From LOFAR to SKA, challenges in distributed computing

Soobash Daiboo
Paris Observatory -LESIA

Overview

- LOFAR telescope
- Data processing with LOFAR
- NenuFar
- SKA
- Summary

LOFAR science drivers

Key science projects

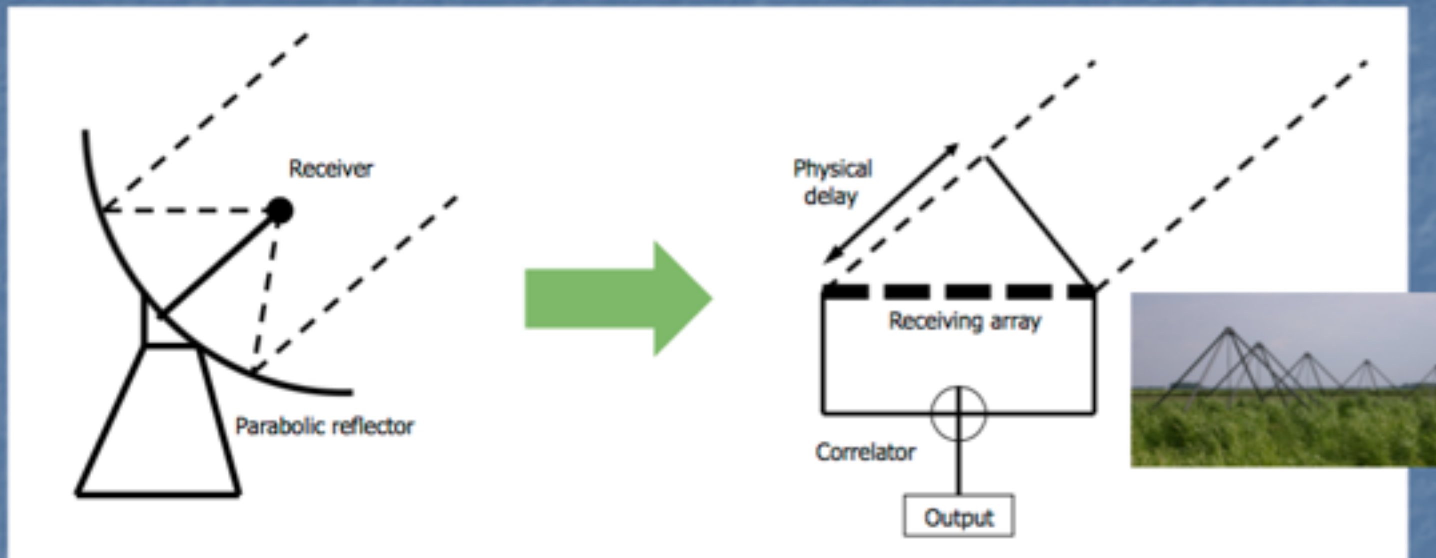
- Epoch of Reionization
- Surveys and Distant Universe
- Transients and Pulsars
- High Energy Cosmic Rays
- Cosmic Magnetism
- Solar Physics and Space Weather

International membership from countries all over the world

International LOFAR Telescope

$$\theta = 1.22\frac{\lambda}{D}$$

International LOFAR Telescope

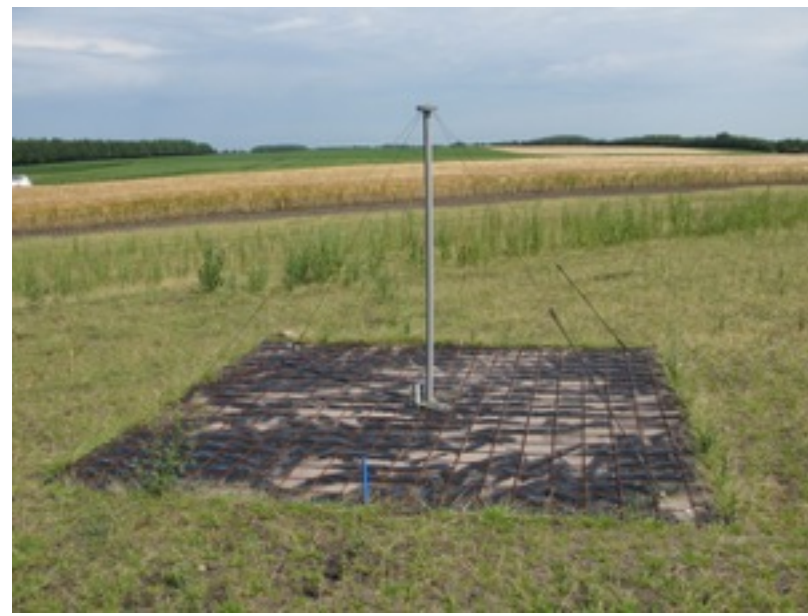


Advantages: a) replacement of mechanical beam forming by electronic signal processing
b) low frequency telescopes become economically affordable
c) multiple and independent beams can be formed at a time

$$\Gamma_{12}(u,v,0) = \iint I(l,m) e^{-2\pi i(ul+vm)}\, dl\, dm$$

LOFAR antennas

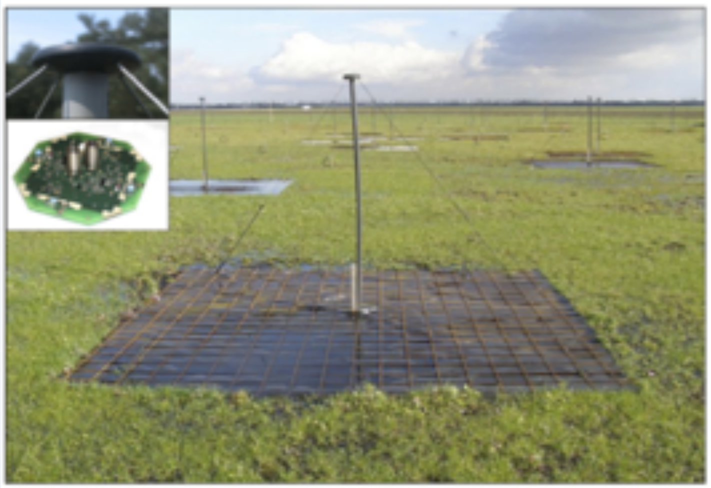
LOFAR group in Bir, Ireland, with an HBA element

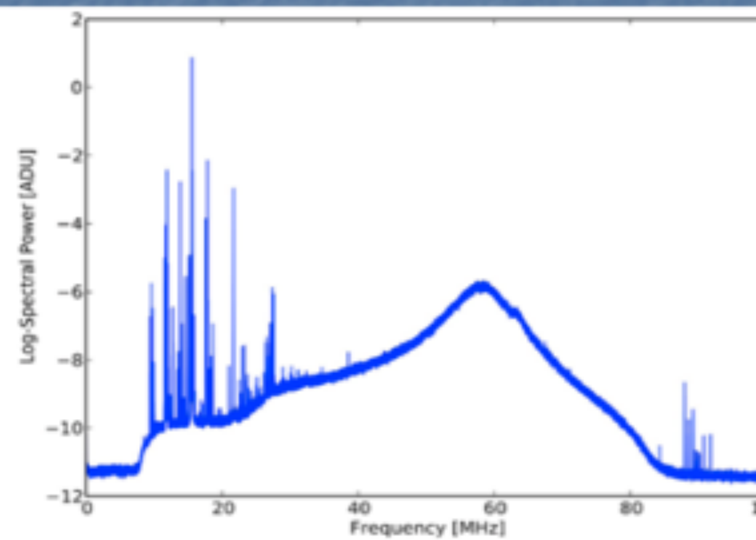


LOFAR LBA antenna in the Netherlands

LOFAR telescope - core stations

- LBA antennas: Cap containing the low noise amplifiers (LNAs), copper wires receive two orthogonal linear polarisations, ground plate
- The response curve: peak close to the resonance frequency (52 MHz) – dipole arms are 1.38 m long

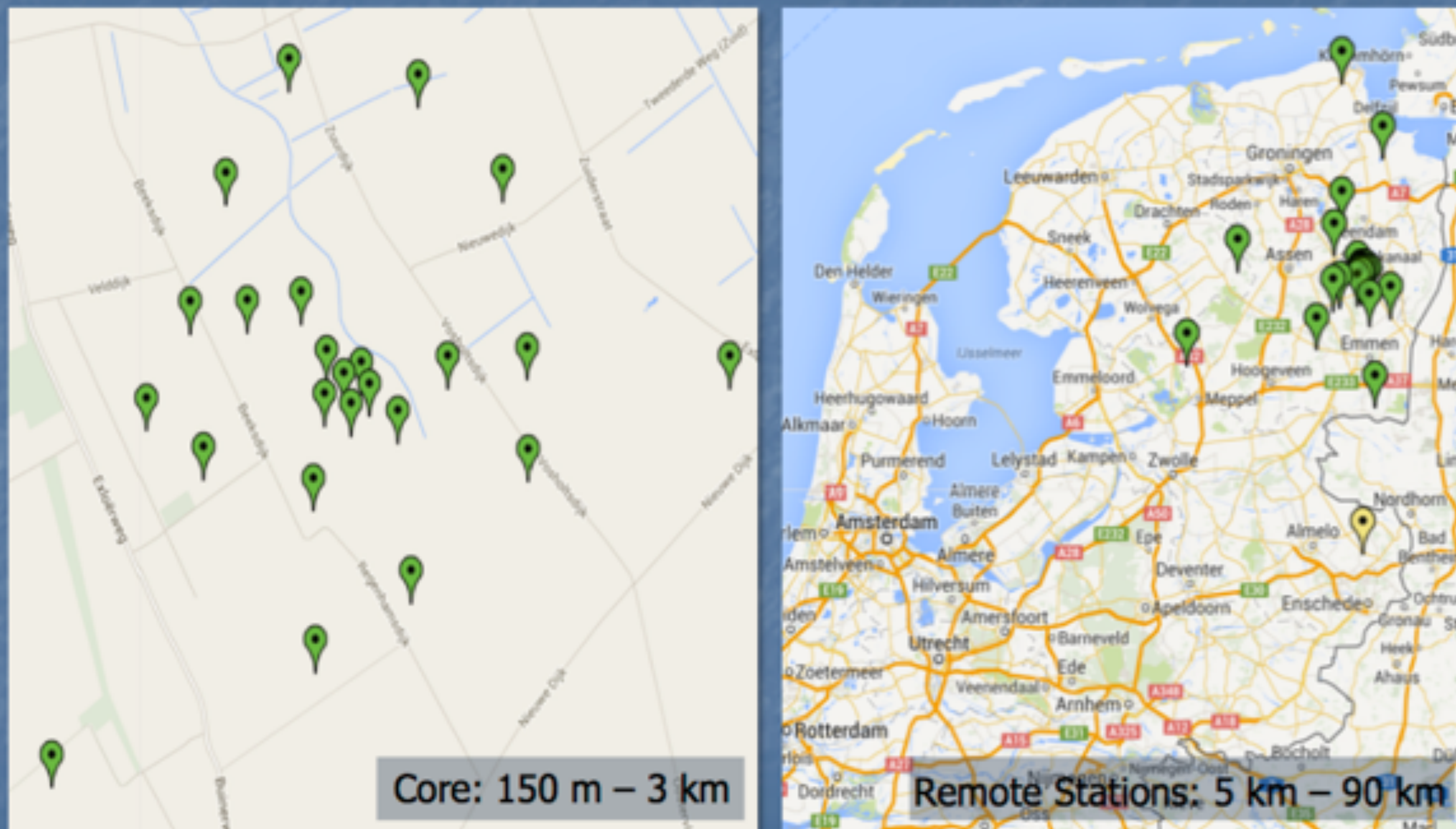




LOFAR telescope - station layout

LOFAR telescope - core stations

LOFAR - The Dutch array



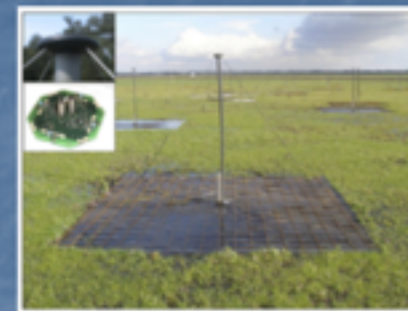
http://www.astron.nl/~heald/lofarStatusMap.html

