Super-Resolution Simulations

Yin Li & Yueying Ni

with Rupert Croft, Tiziana Di Matteo, Simeon Bird, & Yu Feng

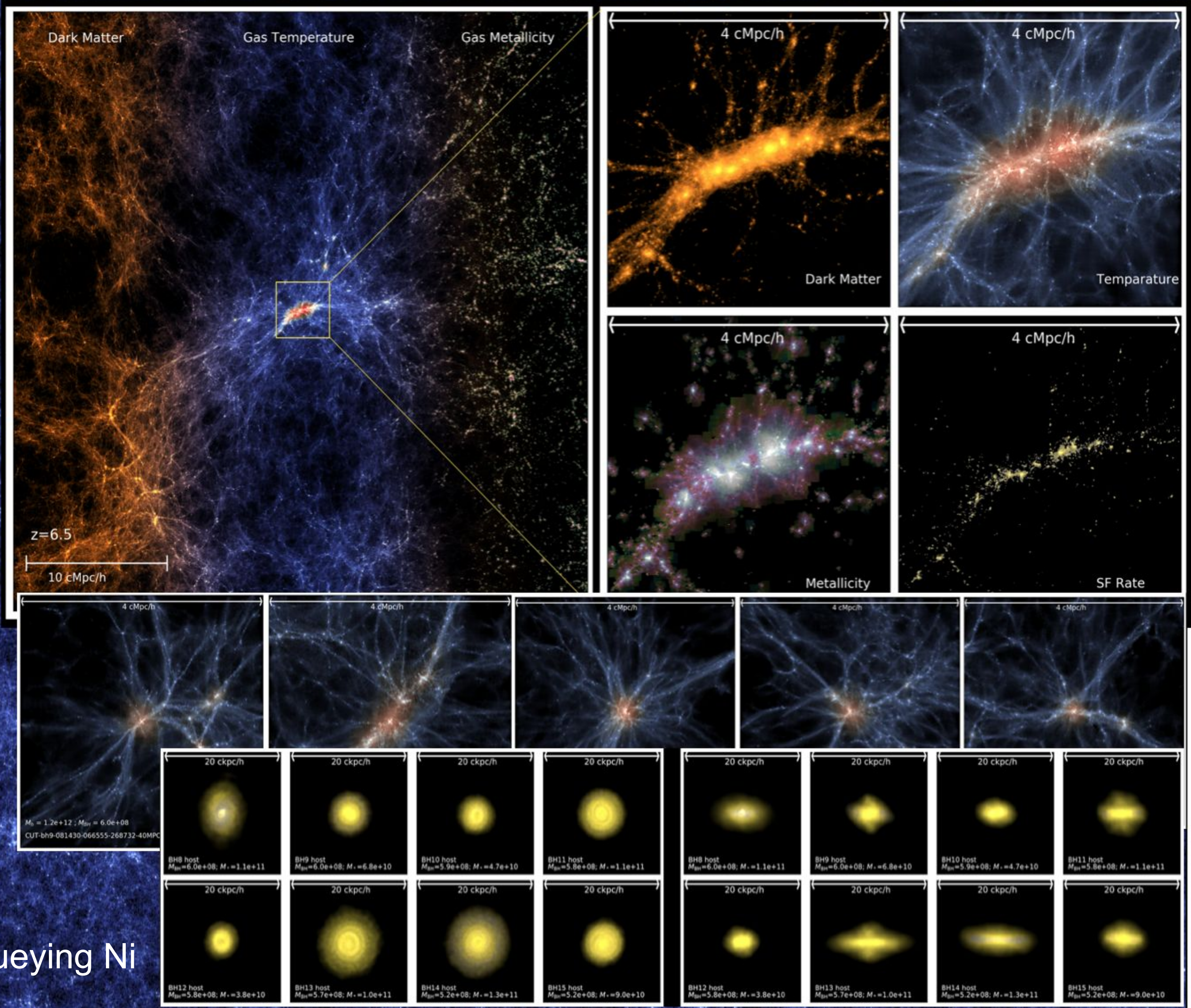
2010.06608 PNAS May 11, 2021 118 (19), 2105.01016

Why super resolution (SR)

- Cosmological and hydrodynamical simulations are expensive

Simulation	Number of particles	Box size
Dark Sky	1 trillion	8 Gpc/h
Outer Rim	1 trillion	4.225 Gpc
QContinuum	0.5 trillion	1.3 Gpc
LastJourney	1 trillion	5.025 Gpc
Uchuu	2 trillion	2 Gpc/h
...		

BlueTides Simulation



$L = 400\mathrm{Mpc}/h$
$N = 2 \times 7040^3$
$z = 6.5$

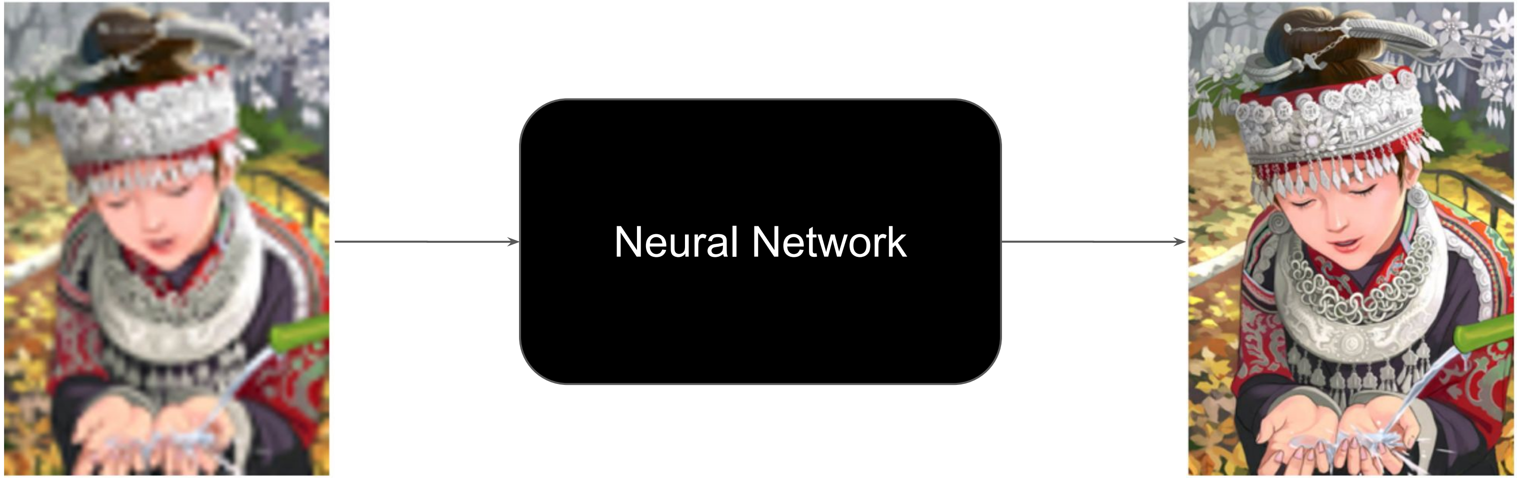
Image by Yueying Ni

Why super resolution (SR)

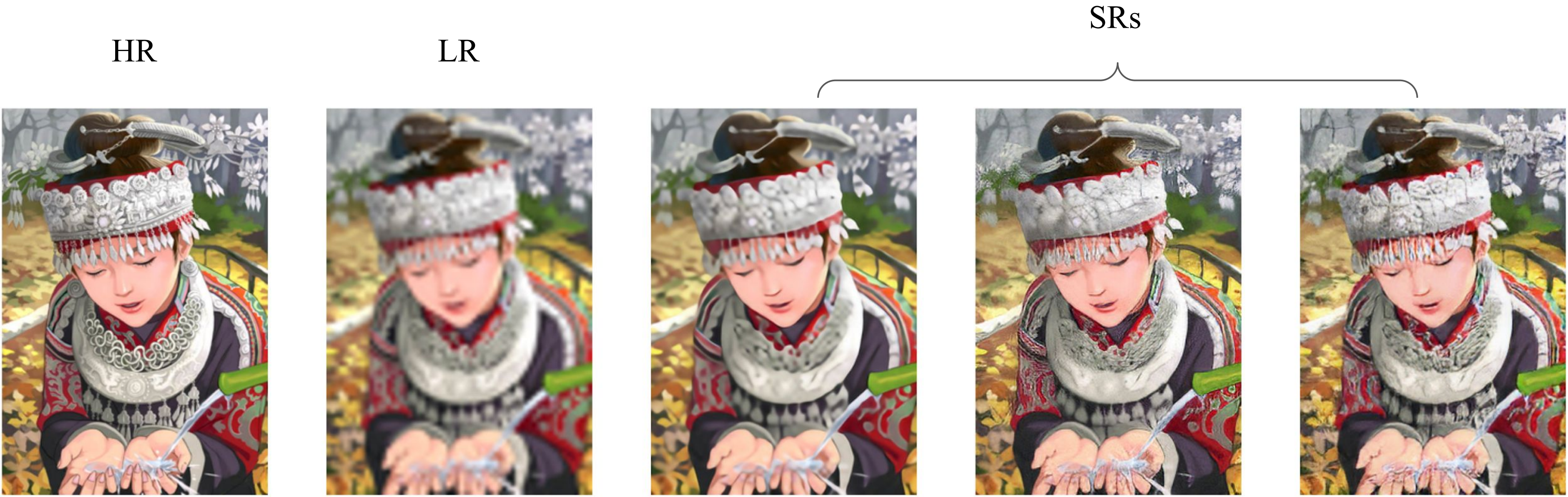
- Cosmological and hydrodynamical simulations are expensive
  - large dynamic range, nonlinear, multiscale
  - time complexity ~ $\mathcal{O}$(num_particles × time_steps)
    - num_particles ~ volume / resolution
- Need for both volume and resolution
  - Larger volume
    - better statistics
    - long-short mode coupling

      (k_{long}->0 limit aka the *super-sample* effect, Takada & Hu 2013, Li, Hu, & Takada 2014)
  - Higher resolution
- Compromise on either, or both?

What is SR — Deep learning image super resolution



What is SR — Deep learning image super resolution

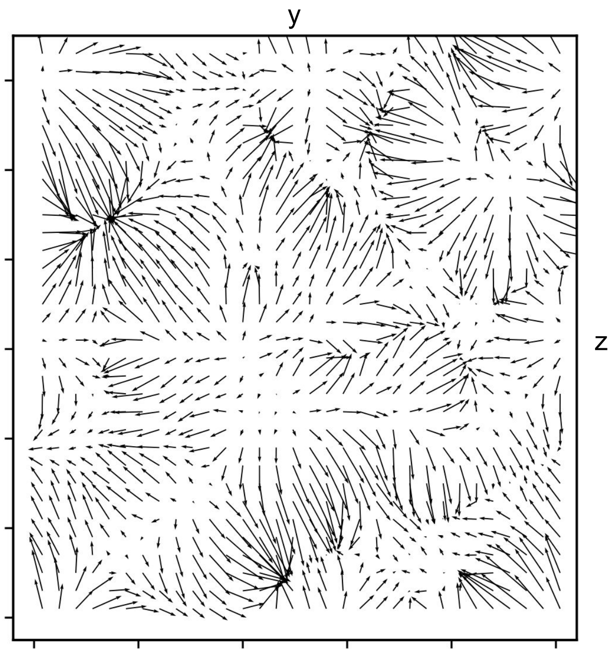


[1808.03344] Deep Learning for Single Image Super-Resolution: A Brief Review
[1902.06068] Deep Learning for Image Super-resolution: A Survey

How to SR an N-body simulation

- Format (N-body) simulations as 3D images
  - Initial conditions typically on *regular grids*
  - In *Lagrangian* description, particle *displacements* are images of 3 channels
    - allows interpreting *results as simulations* (thus the title)
    - *conserves mass* by construction
    - has *non-local* information
  - Likewise for their *velocities*
  - 6-channel images determine the whole phase space

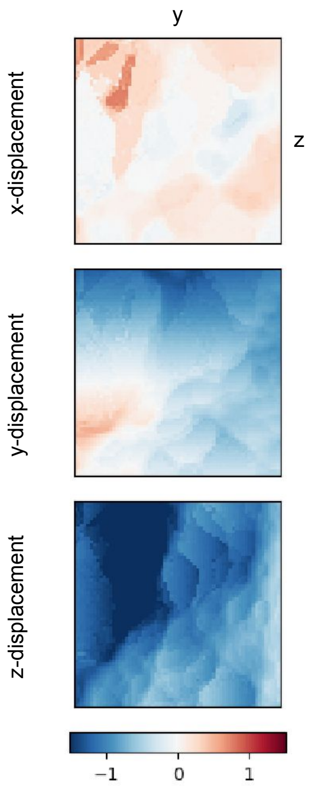
a yz-slice



How to SR an N-body simulation

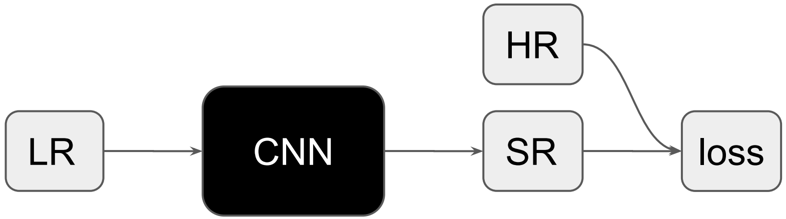
- Format (N-body) simulations as 3D images
  - Initial conditions typically on *regular grids*
  - In *Lagrangian* description, particle *displacements* are images of 3 channels
    - allows interpreting *results as simulations* (thus the title)
    - *conserves mass* by construction
    - has *non-local* information
  - Likewise for their *velocities*
  - 6-channel images determine the whole phase space

a yz-slice

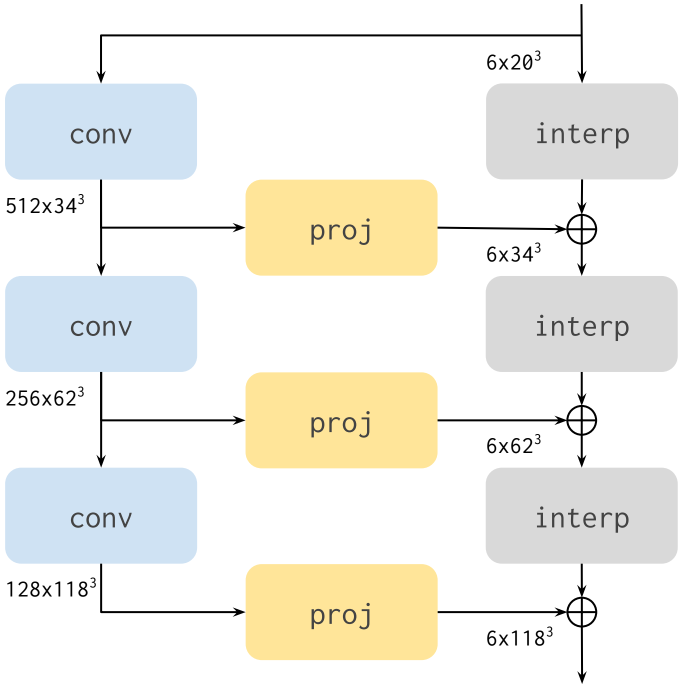


supervised learning?

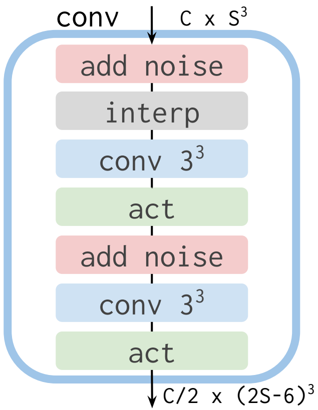
- Simple loss functions aim to minimize the pixel-wise (rather than statistical) difference between SR and HR

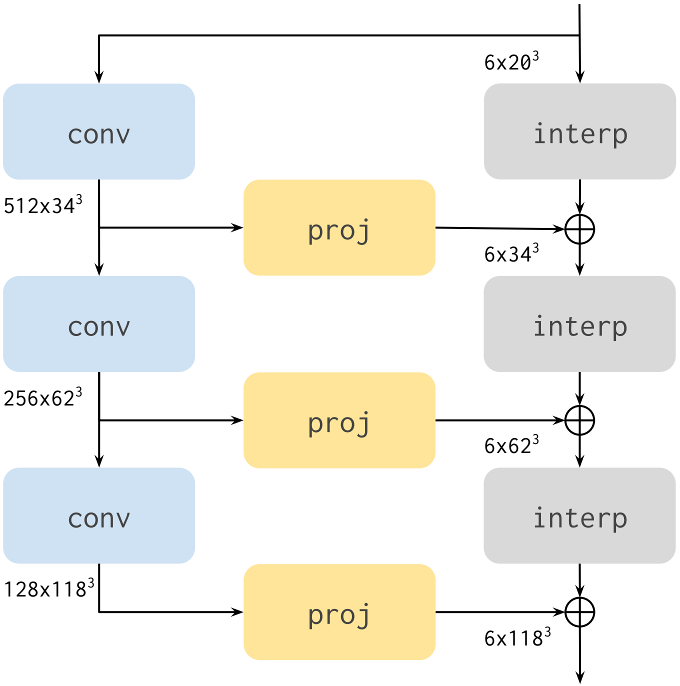


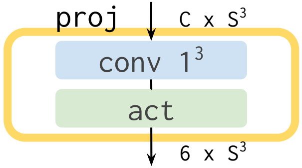
Generator architecture — based on StyleGAN2



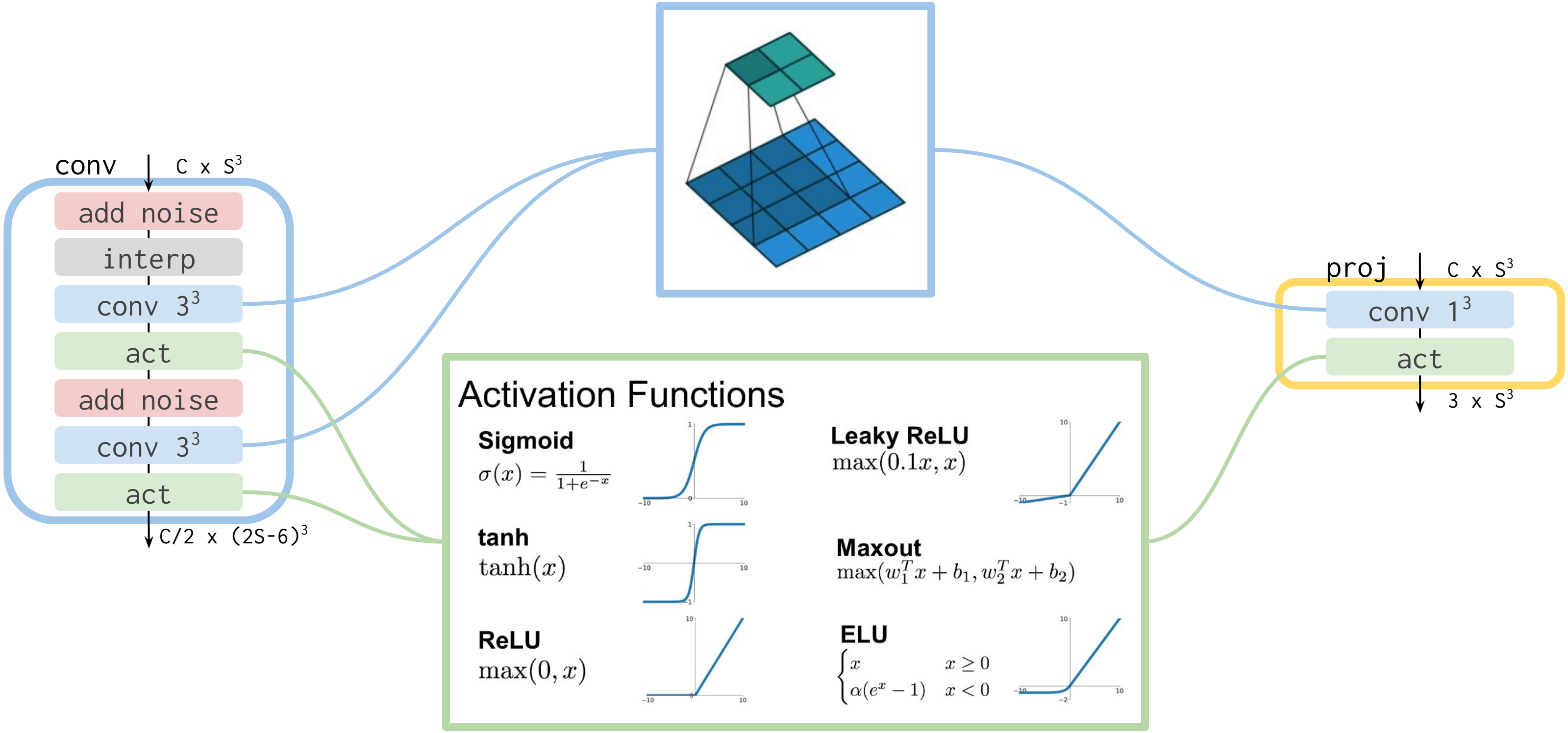
Generator architecture — based on StyleGAN2



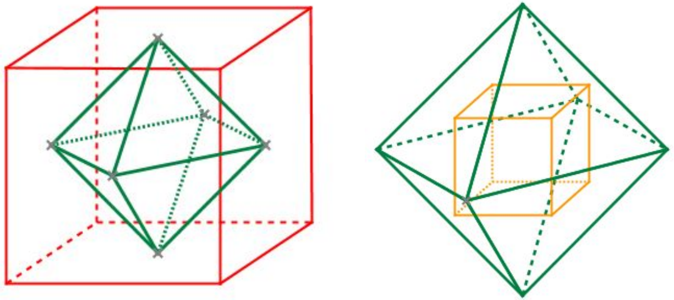




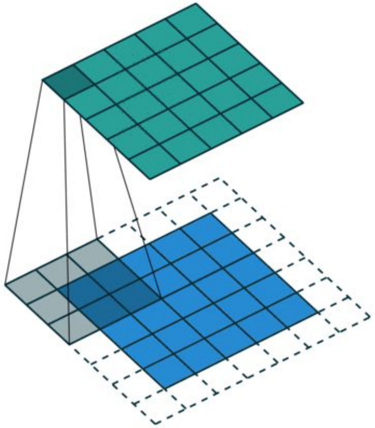
Generator architecture — based on StyleGAN2



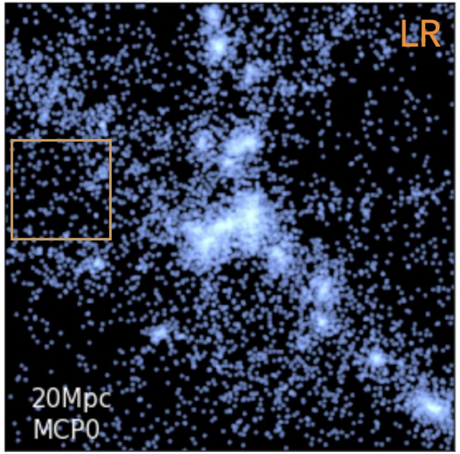
Symmetries



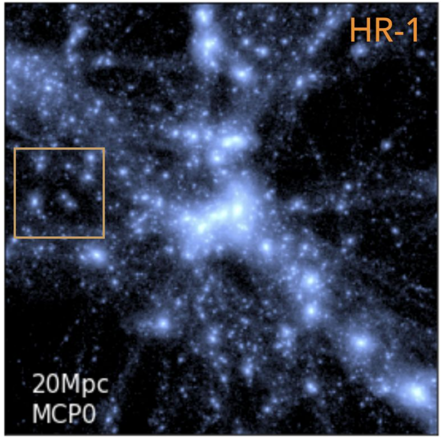
- Rotational symmetry (O_h point group)
  - that of a cube / octahedron
  - Discrete because of periodic geometry, 48 operations
  - Apply *data augmentation*, 48x as many data, and better symmetry in predictions
- Translational symmetry
  - A feature of Convolutional Neural Network (CNN) by construction
  - Padding in the convolution layers can break this
  - Periodic padding (on the LR input)

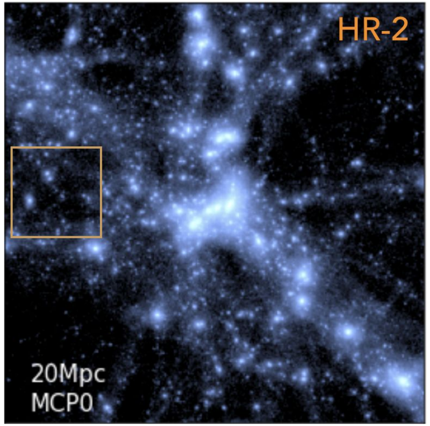


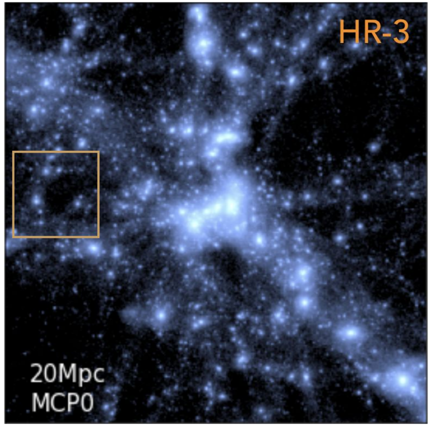
One-to-many mapping



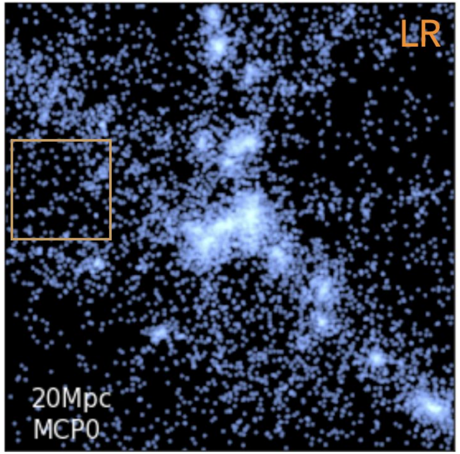
LR —> HR



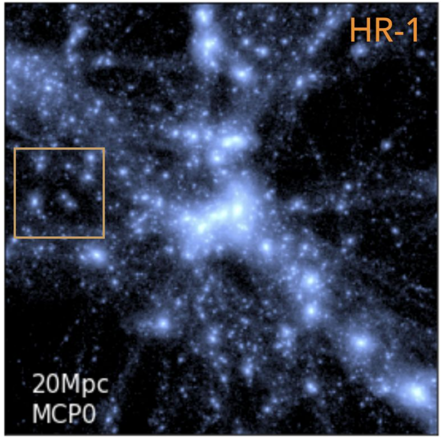


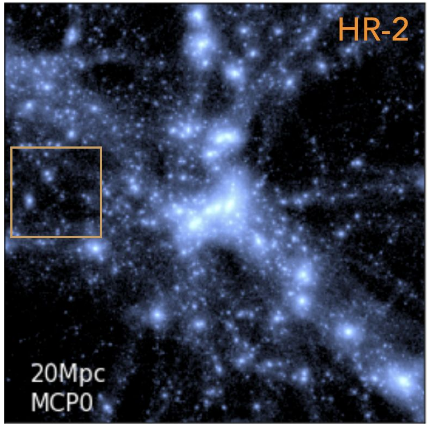


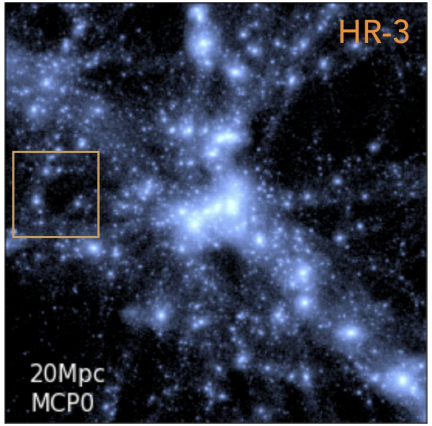
Stochasticity

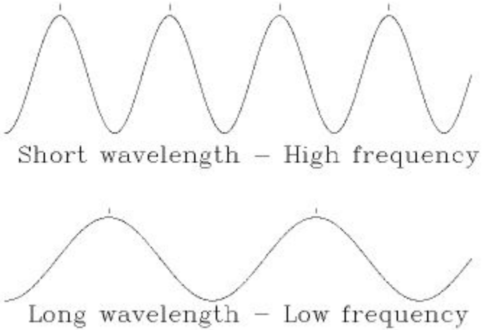


LR —> HR









only in
HR ICs

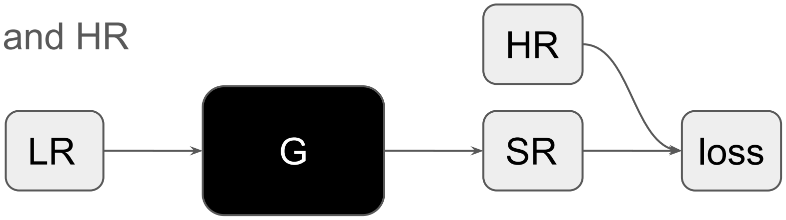
LR ICs

Stochasticity

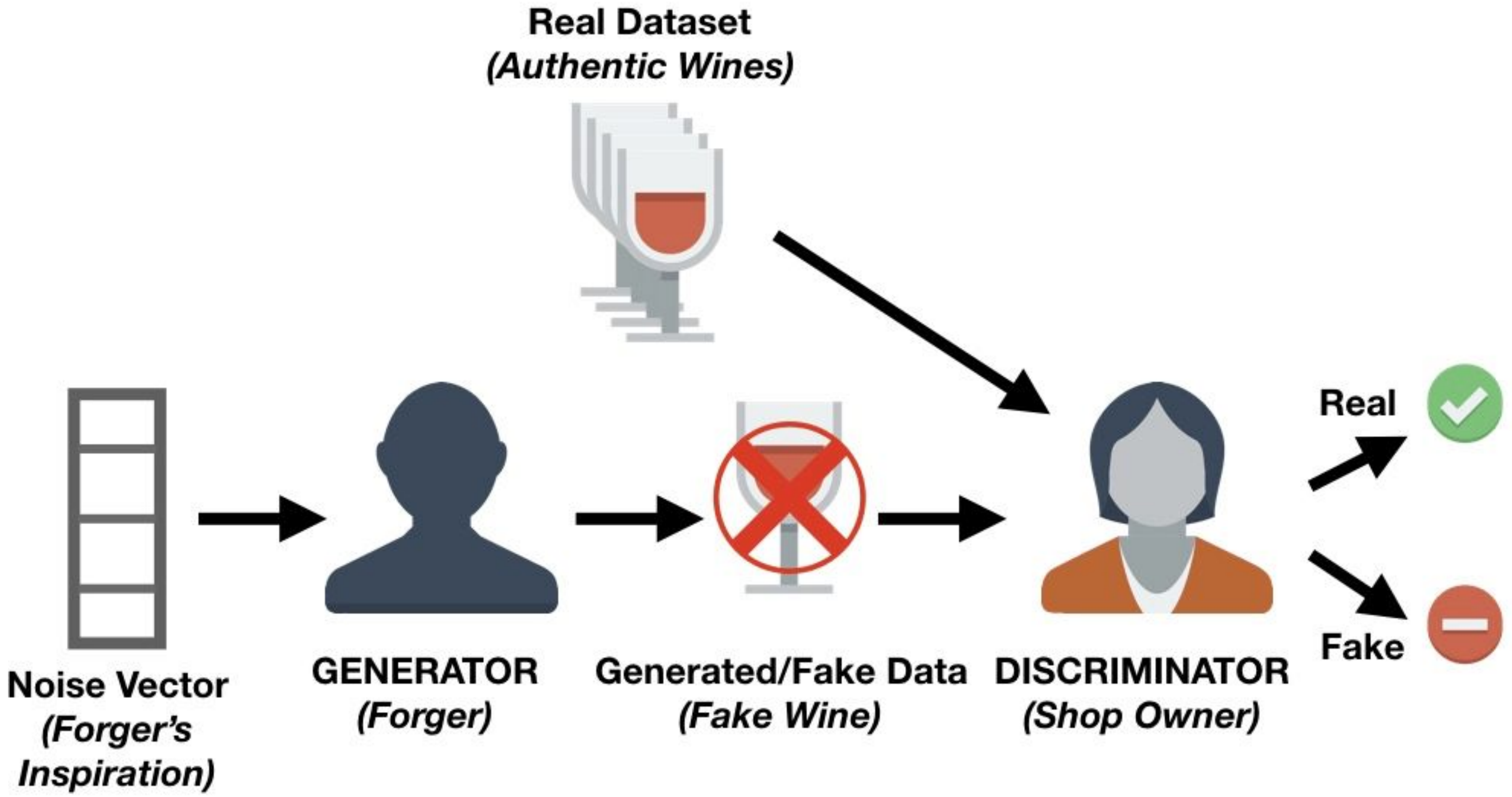
- Stochasticity from short-wavelength modes
  - LR lacks short modes present in HR
  - Initially these modes are statistically independent
  - Later hierarchical structure formation: short SR modes conditioning on the long LR modes
- Stochasticity *consequence 1*: need for randomness
  - Add noises throughout our (generator) neural network
  - An SR realization can be different from the HR one on small scales
    - SR realizations are different among themselves

supervised learning?

