The LOFAR EoR KSP: Data Challenges
Finding an EoR Needle in a Pbyte Haystack

Stockholm, Sweden

August 17-21, 2009

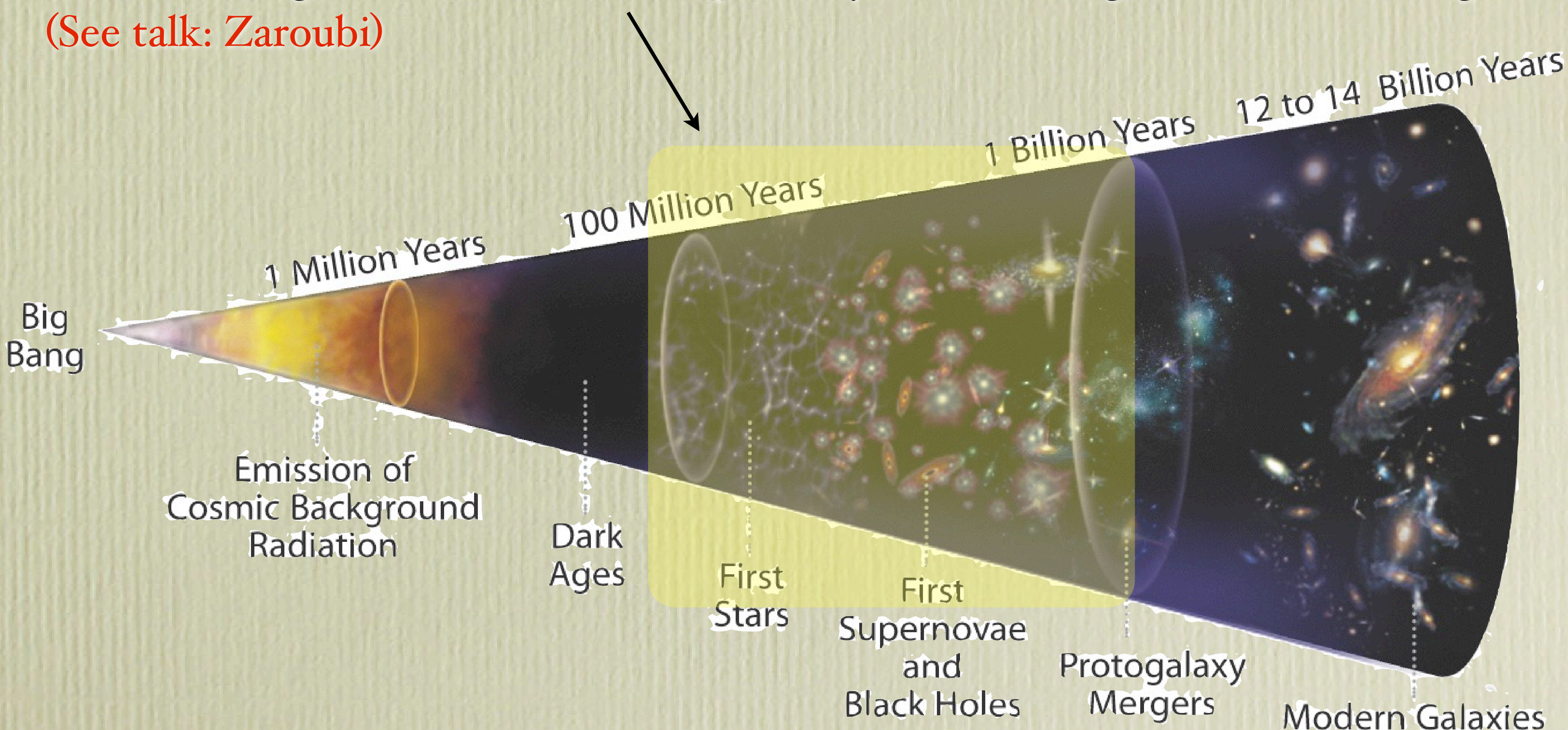
Léon Koopmans

(Kapteyn Astronomical Institute)

Co-PI LOFAR-EoR KSP

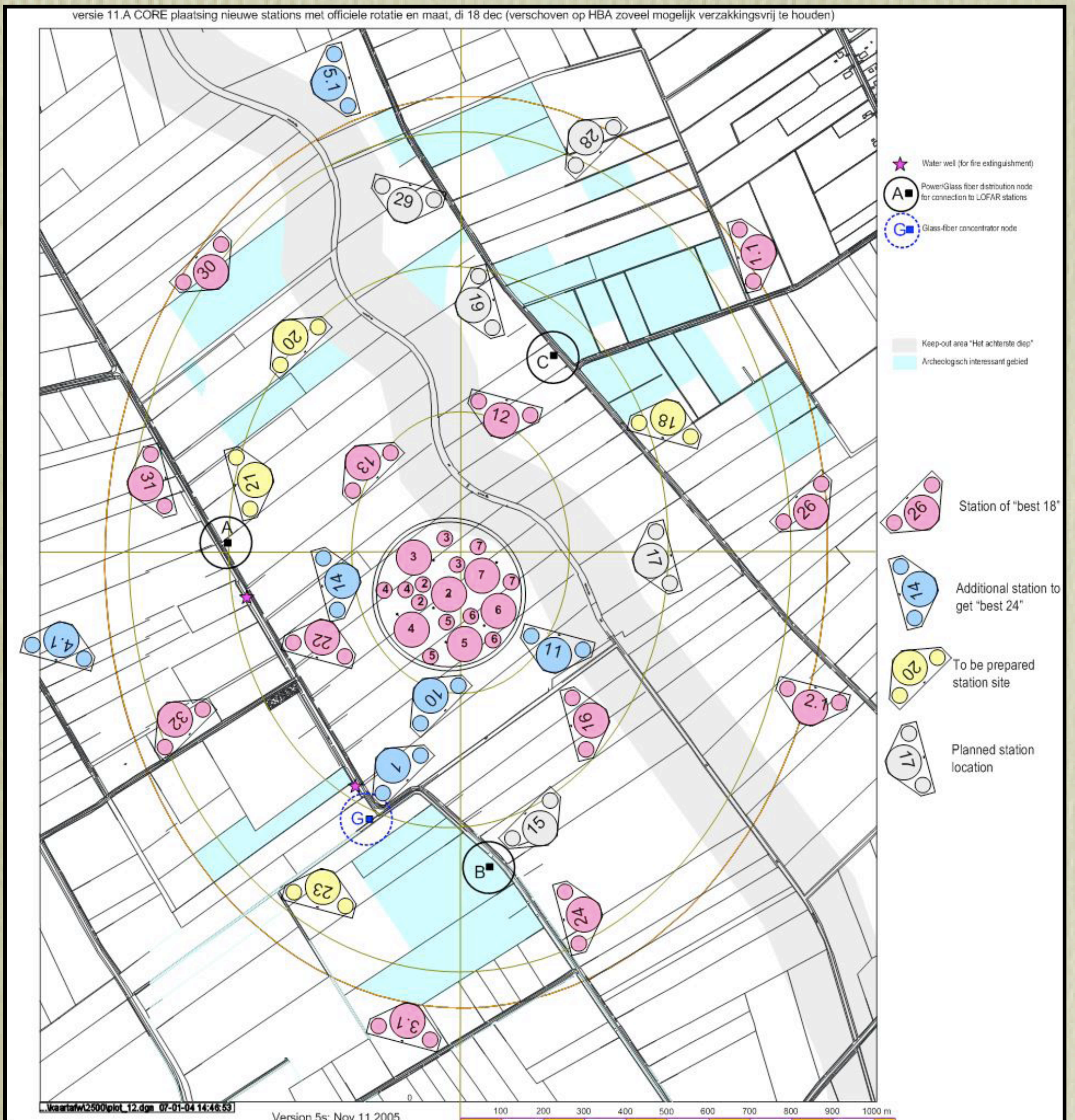
The LOFAR Epoch of Reionization Key-Science Project

Goal: Tracing the EoR in HI and possibly the late stages of the Dark Ages
(See talk: Zaroubi)



Wonderful on paper, but doable in practice?

The LOFAR Central Core - Station Positioning



Core area - 2 km around Exloo

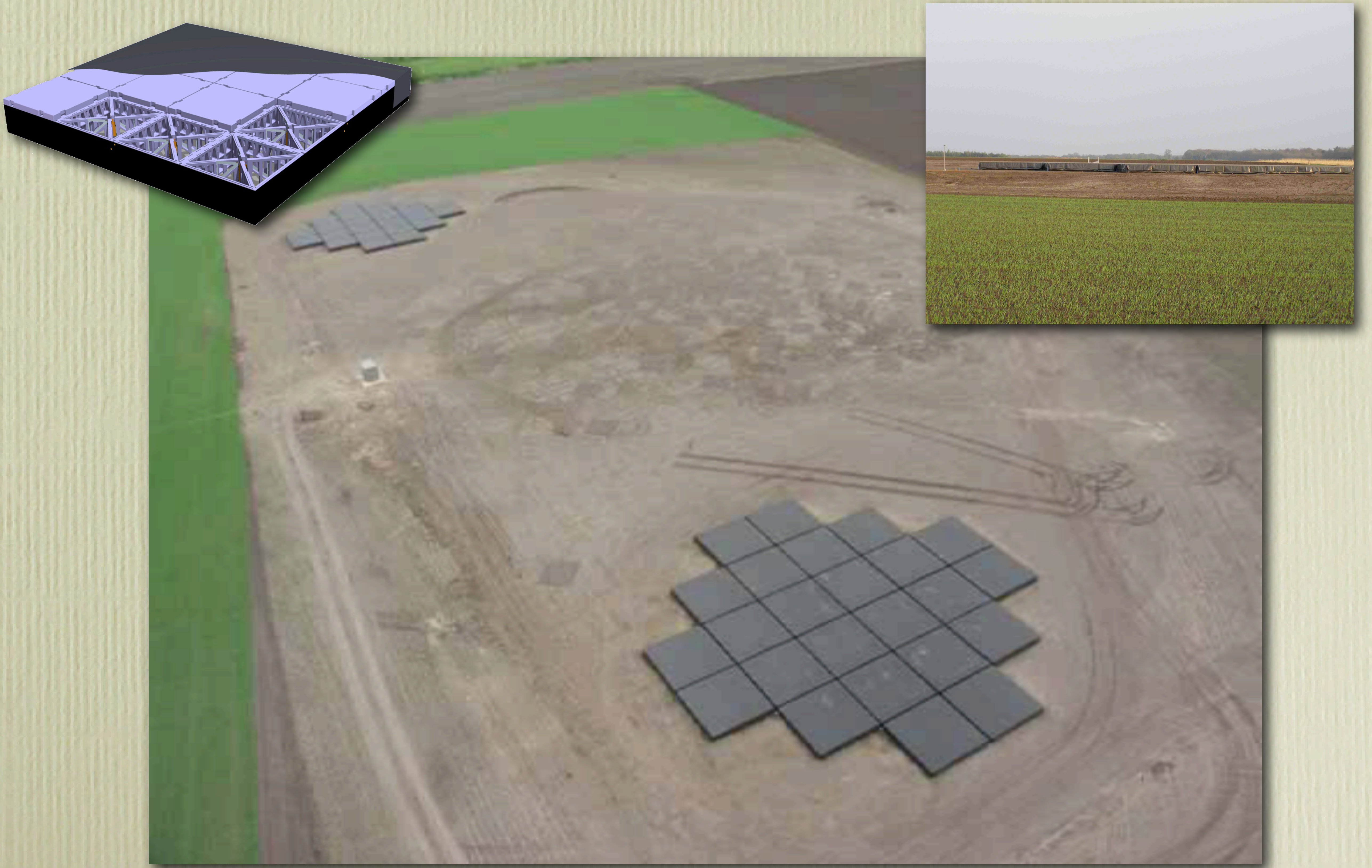
– Min. 18 stations while preparing for more (~22 likely) (note HBA stations are split)

- Goal is to have dense UV-coverage to probe all scales on the sky with few arcmin resolution (~150 MHz)

- Genetic Algorithms were use to optimize UV coverage under pre-conditions

- Area will become nature-reserve.

First Dutch High-Band Antenna Station

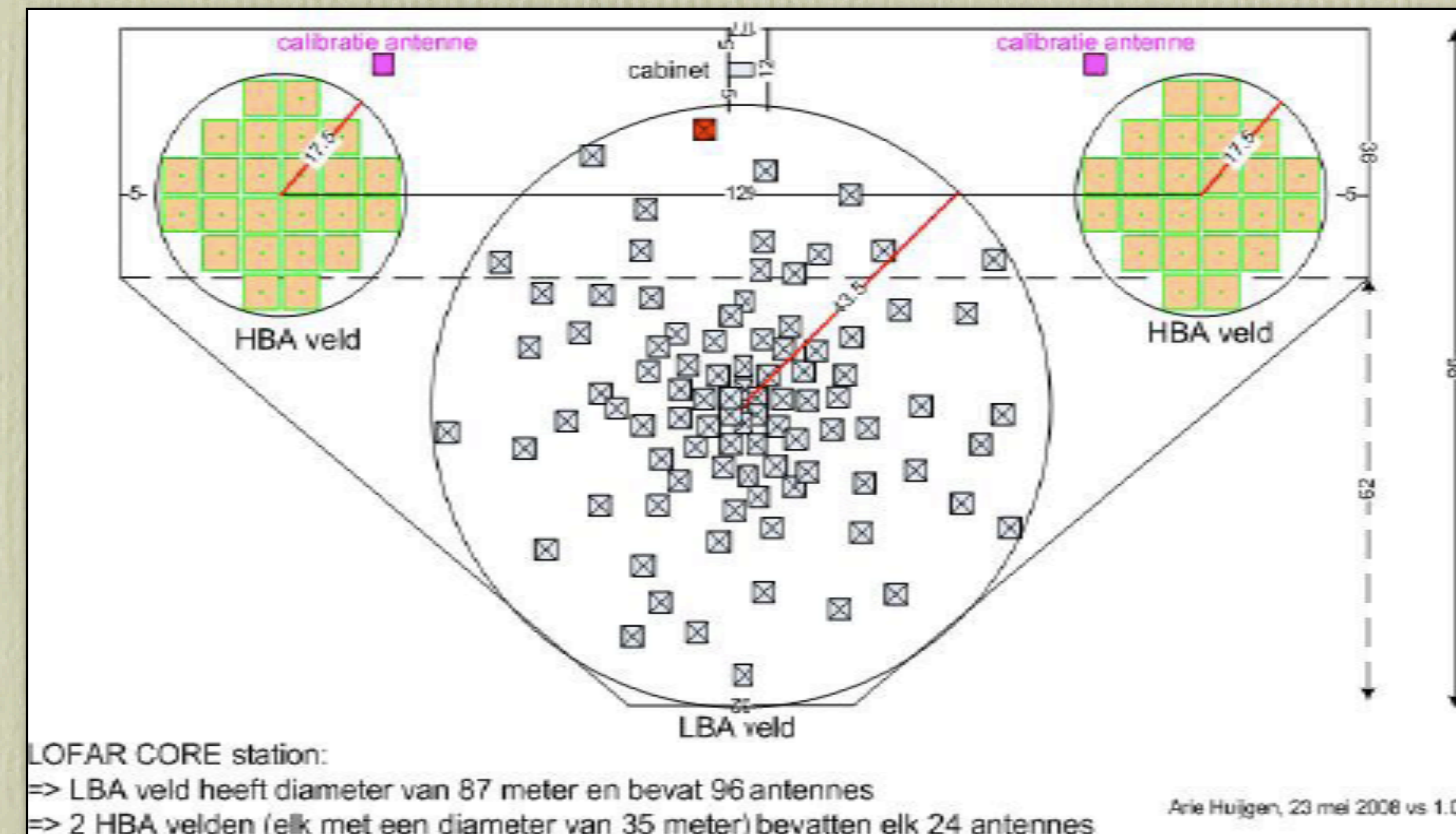


CS302

Credit: Vermeulen

The LOFAR Central Core - Dual HBA Stations

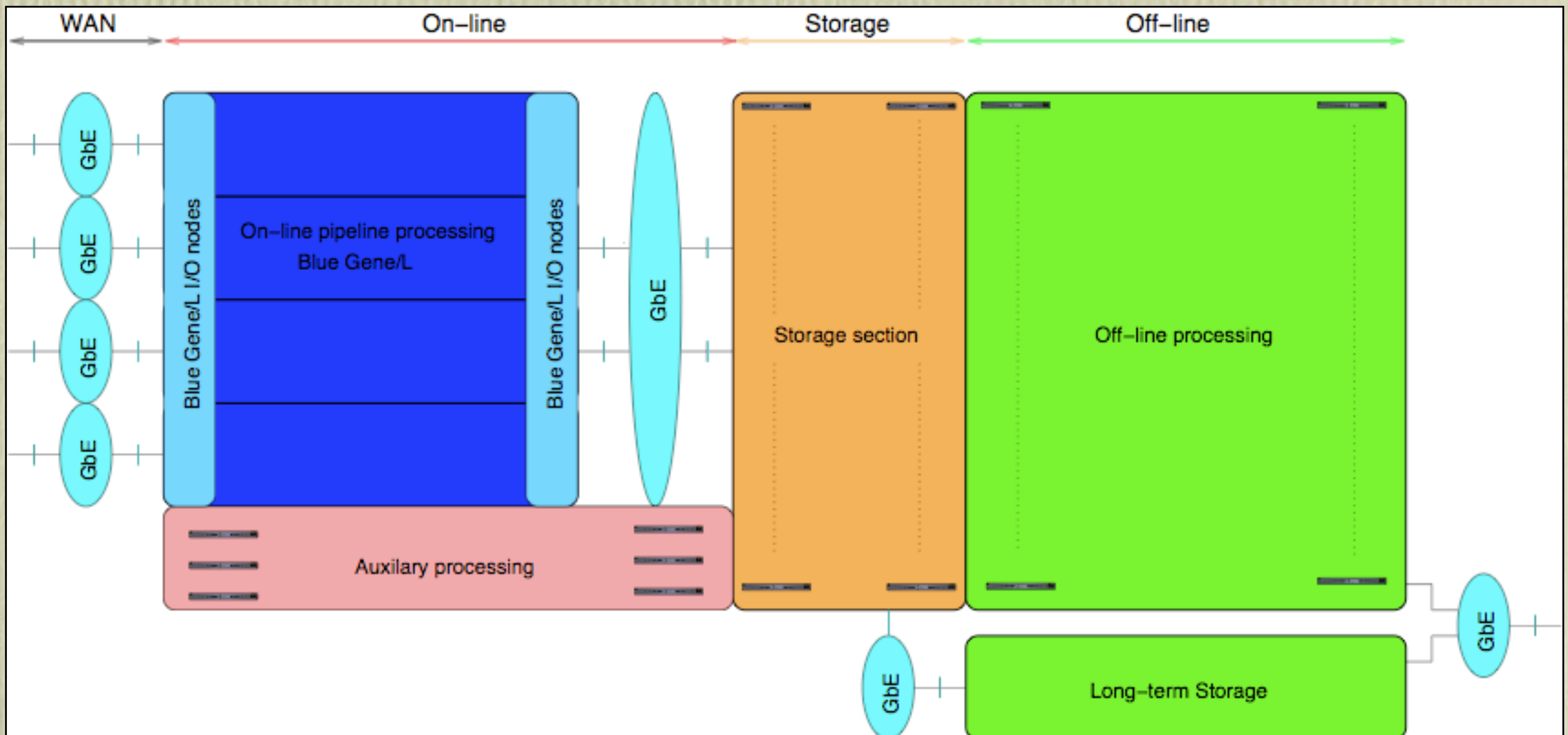
96 LBA & 2x24 HBA tiles



Low Band Antenna
- (10) 30MHz - 80 MHz
- 96 dipoles per station
- within NL only 48 can be used at a time