International LOFAR Telescope

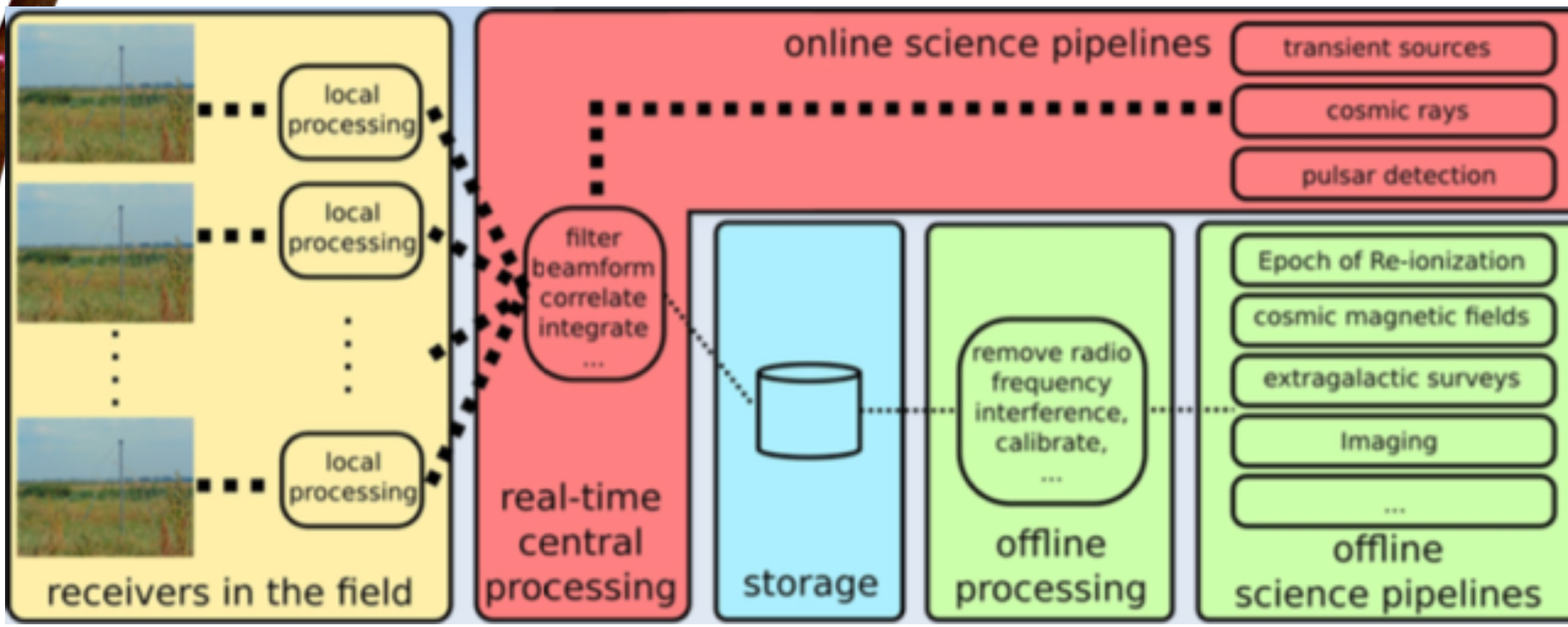
LOFAR station in Nancay- France

LOFAR telescope key facts

- The International LOFAR telescope (ILT) consists of an interferometric array of dipole antenna stations distributed throughout the Netherlands, Germany, France, UK, Sweden (+ Poland, …)
- Operating frequency is 10-250 MHz
- 1 beam with up to 96 MHz total bandwidth, split into 488 sub bands with 64 frequency channels (8-bit mode)
- < 488 beams on the sky with ~ 0,2 MHz bandwidth
- Low band antenna (LBA; Area ~ 75200 m^2; 10-90 MHz)
- High Band Antenna (HBA; Area ~ 57000 m^2; 110-240 MHz)

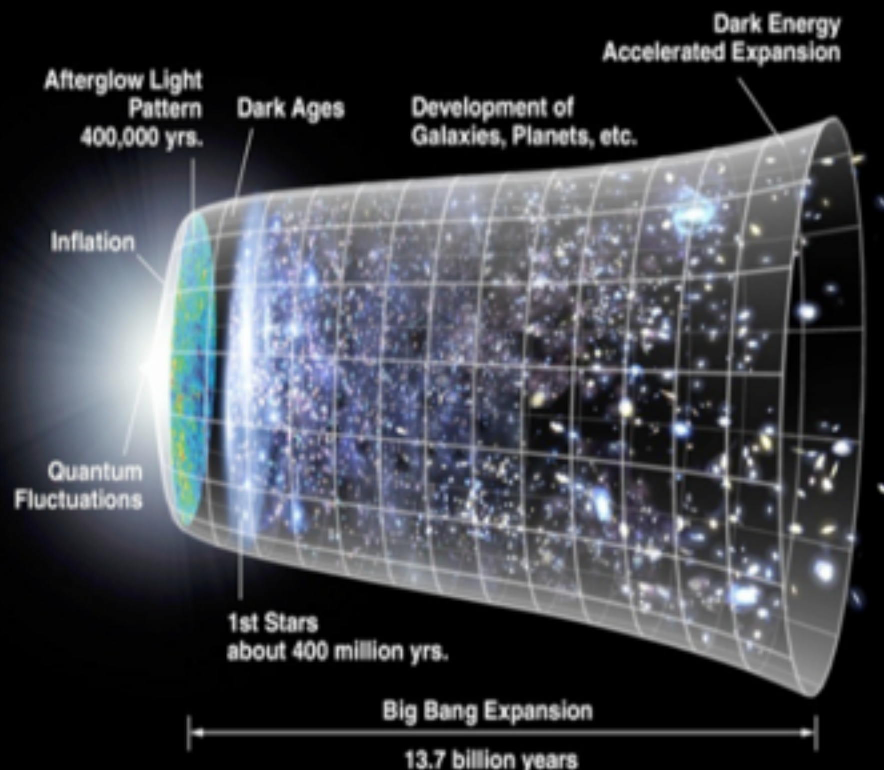


LOFAR data processing



Rob van Nieuwpoort, ASTRON

The LOFAR Epoch of Reionization Key Science Project



- When was the Universe reionized ?
- How (fast) did reionization proceed ?
- Which objects were responsible ?
  stars/galaxies , QSOs, or …

<u>Redshifted HI to frequency mapping</u>

$z = 6.7 \Rightarrow$ 185 MHz

$z = 8.5 \Rightarrow$ 150 MHz

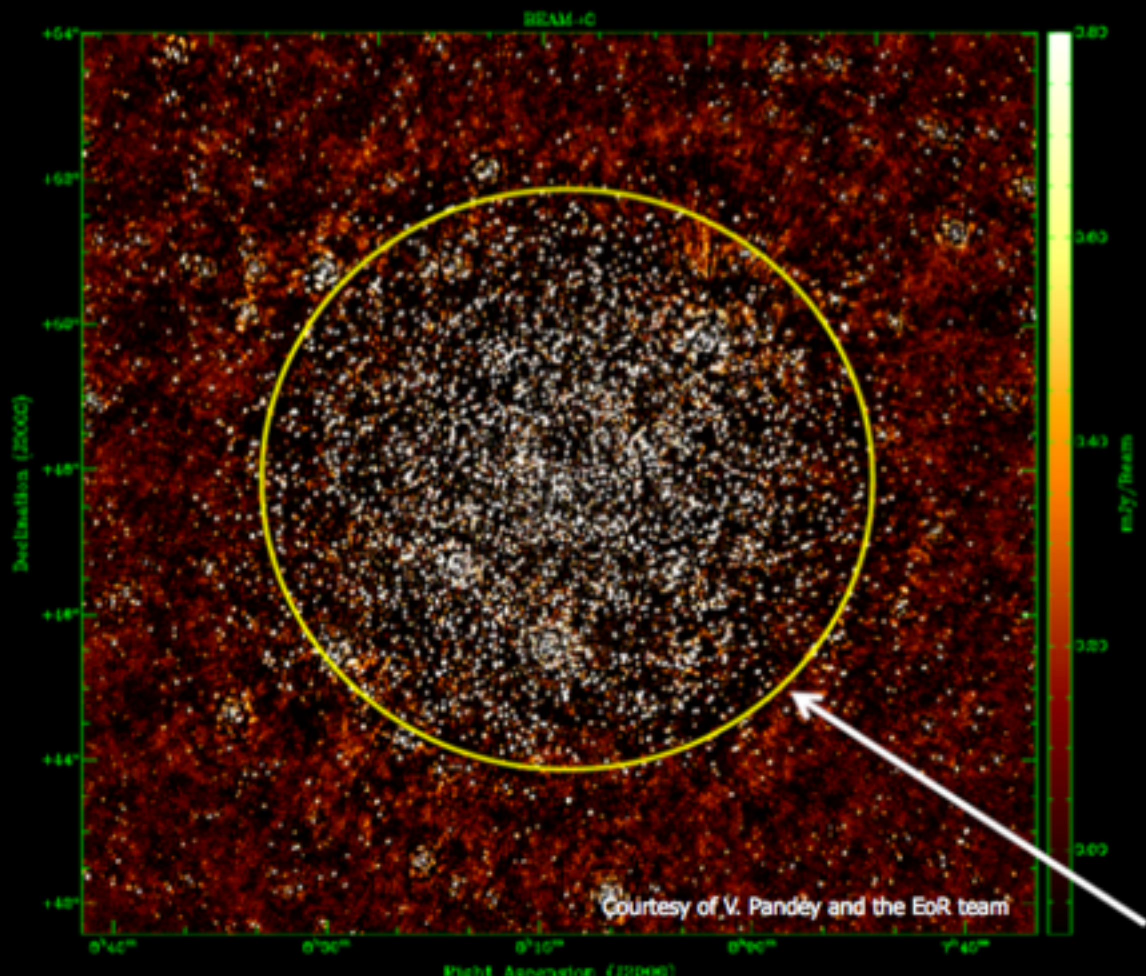
$z = 11.4 \Rightarrow$ 115 MHz

Goal: Detect cosmological 21cm signal (z~6-10) from the Epoch of Reionization

$\Rightarrow$ 1.5 Pbytes and 10^{21}-10^{22} FLOP to extract signal!

