How the SKA will revolutionize studies of the first billion years

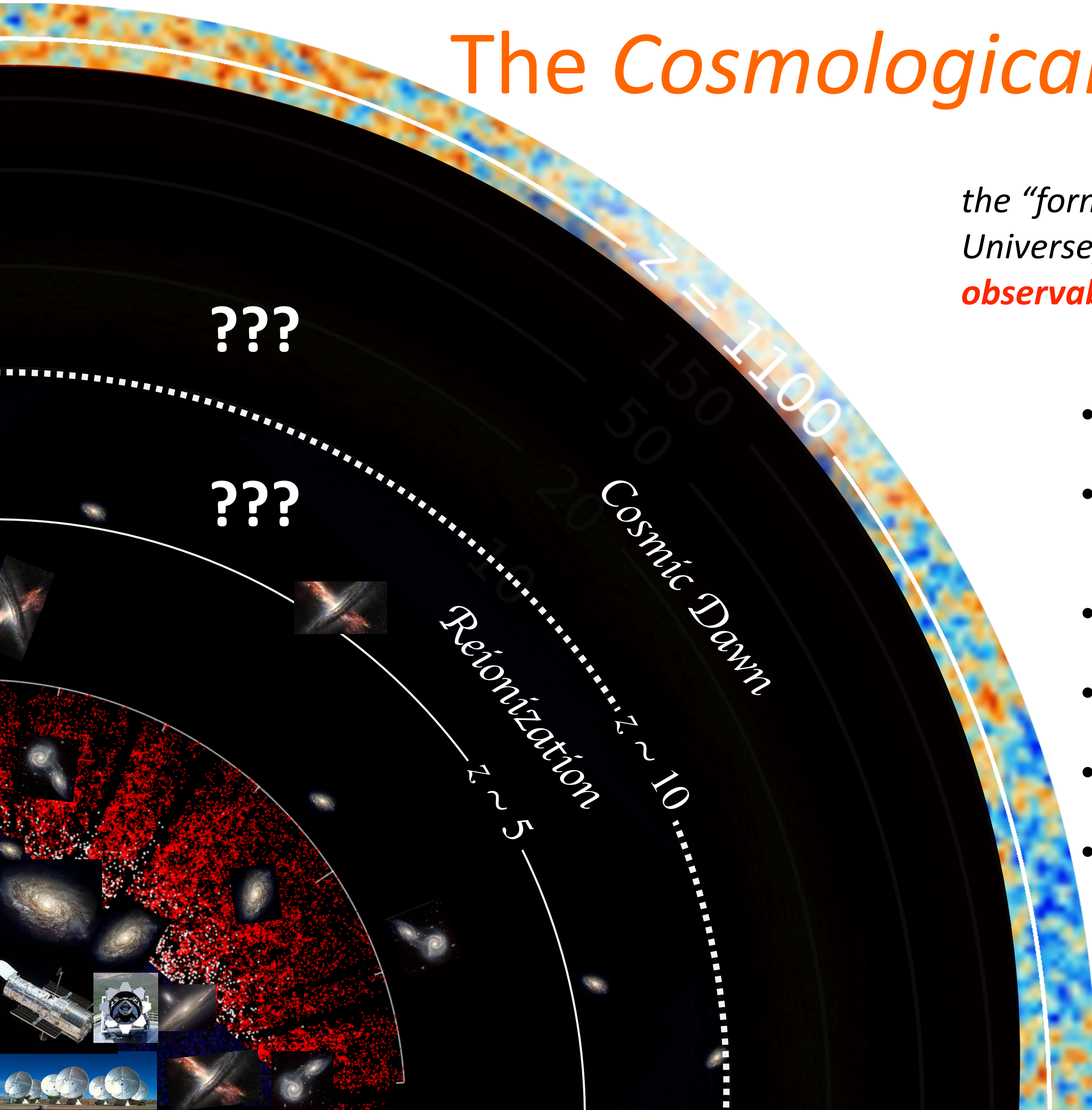
Andrei Mesinger

SCUOLA NORMALE SUPERIORE

MUR Ministero dell'Università e della Ricerca

erc European Research Council

The *Cosmological Frontier…*

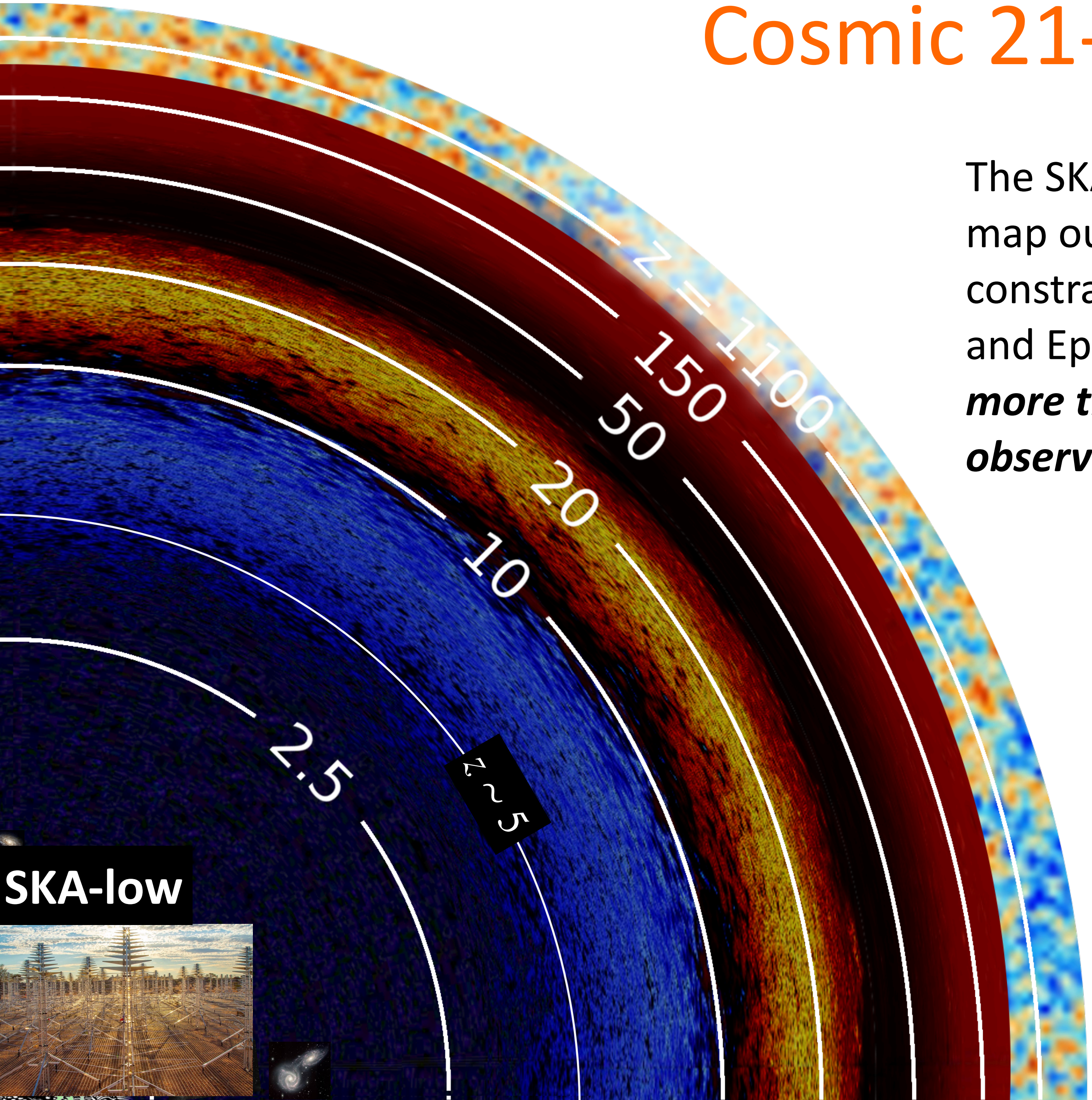


the "formative childhood" of the Universe, yet the ***majority of the observable volume***

- When and how did the first galaxies form?
- How did they impact each other and their surroundings?
- What are the dominant feedback mechanisms?
- Can we learn about Dark Matter properties?
- How does the Hubble parameter evolve?
- What are the properties of the first stars and black holes?

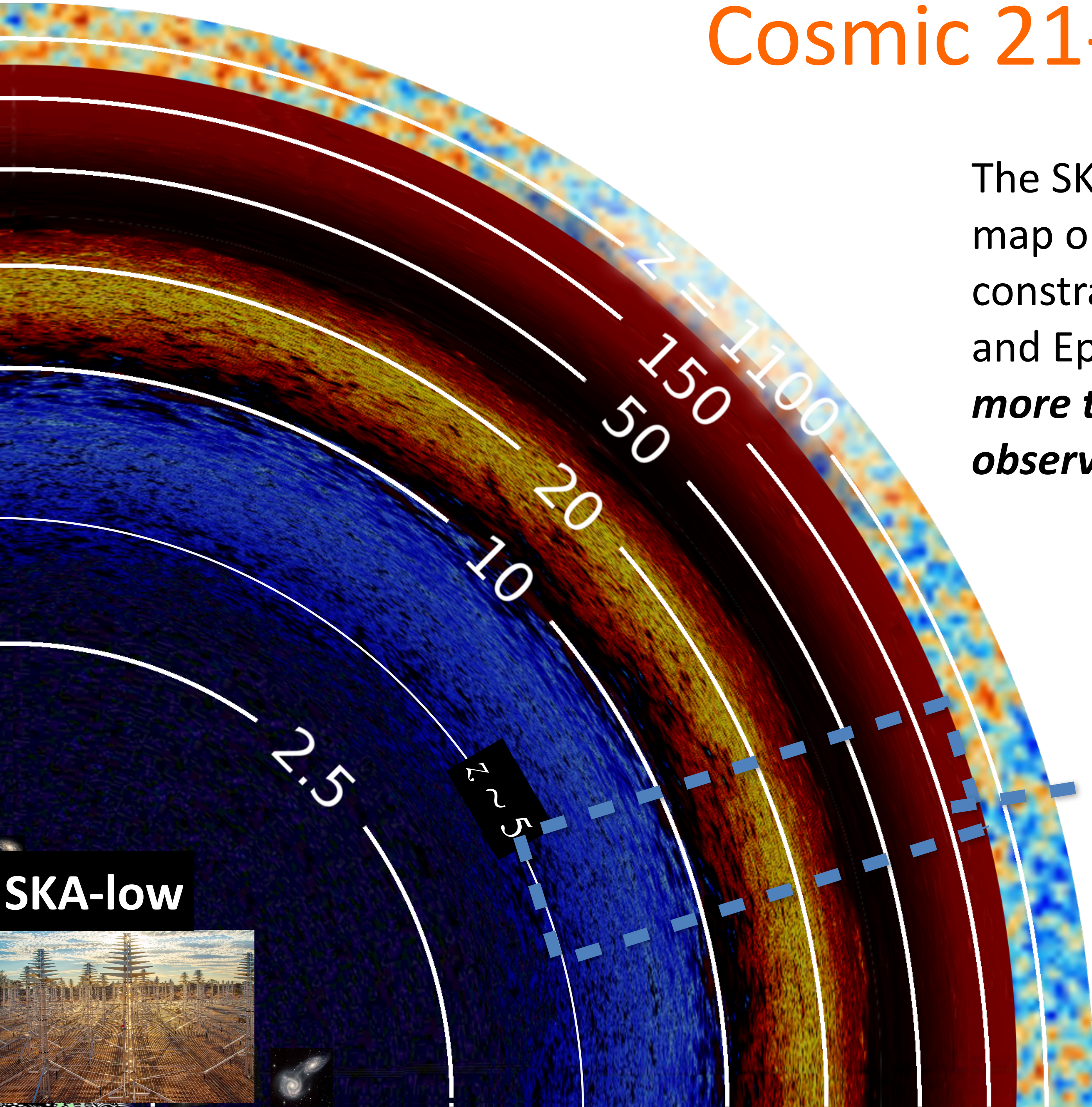
adapted from Cynthia Chiang

Cosmic 21-cm



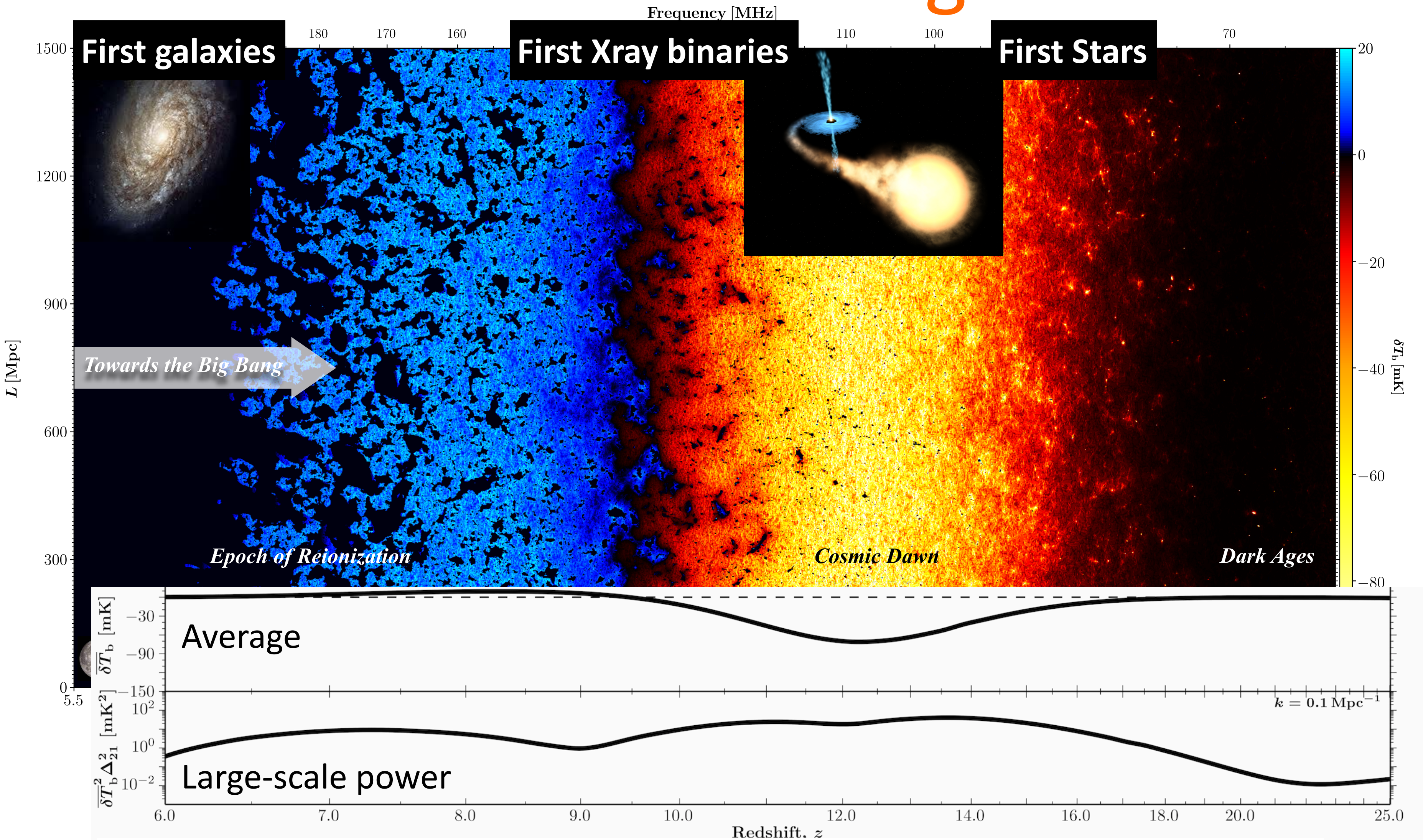
The SKA will eventually map out the poorly constrained Cosmic Dawn and Epoch of Reionization: ***more than 1/2 of our observable Universe***

