SKA Data Challenge 3

Foreground removal update from SKACH team

SKACH

Michele Bianco
Tianyue Chen
Shreyam Krishna
Arpan Das

Rohit Sharma

Sambit Kumar Giri

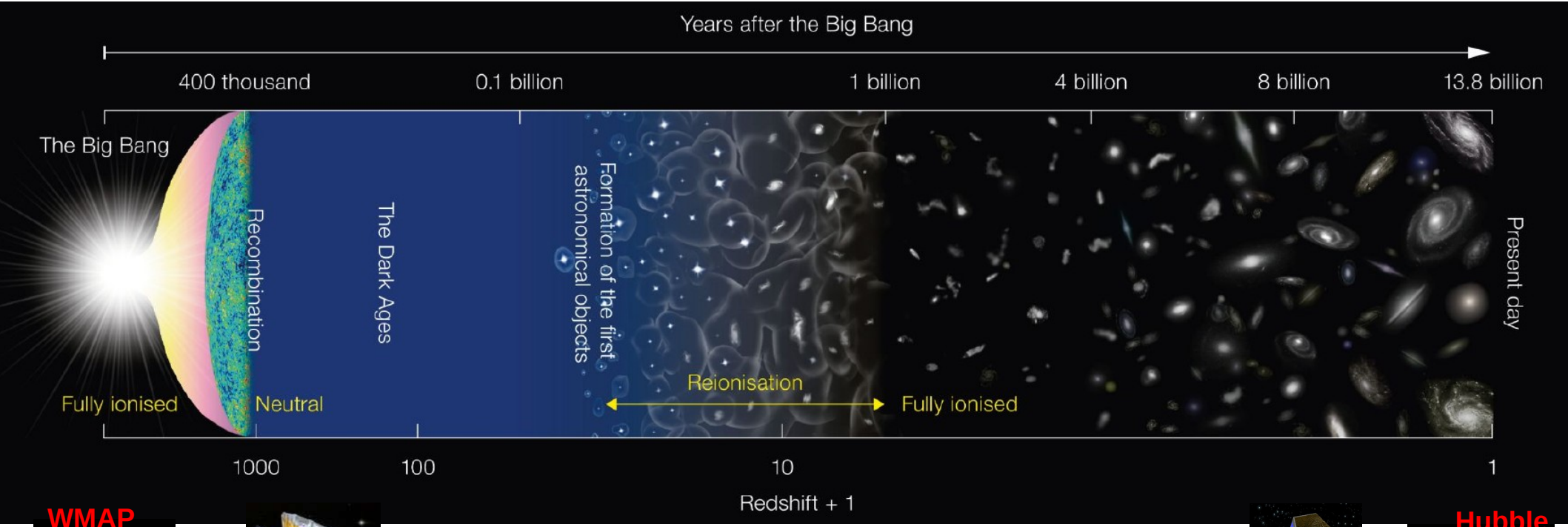
Philipp Denzel

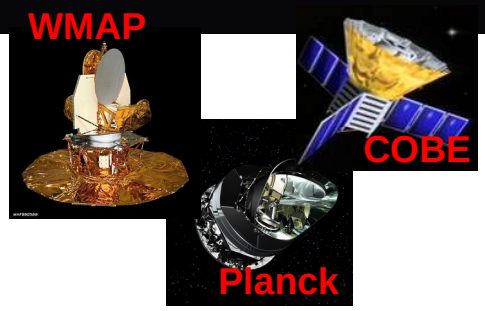
Hatem Ghorbel
Edouard Goffinet
Massimo De Sanctis

Viraj Nistane
Christopher Finlay

Andrei Mesinger
David Prelogović

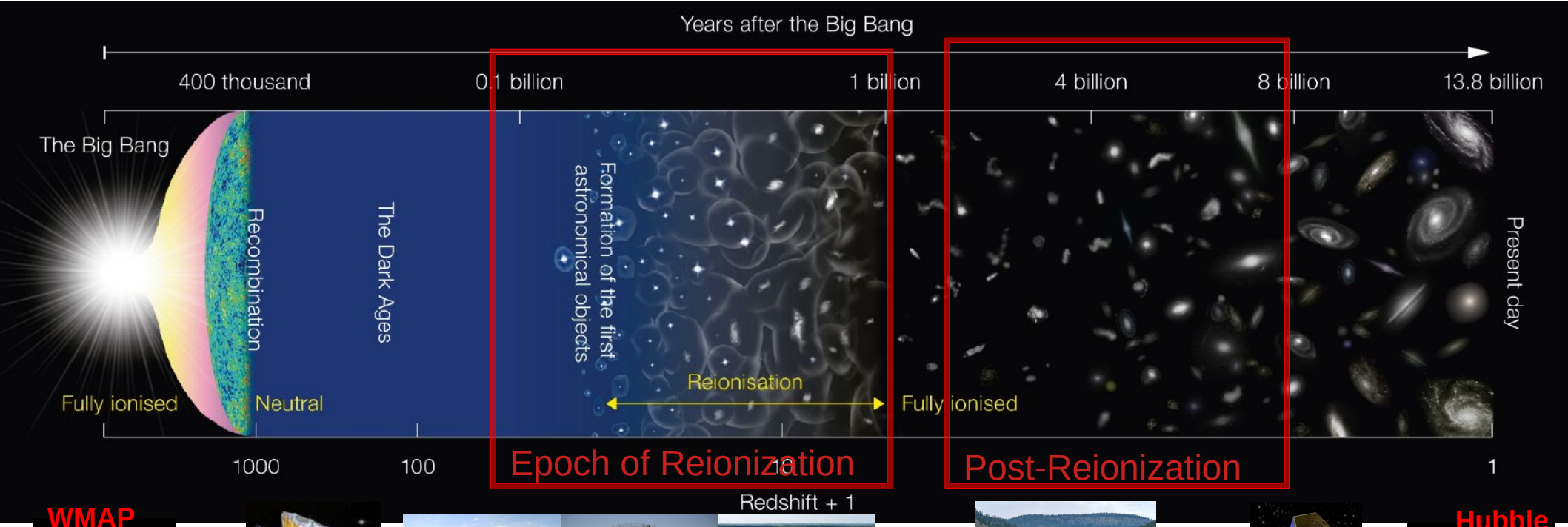
Cosmology with Radio Interferometric Experiments

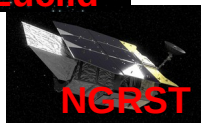




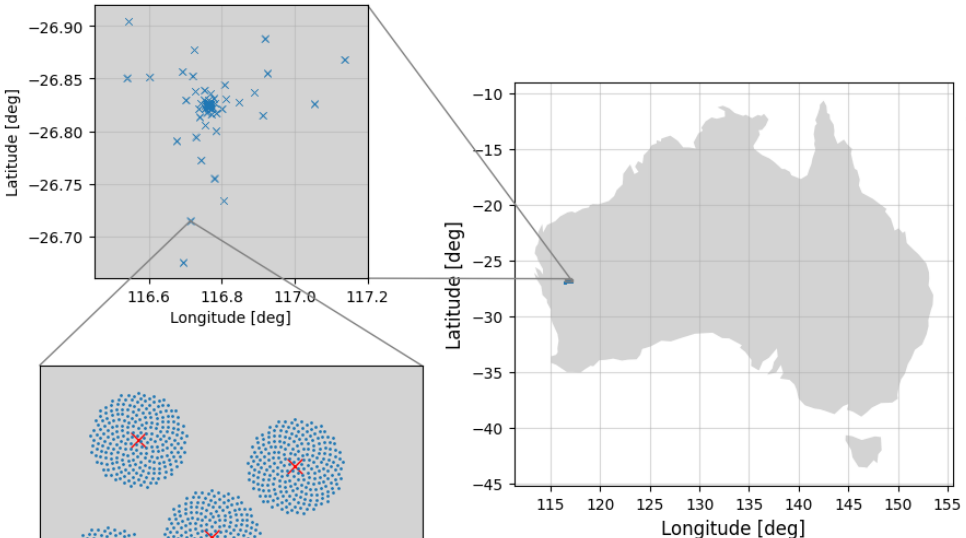
Current cosmology is inferred by either lager scale or local cosmological constraints

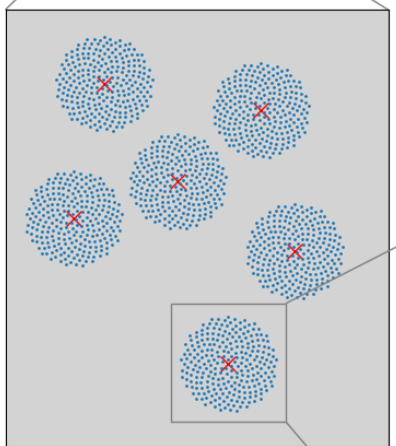
Cosmology with Radio Interferometric Experiments

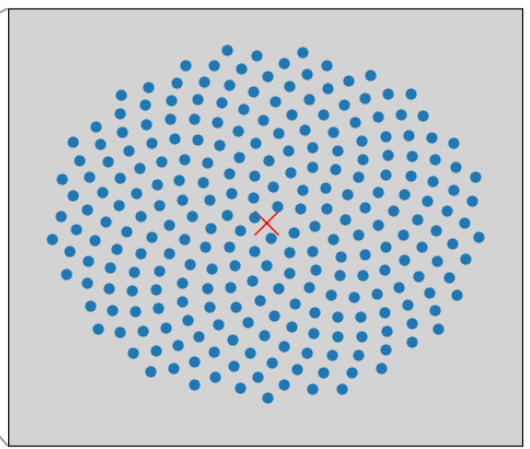




Square Kilometre Array (SKA-Low)







Final layout of the SKA-Low stations and **start of the construction** on the 5th of December 2022.

- 512 stations, each with 128 antennas
- Maximum baseline ~65 km, station diameter 35 m
  - FoV = 10 deg with 16 arcsec resolution
- Frequency range between 50 to 350 MHz (z = 27 – 5)

Ongoing Radio Experiments

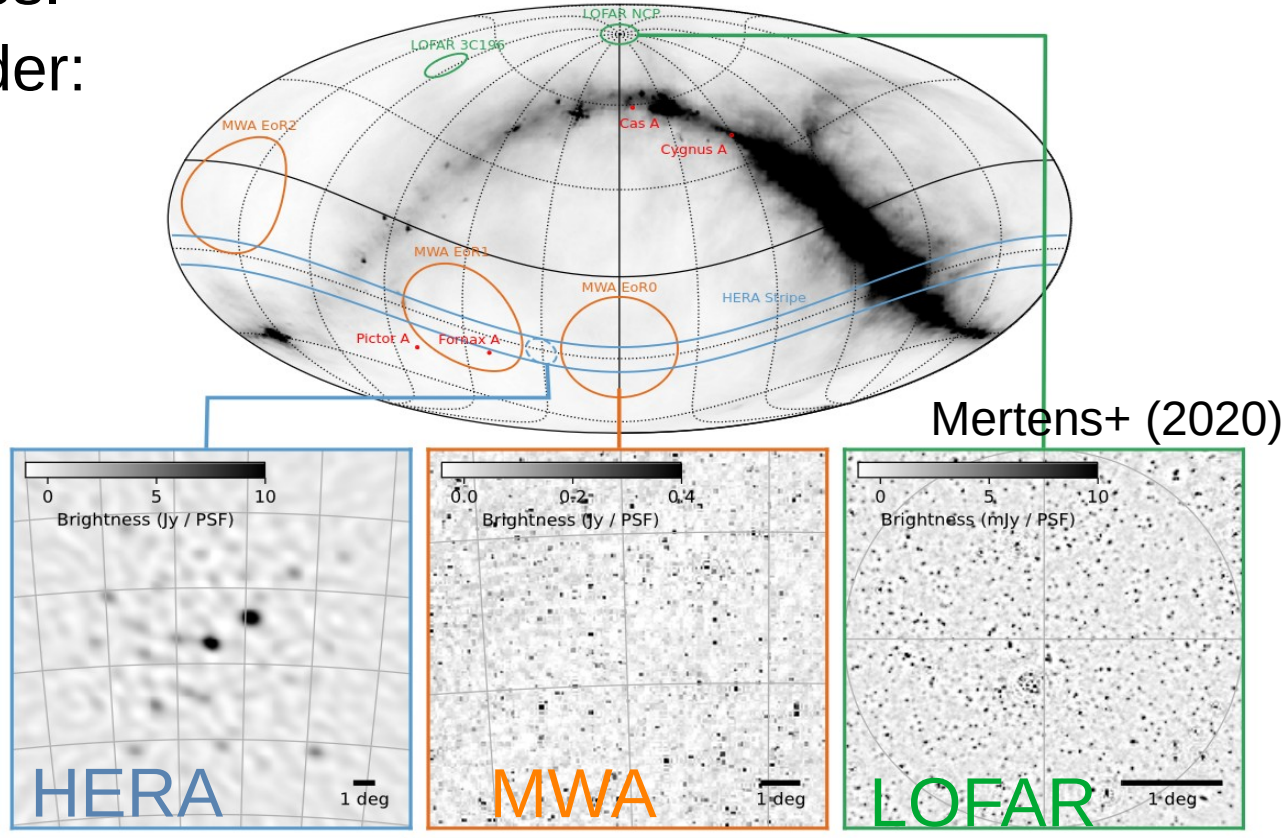
Emission at low frequency of the Galactic foreground sky and radio extra-galactic point sources.

