ARGOS-TOSCA-TITAN meeting - 08/07/2025

📍 Heraklion, Greece

"Foreground cleaning strategies for HI 21 cm signal extraction"

Athanasia GKOGKOU

Institute of Astrophysics/Institute of Computer Science - FORTH, Greece

Collaborators: V. Bonjean, J-L. Starck, M. Spinelli, P. Tsakalides

FORTH INSTITUTE OF COMPUTER SCIENCE

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS

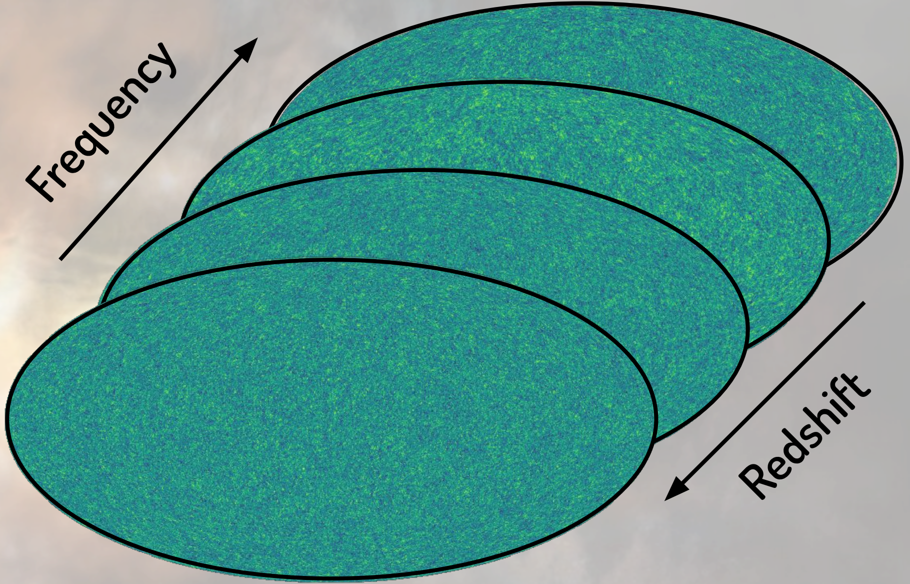
Signal Processing Laboratory

Funded by the European Union

FORTH INSTITUTE OF ASTROPHYSICS

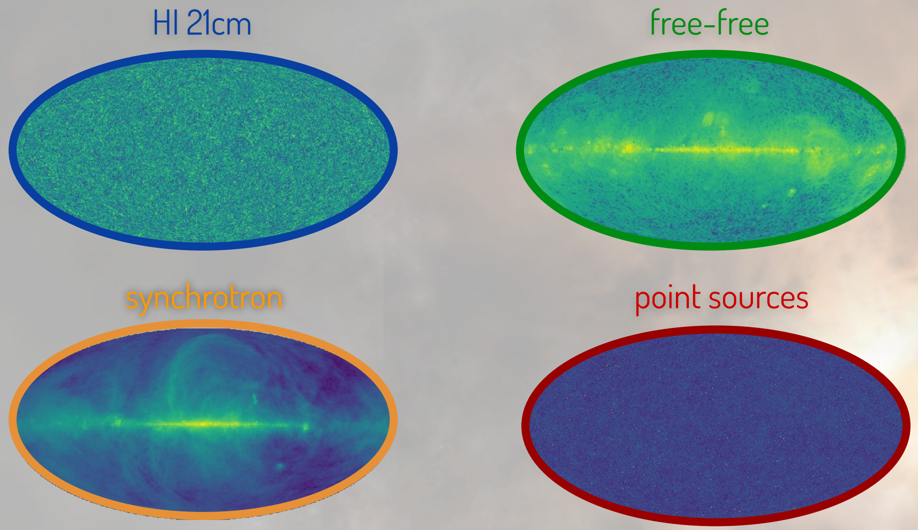
HI 21 cm Intensity Mapping

- Goal: probe the large-scale structure of the Universe
- Observations over large cosmic volumes & high spectral resolution
- HI 21 cm tracer of matter distribution across cosmic time
- Intensity mapping: large volumes & high spectral resolution

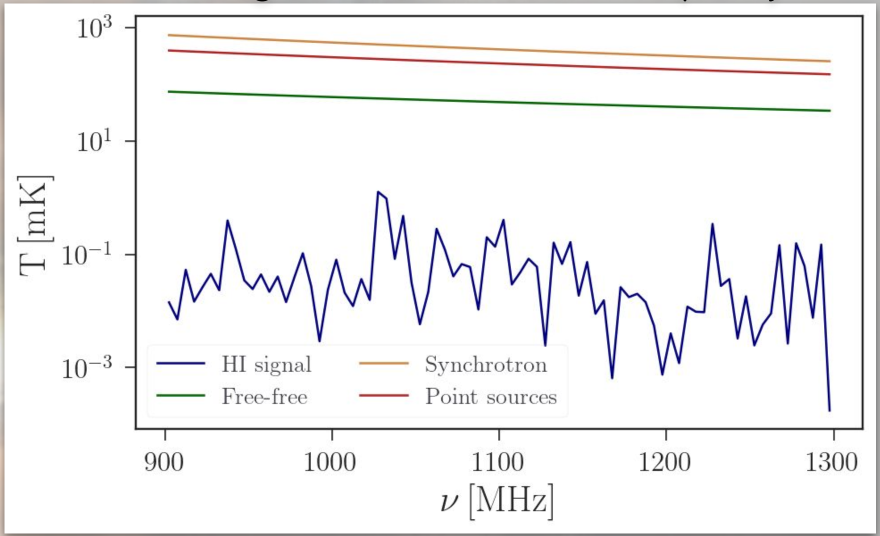


$$\nu = \frac{\nu_{21}}{1+z}$$

Foregrounds of the HI 21 cm signal



Foregrounds are smooth in frequency



Most common ways to extract the HI 21 cm signal:

- Baseline removal (e.g., polynomial fit, parametric fit)
- Blind source separation (BSS)