- Stochasticity consequence 2: need for better loss function
  - Simple loss functions aim to minimize the pixel-wise (rather than statistical) difference between SR and HR

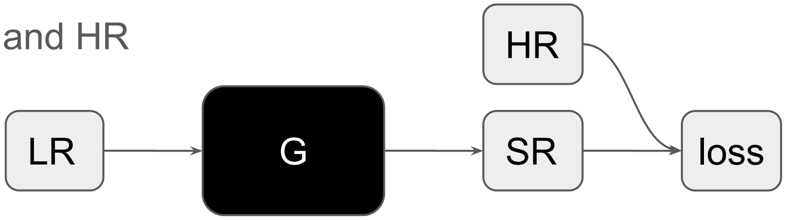


Generative Adversarial Networks (GAN)

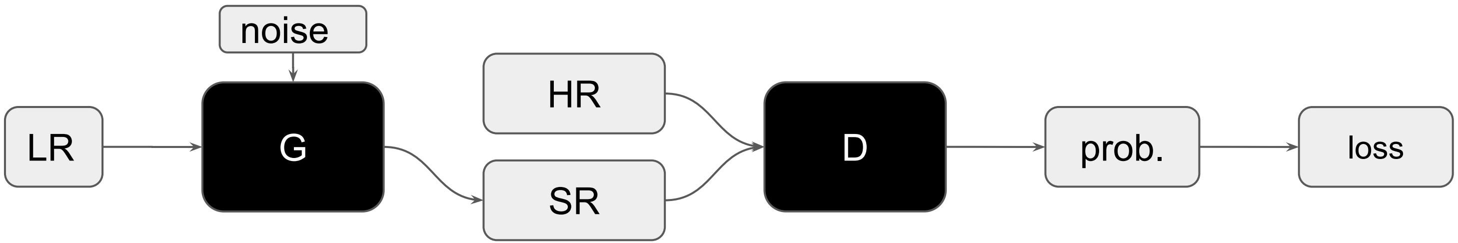


GAN

- Stochasticity consequence 2: need for better loss function
  - Simple loss functions aim to minimize the pixel-wise (rather than statistical) difference between SR and HR

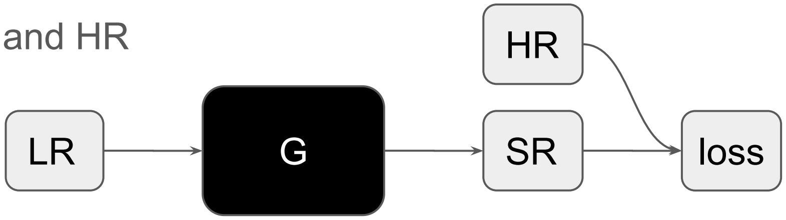


  - Use generative adversarial network (GAN) that adds another (discriminator) network
  - Train generator (G) and discriminator (D) alternately in a *game*
    - Update G to fool D, and update D to distinguish SR from HR

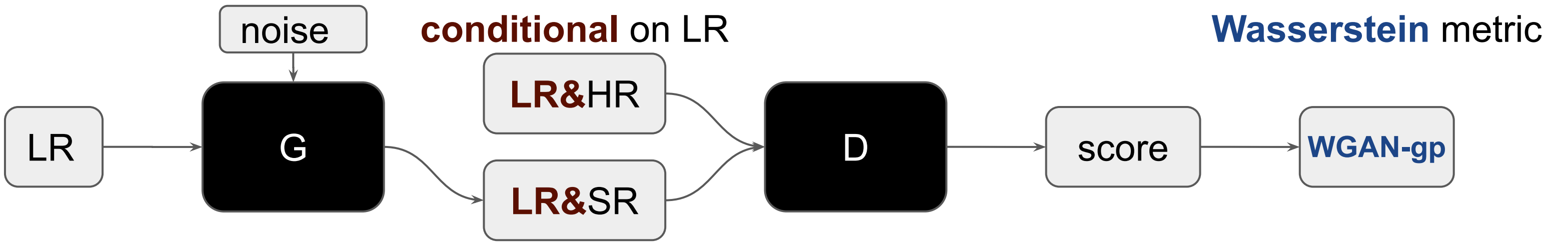


cGAN & WGAN

- Stochasticity consequence 2: need for better loss function
  - Simple loss functions aim to minimize the pixel-wise (rather than statistical) difference between SR and HR

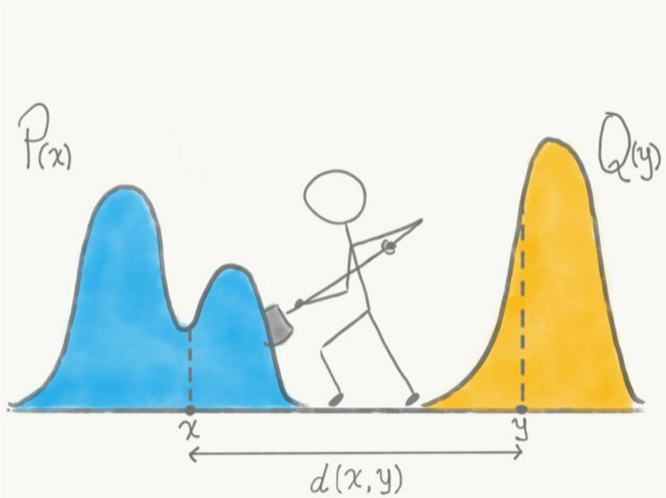


  - Use generative adversarial network (GAN) that adds another (discriminator) network
  - Train generator (G) and discriminator (D) alternately in a *game*
    - Update G to fool D, and update D to distinguish SR from HR

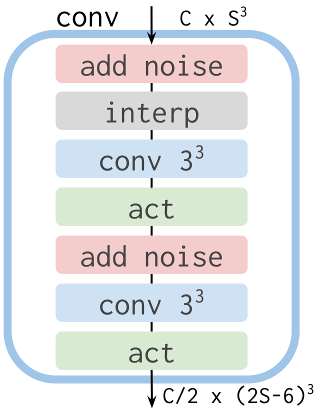


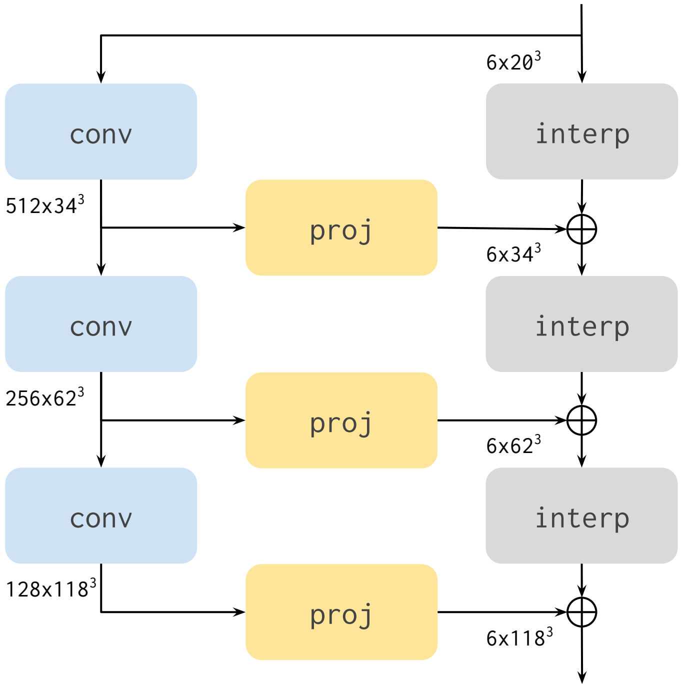
Loss function — Wasserstein distance

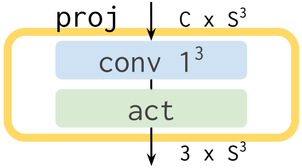
- In the original GAN, D output a probability, and loss function is binary cross entropy (log likelihood)
- Wasserstein GAN (WGAN) minimize Wasserstein distance
    - aka earth mover's distance, measure the difference between two probability distribution by *optimal transport*
        - Minimum work required to move one distribution to another
        - two prob. being that of the real and fake images, in high dimensional space
    - Mathematical *duality* leads to computable expression, however under *1-Lipschitz constraint*
    - Instead of probability, D gives scores to SR and HR
    - WGAN-gp adds *gradient penalty* to the loss for the Lipschitz constraint



Generator architecture — based on StyleGAN2

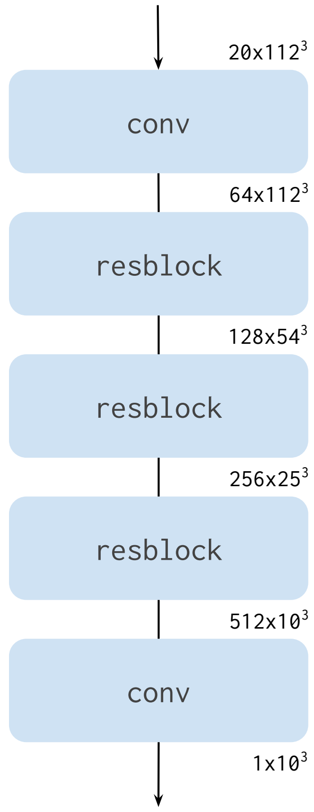




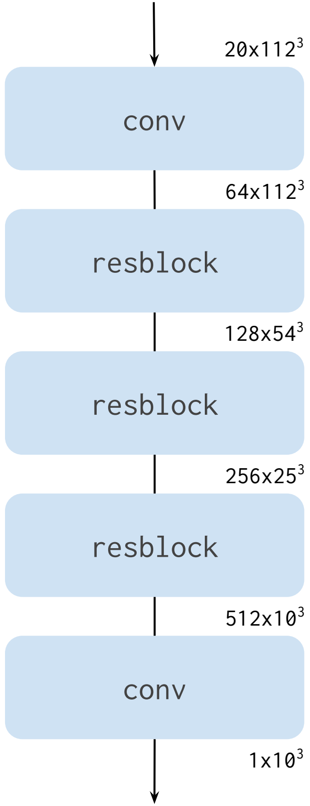


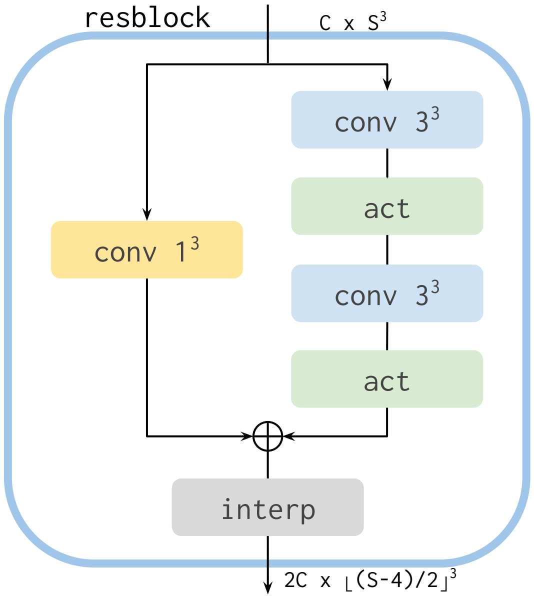
Discriminator architecture — based on StyleGAN2

ResNet



Discriminator architecture — based on StyleGAN2





Code

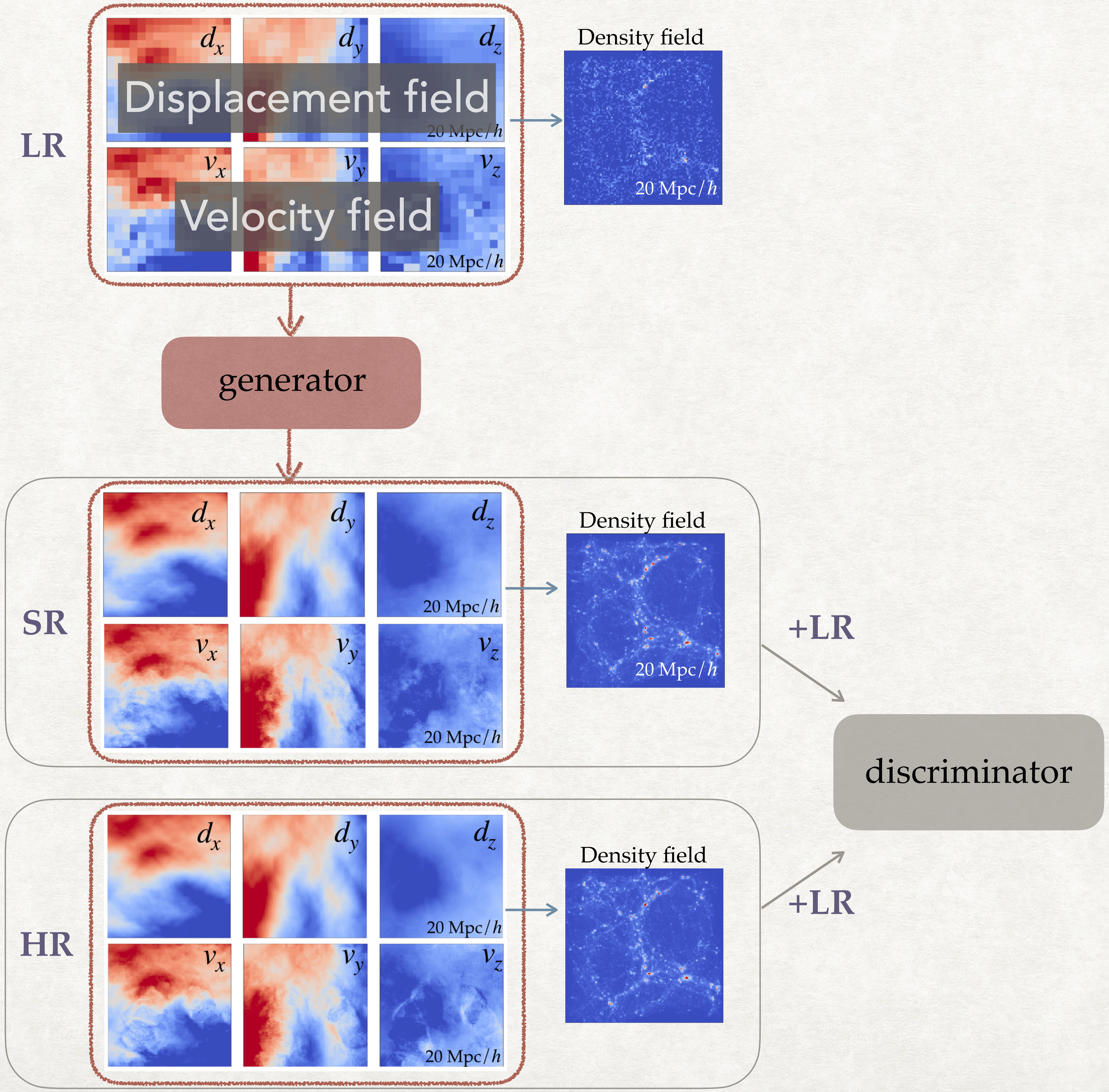
- map2map (https://github.com/eelregit/map2map)
  - supervised and GAN, and more generally field-level emulation
  - automatic data handling
    - loading/sampling/caching: help with I/O to not starve GPUs
    - cropping & padding: translational symmetry
    - augmentation: rotational & translational symmetry
  - PyTorch-based, training tracked with tensorboard

- density field super resolution: Ramanah et al. 2020
  Tensorboard-based: https://github.com/doogesh/super_resolution_emulator

Super-resolution simulations

Trainings and results

Training process



Training Sets:

16 pairs of LR-HR simulations

BoxSize = 100 Mpc/h

LR : Np = 64^3, $M_{\rm DM} = 3 \times 10^{11} M_\odot$

HR : Np = 512^3, $M_{\rm DM} = 5.8 \times 10^{8} M_\odot$

Test Sets:

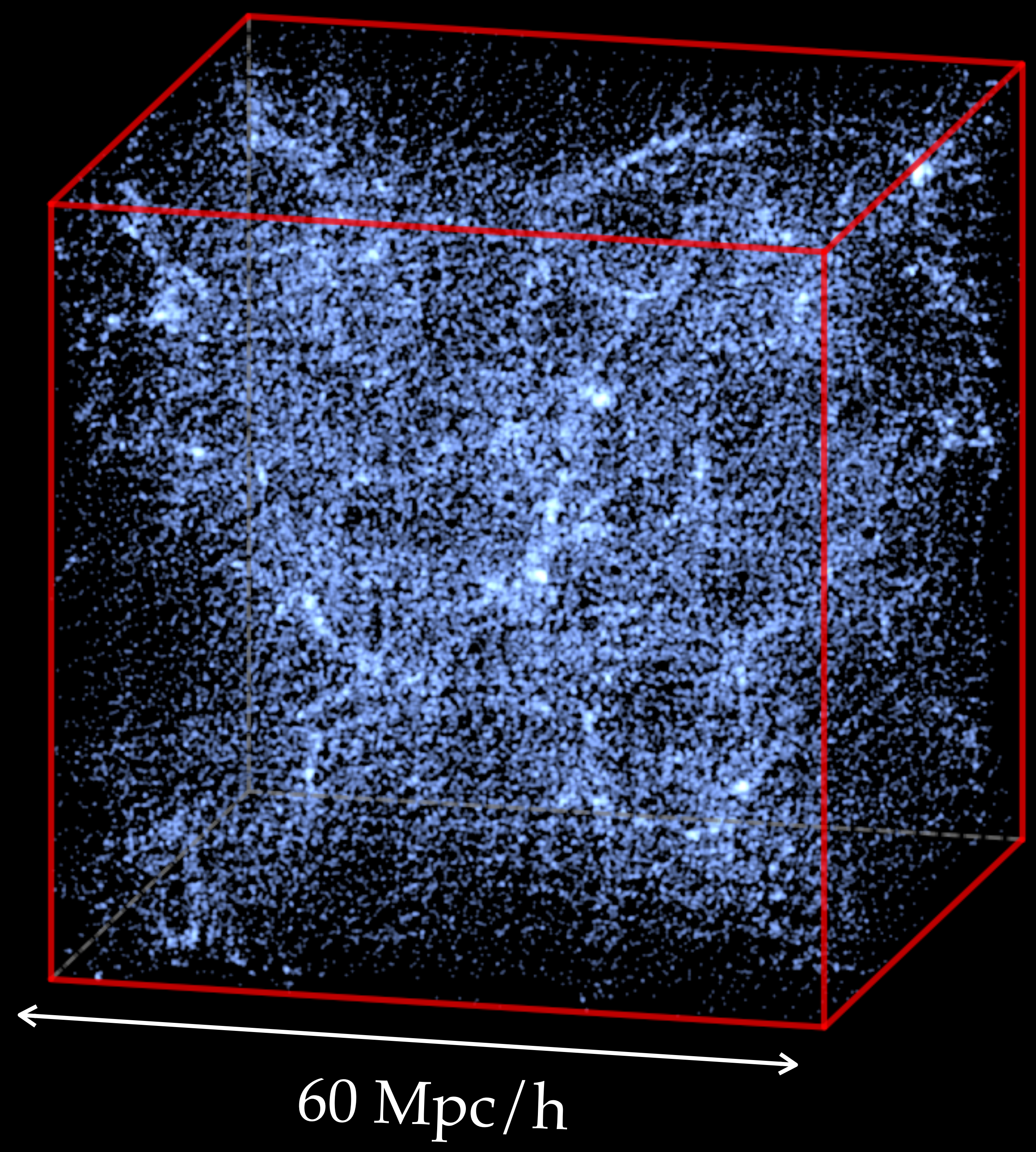
10 pairs of LR-HR simulations

BoxSize = 100 Mpc/h

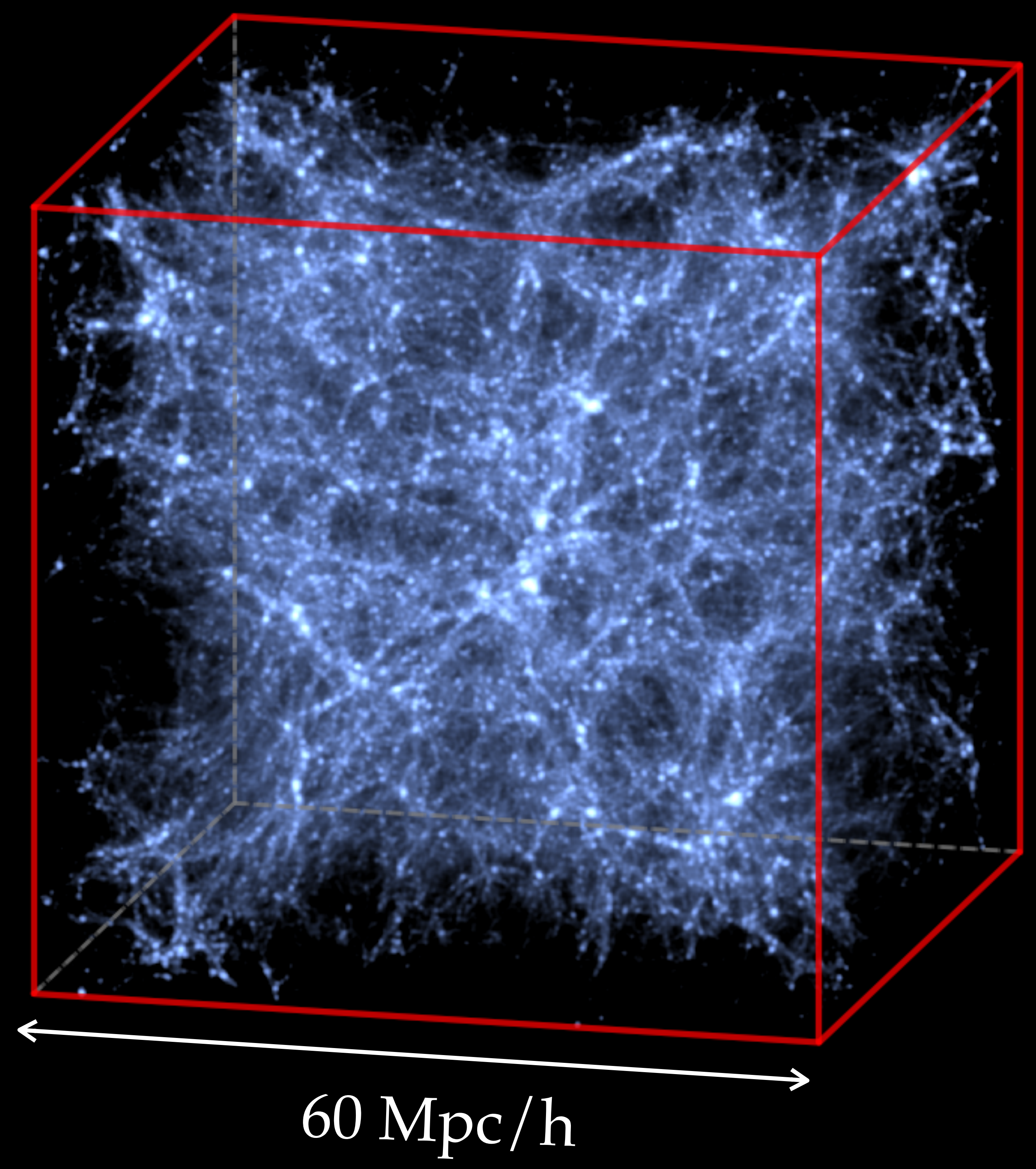
Same cosmology and resolution as the training sets

Density field visualization at z=2

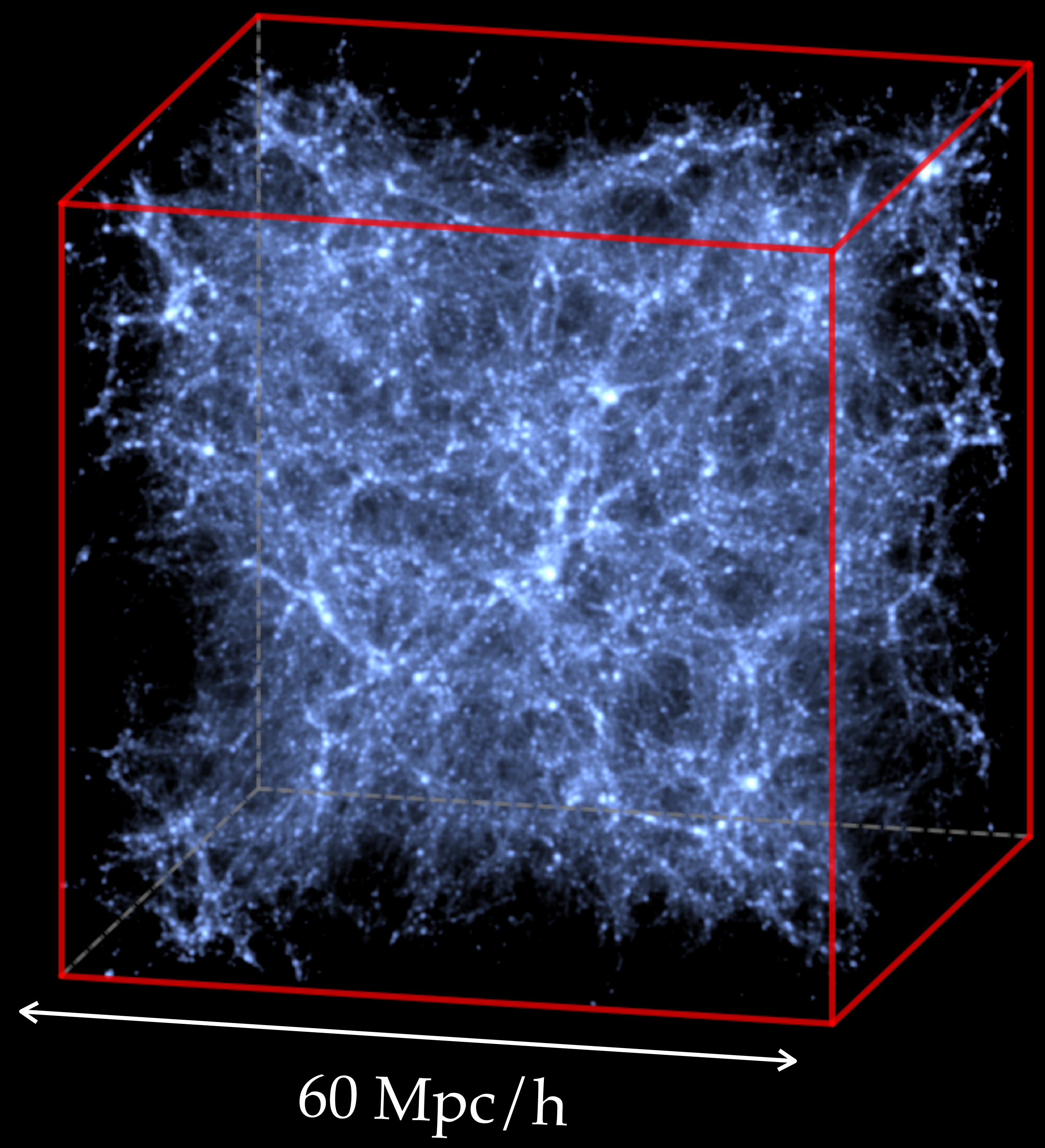
Low Resolution



High Resolution

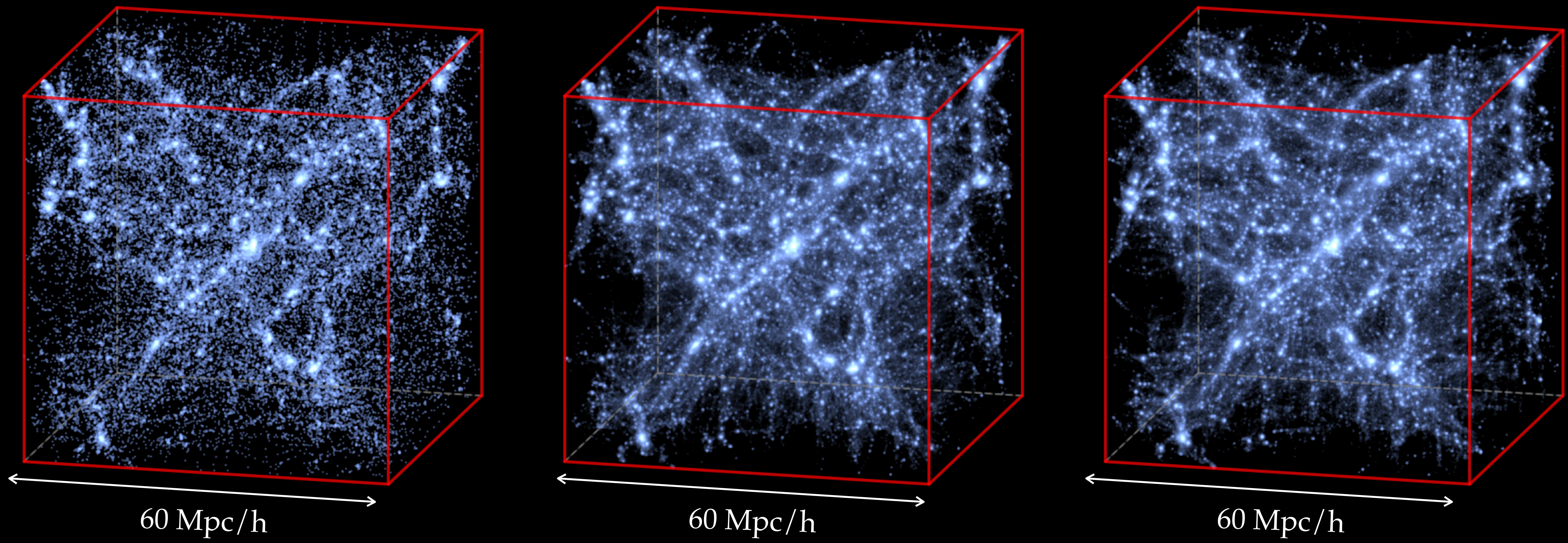


Super resolution



Density field visualization at z=0

Low Resolution | **High Resolution** | **Super resolution**

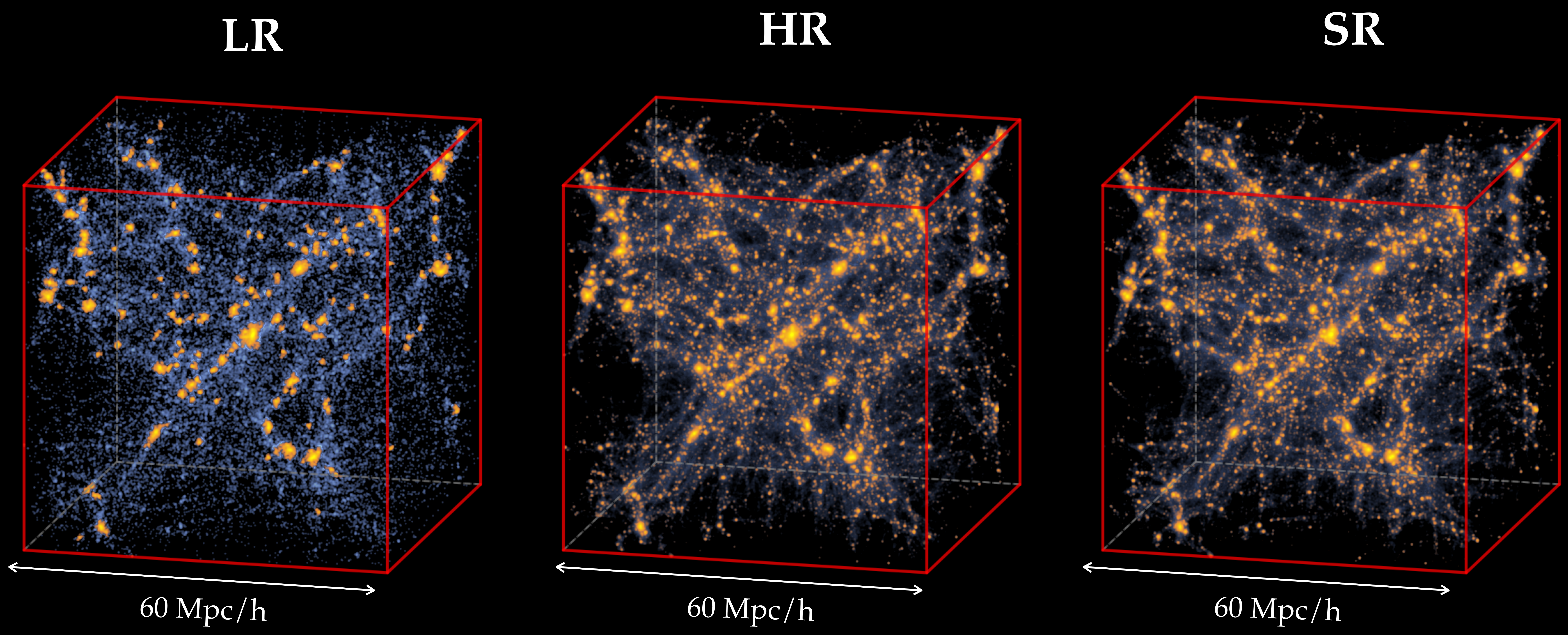


< 1 core hour

~ 2000 core hour

~ 10 seconds!