Results form SKA pathfinder:

- HERA

- MWA

- LOFAR



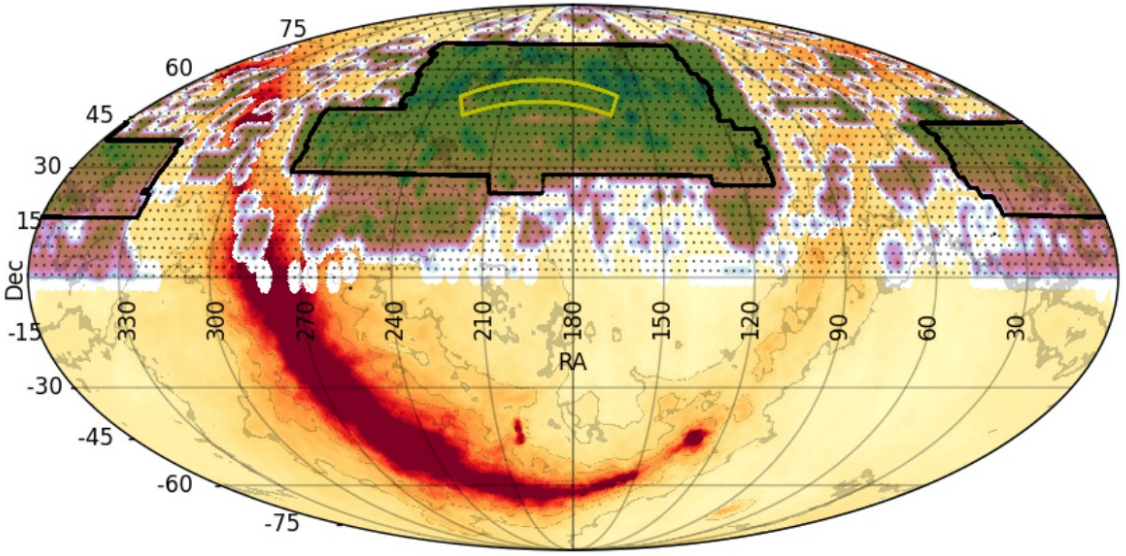
Mertens+ (2020)

Ongoing Radio Experiments

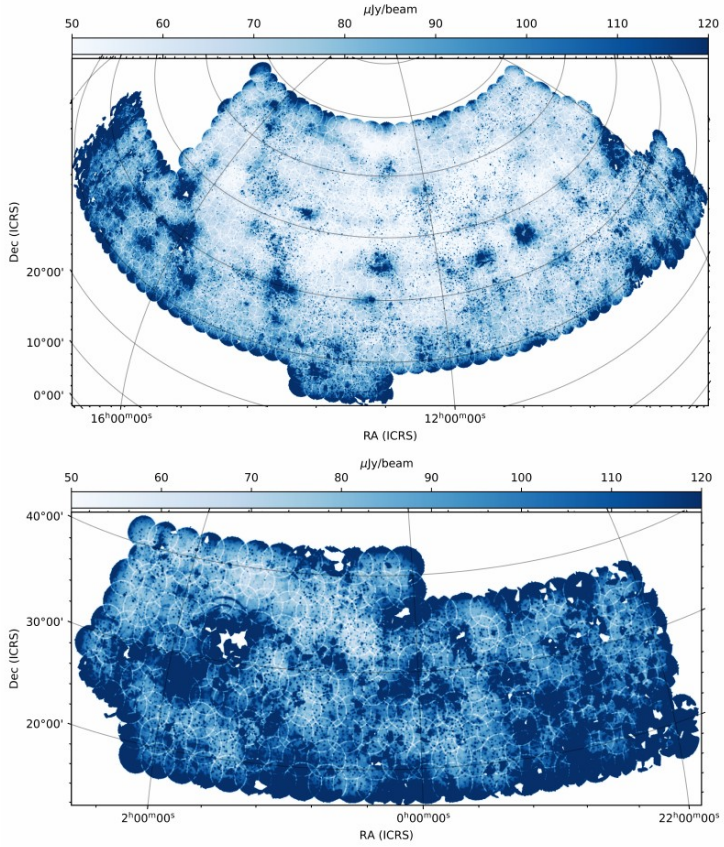
LOFAR is covering 27% of North sky avoiding the Galactic plane.

Recent data release of the **LoTTS-DR2** observation:

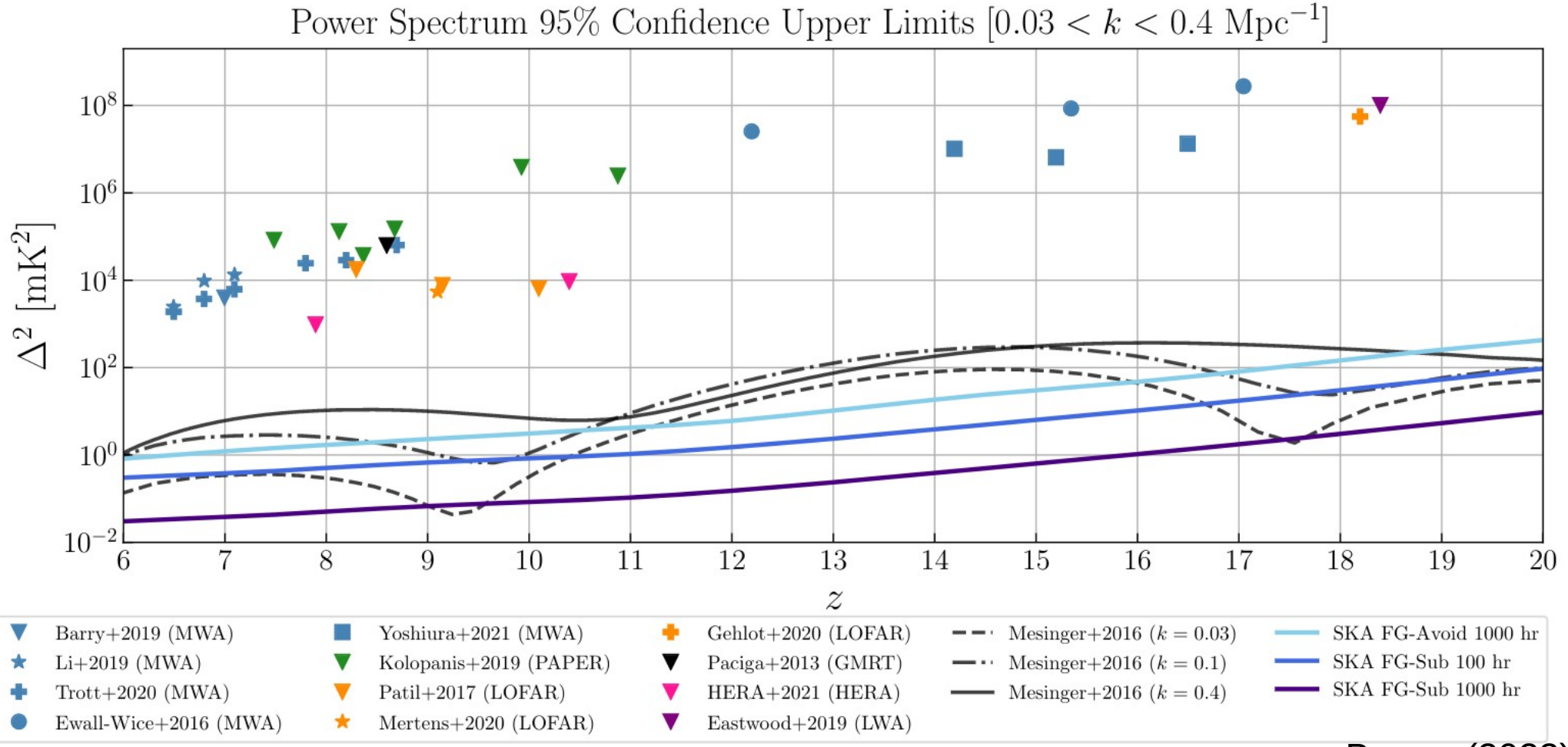
- High Band Antenna: 120 – 168 MHz
- 3.4 hrs obervation (7.6PB)



Shimwell+ (2022)

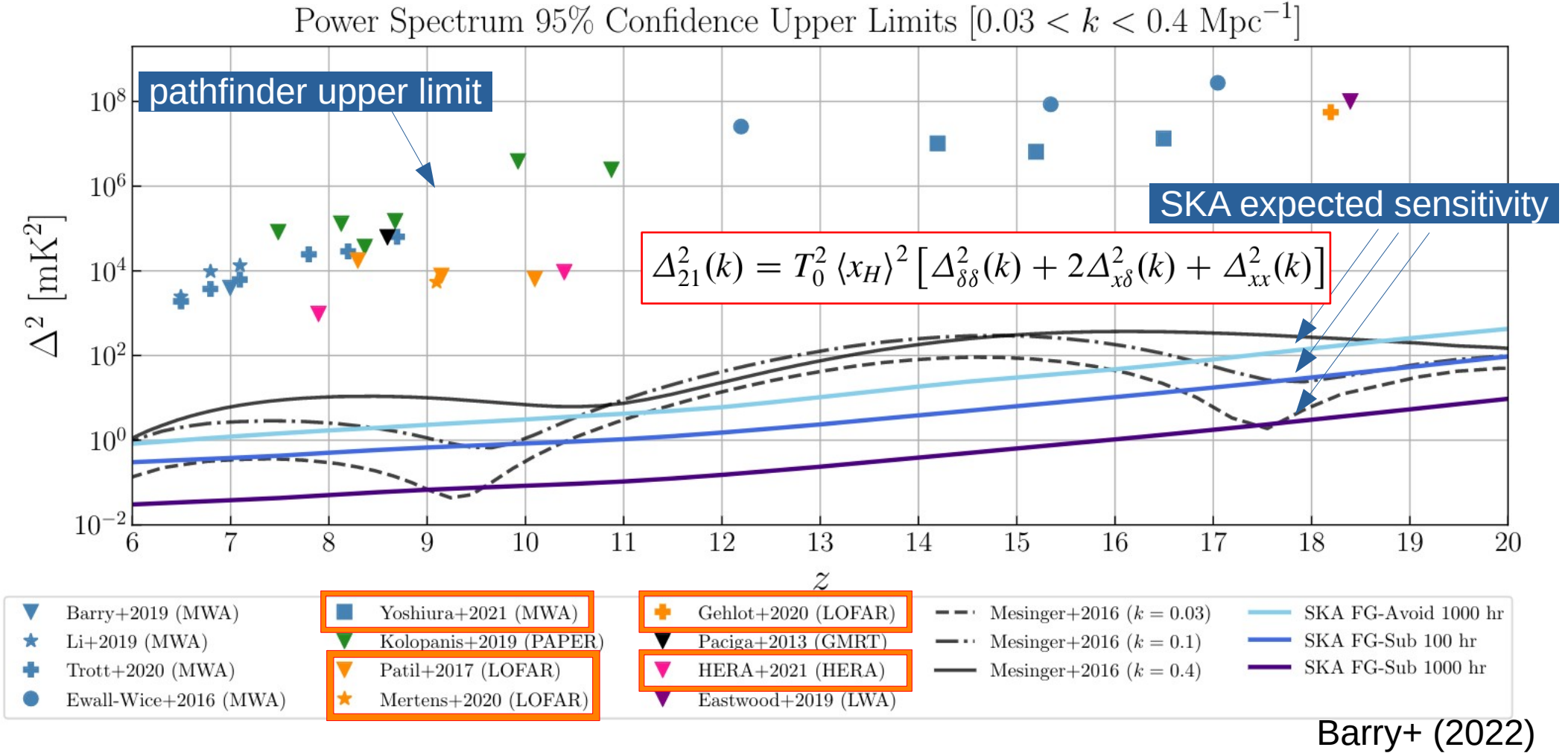


Constrains on the 21-cm Power Spectrum



Barry+ (2022)

Constrains on the 21-cm Power Spectrum



Barry+ (2022)

The 21-cm Signal During EoR

$$\delta T_b = 28\text{mK}\,(1+\delta)\,x_{\text{HI}}\left(1-\frac{T_{CMB}}{T_{spin}}\right)\left(\frac{\Omega_b h^2}{0.0223}\right)\sqrt{\left(\frac{1+z}{10}\right)\left(\frac{0.24}{\Omega_m}\right)}\left[\frac{H(z)/(1+z)}{\mathrm{d}v_{\parallel}/\mathrm{d}r_{\parallel}}\right]$$

