Radio foregrounds and component separation for 21 cm cosmology

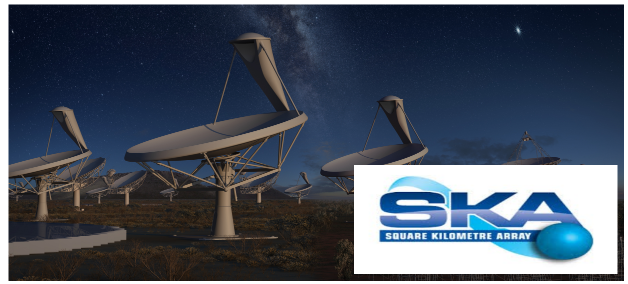
Mathieu Remazeilles

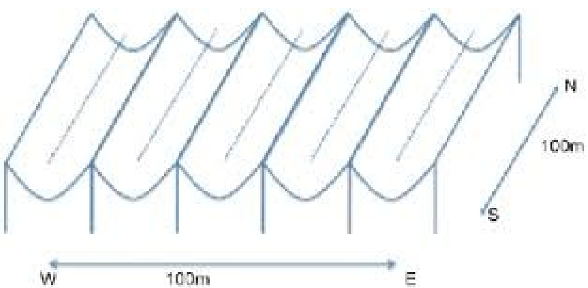
The University of Manchester

"Accurate Astrophysics. Correct Cosmology."
London, 13-16 July 2015

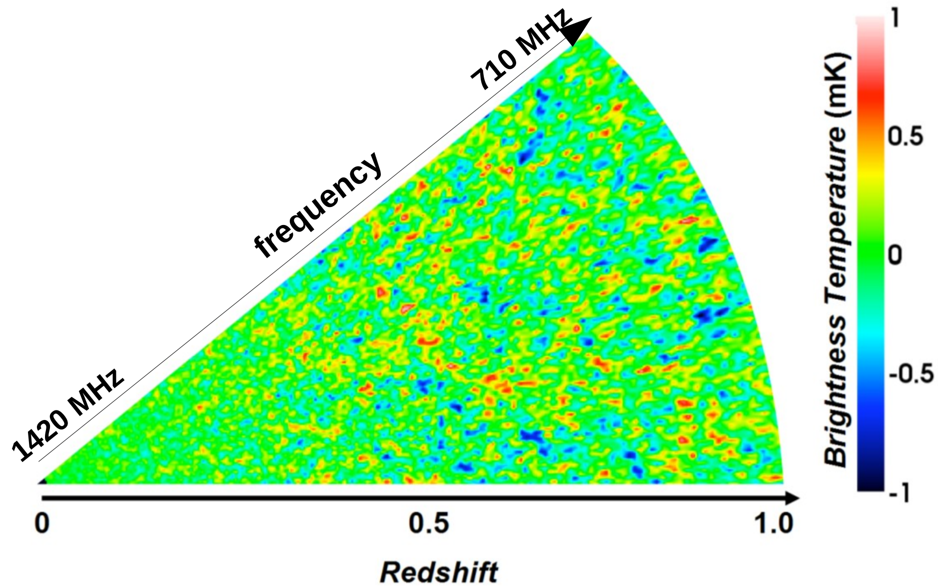
Radio Intensity Mapping

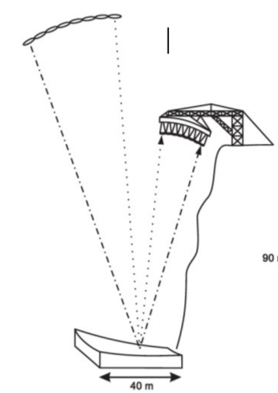
*A new window for **mapping, in the radio domain,***
*large-scale structures **(LSS)** and baryon acoustic oscillations **(BAO)***
*by using the **redshifted 21 cm line emission** from **HI gas***





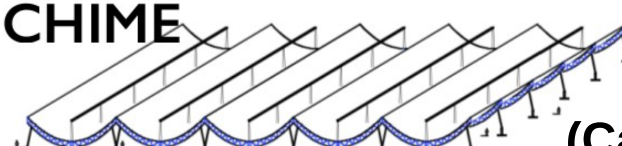
Tianlai (China)





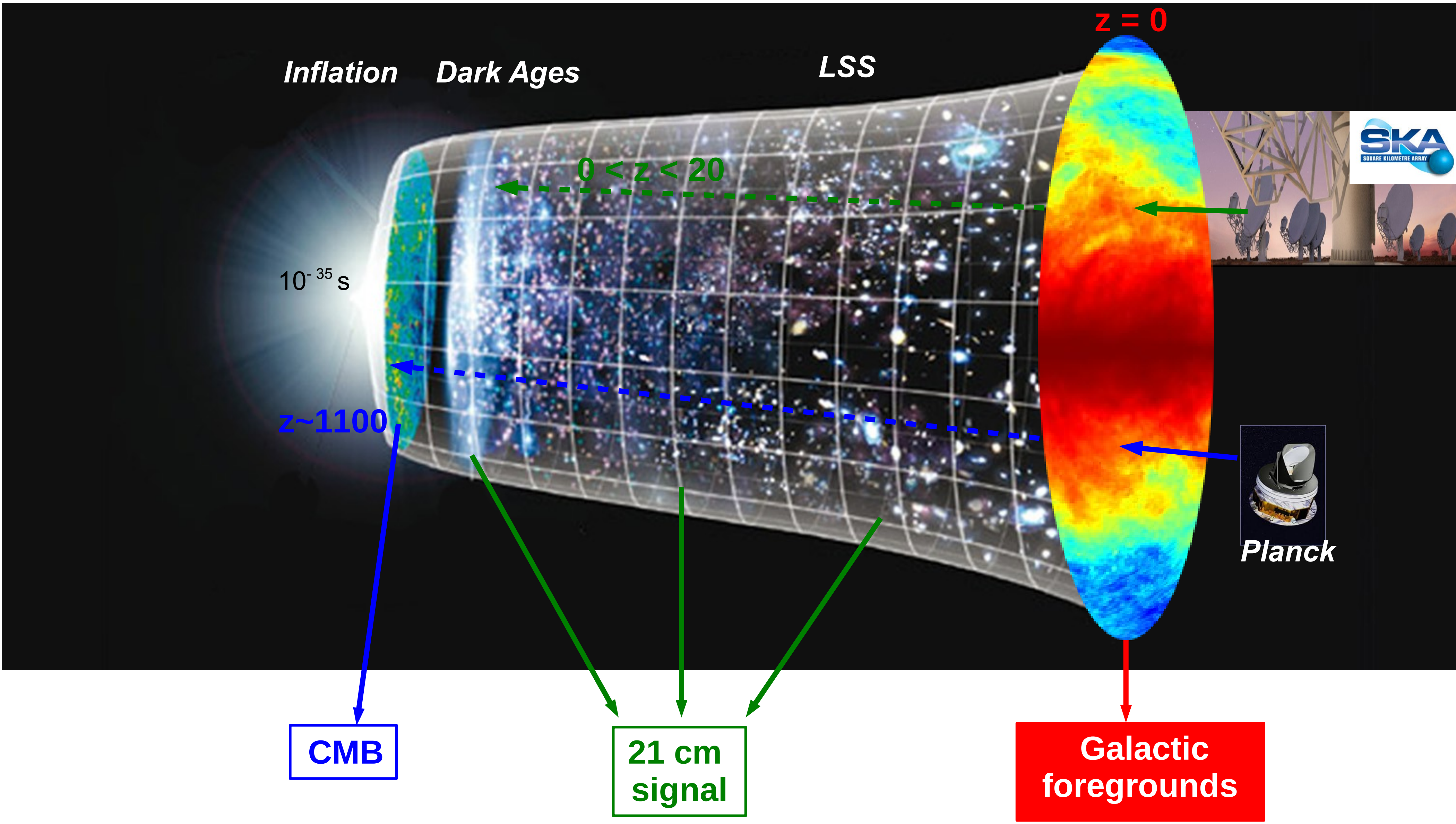
BINGO (UK)

CHIME

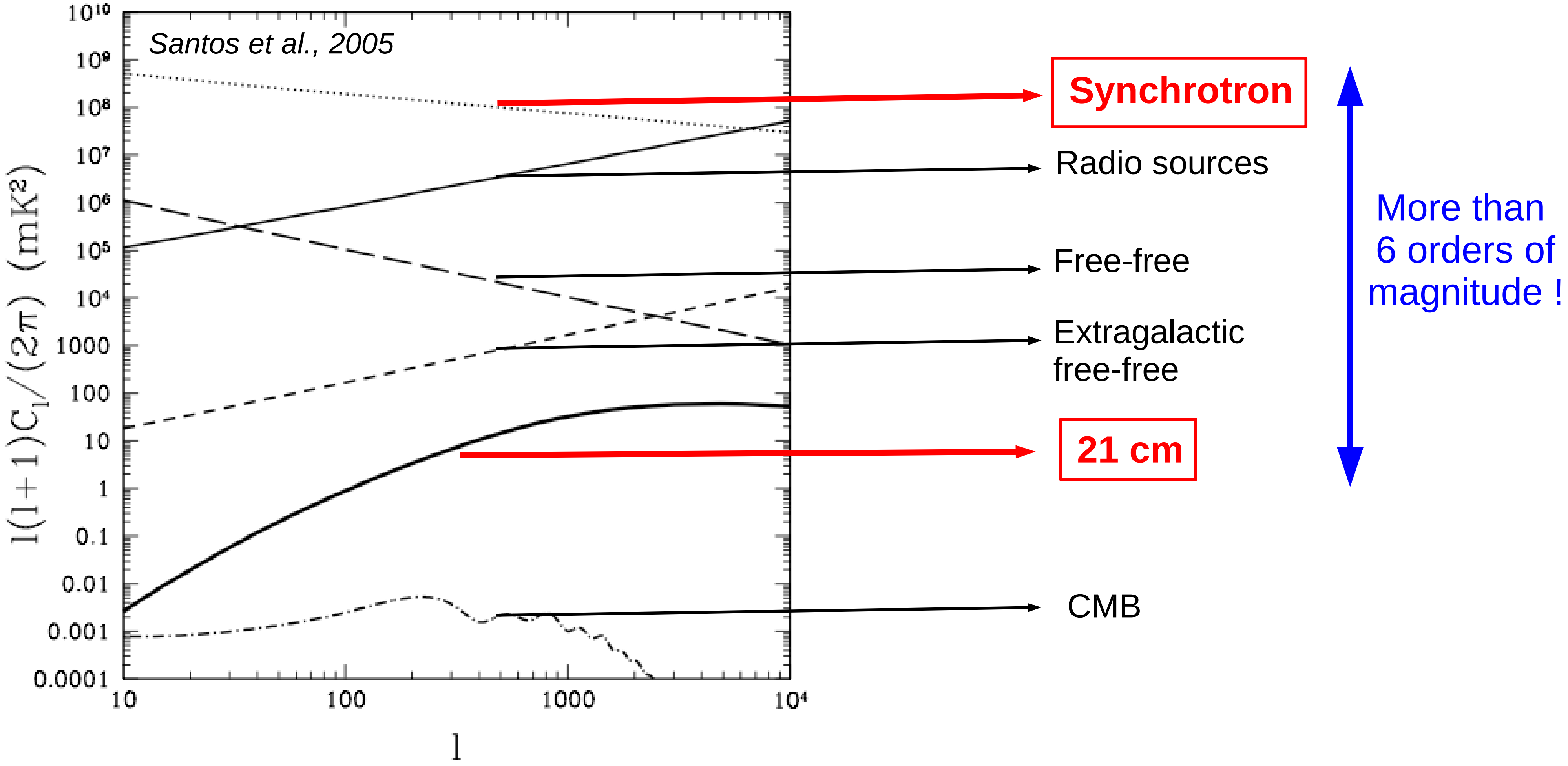


(Canada)

Galactic foregrounds hide cosmological signals



Galactic synchrotron : a major foreground



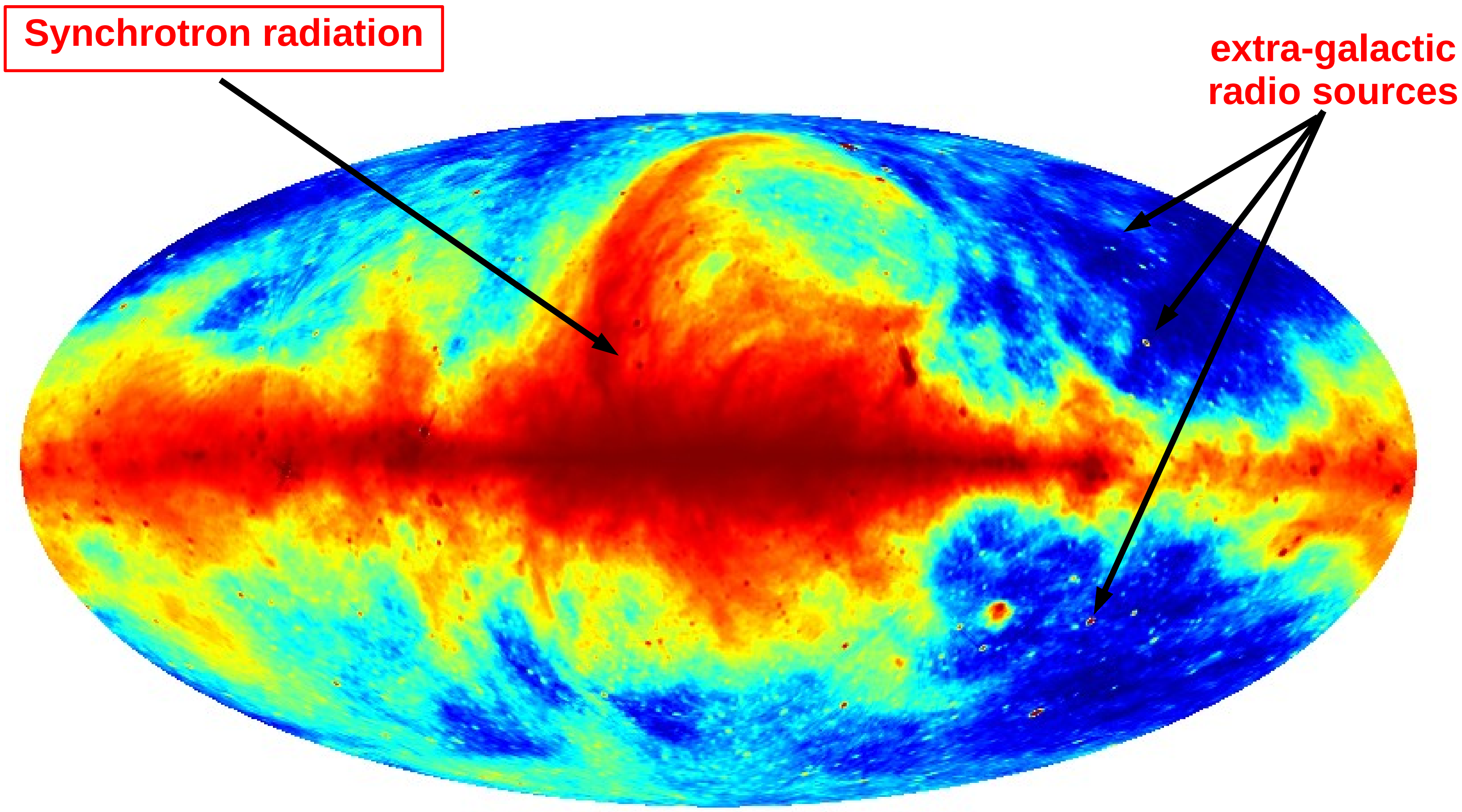
z=9.2 (f = 140 MHz)

Can we guarantee a robust subtraction of foregrounds in 21 cm observations ?

