Optimization for high fidelity imaging with aperture array telescopes

Noorishad, Parisa

rijksuniversiteit
groningen

Optimization for high fidelity imaging with aperture array telescopes

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
vrijdag 7 juni 2013
om 12.45 uur

door

Parisa Noorishad

geboren op 22 januari 1980
te Karaj, Iran

Promotores:	Prof. dr. J. M. van der Hulst Prof. Ir. A. van Ardenne
Beoordelingscommissie:	Prof. dr. A. G. de Bruyn Prof. dr. S. Zaroubi Prof. dr. O. M. Smirnov

ISBN: 978-90-367-6242-7
ISBN: 978-90-367-6243-4 (electronic version)

To my dearest, my mother (1959- 2011),
and all mothers who have been least understood.

Cover: The aperture array beam projection on the celestial hemisphere using the OSKAR simulator (front). The simulated beam pattern of an HBA tile using the CAESAR simulator (back). Both are described in this thesis. Thesis cover design by Niels Bos.

Contents

Chapter 1

Introduction

Aperture arrays have become the design of choice for the low-frequency and mid-frequency parts of the world largest radio telescope, the Square Kilometre Array (SKA)[1] (Hall 2004). An excellent example of an aperture array telescope is the LOw Frequency ARray (LOFAR) (de Vos et al. 2009). It exploits the concept of a phased-array, which is a field of dipoles phased up to synthesize a receiving aperture (so-called a phased-array station shown in Fig.1.3, right panel) with a size corresponding to the dipole field. Multiple stations of this kind can be combined in the classical way to form a conventional synthesis telescope such as Westerbork Radio Synthesis Telescope (WSRT).

Operation of such an unconventional radio telescope as a scientific instrument for e.g. high quality radio imaging is more challenging than ever before (Wijnholds et al. 2010). Their field of view (FoV), sensitivity, and resolution have increased significantly as compared with conventional radio telescopes. They can detect many unknown faint sources as well as the known bright sources in their wide FoV. Their observations are corrupted by variable propagation path effects (e.g. the ionosphere and other layers of the atmosphere) and variations in their primary beam pattern (beam pattern of a phased-array station) in time and frequency. The corresponding corrections have to be made per source in the FoV, and per element of the interferometer. The large number of elements provide high sensitivity on the one hand, but poses a computational challenge in their data processing on the other hand. However, the hierarchical structure in aperture arrays, explained in Wijnholds et al. (2010) allows data reduction to some extent and more control over the data flow e.g. diagnosis of failure in the array elements, station calibration, etc.

With the advent of such new radio telescopes, the 3rd generation of calibration (3GC) has been introduced to the radio astronomical community by Noordam and Smirnov (2010) to enable us to tackle both direction-independent effects (DIE) e.g. the instrumental gain corruptions and direction-dependent effects (DDE) e.g. prop-

[1] http://www.skatelescope.org/

agation path effects and the primary beam pattern. Smirnov (2011a,b,c) has clearly explained how the 3GC trend can be adopted by radio astronomical data reduction packages. Examples of such calibration packges are blackboard self-cal (BBS) (Pandey et al. 2008) and MeqTree (measurement equation tree)[2] with additional diverse simulation and visualization capabilities.

Following the 3GC trend, Bregman (2012) has rigorously outlined the strategies for handling wide-field radio interferometric data calibration in the framework of LOFAR. Taking the ionospheric and tropospheric effects into account, he developes an optimal calibration strategy with a focus on available computational resources, but less so on optimal imaging methods. However, constraining the ionospheric parameters (van der Tol et al. 2007; van der Tol 2009; Intema et al. 2009) and practical steps such as cluster calibration (Kazemi et al. 2011) that solve the ambiguity problem and reduce the parameter space in the calibration process, remain of significant importance for ionospheric calibration.

The efficient representation of the primary beam with a few basis functions is an important step forward for fast and accurate calibration and imaging. The ongoing research aims for such an achievement (De Lera Acedo et al. 2011; Maaskant et al. 2012; Maaskant and Ivashina 2012; Craeye et al. 2012) using different type of basis functions and different consideration for mutual coupling in the models.

In addition to the calibration challenges, we should improve the deconvolution (imaging) process. In this process, the known bright sources are modelled and removed such that the faint unknown sources in the background can appear. Some calibrator sources are resolved and observed as extended structures, owing to the high resolution of the instrument. Consequently, a more accurate source modelling is required for the calibration process, which adds to the computational complexity. Therefore, we require a deconvolution process in which the deconvolution method and its source modelling scheme perform optimally and are computationally efficient.

Radio frequency interference (RFI) has become a significant issue due to the fact that the phased-array stations are sensitive to all directions and observe a wide range of frequencies. However, there are extensive bodies of research available in the field of RFI detection and mitigation, which permit obtaining scientific results in an RFI-harsh environment (Fridman and Baan 2001; Boonstra 2005; Offringa et al. 2010; Offringa 2012; Hellbourg et al. 2012).

To make aperture array telescopes a true success for the SKA, we must understand their particular engineering aspects and the corresponding challenges that they introduce in exploiting the scientific data of the instrument. Imaging is a major scientific application of these instruments, from which other modes of operation of the instrument can benefit. Therefore, an essential first step is to model the data appropriately taking into account all the corrupting factors. The next step is to utilise advanced and appropriate signal processing techniques to effectively and efficiently perform different stages of the data processing i.e. both calibration and deconvolution. Such effort was made in Lampropoulos (2010), with a particular emphasis on the use of LOFAR for detecting the signal of the epoch of reionisation (EoR). In this thesis, we continue

[2]http://www.astron.nl/meqwiki/

this approach with emphasis on the use of a general data model, redundant visibility information, a constrained calibration routine, and an optimal deconvolution process.

In the following section, we briefly explain the fundamentals of radio interferometric (aperture synthesis) image formation and its measures of quality in terms of the *calibration* and *deconvolution* processes. In Sect. 1.2, we revisit the definition of the measures of image quality i.e. the quantities that we would like to improve. Then, we give a short overview of the aperture array features in Sect. 1.3. This helps to point out how each of those new features contribute to the challenge of high quality imaging and also enhance the scientific potentials of the instrument. To achieve excellent scientific results with an aperture array telescope, these effects have to be tackled with sufficient accuracy. In Sect. 1.4, we summarize how this thesis contributes to tackling this problem in the chapters following this introduction.

We use LOFAR as an example to explain the structure of an aperture array telescope, but the arguments remain general.

1.1 Aperture synthesis imaging

The fundamentals of aperture synthesis imaging are based on measuring the spatial coherency or Fourier components of the celestial sources on different baselines in the array. The signals from the celestial sources are corrupted by the propagation path and instrumental effects leading to imperfect measurements. They are the so-called uncalibrated visibilities, which have to be corrected in a process called *calibration*.

Due to the finite number of receivers, we only sample a finite number of the Fourier components. Earth rotation aperture synthesis (Ryle 1962), allows us to sample a larger number of Fourier components given the same number of receivers, but on an irregular grid. Fig.1.1 shows an example of such sampling for the LOFAR telescope. Aperture synthesis imaging would be a simple inverse Fourier transformation of the calibrated visibilities, if they were sampled on a regular grid. Instead, we need to remove the effect of this irregular sampling pattern (often referred to as the synthesised beam or the point spread function (PSF)) from the true intensity (sky brightness distribution) in a process called *deconvolution*.

Traditionally calibration is performed by separately observing strong point sources, solving for the complex antenna gains using the known source flux density and position and then, transferring the calibration solution to the observation of the unknown field. The stability of the instrument and the atmosphere determine how frequently such calibrator observations need to be made. This procedure is commonly known as cross-calibration. In order to improve the dynamic range of an observation, it has proven possible to perform a second order calibration using the sources in the field of observation. This procedure is known as self-calibration (self-cal) (Thompson et al. 1986; Schwab and Bridle 1994). Self-cal has advanced from a process which solves for only DIE e.g. instrument electronic gains to a process which solves for both DIE and DDE e.g. the propagation path effects (Noordam and Smirnov 2010).

Self-cal and deconvolution cannot be disentangled. The self-cal process has an iterative nature and needs to converge in each iteration. It starts with an initial

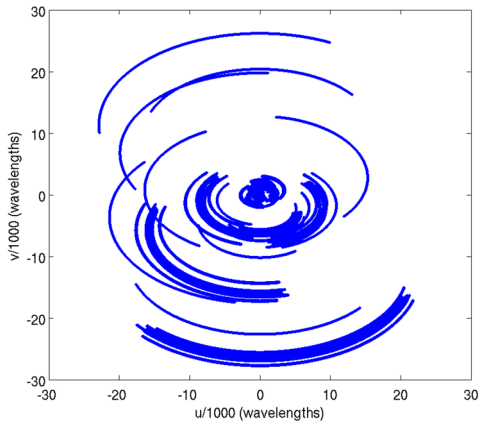


Figure 1.1: An example of irregular sampling of the Fourier components using Earth rotation synthesis technique. The observation is done at frequency $f = 214$ MHz using LOFAR.

sky model, an accurate data model and imperfectly calibrated visibilities. In each calibration iteration, the known bright sources are used to estimate the unknown parameters of instrument and propagation path effects. Then, the sky model is updated by deconvolving and removing the known bright sources such that the faint sources in the background can appear and be characterised. The calibration routine is repeated using the updated sky model and the corrected data. It is continued until the unknown parameters are calibrated for, all the known sources are removed, and ideally the thermal noise limit is reached. The remaining radio image then reveals the faint unknown background with minimum error or high quality. In the next section, we revisit the definitions of the quality measures in imaging i.e. the quantities that we would like to improve.

1.2 Measures of the image quality

Dynamic range and *fidelity* are measures of the quality of an image and are defined in the next subsections.

1.2.1 Dynamic range

The dynamic range is defined as the ratio between the peak intensity and the noise floor. As argued by Wijnholds and van der Veen (2008), the noise floor is the sum of calibration errors, the confusion noise, and the receiver thermal noise. This implies that calibration errors raise the noise floor. Therefore, one way to reduce the noise floor is to perform a more accurate calibration, as the receiver thermal noise is fixed (Yatawatta 2011b).

1.2.2 Fidelity

Essentially, image fidelity is the measure referring to high-quality reproduction of an image. It is mostly derived from the difference between the observed image, and the true image. In aperture synthesis imaging, the true image is usually unknown. Therefore, it is not straightforward to measure the image fidelity precisely this way (Wright 1999). However, it can be defined by the optimality in performing the deconvolution process, by which the true image is estimated. This optimality is measured by the Cramér-Rao bound (CRB). In estimation theory, the CRB gives a theoretical lower bound on the error variance in the estimated parameter. Therefore, reaching the CRB in the deconvolution process means achieving the image fidelity to the fullest possible i.e. the most optimal deconvolution method using the most optimal source modeling scheme. Note that we referred to the deconvolution process as the deconvolution method and its source modelling scheme.

Image fidelity clearly depends on the *uv*-coverage i.e. the pattern by which Fourier components are sampled, that is determined by the array configuration, and the duration and the bandwidth (or number of frequency channels) of the observation. But the *uv*-coverage is not the only factor. We can still optimize our source modelling scheme and methods of deconvolution (imaging), thus increasing the image fidelity. This implies that quantifying the image fidelity only based on the *uv*-coverage is inadequate. We shall therefore refer to a redefinition of image fidelity by Yatawatta (2011b), who uses information theory, based on the Landau-Pollak theorem (Landau and Pollak 1961; Slepian 1976). The Landau-Pollak theorem was derived to find the most optimal representation of the signals which should have the most concentration simultaneously in both time and frequency domains, so called *time-limited band-limited* signals (Slepian 1976), with minimum number of basis functions (minimum model order). Here, it is worth noting that in radio interferometric data, the time and frequency domains translate to image and *uv* domains, respectively. Thus, the time-limited term refers to the source limited boundary in an image (image-domain) and the band-limited term refers to the limited area of the *uv*-coverage (visibility-domain) of the interferometer, basically determined by the longest baseline.

In its simplest form, the Landau-Pollak theorem states that the number of degrees of freedom to model any given source with compact support, N_D, also known as the Shannon number (Percival and Walden 1993; Simons et al. 2006) could be no greater than the product of the area of the source in image domain, A_{lm} and the area of the source observed in the Fourier plane, A_{uv} i.e. $N_D \leq A_{lm} \times A_{uv}$. A_{uv} is limited by the *uv*-coverage whereas A_{lm} is limited by the noise floor (Slepian 1976). The noise floor is determined by self-cal accuracy, which can be improved by more optimal source models and deconvolution methods. At the same time, accurate recognition of the source area for the deconvolution process should be provided by an accurate self-cal process. This implies that image fidelity and dynamic range are intertwined as it is presented in Fig. 1.2. It also implies that we can achieve full dynamic range, even when the *uv*-coverage is poor (Yatawatta 2011b). This is made possible by using optimal source modelling schemes and deconvolution methods.

This argument make us see the concept of redundancy in a new light. Redun-

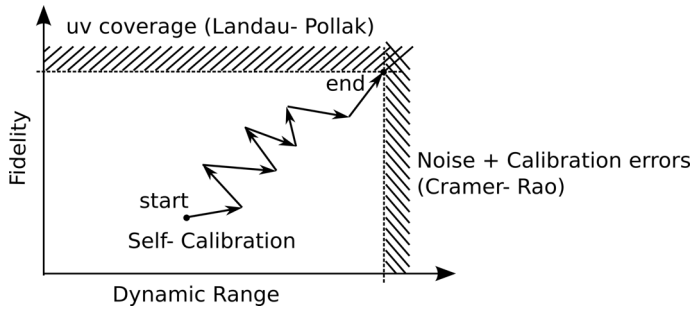


Figure 1.2: The relationship between *fidelity* and *dynamic range* in a typical radio interferometric data reduction. The theoretical limits for fidelity is mainly given by the Landau-Polak theorem. The dynamic range limit is mainly given by the Cramer-Rao lower bounds in calibration. A typical self-cal involves step-wise improvement of dynamic range and fidelity until the limits are reached (Yatawatta 2011b). Courtesy of Sarod Yatawatta.

dancy has proven to improve the dynamic range imgaing with WSRT observations (Noordam and de Bruyn 1982), but it has always been considered as a harm to the image fidelity. On the contrary to this common notion, a new deconvolution method based on using compressed sensing technique has proven to work optimally (thus, high fidelity) for sparse signals (e.g. radio interferometric observation with poor *uv*-coverage) (Candès and Romberg 2007; Levanda and Leshem 2008; Wiaux et al. 2009a; McEwen and Wiaux 2011a). However, the effect of redundancy on the performance of the deconvolution process, thus the quality of the final image using different methods of deconvolution remains a critically important future research direction.

1.3 Aperture array radio telescopes

New generation radio telescopes exploit the concept of phased-arrays i.e. the conventional reflector dish design (Fig.1.3, left panel) is replaced by a field of dipoles (Fig.1.3, right panel). The dipoles are phased up to synthesize a receiving aperture as an element of the interferometer. This is a so-called phased-array station e.g. the low band antennas (LBAs) and high band antennas (HBAs) at a LOFAR station. The type of dipole varies depending on the frequency of operation. Accordingly, the number of them, their arrangements in the field, and their hierarchy of beamforming vary to provide a certain sensitivity for the station.

Hierarchical system architecture is an important conceptual difference between the new generation of radio telescopes and conventional ones. In LOFAR, a station consists of phased-arrays, either sparse or dense. In the sparse phased-array stations, the station elements are dipoles, which are digitally beamformed to synthesize a receiving aperture as a station such as the LBAs in a LOFAR station. Phased-arrays operating above $\sim$ 100 MHz are often implemented as compound elements or tiles such as the HBAs at the LOFAR stations or in the Electronic Multi-Beam Radio As-

Figure 1.3: Left: Westerbork synthesis radio telescope as an example of a conventional radio interferometer (the photo's source: http://www.evlbi.org/network/network.html). Right: Core of LOFAR telescope including low and high frequency dipole fields (the photo's source: http://www.lofar.org/about-lofar/image-gallery/latest-lofar-images)

tronomy ConcEpt (EMBRACE) system (van Ardenne et al. 2004; Kant et al. 2011). The latter is an example of a very dense phased-array station, in which the distance between the dipoles is smaller than half the wavelength. In these stations, the station elements are phased-array tiles (e.g. Fig. 1.4, left most panel). A tile is a regular arrangement of many dipoles, whose signals are added in phase to form an instantaneous beam (the first level of beam-forming). The tile output signals are digitally phased up to synthesize a receiving aperture as a dense phased-array station. At the next level in the beamforming hierarchy, the beam formed output of each station is transported to the central correlator, in order to synthesize the whole telescope. Hierarchical structure also serves the goal of significant reduction of the data to be transferred to, and analyzed at the array central processor.

Fig. 1.4 shows the hierarchy of antenna systems in a phased-array station such as the HBA. Each HBA tile consists of 16 dual polarized identical dipoles. The output of the 16 dipoles of a certain polarization are coherently summed by an analog beamformer to represent a tile. To do this, the output of each dipole is phase shifted, and then summed with the output of all other dipoles in a tile. This makes the beam a complex quantity, that introduces direction-dependent extra phase. The output of each tile is transported to the station correlator to be correlated with the output of other tiles. The resulting correlation matrix is called station visibility which is used for engineering purposes e.g. detection of a failing element, RFI detection, station calibration, etc.

Calibration has to be performed at different levels of this hierarchy to provide a final, high dynamic range image of part of the sky as explained by Wijnholds et al. (2010). Moreover, in closely located elements of a phased-array station, electromagnetic interactions (known as mutual coupling in antenna engineering terminology) between elements leads to dissimilarity between beams of the station elements and correlated receiver noise. Each of these somehow affects the station calibration results.

A great practical advantage of the phased-array design is that the construction

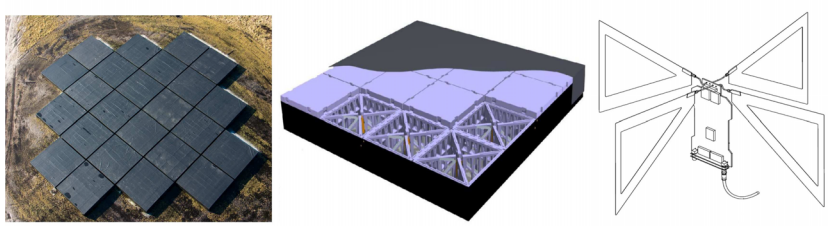


Figure 1.4: Hierarchy of antenna systems in a HBA station. Left: A 24-tiles HBA station. Middle: One HBA tile, that is consist of 16 dual polarized dipoles. Right: A mechanical design of a HBA dual polarized dipole i.e. two orthogonal bow-ties. Courtesy of LOFAR project.

cost drops dramatically and a large radio telescope becomes affordable. But operation of such an unconventional telescope is challenging in several aspects:

1. The dipoles are fixed on the ground and beam steering is done electronically. This has the advantage of fast beam steering and even having multiple beams of the same station, looking at different parts of the sky (potentially for different science cases). But, this also means that the intrinsic beam shape of a station (primary beam) keeps changing over time. The reason is that the beam steering is done electronically by assigning different beamforming weights during a certain observation. This variation is predictable in theory. In fact, there is a *dipole plus array-factor* beam model implemented in the BBS package (Pandey et al. 2008), which tracks such variation during calibration of an observation.

   Such implementation of the beam model in the BBS dismisses the mutual coupling between the dipoles. Therefore, the feasibility and implementation of a *dipole plus array-factor* model, which takes into consideration the mutual coupling effect remains as an important issue to investigate. This becomes more challenging in the case of HBA or EMBRACE like stations with extra level of beamforming hierarchy. Ivashina (2011) presented numerical examples of a multi-level hybrid beam modeling approach for aperture arrays. She concluded that only 3 basis functions are sufficient to model the beam with a good accuracy both in main beam and in sidelobes. One may consider studying how to accomodate such primary beam representation in a 3GC package, either the BBS or the MeqTree.

2. The sky rotates in the FoV of the station owing to the lack of a physical "equatorial" mount. Sky rotation in the FoV has been experienced in Very Large Array (VLA) observation with an "alt-az" (altitude-azimuth) mount contrary to WSRT observations, thanks to the equatorial mount of the telescope (see Perley et al. (1985, Lecture. 8)). Sky rotation causes a variable noise when we observe a certain patch of the sky containing faint sources, and a strong source close to the field of interest comes in and out of the rotating sidelobes. Heywood (2011) has demonstrated that such effect can be treated as a DDE and can be

solved for using the corresponding technique to handle the DDEs. However, one can argue that the rotation of the sky in the FoV is a geometrical effect, which is known a priori and should be corrected for in the software (e.g. by the BBS), instead of estimating for it as a DDE.

3. The dipoles length is seen differently from the point of view of different sources at different locations, as demonstrated in Fig. 1.5. Therefore, their sensitivity varies per source. Due to the same reason, the projection of the beam pattern of a phased-array station (primary beam) on the celestial hemisphere varies from one scanning angle to another, unlike a dish with equatorial mount, whose beam projection remains the same for all directions of observation. Fig. 1.6 demonstrates this effect using the OSKAR[3] simulator. It visualizes the projected beam of a phased-array station, formed toward different directions on the celestial hemisphere.

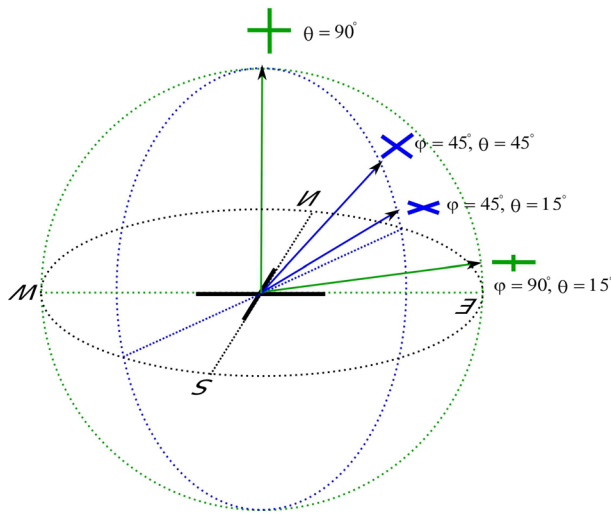


Figure 1.5: The coordinate system, in which locations of the station antenna elements can be expressed for aperture array stations. The x-axis is oriented towards east, the y-axis is oriented towards north and the z-axis is pointing at the local zenith. This illustration is courtesy of Oleg Smirnov.

In the case of dismissing the effect of beam projection, an un-polarized signal is observed as a polarized one, as demonstrated in Lampropoulos (2010, Fig. 3.7) using MeqTree simulation module. Therefore, the projection effect can be seen as the instrumental polarization (Smirnov 2011b). However, if an accurate model of the beam of a phased-array station, that takes the projection effect into account is provided, this effect should not be modelled as a separate effect i.e. the instrumental polarization. Such consideration also reduces the parameter

[3]http://www.oerc.ox.ac.uk/research/oskar/

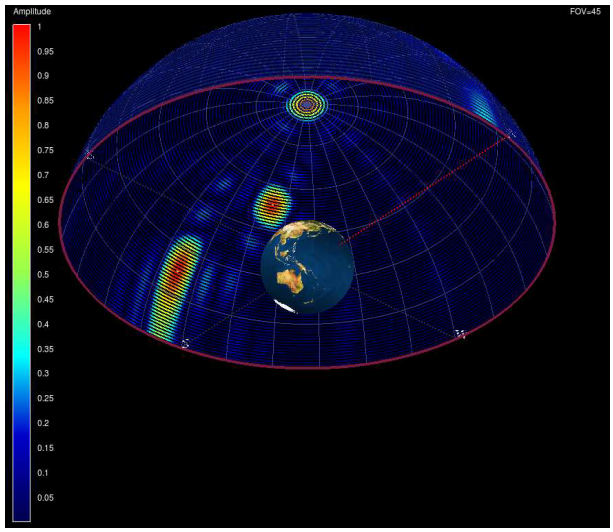


Figure 1.6: Demonstration of the aperture array beam projection on the celestial hemisphere using the OSKAR simulator.

space in calibration, thus improves the calibration efficiency, effectiveness, and speed of convergence.

Ongoing research on efficient representation of the primary beam with a few basis functions (De Lera Acedo et al. 2011; Maaskant et al. 2012; Maaskant and Ivashina 2012; Craeye et al. 2012) can alleviate this problem eventually.

The antenna pattern of a dipole is largely isotropic and hence it can observe the entire sky. A number of them, that are beamformed to represent a phased-array station give a narrower FoV but still large as compared with the FoV of a conventional dish (especially at low frequencies). Through this wide FoV, a large part of the sky is observed. This is scientifically fascinating as we can do more science given the same observing time, but it also means that:

1. the sky is observed through a patch of the ionosphere with spatial irregularity on a scale (S in Fig. 1.7), much smaller than the FoV (V in Fig. 1.7) i.e. calibration falls in the regimes 3 and 4, as illustrated in Fig. 1.7 and explained in Lonsdale (2004). In regime 3, the ionosphere varies significantly over the FoV, but similarly for all the array elements. In regime 4, different array elements view different sections of the ionosphere. In both regimes, the ionosphere is also variable over time and frequency. Aperture array calibration may also fall in regime 1 or regime 2, obviously depending on the size and configuration of the phased-array stations, and the frequency of observation that determine the width of the FoV.

   The propagation path effects and ways to calibrate for them in low frequency

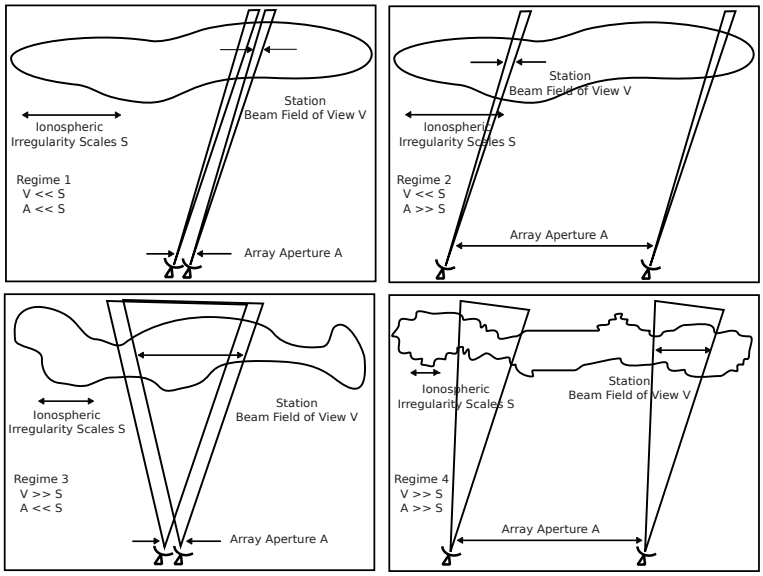


Figure 1.7: Calibration regiems instroduced by Lonsdale (2004).

and wide FoV aperture synthesis data have been extensively discussed by van der Tol et al. (2007); van der Tol (2009); Intema et al. (2009); Bregman (2012). van der Tol et al. (2007); van der Tol (2009) proposed a self-cal process using physically justifiable extra constraints on the ionospheric parameters and the peeling technique (Noordam 2004) to calibrate for such parameters. Peeling technique is sequential self-cal on individual bright sources in the FoV. Their approach was developed further as a technique called source peeling and atmospheric modeling (SPAM) explained in Intema et al. (2009). Bregman (2012) has thoroughly explained the propagation path effects and ways to tackle them in detail. He indicated the direction for futher research in this field.

In addition, cluster calibration (Kazemi et al. 2011), which has a similar approach to the peeling technique constitutes a general amendment to wide FoV calibration. It divides the observed sky into clusters centered at strong sources in the FoV. It assumes certain parameters e.g. the ionospheric effects to be similar for the other weaker sources in the same cluster. This reduces the parameter space in calibration, which is a major contribution to manageable data calibration in wide-field astronomy.

2. the curvature of the celestial sphere is always present in aperture array observations owing to their wide FoV. This is the case regardless of the coplanarity of the baselines. The sky curvature in aperture synthesis observations has been the subject of discussion since the 70's (Brouw 1971; Clark 1973). For an east-west array this is not a problem as nicely shown by Brouw (1971). Otherwise, it is

ignored on co-planar baselines given the narrow FoV of the conventional radio telescopes (and observations in higher frequencies), while it has commonly been tackled using facet-imaging (Perley and Bridle 1994) and w-projection methods (Cornwell et al. 2008), on non-coplanar baselines. w-projection algorithm, that performs faster than facet-imaging converts visibilities in 3D-space, $\widetilde{f}(u,v,w)$ to visibilities in 2D-space $\widetilde{f}(u,v,w=0)$ by projecting out the w-term to $w=0$. This method has clearly been explained in Cornwell et al. (2008). Thus, aperture synthesis imaging has always been done in a 2D-plane. The reasons were the convenience of image display in 2D, and the high computational cost of a full 3D FFT.

The latest work in this area discusses the advantages of doing the data analysis in 3D, on celestial sphere (Carozzi and Woan 2009; McEwen and Wiaux 2011b; Bregman 2012). Carozzi and Woan (2009) suggested performing calibration on celestial hemisphere in (l,m,n) coordinates, as none of the assumptions of the narrow FoV and coplanary holds for new radio telescopes such as LOFAR. This way, all-sky polarimetric calibration and imaging in a single telescope pointing are possible without any step such as the sky faceting or w-projection. In the deconvolution field, there is a growing interest in application of the compressed sensing techniques in radio interferometric imaging owing to its linearity. Its application for radio interferometric image deconvolution was introduced and demonstrated by Levanda and Leshem (2008); Wiaux et al. (2009a). It was enhanced further in research by McEwen and Wiaux (2011b,c,a), where its improved performance in wide-field imaging in the celestial sphere (without w-projection step) was demonstrated.

Moreover, significant ongoing research has focussed on improving the computational efficiency of higher order FFT e.g. 3D FFT (Frigo and Johnson 1998; Frigo 1999; Frigo and Johnson 2005)[4], allowing to reduce the computation cost and speed.

However, this is arguable whether the 3D imaging on celestial sphere is the potential future of aperture synthesis imaging.

All the aforementioned properties of the new radio telescopes affect their imaging capabilities. The signals of interest experience all these effects, which we must take into account in the data model properly. Then, we should correct for each of such effects accurately to obtain the final image with minimum error, thus high dynamic range.

In addition to the calibration challenges, we should improve the deconvolution process. In this process, the known bright sources are modelled and removed such that the faint unknown sources in the background can appear. Some calibrator sources are resolved and observed as extended structures, owing to the high resolution of the instrument. Extended sources should be modelled and removed optimally, which leads to simultaneous improvement in both fidelity and dynamic range of the final image as illustrated in Fig. 1.2.

[4]See the FFTW home page for the latest updates: http://www.fftw.org/

CLEAN (Högbom 1974) is the first and most commonly used deconvolution method. It has greatly served aperture synthesis imaging over the last decades. Its performance, however degrades during the deconvolution of extended sources, owing to its pixel-based source modelling scheme and the nominal resolution (or the PSF width) limit of the interferometer. Fundamental limitations of CLEAN for deconvolution of an extended source were first discussed by Brinks and Shane (1984). However, they had not recognized its pixel-based source modelling scheme as the cause of that limitation. As an improvement to CLEAN performance, Wakker and Schwarz (1988) developed a multi-resolution CLEAN (MRC) by appropiately combining the results of a number of conventional CLEAN operations with optimized parameters, each at a different scale. Later Cornwell (2004); Bhatnagar and Cornwell (2004) recommended a scale sensitive pixel-based source modelling scheme by representing the extended source with a number of properly scaled and centered extended basis function components. However, one can argue that a scale sensitive pixel-based source modelling would still be limited by the resolution limit of the interferometer, when the source extension is larger than the PSF width. Moreover, they acknowledged that they cannot reliably specify the number of significant basis function components (model order). An optimal alternative source modelling scheme is the use of orthonormal basis functions that span the whole source area e.g. shaplets or prolate spheroidal wave functions (PSWF). PSWF have proven to be the only set of basis functions that provide the significant number of basis functions (Shannon number) and perform more effectively than other basis functions e.g. shapelets (Yatawatta 2011b). However, they are computationally more expensive than shapelets. One can compute them efficiently taking the approach suggested in Noorishad and Yatawatta (2011).

1.4 Outline and summary of the thesis

The scope of this thesis is to take a few steps towards high quality imaging with aperture array telescopes, taking into consideration the computational efficiency of each step.

In Chapter 2, we present and refine a general data model, in which the nature of the noise terms and the direction dependent effects are well described. This view reveals some potential improvements, that we suggest to be considered in algorithm development e.g. for the calibration and deconvolution processes. In modelling the data of a phased-array station, we also take into account the mutual coupling between the array elements. We view the mutual coupling effect, both in terms of mutual impedances and as the cause of dissimilarity between the station elements beam pattern. We discuss how each view leads to a different model of the same physical phenomenon i.e. electromagnetic interaction between the array elements. We use the refined data model in the chapters that follow Chapter 2.

In Chapter 3, we study the application of the redundancy method for the calibration of phased-array stations with a regular layout e.g. HBA (Fig.1.4, left-most panel). The initial motivation was that it is computationally cheaper and faster than the commonly used model-based calibration method and it is sky model independent.

Our results using real and simulated data of LOFAR HBAs demonstrate that the redundancy calibration method performs better than model-based calibration method, when redundancy is provided. Therefore, we recommend designing redundancy (or at least exploiting the redundant mesurements) in both the station layout and the array configuration of future aperture arrays. This is particularly true in the case of the SKA aiming at a dynamic range requirement that surpasses any existing array by an order of magnitude (Dewdney et al. 2010). In addition, redundancy provides a powerful tool, from which system diagnostics can be obtained. We also conclude that mutual coupling is the main limitation of the calibratability of a phased-array station, regardless of the calibration method. For this reason, we recommend minimization of the mutual coupling as a figure of merit in station configuration.

In Chapter 4, we present the refractive nature of the ionosphere as a DDE and a source of inaccuracy in self-cal gain estimation results, assuming that the beam model is known. We take advantage of the practical solution of cluster calibration (Kazemi et al. 2011) and suggest a similar approach to that of van der Tol et al. (2007) such that self-cal converges better to the true solution, in presence of the refractive effect of the ionosphere. Our suggestion is to solve a constrained self-cal for an array with an arbitrary geometry, using extra constraints as exploited in redundancy calibration. We demonstrate that those constraints have a clear physical justification, only when we observe a single source or multiple sources, that experience similar ionospheric effect. We discuss how cluster calibration might reduce any other observational scenario (shown in Fig. 1.7) to one, in which these constraints are justified. We conclude that a constrained self-cal performs more optimally based on our numerical CRB analysis and comparing it with the CRB of its unconstrained version. Then, we solve a constrained self-cal taking two different approaches: the sequential quadratic programming (SQP) technique and the logarithmic least squares (LOGLS). Their results show that the SQP performs more accurately as compared with the LOGLS, provided the same initial guess, and that self-cal can benefit from the constraints used in the redundancy calibration formalism to perform more reliably for gain calibration of an array with an arbitrary geometry.

In Chapter 5, we discuss the limitations of a pixel-based source modelling scheme for modelling a source which is more extended than the PSF. We also discuss the alternative use of orthonormal basis functions, that span the whole source area as opposed to the pixel-based basis functions, for optimal modelling of the extended sources in the deconvolution process. We give an overview of existing deconvolution methods and discuss how they can accommodate such source modelling scheme. Following the promising results of using orthonormal basis functions, especially the PSWFs (Yatawatta 2010, 2011b,a), we suggest a solution to reduce the computational cost of their implementation (Noorishad and Yatawatta 2011). Moreover, our extensive discussion on the deconvolution problem, the corresponding methods and source modelling schemes reveals that quantifying the image fidelity based on only *uv*-coverage is inadequate.

Finally in Chapter 6, we summarize the insights, that we obtained throughout this thesis to understand the aperture synthesis data processing of an aperture array telescope. We also list several new questions to be pursued in new research projects. The

clear answer to those questions will hopefully be a major step toward computationally effective and efficient wide-field, high quality aperture synthesis imaging.

Chapter 2

Array data model

In this chapter, we present the array data model (also known as the measurement equation (ME)), which we will use in the next three chapters. It is essential to have an accurate understanding of various physical and geometrical effects that contribute in the array data formation and accommodate them in the data model. Therefore, we dedicate this chapter to precisely present and refine a general data model, in which the nature of the noise terms and the direction-dependent effects are well described. This view reveals some potential improvements, that we suggest to be considered in our algorithms development e.g. for the calibration and deconvolution processes.

2.1 Data model of a phased-array station

A phased-array station is an array element in an aperture array telescope such as LOFAR. It actually synthesizes a receiving aperture, which substitudes a dish in the conventional arrays. It is important to note that data formation both at a phased-array station and at the whole synthesis array is based on the same principles. However, some physical effects gain or lose their importance only due to the parameters e.g. integration time, baseline length, and electromagnetic environment. In this section, we obtain the data model or so-called measurement equation (ME) of a phased-array station, in which all the array elements are included (a full ME). Mutual coupling, which is not a dismissable effect in the phased-array stations is taken into account as a matrix of mutual impedances in the data model in Eq. 2.11. The mutual coupling is also studied as the cause of dissimilarities between the station element beams in Sect. 2.1.3. We explain how these dissimilarities can be taken into account as a DDE in a gain matrix $\mathbf{G}_0$, in the data model in Eq. 2.16. In Sect. 2.1.2, we describe a general noise expression for an array, where the nature of noise in the aperture synthesis data is properly defined in terms of the different contributing factors.

In Sect. 2.2, we explain the reasons why we use a data model based on 2×2 Jones matrices at the whole synthesis array level. We note that 2×2 MEs are in fact, the off-diagonal terms of the full ME, when the full polarization content of incoming celestial signals is taken into account.

2.1.1 Data formation

We assume that we have a phased-array station such as a LBA station of LOFAR presented in Fig. 2.1. The electromagnetic field probed by an antenna at position $\mathbf{r} = (x, y)$ at time t can be described as a superposition of electromagnetic plane waves, $\mathbf{E}_s(\mathbf{k}, t)$ coming from celestial sources at all directions in different frequencies. This is represented in Eq. 2.1.

Figure 2.1: An LBA station of LOFAR as an example of a phased-array station, with the simplest antennas, i.e. droopy dipoles (the photo source: http://www.astron.nl/). The photo is courtesy of Hans Hordijk.

$$\mathbf{E}(\mathbf{r},t) = \int_{-\pi}^{\pi}\int_{0}^{\frac{\pi}{2}}\int_{-\infty}^{\infty} \mathbf{E}_s(\mathbf{k},t)e^{-j(\mathbf{k}.\mathbf{r}-|\mathbf{k}|ct)}df|sin\theta|d\theta d\phi. \tag{2.1}$$

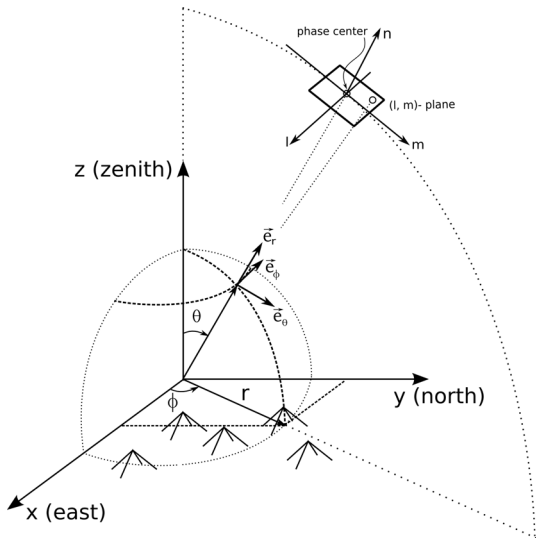


Figure 2.2: Coordinate system to describe the antenna positions and the polarization of the incoming plane waves. The antennas i.e. dipoles are arranged along x- and y- directions to measure different polarization components. (l, m)- plane is tangential to the celestial hemisphere at the phase center.

The wave vector $\mathbf{k} = -2\pi f\boldsymbol{\ell}_s/c$ fully characterizes the plane wave coming from an arbitrary direction $\boldsymbol{\ell}_s = (l, m, n)$, with frequency f. The coordinate system to describe the antenna positions and the polarization of the incoming plane waves is presented in Fig. 2.2. The following coordinates conversions hold: $l = \sin\theta\cos\phi$, $m = \sin\theta\sin\phi$, and $n = \cos\theta$, where $\theta = 90° - el$, $\phi = 90° - az$, el and az denote elevation and azimuth, respectively. The plane wave spectrum $\mathbf{E}_s(\mathbf{k}, t) = [e_\theta, e_\phi, e_r]^T$ describes the complex amplitude distribution along the three polarization directions $\mathbf{e}_\theta$, $\mathbf{e}_\phi$ and $\mathbf{e}_r$. Since a plane wave does not have a field component along the direction of propagation, the field amplitude can be described by just two components, $\mathbf{E}_s(\mathbf{k}, t) = [e_\theta, e_\phi]^T$. The plane waves may come from the following origins:

1. known sources (presumably point-like sources), which are used as calibrator beacons.

2. extended structures such as the galactic plane and north polar spur, which are sampled predominantly on short baselines i.e. the baselines $< 4\lambda$ (λ is the observed wavelength). This upper bound is set empirically and has proven practical by Wijnholds (2008).

3. homogeneous sky background noise.

4. RFI sources in the far-field of the array. We note that the assumption of a plane wave does not hold for an RFI signal in the near-field.

To ease the discussion, that follows, we make several simplifying definitions and assumptions:

1. The plane waves originate from only point-like sources (i.e. a source with compact support) in the sky. A phased-array station does not have enough resolution to resolve the sources owing to its short baselines.

2. The station antenna elements are simple dipole antennas, arranged along orthogonal x- and y- axis to measure different polarization components (see Fig. 2.2). One can construct the polarization information content of the measurement using these two components.

3. We take into account only the dipoles arranged along one axis, either x-axis or y-axis. This way, we can drop some mathematical complexities e.g. the vector $\mathbf{E}$ in Eq. 2.1 is replaced by a scalar signal, $s(\mathbf{k}, t)$.

4. A station element is referred to an antenna including its corresponding electronics. The station elements are assumed to be identical.

