MODELLING AND INVESTIGATING PRIMARY BEAM EFFECTS OF REFLECTOR ANTENNA ARRAYS

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ASTROPHYSICS

By

KELACHUKWU IHEANETU

In the Department of Physics and Electronics
Faculty of Science at Rhodes University

Supervisor: Prof. O. Smirnov

Co-Supervisors: Dr. J. Girard, Dr. K. Thorat, Dr. S. Makhathini

August 2019

Declaration

I, Kelachukwu Joseph Iheanetu, declare that:

- The research presented in this thesis is my original work except where otherwise indicated.
- This thesis has not been submitted in parts or wholly for any degree or examination at any university.
- Where I consulted work published by others, I clearly attributed it to them, else the tables, graphs, pictures, data or other information in this thesis belongs to the author.
- Where in this report I have quoted from the works of others, the source is always given. Except for those quotations, this thesis is entirely my work.
- All the main sources of help have also been acknowledged.

Abstract

Signals received by a radio telescope are always affected by propagation and instrumental effects. These effects need to be modelled and accounted for during the process of calibration. The primary beam (PB) of the antenna is one major instrumental effect that needs to be accounted for during calibration. Producing accurate models of the radio antenna PB is crucial, and many approaches (like electromagnetic and optical simulations) have been used to model it. The $\cos^3$ function, Jacobi-Bessel pattern, characteristic basis function patterns (CBFP) and *Cassbeam* software (which uses optical ray-tracing with antenna parameters) have also been used to model it. These models capture the basic PB effects. Real-life PB patterns differ from these models due to various subtle effects such as mechanical deformation and effects introduced into the PB due to standing waves that exist in reflector antennas. The actual patterns can be measured via a process called astro-holography (or holography), but this is subject to noise, radio frequency interference, and other measurement errors.

In our approach, we use principal component analysis and Zernike polynomials to model the PBs of the Very Large Array (VLA) and the MeerKAT telescopes from their holography measured data. The models have reconstruction errors of less than 5% at a compression factor of approximately 98% for both arrays. We also present steps that can be used to generate accurate beam models for any telescope (independent of its design) based on holography measured data.

Analysis of the VLA measured PBs revealed that the graph of the beam sizes (and centre offset positions) have a fast oscillating trend (superimposed on a slow trend) with frequency. This spectral behaviour we termed *ripple* or *characteristic effects*. Most existing PB models that are used in calibrating VLA data do not incorporate these direction dependent effects (DDEs). We investigate the impact of using PB models that ignore this DDE in continuum calibration and imaging via simulations. Our experiments show that, although these effects translate into less than 10% errors in source flux recovery, they do lead to 30% reduction in the dynamic range. To prepare data for HI and radio halo (faint emissions) science analysis requires carrying out foreground subtraction of bright (continuum) sources. We investigate the impact of using beam models that ignore these ripple effects during continuum subtraction. These show that using PB models which completely ignore the ripple effects in continuum subtraction could translate to error of more to 30% in the recovered HI spectral properties. This implies that science inferences drawn from the results for HI studies could have errors of the same magnitude.

Acknowledgements

I want to express my heartfelt gratitude to a few people whose help made completing this work possible.

I thank my supervisors, Prof. Oleg Smirnov for his inspiration and encouragements, for all the academic advice and support by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation and from the Royal Society and the Newton Fund under grant NA150184. My thanks go to my co-supervisors Dr Julien Girard and Dr Kshitij Thorat, Dr Sphesihle Makhathini, Dr Khan M. B. Asad and Prof. Rick Perley for all their suggestions and contributions throughout this study.

I thank the staff and facilities of the National Radio Astronomy Observatory (NRAO) for the support, clarifications and providing the data for this research.

I thank my wife and family members for all their encouragements and for being with through the study.

My sincere gratitude goes to my colleagues, Ulrich, Theo, Alex, Cyndie, all members of the RATT team (past and present) and the Department for all their support, discussions, reviews and engaging arguments which gave this work a perfect finish.

Finally, I thank the Almighty God for all his blessings and for never letting me down.

Contents

List of Figures

List of Tables

Symbols

Symbol	Name
P	power
ω	Angular frequency
B_{d}	baseline
λ	Wavelenght
Ω	Solid angle
I	Intensity or brightness
A	Area
$\boldsymbol{b}$	Distance vector
$\boldsymbol{R}$	Position vector
t	time
τ_g	Geometric delay
$\boldsymbol{s}$	Unit vector
$\boldsymbol{V}$	Visibility
$()^{\mathrm{T}}$	Transpose
A'	Normalised antenna primary beam or radiation pattern
A_0	The antenna response at the phase center
l, m, n	Sky coordinate systems
u, v, w	Baseline coordinate systems
$\mathcal{F}\{\}$	Fourier transform
$\mathcal{F}^{-1}\{\}$	Inverse Fourier transform
S	Sampling fuction
$\boldsymbol{e}$	Electric field vector
$\boldsymbol{J}$	Jones matrix
$\boldsymbol{u}, \boldsymbol{v}$	Complex vectors
$\mathbf{B}$	Brightness matrix

K	Scalar K-Jones matrix representing the phase delay effect
$\otimes$	Kronecker product
$\mathbf{vec}(.)$	vectorisation operation
$\boldsymbol{\mathcal{J}}$	Jacobian matrix
B_{CassN}	Beams generated with the *Cassbeam* software
B_{CassR}	B_{CassN} scaled and shifted to match the holography beams full width at half maximum (FWHM) and offset (we termed the *ripple beams*)
B_{CassA}	B_{CassN} scaled and shifted to match the average scale and centre offset of the holography beams full width at half maximum (FWHM) and offset (we called the *average beams*)
f_{err}	Flux density error
g_{err}	Mean calibration gain amplitude errors
$g_{\mathrm{applied,err}}$	Error in mean primary beam gain applied on each source
DR	Dynamic range
ZP	Zernike Polynomial
$\boldsymbol{E}^{\mathrm{h}}$	De-noised holography beam models
$\boldsymbol{E}^{\mathrm{c}}$	Principal component analysis (PCA) beam models
$\boldsymbol{E}^{\mathrm{z}}$	Zernike beam models
$\boldsymbol{C}^{\mathrm{c}}$	Decomposition coefficients

To my family...

Chapter 1

Introduction

Astronomy is a science which have been practised since the early civilisation even in the era of the Babylonian empire. It is defined as the study of celestial objects such as the Sun, Moon, the planets, other stars and galaxies. This involved mainly the use of optical telescopes (i.e. optical astronomy) before the accidental detection of radio signals in the 1930s from outside the Solar System. After this discovery, it took about a decade to produce the first radio image of the sky using antennas made of parabolic reflectors (Reber, 1948). Radio astronomy involves studying the celestial objects via the emission they radiate in the radio band of the electromagnetic spectrum. Subsequent decades experienced the development of radio telescopes of different designs and configurations.

Radio signals have wavelengths that are up to six orders of magnitude higher compared to those of the optical regime. The technique of interferometry has played a very important role in the observation of the radio sky. A two-element interferometer allows the signals from two telescopes to be correlated together to achieve higher angular resolution[1] than could be obtained with a single telescope (Thompson *et al.*, 1986).

The signals received by radio antennas are affected by propagation effects as they travel from the sources to the observing antennas and these effects need to be modelled and corrected through a process called "calibration". One of these propagation effects is the antenna's primary beam (PB). Generating accurate PB models of radio antennas is important to the radio astronomy community, particularly for the new generation of radio telescopes, because of their higher sensitivity that forces us to account for the most subtle effects. Many techniques have been developed over many decades to model

[1]Resolution here is a measure of how close objects can be together and still be distinguished by a telescope as separate objects.

the radio antenna PB which include, electromagnetic and geometric optical simulations (Brisken, 2003), analytical pattern models (Popping and Braun, 2008; Braun *et al.*, 2007), characteristic basis function patterns (Young *et al.*, 2013; Mutonkole and de Villiers, 2015), etc. The actual PB deviates from these models because various subtle instrumental effects, such as mechanical deformations of the main reflector, pointing-errors, standing waves[2] etc. are not completely accounted for by these models. Good knowledge of the PB can be obtained by indirect measurements (e.g. using beam holography, etc). These processes are prone to noise, radio frequency interference and other potential effects from the environment.

In this research, we present two sets of holography beams (for Karl G. Jansky Very Large Array (VLA) and MeerKAT antennas) and discuss the task of using them to derive parametrised models for their PB. Analysing the holography PBs of the VLA antennas show that they exhibit complicated frequency behaviour which is due to the standing wave in the optics(Popping and Braun, 2008). We test the impact of ignoring the effect introduced into the PB by the standing waves during calibration and Stokes I imaging.

1.1 The purpose of the research

Having a good model of the PB of the radio antenna is of crucial importance in radio astronomy because it is necessary to correct for direction-dependent effects (DDEs) in the observed data. Achieving high dynamic range[3] (HDR) with minimal bias requires a good knowledge of the realistic behaviour of the antennas PB beyond its first side-lobe. Direction-dependent (DD) calibration algorithms (e.g. differential gain, "source peeling" etc.) can mitigate most artefacts but not source suppression, which compromises data fidelity. (Most of the results presented in this study are taken from Iheanetu *et al.* (2019))

This study hopes to:

- Obtain accurate PB models which can be used for calibration.
- Generate models of the PB using the lowest (possible) degree of freedom that captures essential features of the PBs and their variations.

[2]The *standing waves* here is a semi-periodic signal that keeps re-bouncing within the antenna optics.
[3]Dynamic range in radio astronomy is a measure of the ratio of the brightest pixel to the background noise in an image.

- Produce a set of generic methods to model the spatial and spectral morphology of the PB of any radio telescope, independent of its design.

1.2 The contributions of this study

In this section, we present the two techniques used in this study to model the PB of radio telescopes from measured holography data, we tested the impact of not accounting for DDE introduced into the PB models used for calibration and present the contribution of this study.

The first technique is to use principal component analysis (PCA) (through singular value decomposition (SVD)) to model both spatial and spectral variation of the PB by applying the PCA on the holography measured data. This decomposition provided the first level of data compression because the number of coefficients used for the PB models is much less than the number of measured beams. This produced accurate PB models and compressed the number of data points needed to store the models by more than 99.85% (using only 20 and 15 coefficients for the VLA and MeerKAT antennas, respectively).

Radio frequency interference corrupted the measured data at some frequencies, so the data for the affected frequencies were excluded from the analysis. Discrete cosine transform (DCT) was used to model the spectral behaviour of the PCA coefficients and interpolate PB at the RFI affected frequencies. This further compressed the model by reducing the number coefficients needed to reconstruct the PBs.

The second technique is to use Zernike polynomials to decompose and model the holography data. Zernike polynomials are common orthogonal basis used to represent the morphological variations of optical fields over a circular aperture (Zernike, 1934). Per frequency Zernike decomposition of the holography measured data was done and the dominant Zernike coefficients were selected in the L-band frequency.

The PCA technique was used to model both the VLA and MeerKAT PBs from holography measured data with high accuracy while the Zernike decomposition was applied only on the VLA data. The study presents a set of antenna design independent methods/techniques which can be used to model the PB of any radio telescope from holography measured data with high accuracy.

Finally, we performed a series of experiments to test the impact of ignoring the spectral and spatial behaviour of the primary beams of the VLA antennas which highlight the

value of the approach taken by this study in obtaining accurate primary beams models. This shows that using a PB model that completely ignores the spectral characteristics found in the VLA holography measured beams for calibrating continuum data could degrade the image produced from the result by a factor of up to 30%. This is particularly important in a case were the final aim of the observation is to produce high quality image for science investigation. We also show that performing continuum calibration of bright sources using PB models which completely ignore the ripple effects can translate to errors of more than 30% in the recovered spectral line properties of the HI sources (after continuum subtraction).

1.3 Thesis outline

In Chapter 2, we presents a general introduction to radio antenna design and response, interferometry and calibration followed by a brief introduction to the derivation of the radio interferometry measurement equation (RIME) formalism and calibration algorithms. Discussion on the propagation effects which affect the signal received by radio telescopes is given in this chapter.

We discuss holography measurement technique as one of the proxies that can be used to gain access to the primary beam information (in Chapter 3) as well as the steps which are used to carry out holography measurement (using the VLA as a case study). The measured holography data are discussed and some preliminary analysis presented. These include the main features of the beam pattern derived from holography data - its morphology, its spectral behaviour and its deviation with respect to the ideal beam expected from electromagnetic simulations. It is the first step taken in analysing the data before modelling the PB.

In Chapter 4, we address the primary beam modelling and motivate for an accurate model for the radio antenna beam patterns. Some methods that are used in modelling the primary beam are listed, but we focus on the two that were used to model the primary beam of the VLA and MeerKAT antennas from holography data, namely PCA and Zernike polynomials. Characteristic decomposition coefficients were obtained after the sparse reconstruction of some sub-band that were affected by RFI (hence, excluded from the analysis).

In Chapter 5, we discuss designs of some experiments which were used to test the impact of not accounting for the various characteristic features of the measured beam patterns (e.g the spectral ripples).

Finally, Chapter 6 presents a summary of the results and the conclusions that are drawn from the study, as well as the limitations. We then outlines future studies.

Chapter 2

The response of an interferometer

In astronomy, angular resolution is often a crucial observational limitation. The (spatial) angular resolution θ of a telescope is a measure of the smallest possible detail of an area in the sky the telescope can distinguish. For a conventional circular aperture telescope, it is given by the Rayleigh criteria $\theta = 1.22\frac{\lambda}{D}$ (rad), where λ is the wavelength and D is the diameter of the aperture. At optical wavelengths ($\lambda \approx 10^{-7}$ m), with giant mirrors of a few tens of meters in diameter, one can achieve milliarcseconds scale resolution. However, the radio wavelength ($\lambda \approx 10^{-3}$ – 10 m) regime would require a single reflector size of 10^3 – 10^7 m diameter to achieve similar resolution. Building such a telescope would pose great engineering and economic challenges. To achieve resolutions comparable to those in the optical regime, radio astronomers employ the technique of interferometry. Interferometry in radio astronomy is the technique of combining signals from an array of telescopes to achieve high resolution. A pair of telescopes whose signals are correlated is called an interferometer. The resolution of an interferometer is given by:

$$\theta \approx \frac{\lambda}{B_{\rm d}}, \tag{2.1}$$

where $B_{\rm d}$ is the distance between the interferometer elements called the baseline.

It is for this reason that radio astronomers, having "invented" the field using individual antennas to make the first observations, quickly moved on to constructing interferometric arrays. Success in this led to the invention and development of aperture synthesis which is based on Van Cittert-Zernike theorem (Thompson *et al.*, 2008). Aperture synthesis involves the compilation of signals from many telescopes that are at varying distances which can be used to produce images that have angular resolutions comparable to that

of an instrument that has the size of the whole collection (Ryle *et al.*, 1959; Scott *et al.*, 1961).

To understand the operation of an interferometer, we must first look at its constituent element, the antenna. We shall do this in the next section; this will then be followed by the theory of interferometry proper.

2.1 Antennas

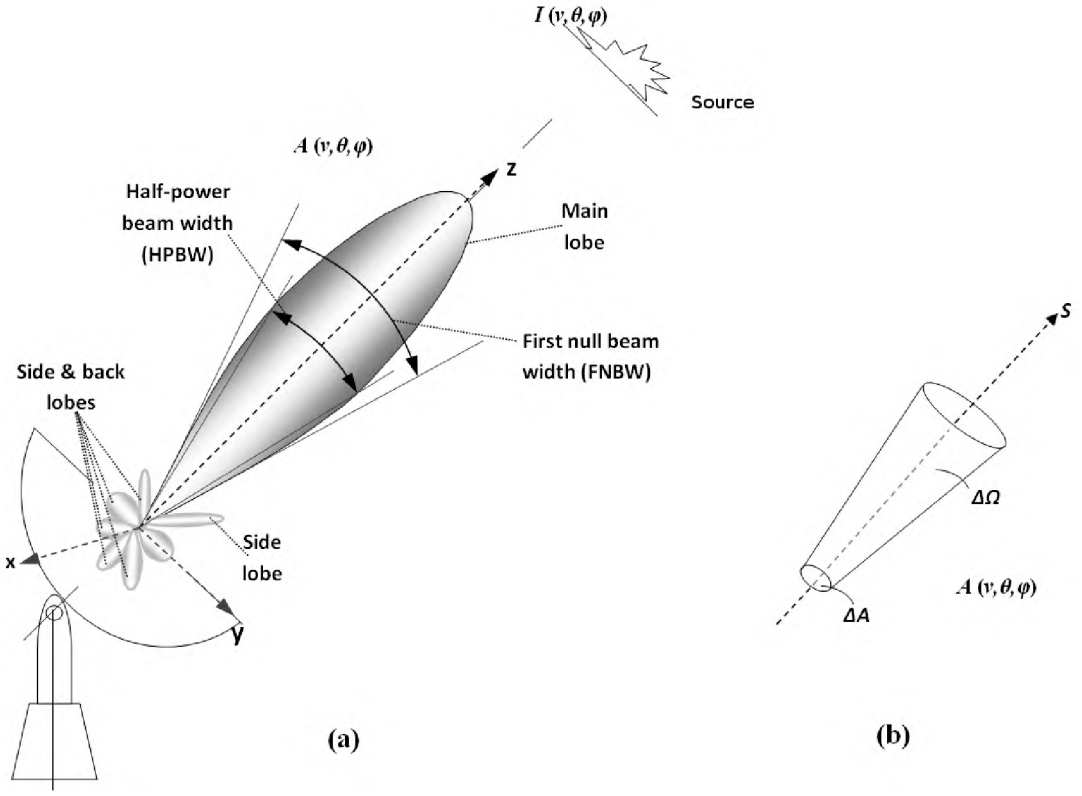

FIGURE 2.1: A schematic drawing (a): of the radiation pattern (primary beam) of an antenna showing major, minor, side and back lobes and having the antenna pointing towards the z-axis and (b): small element of area in the direction of the unit vector $\boldsymbol{s}$ through a small element of solid angle (Balanis, 2005).

In a simplistic sense, a device that converts radio frequency fields into alternating current (or vice versa) is commonly referred to as an antenna. In radio astronomy, signals are the term used to describe the voltages induced on an antenna by radiation coming from cosmic sources. In general, the primary element of an interferometer is the antenna. Consider Figure 2.1(a), let the effective signal-collecting area of the antenna pointing at a source be $A(\nu, \theta, \phi)$ at the frequency, ν. The small element of power, ΔP (in watts)

received by an observer in a small element of area, $\Delta A(\nu, \theta, \phi)$ (Figure 2.1(b)), through a small element of solid angle, $\Delta\Omega$, in a small element of bandwidth $\Delta\nu$, of signals coming from a source in the direction normal to the plane on this area is expressed as $\Delta P = \Delta A(\nu, \theta, \phi) I(\nu, \theta, \phi) \Delta\nu \Delta\Omega$. θ and ϕ are the coordinates of direction, while $I(\nu, \theta, \phi)$ is the brightness or specific intensity of the source (measured in $\mathrm{Wm^{-2}\ Hz^{-1}\ sr^{-1}}$). It gives a measure of the power received per unit area, per unit solid angle, per unit bandwidth. Brightness integrated over the solid angle Ω gives us another interesting quantity of measurement called flux density, which has units of $\mathrm{Wm^{-2}\ Hz^{-1}}$. Astronomers adopted a convenient unit of measuring flux density called Jansky (Jy) because astronomical sources generally have very small values for flux density. ($1\ \mathrm{Jy} = 10^{-26}\ \mathrm{Wm^{-2}\ Hz^{-1}}$).

Basic parameters of antennas: Balanis (2005) defines an antenna as a transitional device between free-space and a transmission line, which can be used in transmit or receive mode. The radiation pattern or primary beam of an antenna represents the far field response of the antenna as a function of direction. An ideal isotropic antenna would have its radiation pattern being the same in all directions, but a real antenna radiates (or receives) more power in some directions as compared to others. A schematic representation of the antenna's primary beam with a circular aperture is presented in Figure 2.1. Antennas have one important property, which is that their radiation pattern is the same when operating in either transmit or receive mode (Wilson *et al.*, 2009). Antennas in radio astronomy are always used in receive mode but it is easier to visualise and analyse them in the transmit mode and the result of this analysis can then be applied to the antenna in the receive mode. The *Beamwidth* is another parameter that is associated with an antenna. The angle between the half-power point of the main-lobe is the beamwidth (beam size). The full width at half maximum (FWHM) is also a common criterion that is used to describe the resolving power of an antenna. It is given as the angle between the point at which the intensity of the radiation falls to half of its maximum (or where the amplitude of beam falls to $1/\sqrt{2}$ of its maximum) value (see Figure 2.1). It also, gives a measure of the size of the telescopes main lobe (i.e. the size of the antenna). The resolution of an antenna, θ is inversely proportional to its FWHM.

There are different types of antennas (and antenna configurations) used for radio astronomy observations. The choice of antenna type is dependent on the frequency band of interest, for example, wire antennas, which may include dipoles, helices, yagis and spirals etc. are mostly used for wavelengths longer than one meter. For wavelengths shorter than one meter, more collecting area is required to increase the power of the received

signal. One common design that is generally used to increase the collecting area is the dish design. The type of dish design can be categorised by its type of mount and type of optics.

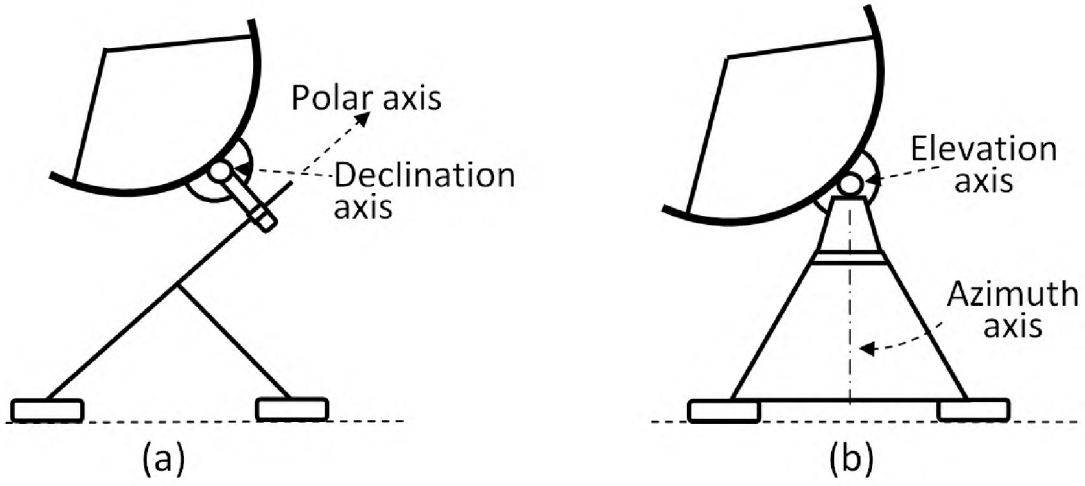

FIGURE 2.2: The cross-section of steerable mounts designs:(a) equatorial and (b) elevation over azimuth (alt-azimuth) mounts (Perley *et al.*, 1989)

A typical radio antenna has three major components, namely optics, mount and electronics receiver systems. Here we only cover the antenna optical and mount systems. These two can be used to classify the antenna.

The antenna mount: Antennas used in radio astronomy can be installed to function as steerable (tracking) or transit (non-tracking, i.e. fixed - observes the sky as it passes over it) etc. but in this discussion, we will focus on steerable antennas. The equatorial and an elevation over azimuth mounts are the commonly used as tracking antenna mount designs. The elevation over azimuth (or alt-azimuth) mount normally has two axes of rotation - azimuth and elevation (Figure 2.2(b)). One major advantage of this mount is that it is simple and hence has relatively lower building cost (i.e. relative to the cost of building an equatorial mount design). Another advantage is that gravity always acts on the same plane so this design keeps the reflector profile accurate as it tracks radio astronomy sources. One disadvantage of this design is that, as the antenna tracks a source, the aperture rotates with respect to the source (Thompson *et al.*, 1986). The result of this is that, if the antennas primary beam is asymmetric, this rotation will cause the source apparent brightness to vary. This rotation also causes beam squint (or offset) rotations with the parallactic angle for antennas that have non-identical dipole receptors that creates a complex instrumental polarisation effect (particularly for antennas with

circularly polarised receptors). If an observation requires this angle to be constant, one needs to compensate for this as the antenna tracks the source. Polar or equatorial mount is another type of mount. One of the prominent advantages of the equatorial mount is that it has a polar axis (Figure 2.2(a)) that is aligned to be parallel with the axis of rotation of the earth. So, to track a source (at sidereal rate) one only needs to rotate the antenna about this axis. Another is that it does not have the beam rotation problem. One major disadvantage of this mount is the complexity added by constructing the polar axis, which increases the cost of the design. A minor disadvantage is that it is difficult to observe sources that are close to the horizon in a direction that is away from the celestial pole using this mount.

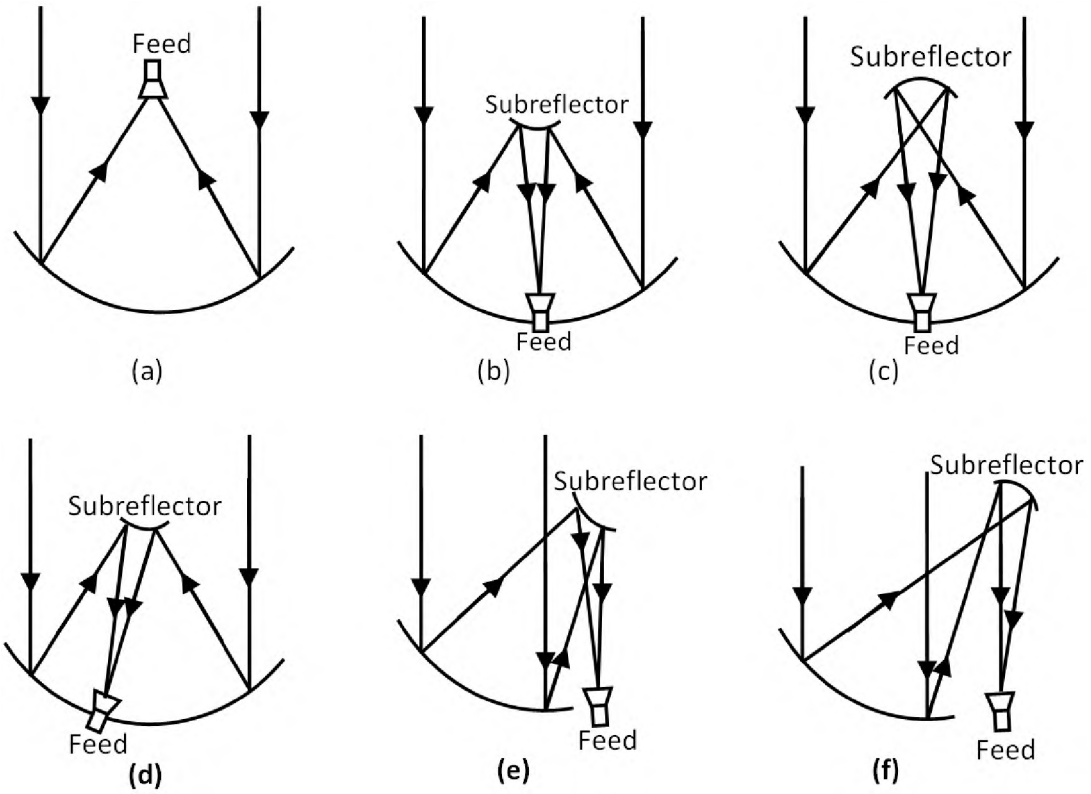

FIGURE 2.3: Cross-section schematics of optical systems used for radio telescope reflectors. (a) Prime focus, (b) Cassegrain, (c) Gregorian, (d) off-axis Cassegrain, (e) offset Cassegrain and (f) offset Gregorian systems (Perley *et al.*, 1989)

Types of antenna reflector or optics: The optical systems used in radio astronomy antenna design have the ability to focus rays (radiations) coming to it parallel to the principal axis of the equipment to its focus. The type of optical system used to feed the radio reflector can also be used to classify the antenna (Rudge *et al.*, 1982) and can be

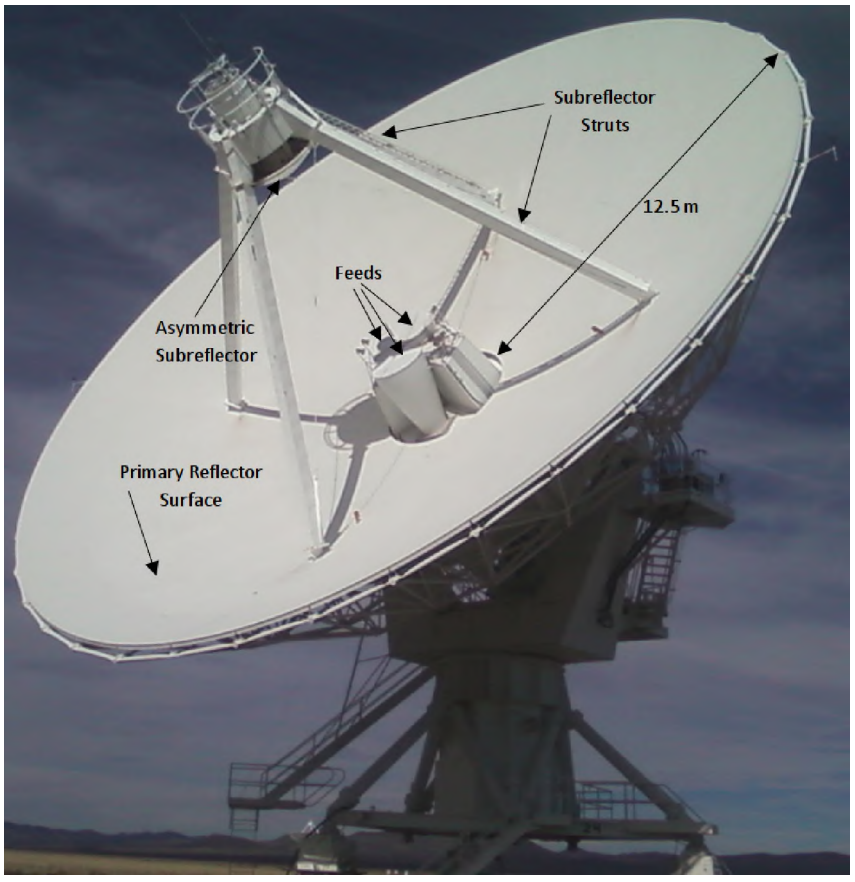

FIGURE 2.4: A picture of the VLA antenna showing the sub-reflector and its supporting strut, the feeds and primary reflector.

grouped as either prime focus or offset configurations. The schematics of the cross-section of a few optical designs (that use paraboloidal main reflector) are presented in Figure 2.3.

The prime focus, the Cassegrain, the Gregorian, off-axis Cassegrain, offset Cassegrain and offset Gregorian systems are some designs of optical systems used in building radio astronomy telescopes. These systems are either single or dual reflector systems. A single reflector design has only the main (or primary) reflector (or collector) while the dual reflector design normally has a main collector and a sub-reflector. Figure 2.3(a) is the schematic cross section of a prime focus design having a single reflector system and having its feed (and electronics) mounted at the focal point in the front (and centre) of the primary (paraboloid) reflector by the support structures. The Cassegrain telescope (Figure 2.3(b)) is a dual-reflector design which has a symmetric hyperboloid sub-reflector placed in the front centre of the paraboloidal main collector to reflect rays coming from the primary collector into the feed. The feed is placed in the optical centre of the main collector. Gregorian design is another dual-reflector configuration (Figure 2.3(c)), which has a paraboloidal main reflector and an ellipsoidal sub-reflector that is mounted in the

front centre position of the main reflector just as in the case of the Cassegrain system. The off-axis Cassegrain system (Figure 2.3(d)) (a dual-reflector configuration) uses an asymmetric hyperboloid sub-reflector which allows for multiple feeds to be positioned in a circular ring around the axis of the (paraboloid) main reflector so that selecting different frequencies only requires a rotation of the sub-reflector. This design is perfect for synthesis telescopes that require frequency flexibility. All these optical systems (or configurations) described above, have one major common disadvantage - the feed or sub-reflector and their support struts both block and scatter the radiation received by the primary collector. This results in these designs having reduced aperture and non-symmetric primary beams, particularly at the side-lobes. The offset Cassegrain (Figure 2.3(e)) and offset Gregorian (Figure 2.3(f)) designs (dual-reflector configuration) both avoid this problem by having their sub-reflectors and support struts away from the line of sight of the primary collector to ensure improved optical efficiency. These offset designs both use a paraboloidal main collector and a hyperboloidal (for offset Cassegrain) and ellipsoidal (for offset Gregorian) sub-reflectors.

2.1.1 The VLA antenna

The Jansky Very Large Array (VLA) is a radio interferometer array consisting of 27 (identical) radio antennas located to the west of Socorro, New Mexico, USA. Each antenna is equipped with eight feed horns installed in a ring around the optical centre of the primary reflector. The VLA antennas initially equipped with receivers to receive signals in the narrow-band L, C, X and Ku bands feed horns, but these were extended in 2016 with a new set of wideband feed horns to include P, L, S, C, X, Ku, K, Ka and Q band frequencies in the EVLA (Expended Very Large Array) project (Perley, 2016). That same year, holography measurement was also conducted for these bands. In this study, we will restrict our discussion to L-band.