1. *Better characterize Galactic foregrounds at radio frequencies*

Remazeilles, Dickinson, Banday, Bigot-Sazy, Ghosh,
“An improved source-subtracted and destriped 408-MHz all-sky map”, MNRAS (2015)

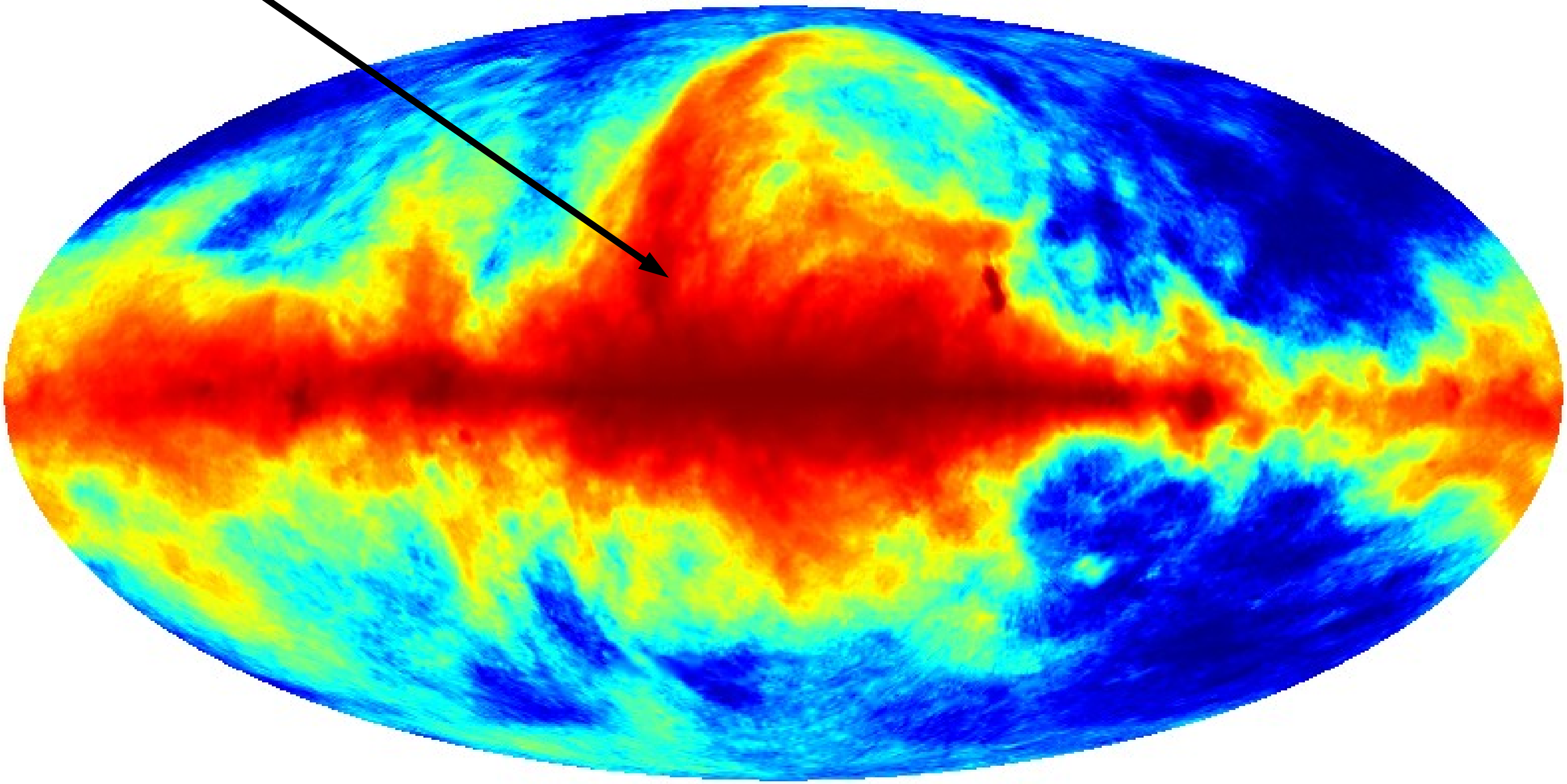
Galactic synchrotron all-sky map at 408 MHz



Haslam et al. (1982)

Galactic synchrotron all-sky map at 408 MHz

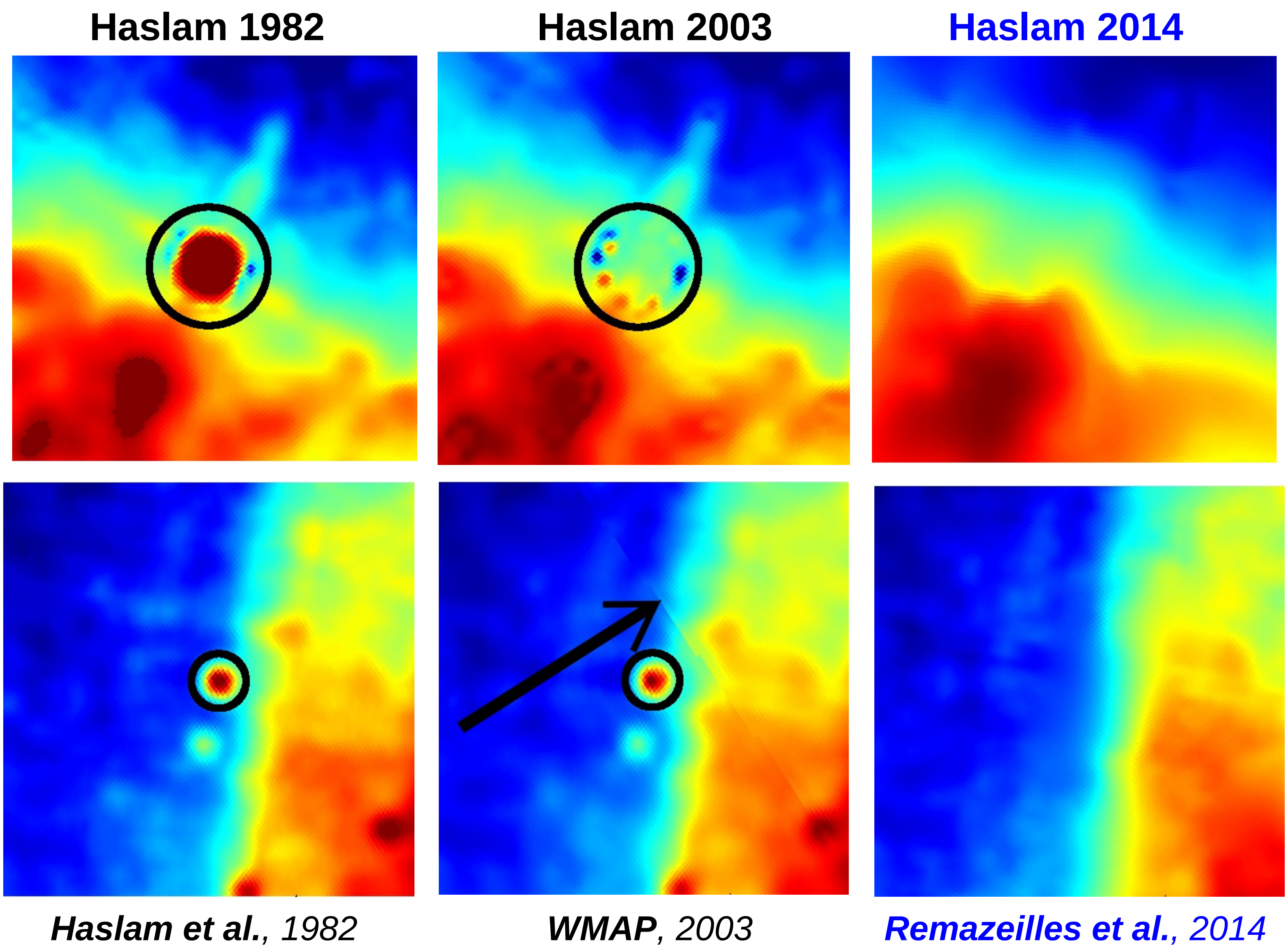
Synchrotron radiation



Remazeilles et al (2014)

Remazeilles, Dickinson, Banday, Bigot-Sazy, Ghosh, "An improved source-subtracted and destriped 408-MHz all-sky map", MNRAS (2015)

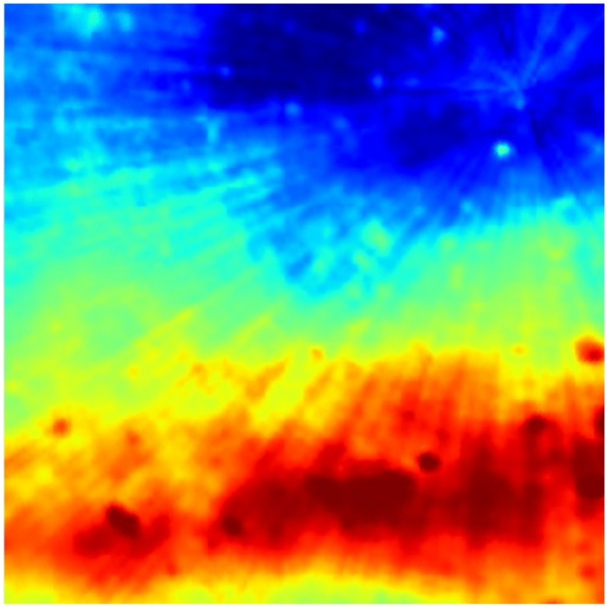
Improvement on desourcing



Remazeilles, Dickinson, Banday, Bigot-Sazy, Ghosh,
"An improved source-subtracted and destriped 408-MHz all-sky map",
MNRAS (2015)