5. The propagation path effects can be dismissed, as the station data is snapshot, with short integration time e.g. one second.[1]

6. We observe in only a narrow frequency subband (channel), thus the inner-most integral in Eq. 2.1 drops.

A plane wave with wave vector $\mathbf{k} = -2\pi f \boldsymbol{\ell}_s / c$ impinging on an array element at position, $\mathbf{r}$ and at time, t is decribed by

$$s(\mathbf{r}, \mathbf{k}, t) = s(\mathbf{k}, t) e^{-j(\mathbf{k}^T \mathbf{r} - |\mathbf{k}| ct)}. \tag{2.2}$$

This plane wave will induce a voltage in each dipole, $v_{wave_i}(\mathbf{k}, t)$. It can be represented in a *Thévenin* equivalent circuit of dipole element number i depicted in Fig. 2.3 (Svantesson 1998). v_{ind} is the voltage induced in the i-th dipole due to the impinging wave and electromagnetic interaction with other dipoles. The receiver device is characterized by an electronic gain, g and the impedance Z_T. Z_A is the antenna impedance, whose definition can be found in any antenna literature e.g. (Balanis 2005). v_T, the voltage over the termination impedance Z_T, is the measured output voltage of the station element number i.

The total induced voltage in the i-th antenna element is then given by

$$v_{ind_i}(\mathbf{k}, t) = v_{wave_i}(\mathbf{k}, t) - \sum_{k \neq i} I_k Z_{ik}, \tag{2.3}$$

[1] Although there is no guarantee that those effects remain unchanged during capturing the snapshot data in all frequency subbands, which takes about 9 minutes for a LOFAR station.

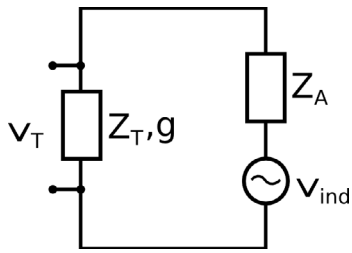


Figure 2.3: *Thévenin* equivalent circuit model of the i-th element of a station (Svantesson 1998).

where the first term is the voltage induced by the plane wave and the second term is the induced voltage due to electromagnetic interaction between antenna i and k. The latter effect is known as mutual coupling and is expressed in terms of a mutual impedance Z_{ik}. I_k is the current at antenna k. Analytical expressions for both terms are obtained by integrating over the surface of the antenna element. Such an expression is somewhat trivial for a dipole but once the antenna element corresponds to a surface,[2] such expression gets a more complicated form (Maaskant 2010). A closed form expression of a dipole beam pattern,[3] $H_i(\boldsymbol{\ell}_s)$ can be found in any antenna reference book e.g. (Balanis 2005). The induced voltage due to the plane wave signal coming from direction $\boldsymbol{\ell}_s$, $s(\mathbf{r}_i, \mathbf{k}, t)$ is simply given by

$$v_{wave_i}(\mathbf{k}, t) = s(\mathbf{r}_i, \mathbf{k}, t) H_i(\boldsymbol{\ell}_s). \tag{2.4}$$

$H_i(\boldsymbol{\ell}_s)$ is the beam pattern of an individual element when it is not electromagnetically interacting with any other element in the array. In an array made of mechanically and electromagnetically identical elements, $H_i(\boldsymbol{\ell}_s)$ is identical for all the elements. It is also assumed to be known from electromagnetic studies of an antenna.

Assuming that we have p elements in an array, we write the circuit equations for all antenna elements using the model in Fig. 2.3. It will result in a system of equations, which can be expressed in form of matrix multiplications

$$\mathbf{v}_T(\mathbf{k}, t) = \mathbf{G} Z_T \mathbf{Z}^{-1} \mathbf{v}_{wave}(\mathbf{k}, t), \tag{2.5}$$

where vector $\mathbf{v}_T = [v_{T1}, v_{T2}, ..., v_{Tp}]^T$ (size $p \times 1$) contains the measured output voltages. $\mathbf{G} = \mathrm{diag}(\mathbf{g})$ (size $p \times p$), $\mathbf{g} = [g_1, g_2, ..., g_p]^T$ (size $p \times 1$) contains the intrinsic complex gain of the reciever elements. $\mathbf{Z}$ (size $p \times p$) is an impedance matrix whose diagonal elements are $Z_{ii} = Z_A + Z_T$ and off-diagonal elements are mutual impedances (Z_{ij}). Vector $\mathbf{v}_{wave} = [v_{wave1}, v_{wave2}, ..., v_{wavep}]^T$ (size $p \times 1$) contains the voltages induced by the plane wave on all antenna elements.

[2]Such as in the case of HBAs in LOFAR. There, an antenna element is a bow-tie (see Fig. 1.4, the right most panel)

[3]Beam pattern is also known as response or sensitivity pattern.

The plane wave signal is assumed to be narrow-band (Zatman 1998). This allows us to present its different times of arrival at each antenna (located at $\mathbf{r}_i$) with a different phase shift. This is represented as geometrical array-factor in

$$\mathbf{v}_T(\mathbf{k},t) = \mathbf{G}\underbrace{(Z_T + Z_A)\mathbf{Z}^{-1}}_{\mathbf{M}}\hat{H}(\boldsymbol{\ell}_s) \underbrace{\begin{bmatrix} e^{-jk\mathbf{r}_1.\boldsymbol{\ell}_s} \\ e^{-jk\mathbf{r}_2.\boldsymbol{\ell}_s} \\ \vdots \\ e^{-jk\mathbf{r}_p.\boldsymbol{\ell}_s} \end{bmatrix}}_{\text{geometrical array-factor}} s(\mathbf{k},t), \tag{2.6}$$

where $\hat{H}$ is the normalized response pattern. The mutual coupling effect is described by the unit-less matrix, $\mathbf{M}$ (size $p \times p$). If no coupling is present, it will reduce to an identity matrix. Often the geometrical array-factor and the normalized response pattern are combined into an array response vector (or steering vector), $\mathbf{a}(\boldsymbol{\ell}_s, \mathbf{r})$

$$\mathbf{v}_T(\mathbf{k},t) = \mathbf{GMa}(\boldsymbol{\ell}_s,\mathbf{r})s(\mathbf{k},t). \tag{2.7}$$

We note that the array response vector carries the information of the element beam patterns.

As formulated earlier in Eq. 2.1, the electromagnetic field probed by the antennas is a superposition of electromagnetic plane waves coming from celestial sources at all directions. If we take into account the contribution of q point sources illuminating the array at time t, we obtain

$$\mathbf{v}_T(t) = \mathbf{GMAs}(\mathbf{k},t), \tag{2.8}$$

where $\mathbf{s}(\mathbf{k},t) = [s_1(\mathbf{k}_1,t), s_2(\mathbf{k}_2,t), ..., s_q(\mathbf{k}_q,t)]^T$ (size $q \times 1$) contains the plane waves or i.d.d. (temporally independent and identically distributed) Gaussian signals coming from q known celestial sources. $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_q]$ (size $p \times q$) is the array response matrix, that contains the array response vectors corresponding to each source on its columns and is defined as

$$\mathbf{A} = exp\left(-j\frac{2\pi}{\lambda}\mathcal{R}\mathcal{L}^T\right), \tag{2.9}$$

where the matrix $\mathcal{R} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_p]^T$ (size $p \times 3$) is the stack of array element positions, if the i-th element of the array is located at $\mathbf{r}_i = [x_i, y_i, z_i]^T$. Likewise, the source positions are stacked in a matrix $\mathcal{L} = [\boldsymbol{\ell}_{s1}, \boldsymbol{\ell}_{s2}, ..., \boldsymbol{\ell}_{sq}]^T$ (size $q \times 3$), if the position of the k-th calibrator source is $\boldsymbol{\ell}_{sk} = [l_k, m_k, n_k]^T$.

The elements of vector $\mathbf{v}_T(t)$ in Eq. 2.8 are the superposition of voltages induced by electromagnetic waves coming from q directions at time t and the voltages induced due to mutual coupling between the station elements. In real observations, $\mathbf{v}_T(t)$ also carries the contribution of the homogeneous sky background noise (T_{sky}) and the unresolved sky (extended sources), which experiences the same mathematical operations as the q celestial signals from the strong point sources. The difference is

that the summation (implemented in matrix multiplication) has to be replaced by an integral. We will explain about the sky noise contribution in Sect. 2.1.2. The receivers thermal noise signals (also assumed to be i.i.d. Gaussian), denoted with vector $\mathbf{n}$ adds to $\mathbf{v}_T(t)$.

$\mathbf{v}_T(t)$ is sampled with period T. The n-th sample of the array signal vector $\mathbf{x}[n]$ is given by

$$\mathbf{x}[n] = \sum_{-\infty}^{\infty} (\mathbf{v}_T(t) + \mathbf{n})\delta(t - nT) = \mathbf{v}[nT] + \mathbf{n}[nT]. \tag{2.10}$$

N samples can be stacked in a matrix $\mathbf{X} = [\mathbf{x}[1], \mathbf{x}[2], ..., \mathbf{x}[N]]$ (size $p \times N$) which denotes the short term integration data set or snapshot.

Radio interferometric measurements are based on measuring the spatial coherency between antenna output signals, so-called visibilities. The array covariance matrix (or correlation between all sampled voltages), $\mathbf{R} = \mathcal{E}\left\{\mathbf{x}[n]\mathbf{x}^T[n]\right\}$, can be estimated by $\hat{\mathbf{R}} = \mathbf{X}\mathbf{X}^T/N$. Its expected value becomes

$$\mathbf{R} = \mathbf{G}\mathbf{M}\mathbf{A}\boldsymbol{\Sigma}_{\mathrm{s}}\mathbf{A}^H\mathbf{M}^H\mathbf{G}^H + \underbrace{\mathbf{G}\mathbf{M}\mathbf{A}\boldsymbol{\Sigma}_{\mathrm{sky}}\mathbf{A}^H\mathbf{M}^H\mathbf{G}^H + \mathbf{R}_{\mathrm{rec}}}_{\boldsymbol{\Sigma}_{\mathrm{n}}}, \tag{2.11}$$

where $\boldsymbol{\Sigma}_{\mathrm{s}} = \mathrm{diag}(\boldsymbol{\sigma}_s)$ (size $q \times q$) is sources spatial coherency or covariance matrix. $\boldsymbol{\sigma}_s$ is a vector of the source fluxes. $\mathbf{R}_{\mathrm{rec}}$ is the receiver noise covariance matrix. $\boldsymbol{\Sigma}_{\mathrm{n}}$ is a general expression of the noise terms, which is discussed in the next subsection. Correlator intrinsic off-sets can be represented as an additive term to the covariance matrix i.e. as a non-diagonal matrix. However, in Eq. 2.11 and throughout this document, it is disregarded assuming that the correlator is designed perfectly.

As it was mentioned earlier, phased-array stations are especially sensitive to extended structures owing to their many short baselines. Since modelling them is computationally hard and expensive, they are not used as calibrator beacons. Wijnholds and van der Veen (2009b) phenomenologically model them as correlated noise. This noise is in addition to the other sources of noise in $\boldsymbol{\Sigma}_{\mathrm{n}}$ e.g. the sky noise and the receiver thermal noise, which are discussed in the next subsection.

Station visibilities ($\mathbf{R}$ in Eq.2.11) are used for engineering purposes e.g. RFI detection, station calibration.

We note that RFI sources in the far-field can be modelled in the same way as celestial point sources, although they may have different statistical properties. However, the RFI sources either in the far-field or the near-field should be detected and mitigated before calibration. Extensive studies have shown that this is possible both at a phased-array station and at the whole synthesis array (Fridman and Baan 2001; Boonstra 2005; Offringa et al. 2010; Hellbourg et al. 2012; Offringa 2012) (and many references therein).

2.1.2 Noise terms

The noise correlation matrix, $\boldsymbol{\Sigma}_{\mathrm{n}}$ in Eq. 2.11 can generally be represented by the following terms

$$\mathbf{\Sigma}_{\mathrm{n}} = \mathbf{R}_{\mathrm{rec}} + \mathbf{R}_{\mathrm{sky}} + \mathbf{R}_{\mathrm{sp}}, \tag{2.12}$$

where $\mathbf{R}_{\mathrm{rec}}$, $\mathbf{R}_{\mathrm{sky}}$, and $\mathbf{R}_{\mathrm{sp}}$ are the correlation matrices of the receivers thermal noise, the background sky noise, and the spillover noise, respectively.

The low noise amplifiers (LNAs) connected to the antennas in a strongly coupled array, generate noise waves towards their outputs, but also send reflected noise waves back into the antenna array. These waves are then coupled into other receiver channels, giving rise to a correlated noise contribution, represented by the non-diagonal matrix, $\mathbf{R}_{\mathrm{rec}}$. It is clear that depending on the antenna and LNA designs, sparsity and density of the station elements layout, the term $\mathbf{R}_{\mathrm{rec}}$ may be reduced to a diagonal matrix, whose diagonal terms are noise power of i.i.d. Gaussian thermal noise signal of the receivers.

$\mathbf{R}_{\mathrm{sky}}$, is defined as

$$\mathbf{R}_{\mathrm{sky}} = \mathbf{GMA}\mathbf{\Sigma}_{\mathrm{sky}}\mathbf{A}^H\mathbf{M}^H\mathbf{G}^H, \tag{2.13}$$

where $\mathbf{\Sigma}_{\mathrm{sky}}$ represents the sky noise coherency. Its elements, $\Sigma_{\mathrm{sky}}^{ij}$ are obtained by integrating the homogeneous sky noise temperature, $T_{\mathrm{sky}}(\Omega)$ through a pattern-overlap integral between the i^{th} and j^{th} array element patterns (Ivashina et al. 2011)

$$\Sigma_{\mathrm{sky}}^{ij} = \frac{k_B}{2\pi^2\eta}\int_{\Omega} T_{\mathrm{sky}}(\Omega)[\hat{H}_i(\Omega)\hat{H}_j(\Omega)]d\Omega, \tag{2.14}$$

where Ω is the FoV, $k_B = 1.38\times10^{-23}$ J/K is Boltzmann constant. η is the free-space impedance. $T_{\mathrm{sky}}(\Omega)$ is assumed to be stationary and constant in $0 < \theta < \pi/2$ and all values of ϕ. Furthermore, it is a function of frequency (see Fig. 2.4) described by the empirical equation in

$$T_{\mathrm{sky}}(f) = T_{1\mathrm{K}}\left((\frac{c}{f.l_0})^{2.55} + (\frac{f}{f_0})^{1.8}\right) + T_{bg}, \tag{2.15}$$

where $c = 3\times10^8$ m/s is the speed of light in vacumu, $T_{1\mathrm{K}}$ is a constant equal to 1 Kelvin, l_0 is a constant equal to 0.2008 m, f_0 is a constant frequency equal to 10 GHz and $T_{bg} = 2.7$ K is the cosmic background noise temperature.

$\mathbf{R}_{\mathrm{sp}}$ can be modelled in the same way as $\mathbf{R}_{\mathrm{sky}}$, but $T_{\mathrm{sky}}(\Omega)$ in Eq. 2.14 should be set as only the ground noise ($\pi/2 < \theta < \pi$ and all values of ϕ). The ground noise is generally set to have constant value of 300 K (Cortès Medellìn 2007). However, $\mathbf{R}_{\mathrm{sp}}$ is negligible as compared with $\mathbf{R}_{\mathrm{sky}}$ and $\mathbf{R}_{\mathrm{rec}}$ in general, and in aperture arrays, in particular.

Depending on the frequency of observation, the antenna and receiver design, and the sensistivity of the array, the term $\mathbf{R}_{\mathrm{sky}}$ may or may not dominate the $\mathbf{R}_{\mathrm{rec}}$. van Cappellen et al. (2004) have shown how an array of short dipoles becomes sky-noise limited at low frequencies (see Fig. 2.4), such that the term $\mathbf{R}_{\mathrm{rec}}$ in Eq. 2.12 becomes negligible.

The noise model in Eq. 2.12 has recently been presented in the literature concerning phased-array feeds (PAF) for example by Jeffs et al. (2008) and Ivashina et al.

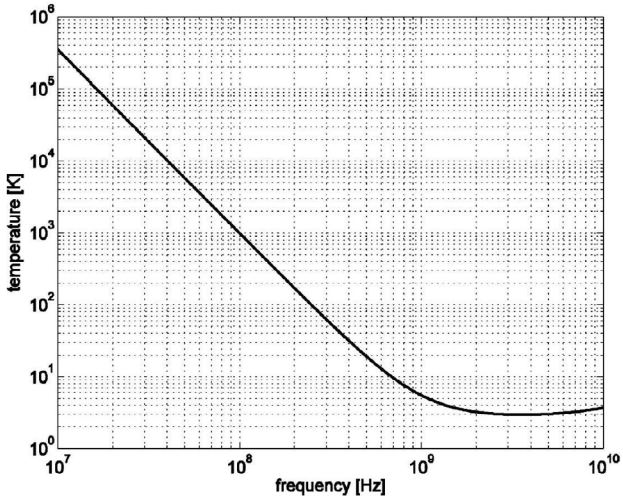


Figure 2.4: Model of the sky noise temperature (van Cappellen et al. 2004).

(2011). But, it can be used as a generic noise model for any antenna array system (Maaskant 2010). Cortès Medellìn (2007) has studied these terms in more details.

Statistical distribution of noise in data model is essential for the CRB derivation and calibration. Usual assumption has always been a Gaussian distribution for the noise, which may not hold. The data that we exploit for e.g. calibration are the visibilities ($\mathbf{R} = \varepsilon\left\{\mathbf{x}[n]\mathbf{x}^T[n]\right\}$), which are the correlation of two Gaussian signals ($\mathbf{x}[n]$ in Eq. 2.10). It is known that the correlation of two arbitrary Gaussian signals has a Wishart distribution. Thus, the elements of noise correlation matrix ($\mathbf{\Sigma}_{\mathrm{n}}$), that are the correlation of two Gaussian signals (station element output) have a Wishart distribution in Eq. 2.11.[4] This fact has also been noted in van der Tol and Wijnholds (2006).

In absence of the mutual noise coupling, the thermal noise of each receiver (term $\mathbf{n}$ in Eq. 2.10) is assumed to be a i.i.d. Gaussian signal, i.e. $\mathbf{R}_{\mathrm{rec}}$ is diagonal. Receiver noise coupling means that a fraction of thermal noise on one receiver is added to the thermal noise of another receiver. It can be shown that noise on each station element has still a normal distribution. This is plausible in view of large number of samples (N) and according to the central limit theorem. However, in either case the noise becomes of a Wishart distribution after correlation in Eq. 2.11.

At the synthesis array level, correlated receiver noise is very unlikely and $\mathbf{\Sigma}_{\mathrm{n}}$ carries the sky noise power on its off-diagonal and the receiver noise power on its diagonal elements. At this level, one might integrate the data longer than that of snapshot data (e.g. station visibilities). The noise pdf may or may not converge to a Gaussian distribution again, depending on the number of visibility samples that are averaged. Potential non-Gaussianity of the noise in the data model implies that we should

[4]Given that signal and noise are assumed to be independent.

reconsider some fundamental assumptions in our calibration algorithms especially in the ones that exploit the statistical properties of the noise e.g. (Yatawatta et al. 2008; Kazemi et al. 2011, 2012).

In this subsection, we note two important issues:

1. Eq. 2.13 implies that the sky noise is convolved with the PSF just as the main signals of interest. This is important in the aperture synthesis imaging. We postpone a more detailed discussion about the consequences of this to Chapter 5.

2. As we explained the elements of $\mathbf{\Sigma}_{\mathrm{n}}$ are non-Gaussian, in general i.e. the noise added to the visibilities is non-Gaussian.

2.1.3 Mutual coupling and beam dissimilarities

The mutual coupling effect was represented in terms of mutual impedances in matrix $\mathbf{M}$ in Sect. 2.1.1. This effect can be viewed differently. In its presence, some part of the power received by the elements is absorbed while the rest of the power is re-radiated to the other elements. The re-radiated power induces new currents in the other elements. Consequently, the radiation pattern of the elements change. This leads to dissimilarity of the element beam patterns despite that they are manufactured identically. The mutual coupling effect depends on the array geometry and the array electrical characteristics. In Sect. 2.1, the mutual impedances in matrix $\mathbf{Z}$ and correspondingly the mutual coupling matrix, $\mathbf{M}$ account for this effect. However, relating $\mathbf{M}$ and dissimilarities of the antenna elements beam is not analytically straightforward, especially for stations with a sophisticated antenna design (e.g. dual polarized, wideband Bow-ties and Vivaldis) and hierarchical beamforming such as in the LOFAR HBA and EMBRACE shown in Fig. 2.5. In such arrays, the mutual coupling between each and individual diples has to be computed and we are interested in their net effect as the mutual coupling between the station elements i.e. between the tiles. Therefore, it has to be studied numerically.

We have done such numerical computations to check the dissimilarity of the element beams for a 24-tile HBA station (Fig.2.5, right panel) using Computationally Advanced and Efficient Simulator for ARrays (CAESAR) (Maaskant and Yang 2006; Maaskant et al. 2008). Numerical computation for the EM simulation was done using the Characteristic Basis Function Method (CBFM) which is the most numerically efficient and accurate method available for large scattering problems (Prakash and Mittra 2003; Yeo et al. 2003). Fig. 2.6 presents the simulation results when the beams are formed toward the local zenith. Each subplot presents the beam patterns of 24 tiles for a particular frequency in the $\phi = 0$ plane. One can see that due to a different mutual coupling environment, each tile has a slightly different beam pattern. The mutual coupling effect is sufficiently small not to disturb the main beam but only the far sidelobes which are at least 12dB lower than the main lobe. This is not a favorable condition for calibration, in particular when a strong source comes from the sidelobes and causes significant disturbances in the measured visibilities (see Fig.3.4, top-panel). This crucially harms the array calibratability, unless the beam model of

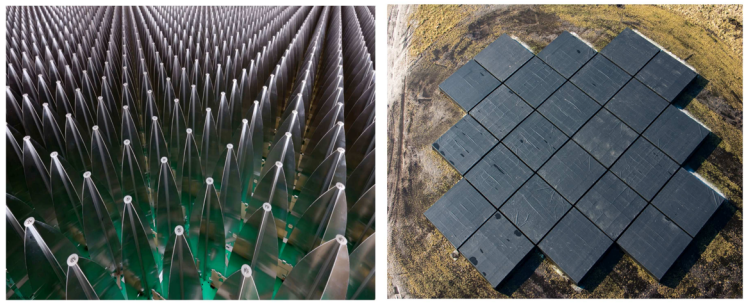


Figure 2.5: Left: An EMBRACE station. The output signals of 72 dual-polarized Vivaldi antennas are coherently summed to form a tile signal. Right: A 24-tile HBA station. The output signals of 16 dual-polarized Bow-ties antennas are coherently summed to form a tile signal. Station elements i.e. tiles are closely packed. The photos are courtesy of ASTRON.

each tile in the station is provided. This model has to be a model of the whole beam i.e. both the main beam and the sidelobes, in all directions.

In this view of mutual coupling, it introduces a direction-dependent gain per source and per array element, which may be modelled as $\mathbf{G}_0$ (size, $p \times q$) being element-wise multiplied to the array response matrix, $\mathbf{A}$. Thus, the data model in Eq. 2.11 becomes:

$$\mathbf{R} = \mathbf{G}(\mathbf{G}_0 \odot \mathbf{A})\mathbf{\Sigma}_{\mathrm{s}}(\mathbf{A}^H \odot \mathbf{G}_0^H)\mathbf{G}^H + \mathbf{\Sigma}_{\mathrm{n}} \tag{2.16}$$

Eq.2.16 and Eq.2.11 can both represent the data model (or the full ME) for a phased-array station. Comparing Eq.2.16 with Eq.2.11, one may notice that direct association of $\mathbf{G}_0$ and $\mathbf{M}$ is analytically impossible. Such comparison implies that the relation $\mathbf{G}_{0p\times q} \odot \mathbf{A} = \mathbf{M}_{p\times p}\mathbf{A}$ must hold, so that both equations model the same effects in the same data. Here, it is clear that $\mathbf{G}_0$ and $\mathbf{M}$ model different effects, that are known to us. $\mathbf{G}_0$ (size $p \times q$) represents the array element beam dissimilarities, which has appeared as a direction-dependant gain being element-wise multiplied with the array response matrix, $\mathbf{A}$ in Eq.2.16. $\mathbf{M}$ (size $p \times p$) represents the mutual impedances between the array elements, which has appeared as a matrix being multiplied with the array response matrix in Eq.2.11. The problem arises when we model any DDE, as a matrix of size $p \times p$ and estimate for it in the calibration process, whereas a matrix of size $p \times q$ actually describes its physical effect. Then, we certainly estimate for incorrect parameters. We study such an example in Chapter 4.

As it has been mentioned ealier, station visibilities are exploited for several engineering purposes. After such processing, the output signals of the station elements are weighted, summed (i.e. beamformed), and sent to the central processor. There, a phased-array station is seen as an element of the whole aperture array telescope, described in Sect. 1.3 and next section.

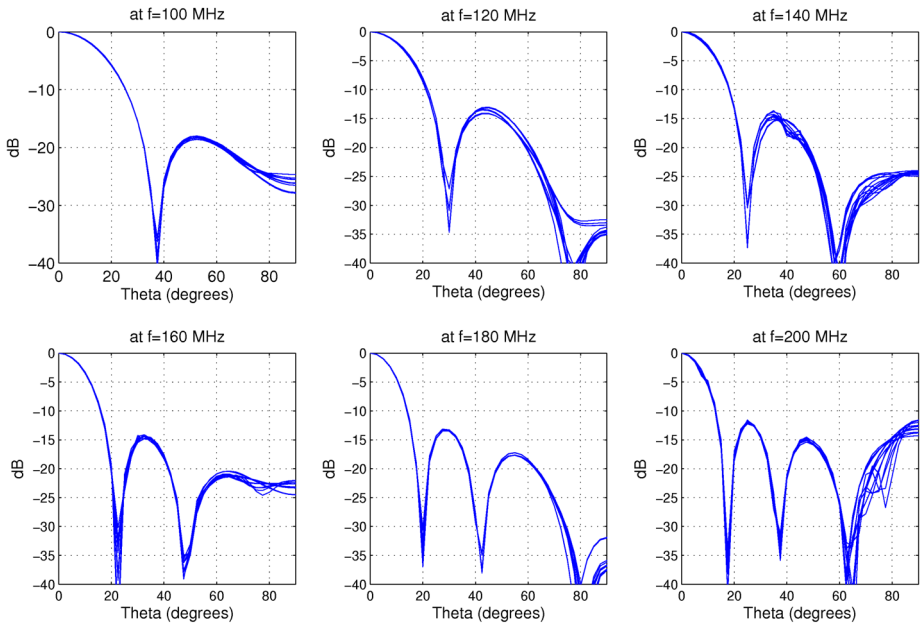


Figure 2.6: Simulated beam pattern of the tiles in a 24-tile HBA station (Fig. 2.5, right) using CAESAR. Each subplot presents the beam patterns of 24 tiles in a particular frequency in $\phi = 0$ plane. The beams are formed toward the local zenith. The dissimilarities in the sidelobes are caused by the different mutual coupling environments of each tile.

2.2 Data model for the aperture array telescope

Obviously, the data model at the synthesis array level follows the same principles as explained in Sect. 2.1 and led to the data model in Eq.2.16. But at this level, certain effects lose their importance e.g. mutual coupling between the array elements and the correlated receiver noise. At this level, the presence of mutual coupling between the phased-array stations is least likely. But, the beam pattern of stations might be different owing to their different size and orientation, which is modelled as a DDE gain ($\mathbf{G}_0$ matrix in Eq.2.16). Some other effects gain more importance owing to the array large baselines, improved sensitivity, and long integration time of observation. Using the array long baselines entails that some point-like calibrator beacons e.g. CasA are resolved and observed as extended sources. In this case, the source flux, σ_s requires a more sophisticated model than a point-like source model (see Chapter 5). Unlike the measured visibilities at a station, the measured visibilities here are used for the ultimate scientific imaging, taking advantage of the array high sensitivity and resolution. Moreover, polarization studies of the images are an important scien-

tific goal. The long integration time makes the presence of some effects important e.g. sky rotation in the FoV, the primary beam variation over time, the propagation path effects etc., which have to be taken into account in the data model. Each of these can be expressed in a corresponding matrix properly accomodated in the full ME such as Eq.2.16 (Boonstra and van der Veen 2003; van der Veen et al. 2004). However, exploiting the full ME for e.g. calibration, when more parameters have to be incorporated is not straightforward and is somewhat impractical. Hamaker et al. (1996); Sault et al. (1996); Hamaker and Bregman (1996); Hamaker (2000) introduced a 2×2 ME for modelling and analysing the radio interferometric data. The 2×2 ME models the visibility measured on a baseline, taking into account the full polarization information of the incoming signals. It uses 2×2 Jones matrices to model different effects, that the signals experience from its origin to the interferometer correlator. The reader is referred to Smirnov (2011a) for more details. Although the original motivation of this way of data modelling was to adopt the optical polarimetric formalism and methodlogies to radio interferomentric data, a 2×2 ME is now commonly used due the following reasons:

- The exploitation of the full ME becomes impractical, as the structure in the matrices are not necessarily exploitable in a special way. For example, an appropiate model of the DDEs is $\mathbf{G}_0$ (size $p \times q$) being element-wise multiplied with the array response matrix, $\mathbf{A}$ in Eq.2.16. The estimation of its parameters is not straightforward using a full ME, as it is extensively discussed by van der Tol et al. (2007).

- The large number of array elements requires larger inversion problems to be solved in the calibration process. They become computationally expensive, given that we must solve for more than two parameters per array element. However, such inversions estimate a certain parameter for all the elements at once.

A 2×2 ME models the visibility measured on a baseline between element i and element j, $\mathbf{R}_{ij}$. It is in principle the off-diagonal term of the full ME in Eq.2.16, when the polarization components of the incoming wave are fully taken into account. $\mathbf{B}_s$ represents the source coherency in terms of its polarization components, for an arbitrary source (a plane wave) in

$$\mathbf{B}_s = \mathbf{E}_s(\mathbf{k},t)^* \mathbf{E}_s(\mathbf{k},t) = [e_\theta, e_\phi]^H [e_\theta, e_\phi] = \begin{bmatrix} \langle |e_\theta|^2 \rangle & \left\langle e_\phi^* e_\theta \right\rangle \\ \langle e_\phi e_\theta^* \rangle & \langle |e_\phi|^2 \rangle \end{bmatrix}. \tag{2.17}$$

Then, the 2×2 ME, when q sources are present in the FoV can be expressed as

$$\mathbf{R}_{ij} = \sum_{s=1}^{q} \mathbf{J}_i(\boldsymbol{\ell}) \mathbf{B}_s(\boldsymbol{\ell}) \mathbf{J}_j{}^H(\boldsymbol{\ell}), \tag{2.18}$$

where $\mathbf{J}$ is the product of the Jones matrices of all the physical effects that a signal experiences,

$$\mathbf{J}_i = \mathbf{J}_{in}\mathbf{J}_{in-1}...\mathbf{J}_{i1}. \tag{2.19}$$

Some $\mathbf{J}$'s do not vary with $(l, m, n = \sqrt{1-l^2-m^2})$. They are so-called uv-plane effects e.g. receiver gain. Some $\mathbf{J}$'s do vary with (l, m, n). They are so-called *image*-plane effects e.g. ionosphere. The order of the terms in Eq. 2.19 are important, because matrices do not commute in general. $\mathbf{J}_{in}$ is the reciever gain and $\mathbf{J}_{i1}$ is the first effect e.g. the ionospheric Faraday rotation, that the source experiences. Eq. 2.19 can be rewritten

$$\mathbf{J}_i = \underbrace{\mathbf{J}_{in}...\mathbf{J}_{ik+1}}_{\text{uv-plane effects: } \mathbf{G}_i} \quad \mathbf{K}_i \quad \underbrace{\mathbf{J}_{ik-1}...\mathbf{J}_{i1}}_{\text{uv- and lm-plane effects: } \mathbf{E}_i}, \tag{2.20}$$

where $\mathbf{K}_i$ carries the phase shift due the different time of arraival of the source signal at the baseline element i. In Eq. 2.18, the uv-plane terms can be taken out of the summation operation changing it to

$$\mathbf{R}_{ij} = \mathbf{G}_i \left(\sum_{s=1}^{q} \mathbf{K}_i \mathbf{E}_i \frac{\mathbf{B}_s}{n} {\mathbf{E}_j}^H {\mathbf{K}_j}^H \right) {\mathbf{G}_j}^H. \tag{2.21}$$

If we expand the $\mathbf{K}$ terms, the Fourier kernel becomes apparent

$$\mathbf{R}_{ij} = \mathbf{G}_i \left(\sum_{s=1}^{q} \underbrace{\mathbf{E}_i \frac{\mathbf{B}_s}{n} {\mathbf{E}_j}^H}_{\text{apparent sky}} \overbrace{e^{2\pi f j (u_{ij} l + v_{ij} m + w_{ij}(n-1))}}^{\text{F.T. kernel}} \right) {\mathbf{G}_j}^H. \tag{2.22}$$

Note that Eq. 2.22 is a 3D Fourier transform of the *apparent sky* in the Cartesian coordinates denoted by (l, m, n), where n is constrained to be $\sqrt{1-l^2-m^2}$. In other words, Eq. 2.22 is a 3D Fourier transform of the projected *apparent sky* on unit celestial sphere. If the FoV is narrow ($n \simeq 1$) and/or the coplanary assumption holds ($w \simeq 0$), Eq. 2.22 is replaced by a 2D Fourier transform in

$$\mathbf{R}_{ij} = \mathbf{G}_i \left(\sum_{s=1}^{q} \mathbf{E}_i \mathbf{B}_s {\mathbf{E}_j}^H \overbrace{e^{2\pi\nu j (u_{ij} l + v_{ij} m)}}^{\text{F.T. kernel}} \right) {\mathbf{G}_j}^H. \tag{2.23}$$

Since none of the aforementioned assumptions holds for new radio telescopes such as LOFAR, Carozzi and Woan (2009) suggested using a 3D ME expressed in Eq. 2.22 with a simple conversion, that converts the $\mathbf{B}_s$ terms from (θ, ϕ, r) to (l, m, n) coordinates. This way, all-sky polarimetric calibration and imaging in a single telescope pointing are possible without any step such as the sky faceting or w-projection. Moreover, significant ongoing research has been carried out on computational efficieny of higher order FFT e.g. 3D FFT (Frigo and Johnson 1998; Frigo 1999; Frigo and Johnson 2005), allowing to reduce its computation cost and speed. We continue this discussion in Chapter 5 and 6.

2.3 Summary

In this chapter, we presented the array data model precisely. The following notes capture a synopsis:

- Mutual coupling in a phased-array station cannot be dismissed. It has consequences for its calibratability, which we will discuss in Chapter 3. Although such effect has been taken into account in array signal processing literature e.g. Manikas and Fistas (1994), it has always been dismissed in the calibration of the phased-array stations in radio astronomical application. Fig.2.6, shows the result of an intensive electromagnetic simulation to make this point for the first time in aperture array community.

- A more rigorous study of the noise statistical properties of the visibility samples is very essential. Noise in the visibility data is non-Gaussian, in general, as also noted earlier by van der Tol and Wijnholds (2006). It may converge to a Gaussian, if a significant number of visibility samples are averaged to represent the data, according to the central limit theorem. Such study is important as the Gaussianity of the noise is a fundamental assumption in many signal processing algorithms, in particular in the calibration algorithms, that are based on the statistical properties of noise e.g. (Yatawatta et al. 2008; Kazemi et al. 2011, 2012). Studying whether the performance of these calibration methods improves with the assumption of noise non-Gaussianity is beyond the scope of this thesis, but remains an essential issue to be studied.

- An appropiate model of the DDE (modelled as $\mathbf{G}_0$) is presented in Eq. 2.16, when we use the full ME. In other words, a data model in which $\mathbf{G}_0$ (size $p \times q$) is element-wise multiplied with the array response matrix ($\mathbf{G}(\mathbf{G}_0 \odot \mathbf{A})\mathbf{\Sigma}_{\mathrm{s}}(\mathbf{G}_0 \odot \mathbf{A})^H\mathbf{G}^H + \mathbf{\Sigma}_{\mathrm{n}}$) correctly accommodates the DDE, rather than a data model in which $\mathbf{G}_0$ (size $p \times p$) is multiplied with the array response matrix ($\mathbf{G}\mathbf{G}_0\mathbf{A}\mathbf{\Sigma}_{\mathrm{s}}\mathbf{A}^H\mathbf{G}_0^H\mathbf{G}^H + \mathbf{\Sigma}_{\mathrm{n}}$). Choosing the latter model leads to estimating for the incorrect parameters, although it makes the estimation process straightforward.

  A data model based on 2×2 ME is more reliable for modelling the DDEs, but it has its own limitations e.g. a 2×2 ME takes the information of an isolated baseline into consideration whereas a full ME takes as much information as possible into consideration in an estimation process. As a result, using the constraints that involve all the array elements for the calibration routine based on 2×2 ME is not straighforward. We will discuss this further in Chapter 4, where we study a constrained self-cal routine using a full ME.

- As it was discussed in Sect. 2.1.2, the sky background noise in the image plane experiences the same physical and geometrical effects, as the signals of interest (signals from the astronomical sources used for self-calibration) received by the interferometer. As a result, the sky background noise (the prominent noise term

at low-frequencies) is convolved with the PSF as well. The consequences of this will be discussed in Chapter 5.

The receivers noise signals do not come from the image plane and are assumed to be i.i.d. Gaussian signals. These imply that they should not correlate between the array elements, thus do not appear in the off-diagonal terms of the visibility matrix (Eq. 2.11). In other words, at the synthesis array, sky noise is the noise in the visibility data, not the receiver thermal noise. However, this condition is met better with longer integration time, in practice.

Chapter 3

Station calibration

In this chapter, we assess the benefits and limitations of using the redundant visibility information in regular phased-array systems to improve the quality of the calibration.

Regular arrays offer the possibility of using redundant visibility information to constrain the calibration of the array independently of a sky model and a beam model of the station elements. This requires a regular arrangement of the configuration of array elements and identical beam patterns.

We revised a previously developed calibration method for phased-array stations using the redundant visibility information in the system and applied it successfully to a LOFAR station. The performance and limitations of the method were demonstrated by comparing its application to real and simulated data. The main limitation is the mutual coupling between the station elements, which leads to non-identical beams and stronger baseline-dependent noise. Comparing the variance in the estimated complex gains with the CRB indicates that redundancy is a stable and optimum method for calibrating the complex gains of the system.

Our study shows that the use of the redundant visibility does improve the quality of the calibration in phased array systems. In addition, it provides a powerful tool, from which system diagnostics can be obtained. Our results demonstrate that designing redundancy in both the station layout and the array configuration of future aperture arrays is recommended. This is particularly true in the case of the SKA aiming at a dynamic range requirement that surpasses any existing array by an order of magnitude (Dewdney et al. 2010).

Redundancy Calibration of Phased Array Stations
P. Noorishad, S. J. Wijnholds, A. van Ardenne, and J. M. van der Hulst
Astronomy & Astrophysics, 545, A108, 2012.

3.1 Introduction

An important conceptual difference between the next generation of radio telescopes and conventional ones is their hierarchical system architecture. An excellent example is the LOFAR (de Vos et al. 2009). In LOFAR, a station consists of phased-arrays that are either sparse or dense. In the sparse phased-array stations such as the LBAs in a LOFAR station, the station elements are dipoles that are digitally beam-formed to synthesize a receiving aperture. Phased-arrays operating above $\sim$ 100 MHz are often implemented as compound elements or tiles, such as the HBAs at the LOFAR stations or in EMBRACE, (van Ardenne et al. 2004; Kant et al. 2011), the latter being is an example of a very dense phased-array station. In these stations, the station elements are phased-array tiles (e.g. Fig. 3.1). A tile is a regular arrangement of many dipoles whose signals are added in phase to form an instantaneous beam (the first level of beam-forming). The tile output signals are digitally phased up to synthesize a receiving aperture. At the next level in the beam-forming hierarchy, the beam-formed output of each station is transported to the central correlator to synthesize the whole telescope. Calibration has to be performed at different levels of this hierarchy to provide a final, high dynamic range image of part of sky, as explained by Wijnholds et al. (2010).

Figure 3.1: Twenty-four-tile HBA station. Each tile is one station element. This is the station configuration for most of LOFAR's HBA stations, including the LOFAR core (CS302). It is clear that a station like this is highly redundant.

In this chapter, we concentrate on the calibration at station level to ensure that the station beam is stable over time and frequency. A robust calibration as part of the beam-forming process should guarantee a stable beam pattern of the station for data going to the central correlator. This is crucial for the dynamic range of the final images made using data from the entire array.

In a phased-array station, the output of all station elements can be correlated. These correlations are called station visibilities, which are used for engineering purposes e.g. station calibration and RFI detection/mitigation. These correlations include many short baselines on which extended structures such as the galactic plane are captured. The most commonly used calibration methods for phased-array stations

are model-based, such as a multi-source calibration method (Wijnholds and van der Veen 2009a). A model-based calibration method requires the presence of one or more relatively unresolved point sources such as CasA and a model of the extended structures (see Fig.3.2). Modelling an extended structure is computationally difficult and expensive. Wijnholds and van der Veen (2009b) phenomenologically model it as correlated noise and estimate the parameters of interest for calibration using a weighted alternative least squares (WALS) approach. However, the model-based methods are in general iterative methods, which usually converge after several iterations.

A regular arrangement of station elements has the advantage that it provides redundant baselines, i.e. baselines of the same physical length and orientation. The use of redundant baseline information for calibration was introduced by Ishiguro (1974). It was exploited further for image sharpness in optical imaging by Hamaker et al. (1977). Noordam and de Bruyn (1982) demonstrated high dynamic range radio imaging using the WSRT redundant baselines information. Wieringa (1991) mathematically formulated a linear calibration algorithm based on using the redundant baseline information. Its linearity, independence of a sky model, low computational cost, and proven efficiency for precision calibration of WSRT observations motivated us to investigate the use of redundancy calibration to phased-array stations. The redundancy calibration algorithm uses the data of all redundant baselines to obtain a convergent calibration solution in a single step.

However, redundancy calibration in phased-array systems requires additional considerations, when compared to the regular WSRT. This is essentially because of their different design concept e.g. the closely located elements of a phased array station experience mutual coupling between elements that leads to non-identical beams of the station elements and to correlated receiver noise.

In this chapter, we refine the standard data model presented in the phased-array signal processing literature to account for baseline-dependent corruptions in terms of the coupling effects. Using this refined data model, we briefly introduce the two calibration methods i.e. model-based and redundancy. This helps us to achieve a clearer understanding of the potential and limitations of both calibration methods. We also revise the redundancy method formalism presented by Wieringa (1991); Liu et al. (2010) to capture the nature of baseline dependent errors that affect the calibration accuracy. Some implementation issues are raised and investigated using observed and simulated data of LOFAR HBA stations. We evaluate the redundancy calibration performance by comparing the variance in its results with the CRB and the plots of residuals for the corrected data after redundancy calibration. We also discuss limiting factors for its applicability.