Cosmic 21-cm



The SKA will eventually map out the poorly constrained Cosmic Dawn and Epoch of Reionization: ***more than 1/2 of our observable Universe***

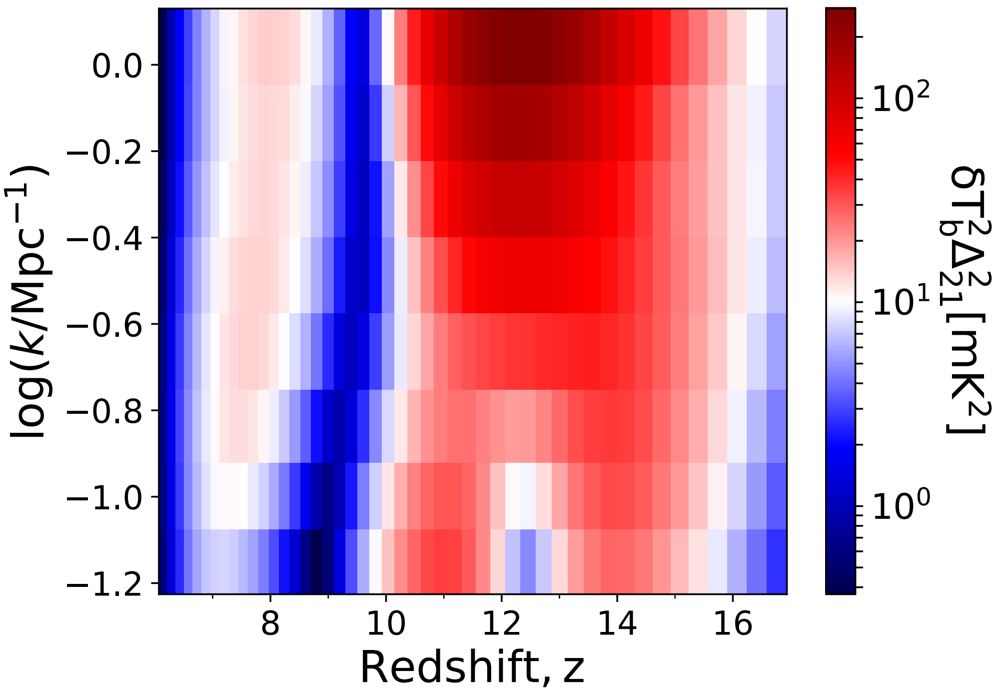
Cosmic 21-cm signal



$$\delta T_b(\nu) \approx 27 x_{HI}(1+\delta_{nl}) \left(\frac{H}{dv_r/dr + H}\right) \left(1 - \frac{T_\gamma}{T_S}\right) \left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)^{1/2} \left(\frac{\Omega_b h^2}{0.023}\right) \mathrm{mK}$$

The SKA will detect the power spectrum of these fluctuations with very high signal to noise

1D power spectrum from "fiducial model"



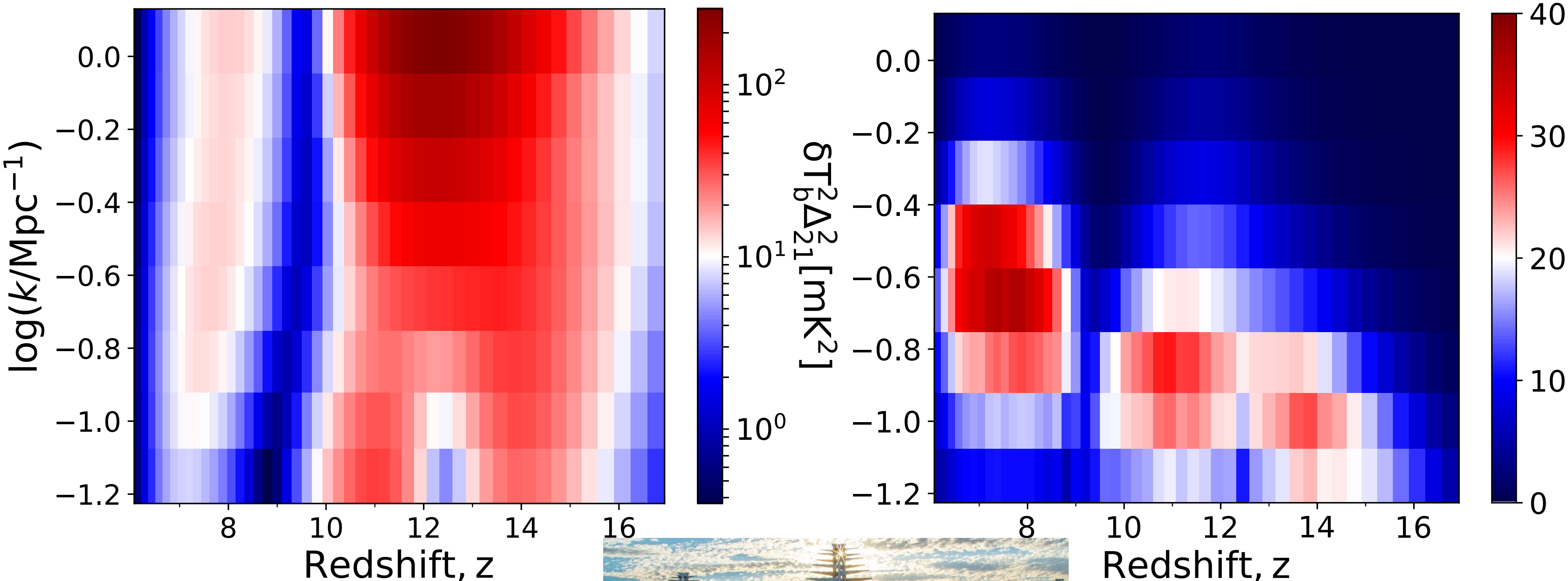
characteristic "three-peak" structure of the cosmic signal

Kaur, Gillet, **AM** (2020)

The SKA will detect the power spectrum of these fluctuations with very high signal to noise

1D power spectrum from "fiducial model"

S/N from a 1000h SKA-low observation

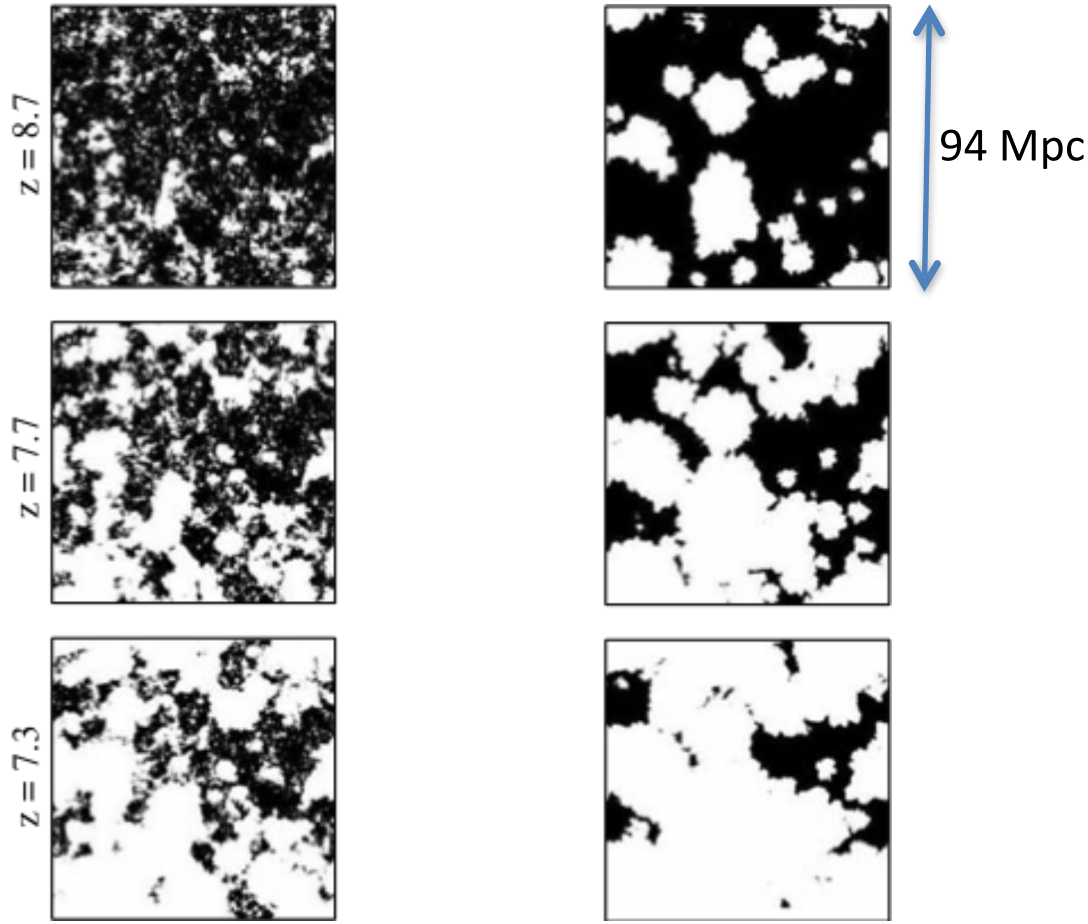


Kaur, Gillet, **AM** (2020)

What can we learn from these patterns?

Timing of reionization and the properties of the (unseen) galaxies that drive it

- Galaxy clustering + stellar properties → *evolution of large-scale EoR/CD structures*

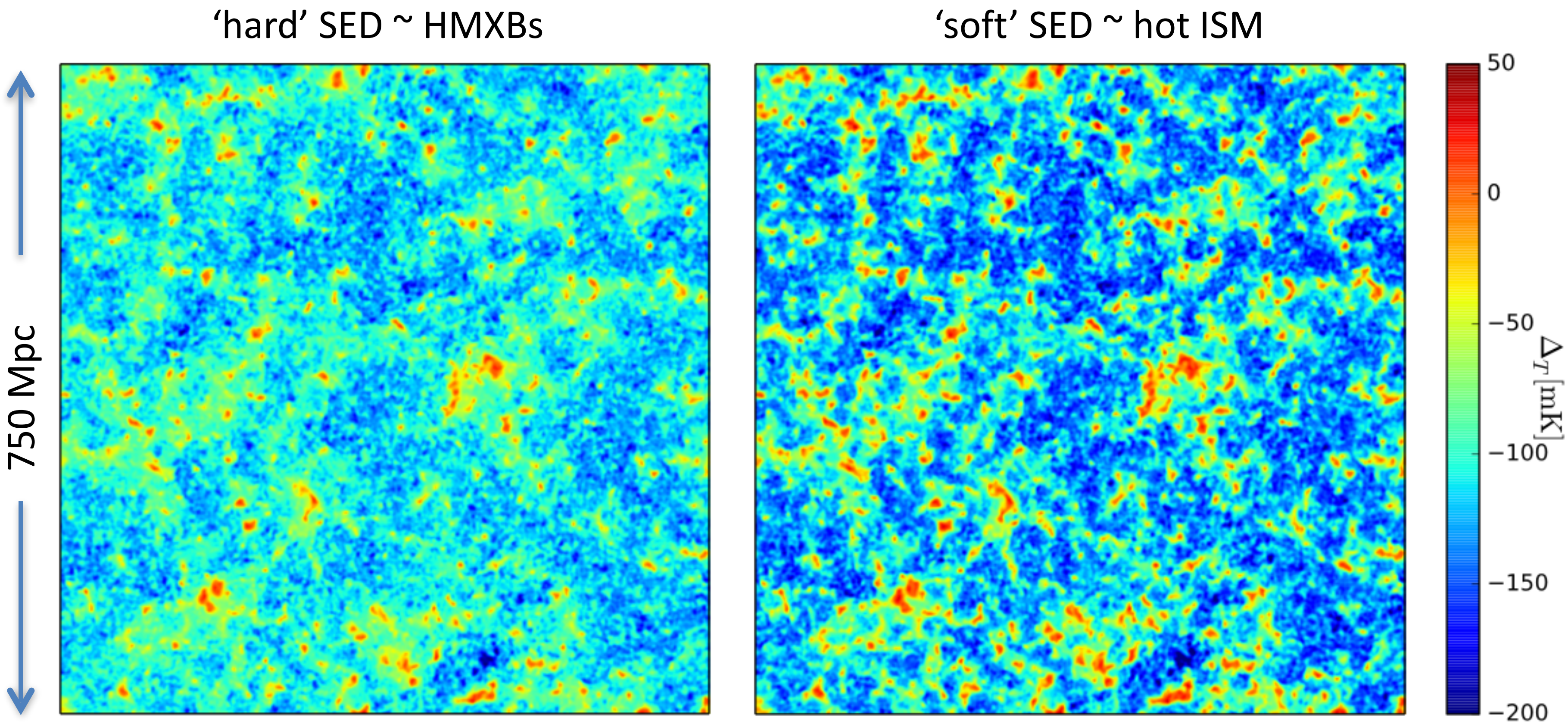


McQuinn+ 2007

Abundant, faint galaxies vs **Rare, bright galaxies**

Patterns in the Epoch of Heating

High-energy processes in the first galaxies are also encoded in the cosmic 21-cm signal