The antennas of the VLA array can be set up to operate in one of four spatial layouts or configurations - namely the A-, B-, C- and D-configurations. The A-configuration has the longest baselines and hence, the highest angular resolution (from Equation (2.1)) at a given frequency, but this gives rise to very low sensitivity to extended structure. The baseline (length) decreases from configuration A through to D. The D-configuration has the shortest baselines, thus the highest surface brightness sensitivity at the expense of resolution that can be achieved with it. So the choice of configuration depends on the science of interest.

The antennas of the VLA are typical examples of off-axis Cassegrain systems (Figure 2.3(d)) with dual reflectors - the primary (or main) and a convex secondary (or sub-) reflectors which have an *alt-azimuth* mount (see Figure 2.3(b)). In Figure 2.4, we present a picture of the VLA antenna, showing the main and secondary reflectors, a number of feeds around the centre of the primary reflector, the sub-reflector support struts and the central blockage. The VLA antennas have a non-symmetric hyperboloid sub-reflector and adding to the disadvantages of this design mentioned above, is the beam squint. Beam squint occurs as a result of the feeds not being positioned at the optical centre of the main reflector giving rise to the beam centre being slightly offset. The VLA antennas are equipped with dual (left-hand and right-hand) circularly polarised receivers that are positioned to be at perpendicular orientation to each other.

Description	Length (m)
x feed position	-0.10026
y feed position	097019
z feed position	1.6764
height of the sub reflector	8.47852
width of the support strut legs	0.27
Strut legs distance from vertex	7.55
support strut length from vertex	10.93876
Diameter of antenna	25.0
Focal length of the main reflector	9
sub reflector angle	9.26^o

TABLE 2.1: VLA antennas geometry/dimension (Napier, 1996)

Standing wave is another phenomenon that is posing a challenge to radio astronomy observations (Briggs *et al.*, 1997). There is an inherent problem of standing waves for these telescope designs built with reflector components. A small percentage of the signals received by the telescope keeps re-bouncing within the optics (feed/electronics) creating a standing wave. This is a term used to describe semi-periodic oscillation in the spectral bandpass of radio telescopes (Popping and Braun, 2008). It causes destructive interference at some frequencies and constructive interferences at others. The prime focus (including the Cassegrain, Gregorian and off-axis Cassegrain) systems are worst affected by this *standing wave*. This standing wave effect introduces a characteristic frequency scaling on the primary beams, which produced a direction-dependent spectral effect, as we will show in Chapter 3.

2.1.2 The MeerKAT antenna

The MeerKAT antennas are located in the Karoo, Northern Cape, South Africa. Originally, the Karoo Array Telescope project planned for a 7-antenna prototype called KAT-7, and a full array of 20+ dishes. A subsequent increase in the project budget expanded the plan to 64 antennas, and the name became "MeerKAT. *Meer* means *more* in Afrikaans, so *MeerKAT* literally means *more KAT* that is *more KAT* antennas. (Meerkat is also the Afrikaans name for a small carnivorous animal (Suricata suricatta) of the mongoose species that is found in South Africa)[1]. At the time of this writing, the MeerKAT telescope features an array of 64 identical antennas. The MeerKAT antennas use the offset Gregorian (dual-reflector) system (Figure 2.3(f)) equipped with a parabolic main reflector and a concave ellipsoidal symmetric sub-reflector, installed with alt-azimuth mounts (Figure 2.2(b)). Each antenna had a projected diameter of the main reflector (minor axis) of 13.5 m, a sub-reflector diameter of 3.4 m and are designed to receive linearly (horizontal (H)/vertical (V) polarised signals (Booth *et al.*, 2009).

2.2 Radio interferometry fundamentals

Let us consider the simplest interferometer (Figure 2.5), consisting of only two identical antennas p and q at positions $\boldsymbol{r}_p$ and $\boldsymbol{r}_q$ respectively. Incident radiation coming from an astronomical source (in the far field) in the direction of the unit vector $\boldsymbol{s}$ arrives at the two antennas separated by a distance vector $\boldsymbol{b} = \boldsymbol{r}_p - \boldsymbol{r}_q$, known as the baseline. The incident radiation is considered to be plane waves (because it is coming from a source whose distance is very far compared to the distance separating the antennas, such that the source is assumed to be at $\approx \infty$) and is received at point $\boldsymbol{r}_q$ at time t and at point $\boldsymbol{r}_p$ at time $t + \tau_g$. The radiation arrives at antenna p later, after travelling an extra path length that results in a delay known as geometric delay, given by:

$$\tau_g = \frac{\boldsymbol{b} \cdot \boldsymbol{s}}{c} \tag{2.2}$$

where $c = 2.99792458 \times 10^8$ m/s is the speed of light in vacuum and $\boldsymbol{s}$ is a unit vector pointing towards the source.

[1]Polity - SKA: Statement by the Square Kilometre Array Organisation, on first MeerKAT antenna and high-tech data centre launched in the Karoo (27/03/2014)
https://www.polity.org.za/print-version/ska-statement-by-the-square-kilometre-array-organisation-on-first-meerkat-antenna-and-high-tech-data-centre-launched-in-the-karoo-27032014-2014-03-27

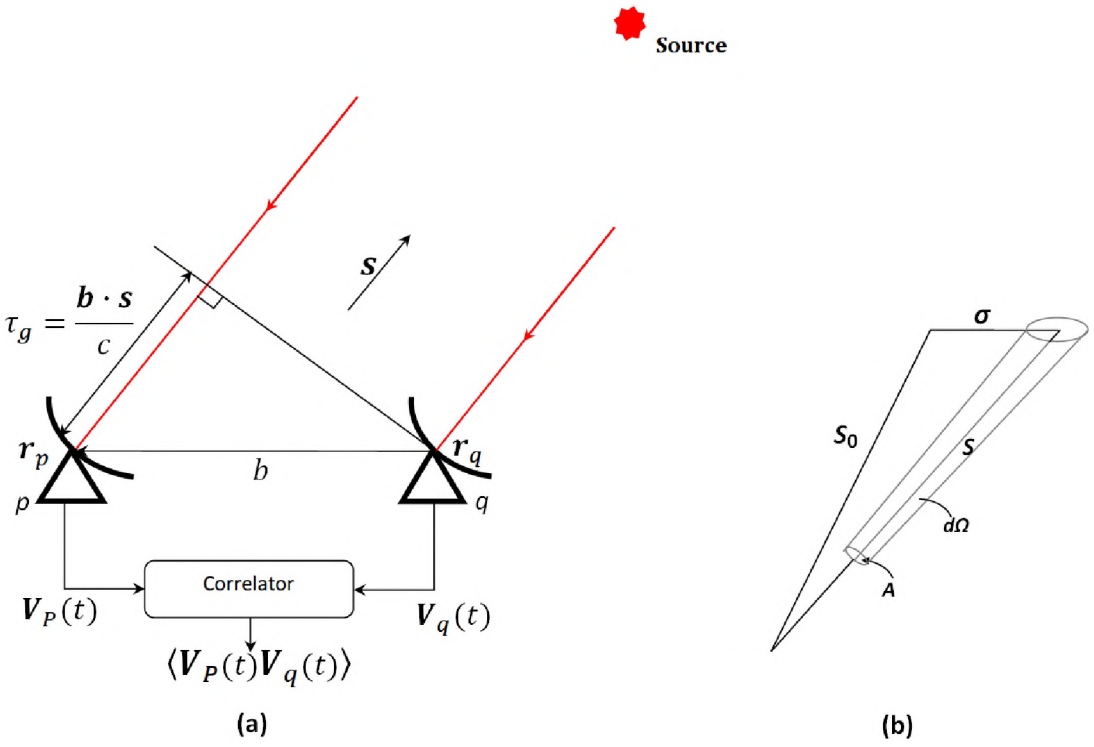

FIGURE 2.5: A schematic of (a) a two element interferometer. The signal coming from a source arrives at antenna q before p hence a delay is introduced into the signals path of q before both signals are correlated to form visibilities and (b) the position vectors that are used in deriving expressions for the interferometer response.

To derive a general formulation for an interferometer, let us ignore polarisation properties of the source for now (and we follow the derivation in Taylor *et al.* (1999)). The voltage responses from each antenna are passed through amplifiers and filters (and other processes) to select the desired frequency bandwidth, $\Delta\nu$ centred on ν. These signals are combined in the correlator by a process of multiplication and averaging. Let the signal received by antenna p and q be $\boldsymbol{V}_p(t)$ and $\boldsymbol{V}_q(t)$ respectively. The output of the correlator is defined as:

$$r = \langle \boldsymbol{V}_p(t)\boldsymbol{V}_q(t)\rangle,$$

where $\langle\cdot\rangle$ denotes the expectation value found by averaging over some integration time. Assuming the signals are quasi-monochromatic, we have from Taylor *et al.* (1999) that:

$$\boldsymbol{V}_p(t) = \boldsymbol{V}_p \cos(2\pi\nu(t - \tau_g)), \tag{2.3a}$$

$$\boldsymbol{V}_q(t) = \boldsymbol{V}_q \cos(2\pi\nu t), \tag{2.3b}$$

where term $2\pi\nu\tau_g$ is a phase term introduced by the correlator to account for the geometric delay. τ_g changes slowly as the earth rotates with time. Using the trigonometric identity $\cos(a-b) = \cos(a)cos(b) + \sin(a)\sin(b)$ and Equations (2.3), the output of the correlator is given by (Taylor *et al.*, 1999):

$$r(\tau_g) = \langle \boldsymbol{V}_p \boldsymbol{V}_q \cos(2\pi\nu\tau_g) \rangle. \tag{2.4}$$

To present the output of the correlator as a function of the radio brightness integrated over the sky, let the effective collecting area of the antennas be $A(\boldsymbol{s})$, the radio sky brightness in the direction of a unit vector $\boldsymbol{s}$ be $I(\boldsymbol{s})$ and assuming it's the same for both antennas. Then, the signal power received in the bandwidth $\Delta\nu$ in the direction, $\boldsymbol{s}$ in the solid angle $d\Omega$ (Figure 2.5(b)) is $A(\boldsymbol{s})I(\boldsymbol{s})\Delta\nu d\Omega$. Assume that $\Delta\nu$ is very small such that variations in A and I as a function of ν can be ignored, the signal is coming from a very far distance ($\approx \infty$ such that the radiation arrives as plane waves on the baseline) and the components of the signal coming from different points on the surface of the source are uncorrelated. Then, the correlator output dr for the signal per unit solid angle element is (Taylor *et al.*, 1999):

$$dr = A(\boldsymbol{s})I(\boldsymbol{s})\Delta\nu \cos(2\pi\nu\tau_g)d\Omega. \tag{2.5}$$

Substituting τ_g from Equation (2.2) into (2.5) and integrating over the surface, S of the celestial sphere gives:

$$r = \int_S A(\boldsymbol{s})I(\boldsymbol{s})\Delta\nu \cos(2\pi\nu\frac{\boldsymbol{b}\cdot\boldsymbol{s}}{c})d\Omega, \tag{2.6}$$

Conventionally, one is expected to specify the centre of the synthesis field of view (FOV), known as the phase tracking centre (or phase centre), $\boldsymbol{s}_0$, given by:

$$\boldsymbol{s} = \boldsymbol{s}_0 + \boldsymbol{\sigma}, \tag{2.7}$$

where $\boldsymbol{\sigma}$ is the position of $\boldsymbol{s}$ w.r.t $\boldsymbol{s}_0$ (see Figure 2.5(b)). Substituting Equation (2.7) into (2.6) and applying the trigonometric identity $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$, then the signal response of the interferometer can be written as:

$$\begin{aligned} r = \Delta\nu \cos(2\pi\nu\frac{\boldsymbol{b}\cdot\boldsymbol{s}_0}{c}) \int_S A(\boldsymbol{\sigma})I(\sigma)\cos(2\pi\nu\frac{\boldsymbol{b}\cdot\boldsymbol{\sigma}}{c})d\Omega \\ - \Delta\nu \sin(2\pi\nu\frac{\boldsymbol{b}\cdot\boldsymbol{s}_0}{c}) \int_S A(\boldsymbol{\sigma})I(\boldsymbol{\sigma})\sin(2\pi\nu\frac{\boldsymbol{b}\cdot\boldsymbol{\sigma}}{c})d\Omega, \end{aligned} \tag{2.8}$$

Let the complex visibility function of the interferometer be defined as (Perley *et al.*, 1989):

$$V = |V|e^{i\phi_V} = \int_s A'(\boldsymbol{\sigma})I(\boldsymbol{\sigma})e^{-2i\pi\nu b\cdot\boldsymbol{\sigma}/c}d\Omega, \tag{2.9}$$

where the phase of the visibilities is ϕ_V, $A'(\boldsymbol{b}) \equiv A(\boldsymbol{\sigma})/A_0$, is the normalised antenna primary beam or radiation pattern and A_0 is the antenna response at the phase centre. Separating Equation (2.8) into its real and imaginary parts, we have:

$$A_0|V|\cos(\phi_V) = \int_s A(\boldsymbol{\sigma})I(\boldsymbol{\sigma})\cos(2\pi\nu\frac{\boldsymbol{b}\cdot\boldsymbol{\sigma}}{c})d\Omega, \tag{2.10a}$$

$$A_0|V|\sin(\phi_V) = \int_s A(\boldsymbol{\sigma})I(\boldsymbol{\sigma})\sin(2\pi\nu\frac{\boldsymbol{b}\cdot\boldsymbol{\sigma}}{c})d\Omega. \tag{2.10b}$$

Substituting equations (2.10) into Equation (2.8) yields:

$$r = A_0\Delta\nu|V|\cos(2\pi\nu\frac{\boldsymbol{b}\cdot\boldsymbol{\sigma}}{c} - \phi_V). \tag{2.11}$$

In practice, to use Equation (2.9) for imaging we need a change of coordinate systems to a more convenient one. (u,v,w) are commonly used to represent the components of the baseline vectors where w is pointing towards the direction of $\boldsymbol{s}_0$ (which becomes the centre of the synthesized image), u towards East and v towards North (w is perpendicular to the uv-plane). u,v,w are measured in wavelengths at the centre frequency of the radio frequency band. Also, the sky positions are in l,m,n coordinate systems, which are direction cosines measured w.r.t u,v,w axes. The synthesised image in (l,m) plane is a

projection of the celestial sphere onto a plane that is tangent at (l, m). These coordinate systems are related through:

$$\nu\frac{\boldsymbol{b}\cdot\boldsymbol{s}}{c} = ul + vm + wn, \quad \nu\frac{\boldsymbol{b}\cdot\boldsymbol{s}_0}{c} = w$$
$$\text{and} \quad d\Omega = \frac{dldm}{n} = \frac{dl\ dm}{\sqrt{1-l^2-m^2}}, = w, \tag{2.12}$$

where $n = \sqrt{1-l^2-m^2}$. Then Equation (2.9) becomes:

$$V(u,v,w) = \iint_{-\infty}^{\infty} A'(l,m)I(l,m)\mathrm{e}^{-2\pi i[ul+vm+w(\sqrt{1-l^2-m^2}-1)]}\frac{dl\ dm}{\sqrt{1-l^2-m^2}}, \tag{2.13}$$

where the integrand in Equation (2.13) is taken to be zero when $l^2 + m^2 \geq 1$. The complex visibility is expressed as a function of (u, v, w) because these are the coordinates that represent the spacing of the antennas with respect to the phase tracking centre, $\boldsymbol{s}_0$. Equation (2.13) expresses $V(u, v, w)$ as a function of $A'I$ or the modified brightness distribution. Equation (2.13) is the fundamental equation in radio interferometry known as the Van Cittert-Zernike theorem. The next section will show that it reduces to a two-dimensional Fourier transform.

2.3 Radio synthesis imaging

Synthesis imaging (in the radio regime) involves using Equation (2.13) to generate a synthesised image. In general, Equation (2.13) is taken to be a Fourier transform by eliminating the w term if one makes the following assumptions.

- If the observation was done on a coplanar array (for very short snap-shot), such that $w = 0$, then Equation (2.13) reduces to:

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A'(l,m)I(l,m)\mathrm{e}^{-2\pi i(ul+vm)}\frac{dl\ dm}{\sqrt{1-l^2-m^2}}. \tag{2.14}$$

  Equation (2.14) is a two-dimensional Fourier transform.

- If the observation was done to image a small region of the sky, such that the term $\sqrt{1-l^2-m^2} \approx 1$, then Equation (2.13) reduces to:

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A'(l,m)I(l,m)\mathrm{e}^{-2\pi i(ul+vm)}dl\ dm, \tag{2.15}$$

  which is also a two-dimensional Fourier transform.

Now for the sake of simplicity, let us assume that the term $A'(l,m) = 1$ (i.e. the case of a small FOV), then the Van Cittert-Zernike theorem becomes:

$$V(u,v) = \iint I(l,m)\mathrm{e}^{-2\pi i(ul+vm)}dl\ dm. \tag{2.16}$$

For the case when $w = 0$, the dependence of $V(u,v)$ on $1/\sqrt{1-l^2-m^2}$, in Equation (2.14) is absorbed into $I(l,m)$. Recall that the Fourier transform, $\mathcal{F}\{\}$ of a function $f(x,y)$ can be written as:

$$\mathcal{F}\{f(x,y)\}(c_1,c_2) = \iint f(x,y)\mathrm{e}^{-2\pi i(xc_1+yc_2)}dx\ dy.$$

Since Equation (2.13) defines $V(u,v)$ as the Fourier transform of $I(u,v)$, we can also define $I(l,m)$ as the inverse Fourier transform of $V(u,v)$ given as:

$$\begin{aligned} I(l,m) &= \iint V(u,v)\mathrm{e}^{2\pi i(ul+vm)}du\ dv \\ &= \mathcal{F}^{-1}V(u,v), \end{aligned} \tag{2.17}$$

where $\mathcal{F}^{-1}$ is inverse Fourier transform and for the remaining part of this section, we take the normalised antenna response to be unity, so that the modified sky brightness $A'(l,m)I(l,m) = I(l,m)$.

Visibilities are not measured continuously but are only sampled at certain (u,v) points in the uv space. Let the sampling function be defined as $S(u,v)$ such that $S(u,v) = 1$ at every position where we have a measurement and 0 elsewhere, Equation (2.17) can be rewritten as:

$$I(l,m) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(u,v)V(u,v,w)\mathrm{e}^{2\pi i(ul+vm)}dl\ dm, \tag{2.18}$$

$$I(l,m) = \mathrm{PSF} \circ I_{\mathrm{true}}(l,m), \tag{2.19}$$

where $I(l,m)$ is the reconstructed image, $\circ$ denotes convolution, PSF (point spread function) $= \iint_{-\infty}^{\infty} S(u,v)\mathrm{e}^{2\pi i(ul+vm)}dl\ dm$ is the interferometer response to a 1 Jy point source and $I_{\mathrm{true}}(l,m)$ is the true image of the sky. The reconstructed image is known as the *dirty image* because it is the convolution of the true image of the sky and the PSF. To recover the true image of the sky, the PSF needs to be removed (deconvolved) from the dirty image. This process is generally called *cleaning* (Högbom, 1974; Cornwell and Wilkinson, 1981).

2.3.1 Visibility weighting

Imaging is an expensive (computational) operation, and most imaging algorithms use the Fast Fourier Transform (FFT) because it makes the process computationally feasible. To use FFT, the uv points need to be placed in a regular grid (called gridding). There are several weighting schemes which can be employed during gridding. The weighting schemes define how we treat the contribution of the different visibility points that fall in the same grid or uv cell.

The *natural weighting* scheme weighs each visibility point with the inverse of the variance, defined as:

$$W_k = \frac{1}{\sigma_k^2},$$

where σ_k is the standard deviation of the kth visibility. This weighting scheme maximises sensitivity with the more sampled short baselines, contributing the most to the final image. In order to optimise resolution, we can employ *uniform weighting*. *Uniform weighting* tries to give equal contribution to every uv-cell, hence the total contribution of each cell is normalised by its density (i.e. the total number of points in each cell). This scheme results in better resolution since it has significant contributions from the long baselines. For uniform weighting, the weights are defined as:

$$W_k = \frac{1}{D_k\sigma_k^2},$$

where σ_k is the standard deviation of the kth visibility and the number of samples in the cell containing the kth visibility is D_k. The *robust weighting* scheme, often called Briggs weighting, is a smooth tunable combination of *natural* and *uniform weighting*. In practice, the choice of weighting scheme is greatly controlled by the science of interest and array configuration. Other weighting schemes exist and may be used for numerous reasons, for example using a *Gaussian taper* to control the shape of the PSF.

2.4 Calibration

Practically speaking, the picture presented in Equation (2.13) is not as simple as shown because several factors corrupt signals received from sources as they travel from these sources to the antennas. We need to correct these visibilities through a process known as calibration before imaging. Calibration is required to correct for ionospheric, tropospheric and instrumental effects that have corrupted the measured visibilities. Many calibration algorithms have been developed over the years. A brief discussion of a few of them will be given in the next sections.

There are two types of propagation effects, namely direction-independent effects (DIEs) and direction-dependent effects (DDEs). DIEs are those propagation effects that remain constant across the field of view. Antenna gains are a good example of DIEs. DDEs are effects that change across the sky and so vary with the direction of each source. Primary beam effects and ionospheric effects are good examples of DDEs.

The next section presents a mathematical formalism which is used to model these propagation effects introduced into the measured visibilities (during observation) and the calibration used to mitigate them. Subsequent sections will discuss two calibration algorithms, then a brief discussion on calibration artefacts (systematics from calibration errors).

2.4.1 The radio interferometer measurement equation

Calibration is the process of finding the gains that minimise the difference between the predicted and measured visibilities. All the propagation effects can be modelled using the radio interferometer measurement equation (RIME) or simply the measurement equation (Hamaker *et al.*, 1996; Smirnov, 2011b). A simple mathematical derivation of the RIME is presented in this section.

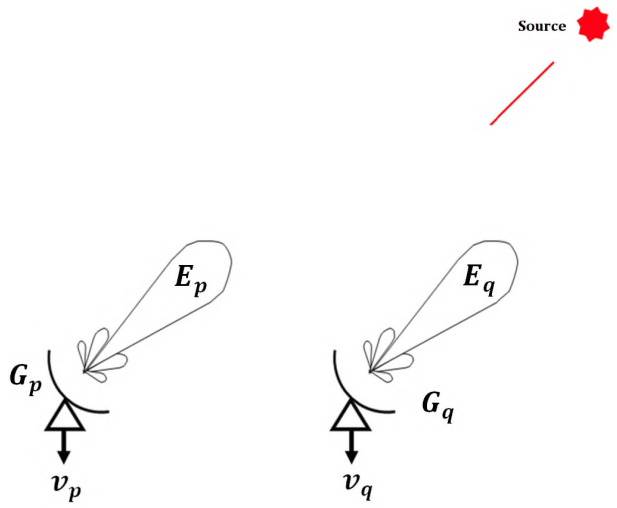

FIGURE 2.6: A schematic drawing of a two element interferometer - showing a source in the sky, the primary beams $(\boldsymbol{E}_p, \boldsymbol{E}_q)$, the antenna gains $(\boldsymbol{G}_p, \boldsymbol{G}_q)$ and the measured visibilities $(\boldsymbol{v}_p, \boldsymbol{v}_q)$.

Let us consider a sky that has a single quasi-monochromatic point source (i.e. a source that has size which is much smaller than the PSF size (Figure 2.6) whose radio signal can be described by a complex field vector, $\boldsymbol{e}$ propagating along the z-axis (i.e from the antennas to the source) in an orthonormal xyz coordinate system. The signal is given by:

$$\boldsymbol{e} = \begin{pmatrix} e_x \\ e_y \end{pmatrix} \tag{2.20}$$

This incoming signal is converted into complex voltages by the antenna feeds. Each antenna normally has two receivers that measure orthonormal polarisation modes. Linearly polarised dipoles or left and right circularly polarised feeds are good examples. The output voltage from the antenna can also be represented by a complex vector, $\boldsymbol{v}$ given as:

$$\boldsymbol{v} = \boldsymbol{J}\boldsymbol{e}, \tag{2.21}$$

where $\boldsymbol{J}$ is a 2×2 matrix called the Jones matrix, used to represent the cumulative product of all the effects along the signal path. A simple two-element interferometer measures the correlation of the output voltages from the two antennas, called the visibilities.

Consider Figure 2.6, $\boldsymbol{v}_p$ and $\boldsymbol{v}_q$ measured by antennas p and q respectively can be expressed as (we follow the terminologies, notations and derivations from Smirnov (2011b)):

$$\begin{aligned} \boldsymbol{v}_p &= \boldsymbol{J}_p \boldsymbol{e} \\ \boldsymbol{v}_q &= \boldsymbol{J}_q \boldsymbol{e}. \end{aligned} \tag{2.22}$$

The visibility matrix is given by:

$$\begin{aligned} \boldsymbol{V}_{pq} &= 2\langle \boldsymbol{v}_p \boldsymbol{v}_q^{\mathrm{H}} \rangle \\ &= 2\left\langle \begin{pmatrix} v_{px} \\ v_{py} \end{pmatrix} (v_{qx}^* v_{qy}^*) \right\rangle \\ &= 2\left\langle \begin{pmatrix} v_{px}v_{qx}^* & v_{px}v_{qy}^* \\ v_{py}v_{qx}^* & v_{py}v_{qy}^* \end{pmatrix} \right\rangle, \end{aligned} \tag{2.23}$$

where $\langle . \rangle$ is the expectation, $(.)^*$ is the complex conjugate and $(.)^{\mathrm{H}}$ is the conjugate transpose. We have followed the convention of multiplying Equation (2.23) by a factor of two as detailed in Smirnov (2011b). Substituting the definition of $\boldsymbol{v}$ in Equation (2.21) into Equation (2.23), we have:

$$\boldsymbol{V}_{pq} = 2\left\langle \begin{pmatrix} \boldsymbol{J}_p e_x e_x^* \boldsymbol{J}_q^{\mathrm{H}} & \boldsymbol{J}_p e_x e_y^* \boldsymbol{J}_q^{\mathrm{H}} \\ \boldsymbol{J}_p e_y e_x^* \boldsymbol{J}_q^{\mathrm{H}} & \boldsymbol{J}_p e_y e_y^* \boldsymbol{J}_q^{\mathrm{H}} \end{pmatrix} \right\rangle. \tag{2.24}$$

Assuming that $\boldsymbol{J}_p$ and $\boldsymbol{J}_q$ do not change over the averaging interval, we can write:

$$\boldsymbol{V}_{pq} = 2\boldsymbol{J}_p \begin{pmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{pmatrix} \boldsymbol{J}_q^{\mathrm{H}}. \tag{2.25}$$

In Equation (2.25), the terms in the brackets are related to the Stokes parameters (I, Q, U, V) and the brightness matrix, B via the expression (Born and Wolf, 1964; Hamaker *et al.*, 1996; Thompson *et al.*, 2008):

$$2\begin{pmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{pmatrix} = \begin{pmatrix} I+Q & U+iV \\ U-iV & I-Q \end{pmatrix} = \mathbf{B} \tag{2.26}$$

.

(taking the case of antennas having linearly polarised receivers). So, the measurement equation for a single source can be expressed as:

$$\boldsymbol{V}_{pq} = \boldsymbol{J}_p \mathbf{B} \boldsymbol{J}_q^{\mathrm{H}}. \tag{2.27}$$

We can expand the Jones matrices to describe the different propagation effects and the RIME for this source then becomes:

$$\boldsymbol{V}_{pq} = \boldsymbol{J}_{pm}(\boldsymbol{J}_{p(m-1)}(\ldots(\boldsymbol{J}_{p2}(\boldsymbol{J}_{p1}\mathbf{B}\boldsymbol{J}_{q1}^{\mathrm{H}})\boldsymbol{J}_{q2}^{\mathrm{H}})\ldots)\boldsymbol{J}_{q(n-1)}^{\mathrm{H}})\boldsymbol{J}_{qn}^{\mathrm{H}}. \tag{2.28}$$

n and m in Equation (2.28) may not always be the same because the number of propagation effects may vary per direction or antenna. Matrix multiplication is generally not commutative, hence there is a need to preserve the order of the matrices which is the order of occurrence of the propagation effects along the signal path. Equation (2.28) can be written for N sources as

$$\boldsymbol{V}_{pq} = \sum_s \boldsymbol{J}_{sp}\mathbf{B}_s\boldsymbol{J}_{sq}^{\mathrm{H}}. \tag{2.29}$$

where $\boldsymbol{J}_{sp}$ are the propagation effects experienced by the signal from source s on its path along the direction towards antenna p.

The propagation effects in Equation (2.28) can be represented with different kinds of matrices. Some of these propagation effects affect both components of the electric field vector, $\boldsymbol{e}$ equally while others affect them differently. Phase delay is a good example of propagation effect that affects both equally and can be represented with a scalar matrix (Smirnov, 2011b) as:

$$\mathrm{K}_p = 2\pi\lambda^{-1}(u_p l + v_p m + w_p(n-1)),$$

where λ is the wavelength of the signal and the direction cosine of $\boldsymbol{u}$ are $l, m, n = \sqrt{1-l^2-m^2}$ are the direction cosines and $\boldsymbol{u}$ is the baseline vector. We can further simplify this by expressing $\boldsymbol{u}$ in units of wavelength, thus dropping λ:

$$\mathrm{K}_p = 2\pi(u_p l + v_p m + w_p(n-1)).$$

Now, let the scalar K-Jones matrix representing the phase delay effect be given as $\mathrm{K}_p = e^{-2\pi i(u_p l + v_p m + w_p(n-1))}$, $\boldsymbol{G}_p$ be the source-independent antenna gain and $\boldsymbol{E}_{sp}$ be direction-dependent primary beam, where $n = \sqrt{1-l^2-m^2}$. The RIME can then be expressed as:

$$\boldsymbol{V}_{pq} = \boldsymbol{G}_p \sum_s \Big(\boldsymbol{E}_{sp} \boldsymbol{K}_{sp} \mathbf{B}_s \boldsymbol{K}_{sq}^{\mathrm{H}} \boldsymbol{E}_{sq}^{\mathrm{H}} \Big) \boldsymbol{G}_q^{\mathrm{H}}. \tag{2.30}$$

$\boldsymbol{G}$ can be brought out of the summation because they are direction-independent gains. Equation (2.30) is the measurement equation of a sky made up of point sources, but the real sky has a continuous brightness distribution, $\boldsymbol{B}(\boldsymbol{\sigma})$. To compute the total visibility measured by an interferometer we change the summation in Equation (2.30) to integration over all possible directions (i.e. over a unit sphere) and we have:

$$\boldsymbol{V}_{pq} = \int_{4\pi} \boldsymbol{J}_p(\boldsymbol{\sigma}) \boldsymbol{B}(\boldsymbol{\sigma}) \boldsymbol{J}_q^{\mathrm{H}}(\boldsymbol{\sigma}) d\Omega. \tag{2.31}$$

Let us perform a sine projection of the sphere onto a plane (l, m) tangential at the field centre. Hence,

$$\boldsymbol{V}_{pq} = \int_l \int_m \boldsymbol{J}_p(l,m) \boldsymbol{B}(l,m) \boldsymbol{J}_q^{\mathrm{H}}(l,m) \mathrm{e}^{-2\pi i(u_{pq} l + v_{pq} m + w_{pq}(n-1))} \frac{dl\ dm}{n}. \tag{2.32}$$

$\boldsymbol{J}_p(l,m)$ in Equation (2.32) can be decomposed into $\bar{\boldsymbol{E}}_p(l,m)$ direction-dependent and $\boldsymbol{G}_p$ direction-independent parts. Then we can write:

$$\boldsymbol{V}_{pq} = \boldsymbol{G}_p \Big(\int_l \int_m \frac{1}{n} \bar{\boldsymbol{E}}_p(l,m) \boldsymbol{B}(l,m) \bar{\boldsymbol{E}}_q^{\mathrm{H}}(l,m) \mathrm{e}^{-2\pi i(u_{pq} l + v_{pq} m + w_{pq}(n-1))} dl\ dm \Big) \mathbf{G}_q. \tag{2.33}$$

We cannot treat Equation (2.33) as a 2-dimensional Fourier transform because of the presence of the non co-planarity term, $w_{pq}(n-1)$ in the exponent, but we can decompose it into per antenna terms $W_p = \frac{1}{\sqrt{n}} \mathrm{e}^{-2\pi i(w_p(n-1))}$, which can be treated as a direction-dependent Jones term. We can combine it with the overall sky-Jones terms. We then define a new Jones term using this decomposition as $\boldsymbol{E}_p = \bar{\boldsymbol{E}}_p W_p$ and we can further define a new term for the apparent sky $\mathbf{B}$ as $\mathbf{B}_{pq} = \boldsymbol{E}_p \mathbf{B} \boldsymbol{E}_q^{\mathrm{H}}$, where $\mathbf{B}_{pq}$ is the apparent sky brightness as seen by an interferometer on the baseline pq. Finally, applying this new definition to Equation (2.33), we have that:

$$\boldsymbol{V}_{pq} = \boldsymbol{G}_p\Big(\int_l\int_m \mathbf{B}_{pq}\mathrm{e}^{-2\pi i(u_{pq}l+v_{pq}m)}dl\ dm\Big)\mathbf{G}_q^{\mathrm{H}}. \tag{2.34}$$

The above equation presents the RIME as a two-dimensional Fourier Transform of the apparent sky as seen by the interferometer baseline pq. Equation (2.34) is the full-sky RIME and is simply the Van Cittert-Zernike theorem (Thompson *et al.*, 2008). As stated earlier, calibration is a (complex) optimisation problem where one uses the modelled (or predicted visibilities obtained from the RIME) and measured visibilities to estimate the propagation effects. The measured visibilities are also contaminated by additive thermal noise, so one can rewrite Equation (2.34) to accommodate this as:

$$\boldsymbol{V}_{pq} = \boldsymbol{G}_p\Big(\int_l\int_m \mathbf{B}_{pq}\mathrm{e}^{-2\pi i(u_{pq}l+v_{pq}m)}dl\ dm\Big)\mathbf{G}_q^{\mathrm{H}} + \mathbf{N}_{pq}. \tag{2.35}$$

The last term $\boldsymbol{N}_{pq}$ in Equation (2.35) is the term representing the thermal noise contribution.

