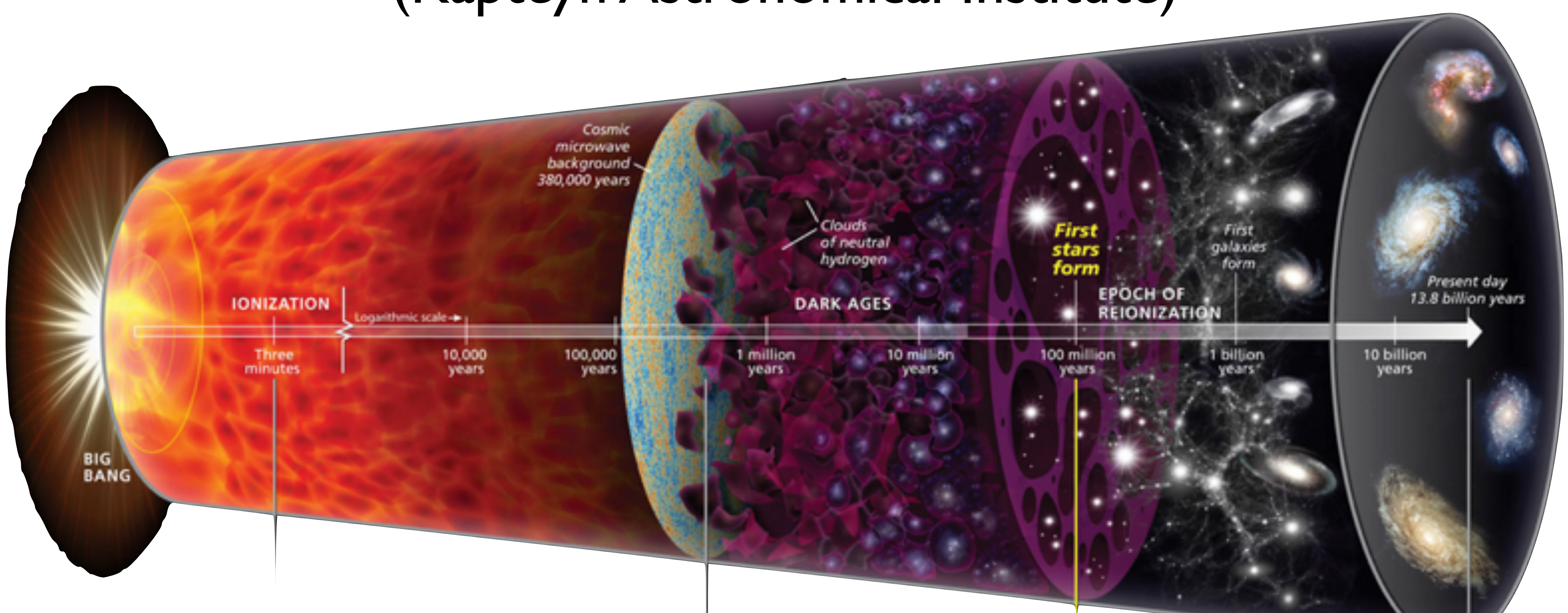
rijksuniversiteit groningen / faculteit wiskunde en natuurwetenschappen / kapteyn instituut

The LOFAR Reionization Key Science Project*: Current Status

*In memory of Ger de Bruyn

Léon V.E. Koopmans
(Kapteyn Astronomical Institute)



Oct.3, 2017, IAU333, Dubrovnik, Croatia

The Low Frequency Array

LOFAR is now a European telescope with its core in the Northern Netherlands, developed by ASTRON+Dutch Universities

(ILT Members: Netherlands, Germany, UK, France, Sweden, Poland, Ireland; interests in Italy, Spain, Austria+Ukraine)

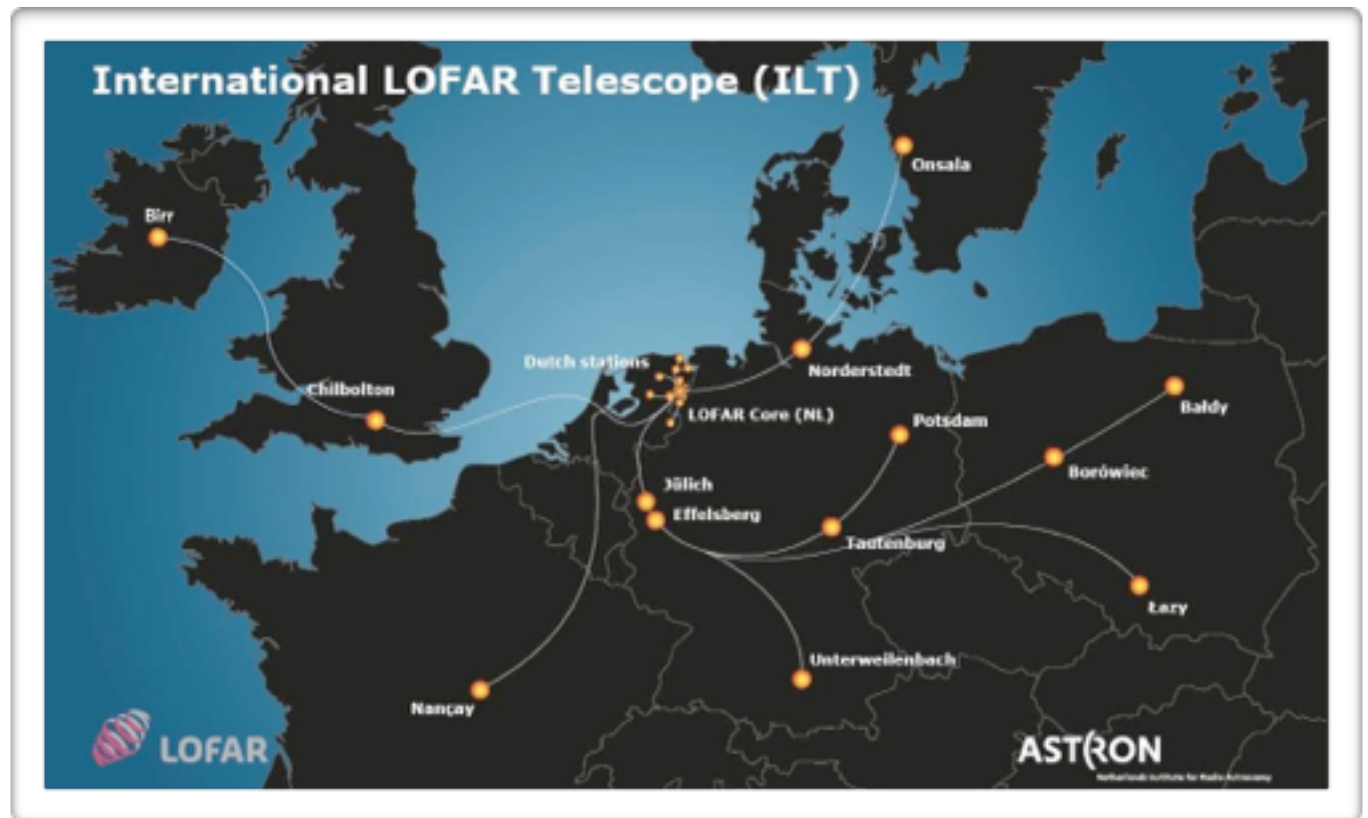
Core	3 km	(2x)24	stations
NL	80 km	14	stations
Europe	>1000 km	12	stations

Stations have 24 – 48 – 96 antennas/tiles

Principle of Aperture Synthesis
Array resolution: sub-arcsec to degrees

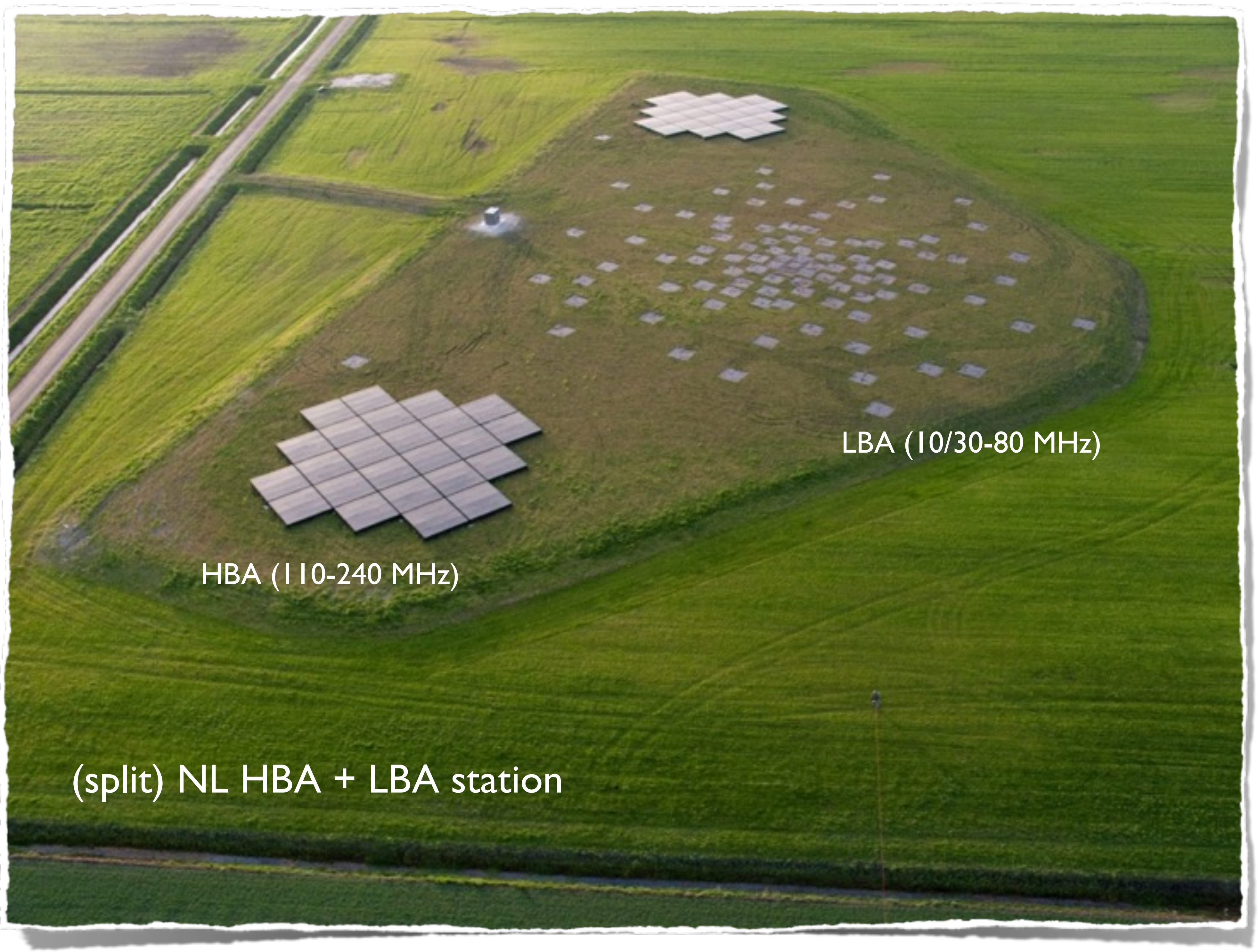
Pulsars: tied-array(s), (in)coherent sums