differences are easily detectable with HERA and the SKA

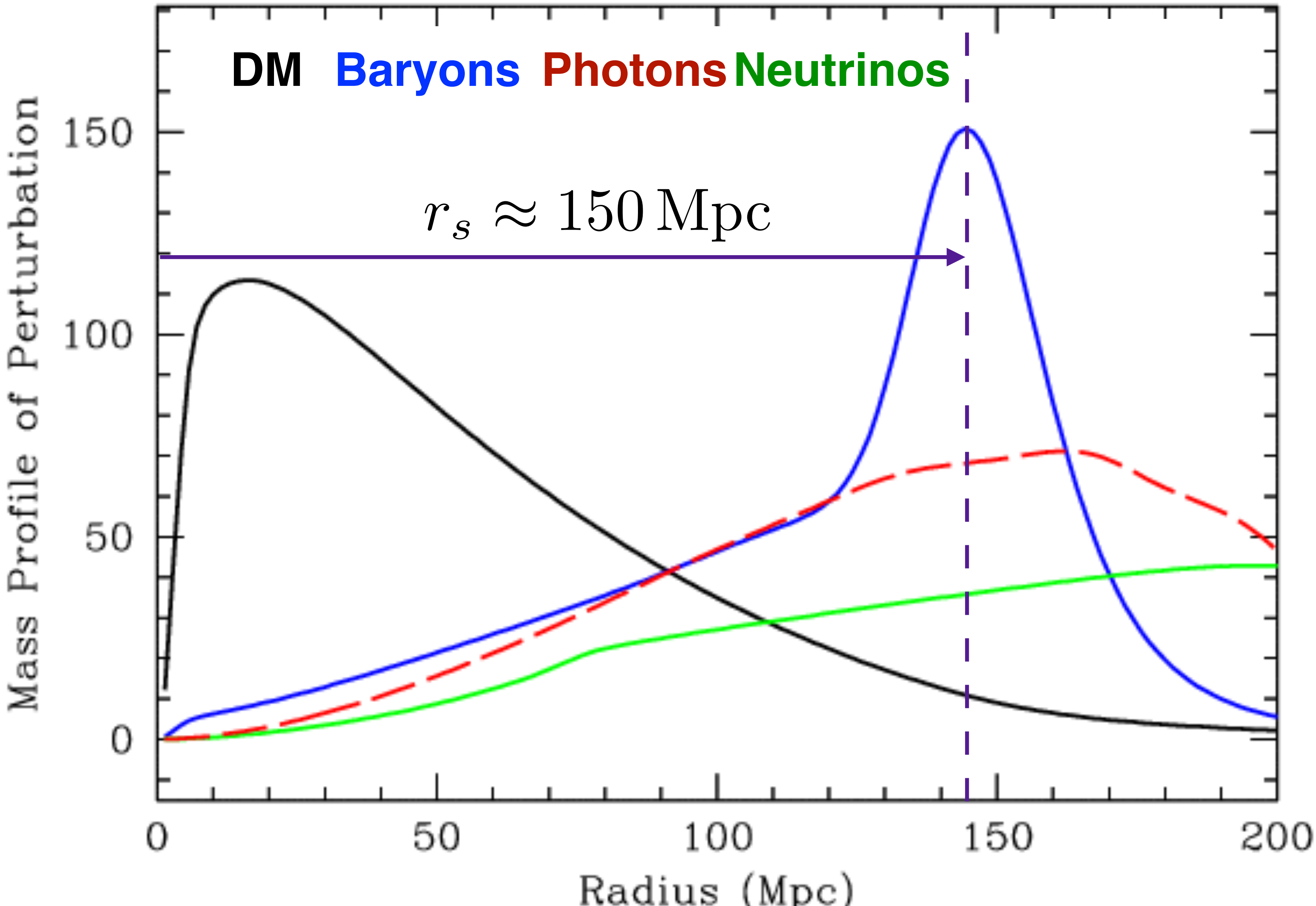
Pacucci, AM + 2014

More exotic sources of early IGM heating?

- Cosmic Rays? (e.g. Leite+2017; Jana and Nath 2018; Gessey-Jones+2023)
- Dark matter annihilations? (e.g. Evoli+2014; Lopez-Honorez+2016)
- Dark matter decay? (e.g. Facchinetti+ 2023)

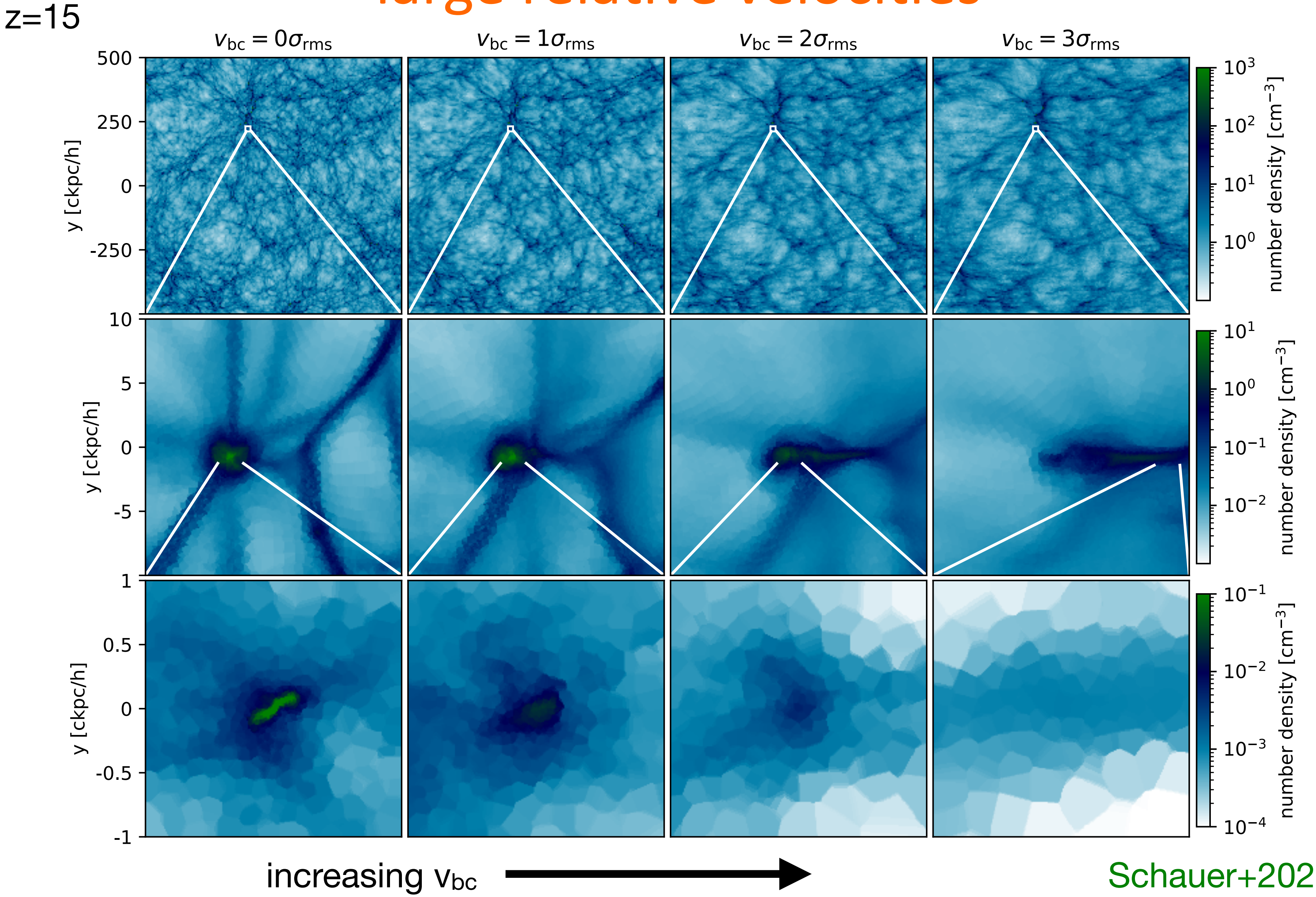
All have different spatial signature

Baryon Acoustic Oscillations



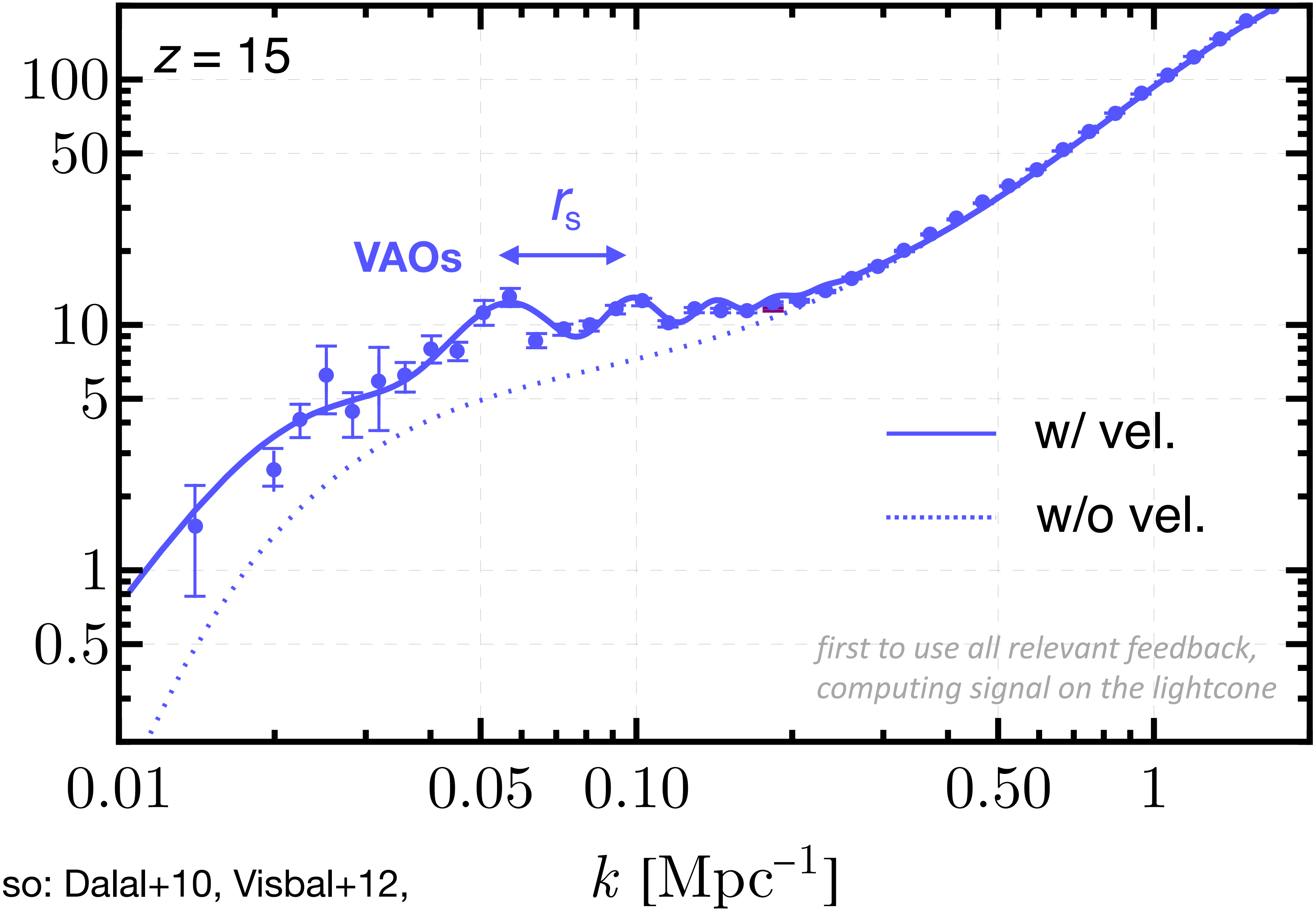
Credit: Daniel Eisenstein

Star formation is suppressed in regions with large relative velocities



Schauer+2021

Standard ruler

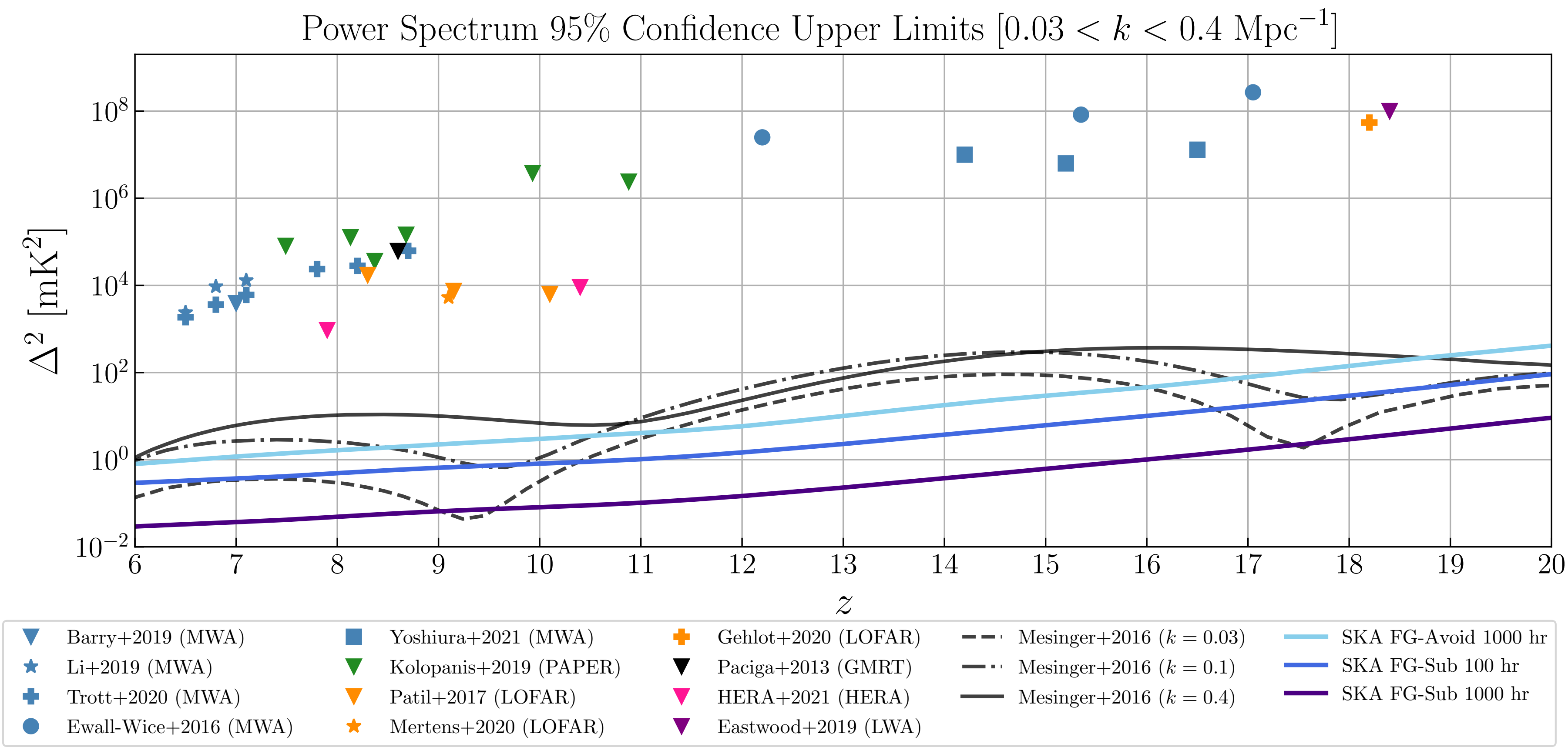


See also: Dalal+10, Visbal+12,
Fialkov+12, McQuinn+12
Munoz 19, Park+19, Cain+20, Sarkar+22

