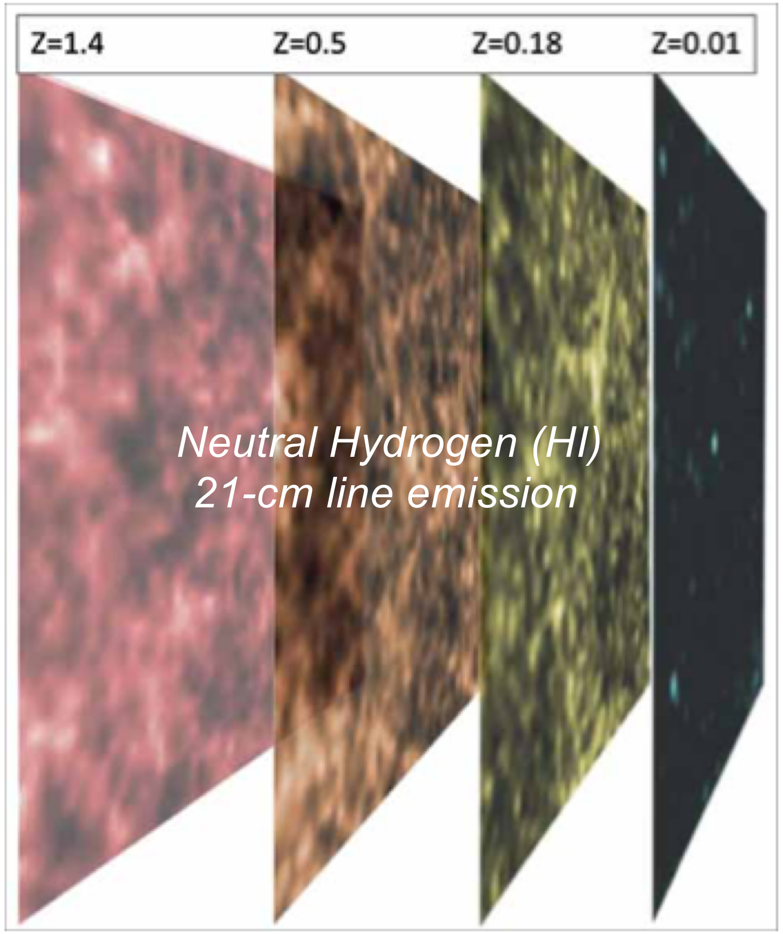
Component separation with GNILC for 21-cm line intensity mapping

Mathieu Remazeilles
Jodrell Bank Centre for Astrophysics
MANCHESTER 1824
The University of Manchester

On behalf of the BINGO collaboration

21-cm line intensity mapping with BINGO



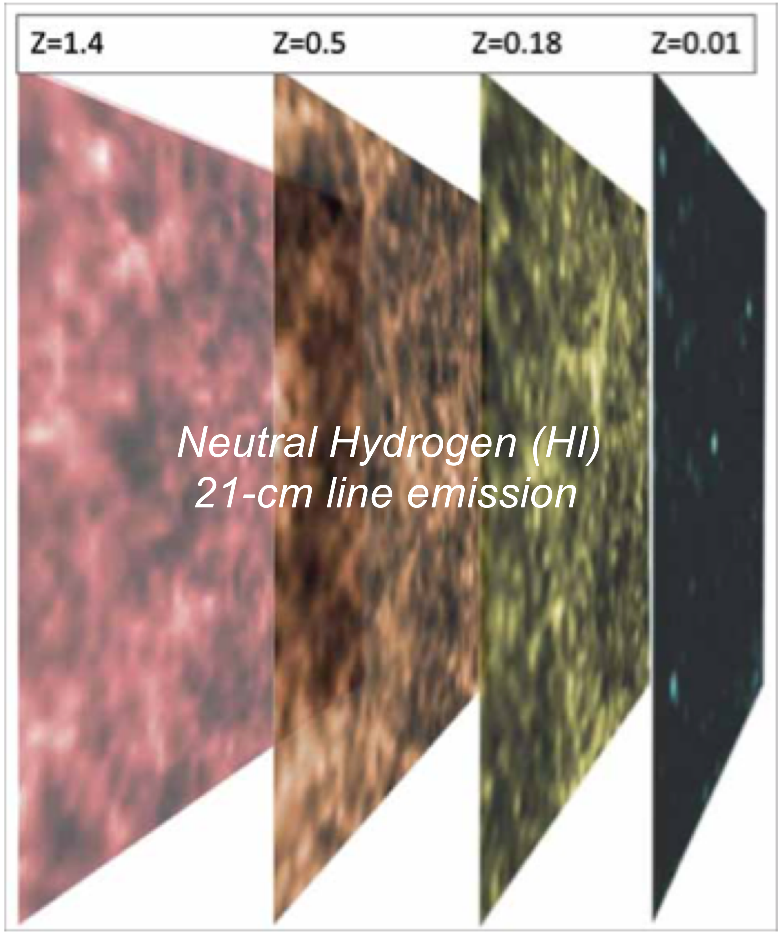
Browne, Astron. Geophys. (2014)

Tomography of the large-scale structure through redshifted HI 21-cm line emission

Credit: BINGO collaboration

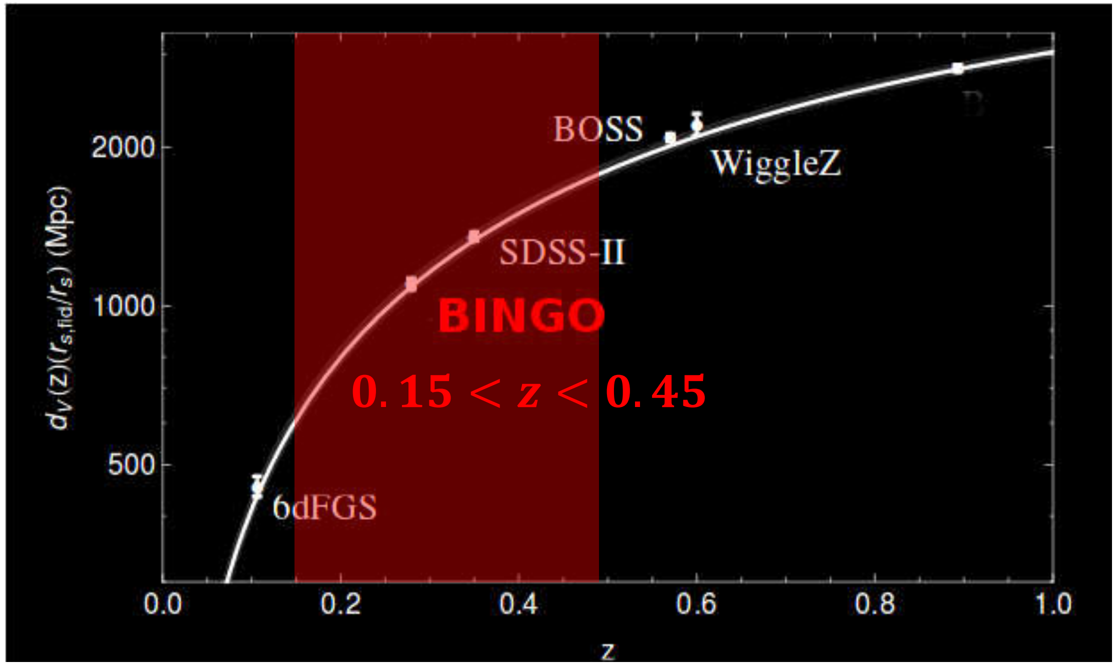
M. Remazeilles

21-cm line intensity mapping with BINGO



Browne, Astron. Geophys. (2014)

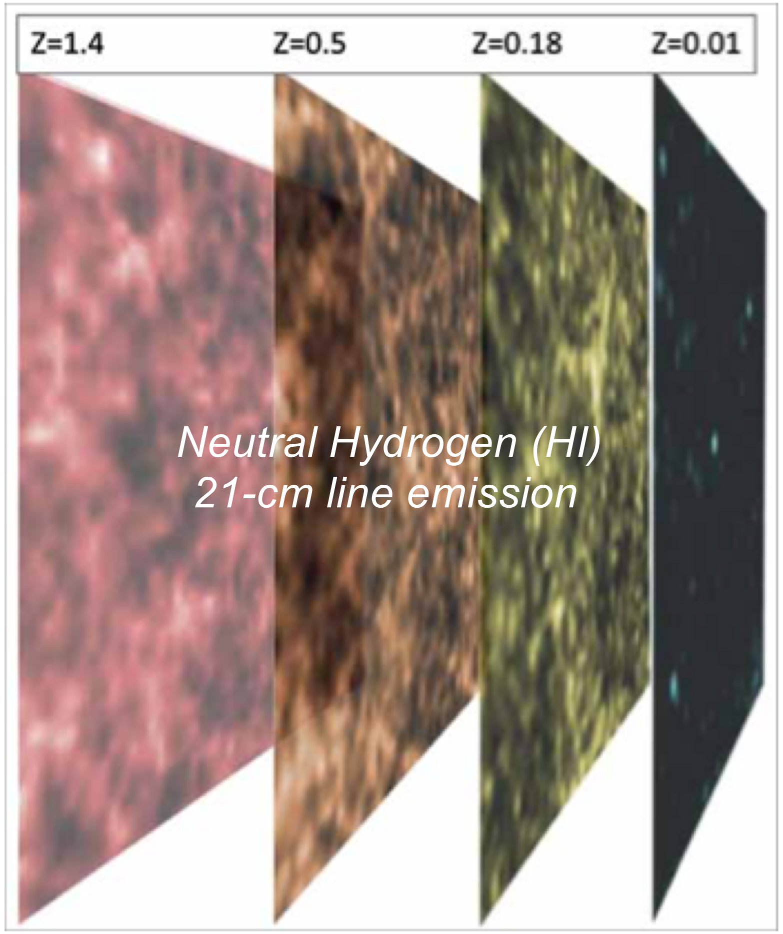
Tomography of the large-scale structure through redshifted HI 21-cm line emission



Radio wavelengths allow to probe larger redshift volumes as compared to optical surveys

M. Remazeilles

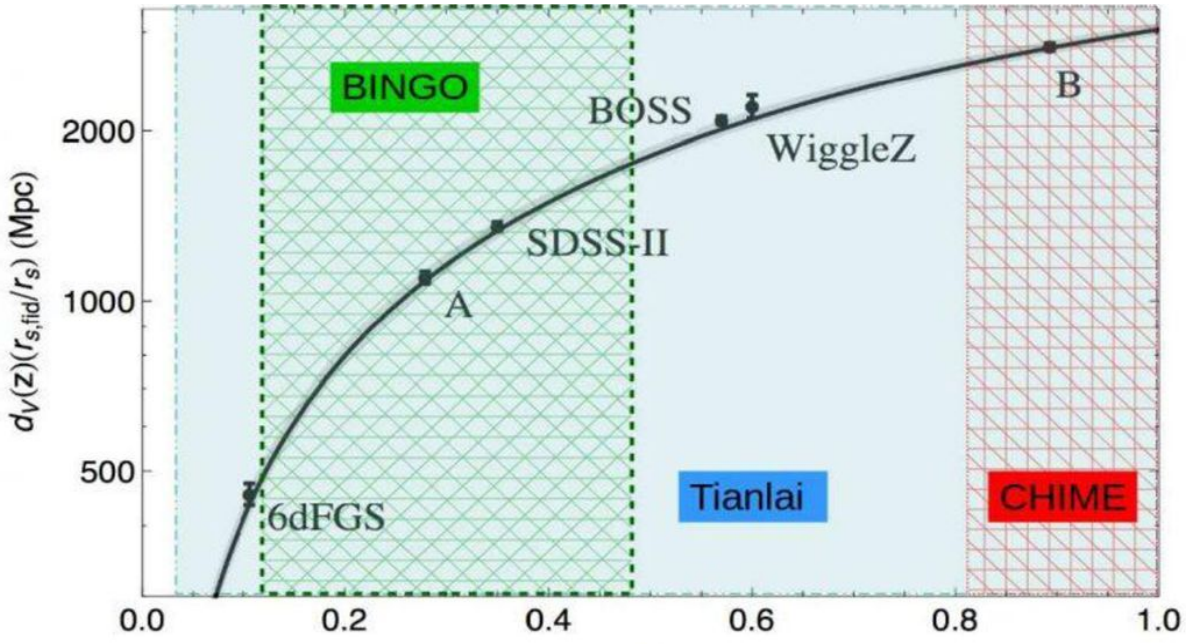
21-cm line intensity mapping with BINGO



Browne, Astron. Geophys. (2014)

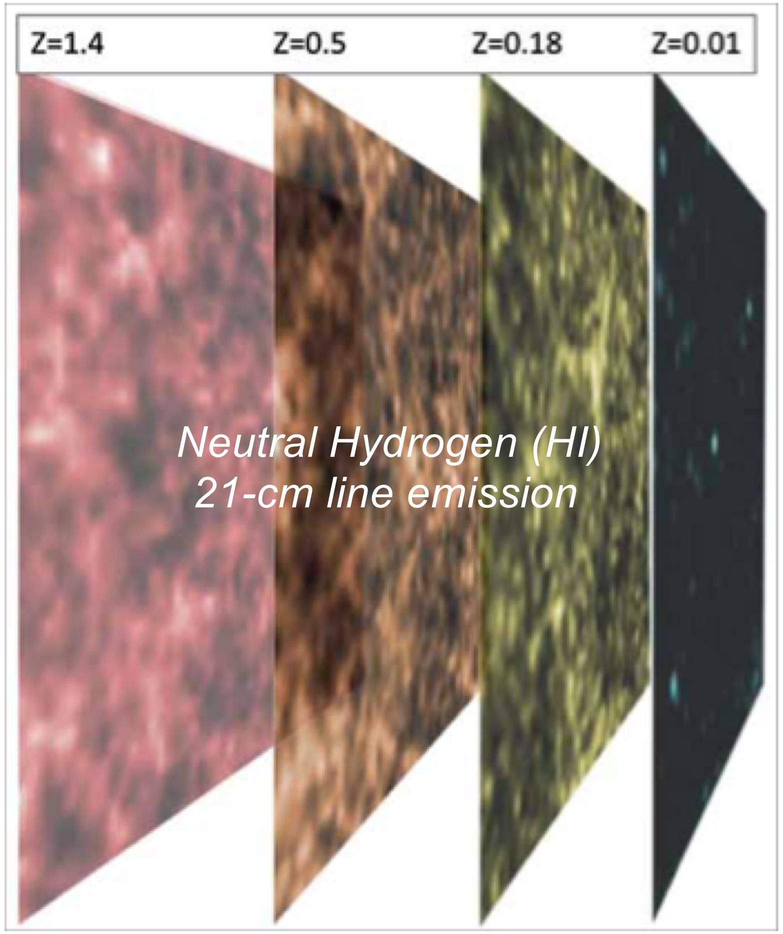
Tomography of the large-scale structure through redshifted HI 21-cm line emission

Wuensche, and the BINGO collaboration (2019)



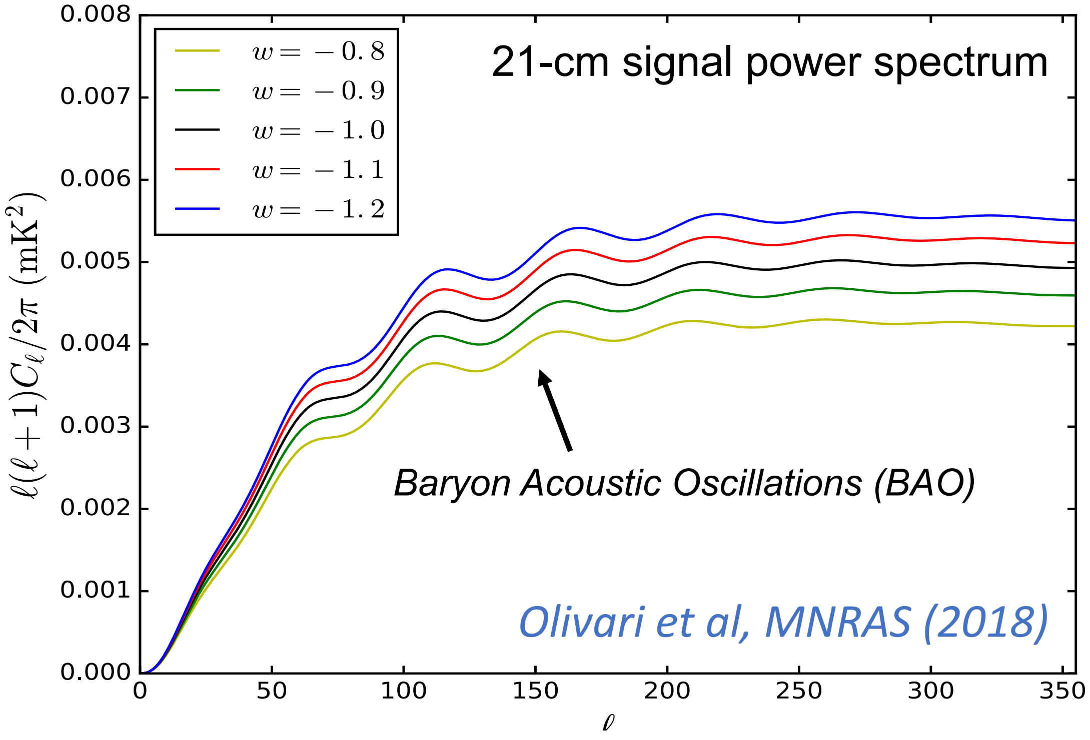
Radio wavelengths allow to probe larger redshift volumes as compared to optical surveys

