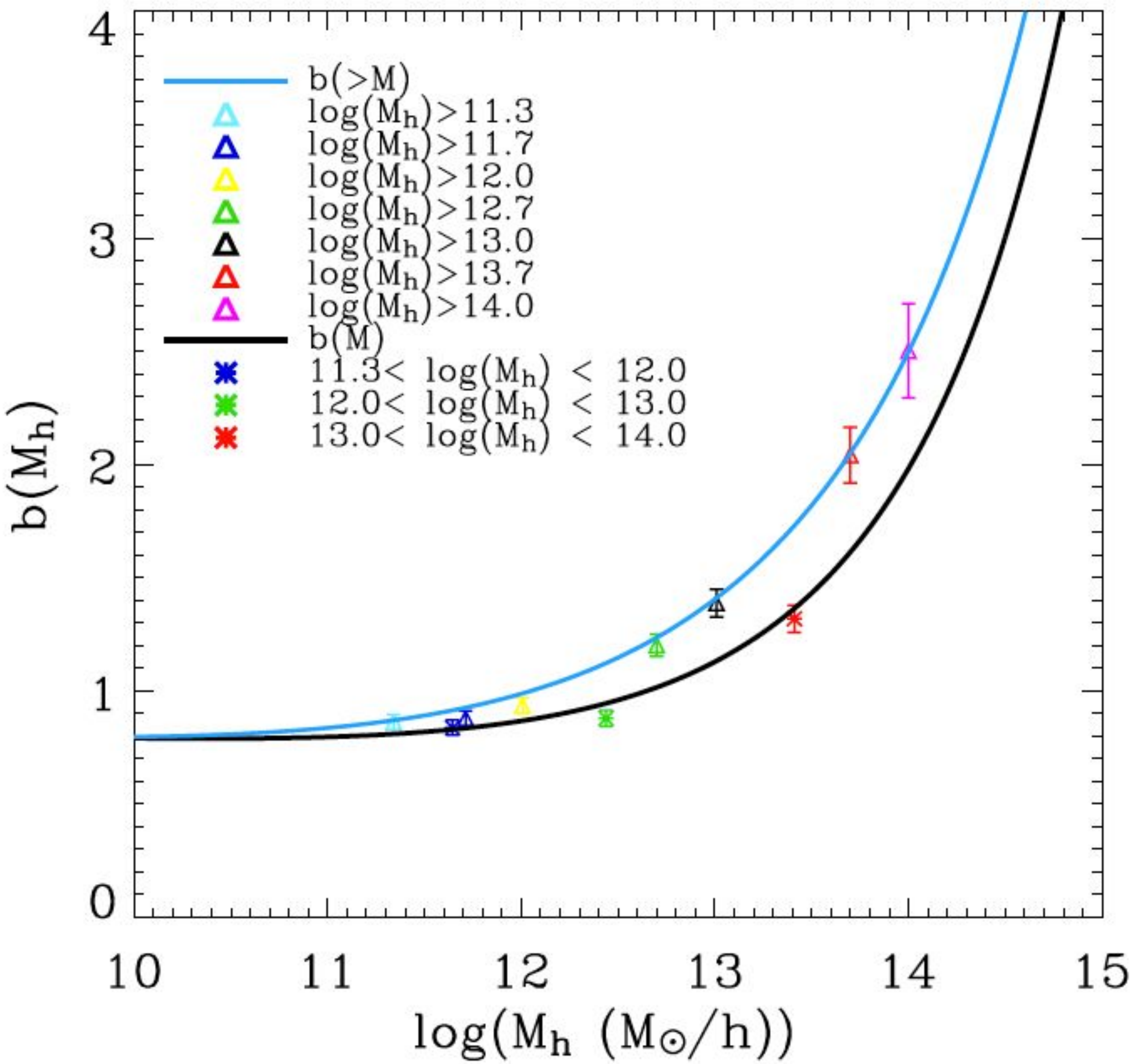
$$\boxed{b^W\left(\nu\left(\sigma\left(M_h\right)\right)\right) = 1 + \frac{c^{'}\nu - 1}{\delta_c} + \frac{2a^{'}b^{'} + b^{'} + \left(c^{'}\nu\right)^{a^{'}}}{\delta_c\left(b^{'} + (c^{'}\nu)^{a^{'}}\right)}}$$

Halo bias

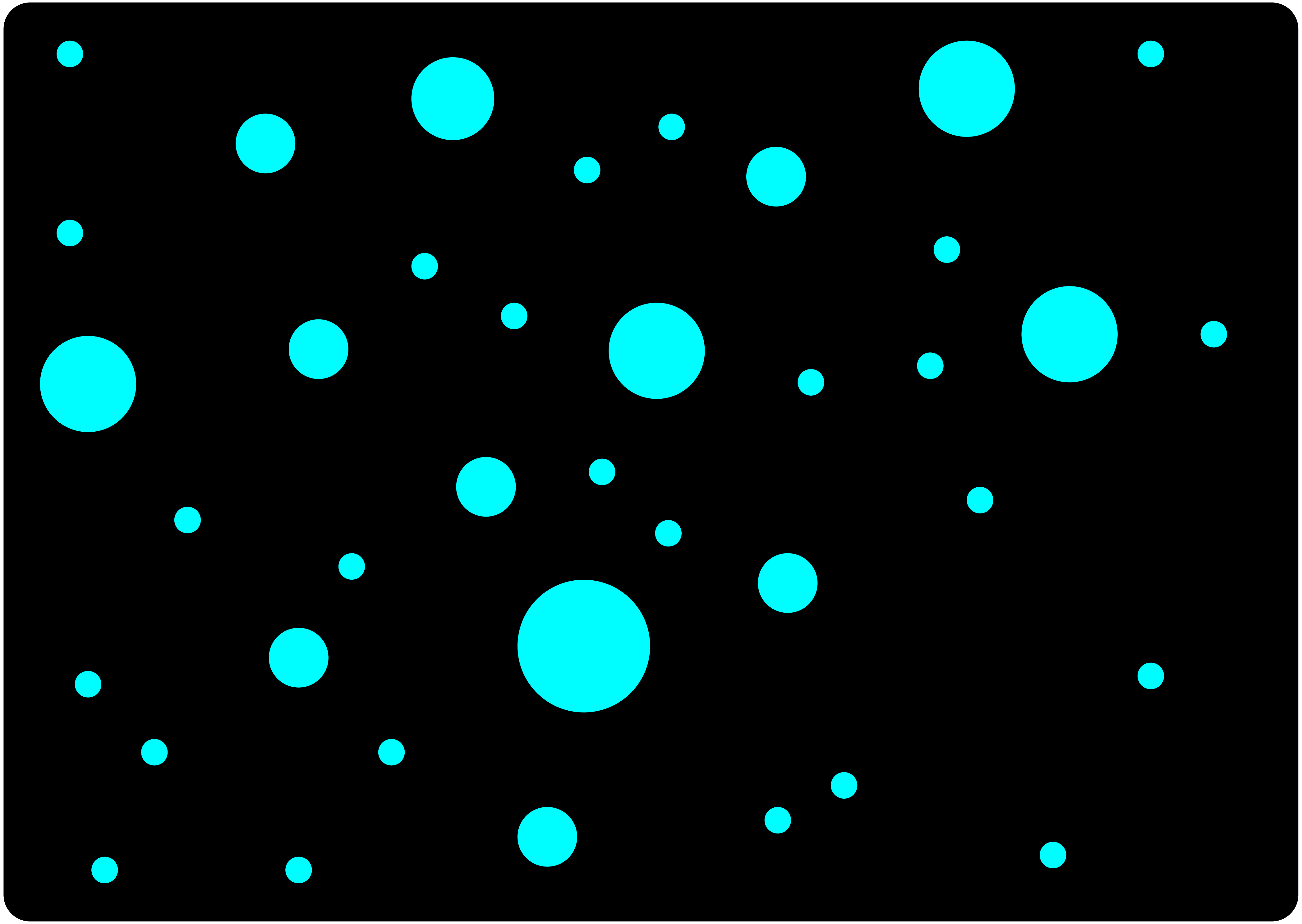


$$\xi_h(r) = \left(b_h^{Lin}\right)^2 \xi_{DM}(r)$$

Halo bias


b(>M)
log(Mh)>11.3
log(Mh)>11.7
log(Mh)>12.0
log(Mh)>12.7
log(Mh)>13.0
log(Mh)>13.7
log(Mh)>14.0
b(M)
11.3< log(Mh) < 12.0
12.0< log(Mh) < 13.0
13.0< log(Mh) < 14.0
b(Mh)
log(Mh (M⊙/h))
0
1
2
3
4
10
11
12
13
14
15

Halo catalogue

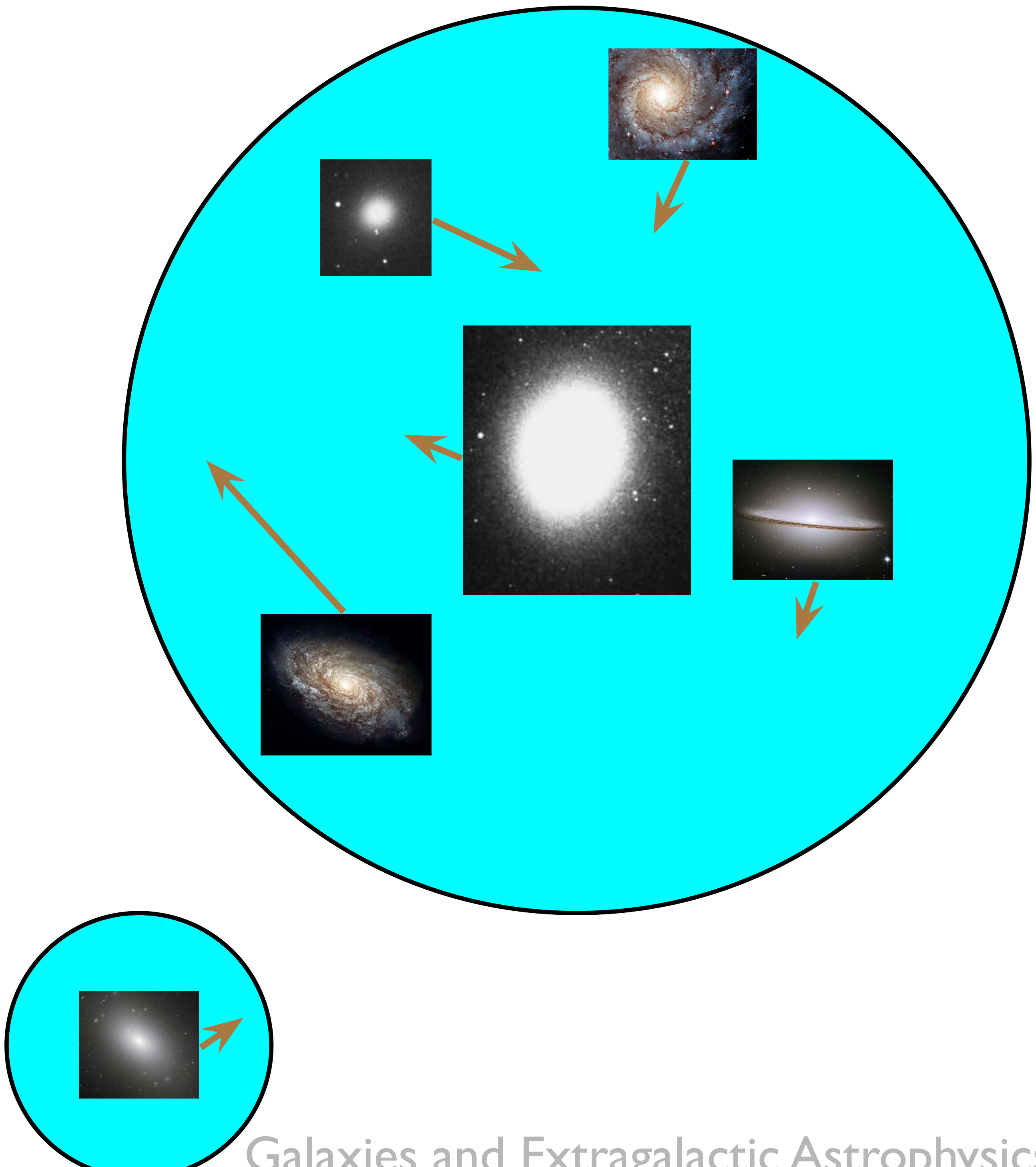


HOD

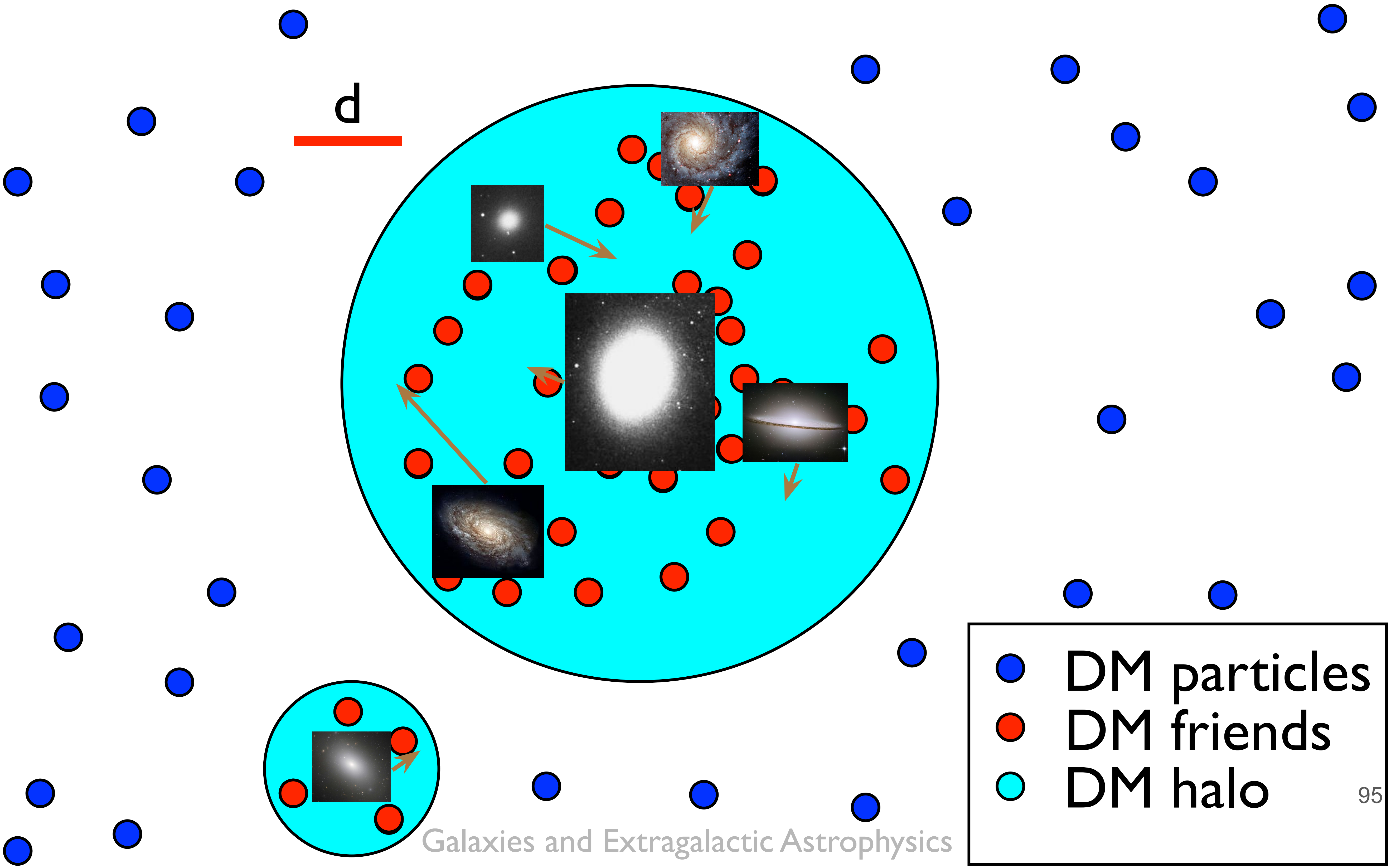
- Relate dark matter and galaxies, "galaxy bias", (following Martinez & Saar 2002):

6. Probability that a DM halo contains Ng galaxies
7. Relation between the galaxy and the DM spatial distribution
8. Relation between the galaxy and the DM velocity distribution

FoF



FoF + HOD



Number of galaxies

- Galaxies separated into centrals or satellites
- $P(N_{gal} \mid M_h)$, different functional forms and number of free parameters, e.g.:

$$\boxed{N_{gal}(>L_r|M_h) = 1 + \left[\frac{M_h}{M_1(L_r)}\right]^{\alpha} \quad \text{if} \quad M_h \geq M_{min}(L_r)}$$

    - M_{min}, minimum mass to form a galaxy
    - M_1, mass that on average the halo contains 1 satellite galaxy
    - α, related to the galaxy formation efficiency
- HOD parameters provide physical meaning (Berlind & Weinberg 2002)

Galaxy spatial distribution

- Different possibilities to place galaxies inside the DM haloes, e.g.:
  - DM particles of the halo “are” galaxies
  - Galaxies follow a NFW density profile (from N-body simulations)
- Mainly affects the small scales of the two-point correlation function

Galaxy velocity distribution

- Haloes are assumed to be virialized (virial theorem holds)
- Two main possibilities:
  - $v_{gal} = v_{DM}$
  - $v_{gal} = v_{vir} + v_{halo}$ (virial motions are well approximated by velocities that are independent Gaussians in each cartesian component with rms values that depends on M_h)

Modeling $\xi_{gal}(r)$

- Two contributions: 1- and 2-halo terms
- 1-halo term: only galaxy pairs that reside in the same halo (small scales)
- 2-halo term: dominated by galaxies that reside in different haloes (>=5Mpc/h)
- 1 halo term: Appendix B of Tinker et al. 2005 (improved the halo exclusion effect)
- 2 halo term: P(k) (in Fourier space)

Algorithm

- This algorithm is described in Carretero et al. (2014). http://arxiv.org/abs/1411.3286

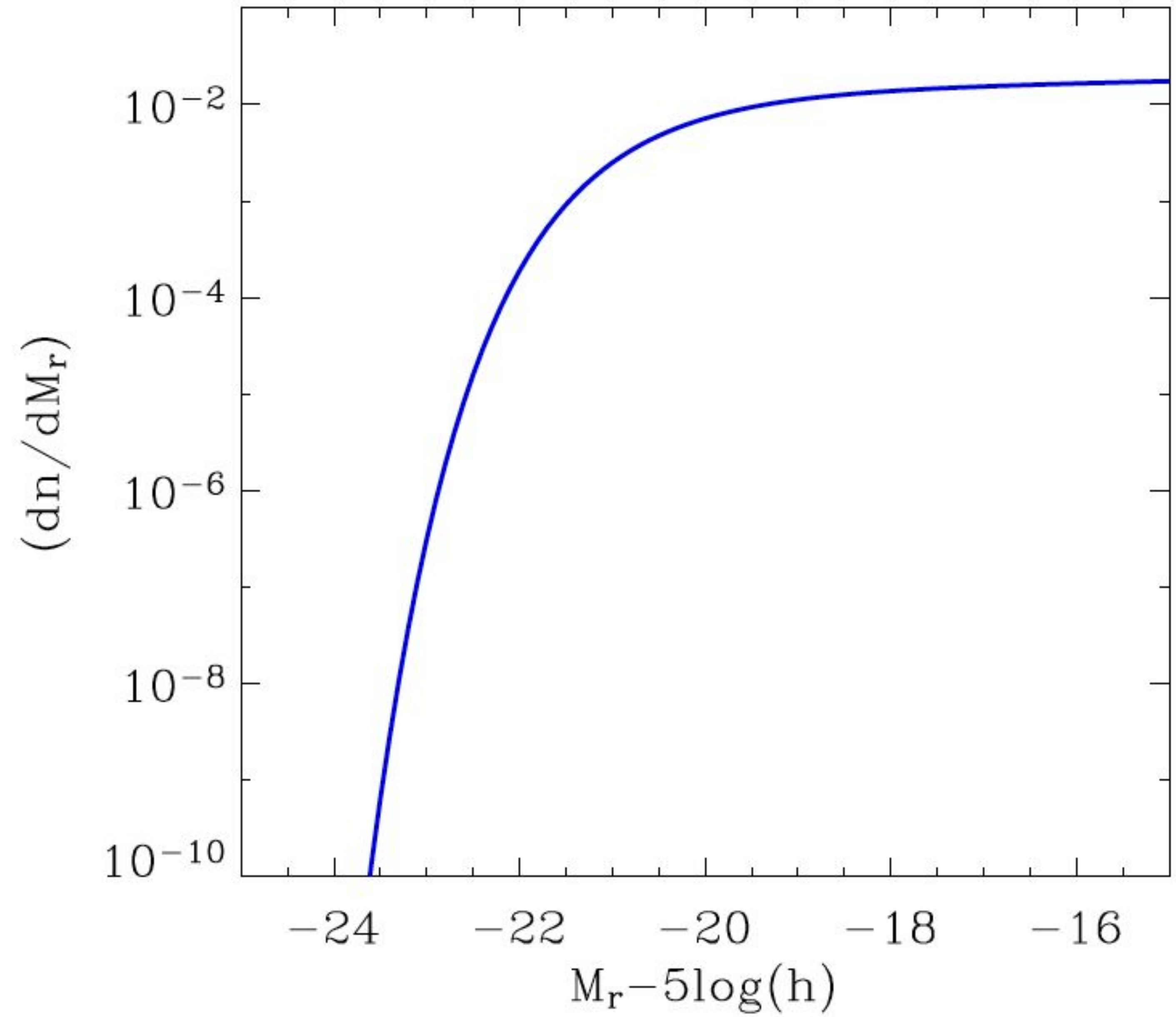
- Hybrid method: HOD + SHAM
- The generated galaxy catalogues follow:
  - Observed luminosity function (Blanton et al. 2003)
  - Observed galaxy clustering as a function of luminosity and colour (Zehavi et al. 2011)
- To download the catalogs: http://cosmohub.pic.es

