Data Processing in the LOFAR Era

Jaap D. Bregman

ASTRON, Oude Hoogeveensedijk 4, P.O. Box 2, 7990 AA Dwingeloo, The Netherlands
Email: bregman@astron.nl, web: http://www.astron.nl

Purpose of presentation

- Make clear the LOFAR postcorrelation processing problem
- Even more serious for Wide Field High Resolution SKA
- Indicate approach
- Start discussion on solutions and implementations

Introduction

- The Problem
  - AIPS can handle ~30 stations & ~300 freq on 1 laptop
  - LOFAR has 90 stations & 50k freq so 1000 laptops needed
  - Wide field imaging ~5^{o} with 1000 km at 150 MHz
    - needs ~1000 facets of 1000^2
    - Another factor 100 in laptops to give ~200 Tflops
  - BlueGene/L ~34 Tflops @ ~1 Tb/s I+O

- Solution
  - Forget conventional facetting and cleaning
  - Fast (U,V,W) facetting approach
  - Geometric (U,V,W) projection onto array snapshot U,V-planes
  - Projection of snapshot planes on tangent sky U,V-plane
  - Solve stations gains for <10 strongest sources per snapshot
  - Subtraction of <50 strongest sources per snapshot with spatial interpolated complex station gains
  - Averaging of spectral channel images into continuum images removes any side and grating lobe

Overview

- LOFAR
- LOFAR correlation data rates & processing
- LOFAR imaging characteristics
- Wide Field Imaging basics
- Source Subtraction in U,V-domain
- Fast Quadrant Facetting
- LOFAR Facet processing requirements
- LOFAR source subtraction requirements
- Conclusions

LOFAR

- Stations in exponential shell distribution
  - 32 stations 486 baselines < 2 km
  - 57 stations 1596 baselines < 14 km
  - 77 stations 2926 baselines < 100 km
  - 90 stations 4005 baselines < 1000 km

- LBA 65 m stations 96 dipoles (10)30 – 80 MHz

- HBA 55 m stations 1536 dipoles 120 – 240 MHz

LOFAR correlation data rates & processing

- From available 80/100 MHz receiver bands
  - 32 MHz processed in 0.61 / 0.76 kHz channels
  - Core stations 24 beams @ 32M CS/s of 2*4 bit
  - Remote stations 8 beams @ 4M CS/s of 2*16 bit
  - Other beam/bit/rate combinations possible

- Agregate input rate
  - 90 st * 2 pol * 1 filed * 2*16 bit * 32M CS/s = 180 Gb/s
  - 32 st * 2 pol * 24 field * 2*4 bit * 32M CS/s = 393 Gb/s

- Output data rate
  - 1 field * 4005 basel * 4 pol * 2*32 bit * 50k CS / 0.1 s = 512 Gb/s
  - 24 field * 486 basel * 4 pol * 2*32 bit * 50k CS / 1 s = 149 Gb/s

- Correlation power
  - 1 field * 4005 inf * 4 pol * 50k CS * 5 Flop/CS * 0.61 kHz = 2.4 TFlop/s
  - 24 field * 486 inf * 4 pol * 50k CS * 5 Flop/CS * 0.61 kHz = 7 TFlop/s

- BG/L with ~34 TFlop/s and ~900 Gb/s is correlator output limited
  - Process and average data before storage to disk
  - ~25 Tflop/s available for post correlation processing

LOFAR Imaging Characteristics

- High Resolution Wide field continuum imaging
  - Spectral bins > 0.61 kHz
  - Time bins > 0.1 sec
- Station main beam
  - Amplitude Varies while tracking a point in the sky
  - Elliptical width varies with zenith angle
  - Hardly influenced by complex gain of dipoles
- Station side lobes
  - -17 dB for randomized expo shell LBA
  - -25 / -60 dB for tapered regular HBA
  - Rotation averaging ~20 dB lower to bring gratings <-20 dB
  - 10^{-2} element complex gain errors give relative errors in all lobes of ~10^{-3} divided by the voltage pattern
  - Only few sources in all sky side lobes need to be solved and subtracted per snapshot of 10 sec & 10 MHz
- Bandwidth smearing for continuum imaging already effective in core

