Tianlai data Analysis and Simulations

R. Ansari

September 2016 - Tianlai Workshop @FermiLab

- Overview of Data flow and data processing
- Map making for wide field transit interferometers
  - Application to PAON-4 & Tianlai
- Calibration, application to PAON-4 data
- Technical considerations , Simulations

Tianlai data processing overview

Level 0 (**L0**)

Visibility data, computed on-site ,
using dedicated hardware (correlator), or
by software
ancillary / housekeeping data

Level 1 (**L1**)

Raw visibility data
[Vij (ν)]

(L0 output)

→

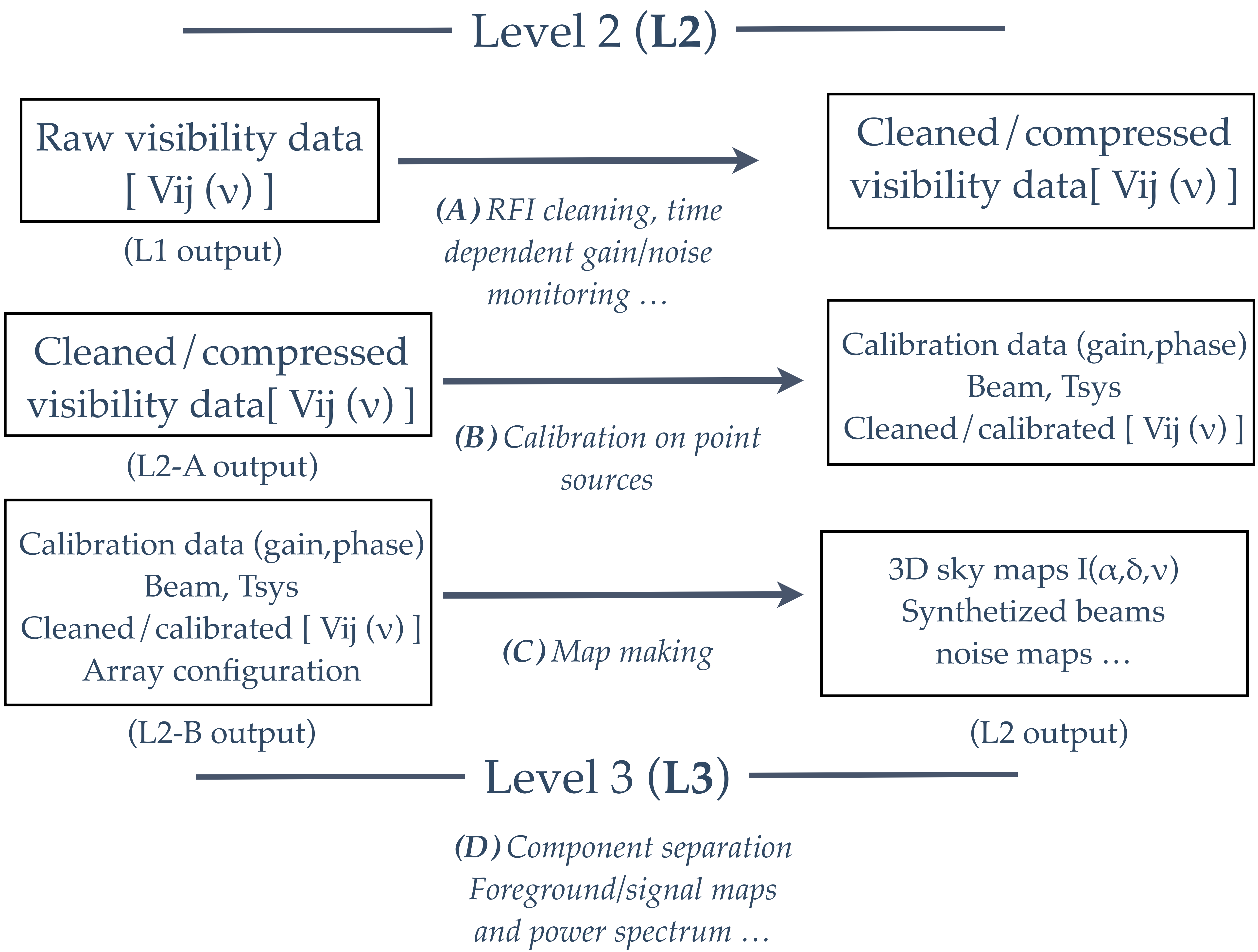
- First stage RFI cleaning,
- data quality monitoring
- data compression, mainly through time averaging
- transfer to TAC

Organized, Compressed, Time sliced visibility data

(L1 output)

L1 output data :

- T-16DA : ~35-70 GB / day , ~24-288 files / day , ~100 TB / year
- T-3Cyl: ~1200-2500 GB / day , ~240-2880 files / day , ~3500 TB / year


Level 2 (L2)
Raw visibility data [Vij (ν)]
(L1 output)
(A) RFI cleaning, time dependent gain/noise monitoring …
Cleaned / compressed visibility data[Vij (ν)]
Cleaned / compressed visibility data[Vij (ν)]
(L2-A output)
(B) Calibration on point sources
Calibration data (gain,phase)
Beam, Tsys
Cleaned / calibrated [Vij (ν)]
Calibration data (gain,phase)
Beam, Tsys
Cleaned / calibrated [Vij (ν)]
Array configuration
(L2-B output)
(C) Map making
3D sky maps I(α,δ,ν)
Synthetized beams
noise maps …
(L2 output)
Level 3 (L3)
(D) Component separation
Foreground/signal maps
and power spectrum …

L1,L2 processings

- (A) : Cleaning raw visibility data (RFI removal, time dependent gain / Tsys monitoring, data compression (rebinning in time)
- (B) : Relative gain/phase calibration using single bright point sources - should be then extended to the use of multiple point sources. Will also provide single dish+feed beam response and Tsys
- (C) : Map making - 3D intensity map reconstruction

Map making - Power spectrum

- Snap shots using (u,v) plane formalism of linear combination of visibilities (for each frequency)
- Full map reconstruction in spherical geometry [J. Zhang, S. Zuo]
- Foreground subtraction in angular/frequency domain
- and power spectrum computation from the cleaned maps
- Or, direct power spectrum computation from visibilities ?

Spherical Harmonics Map reconstruction applied to Tianlai & PAON-4 - Jiao Zhang (NAOC & LAL)

J. Zhang et al. 21cm Map Making Papers

Map Making Paper I - J. Zhang et al (2016), MNRAS, (accepted June 2016, submitted end of March 2016)

MNRAS **000**, 1–17 (2016) Preprint 30 May 2016 Compiled using MNRAS LaTeX style file v3.0

Sky reconstruction from transit visibilities: PAON-4 and Tianlai Dish Array

Jiao Zhang[1,2,3], Reza Ansari[2] ⋆, Xuelei Chen[1,3,4], Jean-Eric Campagne[2], Christophe Magneville[5], and Fengquan Wu[1]

[1]*Key Laboratory of Computational Astrophysics, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*
[2]*Université Paris-Sud, LAL, UMR 8607, F-91898 Orsay Cedex, France & CNRS/IN2P3, F-91405 Orsay, France*
[3]*University of Chinese Academy of Sciences, Beijing 100049, China*
[4]*Center for High Energy Physics, Peking University, Beijing 100871, China*
[5]*CEA, DSM/IRFU, Centre d'Etudes de Saclay, F-91191 Gif-sur-Yvette, France*

Map Making Paper II - J. Zhang et al (2016), accepted for publication in RAA (May 2016, submitted in April)

Research in Astronomy and Astrophysics manuscript no.
(LaTeX: jmapcylinder.tex; printed on May 31, 2016; 18:04)

Sky reconstruction for the Tianlai cylinder array

Jiao Zhang[1,2,3], Shifan Zuo[1,3], Reza Ansari[2], Xuelei Chen[1,3,4] ,Yichao Li[1,3], Fengquan Wu[1,3],Jean-Eric Campagne[2], Christophe Magneville[5]

Slide by J. Zhang

Transit interferometer : map making

Full East-West scan means we could perform Fourier transform on the set of $\mathcal{V}_{ij}(\alpha_p)$ to obtain the $\tilde{\mathcal{V}}_{ij}(m)$ for each m-mode

$$\mathcal{V}_{ij}(\alpha_p) \quad (0 \leqslant \alpha_p < 2\pi) \longrightarrow \tilde{\mathcal{V}}_{ij}(m) \qquad \textit{for all m modes}$$

$$\tilde{\mathcal{V}}_{ij}(m) = \sum_{\ell=|m|}^{+\ell_{\max}} (-1)^m \mathcal{I}(\ell, m) \mathcal{L}_{ij}(\ell, -m)$$

$$\tilde{\mathcal{V}}^*_{ij}(-m) = \sum_{\ell=|m|}^{+\ell_{\max}} \mathcal{I}(\ell, m) \mathcal{L}^*_{ij}(\ell, m)$$

See also R. Shaw et al. Apj 2014, 2015

The full problem of all $\tilde{\mathcal{V}}_{ij}(m)$ could be separated into a set of independent problems for each m.

Block diagonal matrix

$$\begin{pmatrix} \tilde{\mathcal{V}}^{\delta_p}_{ij}(m_0) \\ \tilde{\mathcal{V}}^{\delta_p}_{ij}(m_1) \\ \cdots \\ \tilde{\mathcal{V}}^{\delta_p}_{ij}(m_{max}) \end{pmatrix} = \begin{pmatrix} \mathcal{L}^{\delta_p}_{ij}(m_0, l) & & 0 \\ & \mathcal{L}^{\delta_p}_{ij}(m_1, l) & \\ 0 & \cdots & \\ & & \mathcal{L}^{\delta_p}_{ij}(m_{max}, l) \end{pmatrix} \times \begin{pmatrix} \mathcal{I}(m_0, l) \\ \mathcal{I}(m_1, l) \\ \cdots \\ \mathcal{I}(m_{max}, l) \end{pmatrix} + noise$$

$\mathbf{L}_{m_i}$ Inversion $\mathbf{H}_m$

Slide by J. Zhang

Instrument response and transfer function

- Reconstruction matrix $\mathbf{R}_m \equiv (\mathbf{H}_m \mathbf{L}_m)$: the estimated sky spherical harmonics coefficient $\widehat{\mathcal{I}}(\ell, m)$ are related to the true sky ones $\mathcal{I}(\ell, m)$

- Core response matrix $\mathbf{R}$:

$$\mathbf{R}(\ell, m) = \mathbf{R}_m(\ell, \ell)$$

- Transfer function

$$T(\ell) = \langle |\mathbf{R}(\ell, m)| \rangle_m \quad \textbf{or} \quad T(\ell) = \frac{C^{\text{rec}}(\ell)}{C^{\text{in}}(\ell)}$$

Error variance matrix and noise power spectrum

- Pure noise visibilities → noise covariance matrix
- $\mathbf{H}_m$ matrix and noise covariance matrix → for each m mode, the covariance matrix $\mathbf{Cov}_m(\ell_1, \ell_2)$ of the estimator $\widehat{\mathcal{I}}(\ell, m)$

$$\mathbf{N}_m = \left[\tilde{\mathcal{V}}_{ij}\right]_m \cdot \left[\tilde{\mathcal{V}}_{ij}\right]_m^\dagger$$

$$\mathbf{Cov}_m(\ell_1, \ell_2) = \langle\left[\widehat{\mathcal{I}}(\ell)\right]_m \cdot \left[\widehat{\mathcal{I}}(\ell)\right]_m^\dagger\rangle = \mathbf{H}_m \mathbf{N}_m \mathbf{H}_m^\dagger$$

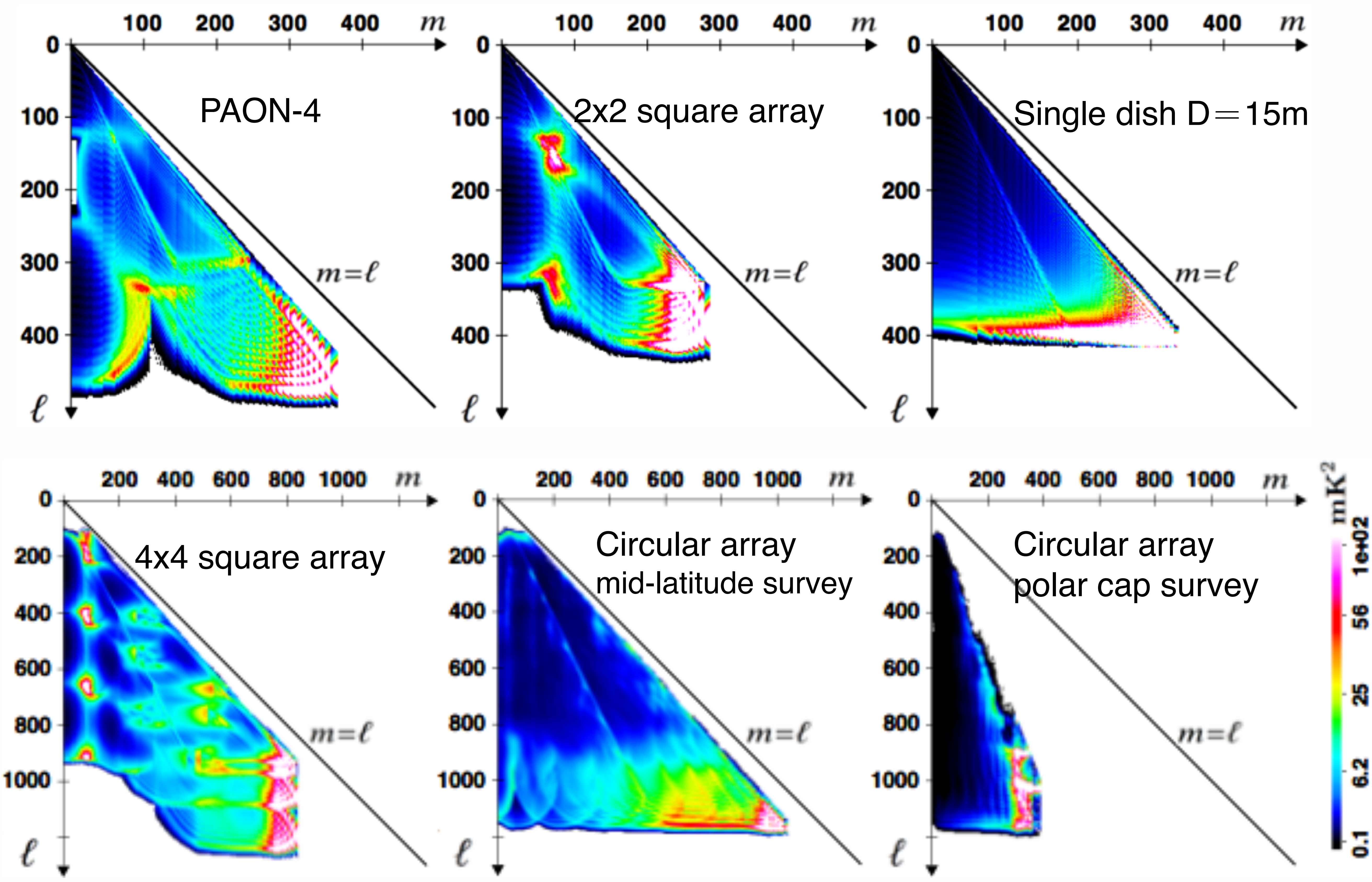
- Error variance matrix and noise power spectrum

