Strategies to identify the Galactic Foreground

Isabella Paola Carucci

Università degli Studi di Torino

Physics
Torino
Department of Excellence

BAM & Radio Synchrotron Background Conference
Barolo, 17th June 2022

1. Cosmology with **Hydrogen Intensity Mapping (IM)**

2. Biggest challenge: weakness of the IM signal compared to contaminants. **Available strategies and ongoing efforts**

3. The **galactic synchrotron** as a bonus: the case of MeerKLASS

4. MeerKLASS X WiggleZ: we detected the *first ever* cosmological signal with an array in single-dish mode. **Getting ready for the SKA Observatory!**

5. Quick interlude on the **gamma-ray sky**

6. Why these strategies could be of use for you

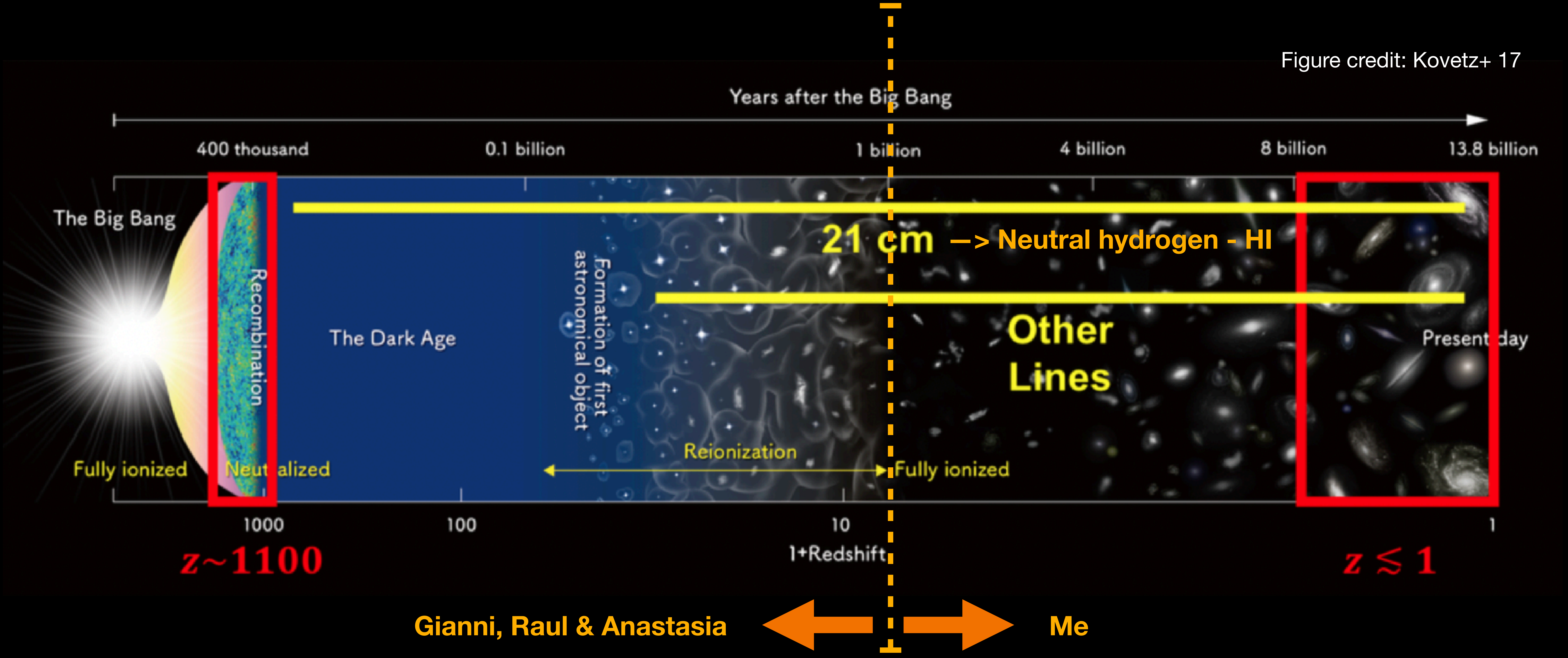
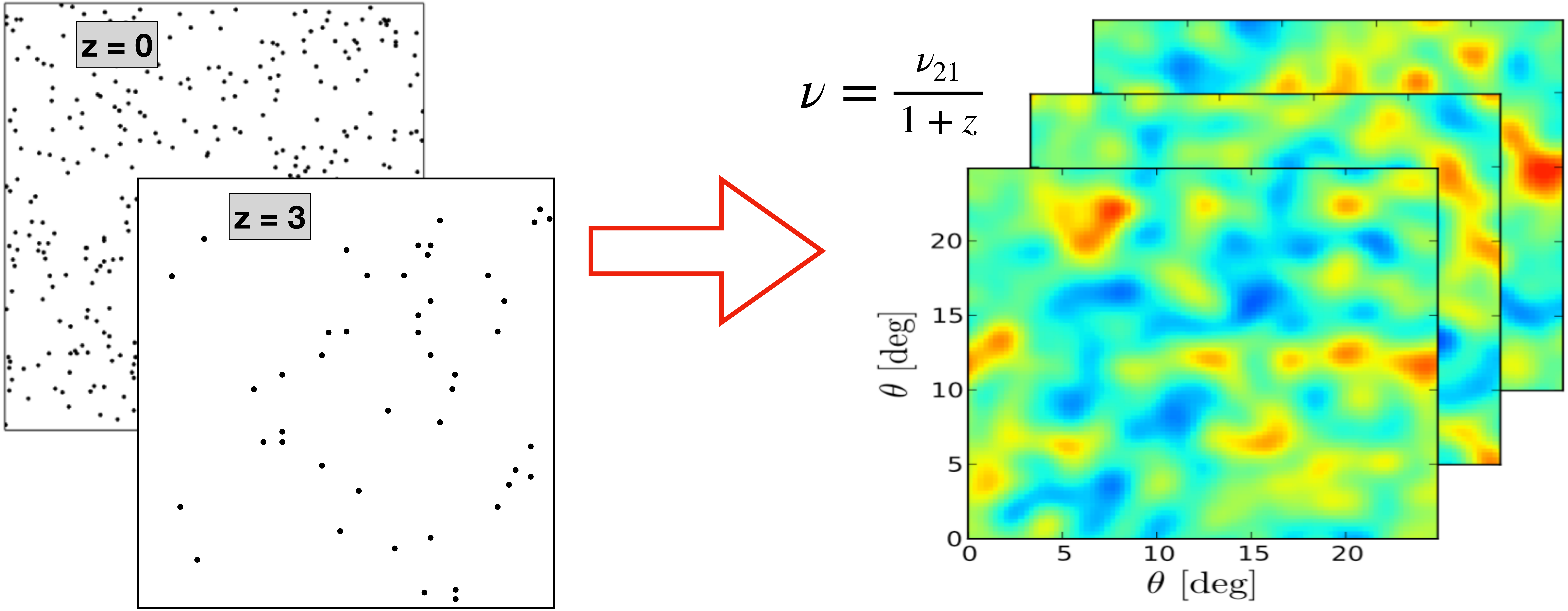


Figure credit: Kovetz+ 17
Years after the Big Bang
400 thousand
0.1 billion
1 billion
4 billion
8 billion
13.8 billion
The Big Bang
Recombination
The Dark Age
Formation of first astronomical object
21 cm —> Neutral hydrogen - HI
Other Lines
Present day
Fully ionized
Neutralized
Reionization
Fully ionized
1000
100
10
1
1+Redshift
z~1100
z ≲ 1
Gianni, Raul & Anastasia
Me

21-cm intensity mapping



Big volumes (for cheap) and high redshift resolution

HI intensity mapping with the SKAO

Proposed SKA1 Cosmology Surveys

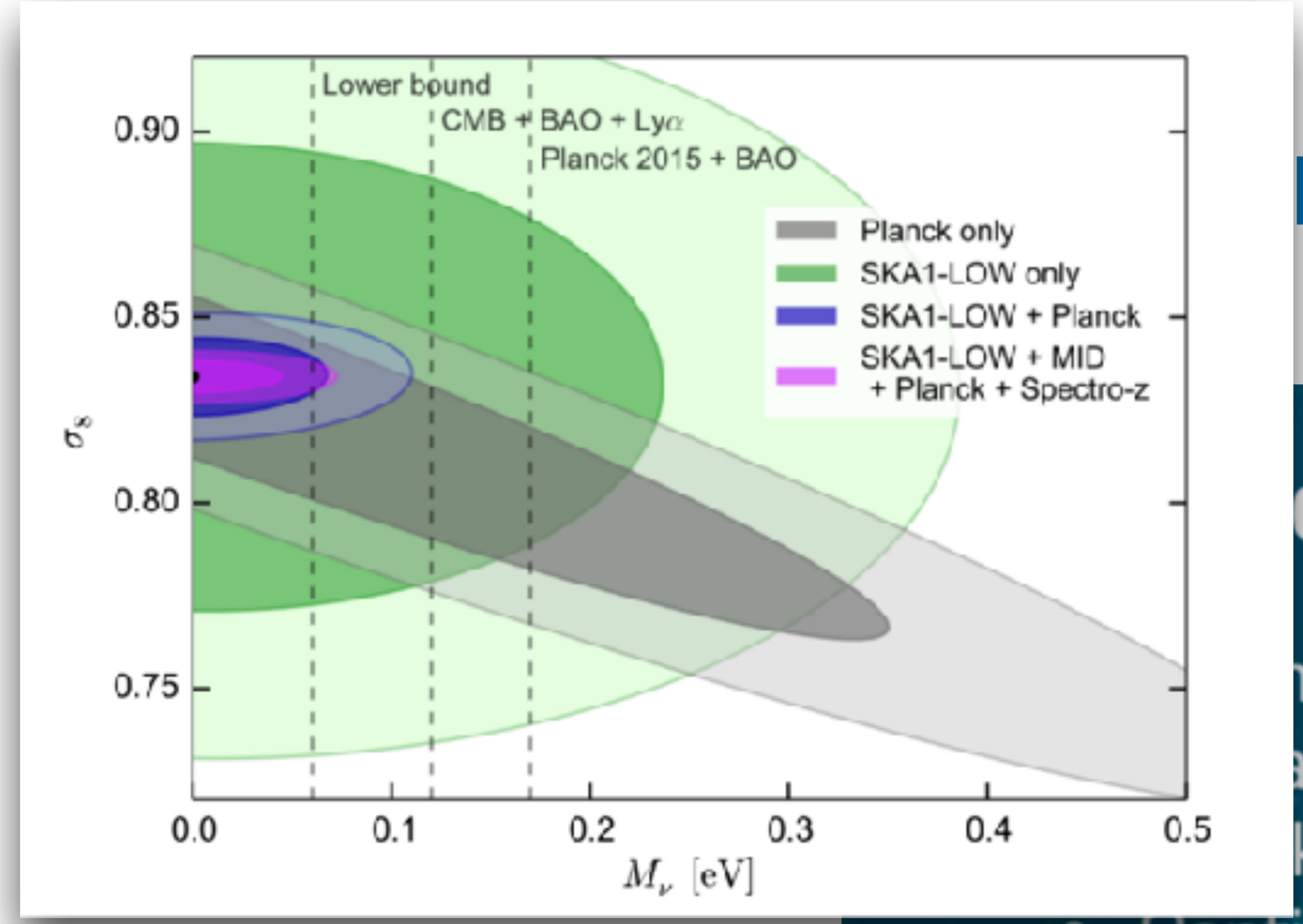
a) Medium–Deep Survey of 5,000 deg^2 at 0.95-1.4 GHz for
- HI galaxy redshift survey with 3.5 million objects
- Weak Lensing shape measurements with ~50 million objects
- Continuum galaxy survey with ~60 million objects

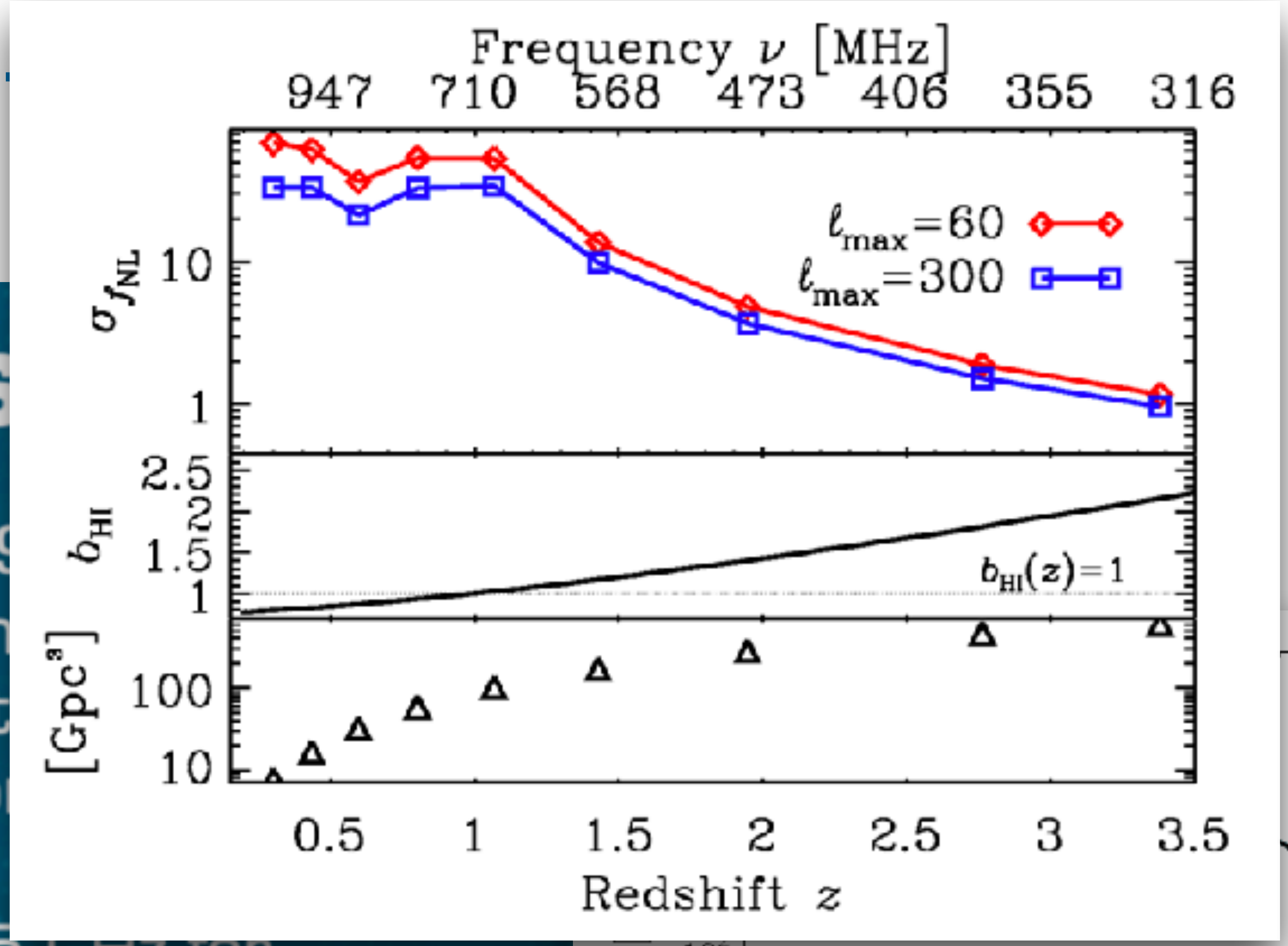
b) Wide Survey of 20,000 deg^2 at 0.35-1.05 GHz for
- Continuum galaxy survey with ~100 million objects
- HI intensity maps for $0.35<z<3$

c) Deep Survey 100 deg^2 at 200-350 MHz for
- HI intensity maps for $3<z<6$

Cosmology with Phase 1 of the Square Kilometre Array **Red Book** 2018:
Technical specifications and performance forecasts

nsity mapping with



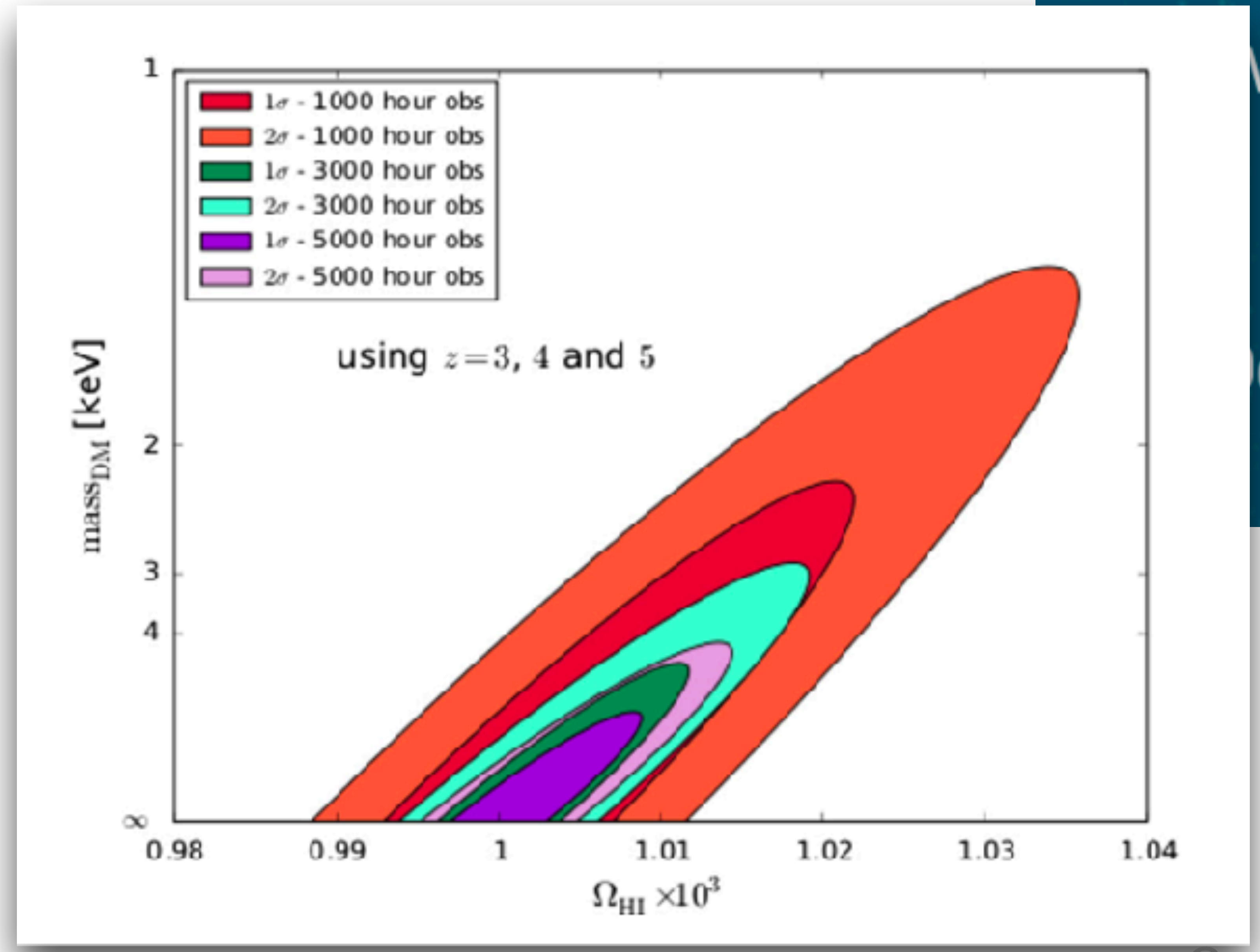


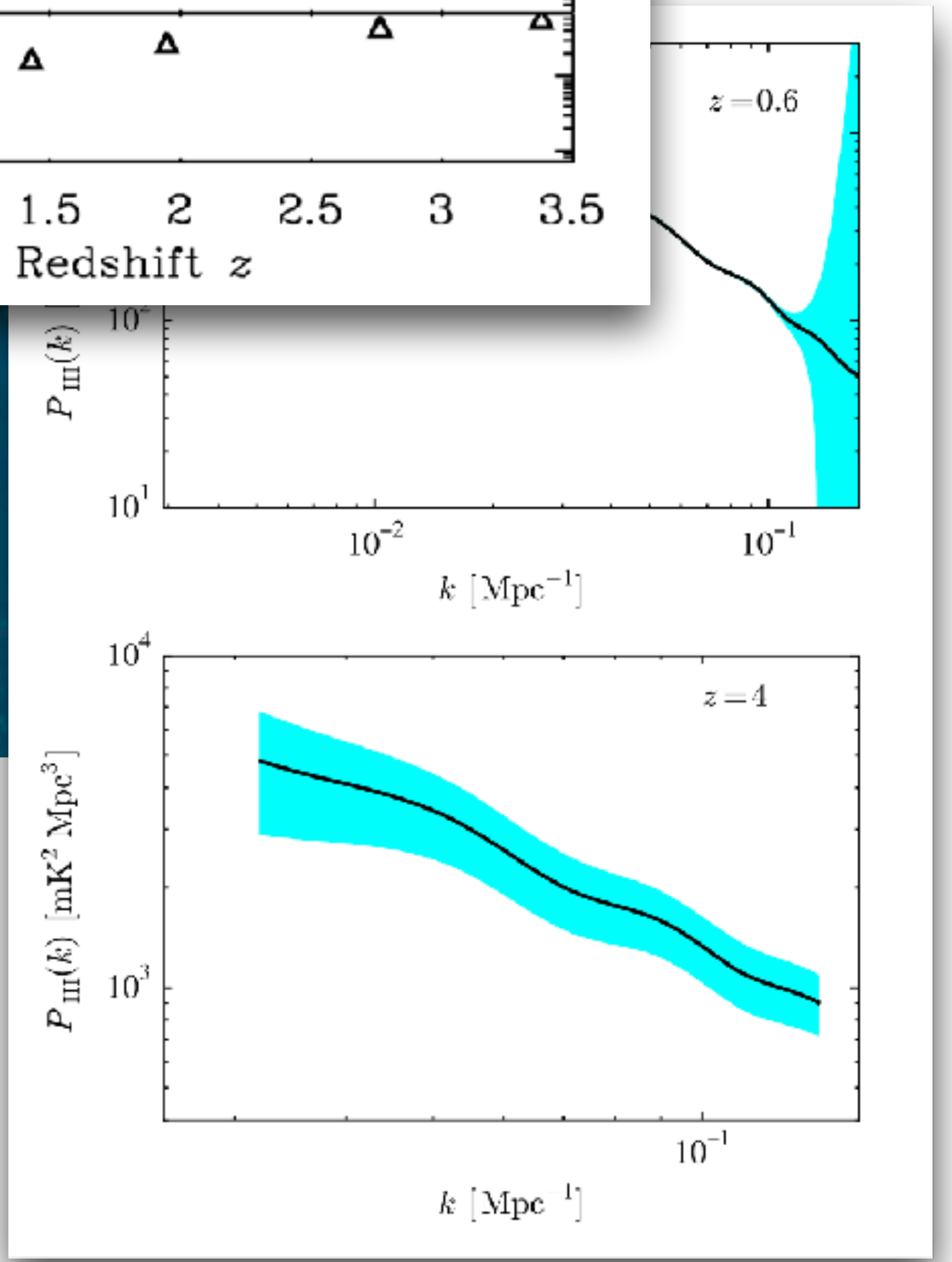
ed SKA1 Cosmology S

n–Deep Survey of 5,000 deg² at 0.9
alaxy redshift survey with 3.5 million
k Lensing shape measurements wit
- Continuum galaxy survey with ~60 millio