Foreground removal: Baseline removal

Polynomial fit

$$T_{\mathrm{fg}}(\nu, p) = \sum_{n=0}^{6} a_n(p)\, \nu^n$$

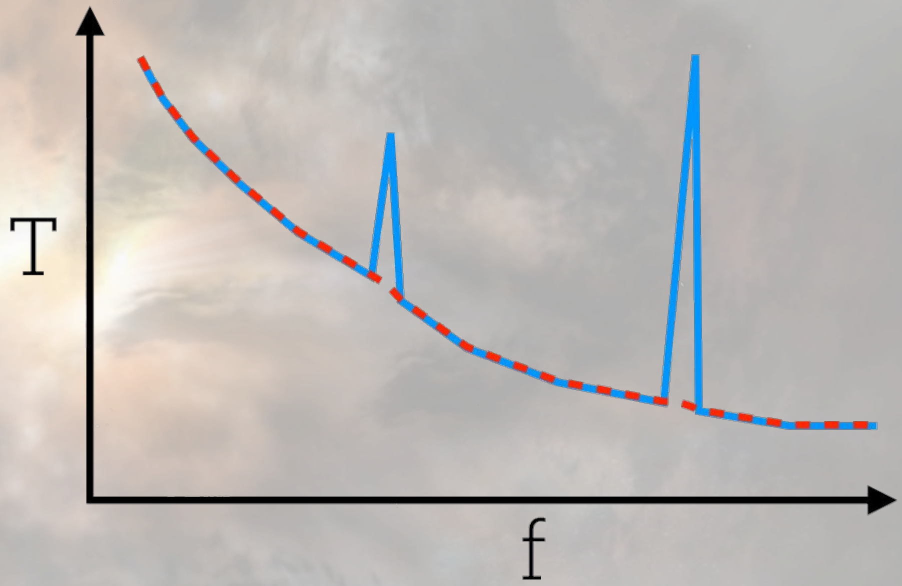
Parametric fit

$$T(\nu) = \alpha_1(p) \left(\frac{\nu}{\nu_0}\right)^{-2.13} + \alpha_2(p) \left(\frac{\nu}{\nu_0}\right)^{\beta(p)}$$

free-free emission

rest of diffuse emission

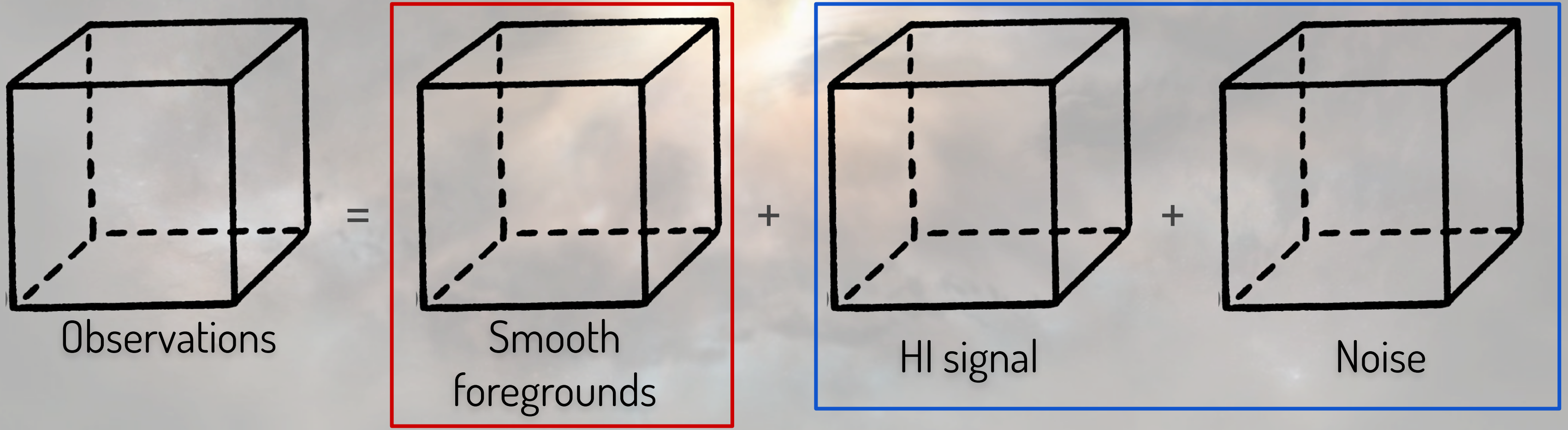
Intensity along a single line of sight



Foreground removal: BSS

- BSS:
  - Principal Component Analysis (**PCA**)
  - Independent Component Analysis (**ICA**)
  - Generalized Morphological Component Analysis (**GMCA**)

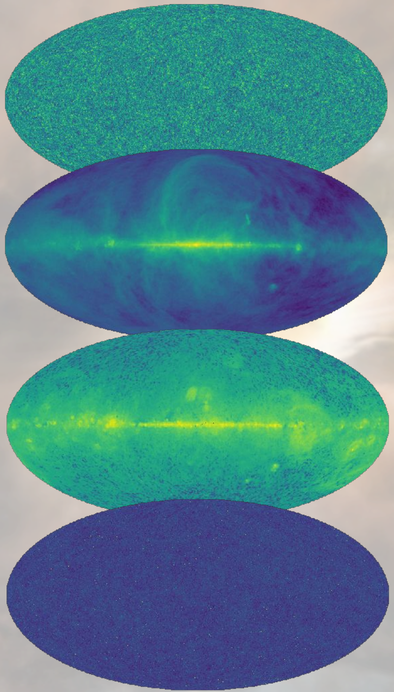
$$X = \boxed{AS} + \boxed{N}$$



Foreground removal pipeline

900-1400 MHz

- HI 21-cm cosmological signal: CRIME (Alonso+2014)
- Synchrotron diffuse emission: Planck Sky Model
- free-free diffuse emission: Planck Sky Model
- Extragalactic point sources: Empirical model (Battye+2013)



Simulation (Carucci+2020)