$$\sigma_{\mathcal{I}}^2(\ell, m) = \mathbf{Cov}_m(\ell, \ell)$$

$$C^{\text{noise}}(\ell) = \langle\sigma_{\mathcal{I}}^2(\ell, m)\rangle_m$$

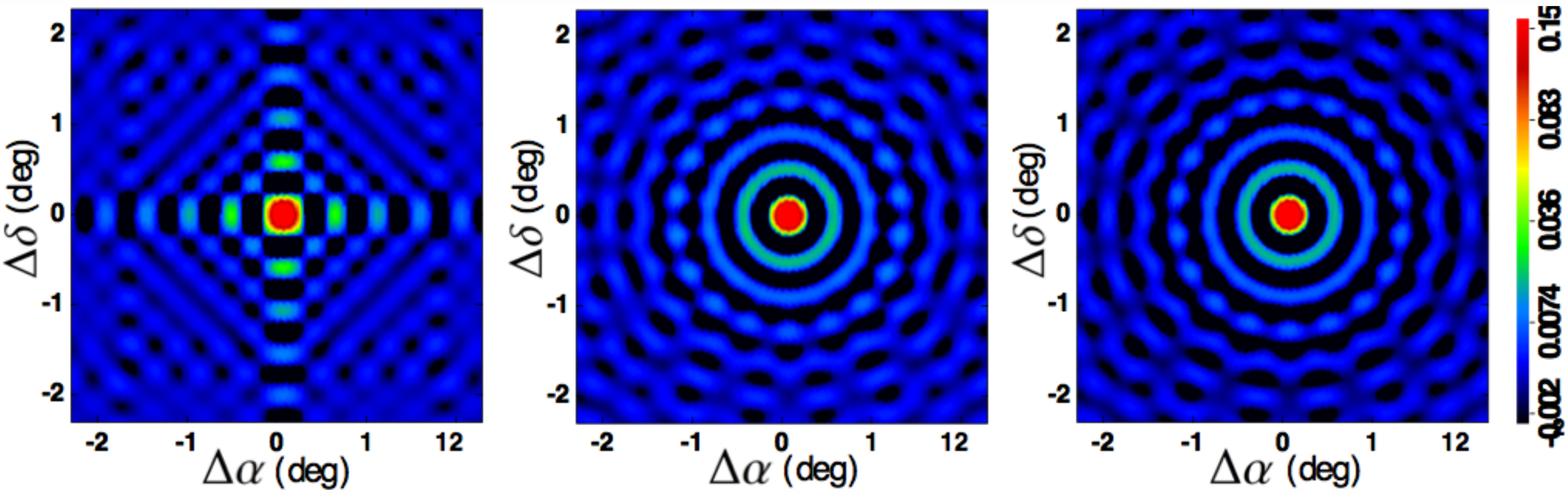
Slide by J. Zhang

Error variance matrix $\sigma_I^2(\ell, m)$



Slide by J. Zhang

Reconstructed beam or PSF for Tianlai dish array



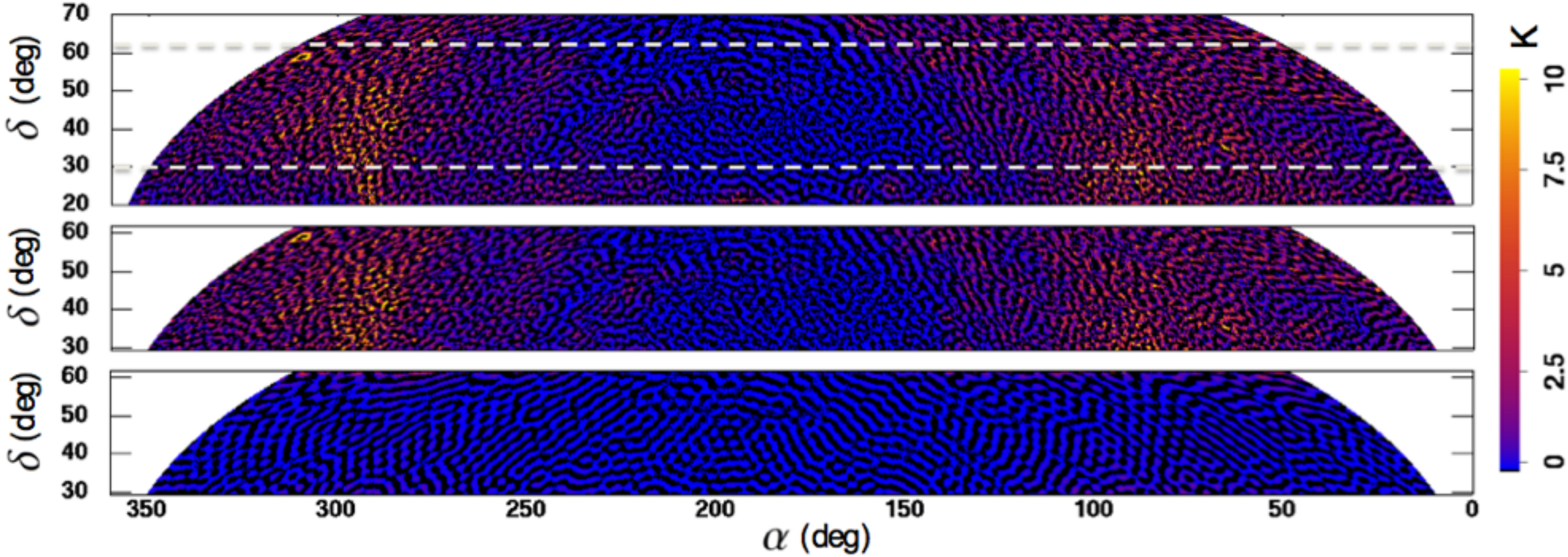
4x4 square array

Circular array
With autocorrelation

Circular array
Without autocorrelation

Slide by J. Zhang

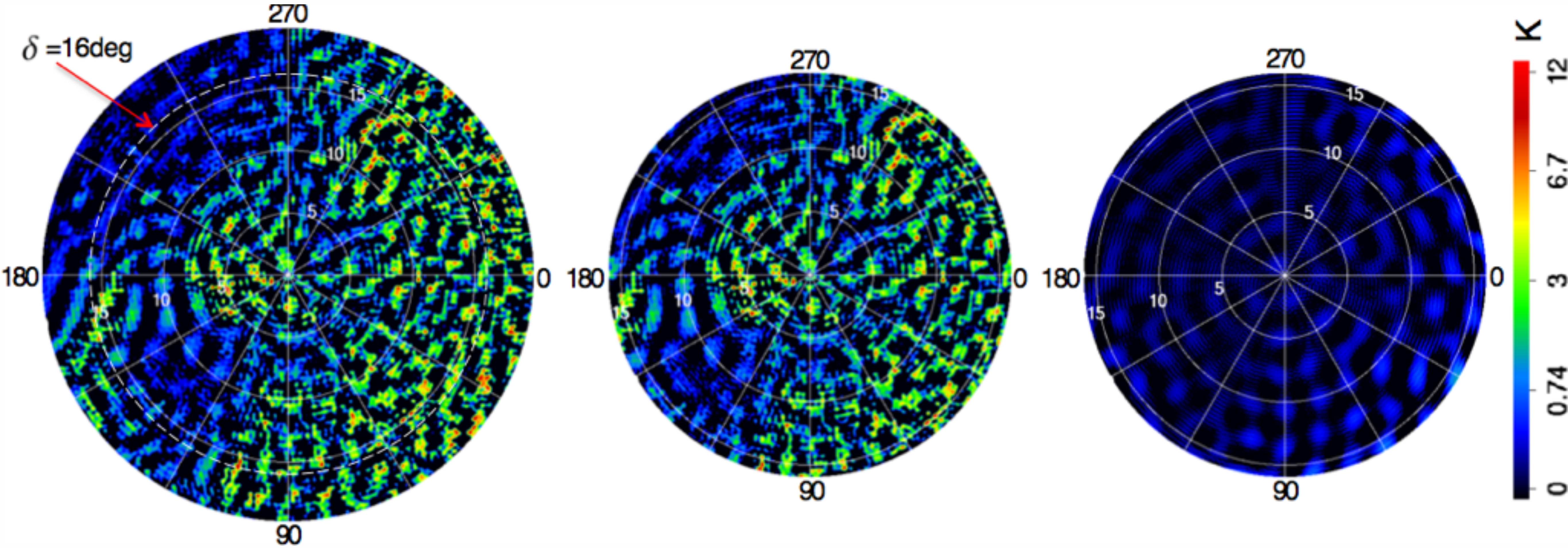
Reconstructed sky map for mid-latitude



- Top: the input map in the latitude range $20^\circ < \delta < 70^\circ$ obtained by applying high pass filter to the LAB map
- Center: the reconstructed map in the range $30^\circ < \delta < 60^\circ$ with filters
- Bottom: the difference between (1) and (2) in the range $30^\circ < \delta < 60^\circ$

Slide by J. Zhang

Reconstructed sky map for polar region

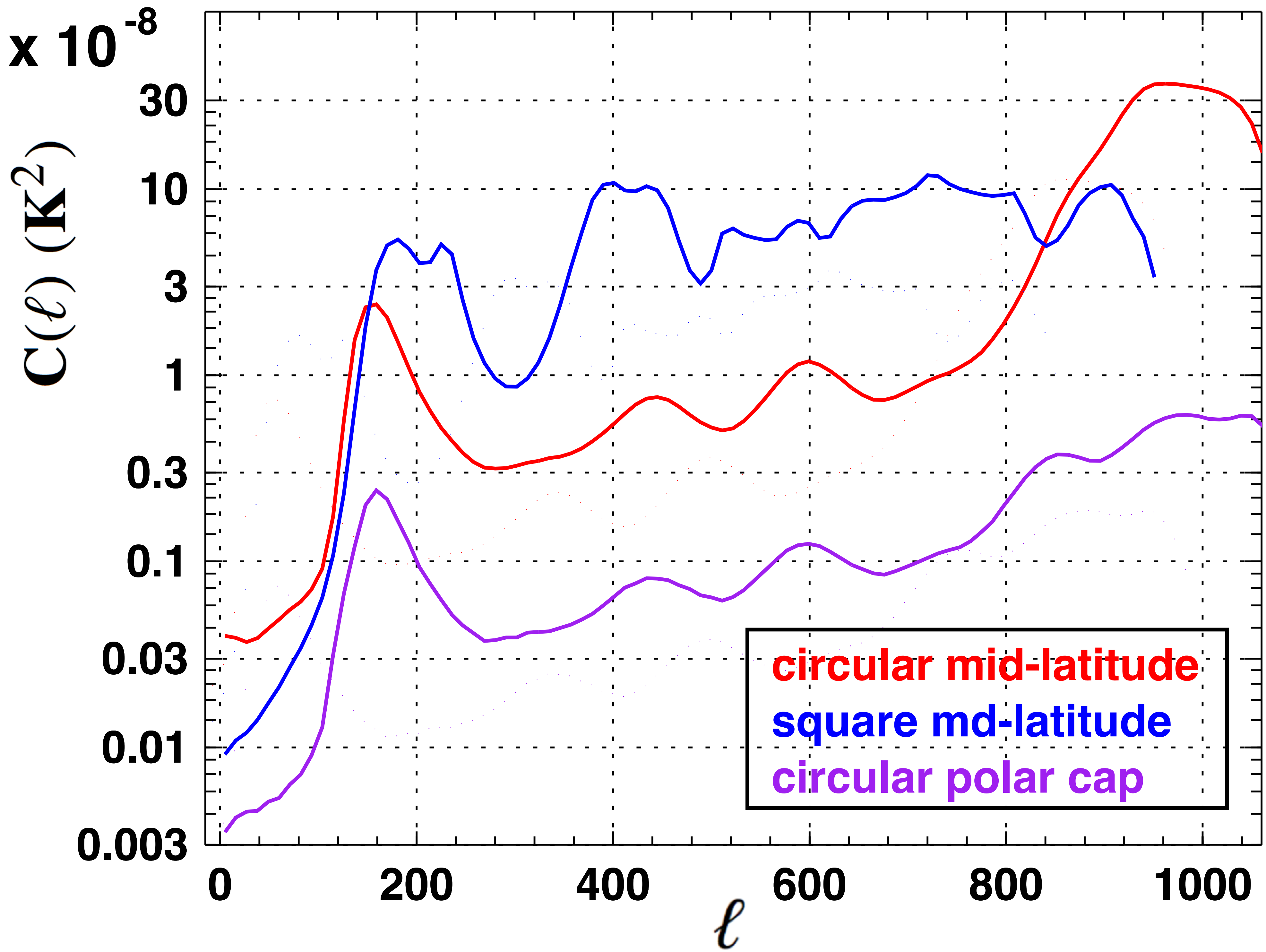


- Left : the input polar cap map after a high pass filtering with a radius of 20°
- Center : the reconstructed map for polar cap survey with a radius of 16°
- Right : The difference between left and center with a radius of 16°

