Data transport in a radio telescope:

Remote Direct Memory Access over Ethernet from FPGA to GPU

Netherlands Institute for Radio Astronomy

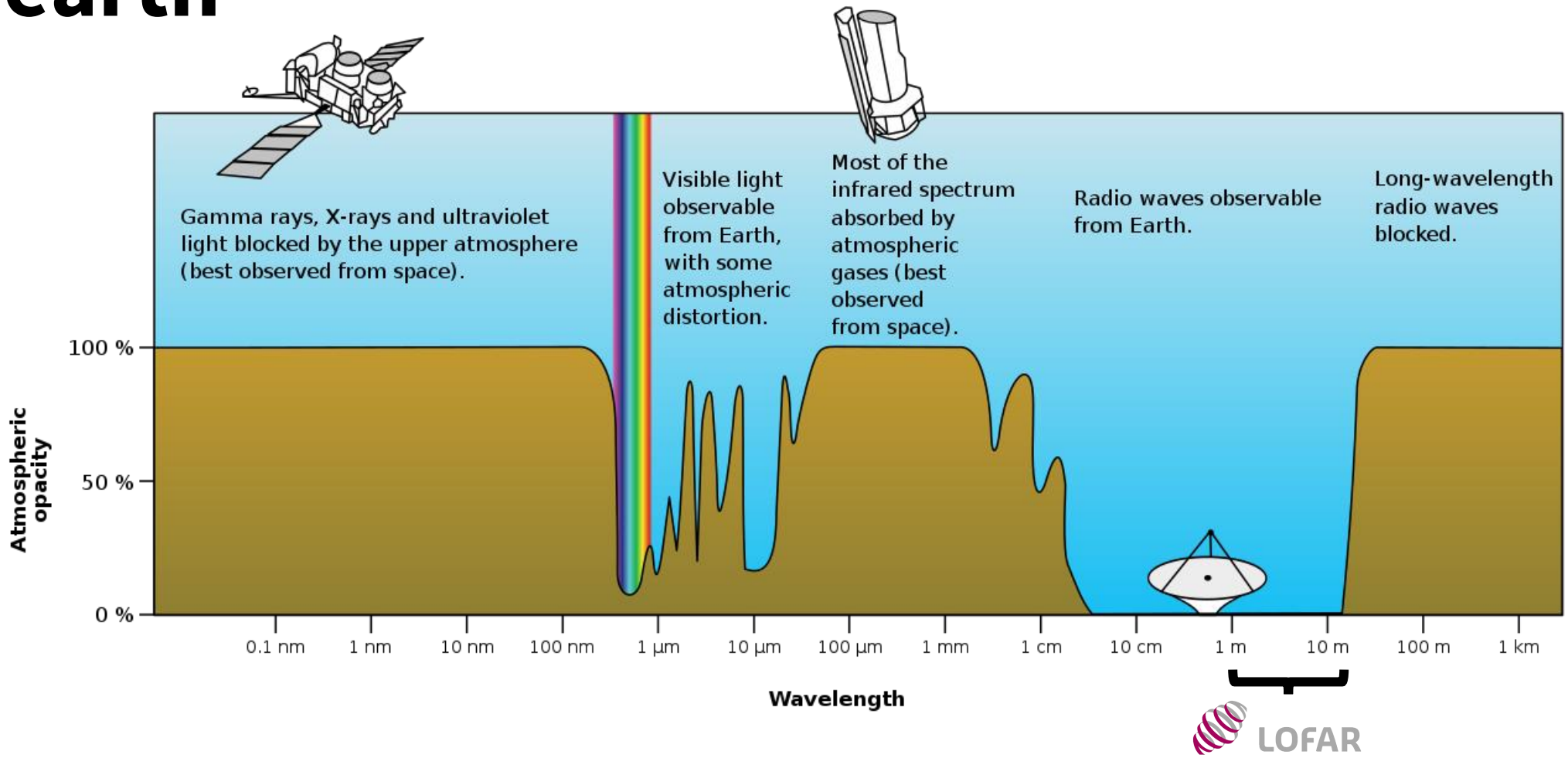
Steven van der Vlugt

vlugt@astron.nl

Link to slides and background materials at the last slide

ORCONF 2023

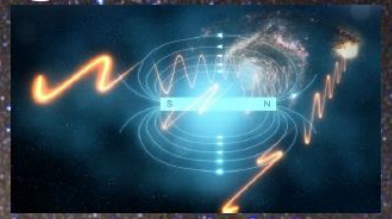
The observable radio spectrum from earth



By NASA (original); SVG by Mysid. - Vectorized by User:Mysid in Inkscape, original NASA image from File:Atmospheric electromagnetic transmittance or opacity.jpg., Public Domain, https://commons.wikimedia.org/w/index.php?curid=5577513

ASTRON
Netherlands Institute for Radio Astronomy

Cosmic magnetism



Supermassive black holes

Early Universe

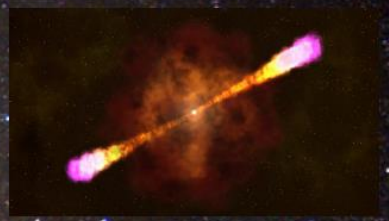
Supernovae

Galaxy clusters

Sun

Pulsars

Gravitational wave events



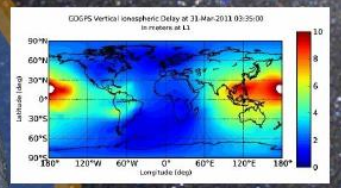
Solar System Planets

Meteors

Nearby galaxies

Cosmic rays

Ionosphere



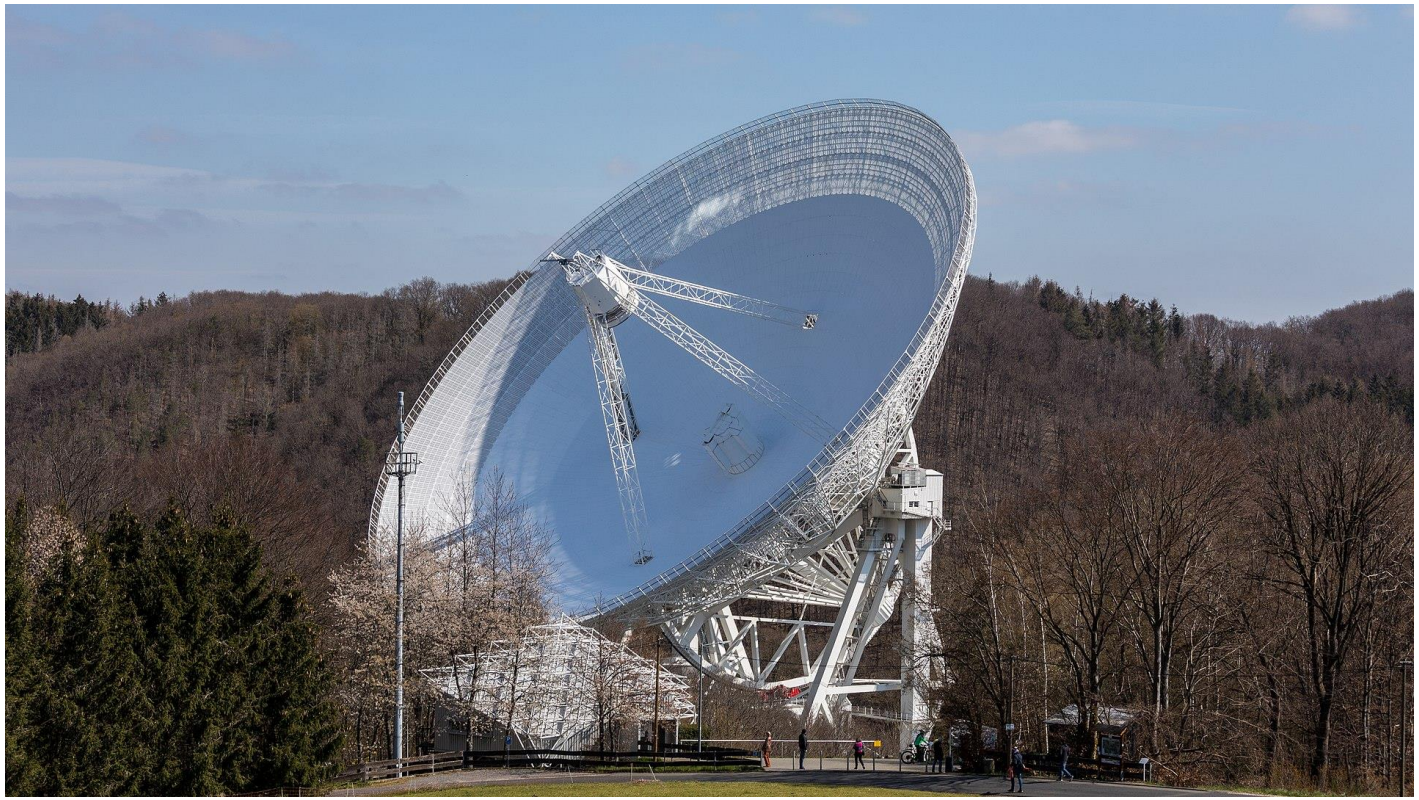
Interstellar medium

Lightning

Space weather

Radio telescopes

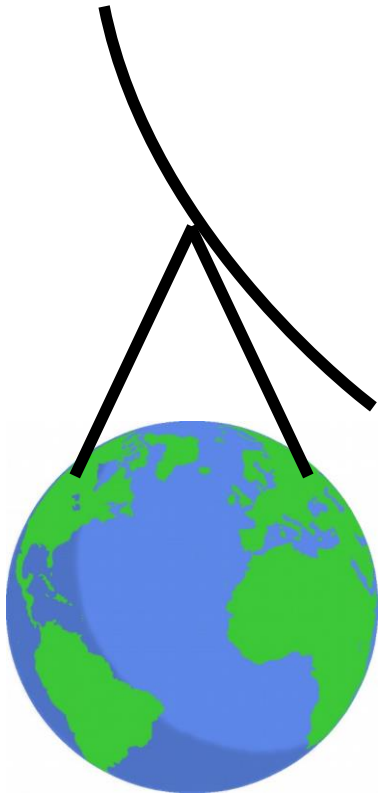
ASTRON Dwingeloo Radio Telescope
By Uberprutser - Own work, CC BY-SA 3.0 nl, https://commons.wikimedia.org/w/index.php?curid=32621318

