Transient and variable radio sources in the LOFAR sky: an architecture for a detection framework
Scheers, L.H.A.

Citation for published version (APA):
Transient and Variable Radio Sources
in the LOFAR Sky

Bart Scheers

An Architecture for a Detection Framework
Transient and Variable Radio Sources in the LOFAR Sky

An Architecture for a Detection Framework

This thesis was typeset in \LaTeX{} by the author and printed in The Netherlands by Ipskamp. Online versions at http://dare.uva.nl/dissertations and http://www.astro.uva.nl/research/theses_phd

Author email: bartscheers@gmail.com

This research has been supported by grants of De Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), De Nederlandse Onderzoekschool Voor Astronomie (NOVA), Het Samenwerkingsverband Noord-Nederland (SNN) and Het Leids Kerkhoven-Bosscha Fonds (LKBF).
Transient and Variable Radio Sources in the LOFAR Sky

An Architecture for a Detection Framework
Promotiecommissie

Promotor: Prof. dr. R. A. M. J. Wijers

Overige leden: Prof. dr. P. J. Groot
               Prof. dr. H. F. Henrichs
               Prof. dr. M. L. Kersten
               Prof. dr. M. B. M. van der Klis
               Dr. B. W. Stappers
               Dr. M. de Vos
               Dr. M. Wise

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The research reported in this thesis was carried out at the Astronomical Institute “Anton Pannekoek”, at the Universiteit van Amsterdam, The Netherlands.
Aan mijn dierbare familie
Contents

1 Introduction .......................................................... 1
  1.1 Radio Astronomy in Historical Perspective .................. 1
  1.2 LOFAR: The Next-Generation Radio Telescope .............. 3
    1.2.1 The International LOFAR Telescope .................... 4
    1.2.2 The Transients Key Science Project ................... 5
  1.3 Expectations of LOFAR ........................................ 9
    1.3.1 Source Counts and Data Rates .......................... 9
    1.3.2 Finding Transients with LOFAR ......................... 10
  1.4 This Thesis .................................................... 10

2 Expected Data Rates and Volumes for the LOFAR Transients Key Project .................. 13
  2.1 Introduction .................................................... 13
  2.2 The Design of the LOFAR Telescope ......................... 14
  2.3 Expected Data Rates & Volumes for some TKP Observation Modes ........................................ 17
    2.3.1 The Expected Number of Sources in the LOFAR Frequency Bands ................................ 17
    2.3.2 Confusion Limited Images ................................ 18
    2.3.3 The Radio Sky Monitor .................................... 19
    2.3.4 The Full Dutch Array ..................................... 23
    2.3.5 Million Sources Sky Survey – Commensal Mode ........ 25
    2.3.6 Summary ................................................... 27
  2.4 TKP Databases ................................................... 28
    2.4.1 The TKP pipeline Database Schema ....................... 29
    2.4.2 MonetDB .................................................... 33
    2.4.3 Associating LOFAR Sources in the Database ............ 35
  2.5 Discussion ...................................................... 41
  2.A Source Counts & Data Rates of LOFAR Modes .............. 43
  2.B Benchmark Queries ............................................ 48
3 LOFAR’s Transients and Variability Detection Algorithms 51
   3.1 Introduction ........................................... 52
   3.2 Database and Internal Algorithms .................... 53
      3.2.1 Column Oriented Storage Model .................. 53
      3.2.2 Database Schema .................................. 54
      3.2.3 Source Association ................................. 55
      3.2.4 Transient and Variability Detection .............. 57
   3.3 Simulated Data ....................................... 58
      3.3.1 The Data ........................................... 58
      3.3.2 Source Association, $r$ and $\log LR$ .......... 59
      3.3.3 Variability ........................................ 61
   3.4 GRB 030329 Field ................................. 64
      3.4.1 The Data ........................................... 64
      3.4.2 Results ........................................... 67
   3.5 Discussion and Conclusions ........................... 78

4 Cross-Matching Multiple Radio Catalogues 81
   4.1 Introduction ........................................... 82
   4.2 Association Parameters ............................... 84
   4.3 Reliability of WENSS–NVSS Associations ............ 85
      4.3.1 Source Selection Criteria ......................... 85
      4.3.2 Parameter Evaluation .............................. 86
   4.4 Cross-Matching Bright VLSS Sources with WENSS and NVSS Sources .......... 91
   4.5 Discussion and Conclusions ........................... 95

5 A New Perspective on GCRT J1745–3009 97
   5.1 Introduction ........................................... 98
   5.2 Data Reduction ....................................... 99
      5.2.1 General ........................................... 99
      5.2.2 The 92 cm WSRT Observations on 2005 March 24 ... 99
      5.2.3 The 21 cm WSRT Observations on 2005 May 14/15 ... 101
      5.2.4 The 92 cm VLA Discovery Dataset of 2002 September 30 / October 1 .......... 102
   5.3 The Source on the Opposite Side of the Supernova Remnant .... 104
   5.4 Overview of Flux Measurements of GCRT J1745–3009 .... 105
   5.5 Reanalysis of the 2002 Discovery Dataset ............. 108
## Contents

5.5.1 Lightcurve .......................................................... 108  
5.5.2 Implications for Other Observations .......................... 112  
5.5.3 Position and Flux Measurements; Spectral Index Determination ........................................ 113  
5.5.4 Circular Polarisation ............................................... 114  
5.5.5 Maximum Source Size and Maximum Distance for Incoherent Emission .................................. 114  
5.6 Discussion ............................................................. 115  
5.7 Conclusions ............................................................ 118  

6 Low-Frequency Observations of the Radio Nebula Produced by the Giant Flare from SGR 1806–20: Polarimetry and Total Intensity Measurements 121  
6.1 Introduction ............................................................ 122  
6.2 Observation and Data Reduction ................................. 122  
   6.2.1 General .......................................................... 122  
   6.2.2 Detailed Description of the Datasets ......................... 125  
6.3 Results ................................................................. 127  
   6.3.1 Total Intensity Measurements ................................. 127  
   6.3.2 Polarimetry ...................................................... 127  
6.4 Discussion ............................................................. 132  
   6.4.1 Total Intensity Measurements ................................. 132  
   6.4.2 Polarimetric Measurements ................................... 134  
6.5 Conclusions ............................................................ 135  

Samenvatting in het Nederlands 137  

List of Publications 145  

Dankwoord 147  

Bibliography 151
Chapter 1

Introduction

The excitement of new instruments is not the old questions they will answer but the new questions they will raise
Ron Ekers

1.1 Radio Astronomy in Historical Perspective

Karl Jansky opened the brand new field of radio astronomy in January 1932 (Kraus, 2005) by recognising that the 20.5 MHz signals received by his rotating antenna array were of extraterrestrial origin. At that time he was appointed at the Bell Telephone Laboratories and given the task to study the nature of the disturbances influencing transoceanic telephone traffic. Minimising antenna responses to certain directions would then increase the signal to noise and improve the connections. More observations made Jansky (1933) realise that the signals came from a fixed location in space, the centre of our own Galaxy. The Milky Way, for the first time, showed up in radio frequencies.

The astronomical community was ignorant of the relatively new radio technology (Verschuur, 2007) and only few others were aware of the significance of the discoveries. Jansky was given no support for a proposed construction of a parabolic antenna to continue his work, after which he was not involved in radio astronomy any more. Jansky’s work attracted Grote Reber, an electronics engineer, who was the first to build a parabolic-reflector antenna, a prototype of the classical dish-like radio telescopes, trying to understand more about the antenna–receiver system. Reber (1940) reported preliminary results of concentrated radiation at 160 MHz from the Galactic Plane and in 1944 he was also the first to carry out a systematic sky survey, producing the first maps of the radio sky (Reber, 1944).

During World War II, Reber’s work reached the Netherlands where it was picked up by Oort, who realised that in combination with Jansky’s dis-
coveries, the radiation was merely a continuum. Seeking for line radiation, Van de Hulst (1945) concluded that a small energy transition related to the electron-spin flip in the ground state of neutral hydrogen was observable at 21.1 cm, or 1420 MHz. A few years later, searches from several sites around the world confirmed the transition in emission (e.g. Ewen & Purcell, 1951; Muller & Oort, 1951; Christiansen & Hindman, 1952), as well as in absorption (Hagen & McClain, 1954), leading to detailed maps of our Galaxy.

Further breakthroughs were lying ahead. Observations at high Galactic latitude, in apparently empty fields, carried out with the horn-reflector antenna at 4080 MHz by Penzias & Wilson (1965) measured the intensities which led to the discovery of the 3 K cosmic microwave background (CMB) radiation, of which predictions were made by Gamow (1946), Alpher, Bethe & Gamow (1948) and Dicke et al. (1965) and pointed out the validity of the Big-Bang theory. Hewish et al. (1968) serendipitously discovered radio pulsars soon after, highly magnetised, rapidly rotating neutron stars. Their instrument, consisting of 2048 dipole antennas sensitive to 80 MHz frequencies, was designed to measure the angular size of radio sources by the rapid intensity fluctuations due to scintillation in the solar corona, but it became clear that it was ideal for detecting pulsars and other detections followed.

Technological development and new design methods were able to improve the observations. An array of multiple telescopes, an interferometer, can synthesise high-resolution images by making use of the aperture synthesis method developed by Ryle & Hewish (1960). Since resolution is inversely proportional to the aperture diameter, now being the diameter of the whole array, larger arrays will have higher resolution. When all telescopes track a source, signals arrive later at those telescopes further away in the array. Correcting for this delay and then adding the signals from all the telescope-pair combinations, will amplify the signal. Furthermore, as seen from a (radio) source point of view, Earth’s rotation varies the distance between telescope pairs (i.e. baselines) continuously, producing a series of measurements – the visibilities or \((u, v)\) data – as function of the hour angle. The visibility data, however, are not in a comfortable format to visualise the sky and make comparison at other wavelengths, and therefore it is converted into image data. In creating a proper image the visibilities need to be calibrated first by removing the instrumental and atmospheric effects from the measurements. Imaging is then done by Fourier transforming the sampled calibrated visibility function, which will give the relationship between the known primary beam pattern and the unknown sky brightness (Briggs et al., 1999).

Some well-known radio interferometers take advantage of the above-mentioned techniques; they were constructed some decades ago and are still operational and highly successful. Examples are the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands (since 1970), the Very Large Array (VLA) in New Mexico, USA (1980), the Giant Metre-wave Radio Tele-
LOFAR: The Next-Generation Radio Telescope

The LOw-Frequency ARray, or LOFAR, is scientifically as well as technologically a next-generation radio telescope. It is sensitive in a not yet well-explored domain of the electromagnetic spectrum, the low frequencies at 30–240 MHz. Relatively simple antennas (see Fig. 1.1) are connected to large computer facilities for further signal processing such as calibration and imaging. In combination with the unprecedented spatial and temporal resolution, sensitivity and large field of view, this makes LOFAR an excellent instrument that is very flexible to carry out observations of many different kinds. Below we will briefly describe the telescope and the radio sources and phenomena that will be studied by the Transients Key Science Project.

Figure 1.1: Left: The Low Band Antennas operate at 30–80 MHz (picture T. Krieg). In the computer cabinet, at the top left, all the antenna signals in the field are combined, before the data are transported to the central processing units in Groningen. Right: The High Band Antennas operate at 120–240 MHz.
1.2.1 The International LOFAR Telescope

The new radio telescope LOFAR is technically of fundamentally different design. Numerous, ordinary, dual-dipole antennas are mounted on the ground and grouped into stations. The antenna elements do not have any steerable parts like dish telescopes, but can be electronically directed by how their signals are combined to specific sky locations. Two different types of antenna design are operational, at frequencies below and above the radio broadcasting FM band, respectively. The former being the Low Band Antennas (LBAs) sensitive at 30–80 MHz and the latter the High Band Antennas (HBAs) performing at 120–240 MHz. Fig. 1.1 shows the two types of antennas.

The stations are distributed over large areas. A dense core of 24 stations is grouped into an area of 2–3 km across, of which 6 stations are located on the circular superterp, having a diameter of 340 m; this area is near the town of Exloo in the province of Drenthe at geographical coordinates 52° 54′ 32″ N, 6° 52′ 8″ E. Fig. 1.2 shows the layout of a core station and the superterp.

Remote stations are spread over the northern part of the Netherlands and have baselines up to \(\sim 100\) km, whereas the international European stations have baselines as large as \(\sim 1000\) km. Fig. 1.3 is a representation of the sites of funded and planned LOFAR stations throughout Europe. Currently funded international stations are located in Germany (Effelsberg, Garching, Jülich, Potsdam, Tautenburg), France (Nançay), England (Chilbolton) and Sweden (Onsala).

---

**Figure 1.2:** Left: Aerial photo of a LOFAR core station on the superterp near Exloo, the Netherlands. A core station consists of two HBA fields made up of 24 tiles, a tile being an assembly of 16 HB Antennas arranged on a 4 \(\times\) 4 grid, and one LBA field with 96 LB Antennas (see inset). Right: An aerial photo of the LOFAR superterp with diameter of 340 m and its 6 core stations. The superterp is located in the heart of the LOFAR core area, having a diameter of 2–3 km and containing 22 core stations. Photo credits: ASTRON.
Signals from the antenna fields are digitised and combined locally at a station level, before they are transported over a dedicated 10 Gb/s ethernet link to the central processing supercomputer in Groningen, that correlates the data coming from all the stations. Then, on the post-processing computing cluster, further operations run on the data streams (still at about 2 Gb/s) in the automated calibration and imaging software pipelines (Wijnholds, 2010). At this point the LOFAR Key Science Projects hook in their software pipelines for dedicated processing.

1.2.2 The Transients Key Science Project

The development of LOFAR is largely driven by the Key Science Projects (KSPs), that will exploit LOFAR to meet their science goals. The Transients Key Science Project aims to search for, detect, and study all the transient events in the LOFAR radio sky. Therefore, the Transients Key Project (TKP) will be using LOFAR intensively to survey the whole sky on a daily basis, monitor large areas, and observe single sources. Fast hemispherical surveys for detecting transients at the earliest stage and full-array modes for observing and tracking targeted sources in detail. All observed sources and
flux measurements will be stored in a light-curve database, gradually building up the largest radio catalogue. Internally it will be used for classifying sources and light curves and it will become available to the astronomical community as soon as possible. Sending and receiving triggers for follow-up strategies with multi-messenger instruments via the Virtual Observatory Event (VOEvent) network are set up for further scientific analysis. For transient sources, any follow-up with other telescopes or particle detectors need to be fast; therefore, LOFAR has been set up to send and receive alerts of new events via the so-called VOEvent network (e.g., Seaman et al., 2006; Swinbank, 2009).

One of the key observational modes of LOFAR for the TKP will be the Radio Sky Monitor (RSM; Fender et al., 2007). This mode will look at large patches of the radio sky for temporal changes in source brightnesses, by comparing current with previous observations, separated logarithmically in time. Multiple beams can be pointed simultaneously from the Core Stations to tile out a large fraction of the sky, while separate beams from the Remote Stations might be formed to observe simultaneously in targeted mode as is shown in Fig. 1.4. Using multiple (100 – 600) pointings, a Rapid All-Sky Monitor is able to survey the visible sky once a day, going as deep as 2 – 180 mJy with resolutions of 3 – 7 arcmin at the HB and LB frequencies, respectively. Another transient survey mode is the Zenith Monitor (ZM), when again a pattern of beams tiles out a large patch of the sky instantaneously. Using Earth’s rotation and keeping the antennas pointing at the zenith, stable beams survey a large strip of declination once a day, including parts of the Galactic Plane, and reaching sensitivities of 0.7 – 8 mJy and resolutions of 3 – 7 arcmin for HB and LB frequencies, respectively. New transients will be picked up quickly in these modes, while swift full-array follow-up observations will reveal more details about the evolution of a detected event.

Of particular interest to the TKP is the class of jet sources, which show radio emission due to synchrotron radiation produced by magnetised relativistic particle outflows interacting with the environment. Many types of objects are known to belong to this class, among which are the black hole and neutron star X-ray binaries (Fender, Belloni & Gallo, 2004), gamma-ray bursts (GRBs; e.g. van Paradijs, Kouveliotou & Wijers, 2000; Mészáros, 2006) and their radio afterglow phases detectable at LOFAR frequencies (van der Horst et al., 2008). Jets are manifested at the largest, galactic, scales as well, in Active Galactic Nuclei (AGNs) where they emanate from the central supermassive black hole. Observing different types of jet sources offers unique insights into the physics that plays a role in the evolution of these jets.

Other expected transient events to be caught by LOFAR are flares from soft gamma-ray repeaters (SGRs), a class of slowly rotating neutron stars.
with high magnetic fields (Gaensler et al., 2005a) and polarised radiation from the radio nebulae produced by these explosions (e.g., Taylor et al., 2005; Spreeuw, Scheers & Wijers, 2010). In addition, the TKP will use LOFAR in its full-array mode in order to carry out systematic deep pulsar surveys that are expected to find many new pulsars and to deliver the true pulsar population in the nearby Galaxy (van Leeuwen & Stappers, 2010). Radio emission from pulsars has a steep spectrum \( S_\nu \propto \nu^\alpha; \alpha \sim -2 \) in frequencies above the LOFAR High Band (see e.g., Burke & Graham-Smith, 2002), whereas a turnover at frequencies below 100 MHz might or might not be seen (e.g. Navarro et al., 1995), right in the LOFAR Bands. Pulsar models will be improved by the large sample of pulsars for which the measured spectral indices of the pulse components and continuum radiation and the turn-over break frequencies are important parameters. Extragalactic pulsar surveys are planned as well. Furthermore, LOFAR will visit some known extraso-
lar planets that are good candidates for predicted low-frequency detections (Grießmeier, Zarka & Spreeuw, 2007).

A primary goal of the Transients Key Project is to fill in the gaps at the relatively unexplored lower end of the radio spectrum for the sources detected by LOFAR. Dedicated wide-field monitors, commensally processing all LOFAR data taken, and targeted source observations with the full array will then provide us with extended light-curve information which will be permanently stored in a database that is accessible and queryable by the astronomical community. Fig. 1.5 displays spectra of some known classes of radio sources, showing that sources of synchrotron emission with or without turn-over frequencies due to synchrotron self-absorption are best observable in the LOFAR Bands.

The large number of classes of sources with hitherto covered behaviour in the low-frequency and time domains makes the quest for a new low-

![Figure 1.5](image)

**Figure 1.5:** Spectra of known classes of synchrotron sources and a thermal source (NGC 7027). Synchrotron radiation is produced by relativistic electrons spiraling in magnetised fields in the radio galaxy Cygnus A and the quasar 3C48, where synchrotron self-absorption in the latter source causes a turnover in its spectrum. NGC 7027 is a planetary nebula, ionised by the radiation of the central star nebula within our Galaxy. The unbound electron–ion (free–free) collisions produce the radio emission, opaque at the lower frequencies (making it invisible in the LOFAR Bands), following the Rayleigh–Jeans law and becoming transparent at the higher frequencies, giving rise to a flat spectrum. Figure adopted from Thompson, Moran & Swenson (2004).
frequency telescope clear. As a software telescope, LOFAR is capable of exploring the new parameter spaces at a wide range of frequencies, timescales, resolutions and sensitivities. The quick reconfiguration of the instrument in flexible modes will optimise observational strategies for detecting events at the earliest stages. Based on experiences with previous new telescopes (see Section 1.1) new types of sources or phenomena are very likely to be discovered, and exploration by LOFAR of the new spectral window on short as well as long timescales, the new parameter space, will almost certainly lead to serendipitous discoveries.

1.3 Expectations of LOFAR

1.3.1 Source Counts and Data Rates

The Transients Key Project inspects the LOFAR data streams of calibrated images for transient sources by means of an automated software pipeline that runs in a Python-based framework (Swinbank et al., 2007). Source fitting routines extract sources and related properties from the images (see Spreeuw, 2010), after which the results are stored into a database. It needs to be emphasised that only extracted data from the images will be stored in the TKP database and not the raw image data. Inside the database, the results are then compared with known, catalogued, source properties. Significant differences are reported back, and triggers for further specific follow-up actions may be sent. Using a database is necessary to have fast access to the large archive of source properties and light-curve data that the TKP gradually builds up.

To optimise the TKP pipeline, we need to think ahead about the specific observation modes, the number of distinct sources likely to be detected, the number of times a source will be observed, and at which pace the data input is streaming. These expected numbers determine the amount of disk space that has to be allocated and the data growth over time. Related to these are the accessibility of the database, containing millions of measurements of millions of sources. Also the response times of the transient detection algorithms should be quick enough to keep pace with the shortest integration times. It was one of the tasks of this thesis to investigate these restrictions and analyse detection algorithms and their applicability in the TKP software pipeline.

Most sources out of the Galactic Plane that LOFAR will detect, are of extragalactic origin and obey, on average, the flux-density–frequency scaling law: $S_\nu \propto \nu^{-0.7}$. Scaling the LOFAR sensitivities to existing source-count models at non-LOFAR frequencies (1.4 GHz, Huynh et al., 2005) will give us predictions on the numbers of sources likely to be detected in certain observation modes.


## 1. Introduction

### 1.3.1 Observation Mode

<table>
<thead>
<tr>
<th>Observation Mode</th>
<th>$N$</th>
<th>$m$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zenith Monitor</td>
<td>0.6</td>
<td>2400</td>
<td>62</td>
</tr>
<tr>
<td>Rapid All-Sky Monitor</td>
<td>1.1</td>
<td>350</td>
<td>9.1</td>
</tr>
<tr>
<td>Full-NL</td>
<td>9.2</td>
<td>4400</td>
<td>114</td>
</tr>
<tr>
<td>MSSS (commensal)</td>
<td>3.4</td>
<td>1800</td>
<td>47</td>
</tr>
</tbody>
</table>

Table 1.1: Expected numbers for some observation modes of LOFAR. We assume a 5$\sigma$ detection threshold in these modes and a 100% allocation time per mode. $N$ is the number of distinct sources likely to be detected, $m$ is the averaged number of measurements to be stored every second, where a (source) measurement is about 300 Bytes. $d$ is the growth of the TKP database disk size per day assuming a 100% time allocation. The RSM Zenith Monitor and Rapid All-Sky Monitor modes and the Full-NL mode will be frequently used by the TKP, whereas the Million Sources Sky Survey (MSSS) is an example of commensally processing LOFAR data.

Some observation modes of LOFAR that will be frequently used by the TKP were briefly mentioned in Section 1.2.2 and are listed in Table 1.1 with their expected number of sources, measurements per second, and data growth per day. From Table 1.1 it can be seen that the highest peaks of the data rate are to be expected from the Full-NL mode.

### 1.3.2 Finding Transients with LOFAR

Successful detection of transient events and variable sources depends on the monitoring strategies as well as on the software algorithms applied to the data streams and it was one of the tasks of this thesis to find out which database system and algorithms fit best to these requirements.

Positions of all the sources detected by LOFAR will be maintained in a database. Known sources from existing catalogues will be loaded as well for positional cross matching. Flux measurements and related properties (e.g. 1$\sigma$ errors, timestamp, frequency, resolution) will be added to the database every time an image has been processed. This means that flux measurements at a given location on the sky might have an association with a known source position in the database. A match will add the measurement to the light curve of the source, whereas no match at all needs further investigation before a reliable alert is sent out.

### 1.4 This Thesis

One of the main scientific goals of this thesis was related to low-frequency behaviour of variable sources in the LOFAR radio sky by reducing existing interferometry data of multiple observations of the decaying intensities from gamma-ray burst afterglows and soft gamma-ray repeaters. LOFAR will outscale the classical observing techniques by making many measure-
ments per source on timescales as short as 1 second. Data will eventually be processed, without manual intervention, in an automated software pipeline optimised to detect transient events and variable sources. This served as the framework for the other scientific software goals: design of a transient database schema that is capable of storing all the source measurements at the high rate at which LOFAR observes. Design and implement fast search and detection algorithms with response times comparable to the image input rate. Chapters 2, 3 & 4 are related to the software development of the Transients Key Project and Chapters 5 & 6 to types of variable sources that are detectable at LOFAR frequencies.

Chapter 2 describes the main observation modes that are relevant to the TKP. Expected numbers for source counts are worked out down to the one-second timescale and up to the envisaged maximum integration times per mode. The TKP will try to commensally process all LOFAR data and for this the Million Sources Sky Survey (MSSS) is discussed as an example. The schema of the database, the main tables and algorithms are elaborated. Benchmarking the speed of the main algorithms concerning source association and transient and variability detection showed a clear preference for the column-store oriented database MonetDB to be implemented in the TKP pipeline.

The source association and transient and variability detection algorithms are described in Chapter 3. Three parameters – distance, weighted positional difference and likelihood ratio – are determined for every pair of sources that is considered as a source association candidate. A reliable and fast cut-off association parameter of two genuinely associated sources is the weighted positional difference, which has a Rayleigh distribution. Furthermore, two variability indices define the absolute flux change and the weighted flux change, i.e., the significance of any variability. This significance gives a quantitative measure for the likelihood of having seen a truly variable event, and thus can be the basis for deciding whether to initiate any follow-up of the source (along with measured source characteristics).

We applied the algorithms to simulated and real data. For the simulated data we used a sample of 1000 VLA 325 MHz noise maps and inserted 64 fixed source positions in each map, meaning that a source position is Rayleigh distributed due to the local noise. The 1000 images were processed to assess the critical values for genuine source associations. These were applied to WSRT observations of the field of GRB030329 made during the years 2003 and 2007 at several frequencies, resolutions, and sensitivities to search for transient and/or variable sources. We found two variable sources.

Chapter 4 describes the usage of multiple major radio catalogues in the TKP database. These catalogues were produced by surveys that were carried out at different frequencies and resolutions. LOFAR will create a catalogue
1. Introduction

as well. The cross correlation of catalogues is based on the same principles as the association of sources as described in Chapters 2 & 3. We derive source lists that include fitted spectral indices, which are available to the calibration processes of LOFAR and the TKP transient detection pipeline. We cross-identified the sources in the VLSS (74 MHz), WENSS (325 MHz) and NVSS (1400 MHz) catalogues and present a high-quality source list of nearly 2800 bright sources (> 1 Jy) at declinations between 44° and 64°, that have unique counterparts in all three catalogues.

Chapter 5 reports on the Galactic Centre Radio Transient source, GCRT J1745–3009. It was discovered at 325 MHz by Hyman et al. (2005), showing five Jy-level bursts of 10 min duration and with 77 min recurrence time. Follow-up observation campaigns showed only two redetections of faint bursts (Hyman et al., 2006, 2007). Reanalysis of the VLA discovery data set improved the burst recurrence times and spectral indices, making clear that such bursts are detectable by LOFAR.

Results of early radio follow-up observations of the giant flare produced by the soft gamma-ray repeater SGR 1806-20 are described in Chapter 6. Low-frequency polarimetry and total intensity measurements of the radio nebula revealed the absence of depolarisation at these frequencies and probably probed different substructures in the nebula as compared to higher frequencies. Again, the results show that LOFAR in full array mode would have been able to detect this burst and follow the evolution of the radio nebula and the magnetic fields.
Chapter 2

Expected Data Rates and Volumes for the LOFAR Transients Key Project

2.1 Introduction

The International LOFAR Telescope is in many aspects a next-generation radio telescope. LOFAR Stations, consisting of dual-dipole antennas, are spread over the northern part of the Netherlands and into the European countries of Germany, France and the United Kingdom. The distribution of the LOFAR Stations allows for scalable baselines. Baselines of up to 3 km are from the core stations alone, up to 100 km when the remote (i.e. Dutch) stations join, and up to 1000 km when the European Stations are included. The large collecting area and high resolution, the large fields of view of the dipole antennas, and the high sensitivities due to the numerous stations and large frequency bandwidth, give unprecedented observation capabilities in the yet unexplored low-frequency domain of the electromagnetic spectrum below 240 MHz. Beam-formed data of a LOFAR station are transported over a 10 Gb/s ethernet link to an IBM BlueGene/P central processing supercomputer that correlates the signals from all the stations, producing raw data at 1 s time resolution and 1 kHz frequency resolution. Calibration and imaging pipelines run on a dedicated post-processing cluster, aiming at producing calibrated snapshots of large patches of the sky every second. At this point the Transients Key Science Project plugs in its automated software pipeline in the continuous data stream of calibrated images in order to detect transient and variable radio sources on various time scales for scientific analysis.

The number of sources likely to be detected and disk size to be allocated
for a given LOFAR observation mode, depend on the sensitivity, observing frequency and field of view of the observations, which in turn depend on the key configuration parameters of LOFAR: the frequency, bandwidth, resolution, integration time, number of beams, number of stations.

To enable searches for transients and wide-field or even full-sky monitoring modes, i.e., to see changes from one observation to another, we want to record all measurements of the observable sources. Comparison with known sources from the major catalogues at multiple frequencies, and in parallel with the running, gradually growing and updated LOFAR catalogue, enables us to classify sources based on their light-curve data at an early stage. The image size and input rate – the number of sources to be processed per unit time – then determine whether modes can reach the desired shortest time scales.

This Chapter focuses on the number of sources and measurements extracted from those images per second, and not on the number of floating point operations per second nor the raw image sizes that are produced by the calibration and imaging pipelines. Nor will we discuss the disk spaces needed for storing the (raw) image data. Instead, we will concentrate on the estimated data growth and disk space needed to build up the large catalogue database containing all LOFAR sources.

This Chapter is organised as follows. Section 2.2 briefly describes the design of the LOFAR Telescope and some of its characteristics. In Section 2.3 the number of sources likely to be detected is estimated for some frequently used observation modes of LOFAR by the Transients Key Project. The details down to the shortest time scales for the modes are reported in Appendix 2.A. Based on the number of measurements made per source, the size and growth of disk space to be allocated is determined. Section 2.4 describes the database schema and the design and performance of the implemented algorithms to enable fast transient and variability detection of which the main benchmark queries are shown in Appendix 2.B. Section 2.5 discusses the results.

2.2 The Design of the LOFAR Telescope

Field of view, resolution and sensitivity are the main players in the on-going race of building more powerful telescopes. While the very expensive classical radio telescopes with steerable dishes have reached their construction limits, cheaper (but not less challenging) alternatives have come into view. During the 1990s, the concept of a software telescope developed and was getting more concrete by the end of that decade (Miley, 2010). LOFAR’s frequency range is divided into two bands bracketing the FM radio broadcasting band. Each band has its own antenna type, the Low
2.2 The Design of the LOFAR Telescope

<table>
<thead>
<tr>
<th>LOFAR Station</th>
<th>( L ) [km]</th>
<th>( N_{\text{station}} )</th>
<th>( N_{\text{LBA}} )</th>
<th>( N_{\text{HBA Tiles}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>3</td>
<td>24</td>
<td>96 (48)</td>
<td>2 \times 24</td>
</tr>
<tr>
<td>Remote</td>
<td>100</td>
<td>16</td>
<td>96 (48)</td>
<td>1 \times 48</td>
</tr>
<tr>
<td>European</td>
<td>1000</td>
<td>8</td>
<td>96</td>
<td>1 \times 96</td>
</tr>
</tbody>
</table>

Table 2.1: Three types of LOFAR stations exist. Core and remote stations are located in the Netherlands, whereas the European Stations are in Germany, France, Sweden and England. For each type the number of stations, \( N_{\text{station}} \), the maximum baseline, \( L \), the number of Low Band Antennas, \( N_{\text{LBA}} \) and the number of High Band Tiles, \( N_{\text{HBA Tiles}} \), is given, where a single HBA Tile contains 16 High Band Antennas, arranged on a \( 4 \times 4 \) grid. The number in parentheses in the column of LBAs represents the maximum number of available LBA antennas for dual polarisation observations.

Band Antennas (LBA) being sensitive from 30 to 80 MHz, and the High Band Antennas (HBA) in the range 120–240 MHz (see Fig. 1.1 in Chapter 1). LBAs and HBAs are most sensitive at 60 and 150 MHz, respectively (e.g. de Vos et al., 2009; Nijboer & Pandey-Pommier, 2009).

Dutch LOFAR stations are located either in the core at Exloo, where multiple stations are densely packed in an area of 3 km, and more sparsely remote, in the north-eastern Dutch provinces, with baselines up to about 100 km. European or international stations are sited in Germany, France, Sweden and the United Kingdom, giving baselines of 1000 km. The number of available LBA antennas and HBA tiles differ depending on the location of the station. All Dutch stations have 96 Low Band Antennas, but when observing in dual polarisation mode only 48 may be selected, whereas in single polarisation mode all 96 may be used. The EU Stations also have 96 Low Band Antennas, but do not have such a restriction. Furthermore, the core stations have two fields (also called "ears") of 24 HBA tiles, whereas a remote station has a single field of 48 tiles, and an EU Station a single field of 96 tiles. Table 2.1 gives the current status of the number of funded and planned LOFAR stations.

The LOFAR Telescope is configurable in many ways (we refer to de Vos et al. (2009) and Nijboer & Pandey-Pommier (2009) for more details), but here we will highlight the most characteristic properties and observation modes relevant to the Transients Key Project. Multiple beams may be constructed with the restriction that the total bandwidth does not exceed the maximum value of 48 MHz. Although LOFAR is able to observe at any frequency in the Low as well as in the High Band, simultaneously observing in both Bands is not possible. However, we will focus here on five frequencies per Band, evenly distributed across the Low and High Band, in accord with the findings of Law & Hessels (2009) of Standard Bands for the Radio Sky Monitor Mode.

Five different LBA antenna configurations are available of which we will
focus on the Inner configuration. This mode uses the innermost 48 antennas, thereby reducing the station size, and increasing the full width half maximum (FWHM) of the station beam

\[ \theta_{\text{FWHM}} = k_1 \frac{\lambda}{D}, \]  

where \( k_1 \) is of order unity, \( \lambda \) the observing wavelength and \( D \) the station diameter, or the distance between the two outermost antennas. And this consequently affects the station field of view (FoV) defined as

\[ \Omega_{\text{FoV}} = \pi \left( \frac{1}{2} \theta_{\text{FWHM}} \right)^2. \]  

The resolution of the LOFAR mode is then determined by

\[ \theta_{\text{res}} = k_2 \frac{\lambda}{L}, \]  

where \( k_2 \) is of order unity, \( \lambda \) the observing wavelength and \( L \) the distance between the two outermost stations. The sensitivity is then defined by

\[ \Delta S = W \left[ 2(2\Delta \nu \tau) \left( \frac{N_C(N_C - 1)/2}{S_C^2} + \frac{N_C N_R}{S_C S_R} + \frac{N_R(N_R - 1)/2}{S_R^2} \right) \right]^{-1/2}, \]  

where \( \Delta \nu \) is the spectral bandwidth (in Hz), \( \tau \) is the integration time (in seconds), \( W \) depends on the imaging weighting scheme used and is unity for core stations only and 1.3 for inclusion of remote stations, \( N_C \) and \( N_R \) are the number of core and remote stations, respectively, and \( S_C \) and \( S_R \) are the measured system equivalent flux densities (SEFDs in Jy; e.g., de Vos et al., 2009; Nijboer & Pandey-Pommier, 2009) for core and remote stations, respectively, which is an indication of the sensitivity of the antenna and receiving system (e.g., Thompson, Moran & Swenson, 2004; Wrobel & Walker, 1999). We adopted the effective bandwidth factor, \( \eta = 0.89 \), based on results from LOFAR test stations (Nijboer & Pandey-Pommier, 2009). If identical core stations are being used, as is the case in some observation modes, the sensitivity is reduced to \( (W = 1) \)

\[ \Delta S = \frac{S_C}{\sqrt{2\Delta \nu \tau N_C(N_C - 1)}}. \]  

From Eqs. 2.4 and 2.5 it can be seen that the sensitivity is inversely proportional to the square root of the bandwidth and the integration time.

Selecting a set of integration times that increases logarithmically enables detection of transient and variable sources on different time scales and flux densities. The Transients Key Project will use thirteen time scales: 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, and 10000 seconds for this. Due to information on the broad frequency range the light-curve catalogue database contains the temporal as well as the spectral characteristics of all sources detected by LOFAR.


2.3 Expected Data Rates & Volumes for some TKP Observation Modes

The Transients Key Project designed and developed an automated software pipeline for detecting transient and variable sources in the input data streams on time scales as short as 1 second. It is clear that the millions of sources and their repeatedly measured properties on different time scales, cannot be simply stored in tuples or arrays of even files. Whether open source databases are capable of processing these large data volumes on those short time scales has to be investigated. Furthermore, we have to find out if we will gain significant processing time by shifting some of the algorithms into the database engine. We want to know whether the database is accessible and scalable to the expected rates and sizes with which the LOFAR catalogue will grow per year.