Slide by J. Zhang

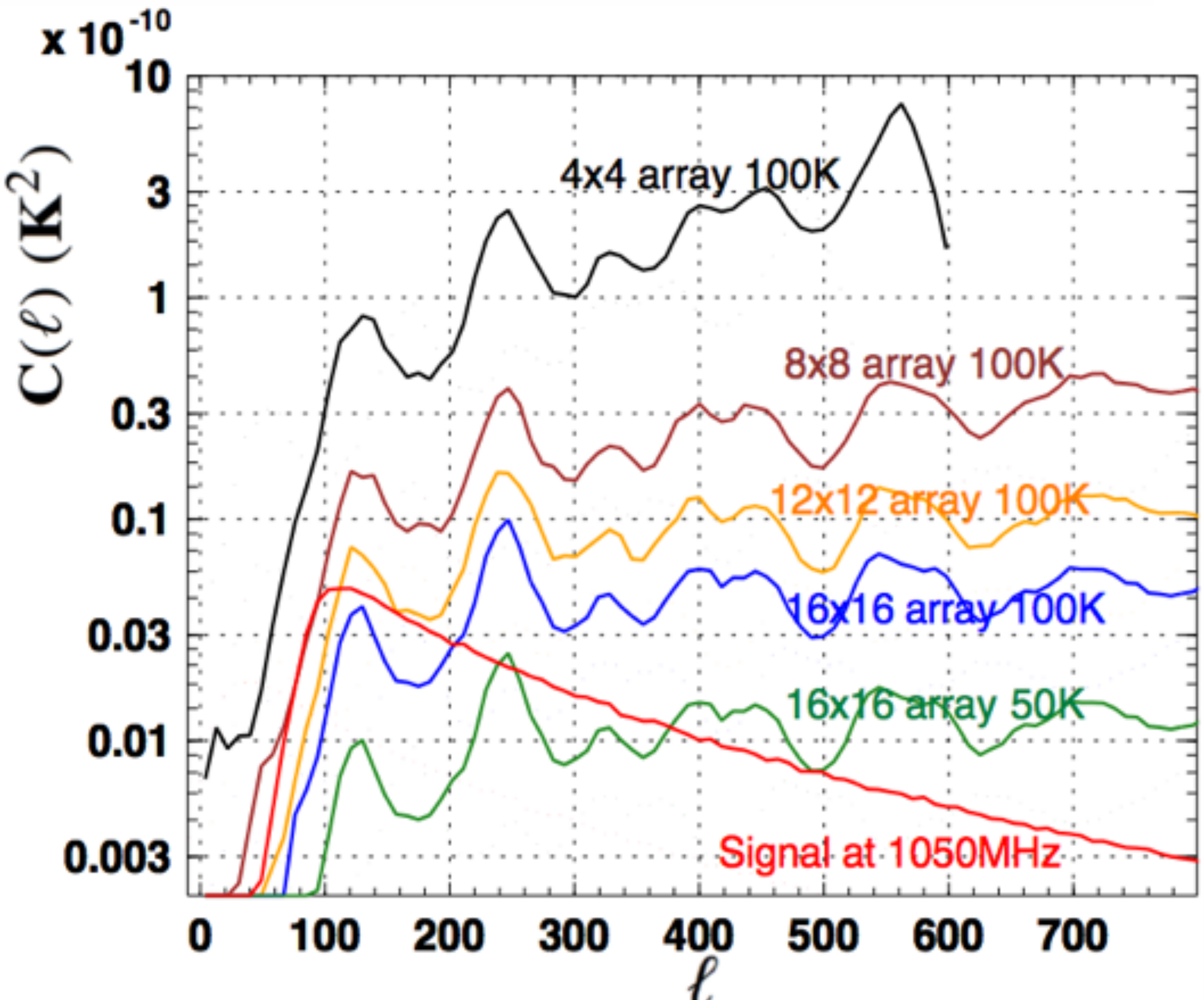
Application to PAON-4 and Tianlai dish array