EoR deep field imaging

EoR deep field imaging



0.8 mJy

- 145 MHz (~2m)
- 60MHz continuum
- 6 powers of 10

- 32 hours on 3C196 (8 hrs x 4 days)
- Dec 21,12-Feb08,13
- 30λ - 5000λ
- Resolution - 50"
- $12^0 \times 12^0$ Image

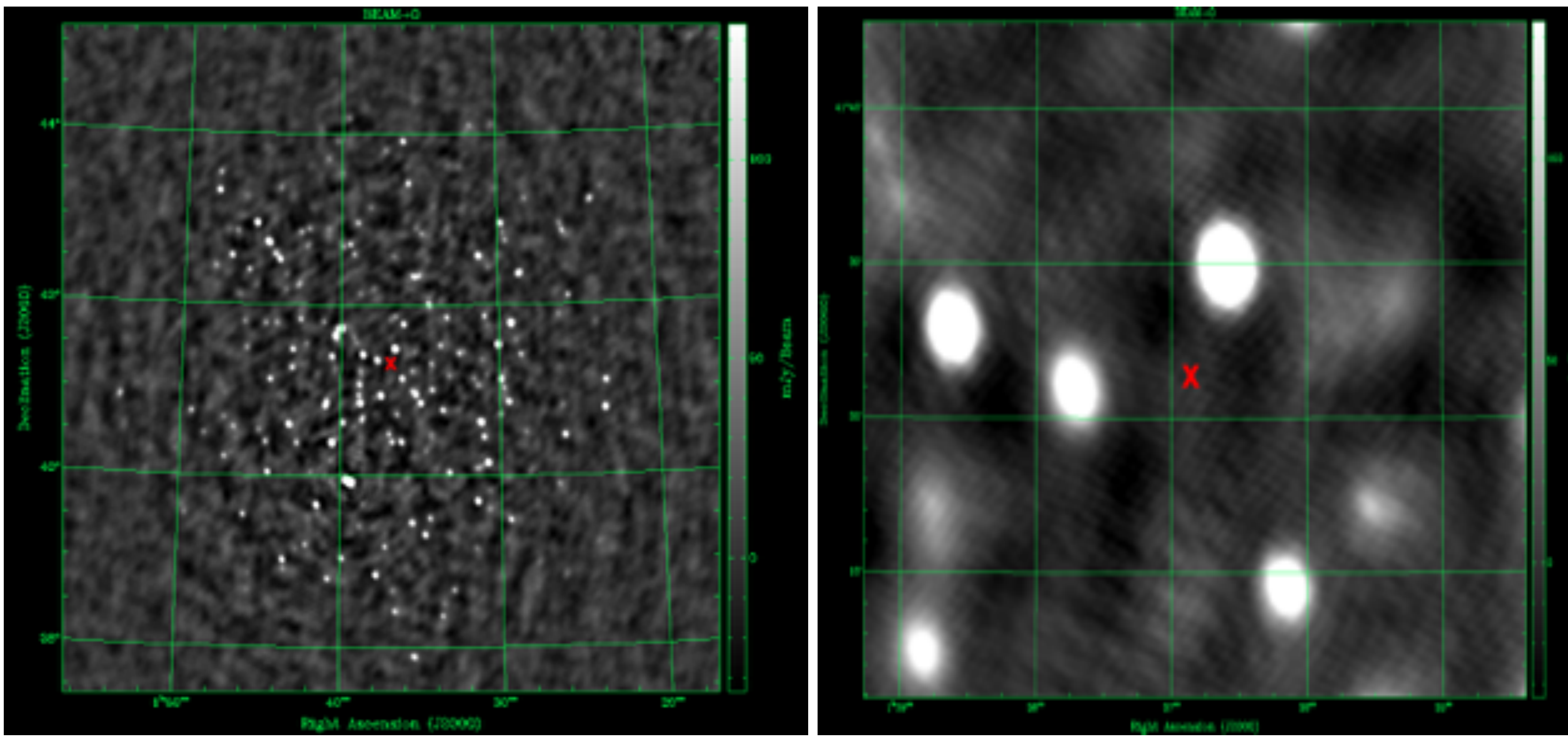
- 'Noise' < 75 μJy
- 3C196 - 79.97 Jy
- **DR: ~ 10^6:1**

Station beam (~8^0)

LOFAR data processing

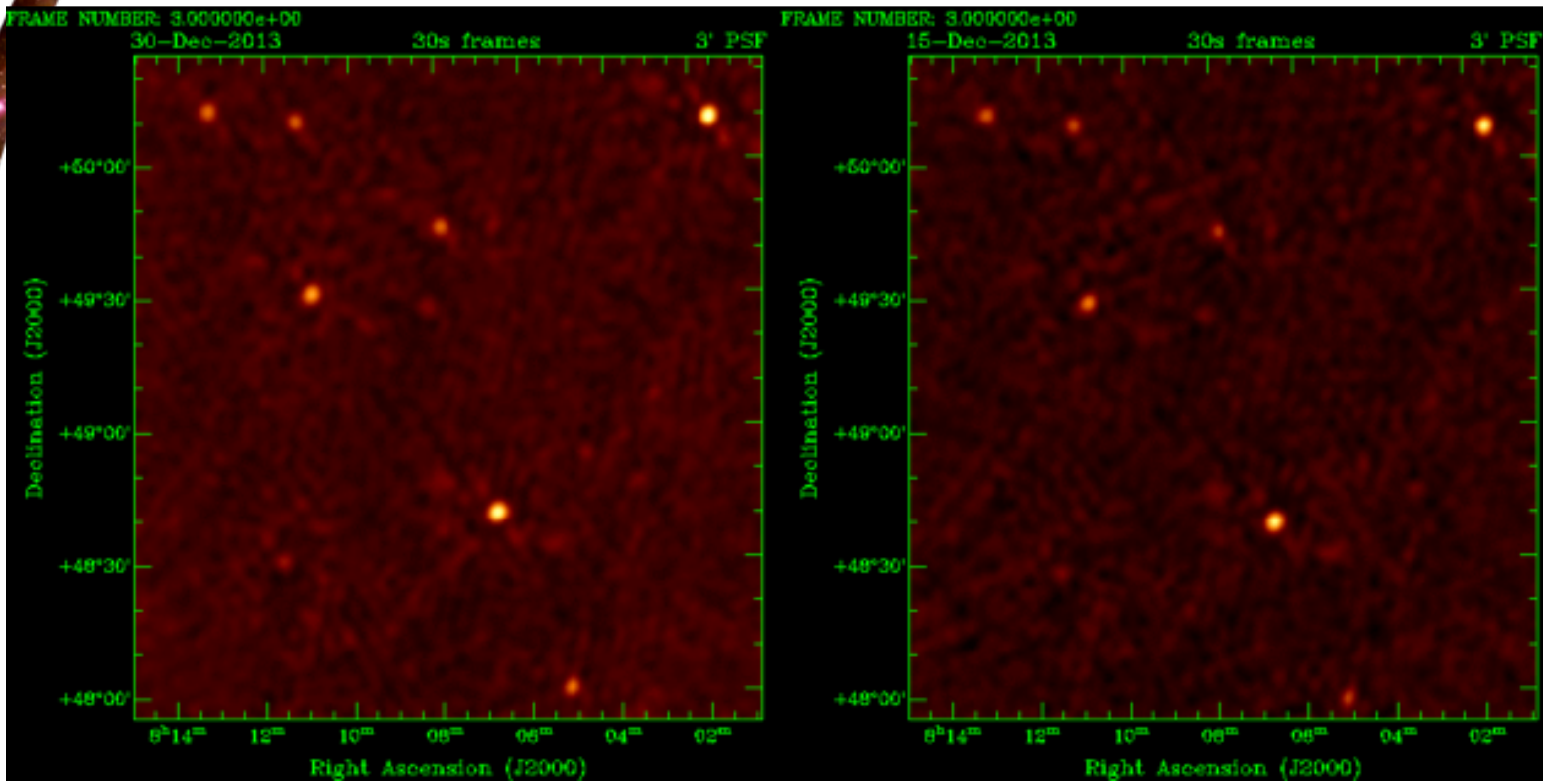
Search for Exoplanets Cyclotron Maser Radio Emission

- LOFAR observations of Upsilon Andromedae at 59 ±3 MHz, 3 hour integration
- Clean image, baselines ≤ 5 kλ, resolution ~ 40", noise ~ 8 mJy, no detection
- Wide-field image obtained by 3 self-cal cycles + zoom
- Upsilon Andromedae (RA=01:36:47.84, δ=+41°24’19.64") is marked by the red "x"



S. Daiboo, P.Zarka

Ionospheric ‘seeing’



Ger de Bruyn, ASTRON

Distributed computing challenges with LOFAR

Station signals collected in the station cabinets

Signal sent to COBALT for correlation

CEP2

CEP3

Data sent to CEP2 for initial RO processing – products might get copied to CEP3

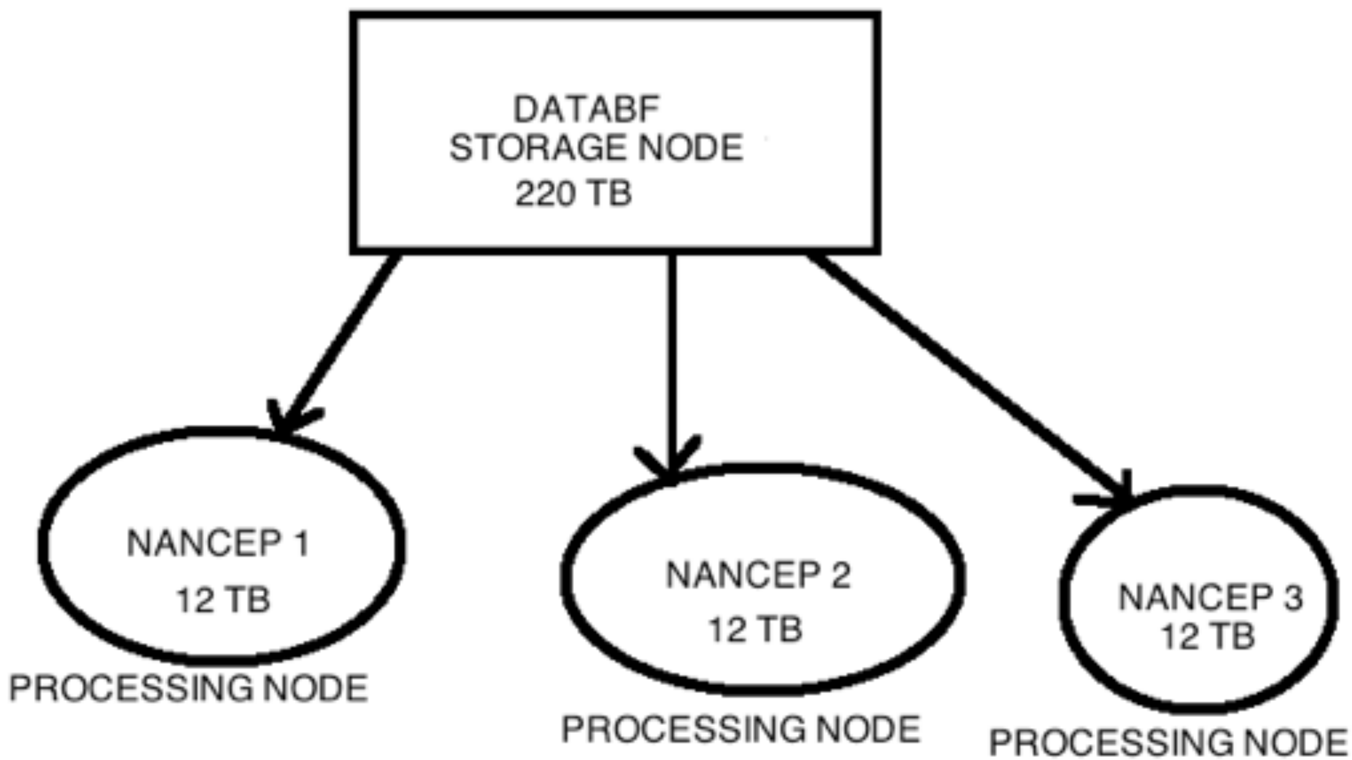
Products sent to the long-term archive

- Large data transport rates → data storage challenges (35 TB /h)
- LOFAR is the first of a number of new astronomical facilities dealing with the transport, processing and storage of these large amounts of data and therefore represents an important technological pathfinder for the SKA