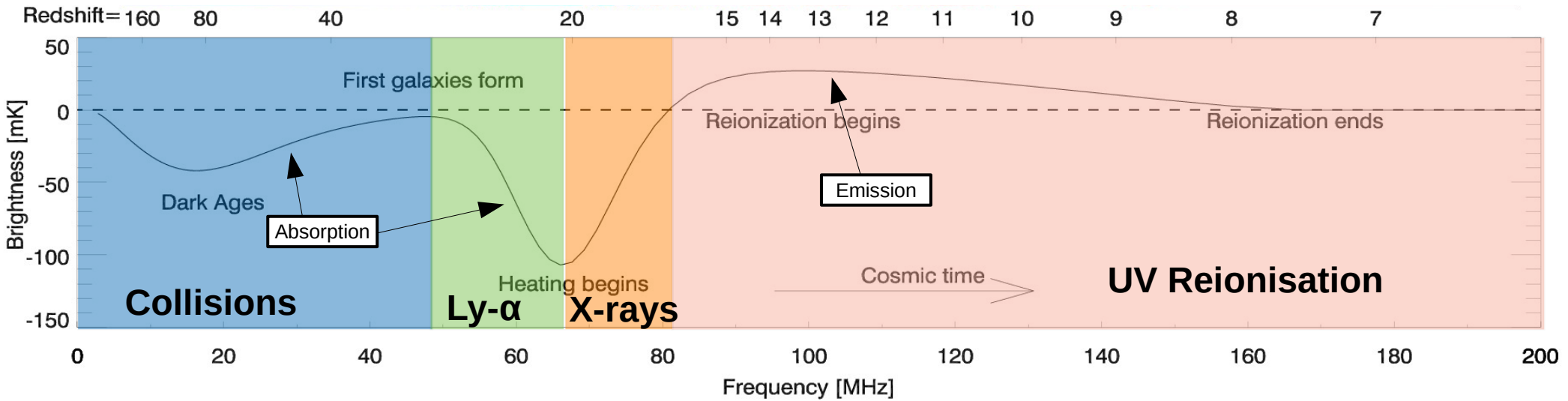
Gas density distribution

Hydrogen Neutral fraction

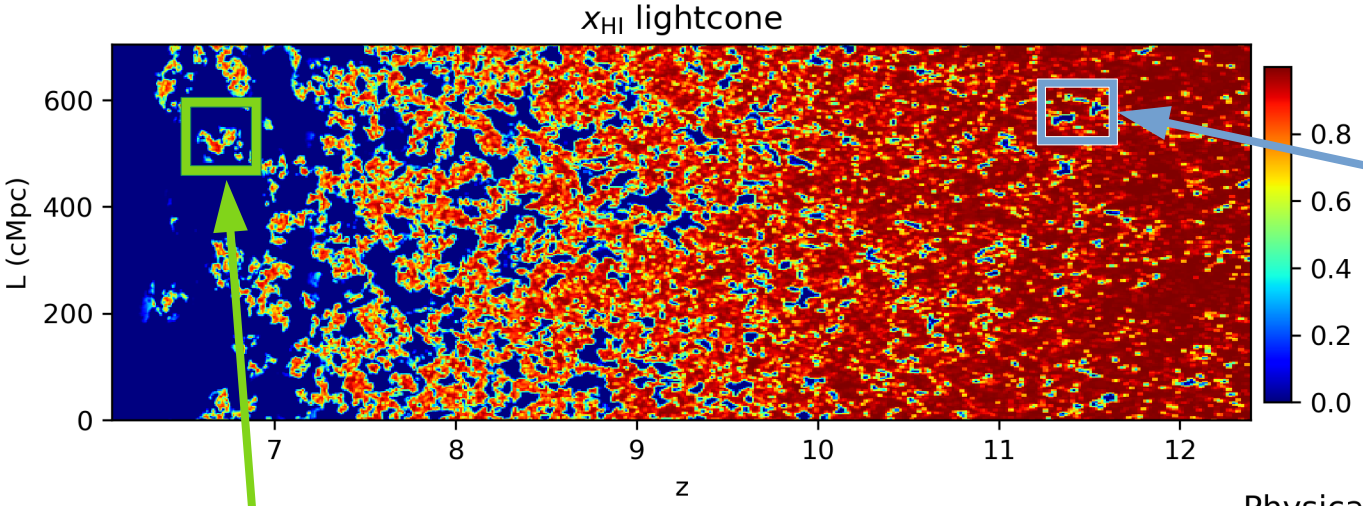
Spin temperature

Cosmology

L.o.S velocity



The 21-cm Signal During EoR

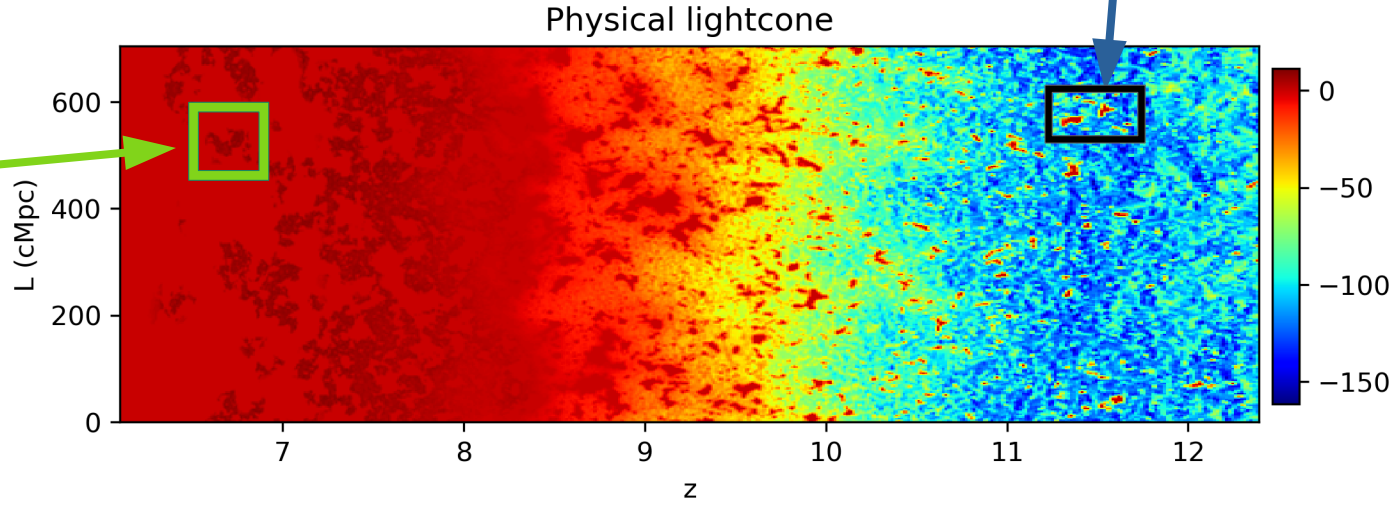


$T_S \ll T_{CMB} \Longrightarrow \delta T_b \ll 0\,\mathrm{mK}$

21-cm in emission in cold IGM

21-cm emission = Neutral hydrogen

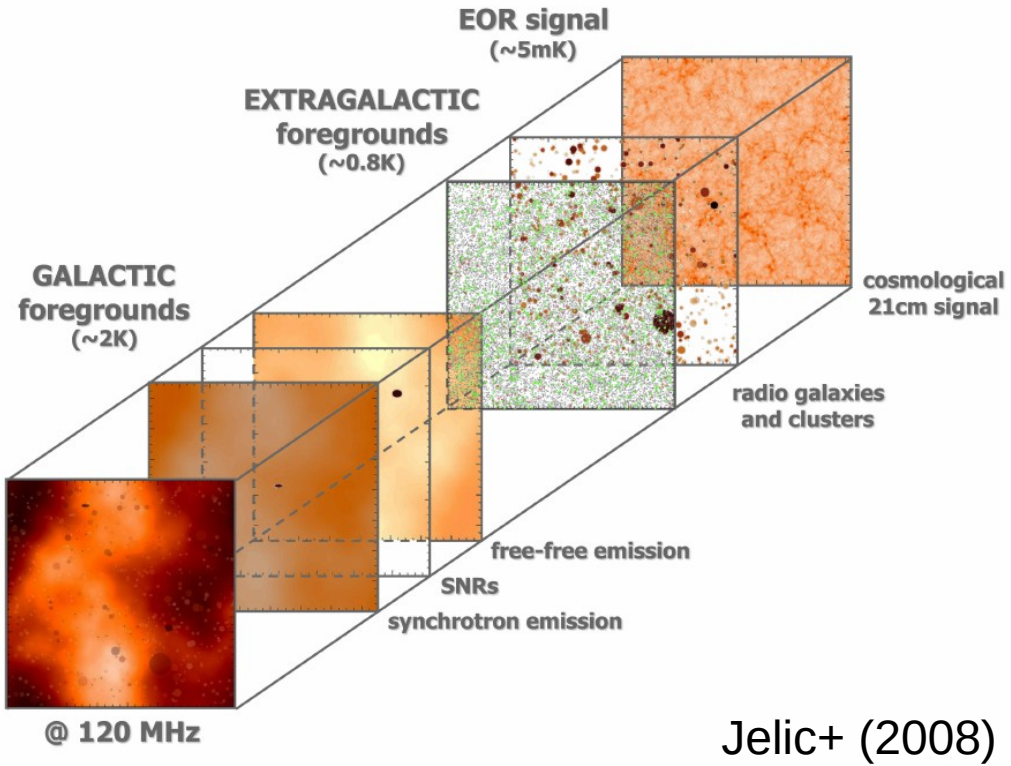
$\delta T_b \propto x_{\rm HI}(z)$



Tomographic Imaging of the 21-cm signal

SKA1-Low tomographic images of redshifted 21-cm signal challenges:

- Instrumental noise (signal ~ 5 K)
- Foreground emission (signal ~ 1 - 1000 K)
- Antennas gain errors
- Ionospheric refraction effects
- Radio frequency interference
- And more ...

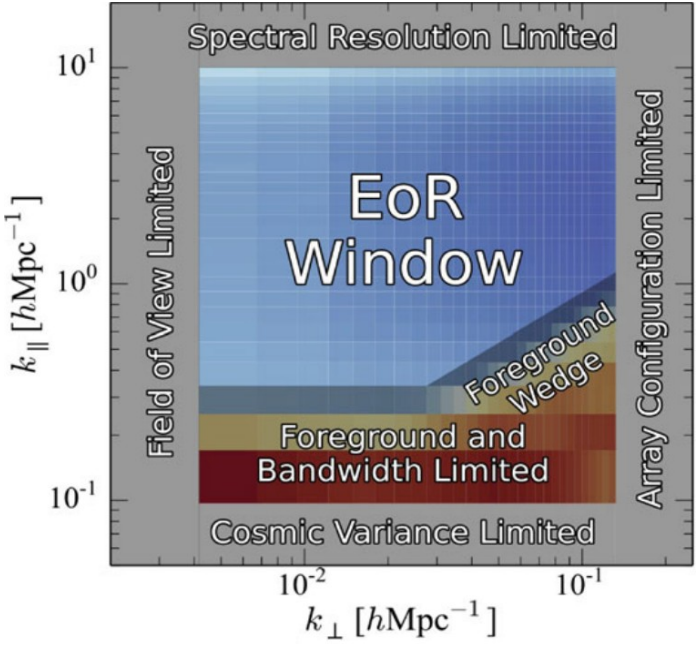


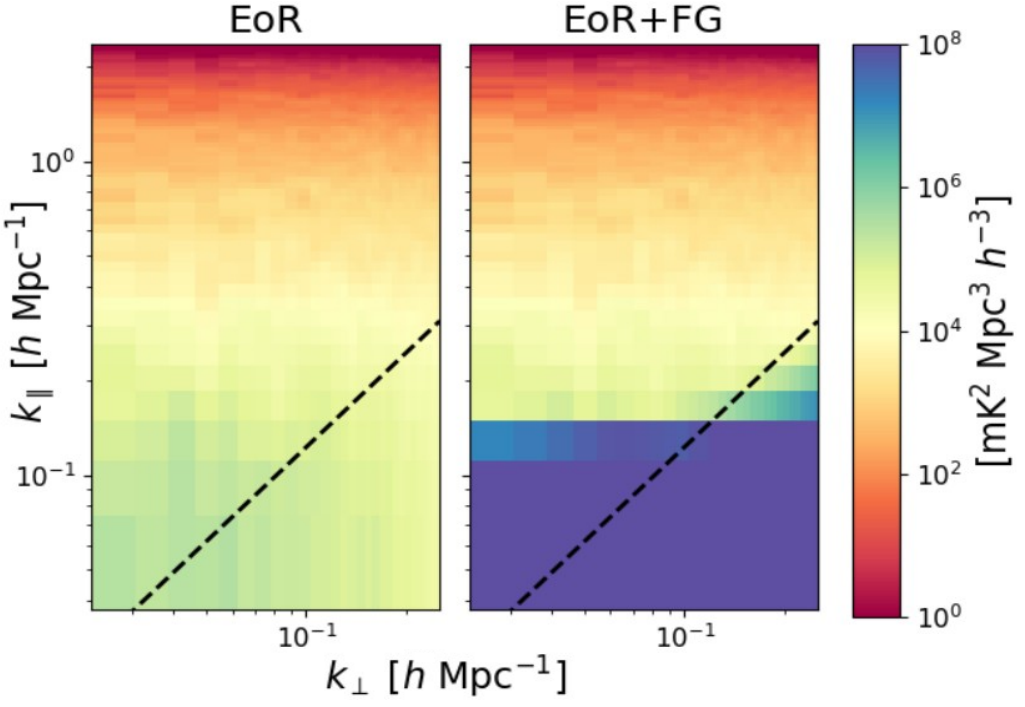
Jelic+ (2008)

Contamination from Galactic & extra-galactic Foregrounds

Most **foregrounds** have **frequency smooth spectra** compared to the 21-cm signal.

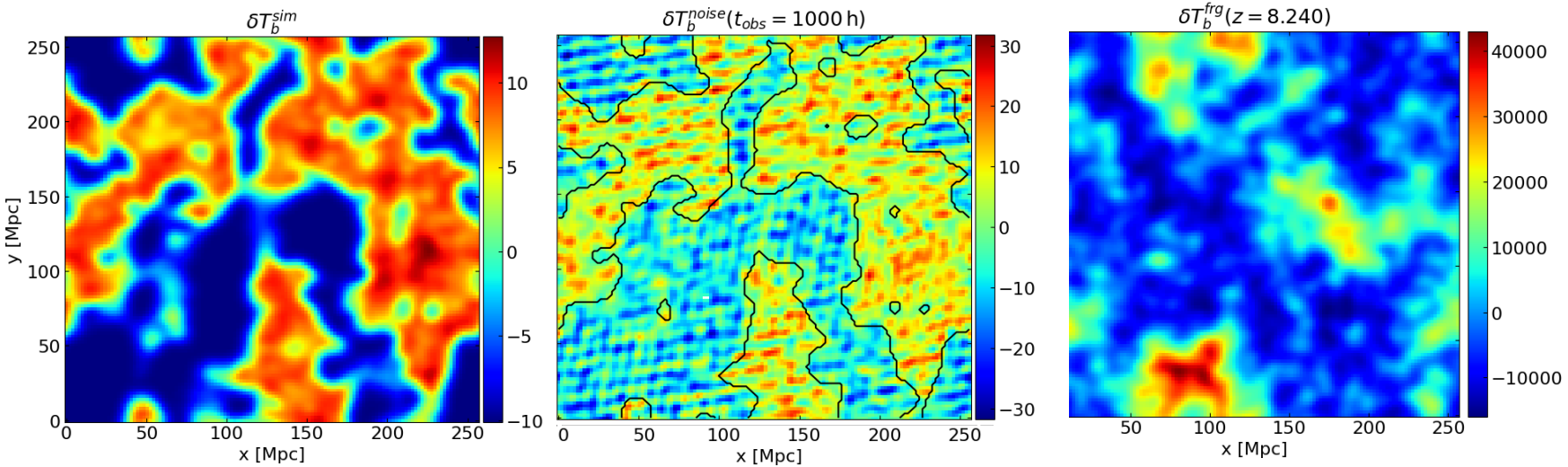
From 2D power spectra, remove $k_\perp$ - $k_\parallel$ modes that are contaminated by the foregrounds as a **avoidance technique** or model them for **substraction**.





Mock Data for 21-cm Observation

Currently we can create EoR mock observation with a combination of numerical models for 21-cm, systematic noise and Galactic foreground.