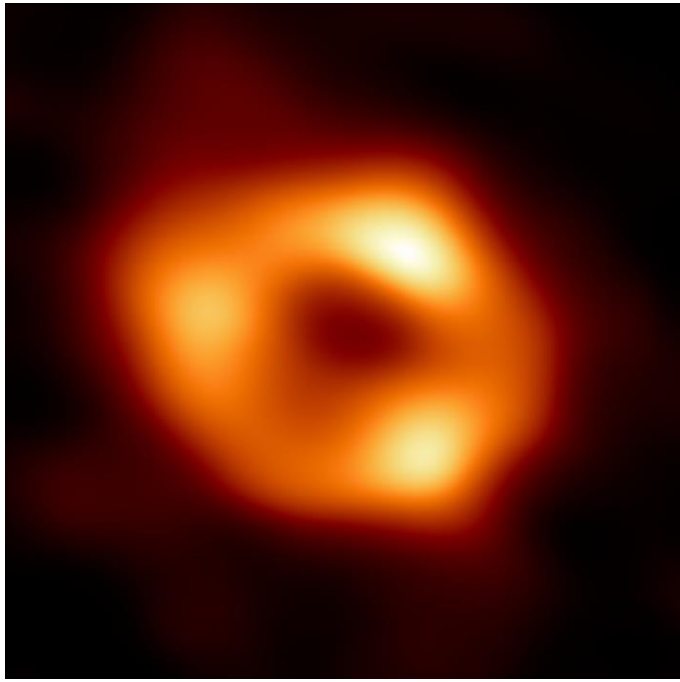


Effelsberg
By © Raimond Spekking / CC BY-SA 4.0 (via Wikimedia Commons), CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=104342305

ASTRON
Netherlands Institute for Radio Astronomy

Radio telescopes

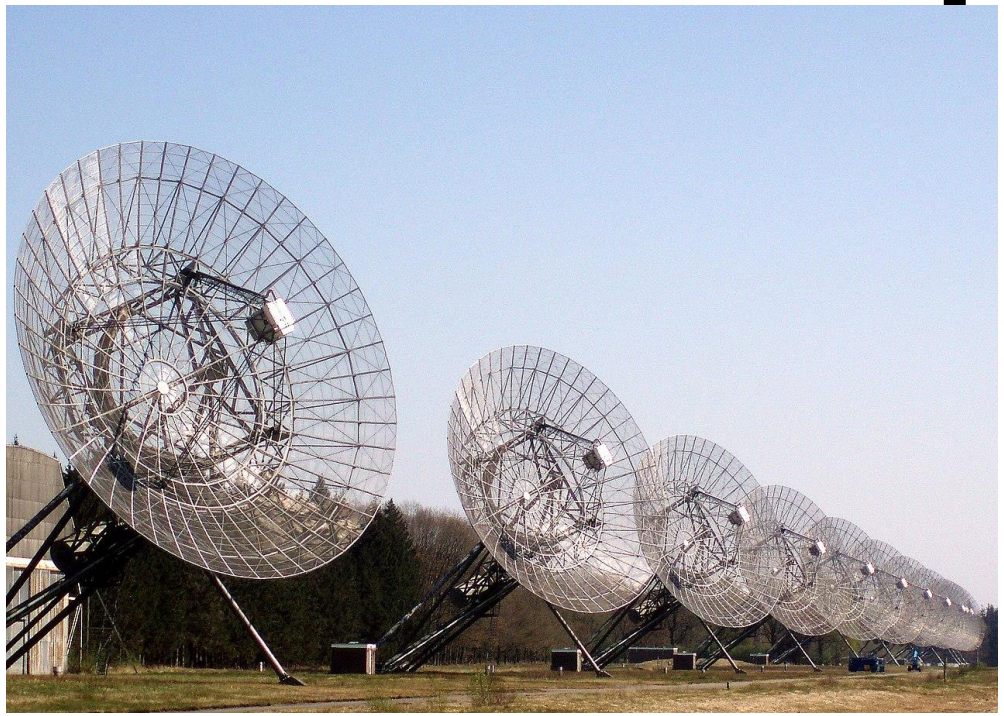




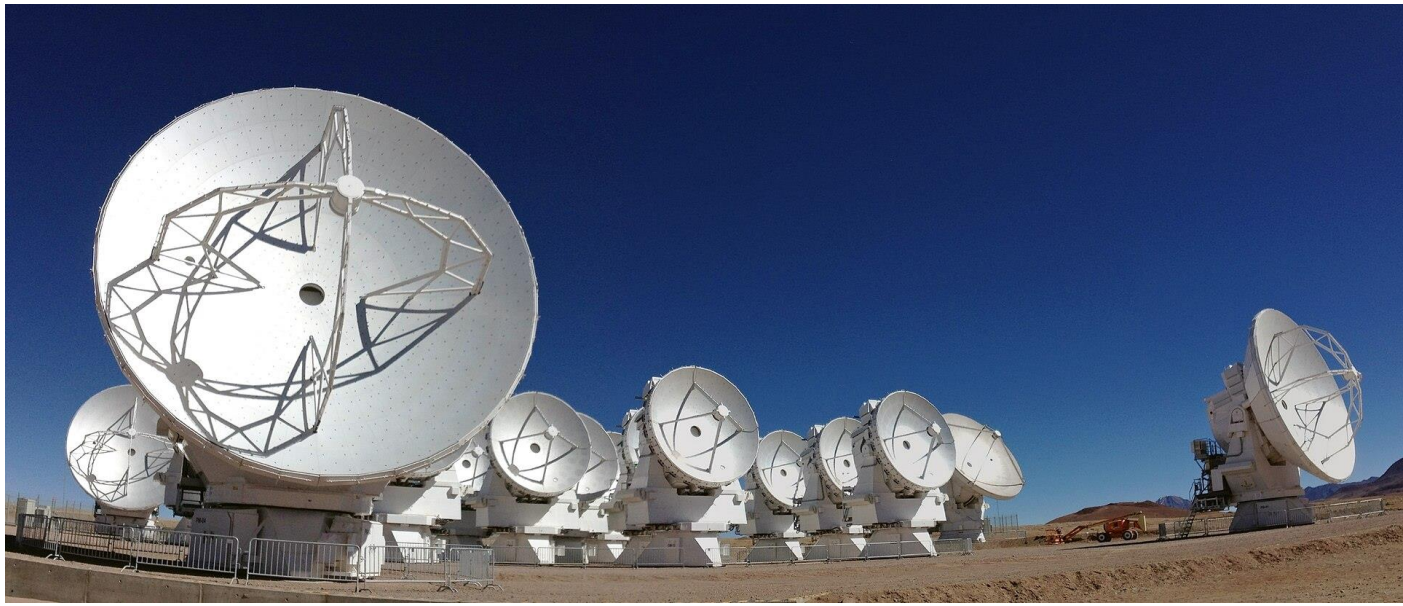
By EHT Collaboration - https://www.eso.org/public/images/eso2208-eht-mwa/ (image link), CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=117933557

ASTRON
Netherlands Institute for Radio Astronomy

Radio telescopes



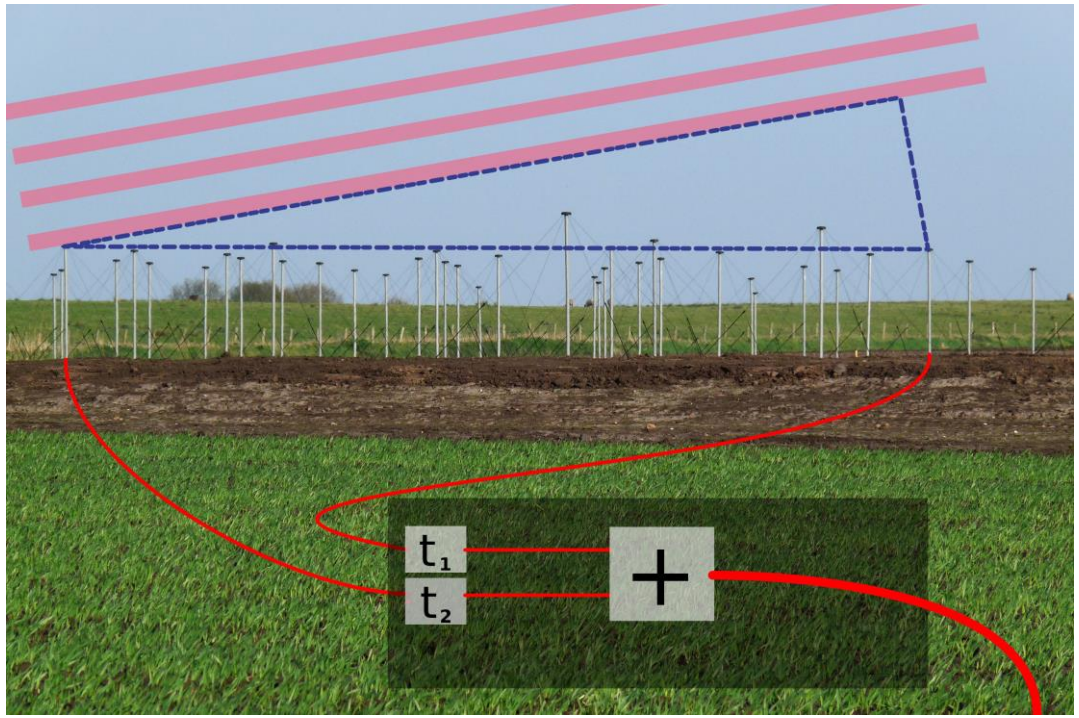
ASTRON WSRT
By Onderwijsgek - Own work, CC BY-SA 2.5 nl, https://commons.wikimedia.org/w/index.php?curid=2098237