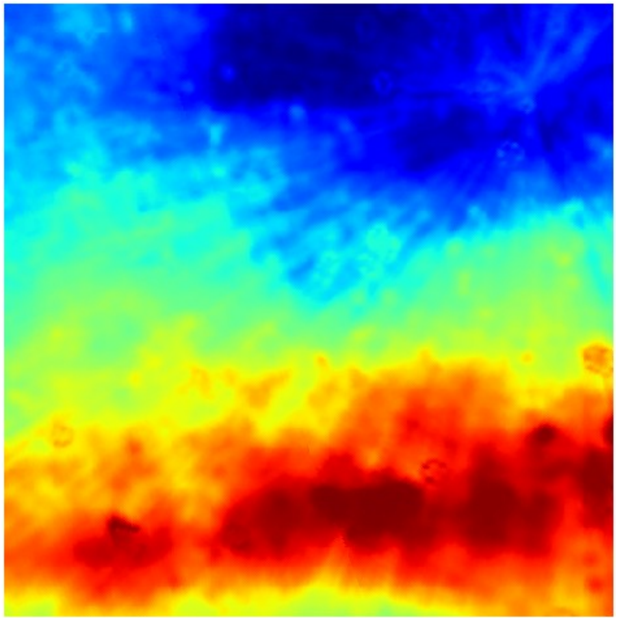
Improvement on destriping

Haslam 1982



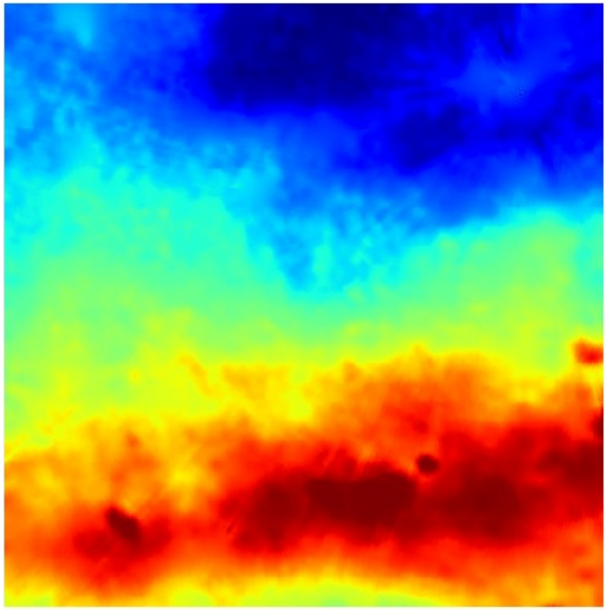
***Haslam et al.**, 1982*

Haslam 2003



***WMAP**, 2003*

Haslam 2014

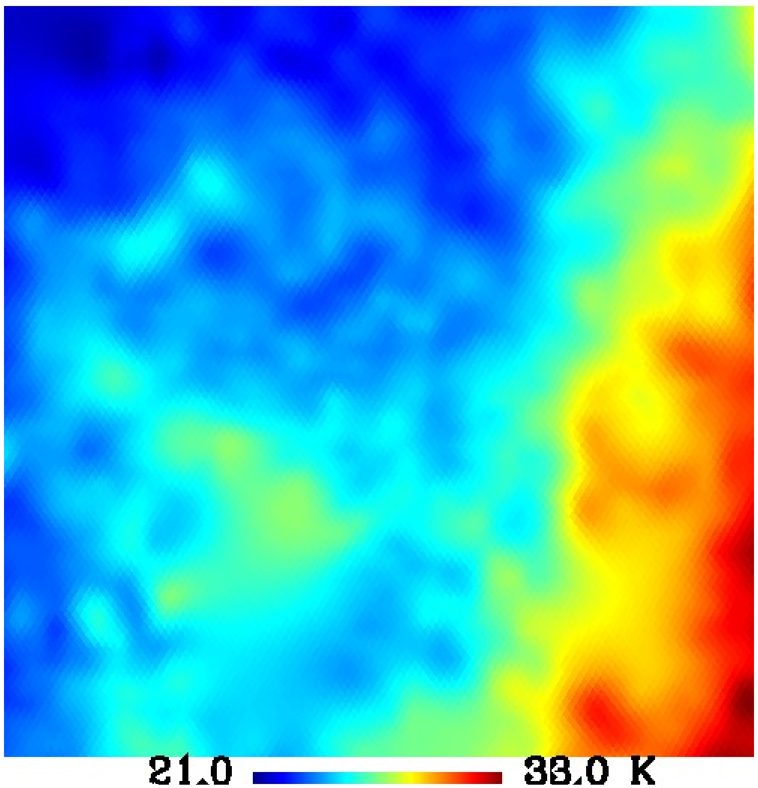


***Remazeilles et al.**, 2014*

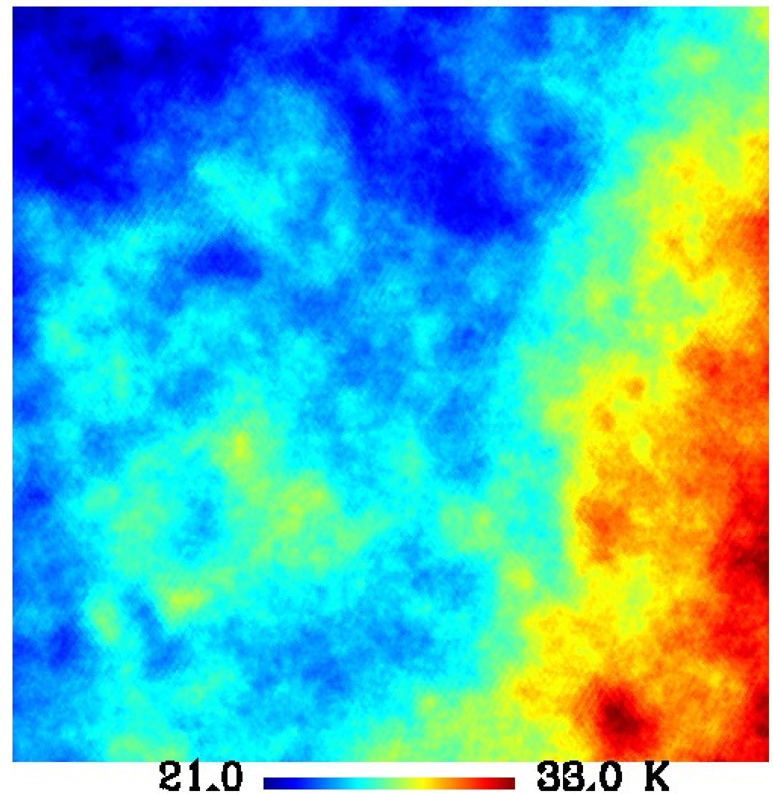
Remazeilles, Dickinson, Banday, Bigot-Sazy, Ghosh,
"An improved source-subtracted and destriped 408-MHz all-sky map",
MNRAS (2015)

Bonus : High-resolution synchrotron template for radio sky simulations

Haslam 2014



High-resolution Haslam 2014



Remazeilles, Dickinson, Banday, Bigot-Sazy, Ghosh,
“An improved source-subtracted and destriped 408-MHz all-sky map”,
MNRAS (2015)

Haslam 408 MHz Data Products

The first file listed below contains HEALPix format data stored in a binary table extension, and its FITS header is linked to in the File Info column. The HPX file contains the same HEALPix data stored as a 2D primary image array in HPX projection (Calabretta and Roukema 2007), and may be displayed with SAOImage DS9. For the convenience of users who do not require the capabilities that HEALPix offers, we also provide the maps in two projected image formats: Mollweide and Zenith Equal Area (ZEA). These files contain the string "_mollweide" and "_zea" respectively, and may be displayed with standard astronomical display software such as DS9 and SAOImage. Note that the Mollweide and ZEA image pixels typically over-sample the HEALPix pixels, thus some amount of data replication will be present in the image data. As a result, care must be taken when using the image data in a regime where the noise properties are important since the statistical weight per pixel will also be replicated.

The HPX files are celestial maps with a FITS-standard HPX coordinate system. The process is described in "Mapping on the HEALPix Grid" by Mark Calabretta and Boud Roukema; see http://www.atnf.csiro.au/people/Mark.Calabretta

Recommended download procedure for '.fits' or '.tar.gz' files:
right click on the file name and save. For more information on the use of Dataset IDs,
see the LAMBDA Products Overview.

Show All Hide All

2014 Reprocessed Haslam Maps

Description	File Info	File Name
Desourced destriped Haslam 408 MHz	12.6 MB FITS Hdr	haslam408_dsds_Remazeilles2014.fits haslam408_HPX_dsds_Remazeilles2014.fits haslam408_mollweide_dsds_Remazeilles2014.fits haslam408_zea_dsds_Remazeilles2014.fits
Desourced destriped Haslam 408 MHz with small scales added	402.7 MB FITS Hdr	haslam408_dsds_Remazeilles2014_ns2048.fits haslam408_HPX_dsds_Remazeilles2014_ns2048.fits
Destriped Haslam 408 MHz	12.6 MB FITS Hdr	haslam408_ds_Remazeilles2014.fits haslam408_HPX_ds_Remazeilles2014.fits haslam408_mollweide_ds_Remazeilles2014.fits haslam408_zea_ds_Remazeilles2014.fits
Bitmask Haslam 408 MHz	6.3 MB FITS Hdr	haslam408_bitmask_Remazeilles2014.fits haslam408_HPX_bitmask_Remazeilles2014.fits haslam408_mollweide_bitmask_Remazeilles2014.fits haslam408_zea_bitmask_Remazeilles2014.fits

Detailed Product Description

Previous Haslam 408 MHz Data Products

Can we guarantee a robust subtraction of foregrounds in 21 cm observations ?

2. *Knowledge transfer from CMB component separation*

GNILC method **("Generalized Needlet Internal Linear Combination")**

Remazeilles, Delabrouille, Cardoso,
"Foreground component separation with Generalized Needlet ILC", MNRAS (2011)

1) CIBA - Galactic dust separation in Planck data

Planck Collaboration et al.,
"Disentangling Galactic dust and CIB anisotropies in Planck observations", in prep. (2015)

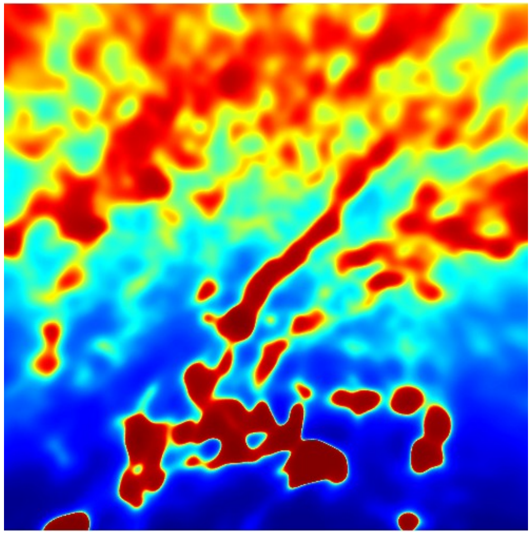
2) 21 cm - foreground separation in radio intensity mapping simulations

Olivari, Remazeilles, Dickinson,
"Extracting 21 cm cosmological signal with Generalized Needlet ILC", in prep. (2015)

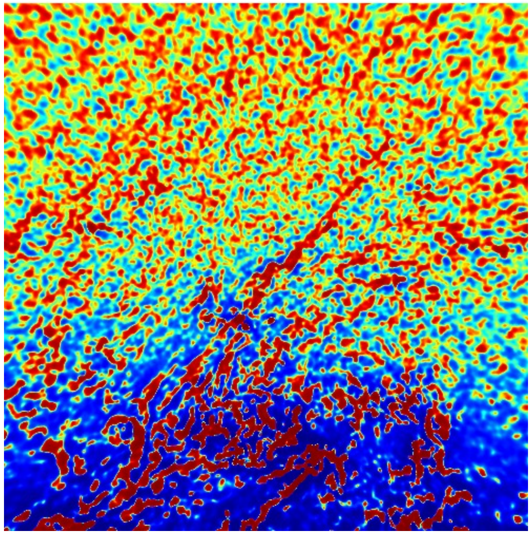
Generalized Needlet ILC (GNILC)

Remazeilles, Delabrouille, and Cardoso, MNRAS (2011)

- **Wavelet decomposition** of the data to deal with **local conditions of contamination**



wavelet scale 3 degrees



wavelet scale 30 arcminutes

- **Data covariance matrix** (n x n frequencies) at each wavelet scale

$$\mathbf{R} = \mathbf{R}_{\text{CIB}} + \mathbf{R}_{\textit{foregrounds}} + \mathbf{R}_{\textit{noise}}$$

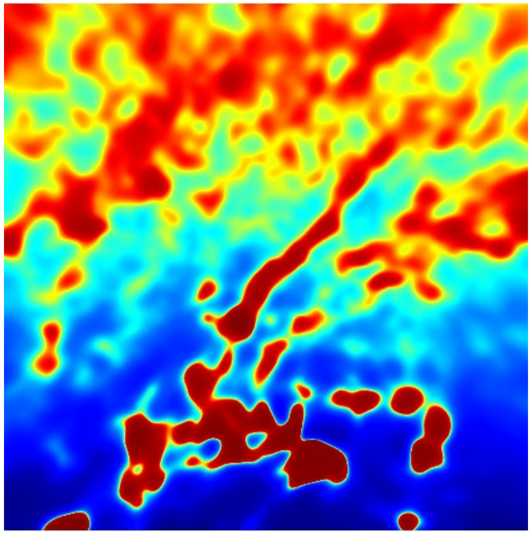
- **Prior on CIB power spectra.** Compute :

$$\left(\mathbf{R}_{\text{CIB}}^{\textit{prior}}\right)^{-\frac{1}{2}} \mathbf{R} \left(\mathbf{R}_{\text{CIB}}^{\textit{prior}}\right)^{-\frac{1}{2}} \approx \mathbf{I} + \mathbf{R}_N$$

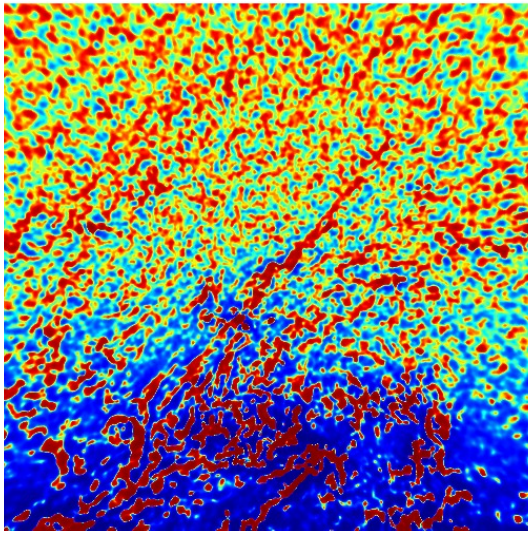
Generalized Needlet ILC (GNILC)

Remazeilles, Delabrouille, and Cardoso, MNRAS (2011)

- **Wavelet decomposition** of the data to deal with **local conditions of contamination**



wavelet scale 3 degrees



wavelet scale 30 arcminutes

- **Data covariance matrix** (n x n frequencies) at each wavelet scale

$$\mathbf{R} = \mathbf{R}_{21cm} + \mathbf{R}_{foregrounds} + \mathbf{R}_{noise}$$

- **Prior on 21 cm power spectra.** Compute :

$$\left(\mathbf{R}_{21cm}^{prior}\right)^{-\frac{1}{2}} \mathbf{R} \left(\mathbf{R}_{21cm}^{prior}\right)^{-\frac{1}{2}} \approx \mathbf{I} + \mathbf{R}_{N}$$

Generalized Needlet ILC (GNILC)

Remazeilles, Delabrouille, and Cardoso, MNRAS (2011)

$$\left(\mathbf{R}_{21cm}^{prior}\right)^{-\frac{1}{2}} \mathbf{R} \left(\mathbf{R}_{21cm}^{prior}\right)^{-\frac{1}{2}} \approx \mathbf{I} + \mathbf{R}_N$$

- **Eigen-decomposition**

$$\left[\begin{array}{c|c} \cdots & \mathbf{U_S} \end{array}\right] \cdot \left[\begin{array}{ccc|c} 1+\lambda_1 & & & \\ & \cdots & & \\ & & 1+\lambda_m & \\ \hline & & & \mathbf{I} \end{array}\right] \cdot \left[\begin{array}{c} \cdots \\ \hline \mathbf{U_S}^T \end{array}\right]$$

m eigenvalues >> 1
→ foreground+noise power

(n-m) eigenvalues ≈ 1
→ 21 cm power

$\mathbf{U_S}$ → eigenmodes of the 21 cm subspace

- **m** estimated by minimizing Akaike Information Criterion : $AIC(m) = 2m - 2\log\mathcal{L}_{max}(m)$

Generalized Needlet ILC (GNILC)

Remazeilles, Delabrouille, and Cardoso, MNRAS (2011)

$$\left(\mathbf{R}^{prior}_{21cm}\right)^{-\frac{1}{2}}\mathbf{R}\left(\mathbf{R}^{prior}_{21cm}\right)^{-\frac{1}{2}} \approx \mathbf{I} + \mathbf{R}_N$$