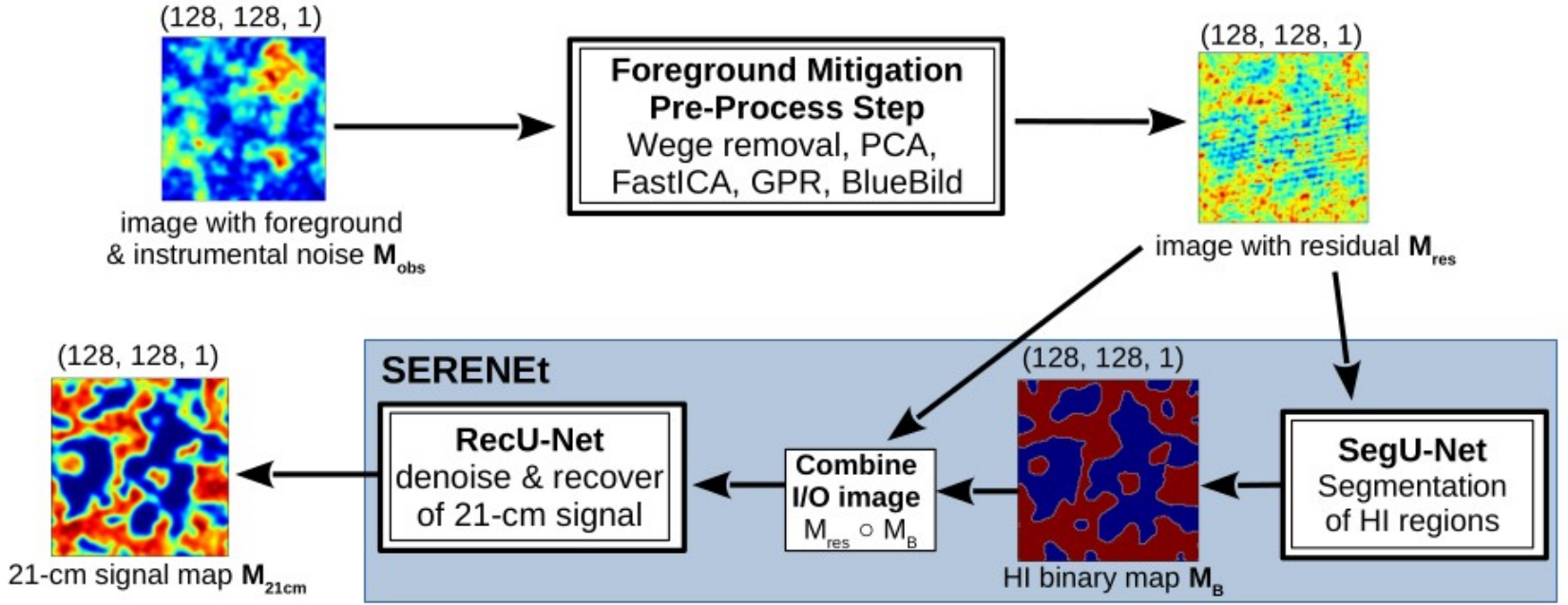


<u>Goal:</u>

pipelines to recover differential brightness and the distributon of netural hydrogen from mock observations.

SERENEt
SEgmentation and **RE**gression **NEt**work

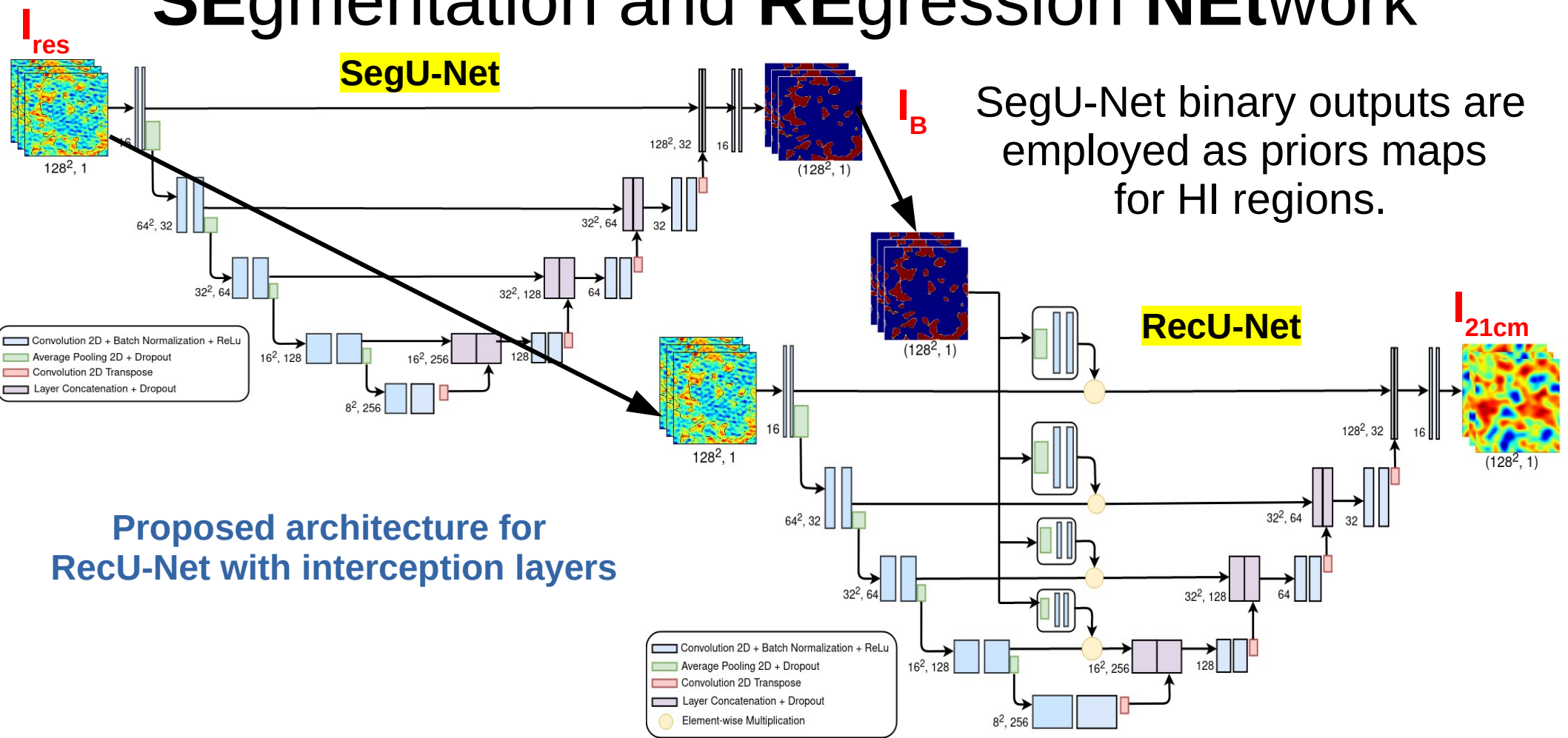
Combine the predicted binary maps of **SegU-Net** as additional input of **Rec-Unet** training step in order to include prior in the network training.



SKACH large HPC **70k hybrid-h** allocation projects at Pitz Daint at CSCS

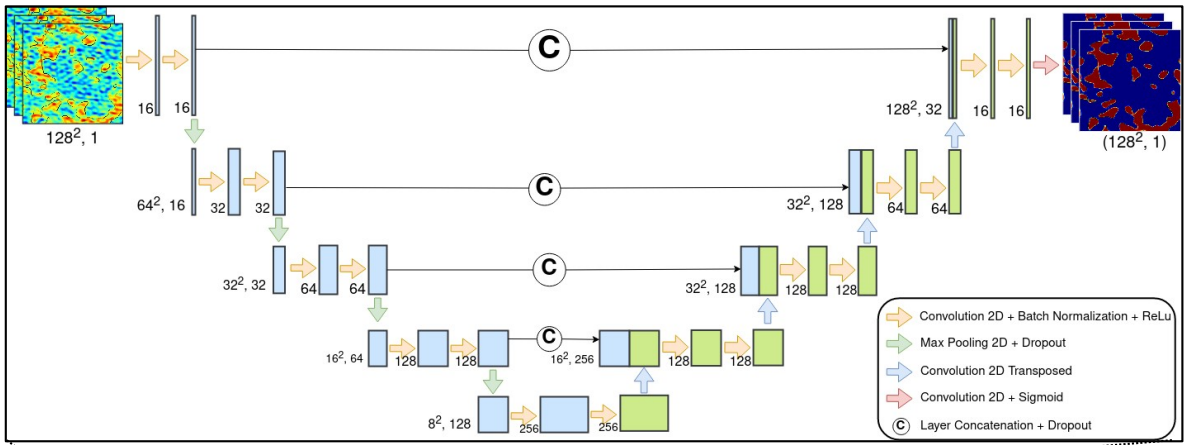
SERENEt

SEgmentation and REgression NEtwork

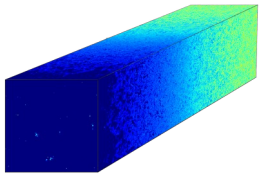


SegU-Net: Segmentation with U-Net
(Bianco+ 2021, 2023)

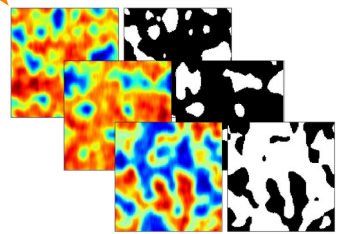
- U-Net:
  Network with interconnected encoder/decoder layers



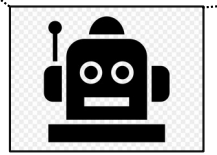
add PCA pre-process step that decreases image dynamic range



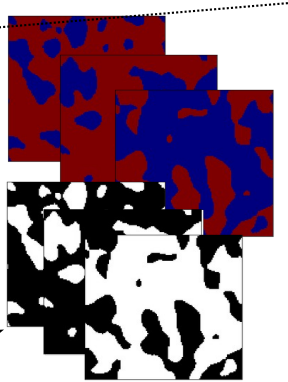
21cm tomography dataset



(z, ν_{obs})
21cm signal = image
x_{HI} field = mask



SegU-Net
(2.5M trainable params.)



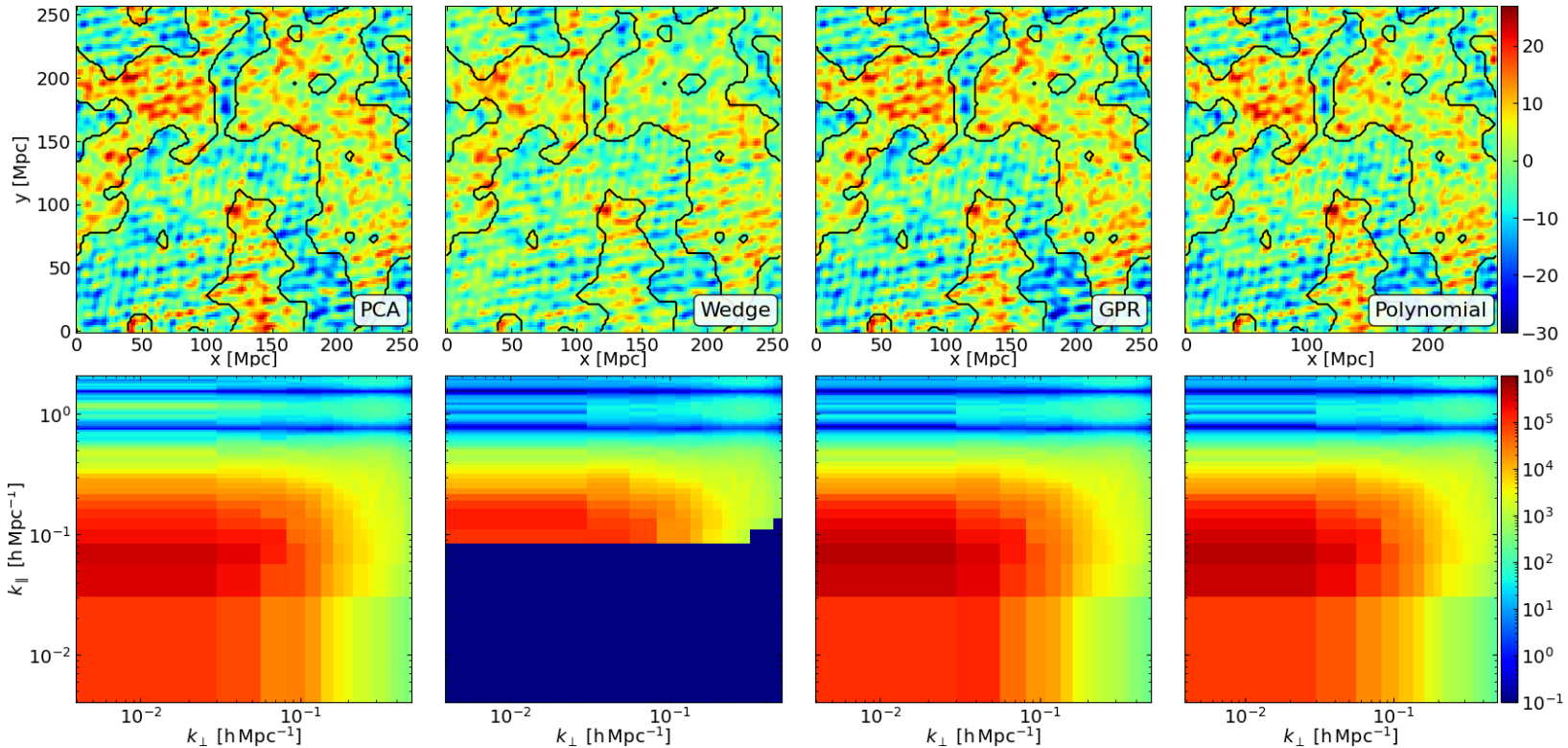
Compare Prediction/True mask

$\mathcal{L}(y, \hat{y})$

Calculate Loss

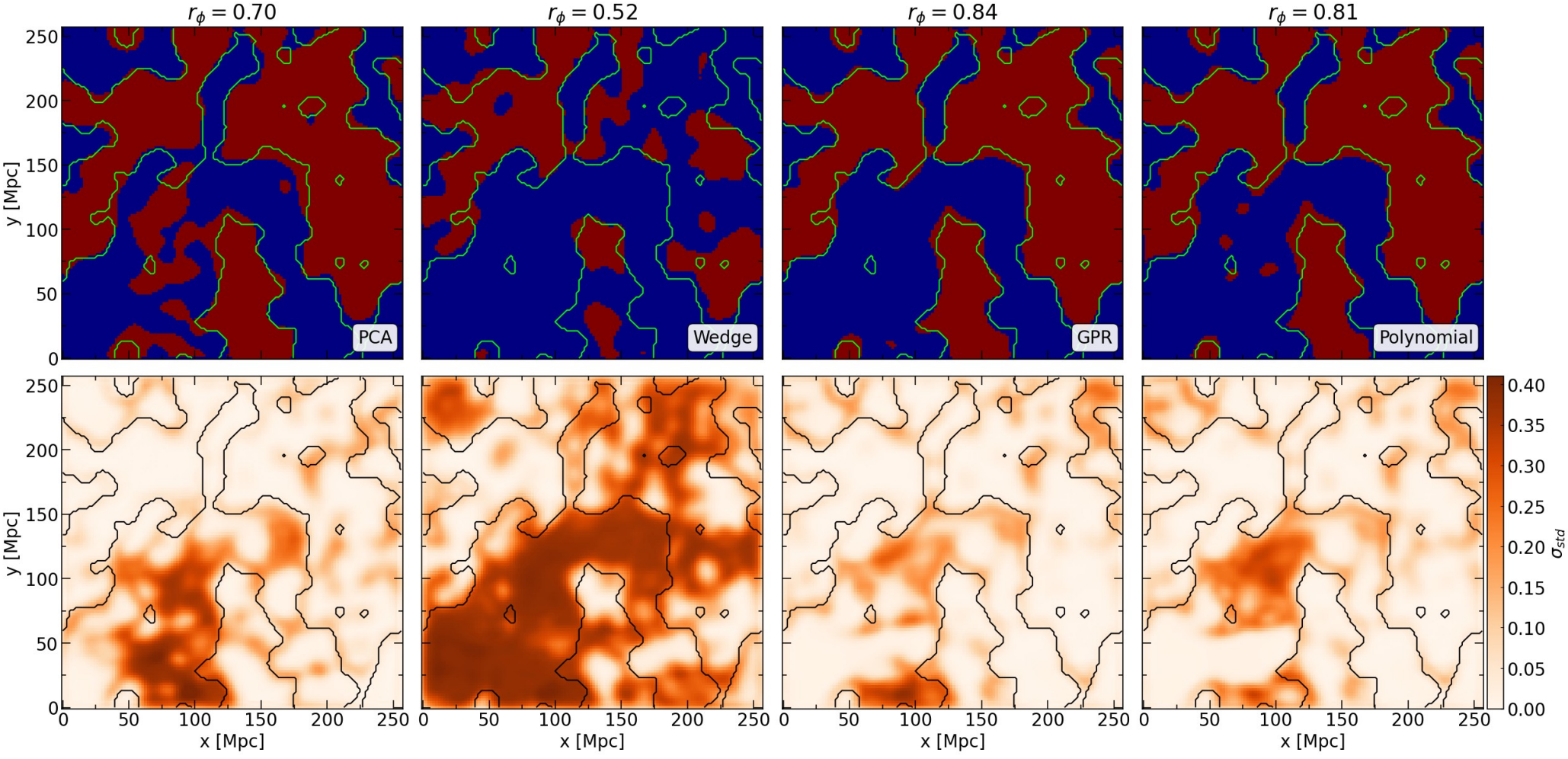
Pre-process:
Foreground Mitigation & Avoidance

