Multi-tracing the Large Scale Structure of the Universe

Forecasts for radio-optical synergies and the state of the art of radio surveys

Matilde Barberi Squarotti

Cerimonia Premio "Milla Baldo Ceolin" – October 8th 2024, GGI

INFN

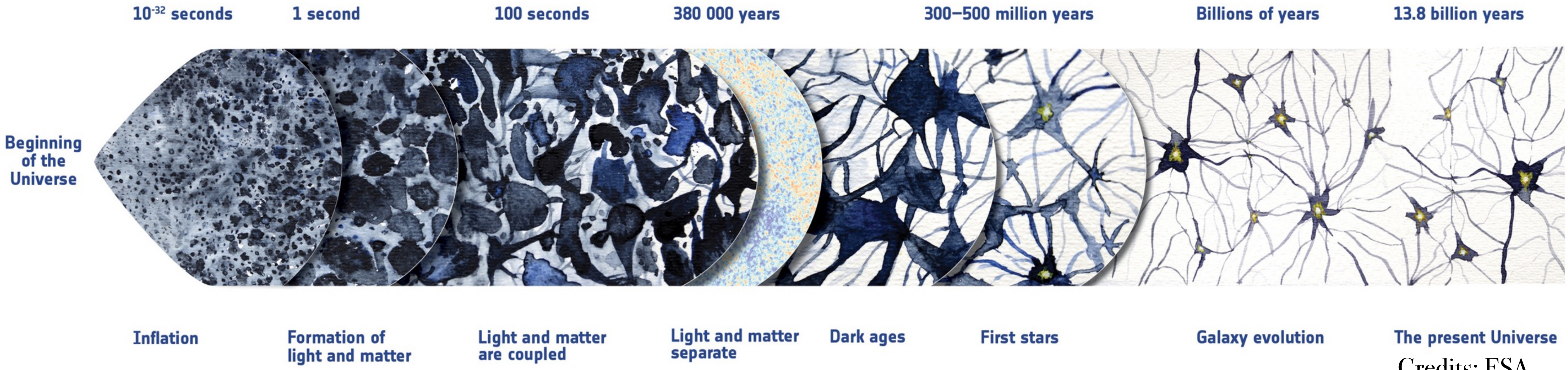
INAF ISTITUTO NAZIONALE DI ASTROFISICA

euclid consortium

SKAO

ΛCDM

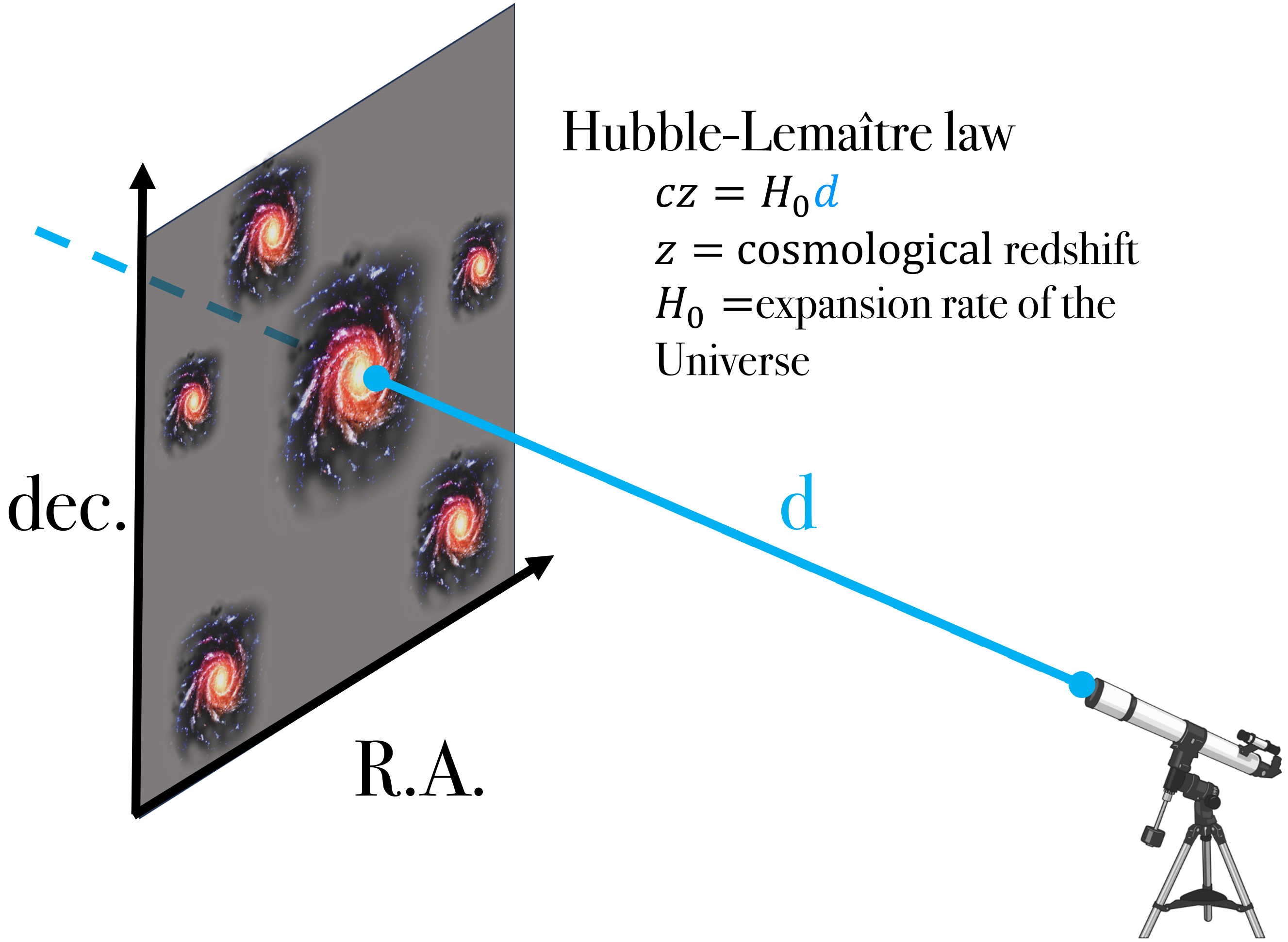
- Evolution of the Universe from primordial fluctuations up to the cosmic web
  - Λ: cosmological constant / dark energy
  - CDM: cold dark matter
  - Inflation



Credits: ESA

The large scale structure

- Matter spatial distribution determined by the cosmological parameters
- Statistical properties of dark matter tracers to constrain the cosmological parameters
  - Cosmological surveys of the three dimensional distribution of dark matter tracers



Hubble-Lemaître law

$cz = H_0 d$

z = cosmological redshift

H_0 =expansion rate of the Universe

Power spectrum

- From dark matter…
  - Density contrast

$$\delta_{\mathrm{m}}(\boldsymbol{x}, z) = \frac{\rho(\boldsymbol{x}, z) - \bar{\rho}(z)}{\bar{\rho}(z)} \rightarrow \tilde{\delta}_{\mathrm{m}}(\boldsymbol{k}, z)$$

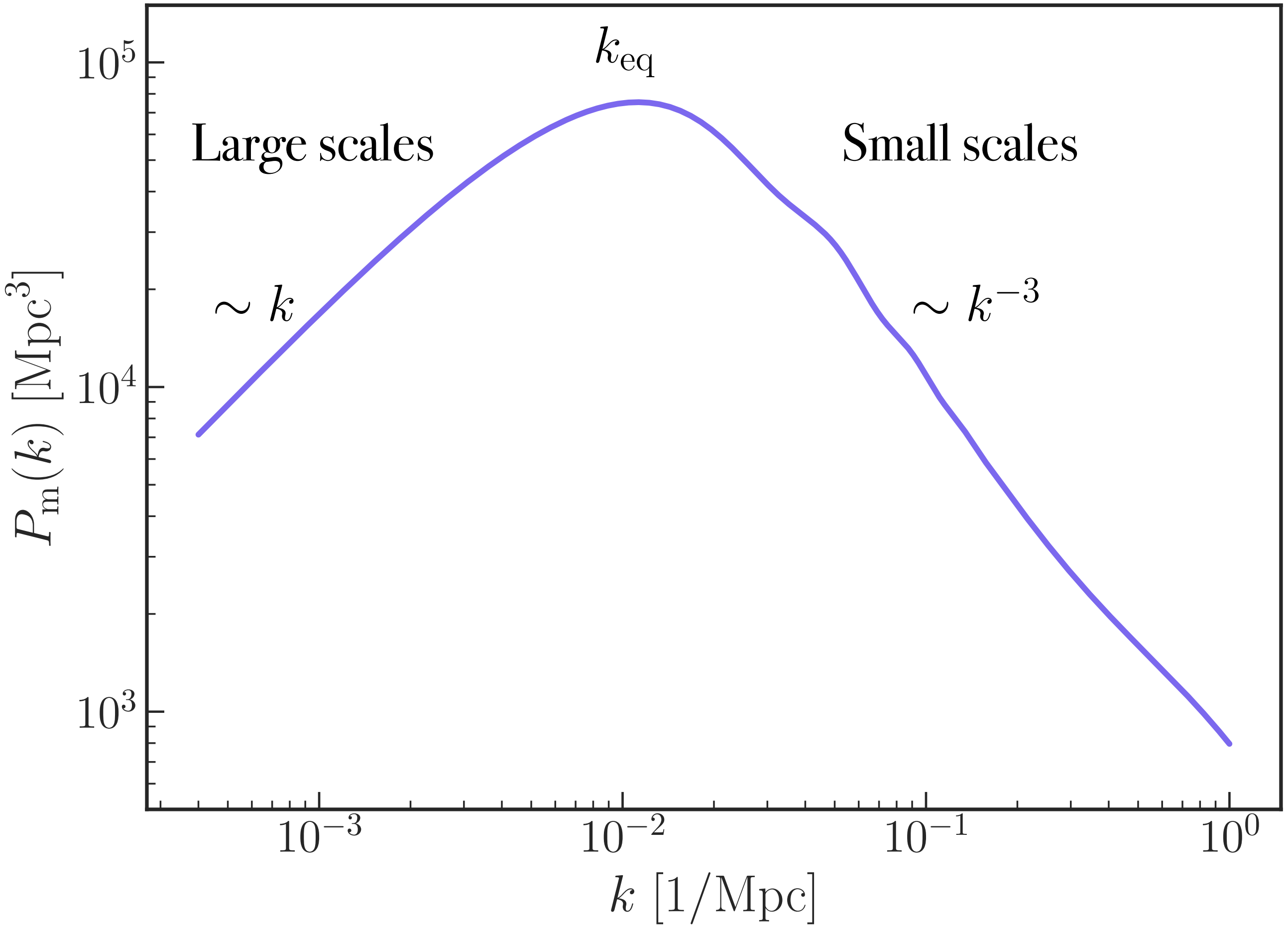
  - Power spectrum

$$P_{\mathrm{m}}(k, z) = \langle \tilde{\delta}_{\mathrm{m}}(k, z) \tilde{\delta}^{*}_{\mathrm{m}}(k, z) \rangle$$

