MINING VIRTUAL UNIVERSES (ONLINE)

Lectures and hands-on sessions

at ISSAC 2012

Gerard Lemson

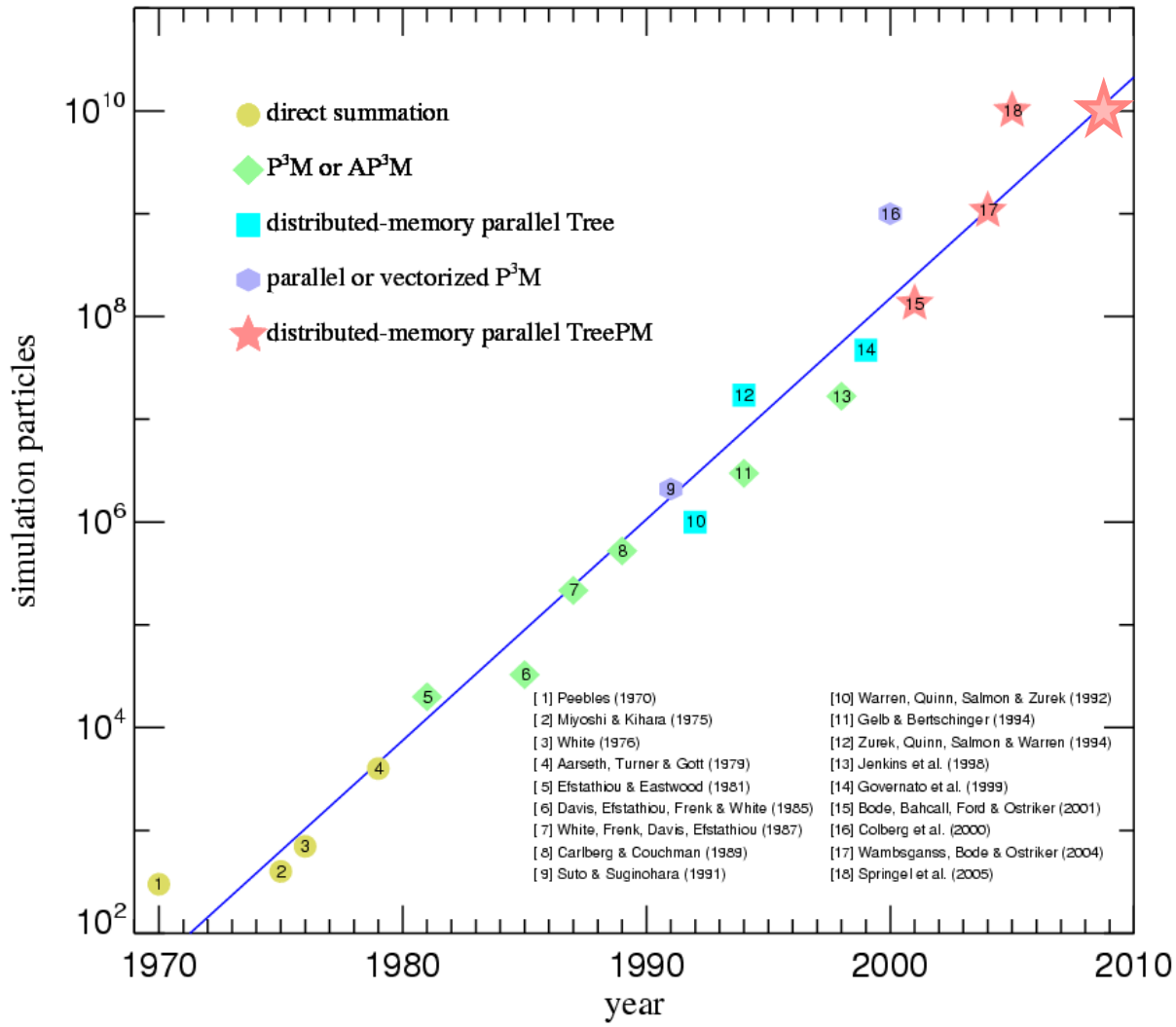
MPA, Garching, Germany

Why me?

- Millennium Run Database (online)
  - arXiv:astro-ph/0608019,
  - gavo.mpa-garching.mpg.de/Millennium
  - also multidark.org
- Millennium Run Observatory (online)
  - arXiv:1206.6923
  - galformod.mpa-garching.mpg.de/mrobs

With (thanks to):
Raul Angulo, Jeremy Blaizot, Emmanuel Bertin, Tamas Budavari, Darren Croton, Gabriella DeLucia, Matthias Egger, GAVO, Bruno Henriques, Gabriel-Dominique Marleau,
Roderik Overzier, Guo Qi, Volker Springel, Alex Szalay,
Virgo Consortium, Simon White.

Why here?
“Moore’s law” for N-body simulations



Courtesy Simon White

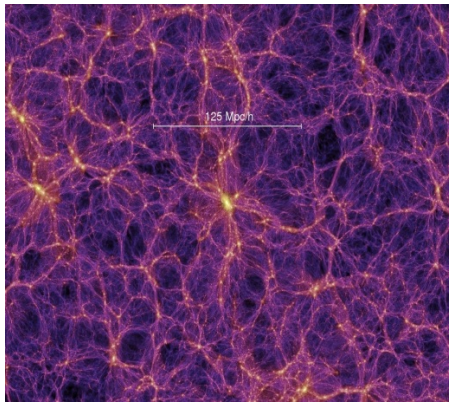
Lectures and hands-on sessions

- **L1: Overview + Phenomenology**
- L2: Virtual Universes in a Database
- L3: Virtual Observatory and Theory
- H1: Querying databases
- H2: Filling databases and publishing them

Max-Planck-Institut für Astrophysik

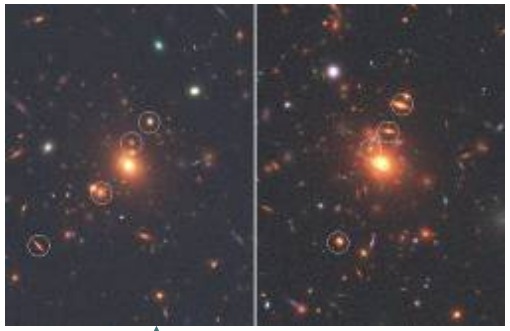
LE-HiPACC

Raw data: Particles

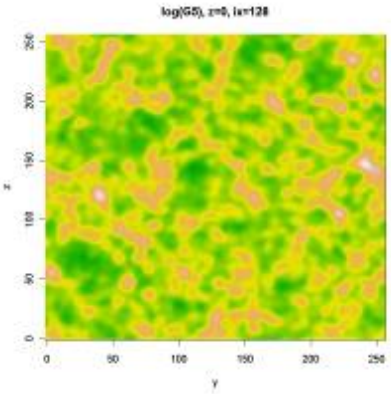


FOF groups and Subhalos

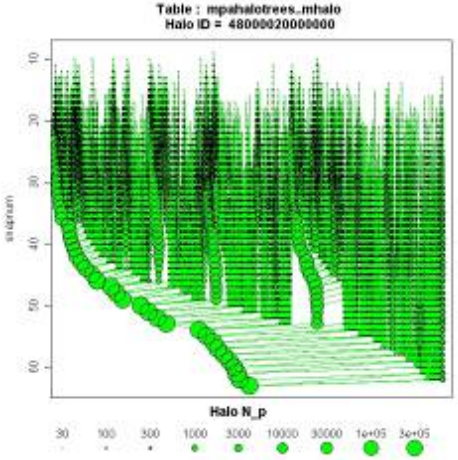
Mock images



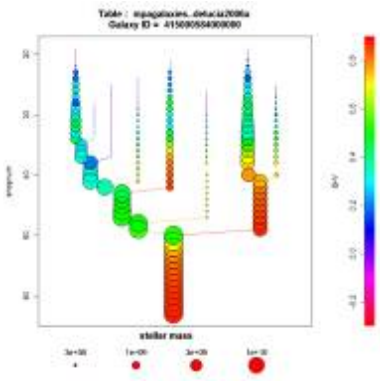
Density fields

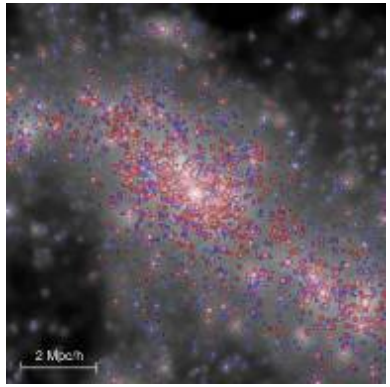


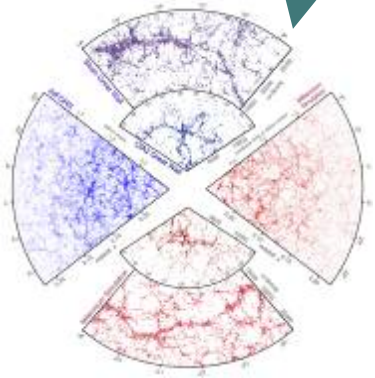
Subhalo merger trees



Synthetic galaxies (SAM)



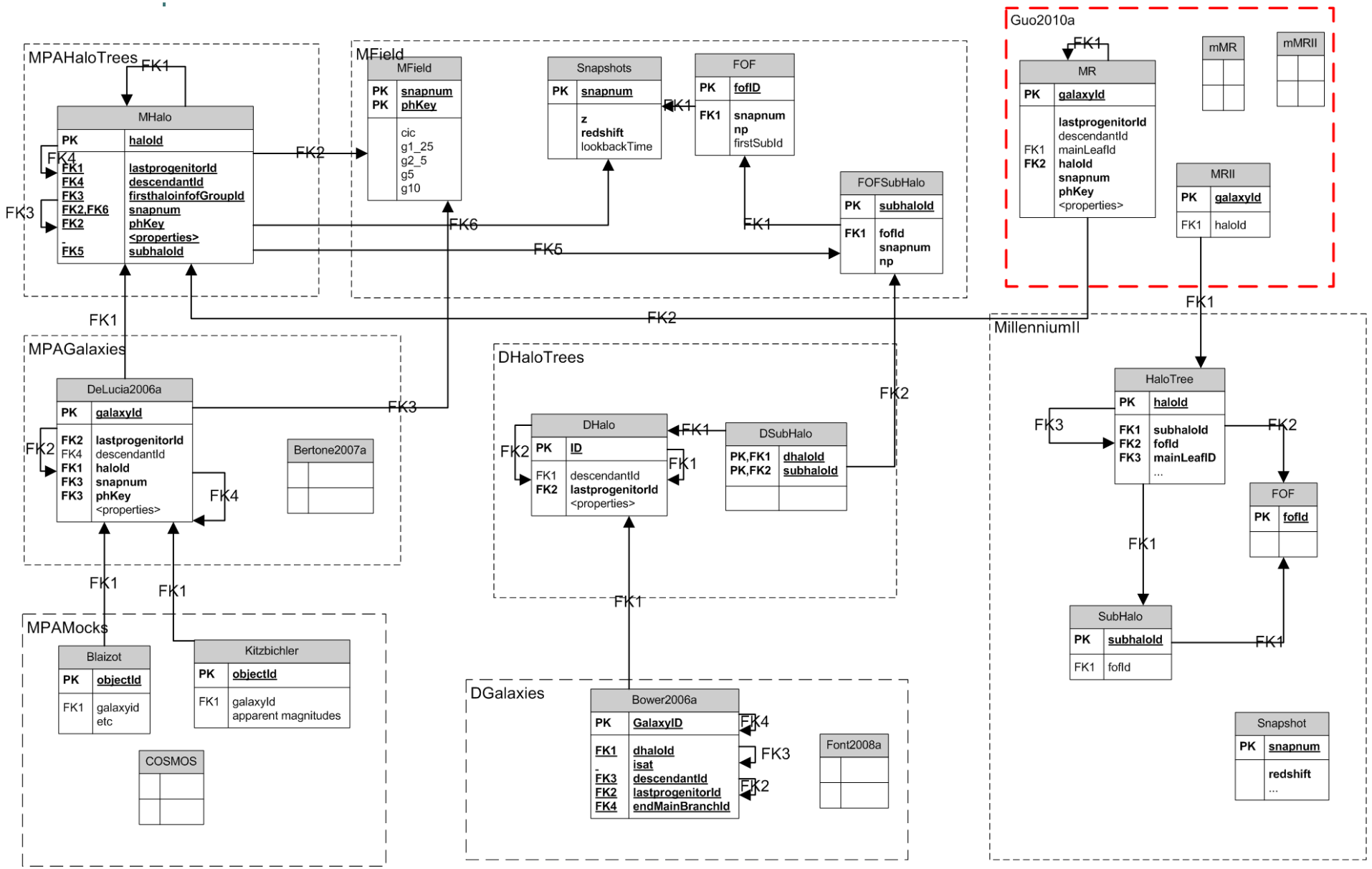




Mock catalogues

Lectures and hands-on sessions

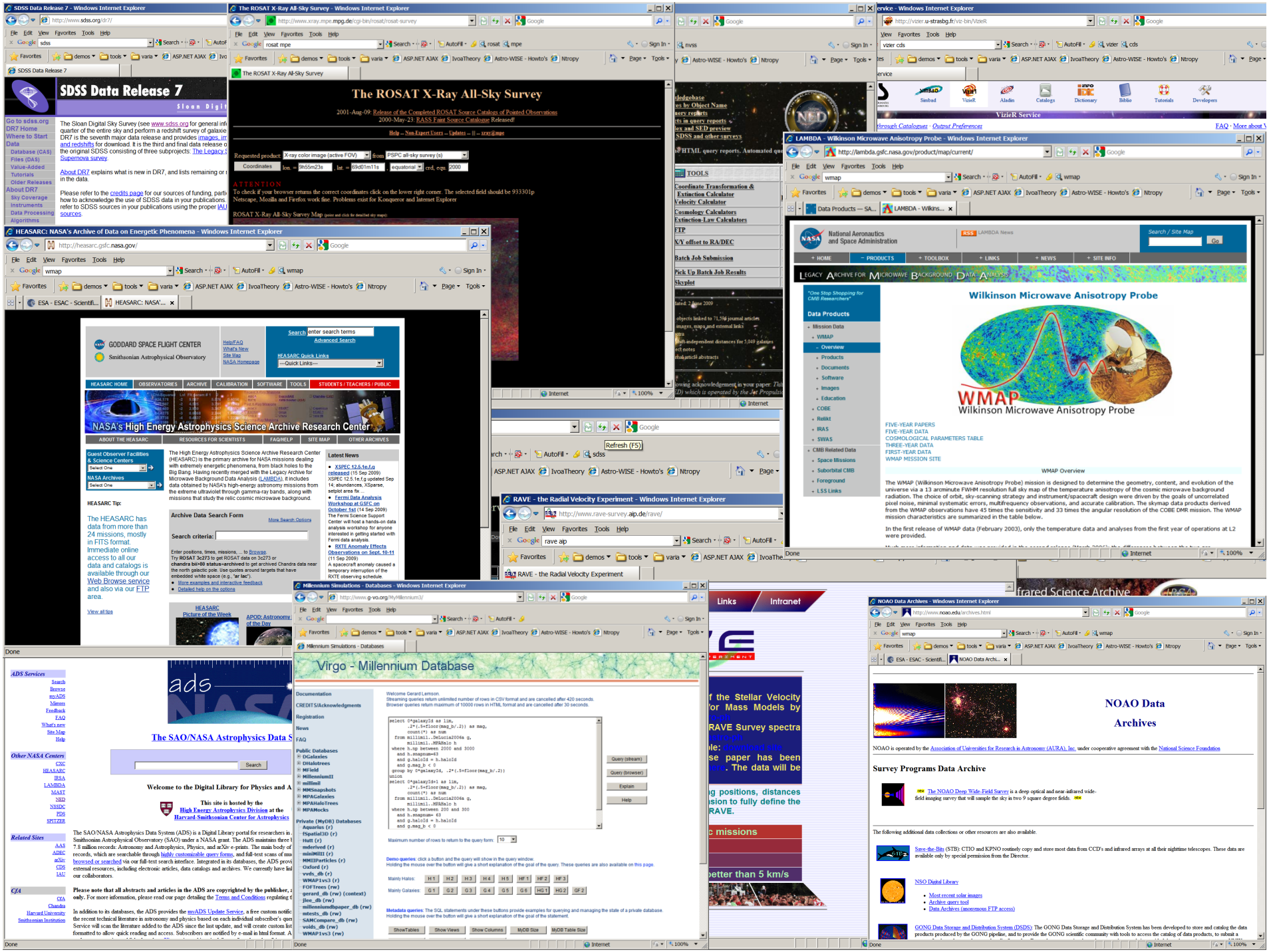
- L1: Overview + Phenomenology
- **L2: Virtual Universes in a Database**
- L3: Virtual Observatory and Theory
- H1: Querying databases
- H2: Filling databases and publishing them


MPAHaloTrees
MHalo
PK haloId
FK1 lastprogenitorId
FK4 descendantId
FK3 firsthaloinfofGroupId
FK2,FK6 snapnum
FK2 phKey
- <properties>
FK5 subhaloId
MField
MField
PK snapnum
PK phKey
cic
g1_25
g2_5
g5
g10
Snapshots
PK snapnum
z
redshift
lookbackTime
FOF
PK fofID
FK1 snapnum
np
firstSubId
FOFSubHalo
PK subhaloId
FK1 fofId
snapnum
np
Guo2010a
MR
PK galaxyId
FK1 lastprogenitorId
descendantId
mainLeafId
FK2 haloId
snapnum
phKey
<properties>
mMR
mMRII
MRII
PK galaxyId
FK1 haloId
MPAGalaxies
DeLucia2006a
PK galaxyId
FK2 lastprogenitorId
FK4 descendantId
FK1 haloId
FK3 snapnum
FK3 phKey
<properties>
Bertone2007a
DHaloTrees
DHalo
PK ID
FK1 descendantId
FK2 lastprogenitorId
<properties>
DSubHalo
PK,FK1 dhaloId
PK,FK2 subhaloId
MillenniumII
HaloTree
PK haloId
FK1 subhaloId
FK2 fofId
FK3 mainLeafID
...
FOF
PK fofId
SubHalo
PK subhaloId
FK1 fofId
Snapshot
PK snapnum
redshift
...
MPAMocks
Blaizot
PK objectId
FK1 galaxyid
etc
Kitzbichler
PK objectId
FK1 galaxyId
apparent magnitudes
COSMOS
DGalaxies
Bower2006a
PK GalaxyID
FK1 dhaloId
isat
FK3 descendantId
FK2 lastprogenitorId
FK4 endMainBranchId
Font2008a
FK1
FK2
FK3
FK4
FK5
FK6

Lectures and hands-on sessions

- L1: Overview + Phenomenology
- L2: Virtual Universes in a Database
- **L3: Virtual Observatory and Theory**
- H1: Querying databases
- H2: Filling databases and publishing them

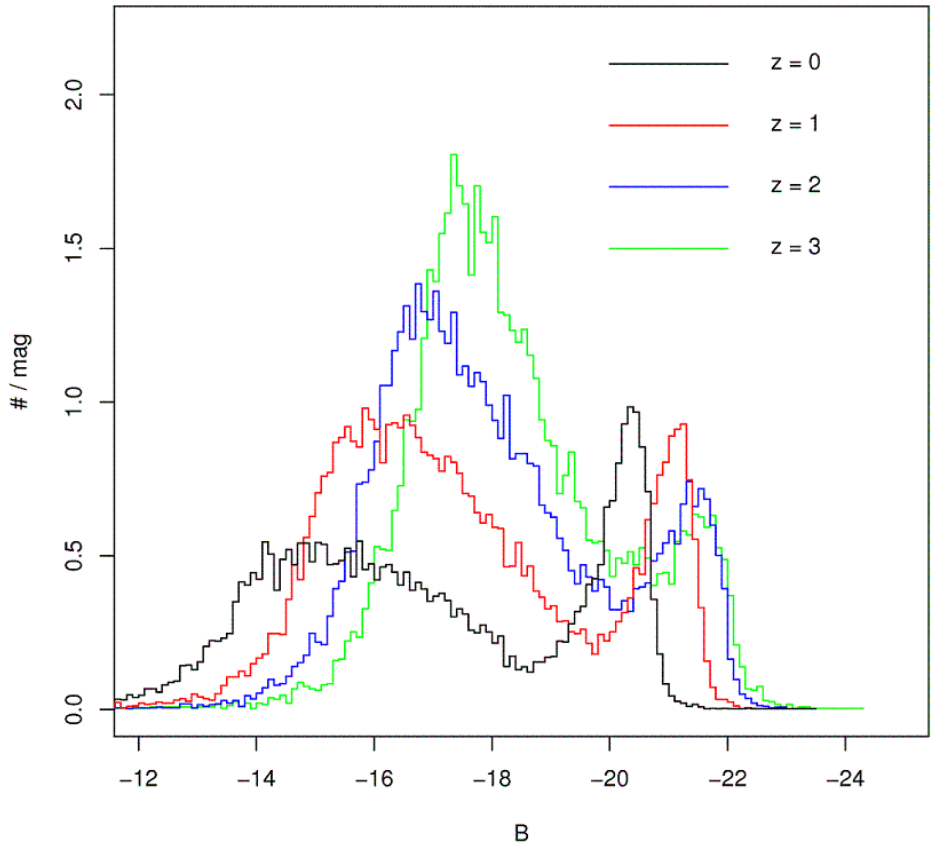
Astro Grid
ArVO
Armenian Virtual Observatory
Aus-VO
BRAVO
China-VO
CVO
VObs.it
esa
VAO
VIRTUAL ASTRONOMICAL OBSERVATORY
IVOA
France
vo-i
EURO VO
UKR VO
GERMAN ASTROPHYSICAL
GAVO
VIRTUAL OBSERVATORY
HVO
Hungarian Virtual Observatory
JVO
NOVA
RVO
SVO
Spanish Virtual Observatory


SDSS Data Release 7 - Windows Internet Explorer
SDSS Data Release 7
The Sloan Digital Sky Survey (see www.sdss.org for general info
quarter of the entire sky and perform a redshift survey of galaxie
DR7 is the seventh major data release and provides images, im
and redshifts for download. It is the third and final data release o
the original SDSS consisting of three subprojects: The Legacy S
Supernova survey.
About DR7 explains what is new in DR7, and lists remaining or
in the data.
Please refer to the credits page for our sources of funding, partic
how to acknowledge the use of SDSS data in your publications.
refer to SDSS sources in your publications using the proper IAU
sources.
The ROSAT X-Ray All-Sky Survey - Windows Internet Explorer
The ROSAT X-Ray All-Sky Survey
2001-Aug-09: Release of the Completed ROSAT Source Catalogs of Pointed Observations
2000-May-23: RASS Faint Source Catalogue Released!
Requested product: X-ray color image (active FOV) from PSPC all-sky survey (s)
Coordinates lon. = 9h55m23s , lat. = 69d01m11s equatorial crd, eqx: 2000
ATTENTION
To check if your browser returns the correct coordinates click on the lower right corner. The selected field should be 933301p
Netscape, Mozilla and Firefox work fine. Problems exist for Konqueror and Internet Explorer
ROSAT X-Ray All-Sky Survey Map
HEASARC: NASA's Archive of Data on Energetic Phenomena - Windows Internet Explorer
GODDARD SPACE FLIGHT CENTER
Smithsonian Astrophysical Observatory
NASA's High Energy Astrophysics Science Archive Research Center
HEASARC HOME
OBSERVATORIES
ARCHIVE
CALIBRATION
SOFTWARE
TOOLS
STUDENTS / TEACHERS / PUBLIC
The High Energy Astrophysics Science Archive Research Center (HEASARC) is the primary archive for NASA missions dealing with extremely energetic phenomena, from black holes to the Big Bang. Having recently merged with the Legacy Archive for Microwave Background Data Analysis (LAMBDA), it includes data obtained by NASA's high-energy astronomy missions from the extreme ultraviolet through gamma-ray bands, along with missions that study the relic cosmic microwave background.
Archive Data Search Form
Search criteria:
HEASARC Tip:
The HEASARC has data from more than 24 missions, mostly in FITS format. Immediate online access to all our data and catalogs is available through our Web Browse service and also via our FTP area.
Latest News
RAVE - the Radial Velocity Experiment - Windows Internet Explorer
RAVE - the Radial Velocity Experiment
LAMBDA - Wilkinson Microwave Anisotropy Probe - Windows Internet Explorer
National Aeronautics and Space Administration
LEGACY ARCHIVE FOR MICROWAVE BACKGROUND DATA ANALYSIS
Wilkinson Microwave Anisotropy Probe
WMAP
Wilkinson Microwave Anisotropy Probe
Data Products
WMAP Overview
The WMAP (Wilkinson Microwave Anisotropy Probe) mission is designed to determine the geometry, content, and evolution of the universe via a 13 arcminute FWHM resolution full sky map of the temperature anisotropy of the cosmic microwave background radiation. The choice of orbit, sky-scanning strategy and instrument/spacecraft design were driven by the goals of uncorrelated pixel noise, minimal systematic errors, multifrequency observations, and accurate calibration. The skymap data products derived from the WMAP observations have 45 times the sensitivity and 33 times the angular resolution of the COBE DMR mission. The WMAP mission characteristics are summarized in the table below.
In the first release of WMAP data (February 2003), only the temperature data and analyses from the first year of operations at L2 were provided.
VizieR Service - Windows Internet Explorer
VizieR Service
Millennium Simulations - Databases - Windows Internet Explorer
Virgo - Millennium Database
Links
Intranet
NOAO Data Archives - Windows Internet Explorer
NOAO Data Archives
Survey Programs Data Archive
ADS Services
The SAO/NASA Astrophysics Data S
Welcome to the Digital Library for Physics and A
This site is hosted by the High Energy Astrophysics Division at the Harvard-Smithsonian Center for Astrophysics
Related Sites
CfA

