University of Groningen

The LOFAR Epoch of Reionization Data Model
Lampropoulos, Panagiotis

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 28-08-2020
Chapter 6

Conclusions

Our investigations have always contributed more to our amusement than they have to knowledge.

Will Rogers

The Epoch of Reionization demarcates the phase transition of the Universe, during which the hydrogen gas turned from the neutral to the ionized state. The last two decades have seen an increased theoretical effort to understand the interplay between the physical processes that characterize the EoR. Nonetheless, observational evidence is still scarce and indirect. The LOw Frequency ARray (LOFAR) is one of the currently designed instruments with the aim to probe the end of the Dark Ages and cosmic reionization and will reach the final stages of its construction at the end of this year (2010).

LOFAR adds a number of new capabilities that were not available in current instruments, such electronic beam-forming, multiple beams and distributed, as well as centralized signal processing. Moreover, a host of instrumental and environmental issues have to be addressed. At the operational frequencies (10-240 MHz) of LOFAR the ionospheric effects as well as the Galactic foregrounds are strong. The diffuse emission from the Galaxy adds another complication to the imaging problem. Furthermore, the station beams vary rapidly and due to the close proximity of the elements and the receivers effects like mutual coupling and cross-talk have to be accounted for. The EoR experiment plans to observe in the frequency range above 110 MHz, that lies just above the FM band. This frequency range is quite populated by terrestrial and air communication channels due to its low atmospheric absorption. Low-power, broadband RFI is particularly difficult to address. Thus, radio frequency interference mitigation is important. Atomic clocks complicate the efficiency of the system. The effective noise, which is the sum of the thermal, receiver noise, confusion noise and calibration noise is also several times stronger than the signal. Finally, the sheer data volume of 1-2 petabytes poses a computational challenge and constrains the selection of a proper algorithm.

Extracting the EoR signal from the corrupted observational dataset requires understanding of the data at a fundamental level. This thesis focuses on two aspects of the
LOFAR EoR Key science project: The first part describes the LOFAR EoR data model and the various components of the detailed LOFAR EoR simulation pipeline that is used to simulate realistic LOFAR EoR observations with all the sky components and instrumental corruptions added. The pipeline is generic enough to accommodate any effect and any telescope configuration. The second part, describes the LOFAR EoR KSP reprocessing step, which is a maximum likelihood inversion of the visibilities after calibration. This will give the final data product of the EoR KSP that will be used to remove the foregrounds and extract the cosmic signal. In the following we summarize the basic conclusions of this thesis and discuss the future prospects.

6.1 Prospects for the EoR Experiment

The success of the LOFAR EoR experiment relies on the detailed understanding of the astrophysical, terrestrial and instrumental distortions of the observed signal. Calibration is the name of the game. Without accurate calibration LOFAR, and for that matter, any of the new generation radio arrays will not be able to achieve better sensitivity than, for example, the Very Large Array at 74 MHz. The foreground emission is several orders of magnitude stronger than the signal and even errors of the order of one per cent or less on the foreground removal (due to calibration, imaging or extraction algorithm) can be comparable with the desired signal strength. In order to understand the observed visibilities and the effects of errors on the signal recovery, the LOFAR EoR team has developed an end-to-end pipeline that creates realistic sky maps with all the sky components included, and then distorts them according to our current understanding of the systematics of LOFAR. Then, the data are processed almost blindly in a manner similar to LOFAR EoR calibration pipeline and finally the signal is extracted from the corrected data.

The pipeline has three main modules: (i) Sky realizations: The cosmic signal is generated with the algorithm described in the thesis work of Rajat Thomas (2009). Diffuse Galactic and discrete extragalactic foregrounds are added using the results of the thesis work of Vibor Jelic (2010). This involves full-polarization realizations of the foreground contaminants. (ii) Systematics: The complex LOFAR instrumental response is simulated (Chapter 2 and 3). This includes ionospheric distortions (thesis) and RFI (thesis work of A. Offringa, in prep.) (iii) Processing: Finally, the data are calibrated using the standard LOFAR-EoR calibration pipeline. The calibration solutions are used as an input to the maximum likelihood inversion (Chapter 4). The output of this step is used to extract the signal (Jelić et al., 2008; Harker et al., 2009a,c). A flow chart is shown in Figure 6.1.

In Chapter 2 we introduced the physics-based data model for the LOFAR EoR Key Science Project. We provided a brief description of the physical connections between the Hamaker–Bregman–Sault formalism (Hamaker et al., 1996) and instrumental/propagation parameters. We use the above physical model to generate mock observations in Chapter 3. In that chapter we also introduced a generic statistical ARMA model for the complex gains based on the theory of time-series. The model is generic enough to include the different contributions to the gain errors and contains the simple case of Gaussian random errors as a trivial case. Our beam modeling includes beam-forming using a delay-and-sum beamformer as well as polarization distortion due to the beam. To model the ionosphere we use a combination of wedges and 3D turbulent fluctuations in the TEC.
6.1 Prospects for the EoR Experiment

**Figure 6.1:** A flow chart of the LOFAR-EoR simulation pipeline, which will help to develop a robust signal extraction scheme for the extremely challenging EoR observations.