Algorithm

- This algorithm is described in Carretero et al. (2014).
  http://arxiv.org/abs/1411.3286

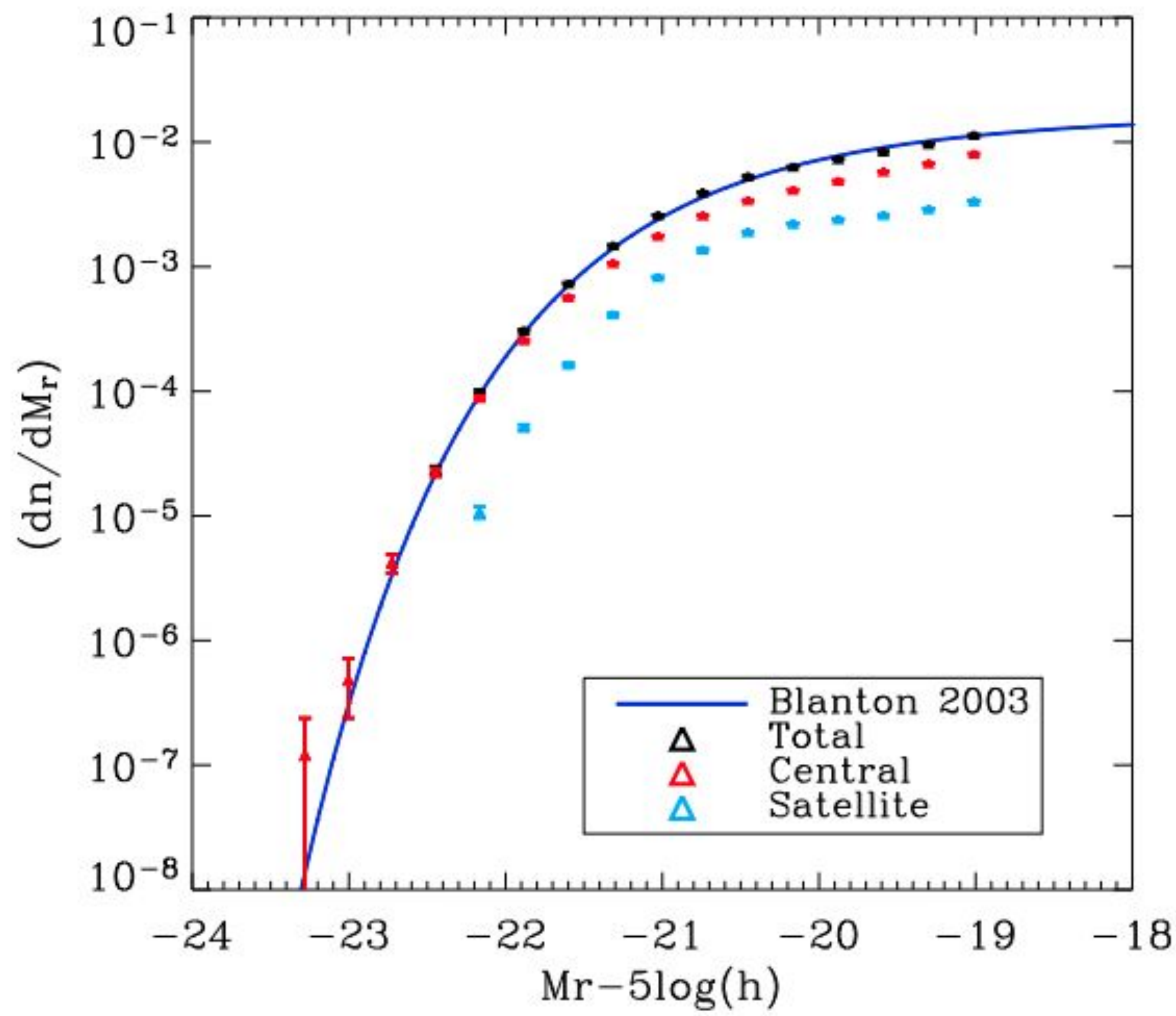
  Updated algorithm in Castander et al. (2024)
  https://arxiv.org/pdf/2405.13495
- Hybrid method: HOD + SHAM
- The generated galaxy catalogues follow:
  - Observed luminosity function (Blanton et al. 2003)
  - Observed galaxy clustering as a function of luminosity and colour (Zehavi et al. 2011)
- To download the catalogs: http://cosmohub.pic.es

SDSS Schechter LF



$$\phi(M_r) = \frac{dn}{dM_r} = \frac{2}{5}\phi^* \ln(10) \left(10^{\frac{2}{5}(M_r^* - M_r)}\right)^{\alpha+1} \exp\left(-10^{\frac{2}{5}(M_r^* - M_r)}\right)$$

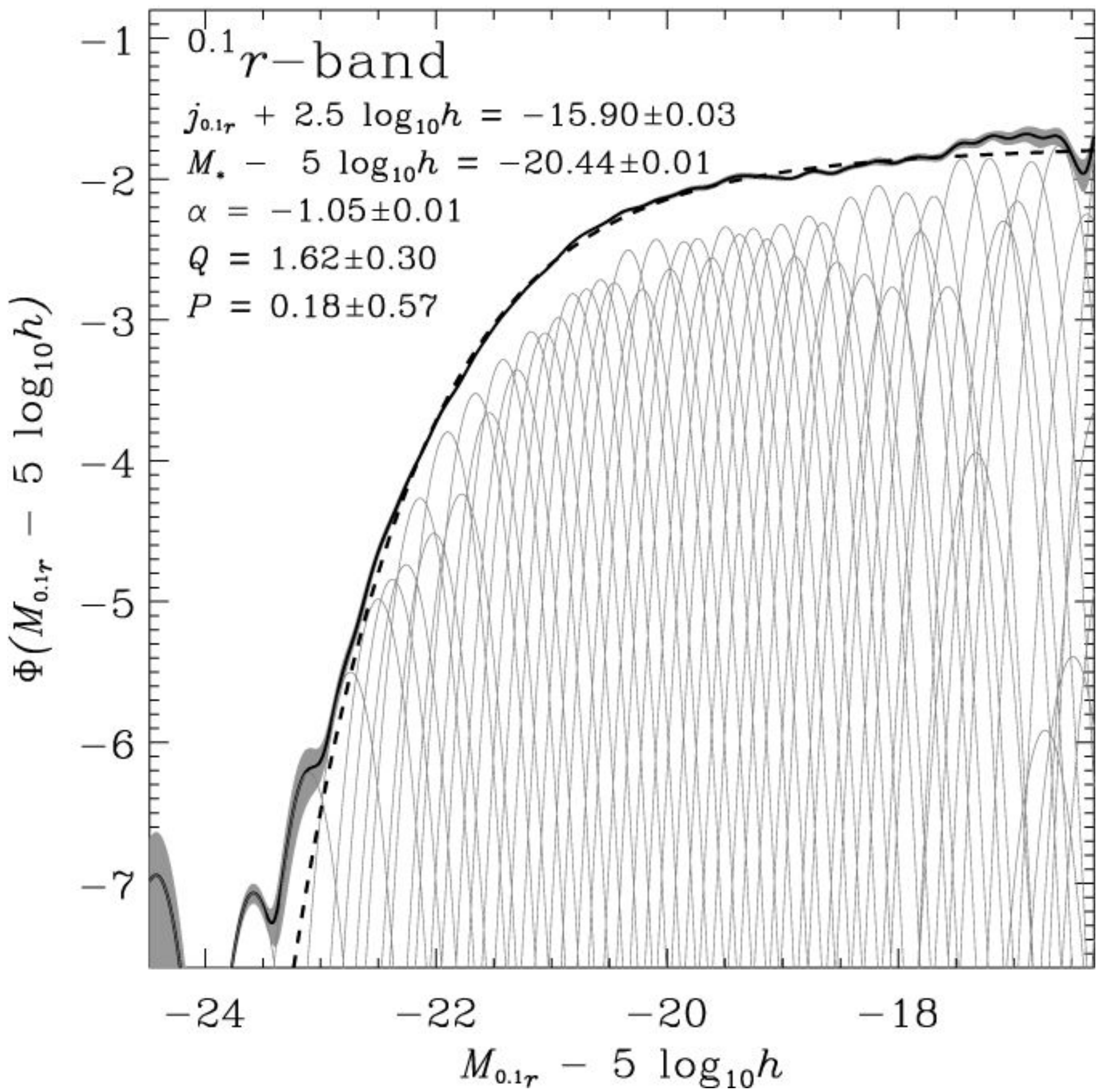
SDSS Schechter LF



$$\phi(M_r) = \frac{dn}{dM_r} = \frac{2}{5}\phi^* \ln(10) \left(10^{\frac{2}{5}(M_r^* - M_r)}\right)^{\alpha+1} \exp\left(-10^{\frac{2}{5}(M_r^* - M_r)}\right)$$

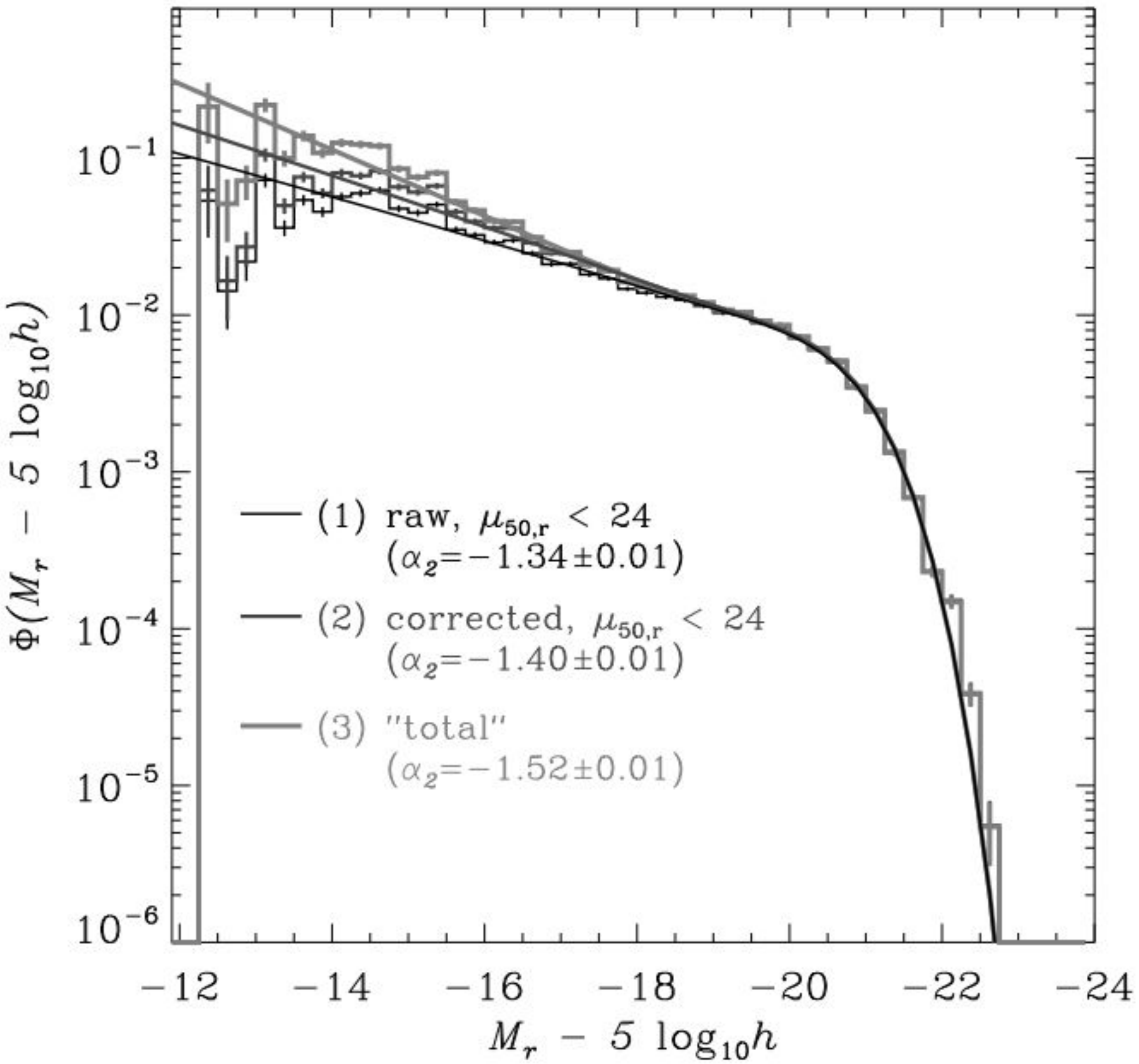
SDSS Schechter LF

Blanton et al. (2003) (astro-ph/0210215)



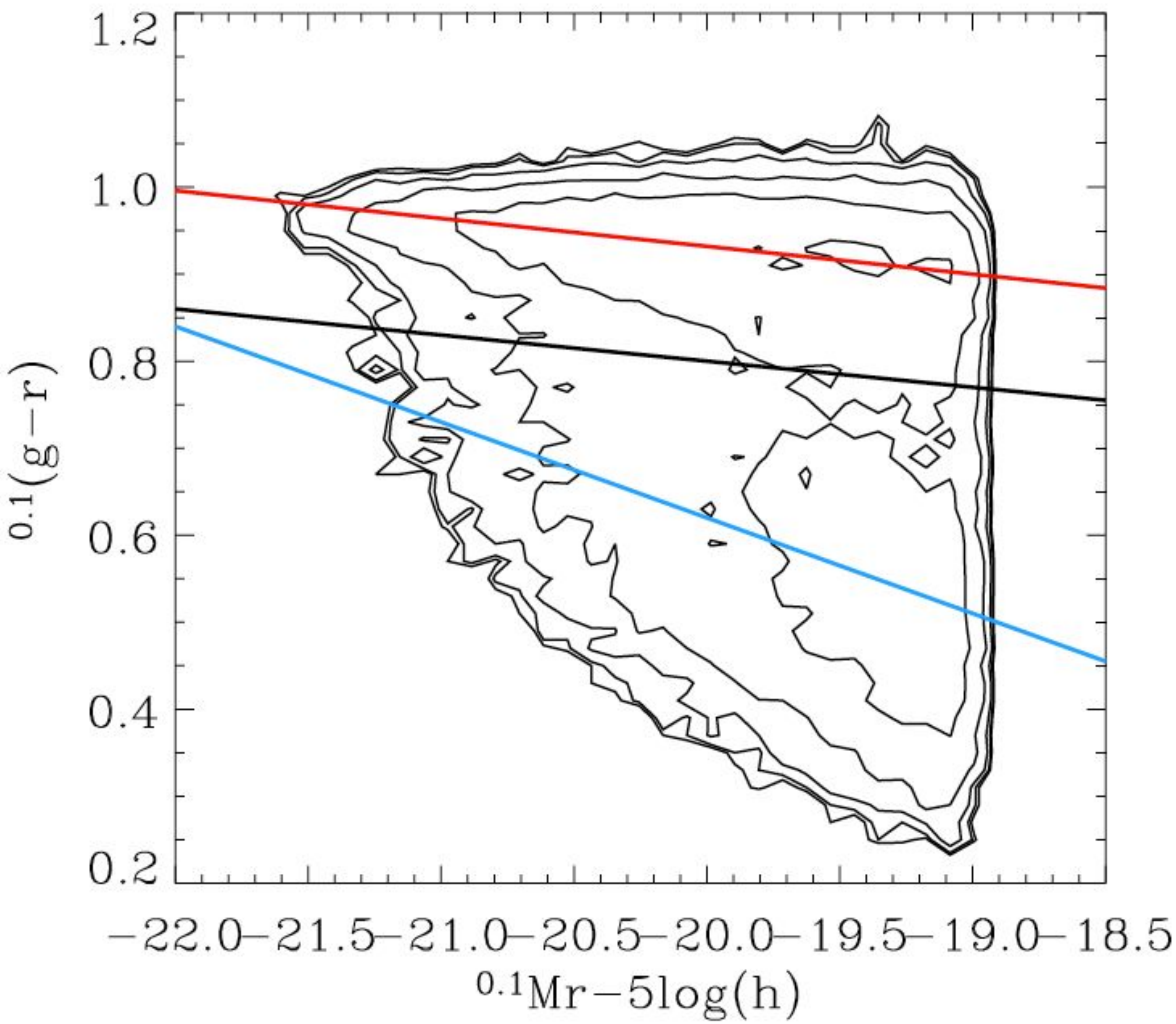
LF Blanton et al. 2003

SDSS Schechter LF

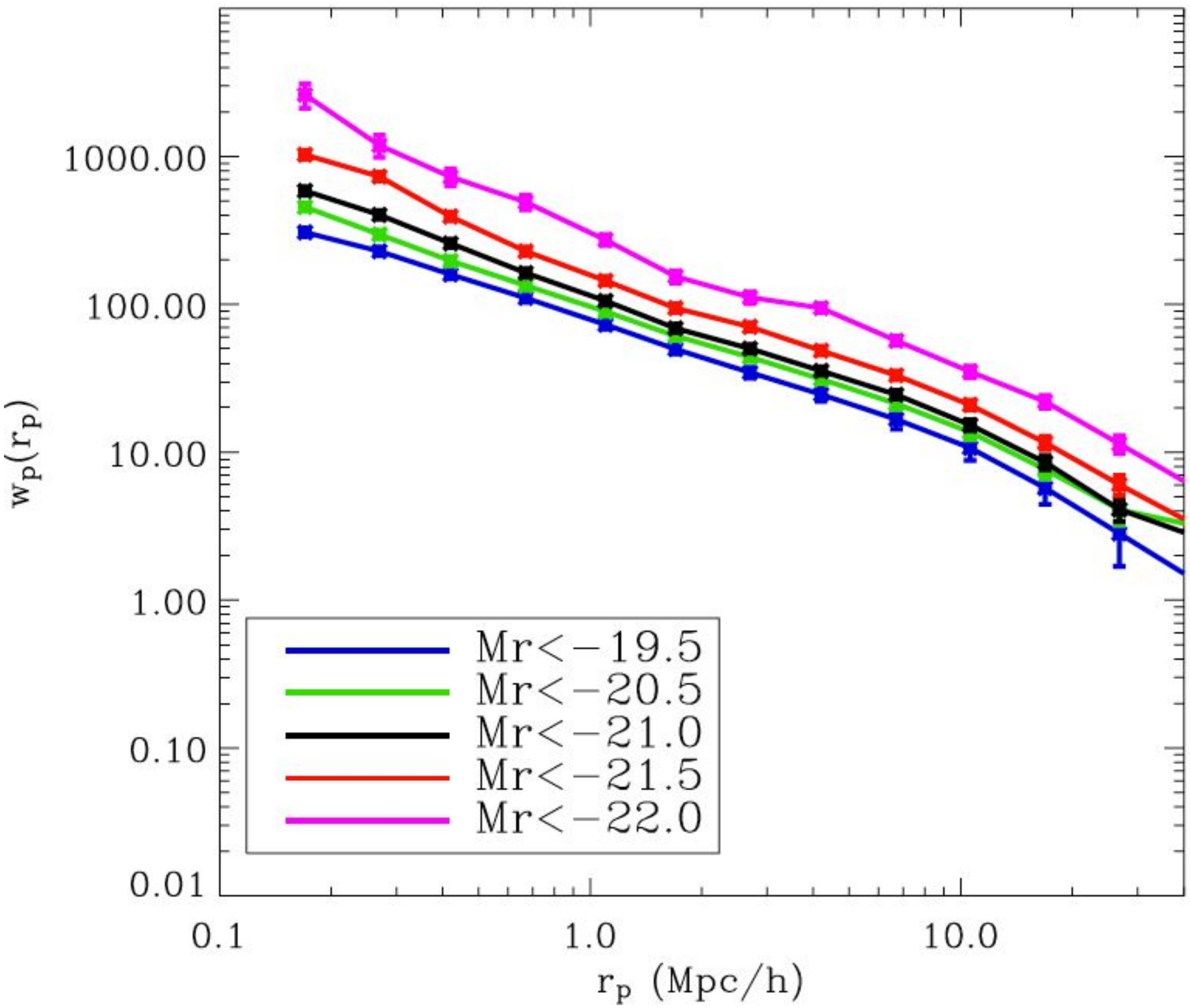


LF Blanton et al. 2005
(faint-end)

