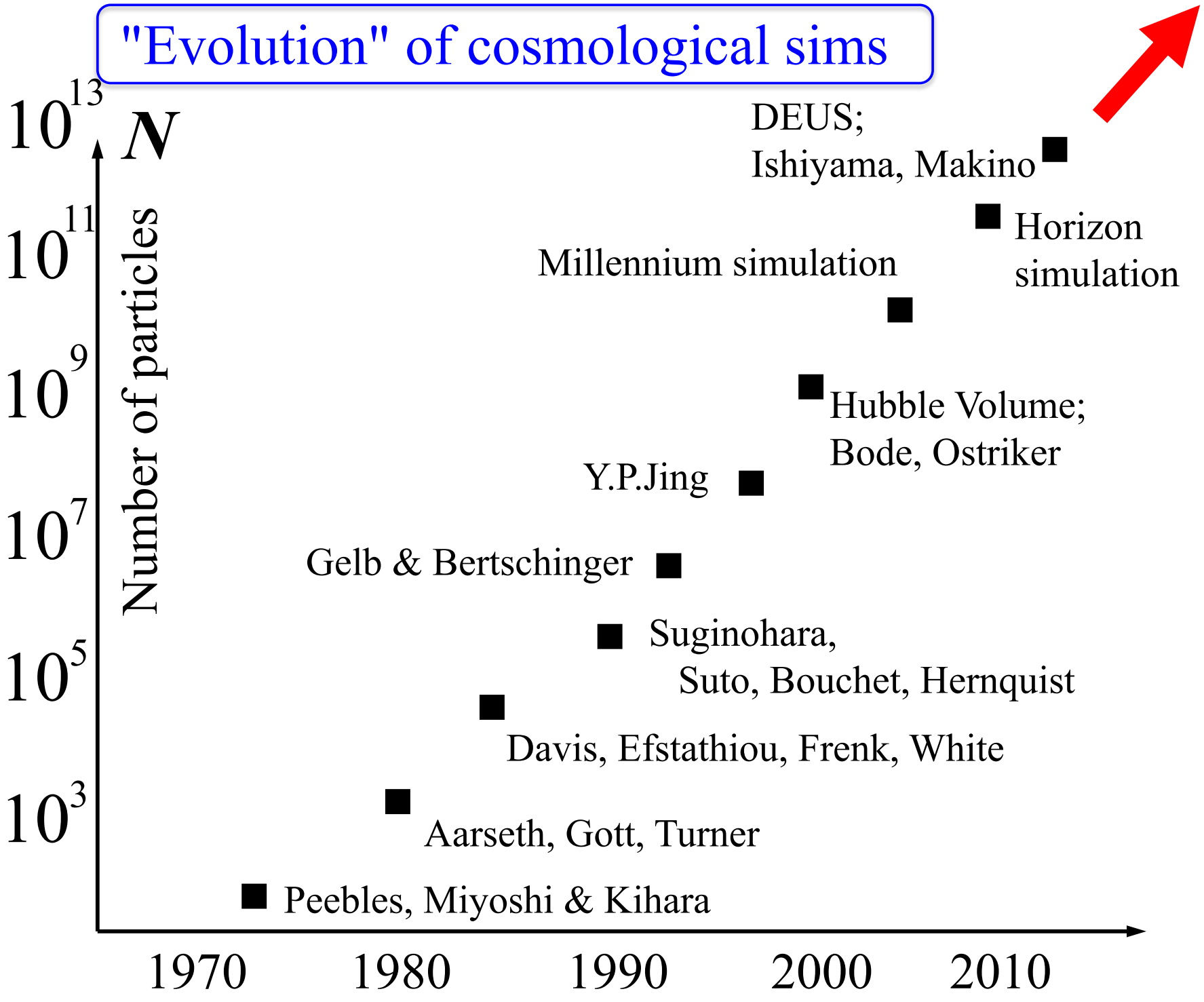
# Statistical Computational Cosmology with Big Astronomical Data

Naoki Yoshida (U-Tokyo)
Takahiro Nishimichi (Kyoto-U)
Satoshi Tanaka (U-Tsukuba)

Priority Issue 9, Sub-project C


"Evolution" of cosmological sims
$10^{13}$
$N$
$10^{11}$
$10^{9}$
$10^{7}$
$10^{5}$
$10^{3}$
Number of particles
DEUS;
Ishiyama, Makino
Horizon
simulation
Millennium simulation
Hubble Volume;
Bode, Ostriker
Y.P.Jing
Gelb & Bertschinger
Suginohara,
Suto, Bouchet, Hernquist
Davis, Efstathiou, Frenk, White
Aarseth, Gott, Turner
Peebles, Miyoshi & Kihara
1970
1980
1990
2000
2010

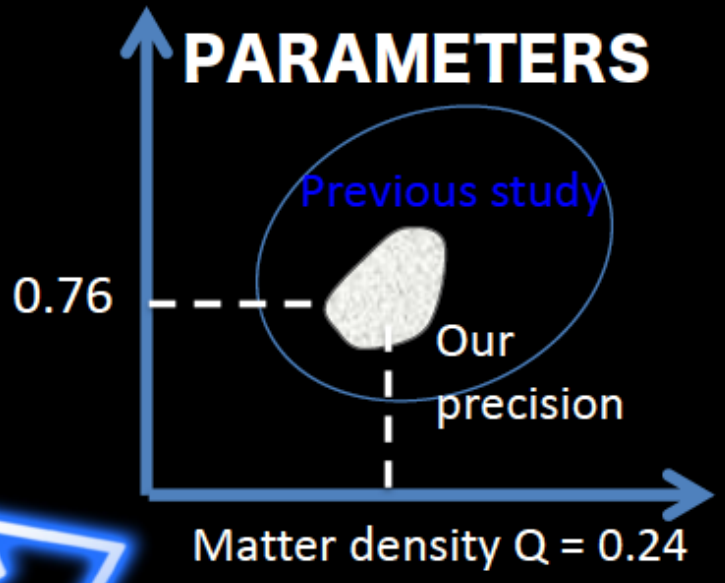
Simulation database (UTokyo)

