Testing the FoF
(not the meeting, but the algorithm)

Ariel Zandivarez

Collaborators

Argentina : Ariel Zandivarez, Eugenia Díaz-Giménez

Brasil: Claudia Mendes de Oliveira, Renato Dupke

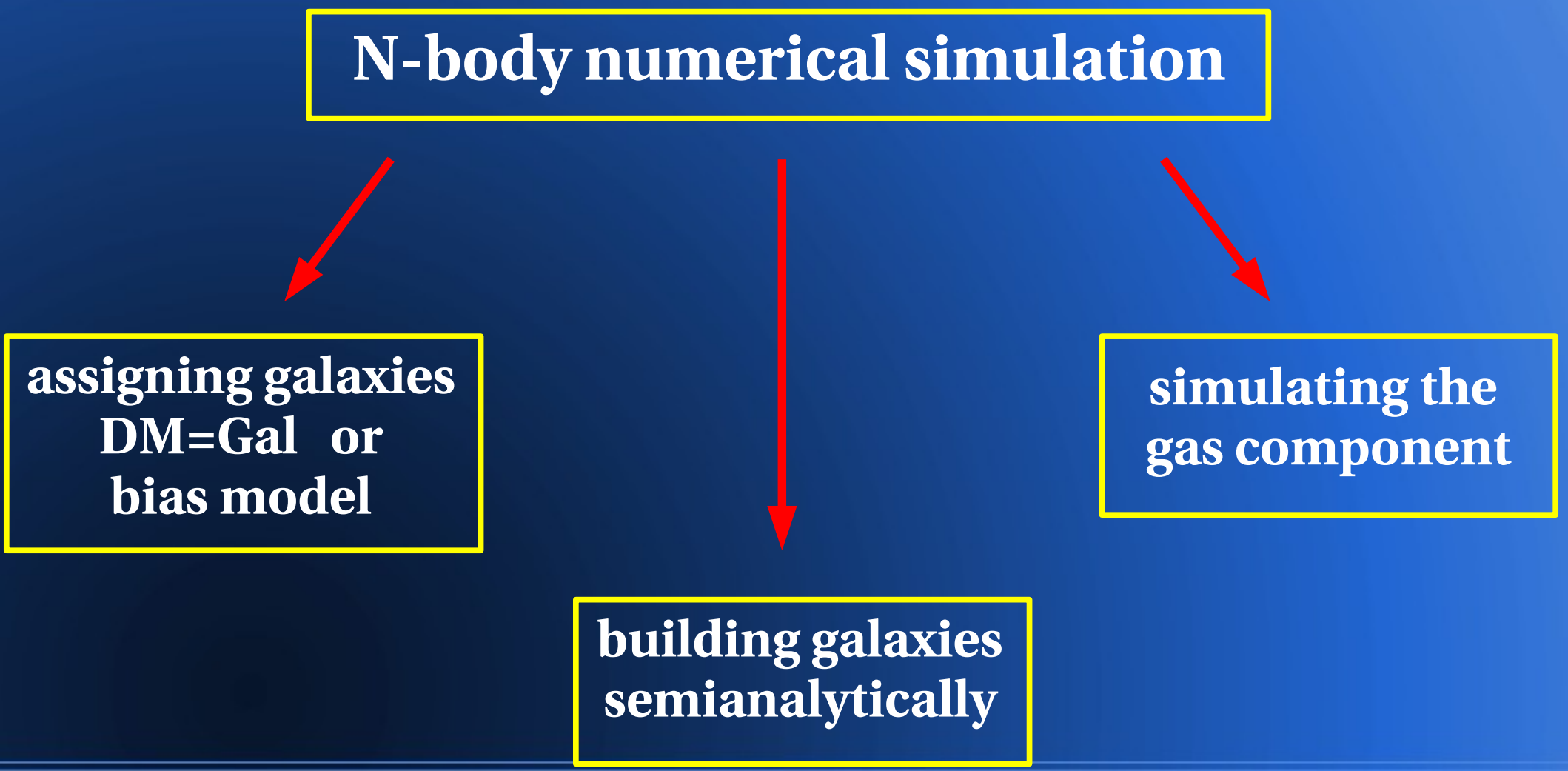
España: Narciso Benitez, Begoña Ascaso

and JPAS-team.

Aim of the work

- Construct mock catalogues for the future JPAS.
- Test algorithms to identify galaxy groups.
- Predict purity and completeness of the resulting groups samples.
- Test different statistical tools.

How to construct a mock catalogue?



How to construct a mock catalogue?

N-body numerical simulation

building galaxies semianalytically

N-body cosmological numerical simulation: The Millennium Run

Virgo Consortium (Springel et al 2005)

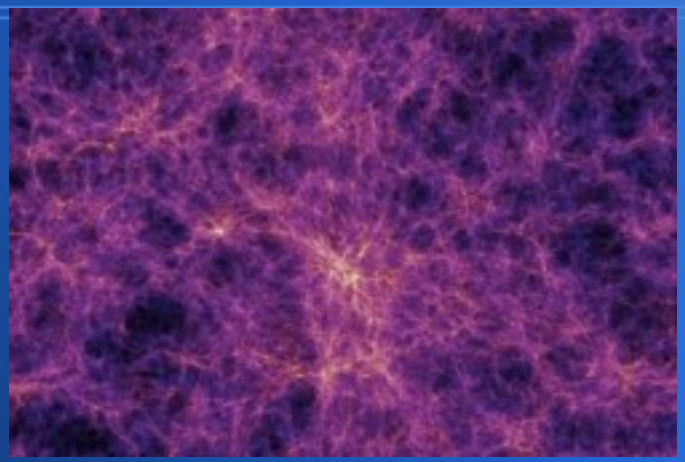
- Cubic Volume of 500 h^{-1} Mpc of side.
- 10^9 dark matter particles.
- $1.7 \times 10^{10}\ h^{-1}\ M_o$ mass resolution.
- Cosmological parameters:

$\Omega_m=0.25, \Omega_b=0.045, \Omega_\Lambda=0.75, h=0.73, \sigma_8=0.9, n=1.$

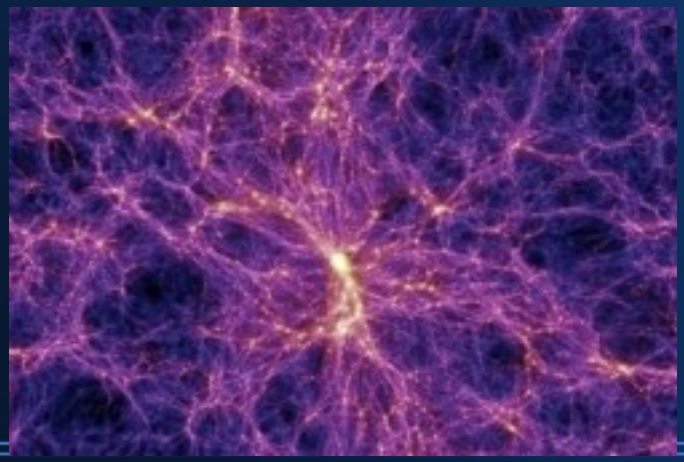
- Running time: a couple of months in a fully dedicated super-computing cluster.

N-body cosmological numerical simulation: The Millennium Run

Z = 18.3



Z = 5.7

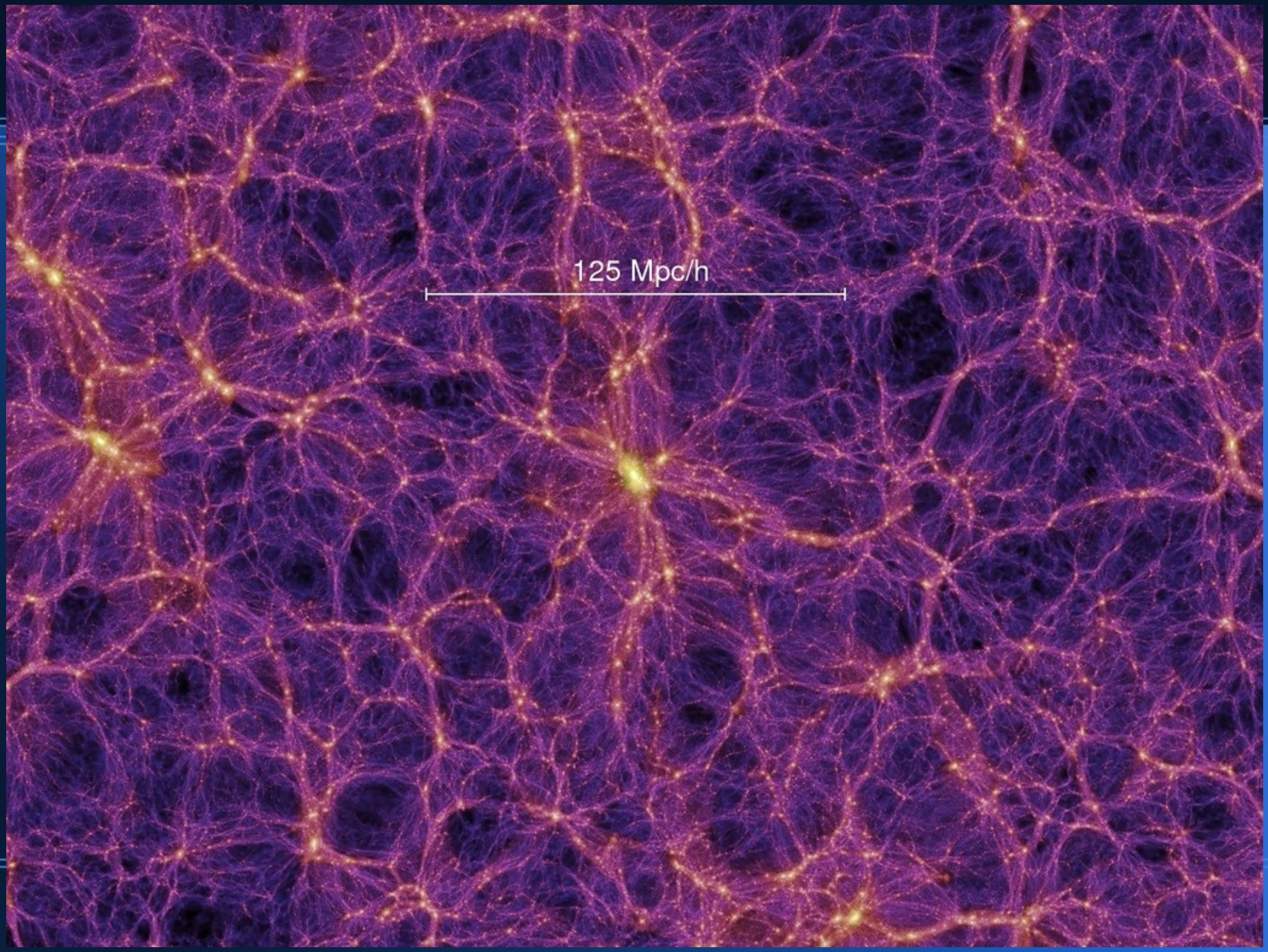


Z = 1.4

Z = 0

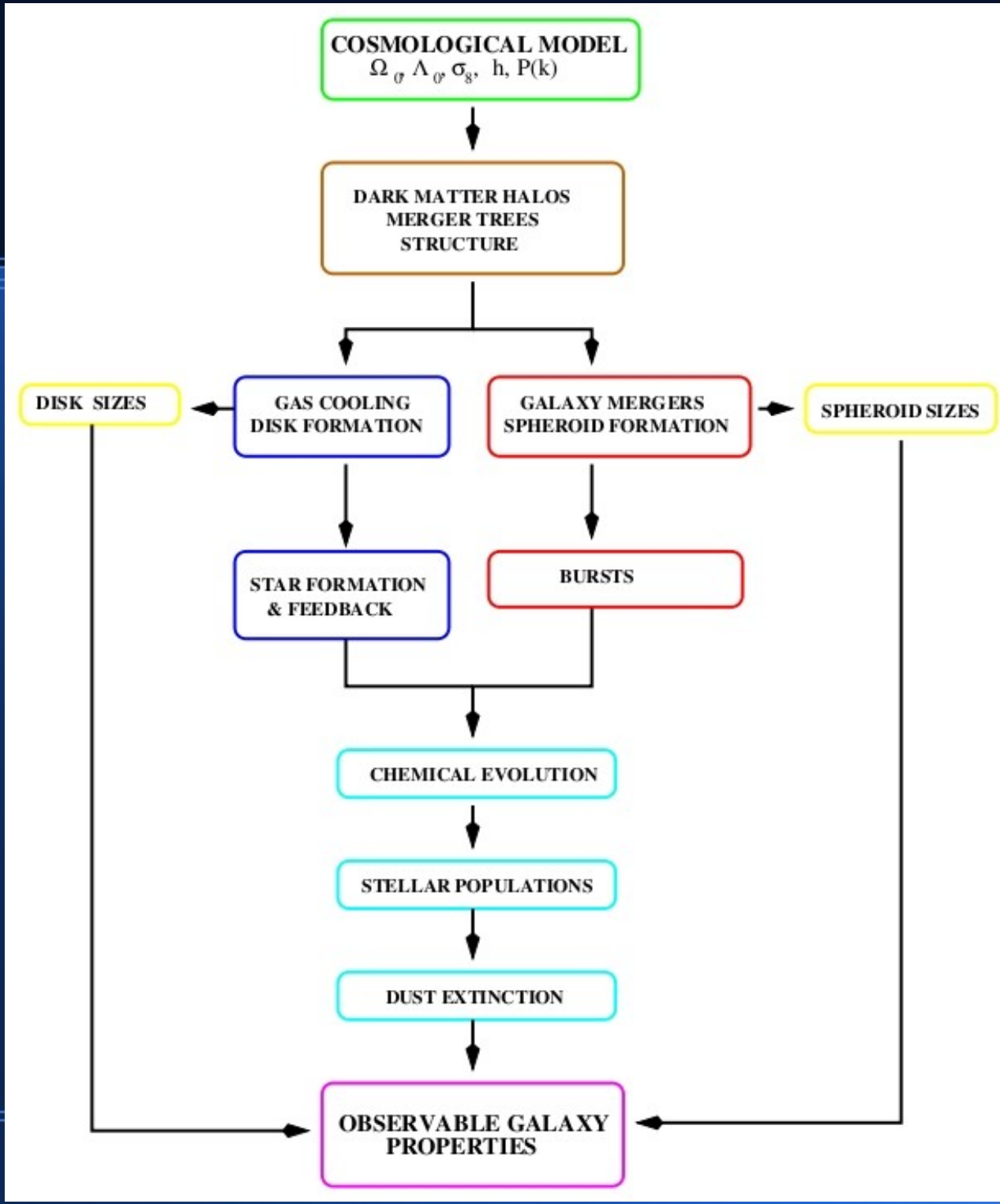
The Millennium Run

Z = 0

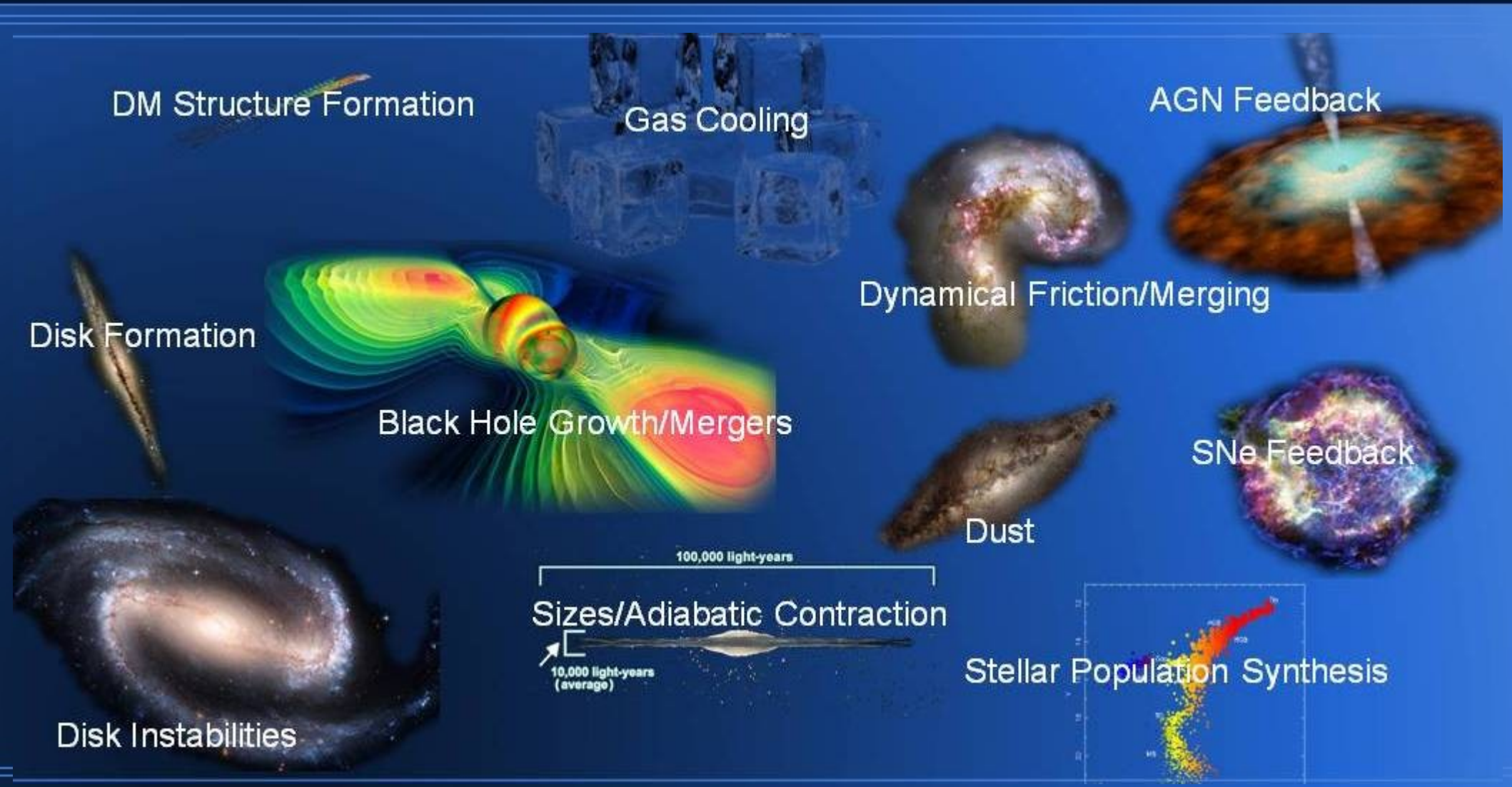


How to build semianalytical galaxies?