Estimates of the number of sources per second likely to be detected in typical LOFAR observation modes will serve as starting points in the choice of the database system, and the design of the database scheme. Benchmarking the most frequently used set of queries will eventually help in managing to process the large amounts of data entering the TKP database.

Miller-Jones (2008) calculated estimates on the number of source counts for the full array and the core configurations. We follow his calculations and investigate some LOFAR- and TKP-specific observing modes.

As will be shown in Section 2.4.1, the upper limit of disk space to be allocated for a source measurement is about 300 Bytes. A single measurement includes source properties like position, all Stokes parameters, timestamps, effective frequency, integration time, beam size, some auxiliary parameters for fast positional look-ups, and Gaussian fit parameters, plus all 1σ errors. All measurements of all sources are stored in an accessible and queryable database.

2.3.1 The Expected Number of Sources in the LOFAR Frequency Bands

Huynh et al. (2005) derived source counts from the 1.4 GHz Australia Telescope Hubble Deep Field–South (ATHDFS) Survey and compared these to other surveys carried out at the same frequency. They found a sixth-order polynomial function, describing the source counts down to the level of 50 μJy

$$\log \left( \frac{dN/dS}{S^{-2.5}} \right) = \sum_{i=0}^{6} a_i \left[ \log \left( \frac{S}{\text{mJy}} \right) \right]^i,$$

(2.6)

where $a_0 = 0.841$, $a_1 = 0.540$, $a_2 = 0.364$, $a_3 = -0.063$, $a_4 = -0.107$, $a_5 = 0.052$ and $a_6 = -0.007$. The differential number of source counts ($dN/dS$) is multiplied by $S^{5/2}$, as to have the ratio of observed numbers
to expected numbers in a simple Euclidean Universe, where the differential source counts should be constant.

Integration of this function will then predict the number of sources likely to be detected at 1.4 GHz, above a lower specified flux limit. However, if we assume that most of the detected sources, out of the Galactic Plane, are of extragalactic origin and obey a spectral flux scaling law described by \( S_\nu \propto \nu^{-0.7} \), we can derive the expected source counts in the LOFAR frequency bands by scaling the predicted LOFAR rms sensitivities up to 1.4 GHz. Nijboer & Pandey-Pommier (2009) describe the general astronomical capabilities of the LOFAR Telescope. From these, and Eqs. 2.1–2.5, sensitivities at detection thresholds of five and thirty times the rms noise level can be computed for particular configurations of LOFAR. This is then used as the lower limit in the integral of the predicted source counts,

\[
N = \int_{(5|30)\sigma_{\text{rms}}}^{S_{\text{up}}} S^{-2.5} 10^{\sum_{i=0}^{6} a_i [\log(S/\text{mJy})]^i} dS, \tag{2.7}
\]

where the upper limit is set high enough to not significantly contribute to the integral anymore; we set \( S_{\text{up}} = 2 \text{ Jy} \).

Extrapolation of the LOFAR sensitivities in the High Band at the longer integration times to the frequency of Huynh’s model are below the 50 \( \mu \text{Jy} \) limit and may introduce uncertainties into the source count numbers. Only the full array mode at the longest logarithmic integration times at 120 and 150 MHz will be affected by this.

### 2.3.2 Confusion Limited Images

At the longer integration times, especially at the lower frequencies in core modes, the LOFAR images will be crowded with sources and the restoring beam can no longer distinguish between sources, which will affect the astrometry and photometry. Confusion becomes a problem when the source densities in the synthesised beam are larger than 1/50 to 1/15 (Hogg, 2001) depending on background noise. A standard rule of thumb is to use 1/30 as the confusion limit. Theoretically, differencing of properly calibrated images can reach the thermal noise levels. In our estimates we therefore use the criterion that an image is *classically* confusion limited, when the source counts times the ratio of the beam area and the field of view is larger than 1/30

\[
\langle N \rangle \frac{\pi (\theta_{\text{res}}/2)^2}{\text{FoV}} > \frac{1}{30}, \tag{2.8}
\]

where \( \langle N \rangle \), \( \theta_{\text{res}} \), and FoV are given for several configuration modes in the next Section.
2.3 Expected Data Rates & Volumes for some TKP Observation Modes

2.3.3 The Radio Sky Monitor

The Radio Sky Monitor (RSM) enables key observational modes of the
LOFAR Telescope for the Transients Key Project. Initial strategies are
described by Fender et al. (2007). The RSM has three observation modes,
of which we will discuss the Zenith Monitor and the Rapid All-Sky Monitor,
and not the Galactic Plane scans. Both modes exploit LOFAR’s capabilities
of simultaneously observing large patches of the sky with multiple beams.
With LOFAR’s large collecting area these survey modes are fast. The goal
of these modes is to detect transients and variable sources at the different
time scales and flux densities.

The Zenith Monitor Mode

The Zenith Monitor mode aims to stare at the zenith and map out the
entire field of view that passes by. By using a hexagonal pattern of 7 beams
formed by 24 core stations (LBA in inner configuration), an instantaneous
field of view of 475 deg$^2$ at 60 MHz and 82.8 deg$^2$ at 150 MHz is constructed.
This hexagonal pattern will scan a declination strip of about 20° wide, cen-
tered at $\delta = +54^\circ$, with a total area of 4211 deg$^2$ or about 10% of the entire
sky. The number of pointings needed to scan the whole strip is about 12
at 60 MHz and 60 at 150 MHz. This means that an integration time for
each field of 2 h in the Low Band and 30 min in the High Band is allowed
to carry it out within a day. Sensitivities for single 4 MHz beams at these
integration times reach the milliJansky level. Table 2.2 gives an overview
of the characteristics of the Zenith Monitor (ZM) and the expected number
of unique sources likely to be detected, data rates as measurements per second
and the storage capacities needed.

<table>
<thead>
<tr>
<th>Freq. [MHz]</th>
<th>$\theta_{res}$ [arcsec]</th>
<th>FoV [deg$^2$]</th>
<th>$\Delta S$ [mJy]</th>
<th>$N$ [x10$^6$]</th>
<th>$m$ [s$^{-1}$]</th>
<th>$d$ [GB/day]</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>413</td>
<td>475</td>
<td>6.0–510</td>
<td>0.16</td>
<td>1785</td>
<td>46</td>
</tr>
<tr>
<td>150</td>
<td>165</td>
<td>82.8</td>
<td>0.5–22</td>
<td>0.60</td>
<td>2996</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 2.2: Characteristics of the RSM Zenith Monitor. The instantaneous field
of view (FoV) of the hexagonal pattern of seven 4 MHz beams is given for the two
observing frequencies. Note that it is not possible to observe at both frequencies
simultaneously. The sensitivities listed range from the longest (7.2 ks at 60 MHz
and 1.8 ks at 150 MHz) to the shortest (1 s) integration times. At the 5$\sigma$ level, $N$
is the number of distinct sources likely to be detected in this mode and $m$ is the
number of (source) measurements to be stored every second. $d$ is the growth of
disk size per day assuming a time allocation of 100%. 
A more detailed description of the expected numbers specified per logarithmic integration time, the way the TKP will search for transients, is given in Table 2.7 in Appendix 2.A. Table 2.7 gives the number of sources likely to be detected, \( \langle N \rangle \), for the given integration time \( \tau_{\text{int}} \) at 5 and 30 times the rms noise, assuming a single 4 MHz beam for the observing frequency. The last column of Table 2.7, \( m \), gives the total number of (source) measurements per second to be stored in the database. A grand total, for a single ZM 4 MHz beam, is given in the last row per frequency, which is simply the sum of the measurements made over all the integration times.

Fig. 2.1 shows the number of measurements per second, \( m \), to be stored at every integration time, \( \tau_{\text{int}} \) for the ZM hexagonal pattern of seven beams. As noticed in Section 2.3.2, from Eq. 2.8 we deduce that the Low Band images are classically confused when \( \langle N \rangle > 338 \), which is at \( \tau_{\text{int}} > 10 \text{ s} \) and \( \tau_{\text{int}} > 500 \text{ s} \) for the 5\( \sigma \) and 30\( \sigma \) detection levels, respectively. In the High Band, the classical confusion limits arise when the integration times exceed 5 and 200 seconds for the two detection levels.

Tables 2.2 and 2.7 show that the RSM Zenith Monitor in LBA mode collects about 1800 measurements per second at the 5\( \sigma \) level, corresponding to about 150 \( \times 10^6 \) measurements per day to be stored. For a single measurement a maximum storage size of 300 Bytes is anticipated (see Section 2.4.1), thus storing about 46 GB/day of 5\( \sigma \) measurements at 60 MHz. For the HBA frequencies, at the 5\( \sigma \) level, the numbers are nearly twice as large. At the

---

**Figure 2.1:** Expected number of measurements per second \( (m) \) to be stored in the RSM Zenith Monitor as related to the integration time \( (\tau_{\text{int}}) \) at the two observational frequencies. For the seven-beam pattern, the thick lines represent the 5\( \sigma \) detection level measurements, whereas the dashed lines represent the 30\( \sigma \) level. A measurement of a source is about 300 Bytes of data.
30σ level this is about \(13 \times 10^6\) measurements per day or 4.0 GB/day in the Low Band and 16 GB/day in the High Band. As the LB and HB cannot operate at the same time the average growth including both frequencies is then about 62 GB/day at the 5σ and 10 GB/day at the 30σ level.

Assuming the noise in an image to be Gaussian, the probability of having a 5σ pixel is \(10^{-6}\). The number of pixels for the full hexagonal pattern per day is about \(3 \times 10^5\) pixels per FoV per second (see Table 2.2) which is approximately \(3 \times 10^{10}\) pixels per day, meaning that there will be \(3 \times 10^4\) false noise spikes catalogued as a measurement. This corresponds to a fraction of about \(3 \times 10^{-4}\) of the total number of 5σ measurements per day. At the 30σ threshold the numbers of false positives can be neglected. More tests with confusion limited images and False Discovery Ratio (FDR) algorithm (Spreeuw, 2010) need to be carried out to choose a proper threshold as starting point for these observations.

Important for the database load, is the number of (source) measurements entering the database every second, \(m \approx 1100\) at 150 MHz and \(\tau_{\text{int}} = 1\) s, or \(m_{\text{max}} \approx 3000\) at 150 MHz, that needs to be processed within 1 second. It is however not specified yet how the pattern of seven beams will be processed in the calibration and imaging pipelines. In Section 2.4.1 and further we use these numbers to benchmark the processing and detection algorithms.

### The Rapid All-Sky Monitoring Mode

The Rapid All-Sky Monitor is a survey mode of LOFAR to search for the rare second-timescale transient events in the whole sky that is visible by LOFAR. Fender et al. (2007) describe the initial strategy for this survey. To have the largest field of view only core stations will be used with the LBAs in the inner configuration mode (Nijboer & Pandey-Pommier, 2009) and the core stations HBAs. In the calculations here, we will adopt a total of 24 core stations, and a bandwidth of 4 MHz per beam at the most sensitive observing frequencies of 60 and 150 MHz. According to Fender et al. (2007), the number of pointings needed to tile out the hemisphere is achieved when the beam pointings are offset by \(\theta_{\text{FWHM}}/\sqrt{2}\), which corresponds to about 100 pointings at 60 MHz and 600 pointings at 150 MHz. The corresponding times to track a field are then 14 minutes and 140 seconds for the LB and HB, respectively. By spacing the integration times logarithmically a set of images is created in order to detect transients at different time scales. Table 2.3 gives an overview of the expected source counts, measurements per second to be stored, and the data growth per day for the RASM mode.

Table 2.8 in Appendix 2.A gives an overview of the configuration parameters and the expected source counts and measurements for the different integration times. From Eq. 2.8 and the numbers in Table 2.8 it can be seen that the classical confusion limit is reached at integration times longer than 21
2. Expected Data Rates and Volumes for the LOFAR Transients Key Project

<table>
<thead>
<tr>
<th>Freq. [MHz]</th>
<th>$\theta_{\text{res}}$ [arcsec]</th>
<th>FoV [deg$^2$]</th>
<th>$\Delta S$ [mJy]</th>
<th>$N \times 10^6$ [s$^{-1}$]</th>
<th>$m$</th>
<th>$d$ [GB/day]</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>413</td>
<td>105</td>
<td>17.6–510</td>
<td>0.36</td>
<td>255</td>
<td>6.6</td>
</tr>
<tr>
<td>150</td>
<td>165</td>
<td>18.4</td>
<td>1.9–22</td>
<td>1.1</td>
<td>419</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 2.3: Characteristics of the RSM Rapid All-Sky Monitor. The instantaneous field of view of a single 4 MHz beam is given for the observing frequencies. The sensitivities listed range from the longest to the shortest integration times. $N$ is the number of distinct sources likely to be detected and $m$ is the number of measurements per second to be stored. $d$ is the growth of disk size per day assuming a time allocation of 100%.

Figure 2.2: Expected measurements per second ($m$) to be stored in the RSM Rapid All-Sky Monitor as related to the integration time ($\tau_{\text{int}}$) at the two observing frequencies. For a single 4 MHz beam (see text), the thick lines represent the $5\sigma$ detection level measurements, whereas the dashed lines represent the $30\sigma$ level. A measurement of a source is about 300 Bytes of data.

10 and 500 seconds in the Low Band for the $5\sigma$ and $30\sigma$ detection levels, respectively, whereas 5 second integration time is just below the limit in the High Band for the $5\sigma$ measurements. At $30\sigma$ no confusion limits are reached. Fig. 2.2 shows the measurements per second, $m$, made at the consecutive integration times for the Rapid All-Sky Monitor mode.

As can be seen from Tables 2.3 and 2.8, the average input rate per day is about 250 and 400 measurements per second at the $5\sigma$ level at the Low Band and High Band frequencies, respectively. The larger ratio at the $30\sigma$ level is caused by the relative higher sensitivity in the High Band at the shorter integration times, where more sources contribute in the sum. A 24 h RASM run collects about $22 \times 10^6$ measurements in the Low and $36 \times 10^6$
in the High Band at the $5\sigma$ level, resulting in storing 6.6 and 11 GB/day, respectively. At $30\sigma$ the storage capacity needed is 500 MB/day for the LB data and 2.3 GB/day for the HB data, assuming a 100% time allocation in a single Band. Again, the Low and High Bands cannot observe simultaneously. The dominating short time scales determine the rate at which the source measurements enter the database, where at the shortest time scale at $5\sigma$, the storage rate is about 160 source measurements per second, with an average of about 350 source measurements per second, assuming that the time allocation is evenly distributed between Bands.

### 2.3.4 The Full Dutch Array

LOFAR’s flexibility admits several configurational set-ups, all leading to slightly different resolutions, sensitivities and fields of view and thus source counts and data rates. We cannot treat them all, but here we will elaborate the full Dutch array mode (Full-NL), where all 24 core and 16 remote stations are used. The longest baseline is about 100 km and multiple beams from the same Band might observe at several frequencies at the same time. The only restrictions are that the LBA and HBA cannot operate at the same time and the total bandwidth of the beams sums to 48 MHz. This 40-station interferometer with its long baselines and large fields of view, achieves unprecedented resolutions and sensitivities at a wide range of time scales. In the calculations here we assume a 4 MHz beam for both Bands (LBA in inner configuration) and deduce the sensitivities (Eq. 2.4) with the logarithmically spaced integration times ranging from 1 to 10,000 seconds. Longer integration times are not treated here, since we focus on the transient detection strategy. Table 2.4 summarises for some selected frequencies the expected numbers, where we take into account using the maximum allowable bandwidth of 48 MHz, meaning we can construct twelve similar beams.

We refer to Tables 2.9 and 2.10 in Appendix 2.A for a more detailed overview of the Low and High Band frequencies of the Full-NL mode, respectively. Fig. 2.3 shows the measurements per second for the range of integration times for the Full-NL mode at the selected frequencies for a single (left) and twelve (right) 4 MHz beams.

From Tables 2.4, 2.9 and 2.10 it can be seen that at the shortest time scales in the lower end of both Bands the source counts are the highest, whereas the Low Band source counts outnumber the ones of the High Band, due to the applied spectral scaling law, $S_\nu \propto \nu^{-0.7}$, the field of view and the high sensitivity. The images will, however, not be confusion limited.

Similarly as in the previous sections, the summed data rates, $m$, for the whole range of integration times in twelve 4 MHz beams at 30 MHz at the $5\sigma$ level give an average of about 9000 source measurements per second that will be detected and have to be stored (see Table 2.4). Consequently, the disk space needed after a full day observing in the Full-NL mode at 30 MHz...
2. Expected Data Rates and Volumes for the LOFAR Transients Key Project

<table>
<thead>
<tr>
<th>Freq. [MHz]</th>
<th>$\theta_{res}$ [arcsec]</th>
<th>FoV [deg$^2$]</th>
<th>$\Delta S$ [mJy]</th>
<th>$N$ [$\times 10^6$]</th>
<th>$m$ [s$^{-1}$]</th>
<th>$d$ [GB/day]</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>20.6</td>
<td>419</td>
<td>11–1100</td>
<td>0.71</td>
<td>0.19</td>
<td>9000</td>
</tr>
<tr>
<td>60</td>
<td>10.3</td>
<td>105</td>
<td>3.9–390</td>
<td>1.0</td>
<td>0.29</td>
<td>4000</td>
</tr>
<tr>
<td>150</td>
<td>4.1</td>
<td>10.3</td>
<td>0.17–17</td>
<td>9.2</td>
<td>1.7</td>
<td>3500</td>
</tr>
<tr>
<td>210</td>
<td>3.0</td>
<td>5.3</td>
<td>0.23–23</td>
<td>5.3</td>
<td>1.2</td>
<td>1200</td>
</tr>
</tbody>
</table>

Table 2.4: Characteristics of the full Dutch Array Mode, assuming 24 core and 16 remote stations, and a beam spectral bandwidth of 4 MHz. Resolutions, $\theta_{res}$, and fields of view, FoV$_{Beam}$, of single 4 MHz beams are given for the selected frequencies. The sensitivities, $\Delta S$, range from the longest (10,000 s) to the shortest (1 s) logarithmic integration times. The number of distinct sources likely to be detected in $2\pi$ steradians, assuming observing only at the specified frequency, is given by $N$, for 5$\sigma$ and 30$\sigma$ detection levels and an integration time of 10,000 seconds. Using the maximum total bandwidth of 48 MHz (i.e. twelve of such 4 MHz beams), the corresponding 5$\sigma$ and 30$\sigma$ data rates, i.e. measurements to be stored per second, are given by $m$. The data growth per day, assuming a 100% time allocation, is given by $d$ for the two detection levels.

Figure 2.3: Expected measurements per second ($m$) to be stored in the Full-NL mode as related to the integration time ($\tau_{int}$) for some selected frequencies, assuming a single beam of spectral beamwidth of 4 MHz, although 12 of these are allowed, giving rise to 12 times higher values. Thick lines represent the 5$\sigma$ detection level measurements, whereas dashed lines represent the 30$\sigma$ level. A measurement of a source is about 300 Bytes of data.
2.3 Expected Data Rates & Volumes for some TKP Observation Modes

is \( \lesssim 230 \) GB. The numbers at the other frequencies yield lower data rates and less storage capacity, as can be seen from Tables 2.4, 2.9 and 2.10.

Furthermore, from the daily averages in Table 2.4, annual estimates can be made based on the data accumulation. Here we take into account that the TKP will be given as much allocation time as the other Science Key Projects, about 20\%, and the assumption that the LBA and HBA observe evenly in time. Then we can conclude, if LOFAR is in the Full-NL mode, the TKP pipeline will process and store about 8 TB/yr at the 5\( \sigma \) level and about 0.8 TB/yr at the 30\( \sigma \) level. Of course, when commensal processing, *piggy-backing*, is enabled and the TKP is granted full data access, these numbers will be 5 times as high, with a maximum of 40 TB/yr.

### 2.3.5 Million Sources Sky Survey – Commensal Mode

The primary goal of the Million Sources Sky Survey\(^1\) (MSSS) is to build a Global Sky Model that serves the calibration of LOFAR. The Global Sky Model (GSM) contains the spectral information of \( 10^5 – 10^6 \) sources in the Northern Hemisphere in the frequency range of roughly 30–200 MHz. It is envisaged that because of the excellent \( uv \)-coverage of 24 LOFAR core stations, a 10 min. observation around transit gives sufficient \( uv \)-coverage to carry out MSSS. The observing frequencies will lie roughly in the middle of the Low and High Band, at 60 and 150 MHz, respectively. About 600 pointings for the LBA and 3500 for the HBA array are needed to Nyquist sample the Northern Hemisphere. The achievable rms sensitivities depend

---


<table>
<thead>
<tr>
<th>Freq. [MHz]</th>
<th>( \theta_{\text{res}} ) [arcsec]</th>
<th>FoV [deg(^2)]</th>
<th>( \Delta S ) [mJy]</th>
<th>( \times 10^6 )</th>
<th>( N )</th>
<th>( M )</th>
<th>( D ) [GB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>413</td>
<td>105</td>
<td>10–255</td>
<td>0.52</td>
<td>0.13</td>
<td>188</td>
<td>26</td>
</tr>
<tr>
<td>150</td>
<td>165</td>
<td>18.4</td>
<td>0.45–11</td>
<td>3.4</td>
<td>0.86</td>
<td>1525</td>
<td>363</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>1713</strong></td>
<td><strong>389</strong></td>
</tr>
</tbody>
</table>

**Table 2.5:** Characteristic parameters of the Million Sources Sky Survey (MSSS) once it has been carried out by using 20 core stations at the LB and HB frequencies of 60 and 150 MHz, respectively, and assuming a 16 MHz beam. The number of pointings needed to Nyquist sample the Northern Hemisphere for the LB and HB frequencies is \( P_{\text{LB}} = 619 \) and \( P_{\text{HB}} = 3515 \), respectively. \( \Delta S \) is the theoretical 1\( \sigma \) rms noise level of the longest (600 s) and shortest (1 s) integration times. The total number of distinct sources likely to be detected by the Survey is given by \( N \) for 5\( \sigma \) and 30\( \sigma \) detection levels. In TKP commensal mode \( M \) is the number of total source measurements made and stored in the TKP database after survey completion and the storage disk size needed is given by \( D \) in the last two columns for 5\( \sigma \) and 30\( \sigma \) detections (see text for the derivations).
on whether the final strategy is to use a single beam of 48 MHz bandwidth (i.e. covering the whole Low Band at once) or three beams of 16 MHz bandwidth each. In our calculations we adopt the latter.

Table 2.5 reports some resulting parameters of MSSS once the survey has been completed in the assumed mode. The number of distinct sources, $N$, likely to be detected is more than three million at the $5\sigma$ level, and about a million at the $30\sigma$ level, if no cross-frequency band source association is taken into account.

However, commensal detection of transients in the MSSS Survey allows the Transients Key Project to inspect the MSSS data at logarithmically spaced time intervals at which the calibrated images will be produced. For this piggybacking mode, the expected number of sources likely to be detected and the number of measurements made in a single 16 MHz MSSS beam at the frequencies 60 and 150 MHz at different integration times are presented in Table 2.11 in Appendix 2.A. As can be seen from Eq. 2.8 and Table 2.11 images are classically confused in the Low Band at integration times longer than 2 and 100 seconds at the $5\sigma$ and $30\sigma$ detection levels, respectively, whereas in the High Band integration times as short as seconds cause the images to be confusion limited. Fig. 2.4 shows the measurements per second, $m$, plotted against the integration times for the MSSS commensal mode.

From Table 2.11 it can be seen that for both Bands the expected source counts, $\langle N \rangle$, are of the same order. The HB, however, needs about 5.5
times more pointings to sample the whole sky, because of its smaller field of view. Furthermore, from the values of \( \langle m \rangle \) at the shortest integration times it can be deduced that when we take into account three similar beams operating at the same time, the average rate is of the order of 1800 sources per second at the 5\( \sigma \) level. The last two columns in Table 2.11 show the total number of measurements made per second in this configuration at the two significance levels. Multiplying the summed value from all integration times by the number of pointings to be made and the time spent on a field (i.e. 600 s), will give the total number of measurements made in the Survey, \( M \), and is given in Table 2.5 for the 5\( \sigma \) and 30\( \sigma \) detection levels. The summed measurements made during the MSSS Survey at the 5\( \sigma \) level has a total of approximately \( 1.7 \times 10^9 \), which corresponds to about 0.5 TB of disk space needed, whereas at 30\( \sigma \) this is about \( 400 \times 10^6 \) measurements and 120 GB of data. It should be noted that these numbers are TKP specific, assuming that the data can be processed commensally, and do not reflect the raw data formats to be stored on the cluster nodes. Furthermore, if the MSSS Survey would produce a real-time data stream of calibrated images, the TKP pipeline should be able to store about 2000 (source) measurements per second, assuming three simultaneously operating 16 MHz beams.

### 2.3.6 Summary

The TKP pipeline faces a wide range of data rates to process and volumes to store, depending strongly on the observation mode of the LOFAR Telescope. Table 2.6 gives a comparable overview of some of the key observation modes of LOFAR relevant to the TKP. Averaged values take into account that the allocation time of the LBA and HBA is evenly distributed during an observation run.

<table>
<thead>
<tr>
<th>Observation Mode</th>
<th>( m ) [s(^{-1})]</th>
<th>( \langle d \rangle ) [GB/day]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZM</td>
<td>2400</td>
<td>400</td>
</tr>
<tr>
<td>RASM</td>
<td>350</td>
<td>50</td>
</tr>
<tr>
<td>Full-NL</td>
<td>4400</td>
<td>500</td>
</tr>
<tr>
<td>MSSS*</td>
<td>1800</td>
<td>360</td>
</tr>
</tbody>
</table>

### Table 2.6: Averaged data rates and volumes expected for some key observational modes of LOFAR. (*) refers to the fact that MSSS will be processed in piggybacking mode by the TKP and is carried out once, while the others are repetitive modes. MSSS will store in total about 0.5 TB of data at the 5\( \sigma \) and about 120 GB at the 30\( \sigma \) level. The Zenith Monitor mode has seven beams scanning the zenith during operation, giving rise to larger rates than the RASM mode where we assumed a single beam operating during observation. Note that for the Full-NL mode the measurement peak values may be as high as 9000 (see Tables 2.4, 2.9 and 2.10 ).
2. Expected Data Rates and Volumes for the LOFAR Transients Key Project

The full Dutch array mode (Full-NL) is the most data intensive configuration, and in Section 2.3.4 the estimated annual growth of the data size was \( \lesssim 8 \text{ TB/yr} \), if LOFAR were in this mode 20\% per year. Other modes of LOFAR are less intensive, and commensal modes will increase this percentage probably by a factor of 5 to \( \approx 40 \text{ TB/yr} \). Note that the presented numbers do not take into account any back-up storage or processing facilities.

Going from "working to working", starting at the 30\(\sigma\) detection threshold will already offer a wealth of information for the various modes, detecting hundreds of thousands of sources, making millions of measurements per day.

2.4 TKP Databases

The data rate and cadence of LOFAR put stringent demands on the TKP pipeline. The time between two subsequent images should not exceed the combined source extraction and database processing time of an image. Image processing times depend strongly on properties like size, integration time, number of frequency bands, resolution and sensitivity. From a database perspective, these may be condensed in the expected number of sources to be detected in an image, which determines the disk space to be allocated, and the query execution times. Spreeuw (2010) gives a rigorous description of the source extraction modules that are implemented in the TKP pipeline, while the previous section of this chapter focuses on the expected database storage sizes of the detected sources. Disk space needed for the storage of the raw data and images is taken care of by the LOFAR Storage Group and will not be discussed here. In this section we will describe the database system and schema used in the TKP pipeline to meet the data flow and retrieval in an optimal way.

The Sloan Digital Sky Survey is by far the most successful astronomical project that uses a database-centric computing approach for their large-scale scientific datasets. Furthermore, it serves as a pathfinder for current and future planned surveys at multiwavelengths, that will be taking enormous amounts of data, e.g. the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Hodapp et al., 2004) and the Large Synoptic Survey Telescope (LSST; Abell et al., 2009) in the optical band and the Allen Telescope Array (ATA; Welch et al., 2009) and Murchison Wide-field Array (MWA; Lonsdale et al., 2009) in the radio band. Their database schema, functions, procedures and algorithms, which has been vastly extended and evaluated over the years, is a good example how the architecture of the environment should be built. The way to approach and start is formulated by one of Gray’s laws: “Bring computations to the data, rather than data to the computations” (Szalay & Blakeley, 2009). This is why procedures and functions run inside the database. The Structured Query Language (SQL)
can access the data directly, giving back summarised or full result sets, instead of transporting the data to other nodes in- or outside the network and process the results, expensively, iteratively. Two other rules from Gray are to start the design with the “top 20 queries” and go from “working to working”. Because we operate in a pipeline framework, in which data are processed in a structured way, we know what our top 20 queries have to do, enabling us to optimise their execution plans.

### 2.4.1 The TKP pipeline Database Schema

Overviews of the TKP pipeline framework in which the calibrated images are processed are given by, e.g., Swinbank et al. (2007) and Swinbank (2010). Spreeuw (2010) developed the Source Finding modules, while here we will focus on the database part. A diagram of the data flow centred at the database interactions in the TKP pipeline is shown in Fig. 2.5. From Fig. 2.5 it can be seen that two separate databases are implemented. One is used in the nearly real-time pipeline during observations and the other is used as a permanent offline catalogue database, collecting all the LOFAR source measurements over time.

At the start of an observation the pipeline database is initialised and loaded with known sources from the catalog database that are in the fields of view during the running observation. These will include sources from the major catalogues as well as the known LOFAR sources once the MSSS Survey has been carried out. The Source Finding modules extract the Gaussian fitted sources and fluxes at positions of particular interest on the sky from the images and pass them on to the database. In this pipeline database sources are then associated to previously detected sources in the observation and to the already known sources, preloaded from the catalogues. After this the transient and variability detection queries run that may send triggers for further actions. Finally, typically on a daily basis, the data will be flushed to the catalog database where the measurements will be appended to the LOFAR catalogue sources. Once LOFAR is fully capable of observing in several different modes, LOFAR’s TKP catalogue is expected to grow with roughly 8 TB/yr, as was worked out in the previous section.

The pipeline database is at the heart of the TKP pipeline system. It should be kept clean and fast to be able to detect differences on the shortest time scales in the images. Therefore, the pipeline database has a temporary character because its content will be flushed at certain periods in times to the permanent catalog database. Furthermore, due to its temporary and limited size, the pipeline database can reside on a single node, close to the input data to follow Gray’s Law.

The goal of the pipeline database is to serve as a tool for the TKP pipeline for detecting transient and variable sources. The properties and measured quantities of all the detected sources will be stored in the database.
In order to respond quickly when necessary, the design should be optimised to these needs. The full and up-to-date data definition and manipulation SQL statements of the database, tables, functions and procedures are maintained in the TKP svn repository\(^2\). A schematic overview of the most relevant tables in the TKP pipeline database is given in Fig. 2.6.

In the design of the pipeline table definitions we adopt Gray’s Laws and incorporate the aspects of the lessons learned from the SDSS SkyServer (Thakar, 2008). Central in the pipeline database are the extracted-sources and catalogedsources tables, together with the basesources and assoccatsources, containing the association information. Extracted

\(^2\)http://svn.transientskp.org/code
Figure 2.6: Schema of the most relevant tables in the TKP pipeline database. Arrows indicate column references between tables.
sources from the Source Finding procedures will be appended to the `extractedsources` table, containing the measured source properties and their $1\sigma$ errors, like position, all Stokes parameters of peak and integrated fluxes, the Gaussian fitting parameters and via `images` the observational frequency and timestamp information, giving rise to about 300 Bytes of data. The `catalogedsources` table contains the preloaded sources from the major catalogues.

The `basesources` table keeps track of the unique sources per band detected in the current observation, and may be regarded as a running catalogue. The band is a standard frequency band as specified in `frequency-bands`. For every new detection of a source, the association parameters (columns) will be updated with the new averaged values. The determination of a source association is treated in Section 2.4.3 and Chapter 3.

Sources in the running LOFAR catalogue, i.e. the `basesources` table, will be mapped to the known sources from the catalogues and stored in the `assoccatsources` table. Similarly associations between current and previously observed sources are maintained in the `assocxtrsources` table as light-curve datapoints. In this way we keep the status of all the (unique) sources detected in the current observation.

The main tables are defined in such a way that the data points from the same source are retrieved in a simple SQL statement, and might therefore be regarded as the light-curve tables. To optimise source association, transient and variability detection, and the data selections for the dumps to facilitate data transport between the two databases, we will create the tables with several columns that also exist in the SkyServer data model as described by Stoughton et al. (2002) and enhanced by Gray et al. (2006).

Ivanova et al. (2007) analysed a typical SkyServer query log of 1.2 Mqueries and found that 83% contained spatial data searches. Gray et al. (2006) replaced the recursive hierarchical triangular mesh (HTM) algorithm by the faster zone algorithm. The latter divides the celestial sphere into declination strips of equal width, a so-called `zone`, defined as an `INT` in the clustered primary key `(zone, ra, id)` of the table. Combined with the Cartesian coordinates, $(x, y, z)$, the dot product is used to calculate distances between sources. To compare the processing times in a MySQL database of both algorithms, we generated 1000 images of $1000^2$ pixels with tens of sources in each image and processed these in the TKP pipeline. From the results shown in Fig. 2.7 it can be seen that the zone algorithm is nearly two orders of magnitude faster, although both increase with the number of images being processed, due to the increasing number of sources needing to be searched.

Because the source association algorithms and related queries rely heavily on spatial searches and are the most intense processing tasks inside the `pipeline` database, we decided to use the zone algorithm.
Figure 2.7: Processing times (vertically) of the HTM (blue circles) and Zone (red squares) algorithms for a series of 1000 simulated images being processed (horizontally). The zone algorithm is nearly two orders of magnitude faster. It was implemented on an Intel(R) Pentium(R) 4 CPU 3.00 GHz with 1 GB RAM, running Fedora 3 (Linux kernel 2.6.12), using MySQL 5.0.22.

2.4.2 MonetDB

The TKP pipeline has implemented the open source column-store database MonetDB\(^3\) (see e.g., Boncz, 2002), developed at the Centre of Mathematics & Informatics (CWI). The column-store model was formerly known as the Decomposition Storage Model (see e.g., Copeland & Khoshafian, 1985; Khoshafian et al., 1987); it splits up a (relational) table vertically into \(c_n\) binary tables, where \(c_n\) is the number of columns. This database system is of a fundamentally different design than the classical relational database systems (RDBMS), such as the open source MySQL and PostgreSQL, or the commercial products Oracle or DB2, but all can be interfaced with the same Structured Query Language (SQL). At CWI, TPC-H\(^4\) benchmark performance tests were executed on the alternative open source database systems of MySQL and PostgreSQL. A series of queries is executed on a predefined and preloaded database. Scaling factors of up to 20 times the initial size of 1 GB reveal that it is not uncommon for the tested classical RDBMSs to give erroneous (empty) result sets and/or that the processing times are extremely long for some of the queries\(^5\).

Full-sized SDSS SkyServer data releases were successfully ported into MonetDB. Although the SkyServer data management system is a tuned

\(^3\)http://monetdb.cwi.nl  
\(^4\)http://www.tpc.org/tpch/  
\(^5\)http://monetdb.cwi.nl/SQL/Benchmark/TPCH
Figure 2.8: A subset of 2 GB of Data Release DR4 was ported into MonetDB. The elapsed times for a test set of 14 queries were compared between MonetDB and a tuned and non-tuned Microsoft SQL Server. 12 Queries are executed faster in MonetDB. Figure adapted from Ivanova et al. (2007).

Microsoft SQL Server, Ivanova et al. (2007) showed that on a subset of the data, the performance of 12 of the 14 most executed queries ran faster in MonetDB, see Fig. 2.8.

In MonetDB every relational table is represented by a group of binary relations, consisting of Binary Association Tables (BATs). A BAT represents a mapping from a unique object id to a single attribute. When the object ids form a dense ascending sequence highly efficient positional lookups are enabled. Direct consequences are that queries only touch the relevant columns, and when in contiguous memory it allows compression and good cache-hit ratios. Furthermore, MonetDB’s kernel is a programmable relational algebra machine operating on "array"-like structures, exactly what CPUs are good at.