Sensitivity (8h, 4 MHz, Core/all NL stations)
@ 60 MHz ~ 6.2/3.9 mJy (LBA)
@ 150 MHz ~ 310/240 μJy (HBA)



van Haarlem et al. 2013

The Low Frequency Array



The Low Frequency Array

LOFAR from Space

LOFAR from Space

EoR windows & Cycles 0-6+8-9 Observations (since 1-Dec-2012)

- Night-time observing, elevation > 50°
- Frequency range 115-190 MHz (Cycle 6: 2-3 beams x 32MHz; Cycle 8-9: 7 beams x 12 MHz on NCP→"Fast track")
- Time/spectral resolution: 2s, 3.2kHz
- Raw data volume: 20 - 70 TB / night

Currently 1st stage processing ongoing (RFI flagging, averaging, initial calibration, imaging)

✓ ~1700 hrs on NCP
✓ ~1100 hrs on 3C196
✓ ~200+520 hrs planned/requested on NCP

>4 PB on disk

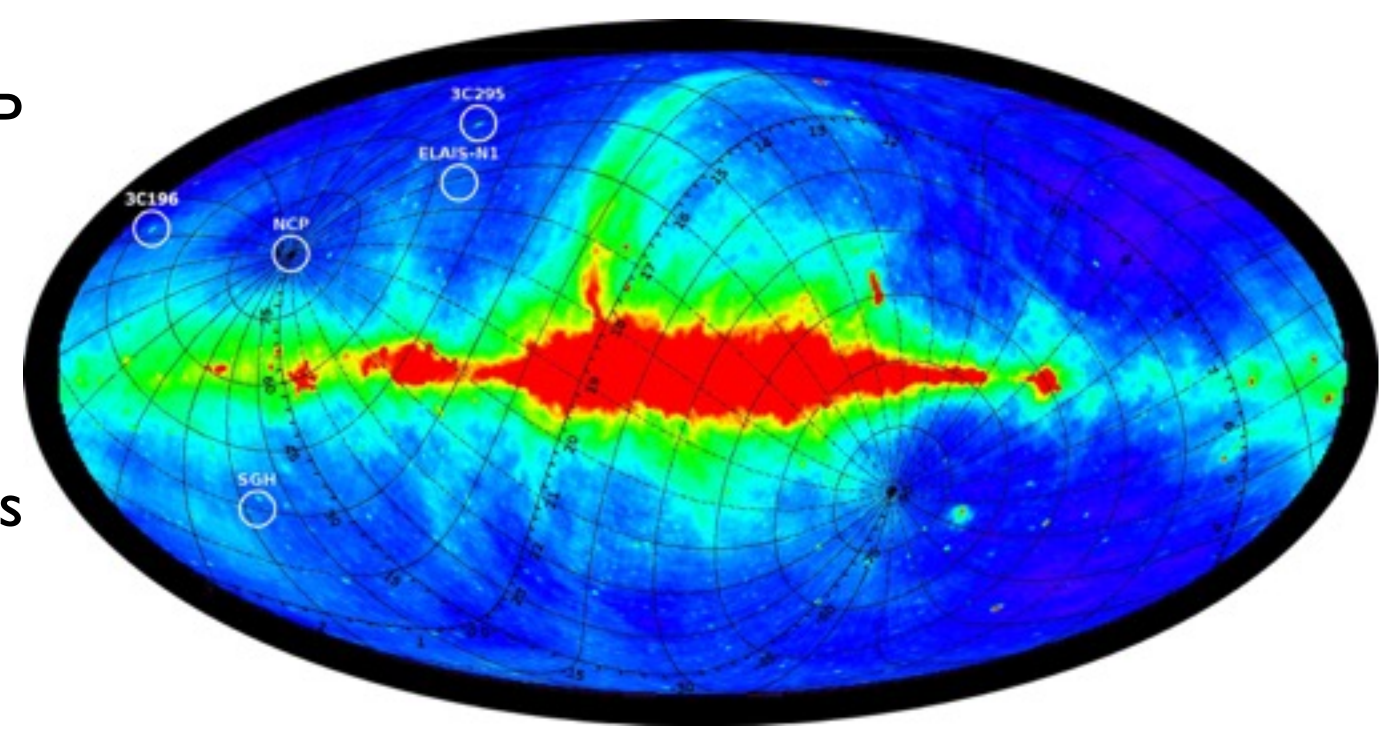
- NCP: constant beam, all-year observable
- 3C196: bright, compact, wintertime

- 2-3 other windows for various other projects

LOFAR spectral capabilities:

- 8-bit mode 488 sub-bands
- 1sub-band = 0.195 MHz
- 96 MHz total bandwidth

One sub-band can have up to 256 ch. We opted to store 64 ch. max. We analyse 3-ch. data (~60kHz).



LOFAR EoR data volume, products and formats

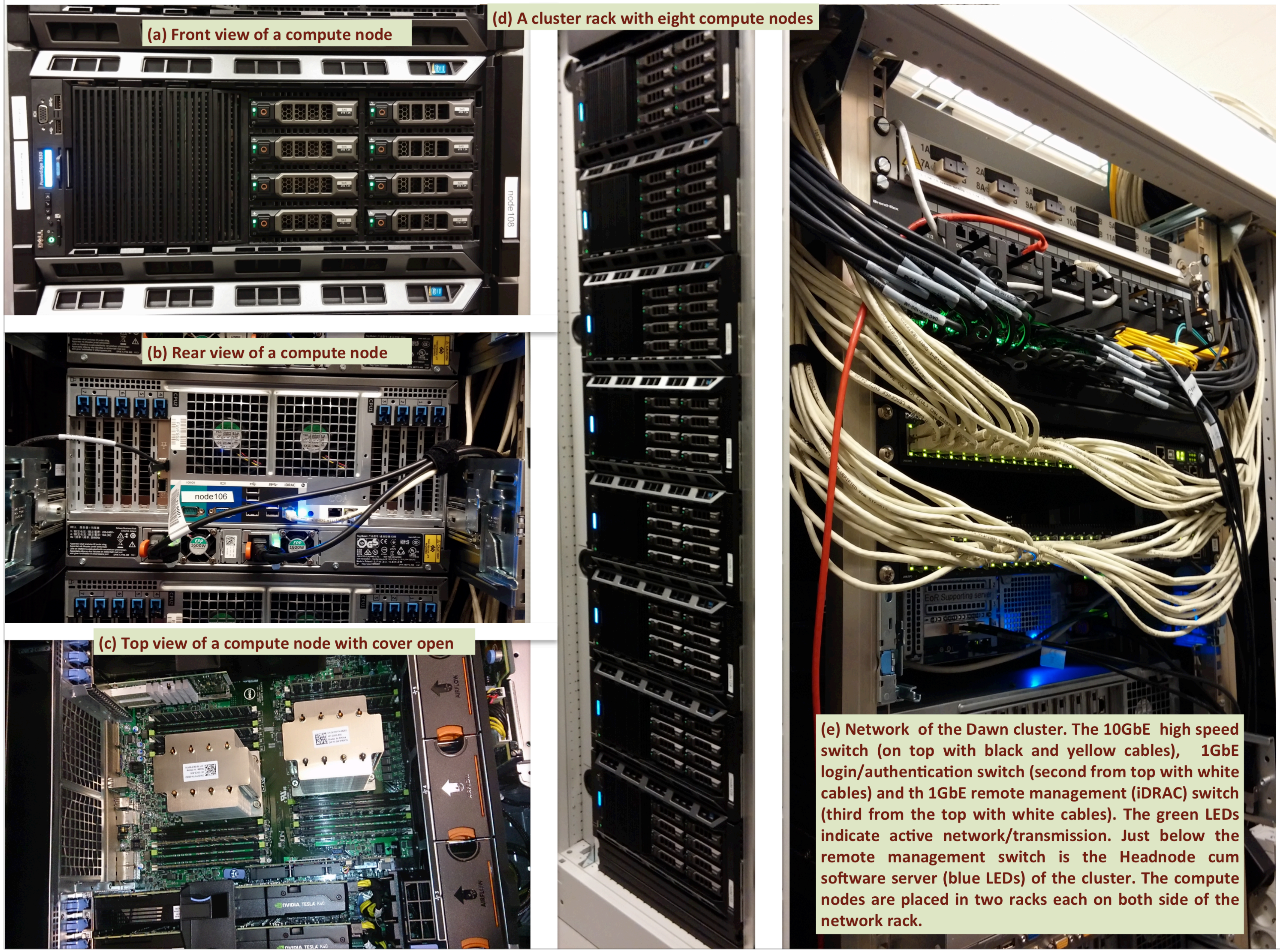
- Measurement Sets: raw, data format 3.2 kHz - 2s ~50 TB/12hr

- Processed data sets/formats:
  15ch/sb = 12 kHz - 2s (12 kHz = 24 km/s at 150 MHz, 21cm forest study)
  3ch = 60 kHz - 2s (currently sufficient for 21-cm EoR PS analyses; 10-15 TB)
  1ch = 180 kHz - 10s (fast starting models)

- Image cubes:
  20x20 deg, 2'' pixels, 6'' PSF 36k x 36k x 380 (488) => ~ 1 TB total (I)
  restored, apparent flux ➠ science analysis (e.g. sky-model building)
  6x6 deg, 40'' pixels, 3' PSF 512 x 512 x 380 (488) => ~ 1 GB (IQUV)

- Residual visibilities [after sky-model subtraction] [see talk by Mertens on Friday]
  - to use in ML inversion
  - to use in Foreground Fitting
  - to use in Power Spectrum estimation

LOFAR EoR-KSP GPU HPC Cluster: *“Dawn”*



(a) Front view of a compute node

(b) Rear view of a compute node

(c) Top view of a compute node with cover open

(d) A cluster rack with eight compute nodes

(e) Network of the Dawn cluster. The 10GbE high speed switch (on top with black and yellow cables), 1GbE login/authentication switch (second from top with white cables) and th 1GbE remote management (iDRAC) switch (third from the top with white cables). The green LEDs indicate active network/transmission. Just below the remote management switch is the Headnode cum software server (blue LEDs) of the cluster. The compute nodes are placed in two racks each on both side of the network rack.