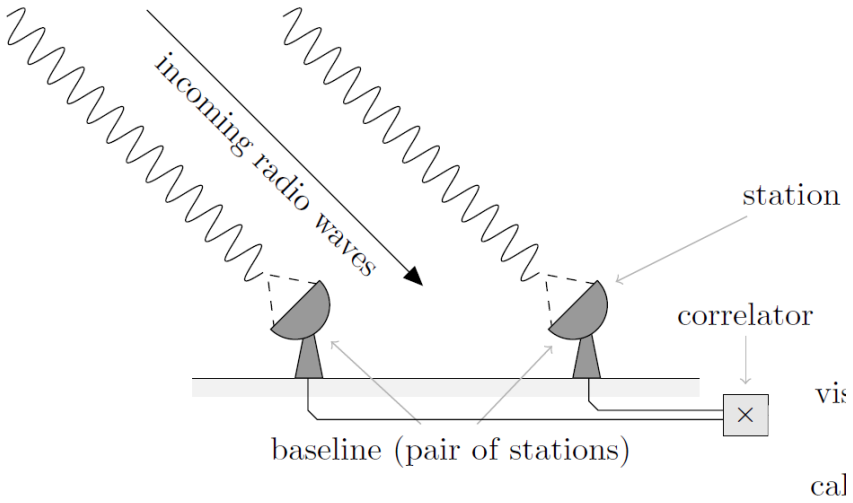


ALMA
By ESO - http://www.eso.org/public/images/ann13040a/, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=25950276

ASTRON
Netherlands Institute for Radio Astronomy

Beamforming and Interferometry





Beamforming many small antennas on station level

- Data reduction
- Define where to point at the sky

Image: M. Brentjens

Radio interferometry, between stations

- Computationally combine multiple stations in to one large telescope

Image: B. Veenboer et al

ASTRON
Netherlands Institute for Radio Astronomy

A generic distributed aperture synthesis radio telescope

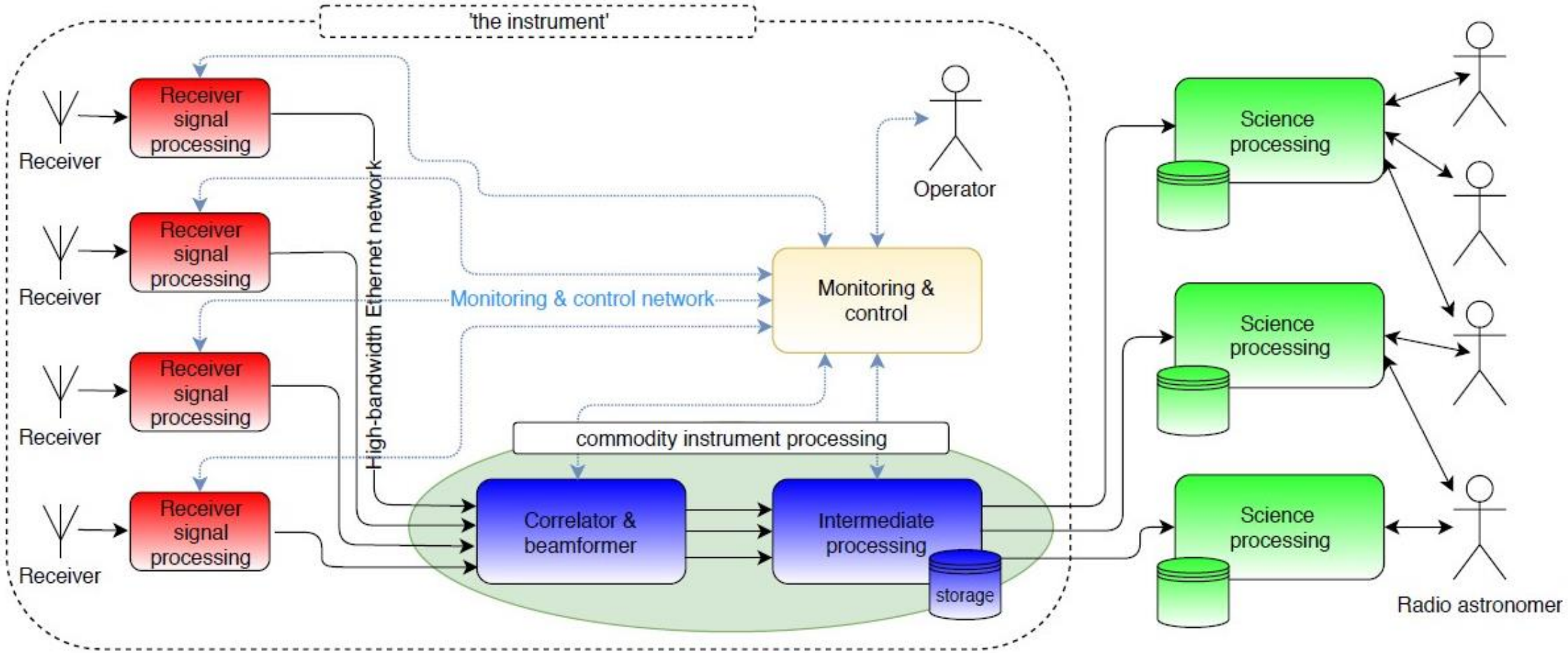
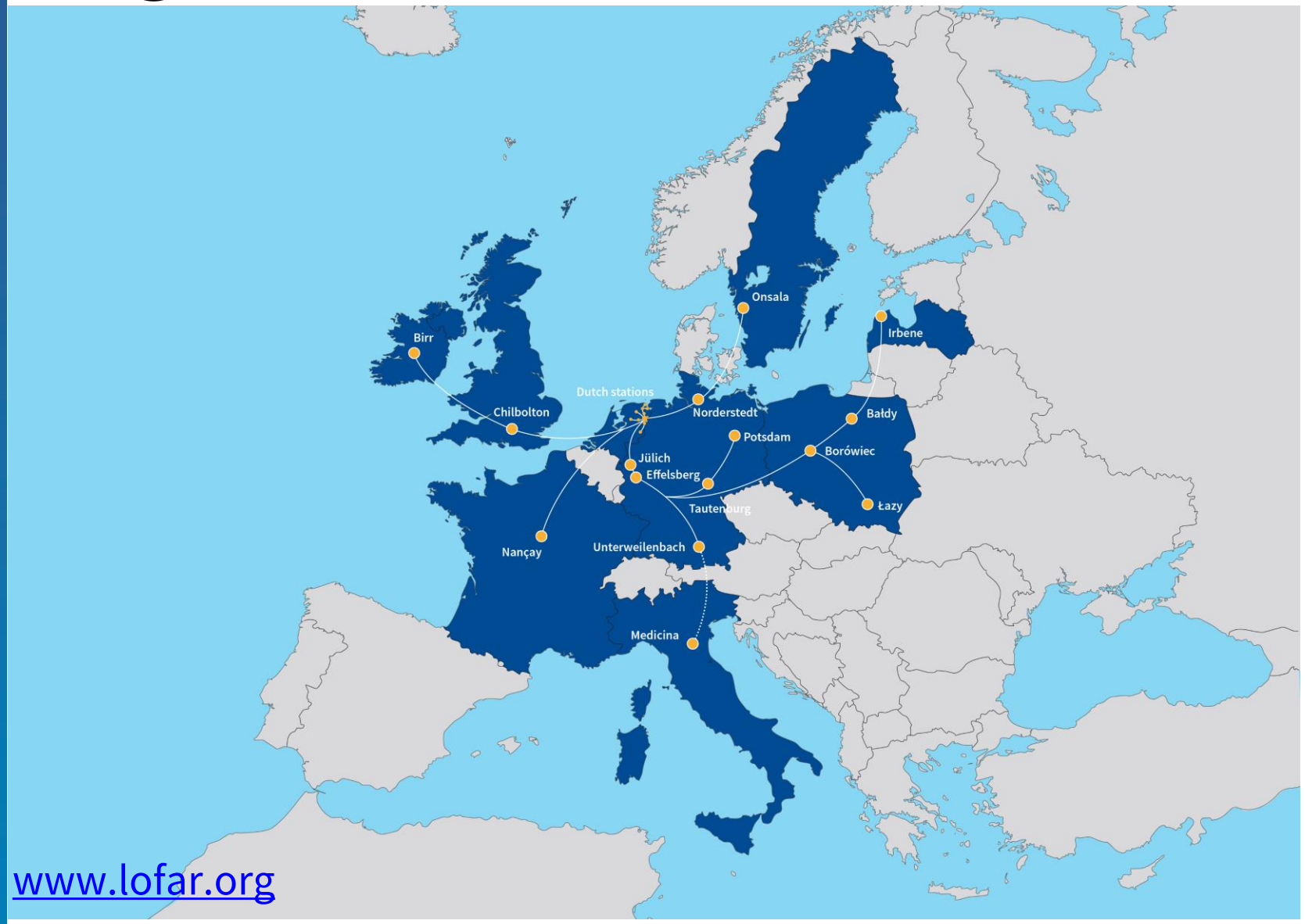