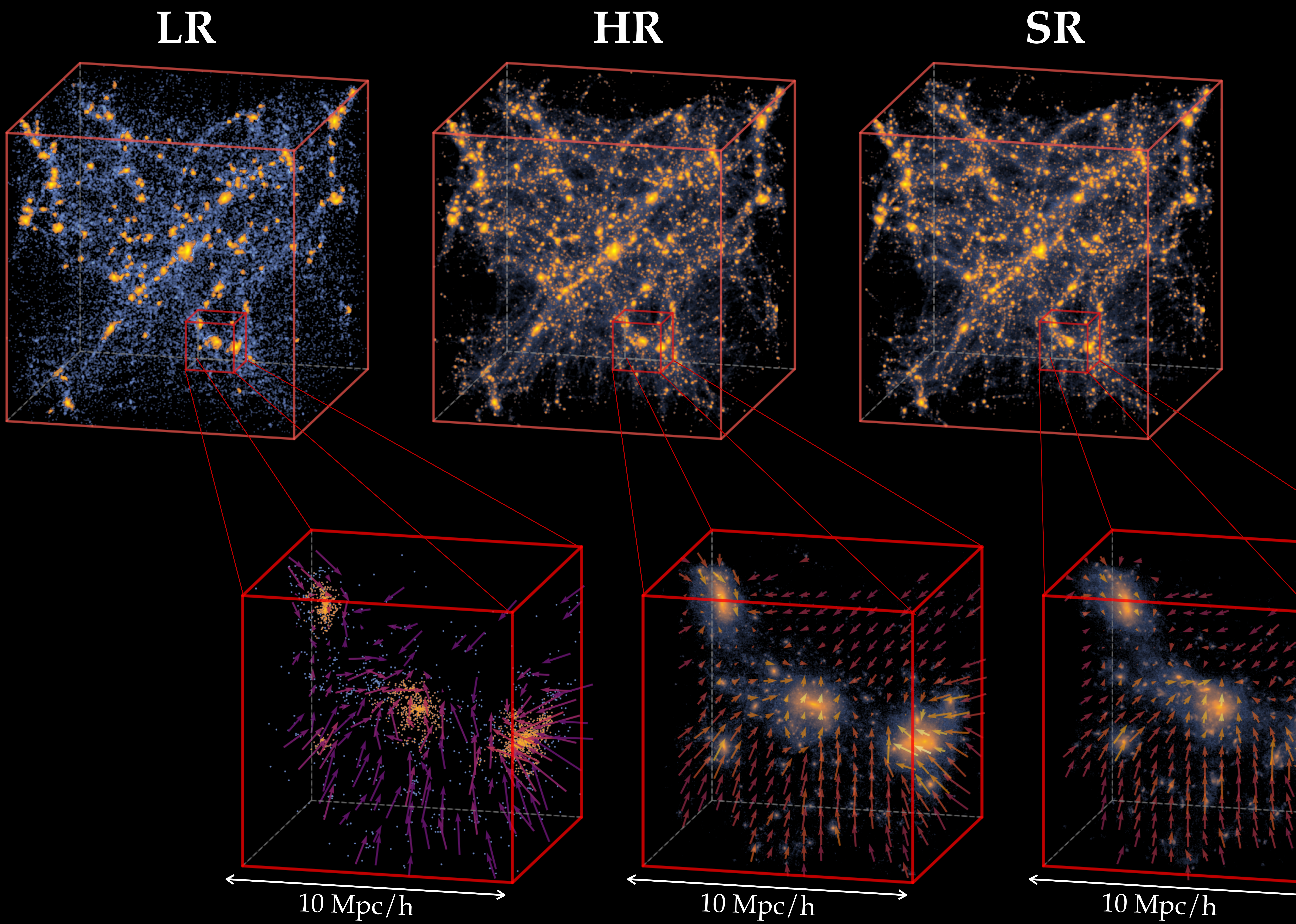
Density field visualization at z=0



< 1 core hour

~ 2000 core hour

~ 10 seconds!



Predicts the full
6D phase space
output

Validation Metrics

Full field statistics :

- Matter power spectrum (two point statistics)
- Bispectrum (three point statistics)
- Redshift space 2D power spectrum (velocity)

Test Sets:
10 pairs of LR-HR simulations
BoxSize = 100 Mpc/h
Same cosmology and resolution as the training sets

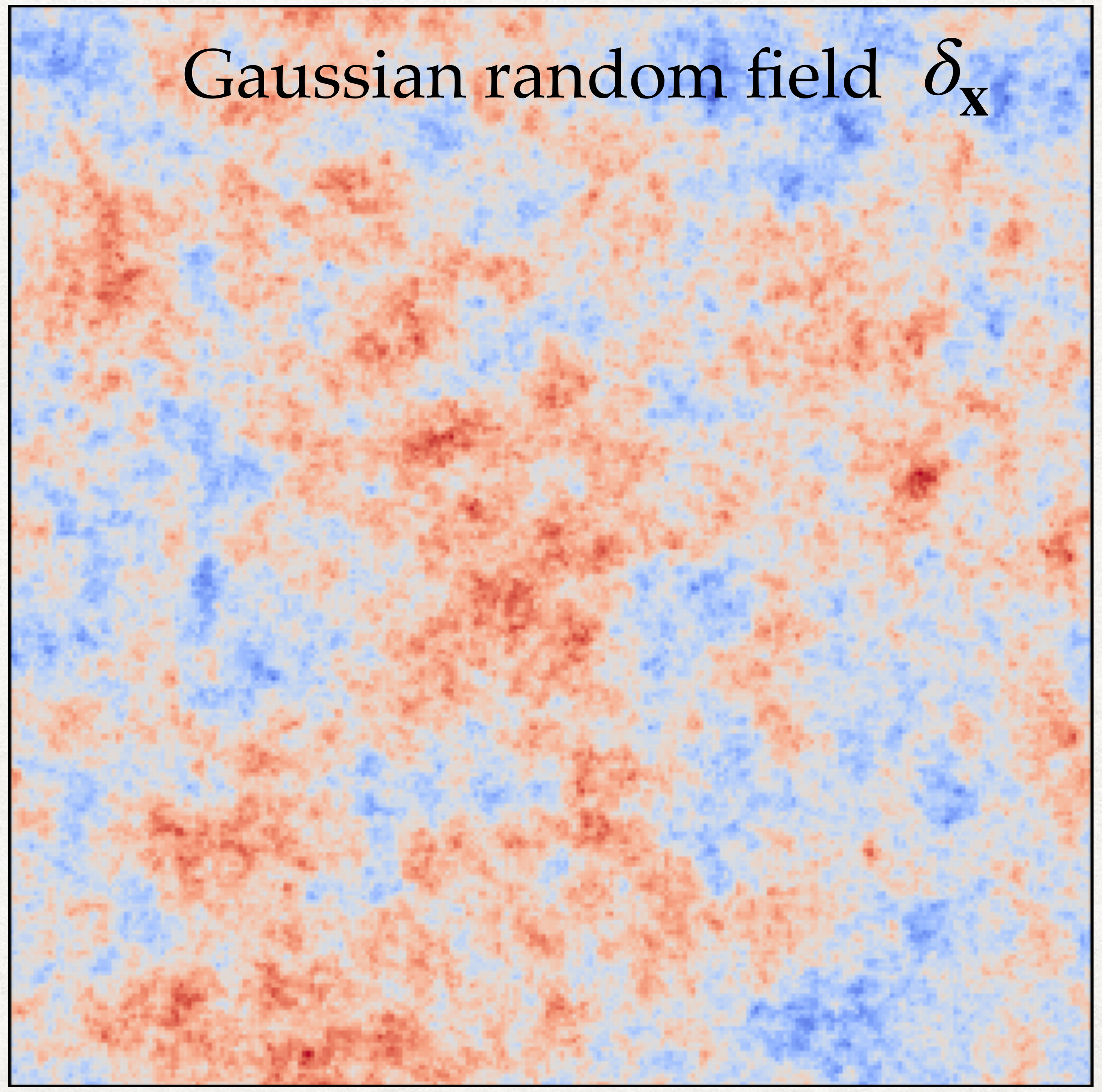
Halo catalog statistics:

- Abundance of halos and subhalos
- Mean occupation distribution of subhalos
- 2-point correlation function
- Redshift-space correlation
- Pairwise velocity distribution

Full field statistics: Matter power spectrum

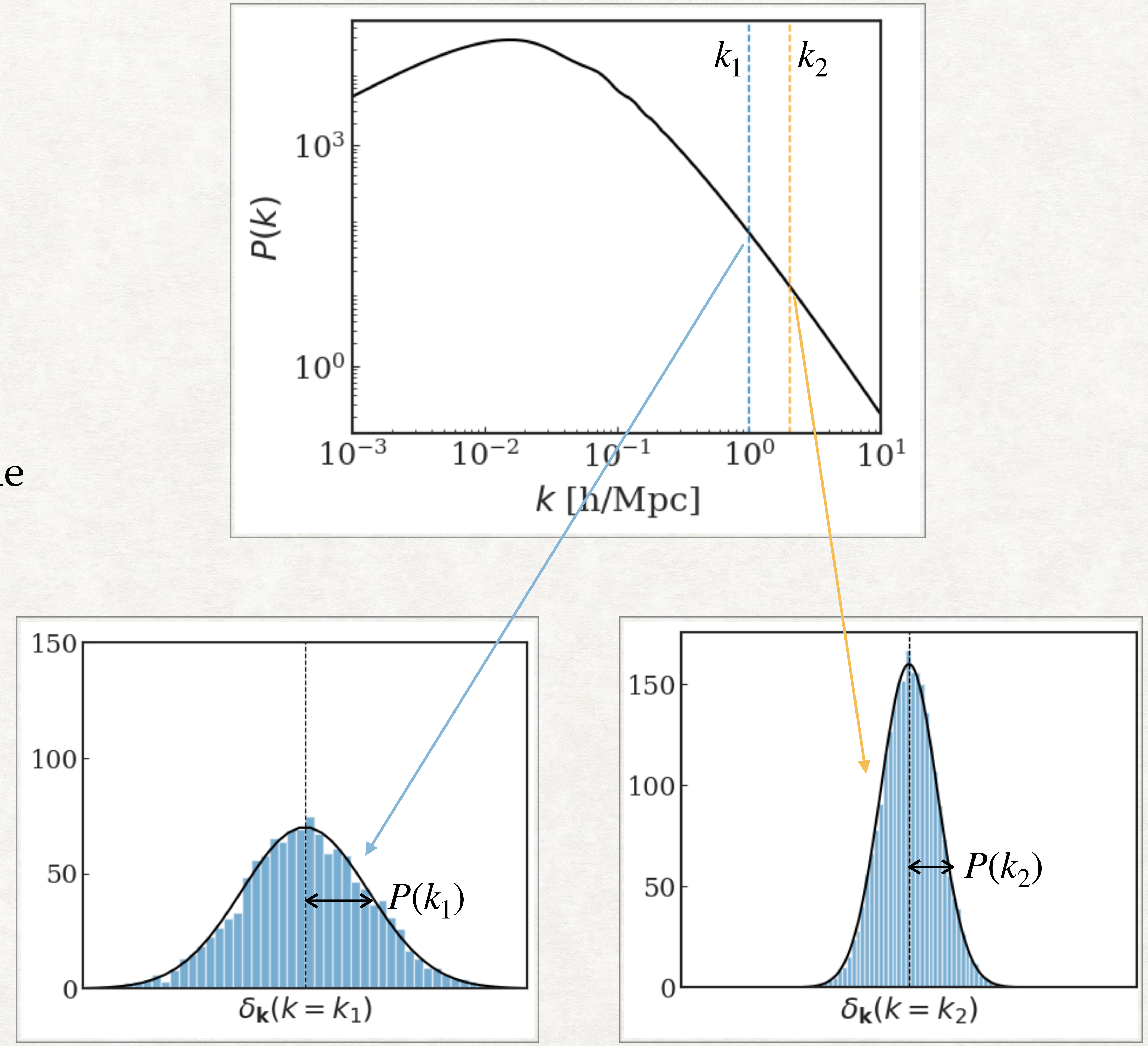
$\delta(\mathbf{r}) = \rho(\mathbf{r})/\bar{\rho} - 1$: spatial overdensity

$$\xi(|\mathbf{r}|) = \langle \delta(\mathbf{r}')\, \delta(\mathbf{r}' + \mathbf{r}) \rangle \xrightarrow{\text{FT}} P(|\mathbf{k}|) = \int \xi(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{r}}\, d^3\mathbf{r}$$



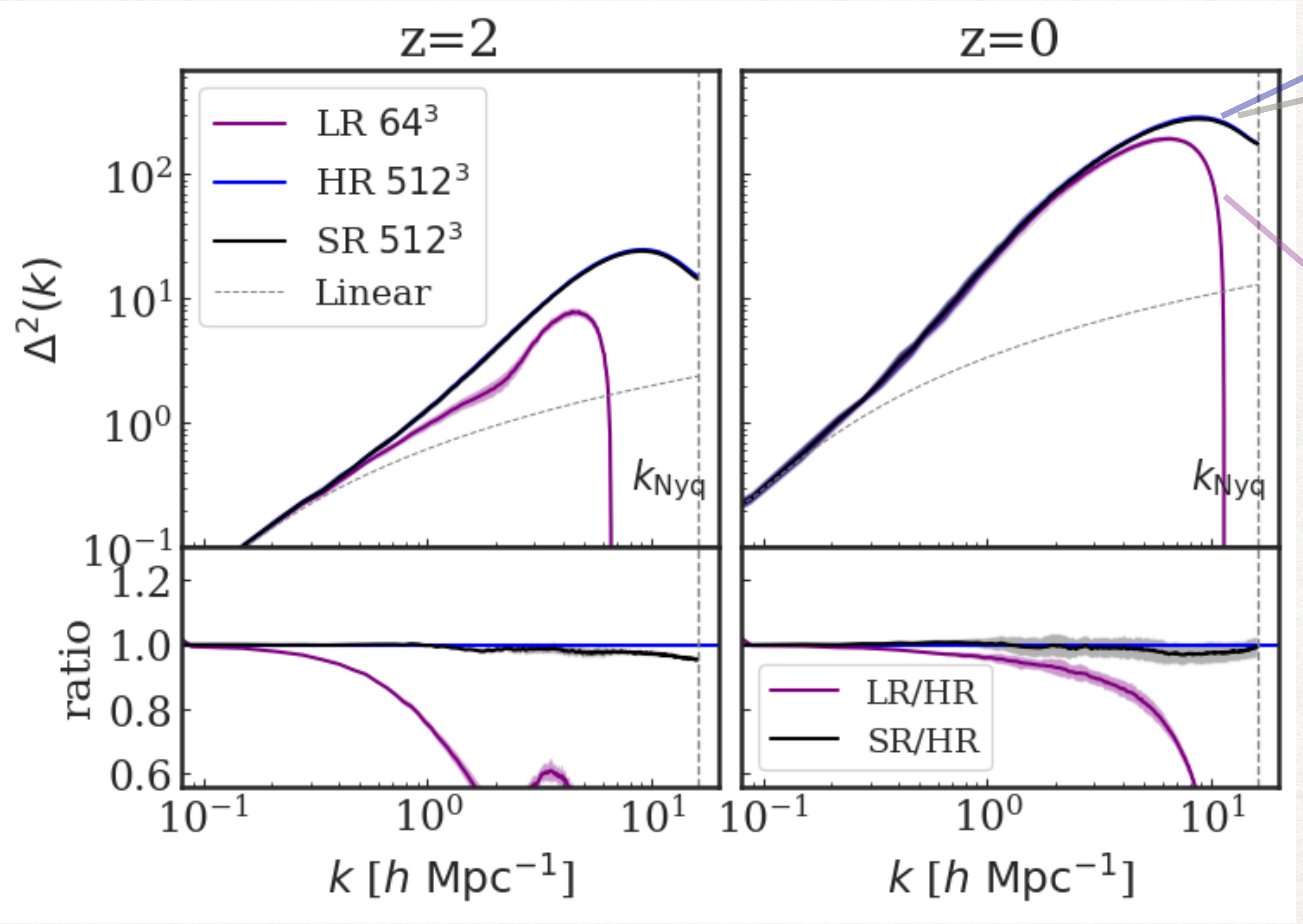
$k = 2\pi/\lambda$
small k —> large scale

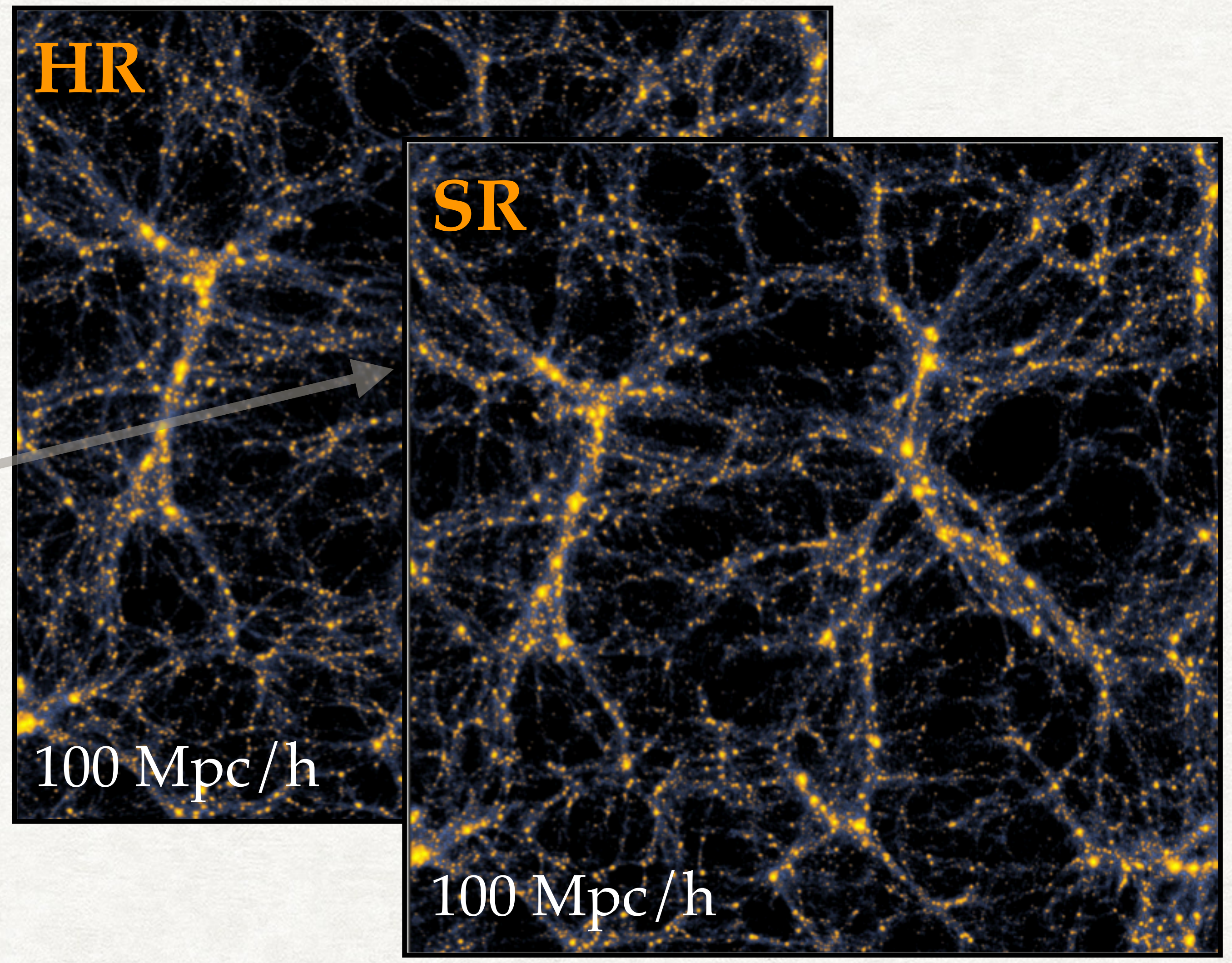
$\xrightarrow{\text{Fourier space}}$

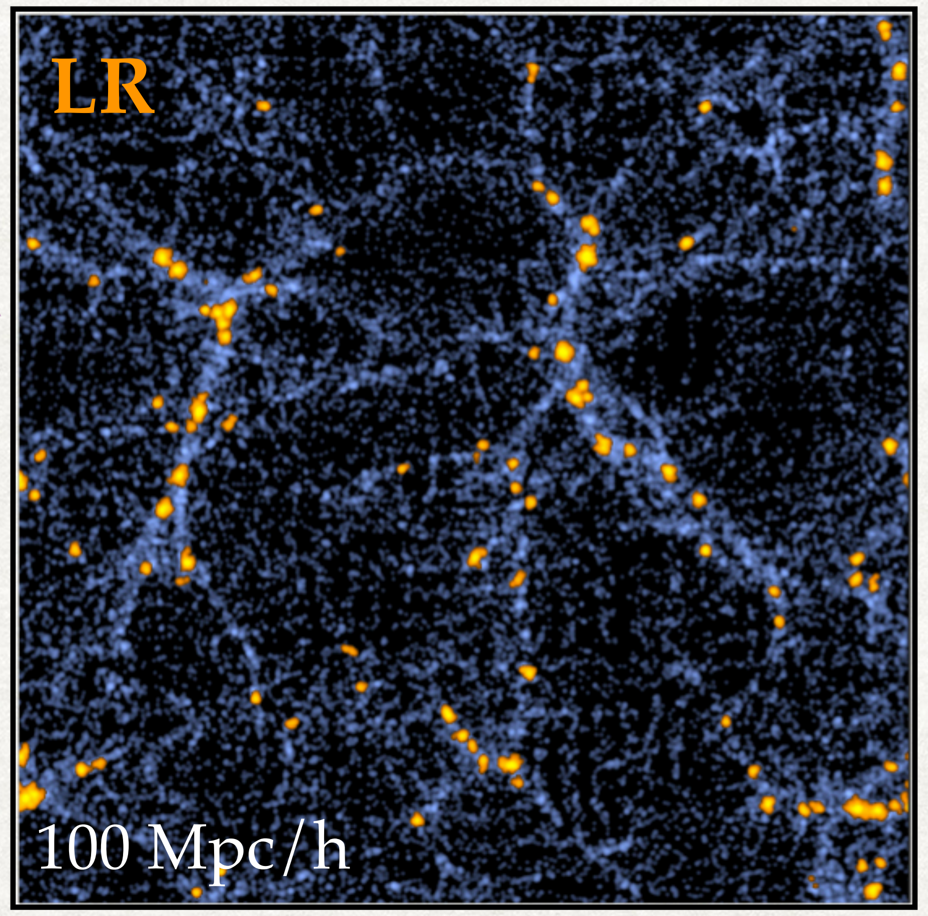


Full field statistics: Matter power spectrum

Dimensionless power $\Delta^2(k) \equiv k^3 P(k)/2\pi^2$

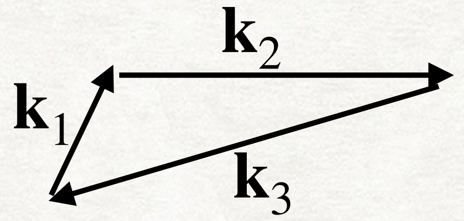






Full field statistics: Bispectra

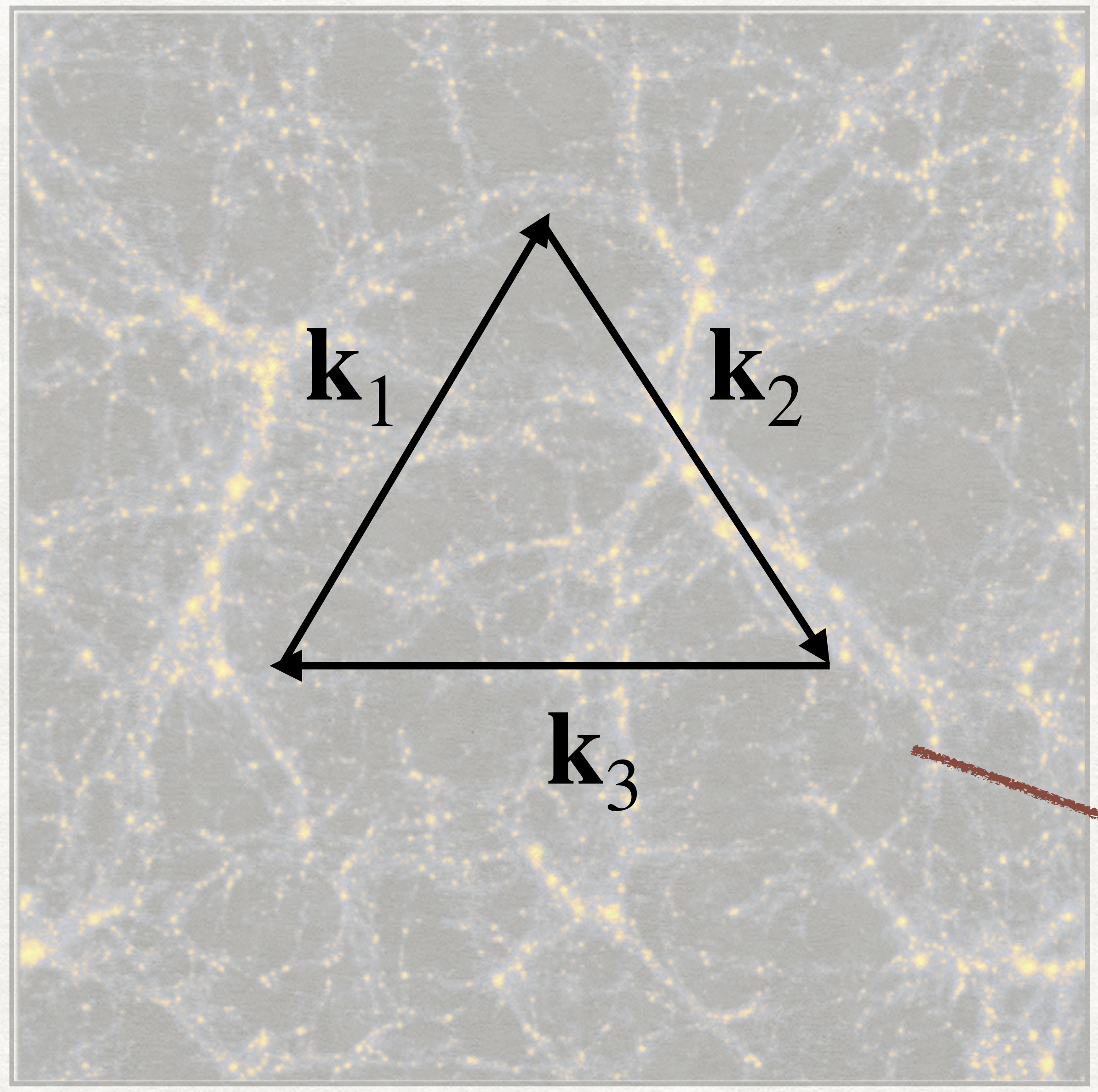
Primary diagnostic for **non-Gaussianity**
Defined for closed triangles (statistical homogeneity and isotropy)