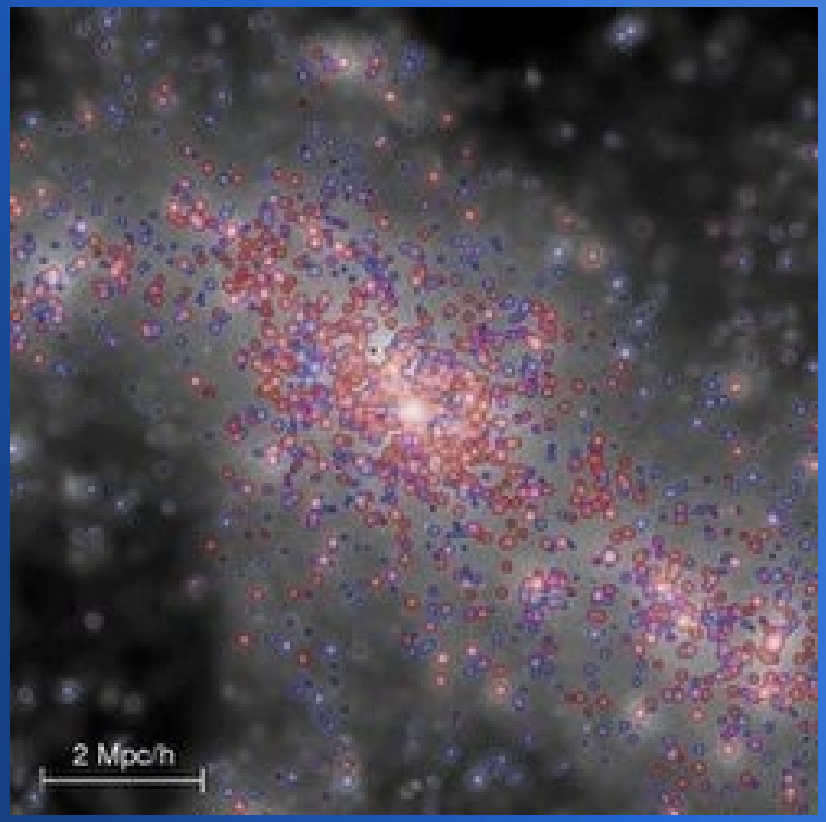
The semianalytical model of galaxy formation



The semianalytical model of galaxy formation



The semianalytical model on top of the Millennium Simulation

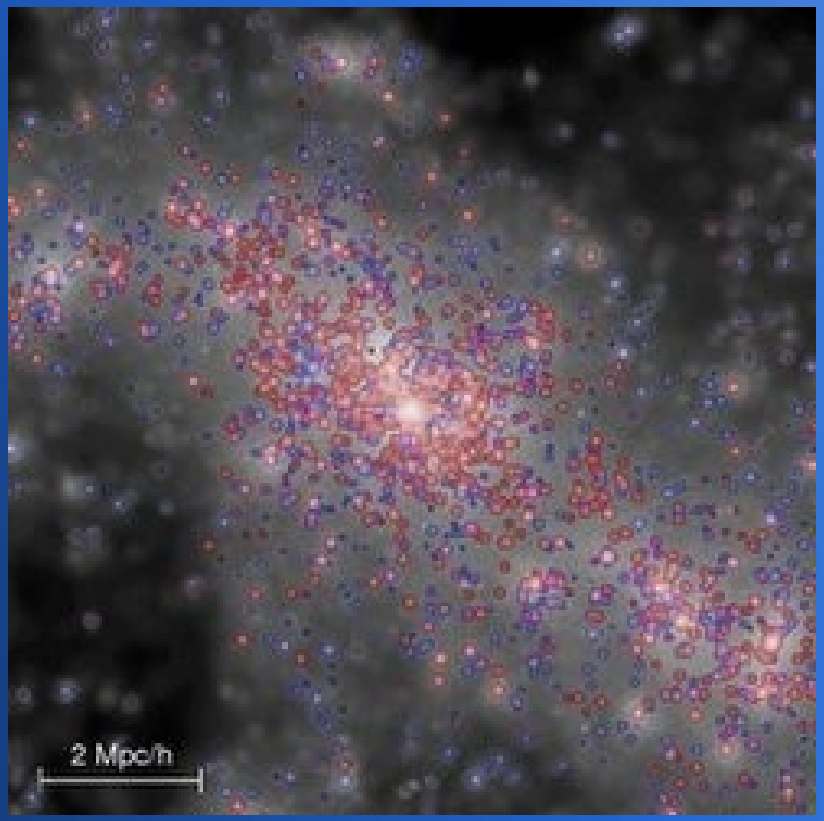


Dark Matter

Galaxies

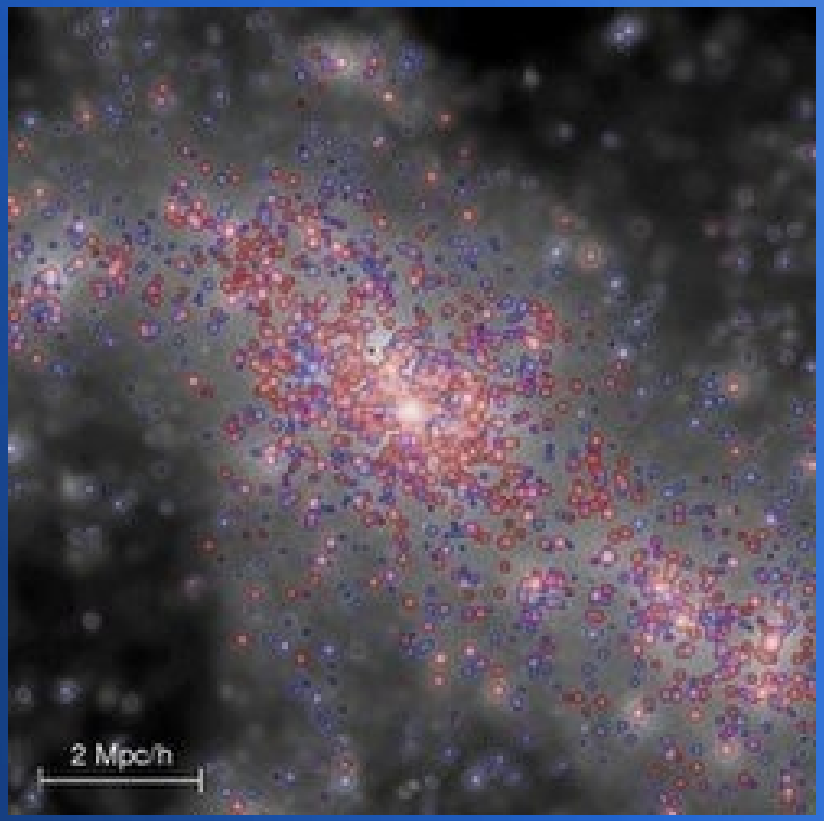
The semianalytical model on top of the Millennium Simulation

- Croton et al (2006)
- Bower et al. (2006)
- De Lucia et al. (2007)
- Bertone et al. (2007)
- Guo et al. (2011)



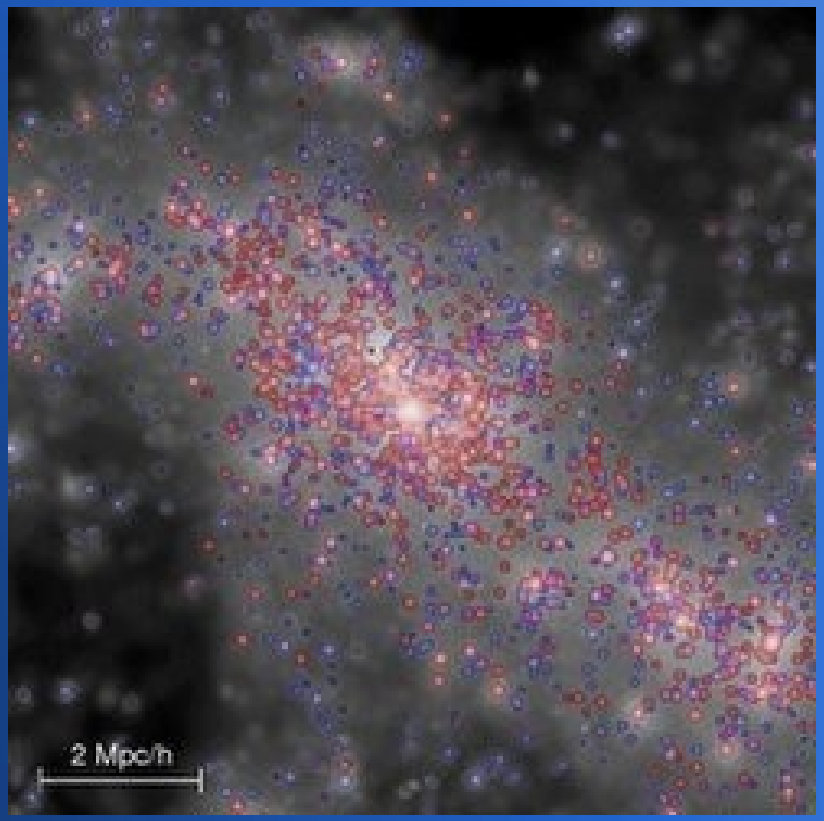
The semianalytical model on top of the Millennium Simulation

- Croton et al (2006)
- Bower et al. (2006)
- De Lucia et al. (2007)
- Bertone et al. (2007)
- Guo et al. (2011)



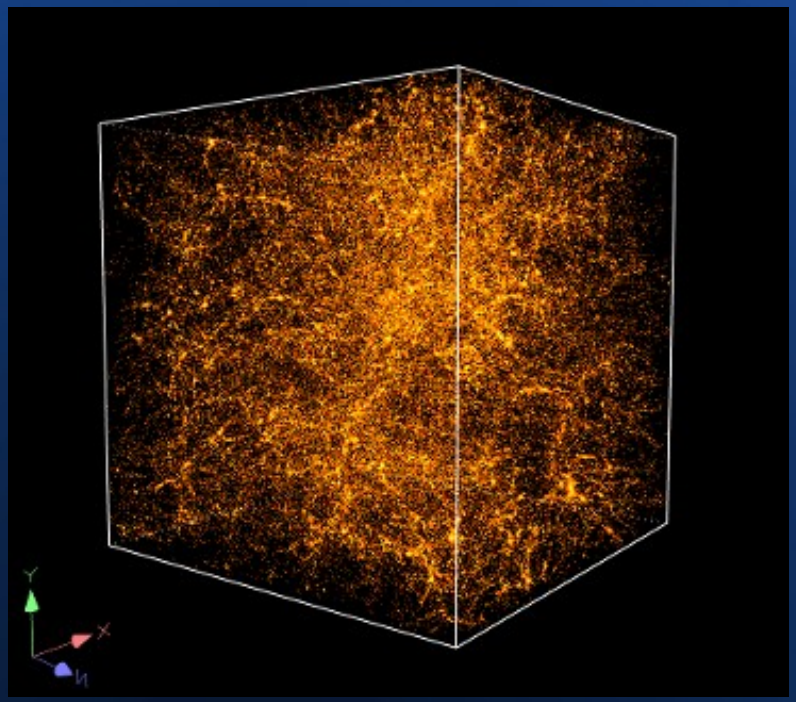
The semianalytical model on top of the Millennium Simulation

Guo et al. (2011)



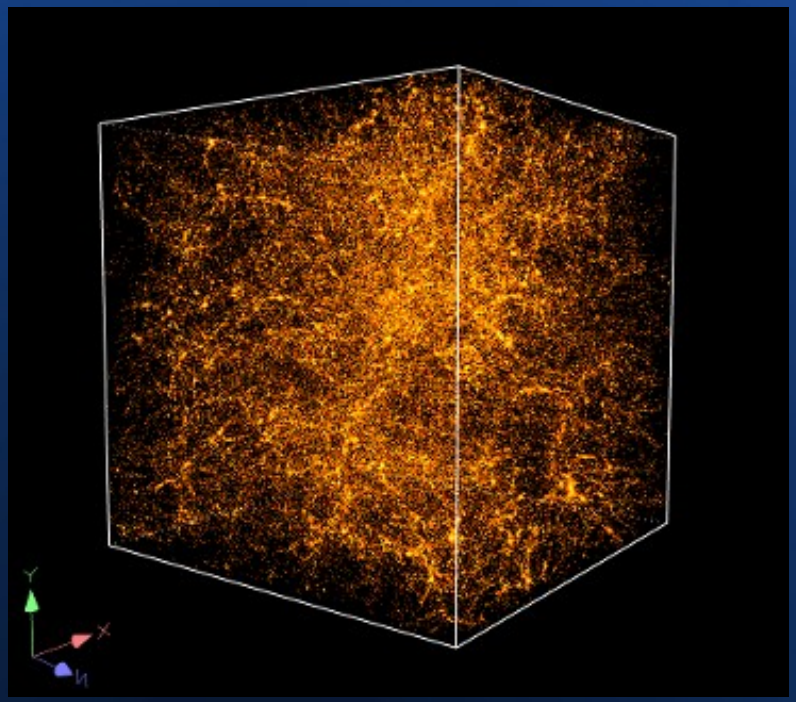
- Increased the eficiency of supernova feedback.
- More realistic satellite galaxy evolution and mergers.
- Tidal disruption of satellite galaxies.