Munoz, Qin, AM+ 2022

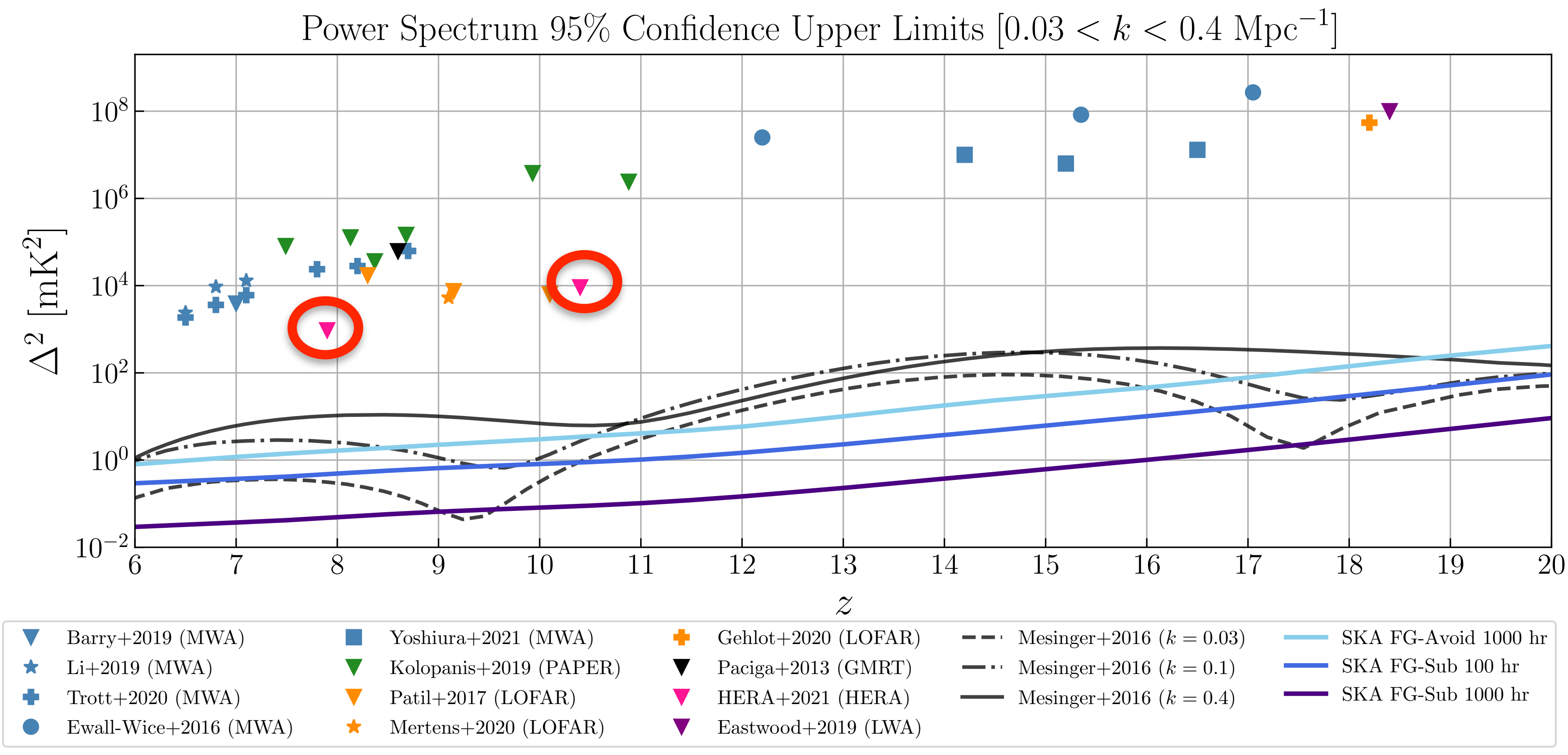
That sounds great, but where are we now?

Measurements are improving, but currently only upper limits on the PS



Barry+ 2022

Currently only upper limits on the PS



Application to **HERA** (HERA collaboration 2022ab).
For similar studies on **LOFAR** and **MWA** data see (Ghara+2020; Mondal+2020; Greig+2020, Greig+2021)

What kind of models are the easiest to rule out (i.e. have the largest power)?

$$\delta T_b(\nu) \approx 27 x_{\rm HI}(1+\delta_{\rm nl})\left(\frac{H}{dv_r/dr + H}\right)\left(1 - \frac{T_\gamma}{T_S}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)^{1/2}\left(\frac{\Omega_b h^2}{0.023}\right) \mathrm{mK}$$

What kind of models are the easiest to rule out (i.e. have the largest power)?

$$\delta T_b(\nu) \approx 27 x_{HI}(1+\delta_{nl})\left(\frac{H}{dv_r/dr + H}\right)\left(1 - \frac{T_\gamma}{T_S}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)^{1/2}\left(\frac{\Omega_b h^2}{0.023}\right) \text{mK}$$

~ 0 — 1

What kind of models are the easiest to rule out (i.e. have the largest power)?

$$\delta T_b(\nu) \approx 27 x_{HI}(1+\delta_{nl})\left(\frac{H}{dv_r/dr + H}\right)\left(1 - \frac{T_\gamma}{T_S}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)^{1/2}\left(\frac{\Omega_b h^2}{0.023}\right) \mathrm{mK}$$

~ 0.1 — 1

What kind of models are the easiest to rule out (i.e. have the largest power)?

$$\delta T_b(\nu) \approx 27 x_{HI}(1+\delta_{nl})\left(\frac{H}{dv_r/dr + H}\right)\left(1 - \frac{T_\gamma}{T_S}\right)\left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)^{1/2}\left(\frac{\Omega_b h^2}{0.023}\right) \text{mK}$$

~ **-10(!)** — 1

What kind of models are the easiest to rule out (i.e. have the largest power)?

$$\delta T_b(\nu) \approx 27 x_{HI}(1+\delta_{nl}) \left(\frac{H}{dv_r/dr + H}\right) \left(1 - \frac{T_\gamma}{T_S}\right) \left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)^{1/2} \left(\frac{\Omega_b h^2}{0.023}\right) \mathrm{mK}$$

Models that are ruled out must have:

COLD IGM: $T_S \ll T_\gamma$

What kind of models are the easiest to rule out (i.e. have the largest power)?

$$\delta T_b(\nu) \approx 27 x_{HI}(1+\delta_{nl}) \left(\frac{H}{dv_r/dr + H}\right) \left(1 - \frac{T_\gamma}{T_S}\right) \left(\frac{1+z}{10}\frac{0.15}{\Omega_M h^2}\right)^{1/2} \left(\frac{\Omega_b h^2}{0.023}\right) \mathrm{mK}$$

Models that are ruled out must have:

COLD IGM: $T_S \ll T_\gamma$

+

Spatial fluctuations *in either:*

- ***ionization fraction (patchy EoR)***
- ***matter density***
- *temperature (requires extremely soft SEDs)*

see also e.g. Ewall-Wice+2013; Ghara+2020; Greig+2020; Mondal+2020; Reis+2020; Greig+2021

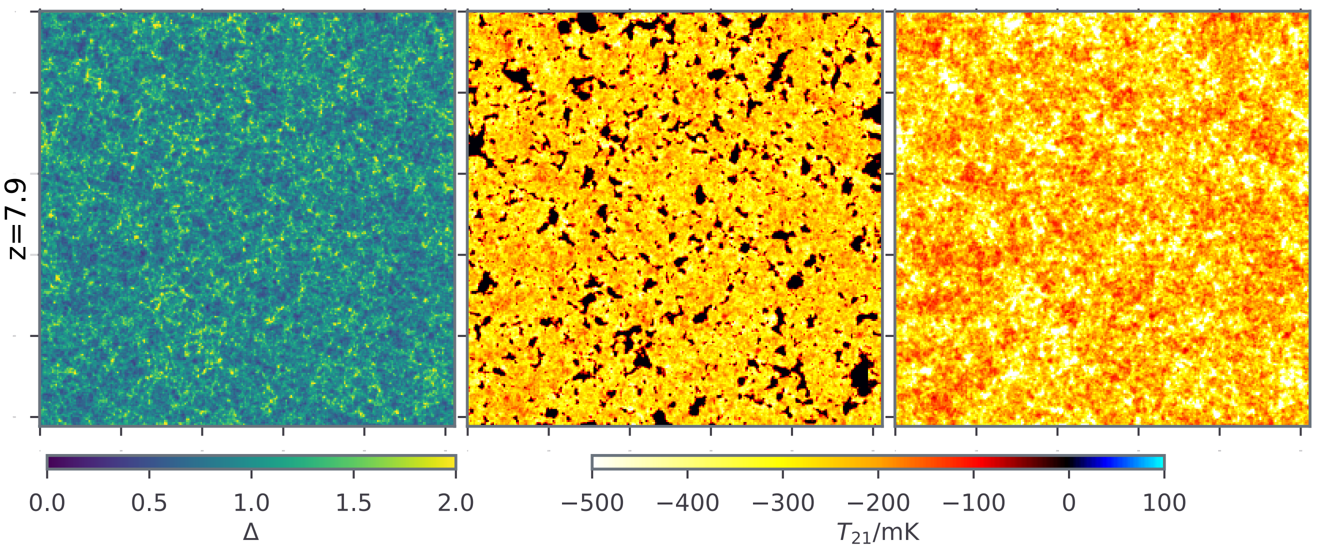
Examples

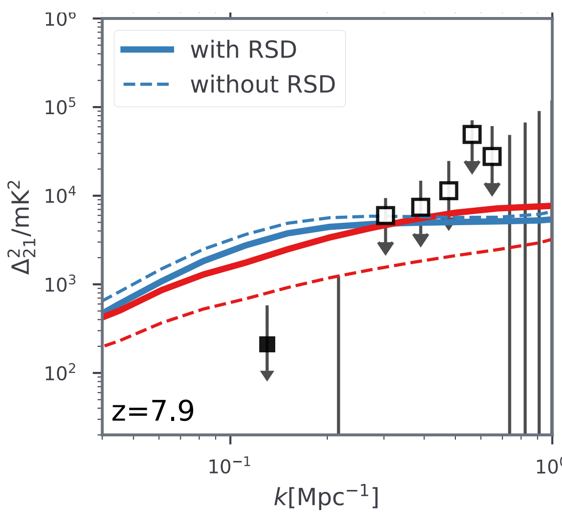
Density

21cm
COLD + EoR

21cm
COLD + density

21cm power





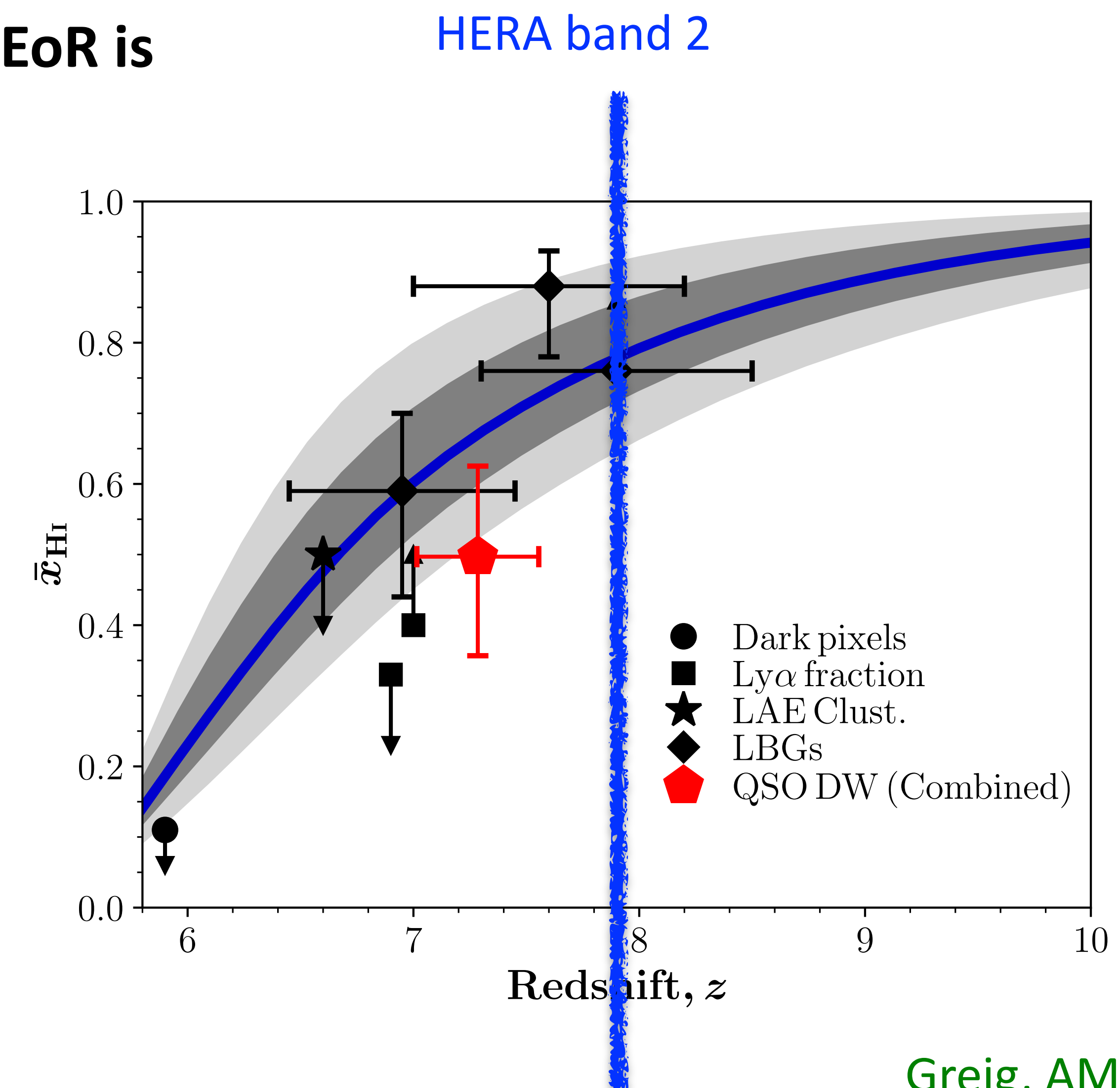
$\bar{x}_{\rm HI} \sim 0.5$

$\bar{x}_{\rm HI} \sim 1$

HERA collaboration (2021)

Current constraints on EoR history

BUT we know the EoR is underway at z~8!