32 nodes in four 19” racks

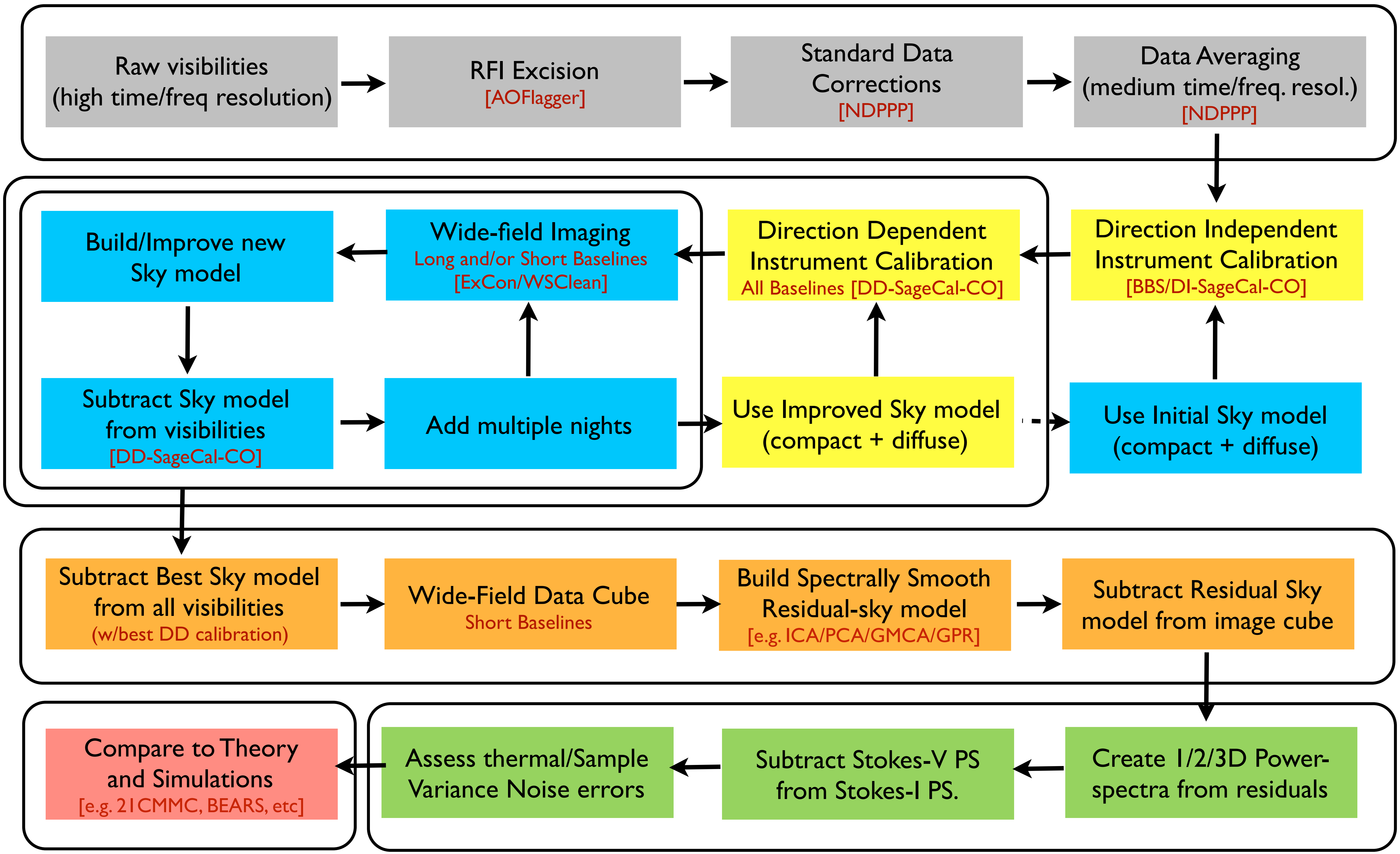
32x4 GPU (K40)
32x48 (HT) cores

0.6/0.2 Petaflops in GPU power

For details:

Pandey etal, ASTRON Newsletter Dec 2015

LOFAR EoR Data-Processing Flow Diagram



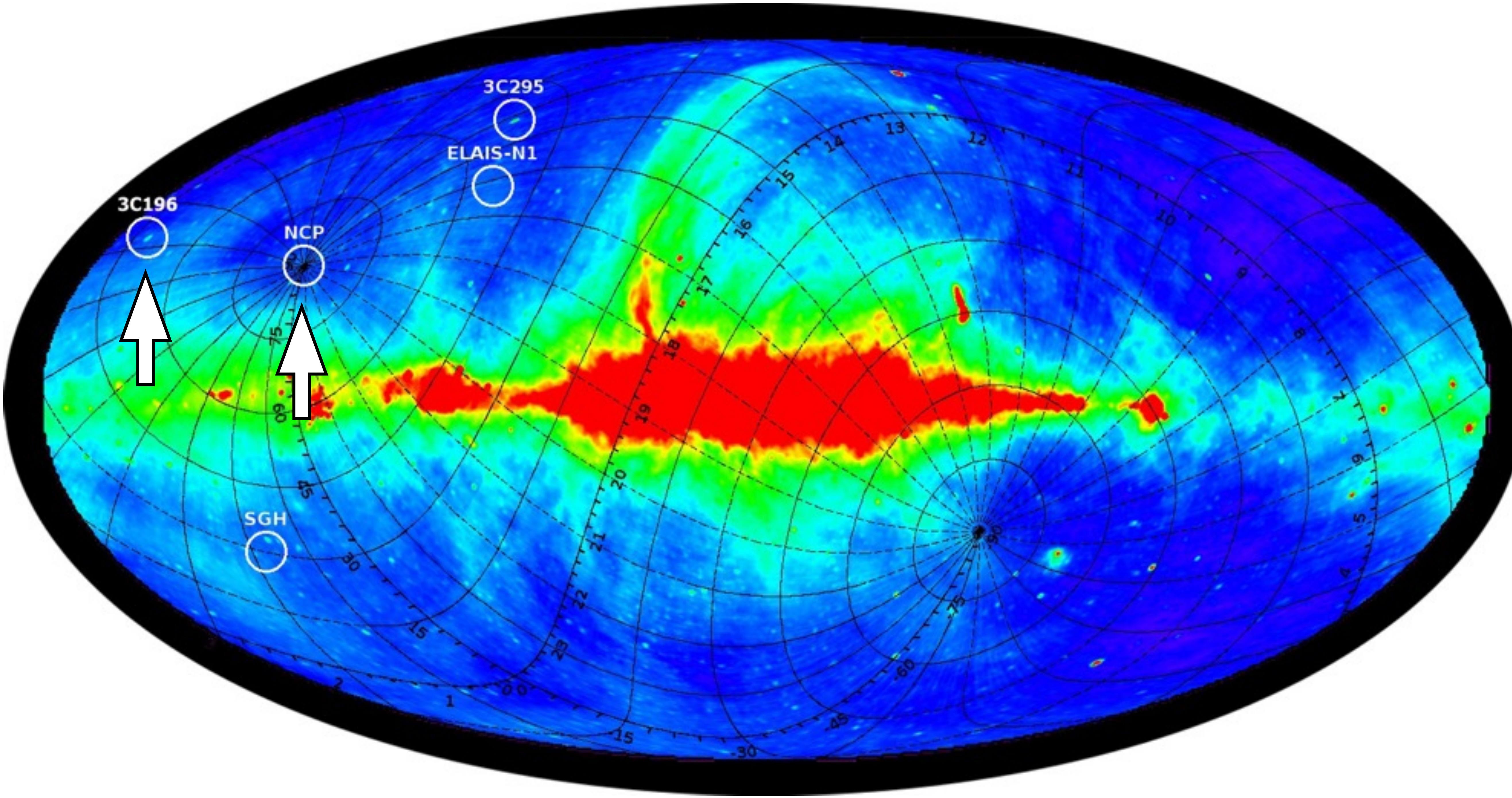
rijksuniversiteit groningen / faculteit wiskunde en natuurwetenschappen / kapteyn instituut

LOFAR EoR Upper Limits on the 21-cm Power-Spectrum during Reionization

UPPER LIMITS ON THE 21-CM EPOCH OF REIONIZATION POWER SPECTRUM FROM ONE NIGHT WITH LOFAR

A.H. Patil[1], S. Yatawatta[1,2], L.V.E. Koopmans[1], A.G. de Bruyn[2,1], M. A. Brentjens[2], S. Zaroubi[1,11], K. M. B. Asad[1], M. Hatef[1], V. Jelić[1,8,2], M. Mevius[1,2], A. R. Offringa[2], V.N. Pandey[1], H. Vedantham[9,1], F. B. Abdalla[7,13], W. N. Brouw[1], E. Chapman[7], B. Ciardi[4], B. K. Gehlot[1], A. Ghosh[1], G. Harker[3,7,1], I. T. Iliev[10], K. Kakiichi[4], S. Majumdar[12], M. B. Silva[1], G. Mellema[5], J. Schaye[6], D. Vrbanec[4], S. J. Wijnholds[2]

LOFAR EoR Deep Fields



We currently focus on two deep windows: NCP and 3C196
This talk: only NCP results.

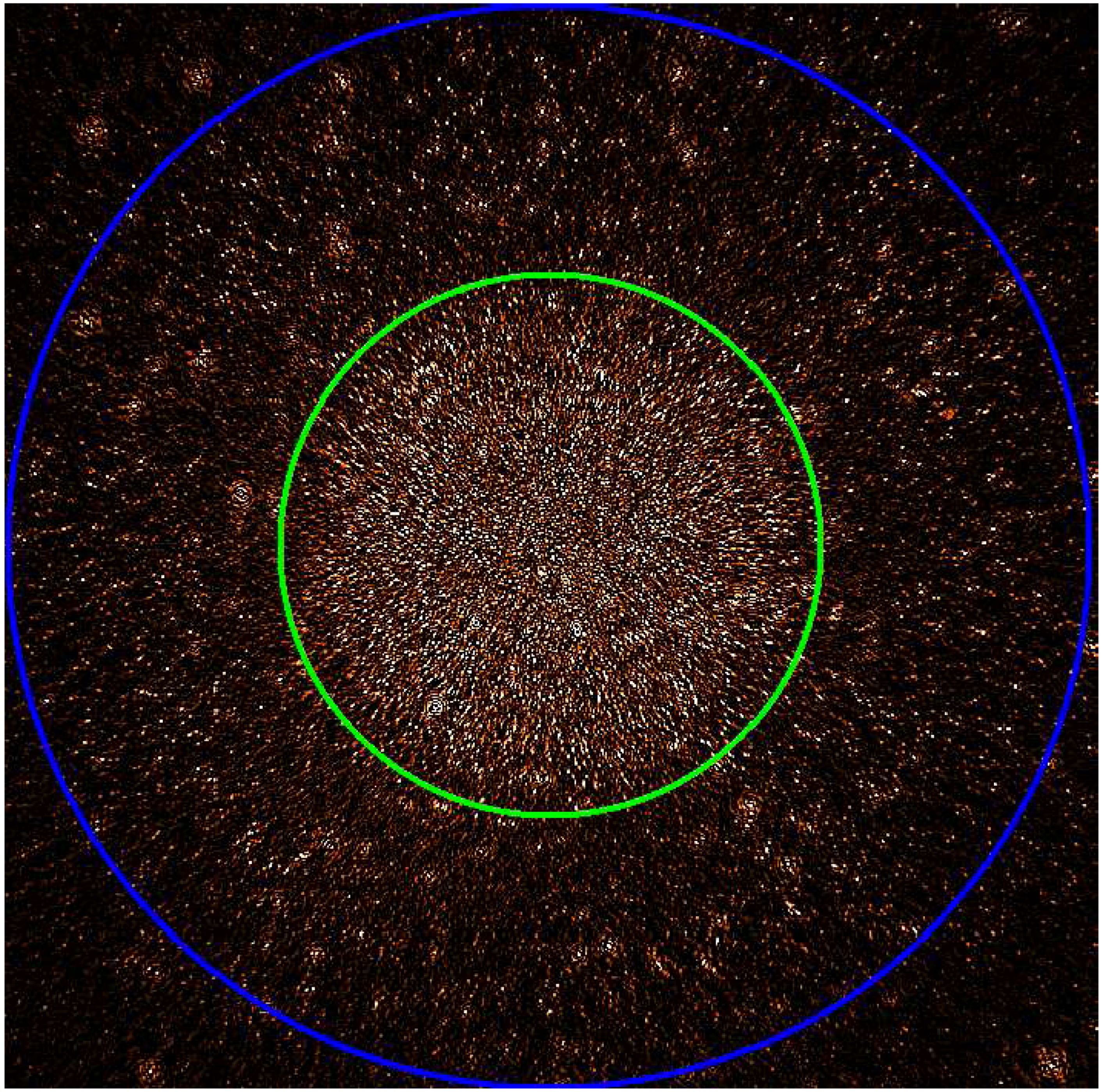
Deep Continuum Image of the NCP

Confusion limited images:

- T_{int}~100h, BW=60 MHz
- 20° x 20° ; 3' FWHM PSF
- 7.2 Jy peak; ~ 30 μJy noise

Note that this image is the sky residual, i.e. not subtracted, emission. 20,000 bright(est) sources were already removed.