The next section will discuss a few techniques that have been used to model the propagation effects and estimate them. A good calibration algorithm will not only model all the Jones matrices and calibrate the measured data but also provide insightful feedback on the state of the antenna at a reduced computational cost.

2.4.2 Optimisation algorithms

The RIME formalism and its derivation were shown in the previous section. Calibration is simply an optimisation problem. Many optimisation algorithms have been developed and used for calibration. Most (least squares optimisation) algorithms optimise by (iteratively) finding a local minimiser of a cost function. Most often, the cost function is the L2-norm of the difference between the measured and model data. In general, for a function $F : \mathtt{R}^n \mapsto \mathtt{R}$, most minimisation problems attempt to find an argument, $\mathbf{x}^*$ that produces the minimum value of the cost function to obtain the optimum estimate of the data (Madsen *et al.*, 2004). In this section, we present a brief introduction to optimisation algorithms with emphasis on Gauss-Newton (GN) and Levenberg-Marquardt (LM) (and we follow the derivation in Madsen *et al.* (2004)).

In non-linear least squares problems, if we assume a vector function $\mathbf{f} : \mathtt{R}^n \mapsto \mathtt{R}^m$ having $m \geq n$, one seeks to minimise $\parallel \mathbf{f}(\mathbf{x}) \parallel$, to get $\mathbf{x}^*$:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}}\{F(\mathbf{x})\}, \text{ where} \tag{2.36a}$$

$$F(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{m}(\mathbf{f}_i(\mathrm{x}))^2 = \frac{1}{2}\mathbf{f}(\mathbf{x})^{\mathrm{T}}\mathbf{f}(\mathbf{x}). \tag{2.36b}$$

Note that the factor of $\frac{1}{2}$ in Equation (2.36b) is added for convenience and $\| \cdot \|$ represents the L2-norm. Assuming that F is differentiable and $\mathbf{f}$ has a continuous second partial derivative then we have its *Taylor expansion* as:

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + \mathcal{J}(\mathbf{x})\mathbf{h} + O(\| \mathbf{h} \|^2), \tag{2.37}$$

where $\mathcal{J} \in \mathtt{R}^{m\times n}$ and $\mathcal{J}_{ij}(\mathbf{x}) = \frac{\partial f_{\mathrm{i}}}{\partial x_{\mathrm{j}}}(\mathbf{x})$ is the *Jacobian.* (For sufficiently small $\mathbf{h}$, the contribution of the higher-order terms of its *Taylor expansion* is very small and could be ignored, so we limited Equation 2.37 to second-order terms.)

From the Equation (2.36b), for a function $F : \mathtt{R}^n \mapsto \mathtt{R}^m$ we have that:

$$\frac{\partial F(\mathbf{x})}{\partial x_i} = f_i(\mathbf{x})\sum_{i=1}^{\mathrm{m}}\frac{\partial f_i}{\partial x_i}(\mathbf{x}), \tag{2.38}$$

which implies that the gradient can be written as:

$$\mathbf{F}'(\mathbf{x}) = \mathcal{J}(\mathbf{x})^{\mathrm{T}}\mathbf{f}(\mathbf{x}). \tag{2.39}$$

Also, the second derivative of Equation (2.36b) is given as:

$$\frac{\partial^2 F(\mathbf{x})}{\partial x_i \partial x_k} = \sum_{i=1}^{m}\Big(\frac{\partial f_i}{\partial x_i}(\mathbf{x})\frac{\partial f_i}{\partial x_k}(\mathbf{x}) + f_i(\mathrm{x})\frac{\partial^2 f_{\mathrm{i}}(\mathbf{x})}{\partial x_i \partial x_k}\Big), \tag{2.40}$$

which implies the Hessian of F can be written as:

$$\mathbf{F}''(\mathbf{x}) = \mathcal{J}(\mathbf{x})^{\mathrm{T}}\mathcal{J}(\mathbf{x}) + \sum_{\mathrm{i}=1}^{\mathrm{m}}\mathrm{f}_{\mathrm{i}}(\mathrm{x})\mathbf{f}_i''(\mathbf{x}). \tag{2.41}$$

2.4.2.1 Gauss-Newton (GN) and Levenberg-Marquardt (LM)

The Gauss-Newton and Levenberg-Marquardt algorithms are two common least-squares minimisation methods. We will give a brief discussion on both.

Gauss-Newton: The GN algorithm is an optimisation algorithm based on the implementation of the first derivatives of the vector function and it optimises by linear approximations of the components of the function F in the region near $\mathbf{x}$: from Equation (2.37), for small update values of $\| \mathbf{h} \|$, the Taylor expansion can be written as:

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) \simeq l(\mathbf{h}) \equiv \mathbf{f}(\mathbf{x}) + \mathcal{J}(\mathbf{x})h. \tag{2.42}$$

Using the definition of F from Equation (2.36) we have:

$$\begin{aligned} \mathbf{F}(\mathbf{x}+\mathbf{h}) \simeq L(\mathbf{h}) &= \frac{1}{2} l(\mathbf{h})^{\mathrm{T}} l(\mathbf{h}) \\ &= \frac{1}{2}\mathbf{f}^{\mathrm{T}}\mathbf{f} + \mathbf{h}^{\mathrm{T}}\mathcal{J}^{\mathrm{T}}\mathbf{f} + \frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathcal{J}^{\mathrm{T}}\mathcal{J}\mathbf{h} \\ &= F(\mathbf{x}) + \mathbf{h}^{\mathrm{T}}\mathcal{J}^{\mathrm{T}}\mathbf{f} + \frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathcal{J}^{\mathrm{T}}\mathcal{J}\mathbf{h}, \end{aligned} \tag{2.43a}$$

where $\mathbf{f} = \mathbf{f}(\mathbf{x})$ and $\mathcal{J} = \mathcal{J}(\mathbf{x})$. For the GN step where $\mathbf{h}_{\mathrm{gn}}$ minimises $L(\mathbf{h})$, we have that:

$$\mathbf{h}_{\mathrm{gn}} = \underset{\mathbf{h}}{\operatorname{argmin}}\{L(\mathbf{h})\}. \tag{2.44}$$

From Equation (2.39) and (2.41) we can write the gradient and Hessian of L as:

$$\mathbf{L}'(\mathbf{h}) = \mathcal{J}^{\mathrm{T}}\mathbf{f} + \mathcal{J}^{\mathrm{T}}\mathcal{J}\mathbf{h} \quad \text{and} \quad \mathbf{L}''(\mathbf{h}) = \mathcal{J}^{\mathrm{T}}\mathcal{J}, \quad \text{respectively.} \tag{2.45}$$

It is interesting to note that at $\mathbf{h} = 0$, $\mathbf{L}'(\mathbf{h} = 0) = \mathbf{F}'(\mathbf{x})$ and also, $\mathbf{L}''(\mathbf{h})$ is independent of $\mathbf{h}$. If the columns of the $\mathcal{J}$ are linearly independent and $\mathbf{L}''(\mathbf{h})$ is positive definite, then one can get a unique minimiser for $L(\mathbf{h})$ by solving:

$$(\mathcal{J}^{\mathrm{T}}\mathcal{J})\mathbf{h} = -\mathcal{J}^{\mathrm{T}}\mathbf{f}. \tag{2.46}$$

Levenberg-Marquardt: The LM method (Levenberg, 1944; Marquardt, 1963) is another minimisation technique used to solve least-squares problems. It follows the same approximation method for the Hessian matrix as the GM but introduces a damping factor. Hence, it can be described as a damped GN method. The damping factor significantly improves the efficiency and the convergence stability of this algorithm. For LM, Equation (2.46) is modified as follows:

$$(\mathcal{J}^{\mathrm{T}}\mathcal{J} + \mu\boldsymbol{H})\mathbf{h} = -\mathcal{J}^{\mathrm{T}}\mathbf{f}, \tag{2.47}$$

where μ is the damping parameter and $\boldsymbol{H}$ is the diagonal matrix of $(\mathcal{J}^{\mathrm{T}}\mathcal{J})$.

Now that a general introduction to optimisation algorithms have been given, let us briefly discuss the uses of GN and LM in the context of radio astronomy calibration.

2.4.2.2 Calibration with Gauss-Newton and Levenberg-Marquardt

Following our discussion of the GN and LM algorithms in Section 2.4.2.1, we now describe how these algorithms can be applied in the context of calibration. Recall from Equation (2.35) that we can write the visibilities for two antenna pairs p and q of an interferometer as:

$$\boldsymbol{V}_{pq} = \sum_{i=1}^{K} \boldsymbol{J}_{pi}\mathbf{B}_{pqi}\boldsymbol{J}_{qi}^{\mathrm{H}} + \boldsymbol{N}_{pq}, \tag{2.48}$$

where $\boldsymbol{V}_{pq}$ is the measured visibilities which is a 2×2 matrix of complex numbers, K is the number of sources in the sky, $\mathbf{B}_{pqi}$ is also a 2×2 matrix of complex numbers representing the ith source brightness and Jones matrices, $\boldsymbol{J}(\boldsymbol{J} \in \mathbb{C}^{2\times 2})$ which represents the propagation effects. An interferometer with N number of antennas, has $\frac{N(N-1)}{2}$ independent pairs whose visibilities are measured with Equation (2.48). Normally, observations are averaged within short time interval, τ and frequency where the components of Equation (2.48) are assumed to remain the same. To formulate our optimisation problem, we start by writing Equation (2.48) in vector form as:

$$\mathbf{vec}(\boldsymbol{V}_{pq}) = \sum_{i=1}^{K} \mathbf{vec}(\boldsymbol{J}_{pi}\mathbf{B}_{pqi}\boldsymbol{J}_{qi}^{\mathrm{H}}) + \mathbf{vec}(\boldsymbol{N}_{pq}), \tag{2.49a}$$

$$\mathbf{vec}(\boldsymbol{V}_{pq}) = \sum_{i=1}^{K} (\boldsymbol{J}_{qi}^{*} \otimes \boldsymbol{J}_{pi})\mathbf{vec}(\mathbf{B}_{pqi}) + \mathbf{vec}(\boldsymbol{N}_{pq}), \tag{2.49b}$$

$$\boldsymbol{u}_{pq} = \sum_{i=1}^{K} \boldsymbol{s}_{pqi} + \boldsymbol{n}_{pq}, \tag{2.49c}$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{u}_{pq} = \mathbf{vec}(\boldsymbol{V}_{pq})$, $\boldsymbol{s}_{pqi} = (\boldsymbol{J}_{qi}^{*} \otimes \boldsymbol{J}_{pi})\mathbf{vec}(\mathbf{B}_{pqi})$, $\boldsymbol{n}_{pq} = \mathbf{vec}(\boldsymbol{N}_{pq})$, $\mathbf{vec}(.)$ is the vectorisation operation used to transform a matrix to a column vector, $(.)^{\mathrm{T}}$ denotes matrix transpose and $(.)^{\mathrm{H}}$ denotes matrix conjugate transpose. To transform Equations (2.49a) to (2.49b), we used the vectorisation property which states that for any three matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ of the correct shapes, $\mathbf{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\mathbf{vec}(\boldsymbol{B})$.

Now, because the optimisation parameters, $\boldsymbol{J}$ are complex numbers, we will split them into their real and imaginary parts to form up the parameter vector:

$$\Theta = [\mathrm{Re}(\boldsymbol{J}_{11}), \mathrm{Im}(\boldsymbol{J}_{11}), ...]. \tag{2.50}$$

Also, in similar manner, we separate the visibility matrix (from Equation (2.49)) into real and imaginary parts.

$$\boldsymbol{s}_i(\Theta) = [\mathrm{Re}(\boldsymbol{s}_{12i}^{\mathrm{T}}), \mathrm{Im}(\boldsymbol{s}_{12i}), \mathrm{Re}(\boldsymbol{s}_{13i}^{\mathrm{T}}), ...]^{\mathrm{T}} \tag{2.51}$$

$$\boldsymbol{y} = [\mathrm{Re}(\boldsymbol{v}_{12}^{T}), \mathrm{Im}(\boldsymbol{v}_{12}), \mathrm{Re}(\boldsymbol{v}_{13}^{T}), ...]^{\mathrm{T}}, \tag{2.52}$$

where $\boldsymbol{s}_i(\Theta)$ and $\boldsymbol{y}$ are the measured and model visibilities respectively. Let the cost function be defined as:

$$\begin{aligned} \mathbf{F}(\boldsymbol{\Theta}) &= \|\boldsymbol{y} - \sum_{i}^{k} \boldsymbol{s}_i(\Theta)\|^2, \\ &= \|f(\Theta)\|^2, \end{aligned} \tag{2.53}$$

where $f(\Theta) = \|\boldsymbol{y} - \sum_i^k \boldsymbol{s}_i(\Theta)\|^2$. From Equations (2.46) and (2.47), the update step for the parameters are defined as follows for the GN and LM algorithms respectively.

$$\text{GN}: \quad (\boldsymbol{\mathcal{J}}^{\mathrm{T}}\boldsymbol{\mathcal{J}})\mathbf{h} = -\boldsymbol{\mathcal{J}}^{\mathrm{T}} f(\Theta), \tag{2.54a}$$

$$\text{LM}: \quad (\boldsymbol{\mathcal{J}}^{\mathrm{T}}\boldsymbol{\mathcal{J}} + \mu\boldsymbol{H})\mathbf{h} = -\boldsymbol{\mathcal{J}}^{\mathrm{T}} f(\Theta), \tag{2.54b}$$

where $\boldsymbol{\mathcal{J}}_{ij} = \frac{\partial f_i(\Theta)}{\partial \Theta}$ represents the Jacobian matrix, μ is the damping parameter and $\boldsymbol{H}$ is the diagonal matrix of $(\boldsymbol{\mathcal{J}}^{\mathrm{T}}\boldsymbol{\mathcal{J}})$. Then, the GN and LM update $\mathbf{h}$ are expressed as:

$$\text{GN}: \quad \mathbf{h} = -(\boldsymbol{\mathcal{J}}^{\mathrm{T}}\boldsymbol{\mathcal{J}})^{-1}\boldsymbol{\mathcal{J}}^{\mathrm{T}} f(\Theta) \tag{2.55a}$$

$$\text{LM}: \quad \mathbf{h} = -(\boldsymbol{\mathcal{J}}^{\mathrm{T}}\boldsymbol{\mathcal{J}} + \mu\boldsymbol{H})^{-1}\boldsymbol{\mathcal{J}}^{\mathrm{T}} f(\Theta), \tag{2.55b}$$

Hence, implying that from some initial value, Θ_0 the estimate Θ^{k+1} at the $(k+1)^{\text{th}}$ iteration is given by:

$$\text{GN}: \quad \Theta^{k+1} = \Theta^k - \alpha(\boldsymbol{\mathcal{J}}^{\mathrm{T}}\boldsymbol{\mathcal{J}})^{-1}\boldsymbol{\mathcal{J}}^{\mathrm{T}} f(\Theta) \tag{2.56a}$$

$$\text{LM}: \quad \Theta^{k+1} = \Theta^k - \alpha(\boldsymbol{\mathcal{J}}^{\mathrm{T}}\boldsymbol{\mathcal{J}} + \mu\boldsymbol{H})^{-1}\boldsymbol{\mathcal{J}}^{\mathrm{T}} f(\Theta), \tag{2.56b}$$

where α is the rate parameter used to control the size of each update value.

2.4.3 The development of calibration

From the early time of the discovery of radio astronomy, many calibration techniques have been developed to solve various challenges that emerged due to the complex design of the new radio astronomy instruments. These techniques may be put into three broad classes; first, second and third generation calibration techniques (Noordam and Smirnov, 2010). We will briefly discourse these categories of calibration techniques.

First generation calibration (1GC) is the first step in data reduction and was the principal calibration procedure used before the 1980s (Noordam and Smirnov, 2010). 1GC calibration techniques address the following errors.

- Absolute flux density calibration - where the observed fluxes of the sources are scaled by some factor to match their true fluxes.
- Bandpass calibration - this is used to correct for the variations in the gain along the frequency.
- Delay calibration - this is used to correct for the phase delay errors.
- Gain calibration - this is used to correct for complex antenna gains.

In this approach, calibrator sources (i.e. sources with known positions and fluxes) are used to compute the gains and then apply these gains to the target field. We obtain this by doing alternating observations of the target source and calibrator source. One or more calibrator sources may be used to address various 1GC calibration errors and must have the following properties (Thompson *et al.*, 2008):

- The source used as the absolute flux density calibrator should be bright, so that high signal-to-noise ratio can be obtained within a short observation time. This reduces the amount of time it takes to observe the calibrator source to enable more time to be used in observing the target source.
- It is preferred that the calibrator source should be unresolved.
- It is required that the calibrator source should be close enough to the target field so that both of them will experience the same atmospheric effects.
- The calibrator used for bandpass calibration should be bright with approximately a flat spectrum.

The second generation calibration (2GC) techniques started in the early eighties, after the development of the self-calibration (selfcal) (Cornwell and Wilkinson, 1981). Self-calibration, as its name implies, is an iterative calibration algorithm where the observed field is used to calibrate itself. The selfcal algorithm is described as follows:

- First, create an initial sky model from the 1GC calibrated data.
- Then, set that model to be your new initial sky model and use it to calibrate the observed visibilities.
- Next, image the corrected visibilities.

- Next, run a source finder on this image and make a new sky model from it (or apply the CLEAN components directly i.e. using the model visibilities generated by the imager directly).
- Then, return to Step 2 with this most recent sky model as the new sky model or stop if no further improvement is seen (or you are satisfied with the dynamic range value you have) from imaging the corrected visibilities.

The first and second generation calibration techniques discussed above are both generally referred to as classical calibration, they concentrate primarily on accounting for direction-independent effects (DIEs) and neglect direction-dependent effects (DDEs). Recently, with the construction of new and future telescopes having a wide field of views like those of the Square Kilometre Array (SKA), Low Frequency Array (LOFAR) (van Haarlem *et al.*, 2013), Expanded Very Large Array (EVLA) (Perley et al., 2011), Low Frequency Array (LOFAR) (van Haarlem et al., 2013), DDEs can not be ignored any more. These brought new calibration challenges to the community which led to the development of the third generation calibration (3GC) techniques. 3GC techniques are those calibration techniques that account for DDEs during calibration. These techniques can be broadly put into two classes (based on the method of development), physics-based and heuristic-based approaches.

Physics-based approaches are applied when the underlying physical phenomena causing the DDEs are known. The known physical phenomena causing the DDEs are modelled, parametrised and incorporated into the RIME formalism during calibration. Pointing-selfcal (Bhatnagar *et al.*, 2004), is one technique which uses a physics-based approach to perform direction-dependent calibration.

Heuristic-based approaches are used when there is no existing physical model for a particular DDE. In this case, sources affected by the DDE are identified and an extra Jones term is added to account for it, which are then solved for during calibration. One of the techniques used in doing this is called peeling (Noordam, 2004). The peeling process solve for DDEs by accounting for the effects towards each source separately in the order of decreasing brightness. Every time the DDEs are solved for towards a source, the source is subtracted and the whole process is repeated for the next brightest source. Peeling is a computationally expensive algorithm and the gain solutions obtained are contaminated sometimes, but a method has been proposed called differential gains, that simultaneously solve for the DIEs and DDEs (Smirnov, 2011b) to mitigate this gain contamination. The RIME that incorporates the differential gain is written as:

$$\mathbf{V}_{pq} = \boldsymbol{G}_p \left(\sum_s \Delta \boldsymbol{E}_{\boldsymbol{sp}} \mathbf{B}_{spq} \Delta \boldsymbol{E}^{\mathrm{H}}_{sq} \right) \boldsymbol{G}^{\mathrm{H}}_q, \tag{2.57}$$

where $\Delta \boldsymbol{E}_{sp}$ and $\Delta \boldsymbol{E}^{\mathrm{H}}_{sq}$ are the differential gains associated with source s and p and q are the interferometer antenna indexes. Mitra *et al.* (2015) used this approach to obtain a dynamic range of 5,000,000:1. They modelled the primary beam, incorporated it into the calibration reduction and applied the differential gain method to the VLA data of the 3C147 field.

Some other calibration algorithms have also been developed. An example is the alternating least squares (ALS) (Boonstra and Van Der Veen, 2001) algorithm which optimises by alternates between solving for the noise and solving for the gains. Robust calibration (Kazemi and Yatawatta, 2013) an approach which models the noise with a t-distribution to account for the unmodelled sources and yet another, is calibration using the expectation maximisation algorithm (Kazemi *et al.*, 2011). Also, another algorithm is the non-linear optimisation on a Riemannian Manifold (Yatawatta, 2013). Recent work by Tasse *et al.* (2018) has revived the old facet-based imaging idea. Here, the field of view is divided into small facets, which are imaged individually while applying a primary beam correction (and an optional DD-gain solution) for the centre of each facet, and the mosaiced into a single image.

There are some artificial features which may be introduced into the data as a result of calibration errors. We will discussion this briefly.

Artefacts: In the context of radio astronomy, artefacts or calibration artefacts are systematic features that result from calibration errors. They produce imperfections in images, suppress or eliminate real source components, generate spurious source components and deform the structure of extended sources etc. (Wilkinson *et al.*, 1988; Linfield, 1986; Martí-Vidal and Marcaide, 2008). In Figure 2.7, we presents a simple case of what artefacts looks like. These are features that we see around the source by the bottom right quadrant of the figure. Work done by Barry *et al.* (2016); Wilkinson *et al.* (1988); Grobler *et al.* (2014); Wijnholds *et al.* (2016); Patil *et al.* (2016) have extensively explained the formation of a particular class of artefacts known as *ghost sources*. These result from errors in the sky model - missing sources or wrong source flux in the calibration sky model and wrong models of various propagation effects, for example, the antennas primary beam or tropospheric and ionospheric effects.

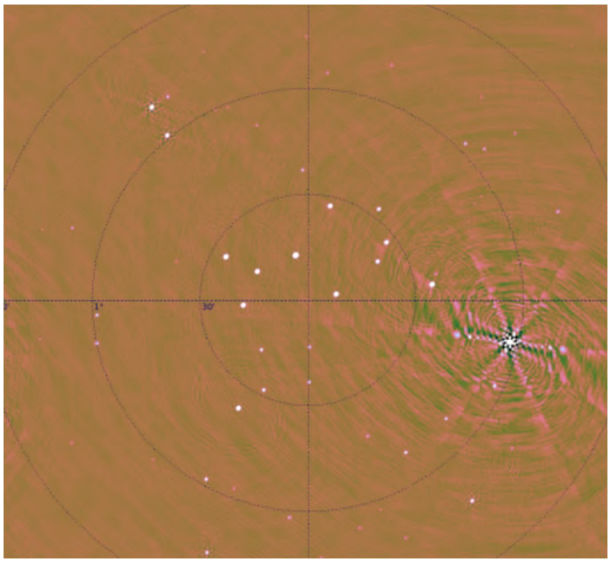

FIGURE 2.7: Image of sources in a field-of-view showing a source by right hand side of the figure having a lot artificial features around it known as artefacts.

2.5 Summary

In this chapter, we introduced and discussed the general calibration problem. A brief introduction and derivation of the RIME was presented. During radio astronomy observation, environmental and instrumental effects perturb the signals as they move from the source to the observing antennas, known as propagation effects. These propagation effects are put into two groups, the DIE and DDE. These need to be modelled and accounted for during calibration to recover the original signal coming from the source. A brief discussion on antennas and its designs was also given. Calibration algorithms were briefly reviewed.

We also discussed the interferometer and identified the primary beam of an antenna as one of the instrumental effects that needs to be modelled and accounted for in calibration. In chapters 3 and 4 we will use these techniques to measure and model the primary beam of a radio telescope.

Chapter 3

Measuring the beam pattern - Holography

3.1 Introduction

Recall that in the RIME formalism, we saw that in order to recover the true source brightness we need to account for propagation effects that affect the signal which is recorded at the feeds of the antenna. The antenna radiation pattern or the primary beam (PB) - is one of these propagation effects. In this chapter, we will discuss a method of estimating the PB so that it can be accounted for during calibration. Most sections of this chapter have been published in Iheanetu *et al.* (2019).

Direction-dependent effects (DDEs) generally affect wide-field observations, and one of the most important DDEs is the PB. Not having a good model of the PB limits the accuracy with which these effects can be removed and hence the dynamic range that can be reached. One approach of obtaining a good knowledge of the antenna PB is by measuring it using holography techniques. This chapter focuses on describing the holography measurements of the Very Large Array (VLA) PB patterns which are used throughout this work. Section 3.2 gives a brief introduction to holography and describe the measurements. The descriptions (of holography measurements) made here are based entirely on EVLA Memo 195 (Perley, 2016). Furthermore, we will also discuss how the measured beams are described mathematically using the Mueller matrix formalism in Section 3.3.1. In Section 3.4, we present our analysis of the measured beams.

3.2 Holography

In its original sense, holography is the practice of recording the interference pattern formed by coherent light that is reflected off a target object and some reference light on film, which can then be used to recreate the optical field of the object with all the three-dimensional information reproduced. In radio astronomy, we use the interferometer technique to perform holography measurement of the antenna primary beam.

The essence of antenna holographic measurements using an interferometer is to infer the actual shape of the normalised complex PB of a radio antenna. The word 'normalised', implies that the resultant measurements are scaled relative to the on-axis intensity which has an amplitude of 1.0 and a 0 phase at the centre. A holography beam measurement is obtained by correlating the voltages produced by 'moving' and 'fixed' antennas. The 'fixed' or reference antenna points to (or traces) a strong point source (known as the calibrator), while the 'moving' or target antenna scans the vicinity of the source in a regular (l, m) two-dimensional grid centred on the moving source. The complex product between voltages from a reference and target antenna gives a direct measure of the normalised voltage beam (see Iheanetu *et al.* (2019) and Perley (2016) for details).

The baselines of tracking antennas can be set up to operate as *fixed-moving* or as *moving-moving* baselines. The operation mode of choice depends on the desired output. A *fixed-moving* mode is one in which one antenna is fixed on a source while the other is allowed to scan regions centred on the source. This mode is used to get antenna parameters and has been used to improve the efficiency of antennas. The *moving-moving* mode is one in which both antennas are moving. This mode allows one to see what the source looks like to the baseline. A *fixed-fixed* mode is used to observe the calibrator source only. The first mode was directly used to perform the holography measurement. Given that the reference antenna tracks a known source that is kept at its centre, the target antenna scans the vicinity of the source in a $\Delta l \Delta m$ grid that is centred on the source. This interferometer setup causes the visibility measurement between one target and one reference antenna to be modulated due to the pattern multiplication of the respective antenna beams. Therefore, the radio source is used to probe the beam of the target antenna. The holography measurement for the beams is obtained by using the correlator product between the fixed and scanning antennas. The correlator output (although a product of the voltages of the two antennas) has variations that give the complex voltage of the target antenna beam, provided the reference antenna remains fixed on the source. Variations in the atmospheric and electronic gains are mostly removed

by on-source calibration of the two antennas. A (technical) detailed description of the holography procedure is given in Iheanetu *et al.* (2019).

3.3 A selected sample of the holography measurements

In the previous section, we presented the method and procedure used in performing the holography measurements. This section will present some of these measured data and a first-order analysis. Detailed analysis will be provided in Chapter 4.

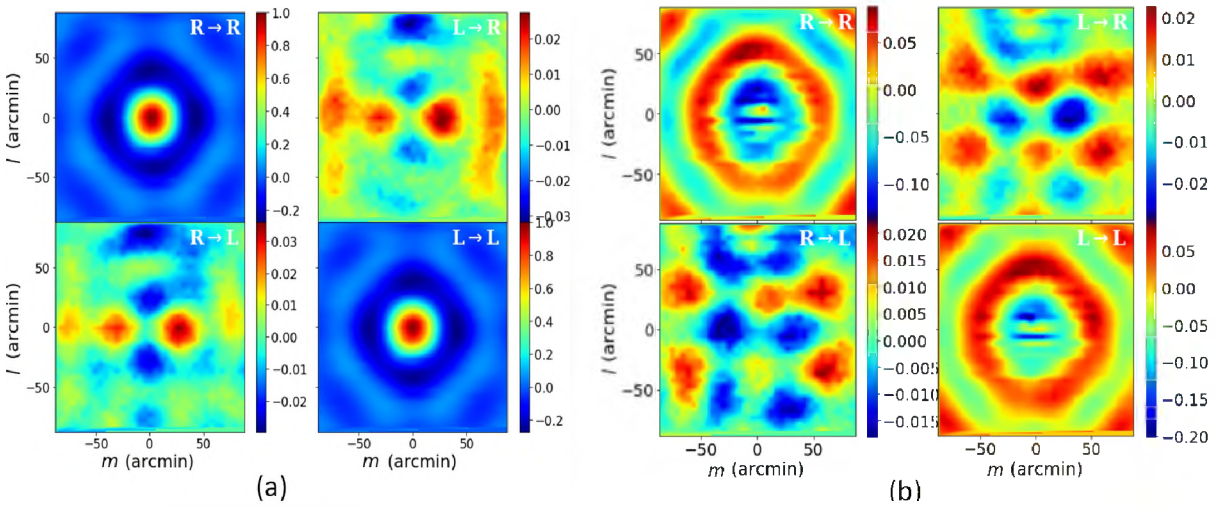

FIGURE 3.1: Jones matrix plots of the two-dimensional holography measured beams for one antenna at 1.708 GHz. The (a) real and (b) imaginary parts. The diagonal plots are RR (top left plot) and LL (bottom right plot) while off-diagonal plots are LR (top right plot) and RL (bottom left plot) in each panel. These beams each have a FOV of approximately 3 degrees.

The feeds of the VLA antennas are circular polarised and thus measure right-circular-polarised (RCP), left-circular-polarised (LCP) signals and their cross-correlations (representative of the crosstalk or polarisation leakage). The PB measurements were stored in the FITS file format. The antennas have complex beams: two files were created for each polarisation (one for real and another for imaginary beams), giving a total of eight files per antenna (for LL, RL, LR and RR polarisations). Two-dimensional plots of the holography measured (or holography) beams for one of the antennas at a randomly selected frequency of 1.708 GHz are presented in Figure 3.1. The morphology of the diagonal plots in Figure 3.1(a) shows approximately circular main lobes, surrounded by a null.