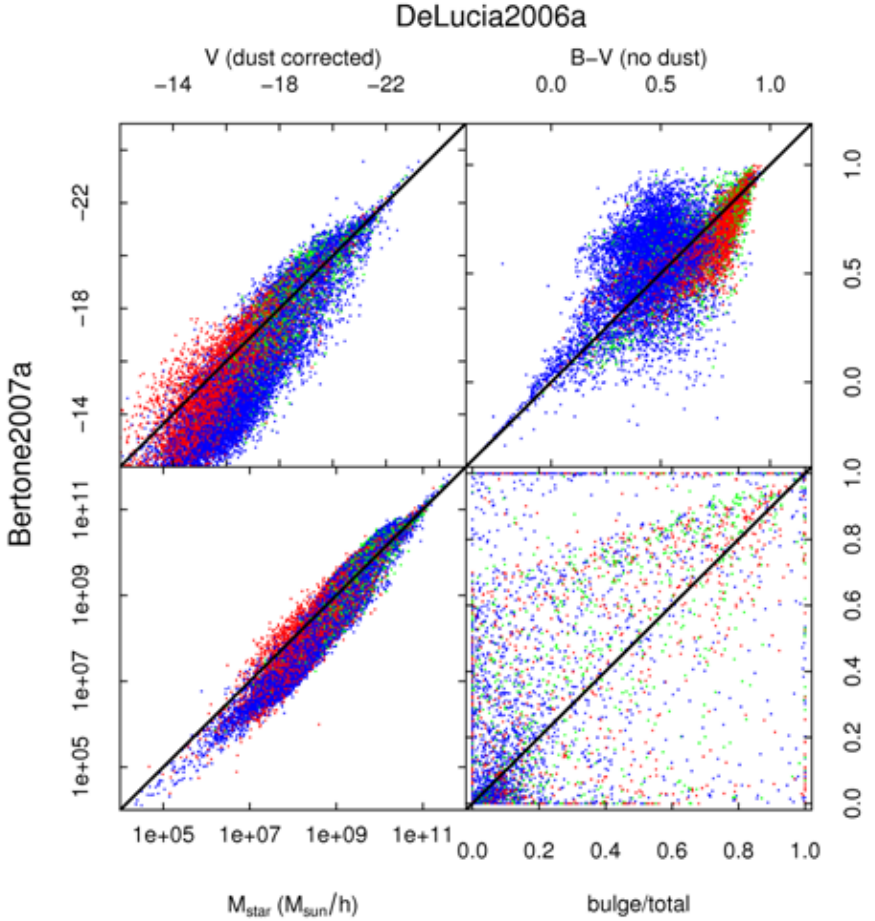
Lectures and hands-on sessions

- L1: Overview + Phenomenology
- L2: Virtual Universes in a Database
- L3: Virtual Observatory and Theory
- **H1: Querying databases**
- H2: Filling databases and publishing them

Querying the database



Conditional luminosity functions:
Galaxies in FOFs ~1000 particles

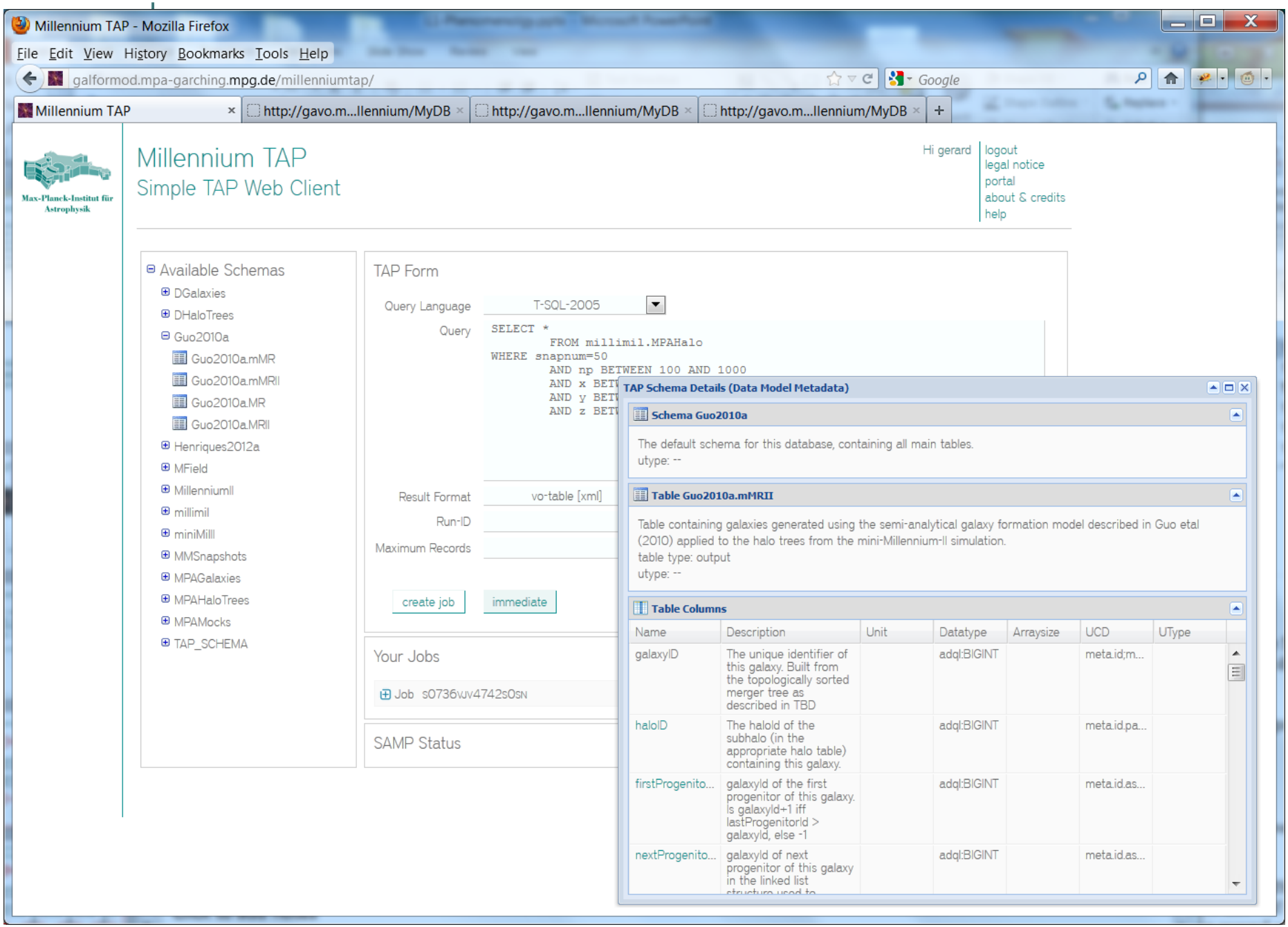


type 012

Comparing SAMs:
DeLucia2006a vs Bertone2007a

Lectures and hands-on sessions

- L1: Overview + Phenomenology
- L2: Virtual Universes in a Database
- L3: Virtual Observatory and Theory
- H1: Querying databases
- **H2: Filling databases and publishing them**
  - Using results from hands-on sessions Darren, Peter and Risa,


Millennium TAP - Mozilla Firefox
File Edit View History Bookmarks Tools Help
galformod.mpa-garching.mpg.de/millenniumtap/
Google
Millennium TAP
http://gavo.m...llennium/MyDB
http://gavo.m...llennium/MyDB
http://gavo.m...llennium/MyDB
Max-Planck-Institut für Astrophysik
Millennium TAP
Simple TAP Web Client
Hi gerard
logout
legal notice
portal
about & credits
help
Available Schemas
DGalaxies
DHaloTrees
Guo2010a
Guo2010a.mMR
Guo2010a.mMRII
Guo2010a.MR
Guo2010a.MRII
Henriques2012a
MField
MillenniumII
millimil
miniMilll
MMSnapshots
MPAGalaxies
MPAHaloTrees
MPAMocks
TAP_SCHEMA
TAP Form
Query Language
T-SQL-2005
Query
SELECT *
FROM millimil.MPAHalo
WHERE snapnum=50
AND np BETWEEN 100 AND 1000
AND x BETW
AND y BETW
AND z BETW
Result Format
vo-table [xml]
Run-ID
Maximum Records
create job
immediate
Your Jobs
Job s0736wjv4742s0sn
SAMP Status
TAP Schema Details (Data Model Metadata)
Schema Guo2010a
The default schema for this database, containing all main tables.
utype: --
Table Guo2010a.mMRII
Table containing galaxies generated using the semi-analytical galaxy formation model described in Guo etal (2010) applied to the halo trees from the mini-Millennium-II simulation.
table type: output
utype: --
Table Columns
Name
Description
Unit
Datatype
Arraysize
UCD
UType
galaxyID
The unique identifier of this galaxy. Built from the topologically sorted merger tree as described in TBD
adql:BIGINT
meta.id;m...
haloID
The haloId of the subhalo (in the appropriate halo table) containing this galaxy.
adql:BIGINT
meta.id.pa...
firstProgenito...
galaxyId of the first progenitor of this galaxy. Is galaxyId+1 iff lastProgenitorId > galaxyId, else -1
adql:BIGINT
meta.id.as...
nextProgenito...
galaxyId of next progenitor of this galaxy in the linked list
adql:BIGINT
meta.id.as...

Max-Planck-Institut für Astrophysik

PHENOMENOLOGY: SIMULATION PIPELINES AND DATA PRODUCTS

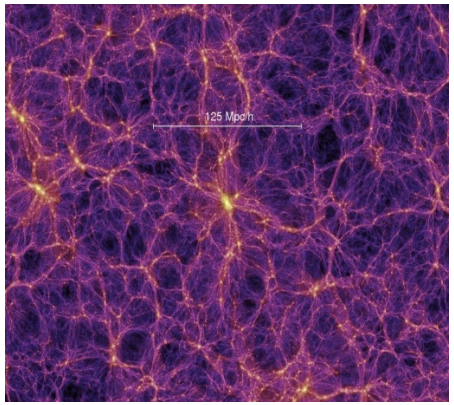
Exemplified by the

Millennium Run Observatory

Max-Planck-Institut für Astrophysik

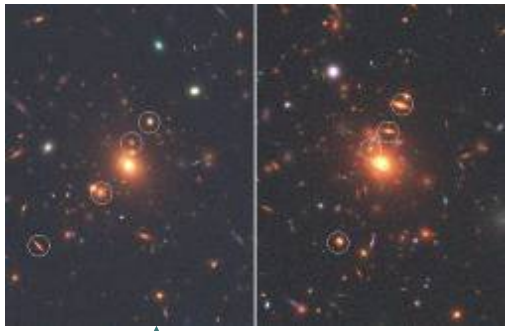
LE-HiPACC

Raw data: Particles

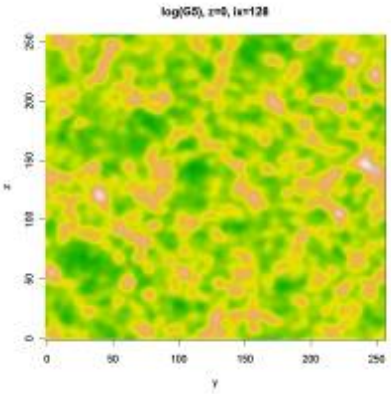


FOF groups and Subhalos

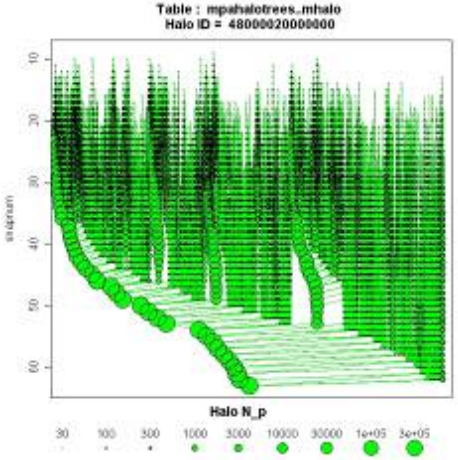
Mock images



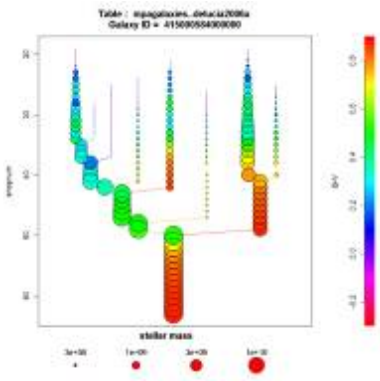
Density fields

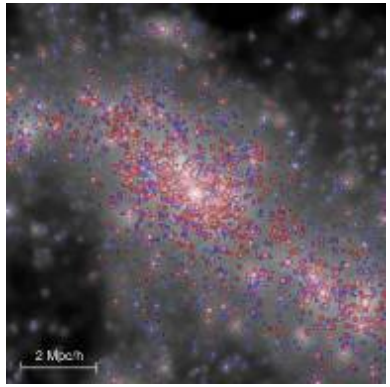


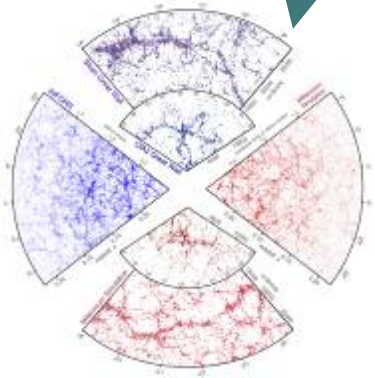
Subhalo merger trees



Synthetic galaxies (SAM)







Mock catalogues

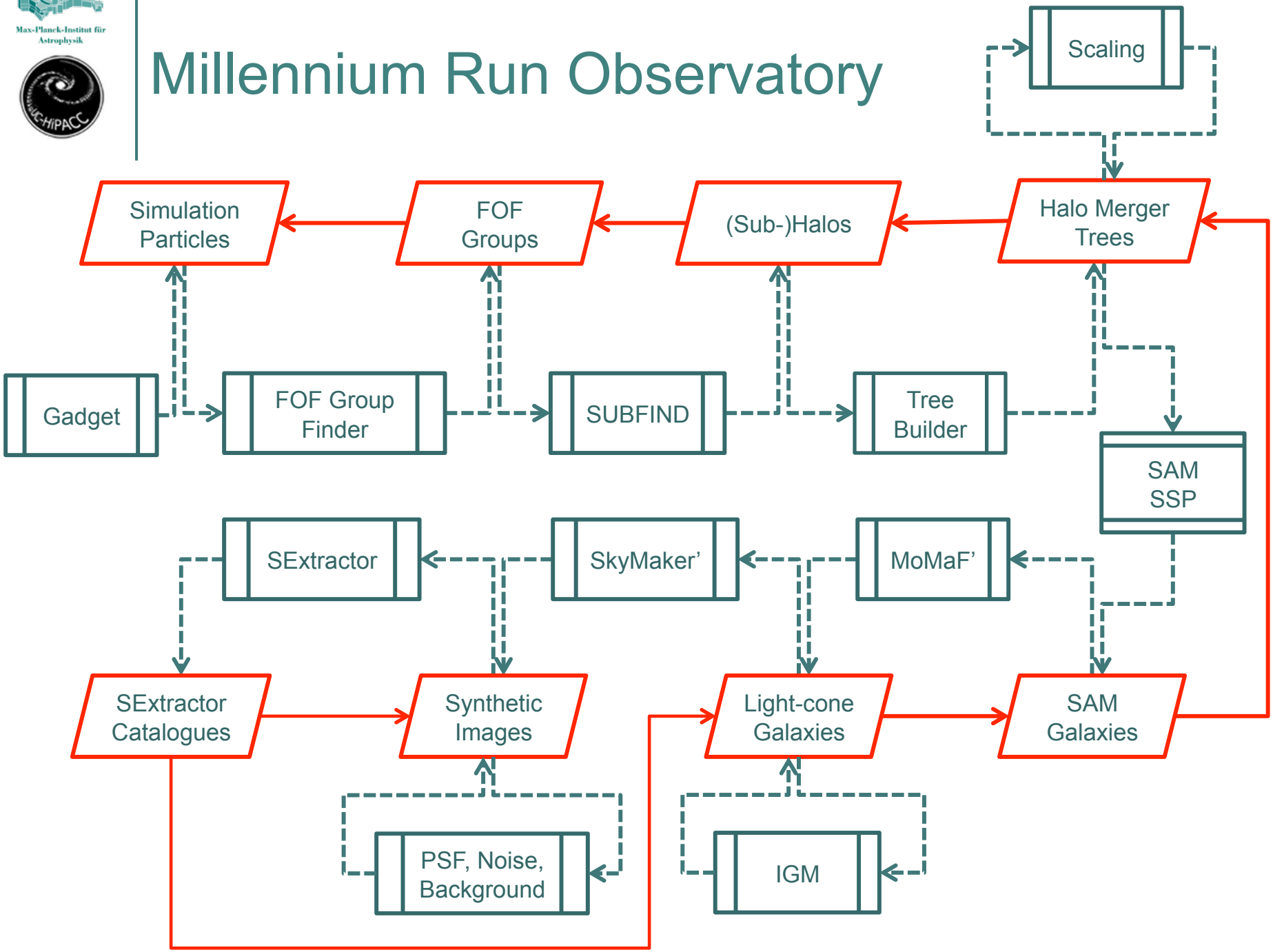
Millennium Run Observatory

Overzier, GL etal 2012, arXiv:1206:6923

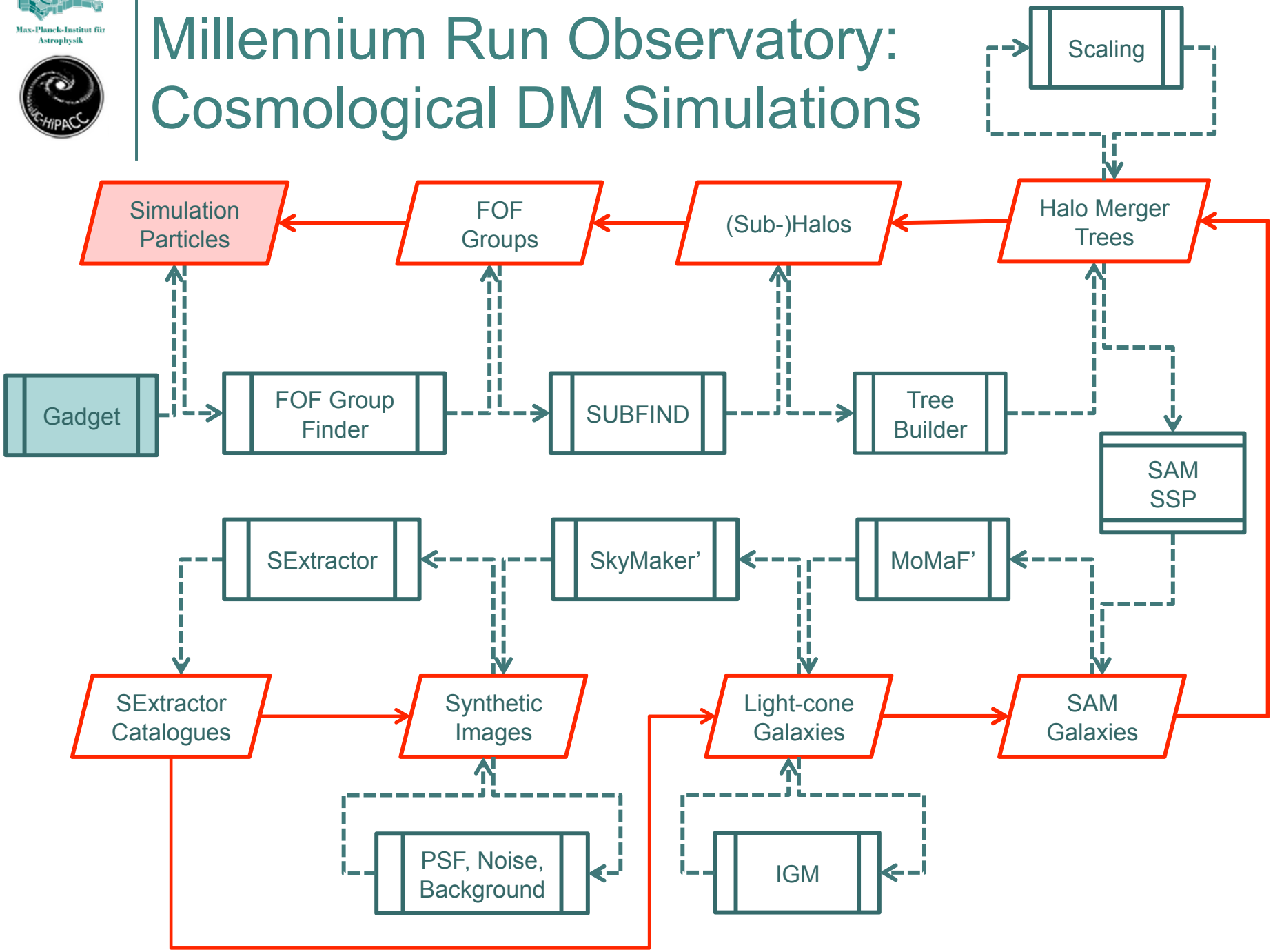
- Extend the Millennium Run project approach by producing data products directly corresponding to observations, namely synthetic images and extracted source catalogs
- Aid theorists in testing analytical models against observations
- Aid observers in making detailed predictions for observations and better analyses of observational data
- Allow the community to subject the models to new kinds of tests
- Allow observers and theorists to work toward each other from either direction with the freedom of where to meet.
- Allow detailed comparisons with synthetic observations produced by other groups performing cosmological simulations
- Allow calibration of observational analysis methods by making available synthetic data for which the entire underlying physical ”reality” is known.
- Extend the realism with which semi-analytic models can address questions such as what is the probability that a z 10 galaxy will be detected within a particular observational data set?
- Provide a framework for future virtual theoretical observatories

