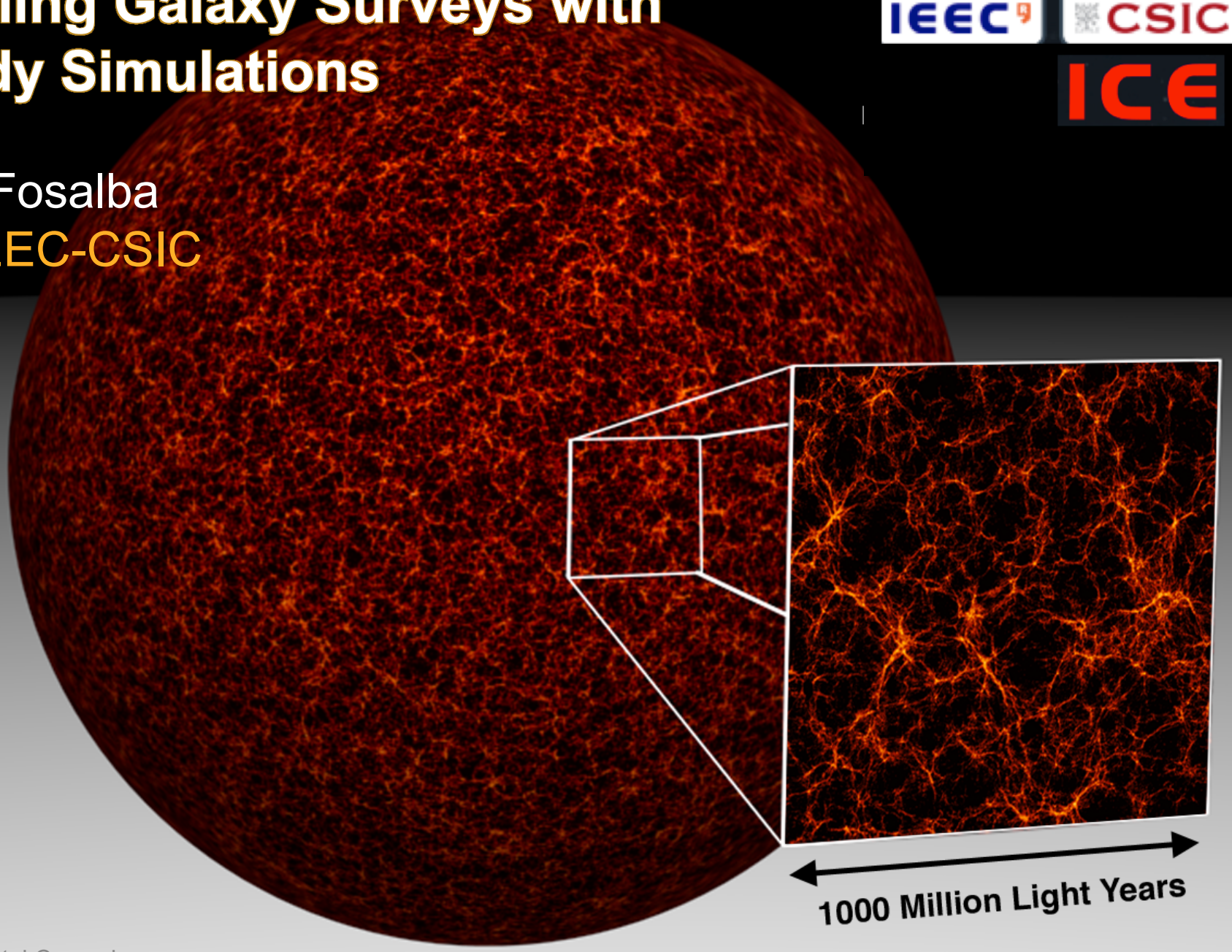
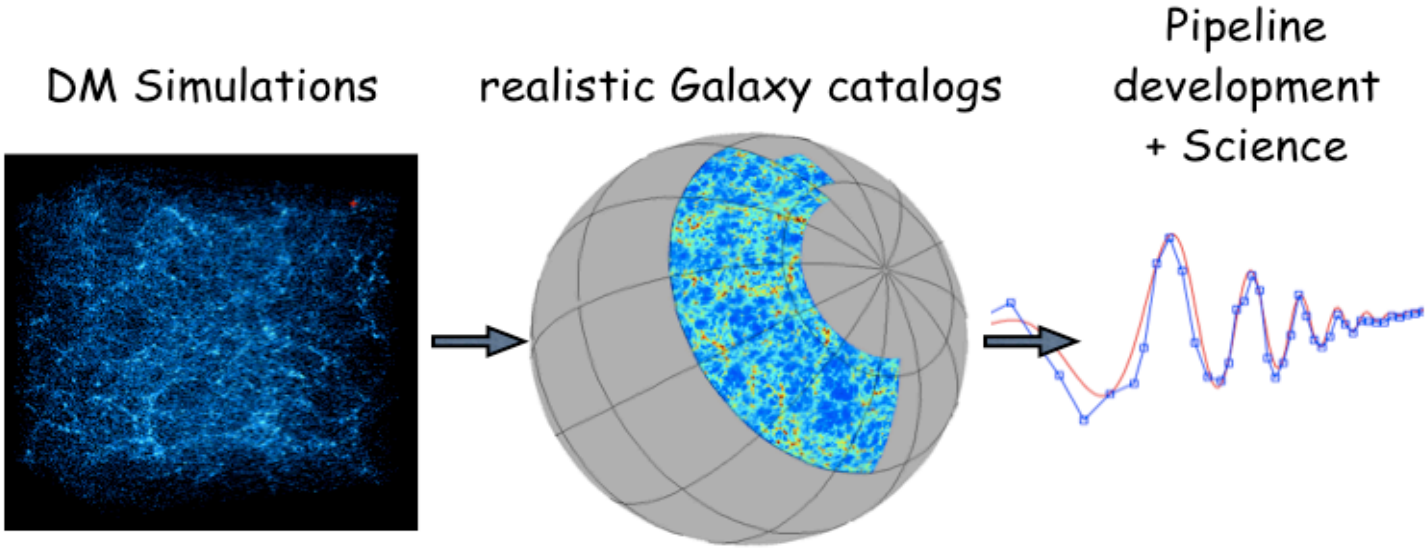
Modeling Galaxy Surveys with N-body Simulations

Pablo Fosalba
ICE, IEEC-CSIC



Fundamental Cosmology,
Fuerteventura June 5, 2014

Why do we need Nbody simulations ?



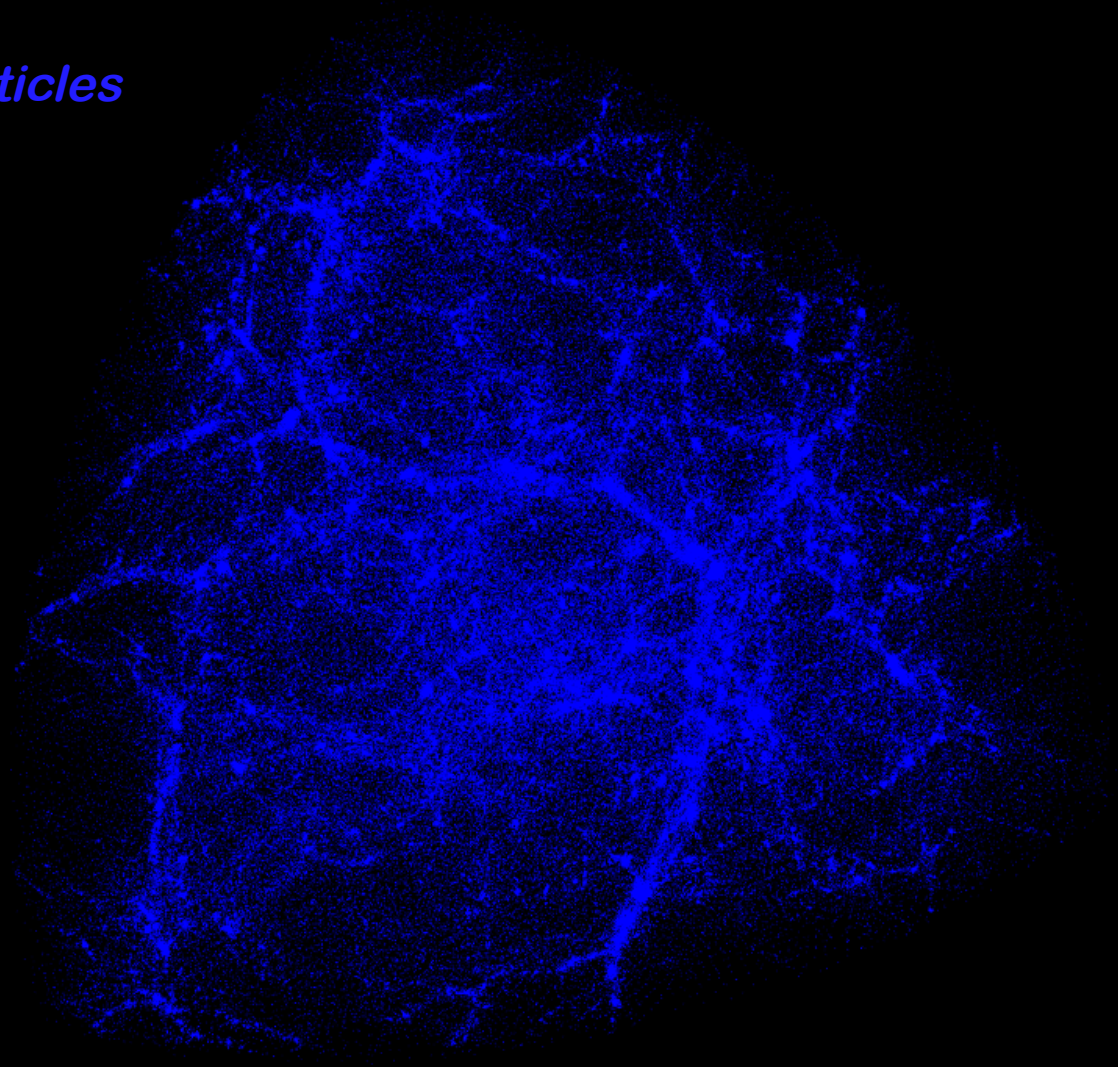
- Unique tool to model evolution of the universe into the Non-Linear regime
  (check validity of analytic theory, construct NL prediction codes “Emulators”)

- Build “Synthetic” Universes
  (input for DM halos, populate them with galaxies, apply survey mask,..)

- Develop reduction and analysis pipelines
  (mimic complexity of real data, extract science)