Image: C. Broekema

ASTRON
Netherlands Institute for Radio Astronomy

LOFAR



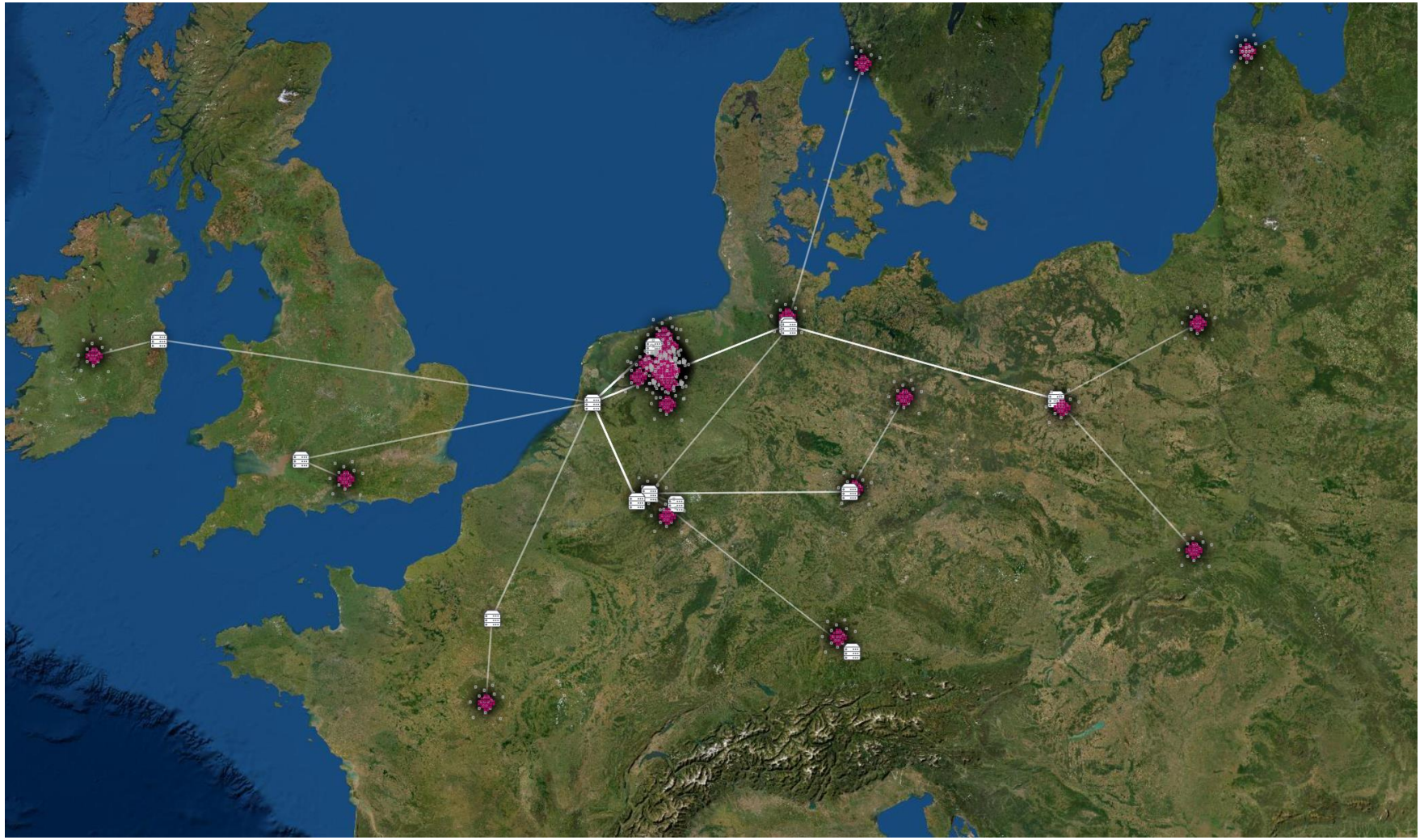
Huge distributed radio telescope!

- 10 – 250 MHz frequency range
- Baselines ≤ 2000 km
- 38 Dutch stations + 14 Intl
- Process all data at a central location

