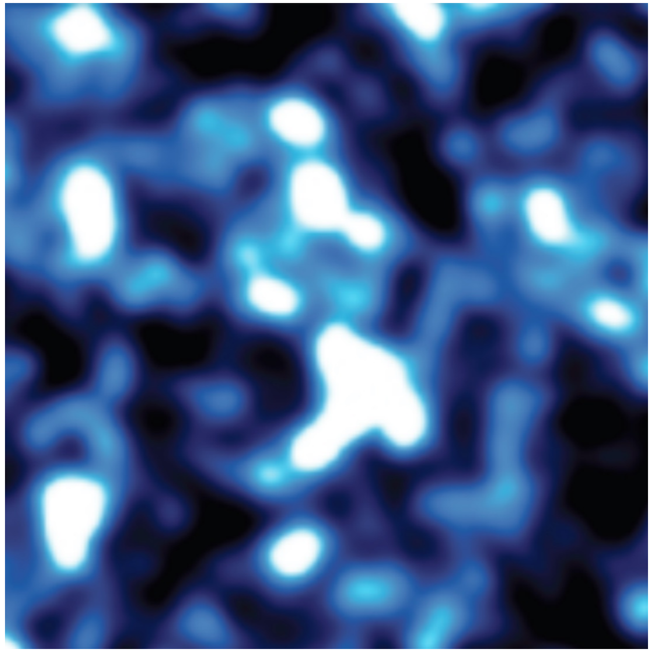
Paving the way for the SKAO
Forecast for radio-optical synergies and the analysis of MeerKAT data

SKA Cosmology SWG Annual Meeting 2024

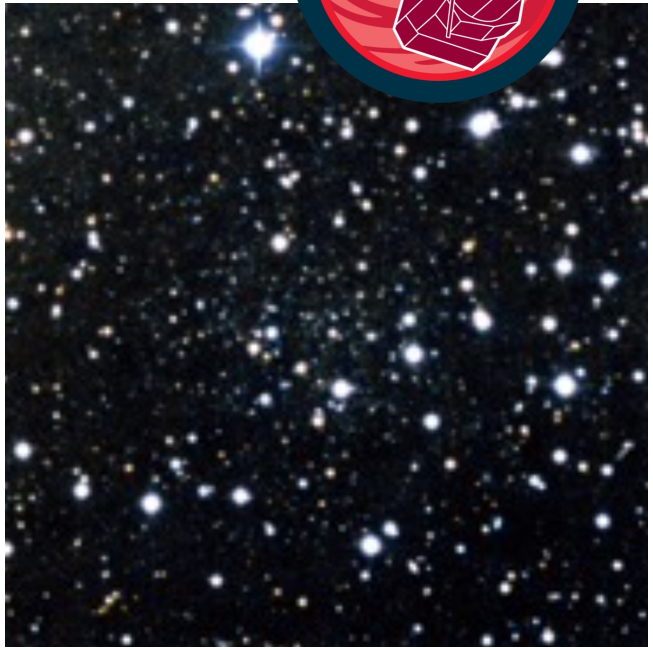
Matilde Barberi Squarotti, Università degli Studi di Milano

A new generation of surveys

HI intensity mapping surveys



Galaxy surveys



A new generation of surveys

- High accuracy (resolution and sensitivity)
- Wide cosmic volumes covered
  - Large sky coverage
  - Extended redshift range
- Multi-wavelength and multi-tracer
  - Independent biased tracers of the dark matter distribution are surveyable
  - Complementary information
  - Convenient to beat the cosmic variance

Primordial non-Gaussianity

- Predicted by inflationary models
  - Direct proof of inflation
  - Information on the dynamics of Early Universe
- Primordial non-Gaussianity of local-type: $f_{\mathrm{NL}} \neq 0$
- Current constraints from CMB [Akrami et al. (2018)]
  - $|f_{\mathrm{NL}}| < 5$
  - $\sigma(f_{\mathrm{NL}}) \sim 5$
- Primordial non-Gaussianity and the Large Scale Structure
  - $b_A(z) \rightarrow b_A(z) + \Delta b_A(k,z) f_{\mathrm{NL}}$

$$\Delta b_A(k,z) = 3\,[b_A(z) - 1\,]\,\frac{\delta_c \Omega_{\mathrm{m},0} H_0^2}{c^2 k^2 T(k) D(z)}$$

Power spectrum

- $P_{AB}(k,z,\mu) = [b_A(z) + \Delta b_A(k,z) f_{\rm NL} + f(z)\mu^2][b_B(z) + \Delta b_B(k,z) f_{\rm NL} + f(z)\mu^2] P_{\rm m}(k,z)$

