2019/10/18 PhD defense

Topics in 21-cm cosmology:

# Foreground models and their subtraction, map reconstruction for wide field of view interferometers and PAON-4 data analysis

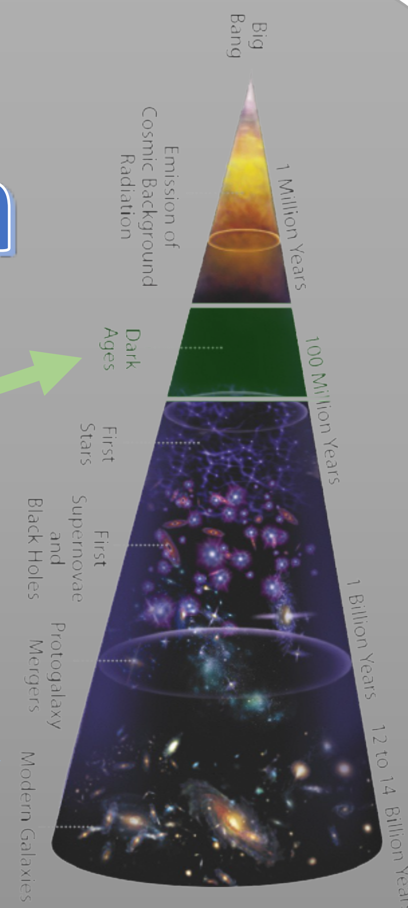

Qizhi HUANG

Supervisors: Prof. Réza ANSARI Prof. Xuelei CHEN

# Outlines

- Introduction
- High resolution foreground model
- 21cm extraction by filtering
- Lunar orbit interferometer imaging
- PAON-4 data analysis

21 cm




- Summary

- Introduction
    - Dark Energy
    - Baryon Acoustic Oscillation (BAO)
    - 21cm Intensity Mapping
    - 21cm line
    - Dark Ages
- High resolution foreground model
- 21cm extraction by filtering
- Lunar orbit interferometer imaging
- PAON-4 data analysis
- Summary

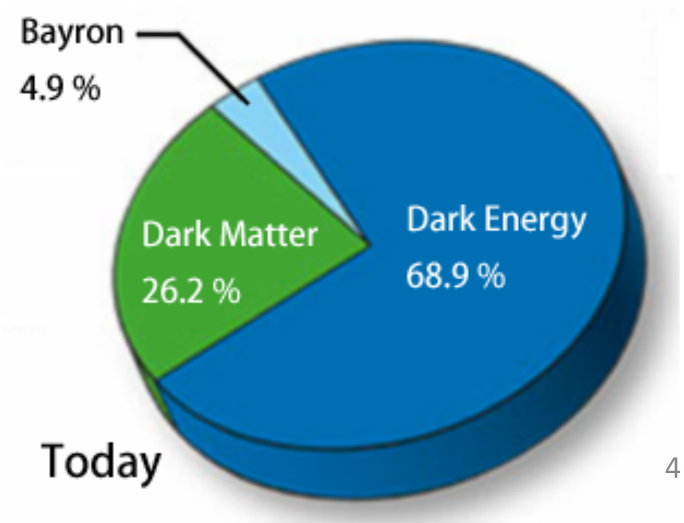
# Dark Energy

- Accelerating expansion of the Universe
  - Type Ia supernovae (Nobel Prize in Physics 2011)
  - Driven by dark energy (~70%)
- Dark energy fluid
  - Equation of state <= H(z) or $D_A$(z) <= BAO

$$H(z) = H_0\sqrt{\Omega_m(1+z)^3+\Omega_r(1+z)^4+\Omega_k(1+z)^2+\Omega_{DE}e^{3\int_0^z\frac{1+w(z')}{1+z'}}}$$

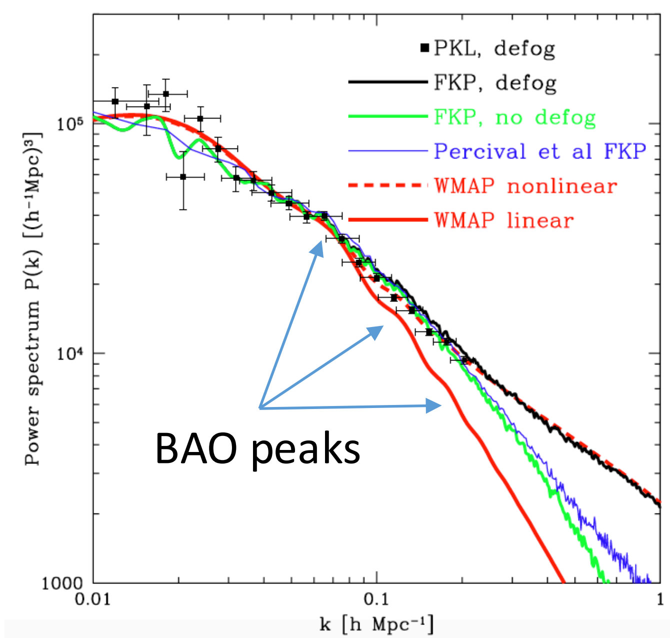
| Cosmological Parameter | TT,TE,EE+lowE+lensing+BAO 68% limits |
|---|---|
| $H_0$ [km s$^{-1}$ Mpc$^{-1}$] . . | $67.66 \pm 0.42$ |
| $\Omega_\Lambda$ . . . . . . . . . . . . | $0.6889 \pm 0.0056$ |
| $\Omega_m$ . . . . . . . . . . . . | $0.3111 \pm 0.0056$ |
| $\Omega_b h^2$ . . . . . . . . . . | $0.02242 \pm 0.00014$ |
| Age [Gyr] . . . . . . . | $13.787 \pm 0.020$ |

(Planck Collaboration et al. 2019)



# Baryon Acoustic Oscillation (BAO)

- Photon-baryon flow oscillation, primordial fluctuation:
  - higher density to lower density
- Standard ruler, 105 Mpc/h
- Obtain $D_A(z)$ or $H(z)$ => universe evolution => dark energy

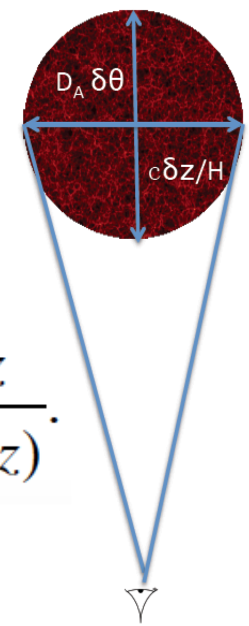



$$r_\perp = (1+z) D_A(z) \Delta\theta,$$

$$r_\parallel = \frac{c\Delta z}{H(z)}.$$

$$D_A(z) = \frac{c}{1+z}\int_0^z \frac{dz}{H(z)}.$$

Angular diameter distance



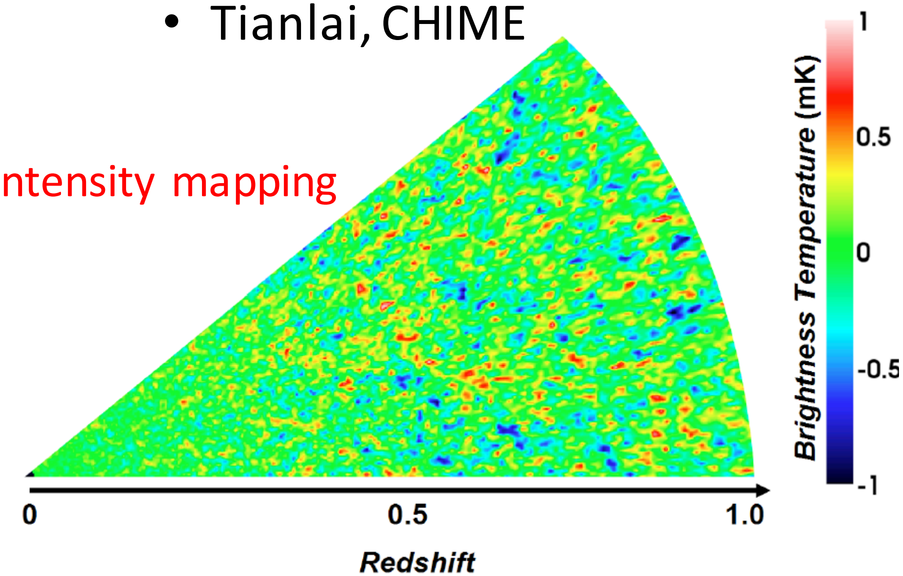

**Standard ruler type probe**

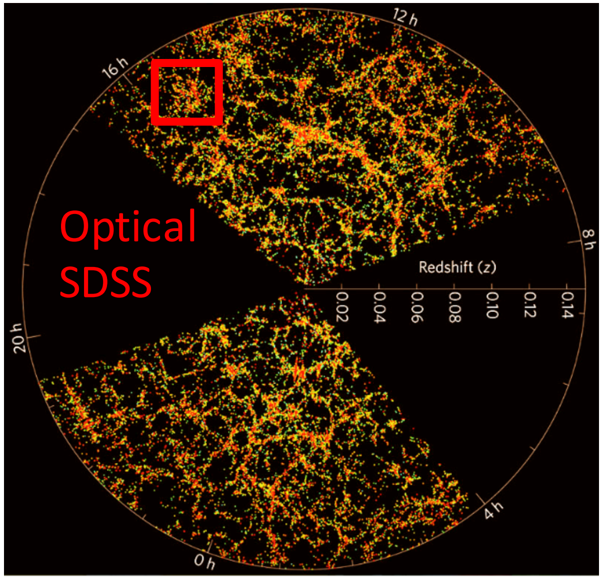
**Measure $d_A(z)$**

# 21cm intensity mapping

- Optical:
  - SDSS, Boss/eBoss, DESI
  - Individual galaxy, long time, only low redshift
- Radio using 21cm:
  - Measure individual galaxy, long time, huge collecting area
  - 21cm intensity mapping:
    - Much faster! Wide FoV; broadband
    - Tianlai, CHIME

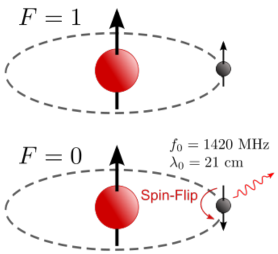
21cm intensity mapping

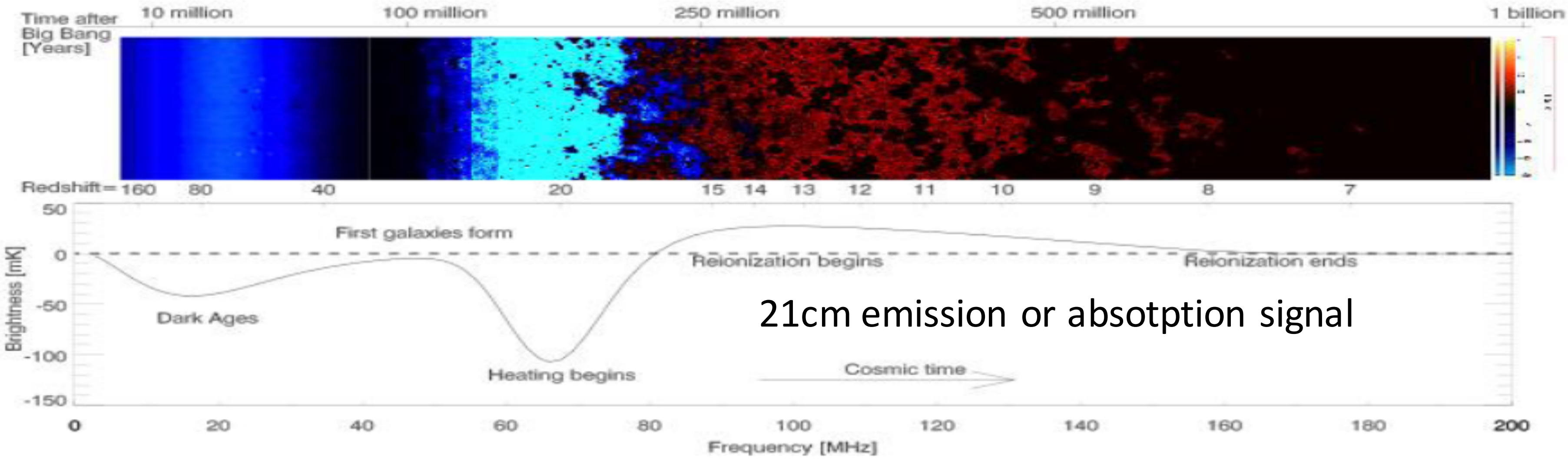


# 21cm signal

- Transition between two hyperfine levels of HI atom
  - proton and electron's spin directions
  - Energy difference => 21cm wavelength



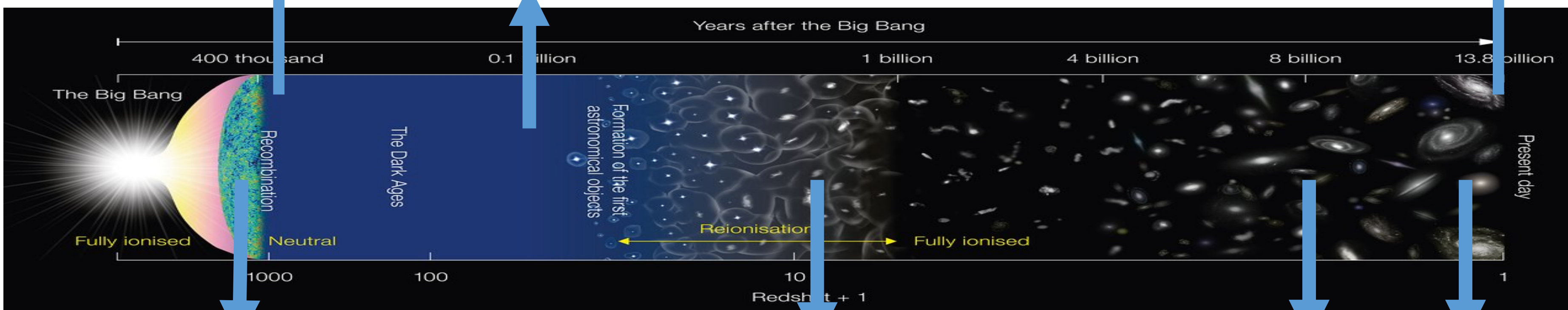




21cm emission or absotption signal

# Experiments

21cm

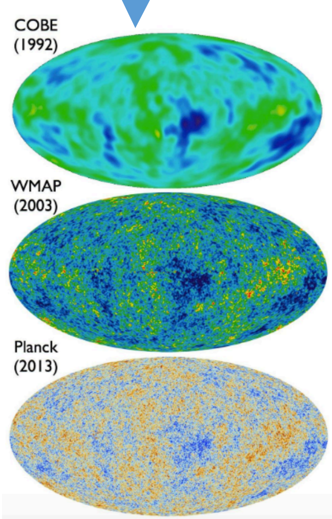
Lunar orbit interferometer DSL



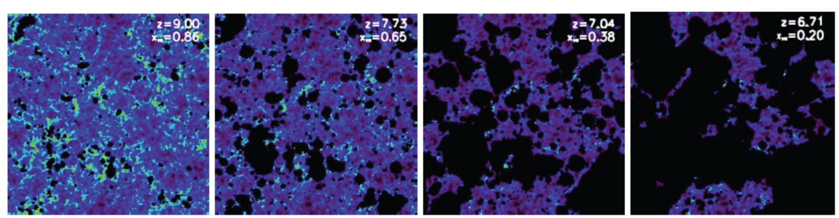

COBE

WMAP

Planck




LOFAR, MWA, etc.



|143 Mpc |

(A. Mesinger et al. 2011)

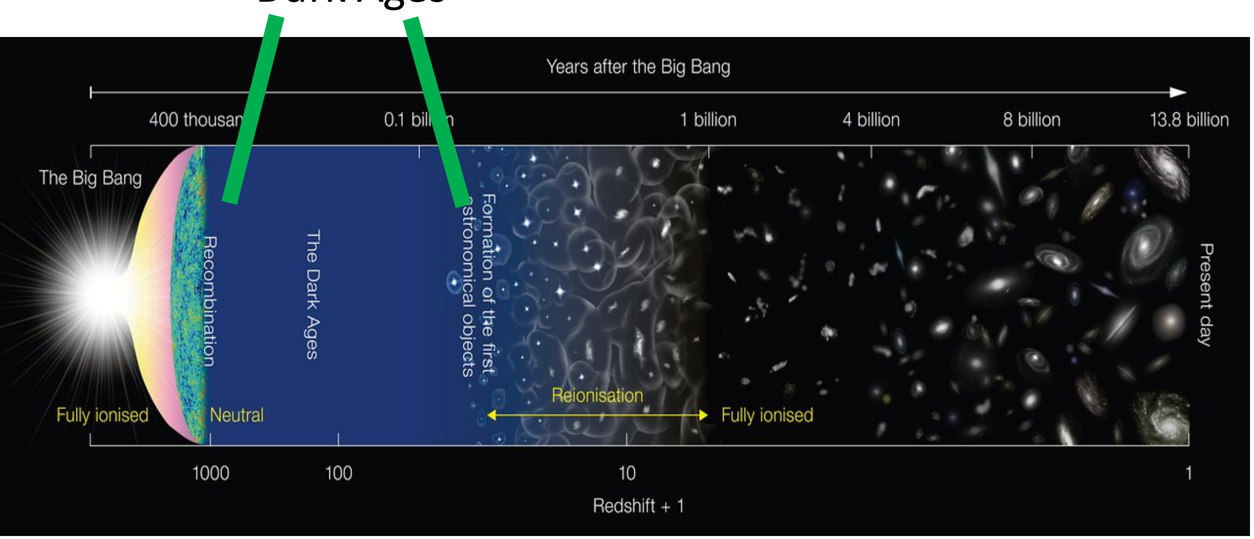
Tianlai, PAON-4

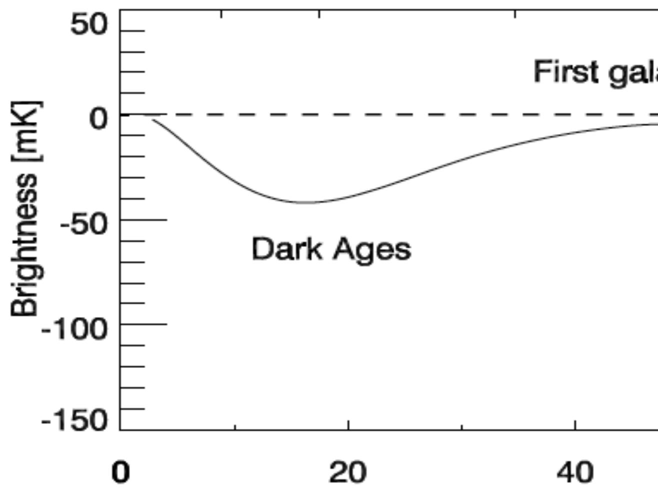
CHIME

# Dark Ages

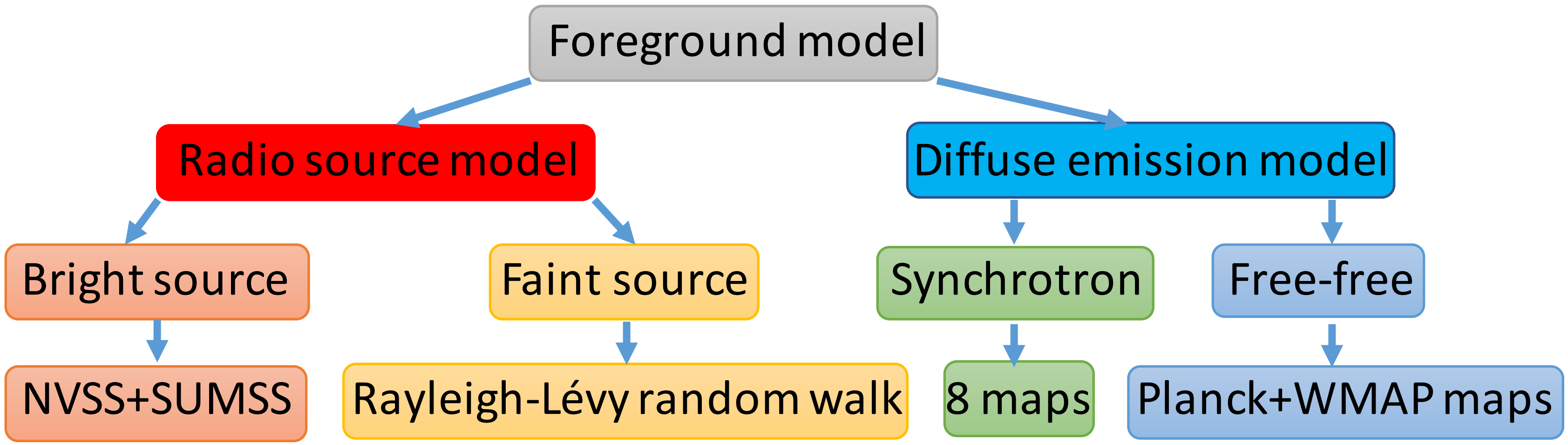
- Highly homogeneous and isotropic
  - most is HI, no luminous star
- Use 21cm line to detect: $T_S < T_{CMB}$: absorption
- High redshift, low frequency, ionosphere => lunar satellites