SDSS CM diagram

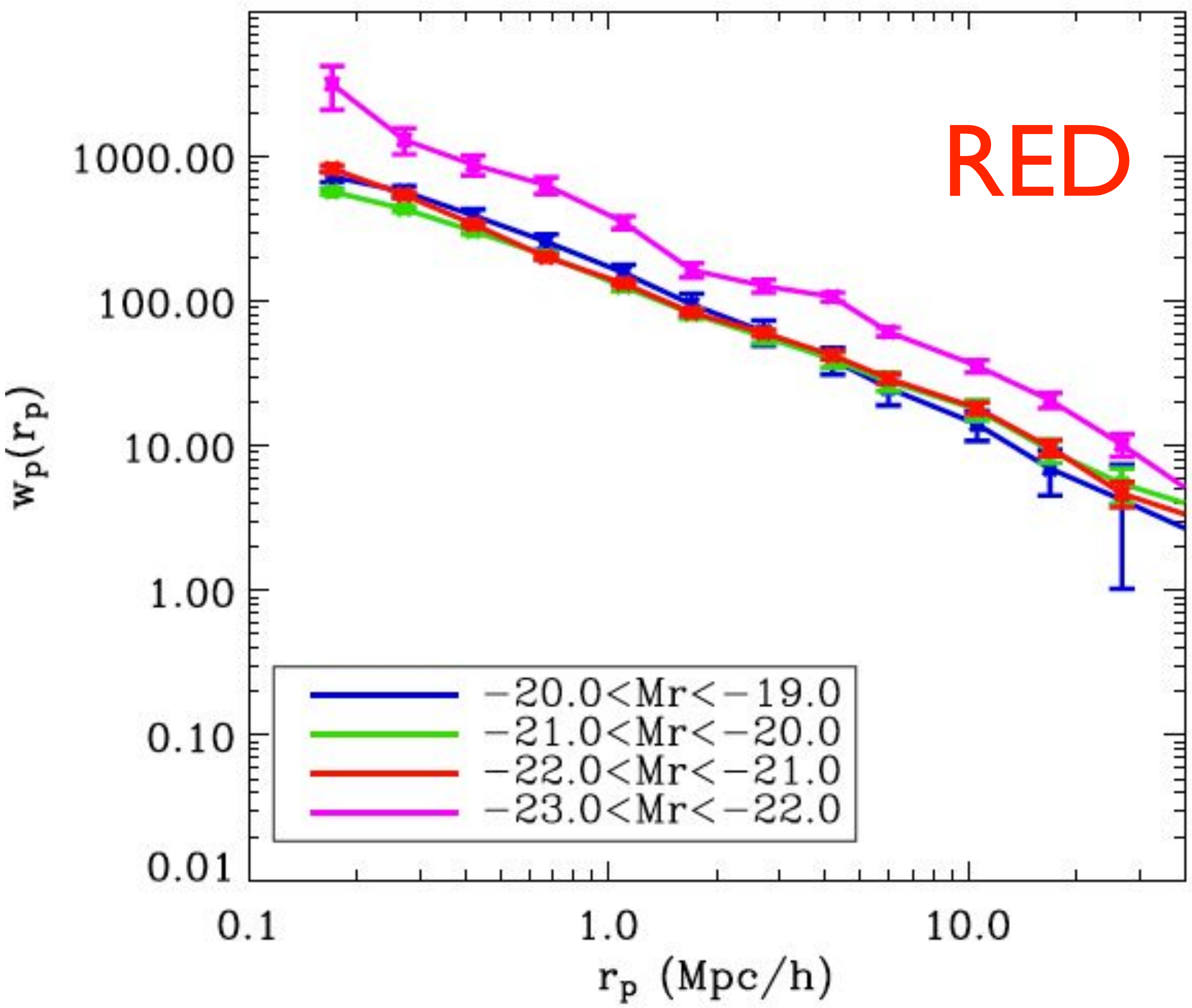


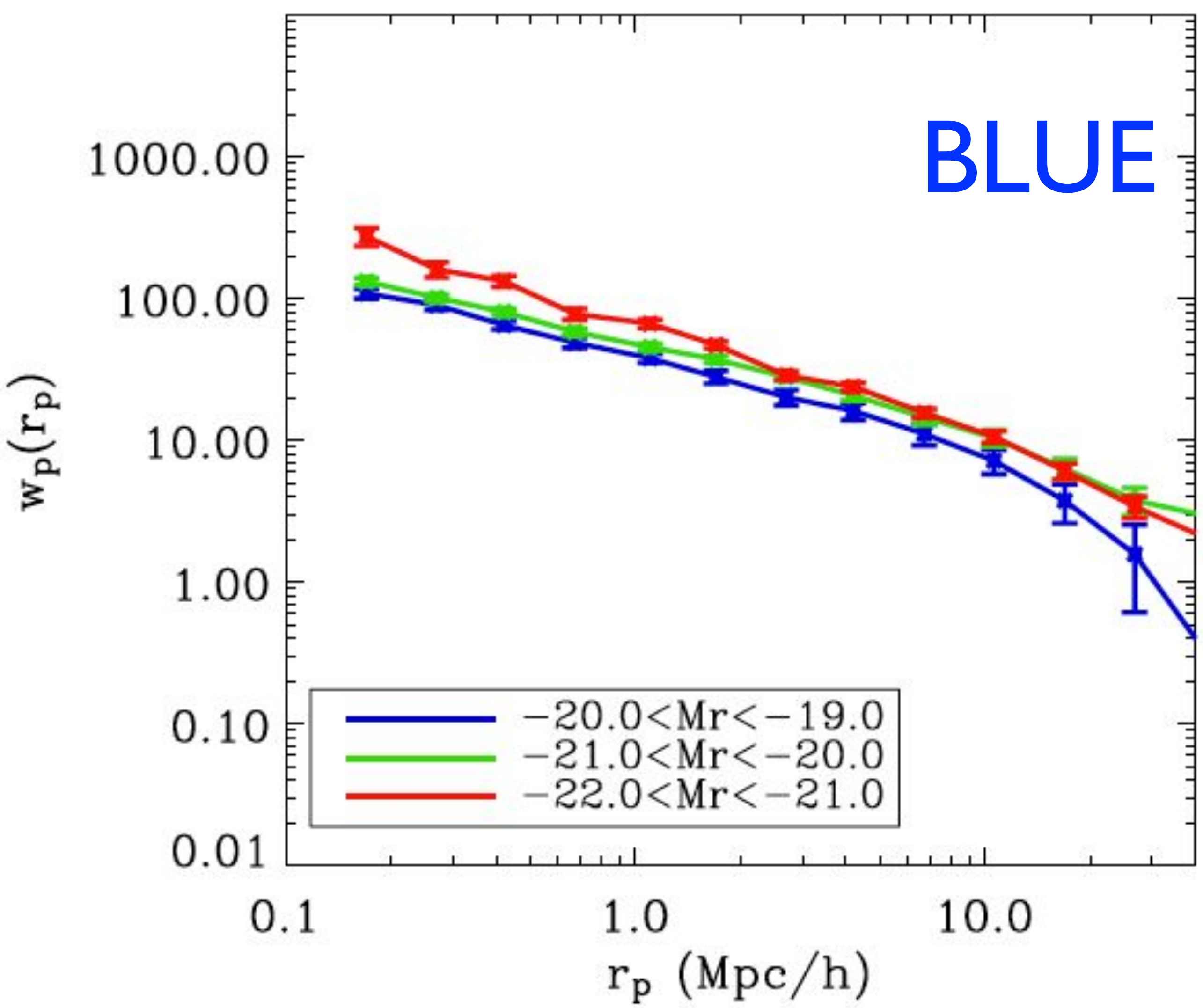
SDSS $w_p(r_p)$ ($<M_r$)



1000.00
100.00
10.00
1.00
0.10
0.01
$w_p(r_p)$
0.1
1.0
10.0
r_p (Mpc/h)
Mr<−19.5
Mr<−20.5
Mr<−21.0
Mr<−21.5
Mr<−22.0

SDSS $w_p(r_p)$ ($<M_r$ | g-r)





SHAM technique

1. Select haloes with $M_h \geq M_{min}$. Each halo will have one central galaxy
2. Satellite galaxies follow a Poisson distribution with mean value $N_{sat}(>L_{min}|M_h)$
3. Assign galaxy luminosities (SHAM technique)
4. Place central galaxies at the center of their host haloes
5. Satellite galaxies follow a modified NFW density profile

SHAM technique

- Simulation + set of HOD parameters:

$$n_{gal}(>M_{min}) = n_{cen}(>M_{min}) + n_{sat}(>M_{min})$$

$$n_{cen}(>M_{min}) = \int_{M_{min}}^{\infty} \frac{dn}{dM_h} dM_h$$

$$n_{sat}(>M_{min}) = \int_{M_{min}}^{\infty} \frac{dn}{dM_h} \left(\frac{M_h}{M_1}\right)^{\alpha} dM_h$$

- Observations (SDSS):

$$n_{gal}(>L_r) = \int_{L_r}^{\infty} \frac{dn}{dL_r} dL_r$$

SHAM technique

$$n_{gal}(> M_{min}) = n_{gal}(> L_r)$$

- We model Mgal - Lgal relation following:

$$\log\left(L_r\right) = c_0 + c_1 \sinh\left[c_2\left(\log\left(M_h - c_3\right)\right)^2 + c_4\left(\log\left(M_h\right) - c_3\right)\right]$$

- We compute the contribution of satellite galaxies to the total galaxy LF introducing Mgal - Lgal:

$$n_{sat}(> M_{min}(L_{gal})) = \int_{M_{min}(>L_{gal})}^{\infty} \frac{dn}{dM_h}\left(\frac{M_h}{M_1}\right)^{\alpha} dM_h$$

Galaxy luminosities

- Central galaxies follow the relation M_{gal} - L_{gal} found
- Since we know N_{sat} and $n_{sat}(>L_r)$, we generate N_{sat} 'available' luminosities sampling randomly the distribution
- We assign one of the available and sampled luminosities for each satellite galaxy if $L_{sat} \leq 1.05\ L_{cen}$ (~central galaxy is the most luminous galaxy in a halo)
- SHAM technique guarantees that each mock galaxy catalogue follows the observed LF

Galaxy positions

- Central galaxies are placed at the center of their host haloes
- Satellite galaxies follow a NFW density profile

$$\frac{\rho(r)}{\bar{\rho}} = \frac{\Delta_{vir}(z)}{3\Omega(z)} \frac{c^3 f(c)}{x(1+x)^2}$$

$$\Delta_{vir}(z) = 18\pi^2 + 82d - 39d^2; \quad d = \Omega(z) - 1.$$

- Concentration parameter provided by Bullock et al. (2001)

$$c(M_h) \approx \frac{9}{(1+z)} \left(\frac{M_h}{M_{*0}}\right)^{-0.13} \qquad r_{vir} = \left(\frac{3M}{4\pi\Delta_{vir}(z)\rho_{crit.}}\right)^{(1/3)}$$

Galaxy positions

In the NFW profile, the density of dark matter as a function of radius is given by:

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s}\left(1 + \frac{r}{R_s}\right)^2}$$

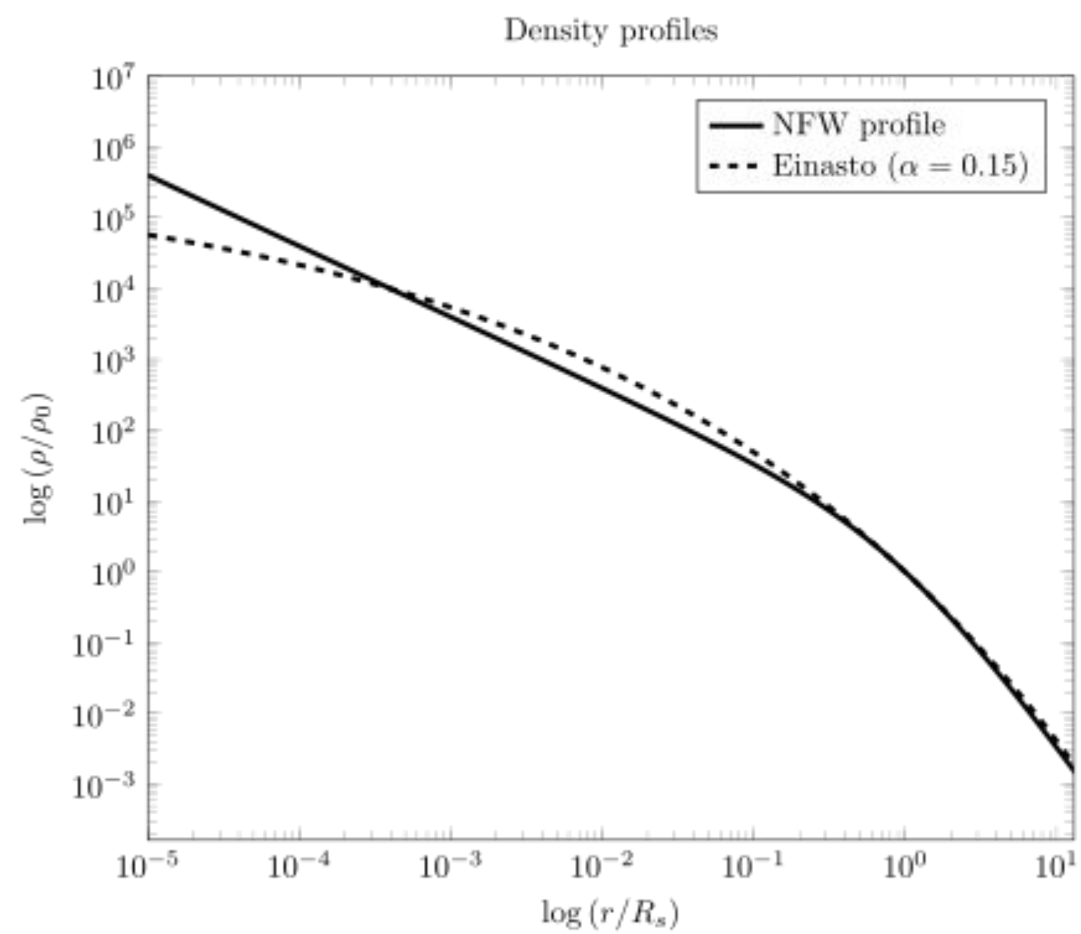
where ρ_0 and the "scale radius", R_s, are parameters which vary from halo to halo.

The integrated mass within some radius R_{max} is

$$M = \int_0^{R_{\max}} 4\pi r^2 \rho(r)\, dr = 4\pi\rho_0 R_s^3 \left[\ln\left(\frac{R_s + R_{\max}}{R_s}\right) - \frac{R_{\max}}{R_s + R_{\max}}\right]$$

The total mass is divergent, but it is often useful to take the edge of the halo to be the virial radius, R_{vir}, which is related to the "concentration parameter", c, and scale radius via

$$R_{\mathrm{vir}} = cR_s$$



Galaxy velocities

- Central galaxies: same velocity as its host halo
- Satellite galaxies: following Seth & Diaferio (2001), the motion of a particle in an N-body simulation can be described as the sum of the virial motion of the particle within its host halo and the bulk motion of the halo as a whole

$$v = v_{vir} + v_h \qquad \sigma_{vir}^2 =< v_{vir}^2 > \propto GM_h/r_{vir} \propto M_h^{2/3}$$

$$\sigma_{vir} = 476 f_{vir} [\Delta_{vir} E^2(z)]^{1/6} \left(\frac{M_h}{10^{15} M_\odot h^{-1}} \right)^{1/3} kms^{-1}$$

Bryan & Norman (1998)

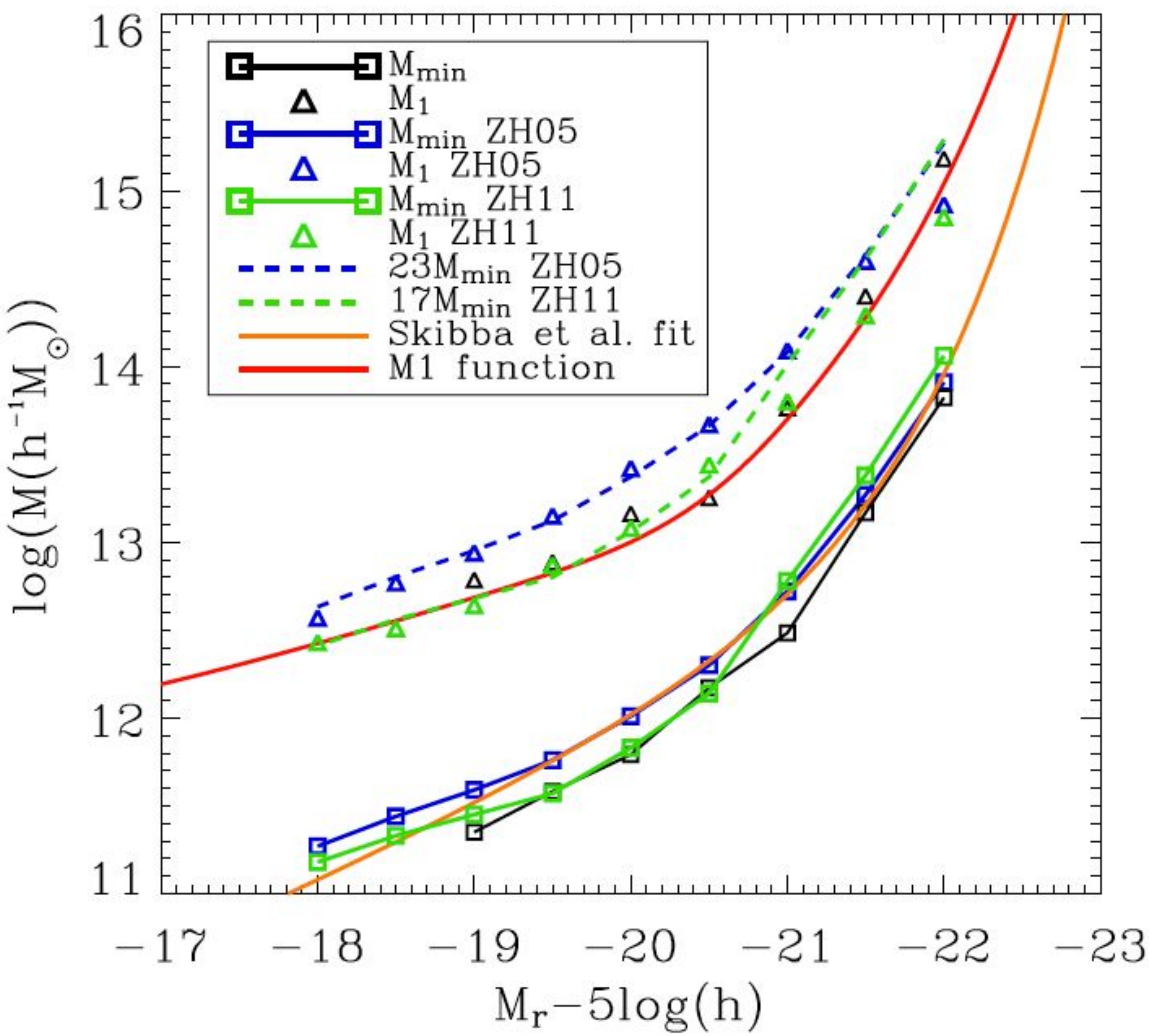
Fine tuning

- M1 is a function of Mh:

$$f_{M_1} = 0.5((a_1 - a_2)\tanh(s_1(b_1 - \log M_h)) + (a_3 - a_2)\tanh(s_2(\log M_h - b_2)) + (a_1 + a_3))$$

