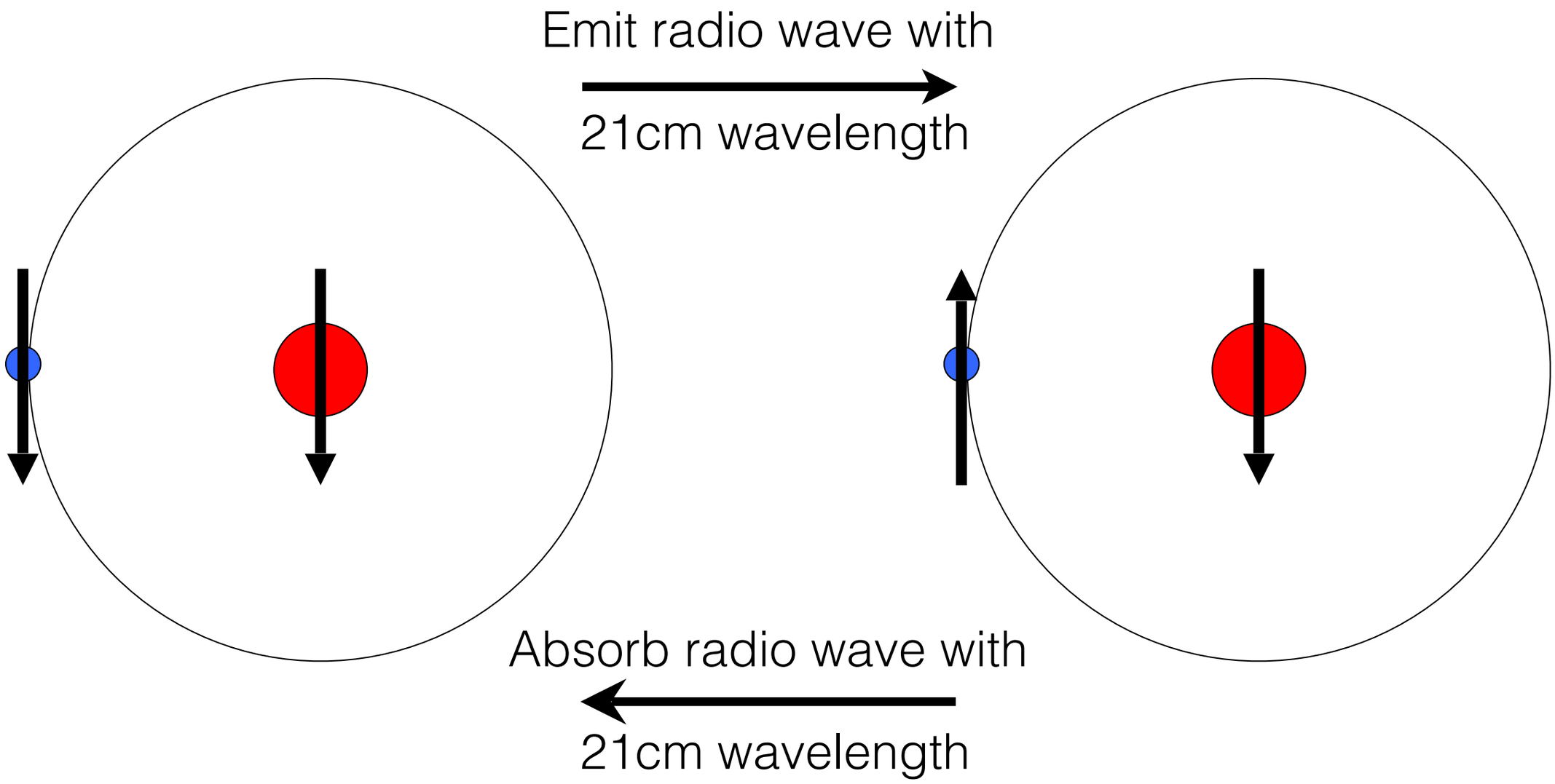
Machine Learning Reconstruction of Epoch of Reionization Bubbles

Adrian Liu
McGill University

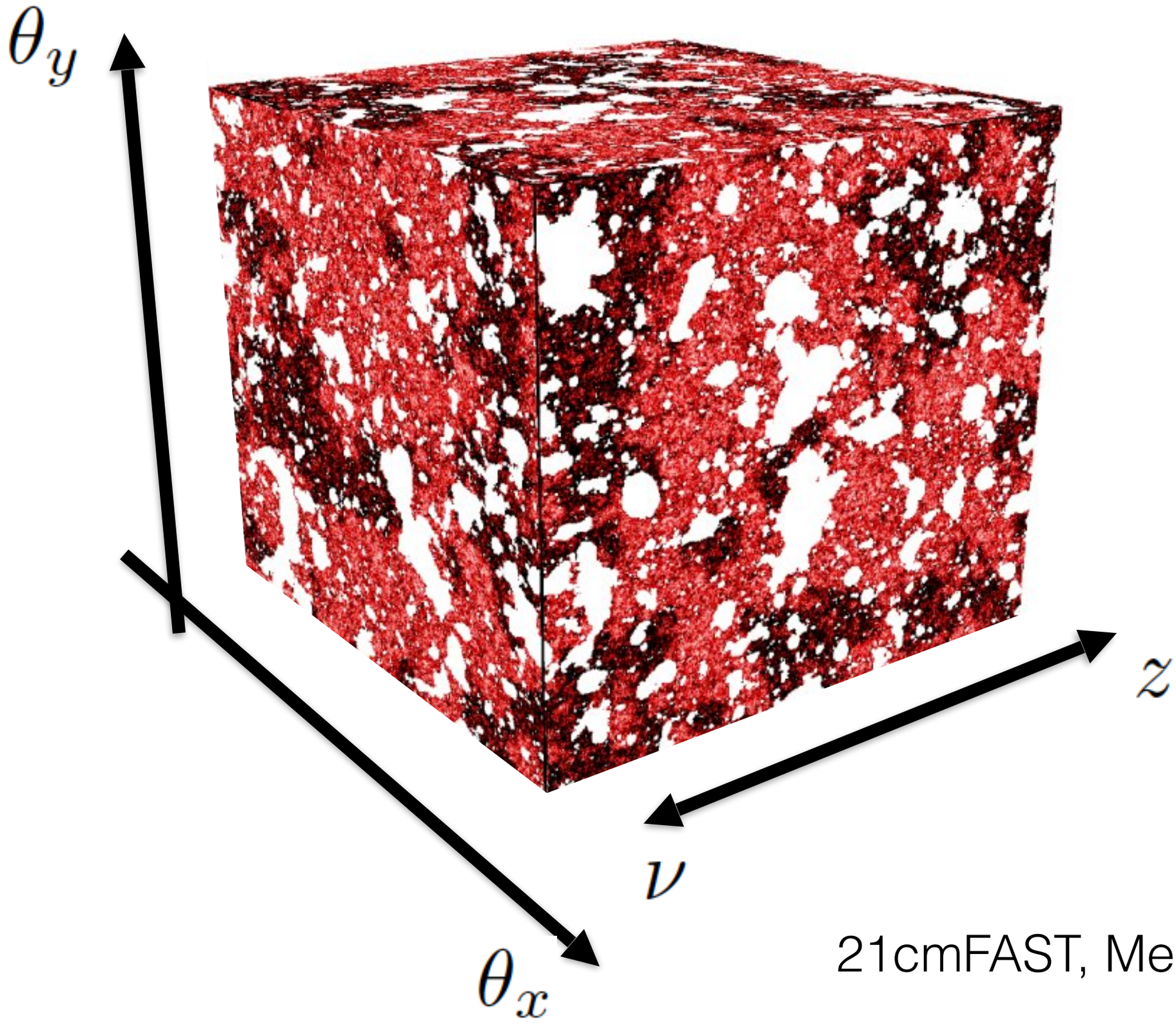
Institut Spatial de McGill
McGill Space Institute
CIFAR
McGill

Vision

Hydrogen is everywhere, and the 21cm line allows us to trace hydrogen

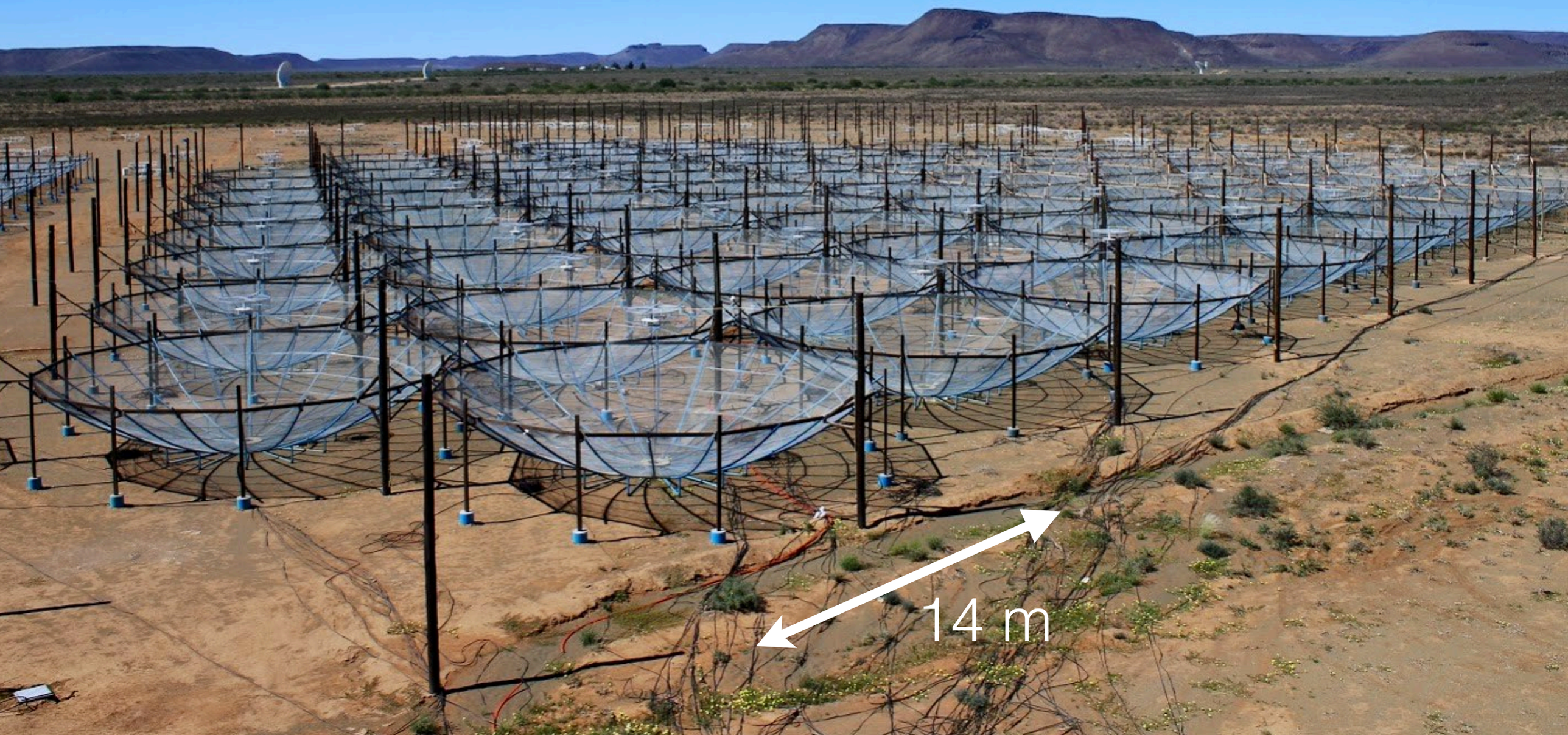


Alvarez et al. (2009)



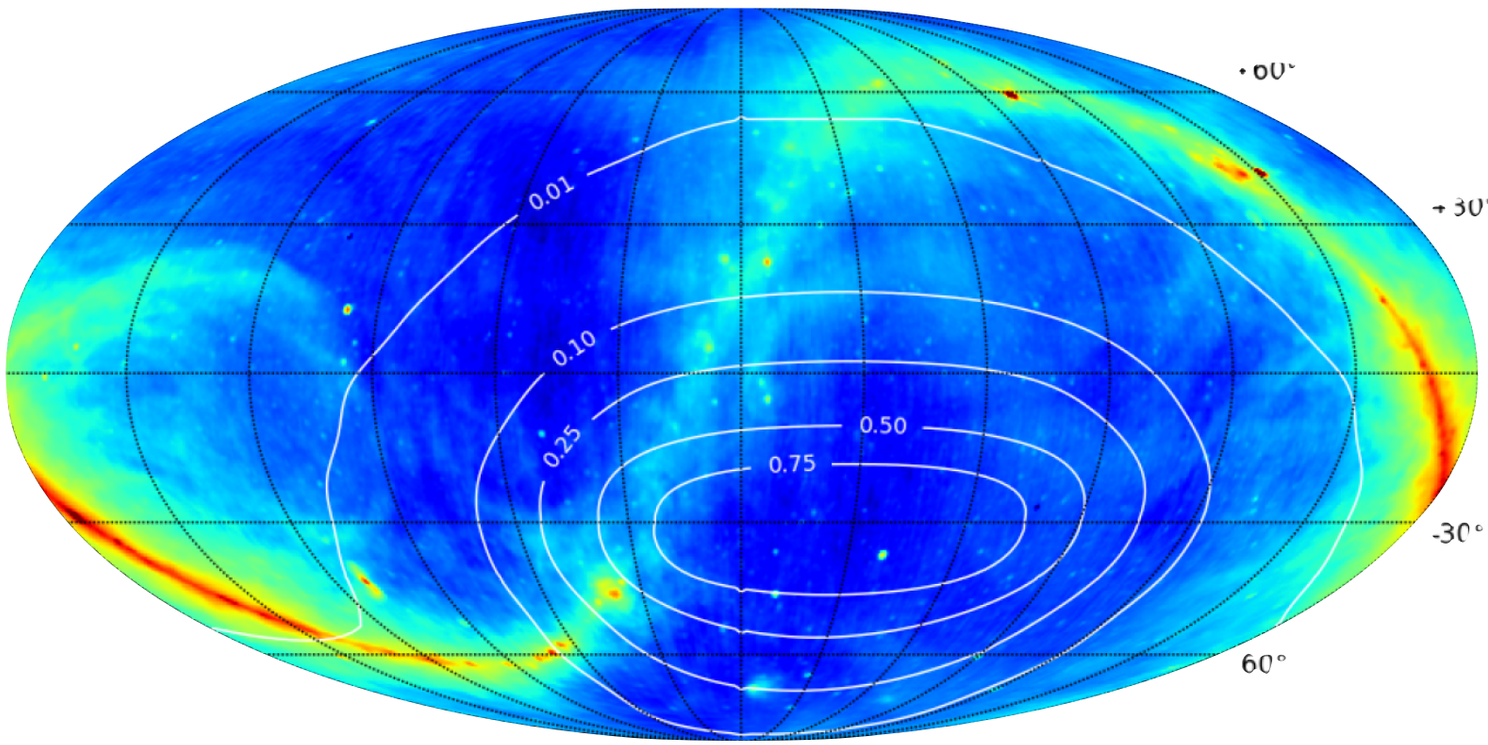
21cmFAST, Mesinger et al.

A new generation of radio observatories such as the **Hydrogen Epoch of Reionization Array (HERA)** are striving to detect these signals