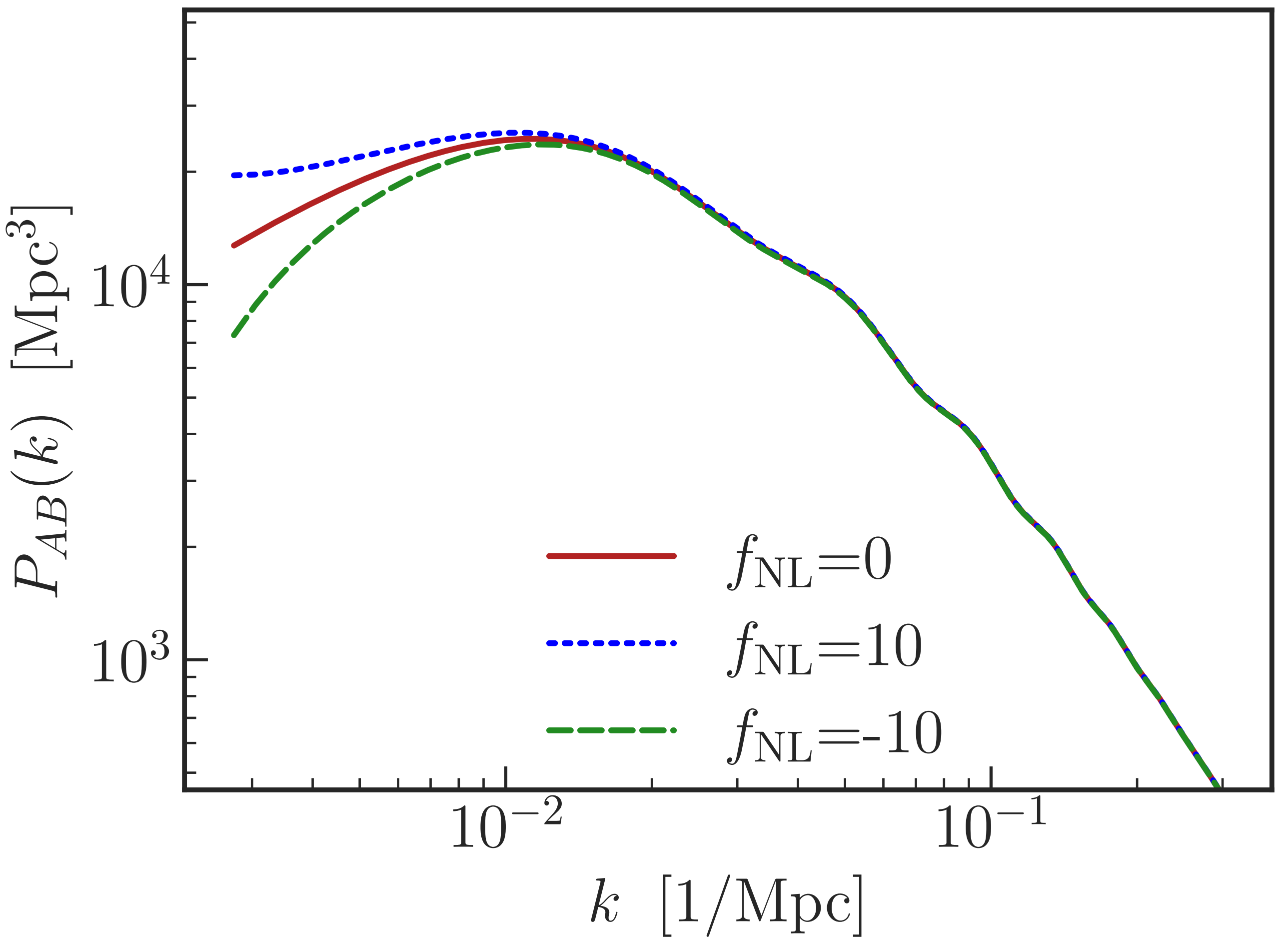
Primordial non-Gaussianity

$$\Delta b_A(k,z) = 3\,[b_A(z) - 1\,] \frac{\delta_c \Omega_{\rm m,0} H_0^2}{c^2 k^2 T(k) D(z)}$$

$$\Delta b_B(k,z) = 3\,[b_B(z) - 1\,] \frac{\delta_c \Omega_{\rm m,0} H_0^2}{c^2 k^2 T(k) D(z)}$$

Power spectrum

- $P_{AB}(k,z,\mu) = [b_A(z) + \Delta b_A(k,z) f_{\rm NL} + f(z)\mu^2][b_B(z) + \Delta b_B(k,z) f_{\rm NL} + f(z)\mu^2] P_{\rm m}(k,z)$



General setup

- 12 redshift bins for $z \in [0.85,4]$
- 10 bins for $k \in [k_{\min}(z), k_{\max}(z)]$
  - $k_{\min}(z) = \frac{2\pi}{V^{1/3}(z)}$
  - $k_{\max}(z) = 0.08(1+z)^{2/(2+n_s)} h\mathrm{Mpc}^{-1}$

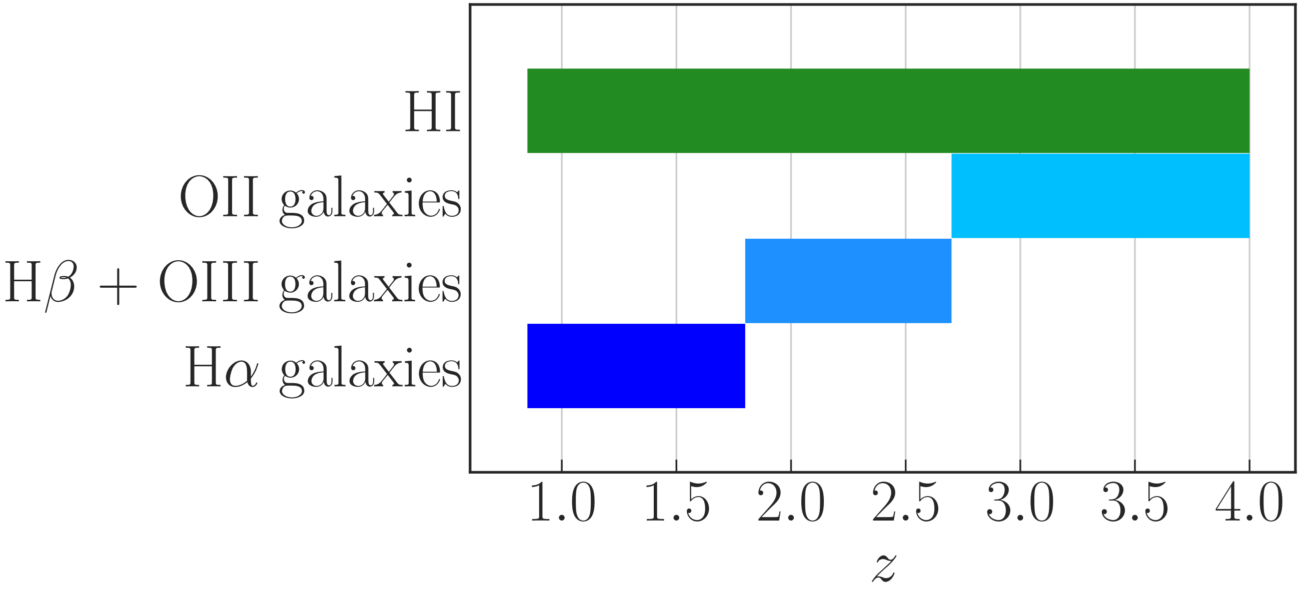
Galaxy survey

- Stage-IV spectroscopic survey
- Redshift range covered with different ELG types [Fonseca, Camera (2020)]
- Flux limit: $F_c = 2 \cdot 10^{-16} \mathrm{erg\ s^{-1}\ cm^{-2}}$
- Sky coverage: $f_{\mathrm{sky}} = 0.36$
- Shot noise