High Band Antenna
- 120 MHz - 240 MHz
- 1 tile = 4 x 4 antennas
- 48 tiles per station (within NL) or 96 tiles

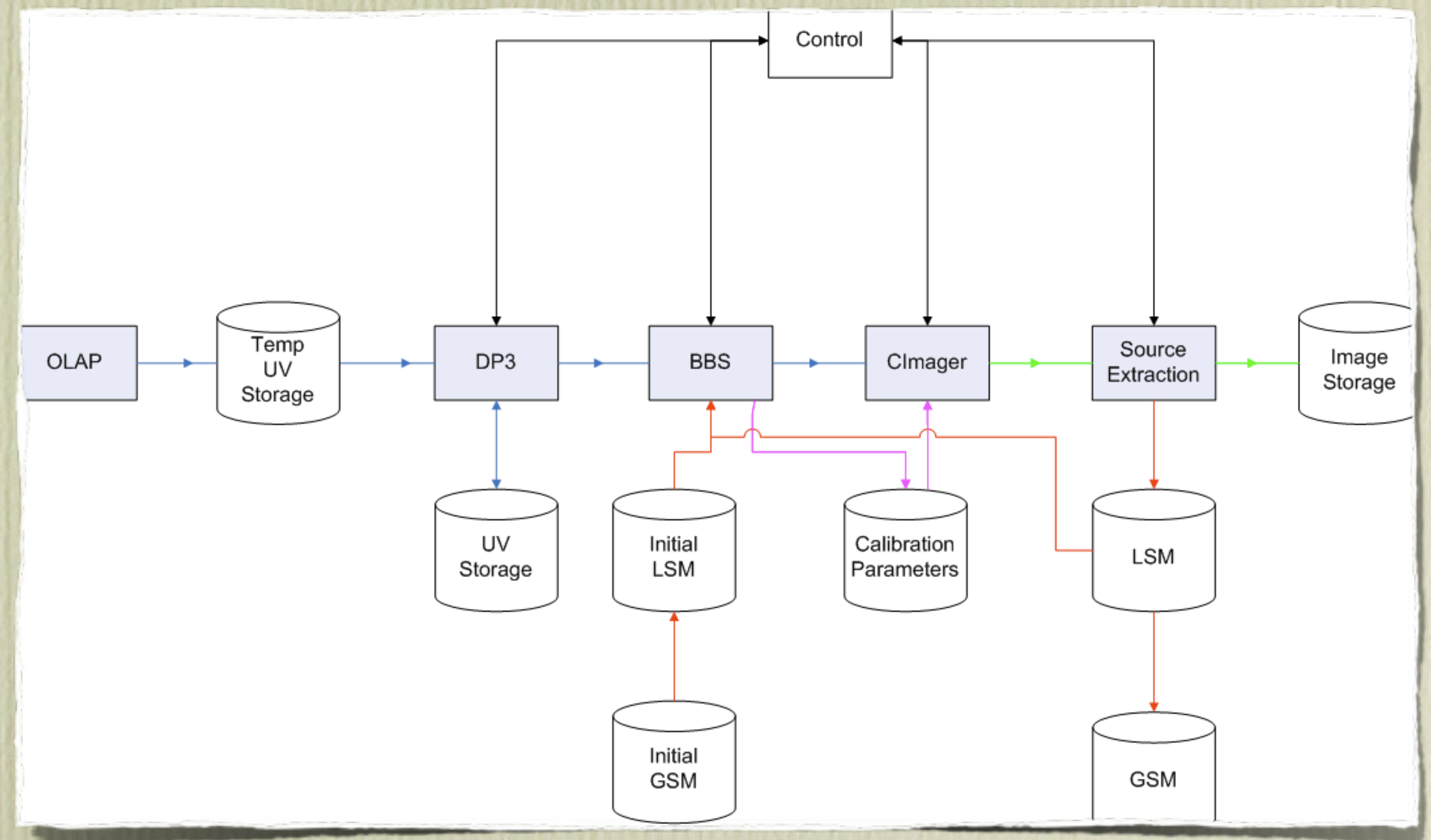
Central Processing at Groningen/RuG CIT



- BG/P Data reception, transpose, correlation, beam-forming, de-dispersion
- Storage system Short term storage of data, ~1 PByte, >100Gbps I/O
- Offline cluster Calibration, data products, off-line analysis, ~1000 nodes

Standard Offline Processing Pipeline

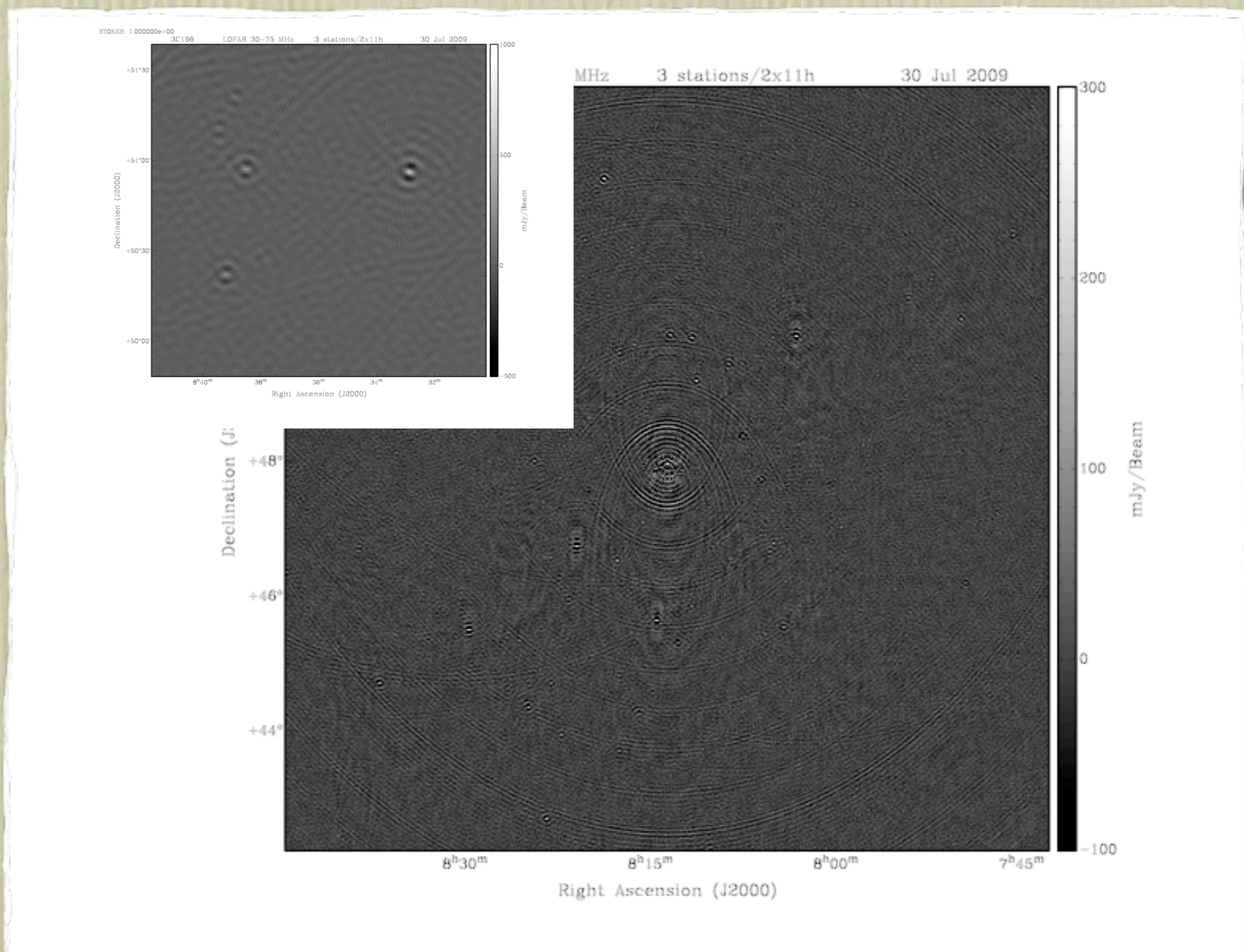
The LOFAR Project provides "standard" modes and data-reduction products;
Complex reduction and analysis the responsibility of the (KSP) projects



Standard and real-time imaging is **not** enough for the detection of the EoR signal: Reprocessing is required

Standard Imaging Pipeline

Even though results are already encouraging (DR ~ 1000), real-time or even reprocessed data are not sufficient to detect the EoR signal.



Calibration: MeqTrees
Imaging: casa/aips++
Hamaker beam-model
Full polarization processing

Credit: Sarod Yatawatta

The EoR Project - Astrophysical & Instrumental Challenges

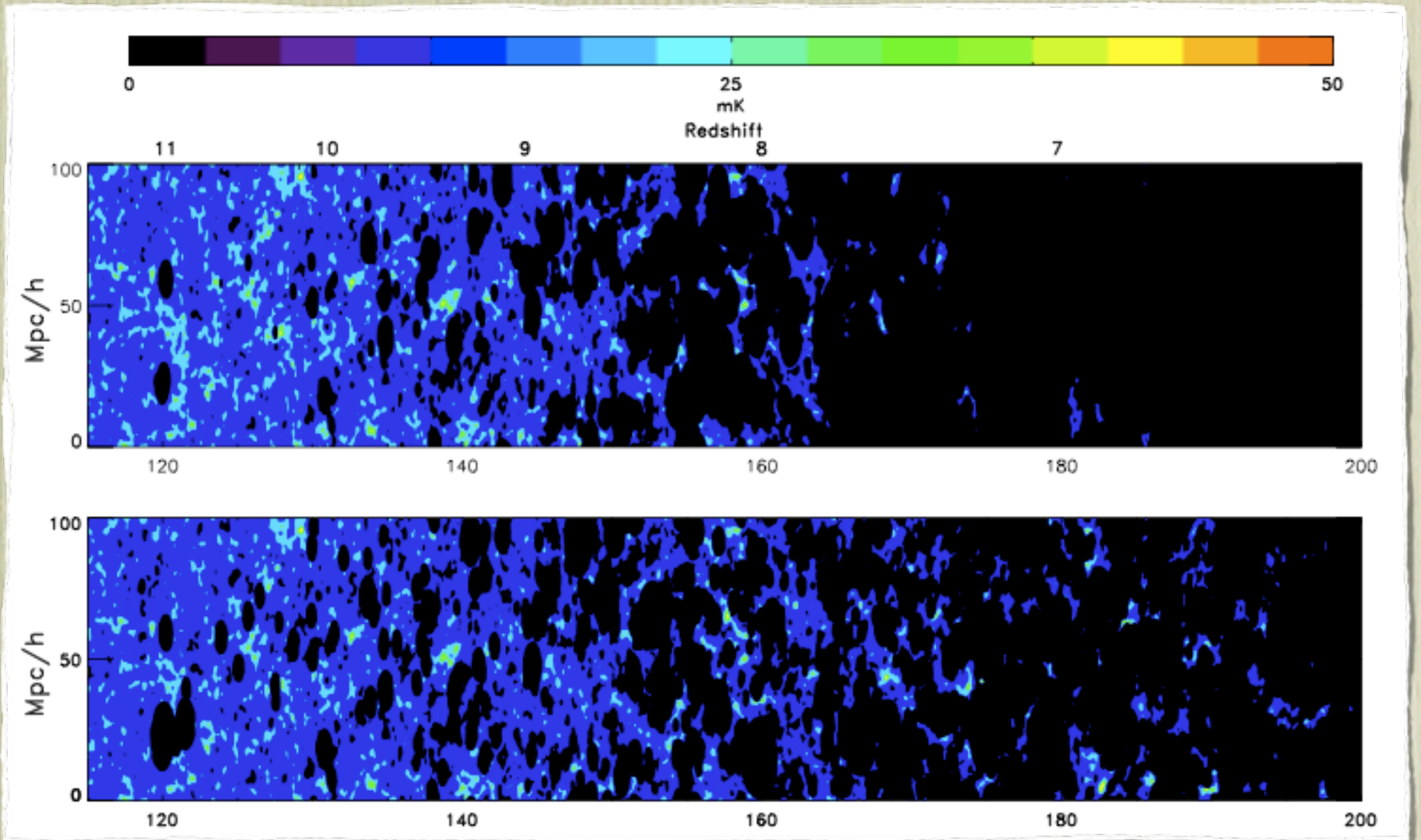
LOFAR EoR-KSP Challenges

The LOFAR EoR KSP is faced by many challenges.

(Astro)physical	Instrumental	Computational
Extragalactic foregrounds	Beam stability	Large data rate
Galactic foregrounds	Sensitivity	~1 Pb raw data
Ionosphere	Dynamic Range	(Non)-linear optim. of coupled equations
Interference	UV Coverage	10s Tflop*year
Polarized emission	Calibratability	...
Variability	...	

LOFAR EoR-KSP Challenges

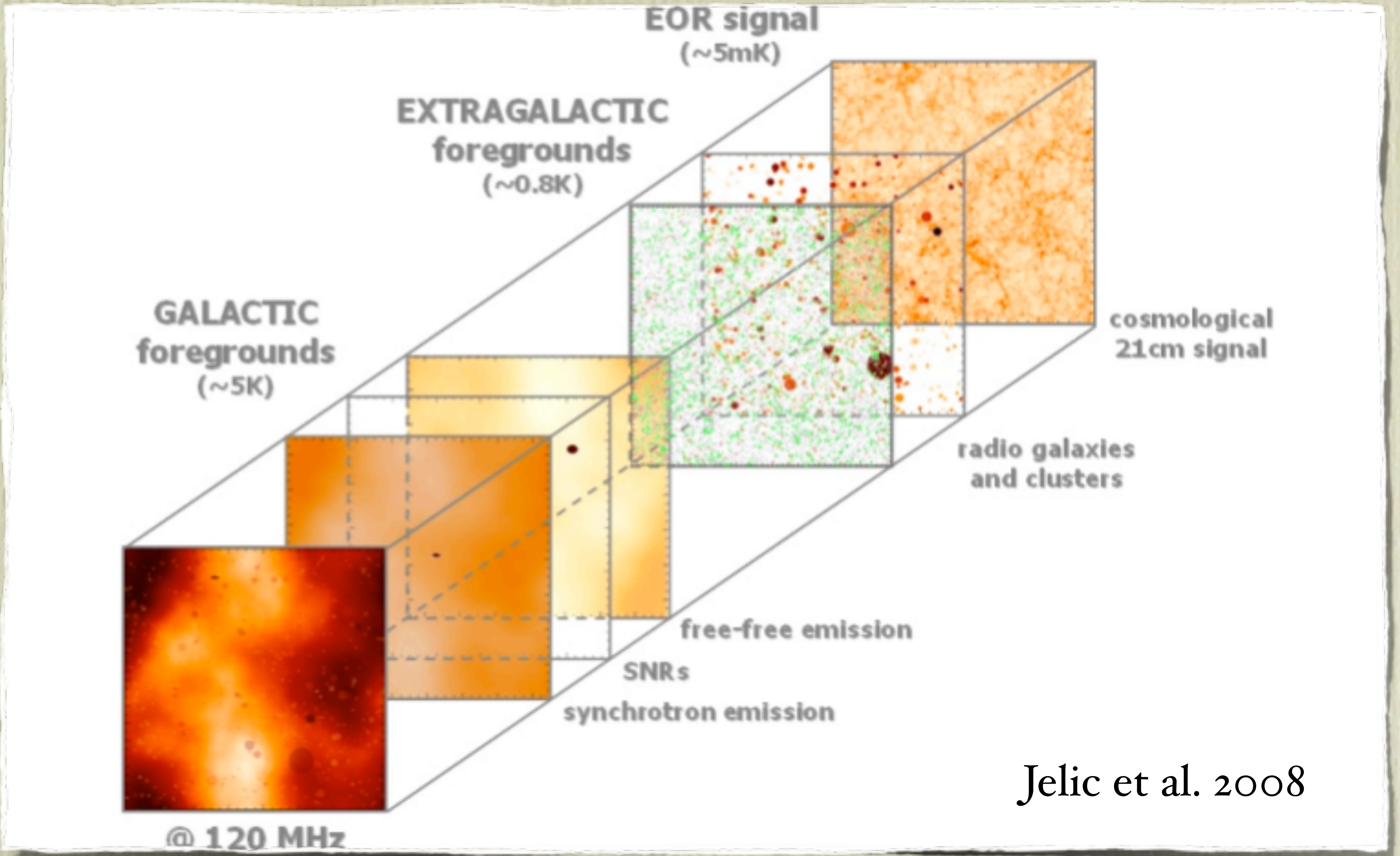
We are searching for a signal of a few mK and only have general ideas on its power-spectrum as function of redshift. (Talk: Rajat Thomas)