- … to its tracers

$$\delta_t(\boldsymbol{x}, z) = b_t(z) \delta_{\mathrm{m}}(\boldsymbol{x}, z)$$

$$P_t(k, z, \mu) = [b_t(z) + f(z)\mu^2]^2 \, P_{\mathrm{m}}(k, z)$$



Matter tracers

Galaxies

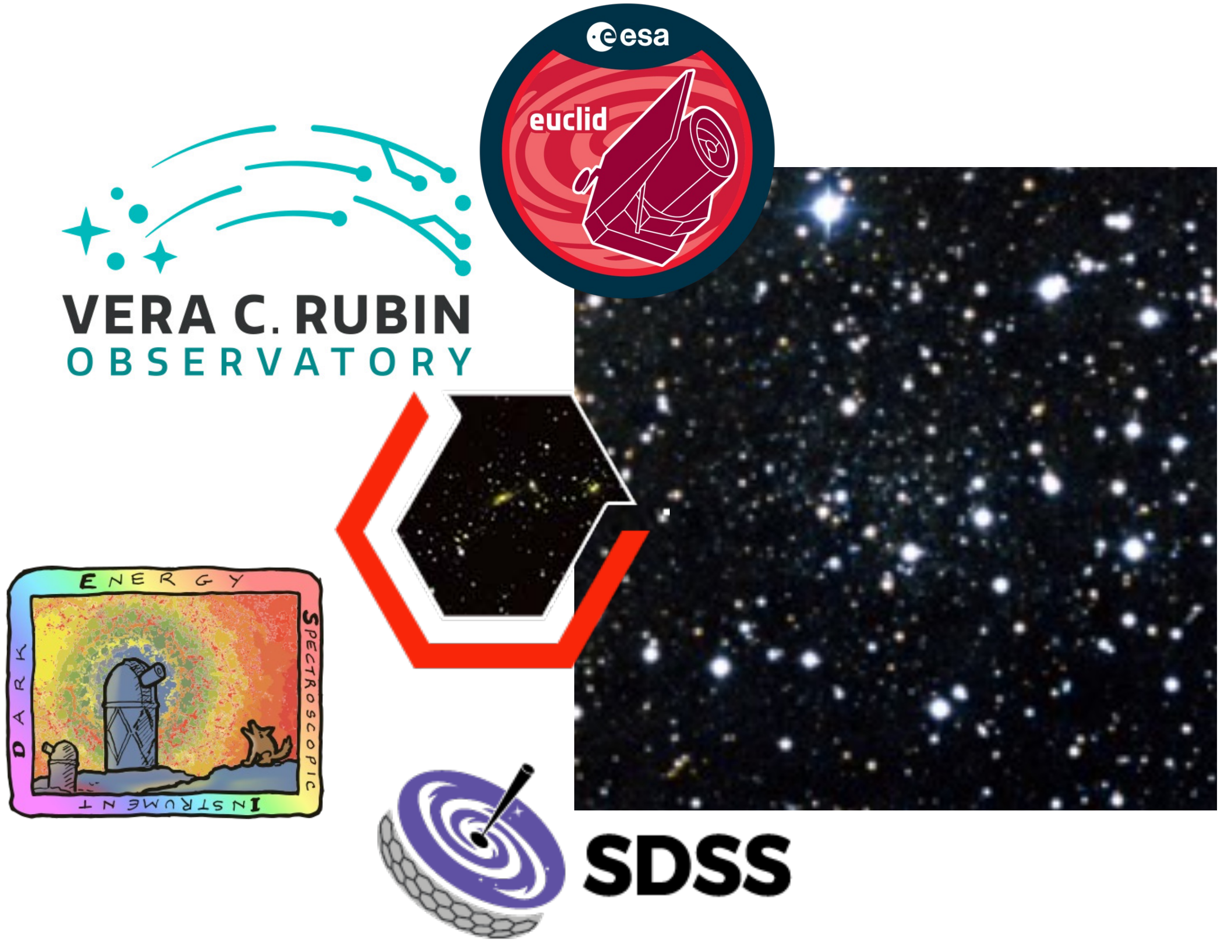
- Emission in caracteristic spectral lines
- Spectroscopic/photometric surveys (optical and NIR bands): single sources resolved

Matter tracers

Galaxies

Matter tracers

Galaxies



Neutral Hydrogen (HI)

- Emission at 21cm from the hyperfine spin-flip transition of neutral hydrogen (HI)
- Intensity mapping (IM) in the radio band: low angular of the total flux from unresolved sources

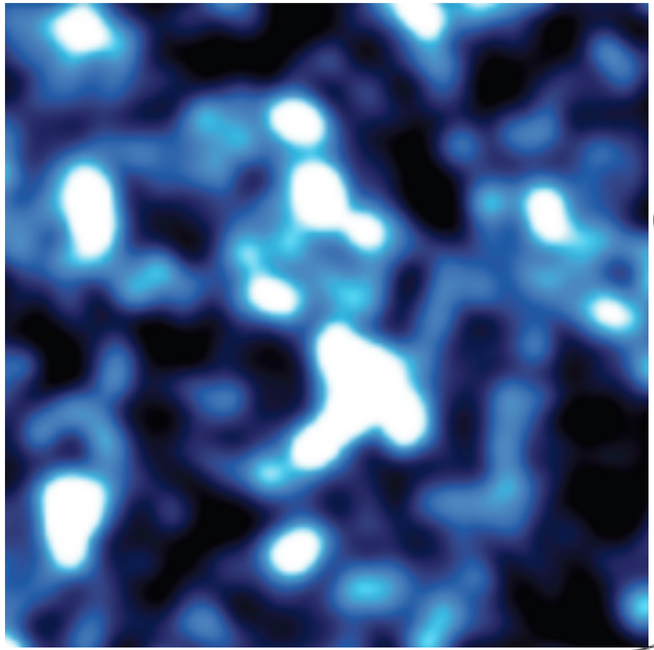
Matter tracers

Galaxies

esa
euclid
VERA C. RUBIN OBSERVATORY
DARK ENERGY SPECTROSCOPIC INSTRUMENT
SDSS

Neutral Hydrogen (HI)

SKAO
Hydrogen Epoch of Reionization Array
HERA
The Search for COSMIC DAWN
HIRAX
GREEN BANK OBSERVATORY
chime



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

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

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

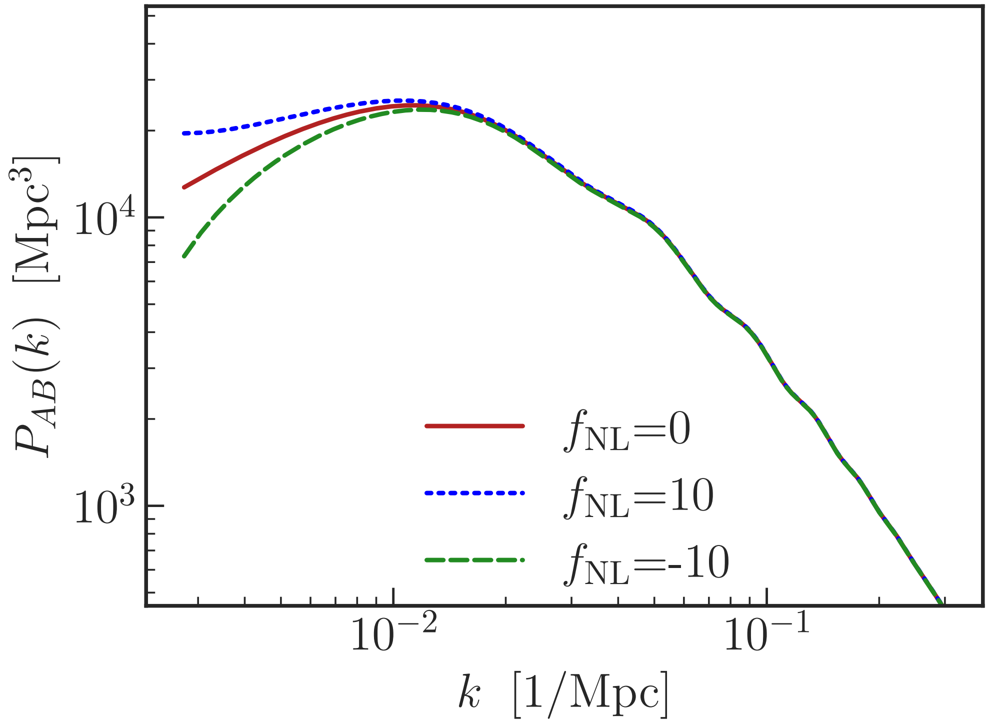
$$\mathrm{Cov}(\boldsymbol{P}) = \frac{2}{N_{\mathrm{m}}} \begin{bmatrix} \tilde{P}_{\mathrm{gg}}^2 & \tilde{P}_{\mathrm{gg}}\tilde{P}_{\mathrm{gH\,I}} & \tilde{P}_{\mathrm{gH\,I}}^2 \\ \tilde{P}_{\mathrm{gg}}\tilde{P}_{\mathrm{gH\,I}} & \frac{1}{2}\left(\tilde{P}_{\mathrm{gg}}\tilde{P}_{\mathrm{H\,IH\,I}}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