www.lofar.org

ASTRON
Netherlands Institute for Radio Astronomy

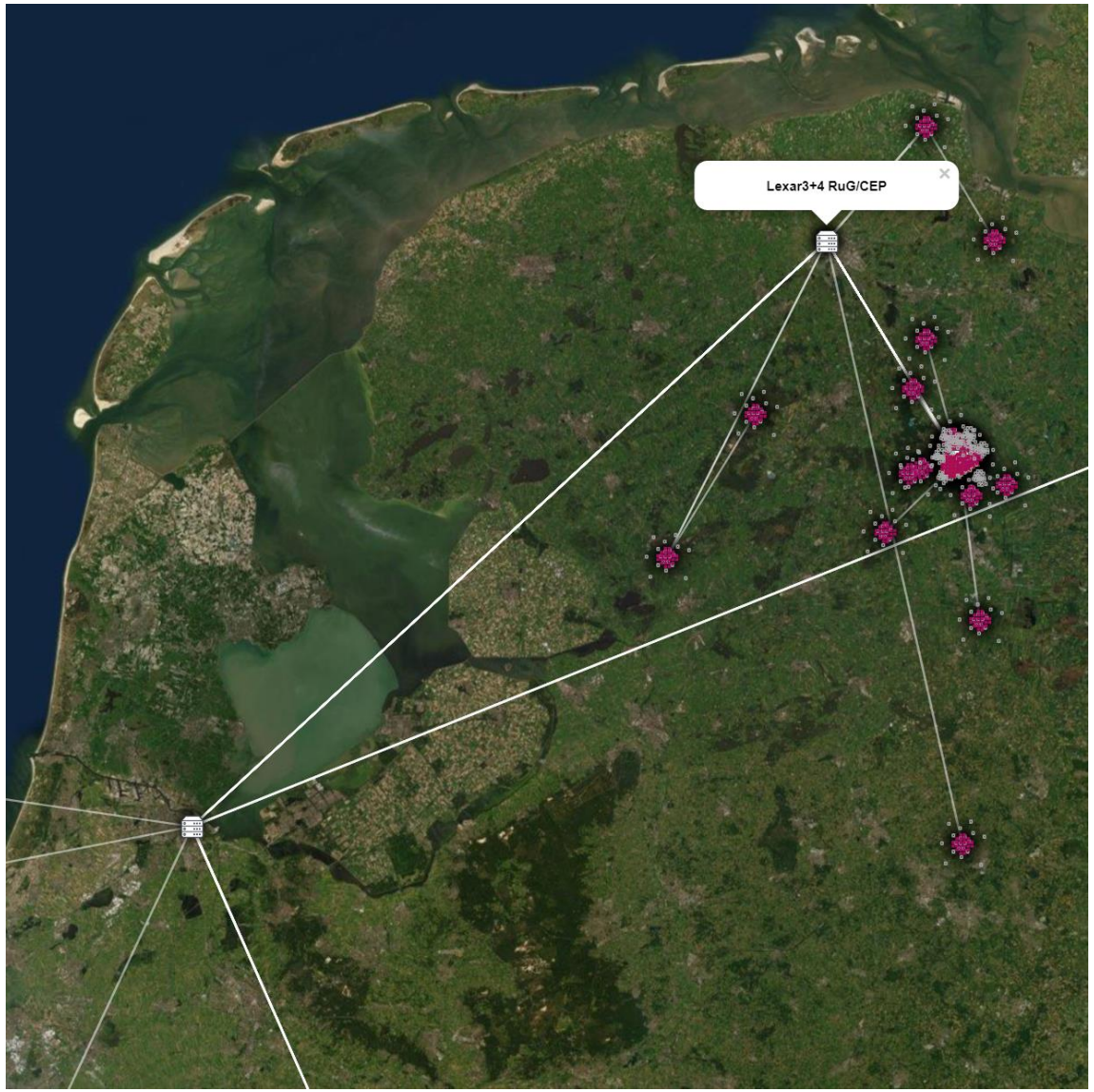
LOFAR stations & network



https://www.astron.nl/lofartools/lofarmap.html

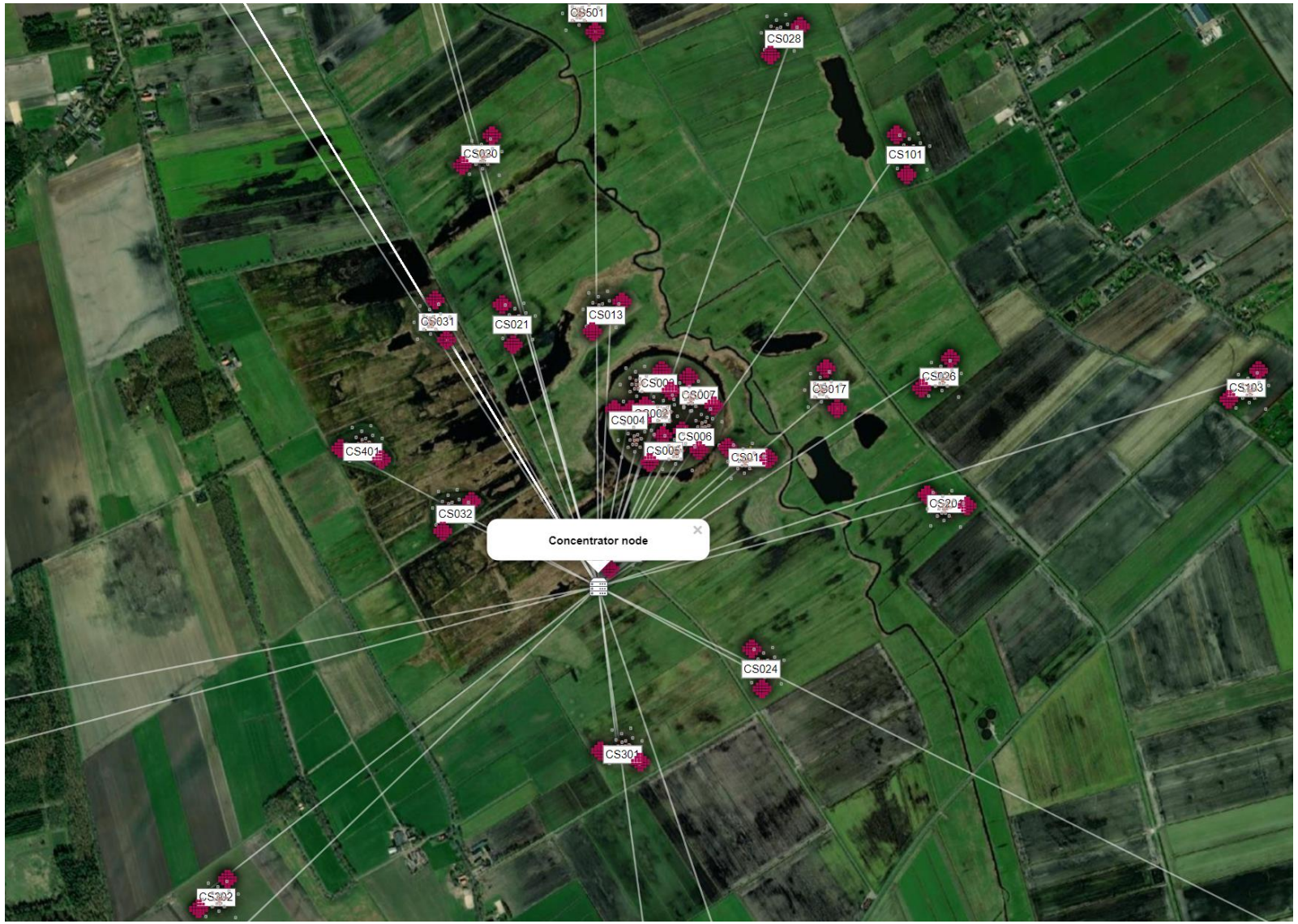
ASTRON
Netherlands Institute for Radio Astronomy

LOFAR Central Processor



ASTRON
Netherlands Institute for Radio Astronomy

Core Stations Concentrator node



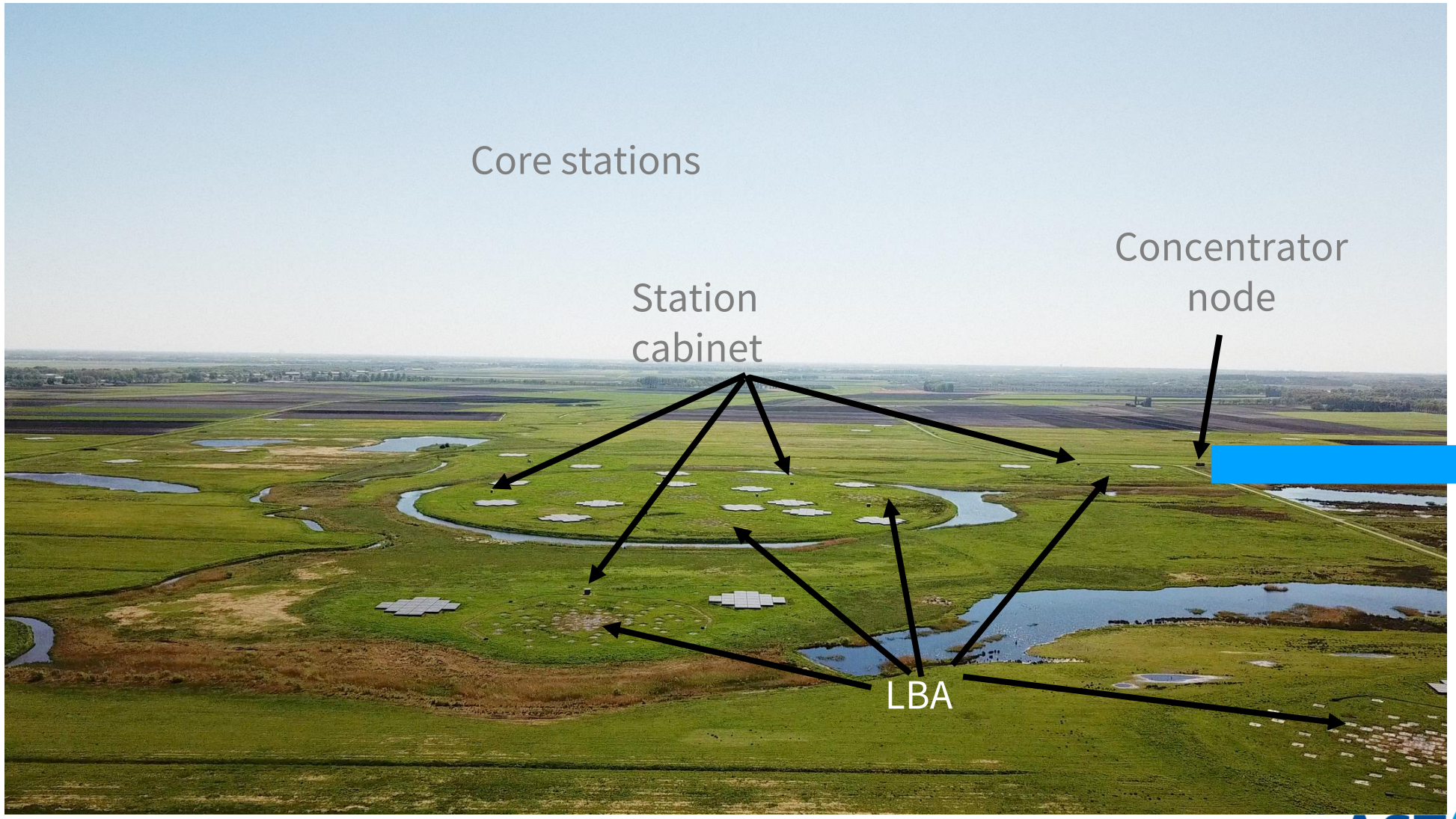
ASTRON
Netherlands Institute for Radio Astronomy

LOFAR Core Stations

Image credit: ASTRON

ASTRON
Netherlands Institute for Radio Astronomy

LOFAR CS

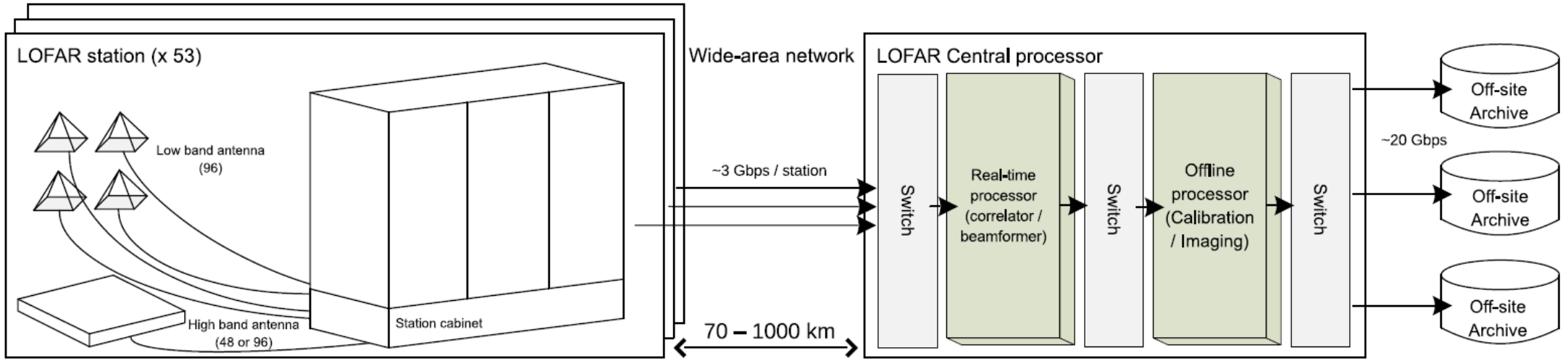


To data center in Groningen

Image credit: ASTRON

ASTRON
Netherlands Institute for Radio Astronomy

LOFAR Central Processor



source: Broekema et al. (2018)

"Standard operation" with beamformed station data

- 3 Gbps Ethernet per station
- LOFAR 2.0 upgrade: approx. 10 Gbps Ethernet

Also, alternative operation modes with 40 Gb/s up to 100 Gb/s data rates and real-time processing requirements

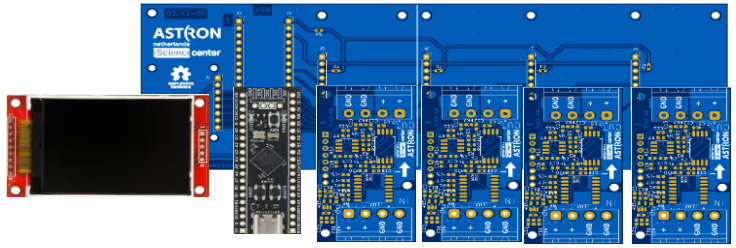
Data rate: 244 beamlets × 16 bits × 2 polarizations × 2 values per sample × 195312.5 samples s^{-1} = 3.05Gb s^{-1}.