→

PCA

FastICA

GMCA

Polinomial Fit

Parametric fit

Methods

→

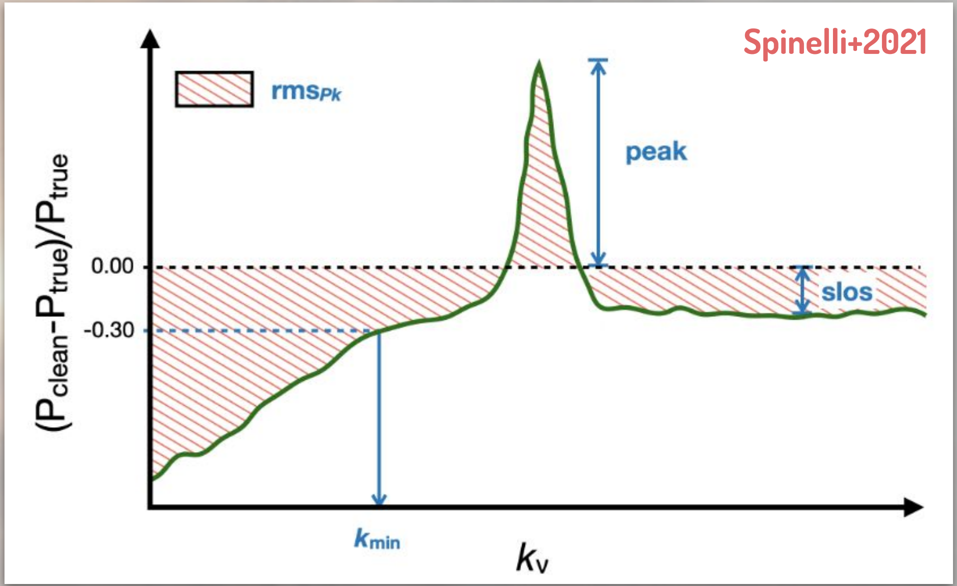
Radial (frequency) power spectrum $P(k_\nu)$

Angular power spectrum C_ℓ

Validation

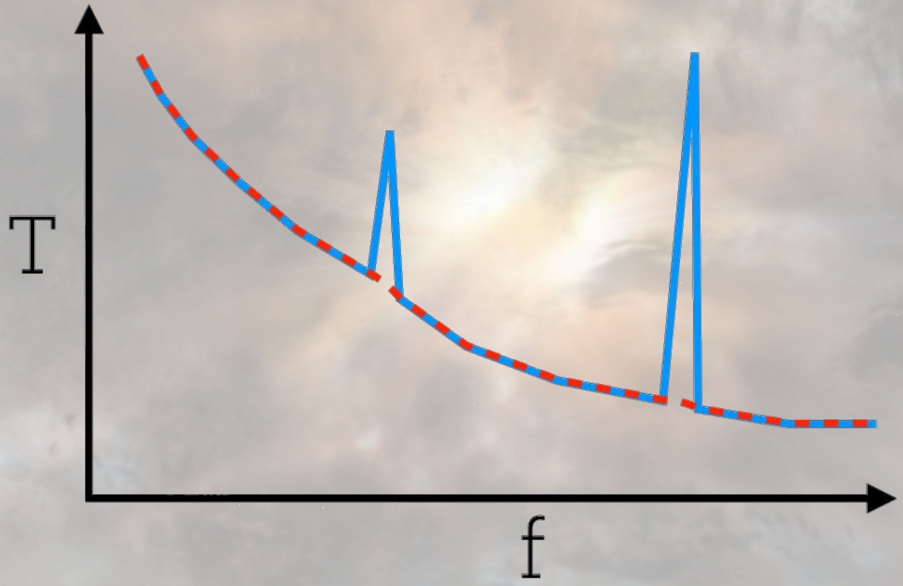
Constraints of current methods

- k_{min}: underestimating the power at large scales
- **slos**: systematic offset at all scales
- **peak**: spike at a certain k_ν