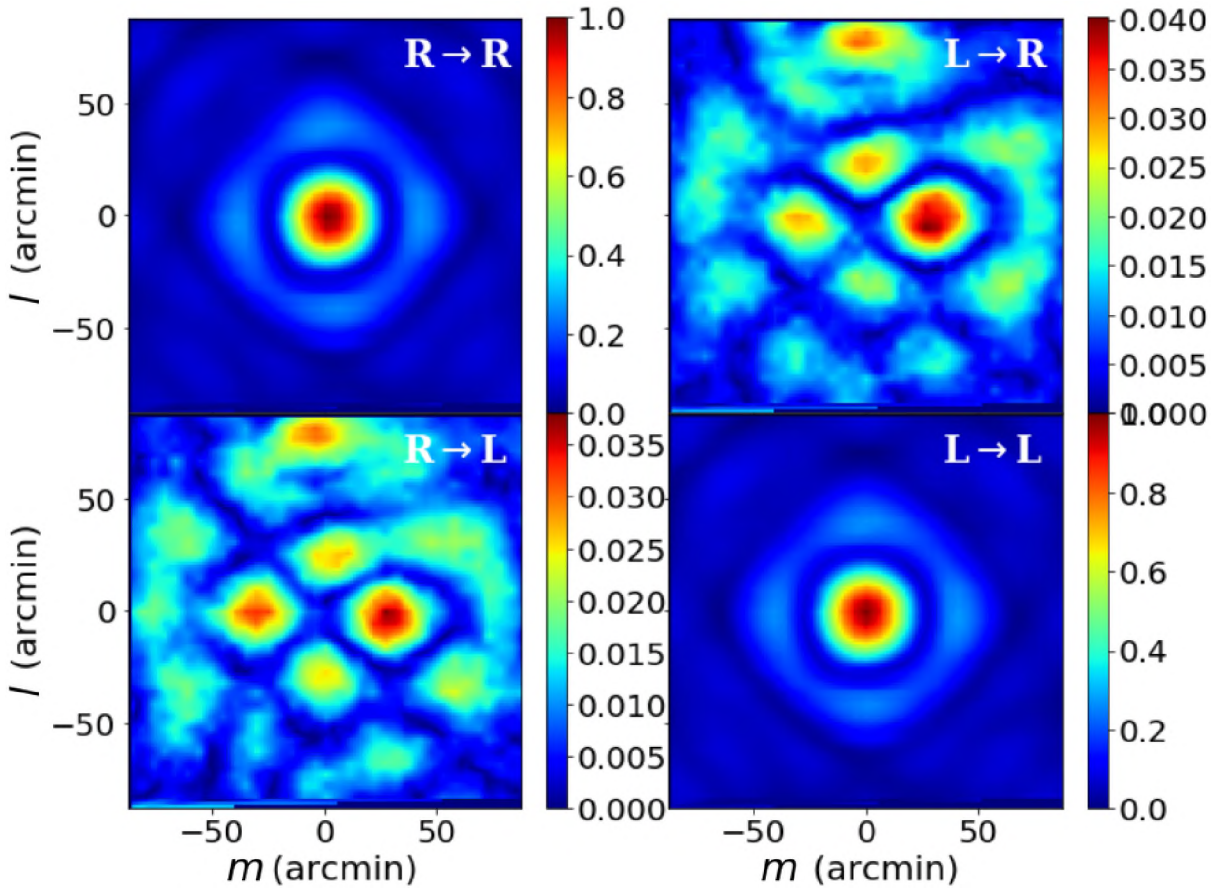

FIGURE 3.2: Jones matrix plots of the two-dimensional holography beams amplitude for one antenna at 1.708 GHz (calculated using figure 3.1). The diagonal plots are RR (top left plot) and LL (bottom right plot) while off-diagonal plots are LR (top right plot) and RL (bottom left plot)

The value of the gain varies from the maximum value (of 1) at the main lobe to a minimum (of approximately -0.2) and back to a relatively high value (of 0.05) in the first side lobe. The morphology of the off-diagonal plots (corresponding to leakage beams) shows a set of four petal-like features, in a diamond shape layout around the centre, and the morphology gets more complex further away from the centre. The imaginary part of the beams are shown in Figure 3.1(b). The morphology of the diagonal terms (of the imaginary beams) can be described to have an asymmetric main lobe surrounded by a doughnut-shaped structure that corresponds to the first side lobe, and other structures at the corners of the beams which may correspond to the third side lobe. The off-diagonal plots have four petal-like structures again arranged in a diamond-shaped layout and other shapes in other lobes.

Figure 3.2 shows plots of the amplitude beams (for the beams displayed in Figure 3.1). The morphology of the diagonal terms of the amplitude beams in Figure 3.2 is asymmetric

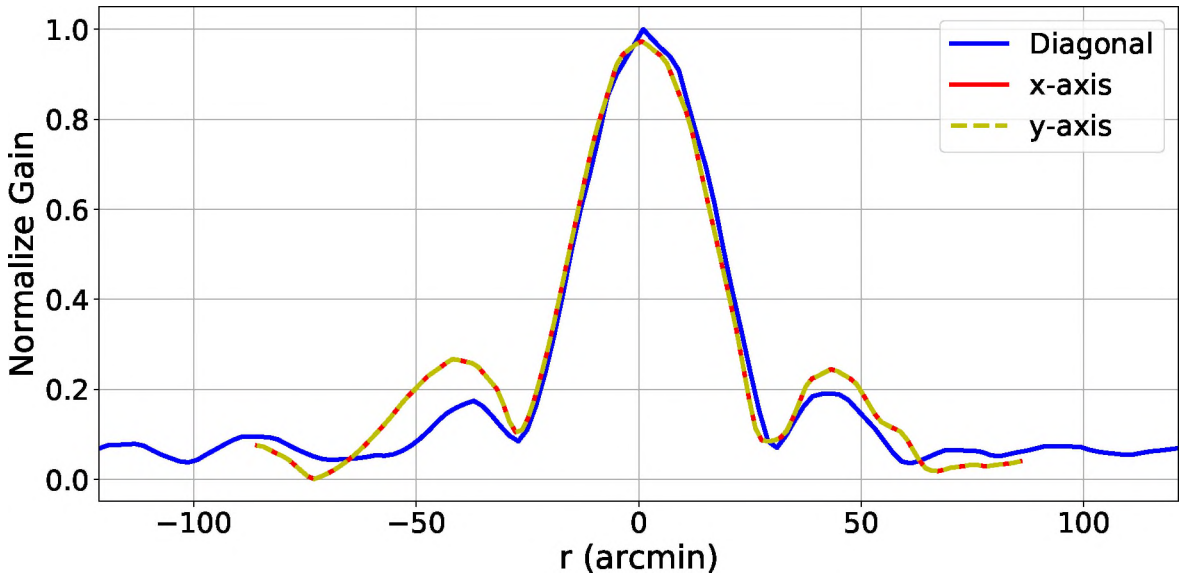

FIGURE 3.3: One-dimensional plots of the normalised amplitude beam at 1.708 GHz along the diagonal (45 degrees), x- and y-axis (for RR).

(only approximately circularly symmetrical in shape at the main lobe but not in the side-lobes). The first side lobe has four-fold symmetry. Recall from Chapter 2 that the VLA antennas are prime focus (off-axis) Cassegrain systems, that have their sub-reflector assembly and its support structures in the line of sight of the signal received by the main collector. The sub-reflector, its assembly and the support structures have both blocked and scattered some signal into the side-lobes. Since the wavelength of the L-band (approximately 0.15 - 0.30 m) is comparable to the cross-sectional size (0.27 m) of the support struct, it constituted a source of diffraction which has spread some signals into the side lobes (that also help to create asymmetry in the beams). All these create a diamond-shaped first side-lobe and structures of other shapes in the other lobes. Surrounding the main lobe is the first null. FWHM at this frequency (1.708 GHz) is 31.96 $\pm$ 0.03 arcmin, and Figure 3.3 presents a one-dimensional plot of the diagonal (at 45 degrees and along the x- and y-axis) intensity distribution profile. The peak gains in the first, second and third side lobes are 17.48 %, 9.50 % and 7.91 % respectively of the peak gain in the main lobe (which shows that the main lobe dominates the overall beam gain). It is important to note that, although the signal in the side-lobes is small (compared to that in the main-lobe), they still have significant contribution to the overall receiving power of the antenna.There are also indications of some asymmetry in the beam since the graphs (in Figure 3.3) are asymmetric about the y-axis. The off-diagonal terms have a structure that is characterised by four petal-like structures arranged in a diamond-shaped layout

around the centre and have more petal-like structures further away from the centre. The off-diagonal terms have much less power compared to the diagonal terms as expected - having a peak value of gain that is 20 % of that in the diagonal terms.

It is worth mentioning that the beams' size is expected from theory to have a smooth variation with frequency, but this is not so as we will show shortly. The beams do not scale smoothly with frequency. The next section will present mathematical representation of the beams.

3.3.1 The Jones and Mueller matrices

Jones formalism (Jones, 1948) has been used over the years to describe the polarised state of electromagnetic (EM) waves. The principle of the Jones formalism was originally formulated to describe optical polarisation. This formalism is generally suitable for analysing light that propagates through an optical system which is composed of many optical components. This formalism was later adapted into radio interferometry (Hamaker *et al.*, 1996; Smirnov, 2011b). The polarisation state of EM waves can also be described by a set of four values known as the Stokes parameters I, Q, U, and V. They are commonly used to represent polarised signals, because they can easily be computed by measuring the EM wave that has passed through different polarisers. The Stokes parameters can be summarised in a column matrix known as the Stokes vector, $[IQUV]^{\mathrm{T}}$. Again, another formalism that is used to describe the polarised state of EM waves is the Mueller formalism. It gives a more general and complete description of the behaviour of media that are interacting with polarised light passing through them. Let us present some of these using mathematical representations.

Let us consider the electric field of an electromagnetic wave that is propagating through a linearly passive optical system along the z-axis. Let us assume that a radio antenna is such a system. The Jones vector Equation (2.20) defined by complex electric field components provides a concise representation of the total polarisation of such a system. The polarisation of the field can be represented with a (2×2) matrix having the correlation of the two orthogonal components. For right (R) and left (L) circular polarisations, the orthogonal modes are expressed as (Perley *et al.*, 1989; Hamaker, 2000):

$$\begin{bmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{bmatrix} = \begin{bmatrix} \langle \mathrm{RR}^* \rangle & \langle \mathrm{RL}^* \rangle \\ \langle \mathrm{LR}^* \rangle & \langle \mathrm{LL}^* \rangle \end{bmatrix} \equiv \begin{bmatrix} \mathrm{RR} & \mathrm{RL} \\ \mathrm{LR} & \mathrm{LL} \end{bmatrix}, \tag{3.1}$$

where $\langle . \rangle$ represents time averaging, and $*$ represents complex conjugate.

The coherency matrix can also be expressed as a column vector. Again, recall from Section 2.4.1, that an electromagnetic (EM) plane wave propagating along the z axis can be represented at any point in space by a vector with elements e_x and e_y (Equation (2.20)). An antenna targets to measure the pairwise coherencies of these amplitudes, given as (O'Neill, 2003):

$$\boldsymbol{x} = \begin{bmatrix} \langle e_x e_x^* \rangle \\ \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle \\ \langle e_y e_y^* \rangle \end{bmatrix} = \langle e \otimes e^* \rangle \,, \tag{3.2}$$

where $\langle . \rangle$ denotes averaging over frequency/time interval and $\otimes$ represents the Kronecker product. For the sake of simplicity, let us consider a feed with linearly polarised receivers. The measured coherences and the Stokes vectors, $\boldsymbol{s}$ $(= [IQUV]^{\mathrm{T}})$ are related as (Born and Wolf, 1980):

$$\boldsymbol{s} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \begin{bmatrix} \langle e_x e_x^* \rangle + \langle e_y e_y^* \rangle \\ \langle e_x e_y^* \rangle - \langle e_y e_y^* \rangle \\ \langle e_x e_y^* \rangle + \langle e_y e_x^* \rangle \\ -i(\langle e_x e_y^* \rangle - \langle e_y e_x^* \rangle) \end{bmatrix} \tag{3.3}$$

The relationship between the polarisation vector and its corresponding Stokes parameters are given as (Parke, 1948; O'Neill, 2003; Azzam and Bashara, 1977; Hamaker, 2000):

$$\boldsymbol{s} = \boldsymbol{S}\boldsymbol{x}, \tag{3.4}$$

where

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 1 & -i & i & 0 \end{bmatrix}.$$

An interferometer measures a set of pairwise correlations between two voltages induced on the two orthogonal components of the feed probes, v_a and v_b by the incident EM fields. Recall from Equation (2.21) that $\boldsymbol{v} = \boldsymbol{J}\boldsymbol{e}$, where $\boldsymbol{v}$ represents the column vector of the

two voltages and $\boldsymbol{J}$ is the 2×2 Jones matrix which describe the propagation effects. One can then write the measured coherencies $\boldsymbol{x}'$ as:

$$\boldsymbol{x}' = \langle \boldsymbol{v} \otimes \boldsymbol{v}^* \rangle = (\boldsymbol{J} \otimes \boldsymbol{J}^*) \langle \boldsymbol{e} \otimes \boldsymbol{e}^* \rangle = (\boldsymbol{J} \otimes \boldsymbol{J}^*)\boldsymbol{x}, \tag{3.5}$$

and the measured Stokes parameter vector $\boldsymbol{s}'$ is related to the original Stokes vector through a formalism known as the Mueller matrix, $\boldsymbol{M}$ that is given as (Parke, 1948; O'Neill, 2003; Azzam and Bashara, 1977):

$$\boldsymbol{s}' = \boldsymbol{M}\boldsymbol{s} = \boldsymbol{S}(\boldsymbol{J} \otimes \boldsymbol{J}^*)\boldsymbol{S}^{-1}\boldsymbol{s}. \tag{3.6}$$

Let us ignore all other propagation effects and consider only the direction-dependent primary beam effect; hence the Mueller matrix here refers to the Mueller matrix of the primary beam. This implies that each direction in the FoV of the antenna has a different matrix associated with it.

Equation (3.6) shows that the Mueller matrix, $\boldsymbol{M}(l, m)$ can be obtained from the Jones matrix , $\boldsymbol{J}(l, m)$. The Jones matrix on-diagonal terms in Figure 3.1 show the response of each feed as a function of direction to the corresponding electromagnetic field component, while the off-diagonal terms show polarisation leakages.

The characterisation of the polarisation elements is given in detail by the Mueller matrix formalism (Bass *et al.*, 2009). Mueller proposed a 4×4 matrix formalism that can be used to measure and quantify the amount of polarisation leakages of one polarised signal into another and is written in terms of the measured Stokes parameters. The Mueller matrix, M is expressed as:

$$\boldsymbol{M} = \begin{bmatrix} I \to I & Q \to I & U \to I & V \to I \\ I \to Q & Q \to Q & U \to Q & V \to Q \\ I \to U & Q \to U & U \to U & V \to U \\ I \to V & Q \to V & U \to V & V \to V \end{bmatrix}, \tag{3.7}$$

where $I \to Q, U, V$ denote how much of Stokes intensity I leaks into the other Stokes polarisation components Q, U, V respectively, while in general, $X \to Y$ represent Stokes X leaking into Y. Figure 3.4 shows plots of the Mueller matrix elements for VLA holography measured beams.

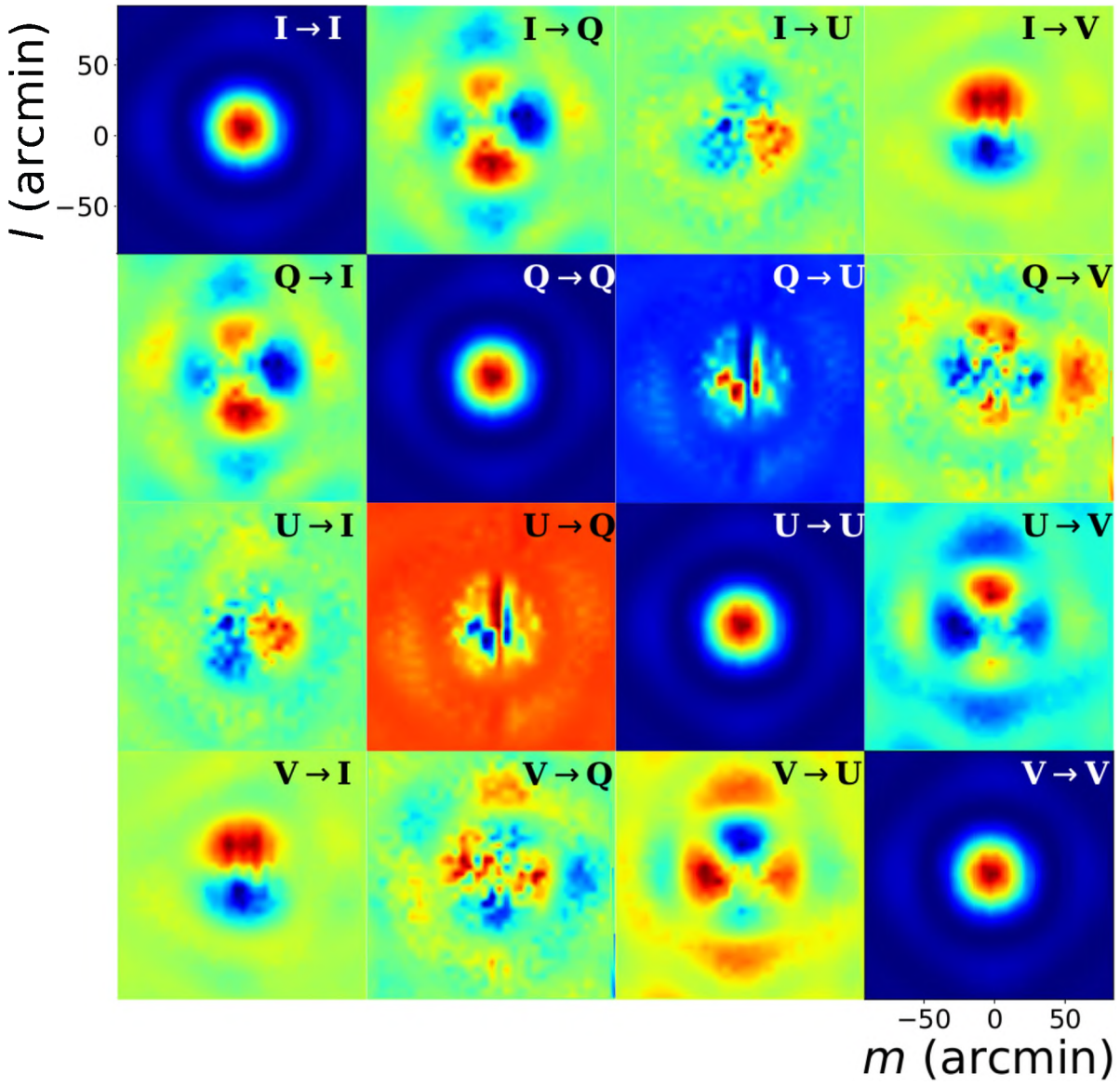

FIGURE 3.4: Mueller matrix (Equation 3.7) plots of the two-dimension holography beams for one antenna at 1.708 GHz.

3.4 Preliminary analysis of the beams

In the previous sections, we presented selected samples of the holography measured beams and the mathematical formalism that can be used to analyse them. Here we will apply the formalism to analyse the holography beams.

Figure 3.5 presents samples of the (Stokes I) amplitude beams at seven frequencies from 1.008 - 1.908 GHz (in steps of 100 MHz).

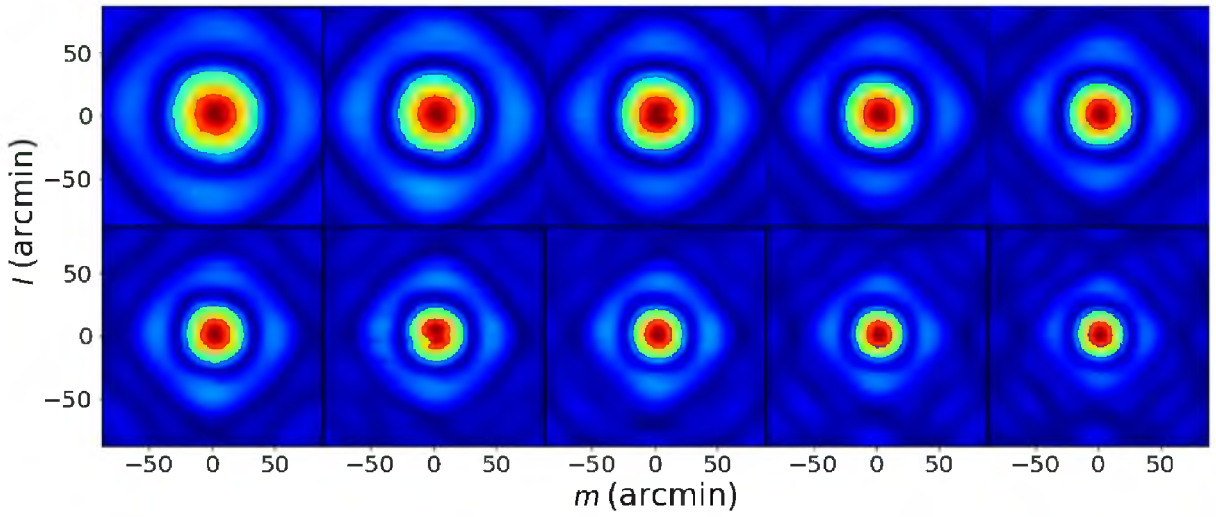

FIGURE 3.5: The amplitude beams (for one antenna) at ten different frequencies for 1.008 - 1.908 GHz, selected in steps of 100 MHz (moving from left to right then top to bottom rows).

There are many ways which the beams' frequency dependence could be presented. Some of the same properties (beams' sizes and offsets) could be visualised by investigating the frequency dependence of the main lobe. This was achieved by fitting a two-dimensional elliptical Gaussian to the main lobe of each (Stokes I amplitude) beam at each frequency (for twelve antennas), using the diagonal terms of the (RR and LL) amplitude beams. This produced the full width at half maximum (FWHM) and centre offset values for the beams along the x and y (i.e. l and m) directions and rotation. The fitted values for the rotation parameters were all effectively zeros, implying that there is no rotation in the Stokes I diagonal terms. Figure 3.5 shows the plots of 12 antenna beams' FWHM (along the x-axis) with frequency, while Figure 3.6 presents one antenna FWHM (top panel) and centre offsets (bottom panel) with frequency.

The plots of the holography beams' FWHM (size) against ν in Figure 3.6 depicts some interesting characteristics:

1. The beams' sizes are inversely proportional to $\frac{1}{\nu}$ in agreement with theory which we call the first-order characteristics.

2. FWHM vs frequency for all of the twelve antennas exhibits a sinusoidal trend which we call a second-order characteristic. As indicated earlier (in Chapter 2), one disadvantage reflector antennas have is the presence of standing-waves in the optics. This standing-waves causes constructive interference (that increases the beams' sizes) at some frequencies, and destructive interference (that decreases the beams' sizes) at

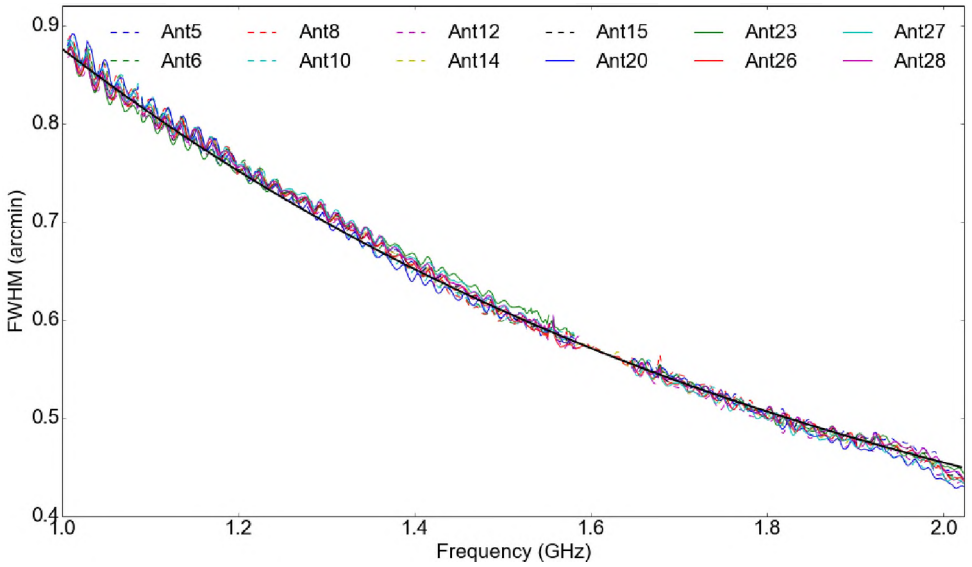

FIGURE 3.6: FWHM against frequency for twelve antennas; The solid black line represents the best-fitting $\frac{1}{\nu}$ frequency scaling law.

others (as is shown in Figure 3.6) producing some sinusoidal trend observed in these plots. Thus, direction-dependent spectral responses are introduced into the beams. We termed this effect the *ripple effect* or *characteristic*.

3. The offsets and shifts in the trend of the curves are not the same across all twelve antennas. This may be explained by the differences in each antenna surface reflectivity error, mechanical deformation (due to gravitational load) etc.

More characteristic effects of the holography beams can be noticed on the other fitted parameters, such as the beams' centre offsets. The VLA antennas are examples of off-axis Cassegrain antenna systems which have their feeds installed in a circular focusing ring around the main reflector's optical centre, hence each feed is offset by 1.05 m. The asymmetric hyperboloidal sub-reflector rotates to focus the beams into each feed. Thus, the holography beams have centre maxima that point slightly offset with respect to their optical centres. This is called the beam squint. Figure 3.7 presents the graphs of one antennas FWHM and offsets against frequency.

We observe the following interesting features of the holography beams (see Figure 3.6):

1. The top panel of Figure 3.7 shows that the beams are virtually exactly symmetrical at the main lobe ($\text{FWHM}_X \approx \text{FWHM}_Y$) for both the diagonal polarisations (RR and LL) beams.

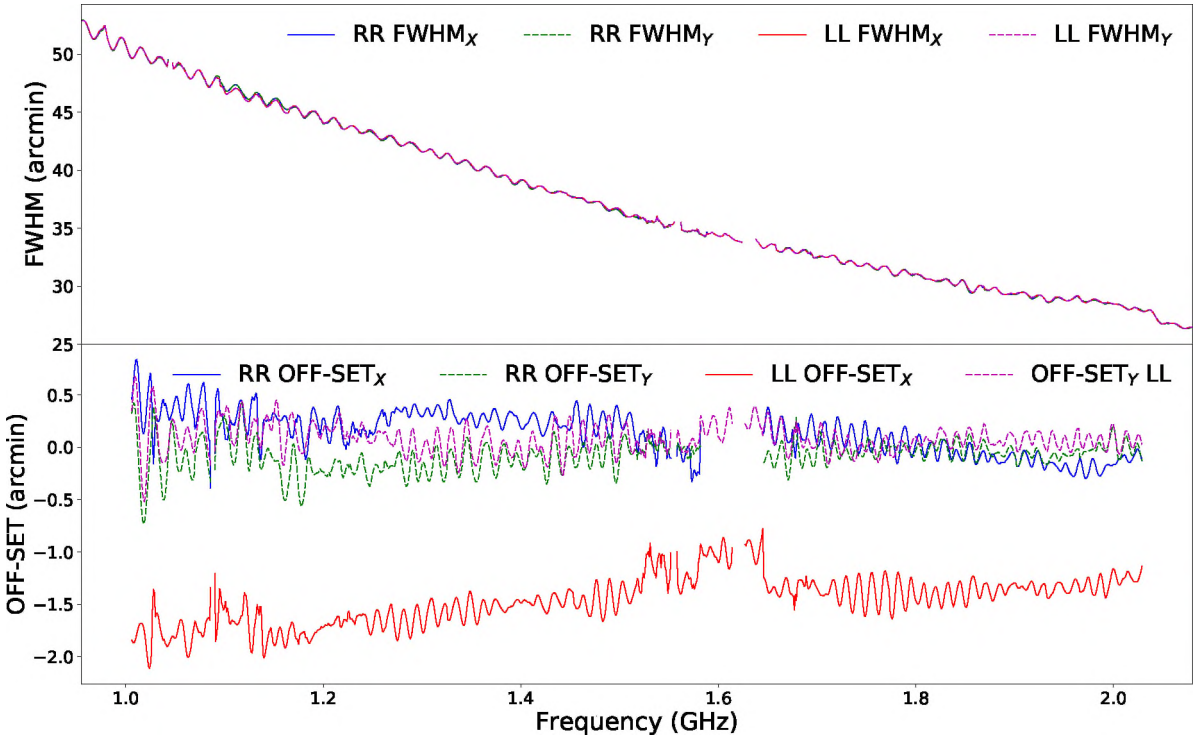

FIGURE 3.7: FWHM (top panel) and offset (bottom panel) against frequency for one antenna (antenna 5).

2. Plots of the beam offset (bottom panel) shows that the *standing wave* in the antenna also induces a jitter in the beam offset vs frequency plots.

3. The scale of the offset shift along the x-axis of the holography beams is higher than along the y-axis, particularly for the LL beams. This fact is made more obvious if one considers Figure 3.8, which presents the two-dimensional plots of antenna 5 beams' squint positions. This is the representation of the bottom plots of Figure 3.7 in two-dimensional form. These graphs are presented in the antenna frame (i.e. plotted with l vs m). For this antenna, the RR beams have mean offset values of approximately 0.34 arcmin and -0.082 arcmin in right ascension and declination respectively, with respect to the optical centre. The LL beams have mean offset values of approximately -1.464 arcmin and 0.082 arcmin in right ascension and declination respectively with respect to the optical centre. The beam centres for the antennas are scattered about the respective average for each polarisation.

Note that places where we have gaps in these plots correspond to flagged frequencies that were affected by high RFI in the measurements.

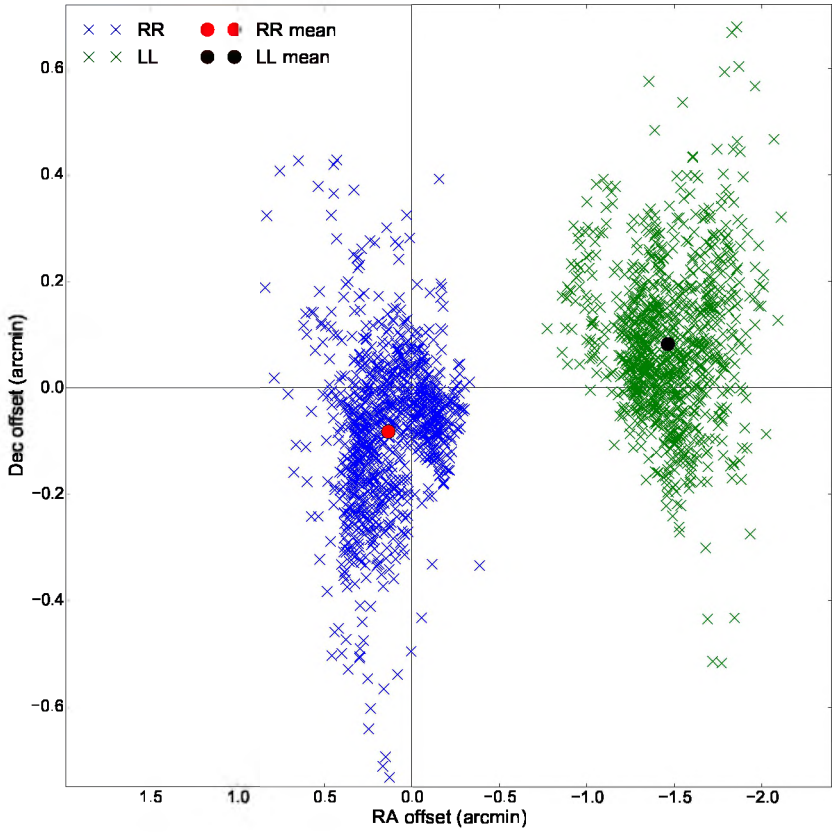

FIGURE 3.8: Two-dimensional plots of the one antennas (antenna 5) beams' center offsets (in the orientation the antenna view the sky)

3.5 Summary

This chapter, presented holography as one of the ways that can be used to obtain the true characteristics of radio antenna primary beams using a specific example of the VLA antennas. It presented the measured data and some preliminary analysis. We saw that the VLA antennas PB have some interesting features (characteristics - asymmetry in the morphology of the beams' patterns, non-smooth beam sizes and offsets scaling with frequency etc.). Further analysis of the properties of the holography beams will be presented in Chapter 4. They are asymmetric, having pointing errors and direction-dependent spectral responses. Accurate PB models for the VLA antennas are required to account for these DDEs. In Chapter 4, we will explore some modelling techniques for beam patterns to see which of them would be the best to account for these features.

Chapter 4

Modelling the VLA and MeerKAT antennas' L-band beams from holography

4.1 Introduction

Recent advancements in the design of radio telescopes have driven the construction of large radio interferometers with improved sensitivity, like LOFAR (van Haarlem *et al.*, 2013), MeerKAT, MWA, VLA etc. These telescopes produce data with very high temporal and spectral resolutions. Faster, more efficient and more accurate techniques are required to meet the challenges of handling the data storage and calibration imposed by the size and complexity of these equipments. The usefulness and quality of the scientific products that one can get from the data are highly dependent on the ability to handle instrumental and environmental distortions that affect the data during observation. As a result of these telescopes having a large field of view (FOV), classical (i.e the first generation, see Section 2.4.3 for more details) calibration and imaging techniques are limited when used to perform data reduction and imaging on data obtained by these instruments. A new mathematical framework was proposed (Hamaker *et al.*, 1996) which efficiently captures the radio-polarimetric inverse problem. This framework, as discussed in Chapter 2, called the radio interferometer measurement equation (RIME), allows for the propagation effects that affect the signal which is coming from the source to be formulated and accounted for. The RIME presents an improved way to model and calibrate complex

DIEs and DDEs, necessary to achieve high dynamic range (HDR) in radio interferometer images.

Incorrect antenna PB models will reduce the accuracy of the true sky recovered at positions far away from the pointing centre. We showed in Chapters 2 and 3 that the PB of off-axis Cassegrain (alt-azimuth mounted) telescopes has spectral and temporal variations (because of parallactic angle rotation) towards off-axis sources. These cause bright radio sources (with constant fluxes) that are at positions close to the boundaries of the side-lobes (and main lobe) where the PB gain has steep variations to appear to have varying brightness. This creates problems for the calibration procedures that assume a sky model made up of sources with constant brightness. Calibration solutions then attempt to compensate for these seemingly variable sources, generating incorrect calibration solutions that produce strong residual artefacts in the resultant images. These artefacts can be bright enough to considerably increase the noise to a level where faint (e.g μJy level) sources cannot be detected in deep observations.

