DarkAI: Reconstructing the large-scale density field of dark matter using AI

史峰

王梓同，杨小虎，李清洋

2023年4月22日 苏州

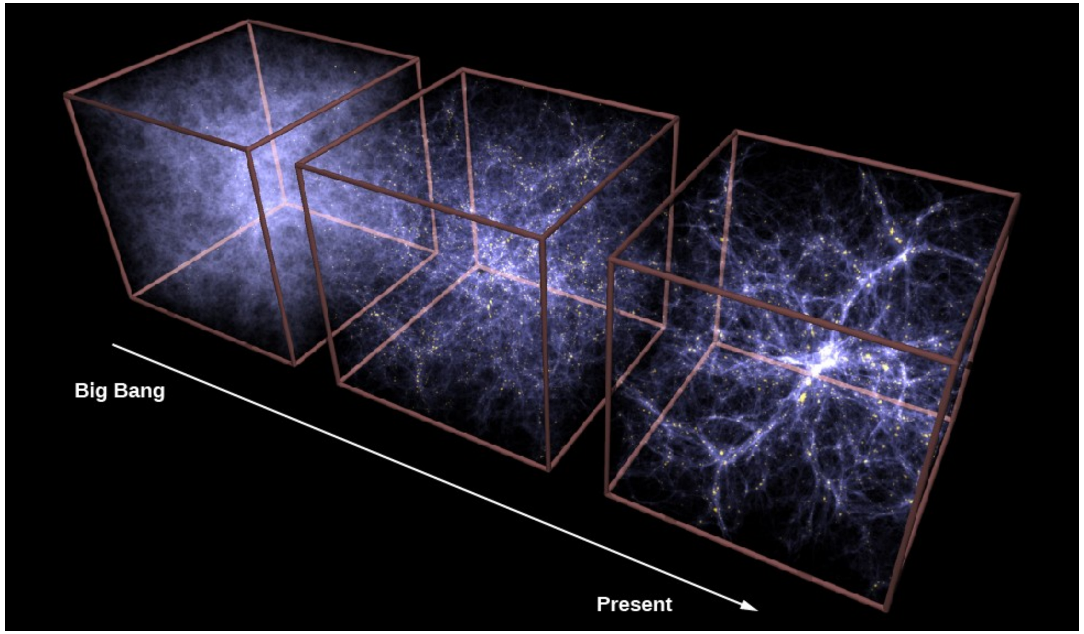
西安电子科技大学
XIDIAN UNIVERSITY
空间科学与技术学院
School of Aerospace Science And Technology
上海交通大学天文系
Astronomy at Shanghai Jiao Tong University

Reconstructing the mass distribution of the Universe

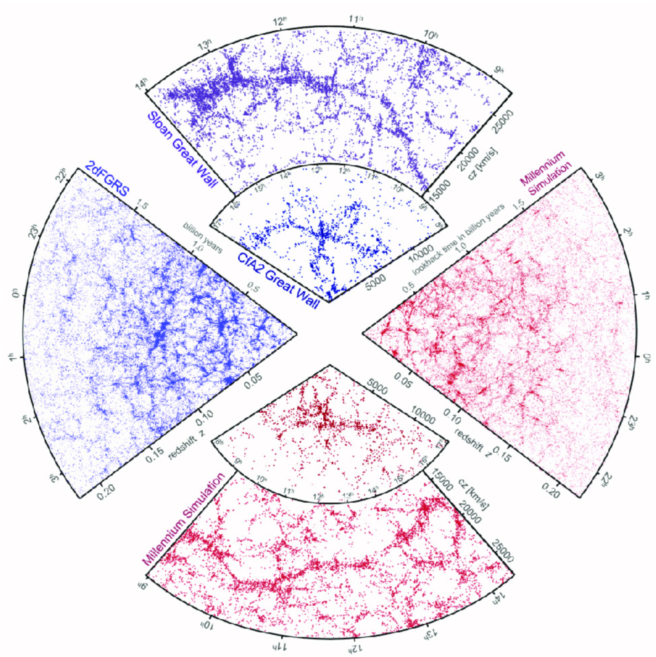
Well-understood

Initial conditions — Dark matter distribution — Dark matter halo — Galaxy distribution

Nbody simulations



Mock galaxy survey



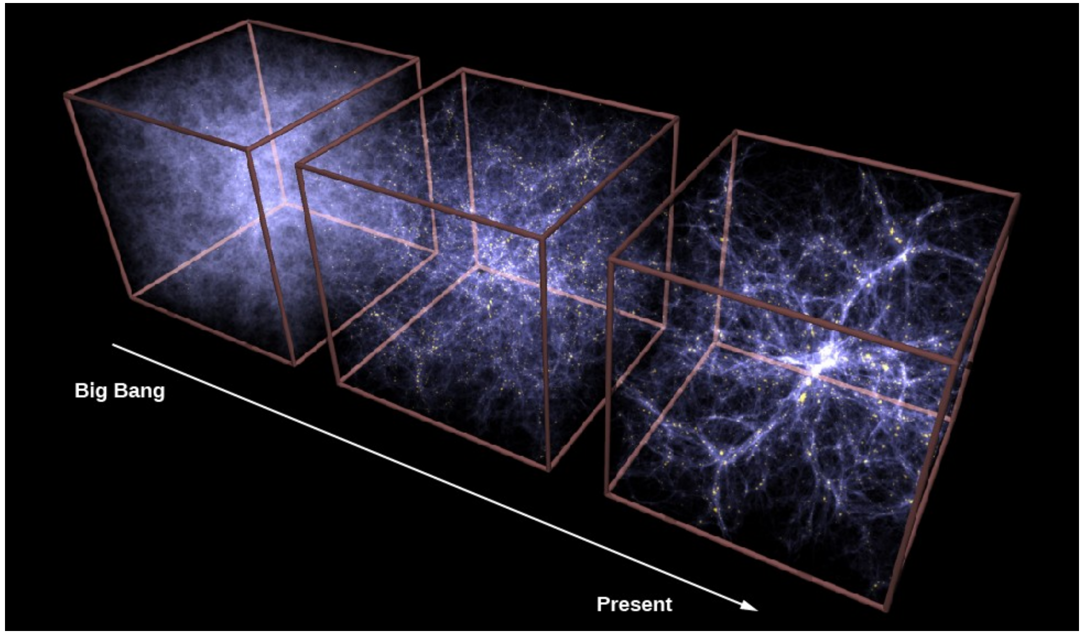
Reconstructing the mass distribution of the Universe

Well-understood

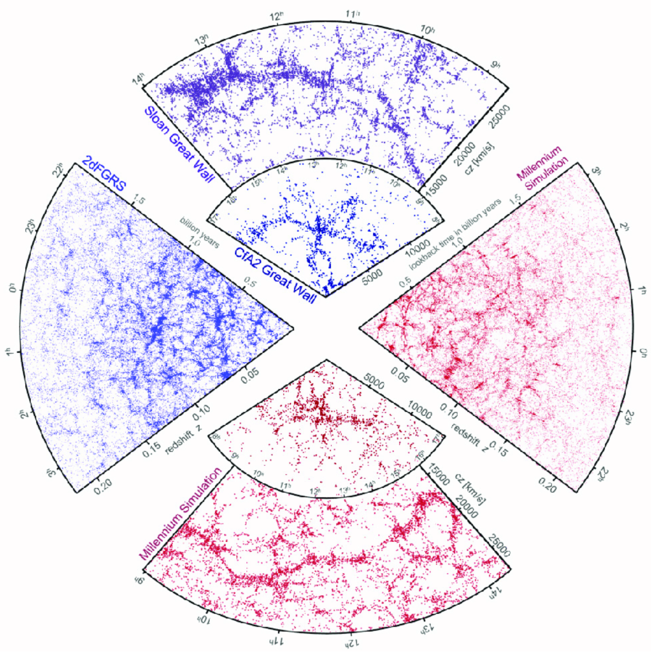
Initial conditions | Dark matter distribution | Dark matter halo | Galaxy distribution

Opportunity to reconstruct the underlying cosmic density field

Nbody simulations

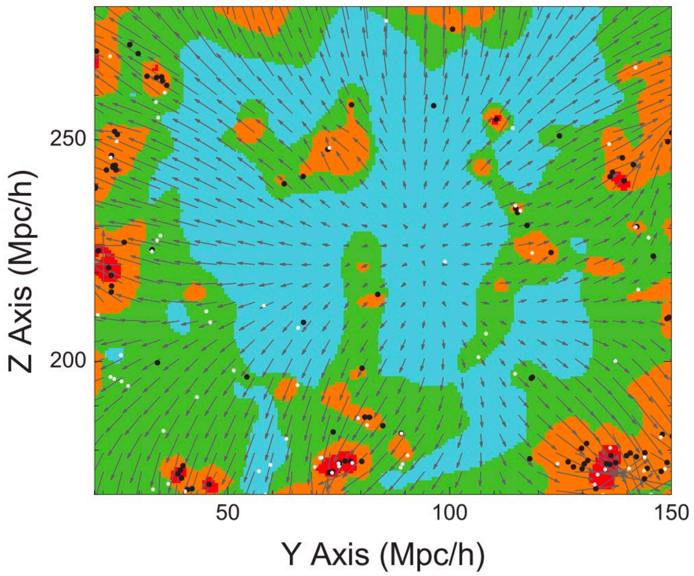


Mock galaxy survey



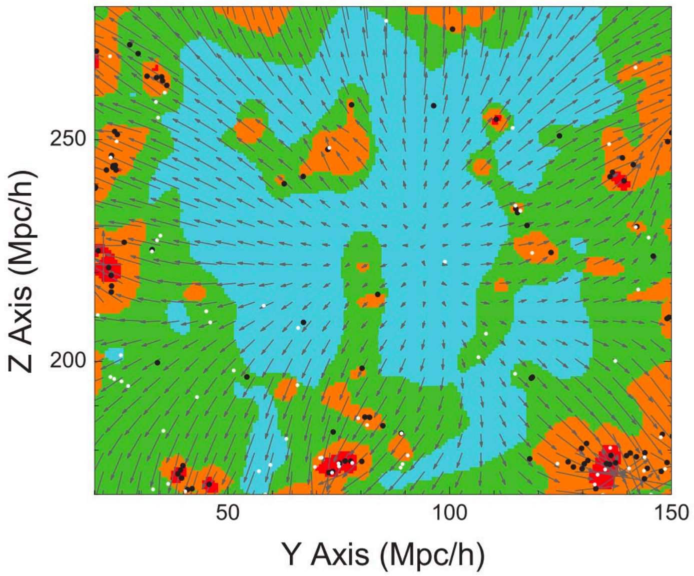
Reconstructing the mass distribution of the Universe can provide us

- Velocity & Tidal field (Wang et al. 2012)

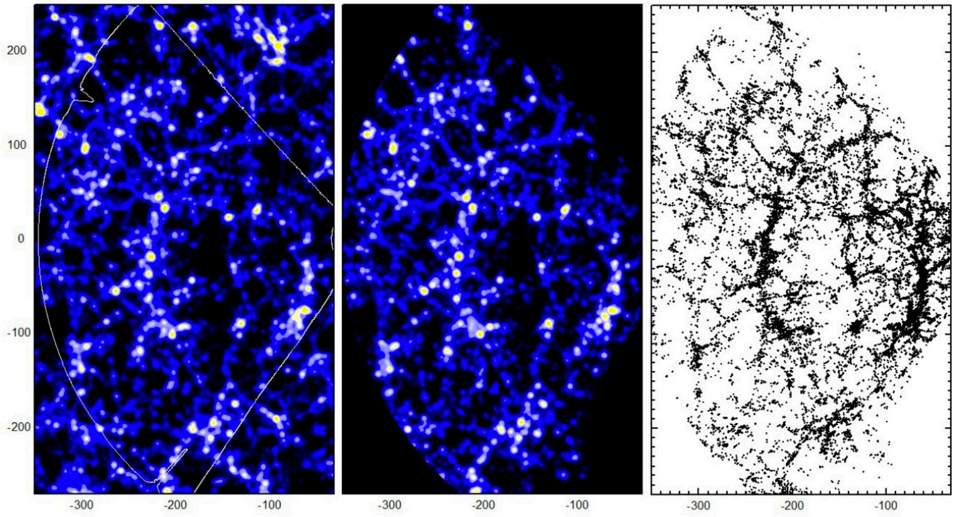


Reconstructing the mass distribution of the Universe can provide us

- Velocity & Tidal field (Wang et al. 2012)

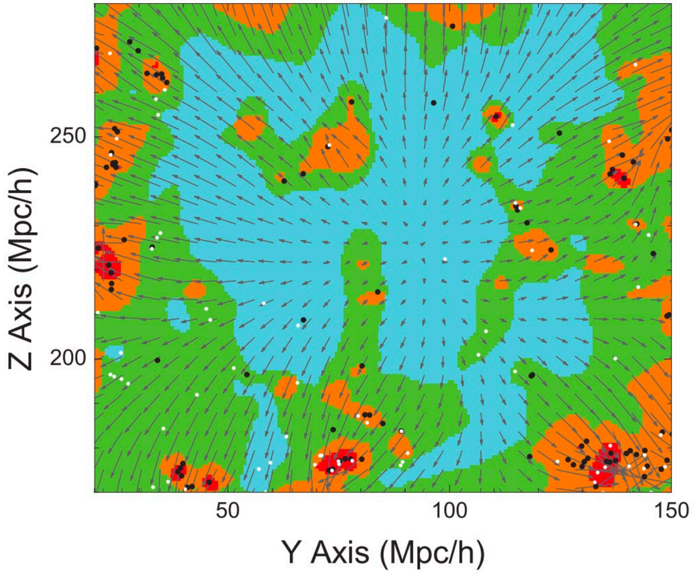


- Initial density field (ELUCID, Wang et al 2016)