ASTRON
Netherlands Institute for Radio Astronomy

Relevance to ORCONF

We strive to be as open as possible: Public money = Public code

- Hardware
  - E.g. PowerSensor3 (demo and poster)
- Firmware
  - More in presentation from Reinier (up next)
- Software
  - Many open source SW packages developed and used in radio astronomy community
  - E.g.: “LOFAR: FOSS HPC across 2000 kilometers; Corne Lukken; FOSDEM2023”
    - https://archive.fosdem.org/2023/schedule/event/lofar_foss_hpc/

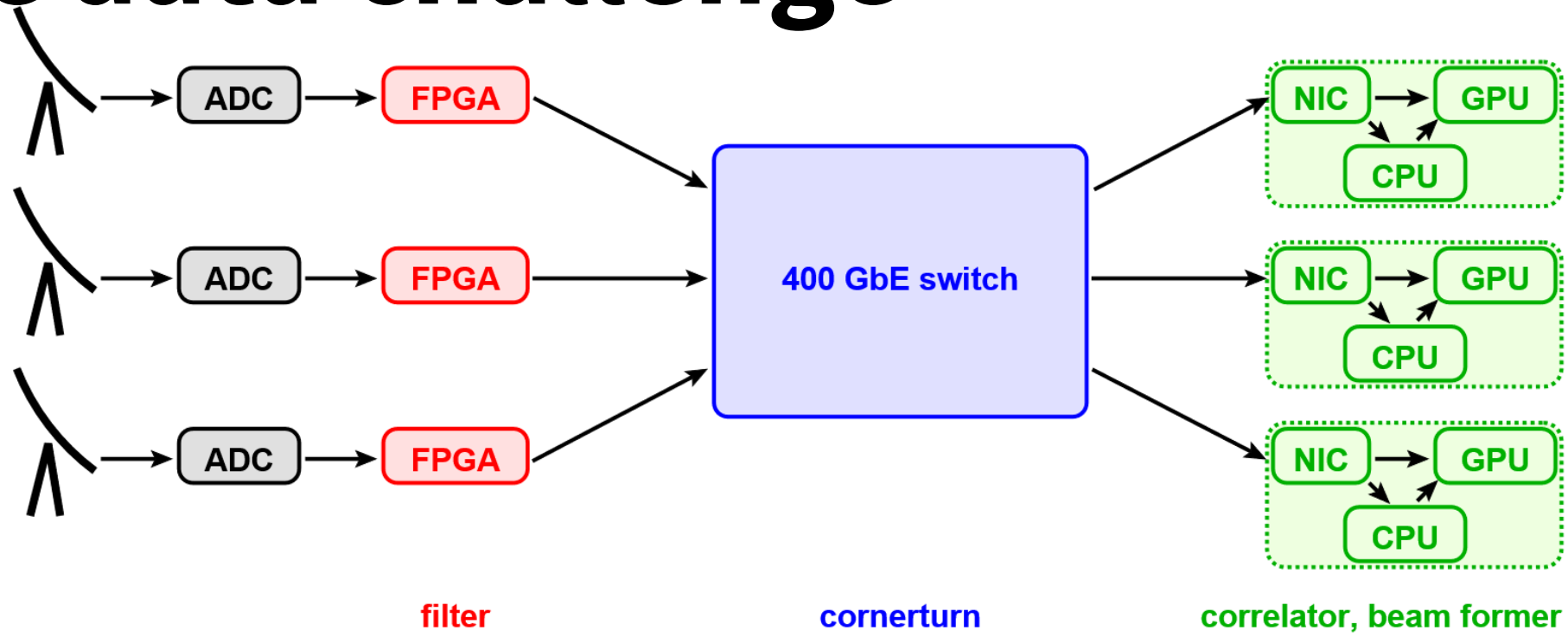


edu.nl/wupy7

ASTRON
Netherlands Institute for Radio Astronomy

RDMA over Converged Ethernet (RoCE)

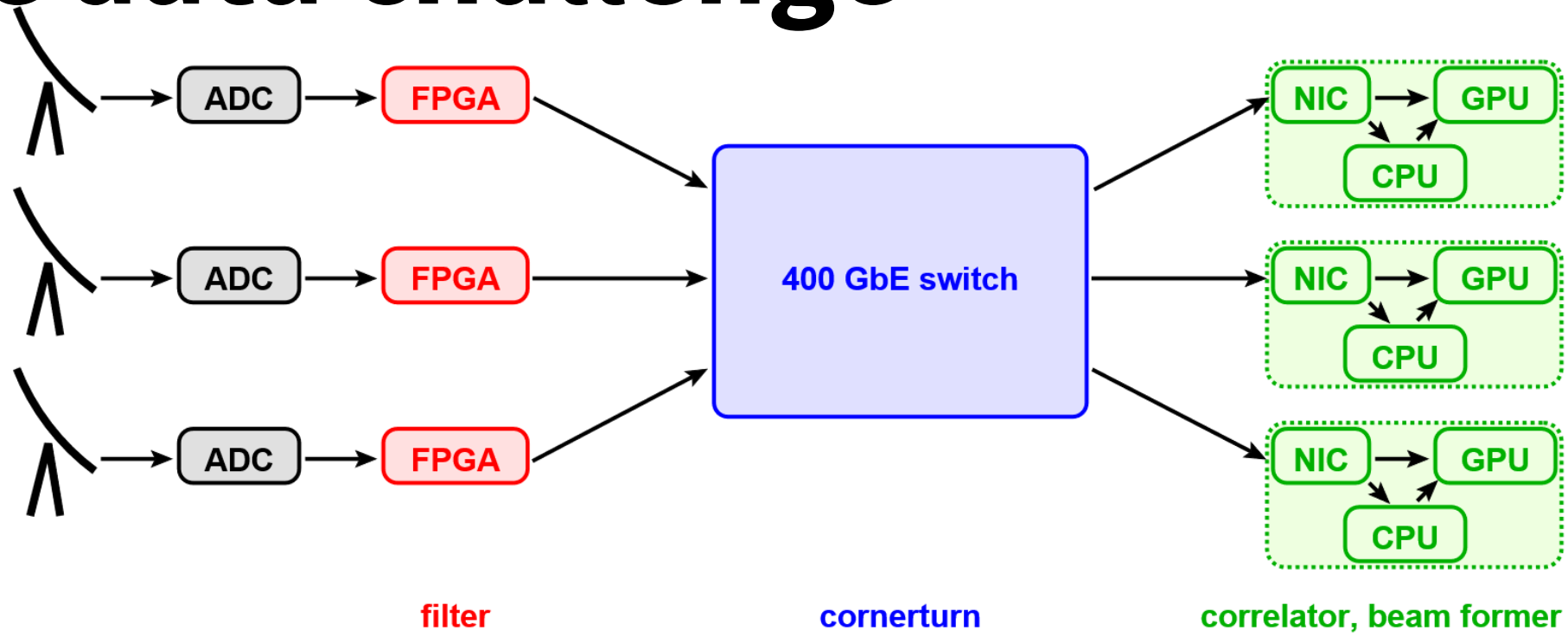
The data challenge



- Building a system with commodity components as much as possible
- Open SW/HW/FW where possible

ASTRON
Netherlands Institute for Radio Astronomy

The data challenge

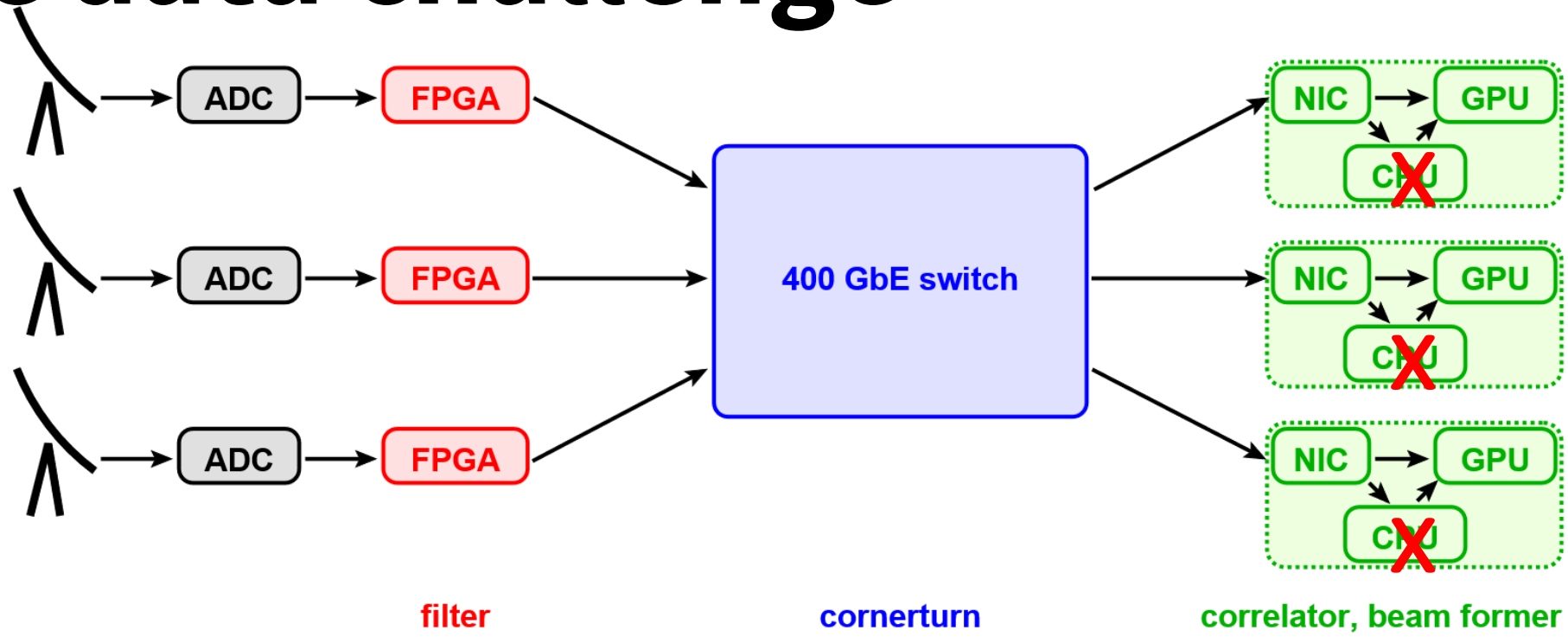


- Stream data from FPGA into server: NIC – CPU - GPU
- Typically UDP/IP over Ethernet
- Next generation systems will use 100 – 400 GbE
- FPGA, CPU – GPU interface can keep up* but the CPU (OS) can't