To speed up query processing further, MonetDB/SQL implements a query processing architecture based on cracking, in which a column is sorted according to the subsequent insert statements that touch the column. In this scheme the first query pays the price, but all the others benefit from previous queries. Because our database starts small the drawback from this initial start-up is small. Idreos, Kersten & Manegold (2002) showed that a simple `count(*)` query with a range predicate of which the two boundaries were randomly chosen, and was fired a 1000 times, each time with a new random range, on a single column table populated with $10^7$ random values between 0 and 9999 achieved response times that were two orders of magnitude faster than those from PostgreSQL, MySQL and MonetDB with cracking disabled. Fig. 2.9 is adopted from Idreos, Kersten & Manegold (2002) and shows clearly that cracking makes sense at an early stage,
and even more cracking will decrease the response times significantly. This technique fits the TKP databases that start small and grow gradually by appending data, since that is the optimal case for applying the cracking algorithms.

### 2.4.3 Associating LOFAR Sources in the Database

Although the format of the input data is not fully specified yet, we can make some assumptions that will hold, because from the database perspective the number of sources detected per image is relevant. Besides LOFAR’s configuration mode, the characteristic properties of an image are the resolution, synthesised beam parameters, integration time, frequency band, and timestamp of observation, all stored in the images table. We envisage a dataset as streams of image cubes arriving at subsequent timestamps. Streams are divided according to their logarithmic integration times ($\tau_1, \tau_2, ..., \tau_{13}$), and each image cube has the same observational timestamp, whereas the individual image planes in this cube fall in different frequency (sub)bands. Fig. 2.10 depicts this, and in this view a dataset might also be regarded as an observation producing the image cubes. From Fig. 2.10 it can be seen that we can search for transient and variability behaviour in the time and frequency domains.

All measured properties of a source, e.g. position, frequency and all Stokes parameters, plus errors, are stored in the extractedsources table, which is essentially the table containing all the measurements made during an observation. The corresponding properties of the image, in which the source
was detected, are retrievable via ID referencing. A list of unique sources in the current observation is maintained in the basesources table, and a list of known sources, from major catalogues and eventually classified LOFAR sources, is kept in the catalogedsources table. Positions of sources detected by LOFAR are checked for uniqueness and reoccurrence against both lists. Finding a positional match – either a genuine or background association – is done by the source association procedure, which is carried out inside the database by a couple of SQL commands. Construction of light curves is moderated by joining the basesources and extractedsources tables.

The goal of the source association is to find for every source detected by LOFAR all its measurements, current and archived, in order to construct light curves that will aid the source classification. The criteria for which an association pair is considered as real or by chance is done by evaluating three association parameters, as described in Chapter 3. One of the association parameters that is very useful is the normalised distance between the two sources $i$ and $j$, weighted by their positional uncertainties:

$$r_{ij}^2 = \frac{(\alpha_i - \alpha_j)^2}{\sigma_{\alpha,i}^2 + \sigma_{\alpha,j}^2} + \frac{(\delta_i - \delta_j)^2}{\sigma_{\delta,i}^2 + \sigma_{\delta,j}^2}, \quad (2.9)$$

which follows a Rayleigh probability distribution. Cutoff values for this dimensionless positional difference were determined by the simulation runs.
Figure 2.11: Extracted source $X$ has two association candidates $C_1$ and $C_2$, that were found in the box with width and height of $2r_w$ at the location of source $X$. Source $C_3$ falls outside the search area and is not considered as a candidate. The measured properties of all the sources, $X, C_1, C_2,$ and $C_3$ are stored in the database. Depending on the association parameters of the candidate(s), $X - C_i$, the pair may be classified as genuine or chance. If genuine, it is recorded as a related measurement to the source. If it could not be associated, it will be recorded as the first measurement of a source. See text for other cases.

The position of an extracted source is placed at the centre of a box that is searched for counterparts. The distance of the source to the edges of the box, $r_w$, is set fixed in the north- and westward directions, to a value of order of 1 arcmin, but will be determined dynamically depending on the image resolution and local source density in future versions. All sources found in the area are considered as candidate associations. Fig. 2.11 gives a sketch of the case where two candidates were found within the search area of an extracted source.

Then, for every pair the association parameters are calculated and based on the criteria the association is considered as real or by chance. The database implementation of the source association algorithm takes care of the situation where extracted sources from an image are matched to previous detections, as depicted in Fig. 2.12 and is as follows:

(i) For associations found to be genuine, the measurements are appended to the corresponding (unique) source in the running source list, i.e. we update the basessources table. This table maintains the source position and frequency dependent averaged values for flux, $\bar{\nu}$, flux squared, $\bar{\nu}^2$, weight of flux, $\bar{w} \equiv 1/\sigma^2_{\nu}$, weighted flux, $\bar{w}\bar{\nu}$, weighted squared flux, $\bar{w}\bar{\nu}^2$, and the number of data points $N$. These values are used in the variability monitoring indices, which are the primary tools for detecting transient and variability behaviour in the LOFAR.
sources. The indices are defined in Chapter 3 and are given by

\[ V_\nu \equiv \frac{s_\nu}{T_\nu} = \frac{1}{T_\nu} \sqrt{\frac{N}{N-1} \left( \frac{1}{N-1} \sum_{i=1}^{N} (I_{\nu i}^2 - I_{\nu}^2) \right)} \]  

(2.10)

and

\[ \eta_\nu = \frac{N}{N-1} \left( \frac{1}{N-1} \sum_{i=1}^{N} (wI_{\nu i}^2 - \frac{1}{N-1} \sum_{i=1}^{N} wI_{\nu i}^2) \right). \]

(2.11)

From Fig. 2.12 it can be seen that case a follows the above described procedures, whereas case b is similar, except that both association candidates are considered as genuine and no discrimination can be made between the two. Both sources in the source list will be updated with the new values from the extracted sources.

(ii) Sources extracted from a higher-resolution image might be resolved compared to the association candidate sources. In these cases, c and d in Fig. 2.12, we replace the lower-resolution source position in the running source list with the higher-resolution source(s). As a consequence, the averaged values needed for the variability indices should be recalculated for this higher-resolution source.

(iii) Other situations are the first detection of a source, case e in Fig. 2.12, in which case the newly detected source will be added to the running source list. Internal triggers should take further analysis actions, in order to classify this new source. Case f in Fig. 2.12 represents the case
where a previously detected source is not detected again, in which case the local noise levels in the image should be reported.

The sources in the image will also be looked up for occurrences in the static cataloged_sources table, following analogous procedures as described above. Associations between sources in the running source list (i.e. sources stored in basesources) and the known sources from the catalog source list (i.e. the sources stored in cataloged_sources) are maintained in the assoccatsources table.

We define a group of SQL statements that execute the above described procedures, trying to associate all the sources detected in a LOFAR image. Parallelised processing of the images is taken care of by atomic transaction-safe storage of all sources detected in an image. Therefore, the association procedure should run on consecutive, but not necessarily chronologous images, and not in parallel.

Spatial searches are the most used predicates in queries that select one or (many) more sources, and they should therefore have fast response times. A location of interest, (\(\text{\texttt{@ra}}, \text{\texttt{@decl}}\)), is set at the centre of a searchable area with radius \(\text{\texttt{@r_search}}\). All sources that are within a distance \(\text{\texttt{@r_search}}\) of this location are found when using the dot product and Cartesian instead of celestial coordinates. They should obey the clause \(\text{\texttt{x}} \cdot \text{\texttt{c}} > \cos r_s\), where \(\text{\texttt{x}}\) and \(\text{\texttt{c}}\) represent the Cartesian vectors of the location of interest and the sources found, respectively, and \(r_s\) is search radius in radians. Refinement of the search is done by excluding the candidates that fall outside the box surrounding the search area, see Fig. 2.11. Candidates having declinations between the minimum and maximum values of the box are included. An extra clause \text{\texttt{zone BETWEEN @zone_min AND @zone_max}}\ is used for the candidates falling in the declination strips of the box. From Gray et al. (2006) we adopted the \text{\texttt{alpha(@r_search, @decl)}}\ function that inflates the RA of the search radius, with increasing declination towards the celestial poles. It is applied in the RA box boundaries as \text{\texttt{ra BETWEEN @ra - alpha() AND @ra + alpha()}}\.

### Database Processing Times

Code snippets of the most intensive queries that make up the source association and detection procedures are shown in Appendix 2.B, where the first query collects the just extracted sources and their association counterparts from the running catalogue and the second query updates the running catalogue with the new values. These queries were timed for the benchmark tests and were executed in a MySQL and MonetDB database. Identical database tables and queries were created, except for a few minor SQL syntax differences, and both were installed on the same machine, a dual-core
2. Expected Data Rates and Volumes for the LOFAR Transients Key Project

Figure 2.13: Comparison of performance tests of the source association procedures in a MySQL 5.0.45 (red line) and MonetDB v5.20.4 Jun2010-SP1 (blue line) database, carried out on a dual-core 64 bit Intel(R) Pentium(R) 4 CPU 3.00 GHz with 1 GB of RAM, running Fedora 8 (Linux kernel 2.6.26.8-57) desk-top computer. We processed a series of 1000 images (horizontal axes), each containing the number of sources as labeled in the bottom right of the subplots. The response times are shown on the vertical axes.

Figure 2.14: Performance tests of the source association procedures on the data server node in the LOFAR computing cluster. The installed database is MonetDB (Feb2010-SP2), and runs on an eight-core 64 bit Intel(R) Xeon(R) CPU L5420 2.50GHz with 16 GB of RAM. We processed a series of 1000 images (horizontal axes), each containing the 1000 sources per images. The response time is shown on the vertical axes.
2.5 Discussion

As was shown in Section 2.3, a fully operational LOFAR Telescope will provide the Transients Key Project with millions of sources that will be detected millions of times, giving peak data rates that may well exceed 200 GB/day and with storage capacities needed per year of the order of 100 TB if we include commensal, piggybacking, processing and data back-ups. A dedicated fast-responding database system should be available and accessible from within the pipeline framework during ongoing observation runs, whereas the database query and procedure response times should be obviously less than the input rates.

The performance tests reported in Section 2.4.3 show the relationship between the number of sources per image and the processing response time for a series of 1000 consecutive images. If the processing time is shorter than the time between two consecutive images, the observation mode is executable from a database point of view. The same is valid for the time to carry out surveys, where the data need to be processed within the survey time. In these figures we have to take into account the processing times of the calibration and imaging pipelines and the source extraction procedures as well. These depend on many parameters and are not in the scope of the investigations here. However, the tests pointed out that MonetDB is able to process 1000 source measurements within 1 second, whereas MySQL processing times exceed ten seconds. This means that monitor modes can be carried out from the database perspective. At the shortest integration
time scales and when using the full 48 MHz bandwidth, the full-array mode (of which an example was discussed in Section 2.3.4), exceeds the limits of the number of source measurements that are still processable at the shortest time scales. Assuming the ultimate image input rate is one per second, we cannot process these at similar time scales yet. Simple data server hardware extensions of the RAM memory and running it in a more dedicated mode in which MonetDB is installed, will improve the performance (Boncz, Kersten & Manegold, 2008), and allow us to process a larger number of sources per image. Minor improvements of the response times are to be expected from converting data type definitions to a smaller number of bits, e.g., conversion of INT to TINYINT gains a storage factor of four.

Acknowledgement

I would like to thank Ronald Nijboer for helpful discussions about the LOFAR configurations and observation modes.
Table 2.7: This table gives the expected source counts, \( \langle N \rangle \), per integration time \( \tau_{\text{int}} \) and number of measurements per second, \( \langle m \rangle \), at the 5\( \sigma \) and 30\( \sigma \) detection levels, in a single Zenith Monitoring beam of assumed bandwidth of 4 MHz. This beam is one of the seven beams that make up the hexagonal pattern that scans the declination strip at the zenith. The last row per observing frequency gives the total number of measurements per second, for a single beam, to be stored during the time spent at a single field at both detection levels. Classical confusion limits arise in the Low and High Band when \( \langle N \rangle \) is larger than 338 and 371, respectively.

<table>
<thead>
<tr>
<th>Freq. [MHz]</th>
<th>( \theta_{\text{res}} ) [arcsec]</th>
<th>FoV [deg(^2)]</th>
<th>Image [pixels]</th>
<th>( \tau_{\text{int}} ) [s]</th>
<th>( \Delta S ) [mJy]</th>
<th>( \langle N \rangle ) [5( \sigma )]</th>
<th>( \langle N \rangle ) [30( \sigma )]</th>
<th>( \langle m \rangle ) [5( \sigma )]</th>
<th>( \langle m \rangle ) [30( \sigma )]</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>413</td>
<td>105</td>
<td>512(^2)</td>
<td>1</td>
<td>510</td>
<td>80</td>
<td>2</td>
<td>80</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>361</td>
<td>118</td>
<td>8</td>
<td>59</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>228</td>
<td>192</td>
<td>20</td>
<td>38</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>162</td>
<td>272</td>
<td>35</td>
<td>27</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>114</td>
<td>379</td>
<td>56</td>
<td>19</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>72</td>
<td>577</td>
<td>97</td>
<td>12</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>51</td>
<td>780</td>
<td>141</td>
<td>7.8</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>36</td>
<td>1041</td>
<td>203</td>
<td>5.2</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>23</td>
<td>1497</td>
<td>319</td>
<td>3.0</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1000</td>
<td>16</td>
<td>1943</td>
<td>442</td>
<td>1.9</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>11</td>
<td>2498</td>
<td>604</td>
<td>1.2</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5000</td>
<td>7.2</td>
<td>3444</td>
<td>896</td>
<td>0.69</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7200</td>
<td>6.0</td>
<td>3905</td>
<td>1041</td>
<td>0.54</td>
<td>0.14</td>
</tr>
</tbody>
</table>

| 150         | 165             | 18.4          | 512\(^2\)       | 1               | 22              | 161             | 30              | 161             | 30              |
|             |                 |               |                 | 2               | 16              | 214             | 43              | 107             | 22              |
|             |                 |               |                 | 5               | 9.9             | 304             | 67              | 61              | 13.4            |
|             |                 |               |                 | 10              | 7.0             | 393             | 93              | 39              | 9.3             |
|             |                 |               |                 | 20              | 4.9             | 503             | 126             | 25              | 6.3             |
|             |                 |               |                 | 50              | 3.1             | 691             | 185             | 14              | 3.7             |
|             |                 |               |                 | 100             | 2.2             | 877             | 243             | 8.8             | 2.4             |
|             |                 |               |                 | 200             | 1.6             | 1118            | 317             | 5.6             | 1.6             |
|             |                 |               |                 | 500             | 0.99            | 1564            | 442             | 3.1             | 0.88            |
|             |                 |               |                 | 1000            | 0.70            | 2055            | 564             | 2.1             | 0.56            |
|             |                 |               |                 | 1800            | 0.52            | 2638            | 691             | 1.5             | 0.38            |

| 150         |                 |               |                 | 428             | 90.5            |
Table 2.8: The RSM Rapid All-Sky Monitor mode. This table gives per integration time $\tau_{\text{int}}$ the expected source counts, $\langle N \rangle$, and measurements per second, $\langle m \rangle$ at the $5\sigma$ and $30\sigma$ detection levels, in a Rapid All-Sky Monitor beam of assumed bandwidth of 4 MHz. The last row per observing frequency gives the total number of measurements to be stored per second during the time spent at a single field at both detection levels. Classical confusion limits arise in the Low and High Band when $\langle N \rangle$ is larger than 338 and 371, respectively.
### Table 2.9: The LOFAR full Dutch array mode (Full-NL), 24 core and 16 remote stations, assuming a 4 MHz beam at each frequency. Beams from Low and High Band cannot operate together, but multiple beams in either the Low or High Band are possible. The rms sensitivities, $\Delta S$, for a few integration times are shown. The number of sources, $\langle N \rangle$, likely to be detected in the corresponding 4 MHz beam is listed for 5 and 30 times the rms noise. The number of measurements to be stored per second, $m$, is reported in the last column, again specified for the two levels of significance.
## 2. Expected Data Rates and Volumes for the LOFAR Transients Key Project

<table>
<thead>
<tr>
<th>Freq. [MHz]</th>
<th>$\theta_{\text{res}}$ [arcsec]</th>
<th>FoV [deg$^2$]</th>
<th>Image Pixels</th>
<th>$\tau_{\text{int}}$ [s]</th>
<th>$\Delta S$ [mJy]</th>
<th>$\langle N \rangle$ [FoV$^{-1} \tau_{\text{int}}^{-1}$]</th>
<th>$(m)$ [s$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>5.16</td>
<td>16.2</td>
<td>16384$^2$</td>
<td>1 22</td>
<td>101</td>
<td>31</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 7.0</td>
<td>387</td>
<td>94</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 2.2</td>
<td>861</td>
<td>242</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1000 0.70</td>
<td>2061</td>
<td>553</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10000 0.22</td>
<td>6506</td>
<td>1242</td>
<td>0.65</td>
</tr>
<tr>
<td>150</td>
<td>4.13</td>
<td>10.3</td>
<td>16384$^2$</td>
<td>1 17</td>
<td>111</td>
<td>22</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 12</td>
<td>146</td>
<td>31</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 5.7</td>
<td>205</td>
<td>48</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50 2.4</td>
<td>460</td>
<td>126</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 1.7</td>
<td>585</td>
<td>165</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200 1.2</td>
<td>750</td>
<td>213</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500 0.77</td>
<td>1064</td>
<td>295</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1000 0.54</td>
<td>1429</td>
<td>375</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 0.38</td>
<td>1953</td>
<td>477</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5000 0.24</td>
<td>3121</td>
<td>657</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10000 0.17</td>
<td>4615</td>
<td>847</td>
<td>0.46</td>
</tr>
<tr>
<td>180</td>
<td>3.44</td>
<td>7.18</td>
<td>16384$^2$</td>
<td>1 20</td>
<td>63</td>
<td>12</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 6.2</td>
<td>153</td>
<td>36</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 2.0</td>
<td>341</td>
<td>94</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1000 0.62</td>
<td>797</td>
<td>219</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10000 0.20</td>
<td>2397</td>
<td>488</td>
<td>0.24</td>
</tr>
<tr>
<td>210</td>
<td>2.95</td>
<td>5.28</td>
<td>16384$^2$</td>
<td>1 23</td>
<td>37</td>
<td>7</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 16</td>
<td>50</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 5.0</td>
<td>72</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20 5.0</td>
<td>121</td>
<td>29</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50 3.2</td>
<td>166</td>
<td>43</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 2.3</td>
<td>211</td>
<td>57</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200 1.6</td>
<td>269</td>
<td>75</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500 1.0</td>
<td>372</td>
<td>106</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1000 0.71</td>
<td>482</td>
<td>136</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 0.50</td>
<td>637</td>
<td>173</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5000 0.32</td>
<td>958</td>
<td>237</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10000 0.23</td>
<td>1353</td>
<td>301</td>
<td>0.135</td>
</tr>
<tr>
<td>240</td>
<td>2.58</td>
<td>4.04</td>
<td>16384$^2$</td>
<td>1 25</td>
<td>24</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 7.9</td>
<td>62</td>
<td>13</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 2.5</td>
<td>141</td>
<td>37</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1000 0.79</td>
<td>318</td>
<td>90</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10000 0.25</td>
<td>849</td>
<td>201</td>
<td>0.085</td>
</tr>
</tbody>
</table>

**Table 2.10:** Same as Table 2.9, but here specified for some of the High Band frequencies. The source count numbers at 120 and 150 MHz for the longest integration times have a higher uncertainty due to the extrapolations from the model from Huynh et al. (2005).
### Table 2.11: The MSSS commensal mode. Characteristic parameters for the Million Sources Sky Survey carried out by 24 LOFAR core stations at the frequencies of 60 MHz and 150 MHz, assuming a single 16 MHz beam. Eventually, MSSS might be carried out with three simultaneous 16 MHz beams. Logarithmically spaced integration times $\tau_{\text{int}}$ are taken into account. $N_{\text{point}}$ is the number of pointings needed to Nyquist sample the Northern Hemisphere for the given field of view. $\Delta S$ is the theoretical 1$\sigma$ rms noise level. $\langle N \rangle$ is the number of sources likely to be detected in the MSSS beam per integration time $\tau_{\text{int}}$, specified for 5$\sigma$ and 30$\sigma$ detections. The total number of measurements made per second, $\langle m \rangle$, is given in the last two columns per detection level. The last row per frequency gives the grand total, which is simply the sum of the measurements made at every integration time. To sample the whole sky with 619 pointings about $190 \times 10^6$ source measurements at 5$\sigma$ are made during the MSSS Survey at 60 MHz, whereas at 150 MHz with $\sim 3500$ pointings about $1.5 \times 10^9$ measurements are made. For the 30$\sigma$ levels, we expect 400 million measurements in total. Classical confusion limits arise in the Low and High Band when $\langle N \rangle$ is larger than 338 and 371, respectively.

<table>
<thead>
<tr>
<th>Freq. [MHz]</th>
<th>$\theta_{\text{res}}$ [arcsec]</th>
<th>FOV [deg$^2$]</th>
<th>$N_{\text{point}}$</th>
<th>Image [pixels]</th>
<th>$\tau_{\text{int}}$ [s]</th>
<th>$\Delta S$ [mJy]</th>
<th>$\langle N \rangle$</th>
<th>$\langle m \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>413</td>
<td>105</td>
<td>619</td>
<td>512$^2$</td>
<td>1</td>
<td>255</td>
<td>164</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>180</td>
<td>244</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>114</td>
<td>380</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>81</td>
<td>523</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>57</td>
<td>710</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>36</td>
<td>1043</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>25</td>
<td>1375</td>
<td>287</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>18</td>
<td>1792</td>
<td>399</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>11</td>
<td>2502</td>
<td>606</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>600</td>
<td>10</td>
<td>2670</td>
<td>656</td>
</tr>
<tr>
<td>150</td>
<td>166</td>
<td>18.4</td>
<td>3515</td>
<td>512$^2$</td>
<td>1</td>
<td>11</td>
<td>280</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>7.8</td>
<td>362</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>4.9</td>
<td>503</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>3.5</td>
<td>640</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>2.5</td>
<td>813</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>1.6</td>
<td>1118</td>
<td>317</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>1.1</td>
<td>1439</td>
<td>408</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td>0.78</td>
<td>1878</td>
<td>522</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td>0.49</td>
<td>2765</td>
<td>717</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>600</td>
<td>0.45</td>
<td>3004</td>
<td>763</td>
</tr>
<tr>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>723</td>
<td>172</td>
<td></td>
</tr>
</tbody>
</table>

5$\sigma$ | 30$\sigma$ | 5$\sigma$ | 30$\sigma$
2. Expected Data Rates and Volumes for the LOFAR Transients Key Project

2.B Benchmark Queries

```
INSERT INTO tempbasesources
(xtrsrc_id, datapoints, I_peak_sum, I_peak_sq_sum, weight_peak_sum, weight_I_peak_sum, weight_I_peak_sq_sum)
SELECT b0.xtrsrc_id,
       b0.datapoints + 1 AS datapoints,
       b0.I_peak_sum + x0.I_peak AS i_peak_sum,
       b0.I_peak_sq_sum + x0.I_peak * x0.I_peak AS i_peak_sq_sum,
       b0.weight_peak_sum + 1 / (x0.I_peak_err * x0.I_peak_err) AS weight_peak_sum,
       b0.weight_I_peak_sum + x0.I_peak / (x0.I_peak_err * x0.I_peak_err) AS weight_i_peak_sum,
       b0.weight_I_peak_sq_sum + x0.I_peak * x0.I_peak / (x0.I_peak_err * x0.I_peak_err) AS weight_i_peak_sq_sum
FROM basesources b0,
     extractedsources x0
WHERE x0.image_id = @imageid
  AND b0.zone BETWEEN CAST(FLOOR((x0.decl - @theta) / x0.zoneheight) AS INTEGER)
                                  AND CAST(FLOOR((x0.decl + @theta) / x0.zoneheight) AS INTEGER)
                                  AND ASIN(SQRT((x0.x - b0.x)*(x0.x - b0.x)
                                 + (x0.y - b0.y)*(x0.y - b0.y)
                                 + (x0.z - b0.z)*(x0.z - b0.z)
                                  ) / 2
                                  )
     / SQRT(x0.ra_err * x0.ra_err + b0.ra_err * b0.ra_err
         + x0.decl_err * x0.decl_err + b0.decl_err * b0.decl_err)
     < @assoc_r;

UPDATE basesources
SET datapoints =
   (SELECT datapoints
    FROM tempbasesources
    WHERE tempbasesources.xtrsrc_id = basesources.xtrsrc_id
   )

,i_peak_sum =
   (SELECT i_peak_sum
```
FROM tempbasesources
    WHERE tempbasesources.xtrsrc_id = basesources.xtrsrc_id
 )
, i_peak_sq_sum =
    (SELECT i_peak_sq_sum
        FROM tempbasesources
        WHERE tempbasesources.xtrsrc_id = basesources.xtrsrc_id
    )
, weight_peak_sum =
    (SELECT weight_peak_sum
        FROM tempbasesources
        WHERE tempbasesources.xtrsrc_id = basesources.xtrsrc_id
    )
, weight_i_peak_sum =
    (SELECT weight_i_peak_sum
        FROM tempbasesources
        WHERE tempbasesources.xtrsrc_id = basesources.xtrsrc_id
    )
, weight_i_peak_sq_sum =
    (SELECT weight_i_peak_sq_sum
        FROM tempbasesources
        WHERE tempbasesources.xtrsrc_id = basesources.xtrsrc_id
    )
WHERE EXISTS (SELECT xtrsrc_id
        FROM tempbasesources
        WHERE tempbasesources.xtrsrc_id = basesources.xtrsrc_id
    );
Chapter 3

LOFAR’s Transients and Variability Detection Algorithms

B. Scheers, H. Spreeuw, A. Kamble, A.J. van der Horst, E. Rol, R.A.M.J. Wijers & the LOFAR Transients Key Science Project
to be submitted

Abstract

LOFAR, the LOw Frequency ARray, will produce vast amounts of radio data that will be inspected by the Transients Key Science Project to search for and study transient and variable sources. An automated software pipeline that processes the data streams of calibrated images of 2 Gb/s is under development, with the aim to identify, classify and monitor astronomical transients in nearly real-time. An implemented light-curve database keeps track of all the sources detected by LOFAR and compares them with previous LOFAR observations and preloaded source lists from external catalogues. To detect transient and variable sources, the responsible algorithms are investigated. Algorithms concerning source association and variability detection are developed inside the database engine as an integrated part of the software pipeline. To quantify the reliability of a source association, we use a likelihood ratio method, whereas for detecting source flux changes we use two variability indices. The methods are applied to simulated data and the intensively monitored field of GRB030329. The criteria we found for classifying a source association as reliable and for detecting a source as variable, were applied to the real data sets of GRB030329. The Transients Key Project software pipeline identified several sources of interest, besides the target object, in the GRB030329 field. Besides GRB030329, we found eight sources to be variable in one or more bands of which two are serious candidates for multiwavelength follow-up observations.
3.1 Introduction

The International LOFAR Telescope currently has 28 stations out in the fields that are operational\(^1\). Core and remote stations are located in the Netherlands with baselines up to 3 and 100 km respectively, whereas the European stations have baselines up to 1000 km. A station field consists of two types of antennas, the Low Band Antennas (LBAs) and the High Band Antennas (HBAs). Both types are dual-dipole antennas that are sensitive in the low-frequency regime of 30–80 MHz and 110–240 MHz, respectively. Submillijansky sensitivities in the High Bands can be achieved in less than 500 seconds of integration time in full-array mode, with fields of view of the order of 10 deg\(^2\). Larger fields of view, >100 deg\(^2\), are formed in the Low Bands, where sensitivities reach approximately 20 mJy for similar integration times. The individual signals are preprocessed and beam-formed at the station site before they are sent to the central processing supercomputer. Here, the data will be filtered and correlated before they are moved to a dedicated cluster for calibration and imaging processing (e.g., de Vos et al., 2009; Wijnholds, 2010). An automated software pipeline, currently under development by the Transients Key Project, is connected to the imaging pipeline in order to inspect and search the calibrated images for transient events and variable sources.

In the near future the Million Sources Sky Survey (MSSS) is planned. It will provide LOFAR with a Global Sky Model (GSM) of the northern hemisphere. The GSM will improve and evolve in time and serves the calibration and imaging process once the source parameters are validated and the astrometry is better than 0.5 arcsec to meet LOFAR’s resolution of the longest baselines. Furthermore, this survey will give more technical insight into the next phases and prepares a fully operational LOFAR for future surveys.

Other dedicated configurations of LOFAR will serve as strategic modes for detecting transient and variable sources. For instance, two Radio Sky Monitor (RSM) modes, the Rapid All-Sky Monitor (RASM) and the Zenith Monitor (ZM) will both scan the northern hemisphere with large fields of view in continuous mode at distinct frequency bands, roughly once a day. Fender et al. (2007) describe the RSM modes, the main scientific foci of the Transients Key Project (TKP) and divides the sources likely to be detected by LOFAR into two types. To the first belongs the class of jet sources, incoherent synchrotron emitters, which is divided into subclasses of Cataclysmic Variables, X-ray binaries, Gamma-Ray Bursts (GRBs), Supernovae, and Active Galactic Nuclei (AGN). The second type contains the classes of the coherent radio emitters, among which are the flare stars, brown dwarfs, active binaries, extrasolar planets, and extragalactic radio bursts. Radio pulsars are also included and van Leeuwen & Stappers (2010) investigate

\(^1\)Current status can be followed at http://www.lofar.org
3.2 Database and Internal Algorithms

Two of the main goals of LOFAR and the TKP are to detect transient and variable sources, and to build a catalogue that contains all the individual measurements of the sources observed by LOFAR, as raw material for further study of the transients.

The growth and volume of this catalogue database and the ability to do data mining at the same time, in combination with user ad-hoc query requests, rule out most of the mainstream relational database systems.

A fundamentally different design of a database system is the column oriented model adopted in MonetDB (Boncz, 2002). In this model the columns of a relational table are represented by binary tables mapping a unique object ID to a single attribute, thereby only retrieving relevant columns. Standard TPC-H Benchmarking for large volumes of data show significant

---

2See http://www.scipy.org/SciPyPackages/Ndimage for documentation
3http://monetdb.cwi.nl

53
differences\(^4\) between relational and column oriented systems. When sorted in a dense sequence, efficient positional lookups are enabled. Idreos, Kersten & Manegold (2002) further optimised the query processing architecture by implementing a cracking scheme that can speed up repetitive queries by two orders of magnitude. Furthermore, MonetDB’s kernel is a programmable relational algebra machine operating on array-like structures, which allows even further optimisation of the pipeline software. The queries that are executed from within the TKP pipeline are known and \textit{fixed}, thereby benefitting most from the characteristics mentioned above.

### 3.2.2 Database Schema

The implemented database schema of the TKP pipeline is relatively simple, and therefore we will only highlight the most relevant tables. Calibrated images that are produced by the imaging pipeline will serve as input for the TKP pipeline. Raw image data are temporarily stored on the storage cluster with a capacity of about a Petabyte. Archival and temporary data storage will not be discussed here. Before the images are processed, an observation entry containing the characteristics is created in the database. Sources from the major catalogues like VLSS, WENSS, NVSS and eventually LOFAR itself are loaded into the \texttt{catalogedsources} table for association purposes. Meta-data from an image, like observation timestamps, frequencies, synthesised beam shapes, etc., are stored in the \texttt{images} table, having a reference to the observation to which it belongs. Sources detected in an image by the Python source extraction modules will be stored directly in the \texttt{extractedsources} table, with a reference to the \texttt{images} table. For these sources, the database will be searched for matching detections in previous images (i.e. in the \texttt{extractedsources} table) and in the \texttt{catalogedsources} table. Only those sources that lie within the search area of an extracted source are taken into account. For all these candidate association pairs found, the association properties that will be explained in Section 3.2.3 are attached to the pair. An extracted–extracted pair will be appended to the \texttt{assocextrsources} table, and analogously the found extracted–catalogued couples to the \texttt{assoccatsources} table. The source association procedures run inside the database, together with auxiliary functions. The association tables are defined such, that retrieval of various kinds of source lists is fairly easy, avoiding multiple and/or recursive joins. These tables can in fact be regarded as the light-curve tables.

Furthermore, newly detected sources and sources flagged as transient or variable by the algorithms described in Section 3.2.4, will be copied to a separate table. Writing to this table activates triggers that undertake a broad range of further actions (deeper evaluation, fitting, classification),

\(^4\)http://monetdb.cwi.nl/SQL/Benchmark/TPCH/
before alerts will be sent out. This last part is work in progress.

### 3.2.3 Source Association

To keep track of sources and fluxes at positions of interest in the sky the TKP pipeline must be able to match sources by position in time as well as in frequency. For this purpose, a large catalogue of sources and their light-curve characteristics must be available all the time with response times less than the shortest integration times. The source association supports the creation of a long-term catalogue, containing millions of distinct sources with their light-curve properties for scientific analysis and serving the calibration of LOFAR.

A simple method to associate two point sources is if their positions coincide within their positional errors. This was explored by Richter (1975). Further development by including spectral information and local source density was done by De Ruiter, Willis & Arp (1977) and Sutherland & Saunders (1992), who calculated and defined probability ratios of true and chance associations. Applications and modifications were made by Rutledge et al. (2000), who cross-correlated the ROSAT/Bright Source Catalog of X-ray Sources with the USNO A-2 Optical Point Sources. They compared on-source fields with source-offset background fields, in order to calculate for each source association pair a probability of the association being unique. Attached probabilities of source associations made at multiple frequencies, will give more precise information about the fluxes extrapolated to the LOFAR bands of the sources that serve as calibrators.

Following mainly Sutherland & Saunders (1992) and Rutledge et al. (2000) we will investigate the association parameters that set the criteria of a reliable association. The distance in arcsec on the sky between source $i$ and its association counterpart $j$, $\theta_{ij}$, is defined as the first association parameter, whereas the second is the normalised distance, weighted by the positional errors of both sources

$$r_{ij} = \frac{(\Delta \alpha)_{ij}^2 + (\Delta \delta)_{ij}^2}{\sigma^2_{\Delta \alpha,ij} + \sigma^2_{\Delta \delta,ij}}, \quad (3.1)$$

where

$$\sigma^2_{\Delta \alpha,ij} = \sigma^2_{\alpha,i} + \sigma^2_{\alpha,j}, \quad (3.2)$$

and $\sigma^2_{\Delta \delta,ij}$ is defined analogously. $\sigma_{\alpha(\delta),i}$ are the 1$\sigma$ rms uncertainties for source $i$ in right ascension (declination), respectively. In these calculations we assume that the $\sigma_{\alpha}$ and $\sigma_{\delta}$ are uncorrelated.

Genuine associations have positional differences due to measurement errors, which follow a Rayleigh distribution (e.g., De Ruiter, Willis & Arp, 1977), that gives the probability of having a true association between $r_{ij}$ and $r_{ij} + dr_{ij}$.
\[ dp_{\text{true},ij} = r_{ij} \exp\left(-\frac{r_{ij}^2}{2}\right) dr_{ij}. \]  \hspace{1cm} (3.3)

From the Rayleigh probability density distribution, \( r \exp(-r^2/2) \), the derived integral probability of finding a source association at \( r \geq \rho \) is theoretically given by \( p_r(r \geq \rho) = \int_{r=\rho}^{\infty} r \exp(-r^2/2) dr = \exp\left(-\frac{\rho^2}{2}\right) \). This may be used for determining the search radius, \( r_s \), of the area that will be scanned for possible counterparts. If we allow missing \( 10^{-4} \) counterparts, this corresponds to setting the search radius to \( r_s \leq 4.29 \), whereas missing a factor of \( 10^{-7} \) limits the search radius to \( r_s \leq 5.68 \). All counterparts falling in the search area will be taken into account as a candidate for association.

The probability of a chance association occurring between \( r_{ij} \) and \( r_{ij} + dr_{ij} \) with a background source follows the Poisson distribution and depends on the local source density, \( n_L \), of the search area

\[ dp_{\text{chance},ij} = 2\pi r_{ij} dr_{ij} \sigma_{\Delta \alpha,ij} \sigma_{\Delta \delta,ij} n_L, \]  \hspace{1cm} (3.4)

where \( n_L \) depends on frequency, region in the sky and the flux level, and where \( \sigma_{\Delta \alpha,ij} \) and \( \sigma_{\Delta \delta,ij} \) are defined as in Eq. 3.2.