- **Eigen-decomposition**

$$\left[\begin{array}{c|c} \cdots & \mathbf{U_S} \end{array}\right] \cdot \left[\begin{array}{ccc|c} 1+\lambda_1 & & & \\ & \cdots & & \\ & & 1+\lambda_m & \\ \hline & & & \mathbf{I} \end{array}\right] \cdot \left[\begin{array}{c} \cdots \\ \hline \mathbf{U_S}^T \end{array}\right]$$

m eigenvalues >> 1
→ foreground+noise power

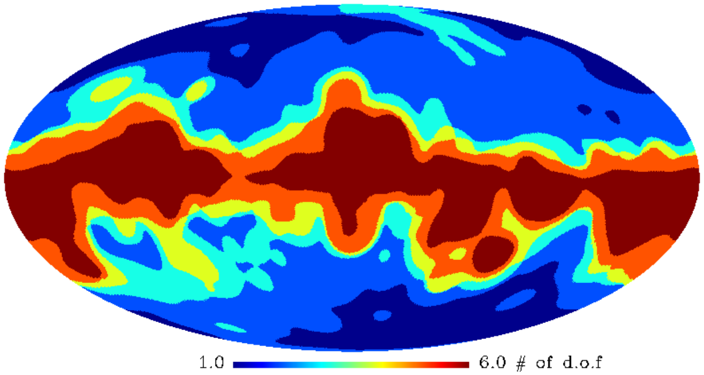
(n-m) eigenvalues ≈ 1
→ 21 cm power

$\mathbf{U_S}$ → eigenmodes of the 21 cm subspace

- **m** estimated by minimizing Akaike Information Criterion : $AIC(m) = 2m - 2\log\mathcal{L}_{max}(m)$

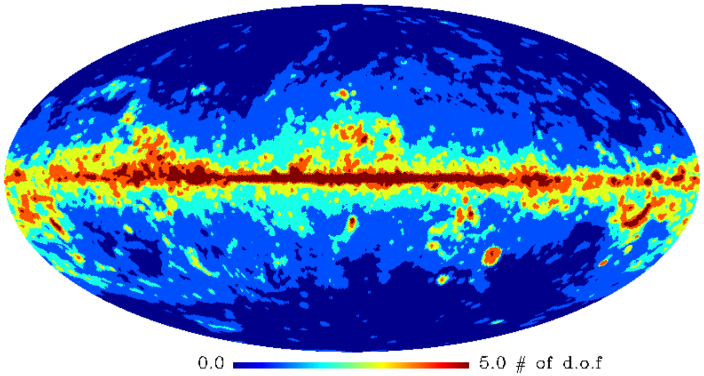
"Sort of PCA driven by the local signal-to-noise ratio"

wavelet scale 1 degree



Number of foreground modes

wavelet scale 5 arcminutes



Number of foreground modes

Generalized Needlet ILC (GNILC)

Remazeilles, Delabrouille, and Cardoso, MNRAS (2011)

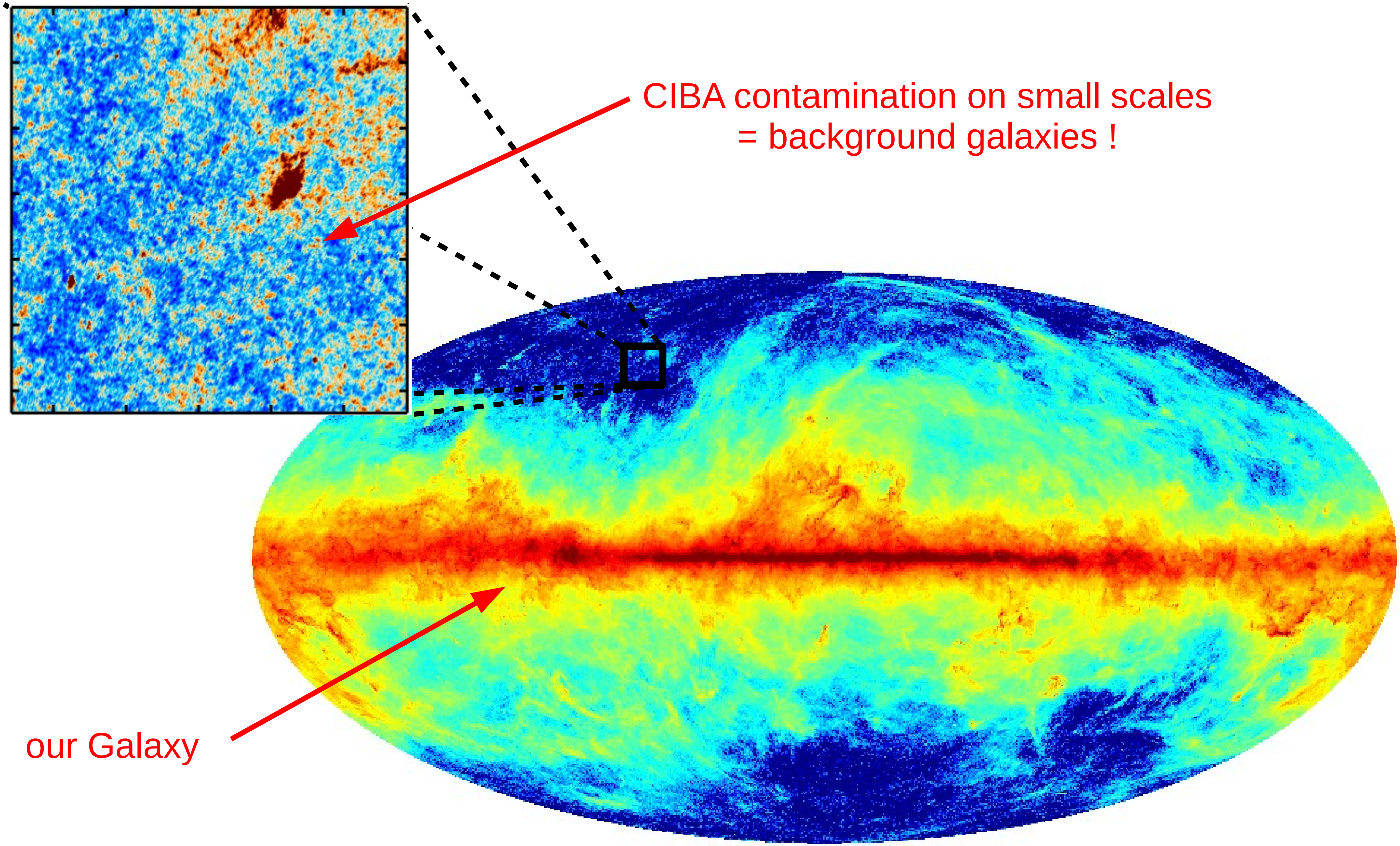
- **Wavelet decomposition** of the data to deal with **local conditions of contamination**
- **Data covariance matrix** (n x n frequencies) $\mathbf{R} = \mathbf{R}_{21cm} + \mathbf{R}_{foregrounds} + \mathbf{R}_{noise}$
- **Prior on 21 cm power spectra** $\left(\mathbf{R}_{21cm}^{prior}\right)^{-\frac{1}{2}} \mathbf{R} \left(\mathbf{R}_{21cm}^{prior}\right)^{-\frac{1}{2}} \approx \mathbf{I} + \mathbf{R}_N$
- **Eigen-decomposition of** $\left(\mathbf{R}_{21cm}^{prior}\right)^{-\frac{1}{2}} \mathbf{R} \left(\mathbf{R}_{21cm}^{prior}\right)^{-\frac{1}{2}}$

$$\left[\begin{array}{c|c} \cdots & \mathbf{U_S} \end{array}\right] \cdot \left[\begin{array}{ccc|c} 1+\lambda_1 & & & \\ & \cdots & & \\ & & 1+\lambda_m & \\ \hline & & & \mathbf{I} \end{array}\right] \cdot \left[\begin{array}{c} \cdots \\ \hline \mathbf{U_S}^T \end{array}\right]$$

- **Multidimensional ILC**

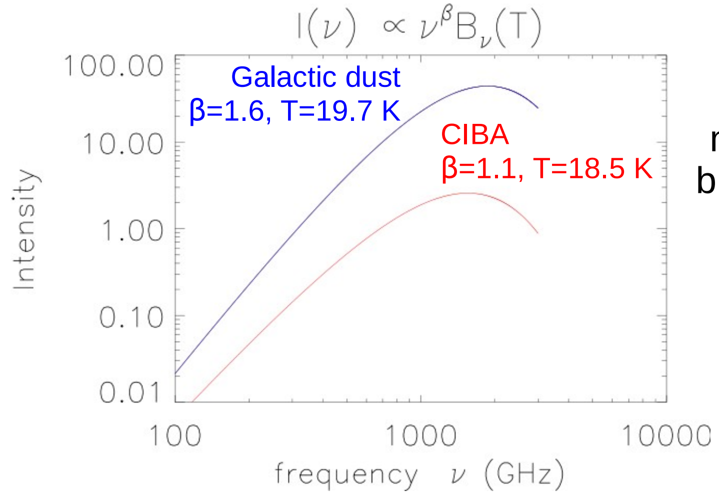
$$\boxed{\mathbf{A}\left(\mathbf{A}^T\mathbf{R}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{R}^{-1}}$$ *where* $\mathbf{A} = \left(\mathbf{R}_{21cm}^{prior}\right)^{1/2} \mathbf{U_S}$

Planck Galactic thermal dust all-sky map



Planck 2013 results. XI, "All-sky model of thermal dust emission", A&A (2014)

Spectral information



modified blackbody

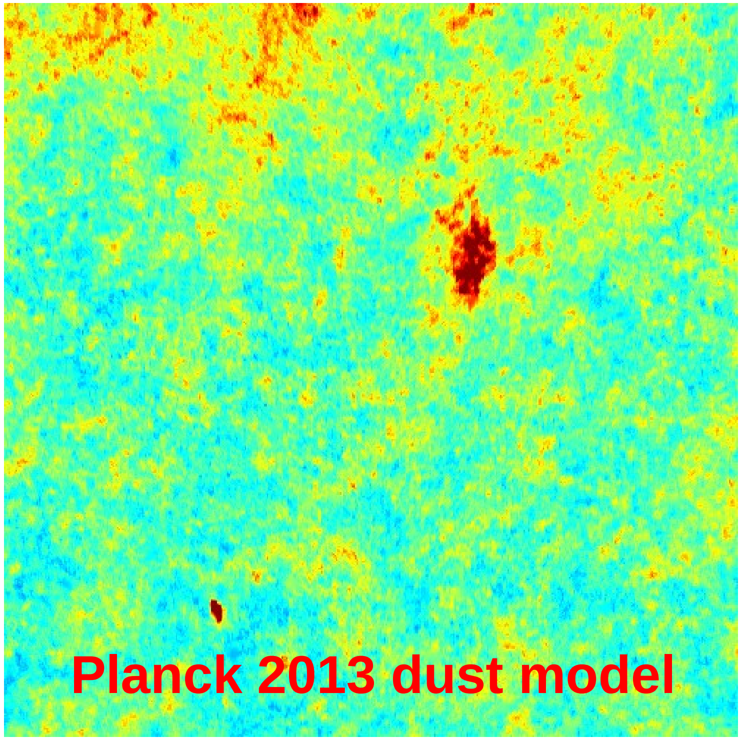
Spectral fitting:

- *Planck 2013 Results. XI. A&A (2014)*
- *Planck 2015 Results. X. arXiv:1502.01588*

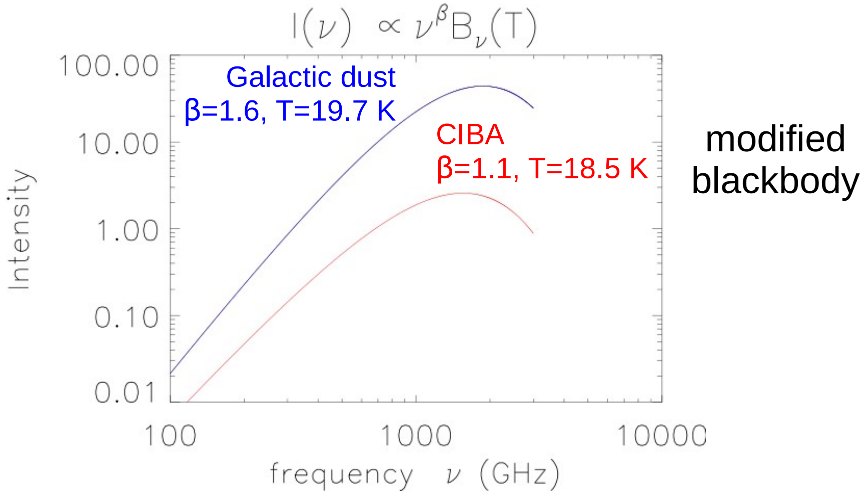
Dust and CIBA share similar spectral distributions

"spectral fitting"

CIBA leak into the thermal dust model !



Spectral information



Spectral fitting:

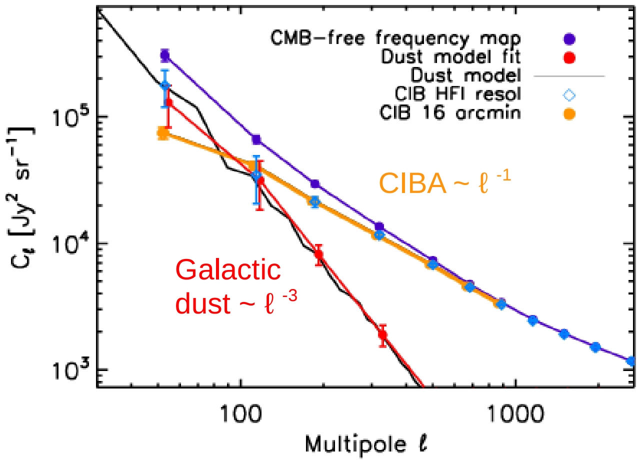
- *Planck 2013 Results. XI, A&A (2014)*
- *Planck 2015 Results. X, arXiv:1502.01588*

Dust and CIBA share similar spectral distributions

↓ "spectral fitting"

CIBA leak into the thermal dust model !