Thomas et al. 2008

LOFAR EoR-KSP Challenges

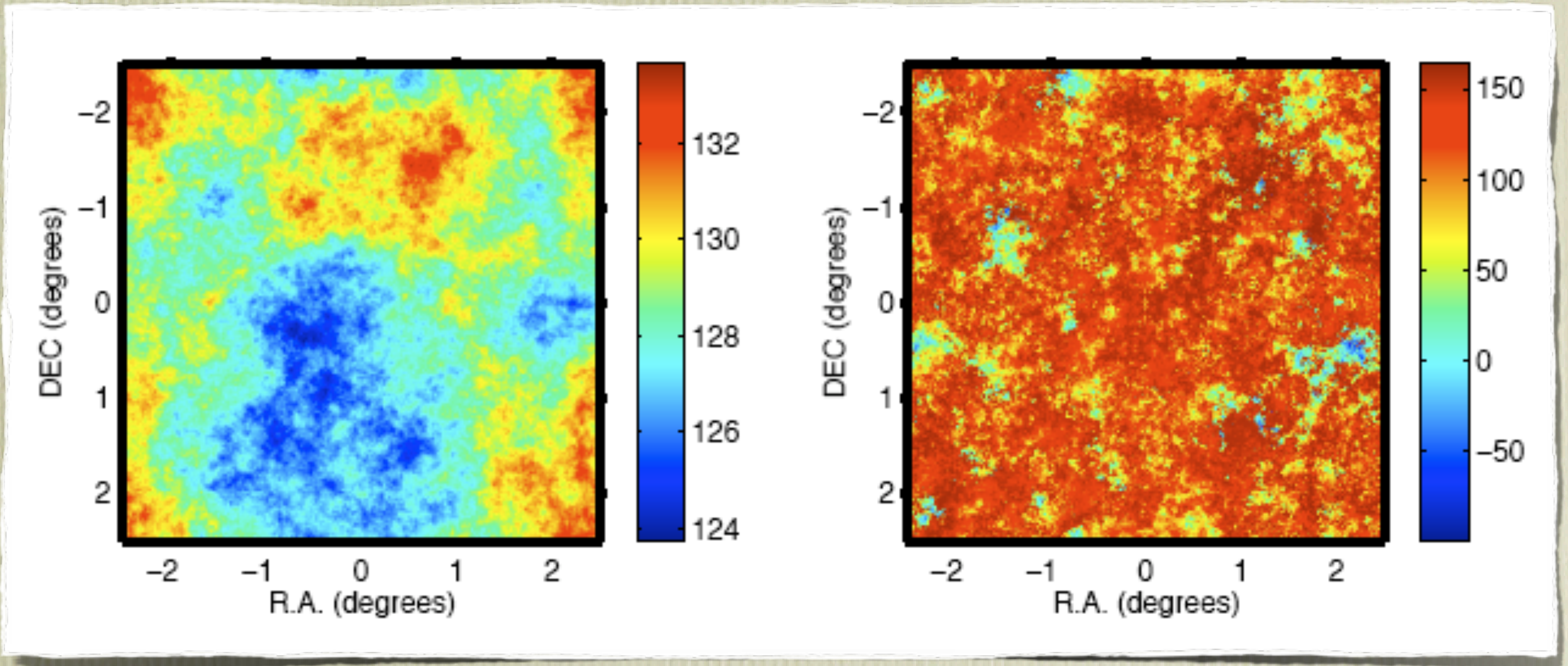
The mK EoR signal is superposed on bright polarized (extra-)Galactic foregrounds (tens to hundreds K), which need to be removed (Talk: Vibor Jelic).



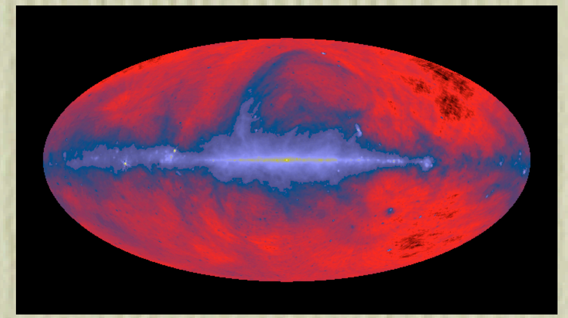
Jelic et al. 2008

LOFAR EoR-KSP Challenges

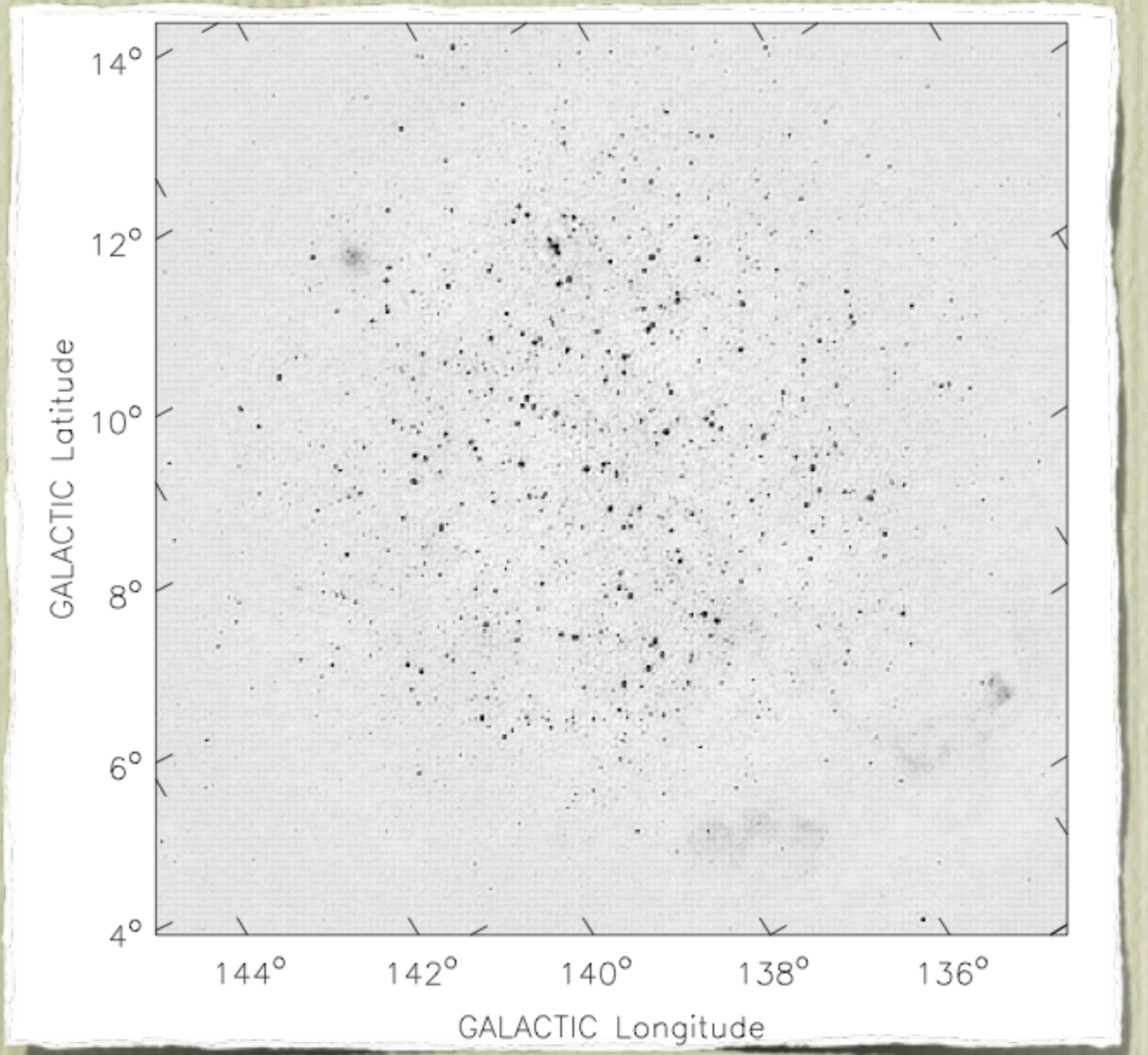
We are looking through (polarized) signals of 10s of K and only have general ideas on its behavior as function of frequency (talk: Gianni Bernardi).

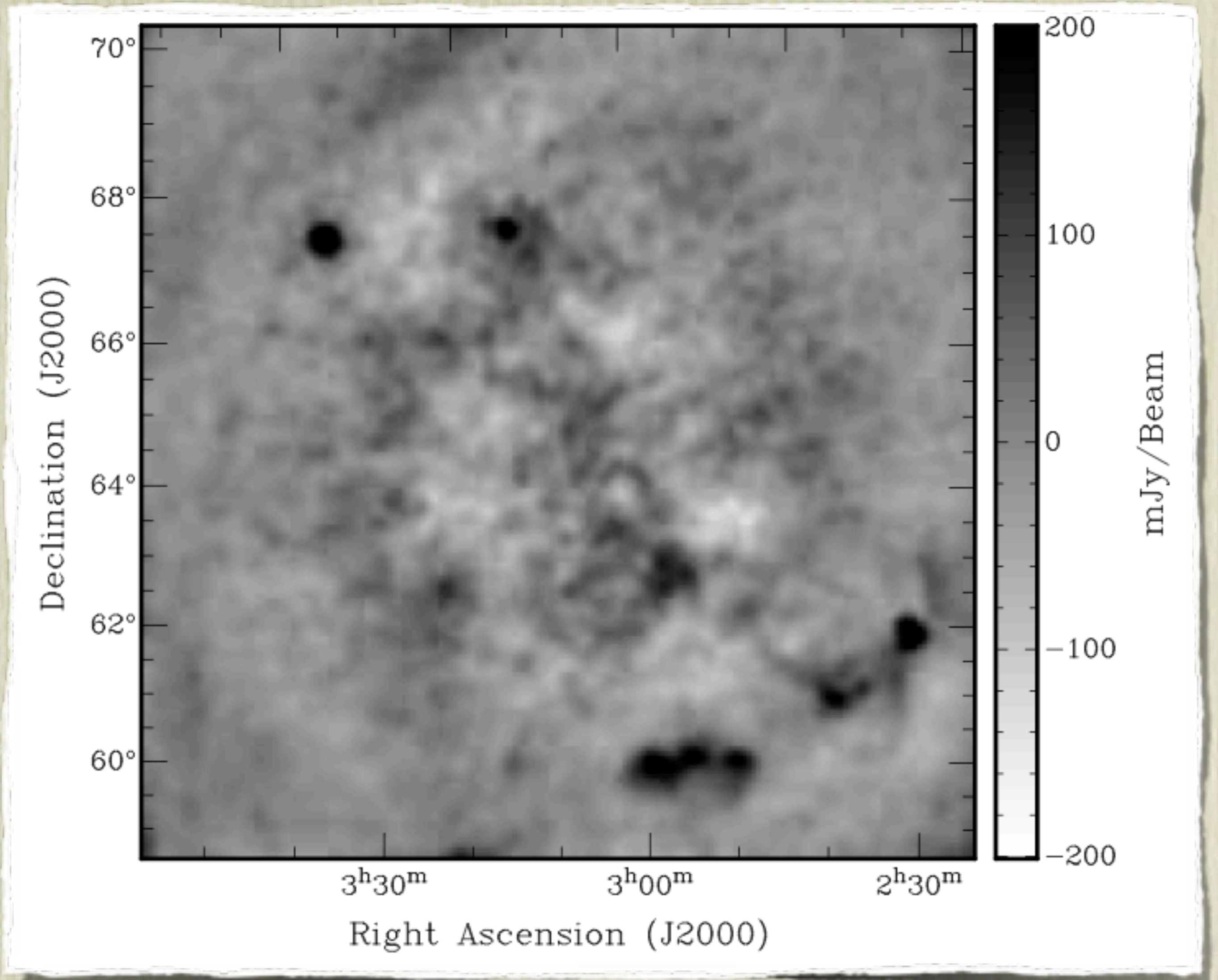


Galactic Foregrounds: Fan Region



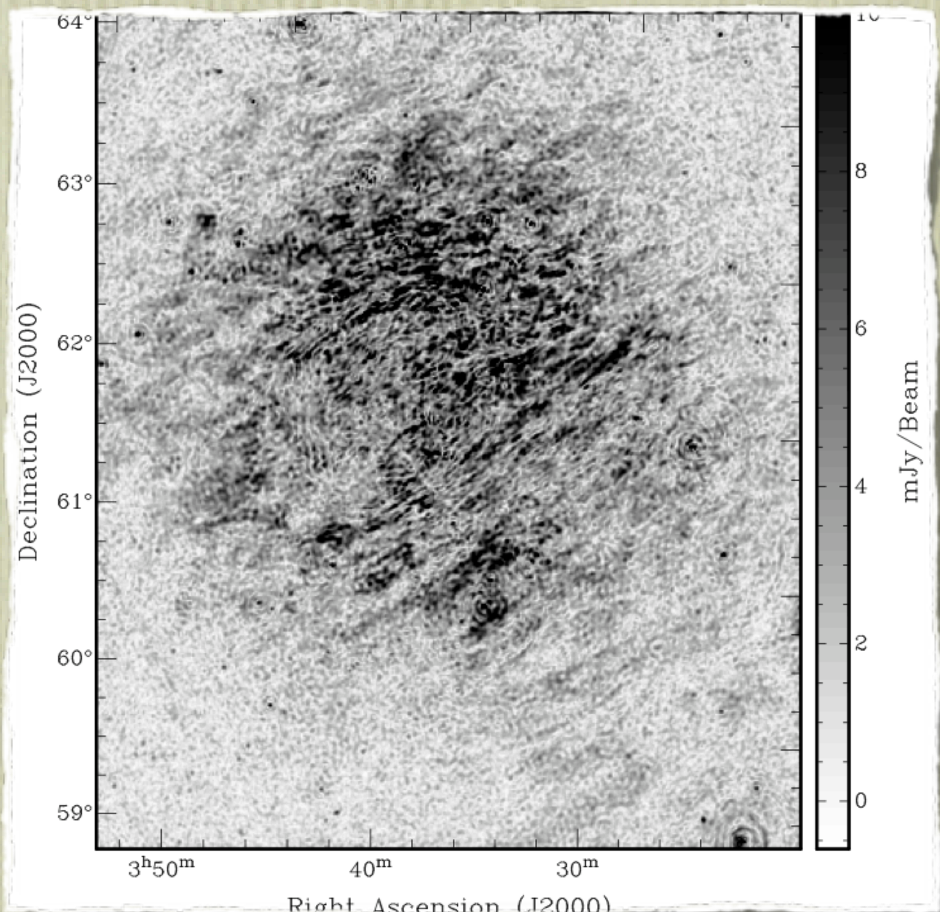
First detection of fluctuations in the synchrotron emission of the Galactic foreground on arc-min scales with WSRT-LFFEs





Bernardi et al. 2009

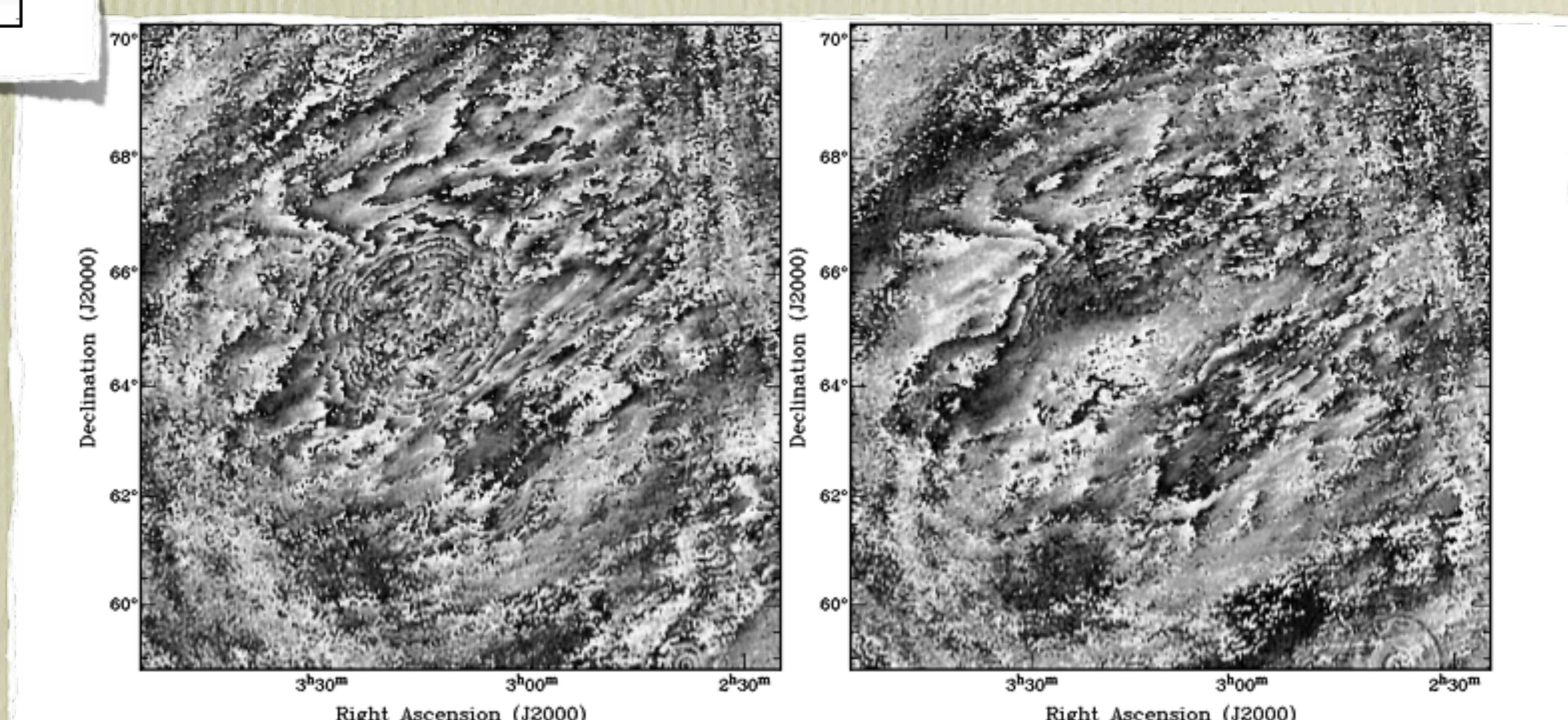
Galactic Foregrounds: Fan Region



Polarization poses a challenge:

Instrumental Polarization can leak Stokes I into Q, U & V and visa versa, causing Stokes I brightness fluctuations as function of wavelength.