Statistics (ISM&NTT)

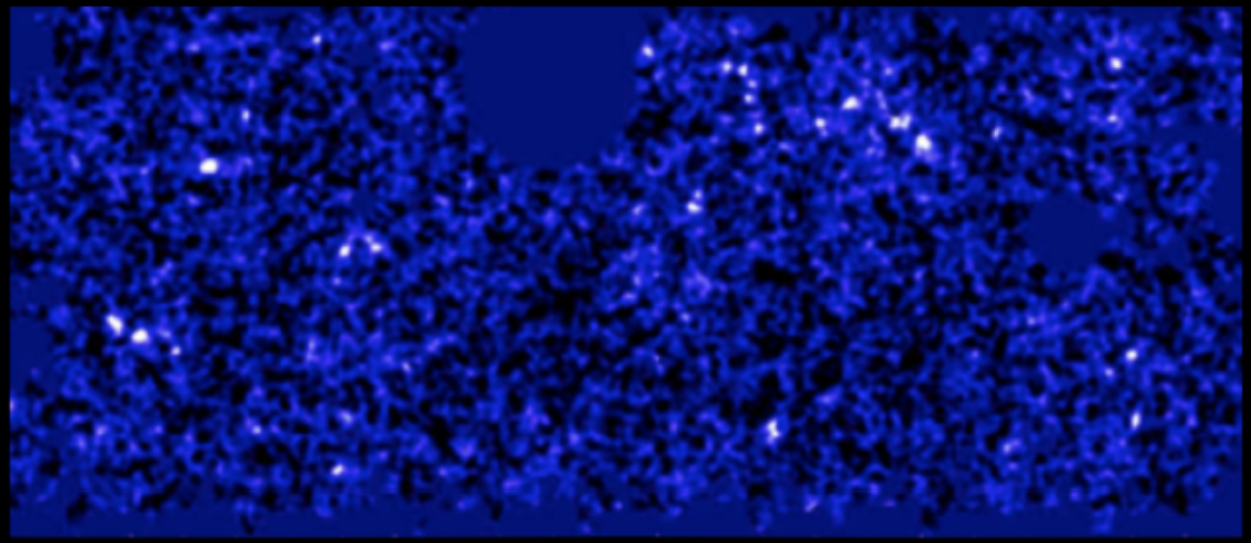
Parallel analysis tools (Tsukuba)