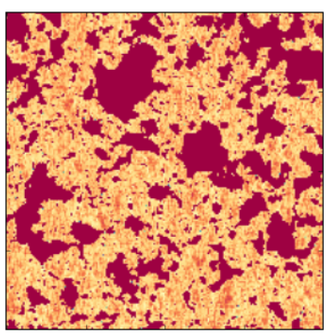


The Challenge…

…is that **foreground** contaminants are **orders of magnitude brighter**



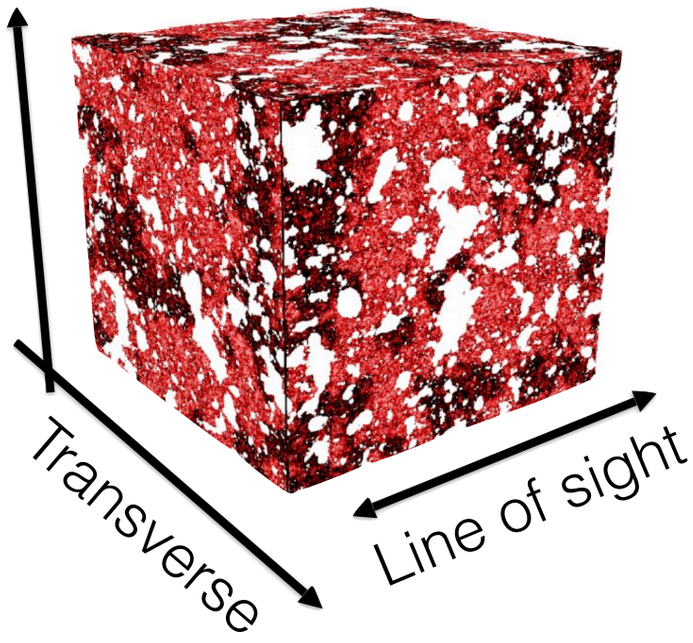
$\sim 10^5$ mK



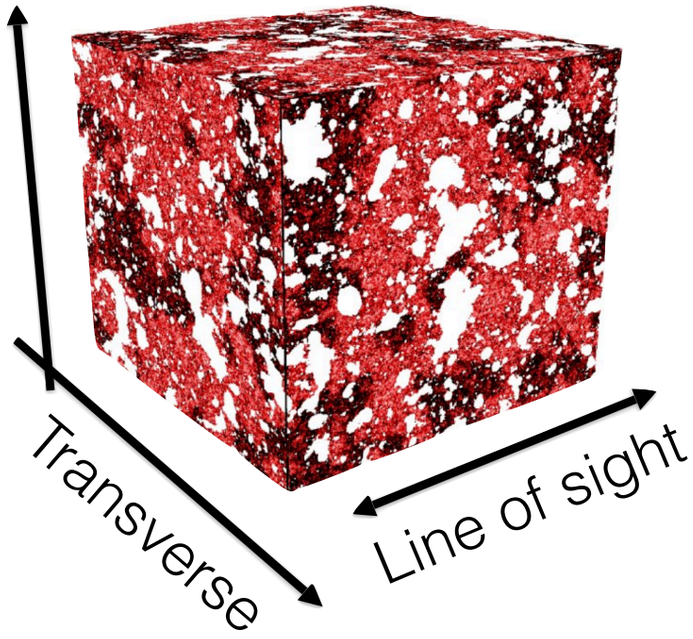
~ a few mK

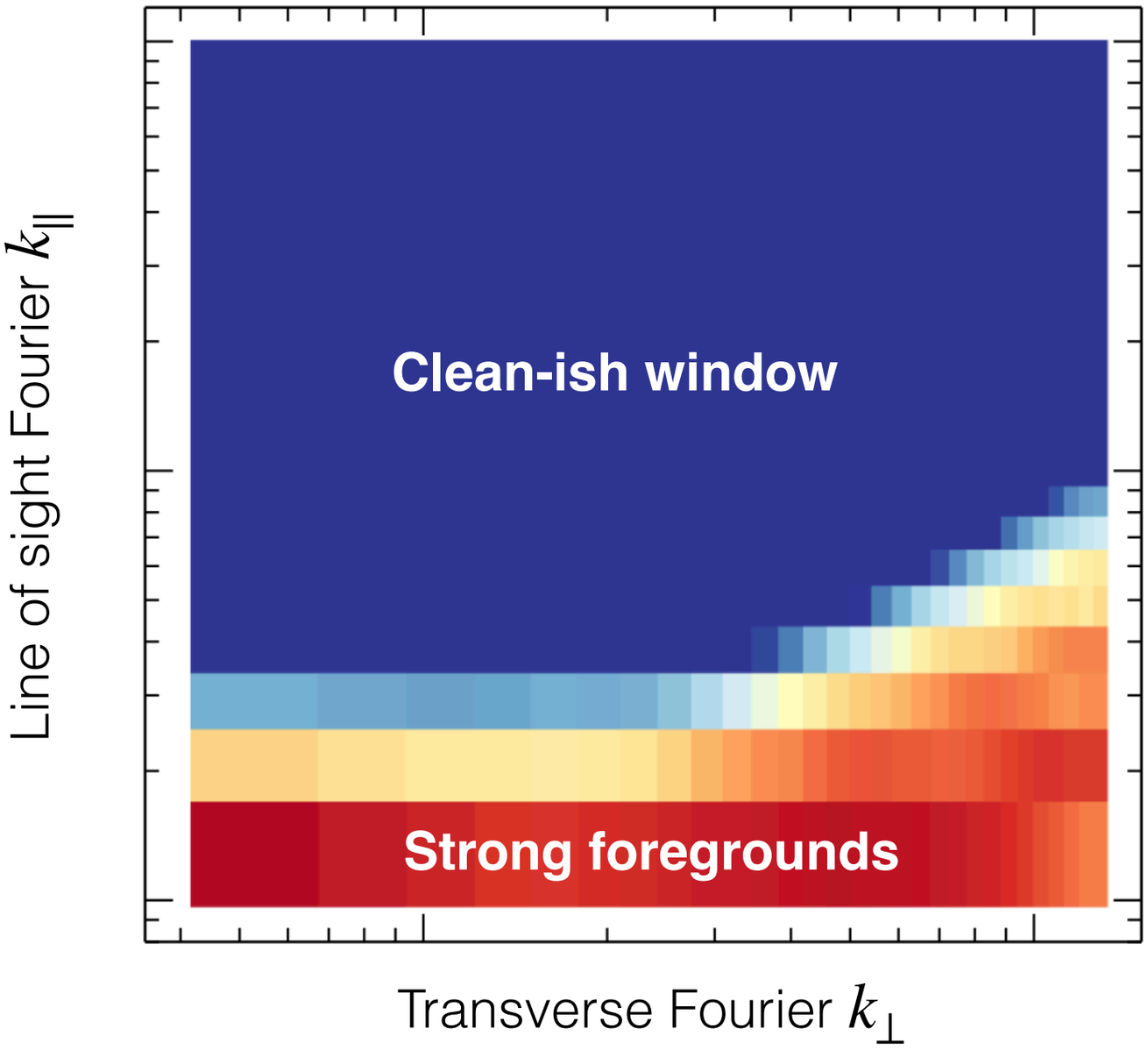
Filtering in Fourier space may be a solution…

Filtering in Fourier space may be a solution…

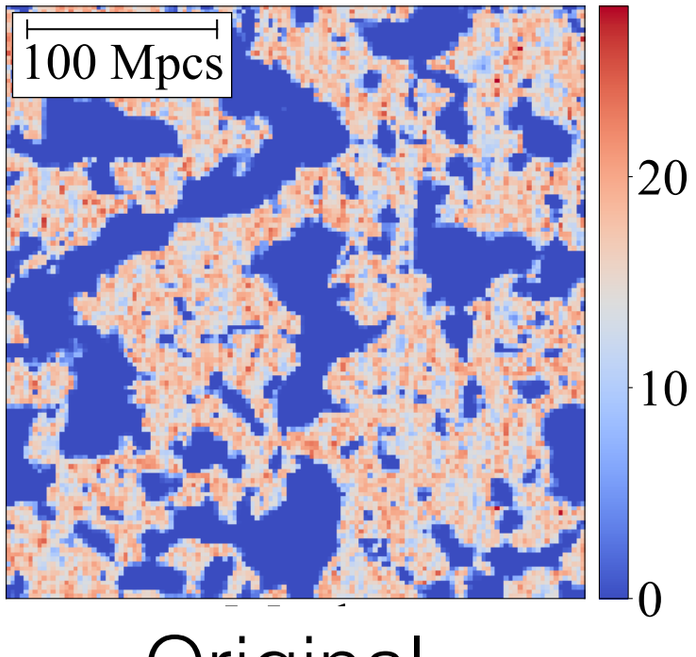


Filtering in Fourier space may be a solution…

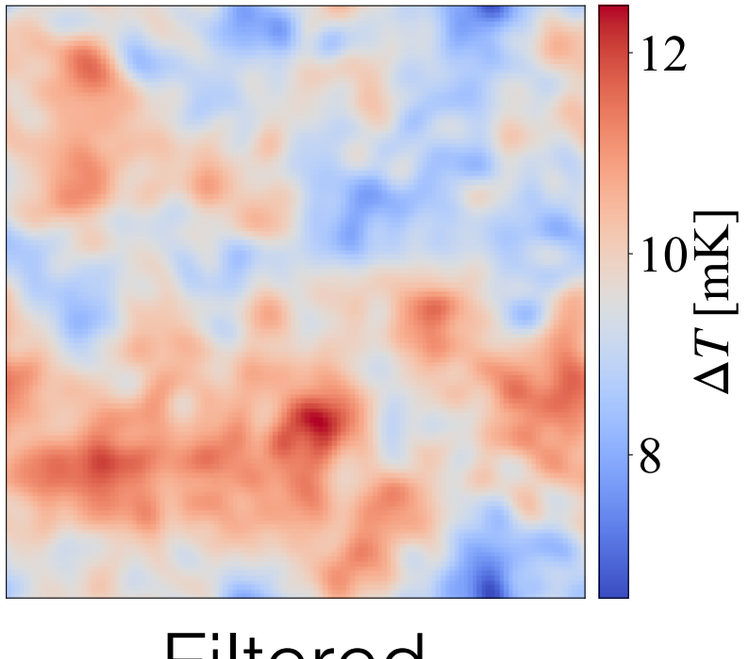




…but in getting rid of contaminants, it could destroy lots of information….

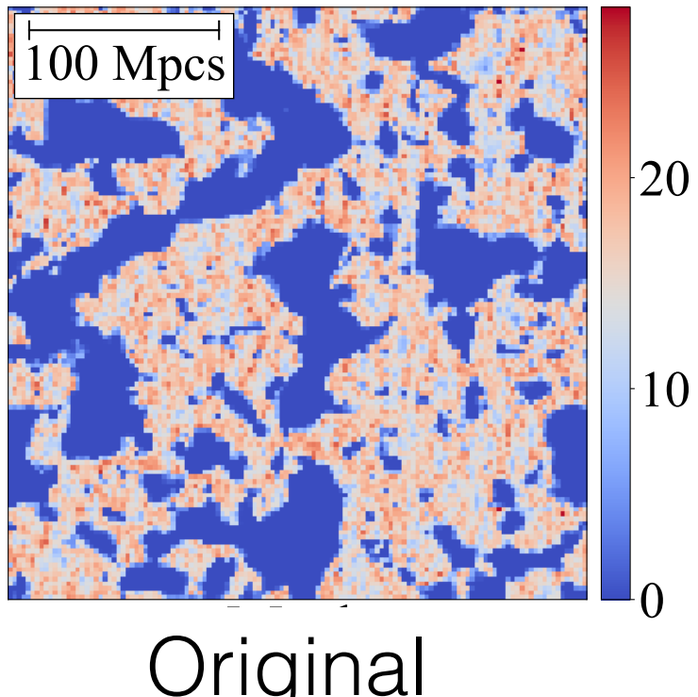


Original

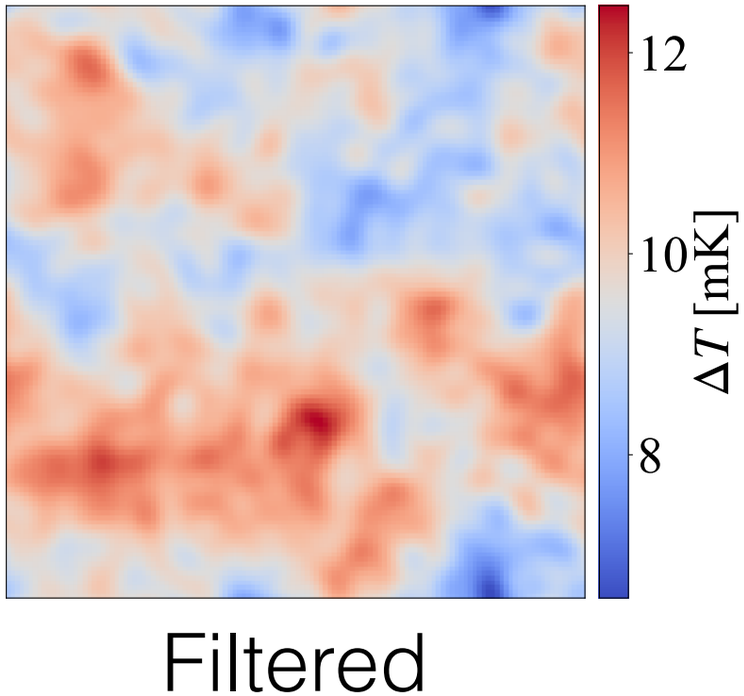


Filtered

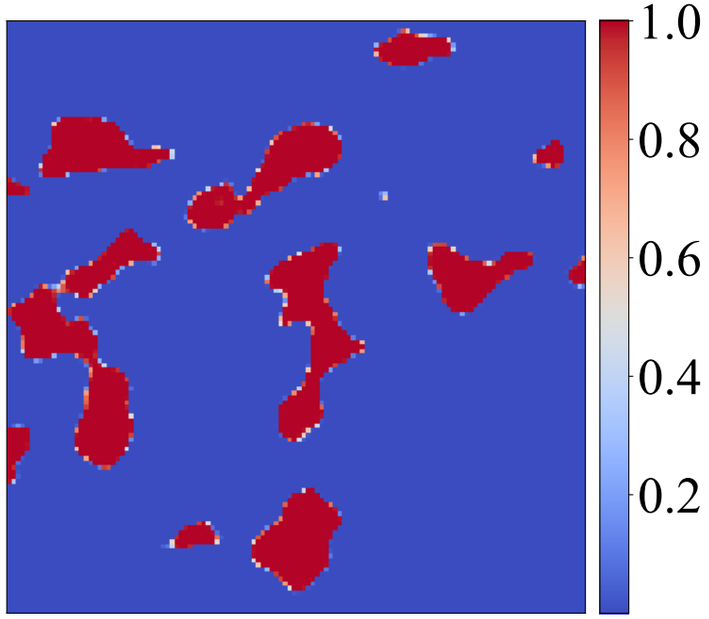
…but in getting rid of contaminants, it could destroy lots of information….



Original



Filtered

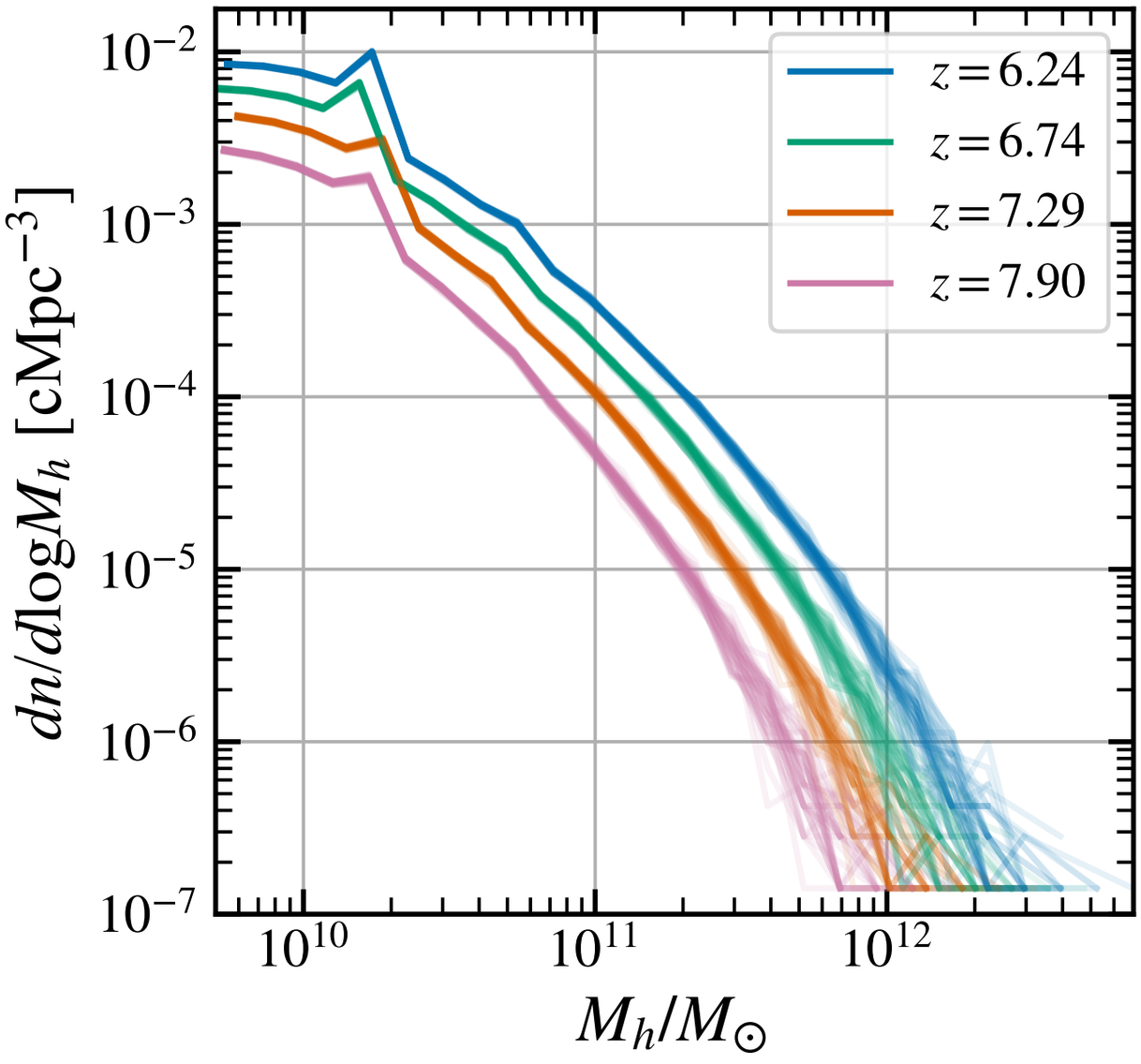


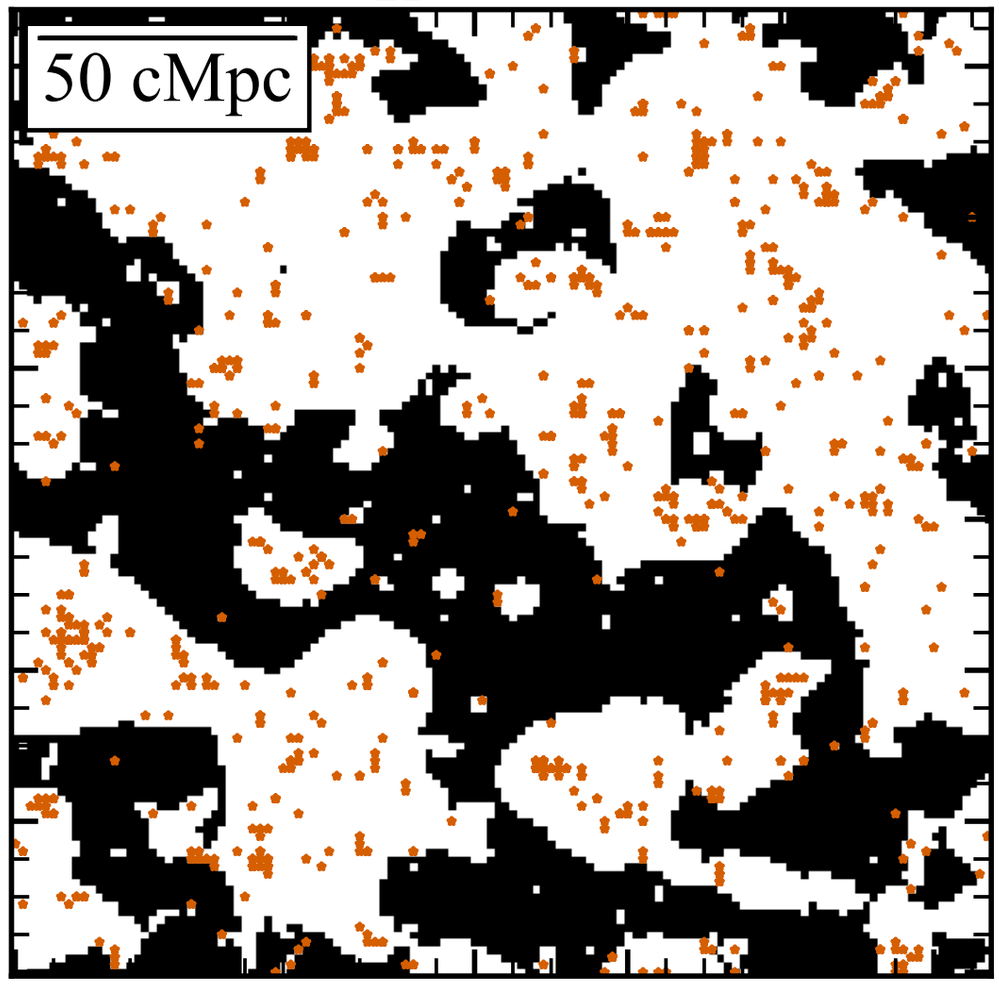
Predicted ionization

….but perhaps machine learning and students can save the day! (Gagnon-Hartman, Cui, **AL**, Ravanbakhsh., 2021, MNRAS **504**, 4716)

Christine Cui

Sam Gagnon-Hartman

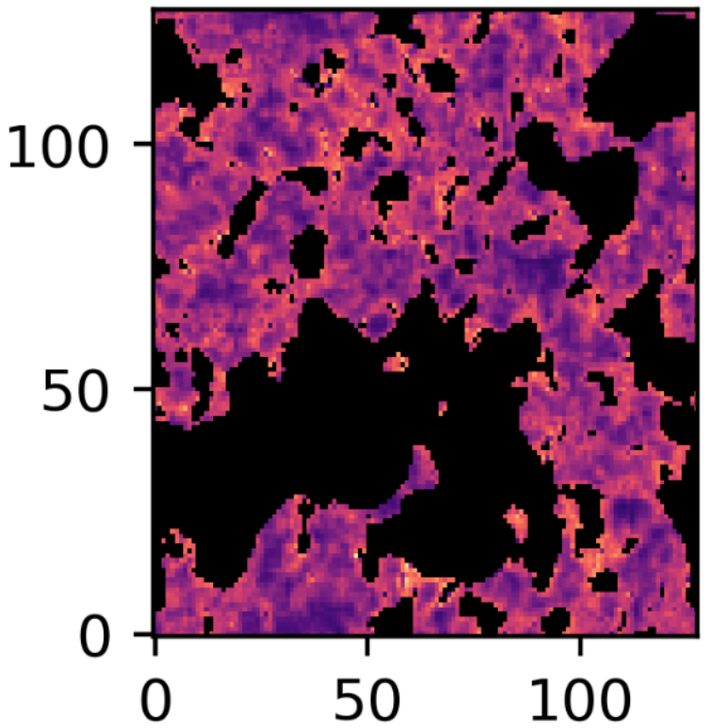




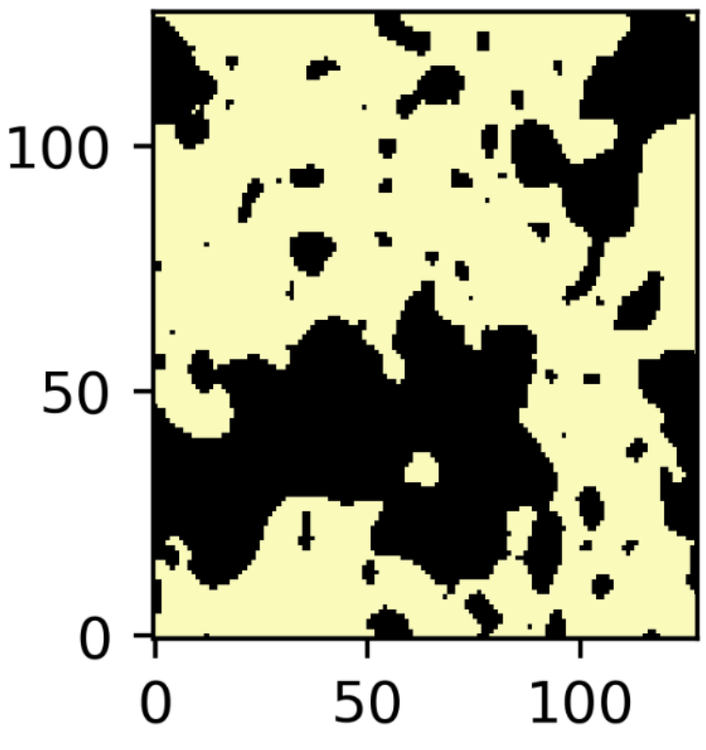
These maps can be used as a guide map for galaxies!