R. Pizzo, ASTRON

LOFAR data processing

Processing setup at Nancay



LOFAR data processing

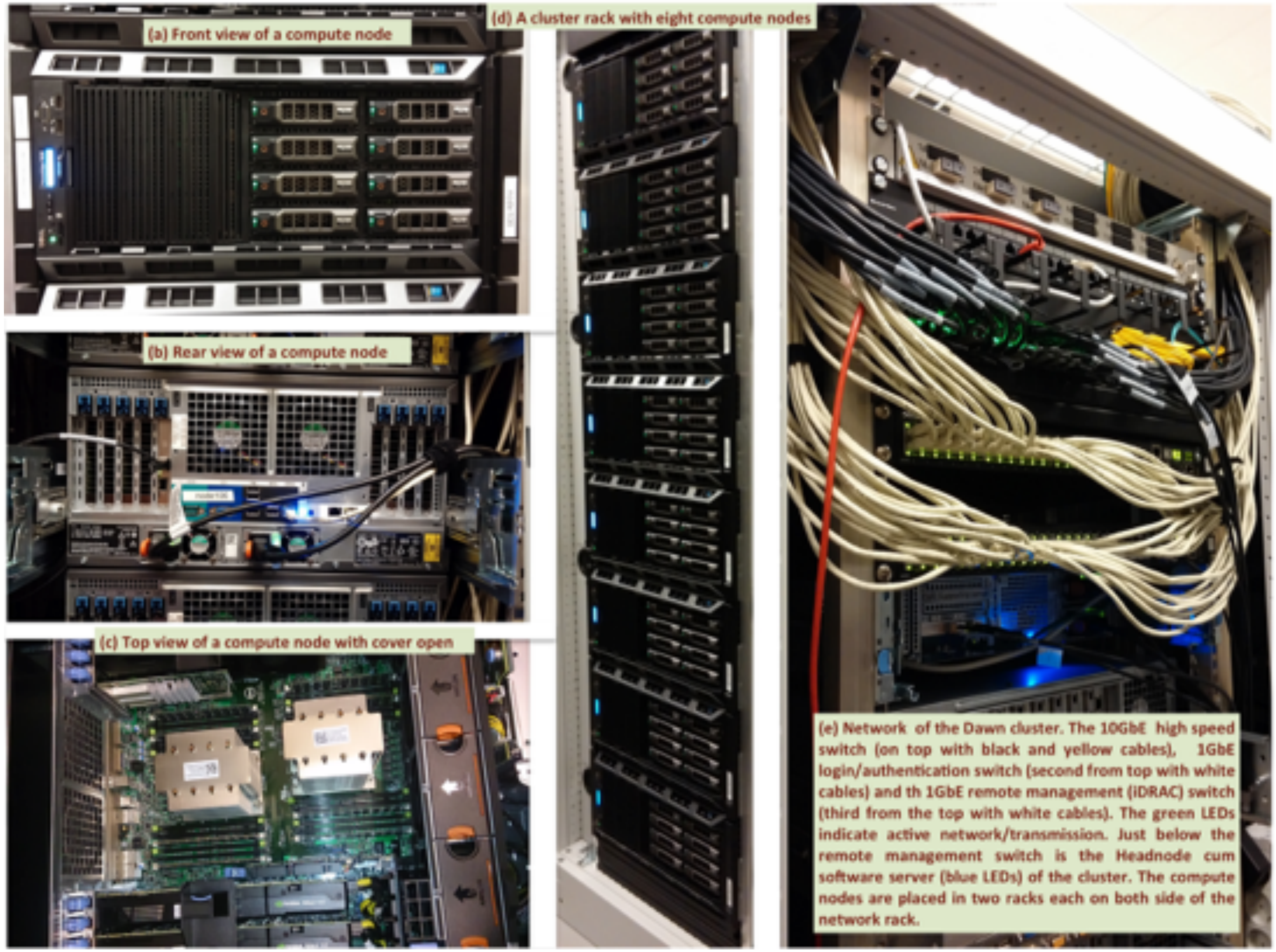
Typical data processing:

- RFI mitigation
- Instrument calibration
- Correction for direction dependent effects like ionosphere and beam.
- Imaging

LOFAR offline data processing challenges

- Large data sets, 1.2 TB per 5 hours observations.
- Takes 1 day to retrieve pre-processed data from observatory and 1 week to retrieve 12 TB of raw data.
- Need to have local copies of data on each processing node due to read/write speed over network is slow. Need faster network connections between nodes to overcome read/write lag.
- Data storage over long term not possible. Time-based processing and deletion of intermediate products.
- Correct for direction dependent effects like ionosphere and beam, computationally intensive.

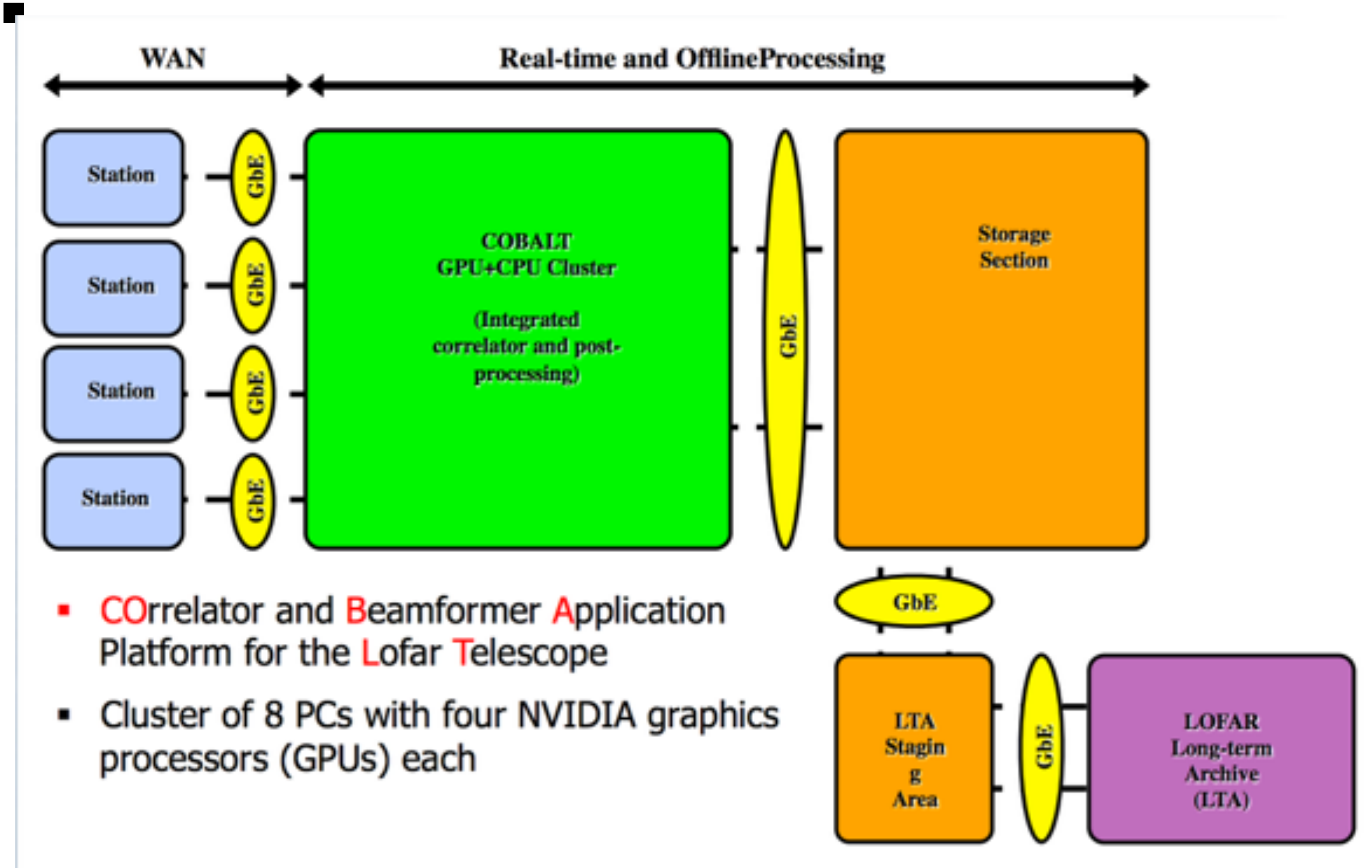
EoR KSP new processing cluster with fast inter-node network to reduce read/write lag.



Pandey, ASTRON

Processing cycle reduced from 4 months to a few weeks !

LOFAR online processing



- COrrelator and Beamformer Application Platform for the Lofar Telescope
- Cluster of 8 PCs with four NVIDIA graphics processors (GPUs) each

Data rate 35 TB/h

Chris Broekema, ASTRON

LOFAR online processing

Chris Broekema, ASTRON

- AARTFAAC, GPU cluster to search for transients
- DRAGNET, GPU cluster to search for fast transients.

J. Hessels, ASTRON

NenuFAR

SKA pathfinder instrument

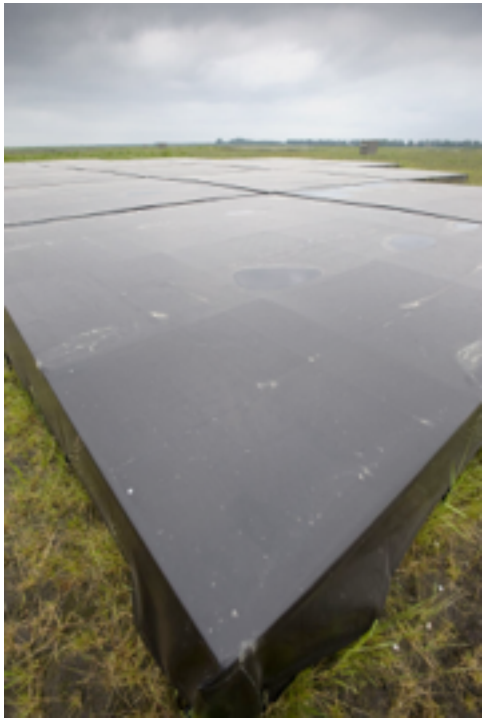
New Extension in Nançay Upgrading LOFAR

P. Zarka[1], M. Tagger[2],
L. Denis[C], S. Corbel[3,4],
J. Girard[1,4], C. Tasse[5],
USN personnel[3] ...
& the NenuFAR-France team[6]

[1]LESIA-OP, [2]LPC2E-Orléans, [3]USN-OP, [4]CEA-Saclay, [5]GEPI, [6]LERMA, ONERA, ENS/IAP, OCA, ...