Building mock galaxy catalogs

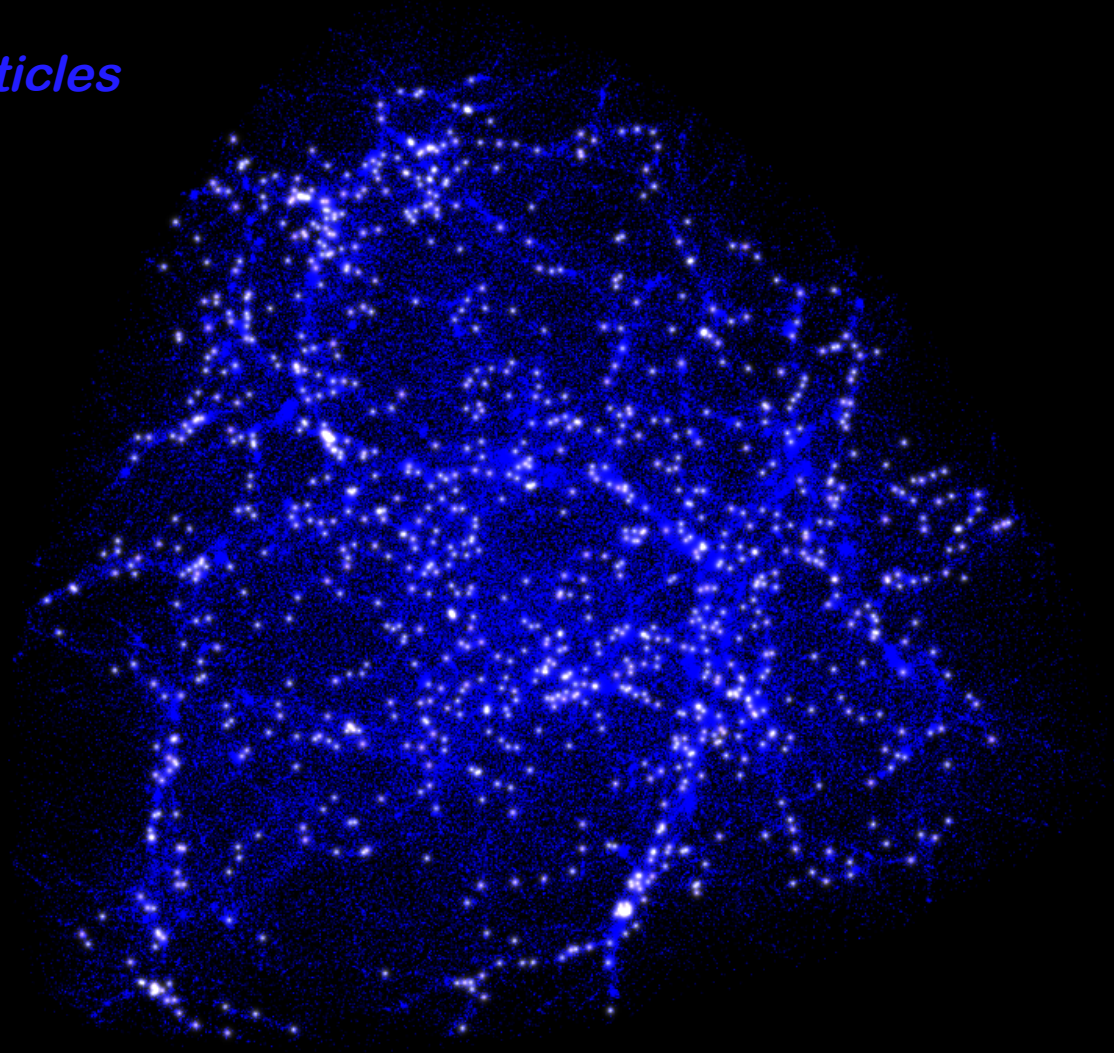
Dark Matter particles



Building mock galaxy catalogs

Dark Matter particles

DM Halos
(FoF / SO)

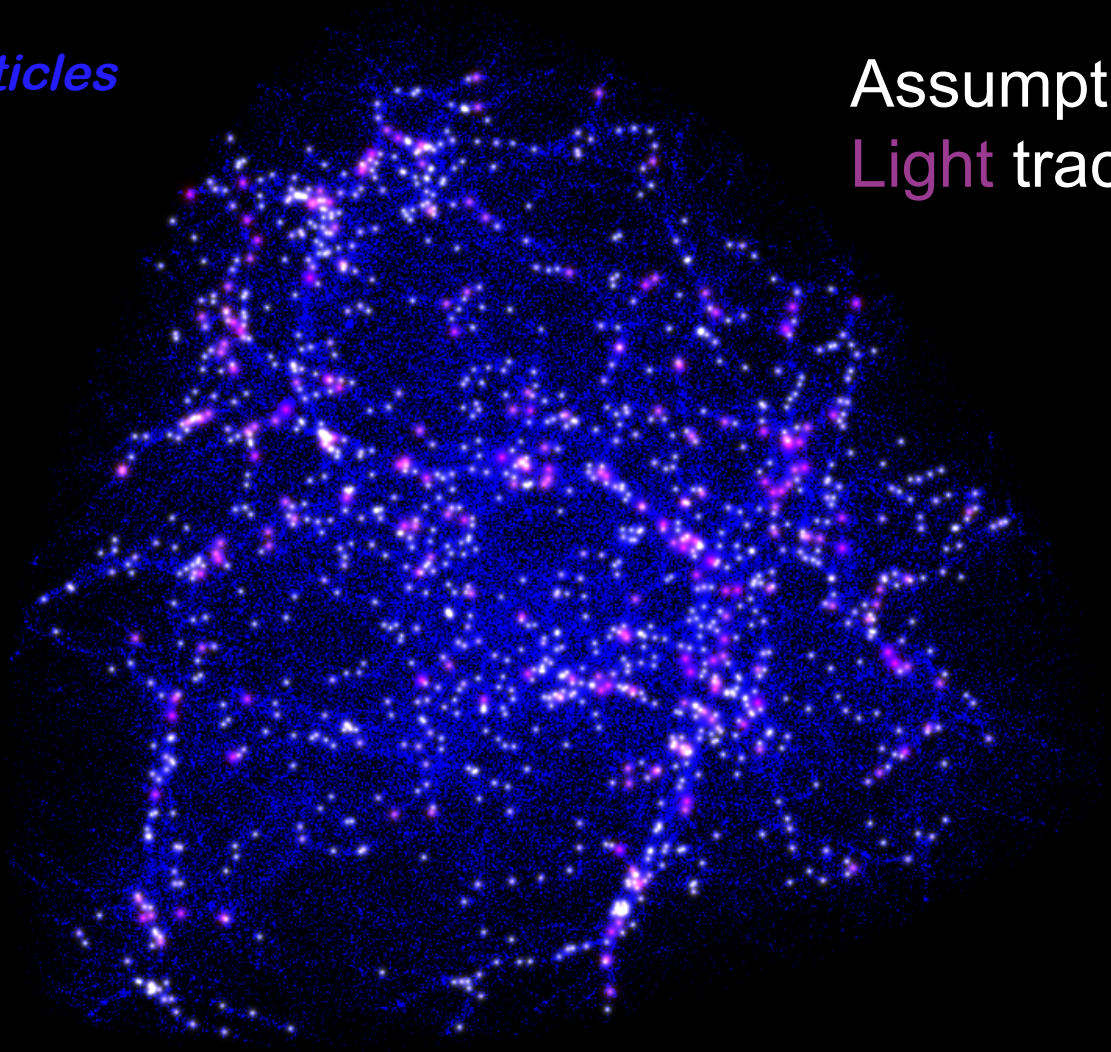


Building mock galaxy catalogs

Dark Matter particles

DM Halos
(FoF / SO)

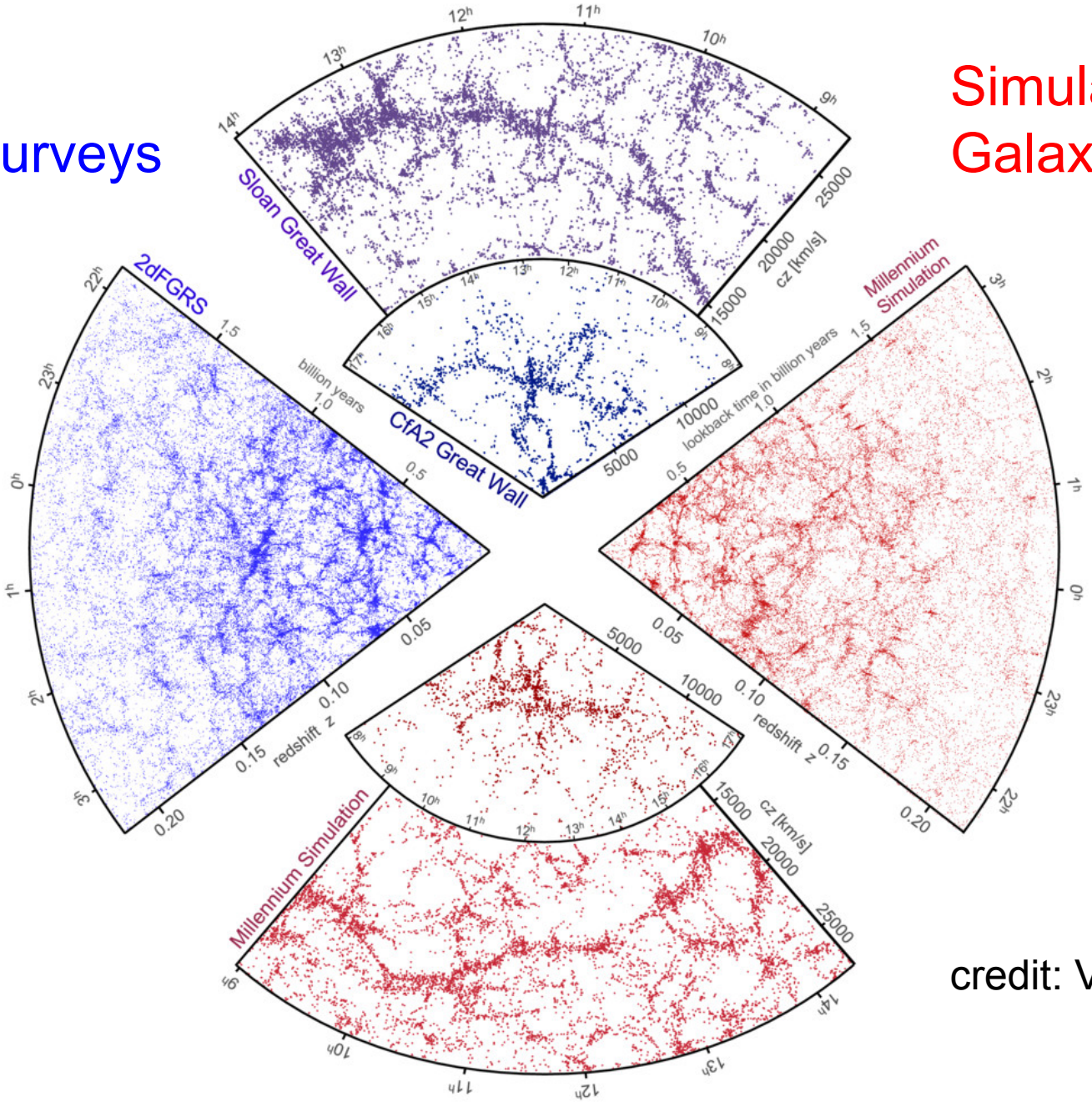
Galaxies
(SAM / HOD)



Assumption:
Light traces mass

True
Galaxy Surveys

Simulated
Galaxy Surveys



credit: V.Springel

Predictions from Simulations

(non-comprehensive list)

- Distribution of matter on large and small scales
- Abundance of collapsed objects (as function of t and M)
- Clustering of halos
- Halo profiles
- Baryon Acoustic Oscillations (BAO)
- Gravitational Lensing statistics (Galaxies and CMB)
- Integrated Sachs-Wolfe effect
- …

"Systematic" error in these predictions has to be kept subdominant wrt expected accuracies from Galaxy Surveys

N-body Algorithms/Simulation Codes

- Start from small initial fluctuations in an expanding universe
- Gravity (Newtonian) only force on large + DM is collisionless
- Self-gravitating fluid: discretize equations into *N-body system*

resolve small structures
+
sample fair fraction of universe

→ Large *N* !