Dark Ages



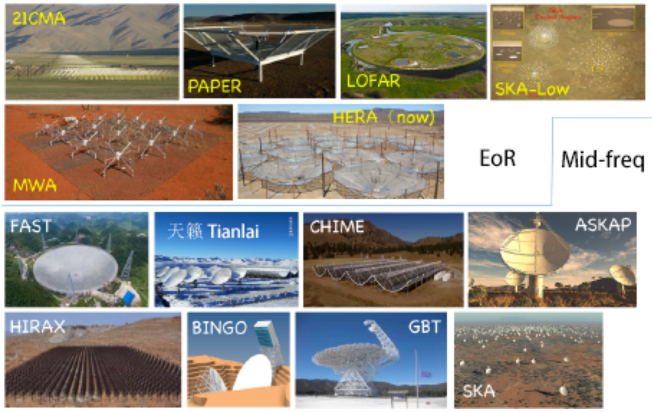
- Introduction
- **High resolution foreground model**




- 21cm extraction by filtering
- Lunar orbit interferometer imaging
- PAON-4 data analysis
- Summary

# 21cm experiments and foreground models

- 21cm experiments
  - EoR: LOFAR, MWA, PAPER, 21CMA, HERA
  - mid-frequency, large scale structure, dark energy: PAON-4, Tianlai, CHIME, BINGO



Some 21cm experiments

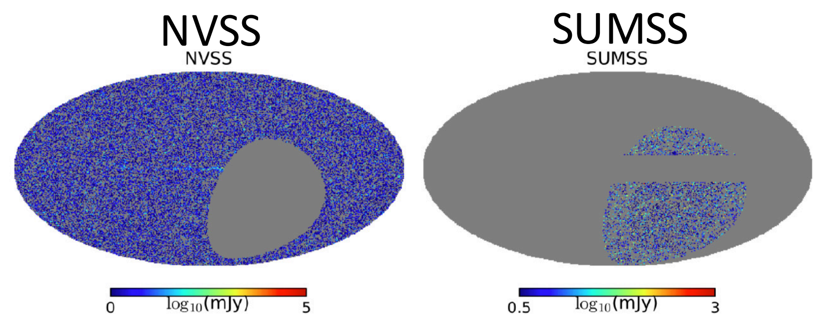
- Foreground models
  - High resolution foreground model is needed
  - GSM 2008 and 2016: low resol, $1^{\circ}$
  - T-RECS: small sky coverage
  - CORA: spectral index---GSM; random sources
  - Our model: spectral index---from observational maps;
    source distribution---Rayleigh-Lévy random walk

# Used data

Foreground model

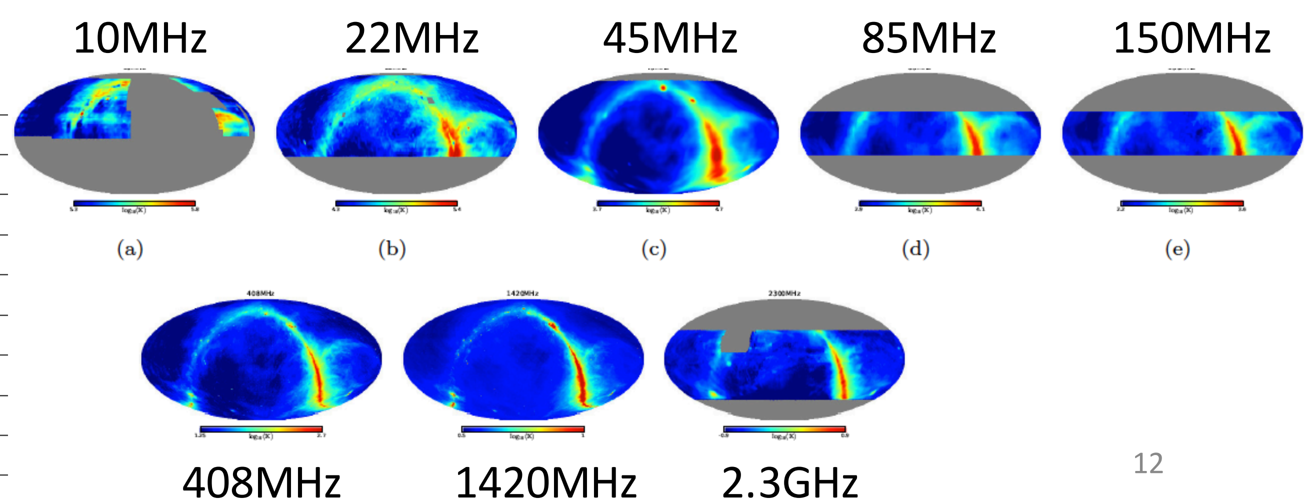
Radio source model | Diffuse emission model

- Radio source data
  - NVSS (1.4 GHz) + SUMSS (843 MHz)
  - Similar angular resolution: ~45''
  - Similar sensitivity: ~mJy
  - Combination is full sky



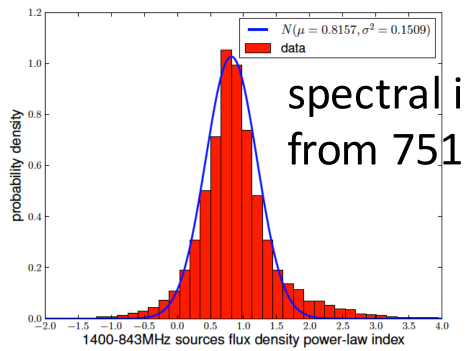

- Diffuse emission data

| Frequency | FWHM |
|---|---|
| 10 MHz | $2.6^\circ \times 1.9^\circ$ |
| 22 MHz | $1.1^\circ \times 1.7^\circ$ |
| 45 MHz | $5^\circ$ |
| 85 MHz | $3.8^\circ \times 3.5^\circ$ |
| 150 MHz | $2.2^\circ$ |
| 408 MHz | $0.85^\circ$ |
| 1420 MHz | $0.6^\circ$ |
| 2300 MHz | $2.3^\circ \times 1.9^\circ$ |



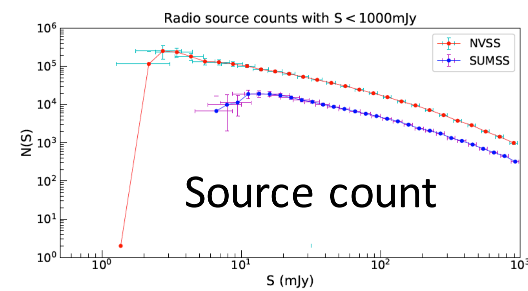
# Bright source model

- Spectral index
  - 10 degree overlap (-40° to -30°)
  - 7515 sources
  - Spectral index: 0.8157 ± 0.3




spectral indice
from 7515 sources

- Complete sample
  - Source count
  - NVSS (1.4GHz): ≥ 2.7 mJy => 15 mJy
  - SUMSS (843MHz): ≥ 12 mJy => 22 mJy

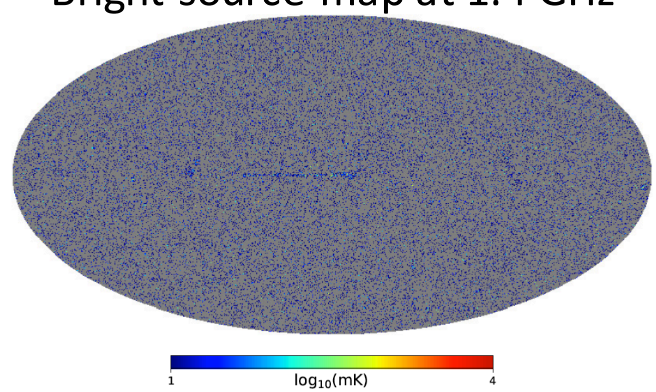



Source count

Fig. 3.5 Source count below 1 Jy. The curves of NVSS and SUMSS drop quickly below $S_{NVSS} = 2.7$ mJy and $S_{SUMSS} = 12$ mJy respectively, because the surveys become incomplete.

- Uniform surface density
  - Combined catalogue at 1.4 GHz: ~30 / deg²

Bright source map at 1.4 GHz



# Faint source model

- Rayleigh-Lévy random walk
  - probability distribution: tell you how the sources apart

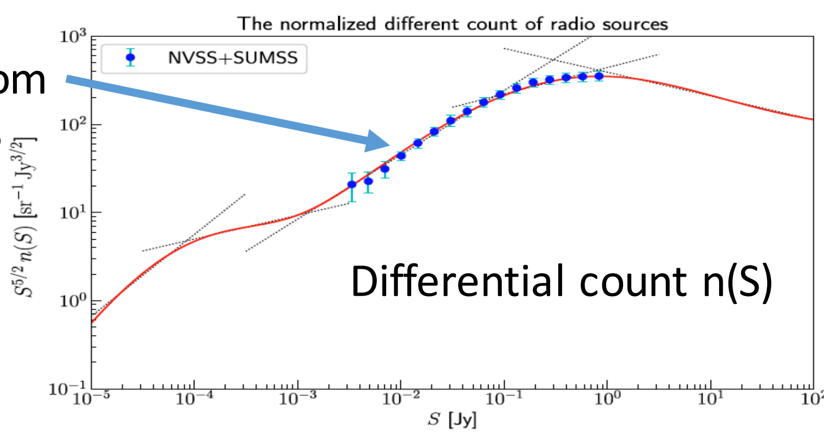
$$P(\Theta > \theta) = \begin{cases} \left(\frac{\theta}{\theta_0}\right)^{-\gamma}, & \theta \geq \theta_0 \\ 1, & \theta < \theta_0 \end{cases}$$

  - From two-point correlation function of source survey:

$\theta_0 = 6'$ $\gamma = 0.8$ (Overzier et al. 2003)

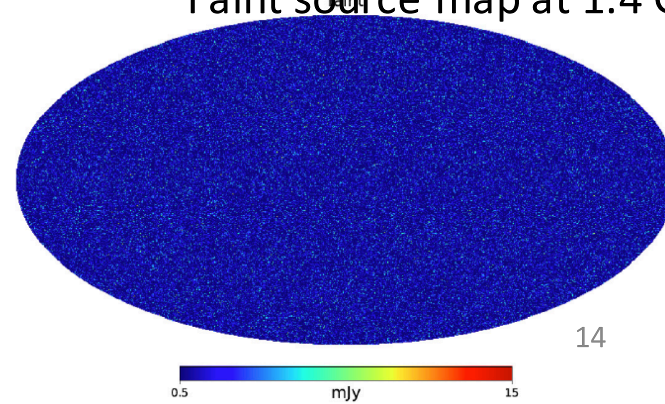
- Need differential source count n(S)
  - Tell you how many sources in a flux density (interval)

Computed from NVSS+SUMSS




Differential count n(S)

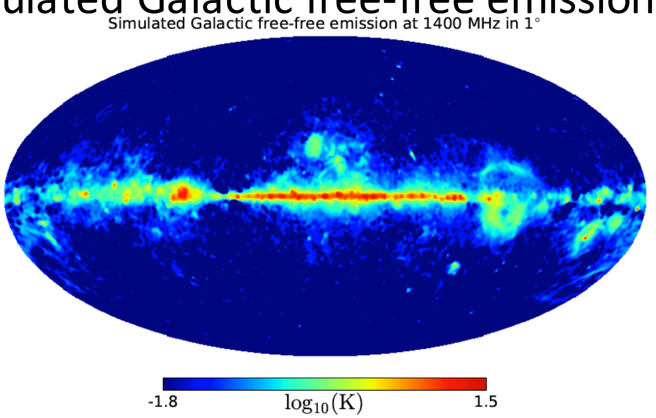
Faint source map at 1.4 GHz



# Galactic free-free emission model

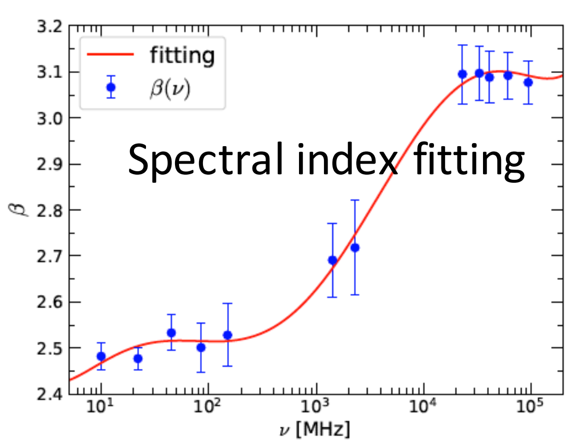
- ~1% of total foreground, but still important
- Well measured at higher frequency
  - At higher freq (>10 GHz), the diffuse warm ionized gas is optical thin
  - Galactic Hα trace Galactic free-free
- Spectral index
  - Observation at higher frequency: -2.13 to -2.17
  - Fitted formula $$\beta_{\rm ff} = -2 - \left[10.48 + 1.5\ln\left(\frac{T_e}{8000\mathrm{K}}\right) - \ln\left(\frac{\nu}{\mathrm{GHz}}\right)\right]^{-1}$$
- Extracted map
  - Combine Planck 2015 data with WMAP-9yr data (Ade et al. 2016)

Simulated Galactic free-free emission at 1.4 GHz

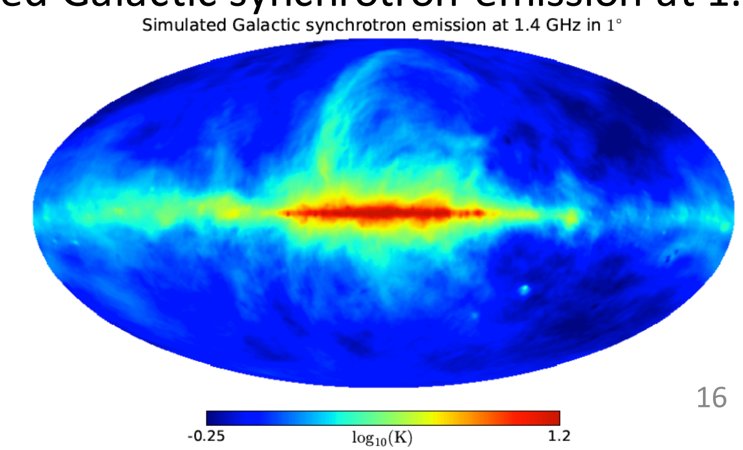


# Galactic synchrotron emission model

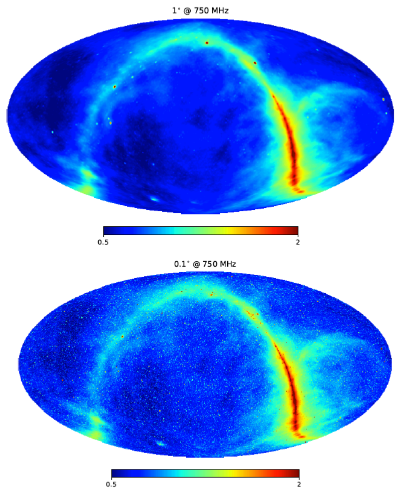
- Dominant component, ~70% of the total foreground
- Extract Galactic synchrotron from observational maps
  - Subtract the CMB: average 2.7255 K and its anisotropy
  - Subtract radio sources: radio source model above
  - Subtract Galactic free-free emission: model above
- Spectral index $N(\mu = \bar{\beta}(\nu), \sigma^2 = 0.024\left(\frac{\nu}{408\,\mathrm{MHz}}\right)^2)$

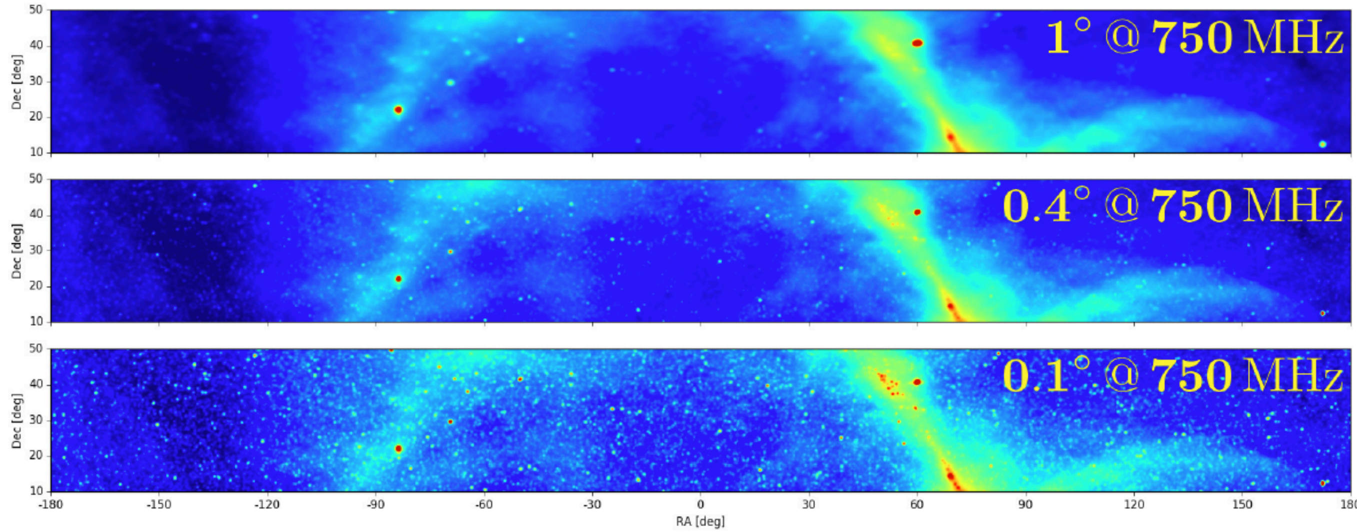



Simulated Galactic synchrotron emission at 1.4 GHz



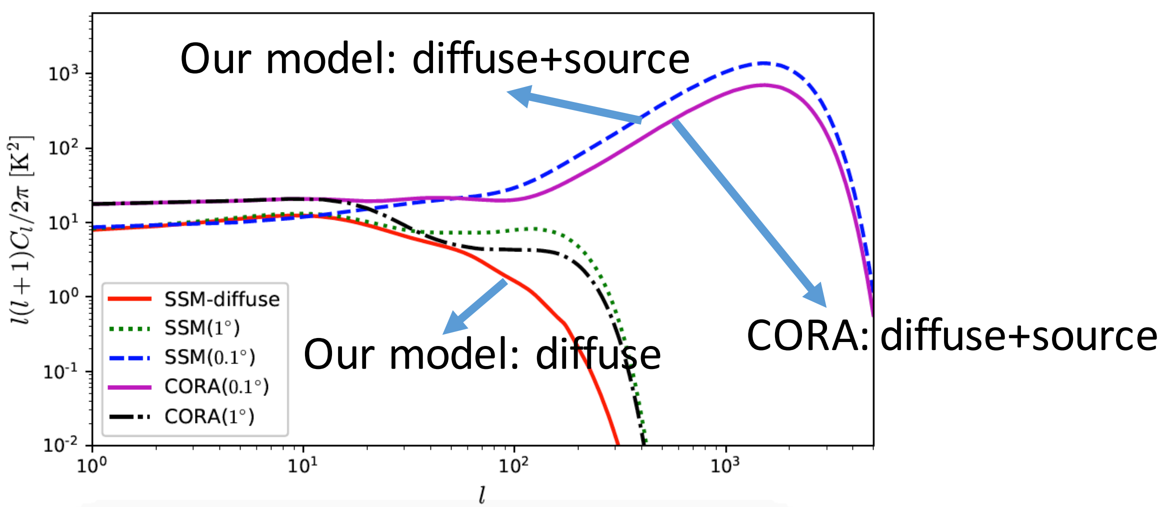
# Total foreground




- Compare with CORA

Angular power spectrum



# Brief summary

- High resolution foreground model
  - Radio source model:
    - bright sources: NVSS+SUMSS, completeness, same surface density, spectral index
    - faint source: differential count, Rayleigh-Levy random walk
  - Free-free emission model: spectral index
  - Synchrotron emission model: preprocess 8 maps, spectral index
  - Compare with GSM, CORA