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

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 [Planck Collaboration: (2018)]
  - $|f_{\mathrm{NL}}| < 5$
  - $\sigma(f_{\mathrm{NL}}) \sim 5$
- Primordial non-Gaussianity and the Large Scale Structure
  - $b_A(z) \rightarrow b_A(z) + \Delta b_A(k,z) f_{\mathrm{NL}}$

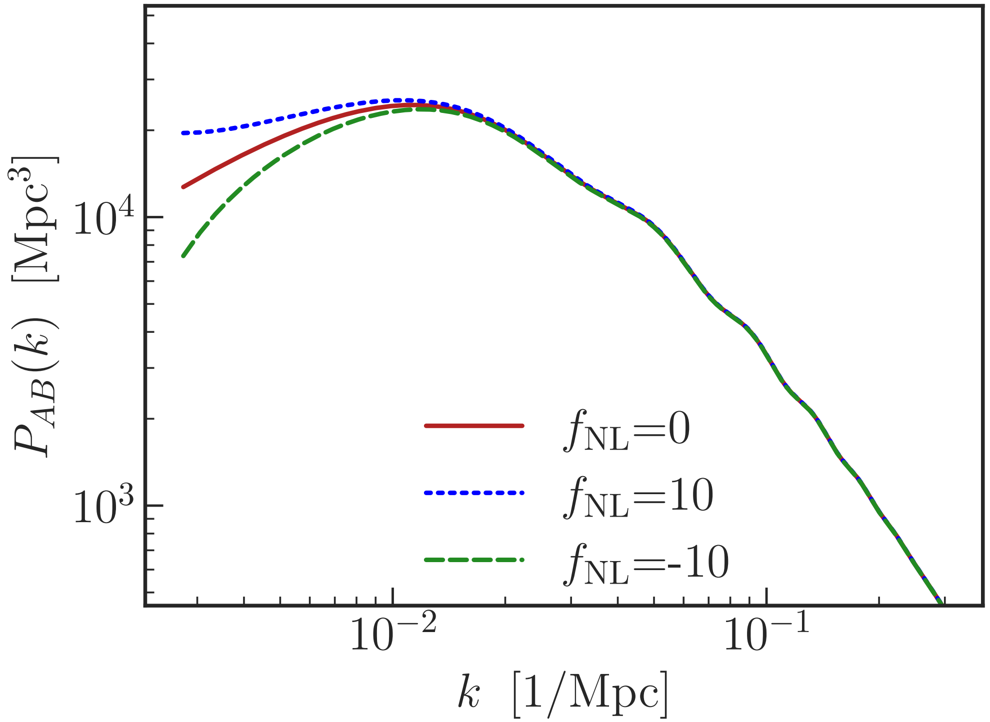
$$\Delta b_A(k,z) = 3\,[b_A(z) - 1]\,\frac{\delta_c \Omega_{\mathrm{m},0} H_0^2}{c^2 k^2 T(k) D(z)}$$



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 [Planck Collaboration: (2018)]
  - $|f_{\mathrm{NL}}| < 5$
  - $\sigma(f_{\mathrm{NL}}) \sim 5$
- Primordial non-Gaussianity and the Large Scale Structure
  - $b_A(z) \;\rightarrow b_A(z) + \Delta b_A(k,z) f_{\mathrm{NL}}$

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



Analysis

- HI and galaxy synthetic data vectors (auto and cross correlations)
  - Stage IV surveys
  - Including systematics

} synthetic multi-tracer data vector

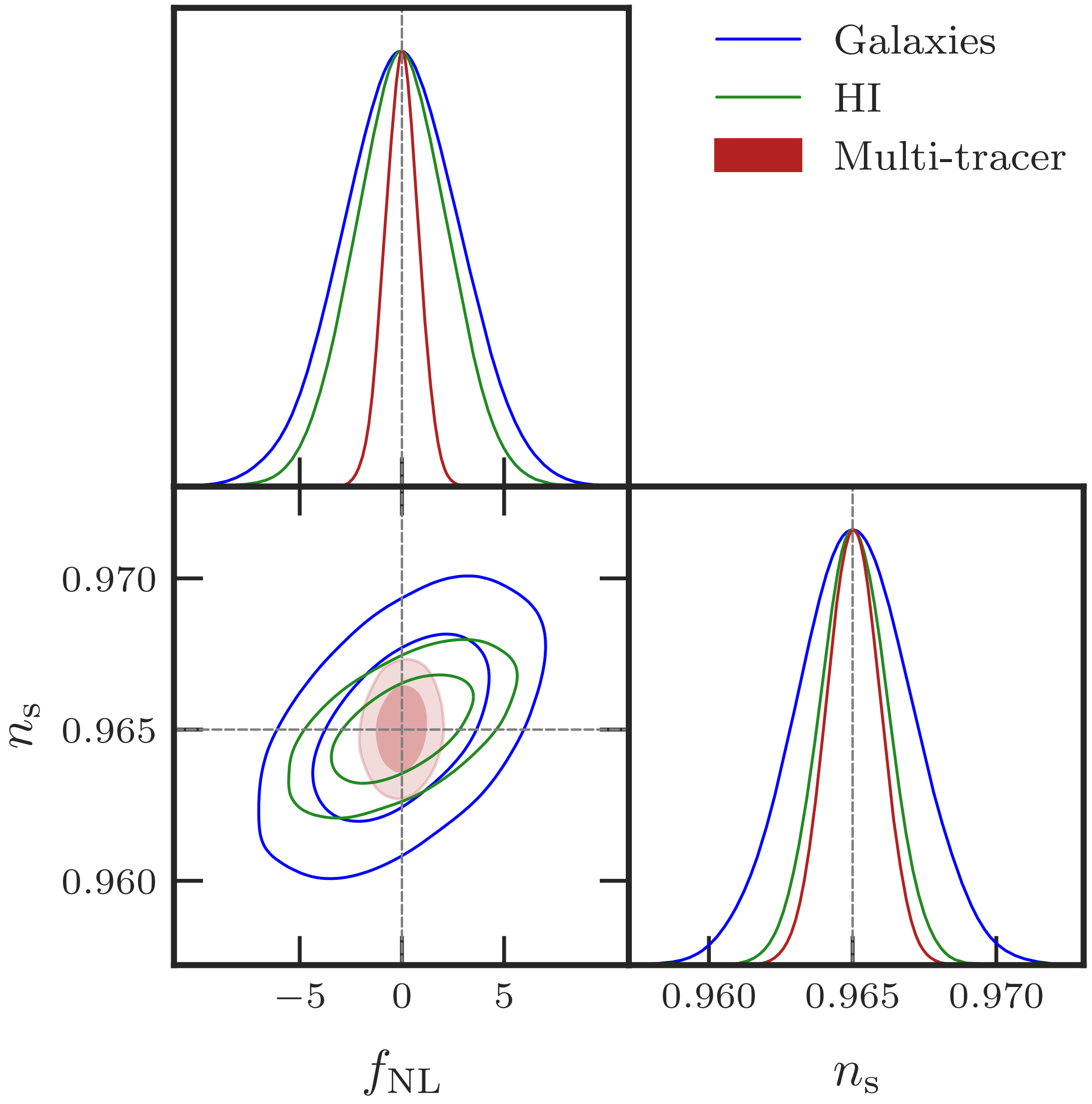
- Multivariate analysis: maximization of the likelihood with MCMC method

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

  - $\boldsymbol{P}(\boldsymbol{\theta}) =$ theoretical multi-tracer data vector
  - $\boldsymbol{\theta} = \{f_{\mathrm{NL}}, n_{\mathrm{s}}, b_{\mathrm{g}}(z), b_{\mathrm{HI}}(z)\}$
- Analysis of the constraints, focusing on f_{NL}

Results

- All fiducial parameters recovered and constrained
- Marginalized uncertainty on f_{NL}:
  - Galaxies auto-correlation: $f_{NL} = 0.0 \pm 2.8$
  - HI auto-correlation: $f_{NL} = 0.0 \pm 2.3$
  - Multi-tracer techique: $f_{NL} = 0.0 \pm 0.76$
- Multi-tracer technique: tightest constraints



