The Data Explosion in Radio-Astronomy
Virtual Instruments and E-LOFAR

Marco de Vos – ASTRON Director of R&D
(devos@astron.nl)

Early history and near future

- Start of radio-astronomy:
  - Grote Reber, Karl Jansky @ 20MHz (~1930)

- Next large new facility:
  - LOFAR @ (20-)30-80 MHz and 120-240 MHz

- Why the gap?
  - Relatively easy to get some signal out
  - Quite difficult to get further than ~300 sources

LOFAR in summary

- Large distributed radio telescope
  - 32 central (24 Gbps) + 45 remote (2 Gbps) phased array antenna stations, each ~4 soccerfields in size
  - Full scale aperture synthesis array, extends 100 km
  - 10 long-distance stations being discussed (E-LOFAR)

- Two main bands
  - High Band ~ 7,500 tiles 120 – 240 MHz
    - 120,000 m²/250 K @ 150 MHz
  - Low Band ~ 7,500 dipoles 10 – 80 MHz
    - 375,000 m²/45,000K @ 20 MHz

- Digital Radio 40 MHz processing bandwidth
- Extreme agility in time/space/frequency
- Large instantaneous sky coverage
- Fibre network, Software Correlator
- New calibration & RFI mitigation schemes
Phased array stations - aperture synthesis array

LOFAR top-level architecture

Streaming towards science

High Band Antenna (120-240 MHz)
Relevance to the SKA

- Science PathFinder
  - In particular for EoR, Pulsars and Transients

Thunderstorm Events

- Does the Electric field of the atmosphere influence CR radio signal?
- For E>100 V/cm E-field force dominates B-field:
  - Fair weather: E=1 V/cm
  - Thunderstorms: E=1 kV/cm
- Select thunderstorm periods from meteorological data:
  - Clear radio excess during thunder storms
  - B-field effect dominates under normal conditions
  - >90% duty cycle possible

Station based processing

- Input data rate: ~ 460 Gbps
- Output data rate: ~ 2 Gbps
- Processing capacity: ~ 1.5 Tmul/s
- Storage capacity: 96 Gbyte

ITS Spatial Filter Experiment

- ITS sky map at 27.85 MHz, no RFI.
  - Two strong sources (Cas A, Cyg A)
- ITS sky map at 27.81 MHz, strong RFI at (Az, el) = (-1.3, 0) rad
- ITS sky map at 27.81 MHz, fixed null at (Az, el) = (-1.3, 0) rad

© ASTRON/LOFAR 2004 - Reproduction in whole or in part is prohibited without written consent of the authors.
37 Tbps raw data (0.5 Tbps per station)

10 Gbps/station

Central Processor Main Modes

- Correlator for all station inputs (32 MHz) - imaging
- Correlator for Core Stations only and more FOV's - EOR
- Tied Array Beamforming using all stations - pulsars
- Tied Array Beamforming for Core Stations only - transients, pulsars

Correlator

Voltage sum

(u,v) data

Time Series

TFlop Correlator demo

- Real-time correlation of predefined analog signals
- Performance: 97.5% of theoretical maximum

CEP Performance

- Storage: 25 TBytes
- >500 TB
- 45 x 10Gbps
- 80 x 8Gbps
- <1 Gbps

CS1 Imaging Pipeline
The LOFAR calibration challenge

- Pathological Ionosphere (1 rad / 10 sec)
  - Affects source subtraction and imaging
- (Very) crowded fields
  - Source confusion
  - PSF sidelobe confusion: increases noise
- Unstable station beam shapes
  - Affects source subtraction and imaging
- High station side-lobes
  - Bright sources (incl galactic plane and Sun)

LOFAR Calibration
- All Sky Calibration

Peeling WSRT data
- Standard selfcal entire field
- 2-patch peeling (miriad)