Greig, AM+2021

Current constraints on EoR history

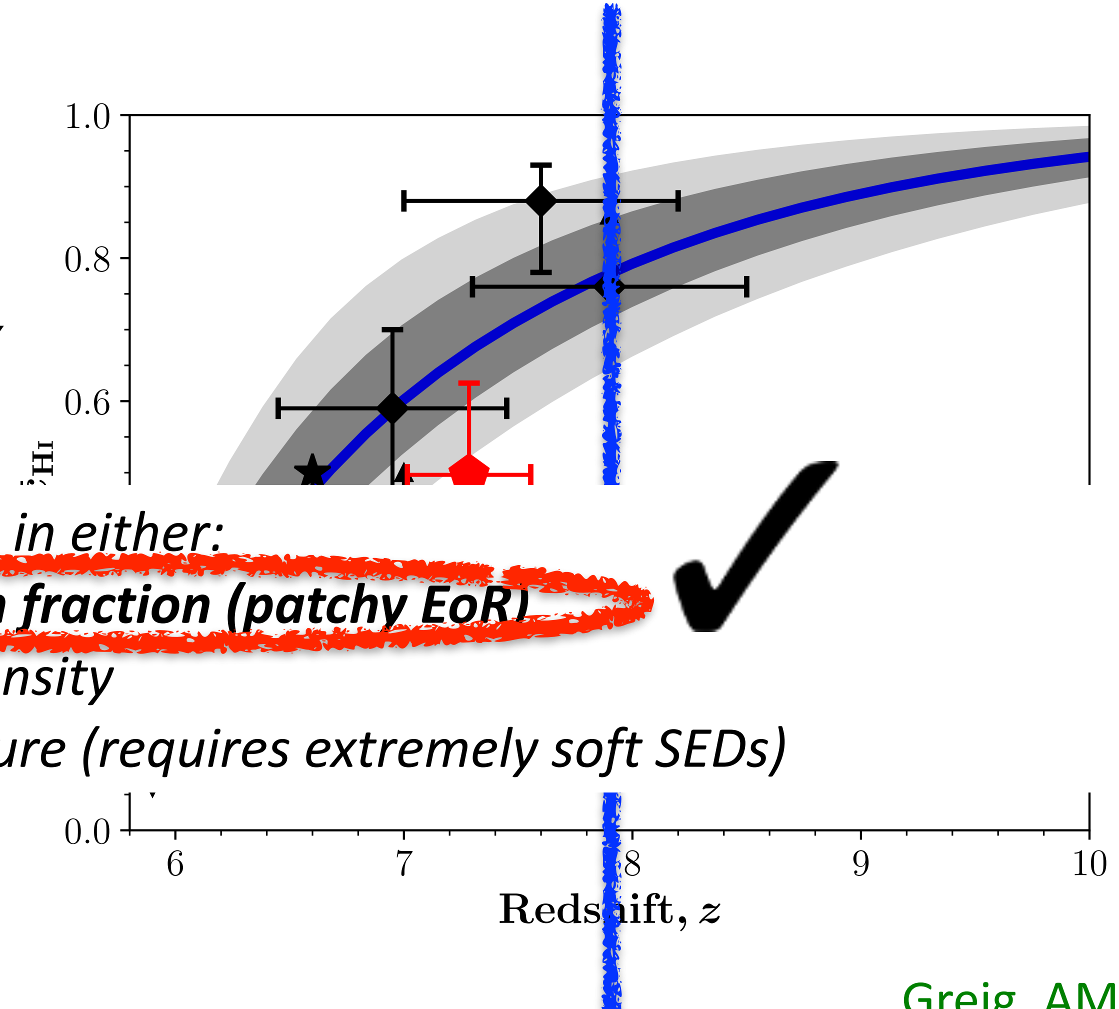
BUT we know the EoR is underway at z~8!

COLD IGM: $T_S \ll T_\gamma$

+

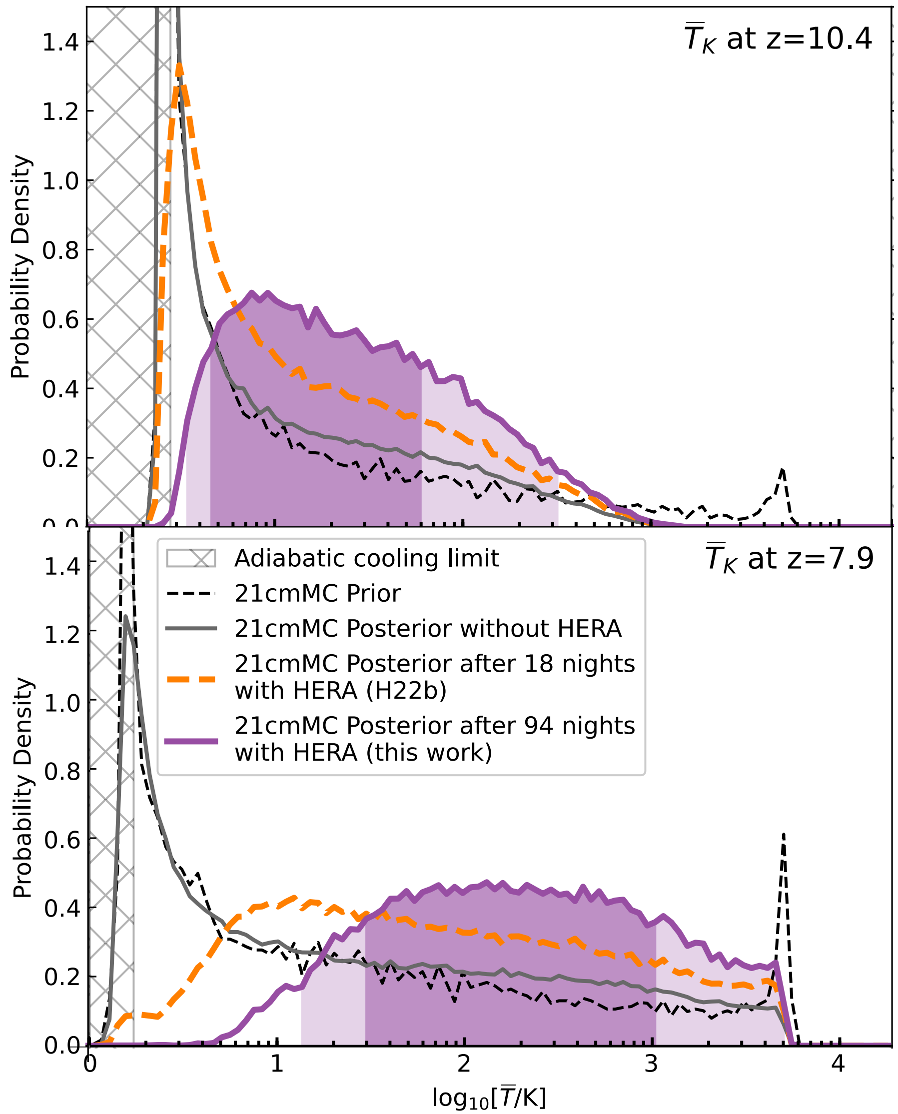
Spatial fluctuations *in either:*

- ***ionization fraction (patchy EoR)*** ✓
- *matter density*
- *temperature (requires extremely soft SEDs)*



Greig, AM+2021

Constraints on IGM properties

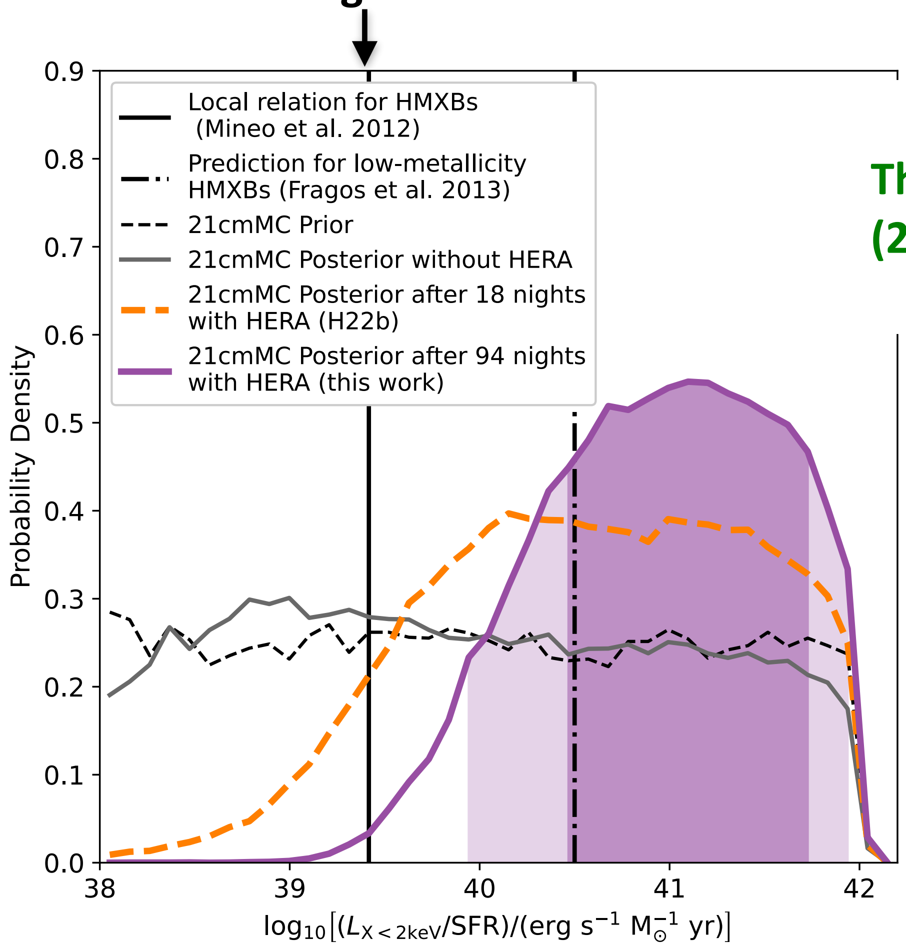


Something must have heated the IGM at z>10!!

The HERA collaboration (2023)

If heating is provided by "normal" galaxies, they would need to be more luminous in X-rays than observed locally

Local galaxies



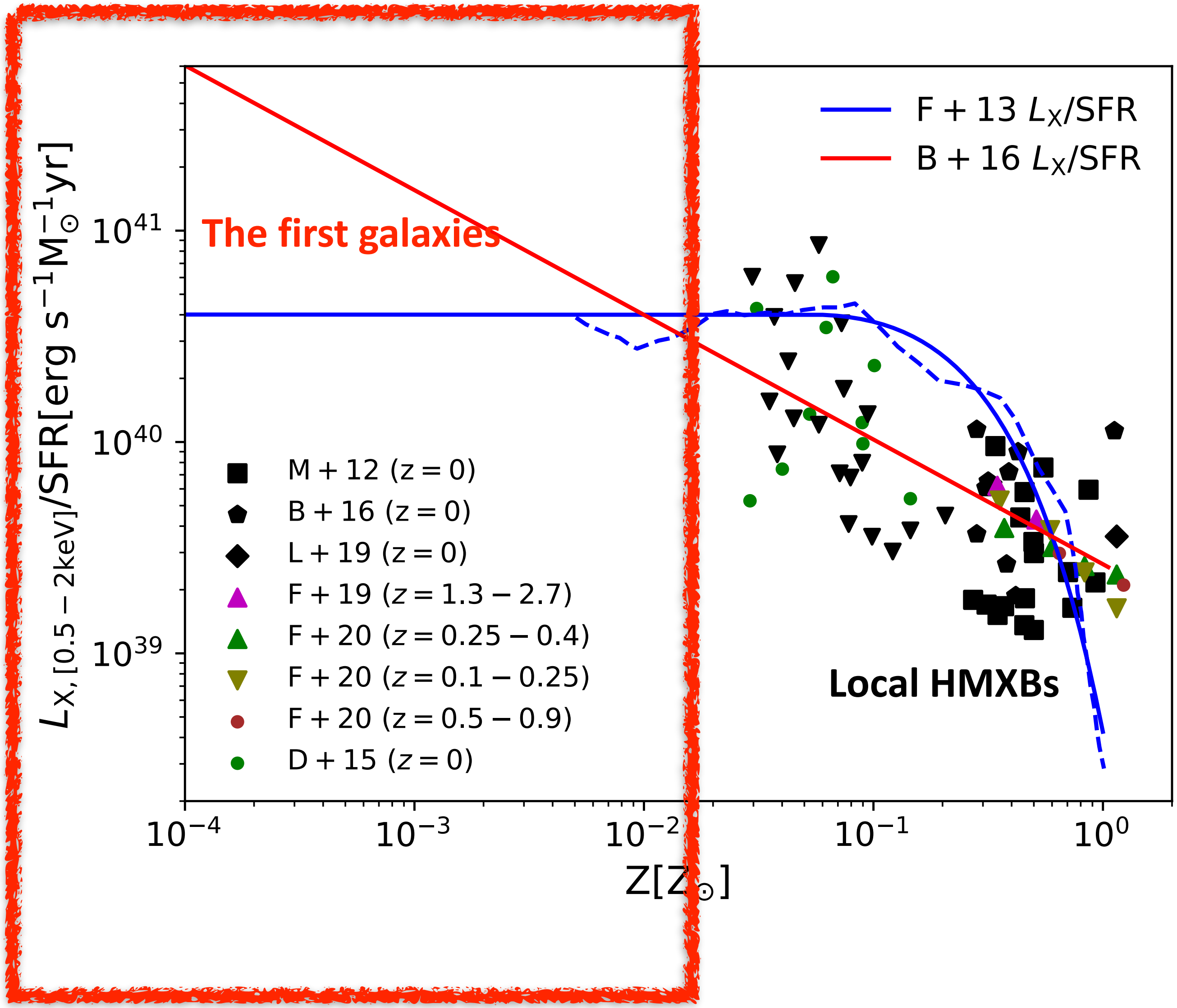
The HERA collaboration (2023) - 94 nights of data

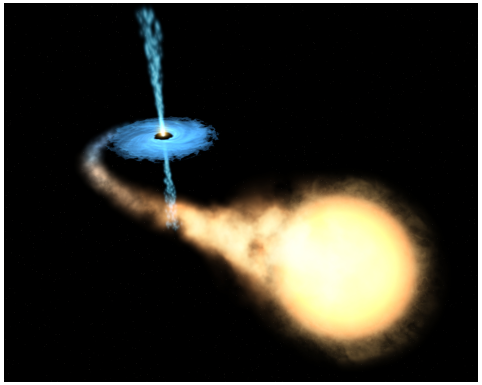
HERA is the first observation to constrain the X-ray luminosities of Cosmic Dawn galaxies (e.g., Fragos+13), *disfavoring the values seen in local, metal-enriched galaxies*

The HERA collaboration (2023; *led by J. Dillon*)

Is this surprising?

