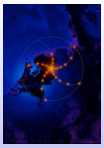
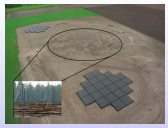


The LOFAR Pulsar Data Pipeline

Anastasia Alexov[1], Jason Hessels[2], Jan David Mol[2], Ben Stappers[3] and Joeri van Leeuwen[2]

1: University of Amsterdam (UvA), 2: Netherlands Institute for Radio Astronomy (ASTRON), 3: The University of Manchester



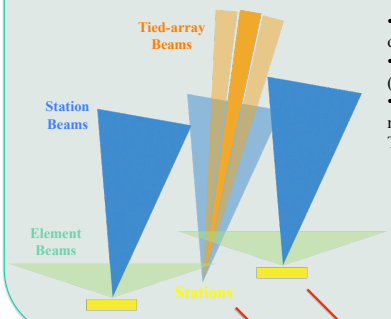
WWW.LOFAR.ORG

Abstract

The LOw Frequency ARray (LOFAR) for radio astronomy is being built in The Netherlands (by ASTRON) and in Europe. LOFAR operates at radio frequencies below 250 MHz. The project is an interferometric array of radio telescopes/stations arranged in clusters that are distributed over an area at least 350 km in diameter. LOFAR will be a breakthrough in the low frequency radio astronomy science domain. Transient radio phenomena and pulsars are one of six LOFAR Key Science Projects (KSPs). As part of the Transients KSP, the Pulsar Working Group has been developing the LOFAR Pulsar Data Pipeline to look at known pulsars as well as search for pulsars within an all-sky survey. The pipeline is being developed for the Blue Gene/P (BG/P) supercomputer and a large Linux cluster in order to utilize enormous amounts of computation capabilities and data streams of up to 23TB/hour. The LOFAR pipeline output will be using the Hierarchical Data Format 5 (HDF5) to efficiently store large amounts of numerical data, and to manage complex data encompassing a variety of data types, across distributed storage and processing architectures. We present the LOFAR Pulsar Data Pipeline overview, the pulsar beam-formed data format, the status of the pipeline processing as well as our future plans of developing additional transient pipelines.

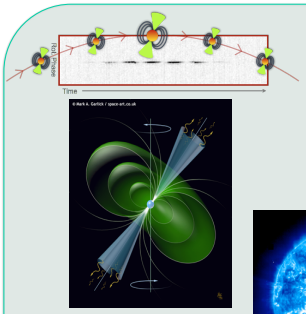
LOFAR Observing

Low Band Antennas (LBA): 10-80Mhz
High Band Antennas (HBA): 100-240Mhz

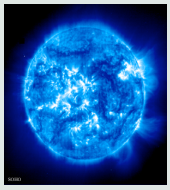


- One to N Stations can be used in an observation
- "Imaging" is the main observing mode (~50% of available telescope time)
- "High Time Resolution" observing mode (~20%) will be widely used for Transient Science (see panel on right):
  - Forming multiple beams is a difficult observing mode
  - Beam-formed modes trade spatial resolution for time resolution
  - Data rates as high as 23TB/hr
  - Pulsar Pipeline effort can be applied to other high timing resolution pipelines

LOFAR "Transient" Science & Requirements on Processing



Pulsar/Beam-Formed Pipeline has many scientific applications. The following areas of study set the requirements on processing for each of the modes:

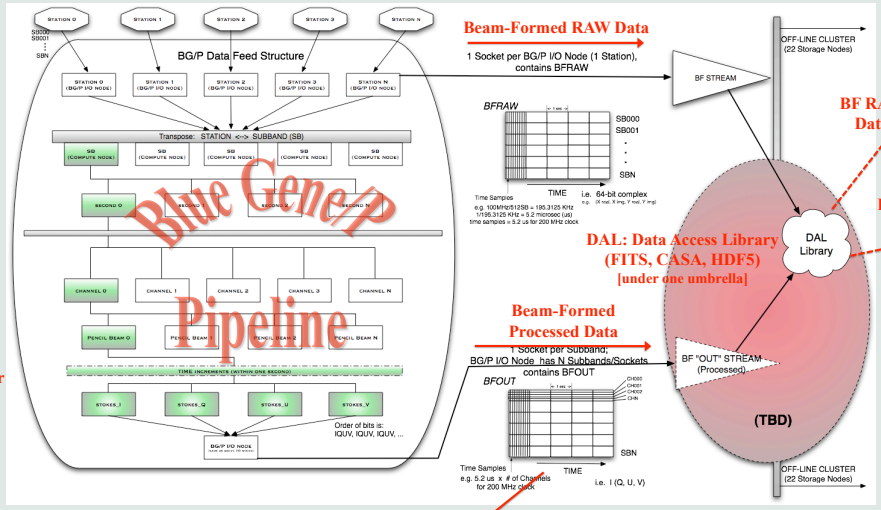


- Known pulsars - channelization, Stokes Parameters, dedispersion, folding, Radio Frequency Interference (RFI) excision
- Pulsar/fast transient survey - channelization, dedispersion [10,000 trial Dispersion Measures (DMs)], RFI excision, searching
- Planets, Sun, flare stars - channelization, RFI excision, dynamic spectra