(Kennedy et al., arXiv:2308.09740)

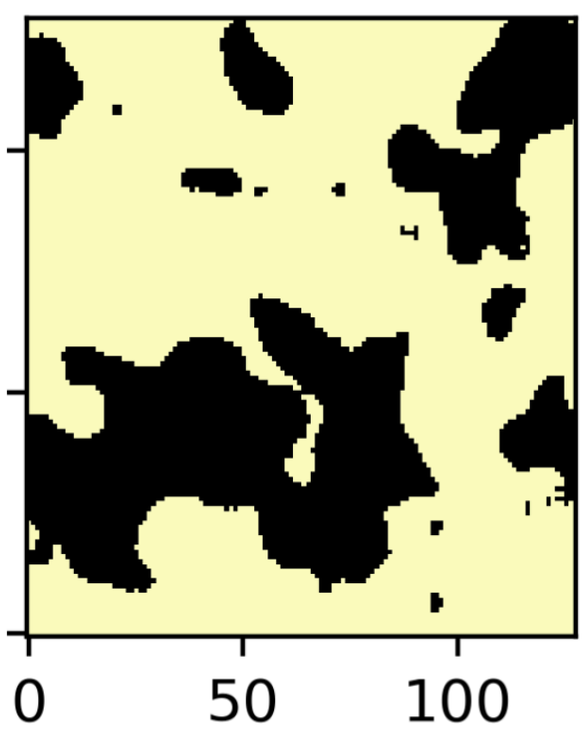
Jacob Kennedy



Or if we have already found the galaxies, we can improve the reconstruction of bubbles!



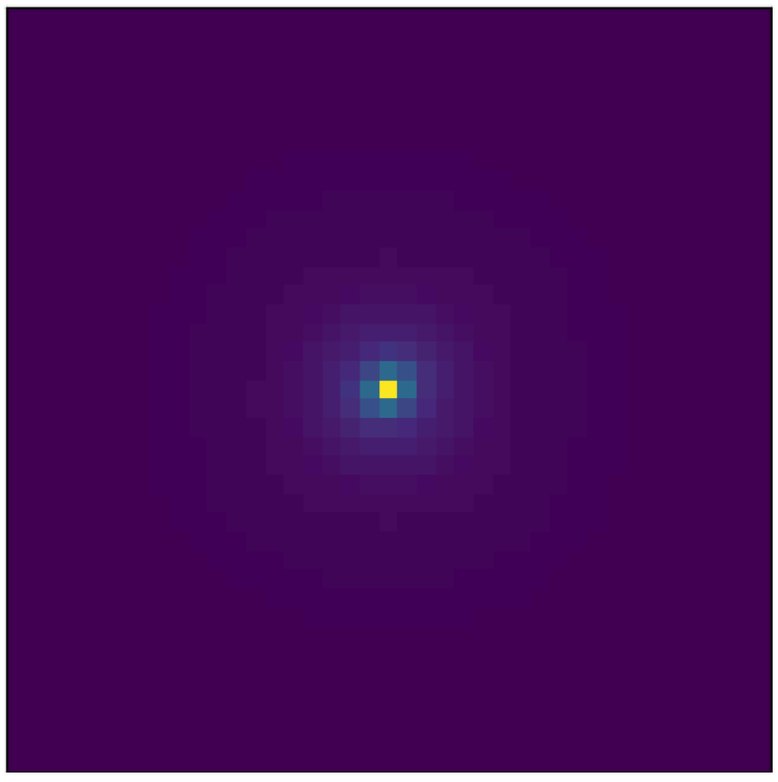
With the help of galaxies



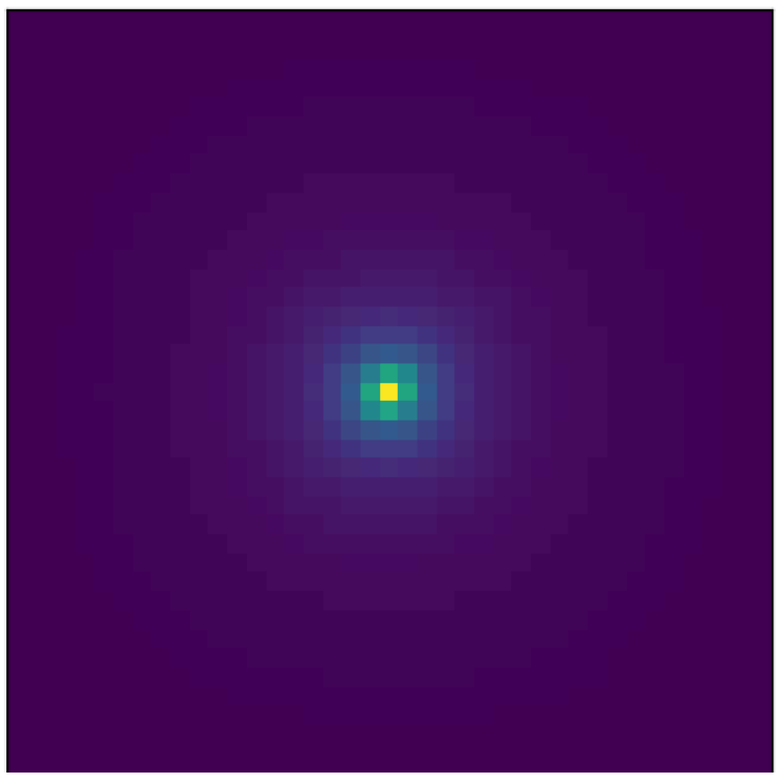
Without using galaxies

Franco del Balso

The **statistical isotropy** of our Universe means that **stacked bubbles** could be used to test the **geometry of spacetime**

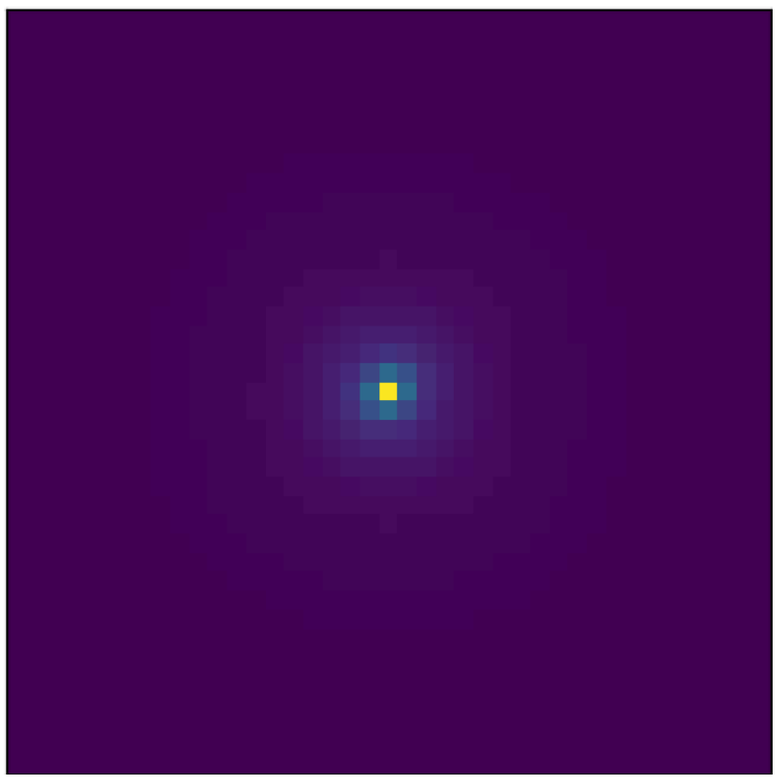


Ground truth bubble stack

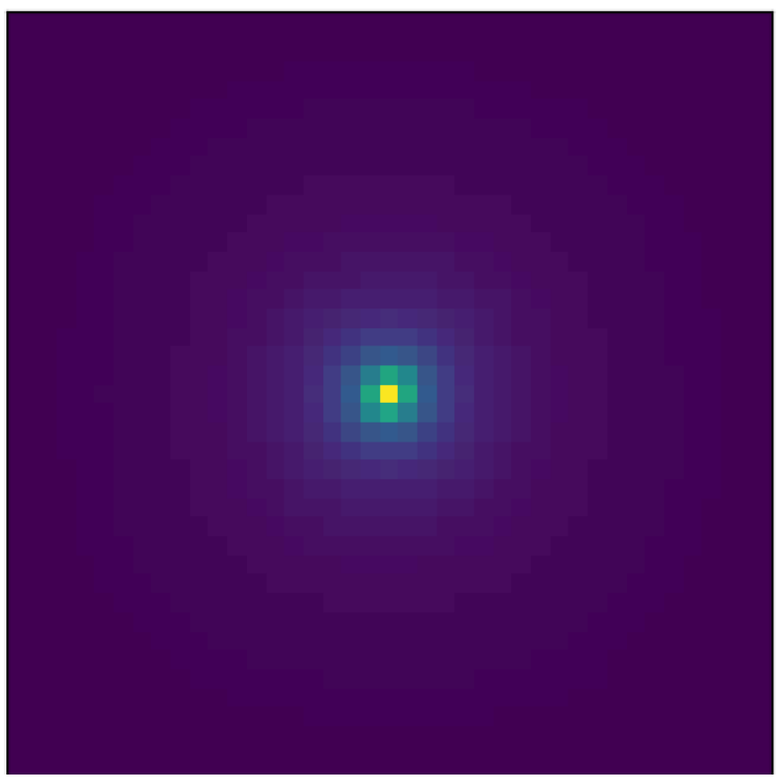


Recovered bubble stack

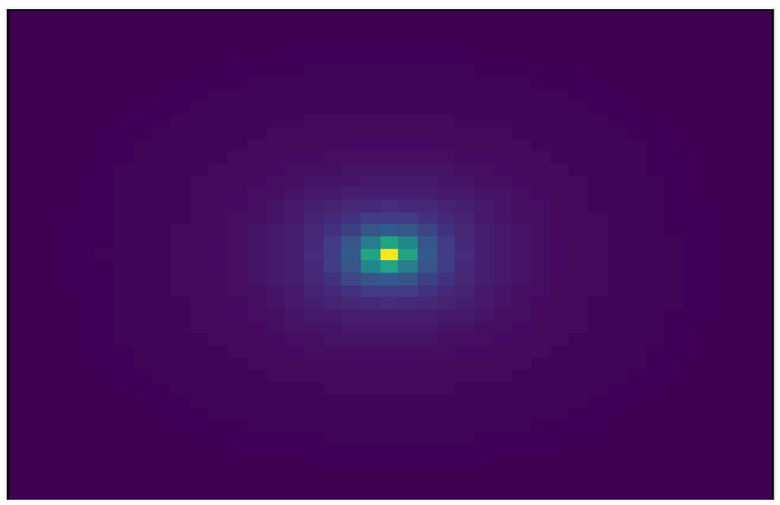
The **statistical isotropy** of our Universe means that **stacked bubbles** could be used to test the **geometry of spacetime**



Ground truth bubble stack

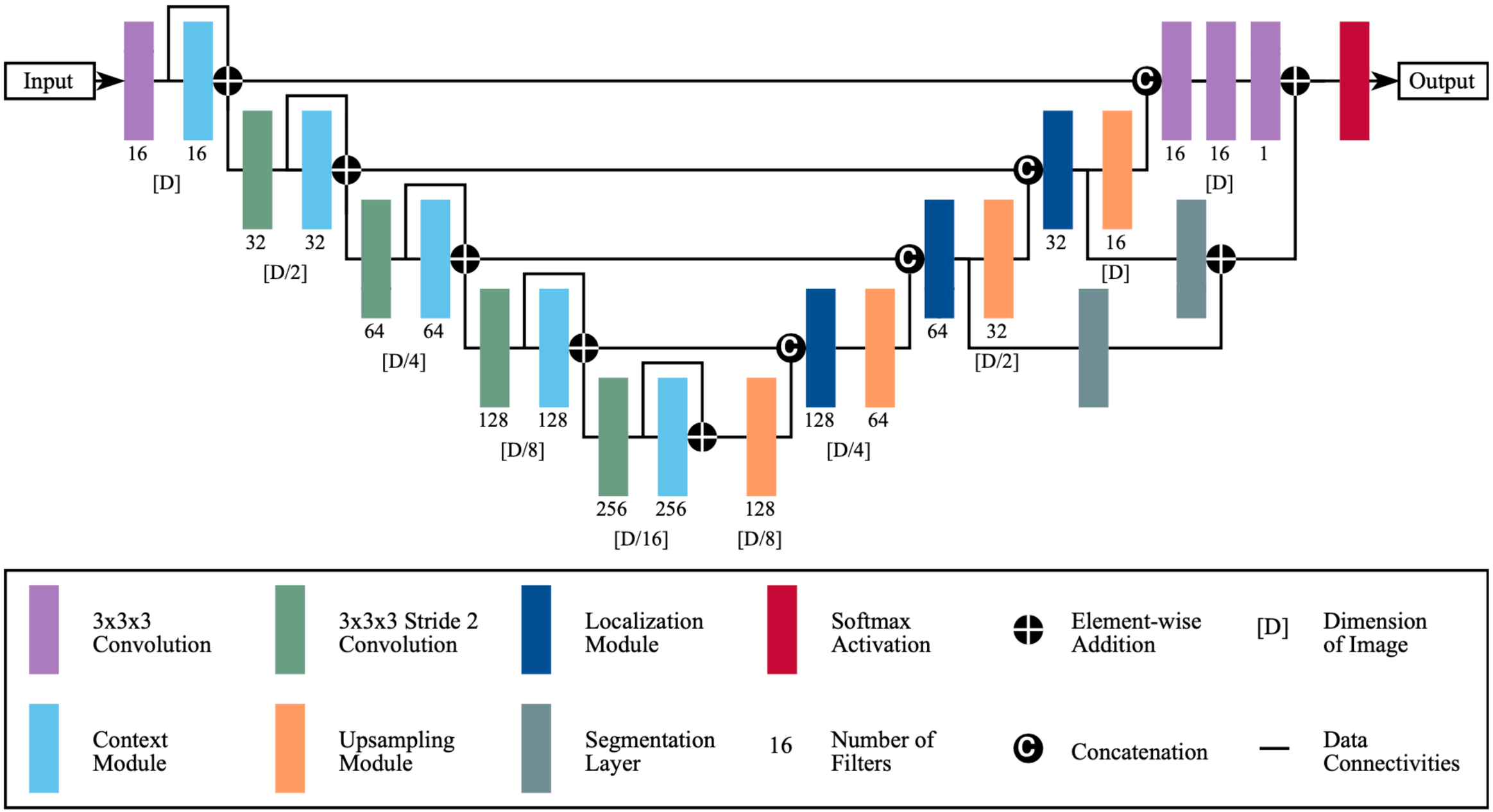


Recovered bubble stack

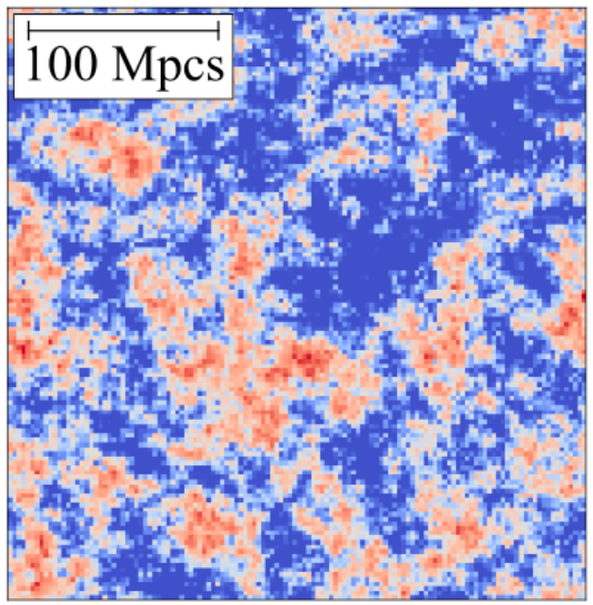


Incorrect assumptions about Universe's geometry

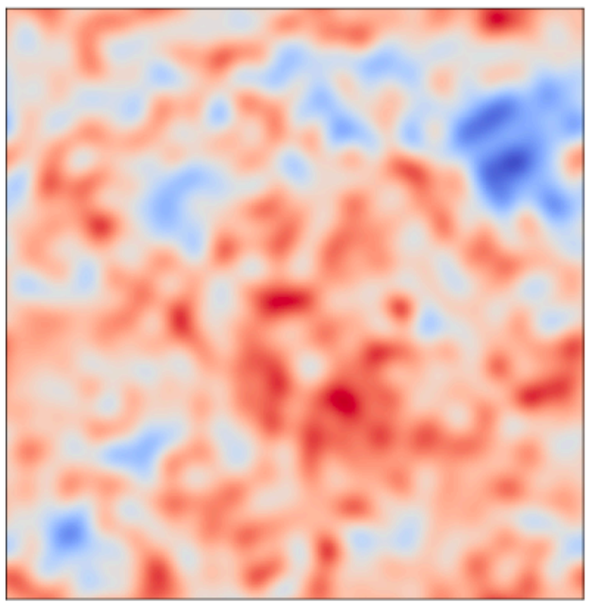
How can this possibly even work?!