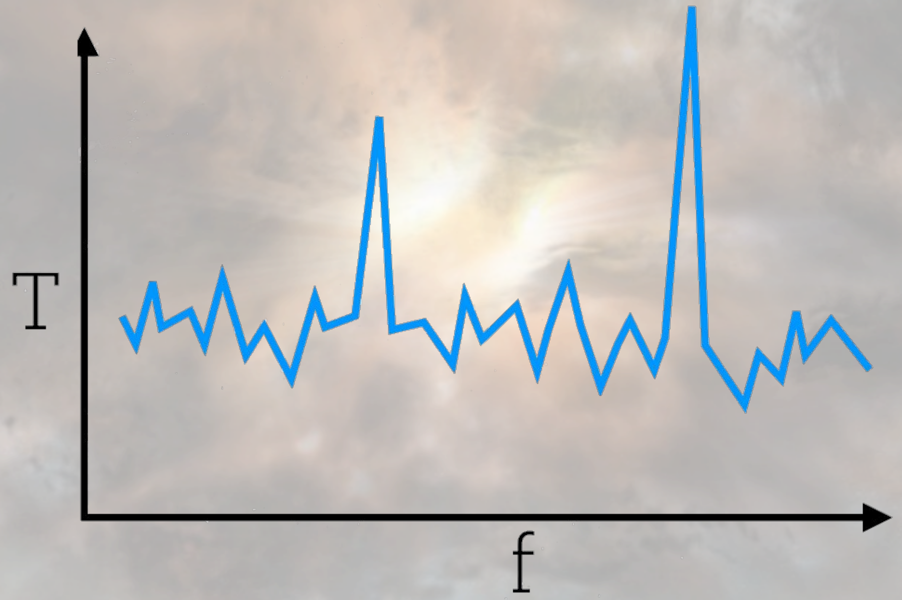
Masking the extremely bright HI sources

Constructing a mask from the data



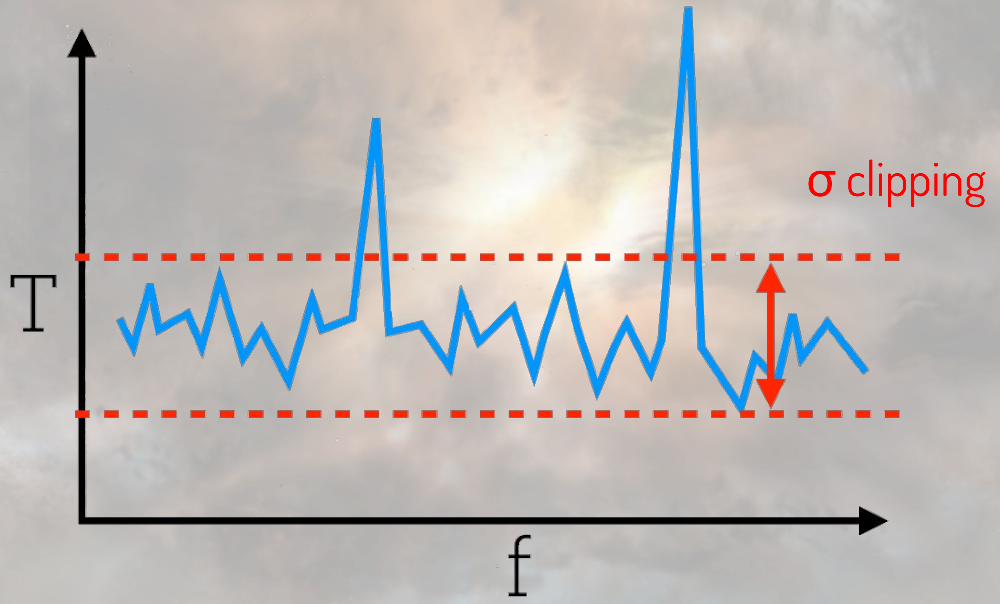
Masking the extremely bright HI sources

Constructing a mask from the data



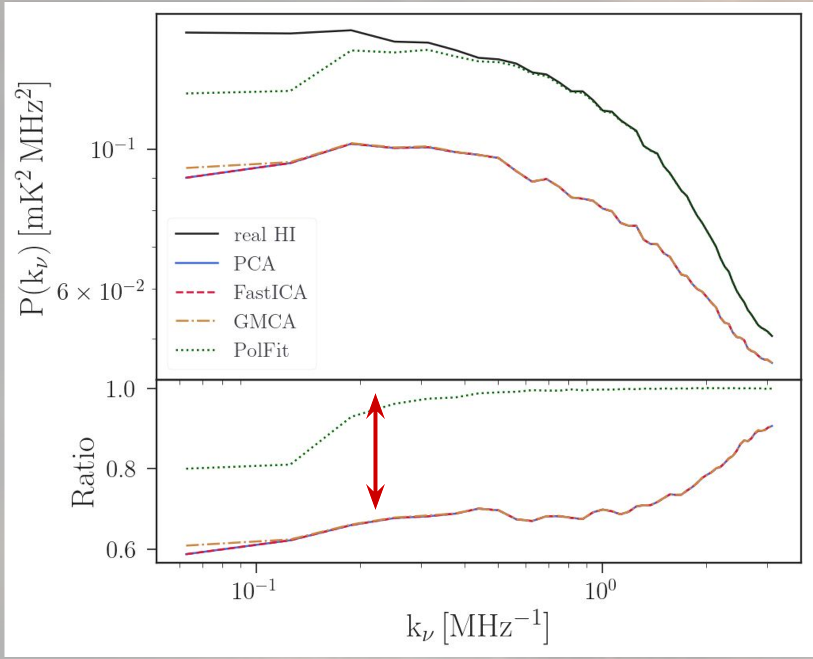
Masking the extremely bright HI sources

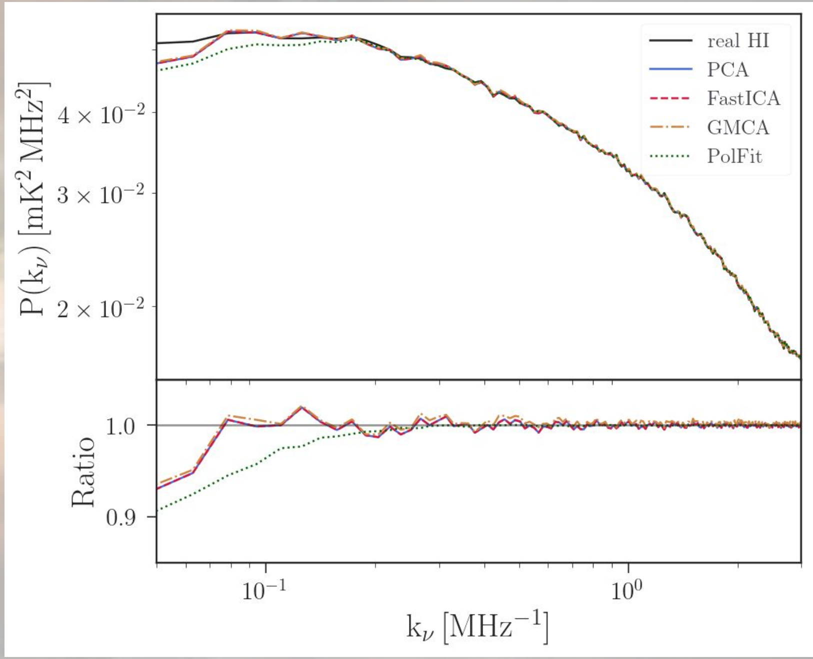
Constructing a mask from the data



Masking the extremely bright HI sources

Frequency power spectrum



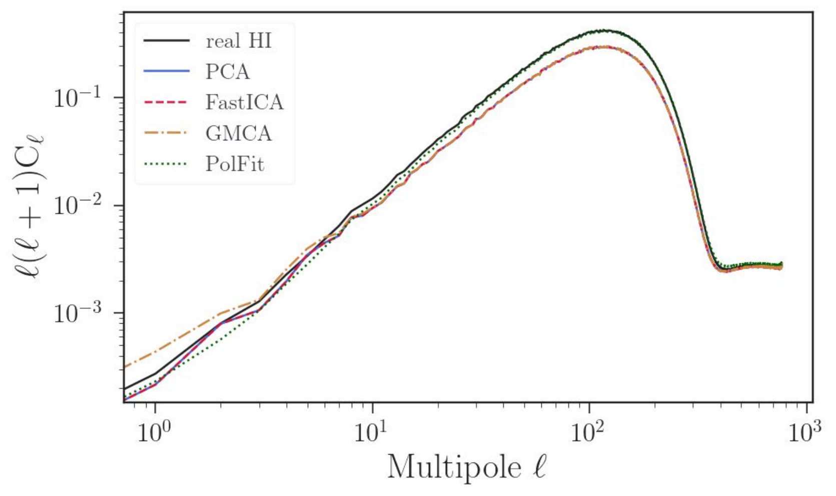


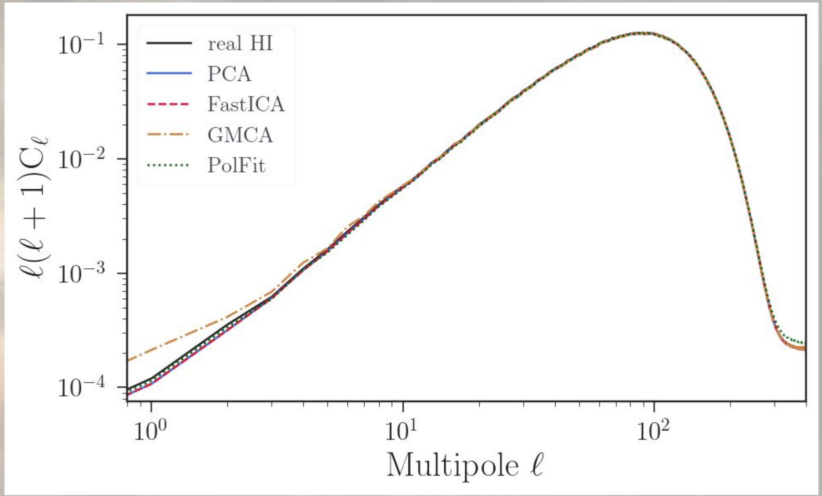
vertical **offset** significantly reduced

k_{min} pushed more to the left

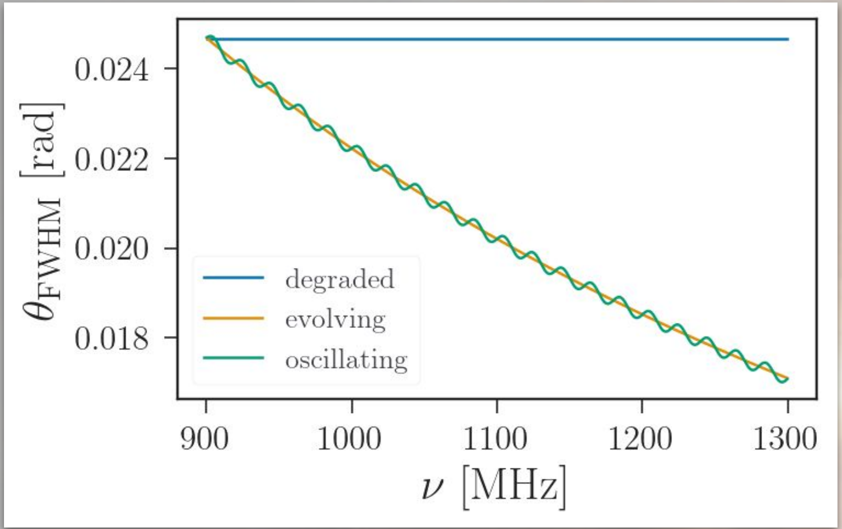
Masking the extremely bright HI sources

Angular power spectrum





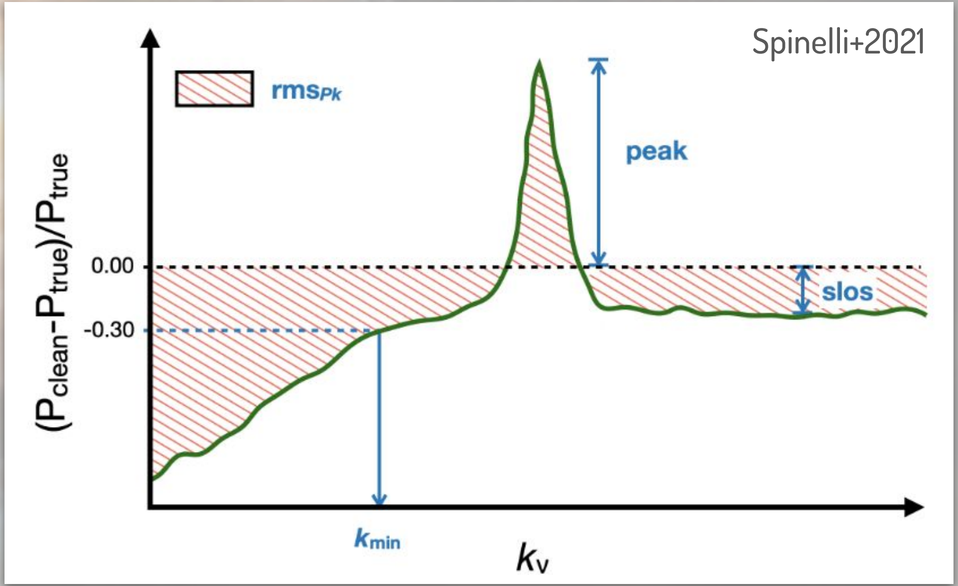
Constraints of current methods



$$\theta = \frac{\lambda}{D}$$

$$\theta = \frac{\lambda}{D} \times A \sin\left(\frac{2\pi\nu}{T}\right)$$

- k_{min}: underestimating the power at large scales
- **slos**: systematic offset at all scales