Statistical information



GNILC method:

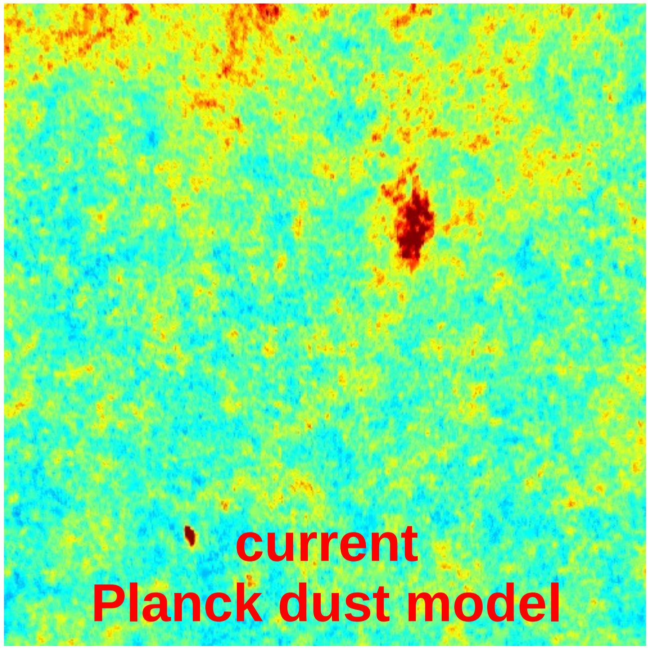
- *Planck 2015 Results. XII, in prep. (2015)*
- *Planck Intermediate Results. 120, in prep. (2015)*

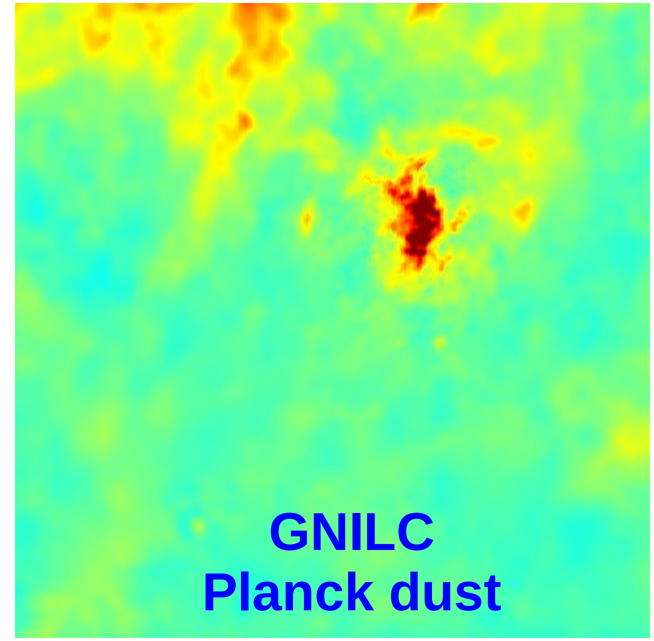
Dust and CIBA have distinct angular power spectra

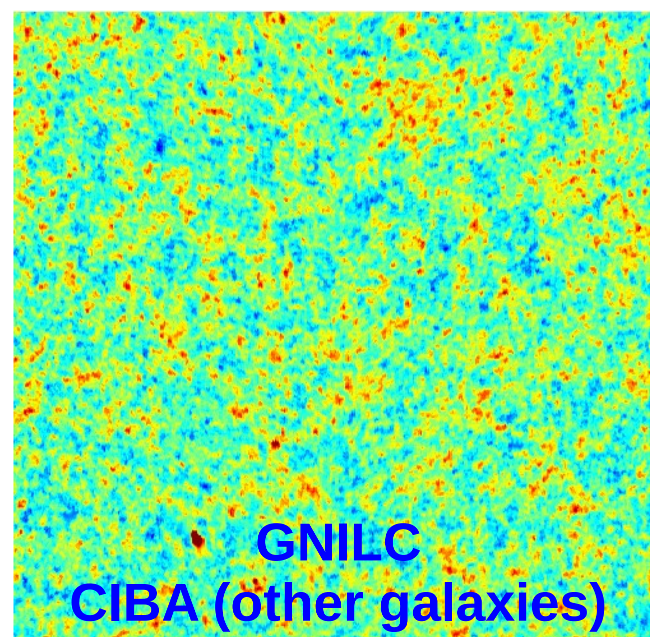
↓ "GNILC"

disentangle thermal dust and CIBA !

Separating dust & CIBA in Planck data

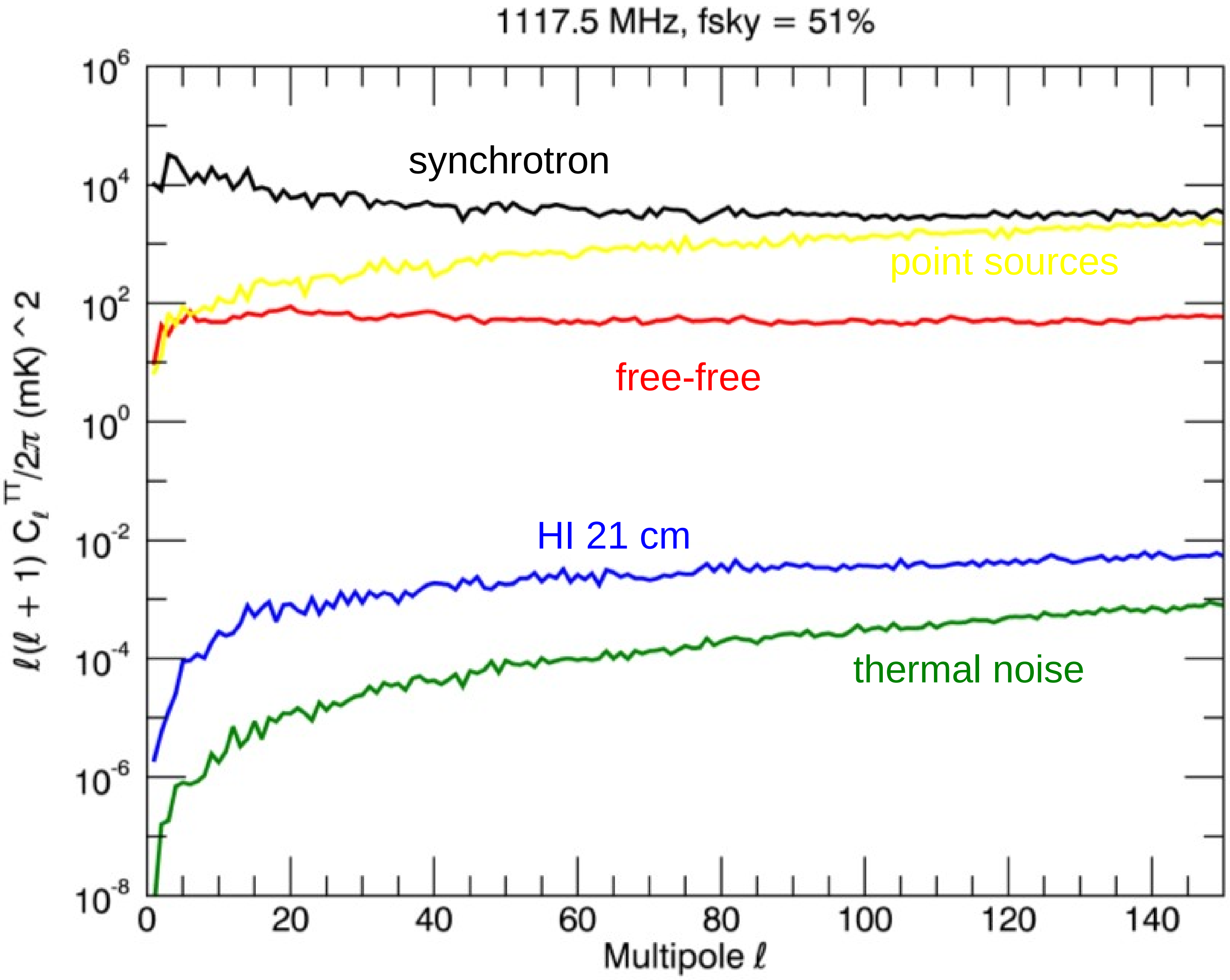






- ***Remazeilles et al.**, "Foreground component separation with Generalised ILC," MNRAS (2011)*
- ***Planck Collaboration**, "Planck 2015 Results. Simulations," in prep. (2015)*
- ***Planck Collaboration**, "Disentangling Dust and CIBA in Planck Observations," in prep. (2015) **[Corresp. Author]***

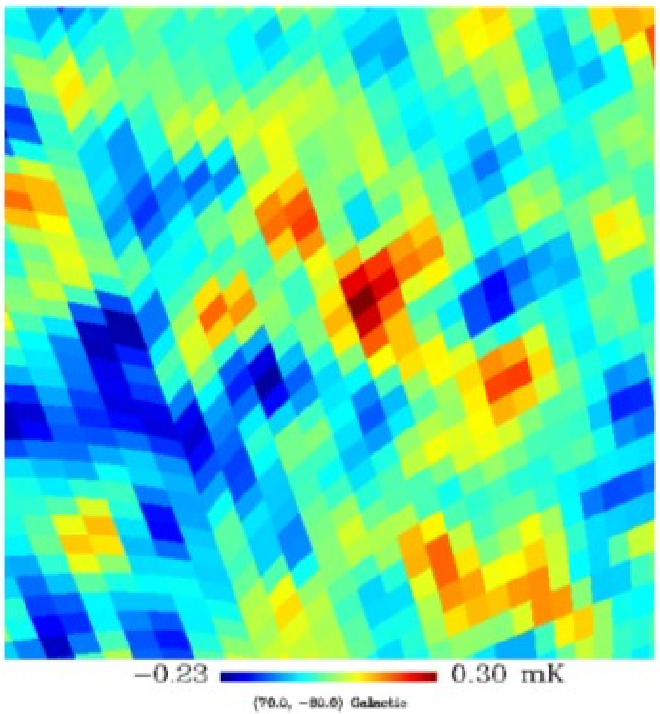
Simulation : radio intensity mapping



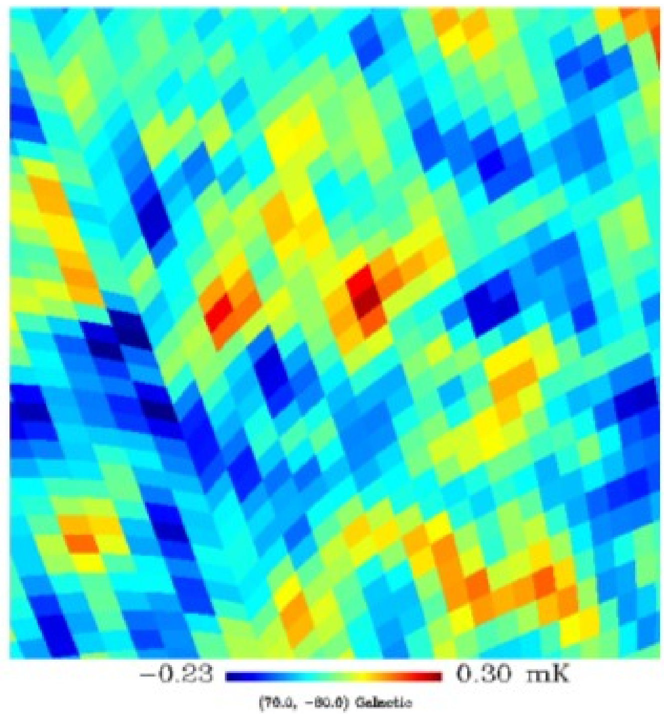
L. Olivari, M. Remazeilles, C. Dickinson,
"Extracting the 21 cm cosmological signal with GNILC", in prep. (2015)