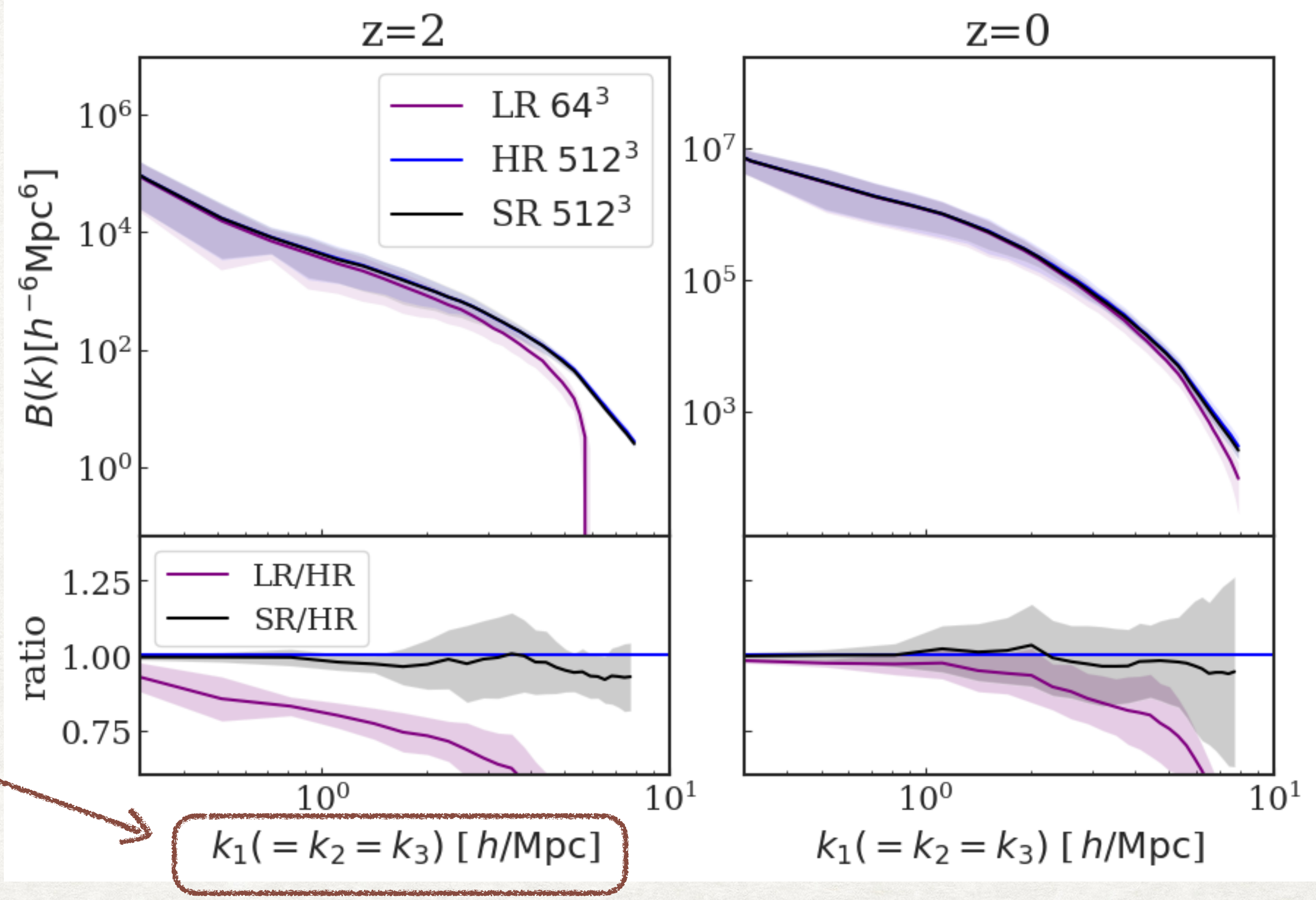


$$(2\pi)^3 B\left(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3\right) \delta_{\mathrm{D}}\left(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3\right)=\left\langle\delta\left(\boldsymbol{k}_1\right) \delta\left(\boldsymbol{k}_2\right) \delta\left(\boldsymbol{k}_3\right)\right\rangle$$

Equilateral triangles

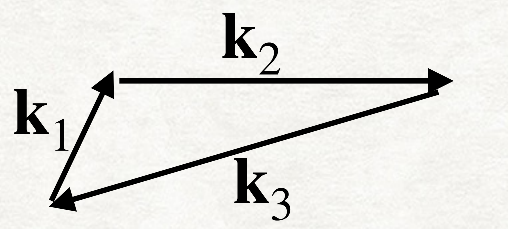
$$k_1 = k_2 = k_3$$





Full field statistics: Bispectra

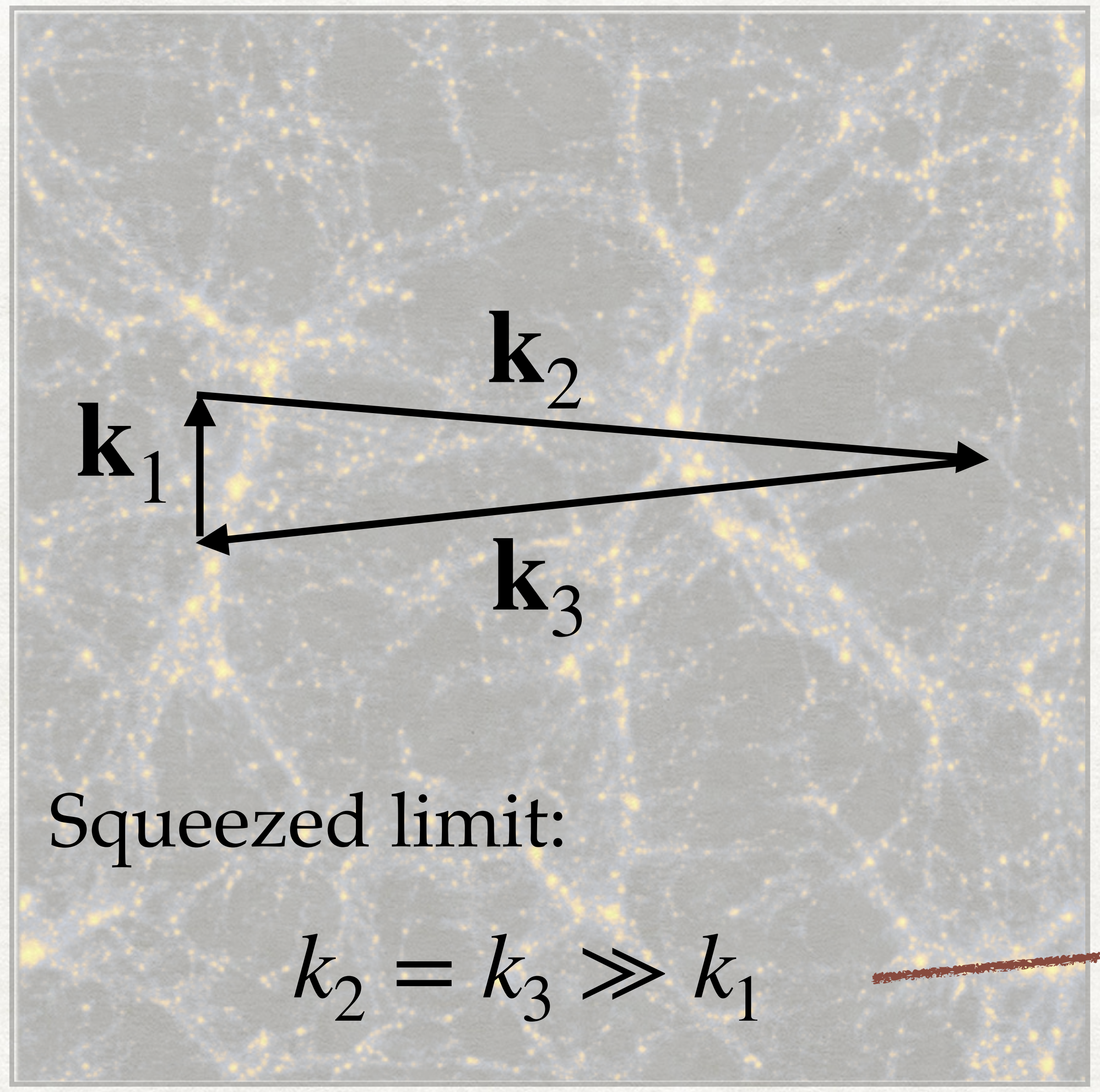
Primary diagnostic for **non-Gaussianity**
Defined for closed triangles (statistical homogeneity and isotropy)

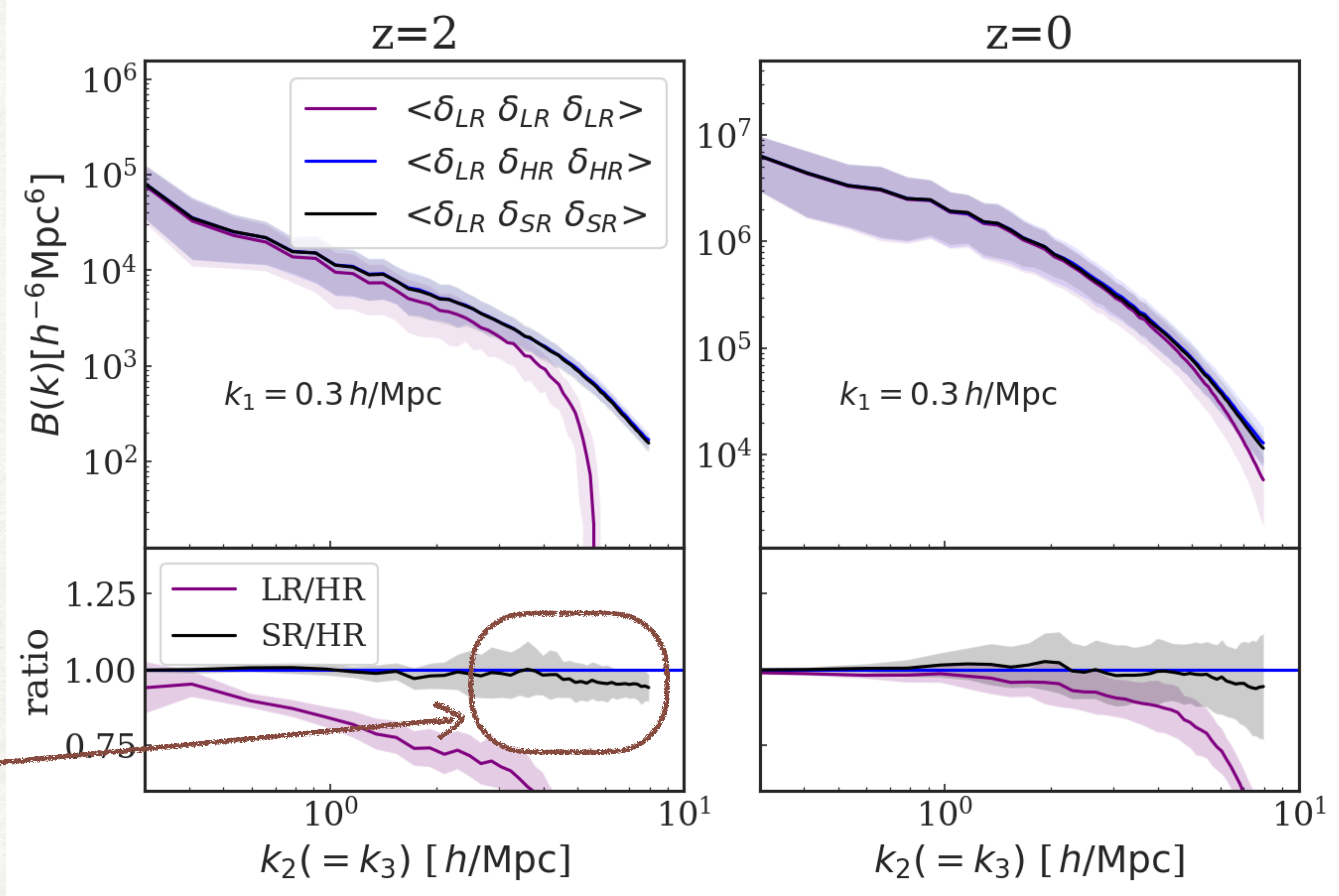


$$(2\pi)^3 B\left(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3\right) \delta_{\mathrm{D}}\left(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3\right)=\left\langle\delta\left(\boldsymbol{k}_1\right) \delta\left(\boldsymbol{k}_2\right) \delta\left(\boldsymbol{k}_3\right)\right\rangle$$

Isosceles triangles

$k_2 = k_3$





Full field statistics: Redshift-space distortion

The peculiar velocity makes the redshift-space clustering anisotropic

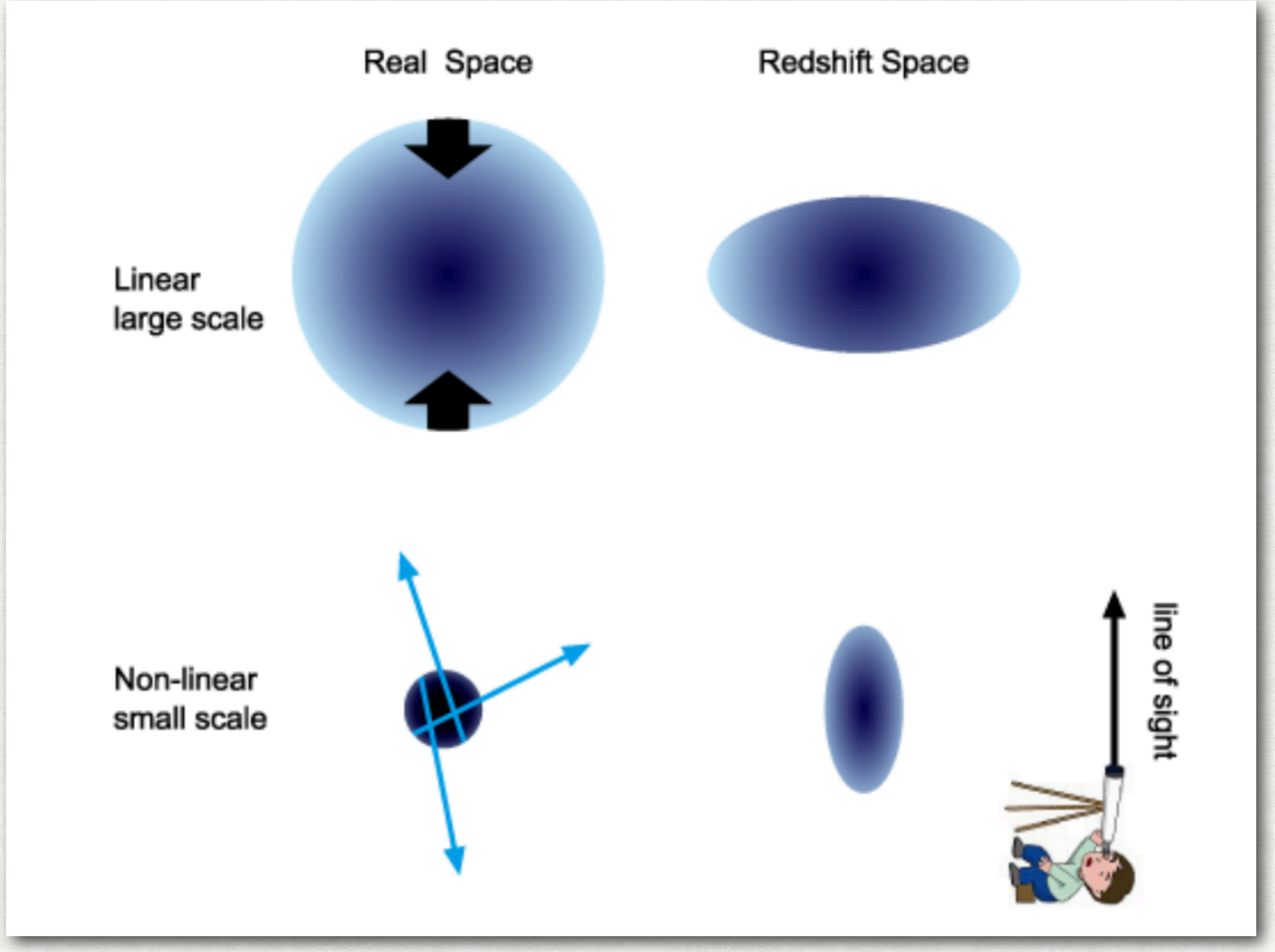


Image from: Shun Saito
RSD lecture note

real-space coordinate

peculiar velocity along the line of sight

$$\boxed{\mathbf{s} = \mathbf{x} + \frac{v_z}{aH(a)}\hat{z}}$$

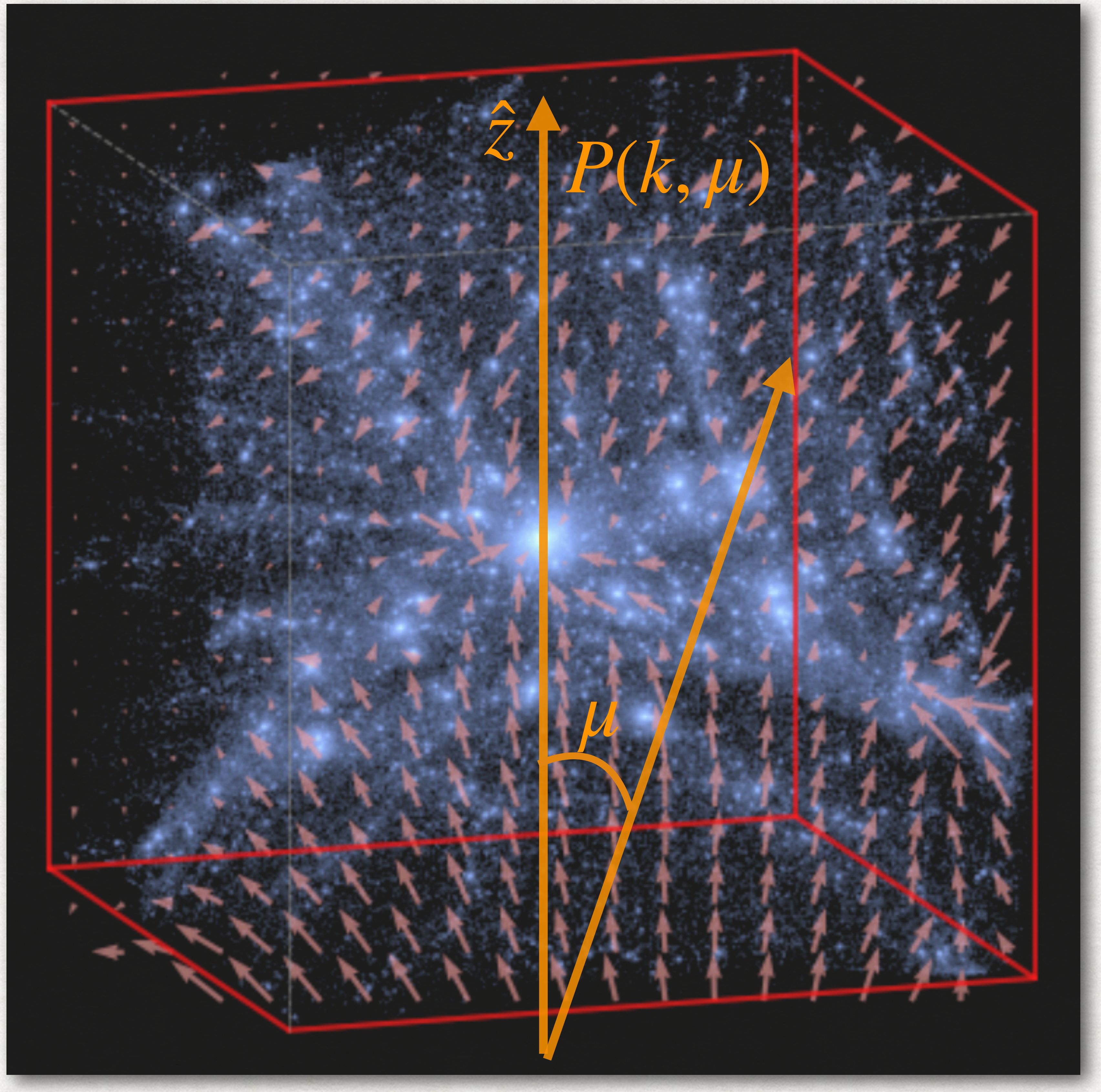
redshift-space coordinate

The line of sight direction

a: scale factor
$H(a)$: Hubble expansion rate

Full field statistics: Redshift-space distortion

The peculiar velocity makes the redshift-space clustering anisotropic —> 2D power spectrum $P(k, \mu)$



real-space coordinate

peculiar velocity along the line of sight

$$\boxed{\mathbf{s} = \mathbf{x} + \frac{v_z}{aH(a)}\hat{z}}$$

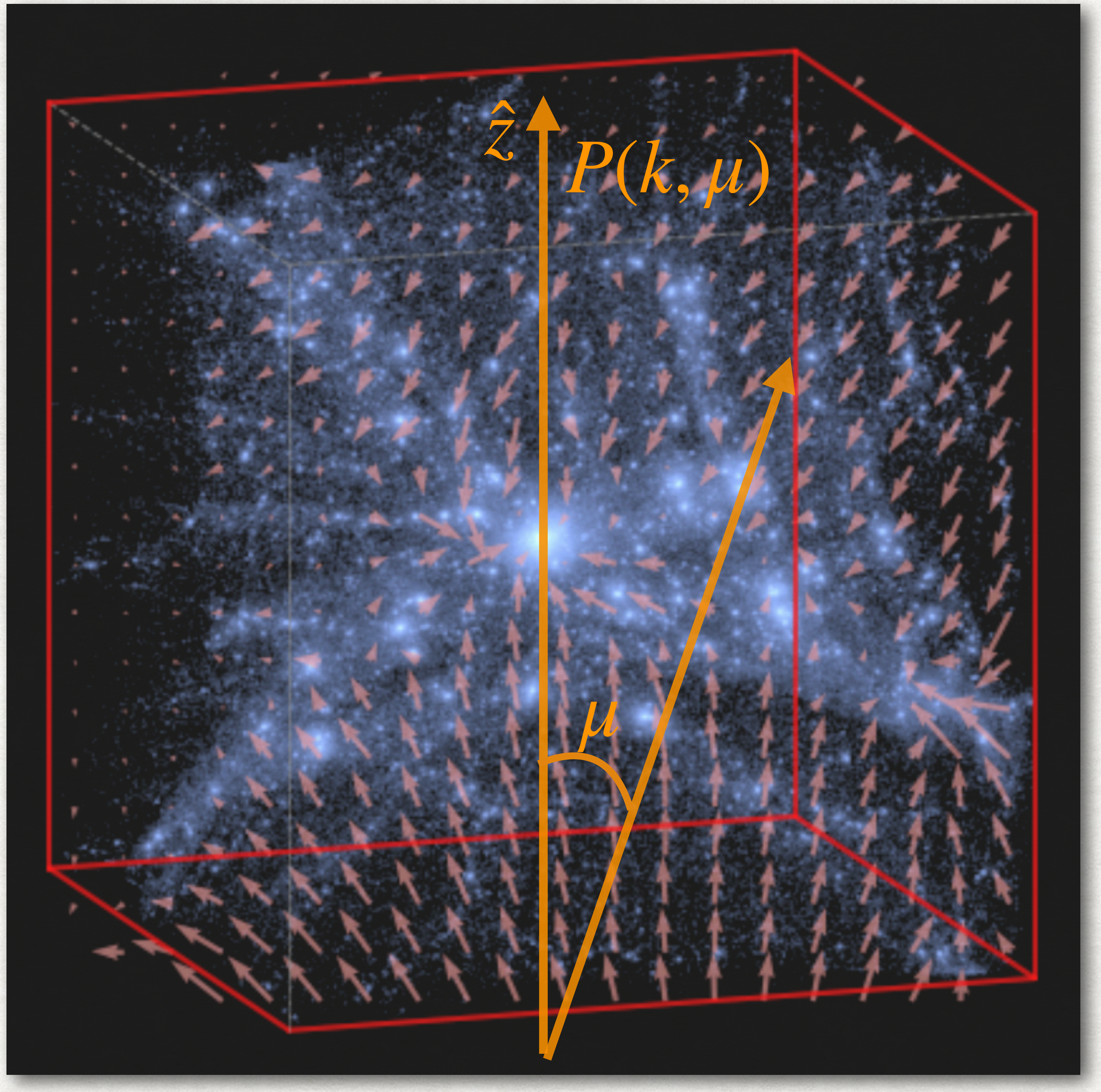
redshift-space coordinate

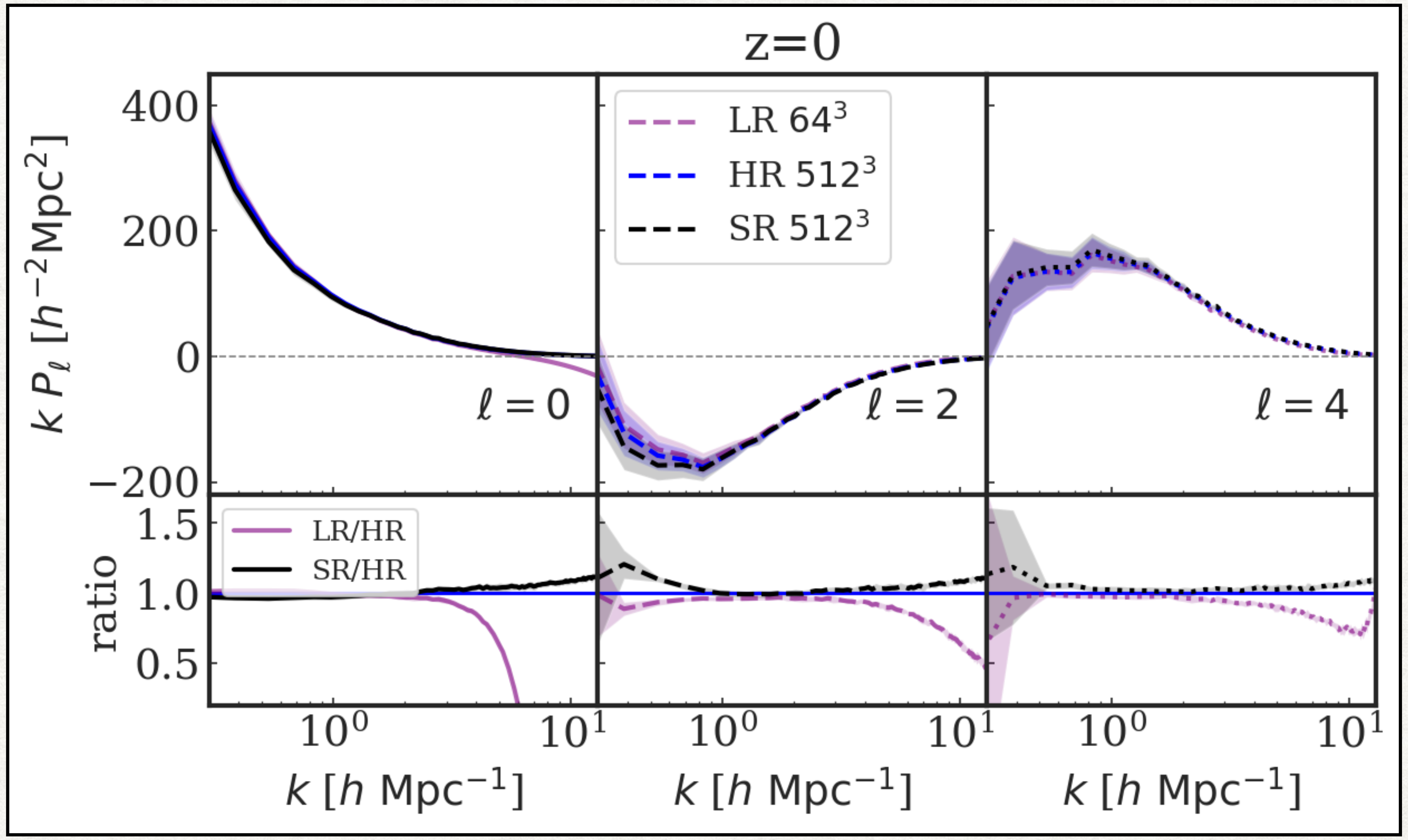
The line of sight direction

a: scale factor
$H(a)$: Hubble expansion rate

Full field statistics: Redshift-space distortion

The peculiar velocity makes the redshift-space clustering anisotropic —> 2D power spectrum $P(k, \mu)$





$$P_\ell(k) = (2\ell + 1) \int_0^1 d\mu P(k, \mu) \mathscr{L}_\ell(\mu)$$

Validation Metrics

Full field statistics :

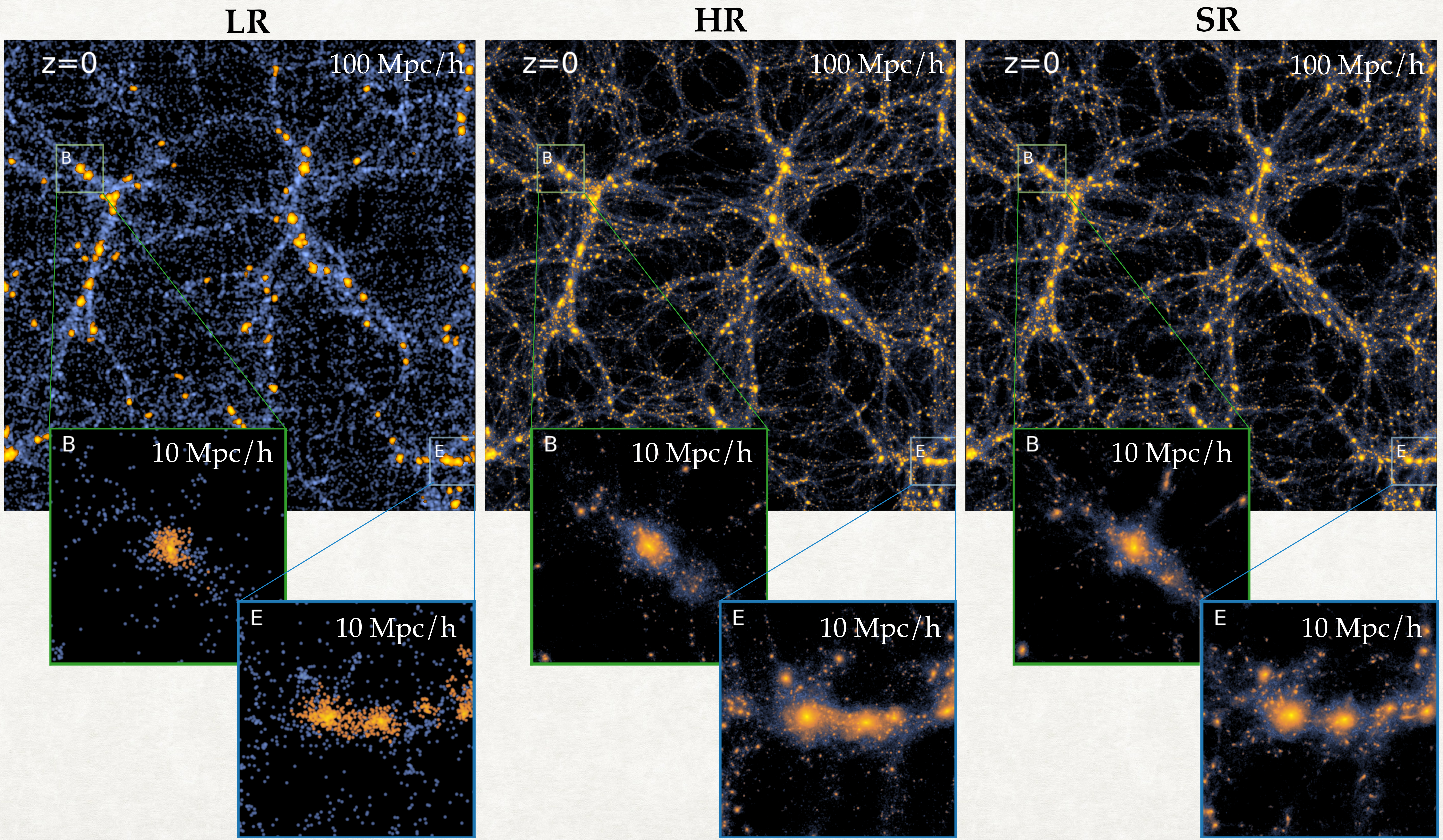
- Matter power spectrum (two point statistics)
- Bispectrum (three point statistics)
- Redshift space 2D power spectrum (velocity)