EMULATOR

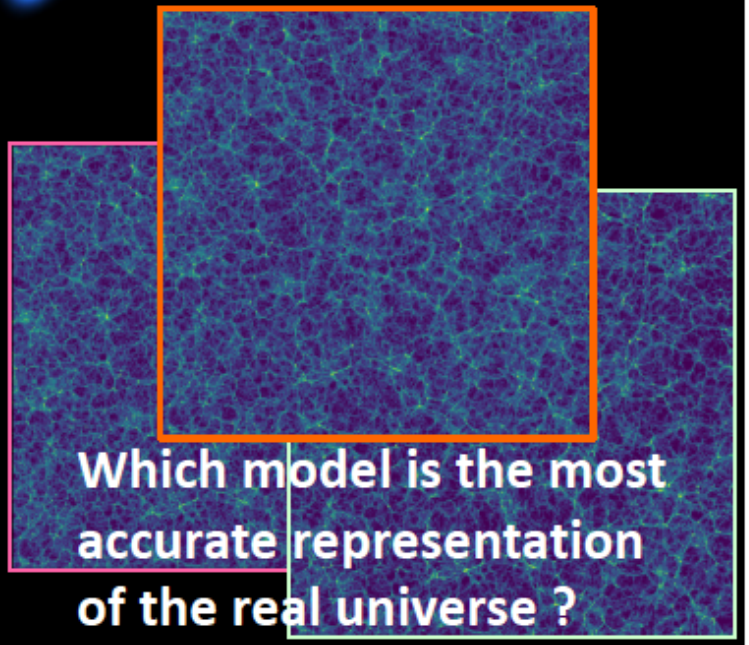
FUNDAMENTAL COSMOLOGICAL PARAMETERS




DARK MATTER MAP

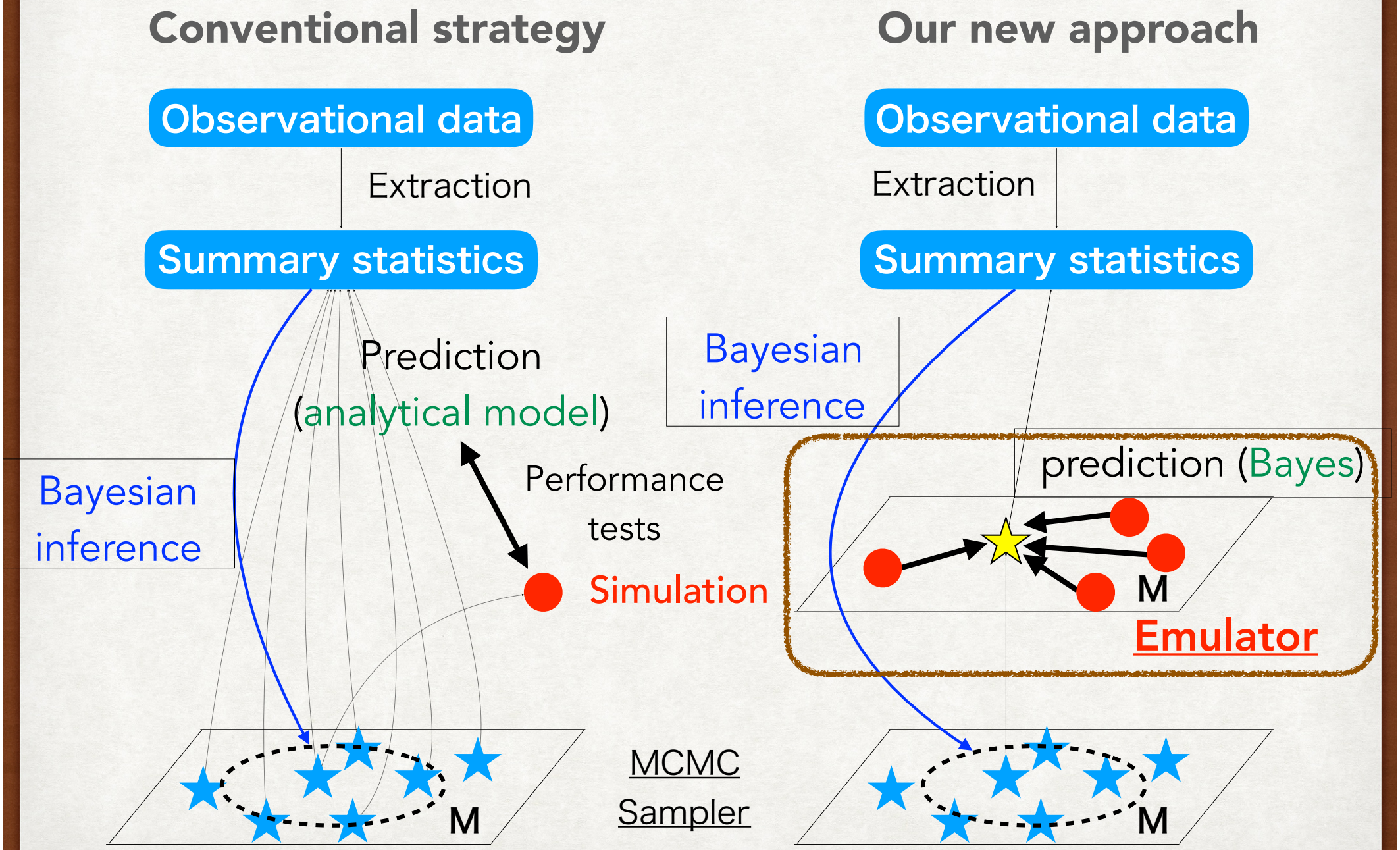


Theory models



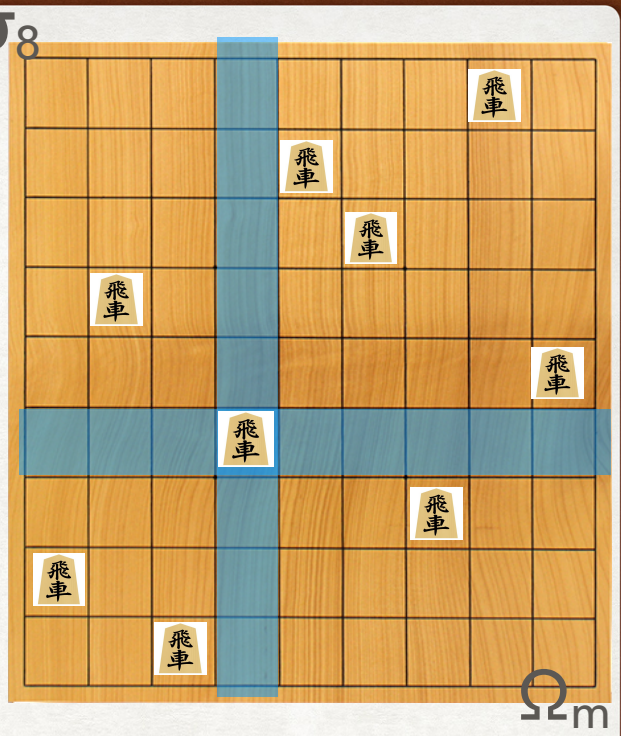
Which model is the most accurate representation of the real universe ?