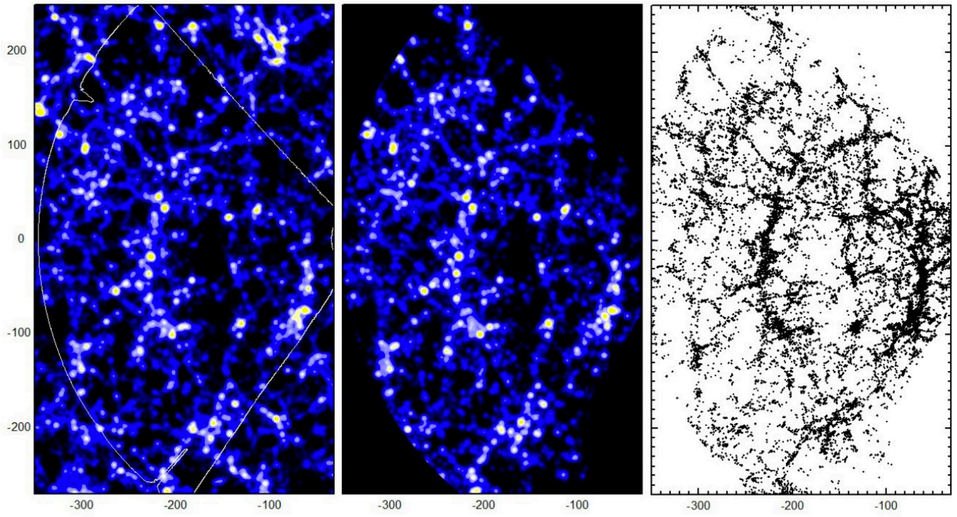


Reconstructing the mass distribution of the Universe can provide us

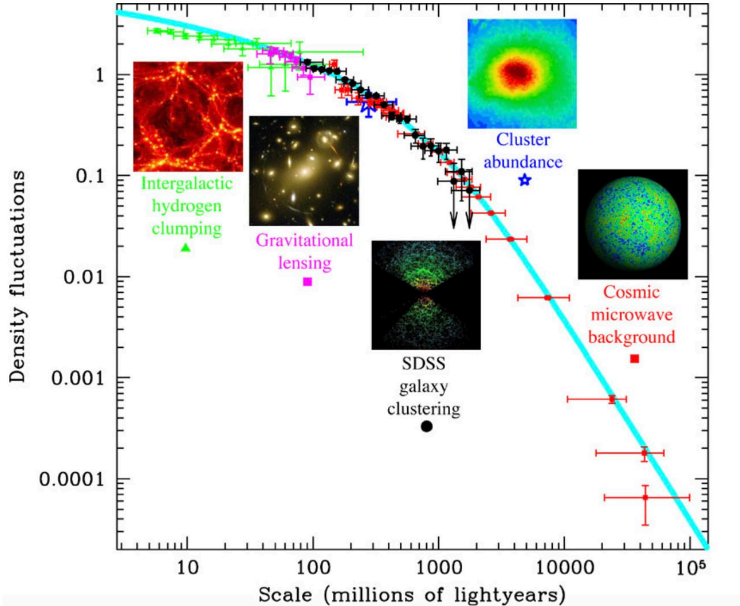
- Velocity & Tidal field (Wang et al. 2012)



- Initial density field (ELUCID, Wang et al 2016)

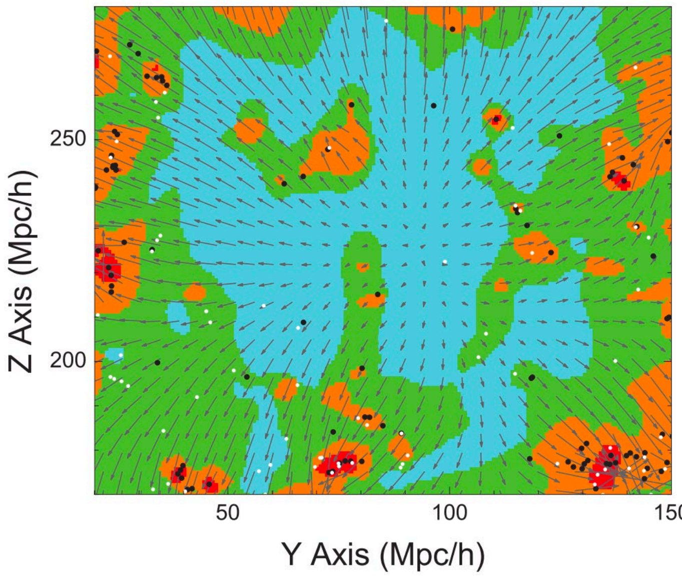


- Real-space dark matter power spectrum (Tegmark et al. 2004)

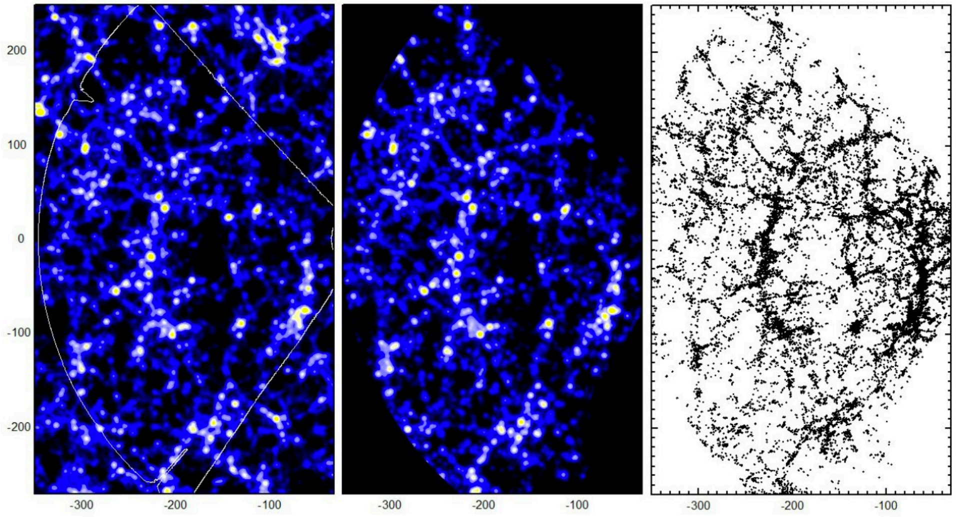


Reconstructing the mass distribution of the Universe can provide us

- Velocity & Tidal field (Wang et al. 2012)

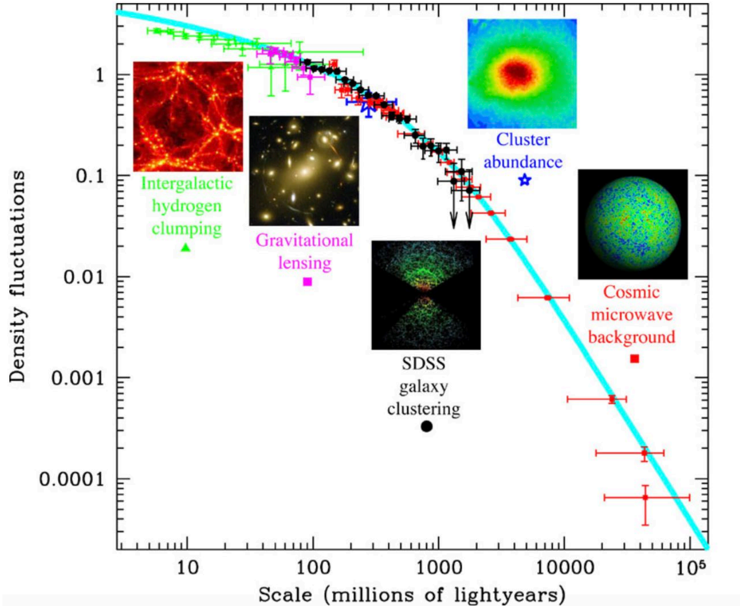


- Initial density field (ELUCID, Wang et al 2016)



Galaxy formation

- Real-space dark matter power spectrum (Tegmark et al. 2004)



Cosmology

The challenge in reconstructing the density field

1) Galaxies bias:

- Biased tracers of the underlying mass distribution
- Exact form of the bias is complicated.
- Linear bias form used to reconstructing velocity field
- Linear bias: only valid for small density fluctuations

  motivated more by simplicity than by physical principles

$$\boldsymbol{v}(\boldsymbol{k}) = H\, a\, f(\Omega)\, \frac{i\boldsymbol{k}}{k^2}\, \frac{\delta_{\rm h}(\boldsymbol{k})}{b_{\rm hm}}.$$

$$\delta_h(\boldsymbol{x}) = b_1\delta(\boldsymbol{x}) + \frac{1}{2}b_2[\delta(\boldsymbol{x})^2 - \sigma_2] + \frac{1}{2}b_{s2}[s(\boldsymbol{x})^2 - \langle s^2\rangle] + \text{higher order terms}\,.$$

The challenge in reconstructing the density field

1) Galaxies bias:

- Biased tracers of the underlying mass distribution
- Exact form of the bias is complicated.
- Linear bias form used to reconstructing velocity field
- Linear bias: only valid for small density fluctuations

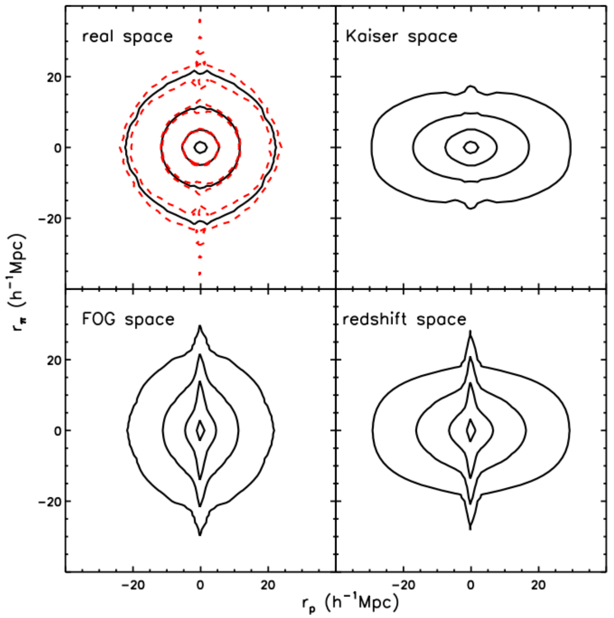
  motivated more by simplicity than by physical principles

$$\boldsymbol{v}(\boldsymbol{k}) = H\, a\, f(\Omega)\, \frac{i\boldsymbol{k}}{k^2}\, \frac{\delta_{\rm h}(\boldsymbol{k})}{b_{\rm hm}}.$$

$$\delta_h(\boldsymbol{x}) = b_1\delta(\boldsymbol{x}) + \frac{1}{2}b_2[\delta(\boldsymbol{x})^2 - \sigma_2] + \frac{1}{2}b_{s2}[s(\boldsymbol{x})^2 - \langle s^2\rangle] + \text{higher order terms}.$$

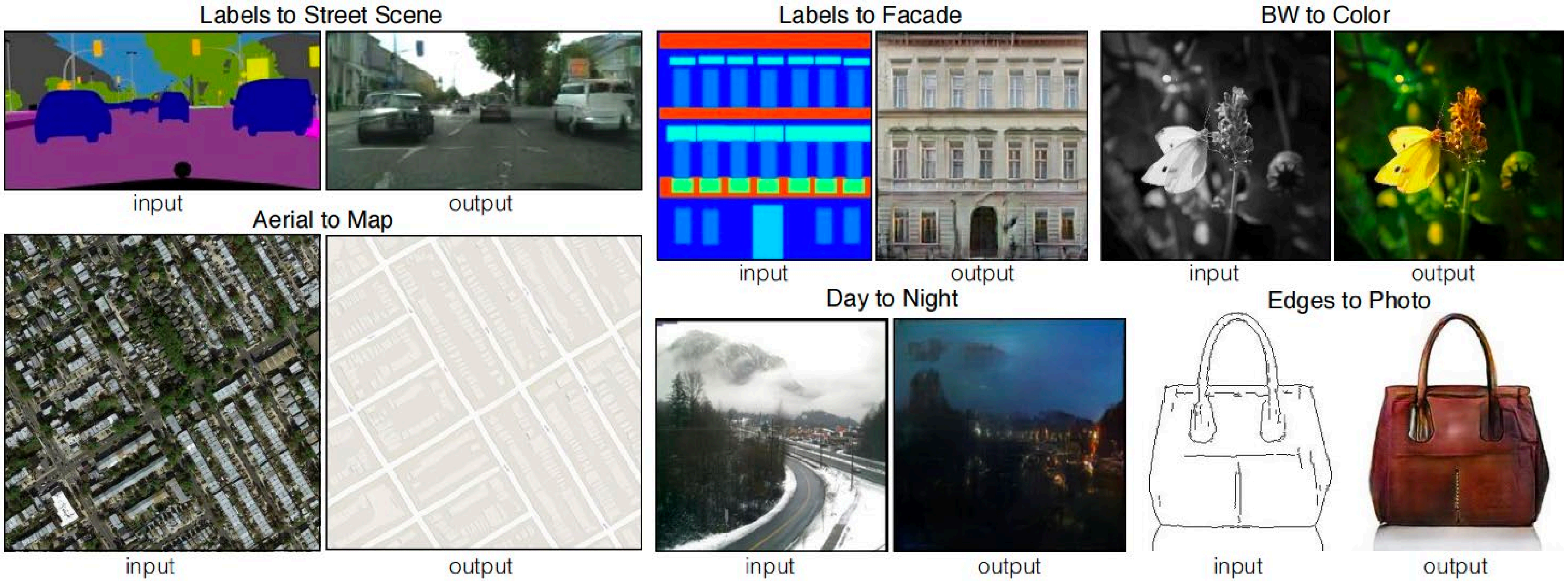
2) Redshift space distortions

- Kaiser effect and FOG effect
- Causing modeling the bias parameters more complicated
- Iteration to make the RSD correction
- Linear theory limited in the high-density regions