21-cm line intensity mapping with BINGO



Browne, Astron. Geophys. (2014)

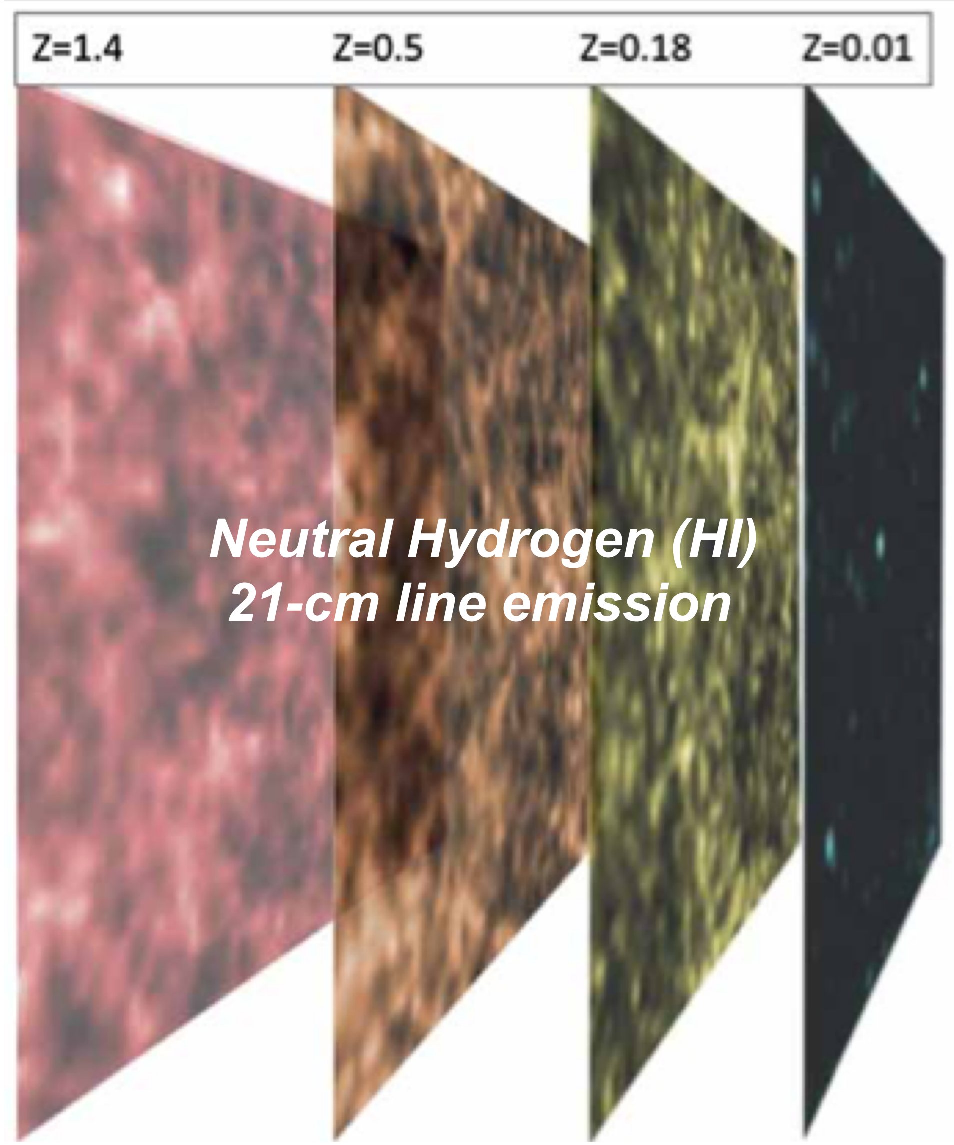
Tomography of the large-scale structure through redshifted HI 21-cm line emission



BINGO will probe BAO and dark energy across redshifts $0.15 < z < 0.45$

M. Remazeilles

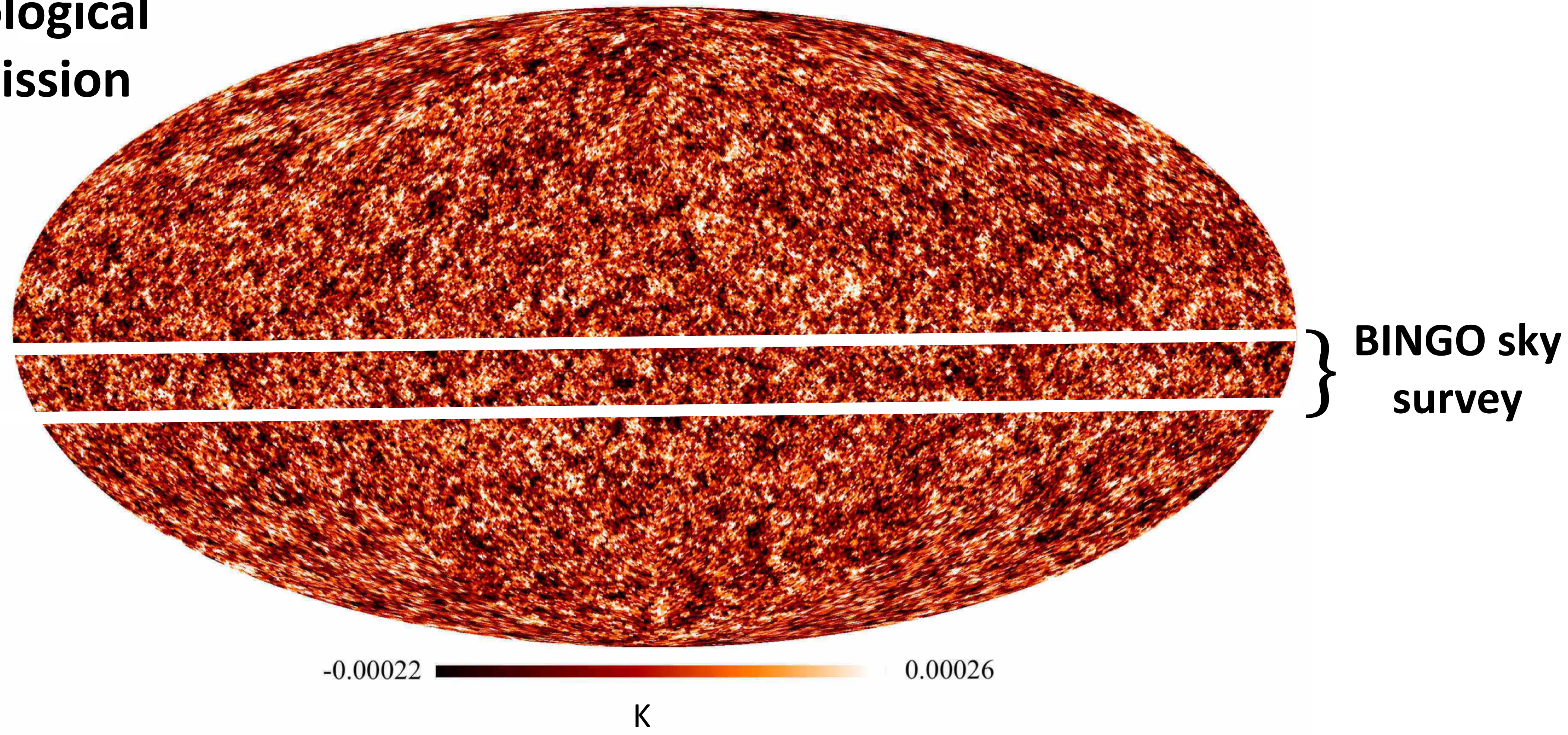
Galactic foregrounds obscure the 21-cm signal



Foreground emission from our Galaxy

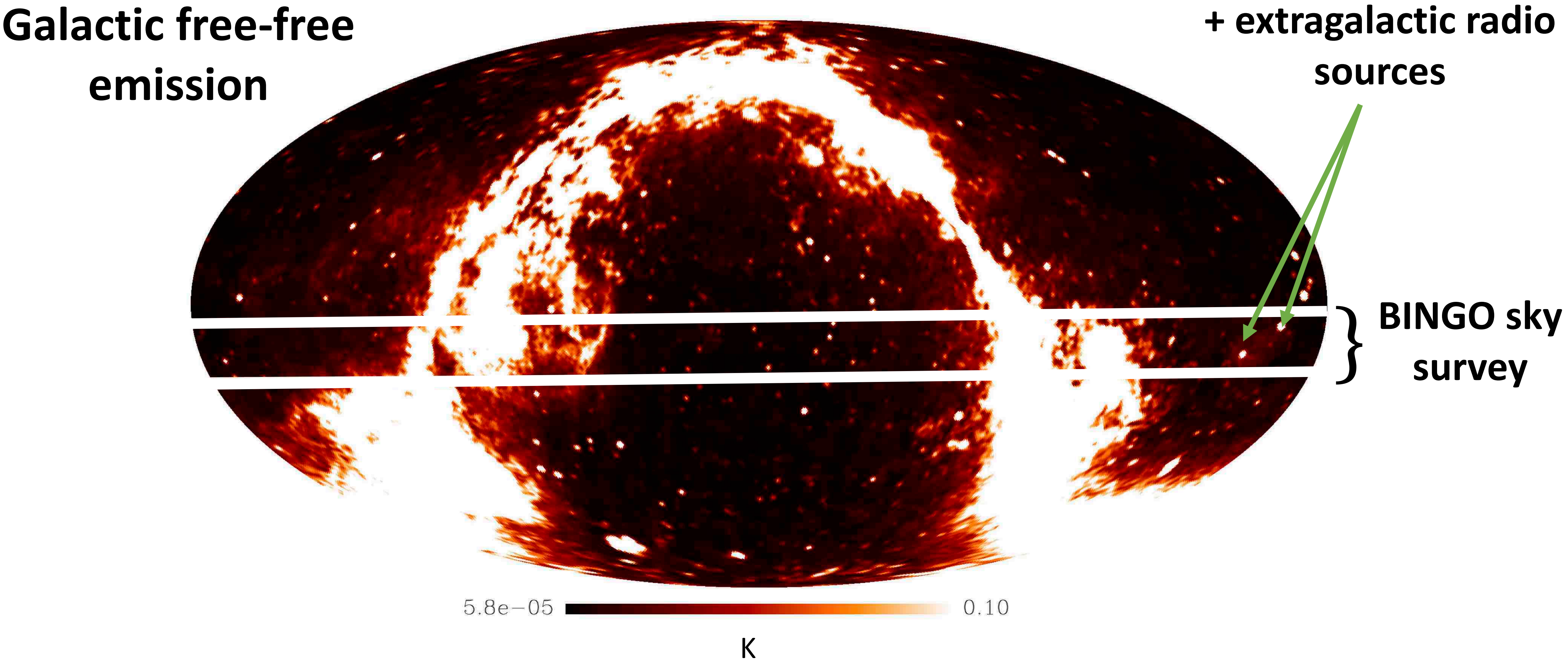
M. Remazeilles

Temperature fluctuations of the 21-cm signal



Liccardo et al, arXiv:2107.01636

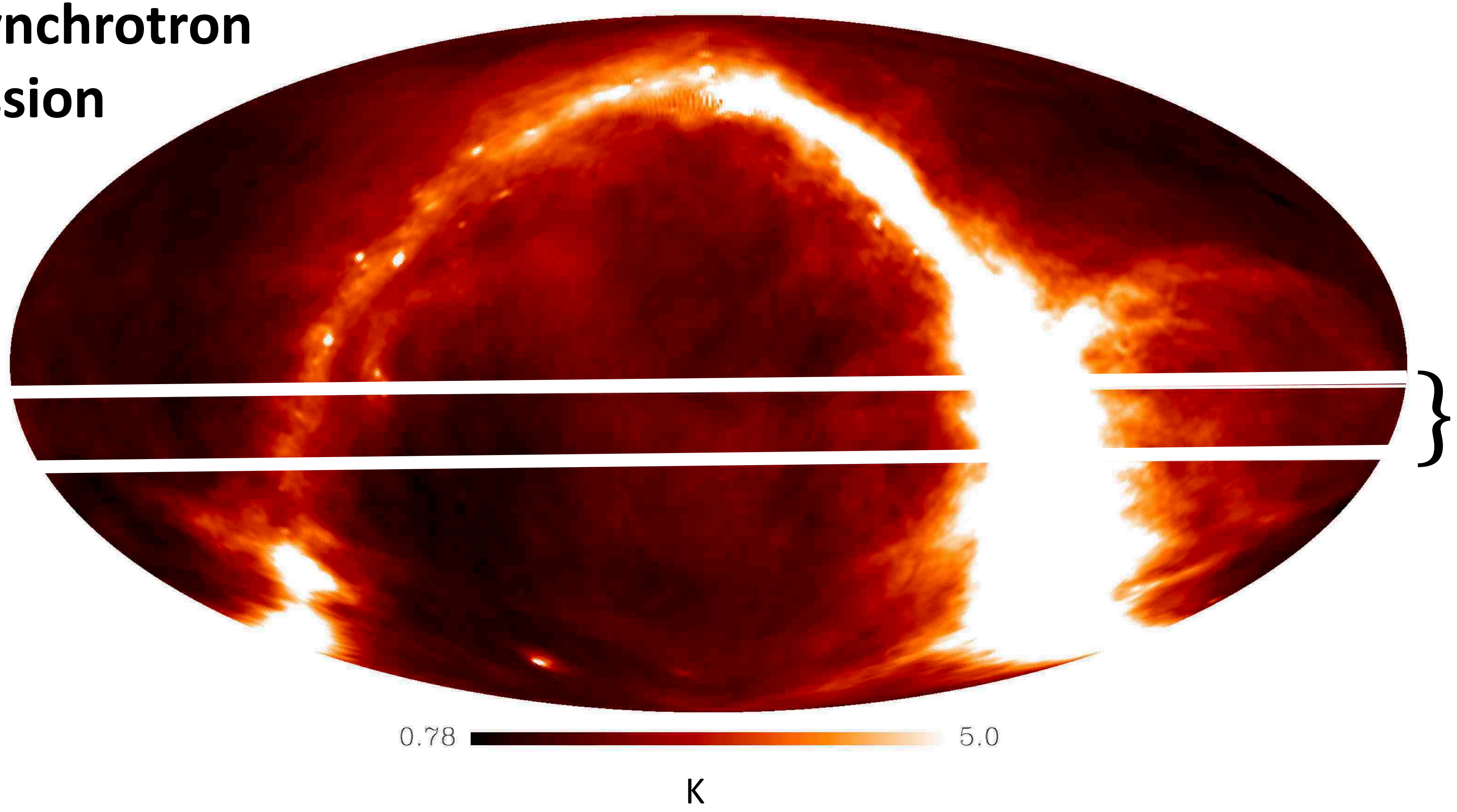
Astrophysical foregrounds obscure the signal