GNILC reconstruction at 1117 MHz

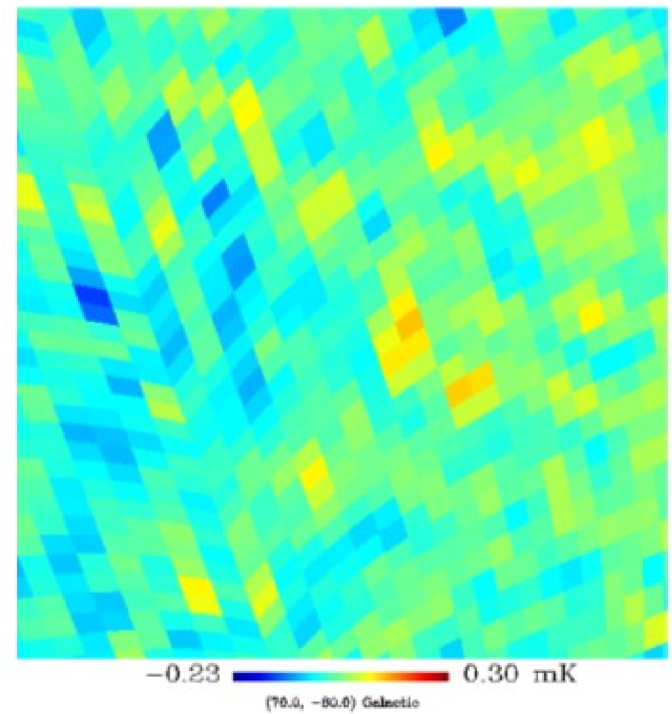
Input 21 cm



Recovered 21 cm



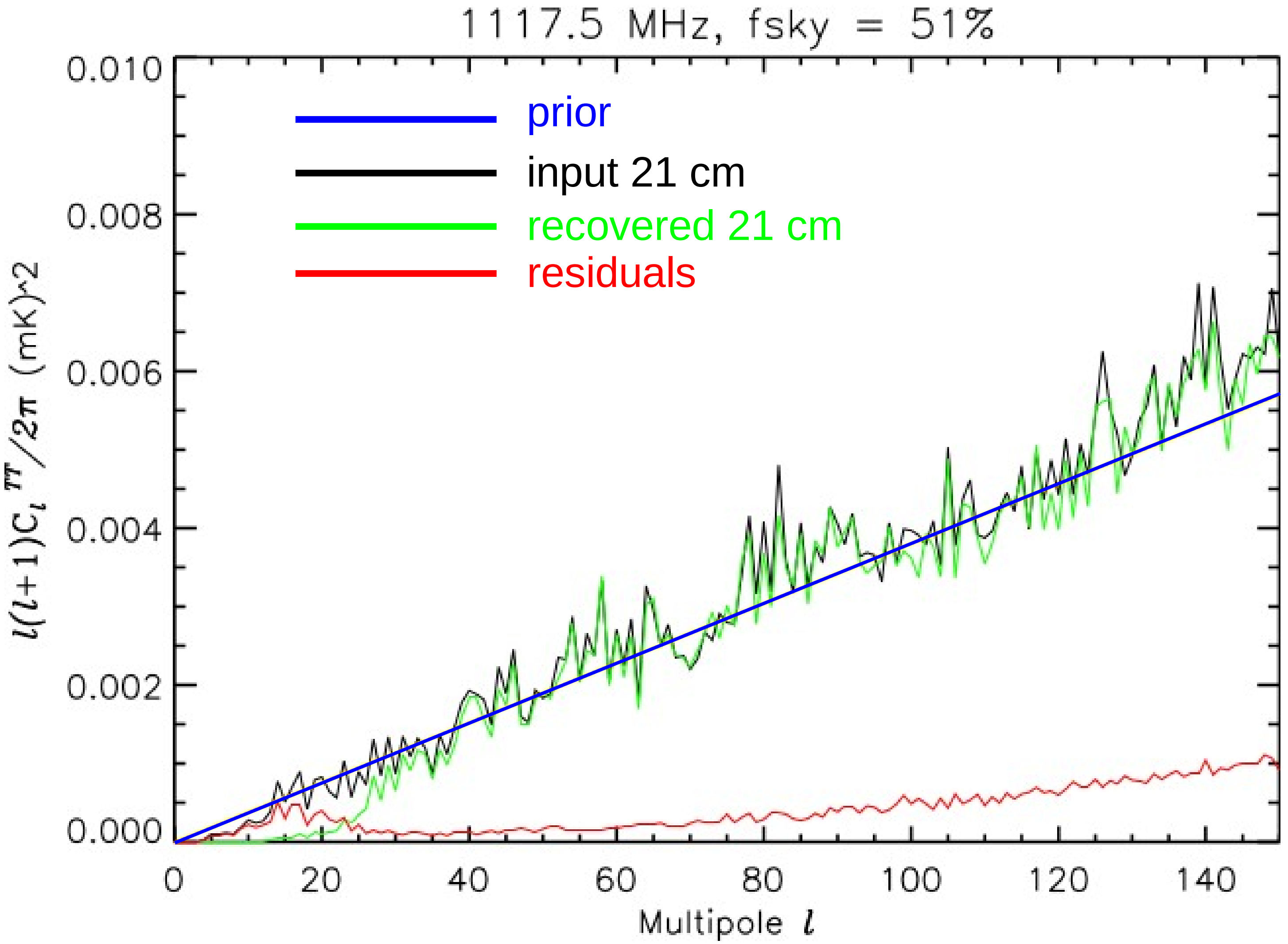
Residuals



12° x 12 °
40 ' resolution

L. Olivari, M. Remazeilles, C. Dickinson,
“Extracting the 21 cm cosmological signal with GNILC”, in prep. (2015)

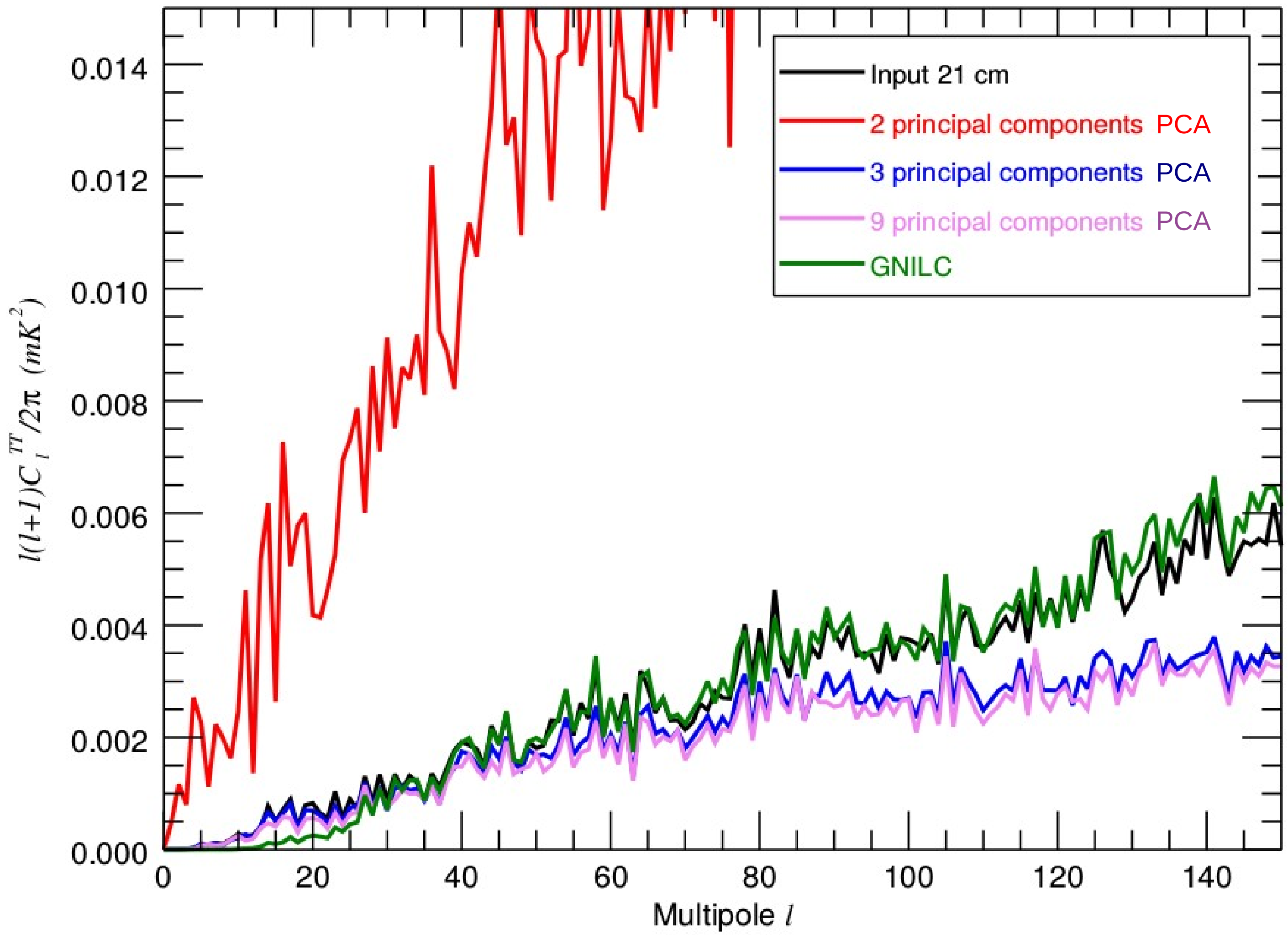
GNILC reconstruction at 1117 MHz



Despite smooth prior, GNILC is able to recover individual features of the 21 cm power spectrum

L. Olivari, M. Remazeilles, C. Dickinson,
"Extracting the 21 cm cosmological signal with GNILC", in prep. (2015)

GNILC versus standard PCA

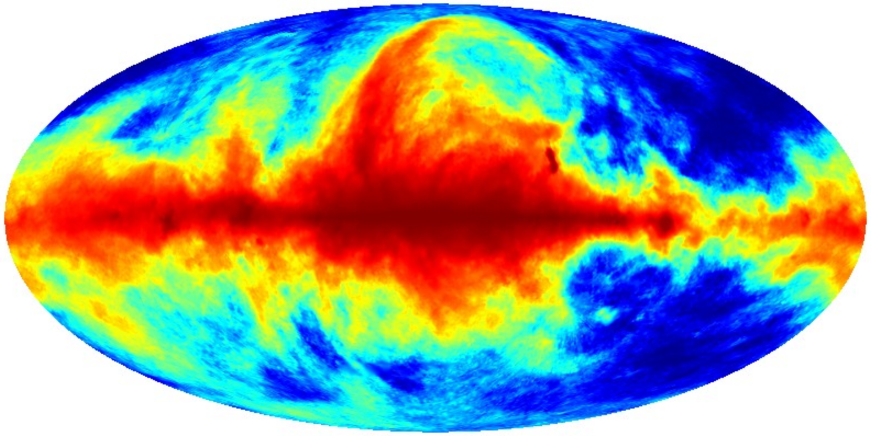


L. Olivari, M. Remazeilles, C. Dickinson,
"Extracting the 21 cm cosmological signal with GNILC", in prep. (2015)

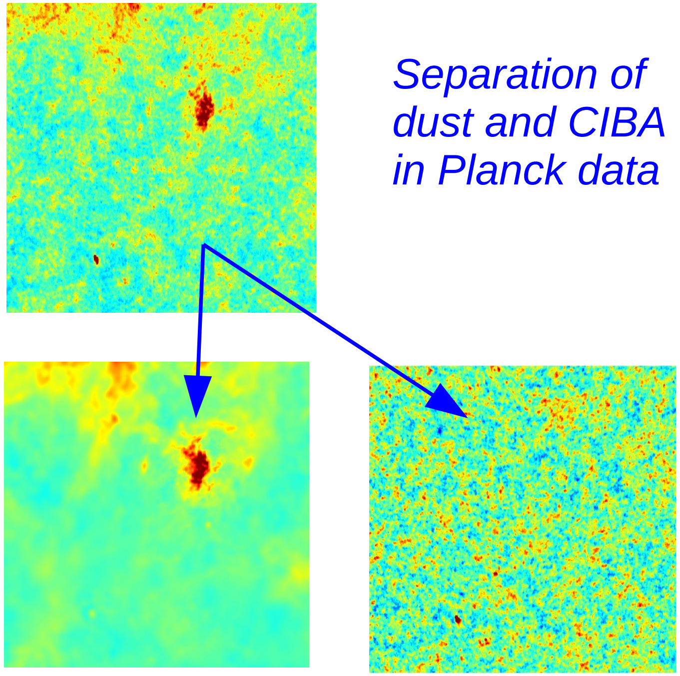
Summary

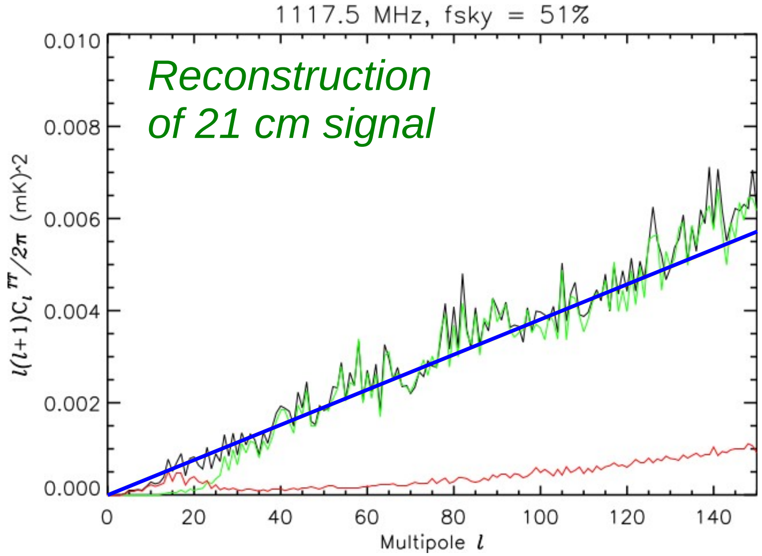
1. Radio foreground characterization : "2014 Reprocessed Haslam 408 MHz map"

Much better desourcing and destriping
(+ high-resolution template)



2. Component separation method : "Generalized Needlet Internal Linear Combination"

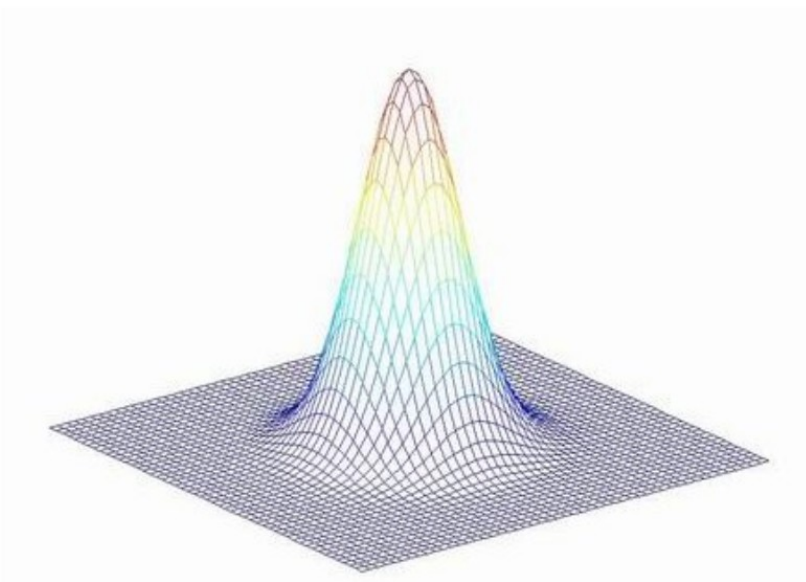




Backup slides

Source removal in the new 2014 Haslam map

$$
\begin{aligned}
F(X, Y; \{A_i\}) &= A_0 + A_1 X + A_2 Y \\
&+ A_3\, e^{-\frac{1}{2}\left(\frac{\cos\left(A_6 \frac{\pi}{180}\right)(X-A_4)-\sin\left(A_6 \frac{\pi}{180}\right)(Y-A_5)}{A_7}\right)^2 - \frac{1}{2}\left(\frac{\sin\left(A_6 \frac{\pi}{180}\right)(X-A_4)+\cos\left(A_6 \frac{\pi}{180}\right)(Y-A_5)}{A_8}\right)^2}
\end{aligned}
$$



9-parameter Gaussian fitting & subtraction

Iterative combination of two techniques depending on :

- *the local SNR*
- *the background geometry around the source*

Remazeilles et al., MNRAS (2015)

Minimum Curvature Spline Surface Inpainting

$$
\sum_{i=1}^{N} \|Z_i - F(X_i, Y_i)\|^2 + \lambda \iint \left[\left(\frac{\partial^2 F}{\partial X^2}\right)^2 + \left(\frac{\partial^2 F}{\partial Y^2}\right)^2 + 2\left(\frac{\partial^2 F}{\partial X \partial Y}\right)^2\right] dXdY
$$