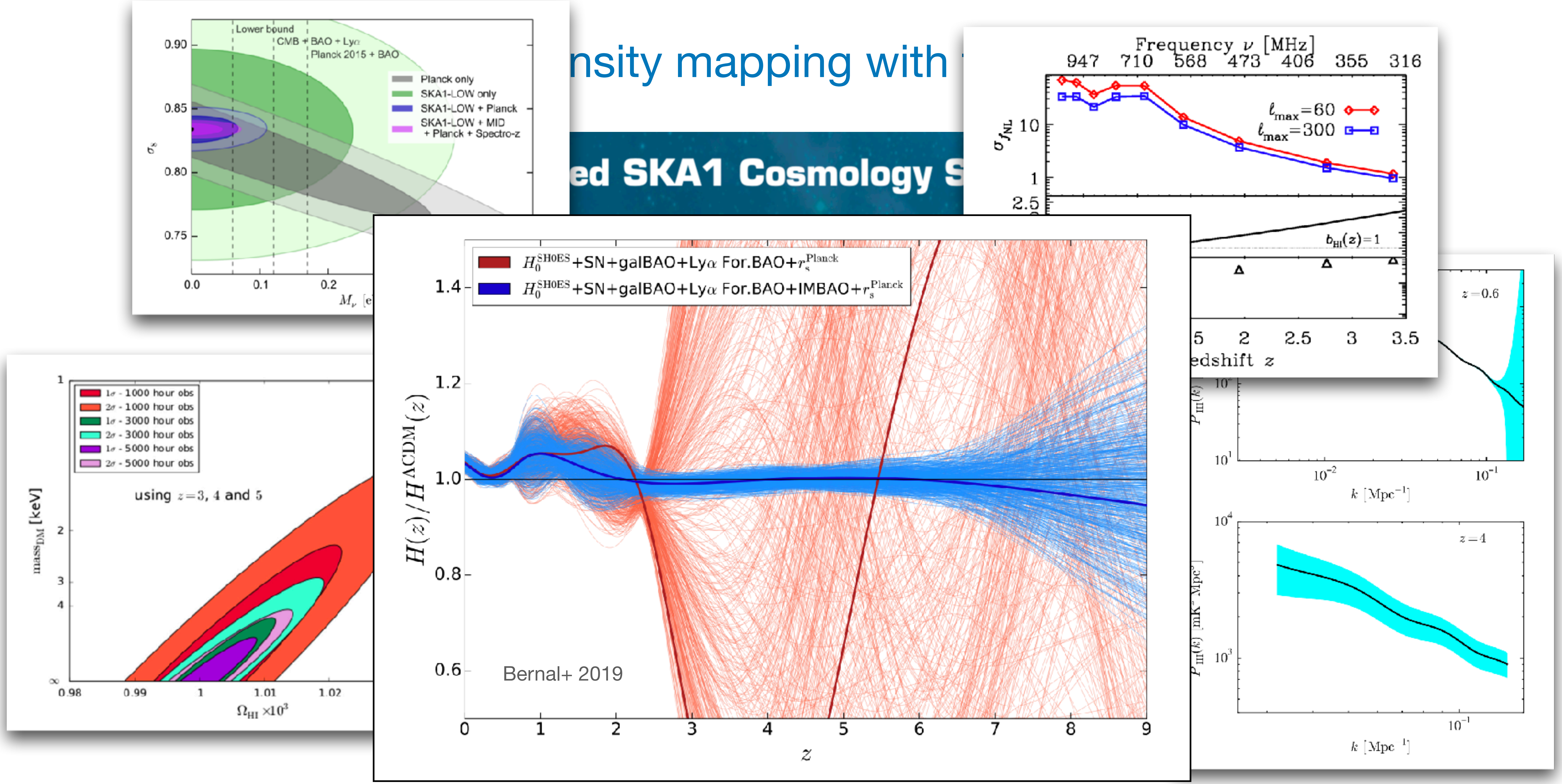
Vide Survey of 20,000 deg² at 0.35-1.05 GHz for
- Continuum galaxy survey with ~100 million objects
- HI intensity maps for 0.35<z<3

eep Survey 100 deg² at 200-350 MHz for
- HI intensity maps for 3<z<6



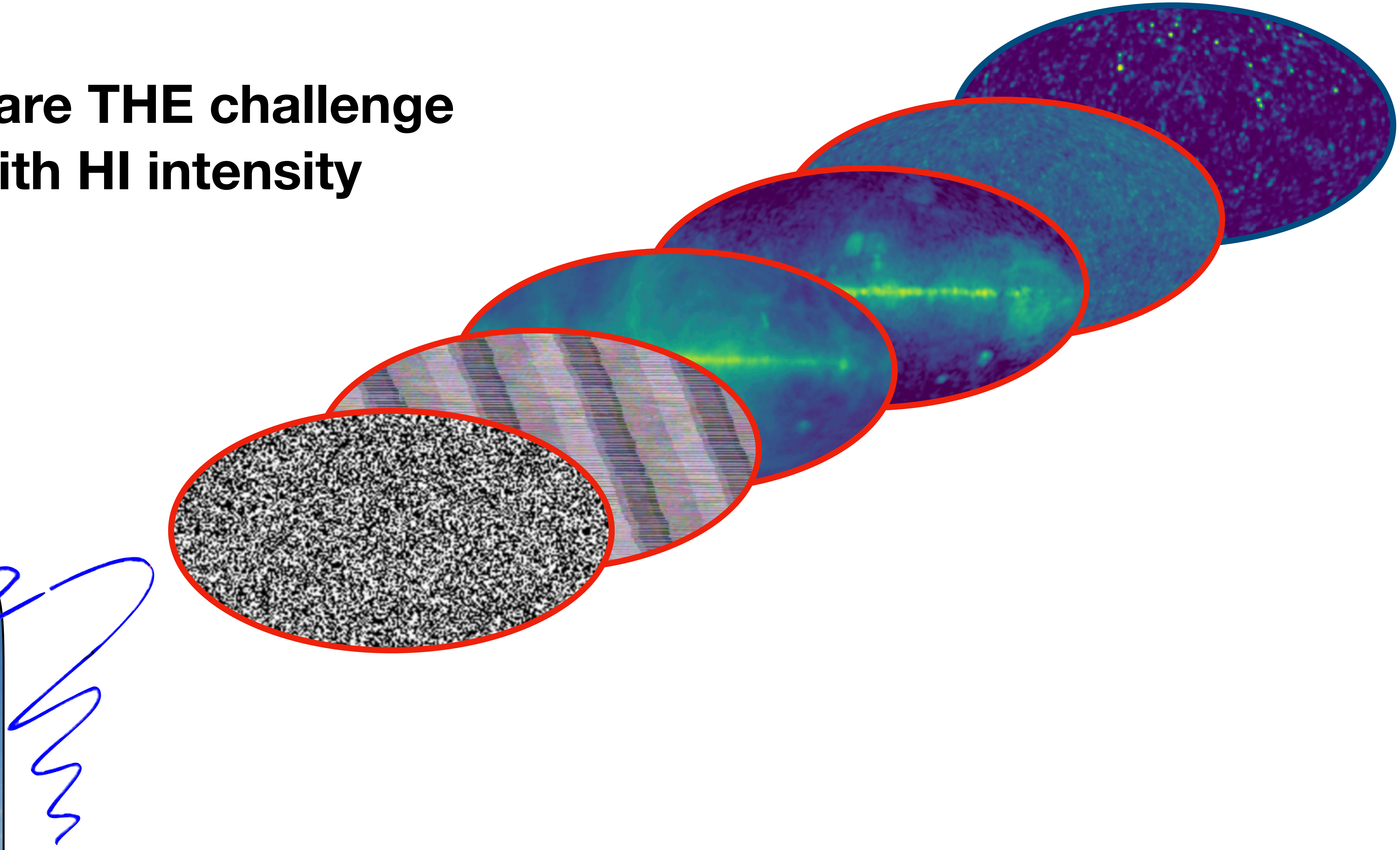


Cosmology with Phase 1 of the Square Kilometre Array **Red Book** 2018:
Technical specifications and performance forecasts


nsity mapping with
ed SKA1 Cosmology S
Lower bound
CMB + BAO + Lyα
Planck 2015 + BAO
Planck only
SKA1-LOW only
SKA1-LOW + Planck
SKA1-LOW + MID
+ Planck + Spectro-z
σ_8
M_ν
Frequency ν [MHz]
947 710 568 473 406 355 316
$\sigma_{f_{NL}}$
$\ell_{max}=60$
$\ell_{max}=300$
$b_{HI}(z)=1$
edshift z
$z=0.6$
$z=4$
k [Mpc^{-1}]
$P_{HI}(k)$
1σ - 1000 hour obs
2σ - 1000 hour obs
1σ - 3000 hour obs
2σ - 3000 hour obs
1σ - 5000 hour obs
2σ - 5000 hour obs
using $z=3$, 4 and 5
mass$_{DM}$ [keV]
$\Omega_{HI} \times 10^3$
H_0^{SH0ES}+SN+galBAO+Lyα For.BAO+r_s^{Planck}
H_0^{SH0ES}+SN+galBAO+Lyα For.BAO+IMBAO+r_s^{Planck}
$H(z)/H^{\Lambda CDM}(z)$
z
Bernal+ 2019

Contaminants are THE challenge to overcome with HI intensity mapping

Contaminants are THE challenge to overcome with HI intensity mapping



HI intensity mapping:

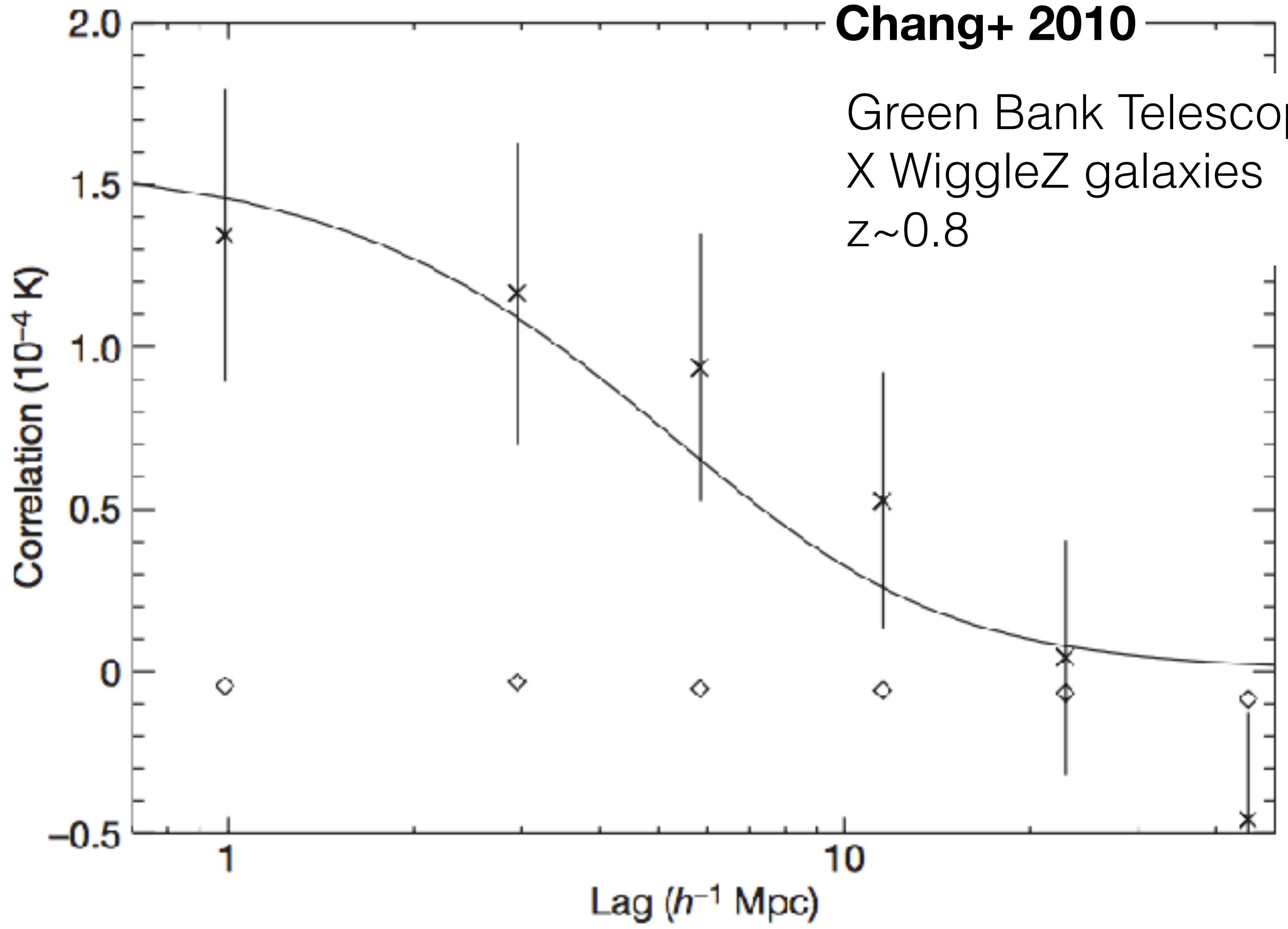
Observational status

HI intensity mapping

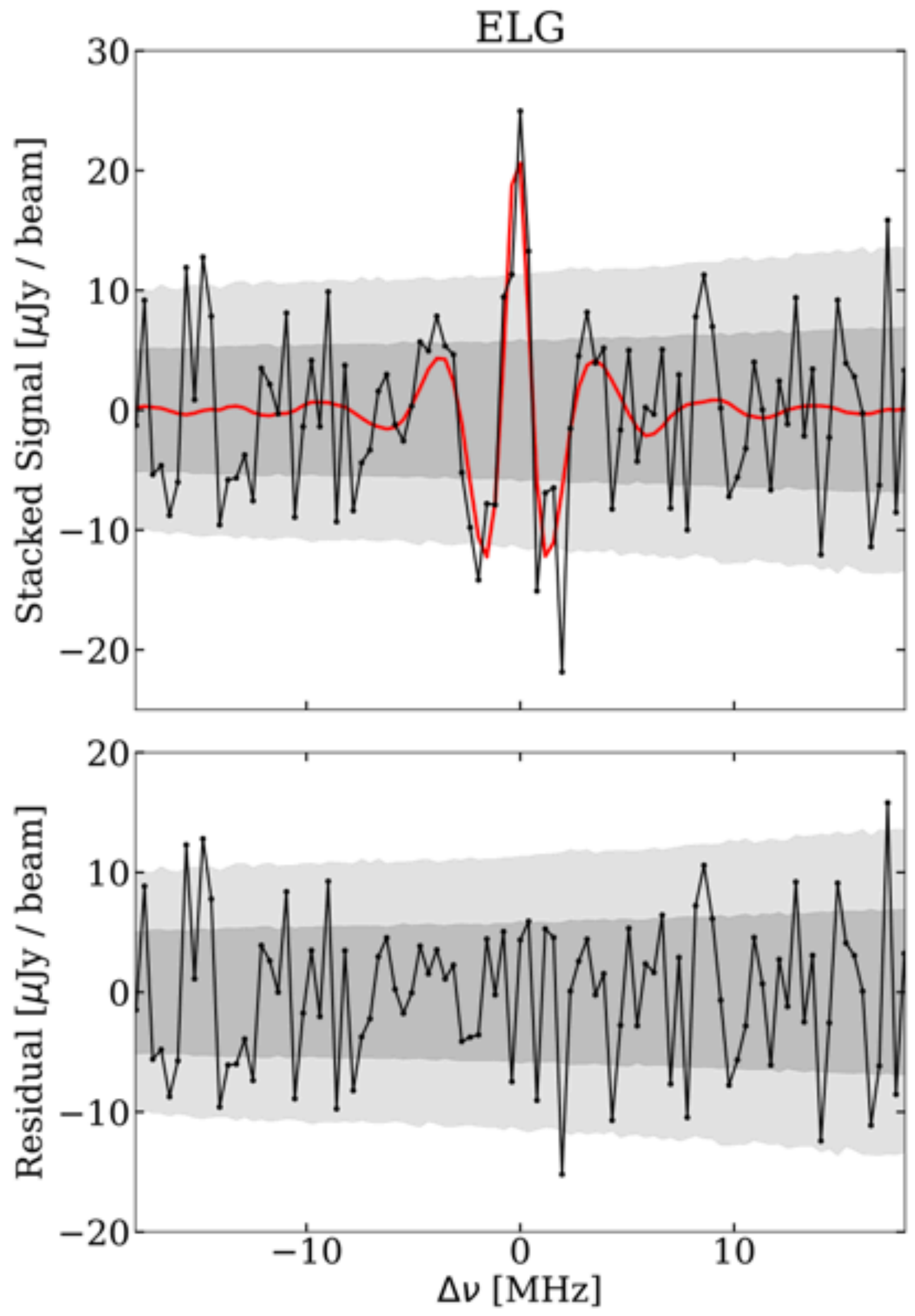
State-of-the-art

Cunnington, Li +
MeerKAT X WiggleZ galaxies
z~0.4

Chang+ 2010
Green Bank Telescope
X WiggleZ galaxies
z~0.8

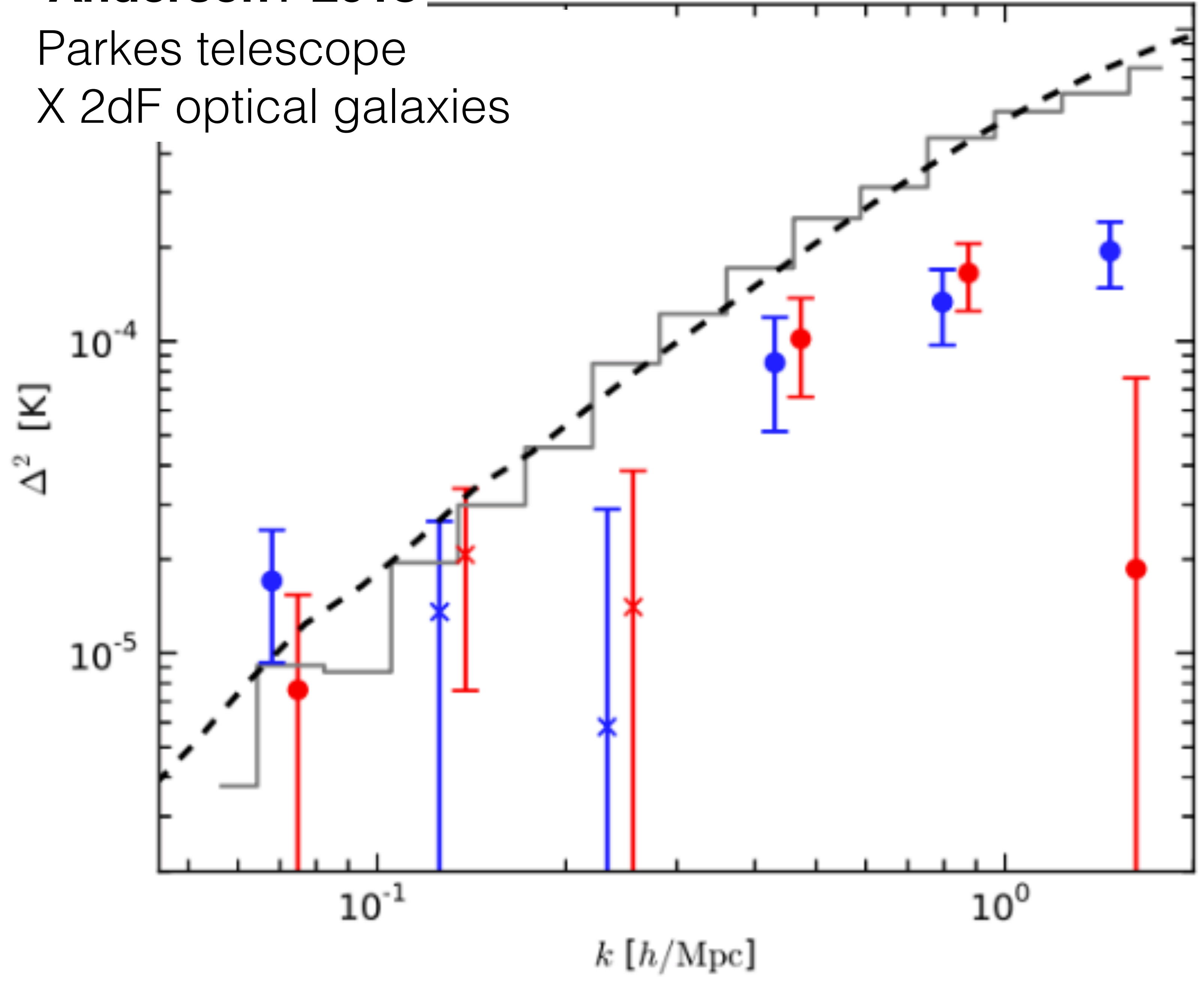


also Masui+ 2013, Switzer+ 2013,
Wolz+ 2017,2022



CHIME collaboration, 2022
stacking LRGs, ELG
and QSOs from eBOSS
0.8<z<1.5

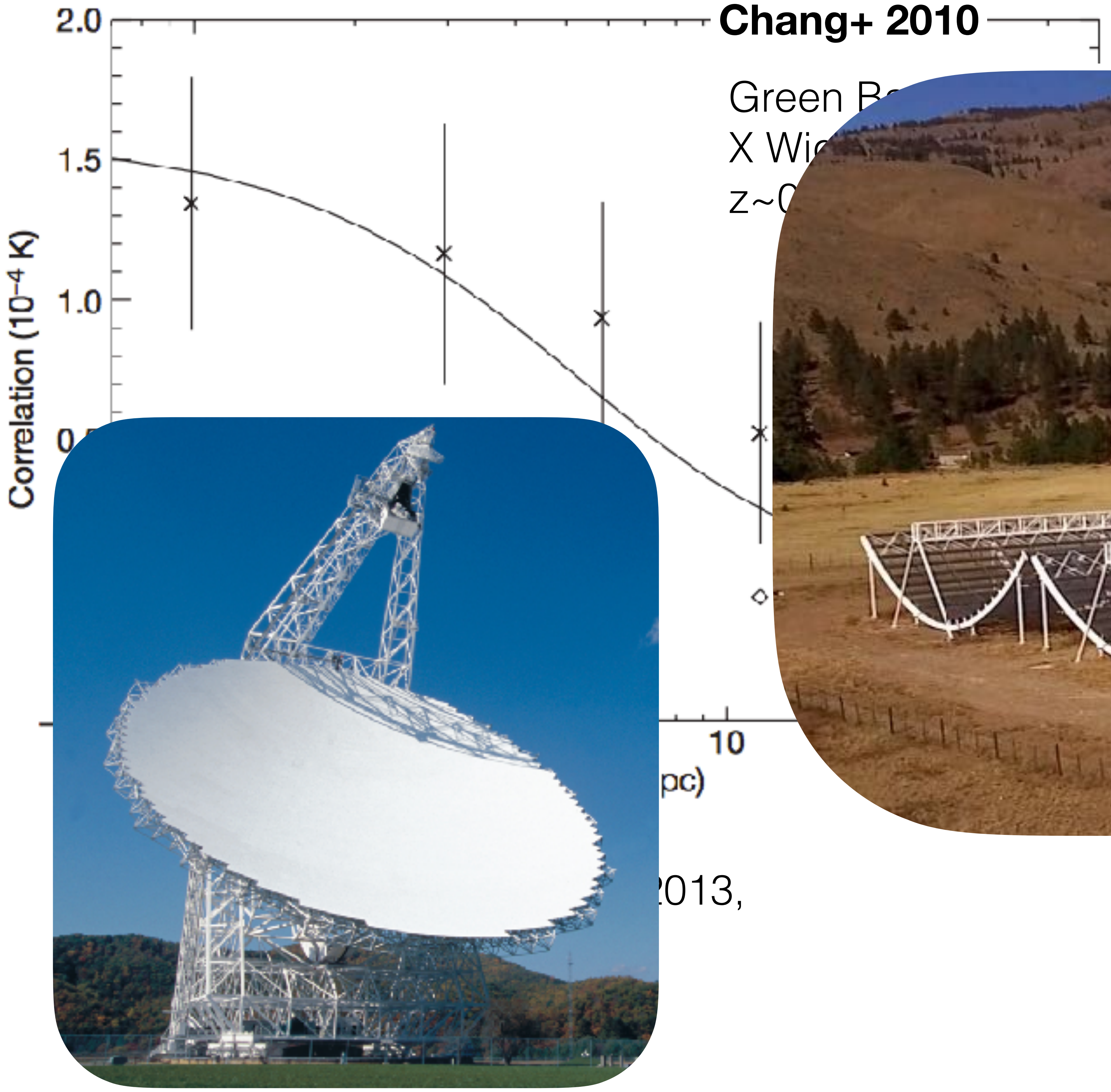
Anderson+ 2018
Parkes telescope
X 2dF optical galaxies