Although we used HBA data to demonstrate the applicability and efficiency of the redundancy method, the analysis in this chapter is relevant to any phased array that is to be calibrated using redundant visibility information.

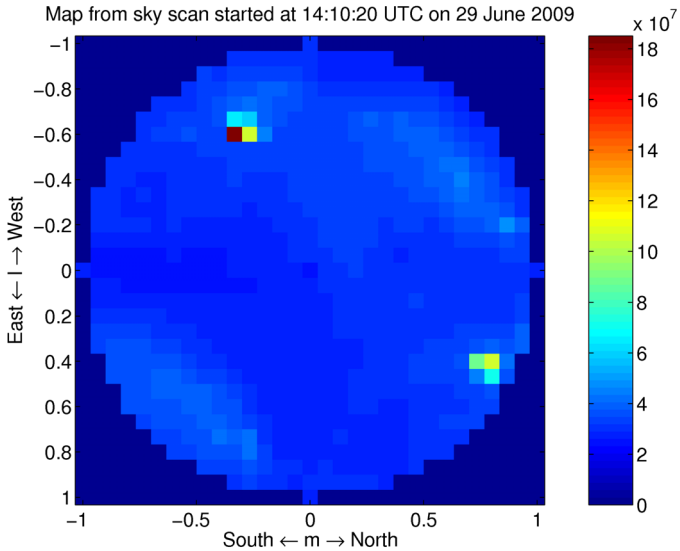


Figure 3.2: The sky imaged by HBA tiles at 14:10:20 UTC on 29 June 2009. The Galactic plane appears in the north-west, the sun appears in the south-west. One can also see their corresponding grating response in the image. The image is presented in the (l, m)-coordinates $l = cos(el)sin(az)$ and $m = cos(el)cos(az)$, where el and az denote elevation and azimuth, respectively.

3.2 Methods

3.2.1 Data model for phased-arrays

The standard data model for phased array stations presented in the literature assumes that in the absence of RFI and any coupling effects, a phased-array of p elements has a signal vector, $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_p(t)]^T$ that can be expressed as

$$\mathbf{x}(t) = \mathbf{\Gamma\Phi}\left(\sum_{k=1}^{q} \mathbf{a}_k s_k(t)\right) + \mathbf{n}(t) = \mathbf{\Gamma\Phi A s}(t) + \mathbf{n}(t), \tag{3.1}$$

where $\mathbf{s}(t)$ is a $q \times 1$ vector containing q mutually independent i.i.d. (temporally independent and identically distributed) Gaussian signals. The covariance matrix of the sources signals is $\mathbf{\Sigma}_s = \mathrm{diag}(\boldsymbol{\sigma}_s)$ (size $q \times q$), where $\boldsymbol{\sigma}_s$ is a vector of the source fluxes. They are also assumed to be narrow band, so we can define the q spatial signature vectors $\mathbf{a}_k$, which include both the phase delays due to the geometry and the directional response of the receiving elements (assumed to be identical). The vectors $\mathbf{a}_k$ are called the array response vectors which, are usually normalized. The receiver noise signals $n_i(t)$ are assumed to be mutually independent i.i.d. Gaussian signals in a $p \times 1$ vector $\mathbf{n}(t)$ and uncorrelated. Thus, the noise covariance matrix

becomes $\mathbf{\Sigma}_\mathrm{n} = \mathrm{diag}(\boldsymbol{\sigma}_n)$ (size $p \times p$). The amplitudes and the phases of direction-independent complex gains (g_i's), that have to be calibrated are $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, ..., \gamma_p]^T$ and $\boldsymbol{\phi} = [e^{j\phi_1}, e^{j\phi_2}, ..., e^{j\phi_p}]^T$. Correspondingly, $\mathbf{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$ and $\mathbf{\Phi} = \mathrm{diag}(\boldsymbol{\phi})$. In addition, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_q]$ (size $p \times q$) is a stack of the array response vectors. Before computing the coherency, the output voltage of each element, $\mathbf{x}(t)$ is sampled at time interval T. The nth sample of the array signal vector $\mathbf{x}[n]$ is given by

$$\mathbf{x}[n] = \sum_{-\infty}^{\infty} \mathbf{x}(t)\delta(t - nT) = \mathbf{x}(nT), \tag{3.2}$$

N samples can be stacked in a matrix $\mathbf{X} = [\mathbf{x}[1], \mathbf{x}[2], ..., \mathbf{x}[N]]$ (size $p \times N$), which denotes the short-term integration data set or snapshot. The array covariance matrix or the visibility matrix describing the correlation between all sampled voltages can be estimated by $\hat{\mathbf{R}} = \mathbf{X}\mathbf{X}^T/N$, whose expected value becomes either

$$\mathbf{R} = \mathbf{\Gamma}\mathbf{\Phi}\mathbf{A}\mathbf{\Sigma}_\mathrm{s}\mathbf{A}^H\mathbf{\Phi}^H\mathbf{\Gamma}^H + \mathbf{\Sigma}_\mathrm{n} \tag{3.3}$$

or

$$\mathbf{R} = \mathbf{G}\mathbf{A}\mathbf{\Sigma}_\mathrm{s}\mathbf{A}^H\mathbf{G}^H + \mathbf{\Sigma}_\mathrm{n}. \tag{3.4}$$

Correlator errors can be represented as an additive term to the covariance matrix, i.e. as a non-diagonal matrix. However, in Eq. 3.3 and throughout this paper, they are disregarded assuming that the correlator is designed perfectly.

We note that γ_i could accommodate the overall amplitude gain of both the receiver system and the atmospheric disturbances, and ϕ_i the corresponding phase shift. In the case of station calibration, we do not calibrate for atmospheric disturbances assuming that the snapshot data of a station is not affected. However, there is no guarantee that those effects remain unchanged during capturing the snapshot data in all frequency subbands, which takes about 9 minutes for a LOFAR station.

In phased-arrays such as the HBA stations of LOFAR, or EMBRACE, the antenna elements are closely packed. This causes mutual coupling between them, i.e. not all the power received by the elements is absorbed but some of the power is reradiated to the other elements. The reradiated power induces additional currents in the other elements. Consequently, the radiation pattern of the elements changes. This leads to a lack of similarity between the element beam patterns, and consequently different array response vectors despite their physical identicalness (as considered in $\mathbf{A}$). Beam dissimilarities can be modeled as a direction-dependent gain, $\mathbf{G}_0$ (size $p \times q$) being element-wise multiplied by the array response vectors

$$\mathbf{R} = \mathbf{G}(\mathbf{G}_0 \odot \mathbf{A})\mathbf{\Sigma}_\mathrm{s}(\mathbf{G}_0 \odot \mathbf{A})^H\mathbf{M}^H\mathbf{G}^H + \mathbf{\Sigma}_\mathrm{n}. \tag{3.5}$$

As studied by Svantesson (1998) and explained in Chapter 2, mutual coupling can be directly represented by a matrix of mutual impedances, $\mathbf{M}$ (size $p \times p$) in the data model

$$\mathbf{R} = \mathbf{G}\mathbf{M}\mathbf{A}\mathbf{\Sigma}_\mathrm{s}\mathbf{A}^H\mathbf{M}^H\mathbf{G}^H + \mathbf{\Sigma}_\mathrm{n}. \tag{3.6}$$

One may note that a direct association of $\mathbf{M}$ and $\mathbf{G}_0$ cannot be expressed analytically. It requires a numerical evaluation that we present in Sect.3.3.1. For simplicity and to continue our argument, we use the data model in Eq. 3.6.

Mutual coupling may not only act on the signal, but also on the system noise. The LNAs connected to the antennas in such an array, generate electromagnetic noise waves towards their outputs, but also send electromagnetic noise waves back into the antenna array. These waves are coupled into other receiver channels, giving rise to a correlated noise contribution. This effect is known as noise coupling, which contributes as $\mathbf{R}_{\rm rec}$ in a general and non-diagonal noise correlation matrix represented as:

$$\mathbf{\Sigma}'_{\rm n} = \mathbf{R}_{\rm sp} + \mathbf{R}_{\rm sky} + \mathbf{R}_{\rm rec} \tag{3.7}$$

where $\mathbf{R}_{\rm sp}$ is the spillover noise correlation matrix that is usually negligible as compared with $\mathbf{R}_{\rm sky}$ which is the sky noise contribution. Crosstalk in the back-end adds another baseline-specific correlated noise term, which we disregard here. The general data model for the visibility matrix becomes:

$$\mathbf{R} = \mathbf{GMA}\mathbf{\Sigma}_{\rm s}\mathbf{A}^H\mathbf{M}^H\mathbf{G}^H + \mathbf{\Sigma}'_{\rm n}. \tag{3.8}$$

We note that $\mathbf{\Sigma}'_{\rm n}$ is not a diagonal matrix, unlike $\mathbf{\Sigma}_{\rm n}$. It can be shown that noise from each station element still has a normal distribution in view of the large number of samples and according to the central limit theorem. Therefore, the off-diagonal elements of $\mathbf{\Sigma}'_{\rm n}$ that appear in Eq. 3.10, have a Wishart distribution.

The noise model in Eq. 3.7 has recently been presented in the literature for PAFs, for example by Jeffs et al. (2008); Ivashina et al. (2011). But, it can be used as a generic noise model for any antenna system (Maaskant 2010). It is clear, however, that depending on the antenna and LNA designs, the electrical characteristics of the array, and both the sparsity and density of the station element layout, the strength of mutual coupling varies. Thus, the term $\mathbf{R}_{\rm rec}$ may be replaced by uncorrelated receiver noise in this definition.

3.2.2 Model-based calibration method

The model-based calibration problem was formulated as a least squares minimization problem by Wijnholds and van der Veen (2009a). Here, we rewrite it using the refined data model given in Eq.3.8

$$\{\hat{\mathbf{g}}, \hat{\boldsymbol{\sigma}}_{\rm n}\} = \underset{\mathbf{g},\boldsymbol{\sigma}_n}{\rm argmin} \|\mathbf{GMA}\mathbf{\Sigma}_{\rm s}\mathbf{A}^H\mathbf{M}^H\mathbf{G}^H + \mathbf{\Sigma}'_{\rm n} - \hat{\mathbf{R}}\|_F^2. \tag{3.9}$$

This estimates the noise and complex gain of each receiver element using the measured visibility, $\hat{\mathbf{R}}$ and the modeled visibility $\mathbf{GMA}\mathbf{\Sigma}_{\rm s}\mathbf{A}^H\mathbf{M}^H\mathbf{G}^H + \mathbf{\Sigma}'_{\rm n}$, where $\mathbf{\Sigma}_{\rm s}$ and $\mathbf{A}$ are assumed to be known. We can calculate them, if we specify the time of observation, the telescope geometry and the known source parameters. In the presence of the coupling effects ($\mathbf{M}$ and $\mathbf{\Sigma}'_{\rm n}$), estimation results are biased, unless either an accurate beam model of each individual element is provided or the matrix $\mathbf{M}$ is known. In

this case, one can handle the effect of the correlated noise using a WALS approach (Wijnholds and van der Veen 2009b).

Fig. 3.2 shows a sky map scanned by HBA tiles. One can see the Galactic plane owing to the many short baselines and the Sun as the dominant radio source. Given the beam models of the station elements, the WALS method treats the extended structure as correlated noise and estimates the parameters of interest. However, the model-based methods are in general iterative methods that usually converge after several iterations.

3.2.3 Redundancy calibration method

The basic assumption of redundancy calibration is that the redundant baselines sample the same Fourier component of the sky ($\mathbf{A}\mathbf{\Sigma}_{\mathrm{s}}\mathbf{A}^H$). This assumption holds if the array response vectors of the redundant baselines are the same i.e. the element beams are identical. We therefore begin with the most general data model given in Eq. 3.8 to understand the limitations of this method for a phased-array station.

To build up the system of equations for the redundancy calibration algorithm, we represent an off-diagonal element of $\mathbf{R}$ in Eq. 3.8 as

$$R_{ij} = g_i g_j^* \left[\mathbf{M}\mathbf{A}\mathbf{\Sigma}_{\mathrm{s}}\mathbf{A}^H\mathbf{M}^H\right]_{ij} + \Sigma'_{\mathrm{n}_{ij}}. \tag{3.10}$$

We note that correlator offsets could contribute as additive corrupting factors in $\Sigma'_{\mathrm{n}_{ij}}$. Since $\mathbf{\Sigma}_{\mathrm{s}}$ in Sect. 3.2.1 is a diagonal matrix, one can expand Eq. 3.10 to

$$R_{ij} = g_i g_j^* M_{ii} M_{jj}^* \overbrace{\sum_{q=1}^{q} A_{iq}\sigma_q A_{jq}^*}^{\text{true visibility}} + \underbrace{g_i g_j^* \sum_{q=1}^{q} \left(\sum_{\substack{p_1=1 \\ p_1\neq i}}^{p} \sum_{\substack{p_2=1 \\ p_2\neq j}}^{p} M_{ip_1} M_{jp_2}^* A_{p_1 q} A_{p_2 q}^* \sigma_q \right)}_{\text{additive term due to mutual coupling}} + \Sigma'_{\mathrm{n}_{ij}}. \tag{3.11}$$

This shows that the mutual coupling produces baseline-dependent multiplicative and additive terms by influencing signals directly and noise indirectly ($\mathbf{R}_{\mathrm{rec}}$ in Eq. 3.7). This violates the fundamental assumption of redundancy i.e. we do not observe redundant visibilities on physically redundant baselines (see Fig. 3.4, top).

Without loss of generality, $M_{ii}M_{jj}^* = 1$ (or they can be absorbed in the gains). To establish an analogy between Eq. 3.11 and the redundancy method formalism in (Wieringa 1991), we define the term

$$e_{ij} = g_i g_j^* \sum_{q=1}^{q} \left(\sum_{\substack{p_1=1 \\ p_1\neq i}}^{p} \sum_{\substack{p_2=1 \\ p_2\neq j}}^{p} M_{ip_1} M_{jp_2}^* A_{p_1 q} A_{p_2 q}^* \sigma_q \right) + \Sigma'_{\mathrm{n},ij} \tag{3.12}$$

and rewrite the Eq. 3.11 as:

$$R_{ij}^{obs} = R_{ij}^{true} g_i g_j^* + e_{ij} \tag{3.13}$$

or

$$R_{ij}^{obs} = R_{ij}^{true} g_i g_j^* \underbrace{(1 + \frac{e_{ij}}{R_{ij}^{true} g_i g_j^*})}_{w_{ij}} \tag{3.14}$$

where R_{ij}^{obs} and R_{ij}^{true} are the observed and the theoretical redundant visibilities, respectively, g_i and g_j are the element complex gains and w_{ij} can be defined as a baseline-dependent error that affects the accuracy of the calibration results.

We take the natural logarithm of both sides of the equation Eq. 3.14 to obtain

$$\ln R_{ij}^{obs} = \ln|\gamma_i| + \ln|\gamma_j| + i(\phi_j - \phi_i) + \ln R_{ij}^{true} + \ln(w_{ij}). \tag{3.15}$$

In the absence of mutual coupling, the first term of e_{ij} drops and the second term reduces the contribution of $R_{\rm rec}$. The only correlated contributing terms come from $R_{\rm sky}$ and $R_{\rm sp}$, which are negligible depending on the signal-to-noise ratio (S/N) of the observation. We assume that we have such a case where we can ignore e_{ij}. We equate the amplitude and the phase values to decouple the system of equations for the phases and the amplitudes

$$\ln|R_{ij}^{obs}| = \gamma_i^{'} + \gamma_j^{'} + \ln|R_{ij}^{true}| \tag{3.16}$$

$$\psi_{ij}^{obs} = \phi_j - \phi_i + \psi_{ij}^{true}, \tag{3.17}$$

where $|R_{ij}|$ and ψ_{ij} are the amplitude and the phase of a complex visibility. Since we have to specify the absolute flux level, we set

$$\Sigma\gamma_i^{'} = 0. \tag{3.18}$$

We also have to constrain the element phase. We can enforce this constraint by specifying that the sum of all element phases is zero

$$\Sigma\phi_i = 0. \tag{3.19}$$

Furthermore, there may also be an arbitrary linear phase slope over the array. This phase slope corresponds to a position shift of the field. This arises because a redundancy solution does not provide an absolute position. This can be absorbed into either the true visibilities or the element phases. For a two-dimensional array, we constrain x and y in the same manner as

$$\Sigma_{i=1}^{p} \phi_i x_i = 0 \tag{3.20}$$

$$\Sigma_{i=1}^{p} \phi_i y_i = 0, \tag{3.21}$$

where x_i and y_i are the (x, y) coordinates of the array elements. Eq. 3.16- Eq. 3.21 formulate the redundancy calibration method as two overdetermined systems of linear equations for phases and amplitudes that can be solved in a single-step least squares solution. For instance, the phase estimator can be symbolized as $\mathbf{E}_{ph}\boldsymbol{\theta} = \boldsymbol{\Psi}^{obs}$ and solved using the pseudo-inverse

$$\hat{\boldsymbol{\theta}} = \left[\mathbf{E}_{ph}^{T}\mathbf{E}_{ph}\right]^{-1}\mathbf{E}_{ph}^{T}\boldsymbol{\Psi}^{obs}, \tag{3.22}$$

where $\boldsymbol{\theta} = [\phi_1, \phi_2, ..., \phi_p, \psi_1^{true}, \psi_2^{true}, ..., \psi_m^{true}]$ is the vector of parameters to be estimated, m is the number of distinct redundant baselines, $\mathbf{E}_{ph}$ is the coefficient matrix, and $\boldsymbol{\Psi}^{obs}$ is the vector of the observed redundant phases and right sides of Eq.3.19-Eq.3.21. Setting a phase reference for element phases, is done after the phase estimation.

We note that we came to the solution in Eq. 3.22, because we ignored the baseline-dependent noise. In its presence, our problem takes the form of $\mathbf{E}_{ph}\boldsymbol{\theta} + \boldsymbol{\beta} = \boldsymbol{\Psi}^{obs}$ (expanded in Eq. 3.24). The estimated parameters will deviate as

$$\boldsymbol{\varepsilon} \equiv \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} = \left[\mathbf{E}_{ph}^{T}\mathbf{E}_{ph}\right]^{-1}\mathbf{E}_{ph}^{T}\boldsymbol{\beta}, \tag{3.23}$$

where $\boldsymbol{\beta} = \mathrm{vec}(\angle(\ln(w_{ij})))$. If the array mutual coupling is significant, redundant baselines do not provide redundant visibilities Theu, redundancy calibration method cannot be applied. estimator. In the case of a weakly coupled array, the vector $\boldsymbol{\beta}$ still carries the correlated noise produced by R_{sp} and R_{sky}. Assuming that each has a Wishart distribution in Eq.3.12, their statistical distribution changes in vector $\boldsymbol{\beta}$. This was taken into account in deriving the results in Sect. 3.3.2.

$$
\begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\
. & . & . & . & . & . & . & . & . & & & & & . & . & . & & 0 \\
. & . & . & . & . & . & . & . & . & & & & & . & . & . & & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & -1 & 1 & 0 & 0 & \dots & 0 \\
. & . & . & . & . & . & . & . & . & & & & & . & . & . & & 0 \\
. & . & . & . & . & . & . & . & . & & & & & . & . & . & & 0 \\
. & . & . & . & . & . & . & . & . & & & & & . & . & . & & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & -1 & 0 & 0 & 0 & 0 & \dots & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & -1 & 0 & 0 & 0 & \dots & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 0 & 0 & 0 & \dots & 0 \\
x_1 & x_2 & x_3 & x_4 & . & . & . & . & . & \dots & . & x_{24} & 0 & 0 & 0 & \dots & 0 \\
y_1 & y_2 & y_3 & y_4 & . & . & . & . & . & \dots & . & y_{24} & 0 & 0 & 0 & \dots & 0
\end{bmatrix}
\Bigg\}\mathbf{E}_{ph}
\begin{bmatrix}
\phi_1 \\ \phi_2 \\ \phi_3 \\ . \\ . \\ . \\ \phi_{24} \\ \psi_1^{true} \\ \psi_2^{true} \\ \psi_3^{true} \\ . \\ . \\ . \\ \psi_{36}^{true}
\end{bmatrix}
\Bigg\}\boldsymbol{\theta}
+
\begin{bmatrix}
\beta_{1,2} \\ \beta_{3,4} \\ \beta_{4,5} \\ \beta_{5,6} \\ . \\ . \\ \beta_{23,24} \\ . \\ . \\ . \\ \beta_{1,23} \\ \beta_{2,24} \\ 0 \\ 0 \\ 0
\end{bmatrix}
\Bigg\}\boldsymbol{\beta}
=
\begin{bmatrix}
\psi_{1,2}^{obs} \\ \psi_{3,4}^{obs} \\ \psi_{4,5}^{obs} \\ \psi_{5,6}^{obs} \\ . \\ . \\ \psi_{23,24}^{obs} \\ . \\ . \\ . \\ \psi_{1,23}^{obs} \\ \psi_{2,24}^{obs} \\ 0 \\ 0 \\ 0
\end{bmatrix}
\Bigg\}\boldsymbol{\Psi}^{obs}
\tag{3.24}
$$

Since the created system of equations e.g. Eq. 3.24 is highly sparse, they are computationally fast. Most importantly, they are independent of a sky model but their accuracy is affected by the S/N of the observed sky as discussed by Liu et al. (2010). They require identical beams of station elements and a regular arrangement of antennas to provide a sufficient number of redundant baselines. One can then, estimate the element complex gains in a single step.

3.3 Implementation using real and simulated data

3.3.1 Verification of redundancy fundamental assumption

As mentioned earlier, redundancy translates into verifying the similarities among the element beams in a station. We checked this for a 24-tile HBA station similar to the one shown in Fig.3.1 using CAESAR (Maaskant and Yang 2006; Maaskant et al. 2008). The numerical computation of the EM simulation was done using the CBFM, which is the most numerically efficient and accurate method available for large scattering problems (Prakash and Mittra 2003; Yeo et al. 2003). Fig. 3.3 presents the simulation results when the beams are formed toward the local zenith. Each subplot presents the beam patterns of 24 tiles for a particular frequency in the $\phi = 0$ plane. One can see that owing to the different mutual coupling environment, each tile has a slightly different beam pattern. The mutual coupling effect is sufficiently small not to disturb the main beam but only the far sidelobes which are at least 12dB weaker than the main lobe. However, it is not a favorable condition in general for redundancy calibration. On the basis of this simulation, we expect to observe non-redundant visibilities on redundant baselines when a strong source falls in the sidelobes. This was confirmed by the real observation presented in Fig. 3.4, top panel. In this observation, owing to the absence of a strong source in the main beam, the non-identical sidelobes have the chance to play a significant role in disturbing the redundancy while in the observation shown in Fig. 3.4, bottom panel, one strong source in the main beam seems to be sufficient to dominate the influence of other possible strong sources observed in the direction of the non-identical sidelobes. There is of course a different contribution from the correlated background noise or sky noise ($\mathbf{R}_{\text{sky}}$) in Fig. 3.4, top and bottom panels.

Fig. 3.4 presents two data sets obtained to verify the use of redundancy in a HBA station. The observations were made on 5 September 2009. The tile beams in a HBA station (in the LOFAR core, known as CS302) were formed toward the local zenith. We let the sky drift over the FoV of the station. We captured the station visibilities approximately every 10 minutes (integration time of one second per frequency channel for 512 frequency channels). The bandwidth of the frequency channels is 195 kHz and the frequency range is $100 - 200$ MHz. On the right panel, the local sky viewed from CS302 is presented to show the sources that have contributed to the visibilities shown in the left panel.

Since the mutual coupling environment changes at lower elevations, dissimilarities between the main beams are consequently found up to $1 - 2$dB. Real observations

have shown less disturbance in the main beam at lower elevations (see Sect. 3.3.2 and Appendix B).

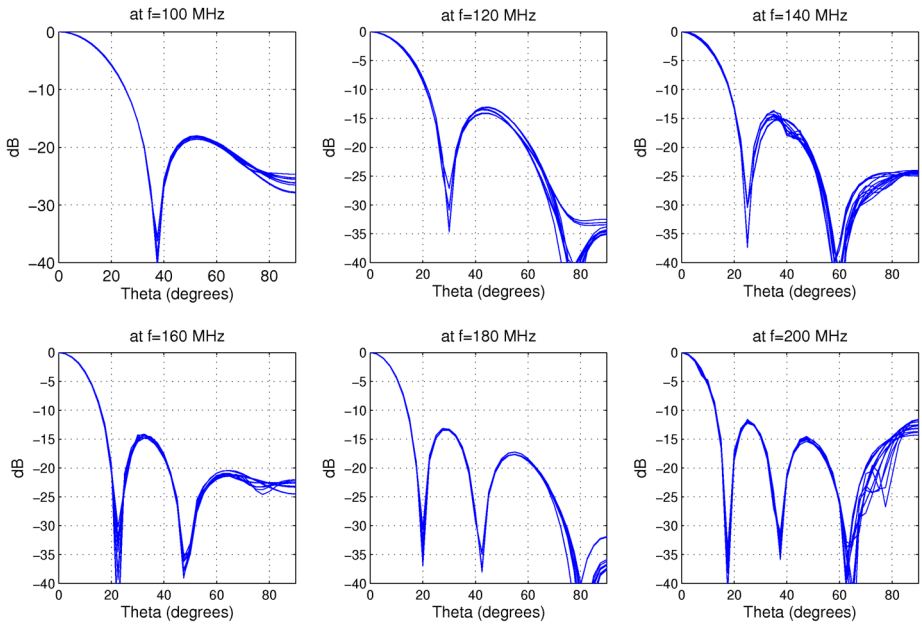


Figure 3.3: Simulated beam pattern of the tiles in a 24-tile HBA station using CAESAR. Each subplot presents the beam patterns of 24 tiles in a particular frequency in $\phi = 0$ plane. The beams are formed toward the local zenith. The dissimilarity between the sidelobes is caused by the different mutual coupling environments of each tile.

On the basis of the EM simulation by CAESAR (Fig. 3.3) and real observations (Fig. 3.4), we conclude that the best case scenario for redundancy calibration of an HBA station is to have a strong source in the main lobe when the beams are identical. However, the dissimilarities of the sidelobes introduce non-redundancy or systematic errors that cannot be eliminated by any statistical method or a longer integration time. Therefore, we must investigate whether the contribution of other strong sources, which may fall in the non-identical sidelobes, is significant. To quantify this systematic error in the visibility measured on a given baseline, we present the following example. At 21:29:04 UTC on 24 November 2009, at an HBA station called RS208 located at ($lon = 6.9196°$, $lat = 52.6699°$), four strong sources were in the FoV, as shown in Fig. 3.5. In Fig. 3.5, the tile beams were formed toward 3C461 (CasA), which was thus at the phase center. The sources 3C405 (CygA), 3C400, and 3C144 (TauA) have fallen in the sidelobes. The tile beams and their standard deviation at 120MHz, at these sources are also depicted in Fig. 3.5. The total complex visibility

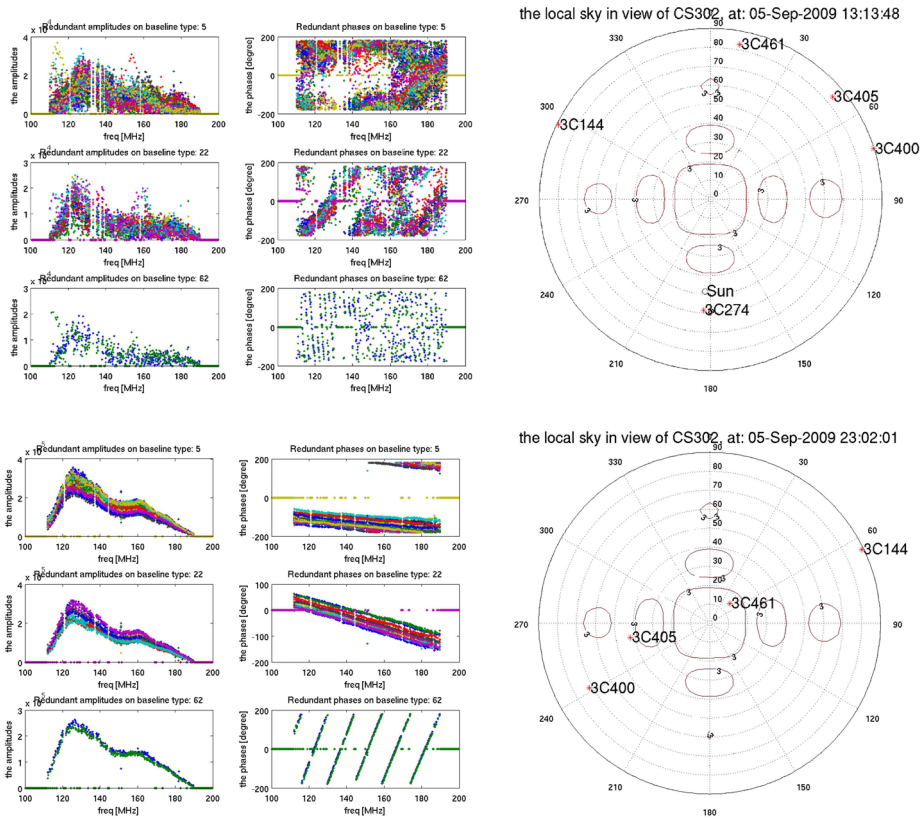


Figure 3.4: An observation performed with a HBA station on September 5, 2009 at 13:13:48 UTC in which redundancy calibration fails (top), as well as an observation with the same station at 23:02:01 UTC in which the redundancy calibration is successful (bottom). The left panels show the measured visibilities for three distinct types of redundant baselines. The right panels show the local sky model at the time of observation, where the strongest sources on the sky are superimposed on a contour plot of the element beam pattern at 170 MHz.

observed on a certain baseline, $\mathbf{D}_\lambda$ is computed as

$$\mathbf{R}^{obs} = |R^{obs}|e^{j\psi^{obs}} = \int_s \Lambda(\boldsymbol{\ell})I(\boldsymbol{\ell})e^{-j2\pi\mathbf{D}_\lambda.\boldsymbol{\ell}}d\Omega \tag{3.25}$$

where $\Lambda(\boldsymbol{\ell}) \equiv A(\boldsymbol{\ell})/A_0$ is the normalized tile reception pattern at σ, A_0 is the response at the beam center and $I(\boldsymbol{\ell})$ is the source flux. Since there are four dominant strong sources in the FoV, we assume that the integral can be replaced by a summation. The complex plane in Fig. 3.6 right panel, shows the summations of the visibility vectors as they were observed through hypothetically identical sidelobes, $\mathbf{R}_{total}$ (solid line), as well as their summation when they were attenuated differently by the actual

non-identical sidelobes, $\mathbf{R}^{'}_{total}$ (dotted line). We computed the systematic errors in the phase and the amplitude of the visibility by assuming the complex vector to be $\mathbf{R}_{total} = \mathbf{R}_{3C461} + \mathbf{R}_{3C144} + \mathbf{R}_{3C405} + \mathbf{R}_{3C400}$ using Eq. 3.25. For this, we choose the baseline type given in Fig. 3.8, rightmost panel ($\mathbf{D}$) and frequency $f = 120$MHz for $\mathbf{D}_\lambda = \mathbf{D}/\lambda$, where λ is the wavelength. We set the average beam pattern of all 48 tiles as the identical beam for the term $A(\boldsymbol{\ell})$ whose values are shown in Fig. 3.5 at different source locations, as the quantity beam level. We use $\Lambda(\boldsymbol{\ell}) = A(\boldsymbol{\ell})/A(\boldsymbol{\ell}_{3C461})$. The source fluxes, $I(\boldsymbol{\ell})$ are given in Table 3.1. We also compute $\mathbf{R}^{'}_{total} = \mathbf{R}^{'}_{3C461} + \mathbf{R}^{'}_{3C144} + \mathbf{R}^{'}_{3C405} + \mathbf{R}^{'}_{3C400}$. The computation settings are as before except that $A(\boldsymbol{\ell})$ is not identical this time, but deviated from the average beam pattern by the standard deviation values shown in Fig. 3.5 at different source locations, as the quantity $std(Beam)$. The bias introduced to the amplitude and the phase of $\mathbf{R}^{'}_{total}$ as compared with $\mathbf{R}_{total}$ are presented in the first row of Table. 3.1.

We steer the beam toward each available strong source and repeat the calculation to predict the systematic error due to the other sources in the sidelobes with different standard deviations. The results are presented in Table 3.1. We can conclude that tracking CasA provides the best redundancy in the observed visibilities.

Table 3.1: A-team radio sources in the FoV of RS208, on 24 of November 2009 at 21:29:04 UTC and the predicted levels of systematic errors in the amplitudes and the phases of the redundant visibility due to the contribution of other strong sources through non-identical sidelobes. The baselines given in Fig. 3.8, rightmost panel and frequency, $f = 120$MHz were chosen for the computations.

Calibrator source	Source flux $[Jy]$	Normalized error in Γ	Error in $\Phi[rad]$
3C461 (CasA)	8609	0.07	0.06
3C405 (CygA)	8100	0.20	0.09
3C400	540	0.93	0.01
3C144 (TauA)	1420	0.84	0.28

Note: Γ and Φ denote amplitudes and phases, respectively.

The plot of the residuals for the corrected redundant visibilities shown in Fig. 3.8 confirms the results in Table 3.1. Fig. 3.8 reveals the non-redundancy in the measured visibilities (in terms of amplitudes and phases). In Table 3.1, we predict the same quantities, which are caused by non-identical sidelobes. There is a few percent of discrepancy between the predicted residuals and their actual values in our observation. This can be explained by our simplifying assumption about the number of strong sources in the FoV. However, Table 3.1 can play an instrumental role for station calibration.

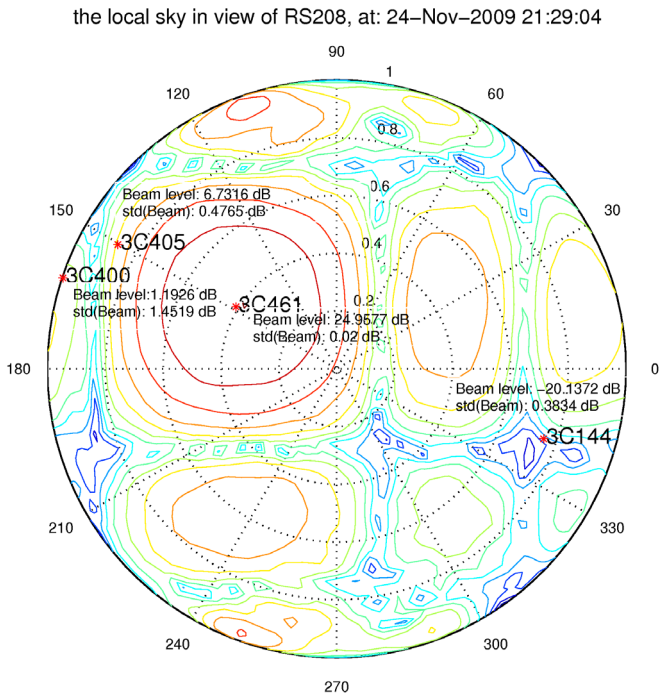


Figure 3.5: A-team radio sources in the FoV of an HBA station called RS208 on 24 November 2009 at 21:29:04 UTC. Tile beams are streered toward 3C461 to provide the most redundant visibilities. The values of the average reception pattern of 48 tiles, $A(\ell)$, and their standard deviation, $std(A(\ell))$ at 120 MHz have also been indicated at other source locations.

3.3.2 Redundancy calibration performance

To study the redundancy calibration performance, we tracked CasA on 24 November 2009 from 15:25:43 UTC until 22:12:19 UTC. Running a redundancy calibration on 48 captured data sets during this observation, provided very stable results for the receiver complex gains over time. This is an indication of the system stability and a working calibration routine, which operated approximately every 10 minutes. We compare the variance in the estimated complex gains over 48 runs of redundancy calibration with the CRB in Fig. 3.7. These quantities for the amplitudes and the phases are presented in two separate plots, as their estimators were decoupled in Sect. 3.2.3. The CRB (or the theoretical minimum variance) and the actual variance in the estimated parameters over time are in good agreement. The small difference between them can be explained by their not having exactly the same S/N from one observation to another during our survey from 15:25:43 UTC until 22:12:19 UTC, while the predicted S/N for this observation is S/N $\simeq 0.75$, following the analysis in (Wijnholds and van Cappellen 2011). This predicted value was used to compute the

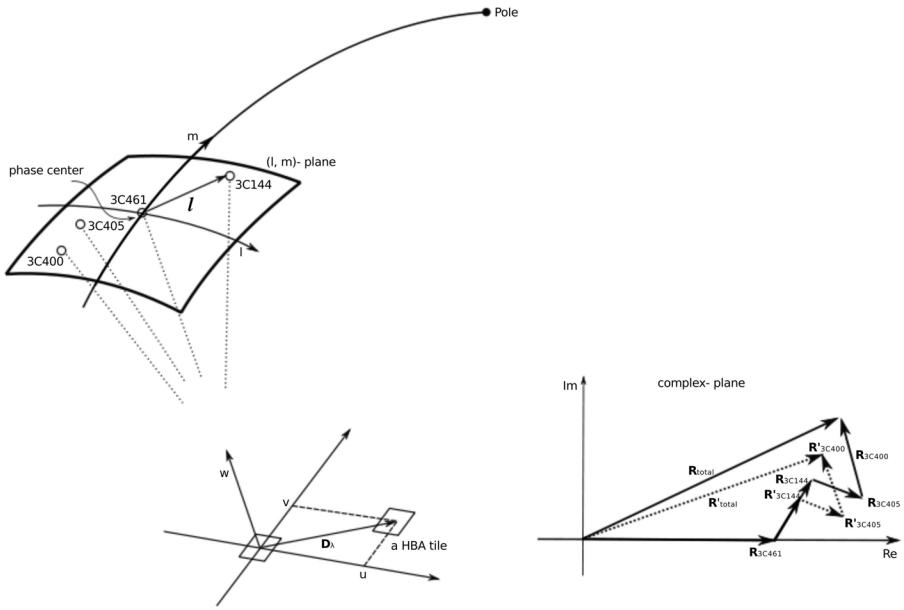


Figure 3.6: Total complex visibility on a certain baseline due to the contribution of all strong sources shown in Fig.3.5. The ones that contribute through non-identical sidelobes introduce non-redundancy or systematic errors into the total redundant visibility. This has been demonstrated in a complex plane on the right. The solid line shows the visibility vectors as they were observed through hypothetically identical sidelobes. The dotted line shows the visibility vectors when they are attenuated differently by the actual non-identical sidelobes.

presented CRB in Fig. 3.7.

We also studied the residuals for the corrected redundant visibilities. Fig. 3.8 shows an example of the residuals in both the amplitudes and phases of the corrected visibilities on a distinct type of redundant baseline. The snapshot is captured at 21:29:04 UTC, when CasA is at high elevation. The first row compares the results after redundancy calibration to those after the model-based calibration in the second row. The type of redundant baseline is depicted on the station configuration in the rightmost panel. The integration time is one second per frequency channel. More plots of the residuals from the same observation are presented in Appendix B, Fig. B.1 and Fig. B.2. By comparing them, one may note that the baseline length does not make a significant difference in the residuals. The residuals are on the order of $2-5\%$ in both phases and amplitudes. Because a strong source such as CasA in the FoV dominates the effect of any possible correlated noise, this can be explained by not having 100% identical main beams or the corruption due to the nonidentical sidelobes (see Table. 3.1). However, these results are satisfactory. Moreover, Fig. 3.7 showed that the estimated complex gains, which are our parameters of interest in a station calibration, are very stable. This is due to the constraints on them in the

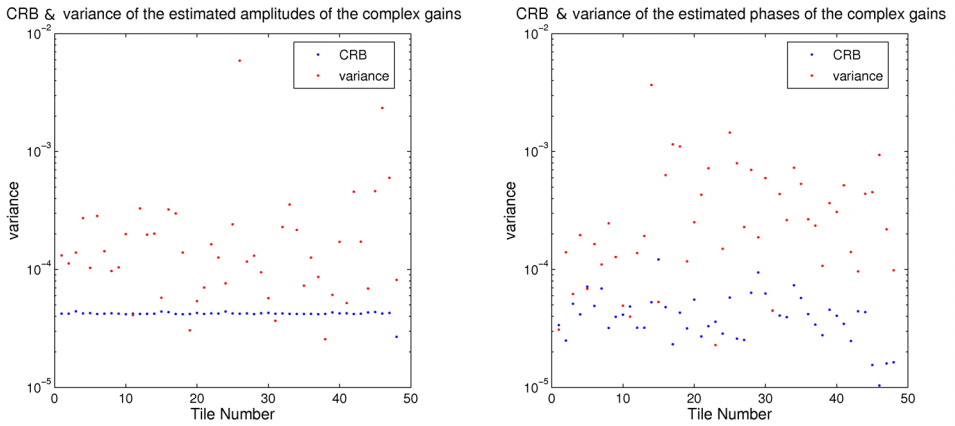


Figure 3.7: Theoretical minimum variances in the estimated amplitudes and phases of the complex gains or CRB compared with their actual variance. These quantities for the amplitudes and the phases are presented in two separate plots as their estimators were decoupled in Sect. 3.2.3.

formalism of the redundancy calibration.

Using the data captured at 15:25:43 UTC, when CasA is at low elevation, reveals slightly larger residuals in the order of $5-10\%$ (see Appendix B, Fig. B.3- Fig. B.5). This is due to a different mutual coupling environment which leads to less identical beams in either the main lobe or the sidelobes.

The residuals also show that the two calibration methods perform almost equally well, although the redundancy method behaves more consistently. The model-based calibration has slightly larger residuals on short baselines, e.g. up to $\sim 0-1\%$ in phases and $\sim 0-2\%$ in amplitudes, whereas it shows similar residuals as the redundancy calibration on long baselines. Extended structures such as the Galactic plane or north polar spur, are captured on short baselines. Modelling them is computationally expensive and impractical. Therefore, in the model-based method, one has to discard the visibilities measured on short baselines to simplify the measured sky for a corresponding simple sky model, while the redundancy method is sky-model-independent. Experiment has shown that discarding the equations of short baselines in the redundancy calibration routine, will not improve its results significantly.

Moreover, the model-based methods are sensitive to RFI sources, as their presence confuses the sky model. Since the redundancy calibration is independent of the sky model, it is less sensitive to RFI in the far-field. Comparing the plots in the first row of Fig. 3.8 with the plots in its second row, one may note that more frequency channels had to be flagged in the model-based calibration.