HI IM survey

- SKAO-like survey
- Sky coverage: $f_{\mathrm{sky}} = 0.48$
- Thermal noise

General setup



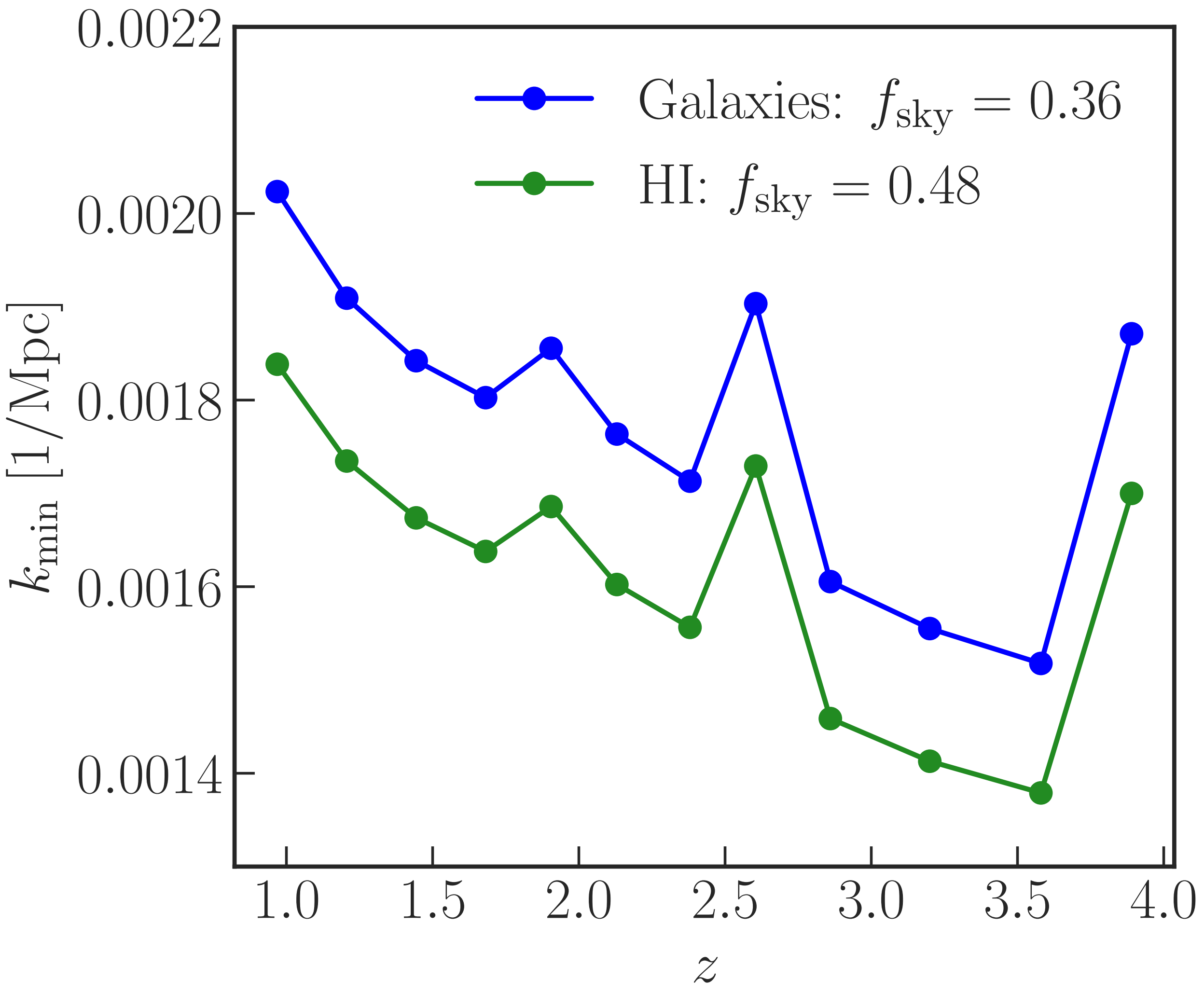
Galaxy survey

- Stage-IV spectroscopic survey
- Redshift range covered with different ELG types [Fonseca, Camera (2020)]
- Flux limit: $F_c = 2 \cdot 10^{-16} \mathrm{erg\ s^{-1}\ cm^{-2}}$
- Sky coverage: $f_{sky} = 0.36$
- Shot noise

HI IM survey

- SKAO-like survey
- Sky coverage: $f_{sky} = 0.48$
- Thermal noise

General setup



Galaxy survey

- Stage-IV spectroscopic survey
- Redshift range covered with different ELG types [Fonseca, Camera (2020)]
- Flux limit: $F_c = 2 \cdot 10^{-16}$erg s^{-1} cm^{-2}
- Sky coverage: $f_{sky} = 0.36$
- Shot noise

HI IM survey

- SKAO-like survey
- Sky coverage: $f_{sky} = 0.48$
- Thermal noise

HI systematics

- HI auto-power spectrum: $P_{\rm HIHI}(k,z,\mu) \longrightarrow \mathcal{D}_{\rm b}{}^2(k,z,\mu)\mathcal{D}_{\rm fg}(k,z,\mu)P_{\rm HIHI}(k,z,\mu)$
- Galaxy-HI cross-power spectrum: $P_{\rm gHI}(k,z,\mu) \longrightarrow \mathcal{D}_{\rm b}(k,z,\mu)\mathcal{D}_{\rm fg}(k,z,\mu)P_{\rm gHI}(k,z,\mu)$
  - Beam damping

    $\mathcal{D}_{\rm b}(k,z,\mu) = \exp\left[-\frac{(1-\mu^2)k^2\chi^2(z)\theta_{\rm b}^2(z)}{16\ln 2}\right]$ with $\theta_{\rm b}$ the beam of the dish
  - Foreground contamination

    $\mathcal{D}_{\rm fg}(k,z,\mu) = 1 - \exp\left[-\left(\frac{\mu k}{k_{\parallel \rm fg}}\right)^2\right]$ with $k_{\parallel \rm fg} = 0.01 h{\rm Mpc}^{-1}$

MULTI-TRACER TECHNIQUE

- Large scales: cosmic variance limited
- Multi-tracer technique: combination of independent tracers of the same underlying matter distribution to overcome cosmic variance [Seljak (2008)]

- Data vector

$$\boldsymbol{P} = \{P_{\mathrm{gg}}, P_{\mathrm{gH I}}, P_{\mathrm{H I H I}}\}$$