We can now define the third association parameter as the likelihood ratio, given by

\[ LR_{ij} = \frac{dp_{\text{true},ij}}{dp_{\text{chance},ij}} = \frac{\exp\left(-\frac{r_{ij}^2}{2}\right)}{2 \pi \sigma_{\Delta \alpha,ij} \sigma_{\Delta \delta,ij} n_L}. \]  \hspace{1cm} (3.5)

Because \( LR \) decreases exponentially with increasing \( r \) it is often more convenient to use the logarithmic value of \( LR \) as the value of the third association parameter. The reliability of a source association pair can be determined by evaluating the log \( LR_{ij} \) value.

In short the processing is as follows. All images are processed in the TKP pipeline, and the detected sources are stored in the database. For every source that is inserted, the database will be searched for candidate sources that lie in an area of radius \( r_s \) around the current source. Previous images as well as the major catalogues (VLSS, WENSS and NVSS) will be included in the association search. For all source pairs found (current source−candidate source) the values defined above will be calculated and the pair will be appended to the association table. A source having multiple counterparts in its search area, will thus have multiple table entries. This table is defined in such a way that light-curve data points can be generated quite easily per source. Simple aggregate functions and selection clauses on the association parameters of this table are used to give statistical information for these sources.
3.2 Database and Internal Algorithms

3.2.4 Transient and Variability Detection

To inspect the detected sources on variability or transient behaviour, we need to have some indicators that will alert us when a flux is changing with time. We will compare the current state of a source with available information stored in the catalogue database. For this, summarised and/or averaged source properties might be useful when the cadence is too high to explore all the individual values. Here we will define two variability indices. The first is based on the flux deviation from the average and can be interpreted as the magnitude of variability. The other index weighs the deviation from the weighted average with the errors and can be seen as the significance of the variability.

The magnitude of the flux variability of a source can be expressed as the ratio of the sample flux standard deviation, \( s_\nu \), and the sample arithmetic mean flux \( I_\nu \). The sample standard deviation is defined as

\[
s_\nu = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (I_{\nu,i} - \bar{I}_\nu)^2},
\]

(3.6)

where \( \bar{I}_\nu = \frac{1}{N} \sum_{i=1}^{N} I_{\nu,i} \) is the mean flux and \( N \) is the number of measurements. The first indicator, the magnitude of the flux variability of a source, written in aggregate form, which is easy to handle in database queries on bulk data, is then defined as

\[
V_\nu = \frac{s_\nu}{\bar{I}_\nu} = \frac{1}{\bar{I}_\nu} \sqrt{\frac{N}{N-1} \left( \frac{\bar{I}_\nu^2 - \bar{I}_\nu^2}{\sigma_{\nu,i}^2} \right)}.
\]

(3.7)

The second indicator, which expresses the significance of the flux variability, is based on reduced \( \chi^2 \) statistics. We assume that the weighted average flux value is a fitted parameter, so that the number of degrees of freedom is \( N - 1 \). It is given by the sum of the squared deviations from the weighted average weighted by the errors, and divided by the number of degrees of freedom

\[
\eta_\nu \equiv \chi_{N-1}^2 = \frac{1}{N-1} \sum_{i=1}^{N} \frac{(I_{\nu,i} - \bar{I}_\nu^*)^2}{\sigma_{\nu,i}^2},
\]

(3.8)

where now, \( \bar{I}_\nu^* \) is the weighted average; if we define the weight as \( w_{\nu,i} = 1/\sigma_{\nu,i}^2 \), with \( \sigma_{\nu,i} \) the 1\( \sigma \) error of the \( i \)-th flux measurement then

\[
\bar{I}_\nu^* = \frac{\sum_{i=1}^{N} w_{\nu,i} I_{\nu,i}}{\sum_{i=1}^{N} w_{\nu,i}}.
\]

(3.9)

In Eq. 3.8, \( \eta_\nu \) is the well-known \( \chi^2 \) probability distribution (see, e.g., Barlow, 1989; Bevington & Robinson, 2003), which, in our case, is expressed as
Furthermore, we choose to test the null hypothesis, $H_0$, that the source under consideration is not variable. Contributing terms to $\eta_\nu$ in the sum will be of the order of unity, giving a value of roughly one after $N$ measurements. With the integral probability, $p_{\eta_\nu} = \int_{\eta_\nu'}=\eta_\nu \, p_{\eta_\nu}(\eta_\nu', N - 1) \, d\eta_\nu'$, we can quantify the probability $p_{\eta_\nu}$ of having a value equal to or larger than the $\eta_\nu$ obtained from the measurements. This probability $p_{\eta_\nu}$ will justify a rejection of $H_0$ once the confidence level is set. It also determines the expected rate of false positives based on the source counts models.

Again, the significance of the flux variability, $\eta_\nu$, in Eq. 3.8, can be written as a simple aggregate function

$$\eta_\nu = \frac{N}{N - 1} \left( \frac{w I_\nu^2 - \overline{w I_\nu^2}}{\overline{w}} \right).$$

(3.10)

$\eta_\nu$ is close to unity when a source does not show significant variability. When a source has an outlier in one of its flux measurements $V_\nu$ will be large, but depending on the error of the flux measurement, the significance of variability $\eta_\nu$ will be either small when $\sigma_{I_\nu}$ is large or large when $\sigma_{I_\nu}$ is small. $V_\nu$ and $\eta_\nu$ will be helpful in the source classification routines. The classification might discriminate between four separate regions of source variability. Those where the magnitude $V_\nu$ is large or small in combination with a large or small significance $\eta_\nu$.

Other descriptive light-curve parameters, like the probability of having a series of consecutive flux measurements deviating above or below from the average, and long-term variability fits are not investigated here, but will be investigated in future work.

## 3.3 Simulated Data

### 3.3.1 The Data

To analyse the behaviour of the source association and variability parameters defined in Sections 3.2.3 and 3.2.4, we used a sample of 1000 images of simulated data. These were processed sequentially in the TKP pipeline. Sources in the images are at fixed positions and only vary due to the background noise. All the source associations that are made are therefore genuine and can be considered as true. These will give us the boundaries of $r$ and log $LR$ in which reliable source associations fall. Furthermore, we inserted a transient source that went into outburst at some random point in time, followed by an exponential decay. By monitoring the variability indices, this source should be picked up by the TKP pipeline as a transient source. The variability indices $V_\nu$ and $\eta_\nu$ (or $p_{\eta_\nu}$) of the sources in the sample may guide us in setting the constraints on the transient alert system.

The images are created from VLA archival data of the discovery observ-
tion of GCR T J1745-3009 at 325 MHz by Hyman et al. (2005). As described in Chapter 2 of Spreeuw (2010), we only used the \((u,v)\) coverage. Gaussian noise was added to the visibilities, by using the data reduction package AIPS and its tasks UVMOD and WTMOD and the adverb FLUX. An average rms noise of 10 mJy/beam was generated in each map. This was chosen to have it comparable to the 325 MHz WENSS 5\(\sigma\) flux limit which is at 18 mJy as reported by Rengelink et al. (1997). The final 1000 noise maps were created with the AIPS task IMAGR in natural-weighting mode, with a pixel size of 12 arcsec in both directions, resulting in images that are \(256 \times 256\) pixels or somewhat less than \(1^\circ \times 1^\circ\) in size.

In every noise map we inserted 64 sources on a rectangular grid. The inserted position of a single source is the same in all the images, although fluctuations occur due to the different local noise in each map. The fluxes of the sources are linearly spaced between 1 mJy and 1 Jy. The lower limit was chosen to have sources below the average noise level as well as sources that have fluxes at the same level of the local noise in a single image, causing this source not to be detected in all of the images. Like the position, the flux of a single source is assumed to fluctuate due to different local noise in the images. The time between two consecutive images is of course arbitrary, but for the purpose of illustration it is set to 1 h.

At a random point in time one of the existing sources is substituted by a transient source. This transient has a random peak flux between 1 and 5 Jy and an exponential decay time chosen randomly to lie between 10 and \(10^5\) s.

### 3.3.2 Source Association, \(r\) and \(\log LR\)

We know that all the source associations made in the sample are true and that if a source could not be associated it is because its flux is below the detection threshold or disappears in the local noise of the image. The normalised number distribution of the dimensionless distance, \(r_{ij}\), and the logarithm of the likelihood ratio, \(\log LR_{ij}\), of the (true) association pairs of the sample are shown in Fig. 3.1. The numbers peak towards low \(r\) and high \(\log LR\). The minimum value of \(\log LR\) is at \(\log LR_{\text{min}} = 0.80\), whereas the maximum value of \(r_{ij}\) for a source association in the sample is at \(r_{\text{max}} = 3.48\), corresponding to a probability of 0.2\%. Fig. 3.2 shows the normalised cumulative distribution of \(r\) of the associations found in the sample. As can be seen, classifications for genuine associations fall well below the theoretical cutoffs as given in Section 3.2.3. From the sample values associations may be classified as genuine for cutoffs at \(r \leq 3.3\) or at \(\log LR > 1.4\), in which cases we omit 0.01\% of the true associations. These numbers differ slightly from the given theoretical values in Section 3.2.3, because here we compare the positional deviations in \(\alpha\) and \(\delta\) of a source association to the first stored detection of a source, which does not change in time during the run, whereas the Rayleigh distribution expects both values to be randomly chosen from
3. LOFAR’s Transients and Variability Detection Algorithms

Figure 3.1: Histogram of the normalised number distribution of the association pairs in the sample of 1000 images. On the horizontal axis the dimensionless normalised distance, $r_{ij}$, of an association pair is plotted, and on the vertical axis the logarithm of the corresponding likelihood ratio ($\log LR_{ij}$). The cell sizes are $dr \times d\log LR = 0.1 \times 0.1$. The color bar represents the logarithm of the normalised number of association pairs in a cell.

Figure 3.2: Normalised cumulative distribution of the normalised distance, $r$, of the source association pairs in the sample as described in Section 3.3.1. The bin width of $r$ is fixed to 0.1.
their Gaussian distributions.

### 3.3.3 Variability

The variability indices $V_\nu$ and $\eta_\nu$ of a source are calculated using all its $N$ flux measurements. From now on we drop the subscript $\nu$ since we work in a single frequency band. The variability indices of all 64 sources in the sample of 1000 images are plotted in Fig. 3.3. As briefly mentioned in Section 3.2.4 it may be divided into four characteristic regions. The lower left containing the sources that do not show pronounced and significant variability. Fluxes may fluctuate, but only in the margins of the rms noise. Sources residing in the lower right part show low, but significant, variability. A source showing periodic variability of, say 1%, where the deviations from the average are well above the rms will be located here. Going to the upper left domain we will encounter sources that show variability, but at low significance. Large deviations from the mean, in combination with large flux errors will result in a lower level of significance. Sources falling in the upper right quarter, display large and significant variability. As can be seen in Fig. 3.3, the transient is located in the upper right area at $V = 0.431$ and $\eta = 365.58$. The $\eta$'s of the other sources, with $\eta_{\text{max}} = 1.119$, all fall in the regime where the probabilities of these values for $\eta$ are $p_\eta > 0.5\%$, meaning that we cannot reject $H_0$ for these sources. Stricter rejection criteria are met when

![Figure 3.3](image_url)

**Figure 3.3:** The significance of variability $\eta_\nu$ against the magnitude of variability $V_\nu$ for the 64 sources in the sample. Each source is represented by a red circle, whereas the number of associations that was made for a source is not shown here. The red circle in the upper right part of the plot is the inserted transient. The two blue squares are plotted for explanatory purposes (see text).
we set cutoffs for the probability at $p_\eta < 10^{-3}$. In our data sample, with $N - 1 = 999$, this would imply finding sources with $\eta_\nu \geq 1.144$. We have to note that the variability indices of the sources, as plotted in Fig. 3.3, were calculated after all the images were processed.

The light curve of the transient as found by the TKP pipeline is plotted in Fig. 3.4, from which it can be seen that the flux was constant during the first 32 days of the sample (774 h after $t_0$), then increased significantly, dropped exponentially until it disappeared in the noise and became undetectable after day 34.7.

Fig. 3.5 shows the development of the variability indices with time of this transient source. Both parameters are fairly constant in time, apart from the beginning when they need to settle. The sudden steep rise occurs at the same moment of outburst of the transient source (775 h after $t_0$) and continues until the end of detection.

Let us assume that we have an intrinsically variable source with an average flux of $\bar{I}$, and that all the measurements were taken with the same error as the rms of the noise $\sigma_{\text{rms}}$, in order to apply a constant weight $w$. For large $N$, Eqs. 3.7 and 3.10 can be rewritten to have for this source $\varpi = \sqrt{\bar{dI}^2 / \bar{I}}$ and $\bar{\eta} = w \bar{dI}^2$, respectively. The latter may be expressed in terms of the rms noise or with the use of $\bar{dI} = n\sigma_{\text{rms}}$ as $\bar{\eta} = n^2$. This means that a well-sampled variable source with an amplitude of the flux variations at 5% of the
average flux \( A = a \bar{T} \), with \( a = 0.05 \) and its \( dI \) all at the \( 10\sigma_{\text{rms}} \) level will be located in the lower right region at \( \eta = n^2 = 100 \) and \( V = a/\sqrt{2} \approx 0.035 \). This example source is plotted in Fig. 3.3 as one of the blue squares. Another example is a source having \( \eta = 3, V = 0.2 \) (see Fig. 3.3). Depending on the number of degrees of freedom, \( N - 1 \), the probability \( p_\eta \) is used to accept or reject \( H_0 \). For six data points, \( N - 1 = 5 \) and \( p_\eta = 1.04 \times 10^{-2} \), which is not that unlikely as could be seen from the sample. However, if we had a larger number of measurements for these values, and thus more degrees of freedom, the probabilities would decrease further, making the rejection of \( H_0 \) very plausible.

The fastest transient and variability triggers will be based on the evaluation of the (simple) variability indices, \( V_\nu \) and \( \eta_\nu \). Other simple triggers are included when the flux of a known, i.e., catalogued, source, deviates more than \( n \) percent from the average value. As soon as one of the parameters rises significantly, the triggers will activate other inspection modules. These might consist of fitting and classifying the light curve, sending alerts (internal as well as external) when a series of consecutive rising data points in the variability curves are observed. Reconfiguration of the telescope to observe with higher time and angular resolution might then be one of the follow-up strategies.
3.4 GRB 030329 Field

3.4.1 The Data

GRB 030329 was one of the nearest gamma-ray bursts detected and was among the brightest and best-sampled radio, optical and X-ray afterglows ever. This led to still ongoing long-term multifrequency radio follow-up campaigns spanning more than seven years now. The afterglow was followed by the Westerbork Synthesis Radio Telescope (WSRT) at frequencies ranging from 350 MHz to 8.4 GHz (e.g., Van der Horst et al., 2008; Kamble et al., 2010, and references therein). The intensively monitored field of view of GRB 030329 makes it a good candidate to search for transients and variable sources.

For this, we collected already reduced and analysed WSRT data from observations taken in the period from the end of 2003 to the end of 2007. We refer to Van der Horst et al. (2008) for the details of the data reduction. In order to further optimise the WSRT data we performed the AIPS task PBCOR to have the images corrected for the primary beam attenuations at the different frequencies. The selection of the datasets for our sample is listed in Table 3.1.

To search for variable sources we apply the algorithms described in Sections 3.2.3 and 3.2.4 while processing all images in the TKP pipeline. We set the source extraction detection threshold to seven times the local rms noise. Spreeuw (2010) showed that this value minimises the number of false source detections near the edges of the images and still finds the fainter sources. The characteristic rms noise in the WSRT 1.4 GHz images is of the order of 0.1 mJy/beam, with a synthesised beam of $62'' \times 21''$. Extracted sources from an image were associated with detected sources from previous processed images in the sample. In this way data points are appended to the light curves of the sources already found in the sample, or when detected for the first time a new branch for light-curve data points is created. The association algorithm only couples sources by positional match and does not take into account any assumptions on the spectral correlation of the associated pair. This enables the routines to find association pairs across multiple frequency bands by just evaluating their positions. The optical or X-ray counterparts positions of AGNs might be offset from the radio up to the level of arcmins. Here we will fix the search to the radio domain, but this should be taken into account when optical and/or non-radio sources are included in the search. In this sample the VLSS (74 MHz) and NVSS (1.4 GHz) catalogues were also looked up to find possible counterparts for the extracted WSRT sources. The WENSS (325 or 352 MHz) catalogue was excluded because it has no overlapping areas with the GRB 030329 field.

Sources far away from the pointing centre will be affected by the primary beam attenuation, which is frequency dependent. Pointing errors may in-
<table>
<thead>
<tr>
<th>Observation start date</th>
<th>Time since burst [days]</th>
<th>Time on target [hours]</th>
<th>Frequency [MHz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003–12–22</td>
<td>268.577</td>
<td>6.5</td>
<td>2300</td>
</tr>
<tr>
<td>2003–12–25</td>
<td>270.612</td>
<td>4.4</td>
<td>1400</td>
</tr>
<tr>
<td>2004–01–25</td>
<td>302.599</td>
<td>12.0</td>
<td>350</td>
</tr>
<tr>
<td>2004–01–29</td>
<td>306.408</td>
<td>3.3</td>
<td>1400</td>
</tr>
<tr>
<td>2004–01–30</td>
<td>306.753</td>
<td>3.3</td>
<td>2300</td>
</tr>
<tr>
<td>2004–02–10</td>
<td>318.555</td>
<td>12.0</td>
<td>840</td>
</tr>
<tr>
<td>2004–03–26</td>
<td>363.816</td>
<td>6.4</td>
<td>4800</td>
</tr>
<tr>
<td>2004–03–27</td>
<td>364.424</td>
<td>11.7</td>
<td>1400</td>
</tr>
<tr>
<td>2004–03–28</td>
<td>365.524</td>
<td>7.0</td>
<td>2300</td>
</tr>
<tr>
<td>2004–04–11</td>
<td>379.267</td>
<td>6.2</td>
<td>840</td>
</tr>
<tr>
<td>2004–05–02</td>
<td>400.331</td>
<td>12.0</td>
<td>350</td>
</tr>
<tr>
<td>2004–05–19</td>
<td>417.342</td>
<td>9.2</td>
<td>1400</td>
</tr>
<tr>
<td>2004–05–22</td>
<td>420.168</td>
<td>6.8</td>
<td>2300</td>
</tr>
<tr>
<td>2004–07–03</td>
<td>462.078</td>
<td>8.0</td>
<td>2300</td>
</tr>
<tr>
<td>2004–07–04</td>
<td>463.076</td>
<td>8.0</td>
<td>4800</td>
</tr>
<tr>
<td>2004–08–01</td>
<td>490.974</td>
<td>6.7</td>
<td>1400</td>
</tr>
<tr>
<td>2004–09–25</td>
<td>546.013</td>
<td>8.1</td>
<td>4800</td>
</tr>
<tr>
<td>2004–11–02</td>
<td>583.683</td>
<td>5.0</td>
<td>2300</td>
</tr>
<tr>
<td>2004–11–11</td>
<td>592.721</td>
<td>8.0</td>
<td>1400</td>
</tr>
<tr>
<td>2004–11–12</td>
<td>593.802</td>
<td>12.0</td>
<td>840</td>
</tr>
<tr>
<td>2005–03–24</td>
<td>726.439</td>
<td>12.0</td>
<td>840</td>
</tr>
<tr>
<td>2005–03–26</td>
<td>728.433</td>
<td>12.0</td>
<td>350</td>
</tr>
<tr>
<td>2005–04–09</td>
<td>742.395</td>
<td>12.0</td>
<td>1400</td>
</tr>
<tr>
<td>2005–04–10</td>
<td>743.892</td>
<td>12.0</td>
<td>2300</td>
</tr>
<tr>
<td>2005–05–14</td>
<td>777.242</td>
<td>9.2</td>
<td>4800</td>
</tr>
<tr>
<td>2005–11–27</td>
<td>974.759</td>
<td>12.0</td>
<td>4800</td>
</tr>
<tr>
<td>2005–12–07</td>
<td>984.163</td>
<td>13.8</td>
<td>2300</td>
</tr>
<tr>
<td>2005–12–09</td>
<td>986.132</td>
<td>12.5</td>
<td>1400</td>
</tr>
<tr>
<td>2006–04–09</td>
<td>1107.399</td>
<td>11.8</td>
<td>840</td>
</tr>
<tr>
<td>2006–04–30</td>
<td>1128.338</td>
<td>12.0</td>
<td>4800</td>
</tr>
<tr>
<td>2006–10–08</td>
<td>1288.899</td>
<td>12.0</td>
<td>1400</td>
</tr>
<tr>
<td>2006–10–29</td>
<td>1309.842</td>
<td>12.0</td>
<td>4800</td>
</tr>
<tr>
<td>2006–11–12</td>
<td>1323.803</td>
<td>12.0</td>
<td>840</td>
</tr>
<tr>
<td>2007–11–03</td>
<td>1679.828</td>
<td>12.0</td>
<td>4800</td>
</tr>
<tr>
<td>2007–11–18</td>
<td>1694.787</td>
<td>12.0</td>
<td>1400</td>
</tr>
<tr>
<td>2007–11–25</td>
<td>1701.768</td>
<td>12.0</td>
<td>2300</td>
</tr>
</tbody>
</table>

Table 3.1: Log of WSRT observations of GRB 030329 that are used in our sample. The indicated time on target is the difference between the logarithmic average of the start and end of the integration and the time of the burst.
3. LOFAR’s Transients and Variability Detection Algorithms

<table>
<thead>
<tr>
<th>Frequency [MHz]</th>
<th>Resolution [arcsec]</th>
<th>Sensitivity [µJy]</th>
<th>FWHM [degrees]</th>
<th>$\tau_{\text{field}}$ [arcsec]</th>
</tr>
</thead>
<tbody>
<tr>
<td>840</td>
<td>26</td>
<td>400</td>
<td>0.83</td>
<td>1400</td>
</tr>
<tr>
<td>1400</td>
<td>13</td>
<td>100</td>
<td>0.5</td>
<td>900</td>
</tr>
<tr>
<td>2300</td>
<td>7.8</td>
<td>50</td>
<td>0.37</td>
<td>660</td>
</tr>
<tr>
<td>4800</td>
<td>3.7</td>
<td>30</td>
<td>0.17</td>
<td>300</td>
</tr>
</tbody>
</table>

Table 3.2: Properties of the fields that were searched in our sample. Typical values for the spatial resolution and FWHM of the primary beam were adopted from the WSRT Guide at http://www.astron.nl.

Figure 3.6: The GRB 030329 field at 1.4 GHz as observed on 27 March 2004 by WSRT. The ellipses in the outward direction represent the search areas centred around GRB 030329 at 4.8, 2.3, 1.4 GHz, and 840 MHz, respectively. The inner ellipse resembles the search area at 4.8 GHz and has major and minor axes of 300 arcsec. As a reference, the sources that are discussed in the text are marked by the labeled circles A through H, where the white circle in the centre represents GRB 030329. The synthesised beam size is $62'' \times 20.7''$.

Introduce fluctuations of 5 to 10% in the flux of sources at the edge of the full width half maximum (FWHM) of the primary beam (T. Oosterloo, private communication). We therefore select down to only those sources that lie within the FWHM of the primary beam, of which the pointing centre is at the GRB location. Table 3.2 reports the radii of the search areas centred at the pointing centre that will be searched for variables at the corresponding frequencies. Resolutions are listed as well because sources might get resolved towards higher frequencies, causing difficulties in fitting multiple Gaussians, which may in turn lead to higher values in $V_\nu$ and $\eta_\nu$ as will be demonstrated in the next section.
Fig. 3.6 shows the field of GRB 030329 at 1.4 GHz on 27 March 2004. The dashed ellipses resemble the search areas at the frequencies 4.8, 2.3, 1.4 GHz, and 840 MHz (in order of increasing size). The pointing centres of the observations are all at RA = 10h44m50.0s and Dec = 21°31'17.8", which is 0.6 arcsec to the northeast of the VLBI position of GRB 030329, RA = 10h44m49.96s, Dec = 21°31'17.44" as reported by Taylor et al. (2007). The labeled circles mark the sources that are variable to some extent and will be discussed below.

3.4.2 Results

The values for $r$ and log $LR$ as determined in Section 3.3.2 for genuine associations obey the ideal cases. The images in the sample of the GRB 030329 observations all have different characteristics and will affect the association parameters, causing the ideal values to be too strict in some cases (see, e.g., Source B in Section 3.4.2).

Variability

The variability indices $V_\nu$ and $\eta_\nu$ of the sources are determined per frequency band, as a priori the spectral behaviour across the frequency bands is unknown. Fig. 3.7 shows the relation between the two variability indices for the sources that fall within the FWHM of the beam of the corresponding frequency bands.

Associations in the 350 MHz band were excluded in this plot, because the three observations were too few for a statistical interpretation, although this band was included in the source association routines in order to have full coverage of a source’s light curve. A single symbol corresponds to a source and all its associations that were made in the same frequency band. It does, however, not depict the number of associations that was found for the source. Due to some poor $(u, v)$ coverage images and therefore higher noise levels, a source might not get detected in the whole time series. We selected sources having at least 5, 7, 6, and 5 associations in the frequency bands of 840 MHz, 1.4, 2.3, and 4.8 GHz, respectively.

There were no single detections of sources having fluxes larger than 0.5 mJy in the different frequency bands. We focus on the sources that stand out in the variability plot of Fig. 3.7 and combine the fluxes of the associated sources across the frequency bands to get their spectral light curves. An overview of these sources is presented in Table 3.3.

The sources that are alphabetically labeled in Table 3.3 are represented by the circles in Fig. 3.6. They all showed significant variability in at least one band with $\eta_\nu$ values having probabilities of $p_\eta \leq 10^{-6}$. Sources that did show variability in just one band with $p_\eta > 10^{-6}$ were left unlabeled.
Figure 3.7: The variability indices for the sources in the GRB 030329 field at the different frequency bands of the WSRT observations. A single symbol corresponds to a source and all its associations. The number of times a source could be associated is not displayed here. GRB 030329 reveals itself by the three data points in the upper right region.

Individual Sources and Light Curves

GRB 030329. The three data points in the upper right region of Fig. 3.7 originate from the same source: GRB 030329. Fig. 3.8 shows the light curve of the GRB 030329. Because of the higher resolution in the 4.8 GHz observation in November 2005 (approximately 975 days after the burst) in combination with a slightly offset positional fit of GRB 030329 in the initial image, which is considered as the base point of a light curve, two sources could be associated to the GRB. One has a peak flux of $192 \pm 12 \mu$Jy and an integrated flux of $121 \pm 15 \mu$Jy at 50.1 arcsec from the VLBI position of GRB 030329. The other candidate has a peak flux of $162 \pm 23 \mu$Jy and an integrated flux of $149 \pm 38 \mu$Jy at a distance of 0.7 arcsec from the VLBI position of GRB 030329. The integrated flux of the latter source was reported by Van der Horst et al. (2008). As mentioned, the measured position of GRB 030329 in the 1.4 GHz images to which these sources were matched is roughly in between the two association candidates. This arises from the fact that the first positional fit was done in a lower-resolution 1.4 GHz image, resulting in a distance of nearly 30 arcsec from the VLBI position$^5$. Both association candidates’ $r (3.32$ and $2.94$, resp.) and log $LR (0.60$ and $1.1$,

$^5$The current schema that is in place avoids this first-measurement bias. A running catalogue is maintained with mean weighted positions and errors from the associated sources. In effect this is analogous to the static positions of the sources in the external catalogues.
### Table 3.3:
Sources that showed variability in the sample. The source labels correspond to the labels in Fig. 3.6. The Distance column lists the (unweighted) average distance of all the position measurements of a source to the pointing centre. \(N\) is the number of associations that could be made for a source, \(V_\nu\) the magnitude and \(\eta_\nu\) the significance of the variability, all three specified per frequency band. The corresponding plot of the variability indices of all the well-detected sources in the sample is shown in Fig. 3.7. The last column lists the weighted average source position in Right Ascension and Declination as determined by the source extraction procedures and the corresponding catalogue source name if an association could be established. No VLSS counterparts were found for these sources. The SDSS associations were found manually via the SkyServer webform. ⋆: If we include the measurements of Source A that fell outside the 840 MHz FWHM search area, the new variability parameters are: \(N = 6, V_{840 \text{ MHz}} = 0.137\) and \(\eta_{840 \text{ MHz}} = 15.5\). †: From Taylor et al. (2007).

<table>
<thead>
<tr>
<th>Source Label</th>
<th>Distance [arcsec]</th>
<th>840 MHz</th>
<th>1.4 GHz</th>
<th>2.3 GHz</th>
<th>4.8 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N) (V_\nu) (\eta_\nu)</td>
<td>(N) (V_\nu) (\eta_\nu)</td>
<td>(N) (V_\nu) (\eta_\nu)</td>
<td>(N) (V_\nu) (\eta_\nu)</td>
<td></td>
</tr>
<tr>
<td>GRB 030329</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1396</td>
<td>5(^*)</td>
<td>0.148(^*)</td>
<td>17.5(^*)</td>
<td>9</td>
</tr>
<tr>
<td>B</td>
<td>739</td>
<td>6</td>
<td>0.012</td>
<td>5.12</td>
<td>8</td>
</tr>
<tr>
<td>C</td>
<td>799</td>
<td>4</td>
<td>0.149</td>
<td>2.54</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>609</td>
<td>4</td>
<td>0.030</td>
<td>0.132</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>262</td>
<td>6</td>
<td>0.060</td>
<td>0.85</td>
<td>10</td>
</tr>
<tr>
<td>F</td>
<td>174</td>
<td>4</td>
<td>0.005</td>
<td>0.07</td>
<td>8</td>
</tr>
<tr>
<td>G</td>
<td>649</td>
<td>5</td>
<td>0.018</td>
<td>0.22</td>
<td>9</td>
</tr>
<tr>
<td>H</td>
<td>112</td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>
3. LOFAR’s Transients and Variability Detection Algorithms

Figure 3.8: The spectral light curve of GRB 030329.

resp.) values are comparable, so that neither source could be ruled out a priori as a genuine association, although the second candidate is more likely to be the counterpart. Therefore, the fluxes of both sources are plotted in GRB 030329’s light curve in Fig. 3.8.

Knowing the GRB position in hindsight with VLBI accuracy of sub-arcsec level one of the sources can be ruled out. This demonstrates the importance of reliable and up-to-date catalogues being available during observation runs, and can save further time consuming processing analysis once a catalogue counterpart is known.

Furthermore, from the large values of the variability indices, this type of transient would have been picked up by the TKP pipeline.

**Source A.** Source A is so close to the edge of the search area that one of the associations was not taken into account because its distance to the pointing centre was slightly larger than $r_{\text{field}}$ at 840 MHz. Because the search area was chosen conservatively and falls within the FWHM of the beam this data point is included in the light curve that is shown in Fig. 3.9. If this measurement is also included in the calculations of the variability indices, the reported values from Table 3.3 change to $N = 6$, $V_{\text{840 MHz}} = 0.137$ and $\eta_{\text{840 MHz}} = 15.5$. Source A falls outside the 1.4 GHz search area, and therefore only the 330 MHz and 840 MHz data points are plotted. Both probabilities of having $\eta_{\text{840 MHz}} \geq 17.5$ for five or $\eta_{\text{840 MHz}} \geq 15.5$ for six measurements is $p_\eta < 10^{-13}$, making it very plausible to reject $H_0$ for this source, even though the time sampling is sparse and the number of data points are few. Furthermore the 330 MHz data points behave similarly,
Figure 3.9: The light curve of Source A. The 1.4 GHz data are omitted due to the fact that the source lies outside the 1.4 GHz FWHM area. The first data point in the 840 MHz band was omitted in the calculations of the variability indices in Table 3.3, because its distance to the pointing centre was 1402 arcsec.

suggesting this source underwent an outburst.

The NVSS counterpart found for source A is J104444+210801 (see Table 3.3), which could be associated for all the detections in the GRB 030329 field, with an average value of the normalised association distance of $r = 4.0$. We searched the optical catalogue of the Sloan Digital Sky Survey (SDSS) and found a nearest-neighbour source, SDSS J104444.97+210801.8, in the search area ($r_s = 90$ arcsec) centred at Source A at a distance of 0.40 arcsec, but no further (spectral) information about the source is known.

Source B. From Table 3.3 and Fig. 3.7 it can be seen that Source B has the second largest value for $\eta_{1.4\,\text{GHz}}$. Its spectral light curve is shown in Fig. 3.10, where we excluded the 2.3 GHz detections because they were outside the FWHM beam; it was not detected at 4.8 GHz. It is a relatively bright source and therefore the flux variations around day 400 are large when weighted by their relatively small errors and contribute much to the sum in $\eta_{1.4\,\text{GHz}}$.

Closer visual investigation reveals that some associations were left out because of the strict selection of only those pairs having $\log LR > 2$. For these pairs the normalised distance $r$ was large due to small positional errors relative to the distance between the sources of the pair, leading to a negative value for $\log LR$ for a genuine association. If we relax the criterion (only in this case) to also selecting pairs having $\log LR > -10$, we get for the
Figure 3.10: The light curve of Source B. At 1.4 GHz, the first measurement and the one near day 600 could be included in a less strict selection criterion for log $LR$.

1.4 GHz variability indices $V_{1.4 \text{GHz}} = 0.021$ and $\eta_{1.4 \text{GHz}} = 31.27$ for $N = 10$ data points (while previously the values were $V_{1.4 \text{GHz}} = 0.020$ and $\eta_{1.4 \text{GHz}} = 32.91$, see Table 3.3). The corresponding probability is too low to accept $H_0$. The lower value for $\eta_{840 \text{MHz}}$ has a probability of $p_\eta \approx 10^{-4}$, which is too high to justify a rejection of $H_0$. Based on these results, this source shows pronounced variability in the 1.4 GHz band, whereas variability in the other bands could not be measured significantly enough.

Source C. The light curve of Source C is shown in Fig. 3.11. Since there were only four detections in the 840 MHz band, the significance of the variability is well below the threshold level to declare it variable. At 1.4 GHz the probability for the measured $\eta_{1.4 \text{GHz}} = 7.5$ is $p_\eta \approx 10^{-10}$, small enough to reject $H_0$. Source B and C are nearly diametrically opposite to each other, but do not show evidence for correlated opposite flux variations due to beam attenuation effects. Both sources did not have closeby neighbouring sources to test the primary beam attenuation effects in more detail. No reliable counterpart could be found in the SDSS Catalogue.
Figure 3.11: The light curve of Source C.

Figure 3.12: The light curve of Source D.
Source D. The light curve of Source D is shown in Fig. 3.12. This source does not show significant variability outside the 2.3 GHz band, where the probability for the $\eta_{2.3\text{GHz}}$ value is $p_\eta < 10^{-7}$. Source D lies well within the FWHM, and therefore the variations in flux of 20% from average are not likely caused by primary beam attenuations. Although the fluxes in the other bands change similarly as in the 2.3 GHz band, the significance of variability was too low for a $H_0$ rejection. There were no counterparts within the search radius of 90 arcsec in the VLSS, NVSS and SDSS Catalogues.

Source E. This source shows variability in three of the four bands in which it was detected. The flux measurement at 1.4 GHz at day 491 deviates significantly from the average (see Fig. 3.13) and causes the variability index to rise. It has to be noted that the quality of this image was quite low, presumably due to the poor ($u,v$) coverage. The probabilities for the variability indices in the 1.4 and 4.8 GHz bands are both $p_\eta < 10^{-7}$, making a rejection of $H_0$ plausible. At 2.3 GHz the probability of having $\eta_{2.3\text{GHz}}$ does not allow us to reject $H_0$. Therefore, Source E showed significant variability in two bands.

Source F. The large $\eta_\nu$ at 2.3 and 4.8 GHz is explained by the higher resolution at these frequencies, where the source is resolved into two components. This can be seen in Fig. 3.14 where the GRB030329 field is shown at comparable size for the 1.4 (left) and 4.8 GHz (right) images. The resolution of the latter is of the order of 4″, three times higher than the former.
Figure 3.14: The left panel shows the GRB 030329 field at 1.4 GHz from Fig. 3.6, but now zoomed in to the search area of the 4.8 GHz primary beam. The right panel shows the 4.8 GHz observation at 2007, November 3rd. The dashed ellipses both represent the search area of the 4.8 GHz primary beam with a radius of 300″. The resolution of 4″ at 4.8 GHz is three times higher than at 1.4 GHz, causing Source F to be resolved into two components. Source H was detected at 4.8 GHz, but not in the 1.4 GHz images.