Reaching the survey deep

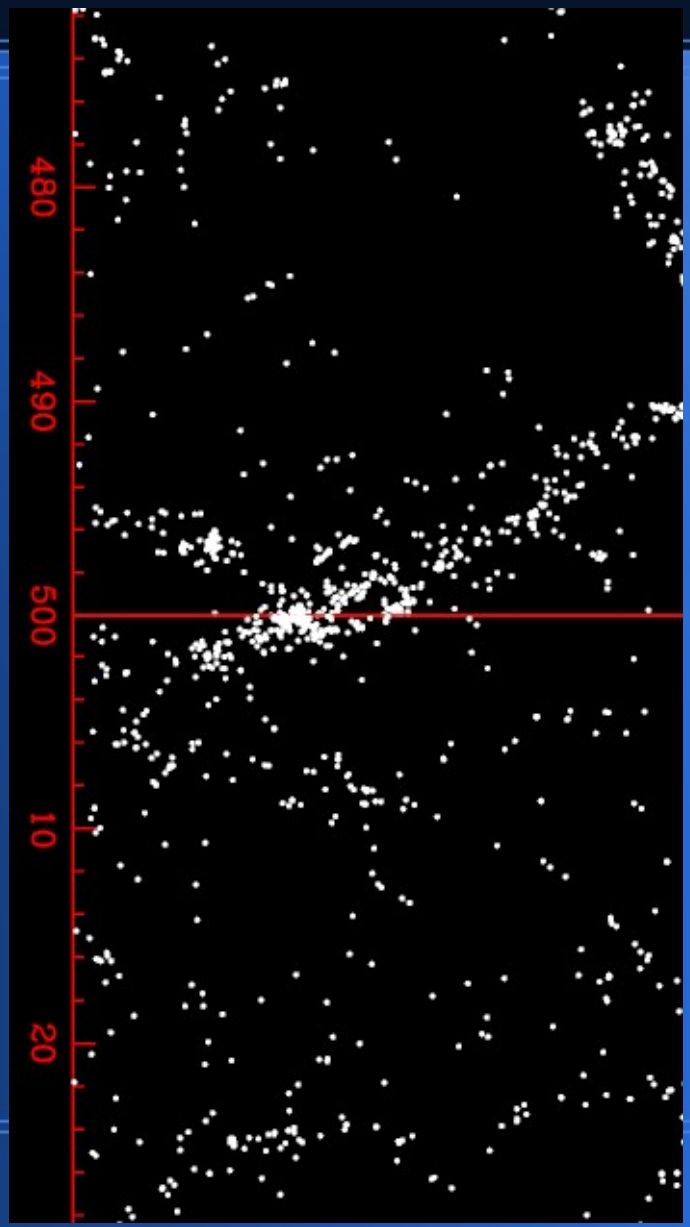


500 h^{-1} Mpc

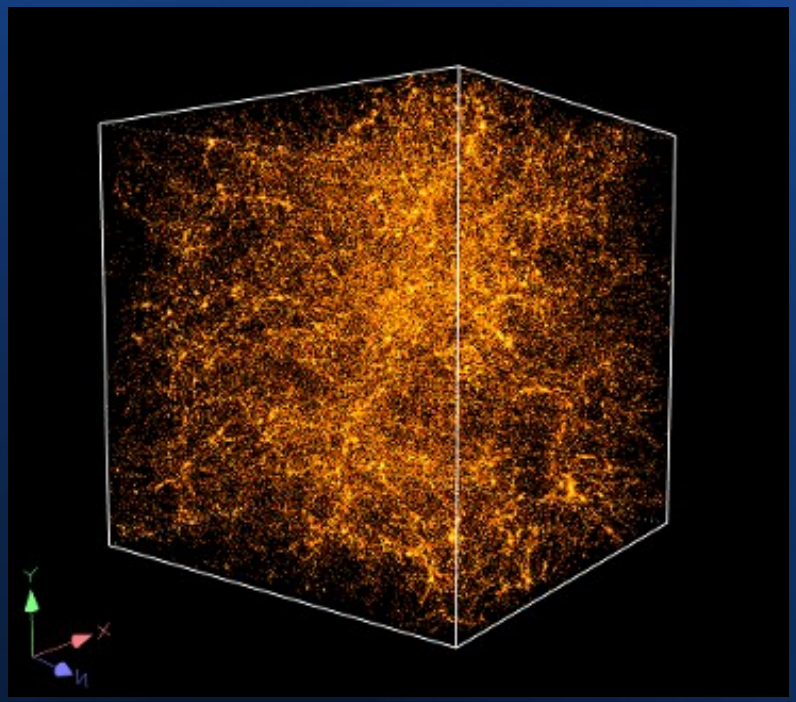
Stacking cubes to reach the survey deep



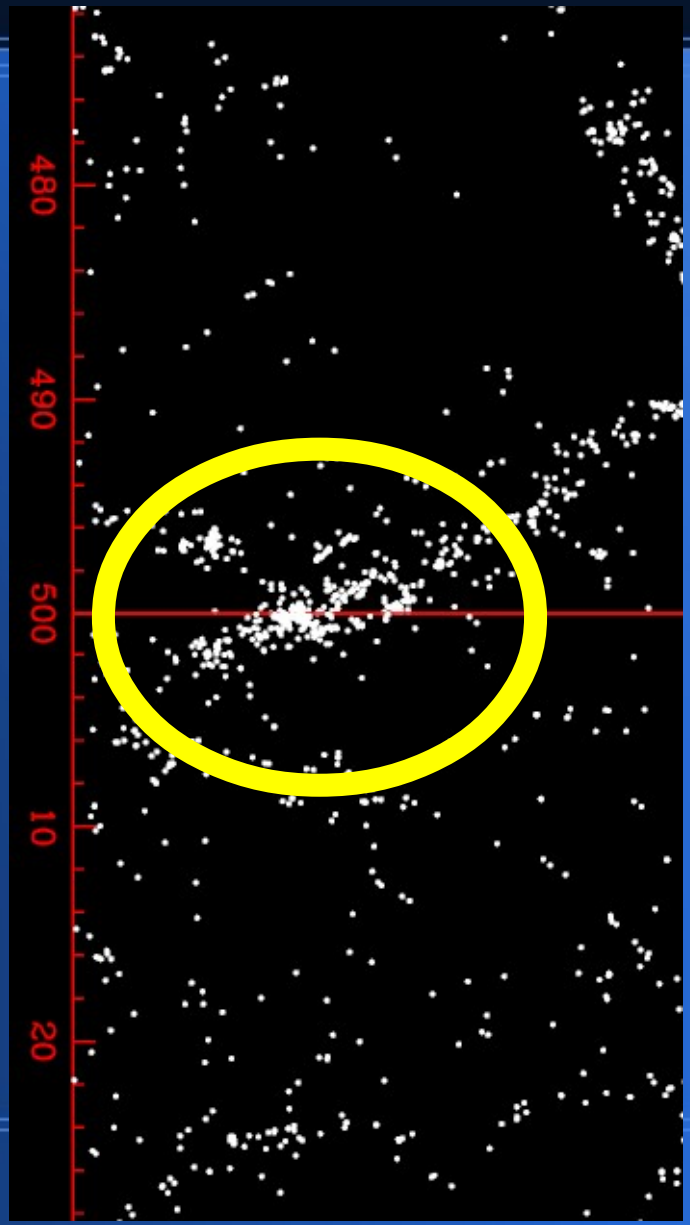
Periodic box



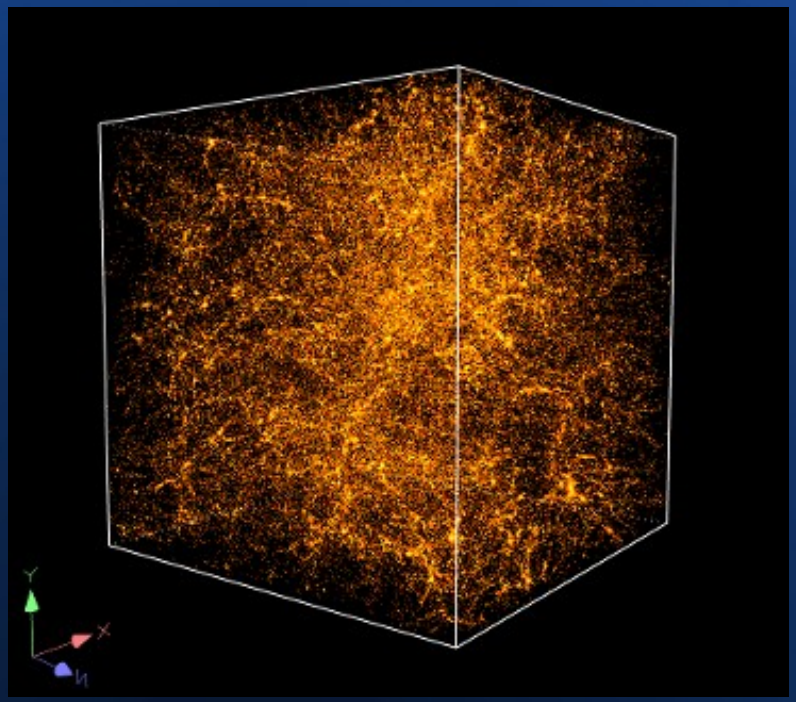
Stacking cubes to reach the survey deep



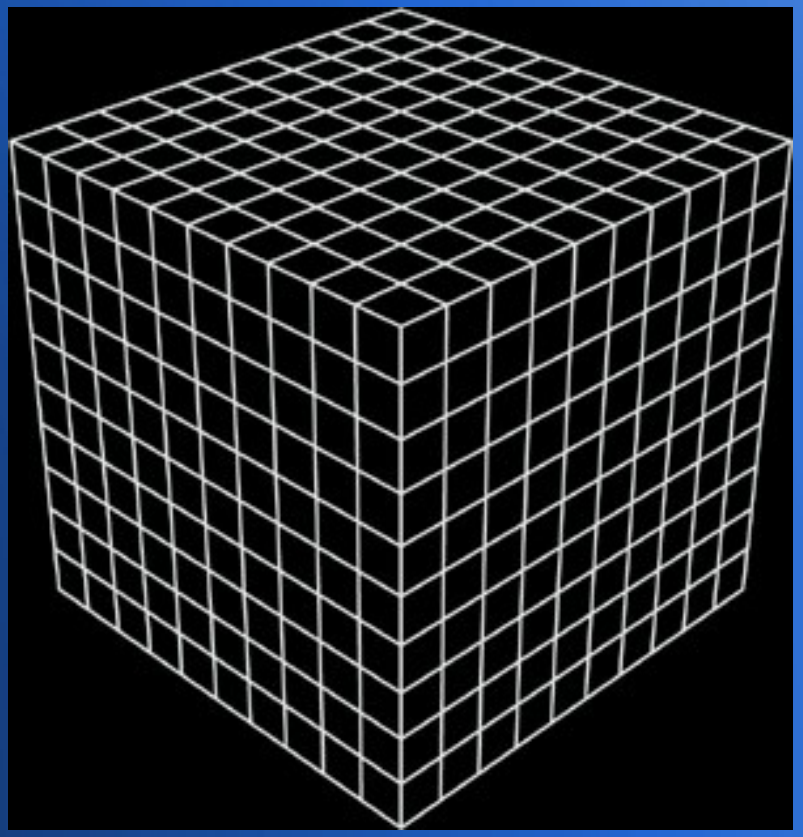
Periodic box



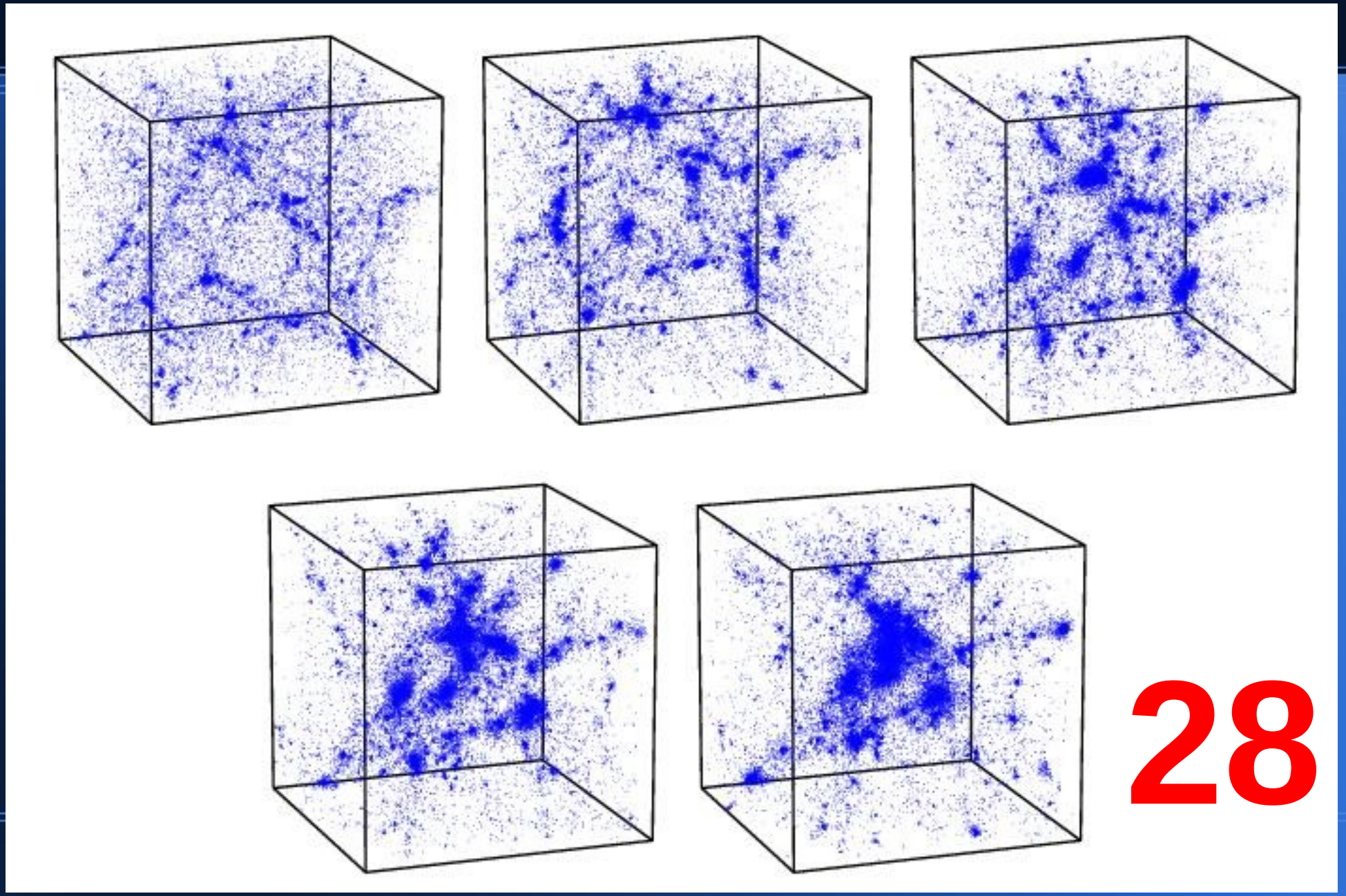
Stacking cubes to reach the survey deep



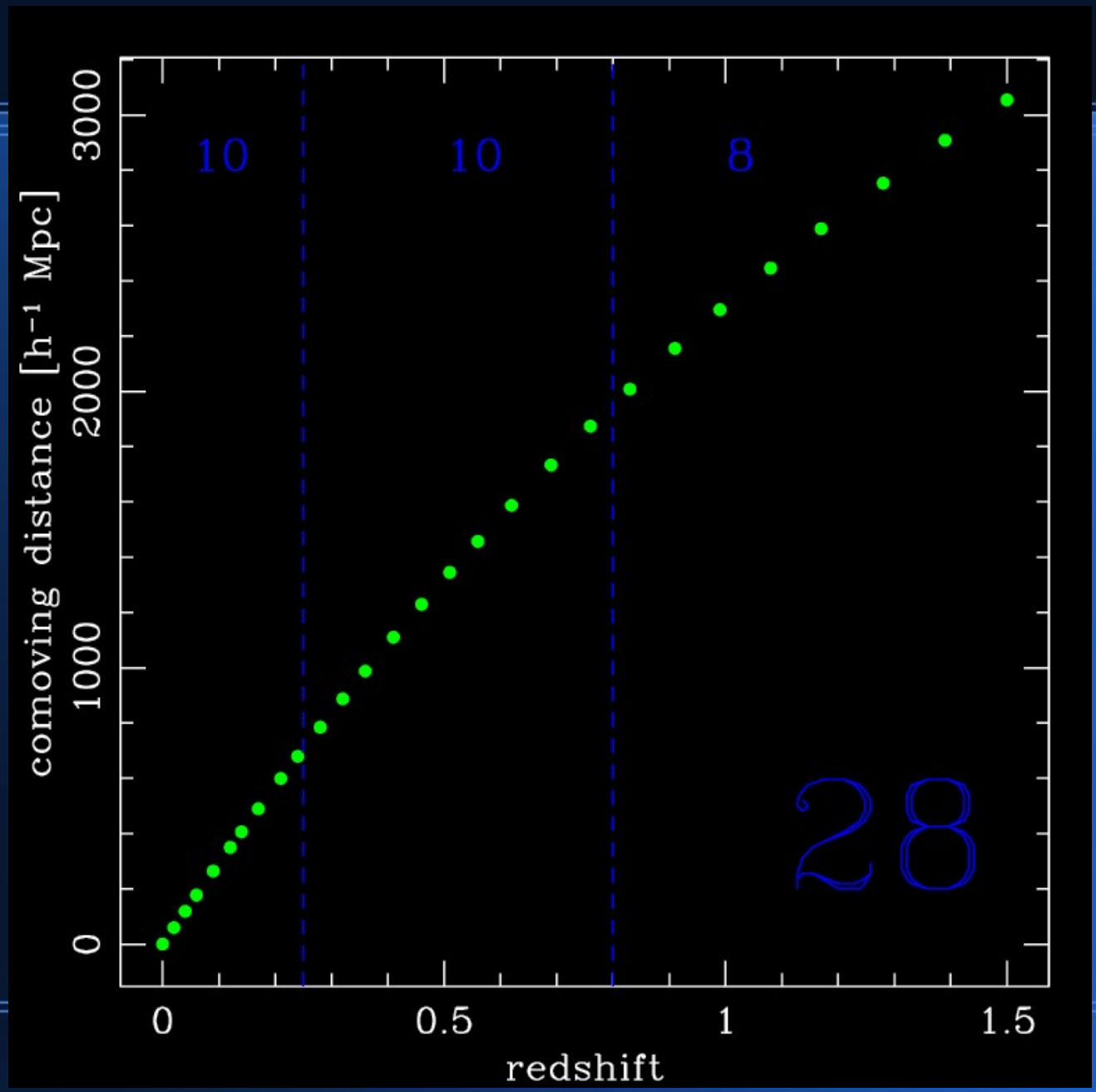
Periodic box



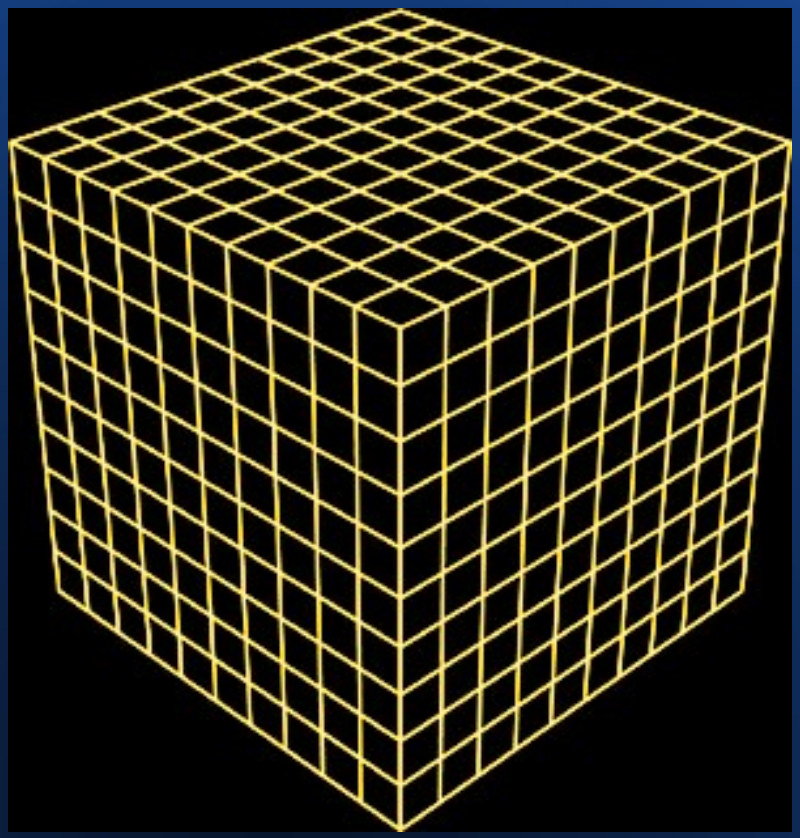
Using different snapshots from the simulation

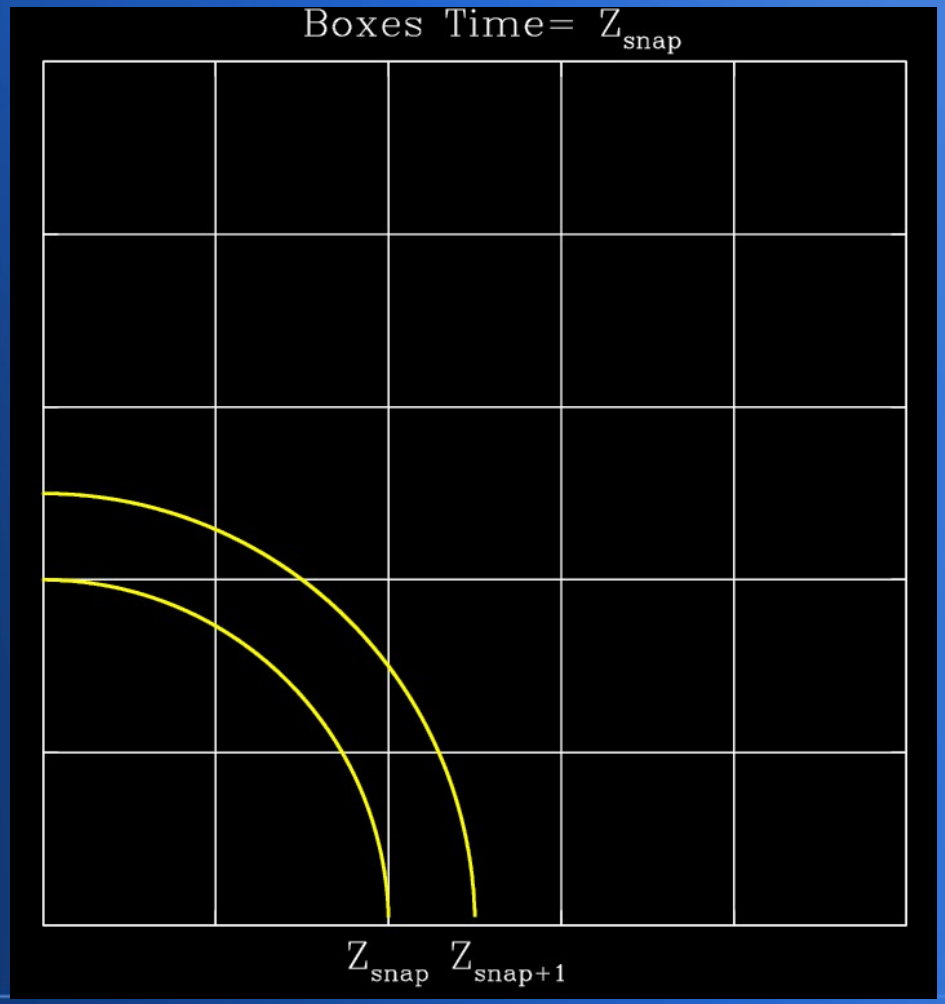


Using different snapshots from the simulation

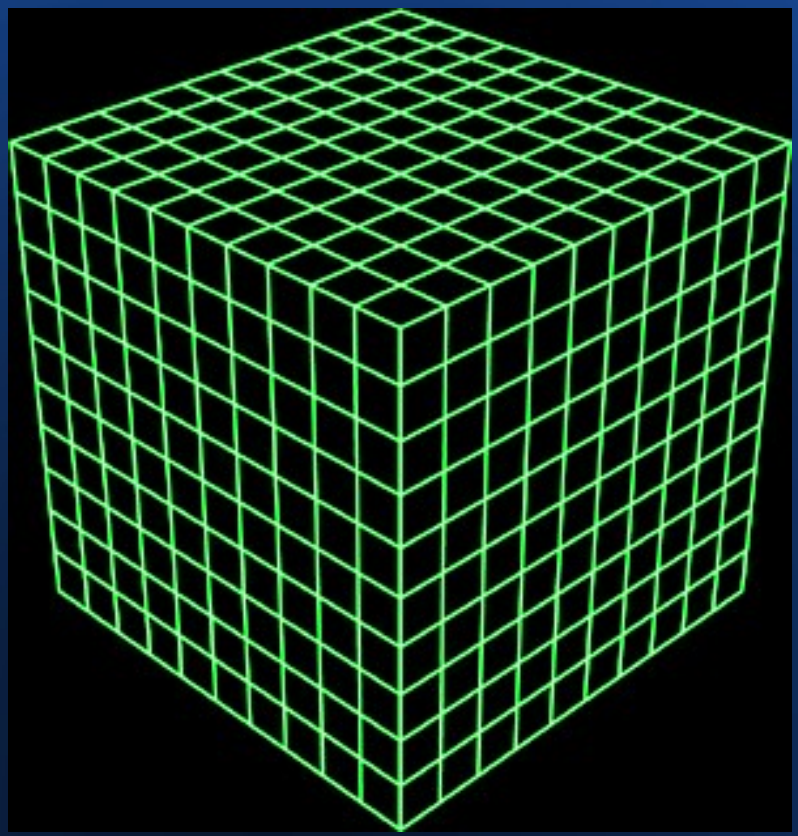


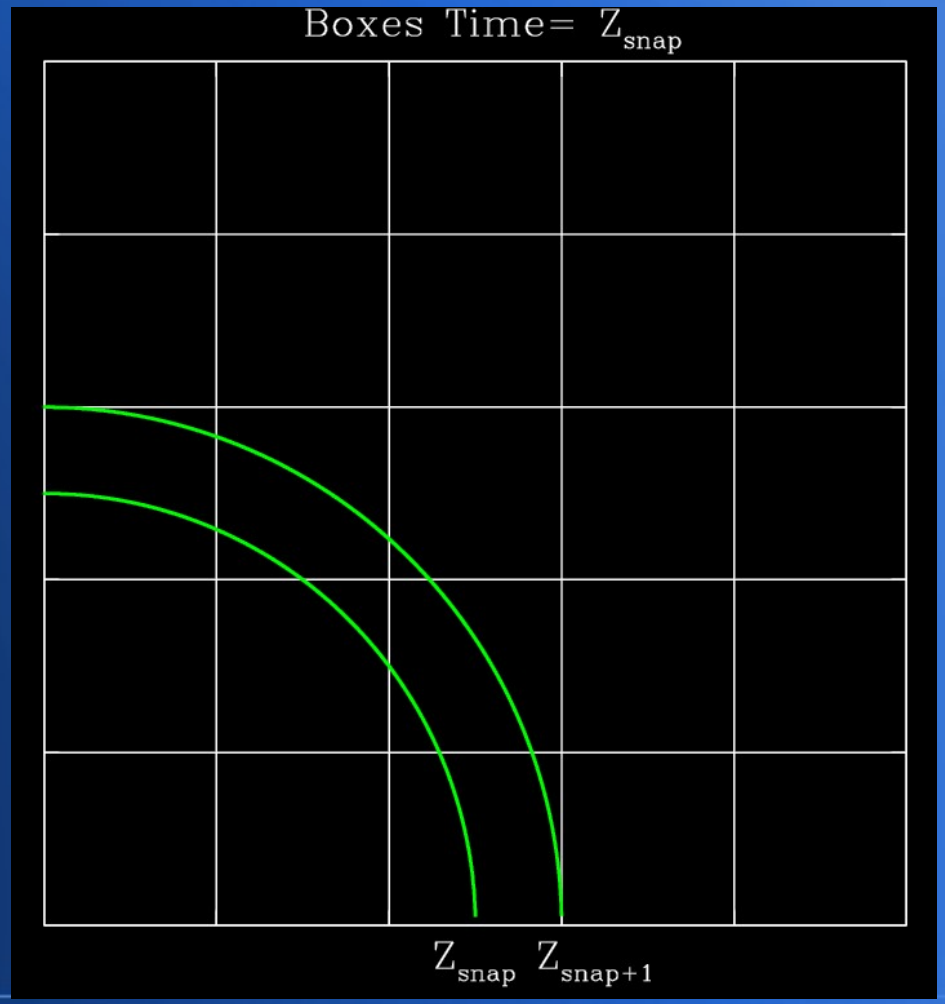
Extracting the strips from different times