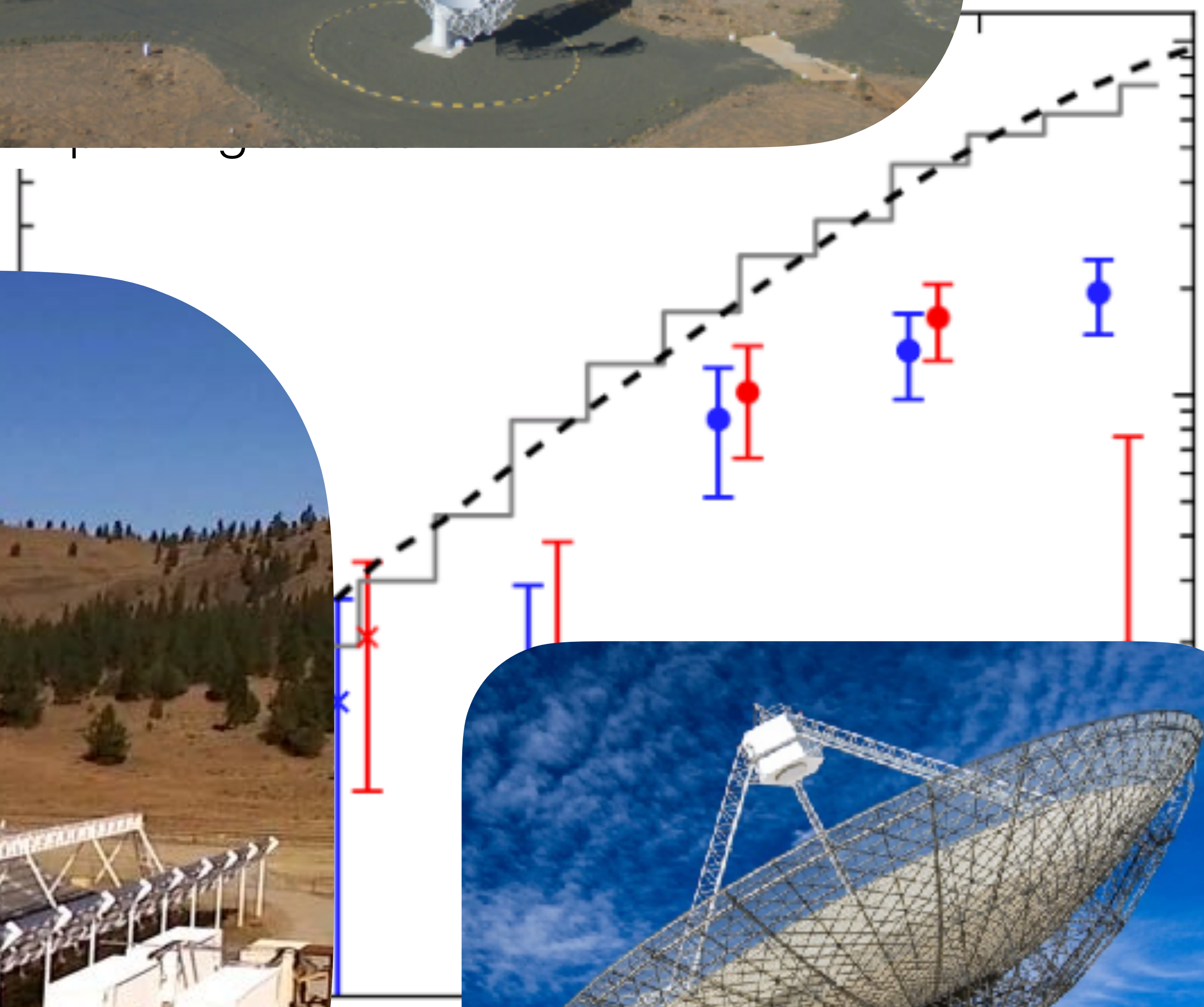


HI intensity mapping

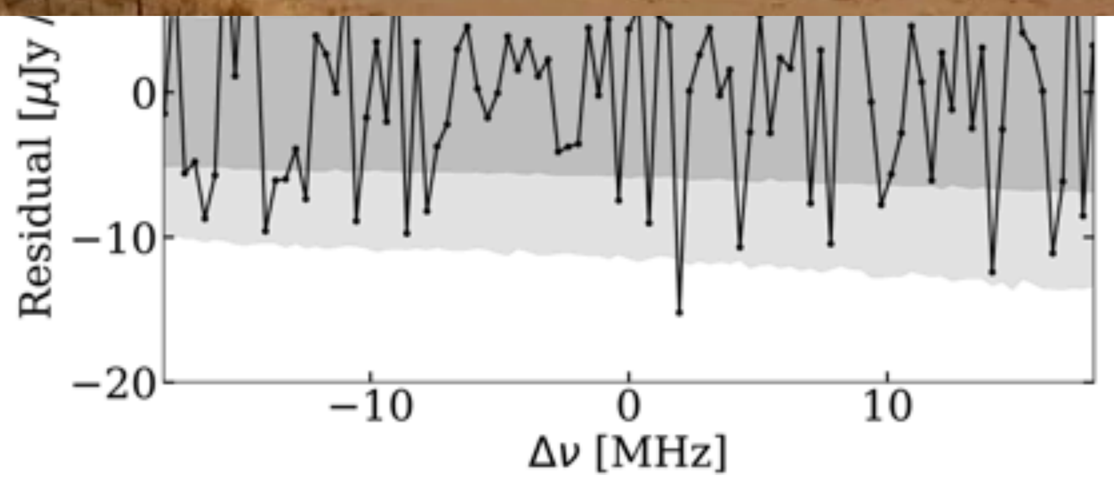
State-of-the-art

Cunnington, Li
MeerKAT X Wig
z~0.4





2013,



CHIME collaboration, 2022
stacking LRGs, ELG
and QSOs from eBOSS
0.8<z<1.5

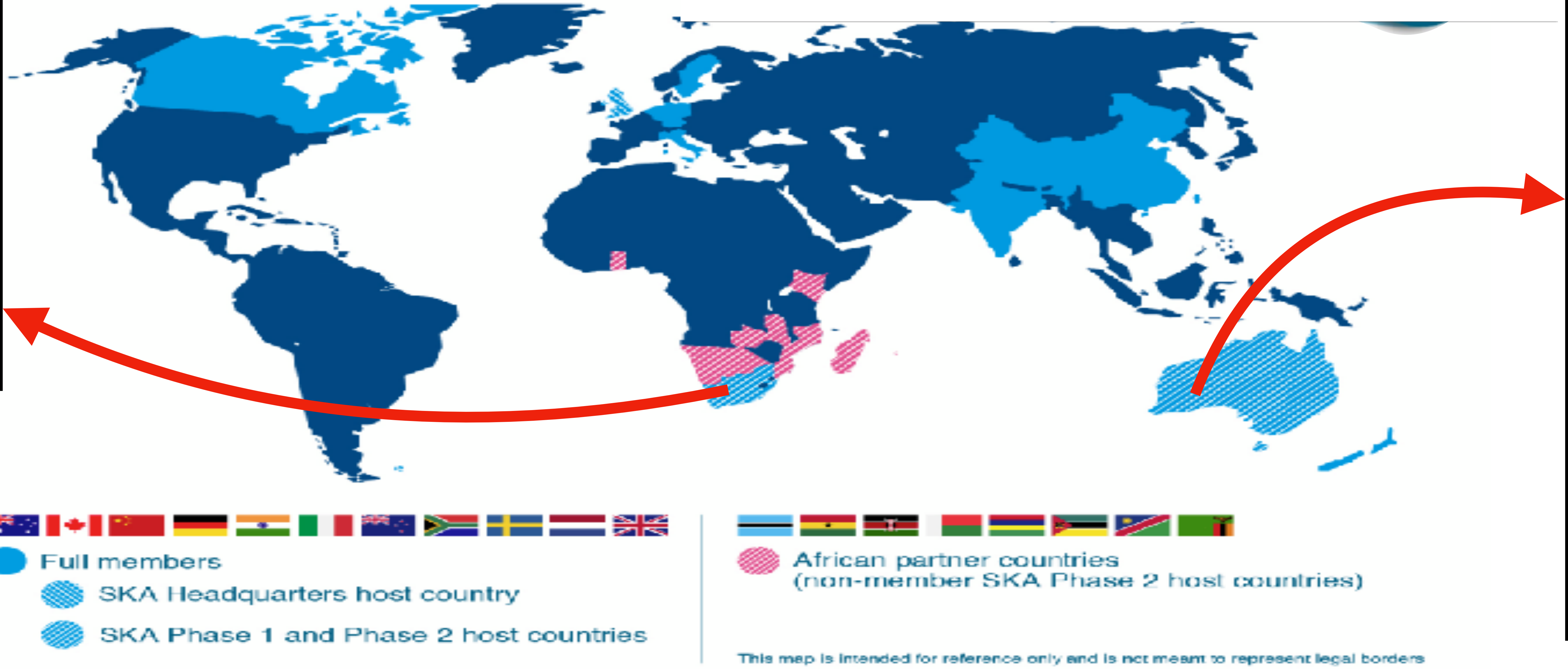
MeerKAT

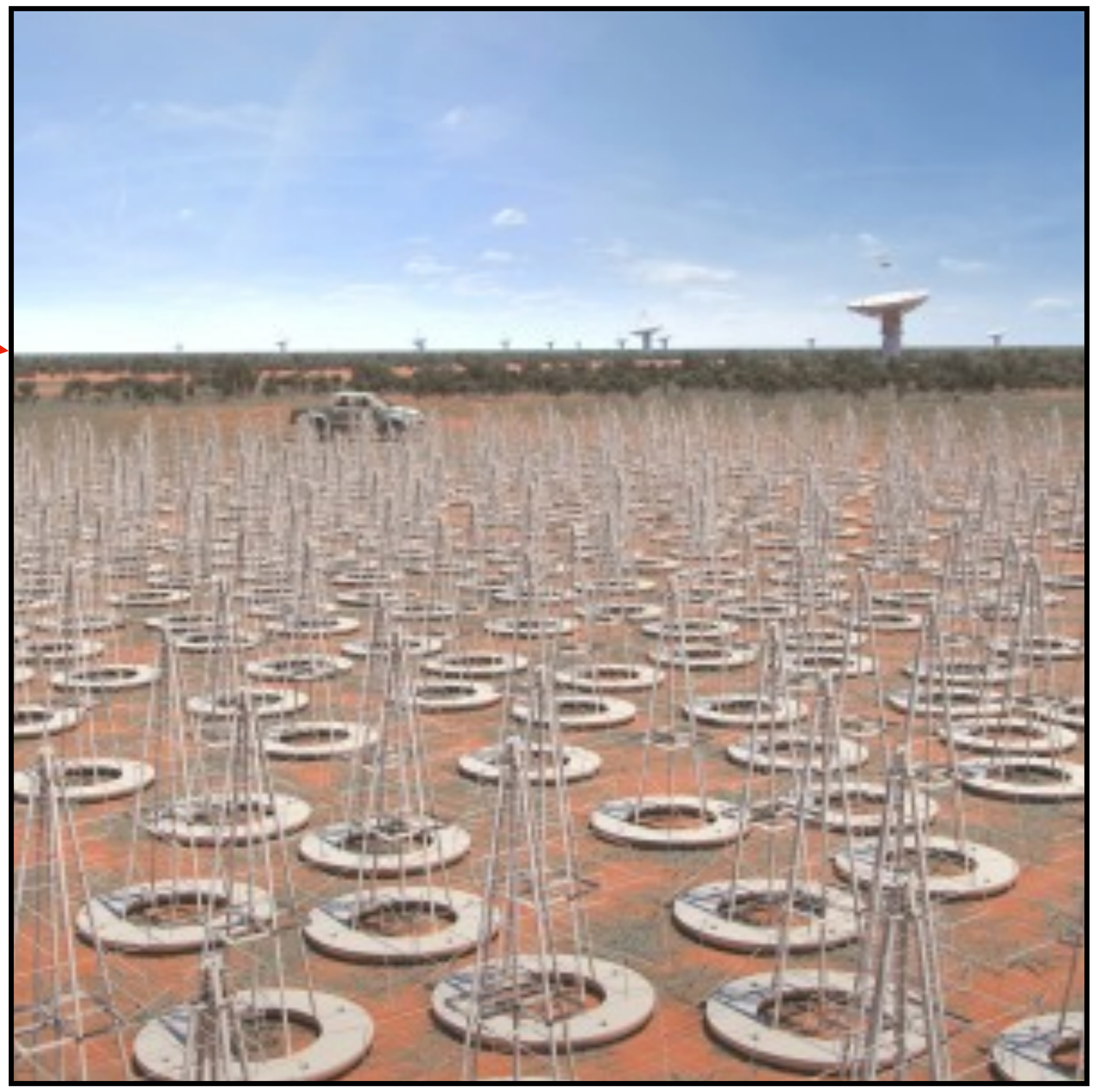
64+ dishes with single pixel feeds

Ongoing **MeerKLASS**: MeerKAT Large Area Synoptic Survey

(Wang+ 2021, Li+ 2021, Irfan+ 2022, Cunnington, Li +, 2206.01579)

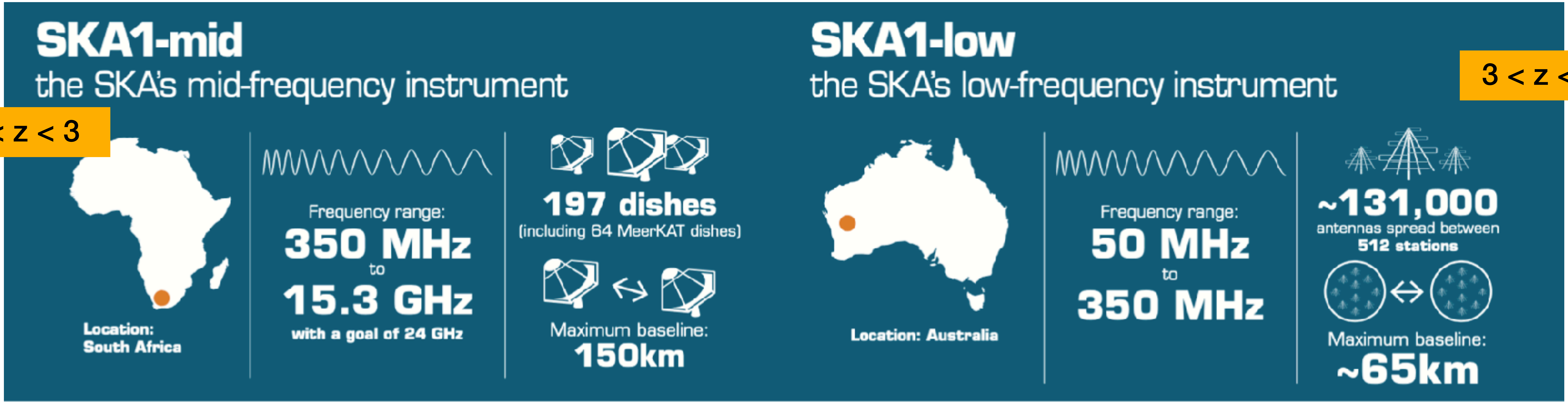
SKAO





$0 < z < 3$

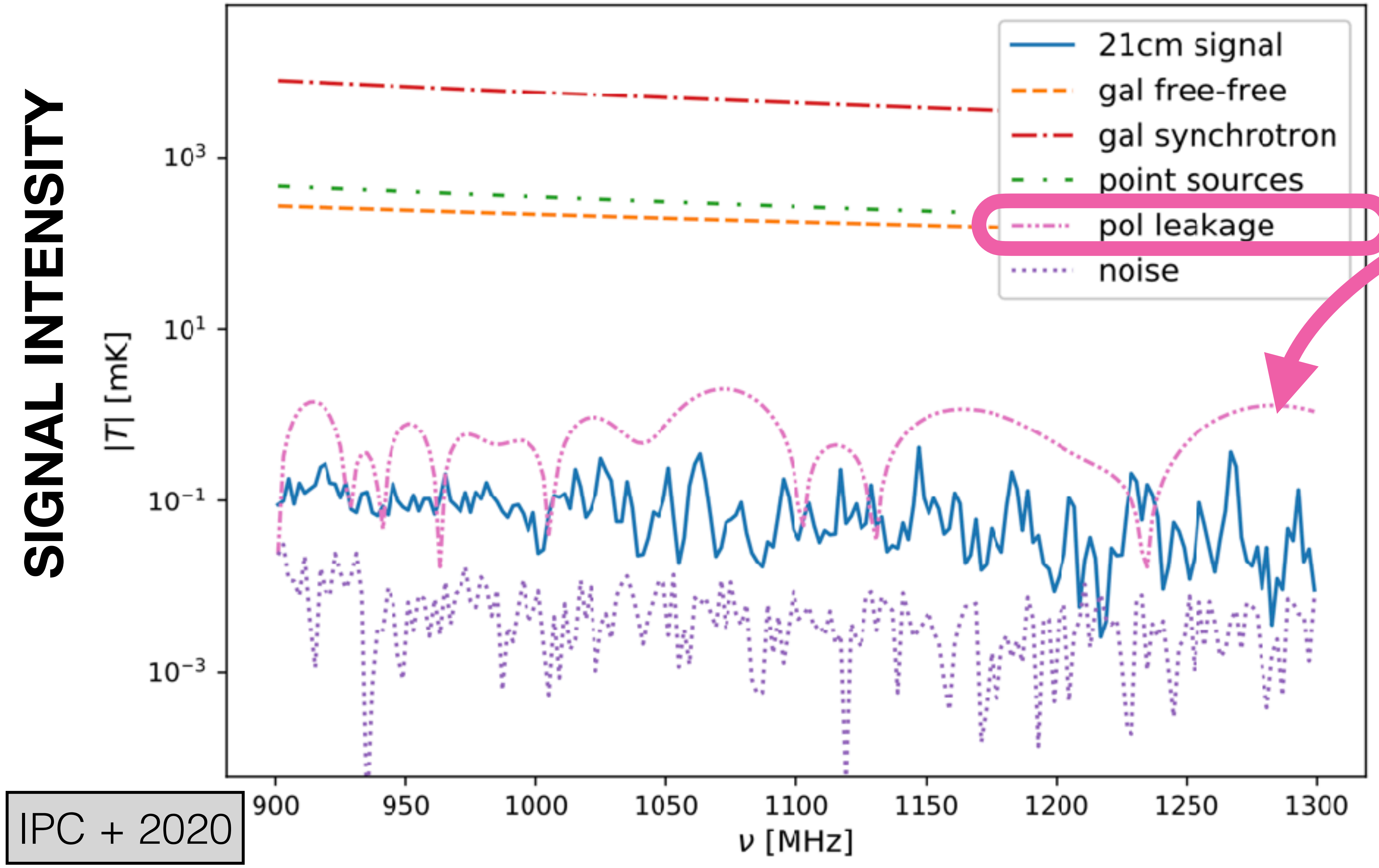
$3 < z < 27$

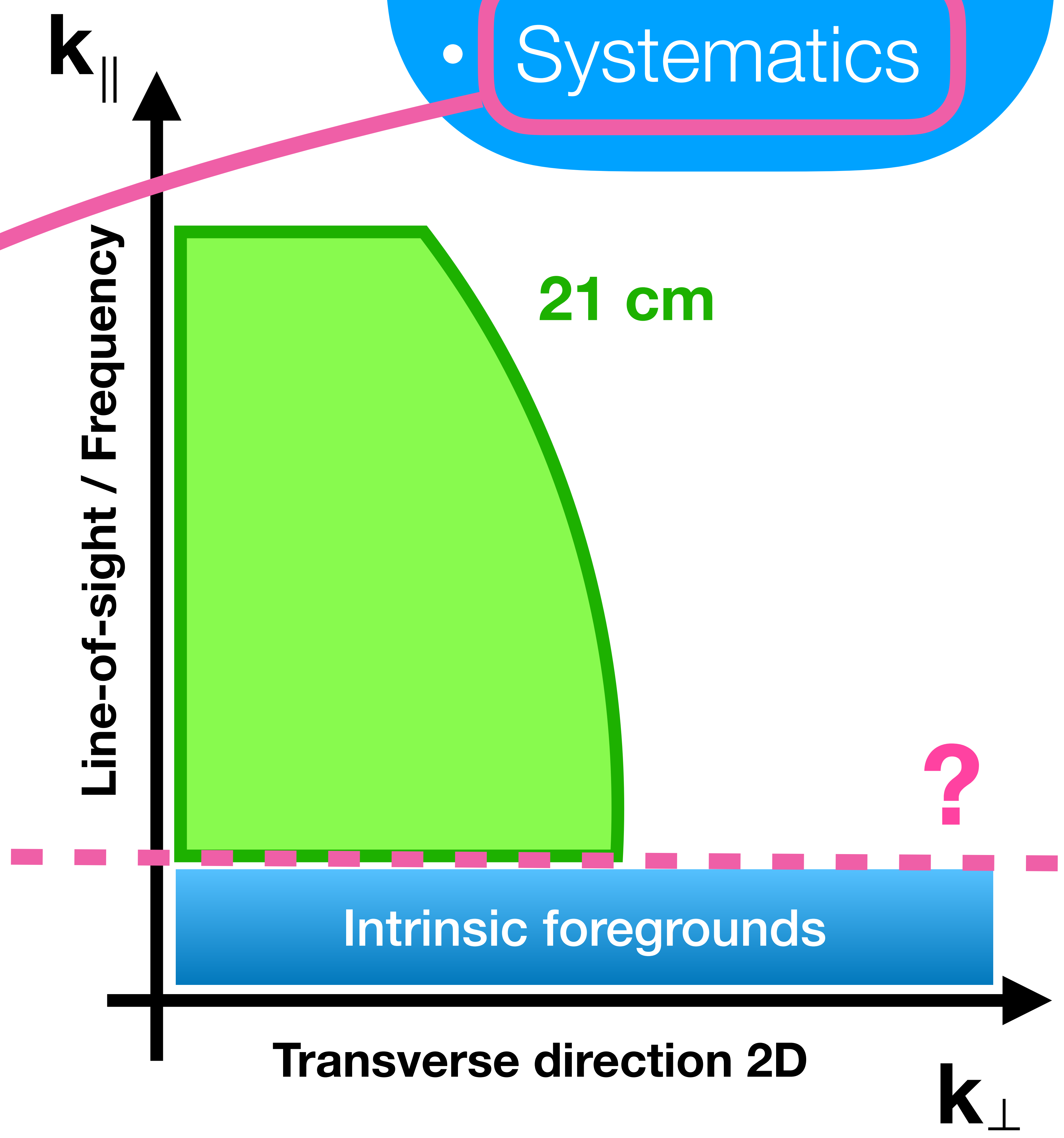


HI intensity mapping: buried under the foregrounds

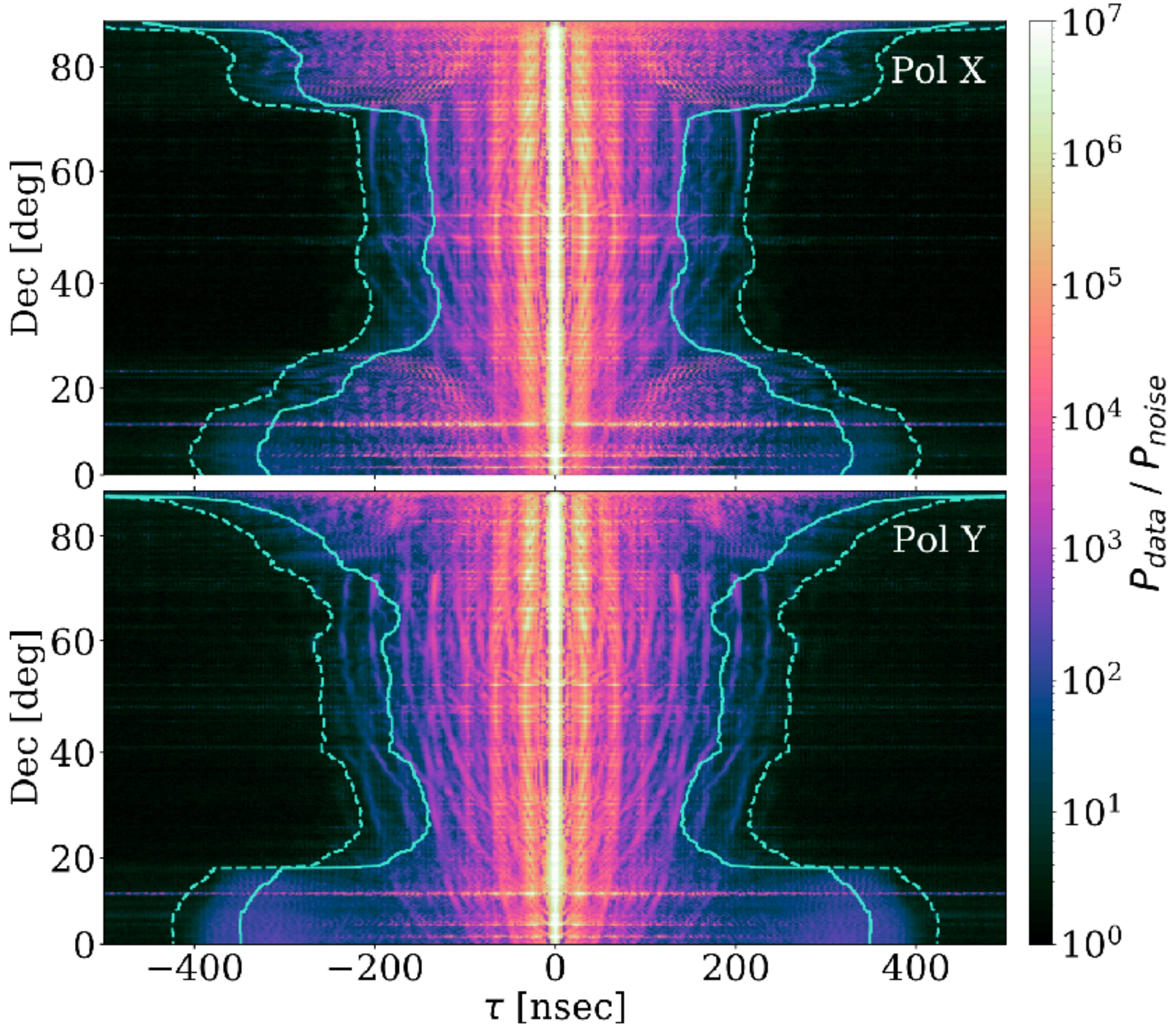
CHALLENGES:

- Foregrounds
- Systematics

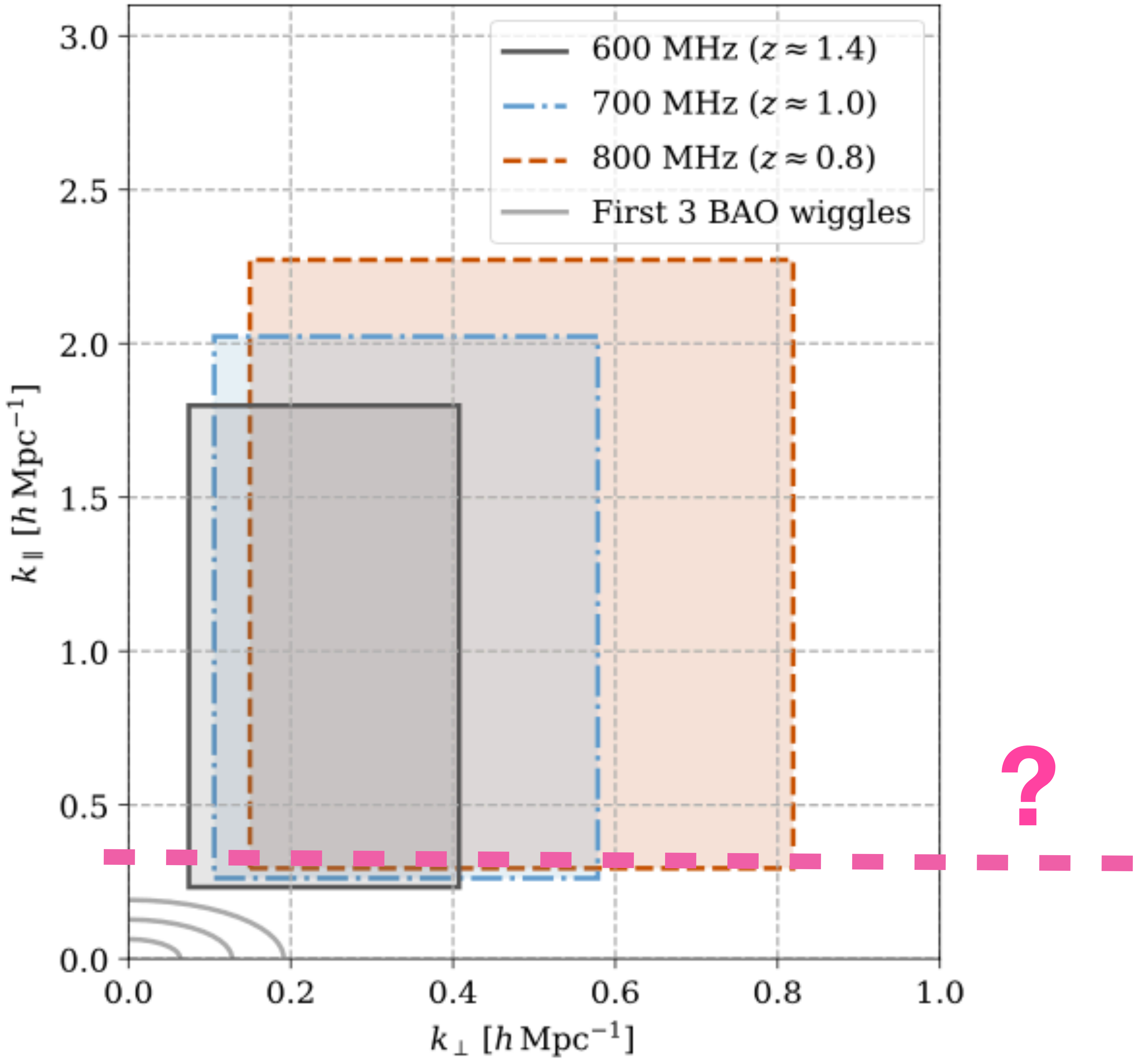




HI intensity mapping:
buried under the foregrounds



Chime Coll. 2022



Blind Source Separation algorithms

The separation of a set of source signals (contaminants) from a set of mixed signals (the maps), with little or no info about the source signal or the mixing process.