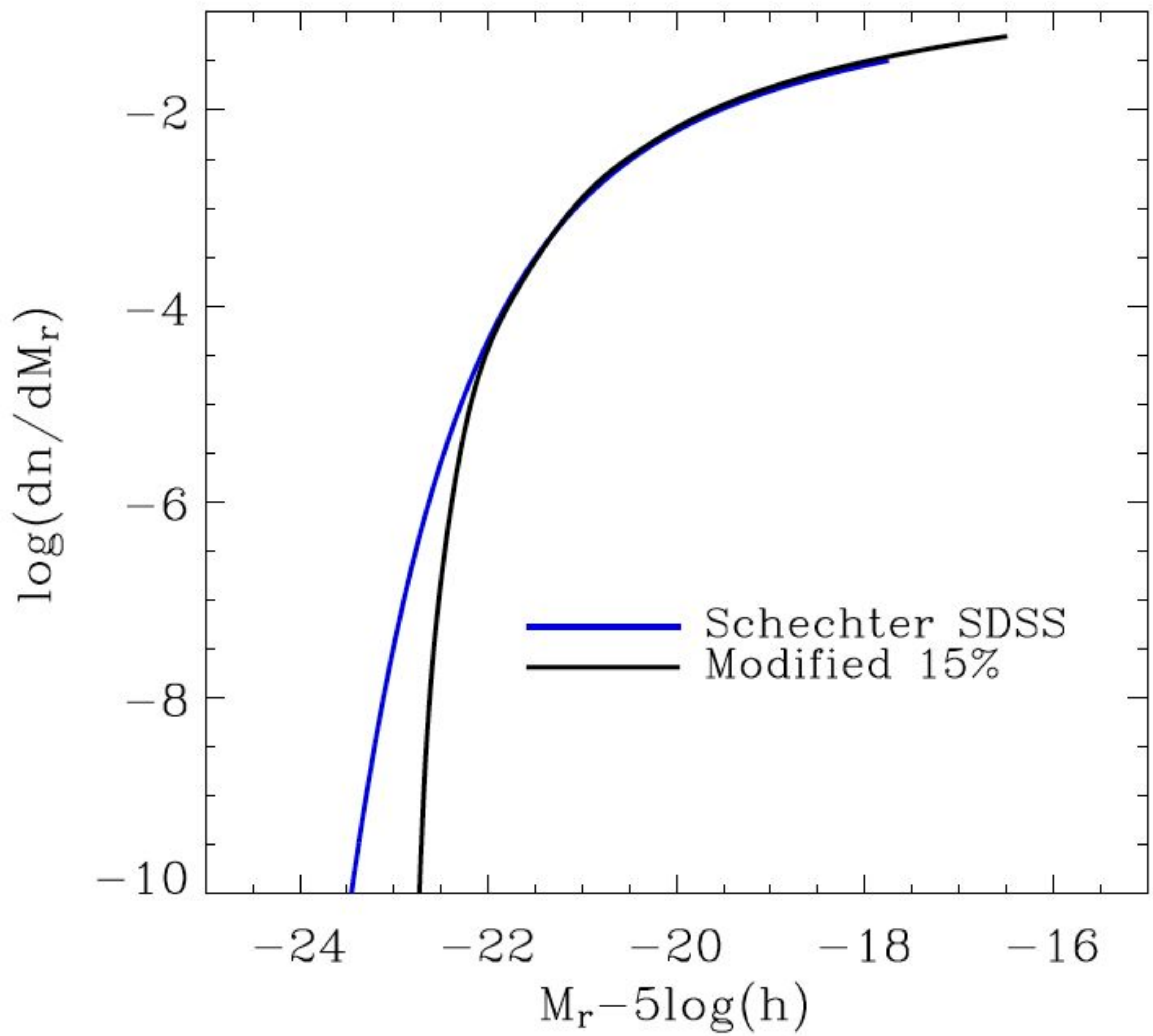
- Scatter in the relation between halo mass and central luminosity relation in order to decrease the clustering of the brightest galaxies to better reproduce their clustering at large scales
- Third Gaussian component to better fit the colour-magnitude diagram

$M_1(M_h(M_r))$

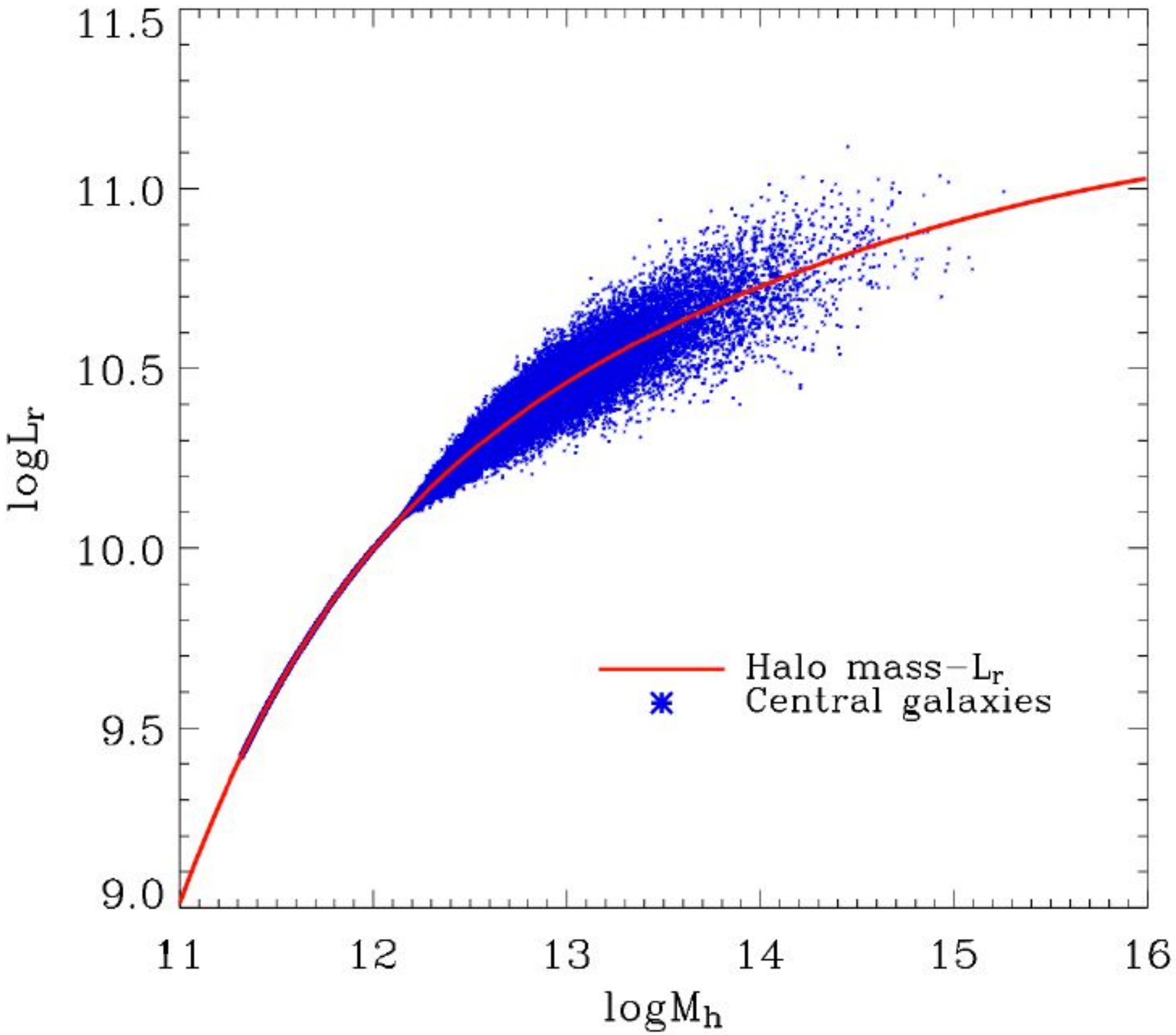


M_{min}
M_1
M_{min} ZH05
M_1 ZH05
M_{min} ZH11
M_1 ZH11
$23M_{min}$ ZH05
$17M_{min}$ ZH11
Skibba et al. fit
M1 function
$\log(M(h^{-1}M_\odot))$
$M_r-5\log(h)$
11
12
13
14
15
16
−17
−18
−19
−20
−21
−22
−23