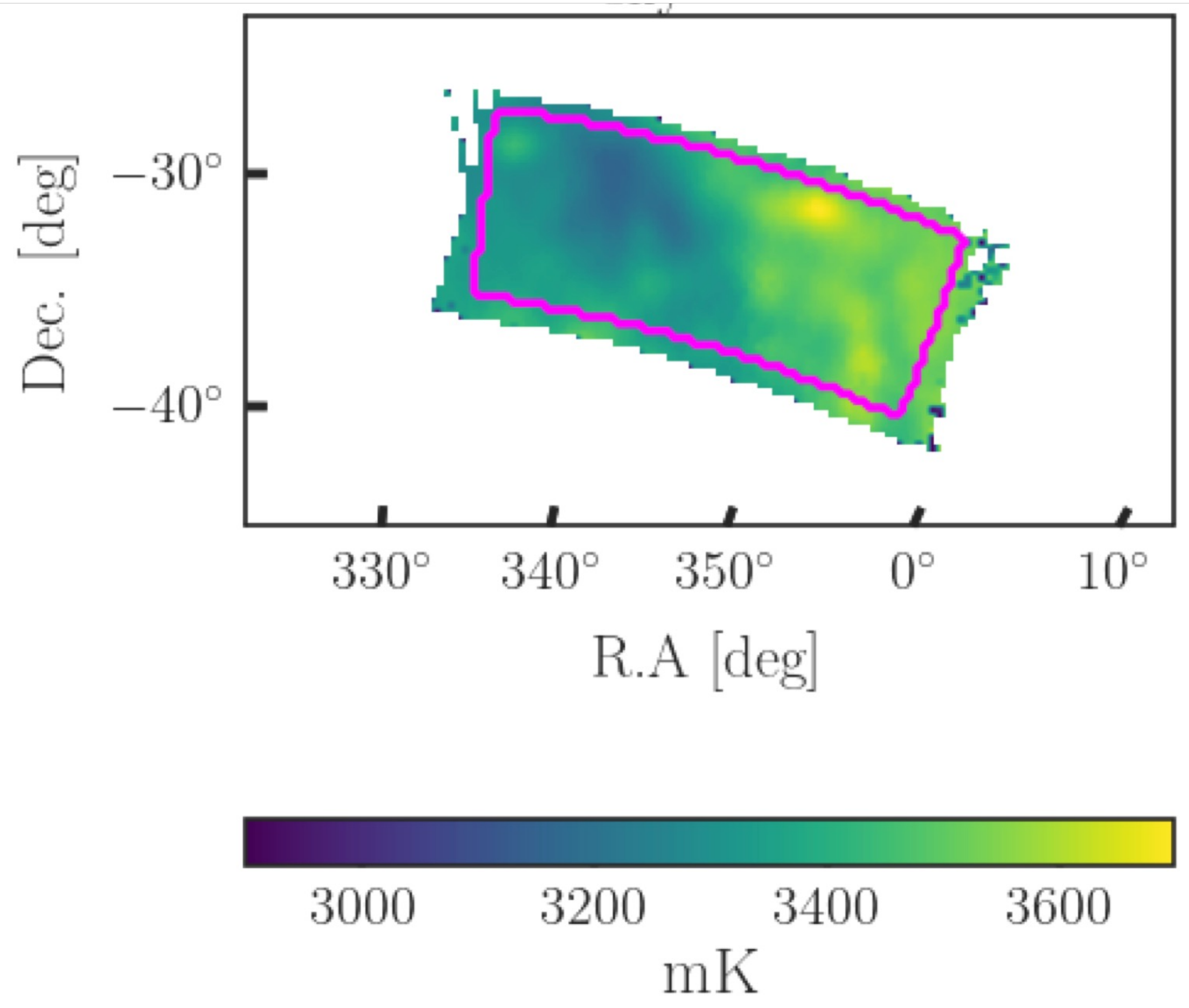
[**MBS**, Camera, Maartens (2023)]

HI IM WITH MEERKAT

- SKAO precursor
- Sud Africa, Karoo
- 64 antennas (diameter:13,5m)
- Frequency bands:
  - L-band: $580 < \nu < 1015$ MHz
  - UHF band: $900 < \nu < 1970$ MHz

- MeerKAT Large Area Synoptic Survey (MeerKLASS) [Santos et al. (2016)]
- 2019 pilot survey (L-band):
  - Validation of the calibration pipeline [Wang et al. (2021), Li et al. (2021)]
  - First detection in cross-correlation with galaxies [Cunnington et al. (2023)]
- 2021 survey (L-band):
  - Re-detection of the cross-correlation [Cunnington, **MBS** et al. (2023)]
  - Auto-power spectrum being extracted [**MBS** et al. (TBS)]
- 2023-2028 (UHF-band): new observing season

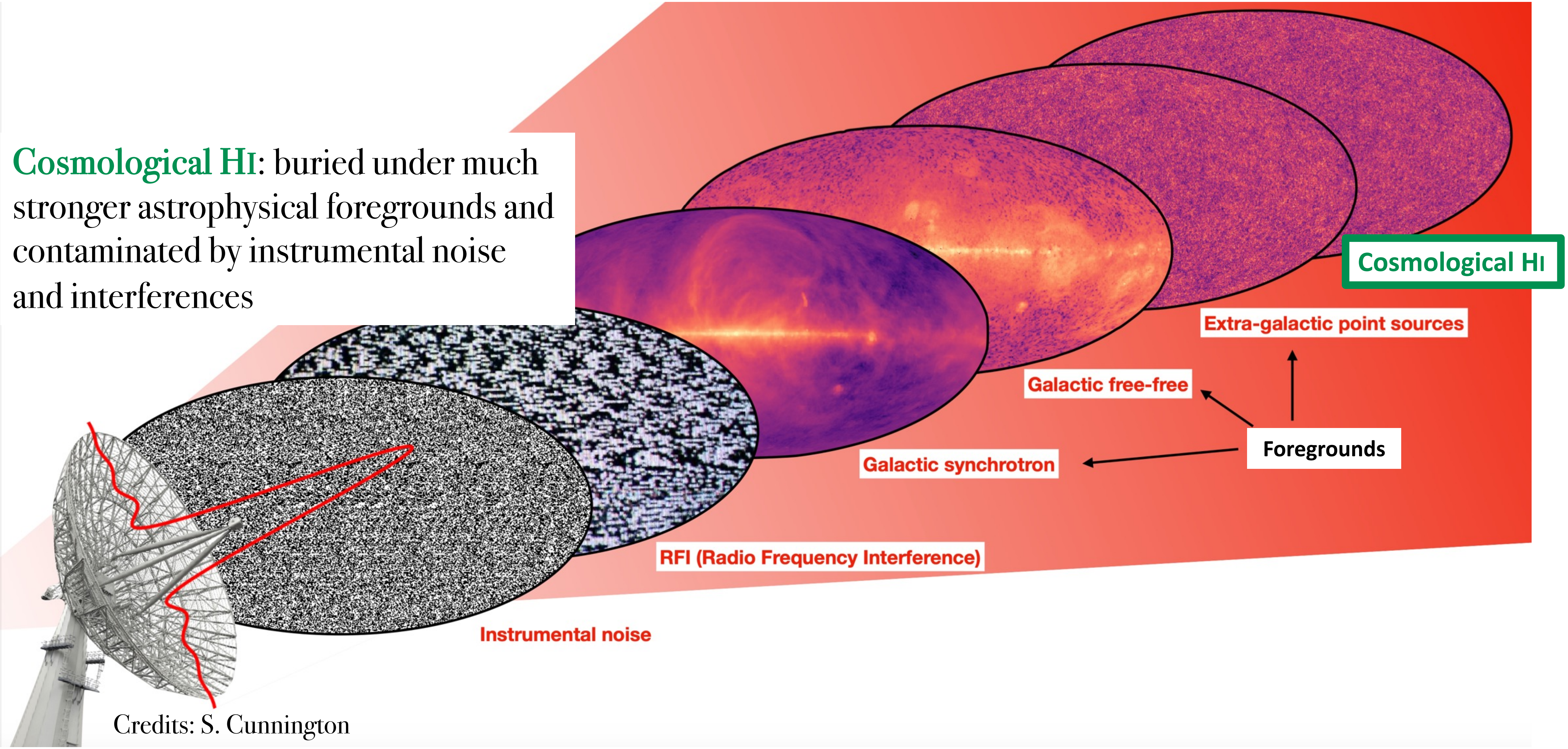
HI IM WITH MEERKAT



- MeerKAT Large Area Synoptic Survey (MeerKLASS) [Santos et al. (2016)]
- 2019 pilot survey (L-band):
  - Validation of the calibration pipeline [Wang et al. (2021), Li et al. (2021)]
  - First detection in cross-correlation with galaxies [Cunnington et al. (2023)]
- 2021 survey (L-band): deepest HI survey
  - Re-detection of the cross-correlation [Cunnington, MBS et al. (2023)]
  - Auto-power spectrum being extracted [MBS et al. (TBS)]
- 2023-2028 (UHF-band): new observing season

Why is HI intensity mapping hard?

Cosmological HI: buried under much stronger astrophysical foregrounds and contaminated by instrumental noise and interferences