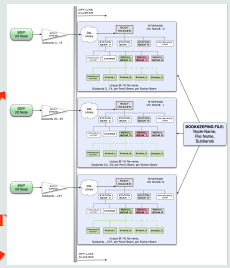
LOFAR Beam-Formed Data Flow

Coherent: Tied-Array (Pencil) Beams Processing:

Stations → Transpose → Polyphase Filter → Beam Forming → Tied-Array Beam Forming → Stokes/Square Power → 2nd Transpose (TBD)



Off-Line Cluster (TBD 2010)
Data Storage using distributed HDF5:



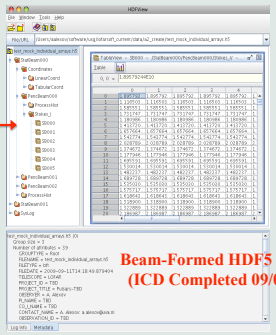
Off-Line Pulsar Pipeline Processing:

- Integrate well-used Pulsar software (*i.e.* PRESTO & TEMPO) into the LOFAR software infrastructure
- Re-use LOFAR Imaging Pipeline Framework for Pulsar Pipelines
- Create "known" pulsar and "survey/ searching" mode pipelines
- Integrate with HDF5 data format
- Reduce 23TB/h raw ->10GB data

Beam-Formed Data Format: HDF5

Radio Astronomy does not have a single data standard (CASA, FITS, Binary, *etc.*) in contrast to Optical (FITS). LOFAR data will use HDF5 (Hierarchical Data Format 5):

- Versatile data model; can represent complex data objects and variety of metadata.
- Portable file format with no limit on the number or size of data objects in the collection.
- Multi-platform software library for single to massively parallel systems.
- Integrated performance and optimization features.
- Handles huge datasets (no 2GB file limit).
- Allows for distributed file structure on file systems
- Tools and applications exist for managing, manipulating, viewing, and analyzing HDF5 data:
  - HDF5 Tools – standalone command-line tools to view and edit data
  - HDFView – file visualization:
    - view a file hierarchy in a tree structure
    - create new file, add or delete groups and datasets
    - view and modify the content of a dataset
    - add, delete and modify attributes
    - replace I/O and GUI components such as table view, image view and metadata view
- LOFAR project has developed the DAL (Data Access Library) which puts HDF5, FITS & CASA under one software library umbrella

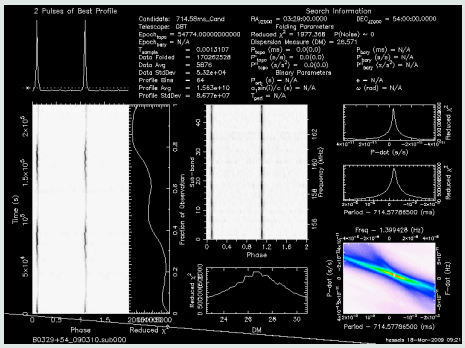


Beam-Formed HDF5 File (ICD Completed 09/09)

HDF5 Home Page: www.**hdfgroup.org**

Beam-Formed Toolset (TBD)

PRESTO: Pulsar software diagnostic plot – pulse found!



PRESTO credit:
www.cv.nrao.edu/~sransom/presto/
Software written by Scott Ransom

PRESTO

- PRESTO is a well-known pulsar software tool suite and is tested within the Pulsar community
- Has quicklook/diagnostic plots, statistical summaries; it allows for quick searches for candidate pulsars
- Using PRESTO in the LOFAR system saves FTEs reuses software
- Any improvements on PRESTO can be fed back into the Pulsar community for other PRESTO data types

•Automation of PRESTO's tools using parallel processing and distributed file-systems will help solve the enormous challenge of reducing 23TB/h of data into 10GB of highly processed data: a pulsar "search survey" observation is reduced to one **new pulsar found**!

LOFAR · University of Amsterdam Faculty of Science · NWO · ASTRON