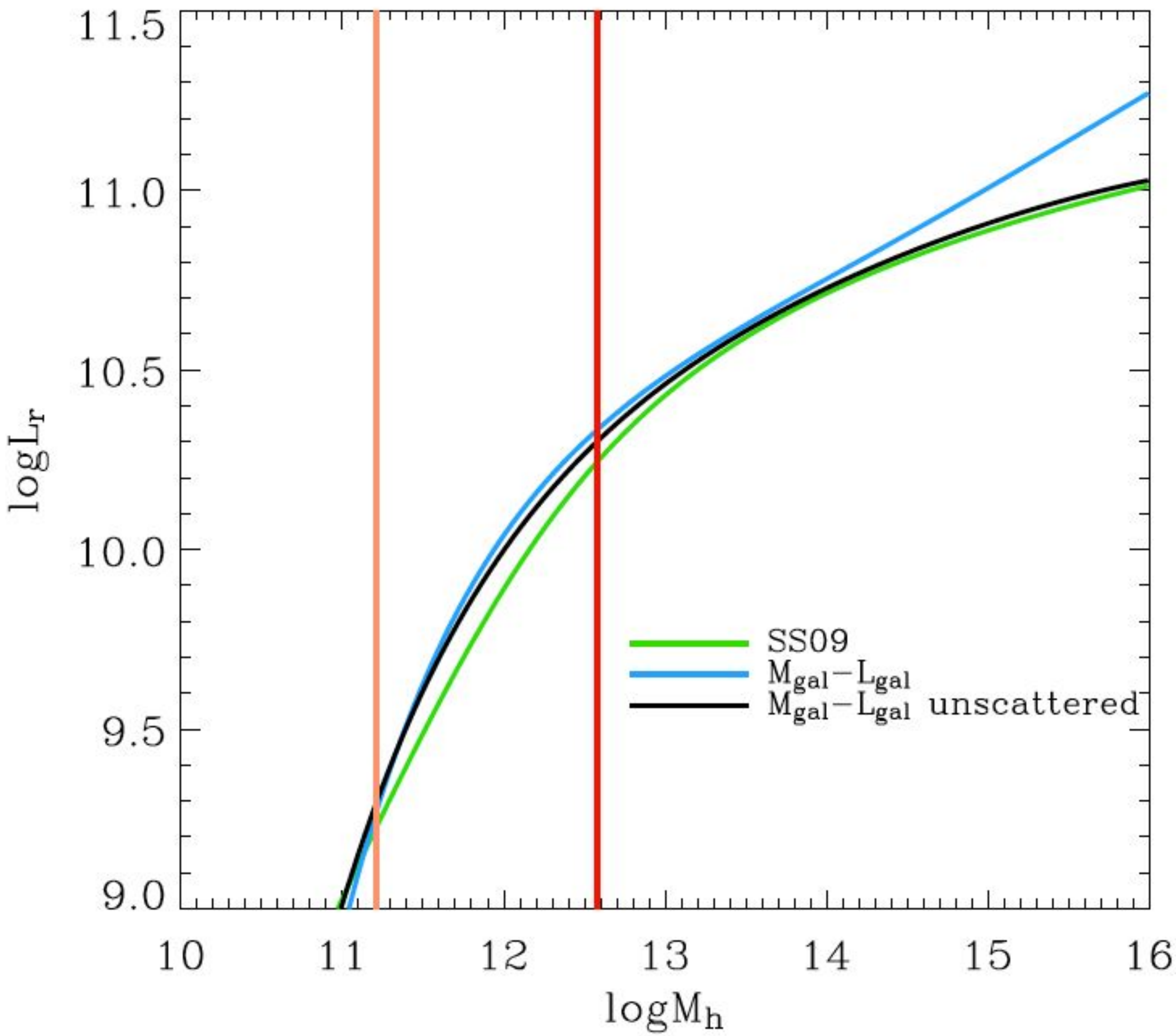
Scatter M_h - M_r



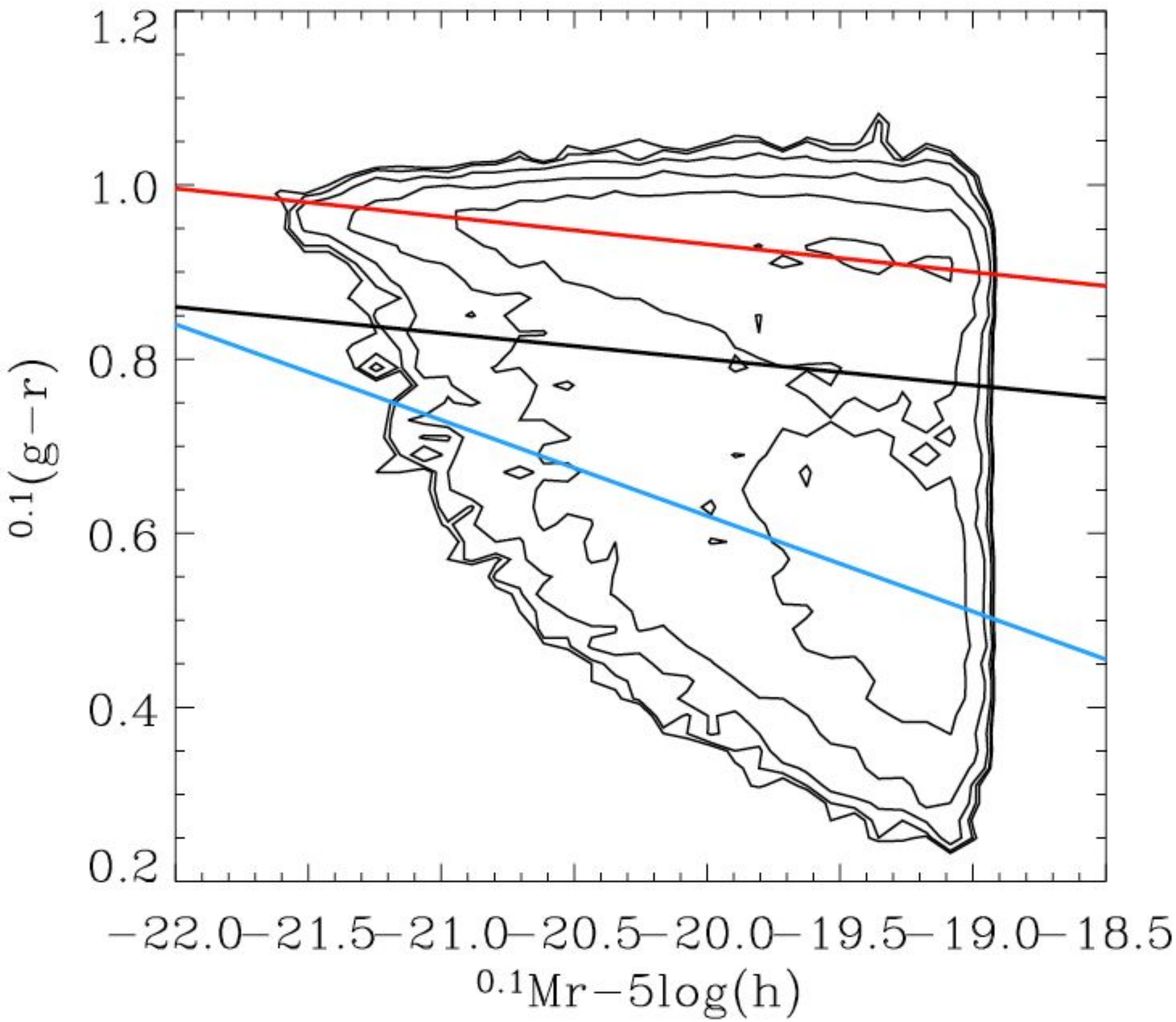
Scatter M_h - M_r



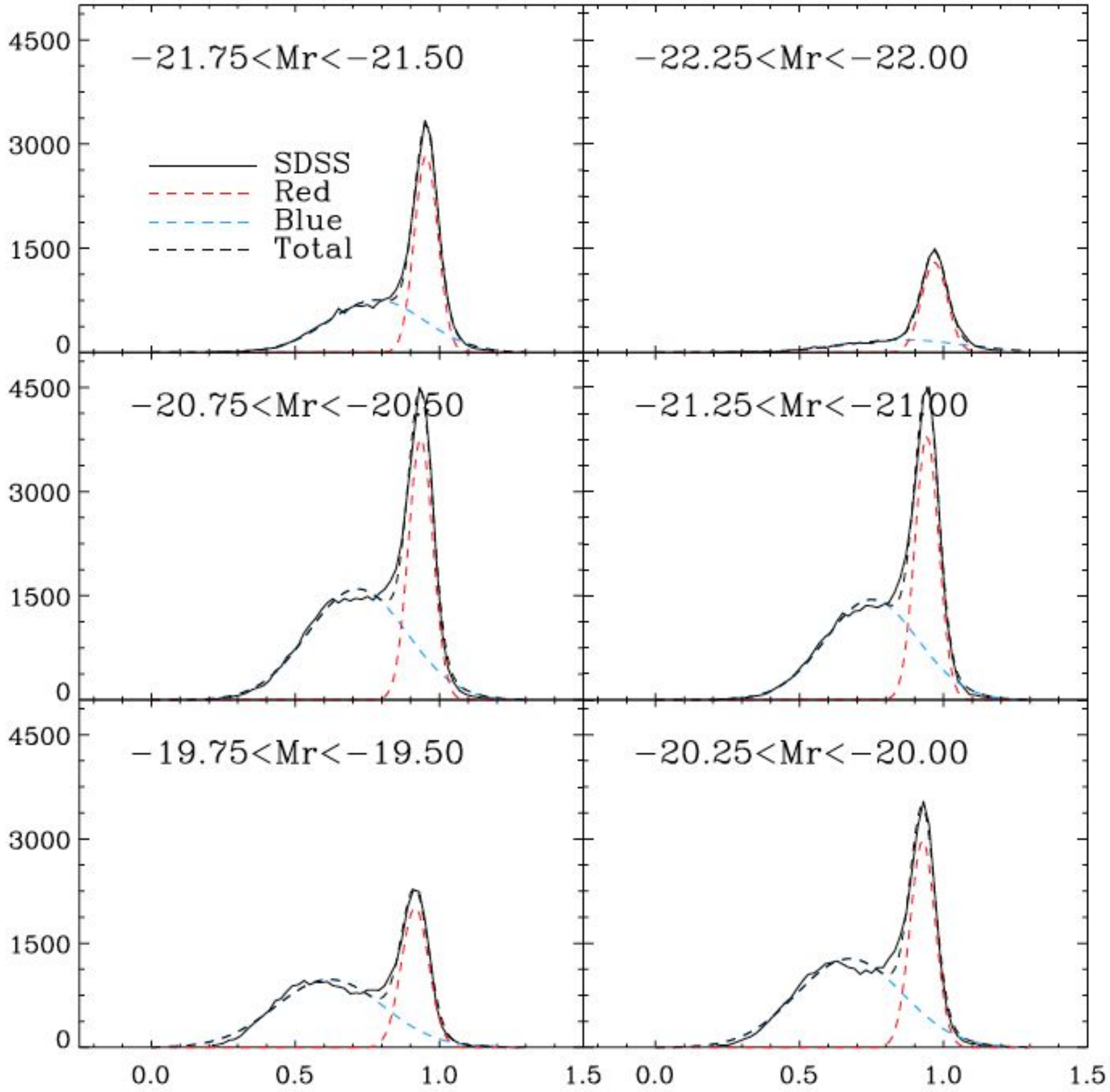
Scatter M_h - M_r



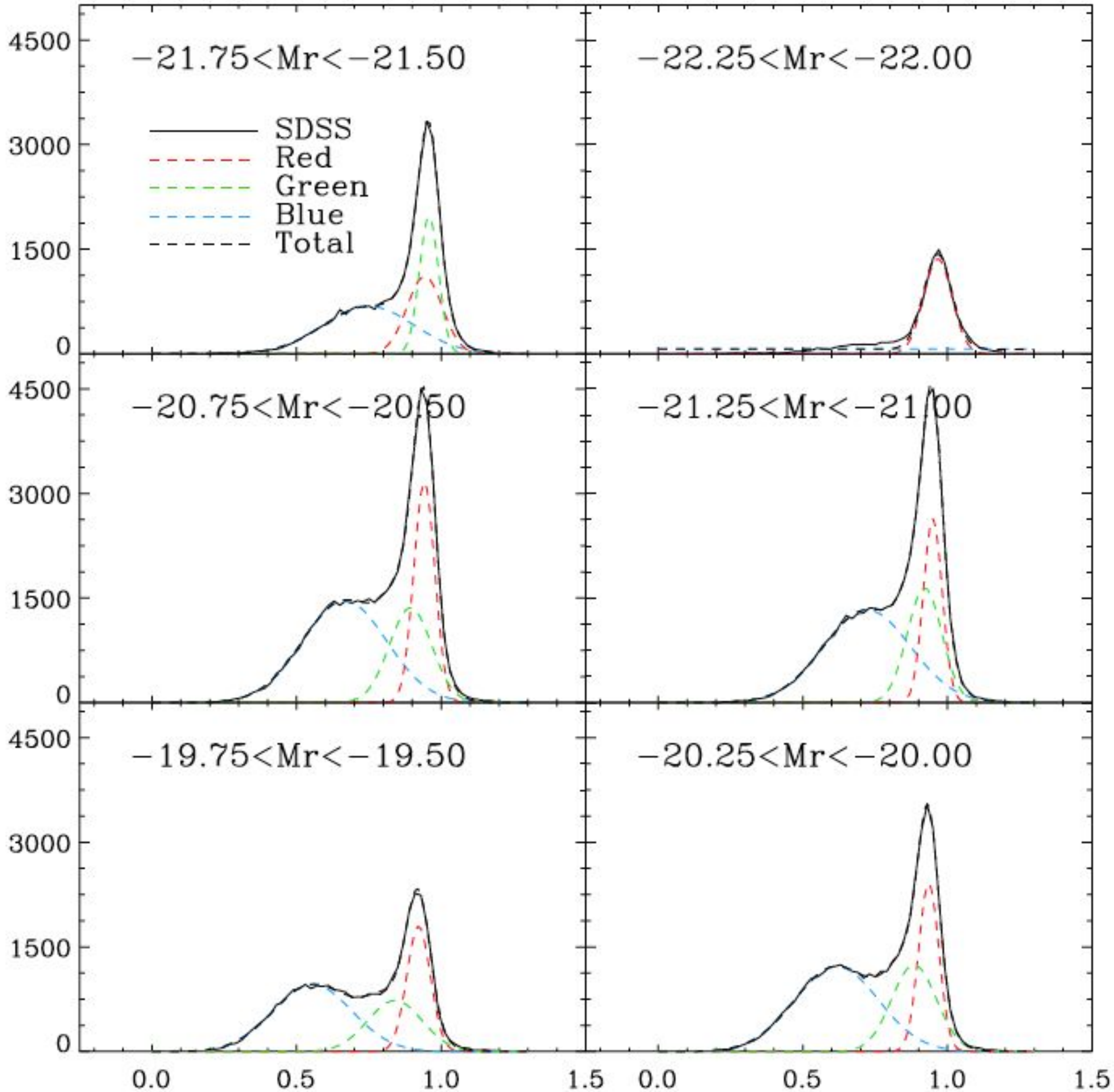
SDSS CM diagram



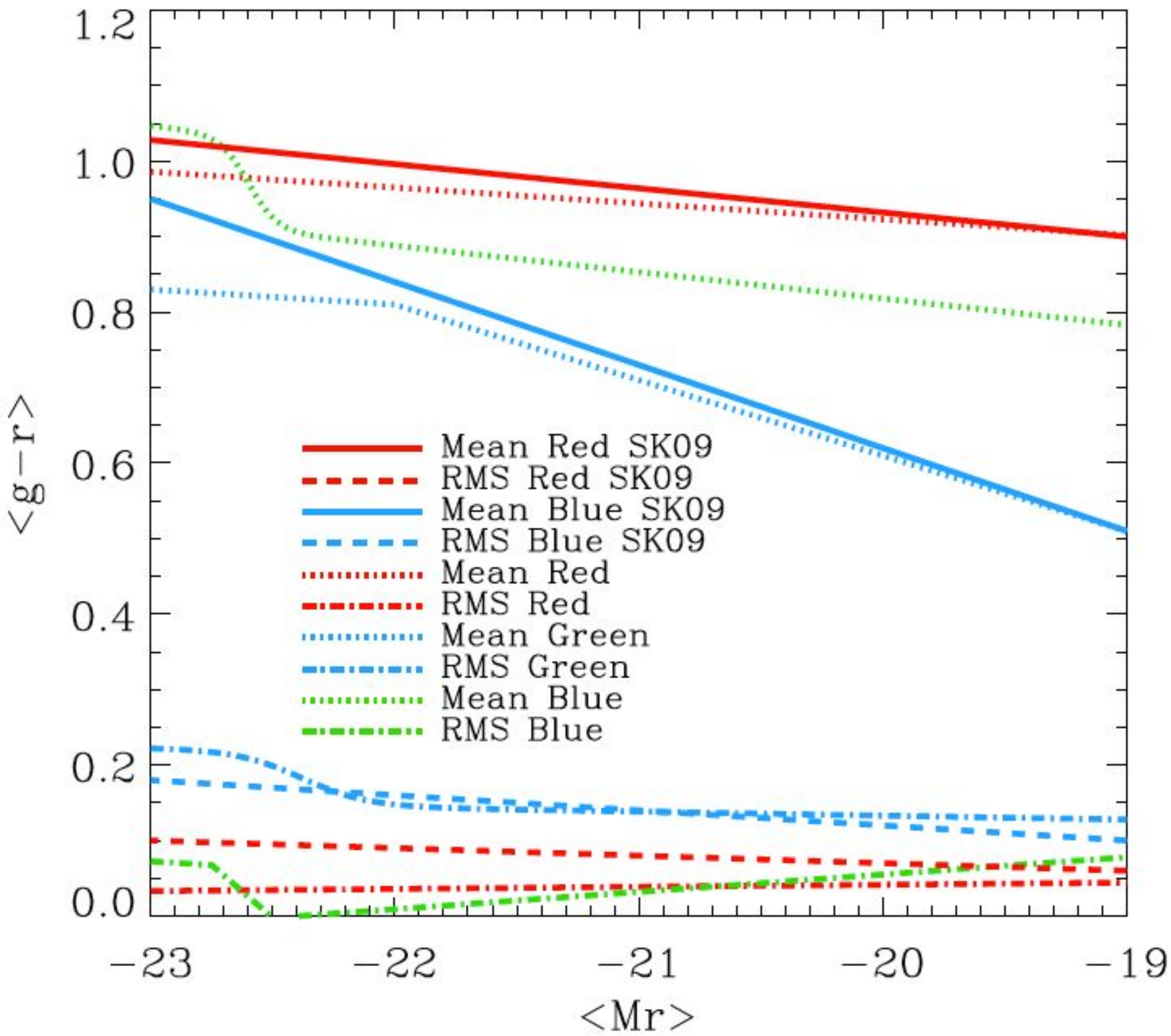
Bimodality (two Gaussians)



Bimodality (three Gaussians)



Bimodality



1.2
1.0
0.8
0.6
0.4
0.2
0.0
<g−r>
−23
−22
−21
−20
−19
<Mr>
Mean Red SK09
RMS Red SK09
Mean Blue SK09
RMS Blue SK09
Mean Red
RMS Red
Mean Green
RMS Green
Mean Blue
RMS Blue

Galaxy colors

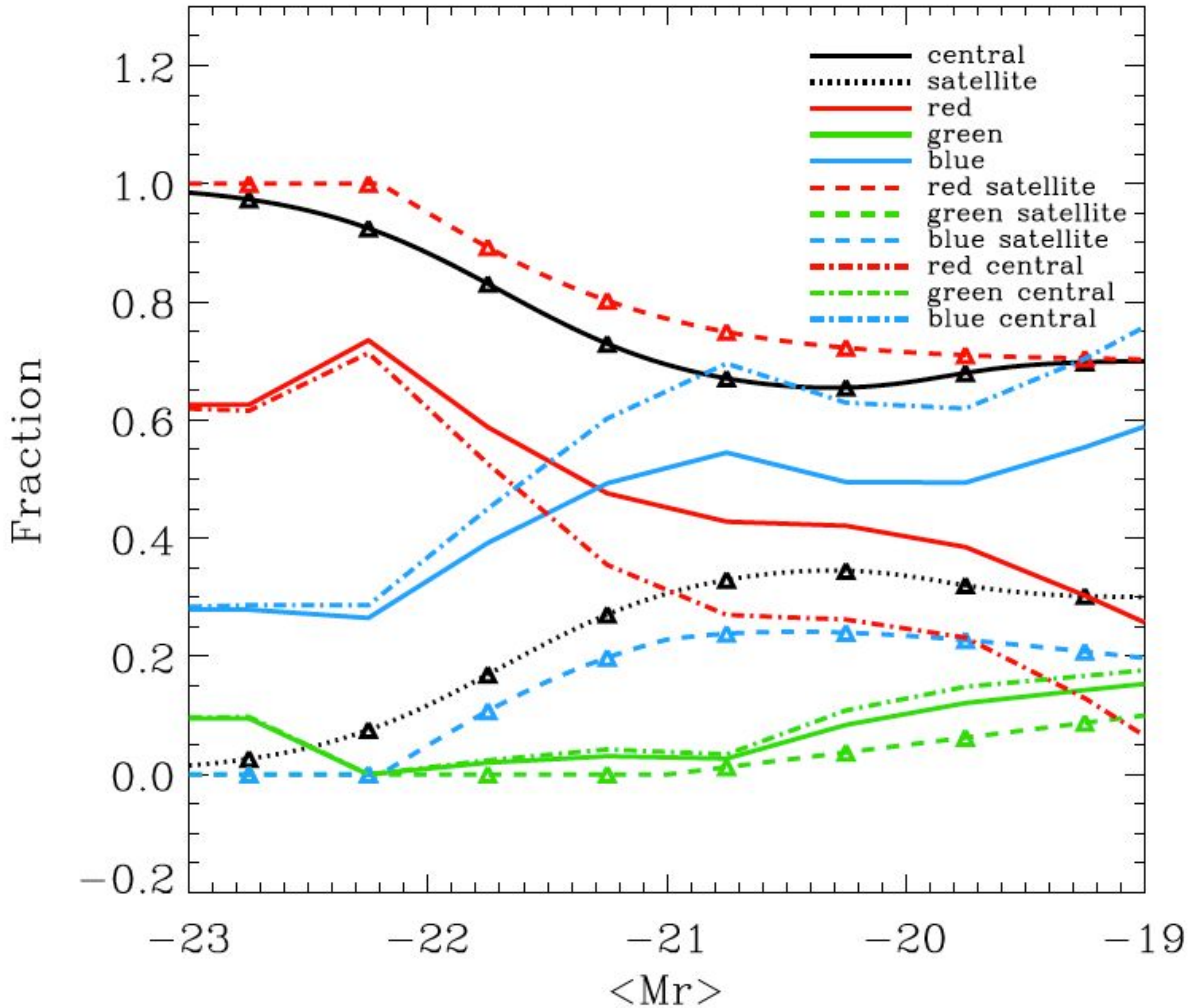
- Skkibba & Sheth 2009 proposed a method to include galaxy colours (main assumption: galaxy colours do not depend explicitly on the halo mass)
- They proposed a function that defines the mean colour of satellite galaxies given Mr. This function is used to derive the fraction of satellite galaxies that belongs to the red sequence
- We directly set the fraction of satellite galaxies that belongs to the red and green sequences:

$$f_{sat}^{red}(M_r) = 1.00 - 0.30 \tanh\left(\frac{M_r + 22.20}{1.20}\right)$$

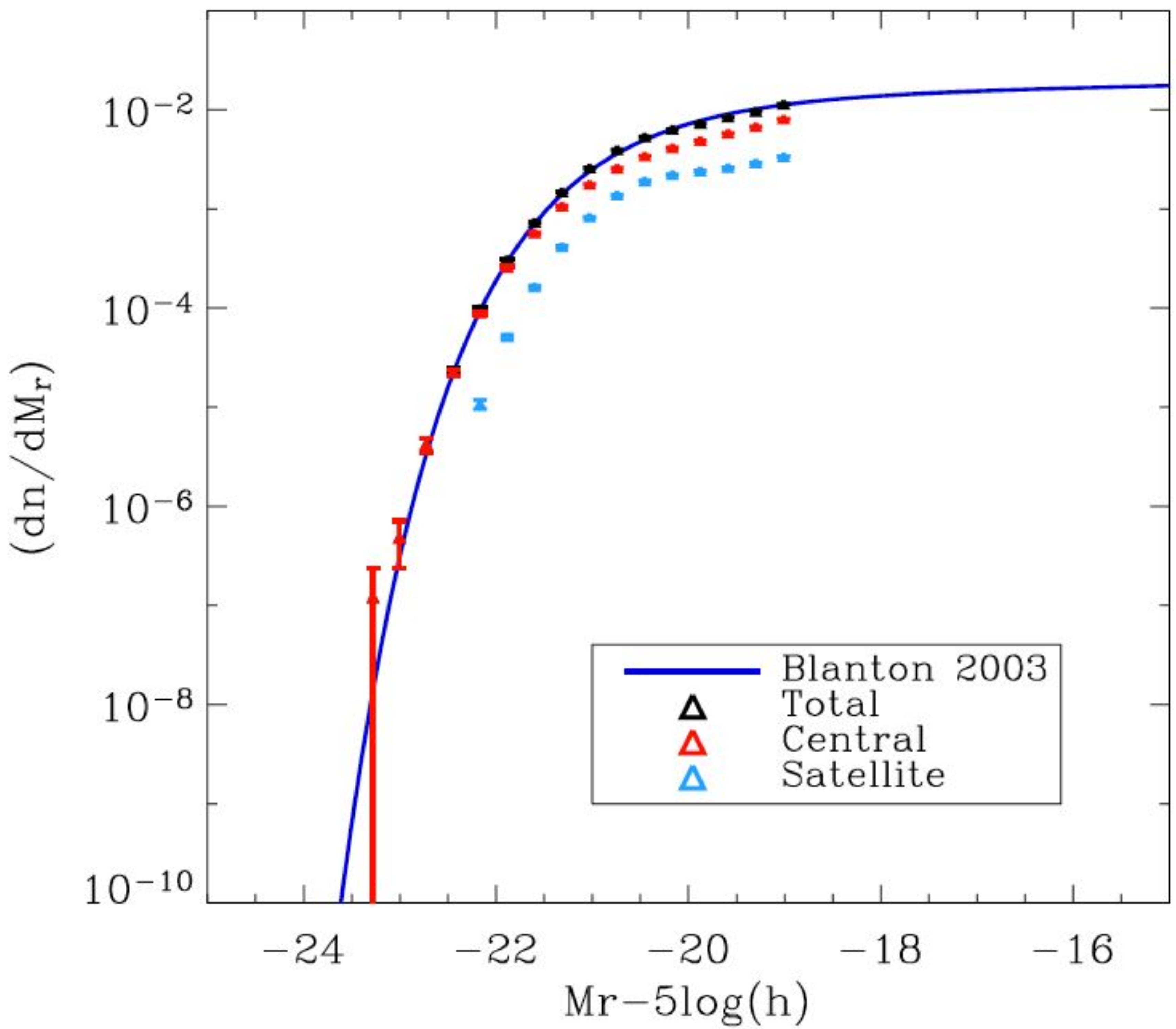
$$f_{sat}^{green}(M_r) = 0.05 - 0.05\,(M_r + 20.0)$$

$$f_{sat}^{blue}(M_r) = 1 - \left(f_{sat}^{red}(M_r) + f_{sat}^{green}(M_r)\right)$$