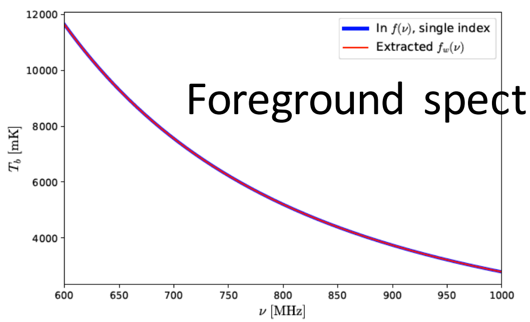
- Introduction
- High resolution foreground model
- 21cm extraction by filtering
    - Simulation data: 21cm, foreground with beam, noise
    - Design cascade of two Wiener filters
    - Test our method
    - Inaccurate beam
- Lunar orbit interferometer imaging
- PAON-4 data analysis
- Summary

# Difficult to extract

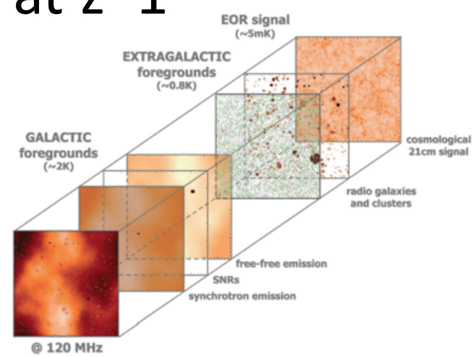
- 21cm signal is very weak: ~0.1 mK at z~1

$$T_b = 0.29 \frac{\Omega_{\mathrm{H\,I}}}{10^{-3}} \left( \frac{\Omega_m + (1+z)^{-3}\Omega_\Lambda}{0.37} \right)^{-\frac{1}{2}} \left( \frac{1+z}{1.8} \right)^{\frac{1}{2}} \mathrm{mK}$$

- Foreground contaminated: ~7000 mK at z~1




Foreground spectrum



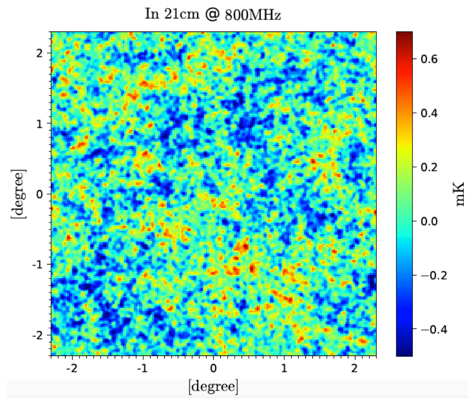

- Frequency dependent antenna beam
    - Leading to fluctuation in frequency direction

- Proposed methods
    - Polynomial fitting, PCA, SVD, ICA, …

# Simulated data

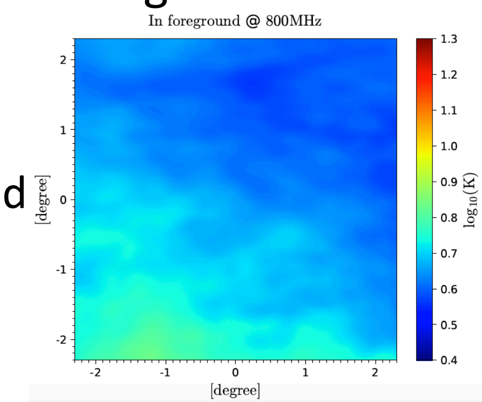
- 21cm signal
  - Bias follows dark matter <= CAMB

simulated 21cm signal at 800 MHz



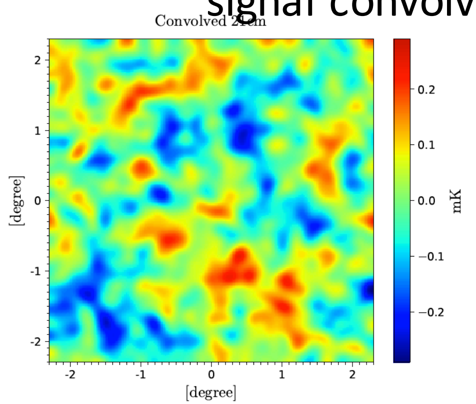

- Foreground
  - Our foreground model

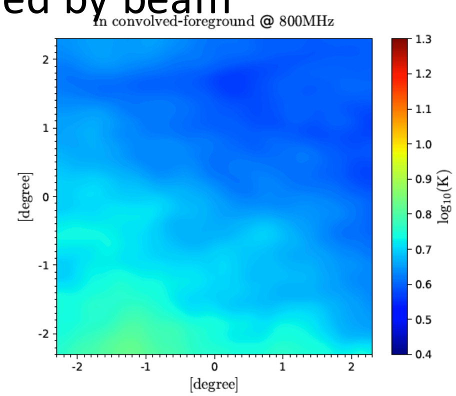
Input foreground




- Beam
  - Gaussian beam with D=100 m

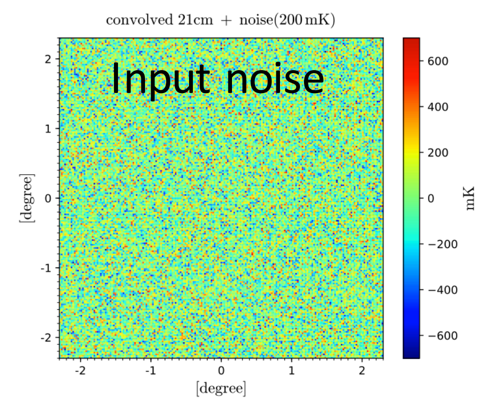
signal convolved by beam




- Noise
  - White Gaussian noise
  - 200 mK, 2000 mK

Input noise

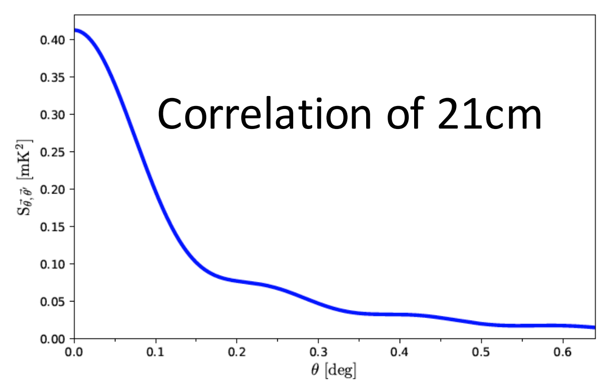


# Our method: Cascade filters

- Make use of statistic properties

- First: 1D filter in frequency domain
  - Remove frequency-smooth foreground, as 21cm signal and noise are fairly random along line of sights.

- Second: 2D filter in angular domain
  - Remove noise from 21cm signal, as the receiver noise is uncorrelated in angular direction, while the 21-cm signals are correlated because it traces the large scale structure of the Universe.



# Wiener filter design

- Wiener filter is an optimal filtering system minimizing the mean square error between the estimated random process and the desired process: $\langle (x-\hat{x})(x^{\mathrm{T}}-\hat{x}^{\mathrm{T}})\rangle$

- Filtering in frequency domain
  - 1D Wiener filter designed to remove the foreground:

$$\mathbf{W}_{\nu}^{f} = \mathbf{F}\left[\mathbf{F}+\mathbf{S}+\mathbf{N}\right]^{-1}$$

The weight for each components

  - Extract the 21cm + noise:

$$\mathbf{W}_{\nu} = \mathbf{I} - \mathbf{W}_{\nu}^{f}$$

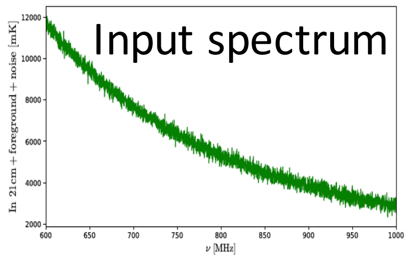
- Filtering in angular domain
  - 2D Wiener filter designed to extract the 21cm signal

$$\hat{x} = \mathbf{W}y \equiv \underline{\mathbf{SA}^{\mathrm{T}}\left[\mathbf{ASA}^{\mathrm{T}}+\mathbf{N}\right]^{-1}}\, y$$

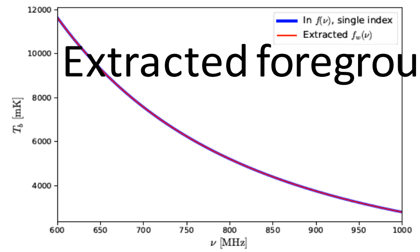
# Filtered signal

- Filtering in frequency domain (1D)

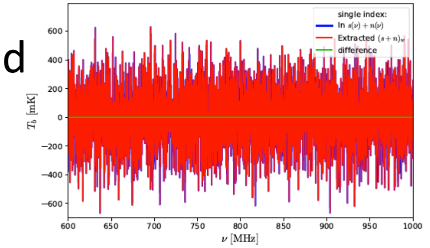
Input spectrum




Extracted foreground

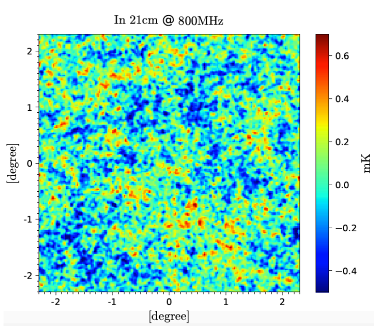



Extracted 21cm+noise

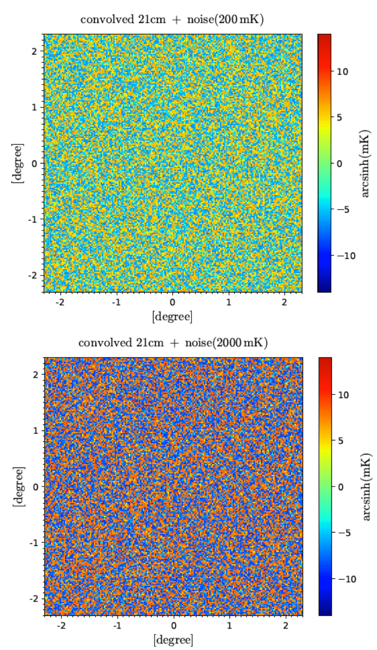



- Filtering in angular domain (2D)

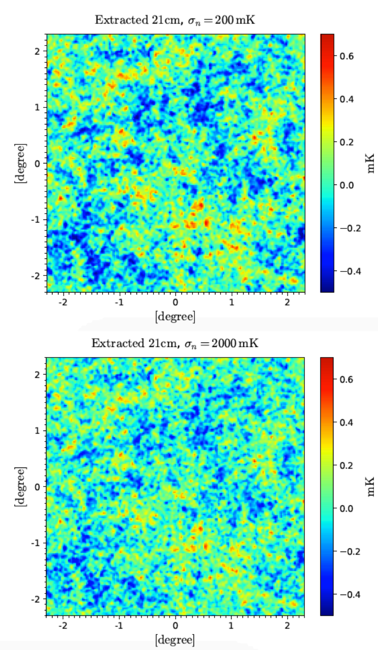
Input 21cm



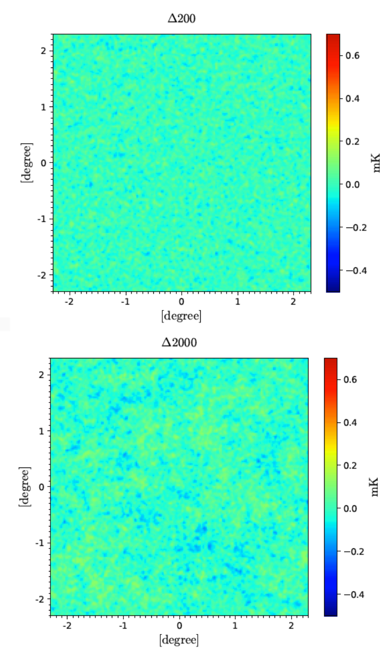

total




Extracted 21cm



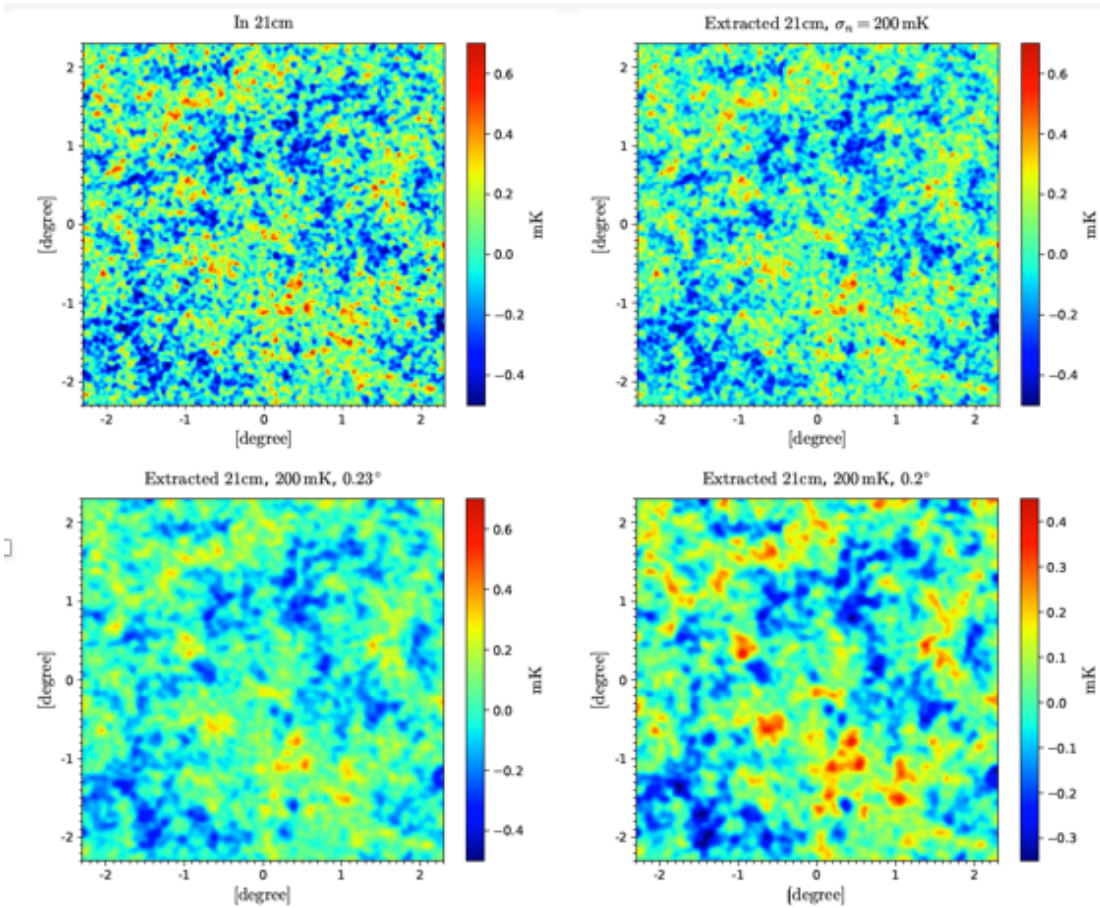

residual



# Effect of inaccurate beam model

Input 21cm

Exact beam 0.25°




Inaccurate beam 0.23°

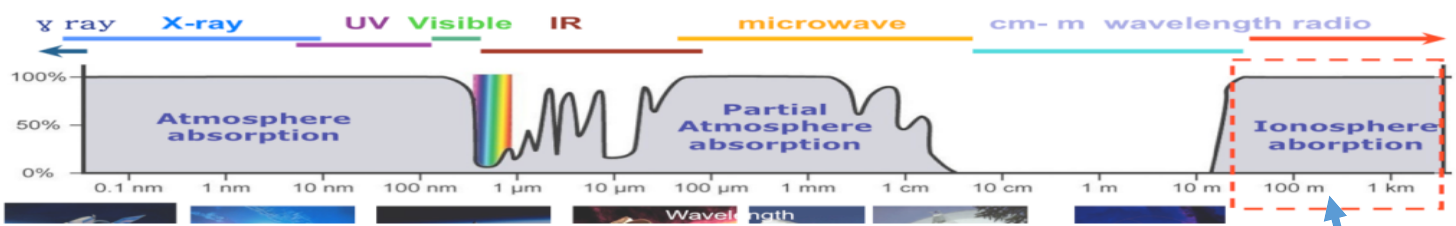
Inaccurate beam 0.20°

# Brief summary

- 21cm extraction by filtering
  - Signal data simulation
  - Design a cascade of two Wiener filters: in frequency then in angular
  - Test two noise level: 200 mK and 2000 mK
  - Try with two inaccurate beam models

- Introduction
- High resolution foreground model
- 21cm extraction by filtering
- Lunar orbit interferometer imaging
  - Hard to observe <30 MHz on the ground => on lunar orbit
  - Advantages and disadvantages
  - Our new imaging algorithm
  - Solve mirror symmetry problem
  - Solve imaging problem with full sky FoV, time-varying blockage, time-varying and noncoplanar baselines
- PAON-4 data analysis
- Summary

# Poorly known below 30 MHz




- Reason
  - strong refraction and absorption by the ionosphere below 30 MHz
  - strong RFI (radio frequency interference)

- Ground observation
  - strong radio phenomena from solar system:
    - solar radio bursts , planetary radio activities

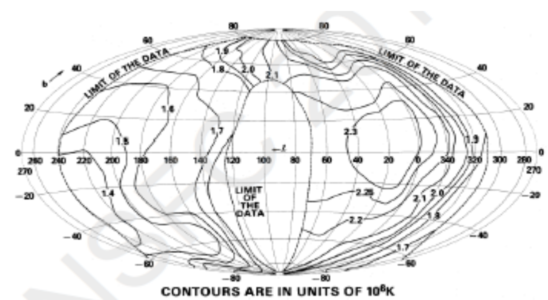



Image from RAE-2

- Space mission
  - Proposed but not realized
    - earth orbit: SunRISE --- strong RFI
    - Sun-Earth L2 point: ALFA, FIRST, SURO-LC --- all-time observation, still RFI
  - IMP-6, RAE-1, RAE-2: the Moon can shield the RFI from the Earth

# DSL (Discovering the Sky at Longest wavelength)

- Lunar orbit interferometer
  - First realized: Longjiang-1 and Longjiang-2, China, Chang’e-4
  - Also known as DSL pathfinder
- Advantages
  - Observe on far-side, shield RFI
  - Transfer data on near-side, don’t need TDRSS
  - Short orbital period, use solar power
  - Interferometry, high resolution
- Disadvantages
  - All sky field of view, sphere
  - Mirror symmetry problem
  - Time-varying noncoplanar baselines
  - Time-varying blockage

No applicable imaging algorithm

- ~~3D FFT, W-Projection, W-Stacking~~

Linear array

uv coverage

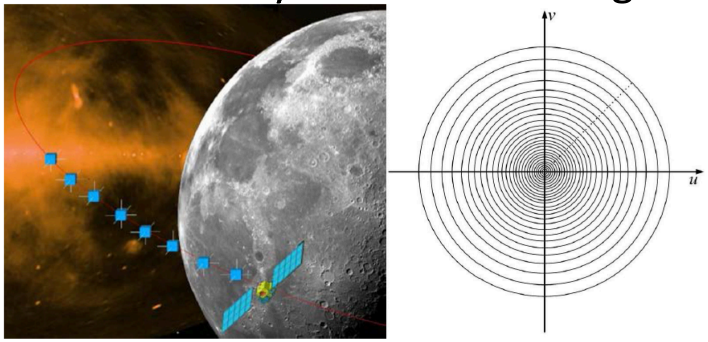
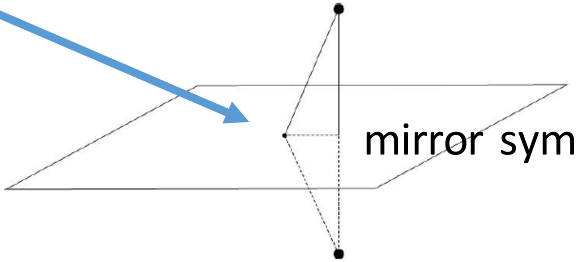



Fig. 5.2 Left: a linear array of flying satellites around the Moon. Right: the baselines between satellite pairs in the array swipe a number of concentric rings on the $(u, v)$ plane.