Need to set number *n* of sources!

mixing matrix (f,n)

HI signal!

$$\mathbf{X} = \mathbf{A}\ \mathbf{S} + \mathbf{N}$$

signal (f,p)

sources (n,p)

- **Decorrelation —>** diagonalise the covariance matrix
- **Independence —>** as more independent sources are mixed the signal becomes more Gaussian (central limit theorem). So, let's maximise the non-gaussianity of the sources to *unmix* them.

Principal Component Analysis (**PCA**)

Independent Component Analysis (**ICA**)

HI intensity mapping:
how to subtract the contaminants?

Polynomial fitting

1. **Chang+ 2010**
Green Bank Telescope
X WiggleZ galaxies
z~0.8

(also Masui+ 2013,
Wolz+ 2017,2022)

10 - 20 **20 - 36**

Independent
Component Analysis
(**ICA**)

2. **Anderson+ 2018** **10**
Parkes telescope
X 2dF optical galaxies

Principal Component
Analysis (**PCA**)

"Instrumental effects such as passband calibration and **polarization leakage** couple bright foregrounds into new degrees of freedom […]. The spectral functions describing these systematics cannot all be modelled in advance, so we take an **empirical approach to foreground removal by estimating dominant modes** from the covariance of the map itself."

Switzer+ 2013

In all theoretical works:

- no noteworthy difference between PCA or ICA
- ~**4** components removed are enough

(e.g., Wolz+ 2014, Alonso+ 2015, Cunnington+ 2019)

HI intensity mapping:
how to subtract the contaminants?

We need:

1. simulations as realistic as possible

Harper+ 2018, Li+ 2020, Matshawule+ 2021, …

2. new BSS algorithms optimised for HI IM

GMCA (sparsity-based) —> **mixGMCA**

(Carucci+ 2020, Cunnington+ 2021, The SKAO Blind Challenge , work in progress…)

a quick interlude on GMCA

Blind Source Separation algorithms

The separation of a set of source signals (contaminants) from a set of mixed signals (the maps), with little or no info about the source signal or the mixing process.

mixing
matrix (f,n)

HI signal!

$$\mathbf{X} = \mathbf{A}\ \mathbf{S} + \mathbf{N}$$

signal
(f,p)

sources
(n,p)

- **Decorrelation —>** diagonalise the covariance matrix
- **Independence —>** as more independent sources are mixed the signal becomes more Gaussian (central limit theorem). So, let's maximise the non-gaussianity of the sources to *unmix* them.
- **Sparsity —>** mixtures are less sparse than sources!

Principal Component Analysis (**PCA**)

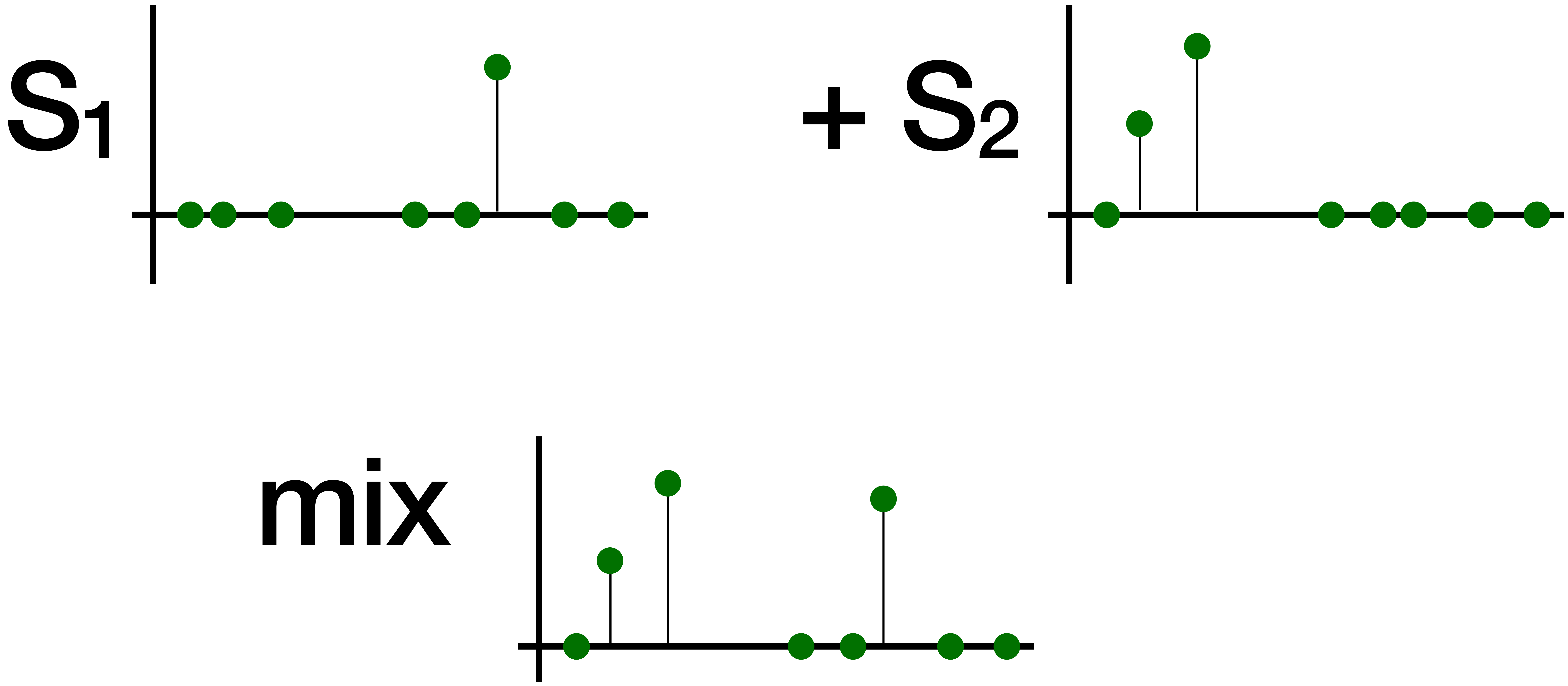
Independent Component Analysis (**ICA**)

Generalised Morphological Component Analysis (**GMCA**)

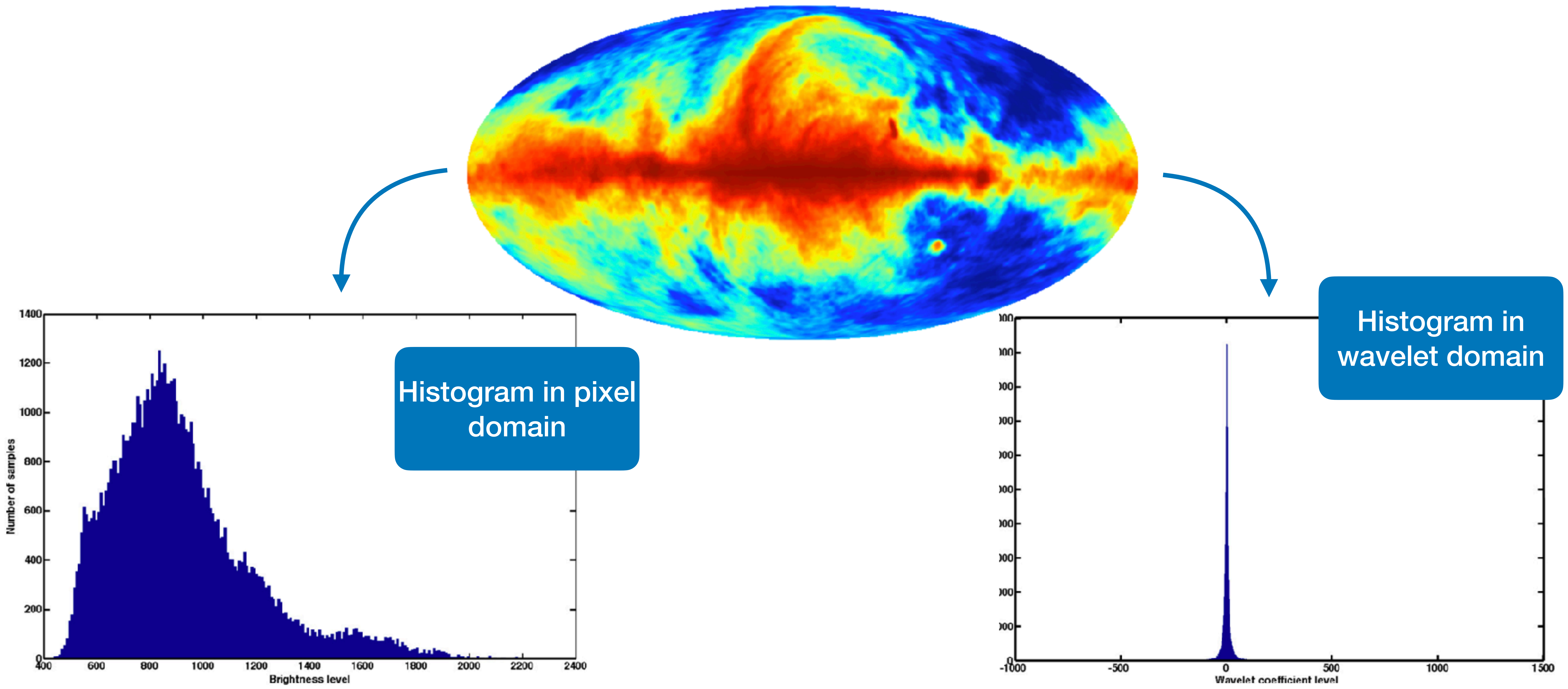
Bobin + 2007, 2008, 2012

why sparsity?

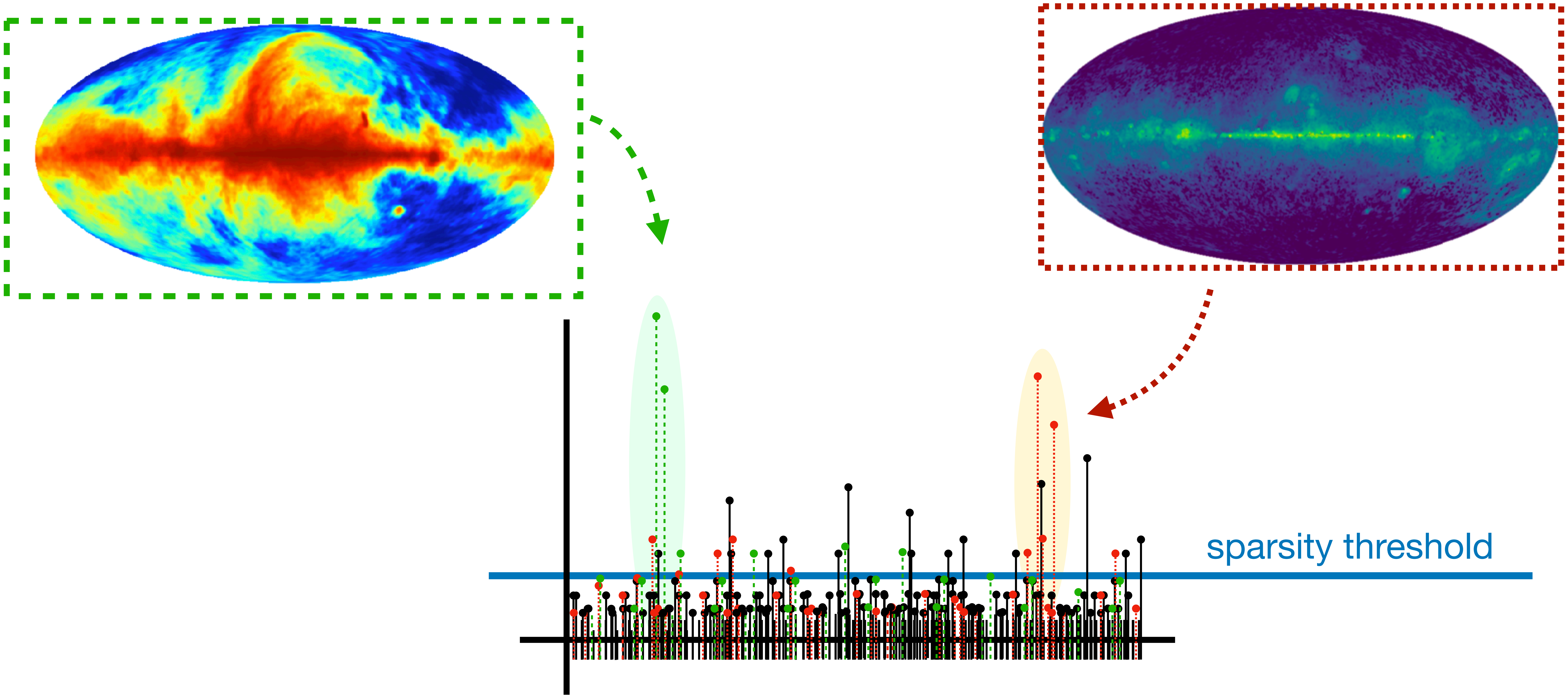
mixtures are less sparse than sources



Enforcing sparsity: in which domain?



Morphological diversity:
more contrast among components



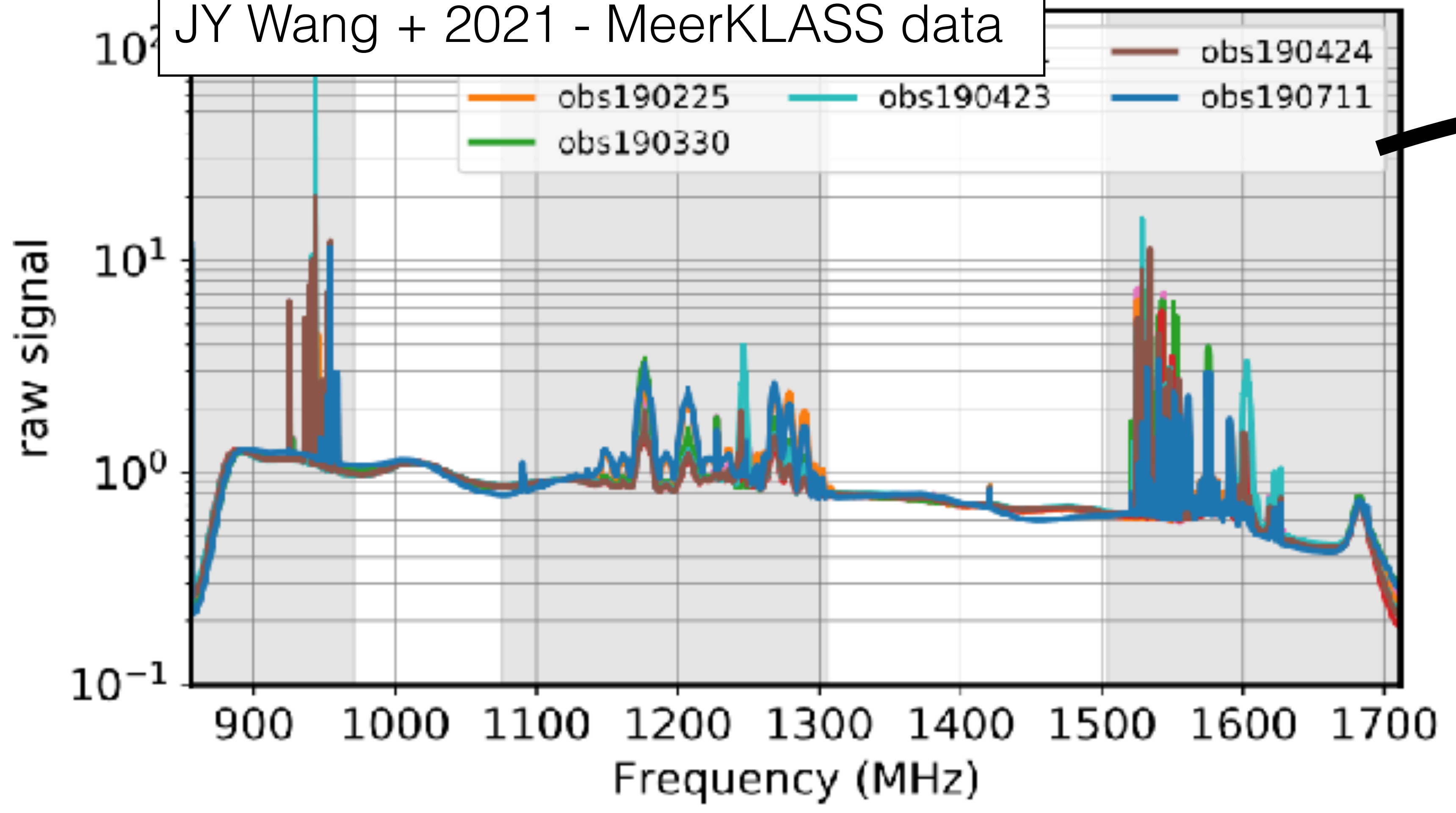
Sparsity-based component-separation for 21-cm IM

GMCA: Generalised Morphological Component Analysis

Bobin+ 2007, 2008, 2012,.. Applied on data in different astro-context: CMB (e.g. Bobin+2016), EoR (e.g. Hothi+2020), X-ray (Picquenot+2019), …

- wavelet decomposition —> **multi-scale** approach
- **No priors on signal**

JY Wang + 2021 - MeerKLASS data

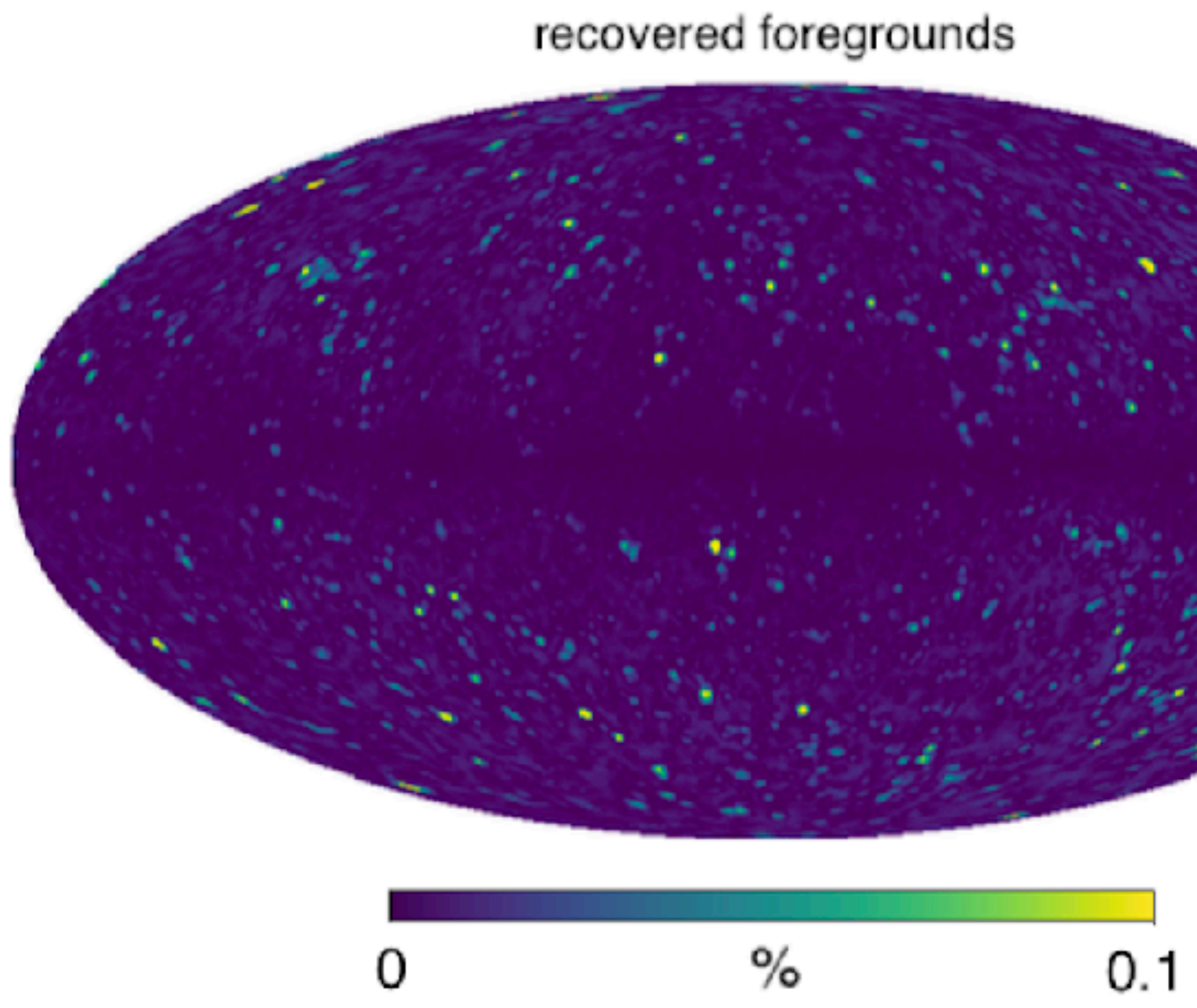


in Carucci+ 2020, for the fist time in the literature:

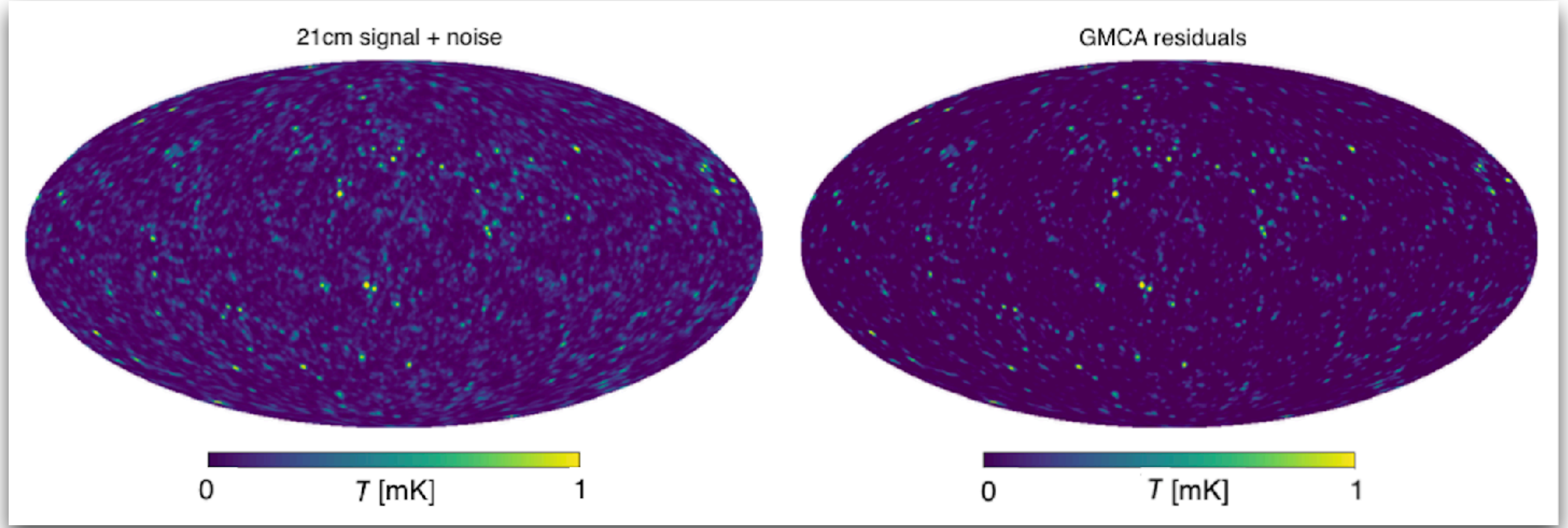
1. Good performance also with **RFI-flagged** data cubes!
   (TV stations, telecommunication, satellites,..)
2. **Pol leakage:** greater complexity of data
   (higher number of sources needed, convergence not assured, mode-mixing assured)