Galaxy colors

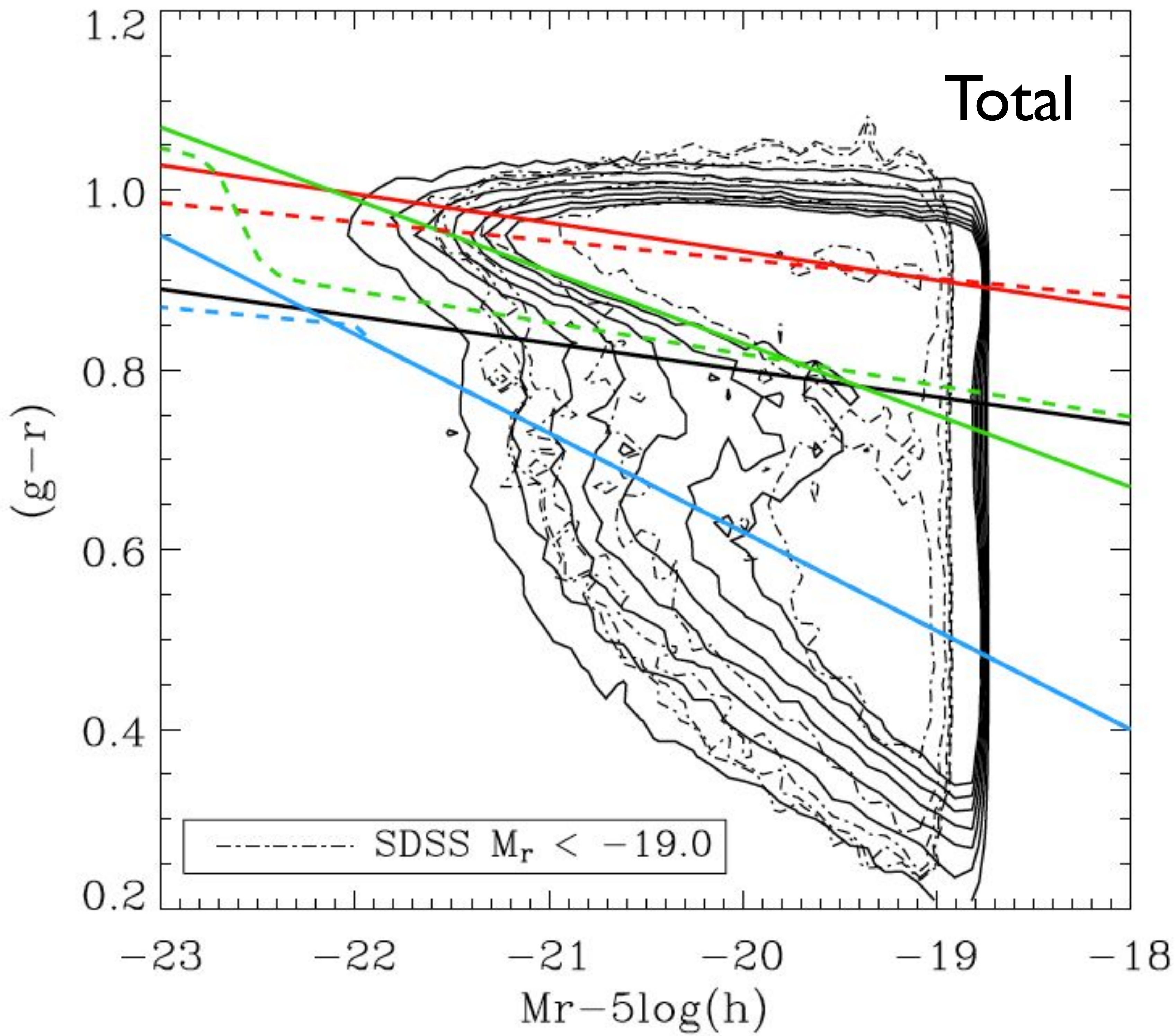


Results | LF

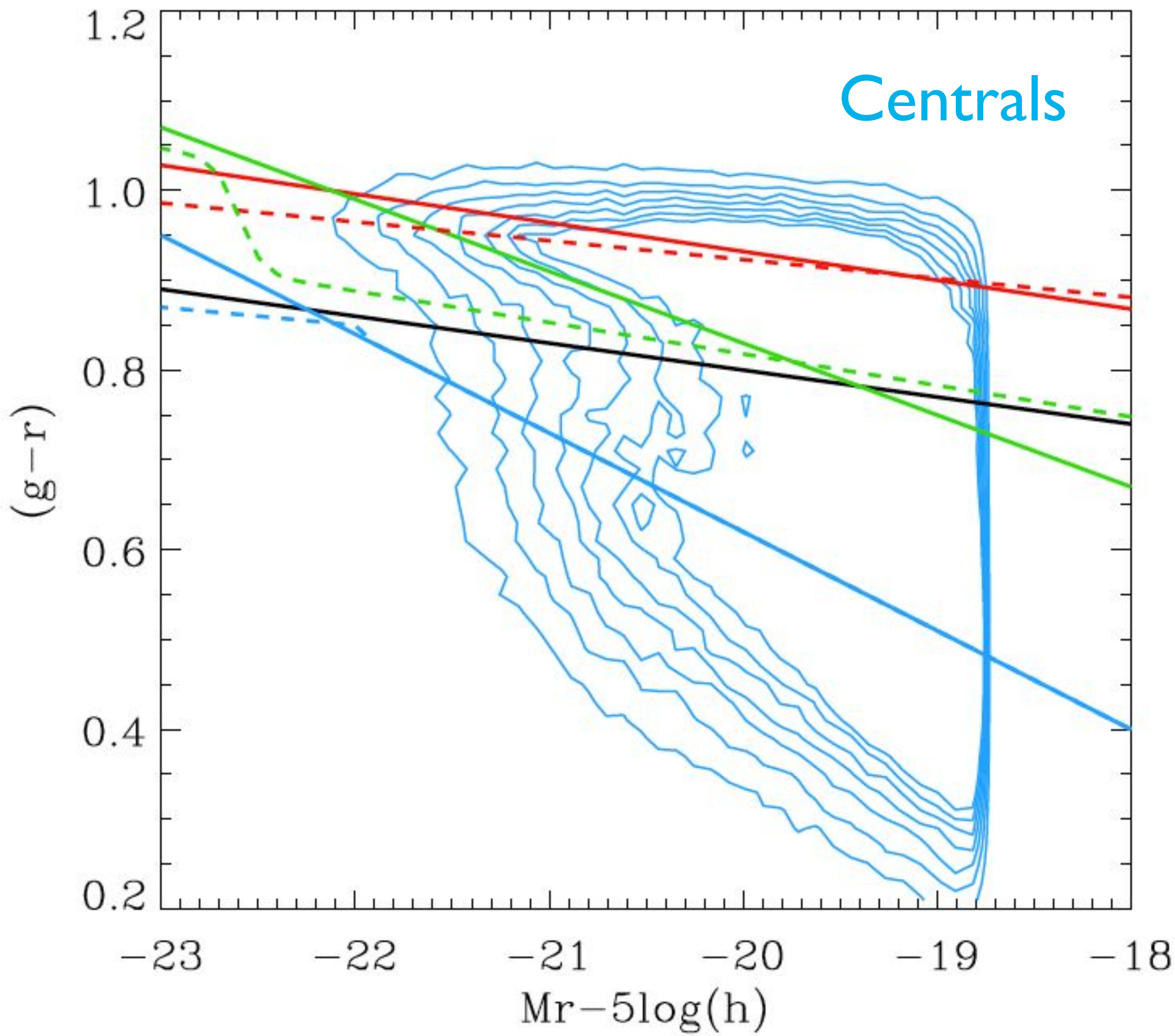


10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
(dn/dM_r)
−24
−22
−20
−18
−16
Mr−5log(h)
Blanton 2003
Total
Central
Satellite