- Full covariance matrix [Karagiannis et al.(2023)]

$$\mathrm{Cov}(\boldsymbol{P}) = \frac{2}{N_{\mathrm{m}}} \begin{bmatrix} \tilde{P}_{\mathrm{gg}}^2 & \tilde{P}_{\mathrm{gg}}\tilde{P}_{\mathrm{gH I}} & \tilde{P}_{\mathrm{gH I}}^2 \\ \tilde{P}_{\mathrm{gg}}\tilde{P}_{\mathrm{gH I}} & \frac{1}{2}\left(\tilde{P}_{\mathrm{gg}}\tilde{P}_{\mathrm{H I H I}} + \tilde{P}_{\mathrm{gH I}}^2\right) & \tilde{P}_{\mathrm{H I H I}}\tilde{P}_{\mathrm{gH I}} \\ \tilde{P}_{\mathrm{gH I}}^2 & \tilde{P}_{\mathrm{H I H I}}\tilde{P}_{\mathrm{gH I}} & \tilde{P}_{\mathrm{H I H I}}^2 \end{bmatrix}$$

$$\tilde{P}_{AB} = P_{AB} + P_{AB}^{\mathrm{noise}}\delta_{AB}$$

Analysis

- Maximization of the likelihood with MCMC method

$$\ln \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{d}) \propto [\boldsymbol{d} - \boldsymbol{P}(\boldsymbol{\theta})]^{\mathrm{T}} \mathrm{Cov}^{-1} [\boldsymbol{d} - \boldsymbol{P}(\boldsymbol{\theta})]$$

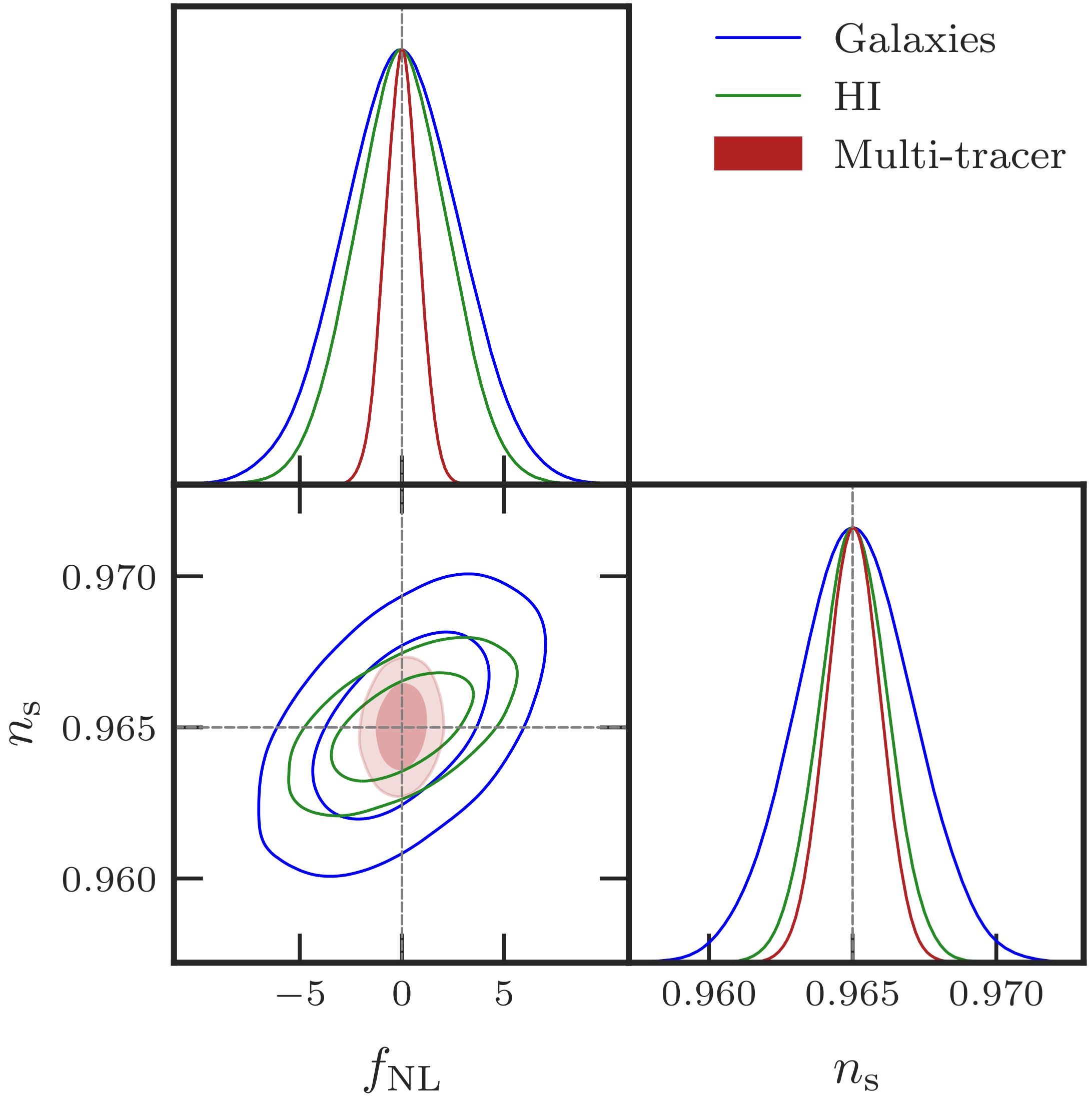
  - $\boldsymbol{P}(\boldsymbol{\theta}) =$ theoretical data vector
  - $\boldsymbol{\theta} = \{f_{\mathrm{NL}}, n_{\mathrm{s}}, b_{\mathrm{g}}(z), b_{\mathrm{HI}}(z)\}$
- Full multi-tracer likelihood [Viljoen et al. (2020)]

$$\ln \mathcal{L}_{\mathrm{MT}}^{\mathrm{tot}} = \ln \mathcal{L}_{\mathrm{MT}}^{\mathrm{overlap}} + \ln \mathcal{L}_{\mathrm{gg}}^{\mathrm{non-overlap}} + \ln \mathcal{L}_{\mathrm{HIHI}}^{\mathrm{non-overlap}}$$

- Analysis of the constraints:
  - All the redshift bins
  - With respect to the redshift (2 redshift bins at a time)
  - With respect to the ELG type of the galaxy surveys (4 redshift bins at a time)

Results

- All fiducial parameters recovered and constrained
- Marginalized uncertainty on f_{NL}:
  - Galaxies auto-correlation: $f_{\mathrm{NL}} = 0.0 \pm 2.8$
  - HI auto-correlation: $f_{\mathrm{NL}} = 0.0 \pm 2.3$
  - Multi-tracer techique: $f_{\mathrm{NL}} = 0.0 \pm 0.76$
- Multi-tracer technique: tightest constraints
- HI more constraining than galaxies:
  - $P^{\mathrm{noise}}_{\mathrm{HIHI}} < P^{\mathrm{noise}}_{\mathrm{gg}}$
  - $f_{\mathrm{sky,HI}} > f_{\mathrm{sky,g}}$