To reproduce these results: codes and sims available online

Sparsity-based component-separation for 21-cm IM

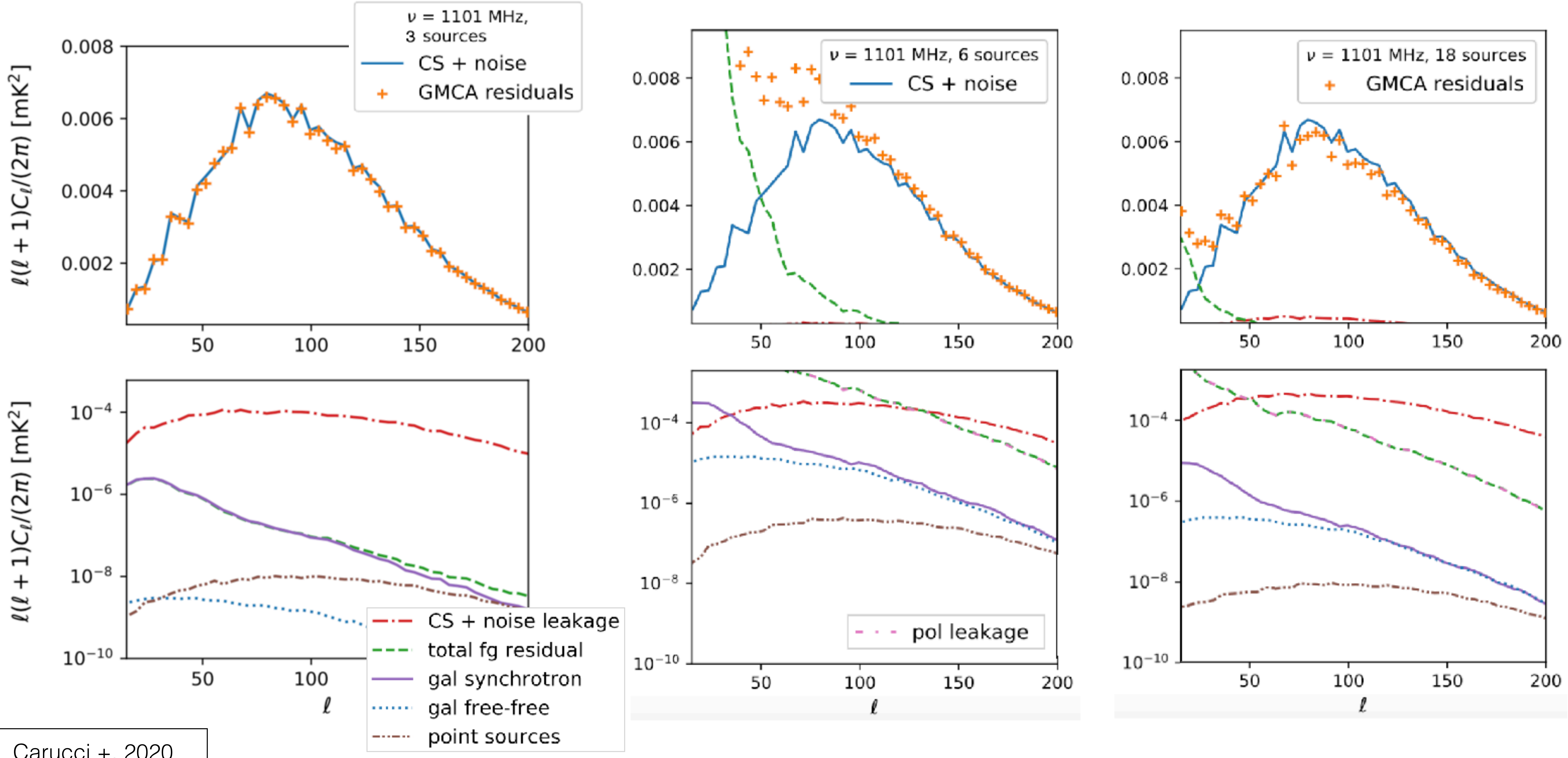


Sparsity-based component-separation for 21-cm IM



- Underestimate by <2% (channel average) the angular PS
- Reproduce at sub percent level the radial PS for $k_{\parallel} > 0.02$ h Mpc^{-1}

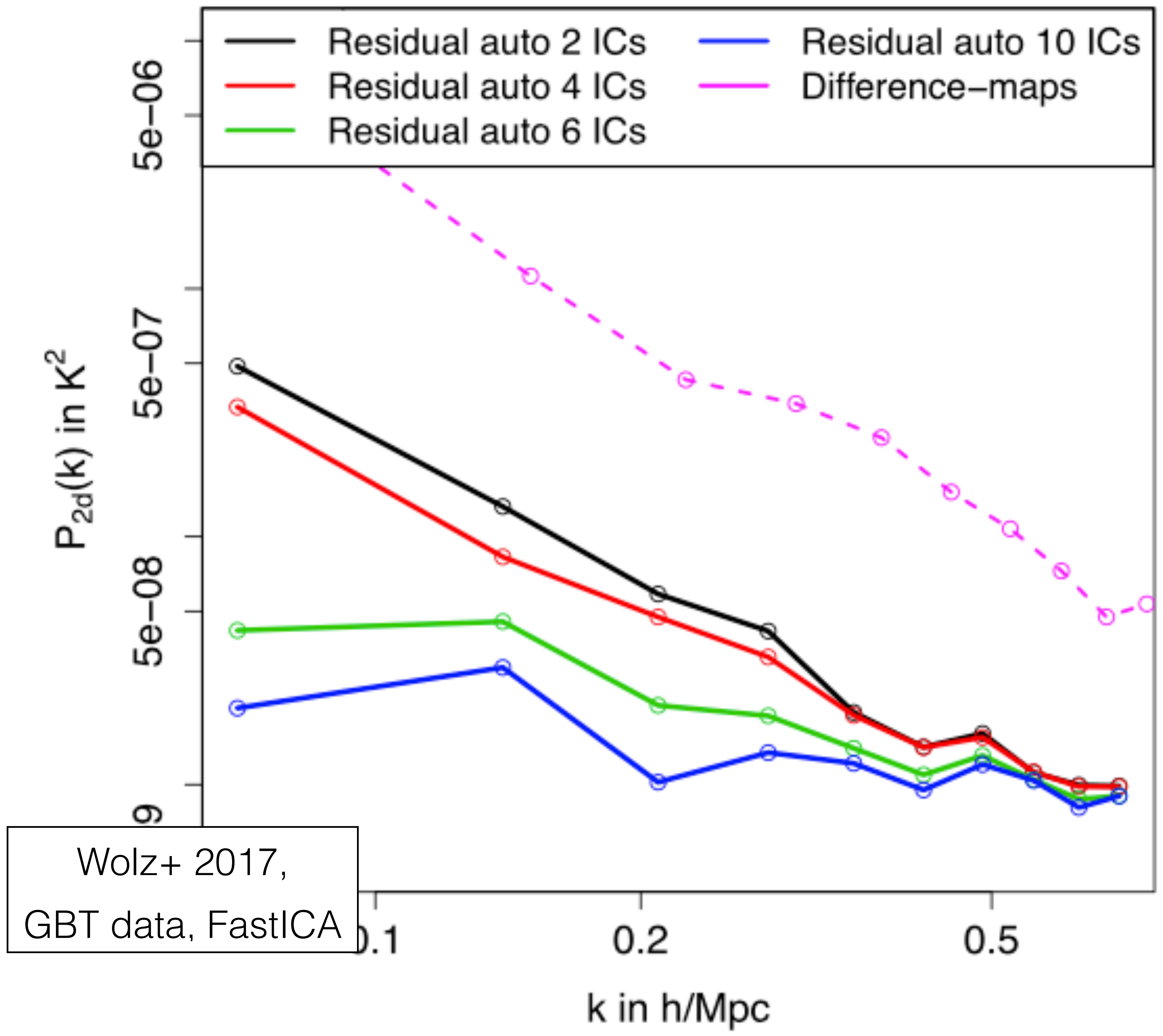
Different scales need different care



Carucci +, 2020

Different scales need different care

The wavelet domain is a multi-scale framework!



Wolz+ 2017, GBT data, FastICA

- GMCA performs very well on small scales, can fail at the large scale
- PCA / ICA —> overfit the large scales

PCA on the large scale
+
GMCA on the small scales

mixGMCA

See also Hothi+2020 with LOFAR data

HI intensity mapping:
how to subtract the contaminants?

We need:

1. simulations as realistic as possible

Harper+ 2018, Spinelli+ 2020, Matshawule+ 2021

2. new BSS algorithms optimised for HI IM

GMCA (sparsity-based) —> **mixGMCA**

(Carucci+ 2020, Cunnington+ 2021, The SKAO Blind Challenge , work in progress…)

3. to test the BSS pipelines on the same set of sims

Started at the 2020 SKA Cosmology SWG meeting, as a collective project of the IM Focus Group

SKAO HI Intensity Mapping: Blind Foreground Subtraction Challenge

Marta Spinelli,[1,2,3★] Isabella P. Carucci,[4,5,6†] Steven Cunnington,[7] Stuart E. Harper,[8] Melis O. Irfan,[3,7]
José Fonseca,[7,3,9,10] Alkistis Pourtsidou,[7,3] Laura Wolz[8]

ABSTRACT

Neutral Hydrogen Intensity Mapping (HI IM) surveys will be a powerful new probe of cosmology. However, strong astrophysical foregrounds contaminate the signal and their coupling with instrumental systematics further increases the data cleaning complexity. In this work, we simulate a realistic single-dish HI IM survey of a 5000 deg^2 patch in the 950 – 1400 MHz range, with both the MID telescope of the SKA Observatory (SKAO) and MeerKAT, its precursor. We include a state-of-the-art HI simulations and explore different foreground models and instrumental effects such as non-homogeneous thermal noise and beam side-lobes. We perform the first Blind Foreground Subtraction Challenge for HI IM on these synthetic data-cubes, aiming to characterise the performance of available foreground cleaning methods with no prior knowledge of the sky components and noise level. Nine foreground cleaning pipelines joined the Challenge, based on statistical source separation algorithms, blind polynomial fitting, and an astrophysical-informed parametric fit to foregrounds. We devise metrics to compare the pipeline performances quantitatively. In general, they can recover the input maps' 2-point statistics within 20 per cent in the range of scales least affected by the telescope beam. However, spurious artefacts appear in the cleaned maps due to interactions between the foreground structure and the beam side-lobes. We conclude that it is fundamental to develop accurate beam deconvolution algorithms and test data post-processing steps carefully before cleaning. This study was performed as part of SKAO preparatory work by the HI IM Focus Group of the SKA Cosmology Science Working Group.

if we were given SKA1-mid IM data today,
what could we achieve in terms of
contaminants subtraction?

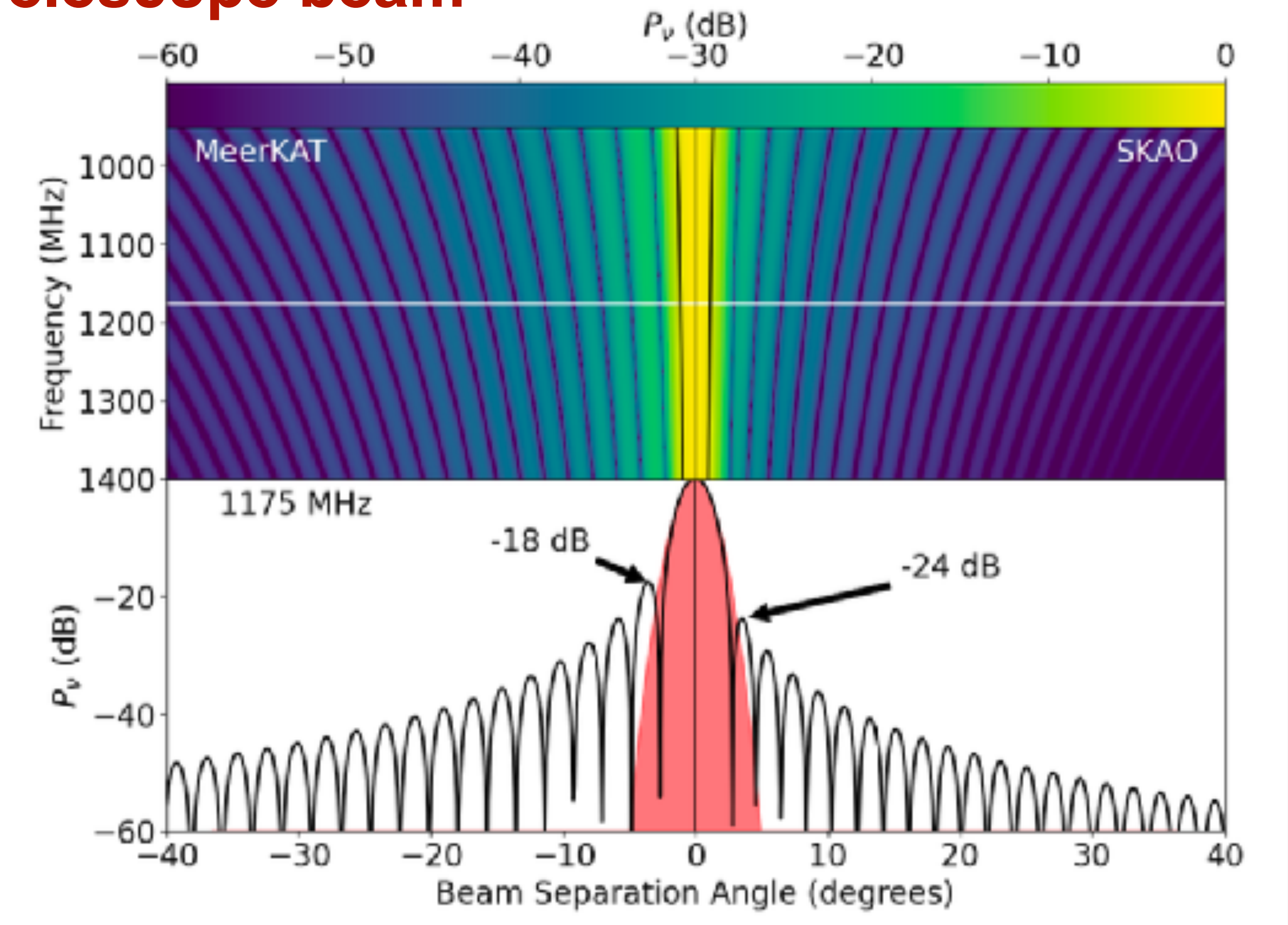
Simulating all we can (up to now)

Sky components:

1. HI
   (SAM coupled to halo catalogue/merger tree)
2. Astrophysical Foregrounds
   - Galactic synchrotron
   - Galactic Free-Free
   - Extragalactic background
   - Point Sources

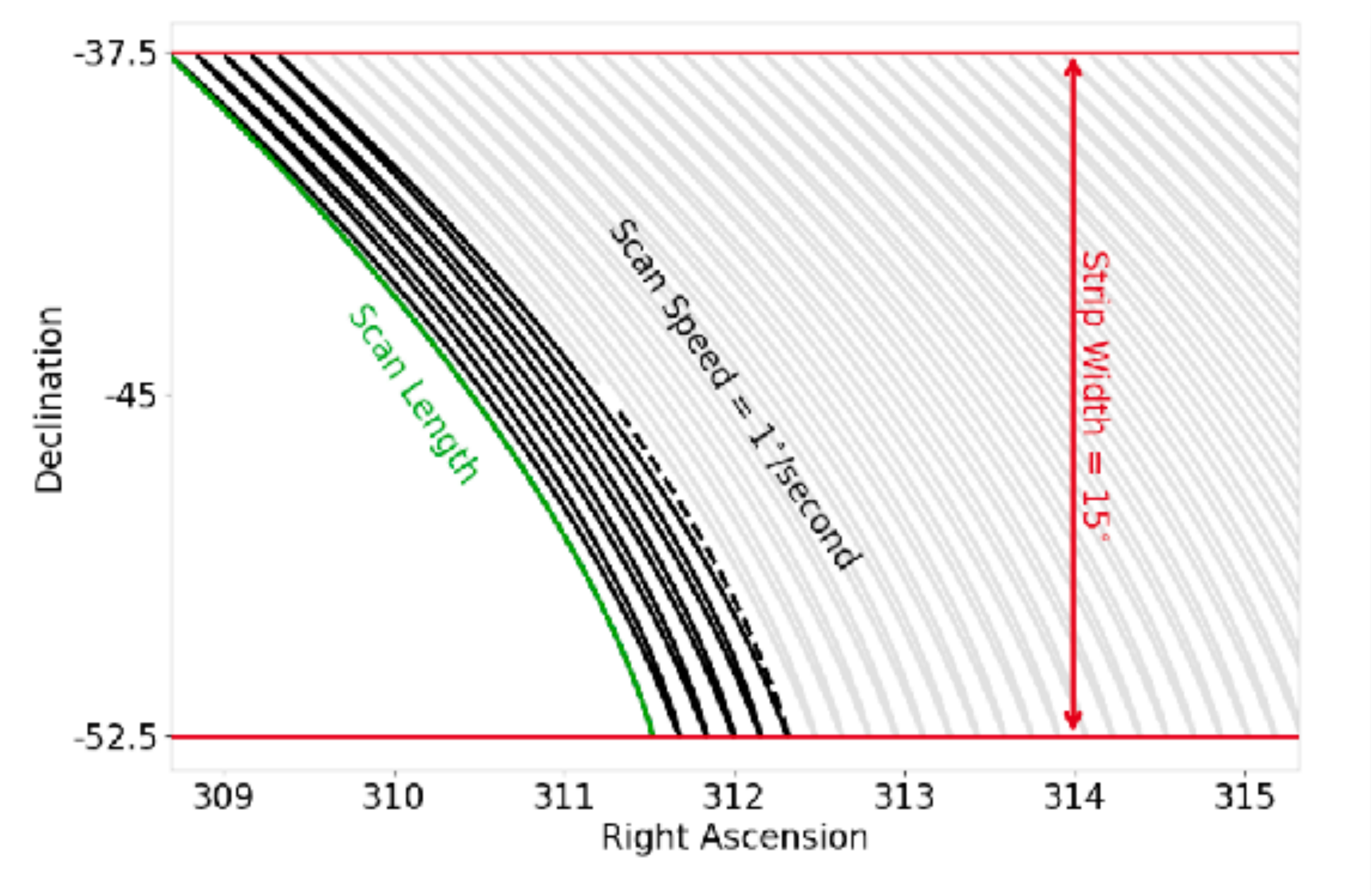
+

Telescope beam



+

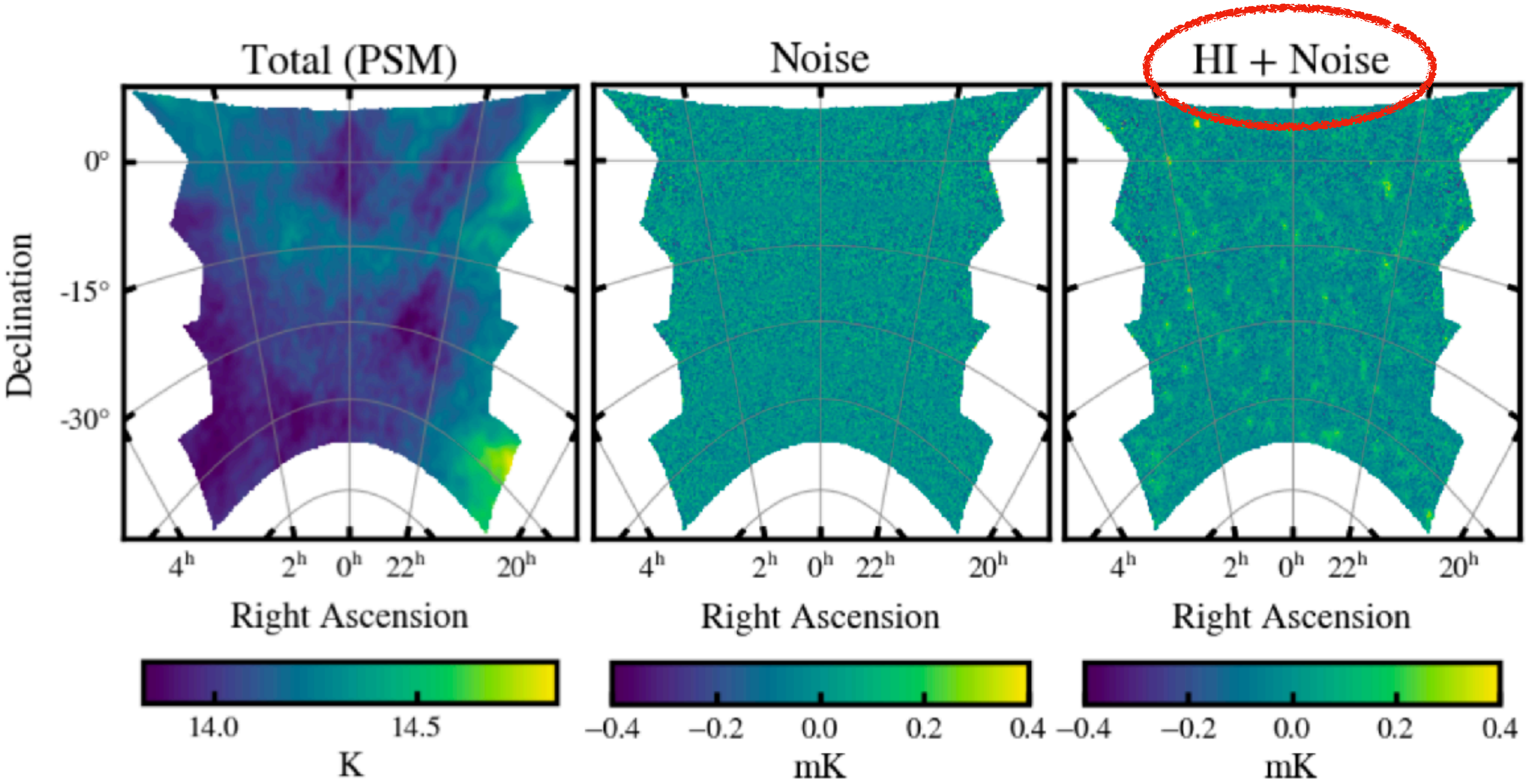
Scanning strategy (non-uniform noise)

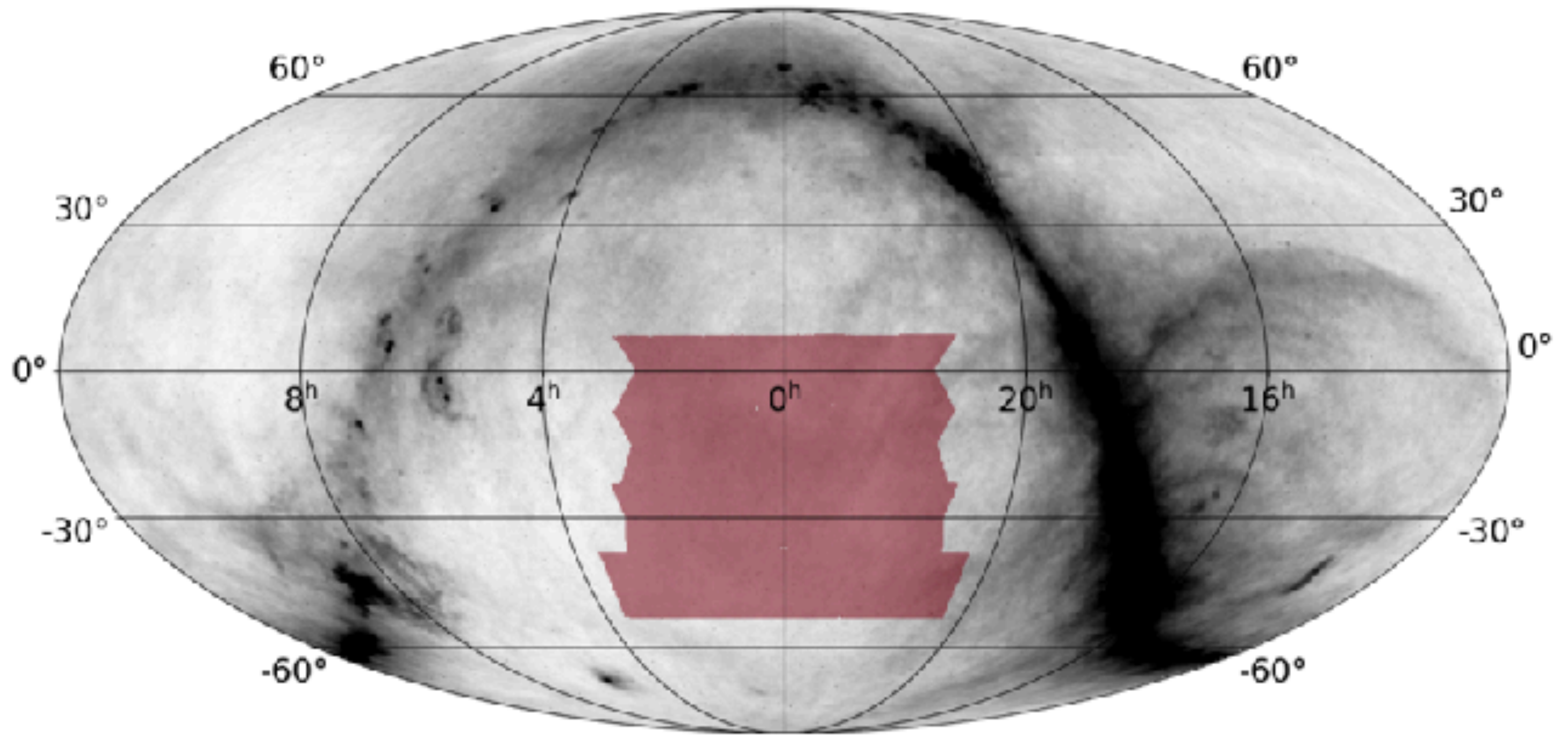


2 FGs models x 2 Beam Models
x 2 Instruments x 2 Deconvolution strategies
= 16 data cubes to clean

Simulating all we can (up to now)

Unknown to participants!