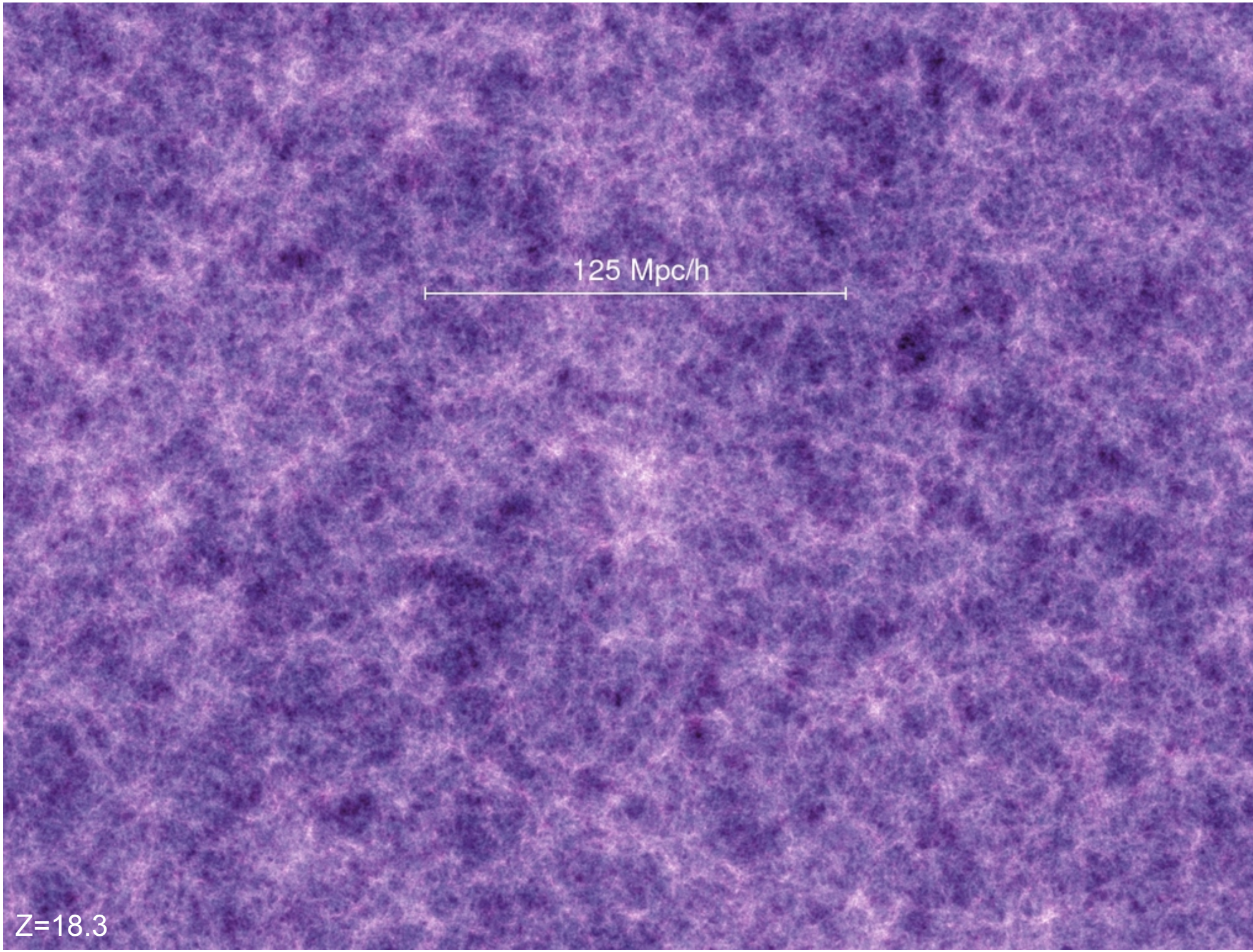
Max-Planck-Institut für Astrophysik

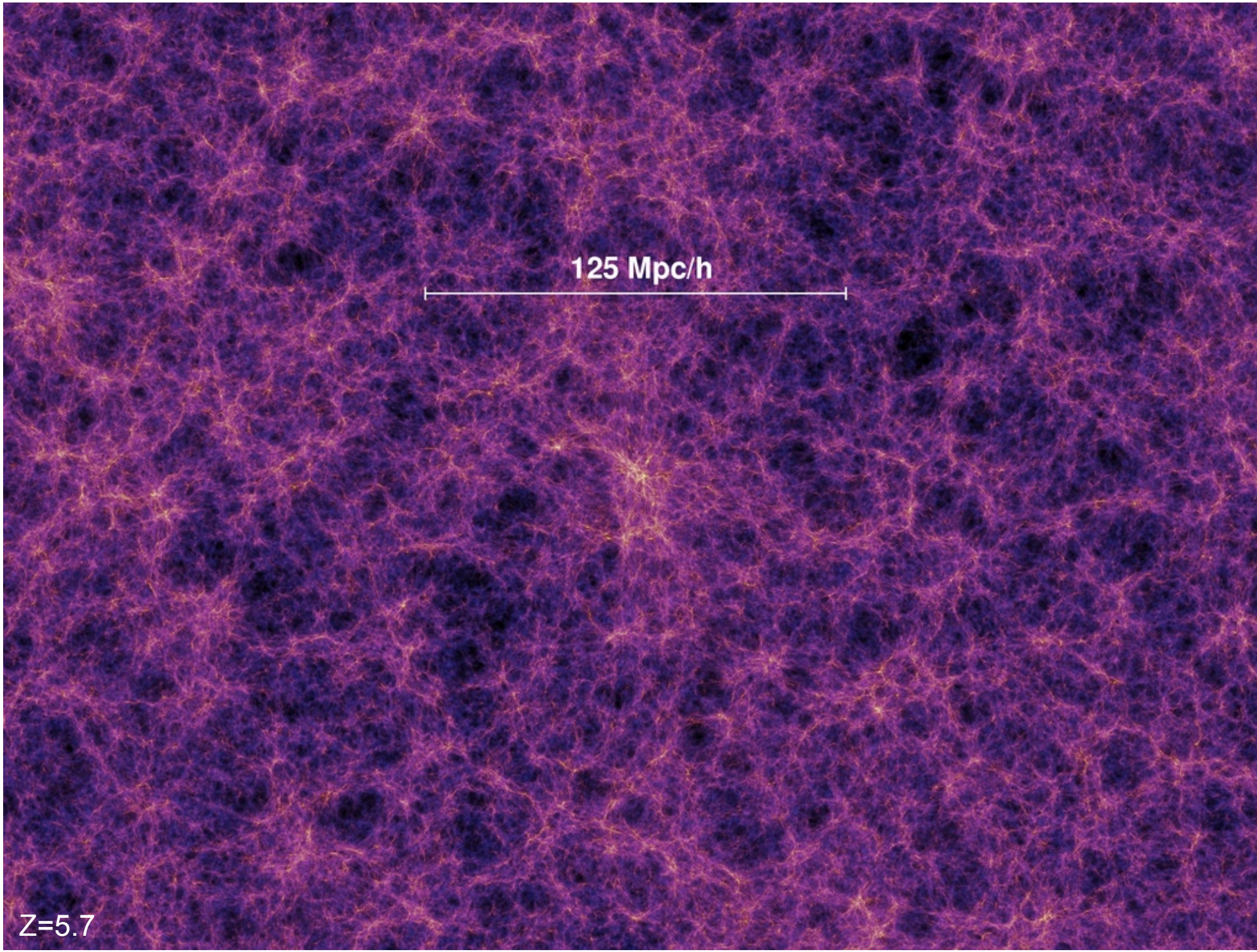
Millennium Run Observatory

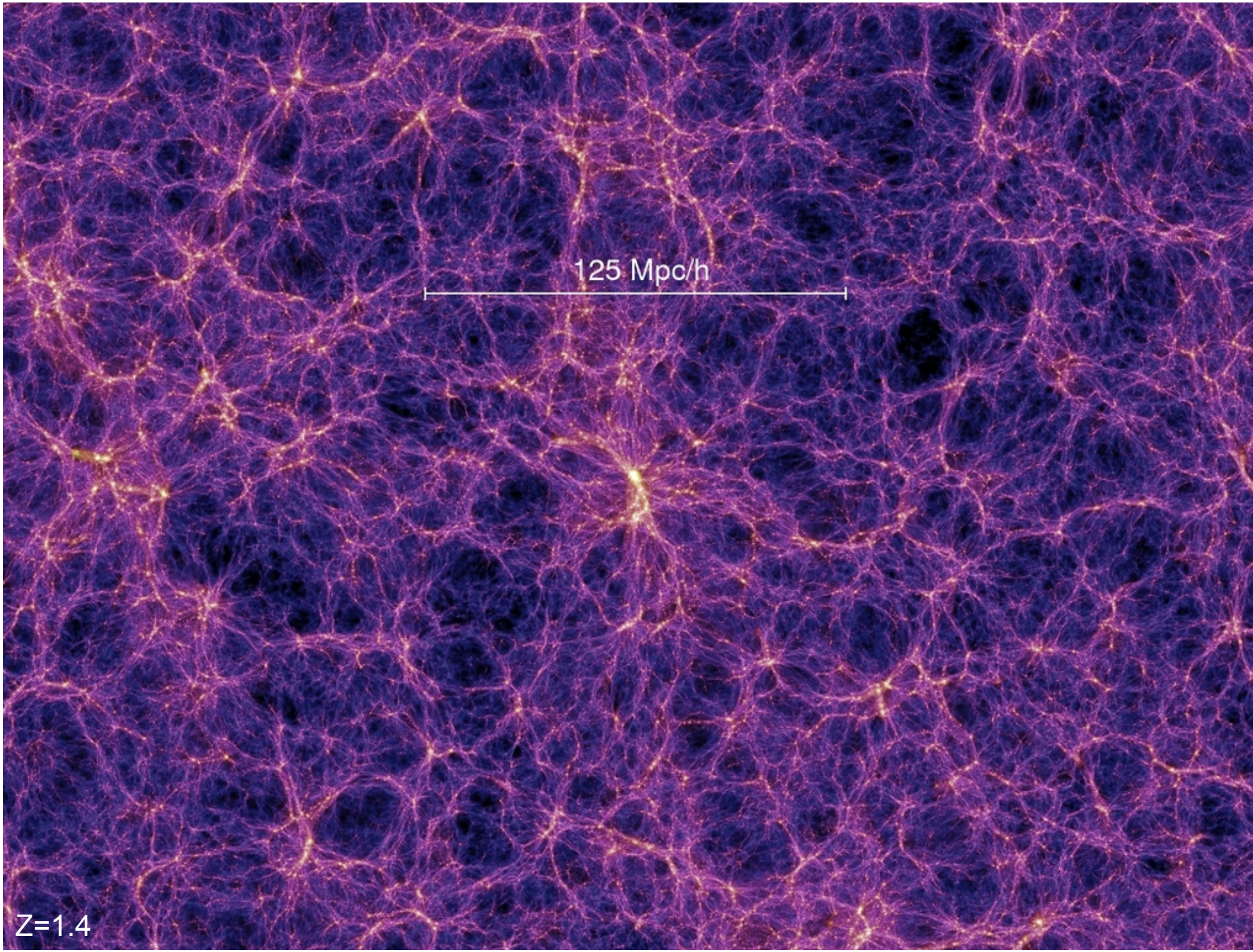


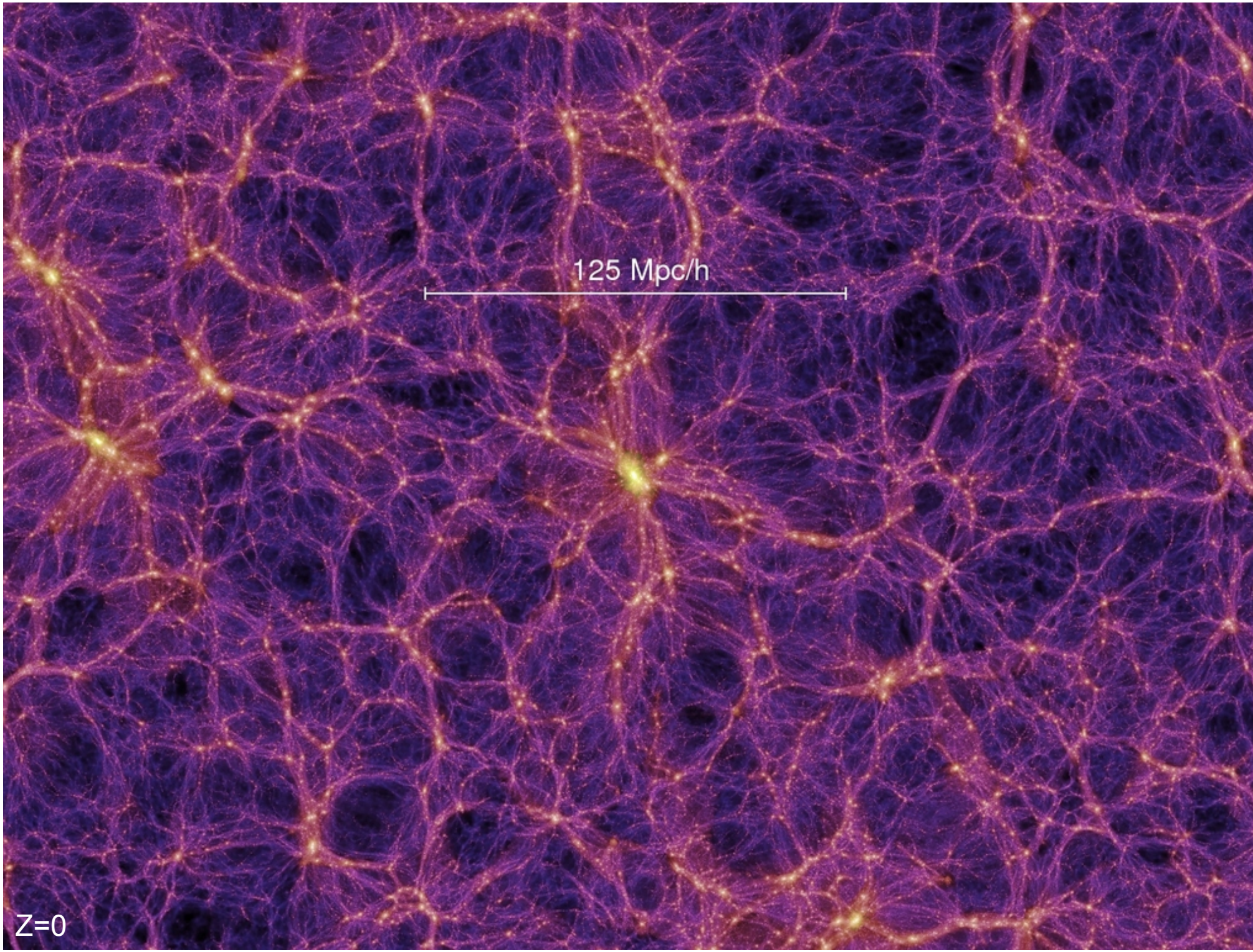
Millennium Run Observatory: Cosmological DM Simulations




125 Mpc/h
Z=18.3

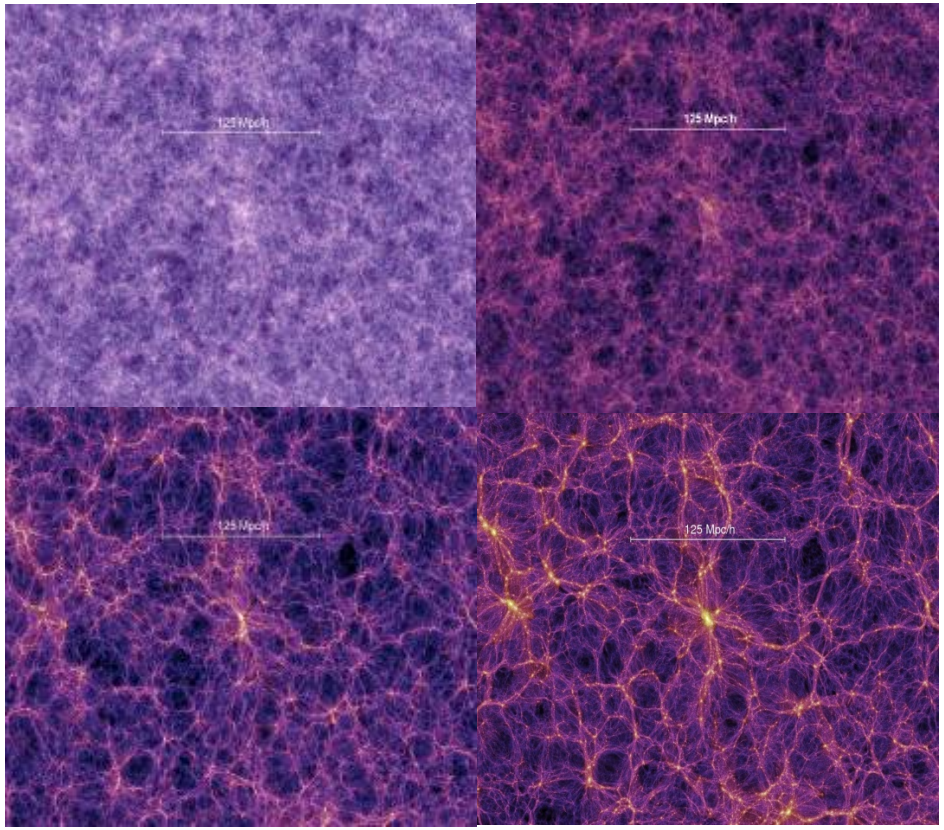

125 Mpc/h
Z=5.7


125 Mpc/h
Z=1.4


125 Mpc/h
Z=0

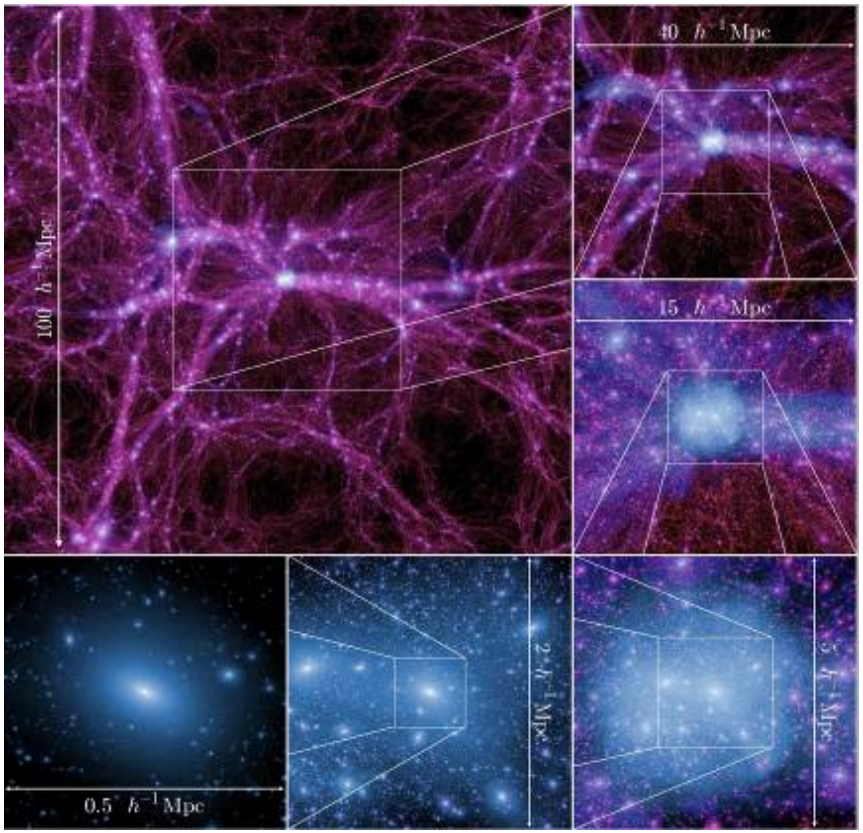
Max-Planck-Institut für Astrophysik

Millennium



V. Springel et al. 2005,
Nature 435, 629
500 Mpc/h, 10^{10} particles
64 snapshots

Millennium-II



M. Boylan-Kolchin et al. 2009,
MNRAS 398, 1150
100 Mpc/h, 10^{10} particles,
68 snapshots

~WMAP1 cosmology

Max-Planck-Institut für Astrophysik

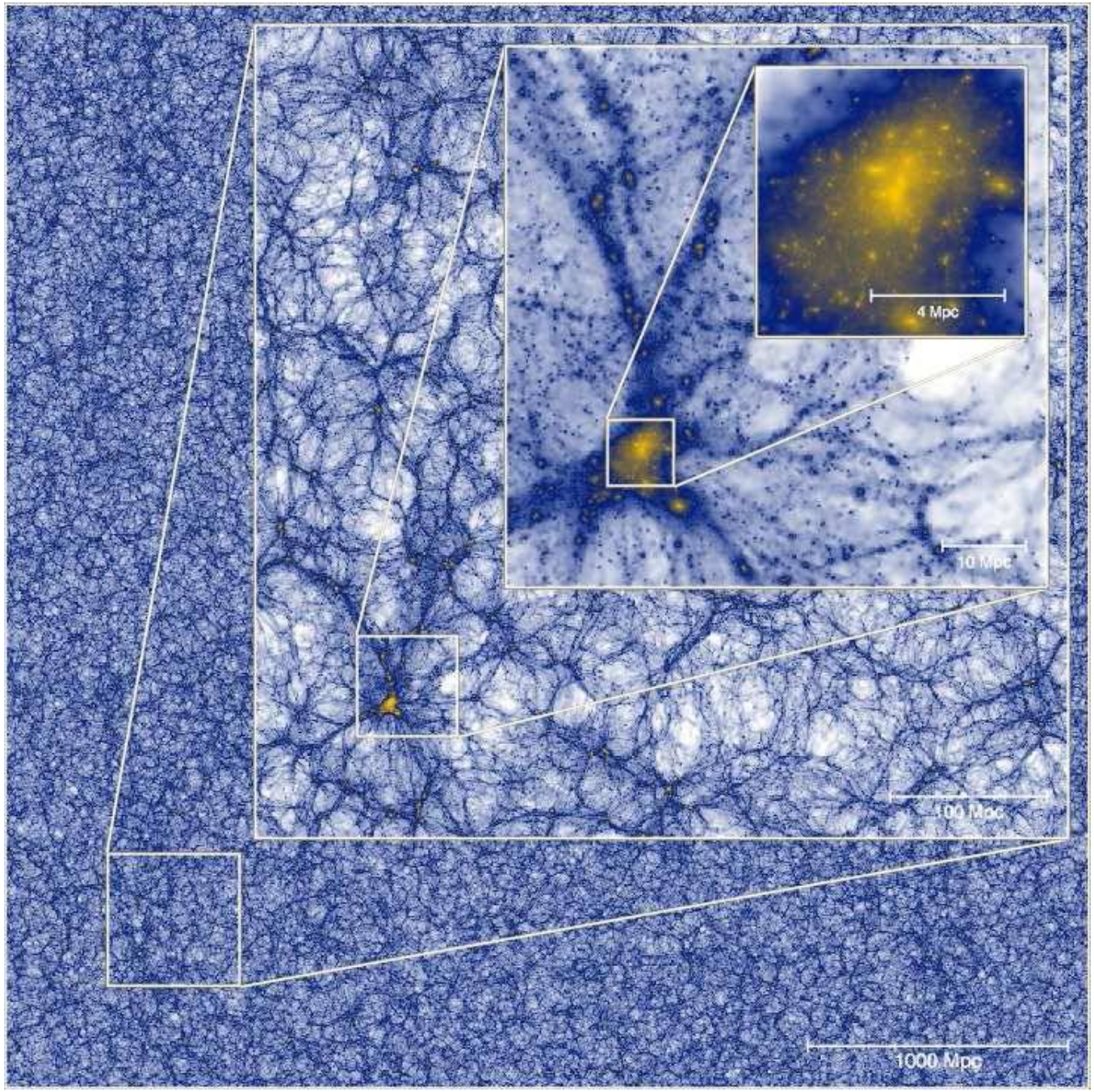
UC-HiPACC

Millennium XXL

3000 Mpc/h,
$3*10^{11}$ particles
64 snapshots
700TB → 70TB
Raw snapshots at
z=0, 0.3, 1.0, 2.0

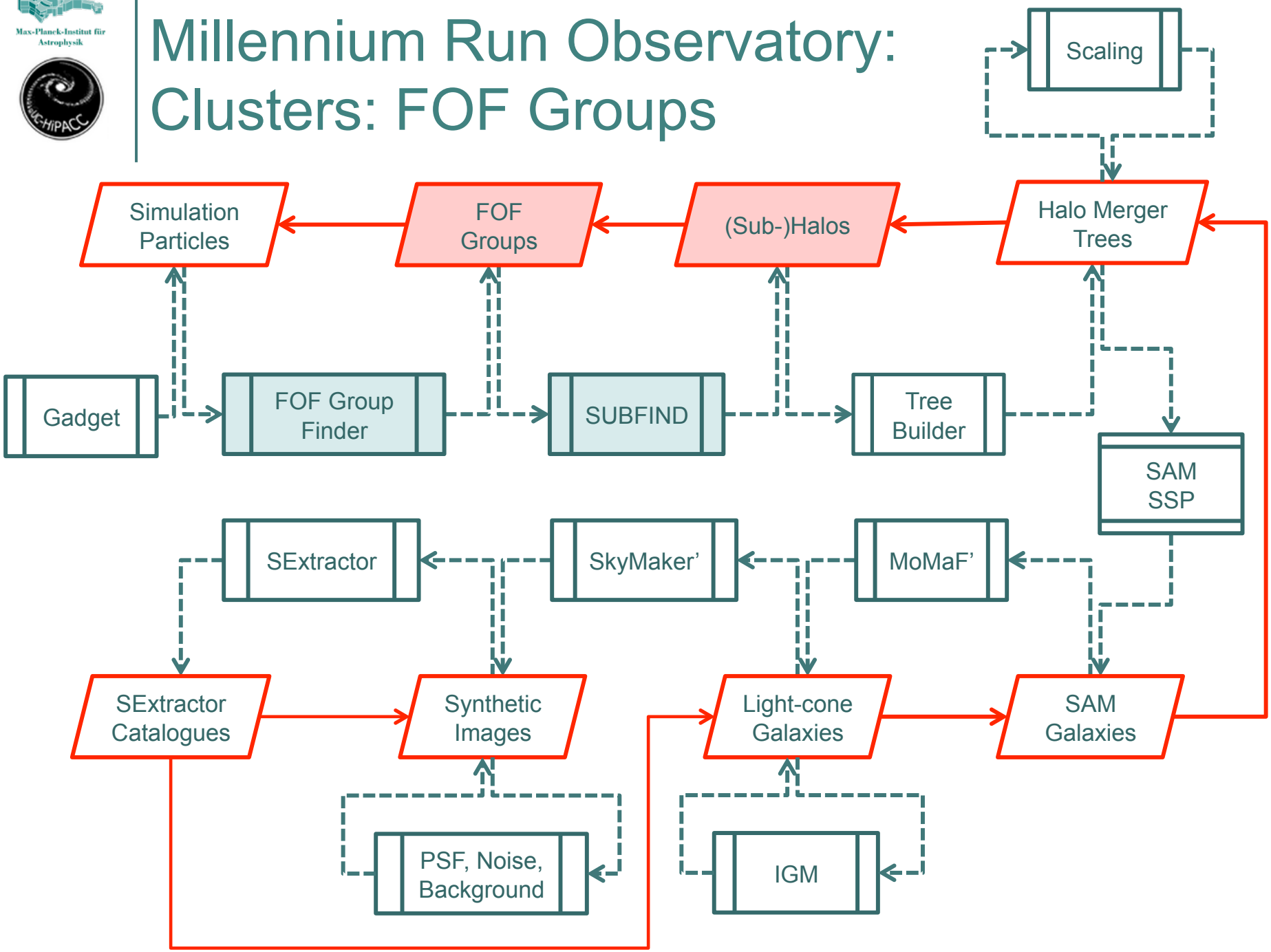
Angulo, R. *etal*, 2012
arXiv:1203.3216

(click for demo)