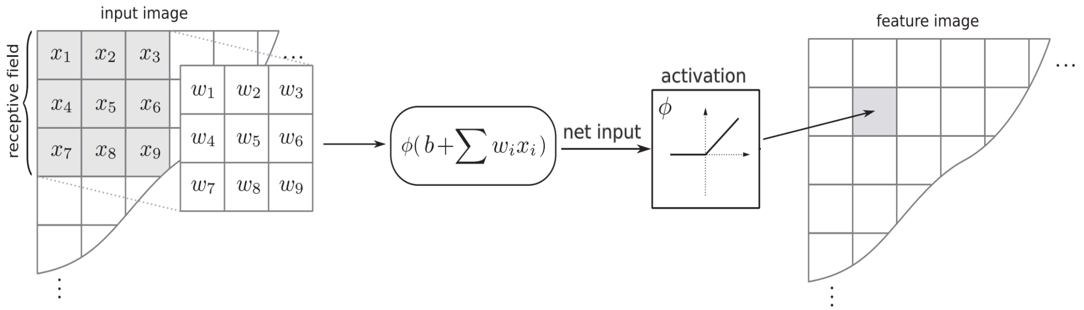
Deeping learning method: UNet model

- Provides a general model for image-to-image translation
- Apply to a wide variety of image generation tasks, including translating photography from day to night and product sketches to photographs

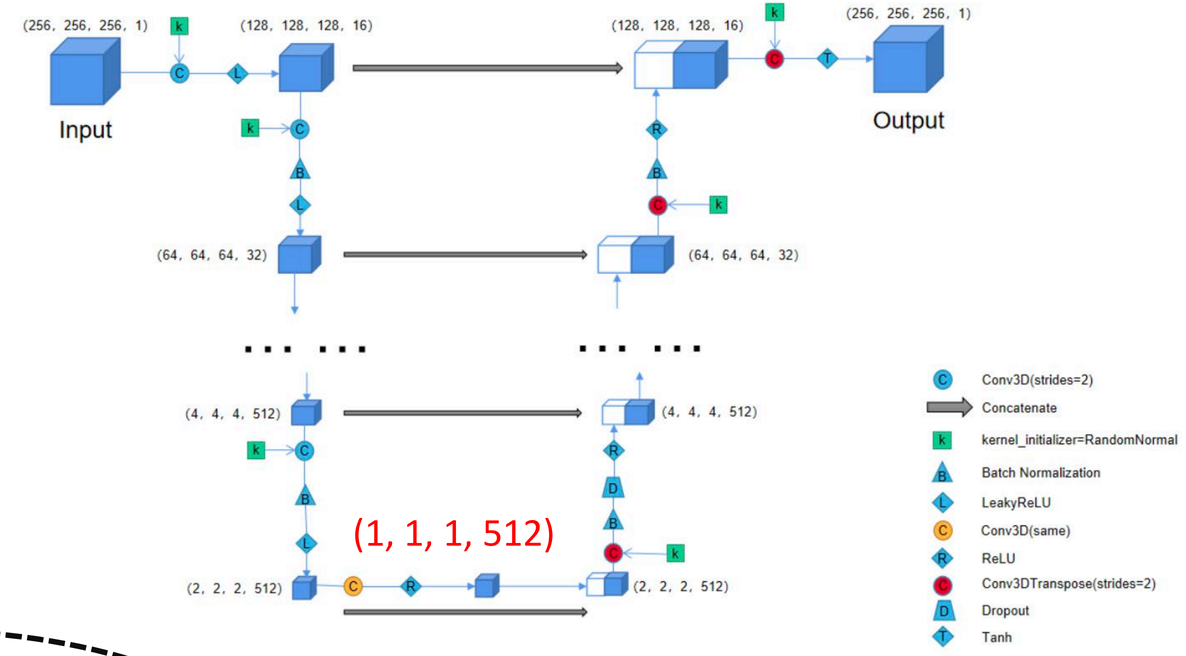


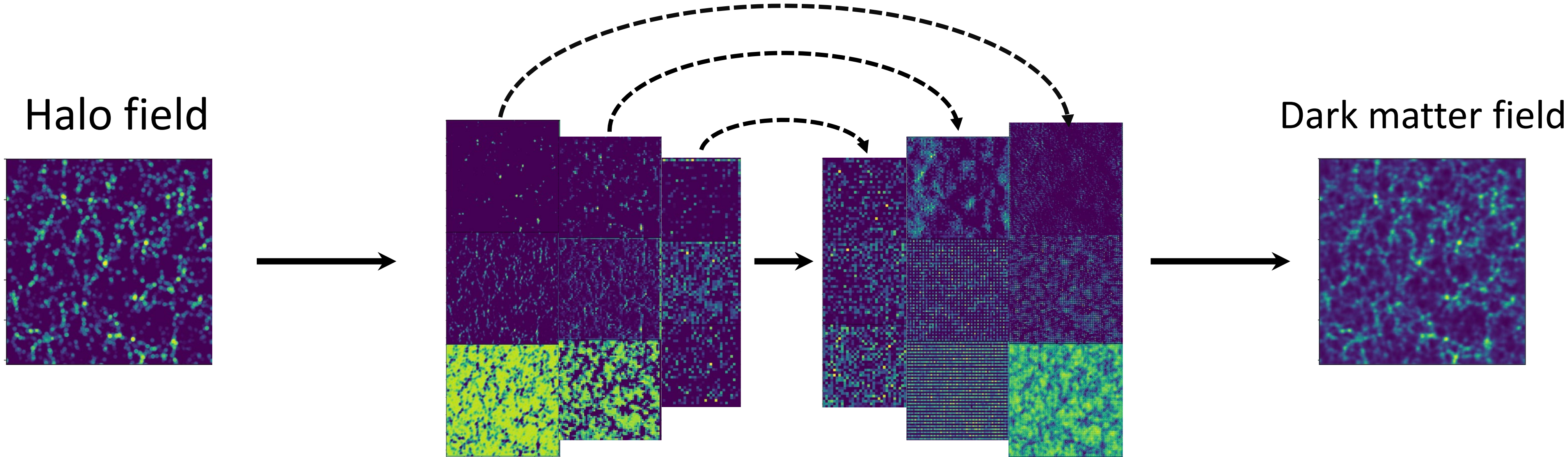
Deeping learning method: UNet model

Convolutional neural network (CNN) :



Encoder-decoder with skip connections:





Generating training and testing data

1) Evolute the particles:

- cola_halo code: COmoving Lagrangian Acceleration (COLA) fast simulation
- Generate 30 simulations: 15 training, 5 validation, and 10 testing samples
- Planck2018 cosmology, $\Omega_m = 0.3111$, $\Omega_\Lambda = 0.6889$, $h = 0.6766$, $\sigma_8 = 0.812$.
- 512^3 particles, 500Mpc/h
- Add RSD along the z-axis for halos. Keep real space for dark matter.

Generating training and testing data

1) Evolute the particles:

- cola_halo code: COmoving Lagrangian Acceleration (COLA) fast simulation
- Generate 30 simulations: 15 training, 5 validation, and 10 testing samples
- Planck2018 cosmology, $\Omega_m = 0.3111$, $\Omega_\Lambda = 0.6889$, $h = 0.6766$, $\sigma_8 = 0.812$.
- 512^3 particles, 500Mpc/h
- Add RSD along the z-axis for halos. Keep real space for dark matter.

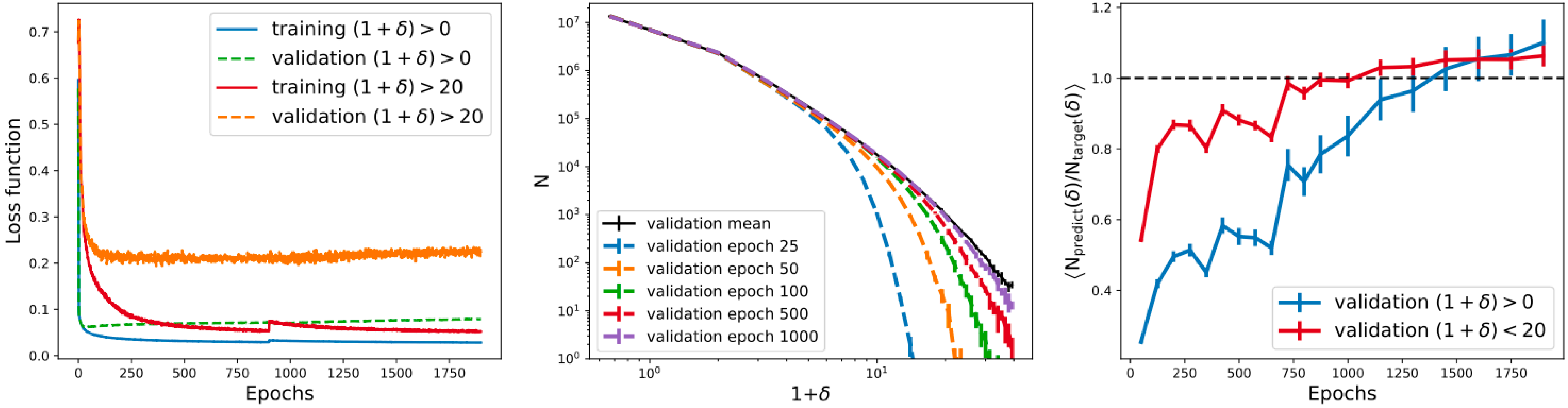
2) Construct the density fields:

- CIC scheme, 256^3 voxels,Top-hat smoothing with $R_s = 5\ h^{-1}$Mpc
- Halo mass weighting
- Rescaled the overdensity values to lie in the interval [−1, 1]

$$s(x) = 2x/(x + a) - 1, \quad a=5$$

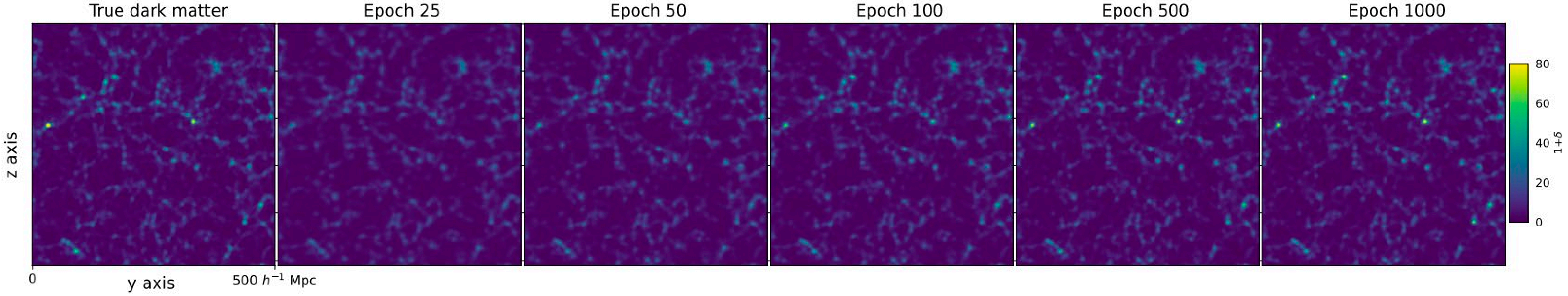
Training process

- Run 1000 epochs: check both training and validating samples
- The differences between the prediction and the target keep less until epoch around epoch 600
- The performance would not be improved after epoch 600
- Save best model during epoch 700-1000 based on validate samples



Training process

- Run 1000 epochs: check both training and validating samples
- The differences between the prediction and the target keep less until epoch around epoch 600
- The performance would not be improved after epoch 600
- Save best model during epoch 700-1000 based on validate samples



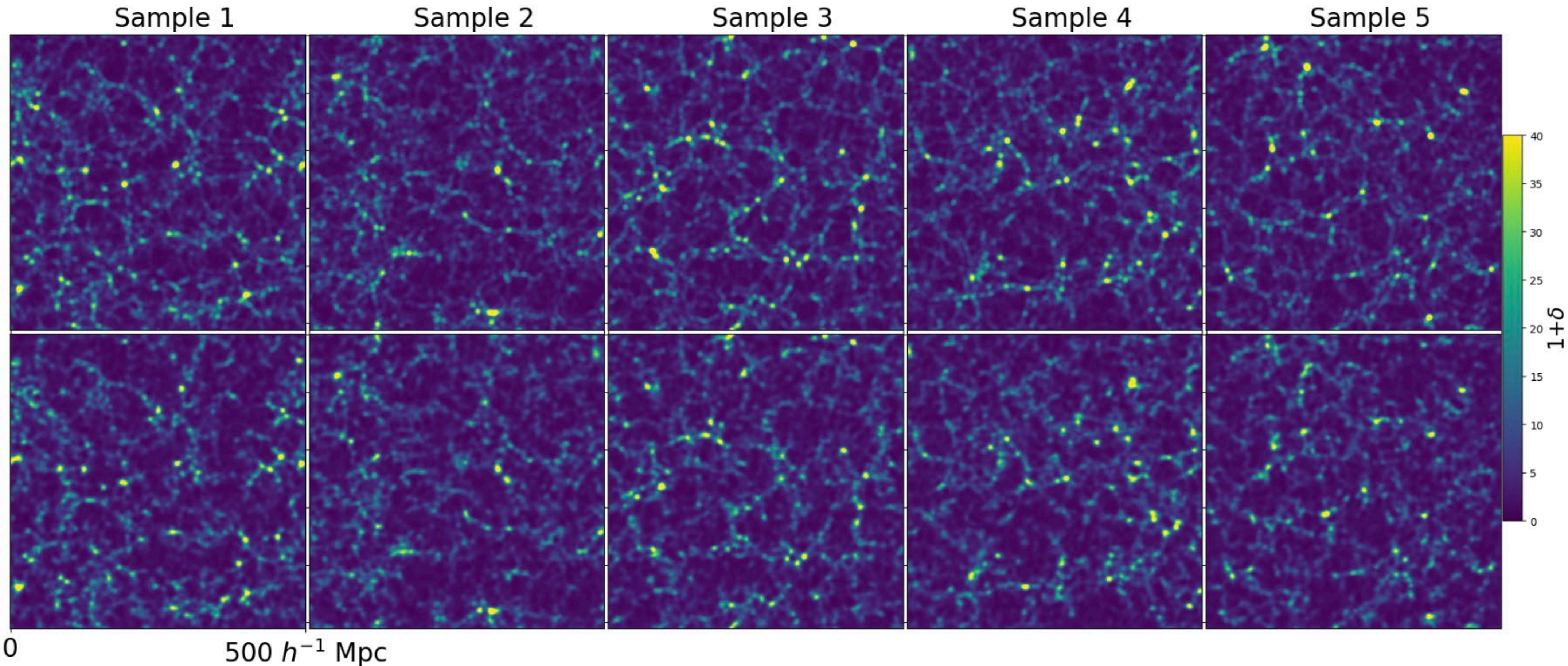
Testing: COLA samples

1) Comparisons of the projected density

- 5 samples randomly selected from the 10 COLA test samples in a slice of 500 × 500 × 9.76 h^{-1}Mpc
- The reconstruction exhibit recognizable, large-scale structures including clusters, filaments, and voids
- The reconstruction is generally very successful over the different scales
- Differ slightly from the target at regions around clusters.