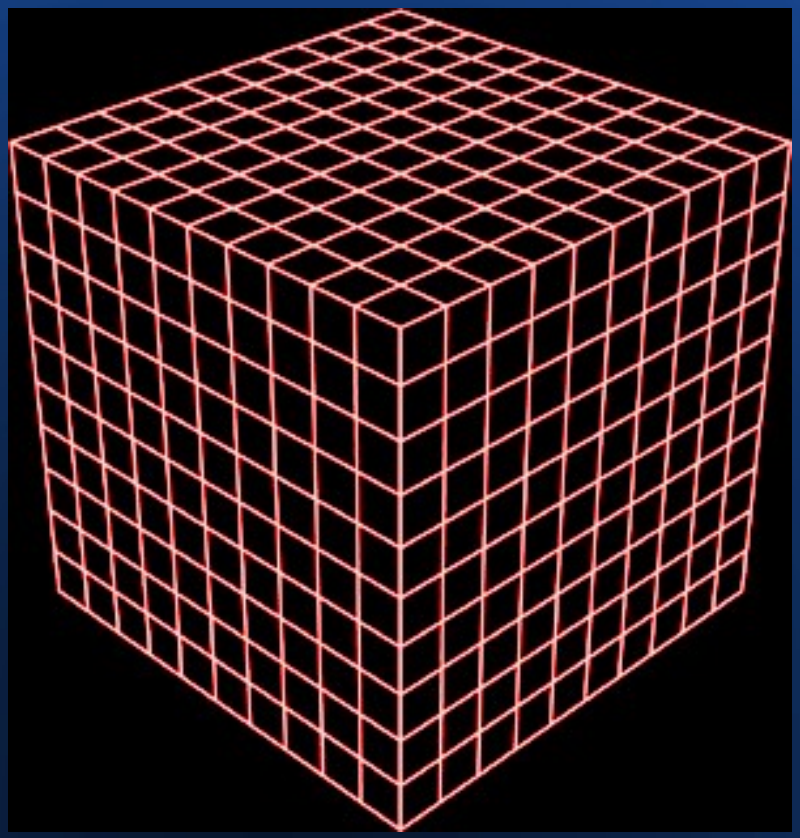


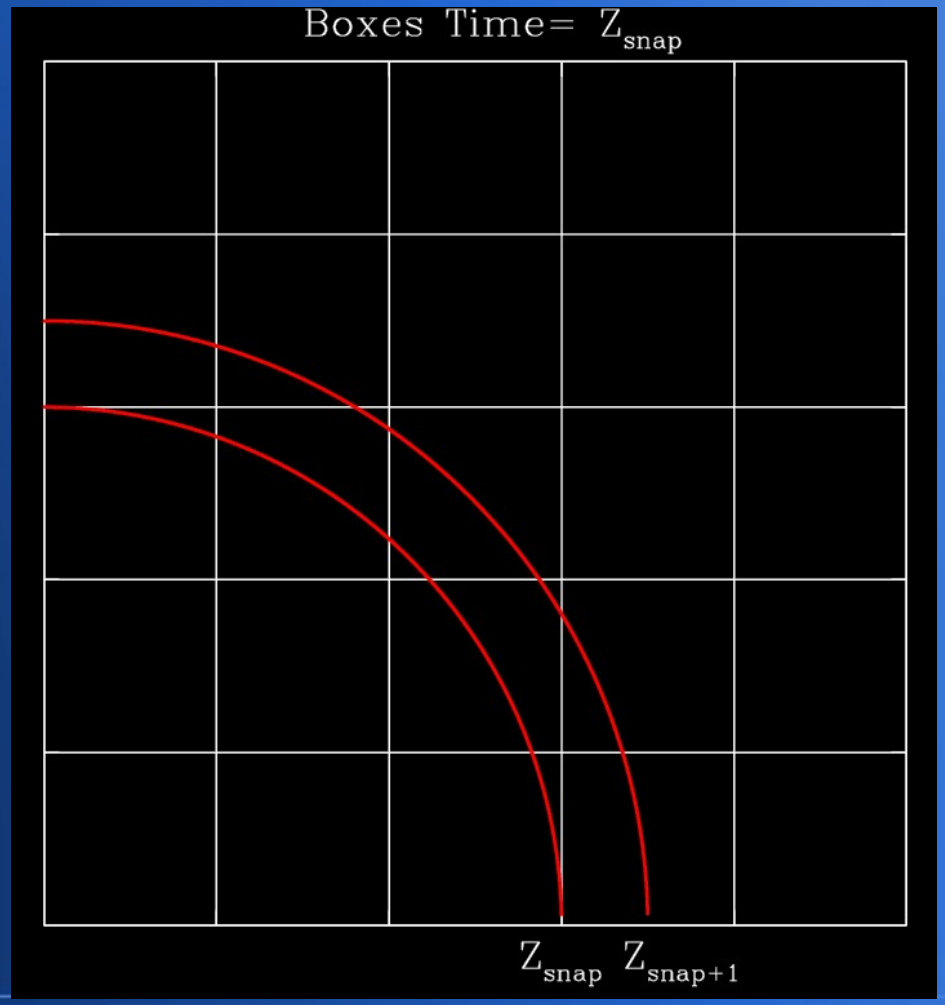
Extracting the strips from different times



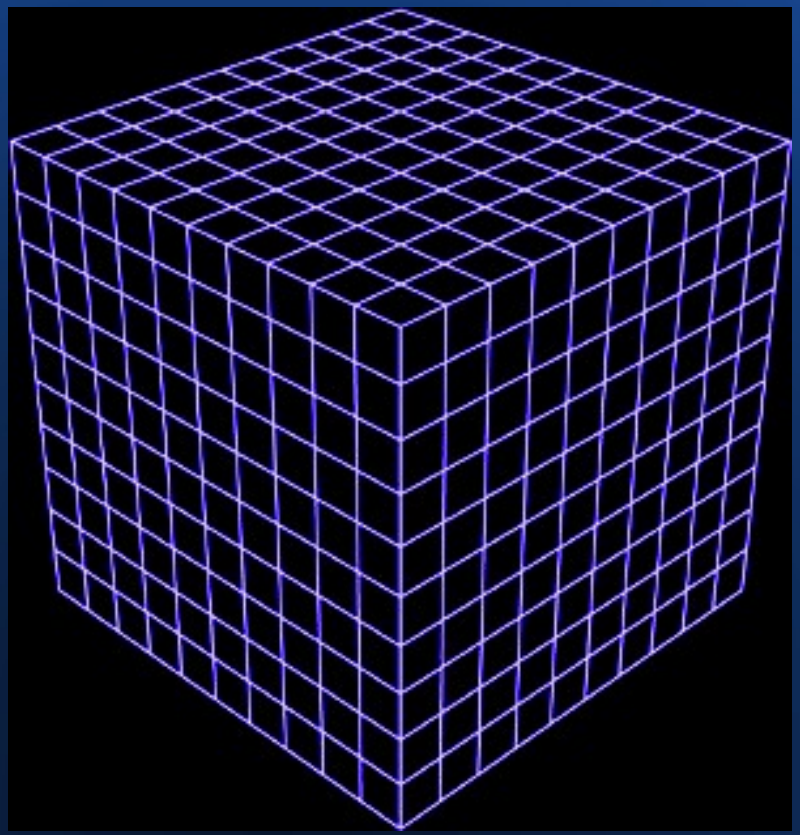


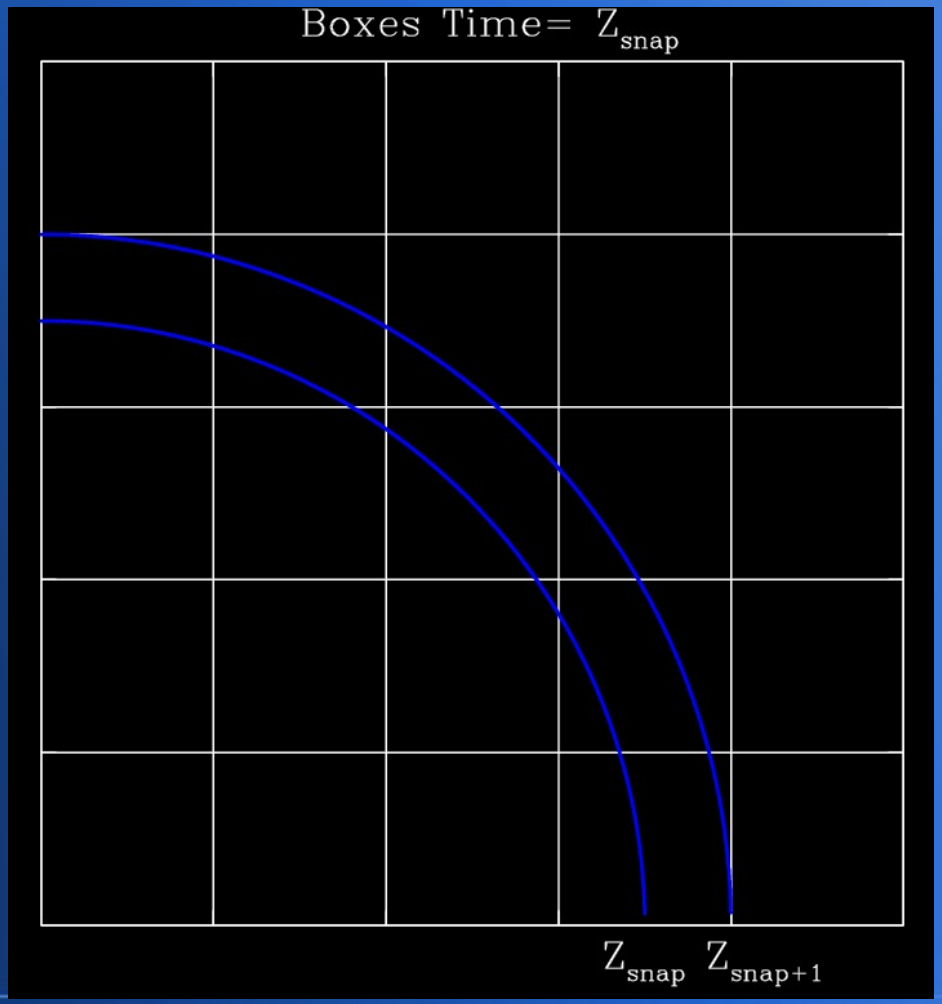
Extracting the strips from different times



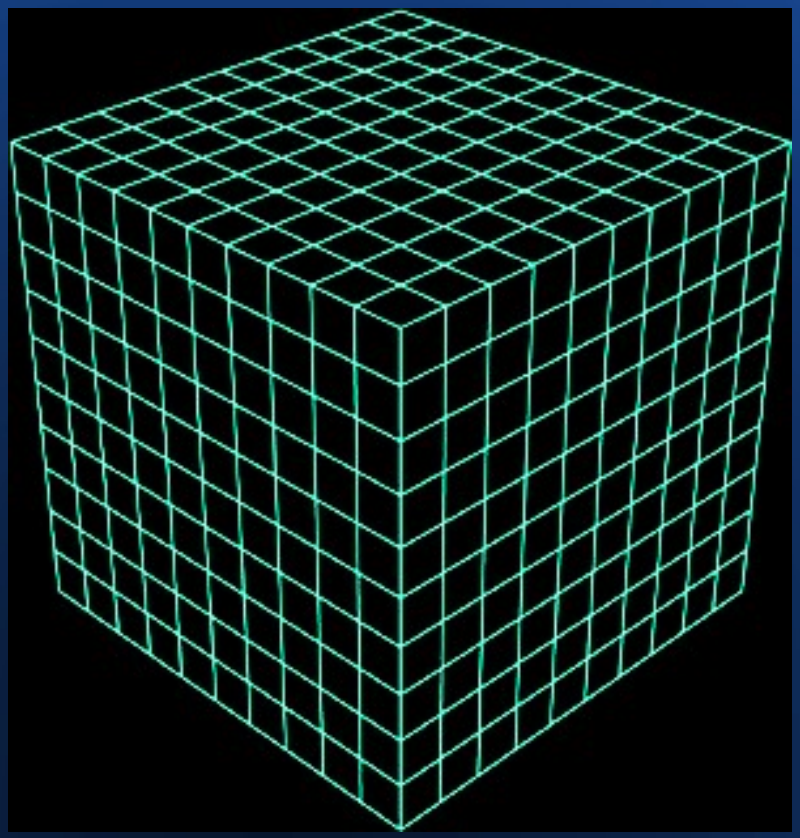


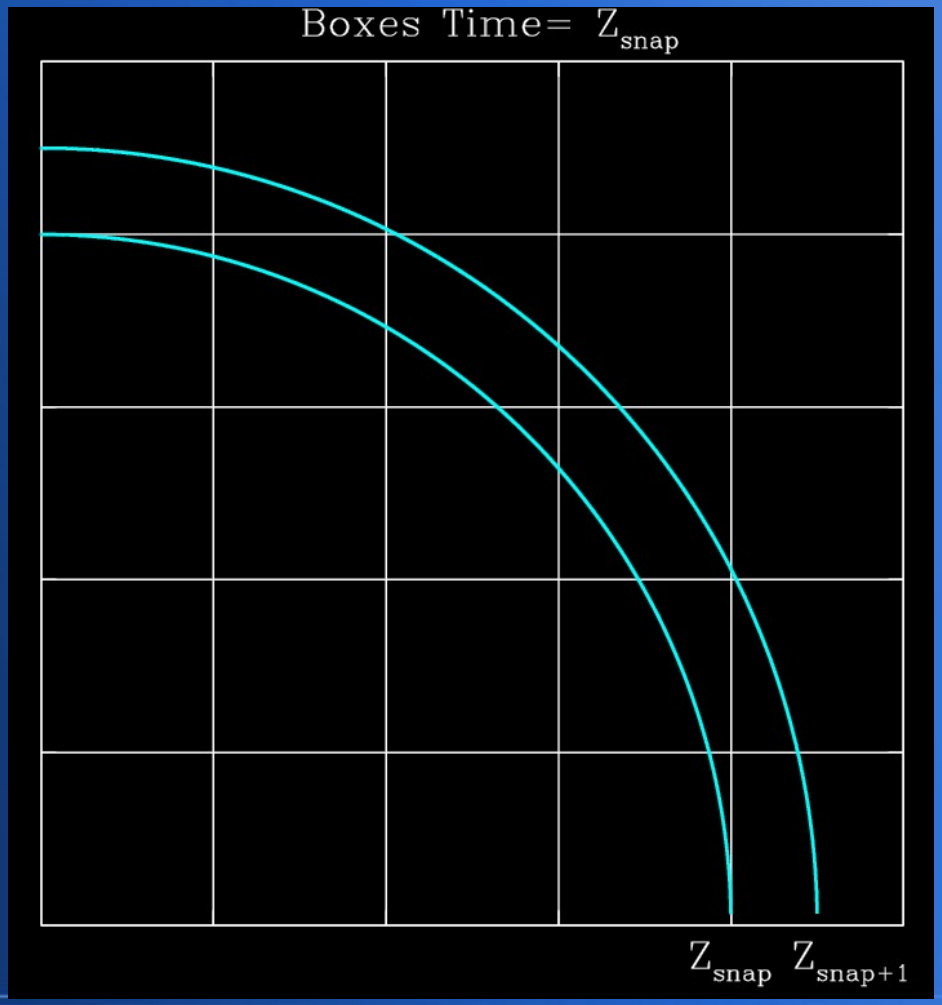
Extracting the strips from different times



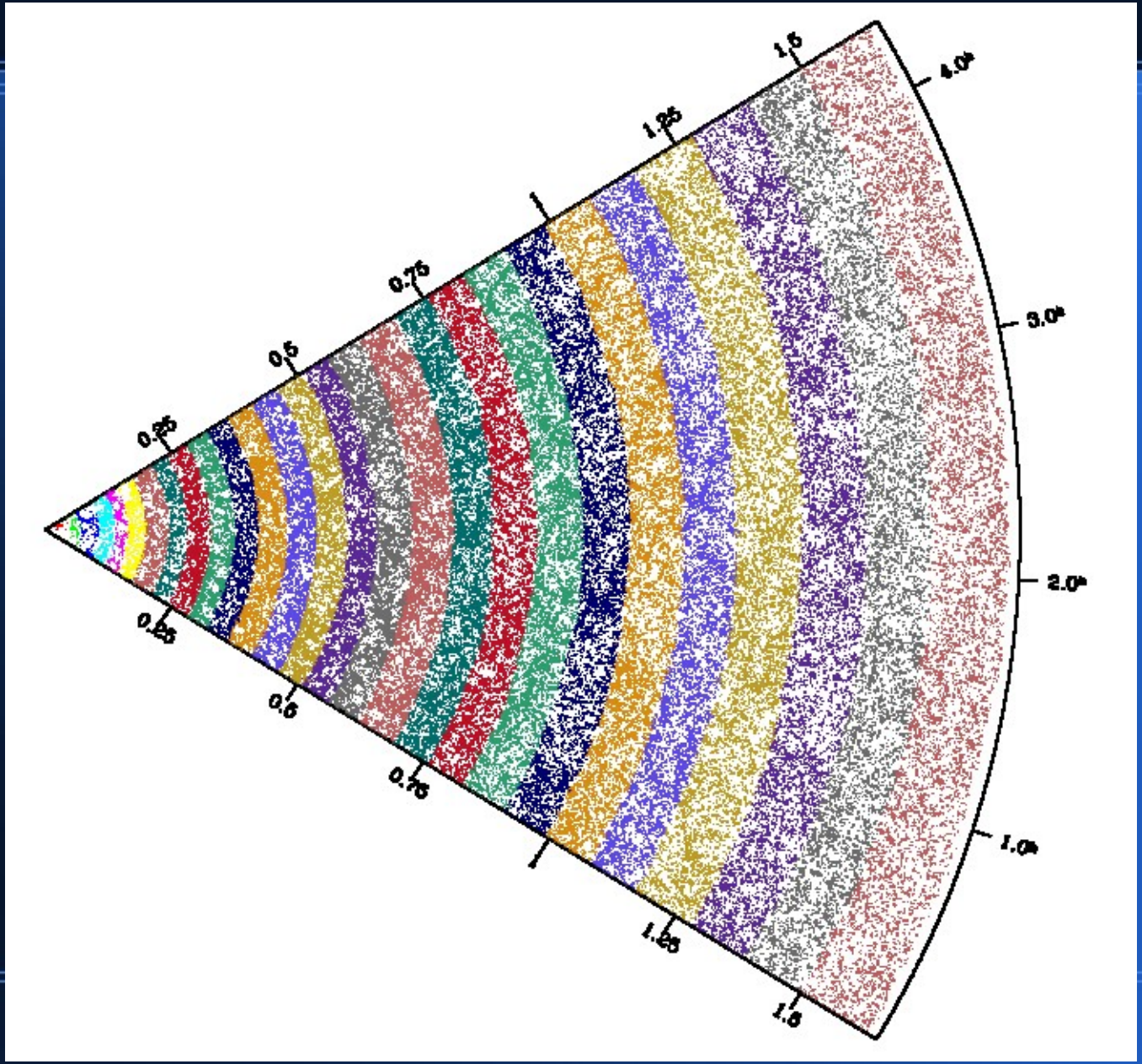


Extracting the strips from different times





Constructing the lightcone



How to overcome the discreteness of the simulation snapshots?