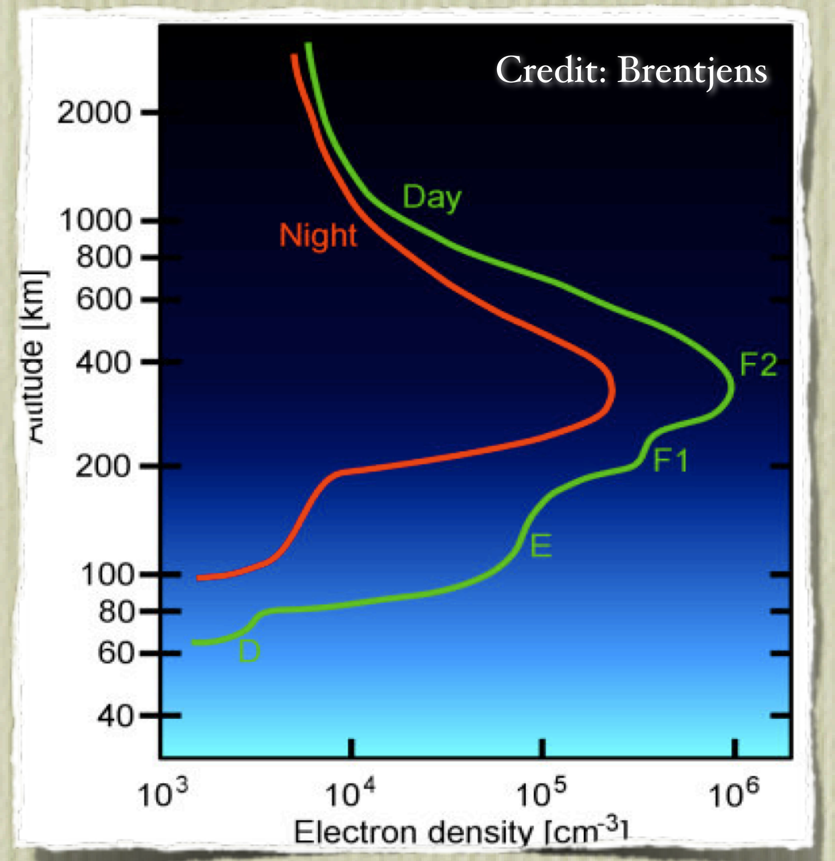
Note: with interferometers, the level of polarization can exceed 100%

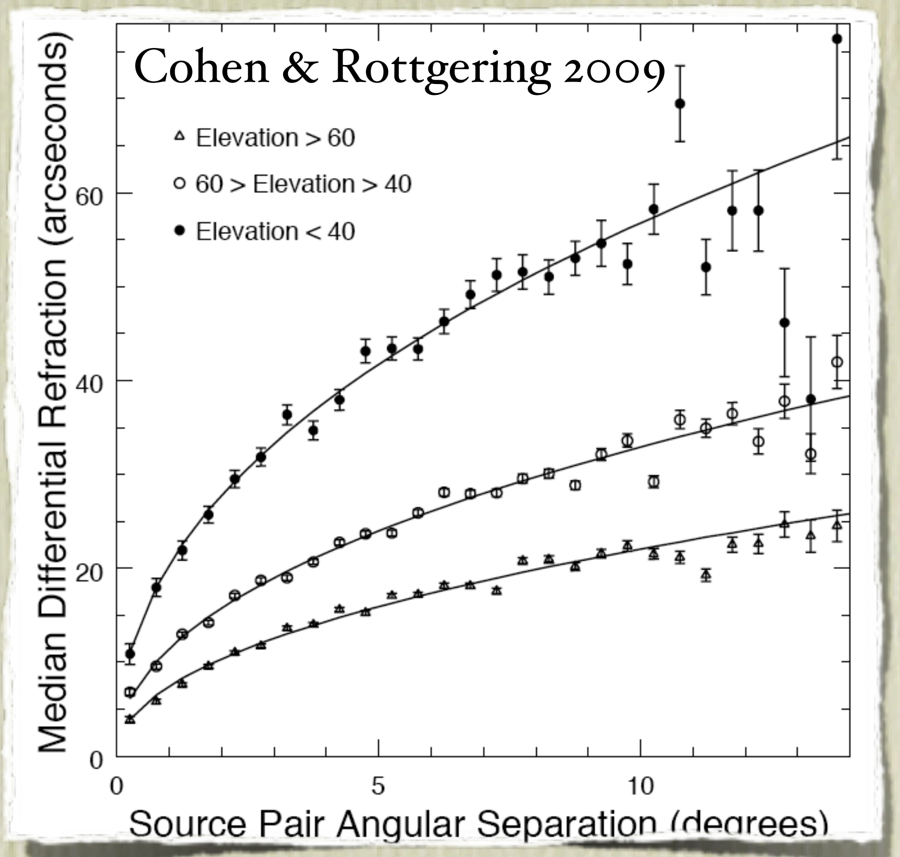


Bernardi et al. 2009

Complex Ionospheric Modeling

Is 3D (slab) or 2D (plane) model sufficient to describe the ionosphere

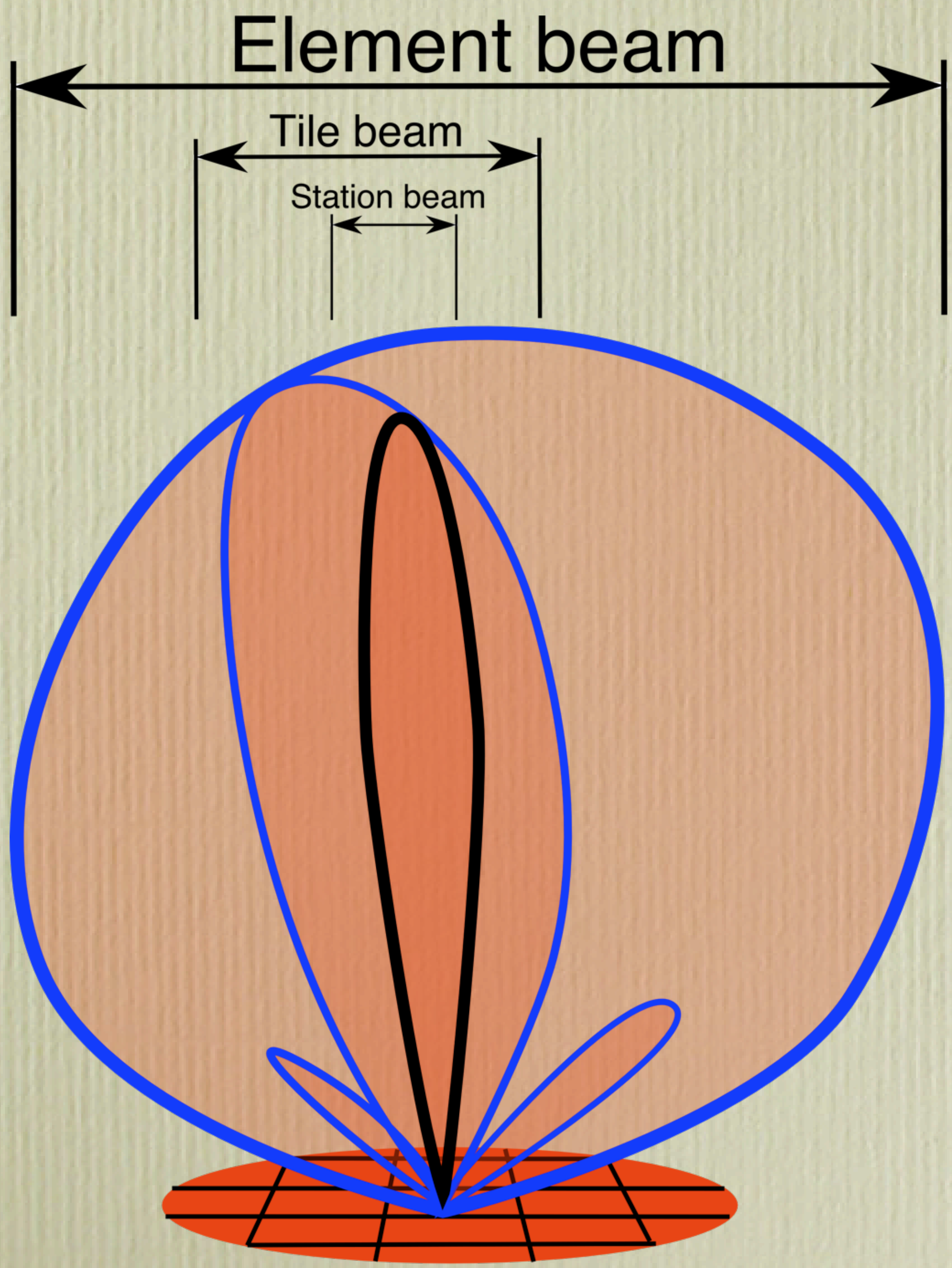




VLA 74 MHz

- F1/F2 responsible for most phase effects and refraction
- Daytime: TIDs
- Sometimes: scintillation
- F1/F2 merge during night
- D layer causes absorption (5% variations at 150 MHz?)
- D layer rapidly recombines during twilight
- E layer recombines more slowly: see RFI along entire terminator during twilight

Complex Beam Modeling



Complex Beam Jones Matrix:

$G(l; l_{rf}; l_d; u; q; q_o; \nu; G_{ant}; X_{ant})$ (1)

Simulate by:

Full EM simulations?
Semi-analytic approximation?
Analytic modeling?
Genetic programming?
. . . ?

Credit: Brentjens

LOFAR EoR-KSP Challenges

Why? Dynamic ranges in the data motivate an accurate & precise calibration, imaging and signal extraction strategy.

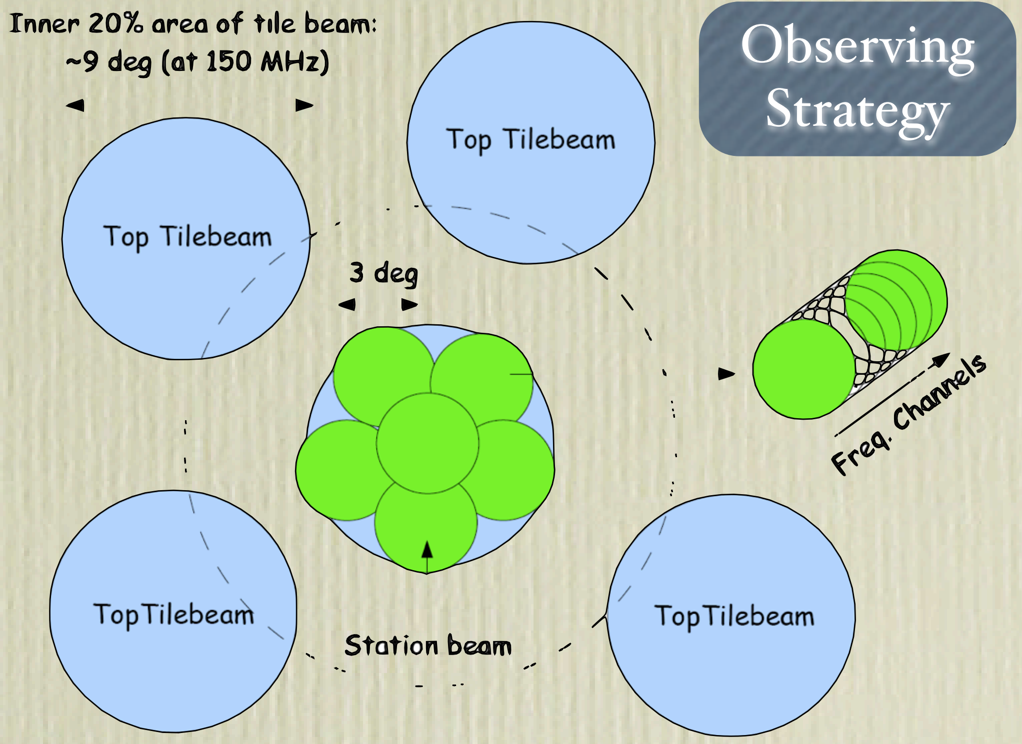
Discrete sources	10^{4-9} μJy/beam
Galactic Foregrounds + Confusion	10^3 μJy/beam
Thermal/Sky Noise	10 μJy/beam
EoR 21-cm signal	1 μJy/beam

(assuming a 3-arcmin resolution beam at 150 Mhz)

The LOFAR Observing Strategy & Data Model

The LOFAR EoR-KSP: Observational Strategy

- Up to 5 independent EoR observing "windows"
- Up to 4-6 independent station-beams/window
- 2 bands of 48-32 MHz/10 kHz = 96-6400 ch./beam
- 1128 visibilities (2x2 matrix) per t_{int} = 10 sec
- t_{obs} = 5 windows x 2 bands x 300 hrs = 3000 hrs
- Full coherency matrix (I, Q, U, V)
- Complex numbers of 2 floats = 8 bytes

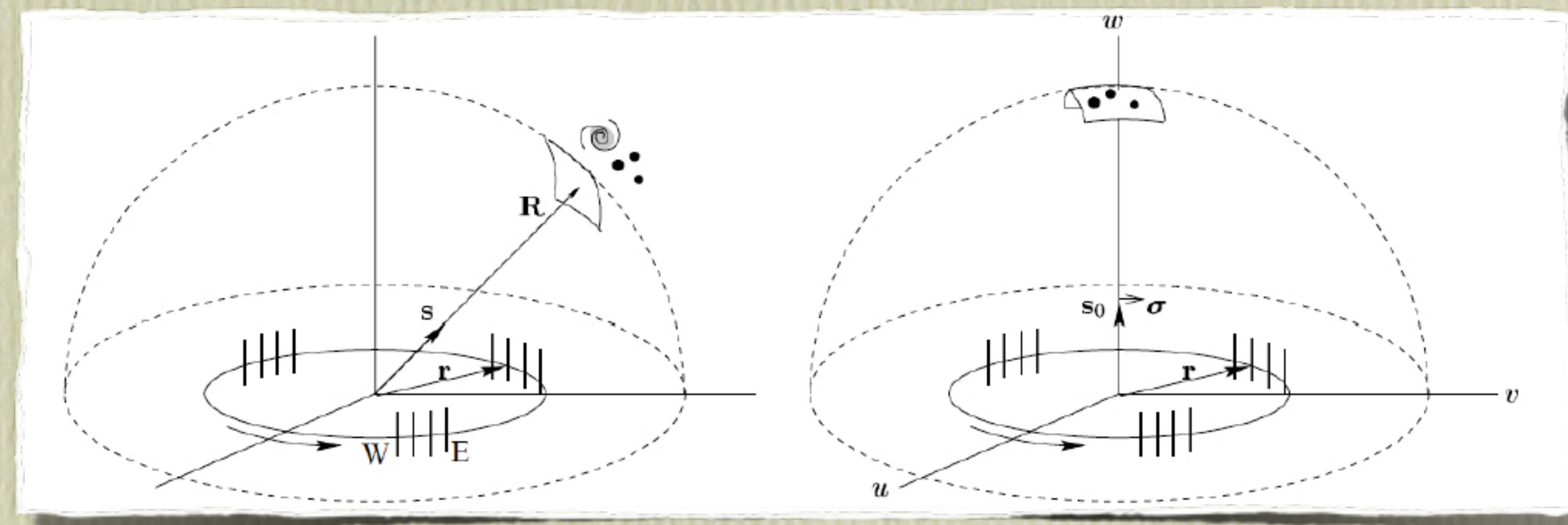


Observing Strategy

Data Volume: 1.5 petabyte = 1500 TB hard-discs

100 TB after averaging to 100 kHz channels

Short Mathematical Data Model: The Measurement Equation



Under the assumption of small angles, far-field, incoherent radiation and **no polarization**, the relation between the visibility (i.e. the correlator output) and the sky-brightness (I) and antenna beam (E) is:

$$V_f\,(u,v) = \iint \mathcal{E}_f(l,m) I_f(l,m) \mathrm{e}^{-i\,(ul+lm)} \mathrm{d}l \mathrm{d}m$$

Classical way of looking at the problem

$$\mathbf{r}_i - \mathbf{r}_j = \lambda[u,v,0], \qquad \lambda \equiv \frac{c}{2\pi\nu}.$$

Short Mathematical Data Model:
The Measurement Equation

However, the sky and instrument are intrinsically polarized and the electric field has to be described by a 2x1 complex vector and the visibility by a complex coherency matrix.