- **peak**: spike at a certain k_ν



SDecGMCA: joint component separation and deconvolution on the sphere

beam model

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \xrightarrow{\text{DecGMCA}} \mathbf{X}_\nu = (\mathbf{A}_\nu \mathbf{S}) * \boxed{\mathbf{H}_\nu} + \mathbf{N}_\nu \xrightarrow{\text{spherical harmonic domain}} \hat{\mathbf{X}}_\nu^{\ell,\mathbf{m}} = \hat{\mathbf{H}}_\nu^\ell \mathbf{A}_\nu \hat{\mathbf{S}}^{\ell,\mathbf{m}} + \hat{\mathbf{N}}_\nu^{\ell,\mathbf{m}}$$

$$\min_{\mathbf{A},\hat{\mathbf{S}}} \|\mathbf{\Lambda} \odot (\hat{\mathbf{S}}\mathcal{F}^\dagger \mathbf{\Phi}^\top)\|_1 + \sum_{(\ell,m)\in\mathcal{D}} \|\hat{\mathbf{X}}^{\ell,m} - \mathrm{diag}(\hat{\mathbf{H}}^\ell)\mathbf{A}\hat{\mathbf{S}}^{\ell,m}\|_2^2$$

sparsity
in wavelet space

data fidelity
in SH domain

algorithm

(i) *Update* $\hat{\mathbf{S}}$: Solved using a Tikhonov-regularized least-squares solution for each (ℓ, m), with $\epsilon_{n\ell}$ the regularization coefficients