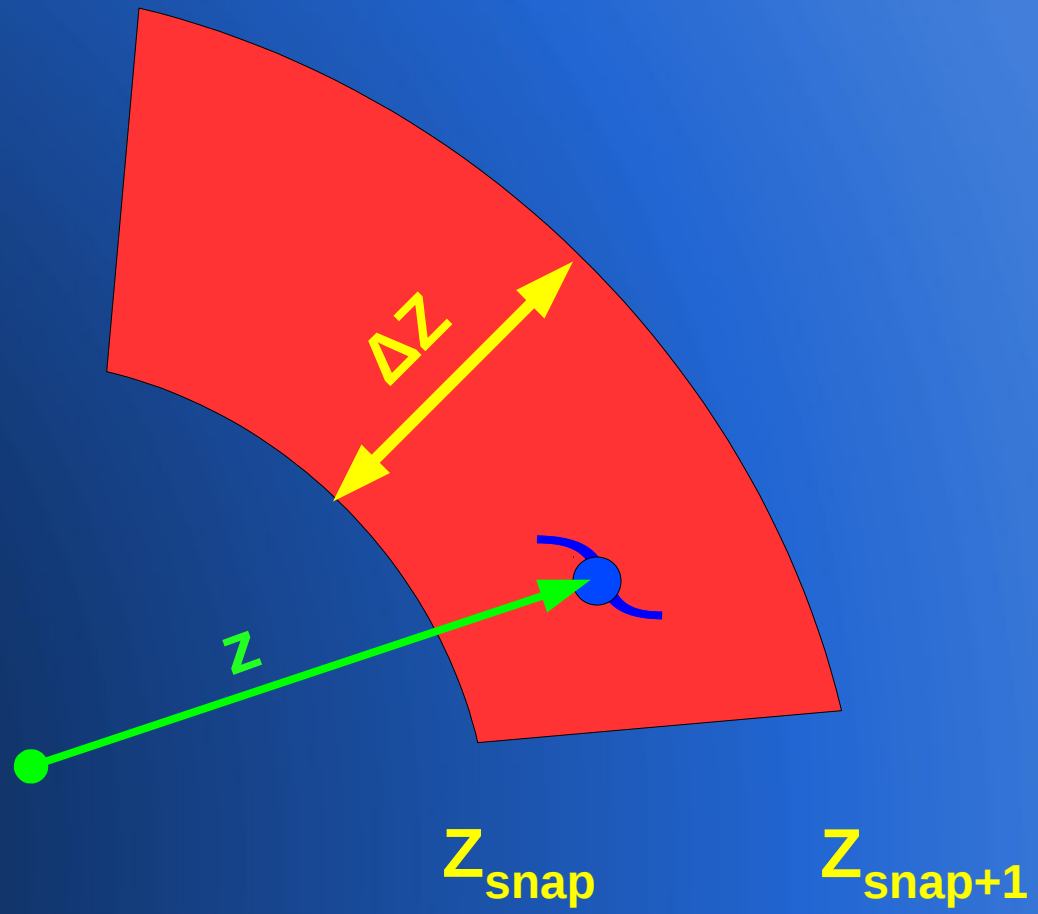
Interpolating positions, velocities and magnitudes:

$$\alpha = \frac{z - Z_{snap}}{Z_{snap+1} - Z_{snap}}$$

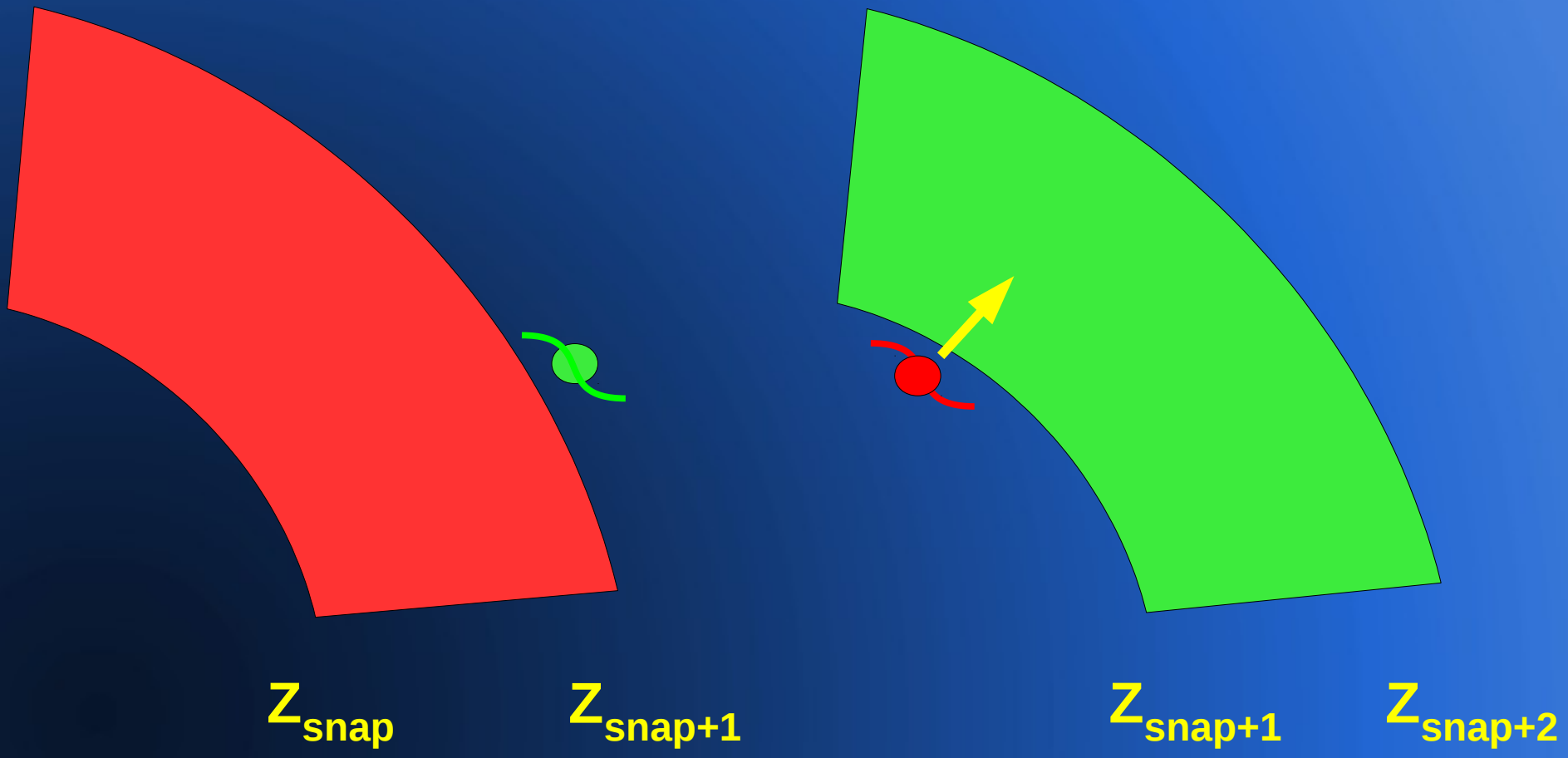
$$X = X_{snap} + \alpha\left(X_{snap+1} - X_{snap}\right)$$

$$V = V_{snap} + \alpha\left(V_{snap+1} - V_{snap}\right)$$

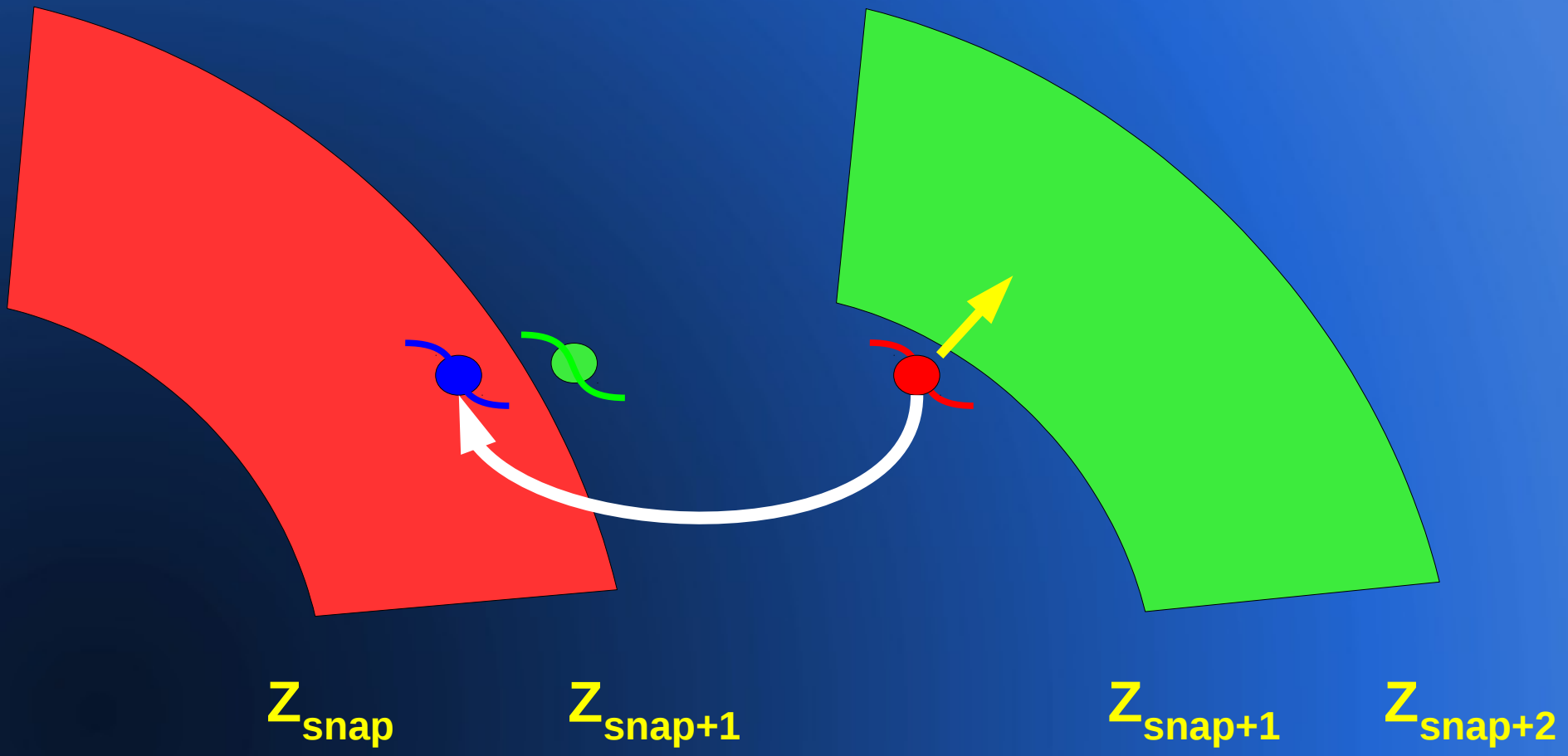
$$M = M_{snap} + \alpha\left(M_{snap+1} - M_{snap}\right)$$