Jones vector

$$\mathbf{e}\left(t\right) = \left(\begin{array}{c} E_x\left(t\right) \\ E_y\left(t\right) e^{i\delta(t)} \end{array} \right)$$

Coherency matrix

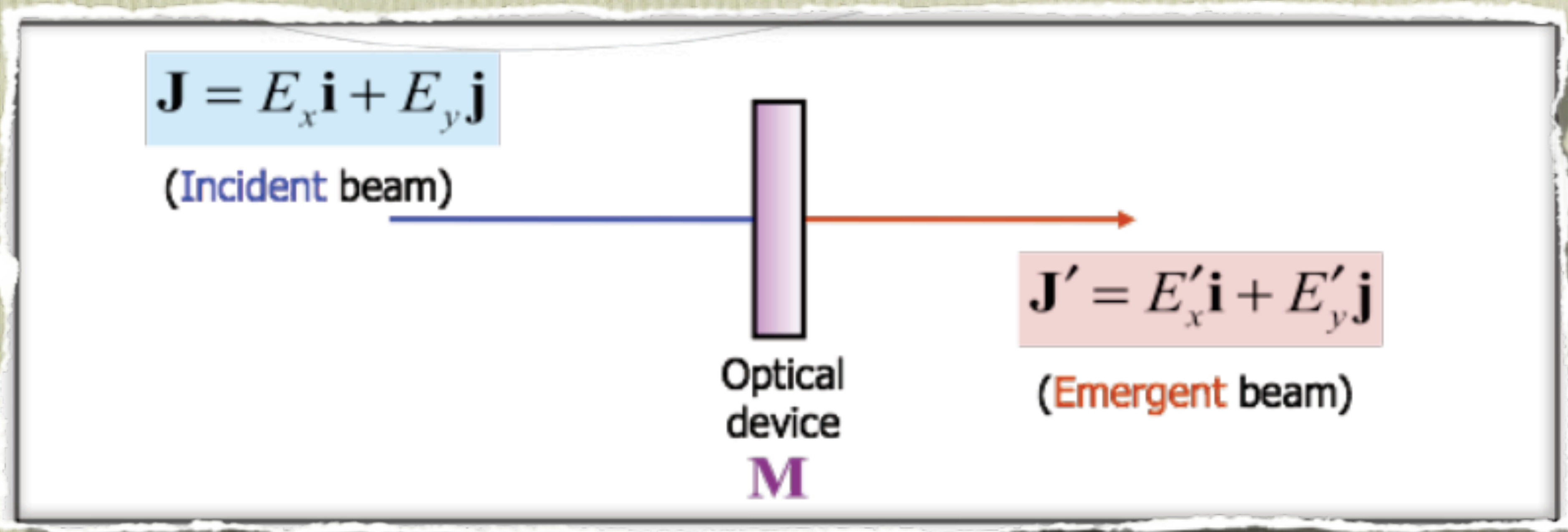
$$\mathbf{C} \equiv \left\langle \mathbf{e}\left(t\right) \otimes \mathbf{e}^{\dagger}\left(t\right) \right\rangle = \left(\begin{array}{cc} \langle e_1\left(t\right) e_1^*\left(t\right)\rangle & \langle e_1\left(t\right) e_2^*\left(t\right)\rangle \\ \langle e_2\left(t\right) e_1^*\left(t\right)\rangle & \langle e_2\left(t\right) e_2^*\left(t\right)\rangle \end{array} \right)$$

Coherency matrix
(in Stokes parameters)

$$\mathbf{S} = \frac{1}{2} \left(\begin{array}{cc} I+Q & U-iV \\ U+iV & I-Q \end{array} \right) \equiv \mathbf{C}$$

Short Mathematical Data Model: The Measurement Equation

Jones Matrices (J) describe physical/instrumental effects that modify the electric fields as (thought to be) measured by the dipoles. The coherency matrix changes accordingly:



$$\mathbf{C}' = \left\langle \mathbf{e}' \otimes \mathbf{e}'^{\dagger} \right\rangle = \left\langle (\mathbf{J}\mathbf{e}) \otimes (\mathbf{J}\mathbf{e})^{\dagger} \right\rangle$$

$$= \left\langle \mathbf{J}\mathbf{e} \otimes \mathbf{e}^{\dagger}\mathbf{J}^{\dagger} \right\rangle = \mathbf{J} \left\langle \mathbf{e} \otimes \mathbf{e}^{\dagger} \right\rangle \mathbf{J}^{\dagger} = \mathbf{J}\mathbf{C}\mathbf{J}^{\dagger}$$

Short Mathematical Data Model:
The Measurement Equation

In a real system, there are many effects that can be described by Jones matrices:

$$\mathbf{A}_i = \mathbf{K}_i \mathbf{B}_i \mathbf{G}_i \mathbf{D}_i \mathbf{E}_i \mathbf{P}_i \mathbf{T}_i \mathbf{F}_i$$

$$\mathbf{A}_j^\dagger = \mathbf{F}_j^\dagger \mathbf{T}_j^\dagger \mathbf{P}_j^\dagger \mathbf{E}_j^\dagger \mathbf{D}_j^\dagger \mathbf{G}_j^\dagger \mathbf{B}_j^\dagger \mathbf{K}_j^\dagger$$

Effective Jones Matrix

F: Ionospheric Faraday Rotation
P: Parallactic Angle
E: Antenna Voltage Patterns
D: Polarization Leakage and Instrumental Polarization
G: Complex baseline-based electronic gain
B: Bandpass
K: Fourier kernel

Short Mathematical Data Model:
The Measurement Equation

Every visibility can then be written as the sum over emission elements (C) distributed over the sky (in this case described by a set of discrete point sources)

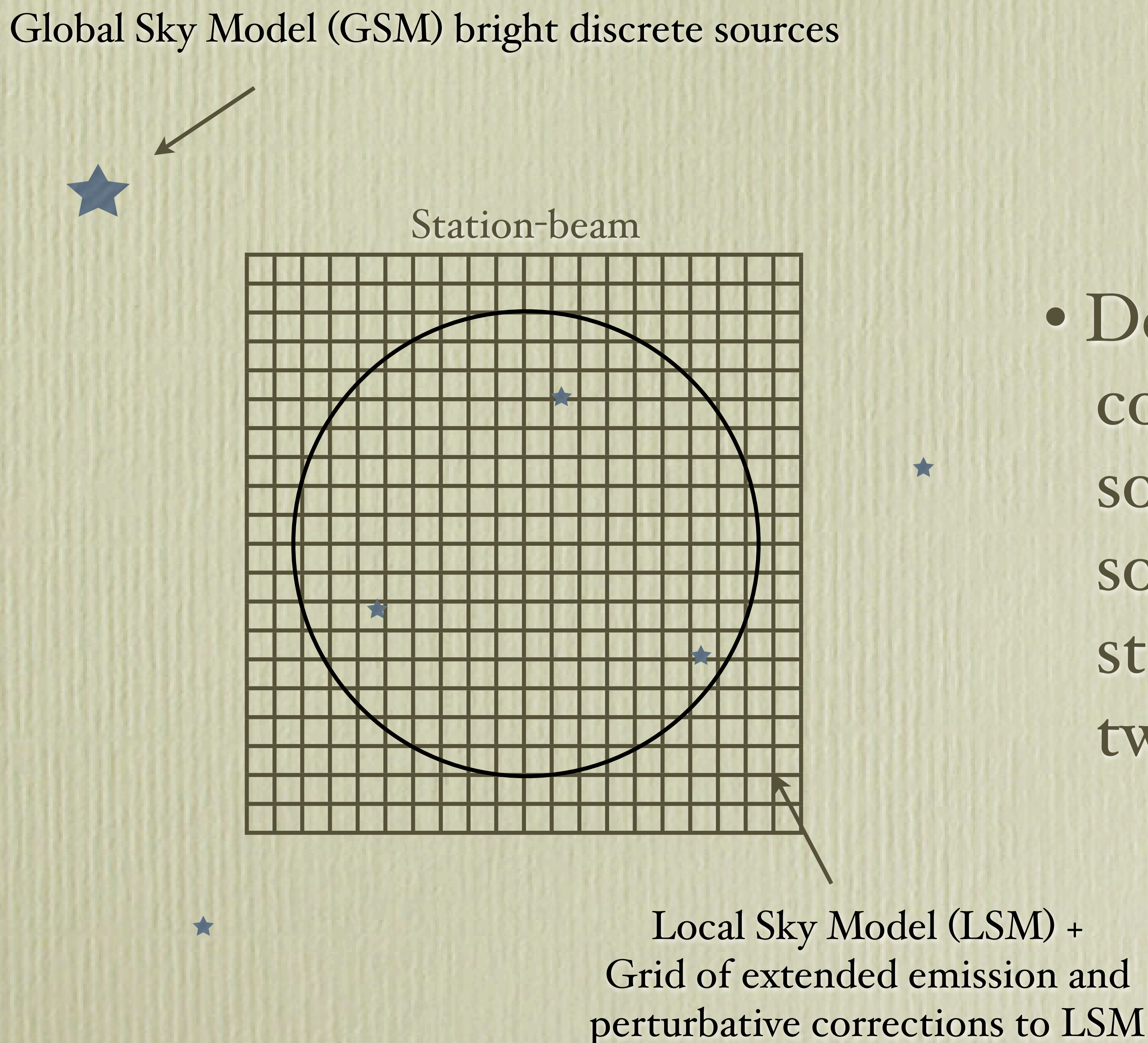
$$V_{ij}^{\mathrm{obs}} = \sum_{l,m} \mathbf{A}_i\,(l,m)\,\mathbf{C}\,(l,m)\,\mathbf{A}_j^{\dagger}\,(l,m)$$

Describes the ionosphere/intrument

The visibility data

Global/Local Sky models
Grid-based models, etc.
(anything that emits EM radiation)

Short Mathematical Data Model



- Description of the sky can consist of brightest GSM sources, discrete LSM sources plus grid for structure not part of these two components

Short Mathematical Data Model

Every visibility (coherency matrix element) can be written as a linear superposition of the entire sky-brightness distribution (in I, Q, U & V), i.e.

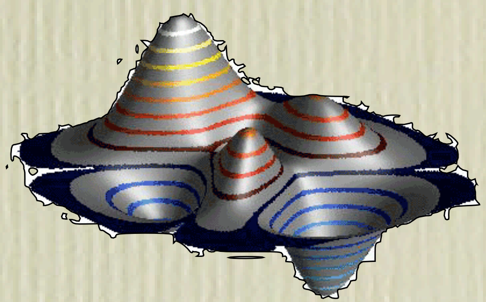
$$\mathbf{v} = \mathbf{A}(\mathbf{p})\mathbf{s} + \mathbf{n} \qquad \mathbf{s} = \begin{pmatrix} \mathbf{s}_{\mathrm{GSM}} \\ \mathbf{s}_{\mathrm{LSM}} \\ \mathbf{s}_{\mathrm{grid}} \end{pmatrix}$$

Classical "Clean-Selfcal" loop iteratively solves **s** (through Cleaning) and **p** (through Self-calibration) until convergence.

This works well if the sky is nearly empty.
However for the EoR KSP the entire sky is filled!

Short Mathematical Data Model: Reprocessing = Calibration

Solving for **p** is a highly non-linear process bound to converge to secondary minima if not carried out carefully.



Reprocessing:
i.e. finding a good initial solution of **p** for all instrument and sky effects using a modified clean-calibration loop and a simple model for **s** (e.g. bright calibrator sources):

What does **p** contain:

1. Bandpass calibration
2. Dipole rotations
3. Complex Telescope Gains
4. Complex omni-directional beam
5. Ionospheric phase fluctuations
6. Faraday rotation
7. What ever else might be out there....

Short Mathematical Data Model: Inversion = Determine Sky Brightness

Once **p** is known within small errors, **s** can be solved even if the sky is filled. The ML solution results from:

$$(\mathbf{A}^{\mathrm{T}}\mathbf{C}_{\mathrm{N}}^{-1}\mathbf{A}_{\mathrm{N}} + \lambda\mathbf{R}^{\mathrm{T}}\mathbf{R})\mathbf{s} = \mathbf{A}^{\mathrm{T}}\mathbf{C}_{\mathrm{N}}^{-1}\mathbf{v}$$

This is computationally <u>very</u> demanding but provides the solution that maximizes the likelihood of the data (i.e. visibilities).

Once **s** is known, **p** can be perturbed along dominant eigen-vectors (e.g. determined through simulations) in several iterations leading to convergence (e.g. subspace techniques).

The LOFAR EoR-KSP: Open Issue: Sparseness

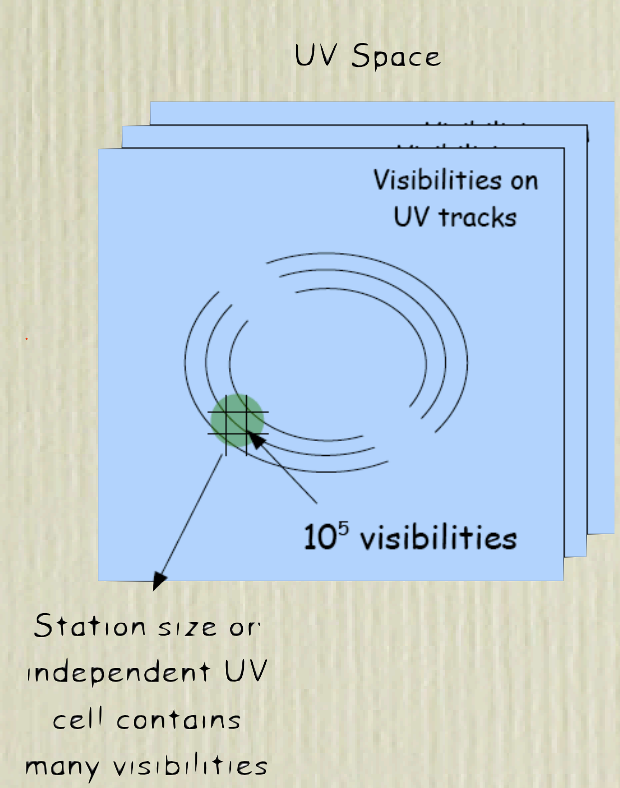
If **s** is regarded, not as the sky, but as the true underlying FT of the sky, then **A(p)** is relatively sparse (True in case of ionosphere?).

Of the 10^9 columns (# coherence elements) in **A**, only a fraction of 2×10^{-4} elements per row are non-zero.

$$V(\vec{u}) = \int \mathcal{A}(\vec{x})B(\vec{x})e^{-2\pi i(\vec{x}\cdot\vec{u})}d\vec{x} = \mathrm{FT}(\mathcal{A}) \star \mathrm{FT}(\mathrm{B})$$

The FT of the complex beam only correlates over a scale of ~ station size / core size ~ 1/80 in UV-plane. We model the UV-plane as a 256 x 256 grid.

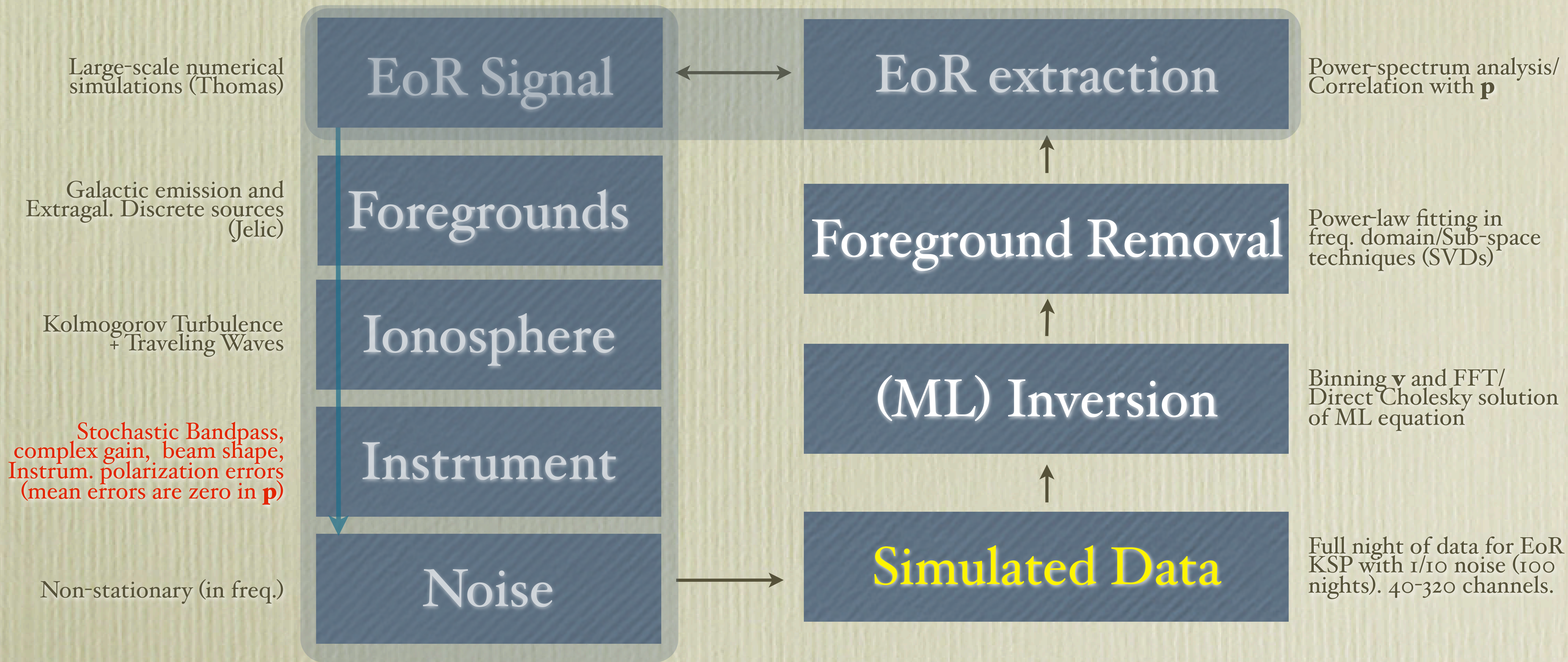
(Hence an average fraction of ~1/6400 ~ 2×10^{-4} is non-zero)



The LOFAR End-to-End Data Pipeline

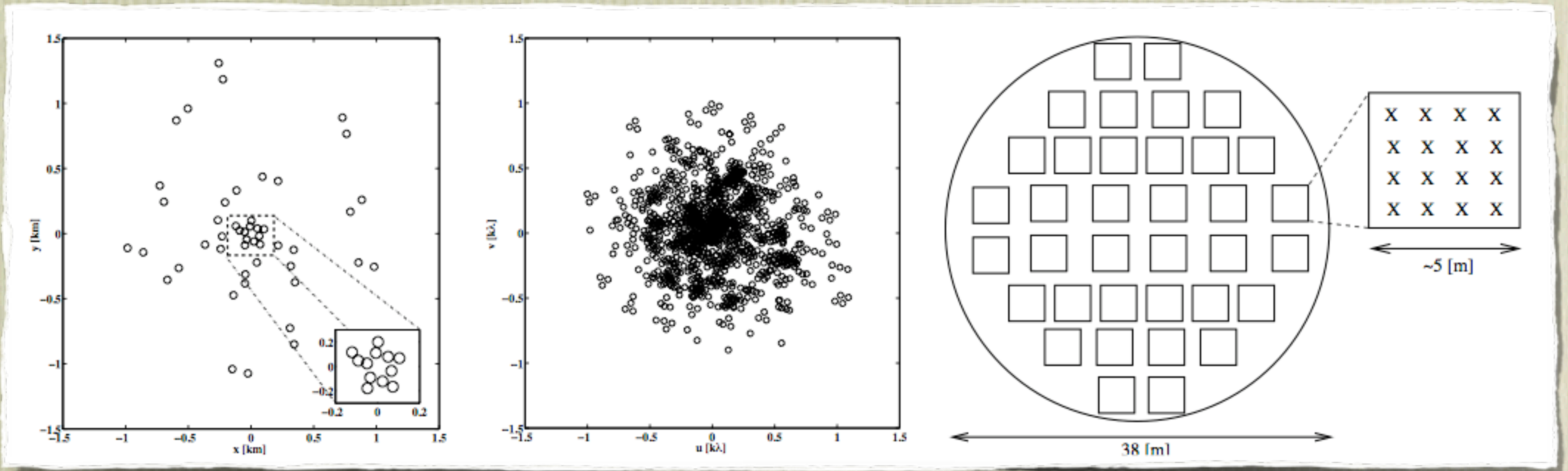
The LOFAR EoR-KSP: Simulations & Inversions

All these challenges require an end-to-end simulation.