Test Sets:
10 pairs of LR-HR simulations
BoxSize = 100 Mpc/h
Same cosmology and resolution as the training sets

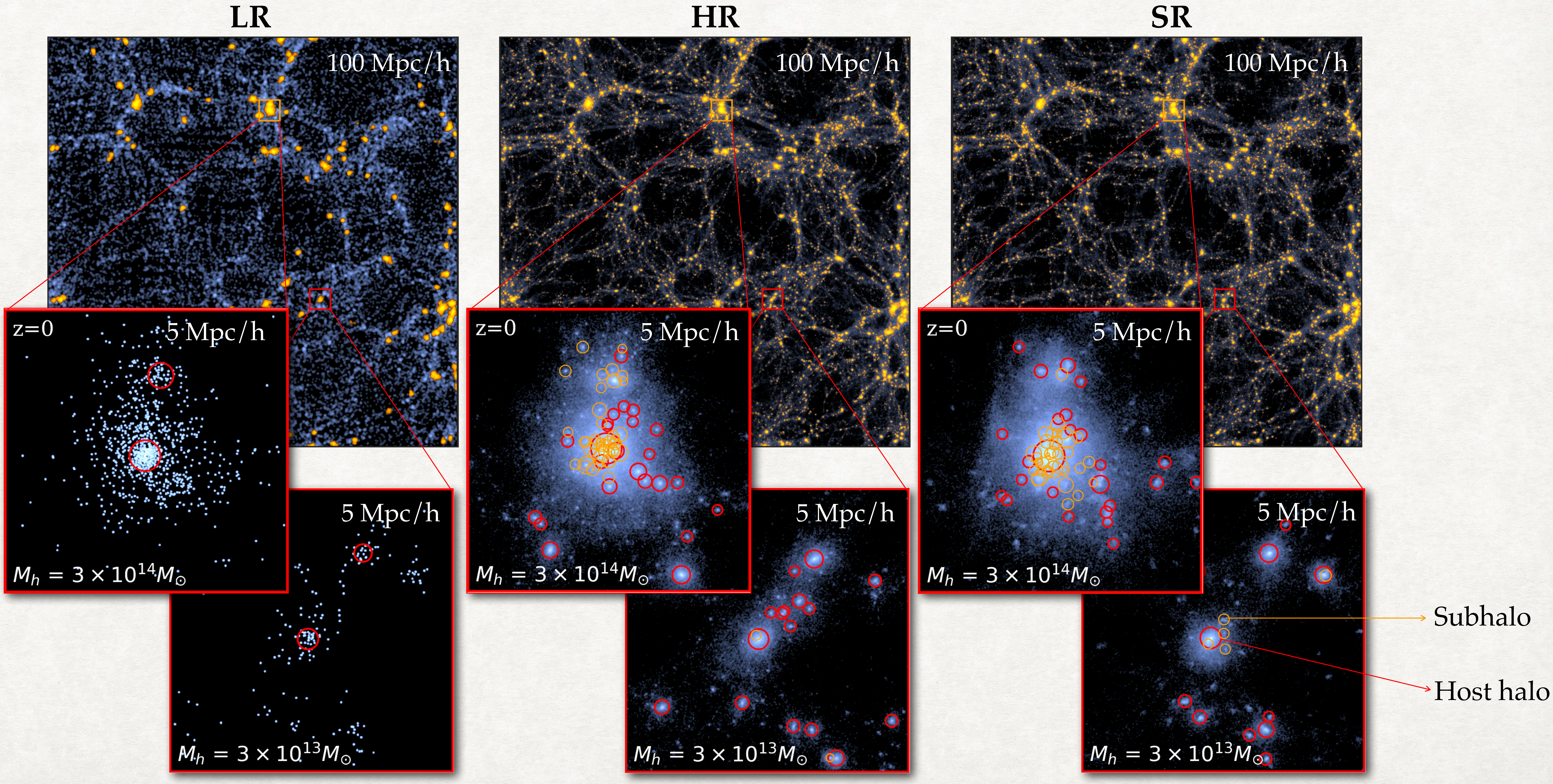
Halo catalog statistics:

- Abundance of halos and subhalos
- Mean occupation distribution of subhalos
- 2-point correlation function
- Redshift-space correlation
- Pairwise velocity distribution

Halo catalogs : halos

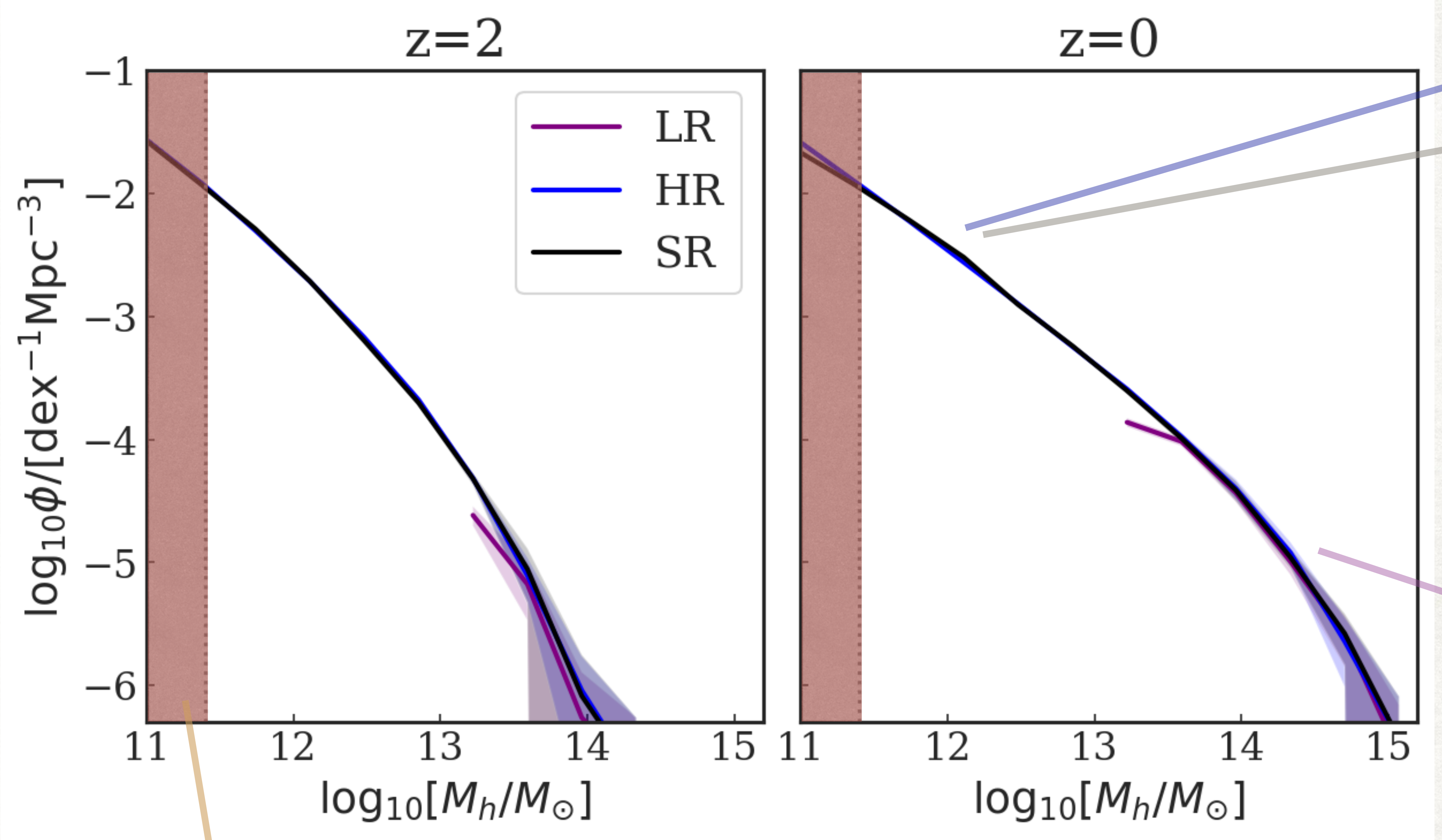


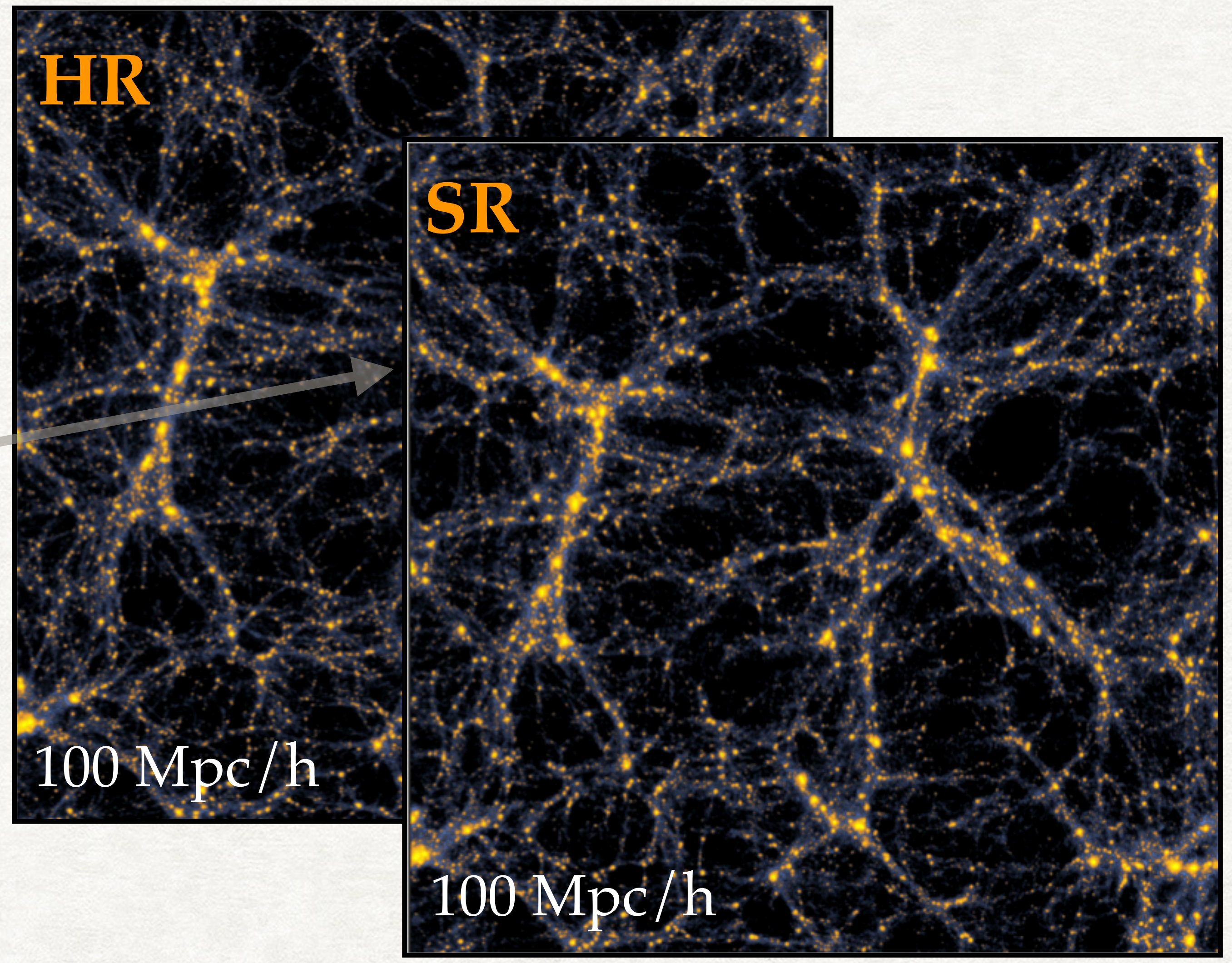
Halo catalogs : subhalos

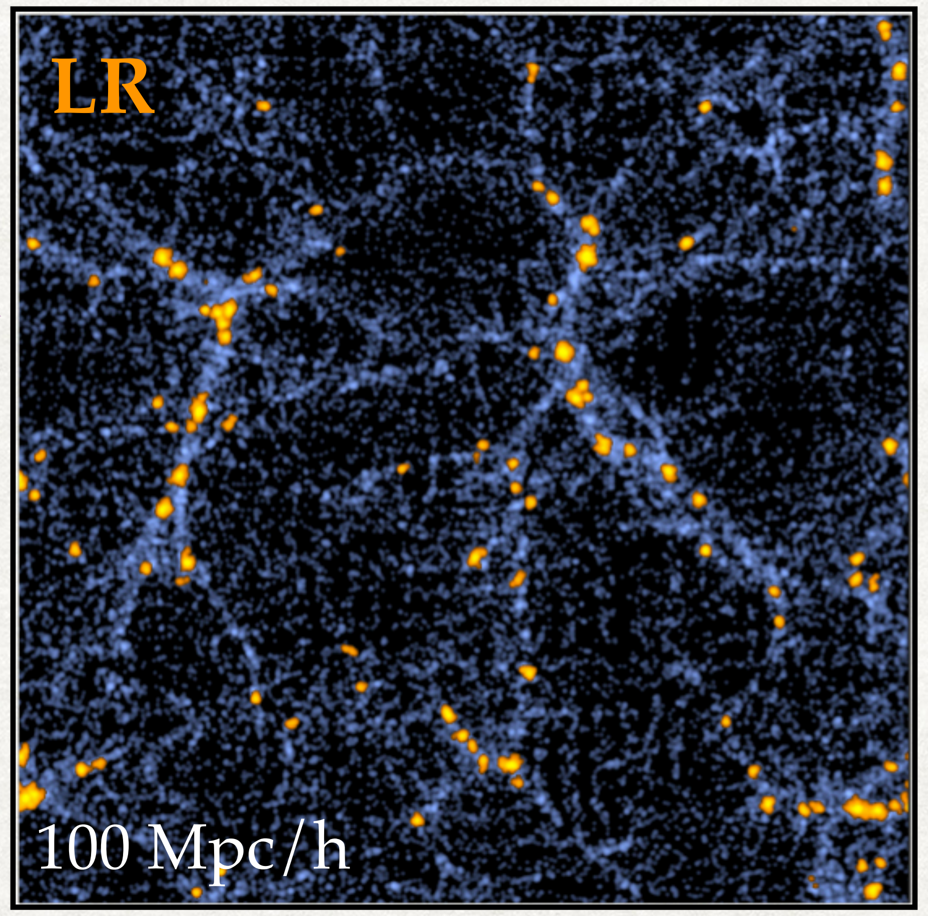


Halo catalog statistics : halo abundance

Abundance of **host halos**



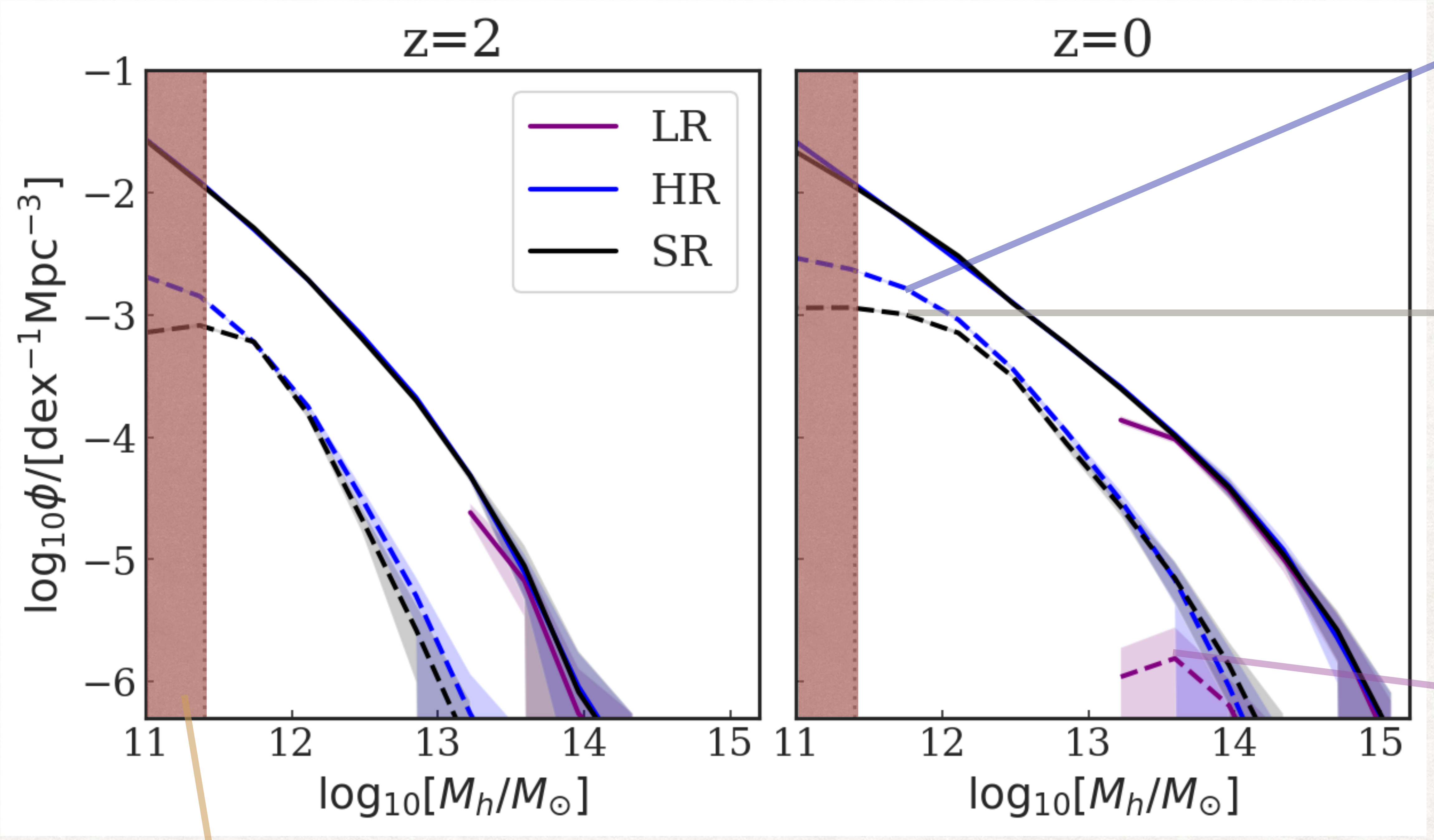




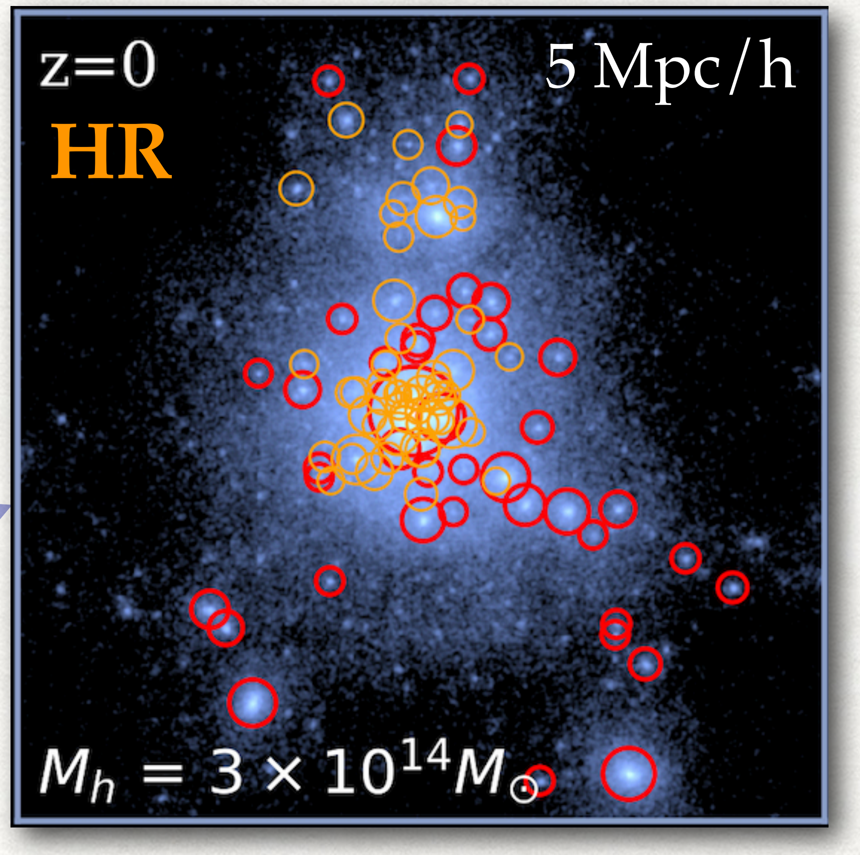
Resolution limit: 300 particles

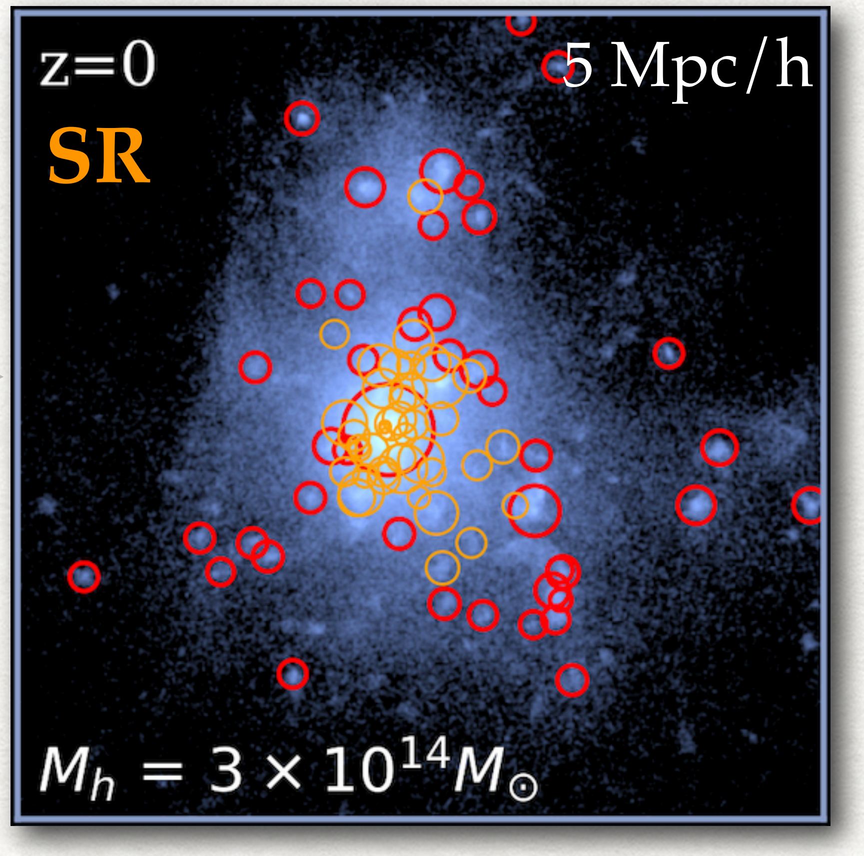
Halo catalog statistics : subhalo abundance