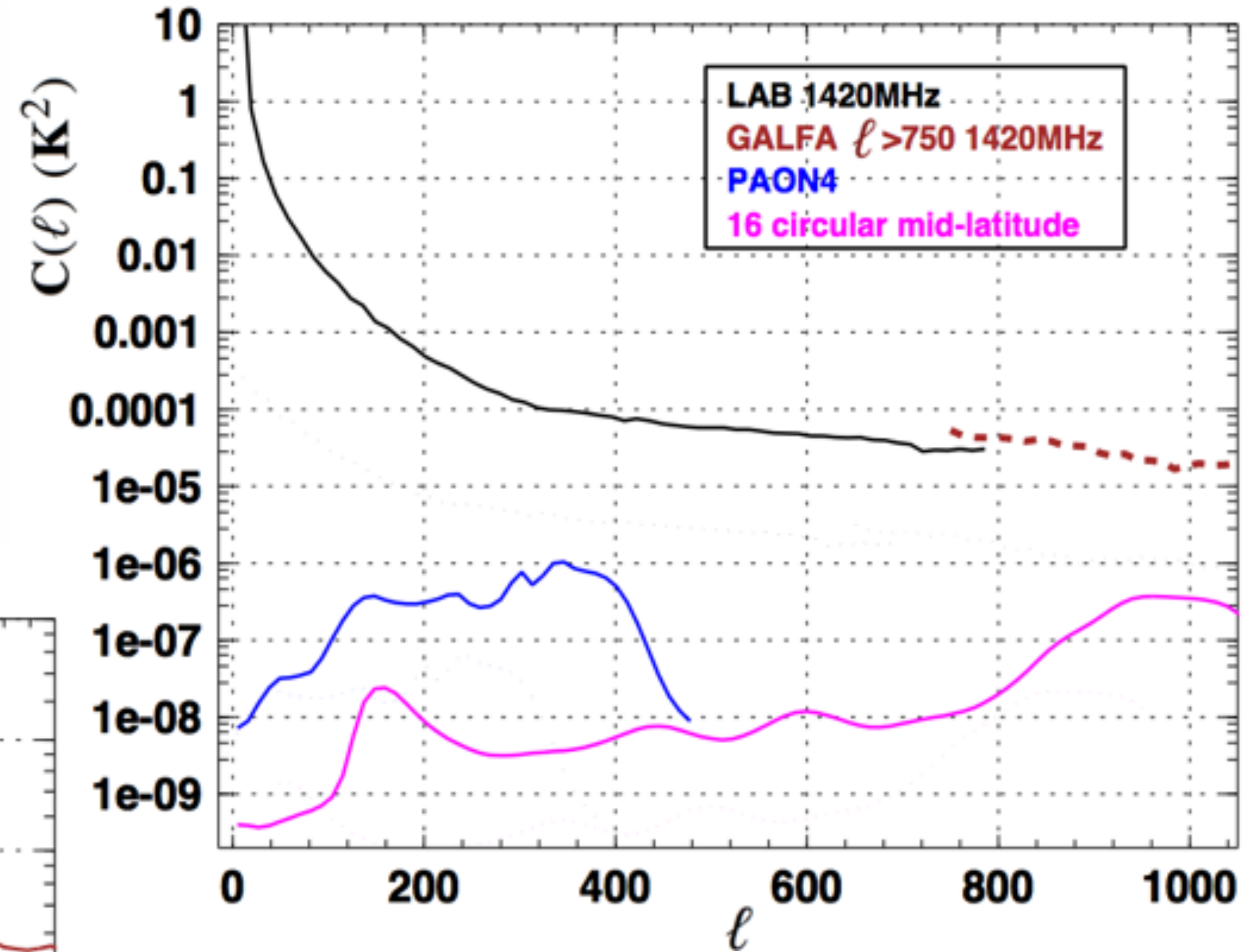
noise power spectrum



Slide by J. Zhang

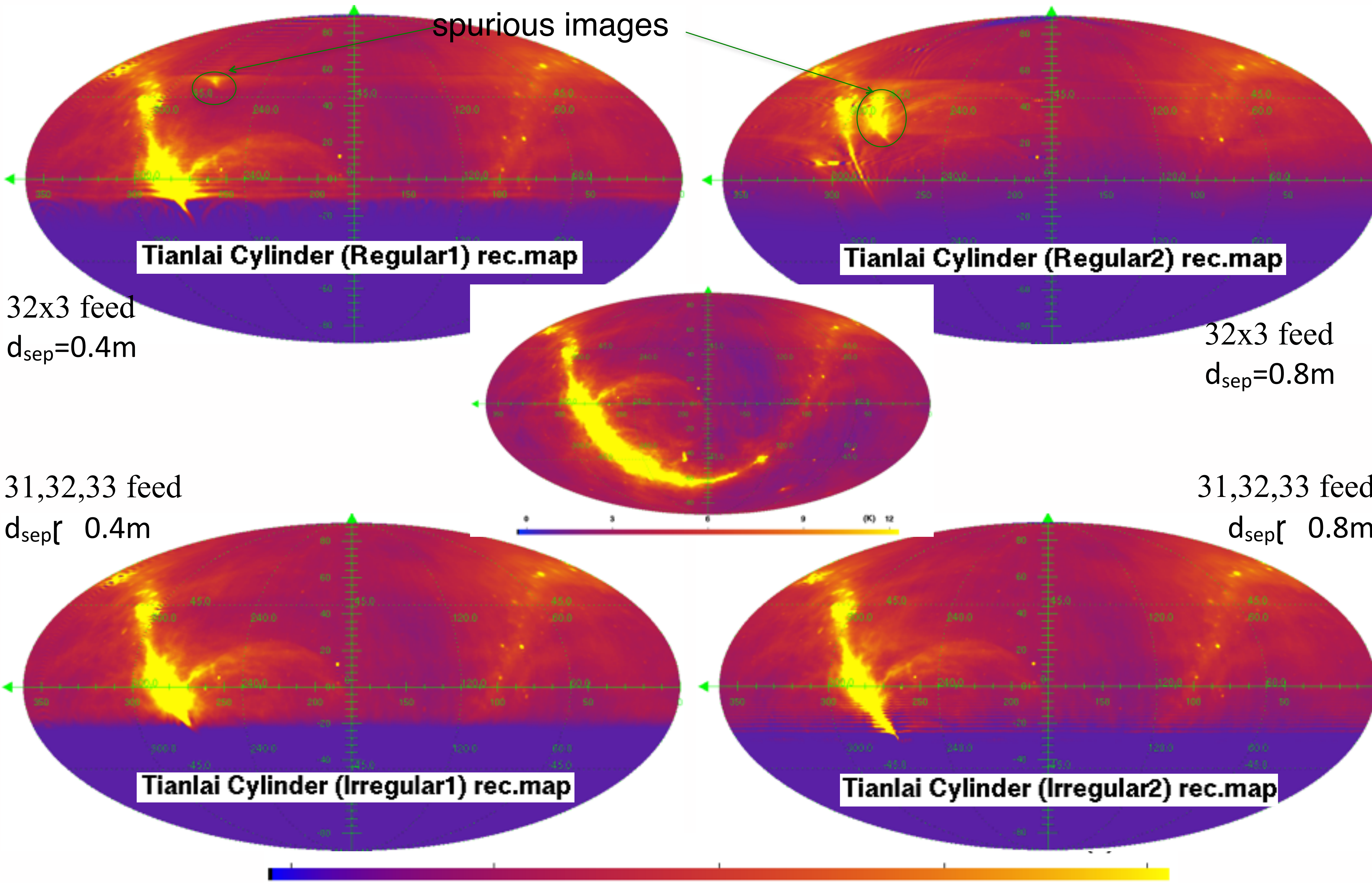
Dish array sensitivity





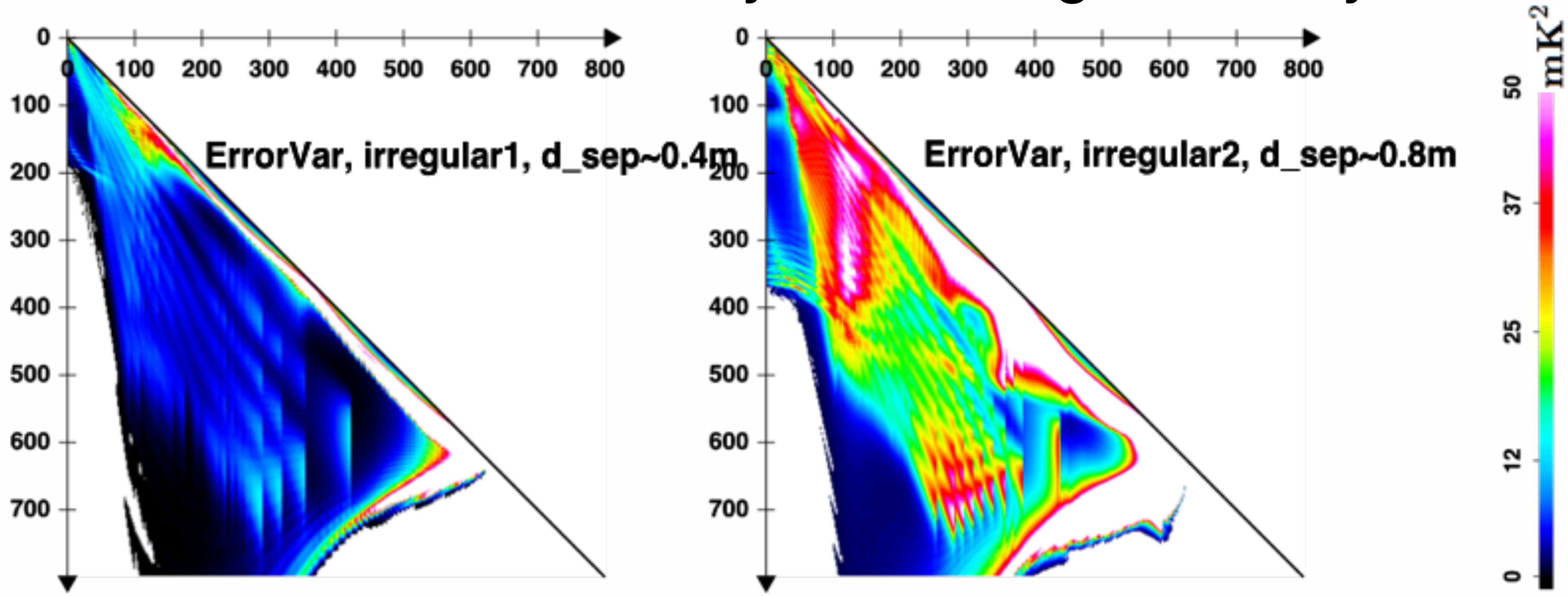
Slide by J. Zhang

Reconstructed sky map for cylinder array

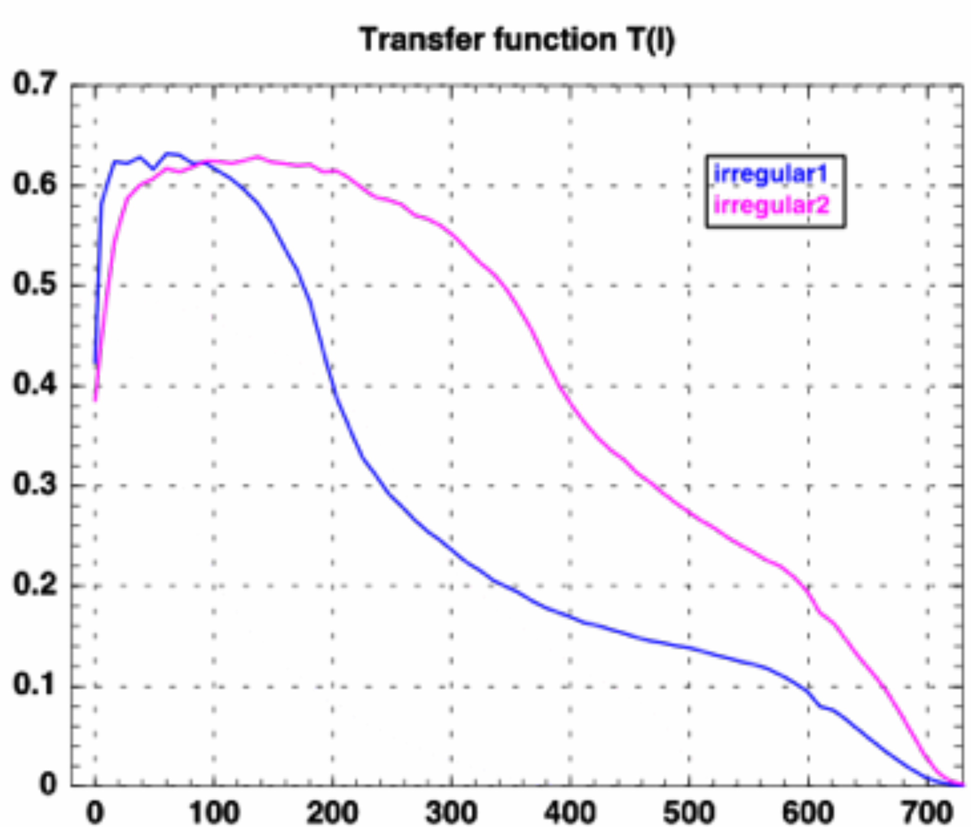


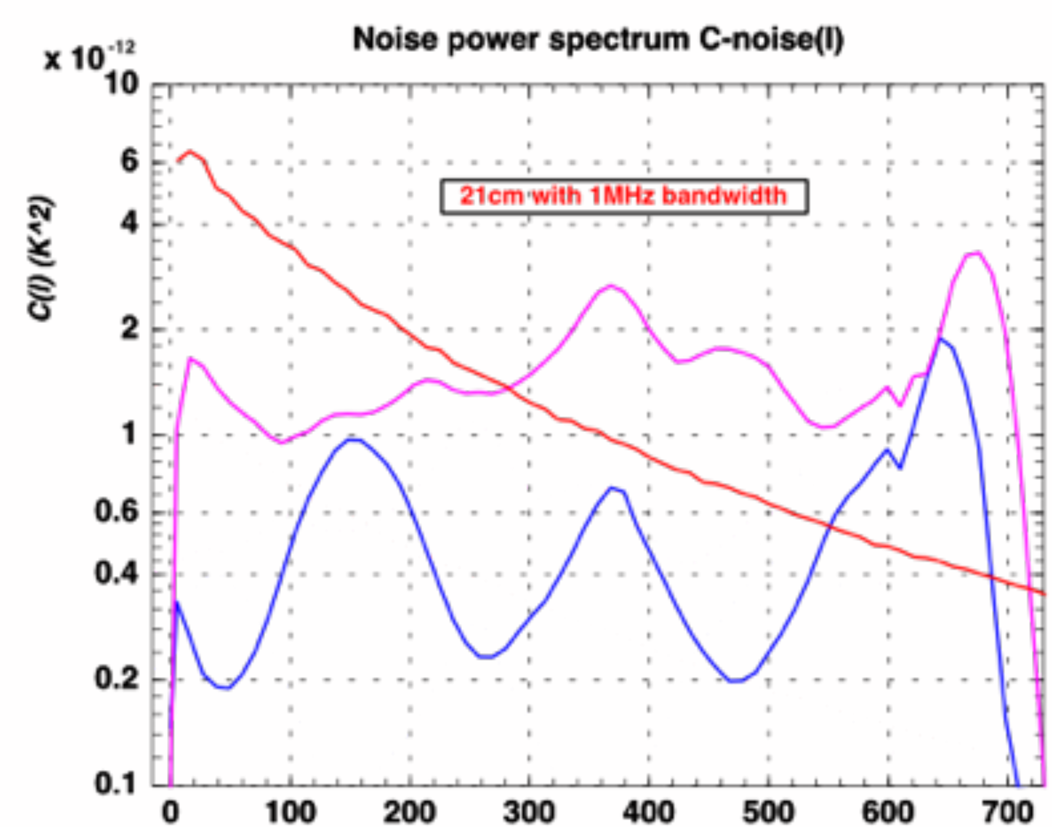
Slide by J. Zhang

Error variance matrix for cylinder irregular array



Transfer function and noise power spectrum





Calibration (PAON-4 & Tianlai)

Analysis mostly done
by Q. Huang, but also T.
Etourneau …

PAON-4 (PI: J.E. Campagne, J.M. Martin) - Technical projet leaders:
F. Rigaud (Mechanics) - D. Charlet (Electronic, Computing, Commissioning)

PAON-4 test interferometer

PAON-2 →

September 2012

Extraction of array characteristics from cross-correlations

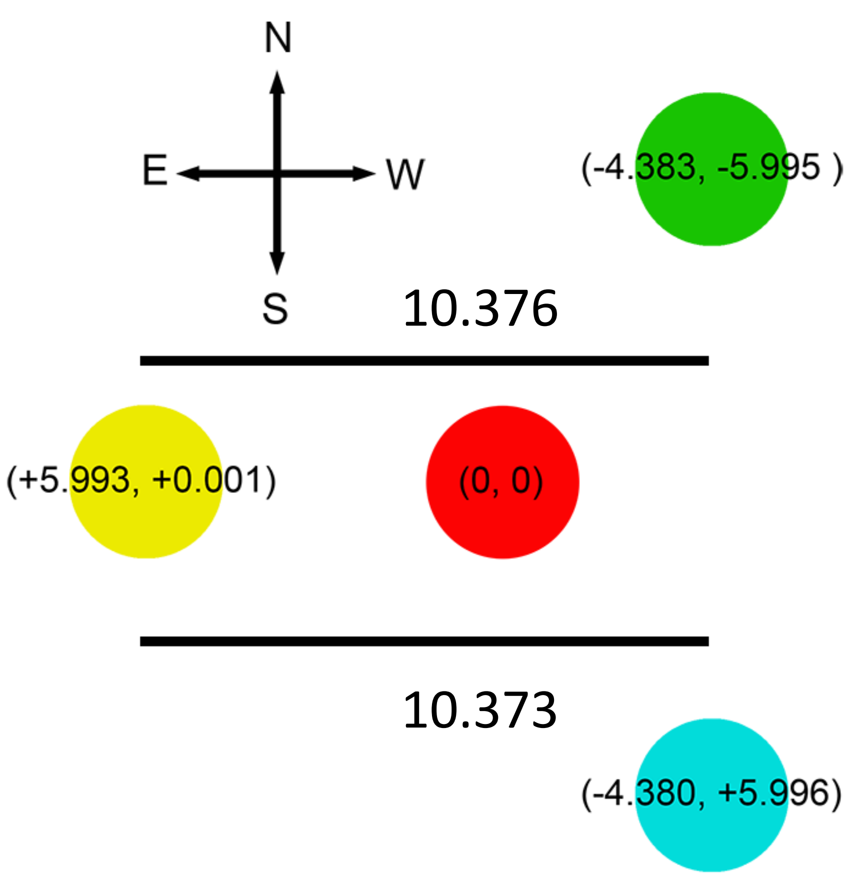
(Qizhi Huang)

- Fit fringes (real & imaginary part) on all antenna pairs (6 cross-correlations for each polarization)
- Extract array geometry (baselines), effective diameters for each antenna, the phase difference term due to electronics, cables ... $\Delta\Phi_{ij}$
- Extract also the antenna alignment (antenna axis direction)

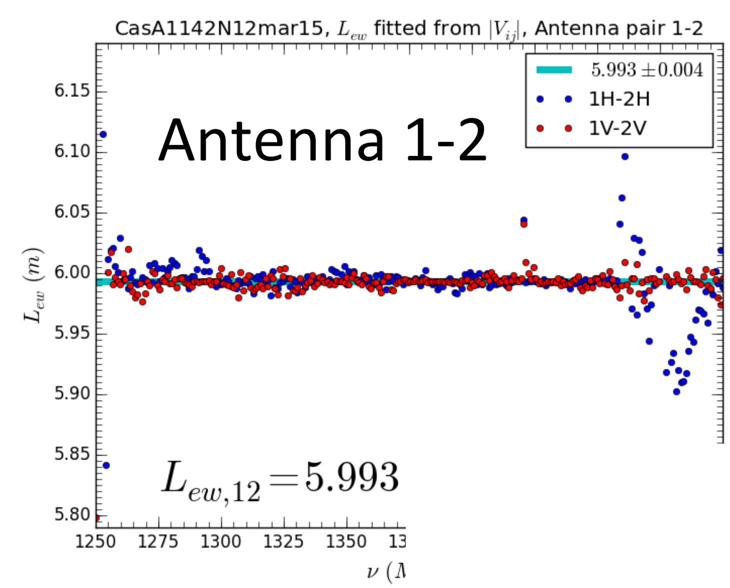
From Q. Huang presentation
https://indico.lal.in2p3.fr/event/2912/contribution/5/material/slides/0.pdf

Array geometry (baselines) determined from fringes in full agreement with on ground measurements of the antenna position (I)

Antenna geometry



East-West baseline of Antenna pair 1-2



From Q. Huang presentation

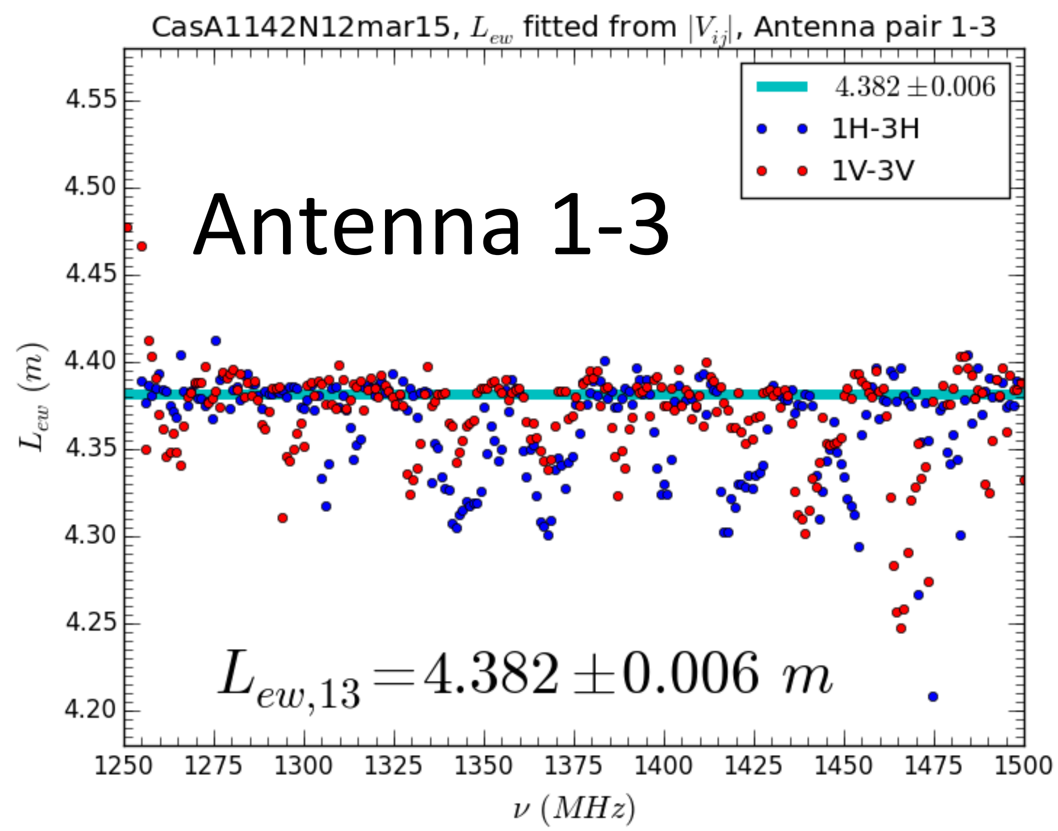
https://indico.lal.in2p3.fr/event/2912/contribution/5/mater

Data I use here

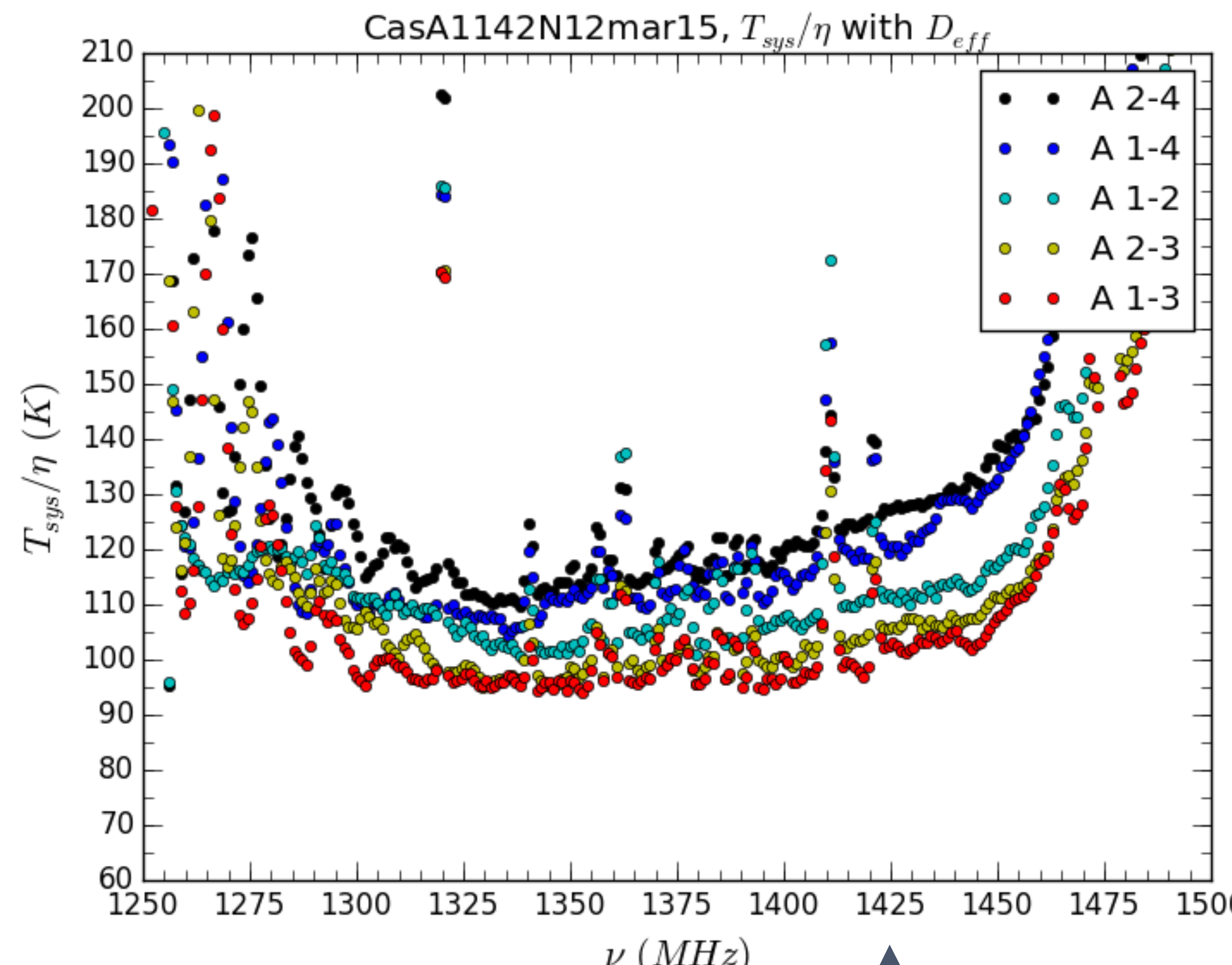
- 7 scans around Cygnus A
- Date of these scans: CygA665S1dec15, CygA365S7dec15, CygA465S5dec15, CygA565S2dec15, CygA765S30nov15, CygA865S3dec15, CygA965S6dec15
- Center declinations of each scans:

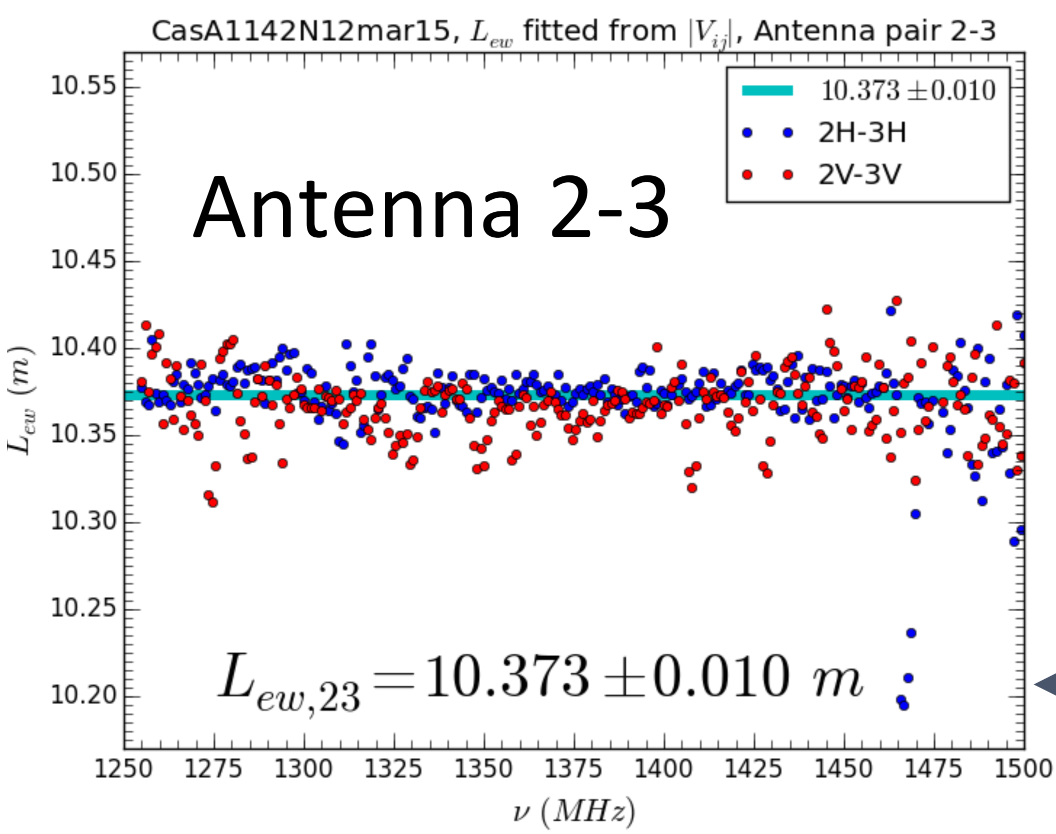
  37.7, 38.7, 39.7, 40.7, 41.7, 42.7, 43.7 degree
- Declination range of the whole data: 7 degree
- Right Ascension range: 360 degree

PAON-4 array geometry, Tsys from calibration



Analysis by Q. Huang
NAOC & LAL





Lew (ν)

Tsys (ν)

From Q. Huang presentation
https://indico.lal.in2p3.fr/event/2912/contribution/5/material/slides/0.pdf

Gain & phase calibration, Map making from PAON-4 data

(Qizhi Huang)

Data I use here

- 7 scans around Cygnus A
- Date of these scans: CygA665S1dec15, CygA365S7dec15, CygA465S5dec15, CygA565S2dec15, CygA765S30nov15, CygA865S3dec15, CygA965S6dec15
- Center declinations of each scans:

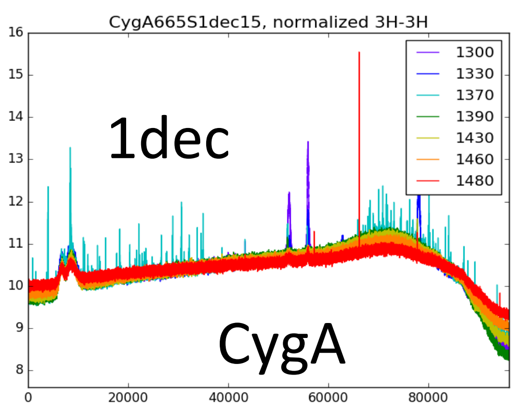
  37.7, 38.7, 39.7, 40.7, 41.7, 42.7, 43.7 degree
- Declination range of the whole data: 7 degree
- Right Ascension range: 360 degree

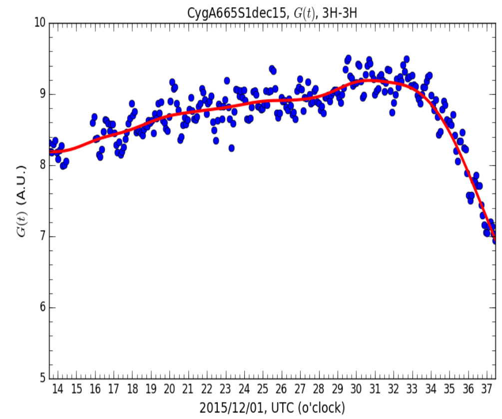
November-December 2015 CygA data

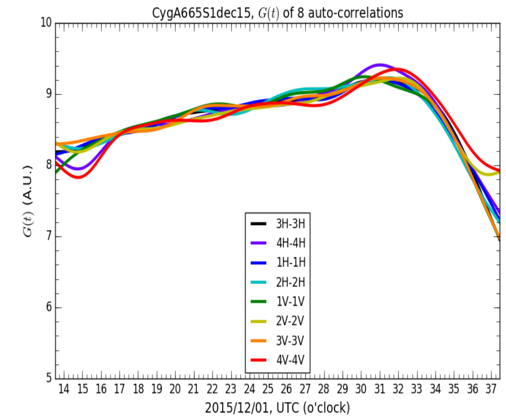
Analysis & slide by Q. Huang
NAOC & LAL

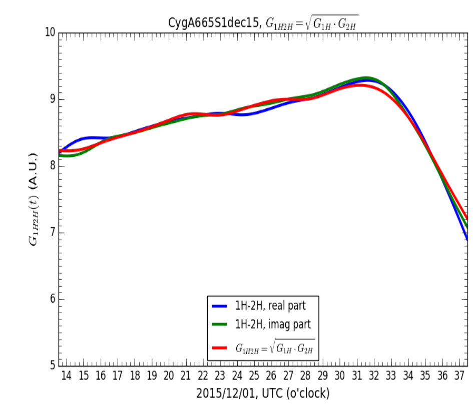
Calibrate gain 2: $G(t)$

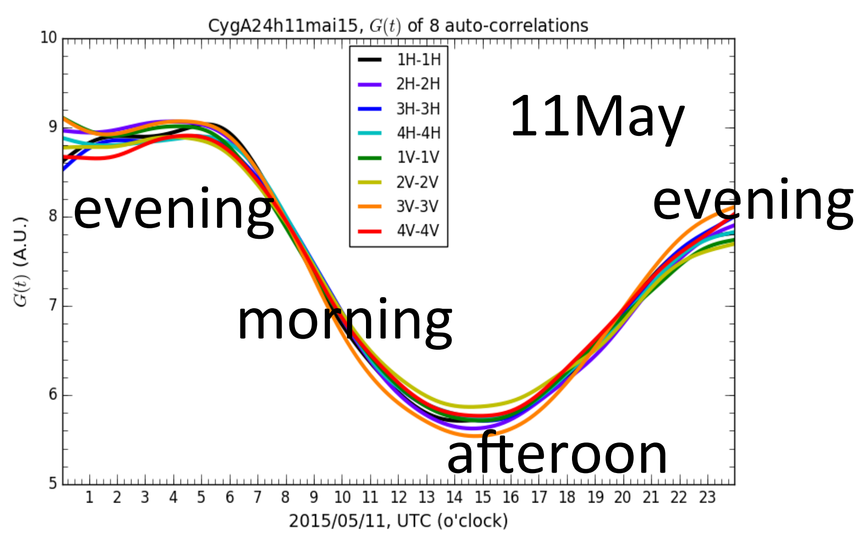
- Same thing but different data set (different day) comparing with the page above.

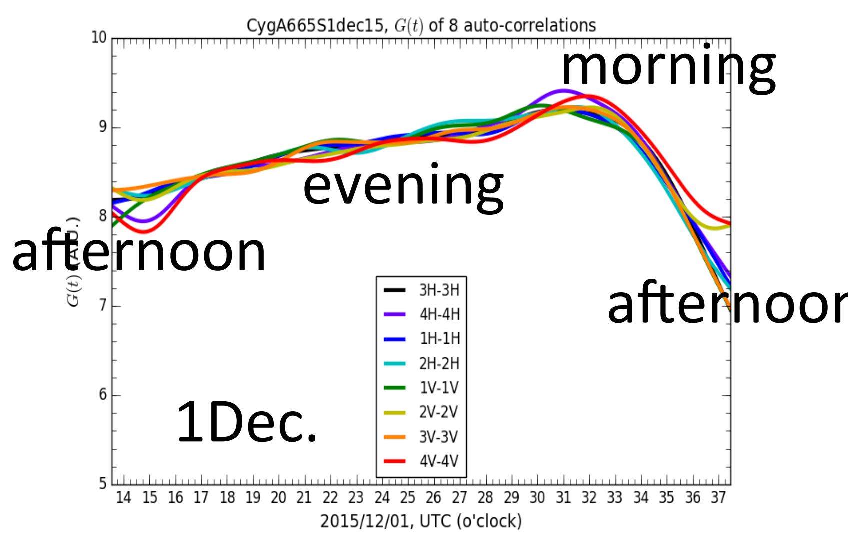












11 May is in mid-late spring, the temperature changes a little large from daytime to night.
While 1 Dec. is in winter, the temperature changes less, then the gain is smoother than that in May.

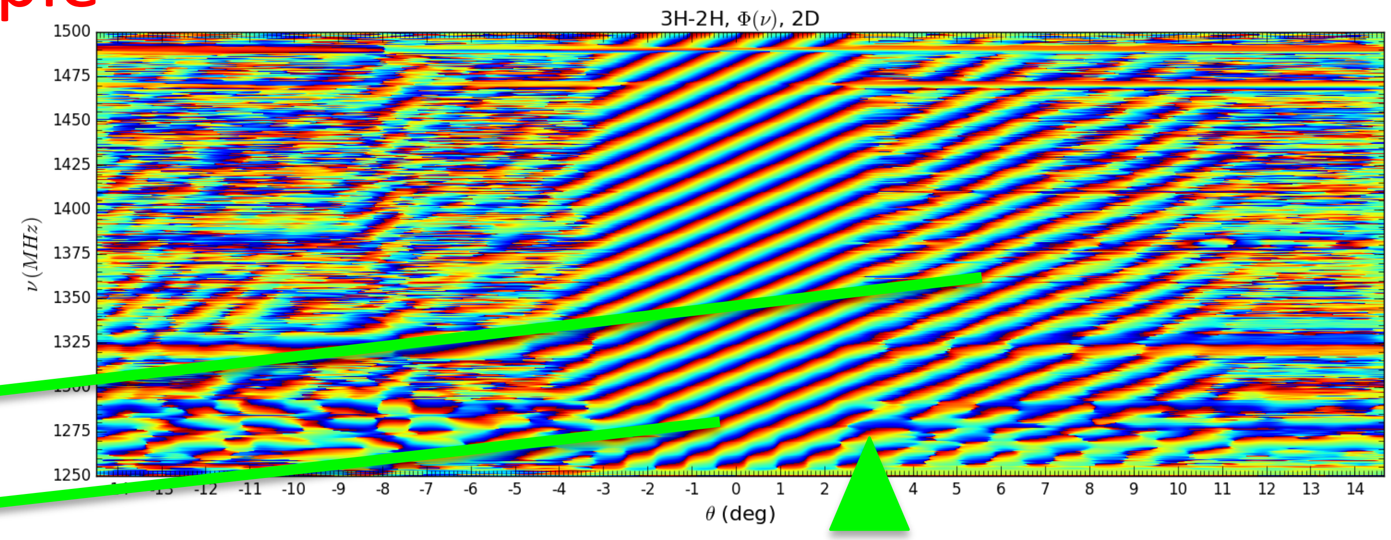
We can see that, the gain decreases from morning to afternoon, and increases in the evening. Corresponding to the temperature. In higher temperature, gain is lower.