All of this emission, should be spectrally smooth, otherwise one would not be able to detect the EoR 21-cm signal.



First null 10 deg. diameter, second null 20 deg. diameter

First power-spectrum results in April 2017

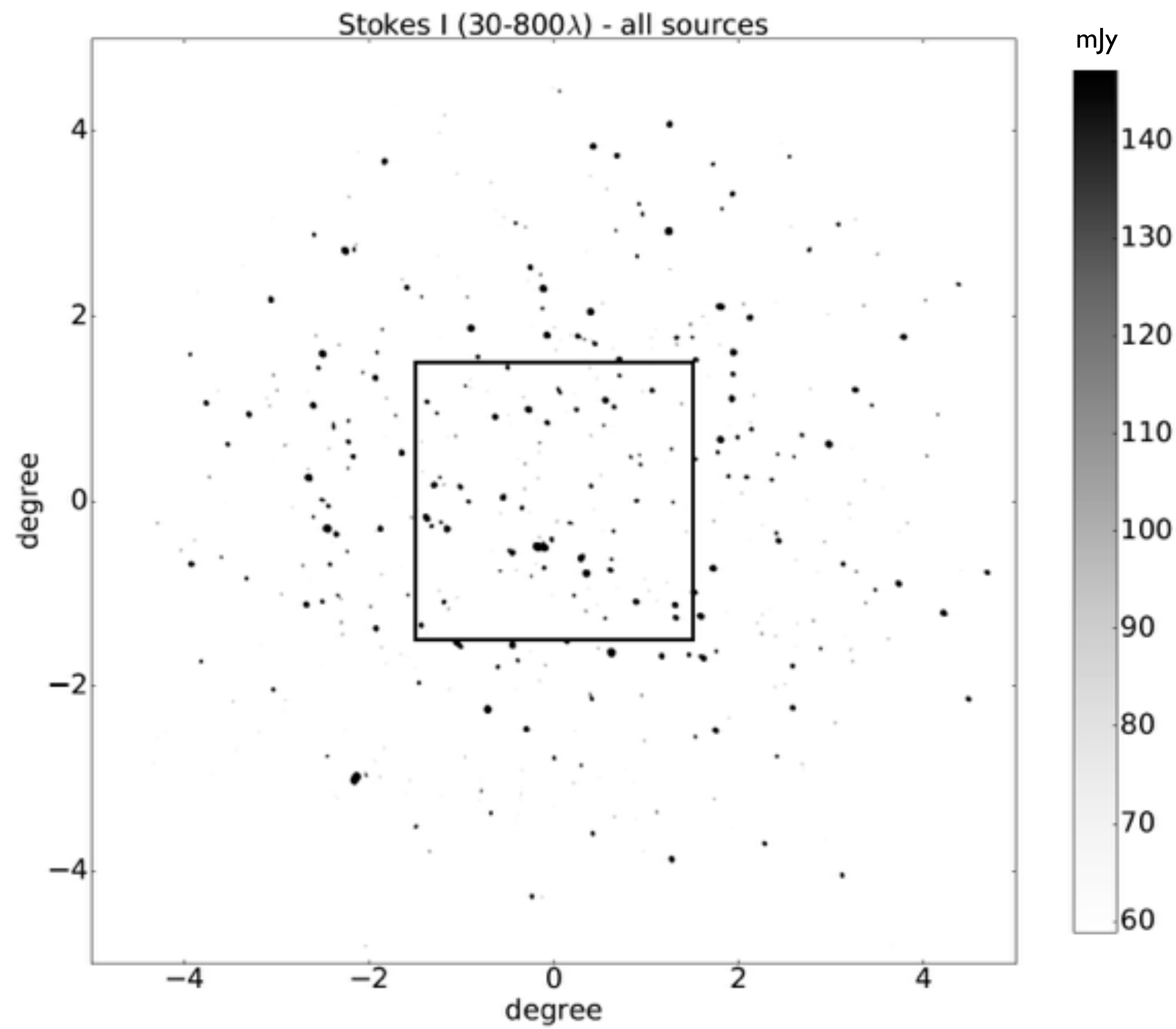
NCP Observations

Results presented today are based on a single 13-hr run
taken 2nd Nov. 2013 @ 17:20:01 (UTC)

Observational and correlator set-up

Phase Centre α, δ	$0^h, +90°$	J2000
Minimum frequency	115.039	MHz
Maximum frequency	189.062	MHz
Target bandwidth	74.249	MHz
Antenna fields	48 / 13	CS / RS
Data size (488 channels)	50	Tbyte
Sub-band (SB) width	195.3125	kHz
Correlator channels per SB	64	
Correlator integration time	2	s
Channels per SB after averaging	1, 3, 3, 15	
Integration time after averaging	10, 10, 2, 2	s
Raw data volume L90490	61	Tbyte

Patil et al. (2017, ApJ)



A continuum (134.5-137.5 MHz) LOFAR-HBA image of 10x10 deg2 centred on the North Celestial Pole (NCP) field. Baselines between 30-800 were included. No sources have been subtracted and the image is partially cleaned. The 3x3d box delineates the area where we measure the power spectra. The bright source to the lower-left of the box is 3C61.1. The units are mJy/PSF. Right Ascension (RA) 00h is towards the bottom and increases clockwise.

NCP Analysis

Calibration of the data is by far the most complex and time-consuming part of process and has taken a decade to develop to its current state.

Parameter	Value	Comments
Sky-model components	~20,800	Compact
Flux-limit sky model	~50	mJy; Inside/Outside PB
Order P_n^S source spectra	3	Polynomial
Calibration directions	122	Source clusters
Calibration baselines	≥250	Wavelength
Order B_n^G gain regul.	3	Bernstein Polynomial
Solution interval	10 min	–
uv-grid cells	4.58×4.58	Wavelength
w-slices	128	–
EoR Imaging baselines	50-250	Wavelength
EoR Imaging FoV	$3^\circ \times 3^\circ$	–
EoR pixel size	$0.5' \times 0.5'$	–
EoR Imaging Resolution	~10'	FWHM
EoR Freq. Resolution	~60	kHz
Redshift range #1	7.9 – 8.7	
Freq. range	146.8 – 159.3	MHz
GMCA components	6/0	Stokes I/V.
Redshift range #2	8.7 – 9.6	
Freq. range	134.3 – 146.8	MHz
GMCA components	6/2	Stokes I/V.
Redshift range #3	9.6 – 10.6	
Freq. range	121.8 – 134.3	MHz
GMCA components	8/2	Stokes I/V.

Patil et al. (2017, ApJ)

Main points:

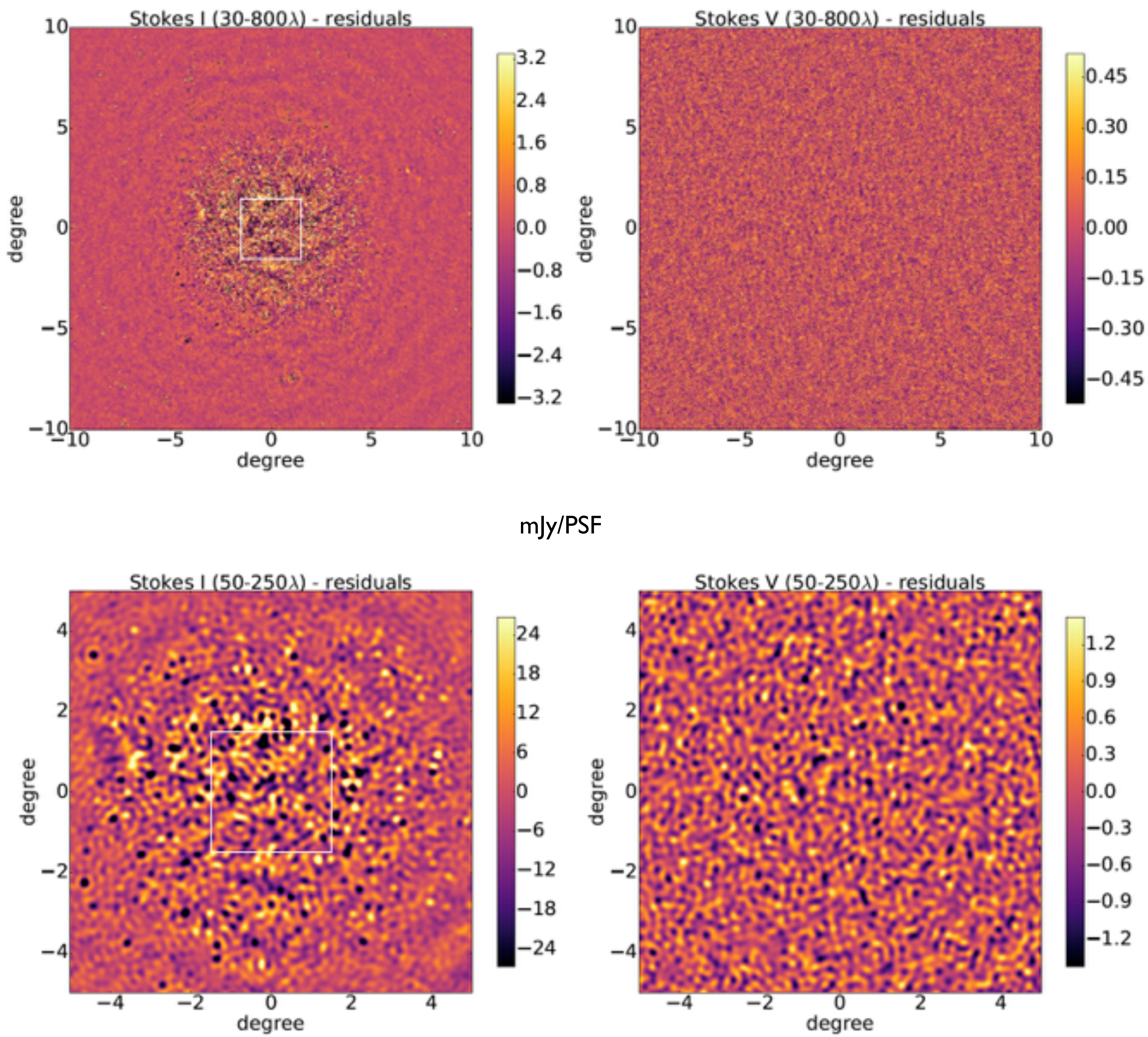
- First calibration step is direction independent (DI) and solutions are applied to the data.