Using the above models, that are statistically as well as physically similar to true components of our data model, we are able to simulate realistic interferometric observations using LOFAR for the EoR experiment. This is the first detailed simulation and processing of LOFAR EoR data cubes, using the standard EoR calibration pipeline.

In Chapter 4 we describe the regularized maximum likelihood inversion step, that is the base-level approach that the LOFAR EoR KSP will use to produce sky maps. As an input we used the simulations generated in Chapter 3. The method includes all image plane and uv-plane effects and the likelihood maximization translates essentially into a linear algebra problem. Diffuse emission is treated naturally within this framework. The total noise which is the superposition of thermal, confusion and calibration noise was used as the effective noise.

In Chapter 5 we briefly discussed the computational issues associated with the simulation and inversion steps, as implemented on Graphics Processin Units.

### 6.1.1 Ability of LOFAR to detect the EoR

To determine the theoretical limits and statistical efficiency of the regularized maximum likelihood inversion, we used a Cramér-Rao bound analysis. We have shown (Chapter 4) that the estimator becomes asymptotically efficient in the case of map-making with the LOFAR core and that imaging would require more than ten times better sensitivity. This is within the grasp of the SKA. Based on our current simulations and understanding of the astrophysical and instrumental processes the LOFAR EoR experiment should be capable to detect the EoR signal, provided that the calibration can increase the dynamic range of the data with an error of less than 0.1 per cent.
6.1.2 The next generation calibration algorithms

Before the 1980s the standard paradigm in calibration was to rely on the stability of the instrument. In the following 20 years the SELFCAL algorithm was introduced that iterated between the calibration and imaging steps, but mostly uv-plane effects were considered. This is the second generation of calibration. From 2000 to the present, direction-dependent effects were considered in calibration software like AIPS++/CASA, MeqTrees, BBS etc. This is the third generation of calibration. However, the increasing complexity and sophistication of the new instruments and the demand for more demanding scientific goals requires astronomers to continuously devise new statistical algorithms to process the data. This is the fourth generation of calibration (Jan Noordam, private communication) and leaves room for research in the mathematical, computational and engineering aspects of this problem. The next generation of calibration algorithms will most likely involve new procedures (computational and statistical) to calibrate and analyze the data and is expected to becomes of major importance for the next generations of interferometers such as SKA.

6.2 Suggestions for future work

In this section we highlight several of the most important future research directions:

- **LOFAR EoR dry-run**: An immediate future goal is to generate a full, single window LOFAR EoR data set that corresponds to the full duration of LOFAR EoR observations. This will be used to study the signal extraction in the presence of day-to-day statistical variation in the noise and instrumental parameters.

- Calculate the global Fisher Information for the given data model using MCMC/nested sampling to assess the level of correlation and degeneracy between calibration parameters (including ionosphere, beam, instrument, etc).

- Study advanced regularization techniques in the deconvolution/imaging of the visibilities that take into account the data model and the observational specification such as spatial and spectral frequency coverage.

- We implicitly assumed the validity of the data model. However, the validity of the ME can be checked against the data in a Bayesian framework and attempts could be made to model the data with other (more simple) data models than those used the generate the data sets.

- Explore the accuracy and numerical complexity of other statistical estimators.

- **Simulating new instruments**: The simulation pipeline is constructed on a modular base. Simulating different instruments requires substituting each module with one relevant for the new instrument. The most important module is the primary beam. Different modules for Focal Plane Arrays (i.e. APERTIF), phased arrays (i.e. EMBRACE) can be easily incorporated.
Figure 6.2: In December 2009 a long observing run was carried out on the extragalactic radio source 3C61.1. For this 60 hour observation a total of 20 LOFAR HBA stations were used, consisting of 16 split core stations and 4 remote stations. For comparison, images of 3C61.1 from other radio telescopes are shown. The VLSS 74 MHz image was made with the Very Large Array (VLA), New Mexico (USA). The VLA 1.5 GHz image is from Leahy and Perley (1991, AJ, 102, 537). The WENSS survey was carried out with the Westerbork Synthesis Radio Telescope located not far away from the central core of LOFAR. (Courtesy of Reinout van Weeren, University of Leiden/ASTRON)

- Square Kilometre Array:
  Whereas in this thesis, we concentrated mostly on the case of the LOFAR EoR KSP, SKA is designed as an array that consists of phased arrays, single dishes and Focal Plane Arrays. The lessons learned from the calibration and data inversion of the LOFAR EoR KSP will be an invaluable asset for similar observations with the SKA and we can thus consider them also as progress made on the SKA calibration.

  Our simulation framework can be extended to handle SKA data. Regularized maximum-likelihood techniques can be used as a standard method for SKA core imaging. However, the increased size of the SKA data products might lead to new strategies that revolutionize astronomical data handling.

Hence, whereas in this thesis a first step was made to simulate, calibrate and invert fully realistic LOFAR EoR data sets, but we expect this work to be far from completed. A continuous learning-process is expected in the coming years, where observations will teach us to improve our data-models and the data-models will further enhance our understanding of the data. This interplay will be critical also in future, with even larger, more complex and ambitious observations of the EoR with the SKA. Exciting times lie ahead as we are experiencing a revolution (as contrasted to evolution) in radio astronomy.