L-band: 950-1400 MHz

Single-dish mode

N_{dish}: 133 (SKAO) and 64 (MeerKAT)

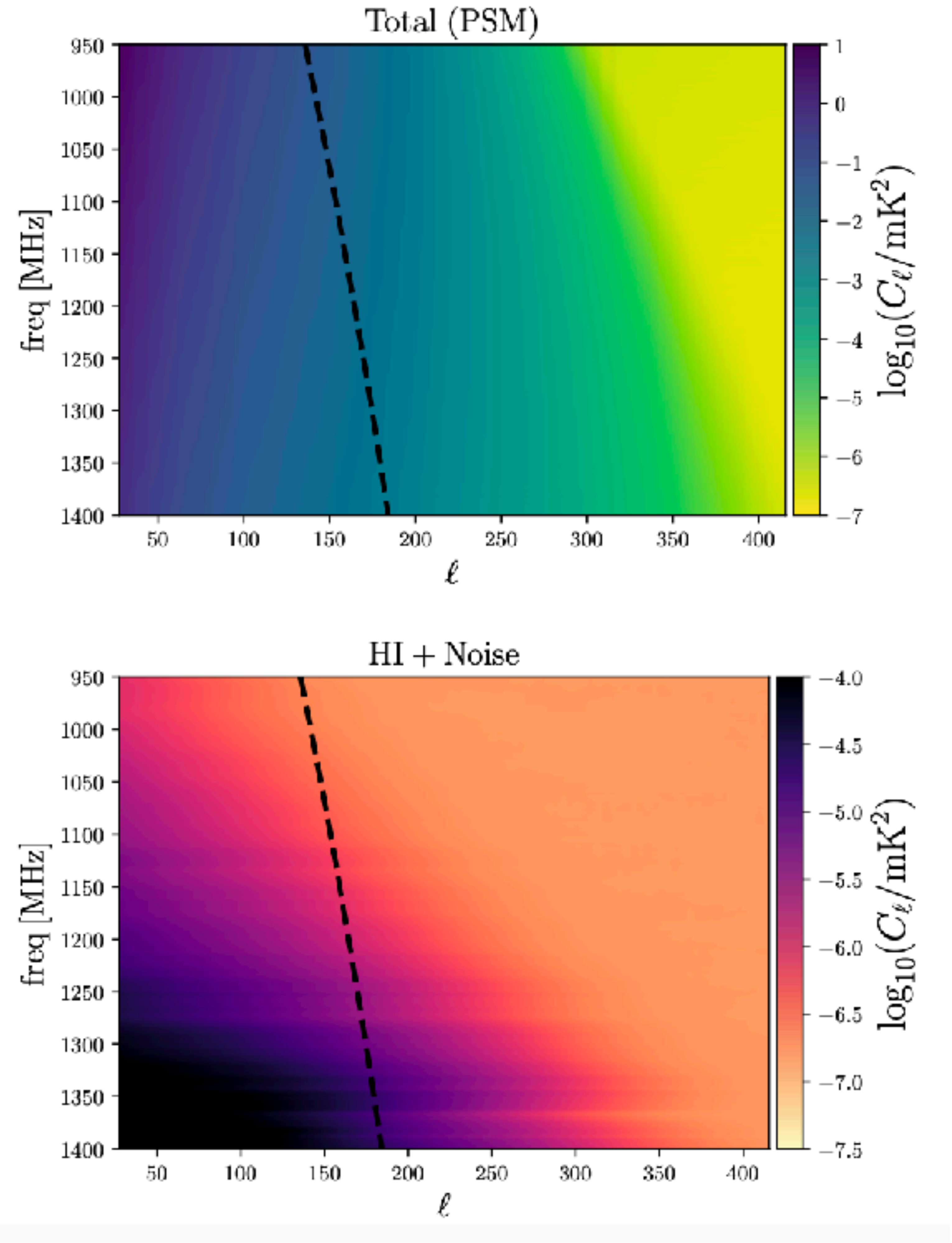
x 512 channels

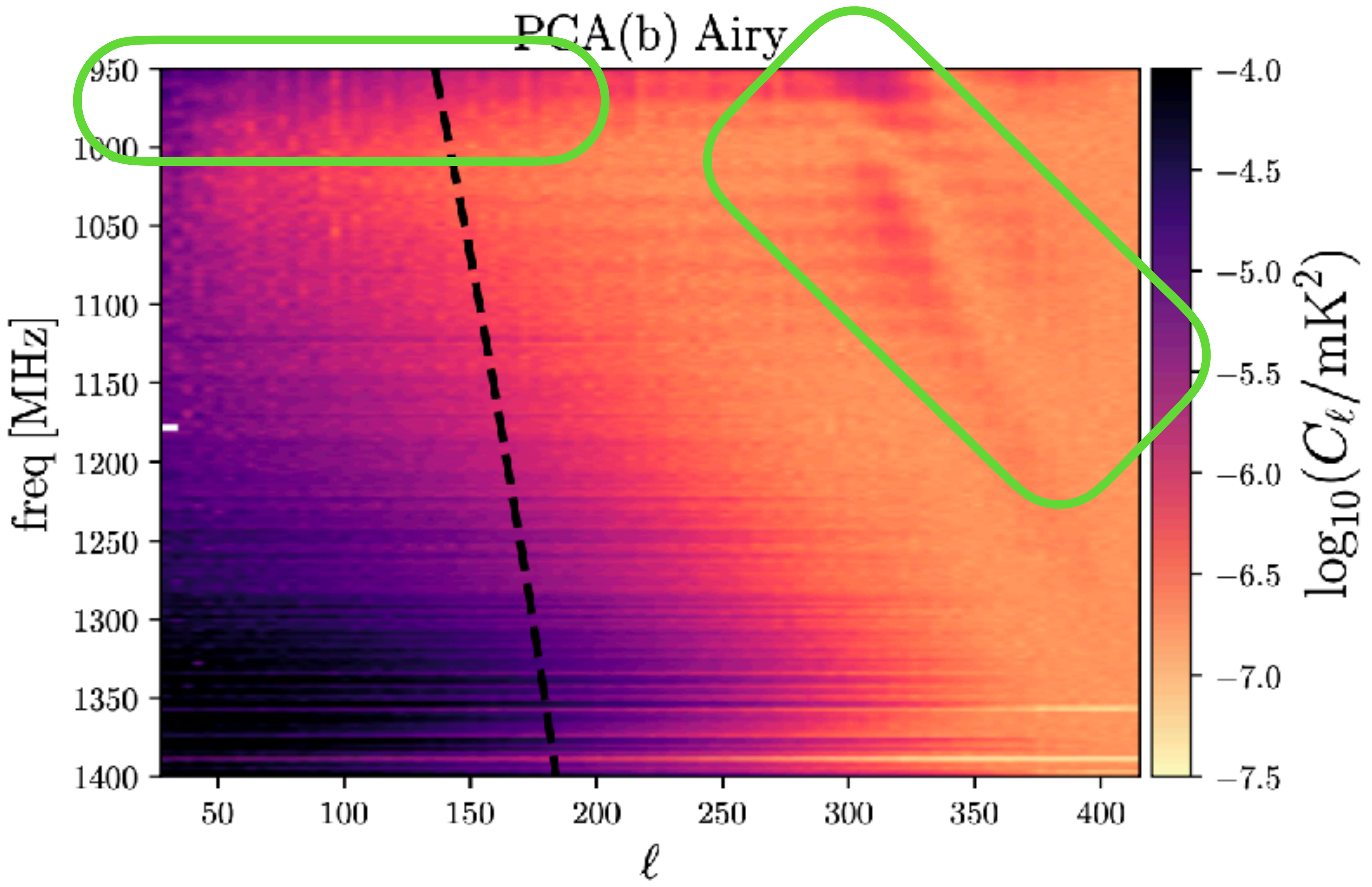
Pipelines that joined the Blind Challenge

Method	Assumption on foreground components	Pipeline	Description
Principal Component Analysis	Statistically uncorrelated	PCA(a) PCA(b) PCAwls	As in Cunnington et al. (2021b) *fg_rm* code (Alonso et al. 2015), with rms weitghing PCA applied on the wavelet-transformed data
Independent Component Analysis	Non-Gaussian	FASTICA(a) FASTICA(b)	Based on `Scikit-learn` package *fg_rm* code (Alonso et al. 2015)
Generalised Morphological Component Analysis	Sparse in a given domain and morphologically diverse	GMCA mixGMCA	As in Carucci et al. (2020) PCA on the coarse scale + GMCA on small scales
Polynomial Fitting	Smooth in frequency	poLOG	In log-log space (Alonso et al. 2015, *fg_rm* code)
Parametric Fitting	Assumptions on spectral indices	LSQ	Fit to individual foregrounds

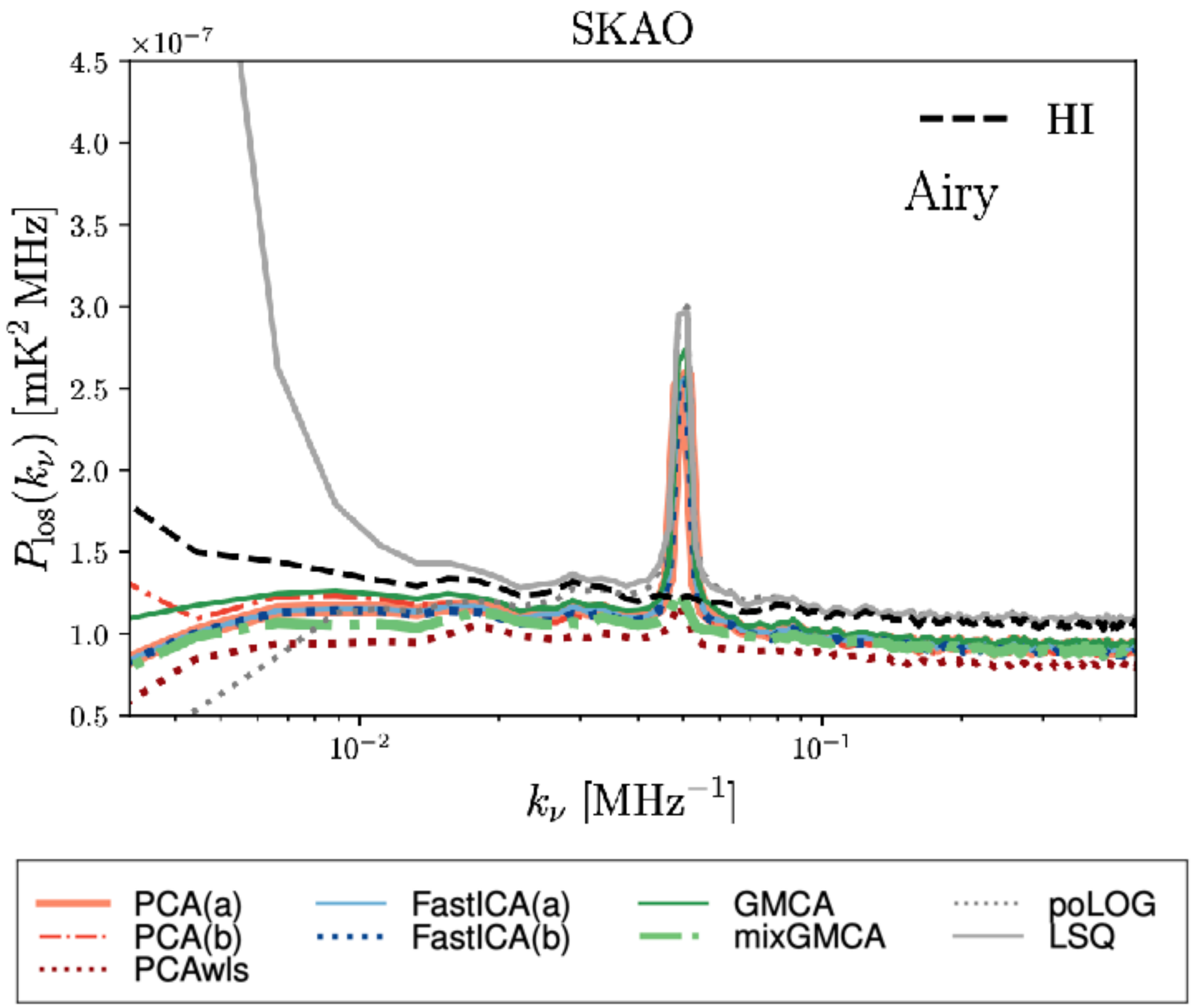
9 pipelines on 16 data cubes

Comparison at the map level: angular and radial power spectra



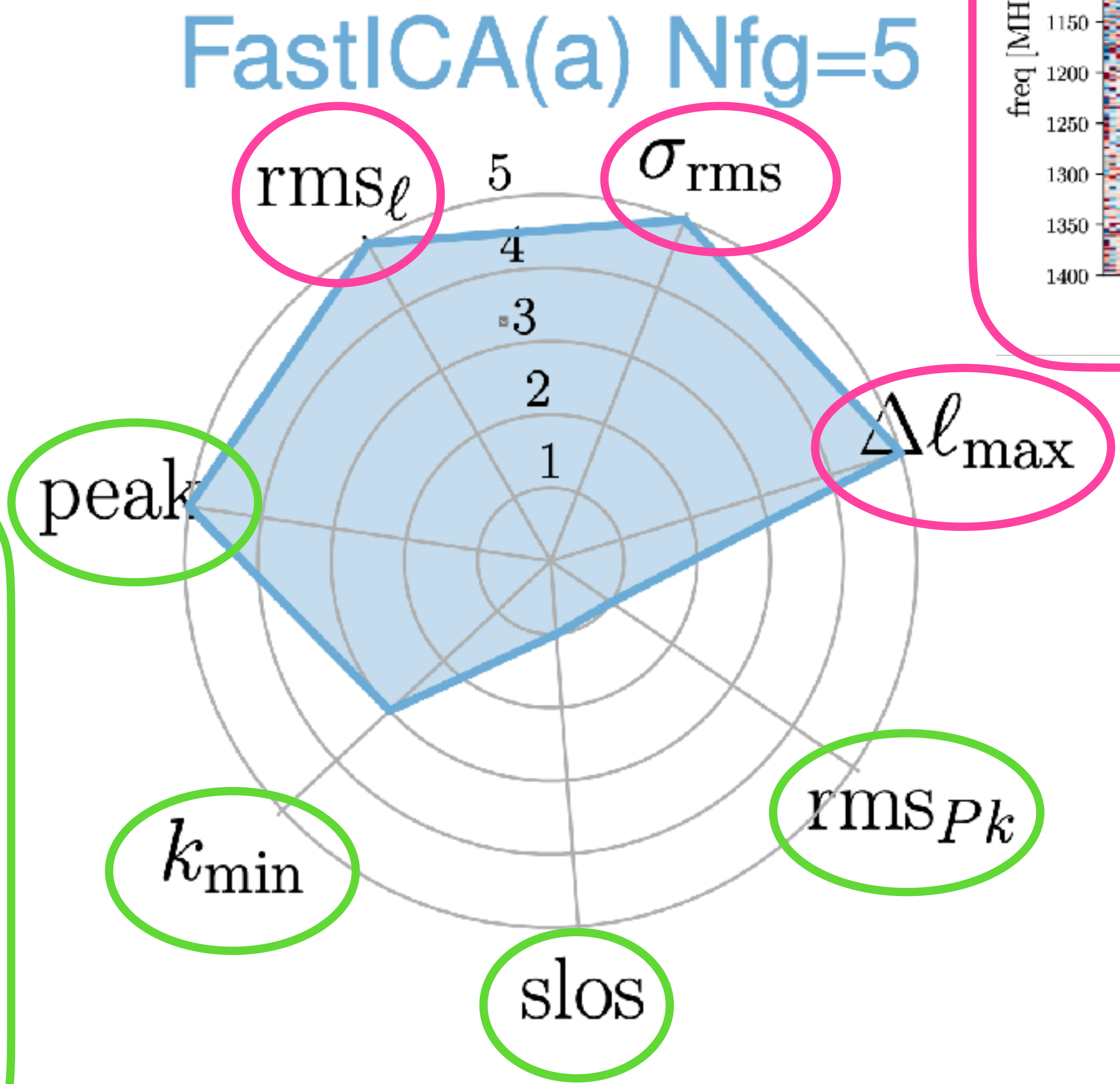


Results: radial power spectra

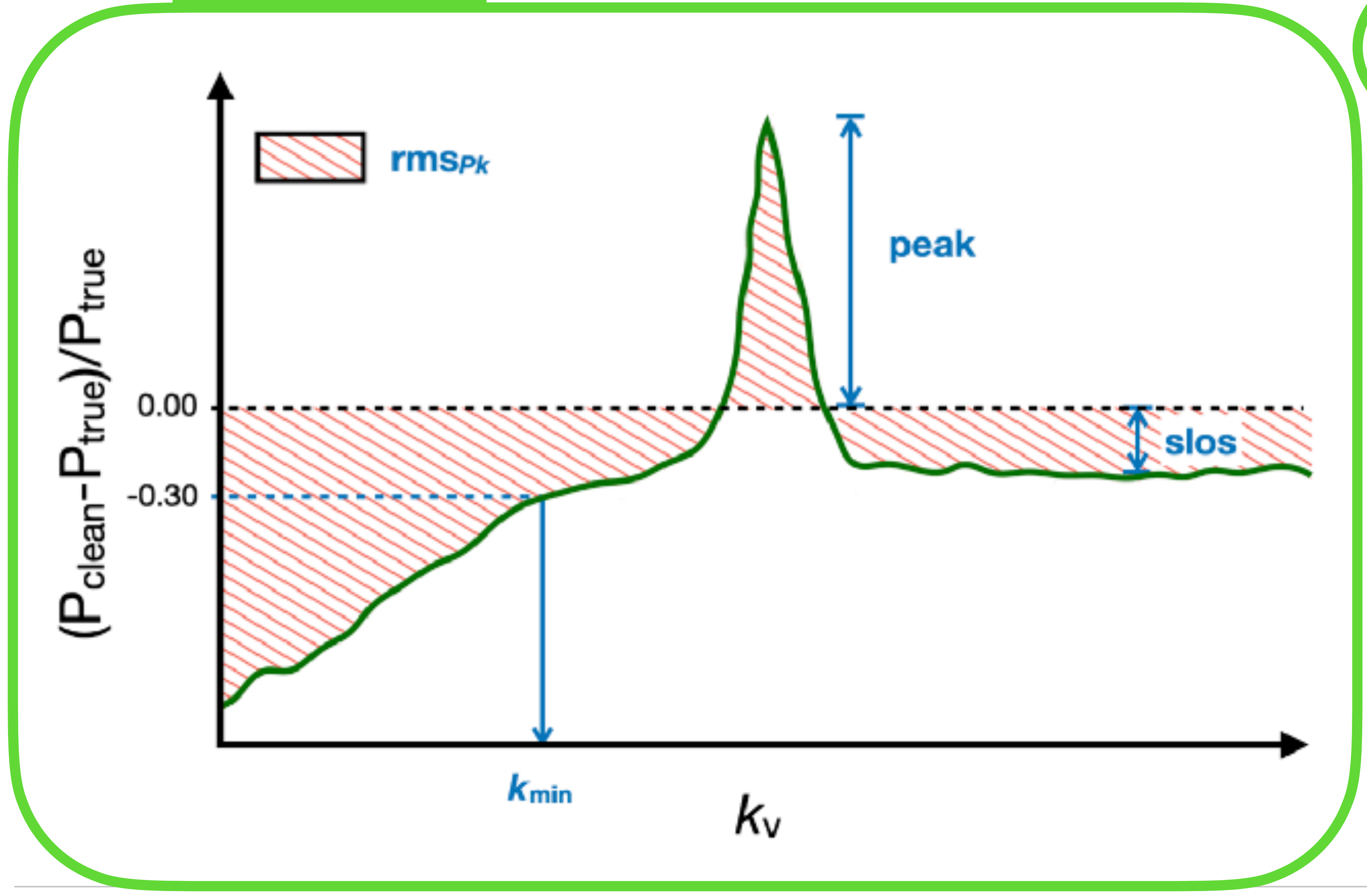


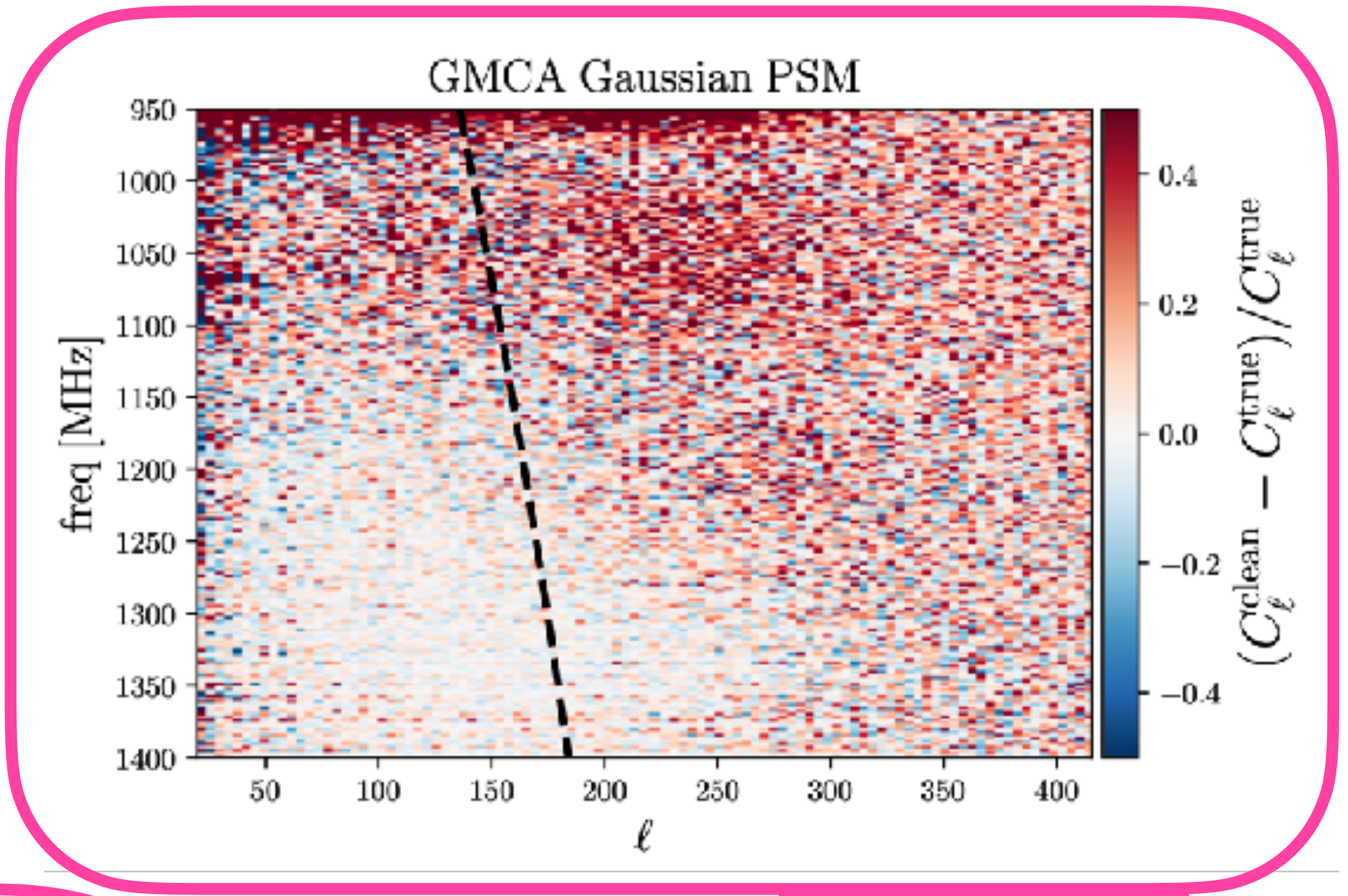
The *peak* feature in the recovered radial PS due to the interaction between the beam and the foregrounds