- Calibration/sky model contains ~21,000 components
- All components have smooth spectra (optimised for)
- Regularised calibration in ~120 directions every 10 min
- Split calibration (>250λ) and imaging (<250λ) baselines sets.
  Reduces suppression diffuse emission but causes excess variance
- Three redshift ranges between z~8 and ~11.

Full-Stokes direction-dependent (DD) complex gains are solved for over the complete frequency range using a Bernstein basis set of 3rd order as regularisation term with quadratic penalty for deviations: necessary to remove bandpass ripples, etc. Some short-cut in regularisation were made [no longer done in 2017]

- Sky model is subtracted including DD calibration.
- Residuals only have DI calibration.

NCP Residuals after Sky-Model Subtraction



Top images shows 20x20d FoV in Stokes I (left) and V (right) with 3' resolution.

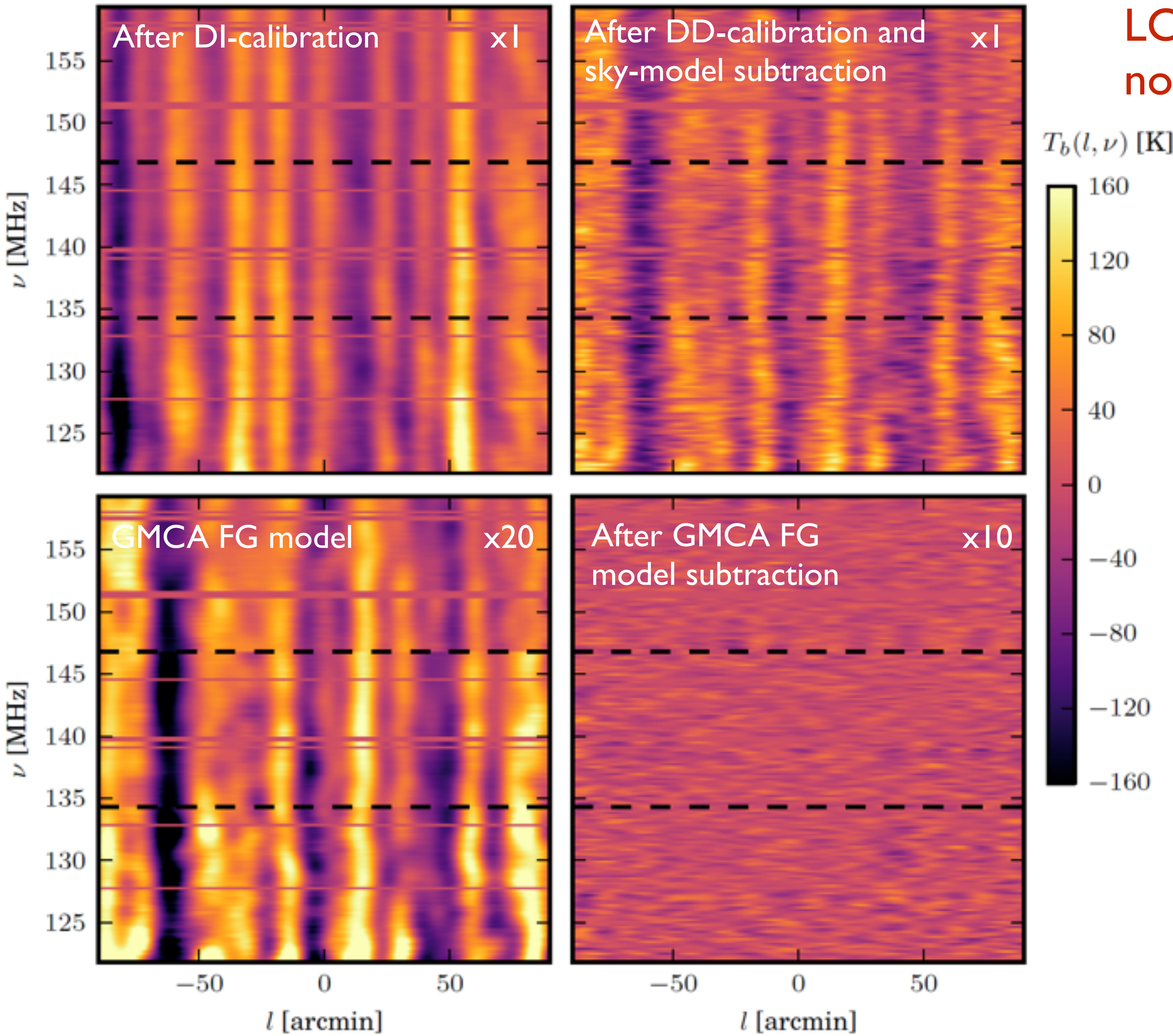
Stokes I shows the primary beam and is confusion limited; Stokes V is consistent with thermal noise to within ~5%.

White box in top of primary beam: region being analysed for power-spectrum

Bottom images shows 10x10d FoV in Stokes I (left) and V (right) with 10' resolution.

Patil et al. (2017, ApJ)

NCP Residuals after Sky-Model Subtraction



LOFAR EoR Approach:
not avoid but remove FGs.

Slice (3°) trough the cube in one spatial direction and frequency
(note scaling in brightness for some panels).

Note the resemblance between GMCA model (Chapman et al. 2013) and data after DD calibration.

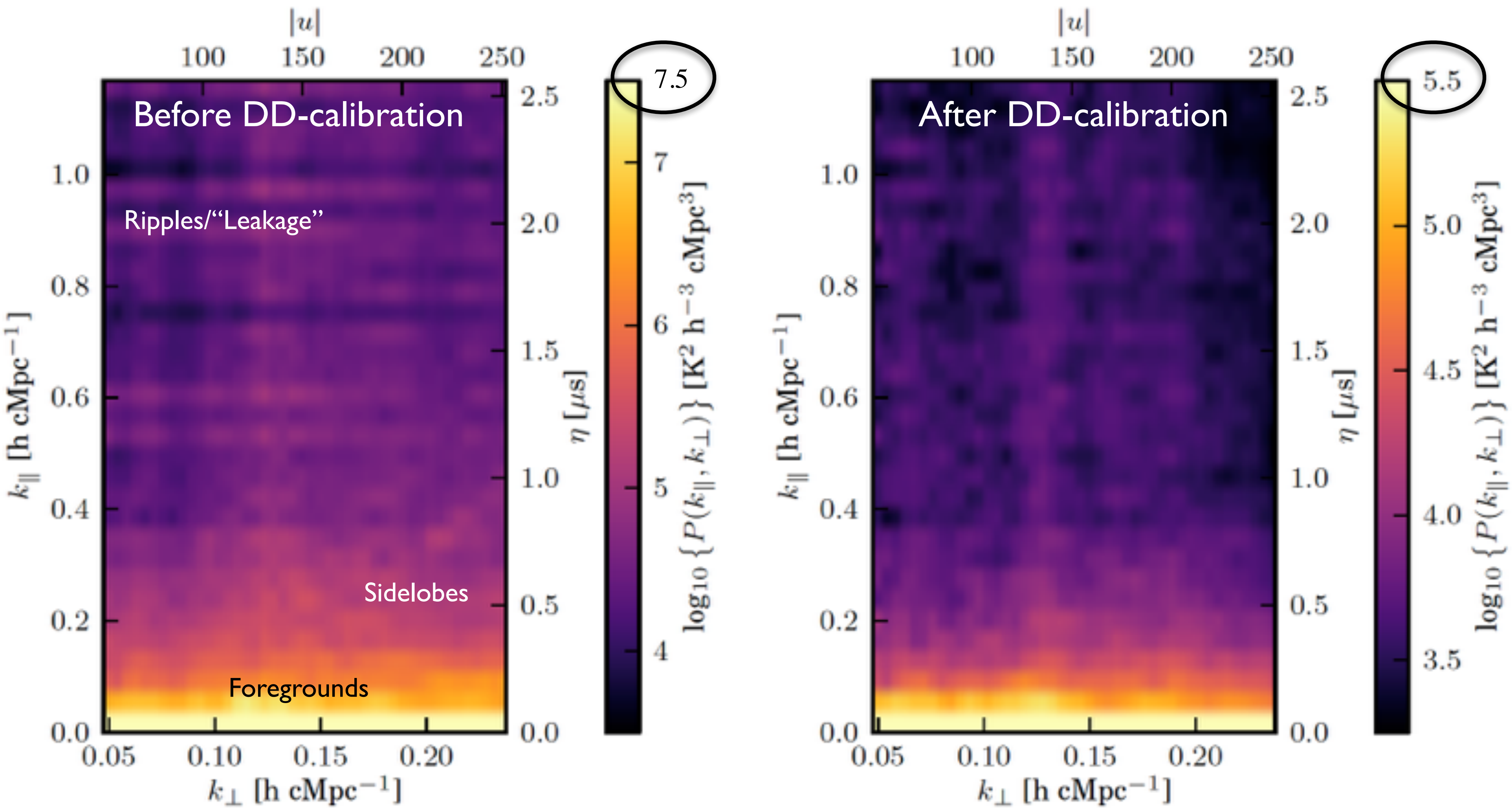
Dashed lines are the three redshift ranges: z=7.9-8.7, 8.7-9.6, 9.6-10.6

Patil et al. (2017, ApJ)

NCP Cylindrical Power-Spectra

The cylindrical power-spectra (PS) are a robust instrument to assess the quality of the data and potential residuals: foreground, side-lobes, band-pass ripples, ionosphere, etc.

We create PS via the FT of the residual data cubes, after turning each axis in to co-moving coordinates (Mpc; frequency ↔ distance), squaring the result and averaging in cylindrical annuli (spatially).