Analysis & slide by Q. Huang
NAOC & LAL

Phase calibration

Here I show 2H3H for example

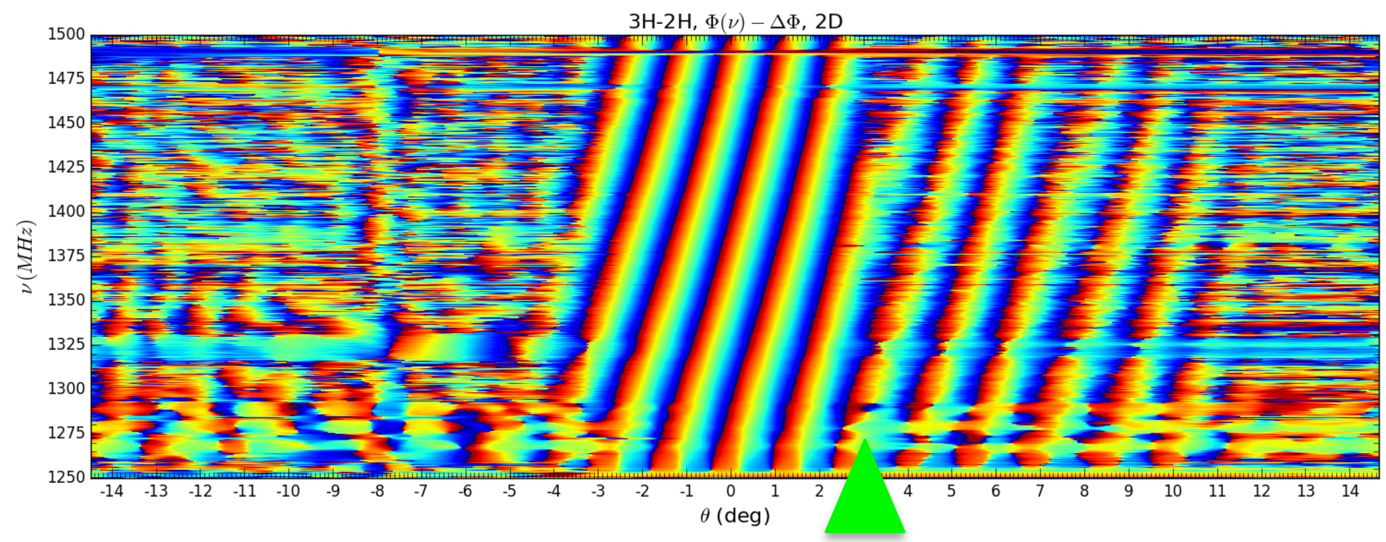
Total phase from
the observation data



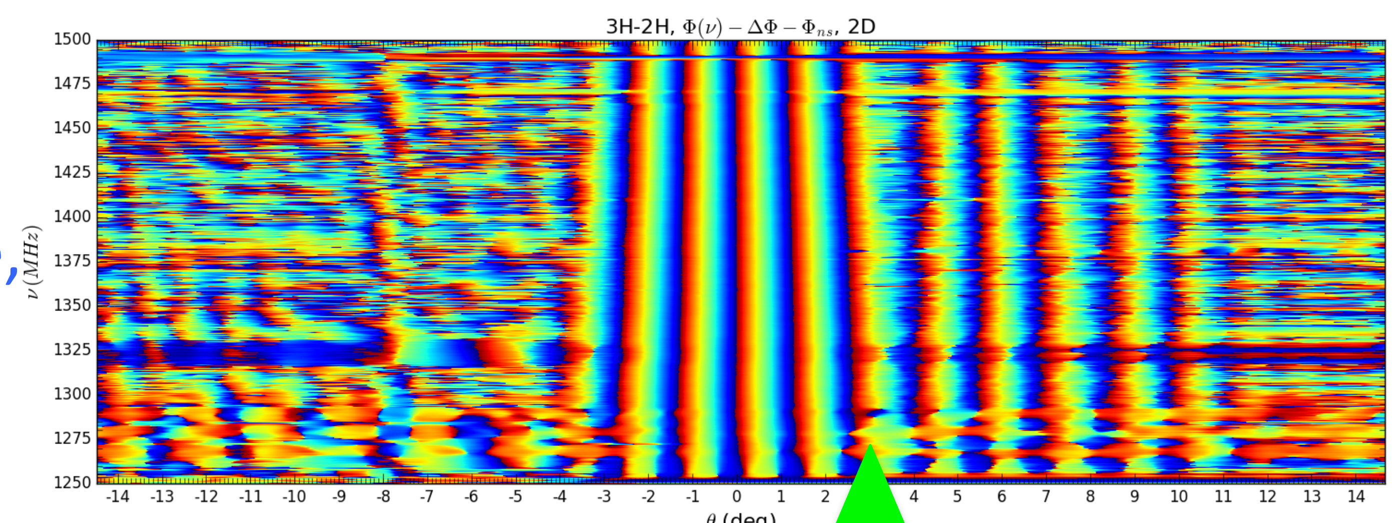
Cygnus X

Cygnus A

After calibrating
the additional phase



After calibrating
the additional phase and
phase from North-South baseline,
just leave the phase from the
East-West baseline



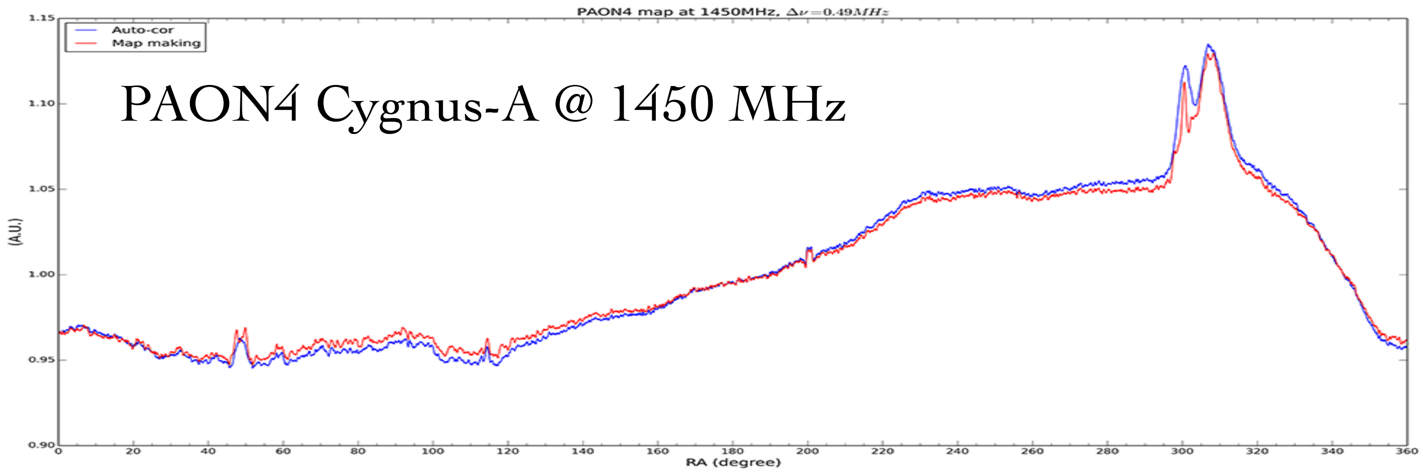
Map from visibilities (single scan) I

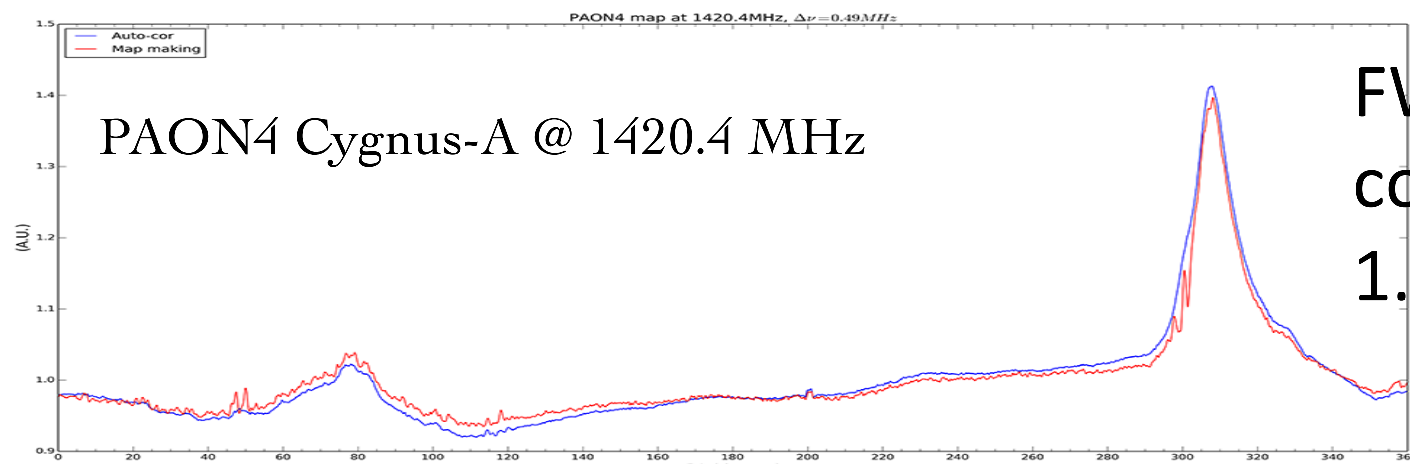
Slide by Qizhi Huang (LAL/NAOC PhD)

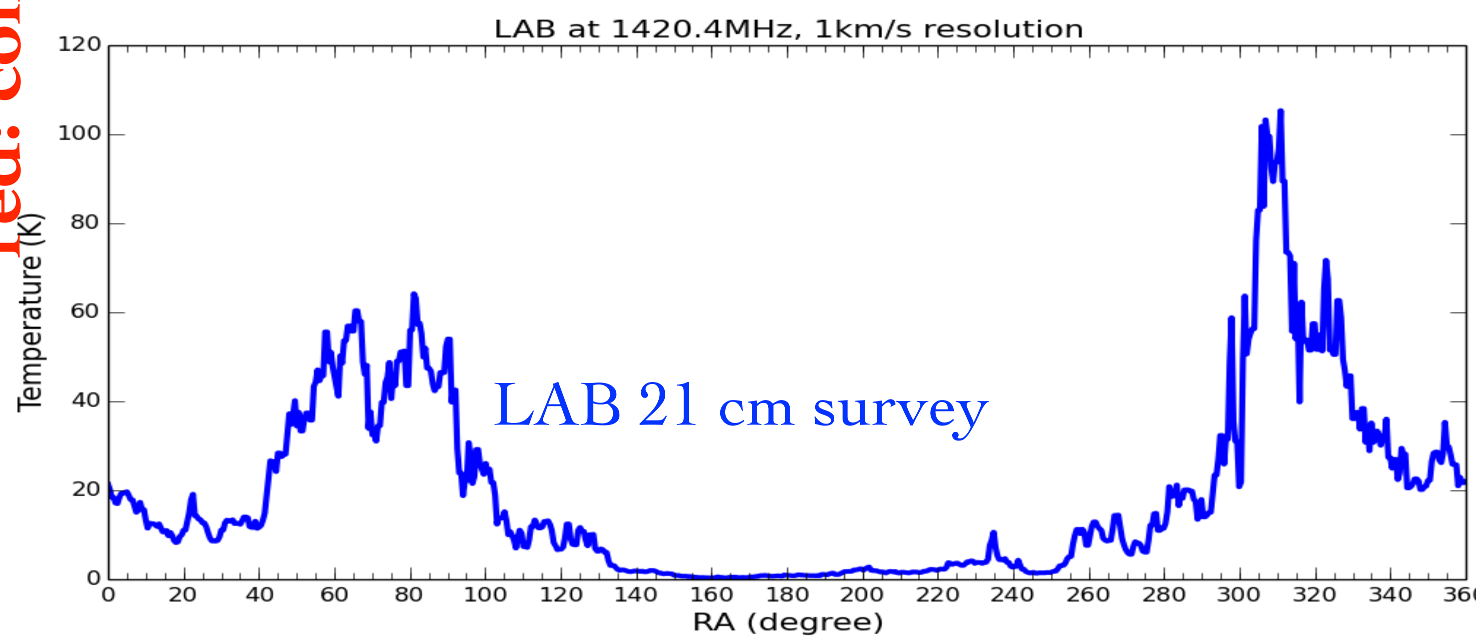
PAON4 CygA

blue : autocorrelation (single dish)

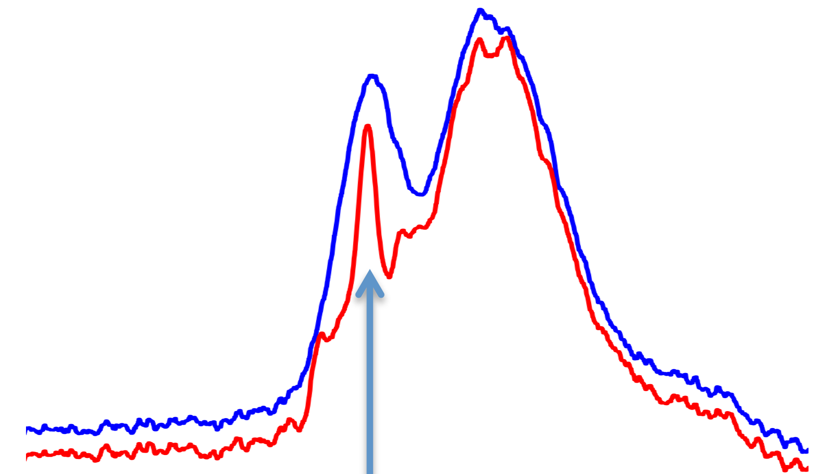
red: combined visibilities (4 dishes)