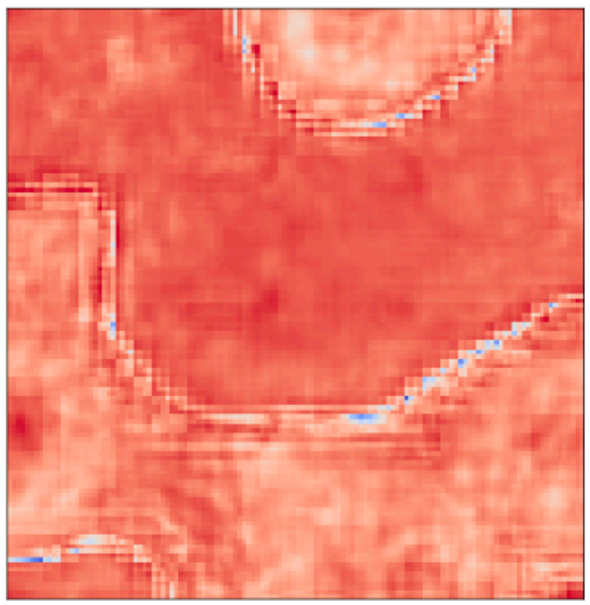
Non-Gaussianity to the rescue: **Gaussian** maps have **uncorrelated Fourier modes** due to stationarity, and thus **no recovery is possible**



Gaussianized ground truth



Filtered input



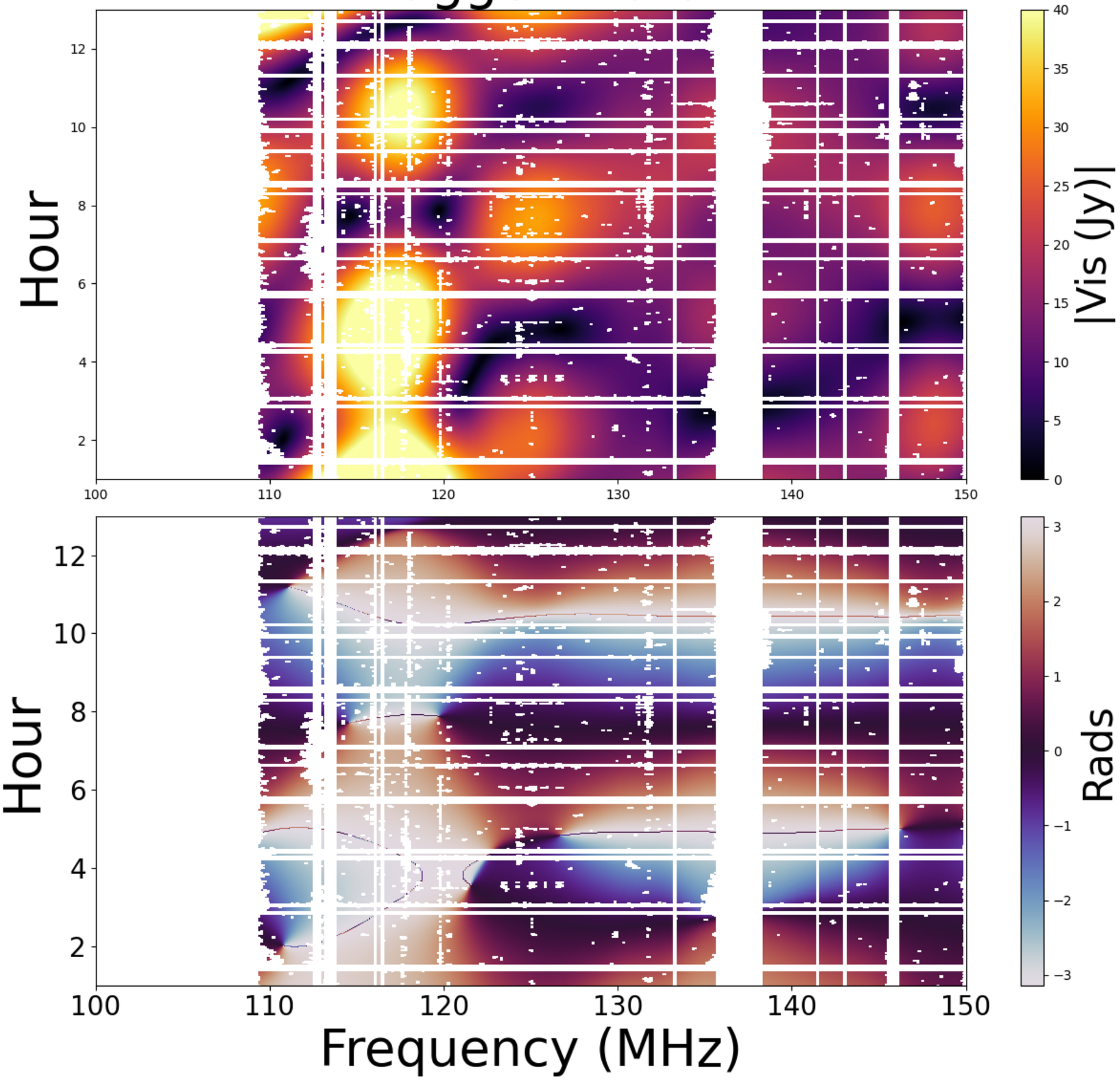
(Failed) prediction

Other astrostats problems in 21cm cosmology that I'd love to talk about

- How do we deal with the avalanche of data **without using data compression techniques that assume Gaussianity?**
- How do we best deal with **missing data?**
- How can we optimally relate "volts" in our instruments to sky signals when **our flux scales or sky models are uncertain?**
- What are the **optimal basis vectors for combining related data** from other probes?
- Can **field-level inferences (bypassing summary statistics)** be a practical path forward in this field?
- Can **symmetries in our data** be effectively utilized?
- What can **simulation-based inference** do for us?

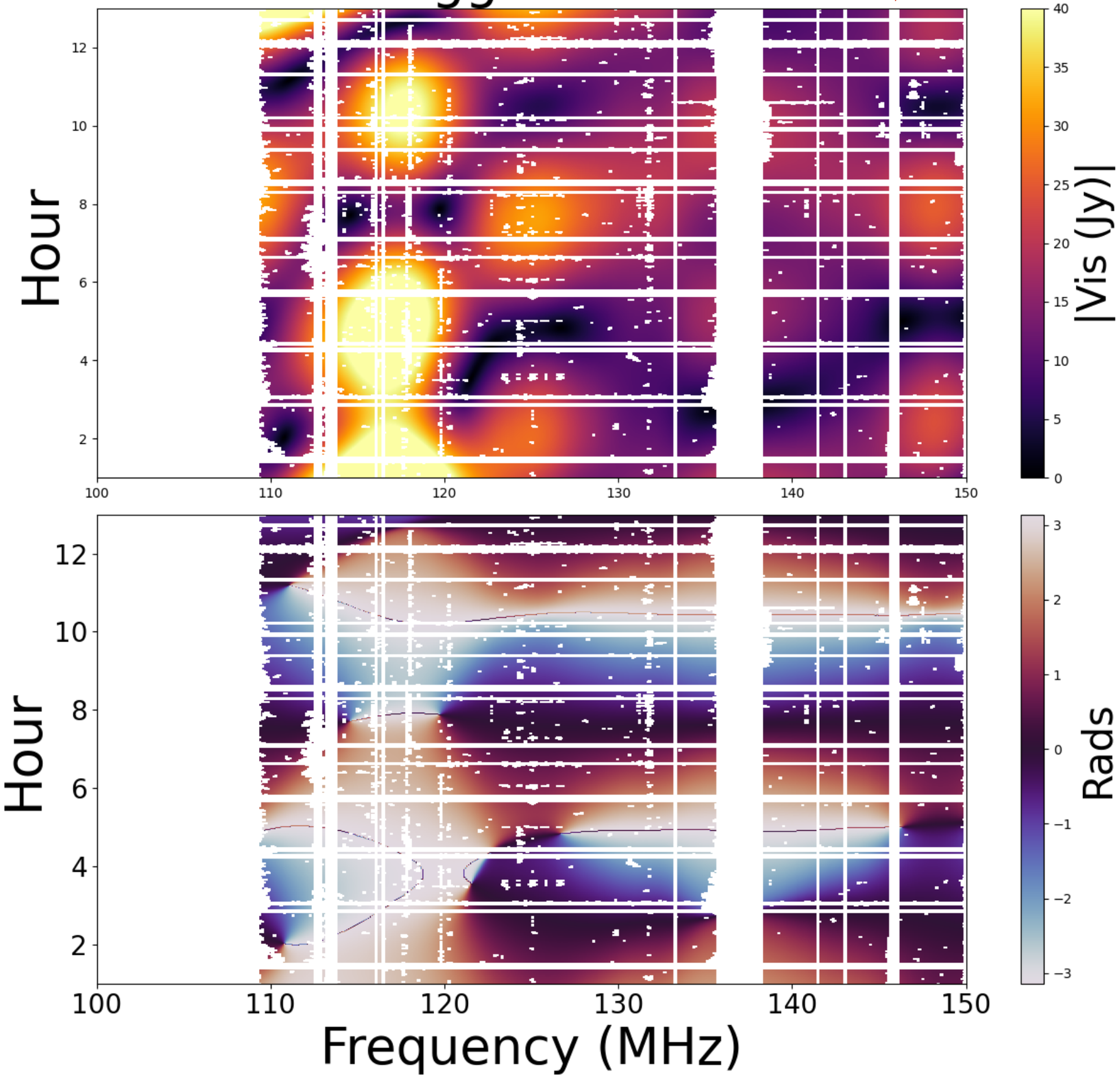
ML

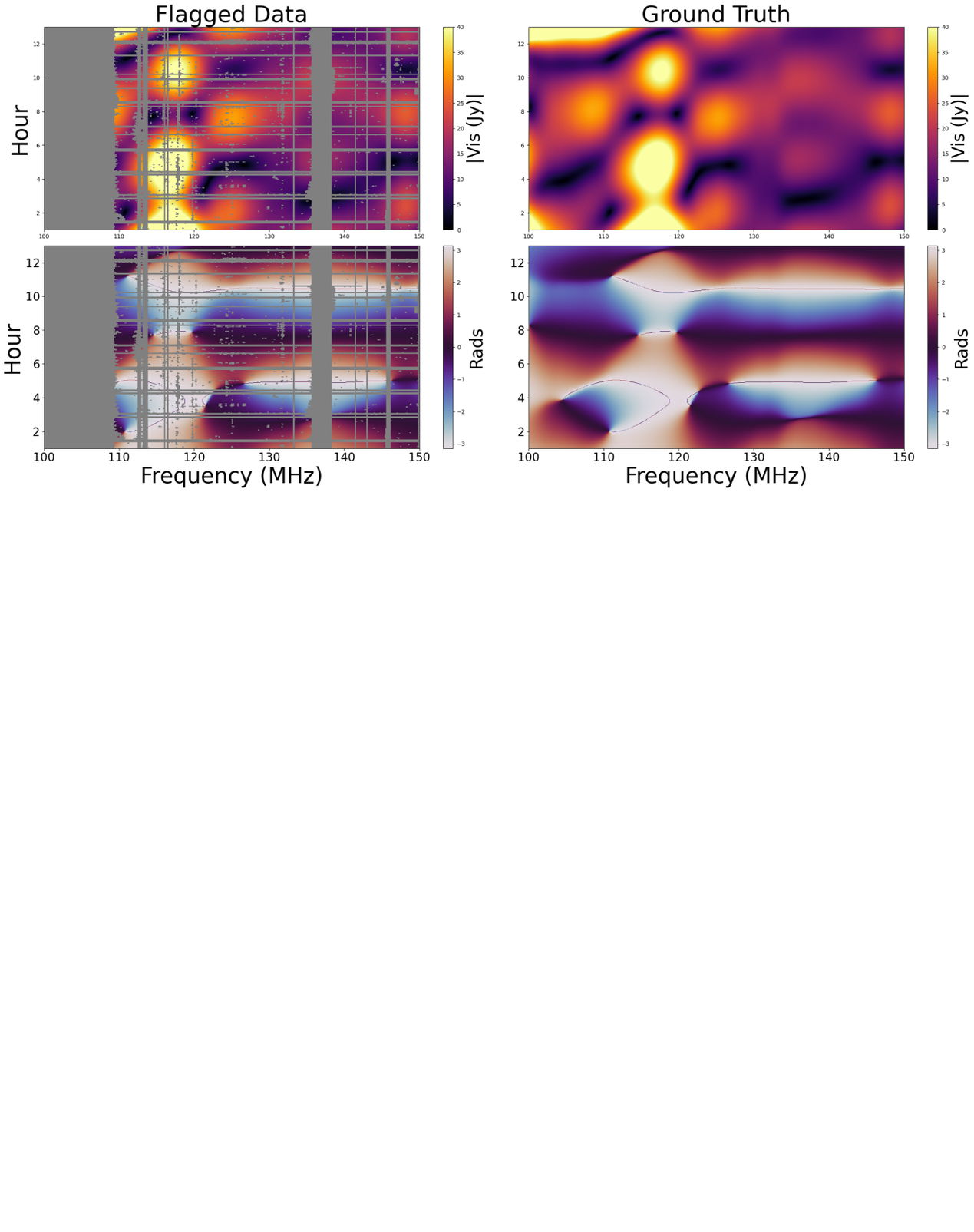
Flagged Data



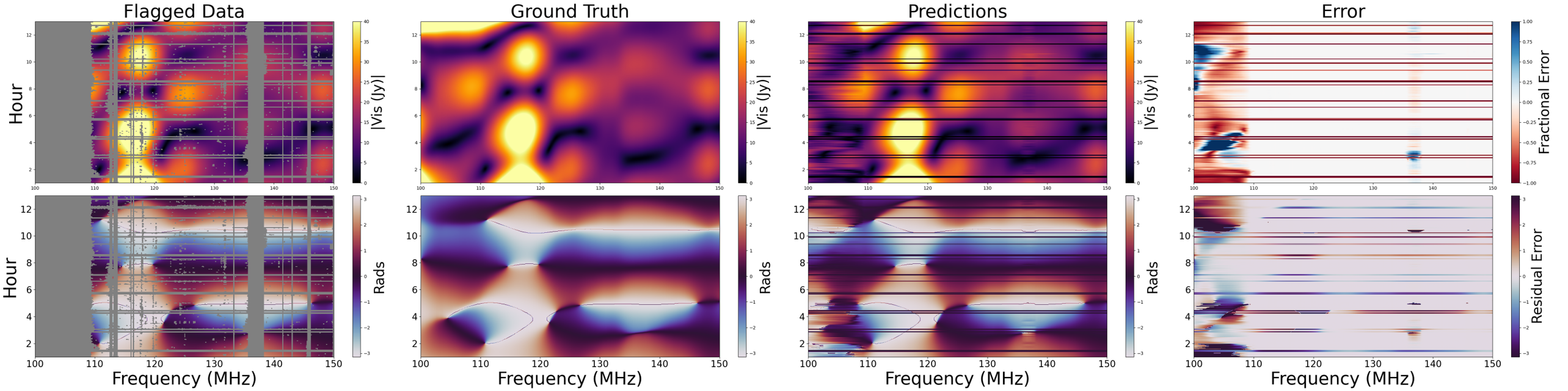
Flagged Data

Radio Frequency Interference (RFI)

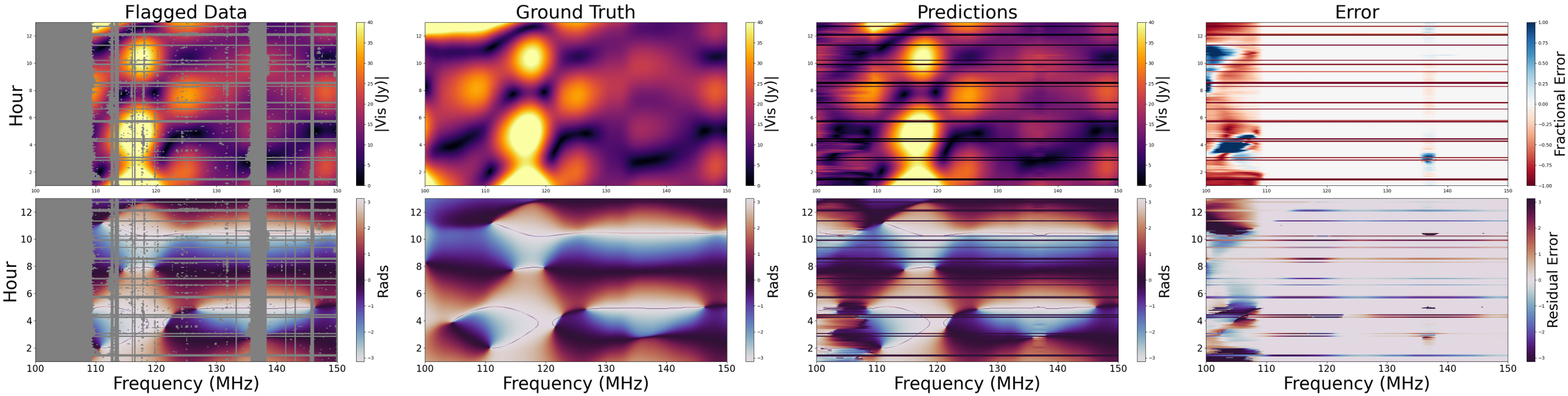



Flagged Data
Ground Truth
Hour
Hour
|Vis (Jy)|
|Vis (Jy)|
Rads
Rads
Frequency (MHz)
Frequency (MHz)