Figure 3.15: The light curve of Source F. At 2.3 GHz and higher frequencies the source gets resolved into two components, of which only one was detected and could not be fitted well, leading to incorrect flux values.
Spreewu (2010) reports the limiting deblending capacities of the current state of the Source Extraction software, making it difficult to fit multiple Gaussians simultaneously. This causes less accuracy on position and flux fitting, giving rise to the flux variations as seen for Source F. Its spectral light curve is shown in Fig. 3.15.

The same effect is present in the 1.4 GHz flux due to the fact that it is barely resolvable at this frequency causing small positional shifts that introduce variations in the fluxes. The variability at 1.4 GHz might be due to the complexity of the source being just unresolved along with poor \((u, v)\) coverage.

It is clear that this type of sources, known to be unresolved in one and resolved in another band, should be flagged for further analysis to avoid false alerts, which is part of the work in progress.

**Source G.** Source G lies near the edge of the 2.3 GHz FWHM field. The flux variations are modest \(V_{2.3\text{GHz}} \leq 0.1\), but two measurements deviate between 8 and 10% from the average. The spectral light curve of Source G is shown in Fig. 3.16. It is not unlikely that the fluctuations for a source at the edge of the FWHM arose from beam attenuation effects in combination with the poor \((u, v)\) coverage of the early observations. The flux variations could be caused by instrumental effects. As before, the 1.4 GHz flux measurement at day 491 was influenced by the poor image quality, resulting in the same effect that caused the flux of Source E to drop. Excluding this single suspicious data point reduces the variability indices to \(V_{1.4\text{GHz}} = 0.019\) and \(\eta_{1.4\text{GHz}} = 2.20\) for \(N = 8\), which are too low to reject \(H_0\), making it not plausible to claim this source variable.

**Source H.** Source H was not detected at and below 1.4 GHz, but the higher resolutions and sensitivities at the higher frequencies make it detectable as can be seen in Fig. 3.14. The spectral light curve of Source H is shown in Fig. 3.17. At both frequencies the magnitude of the flux variability is relatively high at a significant level, giving probabilities \(p_\eta < 10^{-6}\). These values make it plausible to reject \(H_0\) and classify this source as variable. No counterparts in the major radio and the FIRST and USNO catalogues were found, but the Sloan Digital Sky Survey catalogue has an optical candidate, SDSS J104457.09+213210.1 that is at a distance of 0.61 arcsec. According to the SDSS archive\(^7\) this source has too few good detections to be classified and it has not been reported in the literature.

\(^7\)http://cas.sdss.org/astrodr7/en/tools/explore/obj.asp?id=587742014884938668
Figure 3.16: The light curve of Source G.

Figure 3.17: The light curve of Source H.
3.5 Discussion and Conclusions

The source association algorithm presented in Section 3.2.3 is based on a weighted positional difference between two sources and does not take into account any spectral source information, because there is no a priori knowledge of a source’s spectral behaviour. The likelihood ratio, as defined by Eq. 3.5, on the other hand, is frequency (and rms noise) dependent with the use of the (local) source density $n_L$. The source density is determined by the catalogue properties, but existing extragalactic source-count models (e.g., Huynh et al., 2005) may predict the number of sources down to the sub-mJy level, and should then be scaled to the observing frequencies and the expected rms noise. Eventually LOFAR will produce these source count predictions from its own catalogue by usage of the Local and Global Sky Models. A more dynamic determination of $n_L$ will further refine the criteria of the sources that are considered as an association. This is even more needed when the images are confusion limited. Restricting the number of candidate associations can also be done by a search radius $r_s$ that depends on the resolution of the observation.

Continuous monitoring of the variability indices of the extracted and then associated sources will enable LOFAR to detect transients in nearly real time. As was shown in Section 3.4.2 this should definitely include more elaborate analysis of sources that have multiple components at higher frequencies in order to limit the number of false alerts. In principle, the source association is independent of the image resolution. In practice, the current source fitting routines have difficulty in deblending and fitting multiple Gaussians, giving rise to shifted positions and incorrect fluxes for sources near the resolution limit. This also means that the ideal criteria for the association parameters ($r$ and $\log LR$) as determined in Section 3.3.2 are too strict for the practical cases where we have images produced with different beam sizes, $(u, v)$ coverage, resolution, pointing errors, etc. Again, if this is not handled properly, it may cause false alerts.

Furthermore, sources that lie in the outer regions or near the edges of the FWHM of the primary beam may vary due to pointing errors or individual antenna defects. These effects are small on average for the WSRT primary beam, but may be up to 5–10%. A way to overcome this is to do local calibrations by comparing similarities in the flux variations of neighbouring sources.

The sample of WSRT observations of GRB030329 was processed in the TKP pipeline similar in the way the LOFAR images will be. GRB030329 was detected as a transient source based on the (large) values of the variability indices. The light curve of Source A showed a burst-like rise and decay of its flux in two bands. Source H displayed significant variability in two bands. After closer inspection, Sources B, C, D, and E showed significant variability in only one band. At higher frequencies Source F was resolved...
but fitting problems caused the variations in flux to have large contributions to the variability indices. Source G turned out to be suspicious by a bad data point from a low-quality image.

Source H is a good candidate for follow-up observations by LOFAR or other telescopes. A counterpart was detected in the Sloan Digital Sky Survey, where too few good detections could not classify it. The high resolution of the HBA station of the full array will enable a detection provided that the spectral index is less than zero for this source ($S_\nu \propto \nu^{-\alpha}$). A redetection of the burst-like phenomena in Source A is possible if the spectral index is not too steep towards the higher frequencies.

The long-term WSRT observations of the sources from Table 3.3 revealed variability in about 25% of the well-detected sources in the bands at 1.4, 2.3 and 4.8 GHz; at 840 MHz this is about 10%. Although the sources and their counterparts are not classified as known types in any of the other catalogues, LOFAR observations can make a large contribution in doing so. The large fields of view of the full Dutch array (Full-NL; see Section 2.3.4 of Chapter 2) is able to reach in less than three hours of observing, resolutions of $\lesssim 10''$ and sensitivities of 4 and 0.2 mJy at 60 and 150 MHz, respectively. This mode will be very suitable to yield more spectral information about the sources.

Acknowledgements

We thank Tom Oosterloo for helpful discussions about the WSRT beam attenuation effects. The Westerbork Synthesis Radio Telescope is operated by ASTRON (Netherlands Foundation for Research in Astronomy) with support from the Netherlands Foundation for Scientific Research (NWO). This research was supported by NWO NOVA project 10.3.2.02 (BS) and by NWO VICI grant C.2320.0017 (AK) and by NWO VICI 639.043.302 (RAMJW). AJvdH was supported by an appointment to the NASA Post-doctoral Program at the MSFC, administered by Oak Ridge Associated Universities through a contract with NASA.
Chapter 4

Cross-Matching Multiple Radio Catalogues

Abstract

The calibration of LOFAR relies on flux measurements of strong radio sources with accurate positions and well-known spectra, i.e., the calibrator sources. Since the source properties are not well determined in the relatively unexplored frequency range of LOFAR (30–240 MHz), we have to interpolate and/or extrapolate the existing spectra of calibrator sources to the LOFAR frequencies. Therefore, we quantitatively cross-correlate sources from multiple catalogues to collect the spectral information in calibrator source lists.

The aim is to construct high-quality source lists with spectral information at three frequencies, 74, 325 and 1400 MHz, of which the lowest frequency falls in the LOFAR Low Band.

We quantify the likelihood of true and false associations between catalogues, and develop a criterion for accepting associations; we apply this method to cross-identify sources in the VLSS, WENSS, and NVSS catalogues.

We present a high-quality source list of nearly 2800 VLSS sources in the declination strip of $44^\circ \lesssim \delta \lesssim 64^\circ$, that are brighter than $1\ Jy\ (10\sigma_{\text{rms}})$ at 74 MHz and have unique counterparts in the WENSS and NVSS catalogues at 325 and 1400 MHz, respectively. Nearly 90% of the sources show a spectrum that turns over towards lower frequencies, and 27 sources display a peaked spectrum.

The association method can also be used in the search for transient sources, when source detections in repeated scans of the same areas of sky need to be compared with potential counterparts that are catalogued as previously detected; here also, a strict and quantitative criterion must be used to decide whether to accept that a new measurement is of the same source as previous ones.
4. Cross-Matching Multiple Radio Catalogues

4.1 Introduction

Finding which source in one catalogue is genuinely associated to which source in another catalogue is a classical problem in astronomy. Conventional procedures involve the angular distance between sources as a qualification for the association. Prior knowledge about the astrometry and resolution of the catalogues then determines an applicable threshold to be set. However, the flexibility of LOFAR allows many possible configurations of observational modes, resulting in a wide range of sensitivities and resolutions (e.g., De Vos et al., 2009; Nijboer & Pandey-Pommier, 2009). The large number of measurements of sources expected to be detected by LOFAR (see Chapter 2) will consequently produce large catalogues. Cross-matching sources between the LOFAR catalogue and other catalogues, will find either no, one or more possible counterparts. This asks for association procedures that are quantitative and by which reliable decisions can be made to accept or reject an association.

One part in the calibration process of LOFAR depends on flux measurements of strong radio sources, of which the positions and fluxes at given frequencies need to be known accurately beforehand. The large fields of view of LOFAR require many such calibrator sources. However, the low-frequency domain of LOFAR is relatively unexplored, so that the spectral behaviour of radio sources is not well covered. Therefore, we start by building an initial list of calibrator sources by quantitatively cross-correlating existing major catalogues. Sources with matching positions that are classified as a genuine association are added to this list. Calibrator fluxes at any LOFAR frequency are then given by the interpolations and/or extrapolations from this list. The list is, however, not static, but will evolve over time as new measurements made by LOFAR are added, resulting in more (calibrator) sources and a denser sky and frequency coverage. This list of sources represents the so-called Global Sky Model (GSM).

Similar source lists need to be created for the Transients Key Project. Since the lists will contain all sources detected by LOFAR, it is more appropriate to term them "LOFAR catalogue". The Transients Key Project searches for new sources to appear or known sources to change in the LOFAR sky. One way of doing that is by comparing new measurements of sources with the archived measurements in the LOFAR catalogue. This again requires source associations with quantifiable fidelity. Because of the large data volumes and high cadence of LOFAR measurements (see Chapter 2), we embedded the association procedures in the automated software pipeline of the TKP (e.g., Chapter 3 of this thesis; Swinbank et al., 2007; Swinbank, 2010).

Several methods exist that produce source lists by cross-identifying different catalogues, e.g., cross-identifications of radio sources with radio and optical catalogues (Helmboldt et al., 2008; Kimball & Ivezić, 2008), and
Table 4.1: Characteristics of the major radio catalogues, VLSS, WENSS and NVSS.

<table>
<thead>
<tr>
<th></th>
<th>VLSS</th>
<th>WENSS</th>
<th>NVSS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sky Coverage</strong></td>
<td>$\delta \geq -30^\circ$</td>
<td>$28^\circ \leq \delta \leq 76^\circ$</td>
<td>$\delta \geq 72^\circ$</td>
</tr>
<tr>
<td><strong>Frequency [MHz]</strong></td>
<td>73.8</td>
<td>325</td>
<td>352</td>
</tr>
<tr>
<td><strong># Sources</strong></td>
<td>68,311</td>
<td>211,225</td>
<td>18,185</td>
</tr>
<tr>
<td><strong>Resolution</strong></td>
<td>80$''$</td>
<td>$54'' \times 54'' \csc \delta$</td>
<td>45$''$</td>
</tr>
<tr>
<td><strong>Sensitivity [mJy]</strong></td>
<td>100</td>
<td>3.6</td>
<td>0.45</td>
</tr>
<tr>
<td><strong>Astrometry [arcsec]</strong></td>
<td>$3'' - 10''$</td>
<td>$1.5'' - 5''$</td>
<td>$1'' - 7''$</td>
</tr>
<tr>
<td><strong>Source Density [deg$^{-2}$]</strong></td>
<td>2</td>
<td>22</td>
<td>55</td>
</tr>
</tbody>
</table>

of X-ray sources with optical and near-infrared catalogues (Pierre et al., 2007; Rutledge et al., 2000; Haakonsen & Rutledge, 2009). We adopt the quantitative cross-identification analysis of Rutledge et al. (2000) to produce high-quality source lists.

To have source lists with overlap with the LOFAR frequencies, we used the sources that were detected in three major radio surveys, at 74 MHz the VLA Low-Frequency Sky Survey (VLSS), at 325 or 352 MHz the Westerbork Northern Sky Survey (WENSS) and at 1.4 GHz the NRAO VLA Sky Survey (NVSS). Table 4.1 shows the main characteristics of the radio catalogues used in our samples.

The VLSS (Cohen et al., 2007) surveyed the sky north of $\delta = -30^\circ$ at 73.8 MHz. The resolution is 80$''$ and the limiting point source flux density is 0.7 Jy, with typical rms noise levels at about 0.1 Jy. Positional uncertainties are of the order of $3'' - 10''$. The latest version (2007-06-26) contains 68,311 sources.

The WENSS catalogue (Rengelink et al., 1997) contains 229,420 sources of which 18,186 are from the polar catalogue ($\delta \geq 72^\circ$) at 352 MHz and the rest from the main catalogue ($28^\circ \leq \delta \leq 76^\circ$) at 325 MHz. Ten sources in the catalogue are without flux values and they were not taken into account. The resolution of the survey is $54'' \times 54'' \csc \delta$, and for both parts the limiting flux density is 18 mJy (5$\sigma$). The positional accuracy depends on the ratio of flux density and local rms and is about 5.5$''$ for the faint and 1.5$''$ for the stronger sources. The positional uncertainties of the sources were calculated using Eq. 8 in Rengelink et al. (1997).

The NVSS catalogue (Condon et al., 1998) covers the sky north of declination $-40^\circ$ at 1.4 GHz. It contains 1,773,484 sources with a limiting flux

---

1 We extracted the catalogues from the VizieR Astronomical Server at http://cdsarc.u-strasbg.fr/viz-bin/VizieR and loaded them into the database that is implemented in the Transients Key Project software pipeline. The source code of the latter is stored in the password-protected svn repository, available at http://svn.transientskp.org/code.
density of about 2.5 mJy (5σ). The resolution is 45″ and rms positional uncertainties vary from less than 1″ for sources stronger than 15 mJy to 7″ for the faintest ones.

The organisation of this Chapter is as follows. Section 4.2 describes the three association parameters, which are the properties of every source association pair. In Section 4.3 we apply the method of Rutledge et al. (2000) to cross-correlate the WENSS and NVSS catalogues, resulting in a high-quality list of reliable WENSS–NVSS associations. In Section 4.4 we include the VLSS catalogue as well, to construct a high-quality source list of nearly 2,800 sources with flux measurements at 74, 352 and 1400 MHz. We discuss the results in the last section.

4.2 Association Parameters

We repeat the three parameters that are defined for every source association pair (see Chapter 3): The first parameter is the angular distance between the two sources

$$\theta = 2 \arcsin \left( \frac{1}{2} \sqrt{ (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 } \right),$$

where $\Delta x, \Delta y, \Delta z$ are the Cartesian coordinate differences (on a unit sphere) of the sources. The second parameter is the weighted dimensionless distance

$$r = \sqrt{ \left( \frac{(\Delta \alpha)^2}{\sigma_{\Delta \alpha}^2} + \frac{(\Delta \delta)^2}{\sigma_{\Delta \delta}^2} \right) },$$

with $\Delta \alpha = \alpha_i \cos \delta_i - \alpha_j \cos \delta_j$ the measured right ascension difference corrected for declination, and $\Delta \delta$ the measured declination difference. Both are weighted by their positional errors summed in quadrature. A third parameter is based on the ratio of the probability of finding a true association at $r$ and an association by chance due to a background source that happens to be at $r$ from the source. This is the likelihood ratio

$$LR = \frac{\exp\left(-r^2/2\right)}{2 \pi \sigma_{\Delta \alpha} \sigma_{\Delta \delta} n_L(\nu, S_{\text{lim}})},$$

where $n_L$ is the local source density, as function of frequency and flux. Because we only associate sources by positional matches, we consider it as constant here and adopt the average NVSS source density. A more convenient way of expressing the likelihood is in logarithmic form: $\Lambda \equiv \log LR$. 

84
4.3 Reliability of WENSS–NVSS Associations

4.3.1 Source Selection Criteria

From the WENSS catalogue we selected only point sources that were classified as type ‘S’ sources (meaning they could be fitted with a single Gaussian component; Rengelink et al., 1997), and were not flagged as having caused fitting problems. A further selection was set by the declination. We included sources falling in the declination strip of $44^\circ \leq \delta \leq 64^\circ$, as they are all from the main catalogue part of the WENSS catalogue at the single frequency of 325 MHz (see Table 4.1). This sky coverage is identical to the Zenith Monitor mode of LOFAR (see Chapter 2). We join this set of WENSS sources with the NVSS catalogue to search for cross-identifications.

We adopt the quantitative cross-identification method of Rutledge et al. (2000) and apply it as follows. We define a Source Field of radius 90″ around every WENSS source. For every WENSS source, we consider all NVSS sources falling within an angular distance of 90″ as an association candidate and we calculate its likelihood ratio, $\Lambda$. These candidates, in a Source Field, include true and false associations. At the same time, we positionally offset every WENSS source to eight locations, surrounding the original WENSS source to mimic the local source density. We define eight Background Fields centred at the offset locations, all with identical radii as the Source Field. The non-overlapping Background Fields surround the Source Field on a rectangular $3 \times 3$ grid with spacings of 180″ to the neighbouring Fields. A copy of the WENSS source is then placed at the centre of every Background Field. We exclude Background Fields that themselves contain a WENSS source, since we assume that Background Fields do not contain an NVSS source that could be associated with a WENSS source. These offset sources in the Background Fields were then processed analo-

<table>
<thead>
<tr>
<th>$N_{\text{field}}$</th>
<th>$N_{\text{assoc}}$</th>
<th>Source Field</th>
<th>BG Field</th>
<th>All BG Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6,397</td>
<td>67,330</td>
<td>538,637</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>64,504</td>
<td>4,235</td>
<td>33,880</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3,330</td>
<td>158</td>
<td>1,263</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>82</td>
<td>7</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$&gt; 0$</td>
<td>71,414</td>
<td>4,573</td>
<td>36,580</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: The number of Source Fields and Background Fields, $N_{\text{field}}$, that were used for the WENSS–NVSS associations at $44^\circ \leq \delta \leq 64^\circ$. 20,675 Background Fields contained a WENSS source themselves and were excluded. $N_{\text{assoc}}$ is the number of associations that a unique source has with an NVSS source. It can be seen that 6,397 WENSS sources did not have an NVSS source within 90″.
4. Cross-Matching Multiple Radio Catalogues

gously to the original WENSS sources. NVSS counterparts found for the offset sources are regarded as purely false. Statistical analysis of the number counts then determines the likelihood regime in which associations are considered as genuine. We derive the probability for the association being unique.

Table 4.2 shows the total number of Fields used in the sample and the number counts for NVSS associations made in Source and Background Fields.

4.3.2 Parameter Evaluation

We divide the Source and Background Fields into $N_{\text{rings}}$ rings of equal areas: we draw $N_{\text{rings}}$ concentric circles around the Field centre, with the radii of the circles defined as $\rho_i^2 = i \rho_1^2$, $i = 1, 2, \ldots, N_{\text{rings}}$, with $\rho_1 = 10''$. If we then count the number of NVSS sources falling in circular areas between two adjacent circles, $\rho_i$ and $\rho_{i+1}$, we would expect the source counts to approach the NVSS source density in the outer rings of the Source Fields. In Background Fields, where we assume associations by chance, we expect the source counts to fluctuate around the NVSS source density. This is shown in Fig. 4.1. Around $i = 20$, i.e. $\rho \approx 45''$, no distinction between a true or false association with an NVSS source can be made anymore. The deficit of NVSS sources in Source Fields in the more outer rings ($i > 28$) is explained by the fact that most NVSS sources in Source Fields are true

![Source counts in equal areas of Source (red line) and Background (blue dashed line) Fields. The x-axis shows the ring number $i$, or $\rho_i^2/100$, and the y-axis the number of NVSS sources falling in ring $i$. The source counts in the outer rings of the Source Fields approach the average NVSS source density, which is about 52 deg$^{-2}$.](image)
4.3 Reliability of WENSS–NVSS Associations

The positional uncertainties of bright WENSS sources \((S > 10\sigma_{\text{rms}})\) do not have a Gaussian distribution, but are set fixed to \(1.5''\) (Rengelink et al., 1997), whereas the faint WENSS sources do have a Gaussian distribution with typical uncertainties of \(5.5''\). Therefore, the normalised distribution of the dimensionless distance, \(r\), of the source association pairs does not strictly follow a Rayleigh distribution, but is biased towards the bright WENSS–NVSS associations. This can be recognised in Fig. 4.2, where bright WENSS–NVSS (true plus false) associations in the Source Fields tend to have larger \(r\) values, and therefore contribute more towards larger \(r\), which explains the excess of associations in the region of \(12 \leq r \leq 28\). Furthermore, from the Source Field distribution in Fig. 4.2 it can be seen that the positional uncertainties of the bright WENSS sources were set conservatively. Similar effects are seen in the normalised distribution of the logarithmic likelihood ratio, \(\Lambda\), which is displayed in Fig. 4.3 for Source and Background Fields.

Following Rutledge et al. (2000) and Haakonsen & Rutledge (2009), and setting \(\Lambda_{ij} = \log LR_{ij}\), the reliability of a WENSS–NVSS association pair is given by

\[
R_{ij} = R(\Lambda_{ij}) = \frac{N_{\text{true}}(\Lambda_{ij})}{N_{\text{true}}(\Lambda_{ij}) + N_{\text{false}}(\Lambda_{ij})},
\]

where \(N_{\text{true}}(\Lambda_{ij})\) is the number of true associations for the given \(\Lambda_{ij}\) value in
4. Cross-Matching Multiple Radio Catalogues

Figure 4.3: The normalised distribution of the logarithmic likelihood ratio, $\Lambda$, of the WENSS–NVSS associations in Source (red line) and Background (blue dashed line) Fields. The binwidth is $d\Lambda = 2$.

A Source Field, and $N_{\text{false}}(\Lambda_{ij})$ is the number of false associations in a Source Field for the given $\Lambda_{ij}$ value. $R_{ij}$ is the probability that a WENSS–NVSS association with likelihood value $\Lambda_{ij}$ is genuine and not false by a mere coincidental background source. We cannot discriminate a priori between true and false counterparts, but we can approximate $R$ by using the total number counts from Source ($N_{\text{SF}}$) and Background ($N_{\text{BG}}$) Fields by setting $N_{\text{SF}}(\Lambda) = N_{\text{true}}(\Lambda) + N_{\text{false}}(\Lambda)$ and $N_{\text{BG}}(\Lambda) = N_{\text{false}}(\Lambda)$. Substituting this into Eq. 4.4 gives

$$R = R(\Lambda) = 1 - \frac{N_{\text{BG}}(\Lambda)}{N_{\text{SF}}(\Lambda)} \tag{4.5}$$

The reliability is now approximated by the ratio of the normalised number counts for Source and Background Fields. We define a cutoff value in $\Lambda_c$ below which the reliability is set to zero, $R(\Lambda < \Lambda_c) = 0$. The largest value of $\Lambda$ where $R(\Lambda) \leq 0$ is set as the cutoff value, which in our sample is at $\Lambda_c = -7.1$. Associations having $\Lambda < \Lambda_c$ are considered unlikely, and are not taken into account. Fig. 4.4 shows the reliability of WENSS–NVSS associations as function of $\Lambda$.

At $\Lambda_c \geq 1.9$, i.e. $R \geq 0.99$ we only have two WENSS sources that have two NVSS counterparts, and 58,330 WENSS sources having a single NVSS counterpart. In case of no multiple NVSS associations for a WENSS source, the reliability $R$ is at the same time the probability of the association being unique (Rutledge et al., 2000) and from this a high-quality source list can be constructed.
<table>
<thead>
<tr>
<th>RA</th>
<th>Dec</th>
<th>WENSS</th>
<th>NVSS</th>
<th>θ [arcsec]</th>
<th>r</th>
<th>Λ</th>
<th>1 − R</th>
<th>$\alpha_{W-N}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:02:35.988</td>
<td>+63:42:04.336</td>
<td>B0000.0+6325</td>
<td>J000236+634204</td>
<td>5.211</td>
<td>3.216</td>
<td>1.925</td>
<td>0.012</td>
<td>0.992±0.030</td>
</tr>
<tr>
<td>00:02:35.661</td>
<td>+52:55:49.652</td>
<td>B0000.0+5239</td>
<td>J000235+525548</td>
<td>4.234</td>
<td>2.558</td>
<td>2.746</td>
<td>0.002</td>
<td>0.711±0.036</td>
</tr>
<tr>
<td>00:02:36.578</td>
<td>+60:32:20.640</td>
<td>B0000.0+6015</td>
<td>J000237+603220</td>
<td>8.279</td>
<td>1.363</td>
<td>2.595</td>
<td>0.003</td>
<td>1.240±0.108</td>
</tr>
<tr>
<td>00:02:36.845</td>
<td>+50:22:20.242</td>
<td>B0000.0+5005</td>
<td>J000236+502220</td>
<td>6.784</td>
<td>0.915</td>
<td>2.631</td>
<td>0.003</td>
<td>−0.295±0.120</td>
</tr>
<tr>
<td>00:02:37.149</td>
<td>+61:02:36.146</td>
<td>B0000.0+6045</td>
<td>J000237+610236</td>
<td>10.454</td>
<td>1.447</td>
<td>2.453</td>
<td>0.004</td>
<td>0.431±0.081</td>
</tr>
<tr>
<td>00:02:37.753</td>
<td>+45:05:53.649</td>
<td>B0000.0+4449</td>
<td>J000237+450554</td>
<td>10.982</td>
<td>1.682</td>
<td>2.398</td>
<td>0.004</td>
<td>0.405±0.107</td>
</tr>
<tr>
<td>00:02:39.244</td>
<td>+63:13:51.781</td>
<td>B0000.0+6257</td>
<td>J000239+631351</td>
<td>2.863</td>
<td>0.552</td>
<td>3.128</td>
<td>0.001</td>
<td>0.763±0.080</td>
</tr>
<tr>
<td>00:02:38.936</td>
<td>+44:29:45.094</td>
<td>B0000.0+4413</td>
<td>J000238+442945</td>
<td>8.523</td>
<td>1.067</td>
<td>2.436</td>
<td>0.004</td>
<td>0.661±0.100</td>
</tr>
<tr>
<td>00:02:42.722</td>
<td>+46:45:08.974</td>
<td>B0000.1+4628</td>
<td>J000242+464509</td>
<td>10.258</td>
<td>1.854</td>
<td>2.252</td>
<td>0.005</td>
<td>0.002±0.068</td>
</tr>
<tr>
<td>00:02:44.435</td>
<td>+48:02:38.557</td>
<td>B0000.1+4745</td>
<td>J000244+480238</td>
<td>2.297</td>
<td>1.159</td>
<td>3.725</td>
<td>0.000</td>
<td>0.828±0.056</td>
</tr>
<tr>
<td>00:02:44.177</td>
<td>+45:09:28.466</td>
<td>B0000.1+4452</td>
<td>J000244+450928</td>
<td>4.391</td>
<td>0.861</td>
<td>2.978</td>
<td>0.001</td>
<td>0.073±0.086</td>
</tr>
<tr>
<td>00:02:50.393</td>
<td>+51:32:14.595</td>
<td>B0000.2+5115</td>
<td>J000250+513214</td>
<td>1.167</td>
<td>0.726</td>
<td>4.071</td>
<td>0.000</td>
<td>1.005±0.026</td>
</tr>
<tr>
<td>00:02:53.543</td>
<td>+62:19:20.436</td>
<td>B0000.3+6202</td>
<td>J000253+621917</td>
<td>6.085</td>
<td>3.068</td>
<td>1.965</td>
<td>0.011</td>
<td>1.345±0.043</td>
</tr>
<tr>
<td>00:02:54.689</td>
<td>+54:37:03.124</td>
<td>B0000.3+5420</td>
<td>J000254+543702</td>
<td>3.091</td>
<td>1.916</td>
<td>3.392</td>
<td>0.000</td>
<td>0.803±0.025</td>
</tr>
<tr>
<td>00:02:54.648</td>
<td>+52:58:35.308</td>
<td>B0000.3+5241</td>
<td>J000254+525835</td>
<td>3.928</td>
<td>0.612</td>
<td>2.940</td>
<td>0.002</td>
<td>0.768±0.099</td>
</tr>
<tr>
<td>00:02:55.225</td>
<td>+53:22:58.622</td>
<td>B0000.3+5306</td>
<td>J000255+532259</td>
<td>2.598</td>
<td>1.528</td>
<td>3.640</td>
<td>0.000</td>
<td>0.843±0.040</td>
</tr>
<tr>
<td>00:02:58.704</td>
<td>+44:03:07.187</td>
<td>B0000.4+4346</td>
<td>J000258+440306</td>
<td>5.151</td>
<td>0.703</td>
<td>2.763</td>
<td>0.002</td>
<td>1.050±0.107</td>
</tr>
<tr>
<td>00:02:59.960</td>
<td>+44:56:57.856</td>
<td>B0000.4+4440</td>
<td>J000259+445657</td>
<td>3.517</td>
<td>2.033</td>
<td>3.221</td>
<td>0.001</td>
<td>0.580±0.046</td>
</tr>
</tbody>
</table>

Table 4.3: A subset of the WENSS–NVSS association source list. RA and Dec are the weighted averaged coordinates of the associated WENSS (column three) and NVSS (column four) sources. The association parameters $\theta$, $r$ and $\Lambda$ are given for every pair, as well as $1 − R$, the probability of the association being caused by a coincidental background source. The calculated spectral index, $\alpha_{W-N}$, is given with its 1σ error.
4. Cross-Matching Multiple Radio Catalogues

Figure 4.4: The reliability $R(\Lambda)$ for WENSS–NVSS associations having a log-arithmic likelihood ratio value of $\Lambda$. $R$ is the reliability of an association being genuine and not coincidental, caused by a confusing background source. The bin-width is $d\Lambda = 0.1$. The cut-off is at $\Lambda_c = -7.1$. Below this value, associations are considered as unlikely. Depending on the quality of source lists, other cut-off values are at $\Lambda_c = -3.5$, where $R > 0.5$; $\Lambda_c = 0.2$, where $R > 0.9$; $\Lambda_c = 1.6$, where $R > 0.98$; $\Lambda_c = 1.9$, where $R > 0.99$.

According to a spectrum scaling law of $S_\nu \propto \nu^{-\alpha}$ we can determine the spectral index of the association pair as

$$\alpha_{W-N} = -\frac{\log(S_W/S_N)}{\log(\nu_W/\nu_N)}, \quad (4.6)$$

where $S_W$, $S_N$, $\nu_W$ and $\nu_N$ refer to the WENSS and NVSS fluxes and frequencies, respectively. The error on the spectral index is

$$\sigma_\alpha = c \sqrt{\frac{\sigma_{S_W}^2}{S_W^2} + \frac{\sigma_{S_N}^2}{S_N^2}}, \quad (4.7)$$

with $c = \log(\nu_N/\nu_W)$ and $\sigma_{S_W}, \sigma_{S_N}$ are the WENSS and NVSS flux errors.

A small selection of the source list of WENSS–NVSS associations with $R \geq 0.99$ is shown in Table 4.3. This list is implemented in the database system used by the Transients Key Project pipeline (Swinbank et al., 2007), and will become available online.
We selected all VLSS and NVSS sources in the same declination strip as in Section 4.3.1, at $44^\circ \leq \delta \leq 64^\circ$, and all catalogued WENSS point sources for which a single Gaussian component could be fitted without problems (Rengelink et al., 1997). We accepted only counterparts having a dimensionless distance $r \leq r_{\text{lim}}$, where we set $r_{\text{lim}} = 3.717$. This corresponds to accepting that 0.1% of the genuine source associations will be missed. We further reduced the lists by selecting those sources having $\Lambda \leq \Lambda_c$, with the likelihood ratio cutoff value as determined in Section 4.3.2, $\Lambda_c = -7.1$; additionally we specified a lower limit to the VLSS fluxes, $S_{\text{lim}}$. In the resulting source list we adopted $S_{\text{lim}} = 1 \text{ Jy}$, which corresponds to $10\sigma_{\text{rms}}$ sources.

We processed the cross-correlation as follows. First, we loaded all sources from the NVSS catalogue that satisfied the above-mentioned requirements into a so-called running catalogue. This was done to have the catalogue with the highest density as a first reference. Then the WENSS sources were loaded. Positional matches of WENSS sources and sources in the running catalogue (i.e. NVSS sources) obeying $r \leq r_{\text{lim}}$ were considered as an association. The corresponding parameters, $\theta$, $r$ and $\Lambda$ for every pair were calculated and a weighted average position was maintained in the running catalogue. Lastly, the VLSS sources with $S \geq S_{\text{lim}}$ were loaded and searched

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure4_5.png}
\caption{Dual spectral indices plot for the 44 sources that have a counterpart in the three catalogues of VLSS, WENSS and NVSS and of which the VLSS source is brighter than 10 Jy.}
\end{figure}

91
for counterparts in the running catalogue, which was then updated with the new averaged values for associations found.

Since we explore cross-associations between three catalogues, we calculate the spectral indices between two catalogues that are at adjacent frequencies, analogous to the derivations in Section 4.3.2. Plots of low-frequency spectral indices \( \alpha_{V-W} \) (between VLSS and WENSS pairs) vs. high-frequency spectral indices \( \alpha_{W-N} \) (between WENSS and NVSS pairs) give insight into the spectral behaviour of sources. Such dual spectral indices plots are shown in Figs. 4.5, 4.6 and 4.7.

For demonstration purposes, Fig. 4.5 shows the dual spectral indices plot for 44 unique VLSS–WENSS–NVSS associations of which the VLSS source is brighter than 10 Jy \( \sim 100 \sigma_{\text{rms}} \). For the final source list, we choose the VLSS lower flux limit at \( S_{\text{lim}} = 1 \text{Jy} \), to have it at about 10 times the rms noise. Fig. 4.6 shows the corresponding dual spectral indices plot for the 2,791 source associations that meet these requirements. A zoomed-in histogram of the latter source list shows the counts per spectral index cell and is given in Fig. 4.7. Histograms of the two spectral-index distributions are shown in Fig. 4.8, from which it can be seen that the average spectrum turns over towards lower frequencies. Nearly 90% of the sources do show this. The (arithmetic) average spectral indices are \( \langle \alpha_{V-W} \rangle = 0.721 \) and \( \langle \alpha_{W-N} \rangle = 0.923 \). However, Fig. 4.8 is biased against the larger values of the high-

\[ \alpha_{V-W} (74-325 \text{ MHz}) \]

\[ \alpha_{W-N} (325-1400 \text{ MHz}) \]

**Figure 4.6:** Scatter plot of dual spectral indices for the 2,791 sources having a counterpart in all three catalogues of VLSS, WENSS and NVSS and where the VLSS source is brighter than 1 Jy. The fraction of sources having a peaked spectrum resides in the upper left quadrant.
4.4 Cross-Matching Bright VLSS Sources with WENSS and NVSS Sources

**Figure 4.7:** Histogram of the dual spectral indices for the 2,791 sources having a counterpart in the three catalogues of VLSS, WENSS and NVSS and where the VLSS source is brighter than 1 Jy.