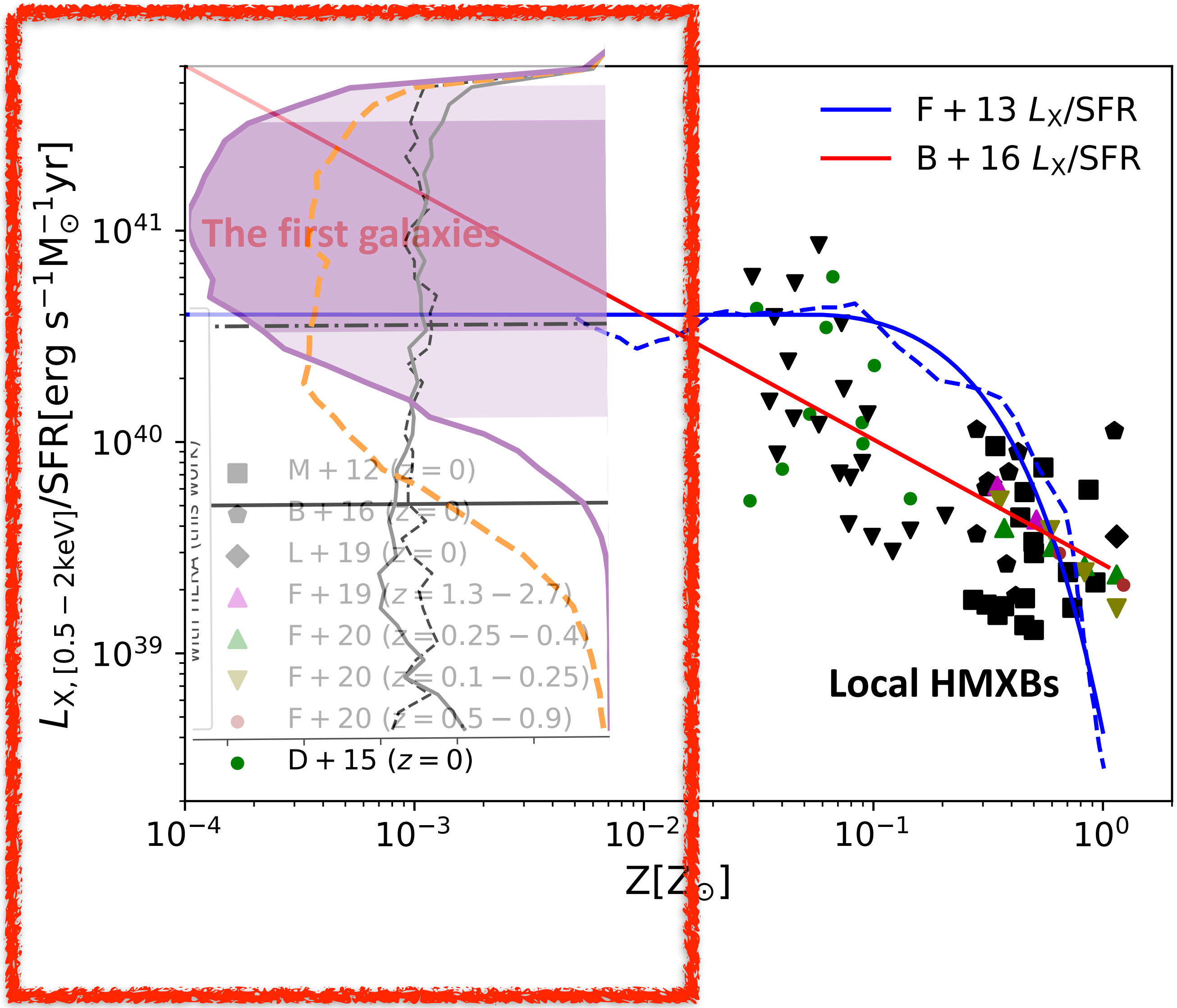
The 21-cm signal probes a new regime for HMXBs: *low mass galaxies + low metalicity*

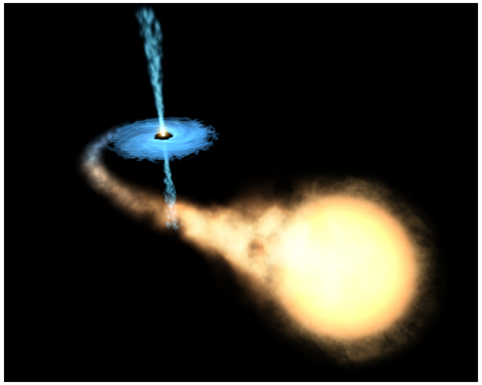




Kaur, Qin, AM+ (2022)

The 21-cm signal probes a new regime for HMXBs: *low mass galaxies + low metalicity*

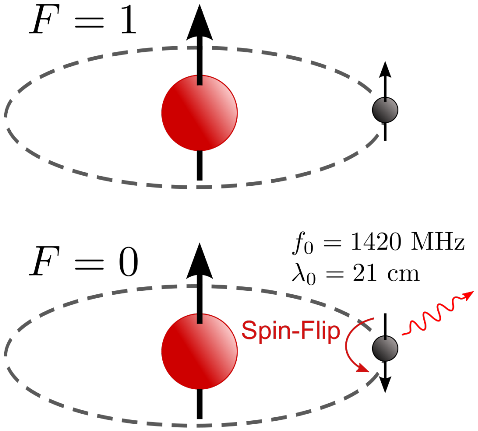




Kaur, Qin, AM+ (2022)

Milestones

aka "The path to the 21-cm revolution"



Where we are now

Upper limits on the 21-cm power spectrum

Where we are now

Upper limits on the 21-cm power spectrum

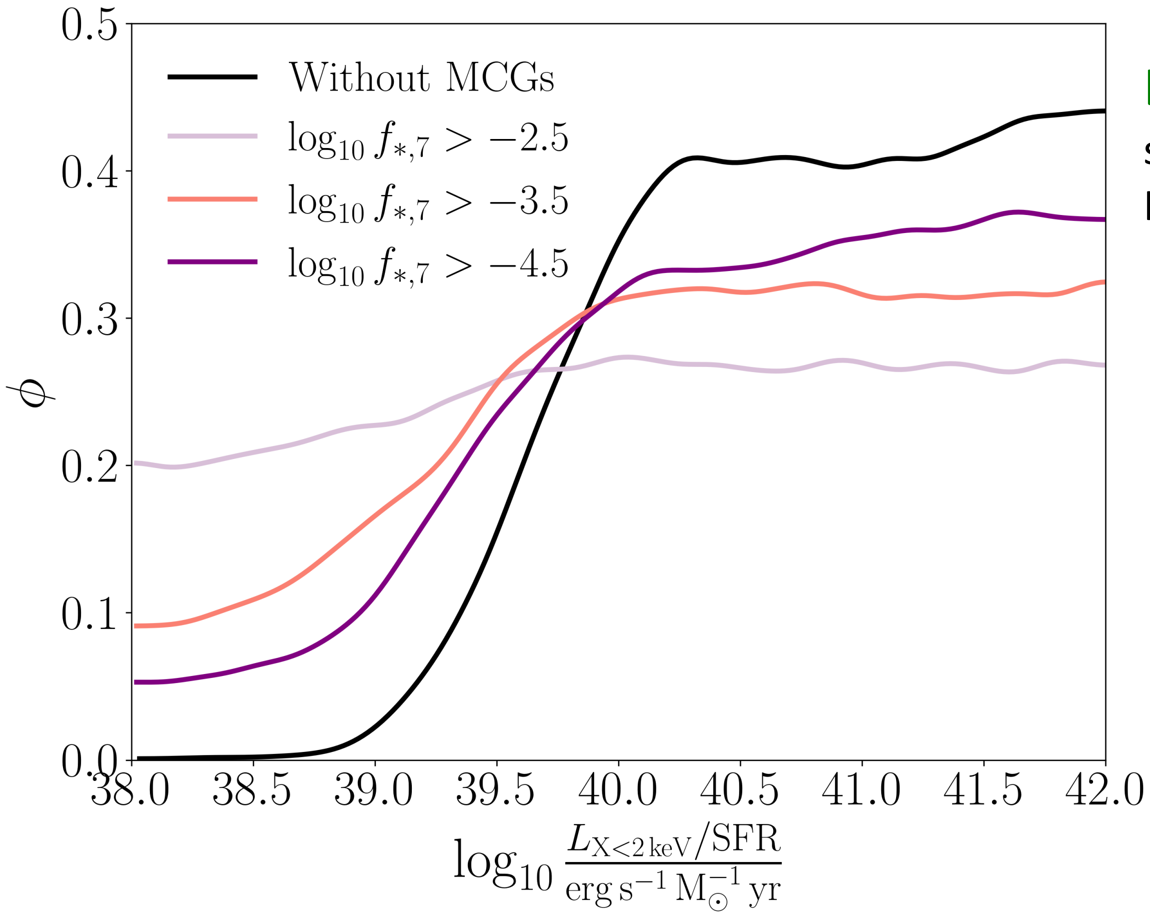
- *understand **systematics**! can we parametrize / sample our uncertainties?*

Where we are now

Upper limits on the 21-cm power spectrum

- *understand* ***systematics****! can we parametrize / sample our uncertainties?*
- *do we have all of the* ***physics*** *we need, especially regarding heating sources?*

Including a contribution from even earlier, molecularly-cooled galaxies (MCGs)?



Lazare+(2023)
see also Qin, AM+2021;
HERA 2022a

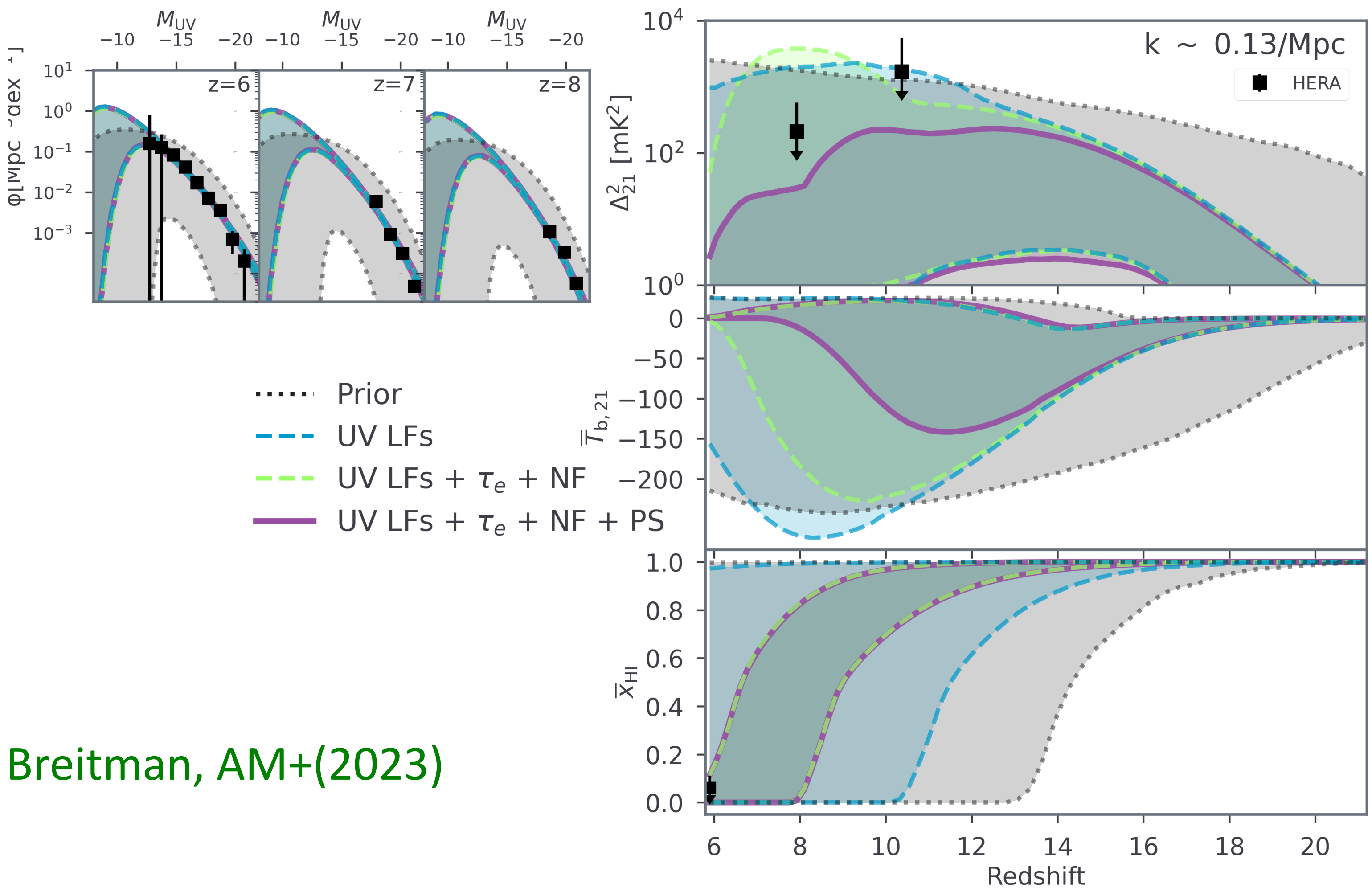
Constraints from HERA can weaken, though results depend strongly on priors