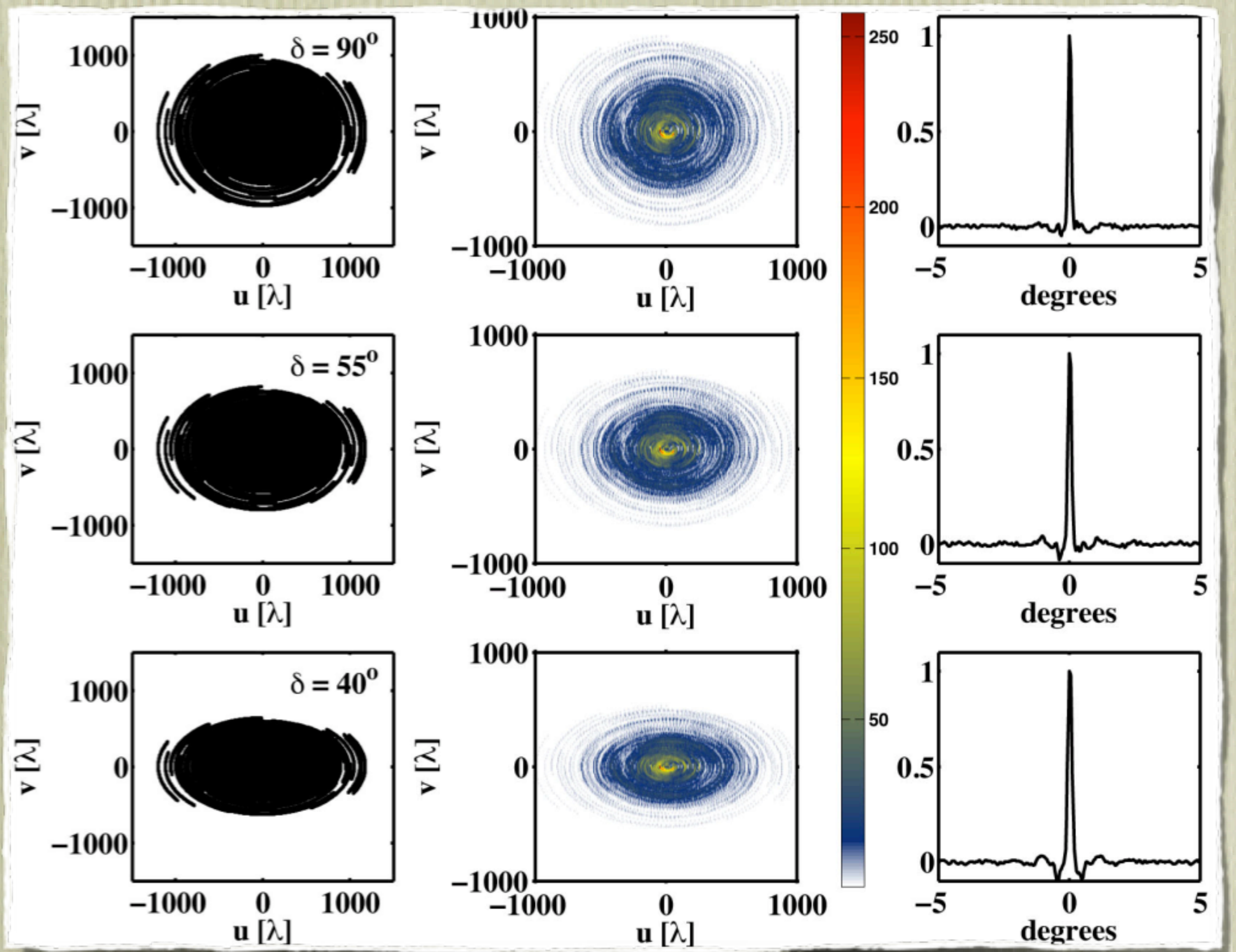


The LOFAR EoR-KSP: Simulations & Inversions

For detection of the EoR 21-cm only the inner 3-5 km baselines are critical (good uv-coverage). Longer baselines are necessary to detect and subtract point-sources below the confusion level (~mJy/beam at 3 arcmin).



The LOFAR EoR-KSP: UV-coverage core + Inner Rings



Good UV coverage is necessary for a well-behaved beam, but also for a proper sampling of the EoR power-spectrum on relevant scales

(note that uv-plane is a Fourier sampling of the sky-brightness power- spectrum)

Labropoulos et al. 2009, subm.

Good UV Coverage

Avoid (moving) UV holes!
Lead to loss of data

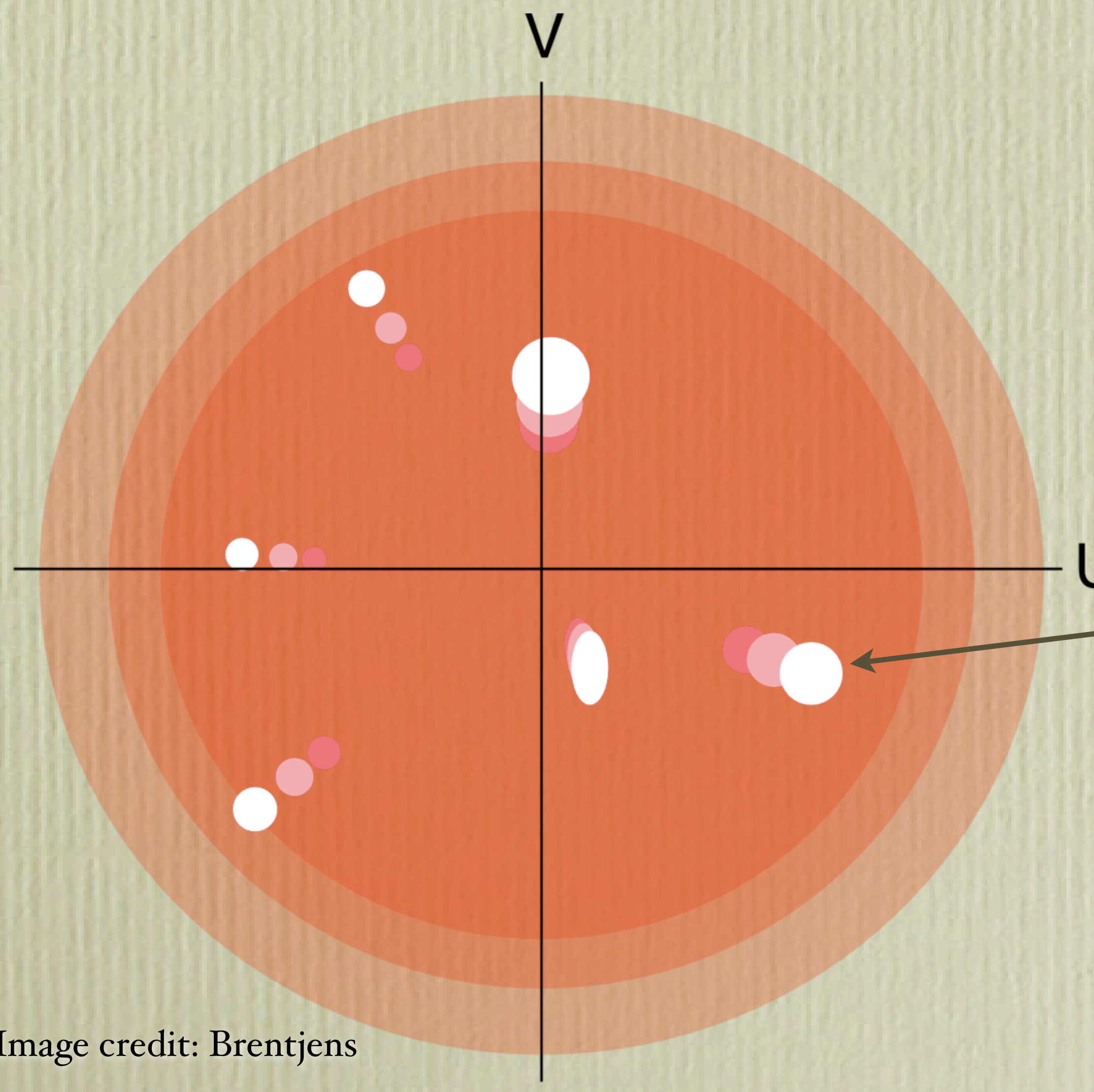
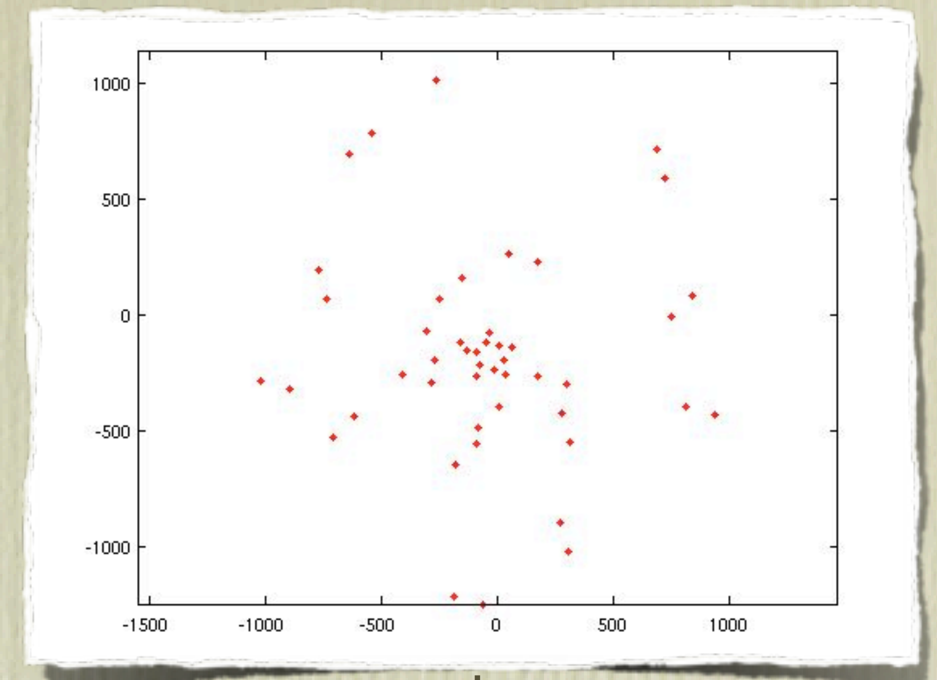
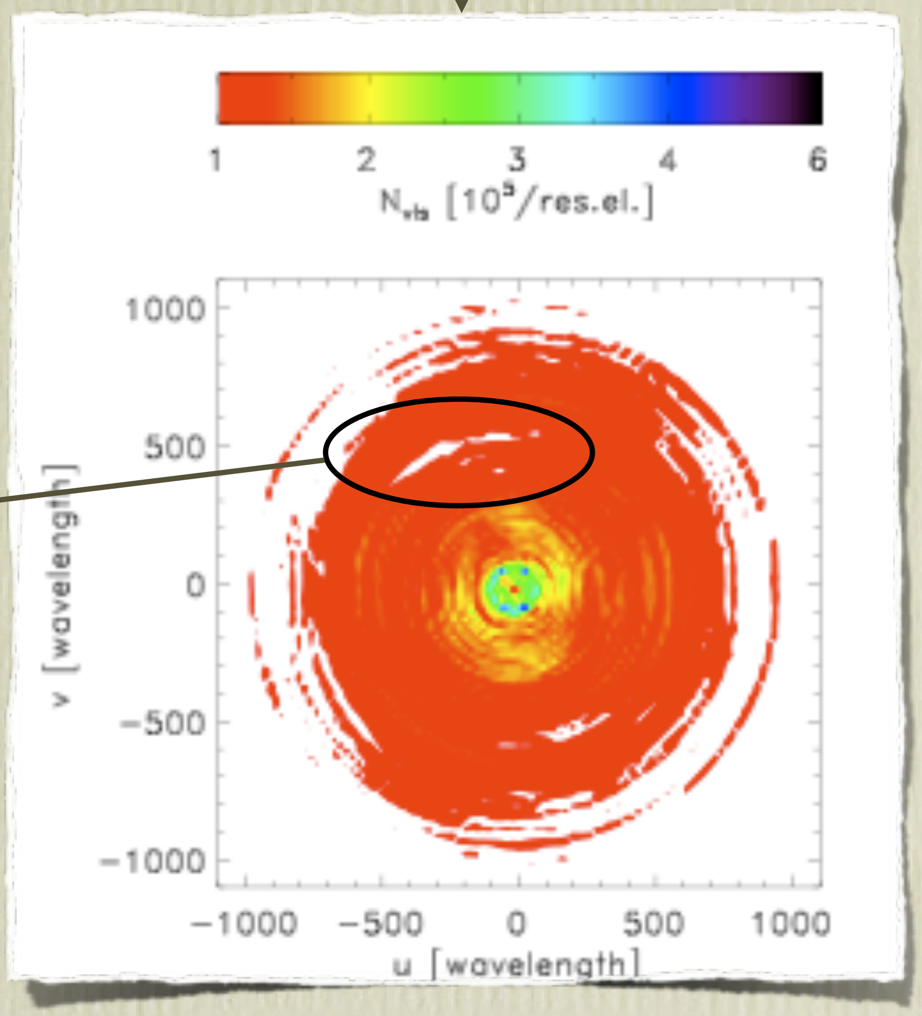


Image credit: Brentjens

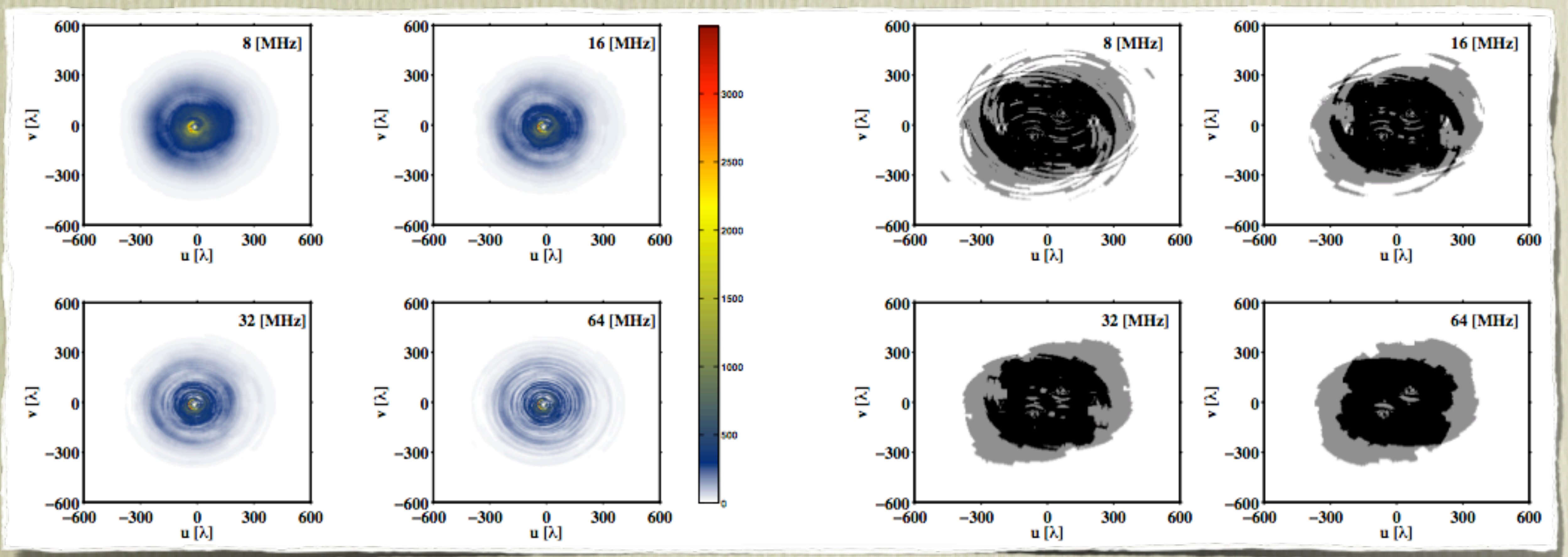




The LOFAR EoR-KSP: UV-coverage core + Inner Rings

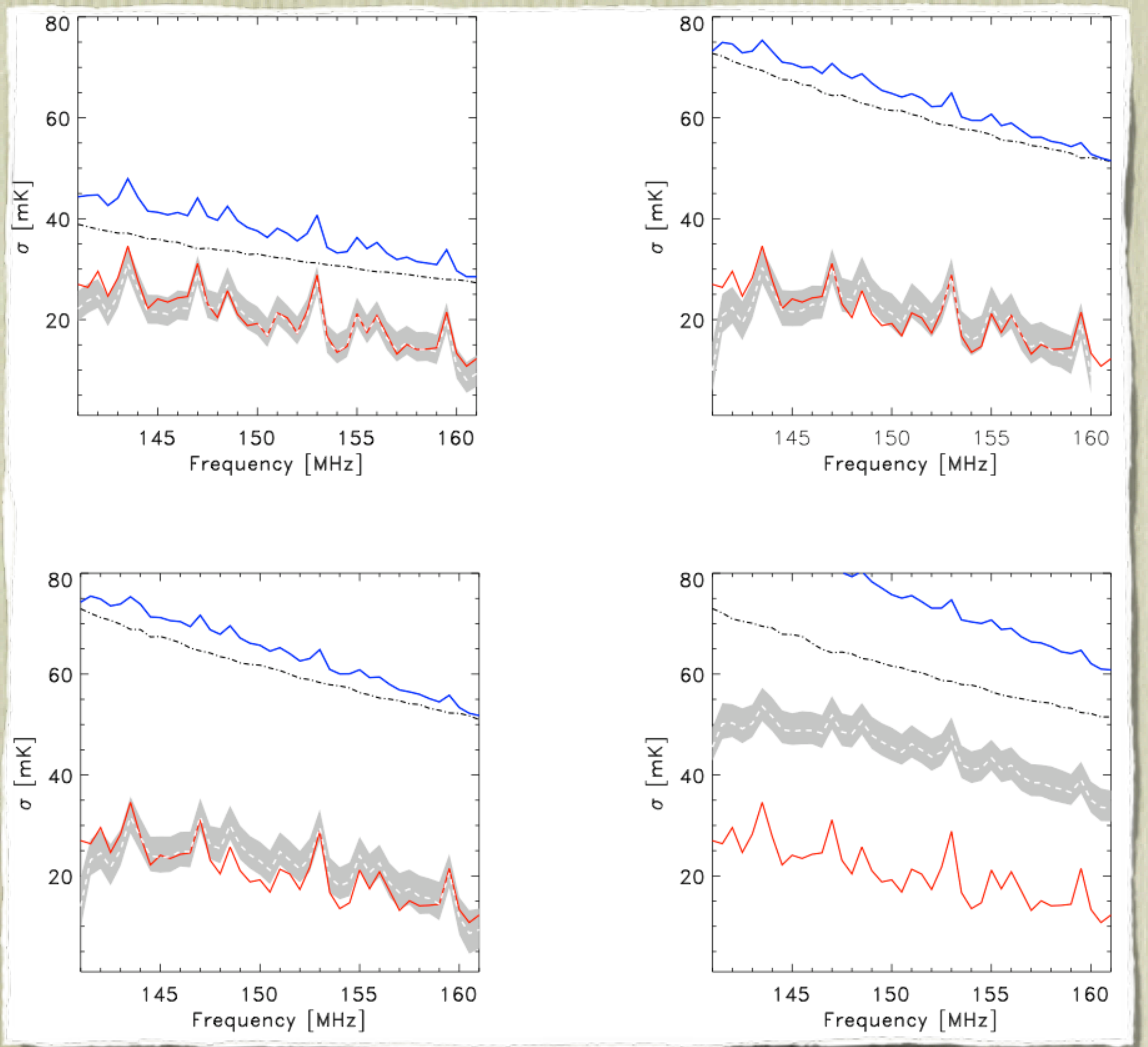
"Holes" in uv-plane move as function of frequency, carving out a slice in the joint redshift-angular-scale space.