**Figure 4.8:** Distribution of the low- ($\alpha_{V-W}$) and high-frequency ($\alpha_{W-N}$) spectral indices of the 2,791 sources that have counterparts in the three catalogues of VLSS, WENSS and NVSS of which the corresponding VLSS source is brighter than 1 Jy. The arithmetic average spectral indices are $\overline{\alpha_{V-W}} = 0.721$ and $\overline{\alpha_{W-N}} = 0.923$. 
4. Cross-Matching Multiple Radio Catalogues

<table>
<thead>
<tr>
<th>Catalogue Name</th>
<th>RA</th>
<th>Dec</th>
<th>$S_\nu$</th>
<th>$\Lambda$</th>
<th>$\alpha_V-W$</th>
<th>$\alpha_W-N$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLSS 0000.4+4409</td>
<td>00:00:24.995</td>
<td>44:09:18.251</td>
<td>1.148</td>
<td>0.860</td>
<td>0.925</td>
<td>5.75</td>
<td></td>
</tr>
<tr>
<td>WNH B2357.8+4352</td>
<td>00:00:28.869</td>
<td>46:54:42.511</td>
<td>1.422</td>
<td>1.583</td>
<td>1.54</td>
<td>2.42</td>
<td></td>
</tr>
<tr>
<td>NVSS J000025+440918</td>
<td>00:00:28.869</td>
<td>46:54:42.511</td>
<td>0.06</td>
<td>0.00</td>
<td>0.32</td>
<td>4.80</td>
<td></td>
</tr>
<tr>
<td>VLSS 0000.4+4654</td>
<td>00:00:28.869</td>
<td>46:54:42.511</td>
<td>1.68</td>
<td>0.20</td>
<td>0.48</td>
<td>3.21</td>
<td></td>
</tr>
<tr>
<td>WNH B2357.9+4637</td>
<td>00:00:28.869</td>
<td>46:54:42.511</td>
<td>0.20</td>
<td>0.01</td>
<td>1.05</td>
<td>3.90</td>
<td></td>
</tr>
<tr>
<td>NVSS J000028+465443</td>
<td>00:00:28.869</td>
<td>46:54:42.511</td>
<td>0.02</td>
<td>0.00</td>
<td>0.59</td>
<td>4.37</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.4:** A subset of the source list of the 2,791 VLSS sources brighter than 1 Jy, that have counterparts in both the WENSS and NVSS catalogues. RA and Dec are the weighted average positional coordinates. The calculated two spectral indices for the associations are given by a low-frequency index, $\alpha_V-W$ (73.8–325 MHz) and a high-frequency index, $\alpha_W-N$ (325–1400 MHz). $\chi^2$ is the least squares of the fitted value. $\alpha_{fit}$ is the fitted index assuming a straight spectrum, whereas $\chi^2$ is the least squares of the fitted value. The names of the counterparts are given together with the rounded flux values from their original catalogues.
4.5 Discussion and Conclusions

We are able to construct high-quality source lists containing spectral information by evaluating the association parameters (Section 4.2) and applying the quantitative cross-association methods developed by Rutledge et al. (2000). Such source lists serve as an initial Global Sky Model for the LOFAR calibration. The source lists are easily fed into the LOFAR calibration and imaging pipelines, due to the integration of the database system that contains the major catalogues from which the source lists are created.

The likelihood ratio cutoff value, $\Lambda_c$, that was determined in Section 4.3.2, is specific to the sample of WENSS–NVSS catalogues. Extensions with the VLSS catalogue were justified, because the local source density, $n_L$, was the same, i.e. the NVSS source density.

The Transients Key Project will maintain a catalogue of all sources detected by LOFAR. The LOFAR catalogue initially comprises the same sources as extracted from the cross-associated catalogues. New measurements of sources are quantitatively associated to the known sources in the LOFAR catalogue, using the same method in order to notice flux changes or detect transient events. The LOFAR catalogue will evolve over time, adding new sources and enhancing the light curve data points of known sources.

In our sample, the majority of the VLSS–WENSS–NVSS cross-associations were unique. If we extend the list of catalogues and are dealing with catalogues of significantly different resolutions, e.g., the optical catalogue of SDSS with $\sim 1''$ resolution, we have to take into account the fact that multiple association candidates will appear in the field of source. In such a field with many possible counterparts, the reliabilities $R_{ij}$ for all the association
pairs will then give the probability for each counterpart being unique or that none can be associated (Rutledge et al., 2000).

From the list of associated sources, we determined the low- and high-frequency spectral indices, which reveal that 90% of the sources have a spectrum that turns over towards lower frequencies, displaying the onset of synchrotron self-absorption. The source list contains 27 sources that have peaked spectra.
Chapter 5

A New Perspective on GCRT J1745–3009


Astronomy and Astrophysics, 502, 549 (2009)

Abstract

Reports on a transient source about 1.25° south of the Galactic Centre motivated these follow-up observations with the WSRT and the reinvestigation of archival VLA data. The source GCRT J1745–3009 was detected during a 2002 Galactic Centre monitoring programme with the VLA at 92 cm by five powerful 10-min bursts with a 77-min recurrence while apparently lacking any interburst emission. The WSRT observations were performed and archival VLA data reduced to detect GCRT J1745–3009 again at different epochs and frequencies, to constrain its distance, and to determine its nature. We attempted to extract a more accurate lightcurve from the discovery dataset of GCRT J1745–3009 to rule out some of the models that have been suggested. We also investigated the transient behaviour of a nearby source. The WSRT data were taken in the “maxi-short” configuration, using 10 s integrations, on 2005 March 24 at 92 cm and on 2005 May 14/15 at 21 cm. Five of the six VLA observations we reduced are the oldest of this field in this band. GCRT J1745–3009 was not redetected. With the WSRT we reached an rms sensitivity of 0.21 mJy beam$^{-1}$ at 21 cm and 3.7 mJy beam$^{-1}$ at 92 cm. Reanalysis of the discovery observation data resulted in a more accurate and more complete lightcurve. The five bursts appear to have the same shape: a steep rise, a more gradual brightening, and a steep decay. We found variations in burst duration of order $\simeq$ 3%. We improved the accuracy of the recurrence period of the bursts by an order of magnitude: 77.012 ± 0.021 min. We found no evidence of aperiodicity. We
derived a very steep spectral index: $\alpha = -6.5 \pm 3.4$. We improved the 5$\sigma$ upper limits for interburst emission and fractional circular polarisation to 31 mJy beam$^{-1}$ and 8%, respectively. Any transient behaviour of a nearby source could not be established. Models that predict symmetric bursts can be ruled out, but rotating systems are favoured, because their periodicity is precise. Scattering constraints imply that GCRT J1745–3009 cannot be located far beyond the GC. If this source is an incoherent emitter and not moving at a relativistic velocity, it must be closer than 14 pc.

### 5.1 Introduction

Reports of a peculiar radio transient, GCRT J1745–3009, about 1.25$^\circ$ south of the Galactic Centre (Hyman et al., 2005, 2006, 2007) and the suggestion that this may be the prototype of a new class of particularly bright, coherently emitting radio transients have led to speculation about its nature. In particular, the 77 minute recurrence of the Jy level bursts was attributed to a period of rotation (Zhang & Gil, 2005), revolution (Turolla et al., 2005) and precession (Zhu & Xu, 2006). A nulling pulsar and an ‘X-ray quiet, radio-loud’ X-ray binary have also been suggested (Kulkarni & Phinney, 2005), as well as an exoplanet and a flaring brown dwarf (Hyman et al., 2005). The discovery has led to follow-up observations and re-examination of archival data at both 92 cm and other bands. Those did not reveal a source (Zhu & Xu, 2006; Hyman et al., 2005, 2006), with two exceptions (Hyman et al., 2006, 2007). Both of the redetections were single bursts, possibly due to the sparse sampling of these observations. The first redetection was possibly the decaying part of a bright (0.5 Jy level) burst that was detected at the first two minutes of a ten minute scan. The second redetection was a faint short (≃2 minute) burst that was completely covered by the observation. The average flux density during the burst was only $57.9 \pm 6.6$ mJy/beam. This redetection also showed evidence for a very steep spectral index ($\alpha = -13.5 \pm 3.0$).

The source has only been detected at three epochs, separated by less than 18 months, all at 92 cm, while the source was not detected in this band at 33 epochs over a period of more than 16 years (see Hyman et al., 2006, Table 1) nor in any other band, ever. We observed the field containing GCRT J1745–3009 using eight 10-MHz IFs in the 92 cm band because its possible association with the supernova remnant G359.1-0.5 would mean that this source is about as far as the Galactic Center. That, in turn, implies a substantial dispersion measure (DM) that will become apparent as a delay of several seconds between the highest frequency IF and the lowest. This would be measurable if the bursts had some sufficiently sharp feature. An observation at 21 cm was performed to make use of the lower Galactic confusion and high sensitivity of the WSRT. We reanalysed five
archival VLA datasets taken between 1986 and 1989 and the 2002 discovery dataset. All of these except the last were pointed at SgrA. Two of them, both obtained in A-configuration, had not been imaged before with the proper three-dimensional image restoration techniques. The complete set of observations we reduced is specified in Table 5.1.

5.2 Data Reduction

5.2.1 General

We used AIPS (Greisen, 2003) for the reduction of all datasets.

5.2.2 The 92 cm WSRT Observations on 2005 March 24

The WSRT 92 cm observations on 2005 March 24 started at UT 01:22 with the observation of the calibration source 3C 295. We acquired data from the target field from 02:33 until 07:50 using 10s integrations, with eight 10-MHz IFs, consisting of 128 channels, each 78.125 kHz wide, separated 8.75 MHz from each other and centered on frequencies ranging from 315.4 to 376.6 MHz. RFI was excised from the spectral line data using the AIPS task SPFLG, while remaining RFI was removed from the continuum data using the AIPS task TVFLG.

Calibration was done in four steps. First we determined the variation in system temperature as a function of time (and therefore also position on the sky), using the intermittent firing of a stable noise source. Next we performed a bandpass calibration using the AIPS task BPASS. We applied the bandpass solution using the AIPS task SPLAT, producing a continuum file with one channel per IF. After that, we performed an external absolute gain calibration using an assumed flux of 61.5 Jy for 3C 295 in the lowest frequency IF, by running the AIPS tasks SETJY and CALIB. SETJY was set to use the absolute flux density calibration determined by Baars et al. (1977) and the latest (epoch 1999.2) polynomial coefficients for interpolating over frequency as determined at the VLA by NRAO staff. Finally, we self-calibrated the data for time variations in the relative complex gain phase and amplitude.

Theoretically, we should be able to reach a thermal noise level of 0.15 mJy/beam in a 5 hour integration, or at least the nominal beam confusion noise limit of 0.3 mJy/beam. However, we did not attain this sensitivity due to the limited $uv$-coverage, RFI, and the existence of bright diffuse emission in the field. The latter compromises both self-calibration and image quality. This could be remedied to some extent by excluding spacings below a certain limit ($uv_{\text{min}} >$ some multiple of $\lambda$, the wavelength). We chose a $uv_{\text{min}}$ of 1.0 $k\lambda$ to eliminate the bulk of the diffuse emission, which could not be deconvolved with the available $uv$-coverage. SgrA and Tornado are the
<table>
<thead>
<tr>
<th>No.</th>
<th>Date</th>
<th>Telescope</th>
<th>Number of antennas</th>
<th>Number of IFs</th>
<th>Number of channels per IF</th>
<th>Number of channels per IF (MHz)</th>
<th>Total BW (MHz)</th>
<th>Chan width (MHz)</th>
<th>Im duration (h)</th>
<th>Antennas (VLA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>860329</td>
<td>VLA A</td>
<td>12</td>
<td>1</td>
<td>127</td>
<td>3.1</td>
<td>98</td>
<td>2.5</td>
<td>4.6</td>
<td>VLA A</td>
</tr>
<tr>
<td>2</td>
<td>860803</td>
<td>VLA B</td>
<td>8</td>
<td>1</td>
<td>127</td>
<td>0.8</td>
<td>98</td>
<td>0.7</td>
<td>6.24</td>
<td>VLA B</td>
</tr>
<tr>
<td>3</td>
<td>860403</td>
<td>VLA C</td>
<td>15</td>
<td>1</td>
<td>63</td>
<td>0.8</td>
<td>98</td>
<td>0.7</td>
<td>3.86</td>
<td>VLA C</td>
</tr>
<tr>
<td>4</td>
<td>050324</td>
<td>WSRT</td>
<td>12</td>
<td>7</td>
<td>128</td>
<td>10.0</td>
<td>6328</td>
<td>89</td>
<td>2</td>
<td>WSRT</td>
</tr>
<tr>
<td>5</td>
<td>050514</td>
<td>WSRT</td>
<td>14</td>
<td>6</td>
<td>128</td>
<td>20.0</td>
<td>12813</td>
<td>154</td>
<td>4.6</td>
<td>WSRT</td>
</tr>
</tbody>
</table>

1. This is the number of antennas after flagging.
2. This is the number of IFs after flagging averaged over the RR and LL polarization products, if both are available.
3. This is total bandwidth for Stokes I imaging, we added RR and LL bandwidths.

Table 5.1: Specifications of these observations.
dominant sources in the field, their sidelobes contributed significantly to the
image noise level of 9.0 mJy beam$^{-1}$ at the location of GCRT J1745–3009.
These and other sources were deconvolved in an image with an asymmet-
rical cell size ($10'' \times 60''$). We chose to do so because a symmetrical cell
size would yield a very elongated synthesized beam, this would hamper the
deconvolution process. We subtracted the clean components of all sources
from the $uv$-data before imaging the residual data with a symmetrical cell
size. To lower the noise from the sidelobes of the two poorly subtracted
extended sources, this final residual image was made by imposing a more
severe lower limit of 2.5k$\lambda$ on the spacings, which resulted in a noise level of
3.7 mJy beam$^{-1}$. That final image was made from only 7% of the recorded
visibilities.

In retrospect, it is possible that the self-calibration process was adversely
affected by bandwidth smearing, particularly because SgrA and Tornado
were located far from the phase tracking center. Bandwidth smearing could
have been diminished by keeping many channels per IF in SPLAT. SgrA
and Tornado were close to the half power beam width (HPBW). This also
hammers self-calibration because the frequency dependence of the primary
beam attenuation is much stronger near the HPBW than near the pointing
center. It could have been fixed to some extent by running self-calibration
per IF, at the expense of signal to noise. These flaws, the poor uv-coverag-
the exclusion of many spacings and the Galactic plane contribution to the
system temperature explain why the achieved noise level is still well above
the thermal noise limit of 0.68 mJy beam$^{-1}$ for this number of visibilities,
imaging bandwidth and IFs (see Table 5.1), for a circular 60$''$ beam towards
cold sky.

5.2.3 The 21 cm WSRT Observations on 2005 May 14/15

The 2005 May 14/15 observations at 21 cm started at UT 22:33 with the
observation of the calibration source 3C 286. We acquired data from the
GCRT J1745–3009 field from 23:09 until 03:46 using 10 s integrations, with
eight 20-MHz IFs, separated 17 MHz from each other and centered on fre-
cuencies ranging from 1265 to 1384 MHz. The calibration was done in the
same way as for the 92 cm WSRT observation. The assumed flux for the
calibrator source 3C 286 in the lowest frequency IF was 15.6 Jy. Theoret-
ically, the rms sensitivity of these observations could be as low as about 21
$\mu$Jy beam$^{-1}$, for a 4.6 h integration. However, as for the 92 cm WSRT data,
we excluded short spacings to eliminate most of the diffuse emission, which
was necessary for successful self-calibration. The rms noise level in the final
residual image was about 210 $\mu$Jy beam$^{-1}$. That noise level is partly due
to the loss of data: the exclusion of spacings below 2.5k$\lambda$ and the excision
of RFI. The total loss of visibilities up to the final image was as high as
55%. With this number of visibilities and with the imaging bandwidth and
IFs as mentioned in Table 5.1, the theoretical thermal noise limit is $45 \, \mu\text{Jy beam}^{-1}$ for a circular $13''$ beam towards cold sky.

### 5.2.4 The 92 cm VLA Discovery Dataset of 2002 September 30 / October 1

The specifics of the 2002 discovery dataset are shown in Table 5.1. We started its reduction with the flagging of 4 of the 27 antennas. Also, we flagged individual spectral channels per baseline, per IF and per polarisation product for all or part of the observing time, using the AIPS task SPFLG. We flagged small portions, of 1 minute or more, of data at the beginning and end of each scan using the AIPS task QUACK. We also clipped data contaminated by RFI using the AIPS task CLIPM. Next, we performed an external absolute gain calibration with an assumed flux of 25.9 Jy for 3C 286 in the lowest frequency IF. This flux was determined by running the AIPS task SETJY, using the absolute flux density calibration determined by Baars et al. (1977) and the latest (epoch 1999.2) VLA polynomial coefficients for interpolating over frequency. We determined gain phase and gain amplitude solutions for both the primary calibrator 3C 286 and the phase calibrator 1711-251, using the AIPS task CALIB. This task was run using all spacings for the primary calibrator and spacings longer than $1 \, \text{kHz}\lambda$ for the phase calibrator. The AIPS task GETJY determines the flux of the secondary calibrator from those gain solutions and the flux of the primary calibrator. GETJY found a flux of 11.1 Jy for 1711-251 at the highest frequency IF (327.5 MHz). The gain solutions were interpolated using the AIPS task CLCAL.

Next, we used 3C 286 to find a bandpass solution. In doing so, we applied the interpolated gain solutions from CLCAL for spacings longer than 500 wavelengths ($uv_{\text{min}} > 0.5 \, \text{k}\lambda$). For one of the antennas no visibilities were recorded during the scan of 3C 286. Hence, no bandpass solution could be found for this antenna and only 22 antennas were left for imaging. We applied the gain and bandpass solution to 20 of the total of 31 available channels using the AIPS task SPLAT. Every two channels were averaged.

Next, we performed 18 iterations of phase only self-calibration, using initial solution intervals of 5 minutes, gradually decreasing down to 1 minute. We used 195 kHz channels for imaging and a cellsize of 4''. We used 85 $512 \times 512$ pixel facets to cover the primary beam and no facets for outlier fields. We performed an amplitude and phase self-calibration and we produced the final model from the spectral averaged dataset. After that, we reran SPLAT on the line data, but this time without spectral averaging, selecting $21 \times 97 \, \text{kHz}$ of the available channels. We phase self-calibrated the new dataset using the acquired model from the spectral averaged data. Next, we imaged and deconvolved our phase self-calibrated dataset using 61 facets to cover the primary beam and 22 facets for the outlier fields. This time we used $256 \times 256$ pixel facets with a pixel size of $10''$. We self-calibrated
Figure 5.1: The supernova remnant G359.1-0.5 with "The Snake" to the northwest, from our reduction of the GCRT J1745–3009 discovery observation on 2002 September 30/October 1 with the VLA in CnB configuration. This observation revealed this transient, indicated by a circle, for the first time (see Hyman et al., 2005). Noise levels in this image vary from 5 to 13 mJy beam\(^{-1}\) across the image. A Gaussian fit to the unresolved GCRT J1745–3009 gives a peak flux density of only 116±14 mJy beam\(^{-1}\) because the five Jy-level bursts have been averaged over about 6h of observation. A Gaussian fit to the source to the northeast of the supernova remnant, indicated by the box, gives a peak flux density of 91±14 mJy beam\(^{-1}\). Correction for primary beam attenuation has been applied.
again, but this time we solved for amplitude and phase, using a solution interval of 1 minute. The total average gain was normalized in this process. We imaged and deconvolved 450 Jy of total flux from the amplitude and phase self-calibrated dataset to make our final model. Fig. 5.1 shows the central facet of this model after correction for primary beam attenuation. We noticed that SgrA is by far the brightest source in the field and that it is near the half power beam point. We anticipated that the calibration of the \( uv \) data could be optimized by applying separate gain solutions to the clean components of the facet with SgrA, so we ran the \textsc{aips} runfile PEELR on the clean components of the facet of SgrA, solving for gain amplitudes and phases on a timescale of 10s. We subtracted the clean components from the peeled data using the \textsc{aips} task \textsc{uvsub} and we determined the position of GCRT J1745–3009 in our final model using the \textsc{aips} task \textsc{imfit}. We shifted the phase stopping centre to this position using the \textsc{aips} task \textsc{uvfix} and we averaged all spectral channels using the \textsc{aips} task \textsc{split}. We did a final edit using the \textsc{aips} task \textsc{clip} and set \( uv_{\text{min}} = 1.0 \text{k}\lambda \). We ran the \textsc{aips} task \textsc{dftppl} on this final residual dataset to produce our lightcurves. We did not correct the output of \textsc{dftppl} for primary beam attenuation because GCRT J1745–3009 was about 13' from the pointing center. Primary beam attenuation for this angular separation is only 1.8%.

In retrospect, it turned out that both the amplitude and phase (A&P) self-calibration and the peeling of SgrA had negligible effect on the burst shapes in the final lightcurves. So the dataset could be reduced in a standard way, except perhaps for the large number of selfcal iterations and the exclusion of a rather large number of antennas, 5 of the 27 antennas being excluded for the entire observation.

5.3 The Source on the Opposite Side of the Supernova Remnant

The source northeast of the supernova remnant G359.1-0.5, indicated by a box in Fig. 5.1 is resolved in VLA A configuration. From a combination of three VLA datasets, two in A configuration and one in B configuration, this source was detected with a peak flux density of \( 17.1\pm2 \text{ mJy beam}^{-1} \) and an integrated flux of 47.6 mJy (see Nord et al., 2004, Table 2, source 72). Apparently the synthesized beam of the combination of these datasets (12'' $\times$ 7'') resolves this source. As noted in the caption of Fig. 5.1, the peak flux density we derived from the 2002 discovery observation is 91\pm14 mJy beam$^{-1}$. A large fraction of the difference with the integrated flux measurement by Nord et al. (2004) is probably caused by extended emission. Indeed, when we exclude the shortest spacings, \( uv_{\text{min}} = 1.0 \text{k}\lambda \), we find a much lower peak flux density of 73\pm5 mJy beam$^{-1}$. The remaining difference may also come from extended emission that is picked up differently by
5.4 Overview of Flux Measurements of GCRT J1745–3009

We hoped to redetect GCRT J1745–3009 with the WSRT, with some of the VLA observations mentioned in the previous section and with two additional A configuration observations from the VLA archive. We did not redetect the source, but we measured its flux at its position in all of the seven maps. Specifics of these observations are shown in Table 5.1. Note that the on-source time for the two WSRT observations is comparable to the VLA observations, despite the limited time for which the WSRT can observe this low declination source. The reason for this is that the WSRT in general does not need to observe secondary calibrators. The results of the flux measurements at these epochs and at the time of the discovery are shown in Table 5.2. For the seven nondetections, we fitted the restoring beam to the position reported by Kaplan et al. (2008). We have also imaged 10 minute subsets of the residual data from the five 1986-1989 observations to look for isolated bursts, but we found none.

We merged our results from Table 5.2 with those from a recent overview of observations since 1989 (see Hyman et al., 2006, Table 1) together with the results from the second redetection (Hyman et al., 2007) to produce a plot of 5σ flux upper limits on quiescent emission from GCRT J1745–3009 in the 92 cm band (see Fig. 5.2).

In order to derive appropriate values, we scaled the 10-minute scan sensitivities mentioned (20 and 10 mJy beam$^{-1}$ for the VLA and the GMRT respectively, after correction for primary beam attenuation) with the square root of the observing bandwidth, taking 6.2 MHz as the base. The sensitivities for complete observations were also scaled with the square root of the total on-source time. We note that the 1989 March 18 observation was
### Table 5.2: Flux measurements at 92 cm (unless otherwise noted) for detections and nondetections of GCRT J1745–3009 at α = 17h45m51.15s, δ = −30°09′52.7″ (Kaplan et al., 2008).

<table>
<thead>
<tr>
<th>No.</th>
<th>Date (yymmdd)</th>
<th>Telescope</th>
<th>Peak flux density</th>
<th>Error on fit</th>
<th>rms noise</th>
<th>Resolution (′ × ′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>860329</td>
<td>VLA A</td>
<td>−0.2</td>
<td>0.02</td>
<td>0.02</td>
<td>4 × 4</td>
</tr>
<tr>
<td>2</td>
<td>860805</td>
<td>VLA B</td>
<td>2</td>
<td>0.12</td>
<td>0.12</td>
<td>110 × 10</td>
</tr>
<tr>
<td>3</td>
<td>861226</td>
<td>VLA C</td>
<td>1</td>
<td>0.10</td>
<td>0.10</td>
<td>64 × 33</td>
</tr>
<tr>
<td>4</td>
<td>881203</td>
<td>VLA A</td>
<td>12</td>
<td>0.15</td>
<td>0.15</td>
<td>27 × 18</td>
</tr>
<tr>
<td>5</td>
<td>890318</td>
<td>VLA B</td>
<td>1</td>
<td>0.10</td>
<td>0.10</td>
<td>120 × 30</td>
</tr>
<tr>
<td>6</td>
<td>020930</td>
<td>VLA CnB</td>
<td>1</td>
<td>0.10</td>
<td>0.10</td>
<td>28 × 28</td>
</tr>
<tr>
<td>7</td>
<td>050324</td>
<td>WSRT</td>
<td>1</td>
<td>0.14</td>
<td>0.14</td>
<td>44 × 27</td>
</tr>
<tr>
<td>8</td>
<td>050514</td>
<td>WSRT</td>
<td>1</td>
<td>0.12</td>
<td>0.12</td>
<td>44 × 33</td>
</tr>
</tbody>
</table>

Note: The formal rms noise levels in these two maps are 19 mJy beam$^{-1}$ and 69 mJy beam$^{-1}$ for the 1986 August 5 and December 26 observations, respectively, much lower than the indicated value of 100 mJy beam$^{-1}$. However, many bright compact sources that should be detectable in these maps are not due to the very poor uv coverage of this observation. We accounted for this by replacing the rms noise by a higher number in these two maps. Here we did not tie the clean beam fit to the position from Kaplan et al. (2008), but set the AIPS task IMFIT to solve for peak flux density.

The formal rms noise levels in these two maps are 19 mJy beam$^{-1}$ and 69 mJy beam$^{-1}$ for the 1986 August 5 and December 26 observations, respectively, much lower than the indicated value of 100 mJy beam$^{-1}$. However, many bright compact sources that should be detectable in these maps are not due to the very poor uv coverage of this observation. We accounted for this by replacing the rms noise by a higher number in these two maps. Here we did not tie the clean beam fit to the position from Kaplan et al. (2008), but set the AIPS task IMFIT to solve for peak flux density.

This is the average noise in the residual image.
already analysed by Hyman et al. (2006) and their reduction led to slightly more constraining values, so we adopted these in Fig. 5.2. Here, we took account of the fact that the total bandwidth of that observation was actually 1.4 MHz instead of the 12.5 MHz mentioned in their Table 1. Consequently, we derived $5\sigma$ upper limits of $5 \times 20 \times \sqrt{6.2/1.4} = 210 \text{ mJy beam}^{-1}$ and $5 \times 20 \times \sqrt{6.2/1.4}/\sqrt{5.3 \times 6} = 37 \text{ mJy beam}^{-1}$, for those 10-minute scans and for that complete observation, respectively.

The lowest noise level of all 92 cm observations, about 6 mJy beam$^{-1}$ in a 2 minute interval, was achieved at the time of the second redetection, with the GMRT on 2004 March 20 (Hyman et al., 2007). This is actually the only observation that could have detected bursts of this kind and only by making 2 minute scan averages. None of the observations included in Fig. 5.2 can detect the 2004 burst (Hyman et al., 2007) in 10 minute averages at the $5\sigma$ noise level.

The WSRT 2005 May 14/15 $5\sigma$ upper limit at 21cm (1.05 mJy beam$^{-1}$) was less constraining than the VLA upper limit at that wavelength on 2005 March 25 (0.4 mJy beam$^{-1}$, see Hyman et al., 2006). 21 cm observations are not included in Fig. 5.2.

Figure 5.2: Approximate detection thresholds ($5\sigma$ noise levels) at the location of GCRT J1745–3009 of 41 Galactic Center observations at 92 cm over two decades. For the WSRT observation at 92 cm, the 10 minute scan sensitivity is not indicated, since the snapshot point spread function (psf) of a linear array does not allow to do this accurately. The observations in this plot start on 1986 March 29 and end on 2005 September 27.
5.5 Reanalysis of the 2002 Discovery Dataset

5.5.1 Lightcurve

The lightcurve that we extracted from the discovery dataset of GCRT J1745–3009 at the position derived in paragraph 5.5.3 is shown in Fig. 5.3. The bursts seem to have similar shapes: a steep rise, a gradual brightening and a steep decay, more consistent than the bursts shown in Fig. 1 of the discovery paper (Hyman et al., 2005). This lightcurve is twice as accurate as the original one. We also ran the AIPS task DFTPL with 5s sampling, this is the integration time for the recording of the visibilities in the discovery dataset. We found no compelling evidence for interburst emission, not even on the shortest (5s) timescale. We determined the recurrence interval between bursts by measuring the times of steepest rise for four of the bursts. Consecutive 1 minute chunks of data were selected by a sliding window. For each chunk of data we determined its average slope by weighted linear regression. The weights come from the reciprocal of the noise variances from DFTPL. The time corresponding to the steepest positive slope was then calculated as the weighted average of the timestamps in the data chunk. For the first burst, this method is illustrated in Fig. 5.4. Weighted linear regression also calculates the error bars of the times of steepest rise from the error bars of the data points. The times of steepest rise and the corresponding error bars are shown in Table 5.3. The times mentioned in that table are relative to 20h50m00s on 2002 September 30 (IAT). We then again applied the formulae for weighted linear regression to find the period between bursts and its 1σ error. We found a period of \(77.012 \pm 0.021\) min from the values in Table 5.3. We have improved the error on the period by an order of magnitude (Hyman et al., 2006, paragraph 3 and caption of Fig. 3), but the period itself agrees with the previously determined period of \(77.1\) min \(\pm 15\)s. However, it is important to note that our method differs from the one used by Hyman et al. (2005). We have made no assumption with regard to the burst shapes in determining the period.

The residuals with respect to that fit are 0.097, −0.114, 0.053 and −0.007 minutes for the first, second, fourth and fifth burst, respectively. The residual for the second burst is the largest, 6.8 s ”too late” with respect to the fit, this corresponds to \(1.9\sigma\), \(\sigma = 0.060\) min, this is the error on the time of steepest rise of the second burst.

We were also able to measure the times of steepest decay for four of the bursts in a similar manner, see Table 5.4. For three bursts we could measure both the time of steepest decay and the time of steepest rise. In this way we found that the time between steepest rise and steepest decay varies. We found intervals of \(8.29\pm 0.08\), \(8.87\pm 0.09\) and \(8.66\pm 0.09\) min for the second, fourth and fifth burst, respectively.
5.5 Reanalysis of the 2002 Discovery Dataset

Figure 5.3: The plot above shows the lightcurve from the discovery dataset of GCRT J1745–3009 with 30s sampling. This plot is setup in the same way as the lightcurve in the discovery paper except for the flux density measurements between bursts. For those nondetections Hyman et al. (2005) showed $3\sigma$ upper limits on interburst emission, we show the actual background flux density measurements. Also, we have folded the lightcurve at intervals of 77.012 minutes instead of 77.130 minutes. The first interval is shown in the bottom panel, starting at 20h50m00s on 2002 September 30 (IAT). The average of all the error bars shown is 74 mJy. The gaps are due to phase calibrator observations.

Table 5.3: Measurements of times of steepest rise for four bursts

<table>
<thead>
<tr>
<th>Burst number</th>
<th>Time of steepest rise (min)</th>
<th>1$\sigma$ error (min)</th>
<th>Slope (Jy/min)</th>
<th>1$\sigma$ error (Jy/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60.624</td>
<td>0.068</td>
<td>0.706</td>
<td>0.158</td>
</tr>
<tr>
<td>2</td>
<td>137.848</td>
<td>0.060</td>
<td>0.828</td>
<td>0.175</td>
</tr>
<tr>
<td>4</td>
<td>291.704</td>
<td>0.065</td>
<td>0.724</td>
<td>0.138</td>
</tr>
<tr>
<td>5</td>
<td>368.776</td>
<td>0.065</td>
<td>0.743</td>
<td>0.150</td>
</tr>
</tbody>
</table>
Figure 5.4: This plot illustrates how the times of steepest rise for four of the bursts are determined. Weighted linear regression is performed on successive one minute chunks of data. The chunks have a maximum of 55s of overlap time. Here the rising part of the first burst is shown.

<table>
<thead>
<tr>
<th>Burst number</th>
<th>Time of steepest decay (min)</th>
<th>1σ error (min)</th>
<th>Slope (Jy/min)</th>
<th>1σ error (Jy/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>146.136</td>
<td>0.041</td>
<td>-1.125</td>
<td>0.146</td>
</tr>
<tr>
<td>3</td>
<td>223.489</td>
<td>0.057</td>
<td>-0.811</td>
<td>0.150</td>
</tr>
<tr>
<td>4</td>
<td>300.578</td>
<td>0.067</td>
<td>-0.717</td>
<td>0.150</td>
</tr>
<tr>
<td>5</td>
<td>377.439</td>
<td>0.066</td>
<td>-0.734</td>
<td>0.148</td>
</tr>
</tbody>
</table>
5.5 Reanalysis of the 2002 Discovery Dataset

So for the second burst the interval between steepest rise and steepest decay is 3.45% less than the weighted mean of those three intervals. The significance of this deviation is 3.0σ.

We can use the derived period to fold the bursts in one plot, see Fig. 5.5. This plot shows that the bursts indeed have similar shapes.

Figure 5.5: The plot above shows the five bursts from the discovery dataset of GCRT J1745-3009 with 30s sampling folded at intervals of 77.012 minutes. Time is relative to 20h50m00s on 2002 September 30 (IAT) (plus multiples of 77.012 minutes).
5. A New Perspective on GCRT J1745–3009

5.5.2 Implications for Other Observations

Now that we have determined the periodicity of the bursts more accurately, we can check if other short GC observations at 92 cm before and after the discovery observation should have detected GCRT J1745–3009. The observation closest in time was taken on 2002 July 21 (see Hyman et al., 2006, Table 1). This was a 59.2 minute scan starting 1719.75 hours before the start of the bright part of the first burst in the discovery dataset. This corresponds to 1339.86 periods of 77.012 minutes. Consequently, the source should not have been seen during that short scan and this was indeed the case (Hyman et al., 2006). However, there is a large uncertainty in calculating burst times over an interval as large as 71 days. The error is $0.021 \times 1339 = 28$ min.

From that uncertainty and Gaussian statistics, we calculated that the chance of having observed at least 5 minutes of bursting activity on 2002 July 21 was 74%, assuming that GCRT J1745–3009 were bursting as during the discovery observation. If GCRT J1745–3009 was indeed active on 2002 July 21, we can infer from the nondetection on that occasion that $P$, the recurrence interval between bursts is tightly constrained: $77.007 \text{ min} < P < 77.021 \text{ min}$.

The next observation closest in time was taken on 2002 June 24. Its duration was only 34.5 min, starting 1842.17 periods of 77.012 minutes before the start of the bright part of the first burst in the discovery dataset. During this observation we should have seen at least 6 minutes of a burst if we take into account the constraints on the period from the nondetection on 2002 July 21. From the fact that we did not detect emission on 2002 June 24 we may conclude that activity started after this 34.5 minute scan.

The first suitably pointed 92 cm observation after the discovery observation was taken on 2003 January 20. The source was not detected, but the data were taken with the VLA in CD configuration. This implies that rms noise levels from 10-min scans are about 250 mJy beam$^{-1}$ (see Hyman et al., 2006, and Fig. 5.2 in this paper). Thus it is likely that GCRT J1745–3009 could not have been detected at the $5\sigma$ level on 2003 January 20, even if an individual ten minute scan were spaced in time such that it completely covered a burst. It may be that the activity continued until the summer of 2003 when three 59 minute and four 34 minute GC observations were performed with the VLA in A configuration. At least two of these scans are spaced in time such that if one covered the interval between two bursts, the other must have covered a complete burst. So we are sure that the recurrent bursting activity of GCRT J1745–3009 stopped before it was redetected on 2003 September 28.

In summary, the bursting activity with a period of 77.012 minutes as seen during the discovery observation must have started after 2002 June 24 and may have continued until the summer of 2003. Unfortunately, we cannot constrain the timespan of a recurrently bursting GCRT J1745–3009 to less than a year.
5.5 Reanalysis of the 2002 Discovery Dataset

5.5.3 Position and Flux Measurements; Spectral Index Determination

The most accurate position measurement, corresponding to the highest signal to noise ratio, can be achieved by selecting just the time intervals that cover the bursts. We found a peak flux density of $900 \pm 23$ mJy beam$^{-1}$ and this J2000 position: $\alpha = 17^{h}45^{m}05.015^{s} \pm 0.045s$, $\delta = -30^\circ09'52.19'' \pm 0.52''$. This position of GCRT J1745–3009 has not yet been corrected for ionospheric-induced refraction (see Nord et al., 2004, for some background). That correction, which is basically, but not exactly, a global position shift of all sources in the field, will significantly increase the uncertainty in the position of GCRT J1745–3009. Here, we just mention that in our maps the bright source SGR E46 is 0.33s west and 0.89" north of the NVSS (Condon et al., 1998) position. The NVSS catalogue mentions a positional accuracy of 0.45" in right ascension and 0.6" in declination for this source. We consider the actual uncertainty for the given position of GCRT J1745–3009 to be 5" in both right ascension and declination.