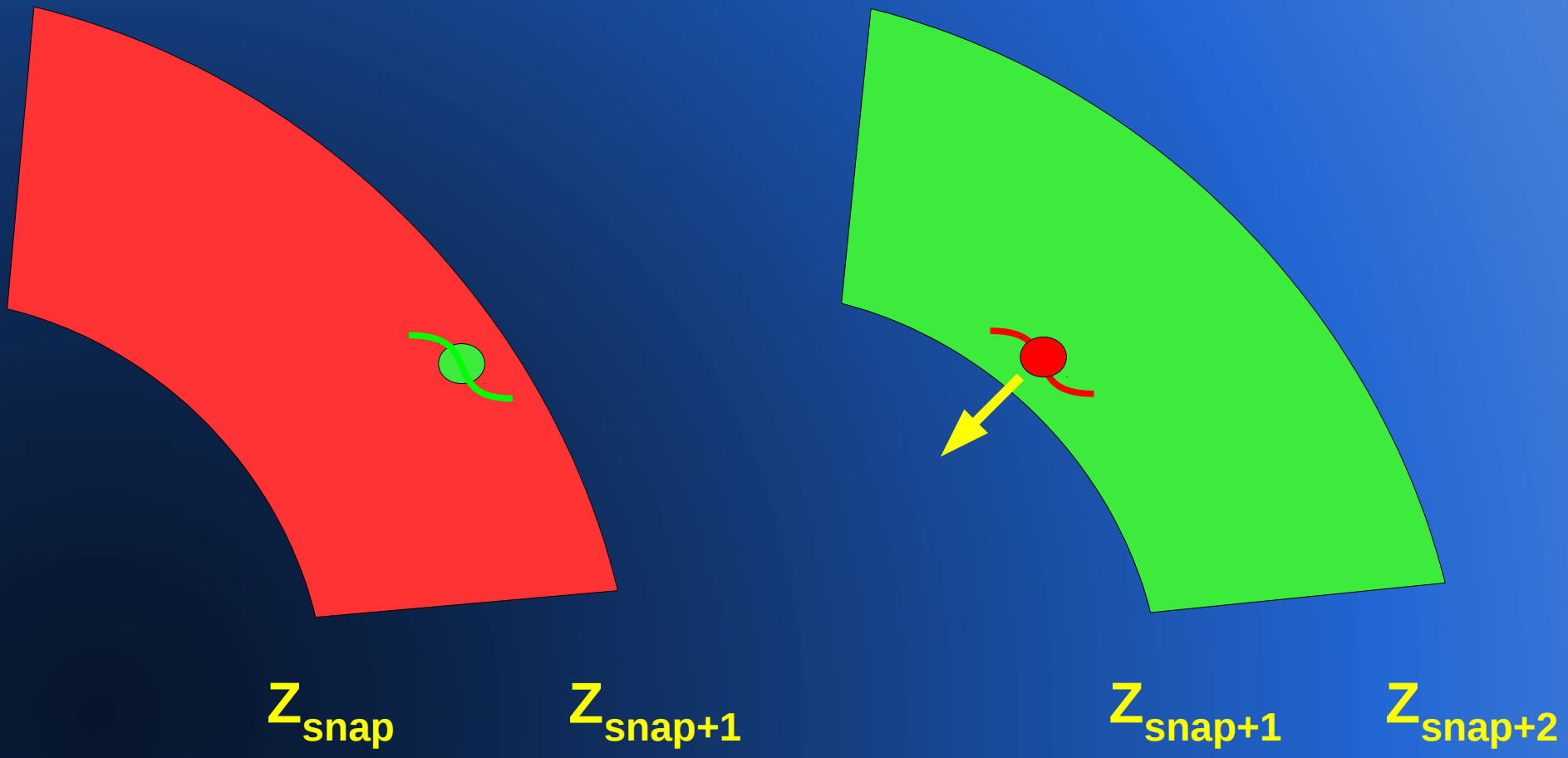
The problem of missing galaxies



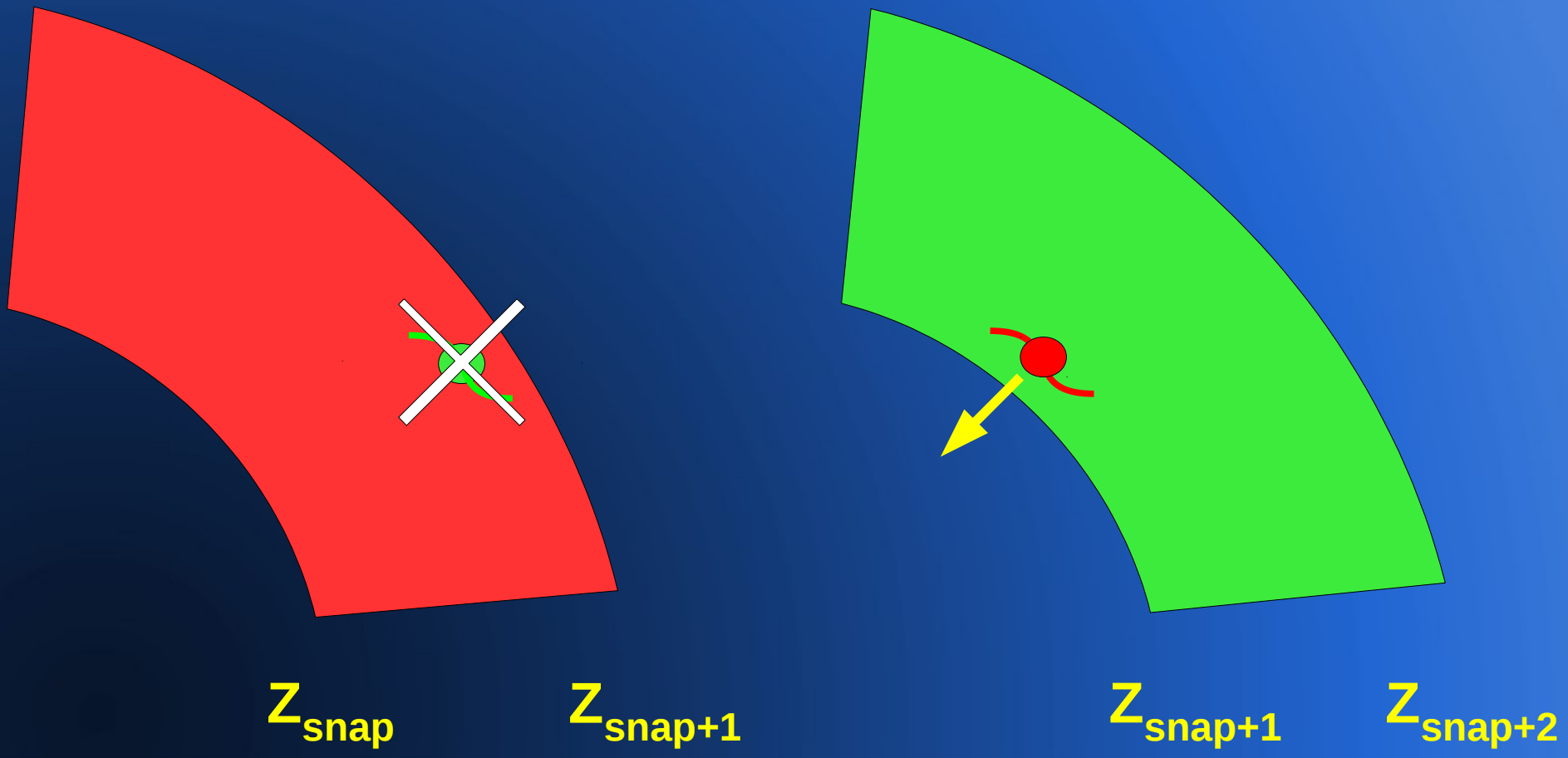
The problem of missing galaxies



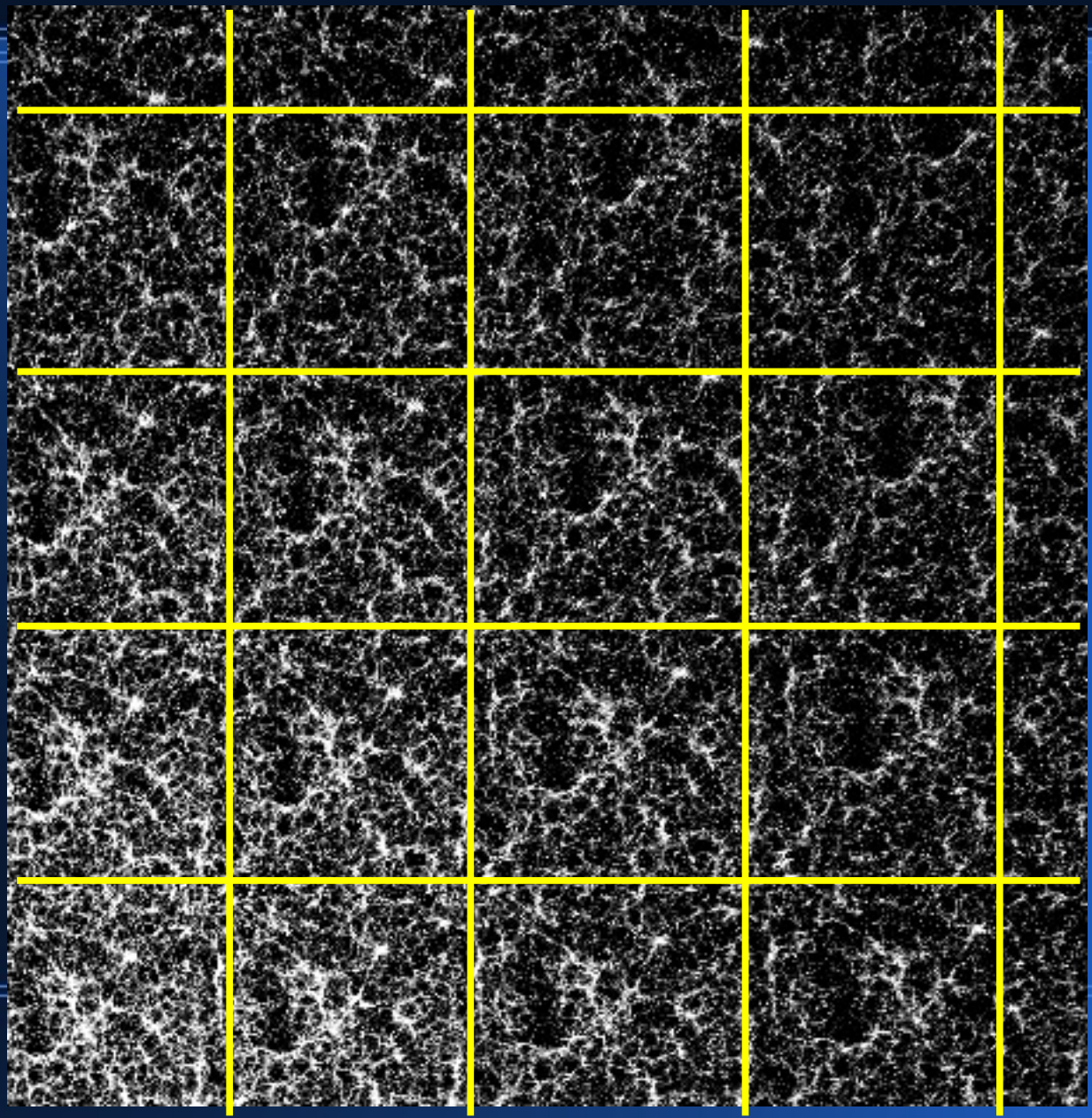
The problem of repeated galaxies



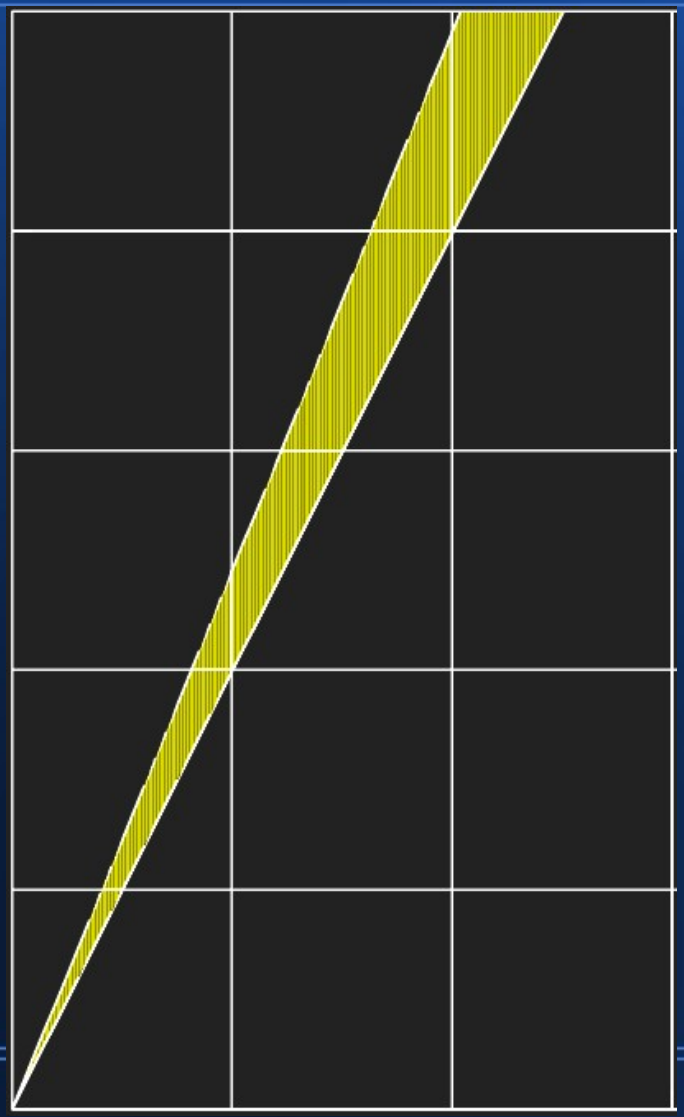
The problem of repeated galaxies

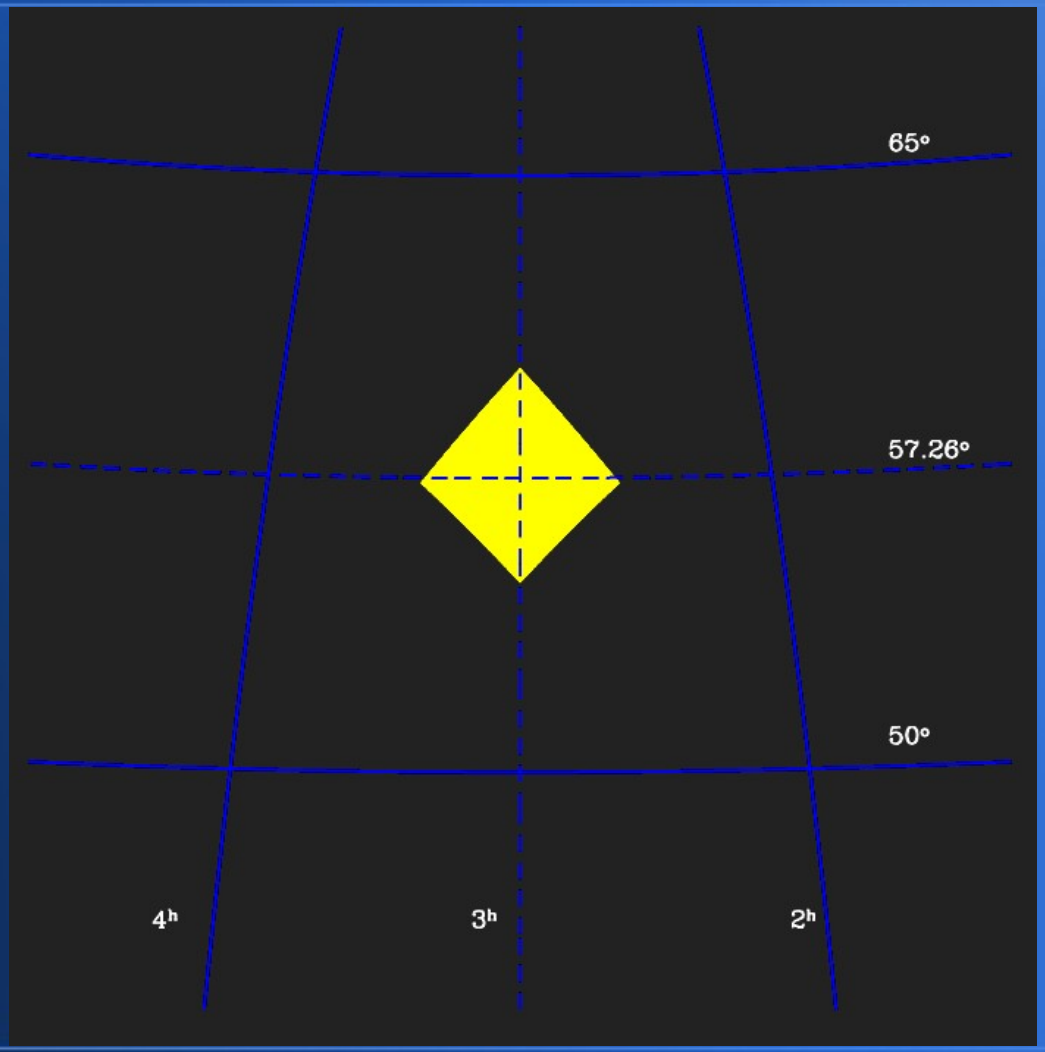


How to overcome the repeated large structures due to the small size of the simulation box?



How to overcome the repeated large structures due to the small size of the simulation box?



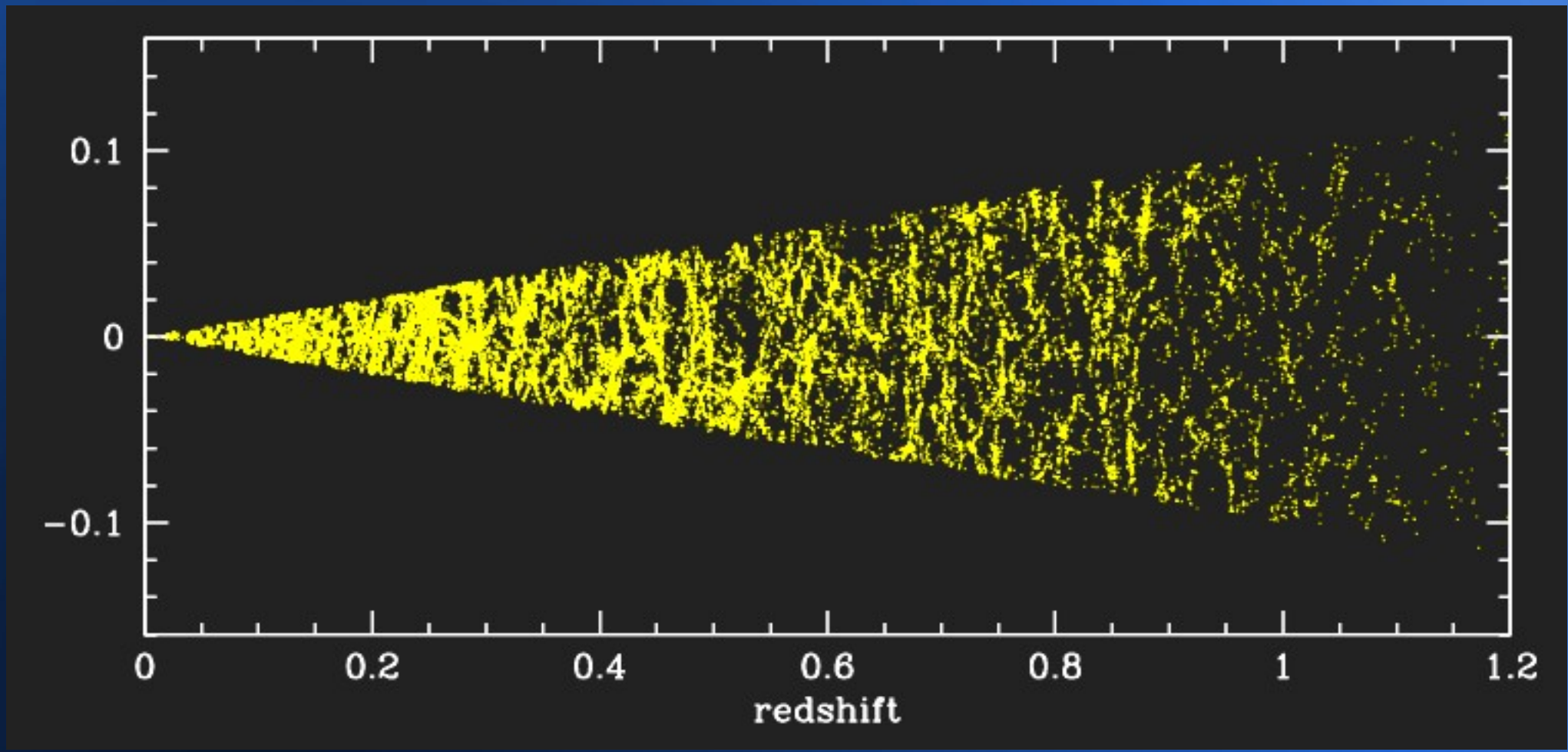


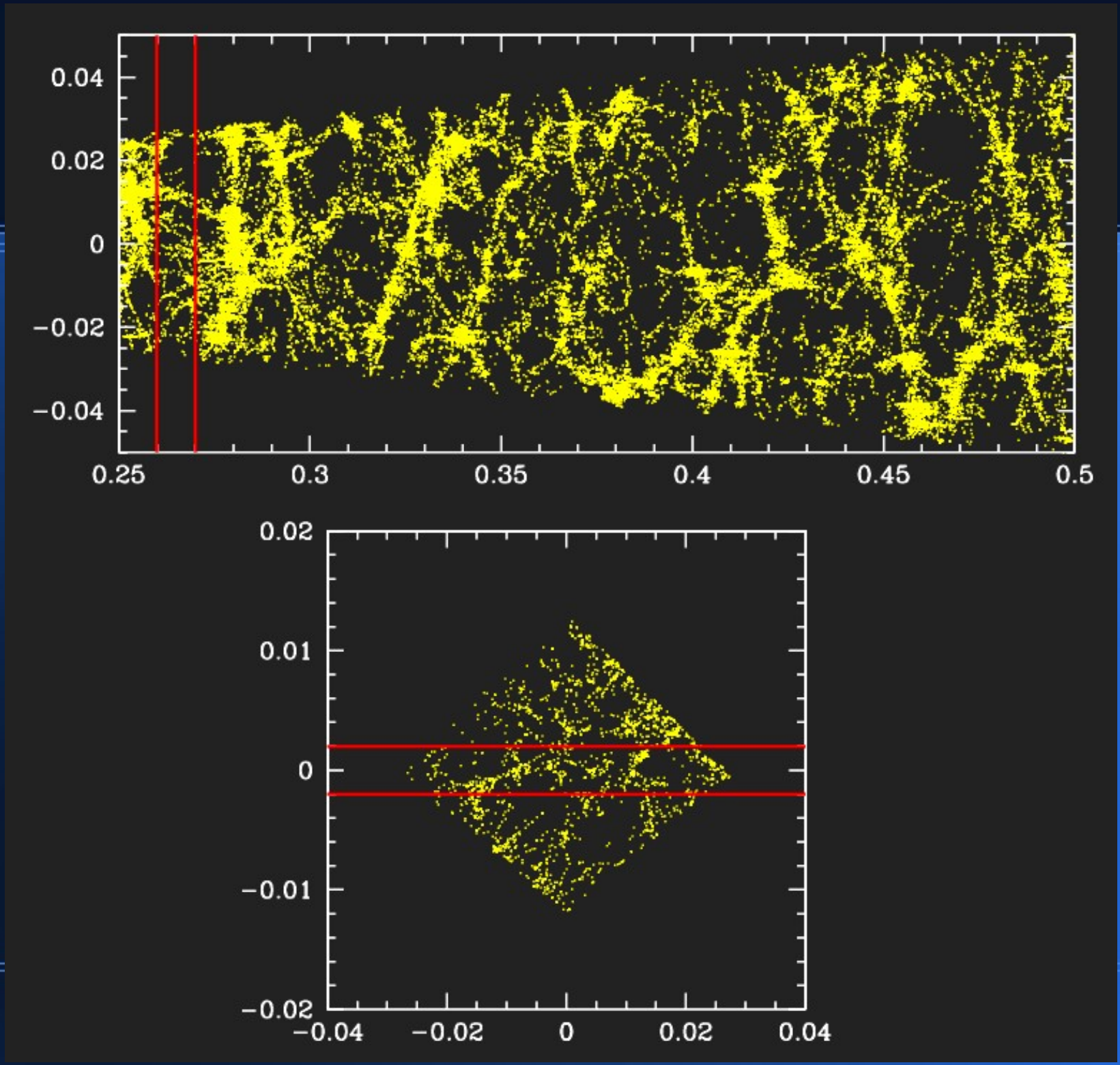
Kitzbichler & White (2007)

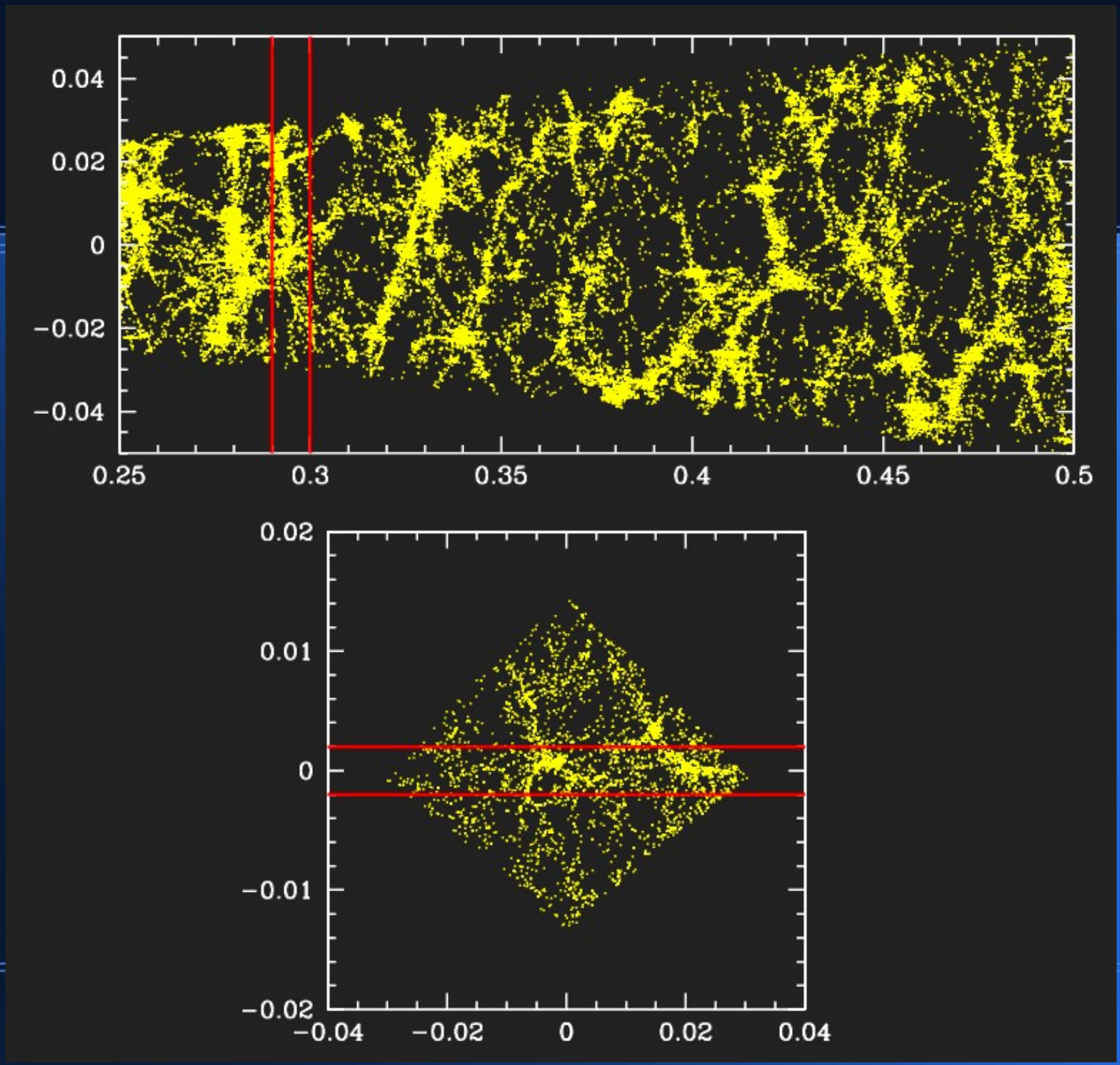
JPAS observational constraints

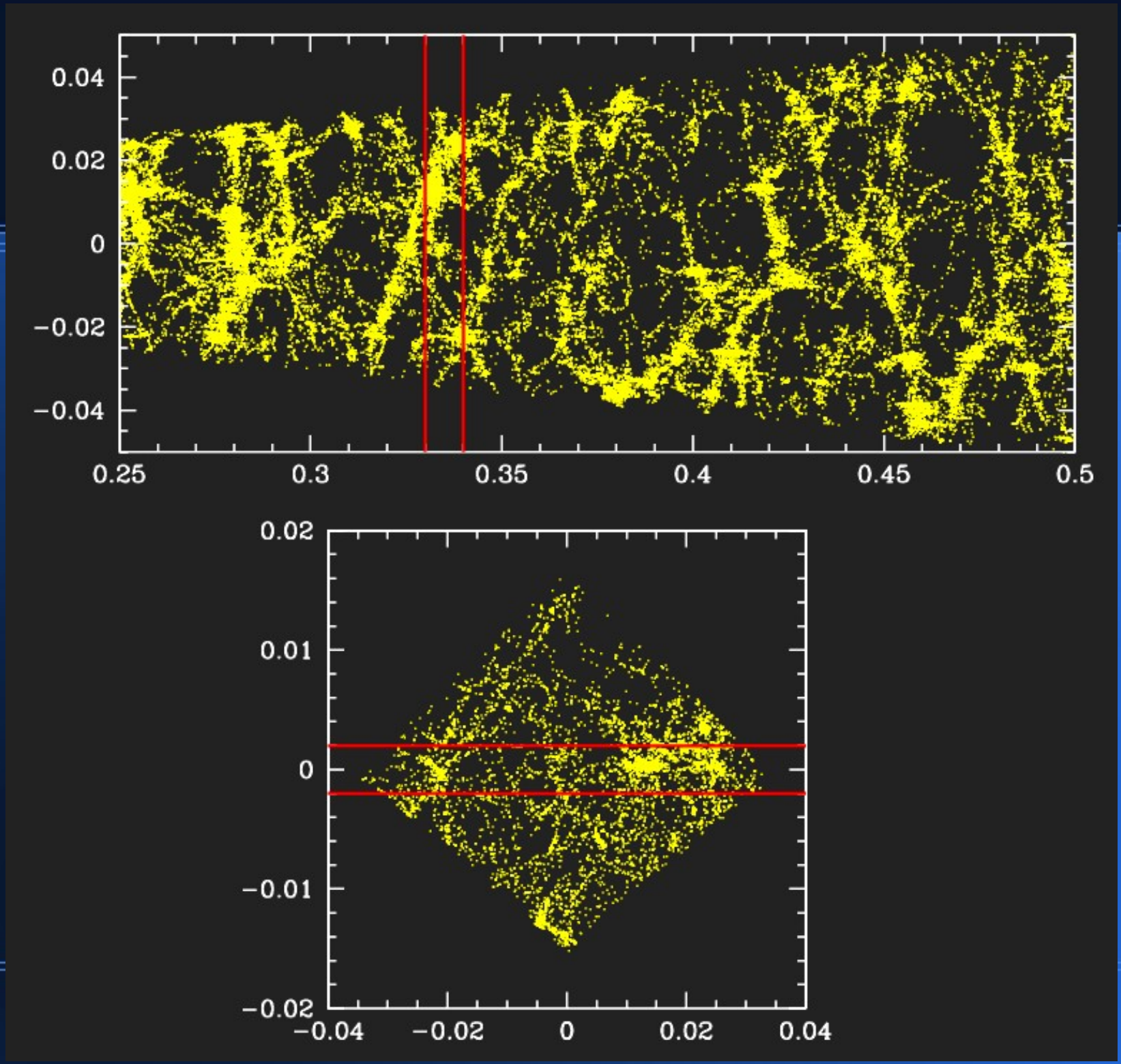
- JPAS will have an i broad band magnitude which will be quite similar to the i-band of the SDSS.
- The Guo et al. (2011) semianalytical model provides the five SDSS rest frame absolute magnitudes. We use i-sdss band to our lightcone.
- We tranform to observed apparent magnitudes using a mean K correction for the i band extracted from Poggianti et al. (1997).
- We include an apparent i band limit of 23, similar to the limit in the future JPAS.