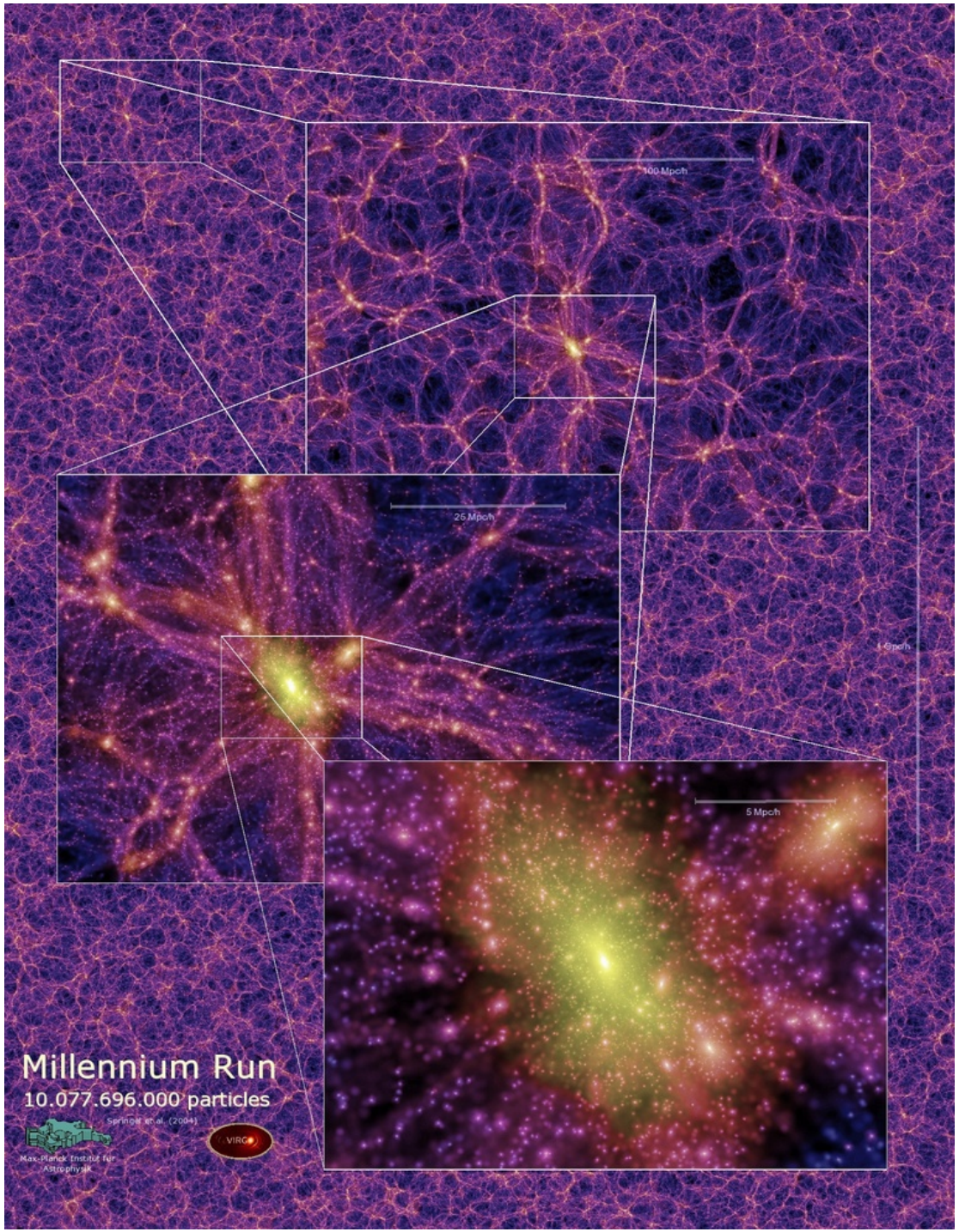
courtesy RaulAngulo

Millennium Run Observatory: Clusters: FOF Groups



Clusters

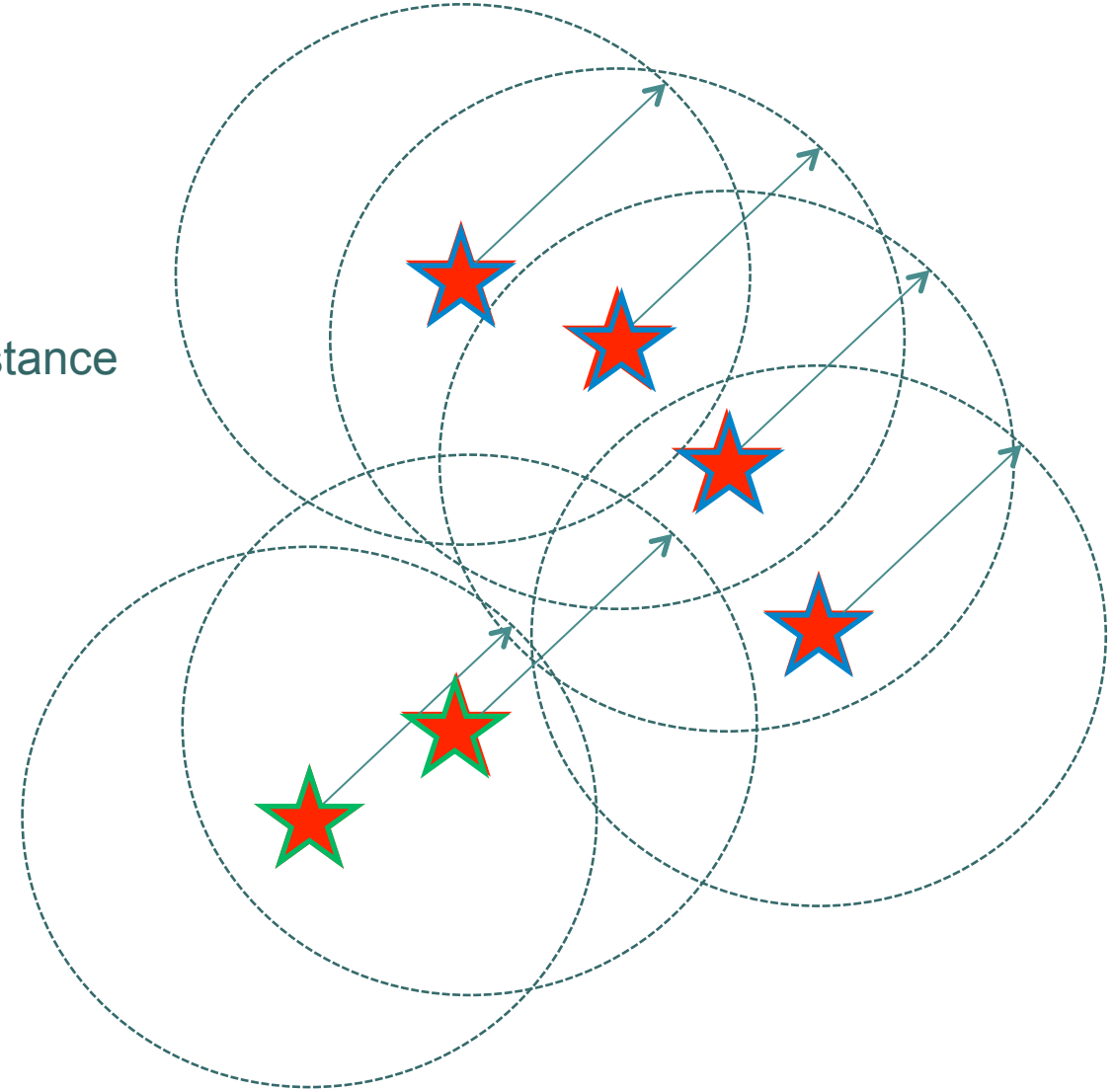
- Particles cluster and collapse to groups/ halos/clusters.
- Many cluster finder algorithms
- All: cluster=group of particles
- Here:
  FOF + SUBFIND
- See also lectures by Risa and Peter



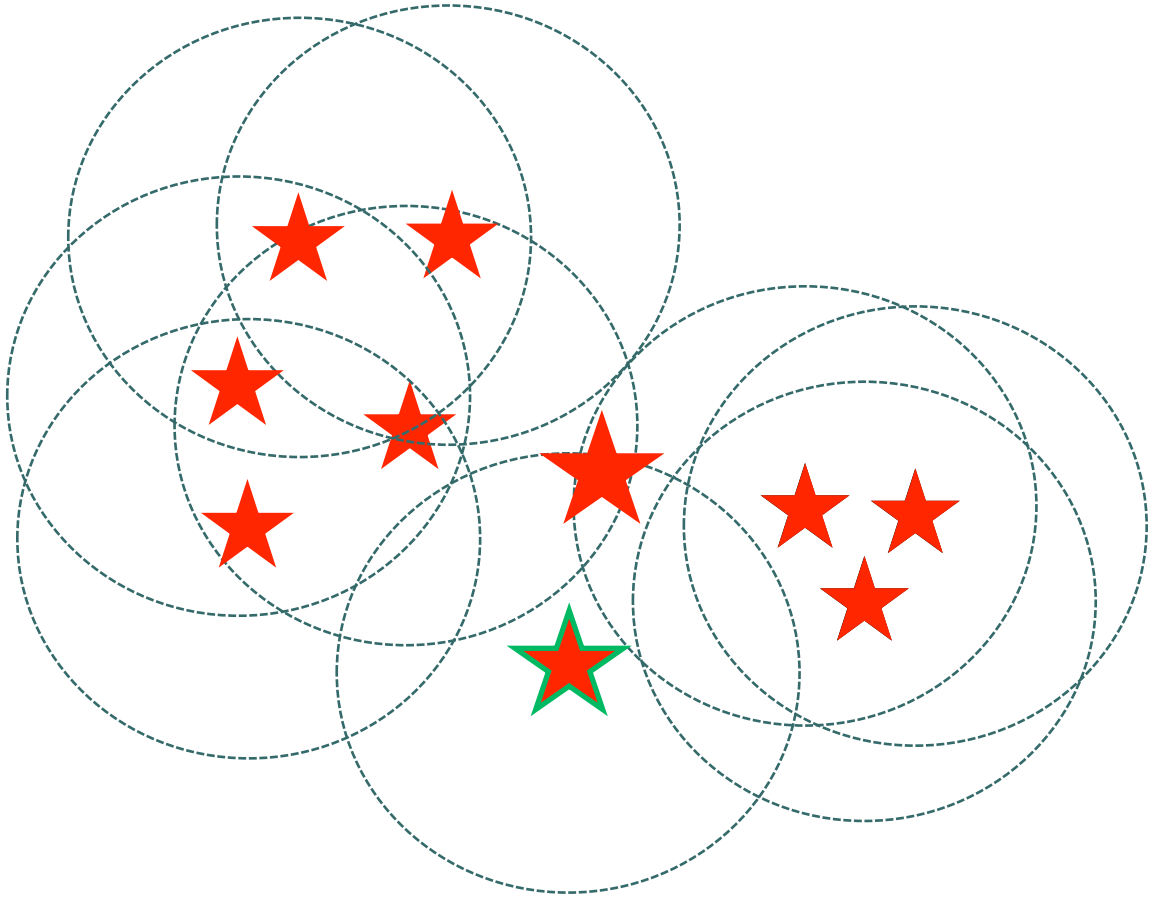
Friends-of-friends algorithm
Davis etal 1985

Linking length :
b=0.2 x mean inter-particle distance

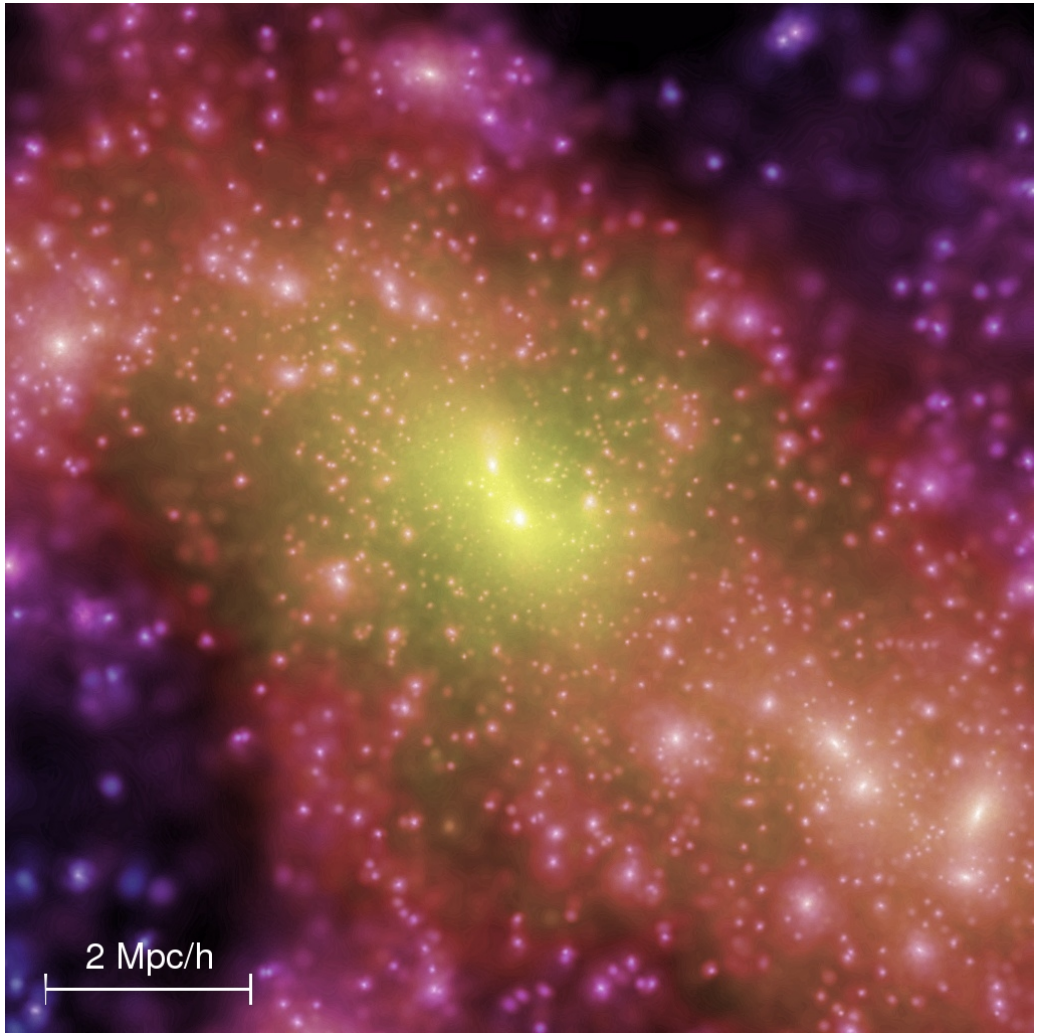
Minimum # particles/group
N_{min} = 20



FOF Groups, sometimes noisy



FOF groups have substructure

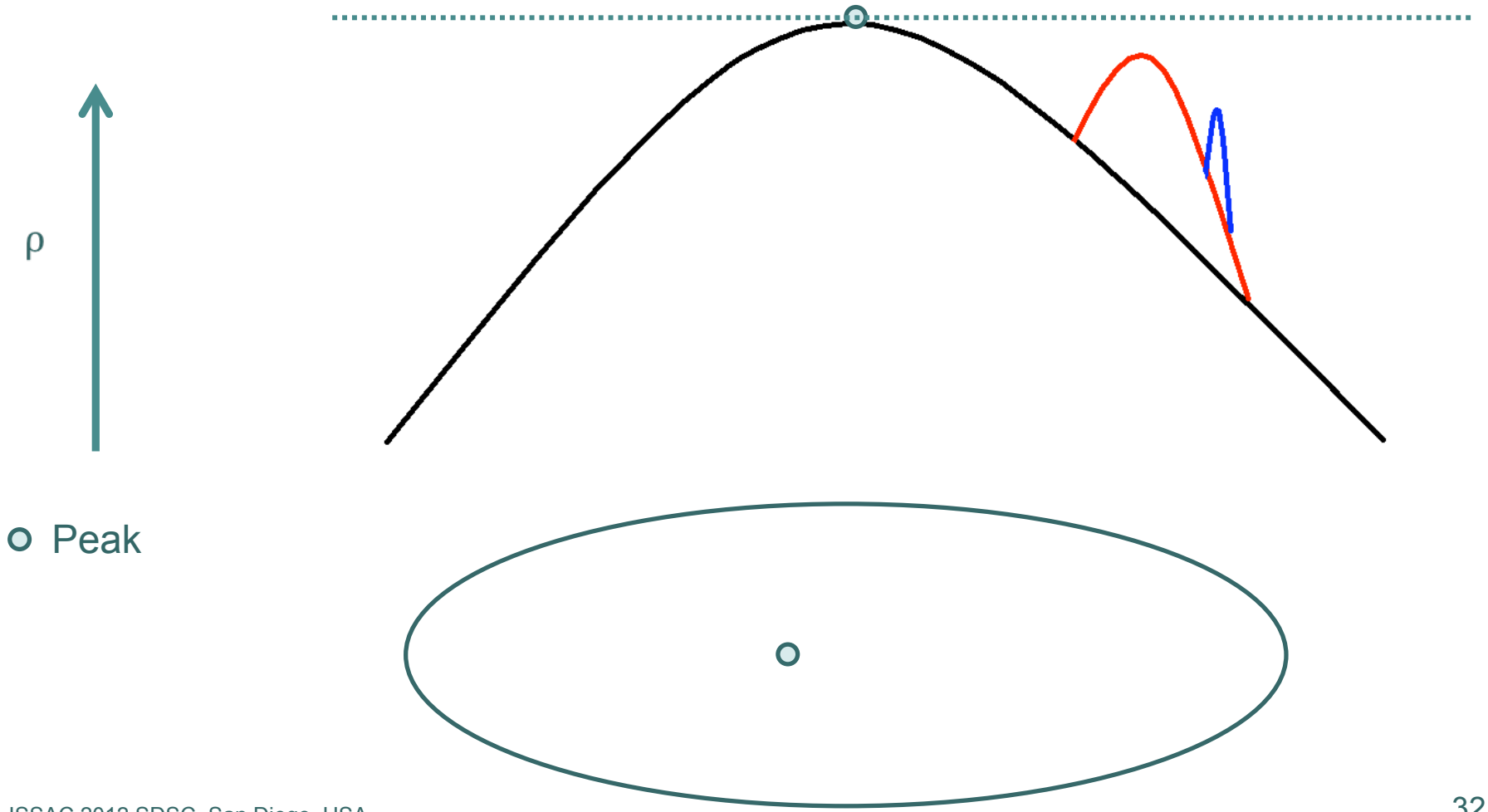


Detect sub-halos

- SUBFIND
  - Springel etal 2001
- One possible algorithm to detect sub-halos
  - See Risa/Peter for alternatives
- Gravitationally bound (more or less)
- Basis of semi-analytical galaxy modeling

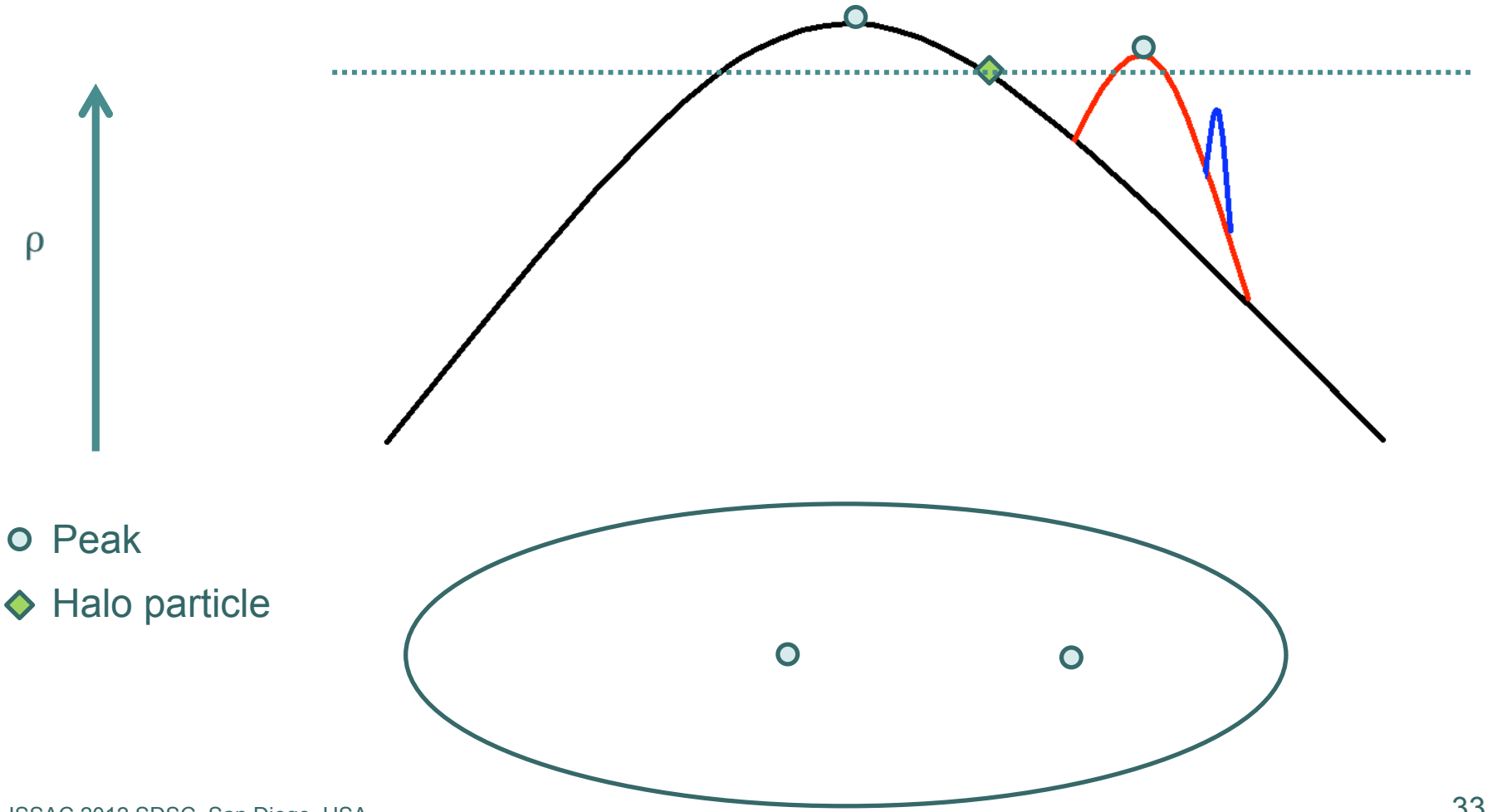
SUBFIND algorithm: Selection of sub-halo candidates

slides courtesy Francesca Iannuzzi



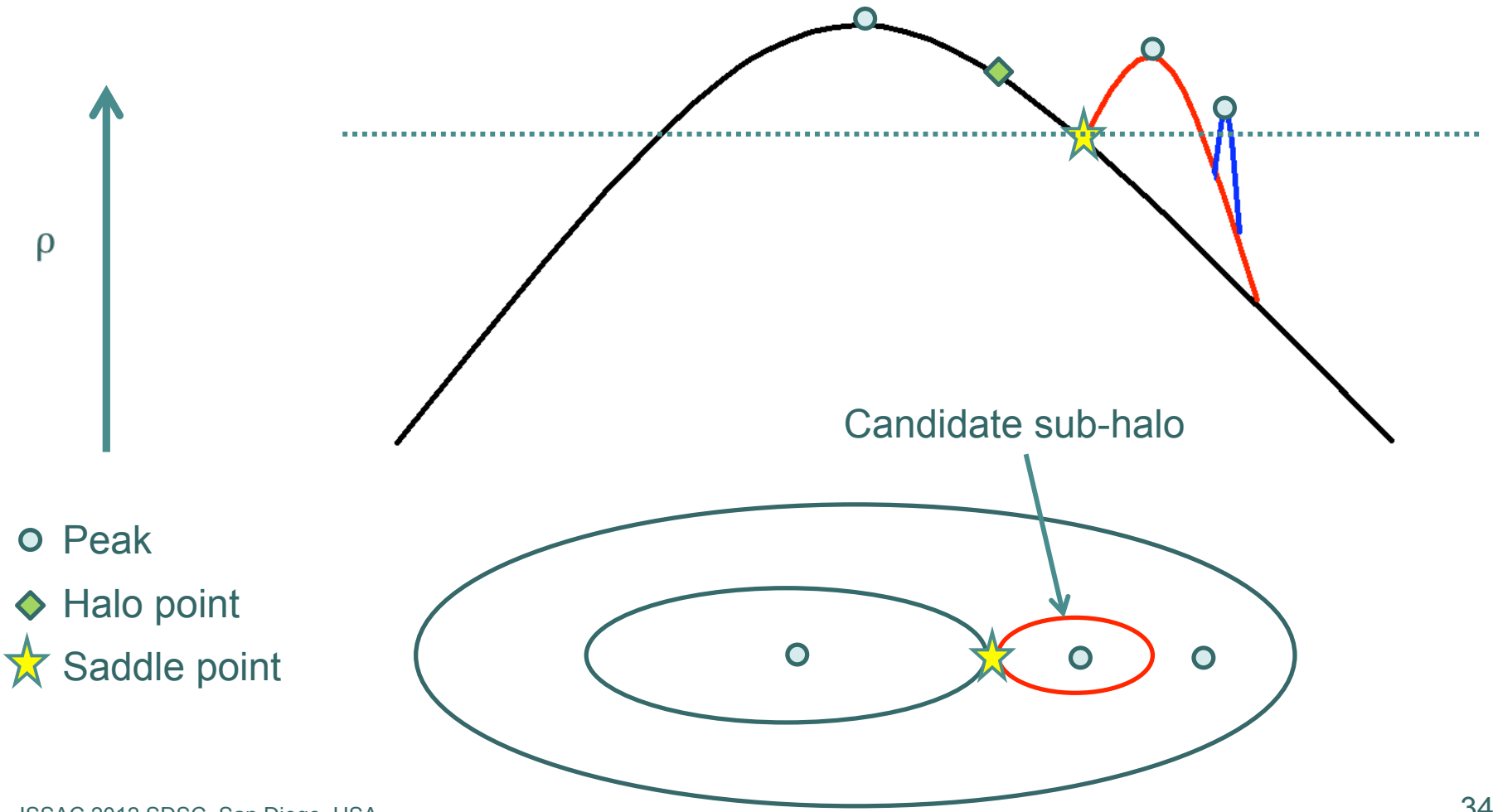
SUBFIND algorithm: Selection of sub-halo candidates