We employ 3 mitigation and 1 avoidance techniques, Bianco+ 2023



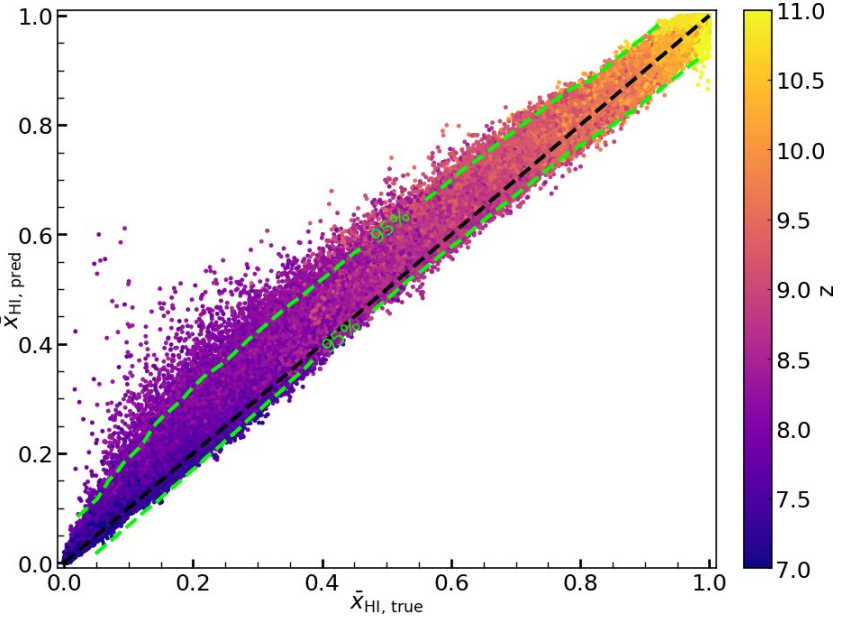
PCA Wedge GPR Polynomial

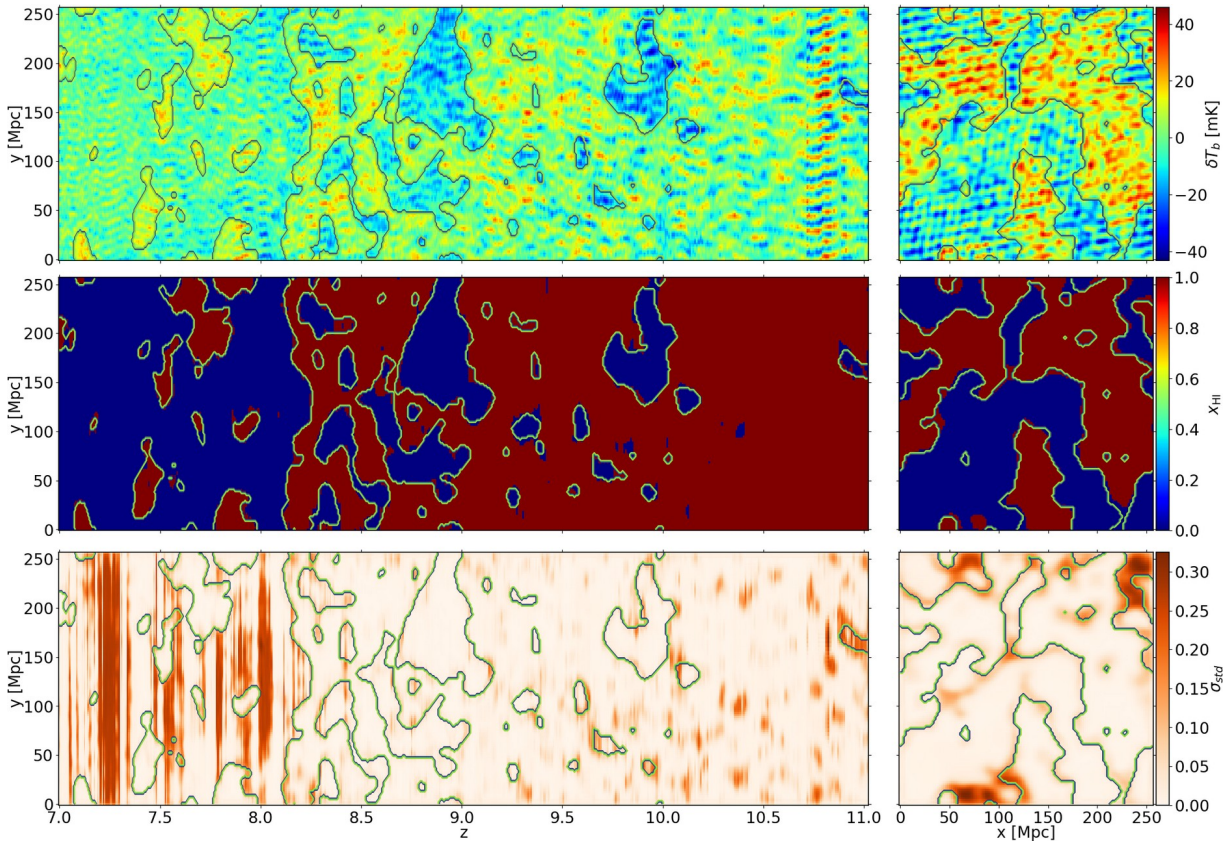
SDC3: General Annoucment



SegU-Net:
Tomographic Data & Reionization History

z_c	pre-process	$\overline{x}_{\rm HI}$
8.24	Ground Truth	**0.45**
	all z PCA	0.48 ± 0.07
	PCA	0.49 ± 0.11





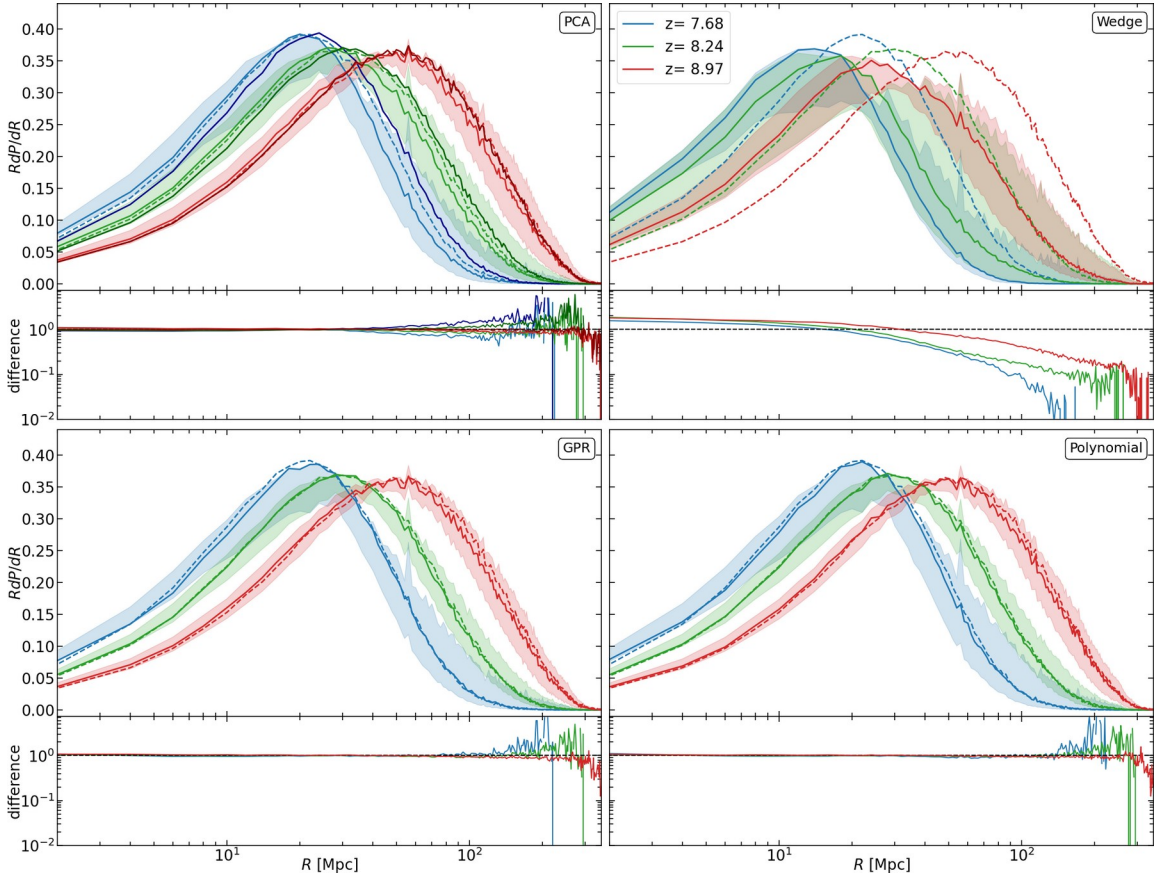
SegU-Net: HI size distribution

The Island (HI regions) size distribution is an powerfull probe of the reionization process. (21-cm non-Gaussianity)

$$\overline{R}_C(z) = \int_{R_{\min}}^{\infty} R \, \frac{dP}{dR}(z) \, dR$$

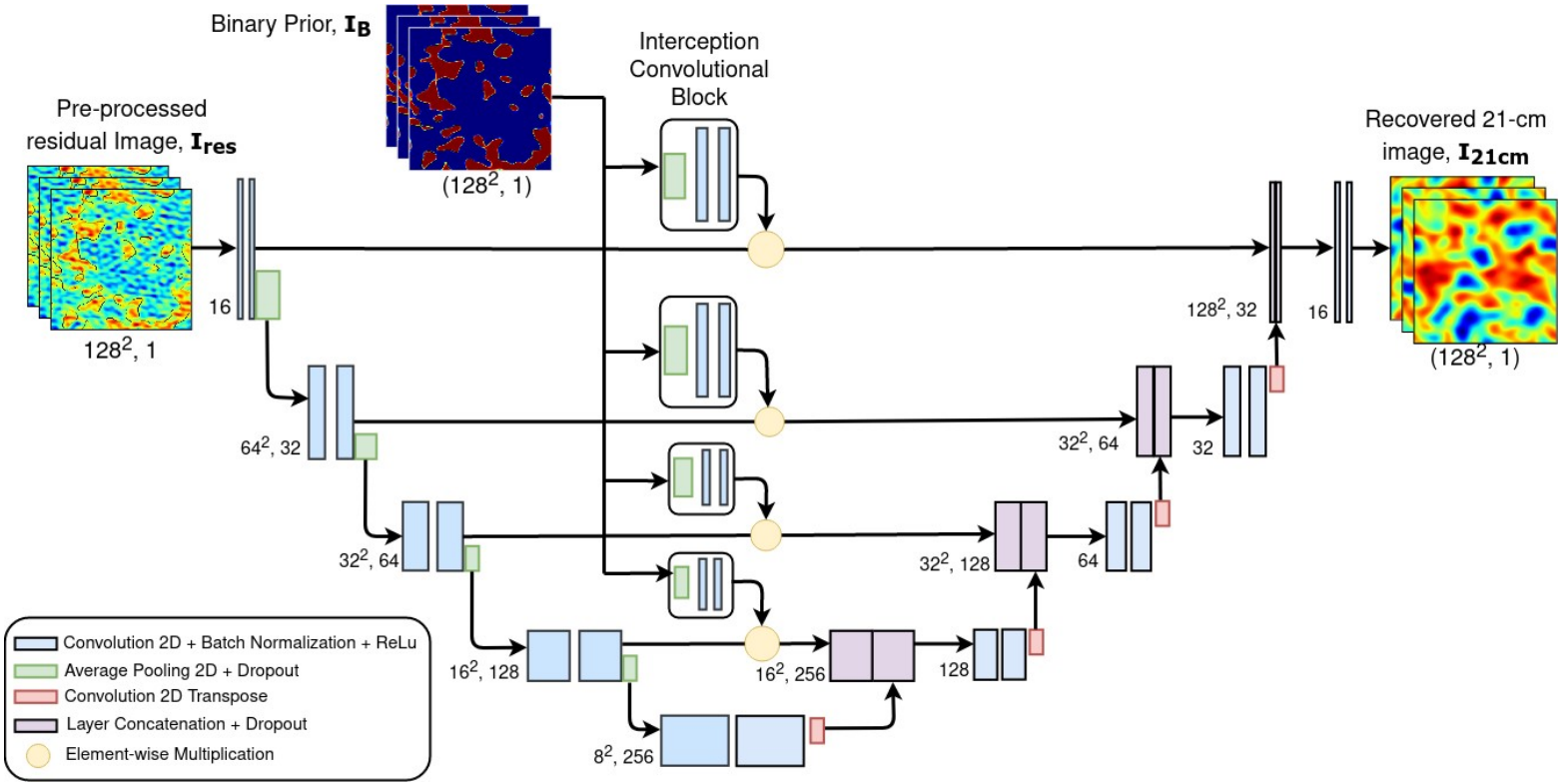
SegU-Net results:

z_c	pre-process	$\overline{R}_C$ [cMpc]
8.24	Ground Truth	**29.54**
	all z PCA	$31.37^{+3.09}_{-3.93}$
	PCA	$27.65^{+9.13}_{-6.12}$
	Wedge	$15.20^{+24.13}_{-6.18}$
	GPR	$29.14^{+5.26}_{-4.89}$
	Polynomial	$29.21^{+5.83}_{-5.21}$