Liccardo et al, arXiv:2107.01636

Astrophysical foregrounds obscure the signal

Galactic synchrotron emission



BINGO sky survey

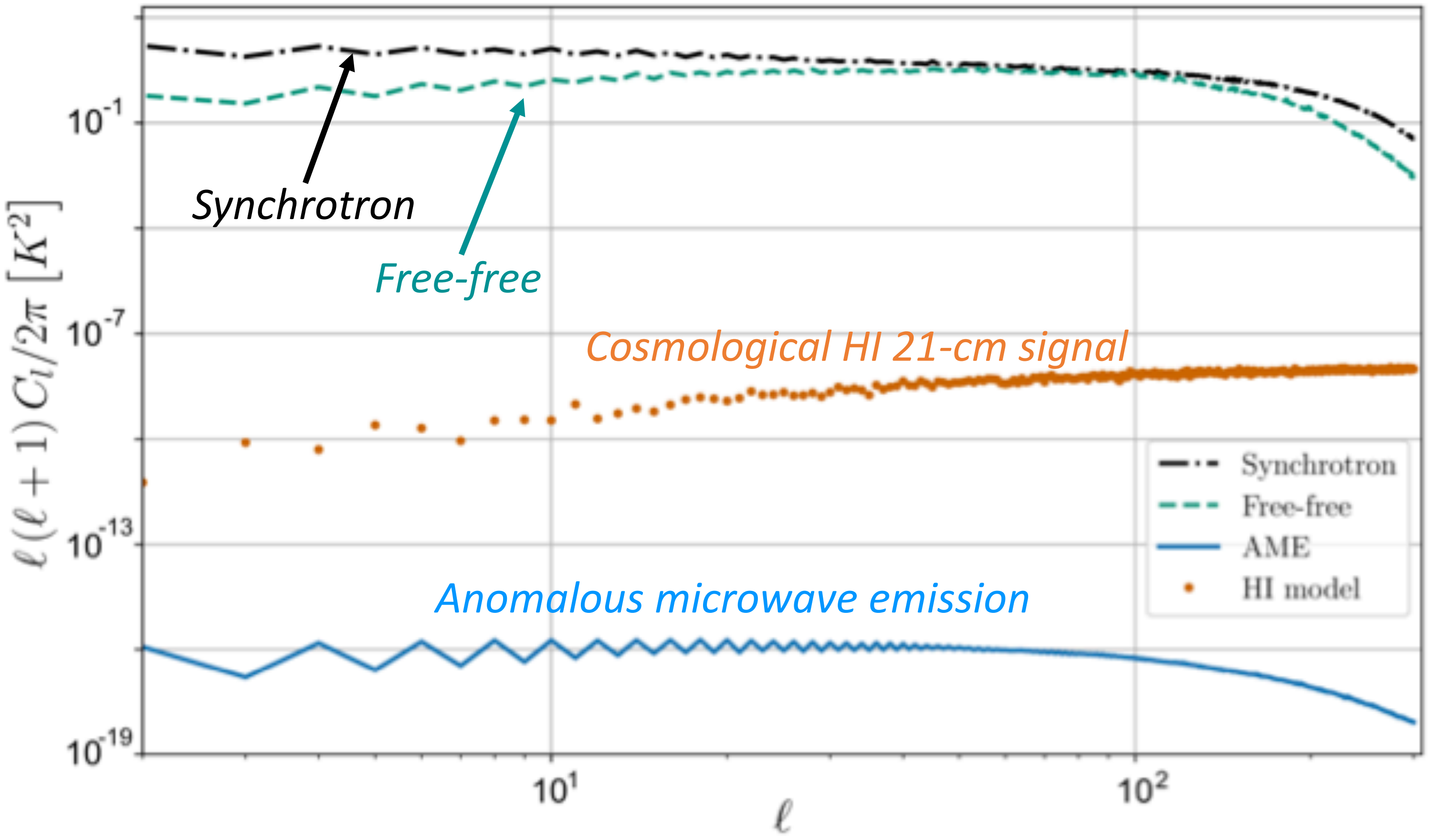
Liccardo et al, arXiv:2107.01636

Astrophysical foregrounds obscure the signal

Huge amplitude discrepancies!

8 to 9 orders of magnitude!

$\nu = 1100$ MHz / $z = 0.29$



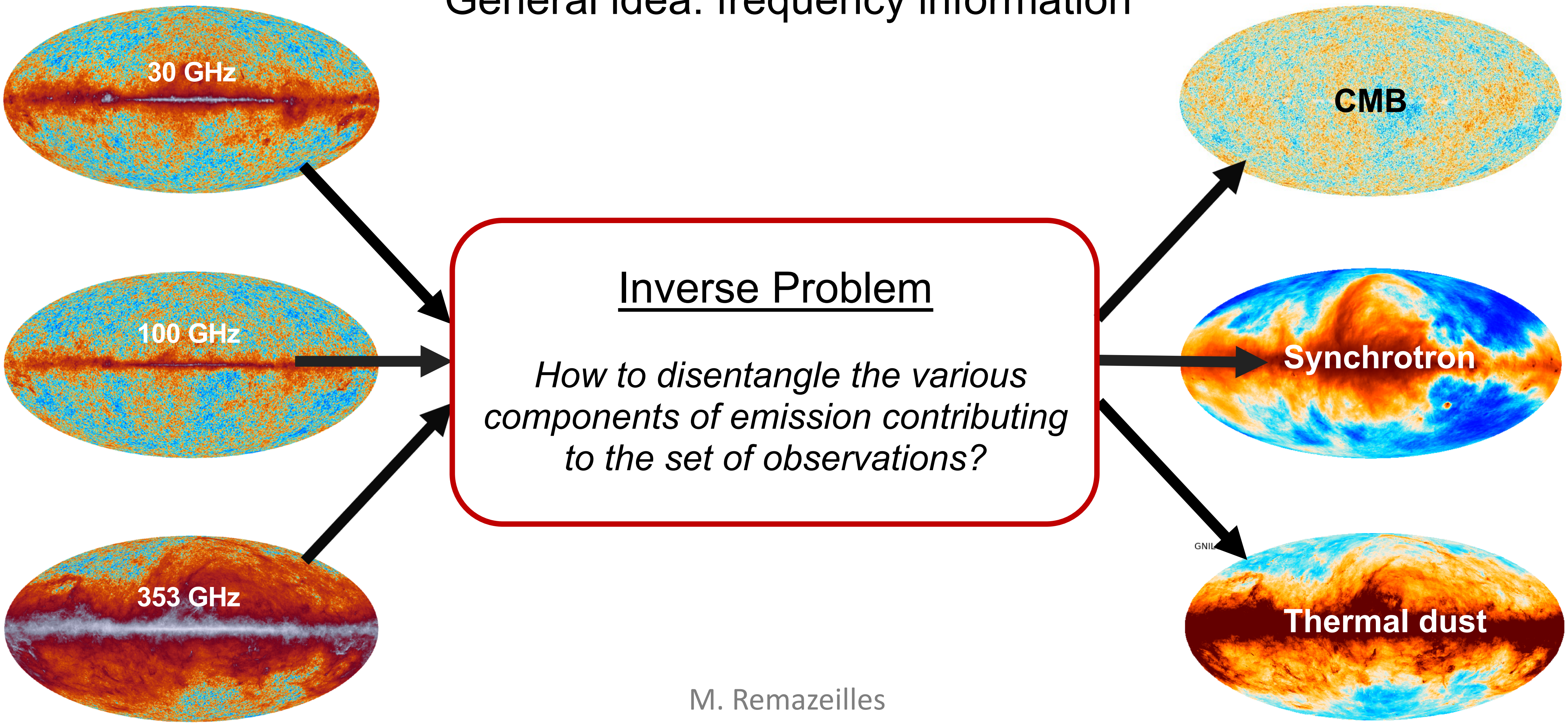
Liccardo et al, arXiv:2107.01636

Component Separation

Similar challenges to that of CMB data

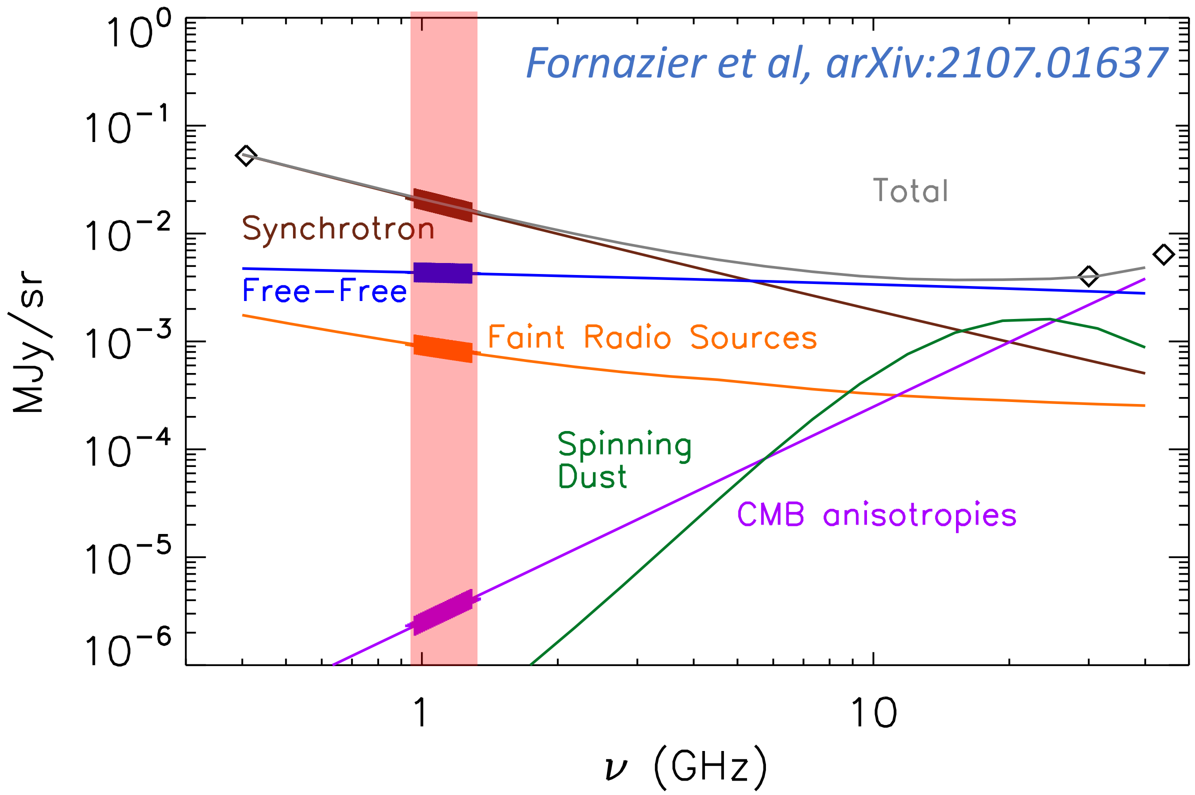
General idea: frequency information

Multi-frequency data

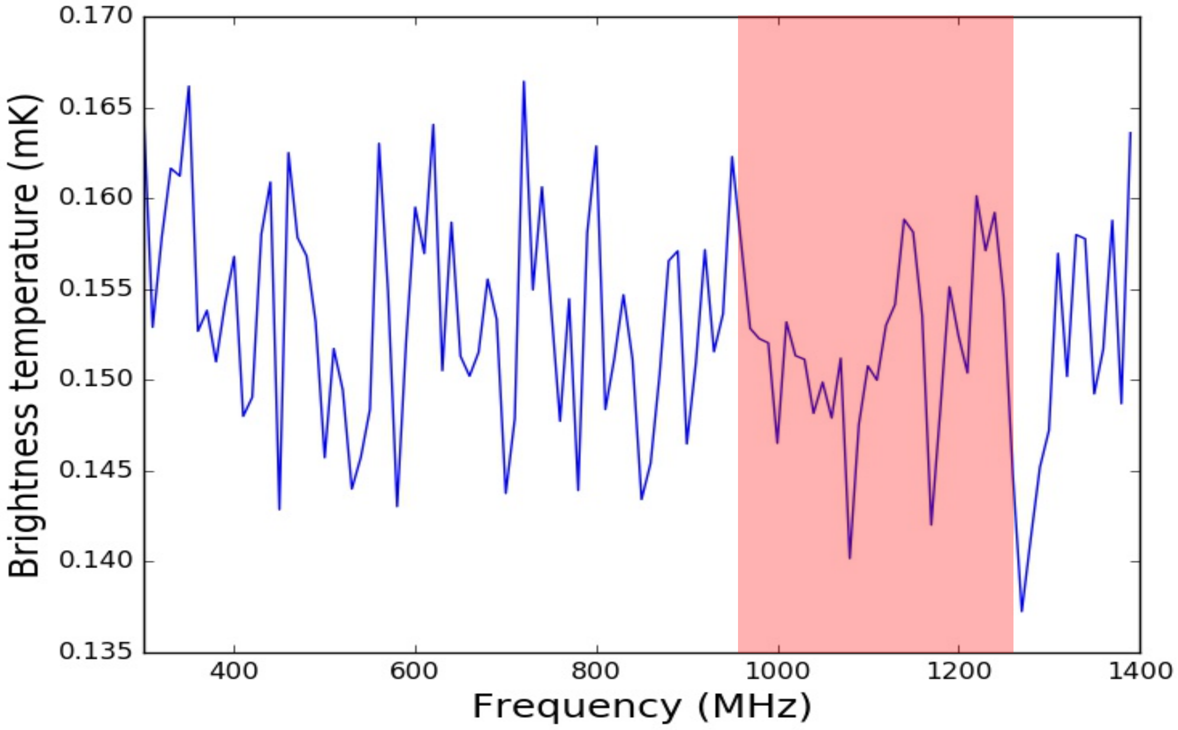


Distinct spectral signatures

Astrophysical foregrounds



Cosmological 21-cm signal



Multi-frequency BINGO observations should allow to disentangle cosmological 21-cm signal and astrophysical foregrounds

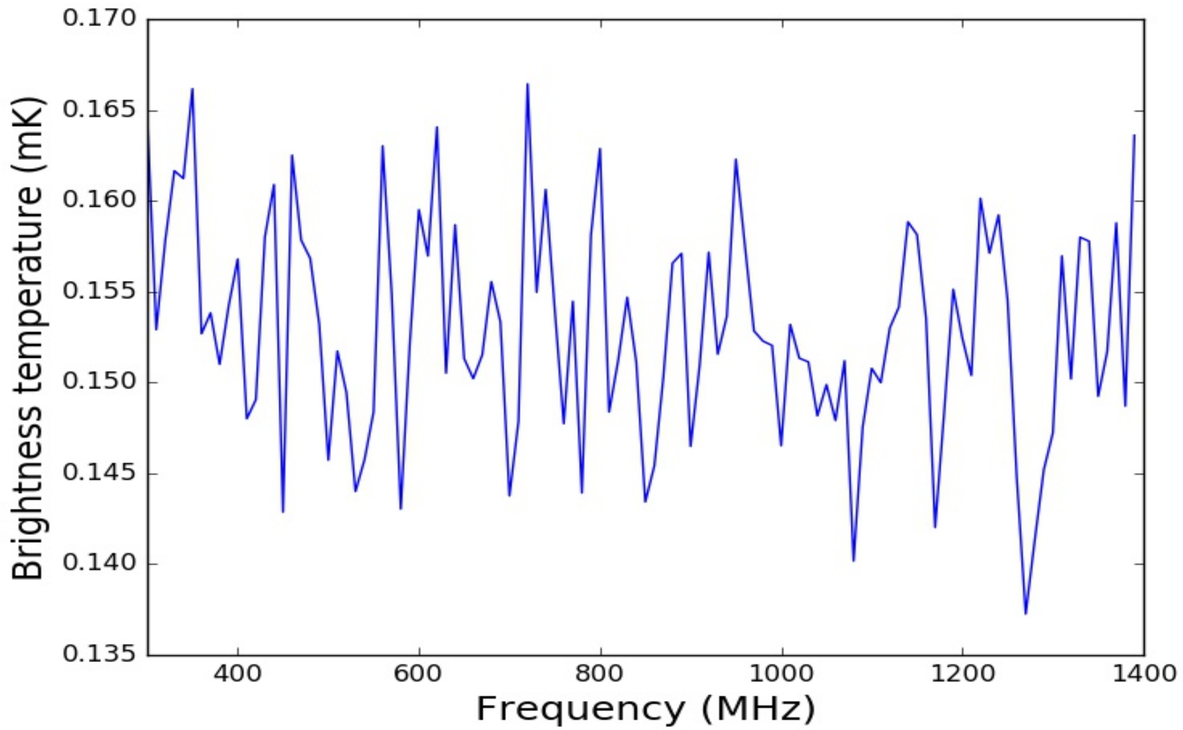
Peculiarities of 21-cm component separation

In contrast to CMB, the spectral signature of the 21-cm signal is unknown/random!

→ 21-cm component separation methods reduce to foreground subtraction techniques

→ risk of losing part of the 21-cm signal by oversubtracting foregrounds

Cosmological 21-cm signal



We need to think beyond spectral modelling to extract the 21-cm signal!