Exact solution: **N^2 problem** ($N<10^6$)
Approximate solvers ($N > 10^6$)*:*

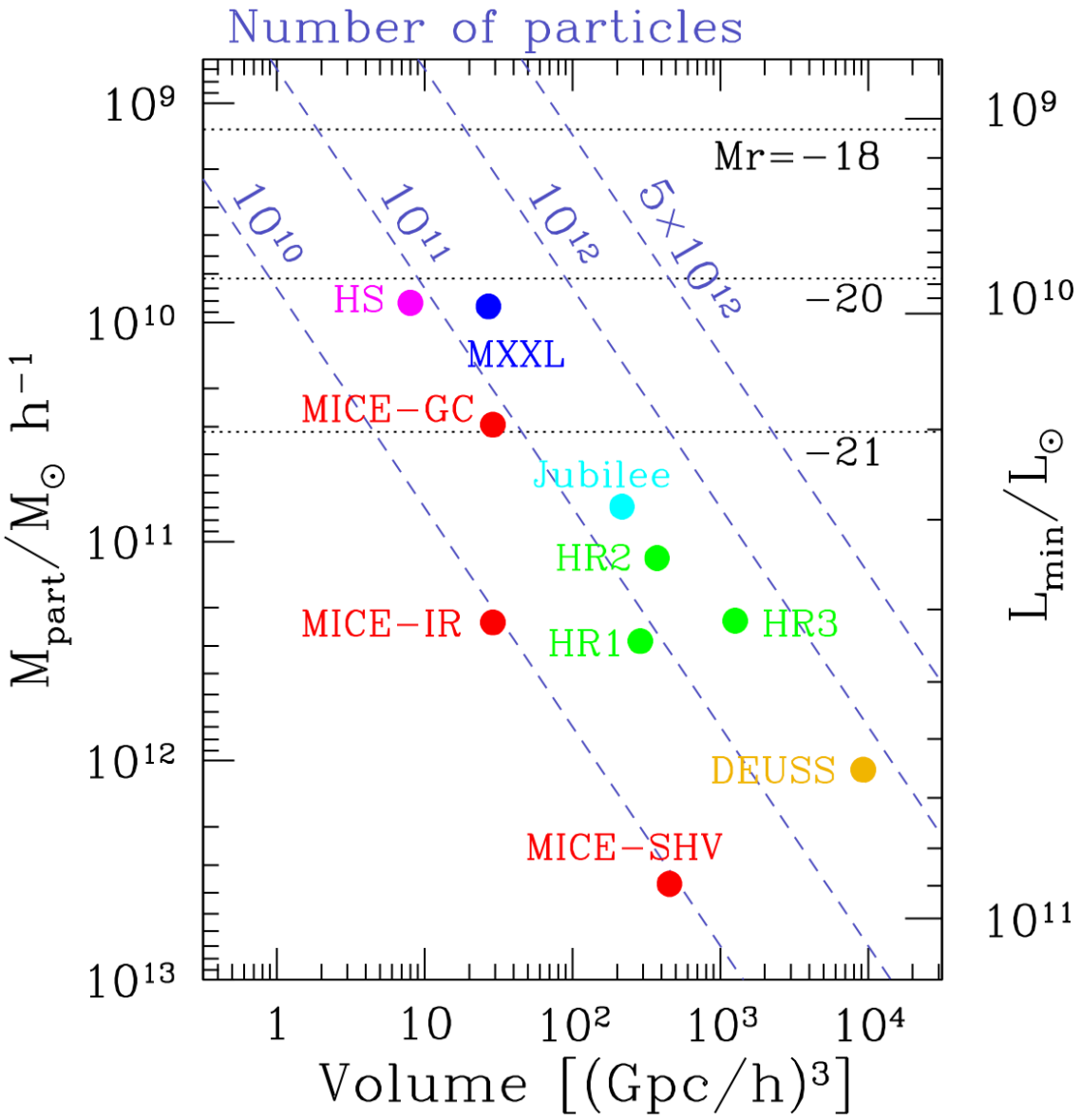
- ✓ Large scales only: "Particle Mesh" (FFT based)
- ✓ Split force computation into large + *small scales*:
  "Tree" methods (**Gadget, PKDGRAV** codes), or
  "Adaptive" mesh methods (**RAMSES, CUBEP^{3}M** codes)

→ reduce to *O*(N logN) + codes fully parallel

Where do we stand ?
What is the next frontier ?

State of the art in Nbody simulations

MXXL: Angulo et al.
Horizon Sim: Teyssier et al.
MICE: Fosalba et al.
HR1,2,3: Kim et al.
DEUSS: Alimi et al.
Jubilee: Watson et al.

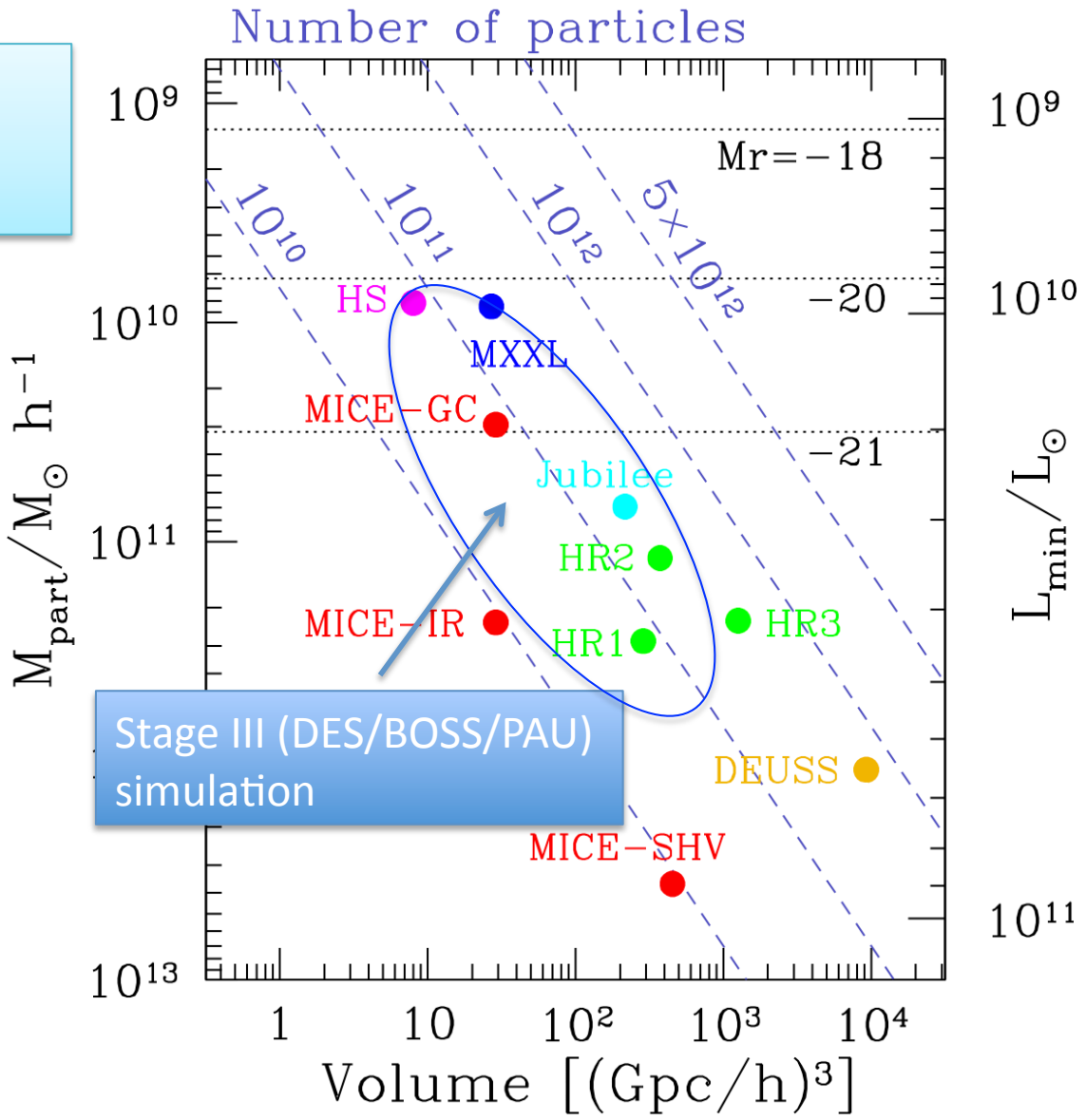


State of the art in Nbody simulations

Current simulations can model MilkyWay like galaxies (10^{10} $L_\odot$) over large volumes

MXXL: Angulo et al.
Horizon Sim: Teyssier et al.
MICE: Fosalba et al.
HR1,2,3: Kim et al.
DEUSS: Alimi et al.
Jubilee: Watson et al.

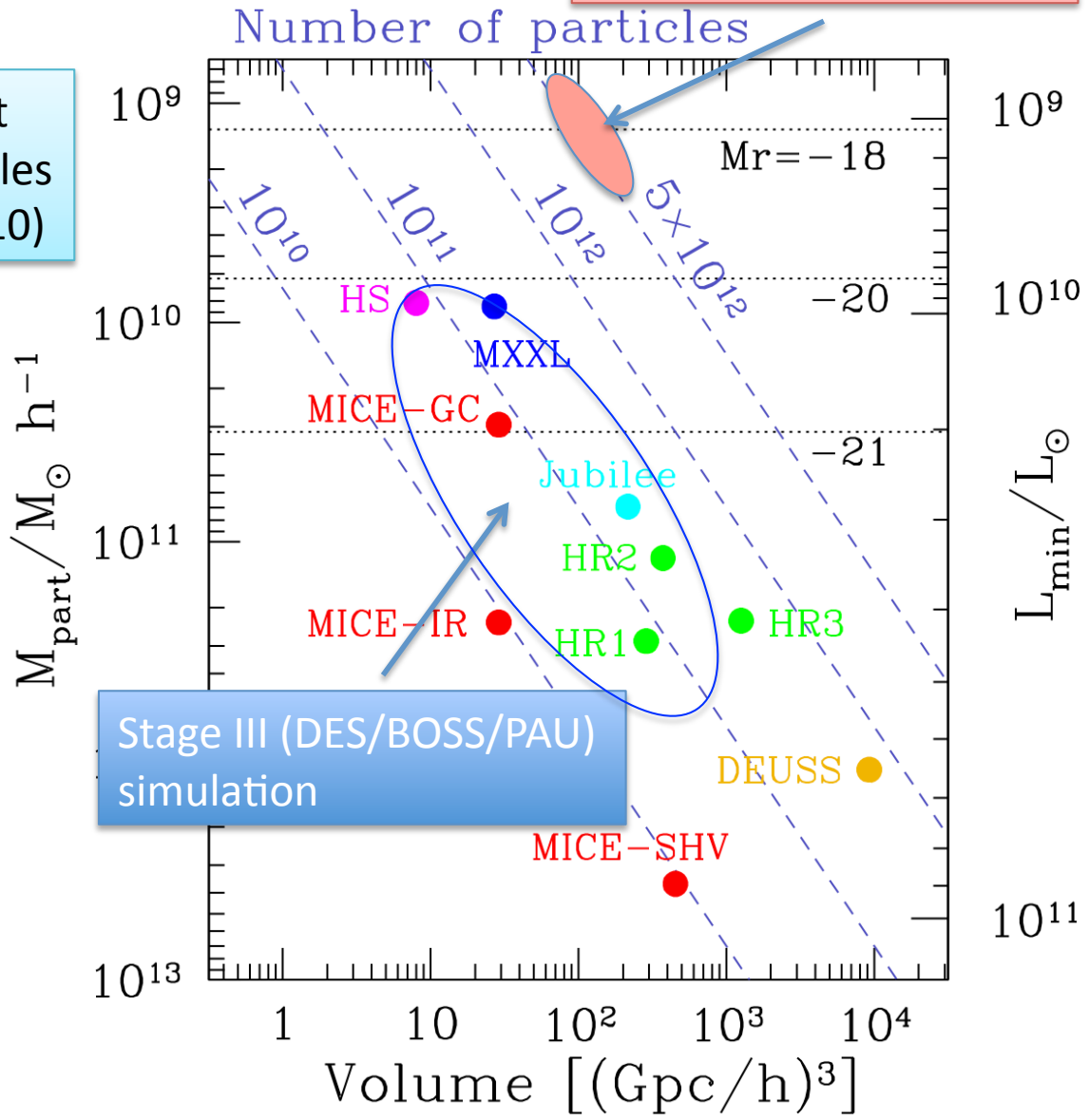
Faintest mock galaxy (L_{min}) hosted by 100-particle halo…



State of the art in Nbody simulations

Need to push current limits by at least one order of magnitude in num.particles (larger volumes, fainter galaxies $L_\star$ /10)

MXXL: Angulo et al.
Horizon Sim: Teyssier et al.
MICE: Fosalba et al.
HR1,2,3: Kim et al.
DEUSS: Alimi et al.
Jubilee: Watson et al.