[MBS, Camera, Maartens (2023)]

Results

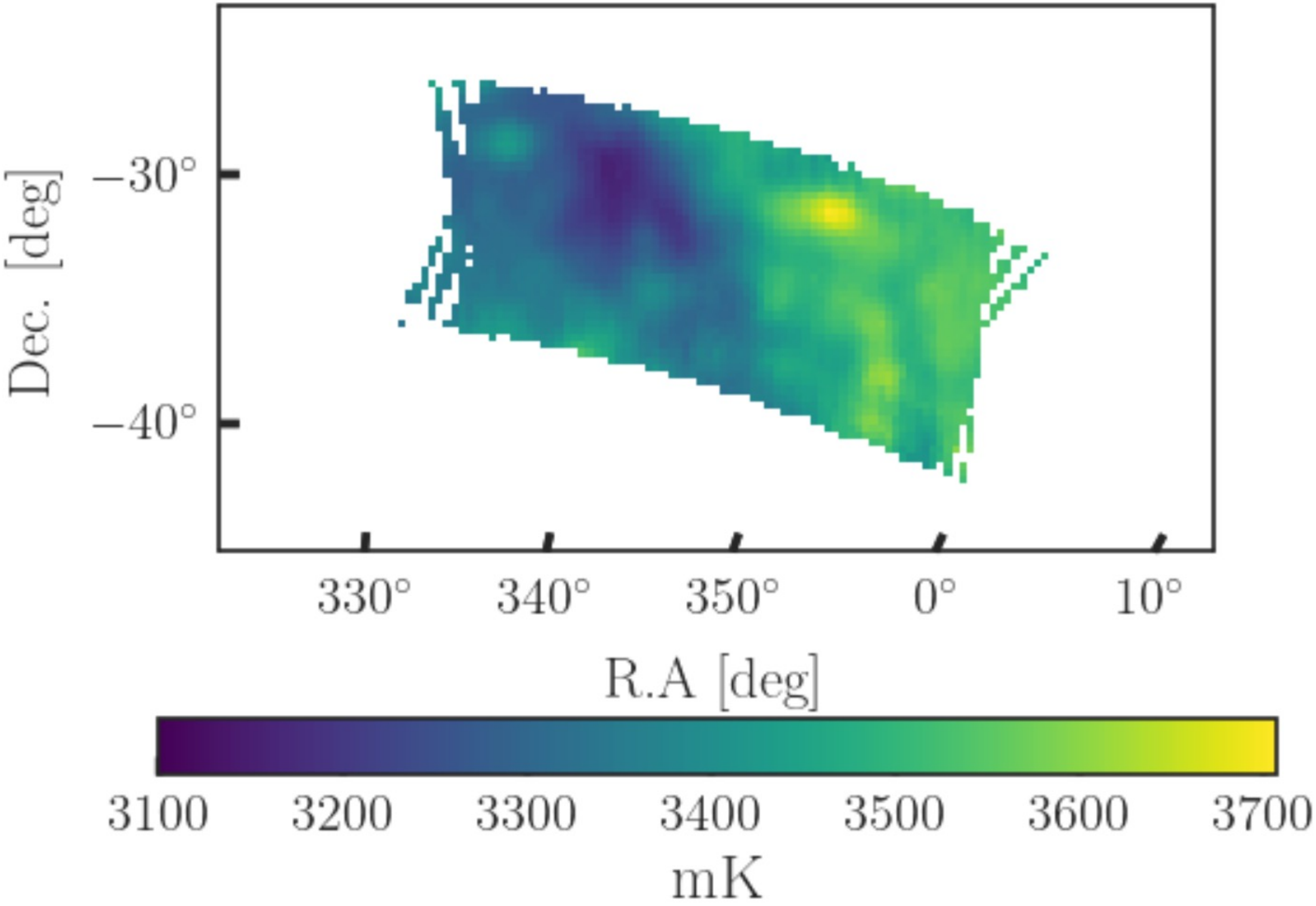
- Multi-tracer technique: $\sigma(f_{\mathrm{NL}}) < 1$
  - Constraints on the dynamics of the primordial Universe
  - Discrimination between inflationary models
- General improvement thanks to the multi-tracer technique: constraints more than halved in the global analysis
- Better constraining power at high redshift
  - Extend the galaxy sample
  - Push down the flux limit
- Importance of the modeling of foregrounds in the 21 cm signal
- Possible extensions
  - Constraining relativistic effects at large scales
  - Applying the multi-tracer techique with three tracers

A step before: MeerKAT

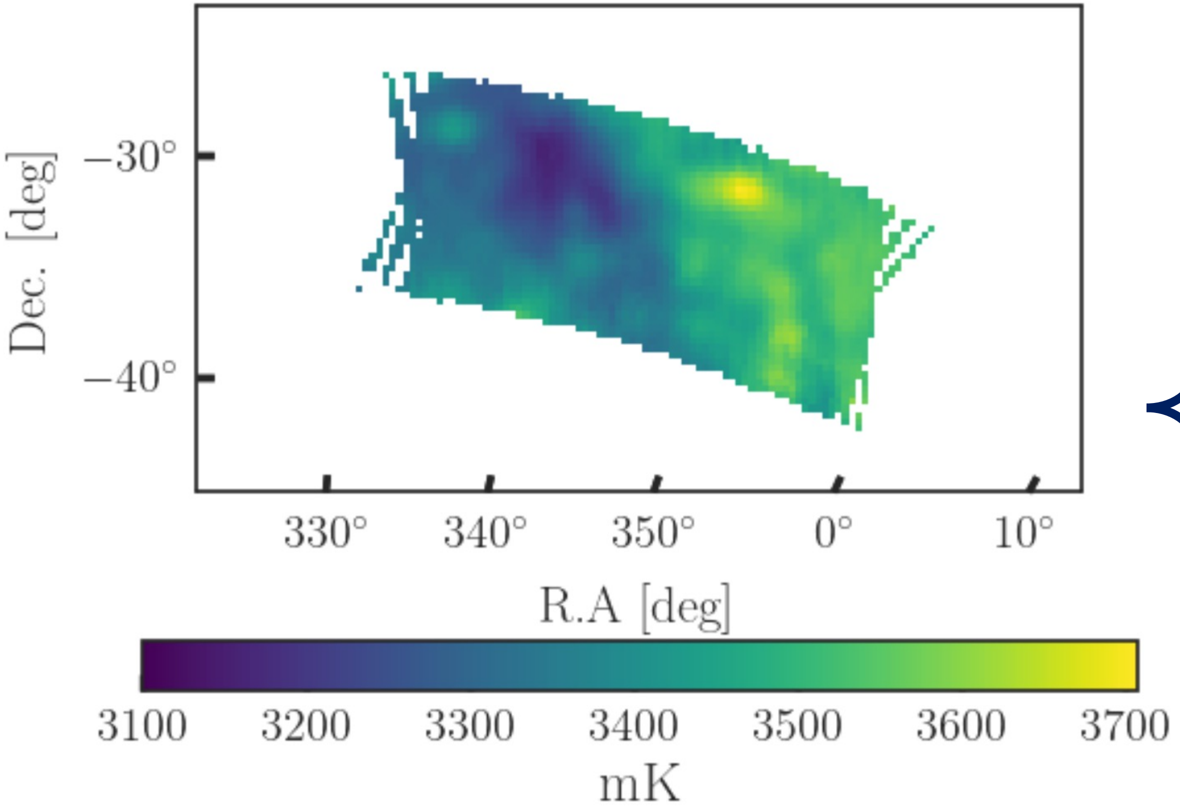
- South Africa, Karoo
- 64 dishes (13,5m diameter)
- Frequency range:
  - L-band: 580 - 1015 MHz
  - UHF band: 900 - 1670 MHz