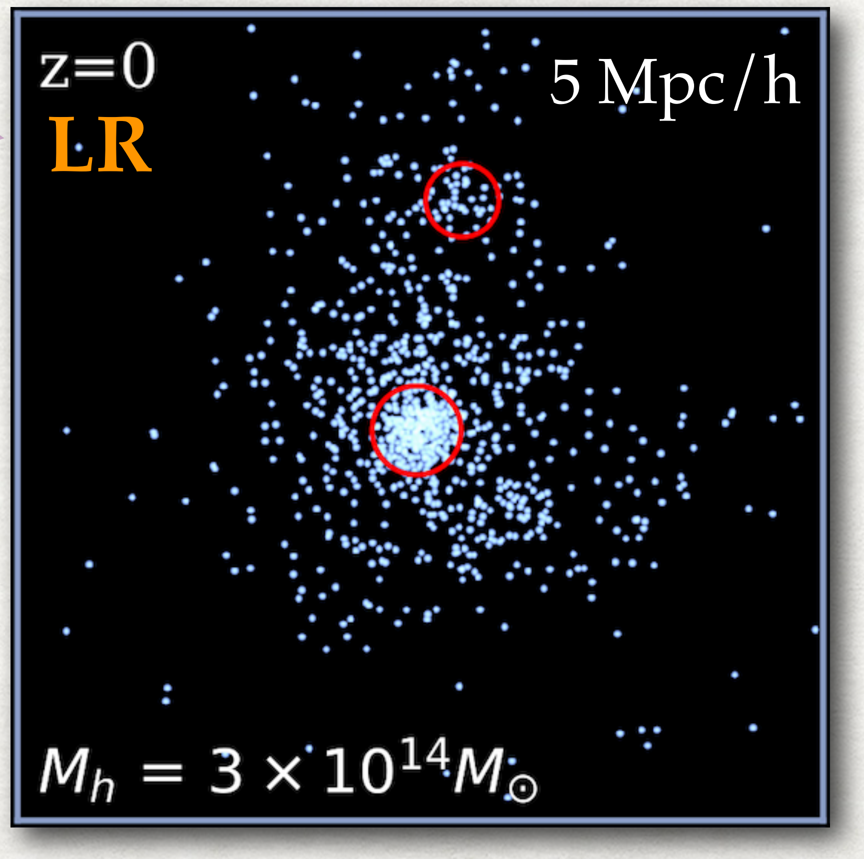
Abundance of **subhalos**



Resolution limit: 300 particles

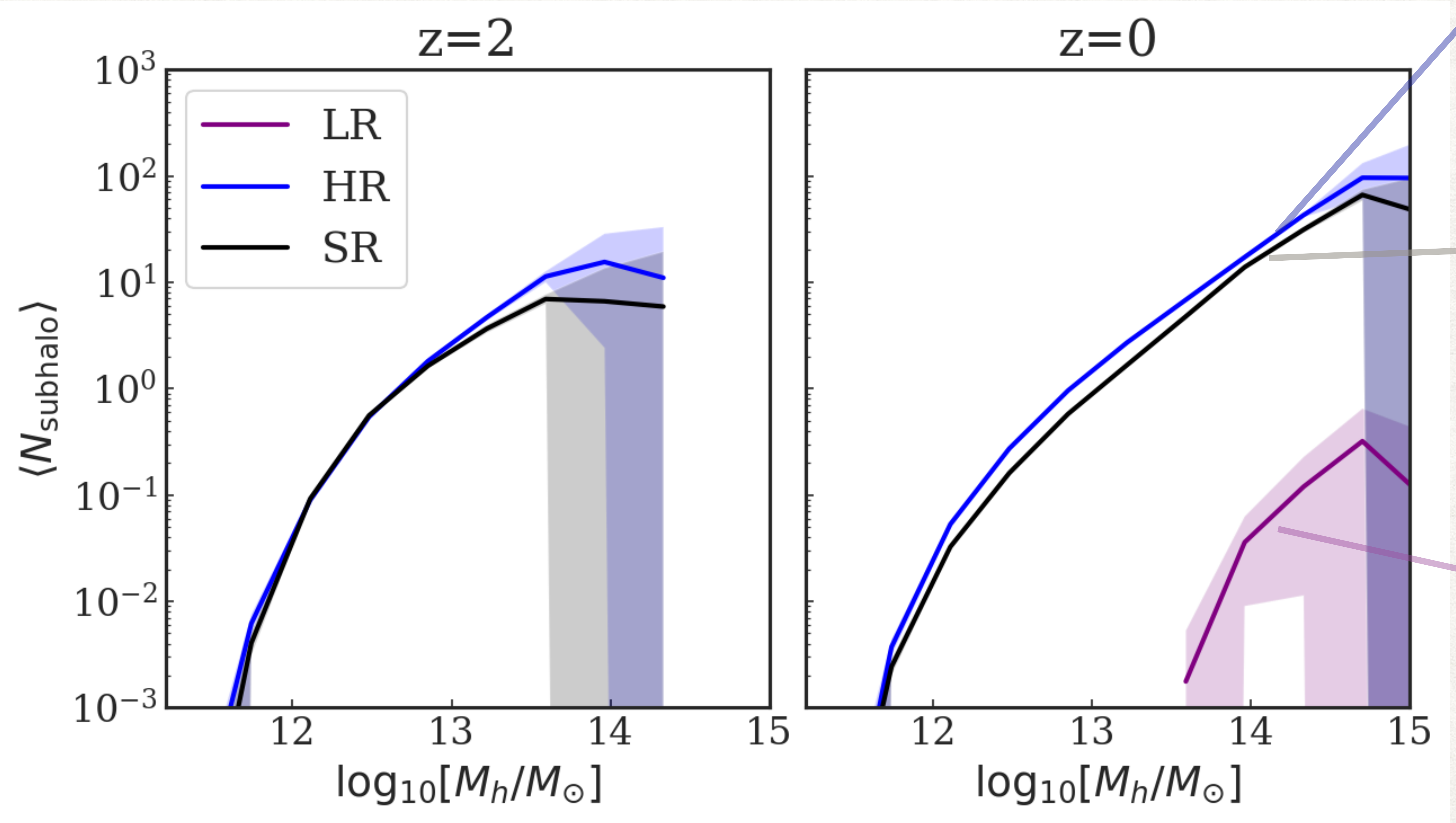


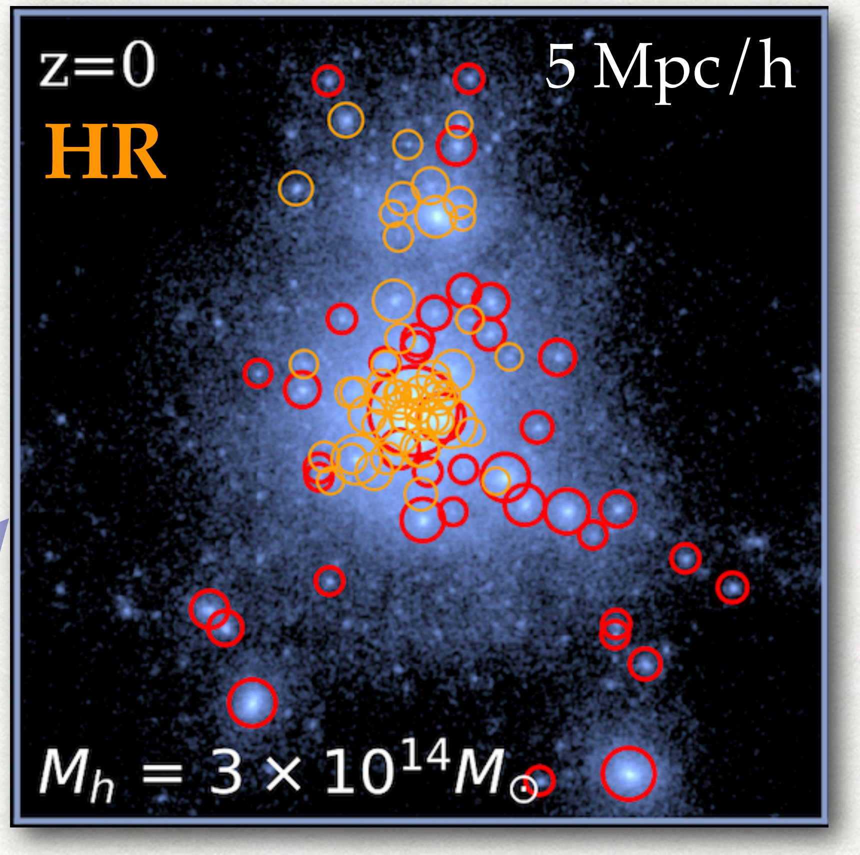


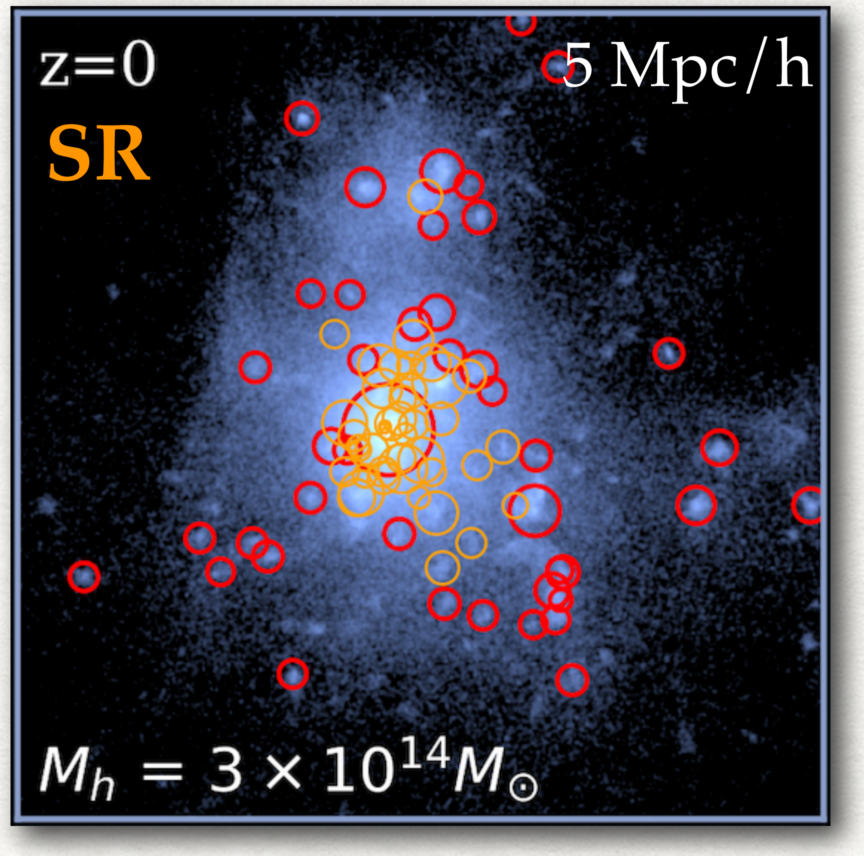


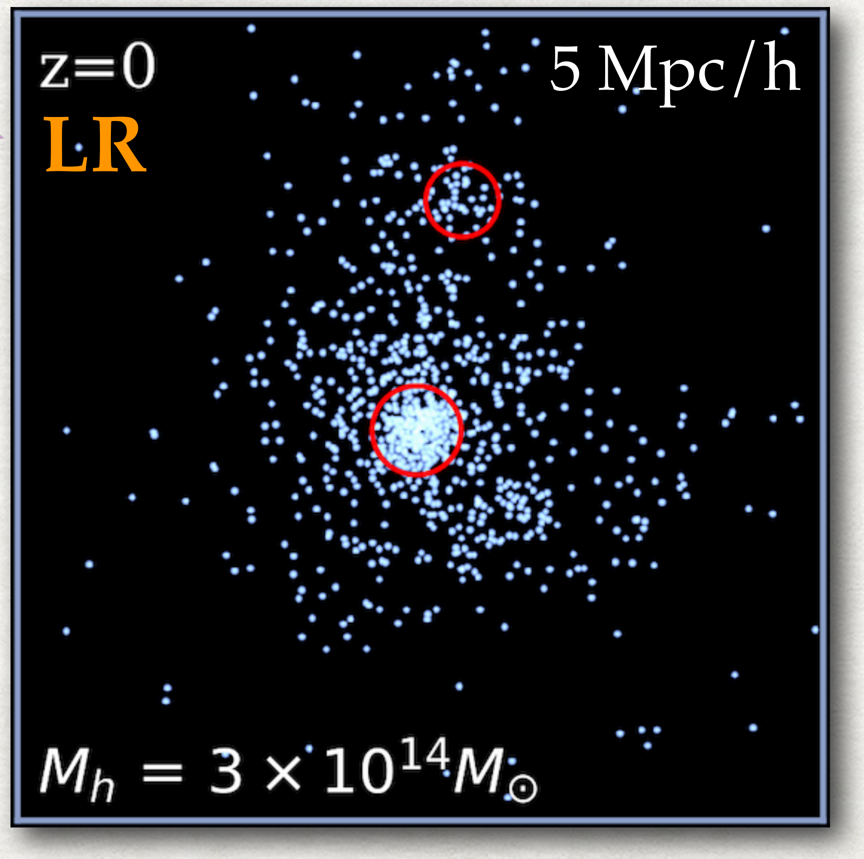
Halo catalog statistics : occupation distribution of subhalo

$\langle N_{\text{subhalo}} \,|\, M_{\text{host}} \rangle$: Number of **subhalos** in the host mass bin









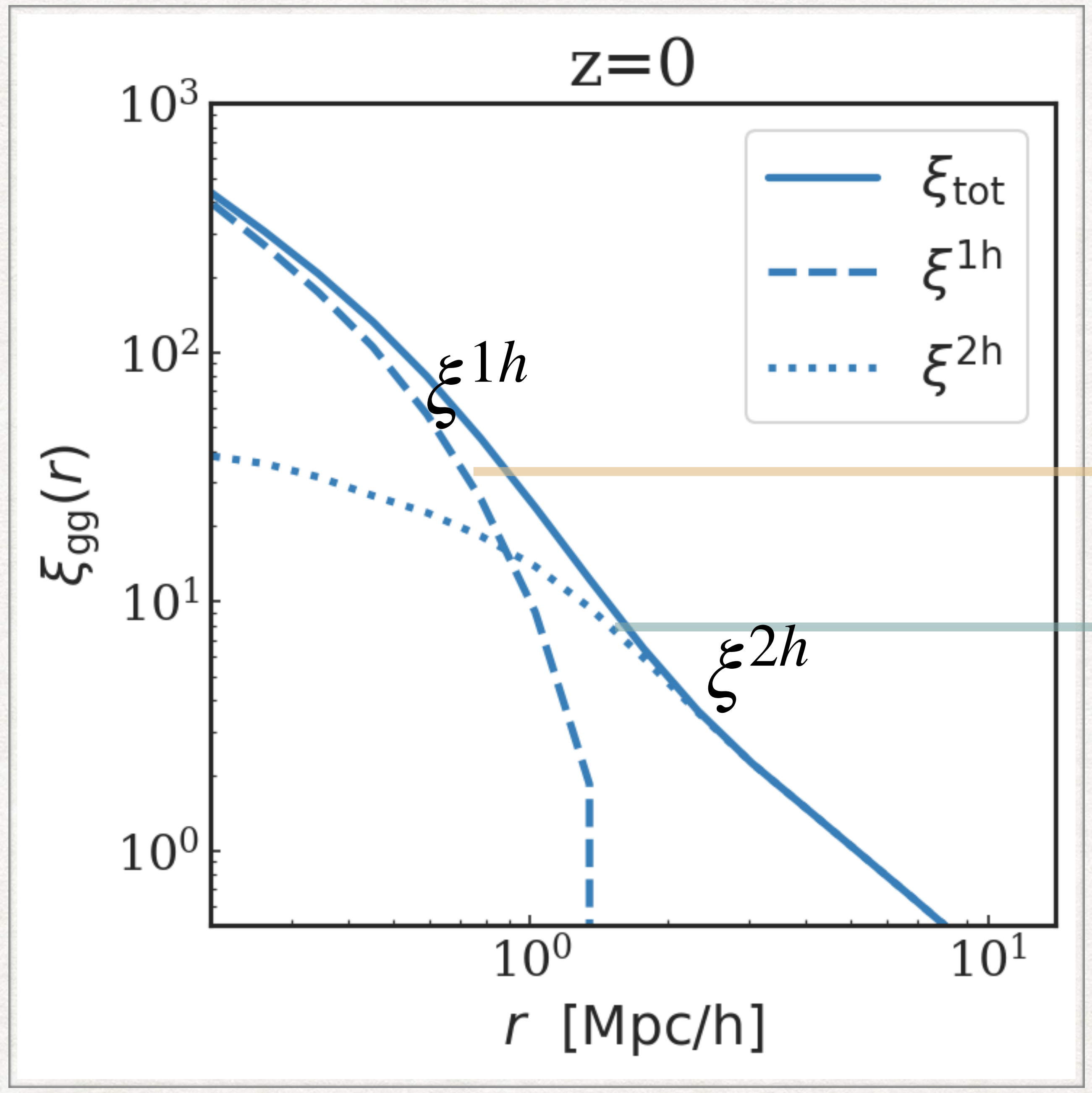
Halo catalog statistics : 2-point correlation of halos

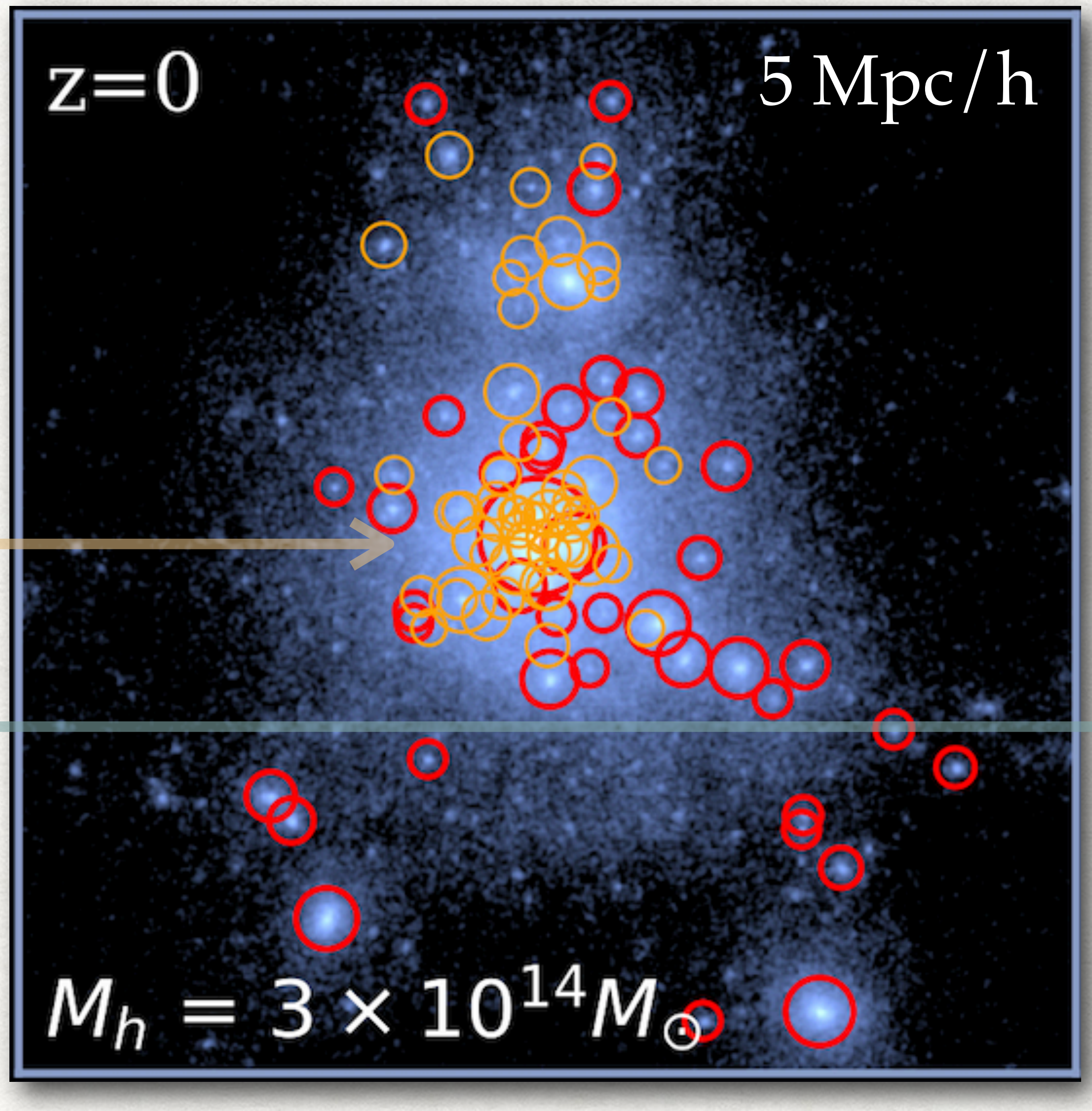
$\xi(r) \equiv \mathrm{DD}(r)/\mathrm{RR}(r) - 1$

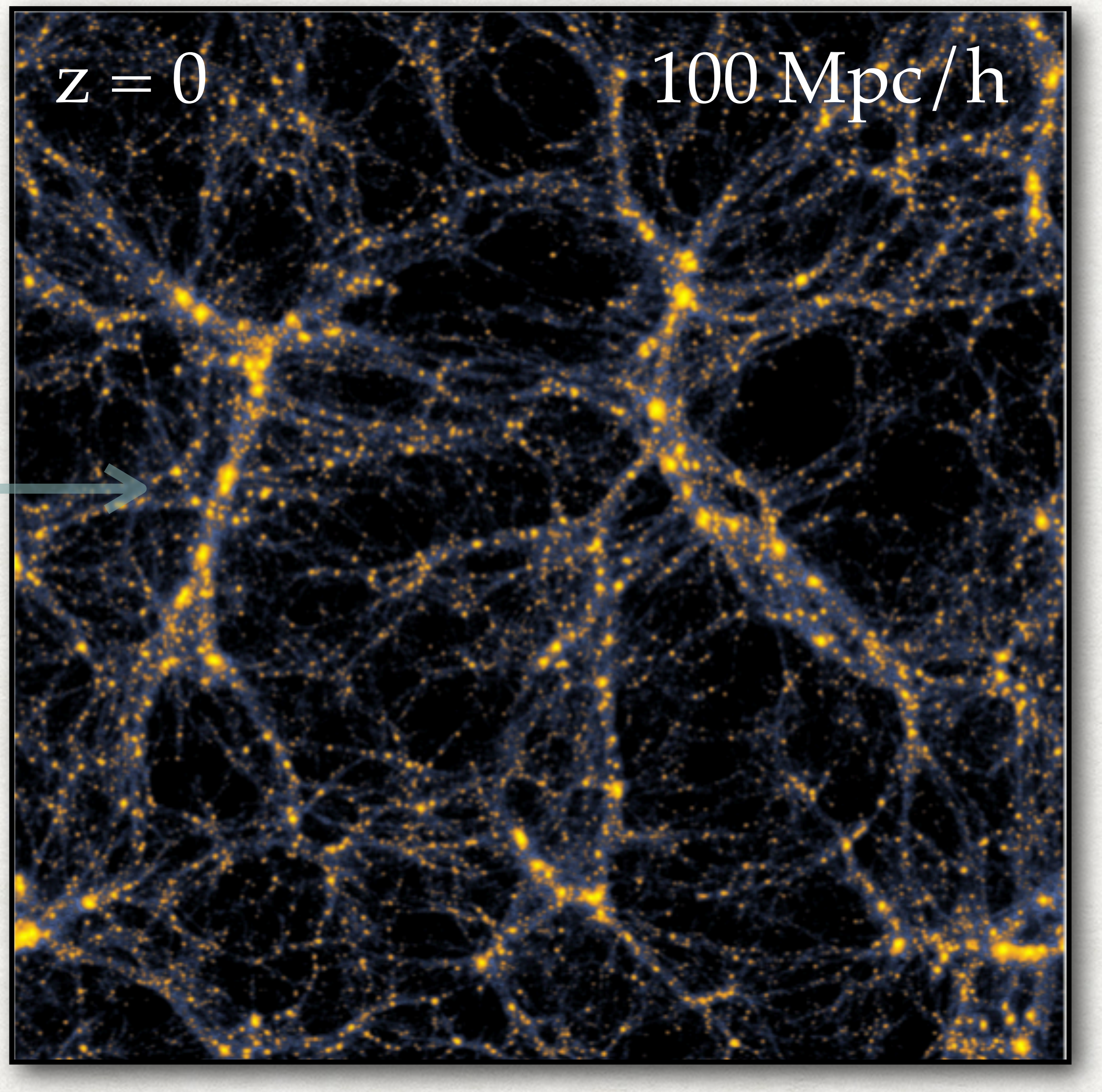
$DD(r)$ $(RR(r))$: the number of **sample pairs (random pairs)** of halos with separations equal to r