As discussed by Liu et al. (2010), redundancy calibration quality depends on the S/N of the observed sky, although it is independent of a sky model. Liu et al. (2010) showed that the estimated parameters are affected differently in the presence of baseline-dependent noise, assuming that it is Gaussian noise. The vector $\boldsymbol{\beta}$ in

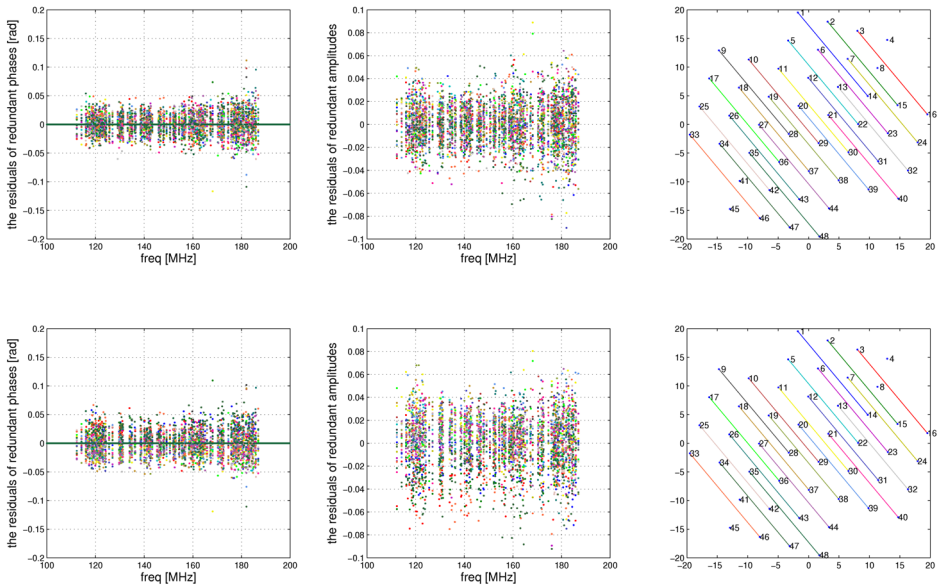


Figure 3.8: Plots of the residuals for corrected redundant visibilities in terms of phases and amplitudes on a given set of redundant baselines. The first row shows the results after redundancy calibration. The second row shows the result after model-based calibration. The data is taken from the observation done on 24 November 2009 at 21:29:04 UTC, when CasA is at high elevation. The station configuration of RS208 is shown in the rightmost panel. The corresponding redundant baselines to each redundant visibility are also depicted on the station layout, with the same color code.

Eq. 3.23 reveals a non-Gaussian baseline dependent noise. We have taken this into consideration in the following results. After adding different levels of a non- i.i.d. Gaussian noise to the output vector of the array (Eq. 3.1), we simulated the station visibility assuming that the station elements have identical beams. We used the station configuration of RS208 (see Fig. 3.8, rightmost panel). The complex gains were estimated over 100 runs of Monte-Carlo simulation for different S/N's per visibility. The variance of the error in the estimated complex gains (over simulation runs with different random noises) versus different S/N's per visibility is shown in Fig. 3.9. These plots indicate that the variances in the estimated complex gains, which are our parameters of interest in a station calibration, are the same for the different tiles. This illustrates the stability of the algorithm. However, an analysis of this nature is very array-configuration-dependent, as the array configuration determines the coefficient matrices by which our estimators are defined.

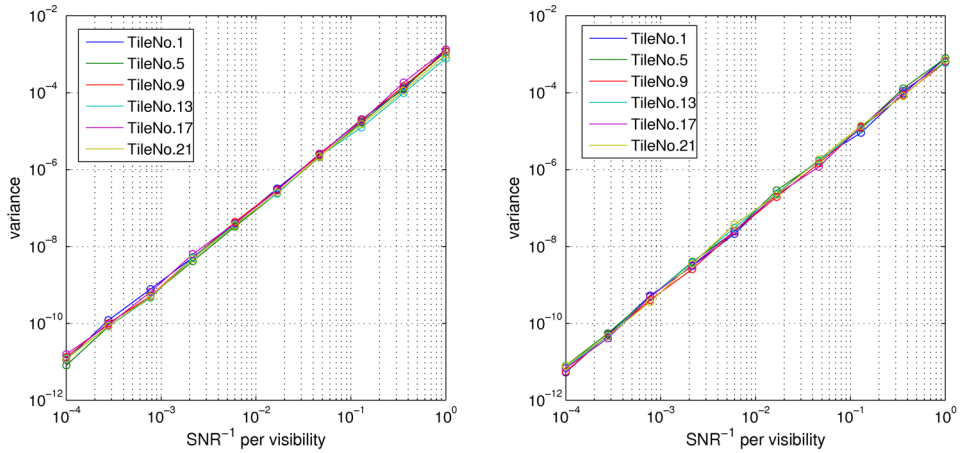


Figure 3.9: The variance of the estimated complex gains in terms of amplitude (left) and phase (right) versus the $(S/N)^{-1}$ per visibility. Each color represents the estimated complex gain of a tile whose number is mentioned in the legend.

3.3.3 Limitations of the redundancy calibration

The applicability of the redundancy calibration is limited by the following factors:

1. The station configuration.
2. Mutual coupling between the station elements, which leads to:
   - non-identical element beams.
   - the presence of baseline dependent noise.
3. the signal-to-noise ratio across the observed sky.

The station configuration can influence the suitability of redundancy in different ways. The redundancy calibration method requires a regular arrangement of station elements. In a station with p elements, one needs a sufficient number of distinct types of redundant baseline to have a system of equations in which all station element gains are involved. The more redundant the station's baselines are, the less information in the measured visibilities will be missed in the computation. This is not a concern in the HBA stations, as they are highly redundant. We denote by I the ratio of the number of the measured visibilities that are used in the computations, to the total number of the measured visibilities. The station configuration also determines the coefficient matrices i.e. $\mathbf{E}_{ph}$ or $\mathbf{E}_{ampl}$ for the phase and amplitude estimators, respectively. The condition number is a relative error magnification factor i.e. errors in the right-hand side of a linear system of equations can cause errors $\kappa(\mathbf{E}_{ph})$ times as large in the solution. Fig. 3.10 shows three different configurations for the HBA

stations in the LOFAR system. After recognizing the redundant baseline in a certain configuration, one can calculate the aforementioned quantities, which are given in Table 3.2. This table shows that RS208 is the most reliable configuration, from the redundancy calibration point of view. However, these quantities can be figures of merit when we design for redundancy within the stations and in terms of the arrangement of the stations within the whole array.

Configuration and element spacing in a station are usually decided based on obtaining low sidelobes of the station beam. Minimizing the mutual coupling between the station elements should be taken into consideration as other important figures of merit in a station layout for the sake of redundancy applicability. This consideration serves a twofold goal, obtaining identical element beams at least in the main beam and minimizing the receiver correlated noise ($\mathbf{R}_{\text{rec}}$). The dissimilarities between the element beams introduces systematic errors that cannot be eliminated by either any statistical method or a longer integration time. In the presence of the baseline dependent noise, the estimated parameters will deviate from their true values as given by Eq. 3.23. Quantifying this deviation for a given array, requires a good understanding of the noise terms in Eq. 3.7, especially $\mathbf{R}_{\text{rec}}$. The correlated noise terms are difficult to model analytically (Maaskant 2010). One may use a powerful and generic tool called CAESAR to compute them numerically using its ElectroMagnetic (EM) and MicroWave (MW) simulators, which is indeed an ongoing project at the HBA and EMBRACE stations. Primary results from the MW-simulation for an HBA station have demonstrated a negligible contribution of $\mathbf{R}_{\text{rec}}$. However, Fig. 3.9 indicates that whatever the effect of the calibration accuracy may be, it does not make the redundancy algorithm unstable.

Table 3.2: Station configuration and the condition of the redundancy calibration

Station	no. of DRB	size($\mathbf{E}_{ph}$)	size($\mathbf{E}_{ampl}$)	$\kappa(\mathbf{E}_{ph})$	$\kappa(\mathbf{E}_{ampl})$	I
RS208	84	1127×132	1125×132	52.72	9.74	99.65%
CS302	113	1123×161	1121×161	6.74e16	9.99	99.29%
CS004	72	547×120	545×120	8.42e16	1.73e15	48.22%

Note: Station column represents some LOFAR station names, each implying a different station configuration. DRB stands for distinct redundant baseline.

3.4 Discussion

We have studied the applicability of redundancy calibration to phased-array stations for the first time. The performance of this calibration method was demonstrated using data acquired with a new telescope, LOFAR. This required new considerations to the original design of conventional arrays e.g. WSRT and VLA. We took them into account by refining the data model. Reformulating the redundancy calibration

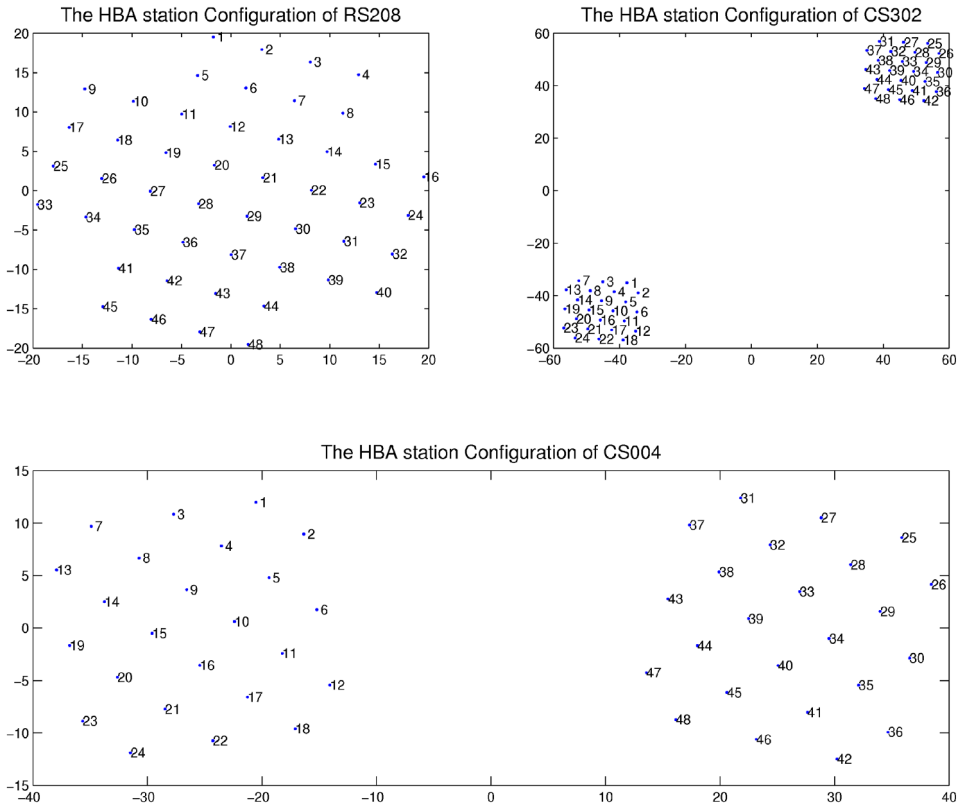


Figure 3.10: Three different station configuration for HBA stations.

formalism using the new data model, helped us to understand its potential, limitations and the effects of non-Gaussian baseline-dependent noise on the calibration accuracy.

We must consider both obtaining identical beams for station elements and minimizing the correlated noise as figures of merit for the configuration and spacings between the elements within a station in SKA pathfinders and the SKA itself. This is definitely necessary, regardless of the calibration method used. For the redundancy method, it is fundamental, while having identical element beams saves computational capacity and time for model-based methods.

In the plots of residuals and Table 3.1, we have demonstrated how non-identicalness of the beams leads to systematic errors that cannot be eliminated by any statistical methods or longer integration time. Since, identical element beams are fundamental especially for the redundancy calibration, similar EM-simulations to those presented here, are highly recommended for any array. Moreover, redundancy turns out to be an extremely suitable diagnostic tool for recognizing and monitoring failing station

elements. This is of concern in arrays with large number of elements, such as LOFAR and SKA. It is currently being used for this purpose in the LOFAR system.

Wide-field science with new radio telescopes such as SKA (van Ardenne et al. 2009), demands an enormous amount of signal processing and computational capacity. This is mainly due to the large number of elements, the variable beams, and the effect of the ionosphere over a large FoV requiring considerable calibration on the raw data. This has extensively been treated in the thesis work of Bregman (2012). The redundancy calibration results encourage us to recommend considering redundancy in the SKA configuration, both at the station and whole array level. In large interferometers such as SKA, where the baselines are non-coplaner, one has to however, think of redundancy in *uvw*-space, however, not only in *uv*-space. Redundancy can be applied to the gain calibration of the whole array in specific regimes where the stations observe the sky, through the same ionospheric patch. It also saves computational capacity and gives a more accurate estimate of the telescope gains than the model-based gain calibration. This is of course a major step forward in achieving radio images of higher dynamic range.

3.5 Conclusion

For the first time, we have studied the applicability and limitations of redundancy calibration in phased array stations using both real and simulated data of a new telescope, that uses aperture array technology such as LOFAR. The results clearly show that the additional contraints provided by redundant baselines do improve the quality of the calibration and in addition provide a powerful tool for system diagnostics at different levels of the telescope phased-array hierarchy (both intra-station and inter-station). The merit of redundancy in the station and full array layout and the additional advantage for diagnostics are clearly demonstrated by this study. We therefore recommend to ensure redundancy in both the station layout and the array configuration of future aperture arrays, in particular the SKA, where the required dynamic range will be an order of magnitude beyond any existing array. Because redundancy gives a better handle on characterizing the state of the system, it provides in addition a model-independent diagnostic tool for subsystems such as a station.

However, the effect of redundancy on the performance of the deconvolution process, thus the quality of the final image using different methods of deconvolution remains a critically important future research direction. Chapter 5 provides a suitable background, based on which one can insightfully plan such research project.

Chapter 4

A constrained self-calibration

In the 3rd generation of calibration (3GC), self-cal solves for more parameters per array element i.e. both the direction-independent effects (DIE) and the direction-dependent effects (DDE) in different stages. Direction-independent gain calibration takes place in the first stage, which should be done accurately. In 3GC, accurate gain calibration becomes even more important as the errors at this stage propagate to the next stages of the calibration. Such error propagation was not of concern in the 2nd generation of calibration (2GC). One way to improve the calibration results is to constrain the calibration routine at each stage such that it converges better to the true solutions.

In this chapter, we examine the refractive nature of the ionosphere as a DDE and a source of inaccuracy in gain calibration results. To avoid such error, we suggest solving a constrained self-cal for an array with an arbitrary geometry, using extra constraints as exploited in redundancy calibration. Those constraints are based on a physical model of the ionosphere refractive effect. We demonstrate that the constraints have a clear physical justification, only when we observe a single source or multiple sources, that experience similar ionospheric effect. We discuss how cluster calibration might reduce any other observational scenario to the one, in which these constraints are justified.

Self-cal is formulated as a non-linear least squares problem. Its constrained version can be solved using sequential quadratic programming (SQP) technique, if it cannot be linearized. We conclude that a constrained self-cal performs more optimally based on our numerical CRB analysis and comparing it with the CRB of its unconstrained version. Then, we solve a constrained self-cal taking two different approaches: the SQP and the logarithmic least squares (LOGLS). Their results show that the SQP performs more accurately as compared with the LOGLS, provided the same initial guess, and that self-cal can benefit from the constraints used in the redundancy calibration formalism to perform more reliably for gain calibration of an array with an arbitrary geometry.

4.1 Introduction

The term self-cal refers to a calibration scheme, that uses the observed known sources as the calibration beacons (Thompson et al. 1986; Schwab and Bridle 1994). Performing an accurate self-cal has been a challenge with the advent of new radio telescopes with wide fields of view, and operating at very low frequencies such as LOFAR (de Vos et al. 2009), a prominent pathfinder towards SKA. Therefore, 3GC[1] has been introduced to the radio astronomical community by Noordam and Smirnov (2010). 3GC enables us to tackle the instrumental gain corruptions (known as DIE) and other corrupting effects such as propagation path effects (known as DDE). Smirnov (2011a,b,c) has clearly explained how the 3GC trend can be implemented in radio astronomical data reduction packages.

In the 3GC approach, self-cal solves for telescope direction-independent complex gain, at its first stage. Those gains are instrument intrinsic properties and ideally should not show a large variation over time or frequency. However, the accuracy of gain calibration is affected by several sources of error. Two major sources of error, which are known as DDE are: a) propagation path effects, of which the Earth's ionosphere is the most dominant one, especially at low frequencies and b) variation of the station "primary" beam pattern over its field of view. The latter introduces a systematic error in gain calibration results, which cannot be eliminated by longer integration time or more iterations. Thus, it must be understood and separated from any other stochastic DDE such as the ionosphere. The variation in the beam pattern of a phased-array station over time is due to different beamforming weightings while observing a target for a long time. Its variation over frequency is due to electromagnetic properties and the design of the station antenna. Both causes of beam variability are deterministic and should be known a priori. In general, having a good knowledge of the instrument beam pattern is an absolute necessity. In this chapter, we assume that the antenna beam pattern of the array elements are identical and known. In contrast, ionosphere acts as a random refractive sheet and introduces a direction-dependent gain per source and per array element, that must be calibrated out. In the 3GC approach, ionospheric calibration takes place after gain calibration. However, the ionosphere disturbances in the measured data affect the accuracy of gain calibration results in the first stage.

The impact of inaccurate direction-independent gain calibration on the dynamic range of the final image has been known and well-understood for decades (Perley et al. 1985, Lecture. 10). In the 3GC approach, accurate gain calibration becomes even more important as those errors at this stage propagate to the next stages of the calibration and are magnified by the condition number of the corresponding Jones matrices (Carozzi 2011). Therefore, it is an incorrect notion that the errors, made at one stage of calibration can be handled at later stages. This emphasizes the importance of performing an accurate estimation at each stage of the calibration routine.

The problem of accurate direction-independent gain calibration in the presence

[1] 1GC relies on instrumental stability. In 2GC, self-cal solves only for two parameters (the amplitude and the phase of a complex gain) per array element (Noordam and Smirnov 2010).

of the DDEs has been acknowledged by all recent literature on radio interferometric data calibration. Here, we summarize the most recent work. First, let us categorize the calibration algorithms in two groups: a) covariance fitting in form of a least squares problem, briefly explained in Sect. 4.2, and b) expectation maximization. Among the first group are the works by Boonstra and van der Veen (2003), van der Tol et al. (2007), and Wijnholds and van der Veen (2009a). Boonstra and van der Veen (2003) study several calibration algorithms using a single calibrator source. They acknowledge that the gain solutions carry the effects of the propagation path. van der Tol et al. (2007) reformulate self-cal by taking into account the DDE and multiple calibrator sources in their data model. The motivation behind this is the wide FoV of the new radio telescopes leading to the presence of multiple calibrator sources in their observations. They recognize a unitary ambiguity in this model-based calibration formalism. They mathematically prove how this unitary ambiguity leads to un-calibratibility of the array (or singularity of the Fisher information matrix (FIM) in the CRB derivation). Their suggested solution is to add deterministic and physically justifiable constraints to a group of parameters, that are estimated in the calibration routine. They constrain the ionospheric parameters. Wijnholds and van der Veen (2009a) also consider the DDE and multiple calibrator sources in the data model, but they model the DDE differently. They suggest a WALS algorithm, that iteratively solves for the direction-independent gains of the array elements, their noise powers, and the direction-dependent gains in the direction of the calibrator sources. But, there is no mathematical proof that taking the WALS approach can sort out the unitary ambiguity in the calibration solutions. Therefore, providing some a priori and physically justifiable information to constrain the calibration parameters, so that the calibration routine converges to the true solutions seems inevitable. This has thoroughly been discussed by van der Tol et al. (2007). Moreover, it is generally known that the more parameters we estimate using the same observation, the larger the variance we allow in the estimated parameters.

Among the second category of calibration algorithms is the space alternating generalized expectation maximization (SAGE) calibration technique (Yatawatta et al. 2008; Kazemi et al. 2011, 2012), which exploits the statistical properties of the noise. Although this technique reaches a superior calibration performance and convergence speed, as compared with the least squares-based (model-based) calibration method, the authors also acknowledged that the use of some justifiable constraints on parameters space is required to force the calibration routine to converge to the true solutions.

An excellent example of such constraints are the ones used in the redundancy calibration (Eq. 3.20 and Eq. 3.21). They are used to avoid ionospheric/tropospheric errors in the gain calibration results, which predominantly appear as a phase shift over the array. In Sect. 4.3, we show that those constraints are physically justifiable only, when we calibrate against a single source or multiple sources, that experience the same ionosphere disturbances. However, as it was discussed by Kazemi et al. (2011), we should always calibrate against a single source in wide-field radio interferometric data processing to reduce the parameters space. It is done by clustering the observed sky in facets centered at the strong sources in the field, and assuming that certain

effects e.g. ionosphere disturbances are the same for all the other sources in that facet. Thus, our constraints are justified, if the fixed parameter in the facet is the ionosphere. Moreover, calibration using a single source does not suffer from the unitary ambiguity (see the mathematical proof in van der Tol et al. (2007)). This condition is provided taking the cluster calibration approach. In Sect. 4.4, we demonstrate that a single ionospheric screen casues a phase shift over the array in gain calibration results. For this demonstration, we use a single calibrator source and a one-dimensional linear equi-spaced array (LEA) only in order to obtain a clear slope and compare it with the predicted slope in Sect. 4.3. Prior to suggesting a solution to avoid such an error in gain calibration results, we show that using the aforementioned constraints indeed improves the optimality of the self-cal gain estimation by the CRB analysis in Sect. 4.5. We note that our constraints involve all the array elements. Therefore, we use self-cal formalism using a full ME as it is done in Boonstra and van der Veen (2003); van der Tol et al. (2007); Wijnholds and van der Veen (2009a), and explained in Sect. 4.2. Then, we suggest solving a constrained self-cal for an array with an arbitrary geometry to avoid such an error in the complex gain solutions in Sect. 4.6. Our approach is very similar to that of (van der Tol et al. 2007) i.e. adding deterministic and physically justifiable constraints to a group of parameters, that are estimated. van der Tol et al. (2007) constrain the ionosphere parameters whereas we propose to constrain the phases of the direction-independent complex gain of the array elements. Our problem becomes a non-linear constrained least squares problem. To solve that, we take two approaches: a) using the linearized version of the model-based calibration called logarithmic least squares (LOGLS) studied in Boonstra and van der Veen (2003) and b) the state of the art approach in non-linear optimization known as sequential quadratic programming (SQP) (Fletcher 1980; Gill et al. 1981). We conclude that a constrained self-cal achieves a more accurate gain estimation result, the impact of which on the final image has been known for decades (Perley et al. 1985, Lecture. 10) and has analytically been calculated by Wijnholds and van der Veen (2008).

4.2 Self-cal, a non-linear least squares problem

In the radio astronomical terminology, self-cal refers to the calibration of a telescope using sky signals in the observation as reference sources (Thompson et al. 1986; Schwab and Bridle 1994). It can solve for the direction-independent electronic complex gain (in terms of amplitude γ_i and phase ϕ_i), and the noise power (σ_i) of the array elements in vector $\boldsymbol{\theta}$, in a least squares covariance fitting

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\| \hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right\|_F^2 \tag{4.1}$$

$$\boldsymbol{\theta} = [\gamma_1, \gamma_1, ..., \gamma_p, \phi_1, \phi_2, ..., \phi_p, \sigma_1^2, \sigma_2^2, ..., \sigma_p^2]^T,$$

where ϕ_1 is assumed to be the phase reference ($\phi_1 = 0$), because an absolute phase definition is meaningless. $\mathbf{R}(\boldsymbol{\theta}) = \mathbf{G}\mathbf{A}\boldsymbol{\Sigma}_{\mathrm{s}}\mathbf{A}^H\mathbf{G}^H + \boldsymbol{\Sigma}_{\mathrm{n}}$ and $\hat{\mathbf{R}}$ are the modelled and the measured array output covariance matrices (visibilities), respectively. $\mathbf{G}$ (size $p \times p$)

is a diagonal matrix whose elements are the complex electronic gains, $g_i = \gamma_i e^{\phi_i}$. It is assumed that the calibrator sources have known position and fluxes ($\mathbf{\Sigma}_\mathrm{s}$). Thus, the array response matrix, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_q]$ is known and defined as

$$\mathbf{A} = exp\left(-j\frac{2\pi}{\lambda}\mathcal{R}\mathcal{L}^T\right), \tag{4.2}$$

where the matrix $\mathcal{R} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_p]^T$ (size $p\times 3$) is the stack of array element positions, if the i-th element of the array is located at $\mathbf{r}_i = [x_i, y_i, z_i]^T$. Likewise, the source positions are stacked in a matrix $\mathcal{L} = [\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \boldsymbol{\ell}_q]^T$ (size $q \times 3$), the position of the k-th calibrator source is $\boldsymbol{\ell}_k = [l_k, m_k, n_k]^T$. λ is the wavelength. p is the number of the array elements and q is the number of the calibrator sources. Then, an adequate sky model can be represented by

$$\mathbf{A}\mathbf{\Sigma}_\mathrm{s}\mathbf{A}^H. \tag{4.3}$$

One can point out two major issues here:

1. The unitary ambiguity is present in Eq. 4.1. The unitary ambiguity means that one obtains different complex gain solutions for the instrument using different observations, while those complex gains are the instrument intrinsic properties and should not depend on what we observe. As it was mathematically described by van der Tol et al. (2007), without further constraint to reduce the degrees of freedom of $\boldsymbol{\theta}$, the solution of Eq. 4.1 is not identifiable through a single measurement.

2. In the presence of a DDE e.g. the ionosphere, an appropiate model of such an effect is a matrix $\mathbf{G}_0$ (size $p \times q$), element-wise multiplied with the array response matrix, $\mathbf{A}$. Thus, the data model becomes: $\mathbf{R}_1(\boldsymbol{\theta}) = \mathbf{G}(\mathbf{G}_0 \odot \mathbf{A})\mathbf{\Sigma}_\mathrm{s}(\mathbf{G}_0 \odot \mathbf{A})^H\mathbf{G}^H + \mathbf{\Sigma}_\mathrm{n}$. This makes the estimation of $\mathbf{G}_0$ (size $p\times q$) parameters not straightforward, as extensively discussed by van der Tol et al. (2007). Modelling it in a matrix $\mathbf{G}_0$ (size $p \times p$) such that the data model becomes: $\mathbf{R}_2(\boldsymbol{\theta}) = \mathbf{G}\mathbf{G}_0\mathbf{A}\mathbf{\Sigma}_\mathrm{s}\mathbf{A}^H\mathbf{G}_0^H\mathbf{G}^H + \mathbf{\Sigma}_\mathrm{n}$ (Wijnholds and van der Veen 2009a) leads to estimation of incorrect parameters, although this data model makes the estimation process straightforward.

Using the data model in $\mathbf{R}_1(\boldsymbol{\theta})$, a supplementary step is to apply physically justifiable constraining models, as discussed by van der Tol et al. (2007). The constraints should fully affect the underlying statistical distribution of the signal samples, that are correlated to form the visibilities, in $\hat{\mathbf{R}}$. van der Tol et al. (2007) use such constraining models on ionospheric parameters. Here, we take a similar approach but constrain the complex gains in the same manner as is done in redundancy calibration (Wieringa 1991). Those constraints are to set the phase gradient over the array caused by the source position shift, to zero. In Sect. 4.3, we demonstrate the physical justification of those constraints.

Without loss of generality, we reduce the vector of parameters to be estimated in self-cal to only complex gains $\boldsymbol{\theta} = [\gamma_1, \gamma_1, ..., \gamma_p, \phi_2, \phi_3, ..., \phi_p]^T$. It was also shown

in van der Tol and Wijnholds (2006) that the impact of unknown diagonal $\mathbf{\Sigma}_{\rm n}$ on complex gain estimation is none. Moreover, for any estimate of the complex gains $\hat{\mathbf{g}}$, the optimal estimate of $\hat{\mathbf{\Sigma}}_{\rm n}$ is still given by $\hat{\mathbf{\Sigma}}_{\rm n} = {\rm diag}(\hat{\mathbf{R}} - \hat{\mathbf{g}}\hat{\mathbf{g}}^H)$.

We note that the array elements are assumed to have the same and known beam pattern towards the q source signals received by the array.

4.3 Ionosphere and phase gradient over the array

The ionosphere acts as a random refractive sheet, shifting the positions of the sources in the FoV. This causes a DDE modelled by the matrix $\mathbf{G}_0$, whose structure depends on the calibration regime (explained in Lonsdale (2004)) of our observation. In the most general case, it is

$$\mathbf{G}_0 = \begin{bmatrix} e^{-j2\pi\frac{\mathbf{r}_1\cdot\Delta\boldsymbol{\ell}_{11}}{\lambda}} & e^{-j2\pi\frac{\mathbf{r}_1\cdot\Delta\boldsymbol{\ell}_{21}}{\lambda}} & \cdots & e^{-j2\pi\frac{\mathbf{r}_1\cdot\Delta\boldsymbol{\ell}_{q1}}{\lambda}} \\ e^{-j2\pi\frac{\mathbf{r}_2\cdot\Delta\boldsymbol{\ell}_{12}}{\lambda}} & e^{-j2\pi\frac{\mathbf{r}_2\cdot\Delta\boldsymbol{\ell}_{22}}{\lambda}} & \cdots & e^{-j2\pi\frac{\mathbf{r}_2\cdot\Delta\boldsymbol{\ell}_{q2}}{\lambda}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi\frac{\mathbf{r}_p\cdot\Delta\boldsymbol{\ell}_{1p}}{\lambda}} & e^{-j2\pi\frac{\mathbf{r}_p\cdot\Delta\boldsymbol{\ell}_{2p}}{\lambda}} & \cdots & e^{-j2\pi\frac{\mathbf{r}_p\cdot\Delta\boldsymbol{\ell}_{qp}}{\lambda}} \end{bmatrix}, \qquad (4.4)$$

in which q sources are shifted differently with respect to each array element. $\Delta\boldsymbol{\ell}_{qp}$ is the position shift of the q-th source with respect to the p-th array element.

We show how the position shifts of the calibrator (reference) sources are transformed into the errors on the self-cal gain estimation results. Due to the diagonality of $\mathbf{G}$ in the data model, $\mathbf{R}_1(\boldsymbol{\theta})$ the position shifts in $\mathbf{G}_0$ can be absorbed in the $\mathbf{G}$ matrix introducing direction-dependent errors on direction-independent complex gains. These errors can be studied by computing a matrix $\mathbf{L}$ (size $p \times p$), that holds the relation $\mathbf{G}_0 \odot \mathbf{A} = \mathbf{L}\mathbf{A}$, and is diagonalized by an eigen-value decomposition. Intuitively, we expect that the position shifts of the calibrator sources in Eq. 4.4 lead to errors in both phase and amplitude of the complex gains. However, the simplest case to study is when we observe a single source, $q = 1$ or the shifts, $\Delta\boldsymbol{\ell}_{qp}$ are all the same for all the sources in the FoV. In both cases, we can write

$$\mathbf{G}_0 \odot \mathbf{A} = \begin{bmatrix} e^{-j2\pi\frac{\mathbf{r}_1\cdot\Delta\boldsymbol{\ell}}{\lambda}} & 0 & \cdots & 0 \\ 0 & e^{-j2\pi\frac{\mathbf{r}_2\cdot\Delta\boldsymbol{\ell}}{\lambda}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-j2\pi\frac{\mathbf{r}_p\cdot\Delta\boldsymbol{\ell}}{\lambda}} \end{bmatrix} \mathbf{A} = \mathbf{L}\mathbf{A}. \qquad (4.5)$$

Then, we obtain a naturally diagonal matrix $\mathbf{L}$. The diagonal elements of $\mathbf{L}$ could be element-wise multiplied to the diagonal terms of matrix $\mathbf{G} = {\rm diag}(\mathbf{g})$ in $\mathbf{R}_1(\boldsymbol{\theta})$, for the cost of keeping an known sky model ($\mathbf{A}\mathbf{\Sigma}_{\rm s}\mathbf{A}^H$ in $\mathbf{R}_1(\boldsymbol{\theta})$) for solving Eq. 4.1. This

way, the error is absorbed in the complex gains resulting to an inaccurate complex gain estimation. In such a case, we estimate for matrix $\mathbf{G}' = \mathbf{GL}$, that is

$$\mathbf{G}' = \begin{bmatrix} \gamma_1 e^{j\phi_1} & 0 & \cdots & 0 \\ 0 & \gamma_2 e^{j\phi_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_p e^{j\phi_p} \end{bmatrix} \mathbf{L} = \\ \begin{bmatrix} \gamma_1 e^{j(\phi_1 - 2\pi \frac{\mathbf{r}_1 . \Delta \boldsymbol{\ell}}{\lambda})} & 0 & \cdots & 0 \\ 0 & \gamma_2 e^{j(\phi_2 - 2\pi \frac{\mathbf{r}_2 . \Delta \boldsymbol{\ell}}{\lambda})} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_p e^{j(\phi_p - 2\pi \frac{\mathbf{r}_p . \Delta \boldsymbol{\ell}}{\lambda})} \end{bmatrix}.$$

This has changed the vector $\mathbf{g} = [\gamma_1 e^{j\phi_1}, \gamma_2 e^{j\phi_2}, ..., \gamma_p e^{j\phi_p}]^T$ to

$$\mathbf{g}' = [\gamma_1 e^{j(\phi_1 - 2\pi \frac{\mathbf{r}_1 . \Delta \boldsymbol{\ell}}{\lambda})}, \gamma_2 e^{j(\phi_2 - 2\pi \frac{\mathbf{r}_2 . \Delta \boldsymbol{\ell}}{\lambda})}, ..., \gamma_p e^{j(\phi_p - 2\pi \frac{\mathbf{r}_p . \Delta \boldsymbol{\ell}}{\lambda})}]^T. \tag{4.6}$$

If we rewrite it in

$$\mathbf{g}' = [\gamma_1 e^{j(\phi_1 - 2\frac{\pi}{\lambda}(\Delta l x_1 + \Delta m y_1))}, \gamma_2 e^{j(\phi_2 - 2\frac{\pi}{\lambda}(\Delta l x_2 + \Delta m y_2))}, ..., \gamma_p e^{j(\phi_p - 2\frac{\pi}{\lambda}(\Delta l x_p + \Delta m y_p))}]^T \tag{4.7}$$

it becomes clear that a phase slope, $-2\frac{\pi}{\lambda}(\Delta l x + \Delta m y)$ has been added to the phases of the complex gains and the amplitudes have remained unaffected. This phase slope is over the position of the array elements (x, y). Therefore, the element numbering is important for recognizing such a slope. In a regular array either 1-D LEA or 2-D (Uniform Linear Array: ULA), it comes naturally. Note that in an array with an arbitrary geometry, one has to number the elements in x or y in ascending order.

Eq. 4.4 is the most realistic case for the aperture array telescopes, as their calibration mainly falls in the regime, where either the ionosphere varies significantly over the FoV (regime 3) or where different array elements view different sections of the ionosphere (regime 4) (see Fig. 1.7 and Lonsdale (2004)). But as discussed earlier, using the cluster calibration (Kazemi et al. 2011), our problem is always reduced to its simplest form in Eq. 4.5, thus the justification presented here can always hold.

In the next subsection, we demonstrate how the source position shift is absorbed in the self-cal gain estimation results, given that a correct sky model is provided for solving Eq. 4.1.

4.4 Predicted phase gradient in self-cal results

We have set up a simple simulation to demonstrate how self-cal gain estimation results are affected by a position shift of the calibrator source $(\Delta l, \Delta m)$. We choose a one-dimensional LEA array of $p = 40$ elements only in order to obtain a clear slope

and compare it with the predicted slope in Eq. 4.7. But, in the following sections, we will use an arbitrary array geometry for our analysis. Here, the gain amplitude γ_i are chosen randomly in the interval $\gamma_0 - \Delta\gamma$ and $\gamma_0 + \Delta\gamma$, where $\gamma_0 = 1$ is the nominal gain amplitude, and $\Delta\gamma = 0.2$ is a spreading parameter. The gain phase ϕ_i lies uniformly distributed between 0 and $\Delta\phi = 0.2$ radian, where $\Delta\phi < 2\pi$. The astronomical calibrator source is a zero mean complex Gaussian signal with unit power. The frequency is $f = 170$ MHz, thus $\lambda = 1.7635$ m. The calibrator is shifted by $\Delta l = 0.02$. We solve for the complex gains using Eq. 4.1. The noise is dismissed for this simulation. Noisy simulated data can be used in further studies. We set the spreading parameter small to obtain a slope close to what we predicted by Eq. 4.7, $-2\pi\frac{\Delta l}{\lambda} = -0.0713$. This is only to get our point across clearly. The results are presented in Fig. 4.1. From these results, it is clear that a shift in position of the calibrator source has been translated to a phase slope over the array, in the phases of the complex gain estimation results. We expected this from the theory explained earlier and captured in Eq. 4.7.

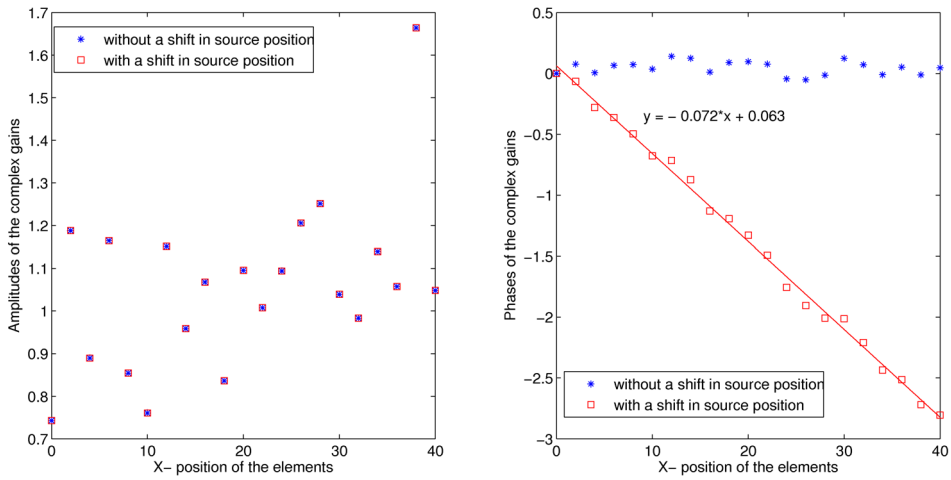


Figure 4.1: Estimated complex gains using Eq. 4.1, in terms of amplitudes (left) and phases (right) using a single source once at $(l, m) = (0.3, -0.7)$ (blue star) and once shifted 0.02 in l direction (red squares). The array geometry is a LEA and the spreading parameter is small. These are only in order to obtain a clear slope and compare it with the predicted slope $-2\pi\frac{\Delta l}{\lambda} = -0.0713$.

The redundancy calibration (Wieringa 1991) constrains the phases of the complex gains, such that the phase slope over both x and y is forced to be zero. Those constraints are expressed in Eq. 3.20 and Eq. 3.21, which we rewrite in

$$\mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{i=1}^{p} \phi_i x_i \\ \sum_{i=1}^{p} \phi_i y_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}\mathbf{x}^T \\ \boldsymbol{\phi}\mathbf{y}^T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}_{2\times 1}. \tag{4.8}$$

Without loss of generality and for mathematical convenience, we rewrite them as

$$\mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\theta}[\mathbf{0}_{1\times p}|\mathbf{x}]^T \\ \boldsymbol{\theta}[\mathbf{0}_{1\times p}|\mathbf{y}]^T \end{bmatrix} = \mathbf{0}_{2\times 1}, \tag{4.9}$$

where x_i and y_i are the (x, y) coordinates of the array elements stacked in the vectors $\mathbf{x}^T$ and $\mathbf{y}^T$ (size $1 \times p$), respectively.

In Sect. 4.6, we suggest solving a constrained self-cal to avoid the phase slope error in its gain estimation results using the constraints in Eq. 4.9. Thus, we intend to incorporate the phase slopes into the calibration formalism as physically accepted constraints, rather than treating them separately in a later stage of the calibration. We use an array with an arbitrary geometry to demonstrate that such slope is observed even when we have an irregular non-liner array and that our suggested solution avoids the corresponding error in the self-cal gain estimation results.

An important assumption is that an intrinsic phase slope over the array is very unlikely.

4.5 Optimality of a constrained self-cal

In the signal processing literature, a standard and robust step to ensure whether a constrained algorithm works more optimally than its unconstrained version is to compute the CRB. We accomplish such computation in this section. The CRB analysis determines limiting estimation error variance under various model assumptions. The CRB on the error variance for any unbiased estimator is given by the diagonal elements of matrix $\mathbf{C}$ (Kay 1993), given in

$$\mathbf{C} = \mathcal{E}\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right) \geq \frac{1}{N}\mathbf{J}^{-1}, \tag{4.10}$$

where $\mathcal{E}(.)$ is expectation operation, N is the number of samples, and $\mathbf{J}$ is the Fisher information matrix (FIM). Stoica and Ng (1998) have proven that under parametric and continuously differentiable constraints, $\mathbf{f}(\boldsymbol{\theta}) = \mathbf{0}$ (such as the constrains in Eq. 4.9), the CRB is expressed by

$$\mathbf{C} = \mathcal{E}\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right) \geq \frac{1}{N}\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^{-1}\mathbf{U}, \tag{4.11}$$

where $\mathbf{U}$ is a matrix whose columns form an orthonormal basis for null-space of the gradient matrix of the constrains matrix

$$\mathbf{F}(\boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}. \tag{4.12}$$

Closed form relations for the CRB for self-cal were derived in earlier work (Boonstra and van der Veen 2003; Wijnholds and Boonstra 2006; van der Tol et al. 2007), thus $\mathbf{J}$ is known. We numerically study the effect of our constraints on the CRB using Eq. 4.11 and the constraints, $\mathbf{f}(\boldsymbol{\theta})$ in Eq. 4.9. The results are presented in Fig.4.2. It shows that using the constraints indeed reduces the CRB or equivalently increases the optimality of the self-cal gain estimation. One may notice that the amplitudes are not affected. This is because we only constrained the phases.