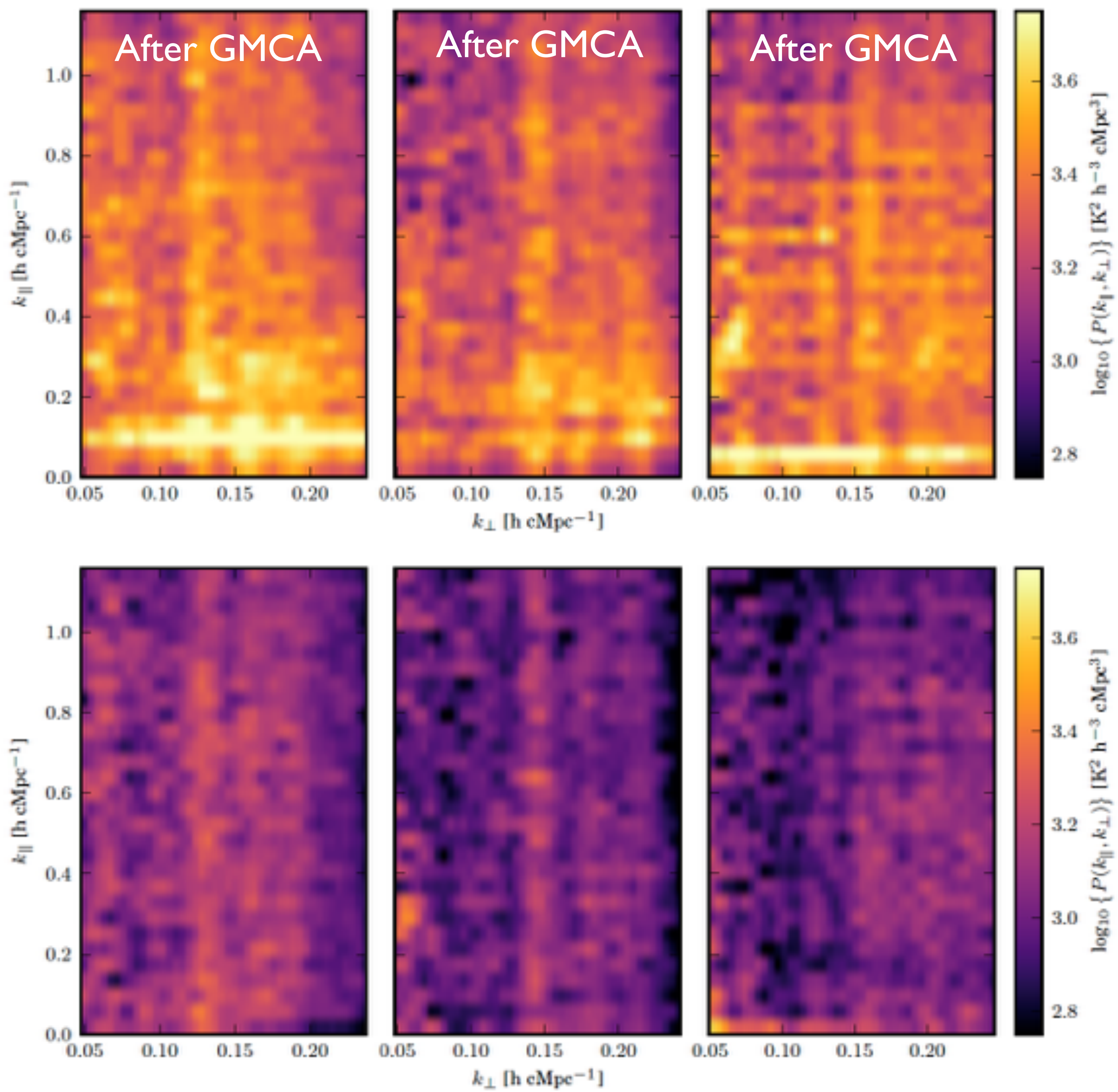
Patil et al. (2017, ApJ)

NCP Cylindrical Power-Spectra

z=9.6-10.6 | z=8.7-9.6 | z=7.9-8.7

Stokes I
(6-8 comp)

Stokes V
(0-2 comp)



After GMCA FG removal we determine the PS in both Stokes I (containing EoR + Noise + more?)

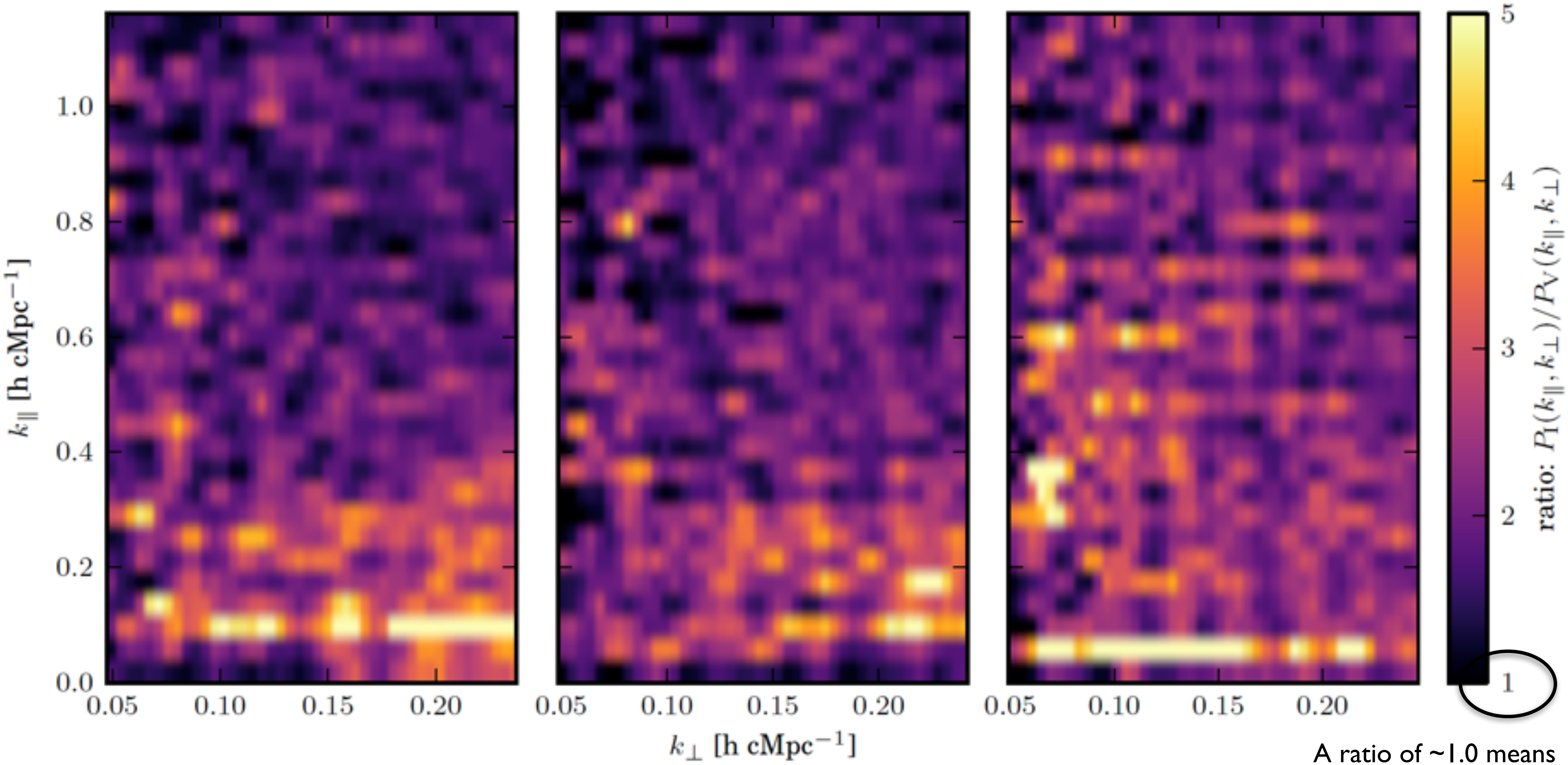
- Stokes V appears consistent with noise (w/i ~5%) and structure is related to uv-density variations.
- Stokes I shows "excess variance". (see next slide). And residual FGs at low $k_{||}$. This could be amplified residual sources, side-lobes, and/or ionosphere (under further study).

Patil et al. (2017, ApJ)

NCP Cylindrical Power-Spectra I/V Ratio

The ratio of the Stokes I over Stokes V power, shows and "excess variance" due to amplified noise, side-lobes and un-modelled structure (see Patil et al. 2016; Sardarabadi & Koopmans 2017, in prep.). Stokes I approaches Stokes V, but does not reach is entirely.

Clearly FG removal has not been perfect yet (but wait a few slides!!)

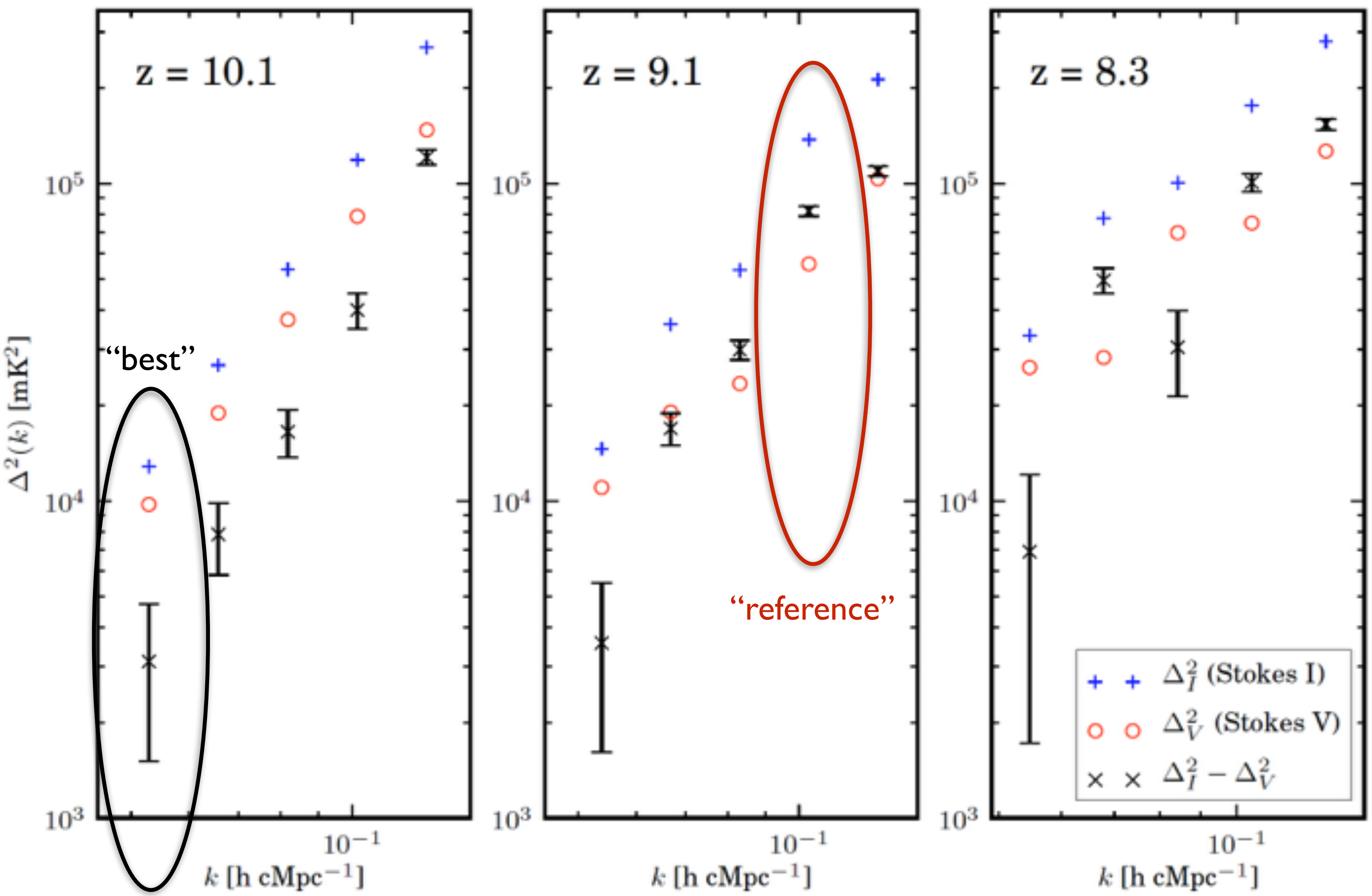


A ratio of ~1.0 means Stokes I (EoR + Noise), reaches Stokes V "noise" within the errors (EoR is assumed much lower still).