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

$\xi^{1h}(r)$: one-halo term; $\xi^{2h}(r)$: two-halo term

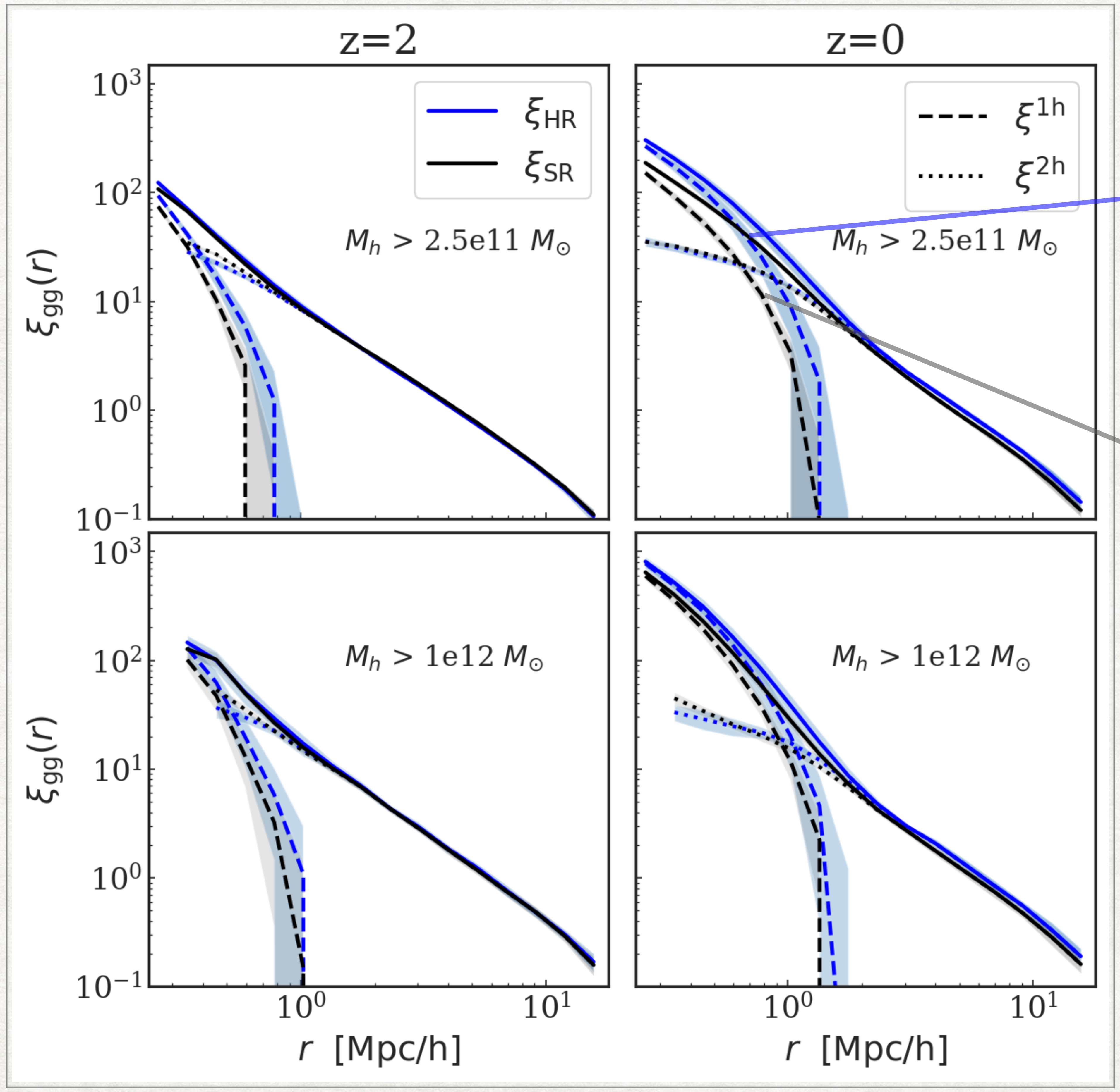


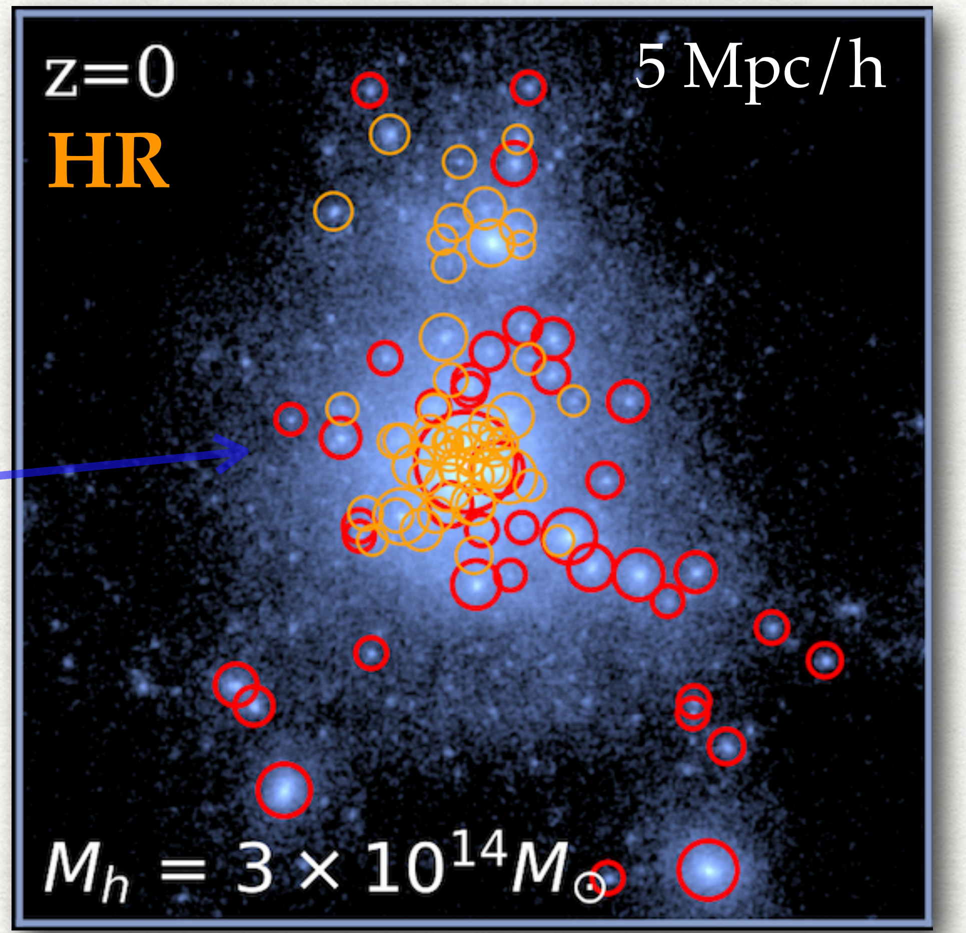


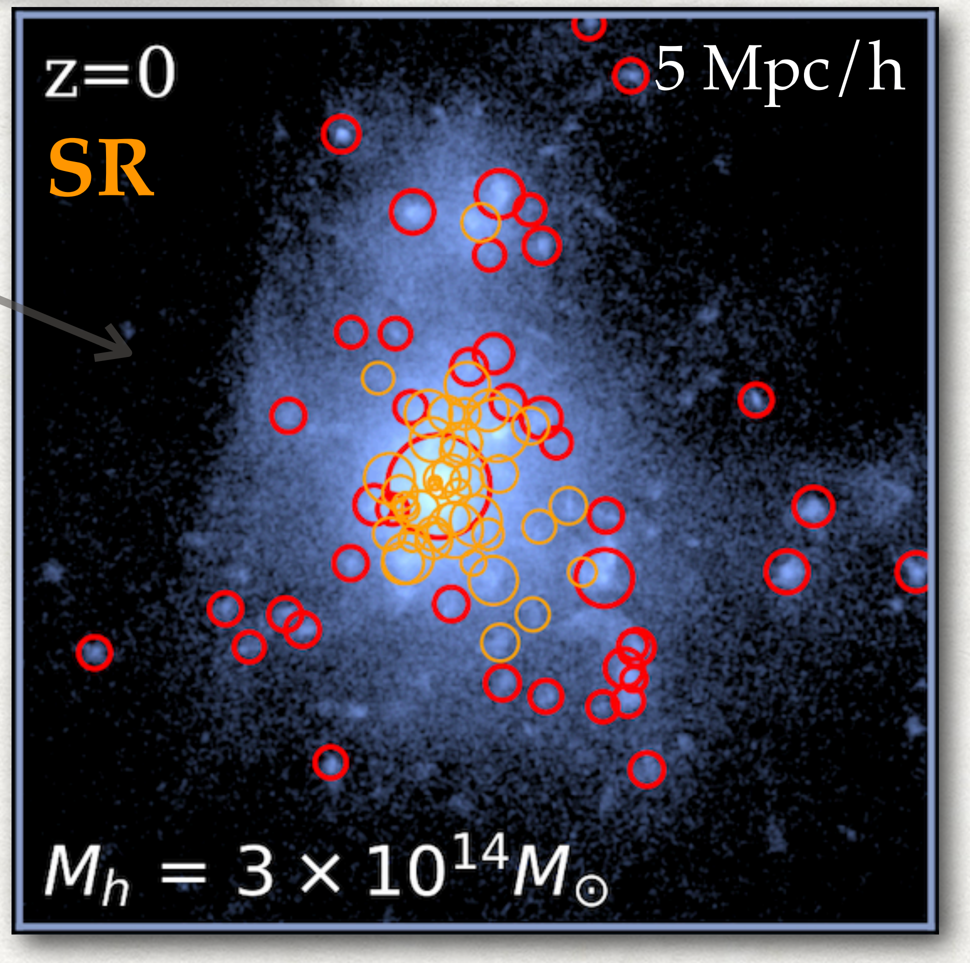


Halo catalog statistics : 2-point correlation of halos

$\xi(r) \equiv \mathrm{DD}(r)/\mathrm{RR}(r) - 1$







Halo catalog statistics : redshift-space correlation

The peculiar velocity makes the redshift-space clustering anisotropic

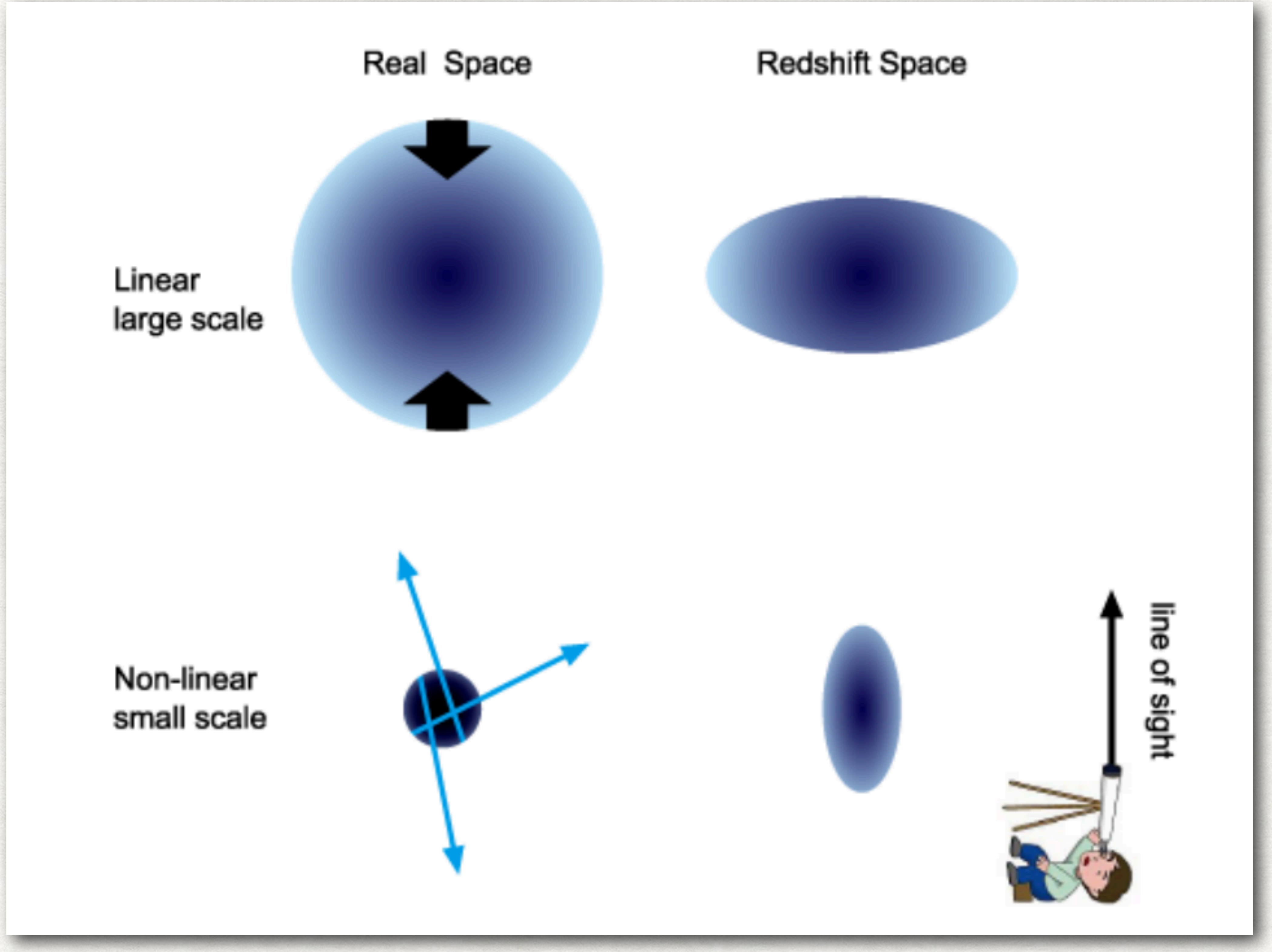


Image from: Shun Saito
RSD lecture note

2D contour of $\xi(\pi, r_p)$

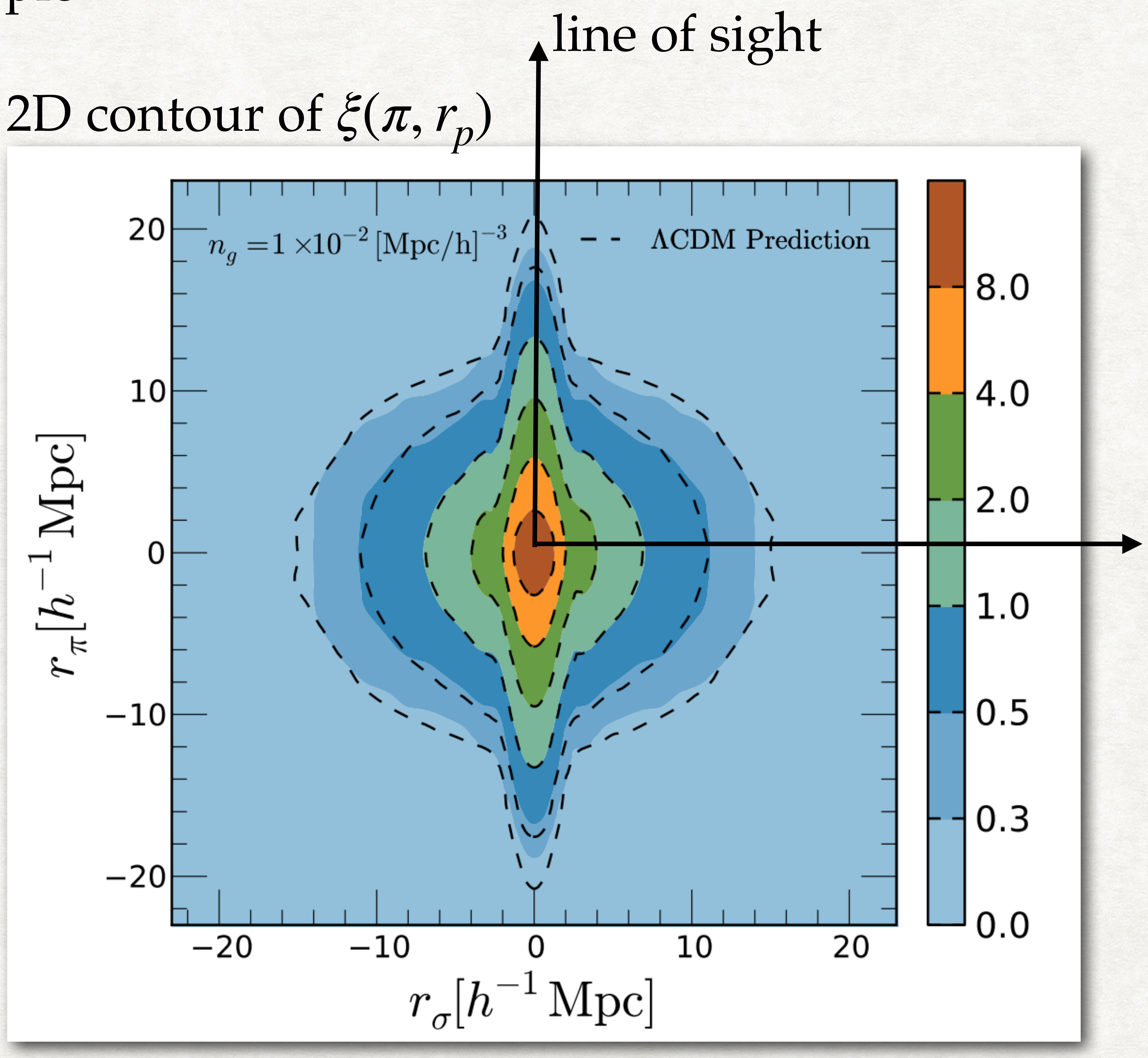
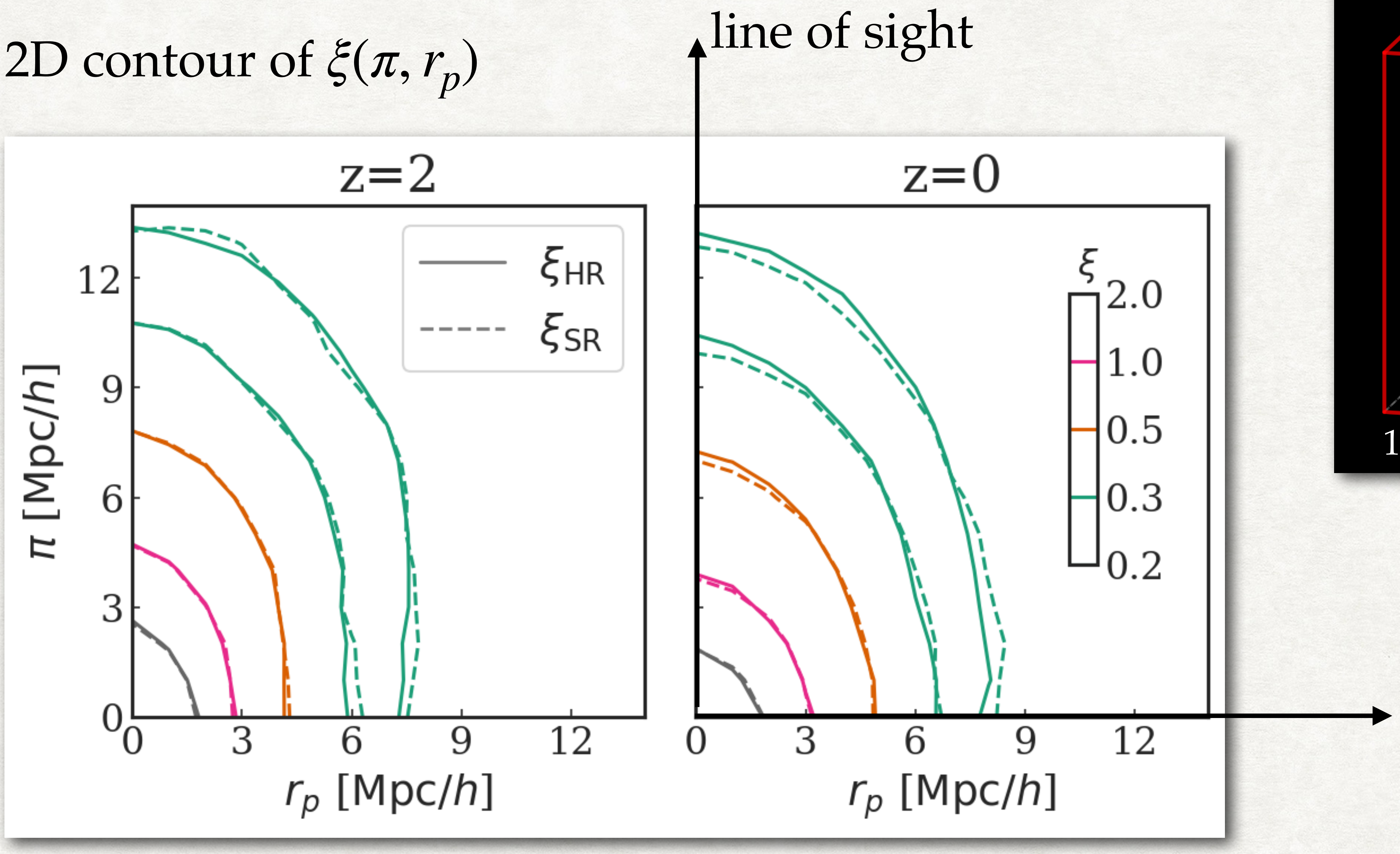
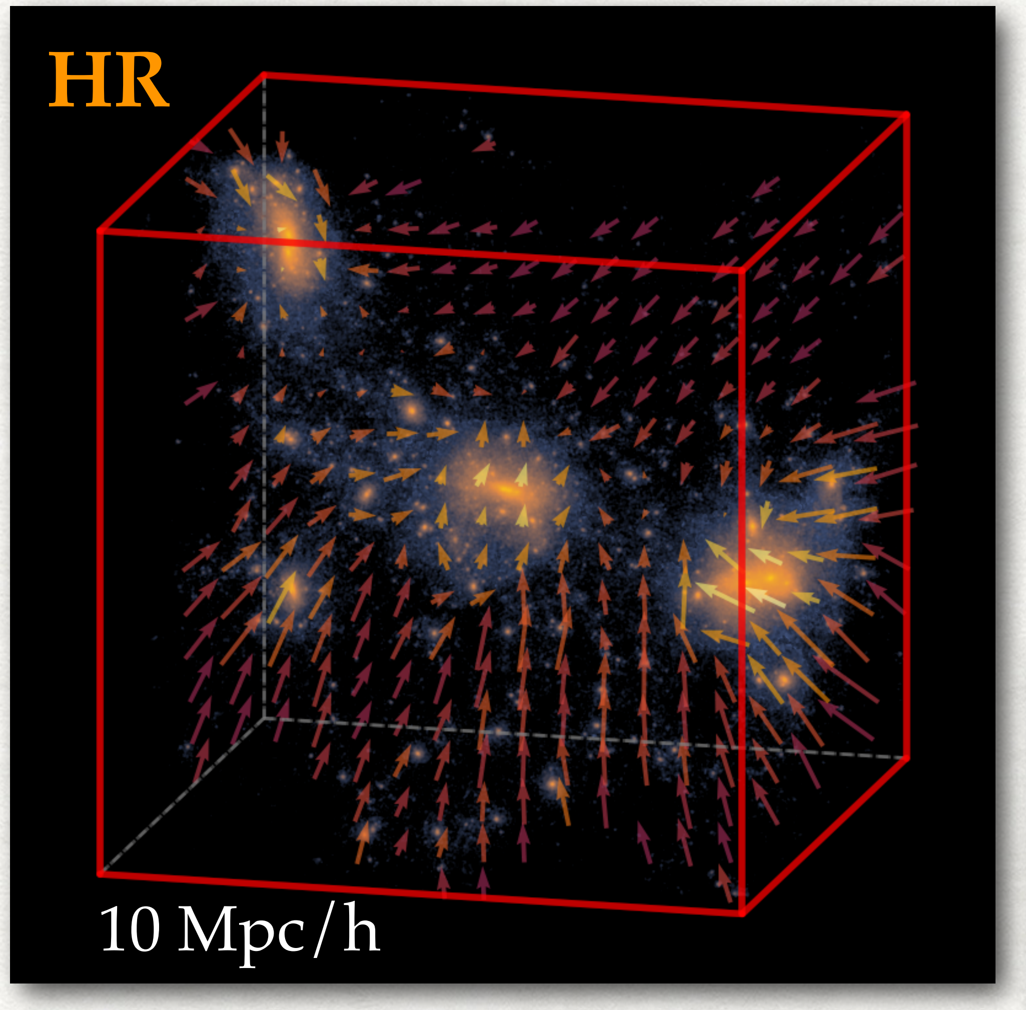


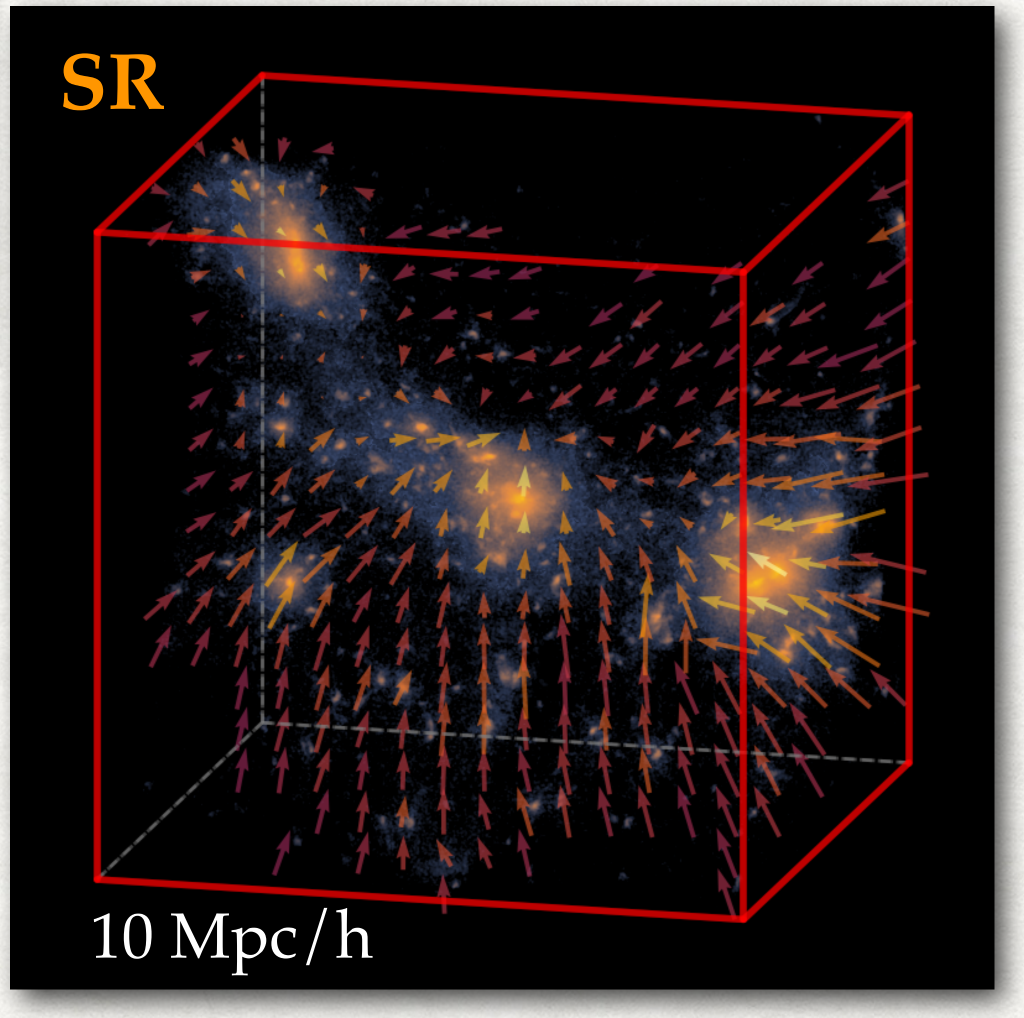
Image from: J. He et al.
Nature Astronomy 2, 967-972(2018)

Halo catalog statistics : redshift-space correlation

2D contour of $\xi(\pi, r_p)$

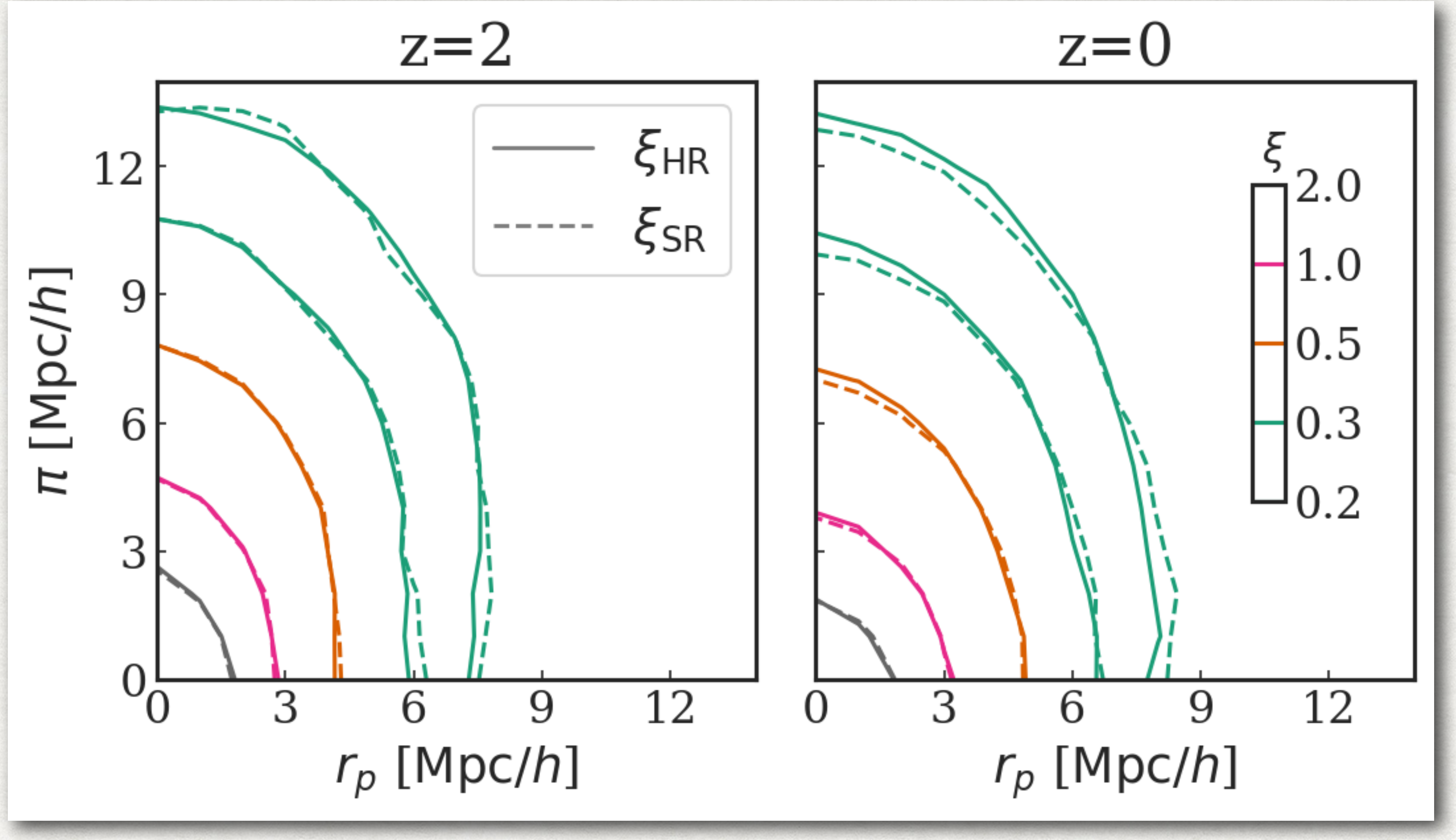




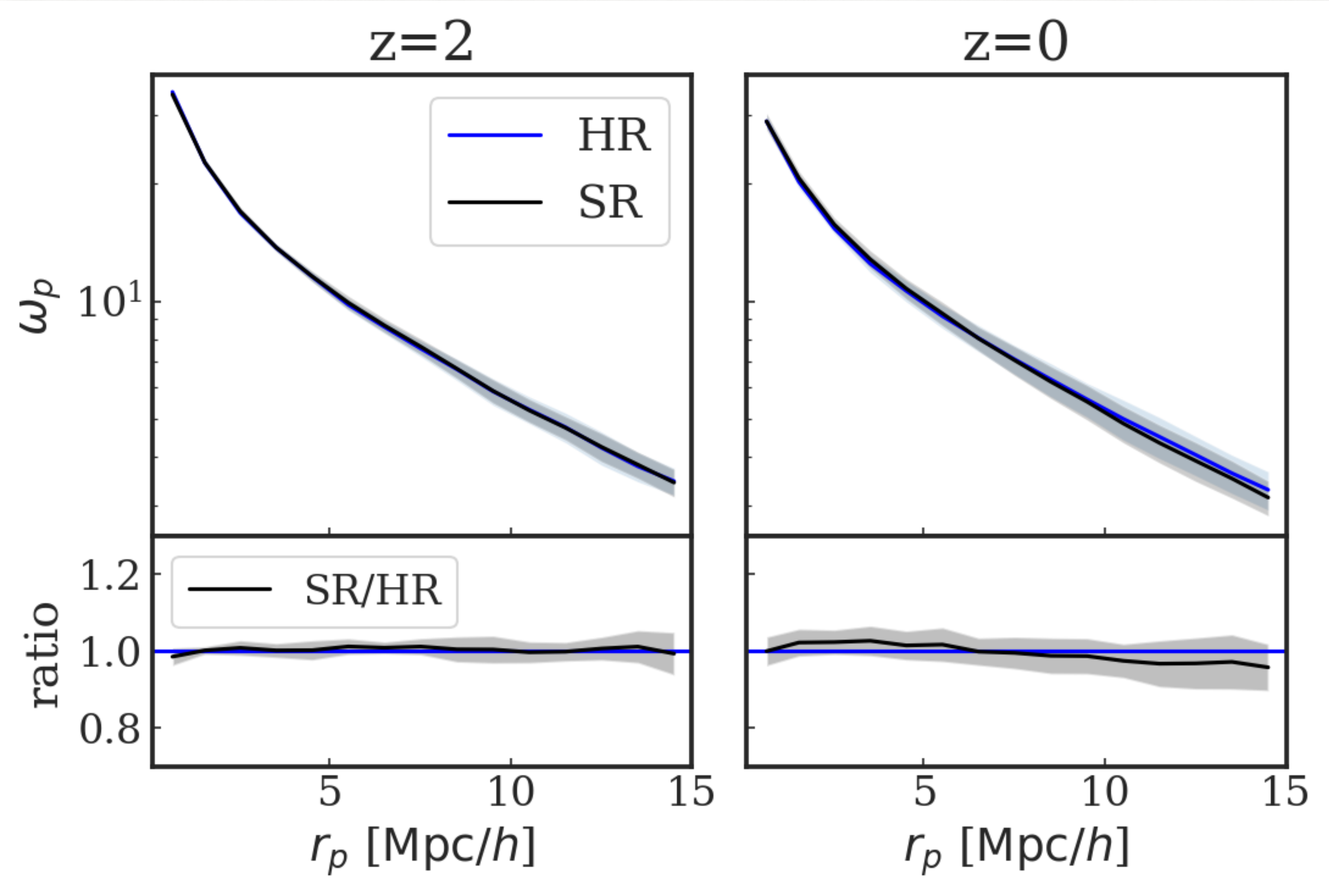


Halo catalog statistics : redshift-space correlation

2D contour of $\xi(\pi, r_p)$



$$\omega_p(r_p) = 2\int_0^\infty d\pi \xi(r_p, \pi)$$

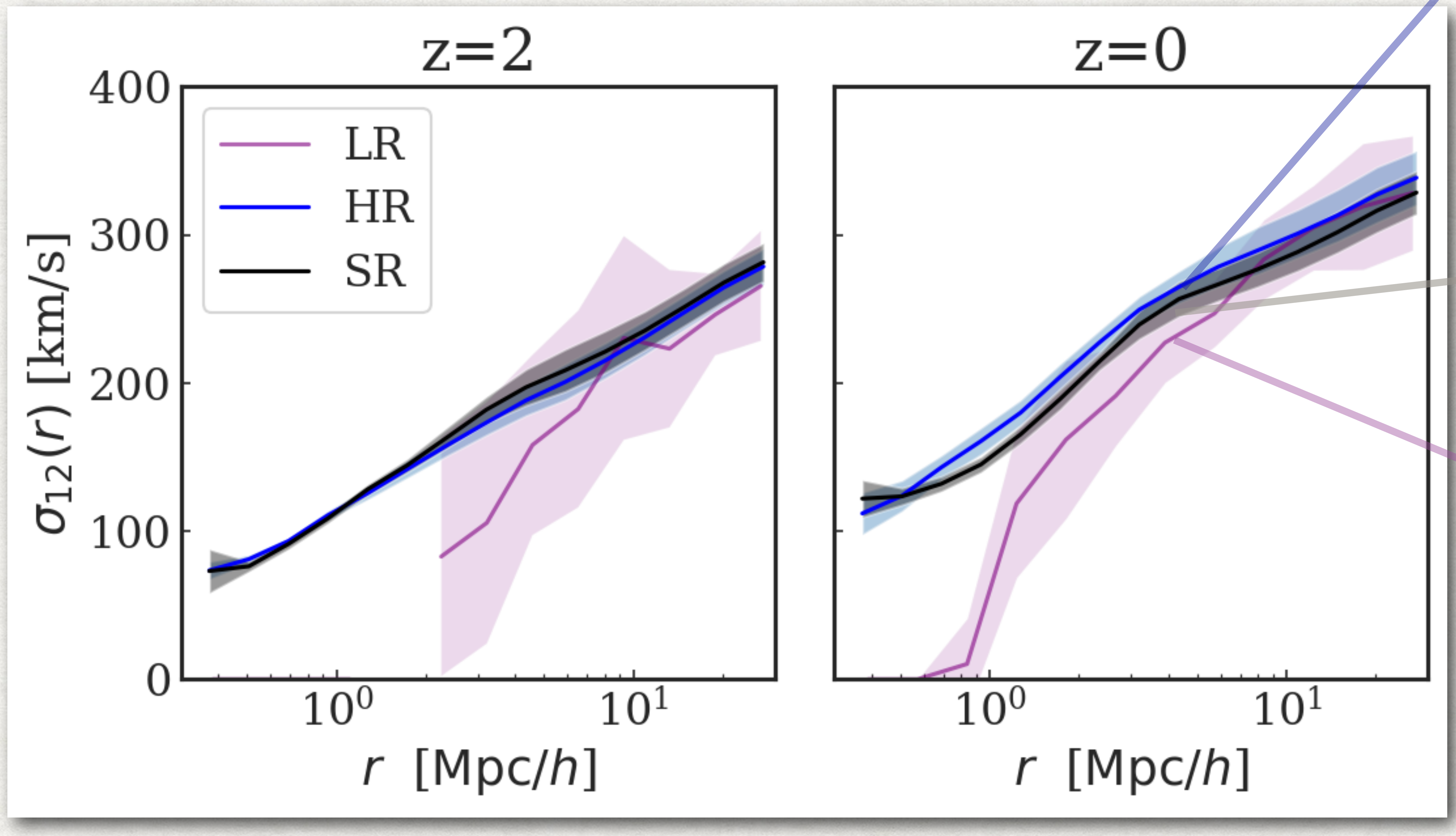


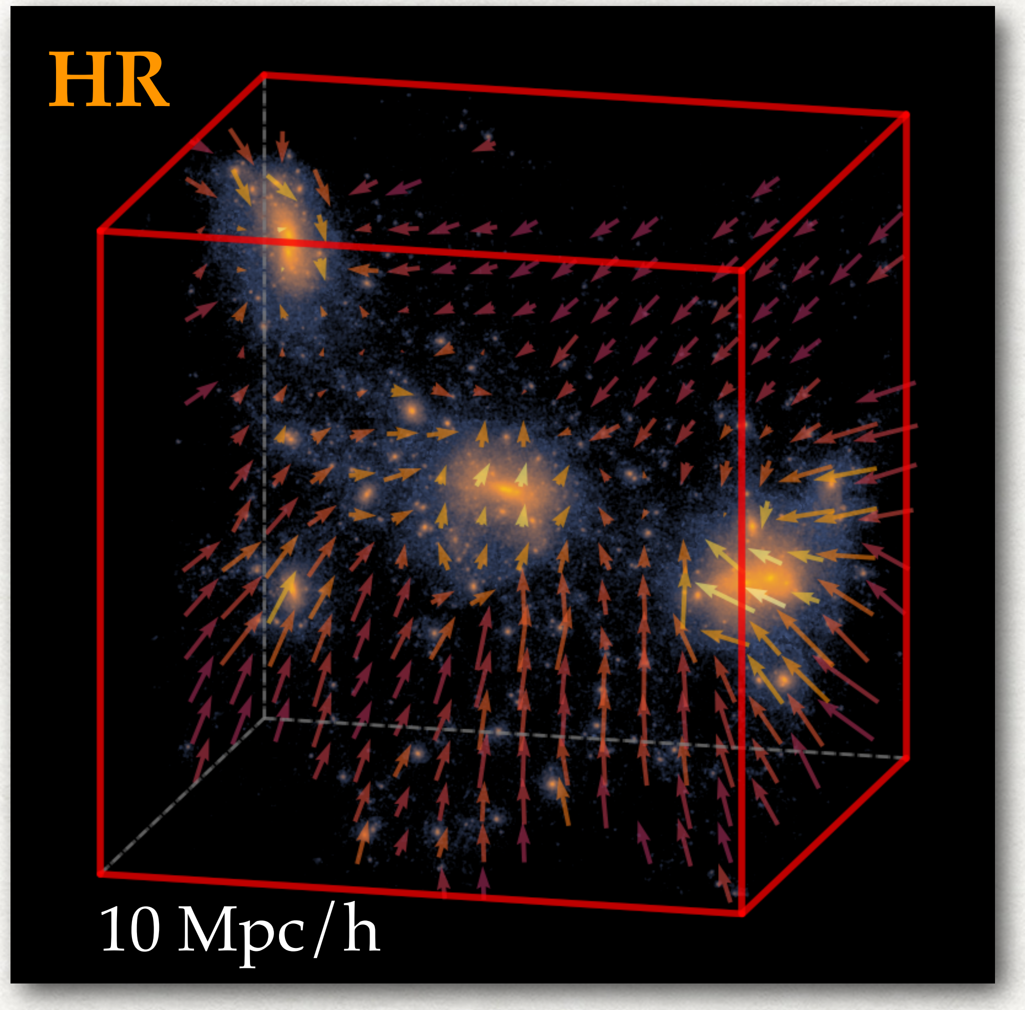
Halo catalog statistics : pairwise velocity of halos