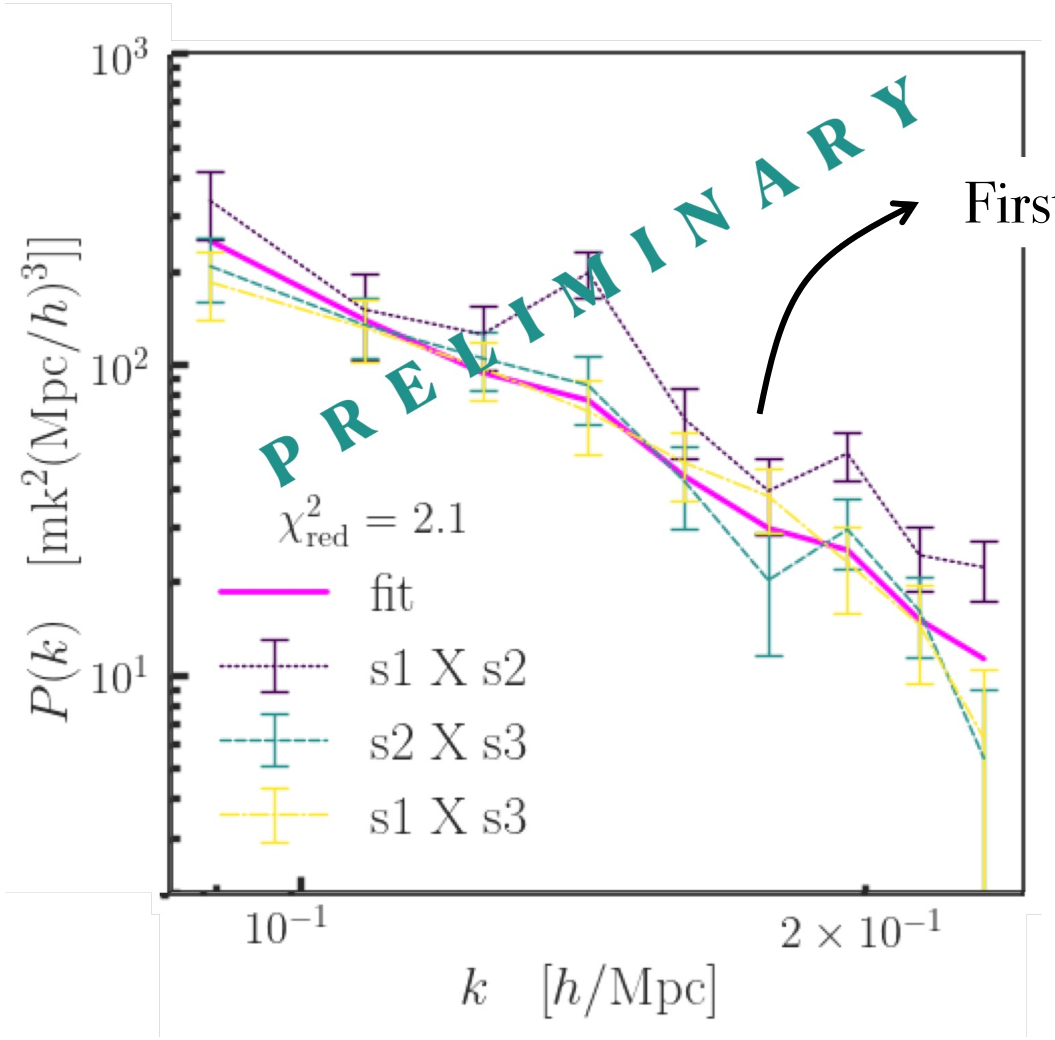


Credits: S. Cunnington

THE HI AUTO POWER SPECTRUM

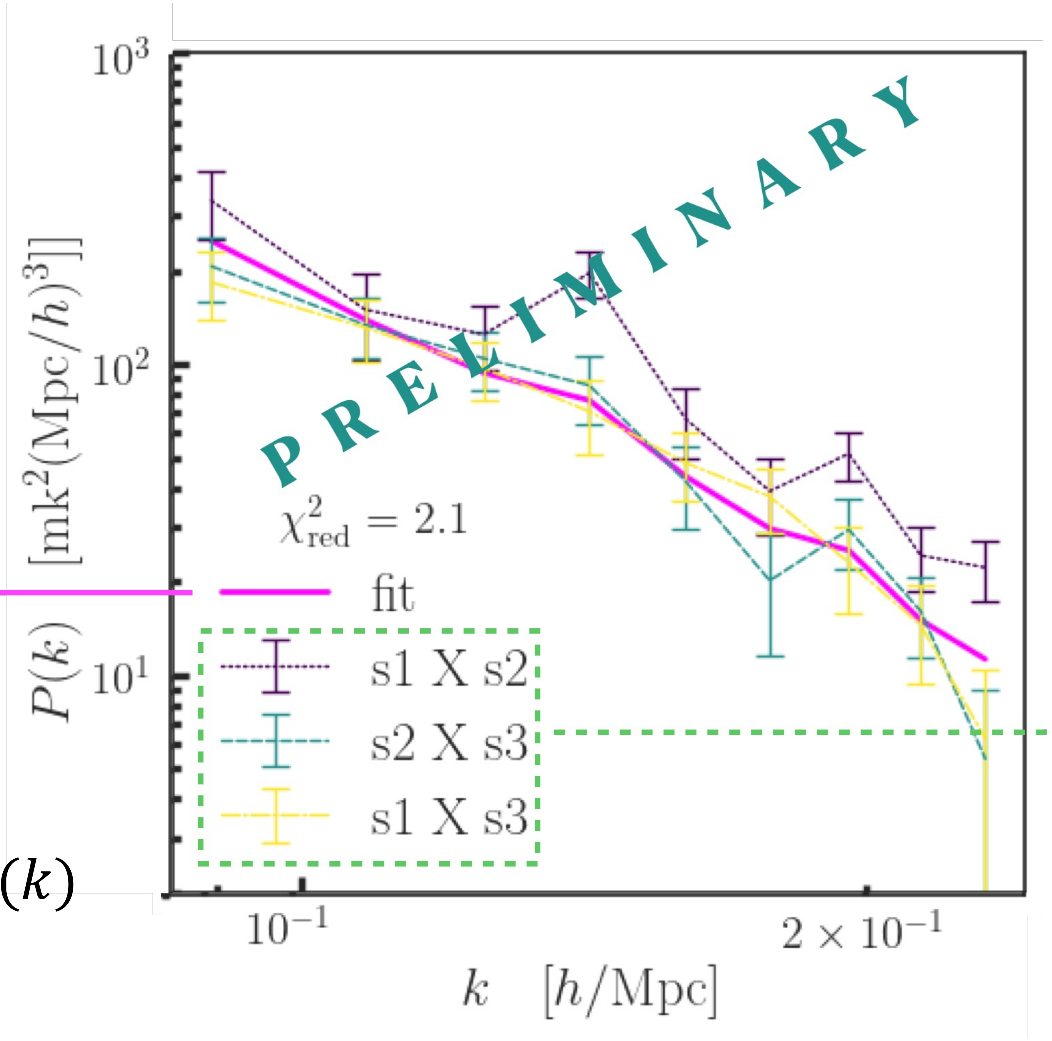
- Internal cross-correlations: building independent subsets from the same survey [Wolz et al. (2021)]
  - The HI cosmological signal is the only correlated part
  - Contaminants not correlated between subsets
  - Noise free cross-subset power spectra

Some preliminary results



First ever HI auto-power spectra

Some preliminary results



One only common fit-curve

Stripe division: three cross-$P(k)$ measured

$$P_{i\times j} \propto \boldsymbol{T^2_{\rm HI} b^2_{\rm HI}} (1 + f\mu^2)^2 P_{\rm m}(k)$$

Conclusions

- The MeerKLASS 2021 survey in the L-band, being the deepest single-dish survey, carry a relaevant amout of information
- Internal cross-correlations:
  - Help in mitigating systematics that affect HI intensity mapping observation
  - Possibility to detect the power spectrum without the need of a cross-correlation with a spectroscopic galaxy survey
- Open road for synergies between surveys, exploiting the complementarities of different tracers

Back-up slides

Power spectrum: Tracers

$$P_{AB}(k,z,\mu) = [b_A(z) + f(z)\mu^2][b_B(z) + f(z)\mu^2]P_{\mathrm{m}}(k,z)$$

Power spectrum: Tracers

$$P_{AB}(k,z,\mu) = [b_A(z) + f(z)\mu^2][b_B(z) + f(z)\mu^2]P_{\mathrm{m}}(k,z)$$

Tracers

$A = B \rightarrow$ auto-correlation

$A \neq B \rightarrow$ cross-correlation

Power spectrum: Tracers

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

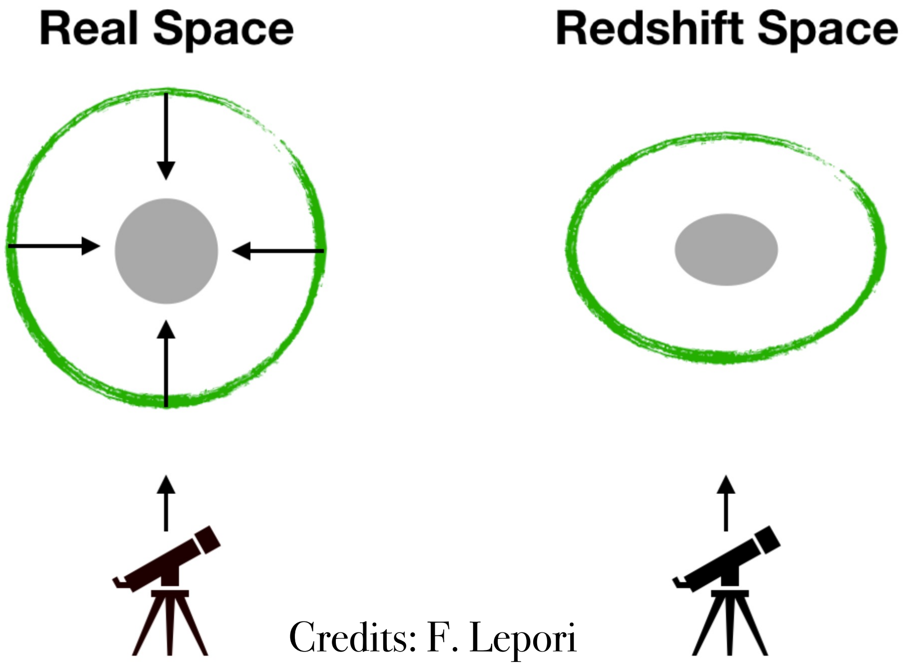
Linear bias

$$\delta_{A,B}(\boldsymbol{x},z) = b_{A,B}(z)\delta_{\rm m}(\boldsymbol{x},z)$$

Power spectrum: Tracers

$$P_{AB}(k,z,\mu) = [b_A(z) + f(z)\mu^2][b_B(z) + f(z)\mu^2]P_\mathrm{m}(k,z)$$

Redshift Space Distortions



Credits: F. Lepori

General setup

HI IM survey

- SKAO-like survey
- Sky coverage: $f_{\text{sky}} = 0.48$
- Thermal noise
- Beam damping
- Foreground avoidance