Truths

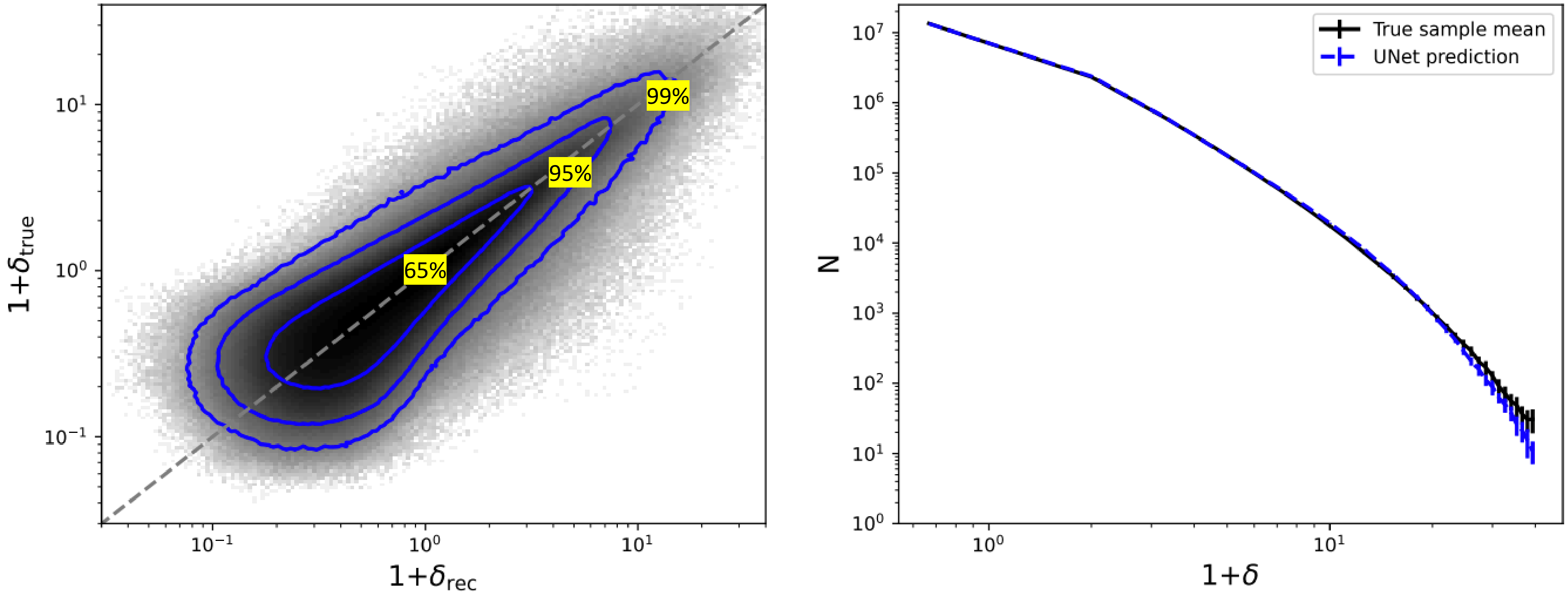
UNet predictions



Testing: COLA samples

2) Density-density relation (left panel) and histogram distribution (right panel)

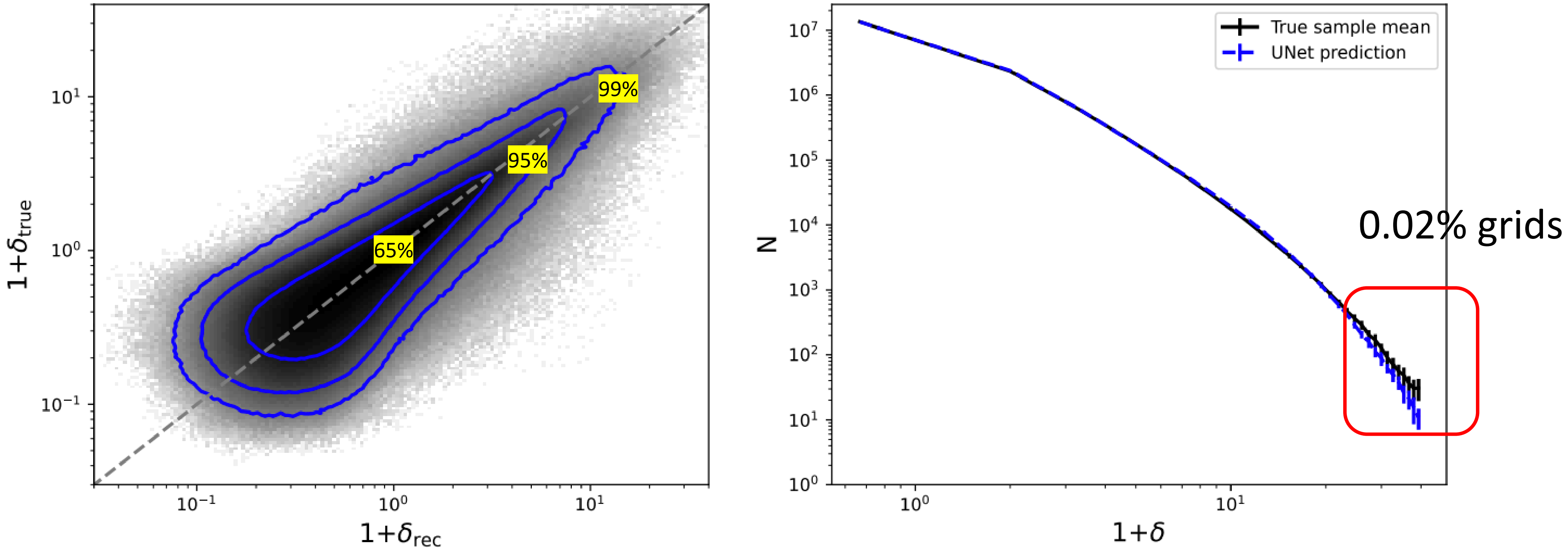
- No significant bias between δ_{rec} and δ_{true}
- 99.98% grids keeps accuracy $\Delta\delta/\delta < 5\%$
- Reason: small number of massive halos in currect trainng volume



Testing: COLA samples

2) Density-density relation (left panel) and histogram distribution (right panel)

- No significant bias between δ_{rec} and δ_{true}
- 99.98% grids keeps accuracy $\Delta\delta/\delta < 5\%$
- Reason: small number of massive halos in currect trainng volume



Testing: COLA samples

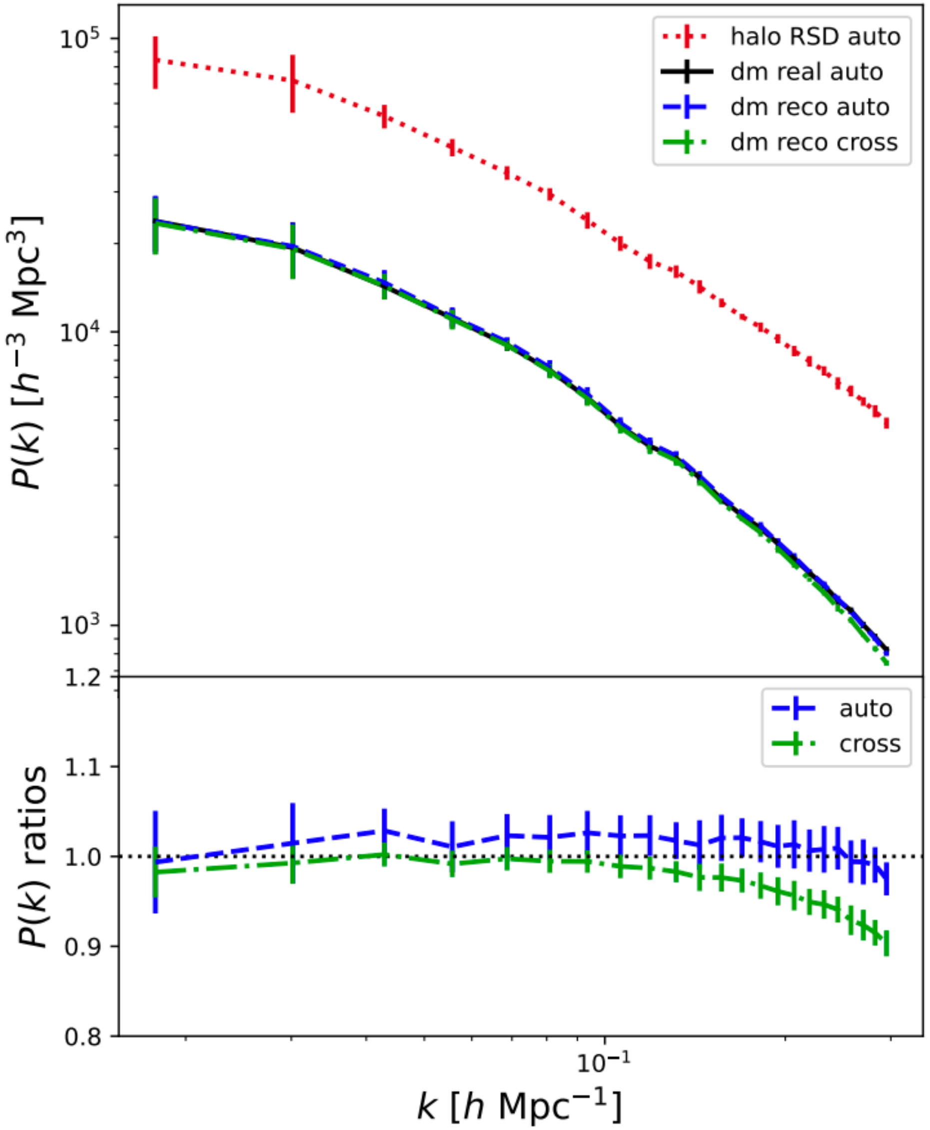
3) Monopole power spectrum

- Cross-correlation:

P(k) ratios is 0.99 $\pm$ 0.01, k < 0.1 h Mpc^{-1}

1% reduction at k = 0.1 h Mpc^{-1}

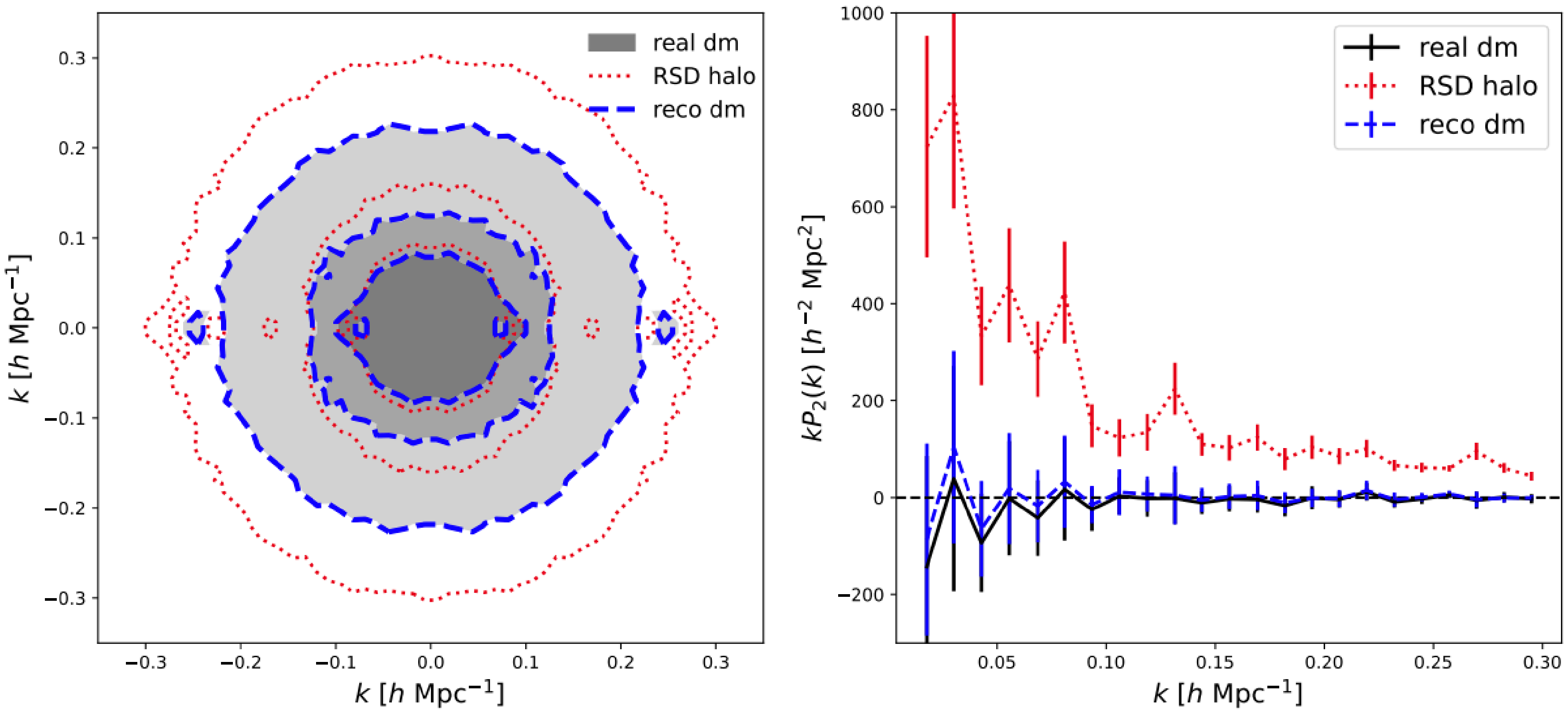
10% reduction at k = 0.3h Mpc^{-1}



Testing: COLA samples

4) 2D power spectrum

- UNet-reconstructed $P\,(k_{\perp}, k_{\parallel})$ is clearly more isotropic and perfectly round.
- Quadrupole $P_2\,(k)$ is very close to zero.
- The correction for the RSDs is overall very successful.