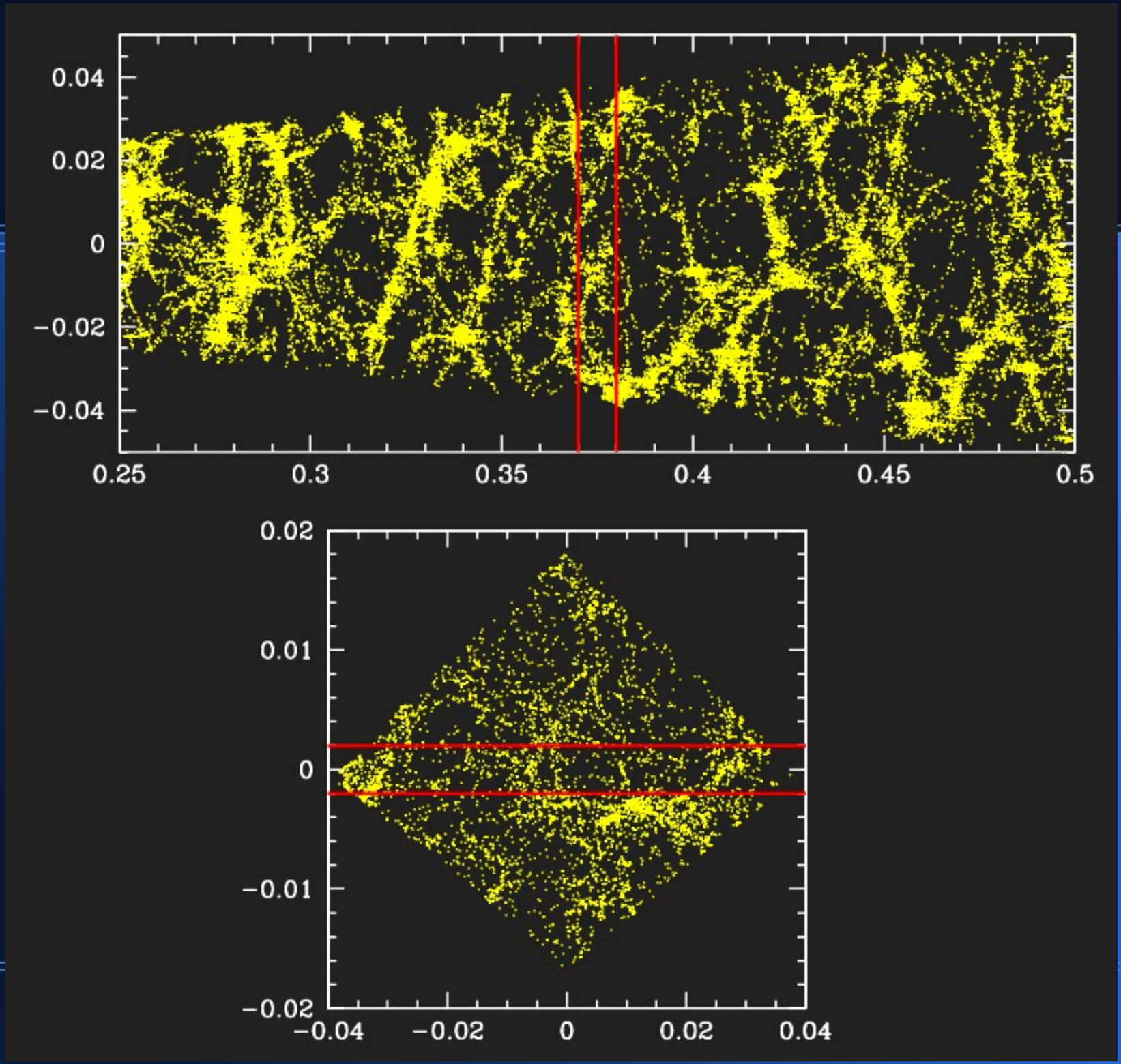
The Final Lightcone

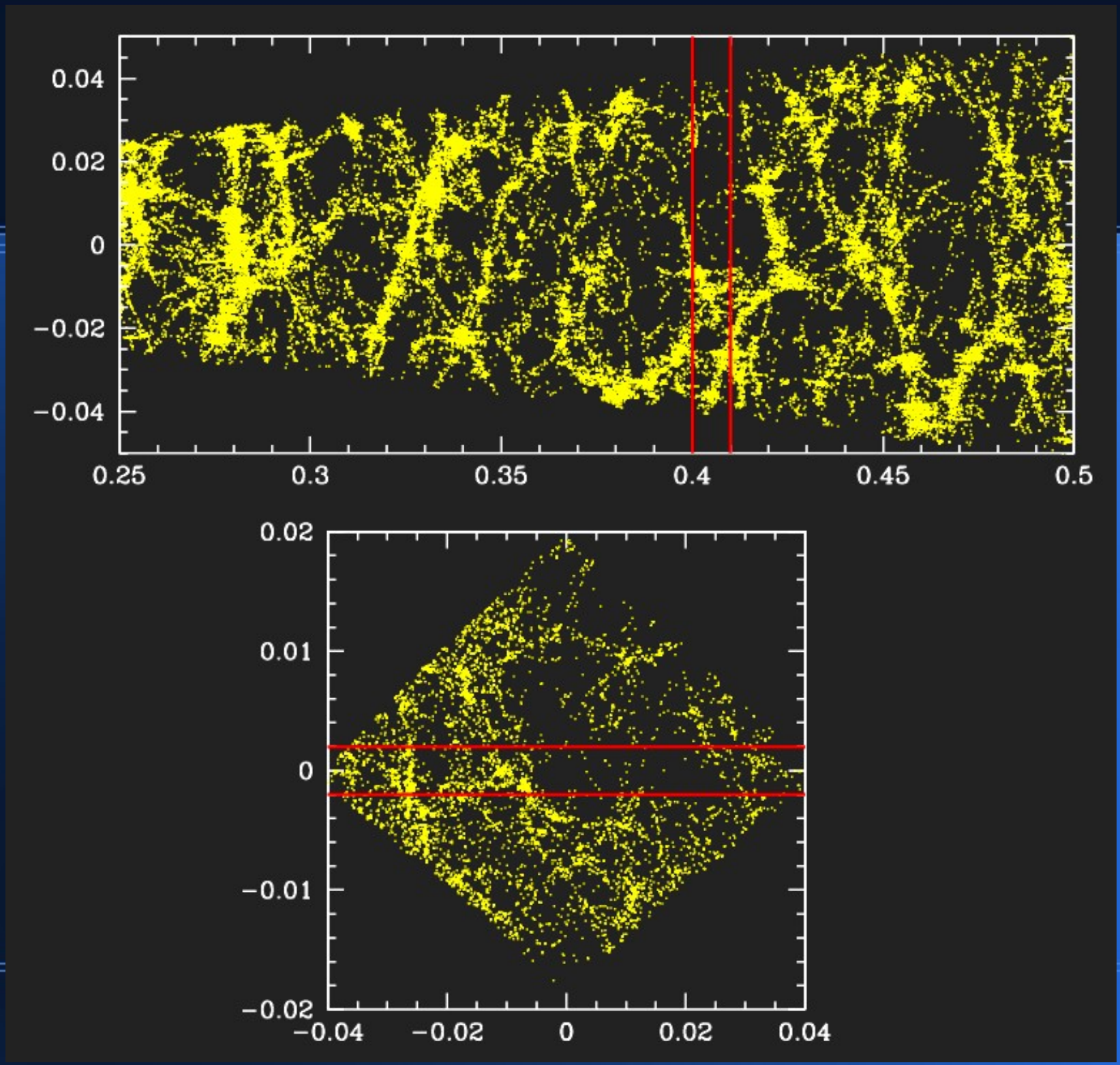


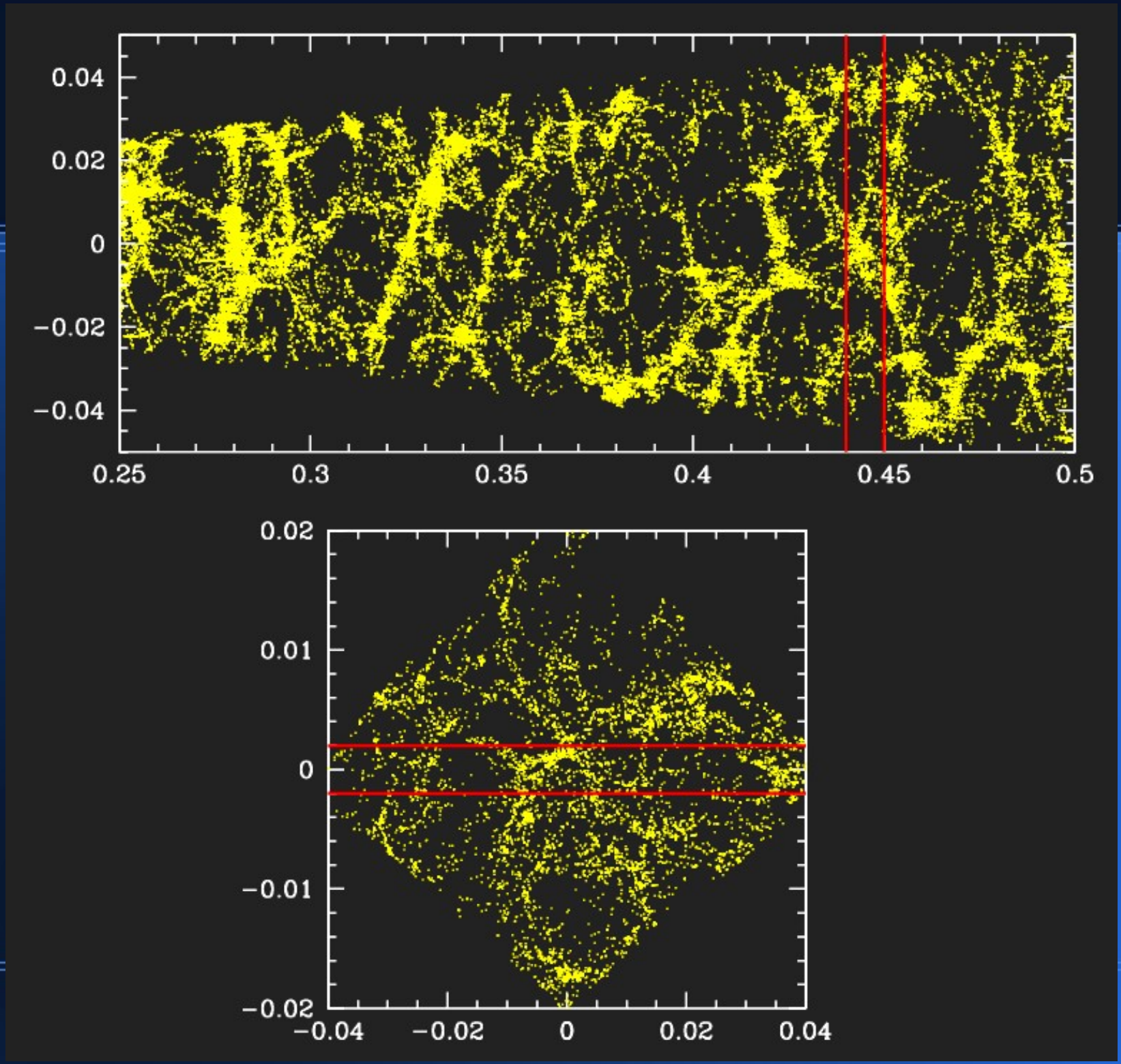

0.04
0.02
0
−0.02
−0.04
0.25
0.3
0.35
0.4
0.45
0.5
0.02
0.01
0
−0.01
−0.02
−0.04
−0.02
0
0.02
0.04

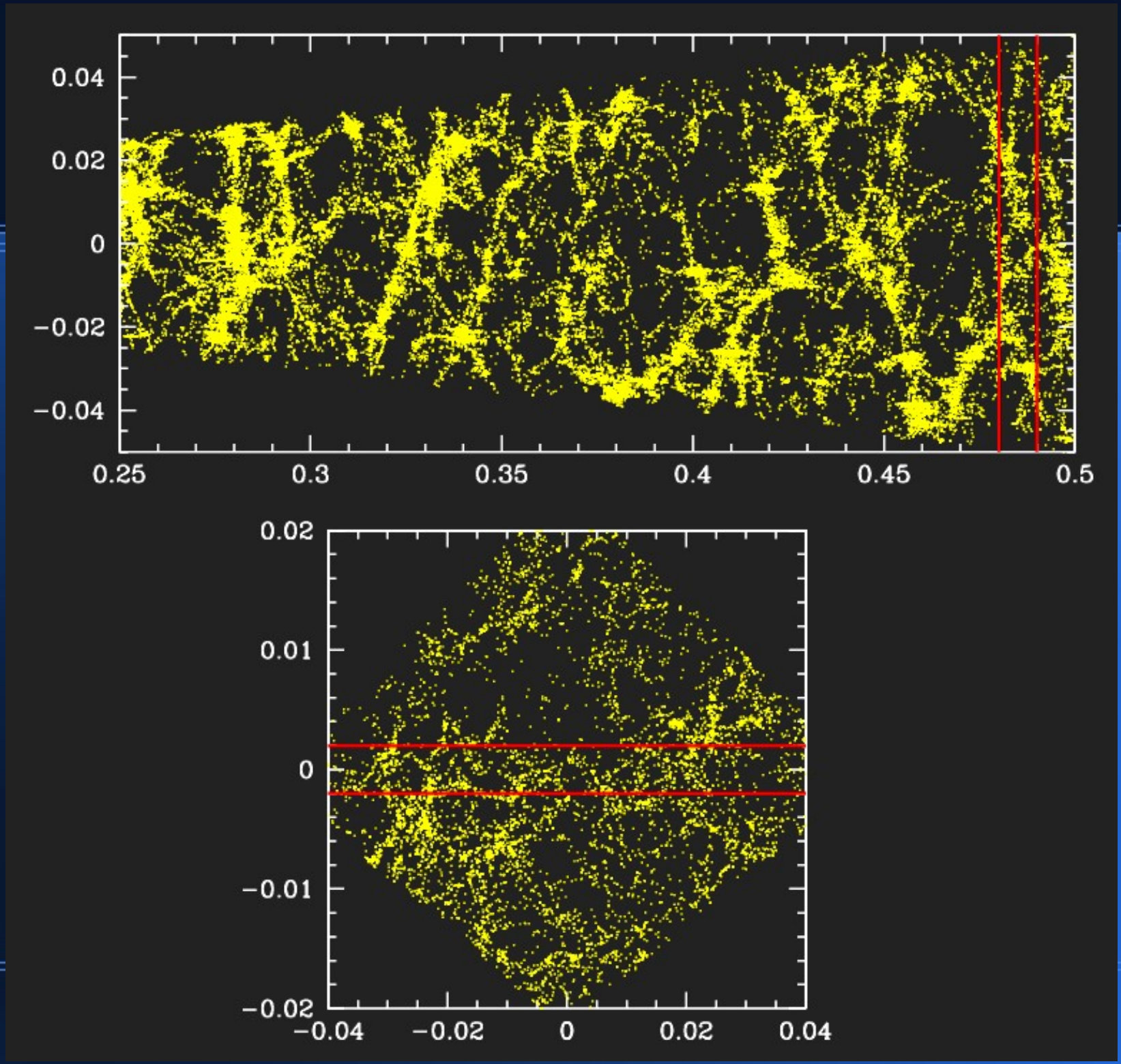

0.04
0.02
0
−0.02
−0.04
0.25
0.3
0.35
0.4
0.45
0.5
0.02
0.01
0
−0.01
−0.02
−0.04
−0.02
0
0.02
0.04


0.04
0.02
0
−0.02
−0.04
0.25
0.3
0.35
0.4
0.45
0.5
0.02
0.01
0
−0.01
−0.02
−0.04
−0.02
0
0.02
0.04


0.04
0.02
0
−0.02
−0.04
0.25
0.3
0.35
0.4
0.45
0.5
0.02
0.01
0
−0.01
−0.02
−0.04
−0.02
0
0.02
0.04


0.04
0.02
0
−0.02
−0.04
0.25
0.3
0.35
0.4
0.45
0.5
0.02
0.01
0
−0.01
−0.02
−0.04
−0.02
0
0.02
0.04


0.04
0.02
0
−0.02
−0.04
0.25
0.3
0.35
0.4
0.45
0.5
0.02
0.01
0
−0.01
−0.02
−0.04
−0.02
0
0.02
0.04


0.04
0.02
0
−0.02
−0.04
0.25
0.3
0.35
0.4
0.45
0.5
0.02
0.01
0
−0.01
−0.02
−0.04
−0.02
0
0.02
0.04

JPAS lightcone: Resutls

- Comprises ~500,000 galaxies
- Maximum redshift: 1.2
- Angular coverage: ~30 deg^2
- Absolute magnitude limit: -16.4 (i_{SDSS} band)
- Stellar mass limit: ~ 10^8 h^{-1} M_o

Friends of Friends algorithm

Huchra & Geller (1982) : Original FoF for observational spectroscopic catalogues.

Davis et al. (1985), Knebe et al. (2011) : Modified FoF for 3D simulations.

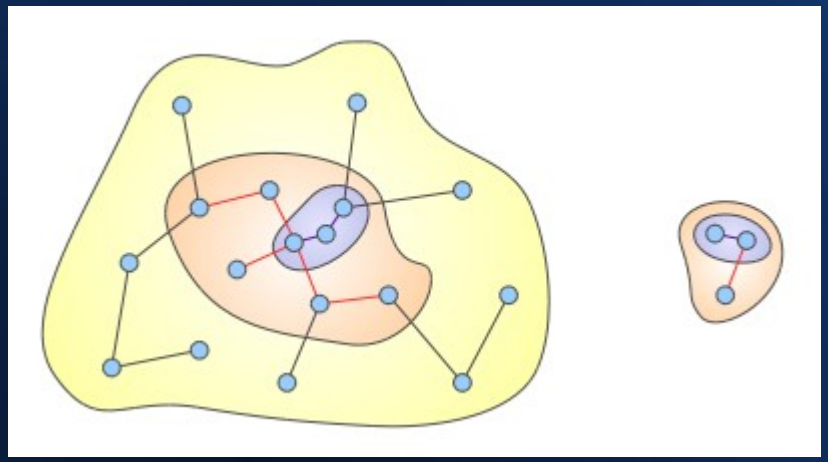
Merchán & Zandivarez (2002,2005): Modified FoF for 2dFGRS & SDSS.

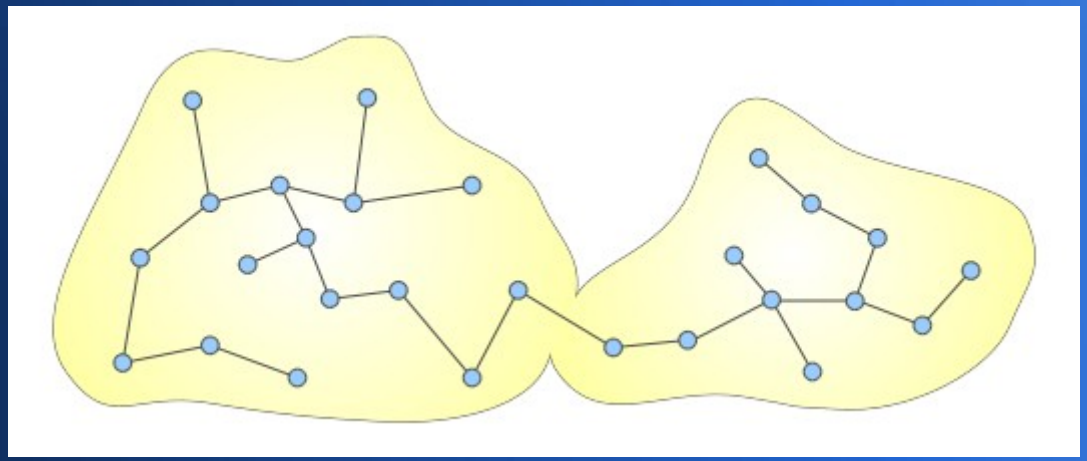
Botzler et al. (2004), Li & Yee (2008), Liu et al. (2008): Modified FoF for photometric catalogues.