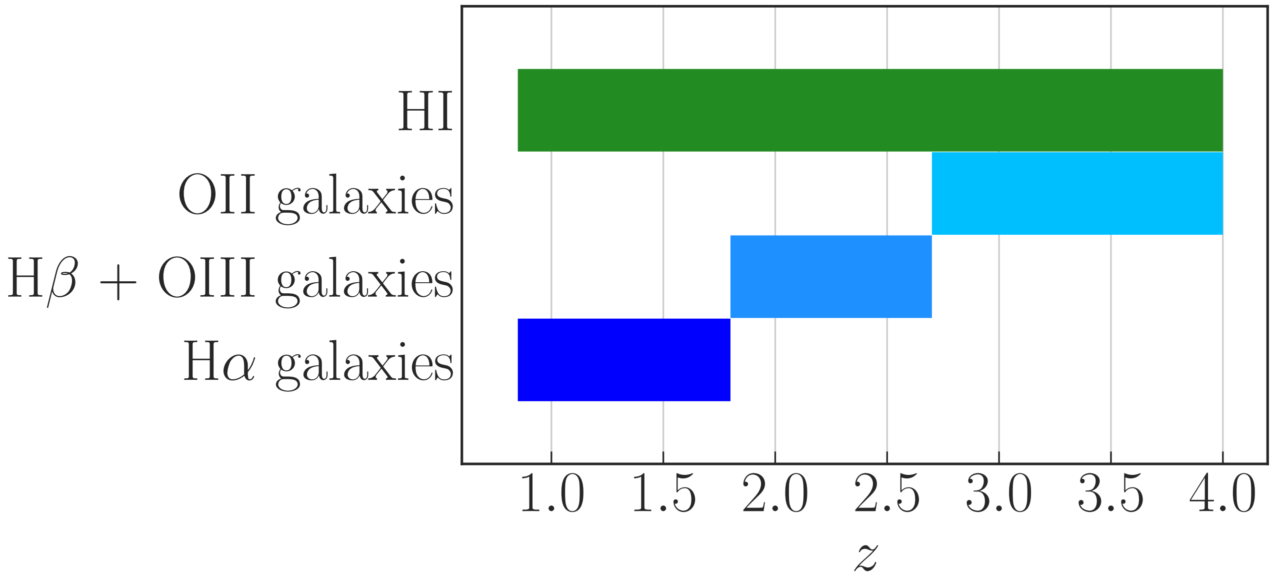
Galaxy survey

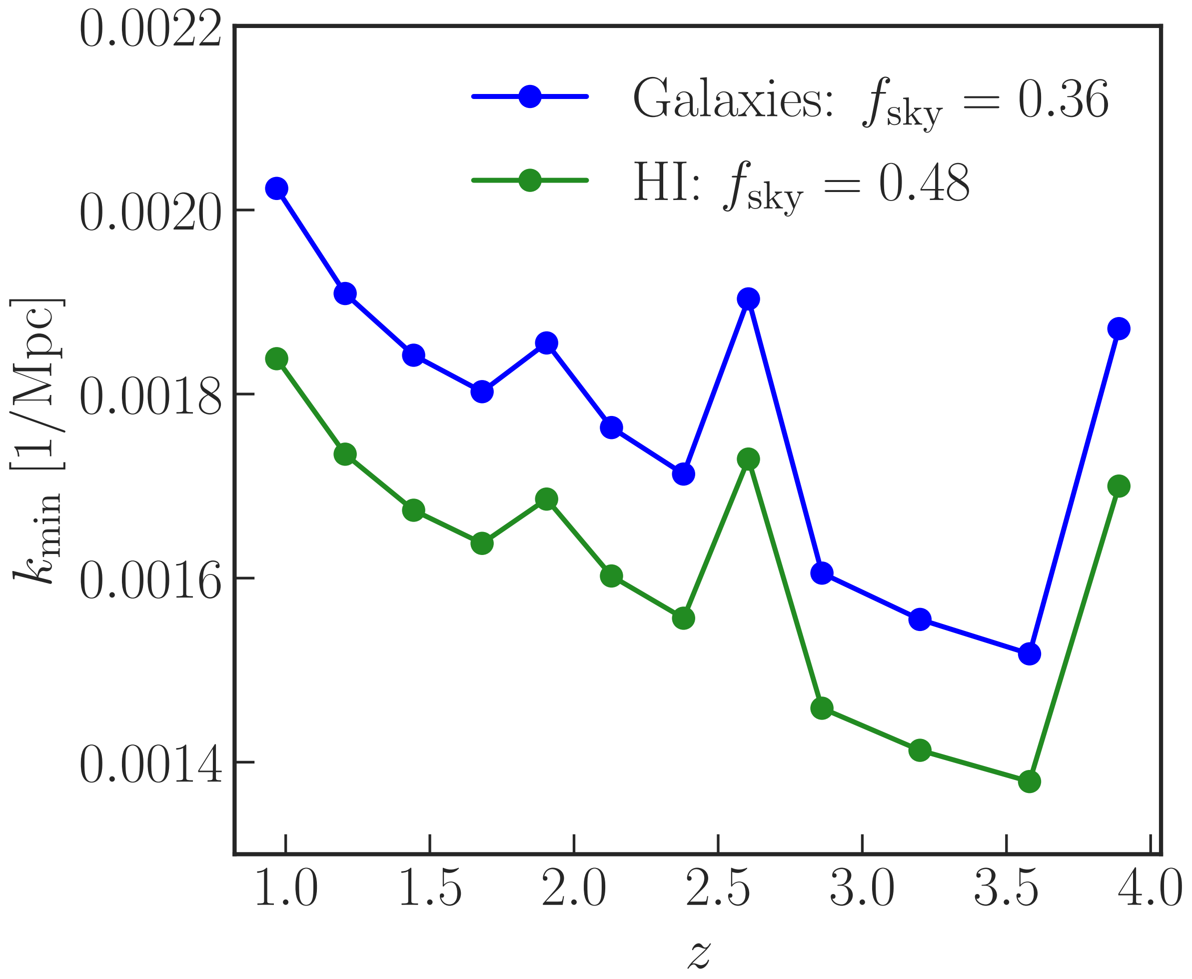
- Stage-IV spectroscopic survey
- Redshift range covered with different ELG types [Fonseca, Camera (2020)]
- Flux limit: $F_{\text{c}} = 2 \cdot 10^{-16} \text{erg s}^{-1}\ \text{cm}^{-2}$
- Sky coverage: $f_{\text{sky}} = 0.36$
- Shot noise

- 12 redshift bins for $z \in [0.85, 4]$

- 10 bins for $k \in [k_{\text{min}}(z), k_{\text{max}}(z)]$
  - $k_{\text{min}}(z) = \frac{2\pi}{V^{1/3}(z)}$
  - $k_{\text{max}}(z) = 0.08(1+z)^{2/(2+n_{\text{s}})} h\text{Mpc}^{-1}$

General setup





HI systematics

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

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

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

Analysis

- Maximization of the likelihood with MCMC method

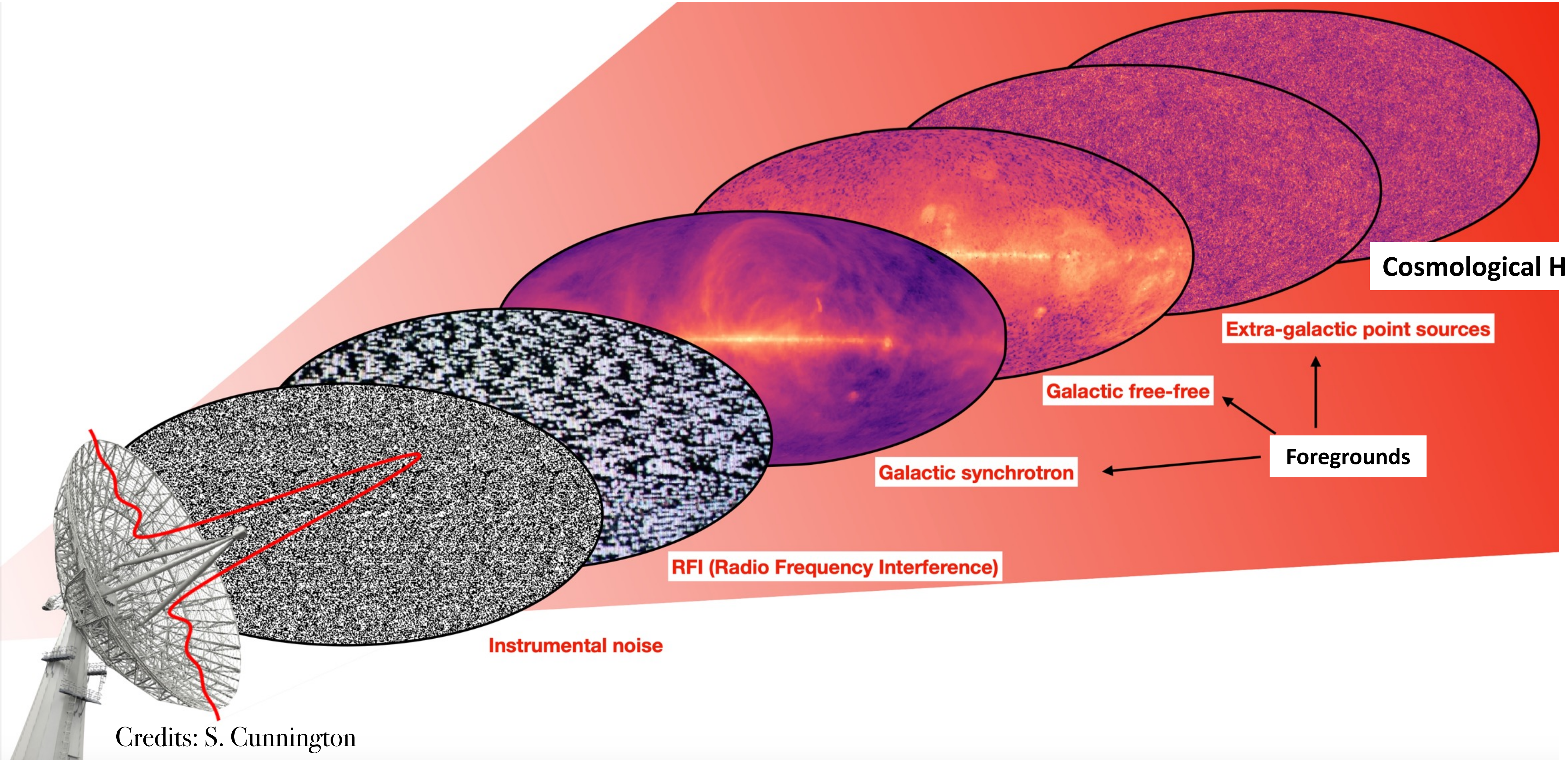
  $$\ln \mathcal{L}\,(\boldsymbol{\theta}; \boldsymbol{d}) \propto [\boldsymbol{d} - \boldsymbol{P}(\boldsymbol{\theta})]^{\mathrm{T}} \mathrm{Cov}^{-1}\,[\boldsymbol{d} - \boldsymbol{P}(\boldsymbol{\theta})]$$

  - $\boldsymbol{P}(\boldsymbol{\theta}) =$ 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)