slides courtesy Francesca Iannuzzi



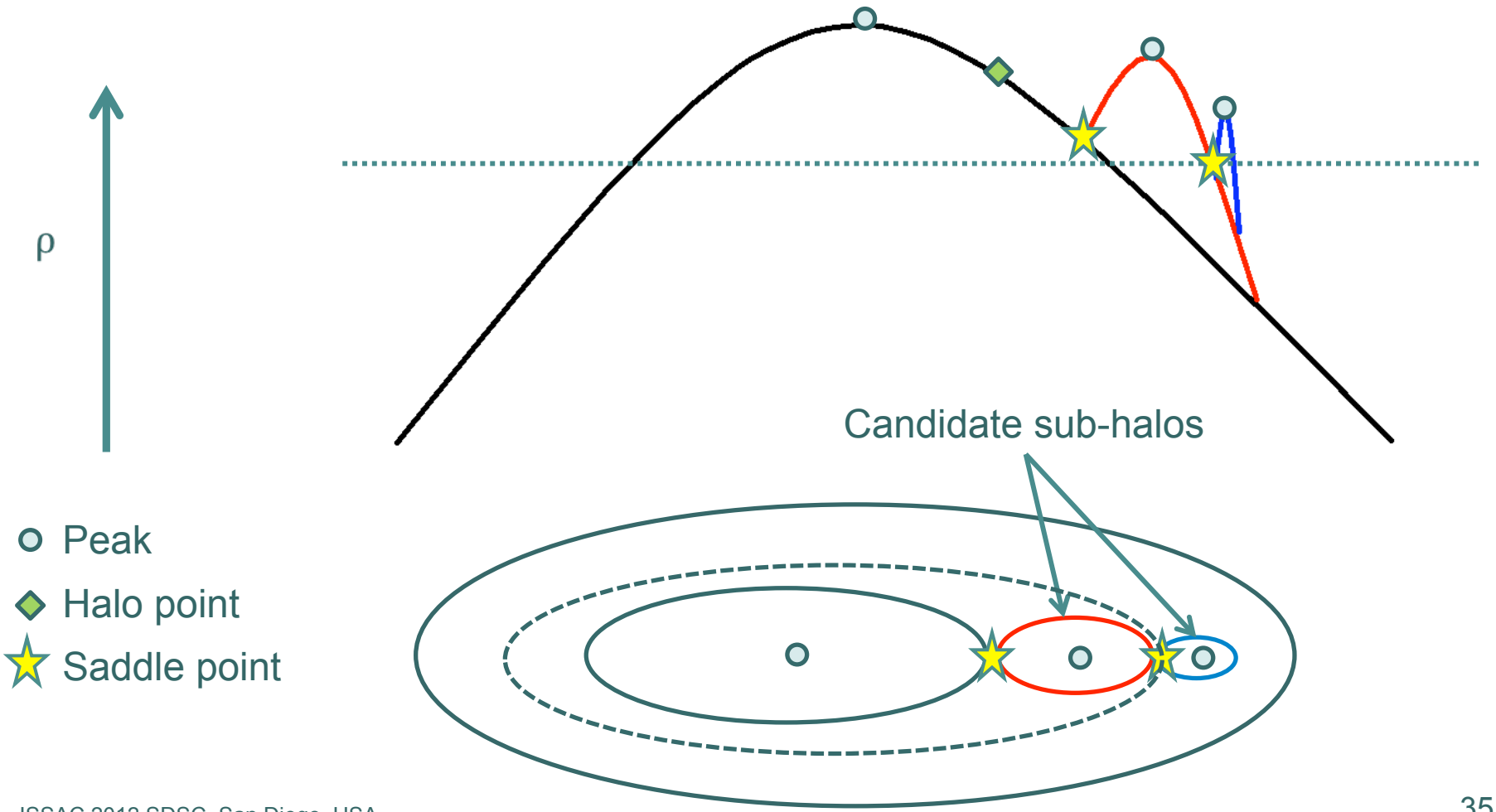
SUBFIND algorithm: Selection of sub-halo candidates

slides courtesy Francesca Iannuzzi



SUBFIND algorithm: Selection of sub-halo candidates

slides courtesy Francesca Iannuzzi



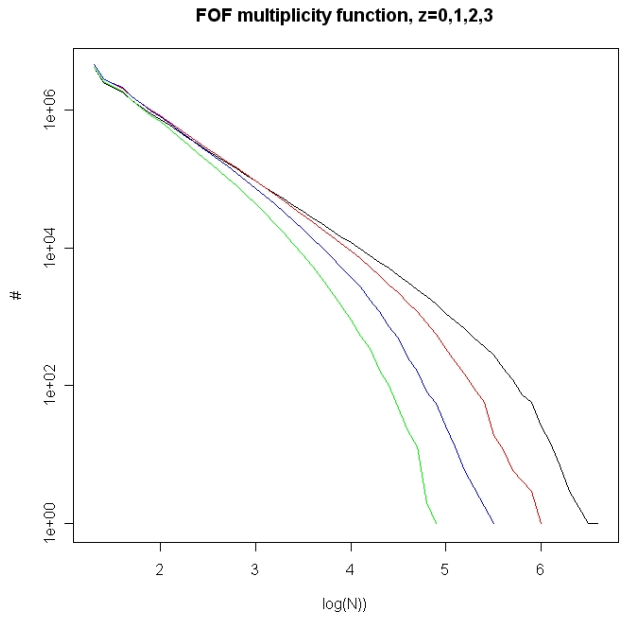
SUBFIND algorithm: Unbinding

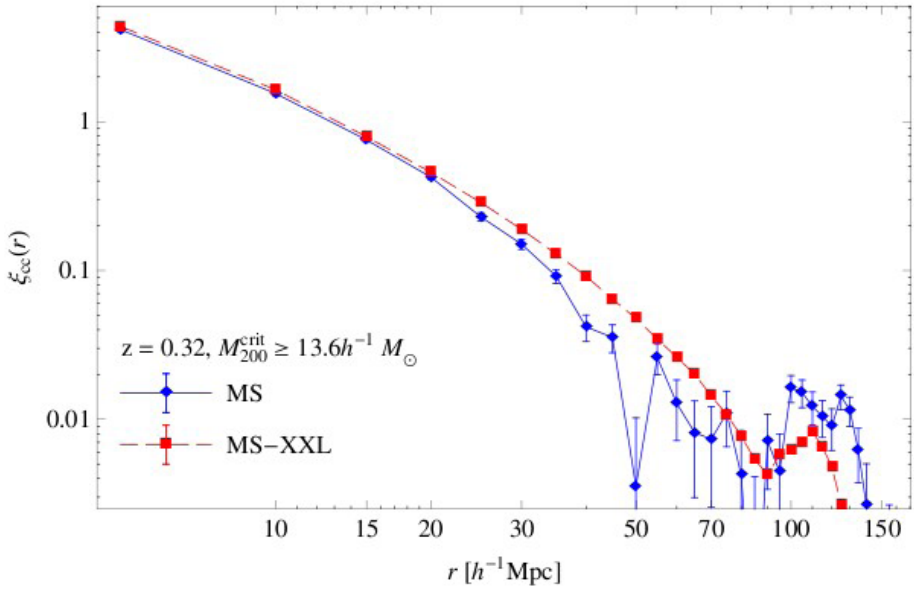
- Process the candidates for self-boundedness
- For every particle compute:
  - potential energy (*positions in physical units*)
  - kinetic energy *(velocities wrt bulk velocity + Hubble flow, physical units)*
- Total energy > 0 ⇨particle removed from the candidate
  - Iteratively - no more than 1/4 of the particles removed at once

slides courtesy Francesca Iannuzzi

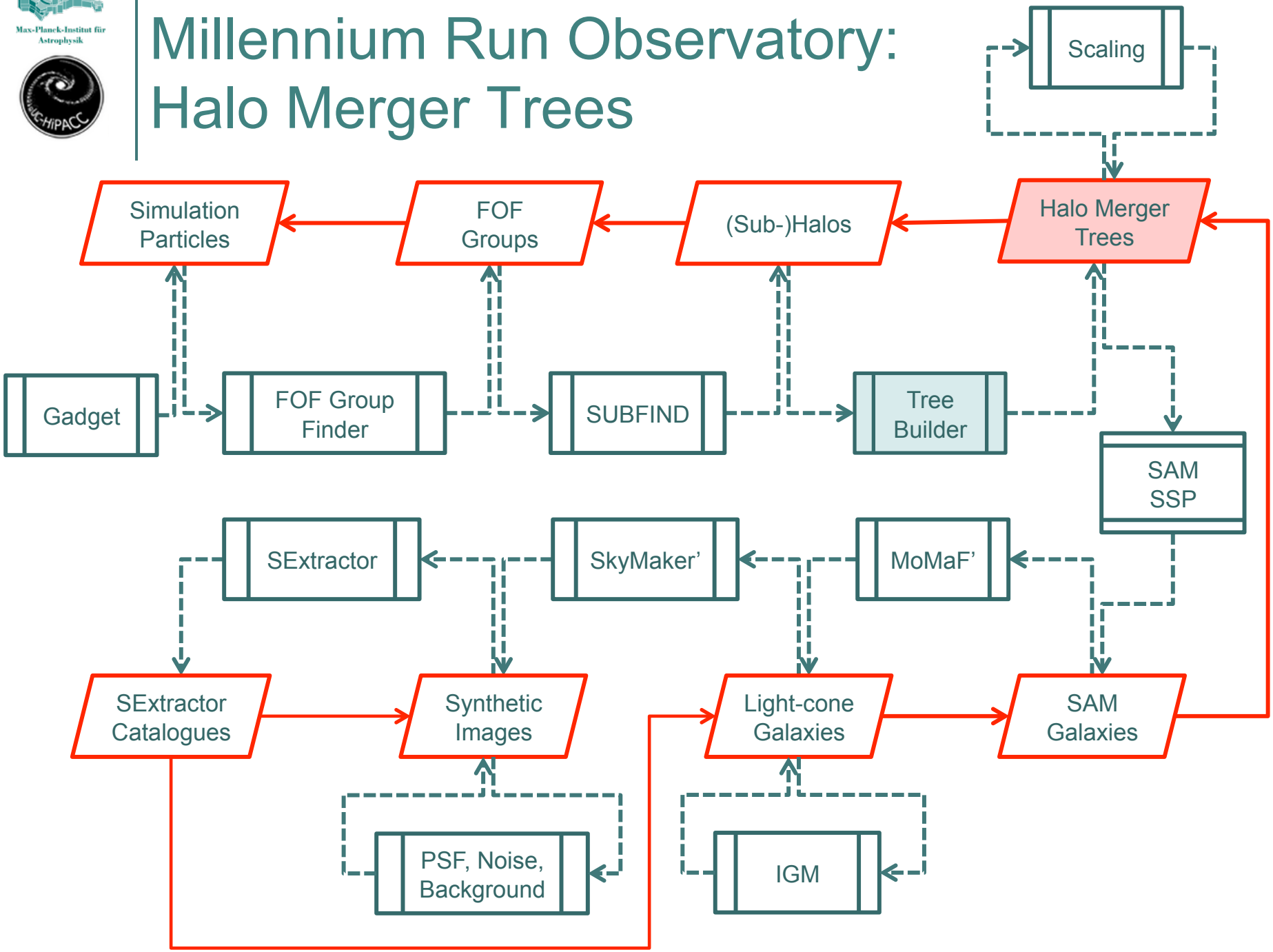
Millennium simulations

- Millennium
  - ~18 million/snapshot @ z=0
  - Total ~800 million
- Millennium-II
  - ~12 million @ z=0
  - Total ~680 million
- Millennium XXL @ z=0
  - FOF ~650 million
  - SH ~ 720 million





Millennium Run Observatory: Halo Merger Trees

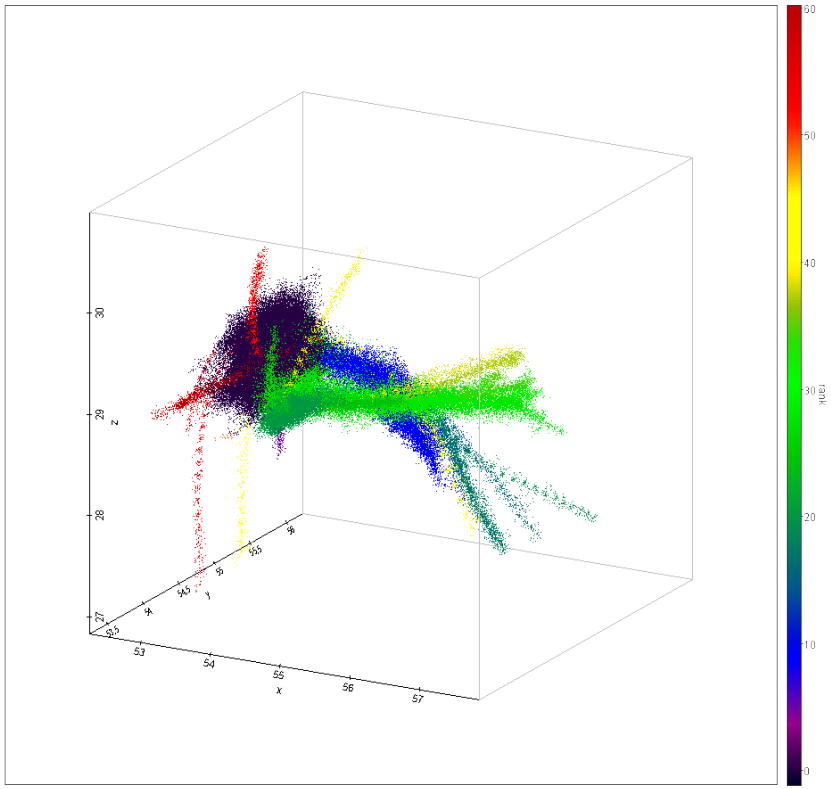


(sub-)halo merger trees

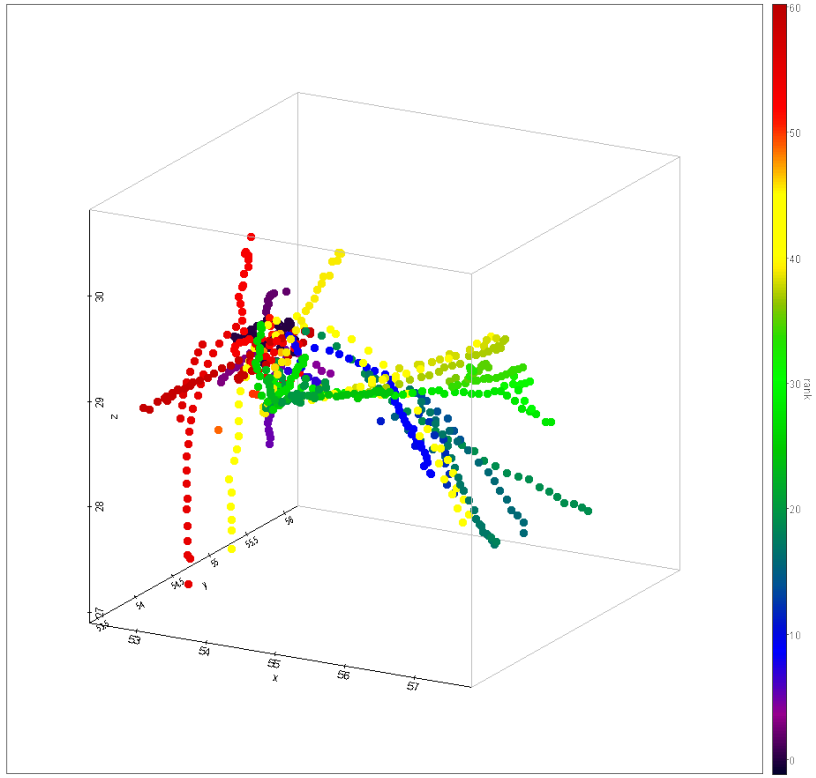
- Evolution of halos:
  - Birth: first time a halo has >=20 particles
  - Growth by
    - Accretion
    - Merging
  - Merge onto other halo
- Also
  - Dissolve
  - Get lost for a while (too close to other subhalo)
  - Are misidentified

Max-Planck-Institut für Astrophysik

Halo evolution: accretion and mergers



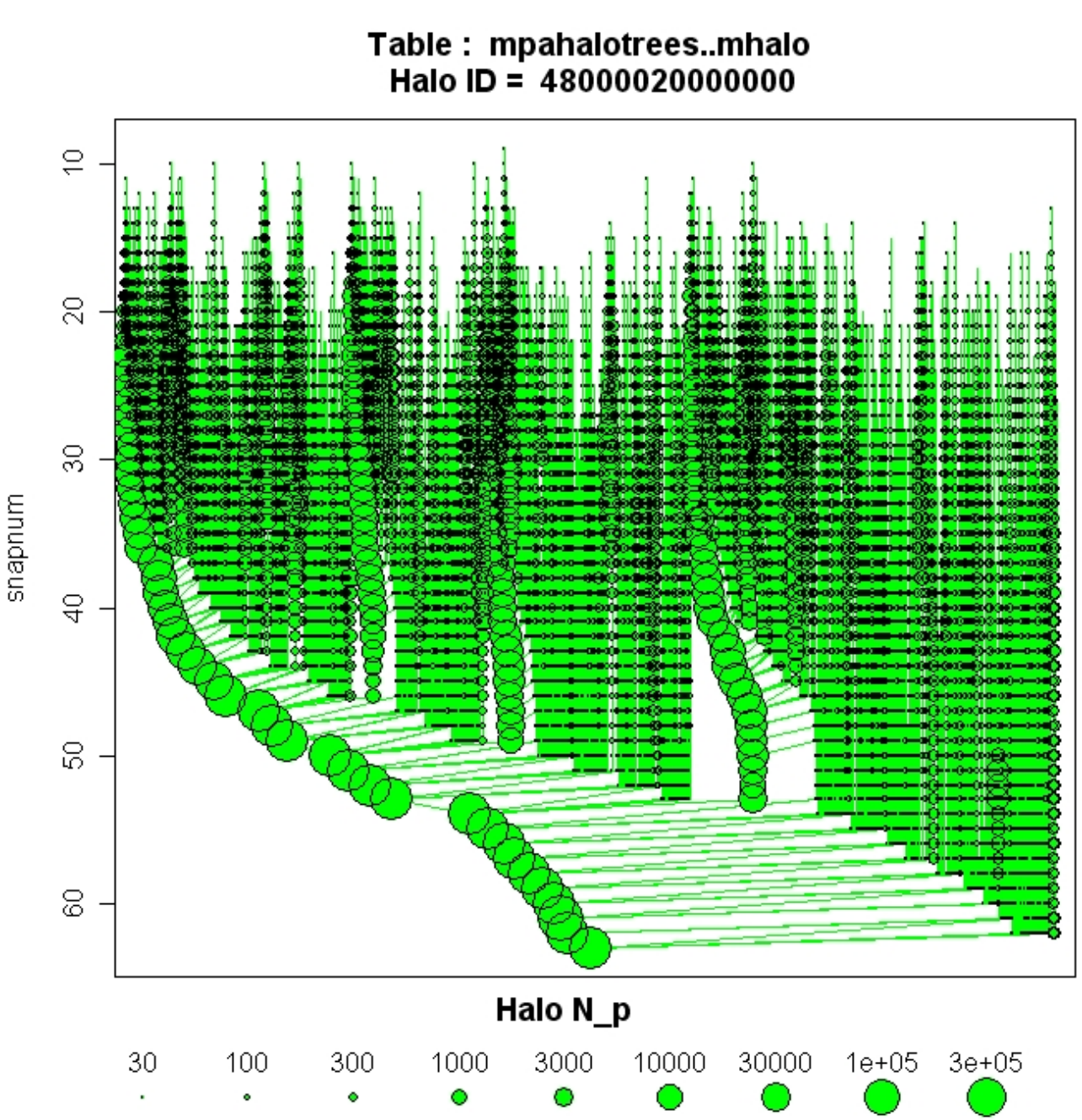
particles



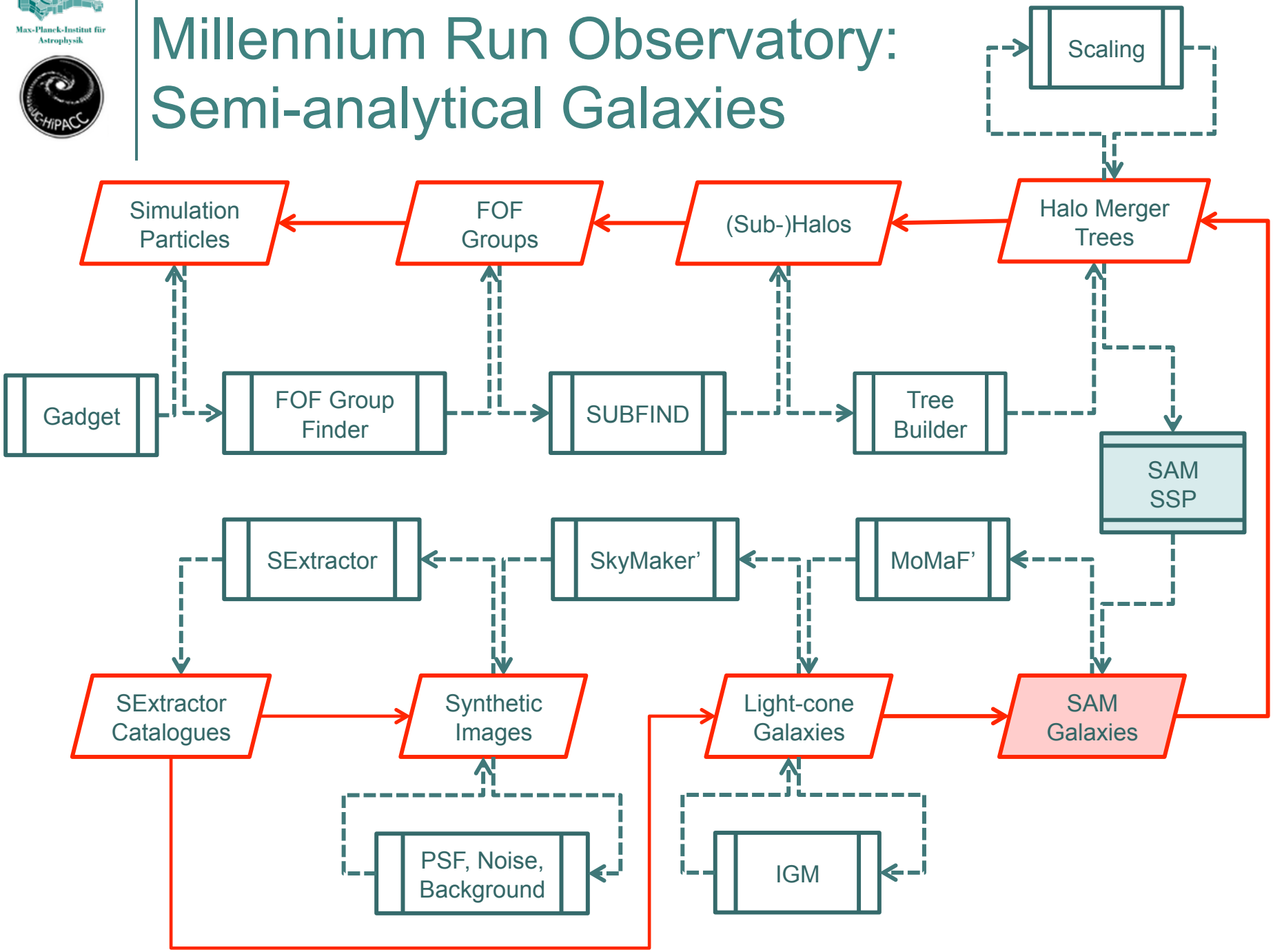
halos

(see TOPCAT)

Max-Planck-Institut für Astrophysik



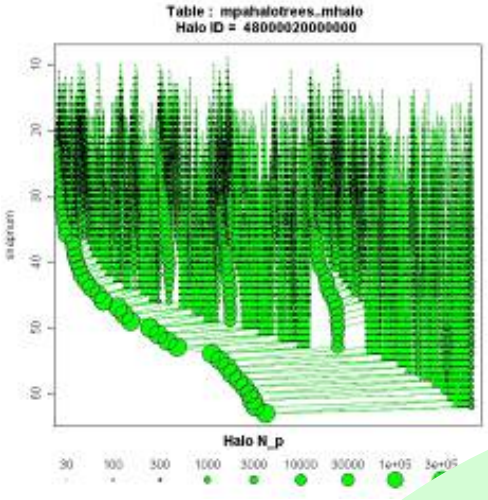
Millennium Run Observatory: Semi-analytical Galaxies

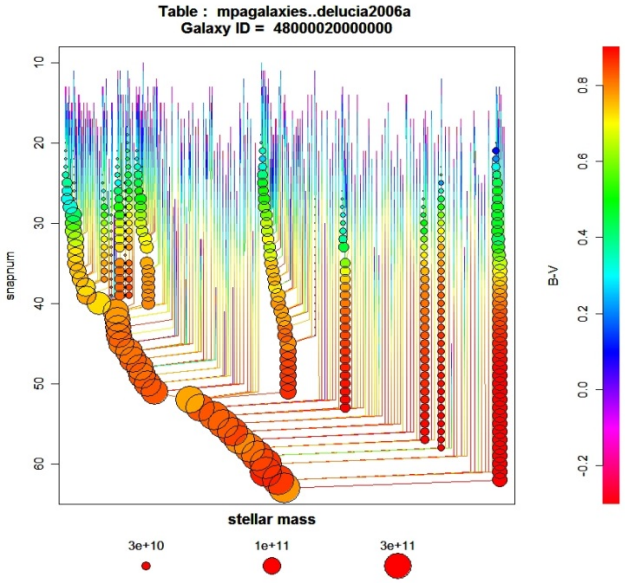


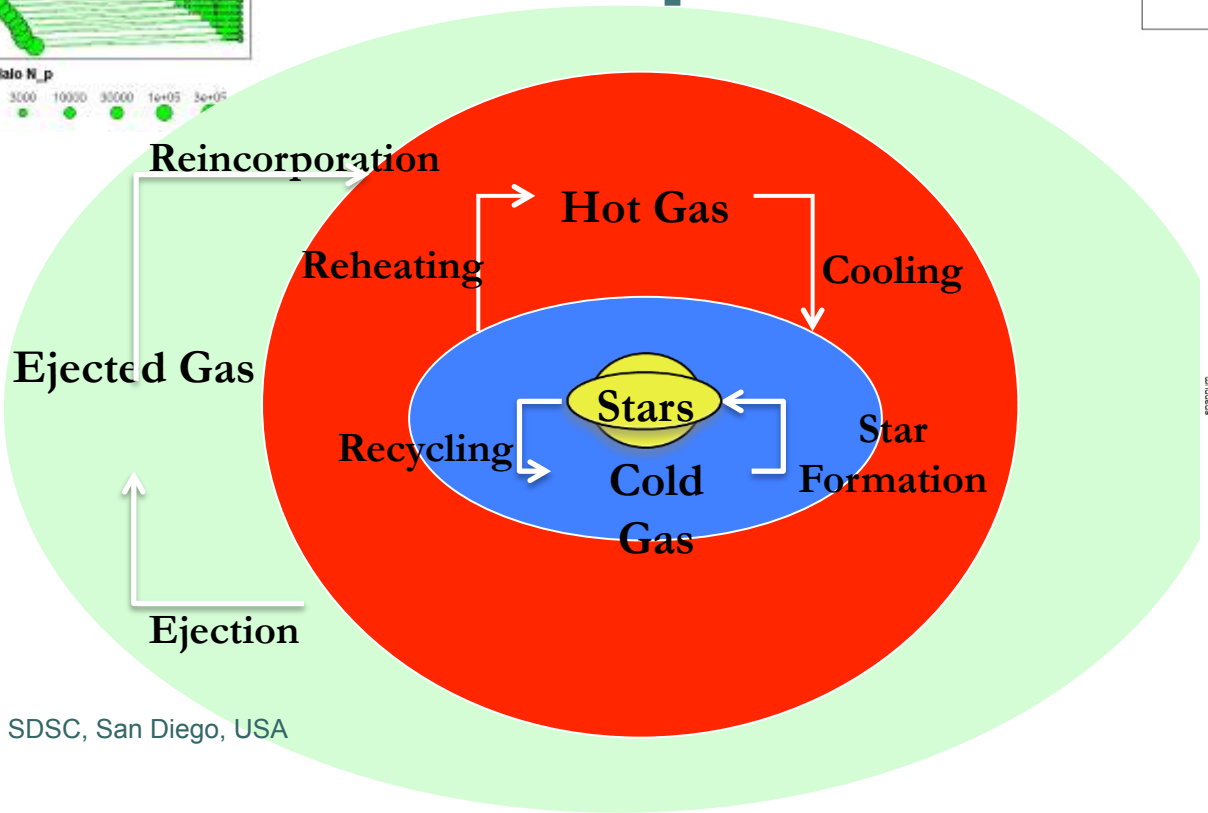
Semi-analytical models for galaxy formation (SAMs)