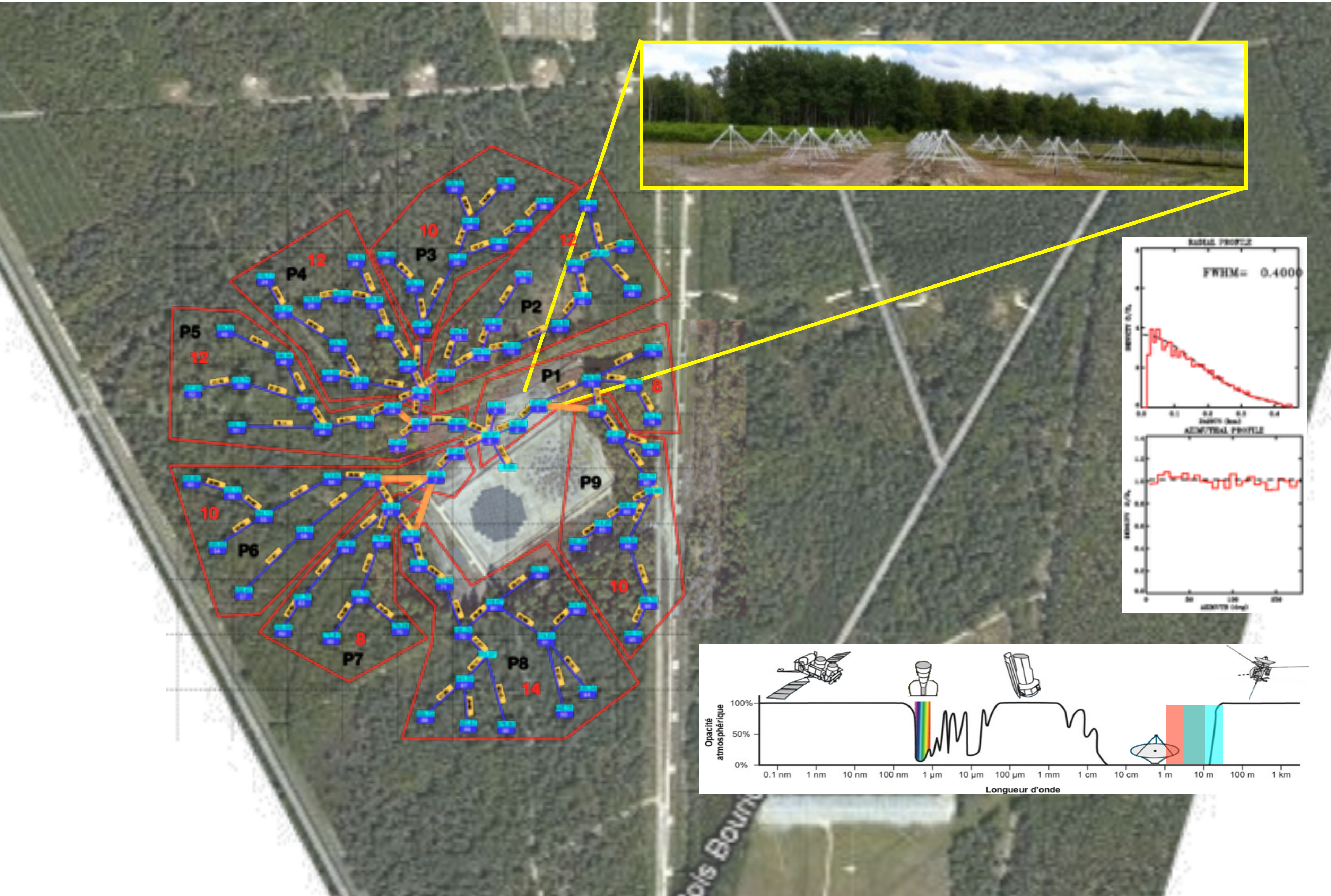
The NenuFAR concept

x

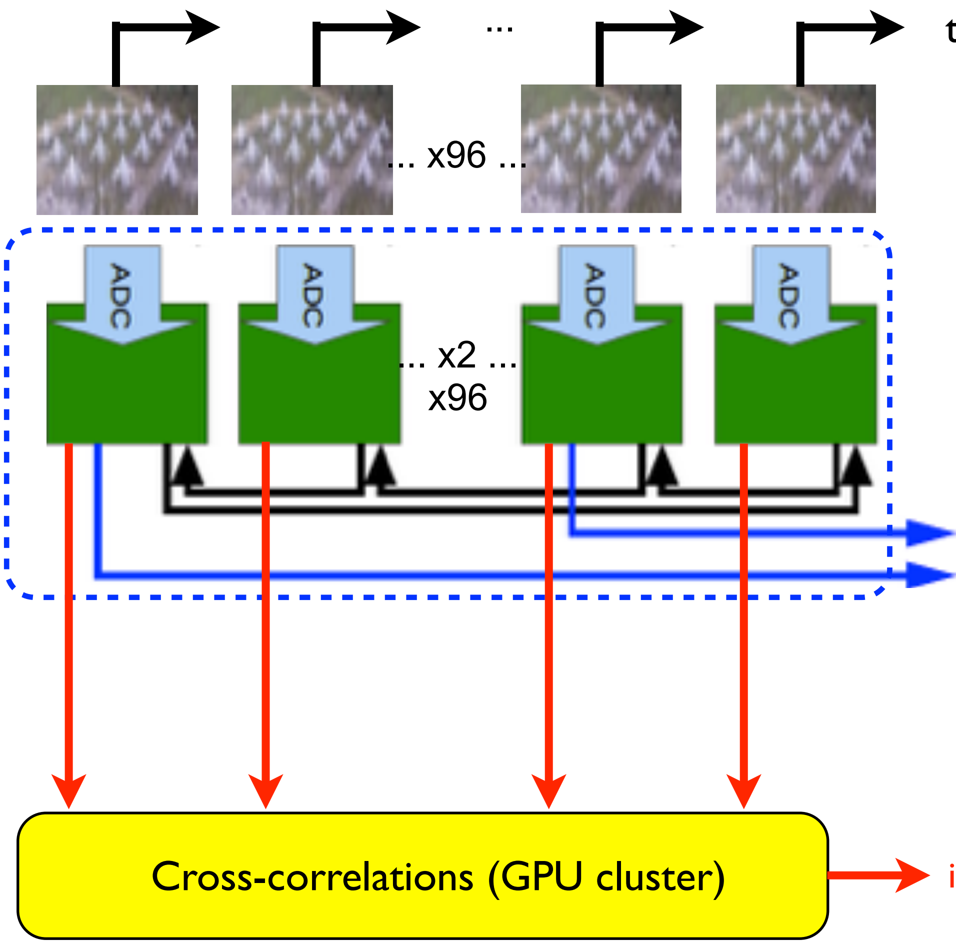


=

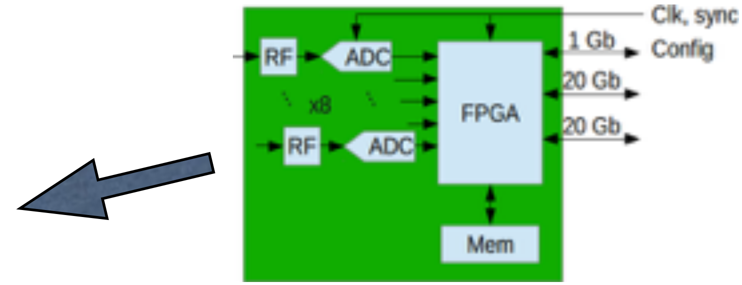
Designing NenuFAR



The main NenuFAR backend : LaNewBa



to LOFAR backend



2 beams full-band (~10-85 MHz), full-polar.
= 768 beamlets x 200 kHz (150 MHz)
multibeam, coherent/incoherent tile summation

channelizer, dedisperser (LPC2E)

imaging, ~every sec, tbd bandwidth, polar.

- NenuFAR = 2 instruments in 1

NenuFAR/LSS
NenuFAR/Standalone

⇒ fully parallel use to LSS mode
⇒ "duty-cycle" ~100% in the analog mini-array beam

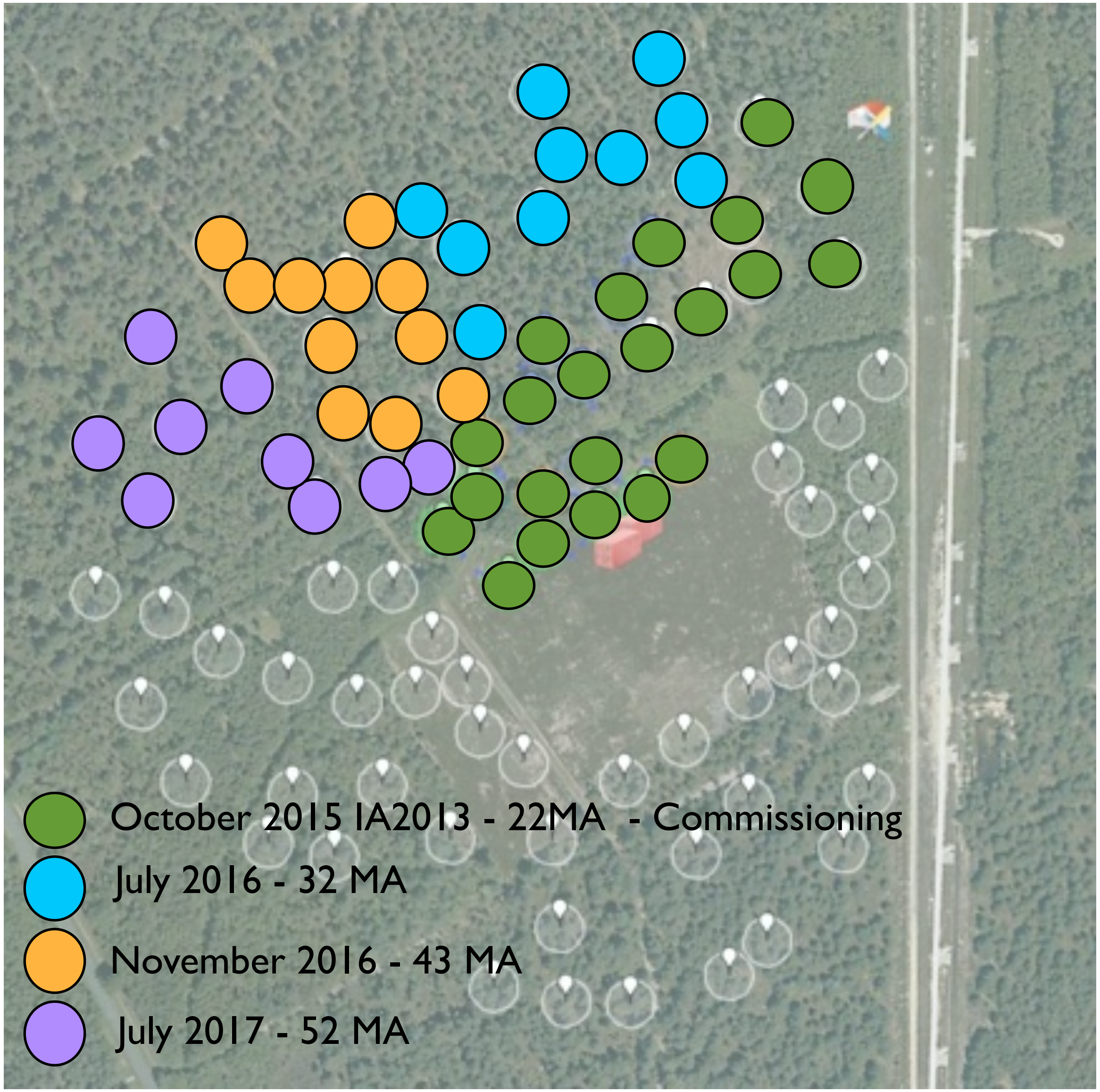
Technical characteristics of NenuFAR

- Giant LOFAR-compatible phased array & interferometer
- 1824 antennas : 96 mini-arrays of 19 antennas each (25 m ∅)
- Diameter ~400 m + 2-6 distant mini-arrays (up to 2-3 km)
- 4560 baselines
- Frequency range = 10-85 MHz (λ=3.5-30m)
- Resolutions: δf = 100 kHz (standalone) → 1 kHz, δt = 5 μsec, TBB @ 5 nsec
- Full polarization (4 Stokes)
- Collective area ~600λ^2 ≤ 62 000 m^2
- FoV = 32° - 8° @ 20-80 MHz ; pointing -23° → +90°
- Angular Resolution 2°-0.5° (Standalone instantaneous),

  5'-40' (Standalone synthesis), 0.1 " (LSS)
- Sensitivity : 2 - 0.5 Jy @ 20-80 MHz (5σ, 1 sec x 10 MHz, polarized signal)

  ~ 31 – 9 mJy " (5σ, 1 h x 10 MHz, polarized signal)