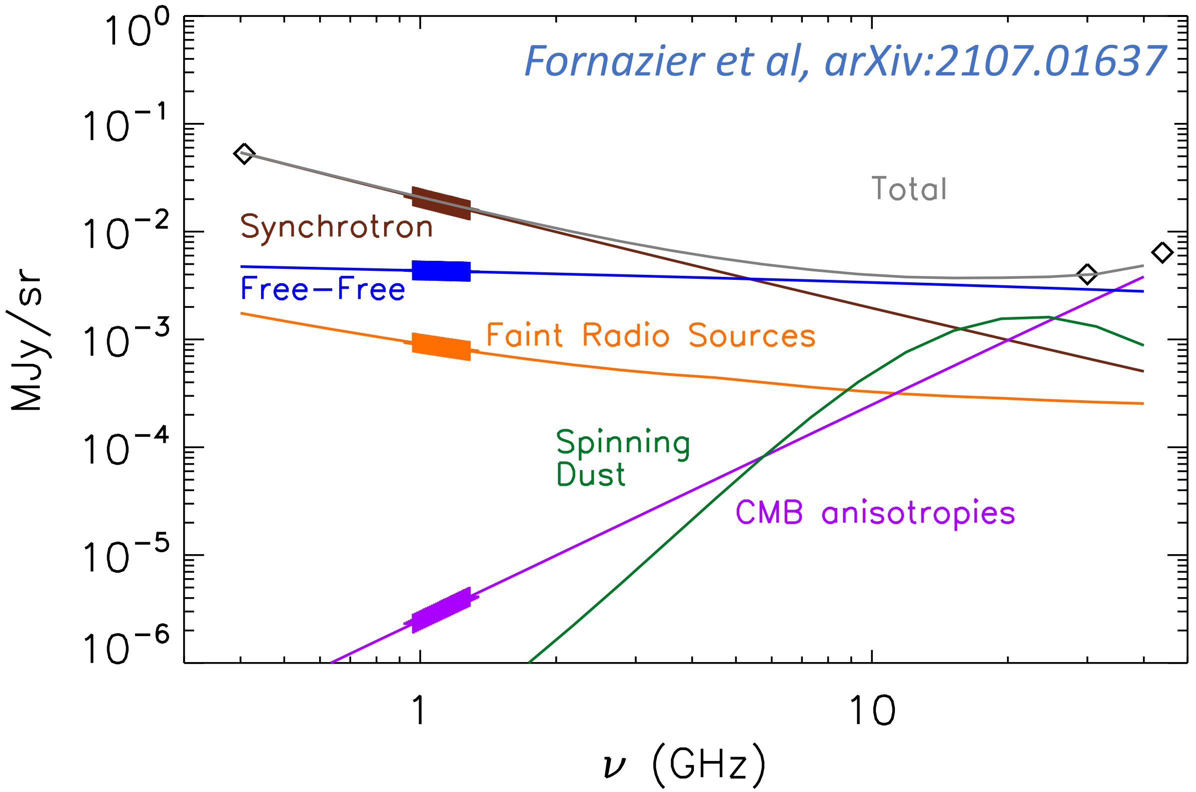
Peculiarities of 21-cm component separation

The so-called "spectral smoothness" of the foregrounds is a myth!

→ *Telescope systematics (e.g. standing waves) break the "smoothness" of the foregrounds*

→ *Given the huge amplitude discrepancy between foregrounds and 21-cm signal, any small mismodeling of the foregrounds will result in large biases on the 21-cm signal*

Astrophysical foregrounds

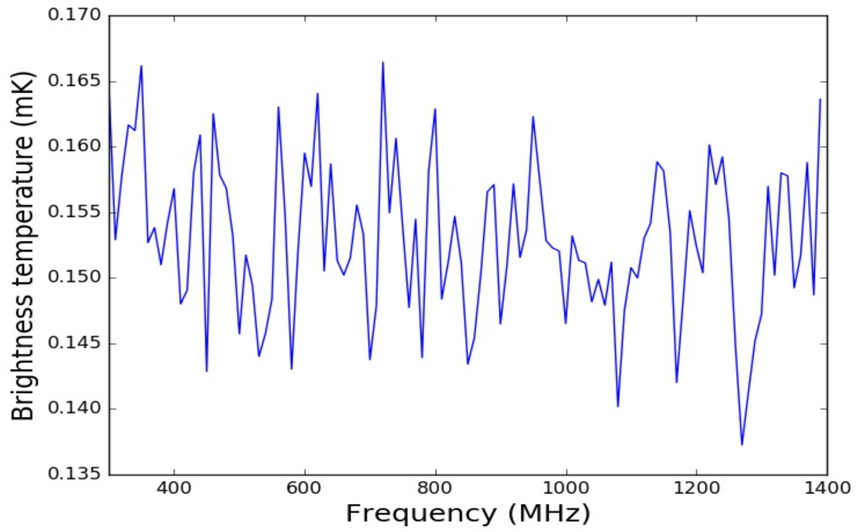


We need to avoid making strong assumptions about foregrounds when the targeted signal is several orders of magnitude lower!

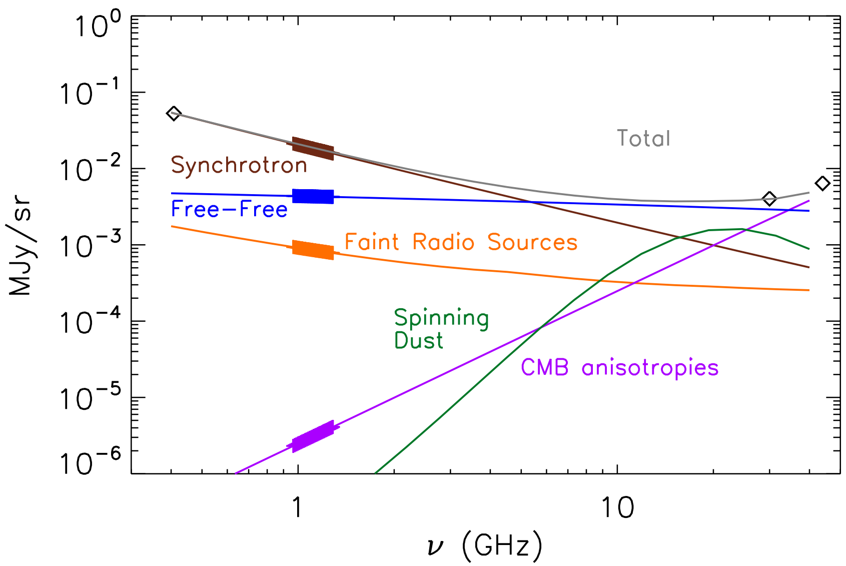
Peculiarities of 21-cm component separation

The 21-cm signal is mostly decorrelated between frequencies, while foreground emissions are strongly correlated across frequencies

Cosmological 21-cm signal



Astrophysical foregrounds



(De)correlation properties should be exploited to discriminate between foregrounds and 21-cm signal!

M. Remazeilles

GNILC

Remazeilles, Delabrouille, Cardoso, MNRAS 2011
Olivari, Remazeilles, Dickinson, MNRAS 2016

GNILC ("Generalized Needlet Internal Linear Combination") is an extension of the blind ILC method which allows

- **to break spectral degeneracies**

  e.g. cosmic infrared background (CIB) and Galactic thermal dust emissions

- **to overcome lack of spectral information**

  e.g. cosmological 21-cm line emission

GNILC

Remazeilles, Delabrouille, Cardoso, MNRAS 2011
Olivari, Remazeilles, Dickinson, MNRAS 2016

- **Use statistical / spatial information** (power spectrum) to compensate any lack of spectral information (e.g. unknown SED, spectral degeneracies)

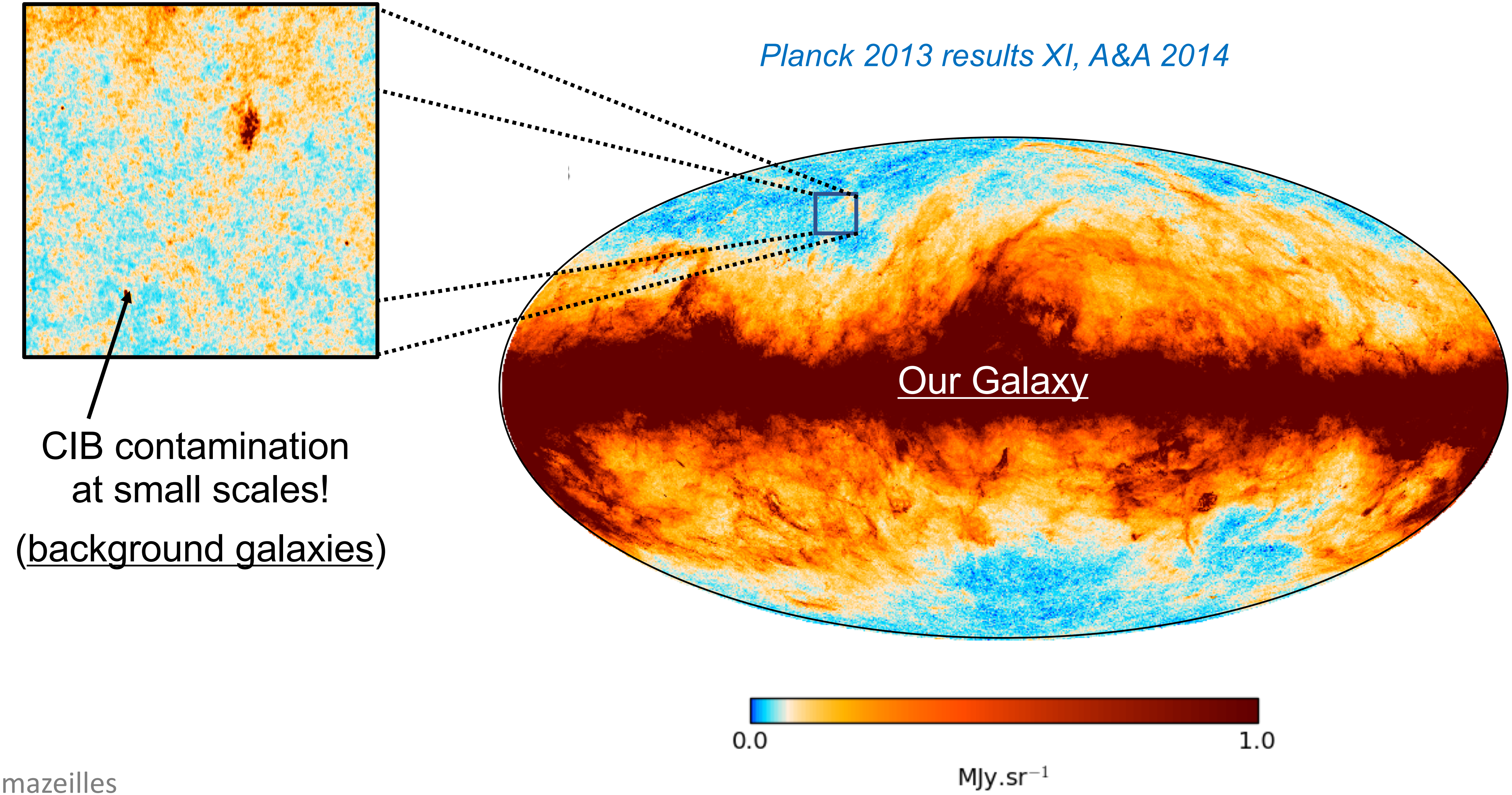
- **Blind,** i.e. no assumption about astrophysical foregrounds

  Sole prior assumption: power spectrum of the cosmological signal

- **Wavelet-based**

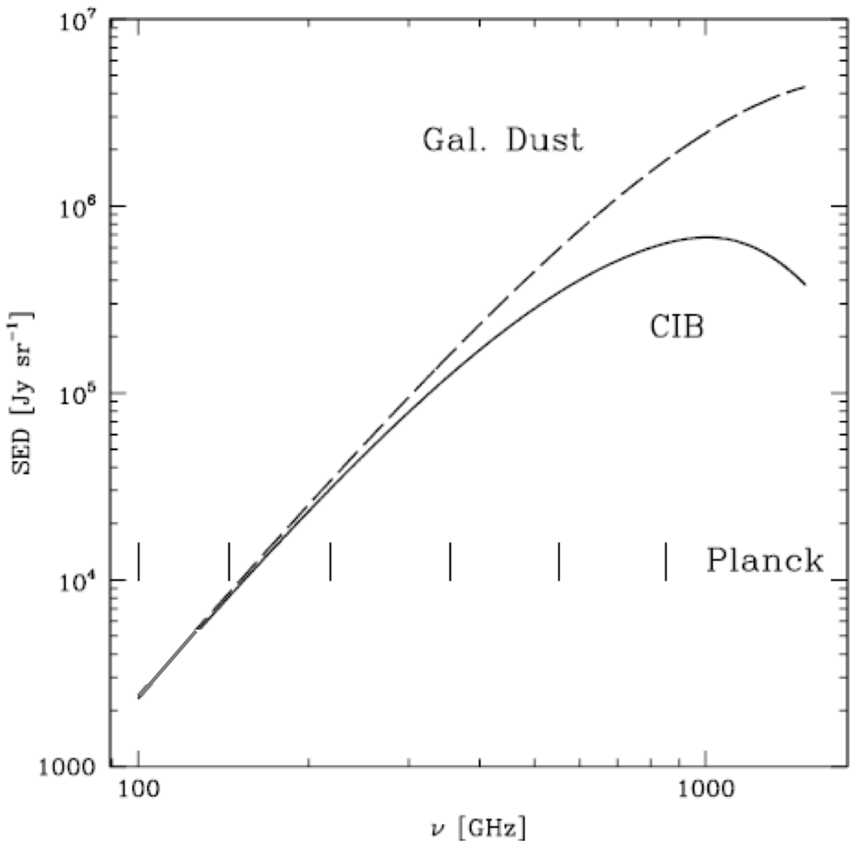
  Allows to optimize component separation depending on the local variations of foregrounds and noise both across the sky and across angular scales

Planck 2013 map of Galactic dust



Dust-CIB spectral degeneracy

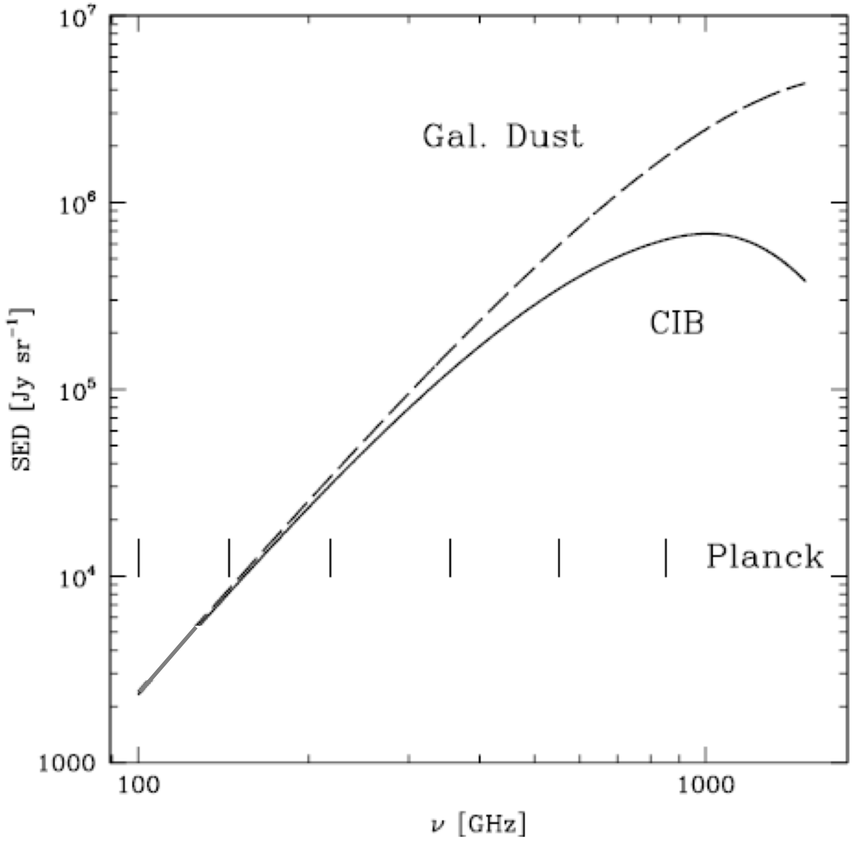
CIB and thermal dust have similar spectral signatures (modified blackbody)