# COSMOLOGICAL ANALYSIS MODEL

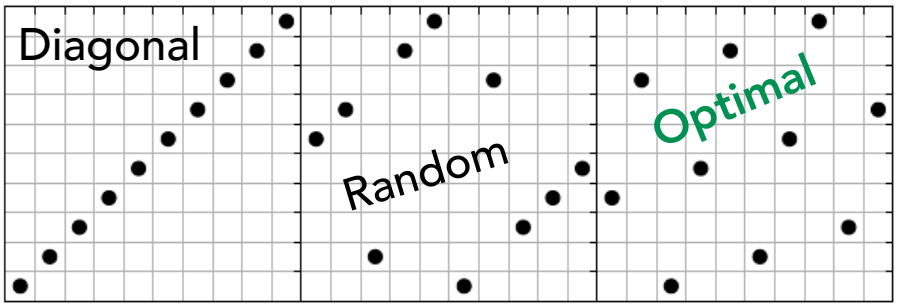


# DARK QUEST: SIMULATION DESIGN

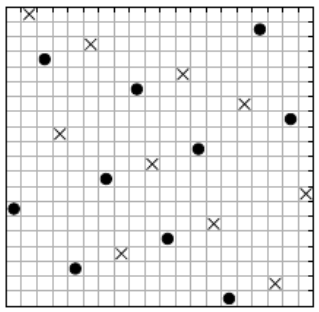
- Curse of dimensionality (input = 6D)
  - Regular lattice is not tractable in high dimensions
- Latin Hypercube Designs (LHDs)
  - Each sample point is the only one both on the row and on the column
  - Uniform sampling when projected onto any one axis
- LHD is not unique and not always efficient
  - One more to add: space filling property: "the closest neighbor should be far"
- A variant useful for ML problems




Possible LHDs



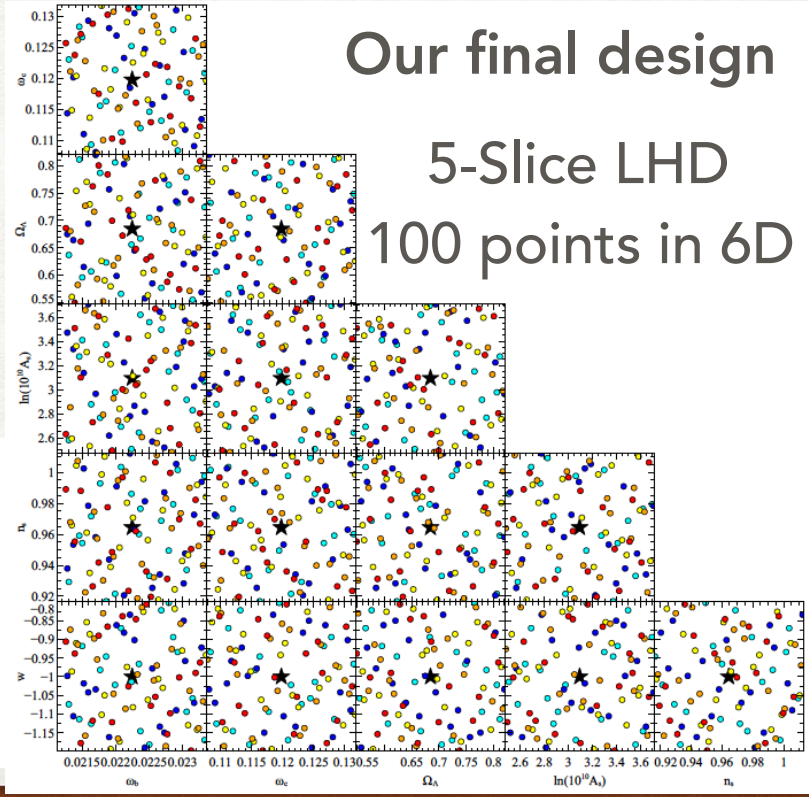

2-Slice LHD



Our final design

5-Slice LHD

100 points in 6D



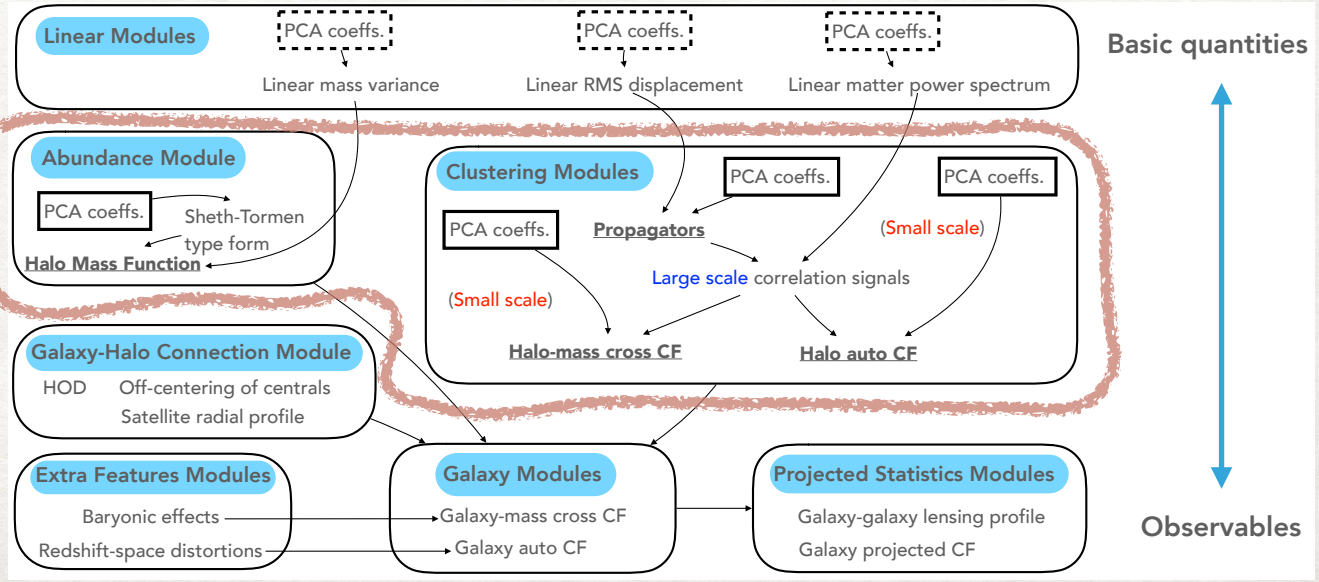
# HYBRID FORWARD MODELING DESIGN

- **Requirements**
  - Accuracy: a few percent level
  - Speed: seconds / evaluation (e.g., 2 days / simulation)
  - Flexibility: capture unknown effects in galaxy-matter connection
- **Our solution: Dark Emulator (= Simulations + Statistics)**
  - Network based on analytical relations
  - Dimension reduction: Principal Component Analysis
  - Core: Gaussian Process Regression