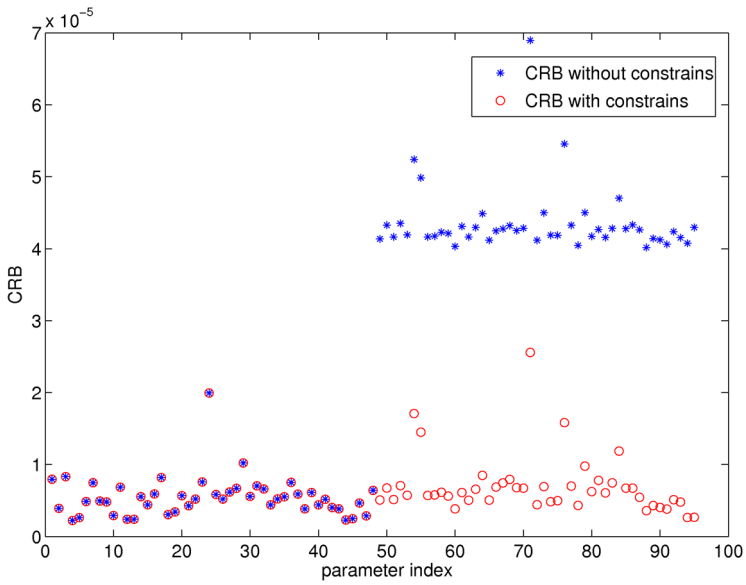


Figure 4.2: The effect of parametric constraints on the error variance in the self-cal gain estimation results. It shows that the level of variance in phases is lower by using the constraints. An array of 48 element with an arbitrary geometry has been used. The x-axis represents the index for the parameters $\boldsymbol{\theta} = [\gamma_1, \gamma_2, ..., \gamma_p, \phi_2, \phi_3, ..., \phi_p]^T$. It shows that using the constraints indeed reduces the CRB or equivalently increases the optimality of the estimation.

Deriving an analytical expression for the constrained CRB remains as a further step to be taken.

4.6 Suggested solutions to a constrained self-cal

Self-cal formulated in Eq. 4.1 is a nonlinear least squares problem. Constraining Eq. 4.1 means solving a constrained nonlinear problem, which is to minimize the objective or cost function $J(\theta)$, subject to the constraints in Eq. 4.9 given in

$$J(\theta) = J(\boldsymbol{\gamma}, \boldsymbol{\phi}) = \left\| \hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right\|_F^2 \tag{4.13}$$
$$subject\ to: \ \mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\theta}[\mathbf{0}_{1\times p}|\mathbf{x}]^T \\ \boldsymbol{\theta}[\mathbf{0}_{1\times p}|\mathbf{y}]^T \end{bmatrix} = \mathbf{0}_{2\times 1}.$$

Here, we face constraining a complex quantity, which boils down to constraining either amplitude and phase or real and imaginary parts, separately. One cannot physically constrain an imaginary quantity as it does not have a physical meaning. Therefore, we take a natural logarithm of the complex quantity, $\boldsymbol{\theta}$ to separate the amplitudes and phases. Since we intend to optimize only the phase estimation, we redefine $\mathbf{f}(\boldsymbol{\theta})$ in Eq. 4.13, reducing it to only the phase parameters, $\boldsymbol{\phi}$ in

$$\mathbf{f}(\boldsymbol{\phi}) = \begin{bmatrix} \boldsymbol{\phi}\mathbf{x}^T \\ \boldsymbol{\phi}\mathbf{y}^T \end{bmatrix} = \mathbf{0}_{2\times 1}. \tag{4.14}$$

All algorithms to solve a problem such as the one in Eq. 4.13 are well-known to be iterative, thus computationally intensive, unless the problem can be linearized. Here, we take both approaches a) the closed-form linearized version of Eq. 4.1, originally presented as LOGLS algorithm in Boonstra and van der Veen (2003) and b) the state of the art method in non-linear optimization known as SQP (Fletcher 1980; Gill et al. 1981).

4.6.1 Constrained logarithmic least squares

Boonstra and van der Veen (2003) propose a closed-form linearized version of the problem in Eq. 4.1, that has the computationally complexity of $2p^2$, where p is the number of elements in an array. They decouple amplitude and phase estimation by taking an element-wise natural logarithm of Eq. 4.1. The phase optimization problem as formulated in Eq. 4.13 becomes

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi},\mathbf{k}}{\operatorname{argmin}} \|\mathbf{I}^s(\mathbf{v}_I + \mathbf{k}2\pi) - \mathbf{I}^s[\mathbf{1}\otimes\mathbf{I} - \mathbf{I}\otimes\mathbf{1}]\boldsymbol{\phi}\|_F^2 \tag{4.15}$$
$$subject\ to: \ \mathbf{f}(\boldsymbol{\phi}),$$

where $\mathbf{v}_I = \text{vec}(\text{Im}\ \ln(\hat{\mathbf{R}}))$, $\mathbf{k}$ (size $1 \times p^2$) is a vector of nuisance integer parameters, and $\mathbf{I}^s$ (size $(p^2 - p) \times p^2$) is a selection matrix, that is constructed from a $p^2 \times p^2$ indentity matrix by removing its rows $1, (p+1)+1, 2(p+1)+1, ..., p^2$. Eq. 4.15 is solved in a closed form in

$$\hat{\boldsymbol{\phi}} = (\mathbf{C}_e^H \mathbf{C}_e)^{-1} \mathbf{C}_e^H (\mathbf{v}_e + 2\pi\mathbf{k}_e), \tag{4.16}$$

where

$$\mathbf{C}_e = \begin{bmatrix} \mathbf{C} \\ \mathbf{e}_1^t \\ \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}, \quad \mathbf{v}_e = \begin{bmatrix} \mathbf{v}_I \\ \phi_1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{k}_e = \begin{bmatrix} \mathbf{k} \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{4.17}$$

where $\mathbf{e}_i$ is the i-th unit vector (all zeros except for the i-th element) and $\mathbf{f}_i$ is the i-th row of the constraint matrix, $\mathbf{f}$. Suppose that we have an initial guess $\boldsymbol{\phi}^\circ$ of $\boldsymbol{\phi}$. Then, $\mathbf{k}$ is obtained by rounding the entries of $\mathbf{C}\boldsymbol{\phi}^\circ - \mathbf{v}_I$ to the nearest multiple of 2π: $\mathbf{k} = \text{round}((2\pi)^{-1}(\mathbf{C}\boldsymbol{\phi}^\circ - \mathbf{v}_I))$.

All the conventional solutions for a constrained linearized problem in (Lawson and Hanson 1974) are essentially applicable here. The presented solution in Eq. 4.16, however is the least complicated one.

4.6.2 Sequential quadratic programming

An alternative solution to our problem in Eq. 4.13 is direct optimization without explicit linearization using SQP technique. The principal idea in SQP technique is to formulate a QP sub-problem based on a quadratic approximation of the Lagrangian function given as

$$L(\boldsymbol{\phi}, \boldsymbol{\lambda}) = J(\boldsymbol{\phi}) + \sum_{i=1}^{m} \lambda_i \mathbf{f}_i(\boldsymbol{\phi}), \tag{4.18}$$

where λ_i is the Lagrange multiplier and $\mathbf{f}_i(\boldsymbol{\phi})$ is the i-th row of the constraint matrix. The QP sub-problem is

$$\begin{aligned} \min \ & \frac{1}{2}\mathbf{d}^T \mathbf{H}_k \mathbf{d} + \nabla J(\boldsymbol{\phi})^T \mathbf{d} \\ \textit{subject to}: \ & \nabla \mathbf{f}_i(\boldsymbol{\phi})^T \mathbf{d} + \mathbf{f}_i(\boldsymbol{\phi}) = 0, \end{aligned} \tag{4.19}$$

where $\mathbf{d}_k$ is a search direction to decrease a merit function (Gill et al. 1981). $\mathbf{H}_k$ is the Hessian of the Lagrangian function in Eq. 4.18, that can be computed using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. The solution is used to form a new iteration by

$$\boldsymbol{\phi}_{k+1} = \boldsymbol{\phi}_k + \alpha_k \mathbf{d}_k, \tag{4.20}$$

α_k is the step length parameter, that is determined by an appropriate line search procedure so that a sufficient decrease in the merit function is obtained. This technique has been implemented in a function called *fmincon* in the MATLAB optimization toolbox. The reader is referred to its documentation for further details.

We applied the two aforementioned algorithms (LOGLS and SQP) for the simulated data to solve a constrained self-cal. An array with an arbitrary geometry is used to simulated the array visibilities, $\hat{\mathbf{R}}$ in Eq.4.1. The rest of the simulation settings are the same as the simulation in Sect. 4.4. The results are given in Fig. 4.3. This figure

presents the array geometry in left panel, that is used for obtaining the estimation results in right panel. The right panel shows that one can observe a phase slope in the self-cal gain estimation results for an array with an arbitrary geometry, although the slope is not as clear as it is in Fig. 4.1. To obtain these results, we used the true value of the phase vector as the initial guess for these algorithms. The results presented required only one iteration. It is clear that the SQP performs better as compared with the LOGLS. Although its formalism is more sophisticated and general than the LOGLS, it performs fast due to its efficient implementation in MATLAB.

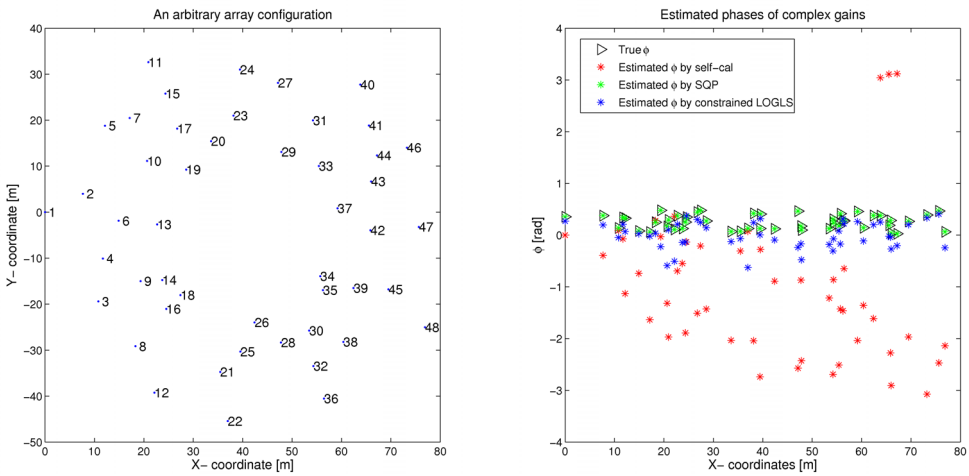


Figure 4.3: An array with arbitrary geometry (left). This array geometry is used for simulating $\hat{\mathbf{R}}$ and $\mathbf{R}$ to solve a constrained self-cal (Eq. 4.13) by running two different methods, a) constrained LOGLS and, b) SQP shown in the right panel. These results are obtained after only one iteration, provided the true value as the initial guess. Note: a 2π jump is seen in a few points in the top-right of the plot. This is taken care of in a more advanced implementation of self-cal.

It is known that the performance of all iterative methods depends on the accuracy of the initial guess. But, more iterations should improve their performance, if an inaccurate initial guess is provided. Further work is required to study the accuracy of the SQP performance versus the accuracy of the initial guess, or the number of required iterations to converge (speed of convergence), when an inaccurate initial guess is provided for this algorithm.

4.7 Conclusion

In this chapter, we examined the refractive feature of the ionosphere as a DDE that is a source of error in the self-cal gain estimation results. We suggested a constrained

self-cal to force the calibration algorithm to converge to the true gain solution, in the presence of this effect. We used the constraints, which the redundancy calibration has exploited (its formalism in Wieringa (1991)). Before solving a constrained self-cal, we took two steps:

1. We showed that those constraints have a clear physical justification, only when we observe a single source or multiple sources, that experience similar ionospheric effect. We discussed how cluster calibration might reduce any other observational scenario to the one, in which these constraints are justified.

   This also means that those constraints are not justified for the station calibration, which falls in calibration regime 3. This is a fair point to make, after the discussion in Sect. 4.3.

2. We performed a numerical CRB analysis. The results proved that our suggested constrained self-cal performs more optimally than unconstrained self-cal.

Finally, we took two different approaches (LOGLS and SQP) to test the performance of our suggested solution. The results are presented in Fig. 4.3. It demonstrates that:

1. the predicted slope in Fig. 4.1 is also observed in an array with an arbitrary geometry (not necessarily with an LEA geometry).

2. the self-cal can benefit from the constraints used in the redundancy calibration formalism to perform more reliably for gain calibration of such an array.

3. for this purpose, the SQP performs more accurately as compared with the LOGLS, provided the same initial guess.

However, we should note that we solved a constrained self-cal in Sect. 4.6 using a full ME, because our constraints involved all the array elements at once instead of individual baselines. Given that 3GC emphasizes the use of a 2×2 ME i.e. use of the data of an isolated baseline in the estimation process, we should seek a feasible way of implementing the constraints such as the ones applied in this chapter, in 3GC packages. This remains as an important challenge to tackle.

We also note that we discussed the first category of calibration i.e. the covariance fitting represented by Eq. 4.1. An extensive research on the second category of calibration algorithms has resulted in developing SAGE (Yatawatta et al. 2008; Kazemi et al. 2011) and Ordered Subsets-SAGE (OS-SAGE) (Kazemi et al. 2012). They are proven to be more accurate and computationally more efficient than the first category. However, this does not reduce the importance of constraining the calibration routine to converge to the true solutions, as Kazemi et al. (2011) concluded as well. This is the main point, that we emphasize in this chapter. Constraining gain calibration is one step.

We recommend constraining the other stages of the calibration routine in 3GC by as much a priori information about the instrument and the propagation path as is possible. The examples are:

1. the use of the ionospheric constraining model in van der Tol et al. (2007).

2. the use of a proper and complete model of the primary beam (both main lobe and sidelobes) that includes the effect of mutual coupling with additional physically justifiable constraints e.g. certain level of main lobe and sidelobes known from the electromagnetic studies, or beam positivity etc. This has been dismissed completely, thus far.

The work presented in this chapter is at a preliminary stage. We propose the following further steps:

1. Derive an analytical expression for the constrained CRB in Eq. 4.11.

2. Compare the performance of the SQP with other competitive methods of optimization such as *trust-region-reflective* or *active-set* (Gill et al. 1981), implemented in the same programming language with the same level of efficiency.

3. Reproduce the results in Fig. 4.3 using noisy data and more than one iteration, to test the convergence accuracy and speed of the optimization methods.

4. Compare the performance of the SQP (as the method of choice thus far) versus the accuracy of the initial guess.

5. Investigate a feasible solution for constraining the 2×2 ME-based calibration packages, not only the direction-independent gains but also other parameters. This can be done by providing the calibration routine with as much a priori information about the instrument and the propagation path features as possible. Examples of such constraints were given earlier.

6. Study whether the SAGE or OS-SAGE can also benefit from the constraints discussed in this chapter for a better performance.

7. Study whether the constrained self-cal produces a result that is comparable to the current SAGE or OS-SAGE, and whether it is computationally more or less expensive. This is an open question and a promising area of future research.

Chapter 5

Optimum modelling of an extended source in image deconvolution

To achieve the ultimate goal of high fidelity high dynamic range radio imaging, the bright known sources in the FoV (either point-like sources or extended structures) have to be modelled and removed accurately such that the faint unknown sources in the background can appear with minimum error. To this end, we discuss the use of orthonormal basis functions, that span the whole source area as opposed to the pixel-based basis functions, for optimal modelling of the extended sources in the deconvolution process. We will give an overview of existing deconvolution methods and discuss how they can benefit from this scheme. Following the promising results of using orthonormal basis functions, especially the prolate spheroidal wave functions (Yatawatta 2010, 2011b,a), we suggest a solution to reduce the computational cost of their implementation (Noorishad and Yatawatta 2011).

Based on:
Efficient computation of prolate spheroidal wave functions in radio astronomical source modeling
P. Noorishad, and S. Yatawatta
11^{th} *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*

5.1 Introduction

Radio interferometers measure the Fourier components of the brightness distribution in the sky on an irregularly filled plane using Earth rotation synthesis technique. Aperture synthesis imaging would reduce to a simple Fourier inversion, if we had the Fourier components sampled on a regular filled grid. Levanda and Leshem (2010) have demonstrated how this sub-Nyquist sampling results in aliasing inside the image after a Fourier inversion and limits the image dynamic range. Aliasing has also been physically interpreted as an effect caused by the high sidelobes of the synthesis array response (also known as the PSF). We remove the effect of the PSF from the true intensity in a process called deconvolution. The deconvolution in interferometric data analysis is basically an ill-posed inversion problem, which can be formulated as follows. We assume that $\widetilde{f}(u_p, v_p)$ is the sampled visibility at the p-th point in the visibility plane (Fig. 5.1, a) and $f(l_q, m_q)$ is the intensity of the q-th pixel in an image (Fig. 5.1, b). The region of interest (ROI) or signal support is the area of a source in an image.

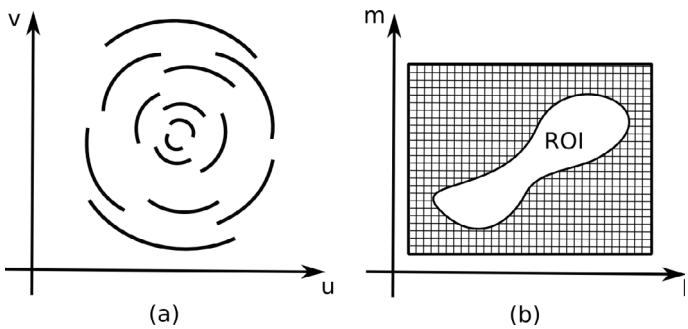


Figure 5.1: (a) Sampling points (total N_a) in the Fourier (visibility) plane. (b) Image of N rectangular pixels, with the support area (or ROI) in white. The ROI has N_b pixels.

With co-planar approximation, the quantities $\widetilde{f}(u_p, v_p)$ and $f(l_q, m_q)$ are related by the van Cittert-Zernike theorem (Brouw 1975) as

$$\widetilde{f}(u_p, v_p) = \sum_{q=0}^{N-1} f(l_q, m_q) e^{-j2\pi(l_q u_p + m_q v_p)}, \tag{5.1}$$

$$f(l_q, m_q) = \sum_{p=0}^{N_a-1} \tilde{f}(u_p, v_p) e^{j2\pi(l_q u_p + m_q v_p)},$$

$$p \in [0, N_a - 1],\ q \in [0, N],$$

where N_a is the total number of sampling points in visibility domain. N is the total number of pixels in the image. Moreover, it is only after the calibration that we can assume that the visibilities are the Fourier components of the true sky image (Eq.

5.1). For simplicity in formulating the deconvolution problem, we assume that we have calibrated visibilities, $\widetilde{f}(u_p, v_p)$ throughout this chapter.

As it was discussed in Chapter 2 (Sect. 2.1.2), the sky background noise in the image plane experiences the same physical and geometrical effects, as the source signal received by the interferometer. As a result, the sky background noise is convolved with the PSF as well. Since it is prominent especially at low-frequencies, a more accurate representation of the relation in Eq. 5.1 is

$$\widetilde{f}(u_p, v_p) = \sum_{q=0}^{N-1} (f(l_q, m_q) + n_q) e^{-j2\pi(l_q u_p + m_q v_p)}. \tag{5.2}$$

where n_q represents the sky background noise added to the image value $f(l_q, m_q)$. If we collect all the visibility points in the vector $\widetilde{\mathbf{f}}$ [1] and all image points (pixels) values in the vector $\mathbf{f}$, we can rewrite Eq. 5.2 in form of matrix multiplication as

$$\begin{aligned} &\widetilde{\mathbf{f}} = \mathbf{T}(\mathbf{f} + \mathbf{n}), \\ &\widetilde{\mathbf{f}} \equiv [\tilde{f}(u_0, v_0), ..., \widetilde{f}(u_{N_a-1}, v_{N_a-1})]^T, \\ &\mathbf{f} \equiv [f(l_0, m_0), ..., f(l_{N-1}, m_{N-1})]^T, \\ &\mathbf{n} \equiv [n_0, ..., n_{N-1}]^T. \end{aligned} \tag{5.3}$$

The matrix $\mathbf{T}$ (size $N_a \times N$) contains $e^{-j2\pi(l_q u_p + m_q v_p)}$ on its q-th row and p-th column. $\mathbf{n}$ is the sky noise vector with the same length as $\mathbf{f}$. For the deconvolution problem, we are interested in optimally extracting information of $\mathbf{f}$ given the observed visibilities, $\widetilde{\mathbf{f}}$ and the PSF. If we multiply both sides of Eq. 5.3 with the term $\mathbf{T}^H$, we obtain

$$\begin{aligned} &\mathbf{T}^H \widetilde{\mathbf{f}} = \mathbf{T}^H \mathbf{T}(\mathbf{f} + \mathbf{n}), \\ &\mathbf{f} = (\mathbf{T}^H \mathbf{T})^{-1} (\mathbf{T}^H \widetilde{\mathbf{f}} - \mathbf{T}^H \mathbf{T} \mathbf{n}). \end{aligned} \tag{5.4}$$

In usual terminology, $\mathbf{T}^H \widetilde{\mathbf{f}}$ is the dirty image, $\mathbf{T}^H \mathbf{T}$ is the dirty beam or the PSF, and $\mathbf{T}^H \mathbf{T} \mathbf{f}$ represents the convolution of the true image with the PSF. Eq. 5.4 gives an approximate solution, when no exact solution exists, provided that $(\mathbf{T}^H \mathbf{T})^{-1}$ exists ($\mathbf{T}$ has full row rank). Eq. 5.4 is an ill-posed problem that performs well, when the first term dominates (Briggs 1995). In the case of weak extended (large scale) emission, the effect of the second term (noise) is significant. It is simple to see that without explicitly incoporating a priori information, which fundamentally separates the two terms, such features in $\mathbf{f}$ cannot be recovered.

[1] Note that the visibility points can be arranged in any order in vector $\widetilde{\mathbf{f}}$. But a suitable vector $\widetilde{\mathbf{f}}$ is a vector made by stacking the columns of the calibrated visibility matrix, $\mathbf{R}$ i.e. $\widetilde{\mathbf{f}} = \text{vec}(\hat{\mathbf{R}} - \mathbf{\Sigma}_{rec})$. This arrangement simplifies the comparison of different deconvolution methods, in particular least squares imaging in Eq. 5.28. with the general formulation of the deconvolution problem in Eq. 5.4. See Appendix. C.

We note that the optimality of image reconstruction performed by the inversion in Eq. 5.4, fundamentally defines the *image fidelity*.

The deconvolution problem has been approached in several different ways, since the inception of radio interferometry. CLEAN (Högbom 1974) is the first and most commonly used deconvolution method. It has greatly served aperture synthesis imaging over the last decades. Its performance, however degrades during the deconvolution of extended sources, owing to its pixel-based source modelling scheme and the nominal resolution (or the PSF width) limit of the interferometer. Fundamental limitations of CLEAN for deconvolution of an extended source were first discussed by Brinks and Shane (1984). However, they had not recognized its pixel-based source modelling scheme as the cause of that limitation. As an improvement to CLEAN performance, Wakker and Schwarz (1988) developed a multi-resolution CLEAN (MRC) by appropiately combining the results of a number of conventional CLEAN operations with optimized parameters, each at a different scale. Later Cornwell (2004); Bhatnagar and Cornwell (2004) recommended a scale sensitive pixel-based source modelling scheme by representing the ROI with a number of properly scaled and centered extended basis function components. However, one can argue that a scale sensitive pixel-based source modelling would still be limited by the resolution limit of the interferometer, when the source extension is larger than the PSF width. Moreover, they acknowledged that they cannot reliably specify the number of significant basis function components. The effect of pixelization on the dynamic range has also been studied numerically by Cotton and Uson (2008), and analytically and statistically by Yatawatta (2010).

For improving the source modelling scheme in the maximum entropy method (MEM), similar effort was made by Pantin and Starck (1996); Maisinger et al. (2004), who recommended decomposing the extended source to the non-orthonormal, redundant discrete à trous wavelets with multiple scales (Pantin and Starck 1996) and with the potential of spanning the whole source area (the ROI) (Maisinger et al. 2004).

Chang and Refregier (2002); Refregier (2003); Berry and Withington (2004) demonstrated the application of shapelets (a set of orthonormal basis functions) for astronomical image analysis. Yatawatta (2010) analytically proved that source modelling based on the use of 2-D orthonormal basis functions that span the whole ROI e.g. shapelets achieves the theoretical noise bound or CRB (Cramér-Rao Bound), regardless of the source structure. He emphasizes the use of such basis functions as opposed to decomposing the image to the pixels with either the same size or scalable sizes, as had been proposed by Cornwell (2004); Bhatnagar and Cornwell (2004). This way, one can overcome the resolution limit of the interferometer for the sources that extend larger than the PSF width. Later, he demonstrated the superiority of the prolate spheroidal wave functions (PSWF) for source modelling as compared with shapelets (Yatawatta 2011a,b), with minimum number of significant basis functions.

Since achieving the CRB means an estimation with minimum error, we conclude that the use of orthonormal basis functions provides an optimum source modelling scheme for inverse problem in Eq. 5.4 or improves the *image fidelity*, from which most of the deconvolution methods can benefit.

In Sect. 5.2, we explain further the underlying reasons for seeking a more optimum source modelling scheme for the extended sources as opposed to the pixel-based source

modelling schemes. We discuss three suitable candidates for the orthonormal basis functions i.e. shapelets, PSWFs, and Franks basis functions in the corresponding subsections. A good overview of most deconvolution techniques and their source modelling scheme is given by Levanda and Leshem (2010). In Sect. 5.3, we extend it from the point of view of their capability in removal of the extended sources. In Sect. 5.3.6, we separate them into two groups of locally and globally optimum solutions, among which the second group can accommodate the use of orthonormal basis for their source modelling scheme. Following the promising results of using orthonormal basis functions, especially the PSWFs (Yatawatta 2010, 2011b,a), we suggest a solution to reduce their computational cost in Sect. 5.4.

5.2 Orthonormal basis functions in source modelling

There are several ways to represent an arbitrary signal $\mathbf{x}(l,m)$ in an arbitrary domain to identify its information content. Wavelets are special sets of basis functions, that allow the efficient representation of a signal. This representation is done by the linear decomposition of the signal into series of localized basis functions such as $\mathbf{p}_i(l,m)$ with different shapes and weights (coefficients), as it is given in

$$\mathbf{x}(l,m) = \sum_{i}^{M} c_i \mathbf{p}_i(l,m), \tag{5.5}$$

In imaging applications, these basis functions can potentially span the whole ROI. Then, Eq.5.5 can be rewritten in a matrix form

$$\mathbf{x} = \mathbf{P}\mathbf{c}, \tag{5.6}$$

where the matrix $\mathbf{P}$ contains M basis functions in its columns and vector $\mathbf{c}$ contains M model coefficients. This representation is as opposed to the pixel-based representation of an image, that models the ROI by associating a weighted delta function, $\delta(.)$ to each point (pixels within the ROI), as it is given in

$$\mathbf{x}(l,m) = \sum_{i=1}^{N_b} x_i \delta(l-l_i, m-m_i). \tag{5.7}$$

In the deconvolution problem in Eq. 5.4, the pixels of the dirty image are coupled due two reasons. First, the pixels of the underlying true image are inherently not independent in the presence of extended emission. The second reason of coupling is the PSF. The main lobe of the PSF has a finite width (proportional to the diffraction limit of the synthesized aperture) and significant wide-spread sidelobes. Ignoring the coupling due to the PSF possibly only results into slower convergence. However, ignoring the inherent coupling of pixels for extended emission leaves the residuals, that are correlated at scales larger than the resolution element (Bhatnagar and Cornwell 2004). Therefore, a source modelling scheme, that fundamentally separates noise from signal is required. From the analytical and statistical studies by Yatawatta

(2010, 2011b,a), we have learned that such an scheme can be optimally implemented using the orthonormal basis functions that span the whole ROI.

Earlier in Chang and Refregier (2002); Refregier (2003); Berry and Withington (2004), it was proven that shapelets provide a way to model extended sources by a closed form and linear construction of the orthonormal basis. But this technique requires additional constraints such as information theoretic bounds to limit the model order (or number of significant basis functions) (Yatawatta 2010). Yatawatta (2011b,a) has shown that the use of PSWF provides an optimal source model for an extended source, with minimum number of basis functions and less artifacts outside the ROI. Unlike shapelets, the PSWF do not require additional explicit constraints from information theory to limit the model order, as it is implicitly considered in their mathematical derivation, see (Slepian 1961; Yatawatta 2011a). But PSWF do not have closed-form expressions and are derived by solving an eigen-function equation (represented by Eq. 5.13). This makes their construction computationally more expensive than shapelets (Yatawatta 2011b). In Sect. 5.4, we propose a solution for computing PSWF in an efficient way.

As an alternative to the PSWF, Wei and Lamahewa (2010) have derived a new set of orthonormal basis functions using the general constrained variational method of Franks (Franks 1969, Ch. 6), which generalizes the PSWF derivation. Their obtained set of orthonormal basis functions exhibits results comparable with the PSWF, but with a flatter main lobe (that may make their use more favorable in imaging). In Sect. 5.2.3, we derive an expression for computing the Franks functions applicable for discrete and 2-D signals i.e. aperture synthesis imaging.

In the following subsections, we give a brief explanation for each set of aforementioned orthonormal basis functions for deconvolution in aperture synthesis imaging, such that we understand their differences and capabilities to handle an extended source modelling.

5.2.1 Shapelets

In image plane, the shapelet basis function $p_n(l, m, \beta)$ is given by

$$p_n(l,m,\beta) = \frac{H_{n_1}(\beta^{-1}l)H_{n_2}(\beta^{-1}m)e^{-\frac{(l^2+m^2)}{2\beta^2}}}{(2^{(n_1+n_2)}\pi\beta^2 n_1!n_2!)^{\frac{1}{2}}}, \tag{5.8}$$

where β is the model scale, and the integers n_1 and n_2 define the model order. H_{n1} and H_{n2} are the Cartesian Hermite polynomials that make the shapelet basis functions invariant within a scale factor, under Fourier transformation i.e. $p_n(u, v, \beta) = p_n(l, m, 1/\pi\beta)$. Thus, the Fourier transform of a basis function is a scaled version of itself. Then, deconvolution can be done both in the image-plane (Yatawatta 2011b) and in the uv-plane (Chang and Refregier 2002) with the same level of complexity and accuracy. In either case, the deconvolution reduces to some linear operations. For instance, in the image-plane implementation, we estimate the model coefficients in vector, $\mathbf{m}$ (size $N_D \times 1$) by a least-squares fitting in

$$\mathbf{f} = \mathbf{P}\mathbf{m},\ \hat{\mathbf{m}} = \mathbf{P}^{\dagger}\mathbf{f}, \tag{5.9}$$

where, each column of $\mathbf{P}$ (size $N \times N_D$) is one basis function in Eq. 5.9 evaluated at the image grid point and $\mathbf{f}$ (size $N \times 1$) is the vectorized value of the source image. Then, we find the equivalent basis function in the Fourier domain: $\widetilde{\mathbf{P}} = \mathbf{T}^H\mathbf{P}$ and subtract the model from the observed data, $\mathbf{z}$

$$\mathbf{r} = \mathbf{z} - \widetilde{\mathbf{P}}\hat{\mathbf{m}}. \tag{5.10}$$

We continue until the residual $\mathbf{r}$ is noise-like i.e. the known source has properly been removed.

N_D is the number of significant basis functions (or model order)[2], which is not known using shapelets for the source modelling. Yatawatta (2010, 2011b) demonstrated that N_D has to be set by the Landau-Pollak theorem (Slepian 1976) using shapelets. In its simplest form, this theorem states that the number of degrees of freedom to model any given source with compact support (the ROI), N_D cannot be greater than the product of the area of the source in the image domain, A_{lm} and the area of the source observed in the Fourier plane, A_{uv} i.e. $N_D \leq A_{lm} \times A_{uv}$.

Yatawatta (2011b) showed that the application of the PSWF for an image representation leads to a better concentration of the signal energy (as a measure of information content within the ROI). He ascribed this to the fact that the Cartesian basis function performs less efficient in representing an arbitrary shape in an image as compared with the spheroidal basis functions. He also showed that PSWF does this with a minimum number of significant basis functions, without explicit use of any information theoric bounds.

5.2.2 Prolate spheroidal wave functions

According to the theory of Fourier transforms, a signal cannot be both time-limited and band-limited simultaneously. However, in many real life applications, we seek signal models that have the greatest energy concentration both in time and the frequency domains (Landau and Pollak 1961; Slepian 1976). This perspective led to the derivation of PSWF (Slepian 1961), from which the astronomical signals (like any time-limited, band-limited signal) can greatly benefit.

It is worth noting that in radio interferometric data, the time and frequency domains translate to image and visibility domains, respectively. Thus, the time-limited term refers to the source's limited boundary (the ROI, or the signal support) in an image and the band-limited term refers to the limited area of the measured visibilities by the interferometer, basically determined by the longest baseline.

For a complete mathematical derivation of the PSWF for a 2-D discrete signal modelling, the reader is referred to Lindquist and Yang (2006); Yatawatta (2011a),

[2]also known as the Shannon number (Percival and Walden 1993; Simons et al. 2006)

and references therein.[3] We will mention a few key mathematical relations of PSWF derivation for studying the problem for its computational cost, as follows: Referring to Fig. 5.1- b and Eq. 5.5, we aim to obtain a representative basis function, $\mathbf{p}(l,m)$ for a source that maximizes the energy in the ROI i.e. the larger the ratio given in 5.11, the less artifacts remain outside the ROI.

$$\lambda = \frac{\sum_{(l,m)\in ROI} |\mathbf{p}(l,m)|^2}{\sum_{(l,m)} |\mathbf{p}(l,m)|^2}, \tag{5.11}$$

the vectorized form of which can be written as

$$\lambda = \frac{\|\mathbf{I}_b^T \mathbf{p}\|^2}{\|\mathbf{p}\|^2}, \tag{5.12}$$

where $\mathbf{I}_b^T$, (size $N_b \times N$) is a selection matrix to select the pixels that belong to the ROI. N_b is the number of pixels in the ROI. N is the total number of pixels in the image. After several mathematical operations, we obtain an eigen-function equation (see Lindquist and Yang (2006))

$$\mathbf{I}_b^T \mathbf{T}^H (\mathbf{T}\mathbf{T}^H)^\dagger \mathbf{T}\mathbf{I}_b \boldsymbol{\eta} = \lambda \boldsymbol{\eta}. \tag{5.13}$$

Then, the PSWF basis can be obtained by eigenvalue-decomposition of the kernel, $\mathbf{K}$ (size $N_b \times N_b$) given as

$$\mathbf{K} = \mathbf{I}_b^T \mathbf{T}^H (\mathbf{T}\mathbf{T}^H)^\dagger \mathbf{T}\mathbf{I}_b, \tag{5.14}$$

$(\lambda_i, \boldsymbol{\eta}_i)$ is the i-th eigenvalue and eigen vector pair of $\mathbf{K}$, that gives the i-th basis vector as

$$\mathbf{p}_i = \frac{1}{\lambda_i} \mathbf{T}^H (\mathbf{T}\mathbf{T}^H)^\dagger \mathbf{T}\mathbf{I}_b \boldsymbol{\eta}_i. \tag{5.15}$$

We stack the N_D significant basis vectors (e.g. the ones with largest eigenvalues) in $\mathbf{P} = [\mathbf{p}_0, \mathbf{p}_1, ..., \mathbf{p}_{N_D-1}]$. Matrix $\mathbf{P}$ can be used in the deconvolution problem just as the shapelets in the computations in Eq. 5.9 and Eq. 5.10.

In Eq. 5.14, most of the computations are required first to find the pseudo-inverse term, $(\mathbf{T}\mathbf{T}^H)^\dagger$ (size $N_a \times N_a$) and secondly, to find the eigenvalue decomposition of $\mathbf{K}$ (size $N_b \times N_b$). Note that the rank of $(\mathbf{T}\mathbf{T}^H)^\dagger$ is at most N and usually $N_a \gg N > N_b$.[4] By properly down-sampling both the Fourier plane sampling points (reducing N_a) as well as the image pixels (reducing N and N_b) we can significantly cut down the cost of computations. In Sect. 5.4, we explain how this can be done efficiently, followed by an example.

[3] Lindquist and Yang (2006) studied the application of the PSWF for Magnetic Resonance Imaging (MRI), where they also have sparse non-uniform Fourier samples, similar to radio interferometric data.

[4] Be reminded that N_a, N_b, and N are number of visibility points, number of pixels in the ROI and total number of image pixels, respectively.

5.2.3 Generalized 2-D Franks basis functions

Franks (1969, Ch. 6), generalizes the PSWF derivation by solving a general constrained variational problem, in which the signal is weighted in both time and frequency domains. It is stated as follows. For a given signal with finite energy ($x(t)$), extermize (minimize or maximize) the arbitrary weighted energy either in the time-domain or in the frequency domain, to obtain the optimal waveform signal. Suppose $w(t) \geq 0$ is an arbitrary weighting function in the time domain, $V(f) \geq 0$ is an arbitrary weighting function in the frequency domain. Following Franks (1969, Ch. 6), this general variational problem is equivalent to extermizing

$$G = \mu_1 I_1 + \mu_2 I_2 + I_3, \tag{5.16}$$

where μ_1 and μ_2 are two Lagrange multipliers and I_1, I_2, and I_3 are three quadratic functionals defined as

$$\begin{aligned} I_1 &\equiv < wx, x > = \int_{-\infty}^{+\infty} w(t)|x(t)|^2 dt, \\ I_2 &\equiv < VX, X > = \int_{-\infty}^{+\infty} V(f)|X(f)|^2 df, \\ I_3 &\equiv < x, x > = < X, X > = 1, \end{aligned}$$

where x and X are the signal of interest and its Fourier transformed, respectively. I_1 and I_2 represent the energy concentration of the signal in time and frequency domains, respectively. I_3 is the signal energy content. Wei and Lamahewa (2010) have solved the problem stated in Eq. 5.16 and obtained a set of orthonormal basis functions by setting the signal weights as $w(t) = t^4$ (or the 4th order moment), and $V(f) = 1$ within the limited frequency band and $V(f) = 0$ outside the frequency band. Their obtained basis functions exhibit a flatter behavior in the main lobe as compared with the PSWF, which may make them a better representative of the source area in an image. Here, we derive the analytical expressions to facilitate its application for the deconvolution problem in radio interferometric imaging (2-D discrete signals). In such a case the terms I_1, I_2, and I_3 are

$$\begin{aligned} I_1 &= \iint w(l,m)|f(l,m)|^2 dldm, \\ I_2 &= \iint V(u,v)|\tilde{f}(u,v)|^2 dudv, \\ I_3 &= \iint |f(l,m)|^2 dldm = \iint \tilde{f}(u,v)|^2 dudv = 1. \end{aligned}$$

Earlier in Eq. 5.3, we represented the image and visibility values in vectors $\mathbf{f} = \mathbf{f}(l,m)$ and $\tilde{\mathbf{f}} = \tilde{\mathbf{f}}(u,v)$, respectively i.e. $\mathbf{f}$ and $\tilde{\mathbf{f}}$ represent 2D functions, the values of which are collected in the corresponding vectors. This allows us to simplify the I_1, I_2, and I_3 expressions for a 2D function in

$$
\begin{aligned}
I_1 &=< \mathbf{W}\mathbf{f}, \mathbf{f} >=\| \mathbf{W}^{\frac{1}{2}}\mathbf{f} \|^2, \\
I_2 &=< \mathbf{V}\tilde{\mathbf{f}}, \tilde{\mathbf{f}} >=\| \mathbf{V}^{\frac{1}{2}}\tilde{\mathbf{f}} \|^2, \\
I_3 &=< \mathbf{f}, \mathbf{f} >=\| \mathbf{f} \|^2 = 1,
\end{aligned}
$$

where $\mathbf{W}$ (size $N \times N$), $\mathbf{V}$ (size $N_a \times N_a$) are diagonal weighting matrices in image and visibility domains, respectively. Like in PSWF derivation, here the objective is to find basis functions (and their conterparts in the visibility domain) that maximize the energy in the ROI. $\mathbf{I}_b^T\mathbf{p}$ represents the selection of the pixels that belong to the ROI from the full image vector $\mathbf{p}$. Using the relation $\tilde{\mathbf{p}} = \mathbf{T}\mathbf{p}$ in Eq. 5.4, I_1, I_2, and I_3 expressions in the ROI are

$$
\begin{aligned}
I_1 &=\| \mathbf{W}^{\frac{1}{2}}\mathbf{I}_b^T\mathbf{p} \|^2=\| \mathbf{W}^{\frac{1}{2}}\mathbf{I}_b^T\mathbf{T}^H\tilde{\mathbf{p}} \|^2= \mathrm{tr}(\tilde{\mathbf{p}}^H\mathbf{T}\mathbf{I}_b\mathbf{W}\mathbf{I}_b^T\mathbf{T}^H\tilde{\mathbf{p}}), \\
I_2 &=\| \mathbf{V}^{\frac{1}{2}}\tilde{\mathbf{p}} \|^2= \mathrm{tr}(\tilde{\mathbf{p}}^H\mathbf{V}\tilde{\mathbf{p}}), \\
I_3 &=\| \mathbf{I}_b^T\mathbf{p} \|^2=\| \mathbf{I}_b^T\mathbf{T}^H\tilde{\mathbf{p}} \|^2 .
\end{aligned}
$$

Note that $\mathbf{W}$, $\mathbf{V}$, and $\mathbf{f}$ are real, thus $(.)^H$ is replaced by $(.)^T$ in the derivations. We define $\boldsymbol{\xi} \triangleq \boldsymbol{\Psi}\tilde{\mathbf{p}}$, where $\boldsymbol{\Psi} \triangleq (\mathbf{T}\mathbf{T}^H)^{\frac{1}{2}}$ and rewrite the I_1, I_2, and I_3 terms

$$
\begin{aligned}
I_1 &= \mathrm{tr}(\boldsymbol{\xi}^H\boldsymbol{\Psi}^{-H}\mathbf{T}\mathbf{I}_b\mathbf{W}\mathbf{I}_b^T\mathbf{T}^H\boldsymbol{\Psi}^{-1}\boldsymbol{\xi}), \\
I_2 &= \mathrm{tr}(\boldsymbol{\xi}^H\boldsymbol{\Psi}^{-H}\mathbf{V}\boldsymbol{\Psi}^{-1}\boldsymbol{\xi}), \\
I_3 &= \mathrm{tr}(\boldsymbol{\xi}^H\boldsymbol{\Psi}^{-H}\mathbf{T}\mathbf{I}_b\mathbf{I}_b^T\mathbf{T}^H\boldsymbol{\Psi}^{-1}\boldsymbol{\xi}).
\end{aligned}
$$

Then, by substituting the terms in Eq. 5.16, the optimization problem becomes

$$
\boldsymbol{\xi} = \underset{\boldsymbol{\xi}}{\mathrm{argmax}}\ \mathrm{tr}(\boldsymbol{\xi}^H \underbrace{\boldsymbol{\Psi}^{-H}\left(\mathbf{T}\mathbf{I}_b(\mu_1\mathbf{W}+\mathbf{I})\mathbf{I}_b^T\mathbf{T}^H + \mu_2\mathbf{V}\right)\boldsymbol{\Psi}^{-1}}_{\mathbf{K}}\boldsymbol{\xi}), \qquad (5.17)
$$

$$
\text{subject to: } \| \boldsymbol{\xi} \| = 1,
$$

The solution to Eq. 5.17 is the largest eigenvalue, eigenvector pair of the matrix $\mathbf{K} = \boldsymbol{\Psi}^{-H}\left(\mathbf{T}\mathbf{I}_b(\mu_1\mathbf{W}+\mathbf{I})\mathbf{I}_b^T\mathbf{T}^H + \mu_2\mathbf{V}\right)\boldsymbol{\Psi}^{-1}$.