BASIC IDEA : linking galaxies that share common neighbours (friends). It starts looking for the friends of an initial galaxy that have separations lower than a given threshold. The surroundings of each friend are examined until no more friends are found.

$$D_{ij} \leq D_l \text{ and } V_{ij} \leq V_l$$

D_l and V_l are called transversal and radial linking lengths, respectively.





Ideal group sample

NO flux limit

NO redshift distortion

GROUP REFERENCE SAMPLE : groups having 4 or more members are identified on the volume limited sample in real space. The linking lengths are specified only by the cosmological model that determines the overdensity of virialization, and the space density of galaxies, both as a function of redshift.

$$D_l = D_0(z) \text{ and } V_l = D_l\, H(z)$$

Flux limit

NO redshift distortion

GROUP RESTRICTED-REFERENCE SAMPLE: is the reference sample only with those galaxies with apparent magnitude <23. This is the *ideal* sample that should be identified in a flux limited catalogue.

In observational samples, there are two main constraints that the algorithm has to take into account:

- **FLUX LIMIT**
- **REDSHIFT SPACE DISTORTIONS**

The linking lengths have to be modified!

WHICH ARE THE PROPER LINKING LENGTHS, D_l and V_l, TO RECOVER THE IDEAL GROUPS WHEN IDENTIFYING IN OBSERVATIONAL CATALOGUES ?

Tuning the FoF linking lengths

Flux limit

NO redshift distortion

FLUX LIMITED SAMPLE : The linking lengths have to consider the diminishing of the galaxy number density due to the missing faint galaxies. Therefore, the linking lengths depend on the galaxy luminosity function.

$$D_l = D_0(z)\, R_s \text{ and } V_l = D_l\, H(z)$$

We tested the algorithm against a fixed luminosity function for nearby galaxies LF(0), or a variable LF(z). Then, $R_s(0)$ or $R_s(z)$, respectively

NO Flux limit

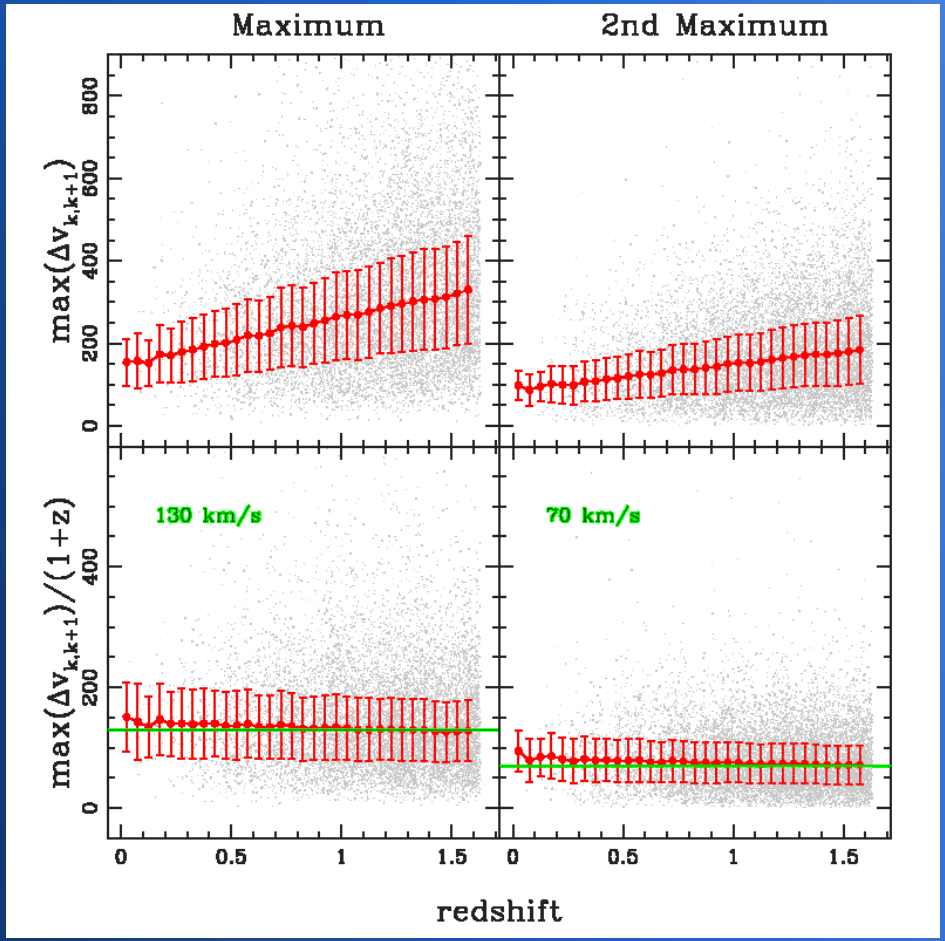
Redshift distortion

REDSHIFT SAMPLE : The radial linking length has to consider the redshift distortion due to peculiar velocities.

$D_l=D_0(z)$ and $V_l=V_0$

We examined the radial distortions of groups in the REFERENCE sample, to look for the most suitable value of V_0 to link all the galaxies in the group.

We analized the maximum radial velocity difference between a galaxy and its closest neighbour, and the second maximum per group.



NO Flux limit

Redshift distortion

REDSHIFT SAMPLE : The radial linking length has to consider the redshift distortion due to peculiar velocities.

$$D_l = D_0(z) \text{ and } V_l = V_0$$

We tested the algorithm against four different recipes for V_0 [km/s]

$$V_0 = 130$$

$$V_0 = 130(1+z)$$

$$V_0 = 70$$

$$V_0 = 70(1+z)$$

Resulting Group Samples

Sample	Linking lengths		Total number of glxs in groups	Groups with $4 \leqslant N < 10$	Groups with $N \geqslant 10$
reference	$D_l = D_0(z)$	$V_l = D_l\, H(z)$	1,825,303	159,258	41,774
restricted-reference			99,230	8,986	2,308
flux limited-LF variable	$D_l = D_0(z)\, R_s(z)$	$V_l = D_l\, H(z)$	107,768	10,255	2,446
flux limited-LF fixed	$D_l = D_0(z)\, R_s(0)$	$V_l = D_l\, H(z)$	116,777	10,923	2,675
Redshift	$D_l = D_0(z)$	$V_l = 130$	1,287,097	160,145	23,572
	$D_l = D_0(z)$	$V_l = 130(1+z)$	2,133,189	203,975	46,557
	$D_l = D_0(z)$	$V_l = 70$	629,841	98,537	8,383
	$D_l = D_0(z)$	$V_l = 70(1+z)$	1,394,091	170,918	26,372
sp-mock catalogue	$D_l = D_0(z)\, R_s(z)$	$V_l = 130(1+z)\, R_s(z)$	120,518	12,694	2,574

Purity & Completeness

PURITY : how good are the identified groups compared to a reference sample?

COMPLETENESS : how many of the original reference groups are recovered in the new identification?

Different categories:

P1: 100% Pure
P2: >70% Pure, <30% interlopers
P3: mostly mergers + few interlopers
P4: <70% Pure, >30% interlopers
P5: <70% mergers + >30% interlopers
P6: 100% False

C1: 100% recovered
C2: >70% recovered, <30% missing glxs
C3: split + few missing glxs
C4: <70% recovered, >30% missing glxs
C5: <70% split + >30% missing glxs
C6: 100% missing glxs

Purity & Completeness

Class	Flux limited		Redshift				Sp-mock
	LF variable	LF fixed	$V_0 = 130$	$V_0 = 130(1+z)$	$V_0 = 70$	$V_0 = 70(1+z)$	$V_0 = 130(1+z)$
P1	62	51	42	35	49	42	28
P2	21	24	21	21	20	21	23
P3	1	2	6	5	8	6	3
P4	3	5	12	12	11	12	13
P5	0	0	2	1	1	1	1
P6	13	18	17	26	11	18	32
P1+P2	83	75	63	56	69	63	51
C1	82	89	14	48	3	17	42
C2	12	8	16	25	5	18	27
C3	2	2	6	8	2	7	8
C4	1	0	21	8	19	21	10
C5	0	0	6	1	7	5	1
C6	3	1	37	10	64	32	12
C1+C2	94	97	30	73	8	35	69

Purity & Completeness

Class	Flux limited		Redshift				Sp-mock
	LF variable	LF fixed	$V_0 = 130$	$V_0 = 130(1+z)$	$V_0 = 70$	$V_0 = 70(1+z)$	$V_0 = 130(1+z)$
P1	62	51	42	35	49	42	28
P2	21	24	21	21	20	21	23
P3	1	2	6	5	8	6	3
P4	3	5	12	12	11	12	13
P5	0	0	2	1	1	1	1
P6	13	18	17	26	11	18	32
P1+P2	83	75	63	56	69	63	51
C1	82	89	14	48	3	17	42
C2	12	8	16	25	5	18	27
C3	2	2	6	8	2	7	8
C4	1	0	21	8	19	21	10
C5	0	0	6	1	7	5	1
C6	3	1	37	10	64	32	12
C1+C2	94	97	30	73	8	35	69

Purity & Completeness

Class	Flux limited		Redshift				Sp-mock
	LF variable	LF fixed	$V_0 = 130$	$V_0 = 130(1+z)$	$V_0 = 70$	$V_0 = 70(1+z)$	$V_0 = 130(1+z)$
P1	62	51	42	35	49	42	28
P2	21	24	21	21	20	21	23
P3	1	2	6	5	8	6	3
P4	3	5	12	12	11	12	13
P5	0	0	2	1	1	1	1
P6	13	18	17	26	11	18	32
P1+P2	83	75	63	56	69	63	51
C1	82	89	14	48	3	17	42
C2	12	8	16	25	5	18	27
C3	2	2	6	8	2	7	8
C4	1	0	21	8	19	21	10
C5	0	0	6	1	7	5	1
C6	3	1	37	10	64	32	12
C1+C2	94	97	30	73	8	35	69

Group Spectroscopic Sample

Flux limit

Redshift distortion

MOCK SP-SAMPLE : based on a compromise between purity and completeness, the final linking lengths are

$$D_l = D_0(z) R_s(z) \quad \text{and} \quad V_l = 130(1+z) R_s(z)$$

Sample	Linking lengths		Total number of glxs in groups	Groups with $4 \leqslant N < 10$	Groups with $N \geqslant 10$
reference	$D_l = D_0(z)$	$V_l = D_l\, H(z)$	1,825,303	159,258	41,774
restricted-reference			99,230	8,986	2,308
flux limited-LF variable	$D_l = D_0(z)\, R_s(z)$	$V_l = D_l\, H(z)$	107,768	10,255	2,446
flux limited-LF fixed	$D_l = D_0(z)\, R_s(0)$	$V_l = D_l\, H(z)$	116,777	10,923	2,675
Redshift	$D_l = D_0(z)$	$V_l = 130$	1,287,097	160,145	23,572
	$D_l = D_0(z)$	$V_l = 130(1+z)$	2,133,189	203,975	46,557
	$D_l = D_0(z)$	$V_l = 70$	629,841	98,537	8,383
	$D_l = D_0(z)$	$V_l = 70(1+z)$	1,394,091	170,918	26,372
sp-mock catalogue	$D_l = D_0(z)\, R_s(z)$	$V_l = 130(1+z)\, R_s(z)$	120,518	12,694	2,574

						$70(1+z)$	Sp-mock $V_0 = 130(1+z)$
						42	28
P2	21	24	21	21	20	21	23
P3	1	2	6	5	8	6	3
P4	3	5	12	12	11	12	13
P5	0	0	2	1	1	1	1
P6	13	18	17	26	11	18	32
P1+P2	83	75	63	56	69	63	51
C1	82	89	14	48	3	17	42
C2	12	8	16	25	5	18	27
C3	2	2	6	8	2	7	8
C4	1	0	21	8	19	21	10
C5	0	0	6	1	7	5	1
C6	3	1	37	10	64	32	12
C1+C2	94	97	30	73	8	35	69

Identification of groups using photometric redshits

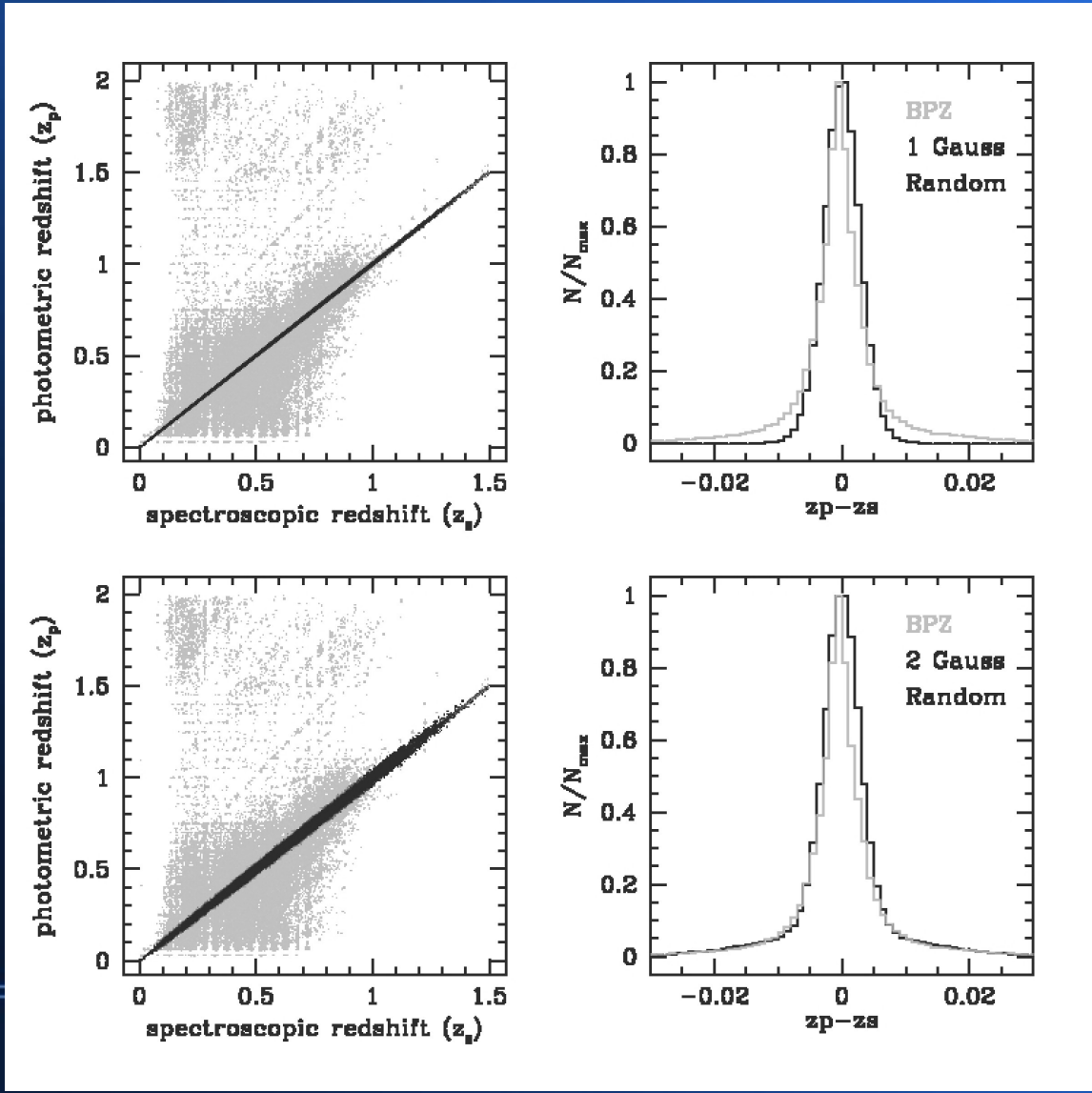
$$P(|V_2 - V_1| \leqslant V_l) \equiv \int_0^{\infty} dz\, G_1(z) \int_{z-V_l}^{z+V_l} dz'\, G_2(z')$$

$$P(|V_2 - V_1| \leqslant V_l) > P_{th}$$

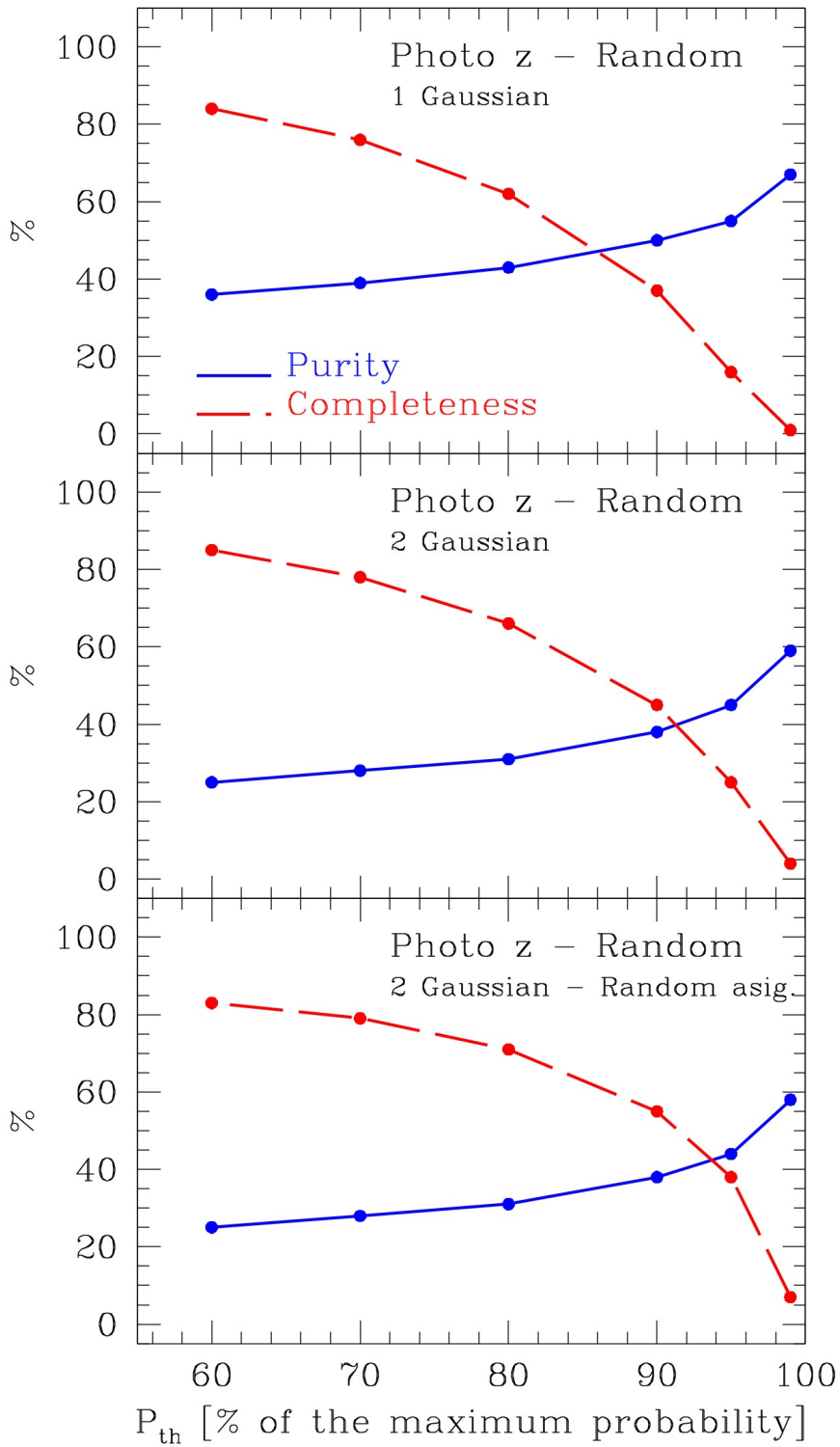
$$G_i(z) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(\frac{-(z - z_i)^2}{2{\sigma_i}^2}\right)$$

Liu et al. 2088

Identification of groups using photometric redshits



Purity and Completeness using random photo-z



Purity and Completeness using BPZ photo-z