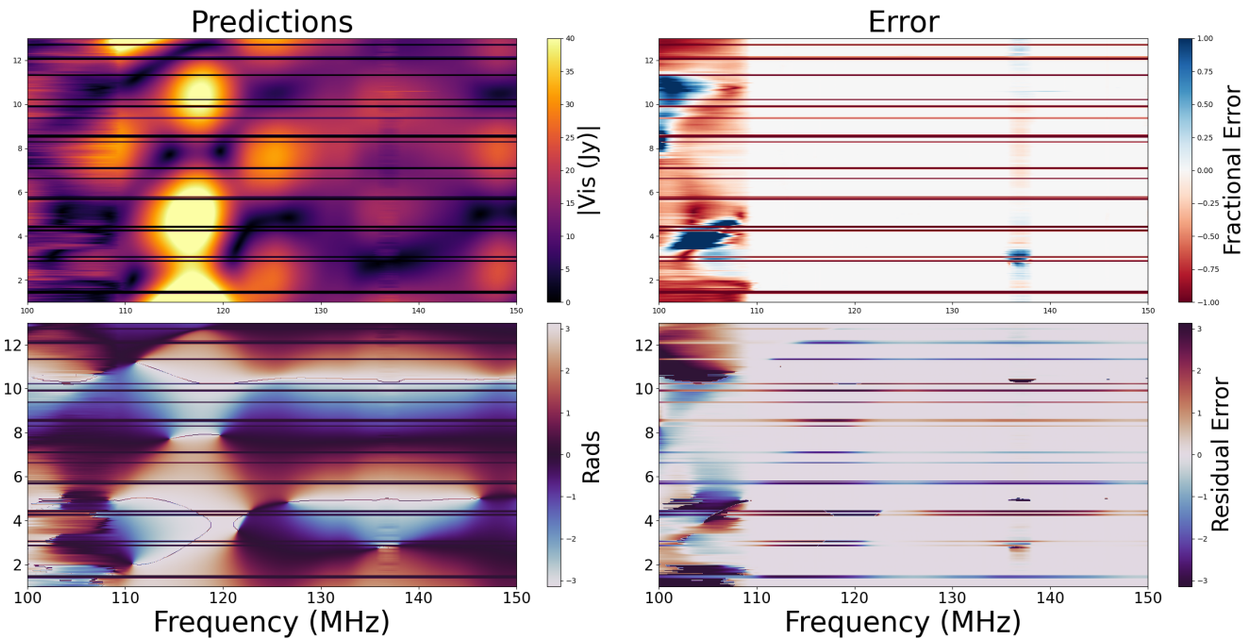
Constrained Gaussian Realizations



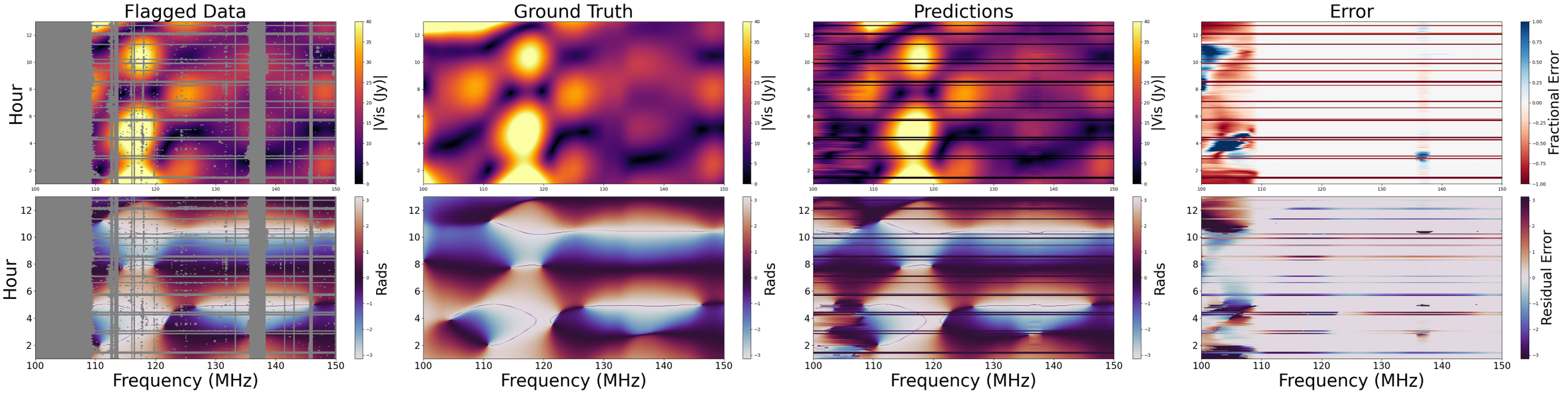
Constrained Gaussian Realizations



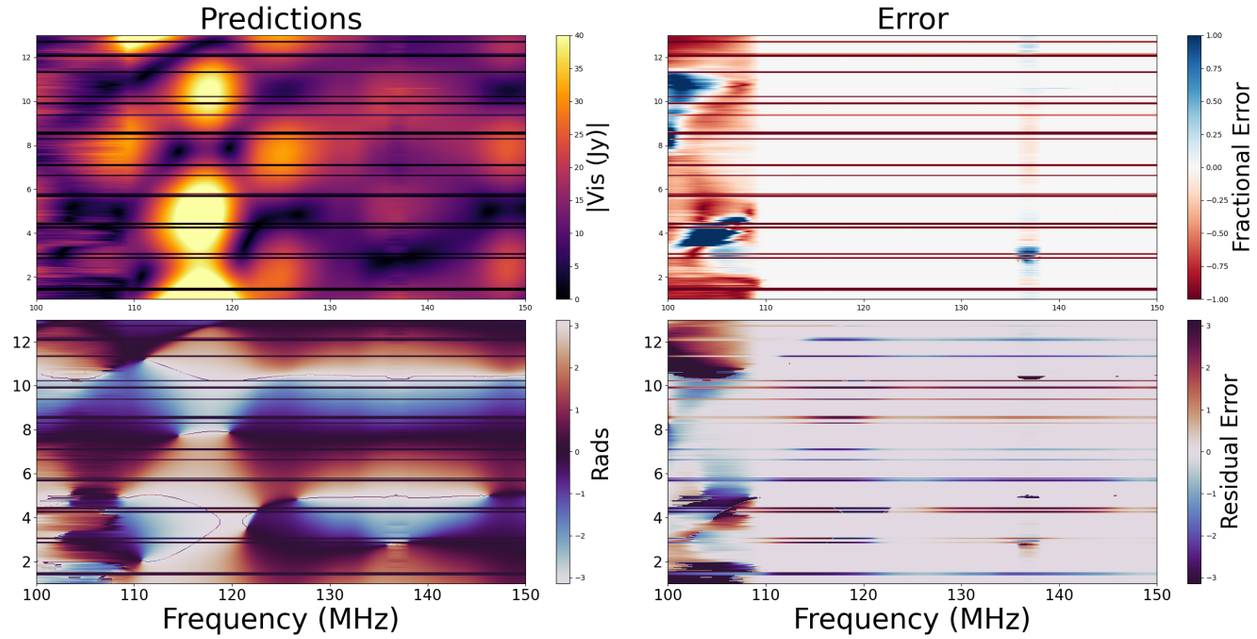
“CLEAN”



Constrained Gaussian Realizations



“CLEAN”

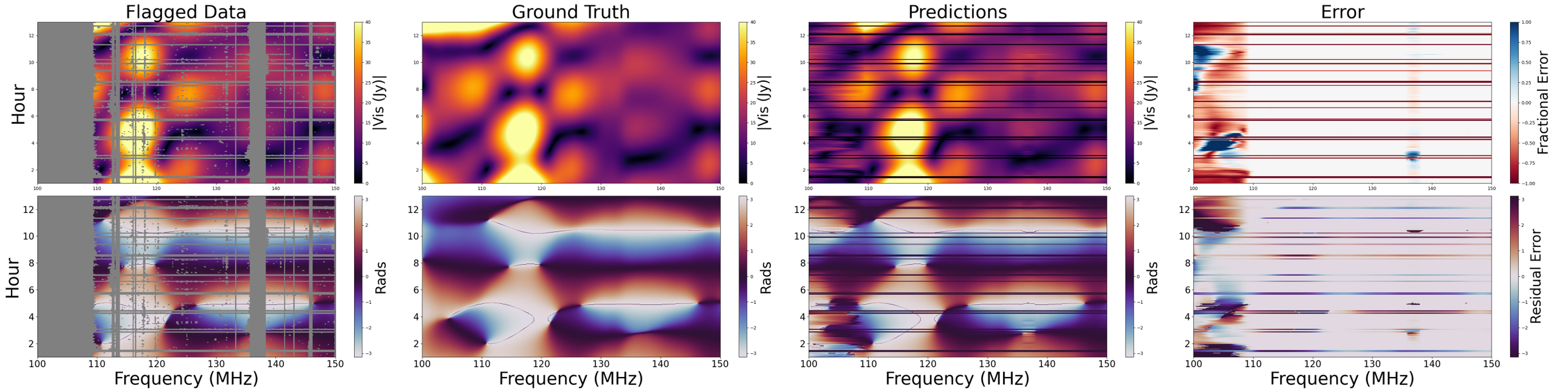


Machine Learning?

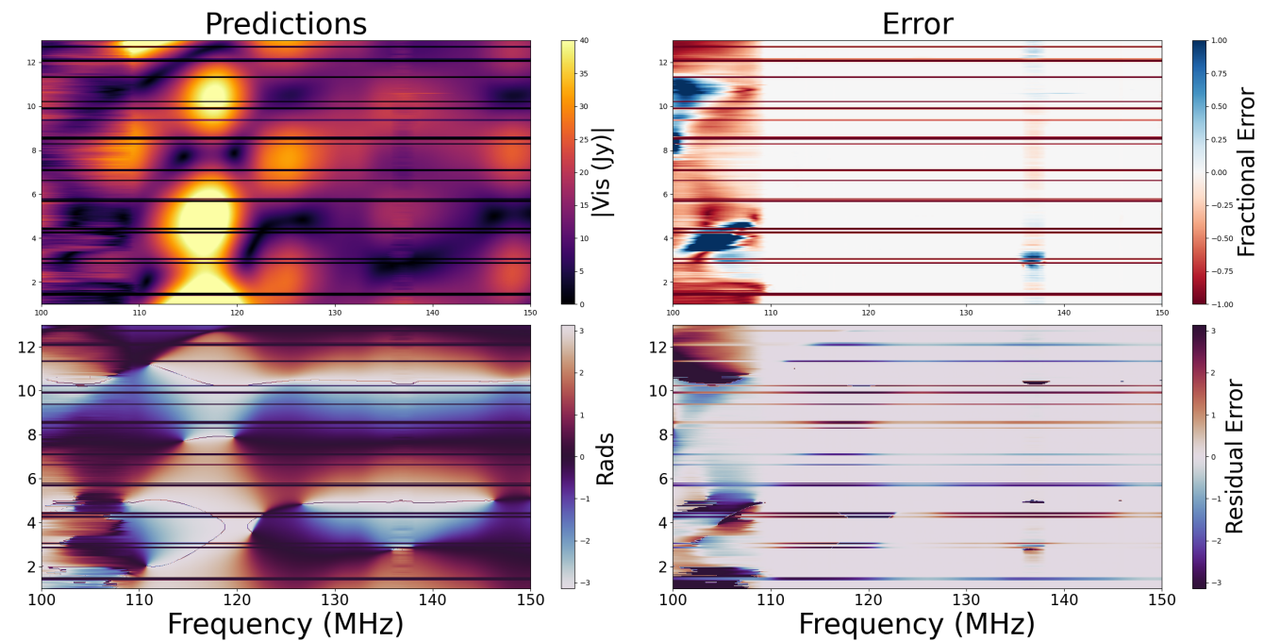
Michael Pagano PhD student

Our data is like a picture with gaps that can be inpainted using machine learning techniques

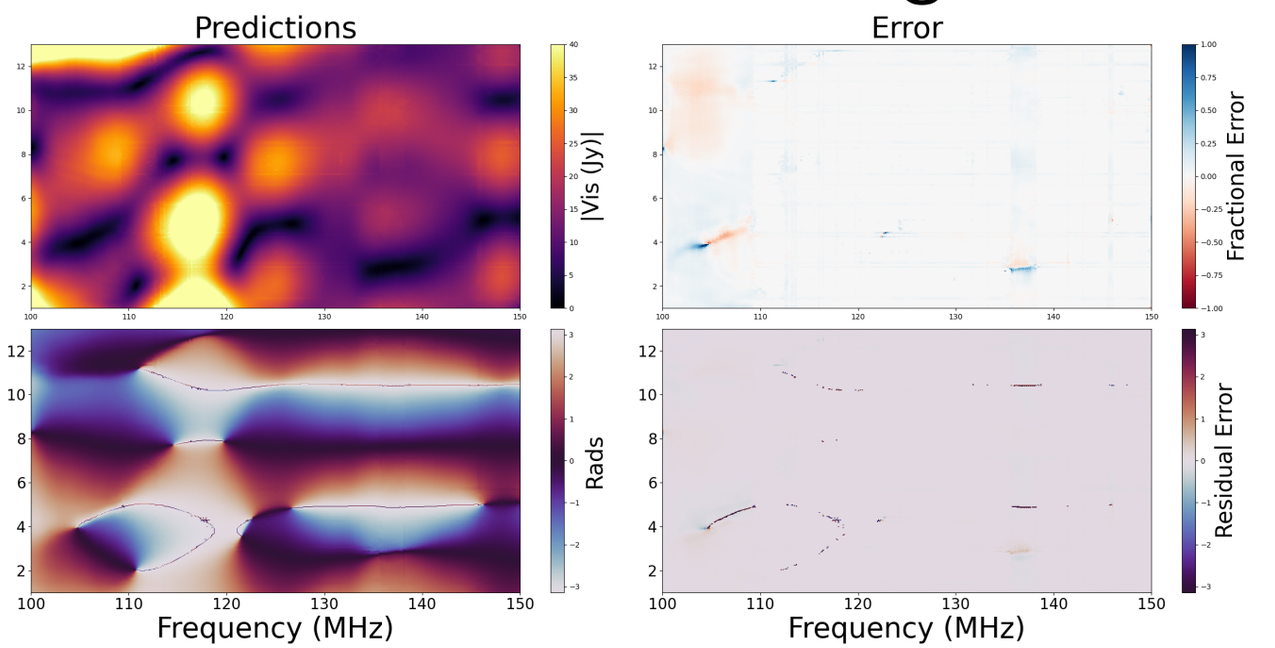
Constrained Gaussian Realizations



“CLEAN”



Machine Learning



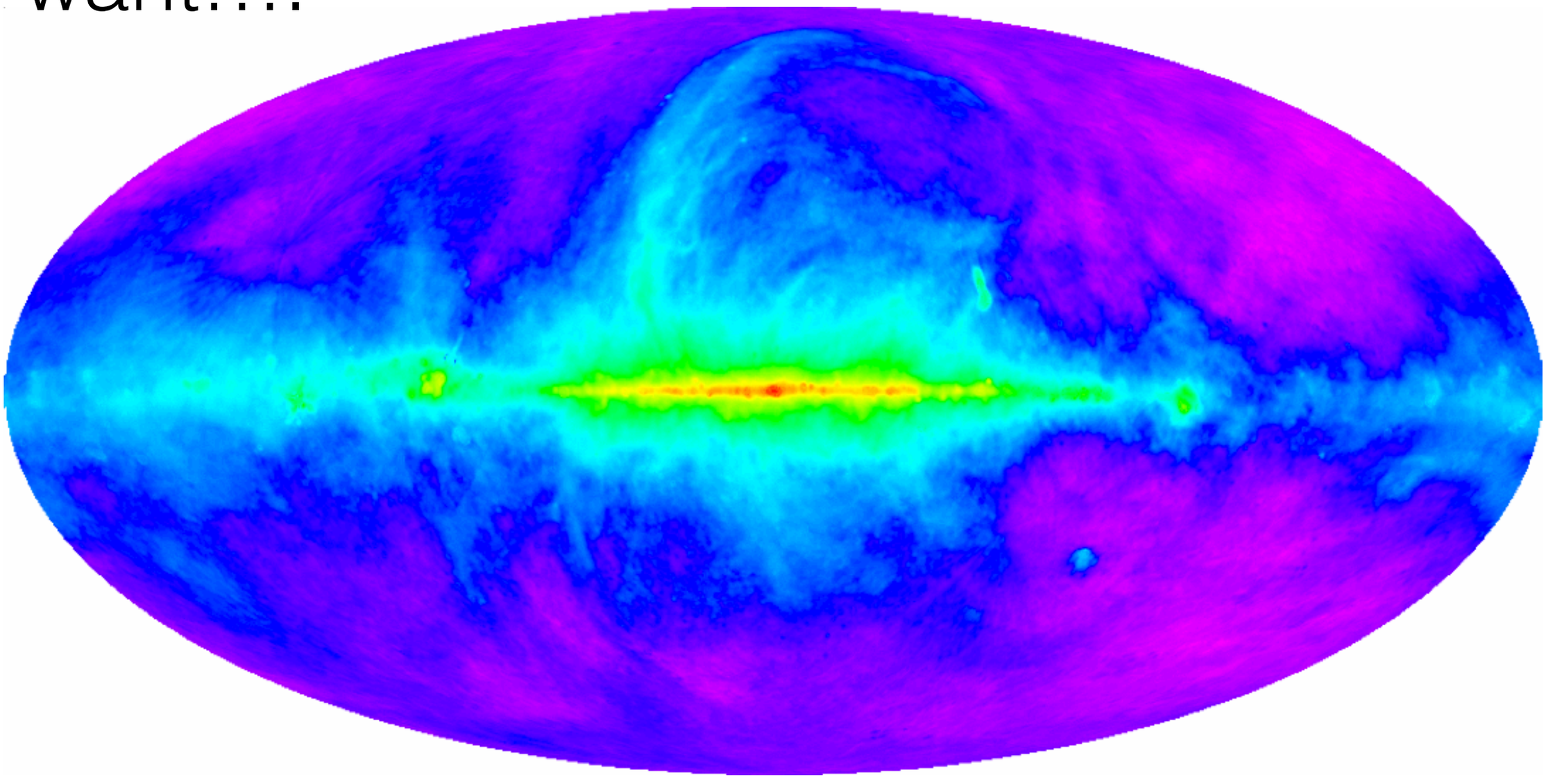
After RFI, there is an even greater challenge…

Our measurements are strongly contaminated by

foreground emission

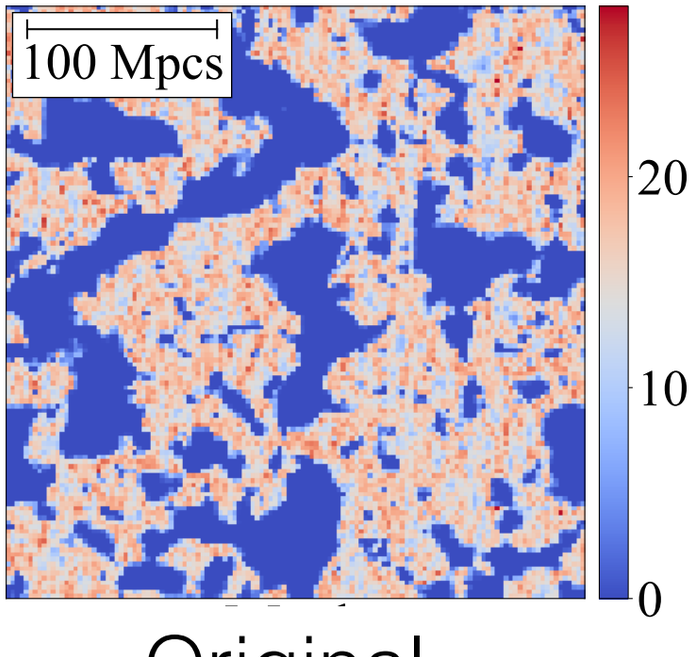
that is ~10^4 to 10^5 times brighter than the cosmological signal

A huge portion is not what you want….