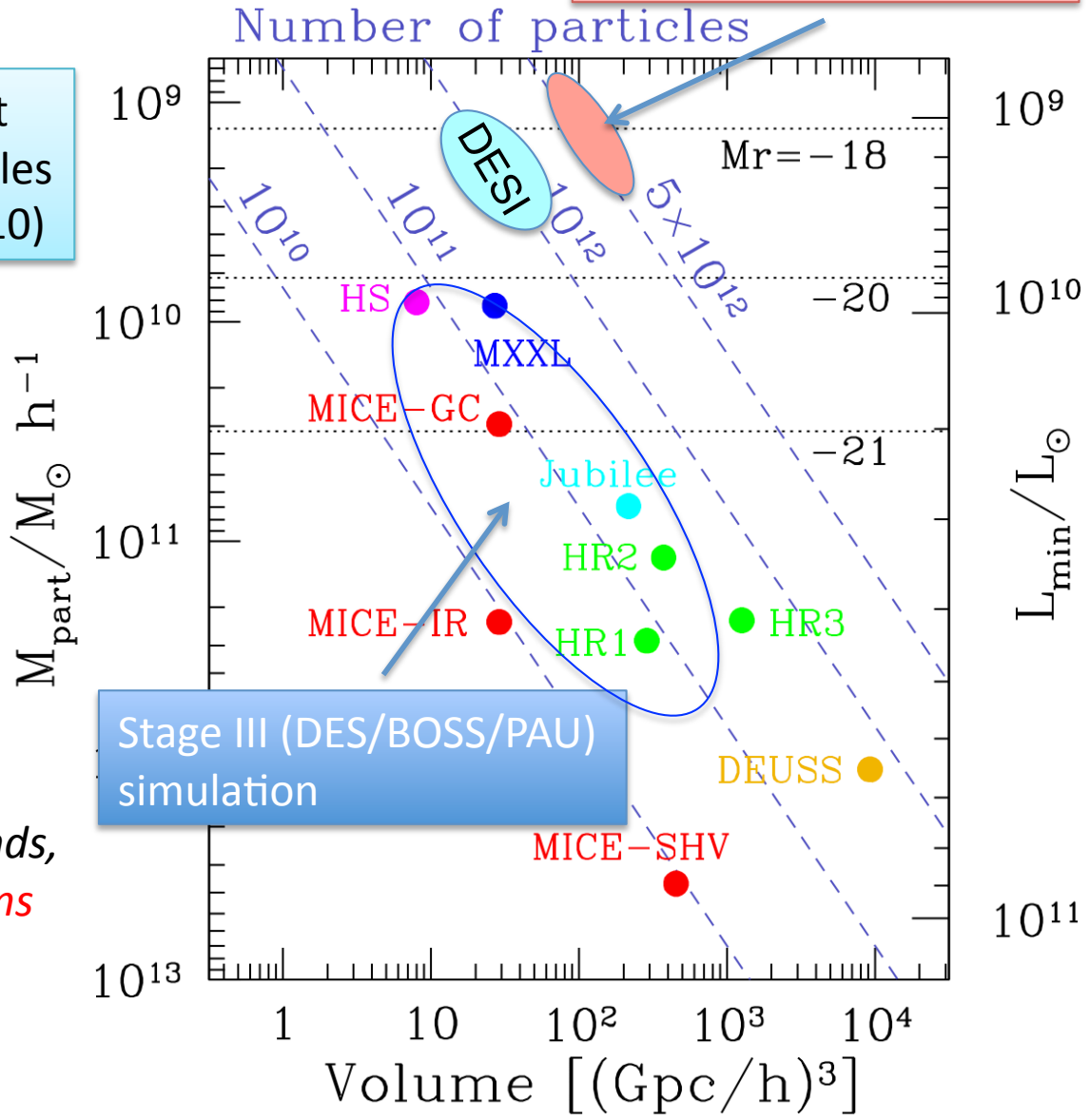


State of the art in Nbody simulations

Need to push current limits by at least one order of magnitude in num.particles (larger volumes, fainter galaxies $L_\star$ /10)

MXXL: Angulo et al.
Horizon Sim: Teyssier et al.
MICE: Fosalba et al.
HR1,2,3: Kim et al.
DEUSS: Alimi et al.
Jubilee: Watson et al.

Good news: *extrapolating current trends, we could run 5x10^{12} particle simulations by 2016*