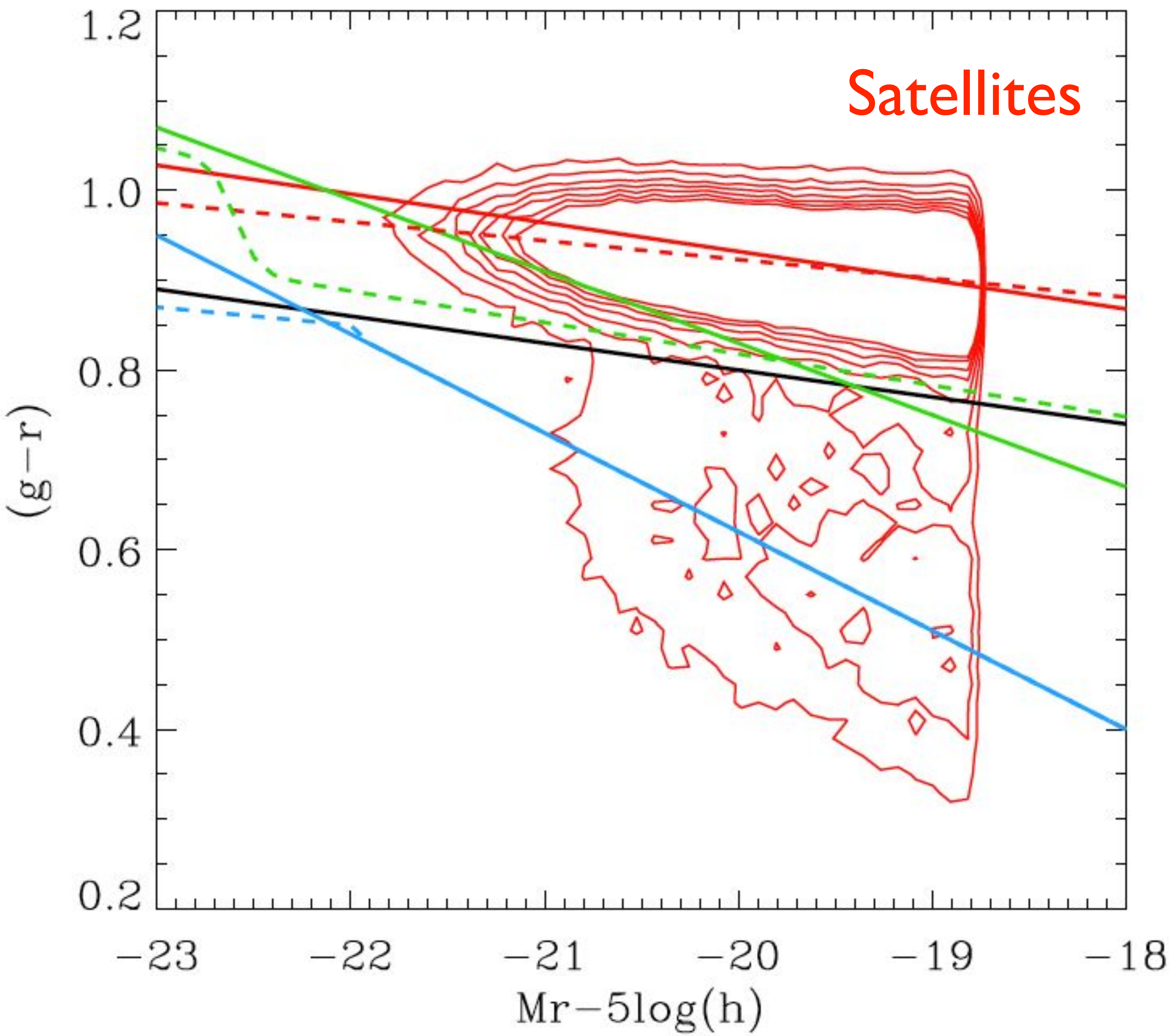
Results: CM-diagram



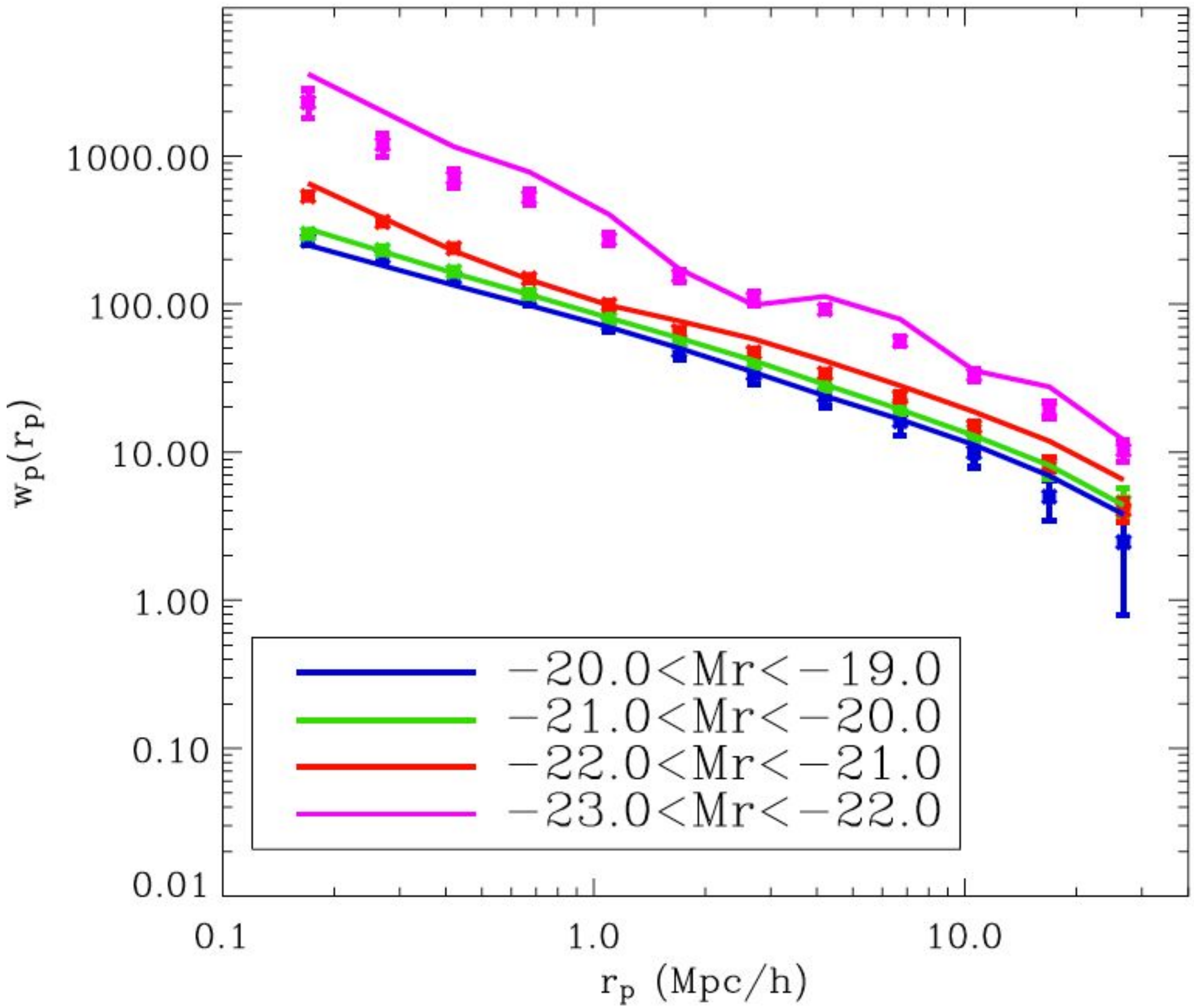
Results: CM-diagram



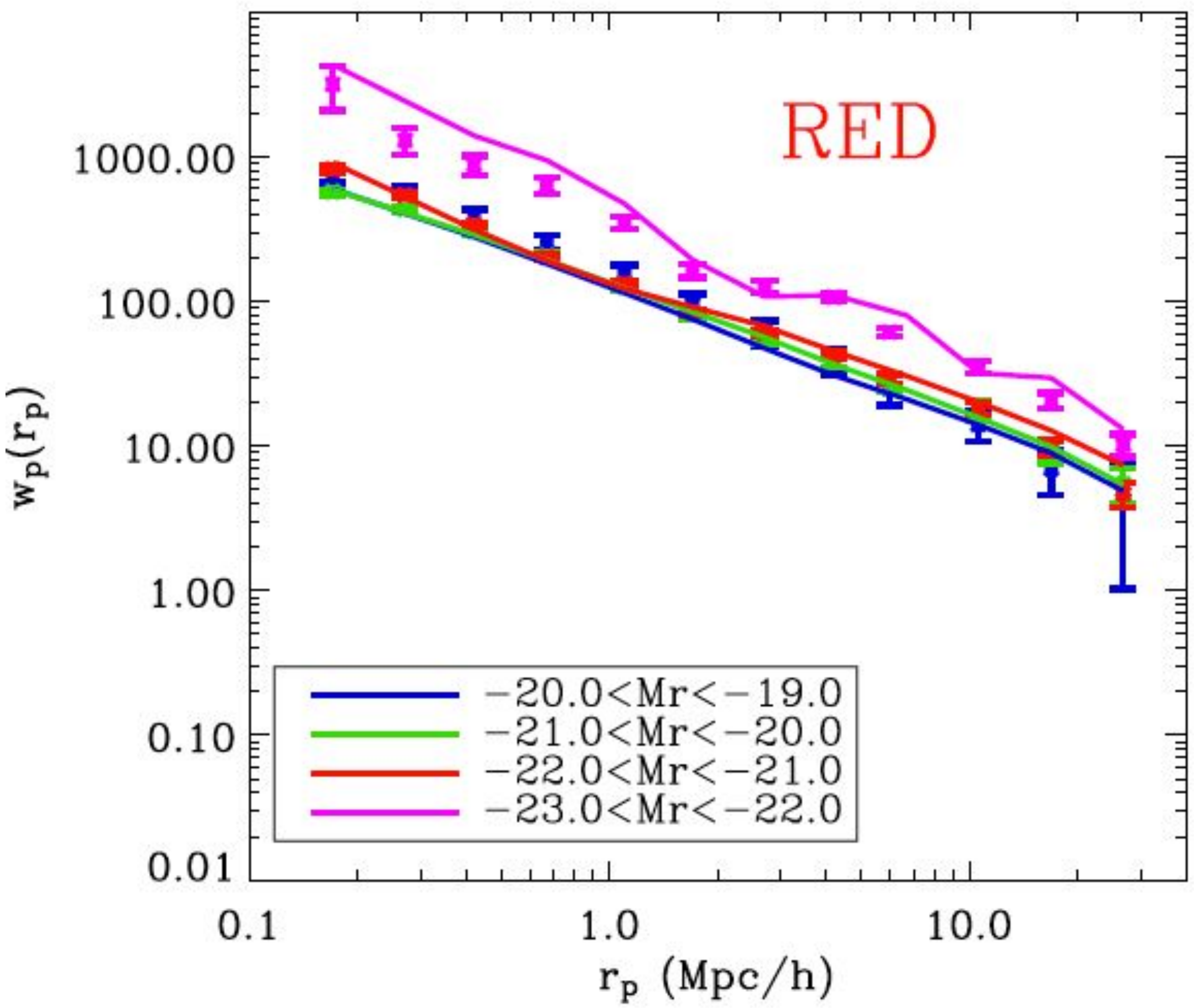
Results: CM-diagram

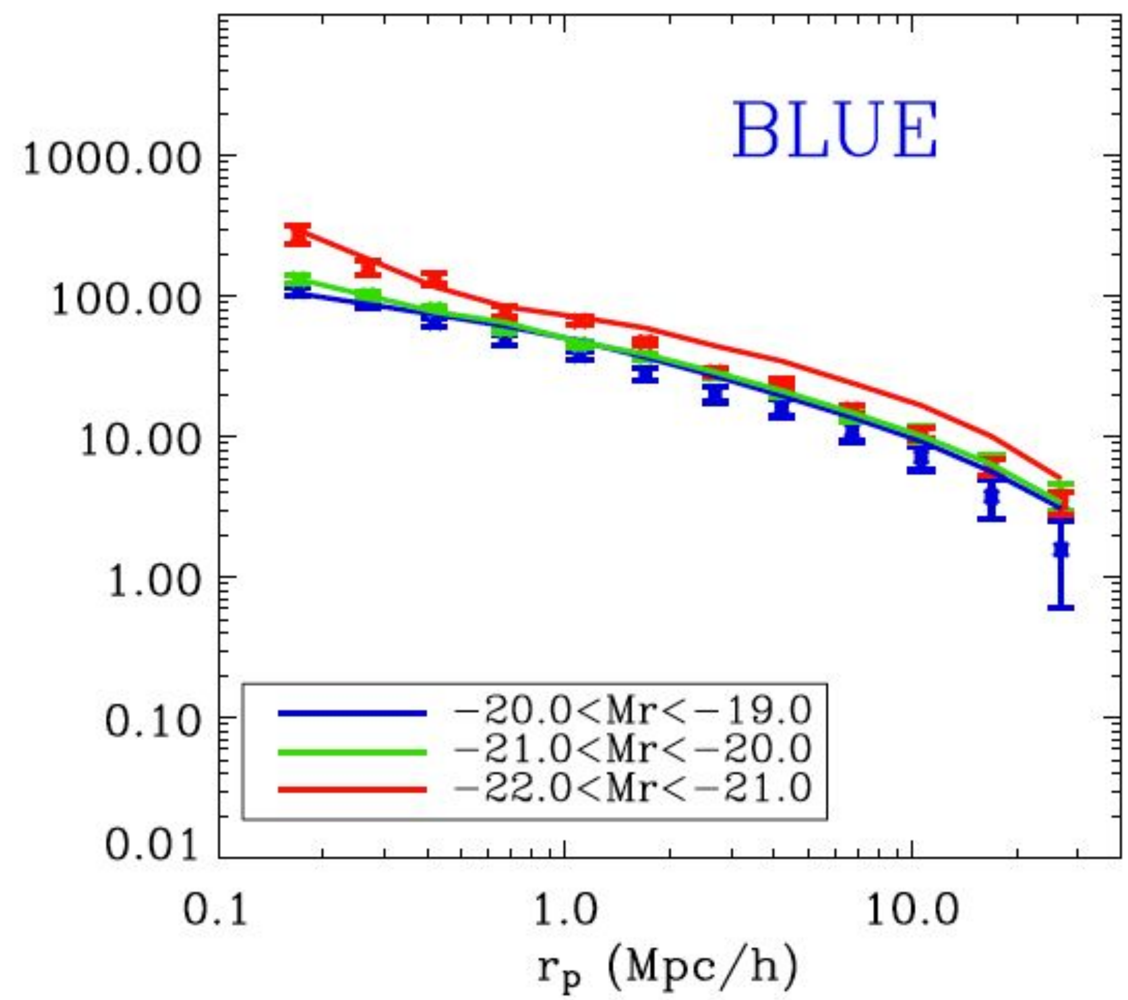


Results: $w_p(r_p)(M_{r1}<M_r<M_{r2})$

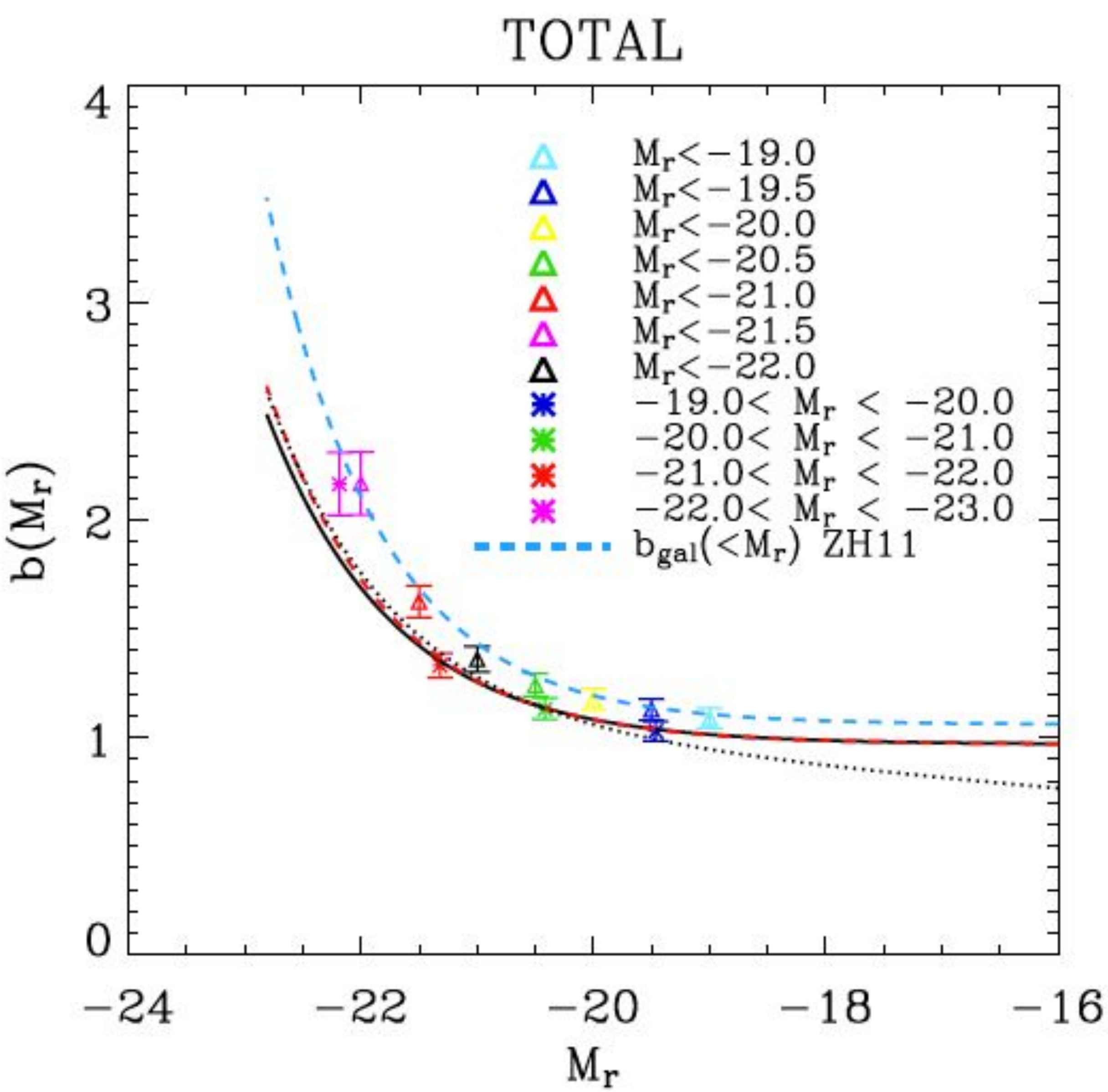


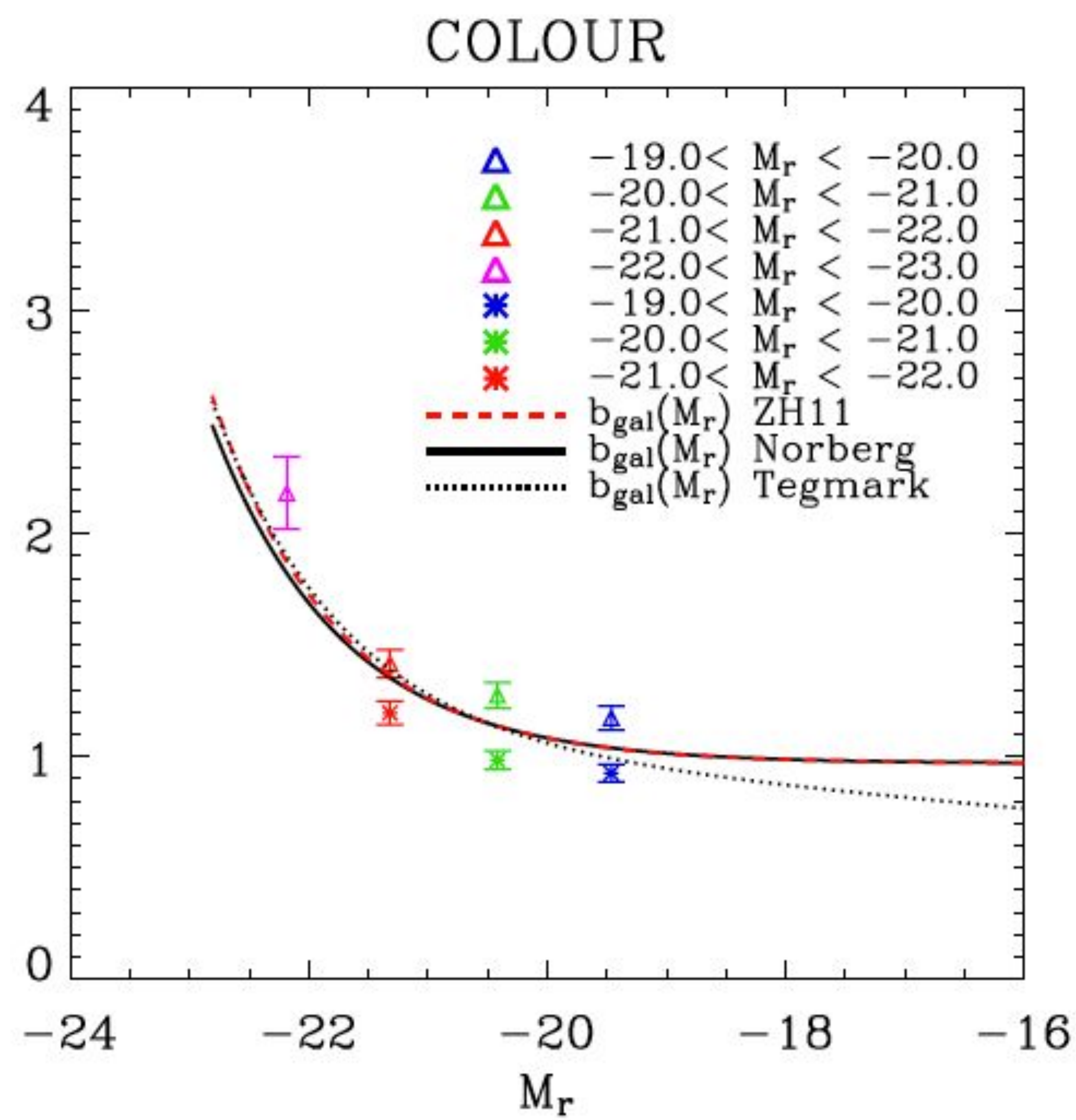
$w_p(r_p)$ $(M_{r1}<M_r<M_{r2})$ (g-r)



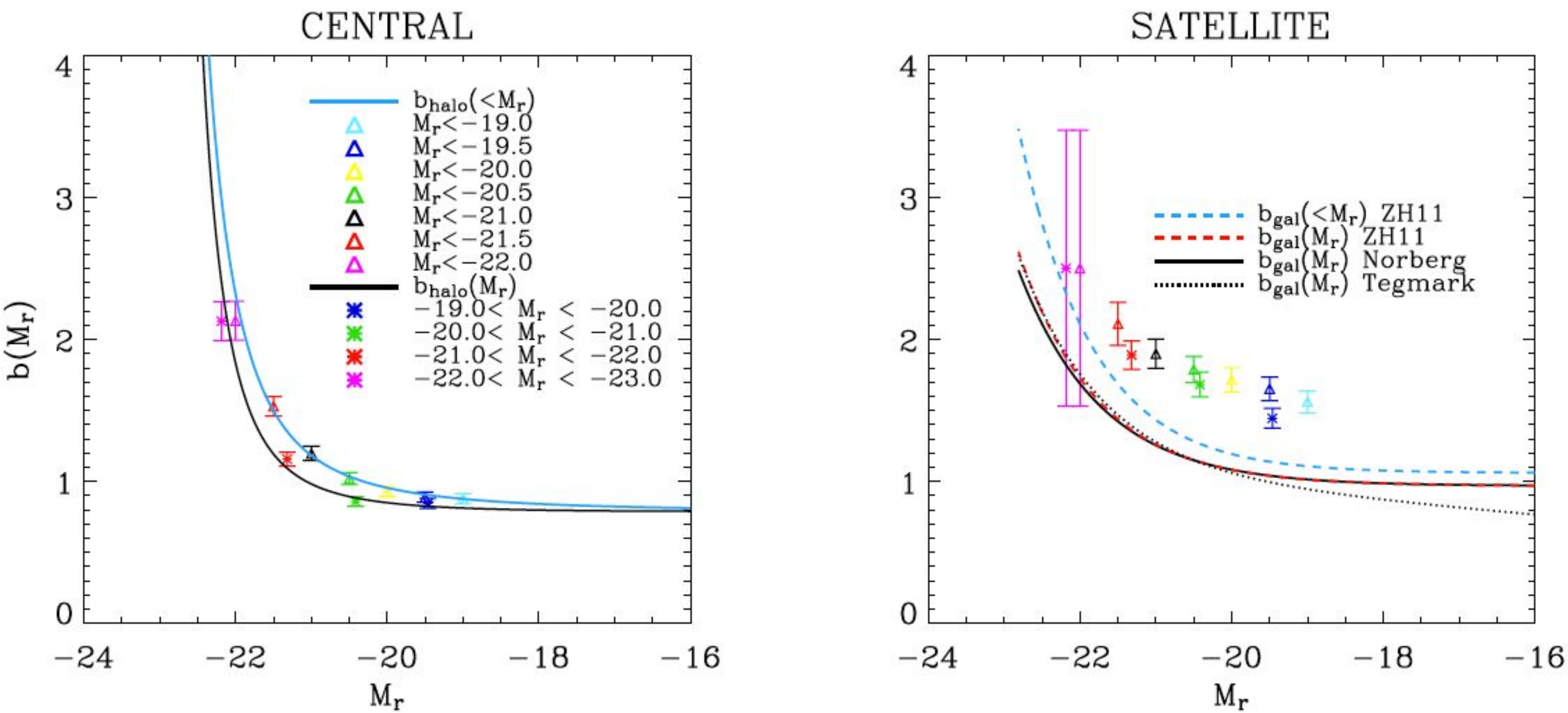


$b_{lin}(r)$ (M_r | (g-r))


TOTAL
b(M_r)
M_r
M_r<−19.0
M_r<−19.5
M_r<−20.0
M_r<−20.5
M_r<−21.0
M_r<−21.5
M_r<−22.0
−19.0< M_r < −20.0
−20.0< M_r < −21.0
−21.0< M_r < −22.0
−22.0< M_r < −23.0
b_gal(<M_r) ZH11

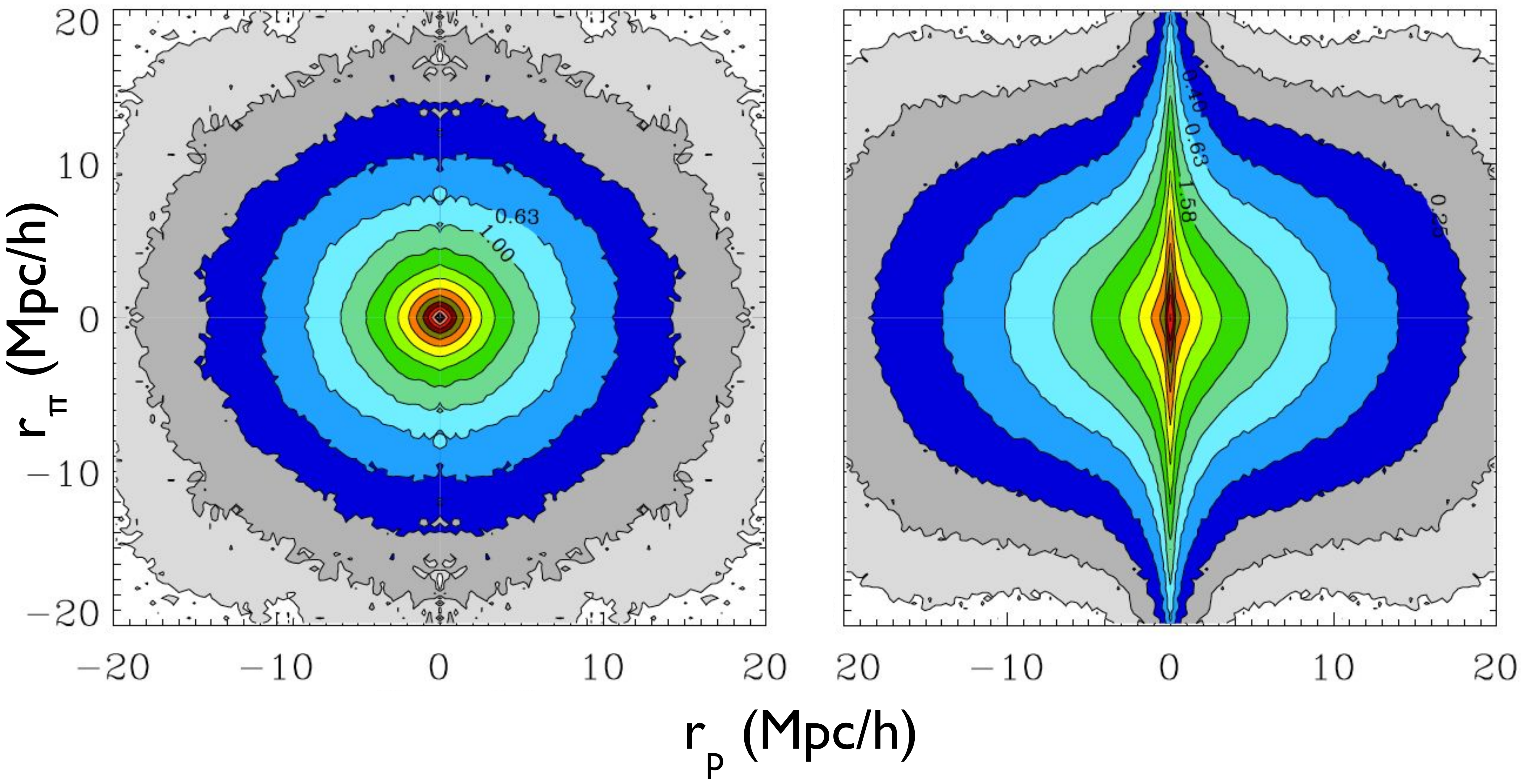

COLOUR
M_r
−19.0< M_r < −20.0
−20.0< M_r < −21.0
−21.0< M_r < −22.0
−22.0< M_r < −23.0
−19.0< M_r < −20.0
−20.0< M_r < −21.0
−21.0< M_r < −22.0
b_gal(M_r) ZH11
b_gal(M_r) Norberg
b_gal(M_r) Tegmark

$b_{lin}(r)$ (M_r | (g-r))

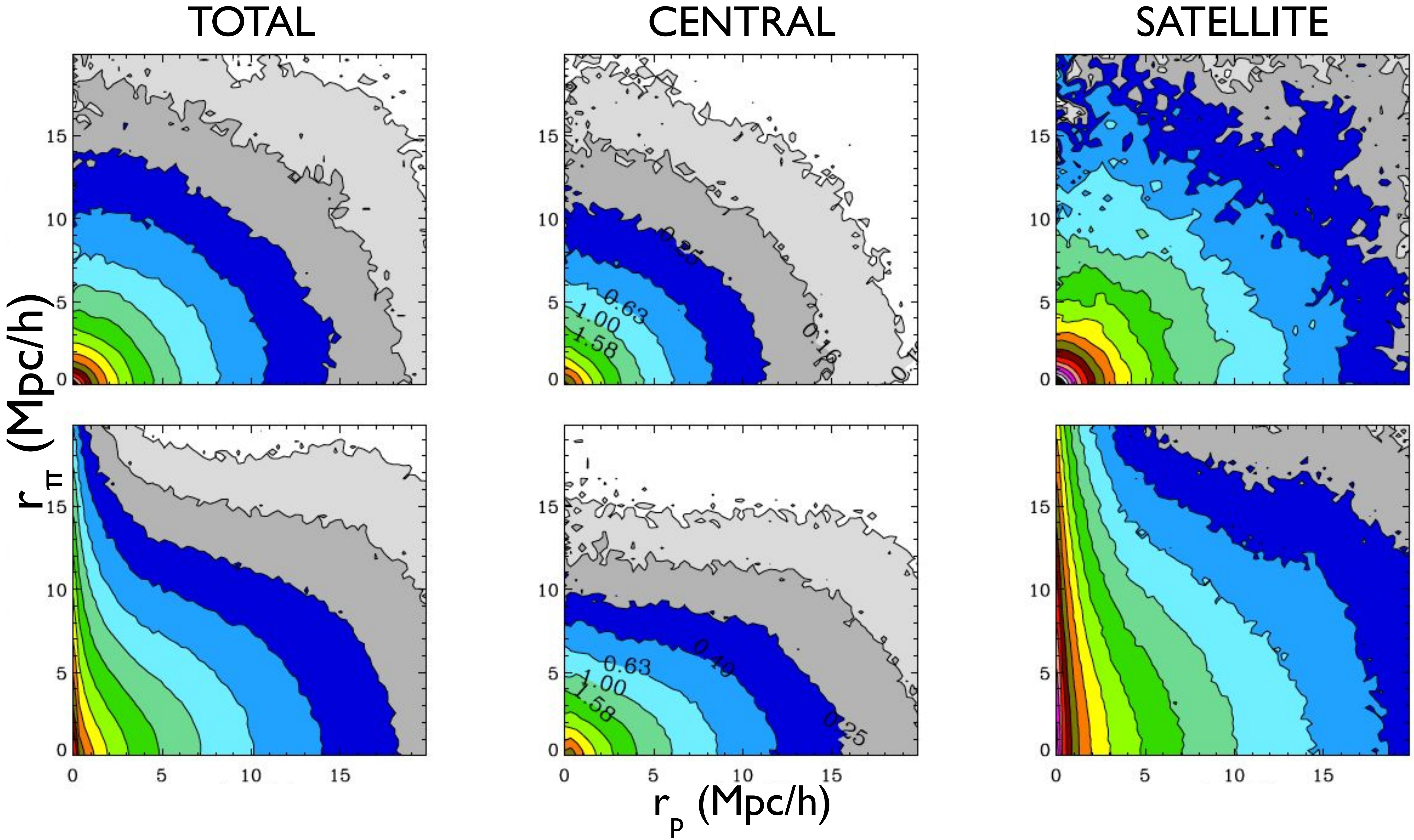


CENTRAL
b(M_r)
M_r
b_halo(<M_r)
M_r<−19.0
M_r<−19.5
M_r<−20.0
M_r<−20.5
M_r<−21.0
M_r<−21.5
M_r<−22.0
b_halo(M_r)
−19.0< M_r < −20.0
−20.0< M_r < −21.0
−21.0< M_r < −22.0
−22.0< M_r < −23.0
SATELLITE
b_gal(<M_r) ZH11
b_gal(M_r) ZH11
b_gal(M_r) Norberg
b_gal(M_r) Tegmark