RecU-Net: Recover 21-cm with U-Net

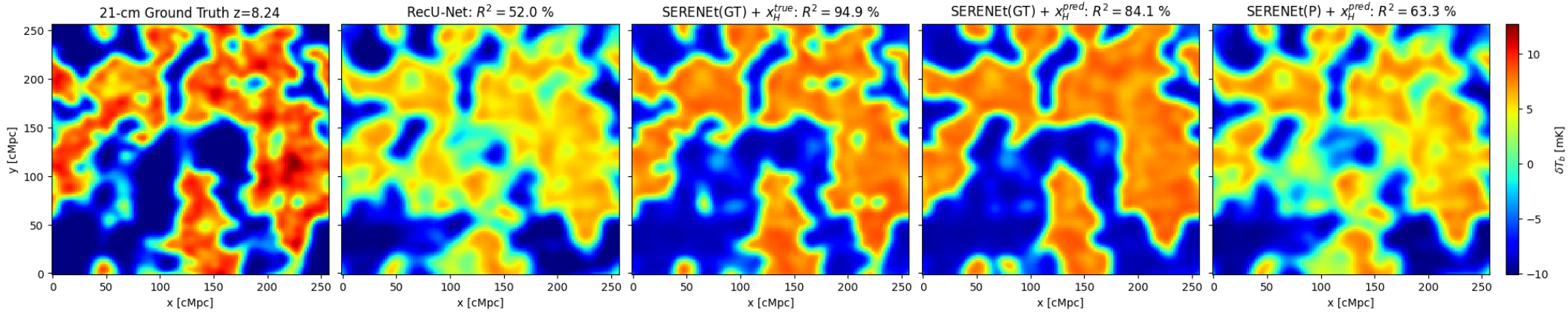
(Bianco+ in prep.)



U-Net architecture with intercepting convolution block to process the binary prior map from SegU-Net

SERENEt: Recover of 21-cm Signal

Recovered 21-cm signal for EoR for image at z = 8.25 (x_{HI}~0.5)



Best performing setup consider:

- Train SERENEt with the prior binary ground truth: **SERENEt(GT)**
- Real application requires the prredicted binary map from SegU-Net for the prediction: **SERENEt(GT) + x^{pred}_H**

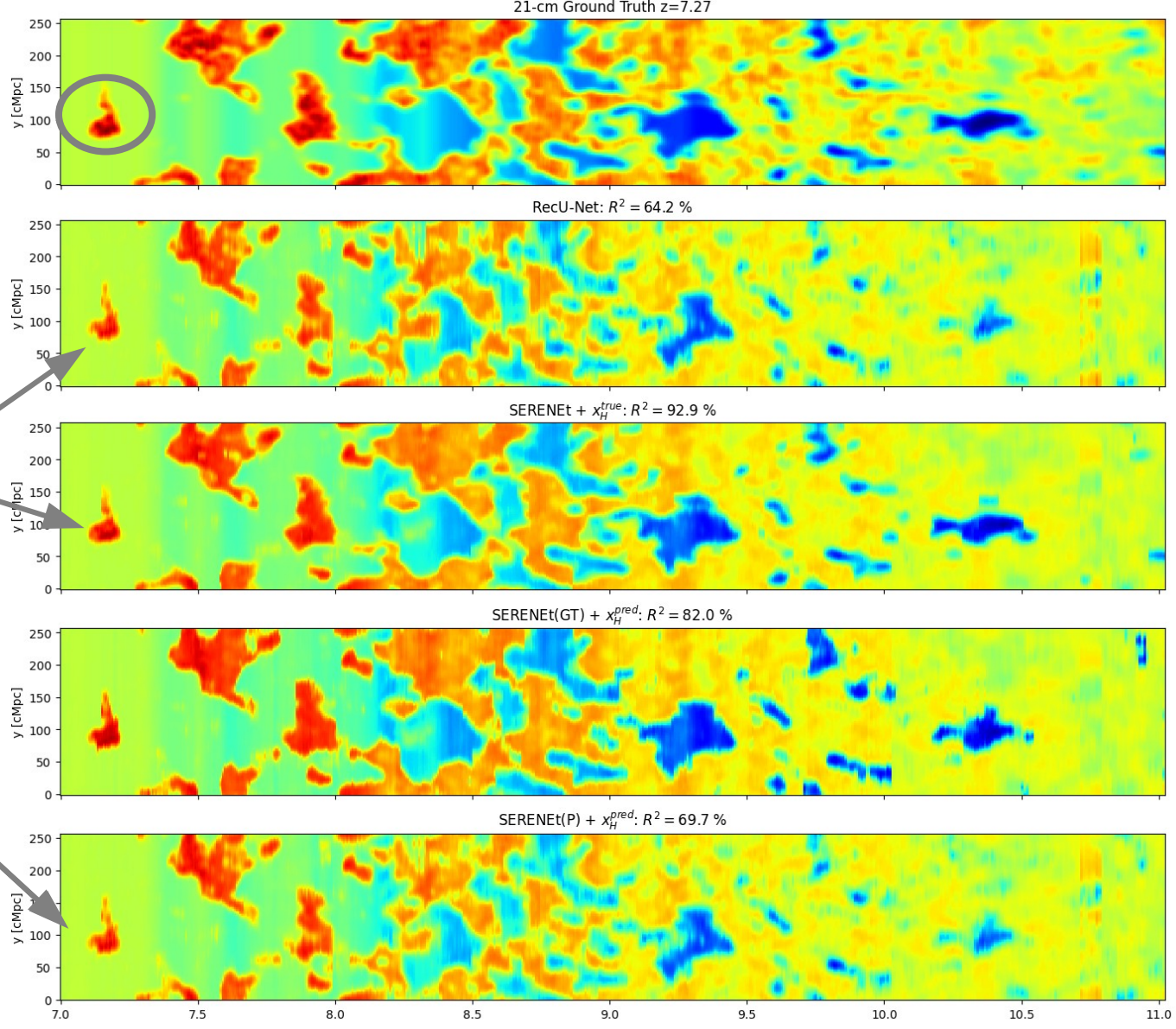
SERENEt

Limitation at lower redshift and decrease in accuracy for:

SERENEt(GT) + x^{pred}_H

When compared to

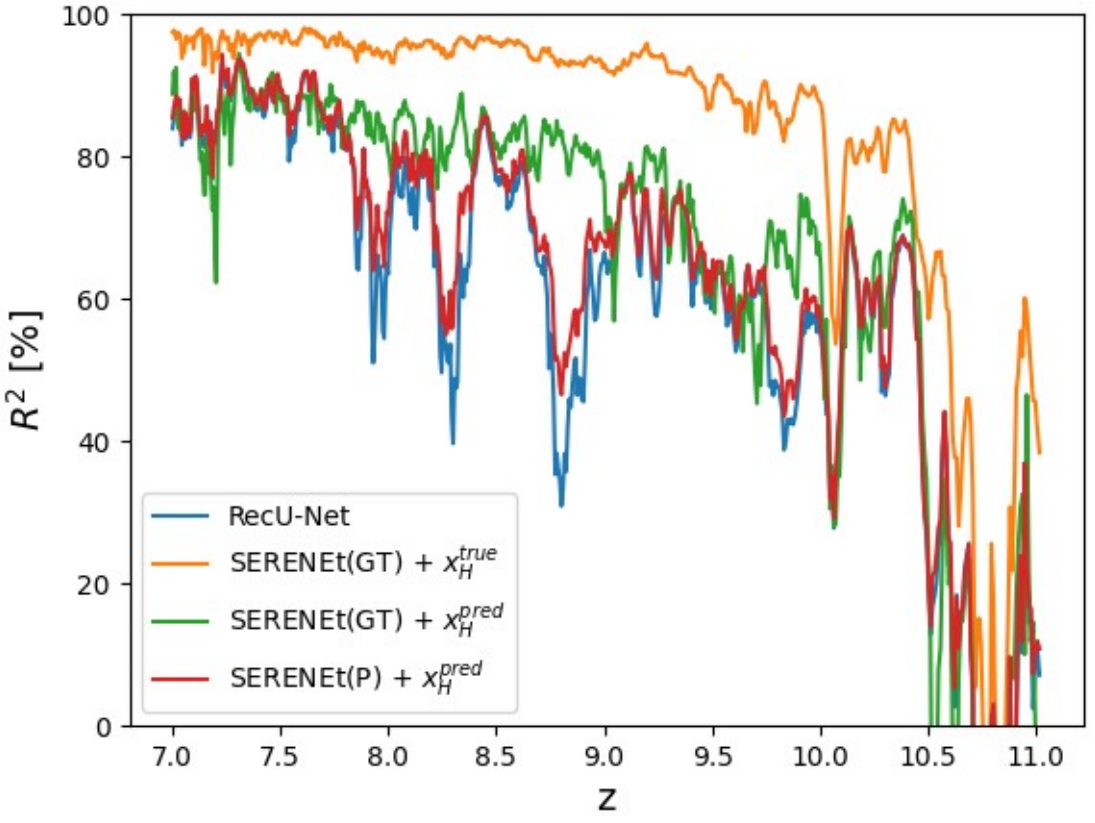
RecU-Net or
SERENEt(P) + x^{pred}_H



SERENEt: Recover of 21-cm Signal

Coefficient of determination (R^2 score) redshift evolution as a temporary Figure of Merit metric (FoM):

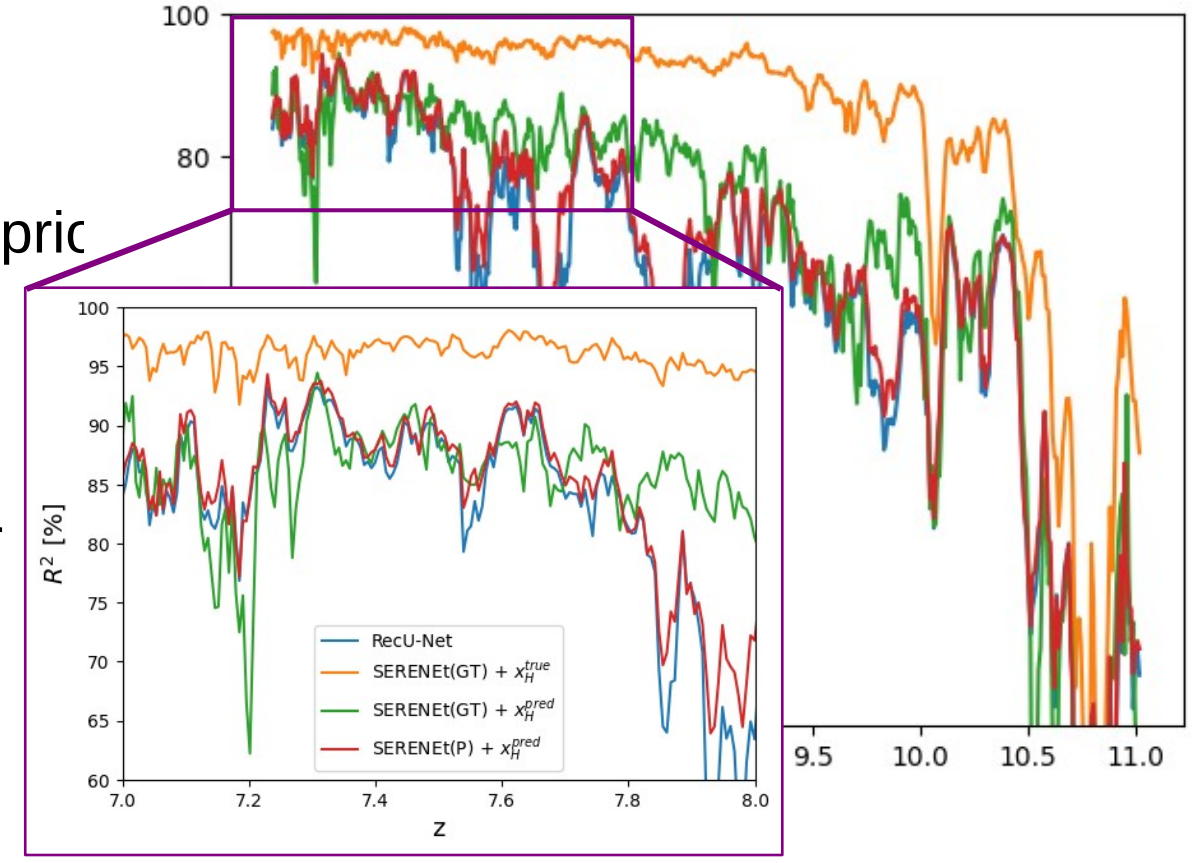
- **SERENEt(GT) + x^{true}_H**
  upper limit based on best pric
  binary map (ground truth)
- **SERENEt(GT) + x^{pred}_H**
  next best results when compared to **RecU-Net** or **SERENEt(P) + x^{pred}_H**