Major Cycle
VO lecture – Kapteyn CMV/200 7/02/09

RUG/EoR Science Center

UvA/Transients Science Center

RUL/Survey Science Center

KUN/UHECR Science Center

Bonn Science Center

Joint LOFAR Operations Center

SurfNet6

Geant

NITG/Geo Science Center

KNMI/Geo Science Center

Mapping on GigaPortNG/Surfnet6

<table>
<thead>
<tr>
<th>Location</th>
<th>Institute/Astronomy</th>
<th>Datarate</th>
<th>Role/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groningen</td>
<td>ASTRON 6, 10 Gbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amsterdam</td>
<td>SARAO 60 Gbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leiden</td>
<td>SURFNET5 10 Gbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alphen</td>
<td>Nuffi/Geochemistry</td>
<td>30 Gbps</td>
<td></td>
</tr>
<tr>
<td>Utrecht</td>
<td>RUG/Astronomy 1 Gbps</td>
<td>1 Gbps</td>
<td></td>
</tr>
<tr>
<td>Bonn</td>
<td>MR19 1 Gbps</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tier-0 (CERN):
- Copy of all raw data
- Tier-1 (~8-10 centres)
- Archive 1/n fraction of raw data & reconstructed data
- Regular re-processing of the raw data
- Archiving data from Tier-2 centers
- Provide central grid services: grid accessible computing and data resources
- Support coordination

Tier-2 (~100 centres):
- Individual scientists
- No data archiving

Multi-tier models for LHC & LOFAR

Towards a European Sensor Grid

<table>
<thead>
<tr>
<th>Sensors</th>
<th>2008</th>
<th>2010</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of sensors</td>
<td>77</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Astronomical “LF” antennas</td>
<td>100</td>
<td>7700</td>
<td>9300</td>
</tr>
<tr>
<td>Astronomical “HF” antenna tiles</td>
<td>100</td>
<td>7700</td>
<td>9300</td>
</tr>
<tr>
<td>Geophysical vibration sensors</td>
<td>18</td>
<td>1386</td>
<td>1674</td>
</tr>
<tr>
<td>Geophysical microbarometers (infrasound)</td>
<td>3</td>
<td>231</td>
<td>279</td>
</tr>
<tr>
<td>Agriculture sensors</td>
<td>80</td>
<td>6160</td>
<td>7440</td>
</tr>
<tr>
<td>Other sensors</td>
<td>8</td>
<td>616 t.b.d.</td>
<td>!!</td>
</tr>
<tr>
<td>Total digitized datarate from sensors</td>
<td>0.5 Tb/s</td>
<td>37 Tb/s</td>
<td>45 Tb/s</td>
</tr>
<tr>
<td>Datarate over LOFAR Backbone</td>
<td>10 Gb/s</td>
<td>0.8 Tb/s</td>
<td>1 Tb/s</td>
</tr>
<tr>
<td>Outgoing datarate over SURFnet6</td>
<td>40 Gb/s</td>
<td>60 Gb/s</td>
<td></td>
</tr>
<tr>
<td>European datarate over Géant</td>
<td>20 Gb/s</td>
<td>0.2 Tb/s</td>
<td></td>
</tr>
<tr>
<td>Total processing power</td>
<td>160 Tops/s</td>
<td>190 Tops/s</td>
<td>230 Tops/s</td>
</tr>
<tr>
<td>Distributed at Sensor Fields</td>
<td>1.5 Tops/s</td>
<td>116 Tops/s</td>
<td>150 Tops/s</td>
</tr>
<tr>
<td>Central Processor (including BlueGene)</td>
<td>43 Tflop/s</td>
<td>60 Tflop/s</td>
<td></td>
</tr>
<tr>
<td>BlueGene</td>
<td>33 Tflop/s</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Very Long Baseline Interferometry
First e-VLBI results