Eq. 5.14 and Eq. 5.17 show that the derivation of the PSWF and Franks basis functions are independent of the signal itself. The shape of the ROI (represented by $\mathbf{I}_b$) and the PSF determine those basis functions. Studying the effect of the array configuration (translated to the PSF) on the number of significant basis functions and their performance in source modelling might be an interesting subject. It may give us a guideline, based on which we can choose the best array configuration (or certain visibility points) to achieve the best image fidelity possible.

5.3 Deconvolution techniques

The deconvolution problem in radio astronomy has been approached in various ways. In this section, we briefly introduce the existing deconvolution methods. This gives us a suitable and unified background to understand and compare their potential and limitations in handling extended sources. In Sect. 5.3.6, we also discuss how they can accommodate the use of orthonormal basis functions in their source modelling scheme to achieve a high image fidelity.

5.3.1 CLEAN and its variants

CLEAN (Högbom 1974) is the earliest method invented for the radio astronomical deconvolution problem. CLEAN and its variants use a pixel-based source modelling scheme with either similar-scale (Högbom 1974; Clark 1980) or multi-scale pixel size (Cornwell 2004; Bhatnagar and Cornwell 2004).

Högbom (1974) and Clark (1980) assume a collection of point sources as a model for an extended source by associating a weighted delta function to each pixel,

$$\mathbf{f}^M = \sum_i^N f_i \delta(\mathbf{r}_i) = \sum_i^N f_i \delta(l - l_i, m - m_i), \tag{5.18}$$

and solves for the strengths, f_i and positions, $\mathbf{r}_i = (l_i, m_i)$ iteratively in a least squares fit. The algorithm is sequential and in each sequence, the pick in the residual image is modelled, fitted to the data, and subtracted.

At first order, the convergence becomes slow in deconvolution of an extended source (as might be expected since one is trying to identify a potentially large number of pixels one by one), and instability occurs (Cornwell 1983). To overcome this limitation, multi-scale CLEAN (Cornwell 2004) and adaptive scale pixel (ASP)-CLEAN (Bhatnagar and Cornwell 2004) were invented, that basically model the extended structures with appropriately scaled and centered extended components, $m(\mathbf{r}_i, \alpha_i)$

$$\mathbf{f}^M = \sum_i w_i m(\mathbf{r}_i, \alpha_i) = \sum_i w_i m(\delta(\mathbf{r}_i), \alpha_i), \tag{5.19}$$

where α_i is the component scale and w_i is an associated weight. The component shape is chosen to be

$$m(\mathbf{r}, \alpha) = \Psi(\mathbf{r}) \left(1 - \left(\frac{|\mathbf{r}|}{\alpha}\right)^2\right), \tag{5.20}$$

where $\Psi(\mathbf{r})$ is an approximated prolate spheroidal wave function (originally proposed by Brouw (1975)) or a suitably scaled Gaussian function at $\mathbf{r} = (l, m)$. Both approaches seek a locally best fit to the observed data for each pixel model. Mutli-scale CLEAN algorithm works simultaneously on all scales, that are pre-computed, instead of computing them sequentially. The pixel scales are arbitrary chosen with a bias towards small scales to enable the algorithm to model the asymmetric source structures. In contrast, ASP-CLEAN estimates the best fit ASP (weight, scale, location)

at the location of the peak in the residual image in each iteration. One major problem is that the significant number of basis functions or degrees of freedom (DOF) is still unknown in these methods.

CLEAN in all its forms is a greedy algorithm i.e. it takes a heuristic approach to make the locally optimal choice at each stage hoping that it achieves a global optimum.

Fundamental limitation of the pixelization in modelling the extended sources is when the centroid of point sources does not exactly match pixel coordinates on a regular grid, and the extended sources require more than a few clean components. The effect of pixelization and choice of grid on the dynamic range in imaging has numerically been studied in Cotton and Uson (2008). Yatawatta (2010) has analytically proven that CLEAN with a pixel-based source modelling scheme (Eq. 5.18) fails to perform satisfactorily, when two point sources are too close (closer than the resolution limit) or if the source is partially resolved. He has also shown that making the image grid smaller (expecting that accuracy of modelling the extended sources improves), is a futile effort by presenting the CRB analysis. In general, using any pixel-based algorithm would still be limited by resolution limit of the interferometer in case of a barely resolved sources (Yatawatta 2010). It is also known that in such cases, the residuals are correlated with the source structure (Bhatnagar and Cornwell 2004).

5.3.2 Maximum entropy method

MEM was introduced to radio astronomy by Gull and Daniell (1978), and optimized and implemented in AIPS by Cornwell and Evans (1985). For radio astronomical image deconvolution, the MEM solution is given by solving the following Lagrangian optimization problem (or maxmizing J w.r.t. image values, $f(l,m)$)

$$J = H - \alpha.\chi^2 - \beta.F, \tag{5.21}$$

where α and β are the Lagrange multipliers. H is a relative entropy used as a measure of resemblance between the prior image, $\mathbf{b}$[5] and a constructed image using the observed visibilities, $\mathbf{f}$. It is defined as

$$H(\mathbf{f}|\mathbf{b}) = -\sum_{i=1}^{N} f_i \ln(f_i/b_i). \tag{5.22}$$

In Eq. 5.21, χ^2 is the goodness of the fit in the corresponding visibility domain for $\mathbf{b}$, that is obtained by

$$\chi^2 = \sum_{i=1}^{N_a} w_i |\tilde{b}(u_i,v_i) - \tilde{f}(u_i,v_i)|^2, \tag{5.23}$$

[5]It is usually a low resolution version of the observed source.

where $\tilde{\mathbf{b}} = \mathbf{T}\mathbf{b}$. F is total energy[6] in the MEM image that is to be the total energy in the observed image and it also implies the image positivity constraint.

$$F = \sum_{i=1}^{N} b_i = \sum_{i=1}^{N} f_i \tag{5.24}$$

The solution to Eq. 5.21 is discussed by Cornwell and Evans (1985). Although Eq. 5.21 seeks a globally optimum solution, its quality and speed of convergence depends on the quality of the prior image, $\mathbf{b}$ and how $\mathbf{f}$ is modelled in its reconstruction process using the measured visibilities. $\mathbf{f}$ can be represented by pixels (using CLEAN and its variants), by non-orthogonal, redundant discrete à trous wavelets (Pantin and Starck 1996; Maisinger et al. 2004) or by orthonormal basis functions e.g. shapelets or PSWF (Yatawatta 2011a). The latter remains untested for MEM.

Pantin and Starck (1996); Maisinger et al. (2004) recommended the use of non-orthonormal basis functions as opposed to the orthonormal basis functions, referring to Lang et al. (1996) and Langer et al. (1993), who claim a better denoised results using non-orthonormal basis functions. On the contrary, the analytical argument in combination with demonstrations on real radio astronomical data have proven that the orthonormal basis functions are the optimal choice (Yatawatta 2011a). Future work may compare these two sets of basis functions for MEM performance.

5.3.3 Least squares minimum variance imaging

Least squares minimum variance imaging (LS-MVI) is a matrix based sequential source removal algorithm. It was inspired by the matrix based techniques for direction of arrival (DOA) estimation. It is formulated as a constrained beamforming problem such that we minimize interference from point sources coming from sidelobes while observing a point source in direction of $\boldsymbol{\ell} = (l, m)$

$$\begin{aligned} \hat{\mathbf{w}}(\boldsymbol{\ell}) = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{k=1}^{q-1} \mathbf{w}_k^H(\boldsymbol{\ell}) \hat{\mathbf{R}}_k \mathbf{w}_k(\boldsymbol{\ell}) \\ \text{subject to:} \quad \sum_{k=1}^{q-1} \mathbf{w}_k^H(\boldsymbol{\ell}) \mathbf{a}_k(\boldsymbol{\ell}) = 1, \end{aligned} \tag{5.25}$$

where $\mathbf{w}_k = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K]^T$ is the beamforming vector, $\mathbf{a}_k$ is the array response vector toward the source in direction $\boldsymbol{\ell} = (l, m)$. q is the number of point sources. The intensity estimation is a LS estimation in

$$\begin{aligned} f_i = \underset{f_i}{\operatorname{argmin}} \sum_k \|\hat{\mathbf{R}} - f_i \mathbf{a}_k(l,m) \mathbf{a}_k^H(l,m)\|_F^2, \\ \text{subject to:} \ f_i \geq 0, \end{aligned} \tag{5.26}$$

[6]Note that here, the measure of energy is the summation of the pixel values, given the positivity constraint, whereas in Eq. 5.11, it is the square root of the summation of the squared absolute values of the pixels. The latter is a more general definition without the positivity constraint.

that is basically the same fitting process in CLEAN. Estimation is done for each point recognized as a pick in the residual, f_i in a closed form solution. For further analytical details, the reader is referred to Leshem and van der Veen (2000); Ben-David and Leshem (2008); Levanda and Leshem (2010). The sequential nature of LS-MVI technique makes it a greedy algorithm like CLEAN. Thus, it is not globally optimum. Despite its more sophisticated model for dirty image than CLEAN and a closed-form intensity estimation solution, this method suffers from the same limitations as CLEAN for deconvolving an extended source. This is because an extended source has to be modelled as a superposition of point sources, which have to be tackled one by one. Note that Eq. 5.26 can accommodate the flux estimation for a vector e.g. $\mathbf{f}$ containing all image pixels instead of the pixels values (scalar f_i), sequentially. Least squares imaging (LSI) well-studied by Wijnholds (2010), accommodates that property. This means it can estimate for the source fluxes in one step by seeking a global optimum.

5.3.4 Least squares imaging

Like LS-MVI technique, LSI is a matrix based technique. It was inspired by the model-based calibration algorithm (Wijnholds 2010) i.e. the image parameters (intensities and positions) can be estimated, just as the calibration parameters in the process of fitting the observed data to its model. It seeks a global optimum by solving a least squares covariance fitting, given in

$$\widehat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \|\mathbf{W}_c \left(\widehat{\mathbf{R}} - \mathbf{\Sigma}_{\mathrm{n}}\right) \mathbf{W}_c - \mathbf{W}_c \mathbf{G}\mathbf{A}\mathbf{\Sigma}\mathbf{A}^H\mathbf{G}^H\mathbf{W}_c\|_F^2, \tag{5.27}$$

where $\widehat{\mathbf{R}}$ is the measured visibility matrix. $\mathbf{G}\mathbf{A}\mathbf{\Sigma}\mathbf{A}^H\mathbf{G}^H$ is the array data model. $\mathbf{W}_c$ is a weighting matrix, and $\mathbf{\Sigma}_{\mathrm{n}}$ is the noise correlation matrix. As it was discussed in Chapter 2, $\mathbf{\Sigma}_{\mathrm{n}}$ carries the correlated sky noise on its off-diagonal and the receiver noise on its diagonal.

Eq. 5.27 should be equivalent of a maximum likelihood (ML) estimation problem, given the large number of signal samples. It should give a statistically efficient estimate of the image intensity,[7] with an optimal choice of the weighting, that is proven to be $\mathbf{W}_c = \mathbf{R}^{-1/2}$ (Wijnholds 2010). A closed form solution to the problem in Eq. 5.27 is

$$\widehat{\mathbf{f}} = (\overline{\mathbf{G}\mathbf{A}} \circ \mathbf{G}\mathbf{A})^{\dagger} \mathrm{vec}(\widehat{\mathbf{R}} - \mathbf{\Sigma}_n). \tag{5.28}$$

For imaging, it boils down to estimating the intensity on every image point (pixel) i.e. by assuming a data model with a source present at every pixel. Thus, it suffers from the pixelization limitation, that we discussed earlier in Sect.5.2 and Sect. 5.3.1, although it seeks a globally optimum solution.

Assuming that the array is calibrated i.e. $\mathbf{G}$ is known, Eq. 5.28 resembles Eq. 5.4. If we expand the array response vector, $\mathbf{A}$ in Eq. 5.28, we actually obtain exactly the

[7] The position parameters are usually estimated simultaneously with Eq. 5.27, which is not of interest in our argument here.

same relation, the optimal solution to which is in an orthonormal space (Yatawatta 2010). The reader is referred to Appendix. C for more clarity.

Unlike the LS-MVI technique, LS imaging is a non-sequential algorithm and seeks a global optimum. Nevertheless, it may converge after several iterations.

5.3.5 Compressed sensing

Radio astronomical signals as time-limited, band-limited sparse signals can be optimally reconstructed using compressed-sensing (CS) techniques (Rudelson and Vershynin 2006). In imaging applications, this is done by minimizing a ℓ_1 based cost function, $\|\mathbf{f}(l,m)\|_{\ell_1}$ in the image domain subject to the constraints of image positivity and the measured visibility data. It is fomulated as

$$\min \ \|\mathbf{f}(l,m)\|_{\ell_1} = \min \sum_{l=1}^{N}\sum_{m=1}^{N} f(l,m), \qquad (5.29)$$
$$subject\ to: \ \mathbf{f} = \mathbf{T}^{\dagger}\tilde{\mathbf{f}}.$$

Since image values, $f(l,m)$ are positive quantities, Eq. 5.29 can be solved by linear programming techniques. Several recovery algorithms such as fixed point continuation (FPC) (Hale et al. 2007) and total variation (TV) (Chambolle and Lions 1997) minimization are proposed specifically for solving an l_1-norm minimization problem (e.g. the one in Eq. 5.29). These algorithms are known to be robust but have different speeds of convergence.

As discussed by Candès and Romberg (2007), the CS technique relies on two principles: *sparsity*, which relates to the signal of interest, and *coherency*, which relates to the sensing modality (for our application, it is the *uv*-coverage expressed in matrix $\mathbf{T}$). A signal is called k-sparse if its representation in an orthonormal space $\mathbb{R}^m$ ($\mathbf{P} = [\mathbf{p}_0, \mathbf{p}_1, ..., \mathbf{p}_k, ..., \mathbf{p}_{m-1}]$) has at most k nonzero (or significant) entries. As discussed earlier in Sect. 5.2.1, N_D or the Shannon number defines the significant number of basis functions to optimally represent a signal. Thus, the signal should be N_D-sparse for the best image fidelity.

Coherency serves as a rough characterization of the degree of similarity between the *sparsity* and measurement systems. It is defined as

$$\mu(\mathbf{T},\mathbf{P}) = \sqrt{m}\ .\ \max_{k\geq 1, j\leq m} |\langle \mathbf{T}_k, \mathbf{P}_j \rangle\,|. \qquad (5.30)$$

For *coherency* to be close to its minimum value, each of the measurement vectors or rows of matrix $\mathbf{T}$ must be uncorrelated with the columns of $\mathbf{P}$. *Coherency* guarantees the accuracy of the minimization in Eq. 5.29. Donoho et al. (2006) have shown that accurate reconstruction occurs for k-sparse signal when $k < \frac{1}{2}(1+\frac{1}{\mu})$, where μ is the *coherency* defined in Eq. 5.30.

There is a growing interest in application of the CS technique in radio interferometric imaging owing to its linearity. Its application for radio interferometric image deconvolution was introduced and demonstrated by Levanda and Leshem (2008);

Wiaux et al. (2009a). It was enhanced further in research by McEwen and Wiaux (2011b,c,a), where its capability in wide-field imaging in the celestial sphere (without w-projection step) was demonstrated. But, they used a constant w-term for all the visibilities in their demonstrations, which is not an adequate representation of the real observations. Wiaux et al. (2009b,c) showed that the CS performance improves in (u, v, w)-space as compared with (u, v)-space, by taking advantage of a phenomenon known as spread spectrum, that increases the incoherency.

We note that regardless of the recovery algorithm, the constraint of the measured visibility data ($\mathbf{f} = \mathbf{T}^{\dagger}\tilde{\mathbf{f}}$) is assumed to be done in an orthonormal space of $\mathbb{R}^m$ for an optimal signal recovery (Candès and Romberg 2006, 2007). However, the basis functions are not specified. The PSWF explicitly provide the significant number of the basis functions, N_D (*sparsity* is known) while no other set of orthonormal basis functions does that. The CS performance using the PSWF or Franks basis functions remains unstudied.

5.3.6 Summary

Sect. 5.3 provides a general overview of the studied or developed deconvolution techniques for radio interferometric astronomical imaging. One can summarize them into two major categories:

1. locally optimum methods such as CLEAN and its variants (MRC, multi-scale CLEAN, and ASP-CLEAN), and LS-MVI.

2. globally optimum methods such as MEM, LSI, and CS.

In the first category, a pixel-based source modelling is used, the limitations to which was extensively discussed. Although, multi-scale CLEAN and ASP-CLEAN improved the CLEAN performance by considering a scale-senstive pixel-based source modelling scheme, they are still limited by the resolution limit of the interferometer for the sources, that extend larger than the PSF width. Moreover, the number of significant basis or degrees of freedom (DOF) is still unknown.

In the second category, MEM is a nonlinear problem formalism unlike the LSI and CS. LSI is reduced to the problem formalism in Eq. 5.4. MEM and CS both use the constraint of the measured data, which can be optimally represented in an orthonormal space.

The methods that seek a globally optimum solution such as MEM, LSI and CS can accommodate the use of orthonormal basis functions, that span the whole source area as their source modelling scheme. Moreover, PSWF derivation is constrained by the Shannon number, N_D, providing the number of significant basis functions required for an optimal source representation. Non-orthonormal redundant discrete à trous wavelets, that span the whole source area have previously been considered for radio astronomical source modelling (Maisinger et al. 2004). A robust comparison of these two classes of basis functions for radio astronomical imaging remains to be done.

5.4 Efficient computation of the PSWF

The application of orthonormal basis functions such as PSWF for accurate source modelling of the extended sources in radio astronomy was discussed earlier. They are of great importance for high fidelity, high dynamic range imaging with new radio telescopes as well as conventional ones. But the construction of PSWF is computationally expensive compared to other closed form basis functions. In this section, we suggest a solution to reduce its computational cost by more efficient construction of the matrix kernel (Eq. 5.14), in several steps explained in the following subsections. We achieve this by down-sampling both the Fourier plane visibilities as well as the image pixels.

Most astronomical (as well as other) images are represented by a regular 2-D array of rectangular pixels. This is required for efficient storage and display purposes and moreover, comes naturally as a by product of the FFT in imaging. We propose the use of Delaunay triangulation (Persson and Strang 2004) as opposed to regular gridding of an image for a finer selection of the ROI during the PSWF kernel construction. Delaunay triangulation allows to select fewer number of pixels outside the ROI and make a finer selection of the image inside the ROI to represent the image during the construction of PSWF. We will show that the computational efficiency improves without loss of information. Once the PSWF basis has been constructed in this manner, we revert back to the original pixel grid by interpolation and thereafter, conventional imaging operations can be applied.

5.4.1 Downsampling the visibility

Since we already have an irregular set of N_a sampling points on the Fourier plane, it is straightforward to downsample this to a lower value $N_a^{'}$, which is already presented in (Yatawatta 2011a). However, we can do even better by combining this with imaging weights. Each point on the Fourier plane, (say p) has a weight associated with it, $\rho(u_p, v_p) \in [0, 1]$. Prior to taking the Fourier transform in Eq. 5.3, we multiply each data point $\widetilde{f}(u_p, v_p)$ by this weight. When we downsample, we generate a uniformly distributed random number $r(u_p, v_p) \in [0, 1]$ and if $r(u_p, v_p) \leq \rho(u_p, v_p)$ we select the p-th point.

5.4.2 Reducing the image pixels but finer selection of the ROI

The motivation behind reducing the number of image pixels is that not all the pixels are required to define the ROI (because of the Landau-Pollak theorem) while we construct the PSWF. In Fig. 5.2 (a), we have shown the ROI of the source as well as the PSF. The amount of information can be loosely translated to the number of PSFs that we can pack within the ROI, as shown in Fig. 5.2 (b). The optimal way to pack the PSFs within the ROI boils down to a circle packing problem, which is generally understood as NP-hard.

Therefore, we follow a heuristic approach: Instead of solving a packing problem, we construct an irregular grid of triangles to cover the ROI as well as the area of

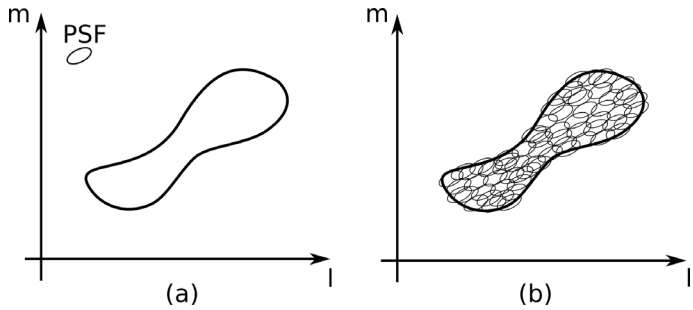


Figure 5.2: (a) The ROI of an extended source and the PSF (the ellipse). (b) The ROI covered by the PSF (an ellipse packing).

the image outside the ROI. Obviously, we employ Delaunay triangulation for this purpose. We shall illustrate this with an example as follows: In Fig. 5.3, we have shown an image of an extended radio source. This image (dimensions 118×64) has 7552 pixels.

In Fig. 5.4, we have shown the ROI for this source, which has 1826 pixels to represent the amount of information present. Therefore, we have generated the Delaunay triangulation for the ROI and its exterior as shown in Fig. 5.5 such that the triangle size within the ROI is smaller than the exterior triangles. We take the centroids of the triangles as our reduced grid to generate the PSWF. Therefore, we end up with a total number of pixels of 1220 while the ROI has 1080 pixels.

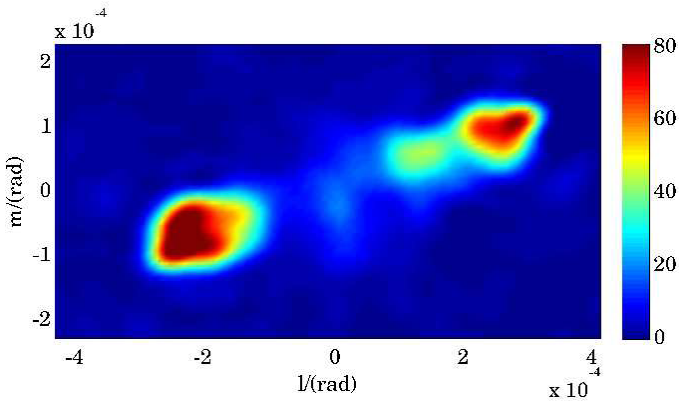


Figure 5.3: A test image of an extended radio source, with dimensions 118×64 pixels.

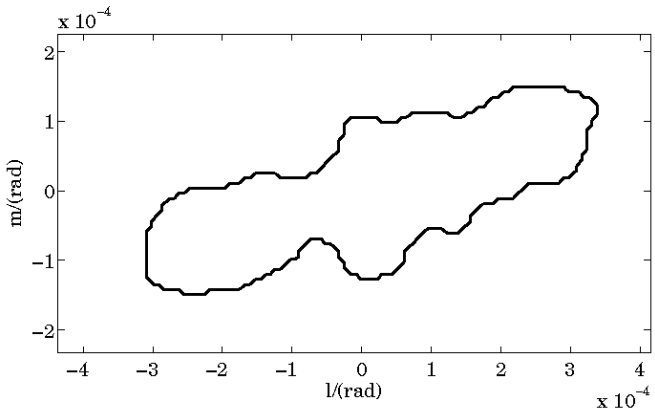


Figure 5.4: The ROI of the source in Fig. 5.3. The total image has 7552 pixels while the ROI has 1826 pixels.

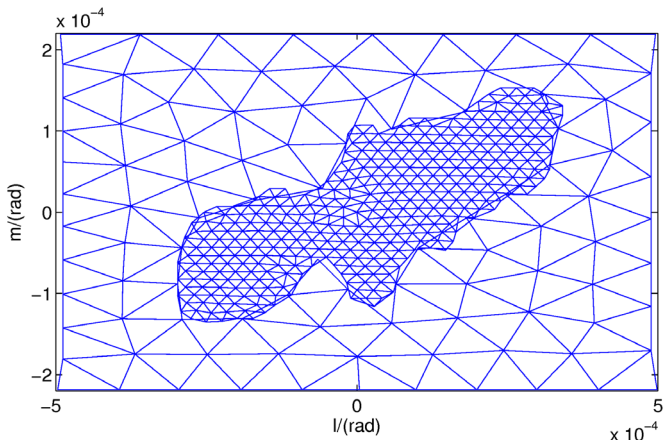


Figure 5.5: Delaunay triangulation of the image in Fig. 5.3. The triangulation inside the ROI has 1080 triangles while the triangulation outside the ROI has 140 triangles. The triangle size within the ROI is made smaller than the exterior triangles.

5.4.3 Reducing cost of pseudoinverse in kernel relation

The computational cost of Eq. 5.14 mostly involves evaluation of the pseudoinverse $(\mathbf{T}\mathbf{T}^H)^\dagger$. Instead of directly evaluating this, we solve a set of linear equations such as:

$$(\mathbf{T}\mathbf{T}^H + \gamma\mathbf{I})\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{y} \tag{5.31}$$

where γ is a small positive regularization parameter. Depending on the context, we

are given $\mathbf{y}$ and we need to find $\mathbf{x}$. The solution of the linear system for $\mathbf{x}$ is done using Q-less QR factorization.[8]

Given $\mathbf{QR} = \mathbf{A}$ (where $\mathbf{Q}$ is unitary and $\mathbf{R}$ is upper triangular), we first solve

$$\mathbf{R}^H\mathbf{R}\widetilde{\mathbf{x}} = \mathbf{A}^H\mathbf{y} \tag{5.32}$$

for $\widetilde{\mathbf{x}}$ and make one step correction:

$$\begin{aligned} \mathbf{r} &= \mathbf{y} - \mathbf{A}\widetilde{\mathbf{x}} \\ solve \quad \mathbf{R}^H\mathbf{R}\widetilde{\mathbf{e}} &= \mathbf{A}^H\mathbf{r} \\ \mathbf{x} &= \widetilde{\mathbf{x}} + \widetilde{\mathbf{e}} \end{aligned} \tag{5.33}$$

to get the solution $\mathbf{x}$. Note that matrix $\mathbf{A}$ represents an arbitrary matrix in this subsection and should not be mistaken with the array response matrix, which appeared in Eq. 5.27.

5.4.4 Results

To demonstrate how our solution works in practice, we give results on the construction of PSWF for the extended source shown in Fig. 5.3. The source was observed by the LOFAR radio telescope during system testing. For an 8 hours of observation, we obtain, $N_a = 7896636$ data points on the Fourier plane. We reduce this to $N_a^{'} = 11041$ points by applying our downsampling scheme as explained in section 3.1.

We use the method of Persson and Strang (2004) to generate conforming Delaunay triangulations for the ROI and the exterior of the image. Selection of triangle scales for inside and outside of the ROI (see Fig. 5.5), has to be done in such a way that the least amount of information is lost. For this observation, the PSF minor diameter is about $d = 4.2 \times 10^{-5}$ radians. To achieve the most appropriate triangulation, we have set different scales for the triangles inside the ROI ($a \times d$) and outside the ROI ($b \times d$), where a, b are the parameters that we vary.

We show the eigenvalue spectrum of the kernel $\mathbf{K}$, for different values of a and b in Fig. 5.6. The eigenvalue spectrum gives us a measure of information a certain basis can represent. By the results shown in Fig. 5.6, we see that in most settings, we get about 100 basis functions that carry significant information, which is close to the limit given by Landau-Pollak theorem for this case. We approximate the geometry of ROI ($\mathbf{I}_b$) by selecting different triangulation scales (a and b). This explains why there are slightly different eigenvalue spectrum. The basis functions are still data independent and they depend only on the shape of the ROI.

Once we obtain the basis functions for the downsampled grid, we interpolate back to the original image grid to get the basis vectors corresponding to the original image. Let $\mathbf{P}$ be the matrix whose columns correspond to the interpolated basis

[8] In linear algebra, a QR factorization (also called a QR decomposition) of a matrix is a decomposition of an arbitrary matrix $\mathbf{A}$ into a product $\mathbf{A} = \mathbf{QR}$ of an orthogonal matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$. QR factorization is often used to solve the linear least squares problem efficiently.

vectors. Then, we study the Frobenious norm of $\mathbf{P}^H\mathbf{P} - \mathbf{I}$ and the condition number of $\mathbf{P}^H\mathbf{P}$ in Fig. 5.7. This, in addition to the eigenvalue spectrum in Fig. 5.6 give us a benchmark to choose the most suitable triangulation (or values of a and b). If the basis vectors are orthonormal, we will obtain the lowest value for the Frobenious norm. Fig. 5.7 (a) shows that the optimality of the source modelling is not affected significantly from one setting to another, because the basis vectors are orthonormal with negligible errors, thus the CRB remains minimum (Yatawatta 2010). Moreover, the condition number of $\mathbf{P}^H\mathbf{P}$ helps us to understand how its condition affects the further computation. Ultimately, these conclude that the setting with $a = 0.2$ and $b = 2.1$ is the most suitable one to fulfill the aforementioned criteria. By this setting, we downsample to $N^{'} = 3270$ pixels in which the ROI has $N_b^{'} = 3199$ pixels. In contrast, the original pixel grid has $N = 7552$ and $N_b = 1826$.

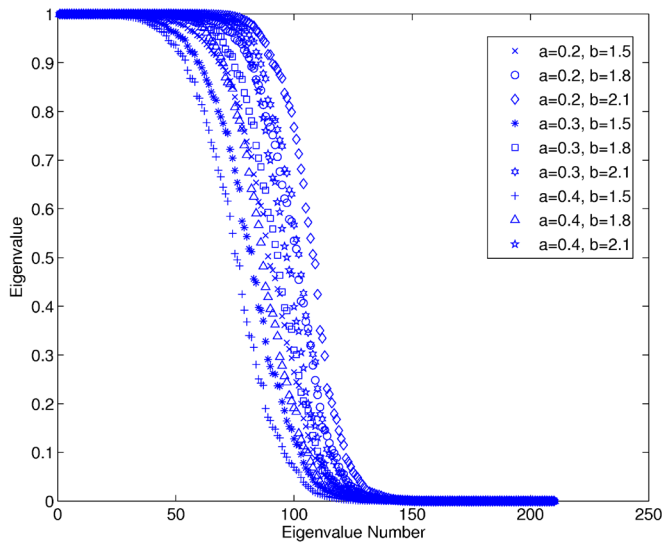


Figure 5.6: Eigenvalue spectrums for different triangulations, with varying a and b.

The use of PSWF for source modelling and deconvolution of extended sources was proposed earlier in (Yatawatta 2011b,a). Here, we suggested a scheme to reduce its computational cost. By using the Delaunay triangulation, we can significantly reduce the cost of construction of PSWF. We have also shown by a real example that we do not lose information by this cost reduction. Further improvement will focus on finer selection of the ROI i.e. the boundary of the ROI is represented with more triangle pixels.

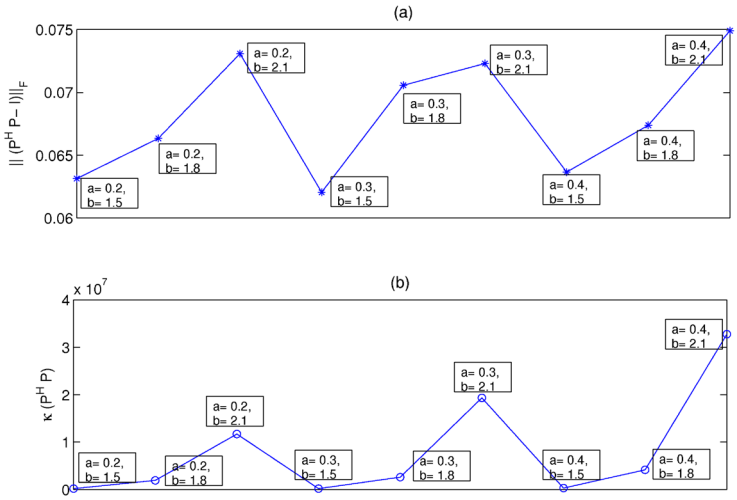


Figure 5.7: (a) Frobenious norm of $\mathbf{P}^H\mathbf{P}-\mathbf{I}$ (scaled by $1/N_{basis}^2$ where N_{basis} is the number of basis functions) for different triangulations (varying a and b) of the image. (b) Condition number of $\mathbf{P}^H\mathbf{P}$ for different triangulations.

5.5 Summary and conclusions

In this chapter we presented the deconvolution problem for an extended source. We explained the limitations of using a pixel-based (with either pixels of the same size or scalable sizes) source modelling scheme for such problem. We explained an alternative way of representing an extended source using the orthonormal basis functions that span the whole ROI in an image. Based on studies (Chang and Refregier 2002; Refregier 2003; Berry and Withington 2004; Yatawatta 2010; Wei and Lamahewa 2010; Yatawatta 2011b,a), that recommend the use of orthogonal basis functions for astronomical source modelling, we presented three sets of orthogonal basis functions i.e. shapelets, PSWFs, and Franks basis functions. The latter remained unstudied in radio astronomical application but has demonstrated promising potential (Wei and Lamahewa 2010) to be a suitable alternative to PSWFs.

We gave a comprehensive overview of the deconvolution techniques that have been used in aperture synthesis imaging. The aim was to provide a suitable and unified platform to discuss their potential and limitations for handling an extended structure. In Sect. 5.3.6, we summarized that globally optimal deconvolution techniques (MEM, LSI, and CS) can accommodate the use orthonotmal basis functions, that span the whole ROI, as their source modelling scheme for an optimal performance, thus high image fidelity. Based on the analytical arguments and demonstration on real data in Yatawatta (2010, 2011a,b), we recommended PSWF as the selected set of orthonormal basis functions. One important reason is the fact that using the PSWF implicitly provides the significant number of the basis functions (model order or DOF), while

no other set of basis functions does that (see the argument in Slepian (1961, 1976)). To support its computationally affordable application, we also suggest a solution to reduce their computational cost in Sect. 5.4. We proposed to use Delaunay triangulation to construct an irregular grid with fewer number of pixels outside the ROI and a finer selection of the image inside the ROI to represent the image during the construction of PSWF with a kernel matrix of smaller size.

Given the extensive overview of the deconvolution methods and the new source modeling scheme, it becomes evident that image fidelity depends on the deconvolution method as well as the source modeling scheme that is considered for the deconvolution process. Therefore, viewing the image fidelity as a function of only the *uv*-coverage (which usually translates into the array configuration) is inadequate. To achieve the image fidelity to the fullest, we must perform a deconvolution process in the most optimal manner i.e. use the most optimal deconvolution method and most optimal source modelling scheme. Theoretically, the CS technique using PSWF should fulfill such expectation.

In summay therefore, the open questions are:

- Is a set of the Franks basis functions a better alternative than the PSWFs in radio astronomical source modelling?
- Does MEM perform better using the shapelets or the PSWFs as compared with its performance using the well-accepted non-orthogonal, redundant discrete à trous wavelets?
- Does CS perform more optimally using the PSWFs or Franks basis functions?
- Given a certain array configuration (thus given a certain *uv*-coverage), which deconvolution process (i.e. deconvolution method and its source modelling scheme) would perform most optimally i.e. provide the best fidelity?
- Does a deconvolution process based on using the CS method and the PSWFs source modeling scheme provide the image fidelity to the fullest?

Chapter 6

Summary and outlook

Here, we summarize the insights obtained in this thesis regarding improving aperture synthesis data processing. We have recognized several issues which need further research. These will be addressed in this chapter and constitute a major step toward computationally effective and efficient wide-field, high quality synthesis imaging.

6.1 Summary

Achieving high fidelity, high dynamic range images has been a never-ending quest in radio astronomy. The advent of aperture array radio interferometers such as LOFAR (de Vos et al. 2009), now operational as an important pathfinder for the SKA (Hall 2004), crucially demands the continuation of this quest so that their ground-breaking scientific outcome can be delivered. To make such unconventional telescopes a true success for the SKA, we must understand their particular engineering aspects and the corresponding challenges that they introduce in exploiting the scientific data of the instrument. Imaging is a major scientific application of these instruments, from which other modes of operation of the instrument can benefit. For this reason, we aimed to take some steps towards computationally effective and efficient high quality imaging.

In radio astronomy, self-cal and deconvolution are two intertwined processes. This implicitly means that the two measures of image quality i.e. fidelity and dynamic range are also two intertwined quantities as discussed in Yatawatta (2011b). Fidelity is the optimality in the image reconstruction process known as deconvolution and dynamic range is defined as the ratio between the peak intensity and the noise floor in the image. The noise floor is the net summation of calibration errors, confusion noise, and thermal noise (Wijnholds and van der Veen 2008). Accurate calibration reduces the noise floor (the thermal noise is fixed), thus improves dynamic range while performing an accurate calibration depends on the optimality (fidelity) of the deconvolution process i.e. the deconvolution method and its source modelling scheme.

At the same time, accurate recognition of the source area for the deconvolution process should be provided by an accurate self-cal process. This is an important insight, as achieving high fidelity, high dynamic range imaging with new radio telescopes is more challenging than ever before.

Aperture array telescopes have high sensitivity, wide FoV, and high resolution as compared with the conventional ones. They can detect many unknown faint sources as well as the known bright sources in their wide FoV. Their observations are corrupted by variable propagation path effects (e.g. the ionosphere) and variations in their primary beam pattern in time and frequency. These are known as the DDEs. The corresponding corrections have to be made per source in the FoV, and per element of the interferometer. All these entail a large computational load in the data processing as discussed in Wijnholds et al. (2010). Therefore, any new algorithm to exploit the data of such instrument (e.g. imaging and calibration algorithms) should be both optimal and computationally cheap.

An initial step to reduce the computational load is by exploiting the hierarchical structure in an aperture array telescope. This structure serves a two-fold goal, a) data reduction b) system health management e.g. station beam stability for the central processor. Each phased-array station has a local processor, at which station visibilities are computed for some engineering purposes e.g. station calibration and RFI detection. In Chapter 3, we investigated the application of redundancy method for the calibration of the phased-array stations with a regular configuration e.g. HBAs, as a sky model independent method that is computationally cheaper and faster as compared with the model-based calibration methods. The main results are:

- Use of redundant information in the visibilities allows a more accurate calibration than the methods that are only model-based. In addition, we can improve the phase solution with additional constraints on the ionospheric refraction.
- Redundancy also provides a powerful model-independent tool for system diagnostics at different levels of the telescope phased-array hierarchy (both intra-station and inter-station).
- Mutual coupling between the elements of a phased-array station seriously affects the calibratability of the station, regardless of the calibration method. Therefore, we recommend minimization of the mutual coupling as a figure of merit in a station configuration. Otherwise, a rigorous electromagnetic study is required that results in either the beam model (including both main lobe and sidelobes) of each and individual station element or the model of mutual impedances between the station elements. This should be avoided, as it calls for more computational resources at the station local processor.

For the two first reasons, we recommend redundancy in the configuration of both the phased-array stations and the whole array. SKA configurations must offer redundant baselines, that allows us to use the redundancy calibration to improve the dynamic range of the imaging. Redundant baselines in the SKA core are also recommend in Bunton (2011) for power saving in the correlator and image processing. Wijnholds and Noorishad (2012) also proposed an optimal calibration scheme for regular

arrays. Moreover, there are several variants of redundancy calibration (Li and Er 2006; Heidenreich and Zoubir 2011; Heidenreich et al. 2012), that are more optimal and computationally cheaper than its formalism in Wieringa (1991); Noorishad et al. (2012); Wijnholds and Noorishad (2012), but have never been tested for radio astronomy. They exploit the Toeplitz structure in the visibility matrix and use the diagonal and the super-diagonal terms of the visibility matrix as opposed to its original formalism (Wieringa 1991), that exploits only and all the off-diagonal terms of the visibility matrix.

After station calibration the outputs of the station elements are beamformed to represent an element of the whole aperture array telescope e.g. a HBA station of LOFAR.

Following the 3GC trend, calibration at the whole array level involves correction for the corruptions due to the propagation path effects e.g. ionosphere and the instrument properties e.g. beam pattern, known as the DDEs. Although these effects are tackled by estimating a DDE gain, we believe that we should treat them differently, because of two important reasons a) these effects are of different natures i.e. the ionosphere (the propagation path effects, in general) is stochastic in nature but beam pattern and even its variability are deterministic effects b) the more parameters we estimate in the calibration process, the larger inaccuracy we allow in the estimated parameters. Our suggestion is to model the beam pattern of a phased-array station taking into account the mutual coupling effect and understand its variability as a function of beamforming weights and the failing station elements. Such model of beam and beam variability should include the whole beam i.e. both the main beam and the sidelobes. This is highly recommended because of high sensitivity of the array, thus detecting also weak sources from any angle. It should also take the projection effect (Fig. 1.6) into account, so that we do not need to model that effect as an instrumental polarization. Such beam modelling approach will save computational capacity and time, if the beam is modelled efficiently with a few basis functions as studied by De Lera Acedo et al. (2011); Maaskant et al. (2012); Maaskant and Ivashina (2012); Craeye et al. (2012). We should also constrain such beam model with additional physically justifiable constraints e.g. certain level of main lobe and sidelobes known from the electromagnetic studies, or beam positivity etc. Then, we can provide our acquired knowledge of beam as a prior information to the calibration routine instead of estimating for it as a DDE gain. Providing such beam model is a major step towards convergence of the calibration algorithm to the true solutions.