Consequently, good knowledge of the PB is essential to be able to generate images with high dynamic range values and minimal bias. Direction-dependent calibration techniques (for example peeling, differential gain calibration etc.) can mitigate artefacts around bright sources but the source flux suppression that they produce compromises the fidelity of the data (Mulcahy, 2014; Intema *et al.*, 2009).

The primary beams of radio telescopes have been modelled with different methods. These can be grouped into analytic, simulation-based and data-driven models. Analytic models are generated by using some analytic functions to approximate the primary beam, for example, $\cos^3$ for WSRT (Smirnov, 2011c; Popping and Braun, 2008; Braun *et al.*, 2007), Jacobi-Bessel pattern model (Galindo-Israel and Mittra, 1977; Rahmat-Samii and Galindo-Israel, 1980), etc. Simulation-based models are developed by following the physics of the problem, for example, *Cassbeam* software. The data-driven method models the PB from data - simulated or measured. An example is the use of characteristic basis function patterns (Young, 2013). In this chapter, we will present two techniques for modelling the radio antenna primary beams from holography measured data using principal component analysis (PCA) and then Zernike polynomials. We will present a brief overview of SVD, principal component analysis (PCA) and Zernike polynomials, then apply these techniques to holography data to model the VLA and MeerKAT antennas' primary beams.

4.2 The fundamentals of singular value decomposition and principal component analysis

SVD and PCA are techniques that can be used, among other things, for carrying out dimensionality reduction to produce a compact representation of data with minimal loss of information.

4.2.1 The fundamentals of SVD

SVD is one of the most commonly used matrix factorisation tools. It provides a method to split a matrix into its fundamental subspaces and shows the relative importance of the directions in those subspaces. Hence, it is an excellent tool for data analysis, e.g. solving least squares problems and provides a way to approximate a matrix using lower rank matrices.

To lay out foundations of this subject, let us define some basic terms. The spectral theorem states that if $\boldsymbol{\mathcal{S}}$ is an $n \times n$ symmetric matrix (i.e $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}^{\mathrm{T}}$), then there exists a set of eigenvectors and their corresponding eigenvalues $\boldsymbol{u}_i$ and λ_i respectively ($i = 1, 2, 3, ...n$) such that:

$$\boldsymbol{\mathcal{S}}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i. \tag{4.1}$$

The set of *eigenvectors* form an orthonormal basis for $\boldsymbol{\mathcal{S}}$:

$$\boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{u}_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \tag{4.2}$$

A matrix (with dimension $m \times m$) is diagonalisable if it has m distinct number of eigenvectors. Let us consider an $m \times m$ diagonalisable matrix $\boldsymbol{\mathcal{A}}$. The square root of the *eigenvalues* of $\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{A}}$ are the *singular values* of $\boldsymbol{\mathcal{A}}$ and the ratio of the largest to the smallest *singular values* is the *condition number* of $\boldsymbol{\mathcal{A}}$. A matrix is said to be *ill-conditioned* if the *condition number* is very large. A matrix is singular if the condition number is infinity. A singular matrix has a zero determinant, the latter being the product of the eigenvalues. The *rank* of matrix $\boldsymbol{\mathcal{A}}$ is the number of linearly independent rows or columns of $\boldsymbol{\mathcal{A}}$.

The spectral decomposition of matrix $\boldsymbol{\mathcal{A}}$ is of the form:

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{U}}^{\mathrm{T}}, \tag{4.3}$$

where $\boldsymbol{\mathcal{U}}^{\mathrm{T}}\boldsymbol{\mathcal{U}} = \mathbf{I}$ i.e $\boldsymbol{\mathcal{U}}$ is made up of orthonormal columns and these columns, are the *eigenvectors* of $\boldsymbol{\mathcal{A}}$. $\boldsymbol{\mathcal{S}}$ is a diagonal matrix whose diagonal elements are the *eigenvalues* of $\boldsymbol{\mathcal{A}}$ corresponding to the eigenvectors of $\boldsymbol{\mathcal{U}}$ and $\mathbf{I}$ is an identity matrix. Multiplying Equation (4.3) by $\boldsymbol{\mathcal{U}}$ gives:

$$\begin{aligned} \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{U}} &= \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{U}}^{\mathrm{T}}\boldsymbol{\mathcal{U}} \\ &= \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{S}}\mathbf{I} \\ &= \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{S}}. \end{aligned} \tag{4.4}$$

From Equation (4.1), using jth *eigenvectors* from $\boldsymbol{\mathcal{U}}$ ($\boldsymbol{u}_j$) and jth *eigenvalues* $\boldsymbol{\mathcal{S}}$ (s_j), we can write:

$$\boldsymbol{\mathcal{A}}\boldsymbol{u}_j = s_j\boldsymbol{u}_j, \tag{4.5}$$

where $j = 1, 2, ..., n$.

We define the SVD of a matrix $\boldsymbol{\mathcal{B}}$ with dimension $n \times m$, $n \geq m$ as:

$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{V}}^{\mathrm{T}}, \tag{4.6}$$

where $\boldsymbol{\mathcal{U}}$ is an $n \times m$ matrix with orthonormal columns and are *eigenvectors* of $\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}^{\mathrm{T}}$ ($\boldsymbol{\mathcal{U}}^{\mathrm{T}}\boldsymbol{\mathcal{U}} = \mathbf{I}$). $\boldsymbol{\mathcal{V}}$ is an $m \times m$ matrix whose orthonormal columns are *eigenvectors* $\boldsymbol{\mathcal{B}}^{\mathrm{T}}\boldsymbol{\mathcal{B}}$ ($\boldsymbol{\mathcal{V}}^{\mathrm{T}}\boldsymbol{\mathcal{V}} = \mathbf{I}$), while $\boldsymbol{\Upsilon}$ represents an $m \times m$ diagonal matrix having zero or positive values called the *singular values* of $\boldsymbol{\mathcal{B}}$ (hence, the square root of the eigenvalues of $\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}^{\mathrm{T}}$ or $\boldsymbol{\mathcal{B}}^{\mathrm{T}}\boldsymbol{\mathcal{B}}$).

Now that we have defined the SVD of a matrix, we will apply it to the derivation of the PCA.

4.2.2 The fundamentals of PCA

PCA is a factorisation and dimensionality reduction method. To present the mathematical formalism of PCA, let $\boldsymbol{X}_{\text{raw}}$ be an $n \times m$ matrix of sampled data such that n is the number of samples while m is the number of features in the data ($\boldsymbol{x}_i \in \boldsymbol{X}; i = 1, 2, ...m$). PCA is normally performed on centred data. A centred form of $\boldsymbol{X}_{\text{raw}}$, $\boldsymbol{X}$ is generated by subtracting each column mean from the respective column data. PCA is formed around two principal moments of the sampled data matrix, namely the mean vector, $\langle \boldsymbol{x} \rangle$ and covariance matrix, $\boldsymbol{C}$, given as:

$$\langle \boldsymbol{x} \rangle = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \quad \text{and} \quad \boldsymbol{C} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \langle \boldsymbol{x} \rangle)(\boldsymbol{x}_i - \langle \boldsymbol{x} \rangle)^{\mathrm{T}}. \tag{4.7}$$

$\boldsymbol{C}$ can also be written in terms of the data matrix, $\boldsymbol{X}$ as:

$$\boldsymbol{C} = \frac{1}{n-1} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{X}. \tag{4.8}$$

$\boldsymbol{C}$ is a symmetric matrix and can be factorised as:

$$\boldsymbol{C} = \boldsymbol{\mathcal{V}} \boldsymbol{\mathcal{L}} \boldsymbol{\mathcal{V}}^{\mathrm{T}}, \tag{4.9}$$

where $\boldsymbol{\mathcal{V}}$ is the matrix of eigenvectors, $\boldsymbol{\mathcal{L}}$ is the diagonal matrix of eigenvalues sorted in decreasing order. Projection of the data in the direction of eigenvectors (known as the principal axis) is called the principal components (PCs) given as:

$$\tilde{\boldsymbol{Z}} = \boldsymbol{X} \boldsymbol{\mathcal{V}}, \tag{4.10}$$

By selecting only the first kth rows of $\tilde{\boldsymbol{Z}}$, we project the data from m to a reduced k number of dimensions.

The choice of what number k of principal components is appropriate to be used to approximate the data is often not easy to determine. One can justify the use of a certain number of k principal components by analysing the energy fraction left in the k low-rank approximation. The characteristics of *singular values* can be used to justify the choice of the number of k PCs. A plot of the *singular values* show when it has stabilised (i.e when the

values become approximately constant) and the remaining PCs can be discarded because they are mostly noise dominated. This is made evident in Figure 4.1, which presents a sample plot of the singular values of an SVD decomposition. We can see that after number 10, the singular values are approximately constant which implies that adding more of these components (with approximately constant values) will not significantly improve the data approximation and the associated weight can be forced to be zero to enable the k low-rank approximation. The region in the graph before the singular values become approximately constant is normally referred to as the *knee*. One common way of knowing what value of k is sufficient for data reconstruction is to observe the *knee* and choose a value for k at the point where the singular values become approximately constant.

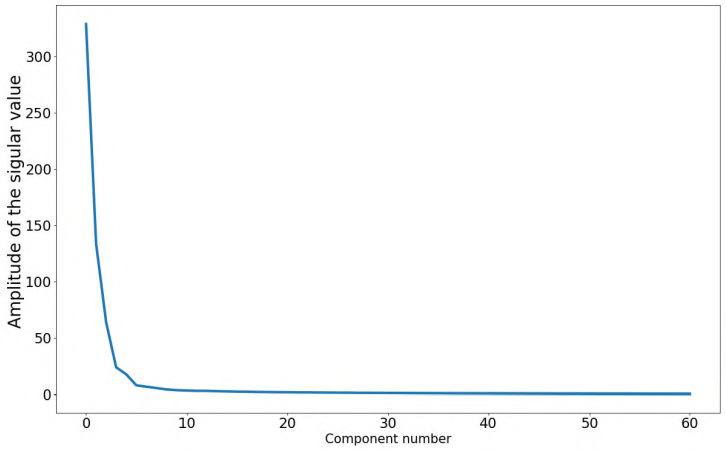

FIGURE 4.1: Example plot of an arbitrary sample singular values of an SVD (to illustrate the distribution of the values of the singular value components).

4.2.3 Performing PCA through SVD

PCA is, in practice, the SVD of a centred data matrix (Jolliffe and Cadima, 2016). Let us continue with the matrix $\boldsymbol{X}$ (from the previous section) and use SVD to decompose $\boldsymbol{X}$. From Equation (4.6) we can perform the SVD decomposition of $\boldsymbol{X}$ as:

$$\boldsymbol{X} = \boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{V}}^{\mathrm{T}}, \tag{4.11}$$

and from Equations (4.8) and (4.11) the covariance matrix can be written as:

$$\begin{aligned} \boldsymbol{C} &= \frac{1}{1-n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \\ &= \frac{1}{n-1}(\boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{V}}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{V}}^{\mathrm{T}}) = \frac{1}{n-1}(\boldsymbol{\mathcal{V}}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{U}}^{\mathrm{T}})(\boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{V}}^{\mathrm{T}}) \\ &= \frac{1}{n-1}\boldsymbol{\mathcal{V}}\boldsymbol{\Upsilon}^2\boldsymbol{\mathcal{V}}^{\mathrm{T}} = \boldsymbol{\mathcal{V}}\frac{\boldsymbol{\Upsilon}^2}{n-1}\boldsymbol{\mathcal{V}}^{\mathrm{T}}. \end{aligned} \tag{4.12}$$

From Equations (4.9) and (4.12) we see that the eigenvalues $\boldsymbol{\mathcal{L}}$ of the covariance matrix are equal to $\frac{\boldsymbol{\Upsilon}^2}{n-1}$.

Also, from Equation (4.11), we see that:

$$\boldsymbol{X}\boldsymbol{\mathcal{V}} = \boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{V}}^{\mathrm{T}}\boldsymbol{\mathcal{V}} = \boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}, \tag{4.13}$$

which implies that one can obtain the PCs of $\boldsymbol{X}$ from its SVD. The data is reconstructed as:

$$\begin{aligned} \boldsymbol{X}_{\mathrm{raw}} &= \langle \boldsymbol{x} \rangle + \boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{V}}^{\mathrm{T}} \\ &= \langle \boldsymbol{x} \rangle + \boldsymbol{a}\boldsymbol{\mathcal{V}}^{\mathrm{T}}. \end{aligned} \tag{4.14}$$

Hence, SVD of $\boldsymbol{X}$ can be used to perform the PCA of $\boldsymbol{X}_{\mathrm{raw}}$. The principal components (PCs), $\boldsymbol{X}\boldsymbol{\mathcal{V}} = \boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon} = \boldsymbol{a}$.

4.3 Overview of Zernike polynomials

It is often common in signal processing to find the need to search for a set of orthonormal bases that can be used to represent functions defined on some finite support having a certain geometry. Zernike polynomials (ZPs) have been used to capture optical path difference in wavefronts (Zernike, 1934) and deformations in optical surfaces like in the human eye (e.g. the cornea in (Iskander *et al.*, 2001; Schwiegerling *et al.*, 1995)). ZPs have also been used to describe the thermal blooming effect in the atmosphere (Bradley and Herrmann, 1974) etc. These applications were made possible because ZPs can easily be used to represent properties of many circular shaped optical systems. There are many possible linear combinations of these polynomials which can be used to represent a system because of some of their unique properties and advantages (Mahajan, 1994). These include but are not limited to the fact that they allow for links to be established

with Seidel theory of aberration (Born and Wolf, 1999), that relies on a third-order Taylor series expansion. The orthonormal property of these polynomials gives them practical and theoretical advantages and makes them a first choice option to be used in describing the phase of wavefronts. In this study, ZPs will be used to represent the morphology of the primary beams of radio astronomy telescopes.

Let the wavefront, $\phi(\rho,\theta)$ be a linear combination of ZPs (in polar coordinates (ρ,θ)) over a unit circle. This can be written mathematically as:

$$\phi(\rho,\theta) = \sum_{n,m}^{M} a_n^m Z_n^m(\rho,\theta), \tag{4.15}$$

where the general definition of $Z_n^m(\rho,\theta)$, the basis of Zernike moments is given as (Zernike, 1934; Ferreira *et al.*, 2015):

$$Z_n^m(\rho,\theta) = \begin{cases} N_n^m R_n^{|m|}(\rho)\cos(\rho m\theta), & m \geq 0 \\ -N_n^m R_n^{|m|}(\rho)\sin(\rho m\theta), & m < 0, \end{cases} \tag{4.16}$$

where N_n^m is the *normalisation factor* and $R_n^{|m|}$ is the *radial polynomial*, and they are respectively given by:

$$N_n^m = \sqrt{\frac{2(n+1)}{1+\delta_{m,0}}} \quad \text{and} \quad R_n^{|m|} = \sum_{s=0}^{0.5(n-|m|)} \frac{(-1)^s (n-s)!\rho^{n-2s}}{s!\left(\frac{n-|m|}{2}-s\right)!\left(\frac{n+|m|}{2}-s\right)!},$$

where the Kronecker delta function $\delta_{m,0} = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0 \end{cases}$, n =1,2,3,.. and for a given value of n, the index m =-n, -n + 2, -n + 4,...,n.

$Z_n^m(\rho,\theta)$ in Equation (4.16) is made of a family of a completely orthonormal set of polynomials over a unit circle and is represented by the product of radial polynomials and angular functions. This orthonormality property makes the basis vectors to be independent of one another (Wyant and Creath, 1992; Charman, 2005) and are generally represented with double (n,m) notation or single j notation mode indices. The mode m denotes the azimuthal frequency of the sinusoidal while the n mode characterises the degree of aberration. ρ (the radial parameter) is continuous within the range (0,1), which implies that the azimuthal component is continuous over θ so that $0 \leq \theta \leq 2\pi$. One can

see from the plots of radial responses in Figure (4.2), that the value of the polynomials converges at the edge of the disc. We can say that all orders of Zernike polynomials are confined within the (-1, 1) interval and one can use them as basis functions for image moments as presented in Figure (4.2).

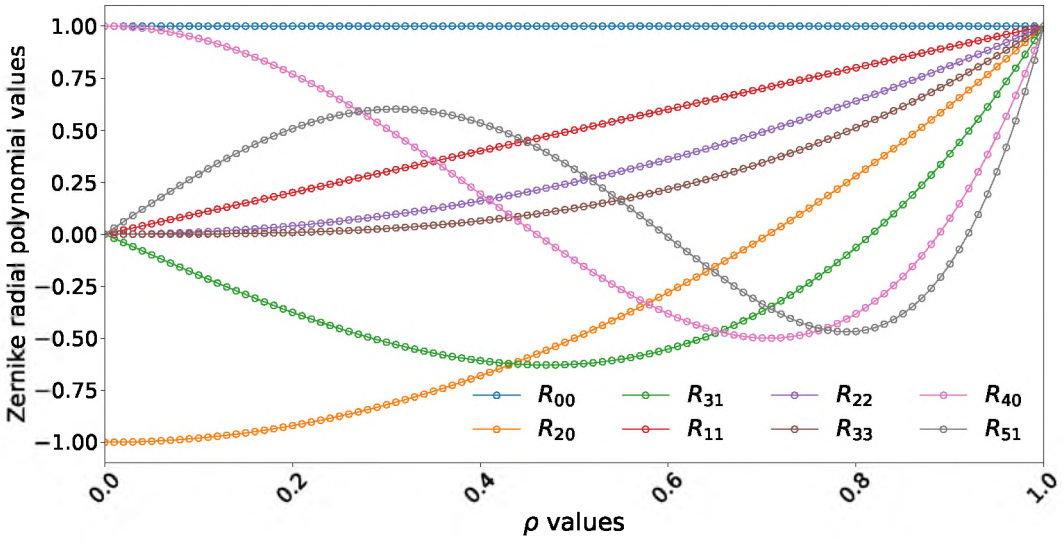

FIGURE 4.2: Expansion of $R_n^{|m|}(\rho)$ for 8 orthonormal radial polynomials.

Two-dimensional plots for the first nine levels of the Zernike pyramid are presented in Figure (4.3). We can also observe that for $m = 0$ (the middle column) that the modes are rotation invariant which makes them symmetric. The Zernike modes at each row, with opposite value of azimuthal frequency (i.e equal $|m|$), generally have the same morphology but different orientation (i.e having a rotation of about 45° or 90° e.t.c). We need these pairs so that we can freely move any mode around 2π, to get the desired orientation by selectively adjusting the weight of the mode.

We have presented the Zernike expansion coefficients with the double mode scheme (or notation) (n, m) in Figure 4.3. A single mode scheme is sometimes used to represent the expansion coefficients (Wyant and Creath, 1992; Noll, 1976). The ordering of the single mode scheme is completely arbitrary because the coefficient depends on two parameters. Let us denote the single mode index scheme with k. It is often convenient to present the polynomials in a pyramidal layout having n row and m column numbers. The relationship between these two mode schemes can be expressed as:

$$Z_n^m(\rho, \theta) \equiv Z_k, \tag{4.17}$$

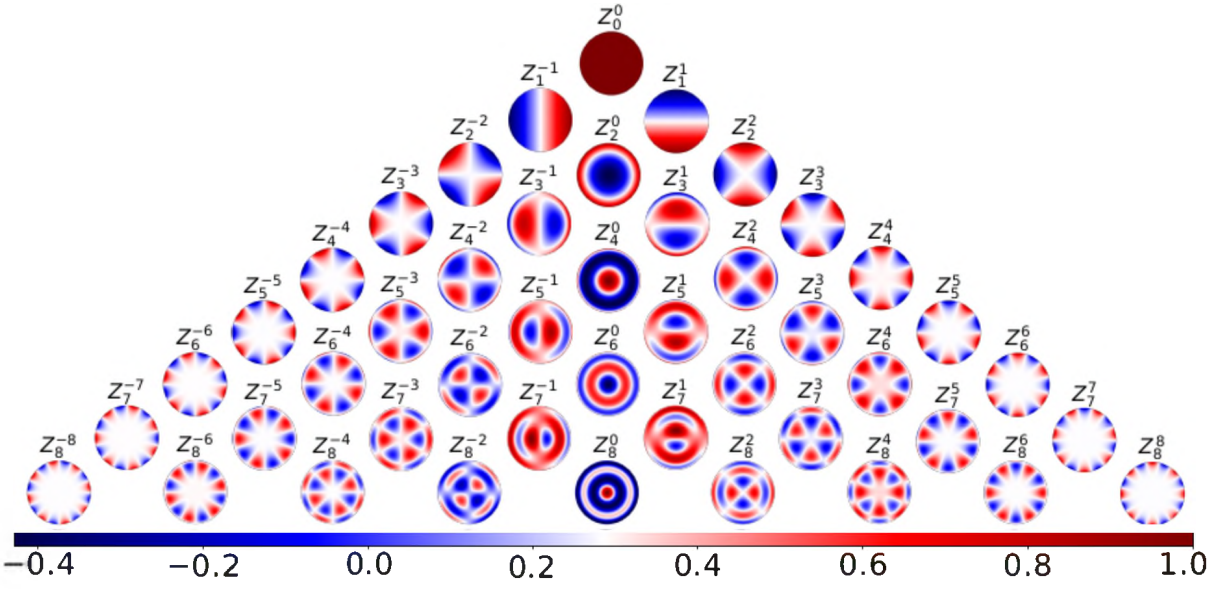

FIGURE 4.3: Two-dimensional plots of $Z_n^m(\rho, \theta)$ (the basis patterns of Zernike moments) of order 9, plotted on a unit disc.

where $n = \lfloor 0.5(-3 + \sqrt{9 + 8k}) \rfloor$, $m = 2k - n(n + 2)$, $k = 0.5(m + n(n + 2))$ and the notation $\lfloor . \rfloor$ denotes the Ceil function. Table 4.1 shows the relationship between these two notations.

n \ m	-4	-3	-2	-1	0	1	2	3	4
0					$k = 0$				
1				$k = 1$		$k = 2$			
2			$k = 3$		$k = 4$		$k = 5$		
3		$k = 6$		$k = 7$		$k = 8$		$k = 9$	
4	$k = 10$		$k = 11$		$k = 12$		$k = 13$		$k = 14$

TABLE 4.1: Tabular representation of the relationship between double and single index notations to the 4th order (where m is the angular frequency and n radial order)

4.3.1 Computing Zernike coefficients

The general expression for approximating data using a Zernike system with a definite function of ith degree is presented in Equation (4.15). In this study, the holography measured beams will be analogues to the wavefront (in optical) that we want to describe with ZPs. To develop an expression for approximating the coefficients C_n^m, let D be the number of measured beams represented by $\boldsymbol{\phi}^d(\rho_d, \theta_d)|_{d=1,2,...,D}$, where (ρ_d, θ_d) and $\boldsymbol{\phi}^d$ represent two-dimensional polar coordinates and the corresponding normalised values of

the measured beams respectively. Rewriting Equation (4.15) in a slightly simpler form and solving this least squares problem, we have:

$$\boldsymbol{Z}c = \boldsymbol{\phi}, \tag{4.18}$$

where ϕ is a D dimensional array that is composed of sampled beam values, c represents an array of D dimensional Zernike coefficients and the matrix (of dimension $P \times D$) made up of the values of the Zernike polynomials in each of the sampled points is $\boldsymbol{Z}$. Multiplying both sides of Equation (4.18) by $\boldsymbol{Z}^{\mathrm{T}}$ gives:

$$\begin{aligned} &\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}c = \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\phi}, \qquad \text{which implies that} \\ &c = \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\phi}. \end{aligned} \tag{4.19}$$

Obtaining the coefficients for each mode only requires a scalar product of the PB data matrix with each basis vector. Having the basis patterns of Zernike polynomials in Figure 4.3 with their corresponding coefficients, one can approximately reconstruct the desired beam.

After modelling the PB and computing the coefficients we also tried to compress it further, by modelling spectral behaviour of the coefficients. One of the techniques we used is the discrete cosine transform (DCT).

4.4 DCT

After producing models of the primary beams with a reduced number of coefficients (for SVD, PCA and Zernike decompositions) we further compress their coefficients in frequency using the DCT. The DCT algorithm is one of the algorithms used to represent an arbitrary signal using an orthonormal basis. This algorithm uses cosine functions to represent signals. We have type I, II and III DCT algorithms and there are different definitions for them but we will focus on type II and III because these are the ones used in this study. A typical type II orthonormalised DCT can be defined as (Makhoul, 1980):

$$\begin{aligned} \boldsymbol{Y}'_k &= \frac{\Upsilon_k}{N} \sum_{n=0}^{N-1} \boldsymbol{x}_n \cos\Big[\frac{\pi k}{N}(n+0.5)\Big], \quad k = 0, 1..., (N-1), \\ \Upsilon_k &= \begin{cases} 1, & \text{if } k = 0 \\ \sqrt{2}, & \text{otherwise,} \end{cases} \end{aligned} \tag{4.20}$$

where $\boldsymbol{x}$ is the input array (of data that is to be modelled), the index of the output coefficients is represented by k, $\boldsymbol{Y}'$ denotes the DCT output while Υ is the scaling function. We will use this in a later section to model the spectral characteristics of the decomposition coefficients.

A type III DCT (DCT inverse of type II DCT) can be defined as (Makhoul, 1980):

$$\boldsymbol{Y}_k = \frac{\boldsymbol{Y}'_0}{N} + \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} \boldsymbol{Y}'_{\boldsymbol{n}} \cos\Big[\frac{\pi k}{N}(n+0.5)\Big], \quad k = 0, 1..., (N-1). \tag{4.21}$$

Equation (4.20) is used to perform the DCT decomposition and get the DCT coefficients, $\boldsymbol{Y}'$ while using the DCT coefficients in Equation (4.21) one can reconstruct the original signal (with a number of DCT coefficients less than the input data).

4.5 Modelling the VLA beams

The terminologies, derivations, figures and notations in this section are taken from Iheanetu *et al.* (2019).

4.5.1 The *Cassbeam* software

The VLA antennas are typical off-axis Cassegrain systems (as we have discussed in Chapter 2) and are composed of a paraboloid main reflector, an asymmetric hyperboloidal sub-reflector and a set of circularly polarised feed horns installed in a circular ring around the optical centre of the main reflector and all mounted in an alt-azimuth mount. The antenna is designed such that changing observing frequency requires only a rotation of the (asymmetric) sub-reflector to direct received signals to the appropriate feed horn. We also saw in Chapter 2 that to ensure a robust and undistorted measurement of radio signals, the various DDEs that affects the data need to be modelled and removed through

DD calibration. One of these effects is the primary beam effect. As we have already explained in Chapter 2, It very important to generate accurate PB model to account for them during calibration to recover the true sky brightness distribution. The VLA antennas primary beams have been modelled using a software called *Cassbeam* (Brisken, 2003).

The Wiener-Khintchin theorem states that the point spread function (PSF) of an antenna (or its far-field radiation pattern) is linked to the geometry of the aperture. Hence, any beam modelling technique designed to model the PB of an antenna (using the Wiener-Khintchin theorem) should take into account the physical parameters associated with the geometry of the antenna dish. *Cassbeam* modelling software was designed to use this framework. The software is a Cassegrain antenna optical ray tracer (Brisken, 2003). It uses the principle of optical ray tracing with some of the antenna physical geometry properties (contained in an input configuration file) to simulate the primary beams of a Cassegrain antenna for all polarisations. The input file contains the position of the feed, the shapes of the main reflector and sub-reflector and the position of the support structures (shown in Table 2.1 in Chapter 2). The software uses this file and then applies the concept of diffraction theory to derive the far-field pattern. It was initially designed for the VLA antennas but can be used to generate a PB model for any Cassegrain radio telescope. A set of *Cassbeam* beams were simulated using the default parameters of the VLA telescopes at the L-band frequencies which represents the ideal beams at this band. We refer to them as the Electromagnetic (EM), simulated or theoretical (or sometimes *Cassbeam*) beams. These terms may be used interchangeably to refer to beams generated with *Cassbeam* software.

4.5.2 Holography primary beams analysis

4.5.2.1 Two-dimensional power distribution of the beams

In Chapter 3, we presented some samples of the 2D morphology plots of holography beams, but here we will compare the holography beams with the EM beams. The Jones terms of the holography and EM beams were used to compute their 4×4 Mueller matrix terms (representing Equation (3.7)). In Figure 4.4[1], we present the $I \rightarrow I, I \rightarrow Q, I \rightarrow U$ and $I \rightarrow V$ terms at the frequency of 1.118 GHz (randomly selected) for one antenna (antenna 5).

[1]Note that $I \rightarrow I \equiv I$ in the first column.

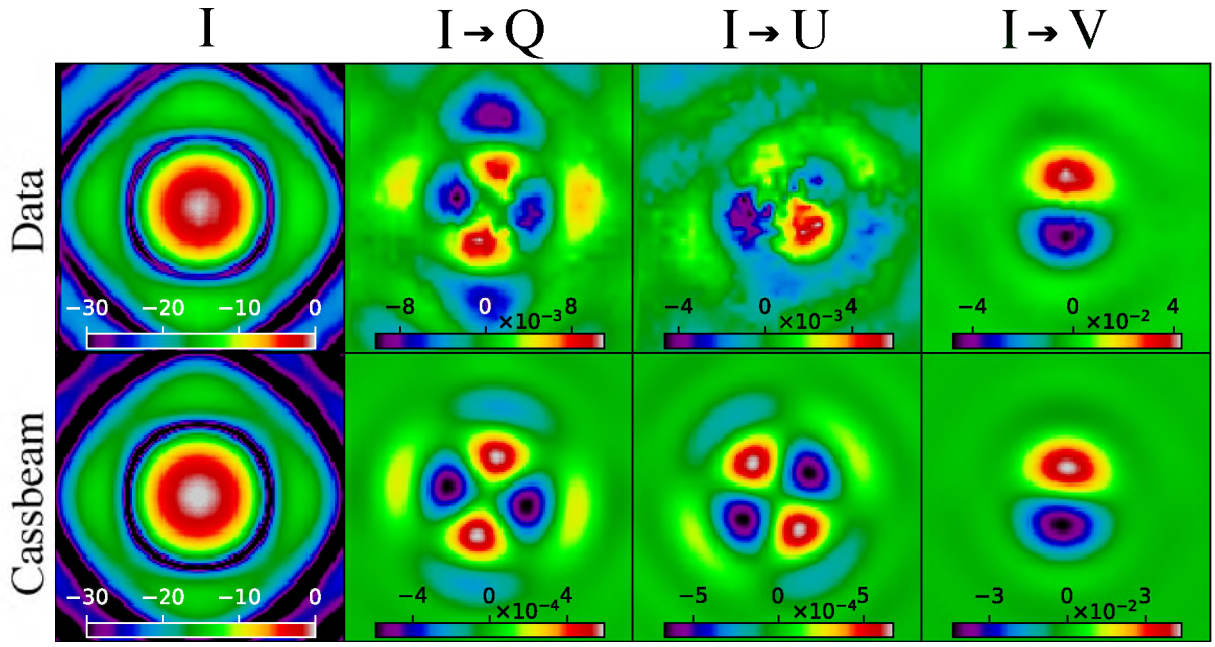

FIGURE 4.4: Comparison of Mueller matrix (Equation (3.7) first row) terms of the two-dimension holography plots (bottom row) and the EM (bottom row) beams images at 1.118 GHz. (Images are from Iheanetu *et al.* (2019) and $I \equiv I \rightarrow I$). The colour bars show the amplitude in dB.

These plots illustrate the difference between the expected EM polarised beam components and the measured holography components. These beams were normalised to their maximum gains and plotted in dB (or linear scale) to allow for coarse visual comparison.

The intensity distribution profile of the holography and EM beams in Stokes I are in good agreement and can be described to have (an approximately) circular main lobe, surrounded by a region known as the first null and then the first side lobe arranged in a set of *petals* arranged with 4-fold symmetry. These show that the major instrumental effects are reasonably captured by the *Cassbeam* modelling software. The holography primary beams Jones (LL and RR) terms have some distortions in shape and distribution of the side lobes as well as their extent. The position and the distribution of the side lobes and nulls are frequency dependent.

The leakage terms, $I \rightarrow Q$ and $I \rightarrow U$ have 4-fold petals shape while $I \rightarrow V$ have 2-fold *petals*. The holography and EM beams are in good agreement, particularly for $I \rightarrow Q$ and $I \rightarrow V$ terms, but $I \rightarrow U$ term appears to be a lot weaker in the holography beam. We suspect that this observed difference may have originated from the effects caused by sub-reflector support structures which effected Stokes Q more than U and the

fact that the antenna has an asymmetrical sub-reflector. Investigating and explaining the reason for this difference is beyond the scope of this work.

There is an overall beating effect in frequency shown by all the polarisation terms (RR, LR, RL and LL) of the holography beams (which are caused by a standing wave), which we investigated (in Section 3.4). The EM beams show no such effect (because the ripple effect is not incorporated into its model). The beams' sizes slowly decrease with increasing frequency, while the positions of the beams centres, main lobes and distribution of the side lobes show some frequency beating.