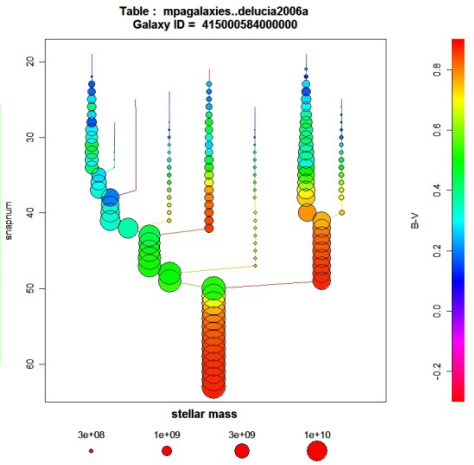
- SAM:
  - Add baryonic physics *after* gravitational evolution on skeleton of halo merger trees
  - Advantage: can sample parameter space more efficiently
  - Disadvantage: no feedback baryonic physics on clustering.
- L-Galaxies: MPA's semi-analytical models (SAM) for galaxy formation
  - Others: Benson, Croton, Durham , GalICS, Somerville, … (etal)
- See lectures by Darren Croton for **all** the details

SAMs: baryons in halo merger trees





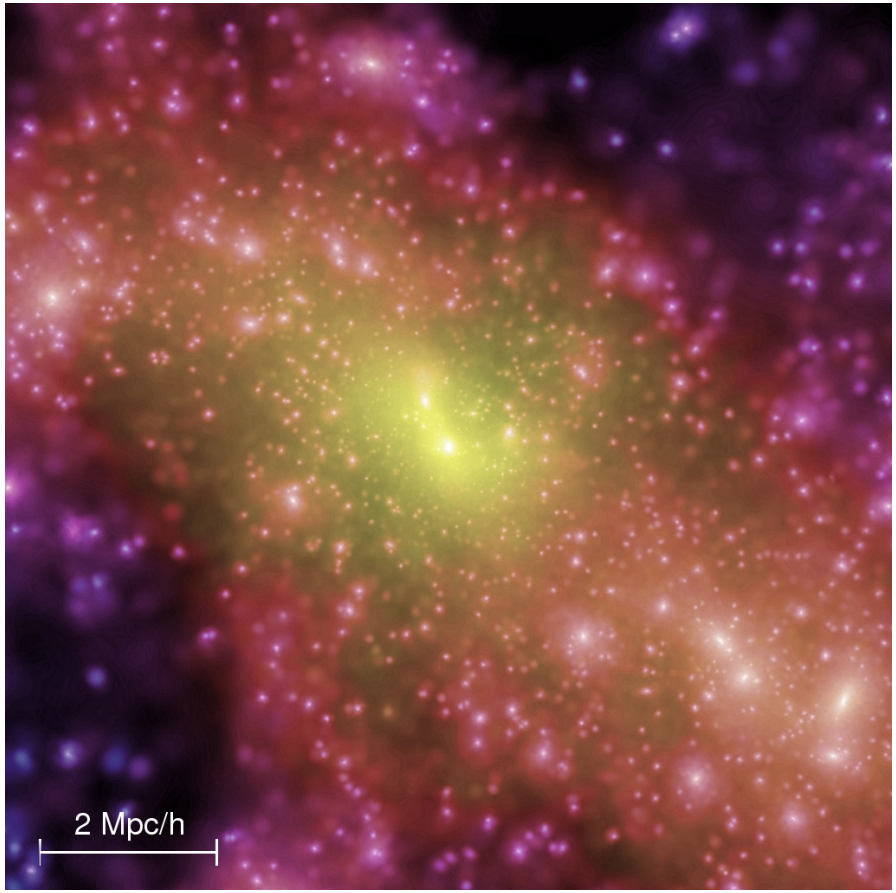


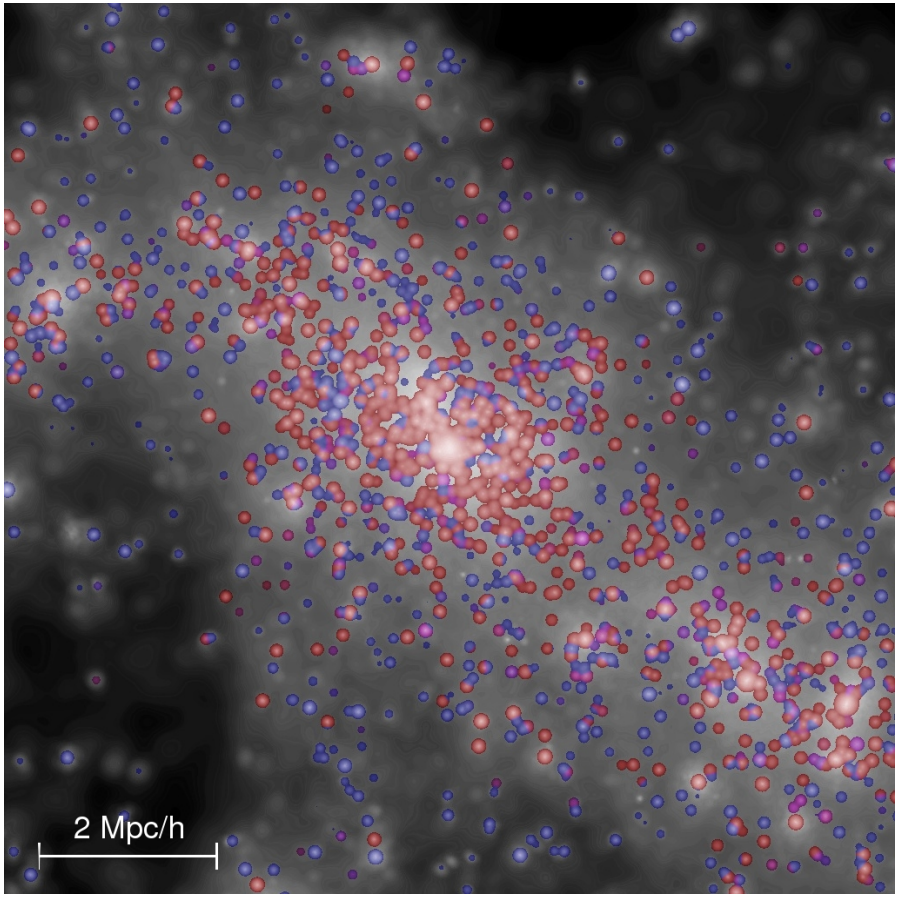


courtesy Bruno Henriques

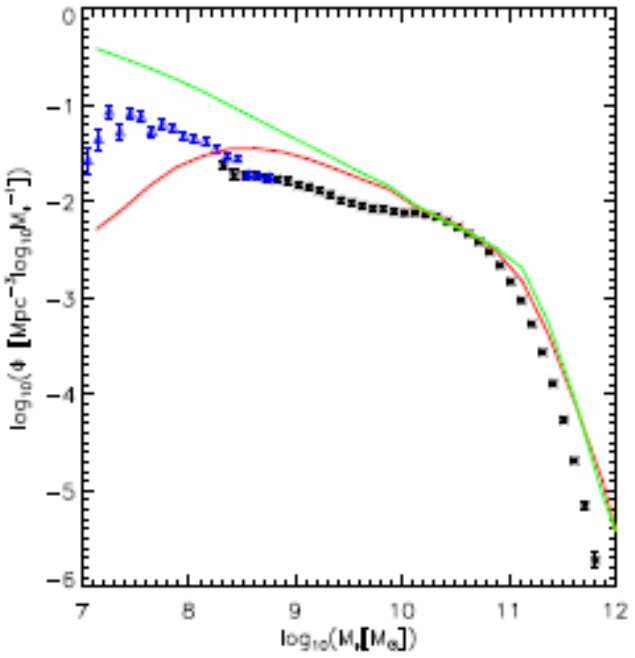
Max-Planck-Institut für Astrophysik

Halos ⇨ Galaxies

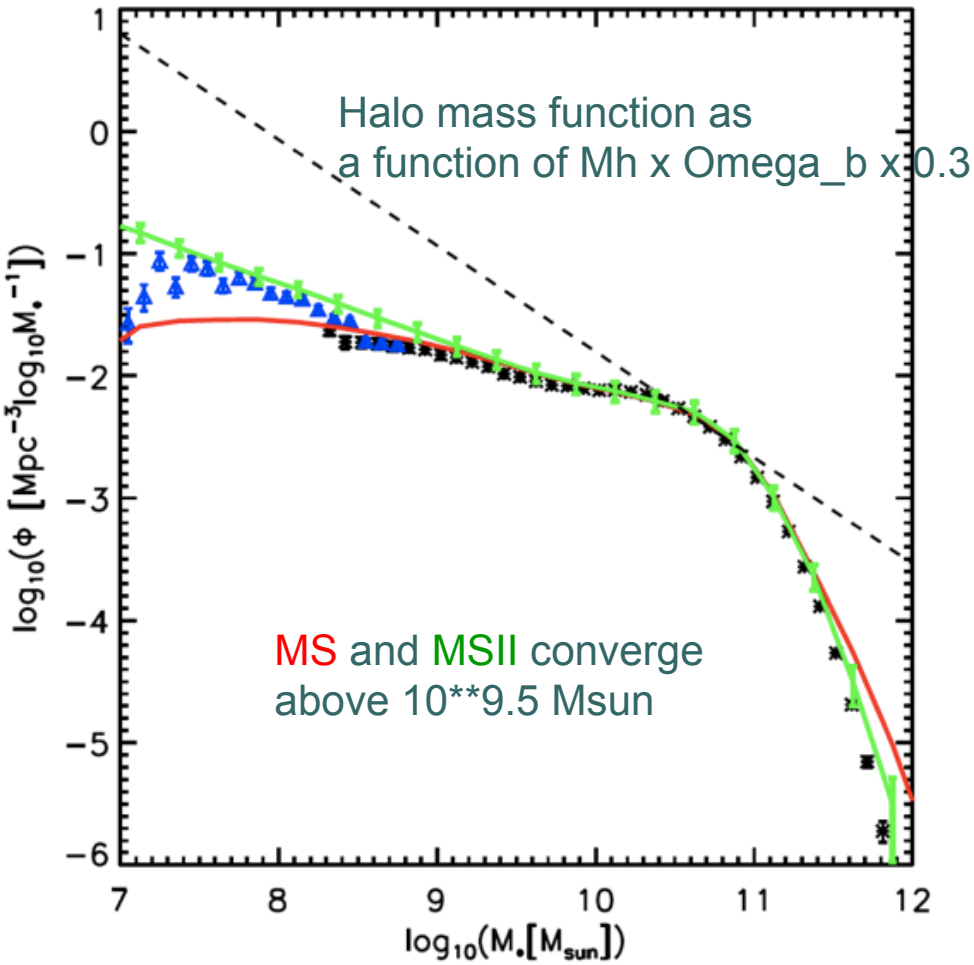




Latest models



DeLucia & Blaizot, 2006

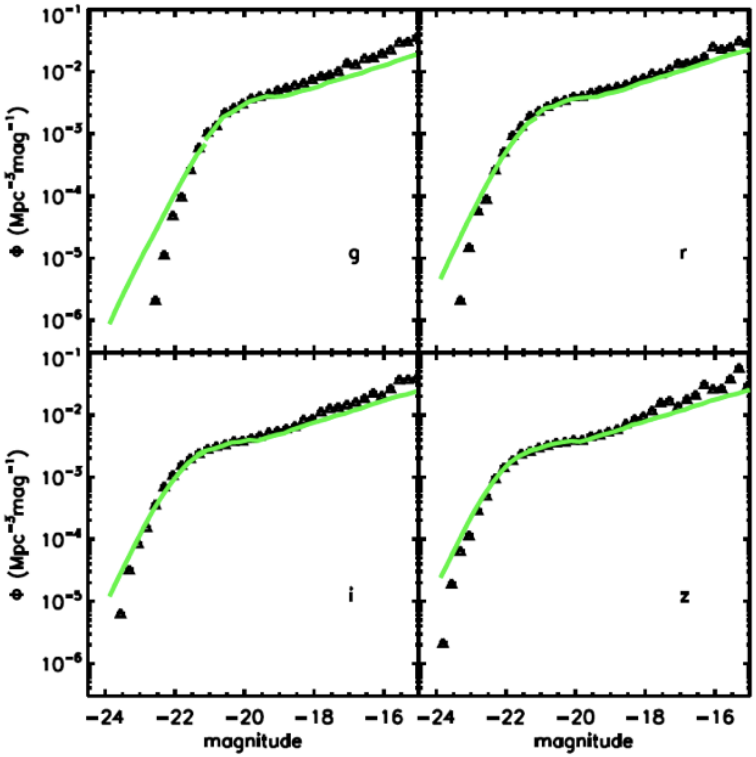


Guo etal, 2011

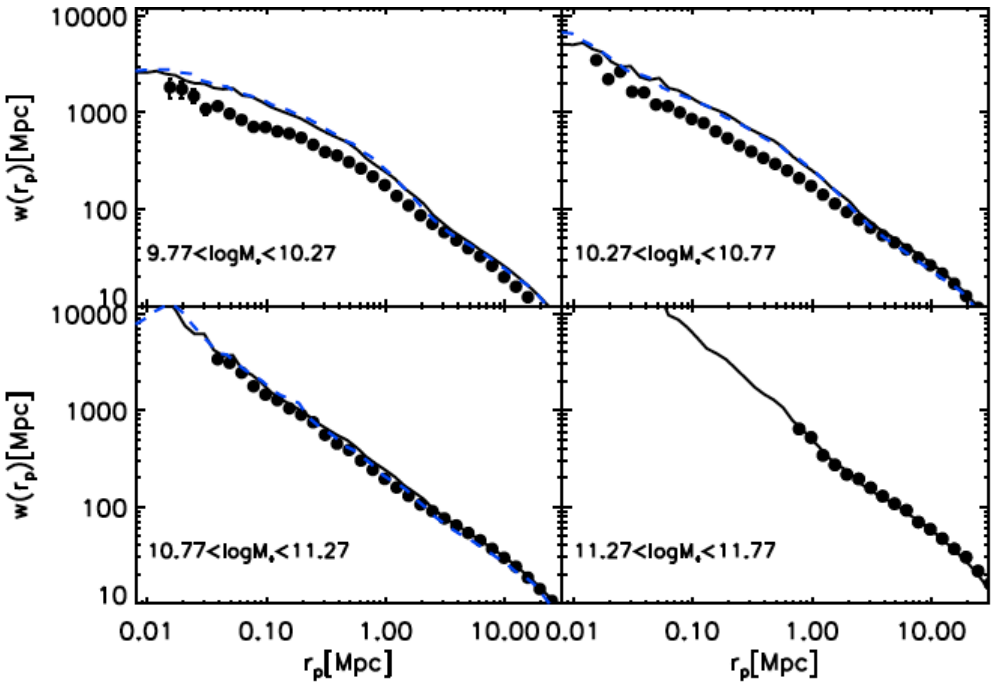
courtesy Qi Guo

Max-Planck-Institut für Astrophysik

Luminosity Functions



Correlation functions

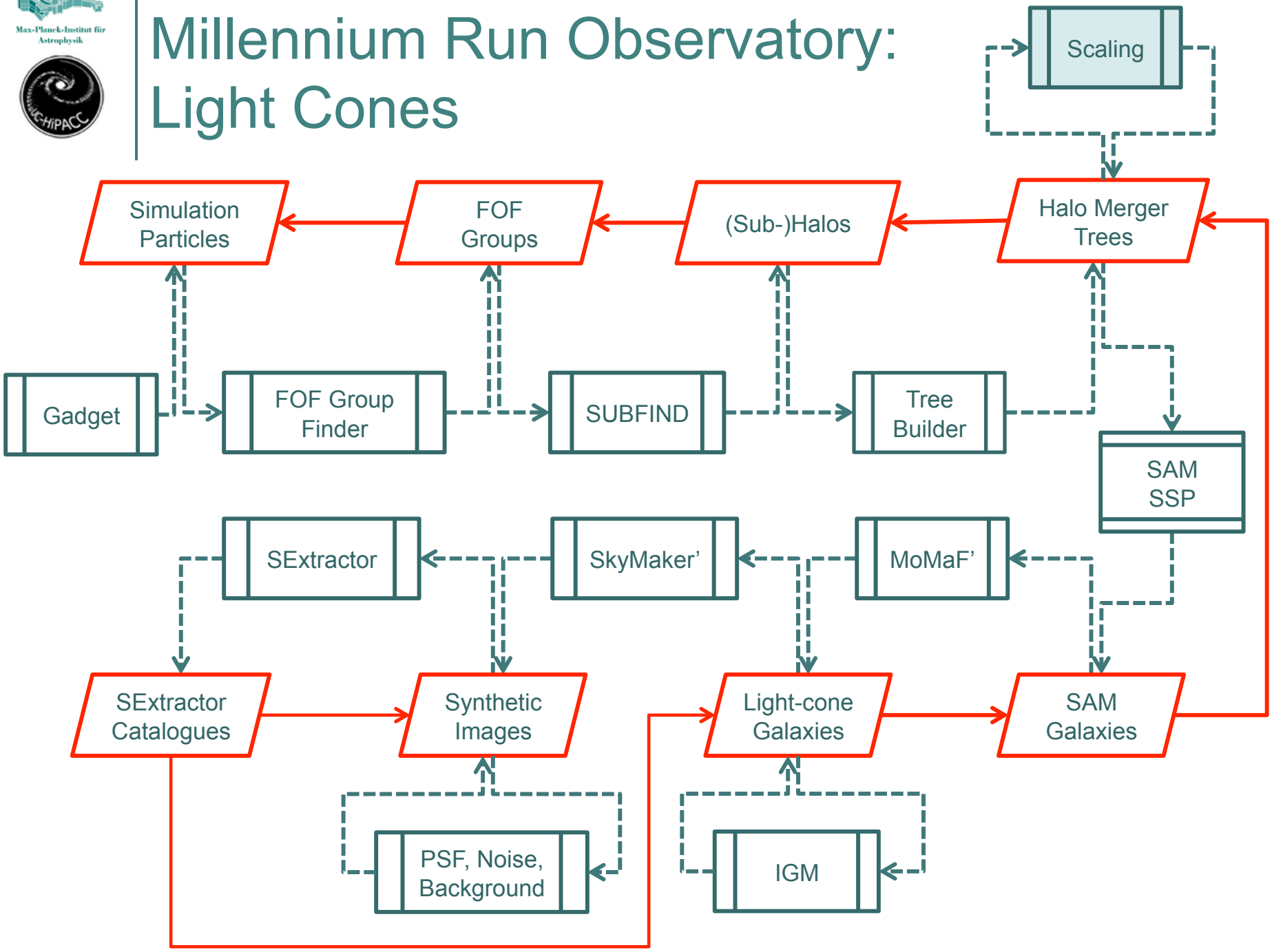


- Good convergence between MS and MSII
- Good fits at large scale → central galaxies are formed in the right halos.
- Excess at small scale → too many satellites? sigma_8 (0.9) too big?

Wrong cosmology?

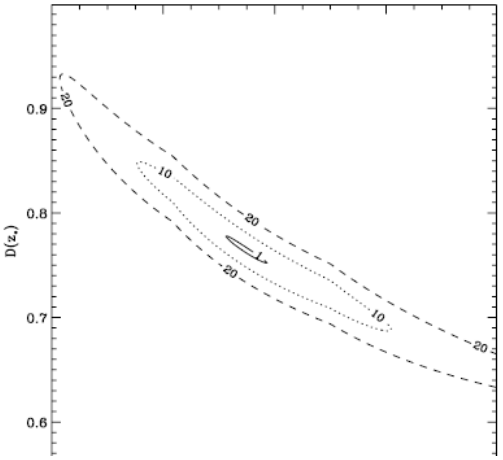
courtesy Qi Guo

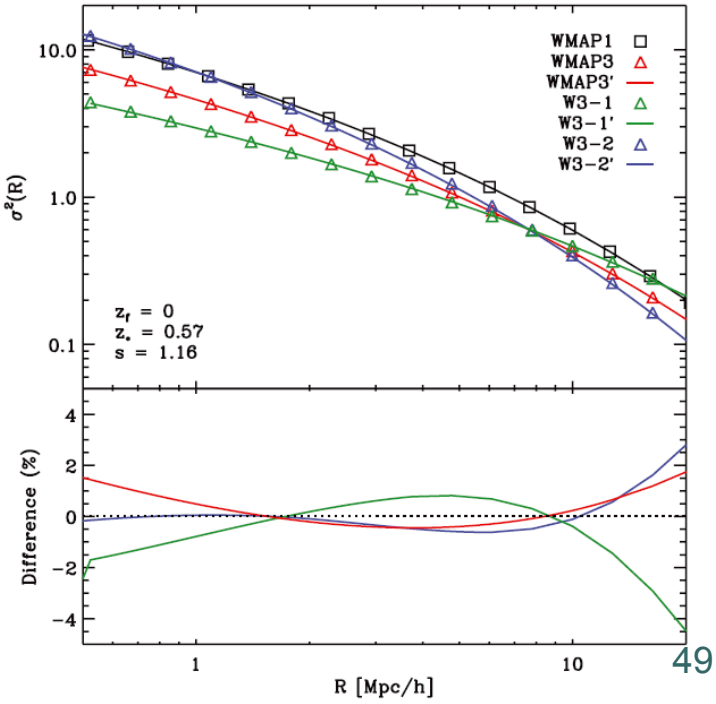
Millennium Run Observatory: Light Cones



Alternative cosmologies: Scaling

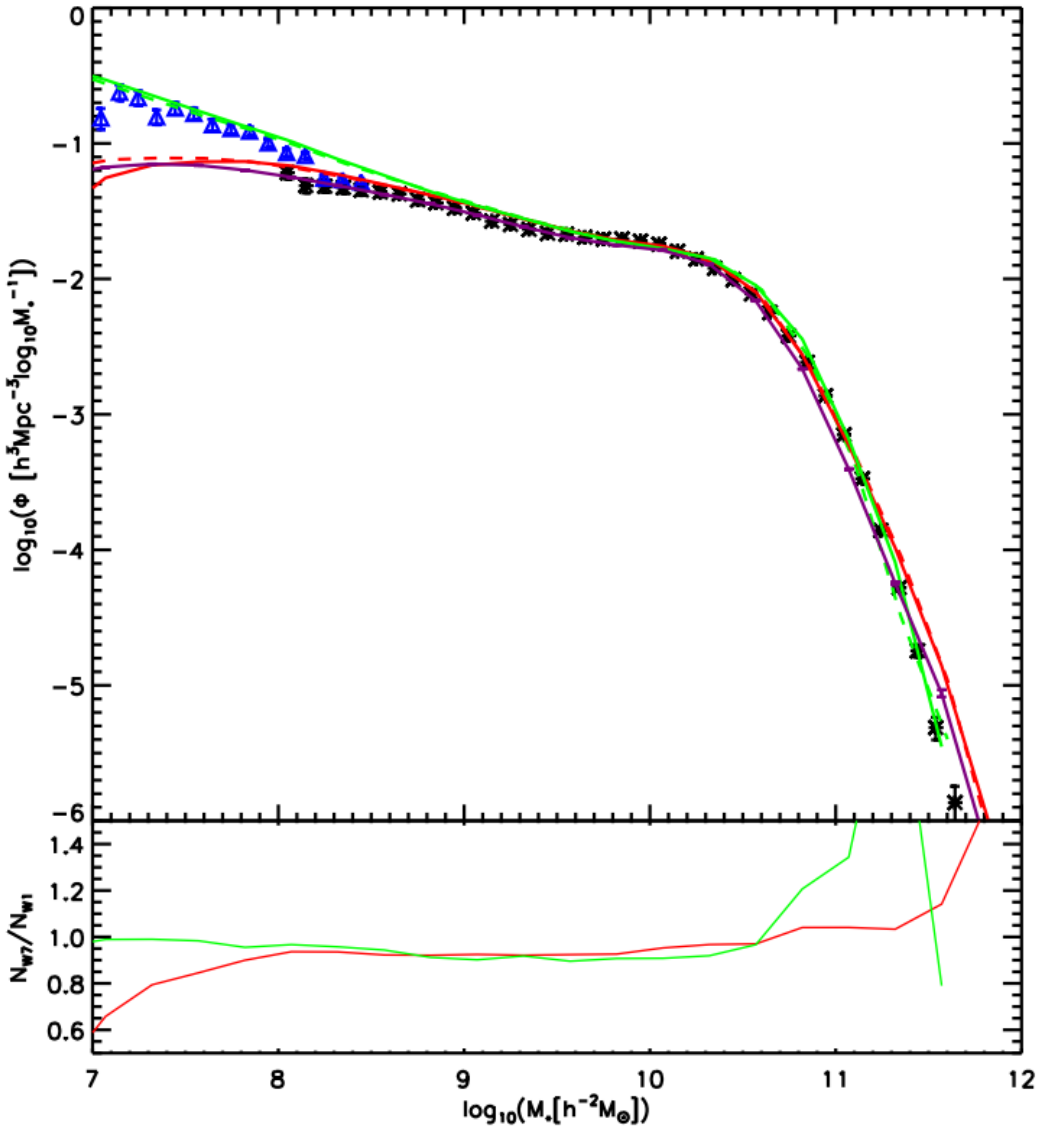
- Beyond WMAP1
  - without needing new simulations
- Angulo & White (2010)
  "One simulation to fit them all"
  - Use Millennium simulation
  - Scale size/mass/time to reproduce mass function/power spectrum of other cosmology
  - Use rescaled, retimed snapshots as input for SAM