LOFAR Performance

<table>
<thead>
<tr>
<th>Frequency (MHz)</th>
<th>$A_{sys}$ (m$^2$)</th>
<th>$T_{sys}$ (m K)</th>
<th>$dS$ 1s (mJy)</th>
<th>$dS$ 10h (mJy)</th>
<th>$dS$ 100h (mJy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>$3.3 \times 10^5$</td>
<td>23k</td>
<td>68</td>
<td>0.35</td>
<td>0.11</td>
</tr>
<tr>
<td>75</td>
<td>$5.2 \times 10^5$</td>
<td>2450</td>
<td>46</td>
<td>0.24</td>
<td>0.07</td>
</tr>
<tr>
<td>120</td>
<td>$3.3 \times 10^5$</td>
<td>820</td>
<td>2.4</td>
<td>0.013</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Approximate sensitivity per beam, with 4 MHz BW and for a single polarization

GLOW – German Long Wavelength Consortium

European Expansion ...

Current discussions:
- Germany ~12 stations
- UK ~2-3 stations
- Italy ~2 stations
- France ~1 station
- EVN t.b.d.
Some actual LOFAR stations

Baselines and Resolution

<table>
<thead>
<tr>
<th>Site</th>
<th>Latitude</th>
<th>Longitude</th>
<th>D(km)</th>
<th>240 MHz</th>
<th>115 MHz</th>
<th>86 MHz</th>
<th>30 MHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS-1</td>
<td>52.947</td>
<td>6.877</td>
<td>0</td>
<td>5.2</td>
<td>10.8</td>
<td>15.5</td>
<td>41.2</td>
</tr>
<tr>
<td>LOFAR 77</td>
<td>50.533</td>
<td>6.883</td>
<td>267</td>
<td>1.0</td>
<td>2.0</td>
<td>2.9</td>
<td>7.7</td>
</tr>
<tr>
<td>Effelsberg</td>
<td>52.400</td>
<td>13.067</td>
<td>422</td>
<td>0.6</td>
<td>1.2</td>
<td>1.8</td>
<td>4.5</td>
</tr>
<tr>
<td>Potsdam</td>
<td>57.400</td>
<td>11.917</td>
<td>595</td>
<td>0.4</td>
<td>0.9</td>
<td>1.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Onsala</td>
<td>51.144</td>
<td>-1.436</td>
<td>602</td>
<td>0.4</td>
<td>0.9</td>
<td>1.3</td>
<td>3.4</td>
</tr>
<tr>
<td>Chilbolton</td>
<td>47.383</td>
<td>1.760</td>
<td>702</td>
<td>0.4</td>
<td>0.8</td>
<td>1.1</td>
<td>2.9</td>
</tr>
<tr>
<td>Nancay</td>
<td>53.017</td>
<td>18.617</td>
<td>786</td>
<td>0.3</td>
<td>0.7</td>
<td>1.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Torun</td>
<td>44.521</td>
<td>11.645</td>
<td>997</td>
<td>0.3</td>
<td>0.5</td>
<td>0.8</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Pioneering new operational models

The path to the SKA

<table>
<thead>
<tr>
<th>Long baseline (km)</th>
<th>NO (km)</th>
<th>NO data rate (Hz)</th>
<th>Technology</th>
<th>Central processor</th>
<th>Other data rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOFAR</td>
<td>142</td>
<td>1.5</td>
<td>EOR</td>
<td>40 Tbps</td>
<td>10 Tbps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-LOFAR</td>
<td>1000</td>
<td>1.2</td>
<td>EOR</td>
<td>40 Tbps</td>
<td>20 Tbps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKA</td>
<td>3000</td>
<td>1.5</td>
<td>EOR</td>
<td>40 Tbps</td>
<td>20 Tbps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKA Pathfinders</td>
<td>9000</td>
<td>1.5</td>
<td>EOR</td>
<td>40 Tbps</td>
<td>20 Tbps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKA</td>
<td>10000</td>
<td>1.5</td>
<td>EOR</td>
<td>40 Tbps</td>
<td>20 Tbps</td>
</tr>
</tbody>
</table>