(ii) *Enforce sparsity*: After the inverse harmonic transform, $\hat{\mathbf{S}}$ is projected onto the wavelet basis and sparsity is enforced via a soft-thresholding procedure

(iii) *Update* $\mathbf{A}$: The mixing matrix is estimated via least squares over all spherical harmonics

$$\mathbf{M}[l] = \mathbf{A}^\top \mathrm{diag}\left(\hat{\mathbf{H}}^l\right)^2 \mathbf{A}$$

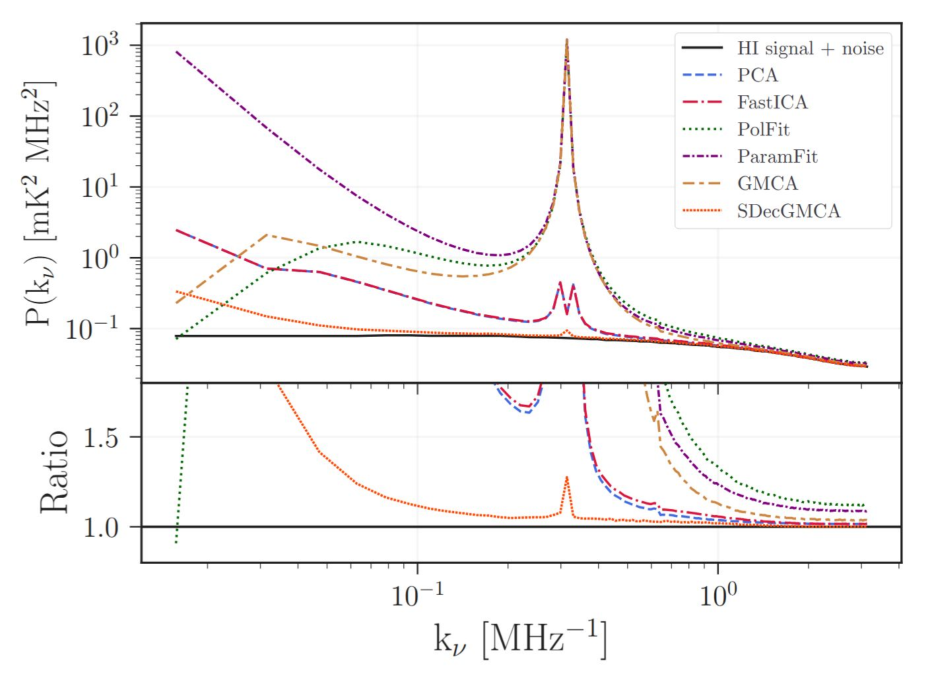
$$\hat{\mathbf{S}}^{l,m} \leftarrow \left(\boxed{\mathbf{M}[l]} + \underset{n\in[1,N_s]}{\mathrm{diag}} \left(\varepsilon_{n,l}\right)\right)^{-1} \mathbf{A}^\top \mathrm{diag}\left(\hat{\mathbf{H}}^l\right) \hat{\mathbf{X}}^{l,m}$$

$$\mathbf{A}_\nu \leftarrow \left(\sum_{(l,m)\in\mathcal{D}} \hat{\mathbf{X}}_\nu^{l,m} \hat{\mathbf{H}}_\nu^l \hat{\mathbf{S}}^{l,m\dagger}\right) \left(\sum_{(l,m)\in\mathcal{D}} {\hat{\mathbf{H}}_\nu^l}^2 \hat{\mathbf{S}}^{l,m} \hat{\mathbf{S}}^{l,m\dagger}\right)^{-1}$$