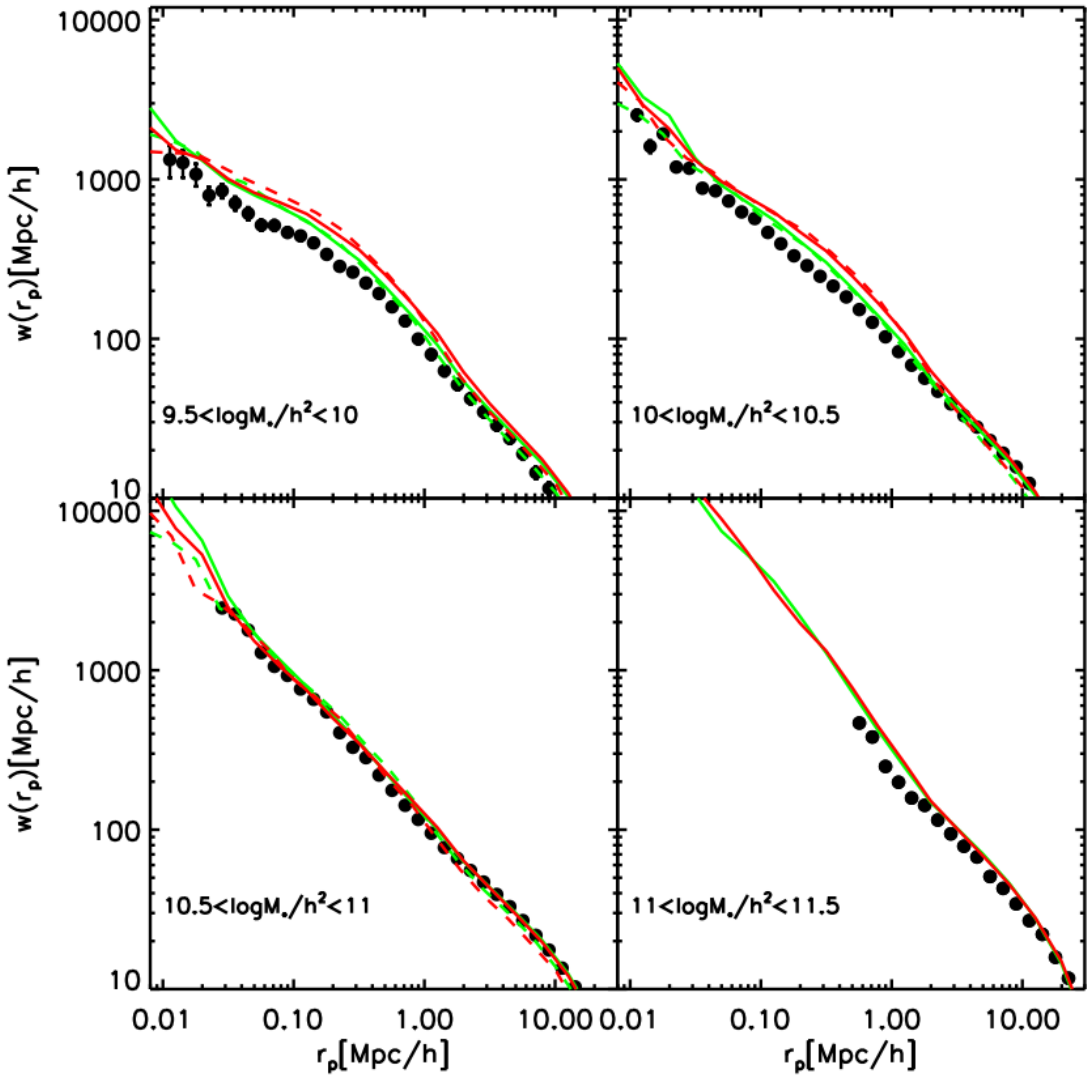
Guo etal 2012 arXiv:12060052

Solid: WMAP 7
Dashed WMAP 1
Red: MR
Green MR-II
Purple: MR-WMAP7

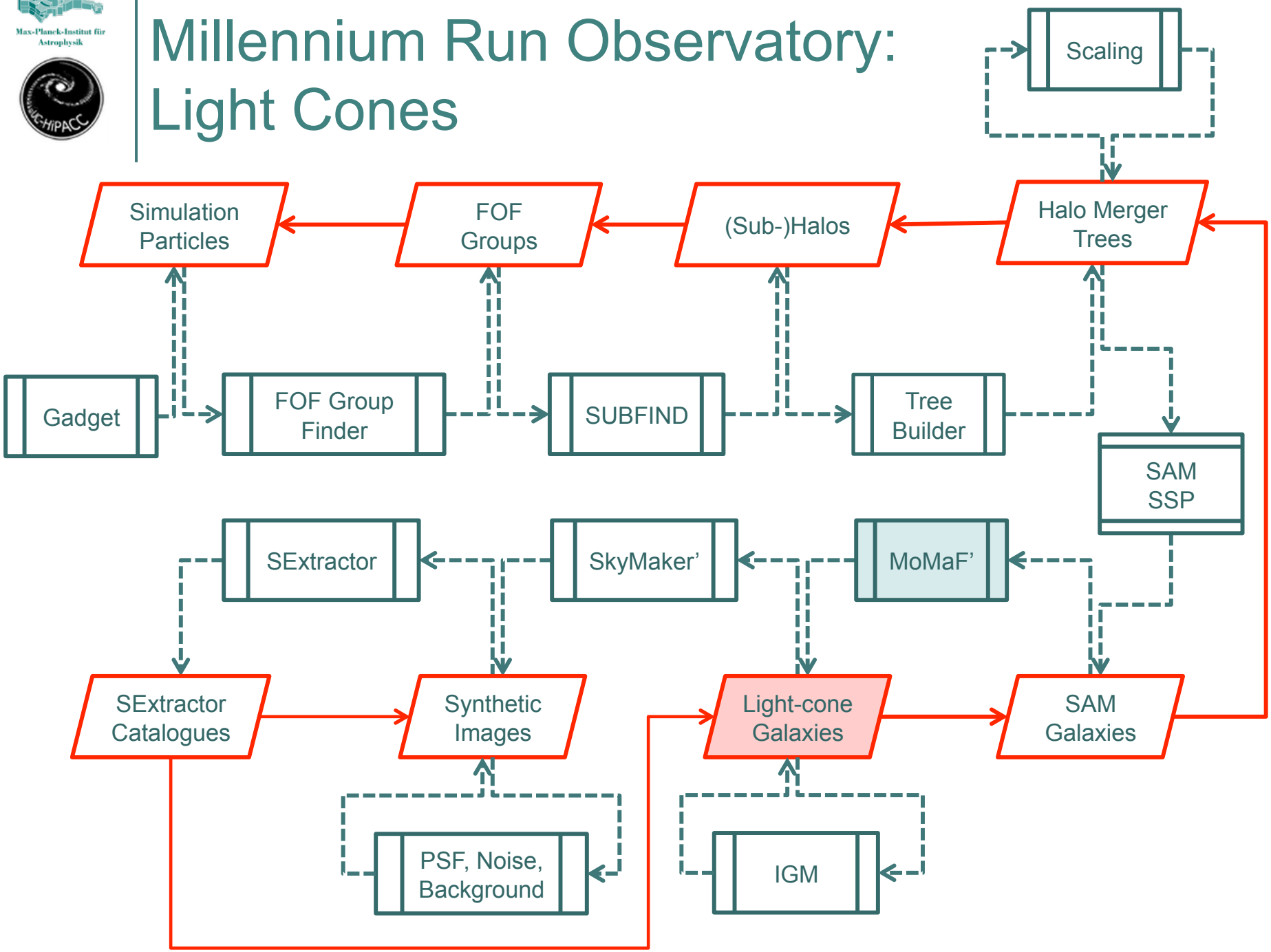


Correlation functions

Solid: MR
Dashed MR-II
Red: WMAP1
Green WMAP 7

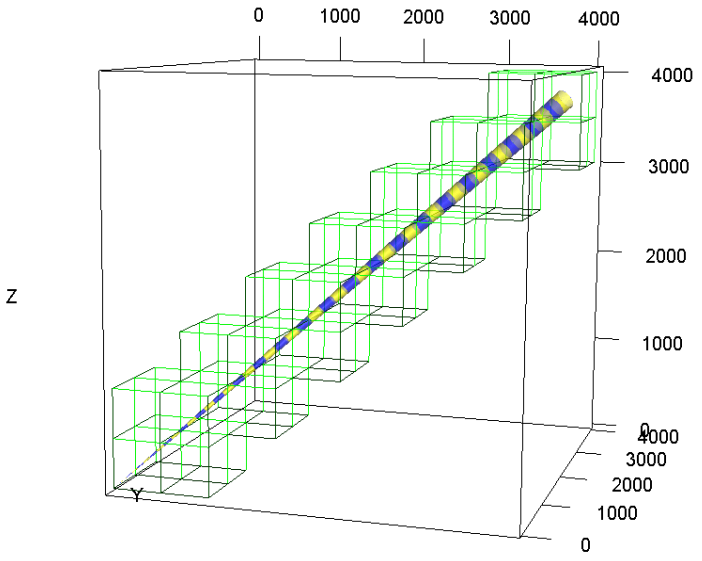


Millennium Run Observatory: Light Cones

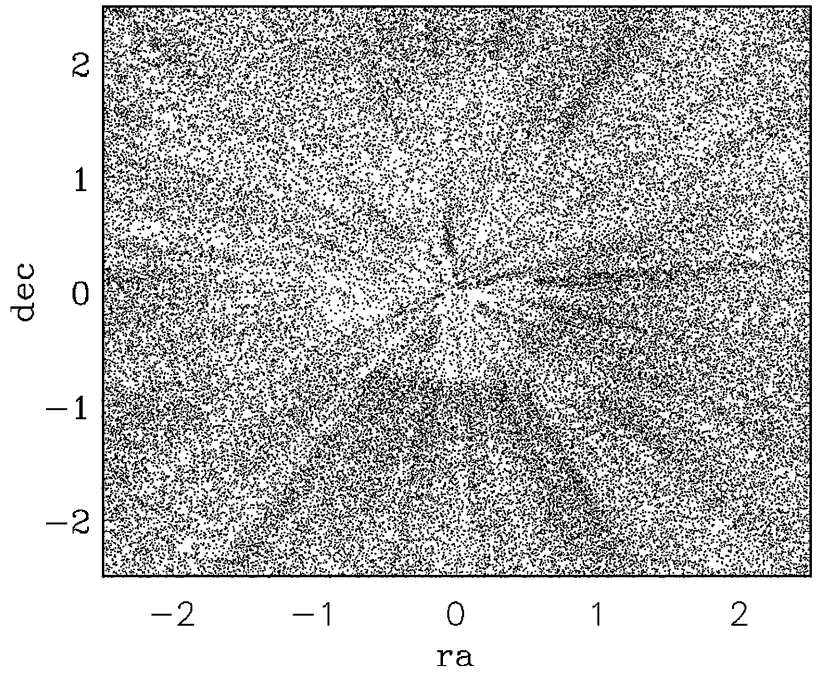


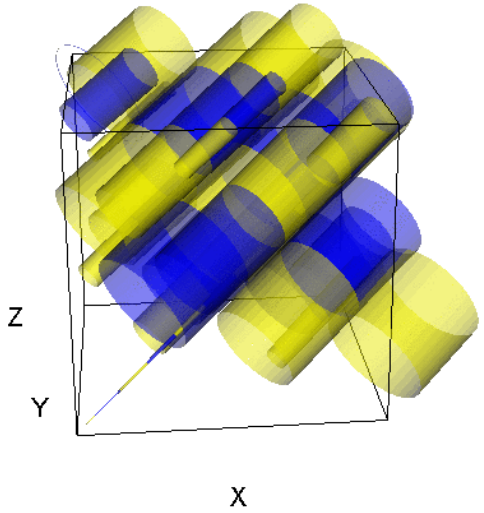
Light cones

- SAMs produce galaxies in snapshots
- To mimic real observations, need to produce “light cones”:
  - Observational properties: apparent magnitudes, sky positions, apparent redshifts, sizes, etc.
- Millennium box too small
  - Periodic replication needed
  - E.g. MoMaF, Blaizot etal 2005



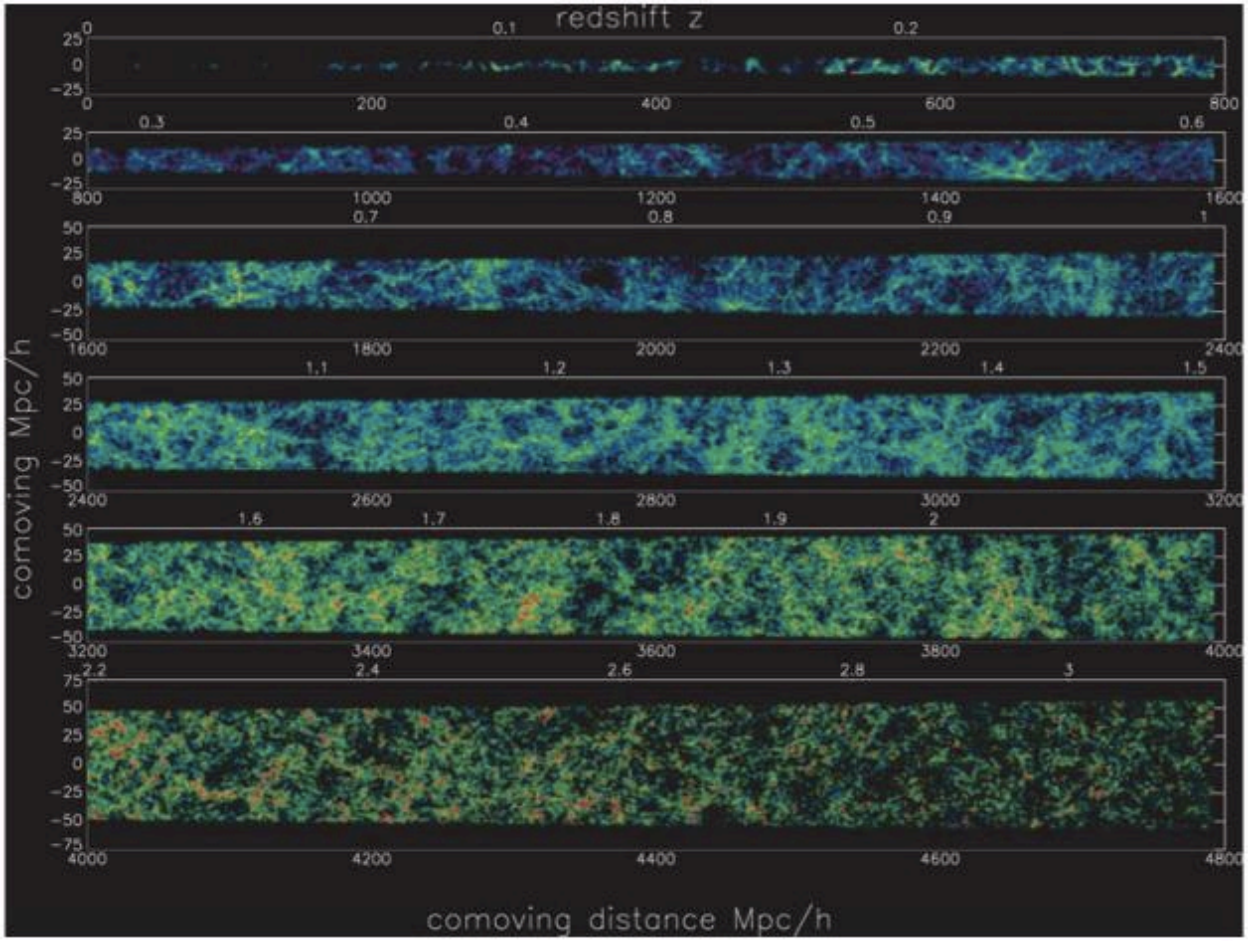
Perspective effect





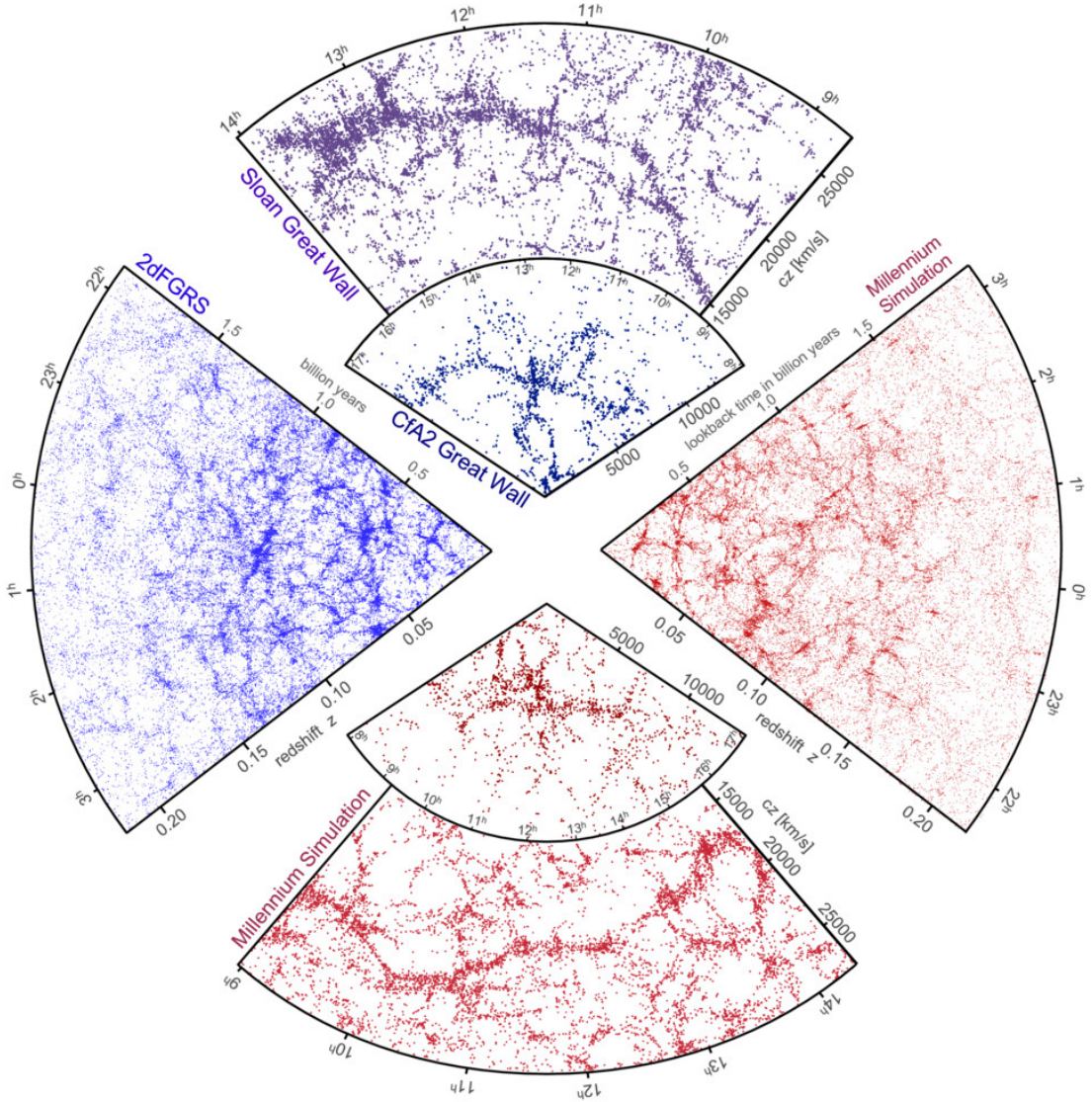
Max-Planck-Institut für Astrophysik

Pencil beams
(Kitzbichler & White 2007, Henriques etal 2012)

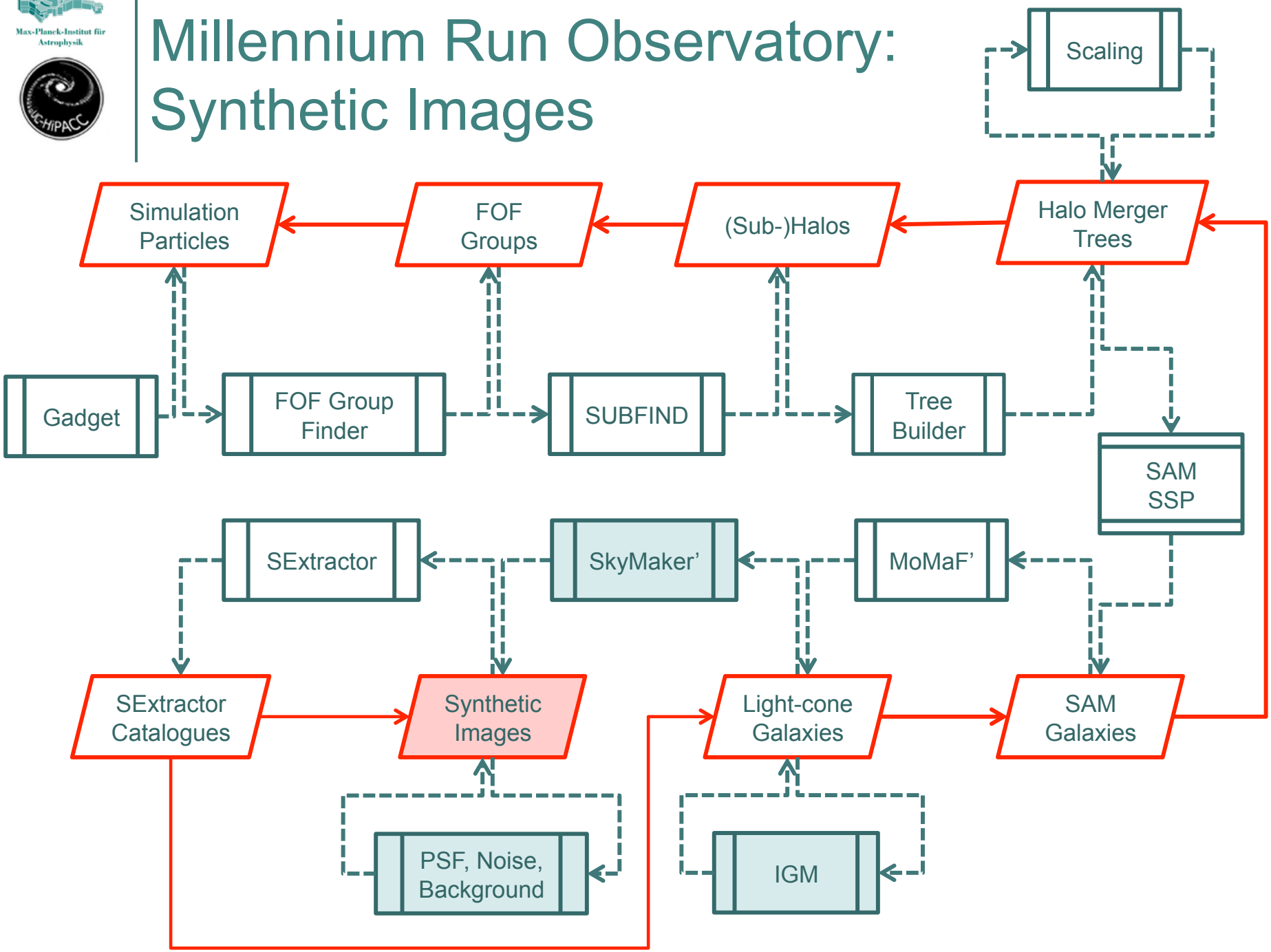


Max-Planck-Institut für Astrophysik

“All” Sky



Millennium Run Observatory: Synthetic Images



Fully known / No assumptions

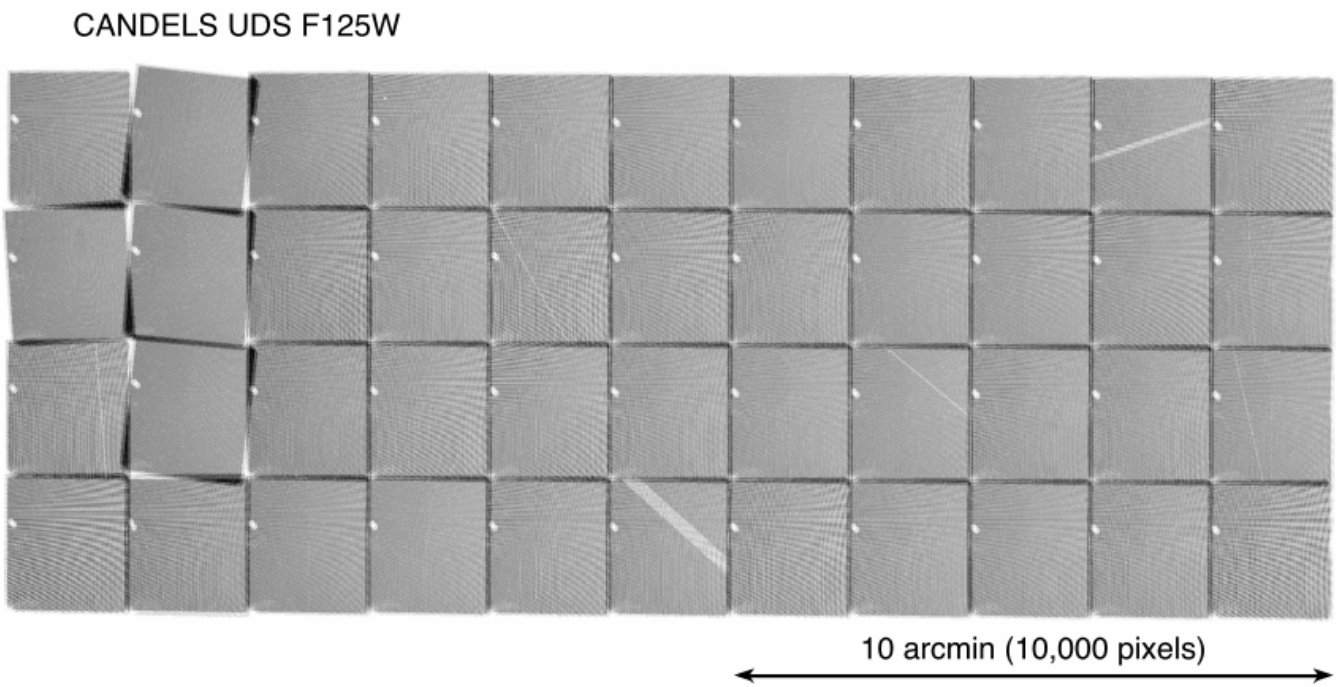
Simulations	*Comparison*	Observations
N-body Dark Matter (M_{halo}, pos, vel)	number counts/clustering ←	
SAM/Hydro (M_{stars}, SFR, size, ...)	SED fitting of Data ←	
SAM+synthesis modeling (rest-frame mags, colors)	SEDs applied to Simulation, selection function affects data →	Detected Objects ("Galaxies") (mags, colors, sizes, (z))
Mock/Lightcone (obs-frame mags, colors)	Realistic comparison **but still idealized** ↔	
Simulated (Noisy) Image (mags, colors, sizes)	Mock images "True Comparison" ↔	