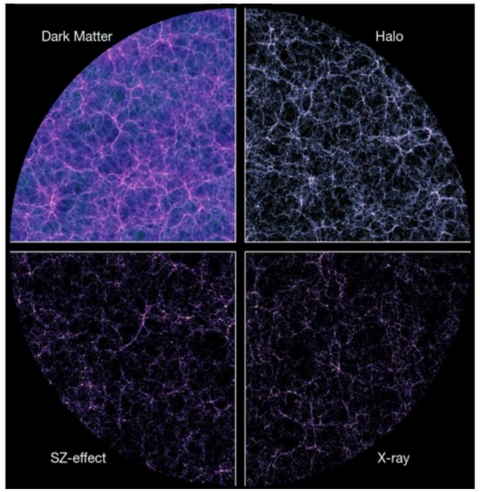


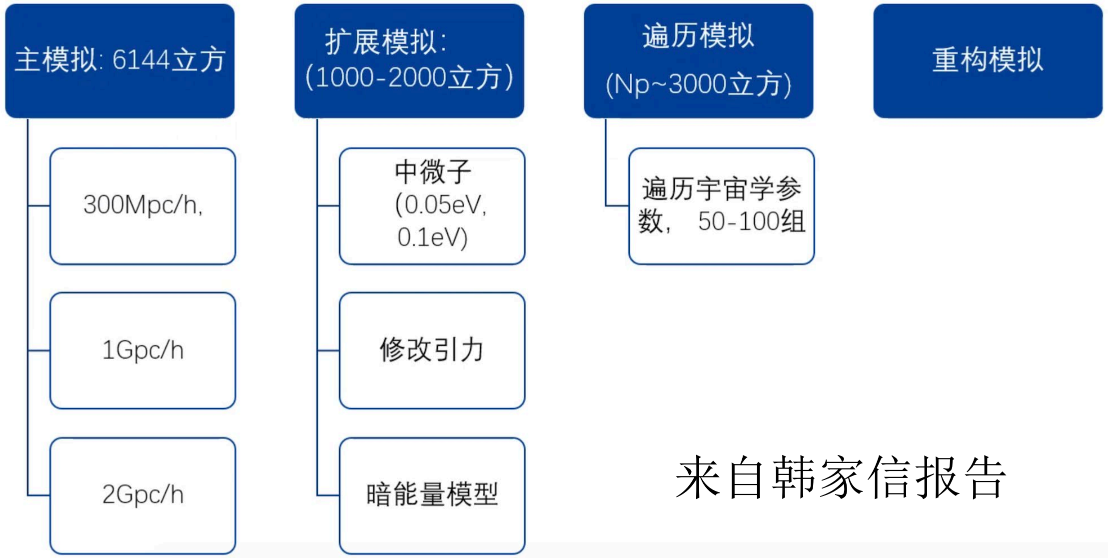
Application to Jiutian

Simulation for CSST

- 标准宇宙学模型 （Planck2018）
- 分层互补，匹配分辨率和尺度需求
- 6144^3 粒子，纯暗物质
- 已完成1Gpc/h
  - L-GADGET3 (李明)
  - 1万核，28天（14天模拟+14天后处理）
  ~700万核时，22TB+内存
  - 6.8TB/snapshot， 共900TB+
  (Millennium: 2160^3, 500Mpc/h, mp=8x10^8, 34万核时， 25TB数据)
  - 完成初步测试和半解析星系建模

边长(Mpc/h)	粒子质量($M_\odot$/h)
300	10^7
1000	3.6×10^8
2000	2.8×10^9





来自韩家信报告

Application to Jiutian

主模拟：九天

盒子大小：1000Mpc/h

粒子数：6144^3

计算：1万核，28天

内存：22TB+

存储：900TB+

Application to Jiutian

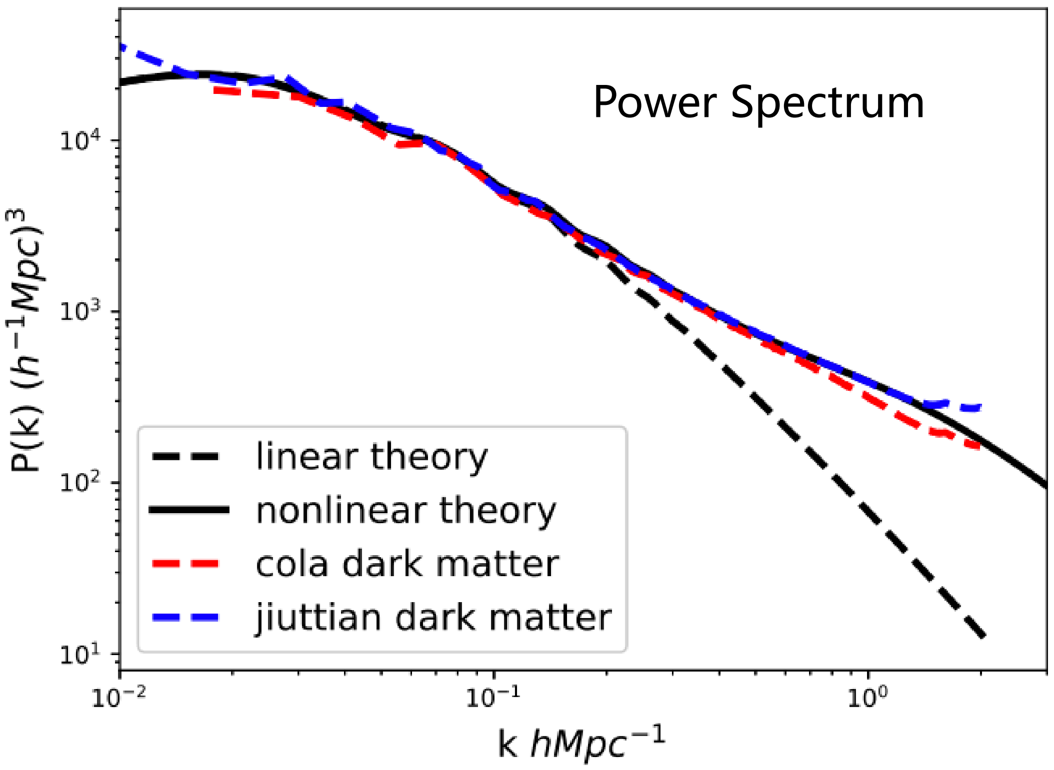
训练数据：COLA
盒子大小：500Mpc/h
粒子数：512^3
计算：28核，0.5小时
内存：<3GB
存储：10GB

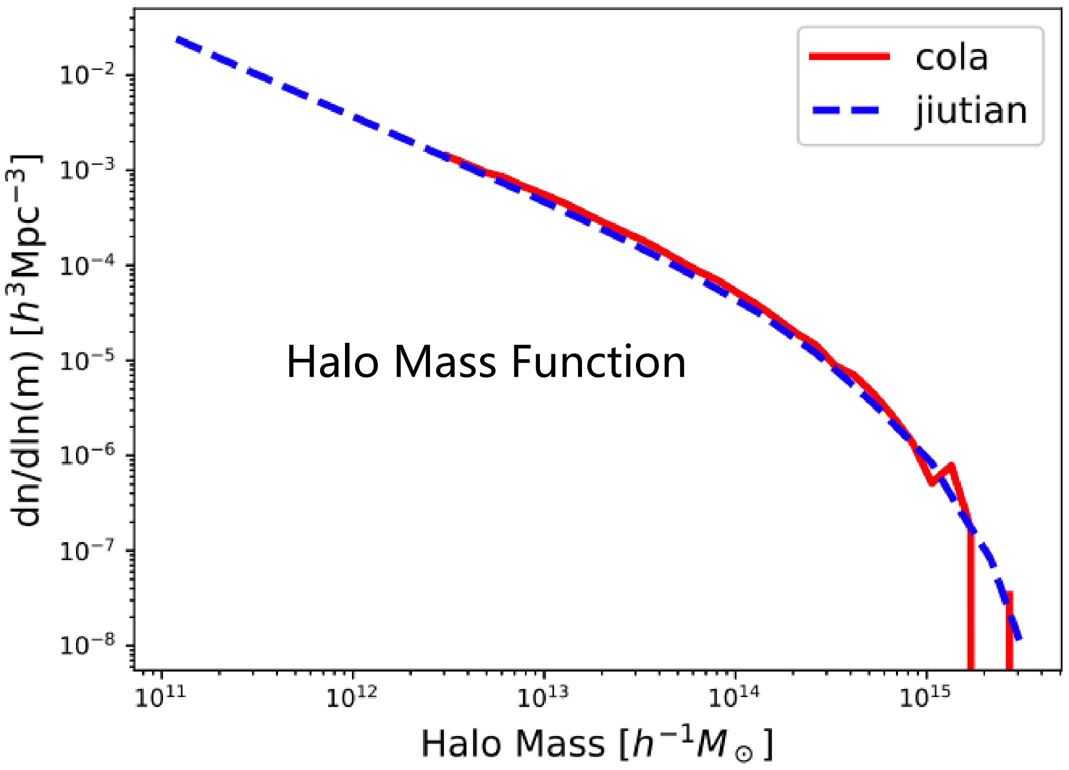
主模拟：九天
盒子大小：1000Mpc/h
粒子数：6144^3
计算：1万核，28天
内存：22TB+
存储：900TB+

Application to Jiutian

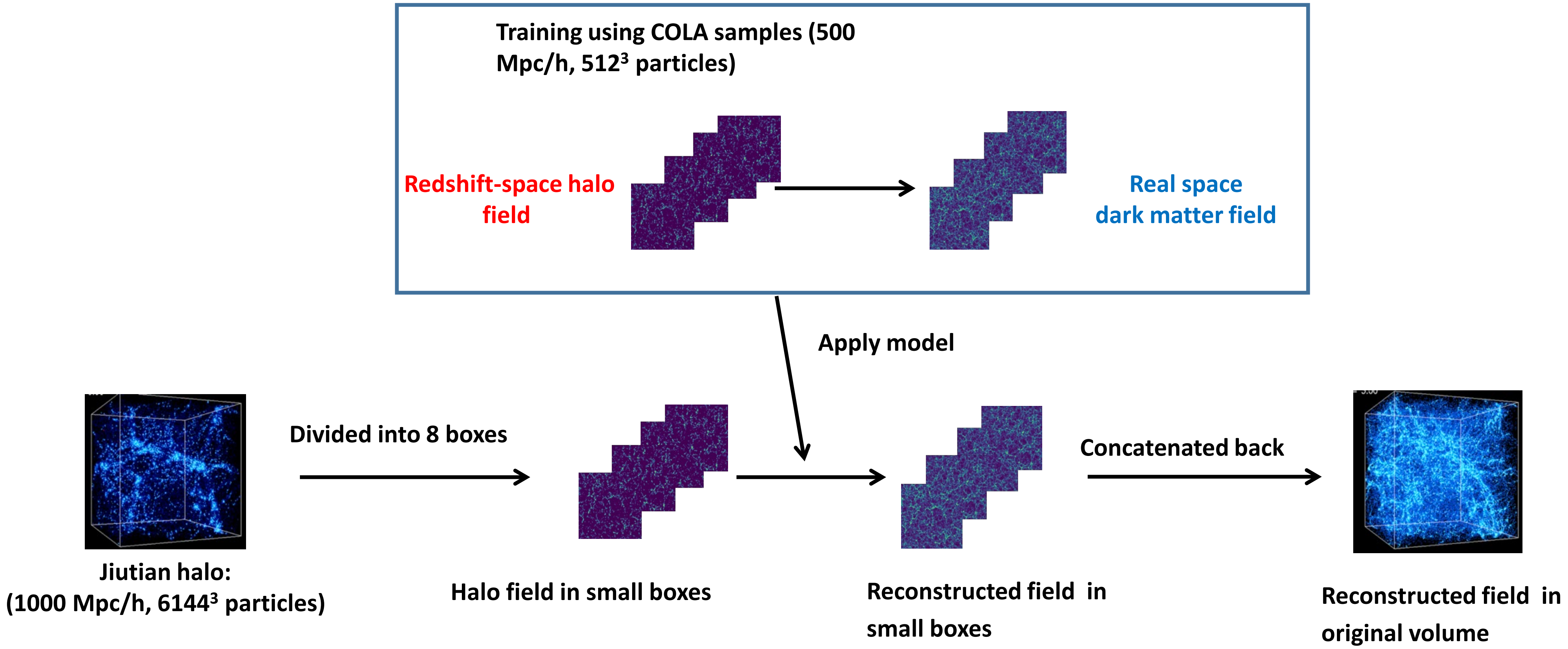
训练数据：COLA
盒子大小：500Mpc/h
粒子数：512^3
计算：28核，0.5小时
内存：<3GB
存储：10GB