Where we are now

Upper limits on the 21-cm power spectrum

- *understand* ***systematics****! can we parametrize / sample our uncertainties?*
- *do we have all of the* ***physics*** *we need, especially regarding heating sources?*
- *posteriors will be* ***prior-dominated*** *UNLESS we have "realistic" galaxy models that can be constrained by other observations*

Contribution of different data



Breitman, AM+(2023)

Where we are now

Upper limits on the 21-cm power spectrum

- *understand* ***systematics****! can we parametrize / sample our uncertainties?*
- *do we have all of the* ***physics*** *we need, especially regarding heating sources?*
- *posteriors will be* ***prior-dominated*** *UNLESS we have "realistic" galaxy models that can be constrained by other observations*
- ***emulators*** *are useful! error is currently sub-dominant*

  (e.g. Kern+2017; Schmit & Pritchard 2017; Shimabukuro & Semelin 2017; Jennings+2019; Ghara+2020; Mondal+2022; Bye+2022a; Lazare+2023; Breitman, ***AM****+2023)*

Where we will be soon

Low S/N detection of the 21-cm PS

Where we will be soon

Low S/N detection of the 21-cm PS

- *understand **systematics**! can we parametrize / sample our uncertainties?*

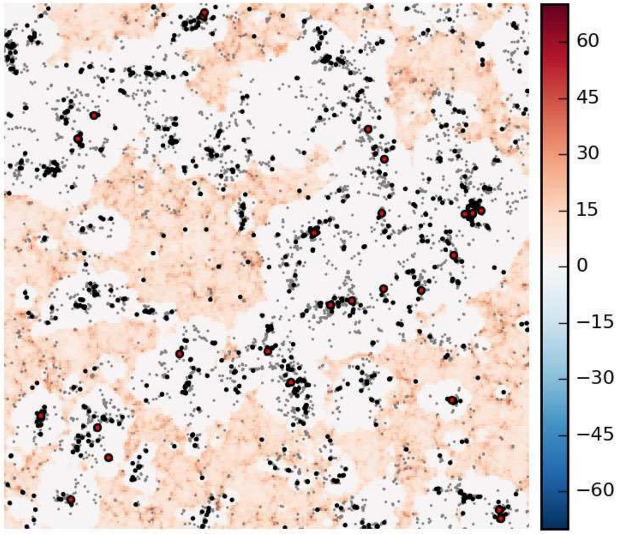
Where we will be soon

Low S/N detection of the 21-cm PS

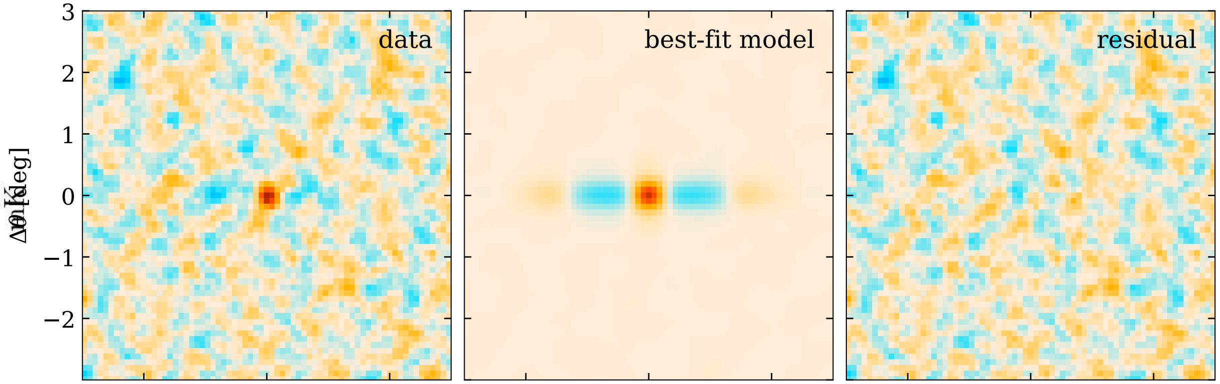
- *understand* ***systematics****! can we parametrize / sample our uncertainties?*
- *how can we convince ourselves and everyone else that the detection is* ***REAL —> cross-correlation with signal of known cosmic origin***

The importance of cross-correlations

- It is an important sanity check to verify claims of detection/analysis pipeline
- improves S/N for preliminary detections (systematics and noise are uncorrelated in cross)
- with images, it lets us study individual HII (or heated) regions, comparing them to their host galaxy properties



Moriwaki+2019



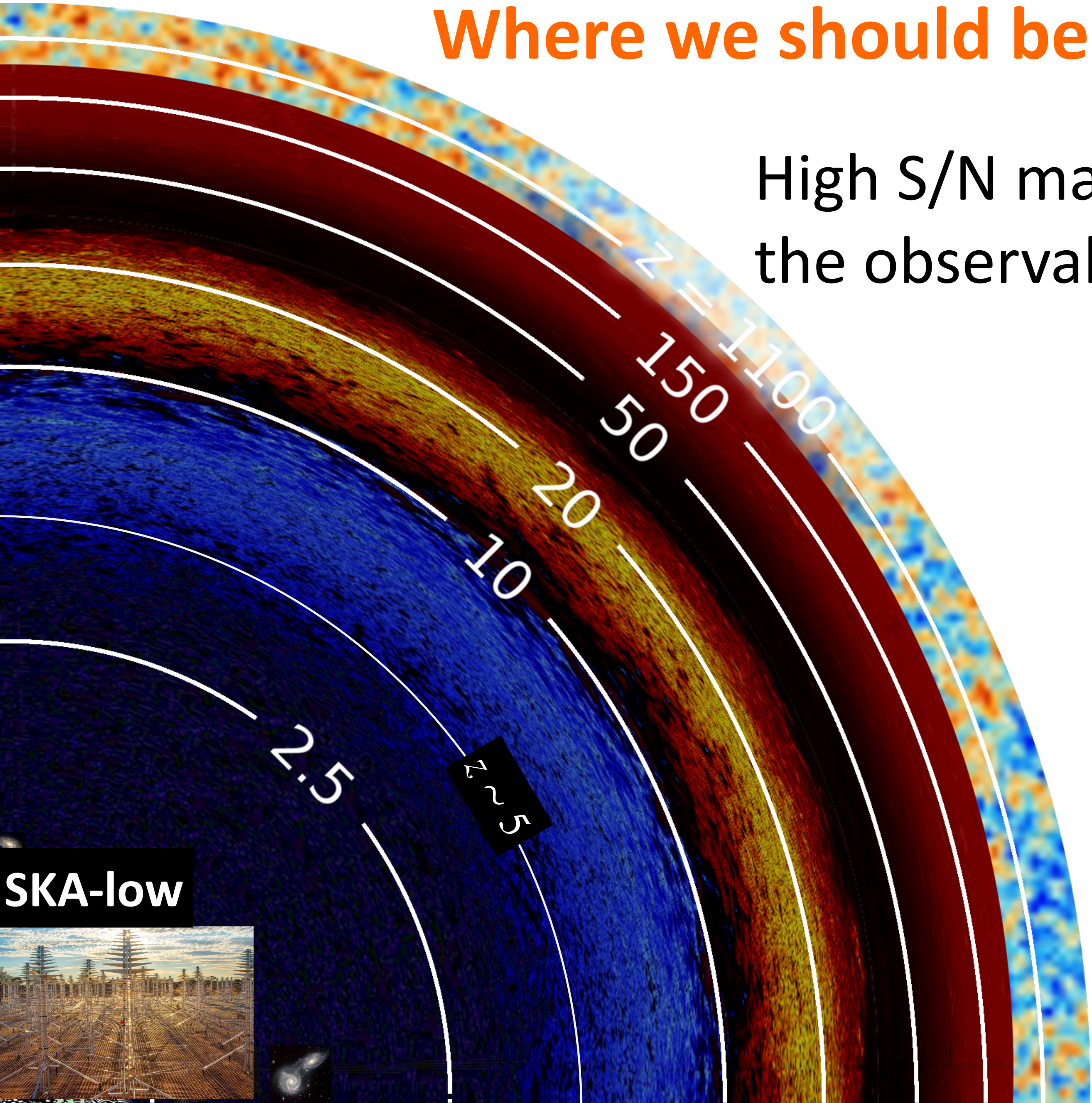
The CHIME collaboration 2022

Signals to cross with 21cm during EoR/CD

1. **Cosmic Backgrounds** (difficult to get good S/N because signal integrates over redshift)
   (i) CMB (e.g. kSZ with SPT/ACT/SO; e.g. Ma+2018; LaPlante+2022)
   (ii) NIR (e.g. CIBERII Mao 2014)
   (iii) XRB (Athena) e.g. Ma+2018
2. Resolved **Galaxies** (need wide and deep, and redshifts to better than percent precision-> grism or multi-object spectroscopy)
   (i) ROMAN grism (e.g. Vrbanec+2020; LaPlante+2023)
   (ii) SUBARU narrow-band (e.g. Sobacchi+ 2016; Vrbanec+2020; Hutter+2017; Kubota+ 2020; Heneka & Mesinger 2020);
   (iii) SUBARU spectroscopy with PFS
   (iv) ELT spectroscopy (Gagnon Hartman+ in prep)
3. **Intensity mapping** (best footprint overlap; signal is generally faint at z>6)
   (i) Lya - SPHEREx (e.g. Heneka & Cooray 2021) CDIM (Cooray+2016)
   (ii) OIII - SPHEREx (Kana+ 2019; Moriwaki+2019; Schengqi+2021)
   (iii) CII - CONCERTO (Lagache+2017), TIME-Pilot (Crites+2014), CCAT-prime (Parshley+2018)

Where we should be >2030-2040

High S/N map of ~50% of the observable Universe

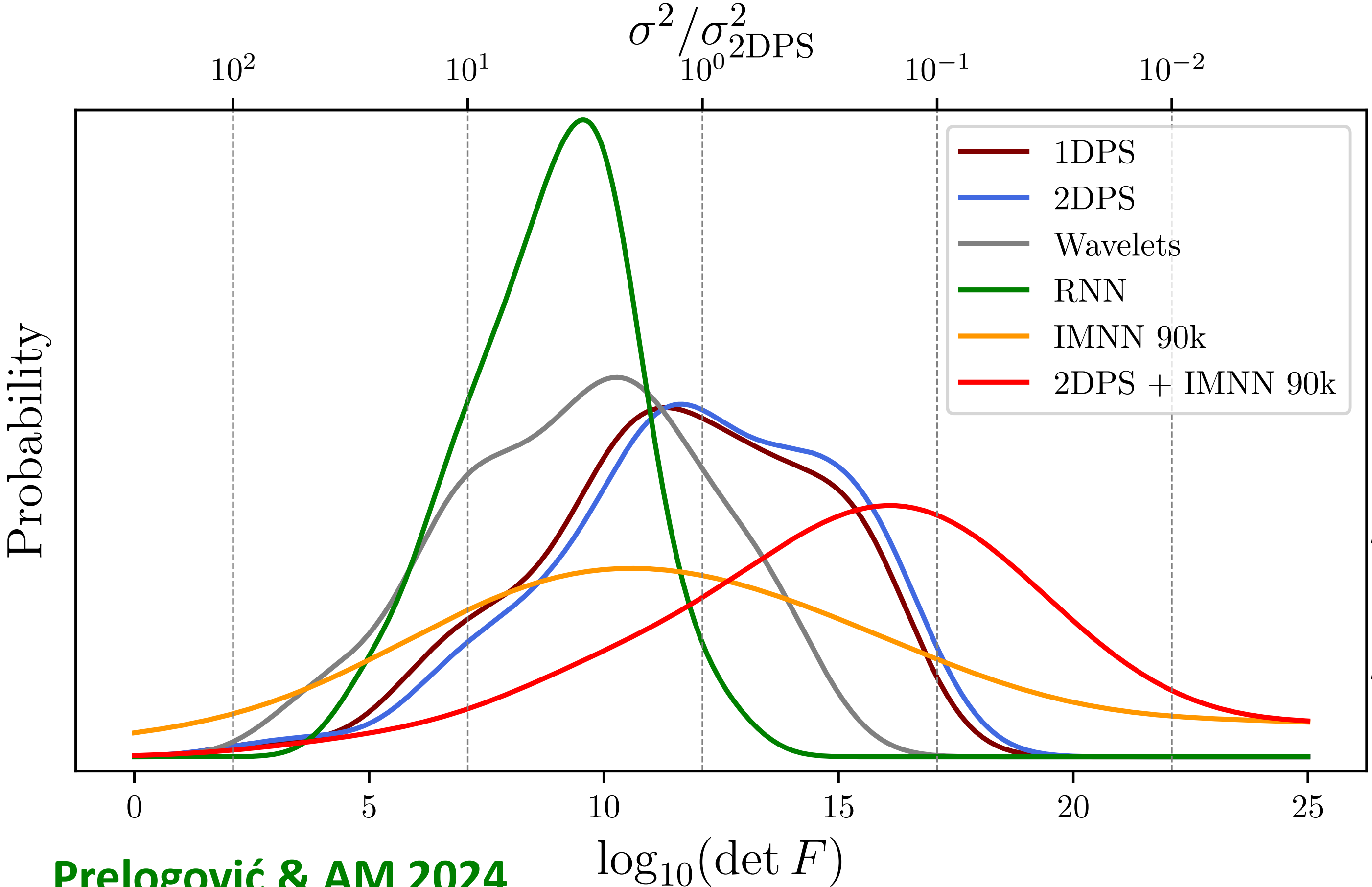


adapted from C. Chiang

Where we should be >2030-2040

High S/N map with the SKA

- *optimal compression of **non-Gaussian** signal (e.g. bispectrum, Minkowski functionals, wavelets, data-driven compression…)*



Compare constraining power of different summaries across prior volume

Prelogović & AM 2024

(see also, e.g. Watkinson+2017; Majumdan+2020; Chen+2019; Giri&Mellema2021; Kamran+2023…)