We started our beam analysis by comparing the Mueller terms of the holography beams with the EM beams. The EM beams, on average, capture the overall power distribution of the holography beams, up to the second null. Here we will focus on using the Jones RR, LR, RL and LL representations of the measured and simulated beams. The holography measurements were done by observing polarised, and un-polarised sources and the resultant (measured) beams display a high degree of distortion, particularly the cross-terms.

To quantify the behaviour of the target antennas, we calculate the mean beam at each frequency separately for each polarisation over the 12 antennas. We also evaluate the dispersion of the beams of the different antennas with respect to the normalised average beam. To achieve this, we first calculate the mean holography beam of the 1024 beams (1024 MHz frequency band from 1008 - 2031 MHz in steps 1 MHz). Then we calculate the mean square error (MSE) between the mean and normalised beams. Conventionally, one uses the MSE to evaluate the quality of model data w.r.t the reference data. We used it to serve here as a measure of reference to quantify/evaluate the dispersion of each antenna's beams about the average behaviour as a function of frequency for each of the correlations. Figure 4.5 shows the variation of the MSE (in a logarithmic scale along the y-axis).

The variations of the MSE against frequency (ν) for each antenna for RR and LL have a smooth linear trend that follows a decreasing power law. These plots indicate that the beams are closer to the mean beam at higher frequencies than at lower ones. In general, the RR and LL MSE varies between 0.01-1% while LR and RL are approximately 1%.

We observed that the LL polarisation could be split into two groups of antennas. Group 1 consists of antennas 10, 20 and 27, with MSE of about $5 \times 10^{-3} - 8 \times 10^{-3}$ and group 2 consists of antennas 5, 8, 12, 14, 15, 26 and 28 with MSE of about 10^{-3} showing approximately equal behaviour. Antenna 6 stands out alone having an MSE of about

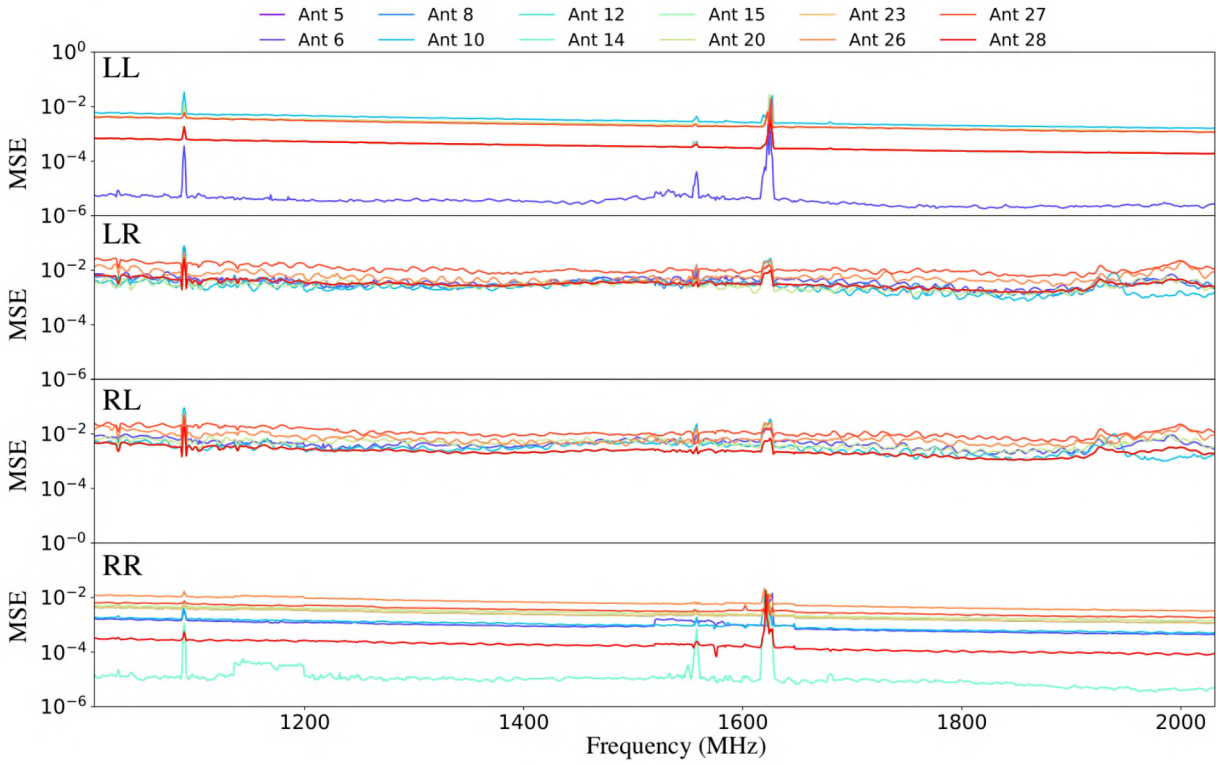

FIGURE 4.5: Mean square error across 12 antennas holography beams as a function of frequency (images are from Iheanetu *et al.* (2019))

8×10^{-3}. There is no obvious explanation per this grouping w.r.t the target antenna position or location in the array. The dispersion of the behaviour of the RR beams for 12 antennas was higher ranging from $10^{-5} - 10^{-2}$, which made it difficult to have any form of grouping. Using the same figure of merit, we do not notice any groupings in the MSE between the RR and LL beams.

The LR and RL (beams) MSE range at about 10^{-2} on the average. The MSE dispersion is more compact except for antennas 26 and 27, with a noisy pattern across frequencies. These plots give insight to the general behaviour of the holography beams with reference to the mean beams.

This rough investigation is the first step towards a better understanding of the relevant fluctuations in the measured data. However, a deeper phenomenological beam modelling is required for better representation of the beams. The next section presents how we unfolded the contributions of the different fluctuations in the beams and reconstruct the beams with a small number of coefficients with high fidelity.

4.5.2.2 Effective FWHM as a function of frequency

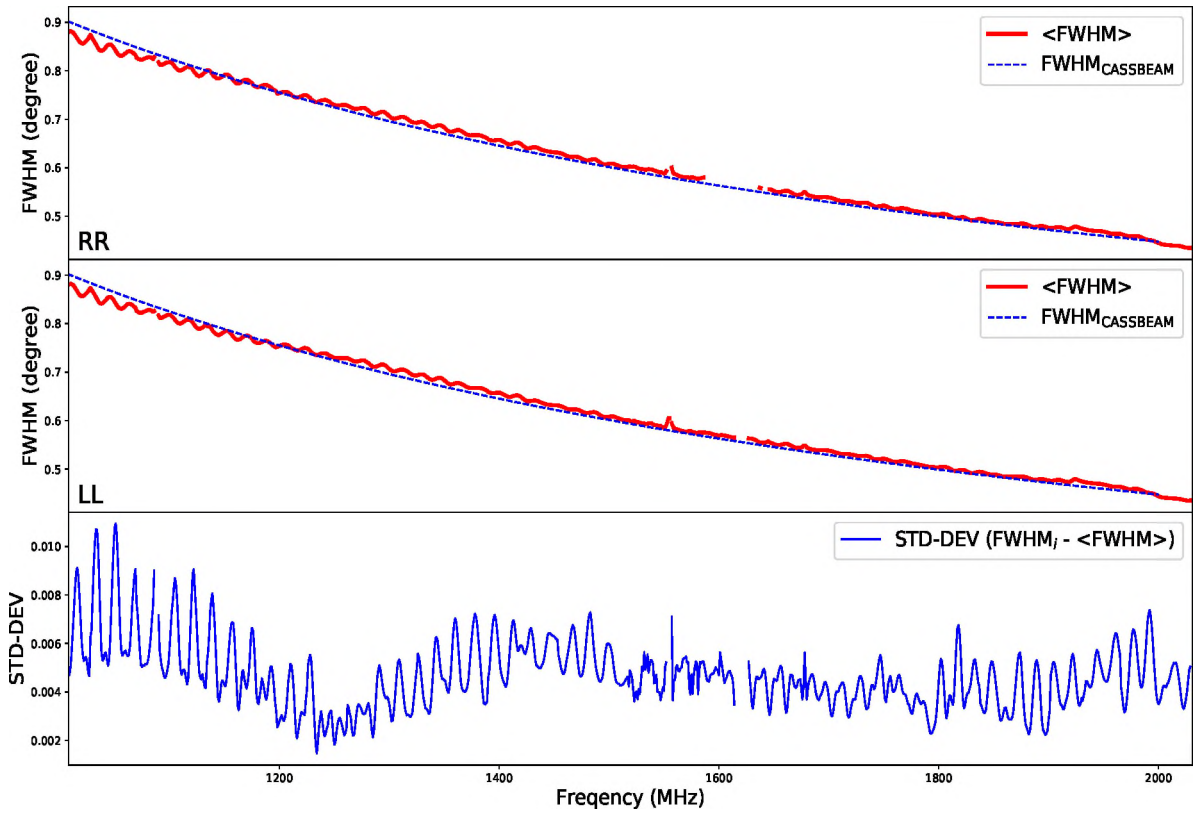

FIGURE 4.6: Plots of the mean FWHM of the holography beams against frequency along side the FWHM of simulated EM models (labelled *CASSBEAM*) against frequency for the RR top panel and LL middle panel. The standard deviation of the residual after removing the mean across the 12 antennas for LL beams is presented in the bottom panel.

In chapter 3, we have shown spectral plots of the holography beams sizes and centre offset positions which demonstrates that the beams exhibit some oscillatory characteristics (called the *ripple*). To compare the FWHM and offset of the holography with that of the simulated EM beams, we again extract the EM beams parameters. A detailed description of the spectral behaviour of the holography beams' FWHM and offset have already been given in Section 3.4. In Figure 4.6, we present the variation of the holography and EM beams FWHM as a function of frequency for 12 antennas, together with the average FWHM. Gaps in the plots' correspond to the frequencies that were affected by high RFI.

Figure 4.6 shows that the FWHM of the EM model captures the first order characteristic behaviour of the holography beams well but misses the (second order) ripple characteristics.

4.5.2.3 Beam squint and pointing error

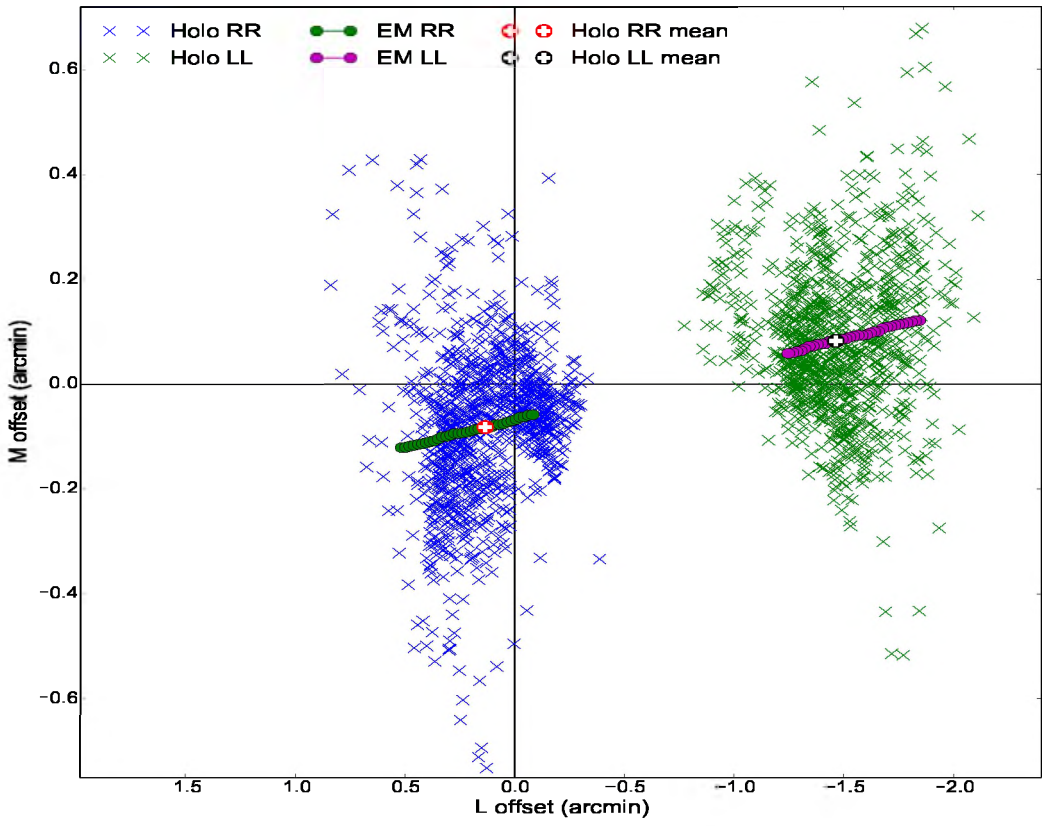

FIGURE 4.7: Two dimensional plots of the one antennas (antenna 5) pointing offsets for LL (green x markers) and RR (blue x markers) polarisations with their mean off-sets (with cross on black and red background respectively). The green and magenta filled circle makers are for EM RR and LL polarisation beams respectively. This is plotted in the antenna frame.

Recall that the VLA antennas are off-axis Cassegrain antennas, so their feeds are installed in an off-axis position in a ring around the optical centre of the main collector. This results in beams whose centres are slightly offset w.r.t the optical centre. This is known as beam squint. The *Cassbeam* software takes this into consideration when modelling the primary beams of a Cassegrain antenna. We presented in the lower panel of Figure 3.7, one antennas' holography beam centre offset as a function of frequency. Let us now compare the same antennas beams centre offset with that of the EM model. Figure 4.7 presents exactly this in two dimensions. This shows that there is good agreement between the holography and EM beams offsets. The EM model takes into account this small shift of the antenna feeds horn to produce a reasonable approximation of the antenna beam squint.

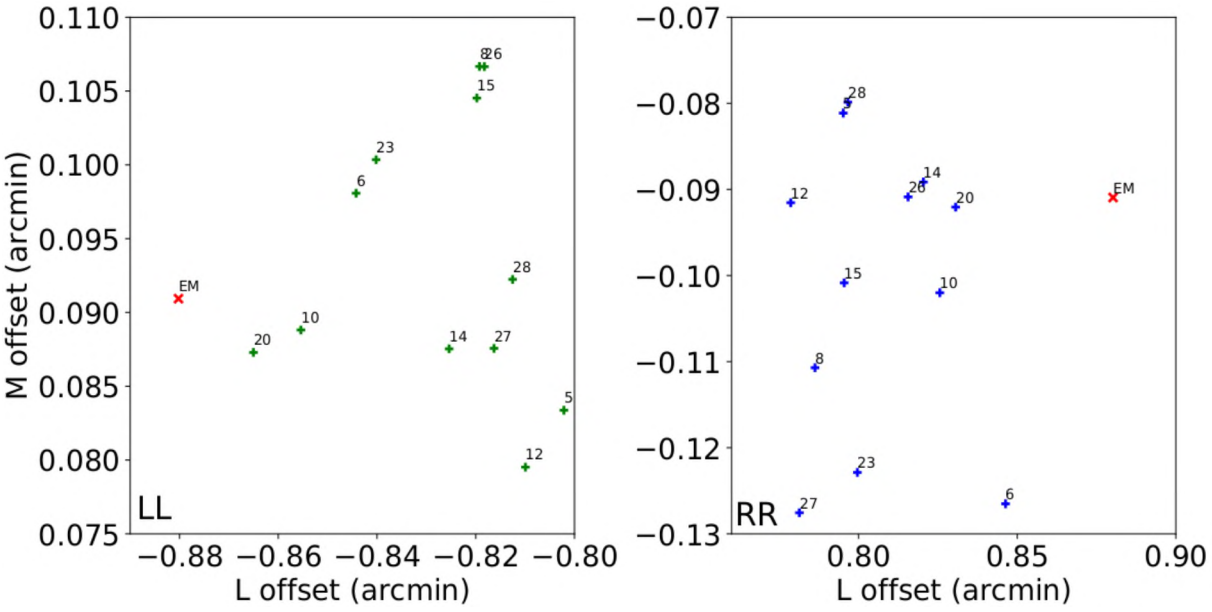

FIGURE 4.8: Two-dimensional plots of each antenna mean pointing offsets (across frequencies) for LL (green) and RR (blue) elements. The red mark is the mean pointing offsets for corresponding EM polarisation beams.

The average plot of each antenna mean offset position (in two dimensions) with that of the EM beams (across frequencies - 1.008 to 2.031 GHz) is presented in Figure 4.8. The mean dispersion of the holography beams LL pointing offset compared to the that of the EM beams are -49.64 ± 1.11 arcsec and 5.62 ± 0.54 arcsec along l and m axis respectively. While that of the RR Jones elements are -48.36 ± 1.22 arcsec and -6.07 ± 0.98 arcsec along l and m axis respectively for RR. The mean pointing offset for EM beams are 52.81 and -5.47 arcsec along the l and m axis respectively for the RR and -5.47 arcsec and 52.81 arcsec along the l and m axis respectively for the LL elements.

We found no correlation between the physical location of the antenna on the ground and the amplitude of the average beams of the LL or RR or their offsets for the antennas. And we did not observe any form of grouping for the antennas from Figure 4.5. We also did not notice any form of grouping in Figure 4.8 which suggested that the beam centre offsets are random and are entirely due to pointing errors. Plots of the 12 antennas l and m offset as a function of frequency are presented in Figure 4.9.

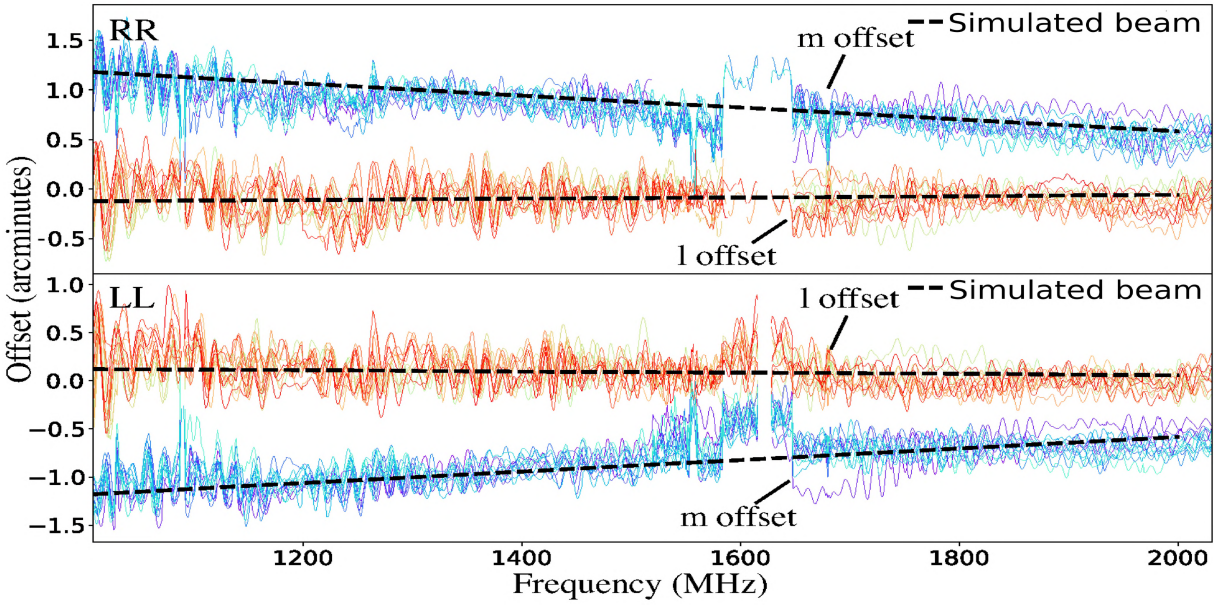

FIGURE 4.9: Plots of the l and m pointing offsets against frequency for RR (top panel) and LL (bottom panel) polarisations for 12 antennas together with that of the EM simulated ones (dashed lines) for comparison.

4.5.3 Search for sparse beam representations

4.5.3.1 Sparse representations

Looking for sparse representation of a signal gives a good insight into the nature and the spreading of energies in the signal. A signal in one space can be efficiently represented in another space (in a compressed form) if the number of (dominant) features that describe the signal is small. A typical instance of this is a cosine function that is dependent on time and frequency (ν). In the signal domain, a lot of coefficients with the Nyquist frequency sampling rate of 2ν are needed to represent sinusoidal behaviour of this function in time, without a priori knowledge of the signal nature. If one moves to the Fourier domain, only two complex coefficients are required to represent this signal (i.e two delta Dirac functions). In this case, the sparsifying basis is the Fourier kernel.

In this work, we attempt to study the variations in the holography measured beams and model these variations by using a set of decompositions in a sparse basis. Each file of the holography beam is a three-dimensional cube that has two-dimensional spatial frames at every MHz step in frequency for each co- and cross-polarisations (real and imaginary per polarisation and for each antenna). We will search if there exists a basis through which the data could be represented through a set of relevant coefficients that can be used to

represent the spatial and spectral variations. We seek to find a set of degrees of freedom which will be *new* unknown parameters for the beams representation. These parameters will first be incorporated into and solved for during the process of direction-dependent calibration.

Two major approaches can be used to find suitable sparsifying bases. One allows us to get bases directly from data, i.e derived representation from data (examples are using SVD, PCA, dictionary learning techniques, etc.). Using already existing bases which gives a good prior picture of the signal. Examples are Zernike modes, wavelets, spherical harmonics, Fourier components to mention but a few.

The Characteristic Basis Function Patterns (CBFP) method is one of the data-driven techniques that can be used to model the primary beams of an antenna from measurement and/or from simulation.CBFPs have been used by Young *et al.* (2013) and Mutonkole and de Villiers (2015) to model the MeerKAT beams, where they applied PCA on simulated beams. The former started with a primary CBFP model for ideal antenna configuration and modelled a secondary CBFP for each expected perturbation in antenna configuration. The latter starts with a simulated model of the primary beam at a single frequency, then sampled it in a few directions, applied the SVD, extracted expansion coefficients and used linear algebra (on the coefficients) to extrapolate the beam at any other frequency (within some chosen frequency band).

Modified *Cassbeam* software: *Cassbeam* has been used to model the primary beams of the VLA antennas before 2017, but recently a modified version of the *Cassbeam* software has been developed (Jagannathan *et al.*, 2018) to model the primary beams of the VLA antennas. In this new version, they developed a fitting algorithm that wrapped the *Cassbeam* input parameters of the EM model to match the frequency behaviour of the holography beams. In other words, they generated a set of new optimal parameters for the feed x- and y-offset, the radius of the sub-reflector - central hole (*Rhole*) the aperture illumination taper function (*Taper1, Taper2* and *Taper3*) for each frequency (for R and L feeds). The *ripple* characteristic of the holography beams was then absorbed by these geometric parameters. In Figure 4.10, we show plots of the spectral variation of some of these parameters.

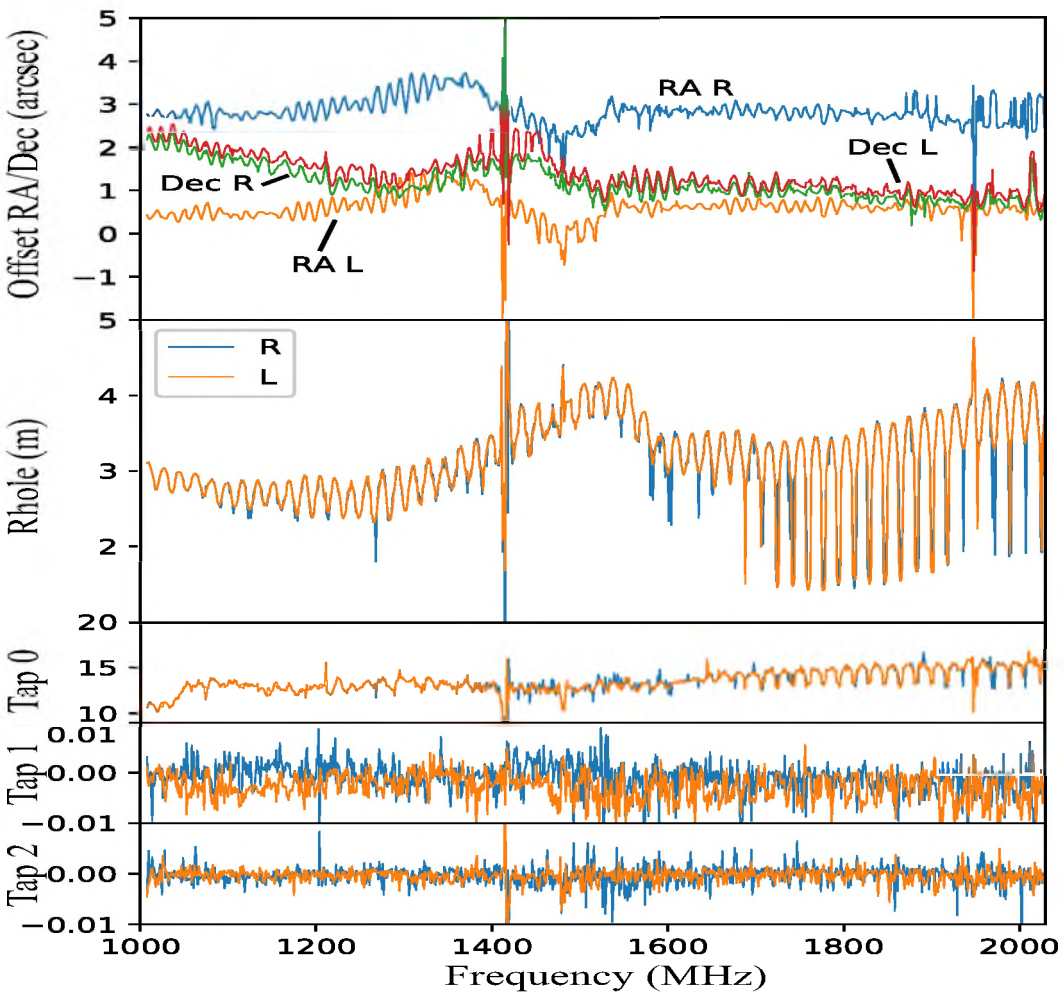

FIGURE 4.10: Plots of the optimised fitted new *Cassbeam* geometric parameters against frequency: for X- and Y-offset (top panel), central hole radius - Rhole (second panel) and aperture illumination taper coefficient (0, 1 & 2) for L and R feed sensors.

4.5.3.2 Data-driven model of the primary beam

Holography measurements of an antenna were originally designed for inspecting surface accuracy of the antennas aperture. Amplitude and phase distribution over the aperture can be related to the surface errors which could be adjusted physically on the reflector surface to correct the errors. Computing the inverse Fourier transform of the holography measured beams, we get the power distribution over the aperture. From its definition, all physical information of the aperture illumination resides inside the aperture. So, one can use hard-thresholding (HT) or aperture plane masking (AP) operation to filter out all information lying outside the synthetic aperture by putting the values of those pixels to zeros, thereby removing distortions because it retains only measured data which lie within the aperture plane. This is used as the first step in processing the data and is

the denoising step which produces smooth beams. To present this mathematically, let $\boldsymbol{E}(\alpha, \beta)$ be the noise dominated far-field radiation pattern at some point in space and the direction cosines be $\alpha = \cos\phi \sin\theta$ and $\beta = \sin\phi \sin\theta$. The aperture illumination function of the antennas reflector surface (from the "beams holography") is given as (Baars *et al.*, 2007):

$$\boldsymbol{F}(\xi, \eta) = \frac{-iRe^{ikr}}{4\pi^2\lambda} \iint \boldsymbol{E}(\alpha, \beta) e^{ik(\alpha\xi+\beta\eta)} d\alpha d\beta, \tag{4.22}$$

where $k = \frac{2\pi}{\lambda}$, the distance of the point from the aperture is R and the integral is taken over the finite support of the aperture function. Measuring $\boldsymbol{F}$ this way is known as 'dish holography'. Hence, to get $\boldsymbol{E}^{\mathrm{a}}$ (a denoised beam), we can apply Fourier transform analysis to $\boldsymbol{F}$ and throw away all the Fourier modes that fall outside the physical aperture plane (A). Mathematically, this is expressed as:

$$\boldsymbol{E}^{\mathrm{a}}(\alpha, \beta) = \frac{i\lambda e^{ikr}}{R} \iint_A \boldsymbol{F}(\xi, \eta) e^{ik(\alpha\xi+\beta\eta)} d\xi d\eta). \tag{4.23}$$

Recall that each 3-dimensional holography cube (for one antenna) has one frequency and two spatial axes. The beam at each frequency has an $N^2_{\mathrm{pix}} = N_{\mathrm{x_{pix}}} \times N_{\mathrm{y_{pix}}}$ pixels ($N_{\mathrm{x_{pix}}} = N_{\mathrm{y_{pix}}} = N_{\mathrm{pix}}$). We flatten each 2D spatial beam frame into a 1D vector, that is stacked together to form a rectangular matrix $\boldsymbol{D}$ with dimension $N_\nu \times N^2_{\mathrm{pix}}$. Where N_ν is the number of samples and N^2_{pix} is the number of features (i.e N_ν and N^2_{pix} are the spectral and spatial axes respectively). We then perform PCA on $\boldsymbol{D}$.

We stack each antenna's complex-valued holography beams together for each polarisation separately before performing PCA on it as outlined in Section 4.2.3. The first twelve strongest *eigenbeams* are presented in Figure 4.11.

The frequency behaviour of some of these complex coefficients are shown in Figure 4.12 (the real and imaginary parts). These plots show that the amplitude coefficients decrease with increasing order of i. This implies that increasing accuracy can be obtained by increasing the number of coefficients. There is a smooth trend with the first three coefficients, but for $i > 3$, it starts to manifest some ripple behaviour. The ripples amplitude in the first three PCA components has a peak-to-peak variation of about ≈ 1.3 arbitrary units. Starting from the 4^{th} order onwards has a ripple frequency of ≈ 17 MHz in addition to having a smooth trend in a noisy background.

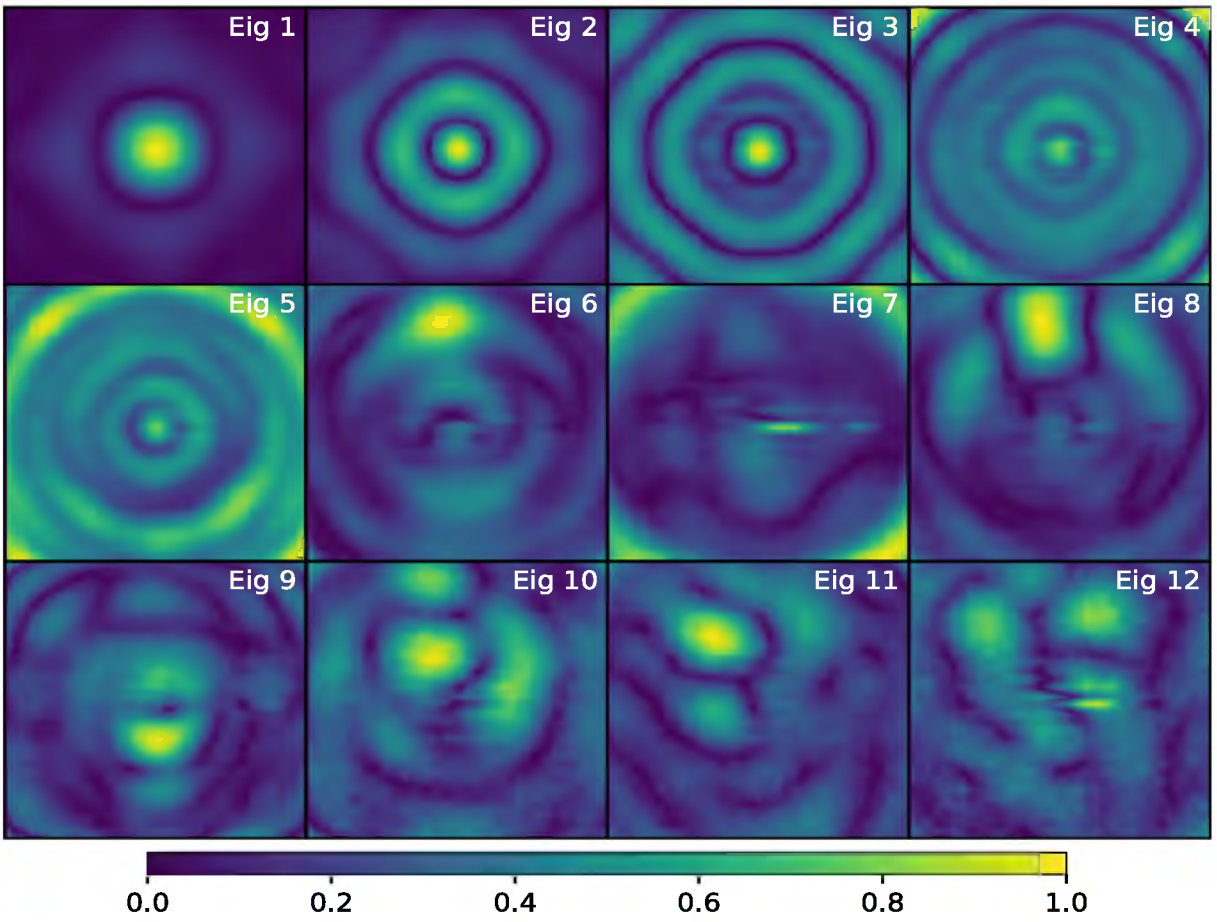

FIGURE 4.11: 2D plots of antenna 6 first 12 (normalised) dominant PCA modes which were ranked among 5th in the L-band.