# Our new imaging algorithm

- Start from visibility definition: $V_{ij} = \int A_{ij}(\vec{n})\, T(\vec{n})\, e^{-i\vec{k}\cdot\vec{r}_{ij}}\, \mathrm{d}^2\vec{n}$
- Integral (continuous) => sum (discrete): $V_{ij}(t) = \sum_{n}^{N_{\mathrm{pix}}} H(n,t)\, T(n)\, \Delta\Omega$
- Matrix formalism in pixel: $\mathbf{V} = \mathbf{H}\,\mathbf{T} + \mathbf{n}$
- Matrix formalism in spherical harmonic:

$$
\begin{aligned}
V_{ij}(t) &= \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} (-1)^m \mathcal{H}_{l,-m}(t)\, \mathcal{T}_{lm} \\
V^*_{ij}(t) &= \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} \mathcal{H}^*_{l,m}(t)\, \mathcal{T}_{lm}
\end{aligned}
$$

$$\mathbf{V} = \mathcal{H}\,\mathcal{T} + \mathbf{n}$$

- Estimator
  - in pixel: $\hat{\mathbf{T}} = (\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{V} \equiv \mathbf{B}^{-1}\mathbf{V}$
  - in spherical harnomic: $\hat{\mathcal{T}} = (\mathcal{H}^\dagger \mathbf{N}^{-1} \mathcal{H})^{-1} \mathcal{H}^\dagger \mathbf{N}^{-1} \mathbf{V} \equiv \mathcal{B}^{-1}\mathbf{V}$

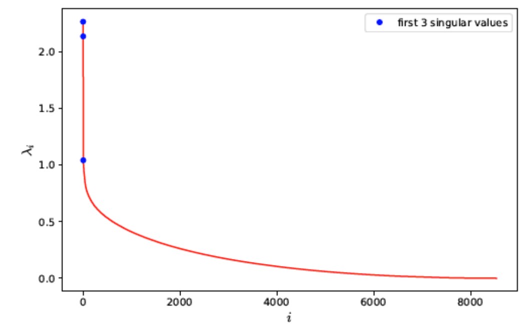
# Pseudo-inverse through SVD

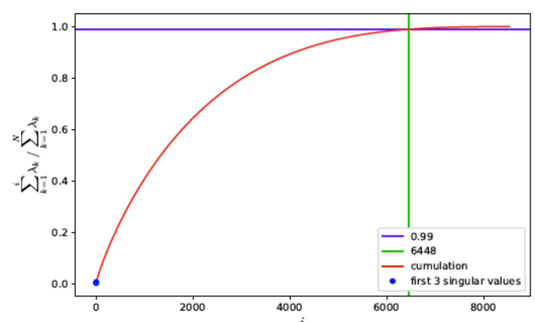
- Use SVD to compute $B^{-1}$:

$$\mathbf{B} = \mathbf{U}\Sigma\mathbf{W}^{\dagger} \Rightarrow \bar{\mathbf{B}}^{-1} = \left(\mathbf{U}\bar{\Sigma}^{-1}\mathbf{W}^{\dagger}\right)^{\dagger}$$

- Deal with very small singular values:
  - General method: absolute threshold, relative threshold to vmax
    - Singular matrix dependent
    - Experience (Jiao Zhang et al. 2016)
  - Automatic method: cumulation ratio

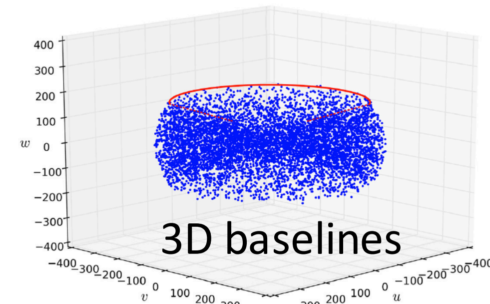
$$\frac{\sum_{i=1}^{N_{\mathrm{thr}}} \lambda_i}{\sum_{N_{\mathrm{all}}} \lambda_i} = 0.99$$



# Baselines and Moon’s blockage

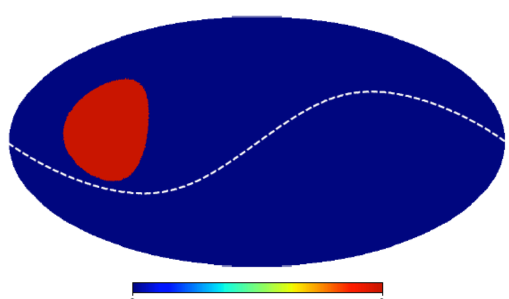
- Orbit and baseline
  - Angle between orbital plane and Moon’s equatorial plane is 30$^\circ$
    - Precession amplitude and rate are large: 360$^\circ$ in 1.29 years
  - Orbital plane precision will produce 3D baselines.



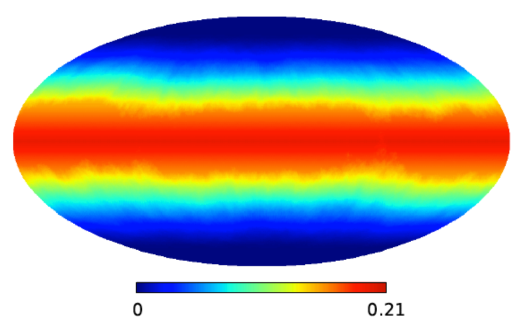
3D baselines

- Blockage
  - Angular size:
    - h is the height to the lunar surface
    - Change with time

$$\theta_m \approx 2\arcsin\left(\frac{R_\mathrm{m}}{R_\mathrm{m}+h}\right)$$



Blocked sky

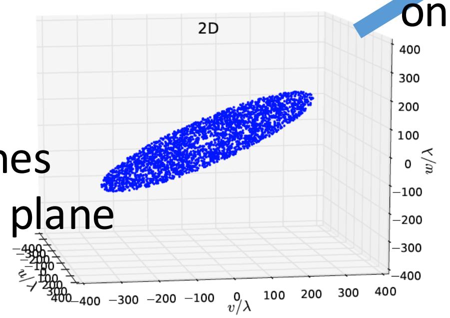


Blockage fraction

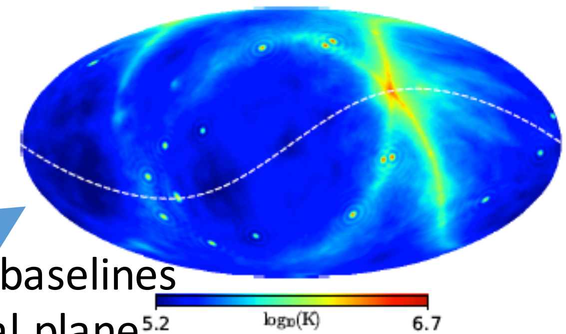
# Solution of mirror symmetry problem

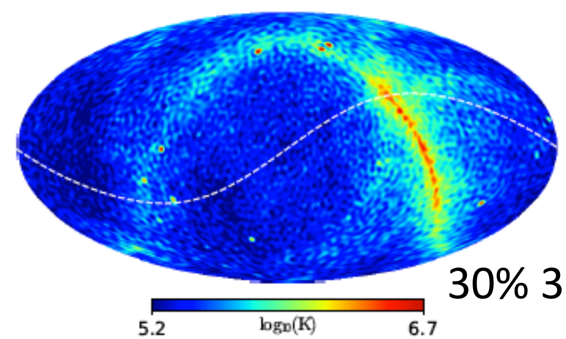
- Orbit and baseline
  - Pitch 30° : precess fast: 360° in 1.29 years
  - Large elliptic orbit: save energy
- All baselines on orbital plane
  - Symmetry images

2D baselines on orbital plane



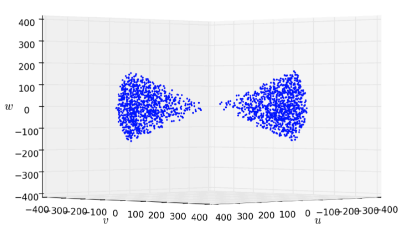

Only 2D baselines on orbital plane



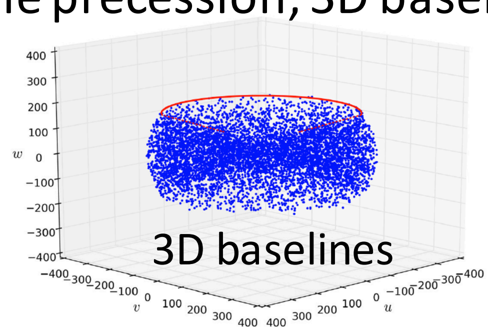




30% 3D

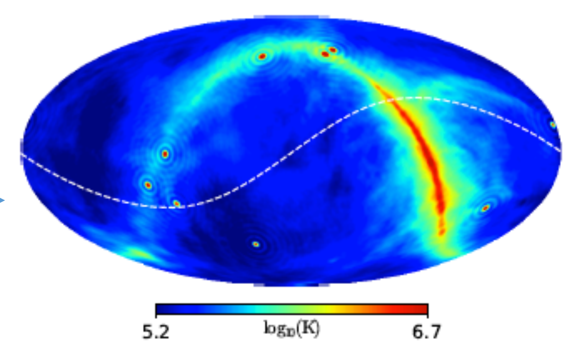
- Solve symmetry problem
  - Orbital plane precession, 3D baselines



Cross-section



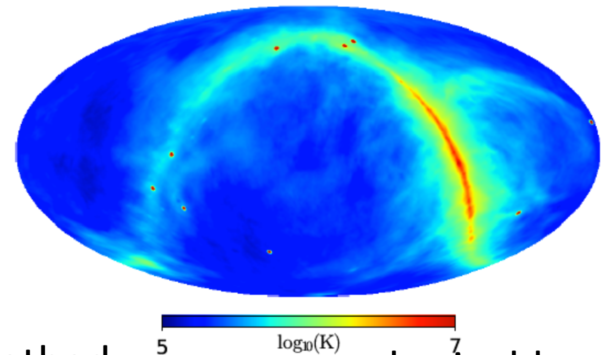
3D baselines




Full 3D baselines

# Spherical harmonic v.s. pixel

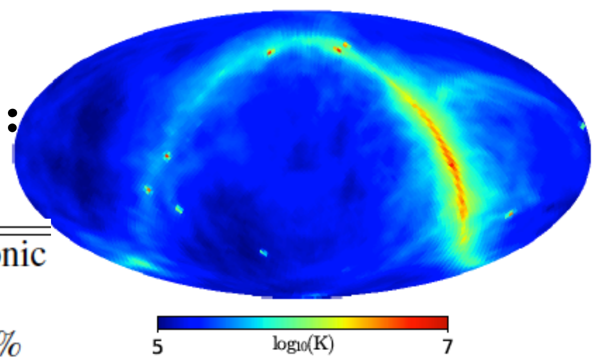
- Input map:

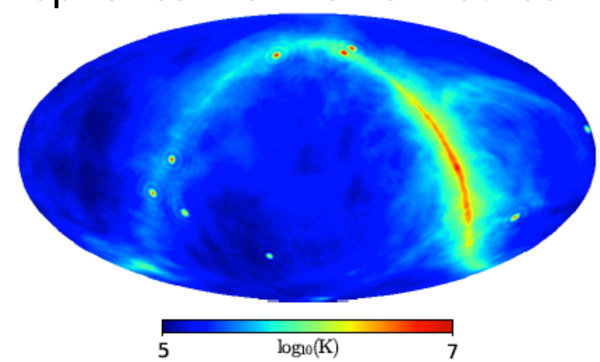



pixel method

spherical harmonic method

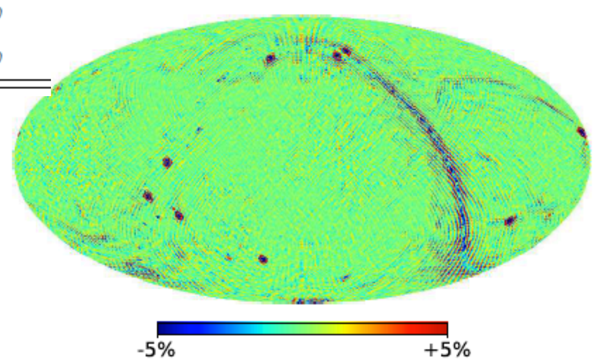
- Reconstructed map:

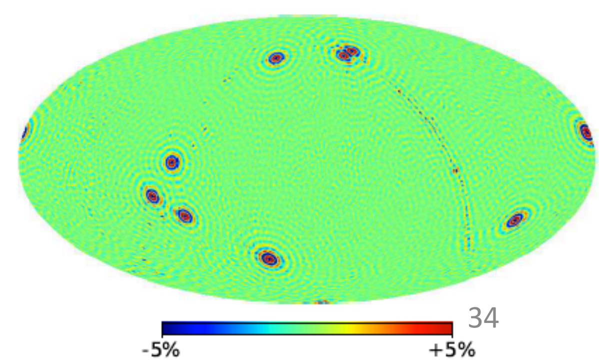



| | pixel method | spherical harmonic |
|---|---|---|
| maximum | 45% | 20% |
| minimum | 0.02% | 0.006% |
| mean | 2.3% | 1.6% |
| median | 1.7% | 1.2% |

- Relative error:



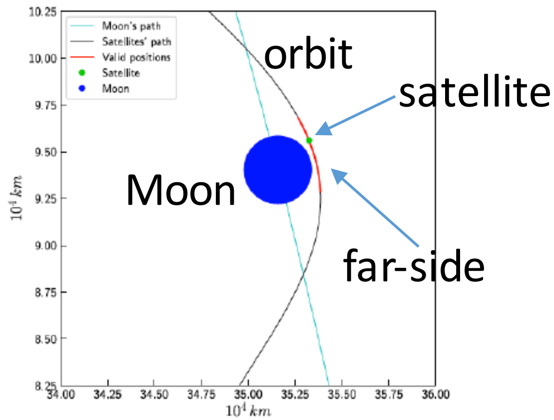
# Blockage effect

Can still image well

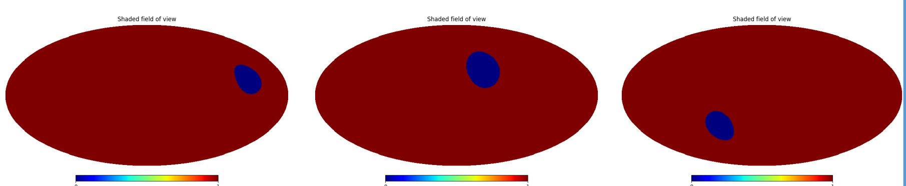
- Blockage angular size

$$\theta_m \approx 2\arcsin\left(\frac{R_\mathrm{m}}{R_\mathrm{m}+h}\right)$$

- Valid observation time

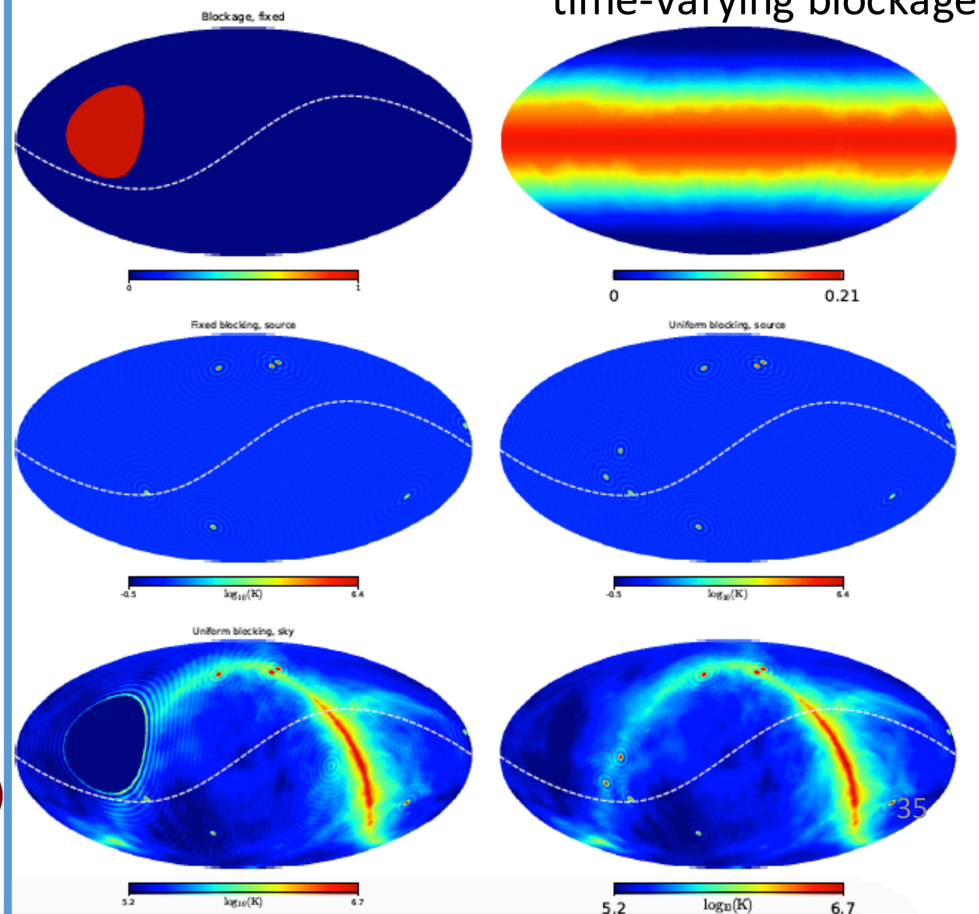



- Time-varying blocked sky



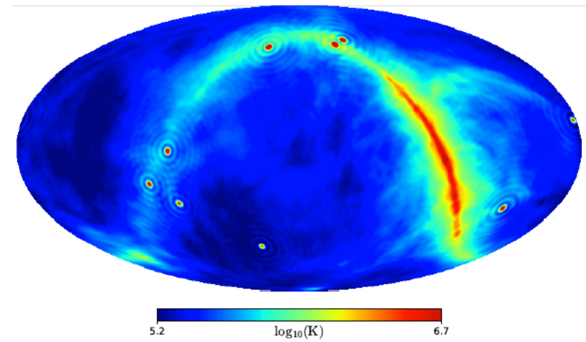

Test our algorithm
fix blockage

time-varying blockage



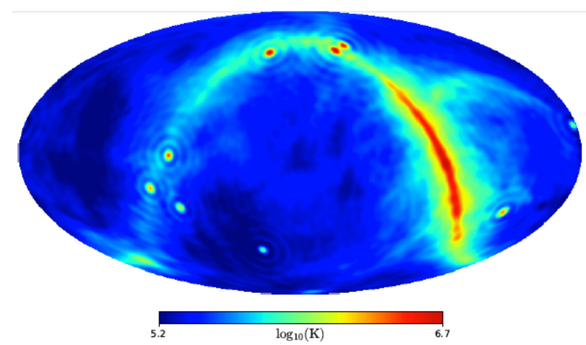
# Baseline distribution effect

- Only long baselines, >6 km
  - For safety
  - Larger sidelobe



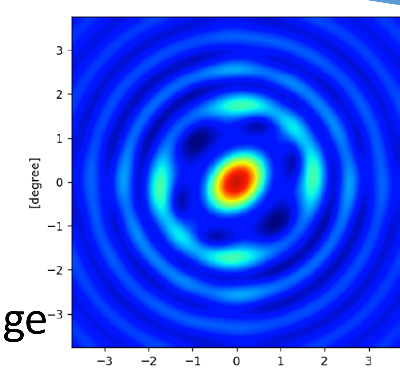

Long baselines

- Only short baselines, <0.5 km
  - More complete uv coverage
  - Lower resolution