Analytical calculation

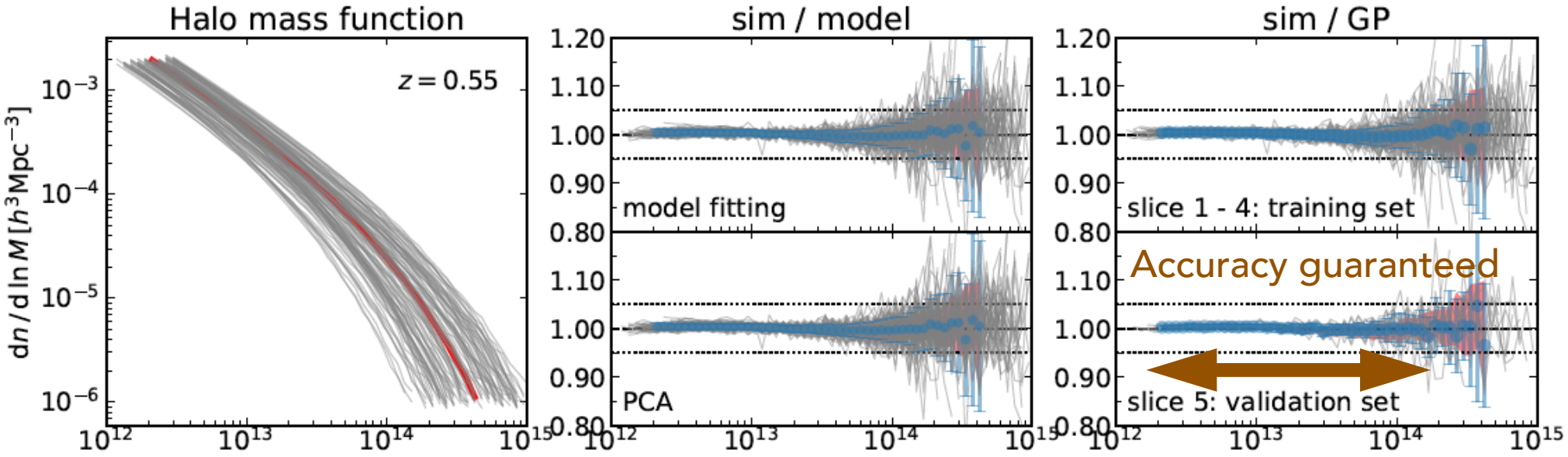
**Predictions by simulations**

Our recipe to connect the simulation world to the reality

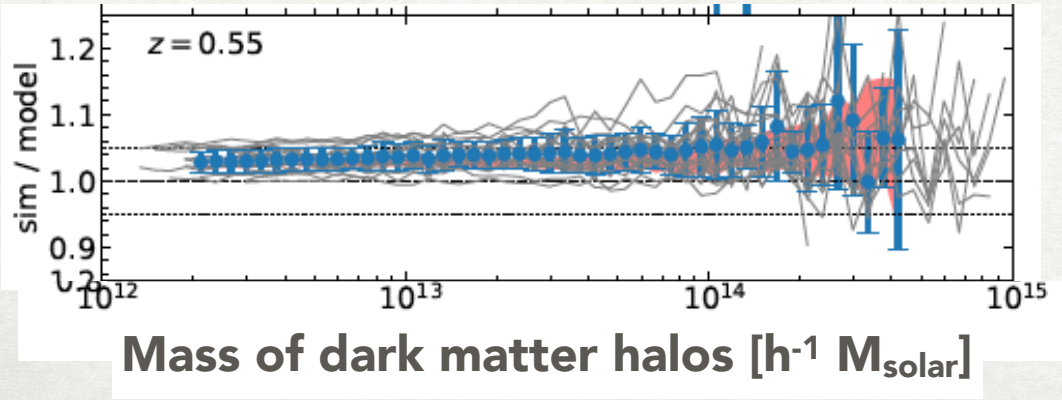


# CROSS VALIDATION STUDY EXAMPLE

## Abundance of structures (80 training, 20 validation)




**(vs model by Tinker et al. 08')**




**Accuracy**: better than 3% for the relevant statistics

vs. ~10 - 15% from existing best models

# A NOVEL APPROACH IN A SIX-DIMENSIONAL PHASE-SPACE

# Physics and math of a self-gravitating system

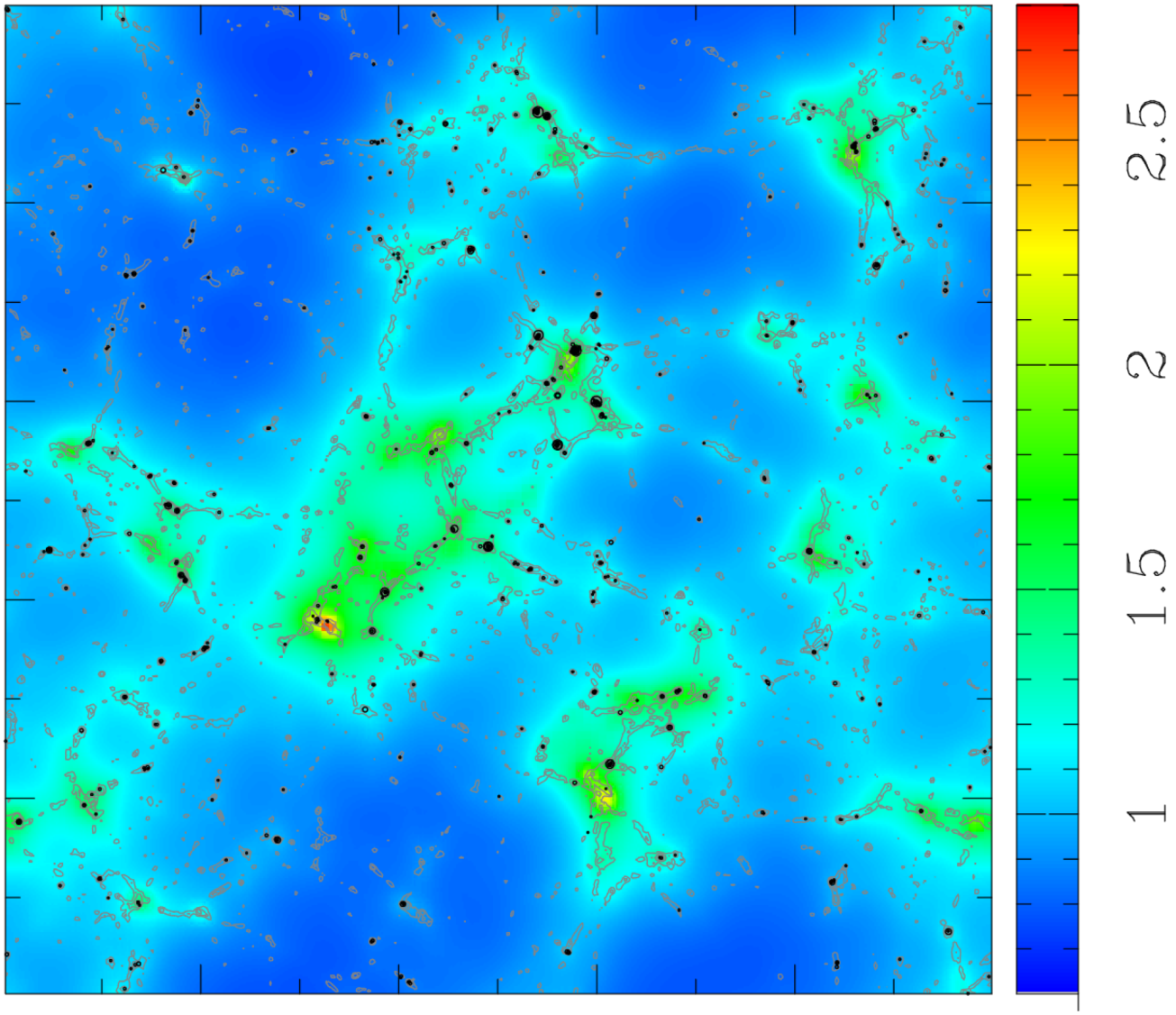
Collisionless *N*-body simulations closely follow the *derivation* of the collisionless Boltzmann equation, but do not directly solve

$$\frac{\partial f}{\partial t} + \vec{v} \cdot \frac{\partial f}{\partial \vec{x}} - \nabla\phi \cdot \frac{\partial f}{\partial \vec{v}} = 0$$

$$\nabla^2\phi = 4\pi G\rho = 4\pi G \int f d^3\vec{v}$$

It'd be nice if the evolution of $f(x, y, z, u, v, w)$ is directly followed in 6D phase-space.