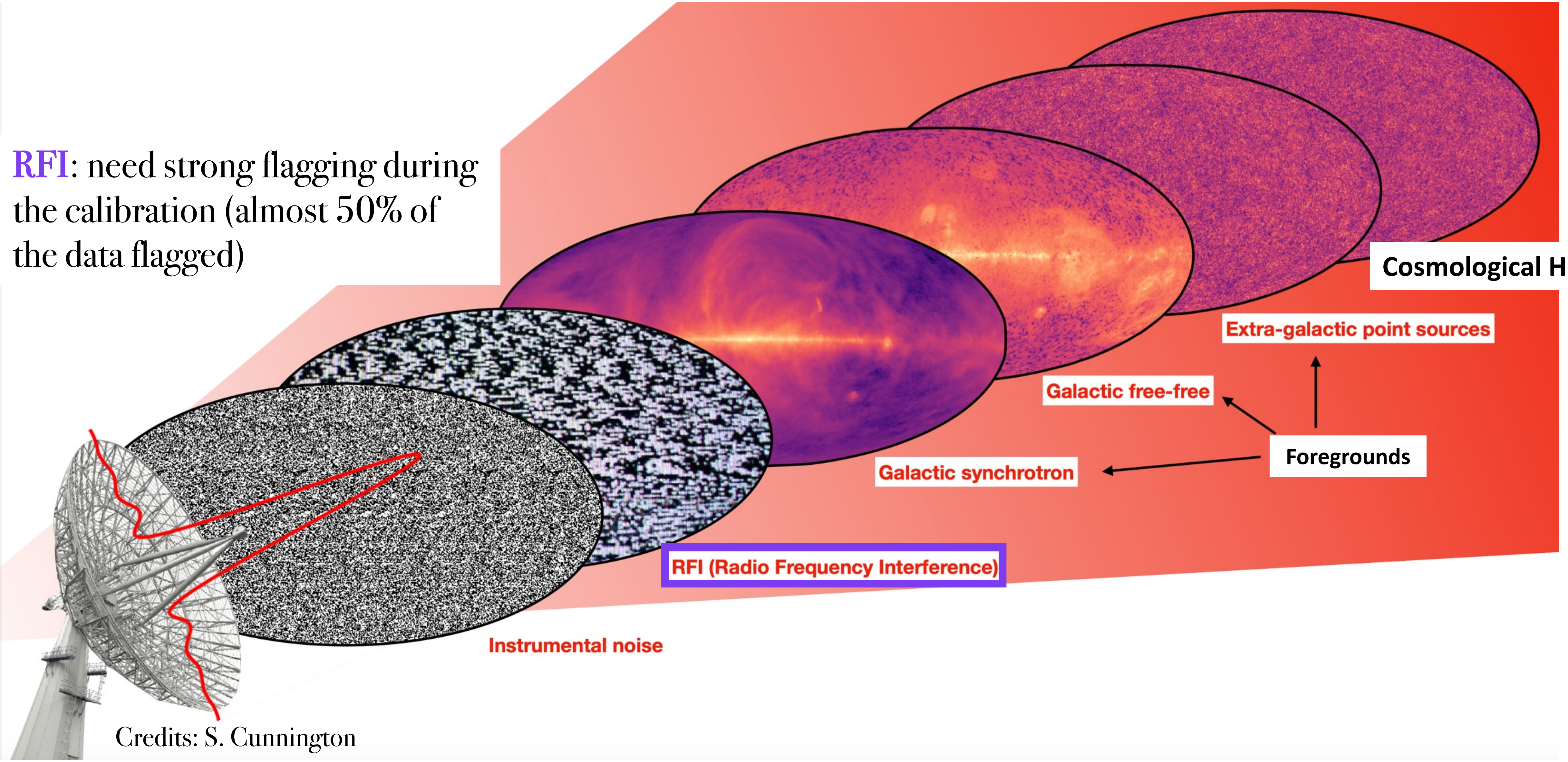
Why is Hi intensity mapping hard?



Credits: S. Cunnington

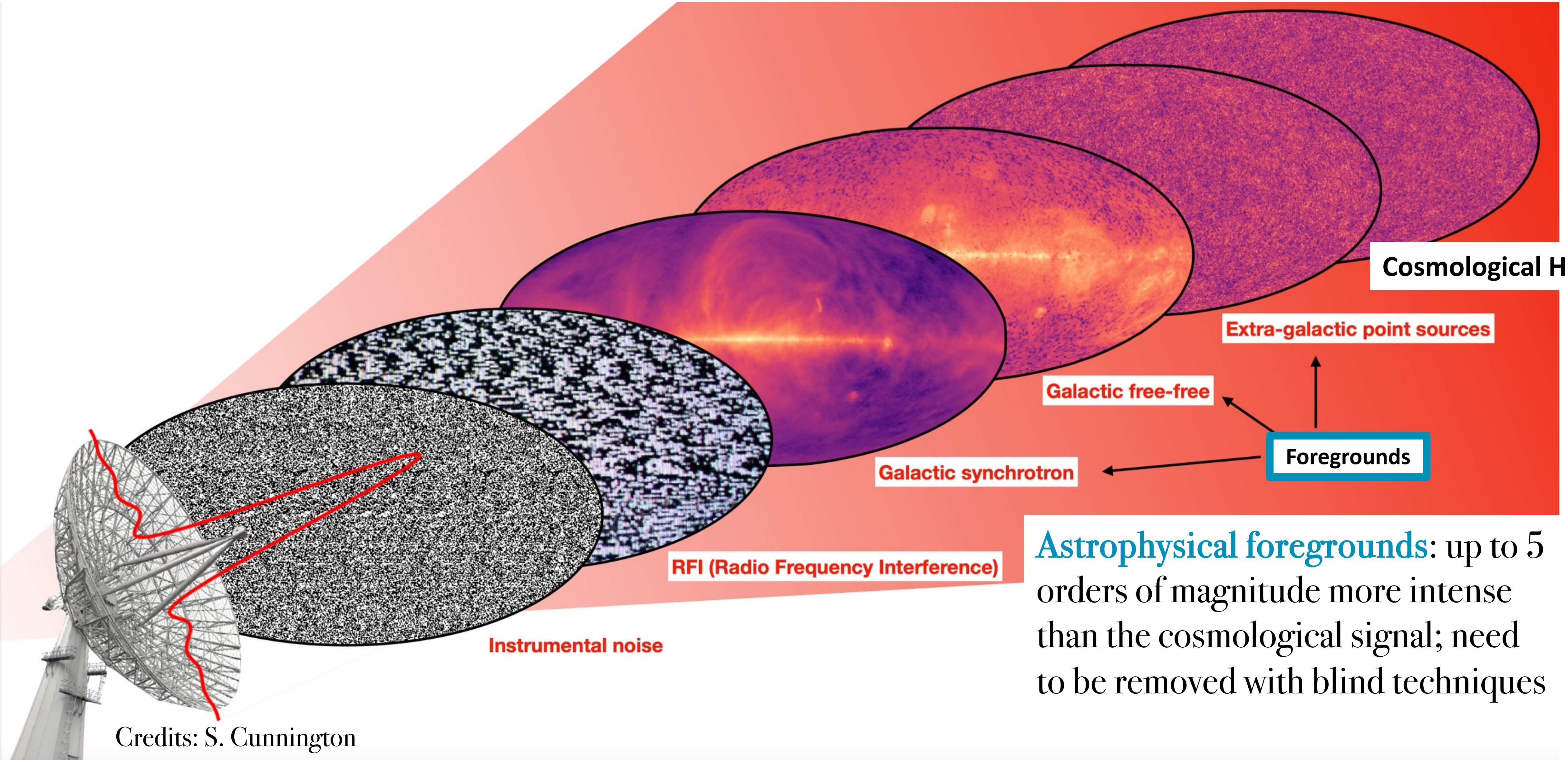
Why is HI intensity mapping hard?

RFI: need strong flagging during the calibration (almost 50% of the data flagged)



Credits: S. Cunnington

Why is HI intensity mapping hard?

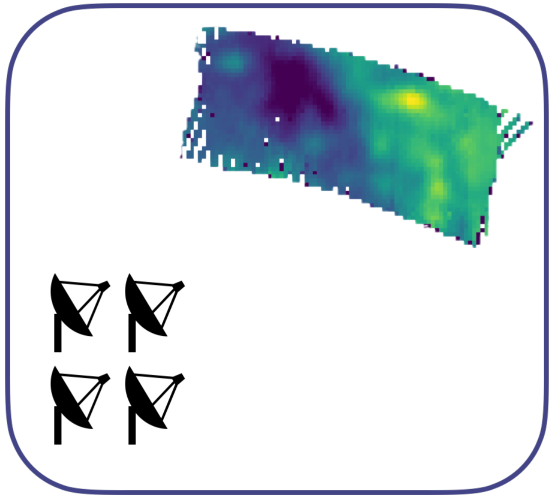


Credits: S. Cunnington

Astrophysical foregrounds: up to 5 orders of magnitude more intense than the cosmological signal; need to be removed with blind techniques

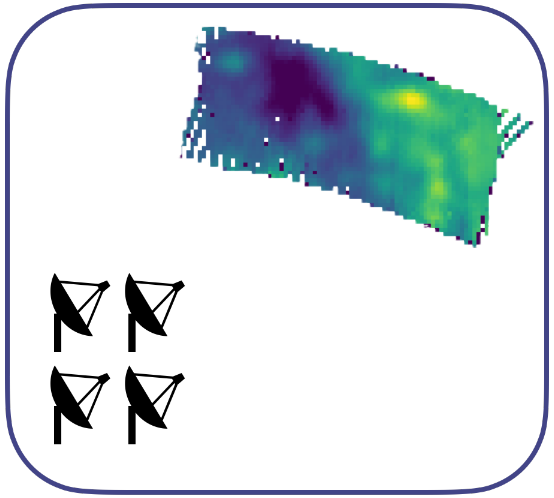
Single dish technique

- All the antennas of the array observe the same region at the same
- Low angular-resolution survey of the total 21cm flux from unresolved sources
- High signal-to-noise ratio
- Large cosmic volumes covered