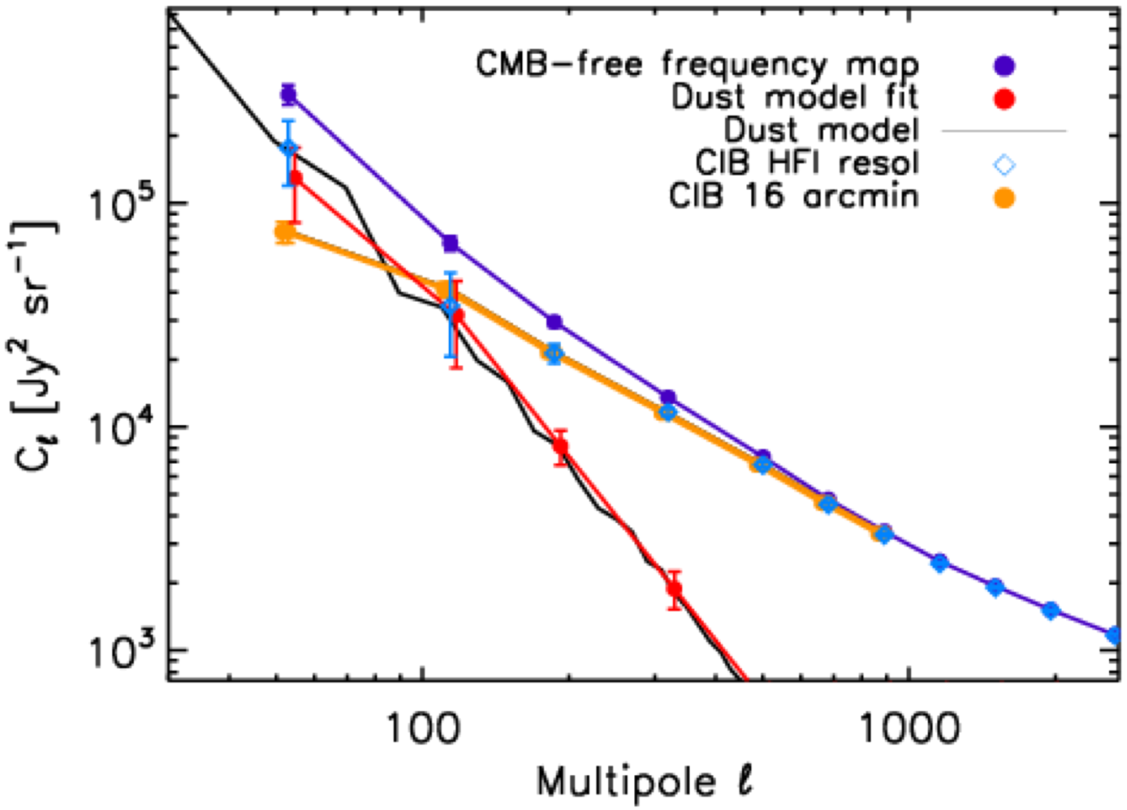
- Fitting a modified blackbody spectrum to *Planck* multi-frequency data can't help to disentangle thermal dust and CIB emissions
- GNILC goes beyond spectral modelling for component separation
- Unlike other methods which rely solely on spectral information, GNILC uses statistical information to discriminate dust and CIB

Breaking the dust-CIB spectral degeneracy

CIB and thermal dust have similar spectral signatures (modified blackbody)

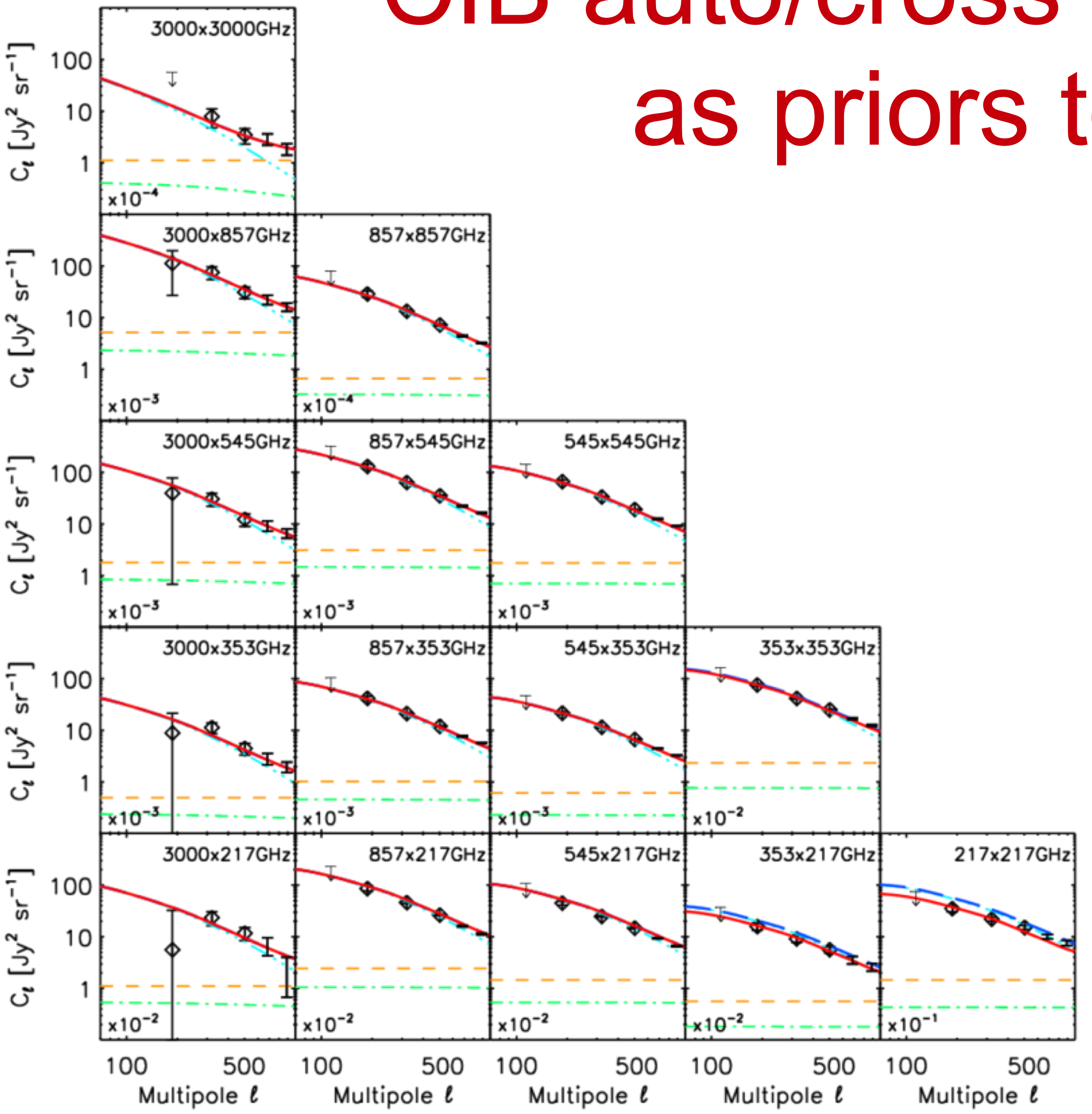


But thermal dust and CIB have distinct angular power spectra!



M. Remazeilles

CIB auto/cross power spectra as priors to GNILC



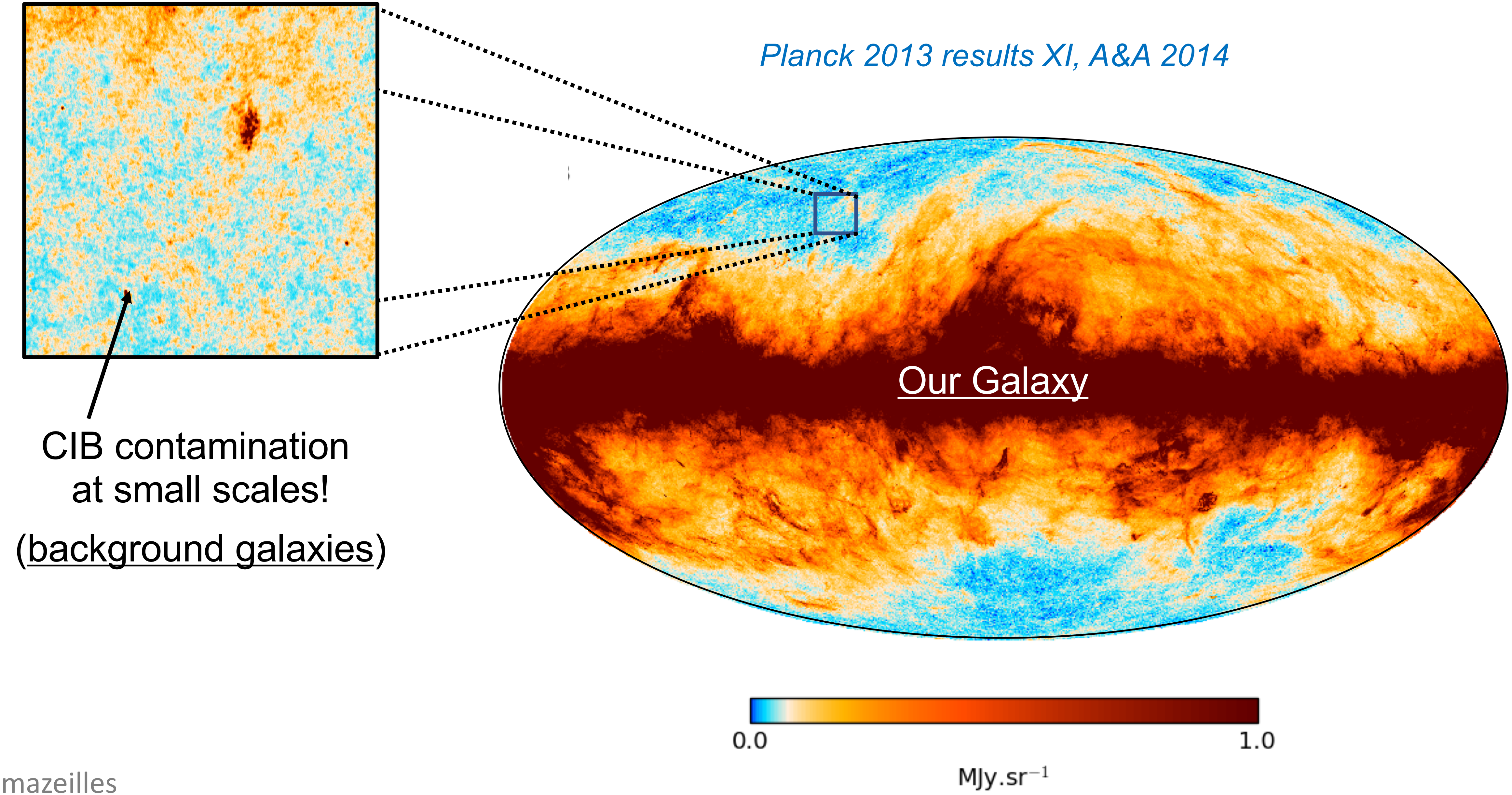
The statistics of CIB is significantly different from that of Galactic dust

(no assumption about Galactic dust)

Planck 2013 results XXX, A&A 2014

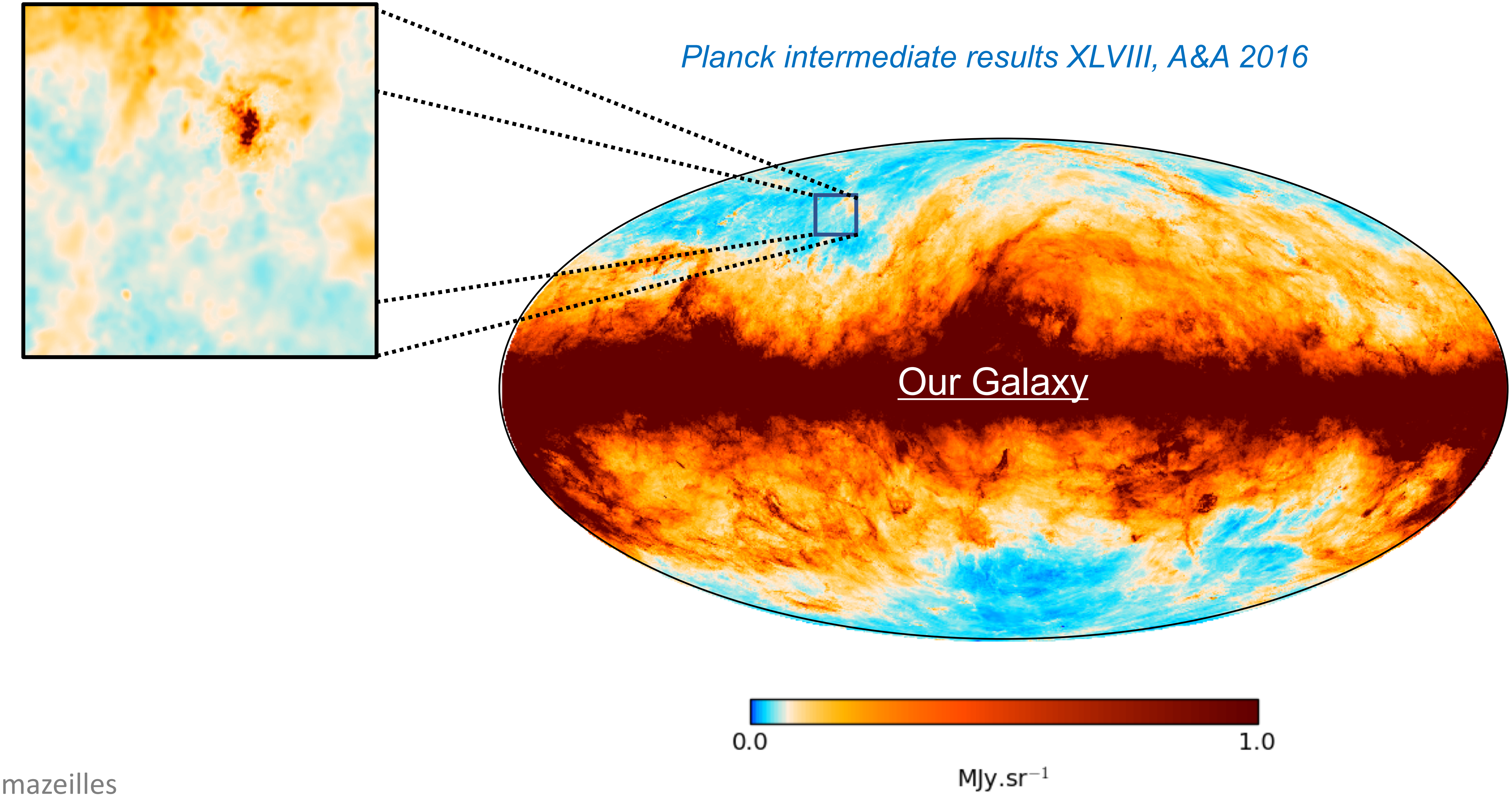
M. Remazeilles

Planck 2013 map of Galactic dust



Planck GNILC map of Galactic dust

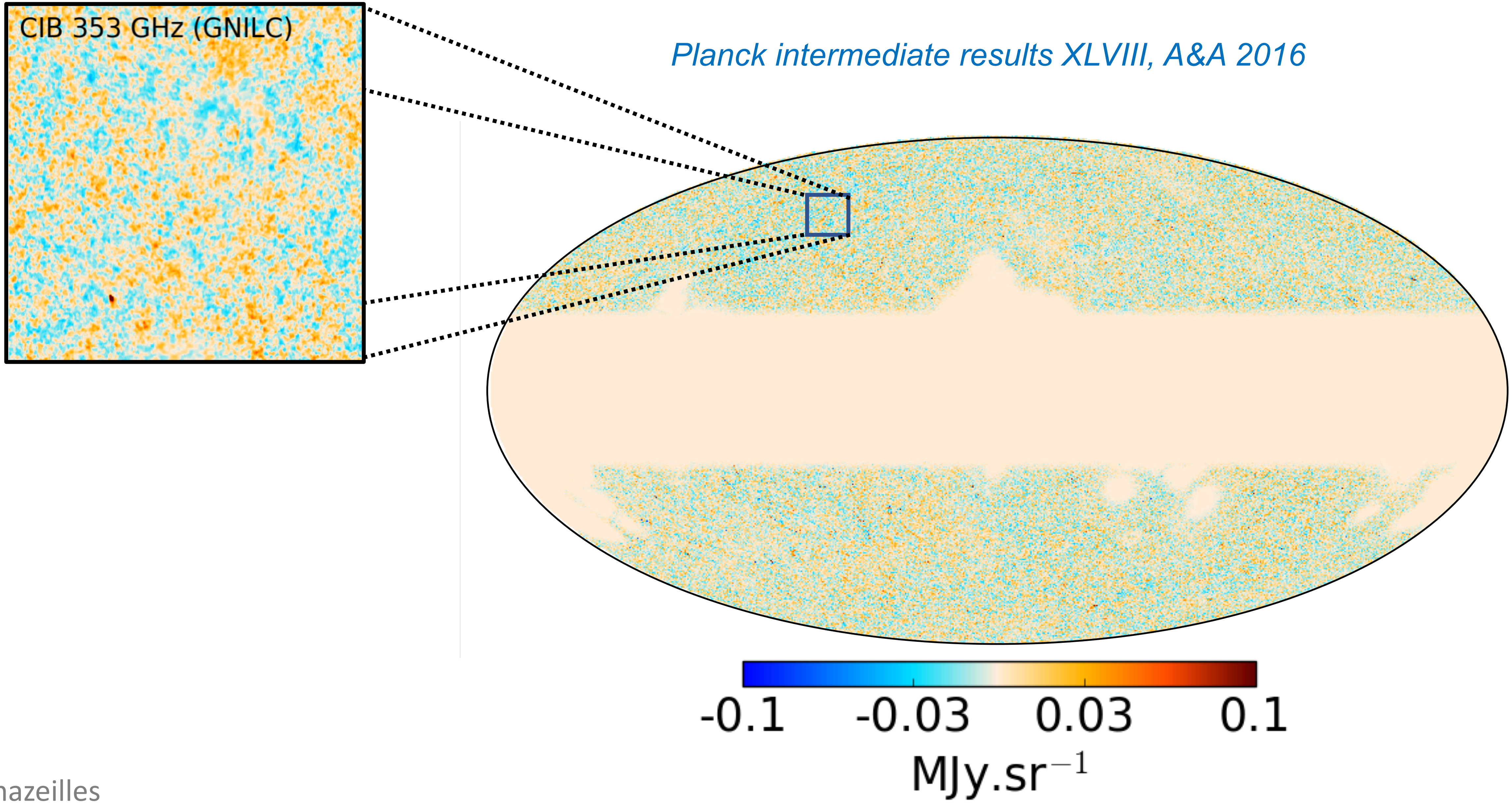
Planck intermediate results XLVIII, A&A 2016