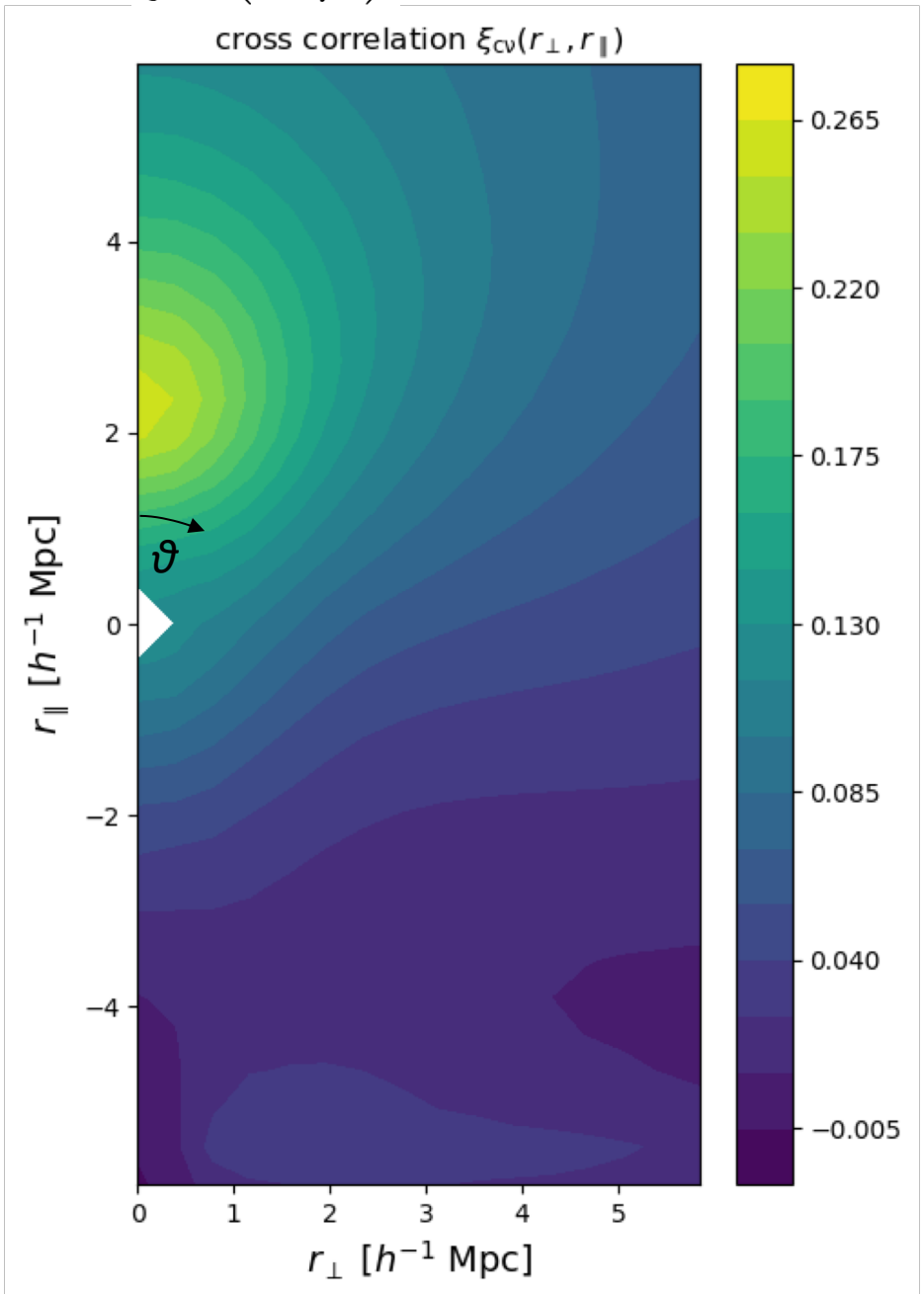
# Neutrino Distribution

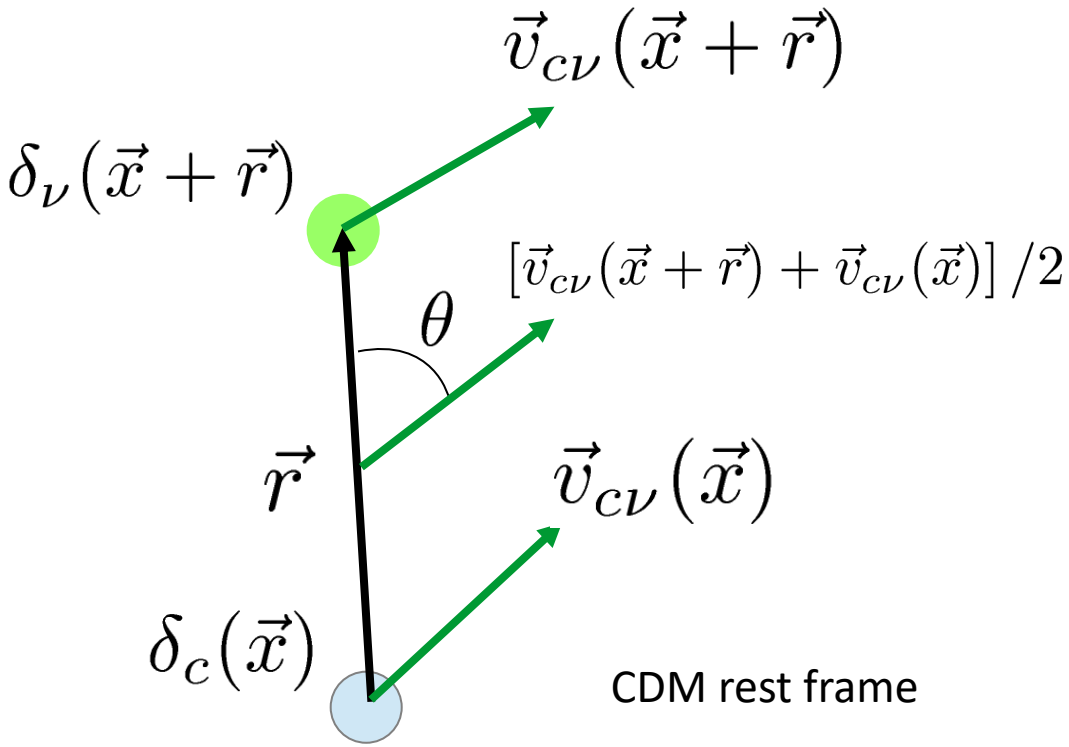



- color : neutrino overdensity
- contour: CDM overdensity
- black circle: DM halo with $M>10^{11}$ solar mass