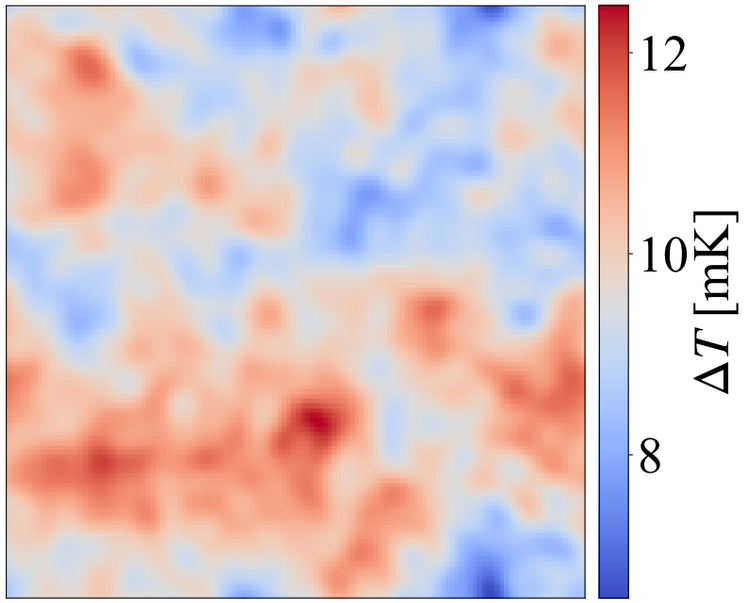


….and just a tiny portion is what you're interested in 

Filtering the data can get rid of contaminants, but destroy lots of information….

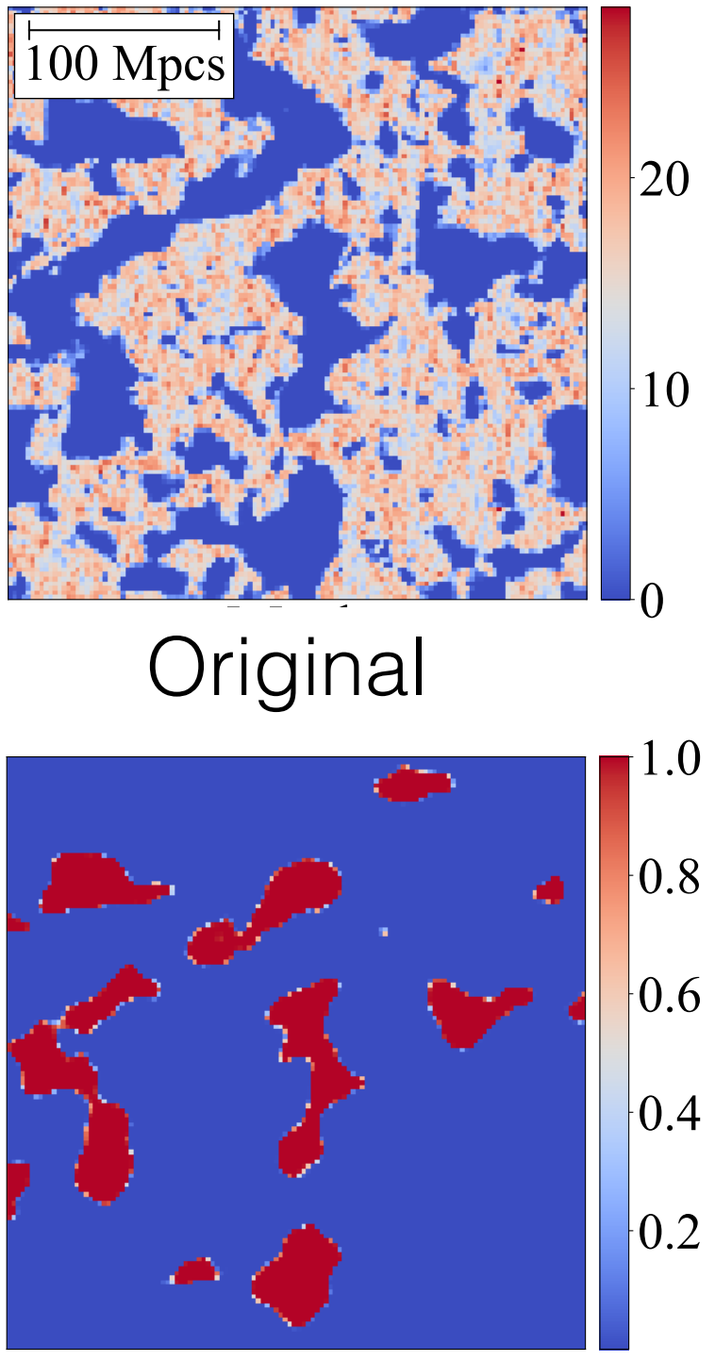


Original



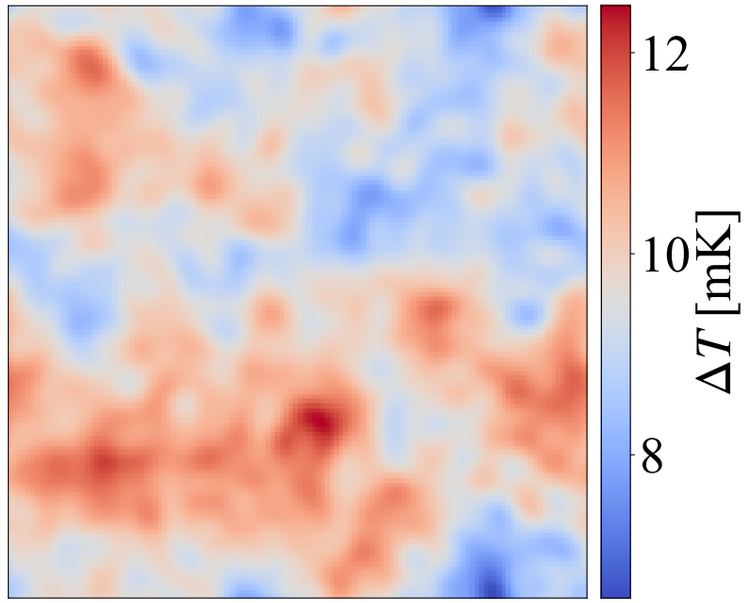
Filtered

Filtering the data can get rid of contaminants, but destroy lots of information….



Original

Predicted ionization



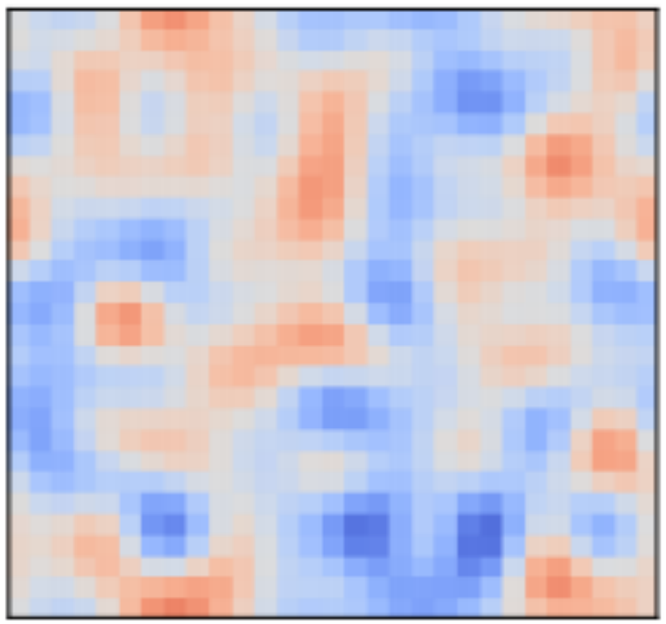
Filtered

….but perhaps machine learning and students can save the day! (Gagnon-Hartman, Cui, **AL**, Ravanbakhsh., 2021, MNRAS **504**, 4716)

Christine Cui

Sam Gagnon-Hartman

It may even work for current-generation instruments like HERA that aren't optimized for imaging



Filtered

Original

Predicted ionization

Gagnon-Hartman, Cui, **AL**, Ravanbakhsh (2021) MNRAS **504**, 4716

Jacob Kennedy

In both the short- and the long-term, the best constraints will come from

a combination of different probes

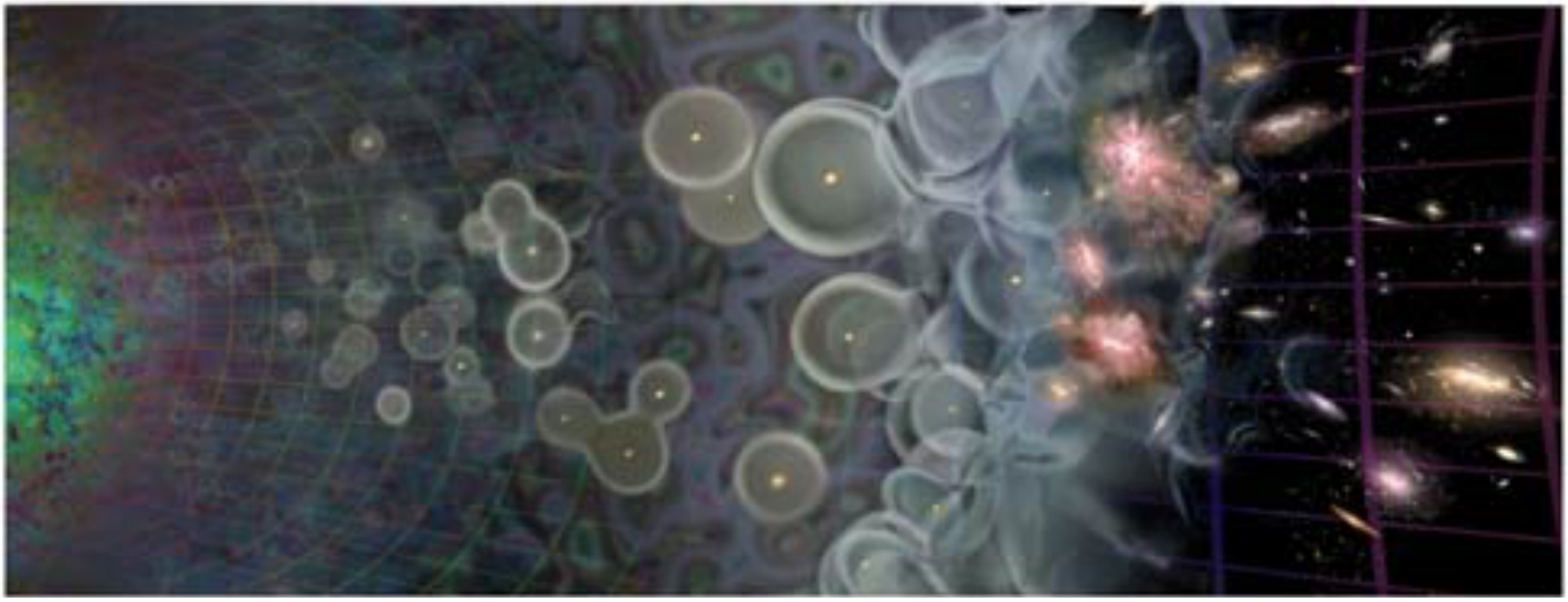
Case Study #1:

Combining Global 21cm Signal Measurements with CMB Kinetic Sunyaev-Zel'dovich Effect Measurements

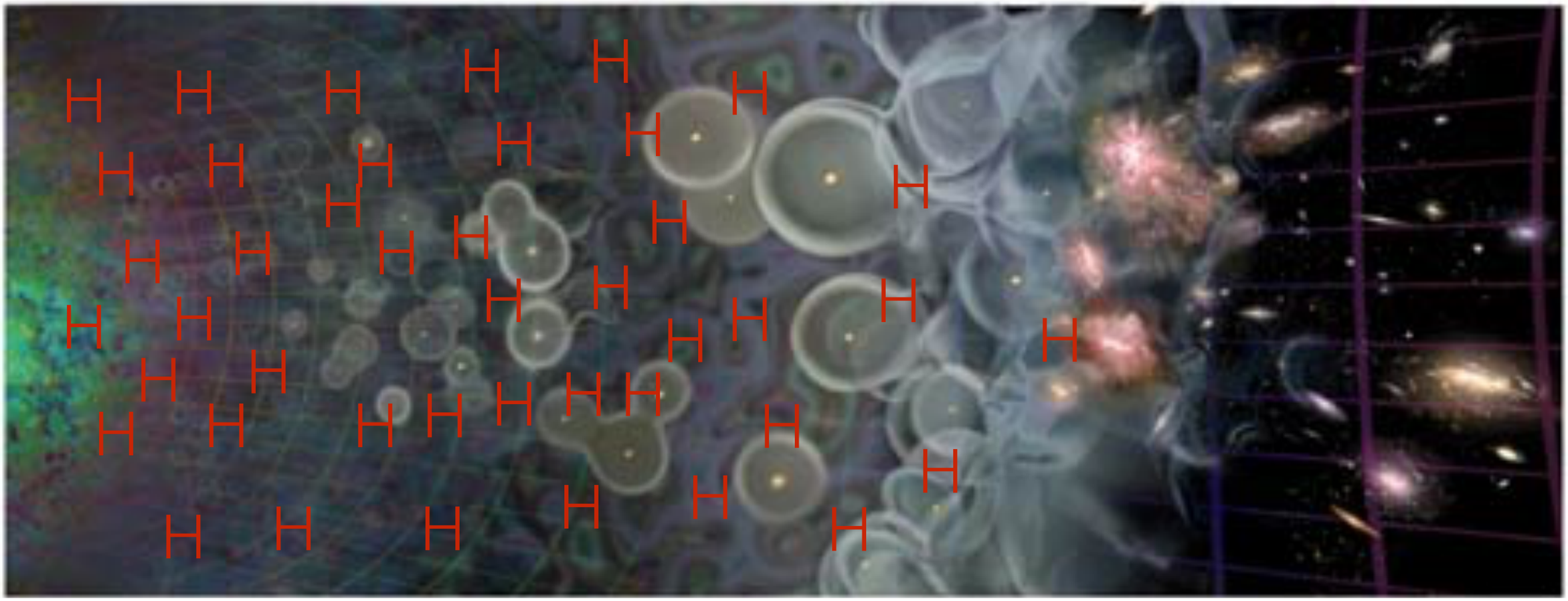
Joëlle-Marie Bégin

Bégin, **AL**, Gorce (2022), PRD **105**, 083503

Adélie Gorce



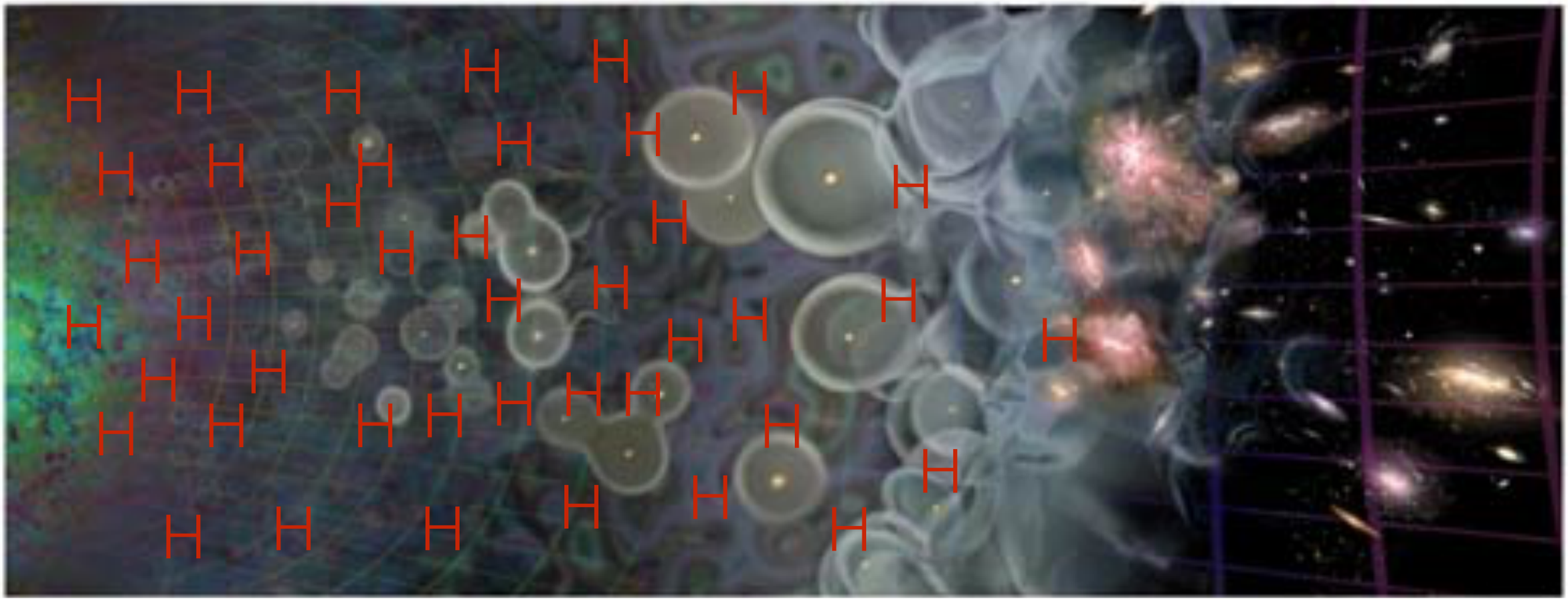
Epoch of
Reionization



Epoch of Reionization

Neutral hydrogen (21cm)

z



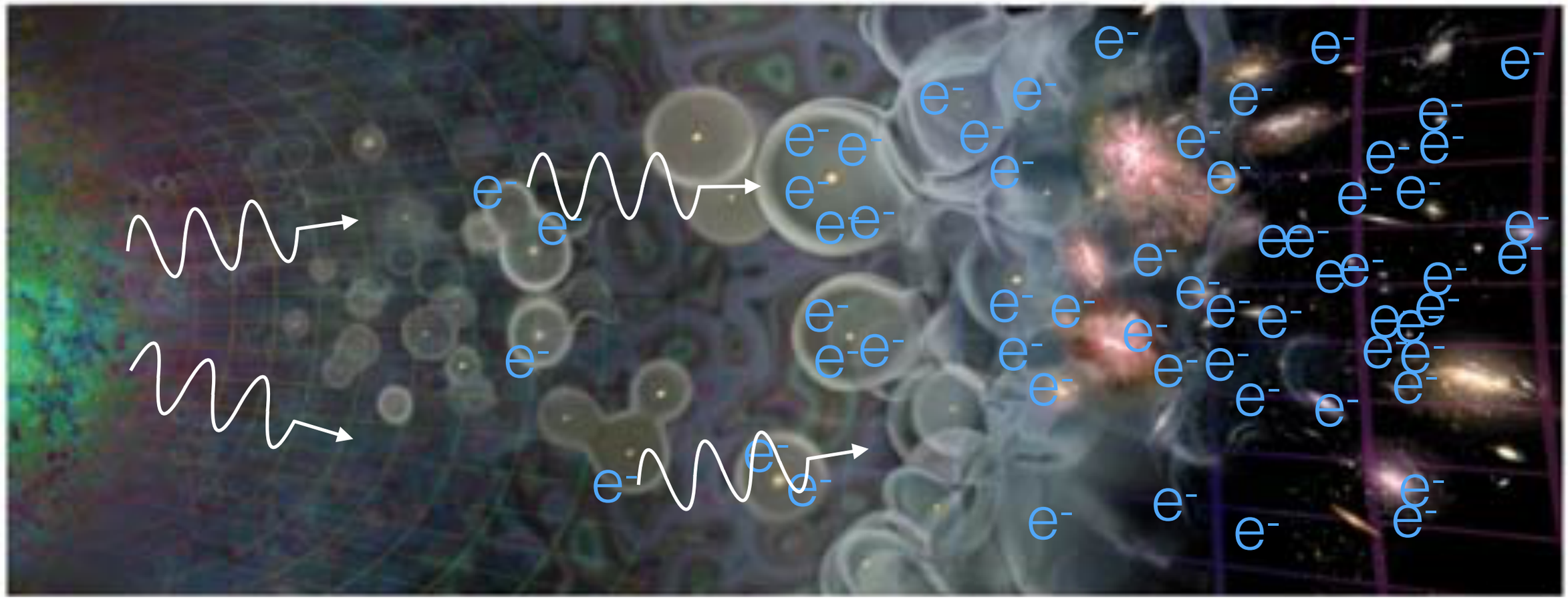
Epoch of Reionization

Neutral hydrogen (21cm)