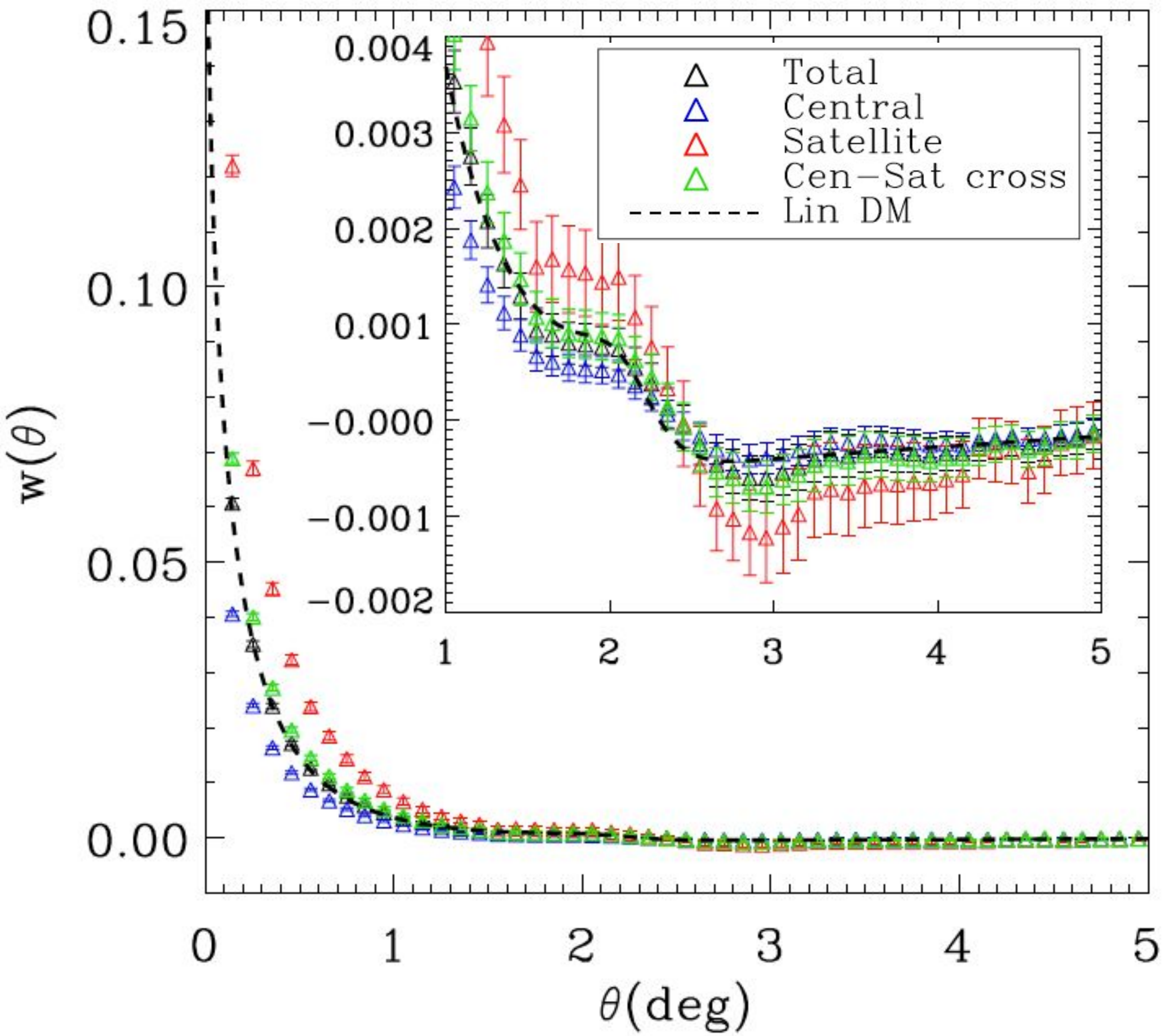
RSD


20
10
0
−10
−20
r_π (Mpc/h)
0.63
1.00
0.40
0.63
1.58
0.25
−20
−10
0
10
20
20
−10
0
10
20
r_p (Mpc/h)

RSD

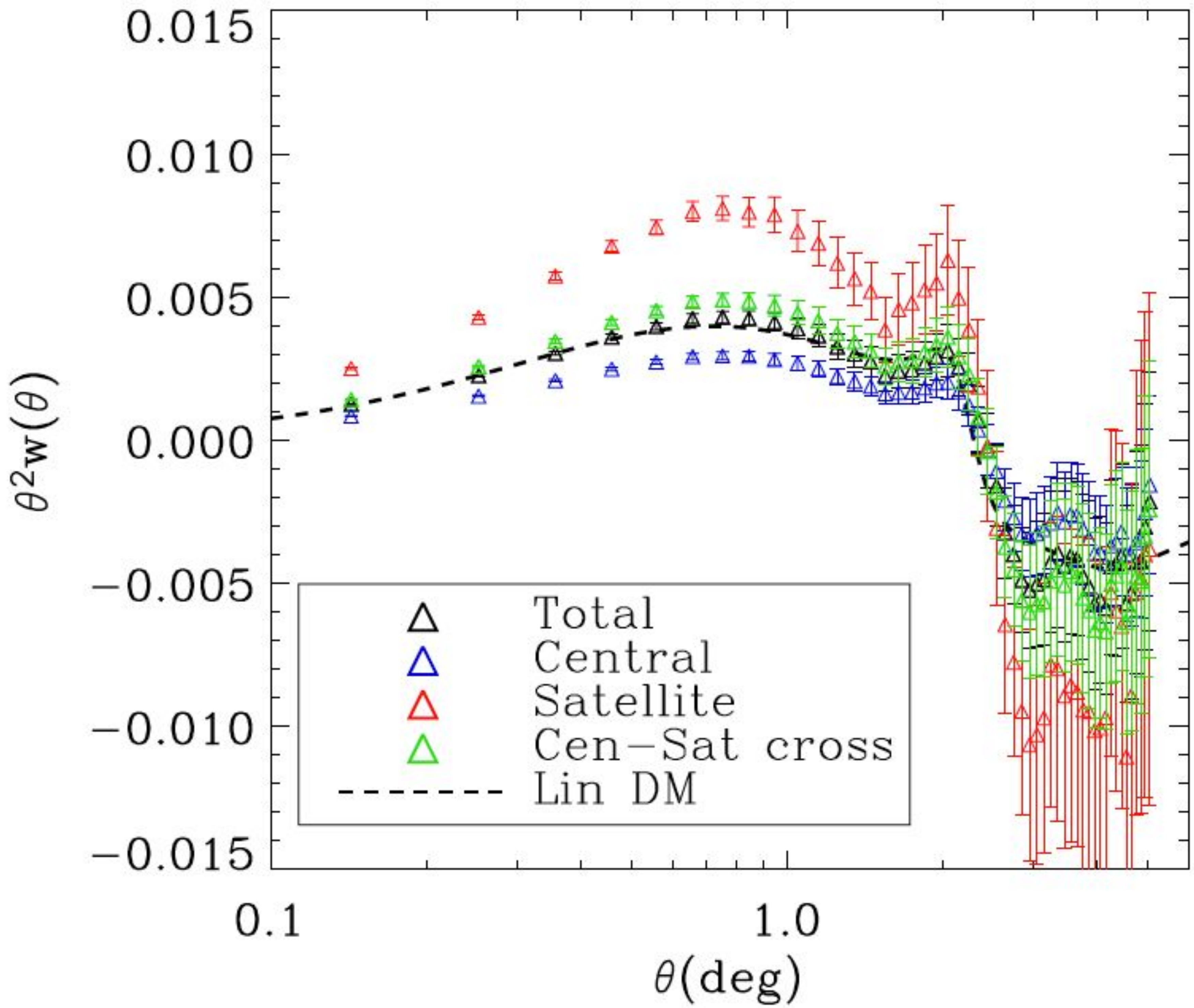

TOTAL
CENTRAL
SATELLITE
r_π (Mpc/h)
r_p (Mpc/h)
0.25
0.63
1.00
1.58
0.16
0.40
0.25
0.63
1.00
1.58

w(θ) BAO

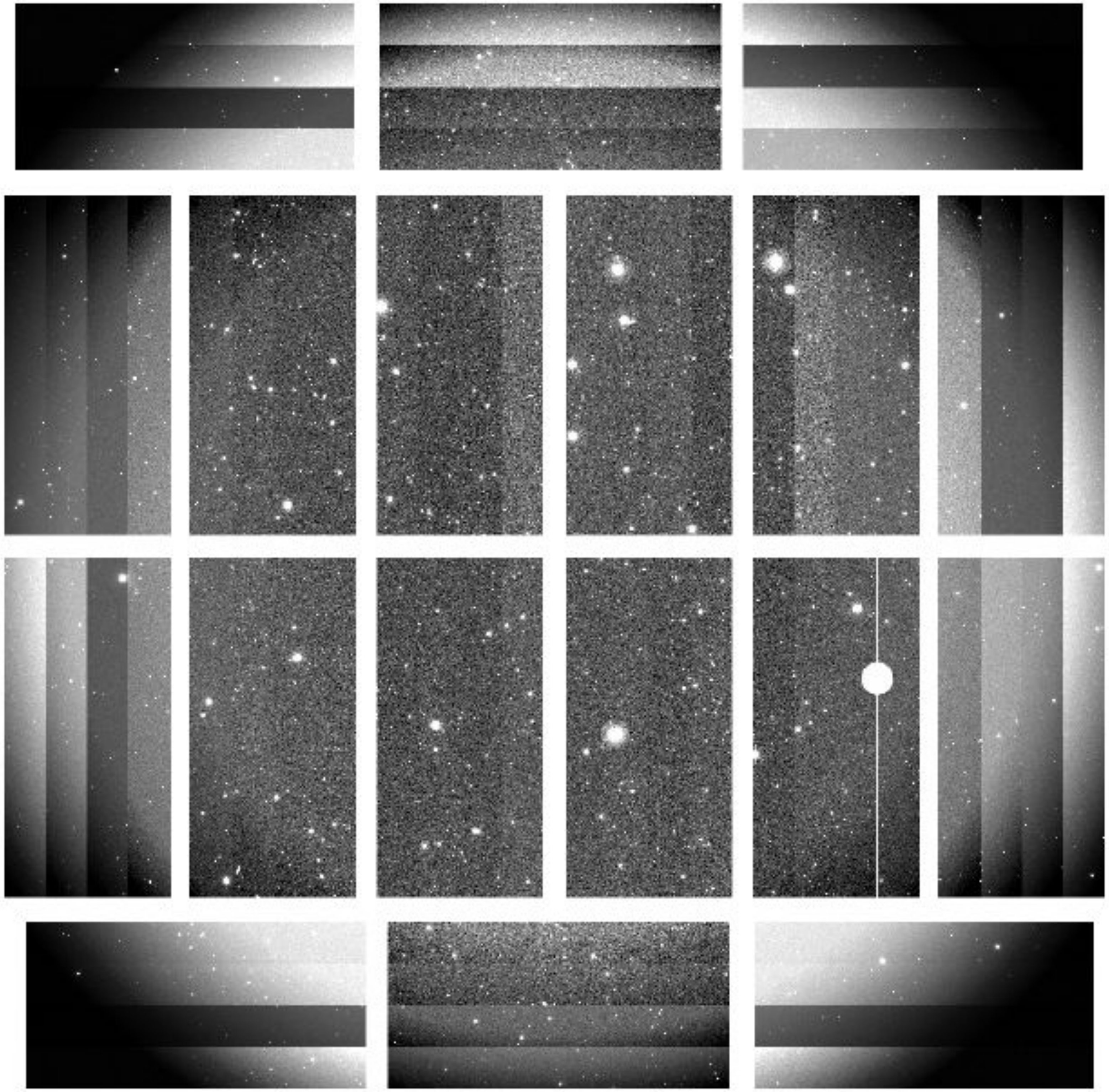


0.15
0.10
0.05
0.00
w(θ)
0
1
2
3
4
5
θ(deg)
0.004
0.003
0.002
0.001
−0.000
−0.001
−0.002
Total
Central
Satellite
Cen−Sat cross
Lin DM

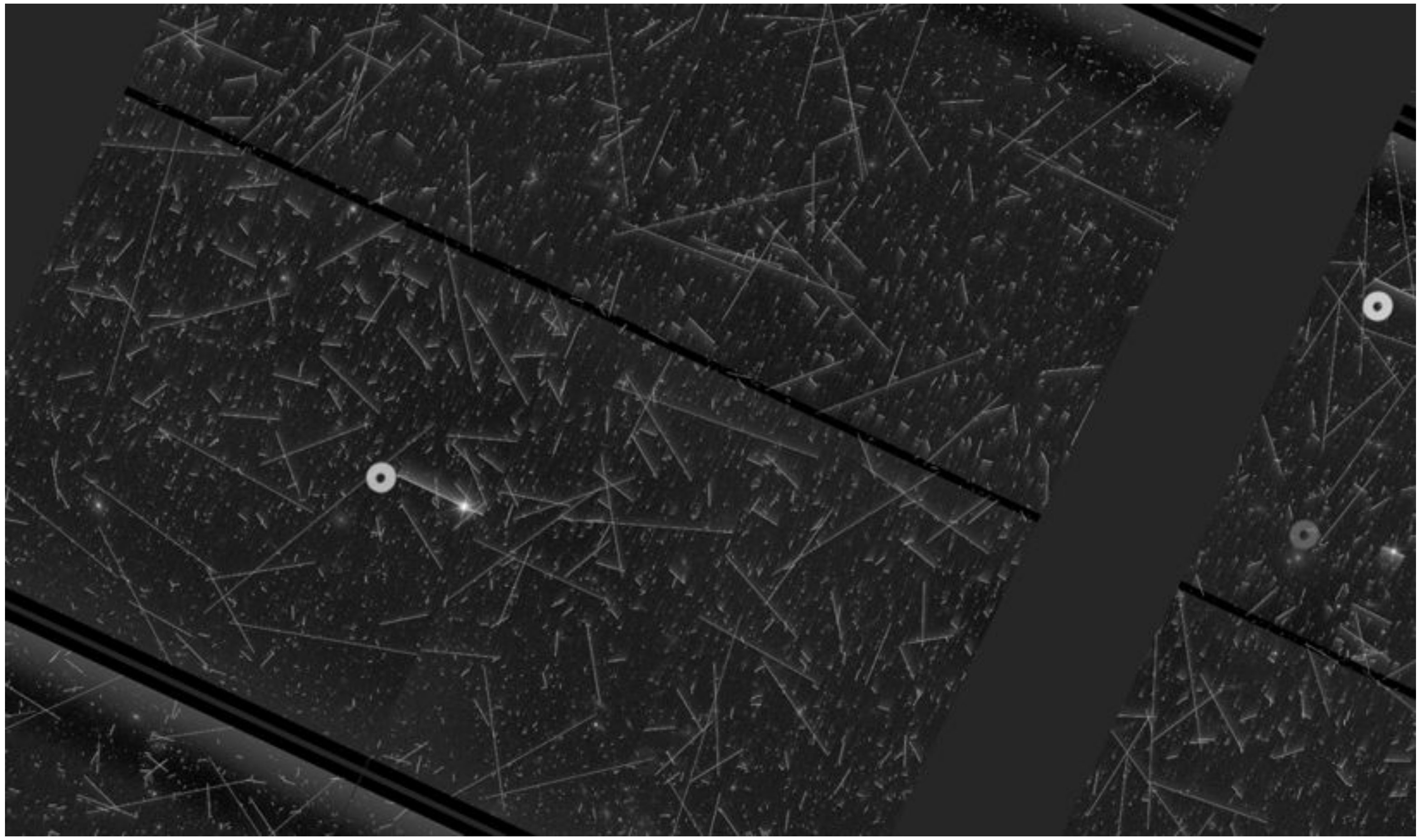
w(θ) BAO


0.015
0.010
0.005
0.000
−0.005
−0.010
−0.015
$\theta^2 w(\theta)$
0.1
1.0
θ(deg)
Total
Central
Satellite
Cen−Sat cross
Lin DM

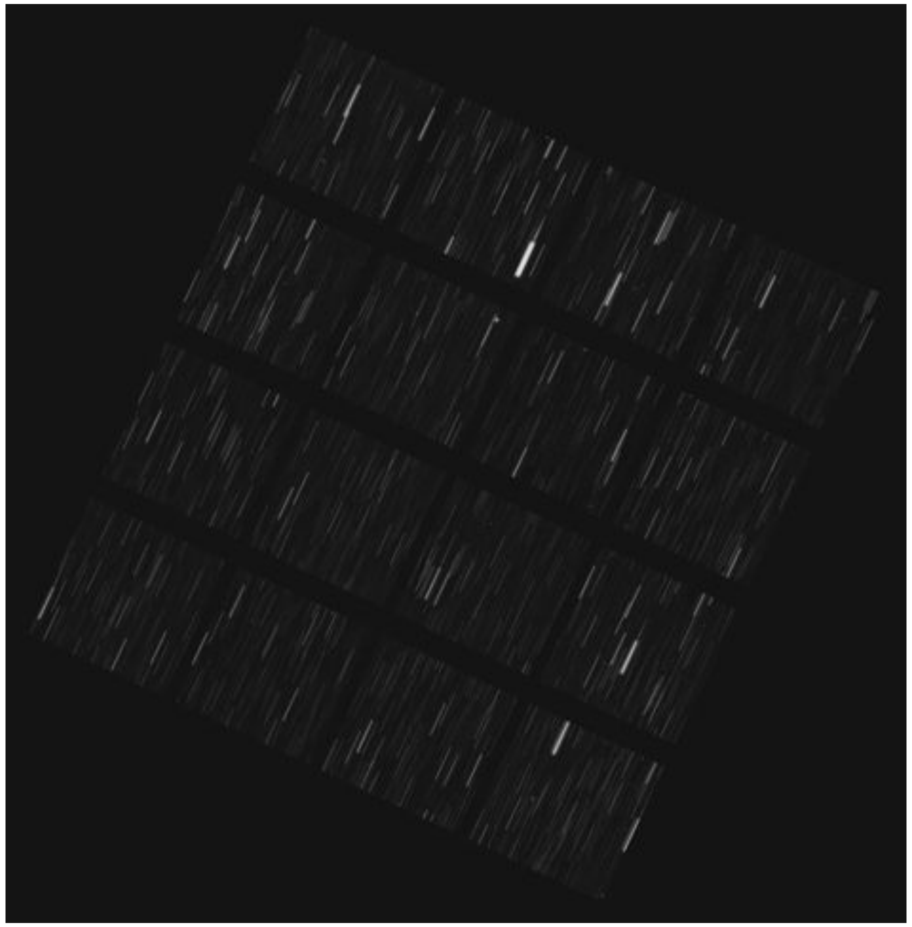
Pixel Simulation (PAU)



Pixel Simulation (Euclid)



VIS:image; credit P.Hudelot OU-SIM



NISP-S full focal plane
credit: N. Fourmanoit; J Zoubian (OU-SIM)

(some) References

Didactic (several papers from David W. Hogg), e.g.:

- Distance measures in Cosmology. David W. Hogg (2000). http://arxiv.org/abs/astro-ph/9905116
- Data analysis recipes: Fitting a model to data. David W. Hogg, Jo Bovy and Dustin Lang (2010). http://arxiv.org/abs/1008.4686
- The K correction. David W. Hogg et al. (2002). http://arxiv.org/abs/astro-ph/0210394

(some) References

Dark Energy:

- Dark Energy: A Short Review. M. J. Mortonson, D. H. Weinberg and M. White (2014). http://arxiv.org/abs/1401.0046
- Observational probes of Cosmic Acceleration. Weinberg et al. (2013). http://arxiv.org/pdf/1201.2434v2.pdf

Galaxy Formation:

- Galaxy Formation Theory. Benson (2010). http://arxiv.org/abs/1006.5394
- A primer on hierarchichal galaxy formation: the semi-analytical approach. Baugh (2006). http://arxiv.org/abs/astro-ph/0610031

(some) References

Halo model:

- Halo models of large scale structure. Asantha Cooray and Ravi Sheth (2002). http://arxiv.org/abs/astro-ph/0206508

Halo Occupation Distribution model (HOD):

- The halo occupation distribution: towards an empirical determination of the relation between galaxies and mass. Andreas A. Berlind and David H. Weinberg (2002). http://arxiv.org/abs/astro-ph/0109001
- The halo occupation distribution and the physics of galaxy formation. Andreas A. Berlind et al. (2002). http://arxiv.org/abs/astro-ph/0212357

(some) References

Semi analytical models:

- Physical Models of Galaxy Formation in a Cosmological Framework. Rachel S. Somerville (2014). https://arxiv.org/pdf/1412.2712.pdf