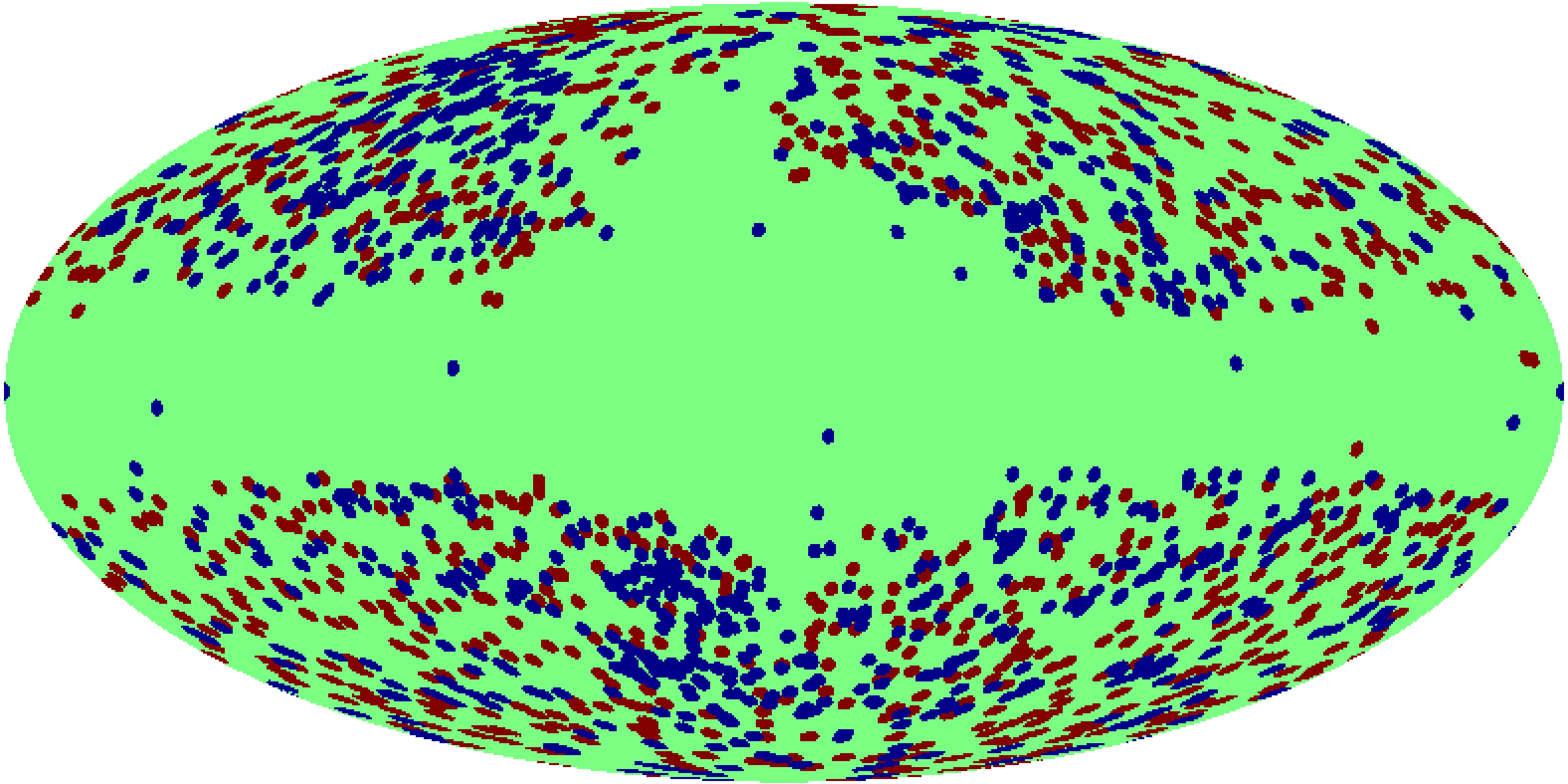
$$
F(X, Y) = \sum A_i d_i^2 \log d_i + a + bX + cY
$$

$$
\text{where } \; d_i^2 = (X - X_i)^2 + (Y - Y_i)^2
$$

Removed extragalactic sources (2 < S < 5 Jy)

RED : GAUSSIAN FITTING

BLUE : INPAINTING

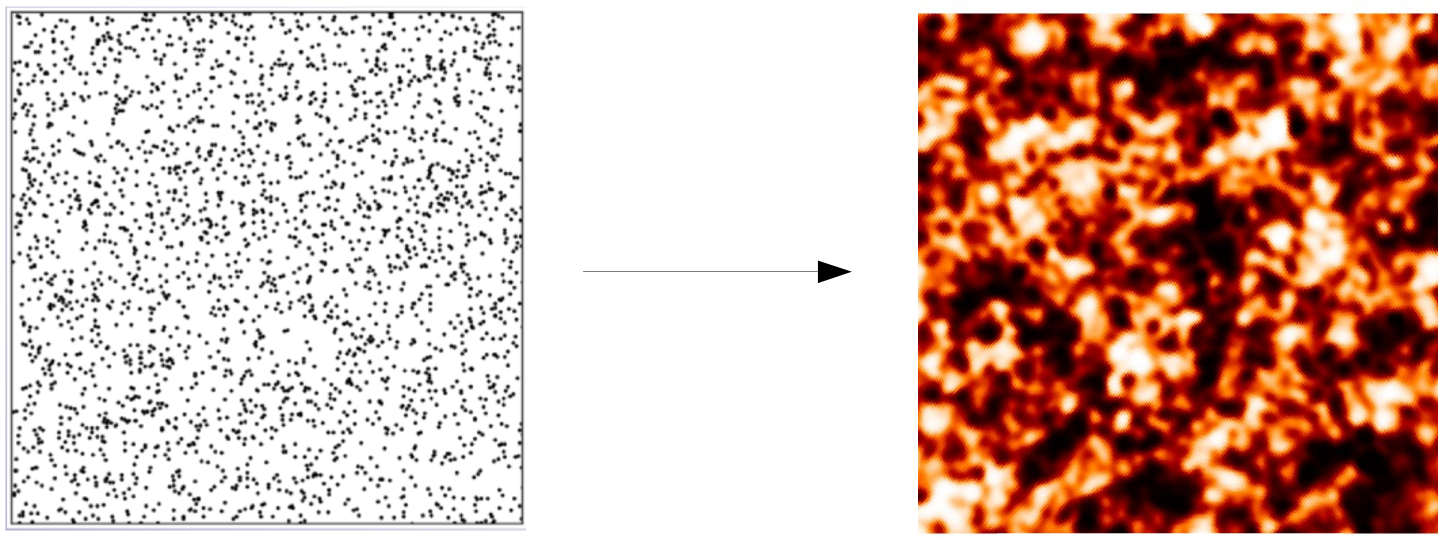


Remazeilles, Dickinson, Banday, Bigot-Sazy, Ghosh,
"An improved source-subtracted and destriped 408-MHz all-sky map",
MNRAS (2015)

HI 21 cm intensity mapping

- **Mapping unresolved 21 cm emission density field at different frequencies/redshifts**

  → *to probe the LSS we do not need to resolve individual galaxies (modest resolution)*

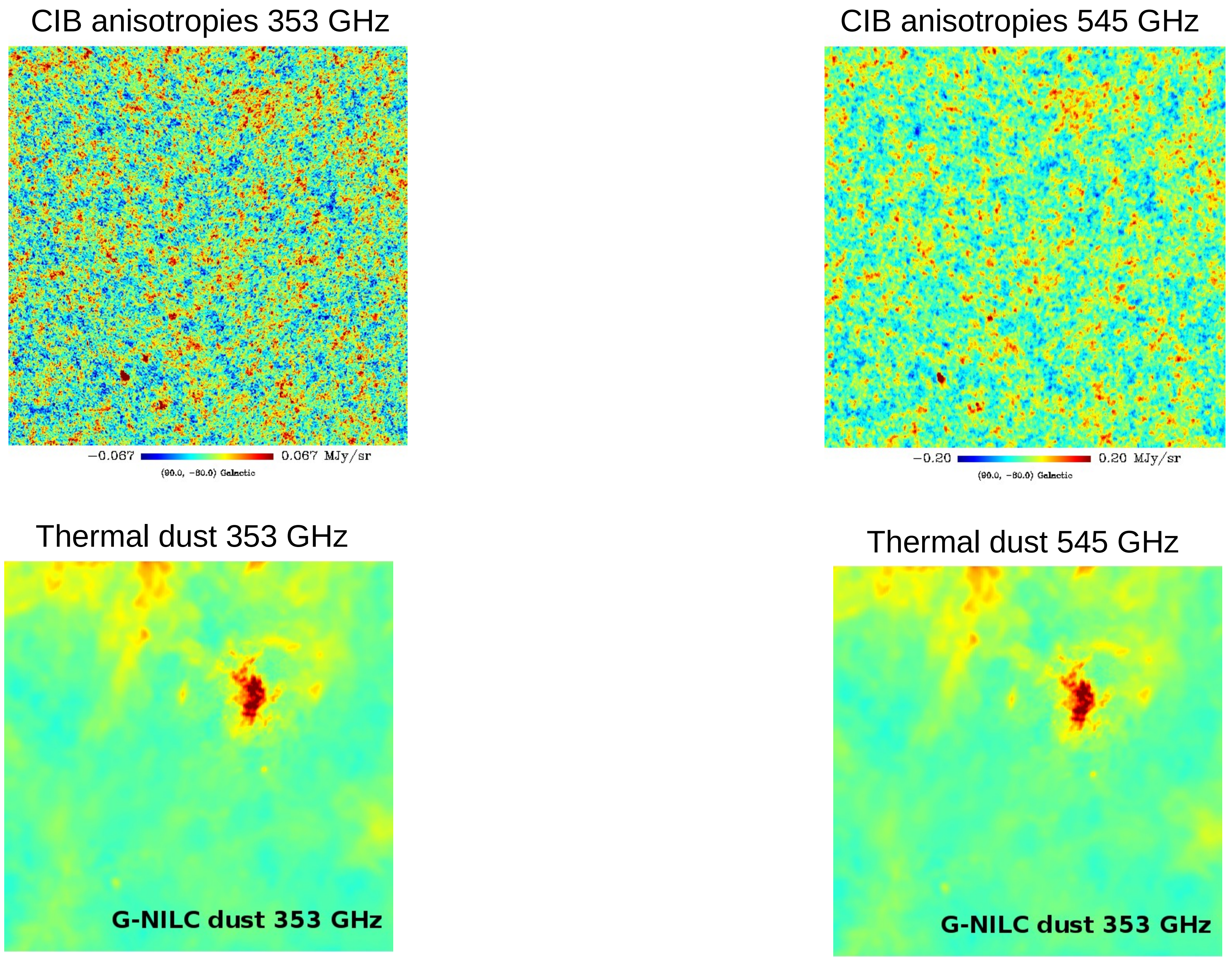


- **21 cm intensity mapping similar to CMB mapping**

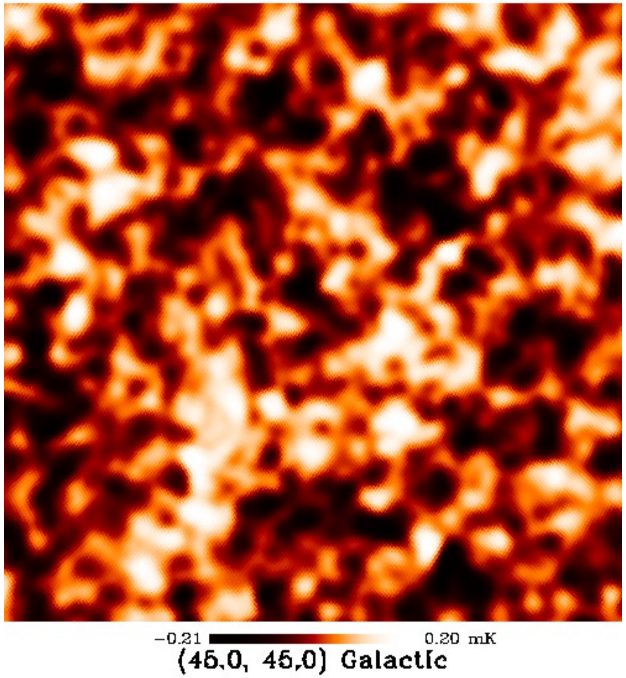
  → *possible knowledge transfer on component separation / foreground cleaning*

- **Alternative to optical surveys for constraining BAO and dark energy**

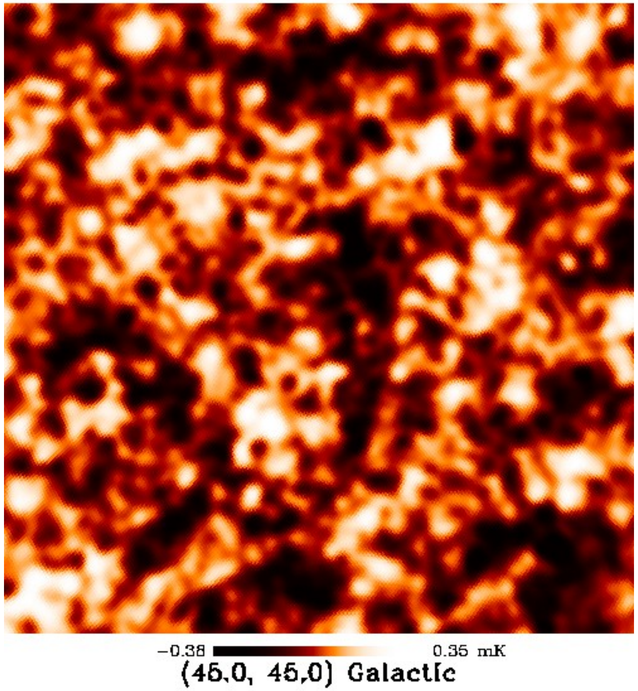
  → *as long as foregrounds can be accurately subtracted (Ansari et al. 2011)*


CIB anisotropies 353 GHz
−0.067
0.067 MJy/sr
(90.0, −80.0) Galactic
CIB anisotropies 545 GHz
−0.20
0.20 MJy/sr
(90.0, −80.0) Galactic
Thermal dust 353 GHz
G-NILC dust 353 GHz
Thermal dust 545 GHz
G-NILC dust 353 GHz

21 cm map 960 MHz (z=0.5)