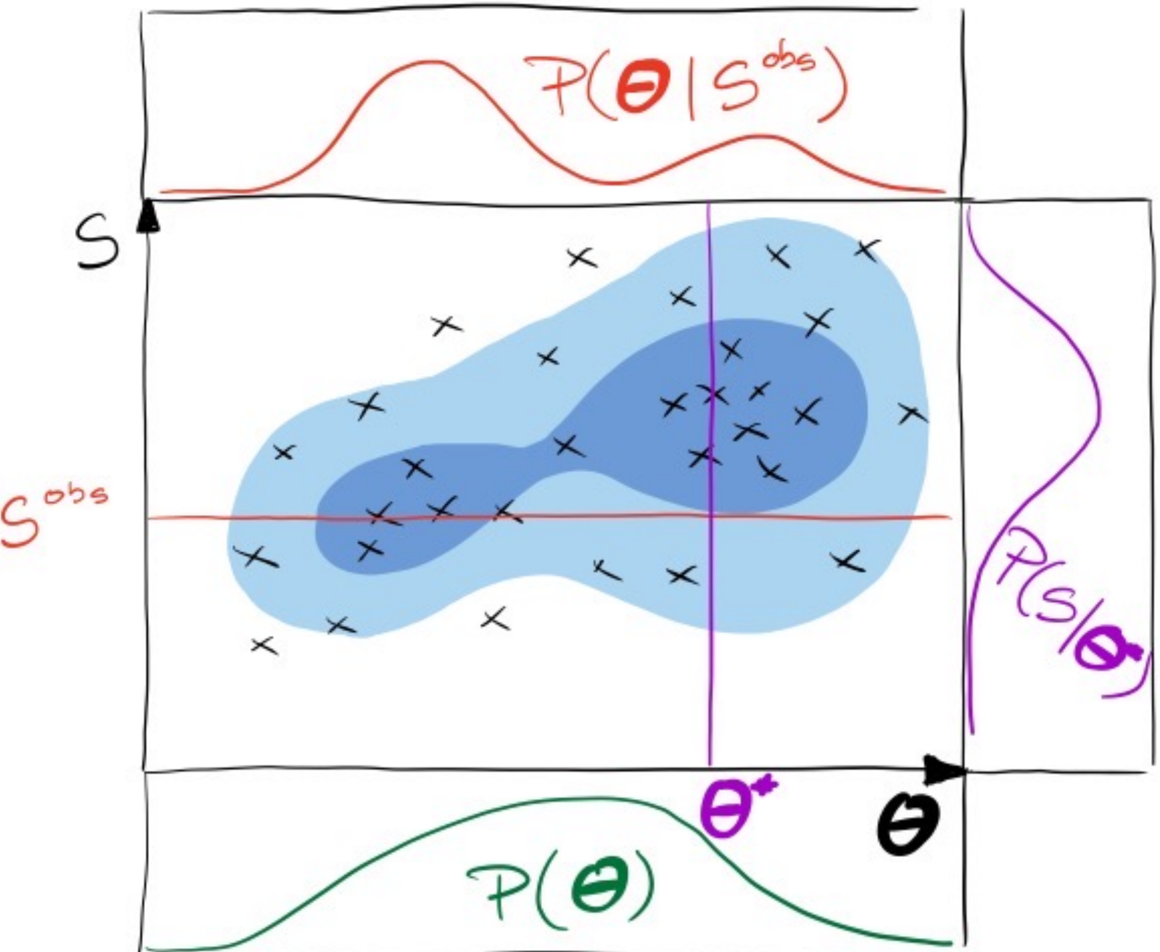
Where we should be >2030-2040

High S/N map with the SKA

- *optimal compression of* ***non-Gaussian*** *signal (e.g. bispectrum, Minkowski functionals, wavelets, data-driven compression…)*
- *do we actually know the likelihood analytically? —>* ***Simulation Based Inference (SBI)***

Simulation Based Inference (SBI)

Inference using SBI: *if including all main sources of stochasticity, each forward model is a sample from the joint distribution of model & data. The* ***likelihood*** *can just be fit with NDEs.*



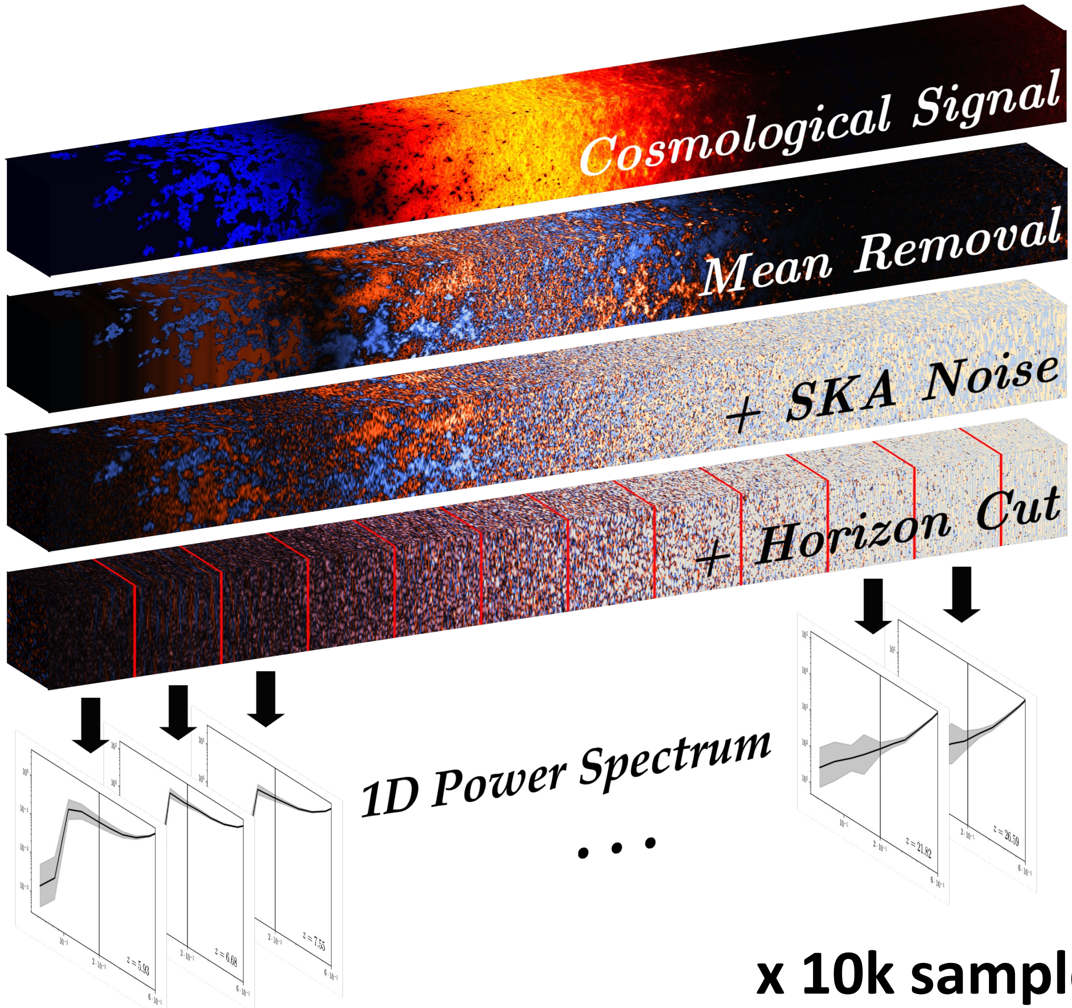
Credit: Tom Charnock

No need for an analytic likelihood!!!

difficult to write down for non-Gaussian and correlated observations

Simulation Based Inference (SBI)

Inference using SBI: *if including all main sources of stochasticity, each forward model is a sample from the joint distribution of model & data. The **likelihood** can just be fit with NDEs.*



No need for an analytic likelihood!!!

difficult to write down for non-Gaussian and correlated observations

Prelogović & AM (2023)

(see also Zhao+2022, Saxena+2023)

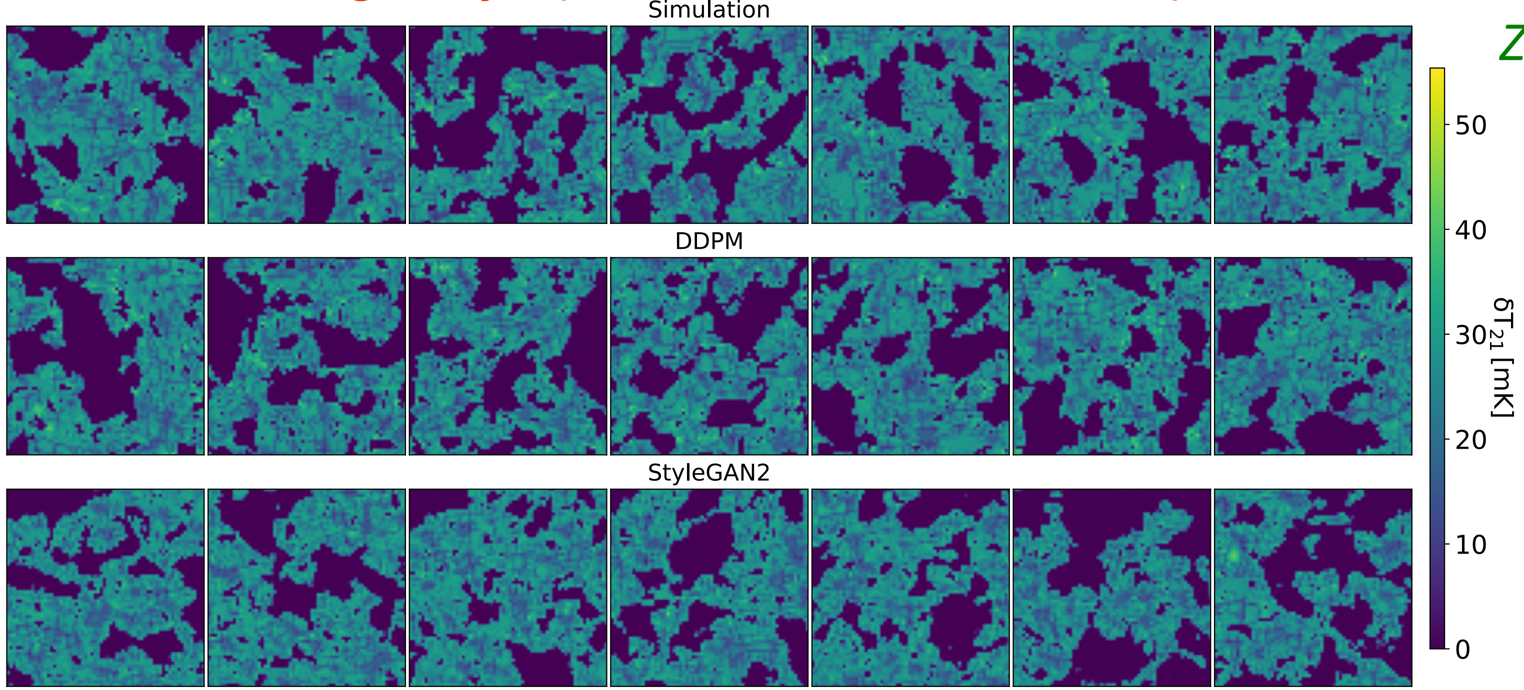
x 10k samples

Where we should be >2030-2040

High S/N map with the SKA

- *optimal compression of **non-Gaussian** signal (e.g. bispectrum, Minkowski functionals, wavelets, data-driven compression…)*
- *do we actually know the likelihood analytically? —> **Simulation Based Inference (SBI)***
- *emulating **maps** (do we trust emulators?)*

Zhao+ (2023)

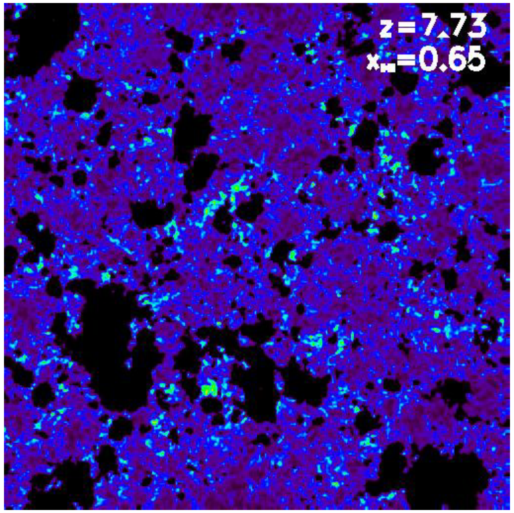


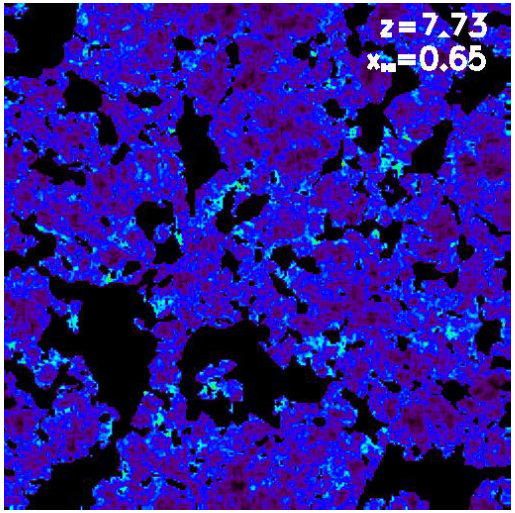
Where we should be >2030-2040

High S/N map with the SKA

- *optimal compression of **non-Gaussian** signal (e.g. bispectrum, Minkowski functionals, wavelets, data-driven compression…)*
- *do we actually know the likelihood analytically? —>* ***Simulation Based Inference (SBI)***
- *emulating **maps** (do we trust emulators?)*
- *how well do we trust our **simulators** (analytic, semi-numeric, moment-based RT, ray tracing, hydro…)**??***

AM+ (2011)





Conclusions

- The cosmic 21cm signal will allow us to learn the **average UV and Xray properties of the unseen first galaxies.**
- SKA will also open a new window on **physical cosmology**, e.g.
  - exotic heating processes, e.g. DM annihilations and decay
  - standard ruler at z=10-15 from velocity-induced feedback on galaxies
- **Upper limits** on the 21-cm power spectrum by SKA precursor, HERA, imply some **heating of the IGM by z>10.**
- If heating is provided by high mass X-ray binary stars, they are likely **more luminous** then local ones, likely due to their **low-metallicities**.
- Future detections will need **cross-correlations** with signals of known origin in order to be believed.
- **High S/N maps** of half of our observable Universe should be enabled by the **SKA** over the next couple of decades, ushering in a Big Data revolution