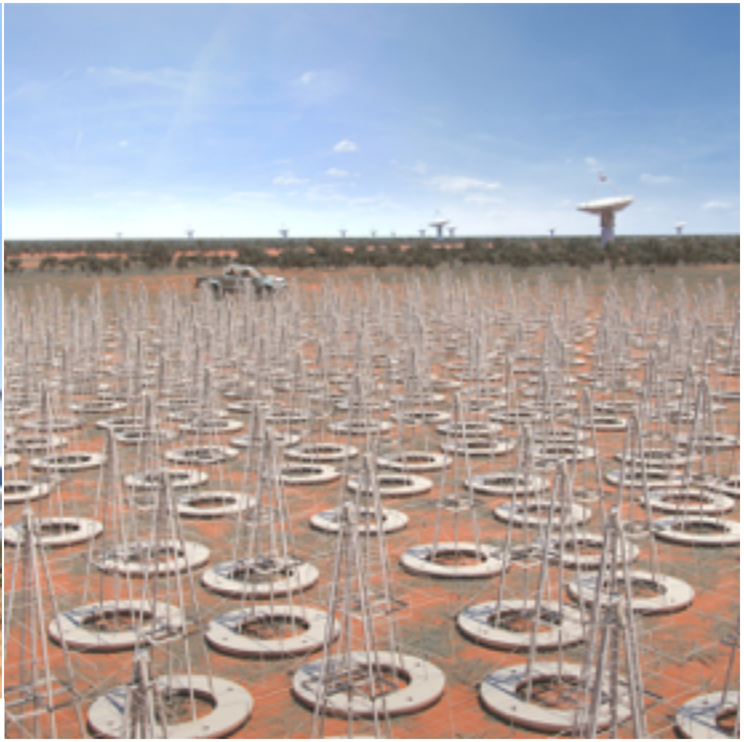
Construction



NenuFAR today

Square Kilometre Array

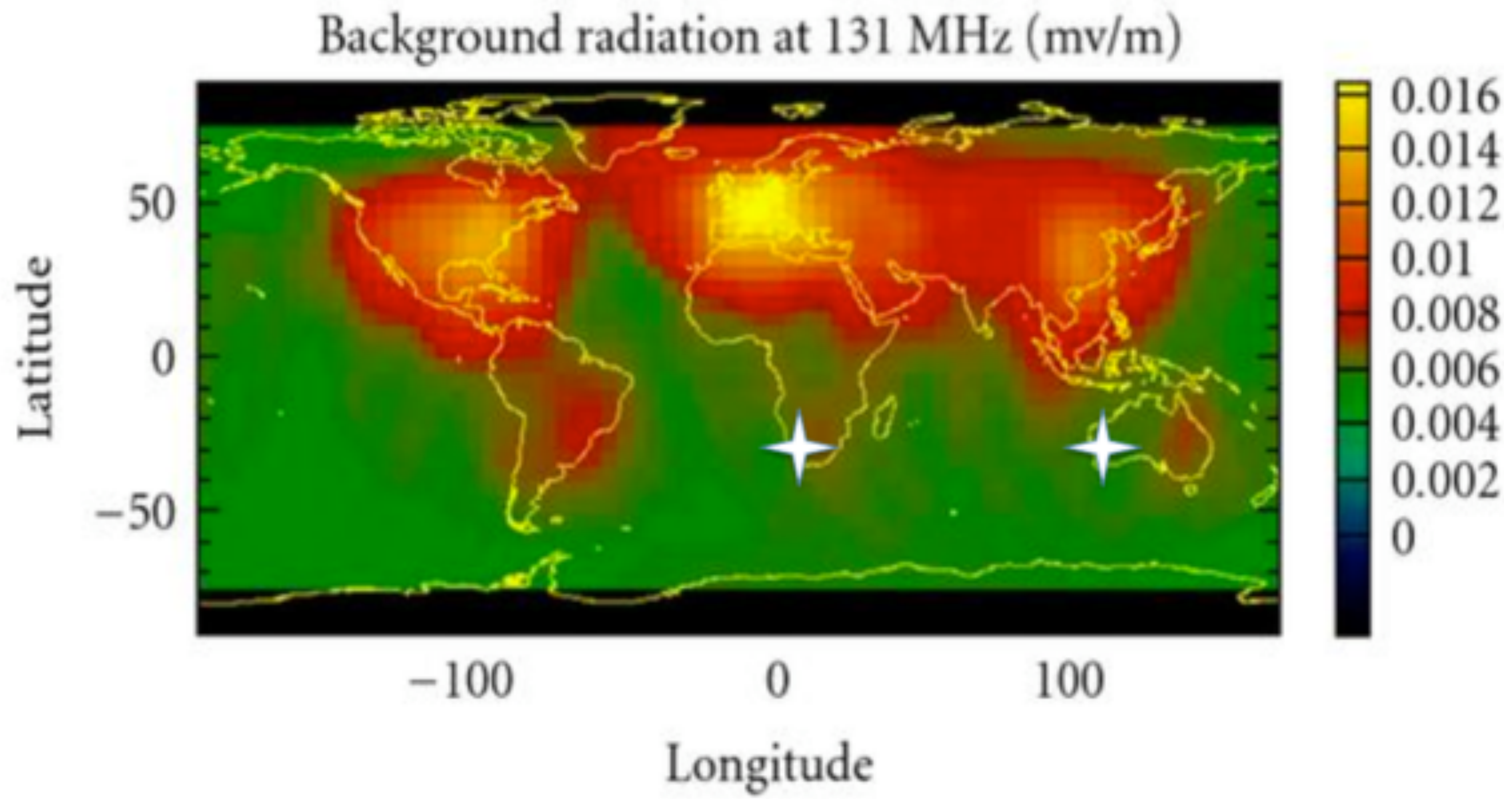
An international consortium



SKA organisation

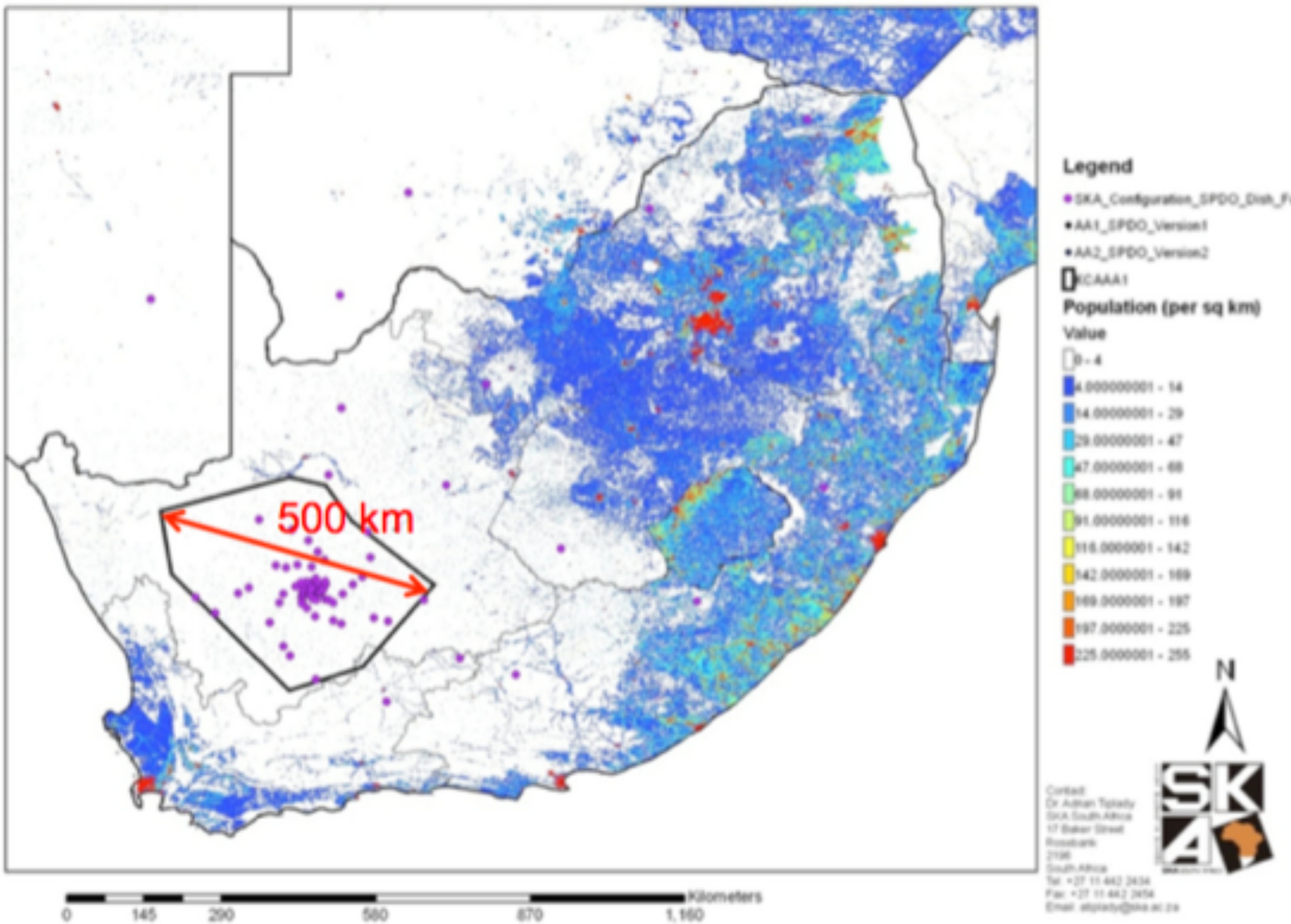
- Australia: Department of Industry and Science
- Canada: National Research Council
- China: Ministry of Science and Technology of the People's Republic of China
- India: National Centre for Radio Astrophysics
- Italy: National Institute for Astrophysics
- New Zealand: Ministry of Economic Development
- South Africa: National Research Foundation
- Sweden: Onsala Space Observatory
- The Netherlands: Netherlands Organisation for Scientific Research
- United Kingdom: Science and Technology Facilities Council

Square Kilometre Array sites



T. Bourke, SKA organisation

Square Kilometre Array Karoo site in South Africa



Square Kilometre Array site in Australia

SKA organisation

Square Kilometre Array construction phases

The SKA will be built in 2 phases and will cover the frequency range 50 MHz - 13.8 GHz:

Construction cost : 650M Euros

Early science : 2020+

- Phase 1 - 2018-2023

SKA1- Low : 131,000 low-freq dipoles, Australia.

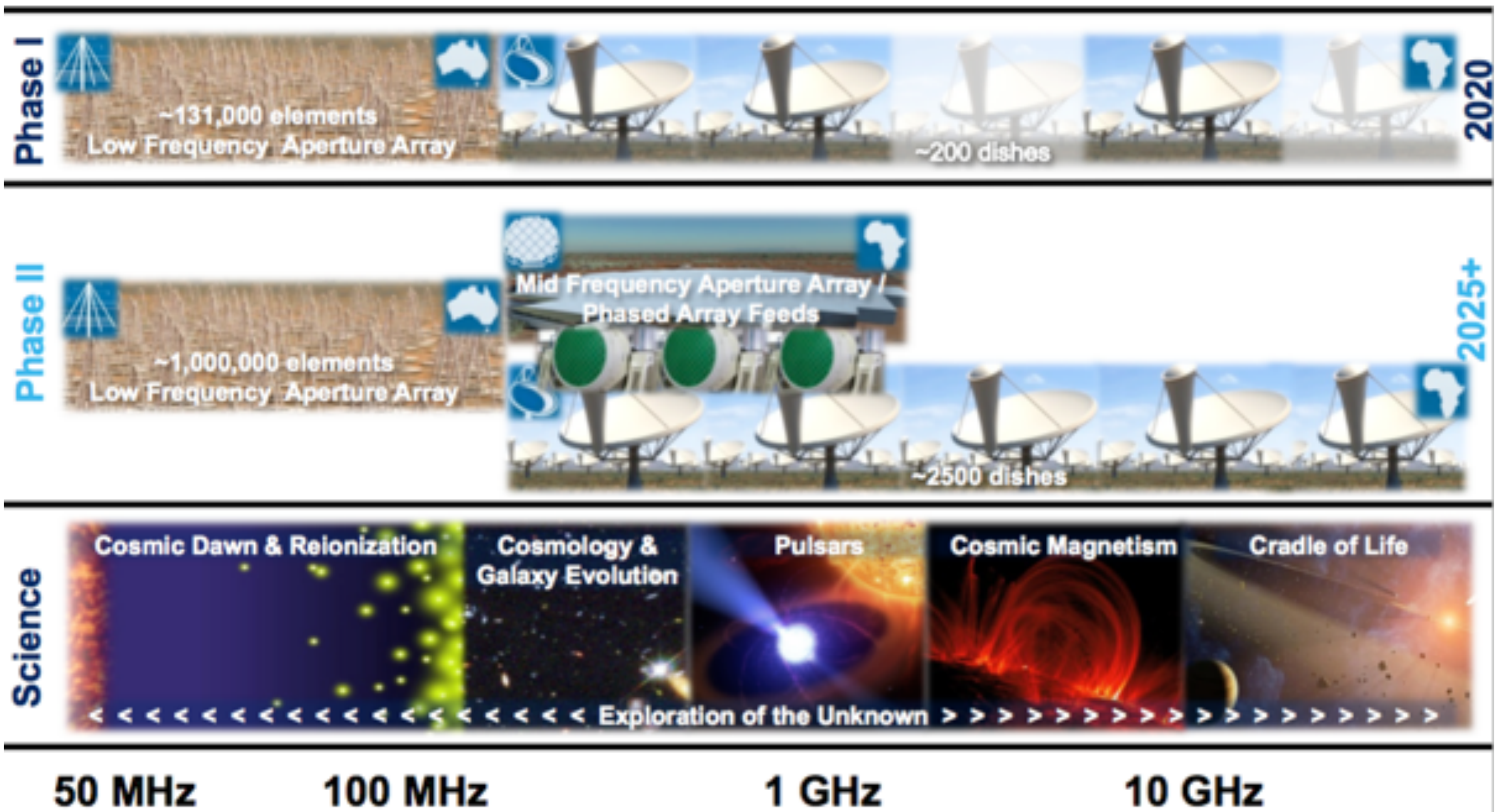
SKA1-Mid: 133x15m dishes + MeerKAT, South Africa

- Phase 2 - 2020+

1,000,000 elements low-frequency aperture array

2500 dishes.