SERENEt: Recover of 21-cm Signal

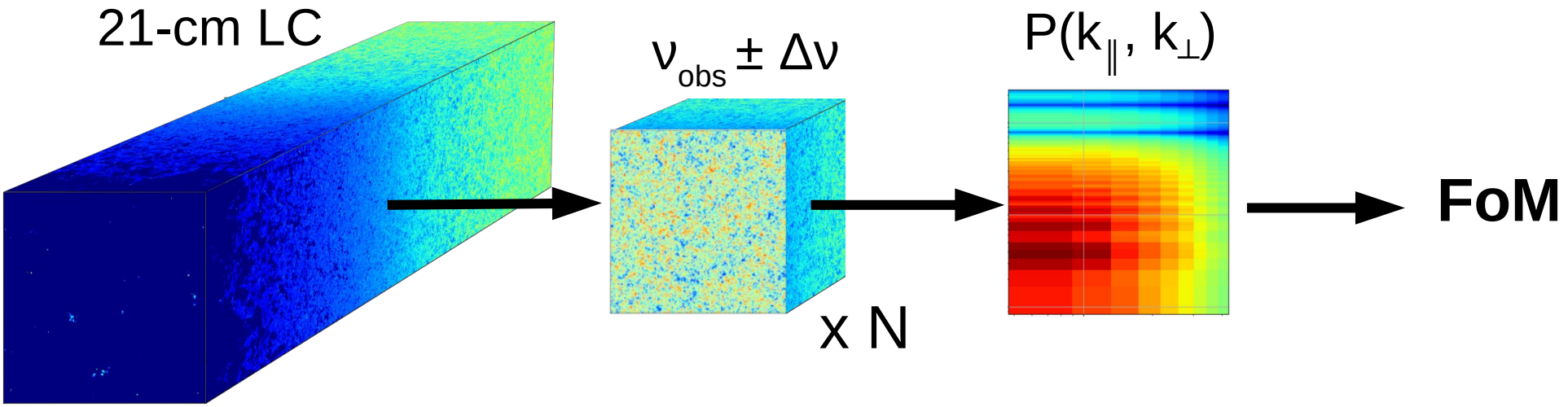
Coefficient of determination (R^2 score) redshift evolution as a temporary Figure of Merit metric (FoM):

- **SERENEt(GT) + x^{true}_H** upper limit based on best pric binary map (ground truth)
- **SERENEt(GT) + x^{pred}_H** next best results when compared to **RecU-Net** or **SERENEt(P) + x^{pred}_H**



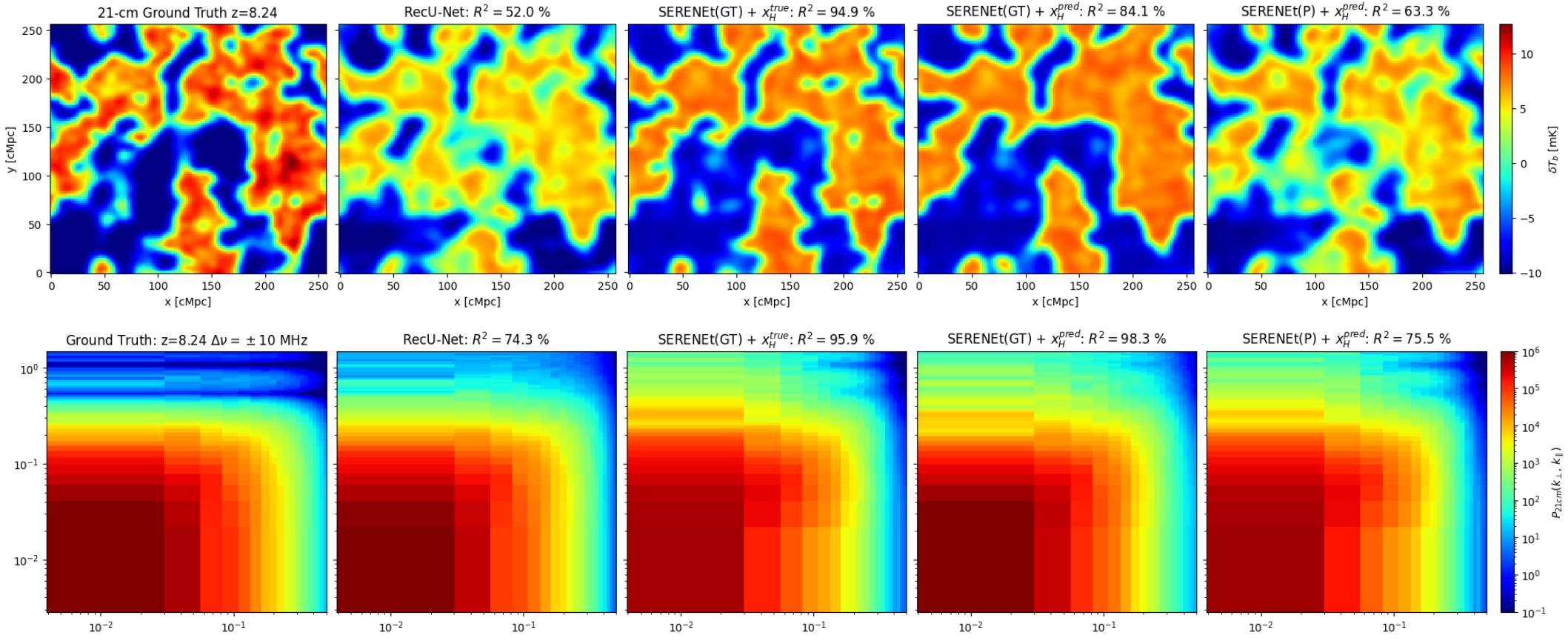
SDC3 Scoring System

The Figure of Merit (FoM) based on *"metric distance"* between the true and recovered spectra for **N sub-volumes** at a central frequency ν_{obs} and width $\Delta\nu$



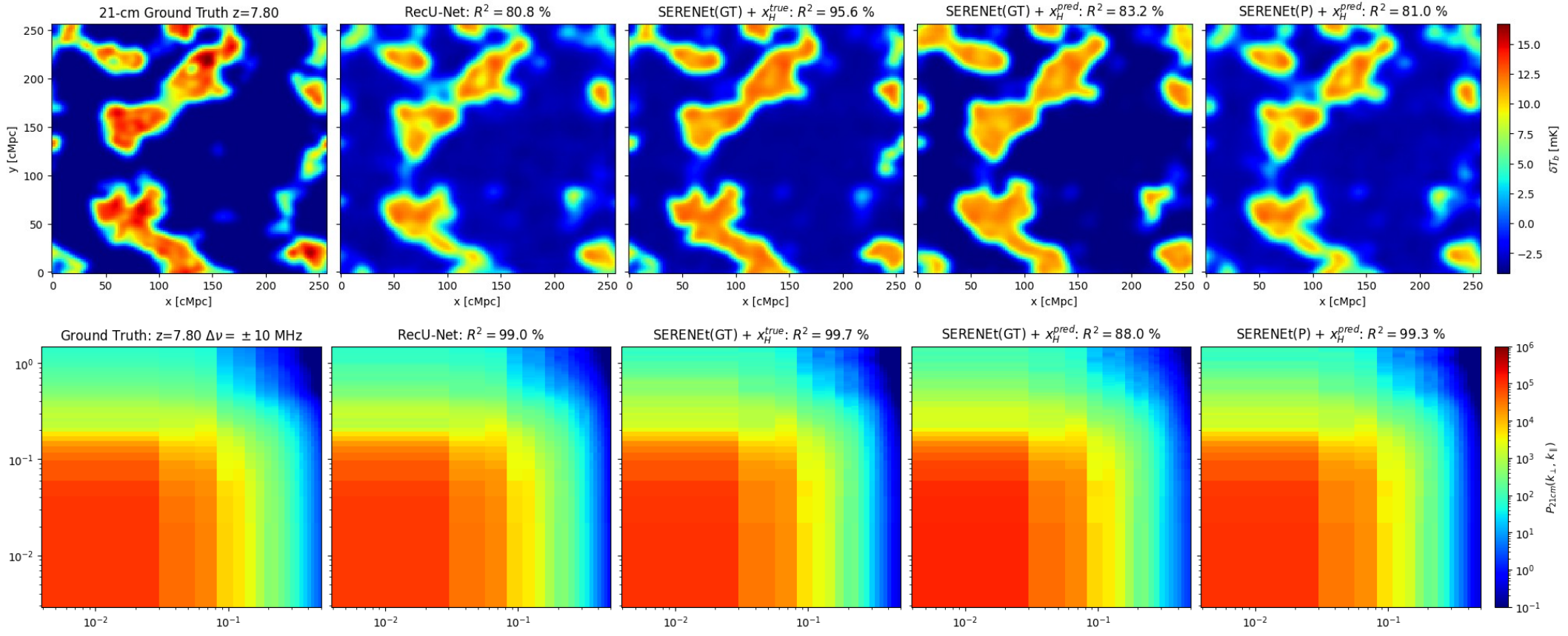
SERENEt: Recover of 21-cm Signal

Recovered 21-cm signal for EoR for lightcone subvolume centered at redshift **z = 8.25** (x_{HI}~0.5) and Δν +/- 10 MHz

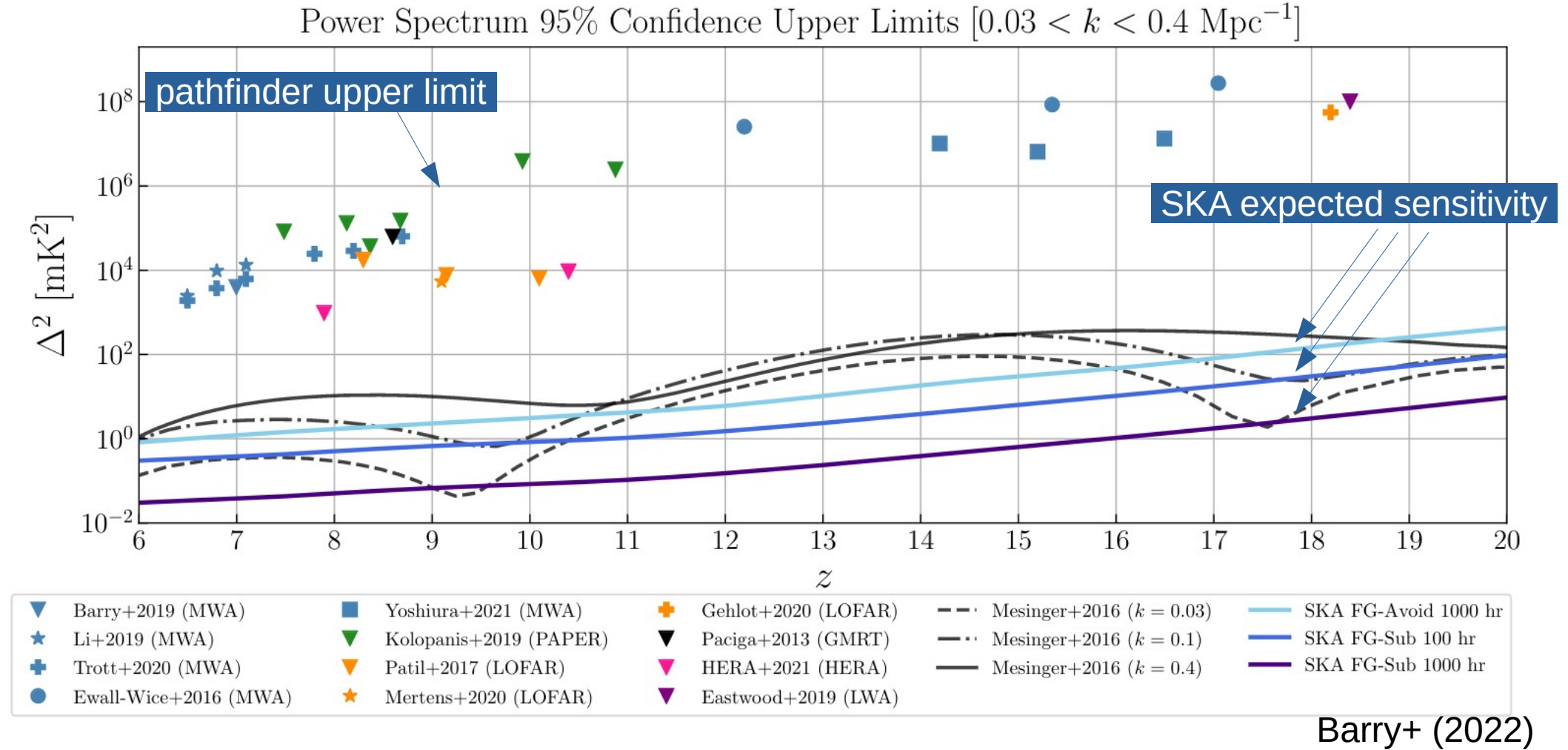


SERENEt: Recover of 21-cm Signal

Recovered 21-cm signal for EoR for lightcone subvolume centered at redshift **z = 7.8** (x_{HI}~0.3) and Δν +/- 10 MHz

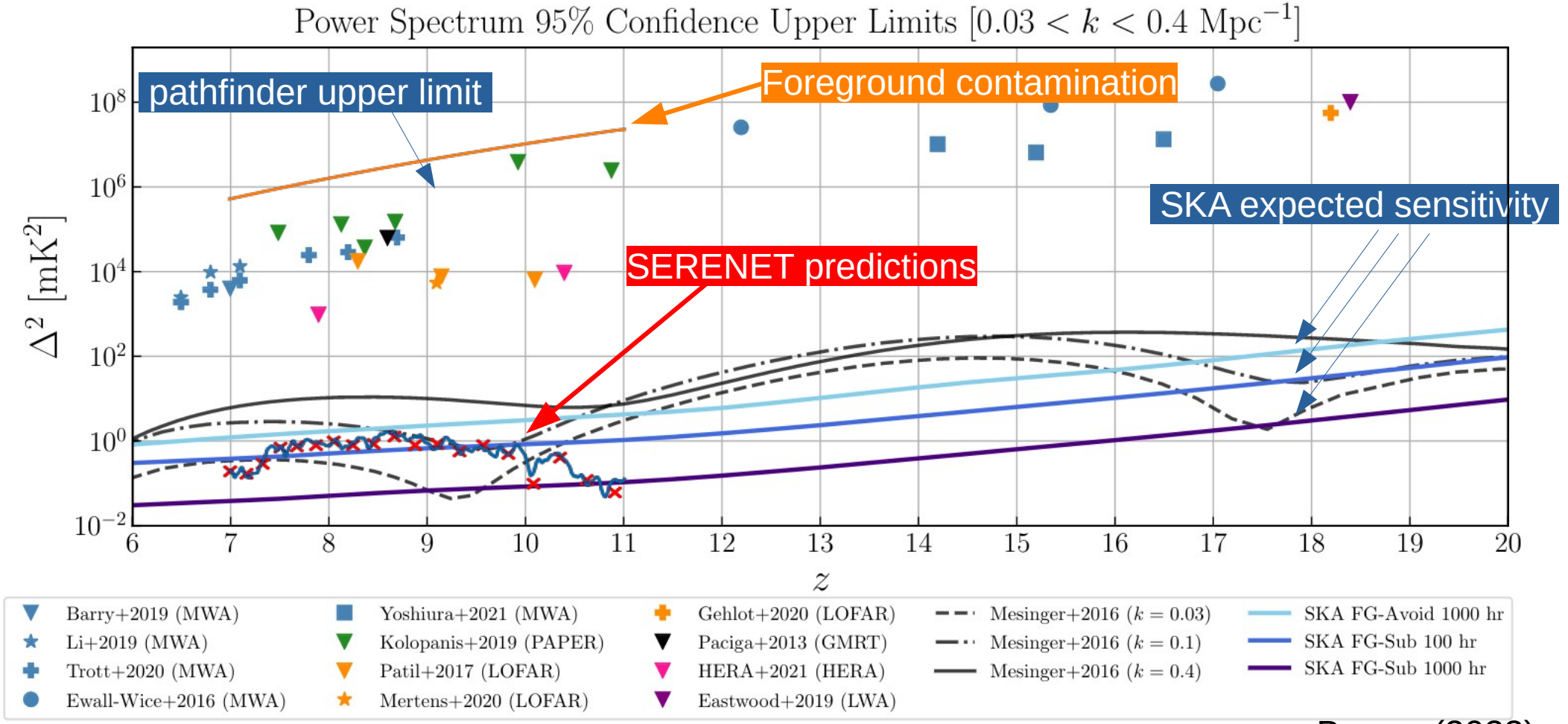


SERENEt: Comparison with Current Data



Barry+ (2022)