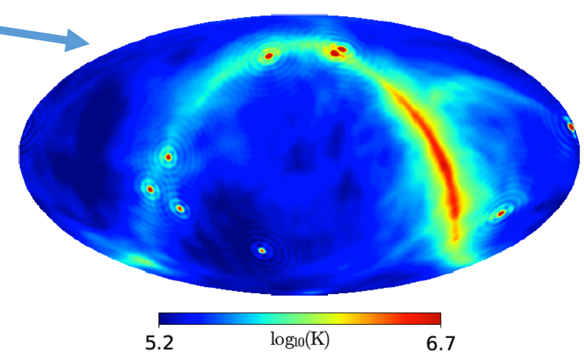



Short baselines

- Nonuniform distribution
  - 1/4 sphere
  - Short time observation




Distorted image




nonuniform baselines


36

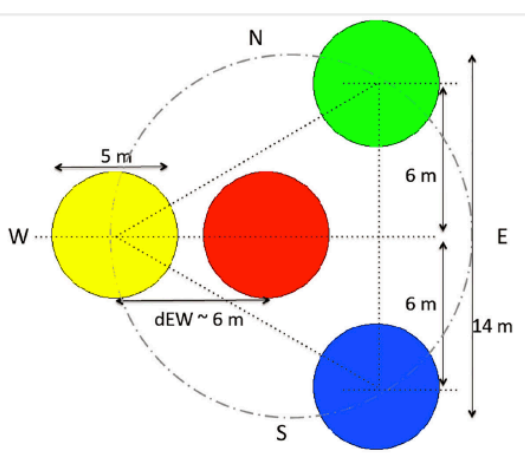
# Brief summary

- Lunar orbit interferometer imaging
  - <30 MHz, advantage
  - Whole sky FoV, symmetry problem => orbital plane precessio
  - Time-varying blockage, time-varying and noncoplanar baselines + whole sky FoV => no applicable imaging algorithm
  - Our imaging algorithm in spherical harnomic
  - Blockage effect, baseline distribution effect, sph harnomic v.s. pixel

- Introduction
- High resolution foreground model
- 21cm extraction by filtering
- Lunar orbit interferometer imaging
- PAON-4 data analysis
  - Surface fitting and RFI mitigation
  - System gain calibration
  - Pointing calibration
  - Phase calibration
  - Amplitude calibration
  - Map-making
- Summary

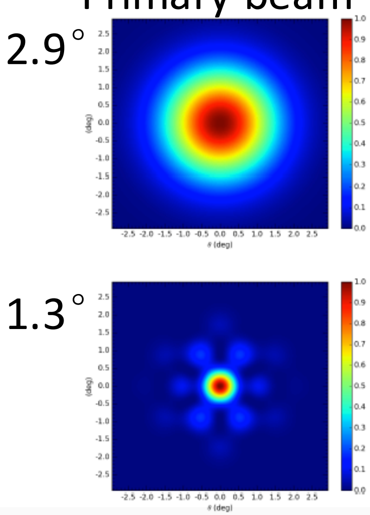
# PAON-4

- Transit interferometer array
  - 4 dishes with D=5 m
  - Located at Nançay radio observatory
  - 1250—1500 MHz, 4096 bins -> freq resol 61 kHz
  - Longest baseline: 12 m, 1.3 degree at 1.4 GHz
  - Total receiving area: 75 $m^2$




PAON-4 configuration

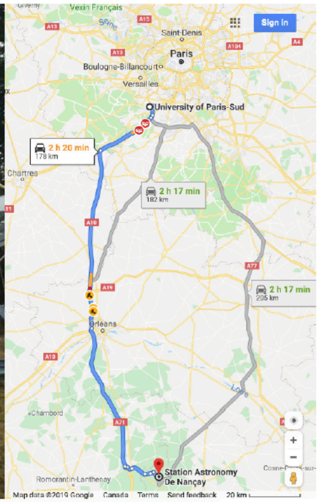
Primary beam




Synthetic beam

PAON-4

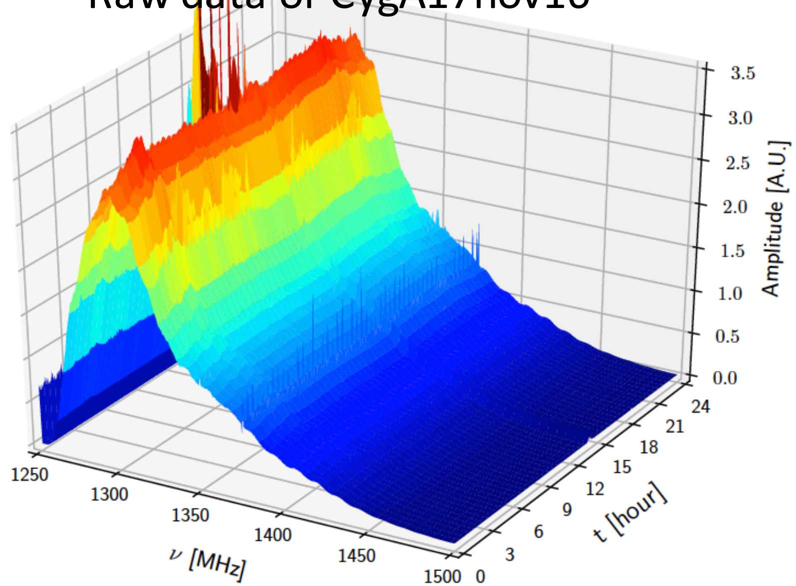
Location



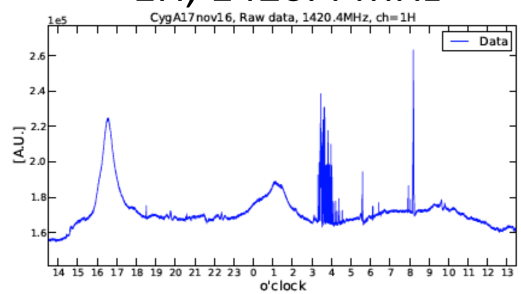
# Raw data

- 4 dual polarization
- 36 visibilities
  - 8 auto
  - 6 H-cross
  - 6 V-cross
  - 16 HV-cross

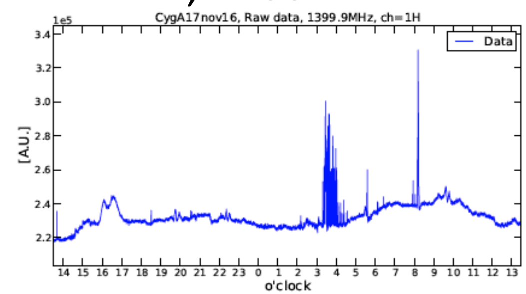
Raw data of CygA17nov16

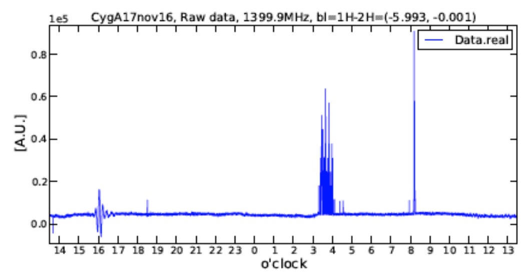



Auto-correlation
1H, 1420.4 MHz

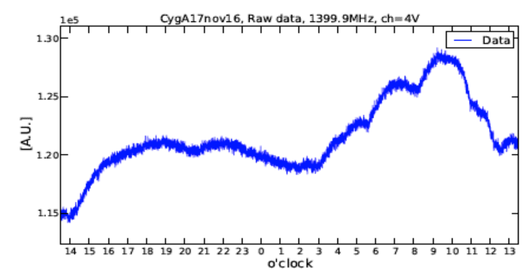



Auto-correlation
1H, 1400 MHz



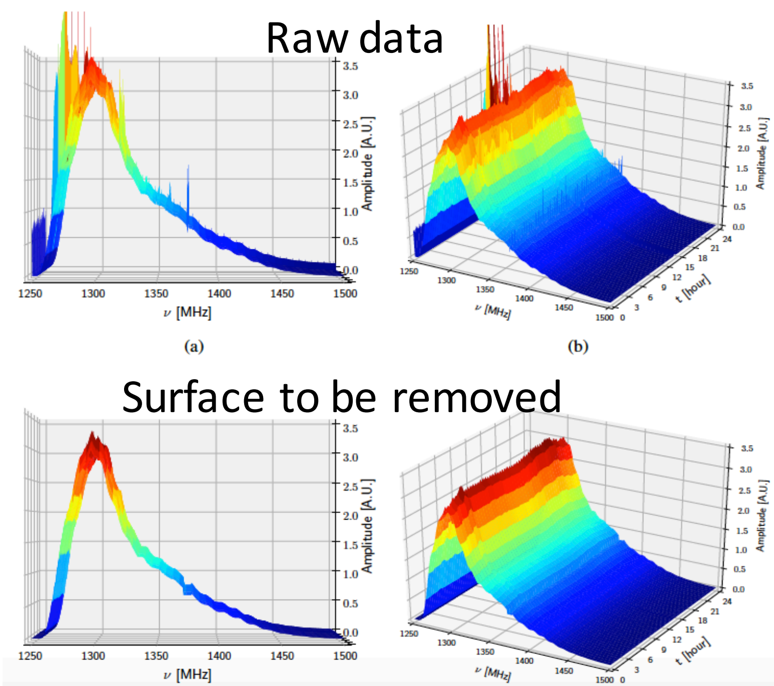

Cross-correlation
real part, 1H2H, 1400 MHz

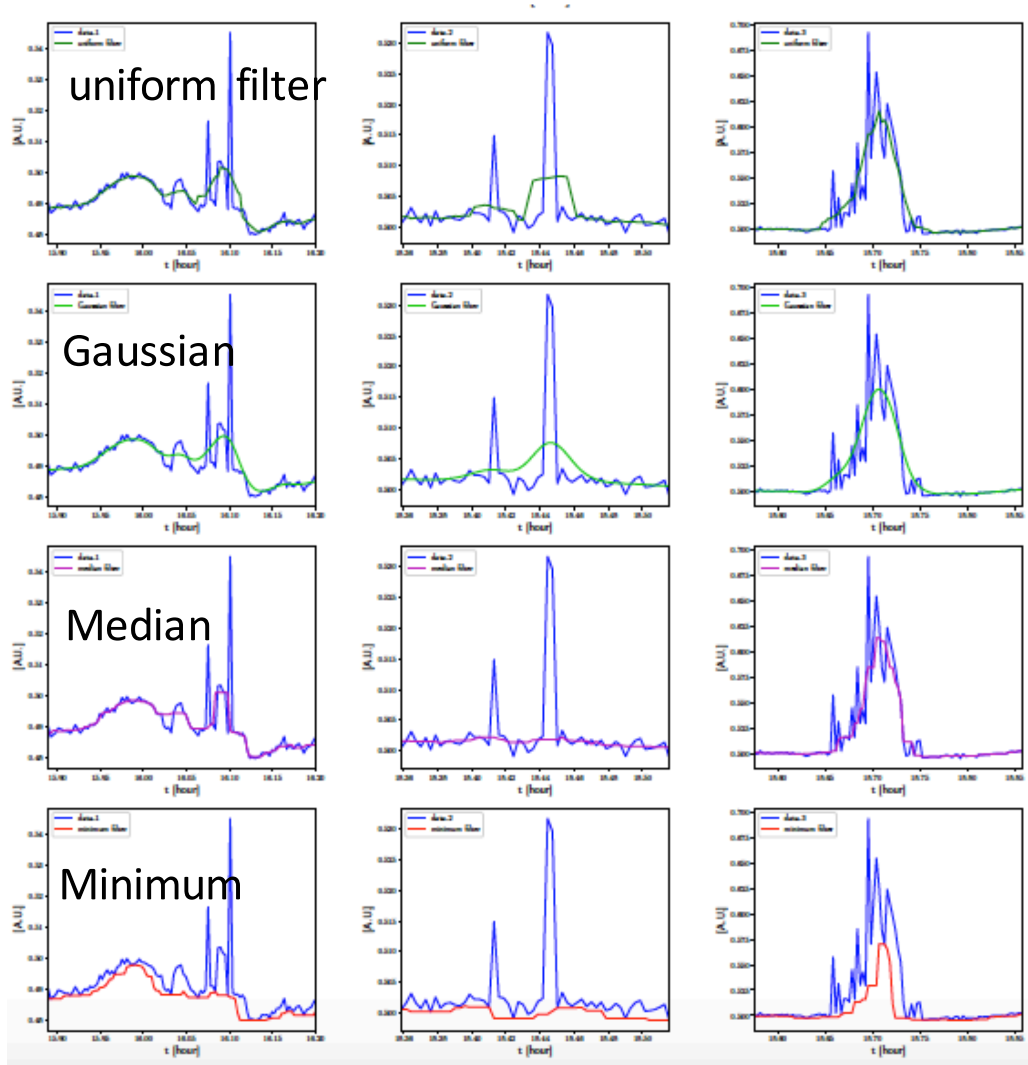



Auto-correlation
4V, 1400 MHz,
thermal noise

# Before RFI mitigation: surface removal

- Methods
  - Uniform window filtering
  - Gaussian filtering
  - Median filtering
  - Minimum filtering

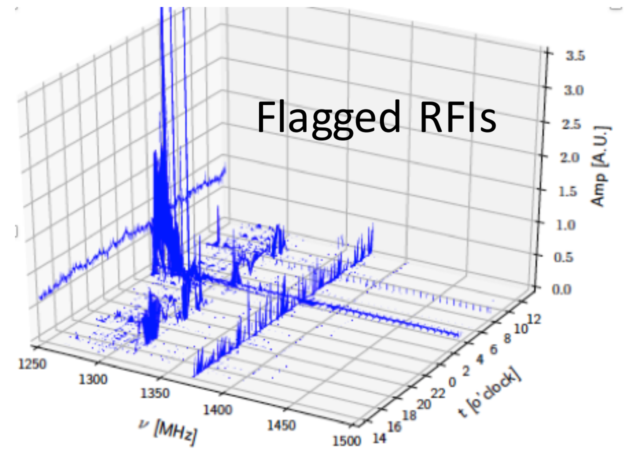


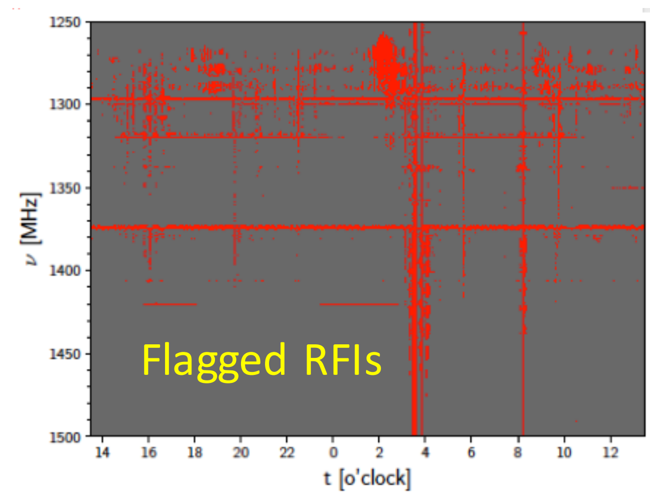



# RFI mitigation

- SumThreshold
  - Works well on LOFAR and WSRT data
    - (Offringa et al. 2010)
  - Connected samples
    - 1, 2, 4, 8, 16 (experiential)
  - Thresholds

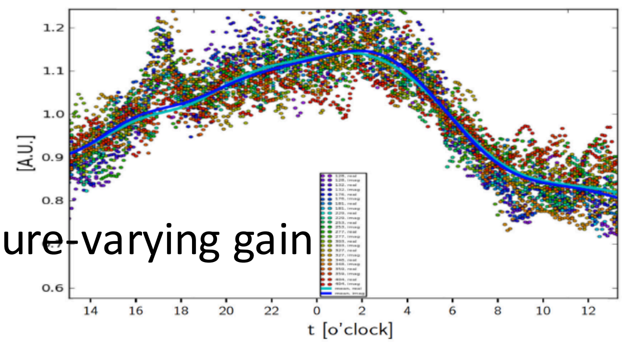
    $\geq \chi \cdot \mathrm{M} \cdot \sigma$
  - $\chi$
    - 4, 3, 2.5, 2, 1.8
  - Two rules
    - Replace, recompute



# System gain calibration

- Temperature-dependent
  - From the data




temperature-varying gain

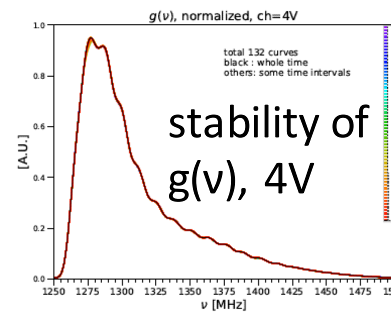
- Assume gain variations of time and frequency are mutual independent

$$V_a(t,\nu) = G(t,\nu)\cdot[I(t,\nu)+N(t,\nu)]$$
$$= g(t)\,g(\nu)\cdot[I(t)\,I(\nu)+n(t)\,n(\nu)]$$

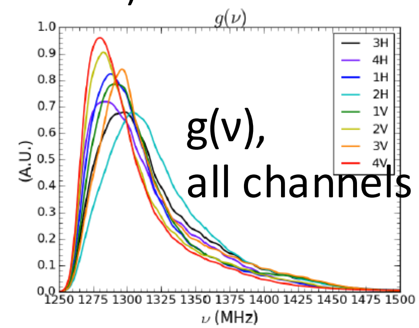
- From the laboratory measurement

4dB gain variation from $0°\mathrm{C}$ to $55°\mathrm{C}$

(PAON-4 paper, Ansari et al. 2019)