* (100 Gbit/s PCIe gen4, 400 Gbit/s PCIe gen5)

ASTRON
Netherlands Institute for Radio Astronomy

The data challenge



- Skip OS and CPU: NIC – GPU
- Remote Direct Memory Access over Ethernet (RoCE)

ASTRON
Netherlands Institute for Radio Astronomy

Goal

1. Enable (server side) receive bandwidth beyond 40 Gbps between FPGA and GPU
2. Reduce load on receiving server for < 40 Gbps data rates
   (might use smaller server, reduce power)
3. Reduce power consumption of the total system

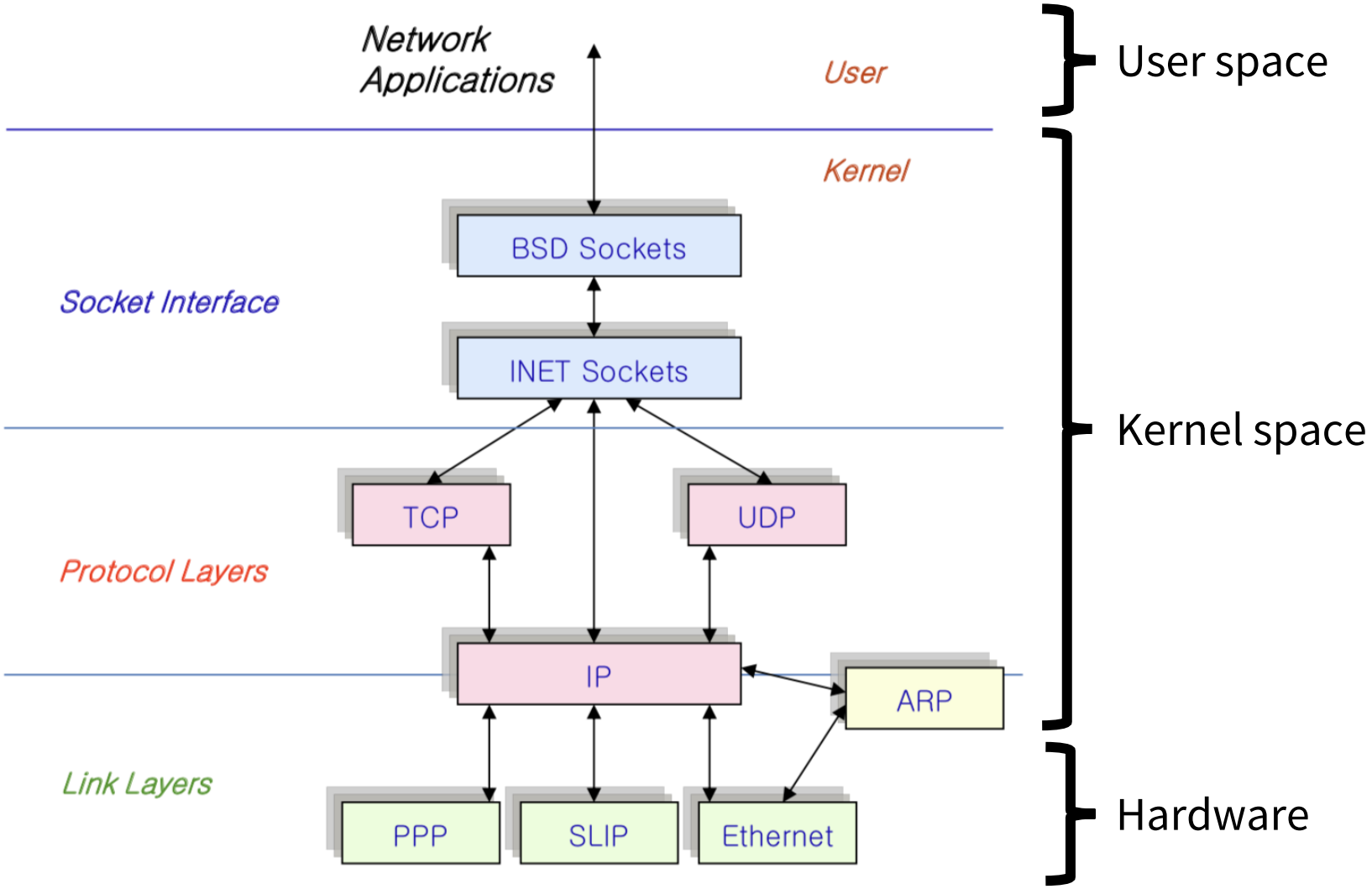
Requirements

- 1) COTS components; 2) Open SW/HW/FW where possible
- Streaming data from FPGA to GPU, no re-transmit possible
- Performance: 1) goodput, 2) CPU load, 3) energy

Several implementations available, but nothing available yet that meets our requirements!

ASTRON
Netherlands Institute for Radio Astronomy

Linux Ethernet stack



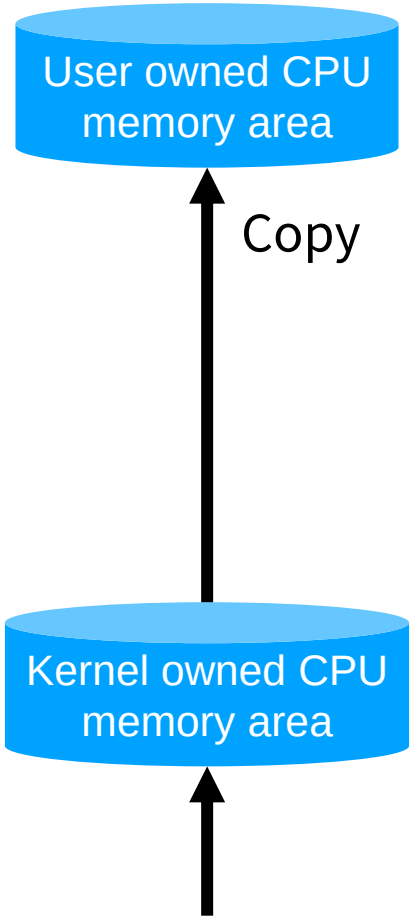


Image: https://wongsingfo.github.io/comp-sys/docs/net

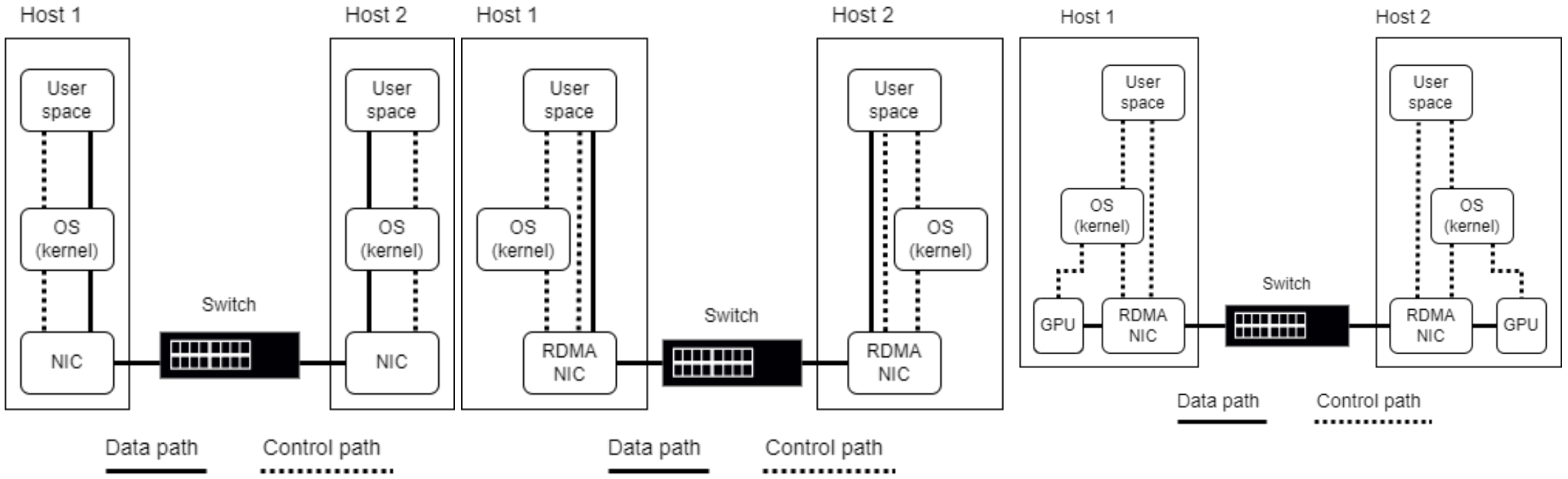
ASTRON
Netherlands Institute for Radio Astronomy