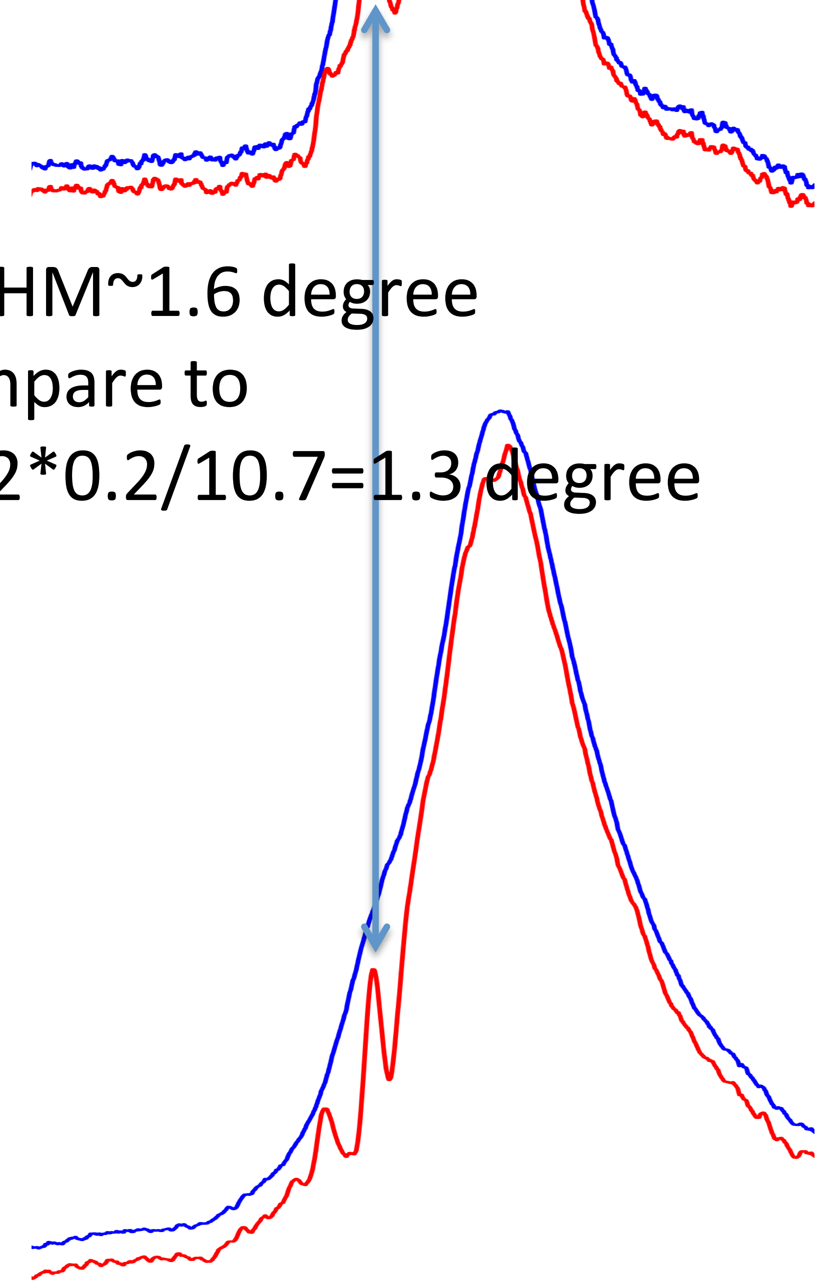




PAON4 CygA @ 1450 MHz



FWHM~1.6 degree
compare to
1.22*0.2/10.7=1.3 degree



PAON4 CygA @ 1420.4 MHz

Nov. 2015 - Data from May 2015

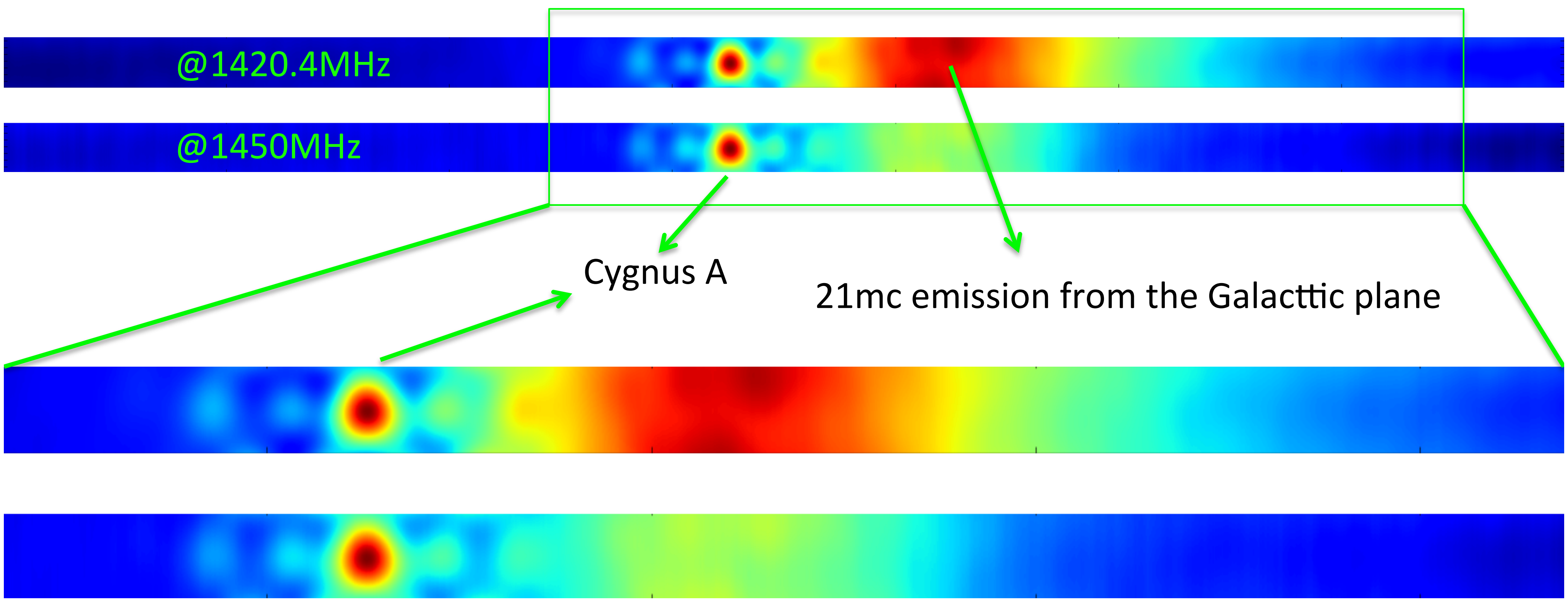
Map from visibilities (single scan) II

Slide by Qizhi Huang (LAL/NAOC PhD)

Nov. 2015 - Data from May 2015

Final maps

CygA24h11mai15 is one day 24 hours observation, we just have one transition. Hower, we can add and special phase to the visibilities to simulate the case that turnning the antennas to other declinations and observate.



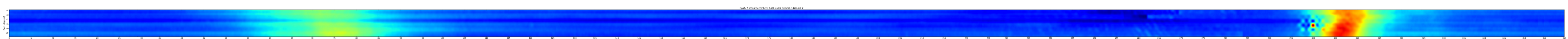
Map from 7 PAON-4 scans of CygA at 1420 MHz (I)

7 x 360 deg^2 map

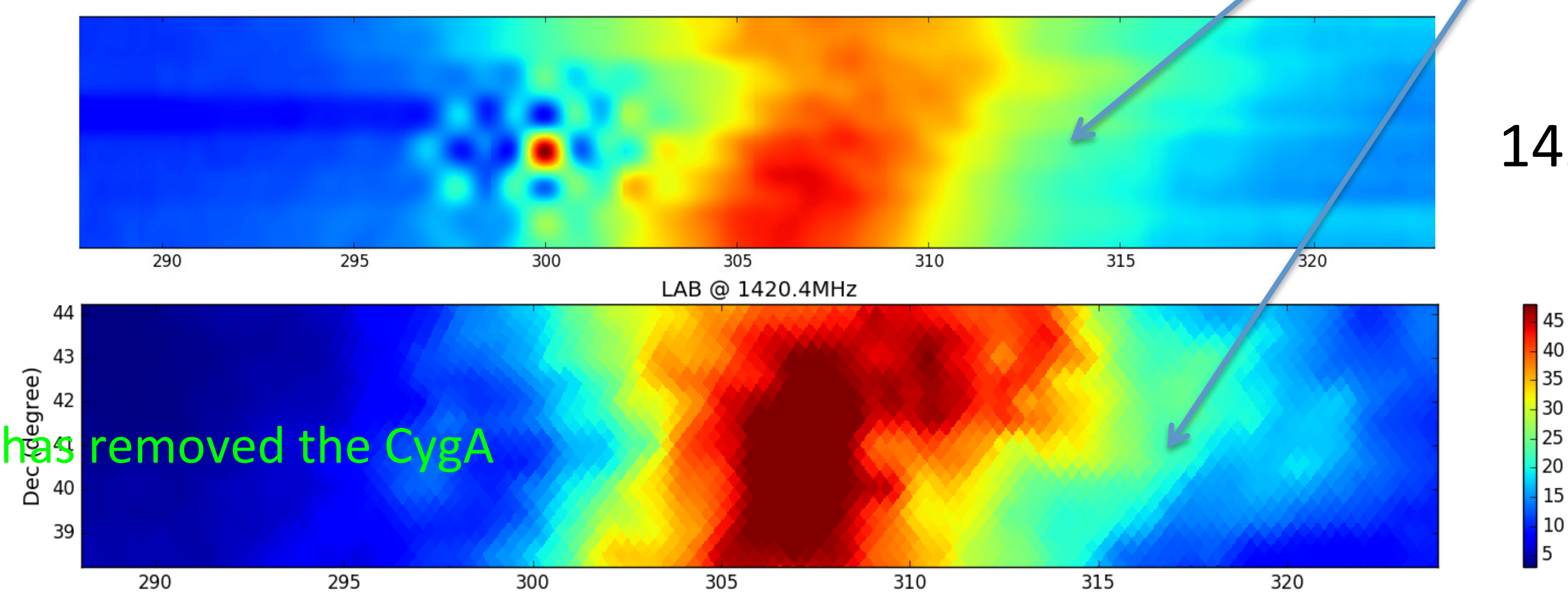
Analysis & slide by Q. Huang
NAOC & LAL

I use 7 scans to make this map.
For each scan, I get 200 mock lines by add a phase exp(i*2pi/l*Lns*sin(delta)), delta=1 degree for 200 pixel.

7dx360d, 1420.4MHz (you also find the large image in .jpe)



The same sky region of LAB data



1420.4MHz

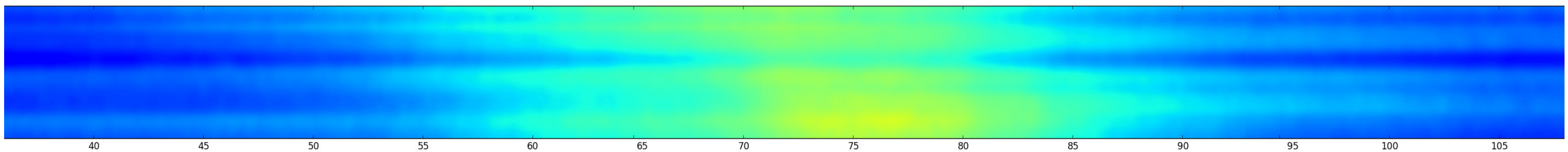
LAB has removed the CygA

Map from 7 PAON-4 scans of CygA at 1420 MHz (II)