Such constraining beam model is required regardless of the calibration algorithm, whether based on a least squares approach (Boonstra and van der Veen 2003; van der Tol et al. 2007; Wijnholds and van der Veen 2009a) or on the expectation maximization approach (Yatawatta et al. 2008; Kazemi et al. 2011). Although the need of providing the calibration algorithm with some extra constraints has been recognized as a necessity, not much has been done to put this idea into practice. Investigating such an issue becomes even more important, because the beam model and its variation affect the data of different mode of operation of an instrument, differently. Thus, they must be understood very well and used as an extra constraint in the calibration routine.

In Chapter 4, we assumed that the beam model is known well and provided for self-cal. We studied the effect of the refractive property of ionosphere as a DDE that is a source of inaccuracy in self-cal gain estimation results. We suggested solving a constrained self-cal for an array with an arbitrary geometry, using extra constraints as exploited in redundancy calibration (Wieringa 1991). We demonstrated that those constraints have a clear physical justification, only when we observe a single source or multiple sources, that experience similar ionospheric effect. We discussed how cluster calibration might reduce any other observational scenario (illustrated in Fig.1.7) to one, in which these constraints are justified.

Self-cal is formulated as a non-linear least squares problem. Its constrained version can be solved using SQP technique, if it cannot be linearized. We concluded that a constrained self-cal gain estimation is performed more optimally based on our numerical CRB analysis and comparing it with the CRB of its unconstrained version. Then, we solved a constrained self-cal taking two different approaches: the SQP and the LOGLS. The main results are:

- The self-cal can benefit from the constraints used in the redundancy calibration formalism to perform more reliable gain estimation for an array with an arbitrary geometry. We note that this became possible using cluster calibration in wide-field observations.

- It is worth noting that our suggested solution, which is similar to that of van der Tol et al. (2007), also avoids the unitary ambiguity in self-cal results.

We explained the problem using a full ME. One may argue that constraining the calibration routine is not straightforward, when it is formulated in 2×2 MEs, supported by the 3GC. One potentially feasible way of constraining such calibration formalism is to provide accurate model of parameters e.g. beam model (Maaskant et al. 2012; Maaskant and Ivashina 2012) and to use cluster calibration to avoid the ambiguity problem.

Calibrated visibilities (the Fourier samples of the true sky brightness) are used in the deconvolution process. To achieve the ultimate goal of high fidelity high dynamic range radio imaging, the bright known sources in the FoV (either point-like sources or extended structures) have to be modelled and removed accurately such that the faint unknown sources in the background can appear with minimum error. To this end, we gave an overview of the deconvolution methods and their corresponding source modelling scheme in Chapter 5. We discussed the main limitations of a pixel-based (either with pixels of the same size (Högbom 1974; Clark 1980) or multi-scale (Cornwell 2004; Bhatnagar and Cornwell 2004)) source modelling scheme for a source that extends larger than the PSF. To overcome such limitations, the use of orthonormal basis functions, that span the whole source area has been proposed as an optimal source modelling scheme (Yatawatta 2010). Following the promising results of using orthonormal basis functions, especially the PSWFs, which also provide the significant number of the required basis functions (Yatawatta 2011b,a), we suggest a solution to reduce their computational cost (Noorishad and Yatawatta 2011). The main results are:

- Globally optimal deconvolution methods such as MEM, LSI, and CS can accommodate the use of orthonormal basis functions (that span the whole ROI for the extended sources) as their source modelling scheme, that leads to a deconvolution process with minimum error, thus maximum fidelity.
- PSWF is the only set of orthonormal basis functions that provide the significant number of required basis functions (Shannon number) or the model order. This is an important property as compared to other sets of basis functions.
- We have a solution in hand for significantly reducing the computational cost the PSWFs in several steps. We proposed to use Delaunay triangulation to construct an irregular grid with fewer number of pixels outside the ROI and a finer selection of the image inside the ROI to represent the image during the construction of PSWF. We can control the size of the triangulation without losing information. Once the PSWF basis has been constructed in this manner, we revert back to the original pixel grid by interpolation. We showed that the computational efficiency improves without loss of information.
- Our extensive discussion on the deconvolution problem, the corresponding methods and source modelling schemes reveals that quantifying the image fidelity based on only *uv*-coverage is inadequate.
- The application of orthonormal Franks basis functions (Franks 1969, Ch. 6), as the generalized form of PSWFs remained in a very early stage of study in Chapter 5. They are expected to have a flatter lobe as compared with the PSWFs, based on the analysis and weighting scheme presented in Wei and Lamahewa (2010), thus better representation of the source area in an image.

Although the urge of using a source modelling scheme based on the scalable basis functions for the extended sources, as opposed to a pixel-based source modelling scheme has already been discussed in several bodies of research (Refregier 2003; Maisinger et al. 2004; Cornwell 2004; Bhatnagar and Cornwell 2004; Cotton and Uson 2008; Yatawatta 2010), orthonormal basis functions, that potentially span the whole ROI are not yet well-accepted in the radio astronomical community, as the optimal choice. This can be done by investing some time in the application of PSWF and Franks basis functions for the radio astronomical source modelling and demonstrate their superiority to the other sets of basis functions.

6.2 Outlook

The radio interferometric data (visibilities) are the irregular and sparse samples of a time-limited, band-limited signal, that can be optimally modelled using orthonormal basis function (Slepian 1961) (especially in the case of modelling an extended source), and can be optimally recovered using CS technique with far less samples (Rudelson and Vershynin 2006). This unconventional view of radio astronomical data has proven promising for a significant improvement in the image quality (Levanda and Leshem

2008; Yatawatta 2010; McEwen and Wiaux 2011a) as compared to standard CLEAN and its variants (Högbom 1974; Clark 1980; Brinks and Shane 1984; Wakker and Schwarz 1988; Cornwell 2004; Bhatnagar and Cornwell 2004). The CS technique assumes an orthonormal space for the optimum signal recovery (Candès and Romberg 2007). However, the set of orthonormal basis function is not specified. The application of PSWF and Franks basis functions as an potential improvement in the performance of the CS technique becomes an interesting research question.

Given the wide FoV of phased-array stations, the curvature of the celestial sphere cannot be dismissed in aperture array observations, thus in their calibration and deconvolution processes. In the conventional arrays, this effect has been visible only on non-coplanar baselines and tackled using facet-imaging (Perley and Bridle 1994) and w-projection methods (Cornwell et al. 2008) such that the aperture synthesis imaging has always been done in a 2D-plane. Matrix-based data modelling and data analyzing (van der Veen et al. 2004) allows us to incorporate the w-term of the visibility domain, and seek a globally optimal solution in the deconvolution process itself, thus in the calibration process. Taking into account the presence of the w-term in the visibility relation in Eq. 5.1, one may simply notice that the relation in Eq. 5.3 can still hold, except that matrix $\mathbf{T}$ contains $e^{-j2\pi(l_q u_p + m_q v_p + (n_q - 1)w_p)}$ on its q-th row and p-th column, in this case. The size of the matrix $\mathbf{T}$, thus its corresponding computational cost remains unchanged. This implies that we are able to do 3D imaging in celestial sphere as opposed to w-projection and that 3D imaging should reduce the computational load of the deconvolution process. However, there has been no analytical or numerical evaluation of the computational cost or performance of this approach as opposed to the w-projection using real data or even realistic simulated data.

Although this possibility was initially discussed in the 70's (Brouw 1971; Clark 1973), it did not attracted enough attention due to the convenience of image display in 2D, and high computational cost of FFT in 3D space, at the time. The latest reseach in the field of radio astronomical data analysis have discussed the advantages of the data analysis in 3D on celestial sphere (Carozzi and Woan 2009; McEwen and Wiaux 2011b; Bregman 2012). Moreover, significant ongoing research has been carried out on the computational efficiency of higher order FFT e.g. 3D FFT (Frigo and Johnson 1998; Frigo 1999; Frigo and Johnson 2005), allowing to reduce the computation cost and speed.

While the deconvolution process in 3D (on the celestial hemisphere) has not been tested for MEM or LSI, it has been tested for the CS (Wiaux et al. 2009b; McEwen and Wiaux 2011b). In fact, the CS technique benefits from an phenomenon known as spread spectrum in (u, v, w)-space instead of in (u, v)-space with their projected w-term. Wiaux et al. (2009b,c) showed that spread spectrum increases the data incoherency, thus improves the CS algorithm performance. In their analysis, they used a constant w-term for all visibility points, which is not a proper representation of real observations in wide-field imaging. Such realistic analysis remain an important research question, given that the CS technique has attracted some attention in the radio astronomical community and perform better in 3D space.

In view of the CS technique, redundancy in the measured data should not harm the

image fidelity,[1] to the contrary of the common notion, that seeks a filled *uv*-coverage to guarantee a higher image fidelity. An interesting question is how the image fidelity gets affected, having redundancy in the measured data, thus sparser samples of the visibility points. The answer to this question is important as the redundant baselines in the SKA core are recommend in several bodies of research e.g. (Bunton 2011; Noorishad et al. 2012). Moreover, the SKA configuration provides so many redundant baselines by default, that allows us to use the redundancy calibration, and improve the dynamic range imaging (Noordam and de Bruyn 1982).

The argument of array configuration for an optimal operational performance of the interferometer (in particular in imaging) is very important. However, all attempts in this regards have been made without keeping an eye on the methods of deconvolution (Lal et al. 2009; Millenaar and Bolton 2010; Wijnholds 2012). Given the extensive overview of the deconvolution methods and the new source modelling scheme, it becomes evident that viewing the image fidelity as a function of only the *uv*-coverage (which usually translates to the array configuration) is inadequate.

Last but not least, the data model discussed in Chapter 2 revealed some important issues including the non-Gaussianity of the noise, that is a fundamental assumption for many algorithms in the calibration and the deconvolution, especially in the calibration methods exploiting the statistical properties of the noise such as (Yatawatta et al. 2008; Kazemi et al. 2011, 2012). An interesting research question is whether the performance of those calibration methods would change taking non-Gaussianity into account.

In summary, the open questions are:

- Is 3D imaging the future of wide-field aperture synthesis imaging as a computationally more efficient and effective choice as compared with the conventional step of *w*-projection in decovolution process?
- Will the CS technique provide the image fidelity to the fullest using PSWFs or Franks basis functions in deconvolution of the extended sources?
- How does redundancy (i.e. larger sparsity in the observed data) affect the fidelity of the final image when using the CS technique?
- How applicable are the other implementations of the redundancy calibration for aperture synthesis data processing, given that they are proven to be more optimal than the redundancy calibration currently in use in aperture synthesis community?
- Are the Franks orthonormal basis functions a better choice than the PSWF for modelling the extended sources?
- How do the speed and accuracy of convergence in a calibration routine improve by providing a complete beam model as explained earlier with some extra physically justifiable constraints known a prior e.g. expected level of sidelobes and main lobe from its electromagnetic studies?

[1] Given that after calibration, only one data point represents the redundant measurements.

- How effectively do we constrain the calibration routine formulated using 2×2 MEs?
- Will the performance of the calibration methods, that exploit the statistical properties of noise change, considering the non-Gaussianity of the noise?
- How to incorporate all the essential criteria in a station configuration to meet its calibratability (e.g. minimum mutual coupling) and yet fulfill other criteria as an interferometer element (e.g having a certain level of sidelobes)?

To achieve the ultimate goal of high quality imaging, thus accurate calibration (from which all other instrument modes of operation can benefit), questions such as the ones raised here should be answered to serve the scientific goals of the radio telescopes of the future. We expect that further research in this area will prove to be very valuable if not essential.

Appendix A

Abbreviations and notations

A.1 Abbreviations and acronyms

3C	third Cambridge (catalog)
1GC	first Generation of Calibration
2GC	second Generation of Calibration
3GC	third Generation of Calibration
AIPS	Astronomical Image Processing System
ASTRON	Netherlands Institute for Radio Astronomy
ASP-CLEAN	Adaptive Scale Pixel CLEAN
BBS	BlackBoard Selfcal
BFGS	Broyden-Fletcher-Goldfarb-Shanno
CLEAN	iterative beam removal technique (Högbom 1974)
CBFM	Characteristic Basis Function Method
CAESAR	Computationally Advanced and Efficient Simulator for ARrays
CRB	Cramér-Rao Bound
CS	Compressed Sensing
DDE	Direction Dependent Effect
DIE	Direction Independent Effect

DOA	Direction of Arrival
EM	ElectroMagnetic
EMBRACE	Electronic Multi-beam RAdio Astronomy Concept
EoR	Epoch of Reionozation
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix
FoV/FOV	Field-of-View
FPA	Focal Plane Array
FPC	Fixed Point Continuation
HBA	High Band Antenna
HPBW	Half Power Beam Width
i.i.d.	independent identically distributed
LBA	Low Band Antenna
LEA	Linear Equi-spaced Array
LOFAR	Low Frequency Array
LOGLS	logarithmic least squares
LS	Least Squares
LS-MVI	Least Squares Minimum Variance Imaging
ME	Measurement Equation
MEM	Maximum Entropy Method
ML	Maximum Likelihood
MeqTree	measurement equation tree
MRC	Multi-Resolution CLEAN
MRI	Magnetic Resonance Imaging
MW	MicroWave
MWA	Murchison Wide-field Array
OSKAR	Beamforming simulation for the SKA aperture array

PAF	Phased Array Feed
pdf	probability distribution function
PSF	Point Spread Function
PSWF	Prolate Spheroidal Wave Function
RFI	Radio Frequecny Interference
SAGE	Space Alternating Generalized Expectation maximization
SDR	Spatial Dynamic Range
ROI	Region of Interest
SKA	Square Kilometre Array
SPAM	Source Peeling and Atmospheric Modeling
SQP	Sequential Quadratic Programming
SNR	Signal-to-Noise Ratio
ThEA	Thousands Eelement Array
TV	Toatal Variation
ULA	Uniform Linear Array
URA	Uniform Rectangular Array
UTC	Coordinated Universal Time
VLA	Very Large Array
WALS	Weighted Alternating Least Squares
WLS	Weighted Least Squares
WSRT	Westerbork Synthesis Radio Telescope

A.2 Notations

A.2.1 Symbols

a, A	italic lower or upper case characters denote scalars
$\mathbf{A}$	bold upper case characters denote matrices
$\mathbf{1}_{m\times n}$	$m \times n$ matrix filled with ones
$\mathbf{0}_{m\times n}$	$m \times n$ matrix filled with zeros
$\mathbf{I}$	idendity matrix of appropiate size
$\hat{.}$	estimated value

A.2.2 Operations

$\mathcal{E}(.)$	expectation operation
$\delta(.)$	delta or Dirac function
$\frac{\partial}{\partial x}(.)$, ∇	gradient of a function w.r.t. x
$(.)^T$	transpose
$(.)^H$	Hermitian or conjugate transpose
$(.)^{-1}$	inverse
$(.)^\dagger$	pseudo-inverse
$e^{(.)}$	element-wise application of the exponential function
$< . >$	inner product of two vestors
$\bar{(.)}$ or $(.)^*$	conjugate
vec(.)	stacks the columns of a matrix $\mathbf{A}$ into a vector denoted by $\mathbf{b} = \text{vec}(\mathbf{A})$
diag(.)	converts a vector to a diagonal matrix, with vector placed on the main diagonal
round(.)	rounds up to an integer
$\odot$	Hadamard or element-wise product of two matrices or vectors
$\otimes$	Kronecker product of two matrices or vectors
$\circ$	Khatri-Rao or column-wise Kronecker product of two matrices
tr(.)	trace of a matrix
A_{ij}	(i, j) element of a matrix $\mathbf{A}$
ln(.)	natural logarithm
$\|.\|_F$	Frobenius norm
$\|.\|_{\ell_1}$	ℓ_1-norm
$\|.\|$	Euclidian norm
$\|.\|$	absolute value
$\kappa(.)$	returns the condition number of a matrix
$\angle(.)$	returns the angle or phase of a complex number

Appendix B

Plots of residuals for corrected redundant visibilities

In order to study the performance of the redundancy and model-based methods for phased-array station calibration, we present additional plots of the residuals for the corrected redundant visibilities here. The data is taken from our observational campaign on 24 November 2009. Fig. B.1 and Fig. B.2 show similar results of the same data set (captured at 21:29:04 UTC, when CasA was at high elevation) presented in Fig.3.8 but for different baselines, shown in the rightmost panel.

Fig. B.3- B.5 show similar results using the data captured at 15:25:43 UTC, when CasA was at low elevation. They reveal slightly larger residuals in the order of $5-10\%$ than the results in Fig. 3.8, Fig. B.1 and Fig. B.2. This is due to a different mutual coupling environment which leads to less identical beams in either the main lobe or the sidelobes. As discussed in Sect. 3.3.2, we still see smaller residuals after redundancy calibration as compared with the model-based method.

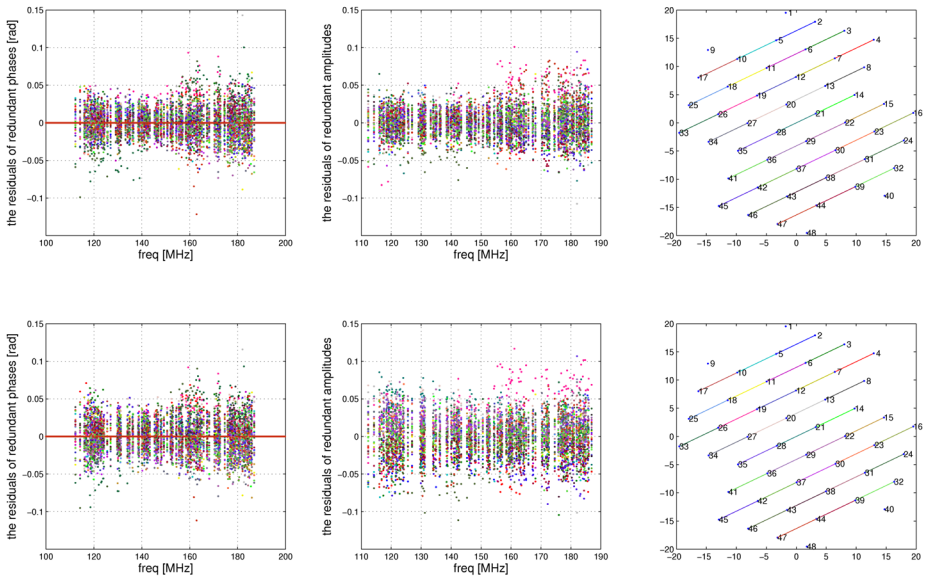


Figure B.1: Plots of the residuals for corrected redundant visibilities in terms of phases and amplitudes on a given set of redundant baselines. The first row shows the results after redundancy calibration. The second row shows the result afters model-based calibration. The data is taken from an observation done on 24 November 2009 at 21:29:04 UTC, when CasA was at high elevation. The station configuration of RS208 is shown in the rightmost panel. The corresponding redundant baselines to each redundant visibility are also depicted on the station layout, with the same color code.

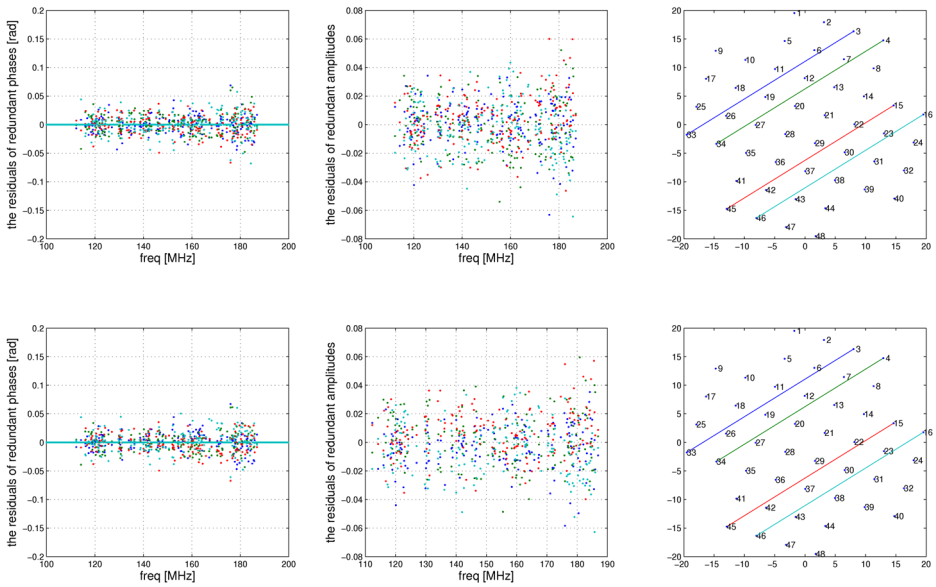


Figure B.2: Continued from Fig. B.1.

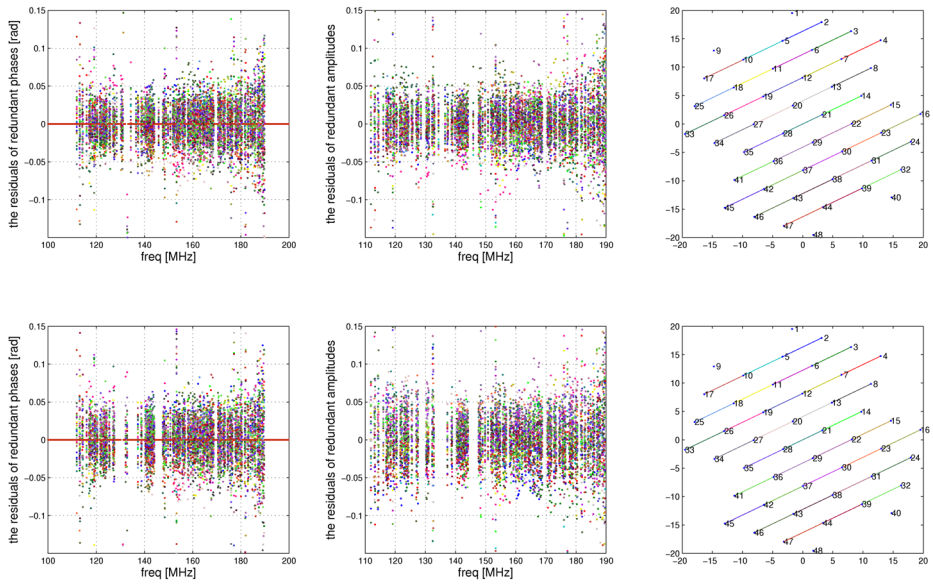


Figure B.3: The plots of the residuals for corrected redundant visibilities in terms of phases and amplitudes on a given set of redundant baselines. The first row shows the results after redundancy calibration. The second row shows the result afters model-based calibration. The data is taken from the observation done on 24 November 2009 at 15:25:43 UTC, when CasA was at low elevation. The station configuration of RS208 is shown in the rightmost panel. The corresponding redundant baselines to each redundant visibility are also depicted on the station layout, with the same color code.

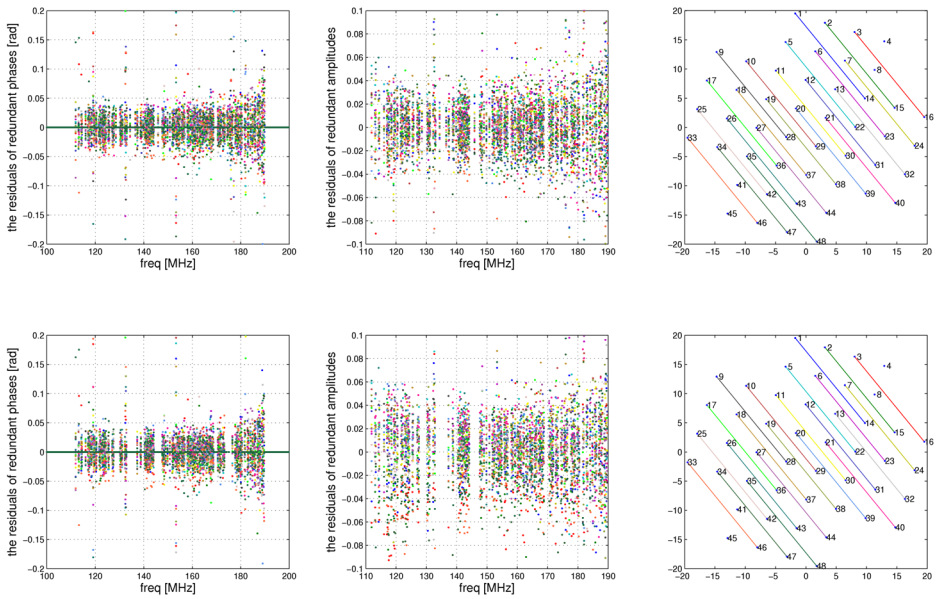


Figure B.4: Continued from Fig. B.3.

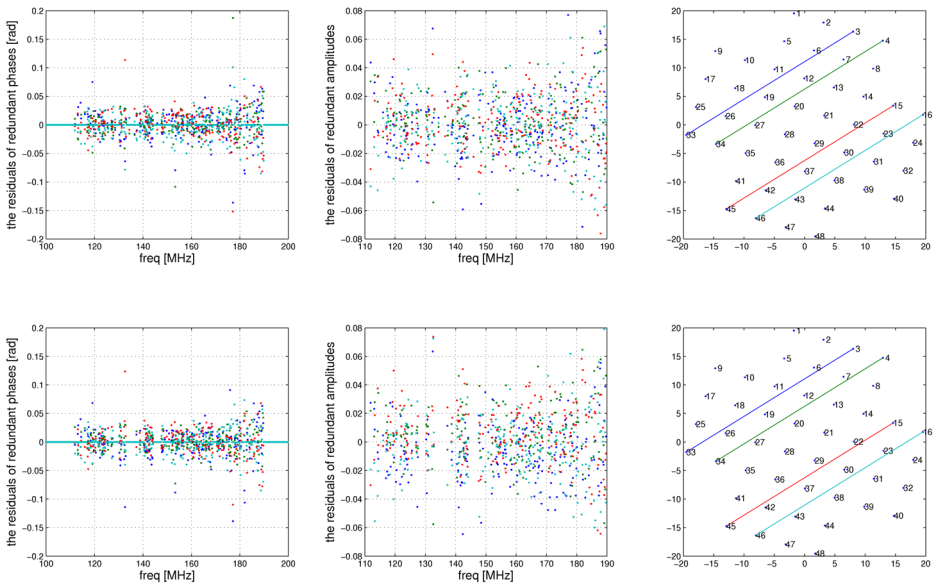


Figure B.5: Continued from Fig. B.4.

Appendix C

Vectorizing the calibrated visibilities

In Chapter 5, we assume that we have calibrated visibilities, in order to explain the deconvolution techniques. The calibrated data is modeled as

$$\mathbf{R} = \mathbf{A}\mathbf{\Sigma}_{\mathrm{s}}\mathbf{A}^H + \mathbf{\Sigma}_{\mathrm{n}}, \tag{C.1}$$

where all the corruptinmg factors have been corrected for. $\mathbf{A}$ is the array response matrix (described in Eq. 2.9 in Chapter 2) and $\mathbf{\Sigma}_n$ is the general noise term (described in Sect. 2.1.2 in Chapter 2). We remember that the term $\mathbf{\Sigma}_{\mathrm{n}}$ carries the background sky noise ($\mathbf{R}_{\mathrm{sky}}$) on its off-diagonal terms and the receiver thermal noise ($\mathbf{R}_{\mathrm{rec}}$) on its diagonal terms. It was discussed that the sky noise is convoluted by the PSF, in the same way as the celestial signals of interest. To make this point clear, we expand $\mathbf{\Sigma}_{\mathrm{n}}$ to its terms, and rewrite Eq. C.1 as

$$\mathbf{R} = \mathbf{A}\mathbf{\Sigma}_{\mathrm{s}}\mathbf{A}^H + \mathbf{A}\mathbf{\Sigma}_{sky}\mathbf{A}^H + \mathbf{R}_{rec} = \mathbf{A}(\mathbf{\Sigma}_{\mathrm{s}} + \mathbf{\Sigma}_{\mathrm{sky}})\mathbf{A}^H + \mathbf{R}_{\mathrm{rec}}, \tag{C.2}$$

where, the first term implies the convolution of signals and the sky noise with the PSF. In fact, a more accurate definition of the calibrated data is when the thermal reciever noise is also known and removed from the data, thus it is

$$\mathbf{R} - \mathbf{R}_{\mathrm{rec}} = \mathbf{A}(\mathbf{\Sigma}_{\mathrm{s}} + \mathbf{\Sigma}_{\mathrm{sky}})\mathbf{A}^H. \tag{C.3}$$

In Chapter 5 in Eq. 5.3, we reformated the calibrated data into a vector $\widetilde{\mathbf{f}}$, which can be obtained by applying the operator vec(.) to Eq. C.3 as

$$\widetilde{\mathbf{f}} = \mathrm{vec}(\mathbf{R} - \mathbf{R}_{rec}) = \mathrm{vec}(\mathbf{A}(\mathbf{\Sigma}_{\mathrm{s}} + \mathbf{\Sigma}_{sky})\mathbf{A}^H). \tag{C.4}$$

Given the following linear algebric relation for arbitrary matrices of $\mathbf{B}$, $\mathbf{X}$, and $\mathbf{C}$ in

$$\mathrm{vec}(\mathbf{BXC}) = (\mathbf{C}^T \otimes \mathbf{B})\mathrm{vec}(\mathbf{X}), \tag{C.5}$$

we can subsitude the terms in Eq. C.4 as follows: $\mathbf{B} = \mathbf{A}$, $\mathbf{X} = \mathbf{\Sigma}_{\mathrm{s}} + \mathbf{\Sigma}_{sky}$, and $\mathbf{C} = \mathbf{A}^H$. Then, we obtain the following relation for the vector of visibilities, $\widetilde{\mathbf{f}}$

$$\widetilde{\mathbf{f}} = (\overline{\mathbf{A}} \otimes \mathbf{A}) \quad \underbrace{\mathrm{vec}(\mathbf{\Sigma}_{\mathrm{s}} + \mathbf{\Sigma}_{sky})}_{\text{image-plane signals: } \mathbf{f}+\mathbf{n}} \quad . \tag{C.6}$$

Comparing Eq. 5.4 with Eq. C.6, the relation between the array response matrix, $\mathbf{A}$ and the measurement matrix, $\mathbf{T}$ becomes clear, and that is

$$\mathbf{T} = \overline{\mathbf{A}} \otimes \mathbf{A}. \tag{C.7}$$

Moreover, in Sect, 5.3.4, we concluded that the formalism of the LSI deconvolution technique in Eq. 5.28 reduces to Eq. 5.4 assuming a calibrated array i.e. $\mathbf{G}$ in Eq. 5.28 is known and corrected for. If we replace the gain matrix $\mathbf{G}$ with a unity matrix in Eq. 5.28, it will reduce to Eq. C.6.

Bibliography

Balanis, C. A. (2005). *Antenna Theory: Analysis and Design: 3rd edition*. John Wiley & Sons. Inc., Hoboken, New Jersey.

Ben-David, C. and Leshem, A. (2008). Parametric High Resolution Techniques for Radio Astronomical Imaging. *Selected Topics in Signal Processing, IEEE journal of*, 2:670–684.

Berry, R. H., H. M. P. and Withington, S. (2004). Modal decomposition of astronomical images with application to shapelets. *MNRAS*, 354:199–211.

Bhatnagar, S. and Cornwell, T. J. (2004). Scale sensitive deconvolution of interferometric images. I. Adaptive Scale Pixel (Asp) decomposition. *A&A*, 426:747–754.

Boonstra, A.-J. (2005). *Radio Frequency Interference Mitigation in Radio Astronomy*. PhD thesis, Technische Universiteit Delft.

Boonstra, A.-J. and van der Veen, A.-J. (2003). Gain calibration methods for radio telescope arrays. *IEEE Transactions on Signal Processing*, 51:25–38.

Bregman, J. (2012). *System Design and Wide-field Imaging Aspects of Synthesis Arrays with Phased Array Stations*. PhD thesis, Universiteit van Groningen.

Briggs, D. S. (1995). High Fidelity Interferometric Imaging: Robust Weighting and NNLS Deconvolution. In *American Astronomical Society Meeting Abstracts*, volume 27 of *Bulletin of the American Astronomical Society*, page 112.02.

Brinks, E. and Shane, W. W. (1984). A high resolution hydrogen line survey of Messier 31. I Observations and data reduction. *Astronomy & Astrophysics suppl.*, 55:179–251.

Brouw, W. N. (1971). *Data processing for the Westerbork synthesis radio telescope*. PhD thesis, Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands.

Brouw, W. N. (1975). Aperture Synthesis. *Methods in Computational Physics*, 14:131–175.

Bunton, J. D. (2011). Antenna Array Geometries to Reduce the Compute Load in Radio Telescopes. *Antennas and Propagation, IEEE Transactions on*, 59(6):2041–2046.

Candès, E. J. and Romberg, J. (2006). Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. Comput. Math.*, 6(2):227–254.

Candès, E. J. and Romberg, J. (2007). Sparsity and incoherence in compressive sampling. *Journal of Inverse Problems*, 23(3).

Carozzi, T. (2011). Talk in 3GC-II workshop: Radio interferometer calibrability and its limits. Online @ *http://www.astron.nl/meqwiki/3GCII-Workshop/*.

Carozzi, T. D. and Woan, G. (2009). A generalized measurement equation and van Cittert-Zernike theorem for wide-field radio astronomical interferometry. *MNRAS*, 395:1558–1568.

Chambolle, A. and Lions, P. (1997). Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188.

Chang, T. C. and Refregier, A. (2002). Shape Reconstruction and Weak Lensing Measurement with Interferometers: A Shapelet Approach. *Astrophysics Journal*, 570:447–456.

Clark, B. (1973). Curvature of the sky. *Memo 170, NRAO.*

Clark, B. G. (1980). An efficient implementation of the algorithm CLEAN. *A&A*, 89:377–378.

Cornwell, T., Golap, K., and Bhatnagar, S. (2008). The non-coplanar baselines effect in radio interferometry: The w-projection algorithm. *IEEE Journal of Selected Topics in Signal Processing*, 2:647–657.

Cornwell, T. J. (1983). A method of stabilizing the CLEAN algorithm. *A & A*, 121:281–285.

Cornwell, T. J. (2004). Multi-Scale CLEAN deconvolution of radio synthesis images. *IEEE Trans. Sig. Proc.*, 354(199):42–55.

Cornwell, T. J. and Evans, K. F. (1985). A simple maximum entropy deconvolution algorithm. *A&A*, 143:77–83.

Cortès Medellìn, G. (2007). Antenna Noise Temperature Calculation. *SKA Memo 95.*

Cotton, W. D. and Uson, J. M. (2008). Pixelization and dynamic range in radio interferometry. *A&A*, 490:455–460.

Craeye, C., Gonzalez-Ovejero, D., Ghods, N. R., and De Lera Acedo, E. (2012). A projection approach to model the main beam of non-regular arrays in presence of mutual coupling. In *Antennas and Propagation (EUCAP), 2012 6th European Conference on*, pages 2609–2612.

De Lera Acedo, E., Razavi-Ghods, N., Ovejero, D. G., Sarkis, R., and Craeye, C. (2011). Compact representation of the effects of mutual coupling in non-regular arrays devoted to the SKA telescope. In *Electromagnetics in Advanced Applications (ICEAA), 2011 International Conference on*, pages 390–393.

de Vos, M., Gunst, A. W., and Nijboer, R. (2009). The LOFAR Telescope: System Architecture and Signal Processing. *IEEE Proceedings*, 97:1431–1437.

Dewdney, P., bij de Vaate, J.-G., Cloete, K., Gunst, A., Hall, D., McCool, R., Roddis, N., and Turner, W. (2010). SKA Phase 1: Preliminary System Description. *SKA Memo 130*.

Donoho, D. L., Elad, M., and Temlyakov, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *Information Theory, IEEE Transactions on*, 52(1):6– 18.

Fletcher, R. (1980). *Practical Methods of Optimization, Vol. 1, Unconstrained Optimization, and Vol. 2, Constrained Optimization,*. Prentice Hall, John Wiley and Sons.

Franks, L. E. (1969). *Signal Theory*. Prentice Hall, Englewood Cliffs, New Jersey.

Fridman, P. A. and Baan, W. A. (2001). RFI mitigation methods in radio astronomy. *A & A*, 378:327–344.

Frigo, M. (1999). A fast Fourier transform compiler. In *Proceedings of the ACM SIGPLAN 1999 conference on Programming language design and implementation*, PLDI '99, pages 169–180, New York, NY, USA. ACM.

Frigo, M. and Johnson, S. G. (1998). FFTW: an adaptive software architecture for the FFT. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 3, pages 1381–1384.

Frigo, M. and Johnson, S. G. (2005). The Design and Implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231.

Gill, P. E., Murray, W., and Wright, M. H. (1981). *Practical Optimization*. Academic Press, London.

Gull, S. F. and Daniell, G. J. (1978). Image reconstruction from incomplete and noisy data. *Nature*, 272:686–690.

Hale, E. T., Yin, W., and Zhang, Y. (2007). A Fixed-Point Continuation Method for l1-Regularized Minimization with Applications to Compressed Sensing. *Department of Computational and Applied Mathematics Rice University, Houston, Texas, 77005, U.S.A.*

Hall, P. J. (2004). The Square Kilometre Array: An Engineering Perspective. *Experimental Astronomy*, 17(1- 3):5– 16.

Hamaker, J. P. (2000). Understanding radio polarimetry. IV. The full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity. *A&A*, 143:515–534.

Hamaker, J. P. and Bregman, J. D. (1996). Understanding radio polarimetry. III. Interpreting the IAU/IEEE definitions of the Stokes parameters. *A&A*, 117:161–165.

Hamaker, J. P., Bregman, J. D., and Sault, R. J. (1996). Understanding radio polarimetry. I. Mathematical foundations. *A&A*, 117:137–147.

Hamaker, J. P., Osullivan, J. D., and Noordam, J. E. (1977). Image sharpness, Fourier optics, and redundant-spacing interferometry. *Journal of the Optical Society of America (1917-1983)*, 67:1122.

Heidenreich, P., Zoubir, A., and Rubsamen, M. (2012). Joint 2-D DOA Estimation and Phase Calibration for Uniform Rectangular Arrays. *Signal Processing, IEEE Transactions on*, 60(9):4683 –4693.

Heidenreich, P. and Zoubir, A. M. (2011). Gain and phase autocalibration for uniform rectangular arrays. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 2568 –2571.

Hellbourg, G., Weber, R., Capdessus, C., and Boonstra., A.-J. (2012). Cyclostationary approaches for spatial RFI mitigation in radio astronomy. *Comptes Rendus Physique*, 13(1):71 – 79.

Heywood, I. (2011). Talk in 3GC-II workshop: EVLA DDEs and the WHDF at 8.4 GHz. Online @ *http://www.astron.nl/meqwiki/3GCII-Workshop/*.

Högbom, J. A. (1974). Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy and Astrophysics Suppl.*, 15:417–426.

Intema, H. T., van der Tol, S., Cotton, W. D., Cohen, A. S., van Bemmel, I. M., and Röttgering, H. J. A. (2009). Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results. *A & A*, 501:1185–1205.

Ishiguro, M. (1974). Phase Error Correction in Multi-Element Radio Interferometer by Data Processing. *Astronomy & Astrophysics suppl.*, 15:431.

Ivashina, M. (2011). Talk in 3GC-II workshop: An overview of PAF beamforming methods, and a novel beam modeling concept. Online @ *http://www.astron.nl/meqwiki/3GCII-Workshop/*.

Ivashina, M. V., Iupikov, O., Maaskant, R., van Cappellen, W. A., and Oosterloo, T. (2011). An optimal beamforming strategy for wide-field surveys with phased-array-fed reflector antennas. *IEEE Transactions on Antennas and Propagation*, 59:nn.

Jeffs, B. D., Warnick, K. F., Landon, J., Waldron, J., Jones, D., Fisher, J. R., and Norrod, R. D. (2008). Signal Processing for Phased Array Feeds in Radio Astronomical Telescopes. *IEEE Journal of Selected Topics in Signal Processing*, 2:635–646.

Kant, G. W., Patel, P. D., Wijnholds, S. J., Ruiter, M., and van der Wal, E. (2011). EMRBACE: A Multi-Beam 20,000-Element Radio Astronomical Phased Array Antenna Demonstrator. *IEEE Transactions on Antennas and Propagation*, 59:1990–2003.

Kay, S. (1993). *Fundamentals of statistical Signal Processing: Estimation Theory*, volume 1. Prentice Hall, Englewood Cliffs, New Jersey.

Kazemi, S., Yatawatta, S., and Zaroubi, S. (2011). Clustered radio interferometric calibration. In *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pages 597 –600.

Kazemi, S., Yatawatta, S., and Zaroubi, S. (2012). Ordered-Subsets Acceleration of Radio Interferometric Calibration: OS-SAGE calibration algorithm. In *Signal Processing and Information Technology (ISSPIT), 2012 IEEE International Symposium, Ho Chi Minh City- Vietnam*.

Kazemi, S., Yatawatta, S., Zaroubi, S., Lampropoulos, P., de Bruyn, A. G., Koopmans, L. V. E., and Noordam, J. (2011). Radio interferometric calibration using the SAGE algorithm. *MNRAS*, 414:1656–1666.

Lal, D. V., Lobanov, A. P., and Jimnez-Monferrer, S. (2009). Array configuration studies for the Square Kilometre Array- Implementation of figures of merit based on spatial dynamic range. *SKA memo 107*.

Lampropoulos, P. (2010). *The LOFAR epoch of reionization experiment data model*. PhD thesis, Rijksuniversiteit van Groningen.

Landau, H. J. and Pollak, H. O. (1961). Prolate spheroidal wave functions, Fourier analysis, and uncertainty-II. *Bell Syst. Tech. J.*, 40(2):6584.

Lang, M., Guo, H., Odegard, J. E., Burrus, C. S., and Wells, R. O., J. (1996). Noise reduction using an undecimated discrete wavelet transform. *Signal Processing Letters, IEEE*, 3(1):10 –12.

Langer, W. D., Wilson, R. W., and Anderson, C. H. (1993). Hierarchical structure analysis of interstellar clouds using nonorthogonal wavelets. *apj*, 408:L45–L48.

Lawson, C. L. and Hanson, R. J. (1974). *Solving least squares problems*. Prentice Hall, Englewood Cliffs, New Jersey.

Leshem, A. and van der Veen, A. J. (2000). Radio-Astronomical Imaging in the Presence of Strong Radio Interference. *IEEE Trans. on Information Theory, Special issue on information theoric imaging.*, 40(2):1730–1747.

Levanda, R. and Leshem, A. (2008). Radio astronomical image formation using sparse reconstruction techniques. In *Electrical and Electronics Engineers in Israel, 2008. IEEEI 2008. IEEE 25th Convention of*, pages 716 –720.