Results from state-of-the-art (STAGE III) Survey Simulations

Millennium-XXL

Largest high-resolution N-body simulation

303 billion particles

L = 3 Gpc/h

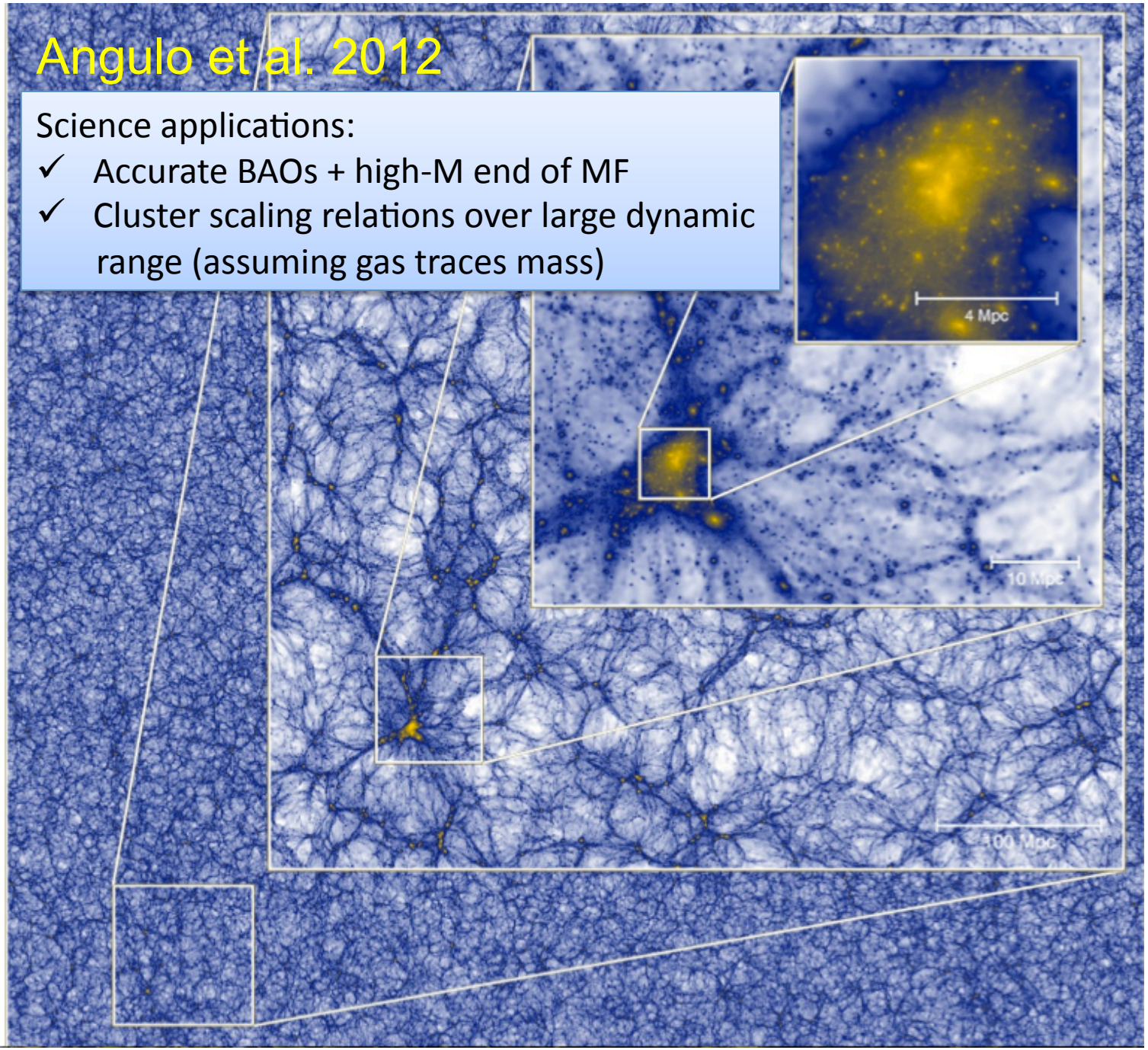
~700 million halos at z=0

~25 billion (sub)halos in mergers trees

$m_p = 6.1 \times 10^9 M_\odot/h$

12288 cores, 30 TB RAM on Supercomputer JuRoPa in Juelich

2.7 million CPU-hours



Millennium-XXL

Largest
high-resolution
N-body simulation

303 billion particles

L = 3 Gpc/h

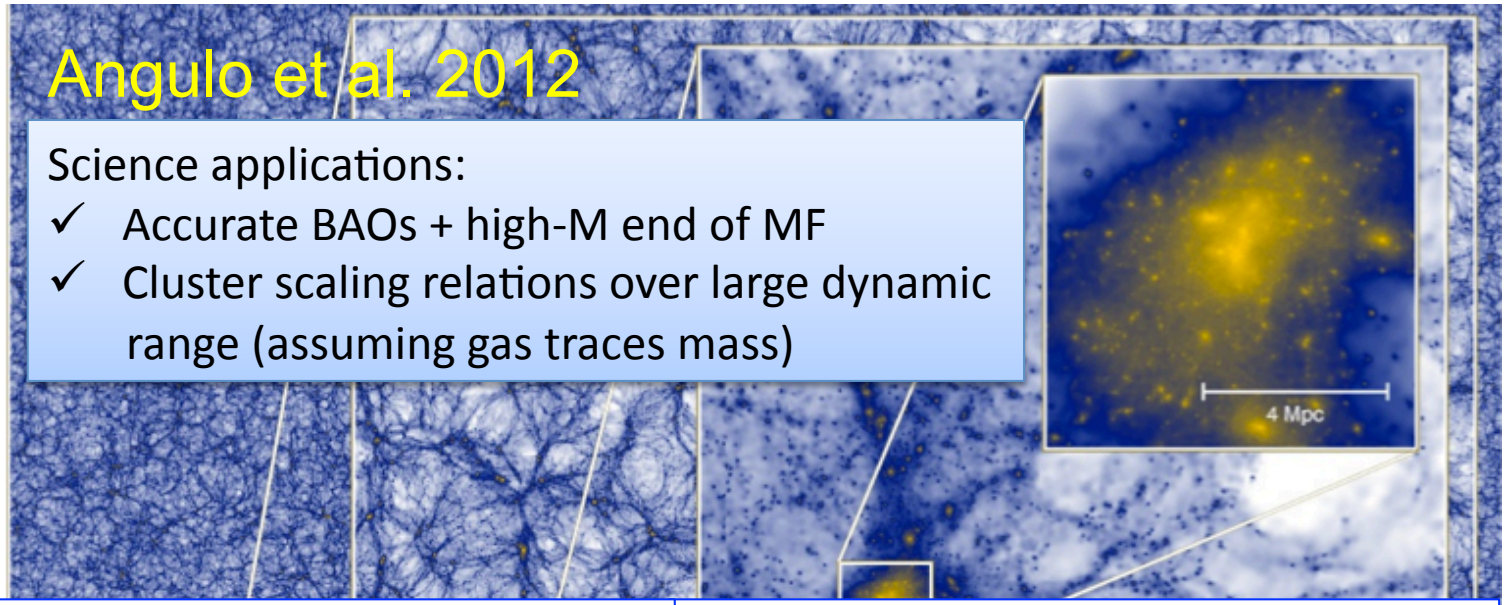
~700 millior
at z=0

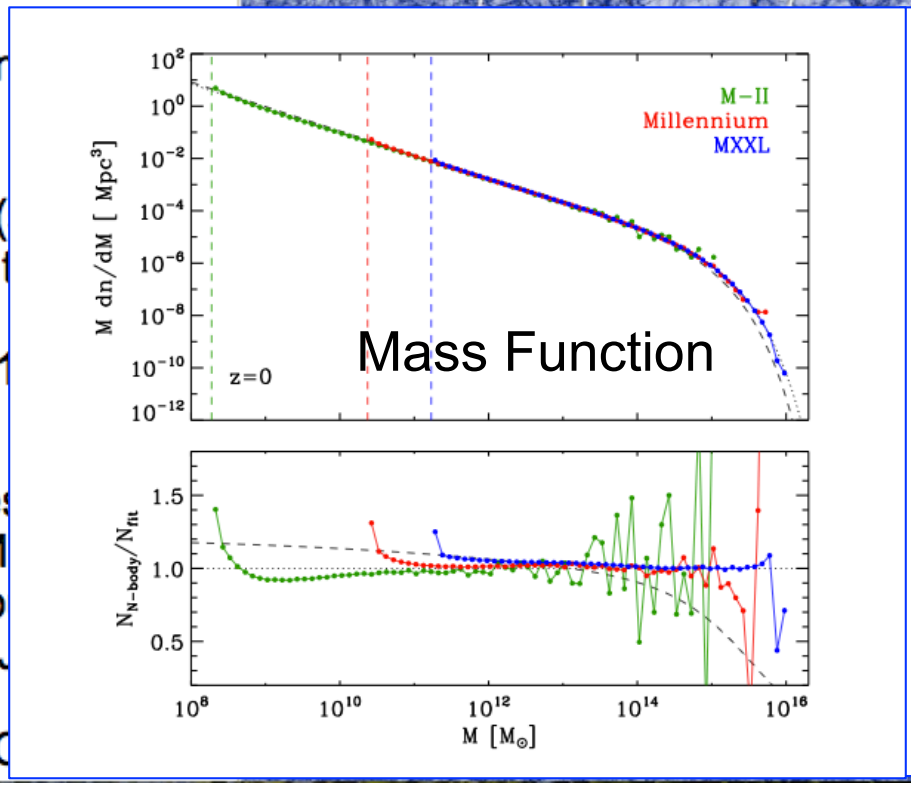
~25 billion (
in mergers t

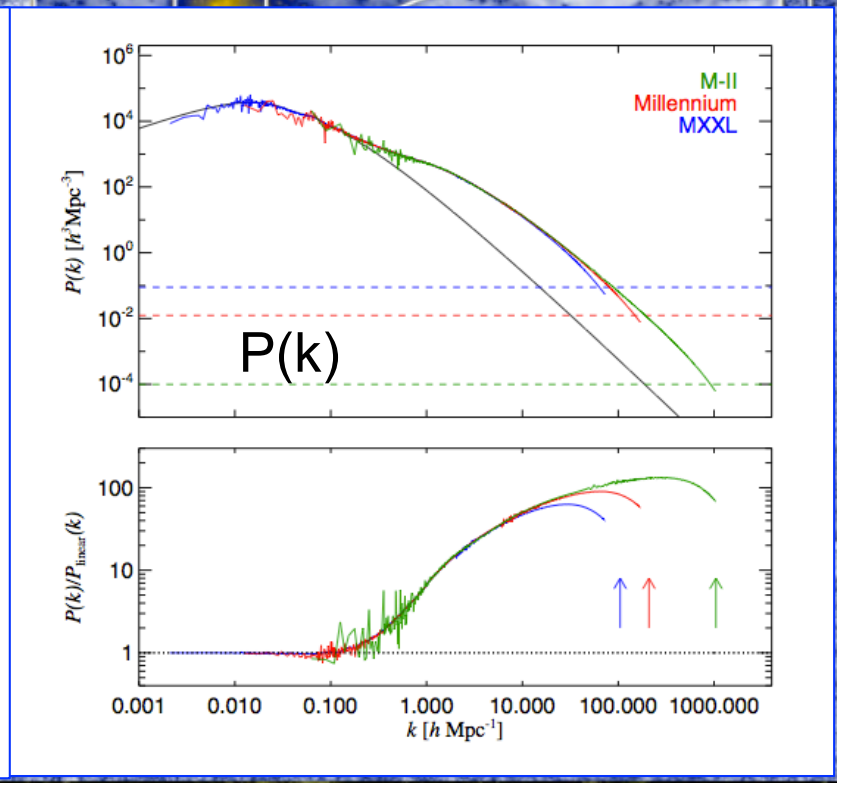
m_p = 6.1 x 1

12288 cores
30 TB RAM
Supercomp
JuRoPa in J

2.7 million C







The Big-MULTIDARK Simulation Suite

PRACE proposal 2012: G. Yepes (PI), F. Prada, S. Gottloeber, A. Klypin, S. Hess
22.5 million cpu hours in SuperMUC@ LRZ.

DATABASES
http://www.multidark.org
BigMD products
Publicly available soon

BigMD suite:

Volume: 2.5/h Gpc volume.
N particles: 3840^3 particles
Force Resolution: 10 kpc/h
N timesteps > 6000 (max $\Delta t < 0.001$)
Cosmological Parameters:

Ω_M=0.27, 0.29, 0.31
σ_8= 0.82, 0.9
n_s = 0.95
zinit= 100

First Planck Cosmology runs:

Ω_M=0.3071, σ_8= 0.8228, n_s=0.96
Zinit=100

Multidark & Small MD Planck1

Box = 1/h Gpc and 400/h Mpc@ 3840^3
Force = 5 kpc/h – 3kpc
Zinit=120-150
88 timesteps stored in each run
FOF and BDM halos. Light Cones

New Mass function calibration

MNRAS , (2013), 433, 1230.

credit: G.Yepes

The halo mass function through the cosmic ages

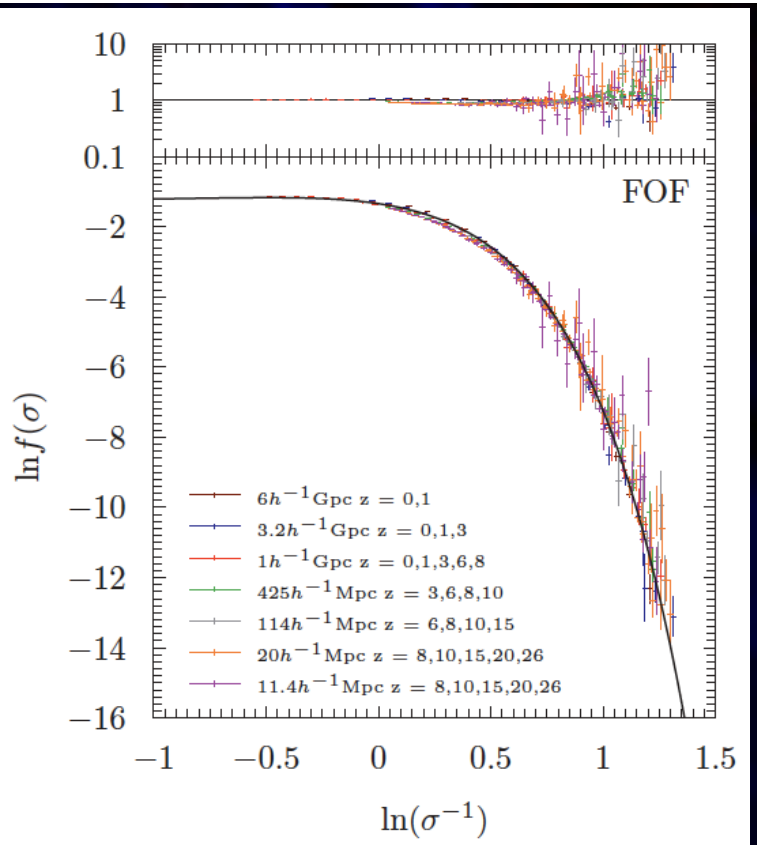
arXiv:1212.0095

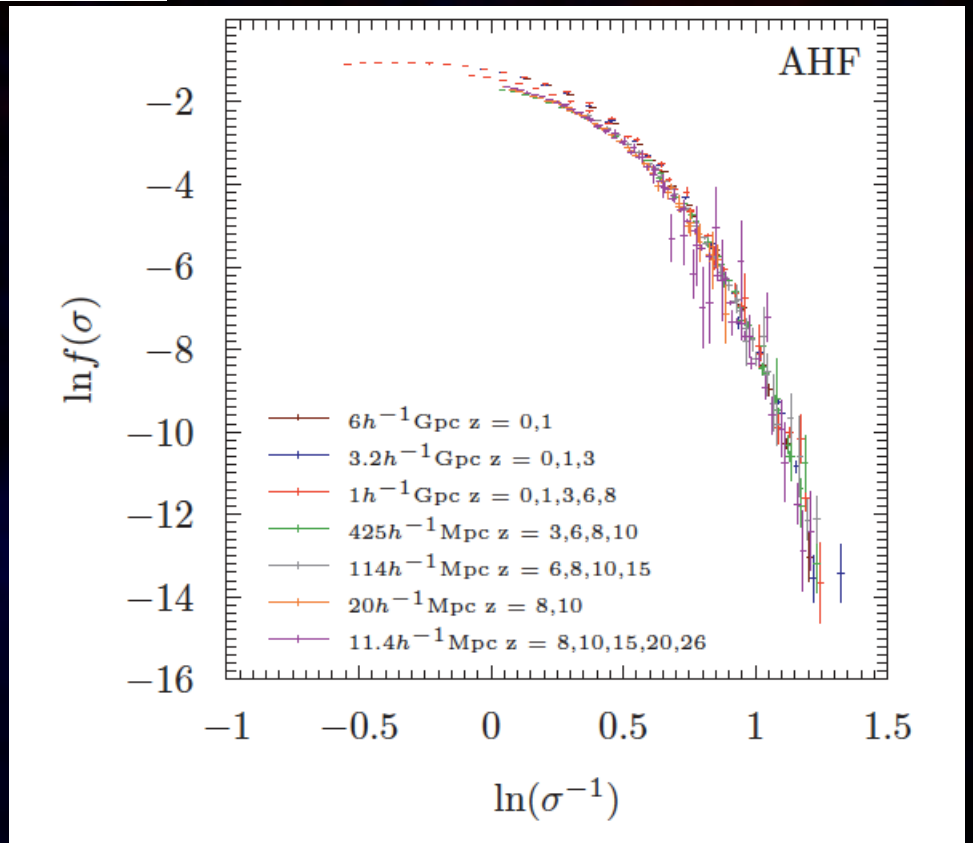
William A. Watson[1⋆], Ilian T. Iliev[1], Anson D'Aloisio[2], Alexander Knebe[3], Paul R. Shapiro[2] and Gustavo Yepes[3]

[1] *Astronomy Centre, Department of Physics & Astronomy, Pevensey II Building, University of Sussex, Falmer, Brighton, BN1 9QH, United Kingdom*

[2] *Department of Astronomy and Texas Cosmology Center, University of Texas, Austin, TX 78712, USA*

[3] *Departamento de Física Teórica, Modulo C-XI, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain*





Leibniz-Institut für Astrophysik Potsdam (AIP)