SKA construction phases

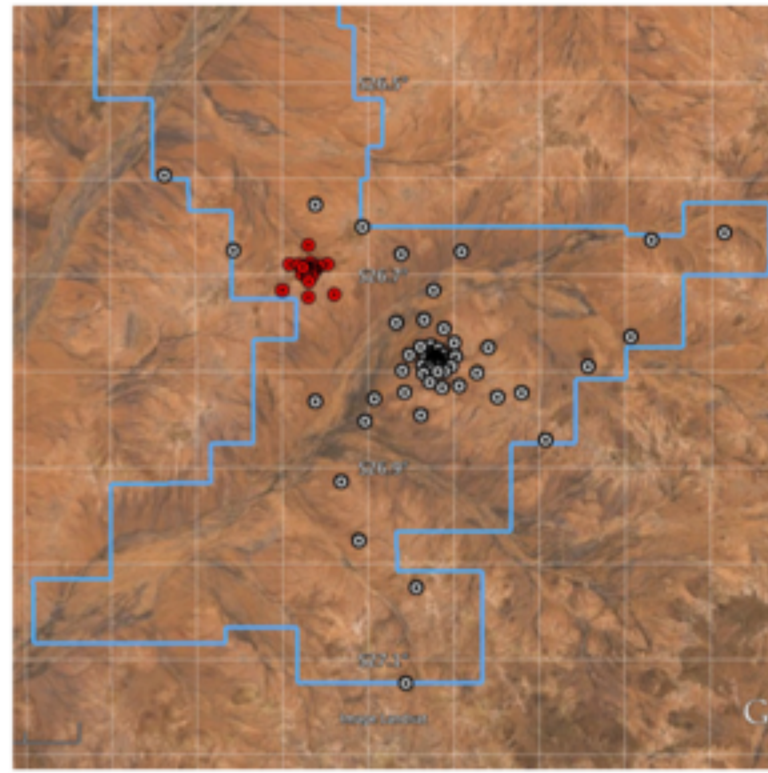


T.Bourke, SKA Organisation

SKA1 configuration

-

Red = MID; white = MeerKat
~150 km maximum baseline

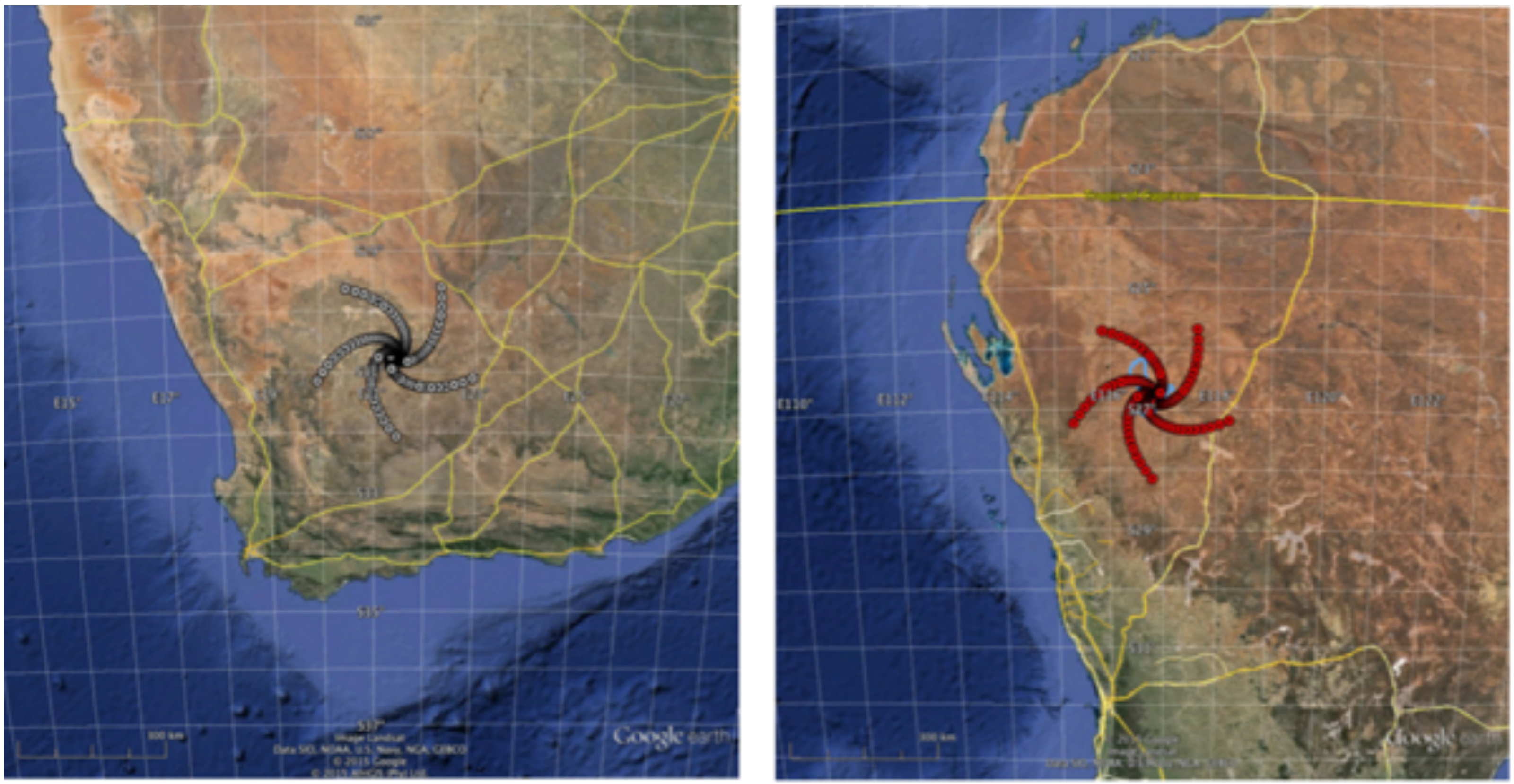


Red = ASKAP; White = LOW
~65 km maximum baseline

T. Bourke, SKA organisation

SKA2 configuration

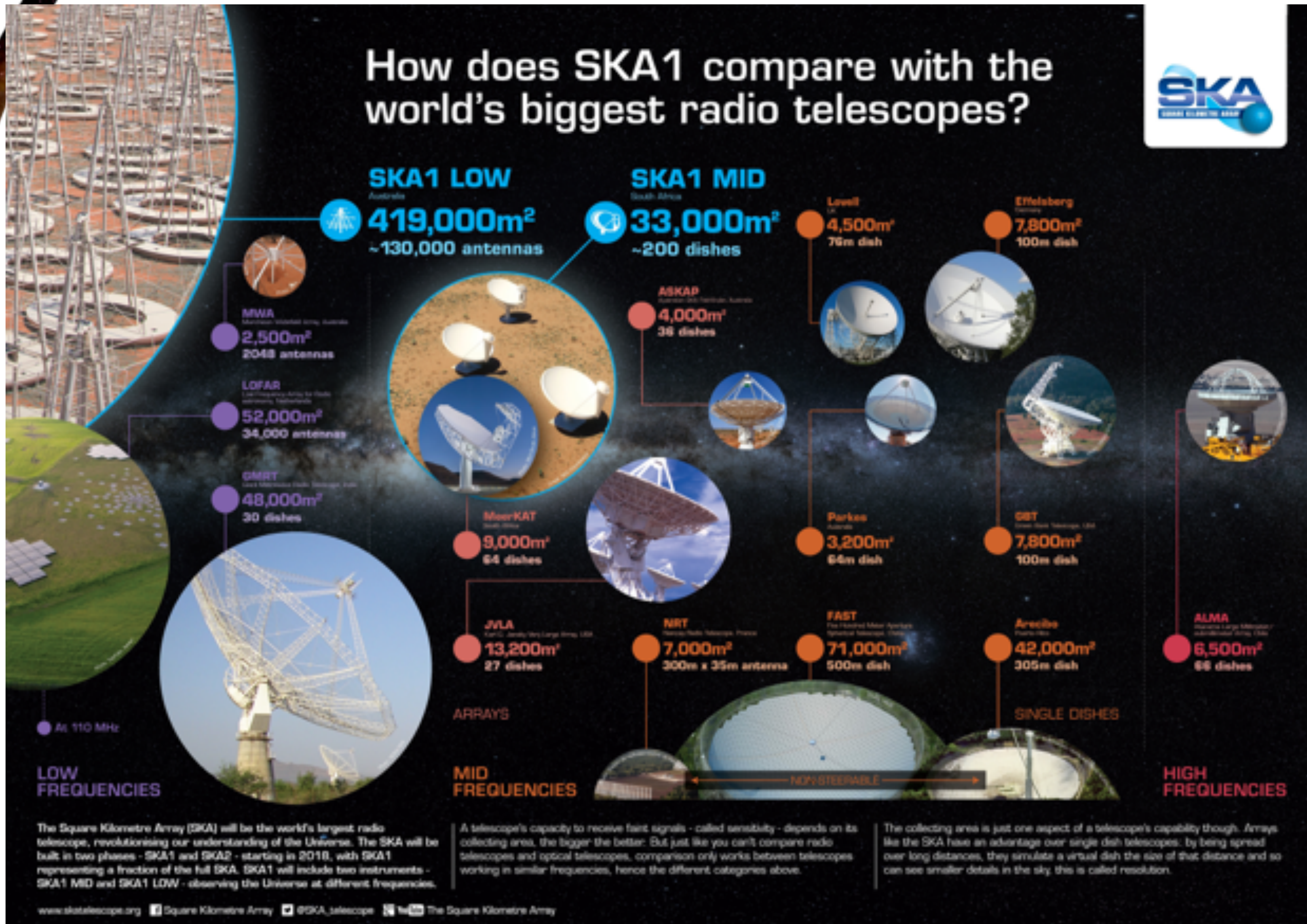
-



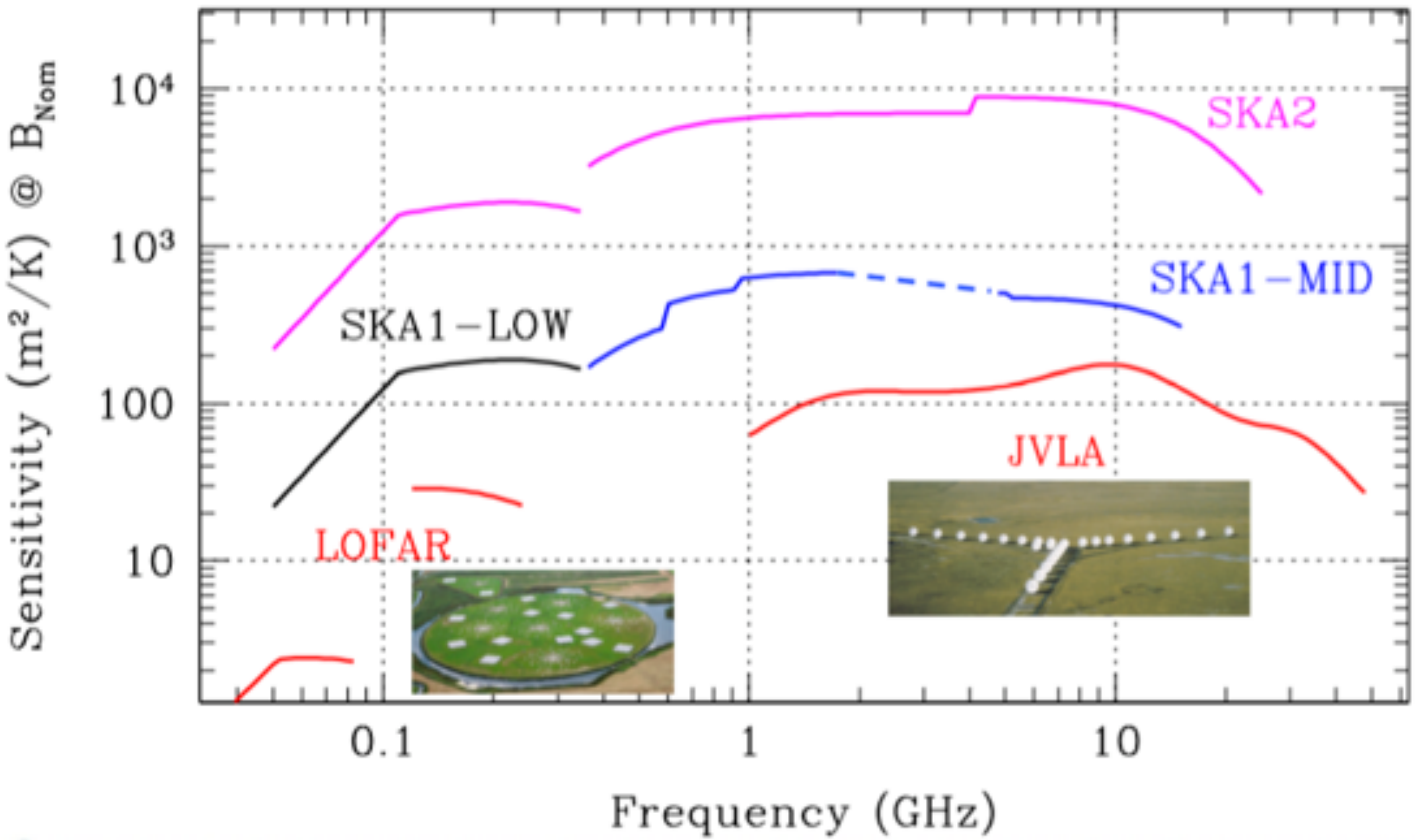
SKA2–Dish, –LOW: B_{max} ≈300 km "core", ≈3000 km remote