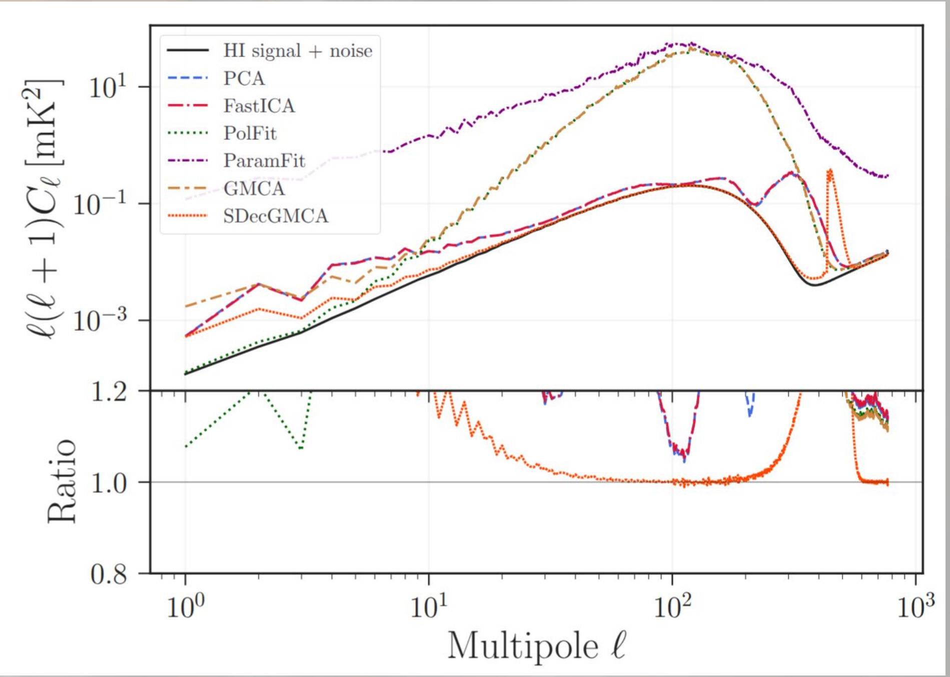
Carloni Gertosio & Bobin 2021

Gaussian oscillating beam

Frequency power spectrum



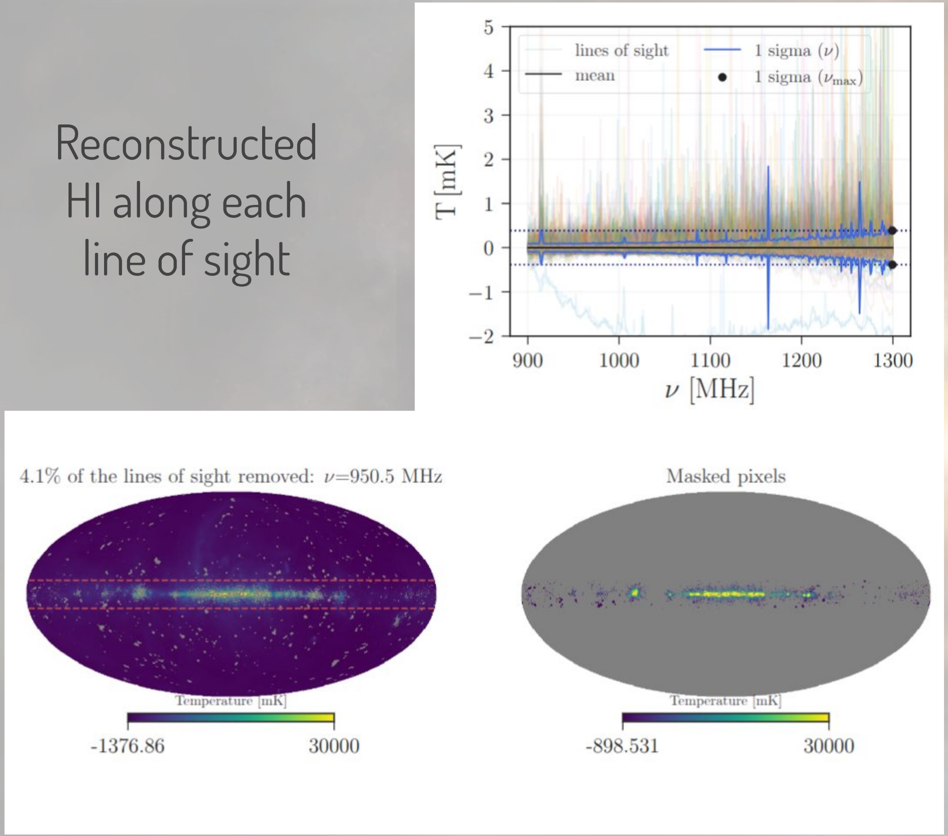
Angular power spectrum

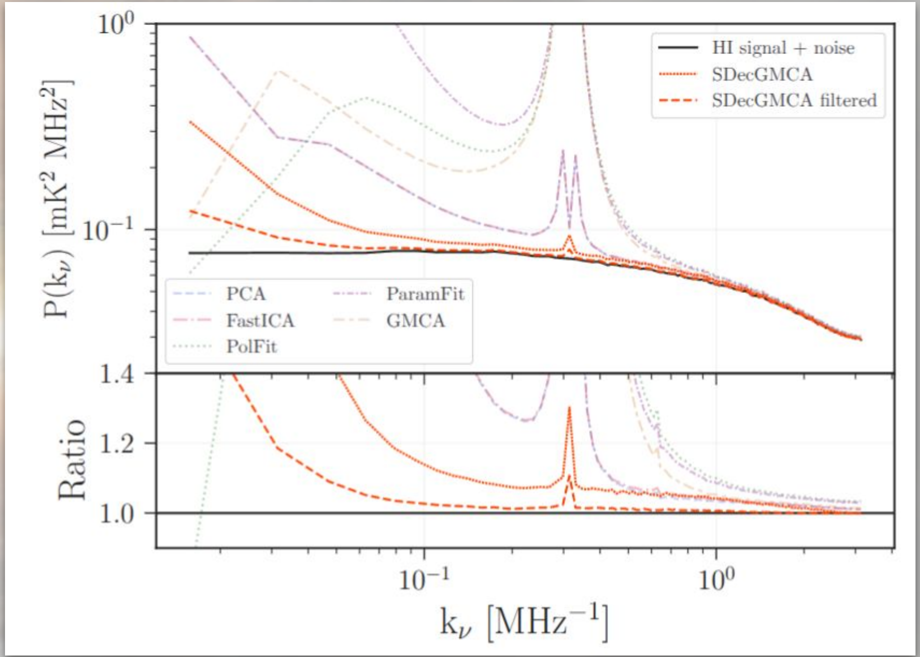


Gkogkou+ to be submitted

Effect of masking : (i) post-processing

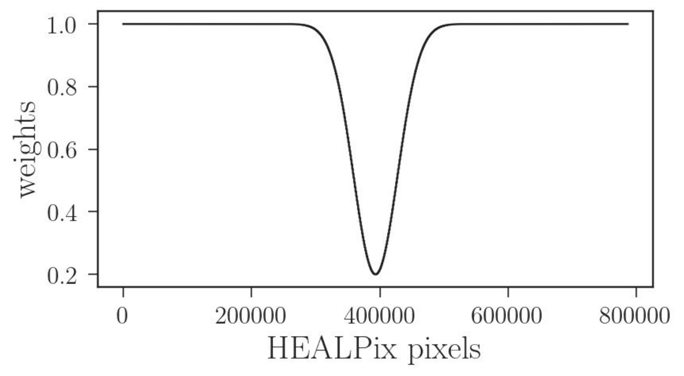
Reconstructed HI along each line of sight