Rms noise values in the map that constitutes our final model range between 5 and 13 mJy beam$^{-1}$. We also made a map from the same data, but without short spacings ($uv_{\text{min}} = 1.0\,\lambda$). Noise levels then drop significantly, varying between 4 and 6 mJy beam$^{-1}$ across the image. We removed the bursts and we made a cleaned image with the same spacings. The noise levels are somewhat higher now: between 5 and 7 mJy beam$^{-1}$.

In order to derive an upper limit on interburst emission we fitted the clean beam to the position measured above. We found a peak flux density of $-0.6 \pm 6.4$ mJy beam$^{-1}$, after correction for primary beam attenuation (1.8%). This gives a 5\(\sigma\) upper limit on interburst emission of 31 mJy beam$^{-1}$. This is more than twice as constraining as the original upper limit.

Neglecting primary beam attenuation, we found a weighted mean flux of $103.5 \pm 2.9$ mJy beam$^{-1}$ from the output of the AIPS task DFTPL on the residual data with full (5s) sampling. We also ran DFTPL on this data for each of the five bursts and for each of the two IFs separately. We only selected times for which both IFs had fluxes and then calculated the natural logarithm of the ratio of the fluxes for each timestamp and the variance of that quantity. We then calculated the weighted mean of these logarithms for each burst. The spectral index and error bar for each burst are shown in Table 5.5, using the average frequencies of IF1 (327.5000 MHz) and IF2 (321.5625 MHz). The spectral indices and error bars of the individual bursts do not seem inconsistent with Gaussian statistics, so we calculated the weighted mean spectral index as well: $\alpha = -6.5 \pm 3.4$. This is not incompatible with the spectral indices found by Hyman et al. (2006, 2007, $\alpha = -4 \pm 5$ and $\alpha = -13.5 \pm 3.0$), given the large error bars. The weighted mean of these three measurements is $\alpha = -9.4 \pm 2.1$. 

113
Table 5.5: Measurement of spectral index for each burst

<table>
<thead>
<tr>
<th>Burst number</th>
<th>$\alpha$ ($S_\nu \propto \nu^\alpha$)</th>
<th>1$\sigma$ error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-9.9</td>
<td>6.7</td>
</tr>
<tr>
<td>2</td>
<td>-9.0</td>
<td>9.3</td>
</tr>
<tr>
<td>3</td>
<td>0.9</td>
<td>8.7</td>
</tr>
<tr>
<td>4</td>
<td>-0.4</td>
<td>6.9</td>
</tr>
<tr>
<td>5</td>
<td>-12.3</td>
<td>6.9</td>
</tr>
</tbody>
</table>

5.5.4 Circular Polarisation

We compared the lightcurves for left (LL) and right (RR) circular polarisation with 30s sampling. Although there are occasional LL and RR flux differences during the bursts larger than the sums of the respective error bars, this is also seen in between the bursts. There is no compelling evidence for circularly polarised emission during any particular phase of the burst cycle. On the other hand, we cannot exclude it completely, because we have insufficient signal to noise in Stokes $V$.

From the residual data, we selected the times corresponding to the bursts and we made a Stokes $V$ dirty image. We corrected for primary beam attenuation and fitted the clean beam to the position of GCRT J1745–3009 as we did in the previous paragraph to determine the upper limit for interburst emission. We measured a Stokes $V$ of $-20 \pm 10$ mJy beam$^{-1}$. Using the total intensity averaged over the bursts, $900 \pm 23$ mJy beam$^{-1}$, we found that the 5$\sigma$ upper limit on the fractional circular polarisation, $|V|/I$, is 8%. Hyman et al. (2005) derived a weaker constraint of 15% on the fractional circular polarisation averaged over the bursts.

Despite the lack of evidence for circularly polarised emission in the discovery observation, it has been detected in the data from the 2003 recovery observation (Roy et al., 2008). Here, only the last part a single burst was covered. From this detection and the fact that the average of Stokes $V$ over a complete burst (almost completely) vanishes we infer that during an earlier part of the burst, Stokes $V$ must have the opposite sign. In other words, if we can assume that the 2003 burst is similar to the 2002 bursts with regard to circularly polarised emission, there must be a sign change in the circular polarisation during the bursts.

5.5.5 Maximum Source Size and Maximum Distance for Incoherent Emission

All of the steep rising part of the bursts can be well approximated by a straight line. This is true even at the very beginning of the bursts, when the flux is at or just above the noise level. It can be seen in the lightcurve
down to 10 s sampling, but at full (5 s) sampling we have insufficient signal to noise to trace any possible slope flattening down to the first 5 s of the beginning of the bursts. The average slope of the bursts in Table 5.3 is 0.75 Jy/min or 0.125 Jy/10 s. This implies a flux doubling time of $\Delta t = 10$ s at the beginning of the bursts, when the flux is 125 mJy. The maximum source size at that time is then 10 lightseconds, if we assume that the source is not moving at a relativistic velocity (see, e.g., Harris et al., 2006, for some background). We can use the maximum source size $c \Delta t$ to link the brightness temperature $T_b(\text{K})$ to the flux $F$ and maximum distance $D$ (see, e.g., Rybicki & Lightman, 1979):

$$T_b = \frac{\lambda^2 I_\nu}{2k} = \frac{\lambda^2}{2k} \frac{F}{\pi \theta^2} = \frac{2}{\pi k} \left( \frac{D}{\nu \Delta t} \right)^2$$

(5.1)

where $\lambda$, $I_\nu$, $\nu$, $k$ and $\theta$ are the wavelength, the specific intensity, the frequency, Boltzmann’s constant and the angle subtended by the radius of the source, respectively. If we express the distance in pc, the flux in Jy and the frequency in GHz, we get:

$$T_b = 4.39 \times 10^{11} \ F \left( \frac{D}{\nu \Delta t} \right)^2$$

(5.2)

If synchrotron self-Compton radiation limits the brightness temperature to $10^{12}$ K, the maximum distance for a source of size ten lightseconds and a flux of 0.125 Jy emitting incoherently at 325 MHz is 14 pc, assuming it is not moving at a relativistic velocity. Hyman et al. (2005) used the decay time of the bursts (conservatively estimated at $\simeq 2$ min.) to calculate a maximum distance of 70 pc. So we have improved this upper limit by a factor 5.

5.6 Discussion

Five of these upper limits on the flux of GCRT J1745–3009 come from the oldest observations of this field in the 92 cm band. This may provide interesting constraints on the feasibility of the double neutron star binary model (Turolla et al., 2005) in the near future. In this model, similar to J0737–3039, the period of recurrence of the 2002 bursts is explained by an orbital period of 77 minutes. The lack of activity for many years is explained by geodetic precession, which could have caused the wind beam of the most luminous pulsar not to intercept the magnetosphere of the other pulsar for decades. Zhu & Xu (2006) claim that the redetection in 2003 (Hyman et al., 2006) does not support this model. Their remark was, however, erroneously based on a geodetic precession period of $\simeq 3$ yr, but this is actually $\simeq 21$ years\(^1\). The last redetection (2004 March 20) and the first observation (1986

\(^{1}\)The ”characteristic time for changing the system geometry” as mentioned by Turolla et al. (2005) differs from the period of geodetic precession by a factor $2\pi$. 

115
March 29) are 18 years apart. Unfortunately this timespan is too short
to test the double neutron star binary model, but not if we redeetect the
system in the near future. More constraining are the results from population
synthesis models (see, e.g., Portegies Zwart & Spreeuw, 1996, Fig. 2): fairly
eccentric \(0.3 < e < 0.6\) double neutron star binaries with an orbital period
of 77 minutes are scarce, even compared to systems like J0737–3039. Also,
the unpulsed emission needed for this model has not been detected in J0737–
3039 (Chatterjee et al., 2005).

The lightcurve from our reduction of the 2002 discovery dataset shows
that the bursts have similar shapes. There are three distinct parts separated
by breaks, a steep rise, a gradual brightening and a steep decay. The main
differences with the lightcurve from the discovery paper (Hyman et al., 2005)
can probably be explained by sidelobes from SgrA (Roy et al., 2007, end of
section 2). These sidelobes are not seen in our images. Apparently, the
lightcurve from the discovery paper was made by compiling fluxes from
successive snapshot images (Hyman et al., 2006, paragraph 2). Therefore,
we also made a lightcurve with 30 s sampling of the fourth burst using the
AIPS task IMAGR and natural weighting, but the differences were negligible.
We also learned that the output from DFTPL is likely to be more accurate
than fluxes from snapshots (Eric Greisen 2009, private communication).

Our refined reduction of the discovery data seems to support the transient
white dwarf model pulsar proposed by Zhang & Gil (2005). A light-house
beam associated with a highly magnetized white dwarf can emit radio emis-
sion with a 77 minute period while maintaining an accuracy better than
one second. The duty cycle \(9/77 \sim 0.1\) (with a few percent jitter from one
pulse to another) is typical for pulsars. Moreover, an intensity asymmetry
between the opposite sides of single pulses is typical in normal pulsars, so it
can be expected also in white dwarf pulsars.

On the other hand, if the bursts we see are actually convolved with some
scattering function, the intrinsic shape of the bursts could be different. In-
terstellar scattering can cause bursts to decay exponentially. We compared
exponential fits to weighted linear regression for the 1 min data chunks that
we used to determine the times of steepest decay. We found that residuals
for linear fits are slightly smaller (12% overall) than for exponential fits. The
exponential fit was better than the linear fit for the tail of one of the four
bursts only. From the exponential fits we found decay times of 0.56, 0.77,
0.73 and 0.81 min for the second, third, fourth and fifth burst, respectively.
These values are rather large for a source near the GC. For an observing
frequency of 325 MHz and for the position of GCRT J1745–3009 on the
sky, pulse broadening times of 3.96–8.72 s and a DM of 567–751 cm\(^{-3}\) pc are
estimated from the NE2001 model of Cordes & Lazio (2003), assuming a
distance (to the GC) of 8 kpc (Reid, 1993). We also checked what disper-
sion measure would follow from our average scattering timescale (0.72 min)
and the empirical relation found by Mitra & Ramachandran (2001):

$$\tau_{sc} = 4.5 \times 10^{-5} \text{ DM}^{1.6} \left(1 + 3.1 \times 10^{-5} \text{ DM}^3\right) \lambda^{4.4},$$  \hspace{1cm} (5.3)

with the scattering time ($\tau_{sc}$) in ms, the dispersion measure (DM) in cm$^{-3}$ pc and the observation wavelength ($\lambda$) in meters. From this relation we find a dispersion measure of $\simeq 925$ cm$^{-3}$ pc. This would imply that GCRT J1745–3009 is located beyond the GC. For a check on consistency we compared this dispersion measure with the DM that can be found from the formula for the dispersion delay $\Delta t$ (in seconds):

$$\Delta t = 4150 \text{ DM} \left(\frac{1}{f_1^2} - \frac{1}{f_2^2}\right)$$  \hspace{1cm} (5.4)

between the highest ($f_2 = 327.50$ MHz) and lowest frequency IF ($f_1 = 321.56$ MHz) using the times of steepest rise for four of the bursts. The delay we found was $-0.94 \pm 3.65$ s corresponding to a DM of $-653 \pm 2530$ cm$^{-3}$ pc, consistent with the value above, but a very weak constraint.

From the poorer quality of the exponential fits relative to the linear fits we are inclined to conclude that the shape of the tails of the observed bursts are dominated by tails in the intrinsic emission. It seems justified that the average decay time from the exponential fits (0.72 min) is merely an upper limit for the true scattering time. In general we can state that for scattering times corresponding to distances not far beyond the Galactic Center the intrinsic burst shape will not differ greatly from the observed burst shape, besides any unresolved variability on very short timescales. The reason for this is that the duration of the observed bursts is much longer ($\simeq 10$ min) than any reasonable scattering time for sources near the GC.

We can work out the original burst profile using theorems for Laplace transforms. The intrinsic emission $I(t)$ is convolved with the scattering function $\zeta(t)$. This gives the observed burst $O(t)$:

$$O = I \ast \zeta,$$  \hspace{1cm} (5.5)

where $\ast$ denotes convolution. For simple scattering, $\zeta$ is the product of the Heaviside step function $\Pi$ and an exponential:

$$\zeta(t) = \Pi(t) \exp(-t/\tau_{sc})$$  \hspace{1cm} (5.6)

The Laplace transform of this product is equal to $(s + \alpha)^{-1}$, with $s$ the transformed coordinate and $\alpha = 1/\tau_{sc}$. Now, using the theorems for Laplace transforms of convolved functions and derivatives we find:

$$I \kappa = \alpha O + \frac{dO}{dt},$$  \hspace{1cm} (5.7)

with $\kappa$ a constant for normalization. If no emission is absorbed, it follows that $\kappa = \alpha$. Thus, we could reconstruct the intrinsic, unscattered burst
from the observed burst if we knew the scattering time $\tau_{sc}$. If the observed burst is represented very accurately by three straight lines for the steep rise, the gradual brightening and the steep decay, the original burst must have the same slopes. It then follows that $\tau_{sc} = 1/\alpha = 0$, hence no scattering, unless there are faults, i.e. sudden "jumps", in the intensity of the intrinsic emission. So the breaks link scattering times and fault sizes.

Without any assumptions on the possible degree of faulting in the intrinsic emission, we can find an upper limit for the scattering times using the end of the tails of the observed bursts. The slopes seem constant until the flux is essentially zero for at least three of the bursts. For the end of the tail of the second burst, which is relatively noisy, this is not so clear. Equation 5.7 then imposes an upper limit on the scattering time $\tau_{sc}$ from the condition that the intrinsic emission cannot be negative. This means that the scattering time must be smaller than the time resolution for which we can determine the slopes with confidence: 10s. This implies that GCRT J1745–3009 cannot be located far beyond the GC. From the NE2001 model of Cordes & Lazio (2003) we find a pulse broadening time of 9.93s at 325 MHz for a distance of 11 kpc in the direction of GCRT J1745–3009.

We conclude from this discussion that the observed bursts depicted in Fig. 5.5 will closely resemble the intrinsic bursts. Models will need to explain the asymmetry of the bursts, the steep rise, the more gradual brightening and the steep decay and the breaks between them as well as the fact that the brightest emission is seen just before the steep decay.

## 5.7 Conclusions

We have derived new upper limits on the quiescent emission of GCRT J1745–3009 at seven epochs. Six observations were made in the 92 cm band and one in the 21 cm band. The 92 cm observation of GCRT J1745–3009 on 2005 March 24 with the WSRT was the second deepest until that time. Five of these seven epochs constitute the oldest set of 92 cm observations taken of the Galactic Center. The nondetections at those epochs do not provide evidence for the double neutron star binary model (Turolla et al., 2005) with a geodetic precession period close to 18 years. However, geodetic precession times could well be somewhat longer.

We have reproduced the lightcurve of the discovery dataset of GCRT J1745–3009 more accurately and more completely than in the discovery paper. We see that the shapes of the five bursts are consistent: a steep rise, a gradual brightening and a steep decay. We have improved the $5\sigma$ upper limit on interburst emission from 75 mJy beam$^{-1}$ to 31 mJy beam$^{-1}$. Also, we further constrained the $5\sigma$ upper limit on the fractional circular polarisation from 15% to 8%. We determined the recurrence interval between bursts more accurately: $77.012 \pm 0.021$ min. We see no evidence for aperiodicity,
but we do find that the duration of the bursts varies at the level of a few %. We derived a very steep spectral index, $\alpha = -6.5 \pm 3.4$. We have investigated scattering and we have shown that scattering times must be less than 10s. This implies that GCRT J1745–3009 cannot be located far beyond the GC. It also means that the shape of the observed bursts will differ little from the intrinsic emission. Models for GCRT J1745–3009 have to explain the asymmetry in the shape of the bursts and in particular the gradual brightening until the steep decay. Some of the suggested models (Turolla et al., 2005; Zhu & Xu, 2006) predict symmetric bursts. The simplest interpretations of those models can now be ruled out, but it is conceivable that the asymmetry in the bursts could be achieved by adding some complexity to those models. Our results favour a rotating system, like the white dwarf pulsar (Zhang & Gil, 2005), because that can explain the high level of periodicity we see. We have shown that it is very unlikely that this transient is an incoherent synchrotron emitter, because it would have to be closer than 14 pc, unless the emitting region is moving at a relativistic velocity. Although we now have more contraints on the properties of this source, we are still unsure about its basic model.

A better understanding of its nature should come from more detections by long time monitoring with high sensitivity and high angular resolution, to tackle the confusion limit and to reduce the number of possible optical counterparts. The next generation of radio telescopes, like LOFAR (see, e.g., Fender et al., 2006), will help to do so. The most pressing issue in revealing the nature of GCRT J1745–3009 is still the determination of its distance, which could be achieved by a new detection with sufficient bandwidth between sidebands, in order to measure the time delay from dispersion towards the Galactic Center.

We have also investigated possible transient behaviour of a source on the opposite side of the supernova remnant G359.1-0.5 but we found no compelling evidence for variability.

Acknowledgements

We thank Tao An from Shanghai Astronomical Observatory for supplying us with an A-configuration model of the radio galaxy 1938-155. This model was necessary to phase calibrate our 1986 March 29 observations. We thank the anonymous referee for encouraging us to reproduce the lightcurve from the 2002 discovery dataset of GCRT J1745–3009 after we reported low noise levels in a map from that dataset. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.
Chapter 6

Low-Frequency Observations of the Radio Nebula Produced by the Giant Flare from SGR 1806–20: Polarimetry and Total Intensity Measurements

H. Spreeuw, B. Scheers & R.A.M.J. Wijers

Abstract

The 2004 December 27 giant flare from SGR 1806–20 produced a radio nebula that was detectable for weeks. It was observed at a wide range of radio frequencies. We aim to investigate the polarized signal from the radio nebula at low frequencies and to perform precise total intensity measurements. We made a total of 19 WSRT observations. Most of these were performed quasi-simultaneously at either two or three frequencies, starting 2005 January 4 and ending 2005 January 29. We reobserved the field in 2005 April/May, which facilitated an accurate subtraction of background sources. At 350 MHz, we find that the total intensity of the source is lower than expected from the GMRT 240 MHz and 610 MHz measurements and inconsistent with spectral indices published previously. Our 850 MHz flux densities, however, are consistent with earlier results. There is no compelling evidence for significant depolarization at any frequency. We do, however, find that polarization angles differ substantially from those at higher frequencies. Low-frequency polarimetry and total intensity measurements provide a number of clues with regard to substructure in the radio nebula associated with SGR 1806–20. In general, for a more complete understanding of similar events, low-frequency observations can provide new insights into the physics of the radio source.
6. The Giant Flare from SGR1806–20

6.1 Introduction

The 2004 December 27 flare from the Soft-\(\gamma\)-ray Repeater SGR 1806–20 was a major event in astronomy in a number of ways. First of all by the energy of the explosion: the brightest flash of radiation from beyond our solar system ever recorded. This is how it caught the attention of a larger audience. Secondly, because the flare provided new observational data about a known class of objects: magnetars, i.e., strongly magnetized neutron stars (see, e.g., Hurley et al., 2005). Also, it led to speculation about a possible link with \(\gamma\)-ray bursts (GRBs) (see, e.g., Tanvir et al., 2005). Theorists investigated the connection between the magnetic field and the explosion (see, e.g., Blandford, 2005). Others focused on modeling the fireball and the afterglow (see, e.g., Nakar, Piran & Sari, 2005; Dai et al., 2005; Wang et al., 2005). Astronomers performed a number of follow-up observations at various wavelengths (Rea et al., 2005; Israel et al., 2005; Palmer et al., 2005; Schwartz et al., 2005; Fender et al., 2006). In particular, the flux from the radio nebula produced by the explosion (Gaensler et al., 2005a; Cameron et al., 2005; Taylor et al., 2005) was measured very frequently in 2005 January. These observations focused on total intensity measurements at various radio wavelengths and on polarimetry at 8.5 GHz. Some polarimetry was done at lower frequencies, but without the proper correction for the leakages (Gaensler et al., 2005b). We have performed accurate polarimetry at 350, 850 and 1300 MHz. Also, we were able to measure the Stokes I flux from the radio nebula at 350 and 850 MHz more precisely by observing the same field again in 2005 April/May. In this way, we could properly subtract the background sources from the \((u,v)\) data of the 2005 January observations. We compare our measurements with those at nearby frequencies.

6.2 Observation and Data Reduction

6.2.1 General

A total of 19 observations were performed in January, April and May of 2005. Four of these, on January 16, 20, 23 and 29 were alternating between 350 and 850 MHz. On January 7 and 10 scans at 1300 MHz were also included. On January 4 we observed at 350, 650 and 1300 MHz, but the 650 MHz data were not used. A summary is shown in Table 6.1.

We used AIPS (Greisen, 2003) and ParselTongue (Kettenis et al., 2006) scripts for the reduction of all 19 datasets. The Westerbork Synthesis Radio Telescope (WSRT) was used for all observations. The WSRT is a linear array with 14 equatorially mounted 25-m dishes equipped with linear feeds. Its maximum baseline is 2.7 km. All datasets recorded four polarization products with 8 IFs. 3C 286 was observed before the target and 3C 48 after. RFI was excised from the spectral line data using the AIPS task SPFLG.
6.2 Observation and Data Reduction

<table>
<thead>
<tr>
<th>Epoch</th>
<th>Days since burst (min)</th>
<th>Time on source (min)</th>
<th>Frequency (MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 4</td>
<td>7.6</td>
<td>54</td>
<td>1300</td>
</tr>
<tr>
<td>January 4</td>
<td>7.6</td>
<td>91</td>
<td>350</td>
</tr>
<tr>
<td>January 5</td>
<td>8.6</td>
<td>462</td>
<td>850</td>
</tr>
<tr>
<td>January 7</td>
<td>10.5</td>
<td>107</td>
<td>1300</td>
</tr>
<tr>
<td>January 7</td>
<td>10.5</td>
<td>107</td>
<td>350</td>
</tr>
<tr>
<td>January 7</td>
<td>10.5</td>
<td>107</td>
<td>850</td>
</tr>
<tr>
<td>January 10</td>
<td>13.6</td>
<td>78</td>
<td>1300</td>
</tr>
<tr>
<td>January 10</td>
<td>13.6</td>
<td>71</td>
<td>350</td>
</tr>
<tr>
<td>January 10</td>
<td>13.6</td>
<td>71</td>
<td>850</td>
</tr>
<tr>
<td>January 16</td>
<td>19.6</td>
<td>181</td>
<td>350</td>
</tr>
<tr>
<td>January 16</td>
<td>19.6</td>
<td>181</td>
<td>850</td>
</tr>
<tr>
<td>January 20</td>
<td>23.6</td>
<td>181</td>
<td>350</td>
</tr>
<tr>
<td>January 20</td>
<td>23.6</td>
<td>165</td>
<td>850</td>
</tr>
<tr>
<td>January 23</td>
<td>26.6</td>
<td>198</td>
<td>350</td>
</tr>
<tr>
<td>January 23</td>
<td>26.6</td>
<td>198</td>
<td>850</td>
</tr>
<tr>
<td>January 29</td>
<td>32.6</td>
<td>196</td>
<td>350</td>
</tr>
<tr>
<td>January 29</td>
<td>32.6</td>
<td>196</td>
<td>850</td>
</tr>
<tr>
<td>April 30/May 1</td>
<td>124.3</td>
<td>444</td>
<td>350</td>
</tr>
<tr>
<td>May 2</td>
<td>125.2</td>
<td>464</td>
<td>850</td>
</tr>
</tbody>
</table>

Table 6.1: Summary of these 19 WSRT observations.

Calibration was done in four steps. First we determined the variation in system temperature as a function of time (and therefore also as a function of position on the sky), using the intermittent firing of a stable noise source. Next we performed a bandpass calibration using the AIPS task BPASS using either 3C 48 or 3C 286 or both. We applied the bandpass solution using the AIPS task SPLAT. After that, we performed an external absolute gain calibration using an assumed flux for 3C 48 by running the AIPS tasks SETJY and CALIB. SETJY was set to use the absolute flux density calibration determined by Baars et al. (1977) and the latest (epoch 1999.2) polynomial coefficients for interpolating over frequency as determined at the VLA by NRAO staff. Finally, we self-calibrated the data for time variations in the relative complex gain phase and amplitude.

Polarization calibration was performed by running the AIPS task LPCAL on 3C 48 and CLCOR to correct for the instrumental XY phase offset. Generally, we followed the scheme for data reduction of WSRT data in AIPS as outlined by Robert Braun\(^1\), although we ran some AIPS tasks differently depending on frequency. Those differences mainly involved the details of polarization.

calibration. For instance, the leakage terms ("D terms") of the WSRT IFs are channel dependent, as pointed out by Brentjens (2008, paragraph 3.2). We took account of this, by first averaging groups of 5 channels through the AIPS task SPLAT. Next, we ran UVCOP to make separate datasets from the averaged channels. After that, we ran LPCAL and CLCOR on each of these separately before applying the feed and XY instrumental phase offset corrections by again running SPLAT.

Before imaging Stokes $Q$ and Stokes $U$ and before merging the datasets from 5 channel averaging back together through DBCON, we applied a Parsel-Tongue script for ”derotation” to the residual data, i.e., the $(u, v)$ data where all sources except the target were removed, by running the AIPS task UVSUB. The original Aips++ glish script was kindly given to us by G. Bernardi; we modified and translated it to a Python/ParselTongue script on a channel by channel basis. The ”derotation” of the visibilities is absolutely necessary, since the rotation measure (RM) of SGR 1806–20 is large, 272 rad/m² (Gaensler et al., 2005a). This means that the polarized signal would vanish if all IFs were imaged simultaneously. For the 350 MHz data, one really needs the derotation of the visibilities to be performed on a channel per channel basis, because the imaging of even one single IF would result in a severely corrupted measurement and underestimate of the fractional linear polarization. The uncertainty in this RM (10 rad/m², see Gaensler et al., 2005a) is too large for accurate polarization angle measurements, especially at frequencies below 1 GHz. For this reason we determined the RM more accurately, by fitting the $\sin^2 \cdot \mathrm{RM} \cdot \lambda^2$ spectrum of either Stokes $U$ or Stokes $Q$ to its measured values at the 8 wavelengths $\lambda$ corresponding to the IFs near 350 MHz and 850 MHz. The contribution to this RM from the ionosphere is naturally included in this fit, at least the part that did not vary during the observation run. We checked the output of the AIPS task TECOR for any significant variations in the ionospheric Faraday rotation during every observing run. The ionospheric Faraday rotation computed by TECOR is considered accurate since it does not use a model for the ionosphere but actual data from the CDDIS archive. We did not apply the ionospheric corrections from TECOR to our data because it implicitly assumes that one has recorded data from circular feeds.

It should be clear from Table 6.1 that the maximum observing time is 7.7 h due to the low declination of the source. Hence, the $(u, v)$ coverage is sparse for all observations, since linear arrays like the WSRT ideally have 12h runs. The worst coverage was at three epochs when we alternated between three frequencies.
6.2 Observation and Data Reduction

6.2.2 Detailed Description of the Datasets

Observations at 350 MHz

The 350 MHz observations were performed on January 4, 7, 10, 16, 20, 23 and 29 and April 30/May 1 of 2005. The last observation was made to make an accurate subtraction of background sources possible. This mainly concerns the subtraction of the Luminous Blue Variable discussed in Supplementary Table 1 of Gaensler et al. (2005a). The time resolution of all observations, except the first and the last was 30s. On January 4 and April 30/May 1 the sampling times of the visibilities was 60s. The bandwidth per IF was 10 MHz, separated 8.75 MHz from each other and centered on frequencies of 315.00, 323.75, 332.50, 341.25, 350.00, 358.75, 367.50 and 376.25 MHz. The IFs were split into 64 channels, each 156.25 kHz wide, except for the April 30/May 1 observation. For that observation, the IFs were split into 128 channels, each 78.125 kHz wide. We used an automated flagger for the initial editing of our data: WSRT flagger\(^2\). 3C 286 was included in the external gain calibration, along with 3C 48. This was trivial, since 3C 286 is unpolarized at this frequency. The assumed fluxes for 3C 48 and 3C 286 in the lowest frequency IF were 43.889 and 26.106 Jy, respectively.

The April 30/May 1 observation has the best \((u,v)\) coverage. After performing 10 iterations of self calibration on this dataset the rms noise in the final image was 2.5 mJy/beam. Its clean components were used to solve for the gain phases and amplitudes of the other datasets using a rather sophisticated scheme. First, a deconvolution of each of the 2005 January datasets was done in order to subtract the central region containing the radio nebula and the LBV, using the AIPS tasks IMAGR, CCEDT and UVSUB. The residual data were calibrated on the April 30/May 1 model which had the clean components from the central region removed. The gain phase and amplitude solutions were then copied and applied to the original 2005 January datasets. It this way we made sure that the Stokes \(I\) flux from SGR 1806–20 would not be reduced by calibrating on a model from an observation months after the flare. As explained in Section 6.2.2, amplitude self calibration could also reduce the Stokes \(Q\) flux. However, due to the large RM of the source and because we use 45 of the available 64 channels, the Stokes \(Q\) flux almost completely vanishes in a single IF at 350 MHz. Thus this problem does not occur, at least not before "derotation".

PSR 1937+21 was observed in between SGR 1806–20 and 3C 48 for polarization calibration. This polarization calibration technique is decribed in detail by Brentjens (2008, paragraph 3.2). Since the RM of this pulsar is positive, Stokes \(Q\) should be 90° ahead of Stokes \(U\) with increasing \(\lambda^2\), as noted by Brown & Rudnick (2009, paragraph 2.3).

\(^2\)http://www.astron.nl/~renting/
Observations at 850 MHz ("UHF High")

We observed SGR 1806–20 on January 5, 7, 10, 16, 20, 23, 29 and May 5 of 2005. The last observation was performed to make an accurate subtraction of background sources possible. The time resolution of all observations, except for the first and the last, was 30s. The sampling time of the visibilities on January 5 and May 5 was 60s. The bandwidth of the eight IFs is 10 MHz, they were separated exactly 10 MHz from each other and ranging from 805 to 875 MHz. Each IF was split into 64 channels with a width of 156.25 kHz, except for the May 1/2 data that were split into 128 channels of 78.125 kHz. The external gain calibration was performed using an assumed flux for 3C 48 of 24.240 Jy for the lowest frequency IF. The 850 MHz data were reduced in almost the same way as the 1300 MHz data. Only polarization calibration was performed slightly differently. Since the Stokes \( Q \) and \( U \) of 3C 286 are not known for the "UHF high" frequencies, when the task CALIB was run on this calibrator, it was set to solve for gain phases only and not for gain amplitudes.

Observations at 1300 MHz

We observed SGR 1806–20 at 1300 MHz on January 4, 7 and 10 of 2005. The total intensity measurements have already been published (see Gaensler et al., 2005a), so we focused on the polarized signal. However, we did check that our Stokes \( I \) fluxes agreed with those previously published.

On 2005 January 4 visibilities were recorded every 60s, on January 7 and 10 every 30s. The eight 20-MHz IFs were centered on frequencies of 1255, 1272, 1289, 1306, 1323, 1340 and 1357 MHz. Each IF was split into 64 channels with a width of 312.5 kHz. The external gain calibration was performed using an assumed flux for 3C 48 of 17.388 Jy for the lowest frequency IF. 3C 286 was also included in the external gain amplitude and phase calibration using an assumed flux of 15.550 at 1255 MHz. 3C 286 is linearly polarized. We took account of this and of the usual "AIPS for linear feeds" projection (\( R \rightarrow X, L \rightarrow Y \)) by placing the assumed Stokes \( Q \) flux of 3C 286 (0.594 Jy at the lowest frequency IF) with a minus sign at the position of Stokes \( V \) in the AIPS SU table. For the other IFs we kept the same ratio between Stokes \( I \) and Stokes \( Q \). In this way we could use 3C 286 not only for fixing the instrumental XY phase offset, but also for external gain calibration.

Self-calibration was run to solve for the gain phases only, since solving for the amplitudes could reduce the Stokes \( Q \) flux. The AIPS task CALIB cannot be set to run simultaneously on a Stokes \( I \) and Stokes \( Q \) model. Obviously, when CALIB is run on a Stokes \( I \) model, it implicitly assumes that \( Q = 0 \). Consequently, the same model is used to derive the X gains from the XX visibilities as the Y gains from the YY visibilities, while in fact \( XX = I - Q \).
and $YY = I + Q$, so different models should be used. When solving for gain phases only, the error made is generally considered acceptable.

6.3 Results

6.3.1 Total Intensity Measurements

The total intensity flux measurements at 350 MHz were done by fitting a Gaussian of the same shape and size as the restoring beam to the (fixed) location of SGR 1806–20 in the Stokes $I$ images. This was done by the AIPS task IMFIT. We used the position, $\alpha = 18^h08^m39.343^s, \delta = -20^\circ24'39.8''$, from Gaensler et al. (2005a) for the fits. The results are summarized in Table 6.2. The error bars are conservative estimates from measurements of the residuals of bright sources in the field. The actual rms noise in these images is much lower, around 3mJy/beam, which is about the same as the error from IMFIT.

6.3.2 Polarimetry

General

Polarimetry was performed on 2005 January 4, 5, 7 and 10. Although all of our observations recorded full Stokes, we anticipated that it would not be possible to detect the polarized signal from SGR 1806–20 on later dates, since the total intensity drops rapidly. Also, we did not expect polarization fractions to exceed the values given by Taylor et al. (2005, Table 2).

<table>
<thead>
<tr>
<th>Epoch (2005 date)</th>
<th>Days since burst</th>
<th>Stokes $I$ (2005 since flux dens.)</th>
<th>1 $\sigma$ error</th>
<th>Stokes $I$ (850 MHz flux dens.)</th>
<th>1 $\sigma$ error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 4</td>
<td>7.6</td>
<td>186</td>
<td>20</td>
<td>157</td>
<td>10</td>
</tr>
<tr>
<td>Jan. 5</td>
<td>8.6</td>
<td>157</td>
<td>10</td>
<td>97</td>
<td>28</td>
</tr>
<tr>
<td>Jan. 7</td>
<td>10.5</td>
<td>78</td>
<td>10</td>
<td>50</td>
<td>22</td>
</tr>
<tr>
<td>Jan. 10</td>
<td>13.6</td>
<td>78</td>
<td>10</td>
<td>35</td>
<td>14</td>
</tr>
<tr>
<td>Jan. 16</td>
<td>19.6</td>
<td>78</td>
<td>10</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>Jan. 20</td>
<td>23.6</td>
<td>78</td>
<td>10</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>Jan. 23</td>
<td>26.6</td>
<td>78</td>
<td>10</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>Jan. 29</td>
<td>32.6</td>
<td>78</td>
<td>10</td>
<td>22</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 6.2: Stokes $I$ flux measurements at 350 and 850 MHz: clean components from the 2005 April 30/May 1 and May 1/2 observations were subtracted.
6. The Giant Flare from SGR1806–20

Determining the RM of SGR1806–20

As noted before, the rotation measure (RM) as measured by Gaensler et al. (2005a) \((272 \pm 10 \text{ rad/m}^2)\) has a rather large error bar which translates into a polarization angle uncertainty at 1300 MHz of \(30.5^\circ\). At 850 MHz this is even \(71.4^\circ\). Naturally, the RM should be determined more accurately before polarization angles are to be measured.

This can be done by plotting Stokes \(U\) or \(Q\) fluxes of SGR1806–20 as a function of frequency and fit for the RM. We are in the advantageous position that these WSRT observations were performed with eight IFs. Over a wide span of frequencies there are many turns of Stokes \(U\) (or \(Q\)) since its spectrum is sinusoidal as a function of \(\lambda^2\). This effect is largest at low frequencies: at 1300 MHz, there is less than one cycle of \(A \cdot \sin(2 \cdot \text{RM} \cdot \lambda^2 + \phi)\), at 850 MHz there are almost two cycles and at 350 MHz there are 23 cycles. It is evident that the most accurate measurement can be made at the lowest frequency, if there is sufficient signal to noise. Fortunately, we could detect polarized signal at 350 MHz from all three observations on 2005 January 4, 7 and 10 after an initial "derotation" of our visibilities using the RM from Gaensler et al. (2005a, 272 rad/m²). This initial derotation prevents diminution of the polarized signal in a single IF. At 850 MHz this initial derotation was not necessary. The noise levels at that frequency were such that detecting a polarized signal was only possible on 2005 January 5 and 7, but the latter observation yielded a very poor constraint on the RM, so we left it out. The 1300 MHz data also gave very poor constraints on the RM, thus in determining the weighted mean RM we ignored those, too. For the other observations, we plotted Stokes \(U\) per IF and solved for the RM (850 MHz) or the correction to the RM (350 MHz), as illustrated in Fig. 6.1. The results are shown in Table 6.3. It turned out that the noise levels in all of the Stokes \(Q\) maps were much higher than in the Stokes \(U\) maps, so we did not use them. In determining the weighted mean RM we also took into account the measurement by Gaensler et al. (2005a, 272±10 rad/m²). From this set of five measurements we derived an RM of \(255.01 \pm 0.83 \text{ rad/m}^2\). It should be clear that, with regard to the 350 MHz RM measurements, the fits give the same reduced \(\chi^2\) for both the positive and the negative correction to the initial "derotation". We removed those ambiguities by considering the Stokes \(U\) measurements near 850 MHz data on 2005 January 5. The fit to these data gave an RM of \(253.14 \pm 12.43 \text{ rad/m}^2\) which made all of the positive RM solutions to the 350 MHz data very unlikely (≃ 3.0\(\sigma\) level for January 4 and 7).