主模拟：九天
盒子大小：1000Mpc/h
粒子数：6144^3
计算：1万核，28天
内存：22TB+
存储：900TB+



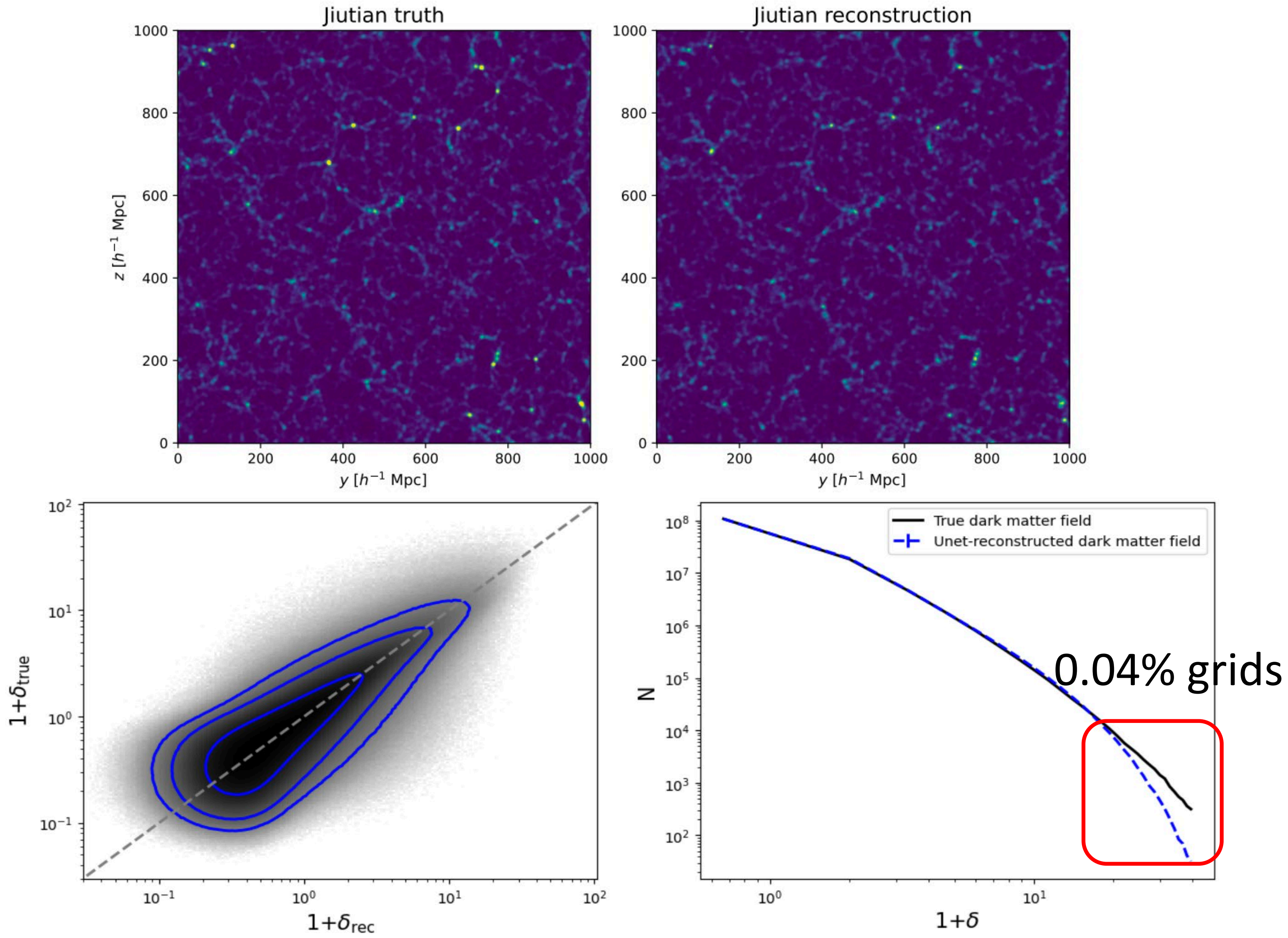


Application to Jiutian



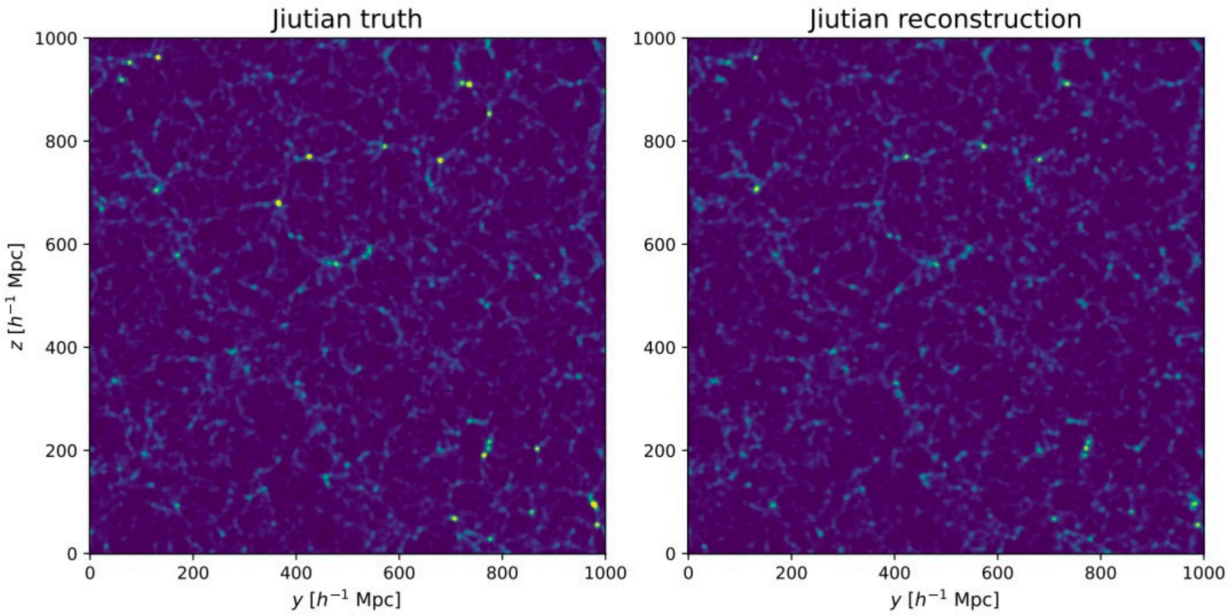
Results: Jiutian simulation

- Reconstructing dark matter density field based on UNet

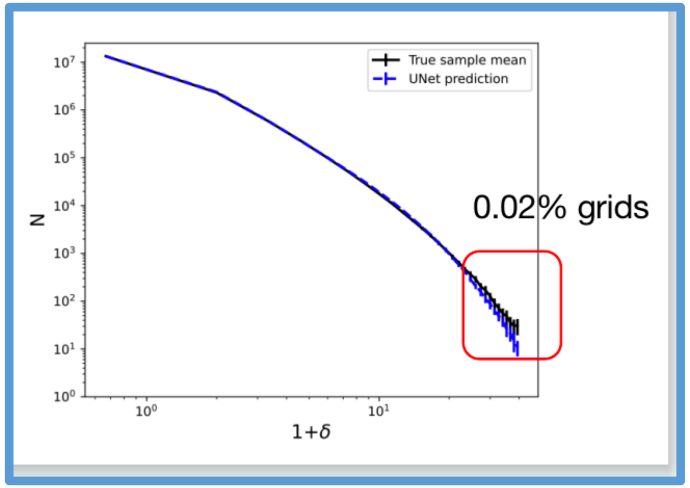


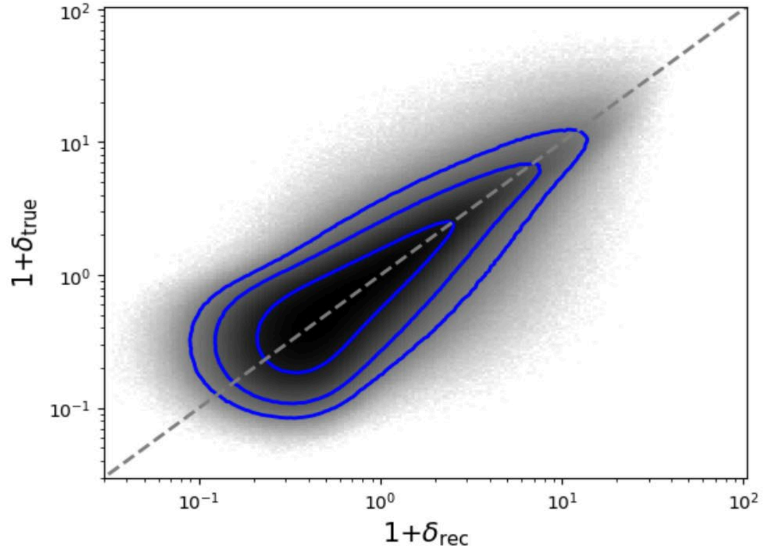
Results: Jiutian simulation

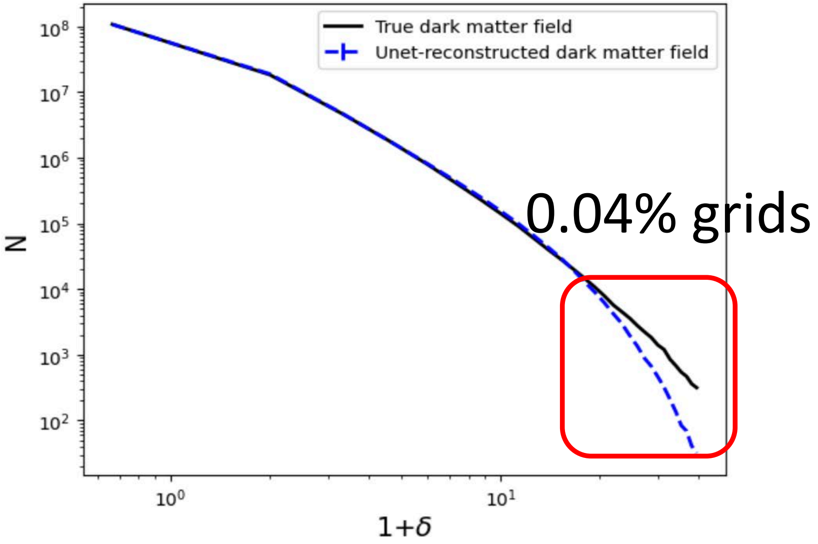
- Reconstructing dark matter density field based on UNet