Patil et al. (2017, ApJ)

NCP Spherical Power-Spectra

Averaging spherically provides the lowest errors (maximum # of samples per shell).



“best” upper limits

Patil et al. (2017, ApJ)

The “best” upper limit (2-sigma) on the 21-cm PS is currently reached at z~10 at the smallest k-values (scales ~120 cMpc) and reach levels of about 80 mK;

EoR signal are expected to be somewhere between few to ten mK so still much deeper.

2-sigma upper limits

k h cMpc^{-1}	$z = 7.9 - 8.7$ mK^2	$z = 8.7 - 9.6$ mK^2	$z = 9.6 - 10.6$ mK^2
0.06	$(131.5)^2$	$(86.4)^2$	$(79.6)^2$
0.07	$(242.1)^2$	$(144.2)^2$	[illegible]
0.08	$(220.9)^2$	[illegible]	$(148.6)^2$
0.10	$(337.4)^2$	$(296.1)^2$	$(224.0)^2$
0.13	$(407.7)^2$	[illegible]	$(366.1)^2$

“Next Steps” as defined in late 2016

Improvements since April 2017

Improve calibration

- ➤ Remove/reduce “excess variance” (3-4x thermal variance).
- ➤ Improve the sky/calibration model further reducing gain errors transferred to shorter baselines; Improve DD calibration; Improve beam-model
- ➤ Include diffuse emission from Stokes, Q, U and possible I to enable including short baselines in calibration (currently not possible)
- ➤ Improve diffuse FG subtraction via various methods (e.g. above).
- Use cross-variance methods to avoid the noise bias in PS analysis.
- Improve cross-correlation of gain solutions with various metrics to gain insight.

Improve sensitivity

- If OK, include previously flagged short (30-60 lambda) baselines that have very high PS sensitivity (~10x deeper at k~0.03, vs k~0.05 at the moment).
- ➤ Analyse and combine more of the data (1➟10➟100 nights, rather than ~1 night).

Second window/more data

- Add second field to the processing/results: 3C196
- ➤ Keep collecting data (~3000hr total)

Current power-spectrum results in October 2017

Going ~30-40x deeper…

Changes implemented since late 2016

- The fundamental issue in 2012-2016 (see talk of Saleem Zaroubi):

  To avoid calibrating away diffuse FGs (and possibly EoR) in DD calibration, one needs to make a cut in baselines used for calibration versus those used for the 21-cm analysis (><250λ). This leads to excess noise <250λ as shown by Patil et al. 2016, 2017. This excess noise are amplified residuals of the sky and noise, that are not part of the calibration model.

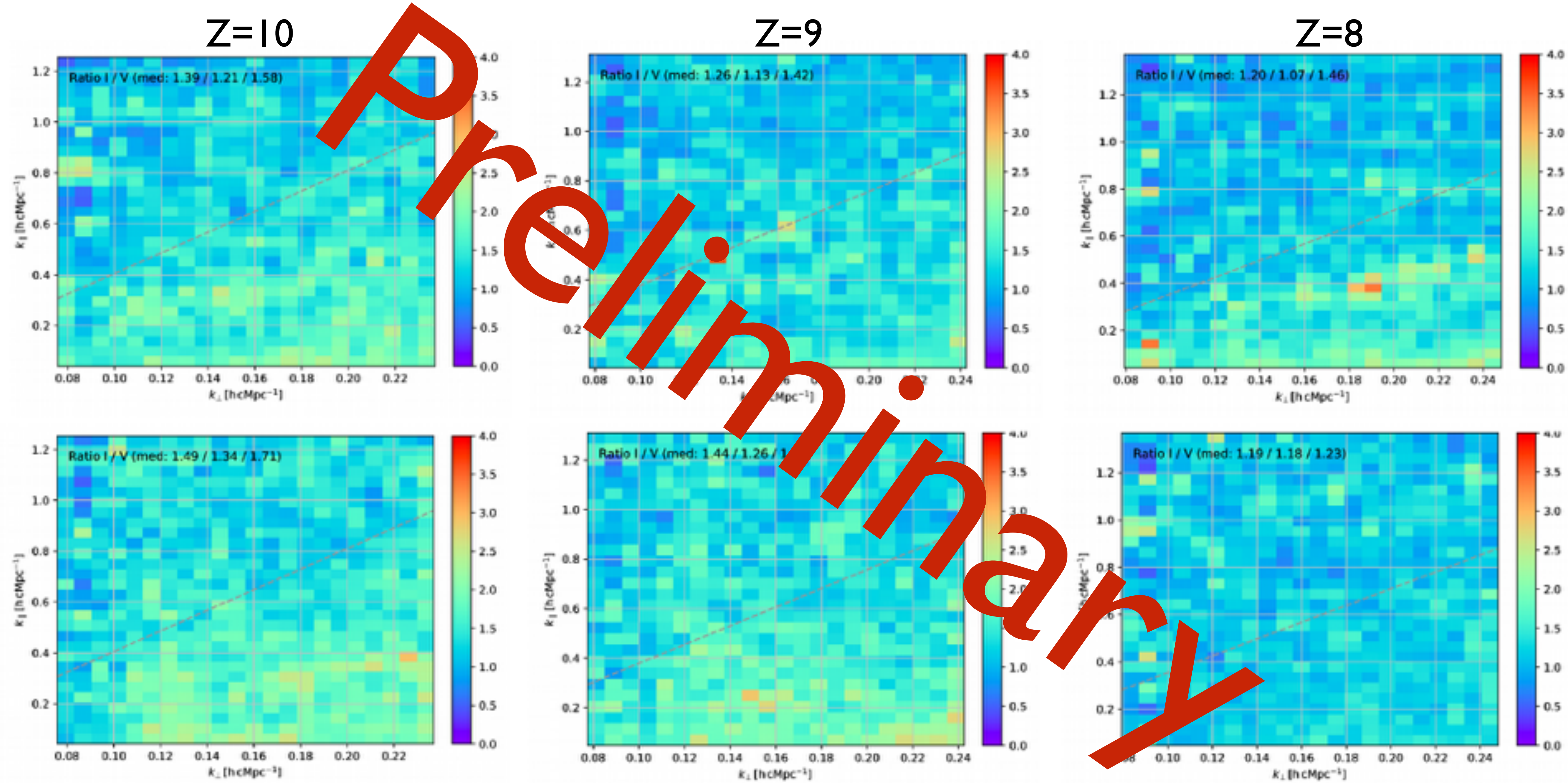
- To mitigate these coupled effects, since late 2016 we:

  ✓ Add diffuse DI-calibrated sky models in I, Q, U (shapelets) to the all calibration steps.
  ✓ Calibrate using all frequency channels together, w/regularisation (DD-SageCal-CO)
  ✓ Remove the baseline cut between calibration and imaging.
  ✓ Remove residual FGs using Gaussian Process Regression (talk: Florent Mertens)

This has reduced excess variance by a factor 3-4, without appreciable suppression of the diffuse emission, as was earlier seen (Patil et al. 2016). We are testing suppression/bias further by injecting signals in to the data.

New PS Results in October 2017

Examples of cylindrical power-spectra from two nights in Stokes I/V



- Ratio I/V is quite flat; clear of obvious structure.
- Stokes I has r.m.s. ~ 10-20% higher than Stokes V (~thermal noise).
- No obvious wedge; maybe some incoherent chromatic effects (ionosphere?)
- In some cases RFI or residual (A-team) sources are seen

New PS Results in October 2017

Spherical power-spectra after five nights give $\Delta^2 < (49\ \mathrm{mK})^2$ @ $k=0.1$, $z=9$ $(dk \sim k/4)$

Idem we have PS for z=8 and 10 (not shown)

Preliminary

Figure credit: F. Mertens

Conservative still: no optimal weighting, residual RFI rejection, noise-bias correction

Forecast (NPC) based on extrapolation of current results.

If the data keeps averaging down, as it has been in 5 nights, detection levels of ~few mK can be reached with the data in hand (~1700 hrs on NCP alone) in a dk~k (1 dex) range for k~0.1).