This leads to loss of information and possible leakage of spatial and redshift information



Labropoulos et al. 2009, subm.

The LOFAR EoR-KSP: Noise & Calibration Errors



Labropoulos et al. 2009, subm.

Signal extraction with 0.5, 1-2, 4 and 10% calibration errors
(top left to bottom right)

Blue: Measured signal
Dashed: Instr. Noise
Red: True EoR Signal
White/Gray: Extracted signal plus rms error

The LOFAR EoR-KSP: A full end-to-end simulation

Current status of end-to-end simulations:

- EoR Signal is 5 x 5 deg. simulation (simple) or based on BEARS model/num. sims.
- Sky: same as before, but all foregrounds are included - point sources (based on 3C196 field) plus model of extended (extra)galactic emission
- Polarization is included in the sky models (and instrument!)
- Many ionospheric effects are taken into account (Kolmogorov turbulence, TIDs etc)
- Analytic complex beam model (with differences between dipoles accounted for) varying with time/freq. (and stochastically).

- Jones matrix parameters are set according to statistical (ARMA) model based on current knowledge from WSRT/LOFAR.
- UV plane is sampled for LOFAR baselines with precise geometric model of the instrument (all Jones matrices are updates for each UV point in all directions).

- Calibration is done blindly in Meqtree to recover the input values (ongoing!)

- Next step is to extract the EoR signal to see how well it is recovered.

The LOFAR Computational Challenge

LOFAR EoR-KSP Challenges

The LOFAR Epoch-of-Reionization Key Science Project, with the goal to detect neutral hydrogen at z=6-11, is one of the most challenging radio-astronomical project to date.

Data volume	~1.5 petabyte raw data
Data model	very complex ME
Data reprocessing and analysis	flops ~10^{21-22}
Signal Extraction	dynamic range ~ 10^{6-7}

The LOFAR EoR-KSP: Computational Effort

Per channel (all Stokes) and per iteration

Operation	Computations	Time
$A^TC^{-1}A$	10^{15} cmad	...
$A^TC^{-1}v$	2×10^{10} cmad	...
Solving ML eqn.	5×10^{14} cmad	...
Per channel*	6×10^{16} flops	10 min
All channels*	1.2×10^{21} flops	140 days

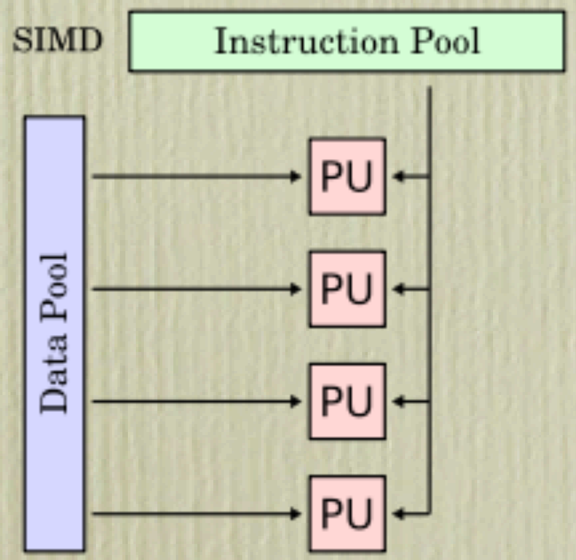
*We assume a computational efficiency of 10%, a 100 Tflop/sec computer (100% up-time), 4 flops per complex multiply/add (cmad).

Note that during reprocessing, for every calibrator source, only 10^{10} visibilities/channel need to be predicted ($\sim 10^{11-12}$ flops) and based on **p** is modified. Much faster than the inversion because the sky is assumed to be nearly empty in this process.

The LOFAR EoR-KSP: Computational Solution

Where to get 100 Tflop/s sustained computing power for 1-2 years ?

Since our problem is largely linear,
Graphics Processor Units (GPU)
are the ideal hardware solution



- Single-Instruction-Multiple-Data (SIMD) architecture
- GPUs have already up to 240 streaming processors and are cheap
- Each processor can perform one multiply/add per clock-cycle
- Linear algebra is multiply/adds on parallel data streams (i.e. vectors)
- I/O is not an issue for computationally-dominant problems

The LOFAR EoR-KSP: Computational Solution

Currently proposed solution

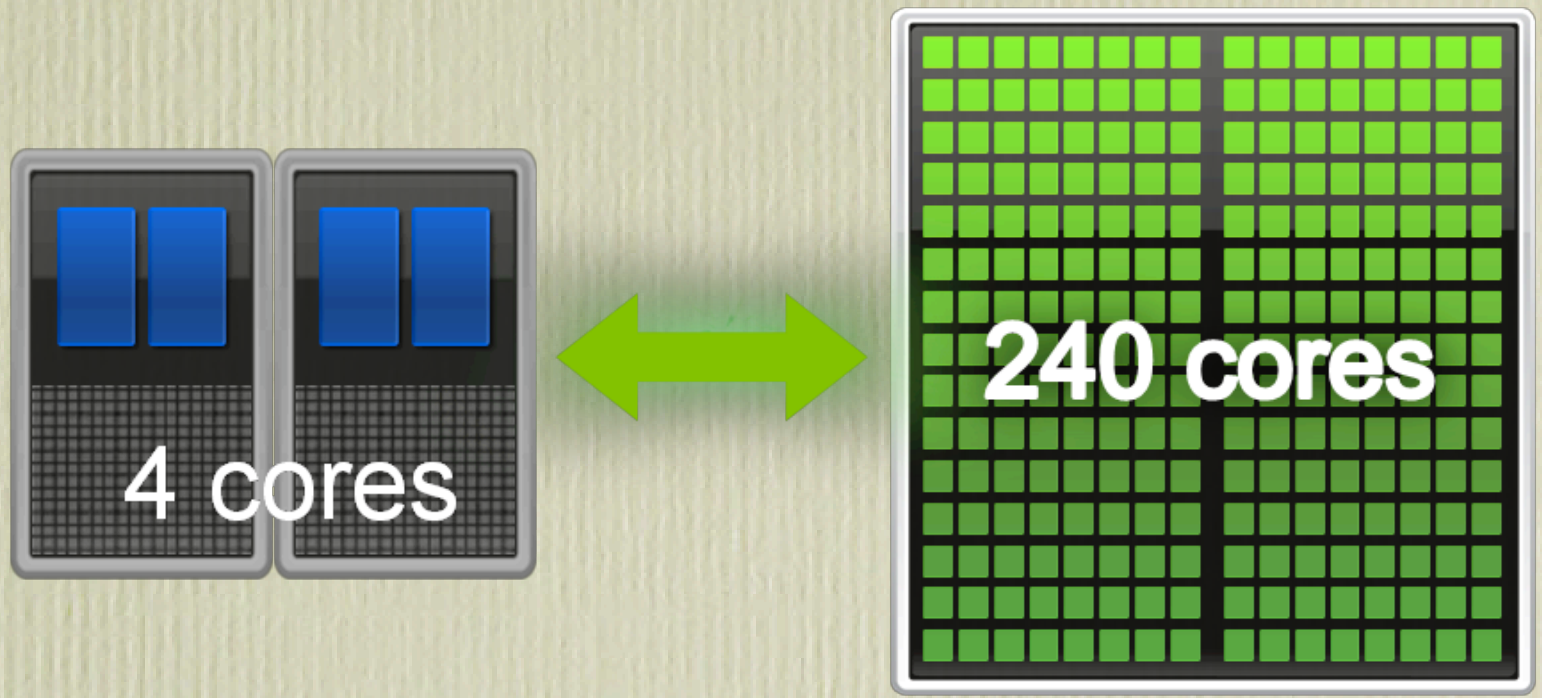
(Dutch Science Foundation (NWO-m) proposal has been accepted; PI: Koopmans)

- 40 NVIDIA Tesla-S1070 units
- 40 quad-core PC + 2 x PCI 16x
- ~400+ GB memory CPU/GPU
- ~160 TB HD space

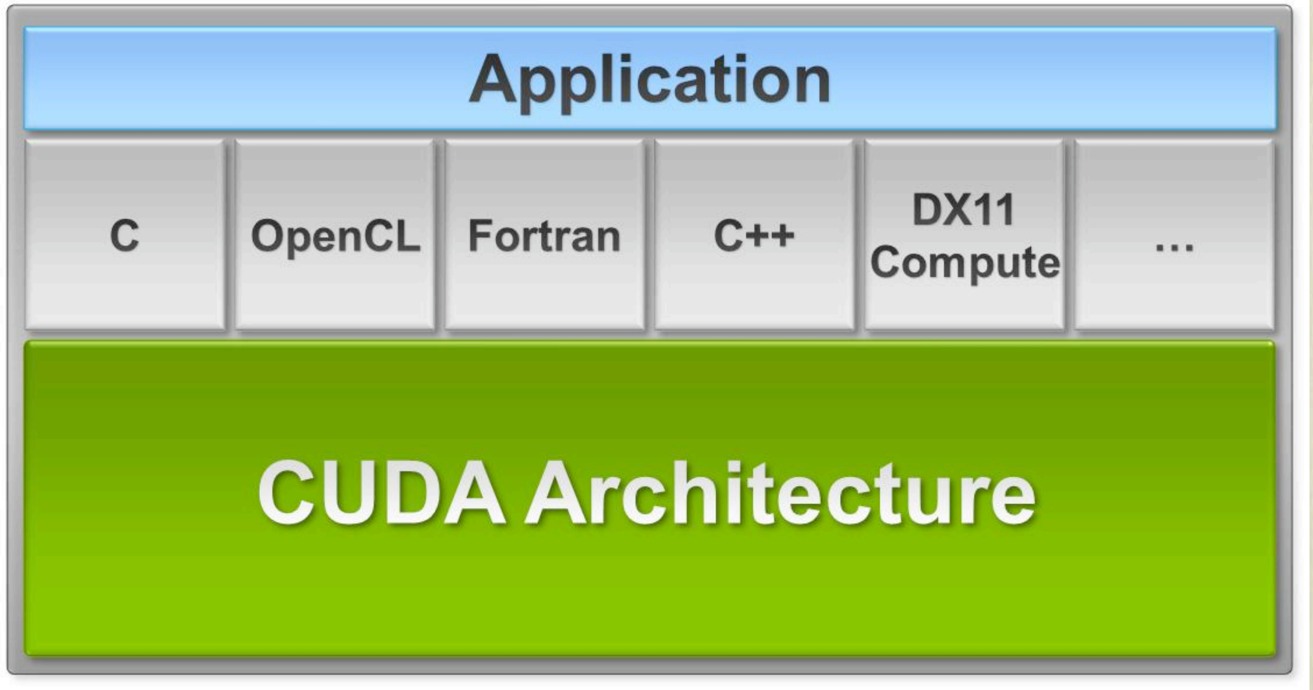
With this cluster we can reach a sustained ~160 Tflop/s peak computing power!
(40 x 4 Tflop/s/unit)

CPU+GPU Cluster: based on new NVIDIA S1070

Hardware:
Map applications of CPU-GPU architecture

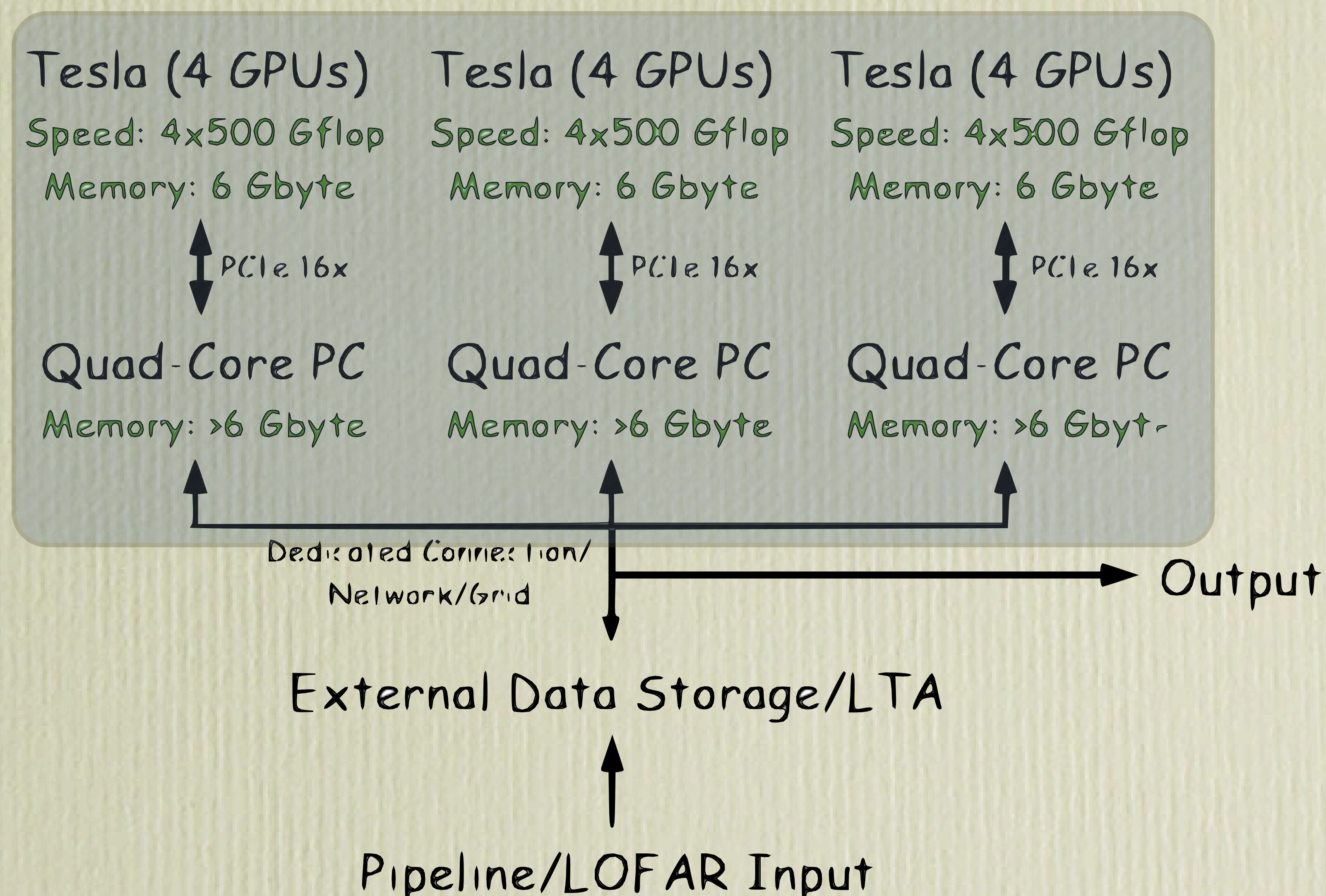


Software:
CUDA parallel computing model



The LOFAR EoR-KSP: Current CPU+GPU Test-Setup

3-unit NVIDIA S870 cluster at the Kapteyn Astronomical Institute to be scaled to ~40 units (4 TFlop/sec each)



CPU+GPU Test-Setup: Speed-up

Speedup
(S870/1CPU)