Results: compressed in radar charts, example



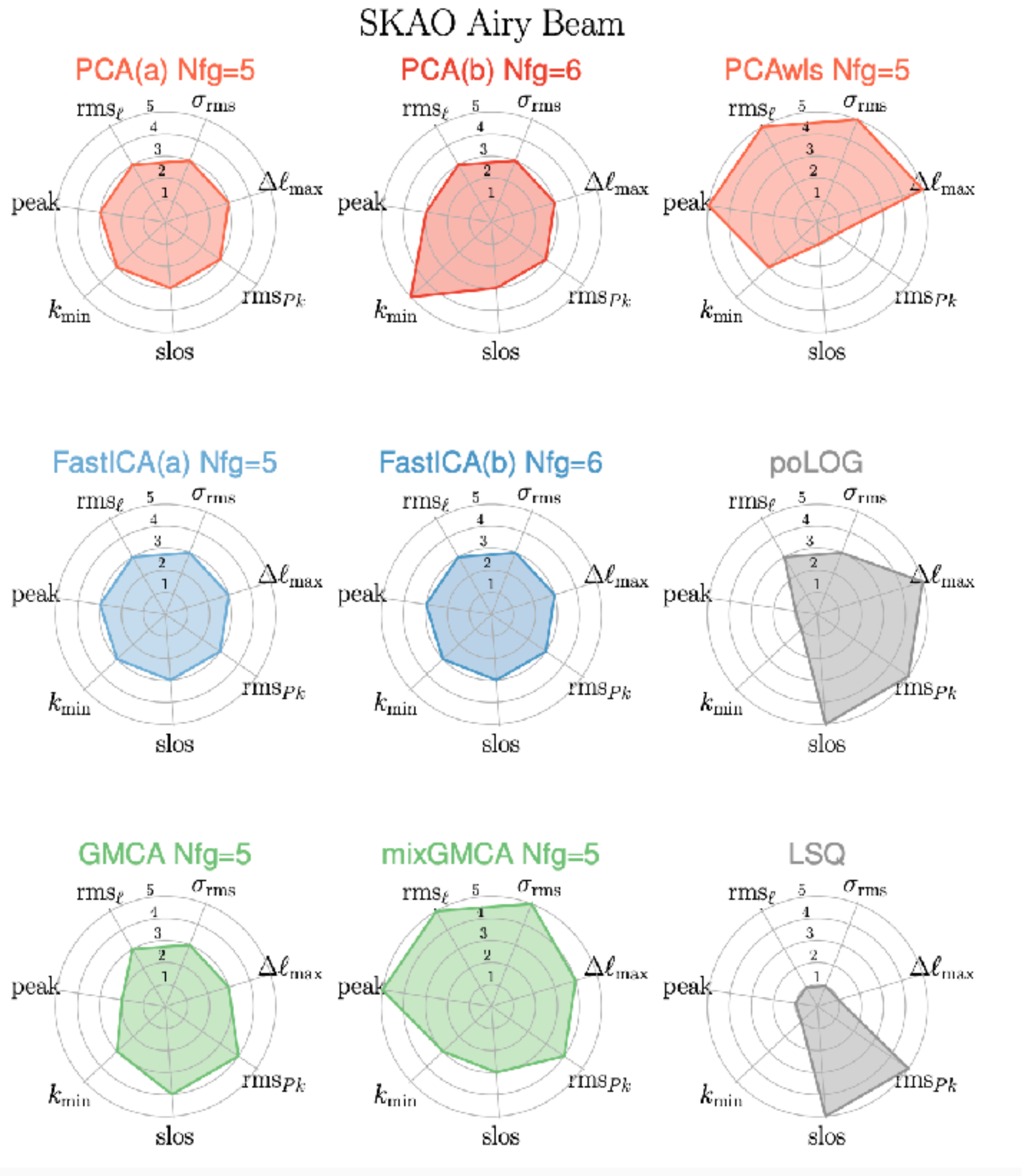
Radial PS

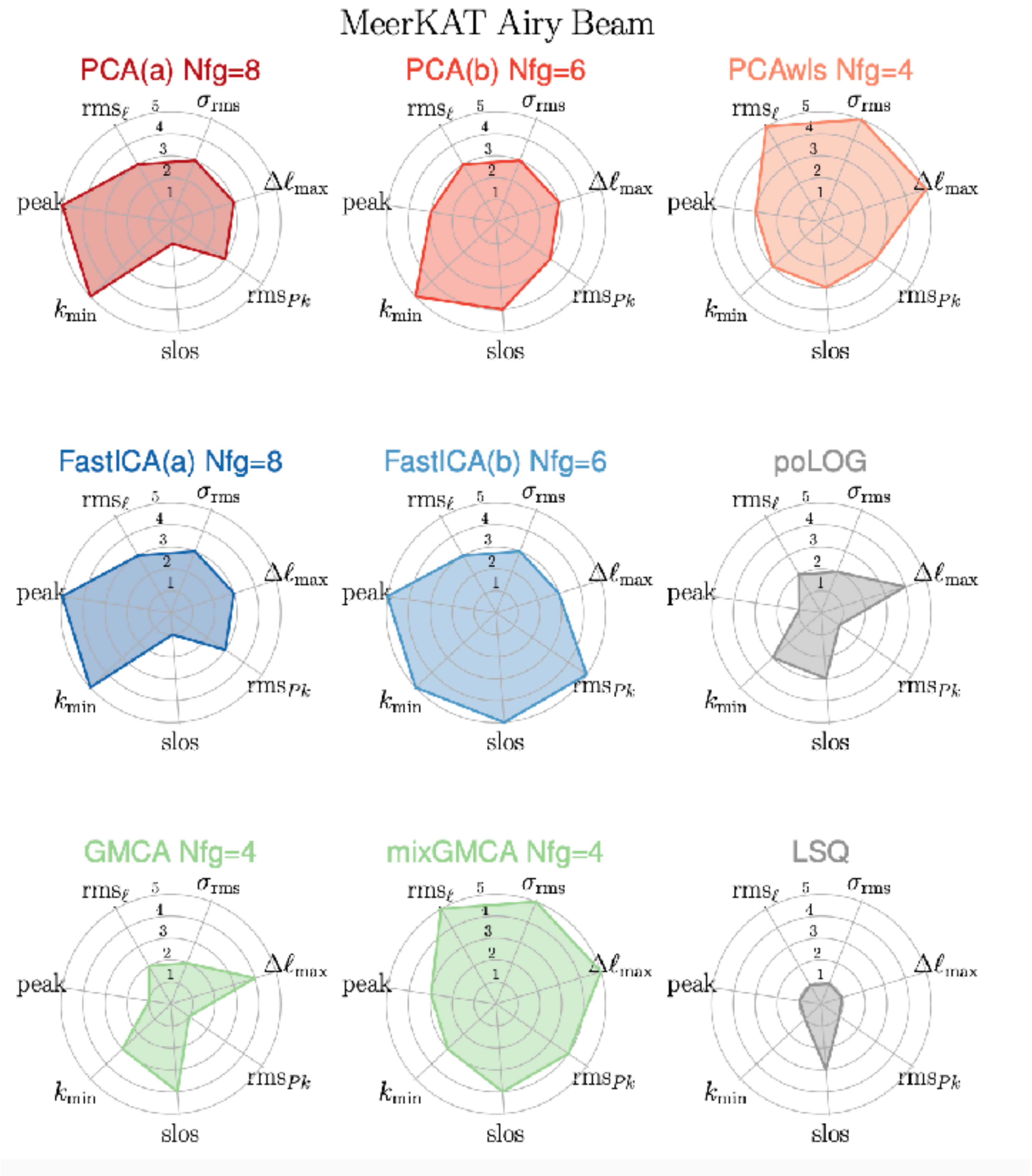




Angular PS

Results


SKAO Airy Beam
PCA(a) Nfg=5
PCA(b) Nfg=6
PCAwls Nfg=5
FastICA(a) Nfg=5
FastICA(b) Nfg=6
poLOG
GMCA Nfg=5
mixGMCA Nfg=5
LSQ
rms_ℓ
σ_{rms}
$\Delta\ell_{\max}$
rms_{Pk}
slos
$k_{\min}$
peak


MeerKAT Airy Beam
PCA(a) Nfg=8
PCA(b) Nfg=6
PCAwls Nfg=4
FastICA(a) Nfg=8
FastICA(b) Nfg=6
poLOG
GMCA Nfg=4
mixGMCA Nfg=4
LSQ
rms_ℓ
σ_{rms}
$\Delta\ell_{\max}$
rms_{Pk}
slos
$k_{\min}$
peak

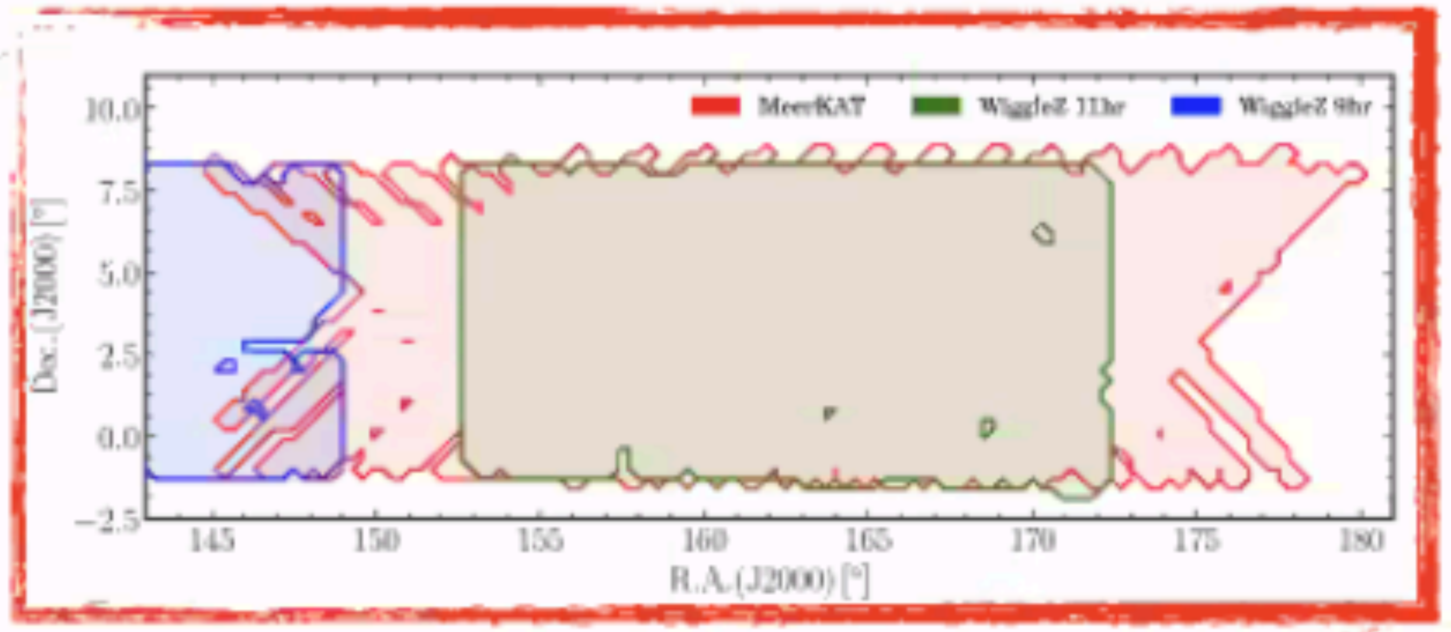
The ongoing MeerKLASS:

HI IM with the MeerKAT telescope

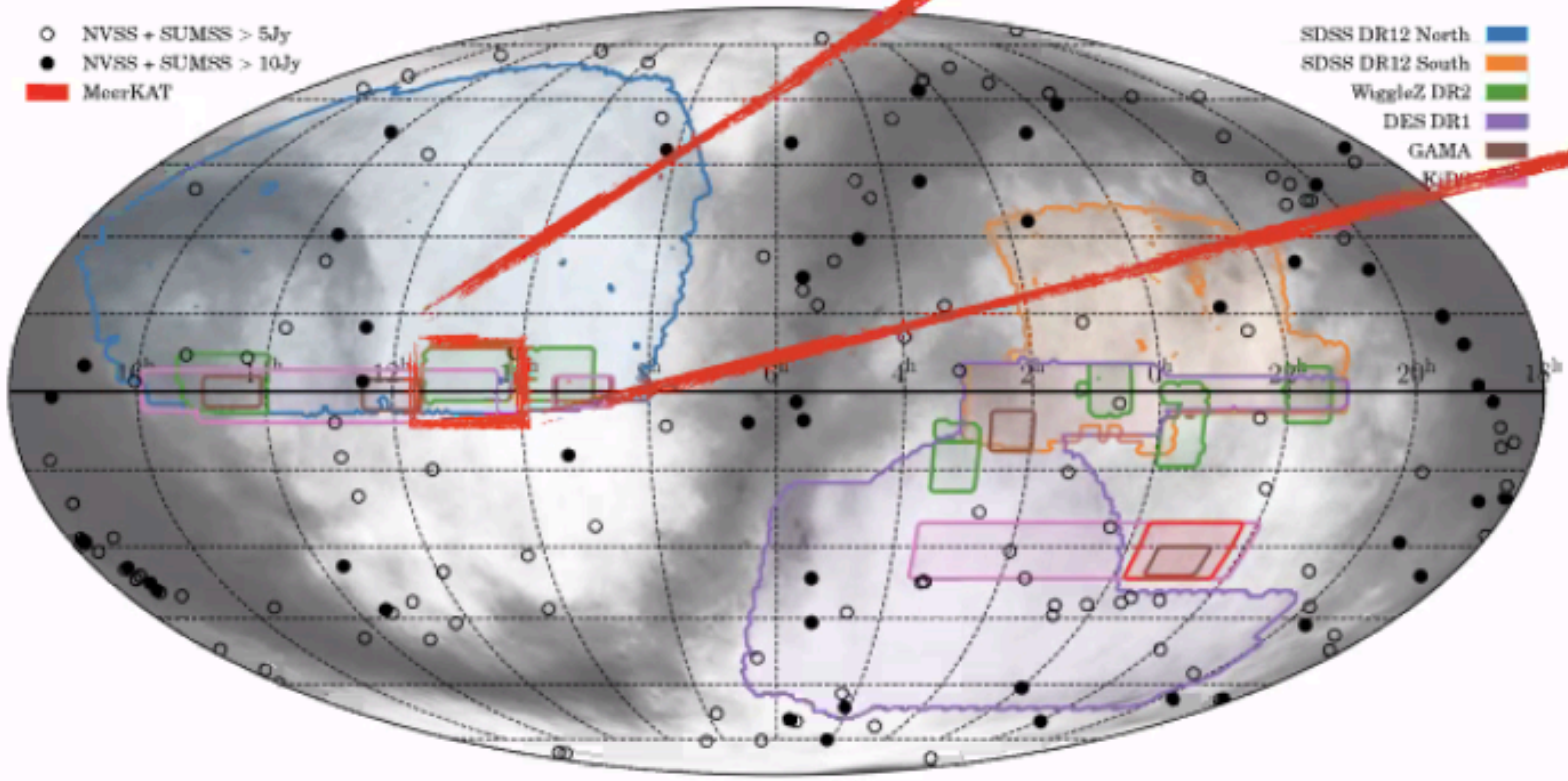
- **MeerKAT Large Area Synoptic Survey** (MeerKLASS, PI: Mario Santos)

M. G. Santos et. al. arXiv:1709.06099

- MeerKAT HI IM Pilot survey
  - **~170 square deg, ~10 hours, ~60 dishes**
  - Fix Alt ~ 45deg
  - L-band (856-1712MHz)
  - Overlap with WiggleZ/SDSS
  - Test system, training pipeline



J. Wang et. al. arXiv:2011.13789



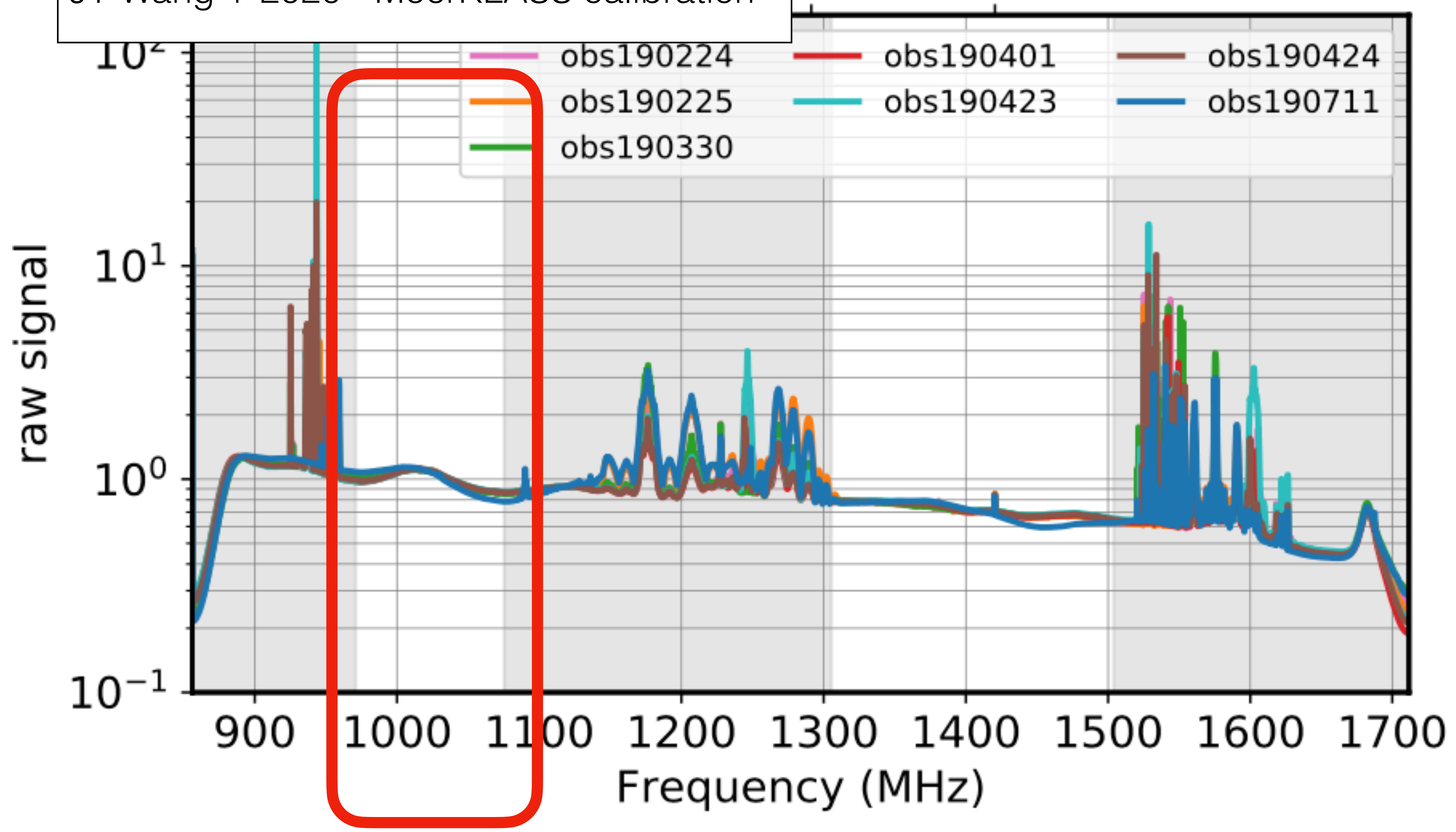
Credit: Yichao Li

2019 Data :

500 channels, from
970.97 to 1075.28 MHz,
level 6

- Satellites: big concern
- RFI-free regions 0 <z< 0.09 and 0.32 <z< 0.46
- Several rounds of RFI cleaning

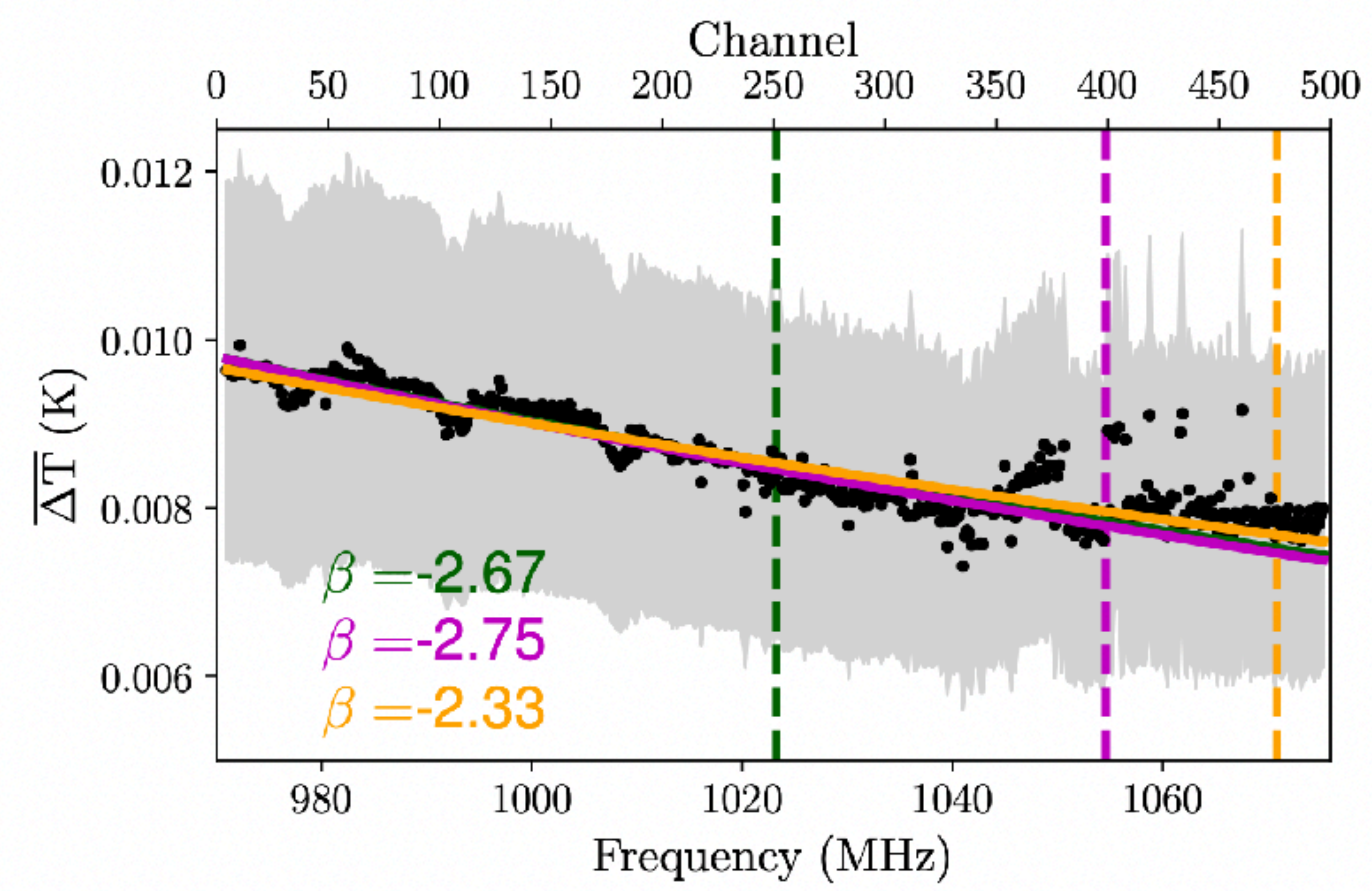
JY Wang + 2020 - MeerKLASS calibration



Measurements of the diffuse Galactic synchrotron spectral index and curvature from MeerKLASS pilot data

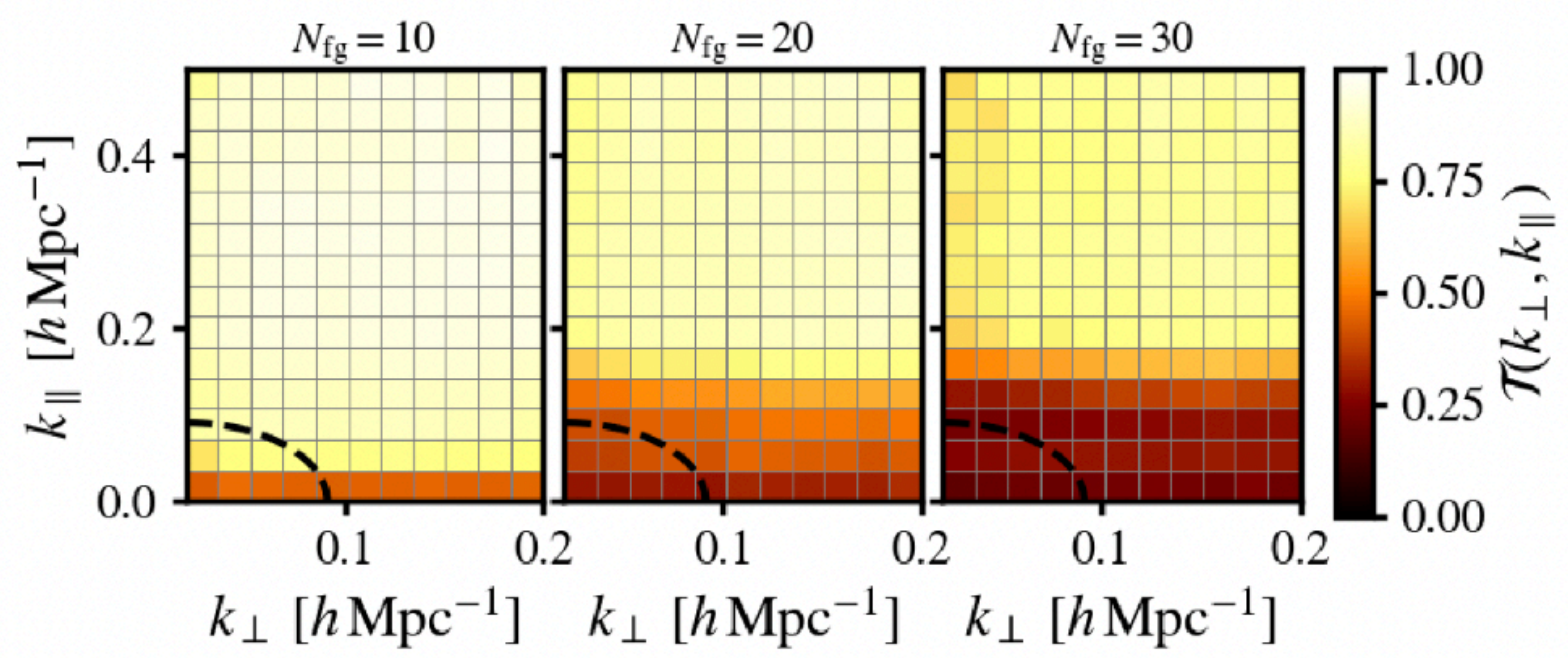
Melis O. Irfan,[1,2★] Philip Bull,[2,1] Mario G. Santos,[1,3] Jingying Wang,[1] Keith Grainge,[4]
Yichao Li,[9,1] Isabella P. Carucci,[5,6] Marta Spinelli,[7,8,1] Steven Cunnington[2]

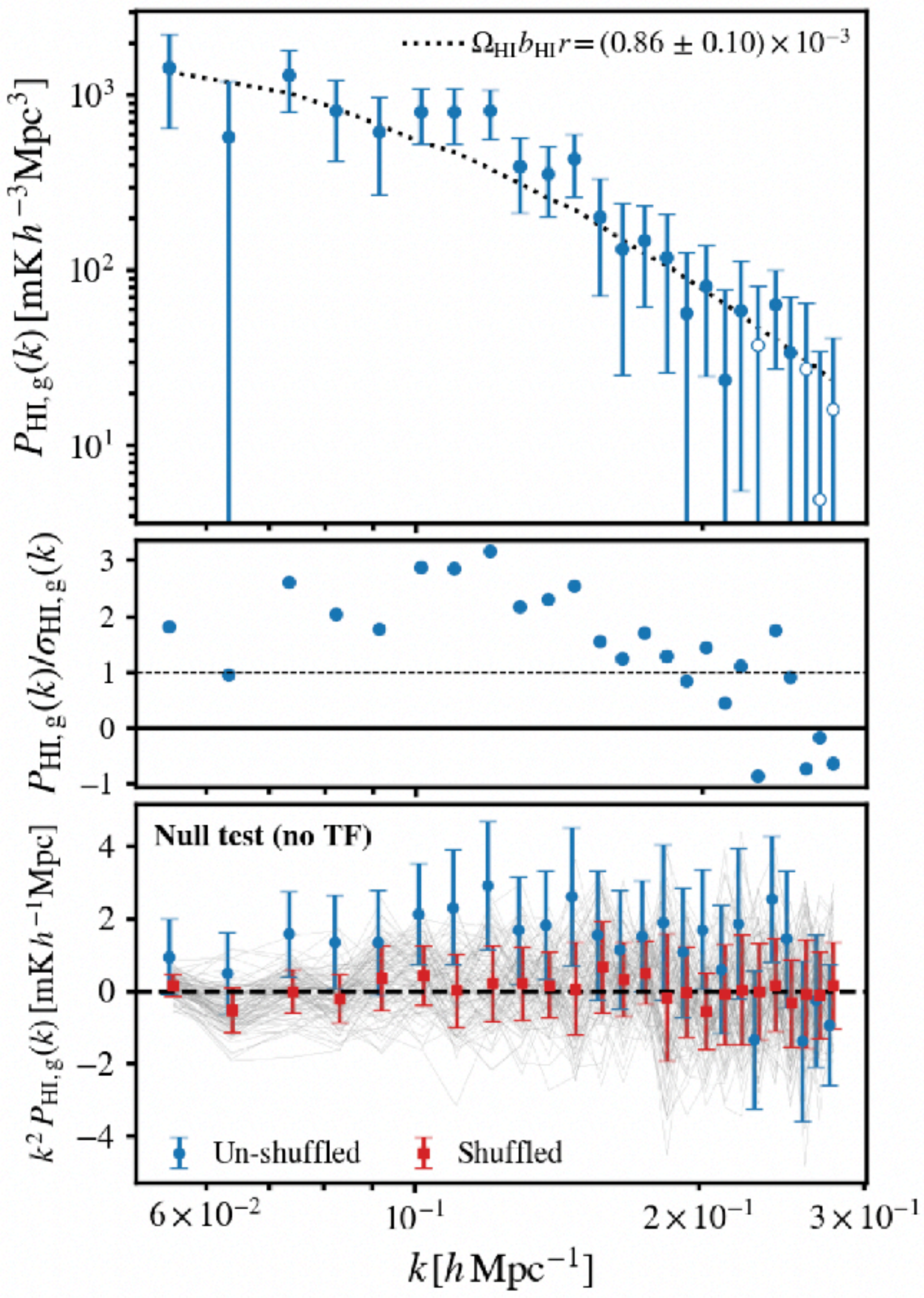
21cm intensity mapping experiments are bringing an influx of high spectral resolution observational data in the $\sim$ 100 MHz – 1 GHz regime. We use pilot 971 – 1075 MHz data from MeerKAT in single-dish mode, recently used to test the calibration and data reduction scheme of the upcoming MeerKLASS survey, to probe the spectral index of diffuse synchrotron emission below 1 GHz within $145° < \alpha < 180°$, $-1° < \delta < 8°$. Through comparisons with data from the OVRO Long Wavelength Array and the Maipu and MU surveys, we find an average spectral index of $-2.75 < \beta < -2.71$ between 45 and 1055 MHz. By fitting for spectral curvature with a spectral index of the form $\beta + c\ln(\nu/73\,\mathrm{MHz})$, we measure $\beta = -2.55 \pm 0.13$ and $c = -0.12 \pm 0.05$ within our target field. Our results are in good agreement (within 1σ) with existing measurements from experiments such as ARCADE2 and EDGES. These results show the calibration accuracy of current data and demonstrate that MeerKLASS will also be capable of achieving a secondary science goal of probing the interstellar medium.