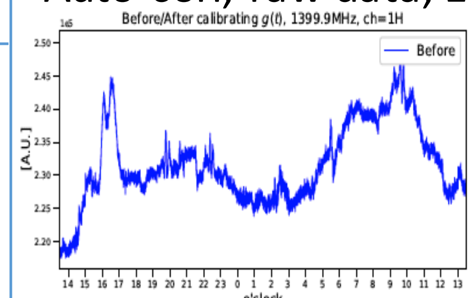


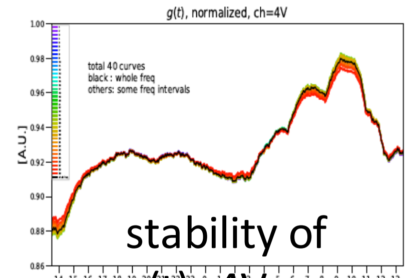

stability of g(ν), 4V



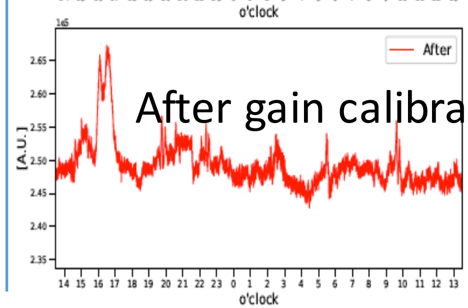

g(ν), all channels

Auto-corr, raw data, 1H



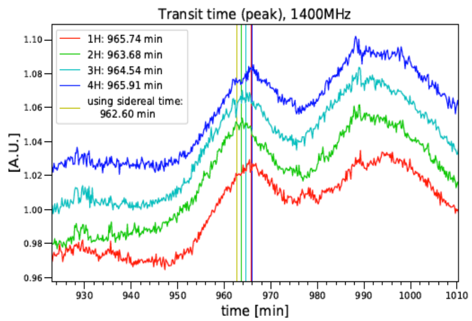

stability of g(t), 4V




After gain calibration

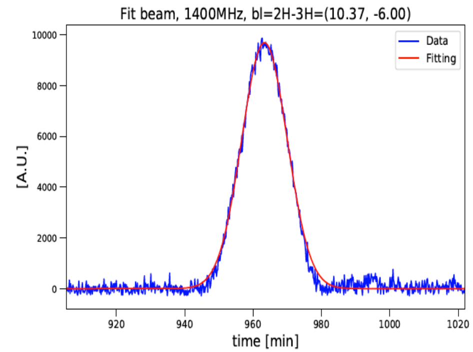
# Pointing calibration and Effective diameter

- Not use auto-corration data




Peaks of CygA in Auto-corr

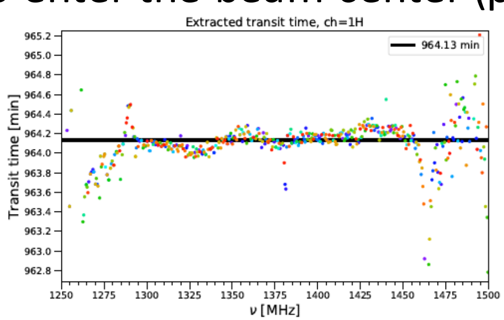
- Use cross-correlation data



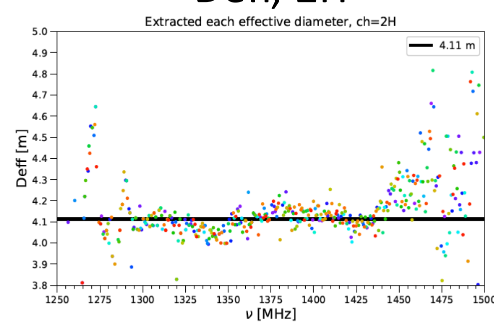

Peak of CygA in Cross-corr

- Effective diameters: 4.22m, 4.11m, 4.23m, 4.36m
- Pointing offsets: 1.3°, 0.5°, 0.7°, 1.2°

Time to enter the beam center (pointing)




Deff, 2H



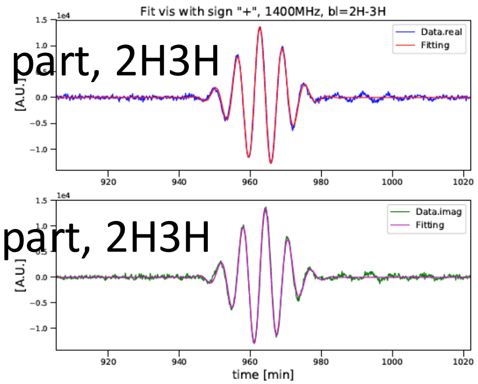
# Phase calibration

- Use bright source Cygnus A
  - Method 1: fit the visibility

$$V_{ij} = I_0\, A_{ij}\, e^{i\,\frac{2\pi}{\lambda}\,\vec{B}\cdot\vec{s}}$$

Real part, 2H3H

Imaginary part, 2H3H

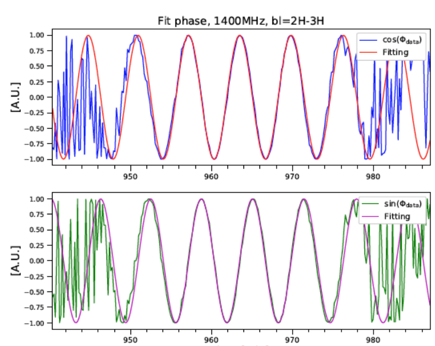



  - Method 2: fit the sin(phase)

$$\Phi_{ij} = \frac{2\pi}{\lambda}\vec{B}\cdot\vec{s} = \arctan\left(\frac{V_{ij}^{\mathrm{imag}}}{V_{ij}^{\mathrm{real}}}\right) \qquad \cos(\Phi_{ij}) + i\sin(\Phi_{ij})$$

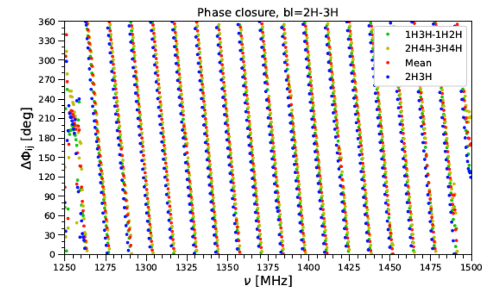
cos(phase), 2H3H

sin(phase), 2H3H

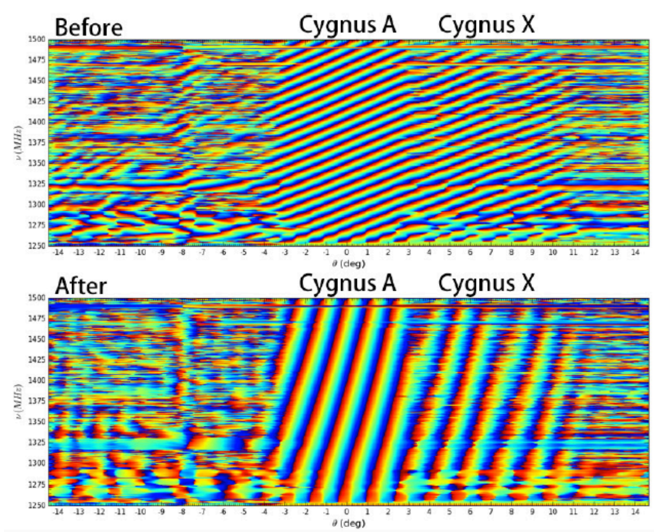



- Phase closure checking

$$\Phi_{ij} = \Phi_{ik} + \Phi_{kj}$$




- Calibrated




Before

After

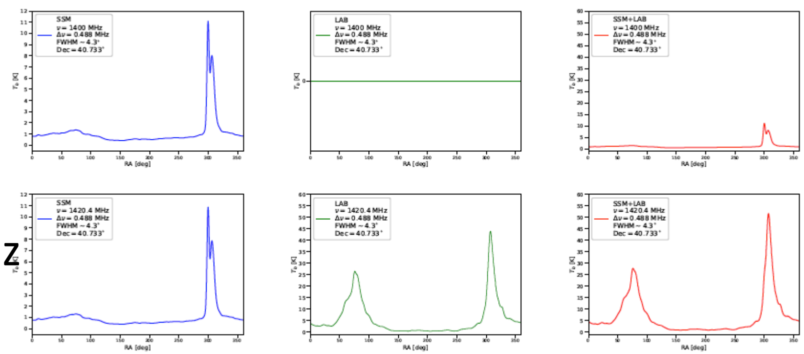
# Amplitude calibration

- Calibrate Auto-corr data
  - Our foreground model + LAB

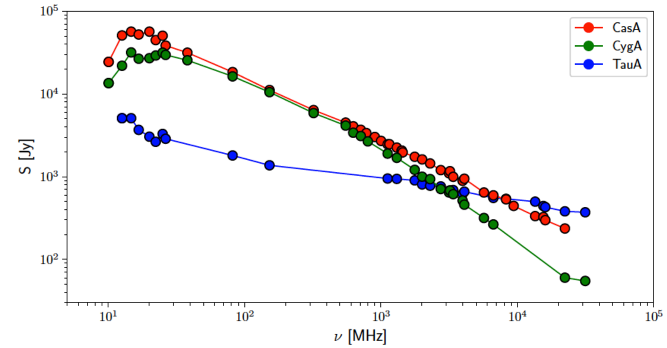
foreground | Galacitc 21cm | total

1400 MHz

1420.4 MHz



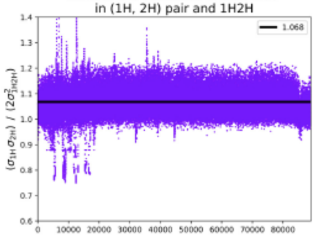

- Calibrate Cross-corr data
  - Use Cygnus A



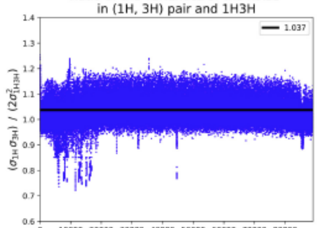

- Data noise checking

$$2\sigma^2_{\mathbf{C}_{XY}} = \sigma_{\mathbf{A}_X}\,\sigma_{\mathbf{A}_Y}$$

Ratio of the variances of the noise in (1H, 2H) pair and 1H2H




Ratio of the variances of the noise in (1H, 3H) pair and 1H3H

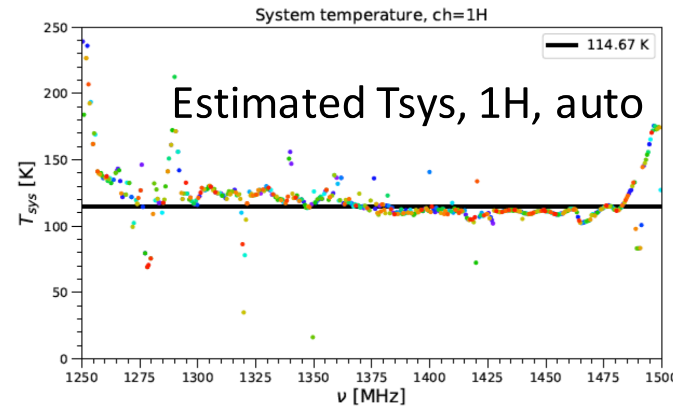


# System temperature

- Assume a thermal noise
- For auto-correlation

$$T^{i}_{\text{sys}} = \sigma_i \sqrt{\Delta t \cdot \Delta \nu}$$

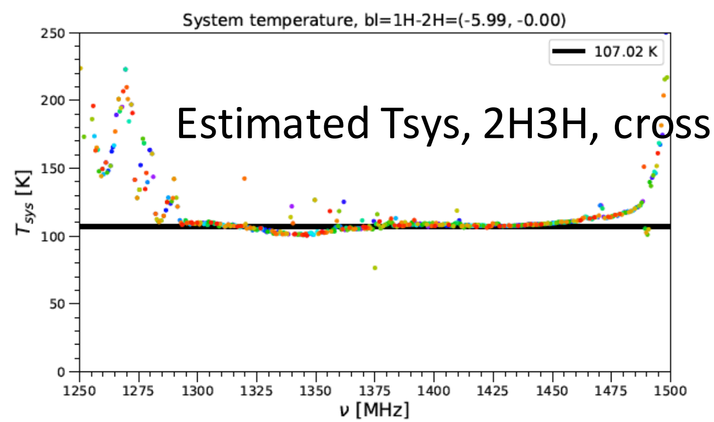
  - 114K、104K、125K、126K




- For cross-correlation

$$T^{ij}_{\text{sys}} = \sigma_{ij} \sqrt{2\,\Delta t \cdot \Delta \nu}$$

  - 107K, 120K, 116K, 108K, 118K, 127K




- The average system temperature: 120 K

# Map-making

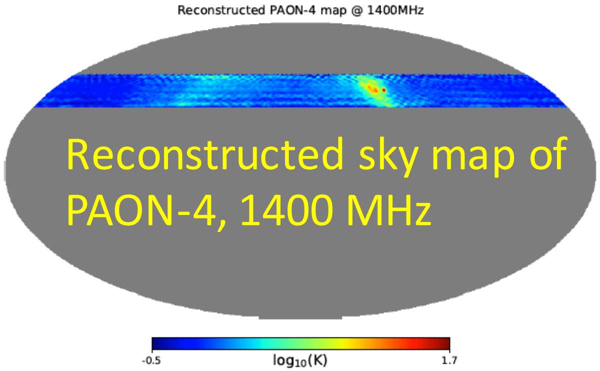
- PAON-4 scan
  - drift scan (transit)
  - 11 scans $\delta = +35.7^\circ$ to $+45.7^\circ$

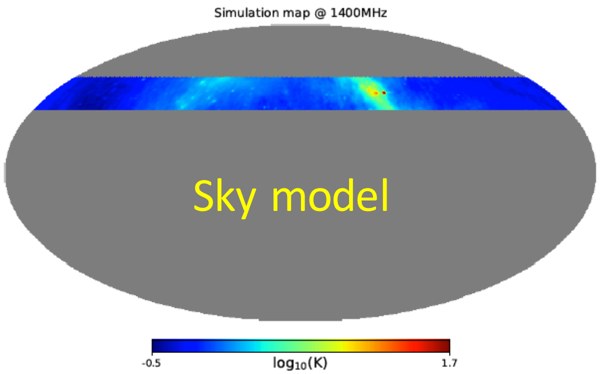
| Dataset | Dec. | Freq. [MHz] | Begin (UTC) | Duration (Sidereal time) |
|---|---|---|---|---|
| CygA17nov16 | +40.733° | 1250–1500 | 13:30 | 24 h |
| CygA18nov16 | +41.733° | 1250–1500 | 14:00 | 24 h |
| CygA19nov16 | +39.733° | 1250–1500 | 14:45 | 24 h |
| CygA21nov16 | +42.733° | 1250–1500 | 10:00 | 24 h |
| CygA22nov16 | +38.733° | 1250–1500 | 10:30 | 24 h |
| CygA23nov16 | +43.733° | 1250–1500 | 11:00 | 24 h |
| CygA24nov16 | +37.733° | 1250–1500 | 11:30 | 24 h |
| CygA25nov16 | +44.733° | 1250–1500 | 12:00 | 24 h |
| CygA26nov16 | +36.733° | 1250–1500 | 12:30 | 24 h |
| CygA27nov16 | +45.733° | 1250–1500 | 13:00 | 24 h |
| CygA1dec16 | +35.733° | 1250–1500 | 13:30 | 24 h |

- m-mode algorithm
  - (Zhang J et al. 2016)

$$\mathcal{V}_m = \sum_{l=|m|}^{l_{\max}} (-1)^m \mathcal{B}_{l,-m} \mathcal{T}_{lm} \qquad \mathcal{V}^*_{-m} = \sum_{l=|m|}^{l_{\max}} \mathcal{B}^*_{lm} \mathcal{T}_{lm}$$

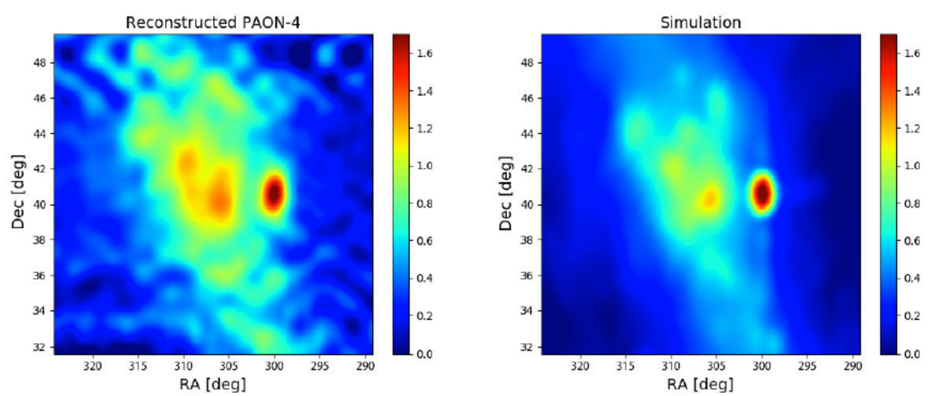
$$\begin{bmatrix} \mathcal{V}(m_0) \\ \mathcal{V}(m_1) \\ \cdots \\ \mathcal{V}(m_{\max}) \end{bmatrix} = \begin{bmatrix} \mathcal{B}(m_0,l) & & & \\ & \mathcal{B}(m_1,l) & & \\ & & \cdots & \\ & & & \mathcal{B}(m_{\max},l) \end{bmatrix} \begin{bmatrix} \mathcal{T}(l,m_0) \\ \mathcal{T}(l,m_0) \\ \cdots \\ \mathcal{T}(l,m_{\max}) \end{bmatrix}$$




Around Cygnus A

PAON-4 Model



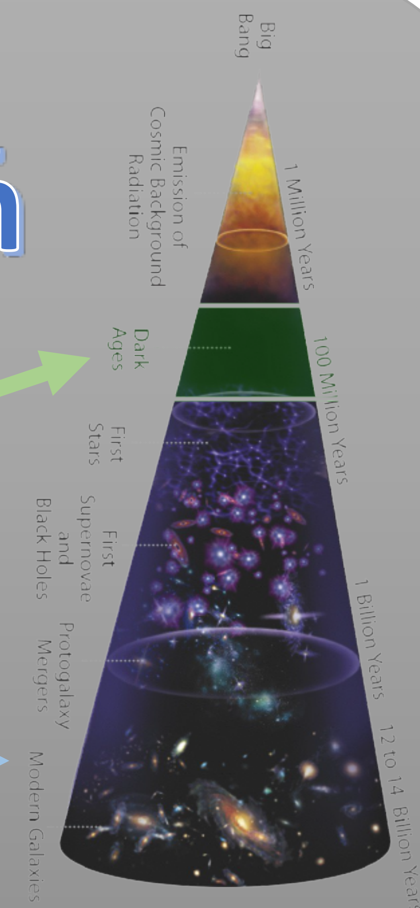
# Brief summary

- PAON-4 data analysis
  - RFI mitigation: surface fitting by minimum filter, SumThreshold
  - System gain calibration: 4V, g(t) and g(ν) independent, stability
  - Pointing calibration and Deff: solve linear system of equations
  - Phase calibration: Cygnus, visibility directly, sin(phase)
  - Amplitude calibration:
    - Cross: Cygnus A
    - Auto: our foreground+LAB, data noise relationship
  - Estimate system temperature
  - Map-making using m-mode algorithm

# Summary: relationship among works

- Introduction
- High resolution foreground model
- 21cm extraction by filtering
- Lunar orbit interferometer imaging
- PAON-4 data analysis / Tianlai

21 cm




- Summary

# My publications

- An Imaging algorithm for a lunar orbit interferometer array

  **Qizhi Huang** et al. 2018. Published in AJ

- Extracting 21cm signal by frequency and angular filtering

  **Qizhi Huang** et al. 2018. Published in RAA

- A high-resolution self-consistent whole sky foreground model
  **Qizhi Huang** et al. 2019. Published in Sci.China Phys.Mech.Astron.