# Cross correlation of CDM and neutrinos

$\xi_{c\nu}(r,\mu)$ $\mu \equiv \cos\theta$



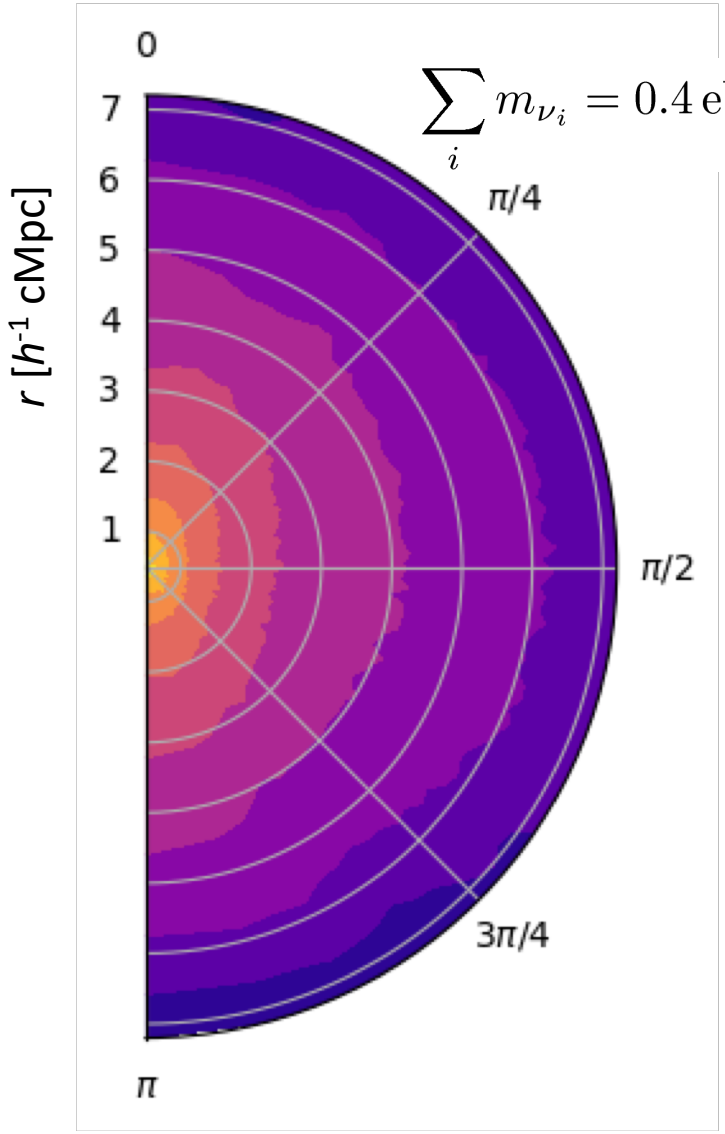



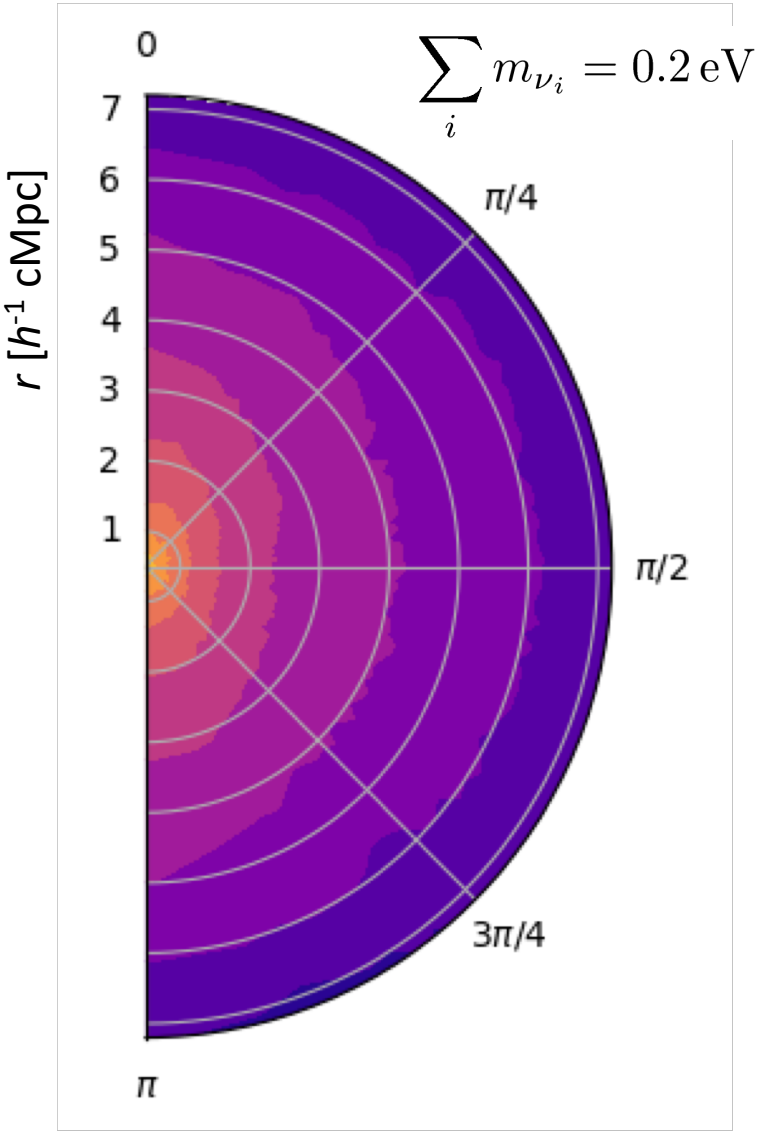

$r_\perp = r\sin\theta$

$r_\parallel = r\cos\theta = r\mu$

- Excess of cross-correlation in the down stream side of relative velocity due to neutrino wakes

# Probing the neutrino mass with cross-correlation



# SUMMARY

Wide-field sky survey probes a large volume of our universe

Numerical simulations play a vital role in determining cosmology

There are a variety of new approaches to reveal cosmic structure formation