It is evident that the contribution of the ionosphere to the RM, \(\text{RM}_{\text{ion}}\), is included in all fits. For the 2005 January 4, 5, 7 and 10 observations, \(\text{RM}_{\text{ion}}\) as reported by the AIPS task TECOR, is the range 2.1 ± 0.4 rad/m². Consequently, the interstellar RM is given by \(\text{RM}_{\text{int}} = 255.01 - 2.1 = 252.91 \pm 0.92\) rad/m².

128
Figure 6.1: Determining the rotation measure of SGR 1806–20 by fitting the sinusoidal Stokes \( U \) spectrum. Here, the fit was made to the values of Stokes \( U \) on 2005 January 04 at the wavelengths corresponding to the 8 IFs near 350 MHz after the visibilities were "derotated" by an angle corresponding to an RM of \(-272 \pm 10\) rad/m\(^2\). We used gnuplot to fit the function \( A \cdot \sin(2 \cdot \text{RM} \cdot \lambda^2 + \theta) \) for three free parameters \( A \), \( \text{RM} \) and \( \theta \). The correction to the RM from this fit is \(-18.76 \pm 1.82\) rad/m\(^2\). The reported reduced \( \chi^2 \) is 0.69.

<table>
<thead>
<tr>
<th>Epoch [2005 date]</th>
<th>Days since burst</th>
<th>Frequency [MHz]</th>
<th>Measured RM [rad/m(^2)]</th>
<th>1( \sigma ) error [rad/m(^2)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 4</td>
<td>7.6</td>
<td>350</td>
<td>253.24</td>
<td>1.82</td>
</tr>
<tr>
<td>Jan. 5</td>
<td>8.6</td>
<td>850</td>
<td>253.14</td>
<td>12.43</td>
</tr>
<tr>
<td>Jan. 7</td>
<td>10.5</td>
<td>350</td>
<td>253.65</td>
<td>1.05</td>
</tr>
<tr>
<td>Jan. 10</td>
<td>13.6</td>
<td>350</td>
<td>261.74</td>
<td>2.04</td>
</tr>
</tbody>
</table>

Table 6.3: RM measurements of SGR1806-20.
Polarization Fractions and Position Angles

We were able to measure the fractional linear polarization on all of the four epochs mentioned in paragraph 6.3.2. At 850 MHz, we were not able to measure polarization on 2005 January 10. For the other occasions, the measured polarized fluxes, \( P = \sqrt{Q^2 + U^2} \), fractions and their error bars are listed in Table 6.4. The latter two quantities are depicted in Fig. 6.2. The overall conclusion is that there is no compelling evidence for any significant depolarization at any frequency. Only the polarization fraction at 1300 MHz on January 4 is low compared to the 8.4 GHz measurements, but this fraction was determined from our worst fit, i.e., the fit with the highest reduced \( \chi^2 \).

![Figure 6.2:](image)

**Figure 6.2:** Comparison between linear polarization fractions at 350, 850, 1300 and 8400 MHz.
### Table 6.4: Polarimetric measurements of SGR 1806–20

<table>
<thead>
<tr>
<th>Epoch (2005) burst</th>
<th>Days since</th>
<th>Frequency (MHz)</th>
<th>$\sqrt{Q^2 + U^2}$ (mJy/beam)</th>
<th>1 $\sigma$ Polarization error (mJy/beam)</th>
<th>Polarization fraction (%)</th>
<th>1 $\sigma$ error (%)</th>
<th>Polarization angle (°)</th>
<th>1 $\sigma$ error (°)</th>
<th>Reduced $\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 4</td>
<td>7.6</td>
<td>350</td>
<td>2.68</td>
<td>0.81</td>
<td>1.44</td>
<td>0.46</td>
<td>103</td>
<td>39</td>
<td>0.70</td>
</tr>
<tr>
<td>Jan. 4</td>
<td>7.6</td>
<td>1300</td>
<td>0.71</td>
<td>0.66</td>
<td>0.47</td>
<td>0.44</td>
<td>31</td>
<td>26</td>
<td>1.57</td>
</tr>
<tr>
<td>Jan. 5</td>
<td>8.6</td>
<td>850</td>
<td>2.22</td>
<td>0.28</td>
<td>1.41</td>
<td>0.20</td>
<td>96</td>
<td>7</td>
<td>0.85</td>
</tr>
<tr>
<td>Jan. 7</td>
<td>10.5</td>
<td>350</td>
<td>2.30</td>
<td>0.64</td>
<td>2.73</td>
<td>0.82</td>
<td>69</td>
<td>38</td>
<td>0.33</td>
</tr>
<tr>
<td>Jan. 7</td>
<td>10.5</td>
<td>850</td>
<td>1.71</td>
<td>0.52</td>
<td>1.76</td>
<td>0.74</td>
<td>44</td>
<td>9</td>
<td>0.65</td>
</tr>
<tr>
<td>Jan. 7</td>
<td>10.5</td>
<td>1300</td>
<td>1.90</td>
<td>0.37</td>
<td>2.29</td>
<td>0.48</td>
<td>50</td>
<td>7</td>
<td>1.06</td>
</tr>
<tr>
<td>Jan. 10</td>
<td>13.6</td>
<td>350</td>
<td>1.14</td>
<td>0.87</td>
<td>1.46</td>
<td>1.13</td>
<td>36</td>
<td>47</td>
<td>0.79</td>
</tr>
<tr>
<td>Jan. 10</td>
<td>13.6</td>
<td>1300</td>
<td>1.31</td>
<td>0.40</td>
<td>3.65</td>
<td>1.38</td>
<td>69</td>
<td>12</td>
<td>1.28</td>
</tr>
</tbody>
</table>

1 Poor fit.
6. The Giant Flare from SGR 1806–20

The polarization angles and their uncertainties are also listed in Table 6.4. The observations at 850 and 1300 MHz gave the most accurate position angles, with typical uncertainties of order $10^\circ$. They are depicted in Fig. 6.3. Here, we see compelling evidence for significantly different polarization angles with respect to the 8.4 GHz observations from Taylor et al. (2005), particularly on January 5 at 850 MHz and on January 7 at both 850 and 1300 MHz.

6.4 Discussion

6.4.1 Total Intensity Measurements

It is clear from Fig. 6.4 that SGR 1806-20 is much dimmer at 350 MHz than what would be expected from the GMRT observations at 240 and 610 MHz (Cameron et al., 2005). In principle the Luminous Blue Variable, 14" to the east of SGR 1806–20 (see the Supplementary Information to Gaensler et al., 2005a) should be easily distinguishable from the Soft Gamma Repeater in the GMRT images, even at 240 MHz. The FWHM beamsize reported at that frequency is $12'' \times 18''$ (Chandra, 2005b). This makes it hard to understand
6.4 Discussion

Figure 6.4: Comparison between the 240, 350, 610 and 1300 MHz fluxes of the radio nebula associated with SGR 1806–20. The 1300 MHz fluxes were published previously (Gaensler et al., 2005a). The fluxes at 350 MHz and 1300 MHz are approximately equal on 2005 January 7, they coincide in this plot. For the 350 MHz flux on 2005 January 29, instead of the flux, the $3\sigma$ upper limit on the flux is depicted in order to enable a more appropriate vertical scale.

In principle the discrepancy cannot originate from the inclusion or exclusion of extended emission. The GMRT data were corrected for this (Chandra, 2005a,b). We excluded short spacings ($< 1k\lambda$) from our 350 MHz WSRT observations. This was actually a necessity since these were daytime observations and solar interference would otherwise compromise our calibration (see also Brentjens, 2008, end of paragraph 3.2).

Also, it is possible that the LBV radio nebula is variable and that it was much brighter on 2005 April 30/May 1 than on some occasions in 2005 January. We ran the AIPS task IMFIT on the map from our 2005 April 30/May 1 observation and we found a peak flux density of $138 \pm 1\text{mJy/beam}$ and an integrated flux of $189 \pm 2\text{mJy/beam}$ at the location of the LBV. The NVSS (Condon et al., 1998) image of this field shows this source at the 15 mJy level. This would indicate that the LBV has a spectral index of about $-1.8$, which is almost the index for thermal radio radiation. It should be
noted that, at the times of the latest observations in January 2005, when the radio nebula was relatively dim, there is no evidence for negative residuals in our maps that could be caused by the subtraction of the LBV. This indicates that, most likely, the LBV had the same brightness at the times of at least some of the 2005 January measurements as on 2005 April 30/May 1.

Variability at radio wavelengths of the radio nebulae from LBVs has been known for quite some time (see, e.g., Abbott et al., 1981). For the P Cygni nebula variability at timescales of days was established at cm wavelengths (Skinner et al., 1996). These authors report a 50% increase in flux in less than two days on one occasion during three months of observations on every other day. It is unknown how these variations translate to lower frequencies. We therefore cannot completely exclude that the LBV was brighter at the time of the 2005 April 30/May 1 observation than on some occasions in January 2005. Also, the spectral index derived above does not agree with any of the spectral indices of the four LBVs observed by Duncan & White (2002) at 3 and 6 cm. Two of those spectral indices are close to that of a spherically symmetric radially expanding stellar wind (+0.6, see Panagia & Felli, 1975; Wright & Barlow, 1975). However, at these wavelengths, those systems may well be described as optically thin, which may not be the case at the frequencies we are considering.

The WSRT 850 MHz Stokes $I$ measurements are not inconsistent with the 840 MHz MOST data published earlier (Gaensler et al., 2005a), given the rather large noise levels in the data from both telescopes. The last MOST observation was taken 15 days after the Giant Flare (GF). Consequently, the 850 MHz WSRT observations after 2005 January 10 cannot be compared with other observations in this band. The last three of the January 2005 observations at 850 MHz were less contaminated by RFI than the first four, which resulted in smaller error bars on the fluxes. There is evidence ($> 2\sigma$ level) for a deviation from a power-law decay from about 15 days after the GF, analogous to the 4.8 GHz observations by Gelfand et al. (2005, paragraph 2). These authors also mention a gradual rebrightening from about 25 days after the GF, as a result of swept up ambient material. We can also see that in the WSRT 850 MHz data, but the evidence for this is less compelling, since the sampling of these observations is sparse in time. Consequently, it is shown only in one of our observations, on 2005 January 29, 32.6 days after the GF.

6.4.2 Polarimetric Measurements

In Fig. 6.2 we compare the polarization fractions as listed in Table 6.4 with the measurements at 8.4 GHz by Taylor et al. (2005, Table 2). In Fig. 6.3 we have done the same for the polarization angles. It is clear that the observations at 8.4 GHz are much more accurate. Still, we do not see any significant discrepancies in the polarization fractions.
Our observations reveal larger polarization position angles than the 8.4 GHz observations. Most compelling are the observations on 2005 January 5 at 850 MHz and on January 7 at both 850 and 1300 MHz. The error bar on the polarization angle at 350 MHz on January 4 is rather large, but this measurement and the 850 MHz measurement on January 5 show the largest differences with the 8.4 GHz observation, about 85°. At these times, the polarization angles from the 8.4 GHz observations suggest that the magnetic field in the emitting plasma is aligned preferentially along the axis of the radio source, on average (Gaensler et al., 2005a). Thus, the January 4 and 5 polarization angles at 350 and 850 MHz indicate that the magnetic field in the emitting plasma that causes linearly polarized radiation at these low frequencies is close to perpendicular to the axis of the radio source, within ≃ 20°. Possibly a different substructure in the radio nebula is being probed. It seems hard to explain this feature without a complex model of the radio source.

6.5 Conclusions

It is striking that depolarization at low frequencies is absent. Also, we have shown that low-frequency polarimetry of SGR 1806–20 provides hints with respect to the detailed substructure of the radio nebula which cannot be derived from the extrapolation of high frequency measurements. Models for the radio nebula need to take into account a distinct source of linearly polarized low-frequency radiation with magnetic fields in the emitting plasmas aligned quite differently from the fields that are associated with radiation at high frequencies.

Acknowledgements

We thank Michiel Brentjens, James Miller-Jones and Gianni Bernardi for helpful discussions about polarization calibration. We thank Eric Greisen for providing background information about many AIPS tasks. The Westerbork Synthesis Radio Telescope is operated by ASTRON (Netherlands Foundation for Research in Astronomy) with support from the Netherlands Foundation for Scientific Research (NWO). This research was supported by NWO Vici grant 639.043.302 (HS and RAMJW) and by NWO NOVA project 10.3.2.02 (BS).
Samenvatting in het Nederlands

Lang was het blote oog het enige instrument van de mens om de sterrenhemel te kunnen waarnemen, totdat Galilei aan het begin van de 17de eeuw de telescoop introduceerde. In de loop der eeuwen is dit instrument in velerlei opzichten enorm verbeterd en heeft het astronomen in de gelegenheid gesteld om het firmament in steeds meer detail te bestuderen. Hoe het ook zij, met het oog of een telescoop, de waarnemingen aan de sterrenhemel werden tot 1932 slechts in het zichtbare deel van het elektromagnetisch spektrum gedaan. De opkomst van het transatlantisch telefoonverkeer aan het begin van de 20ste eeuw maakte de ontdekking van radiostraling van kosmische oorsprong mogelijk. Radiostraling is ook elektromagnetisch van aard en heeft langere golflengten en lagere energieën en frequenties dan straling van zichtbaar licht en kan met radiotelescopen worden gedetecteerd. Karl Jansky toonde in 1932 aan dat de door hem waargenomen radiostraling uit het centrum van onze eigen Melkweg afkomstig was. In rap tempo werden na de Tweede Wereldoorlog nog meer bronnen van radiostraling ontdekt, uit-eenlopend van verre sterrenstelsels op miljarden lichtjaren afstand, tot de nabije overblijfselen van ontplofte sterren (supernova’s), de snelroterende en pulserende neutronenstern (pulsars) in onze Melkweg. Ook de alomvertegenwoordigde 3 Kelvin achtergrondstraling, de nagalm van de Oerknal, is voor het eerst waargenomen met radiotelescopen.

De huidige radiotelescopen bestaan vaak uit een reeks identieke schotelantennes, zoals te zien in Figuur 1, die tezamen een groter oppervlak hebben en zodoende hogere resoluties halen. Omdat het signaal van een ver verwijderde radiobron de voorste telescoop als eerste bereikt en de achterste als laatste, kan door het inbouwen van tijdvertragingen per telescoop het signaal door interferentie versterkt worden. Op deze manier kunnen zwakkere signalen opvangen worden. Door verder tijdens een waarneming gebruik te maken van de draaiing van de Aarde zorgt een goede uv-bedekking voor betere afbeeldingen dan in het geval van een enkele schoteltelescoop. De zware constructies en beweegbare onderdelen van deze instrumenten zijn echter complex en kostbaar en leggen beperkingen op aan de ontwerpen ervan.
Het rechtstreeks koppelen van krachtige computers aan de instrumenten, maakt het nu mogelijk alternatieve ontwerpen in ogenschouw te nemen, waarbij de signaalverwerking en gegevensanalyse softwarematig afgehandeld worden. LOFAR, de Low Frequency ARray, is het eerste instrument in zijn soort dat op deze manier ontworpen is en verschilt wezenlijk van de klassieke radiotelescopen. LOFAR bestaat uit vele relatief simpele radioantennes die gegroepeerd zijn in stations, welke weer verspreid zijn over een oppervlak ter grootte van Noordwest-Europa. Elk station bestaat uit twee typen antennes: de lage-frequentie Low Band Antennas (LBA) zijn gevoelig in het frequentiegebied van 30 tot 80 MHz en de hoge-frequentie High Band Antennas (HBA) zijn gevoelig van 120 tot 240 MHz. Figuur 2 laat de twee antennetypen van LOFAR zien. De kern van het instrument bevindt zich bij Exloo, Drenthe. Hier liggen zes kernstations dicht bij elkaar op een super-
Samenvatting in het Nederlands

Figuur 2: Links: De lage-frequentie Low Band Antennas (LBA) zijn operationeel tussen 30 en 80 MHz (foto T. Krieg). De door de antennes ontvangen signalen worden in een computercontainer, linksboven op de foto, gecombineerd, voordat de gegevens naar de supercomputer in Groningen getransporteerd worden voor verdere verwerking. Rechts: De hoge-frequentie High Band Antennas (HBA) zijn operationeel tussen 120 – 240 MHz.

Figuur 3: Een LOFAR kernstation bestaat uit een LBA veld en twee HBA velden. Het LBA veld bevat 96 dipoolantennas. Een (kernstation) HBA veld bevat 24 tegels met elk 16 HBA antennes. NB. Een kernstation is iets anders van ontwerp dan een remote of EU station.
Samenvatting in het Nederlands

Figuur 4: Impressie van de gerealiseerde en geplande lokaties van de LOFAR stations in Europa (Spektrum der Wissenschaft/Emde-Grafik).

...terp van 340 m doorsnede. Nog eens zestien kernstations liggen hieromheen verspreid over een gebied van 2 tot 3 km. Een voorbeeld van een kernstation is te zien in Figuur 3. De stations in de verder naar buiten gelegen gebieden (remote stations) zijn dunner verspreid over Noord-Nederland, met afstanden tot de kern tot ongeveer 100 km. De gefinancierde Europese stations liggen momenteel in Duitsland, Frankrijk, Engeland en Zweden en hebben basislijnen tot 1000 km. Figuur 4 laat een overzicht zien van gerealiseerde en geplande LOFAR stations.

De signalen die door de antennes opgevangen zijn worden eerst op stationsniveau verwerkt voordat ze worden verzonden naar de supercomputer in Groningen, waar de gegevens van alle LOFAR stations tezamenkomen om te worden gecorreleerd. Hierna zorgt een zogenaamd post-processing computercluster voor ijking en imaging van de gegevens. De gegevens die na deze stappen beschikbaar komen, worden voor wetenschappelijke doeleinden verder onderzocht. Eén van LOFARs onderzoeksprojecten is het Transients Key Science Project (TKP) dat zich ten doel stelt alle veranderlijke en vergankelijke objecten met LOFAR waar te nemen en wetenschappelijk te onderzoeken. Objecten en fenomenen van speciale aandacht zijn o.a. de pulsars, uitbarstingen van magnetars (neutronensterren met extreem ster-
ke magneetvelden), de radiostraalstromen (Eng.: radio jets) afkomstig van zwarte gaten in röntgendubbelsterren en superzware zwarte gaten in actieve kernen van sterrenstelsels, de nagloeiers van gammaflitsers en exoplaneten. Al deze objecten zijn veranderlijk op tijdschalen variërend van seconden tot zelfs jaren.

De eigenschappen van LOFAR zijn uitermate geschikt om bovengenoemde typen bronnen waar te nemen. Het grote gezichtsveld van de antennes, de lange afstanden tussen de stations en de grote bandbreedte van 48 MHz maken het mogelijk om grote gebieden van de hemel met een ongekende gevoeligheid en resolutie waar te nemen, veel groter vergeleken met de klassieke radiotelescopen. Door verschillende bundels te vormen kunnen zelfs nog grotere oppervlakken, al dan niet aan elkaar grenzend, tegelijkertijd waargenomen worden, hoewel dit ten koste gaat van de gevoeligheid. Bronnen in het grote frequentiebereik van LOFAR, 30–240 MHz, zijn nog relatief weinig in kaart gebracht, en van vele buigt het spectrum om in dit gebied.

Omdat LOFAR softwarematig bestuurd kan worden zijn inzicht in de gebeurtenissen die een specifieke resolutie of gevoeligheid vereisen. Een andere ongeëvenaarde eigenschap van LOFAR is de korte integratietijd van de observaties. De supercomputer en het computercluster zijn erop gericht om tijden zo kort als 1 seconde te halen, oplopend tot de gewenste waarneming-specifieke integratietijden. Door deze “snapshots” van één seconde is het mogelijk om de hele hemel in een korte tijd, ongeveer een dag, in kaart te brengen.

LOFAR luidt het begin in van een nieuw tijdperk voor de radioastronomie en zelfs de sterrenkunde. De nieuwe generatie telescopen zullen in belangrijke mate volgens de principes van een nieuw paradigma ontworpen worden, waarbij sterrenkunde, electrotechniek en informatica met elkaar verstrengeld zijn.

Eén van de belangrijkste doelen van dit proefschrift was inzicht te verschaffen in het gedrag van veranderlijke bronnen in de LOFAR frequenties. Hiertoe zijn interferometriedata van verschillende observaties van de nagloeier van een gammaflitser en de uitbarsting van een zogenaamde soft gamma-ray repeater (SGR) gereduceerd en geanalyseerd. LOFAR zal de klassieke observatiemethoden voorbijgaan in het feit dat vele (miljoenen) metingen per object gemaakt zullen worden met tussenpozen zo kort als één seconde. Deze enorme hoeveelheden gegevens kunnen dan niet meer handmatig worden geanalyseerd, maar zullen uiteindelijk automatisch verwerkt moeten worden in zogenaamde software pipelines, die er speciaal op gericht zijn veranderingen in de helderheden van bronnen te detecteren. Dit diende als het kader voor de wetenschappelijke software: het ontwerpen van een database die in staat is alle metingen van alle waargenomen bronnen op te slaan tijdens LOFAR waarnemingen; en het opstellen van snelle re-
kenkundige regels om bronnen te localiseren en veranderlijke helderheden te registreren. Hoofdstukken 2, 3 & 4 zijn gerelateerd aan de TKP software en Hoofdstukken 5 & 6 aan de typen veranderlijke bronnen die te detecteren zijn met LOFAR.

Hoofdstuk 2 beschrijft de belangrijkste waarneemstrategieën van LOFAR die relevant zijn voor de TKP. Hiertoe behoren de Zenith Monitor en de Rapid All-Sky Monitor. Ook is de Million Sources Sky Survey (MSSS) behandeld en de modus waarbij alle Nederlandse LOFAR stations betrokken zijn, de Full-NL Array. Voor elke configuratie hebben we voor de kortste (1 seconde) tot de langste integratietijden de te verwachte aantallen bronnen uitgerekend en de hiervan afgeleide benodigde schijfruimte en gegevensverwerkingsstijden. We hebben het databaseschema, de belangrijkste tabellen en algoritmes uitgewerkt en daarbij de snelheid gemeten van de algoritmes die bronnen associëren en veranderlijke helderheden detecteren in twee typen databases, MySQL en MonetDB. De resultaten geven een duidelijke voorkeur om het kolomgeoriënteerde databasesysteem van MonetDB in de TKP software pipeline te gebruiken. Bovendien is uit Figuren 2.13 en 2.14 op te maken dat het verwerken van de grote datastromen in de database zelf voor vrijwel alle gevallen binnen de te verwachte verwerkingsstijden blijft.

De algoritmes van het associëren van bronnen – het bepalen van de kans dat een bron in de ene waarneming dezelfde is als een bron in de andere waarneming – is beschreven in Hoofdstuk 3. Ook de algoritmes om variabiliteit van bronnen te kunnen detecteren zijn hier uitgewerkt. Voor ieder paar bronnen dat mogelijk met elkaar geassocieerd is, worden drie parameters bepaald. Dat zijn hun onderlinge afstand, het gewogen positieverschil en de waarschijnlijkheidsverhouding. De laatste is de verhouding van de kans dat een associatie een positieverschil heeft door meetfouten (dit volgt de bekende Rayleigh-verdeling) en de kans dat de associatie puur toevallig is met een ongerelateerde bron op de achtergrond. Evaluatie van deze parameters bepaalt op een betrouwbare en snelle manier of een associatieschaal echt of toevallig is. Verder hebben we twee indices gedefinieerd die een maat voor de variabiliteit zijn. De eerste index beschrijft de absolute variabiliteit, terwijl de tweede de gewogen fluxveranderingen, oftewel de significantie, weergeeft. De significantie van de variabiliteit is een kwantitatieve meetmethode voor de waarschijnlijkheid die uitmaakt of een gebeurtenis daadwerkelijk variabel was en aldus de basis vormt voor beslissingen om deze bron eventueel op te volgen met nieuwe waarnemingen.

We hebben de algoritmes toegepast op gesimuleerde en echte gegevens. De gesimuleerde data maakten gebruik van ruiskaarten van VLA waarnemingen op 325 MHz. Per kaart werden 64 bronnen op vaste posities ingevoerd, wat betekent dat positieafwijkingen slechts veroorzaakt worden door variaties in de locale ruis, en een Rayleigh-verdeling volgen. Op deze manier zijn 1000 kaarten verwerkt en konden we de kritieke waarden van de as-
sociatieparameters bepalen. Deze criteria zijn vervolgens toegepast op een serie waarnemingen van de nagloeier van gammaflitser GRB 030329. Tussen 2003 en 2007 heeft de Westerbork Radio Telescoop het bijbehorende stuk van de hemel in verschillende frequenties, resoluties en gevoeligheden waargenomen. Hierin vonden we acht veranderlijke bronnen, waarvan er twee als serieuze kandidaat beschouwd worden om met LOFAR waargenomen te worden.

Hoofdstuk 4 beschrijft het gebruik van verschillende catalogi in de TKP database. Deze catalogi zijn geproduceerd naar aanleiding van waarnemingen met verschillende observatoria om de hemel in kaart te brengen en bevatten alle gevonden bronnen. Uiteindelijk zal LOFAR zelf ook een catalogus met alle waargenomen bronnen (en hun lichtkrommen) produceren en deze reeds in een vroeg stadium beschikbaar stellen aan de astronomen. Deze lijst van ijkbronnen is van belang voor het calibratieproces van LOFAR en voor de TKP om bronnen op te sporen die in vergelijking met eerdere of andere waarnemingen veranderlijk zijn. Het identificeren van de bronnen gebeurt volgens de principes zoals beschreven in Hoofdstukken 2 & 3. Verder hebben we de betrouwbaarheid bepaald van waarden voor de waarschijnlijkheidsverhouding. Op basis hiervan hebben we een lijst van hoge kwaliteit kunnen samenstellen. De initiële lijst bevat bijna 2800 bronnen die in de drie belangrijkste radiocatalogi voorkomen, VLSS (74 MHz), WENSS (325 MHz) en NVSS (1400 MHz) en helderder zijn dan 1 Jy op 74 MHz en liggen tussen declinaties 44° en 64°.

Hoofdstuk 5 rapporteert over de Galactic Centre Radio Transient source, GCRT J1745–3009. Deze was ontdekt in maart 2005 op 325 MHz door Hyman et al. (2005), en liet vijf uitbarstingen op Jansky-niveau zien die elk 10 minuten duurden met tussenpozen van 77 minuten. Uitgezette campagnes om dit onbekende object opnieuw waar te nemen hadden weinig succes, op twee herdetecties van zwakke uitbarstingen na. De dataset van de ontdekking is opnieuw geanalyseerd, wat leidde tot een nauwkeurigere bepaling van het verloop van de uitbarsting als ook een verbeterde waarde voor de periodiciteit van de uitbarstingen. Tot nu toe is deze bron alleen bij 325 MHz waargenomen en het is nog steeds niet duidelijk wat voor object het is. LOFAR kan een belangrijke rol spelen bij het ontrafelen van de aard van dit object, hoewel de lage declinatie parten zal spelen, door de mogelijkheid om langdurige waarnemingen uit te voeren met hoge resolutie en gevoeligheid.

De resultaten van waarnemingen snel na de Grote Uitbarsting op 27 december 2004 van de soft gamma-ray repeater SGR 1806–20 zijn beschreven in Hoofdstuk 6. Een soft gamma-ray repeater is een neutronenster met zeer sterk magneetveld, die met enige regelmaat heldere fits in uitstoot van gammastraling, veel energetischer dan radiostraling. Dat we deze SGR toch met radiotelescopen konden waarnemen, komt doordat de electronen in de omheenliggende nevel van geioniseerd gas in het magneetveld spiraliseren
en daarbij synchrotron radiostraling uitzenden. Metingen van de Westerbork radiotelescoop aan de polarisatie en de intensiteit van SGR in de lage frequenties, wezen uit dat de depolarisatie op deze lage frequenties afwezig is, wat erop kan duiden dat andere substructuren bekeken worden dan bij hogere frequenties. Ook hier zou LOFAR de uitbarsting hebben kunnen detecteren en de evolutie van de radionevel en het magnetisch veld nauwkeurig hebben kunnen volgen.
List of Publications


* Database Techniques within LOFAR’s Transients Key Project, Scheers, B., 2009, ASPC, 411, 143


* The LOFAR Transients Key Project, Rob Fender, Ralph Wijers, Ben Stappers, Robert Braun, Michael Wise, Thijs Coenen, Heino Falcke, Jean-Mathias Griessmeier, Michiel van Haarlem, Peter Jonker, Casey Law, Sera Markoff, Joseph Masters, James Miller-Jones, Rachel Osten, Bart Scheers, Hanno Spreeuw, John Swinbank, Corina Vogt, Rudy Wijnands & Philippe Zarka, Proceedings of the VI Microquasar
List of Publications

Workshop: Microquasars and Beyond. September 18-22, 2006, Como, Italy., p.104.1
Dankwoord

Zo, het is af! Gefeliciteerd, en wat nu?
Ja, dank u. Ik zou graag van de gelegenheid gebruik willen maken om een aantal mensen te bedanken. Zij hebben het mogelijk gemaakt dat dit proefschrift er is gekomen. Daarom wil ik allereerst mijn hoogge...

Ho maar, dat doen we straks wel. Ik ben benieuwd wat u heeft bewogen om na zovele jaren terug te keren in de sterrenkunde, een halvering van uw salaris op de koop toe te nemen en nog net voor uw pensioengerechtigde leeftijd te promoveren.
Nou, zo oud ben ik nog niet, en ik loop nog steeds zonder stok. Maar ik had natuurlijk ontzettende heimwee. Toen ik aan de UvA met natuurkunde begon, eind jaren tachtig, ontmoette ik daar Paul (ooit was ik zijn paranimf en nu zit hij in mijn promotiecommissie, hoe bijzonder!), Jeroen (onze avonturen op en rond de Bajkal–Amoer–Magistrale zijn onvergetelijk), Ernest en Mike, en later Haakon en Hanno. We zijn elkaar na al die jaren niet echt uit het oog verloren en de contacten met de sterrenkunde bleven.

Hanno? Bedoelt u Hanno Spreeuw? Ja inderdaad, ik heb veel aan hem te danken. Ik ontwierp al een flink aantal jaren software voor de grotere gegevensbanken, toen hij begin 2005 enthousiast vertelde over de nieuwe radiotelescoop LOFAR. Eenmaal begonnen, eind 2005, maakte hij me wegwijst in de datareductie van radiowaaarnemingen. We hebben veel samengewerkt en ik heb erg genoten van de tijden die we samen doorbrachten. Hanno, bedankt!

Maar wat vond u er eigenlijk van, de afgelopen vier jaar?
Eh, het waren er vijf.

Oh pardon, vijf. Kunt u wat meer vertellen over die periode?
Het eerste jaar was behoorlijk wennen, ik ging samenwonen met Marja en we kregen een prachtige zoon, Brecht. Als het komt, dan komt het ook allemaal tegelijk. Maar goed, dat is natuurlijk op het persoonlijke vlak. Zakelijk ge-
zien vond ik het een enorme uitdaging...

**Zakelijk? U bent toch geen zakenman?**
Nee, ik bedoel natuurlijk, werkelijk, wat het werk betreft. Ik vond het geweldig interessant om aan de ene kant een systeem te ontwerpen dat het mogelijk maakt om alle gegevens van alle LOFAR-waarnemingen te behouden en zodanig op te slaan dat ze ook weer zo snel mogelijk opgevraagd en hergebruikt kunnen worden. En om anderzijds bezig te zijn met wetenschap, de radiobronnen die LOFAR op het punt staat te gaan waarnemen. Wat ik als heel prettig heb ervaren is dat Ralph mij hierin vrij gelaten heeft en op de juiste momenten, als ik bijvoorbeeld strandde in details, weer de goede richting op wees.

**Ralph? Dat moet uw promotor zijn.**
Klopt. Ralph, heel erg bedankt!

**Wat verwacht u dat er met uw werk zal gebeuren?**
Ik hoop dat het als basis kan dienen om vergankelijke en veranderlijke bronnen te detecteren en te volgen, en zo kan bijdragen aan de spectaculaire ontdekkingen die LOFAR ongetwijfeld gaat doen.

**De sfeer bij het API schijnt zo bijzonder te zijn. Kunt u dat bevestigen?**

**U draagt uw proefschrift op aan uw familie. Gaat u nog namen noemen?**
Het leven met een kind speelt zich toch voornamelijk in en rond het huis af en de borrels met de verticale, horizontale en diagonale buren Renske & Paul, Sander en Jowanneke passen daar precies bij.

Wat je ook vaak ziet in proefschriften zijn al die Latijnse citaten en Chinese wijsheden. Hebt u nog een mooie uitsmijter voor de lezers?
Ik zou natuurlijk veel van Richard Feynman of Marten Toonder kunnen aanhalen, maar laat ik voor de verandering eens wat uit eigen werk citeren:

Sinds professor De Bie
tot in de precisie en uitstekend
de maten van het heelal had uitgereked
lijdt hij aan claustrofobie.

Leuk hoor. Nog meer? Of nog liefdesverklaringen af te leggen?
Nee dank u, dat weten ze wel; nu is het tijd voor een biertje. U ook één?
Bibliography

Abell et al., 2009, LSST Science Collaborations, arXiv, arXiv:0912.0201L
Barlow, R.J., 1989, Statistics, John Wiley & Sons Ltd., England
Blandford, R. D. 2005, in BAAS, 37, 1220
Chandra, P. 2005a, GRB Coordinates Network, 2947, 1
Chandra, P. 2005b, GRB Coordinates Network, 2995, 1


Fender, R.P. et al., in Proceedings of the VI Microquasar Workshop: Microquasars and Beyond, 2006, 104


Gamow, G., 1946, Phys.Rev., 70, 572


Hodapp, K.W., et al., 2004, AN, 325, 636


van de Hulst, H.C., 1945, Ned. Tijdschr. Natuurk., 11, 201


Idreos, S., Kersten M. L., Manegold, S., 2007, CIDR 2007, January 7-10, Asilomar, CA, USA, p. 68


Kamble A., et al., 2010, in preparation


Law C. & Hessels, J. 2009, Conclusions on Standard Bands for RSM, Internal document


Mészáros, P., 2006, RPPh, 69, 2259


Miller-Jones J., 2006, On frequency bands for the Transient database, Internal document

Miller-Jones J., 2008, On the number of sources expected to be detected by the LOFAR Transients Key Project: update following descope, Internal document

Muller, C.A. & Oort, J.H., 1951, Nature, 168, 357


Richter, G. A. 1975, AN, 296, 65


Bibliography


156
Taylor, G.B., Gelfand, J.D., Gaensler, B.M., Granot, J., Kouveliotou, C.,
Fender, R.P., Ramirez-Ruiz, E., Eichler, D., Lyubarsky, Y.E., Garrett,

622, 86

Thakar, A.R., 2008, Computing in Science and Engineering, Volume 10,
Issue 6, 65-71

Thompson, A.R., Moran, J.M. & Swenson Jr., G.W., 2004, Interferome-
try and Synthesis in Radio Astronomy, 2nd Ed., Wiley-VCH, Weinheim,
Germany


97, No. 8, p. 1431

L29


Radio Telescopes, TU Delft, The Netherlands


Wrobel, J.M. & Walker, R.C., 1999, Ch. 9 in Synthesis Imaging in Radio
Astronomy II, ed. G.B. Taylor, C.L. Carilli & R. A. Perley (San Francisco,
CA: ASP), 180, 171