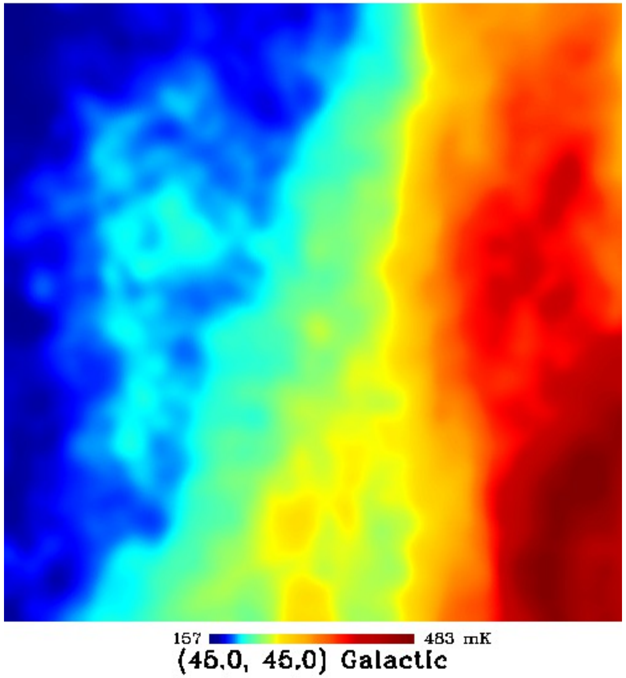


21 cm map 1250 MHz (z=0.1)

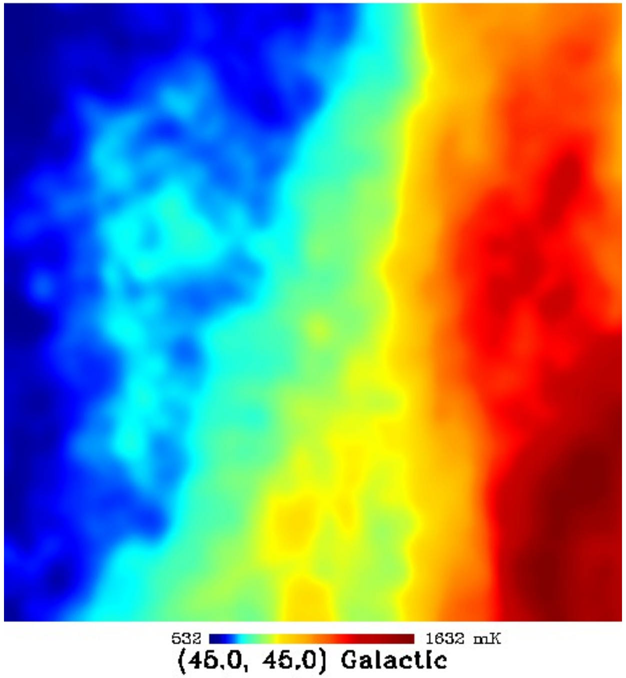


Same fight !

Synchrotron map 960 MHz



Synchrotron map 1250 MHz



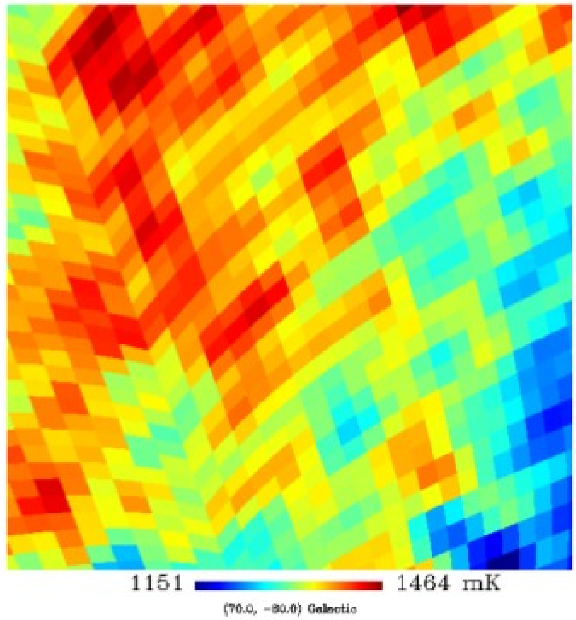
Simulation HI intensity mapping

Redshift range $[z_{\rm min}, z_{\rm max}]$	[0.13, 0.48]
Bandwidth $[\nu_{\rm min}, \nu_{\rm max}]$ (MHz)	[960, 1260]
Number of feed horns $n_{\rm f}$	60
Sky coverage $\Omega_{\rm sur}$ (deg^2)	21000
Observation time $t_{\rm obs}$ (yrs)	1
System temperature $T_{\rm sys}$ (K)	40

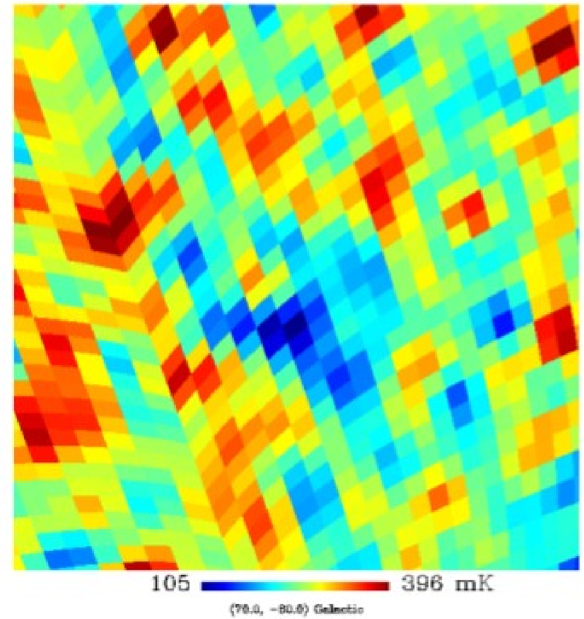
Simulation : radio intensity mapping

Remazeilles et al., 2015

Synchrotron – 1117 MHz



Point sources – 1117 MHz

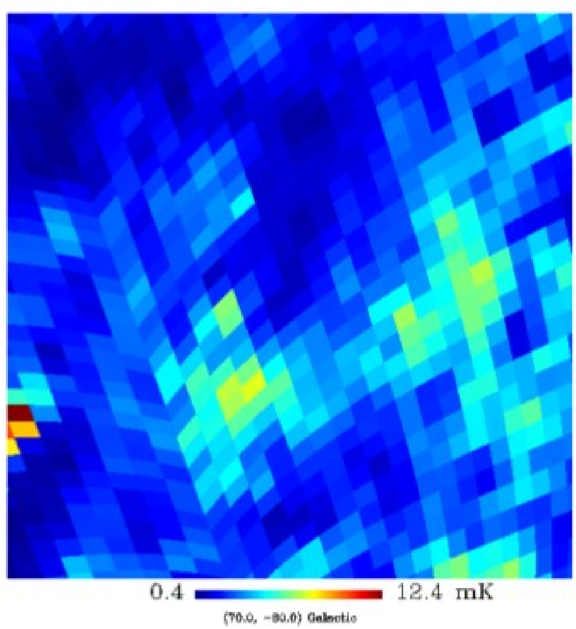


Battye et al., 2013

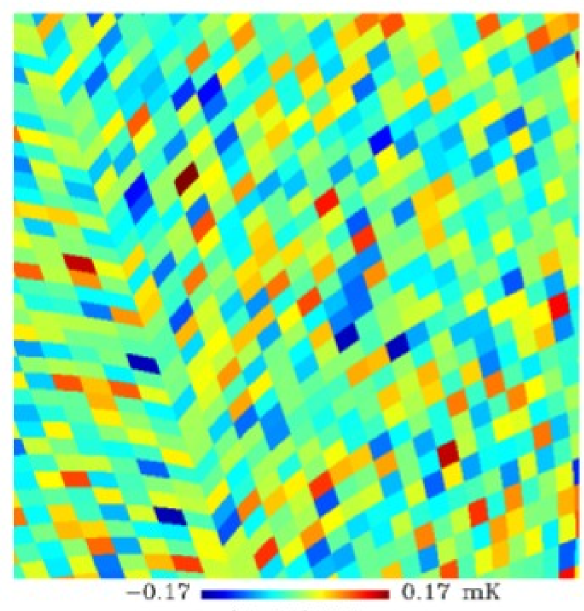
12° x 12 °

~ 40 ' resolution

Free-free – 1117 MHz



Thermal noise – 1117 MHz



Dickinson et al., 2003

BINGO noise level

Generalized Needlet ILC (GNILC)

Remazeilles, Delabrouille, and Cardoso, MNRAS (2011)

- **Wavelet decomposition** of the data to deal with **local conditions of contamination**
- **Data covariance matrix** (n x n frequencies) $\mathbf{R} = \mathbf{R}_{21cm} + \mathbf{R}_{foregrounds} + \mathbf{R}_{noise}$
- **Prior on 21 cm** power spectra $\left(\mathbf{R}^{prior}_{21cm}\right)^{-\frac{1}{2}} \mathbf{R} \left(\mathbf{R}^{prior}_{21cm}\right)^{-\frac{1}{2}} \approx \mathbf{I} + \mathbf{R}_N$
- **Eigen-decomposition** (PCA driven by local SNR = data / 21cm power)

$$\left[\begin{array}{c|c} \cdots & \mathbf{U}_S \end{array}\right] \cdot \left[\begin{array}{ccc|c} 1+\lambda_1 & & & \\ & \cdots & & \\ & & 1+\lambda_m & \\ \hline & & & \mathbf{I} \end{array}\right] \cdot \left[\begin{array}{c} \cdots \\ \hline \mathbf{U}_S^T \end{array}\right]$$

m eigenvalues >> 1
→ foreground+noise power

(n-m) eigenvalues ≈ 1
→ 21 cm power

$\mathbf{U_S}$ → eigenmodes of the 21 cm subspace

- **m** estimated by minimizing Akaike Information Criterion $AIC(m) = 2m - 2\log \mathcal{L}_{max}(m)$

- **Multidimensional ILC**

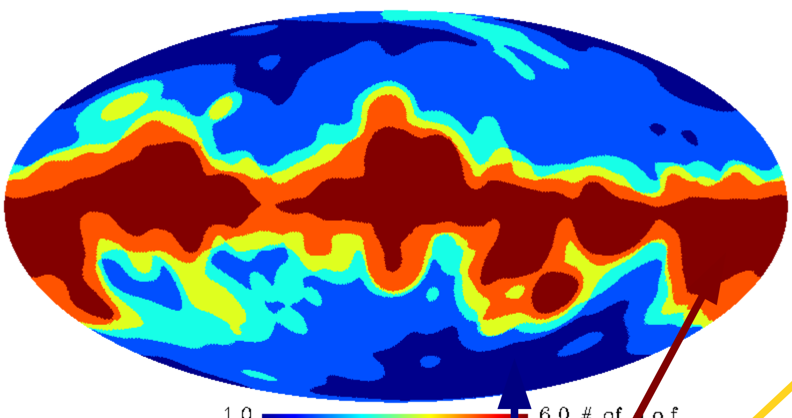
$$\mathbf{A}\left(\mathbf{A}^T\mathbf{R}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{R}^{-1} \quad \textit{where} \quad \mathbf{A} = \left(\mathbf{R}^{prior}_{21cm}\right)^{1/2} \mathbf{U_S}$$

Generalized Needlet ILC (GNILC)

Remazeilles, Delabrouille, and Cardoso, MNRAS (2011)

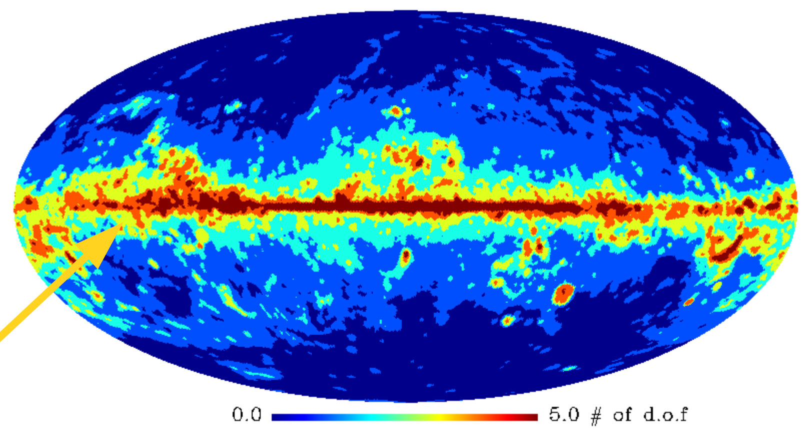
- **Wavelet decomposition** of the data to deal with **local conditions of contamination**
- **Data covariance matrix** (n x n frequencies) $\mathbf{R} = \mathbf{R}_{21cm} + \mathbf{R}_{foregrounds} + \mathbf{R}_{noise}$
- **Prior on 21 cm** power spectra $\left(\mathbf{R}^{prior}_{21cm}\right)^{-\frac{1}{2}} \mathbf{R} \left(\mathbf{R}^{prior}_{21cm}\right)^{-\frac{1}{2}} \approx \mathbf{I} + \mathbf{R}_N$
- **Eigen-decomposition** (PCA driven by local SNR = data / 21cm power)

Wavelet scale 1 degree



Number of foreground modes

Wavelet scale 5 arcminute



Number of foreground modes

- **Multidimensional ILC**

$$\boxed{\mathbf{A}\left(\mathbf{A}^T\mathbf{R}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{R}^{-1}}$$ *where* $\mathbf{A} = \left(\mathbf{R}^{prior}_{21cm}\right)^{1/2} \mathbf{U_S}$

Table 3. The average difference between the input and the reconstructed HI power spectrum normalized by the input HI power spectrum for the GNILC method and the PCA with 3 principal components.

$$\frac{C_\ell^{input} - C_\ell^{output}}{C_\ell^{input}}$$

Range of multipoles	GNILC	PCA
0 - 15	0.91	0.06
15 - 30	0.53	0.12
30 - 45	0.06	0.14
45 - 60	0.07	0.18
60 - 75	0.04	0.19
75 - 90	0.05	0.26
90 - 105	0.04	0.29
105 - 120	-0.001	0.33
120 - 135	-0.006	0.39
135 - 150	-0.01	0.46

Impact of prior