M. Remazeilles

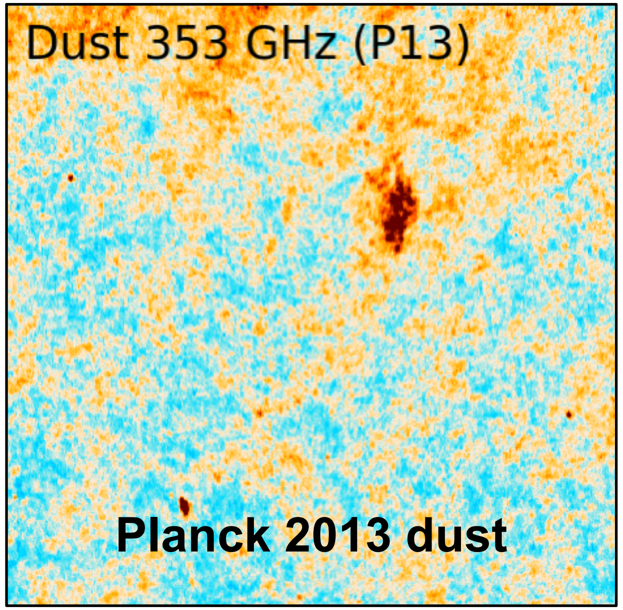
Planck GNILC map of CIB fluctuations

Planck intermediate results XLVIII, A&A 2016

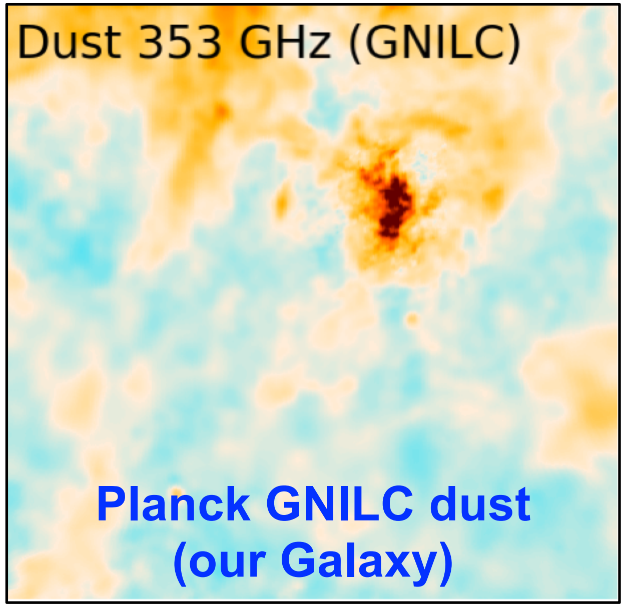


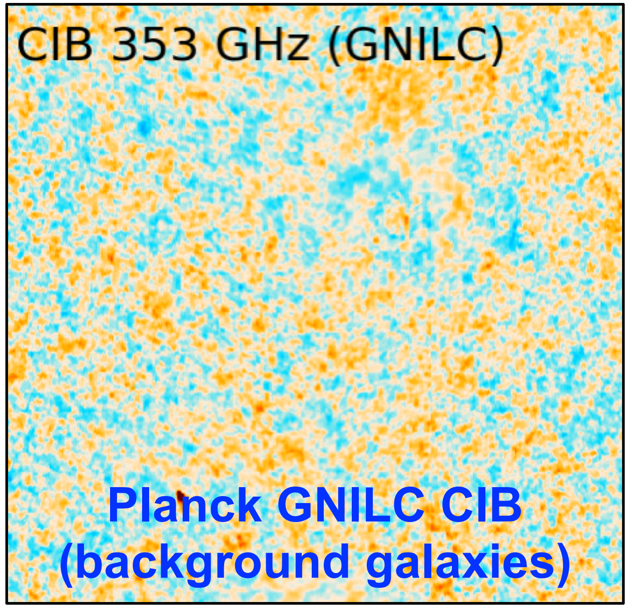
M. Remazeilles

GNILC disentangles Galactic dust and CIB



GNILC





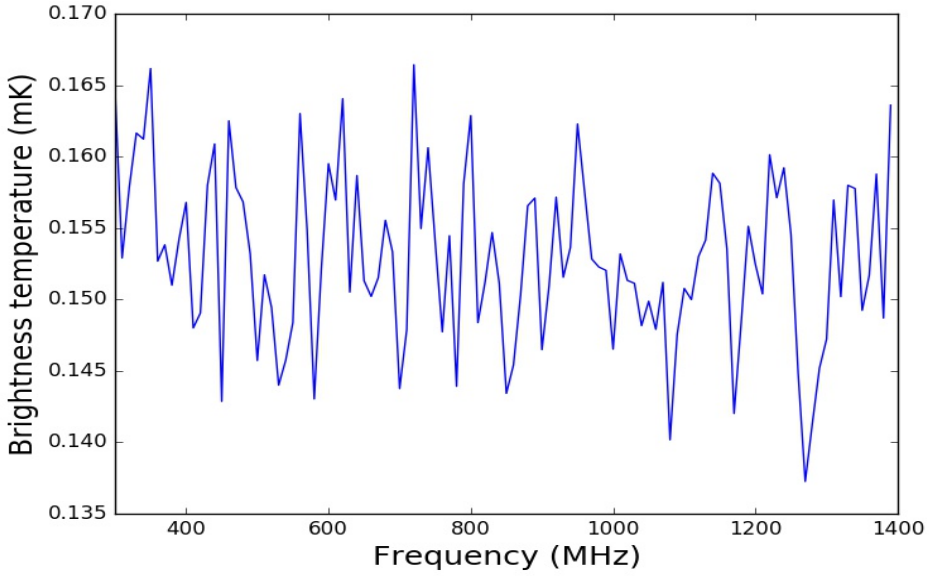
Planck intermediate results XLVIII, A&A (2016)

Remazeilles, Delabrouille, Cardoso, MNRAS (2011)

GNILC for 21-cm intensity mapping

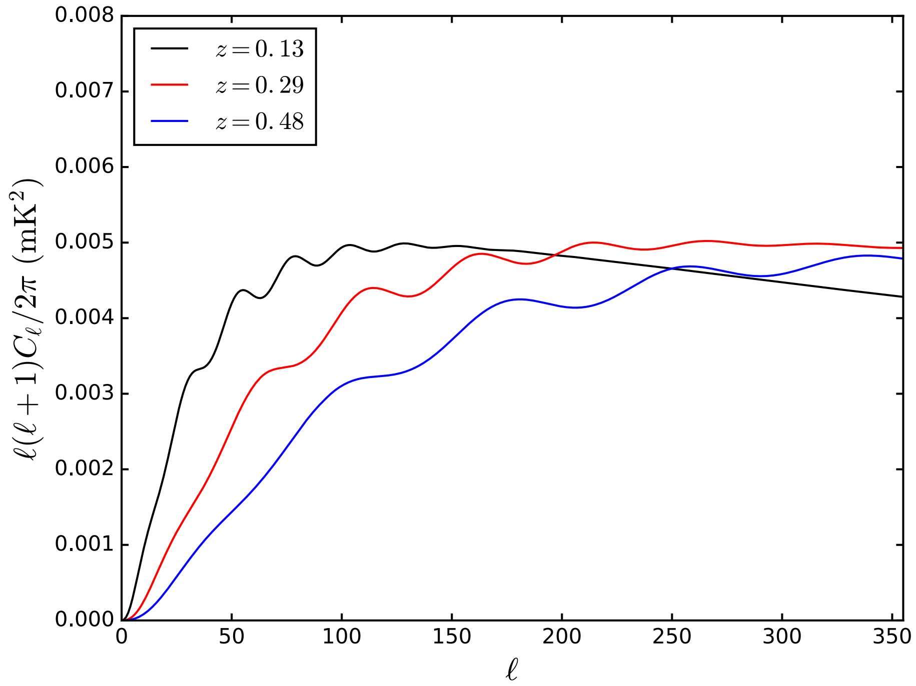
Olivari, Remazeilles, Dickinson, MNRAS 2016

Non-trivial spectral response (SED) of 21-cm signal (mostly decorrelated across frequencies)



GNILC →

Use prior information on 21-cm power spectrum



M. Remazeilles

GNILC in 6 main steps

GNILC

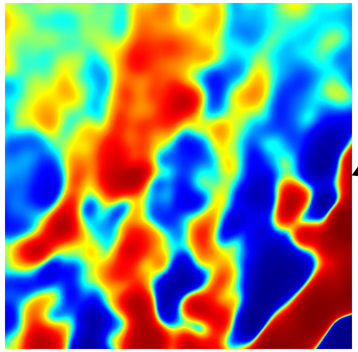
Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

1. Needlet (spherical wavelet) decomposition of the BINGO sky maps

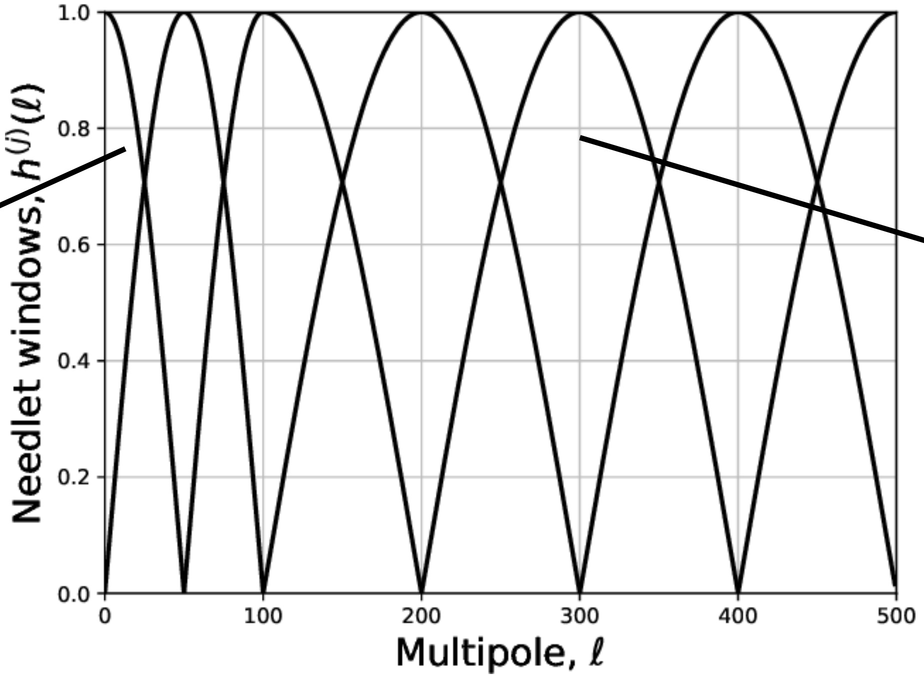
$$d_\nu(\vec{n}) \xrightarrow{\text{SHT}} d_\nu(\ell, m) \xrightarrow{\times h^{(j)}(\ell)} d_\nu(\ell, m) \times h^{(j)}(\ell) \xrightarrow{\text{SHT}^{-1}} d_\nu^{(j)}(\vec{n})$$

Bandpass filtering in harmonic space through needlet windows

Wavelet scale: 1

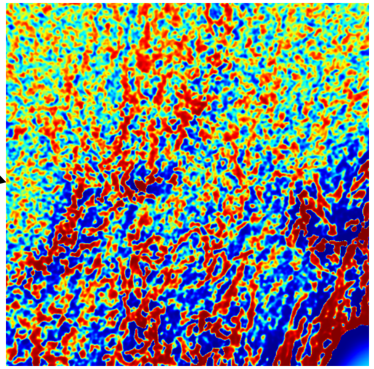


$\times h^{(1)}(\ell)$



$\times h^{(5)}(\ell)$

Wavelet scale: 5



Component separation performed locally both across the sky and across the scales

M. Remazeilles

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

2. **For each needlet scale (j) and pixel $\vec{n}$, compute the data covariance matrix across all pairs of frequencies a, b**

$$\boldsymbol{C}_{ab}^{(j)}(\vec{\boldsymbol{n}}) = \sum_{\vec{\boldsymbol{n}}' \in \mathfrak{D}(\vec{\boldsymbol{n}})} \boldsymbol{d}_a^{(j)}(\vec{\boldsymbol{n}}')\, \boldsymbol{d}_b^{(j)}(\vec{\boldsymbol{n}}')$$

For each pixel $\vec{n}$ and scale (j), $\boldsymbol{C}^{(j)}(\vec{\boldsymbol{n}})$ is a $\boldsymbol{N}\times\boldsymbol{N}$ matrix,

where $\boldsymbol{N}$ is the number of frequency channels

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

3. Use theoretical priors on 21cm signal power spectra, $C_\ell^{\rm 21cm,prior}(\nu)$, across frequencies/redshifts to model the signal covariance matrix

- Use priors $C_\ell^{\rm 21cm,prior}(\nu)$ to simulate realisations of 21-cm signal maps $s_\nu^{\rm prior}(\vec{n})$
- Similarly to the data, the prior 21-cm realisations go through needlet decomposition: $s_\nu^{\rm prior}(\vec{n}) \rightarrow s_\nu^{{\rm prior},(j)}(\vec{n})$
- For each needlet scale (j) and each pixel $\vec{n}$, compute the **prior 21-cm signal covariance matrix:**

$$S_{ab}^{{\rm prior},(j)}(\vec{n}) = \sum_{\vec{n}' \in \mathfrak{D}(\vec{n})} s_a^{{\rm prior},(j)}(\vec{n}')\, s_b^{{\rm prior},(j)}(\vec{n}')$$

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

4. Eigenvalue decomposition of the whitened data covariance matrix:

$$C = S + N \Longrightarrow \left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq I + \left(S^{\text{prior}}\right)^{-\frac{1}{2}} N \left(S^{\text{prior}}\right)^{-\frac{1}{2}}$$

C: data; S: signal; N: foregrounds + noise

$$\left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq \left(U_N | U_S\right) \left(\begin{array}{ccc|ccc} 1+\lambda_1 & & & & & \\ & \ddots & & & & \\ & & 1+\lambda_m & & & \\ \hline & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{array}\right) \left(\frac{U_N^T}{U_S^T}\right)$$

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

4. Eigenvalue decomposition of the whitened data covariance matrix:

$$C = S + N \Longrightarrow \left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq I + \left(S^{\text{prior}}\right)^{-\frac{1}{2}} N \left(S^{\text{prior}}\right)^{-\frac{1}{2}}$$

C → data, S → signal, N → foregrounds + noise

$$\left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq \left(U_N | U_S\right) \left(\begin{array}{ccc|ccc} 1+\lambda_1 & & & & & \\ & \ddots & & & & \\ & & 1+\lambda_m & & & \\ \hline & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{array}\right) \left(\frac{U_N^T}{U_S^T}\right)$$

Those eigenvalues do not contain relevant foreground power ($\boldsymbol{C} \simeq \boldsymbol{I}$). Data are consistent with **21-cm signal**.