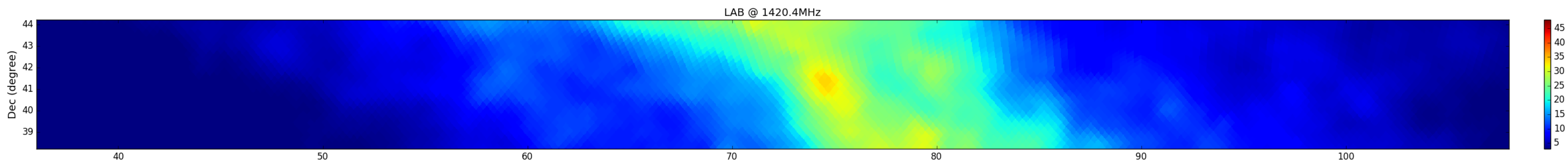
7 x 40 deg^2 map

Analysis & slide by Q. Huang
NAOC & LAL

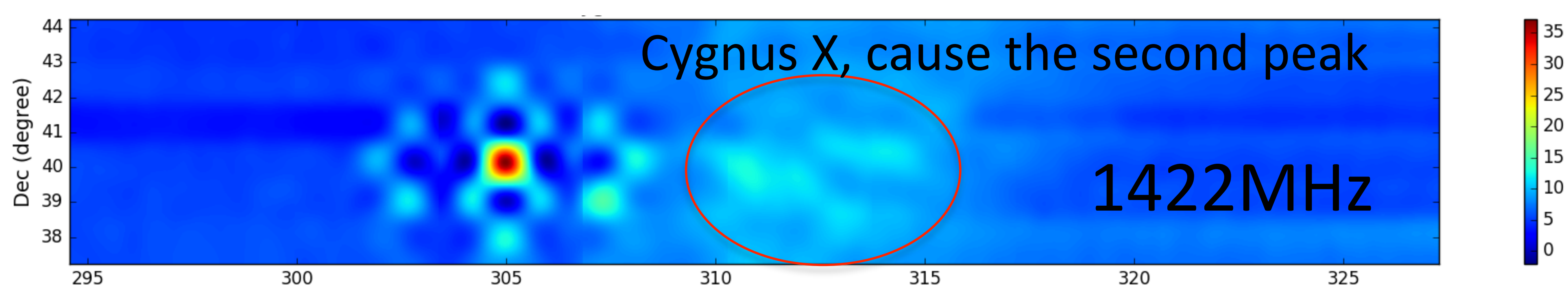
PAON4, galactic anticenter, 1420.4MHz



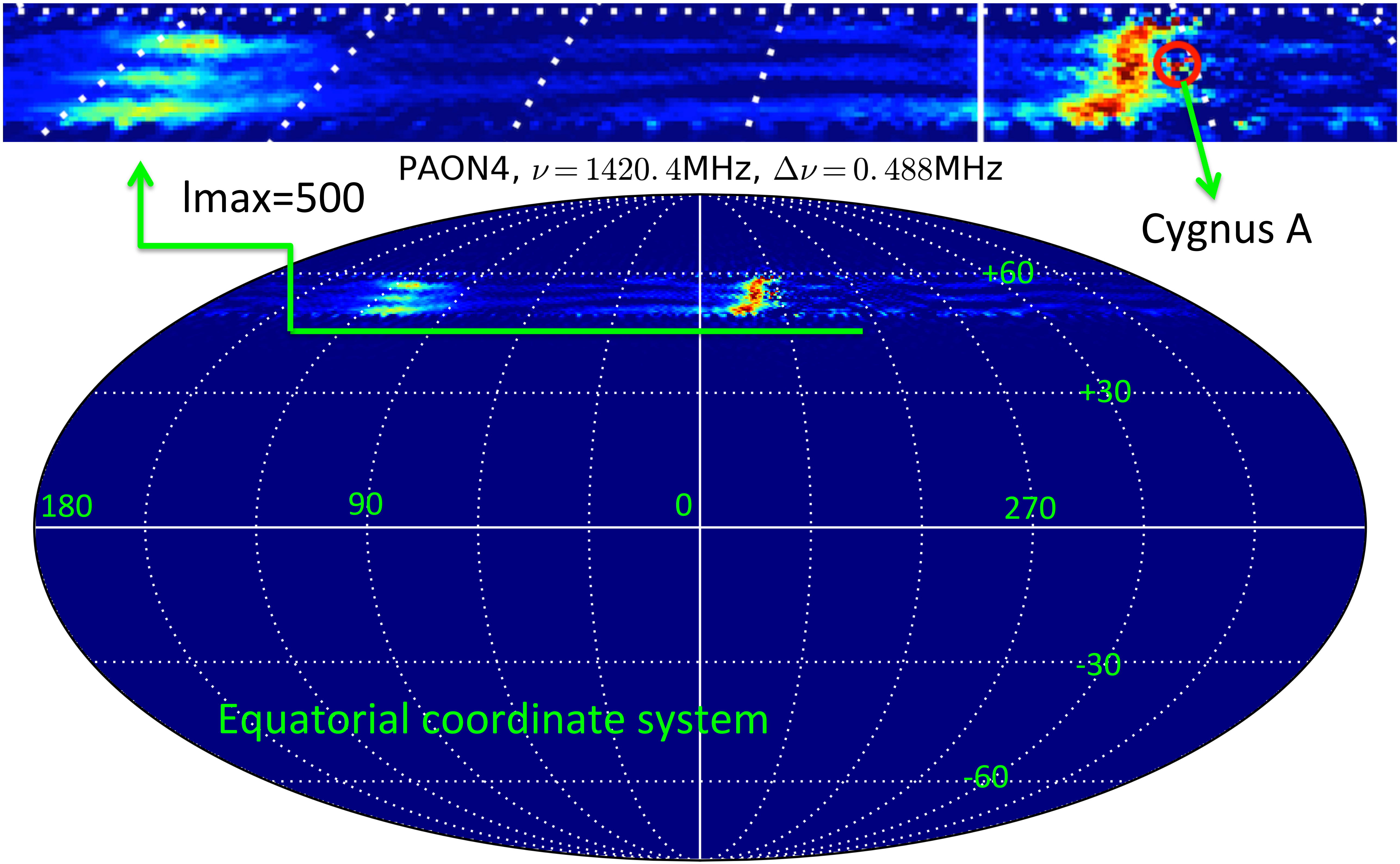
LAB, galactic anticenter, 1420.4MHz



When we observe the CygA, we will usually see two peak, one is the CygA, the other is the so called Cygnus X.
I make the map at 1422MHz where HI emission is weak, we can see the structure of the Cygnus X.



Map making using Jiao's program (2)

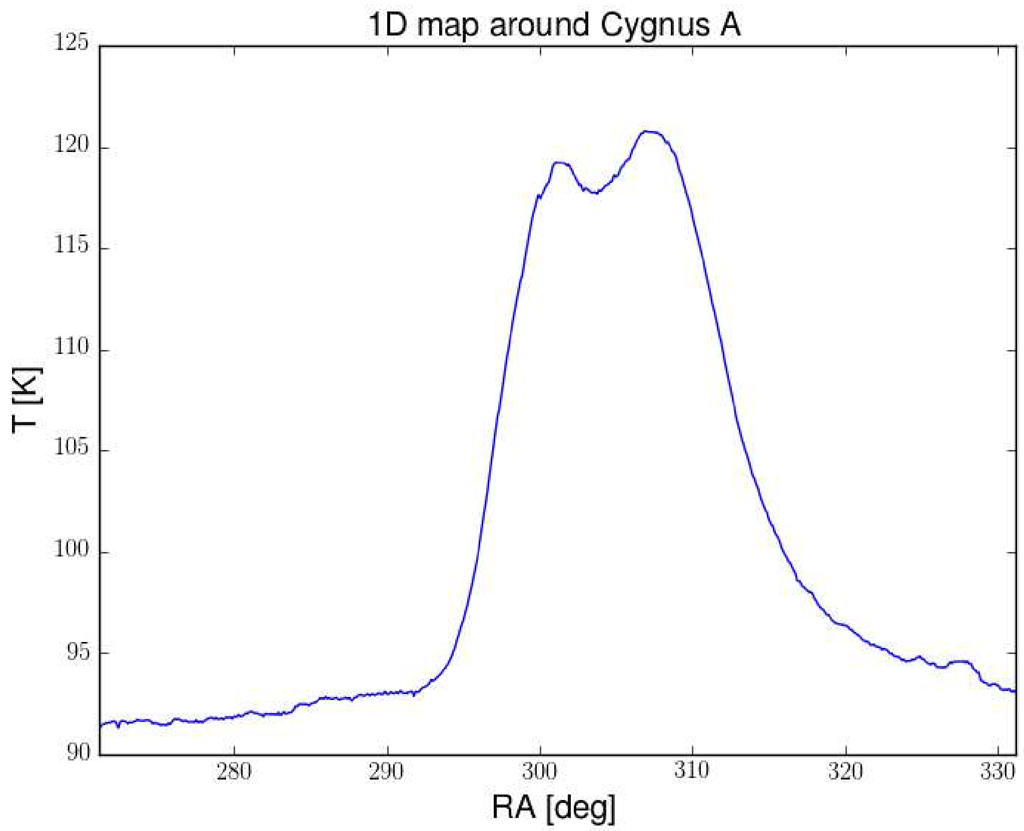


Analysis by Q. Huang & J. Zhang (NAOC & LAL)

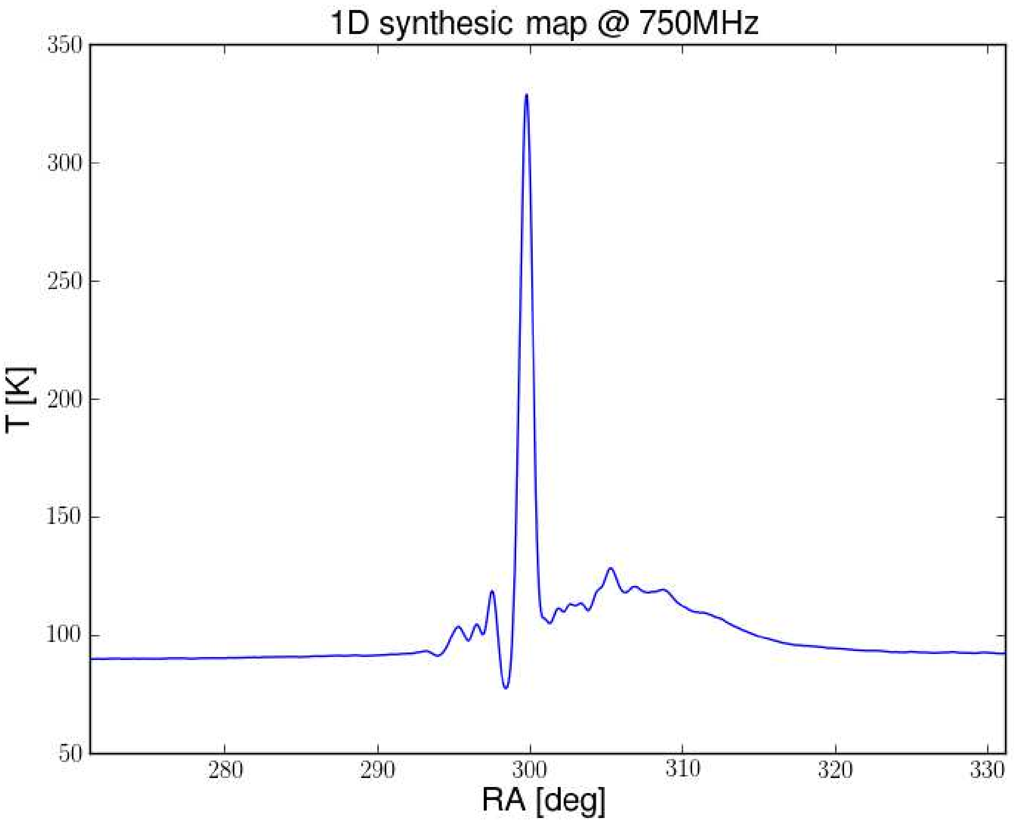
Calibration & map with Tianlai Dish array

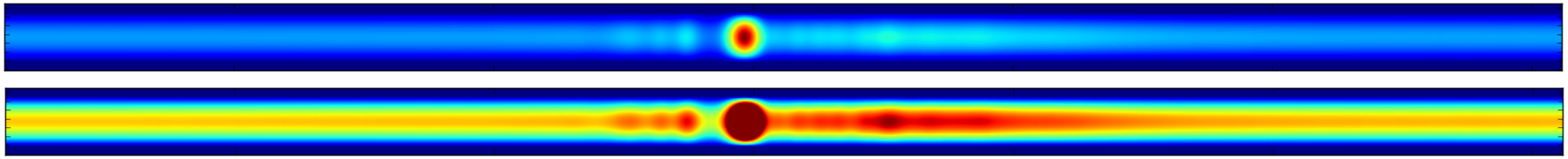
- Q. Huang applied similar methods to Tianlai 16 dish array
- Compatible phase calibration from ky (sources) and noise source

Cygnus A map (1D - single dish)



Cygnus A map (1D - 16-dish)





Calibration

- Initial calibration: use bright source transit to determine array geometry, beam, gain & phase [Q. Huang]
- Antenna beam determination using a dedicated measurements (drone ?)
- Array gain & phase calibration/monitoring using the noise source
- Gain & phase calibration and monitoring using the known sky (Radio sources & Synchrotron)
- 2 N unknown (or N complex) - N (N+1)/2 complex constraints x number of time samples - But NOT (?) a linear fit …
- Should be computed for each frequency, and freq. behaviour should be checked against a model.

Technical considerations, Simulations

Level 0 (**L0**) output data (on site)

Tianlai 16-Dish Array (T-16DA)

- 2 x 16 = 32 receivers, 1000 freq. channels, 528 visibilities
- ~ 4MB/sec visibility data (averaged @ 1 sec. time interval)
- 350 GB/day ➟ 1000-1500 TB L0 data / year
- ~ 288 x 5 min time slice files / day (1.2 GB/file) - or 24 x 1 h time slice files (15 GB/file)

Tianlai 3-Cylinder Array (T-3Cyl)

- 2 x 3 x 92 = 192 receivers, 1000 freq. channels, 18 528 visibilities
- ~ 140 MB/sec visibility data (averaged @ 1 sec. time interval)
- 12 TB/day , ➟ 35 PB L0 data / year
- ~ 10x288 x 5 min time slice files / day (4.2 GB/file) - or 10x24 x 1 h time slice files (50 GB/file)

L1: NAOC (Beijing)

- L0 output : Visibility data $V_{ij}(\nu)$ computed on-line (in HW)
- Organize data sets as time sliced files , grouped with auxiliary (housekeeping) data
- Perform a first step, simple RFI mitigation
- Data compression, mainly through time averaging after RFI cleaning (factor 5-10)
- L1 output data :
  - T-16DA : 35-50 GB / day, ~100 files / day , ~100 TB / year
  - T-3Cyl : ~1000 GB / day, ~1000 files / day , 2-5 PB / year
- Transfer raw data (L1 output) to TAC

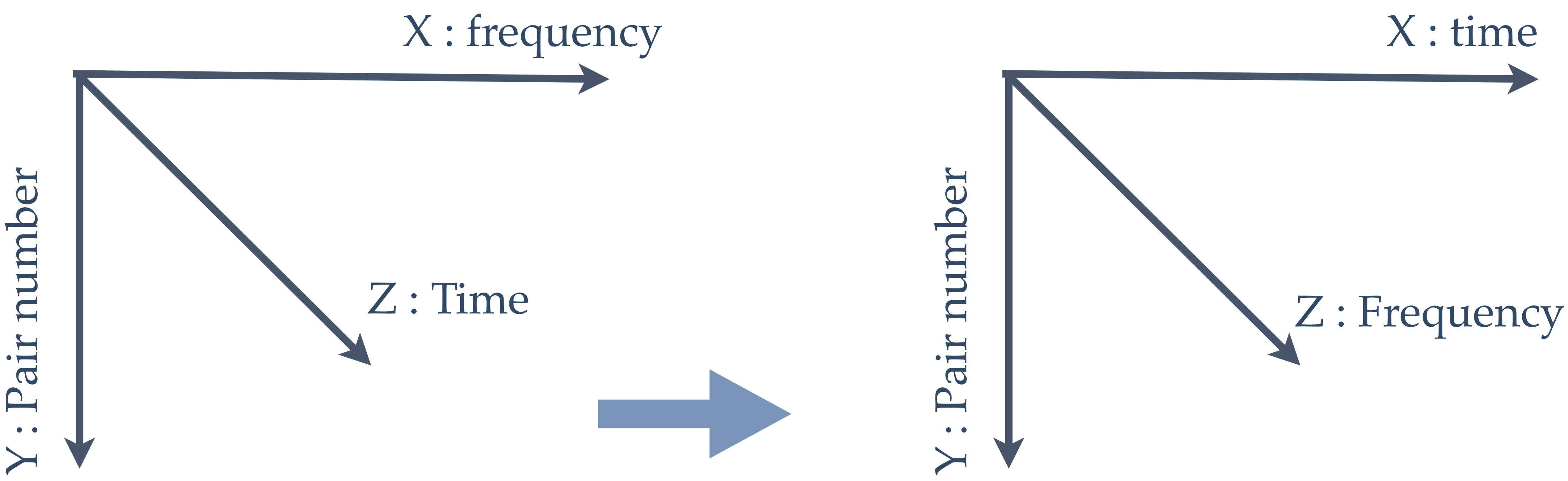
TAC : Tianlai **A**rchive and **A**nalysis **C**enter
Fermilab (Batavia, IL)

- **L2-A**: second stage RFI cleaning, gain/noise monitoring
- **L2-B**: phase/gain calibration
- **L2-C**: 3D map making
- **L3 :** Component separation, power spectrum estimation
- **TAC**:
  - data organization and data access services
  - computation ressources for L2 &L3 (~ few x 10 CPU-cores / MB/sec L1 data rate)

File format, file organisation

- HDF5 for visibilty data
- Maps (HEALPix …) in FITS or HDF5 ?
- Source catalog in FITS or HDF5 ?
- We need to setup a ‘Data Base’ to track the data files (visibilty) and other instrument related files

Visibility array structure



Vij(ν, t)

Vij(t, ν)

Might be broken in part according to frequency or pair number

Should we rearrange the visibility data at some stage of the processing ?

Simulations

- Full sky signal simulation tools , or 3D maps
- Foreground (synchrotron) model / maps
- Comprehensive known Radio Source catalog
- Simulate Visibility time streams (which is part of the map making software)
- Include calibration error models (beam, geometry)
- Noise & RFI model
- reconstruct maps, apply component separation …