Foregrounds make it hard for 21cm measurements to capture extended/smooth features in the ionization history

z

ν

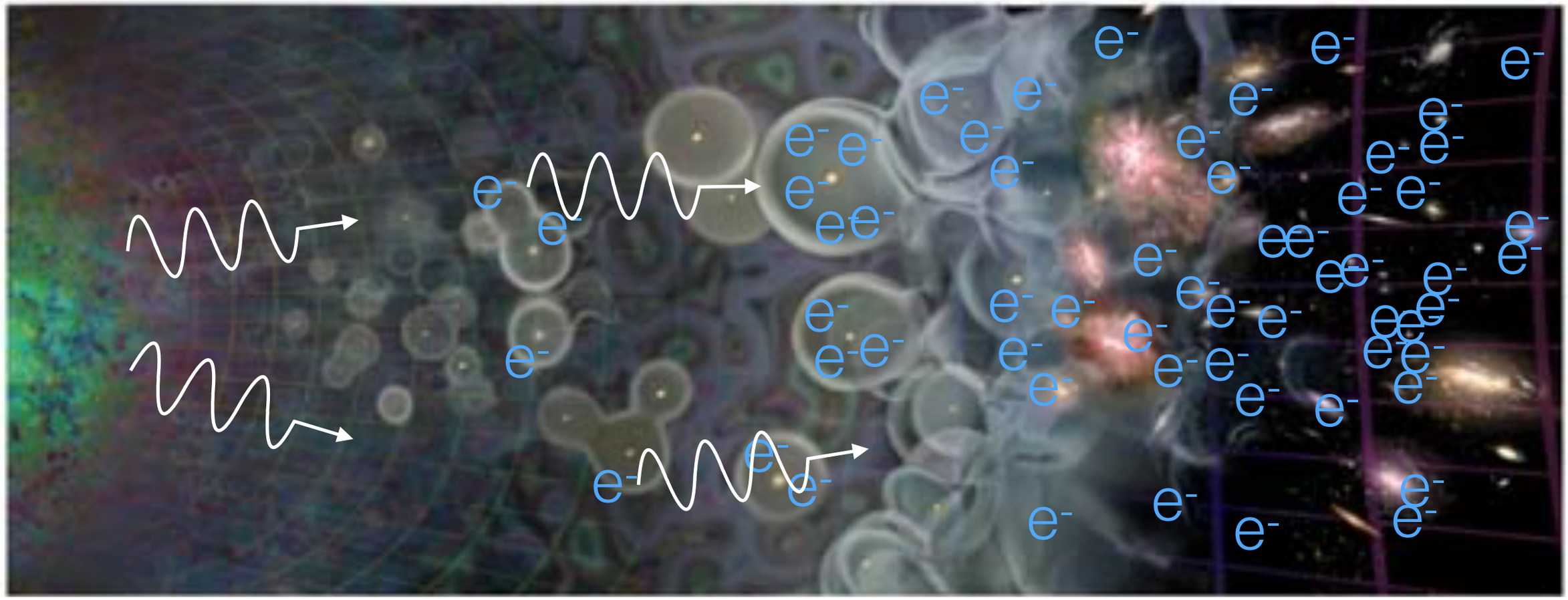


Epoch of Reionization

Neutral hydrogen (21cm)

Free electrons (CMB)

z



Epoch of Reionization

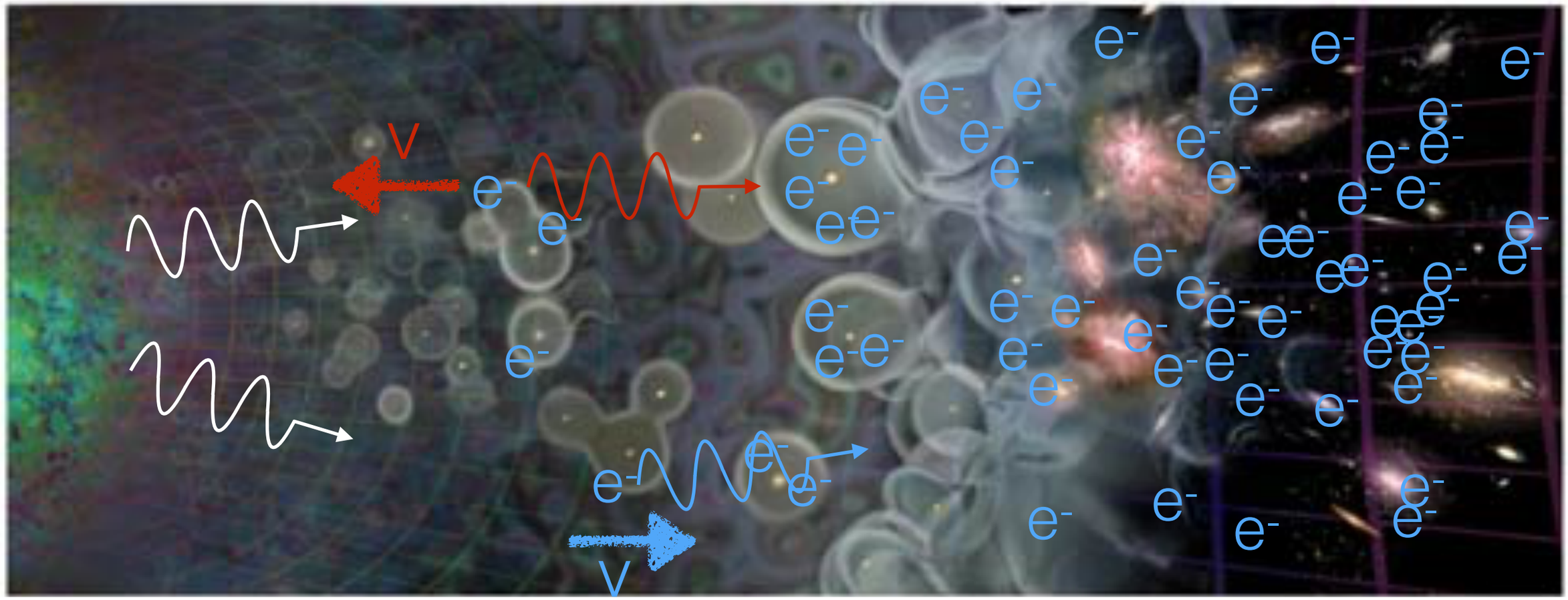
Neutral hydrogen (21cm)

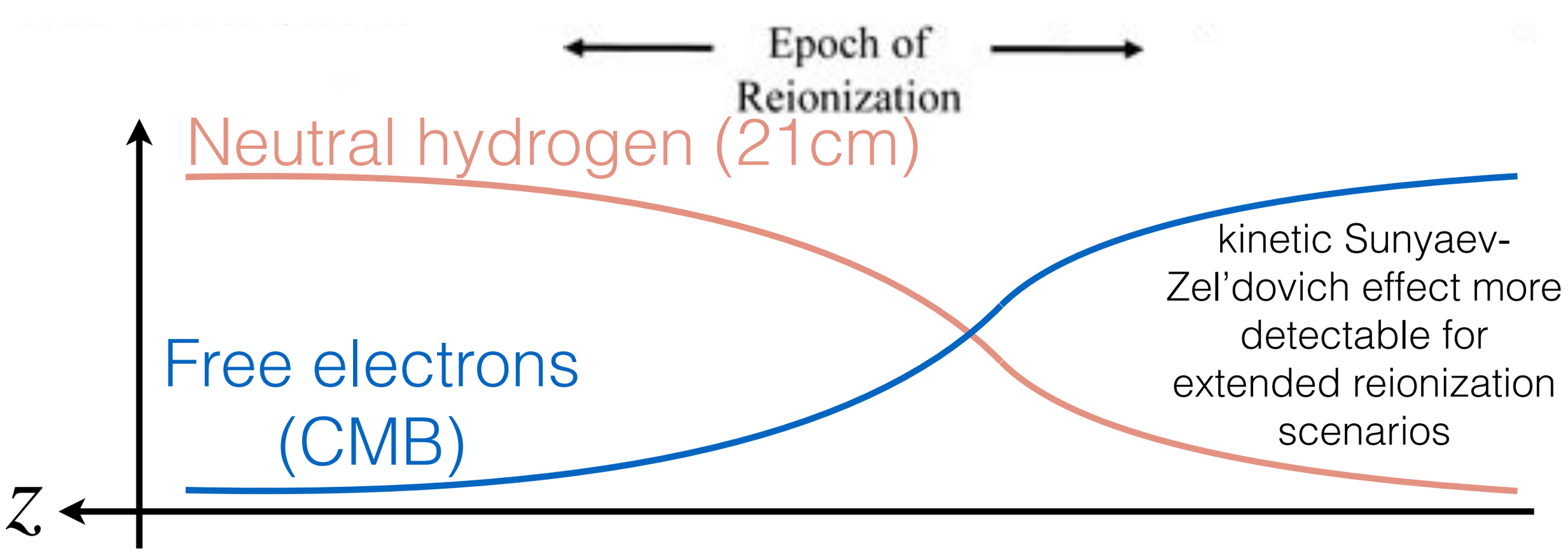
Free electrons (CMB)

z

Optical depth of scattering

$$\tau = \sigma_T \int n_e(l) dl$$


v
e-
v


Epoch of Reionization
Neutral hydrogen (21cm)
Free electrons (CMB)
z
kinetic Sunyaev-Zel'dovich effect more detectable for extended reionization scenarios

The complementarity of our two probes can be brought out using a **Karhunen-Loève basis to describe** reionization histories

The complementarity of our two probes can be brought out using a **Karhunen-Loève basis to describe** reionization histories

$$\mathbf{F}_{\mathrm{kSZ}}\mathbf{v} = \lambda \mathbf{F}_{21}\mathbf{v}$$

$N_z \times N_z$ Fisher information matrix for kSZ ionization history measurements

The complementarity of our two probes can be brought out using a **Karhunen-Loève basis to describe** reionization histories

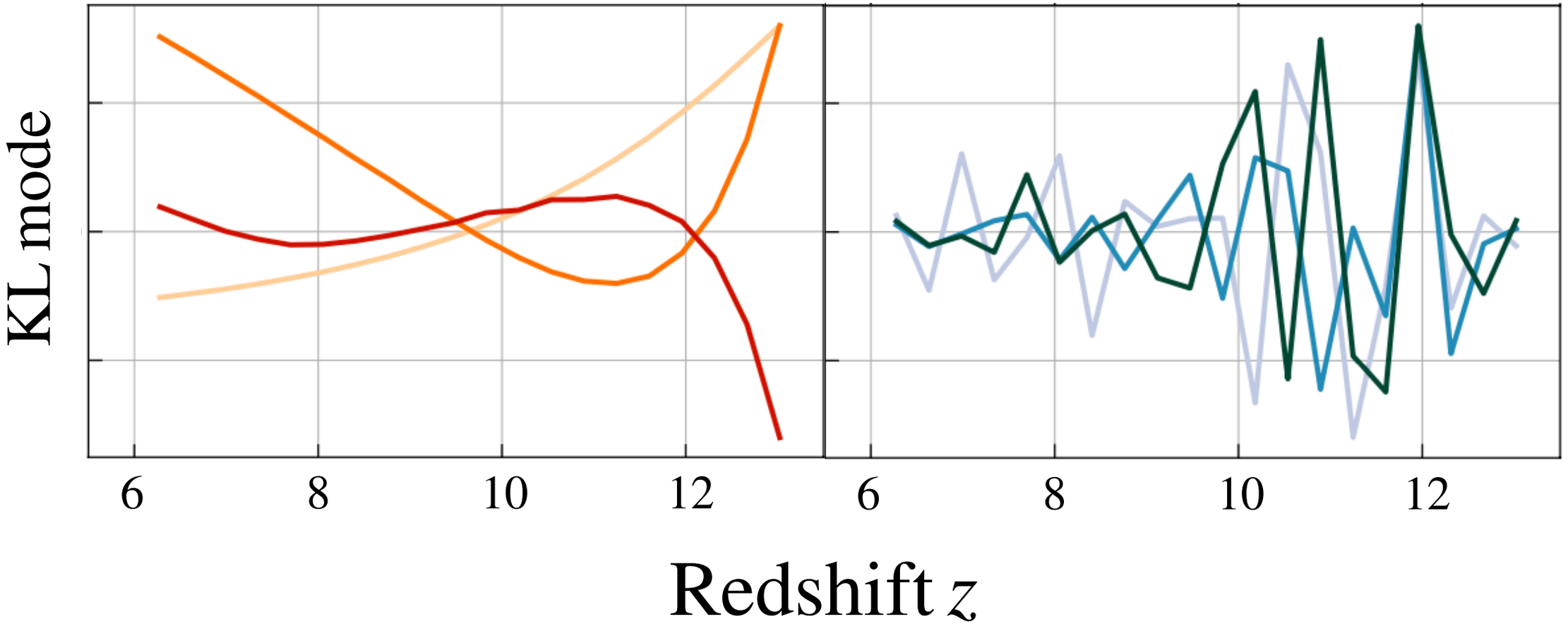
$$\mathbf{F}_{\mathrm{kSZ}}\mathbf{v} = \lambda \mathbf{F}_{21}\mathbf{v}$$

$N_z \times N_z$ Fisher information matrix for kSZ ionization history measurements

Solving this generalized eigenvalue problem gives a set of basis vectors **rank-ordered by how well they can be probed by 21cm vs kSZ**

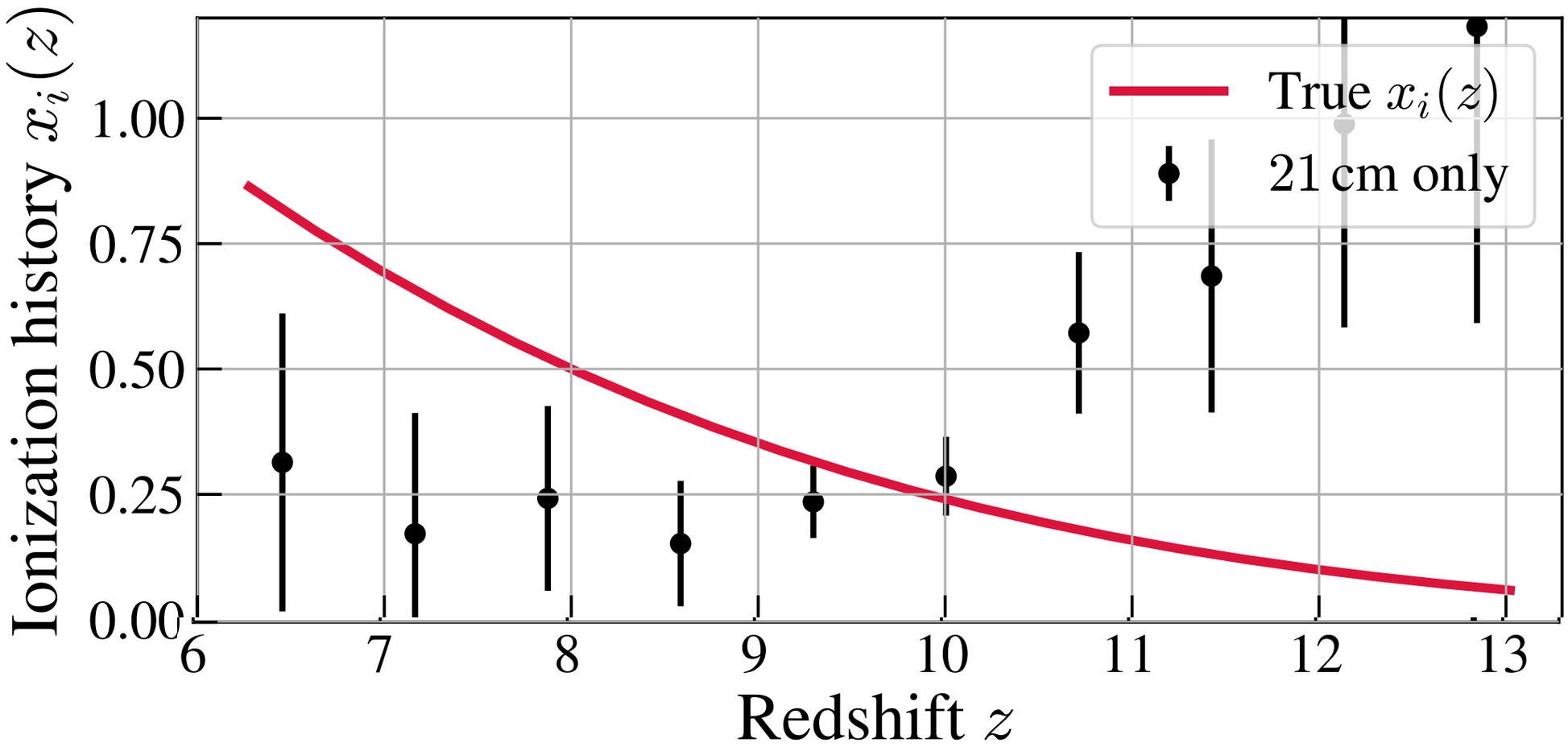
The complementarity of our two probes can be brought out using a **Karhunen-Loève basis to describe** reionization histories