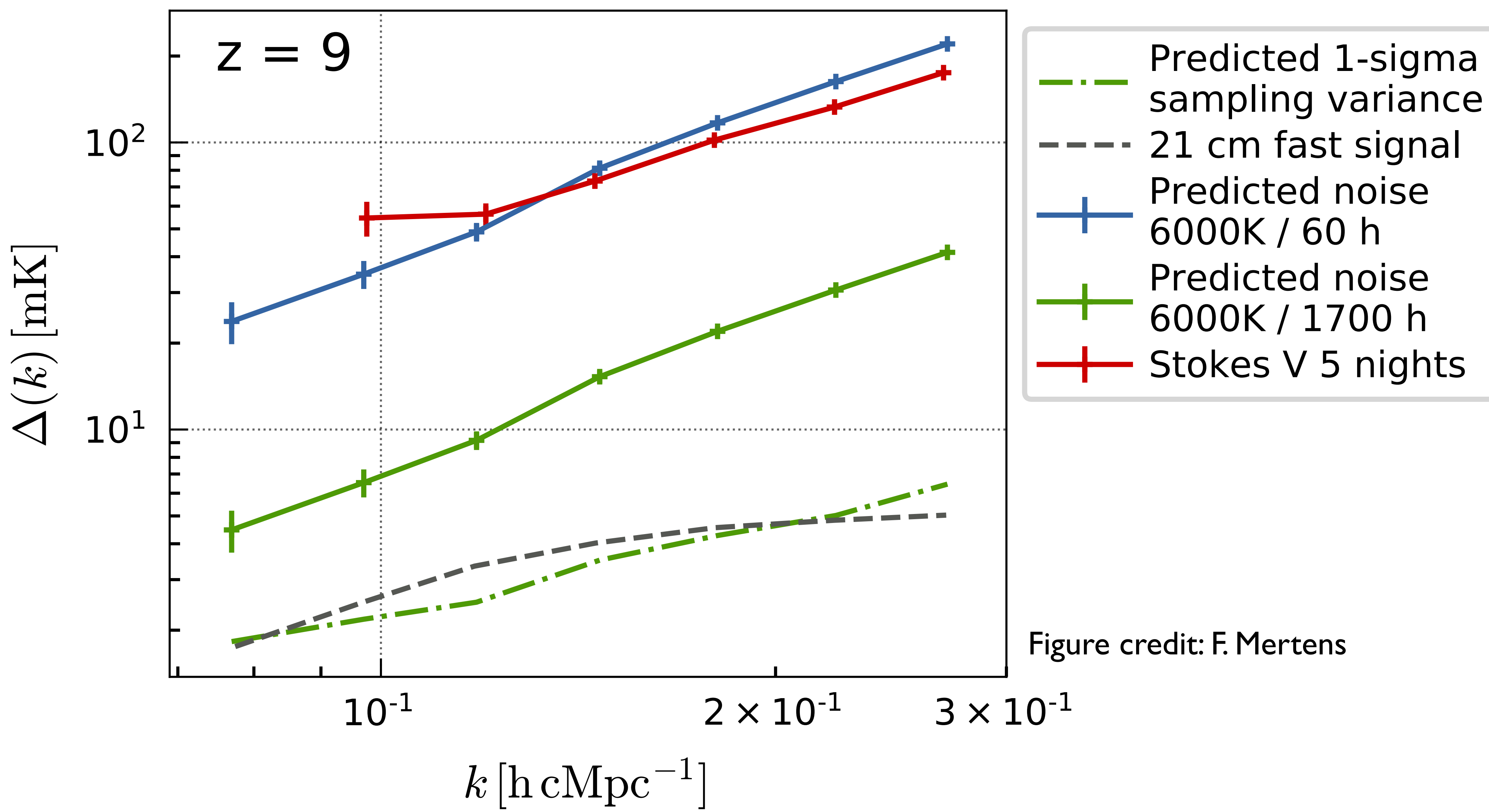


Figure credit: F. Mertens

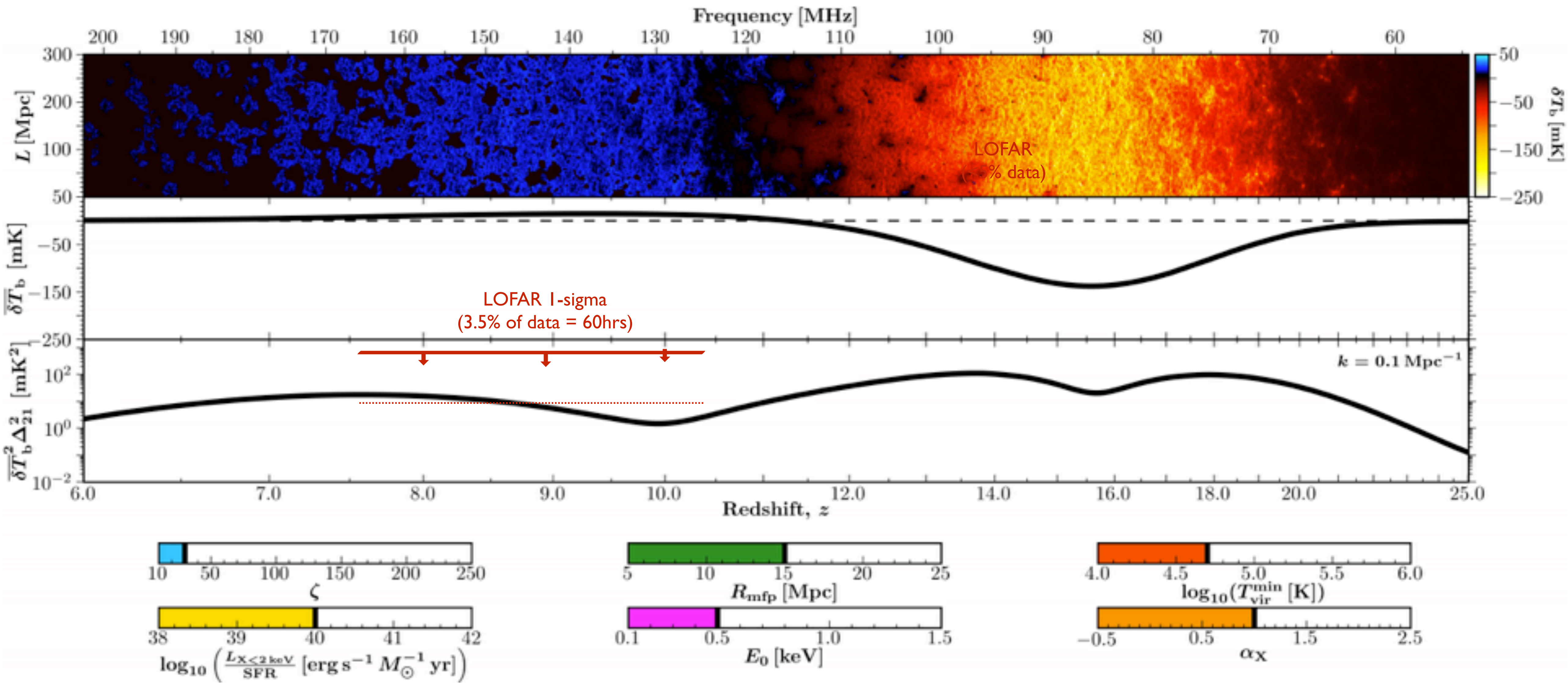
rijksuniversiteit groningen / faculteit wiskunde en natuurwetenschappen / kapteyn instituut

Constraints on the Cosmic Dawn/ Heating Epoch

Physical Limits on the Cosmic Dawn

21CMFAST - varying the X-ray heating luminosity.

Credit: Mesinger & Greig

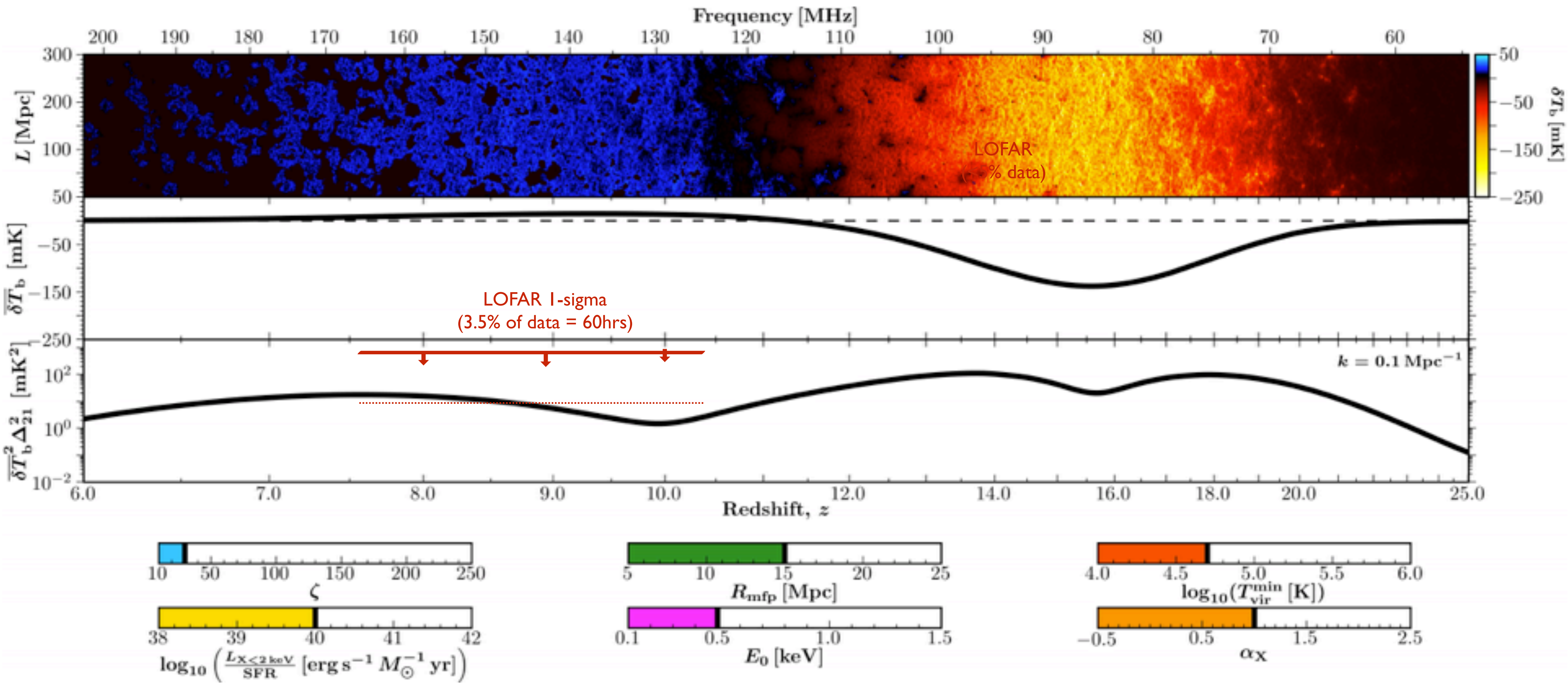


Note: Excess noise (I over V) is incoherent (averages away): we assume it drops in the cross-variance and plot the 1-sigma upper limits (in 21CMMC, we double the errors)

Physical Limits on the Cosmic Dawn

21CMFAST - varying the X-ray heating luminosity.

Credit: Mesinger & Greig



Note: Excess noise (I over V) is incoherent (averages away): we assume it drops in the cross-variance and plot the 1-sigma upper limits (in 21CMMC, we double the errors)

Summary & Next Steps

- After Cycle 9: ~3000 hours of data (NCP, 3C196) on disk of which 70-80% of good quality Ionospheric scintillation leads to (only) <20-30% data 'loss'; new GPU cluster to process this.

- Last years were were all about understanding 'systematics', improve wide-field broad-band calibration (SageCal-CO), separation imaging/calibration, working on compact/diffuse sky models, polarisation and ionospheric effects.

- Since 2016/2017 we are close to reaching the thermal noise w/o appreciable (known) systematics. Some incoherent (~15%) excess noise is still seen, but it averages down largely

  ✓ Current published EoR limits (13hr) at z=9 is at Δ_{21}<296 mK (2-sigma) levels at k ~ 0.1

  ✓ Current unpublished EoR limits (~60hr) at z=9 is at Δ_{21}<49 mK (2-sigma) levels at k ~ 0.1, some 36x deeper in power than before at same k-value. We are able to start testing CD/EoR model spaces. This is only based on a ~3.5 percent of our NCP data in hand.

  Preliminary

- NCP data should be sufficient for a few mK (over dk~k) depths at k~0.1 in the z=8-10 range.

- Next steps: (a) process more nights aiming for ~10 first, (b) optimal power-spectrum extraction, (c) excise residual RFI/sky-power, (d) use cross-variance, (e) include baselines in common electronics cabinet, (f) start exploring CD/EoR model parameter space.