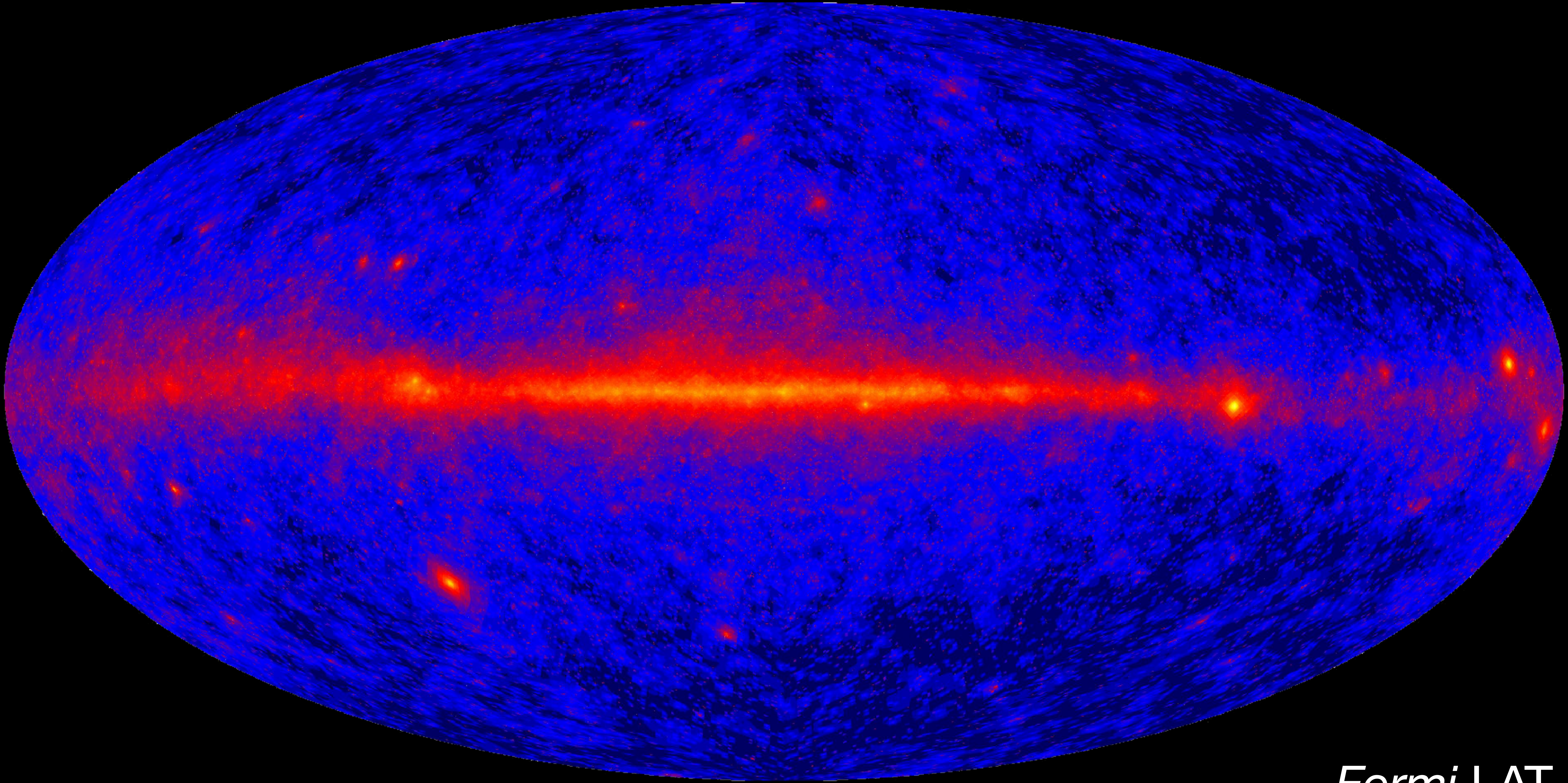
2019 Data :

Two groups with two independent pipelines:
Signal in cross with WiggleZ-11h galaxies is there!

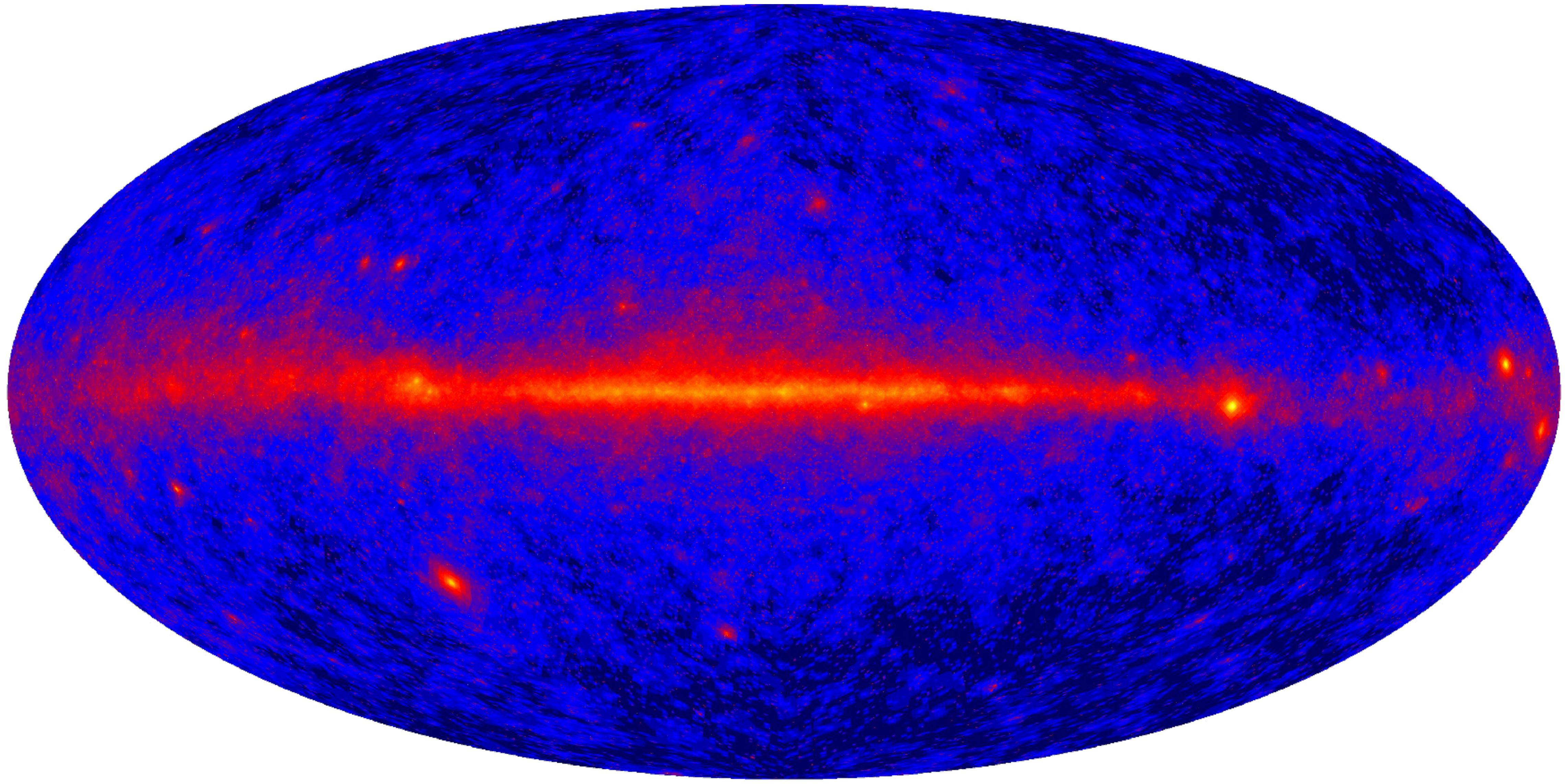




Gamma - rays ?

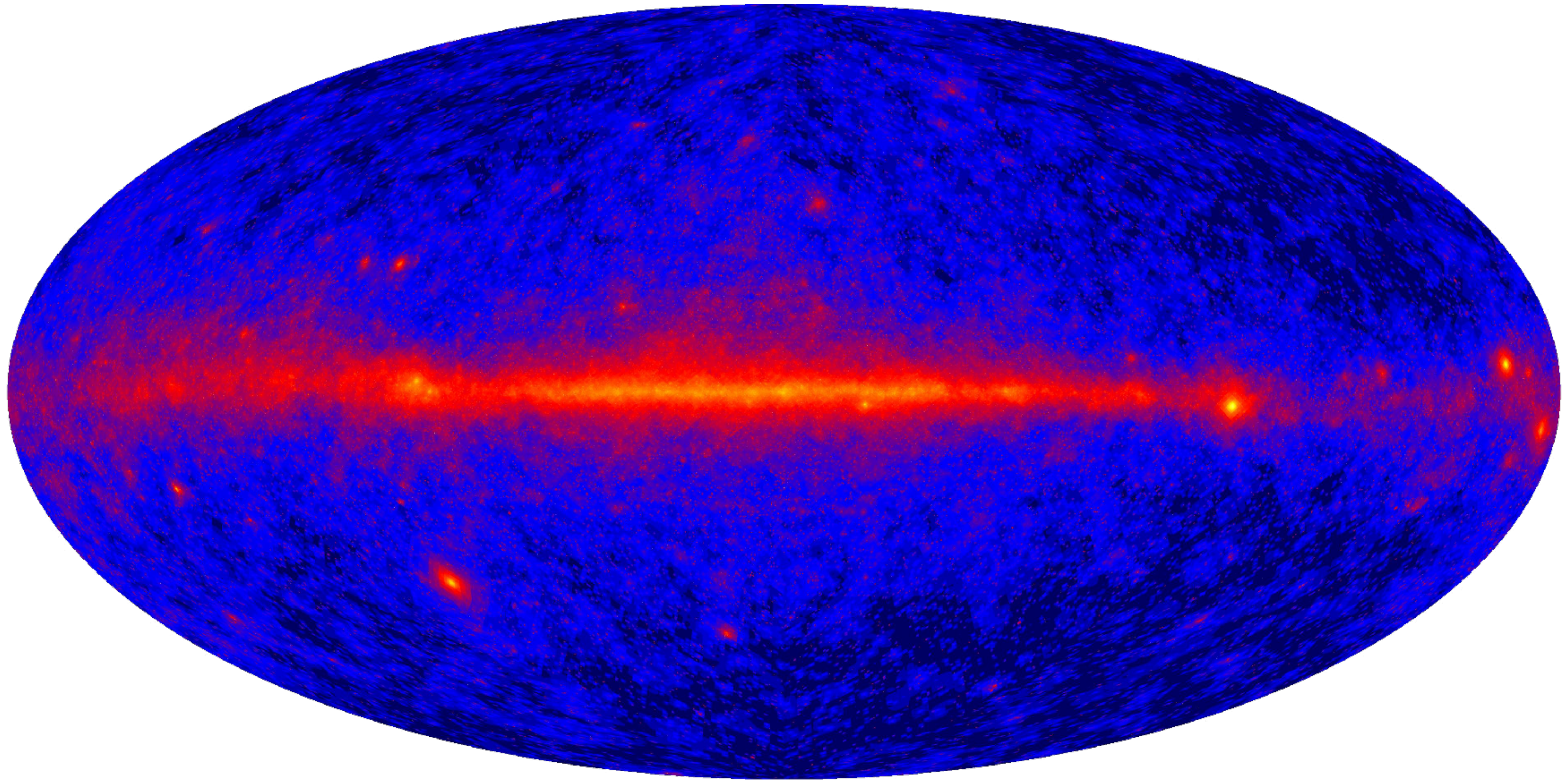


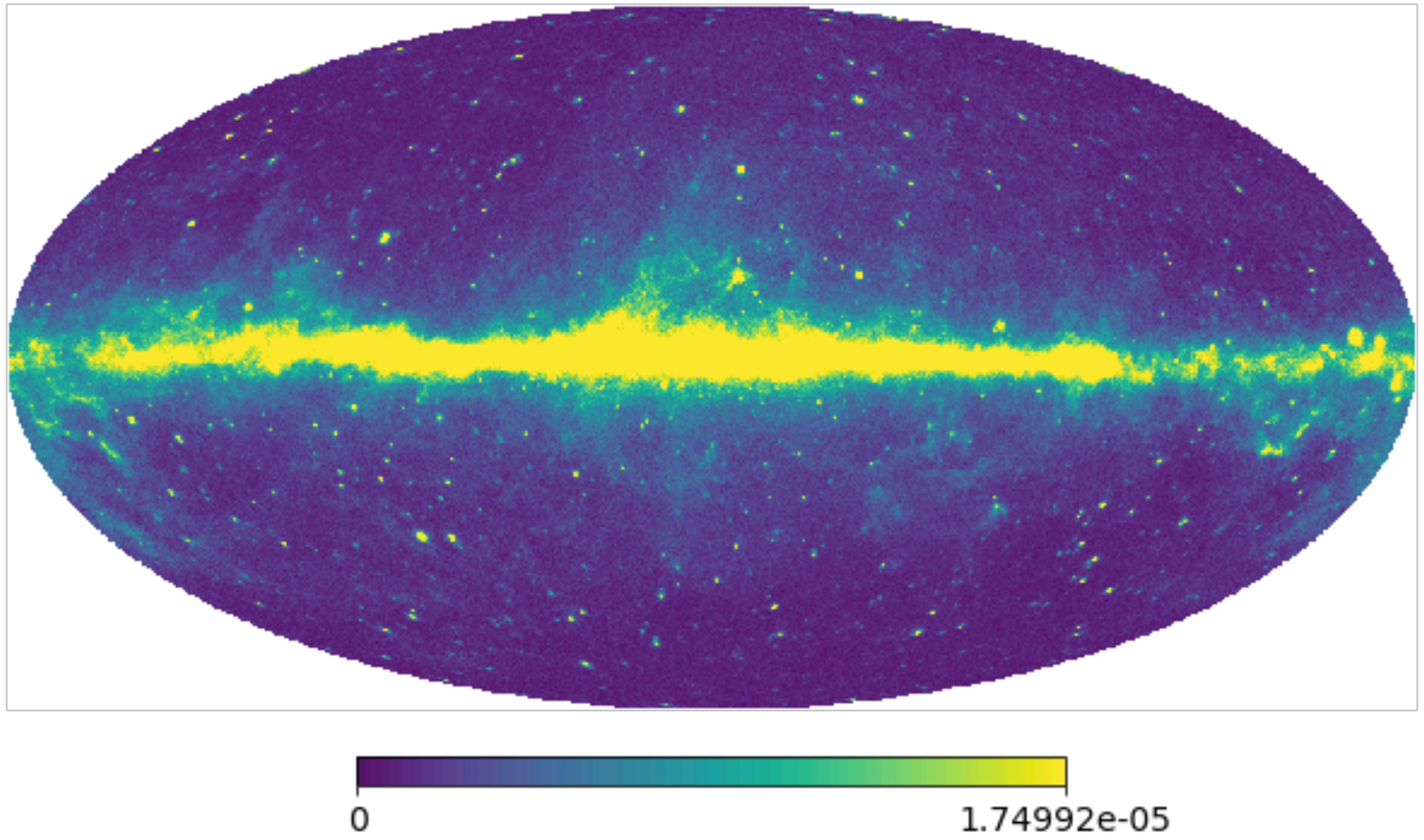
Fermi LAT



the galactic diffuse emission

- Interaction of cosmic rays (CR) with interstellar medium
  - protons+nuclei -> decay of secondary pions
  - Bremsstrahlung and inverse Colton scattering of CR electrons with IR/UV photons of the interstellar radiation field




0
1.74992e-05

1. Blind analysis of data

mixing matrix (f,n)

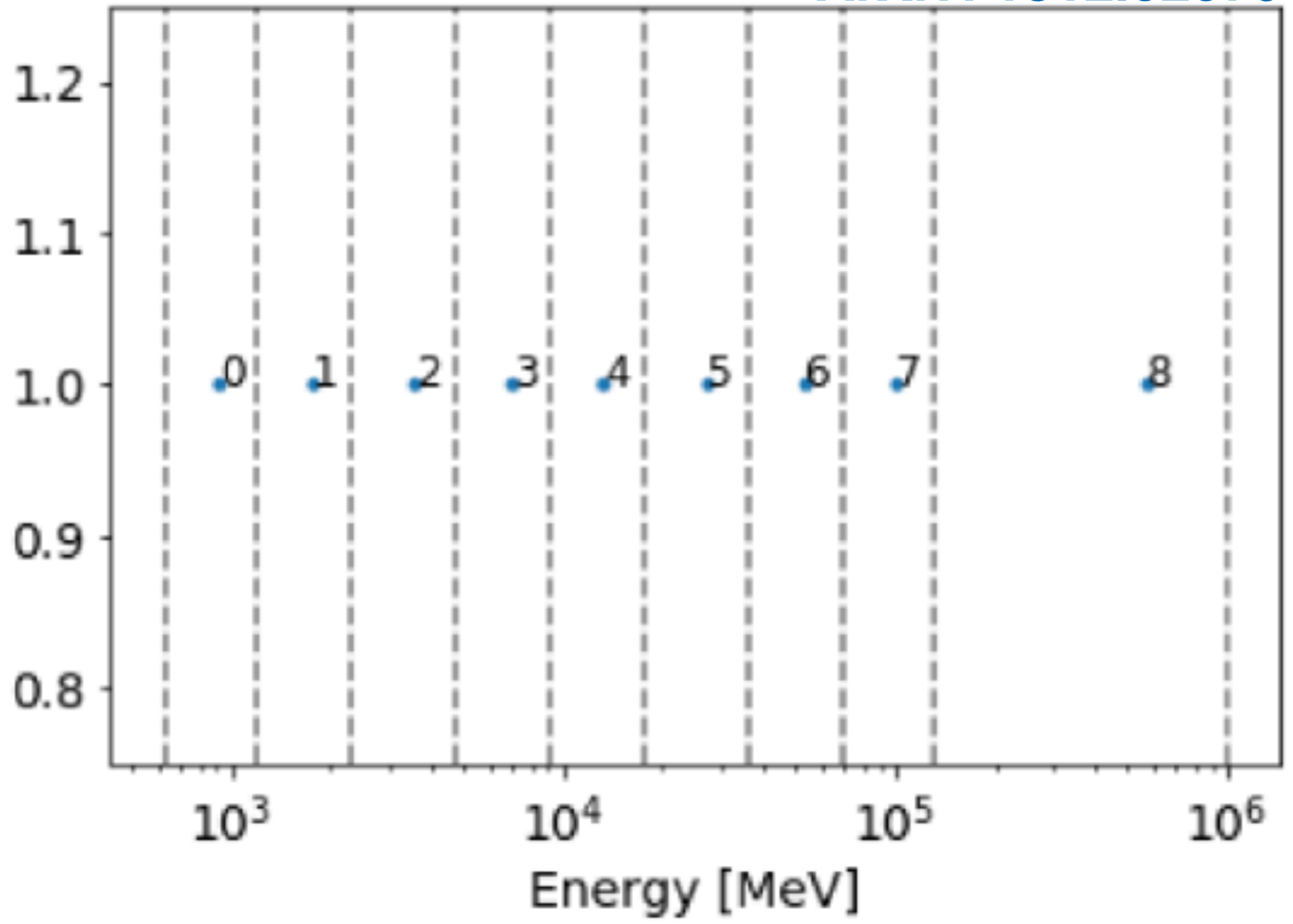
$$\mathbf{X} = \mathbf{A}\ \mathbf{S} + \mathbf{N}$$

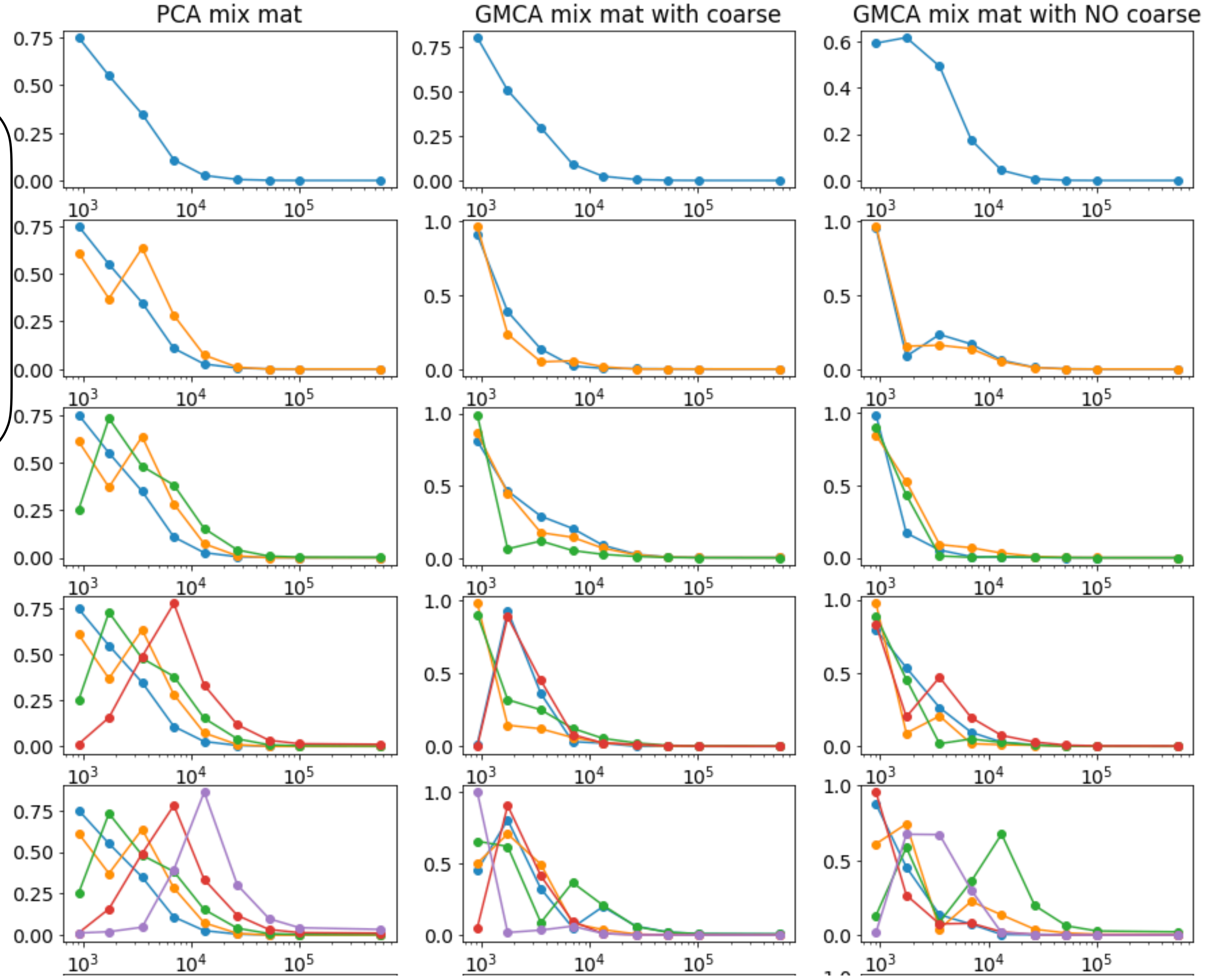
signal (f,p)

sources (n,p)

Unresolved gamma-ray background!

ArXiv: 1812.02079





Strategies to identify the Galactic Foreground

- HI IM will bring new radio data in the ~100 MHz - 1 GHz regime
- Go statistical, let the signal processing scientists do the job!
  (In HI intensity mapping we are using these techniques successfully)
  - When things like the slab model is not enough
  - Measuring the radio SZ: getting read of relics, halos and whatever has structure (compared to the background)
  - …