At this stage, we need to mention that we observed a similar trend for the spectral behaviour of the coefficients for all antennas (which make describing all of them redundant) and points towards the direction of possibly using one model to describe all. To model the beams of each antenna, we compute a common set of *eigenbeams* which can be used to reconstruct all the beams. Now, having obtained the common basis for each antenna, we can generate new coefficients by calculating the dot product of each antenna's beam with its corresponding coefficients. This common basis produces a relative reconstruction error that ranges from 2 to 4% across the antennas. In the next section, we will use a different technique that decomposes the beams morphology into bases at each frequency.

A plot of the eigenvalues of the PCA decomposition of $\boldsymbol{D}$ is present in Figure 4.13, which shows the graph of the first 200 singular values for this decomposition. We can see from this graph that there is no significant change in the amplitude of the singular values after the 20^{th} singular value and the first 20 singular values contributes $\approx$ 88 % $\left(\left(= \frac{\sum_{(i=1)}^{20} \sigma_i}{\sum_{(i=1)}^{N} \sigma_i} \times 100\right)\right.$ of the total amplitude, where σ is singular value and N is the total

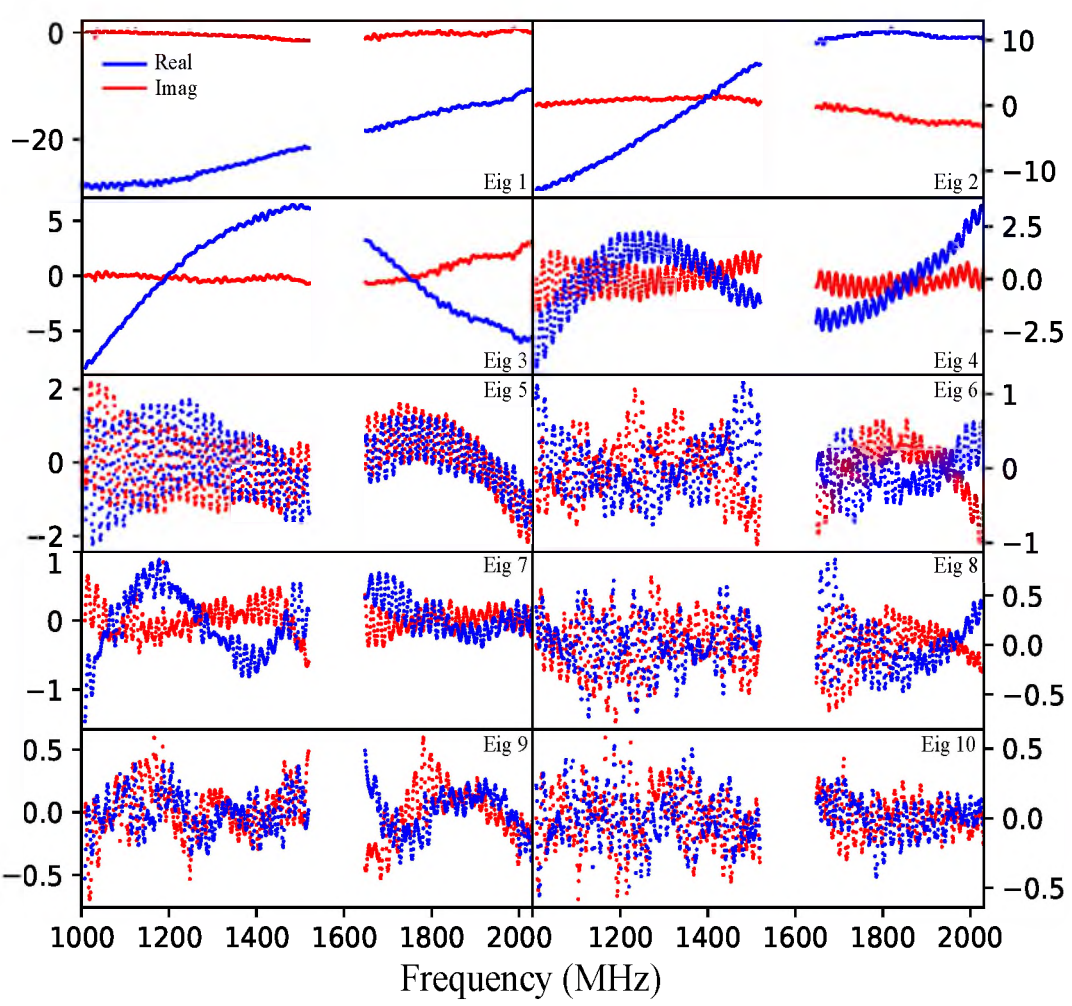

FIGURE 4.12: Plots of the frequency behaviour of the first 10 PCA components (real & imaginary) for antenna 6.

number of singular values) of the sum of all singular values. This implies that using only 20 PCs is sufficient to produce a reasonable reconstruction of the holography beams.

4.5.3.3 VLA beam model using ZP

A brief review of Zernike decomposition has been presented in Section 4.3 and this recipe is applied here to model the VLA L-band PB. We will start our application of Zernike with the holography beam cube that has its FWHM scaling and offset spectra features intact.

Figure 4.14 presents plots of the first sixteen dominant Zernike polynomials used to model the beams for antenna 6. Each Zernike mode has a corresponding mode index called the Noll index. One can use small (respect to large) Noll indices to model a typical frequency

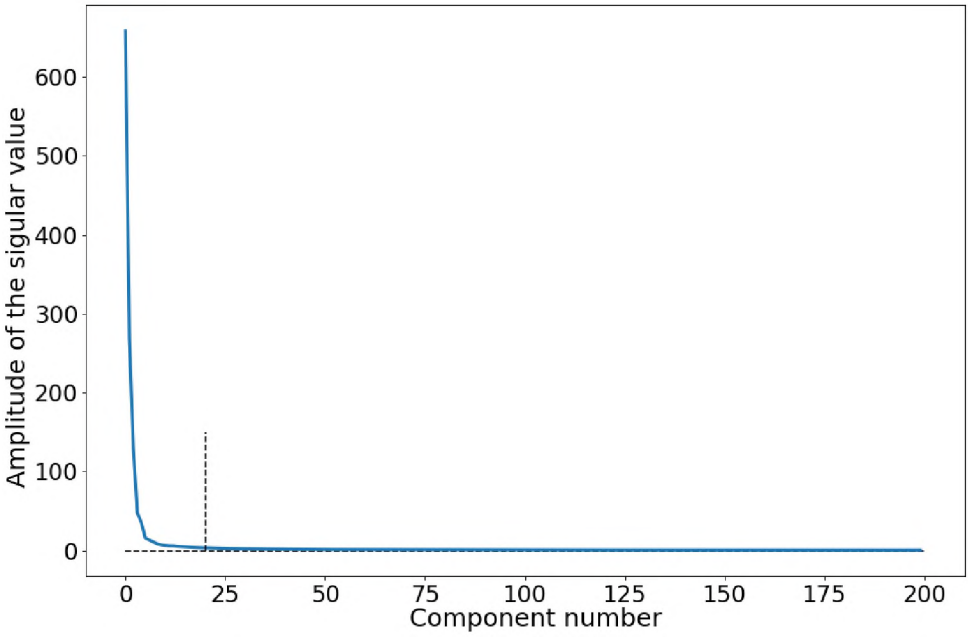

FIGURE 4.13: Plot of the first 200 singular value components of the PCA decomposition for antenna 6. The vertical dash line indicates the position of the 20^{th} singular value.

varying beam at low (w.r.t high) frequencies. A set of Zernike modes are found to be dominant in representing the beam at specific set of frequencies. Owing to the frequency scaling nature of the beams across the L-band, different sets of Zernike modes may be required to represent each beam shape. To model the beams across the whole L-band, one needs to monitor the Noll indices that are associated with the dominant modes. To implement this, Zernike decomposition was performed at each channel over the first hundred Zernike modes and the Zernike modes were ranked in order of decreasing energy. Figure 4.15 presents the colour map which represents the distribution of the ranks of the first 23 dominant Zernike modes plotted against frequency. The y-axis is ranked from 0 (= most dominant) to 100 (= least dominant). Each colour represents a single tracked Zernike mode. Small Noll indices dominated at low frequencies hence, occupying the top ranks at the start of the band but not so at high frequencies where they quickly loose dominance. Low frequencies were dominated by small values of Zernike mode indices (represented by red, orange, brown colours etc.) thereby filling the top ranks at the low frequencies. On the other hand, large values of mode indices dominated at high frequencies (represented by for examples blue, light blue, cyan colours etc.) and filling up the top rank. At about 1.5 GHz, there is a crossing of the dominance of the mode indices from small to large. The main reason for these variations in Zernike mode dominance as one moves from low to high frequencies is that (seeing as we are keeping the field of view fixed) at high frequencies the method is forced to reconstruct more sidelobes, which

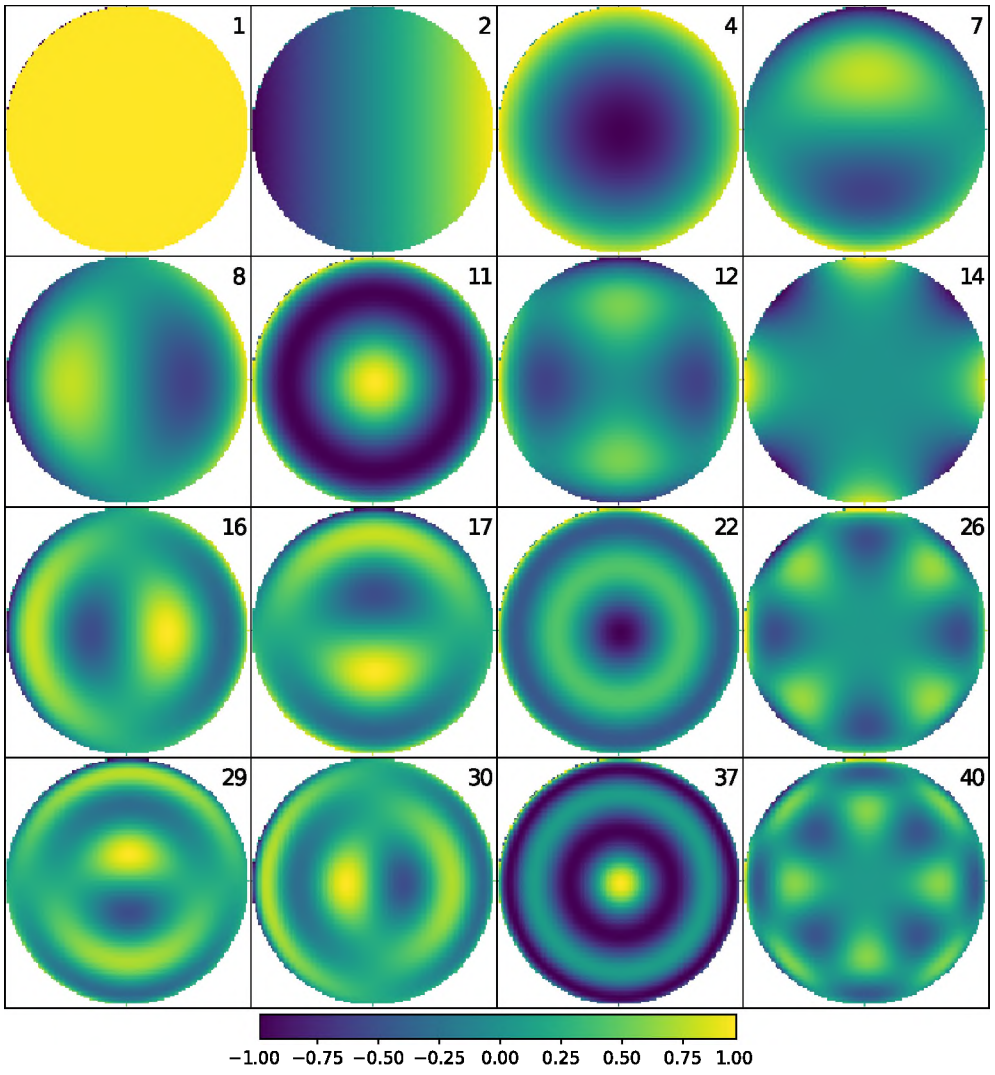

FIGURE 4.14: 2D plots of antenna 6 first 16 (normalised) dominant Zernike modes. The insert number is the associated Zernike polynomials mode index.

require a larger contribution from higher-order Zernike modes.

Selection of relevant Zernike mode indices is required for one to be able to represent the beams with a high level of accuracy. One can divide the band into two sub-bands and generate two sets of Zernike modes, one for each sub-band.

The selected Zernike modes have complex coefficients which have their unique spectral characteristics. A plot of these coefficients is presented in Figure 4.16, the values inserted are their corresponding Zernike mode indices. To demonstrate our earlier point, if one considers the graphs for the mode indices 4, 11 and 14, they are dominant at low frequencies while indices 56, 64 and 79 are dominant at high frequencies. Zernike mode indices 22 and 26 for example, are dominant at the middle frequencies.

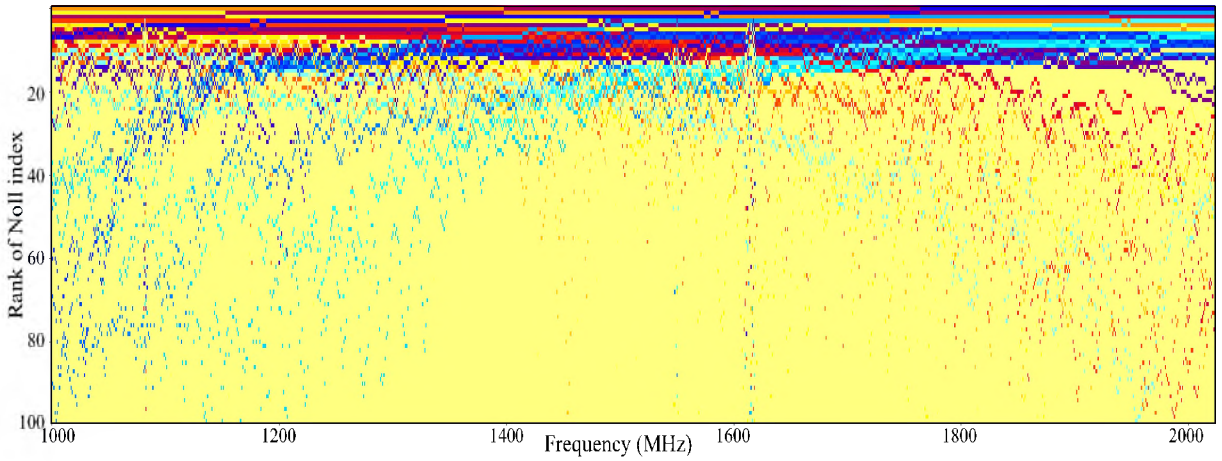

FIGURE 4.15: Plots of 23 dominant Zernike modes against frequency (mode indices = 1, 2, 4, 7, 8, 11, 12, 14, 16, 17, 22, 26, 29, 30, 37, 40, 44, 47, 56, 60, 64, 79, 80) for antenna 6. Along the y-axis is plotted with the rank of the dominance - starting from the most (0^{th}) to the least (100^{th}) dominant modes. Each colour represents a single Zernike mode index.

It is important to mention that Zernike mode index 1 is the piston term that corresponds to the DC component of the beam-mode index 2 describes an east-west tip-tilt error which is related with the pointing error. Mode index 3 describes the north-south tip-tilt and is not part of the dominant mode indices in the complete band, which is consistent with the offset of the LL beams for antenna 6. The morphology of the beams at each frequency is unique and so is the associated set of dominant Zernike mode indices and the corresponding coefficients. The beams scale down smoothly with frequency and there is a transfer of energy to subsequent Zernike modes.

4.5.4 Spatial and spectral compression

4.5.4.1 Compressing the spectral information

The decomposition coefficients computed from SVD/PCA and Zernike orthonormal basis of all holography beams cubes (and for all target antennas) have been stored for further investigations. Notwithstanding the method of spatial decomposition of choice, we also want to adequately reconstruct the beams with reduced number of coefficients across the L-band. For each decomposition method, we have a series of N_{freq} complex coefficients which describes the spectral spreading of this basis vector contribution. In this section, we use a one-dimensional decomposition and low-rank approximations to perform a spectral encoding of the coefficients. Although the storage requirements for a primary beam model

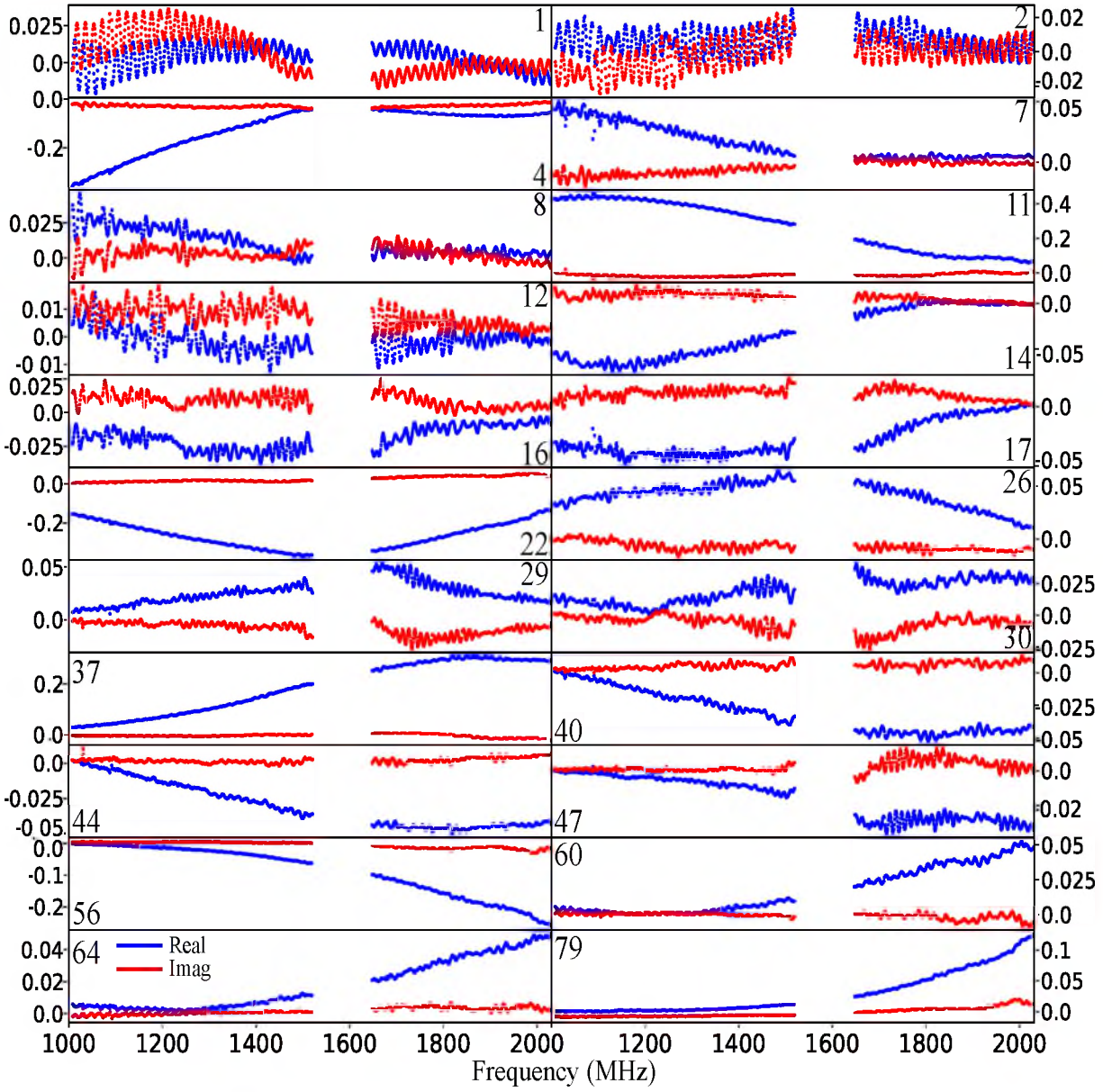

FIGURE 4.16: Spectral plots of the behaviour of the coefficients associated to the first 22nd dominant Zernike modes (real and imaginary - plotted in Figure 4.15) for antenna 6 that were ranked amongst the 20th (in the L-band). The place where we have gaps in these plots corresponds to flagged data.

at full frequency resolution are not particularly onerous compared to the raw data sizes of modern instruments, we still consider a reduction in the number of representation coefficients to be desirable. Such models offer better robustness against measurement noise, are easier to distribute along with software code, and can be accurately recomputed on any given grid. In some computational contexts (such as GPUs), memory bandwidth is limited compared to compute capacity, so computing a model on the spot from relatively few coefficients could offer a real increase in performance.

We stated earlier in chapter 3 (see Section 3.4), that some of the measured data was affected by RFI and we excluded them from our analysis in order to avoid the contribution of non-physical beams features into the decomposition. The whole L-band data (excluding the RFI affected ones) for a single polarisation was decomposed with SVD/PCA (at once) while per frequency decomposition was done with Zernike decomposition. Regardless of the two decomposition methods, we have a set of coefficients of similar size in frequency and we can now compress the spectral information in them.

4.5.4.2 Interpolating/predicting beams at missing frequencies

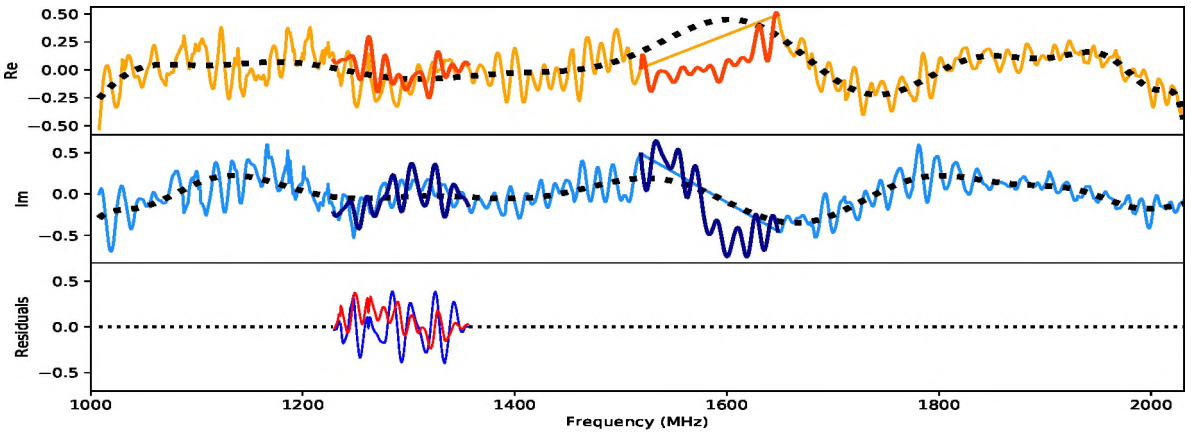

FIGURE 4.17: Plots of the first eigenvectors (real and imaginary in orange respectively blue colours) of the SVD/PCA decomposition and the predictions of the missing frequencies with Papoulis-Gerchberg algorithm (dark orange and navy blue) and DCT (discrete cosine transform) decomposition from 1.519 - 1.648 GHz. The data is in orange and blue, Polynomial interpolation of order 30 in black and Papoulis-Gerchbeerg in dark orange and navy blue - in the top and middle panels (for real and imaginary) respectively. In the bottom panel are the residuals of the Papouli-Gerchberg's interpolations - real and imaginary in red and blue colours respectively.

The plots of the coefficients which are associated with the first SVD/PCA decomposition eigenvectors (real and imaginary) for the holography cube of antenna 6 are shown in Figure 4.17. These graphs show a fast oscillating trend which corresponds to the ripple frequency that is superimposed on by a smooth low order variation. We used the DCT type II algorithm (Equation (4.20)) to decompose the signal and obtain DCT coefficient, then reconstructed this trend with is DCT type III algorithm (Equation (4.21)) using the DCT coefficients. Recall that the gaps (or missing portions) in the graphs (see Figure 4.16) correspond to RFI-affected data that were excluded from our analysis. We need to interpolate these missing regions to generate beam models for those frequencies. Using

a first-order polynomial or linear interpolation technique capture the trend but was only able to interpolate a straight line (with no ripples) in the sub-band where we have some missing information - RFI-corrupted data. Due to the discontinuity of data at flagged frequencies, a low-ranked DCT reproduced a strongly biased amplitude of the ripples with slow variation at those regions. We applied the PGA (Papoulis-Gerchberg algorithm (Papoulis, 1975)). PGA is a well-known algorithm that uses iterative thresholding to get signal restoration by *inpainting*. *Inpainting* is a technique used in image processing to reconstruct damaged or missing regions in an image (Guillemot and Le Meur, 2014). PGA was designed to be used to interpolate band-limited signals. Assuming that the signal is band-limited (i.e the signal frequency is larger than the related ripple frequency), we can apply PGA in combination with DCT to interpolate the signal. In Figure 4.17, we present the successful interpolation of the RFI affected (or missing) part of the signals in comparison with low-order polynomial interpolation. The polynomial captured the average behaviour of the signal only (missing out the ripple oscillations). It only uses a maximum of three thousand iterations for the PGA decomposition to converge for each coefficient keeping the 50 most dominant DCT modes.

For us to test how robust this method is, we blanked out the signal in the regions where we have data from 1.3 to 1.35 GHz band. There is good agreement between the signal and the interpolation at this band (see Figure 4.17), which shows that the PGA approximately estimates the coefficients at the missing frequencies (although, this does not serve as a method to get the perfect original missing underlying signal). The one-dimensional signal can further be compressed with DCT decomposition with the lowest possible bias. The reconstructed residuals in the test gap show that, while our interpolation approach is more accurate than naive linear interpolation, its accuracy is still somewhat limited. In general, this is a difficult interpolation problem, and will require further research to address properly. We provide our scheme on the premise that at least an approximate ripple interpolation scheme is better than straight linear interpolation. Due to its consistently high RFI occupancy at the VLA site, the 1.519-1.648 GHz range is known as the "Zone of Death" (Perley, priv. comm.) and tends to be flagged wholesale in real data, so the need for accurate PB models in this frequency range is largely academic.

In summary, this interpolation captures both slow and fast trends in the signal. This method allows for the selection of the dominant features of the signal. The PGA technique was applied to interpolate both SVD/PCA and the Zernike coefficients at missing frequencies.

4.5.4.3 One-dimensional compression for the data representations

One can further compress the spectral information in all coefficients after signal restoration. The final method of beams representation will involve two steps: one, spatial beams encoding and approximation using representative basis functions and; two, spectral encoding of the coefficients. SVD/PCA or Zernike decomposition was used to perform spatial encoding while spectral encoding of the coefficients was done using DCT (with a maximum of 50 DCT modes). The compressibility of this method was evaluated by counting the total number of coefficients that are required to reconstruct the full beams. A complex-valued matrix of size $4 \times N_{\text{freq}} \times N_{\text{pix}}^2$ can be used to represent the complex full-Stokes holography beams cube. Let our desired ranked approximation of beam be (k_{lm}, k_ν), where k_{lm} is the cut-off spatial and k_ν cut-off spectral ranks. SVD decomposition requires the computation of $\boldsymbol{v}_{lm}^{\text{T}}$ (with dimension $4 \times N_{\text{freq}} \times N_{\text{pix}}^2$) which is the first k dominant *eigenbeams* (where k is the number of selected coefficients or eigenbeams). $\boldsymbol{u}_{k_{lm}}, \boldsymbol{w}_{k_\nu}$ ($N_{lm} \times N_{\text{freq}}$ matrix) are the corresponding coefficients as a function of frequency. Then, modelling the coefficients with the first k_ν DCT modes will require the total number of coefficients per antenna per polarisation given as:

$$N_{\text{SVD}} = 2(k_{lm}k_\nu + k_{lm}N_{\text{freq}}N_{\text{pix}}) \tag{4.24a}$$

$$N_{\text{Zernike}} = 2(k_{lm}k_\nu + k_{lm}), \tag{4.24b}$$

where the multiplication by 2 is to account for the modelling of the real and imaginary parts. The first terms in Equations (4.24) are the contribution of the coefficients in each case; while the second term in Equations (4.24a) and (4.24b) are the contributions of the eigenbasis ($\boldsymbol{v}_{lm}^{\text{T}}$) and list of the Zernike mode indices respectively. The Zernike decomposition also requires an additional list of k_{lm} Zernike mode indices associated with $k_{lm}k_\nu$ coefficients.

One can compute the compression factor (CF) as the ratio of $\frac{N_{\text{freq}}N_{\text{pix}}^2}{N_{\text{2D}}}$, where the number of spatial scale N_{2D} could be N_{SVD} or N_{Zernike} for SVD or Zernike decompositions respectively.

The new version of *Cassbeam* (discussed in Section 4.5.3.1) can compress these coefficients while keeping the other parameters constant, so that:

$$N_{\text{Cassbean}} = 7k_\nu + N_{\text{CB}}, \tag{4.25}$$

where the first term is the contribution of the optimised X- & Y-offsets, Rholes and Taper 1, 2 & 3 and N_{CB} are that of the fixed parameters.

When applying the HT step (see Section 4.5.3.2) on the inverse FT of the holography data, the pixelated coefficients corresponding with the aperture plane are stored. Owing to the FFT frequency dependence, 30, 37 and 57 complex coefficients had to be stored for 1.008, 1.408 and 1.908 GHz respectively. Approximately 1040 coefficients will be required to estimate the data across the 1024 MHz bandwidth in the L-band if one assumes a linear relationship between the number of coefficients (required to reconstruct the data) and frequency. Seeing the low quality of the reconstructed beams (in Figure 4.18) implies that HT is not the best choice to compress the holography beams representation.

4.5.5 Beam reconstruction accuracy compared to compressibility

The accuracy of the reconstructed beams is now compared with one another and the required number of coefficients. All the beam decomposition techniques that we have presented have their advantages and disadvantages. After computing the series of $N_{\text{freq}} \times N_{lm}$ coefficients, they can be used to generate a beam cube that approximates the holography beams. This is achieved by performing a linear sum of the basis vectors and coefficients. To compare the reconstruction quality, we use as a Figure of Merit the normalised root mean squared error (NRMSE) of the relative reconstruction accuracy of the beams. In Figure 4.18, we compare the residuals of the beams that were produced by *Cassbeam* software (with the optimised antenna parameters - 'CASS'), SVD/PCA and Zernike (ZER) with the holography beams at the frequencies of 1.008, 1.404 and 1.908 GHz (using k_{lm} = 20 and k_ν = 50 for antenna 6). These techniques have reasonably captured the beams' spatial distribution, particularly at low frequencies with HT having the worst reconstruction error and also reconstructing with the largest number of coefficients. The *Cassbeam* and Zernike reconstructions have a similar average error of $\approx$ -10 dB. In general, for all the techniques the side-lobes and the nulls are poorly reconstructed. The SVD/PCA has the lowest residual errors (-15 dB on the average) hence can be considered the most robust method of reconstructing the beams.