Wide Field Imaging Basics

- For a planar array we have the (F)FT between U,V and l,m
- For a quasi-planar array
  - project (U,V,W) points for direction θ_0 on horizontal U,V-plane
  - FT valid for radius $\Delta\theta$ around θ_0 with $\Delta\theta = (\Delta\phi / \pi W)^{1/2}$
  - And max tolerated phase error $\Delta\phi \sim 0.03$
  - Earth curvature for distance +/-500 km Z = -20 km
  - For HBA beam at 150 MHz
    - $\Delta\theta = 10^{-3}$ and resolution $\delta\theta = \lambda/B = 2\ 10^{-6}$ for 2 pixels
    - beam $\theta_{fwhm} = 0.044$ so 500 facets of $\sim 2000^2$
  - For LBA beam at 40 MHz
    - $\Delta\theta = 2\ 10^{-3}$ but resolution $8\ 10^{-6}$
    - beam $\theta_{fwhm} = 0.14$ so 1200 facets of $\sim 1000^2$
- Projection corrected snapshot images can be combined by projection of each horizontal U,V-snapshot-plane on sky tangent plane
- Array beam is average of array snapshot beams
- Station beam is also average of station snapshot beams

Source Subtraction in U,V-domain

- Exact relation from sky brightness to (U,V,W) visibility by Measurement Equation
- Inversion by (F)FT only to image facet of limited sizelimited on sky
- Separate projection corrected (U,V) facet for each (l,m) image facet
- Subtract exactly all sources in sky that would give synthesized side lobes above noise in snapshot image
- ~100 sources in main beam >1σ per baseline per 10 MHz & 10 s
- These have also side lobes stronger than 1σ per snapshot
- No more than ~N_{st}/2 <16 independent complex gains can be solved per station is enough for ionospheric phase screen solution

Fast Quadrant Facetting

- For each (U,V,W) facet we need to phase shift all visibilities to centre of corresponding (l,m) facet
- The (U,V,W) facets need less spectral & temporal resolution and can be averaged after phase shifting
- We have 200M visibilities for 200k snapshots in ~12 hour
- We need ~500 facets so time and frequency averaging factor of ~22 from 0.1 to 2.2 sec and from 0.6 to 13 kHz
- Define 4 (l,m) quadrant facets and 4 corresponding (U,V,W) facets
  - Phase centre all (U,V,W) data for each of the 4 (l,m) quadrant centres
  - Average 2 time slices and 2 frequency channels
  - Each interferometer gets 2 station phase corrections
  - Total amount of data remains te same

LOFAR Facet Processing Requirements

- (U,V,W) facet quadrant processing
  - 4005 inf * 4 pol * 50k CS * 4 quad * 2 stations * 5 Flop/CS / 0.1 s = 320 GFlop/s
- Repeat process 5 times for 1.6 TFlop/s which is comparable to correlation
  - and get 1024 horizontal (U,V,W) facets
  - with 4005 baselines averaged over 3.2 s and 19 kHz
  - FFT trick saves factor 32 in processing
- After 20,000 s we have
  - 6250 horizontal facets * 4005 baselines to grid into a 2000*2000 image
  - i.e. Complete U,V filling for each 19 kHz bin
- Projection correction to horizontal (U,V) plane and then 6250 horizontal facets to a tangent sky plane takes less than 1 TFlop/s
- Self calibration involves only the facets of the 10 strongest sources

LOFAR Source Subtraction Requirements

- Each horizontal U,V,W-visibility needs to be corrected for ~100 sources
- We need two station phase corrections per object per baseline
- Processing is comparable to the shift correction
  - 4005 inf * 4 pol * 50k CS * 2 stations * 5 Flop/CS / 0.1 s = 80 GFlop/s
- For sources outside the facet we need time and bandwidth decorrelation corrections
  - Baseline dependent
  - Factor 2-10 more expensive than phase correction only
- For 100 sources we need 16 - 80 Tflop/s

Conclusions

- Forget conventional facetting
  - Geometric (U,V,W) projection onto array snapshot U,V-planes allows relatively large image facets
  - High Resolution Wide Field LOFAR needs ~1000 facets
  - Fast Quadrant Facetting can be handled for the 1000 km
  - Resulting frequency bin width ~20 kHz and 3 sec snapshot facets
  - After 12 h complete U,V sampling
  - Processing power comparable with correlation power

- Forget conventional cleaning
  - No more than ~100 sources need to be subtracted in U,V,W-domain
  - Processing power only available for less sources at highest resolution
  - Selfca
  - libration on ~10 strongest sources in beam can be handled

- Flexible use of General Purpose processing platform for correlation and calibration
  - At 32 MHz bandwidth the correlation power is 10% of the post correlation
  - Valid approach for SKA at 320 MHz and more stations

- Averaging over more bins reduces side lobes to invisible levels

Suggestions ?