SERENEt: Comparison with Current Data



Barry+ (2022)

Thanks

Support Slides

Mesinger+ (2016)

$\delta(\mathbf{r}, z)$: Dark matter density fluctuations depends on structure formation

$x_{\mathrm{HI}}(z)$: Volume neutral fraction depends on SFH in the early Universe

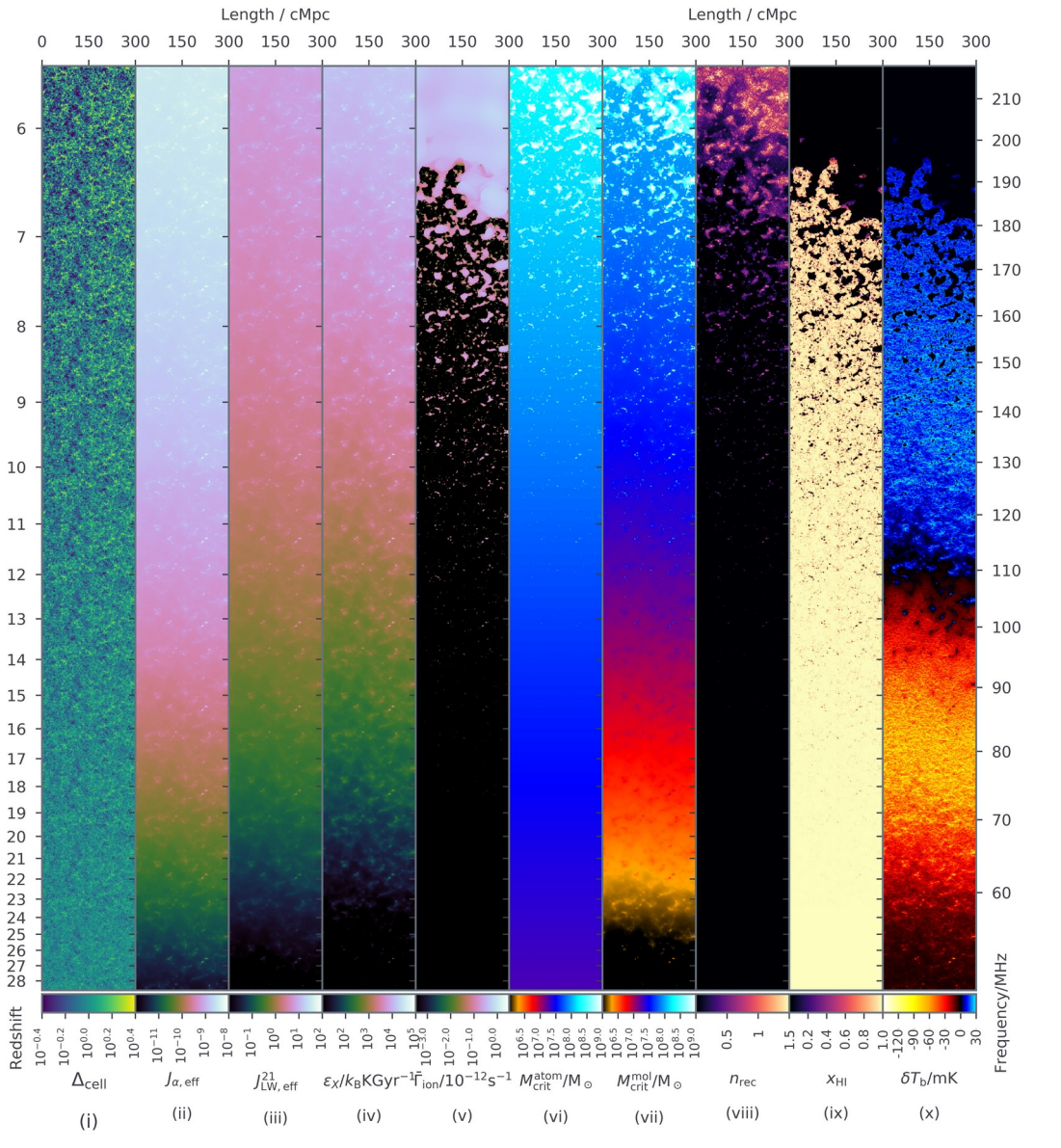
$$T_S^{-1} = \frac{T_{CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}$$

$x_\alpha \propto J_\alpha(z)$

Wouthuysen-Field effect (Lyα pumping)

$$x_c = x_c^{HH} + x_c^{eH} + x_c^{pH}$$

Collision coupling



Mesinger+ (2016)

$\delta(\mathbf{r}, z)$: Dark matter density : fluctuations depends on structure formation

$x_{\rm HI}(z)$: Volume neutral fraction depends on SFH in the early Universe

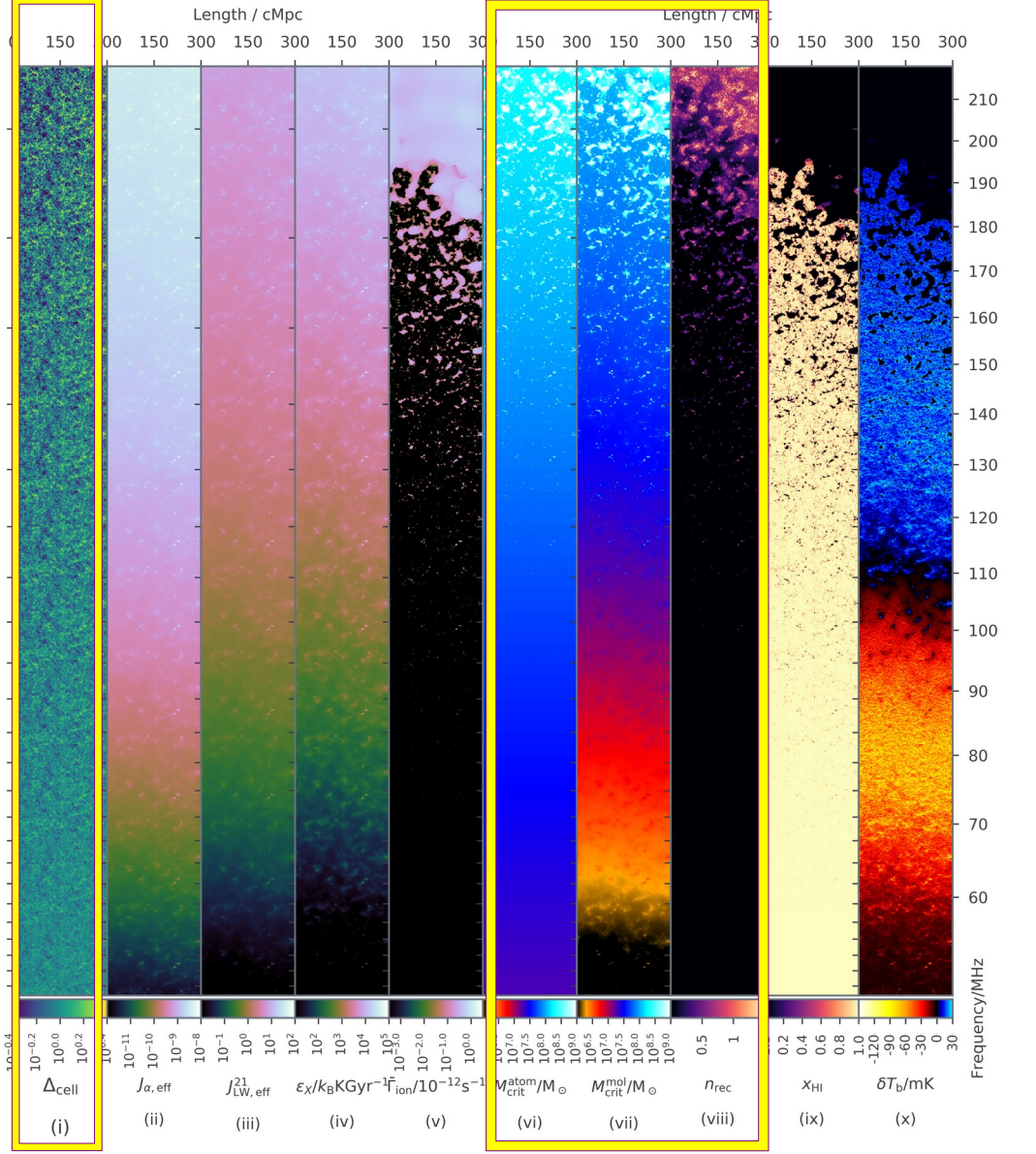
$$T_S^{-1} = \frac{T_{CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}$$

$x_\alpha \propto J_\alpha(z)$

Wouthuysen-Field effect (Lyα pumping)

$$x_c = x_c^{HH} + x_c^{eH} + x_c^{pH}$$

Collision coupling



Mesinger+ (2016)

$\delta(\mathbf{r}, z)$: Dark matter density fluctuations depends on structure formation

$x_{\rm HI}(z)$: Volume neutral fraction depends on SFH in the early Universe

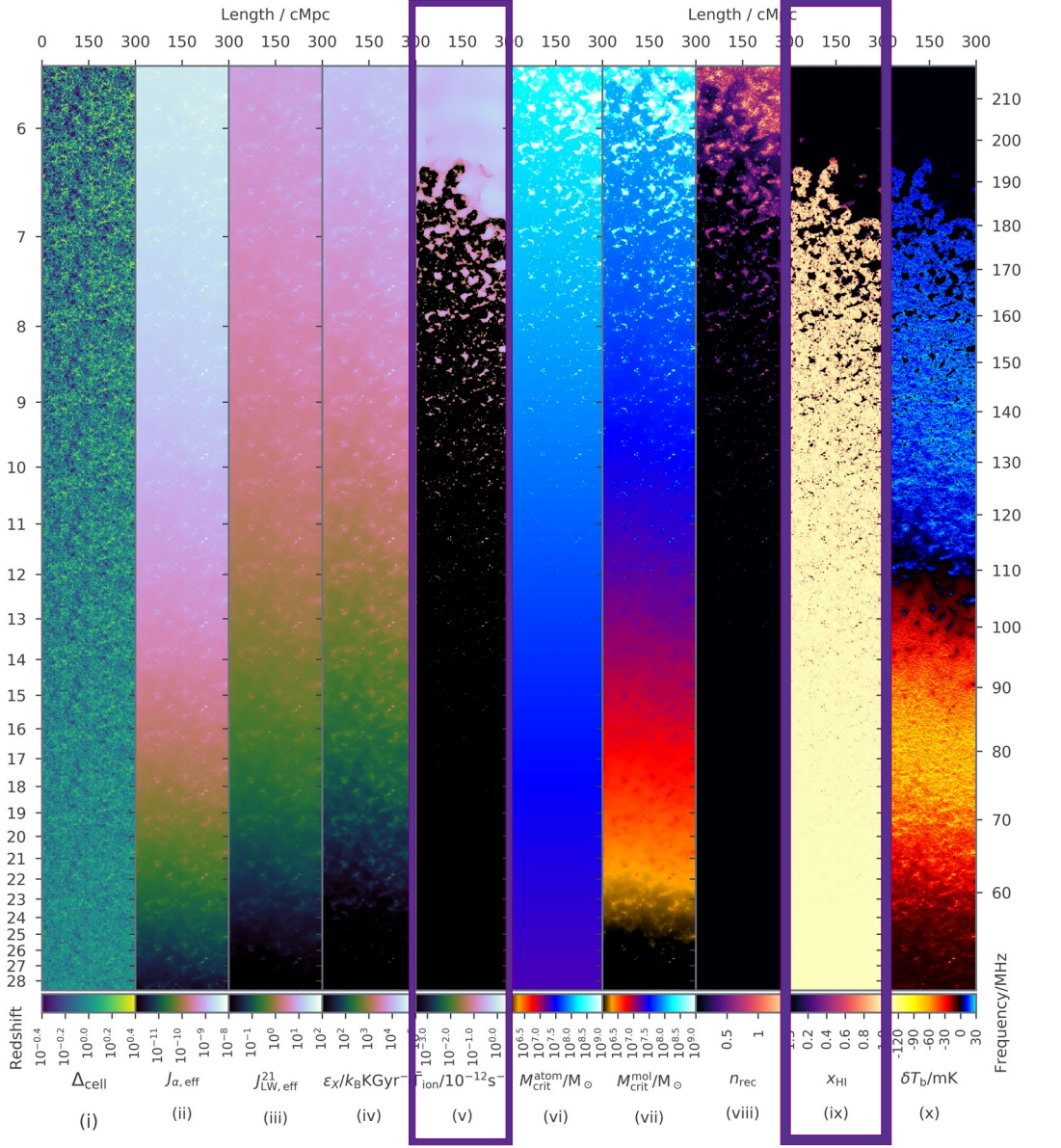
$$T_S^{-1} = \frac{T_{CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}$$

$x_\alpha \propto J_\alpha(z)$

Wouthuysen-Field effect (Lyα pumping)

$x_c = x_c^{HH} + x_c^{eH} + x_c^{pH}$

Collision coupling



Mesinger+ (2016)

$\delta(\mathbf{r}, z)$: Dark matter density : fluctuations depends on structure formation

$x_{\mathrm{HI}}(z)$: Volume neutral fraction depends on SFH in the early Universe

$$T_S^{-1} = \frac{T_{CMB}^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}$$

$x_\alpha \propto J_\alpha(z)$

Wouthuysen-Field effect (Lyα pumping)

$$x_c = x_c^{HH} + x_c^{eH} + x_c^{pH}$$

Collision coupling