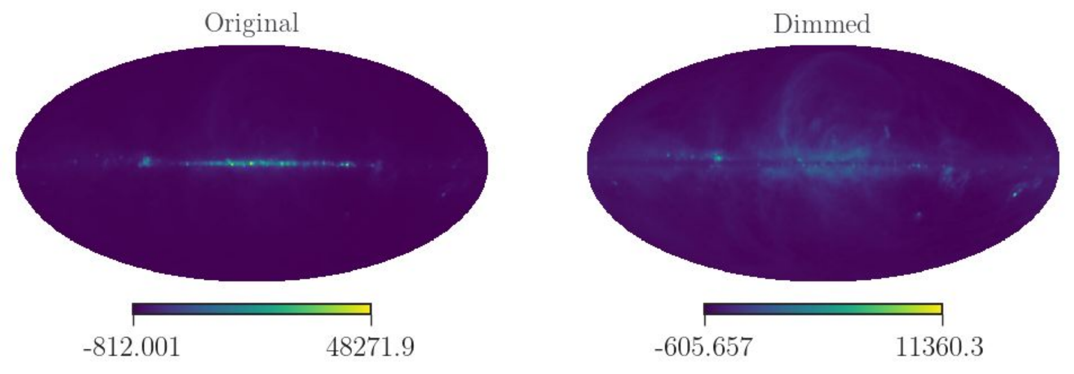


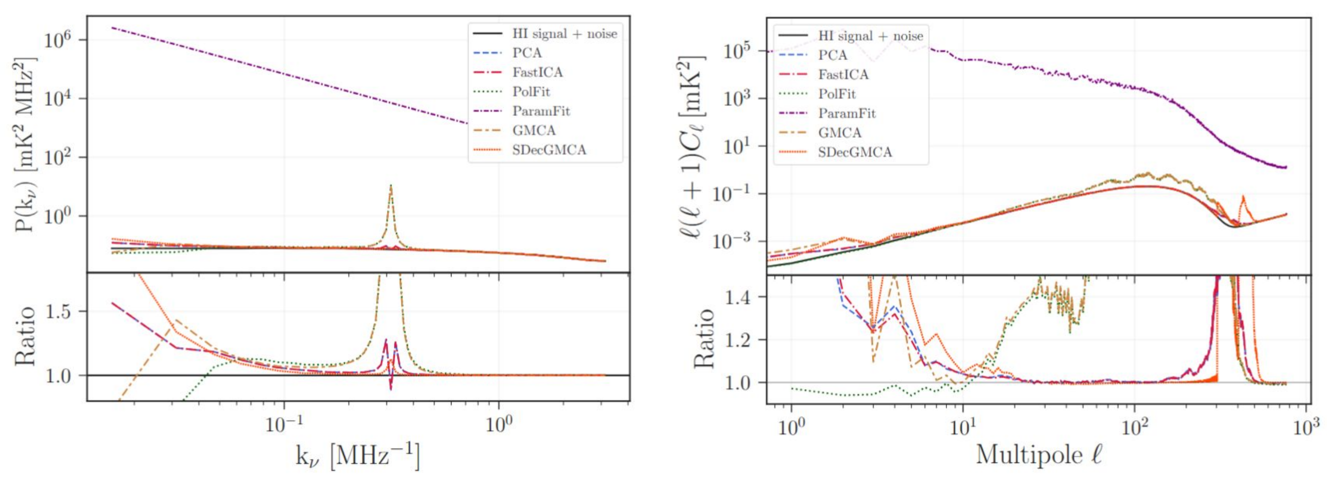


Gkogkou+ to be submitted

Effect of masking : (ii) simulating a Galactic mask







Gkogkou+ to be submitted

SDecGMCA: Future work

- **Masked SDecGMCA**
  - Masked pixels in the algorithm

$$\mathbf{X}_\nu = (\mathbf{A}_\nu \mathbf{S}) * \mathbf{H}_\nu + \mathbf{N}_\nu$$

- **SDecGMCA + Learnlets** (Bonjean+25 submitted)
  - Instead of thresholding in wavelet domain
  - Use learnlets

Summary of `SDecGMCA` iteration:

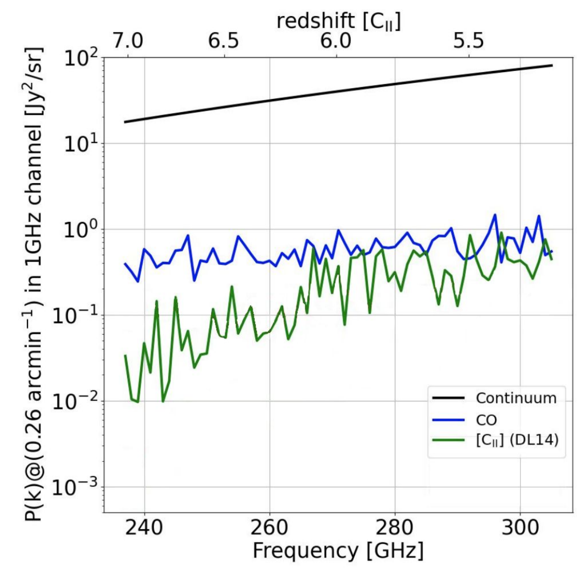
- Pixel space: start with `S`
- convert to spherical harmonics (`Slm`)
- Beam inversion in spherical harmonics (SH)
- apply wavelet transform in SH domain → `Swt`
- threshold in wavelet domain
- inverse wavelet → back to pixel space (`S`)
- convert `S` and `X` to SH domain, extract detail scales → `Slm_det`, `Xlm_det`
- update `A` from `Slm_det`, `Ylm_det` (in SH domain)
- update weights, thresholds

SDecGMCA: Future work

- **Masked SDecGMCA**
  - Masked pixels in the algorithm

$$\mathbf{X}_\nu = (\mathbf{A}_\nu \mathbf{S}) * \mathbf{H}_\nu + \mathbf{N}_\nu$$

- **SDecGMCA + Learnlets** (Bonjean+25 submitted)
  - Instead of thresholding in wavelet domain
  - Use learnlets
- **Application to other spectral lines**
  - e.g., removing CIB from LIM maps



Summary & Conclusions

- Masking the extremely bright sources removes the systematic offset
- Joint component separation and deconvolution (**SDecGMCA**): reduces the peak and improves reconstruction
- Masking the Galactic plane can significantly improve the performance of SDecGMCA
- Future improvements:
  - masked SDecGMCA
  - SDecGMCA + Learnlets