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

4. Eigenvalue decomposition of the whitened data covariance matrix:

$$C = S + N \Longrightarrow \left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq I + \left(S^{\text{prior}}\right)^{-\frac{1}{2}} N \left(S^{\text{prior}}\right)^{-\frac{1}{2}}$$

C: data; S: signal; N: foregrounds + noise

$$\left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq \left(U_N | U_S\right) \left(\begin{array}{ccc|ccc} 1+\lambda_1 & & & & & \\ & \ddots & & & & \\ & & 1+\lambda_m & & & \\ \hline & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{array}\right) \left(\frac{U_N^T}{U_S^T}\right)$$

Those eigenvectors form a basis of independent 21-cm modes

Those eigenvalues do not contain relevant foreground power ($\boldsymbol{C} \simeq \boldsymbol{I}$).
Data are consistent with **21-cm signal**.

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

4. Eigenvalue decomposition of the whitened data covariance matrix:

$$C = S + N \Longrightarrow \left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq I + \left(S^{\text{prior}}\right)^{-\frac{1}{2}} N \left(S^{\text{prior}}\right)^{-\frac{1}{2}}$$

C → data; S → signal; N → foregrounds + noise

$$\left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq \left(U_N | U_S\right) \left(\begin{array}{ccc|ccc} 1+\lambda_1 & & & & & \\ & \ddots & & & & \\ & & 1+\lambda_m & & & \\ \hline & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{array}\right) \left(\frac{U_N^T}{U_S^T}\right)$$

Those eigenvalues contain significant power from the **foregrounds**.

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

4. Eigenvalue decomposition of the whitened data covariance matrix:

$$C = S + N \Longrightarrow \left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq I + \left(S^{\text{prior}}\right)^{-\frac{1}{2}} N \left(S^{\text{prior}}\right)^{-\frac{1}{2}}$$

C → data; S → signal; N → foregrounds + noise

$$\left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq \left(U_N | U_S\right) \left(\begin{array}{ccc|ccc} 1+\lambda_1 & & & & & \\ & \ddots & & & & \\ & & 1+\lambda_m & & & \\ \hline & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{array}\right) \left(\frac{U_N^T}{U_S^T}\right)$$

Those eigenvectors form an orthonormal basis of ***m*** independent foreground modes. (**"foreground subspace"**)

Those eigenvalues contain significant power from the **foregrounds**.

M. Remazeilles

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

4. Eigenvalue decomposition of the whitened data covariance matrix:

$$C = S + N \Longrightarrow \left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq I + \left(S^{\text{prior}}\right)^{-\frac{1}{2}} N \left(S^{\text{prior}}\right)^{-\frac{1}{2}}$$

C → data; S → signal; N → foregrounds + noise

$$\left(S^{\text{prior}}\right)^{-\frac{1}{2}} C \left(S^{\text{prior}}\right)^{-\frac{1}{2}} \simeq \left(U_N | U_S\right) \left(\begin{array}{ccc|ccc} 1+\lambda_1 & & & & & \\ & \ddots & & & & \\ & & 1+\lambda_m & & & \\ \hline & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{array}\right) \left(\frac{U_N^T}{U_S^T}\right)$$

Those eigenvectors form an orthonormal basis of ***m*** independent foreground modes. (**"foreground subspace"**)

As highly correlated components of emission, foregrounds can thus be decomposed on a subset of m independent templates

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

5. **For each needlet scale (j) and pixel $\vec{n}$, estimate the effective dimension $m \equiv m_{\mathrm{AIC}}^{(j)}(\vec{n})$ of the foreground subspace using Akaike Information Criterion (AIC)**

$\boldsymbol{m}_{\mathbf{AIC}}^{(j)}(\vec{\boldsymbol{n}})$ is the minimizer of

$$\mathbf{AIC}[\boldsymbol{m}] = \mathbf{2}\boldsymbol{m} + \sum_{\boldsymbol{i}=\boldsymbol{m}+\mathbf{1}}^{\boldsymbol{N}} (\boldsymbol{\mu}_{\boldsymbol{i}} - \mathbf{log}\,\boldsymbol{\mu}_{\boldsymbol{i}} - \mathbf{1})$$

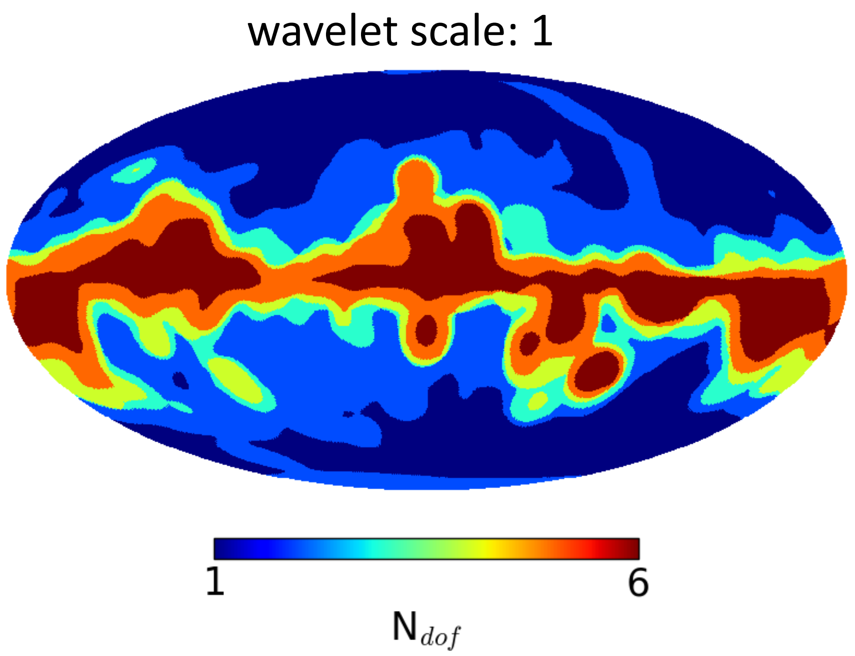
where $\{\mu_i\}_{1\leq i\leq N}$ are the eigenvalues of matrix $\left(S^{\mathrm{prior}}\right)^{-\frac{1}{2}} C \left(S^{\mathrm{prior}}\right)^{-\frac{1}{2}}$

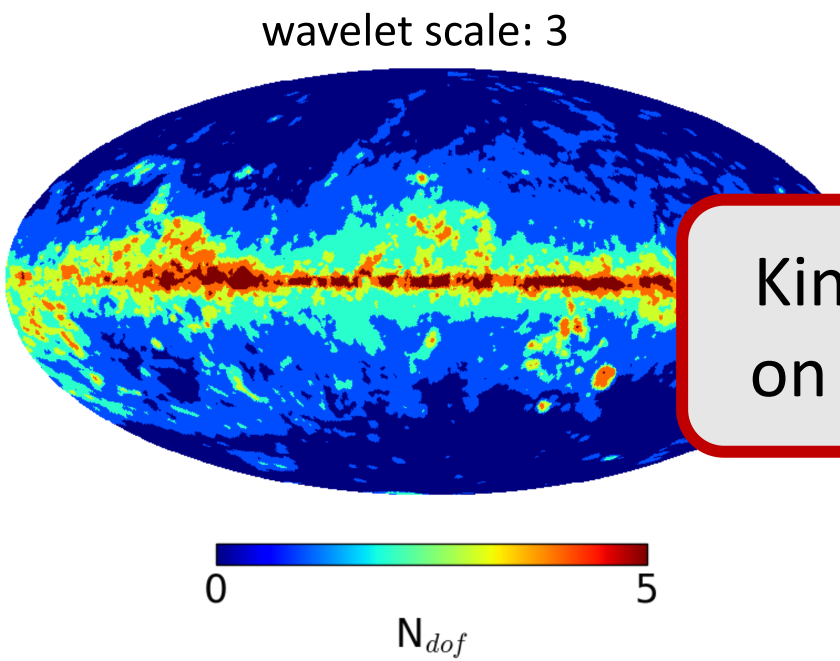
GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

5. For each needlet scale (j) and pixel $\vec{n}$, estimate the effective dimension $m \equiv m_{\rm AIC}^{(j)}(\vec{n})$ of the foreground subspace using Akaike Information Criterion (AIC)

Effective number $m_{\rm AIC}$ of foreground components





Kind of PCA on steroids!

GNILC

Remazeilles et al MNRAS 2011
Olivari et al MNRAS 2016

6. Perform a $(\mathrm{N} - m_{\mathrm{AIC}})$-dimensional ILC in the "21-cm signal subspace" to estimate the GNILC 21-cm maps

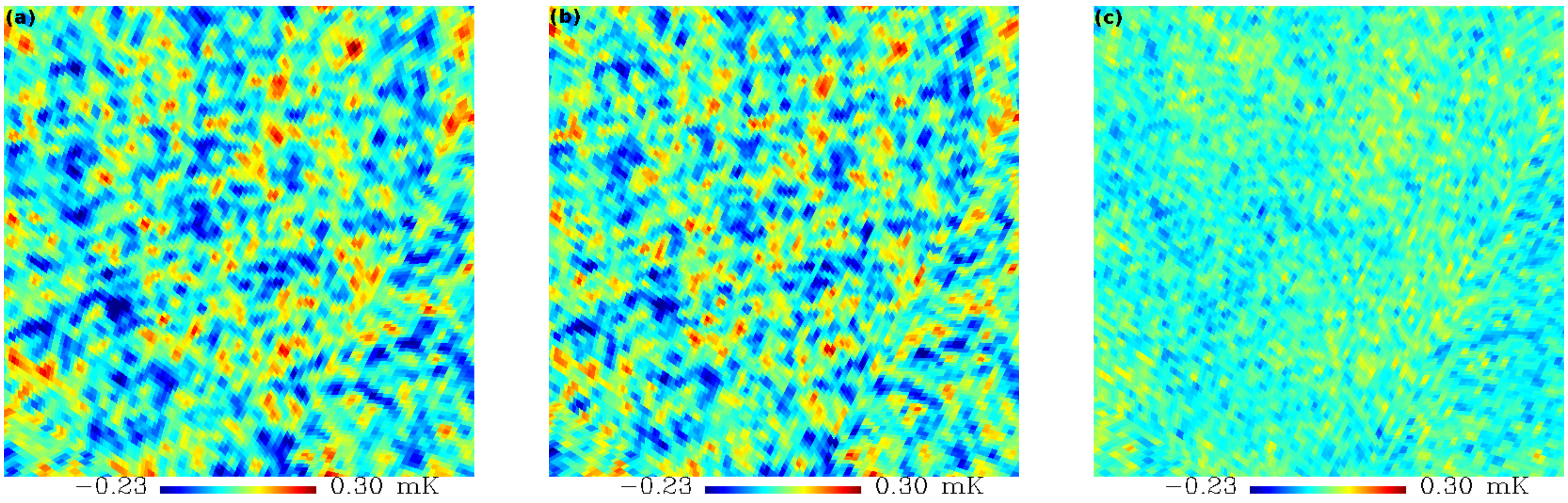
$$\hat{s}_{\nu}^{\mathrm{GNILC}}(\vec{n}) = \sum_{\nu'} W(\nu, \nu')\, d_{\nu'}(p)$$

where $W = A(A^T C^{-1} A)^{-1} A^T C^{-1}$ and $A \equiv S^{1/2}\, U_s$

Foreground-cleaned estimates of 21-cm maps across frequencies!