$$\mathbf{F}_{\mathrm{kSZ}}\mathbf{v} = \lambda \mathbf{F}_{21}\mathbf{v}$$

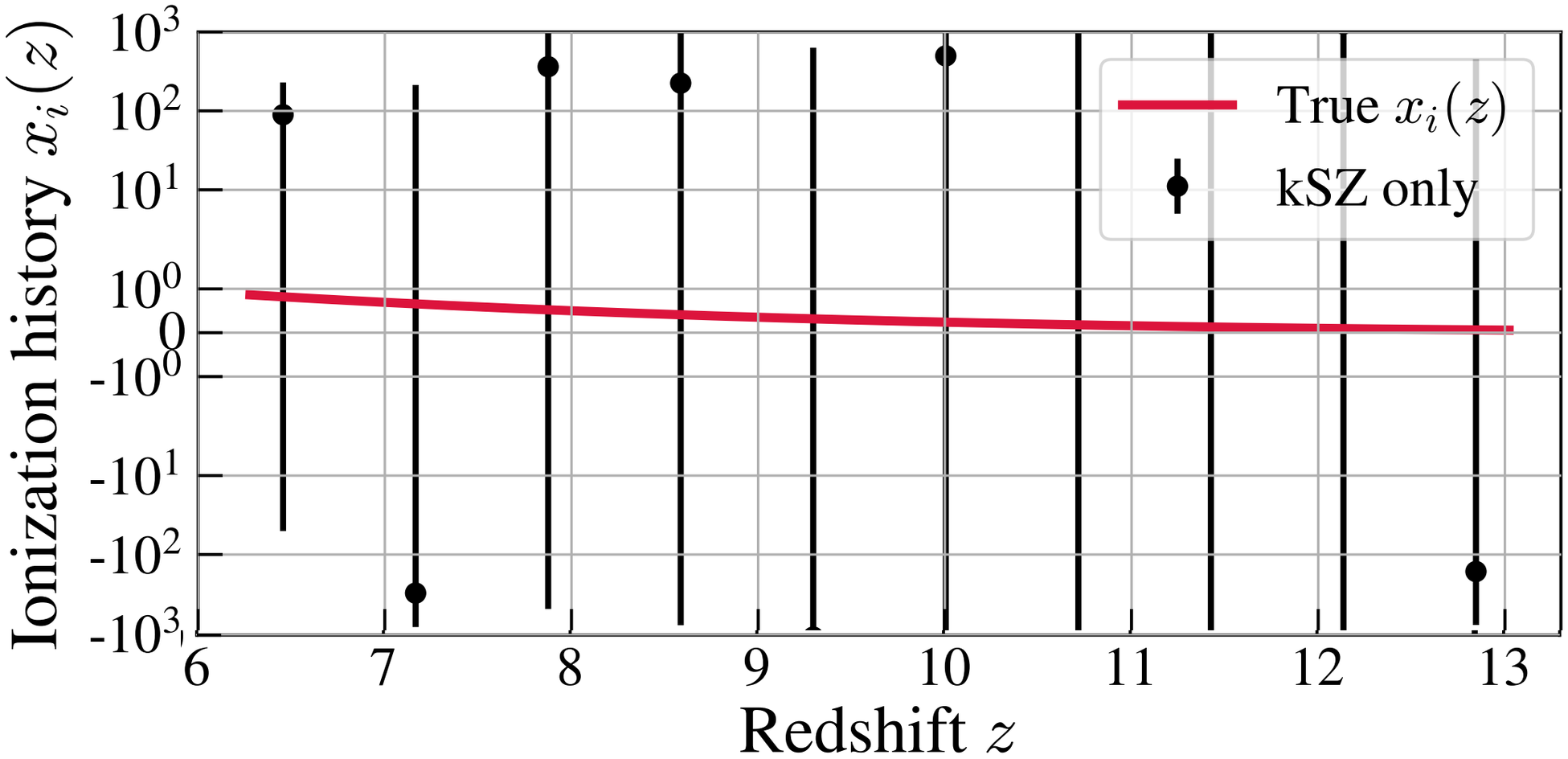


Higher-order CMB information (like from the **kinetic Sunyaev-Zel'dovich** effect) can be used in conjunction with the **global 21cm signal** to product **model-independent constraints** on reionization histories

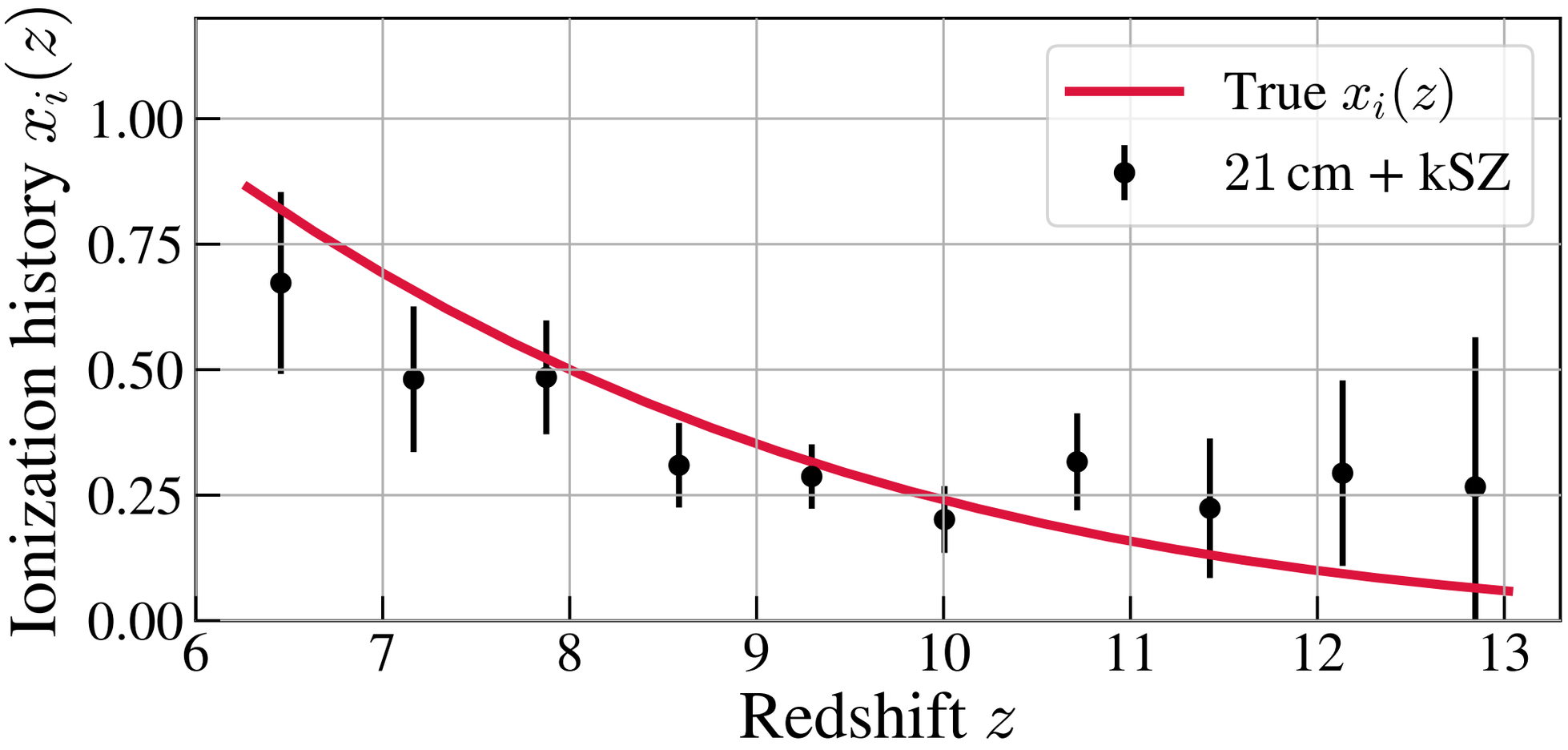
Higher-order CMB information (like from the **kinetic Sunyaev-Zel'dovich** effect) can be used in conjunction with the **global 21cm signal** to product **model-independent constraints** on reionization histories

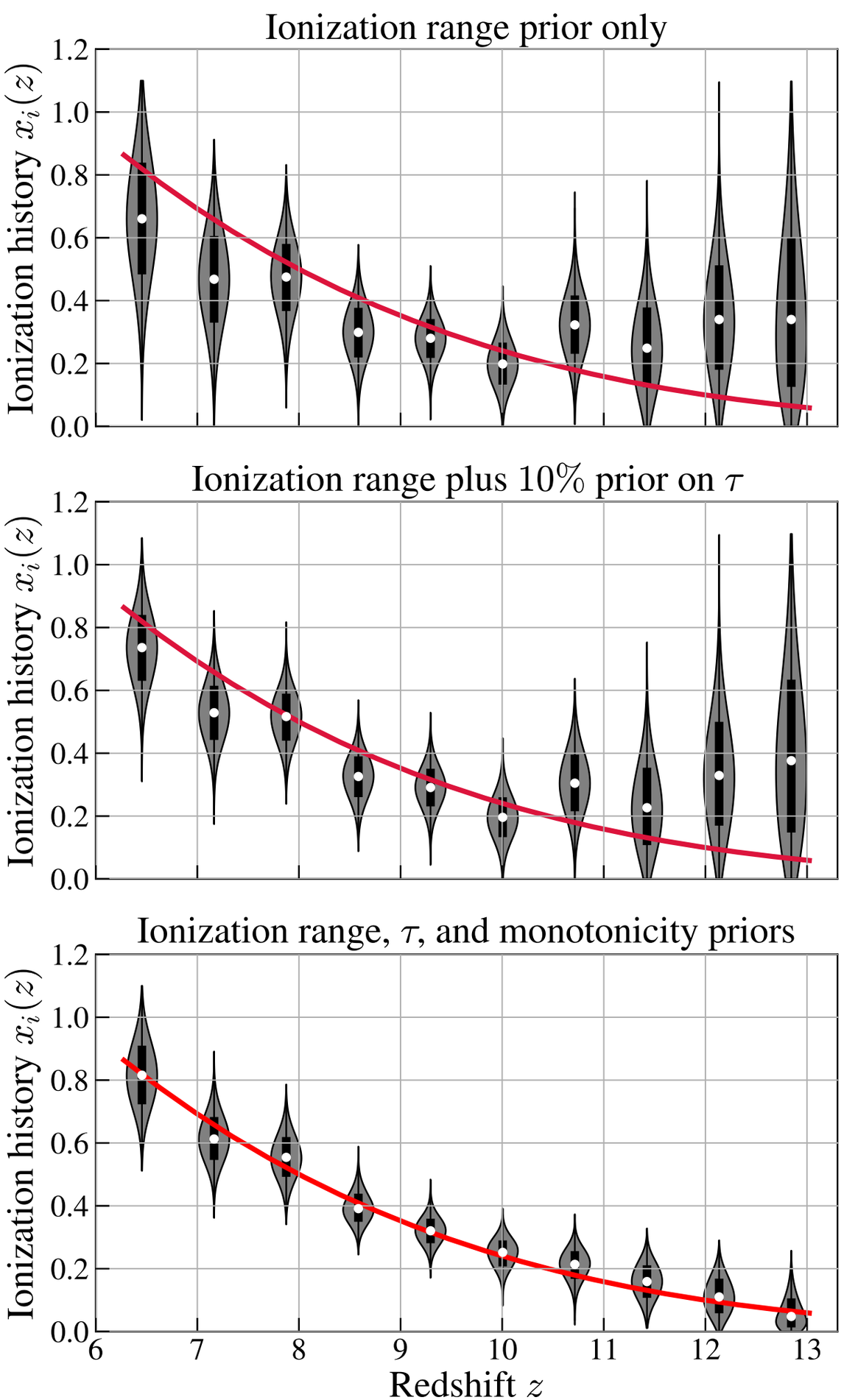


Higher-order CMB information (like from the **kinetic Sunyaev-Zel'dovich** effect) can be used in conjunction with the **global 21cm signal** to product **model-independent constraints** on reionization histories

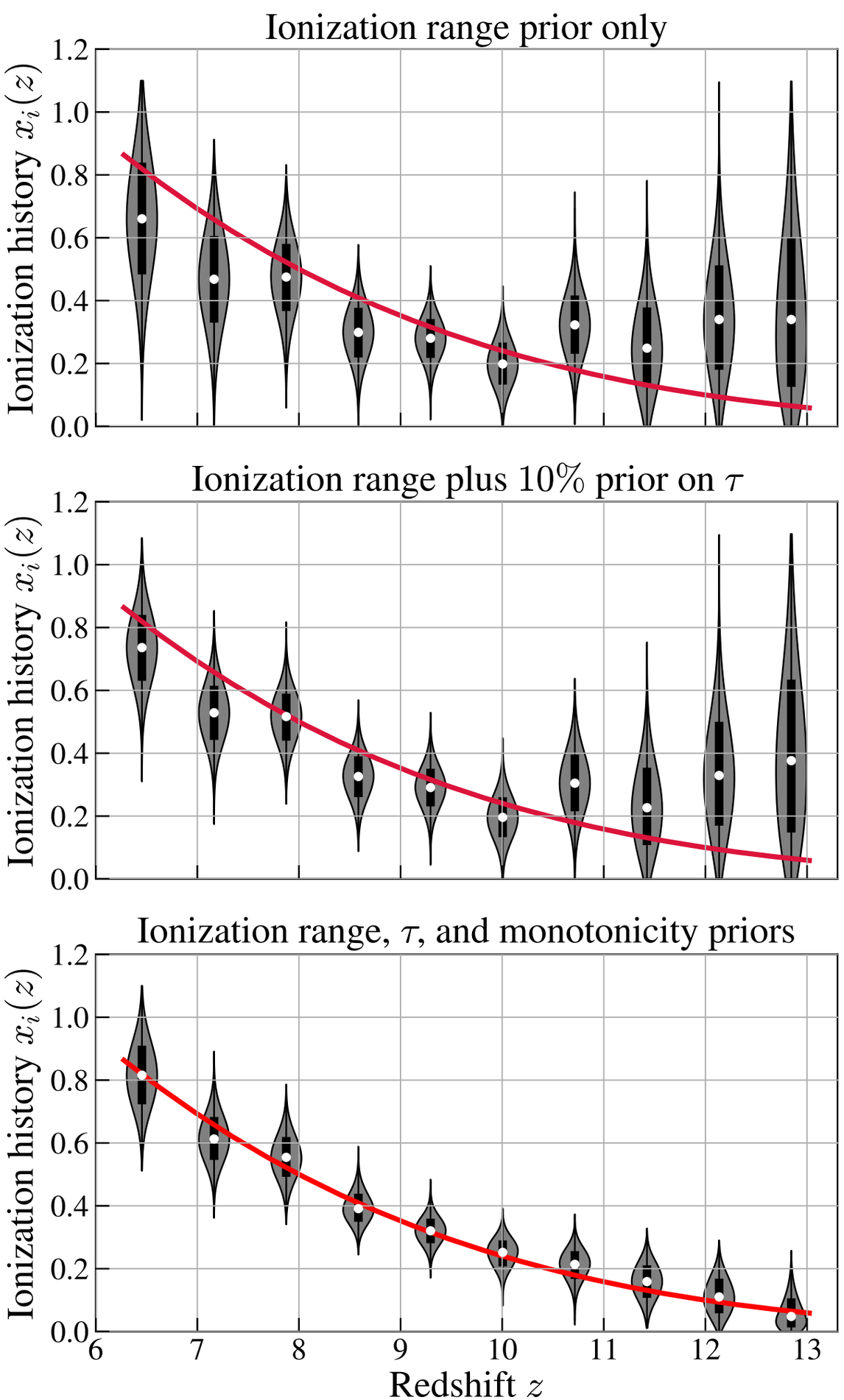


Higher-order CMB information (like from the **kinetic Sunyaev-Zel'dovich** effect) can be used in conjunction with the **global 21cm signal** to product **model-independent constraints** on reionization histories





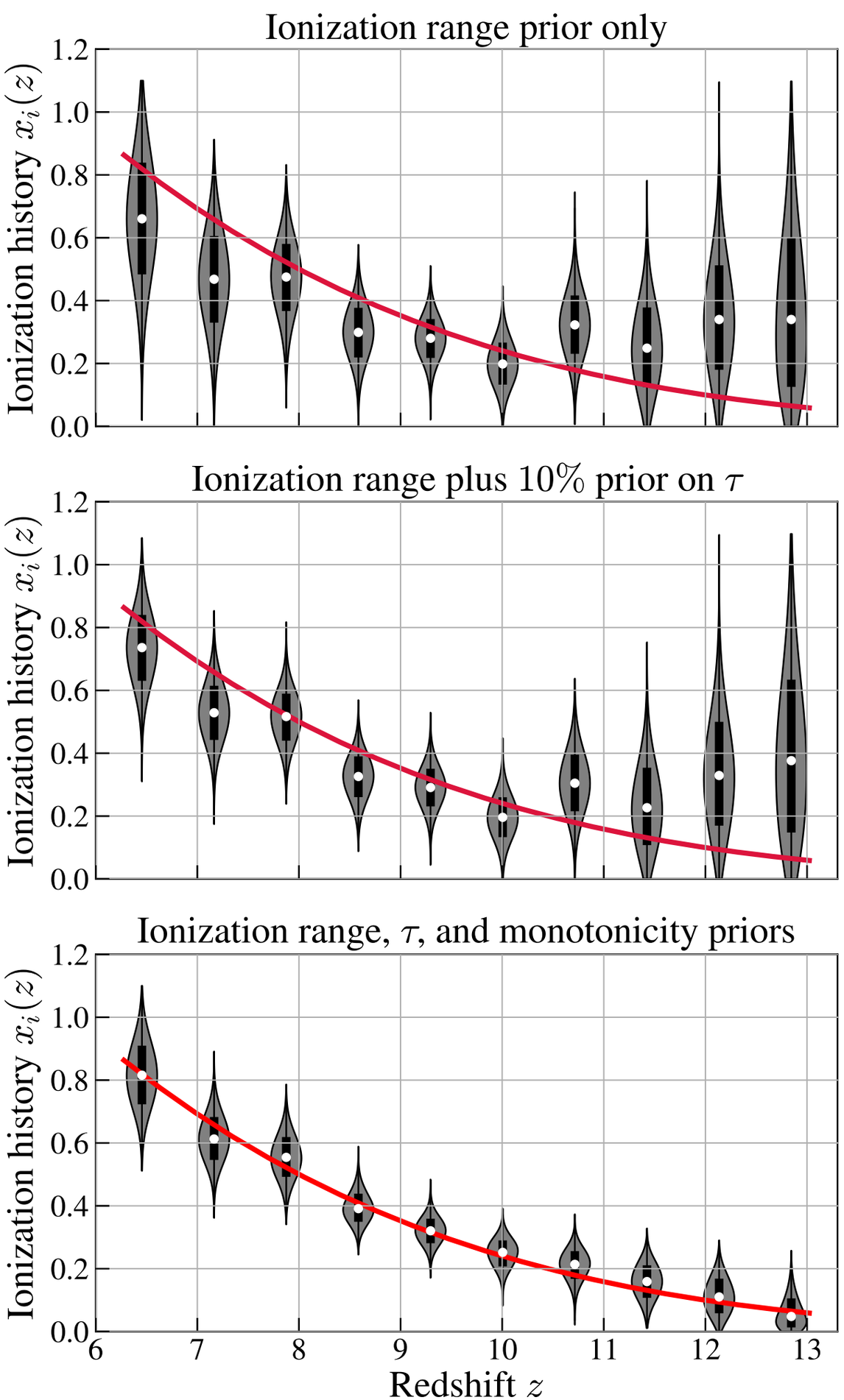
Adding other priors can further sharpen constraints….



Adding other priors can further sharpen constraints….

Optical depth of scattering

$$\tau = \sigma_T \int n_e(l) dl$$



Adding other priors can further sharpen constraints….

…and combining probes also provides safeguards against potential systematics

Bégin, **AL**, Gorce (2022), PRD **105**, 083503