COLA

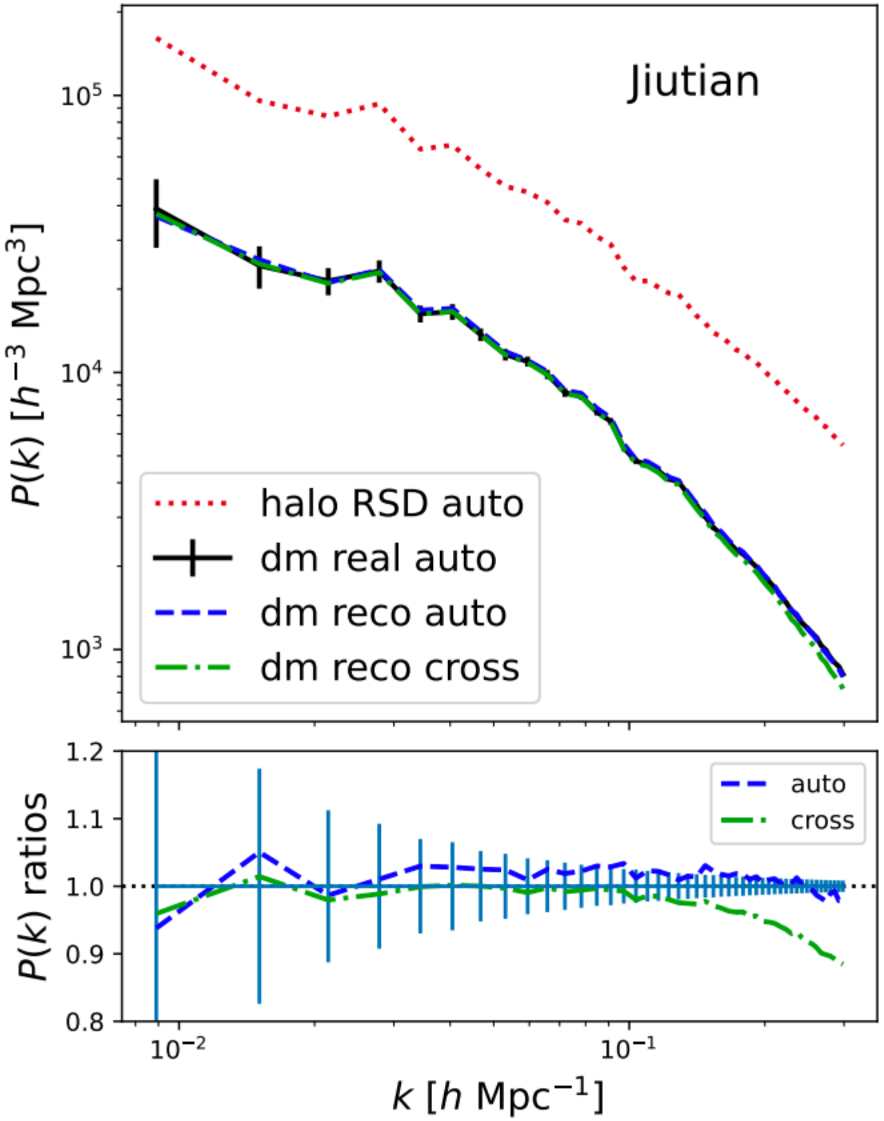


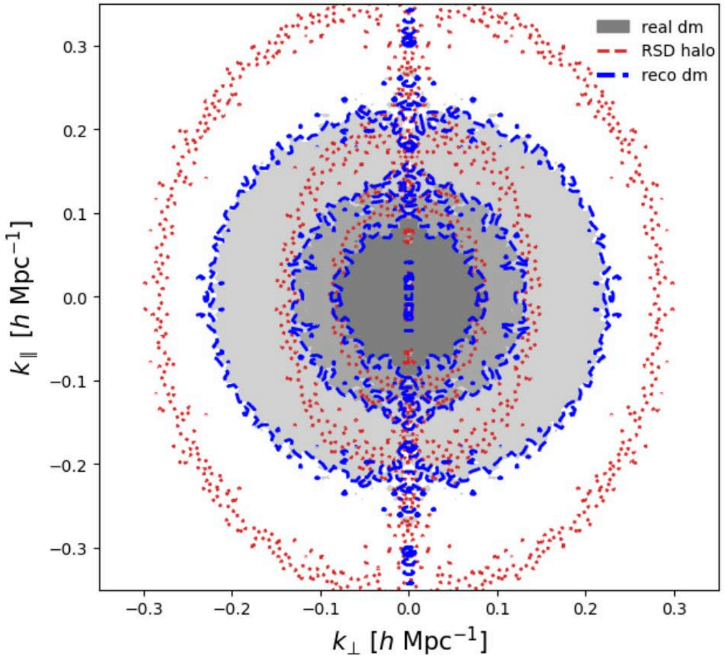


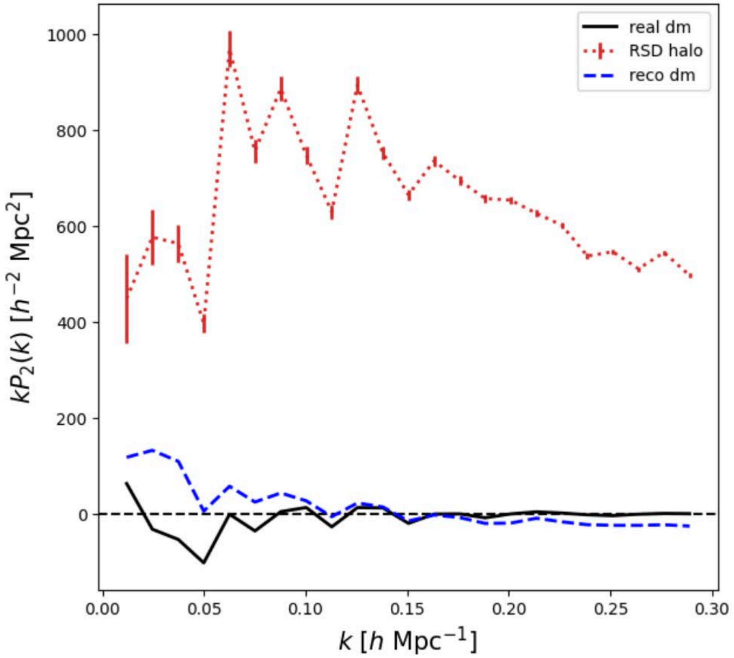


Results: Jiutian simulation

- Reconstructing dark matter density field based on UNet







The real-space P(k) is also recovered accurately, with only small reduction of the cross-correlation power spectrum at 1% and 10% levels at k = 0.1 and 0.3 h Mpc^{-1}, respectively.

Application to ELUCID simulation

Check the impact of cosmology

- COLA and Jiutian simulations :

  Planck2018 cosmology

  $\Omega_m = 0.3111$, $\Omega_\Lambda = 0.6889$, $h = 0.6766$, $\Omega_b = 0.049$, $\sigma_8 = 0.817$.

Application to ELUCID simulation

Check the impact of cosmology

- COLA and Jiutian simulations :

  Planck2018 cosmology

  $\Omega_m = 0.3111$, $\Omega_\Lambda = 0.6889$, $h = 0.6766$, $\Omega_b = 0.049$, $\sigma_8 = 0.817$.

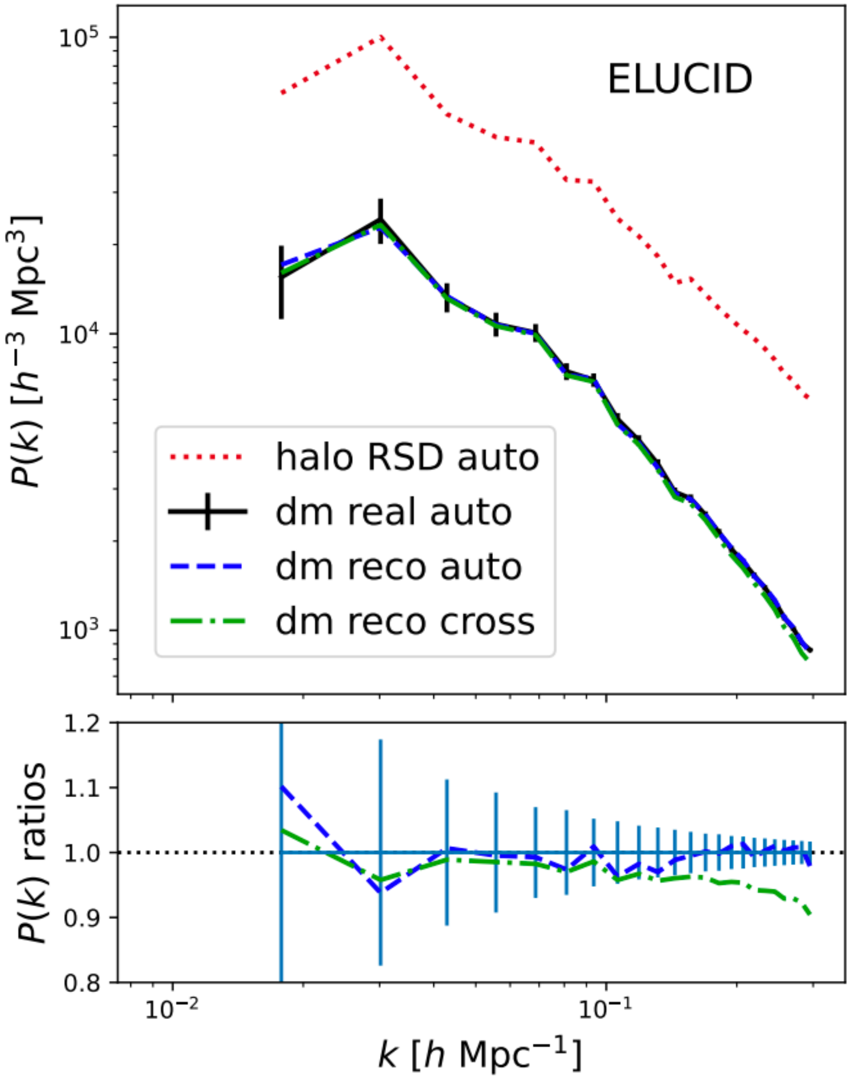
- ELUCID simulation (500Mpc/h, 3072^3 particles):

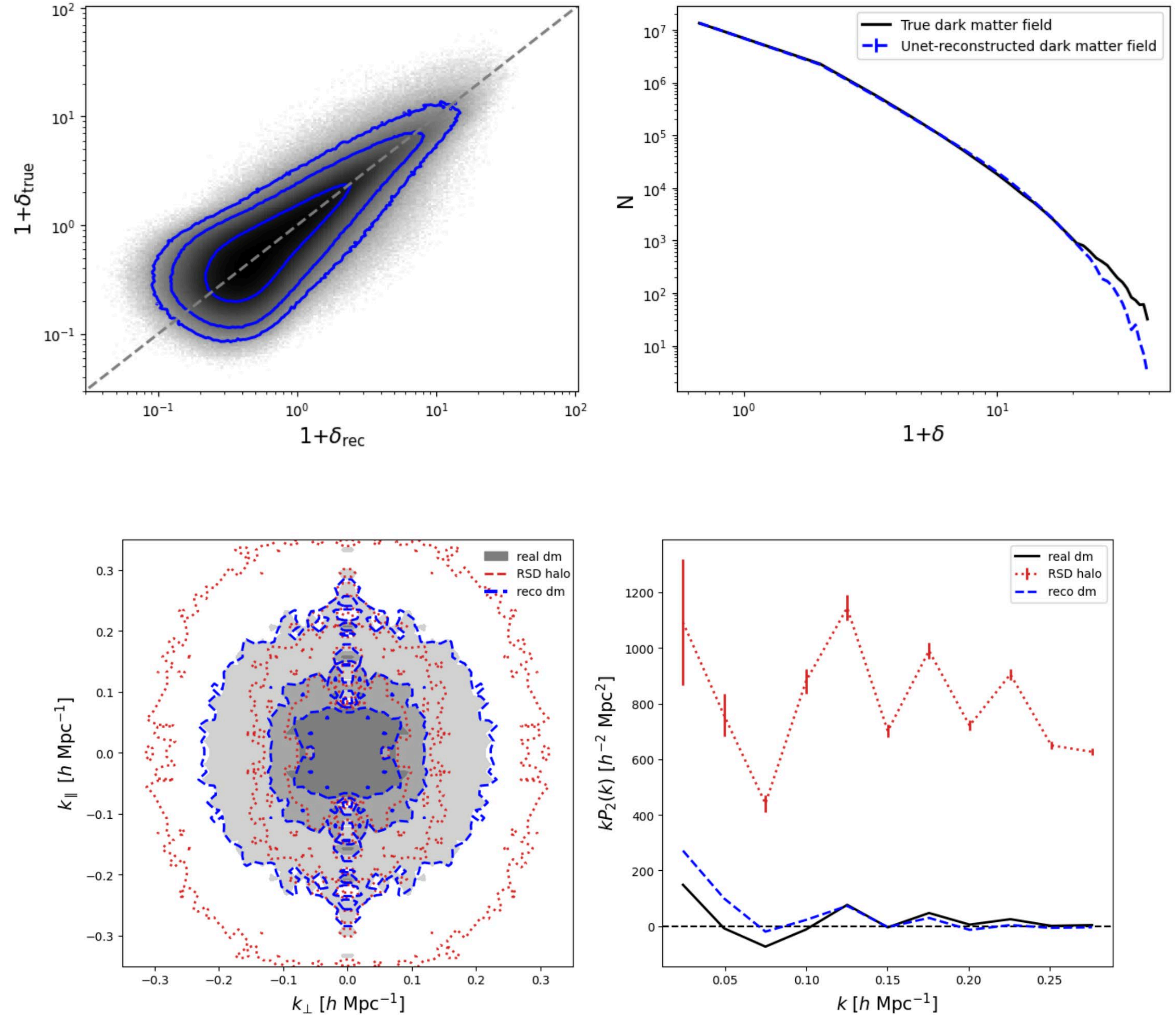
  WMAP5 cosmology

  $\Omega_m = 0.258$, $\Omega_\Lambda = 0.742$, $\Omega_b = 0.044$, h = 0.72, $\sigma_8 = 0.80$

Results: ELUCID simulation

- No large distinction of the results between the WMAP5 and Planck18 cosmology





Testing: velocity field reconstruction

- Reconstruct velocity field

VS.

Halo density field $\delta_h(k)$ with a bias b_{hm}

$$\boldsymbol{v}(\boldsymbol{k}) = H\ a\, f(\Omega)\, \frac{i\boldsymbol{k}}{k^2}\, \frac{\delta_{\rm h}(\boldsymbol{k})}{b_{\rm hm}}$$

(Wang et al 2012, Shi et al 2016)

Testing: velocity field reconstruction

- Reconstruct velocity field

UNet-reconstructed δ(k)

$$\boldsymbol{v}(\boldsymbol{k}) = H\ a\ f(\Omega)\ \frac{i\boldsymbol{k}}{k^2}\ \delta(\boldsymbol{k})$$

VS.

Halo density field $\delta_h(k)$ with a bias b_{hm}

$$\boldsymbol{v}(\boldsymbol{k}) = H\ a\ f(\Omega)\ \frac{i\boldsymbol{k}}{k^2}\ \frac{\delta_{\rm h}(\boldsymbol{k})}{b_{\rm hm}}$$

(Wang et al 2012, Shi et al 2016)

Testing: velocity field reconstruction

- Reconstruct velocity field

UNet-reconstructed δ(k)

$$\boldsymbol{v}(\boldsymbol{k}) = H\, a\, f(\Omega)\, \frac{i\boldsymbol{k}}{k^2}\, \delta(\boldsymbol{k})$$

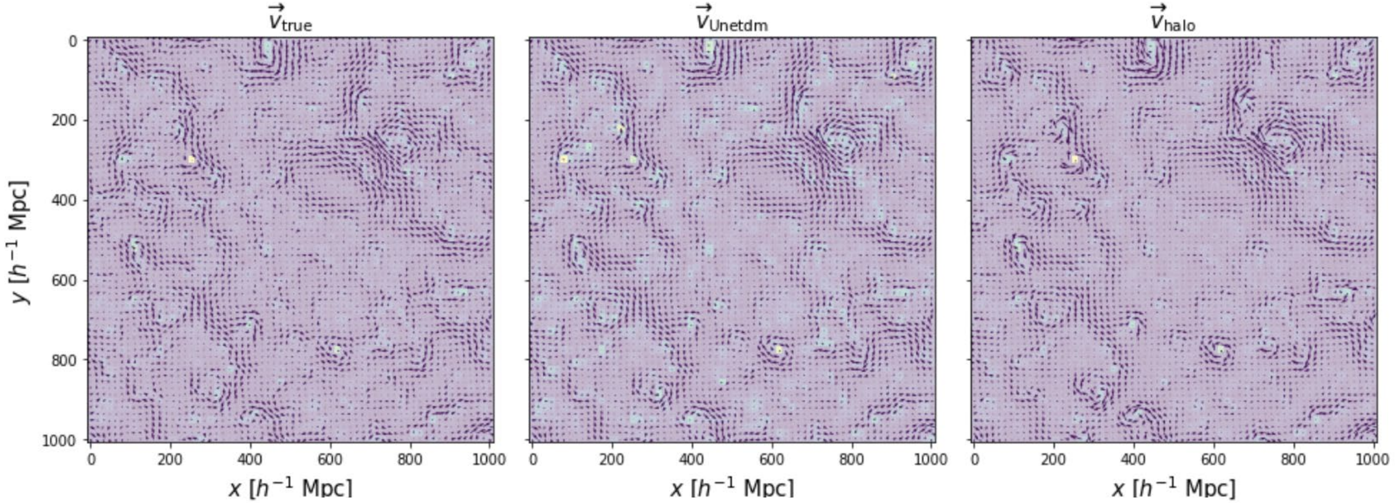
VS.