GNILC reconstruction of 21-cm signal

Input 21-cm map | GNILC 21-cm map | Residuals

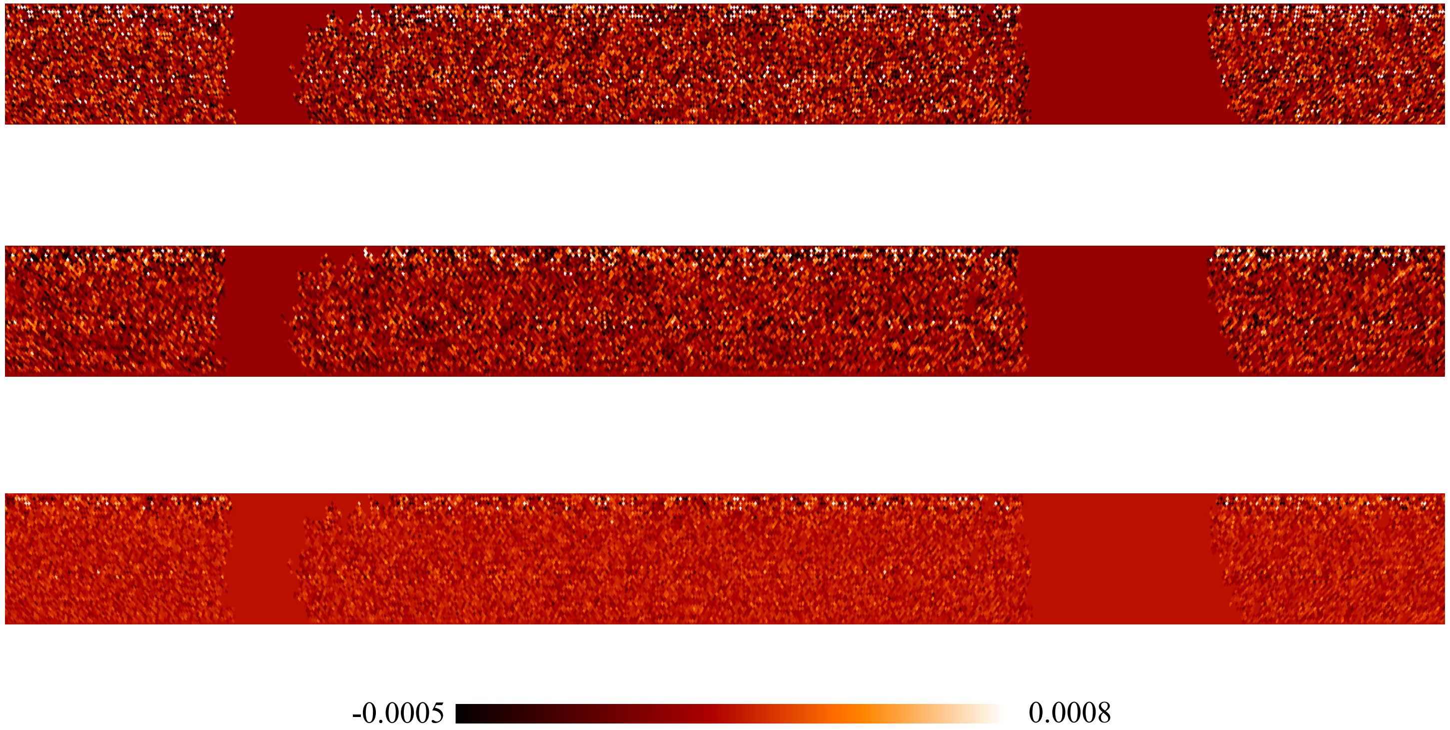


Olivari, Remazeilles, Dickinson, MNRAS 2016

GNILC reconstruction of 21-cm signal

BINGO simulations

BINGO stripe



Input 21-cm map

GNILC 21-cm map

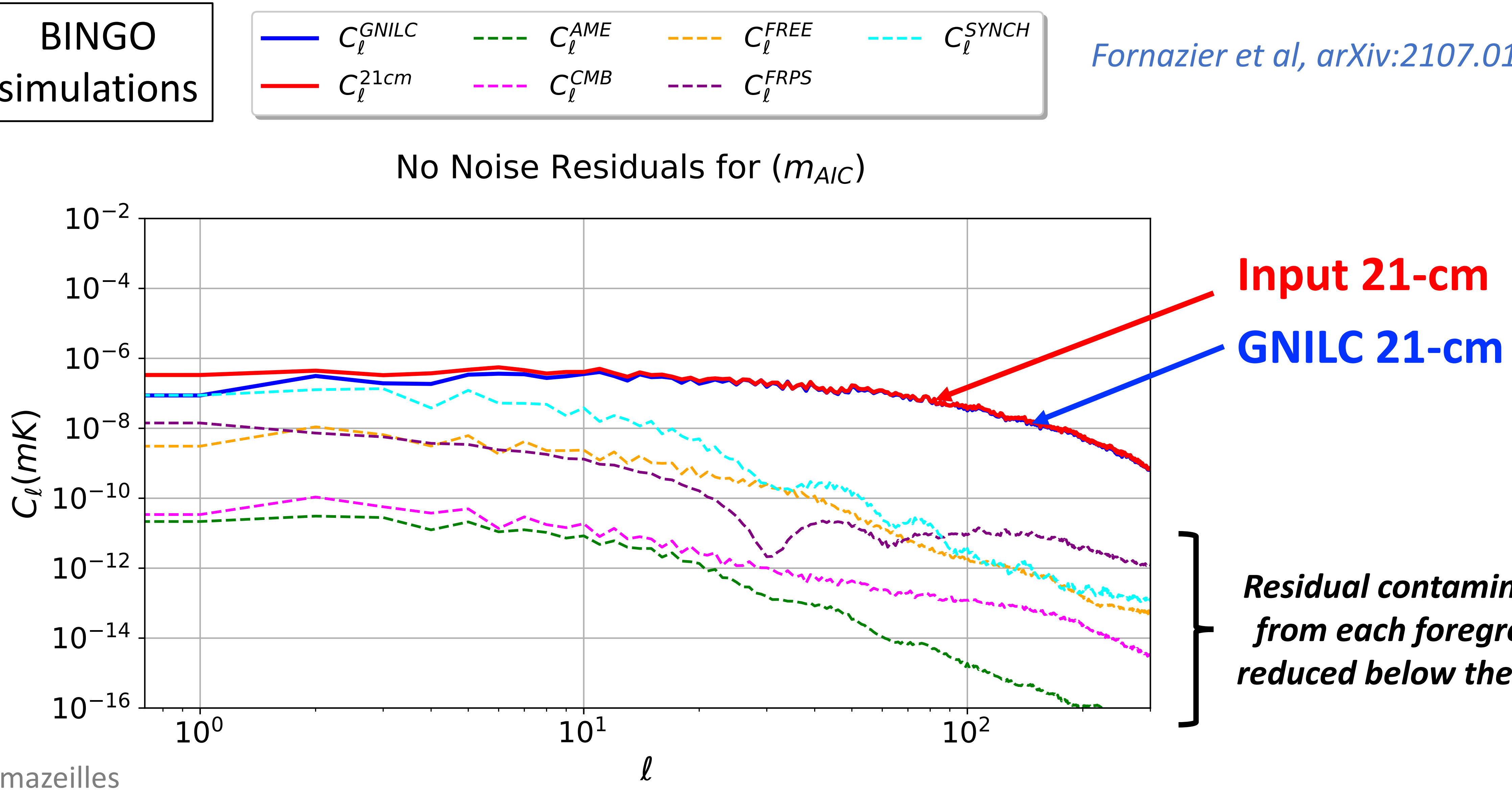
Residuals

Liccardo et al, arXiv:2107.01636

Power spectrum of GNILC 21-cm map

BINGO simulations

Fornazier et al, arXiv:2107.01637

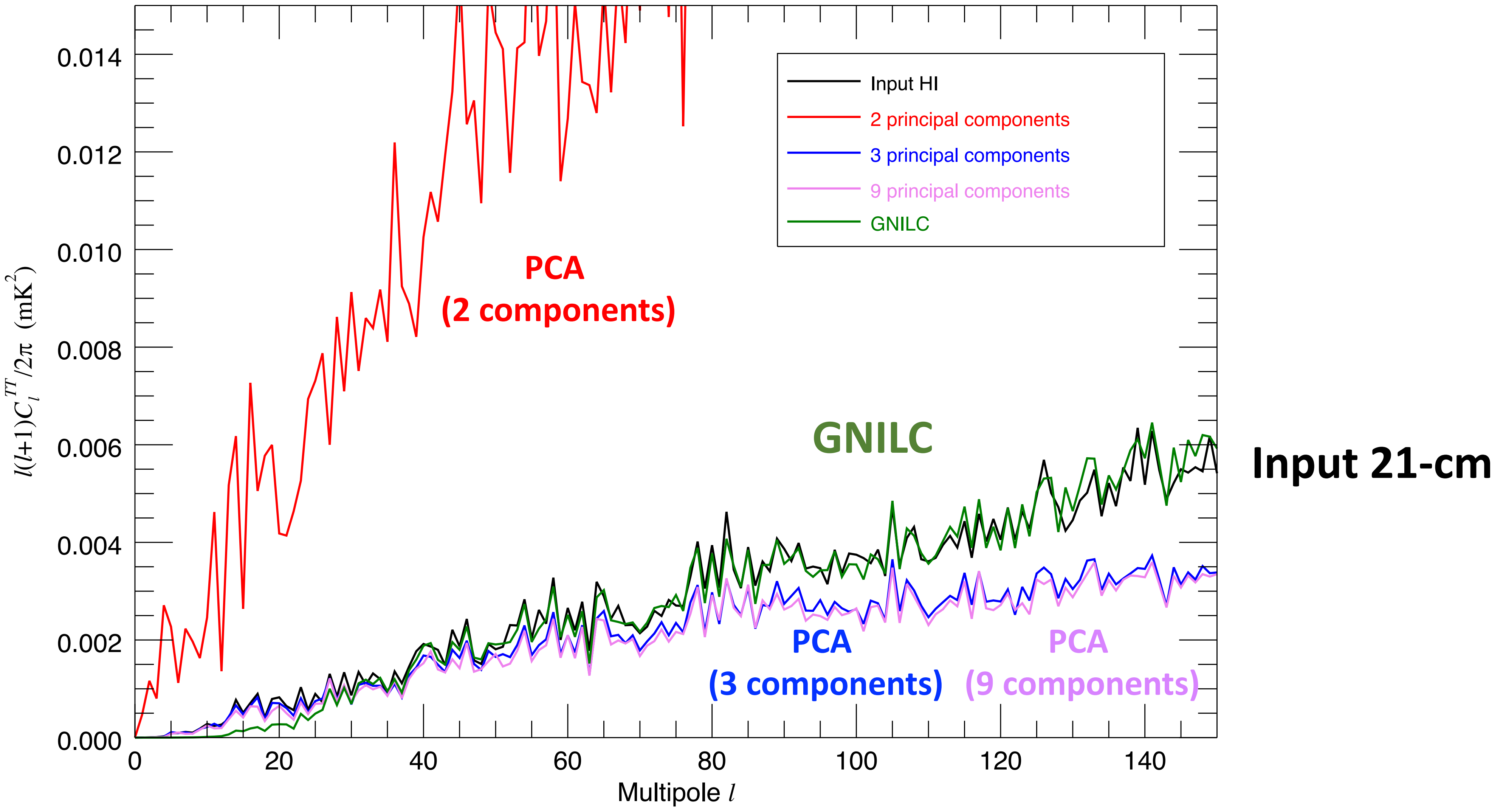


Residual contamination from each foreground reduced below the signal

M. Remazeilles

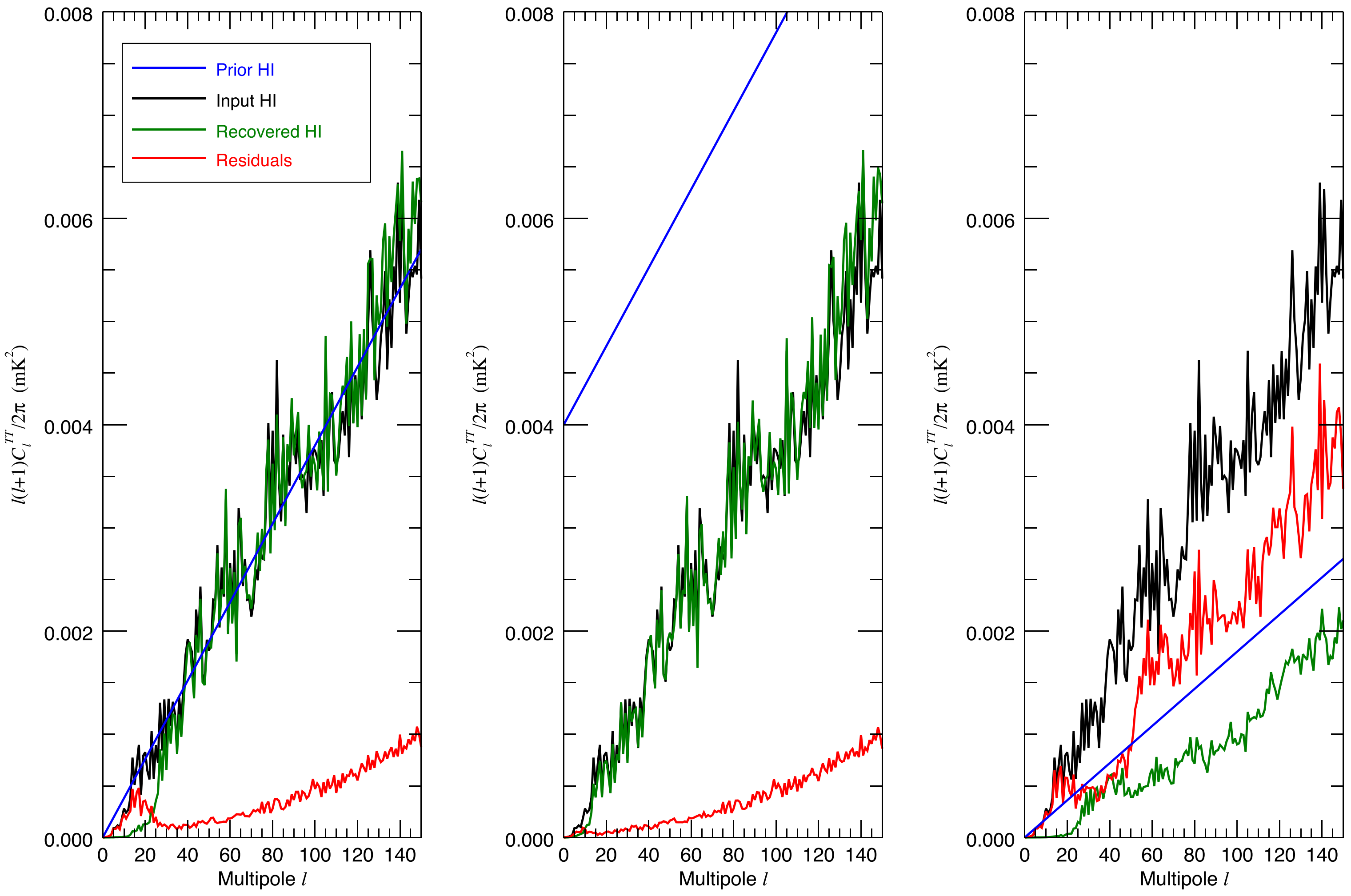
GNILC versus PCA

Olivari, Remazeilles, Dickinson, MNRAS 2016



M. Remazeilles

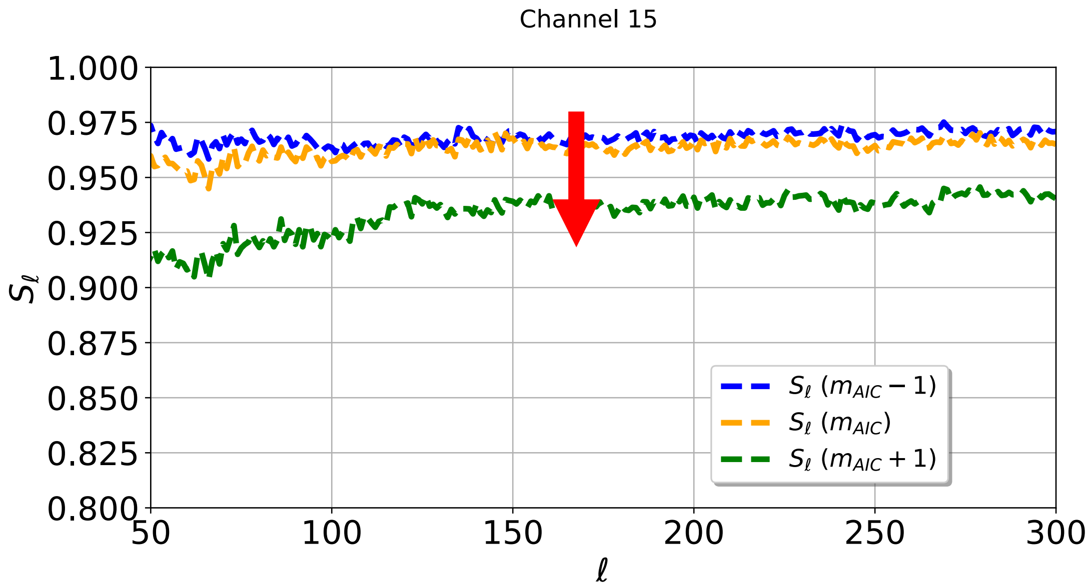
GNILC quite insensitive to 21-cm priors



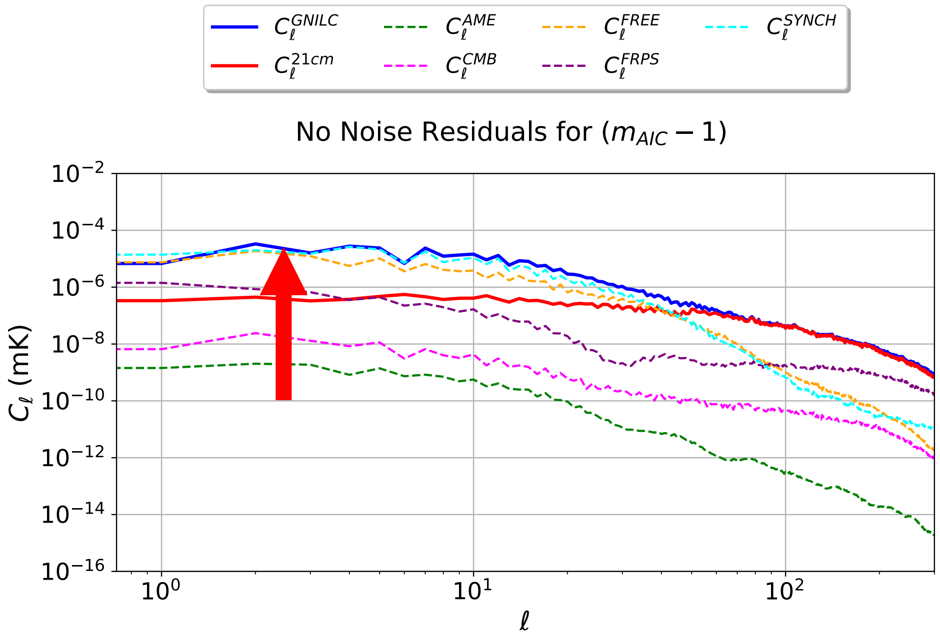
Olivari, Remazeilles, Dickinson, MNRAS 2016

Foreground subtraction vs 21cm signal loss

More aggressive foreground subtraction increases loss of 21-cm signal



Less aggressive foreground subtraction leaves residuals larger than the signal



GNILC with AIC value $m_{\rm AIC}$ finds the sweet spot!

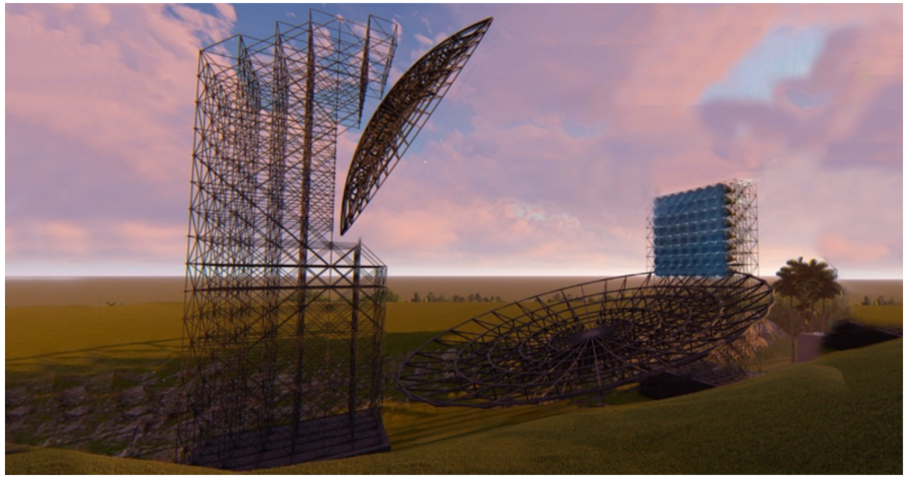
Fornazier et al arXiv:2107.01637

M. Remazeilles

Takeaway

- GNILC goes beyond simple spectral modelling for component separation and 21-cm intensity mapping
- GNILC shows successful 21-cm signal reconstruction on various sky simulations of the BINGO experiment
- GNILC has already been intensively used on real *Planck* data and is at the heart of several *Planck* papers

  Planck intermediate results XLVIII. Disentangling Galactic dust and cosmic infrared background anisotropies, A&A (2016)
  Planck 2018 results IV. Diffuse component separation, A&A (2020)
  Planck 2018 results XII. Galactic astrophysics using polarized dust emission, A&A (2020)



Let us ensure successful science return from BINGO with GNILC!