University of Sussex

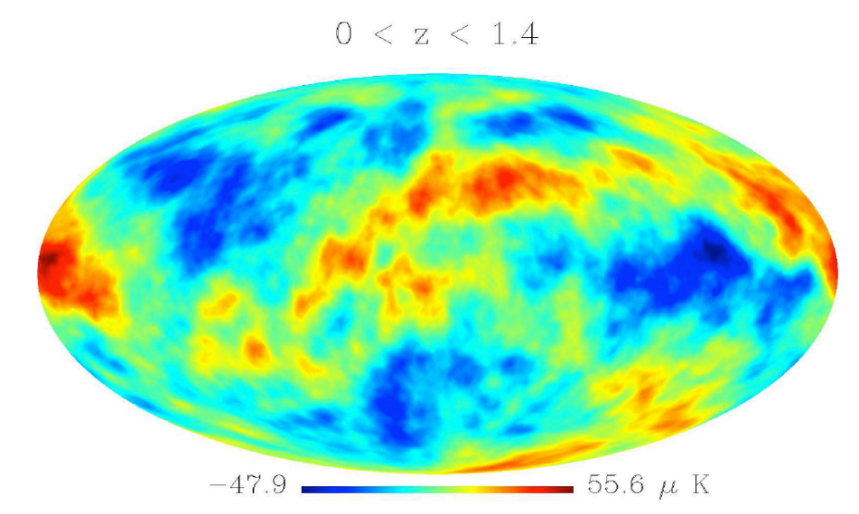
Universidad Autonoma de Madrid

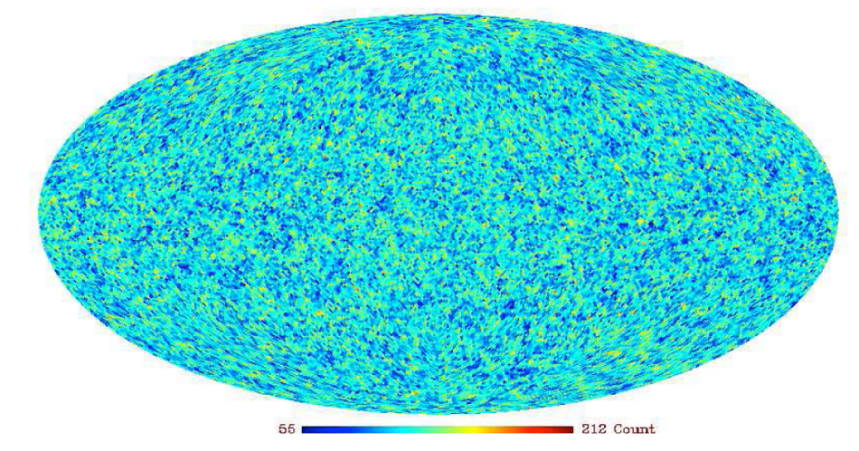
IFCA Instituto de Física de Cantabria

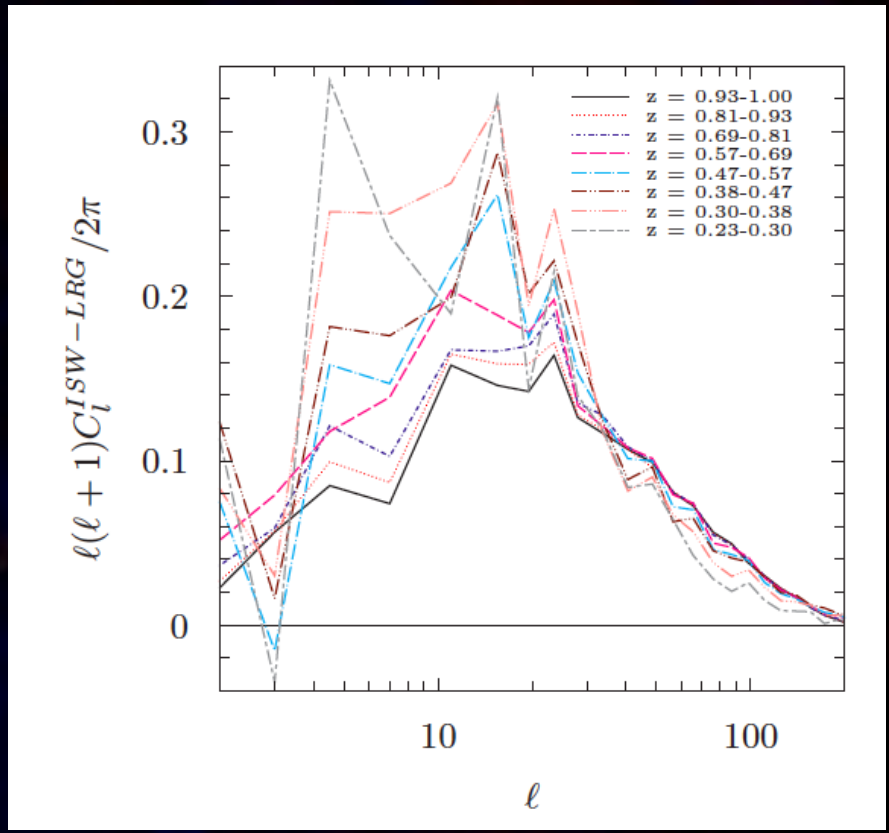
The Jubilee Simulation

Cross-correlation between ISW and LRGs

http://jubilee-project.org







credit: G.Yepes

Marenostrum Institut de Ciències de l'Espai Simulations

MICE

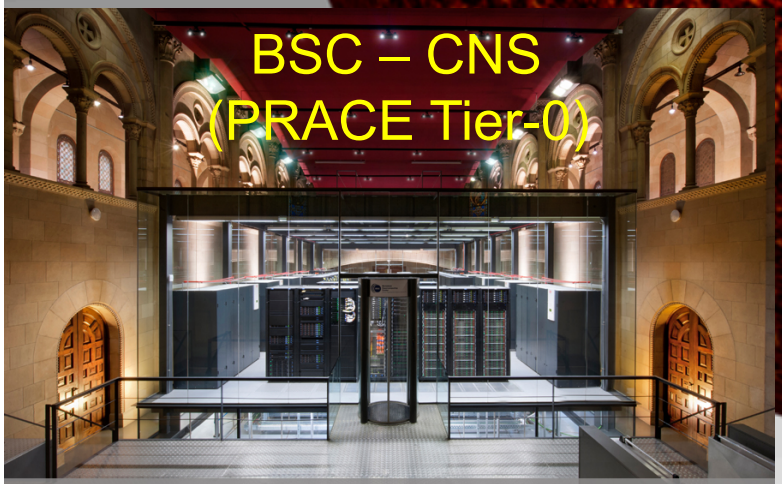
Cosmological Simulations @
Marenostrum Supercomputer
using 4000 processors

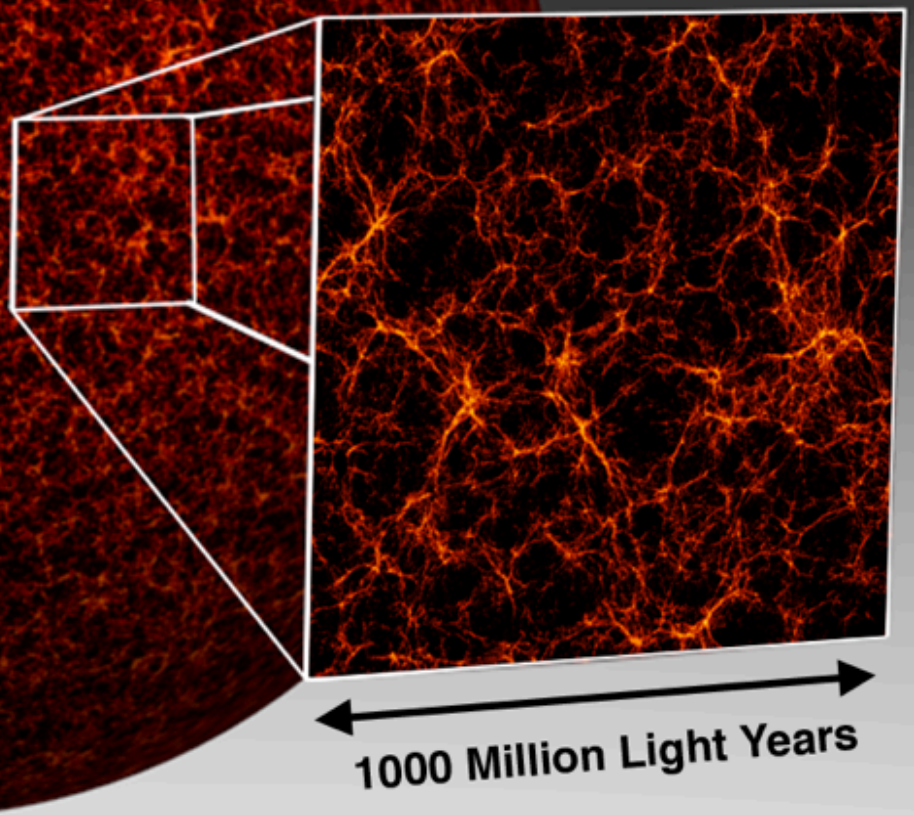
P.Fosalba

and the MICE Collaboration:

M.Crocce
F.Castander
E.Gaztanaga
M.Manera
D.Reed
C.Bonnett
J.Carretero
K.Hoffman

www.ice.cat/mice



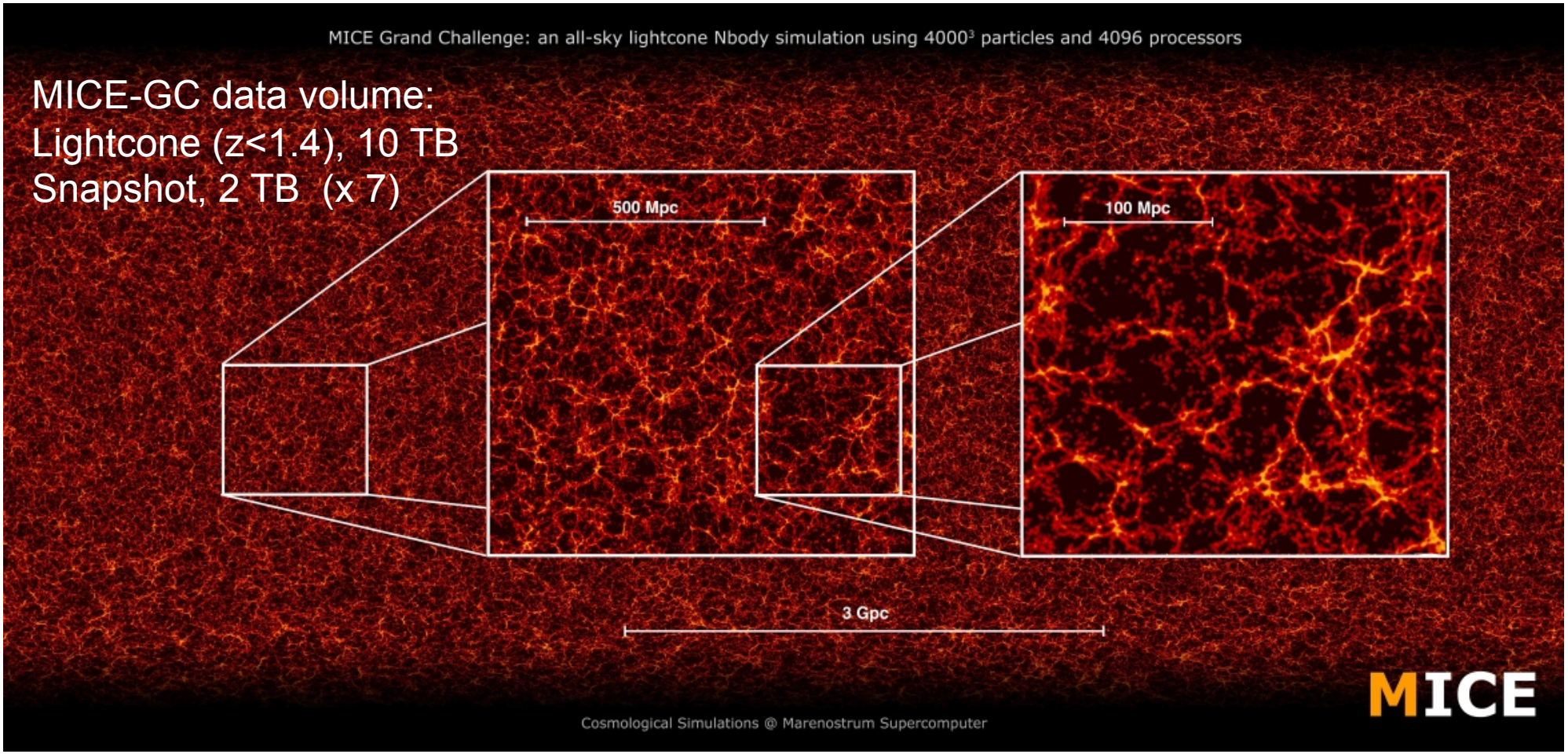


MICE-Grand Challenge simulation

Developed at Marenostrum @ BSC, used Gadget2, 4100 cores, 3 M-hours

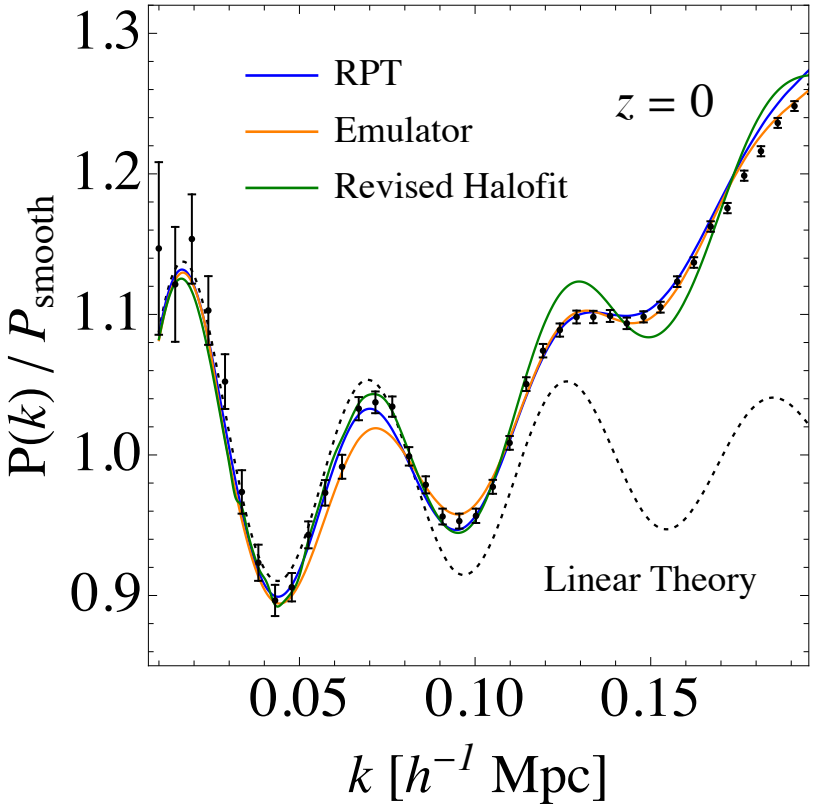
70 billion particles in a 3 Gpc/h box (50 kpc/h soft length)