T. Bourke, SKA organisation

SKA1 compared to present telescopes



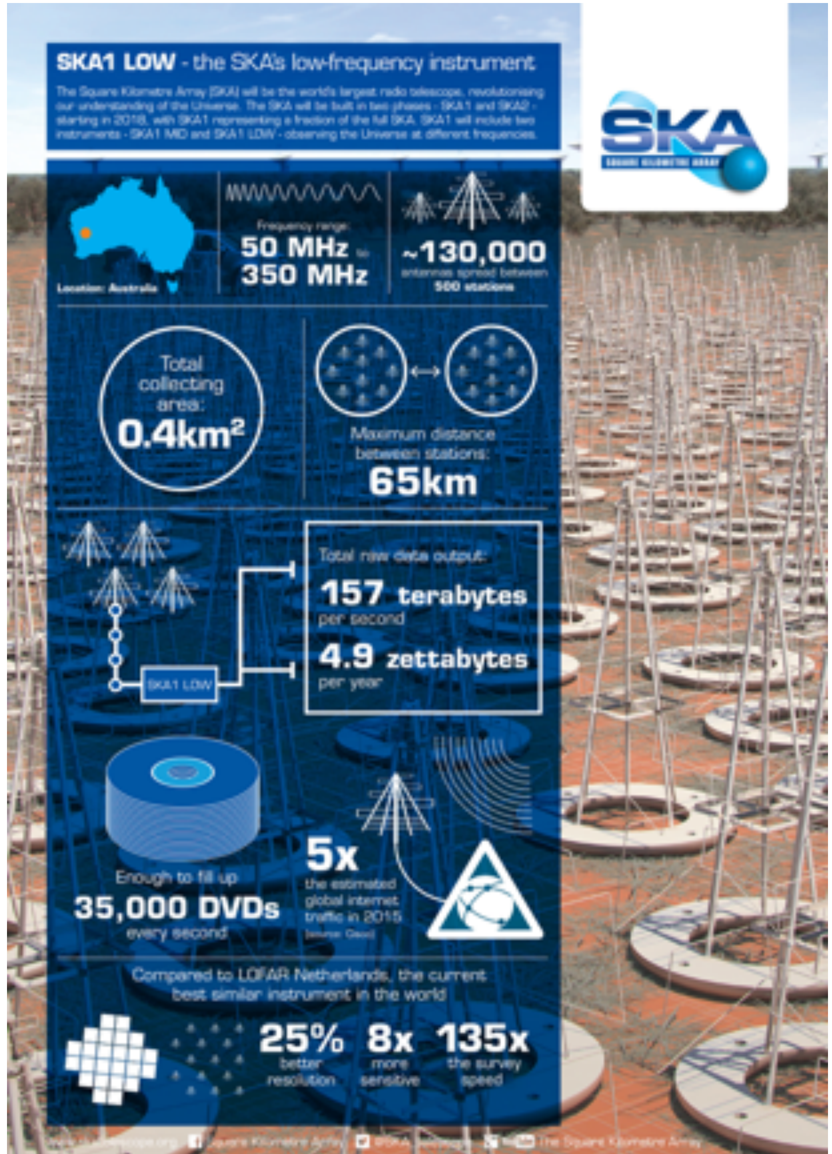
SKA sensitivity comparison



T. Bourke, SKA organisation

SKA data rates

-



SKA organisation

SKA data rates

- Raw Data Rates (Transport)
  - LOW ~150 Tb/s, ~5 Zb/yr
  - MID ~ 2 Tb/s, 62 Eb/yr

- Processing Power (HPC)
  - LOW ~21 PFlops
  - MID ~60 PFlops
- Power (300 PFlops) = Power to run San Fran.
- Remote, power limited, future = renewables

Archive(s)

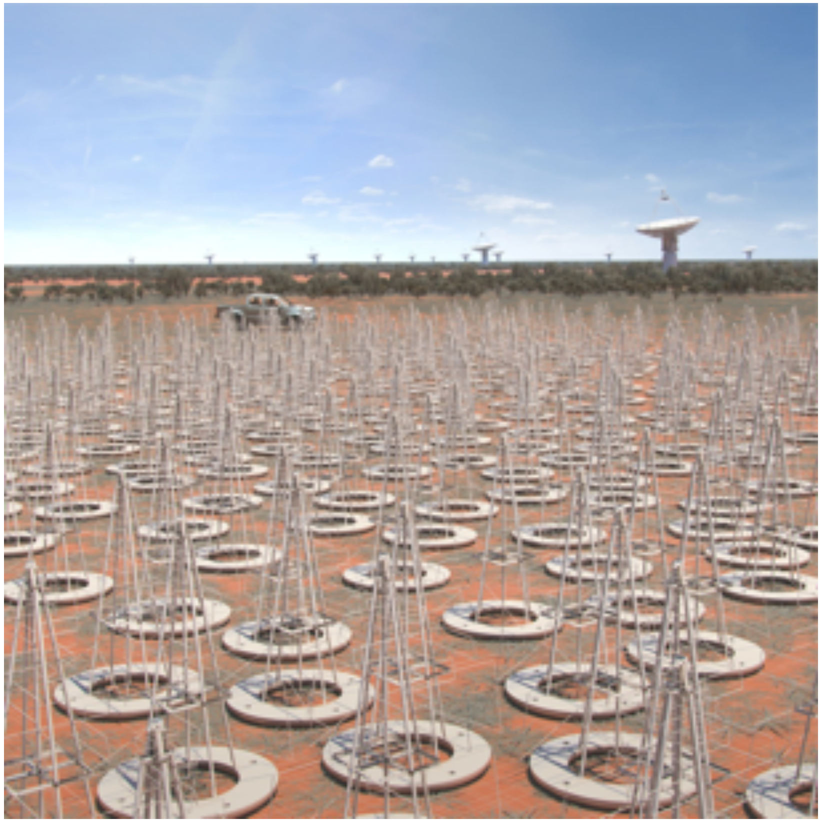
- Exa-byte capacity
- Where (Regional Data Centres?)
- What (… should be in it)
- Growth rates are 10s Pb/year
- SKA2 ~10^6 times worse!

AWS, IBM, Google, Nvidia
SGI, Intel, …

T. Bourke, SKA organisation

Square Kilometre Array data rates

- The SKA will produce 10 times the global internet traffic.
- It will use enough optical fibre to wrap twice around the earth.
- The central computer will have the processing power of one hundred million PCs.



SKA organisation

Square Kilometre Array data rates

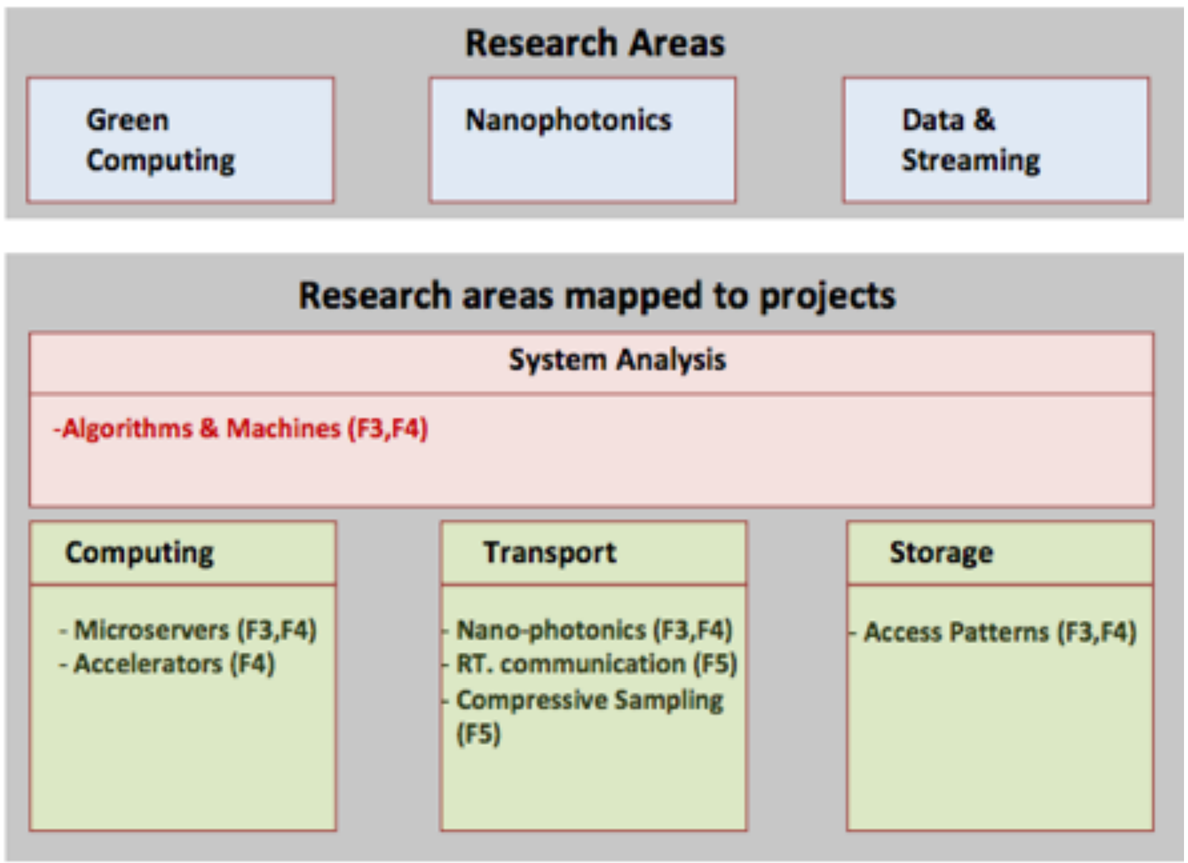
How to address the challenges of large data rates?

- Process data online, means automated pipelines for processing currently done offline.
- Smarter algorithms, e.g. CohJones direction dependent calibration (C. Tasse), Baselines averaging to reduce data size and taking advantage of sparsity in the measurement space.

Square Kilometre Array data rates

How to address the challenges of large data rates?

- Ongoing DOME project by IBM/ASTRON to address:



IBM/DOME project

Summary

- Large telescopes like LOFAR generate massive amount of data and require enormous amount of computing power.
- Becoming increasingly challenging to process and store such large quantity of data.
- Present upgrades to the LOFAR like NenuFar also face the same data storage and processing challenges.
- The SKA will scale up the number of sensors on the ground and proportionally scale up the data and processing rate.
- We need new algorithms, new data transport and processing to cope with this deluge of data.