Levanda, R. and Leshem, A. (2010). Image formation in synthesis radio telescopes. *IEEE Sig. Proc. Magazine*, 27(1):14–29.

Li, Y. and Er, M. H. (2006). Theoretical analyses of gain and phase error calibration with optimal implementation for linear equispaced array. *Signal Processing, IEEE Transactions on*, 54(2):712 – 723.

Lindquist, M. A., Z. C. G. G. S. L. and Yang, Q. X. (2006). A generalization of the two dimentional prolate spheroidal wave function method for nonrectiliner MRI data acquisition methods. *IEEE Trans. on Image Proc.*, 15(9):2792–2804.

Liu, A., Tegmark, M., Morrison, S., Lutomirski, A., and Zaldarriaga, M. (2010). Precision calibration of radio interferometers using redundant baselines. *MNRAS*, 408:1029–1050.

Lonsdale, C. (2004). Calibration Approaches. *Tech. Rep. MIT Low Frequency Program, Memo no.015, MIT Haystack*.

Maaskant, R. (2010). *Analysis of large antenna systems*. PhD thesis, Technische Universiteit Eindhoven.

Maaskant, R. and Ivashina, M. (2012). Characteristic basis function patterns- A novel expansion method for the fast and accurate prediction of antenna array beams. In *Electromagnetics in Advanced Applications (ICEAA), 2012 International Conference on*, pages 796–799.

Maaskant, R., Ivashina, M., Wijnholds, S. J., and Warnick, K. F. (2012). Efficient Prediction of Array Element Patterns Using Physics-Based Expansions and a Single Far-Field Measurement. *Antennas and Propagation, IEEE Transactions on*, 60(8):3614–3621.

Maaskant, R., Tijhuis, A. G., Mittra, R., Ivashina, M. V., van Cappellen, W. A., and Arts, M. J. (2008). Hybridization of Efficient Modeling Techniques for Fast Analysis of Large-Scale Antenna Structures in the Context of Square Kilometre Array Project. *proc. 38th European Microwave Conference, Amsterdam*, pages 837–840.

Maaskant, R. and Yang, B. (2006). A Combined Electromagnetic and Microwave Antenna System Simulator for Radio Astronomy. In *The European Conference on Antennas and Propagation: EuCAP 2006*, volume 626 of *ESA Special Publication*.

Maisinger, K., Hobson, M. P., and Lasenby, A. N. (2004). Maximum-entropy image reconstruction using wavelets. *MNRAS*, 347:339–354.

Manikas, A. and Fistas, N. (1994). Modeling and estimation of mutual coupling between array elements. *Proc. ICASSP*, pages 553–556.

McEwen, J. D. and Wiaux, Y. (2011a). Compressed sensing for radio interferometric imaging: review and future direction. *ArXiv e-prints*.

McEwen, J. D. and Wiaux, Y. (2011b). Compressed sensing for wide-field radio interferometric imaging. *MNRAS*, 413:1318–1332.

McEwen, J. D. and Wiaux, Y. (2011c). Intrinsic advantages of the w component and spherical imaging for wide-field radio interferometry. *URSI 2011*.

Millenaar, R. P. and Bolton, R. C. (2010). Figures of Merit for SKA Configuration Analysis. *SPDO reports: WP3-050.020.000-TR-001*.

Noordam, J. E. (2004). LOFAR calibration challenges. In Oschmann, Jr., J. M., editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5489 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 817–825.

Noordam, J. E. and de Bruyn, A. G. (1982). High dynamic range mapping of strong radio sources, with application to 3C84. *nature*, 299:597–+.

Noordam, J. E. and Smirnov, O. M. (2010). The MeqTrees software system and its use for third-generation calibration of radio interferometers. *A&A*, 524:A61.

Noorishad, P., Wijnholds, S. J., van Ardenne, A., and van der Hulst, J. M. (2012). Redundancy calibration in phased-array stations. *A&A*, 545:A108.

Noorishad, P. and Yatawatta, S. (2011). Efficient computation of prolate spheroidal wave functions in radio astronomical source modeling. In *Signal Processing and Information Technology (ISSPIT), 2011 IEEE International Symposium on*, pages 326 –330.

Offringa, A. R. (2012). *Algorithms for radio interference detection and removal*. PhD thesis, Rijksuniversiteit van Groningen.

Offringa, A. R., de Bruyn, A. G., Biehl, M., Zaroubi, S., Bernardi, G., and Pandey, V. N. (2010). Post-correlation radio frequency interference classification methods. *MNRAS*, 405:155–167.

Pandey, V. N., van Zwieten, J. E., de Bruyn, A. G., and Nijboer, R. (2008). LOFAR calibration with BBS. In *The Low-Frequency Radio Universe, NCRATIFR, Pune-India*, ESA Special Publication.

Pantin, E. and Starck, J. L. (1996). Deconvolution of astronomical images using the multiscale maximum entropy method. *A&A*, 118:575–585.

Percival, D. B. and Walden, A. T. (1993). *Spectral analysis for physical applications: multitaper and conventional univariate techniques*. Cambridge University Press, New York.

Perley, R. A., Schwab, F. R., and Bridle, A. H. (1985). Synthesis Imaging. In *Course notes from an NRAO summer school, held in Socorro, New Mexico*, volume 3.

Perley, R. A. S. F. R. and Bridle, A. H. (1994). Wide Field Imaging II: Imaging with Non-Coplanar Arrays, in Synthesis Imaging in Radio Astronomy. *Astronomical Society of the Pacific Conference Series*, 6:139– 165.

Persson, P. and Strang, G. (2004). A Simple Mesh Generator in MATLAB. *SIAM rev.*, 46(2).

Prakash, V. and Mittra, R. (2003). Characteristic basis function method: A new technique for efficient solution of method of moments matrix equations. *Micr. Opt. Technol*, 36:95–100.

Refregier, A. (2003). Shapelets-I. A method for image analysis. *MNRAS*, 338:35–47.

Rudelson, M. and Vershynin, R. (2006). Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *Information Sciences and Systems, 2006 40th Annual Conference on*, pages 207–212.

Ryle, M. (1962). The New Cambridge Radio Telescope. *Nature*, 194:517–518.

Sault, R. J., Hamaker, J. P., and Bregman, J. D. (1996). Understanding radio polarimetry. II. Instrumental calibration of an interferometer array. *A&A*, 117:149–159.

Schwab, R. P. F. and Bridle, A. H. (1994). Synthesis imaging in radio astronomy. *in proc. Astron. Soc. Pacific Conf. Series*, 117.

Simons, F. J., Dahlen, F. A., and Wieczorek, M. A. (2006). Spatiospectral concentration on a sphere. *SIAM rev.*, 48:504– 536.

Slepian, D. (1976). On bandwidth. *Proc. of the IEEE*, 64(3):292–300.

Slepian, D., P. H. O. (1961). Prolate spheroidal wave functions, Fourier analysis, and uncertainty-I. *Bell Syst. Tech. J.*, 40(2):43–61.

Smirnov, O. M. (2011a). Revisiting the radio interferometer measurement equation I. A full-sky Jones formalism. *A&A, Astronomical instrumentation*, 527.

Smirnov, O. M. (2011b). Revisiting the radio interferometer measurement equation II. Calibration and direction-dependent effects. *A&A, Astronomical instrumentation*, 527.

Smirnov, O. M. (2011c). Revisiting the radio interferometer measurement equation III. Addressing direction-dependent effects in 21 cm WSRT observations of 3C 147. *A&A, Astronomical instrumentation*, 527.

Stoica, P. and Ng, B. C. (1998). On the Cramer-Rao bound under parametric constraints. *IEEE Signal Processing Letters*, 5:177–179.

Svantesson, T. (1998). The effects of mutual coupling using a linear array of thin dipoles of finite length. In *Proceedings of 9th IEEE SP Workshop on Statistical Signal and Array Processing*, pages 232 –235.

Thompson, A., Moran, J., and Swenson, G. (1986). *Interferometry and Synthesis in Radio Astronomy*. Wiley, New York.

van Ardenne, A., Bregman, J. D., van Cappellen, W. A., Kant, G. W., and de Vaate, J. G. B. (2009). Extending the Field of View With Phased Array Techniques: Results of European SKA Research. *IEEE Proceedings*, 97:1531–1542.

van Ardenne, A., Wilkinson, P. N., Patel, P. D., and Vaate, J. G. B. (2004). Electronic Multi-beam Radio Astronomy Concept: Embrace a Demonstrator for the European SKA Program. *Experimental Astronomy*, 17:65–77.

van Cappellen, W. A., Bregman, J. D., and Arts, M. J. (2004). Effective Sensitivity of A Non-uniform Phased Array of Short Dipoles. *Experimental Astronomy*, 17:101–109.

van der Tol, S. (2009). *Bayesian Estimation for Ionosphere Calibration in Radio Astronomy*. PhD thesis, Technische Universiteit Delft.

van der Tol, S., Jeffs, B. D., and van der Veen, A. J. (2007). Self-Calibration for the LOFAR Radio Astronomical Array. *IEEE Transactions on Signal Processing*, 55(9).

van der Tol, S. and Wijnholds, S. J. (2006). CRB Analysis of the Impact of Unknown Receiver Noise in Phased Array Calibration. In *4th IEEE workshop on Sensor Array and Multichannel Processing (SAM)*.

van der Veen, A.-J., Leshem, A., and Boonstra, A.-J. (2004). Signal processing for radio astronomical arrays. In *in IEEE workshop on Sensor Array and Multichannel*.

Wakker, B. P. and Schwarz, U. J. (1988). The Multi-Resolution CLEAN and its application to the short-spacing problem in interferometry. *A&A*, 200:312–322.

Wei, L., K. R. A. and Lamahewa, T. A. (2010). An optimal basis of band-limited functions for signal analysis and design. *IEEE Trans. on Signal Proc.*, 58(11):5744–5755.

Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., and Vandergheynst, P. (2009a). Compressed sensing imaging techniques for radio interferometry. *MNRAS*, 395:1733–1742.

Wiaux, Y., Puy, G., Boursier, Y., and Vandergheynst, P. (2009b). Compressed sensing for radio interferometry: spread spectrum imaging techniques. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7446 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.

Wiaux, Y., Puy, G., Boursier, Y., and Vandergheynst, P. (2009c). Spread spectrum for imaging techniques in radio interferometry. *MNRAS*, 400:1029–1038.

Wieringa, M. (1991). *327 MHz Studies of The High Redshift Universe And The Galactic Foreground*. PhD thesis, Rijksuniversiteit te Leiden.

Wijnholds, S. and Boonstra, A.-J. (2006). A Multisource Calibration Method for Phased Array Radio Telescopes. In *Sensor Array and Multichannel Processing, 2006. Fourth IEEE Workshop on*, pages 200 –204.

Wijnholds, S. J. (2008). Autonomous Online LOFAR Station Calibration. In *URSI GA 2008, Chicago (Illinois, USA)*.

Wijnholds, S. J. (2010). *Fish-eye observing with phased array radio telescopes*. PhD thesis, de Technosche Universiteit Delft.

Wijnholds, S. J. (2012). LOFAR Configuration Considerations as a Design Exercise for SKA. *SKA memo 143*.

Wijnholds, S. J. and Noorishad, P. (2012). Statistically optimal self-calibration of regular imaging arrays. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1304 –1308.

Wijnholds, S. J. and van Cappellen, W. A. (2011). In Situ Antenna Performance Evaluation of the LOFAR Phased Array Radio Telescope. *IEEE Transactions on Antennas and Propagation*, 59:1981–1989.

Wijnholds, S. J., van der Tol, S., Nijboer, R., and van der Veen, A.-J. (2010). Calibration challenges for future radio telescopes. *IEEE Signal Processing Magazine*, 27:30–42.

Wijnholds, S. J. and van der Veen, A.-J. (2008). Fundamental Imaging Limits of Radio Telescope Arrays. *IEEE Journal of Selected Topics in Signal Processing*, 2:613–623.

Wijnholds, S. J. and van der Veen, A.-J. (2009a). Multisource Self-Calibration for Sensor Arrays. *IEEE Transactions on Signal Processing*, 57:3512–3522.

Wijnholds, S. J. and van der Veen, A.-J. (2009b). Self-Calibration of Radio Astronomical Arrays With Non-Diagonal Noise Covariance Matrix. *17th European Signal Processing Conference, Glasgow- the UK*, pages 24–28.

Wright, M. C. H. (1999). Image fidelity. *BIMA Memo 73*.

Yatawatta, S. (2010). Fundamental Limitations of Pixel Based Image Deconvolution in Radio Astronomy. *in proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Jerusalem, Israel*, pages 69–72.

Yatawatta, S. (2011a). Radio Astronomical Image Deconvolution Using Prolate Spheroidal Wave Functions. *in proc. IEEE International Conference on Image Processing (ICIP), Brussels, Belgium*, pages 2813–2816.

Yatawatta, S. (2011b). Shapelets and related techniques in radio astronomical imaging. *in proc. URSI GA 2011, Istanbul, Turkey*.

Yatawatta, S., Zaroubi, S., de Bruyn, G., Koopmans, L., and Noordam, J. (2008). Radio Interferometric Calibration Using The SAGE Algorithm. *ArXiv e-prints*.

Yeo, J., Prakash, V., and Mittra, R. (2003). Efficient analysis of a class of microstrip antennas using the charactreistic basis function method (CBFM). *Micr. Opt. Technol*, 39:456–464.

Zatman, M. (1998). How narrow is narrowband? *IEE Proceedings on Radar, Sonar and Navigation*, 145(2):85 –91.

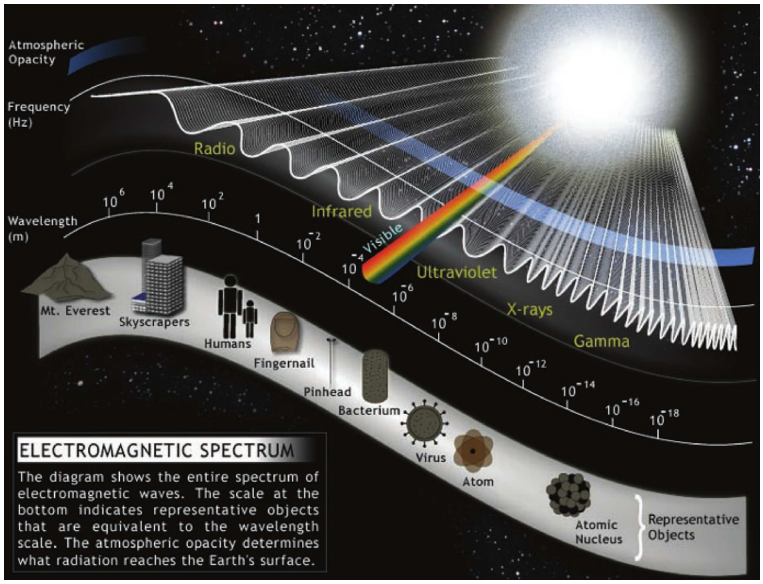
Nederlandse Samenvatting

Het zijn grote vragen die sterrenkundigen bezighouden: “Hoe heeft het huidige Heelal zich gevormd uit het Heelal van vlak na de Oerknal?”, “Hoe ontstaan sterrenstelsels?”, “Hoe worden sterren geboren, en hoe gaan ze weer dood?”, “Hoe is het Zonnestelsel ontstaan?”, “Wat is de wisselwerking tussen sterrenstelsels onderling?”, en “Zijn mensen op Aarde de enigen die zich deze vragen stellen?”. Deze en vele andere sterrenkundige en kosmologische vragen hebben geleid tot grote technologische vooruitgang in zowel het bouwen van verschillende instrumenten, alsook de algoritmes waarmee wij de hiermee vergaarde gegevens efficiënt kunnen verwerken. Hierdoor verleggen we voortdurend de grenzen van ons begrip van het Heelal waarin we leven.

Waarnemingen

Hemellichamen zoals sterrenstelsels stralen voortdurend elektromagnetische golven (licht) uit, over een breed spectrum. Deze golven reizen alle kanten op het Heelal in, waaronder de onze, richting Aarde. Om deze hemellichamen beter te begrijpen, verzamelen we deze golven met de verschillende instrumenten die tot onze beschikking staan. Deze informatie verwerken en bestuderen we aan de hand van zo efficiënt mogelijke algoritmes. Het elektromagnetische spectrum is gellustreerd in figuur C.1, van de langste golflengtes die kennen we in de vorm van radiogolven ($\sim 10^4$ Hz) tot gammastraling ($\sim 10^{20}$ Hz).

De verschillende frequenties licht zijn gerelateerd aan verschillende fysische of chemische processen die zich in deze objecten plaats vinden. Wetenschappers willen deze verschijnselen helemaal leren begrijpen; hoe bijvoorbeeld het ene fenomeen terugkoppeling heeft op het andere. Het is hiervoor dus van belang de objecten in alle verschillende frequenties waar te nemen. De bijbehorende technologie kan echter sterk verschillen per golflengtegebied. Om bijvoorbeeld röntgenstraling waar te nemen, moeten we een telescoop de ruimte insturen, omdat de atmosfeer de meeste rntgenstraling, en diens schadelijke neveneffecten, tegenhoudt. Sterrenkunde in zichtbaar licht kan alleen bedreven worden op plaatsen met mooi weer, omdat wolken al het zichtbare licht verstrooien. Radiogolven boven de 10 MHz hebben geen last van



Figuur C.1: Het diagram toont het hele spectrum der elektromagnetische golven. De onderste schaal toont objecten van een grote die vergelijkbaar is met de bijbehorende golflengte, die daarboven is aangegeven. De bovenste blauwe lijn toont de doorlaatbaarheid van de atmosfeer voor deze golven; dus welk percentage van de straling het aardoppervlak bereikt. Bron: `http://ds9.ssl.berkeley.edu/LWS_gems/2/espec.htm`, ©UC Regents.

wolken en bewegen zich vrij door de atmosfeer. Hierdoor is het ook mogelijk deze waar te nemen vanaf de grond, en in het bijzonder ook vanuit het bewolkte Nederland, zoals bij de telescopen in Westerbork (WSRT) gebeurt.

Radiosterrenkunde

De WSRT is een hele rij van schotelantennes. Elke schotel reflecteert radiogolven naar zijn brandpunt, waar een ontvanger is geïnstalleerd. De techniek om de zo ontvangen signalen van alle veertien schotels te combineren tot een afbeelding heet *interferometrie*. Door deze techniek kunnen we profiteren van het grotere totale oppervlak van alle spiegels samen, en van een sterk verbeterde resolutie. Hoe groter het ontvangstoppervlak, des te groter wordt de gevoeligheid van het instrument, en kunnen we zwakkere objecten waarnemen. De grote afstand van de schotels onderling maakt het mogelijk afbeeldingen te *synthetiseren* van een hoge resolutie. De afstand tussen twee ontvangers wordt ook wel *basislijn* genoemd. Elke combinatie van twee ontvangers in de opstelling vormen een basislijn van de interferometer.

LOFAR

Om dus betere afbeeldingen te maken van sterrenstelsels, hebben we altijd grotere en gevoeligere telescopen nodig. Om zulke grote observatoria te bouwen met schotelantennes is een dure onderneming en de constructie vergt heel veel tijd. ASTRON, het Nederlandse instituut voor radiosterrenkunde, heeft LOFAR ontwikkeld en gerealiseerd; een instrument met een onconventioneel ontwerp: *fasegestuurde antennes*. Dit ontwerp maakt het een stuk goedkoper en sneller om een grote telescoop op te bouwen. Ook zijn de wetenschappelijke toepassingen legio. Eén van de voordelen is dat LOFAR meerdere delen van de hemel in één keer kan waarnemen, doordat de kijkrichting van een fasegestuurde antenne elektronisch bepaald wordt. Deze eigenschap versneld het verzamelen van wetenschappelijke data op een manier die niet mogelijk is met conventionele technieken.

Het mechanische deel van LOFAR heeft een heel eenvoudig ontwerp, het bestaat uit duizenden simpele dipoolantennes. Dit moeten we bekopen in de data-analyse. Het onconventionele ontwerp van LOFAR betekent ook dat we nieuwe methodes moeten ontwikkelen voor het analyseren en begrijpen van de waarneemgegevens, en deze te verwerken tot wetenschappelijke resultaten. Met dit oogpunt, zijn er verschillende studies verricht door ASTRON in nauwe samenwerking met Nederlandse universiteiten, waaronder die in Groningen, Leiden en Delft. LOFAR heeft alleen een succes kunnen worden dankzij deze samenwerking. We hebben fantastische prestaties laten zien in het ontwikkelen en toepassen van vernieuwende methodes in signaal analyse, om zo uitmuntende wetenschappelijke resultaten te kunnen presenteren aan de hand van LOFAR waarnemingen. Dit proefschrift heeft bijgedragen in deze prestaties.

Figuur C.2: Vierentwintigtegelige HBA opstelling: Elke tegel is een element van een HBA station. Deze opstelling wordt gebruikt in veruit de meeste HBA stations die LOFAR rijk is, inclusief de stations op de *superterp*, het centrum van LOFAR nabij Exloo. Een opstelling als deze telt veel *redundante* basislijnen.

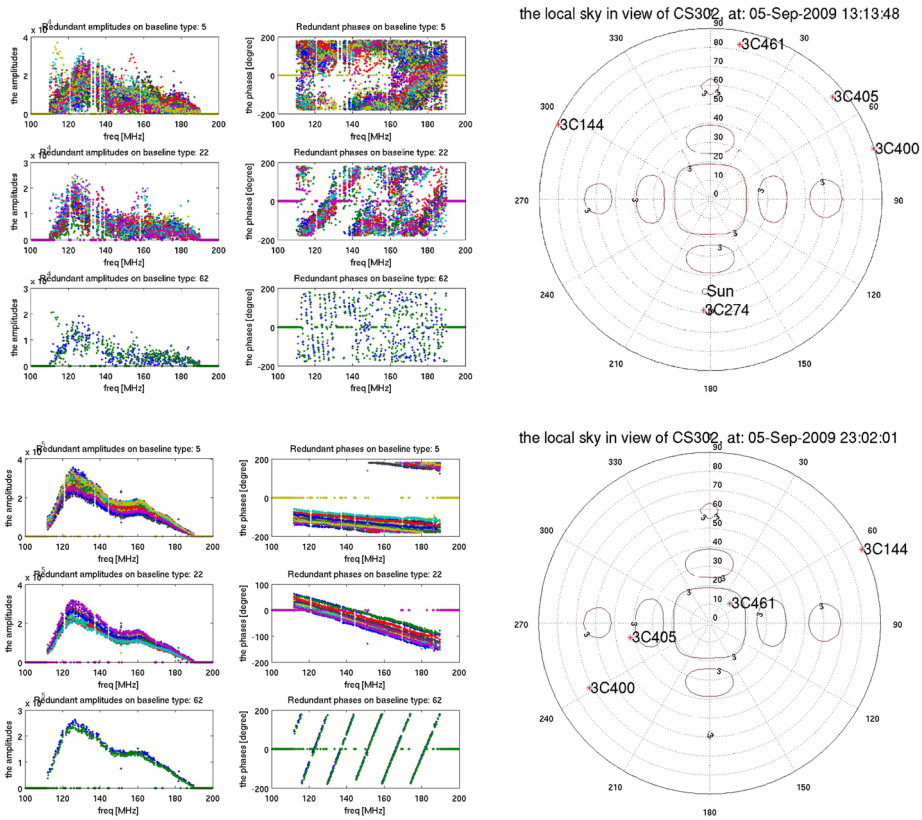
Redundantie

We geven een voorbeeld uit dit proefschrift, dat geholpen heeft LOFAR data beter te begrijpen. De HBA (High Band Antenna) stations vormen een onderdeel van LOFAR dat gericht is op het hogere deel van het frequentieberijk van LOFAR. Ze staan in een regelmatige opstelling. In figuur C.2 zien we het HBA station met aanduiding CS302. Deze opstelling leidt er toe dat veel van de combinaties van antennes precies dezelfde onderlinge afstand en oriëntatie hebben. Deze basislijnen zijn in feite overbodig, ofwel: *redundant*. De gemeten data op een basislijn is de *correlatie* tussen de twee signalen die tegels ontvangen. Van redundante basislijnen krijgen we deze informatie dus dubbel. Deze dubbele informatie wordt niet weggegooid, maar helpt ons om de antennes aan elkaar te ijken. Dit werkt alleen onder de aanname dat de betrokken antennetegels identiek zijn. Omdat de tegels zo dicht op elkaar zitten, zullen ze elkaar enigszins beïnvloeden. Hierdoor gaat deze aanname niet op. Het signaal dat we ontvangen van verschillende redundante basislijnen zal dus niet helemaal identiek zijn. We hebben de HBA tegels gesimuleerd met het programma CAESAR. De resultaten van deze simulaties lieten zien dat de verschillende tegels in de opstelling elkaar inderdaad elektromagnetisch beïnvloeden, zodat hun ontvangstpatronen van elkaar verschillen.

Zoals ook in figuur C.3 te zien is, heeft het ontvangstpatroon een aantal *zijlobben*. Dit betekend dat we de antennes niet perfect op één plek kunnen richten. Het te meten signaal zal altijd 'besmet' worden door objecten die zich in de zijlobben bevinden. Het is in deze zijlobben dat we de grootste verschillen vinden tussen 'ideale' redundantie en het echte instrument. Als een signaal dus afkomstig is uit een van de zijlobben, meten we geen redundantie, waar dat wel het geval is als het signaal afkomstig is uit de hoofdlob. De grafieken in figuur C.3 bevestigen deze voorspelling. Rechts zien we een projectie van de hemel met het ontvangstpatroon van de HBA en een aantal heldere bronnen die we op een gegeven moment kunnen waarnemen. Links zien we de overeenkomstige data die we met een HBA station kunnen meten, op drie verschillende redundante basislijnen. Dit laat zien dat de aanwezigheid van slechts één enkele heldere bron in een van de zijlobben de redundante data aanzienlijk verstoort (vergelijk onder met boven). Ook zien we dat een sterke bron waargenomen in de hoofdlob juist redundante informatie oplevert.

Conclusie

Dit soort studies zijn belangrijk om de IJkbaarheid van een fasegestuurde antenne stations, en dan in het bijzonder die met een regelmatige opstelling (zoals HBA's) te begrijpen. IJkbaarheid van deze instrumenten is een belangrijke stap richting het kunnen synthetiseren van radio-afbeeldingen van hoge kwaliteit. In dit proefschrift hebben we nog meer van dit soort theoretische technieken bestudeerd om dit doel te bereiken. Deze studies leggen ook een belangrijke basis voor verder onderzoek in deze richting.



Figuur C.3: Twee waarnemingen die zijn uitgevoerd met een HBA station op 5 september 2009. De eerste om 13:13:48 UTC, waar zgn. redundantie-ijking mislukt (boven), de tweede met hetzelfde station op de zelfde dag, om 23:02:01 UTC waar de redundantie-ijking slaagt (onder). Links zien we de gemeten signalen voor drie verschillende redundante basislijnen. Rechts is een projectie van de hemel ten tijde van de waarneming afgebeeld. We zien de helderste bronnen aangeduid, alsmede een indicatie van de hoofd- en zijlobben in het ontvangspatroon.

Toekomst

Het succes van LOFAR heeft er toe geleid dat het ontwerp van LOFAR ook overgenomen wordt in delen van de nieuw te bouwen *Square Kilometre Array*, 's Werelds grootste en gevoeligste toekomstige radiotelescoop. SKA wordt ontworpen en gebouwd door een internationaal consortium van instituten waar ook Nederland een grote bijdrage aan levert. Dit maken zowel ASTRON als betrokken Nederlandse universiteiten tot een zeer boeiende werkomgeving.

Acknowledgements

This has surely been an amazing journey, through which I met many differently interesting people and grew more mature by their love, support, and encouragements and sometimes the opposite of these. I am very grateful to every single person that I met during this period, some of whom are explicitly thanked here. Thank you for having been part of this!

My very special thanks goes to Prof. Arnold van Ardenne, whom I met when I was still a master student. My fascination about the unthinkable idea of civilization on the Moon took me to the international symposium of *To Moon and beyond* in March 2007, where I learned about the project of *LOFAR on the Moon*. That introduced me to the world of ASTRON and to Prof. Arnold van Ardenne. He convinced me to work on *LOFAR on the Earth* that was being constructed by ASTRON, which led to two-thirds of the world largest radio telescope, the SKA. The project was novel and challenging and propelled my interest in being part of it. Later, I realized that the idea of civilization on the Moon is a bad idea before we learn how to sustainably civilize on Earth. However, using the Moon as an observational platform is still a great idea. Now, I am very pleased with the decision that opened up even greater opportunities later in my career life. Thanks Arnold! You have a big heart and have also greatly supported me during the past years.

My sincere thanks to Prof. Thijs van der Hulst, my co-promoter who taught me how to be an independent researcher by often challenging me. Thanks Thijs! also for your immense support of me during the past years. The last stage of my thesis preparation was very hectic. Thanks for your quick reactions on my corrections and being available for any question at any time.

Thanks to both of you also for being flexible, when my gipsy-self kicked in and had to go to somewhere new very spontaneously for a break.

Truus van den Brink-Havinga, thanks for doing the not-so-easy job of scheduling

meetings for Thijs, Arnold and me. Your patience and charisma are respectable.

Many thanks to Prof. Ger de Bruyn and Prof. Saleem Zaroubi for accepting to be in the reading committee of my thesis and for your attentive reading of it. Your valuable comments improved the final version of this thesis significantly. Thank you!

Many thanks Oleg! from your *House of Mojito* to being on my reading committee. Thanks for being such a cool party host among the many friendly people whom I met first, when I came to the Netherlands. Thanks also for reading my thesis carefully and for your very constructive comments.

Many thanks to the members of the Promotie Begeleidings Commissie (PBC), for helping me to finish this thesis in spite of various difficulties that I faced in this period.

Jackie, Hennie, Lucia, Christa and Gineke, there is no need to mention that none of us could have done it without you and your patience. You made all seemingly impossible administrative work seem like just a piece of cake. Thank you very much!

Marco van der Vinne, from the International Serive Desk (ISD), thanks for your wholehearted reception of me anytime that I needed your help.

Many thanks to the computer group at Kapteyn for providing such robust IT facilities and being available for any question. Special thanks to Wim, whom I often disturbed with my questions. Thanks also for being such a caring friend.

Many thanks to Prof. dr. Edwin Valentijn, who recognized me as an addition to his uniquely successful group. You supported me in becoming a member of two outstanding groups. Thank you!

I was blessed to have shared an office with three most gentle Dutch men during most of my PhD time: Hugo, Johan, and Niels, who provided very positive vibes in the office. I am proud of you for your modest personalities and your beautiful scientific work. You have a great capabilities for understanding not only various scientific areas but also people. These are great qualities.

Hugo, you are a special soul with no doubt in my heart. In your view of the world, everybody counts. The world needs more people like you. Your generosity and enthusiasm in sharing your knowledge with others and giving opportunities to everybody are commendable. You always came up with more than one answer to my questions, each better than the earlier one over 5-10 minutes;) I am indeed very pleased that I will be working closely with you soon.

Johan, you are one of the most talented and knowledgeable people that I have met. I always enjoyed talking to you and learning from you. You are man of a massive heart and integrity. You helped me with many things including occasional Dutch or math

lectures and doing the Dutch translation of my summary. Thanks also for sharing your chocolates. I also admire your super organized mind, being able to manage your stuff in the pile of papers, food, scientific/non-scientific books, cutlery, coffee machine, clothes, labtop, desktop PC, and several other electronic devices on your desk;)

Niels, I remember the first day that you came to the office and looked around thinking of a quick redecoration. We did it in less than an hour, turning our office into its nicest shape. Thanks for doing that! You are a gentle creative person, who helps almost everybody at the institute. I am very proud of you. Thanks for your warm greeting every morning! Many thanks also for your dedication in making a cool cover design for my thesis. It looks great. Well done!:)

Johan and Niels, thanks also for taking my stressed mood, especially late in my PhD period and working hard for me, the last weekend before my thesis goes to the printer. I felt bad that you had to work for me in your weekend, but you both made me feel just fine. Thank you!

Jackob and Matìas, my two other office-mates, whose presence did not overlap long enough with mine for us to get to know each other better. I wish you all the best in whatever you do, wherever you are now.

Thanks to many good friends that I made over the past years who certainly created a caring global family for me, and whom I truly adore.

Stefania, you are a beautiful person. I think this is self-explanatory. I will never forget the first time that I met you at Kapteyn, when you were a visitor for only 3 days. I could feel the halo of positivity and kindness around you. I knew you would make a good friend. I approached you to show some hospitality to a person who has just arrived in Groningen and to invite you for something. I was astonished when you offered me to join you for a contemporary dance performance in the evening of the next day in Groningen. Then, I knew that you would make a wonderful friend. I simply cannot say how delighted I am to have you as a close friend. I learned a lot from you. Thanks also for taking all my moods over the last year that we shared a flat together. Thanks! Thanks for being you!

Andrea, you are such a calm and gracious person. I am not surprised that you are Stefania's boyfriend. Thanks for making us the best chicken lemon meal.

Esra, thanks for making me feel like a good friend. You are one yourself. You are a considerate person who goes out of her way to make others feel right. Your way of frankness is so cute that it always puts smile on my face. You are a special person with so many capabilities. I am very proud you, Esra. You did a thesis work that not that many people would dare to do. This also explains a lot about you as a strong person. I am very happy for you for being gifted the world coolest baby, William-Kayra next to a peaceful man like Dave.

Giuliana and Piero, you are both such understanding friends. I enjoyed your company and your Italian meals. I will visit you in Italy some time. Promise!

Linda, my special friend! Thanks for making us think about various political, humanitarian, and even philosophical issues almost every time we met. I learned a lot from you. Thank you! You often criticized me for not being crazy enough. I wonder if anybody would meet your standard at all, maybe only Pratyush;)

Pratyush, the brave soul! I always enjoyed talking to you. I remember telling you that you remind me of *highway man* song by Johnny Cash. I think everybody who knows you would agree with me. We all are a *highway man* but we just do not feel it as much as you do.

Burcu, thanks for inviting me to watch very good movies together and for sharing your nice cookery experiences. You are really good at it:)

Stephanie, you are truly a role model for many of us. I admire you for many reasons, being a fantastic mom, a successful scientist, a true friend, a wonderful person, a dedicated supervisor. I would be out of words to describe you. Thanks also for preparing me for my most important job interview. You made me as nervous as I would be in a real one;)

Jan, thanks for being available as a friend at any time. Thanks for helping me with my best move in Groningen.

Federico, the brain of all the party brainstorming sessions! You always came up with the best ideas and things to say. You make a good public speaker with a good sense of humor. Thanks for being such a sweet friend!

Home is where you feel at home. Wieke, you certainly made me feel at home in Groningen when you sent me a long comforting letter of condolences, a much better version of the beautiful Dutch tradition of sending cards. Thanks again!

Annemieke Janssen, I will never forget the day that you came to me with a wide smile and gave me the *Jip and Janneke* book to help me learn Dutch. It was so kind of you! Now, I can almost say that I speak Dutch. Thank you!

Andrè Offringa, Keimpe, and Leon, thanks for introducing me to *blue toes*, the student dance club, which I enjoyed very much.

Sanaz, you are a smart strong woman. I am very happy to see your advancements over the past years. Thanks for sharing a lot of laughter.

Oscar, thanks for the guitar and computer lessons that you gave me. You also generously helped me prepare for the job that you yourself wanted very much. This means a lot to me and explains a lot about you.

Sarod, I believe in you. You are indeed a man of great heart and mind. Wish you all the best for your further advancements. Thanks for being such a tough but compassionate teacher.

Pandey, thanks for your useful hints for how to efficiently incorporate comments in the final version of my thesis manuscript.

Maarten, thanks for helping me with some LaTeX editorial problems, when I was very short in time before I send this thesis to the printer.

Shoko thanks for kindly and carefully correcting my English and letter style for my job-applications. Thanks also for dropping by my office and caring how I was doing when I was very busy doing many things at the same time. It is so kind of you.

There are very special people that one does not need to know for a long time to feel so close to them. Annemarie, you are certainly one of those. Your way of embracing multiculturalism is admirable. I am happy to have met an intelligent and special woman like you.

Bruno, you are another such person. I remember the first time that I met you. The more we talked the happier I became for Silvia. You are an amazing energetic person who feels life and lives a good one. Silvia, thanks for your warmth and also for the delicious Portuguese Bacalhau that you cooked for us.

Aleksandar, you are a funny guy in a special way. Thanks for answering my occasional questions.

Koshy, thanks for your regular interruptions and distractions in the period when I was the most asocial person at the institute because of my thesis preparation and job hunting at the same time. Wish you all the best!

Rowin and Paolo Soleri, thanks for giving me job advice and introducing me to the new possibilities that you came across.

Guobao and Yanping, I admire your courtesy. Thanks for your friendly greeting everyday. Guobao, I will not forget your valuable advice for the defense day. I will certainly put that to practice. For other PhD students who are reading this, his advice was to *enjoy it*. Wish you both all the best in pursuing your scientific career next to each other.

Ivona, you are easy-going and super smart. No, maybe you are easy-going and the

latter comes as its byproduct;). However, you are awesome! I should learn more from you.

Gert, thanks for inviting me to your parties.

Dan, the unique man of heart! I learned a lot from you and with you in a very peculiar period of my life. Wish you all the best wherever you are now.

Durba, thanks for showing me where the confidence takes one in life. Thanks for reminding me of what life is worthy of.

Thanks to Prof. dr. Bob Sanders, Prof. dr. Hugo van Woerden, Henk from the canteen, Elif, Seyit, Stefano, Giamba, Yunhee, Kyle, Chiara, Evgenia, Tjitske, Jeffrey, Rosina, Teresa, Hans, Maarten, Germán, Marius, Harish, Alexandros, Thomas Martinson, Andrea and Giacomo for your smile and warm greeting everyday. That makes a big difference in the atmosphere of a working place, where one spends his/her most efficient period of his/her day. Thank you!

Thanks to all of my sport buddies and trainers at the ACLO sport center. Without you, taking the intensity of the mental workload of the PhD would have been impossible.

Thanks to my buddies from ASTRON whom I also shared an office with, Raj and Greg. Raj, the man of moments! Thanks for being such a gracious selfless friend and for being available any time that I needed your help. Thanks to you and your hospitable house-mates, who always welcomed me for an unplanned stay in Delft. Thanks also for listening to my complaints patiently. Gerg, you were the coolest visitor over the past years that I visited ASTRON. I am sure anybody else at ASTRON would admit that too. You made a music group at ASTRON whose performances we began to enjoy in occasional parties in Dwingeloo. Thank you!

Thanks to Giuseppe, one of the first friendly and interesting people that I met at ASTRON and JIVE. You introduced me to the world of Couchsurfing, via which I still keep meeting so many amazing people, some of whom have become my good friends. Thanks also for giving me a ride to ASTRON regularly and excusing me for my Iranian punctuality. Many thanks to Stephen, Valeriu and Mehreen for making the daily trips to Dwingeloo more enjoyable with interesting discussions about almost anything.

Jelly Kuiters, thanks for being my first Dutch lady friend who helped me a lot when I was very new to life in Groningen.

Daniela, thanks for being such an interesting friend. I enjoyed your company in everything that we did together, from movies and festivals to foods, traveling, and sport. I learned a lot from you.

Many thanks to Dusko and Jelena, whose serenity and wisdom have taught me a lot. Thanks for embracing me genuinely like a family, caring for me and my advancements in life.

Monique, thanks for your friendship. Sorry for being too occupied with my thesis to be available for you most of the time.

Marianna Ivashina, Rob Maaskant, and Michel Arts from ASTRON, thank you for answering all my antenna and electromagnetic questions with patience and care. Thanks for collaborating with me to make a very useful point for understanding the phased-array station calibration.

Thanks to many people form ASTRON and JIVE who have created a scientifically vibrant and socially welcoming environment in Dwingeloo: Mike Garrett, Rafaella Morganti, Tom Oosterloo, Marco de Vos, Albert-Jan Boonstra, Jaap Bregman, Johan Hamaker, Wim Brouw, Jan Noordam, Johanne Noordam, Stefan Wijnholds, Harm Munk, Teun Grit, Antonis Polatidis, Benedetta Fiorelli, John McKean, Paolo Serra, Vibor Jelic, George Heald, Mike Sipior, Maura Pilia, Carmen Toribio, Megan Argo, Jason Hessel, Joeri van Leeuwen, Elizabeth Mahony, Alicia Berciano, Roberto Pizzo, Rob Millenaar, Menno Norden, Michiel Brentjens, Ashish Asgekar, Koos Kegel, Dion Kant, Eric Kooistra, André Kunst, Majid Naeem, Nikta Amiri, Dmitry Duev, Tatiana Bocanegra, Janneke Wubs-Komdeur, Andre van Es, Ilse van Bemmel, Jan Geralt Bij de Vaate, Saswata Bhaumik, Adriaan Renting, Arno Schoenmakers, Harro Verkouter, Zsolt Paragi, Lars Baehren, Erika Timmerman, Diana Verweij, Carin Lubbers, radio observatory staff, Frits Möller, Mrs. reception and many others.

Thanks to Serveh and Farshad, my sunshine friends in Sweden, whose continuous support and care kept me moving forward over the past few years in the Netherlands. Thank you!

Thanks to Christina Romeo, my very kind Mexican friend. We have been through many similar experiences. I always felt so much understood talking to you. I hope you felt the same way. Thanks also for your super hospitality in Granada.

The *little black fish*[1] journey of mine started years before I left Iran. This could not have been possible without support and encouragements of my beloved friends, Sara and Sepideh and their families who became as close as my own family. Thank you!

Thanks to my very special family: my parents, Parvaneh, Babak, Parastoo, Behrooz, Behdad, Majid, Mehdi, Hamed, Mobina, and Jina. Thanks for putting so much trust in me and accepting me as the rebellious person that I was. Thanks for your love and support!

[1]The story can be found on: `http://www.iranchamber.com/literature/sbehrangi/works/the-little-black-fish.php`. The image below is also taken from this link.

Last but not least, my heartfelt thanks to my mother, whom I left behind to be the *little black fish* of my childhood hero and live a life in my very own way, when her life was me. Thanks for loving me so much that you let me go. Thanks mom! for being such a wonderful person whom I did not understand nor appreciate well. Thanks for giving me many life lessons. You left us too soon and so unexpectedly that I still have a hard time believing that you are not actually coming back. I only wish I could have made up a tiny bit of what you had done for me, to you. Although very small, please accept this, the hard work of my past almost five years and the beautiful life that I lived through it.

Yours Sincerely,

May 2013- Groningen