Halo density field $\delta_h(k)$ with a bias b_{hm}

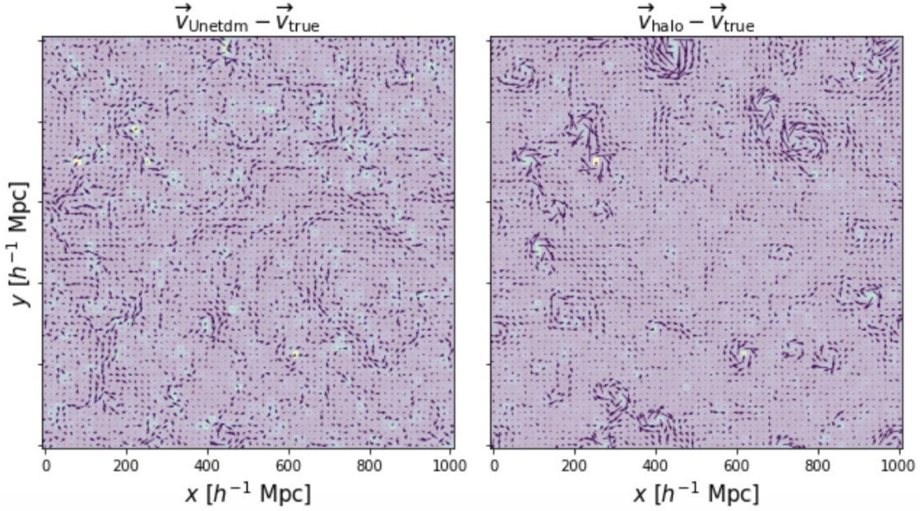
$$\boldsymbol{v}(\boldsymbol{k}) = H\, a\, f(\Omega)\, \frac{i\boldsymbol{k}}{k^2}\, \frac{\delta_h(\boldsymbol{k})}{b_{hm}}$$

(Wang et al 2012, Shi et al 2016)

Velocity field



Velocity difference field



UNet dark matter

Halo-based with linear bias

Testing: velocity field reconstruction

- Reconstruct velocity field

UNet-reconstructed δ(k)

$$\boldsymbol{v}(\boldsymbol{k}) = H\ a\ f(\Omega)\ \frac{i\boldsymbol{k}}{k^2}\ \delta(\boldsymbol{k})$$

VS.

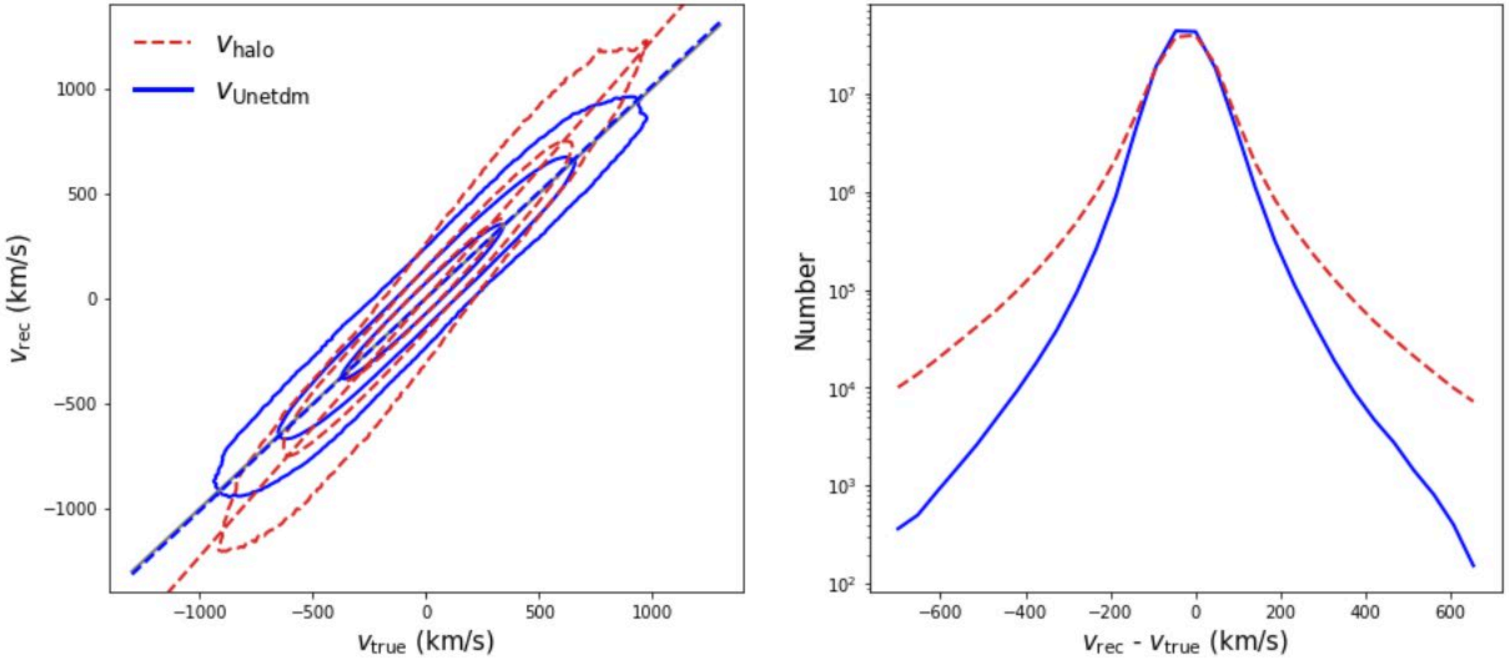
Halo density field $\delta_h(k)$ with a bias b_{hm}

$$\boldsymbol{v}(\boldsymbol{k}) = H\ a\ f(\Omega)\ \frac{i\boldsymbol{k}}{k^2}\ \frac{\delta_{\rm h}(\boldsymbol{k})}{b_{\rm hm}}$$

(Wang et al 2012, Shi et al 2016)

	Slope	Scatter
Halo :	1.15	78.2 km/s
UNet :	1.01	57.0 km/s

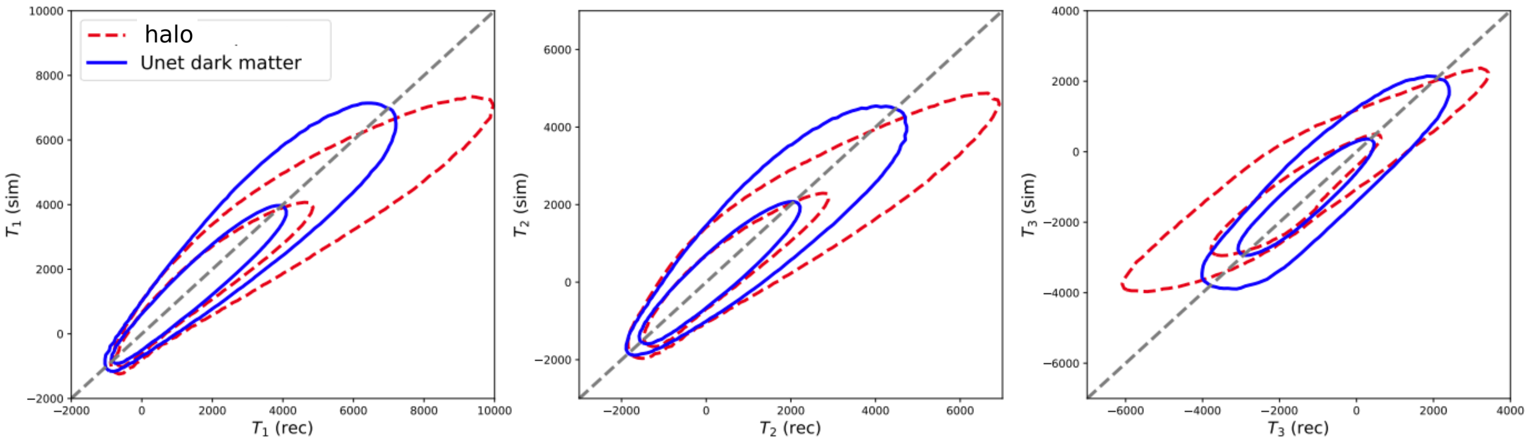
- Unbiased relation
- 21.1% scatter errror reduction



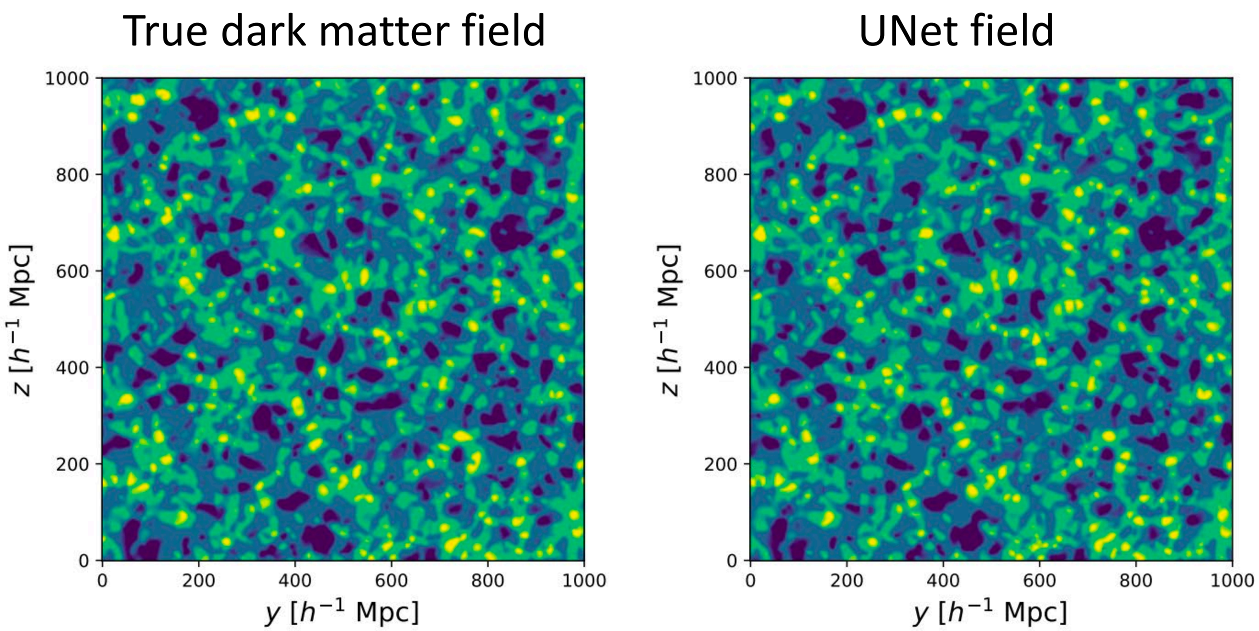
The three contours encompass 67%, 95%, and 99% of the grid cells

Testing: tidal field reconstruction

- Reconstruct tidal field:



- Classification of the large-scale structure:



- cluster : yellow
- filament: yellow-green
- sheet: green
- void: black

Summary & Conlusions

1) Method: Reconstruct the cosmic density field from the redshift-space halo feild based on Unet

2) Testing:

- Three simulations: COLA, Jiutian and ELUCID
- Statistics: projected density, density-density relation, historgram, 1D & 2D P(k)
- Fields: denstiy, velocity and tidal fields

3) Conlusions:

- Accurate reconstruction with only 1% and 10% reduction of the cross P(k) at k = 0.1 and 0.3 h Mpc^{-1}
- RSD corrected successfully
- Low-resolution-COLA-trained network generalizes to the typical high-resolution N-body simulation
- UNet-based field outperforms the traditional method in accurately recovering the velocity & tidal field

Summary & Conlusions

1) Method: Reconstruct the cosmic density field from the redshift-space halo feild based on Unet

2) Testing:

- Three simulations: COLA, Jiutian and ELUCID
- Statistics: projected density, density-density relation, historgram, 1D & 2D P(k)
- Fields: denstiy, velocity and tidal fields

3) Conlusions:

- Accurate reconstruction with only 1% and 10% reduction of the cross P(k) at k = 0.1 and 0.3 h Mpc^{-1}
- RSD corrected successfully
- Low-resolution-COLA-trained network generalizes to the typical high-resolution N-body simulation
- UNet-based field outperforms the traditional method in accurately recovering the velocity & tidal field

Summary & Conlusions

1) Method: Reconstruct the cosmic density field from the redshift-space halo feild based on Unet

2) Testing:

- Three simulations: COLA, Jiutian and ELUCID
- Statistics: projected density, density-density relation, historgram, 1D & 2D P(k)
- Fields: denstiy, velocity and tidal fields

3) Conlusions:

- Accurate reconstruction with only 1% and 10% reduction of the cross P(k) at k = 0.1 and 0.3 h Mpc^{-1}
- RSD corrected successfully
- Low-resolution-COLA-trained network generalizes to the typical high-resolution N-body simulation
- UNet-based field outperforms the traditional method in accurately recovering the velocity & tidal field

Summary & Conlusions

1) Method: Reconstruct the cosmic density field from the redshift-space halo feild based on Unet

2) Testing:

- Three simulations: COLA, Jiutian and ELUCID
- Statistics: projected density, density-density relation, historgram, 1D & 2D P(k)
- Fields: denstiy, velocity and tidal fields

3) Conlusions:

- Accurate reconstruction with only 1% and 10% reduction of the cross P(k) at k = 0.1 and 0.3 h Mpc^{-1}
- RSD corrected successfully
- Low-resolution-COLA-trained network generalizes to the typical high-resolution N-body simulation
- UNet-based field outperforms the traditional method in accurately recovering the velocity & tidal field