Samples 5 decades in dynamic range



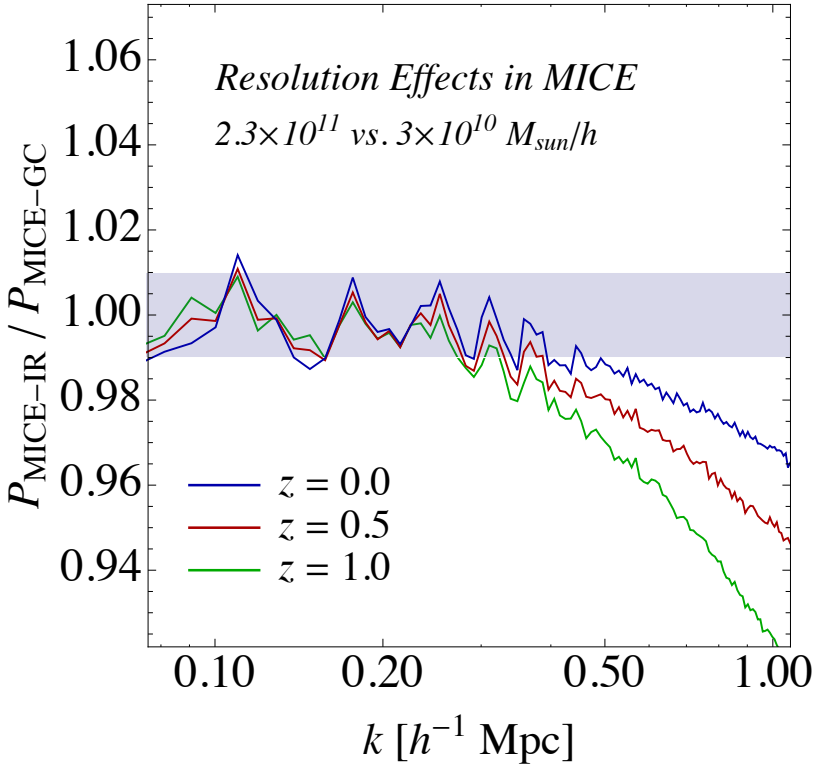
MICE-GC: Dark-Matter clustering

P(k) at BAO scales



✓ Agreement with hi-res numerical fits (Heitmann et al 2013; Takahashi et al 2012) within 2% accuracy

Mass Resolution Effects



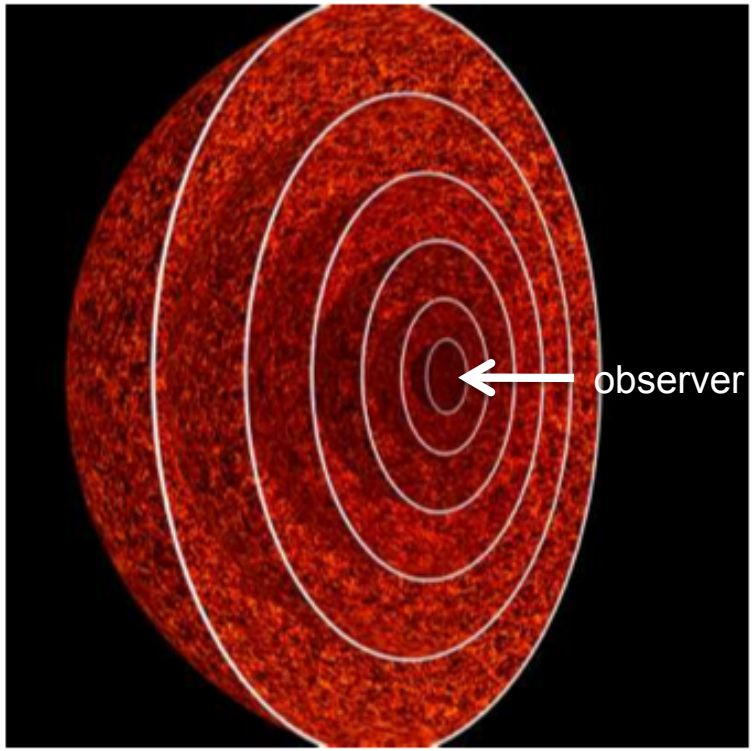
✓ 5-10% resolution effects at k>1 h/Mpc
✓ increase with redshift

Fosalba et al. 2013, arXiv: 1312.1707

MICE-GC: Lensing observables

" The onion universe: all sky light-cone simulations in spherical shells "

Fosalba et al., MNRAS, **391**, 435 (2008)



1. Divide 3D Lightcone data into redshift shells.
2. Project DM particle counts onto (Healpix) pixelized 2D maps
3. Add shells with WL kernel/weight

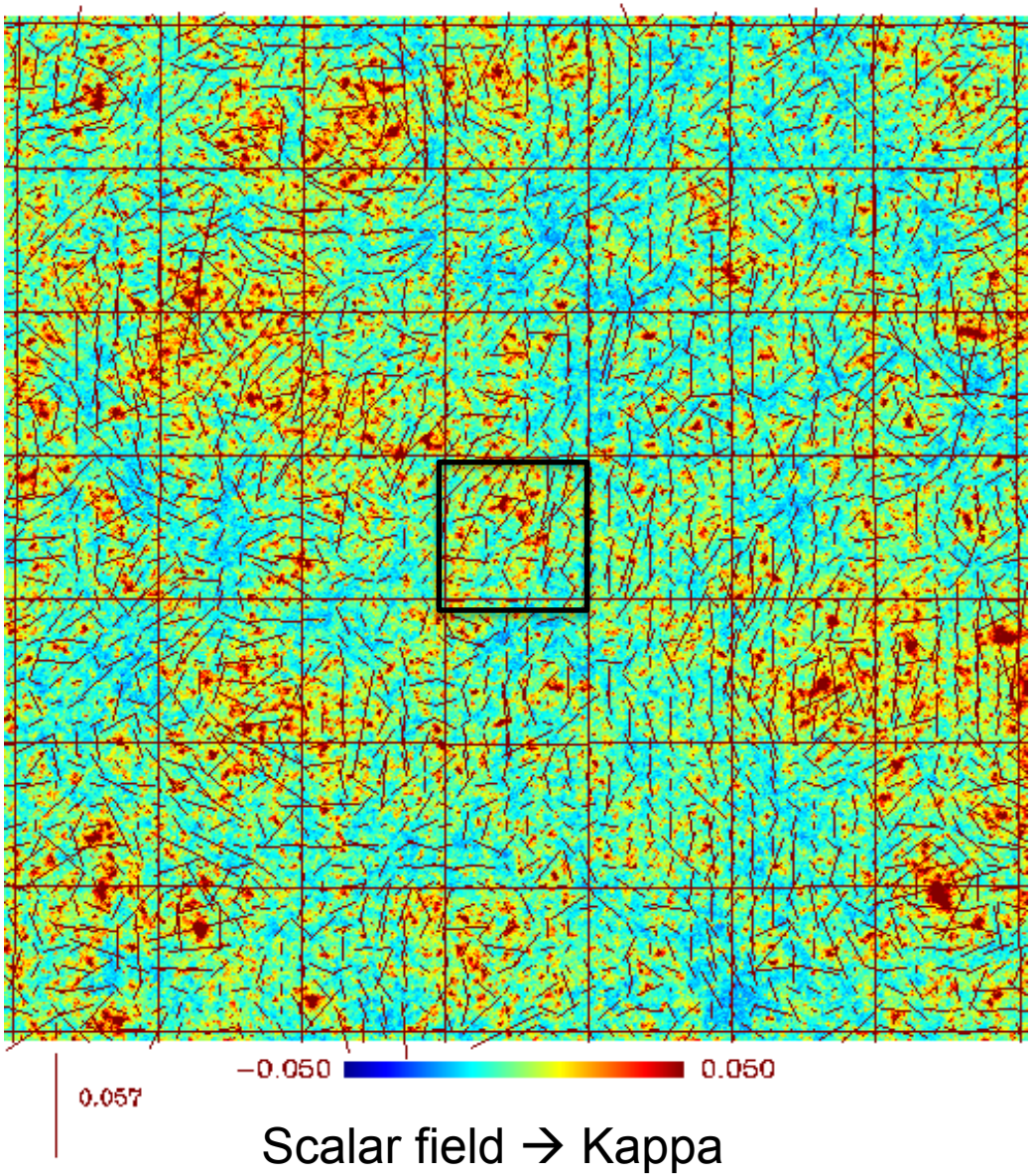
For the convergence/kappa:

$$\kappa(\theta) = \frac{3H_0^2\Omega_m}{2c^2} \int dr\ \delta(r,\theta) \frac{(r_s - r)r}{r_s\ a}$$

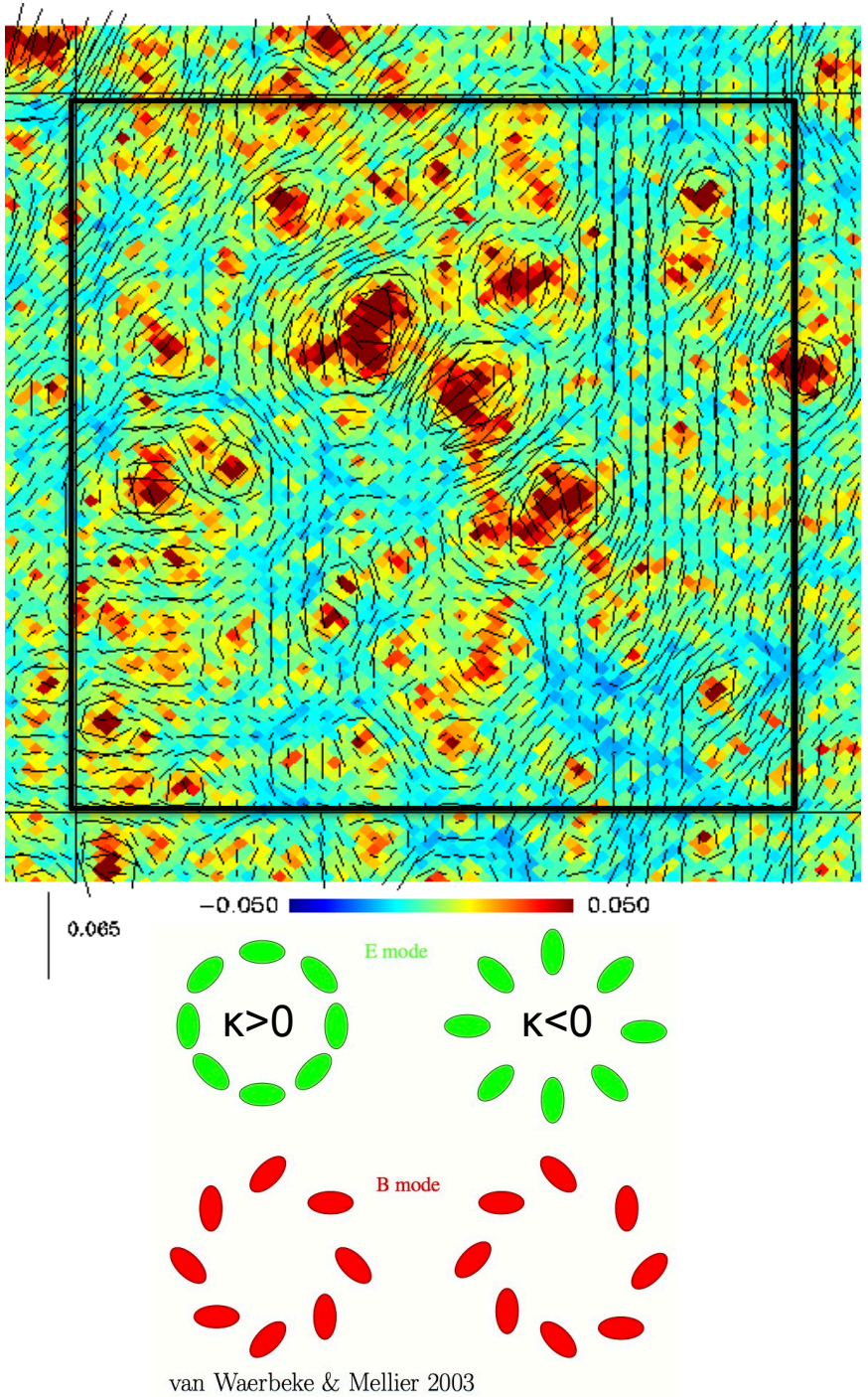
$$\kappa(i) = \frac{3H_0^2\Omega_m}{2c^2} \sum_j \delta(i,j)\ \frac{(r_s - r_j)r_j}{r_s a_j}\ dr_j$$

- Valid *in the Born approximation (ie, linear/uncorrelated deflections)*
  1% accurate at arcmin scale
- *Use simple relations in harmonic space to model other lensing observables*