- MeerKAT Large Area Synoptic Survey (MeerKLASS) [Santos et al. (2016)]
- SKAO precursor
- Operations started in 2018
- 2019 data (L-band): complete calibration pipeline [Wang et al. (2021), Li et al. (2021)], first detection of the power spectrum in cross-correlation [Cunnington et al. (2023)] and in auto-correlation [Paul et al. (2023)]
- 2021 data (L-band): ongoing analysis ←
- Now (UHF band): new observing season

The MeerKAT 2021 data set

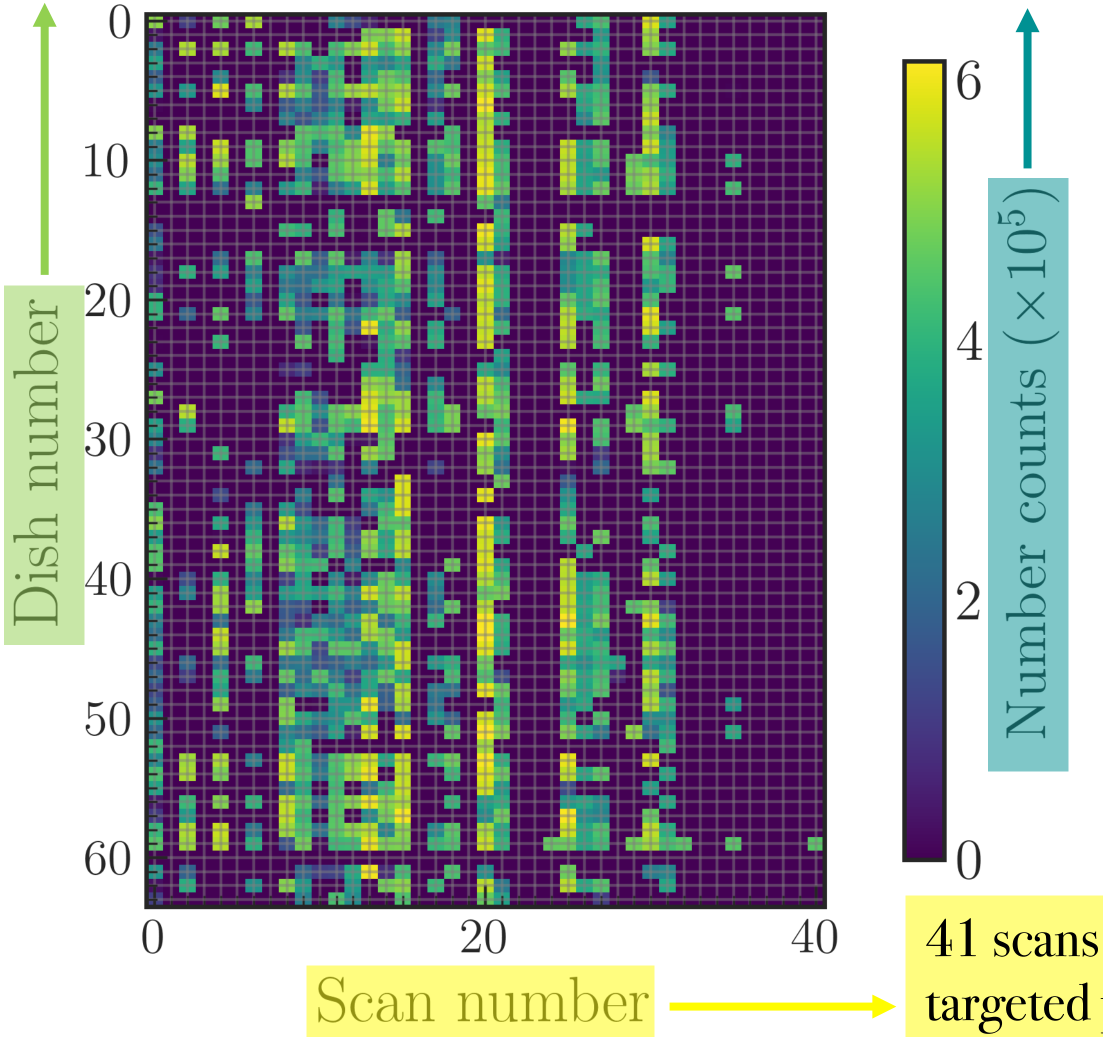


The MeerKAT 2021 data set



64 MeerKAT dishes

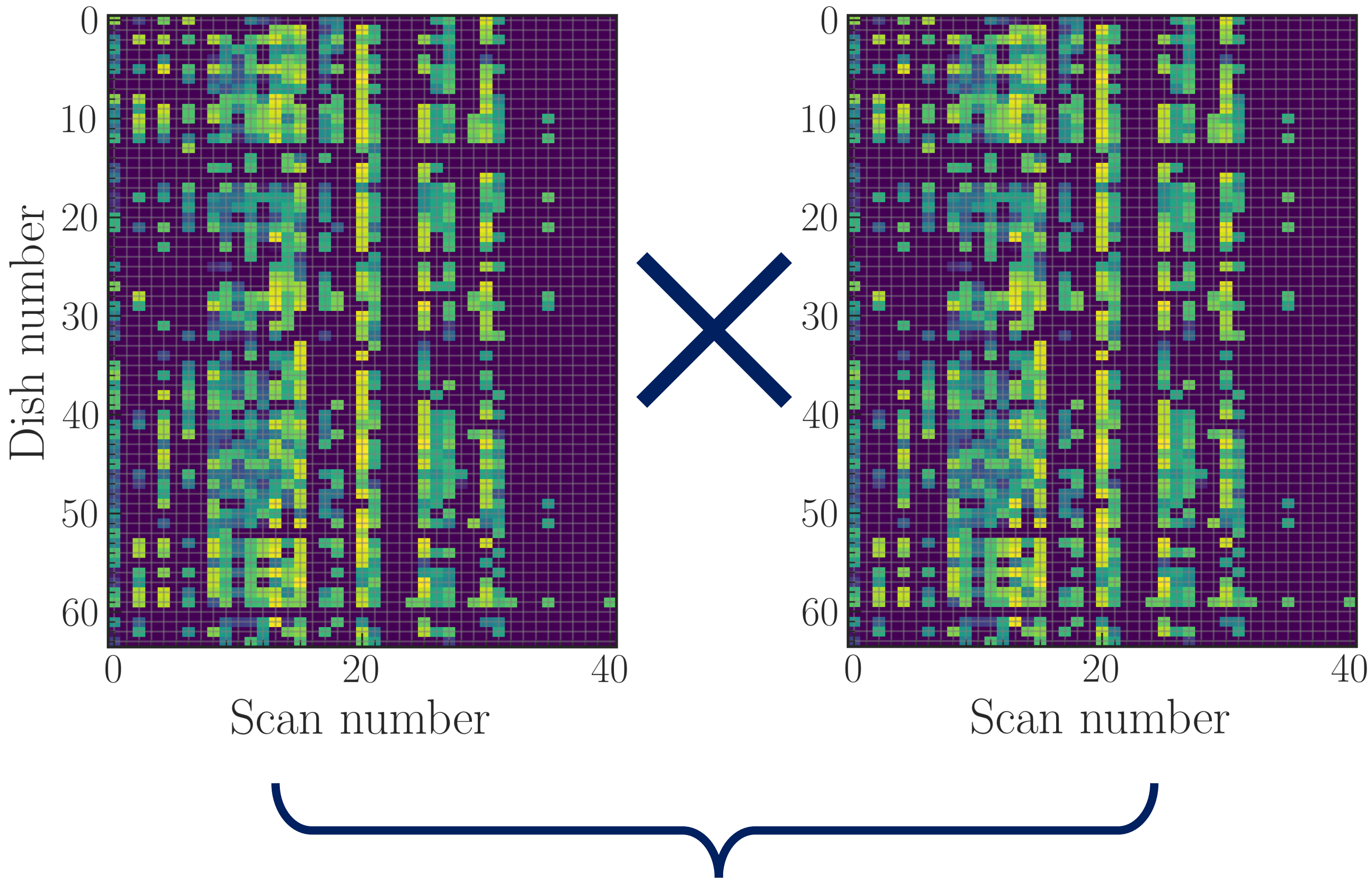
from hit map: ~amount of information



41 scans of the targeted patch

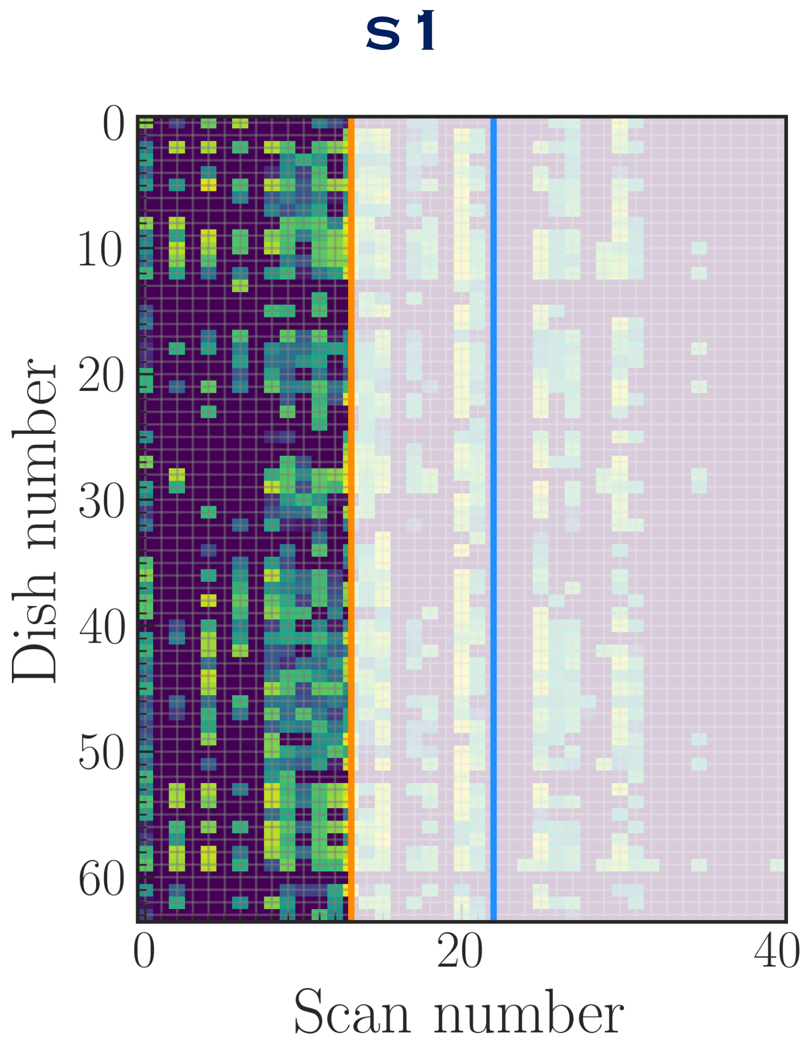
Full data set auto-correlation

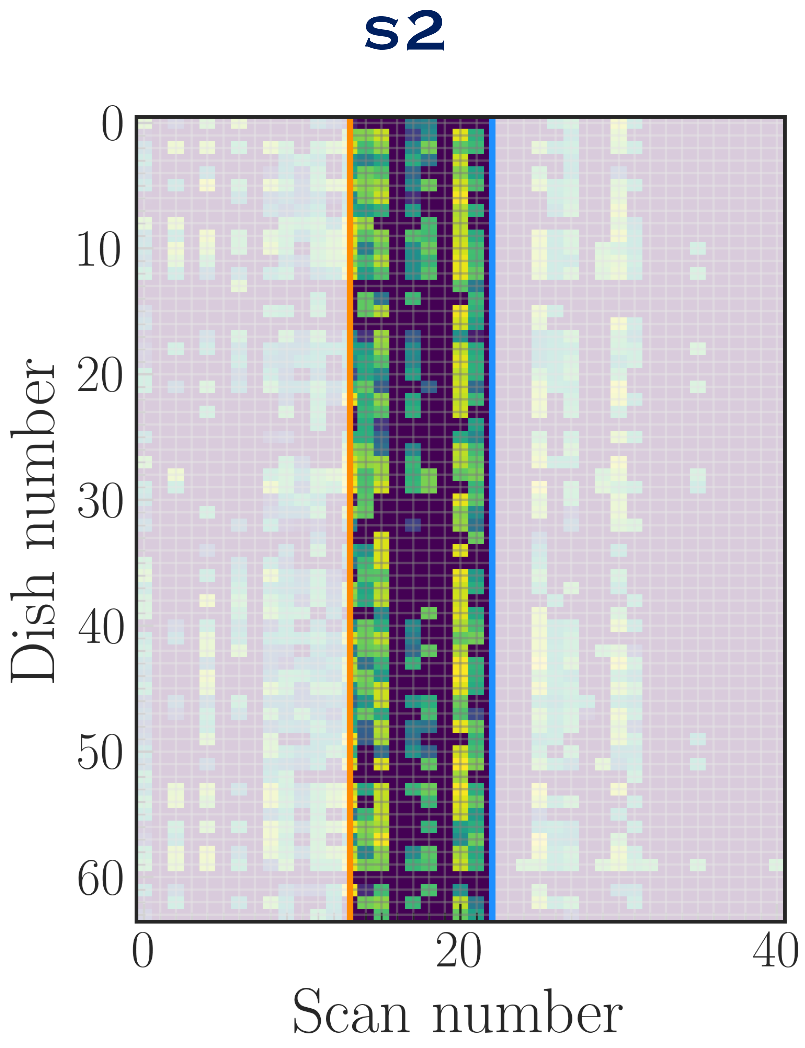