Credit: Panos

1000

100

10

1

0.1

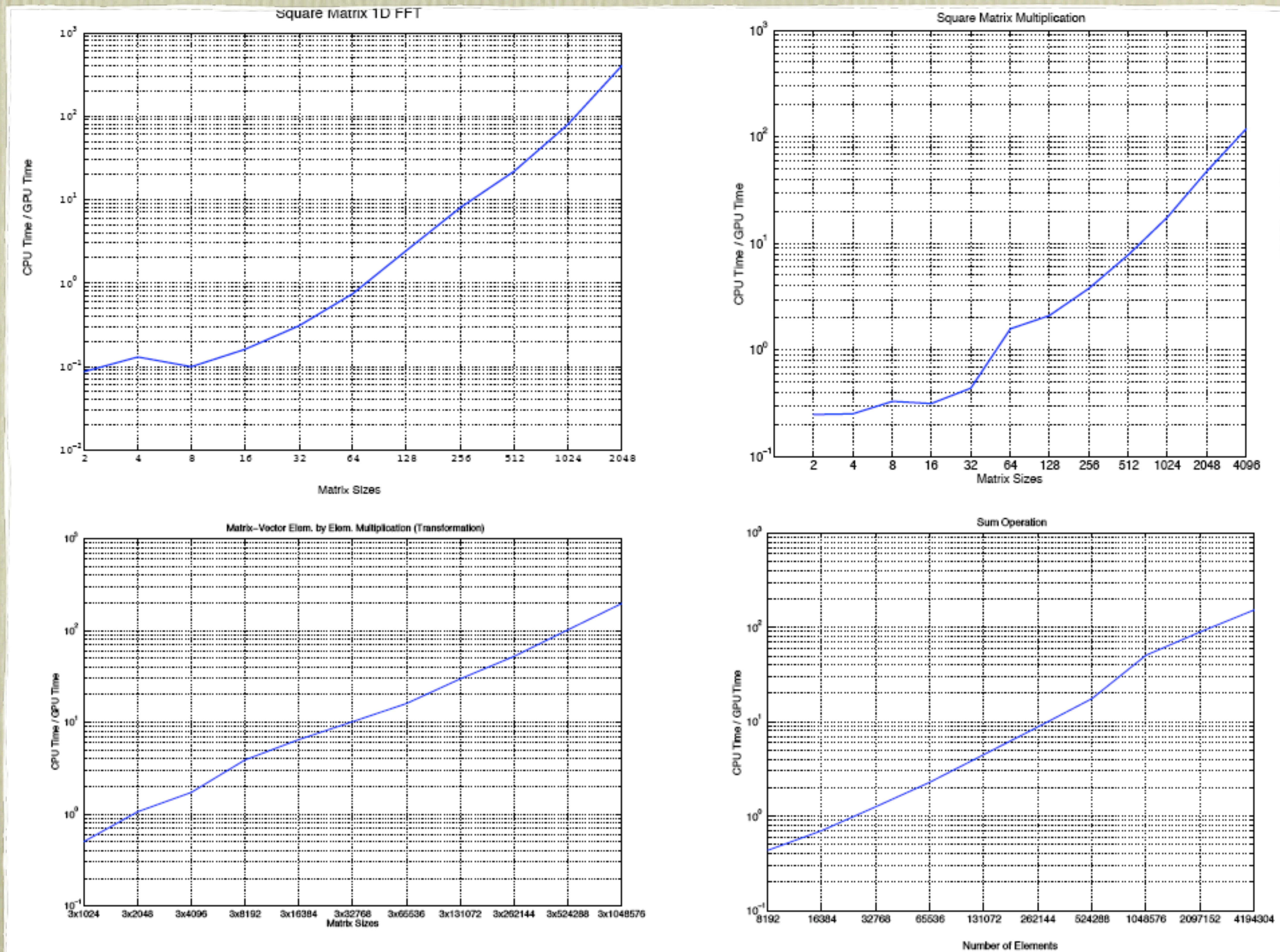
1000

100

10

1

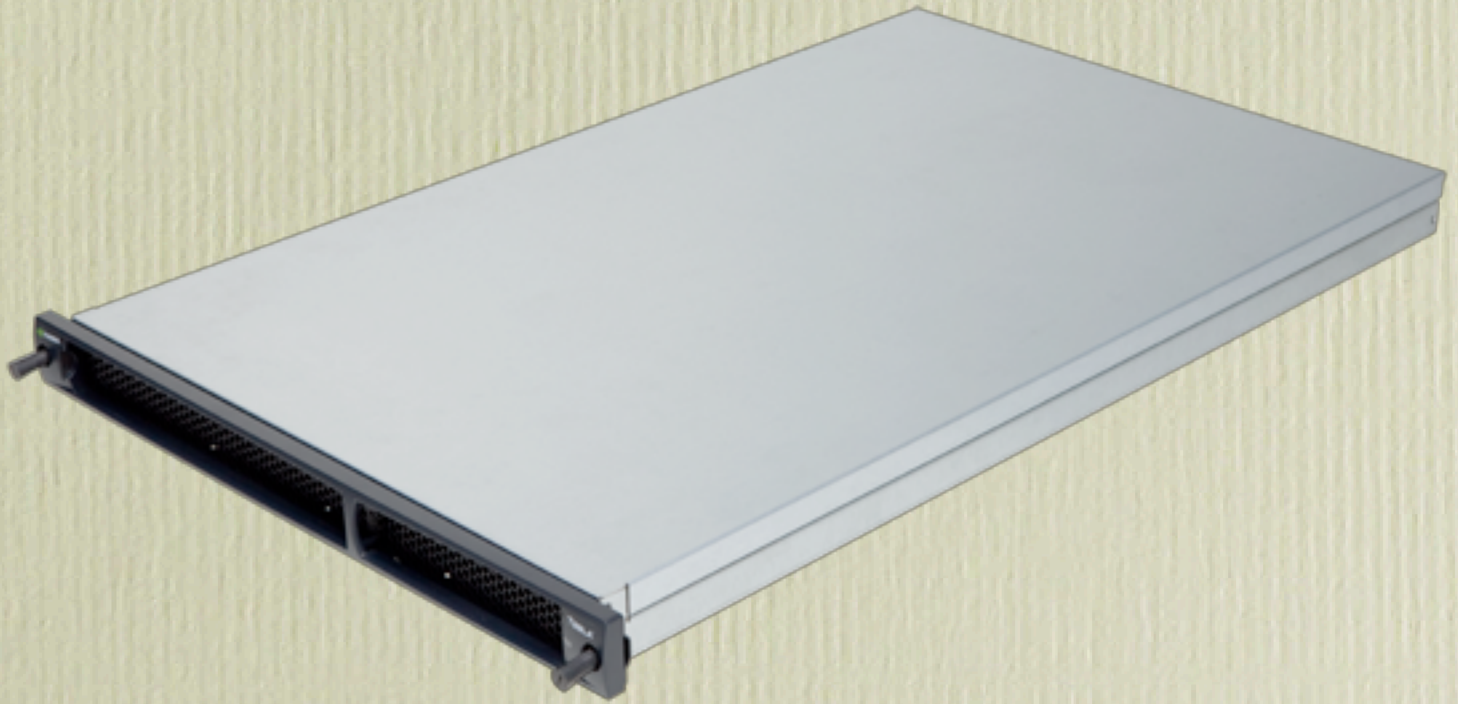
0.1



CPU+GPU Cluster

Tesla S1070

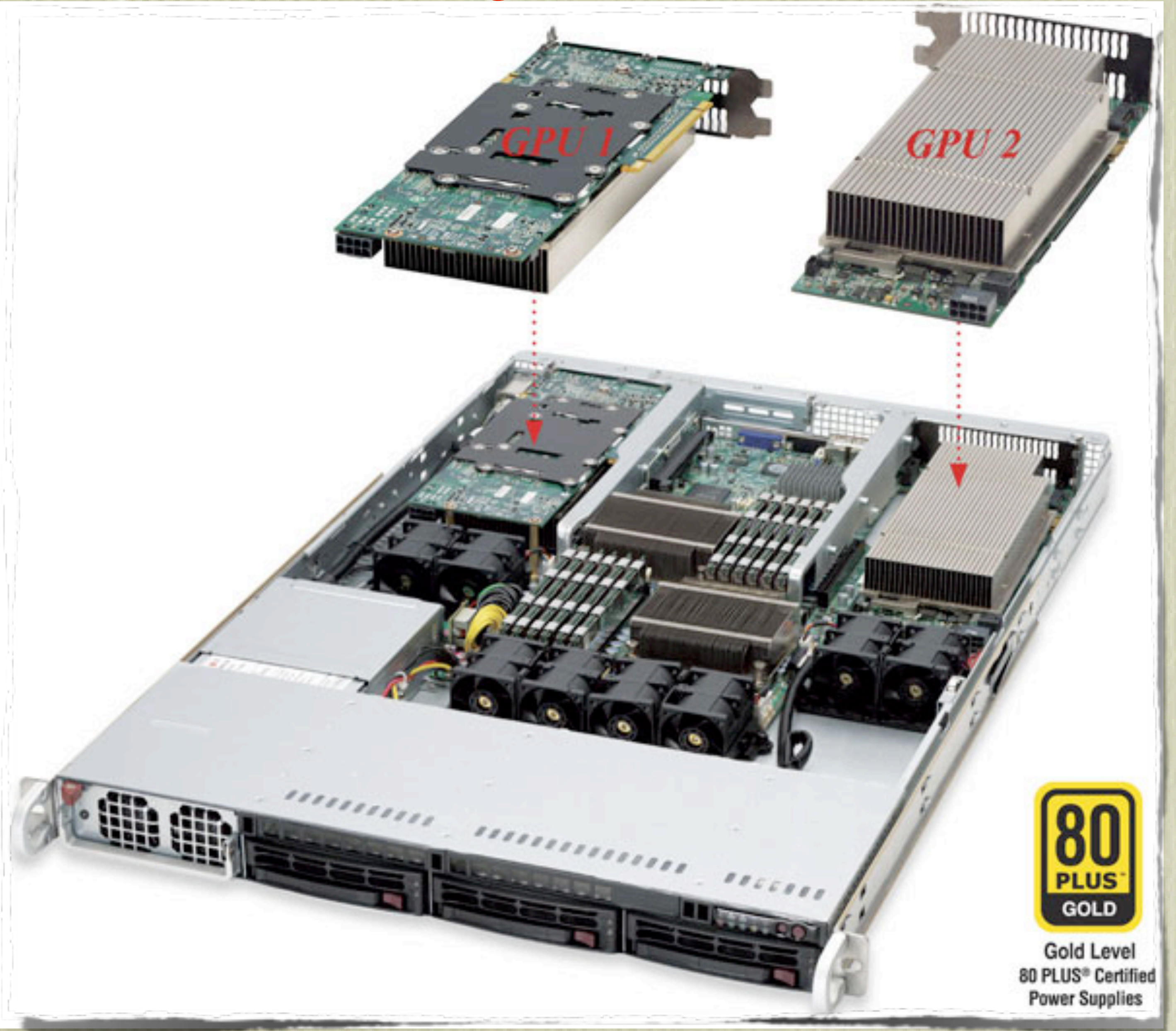
The NVIDIA® Tesla™ S1070 Computing System (Dual PCI Express 2.0 cable connections) is a four-teraflop 1U system powered by the world's first one-teraflop processor.



With 960 processor cores and a standard C compiler that simplifies application development

CPU+GPU Cluster

Supermicro: integrated CPU/GPU unit



Western Digital SATA HDD 2TB WD20EADS Caviar Green Desktop SATA 3.0GB/s 7200rpm 32MB 3.5-in Bulk (3x)

Intel Xeon Processor E5520 LGA1366 Quad-Core 2.26GHz 8MB 5.86GT 80W AT80602002091AA Tray (2x)

Supermicro GPU SuperServer 6016GT-TF-TC2 1U DP Xeon 5500 LGA1366 Quad-Core 2xGPU DDR3 3x3.5-in SATA2 Hot-Swap 1400W 80PLUS Black (1x)

DDR3 PC3-10600 1GB 240-pin DIMM DDR3-1333 Registered ECC 1.5V Brand Name (12x)

1U = 8CPU + 2GPU +12 GB mem + 6 GB HD

New features f the S1070 beyond the S870

Removes possible bottleneck in I/O chain of *Network-HD-Memory-GPU*

FEATURES AND BENEFITS

TERAFLOP PROCESSORS	Delivers up to four teraflops of performance in a 1U rack-mount system giving Tesla products unmatched performance per unit of volume.
MASSIVELY-PARALLEL, MANY-CORE ARCHITECTURE	240 processor cores per processor with the ability to execute thousands of concurrent threads.
IEEE 754 FLOATING POINT PRECISION	Ensures your results meet industry standard precision.
DOUBLE-PRECISION MATH	Meets the precision requirements of your most demanding applications.
ASYNCHRONOUS TRANSFER	Turbocharges system performance by executing data transfers, even when the computing cores are busy.
SYSTEM MONITORING FEATURES	Simple management and monitoring post-installation helps your IT staff manage systems with ease. Remote capabilities as well as status lights on the front and rear of the unit ensure your staff can see the status whether they are on the other side of the rack or the other side of the world.
DUAL GEN2 PCIe CABLE CONNECTIONS	Maximizes bandwidth between the host system and the Tesla processors, with up to 12.8 GB/s peak transfer rates.
GEN2 PCIe CABLE WITH SMALL-FORM-FACTOR (SFF) HOST ADAPTER CARD	Enables Tesla systems to work with virtually any PCIe-compliant host system with an open PCIe slot (x8 or x16).

Tokyo Tech Builds First Tesla GPU Based Heterogeneous Cluster To Reach Top 500

Printer Friendly

NVIDIA Tesla Powers 29th Most Powerful Supercomputer in the World

For further information, contact:

A similar cluster is in existence!

Andrew Humber
NVIDIA Corporation
(408) 486 8138
ahumber@nvidia.com

FOR IMMEDIATE RELEASE

SC08—AUSTIN, TX—NOVEMBER 17, 2008—The Tokyo Institute of Technology (Tokyo Tech) today announced a collaboration with NVIDIA to use NVIDIA® Tesla™ GPUs to boost the computational horsepower of its TSUBAME supercomputer. Through the addition of 170 Tesla S1070 1U systems, the TSUBAME supercomputer now delivers nearly 170 TFLOPS of theoretical peak performance, as well as 77.48 TFLOPS of measured Linpack performance, placing it, again, amongst the top ranks in the world's Top 500 Supercomputers.

"Tokyo Tech is constantly investigating future computing platforms and it had become clear to us that to make the next major leap in performance, TSUBAME had to adopt GPU computing technologies," said Satoshi Matsuoka, division director of the Global Scientific Information and Computing Center at Tokyo Tech. "In testing our key applications, the Tesla GPUs delivered speed-ups that we had never seen before, sometimes even orders of magnitude - a tremendous competitive boost for our scientists and engineers in reducing their time to solution."

Speaking to the ease of implementation, Matsuoka continued,

"The entire upgrade was carried out in 1 week, and the TSUBAME supercomputer remained live throughout. This is an unprecedented feat in top-level supercomputing."

"We are honored to partner with Tokyo Tech - world famous for their supercomputing expertise and success," said Andy Keane, general manager of the GPU Computing business at NVIDIA. "NVIDIA Tesla breaking into the Top 500 marks a milestone in supercomputing history. The massively parallel GPU is now essential for supercomputing centers worldwide."

The first to achieve Top 500 ranking with an NVIDIA Tesla based GPU cluster, Tokyo Tech. is one of hundreds of distinguished universities and supercomputing centers that have adopted GPU based solutions for research. Other leading centers include the National Center of Supercomputing Applications (NCSA) at the University of Illinois, Rice University, University of Heidelberg, University of Maryland, Max Planck Institute and University of North Carolina.

The Tesla S1070 1U GPU system is based on the NVIDIA CUDA™ parallel architecture. This architecture is accessible through an industry standard C language programming environment that allows developers and researchers to tap into the parallel architecture of the GPU more quickly and easily than any other solution shipping today.

For more information on NVIDIA Tesla S1070, please visit: www.nvidia.com/object/tesla_s1070

CPU+GPU Cluster: Budget

The LOFAR EoR KSP aims at 100+ TFlop/sec computing.

Within current budget we can reach ~160 TFlop/sec (peak) !

LOFAR EoR KSP:

Current Computational/Storage Situations

44 TFlops
Bluegene/P ⟷ Temporary storage
0.5-1.0 Pb

5 TFlops
Off-line Cluster ⟷ Long-Term Archive
1.5Pb

1.4/140 TFlops
Reprocessing Cluster
CPU+GPU cluster ⟷ Internal storage
100-200 Tb

Both Reprocessing & Analysis can be possibly be done on this CPU+GPU cluster

Note difference in CPU vs GPU speed!

CPU+GPU Cluster: Complete Pipeline

Investigating how to integrate a full pipeline on this cluster

- Data Simulation/Calibration: Meqtree/Matlab/C code.
- (P)reprocessing: DP3, BBS, BBI code.
  (Less data than on offline cluster; i.e. less CPU power needed.)
- Inversion/Imaging: MATLAB/C codes.
  Gridding/compression, then FFT, ML or other inversion
- Signal Extraction: IDL/Matlab/C codes.
  (Regularized) Fitting/Wp/Filtering?

Current range of different codes/methods could be joint into a single continuous pipeline (programmer).

Summary & Future Work (I)

- A working LOFAR-EoR data-simulation pipeline is in place.
- Inversion algorithms (FFT/ML inversion) are being tested
- Extraction algorithms (pol. fitting/SVD) are being tested
- In the presence of noise and zero-mean calibration errors, the EoR signal extraction remains successful.

- A complete polarization simulations is being done.
- Bright point sources + confusion are being included.
- Corrupted data (non-zero mean) will be calibrated with MeqTree/BBS as test of reprocessing strategy.
- New effects will be added to simulations

Summary & Future Work (II)

- The simulation and inversion/extraction code will be implemented on the current GPU cluster, expandable to the full cluster. This is required for our severe computational/data-volume demands.
- So far no show-stoppers have been identified, but we need to make the simulations increasingly more complex and realistic to test every aspect of the LOFAR EoR KSP.
- Near-term goal: A full data-cube simulation plus analysis with all known effects included (blindly); calibrate in Meqtree/BBS; invert using our ML code; signal extraction with pol. fitting/SVD; compare results for different scenarios of input models.

Summary & Future Work (III)

- LOFAR-20: MSSS Survey commences in 2009/10 and will test our software and hardware and provide initial clues as to where to place our deep EoR windows (e.g. good calibration sources, no very extended sources, etc.)

- Possibly more wide-area surveys are needed to assess the level of foregrounds and polarization before deep windows are selected.

- LOFAR-40+: Deep integrations on 5 windows x 4-6 beams x 2-3 freq. combs x 300 hrs. Aim: detection of EoR in 2011-13