- Design, operation and performance of the PAON4 prototype transit interferometer

  Réza Ansari et al. (I am a co-author) 2019. Submitted to MNRAS


Thank you
TAGS MITSU

# Need to be improved

- High resolution foreground model
  - More radio source data from different surveys at different frequencies
  - High resolution diffuse model
- 21cm extraction by filtering
  - Try more different methods
  - Consider noise correlation
- Lunar orbit interferometer imaging
  - Polarization
  - Realistic beam pattern
  - Different orbits of satellites
- PAON-4 data analysis
  - More data to reconstruct larger sky map
  - More data to reach higher sensitivity
  - More precise calibration method

# Faint source model

- Rayleigh-Lévy random walk
  - probability distribution

$$P(\Theta > \theta) = \begin{cases} \left(\frac{\theta}{\theta_0}\right)^{-\gamma} , & \theta \geq \theta_0 \\ 1 , & \theta < \theta_0 \end{cases}$$

  - From two-point correlation function of source survey:

$\theta_0 = 6'$ $\gamma = 0.8$ (Overzier et al. 2003)

- Detail steps
  - (1) First step: arbitrary position
  - (2) Choose random (n, θ), and uniform random u, if u>P, place
  - (3) Repeat (2) until all sources are placed
  - (4) Set flux densities to each sources, according to the differential count **n(S)**

# Normalized differential count n(S)

- Definition

$$S^{5/2}n(S) \equiv \frac{1}{4\pi}\int_0^{\infty} S^{5/2}\eta(S,z)\mathrm{d}z$$

where $\eta(S,z)$ is the number of radio sources per unit flux densitiy per unit redshift

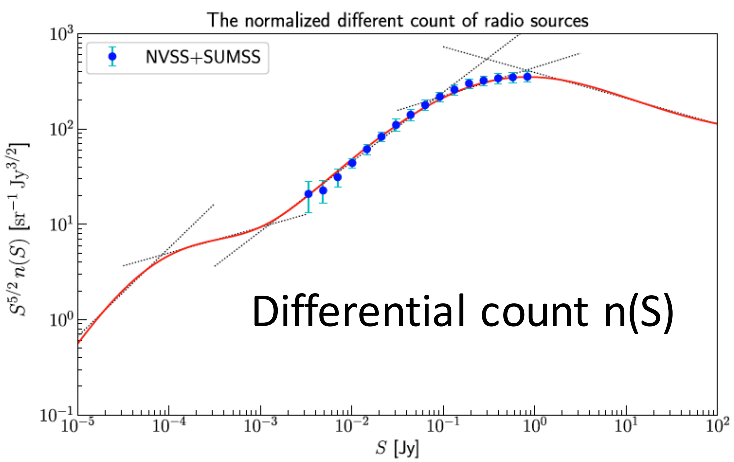
$$\eta(S,z) = \rho(L,z)A\frac{\mathrm{d}L}{dS}\frac{\mathrm{d}r}{\mathrm{d}z} = \frac{cA^2(1+z)^{1+\alpha}\rho(L,z)}{H_0\sqrt{\Omega_\Lambda+\Omega_m(1+z)^3}}$$

- Star-forming galaxy

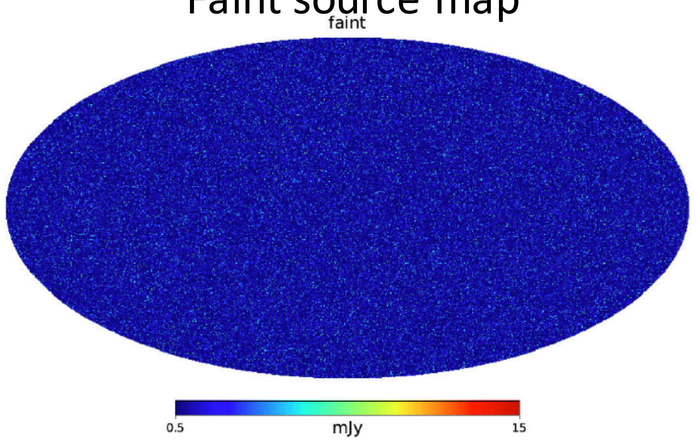
$$\rho_m(L) = C\left[\frac{L}{L_*}\right]^{1-a}\exp\left\{-\frac{1}{2}\left[\frac{\log_{10}\left(1+\frac{L}{L_*}\right)}{\sigma}\right]^2\right\}$$

- Radio loud AGN

$$\log_{10}(\rho_m) = Y - \frac{3}{2}\log_{10}(L) - \sqrt{B^2+\left[\frac{\log_{10}(L)-X}{W}\right]^2}$$




Faint source map



# Confusion limit

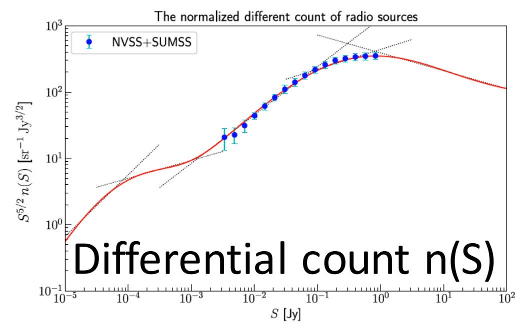
- Radio sources with flux densities below the confusion limit are undetected.

$$N(S_{\rm B}) = \int_{\Omega_{\rm B}} \frac{n\left(\frac{S_{\rm B}}{B(\theta,\varphi)}\right)}{B(\theta,\varphi)} {\rm d}\Omega$$

$$\sigma_{\rm c}^2 = \int_0^{S_{\rm c}} S_{\rm B}^2 N(S_{\rm B})\, {\rm d}S_{\rm B}$$

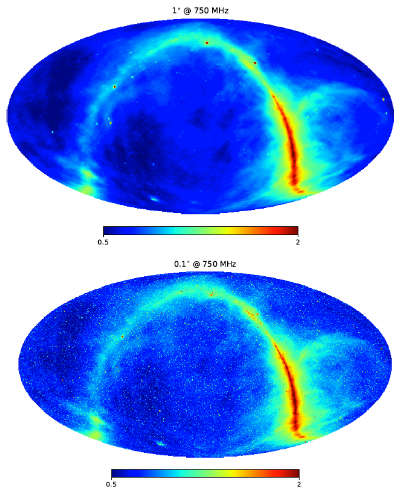
- Our formula

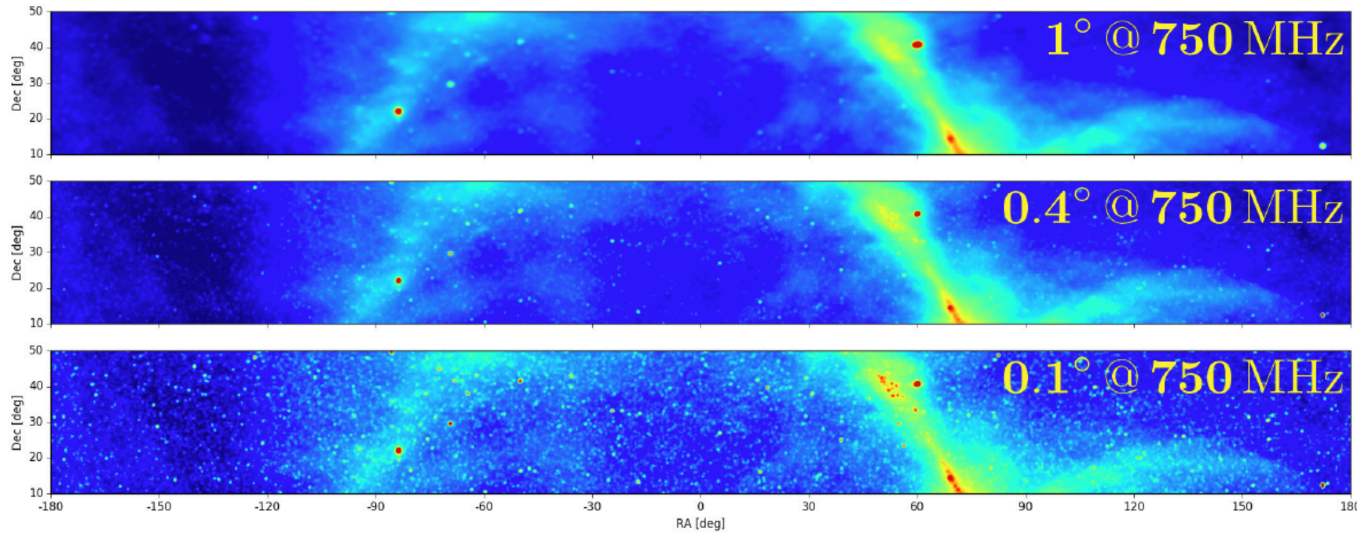
$$S_c = \begin{cases} 0.05\,\mathrm{mJy} \cdot Q^{3.51} \left(\frac{\theta_0}{\mathrm{arcmin}}\right)^{3.51} \left(\frac{\nu}{1.4\mathrm{GHz}}\right)^{-2\alpha}, & \theta_0 \le 0.25\ \mathrm{arcmin} \\ 0.07\,\mathrm{mJy} \cdot Q^{1.53} \left(\frac{\theta_0}{\mathrm{arcmin}}\right)^{1.53} \left(\frac{\nu}{1.4\mathrm{GHz}}\right)^{-2\alpha}, & 0.25 < \theta_0 \le 1.13\ \mathrm{arcmin} \\ 0.01\,\mathrm{mJy} \cdot Q^{2.56} \left(\frac{\theta_0}{\mathrm{arcmin}}\right)^{2.56} \left(\frac{\nu}{1.4\mathrm{GHz}}\right)^{-2\alpha}, & 1.13 < \theta_0 \le 7.18\ \mathrm{arcmin} \\ 0.3\,\mathrm{mJy} \cdot Q^{1.63} \left(\frac{\theta_0}{\mathrm{arcmin}}\right)^{1.63} \left(\frac{\nu}{1.4\mathrm{GHz}}\right)^{-2\alpha}, & 7.18 < \theta_0 \le 30\ \mathrm{arcmin} \\ 3\,\mathrm{mJy} \cdot Q^{1.18} \left(\frac{\theta_0}{\mathrm{arcmin}}\right)^{1.18} \left(\frac{\nu}{1.4\mathrm{GHz}}\right)^{-2\alpha}, & \theta_0 > 30\ \mathrm{arcmin} \end{cases}$$




- Test our confusion limit formula
  - ASKAP EMU survey, 1.3 GHz, 10 arcsec, Sc=5 μJy
    our: 4.95 μJy
  - MWA, 512 antenna tiles, longest baseline 5 km, 300 MHz, Sc=2.3 mJy
    our: 2.6 mJy

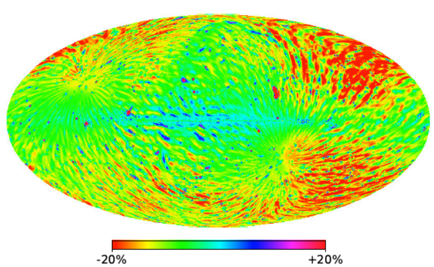
# Total foreground



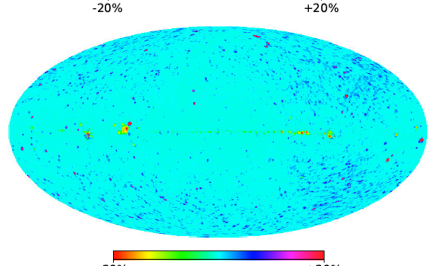




- Compare with GSM

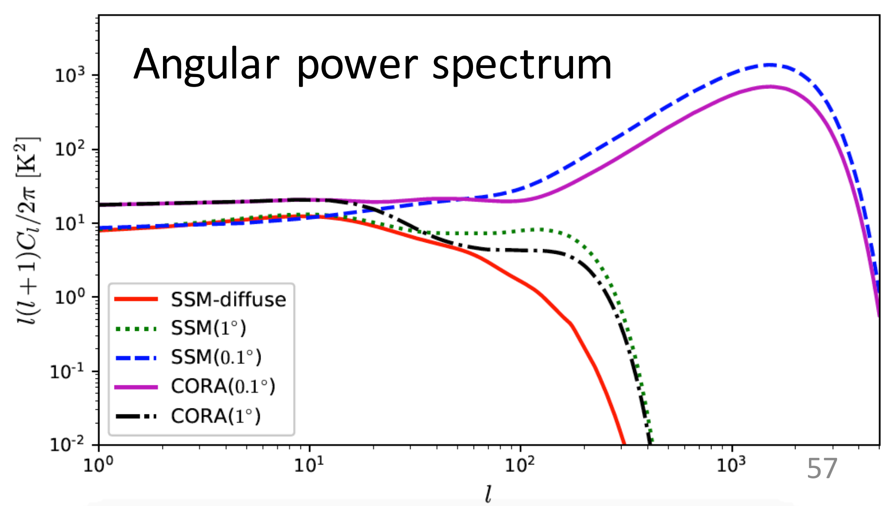
GSM




Our model




- Compare with CORA



# Proposed 21cm extraction methods

- Polynomial fitting

$$\ln T(\widehat{\boldsymbol{r}}_i, \nu) = \alpha(\widehat{\boldsymbol{r}}_i) + \beta(\widehat{\boldsymbol{r}}) \ln \frac{\nu}{\nu_*} + \gamma(\widehat{\boldsymbol{r}}_i)\left(\ln \frac{\nu}{\nu_*}\right)^2$$

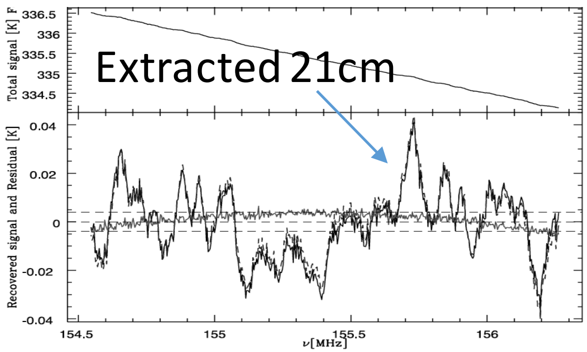
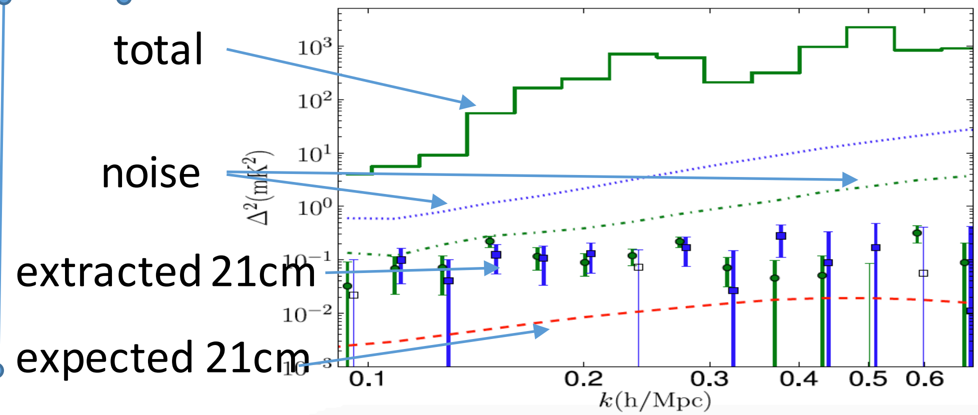



FIG. 3.—Spectrum in a single pixel before and after foreground cleaning. The

- SVD

$$\mathbf{C}_{AB} = (\mathbf{W}_A \circ \mathbf{M}_A)(\mathbf{W}_B \circ \mathbf{M}_B)^{\mathrm{T}}/N_\theta$$

- Cross different days => low noise




- PCA

$$\mathbf{C} \equiv \frac{1}{n_{\rm pix}} \sum_{i=1}^{n_{\rm pix}} \mathbf{y}_i \mathbf{y}_i^{\rm t} \qquad \mathbf{R} = \mathbf{P}\Lambda\mathbf{P}^{\rm t}$$

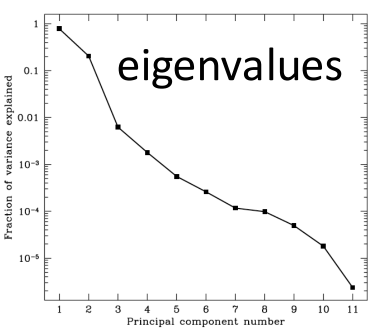
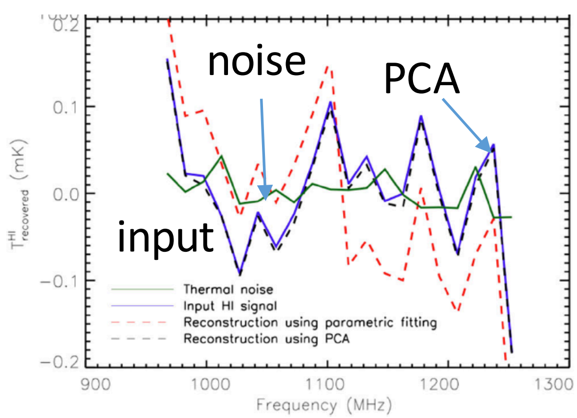



Figure 4. The eigenvalues $\lambda_i/11$ for the 11 principal components, which can be interpreted as the fraction of the total variance at the 11 frequencies that each component explains.