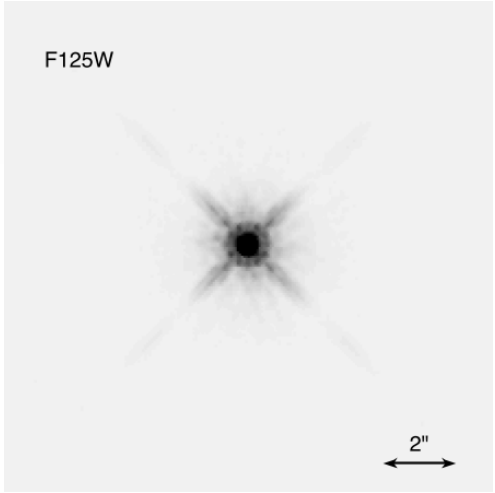
courtesy Roderik Overzier

SkyMaker (E. Bertin, 2008)

- Ideal image: bulges and disks
- Incl and PA from model

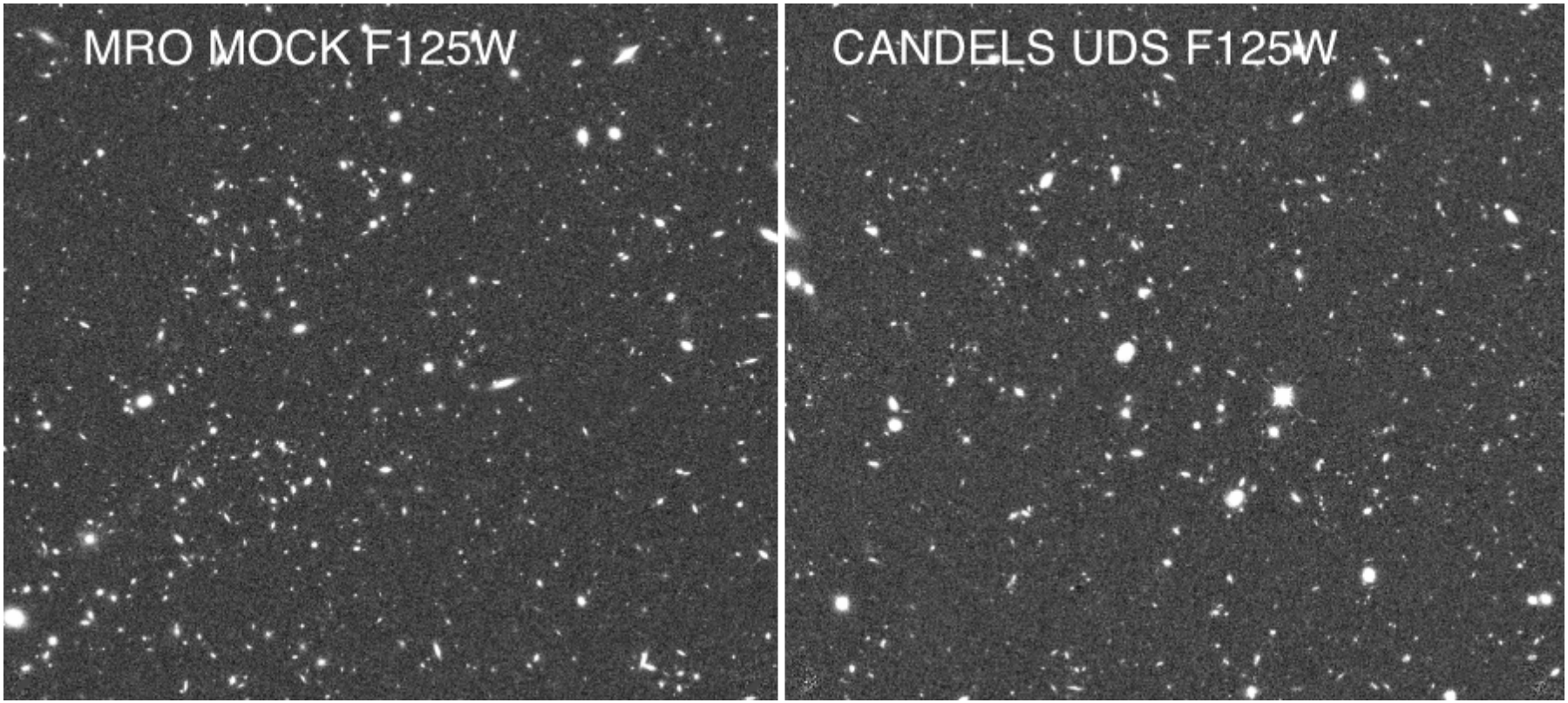
Observe: instrument, PSF, noise, depth, pixel size,…





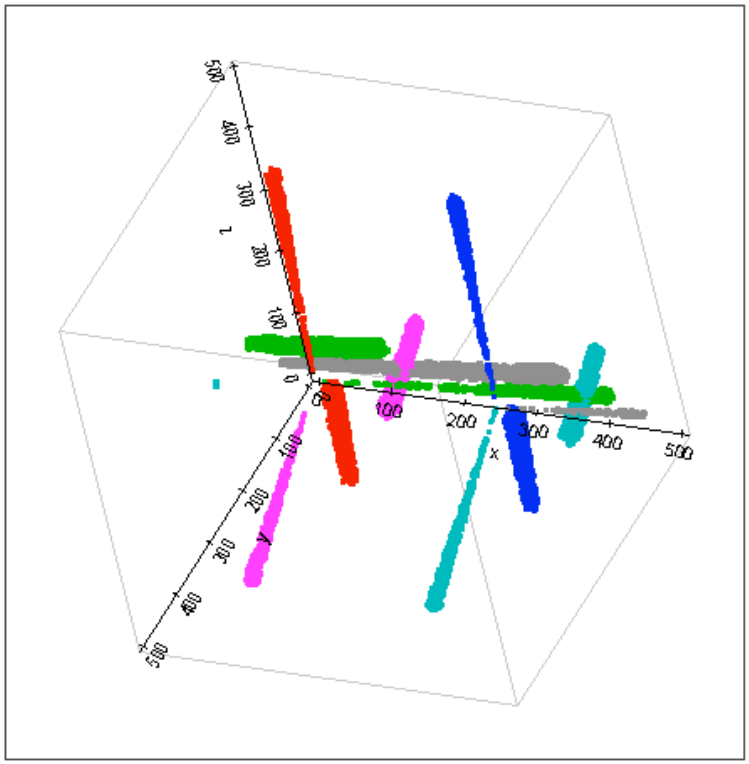
Max-Planck-Institut für Astrophysik

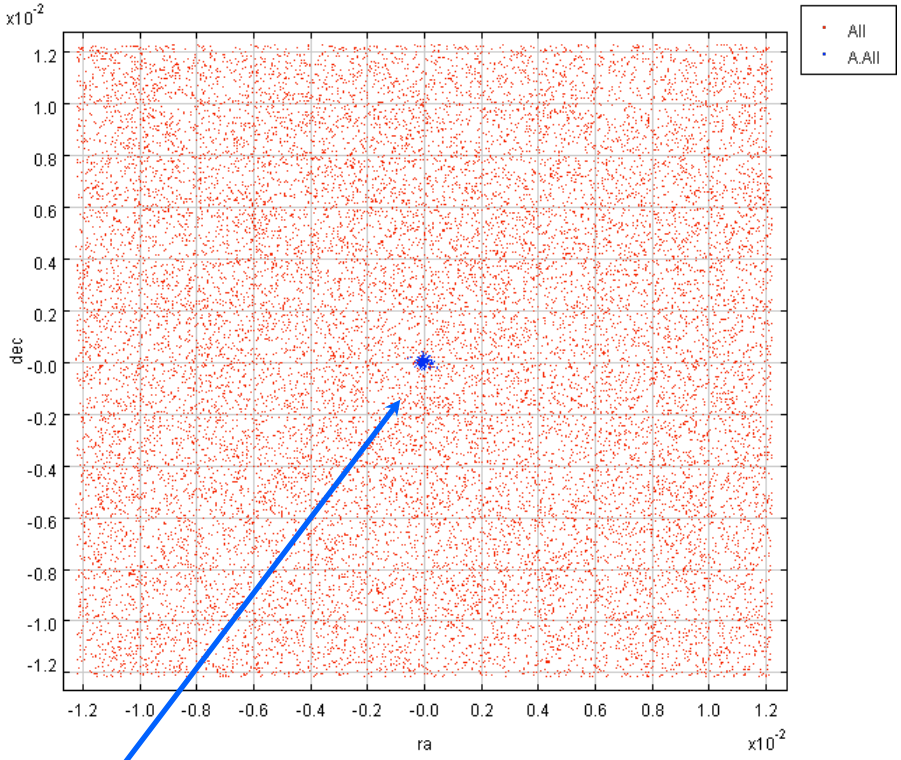
Mock vs real CANDELS



Looking at A random Direction...

Aiming at Specific Objects at Desired Redshift...

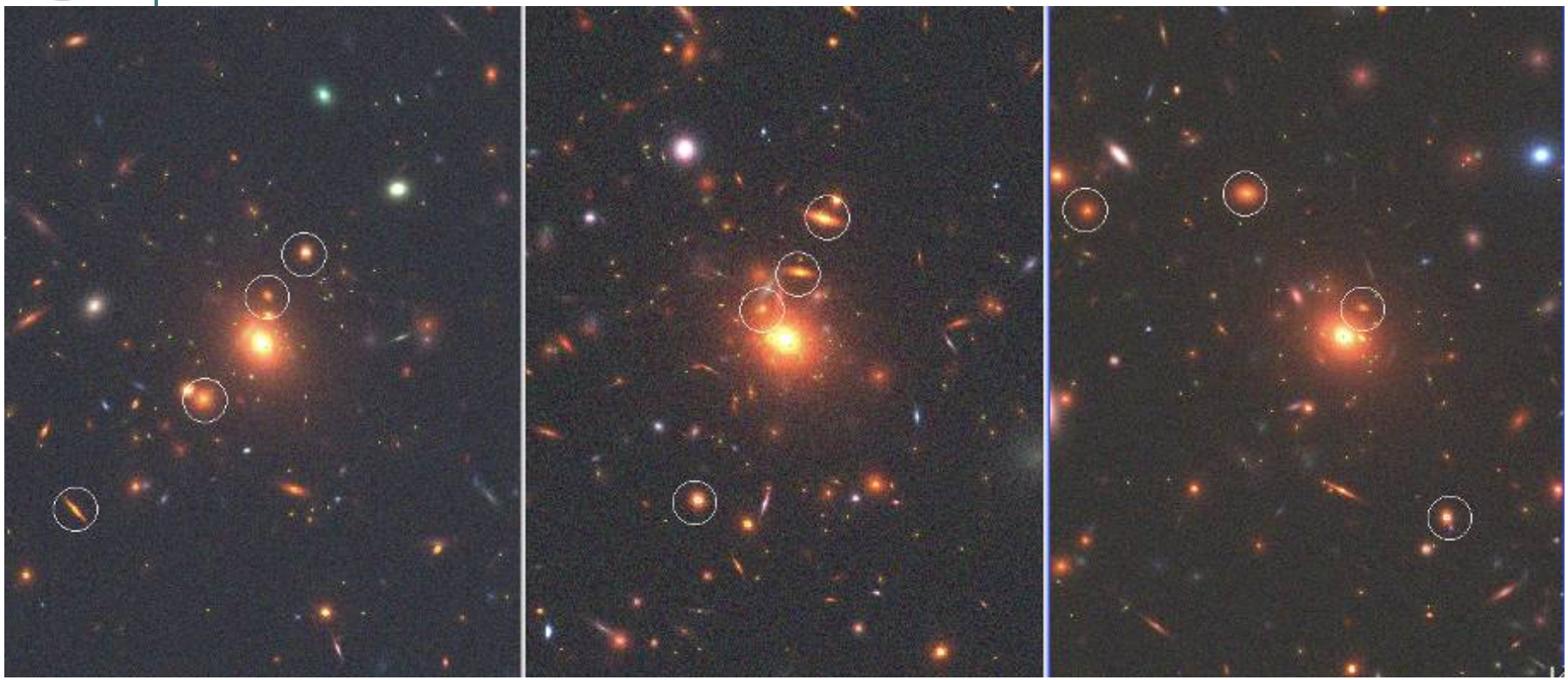




Cluster at z=2

Max-Planck-Institut für Astrophysik

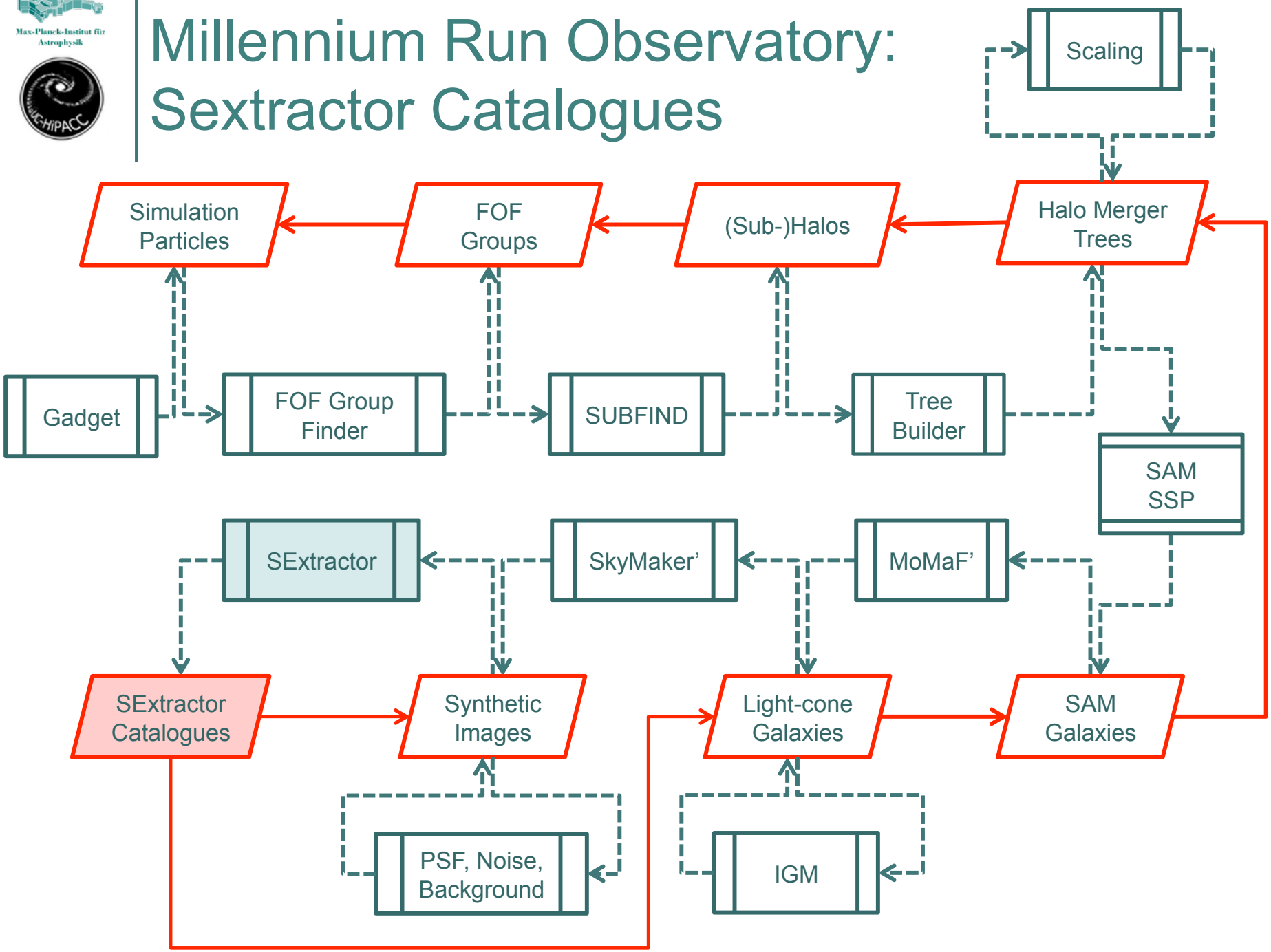
Aiming at a cluster from different directions



(most massive z=0 MR cluster, seen at z=0.83, three different observer's viewing angles)

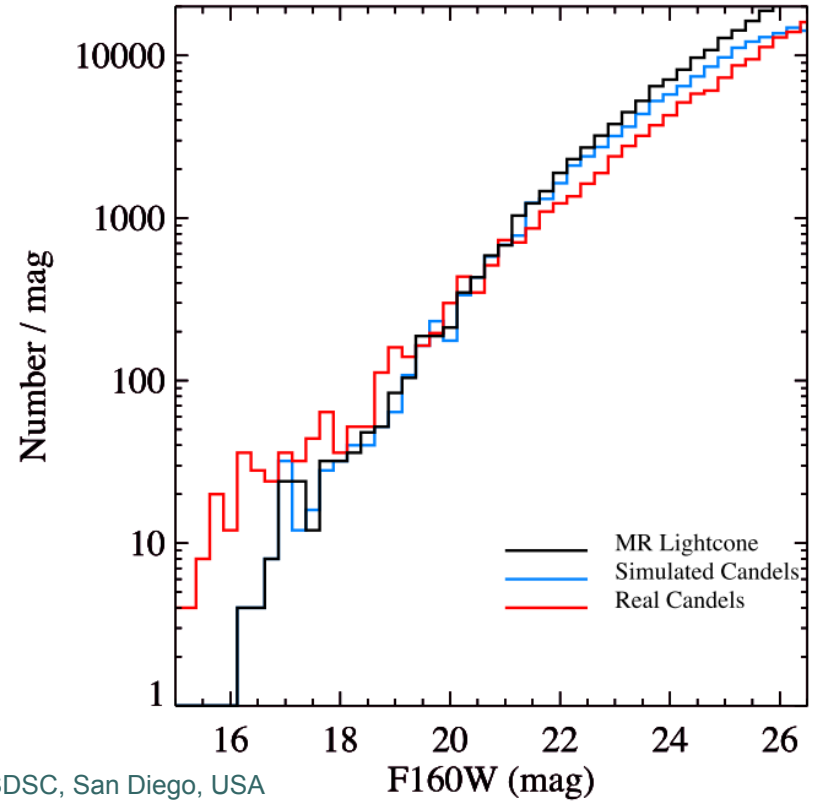
courtesy Roderik Overzier

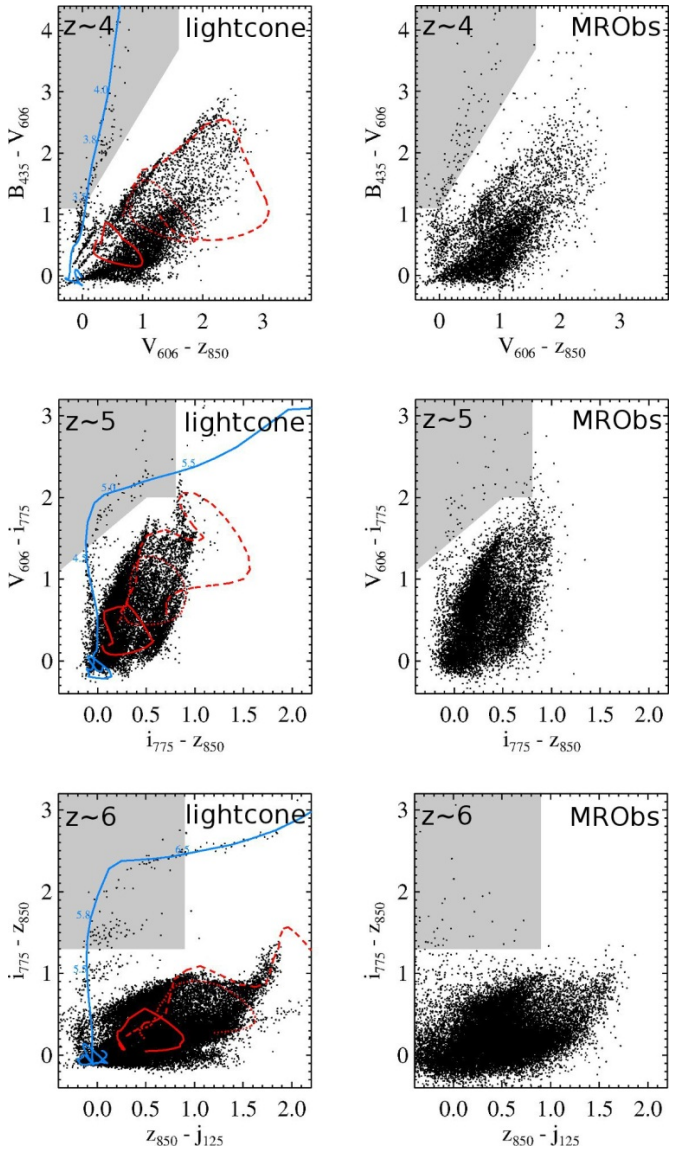
Millennium Run Observatory: Sextractor Catalogues



Analyse like ordinary images

- E.g. SExtractor
- Compare these catalogues with originals





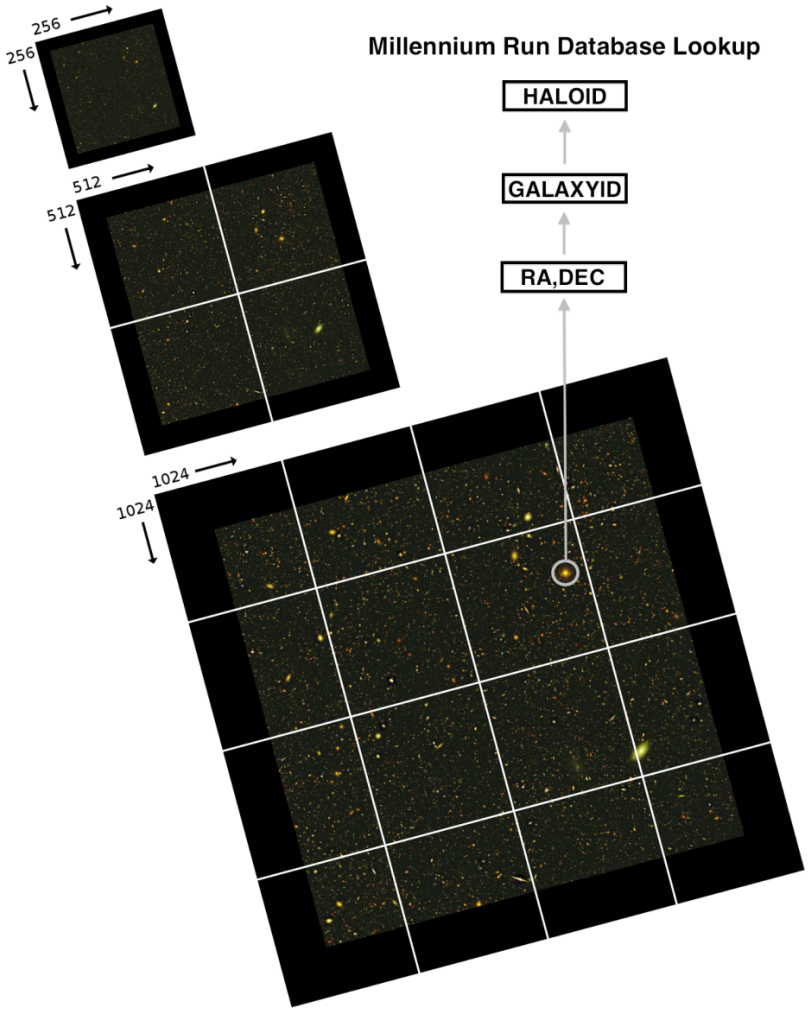
MRObs Online

- Original database site
  - http://gavo.mpa-garching.mpg.de/Millennium
  - http://gavo.mpa-garching.mpg.de/MyMillennium (private with MyDB)
- Download images etc
  - http://galformod.mpa-garching.mpg.de/mrobs

MRObs Image Browser

http://galformod.mpa-garching.mpg.de/mrobs/browser

- Online tool for browsing images “ala GoogleMaps”
- Image pyramids
- Interactive version (private for now) allows click-and-query



Max-Planck-Institut für Astrophysik

THANKS TO THE ORGANIZERS AND THANK YOU.

Acknowledgment:
GL is supported by Advanced Grant 246797 GALFORMOD from the European Research Council.