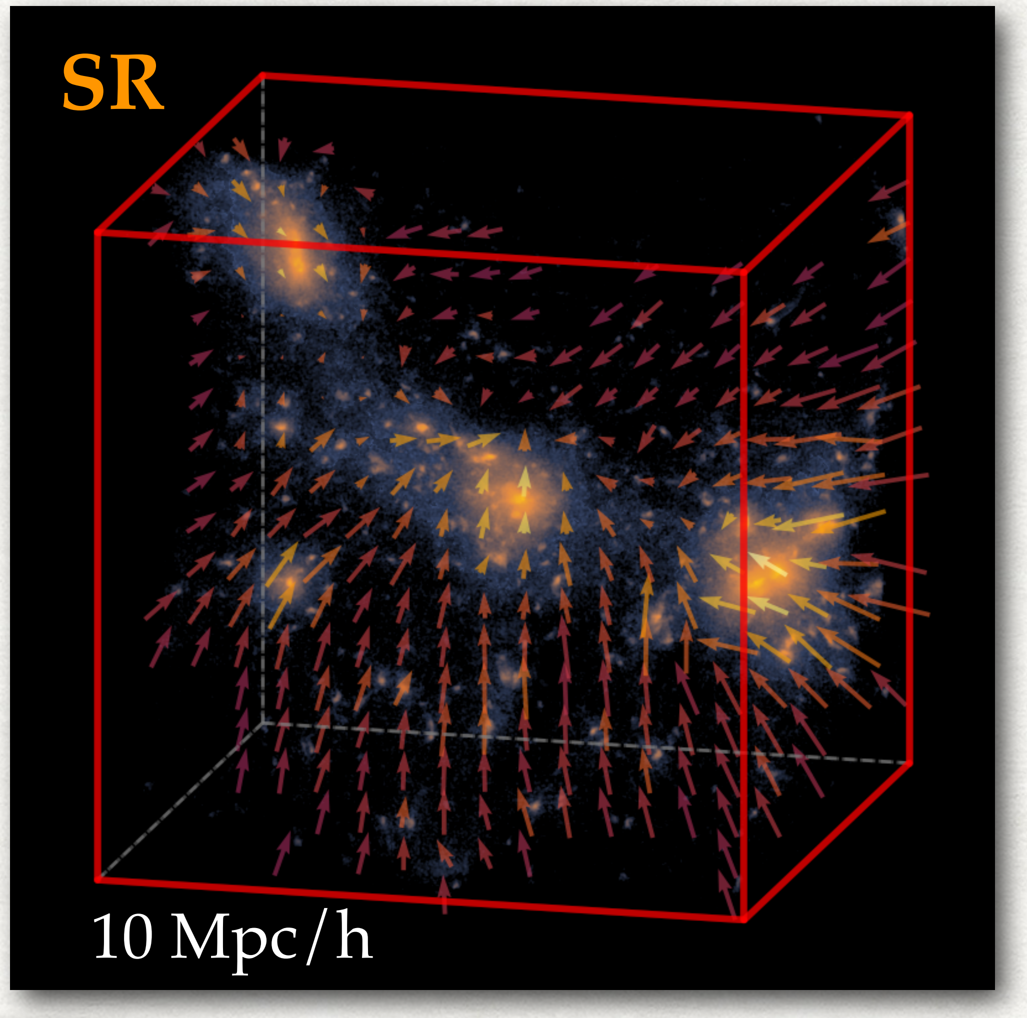
Pairwise velocity dispersion $\sigma_{12}(r)$:

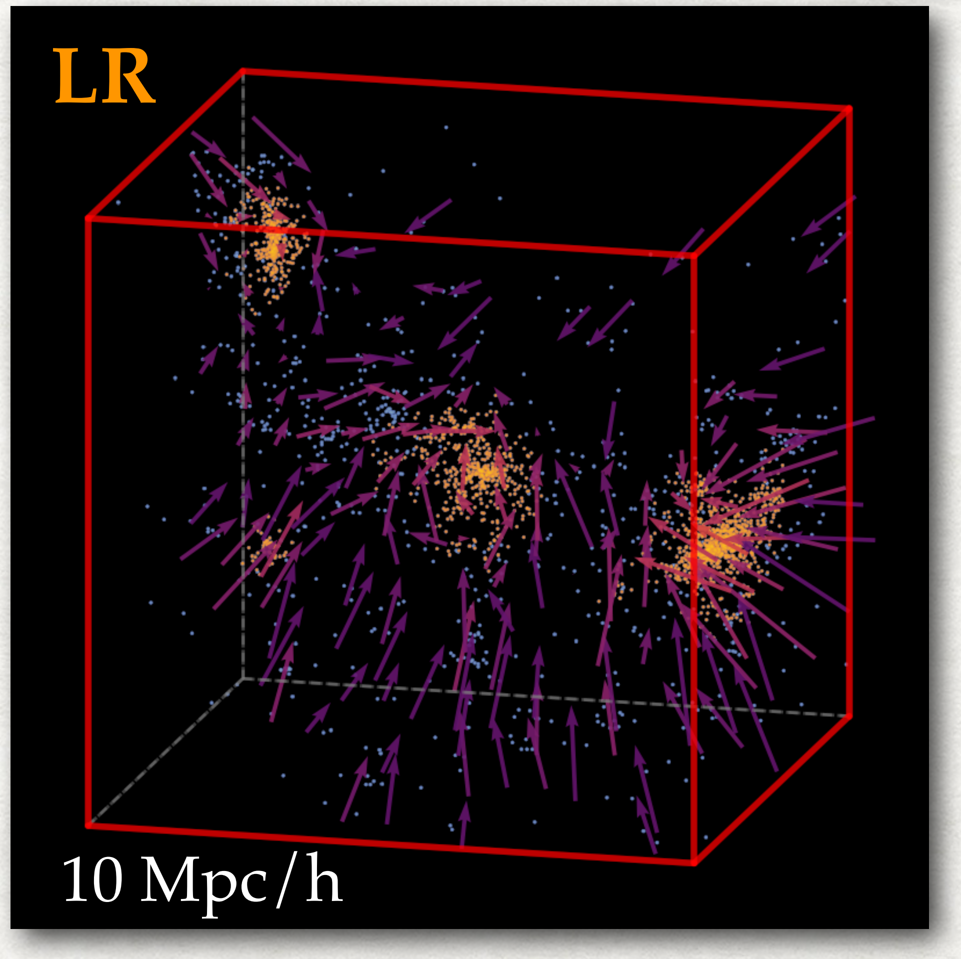
$$v_{12}(r) = \vec{v}_1 \cdot \vec{r}_{12} - \vec{v}_2 \cdot \vec{r}_{12}$$

$$\sigma_{12}(r) = \mathrm{std}(v_{12}(r))$$

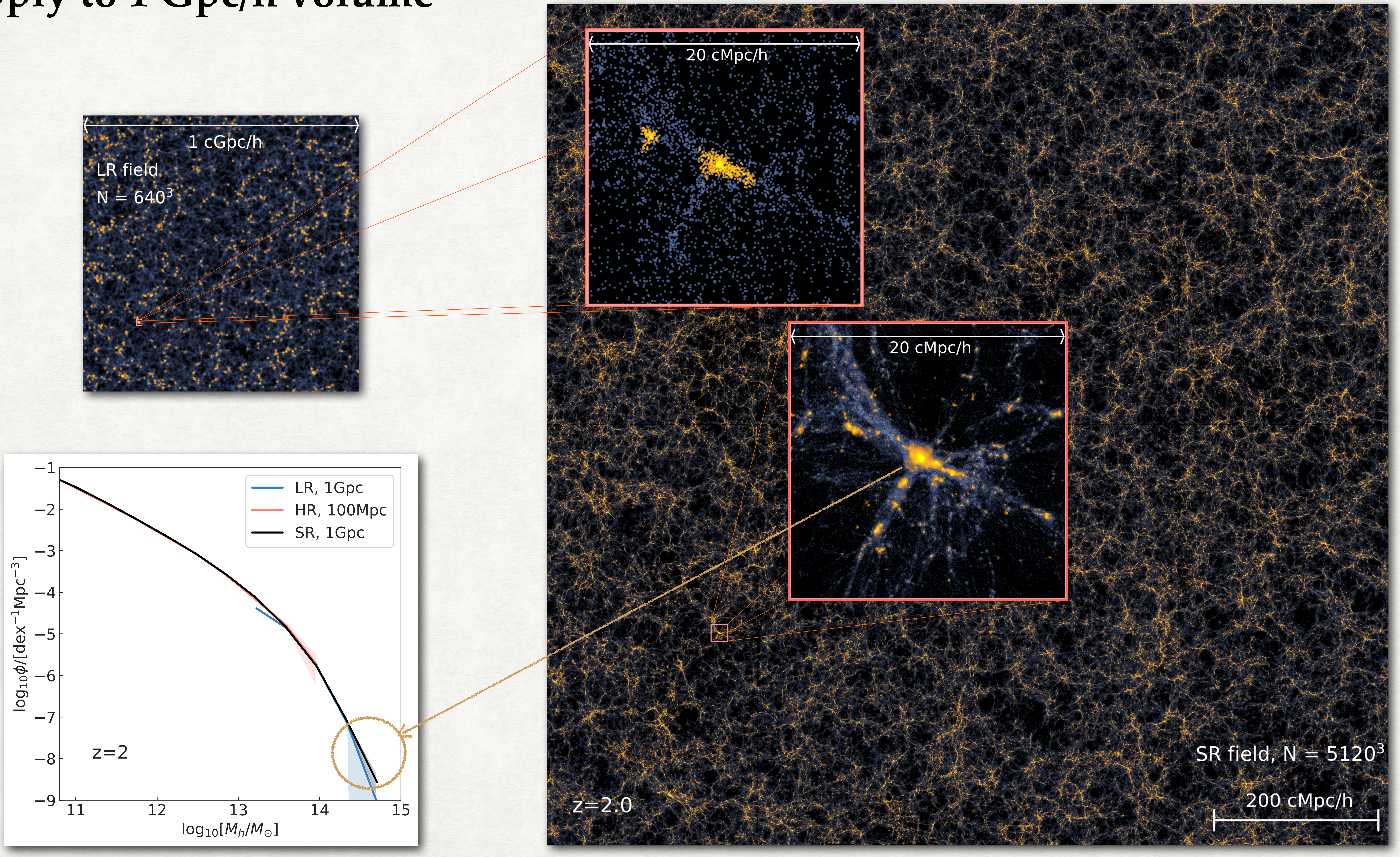








Apply to 1 Gpc/h volume



costs ~ 16 hours
with a single GPU

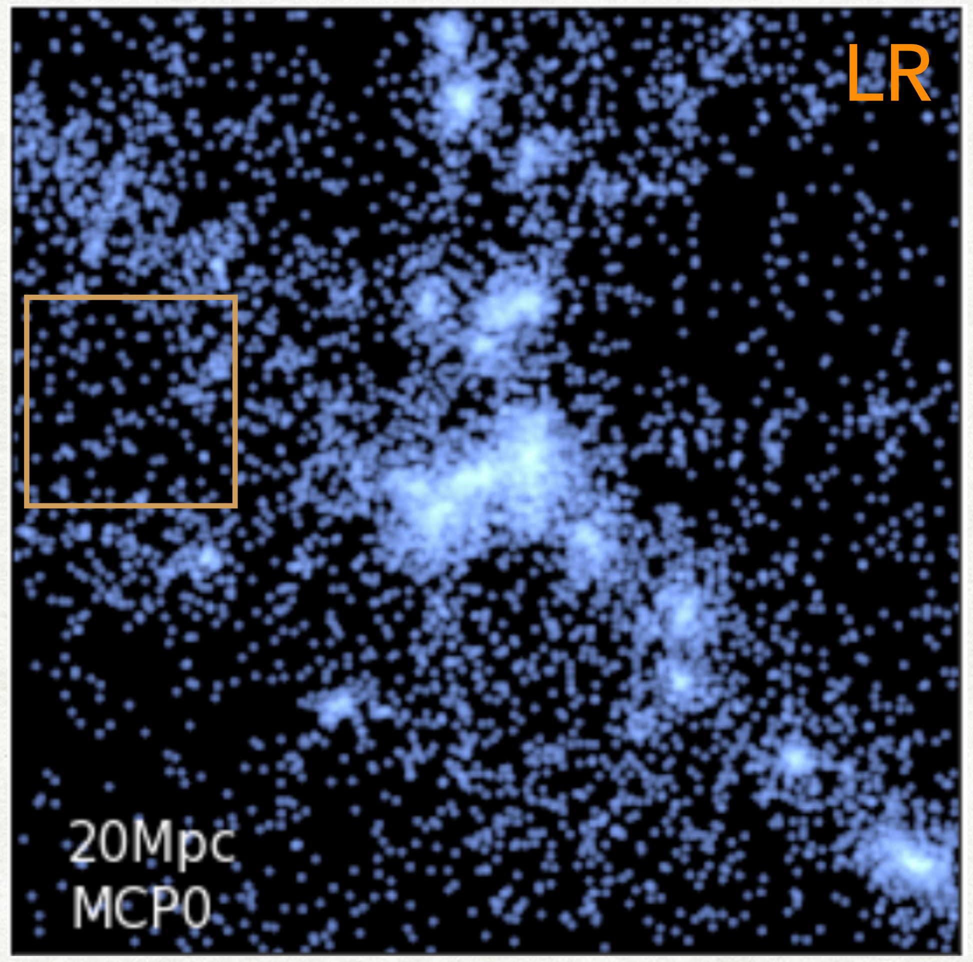
Summary

- SR model: generate the full 6D phase space N-body simulation output with 512 higher mass resolution
- The generated SR fields give statistically good agreement with the authentic HR fields
- Show capability to apply the SR model to large cosmic volume and generate mock catalogs

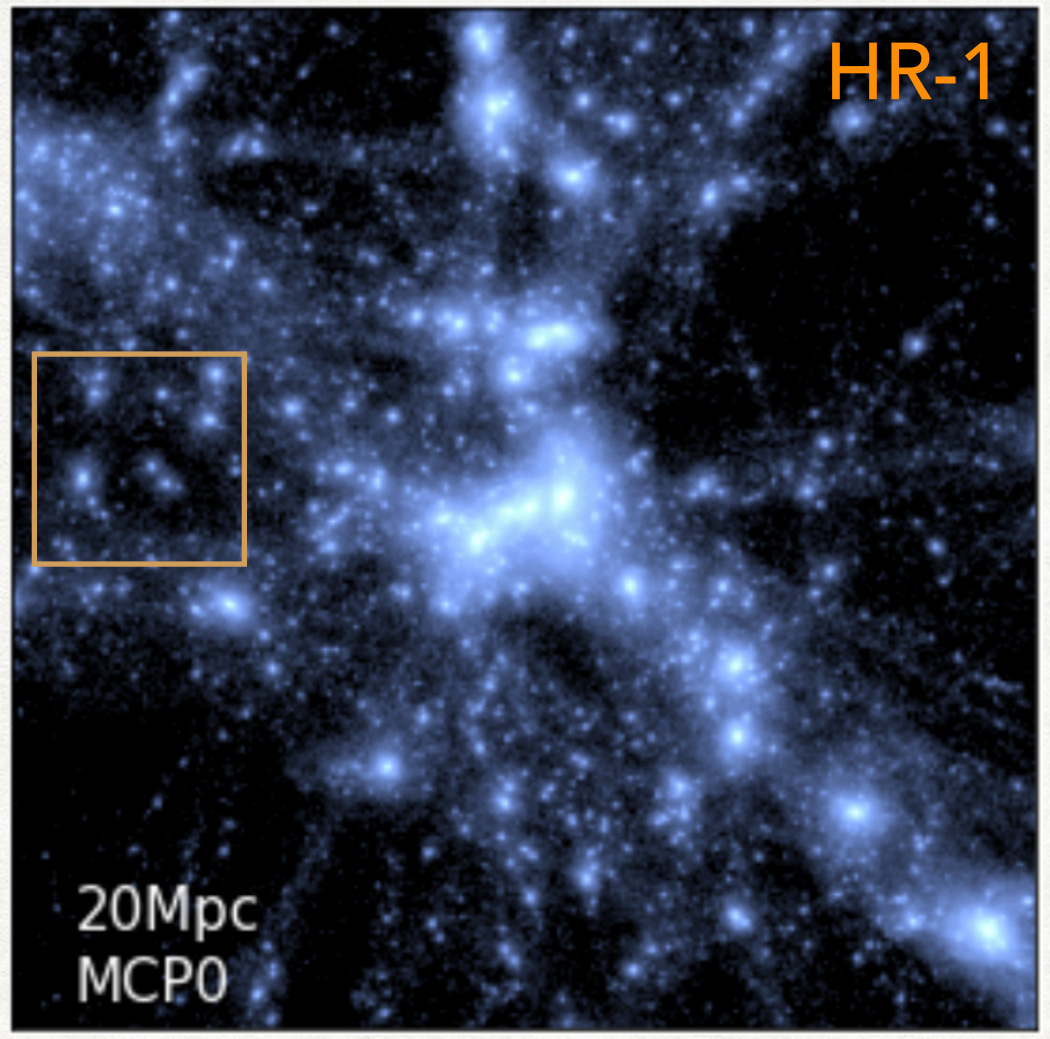
Challenges and future directions

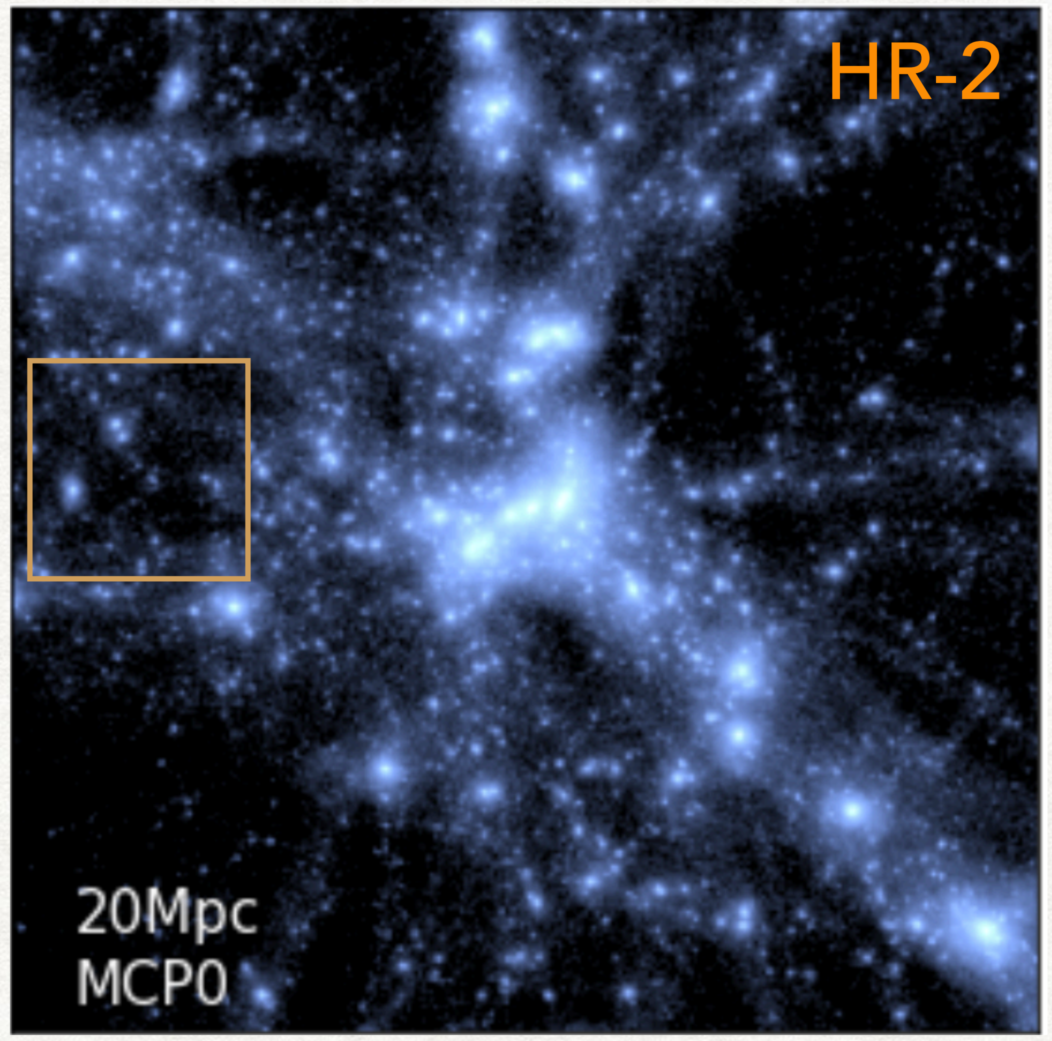
- Improve subhalo statistics
- Accommodate for different cosmology and include the redshift dependency
- From dark matter only to hydrodynamic simulation

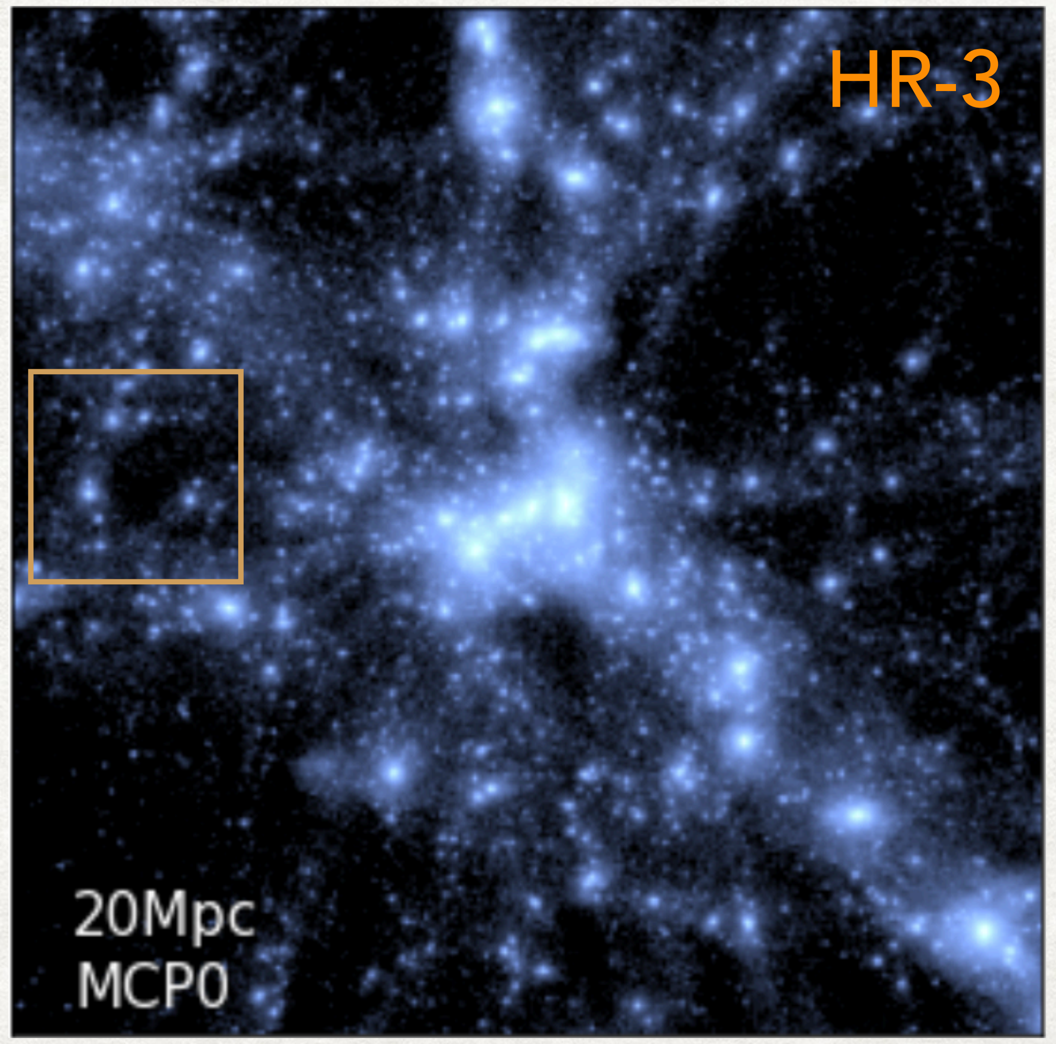
LR —> HR: One to many task



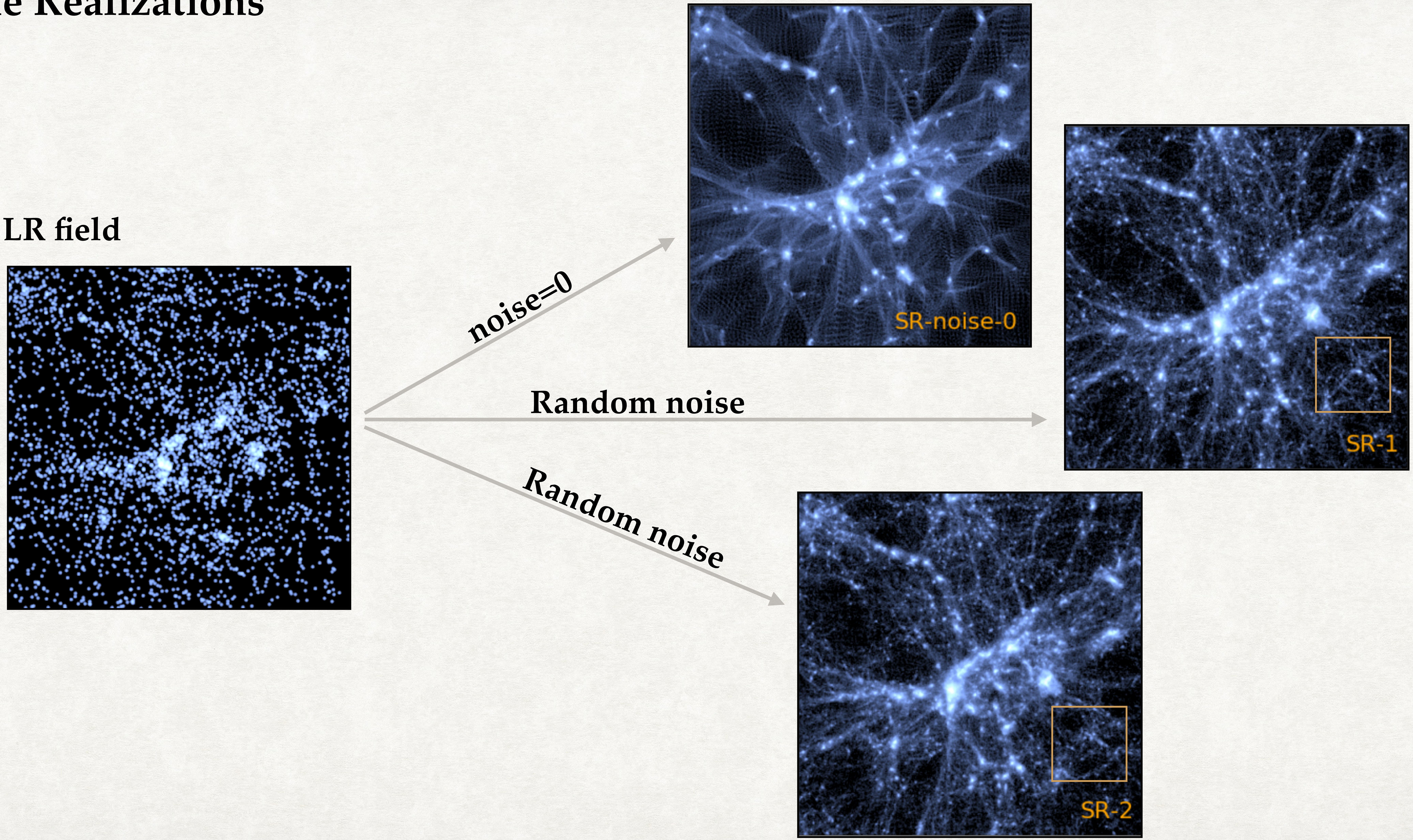
LR —> HR





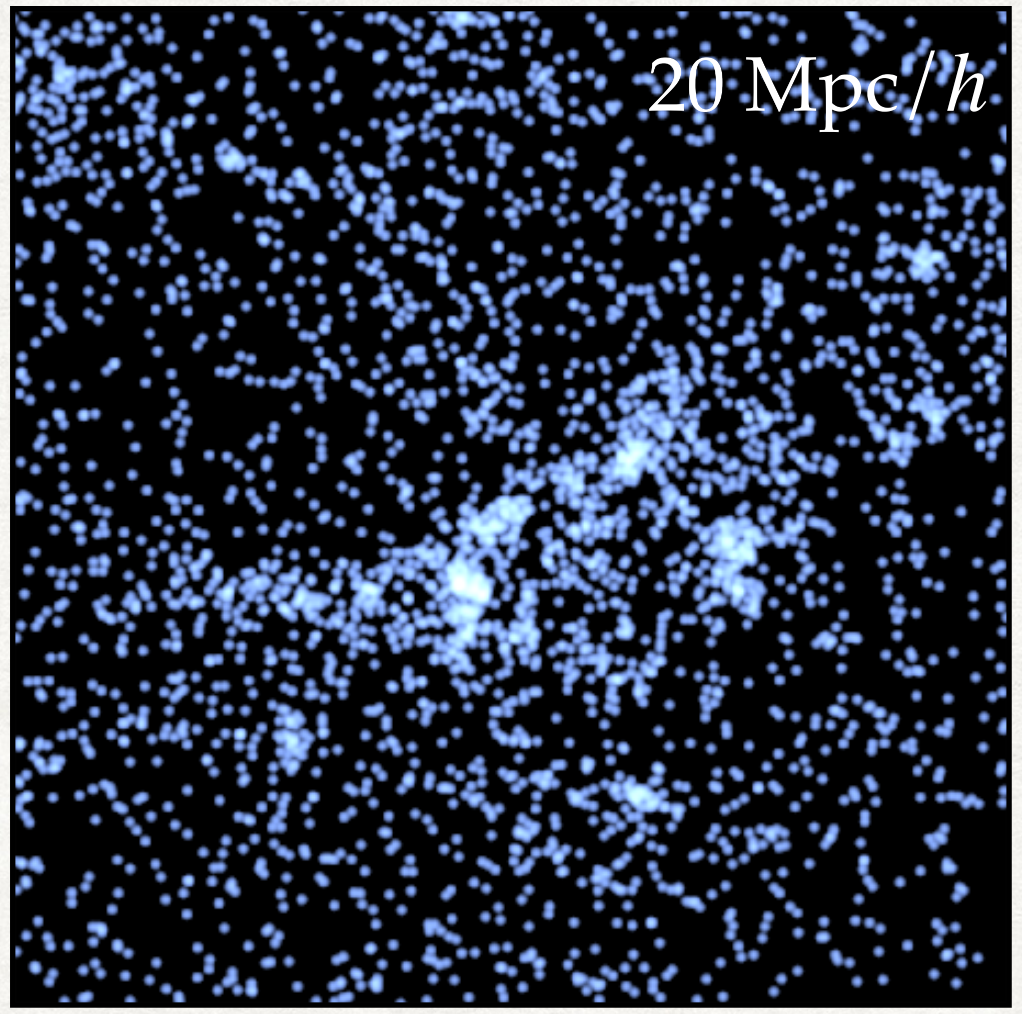


Multiple Realizations



Multiple Realizations

LR field



— Guess which one is HR ?