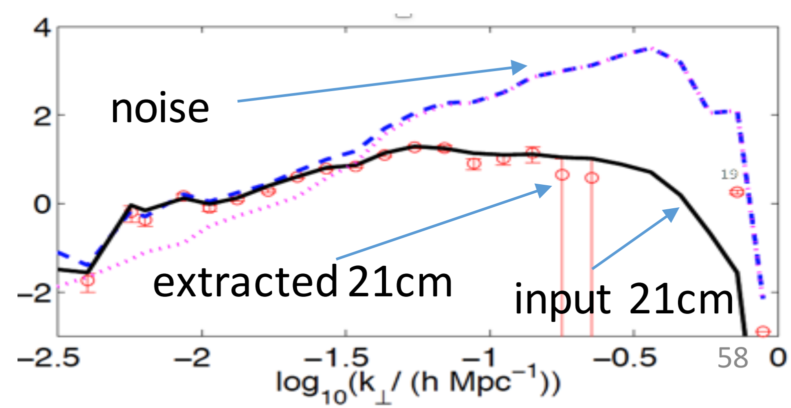



the recovered H I signal from parametric fitting (red dashed line) and from PCA (black dashed line). We show the input H I (blue solid line) and the thermal noise (green solid

- ICA: whitened (uncorrelated)

$$\tilde{x} = \hat{\mathbf{A}}' s \equiv \hat{\Lambda}^{-1/2}\,\hat{\mathbf{U}}\,\hat{\mathbf{A}}\,s \qquad s = \hat{\mathbf{W}}\,\tilde{x}$$

SVD to compute W inverse



# Simulated 21cm signal

- Cube box with $200^3$ voxels
  - Center redshift: 0.7755

$$D_A = 1068.95\ \mathrm{Mpc}\, h^{-1}$$

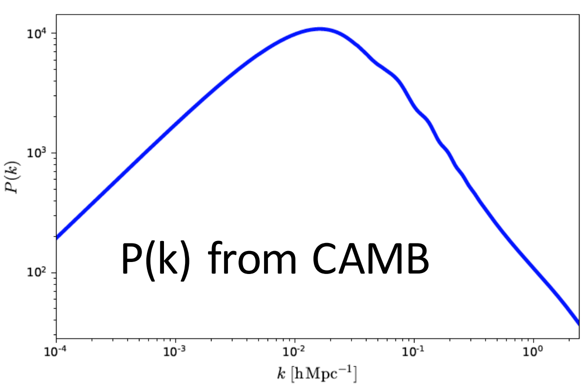
  - Frequency resolution: 0.1 MHz

$$\Delta D_c = 0.43\ \mathrm{Mpc}\, h^{-1}$$

  - Uniform voxel size:

$$\Delta\theta = \Delta D_c / D_A = 0.023^\circ$$

- Nonlinear power spectrum of dark matter from CAMB




- 21cm signal box
  - Gaussian random field

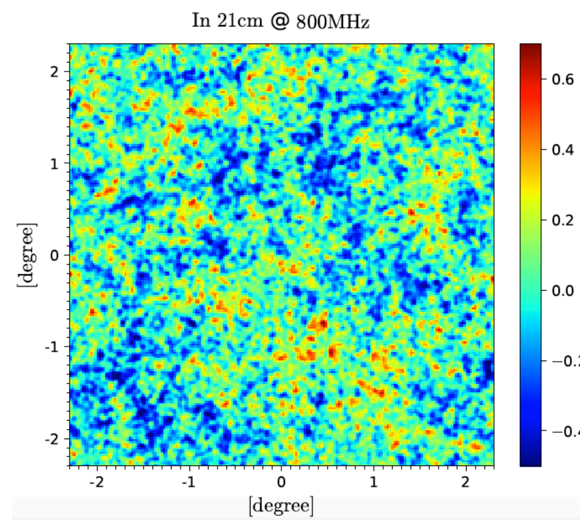
$$\delta(\vec{k}) = \sqrt{\frac{V}{2} P(k)} \left(a_k + i b_k\right)$$

$$\delta(\vec{x}) = \frac{1}{V} \sum \delta(\vec{k})\, e^{i\vec{k}\cdot\vec{x}}$$

  - 21cm temperature

$$\delta T_{21}(\vec{x}) = b_{\mathrm{HI}} \bar{T}_{21} \delta(\vec{x})$$

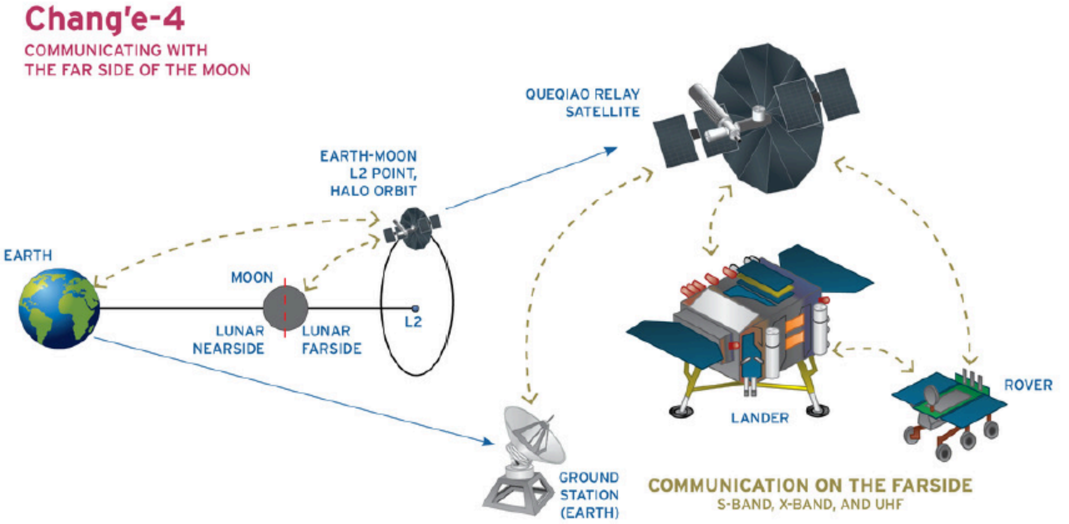
$$\Omega_{\mathrm{HI}} = 6.6 \times 10^{-4} \quad b_{\mathrm{HI}} = 0.70$$




simulated 21cm signal at 800 MHz

# China's Chang'e-4

- First soft landing on the far-side of the Moon (01/2019)
- Tracking and data relay satellite system (TDRSS), Queqiao
- Lunar orbit satellites:
  - Longjiang-1 (lost) and Longjiang-2 (05/2018)
  - Original objective: form a satellite interferometer
  - Whole sky spectrum, single satellite imaging



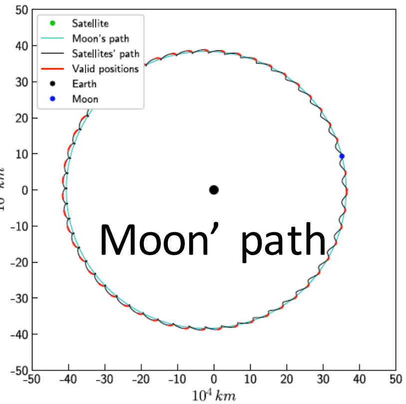
# Blockage effect

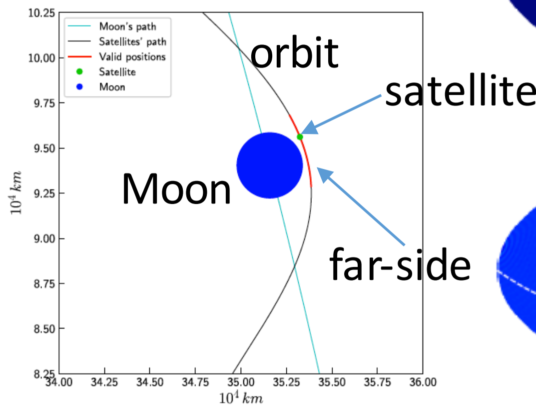
Can still image well

- Blockage angular size

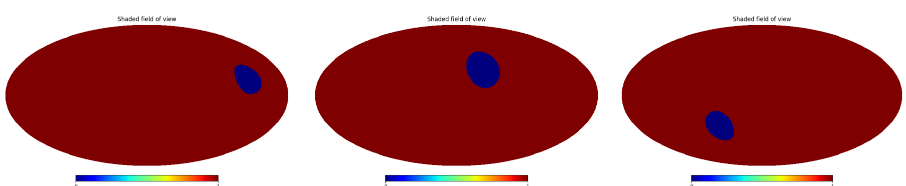
$$\theta_m \approx 2\arcsin\left(\frac{R_\mathrm{m}}{R_\mathrm{m}+h}\right)$$

- Valid observation time



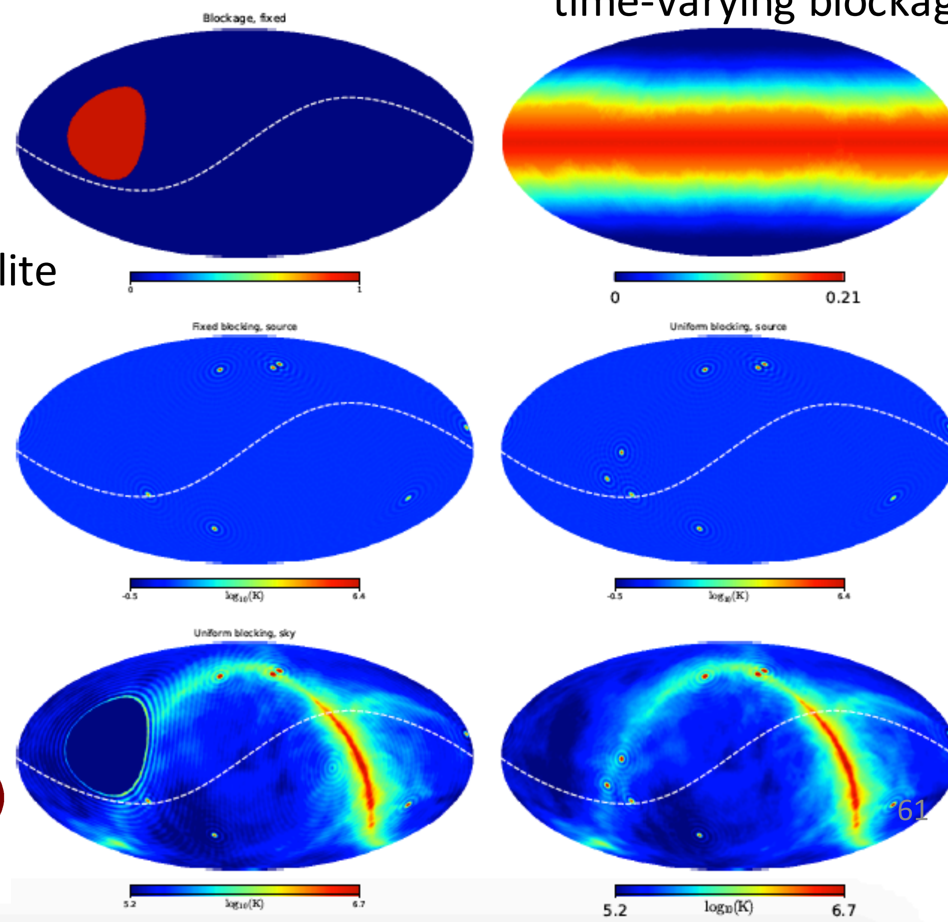




- Time-varying blocked sky




Test our algorithm fix blockage

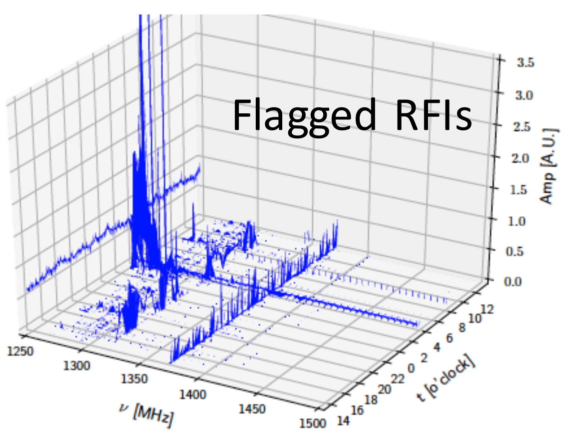
time-varying blockage



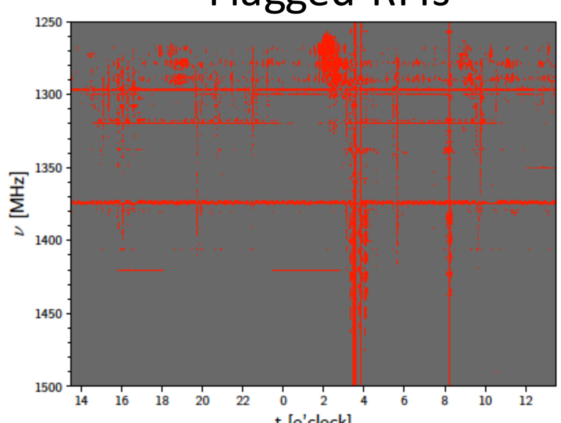
# RFI mitigation

- Proposed methods
  - Component decomposition methods
    - PCA, SVD: good if RFI repeated
    - But not good for stochastic RFI
  - Threshold methods
    - CUMSUM, VarThreshold
    - SumThreshold
  - For LOFAR and WSRT data
    - SumThreshold performs the best (Offringa et al. 2010)

- SumThreshold
  - Connected samples
    - 1, 2, 4, 8, 16 (experiential)
  - Thresholds $\geq \chi \cdot \mathrm{M} \cdot \sigma$
  - $\chi$
    - 4, 3, 2.5, 2, 1.8
  - Two rules
    - Replace, recompute




Flagged RFIs



# Pointing calibration and Effective diameter

- Use cross-correlation data
- Assume Gaussian beam
- Effective diameter

$$\frac{1}{\sigma_{ij}^2} = \frac{1}{2}\left(\frac{1}{\sigma_i^2} + \frac{1}{\sigma_j^2}\right) \Rightarrow D_{ij}^2 = \frac{D_i^2 + D_j^2}{2}$$

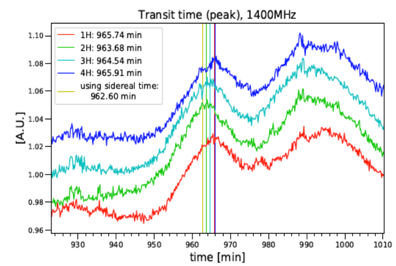
- Pointing

$$\theta_{ij} = \theta_i \left(\frac{D_i^2}{D_i^2 + D_j^2}\right) + \theta_j \left(\frac{D_j^2}{D_i^2 + D_j^2}\right)$$

- Solve

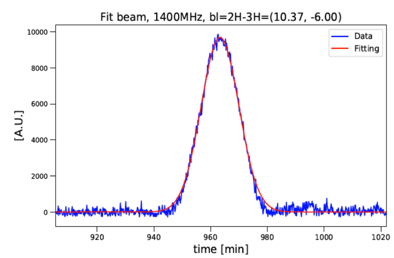
$$\begin{array}{ccccc} D_1^2 & + & D_2^2 & = & 2D_{12}^2 \\ D_1^2 & + & D_3^2 & = & 2D_{13}^2 \\ D_2^2 & + & D_3^2 & = & 2D_{23}^2 \\ & & \dots & & \\ D_i^2 & + & D_j^2 & = & 2D_{ij}^2 \end{array} \Rightarrow \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & \dots & 0 & \dots \\ 1 & 0 & 1 & \dots & 0 & \dots & 0 & \dots \\ 0 & 1 & 1 & \dots & 0 & \dots & 0 & \dots \\ & & & \dots & & & & \\ 0 & 0 & 0 & \dots & 1 & \dots & 1 & \dots \end{bmatrix} \begin{bmatrix} D_1^2 \\ D_2^2 \\ D_3^2 \\ \dots \\ D_i^2 \\ \dots \\ D_j^2 \\ \dots \end{bmatrix} = 2 \begin{bmatrix} D_{12}^2 \\ D_{13}^2 \\ D_{23}^2 \\ \dots \\ D_{ij}^2 \end{bmatrix} \Rightarrow Ax = B$$

- Not use auto-corr data



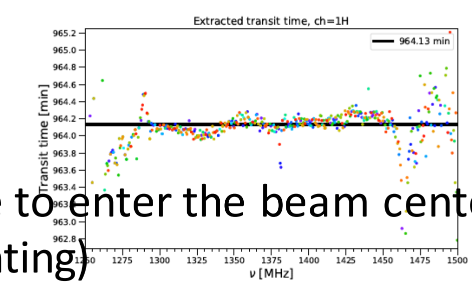

Peaks of CygA in Auto-corr

- Fit beam profile

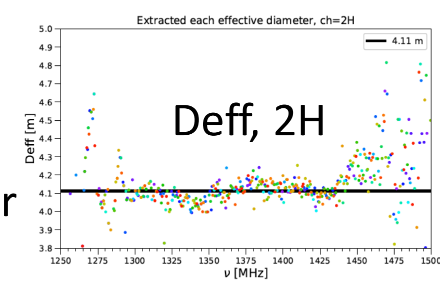



Peak of CygA in Cross-corr

- Deff: 4.22m, 4.11m, 4.23m, 4.36m
- Pointing error: 1.3°, 0.5°, 0.7°, 1.2°




Time to enter the beam center (pointing)



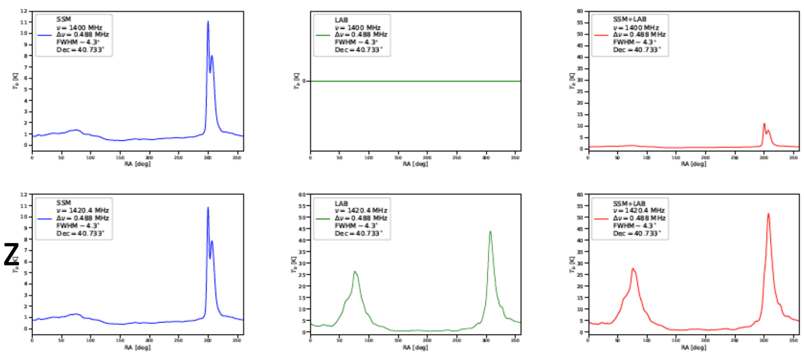
# Amplitude calibration

- Calibrate Auto-corr data
    - Our foreground model + LAB

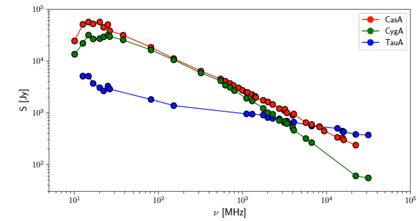
foreground Galacitc 21cm total

1400 MHz

1420.4 MHz

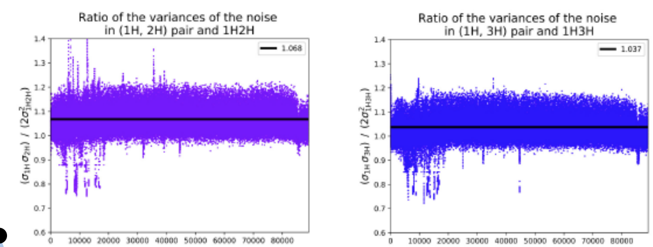


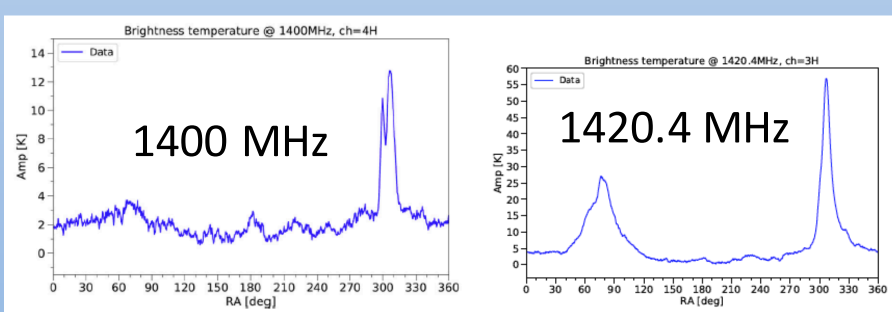
- Calibrate Cross-corr data




- Data noise ckeckin

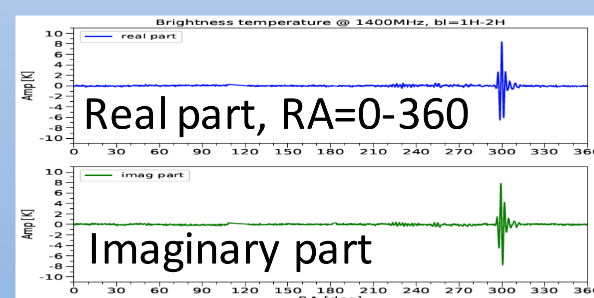
$$2\sigma^2_{\mathbf{C}_{XY}} = \sigma_{\mathbf{A}_X}\,\sigma_{\mathbf{A}_Y}$$




Auto-corr after calibration

- correlation



# Summary

- High resolution foreground model
  - Radio source model:
    - bright sources: NVSS+SUMSS, completeness, same surface density, spectral index
    - faint source: differential count, Rayleigh-Levy random walk
    - formula of comfusion limit: compare with achieved results
  - Free-free emission model: spectral index
  - Synchrotron emission model: preprocess 8 maps, spectral index
  - Compare with GSM, CORA

- 21cm extraction by filtering
  - 21cm signal simulation
  - Design a cascade of two Wiener filters: in frequency then in angular
  - Test two noise level: 200 mK and 2000 mK
  - Try with two inaccurate beam models

# Summary

- Lunar orbit interferometer imaging
  - <30 MHz, advantage
  - Whole sky FoV, symmetry problem => orbital plane precessio
  - Time-varying blockage, time-varying and noncoplanar baselines + whole sky FoV => no applicable imaging algorithm
  - Our imaging algorithm in spherical harmonic
  - Blockage effect, baseline distribution effect, sph harnomic v.s. pixel

- PAON-4 data analysis
  - RFI mitigation: surface fitting by minimum filter, SumThreshold
  - System gain calibration: 4V, g(t) and g(ν) independent, stability
  - Pointing calibration and Deff: solve linear system of equations
  - Phase calibration: Cygnus, visibility directly, sin(phase)
  - Amplitude calibration:
    - Cross: Cygnus A
    - Auto: our foreground+LAB, data noise relationship
  - Estimate system temperature
  - Map-making using m-mode algorithm

2019/10/18 PhD defense

Topics in 21-cm cosmology:

# Foreground models and their subtraction, map reconstruction for wide field of view interferometers and PAON-4 data analysis


Qizhi HUANG

Supervisors: Prof. Réza ANSARI Prof. Xuelei CHEN