- HI cosmological signal
- Noise
  - Thermal noise
  - Shot noise
- Residuals foregrounds
- Contaminants

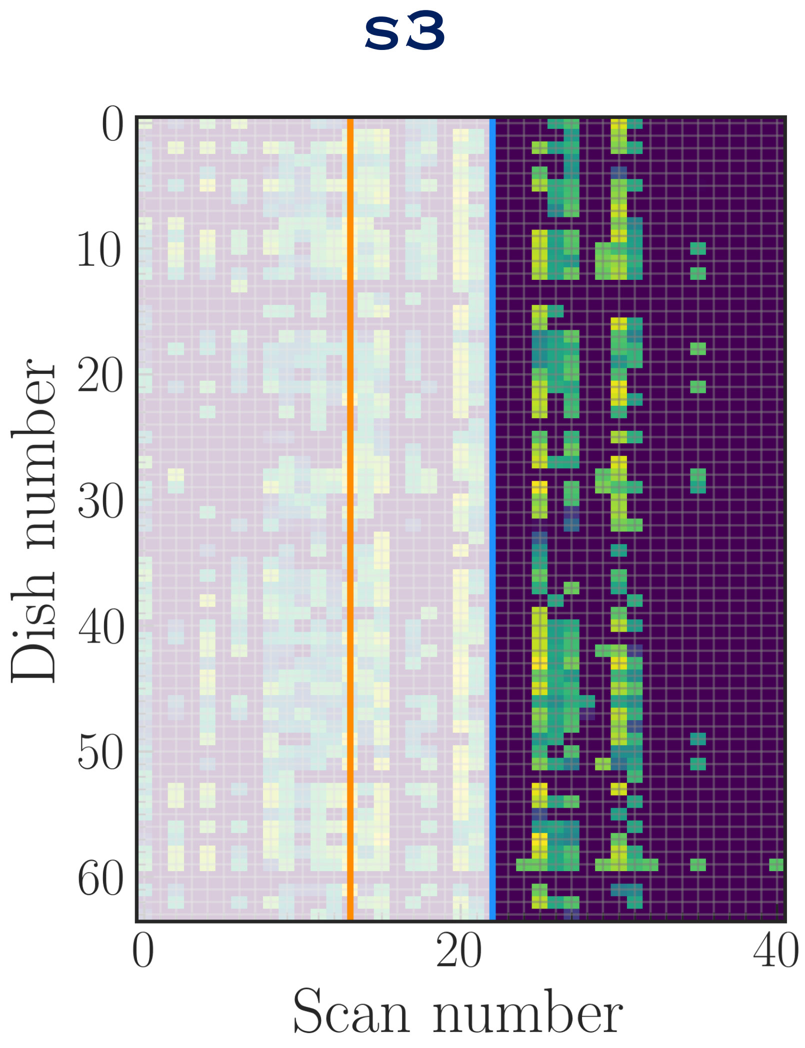


$P_{\mathrm{HIHI+noise}}$

Splitting the data set







Splitting the data set

- Independent foreground cleaning for each subset
  - PCA: $N_{\rm fg} \in [4,9]$
  - Multiscale PCA: $N_{\rm fg,L}, N_{\rm fg,S} \in [4,9]$
- Subset auto-correlation: $P_{\rm HIHI+noise}$
  - $\rm s1\times s1$, $\rm s2\times s2$, $\rm s3\times s3$
- Subset cross-correlation: $P_{\rm HIHI}$
  - $\rm s1\times s2$, $\rm s1\times s3$, $\rm s2\times s3$

- Goals:
  - Detecting the HI power spectrum without the cross-correlation with galaxies
  - Constraining $\Omega_{\rm HI} b_{\rm HI}$
  - Constraining noise terms

Very preliminary results

- mPCA seems to perform better than simple PCA (better fits)
- Consistent noise estimates
- The HI power spectrum is there (null tests being performed)
  - $\Omega_{\rm HI} b_{\rm HI}$ coherent with the results from the cross-correlation with WiggleZ galaxies (2019 MeerKAT data set) [Cunnington (2023)]
  - $\Omega_{\rm HI} b_{\rm HI}$ coherent between themselves

Conclusions

MeerKAT's progress leading to…

- Increased data quality
- Refined data reduction and analysis techniques
- Possible detection of the HI power spectrum

… the realisation of SKAO's full potential

- Constrainig power on very large scales
- Constraining power on the dynamics of the Early Universe: $\sigma(f_{\rm NL}) = 2.3$
- Multi-tracer applications: $\sigma(f_{\rm NL}) = 0.76$