7x7 sq.deg sources @ z=1



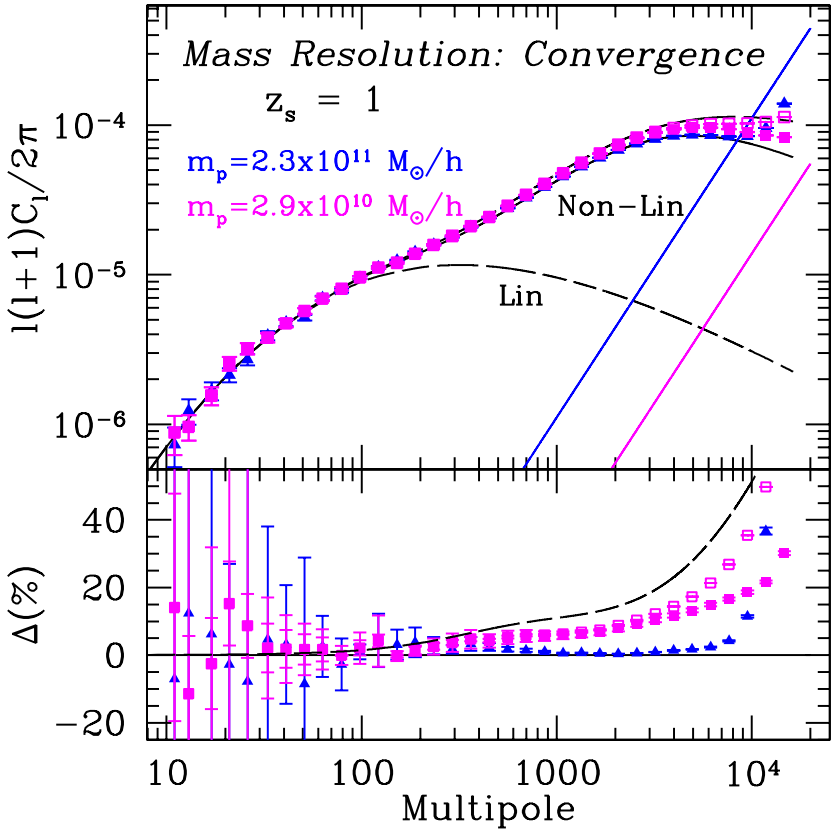
Zoom: central 1 sq.deg



Scalar field → Kappa
Vector field → Shear

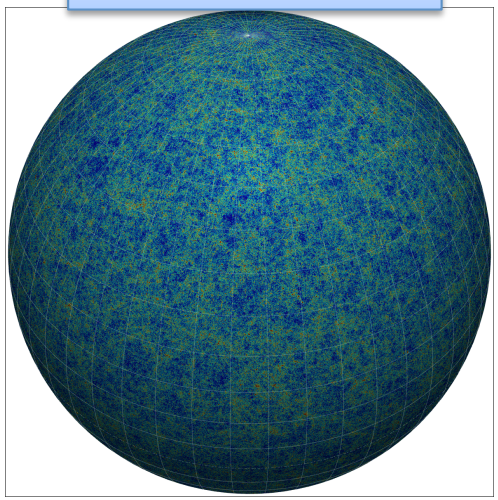
MICE-GC: Lensing statistics

all-sky maps of convergence, shear and magnification with < 1 arcmin resolution

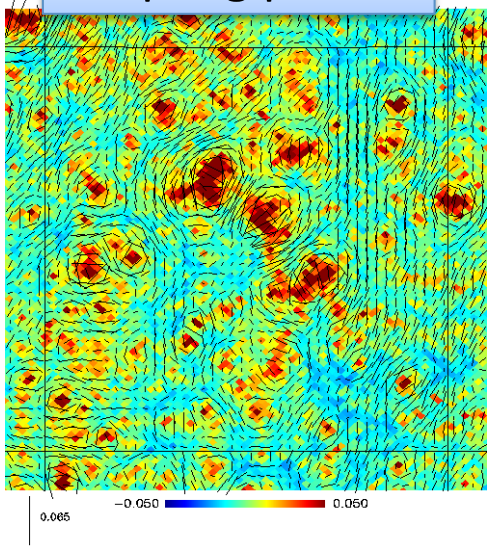


Magenta: MICE-GC
Solid: Smith et al 2003
Dashed: Takahashi 2012

All-sky maps



1 sq.deg patch



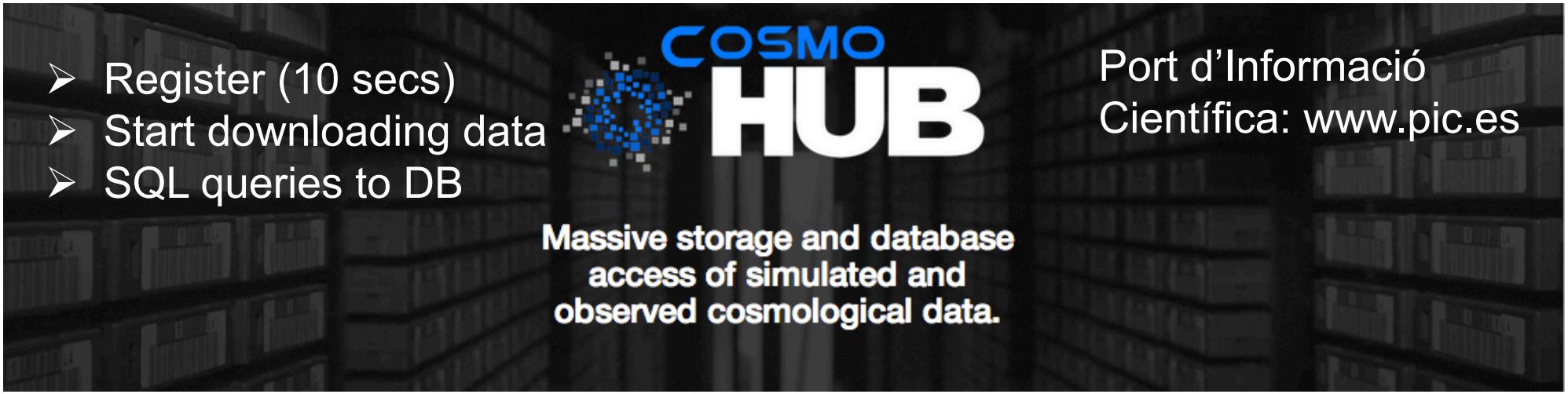
mass resolution < 10^{10} M$_\odot$ needed to resolve lensing power to ~10-20% accuracy at few arcmin scales(multipoles > 1000)

Fosalba et al. 2013, arXiv:1312.2947

MICE-GC simulation papers/data

Paper I: Dark-matter clustering [Fosalba et al. 2013a, **arXiv:1312.1707**]
Paper II: Halo and galaxy clustering [Crocce et al. 2013, **arXiv:1312.2013**]
Paper III: Galaxy lensing mocks [Fosalba et al. 2013b, **arXiv:1312.2947**]

MICE data publicly available at CosmoHub portal:
http://cosmohub.pic.es



Towards STAGE IV Survey Simulations

Fundamental Cosmology,
Fuerteventura June 5, 2014

Simulations of Non-standard models

Status: 2012

[source: Marco Baldi]

1. Quintessence and Early DE
2. Inhomogeneous large-voids (LTB)
3. WDM
4. NG initial conditions
5. Massive neutrinos
6. Self-interacting DM
7. Linear spatial DE fluctuations
8. Non-linear spatial DE fluctuations (MG)

complexity

- Nbody codes already developed
- Partially developed
- Mostly TB developed

Simulations of Non-standard models

Status: 2014

1. Quintessence and Early DE
2. Inhomogeneous large-voids (LTB)
3. WDM
4. NG initial conditions
5. Massive neutrinos
6. Self-interacting DM
7. Linear spatial DE fluctuations
8. Non-linear spatial DE fluctuations (MG)

complexity

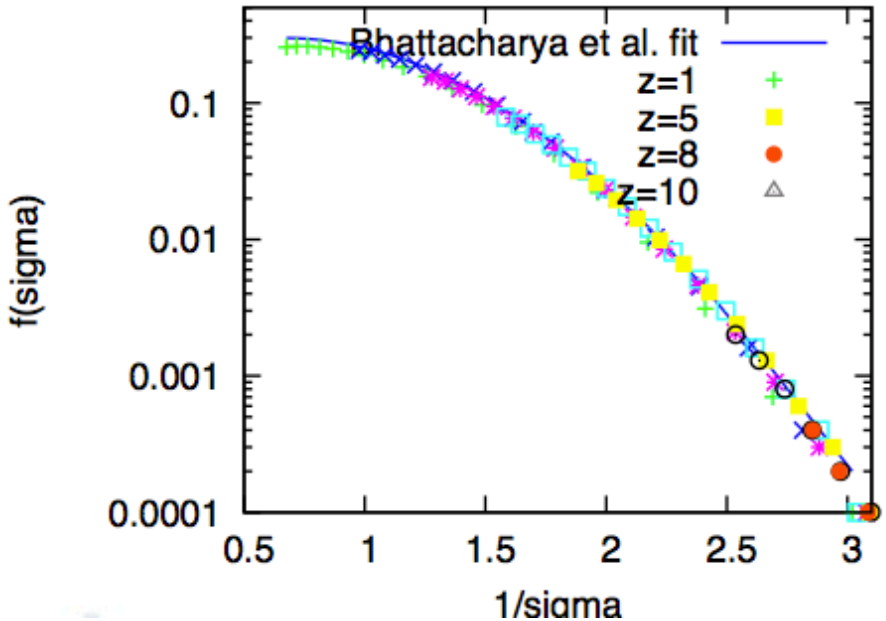
- Nbody codes already developed
- Partially developed
- Mostly TB developed

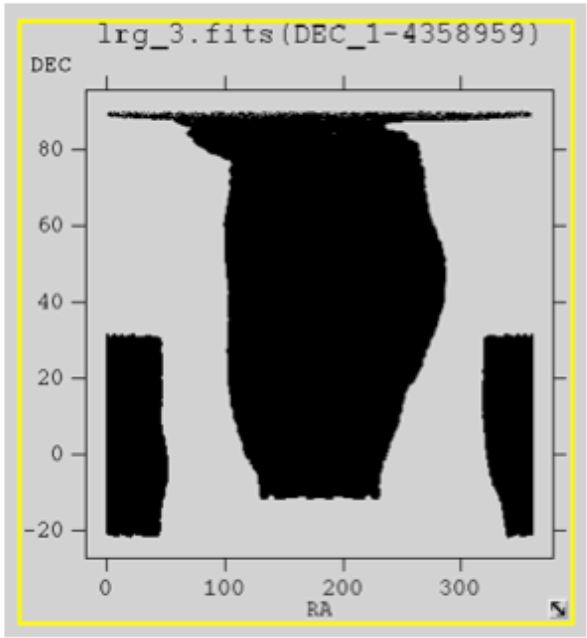
All need to be upgraded for Stage IV experiments!

Heitmann et al.
@ Argonne

The Outer Rim Simulation

- **Simulation parameters:** 4.225 Gpc (3Gpc/h), 1.1 trillion particles, ~$10^9 M_\odot$ (WMAP-7)
- Running on Mira (Argonne BG/Q), reached z~0.4, 100 snapshots starting at z~10, halo catalogs generated
- Results used for creating 8 DESI mocks containing 8M LRGs, 40M ELGs, 3M QSOs, available on NERSC repository (mocks created by Martin White)





BIG DATA

State-of-the art simulations (MXXL, Horizon, MICE, DEUS…):
DM snapshot ~ 2 TB
DM 3D Lightcone ~ 20 TB (up to z=2)

STAGE IV survey simulation is ~100 times larger…
Looking for PetaBytes of data per run

- ✧ Requires HPC facilities for simulation development, storage
- ✧ Nbody codes will need to be upgraded
- ✧ Analysis codes will need to be upgraded + work on the fly to avoid massive data storage and post-processing

We need large HPC access and people to develop this in next few years!!

Summary

- Percent accuracy in NL predictions (k>1 h/Mpc) from Nbody simulations/ Emulators has not yet been achieved (but we are close…)
- Huge challenges ahead: next frontier galaxy surveys demand modeling larger volumes with increasing resolution
- Good News: Few trillion particle simulations feasible within next 2 years
- Simulation and analysis codes to be upgraded to exascale (need HPC experts!)
- Coordinated access to HPC infrastructures is key to successful development plan
- New tools for Science exploitation: analysis increasingly based on efficient access to Simulated data Databases