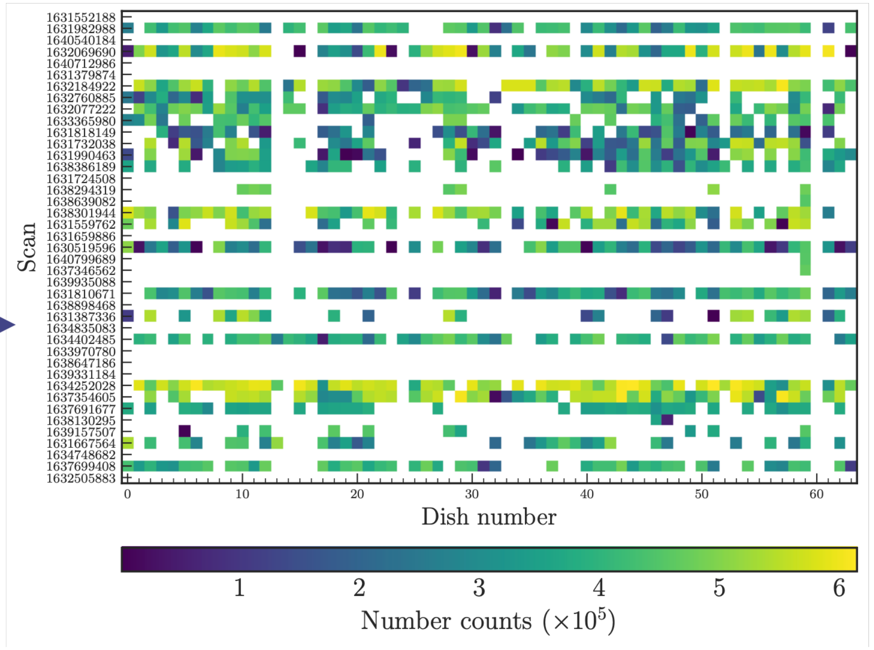


Single dish technique

- All the antennas of the array observe the same region at the same
- Low angular-resolution survey of the total 21cm flux from unresolved sources
- High signal-to-noise ratio
- Large cosmic volumes covered

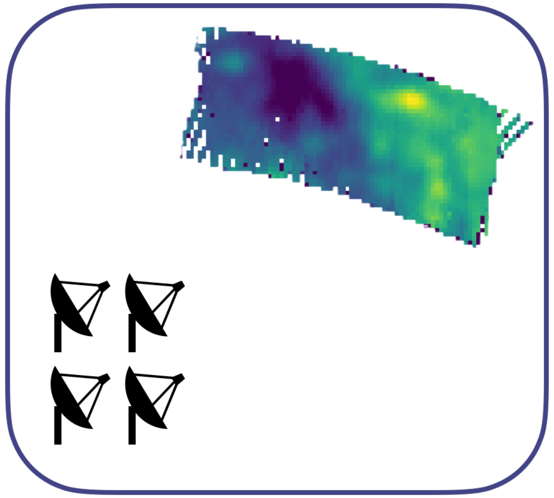


x *n* times

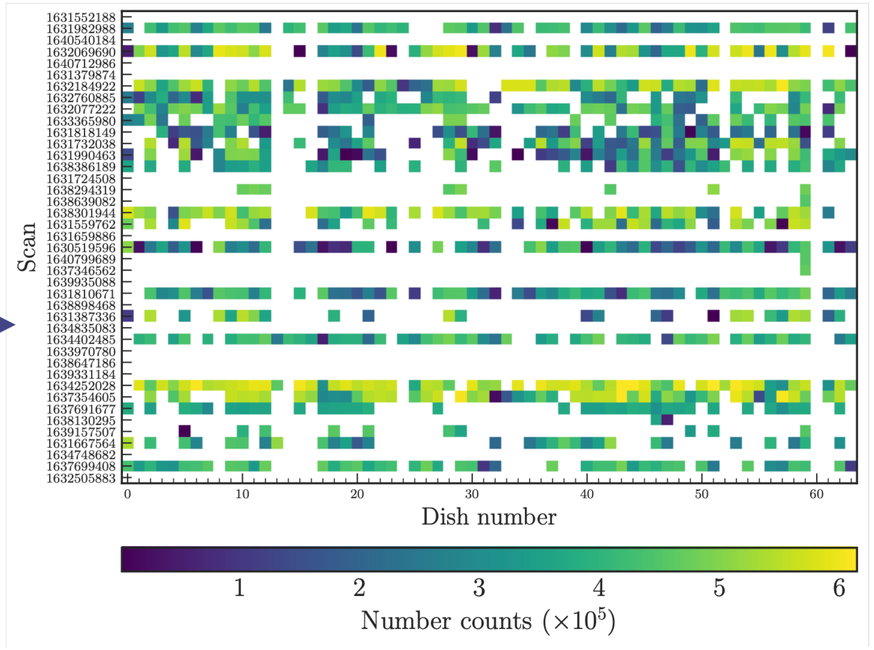


Single dish technique

- All the antennas of the array observe the same region at the same
- Low angular-resolution survey of the total 21cm flux from unresolved sources
- High signal-to-noise ratio
- Large cosmic volumes covered



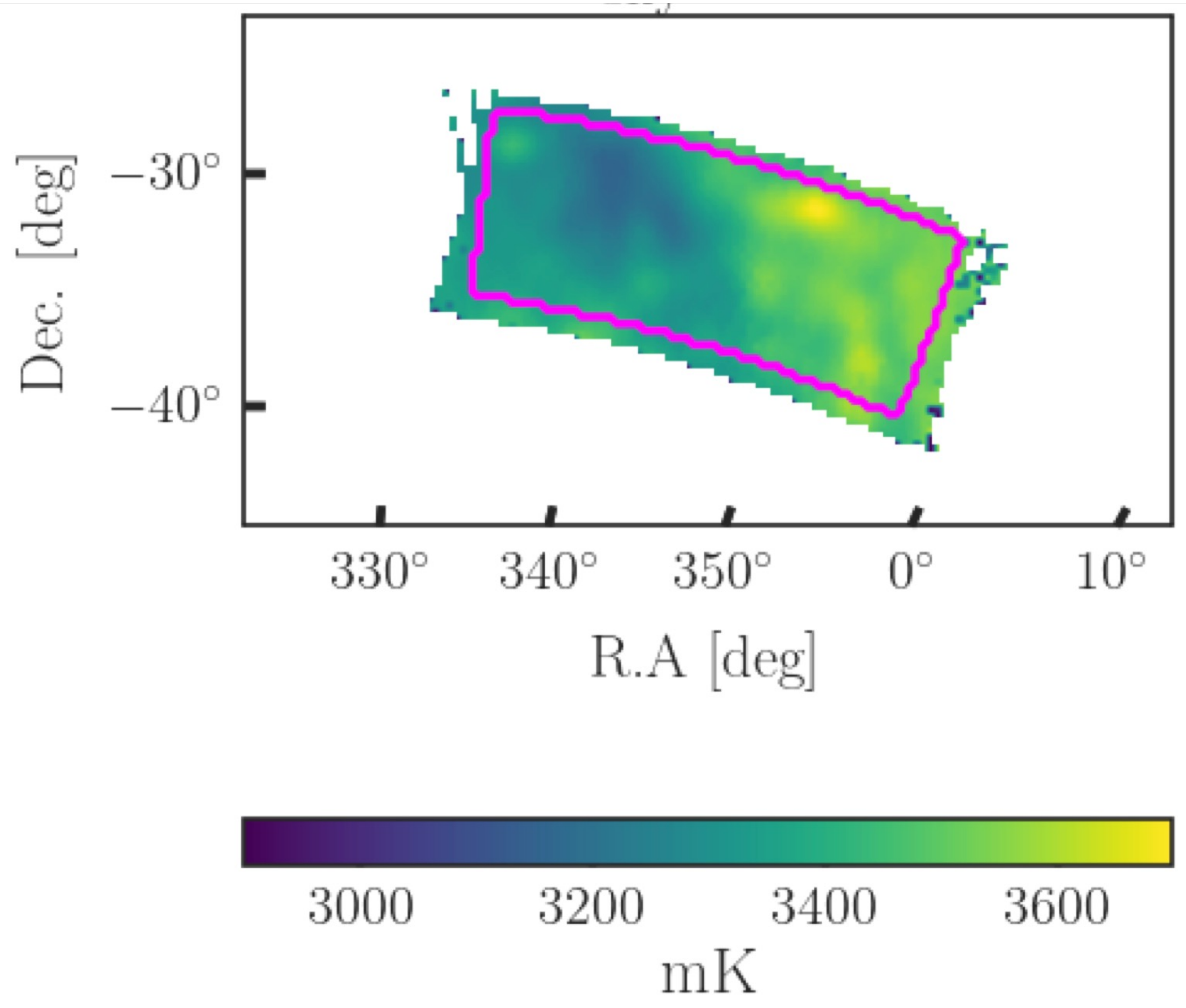
x *n* times →



→ Combined calibrated map

The 2021 data set

- The deepest HI survey
  - 62 hours of observation
  - 236 deg^2 patch
- Suitable to tackle the challenge of pursuing the measurements of the HI power spectrum in auto-correlation



The HI auto power spectrum

Full data set auto-correlation

- HI cosmological signal
- Noise
- Residuals foregrounds
- Contaminants

Internal cross-correlations

- Building independent subsets from the same survey [Wolz et al. (2021)]
- The HI cosmological signal is the only correlated part
- Contaminants not correlated between subsets
- Noise free cross-subset power spectra

MeerKLASS multi-subset fits

- Multi-tracer formalism translated to a multi-scan formalism (with cross power spectra only)

$$\boldsymbol{P} = \{P_{12}, P_{23}, P_{13}\}$$

$$\mathrm{Cov}(\boldsymbol{P},\boldsymbol{P}) \propto \begin{vmatrix} \frac{1}{2}\left(P_{11}P_{22}+P_{12}P_{12}\right) & \frac{1}{2}\left(P_{13}P_{22}+P_{12}P_{23}\right) & \frac{1}{2}\left(P_{11}P_{23}+P_{12}P_{13}\right) \\ & \frac{1}{2}\left(P_{22}P_{33}+P_{23}P_{23}\right) & \frac{1}{2}\left(P_{12}P_{33}+P_{13}P_{23}\right) \\ & & \frac{1}{2}\left(P_{11}P_{33}+P_{13}P_{13}\right) \end{vmatrix}$$