Remote Direct Memory Access

Vanilla software stack

standard RDMA

RDMA with PeerDirect



High level view of RoCE

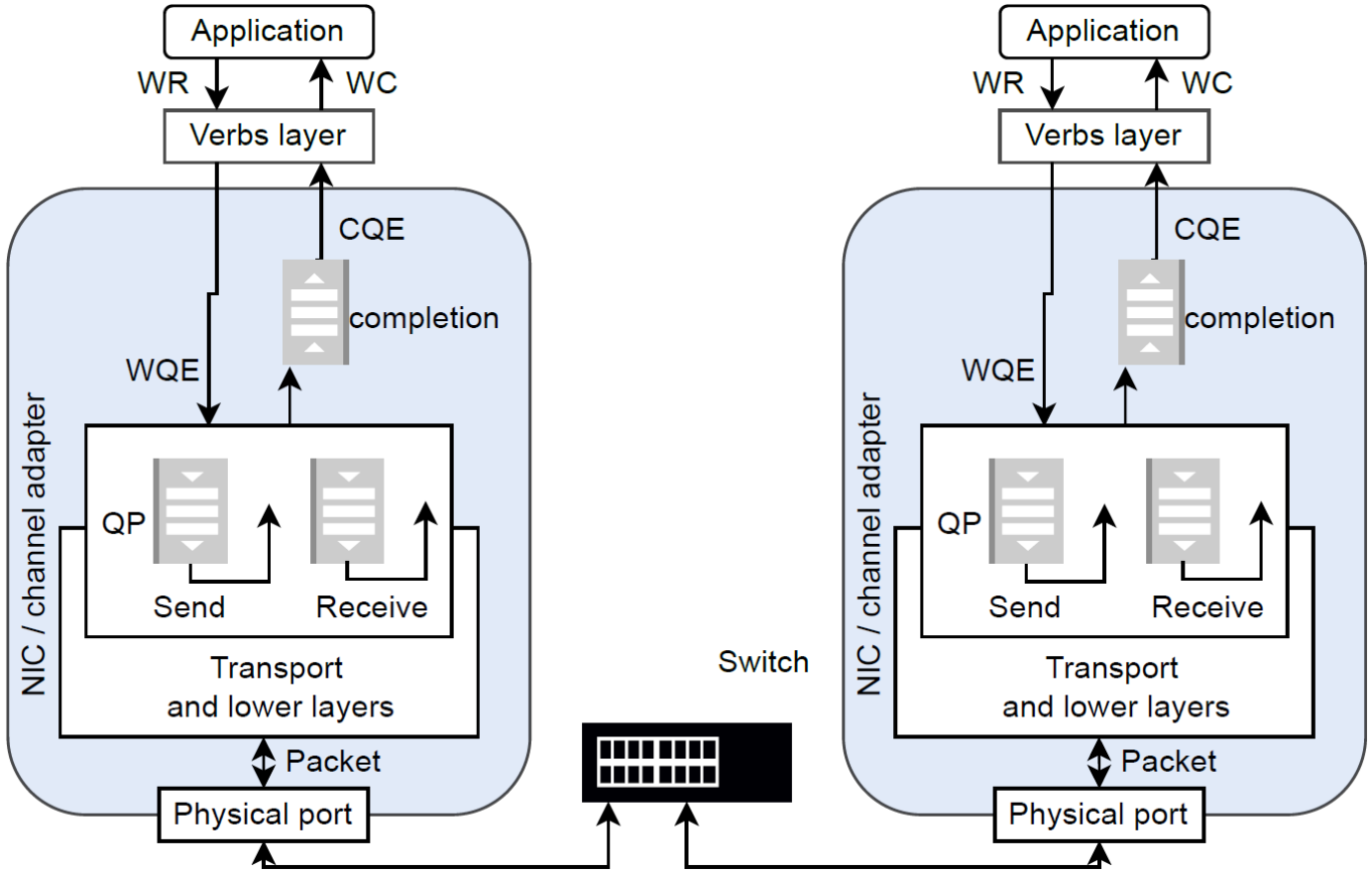


Image: Infiniband specs.

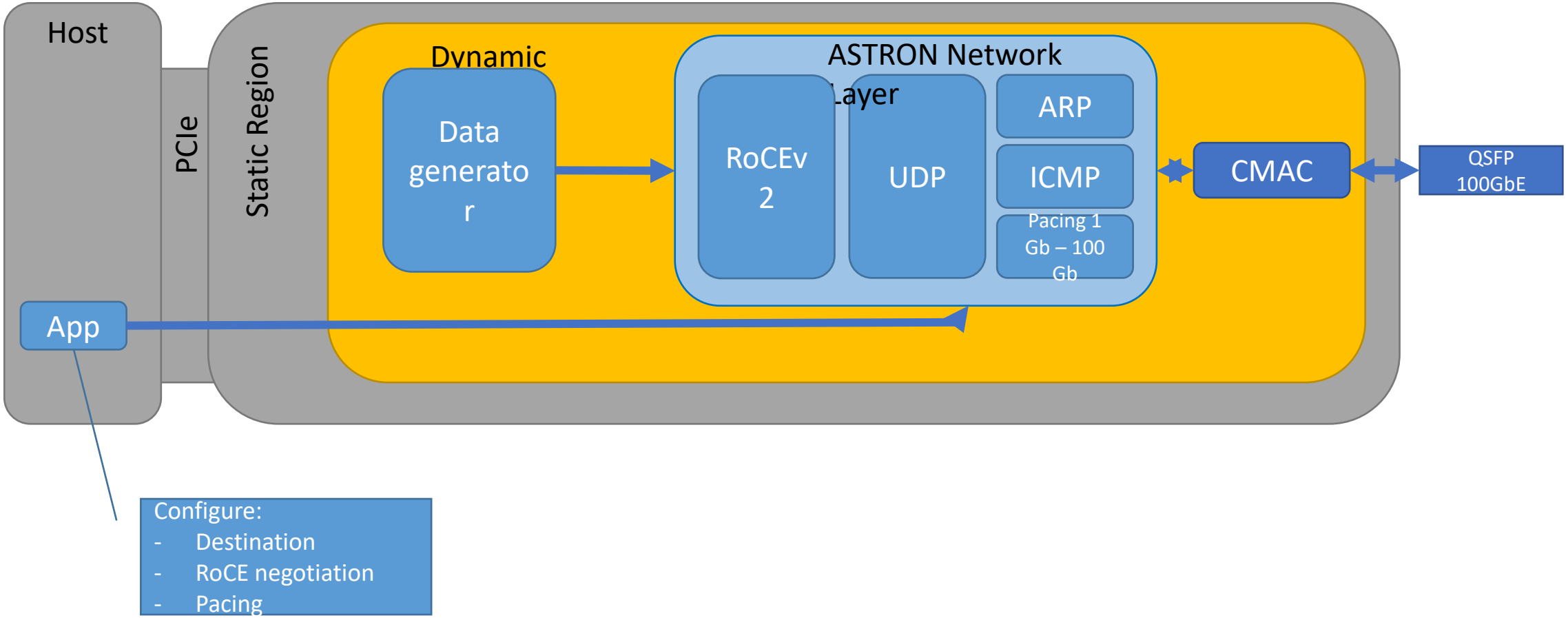
WR → Work Request
WC → Work Completion
WQE → Work Queue Element
CQE → Completion Queue Element
QP → Queue Pair

packetsize ≠ message size
max 4kB max 2GB

Flexible protocol, but many parameters to tune!

ASTRON
Netherlands Institute for Radio Astronomy

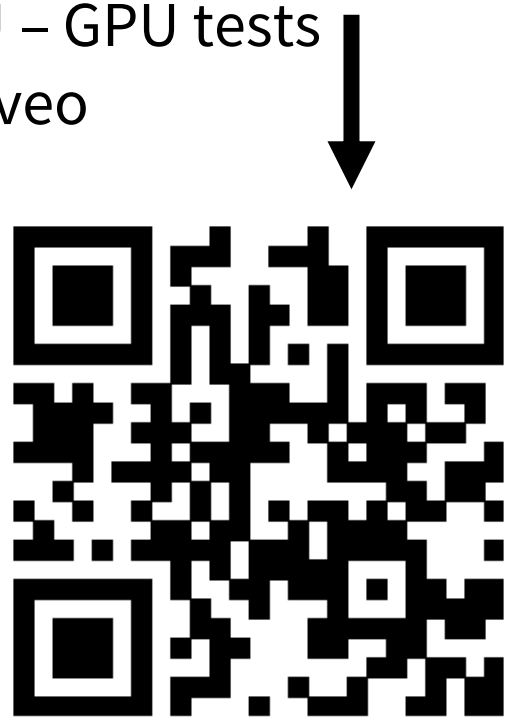
RoCE v2 implementation



Current and Future

- Current
  - Feasibility study RoCEv2 for radio telescope systems, with CPU – GPU tests
  - VHDL components, simulated and partially tested with AMD Alveo
    - UDP network layer + packet pacing, tested up to 100 Gbps
    - Write only header (header + payload + CRC)
    - Open available on ASTRON Git

- Future
  - Add multi packet support
  - Add multi message support
  - Add multi connection (QP) support
  - Move to Intel Agilex 7 (multi 400 GbE support)
  - Scale up to 400 Gbps



edu.nl/7cct8

ASTRON
Netherlands Institute for Radio Astronomy

Questions ?

vlugt@astron.nl

These slides and background, links and background materials:

https://git.astron.nl/vlugt/ORCONF2023

edu.nl/td8du

ASTRON
Netherlands Institute for Radio Astronomy