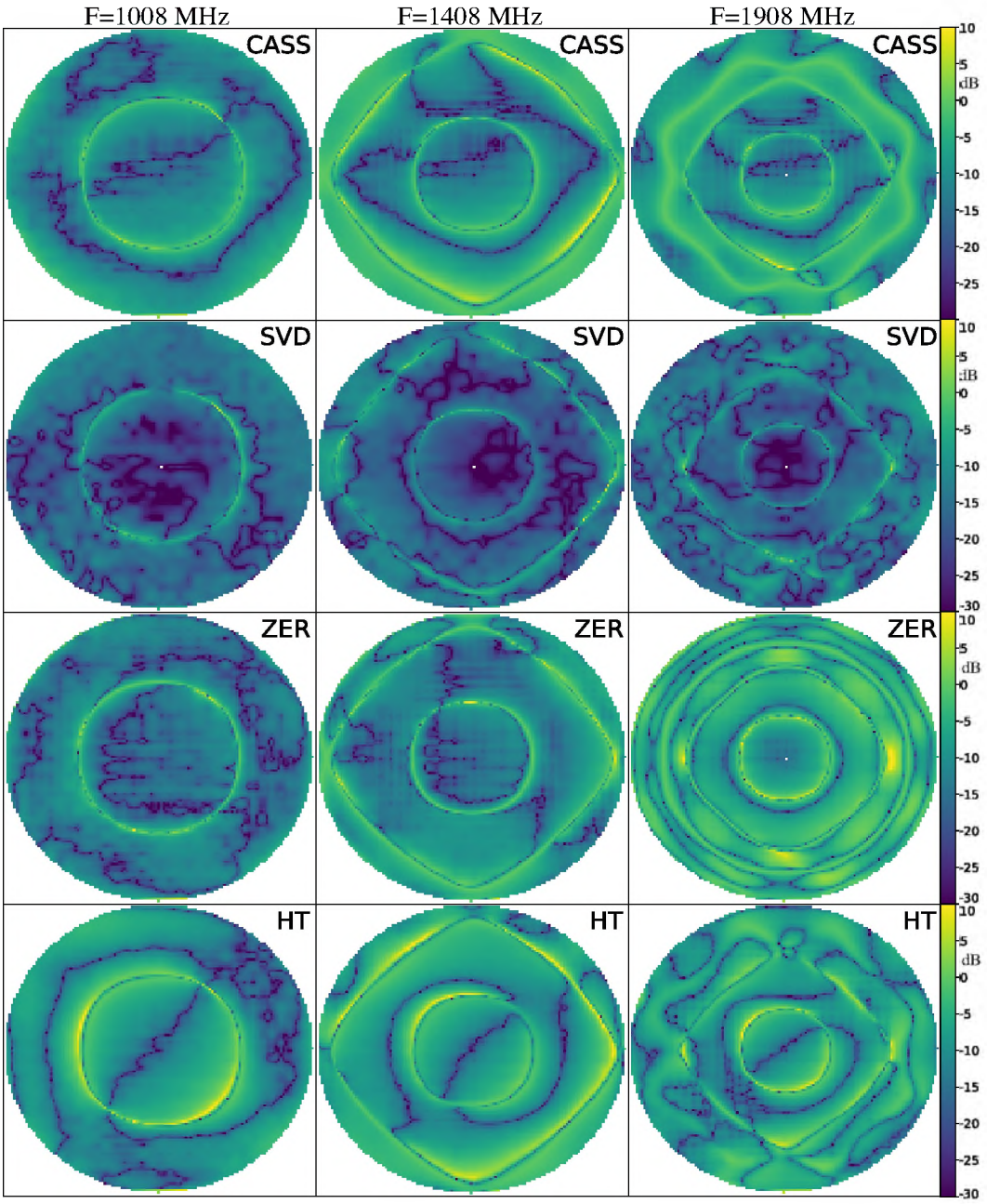

FIGURE 4.18: Plots of the reconstruction residuals (on the same linear colour transfer scale of -30 dB to 10 dB) at frequencies of 1.008, 1.404 and 1.908 GHz. The new *Cassbeam* (RR) is at the top row, followed by SVD (RR), then Zernike (LL) and HT (LL) in that order. The construction error for LL and RR beams are equivalent.

To check the overall reconstruction error compared with each technique compressibility, we perform beams reconstruction for $k_{lm} = 10,..,50$ and $k_\nu = 5,6,..,20$. Figure 4.19 presents the plots of NRMSE of the reconstructed beams as a function of the compression factor for SVD/PCA and Zernike decompositions. The graphs in Figure 4.19 are the graphs of the relative average error across the L-band beams for each technique. It is important

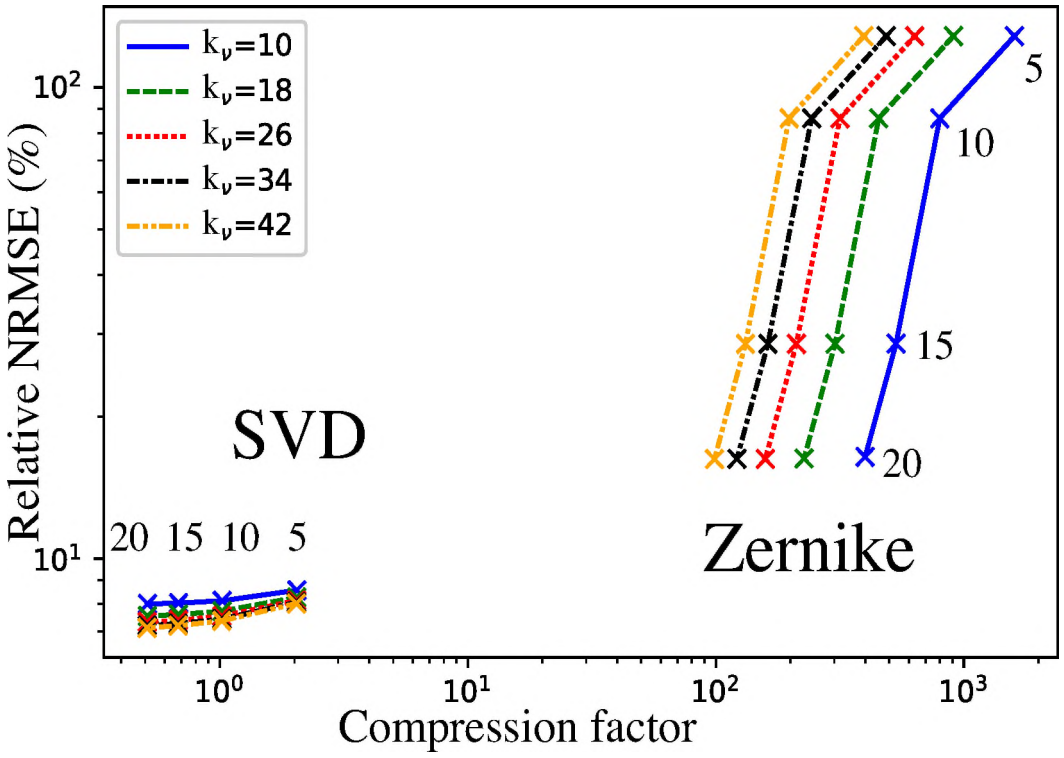

FIGURE 4.19: Graphs of the beams' NRMS error for SVD/PCA and Zernike decomposition for the entire L-band for k_{lm} = 10,..,50 and k_ν = 5,6,..,20.

to note that for the two techniques as k_ν varies from 42 to 10, it has an impact on the compression factor by approximately one order of magnitude but not on the relative NRMSE. Note that variations in k_{lm} impacted these techniques differently.

The SVD method produced the best NRMSE values but has lower values for corresponding compression factor because it requires the storage of the first k_{lm} eigenvectors (N^2_{pix} matrices). Using a large number of SVD coefficients can lead to over-fitting causing it to start fitting for noise (and measurement errors) in the holography measurement (like the horizontal, vertical and slanting stripes seen in the residual plots of *Cassbeam* and Zernike decomposition in Figure 4.18).

The Zernike technique has a larger value for compression factors compared with SVD but less accuracy across the whole band. This comes from the fact that selecting relevant Zernike modes across the whole L-band is difficult because using a small number of Zernike modes produces high NRMSE errors. Notwithstanding, Zernike decomposition produces low average residual errors when performed on each separate frequency slice of the holography cube. In other words, if one needs to work on a very small frequency

sub-band or in cases where compression is not an issue, Zernike beam models would be a very good choice.

4.6 Modelling the MeerKAT beams

After years of planning, designing and building of South Africa's SKA precursor, MeerKAT telescope is now commissioned and is moving towards its full operation. At the later stage of our VLA beams modelling (late 2017) we received holography data for MeerKAT antennas, so we applied the technique of PCA on the data to model the PB. In this section, the process of applying PCA on this data to generate PB models for the MeerKAT antennas will be described. (The terminologies, derivations, figures and notations of this section are taken from Asad *et al.* (2019)).

MeerKAT is an array of 64 offset Gregorian antennas located in the northern Karoo region of South Africa (as mention in Chapter 2). Forty-eight of these antennas are positioned within a one-kilometre diameter core and the 16 remaining antennas are positioned outside of this core, providing a maximum baseline of eight kilometres. Each antenna is planned to be equipped with four linearly polarised feeds to receive signals in the UHF (ultra high frequency), L, S and X[2] (respectively 580–1015, 0.9–0.167, 1.75–3.5 and 8–14.5 GHz (Booth *et al.*, 2009)) frequency bands, but we will focus only on the L-band.

4.6.1 MeerKAT antenna holography measurement

Holography observations of MeerKAT were done using (for example) antenna M009 as scanning antenna while using M011 as the tracking antenna. The source 3C273 was chosen as the observed source and the holography measurements were made using the observational parameters detailed in Table 4.2. The obtained data were processed with the KATHOLOG[3] software. KATHOLOG is a Python module developed and maintained by M. de Villiers (not yet open for public use at the time of writing) and used by SKA-SA to process holography data. We show the amplitude of the de-noised holography beams denoted with $\boldsymbol{E}^{\mathrm{h}}$ in Figure 4.20.

[2]Only the L-band feeds are commission at the time of writing. X-band feeds are not currently funded.
[3]https://github.com/ska-sa/katholog

Description	value(s)
Average elevation	39.82 degrees
Average azimuth	303.22 degrees
Number channels	1024
Bandwidth	856
Channel Width	0.835937 MHz
Observed sources	3C273
Total Time duration	0.5 h
Centre frequency	1.284 GHz
'Target' antennas	M009, M012, M015
Experimental ID	20181206-0010

TABLE 4.2: The main observational parameter which was used in performing the MeerKAT holography measurement

The beam diagonal terms ($\boldsymbol{E}^{\mathrm{h}}_{00}, \boldsymbol{E}^{\mathrm{h}}_{11}$) can be seen to be approximately circularly symmetric (having concentric regions of approximately equal amplitudes similar to that of the VLA) while the off-diagonal ones show a cloverleaf pattern.

We need to mention here that the *ripple characteristics* corresponding to the ones we observed in the VLA beams are very small (and asymmetric) in the case of the MeerKAT beams, which implies that the standing wave effects are very minimal on these antennas. Rather, the ripples originate as a result of the interference between the energies diffracted from behind the sub-reflector and that reflected from behind the main reflector (Morris, 1978; De Villiers, 2013; de Villiers, 2018). The very small standing wave effects can also be explained by the differences in antenna designs for VLA and MeerKAT (aperture blocking and non-blocking respectively, see Section 2.1).

4.6.2 MeerKAT spatial beams modelling

Equation (4.23) is applied (using KATHOLOG software) on the measured holography beams to get a de-noised model $\boldsymbol{E}^{\mathrm{h}}$. The top row of Figure 4.20 presents the two-dimensional plots of the amplitude $|\boldsymbol{E}^{\mathrm{h}}|$ beams for the orthogonal polarisations.

4.6.2.1 PCA model, $\mathbf{E}^{\mathrm{c}}$

The MeerKAT holography beams are complex beams and two beam cubes were produced for each antenna per polarisation - one for real and the other for imaginary data. Each

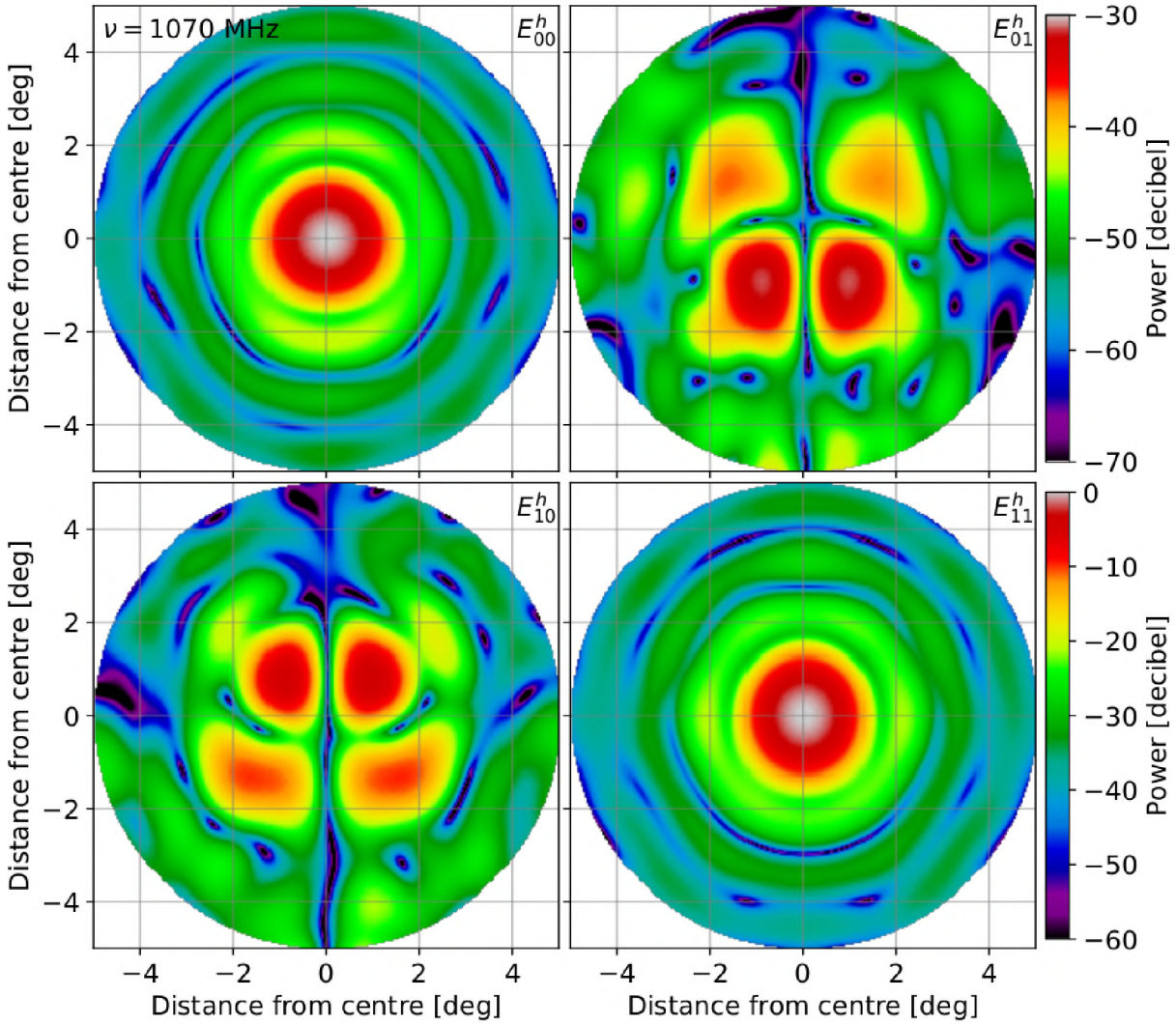

FIGURE 4.20: Two-dimension plots the holography beams for MeerKAT antenna M009 (average of M009, M012 and M015) over the FOV of 10 degrees at 1.07 GHz for the polarisation elements.

has beams from 0.9 to 1.67 GHz and a frequency resolution of 0.1 MHz. Each beam has 128×128 pixels. To generate a PCA model for the data, we take the 3D cube of the de-noised data $\boldsymbol{E}^{\mathrm{h}}$. Then, we flatten the beams for each frequency and stack them together to form 2D complex array, $\boldsymbol{E}^{\mathrm{h}}(\nu)$ with dimensions of $N_\nu \times N^2_{\mathrm{pix}}$, where N_ν is the number of *clean* beams (similar to the VLA beams), that is, the number of frequency beams that are not affected by RFI. The total number of pixels in each beam is N^2_{pix}. Performing SVD on $\boldsymbol{E}^{\mathrm{h}}(\nu)$ produces $N_\nu \times N^2_{\mathrm{pix}}$ complex matrix, $\boldsymbol{\mathcal{U}}$, $N_\nu \times N^2_{\mathrm{pix}}$ diagonal matrix of singular values, $\boldsymbol{\Upsilon}$ and $N^2_{\mathrm{pix}} \times N^2_{\mathrm{pix}}$ unitary matrix $\boldsymbol{\mathcal{V}}^{\mathrm{c}}$. $\boldsymbol{\mathcal{V}}^{\mathrm{c}}$ is made up of a collection of eigenvectors (the *eigenbeams*). The decomposition coefficients, $\boldsymbol{C}^{\mathrm{c}}$ are computed as $\boldsymbol{C}^{\mathrm{c}} = \boldsymbol{\mathcal{U}}\boldsymbol{\Upsilon}$. The L-band beams, $\boldsymbol{E}^{\mathrm{c}}$ are modelled using the *eigenbeams* and coefficients as

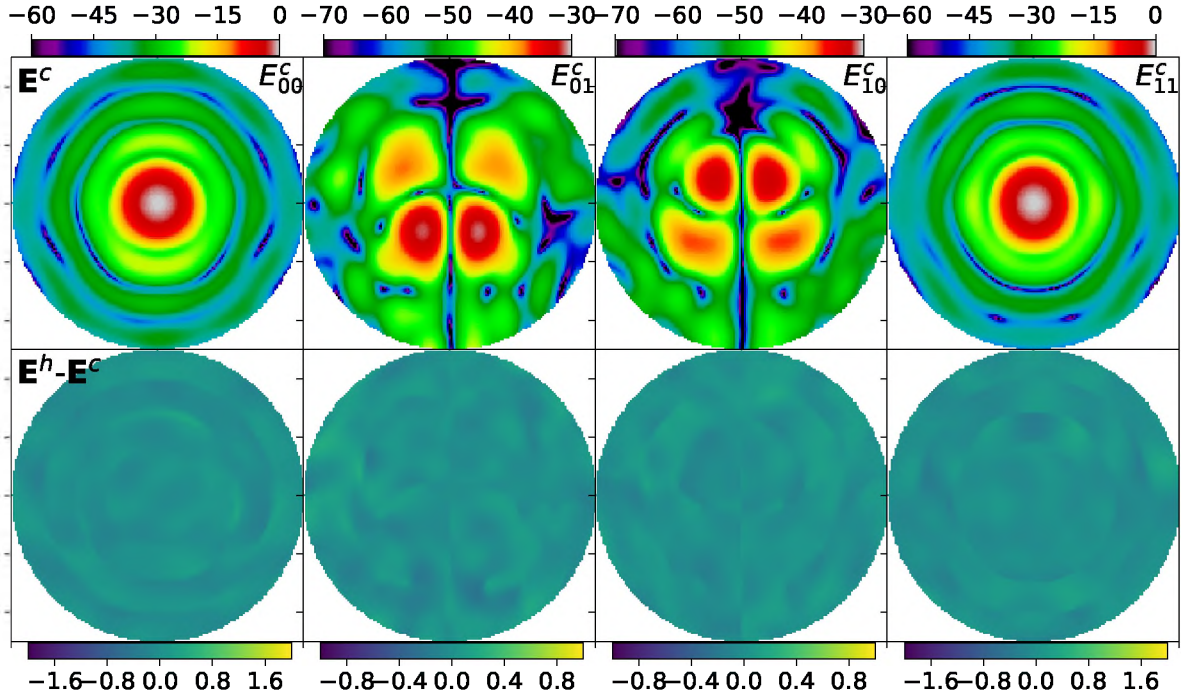

FIGURE 4.21: Two-dimension plots of the amplitude beam models $|\boldsymbol{E}^{\mathrm{c}}|$ generated by performing PCA on $\boldsymbol{E}^{\mathrm{h}}$ at 1.07 GHz while the bottom row shows the residuals of the models, $|\boldsymbol{E}^{\mathrm{h}}| - |\boldsymbol{E}^{\mathrm{c}}|$ (in dB). These were multiplied by a factor of one hundred for better visualisation.

$\boldsymbol{E}^{\mathrm{c}} = \boldsymbol{C}_0^{\mathrm{c}}\boldsymbol{\mathcal{V}}_0^{\mathrm{c}}$. $\boldsymbol{C}_0^{\mathrm{c}}$ is the matrix of the strongest coefficients while the corresponding PCs are $\boldsymbol{\mathcal{V}}_0^{\mathrm{c}}$.

In the top row of Figure 4.21, we present the reconstructed beams, $\boldsymbol{E}^{\mathrm{c}}$(at $\nu = 1.07$ GHz) the four polarisations and the residuals in the bottom row. These models are generated with the first fifteen strongest coefficients (i.e $N^{\mathrm{c}} = 15$) or PCs. This choice of $N^{\mathrm{c}} = 15$ was taken after considering the plots of the normalised root-mean-square error ($\sqrt{|\boldsymbol{E}^{\mathrm{h}}| - |\boldsymbol{E}^{\mathrm{c}}|}/|\boldsymbol{E}^{\mathrm{h}}|$)) first twenty eigenvalues N^{c} presented in Figure 4.23 and Figure 4.24. Figure 4.24 shows that the amplitude of the PCA decomposition *singular values* does not have any significant change after the 15^{th} component. This implies that 15 coefficients are sufficient to produce a meaningful reconstruction. Figure 4.23 shows that the RMSE of PCA reconstructed beams reduce (for $\boldsymbol{E}^{\mathrm{c}}_{00}$ and $\boldsymbol{E}^{\mathrm{c}}_{01}$) beams, with increasing number of coefficients as expected and further increasing the number coefficients after about fifteen strongest coefficients do not provide a significant improvement in the RMSE, (the graph flattens out). This figure also shows that 15 PCs produce RMS residual error level of approximately 5×10^{-5} for the off-diagonal and 10^{-3} for the diagonal Jones elements. We present the radial profile model $|\boldsymbol{E}^{\mathrm{c}}|(\nu = 1.07$ GHz) in Figure 4.22

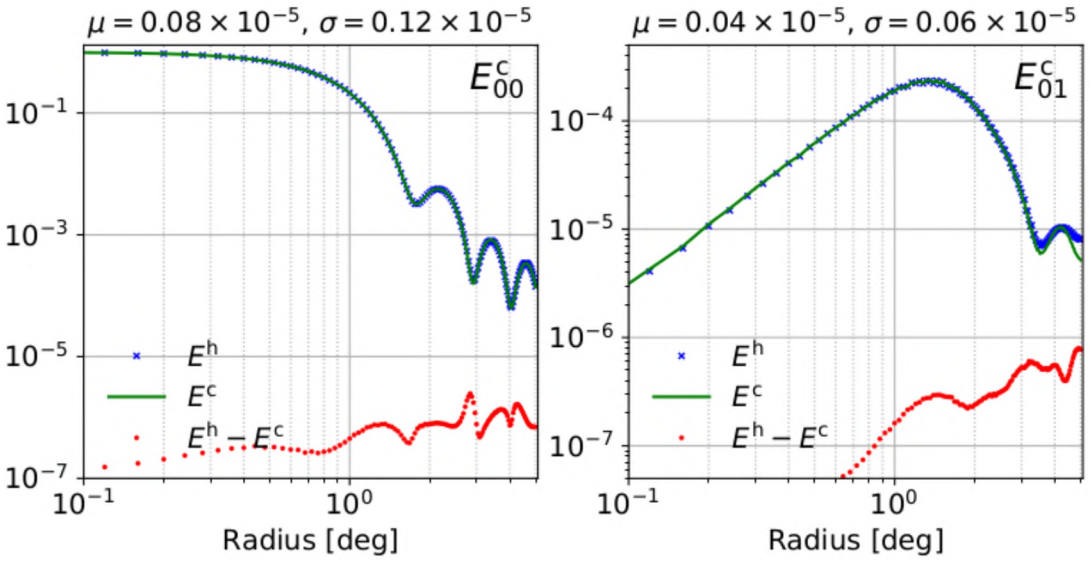

FIGURE 4.22: Plots of the radial profile of data, PCA model of E_{00} (first panel) and E_{01} (second panel) as indicated in the insert and residuals were averaged in concentric annuli rings. μ and σ represent the mean and the standard deviation of the residuals respectively. They have relative units, normalized to 1 at beam centre.

(red), which shows the holography data has been accurately reconstructed by PCA for all polarisations.

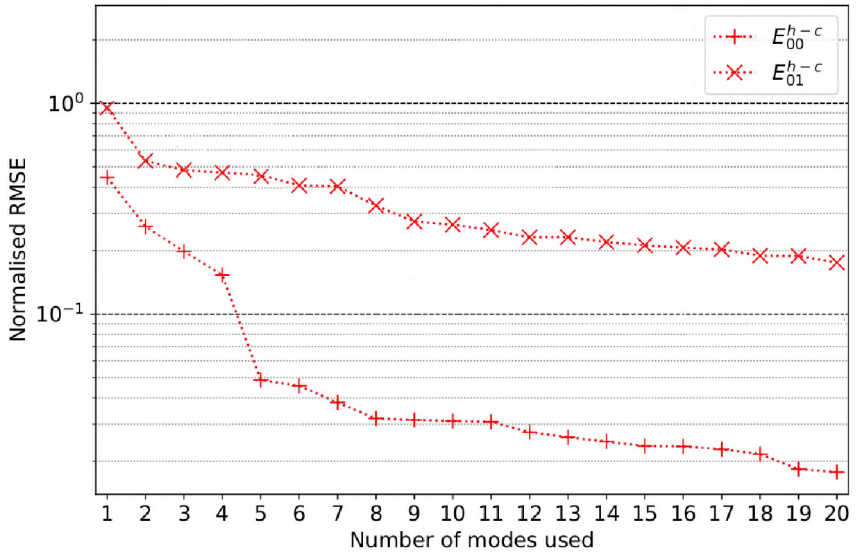

FIGURE 4.23: Plots of the NRMSE for the beams modelled with PCA, $|\boldsymbol{E}^c|$ using different number of PCs. $|\boldsymbol{E}^{h-c}|$ is calculated by comparing with $|\boldsymbol{E}^h|$ and $|\boldsymbol{E}^c|$ ie. $|\boldsymbol{E}^{h-c}| = |\boldsymbol{E}^h| - |\boldsymbol{E}^c|$

Another advantage of representing the beams with an orthogonal basis is that it allows one to reconstruct them with fewer parameters or information. The PCA model, $\boldsymbol{E}^c$

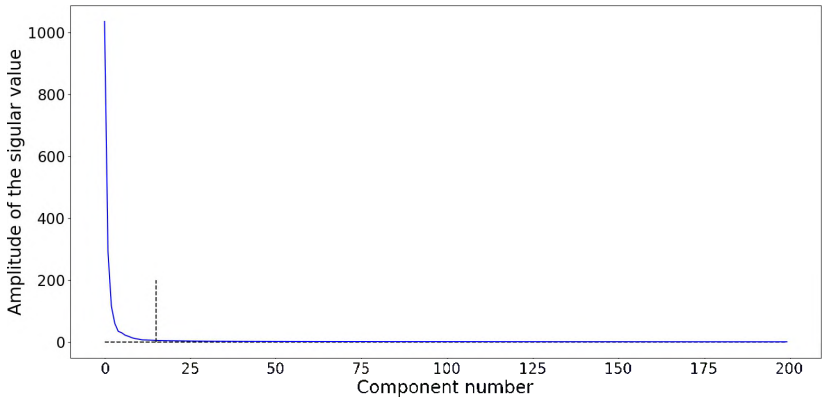

FIGURE 4.24: Plot of the first 200 singular value components of the PCA decomposition. The vertical dash line indicates the position of the 15^{th} singular value for E^{c}_{00}.

requires $4(N^2_{\text{pix}} + N_\nu)$ number of parameters to reconstruct beams for all polarisations. Also, the behaviour of the coefficients with frequency can be modelled with $N^{\text{c}}_\nu < N_\nu$ number of parameters (as is presented in Section 4.6.2.2) which further compresses the number of parameters that are needed to produce $\boldsymbol{E}^{\text{c}}$.

4.6.2.2 Modelling the spectral behaviour of the coefficients

We further compress the spectral behaviour of the strongest decomposition coefficient using DCT decomposition.

RFIs also affected the MeerKAT holography measurements, as it did for that of the VLA beams in the L-band frequencies. Hence, we have bad, completely corrupted or no data at some frequencies. The beams for the affected frequencies were excluded from our analysis. Figure 4.25 presents plots of the PCA coefficients as a function of frequency - real (first two rows) and imaginary (last two rows). The dotted lines represent the PCA decomposition coefficients as obtained with different bases while the interpolated[4] spectral models are represented by the solid lines. To generate models for the RFI affected frequency bands, we reconstruct their (corresponding) coefficients by interpolating them from the surrounding RFI-unaffected channels. We interpolate and compress with DCT exactly as the case of the VLA.

[4]The Python *numpy.interp* was used to perform the interpolation.

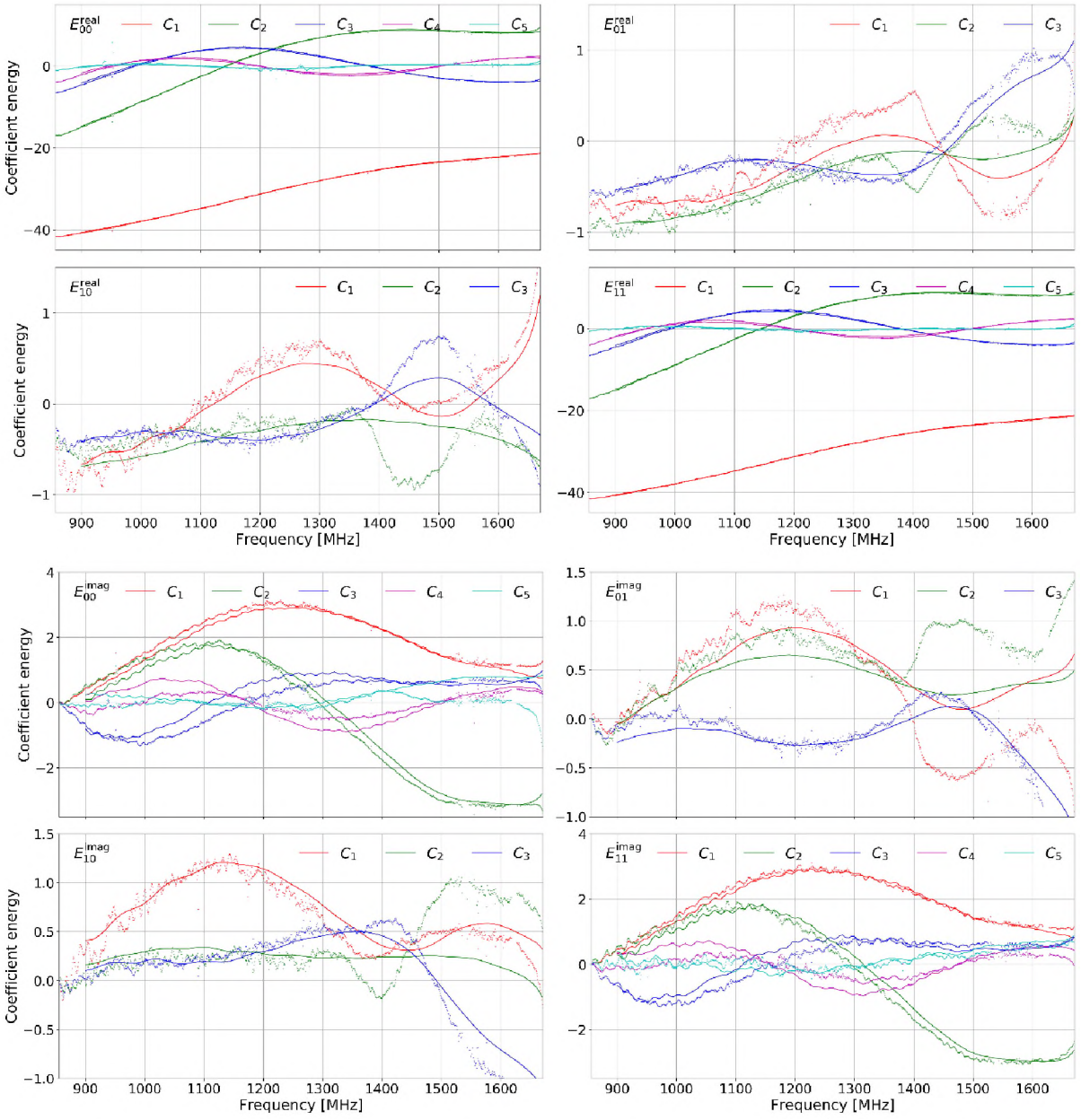

FIGURE 4.25: Plots of the PCA coefficients for the polarisations elements. The first two rows present the graphs of the real (E_{00} (first row, left panel) and E_{01} (first row, right panel) and E_{10} (second row, left panel) and E_{11} (second row, right panel)). The last two rows present the graphs of the imaginary coefficients of the polarisation elements. The dotted and solid lines respectively represent data and models. PCA decomposition coefficients, C_i; i =1, 2, 3...

4.7 Summary

We started this chapter with a brief introduction to primary beam (PB) modelling and why a good model for the PB is required for radio astronomy. Having accurate PB models are necessary to harness the full potentials of modern telescopes by accounting

for their instrumental effects during calibration and imaging. There are different methods for modelling the PB of radio telescopes that can be deduced from analytical derivation, electromagnetic simulation or the observed data. On one hand, use of a $\cos^3$ model for the PB is an example of analytic beam model while, on the other hand, a PB derived from *Cassbeam* software is a representative instance of the simulation-based method which accounts for the underlying physical effects (e.g. constraints, radiation properties etc.) of the antenna.

A third way to model the PB is to rely on characteristic basis function patterns constrained by data. These functions allow for efficient decomposition of PB derived from data. A modified version of the *Cassbeam* software by Jagannathan *et al.* (2018) has also been developed to model the PB (of Cassegrain antennas) by optimising the physical input parameters of this software in order to produce models that match measured data. In our approach, we modelled the PB using two methods: PCA and Zernike decompositions from measured data. First, a brief review of (SVD) was presented. Then a brief review of PCA and Zernike decompositions were also presented. PCA is one of many linear algebra tools used to find important features in the dataset, rank them by importance and also enable dimensionality reduction. We also outlined how one can perform PCA decomposition using SVD. PCA produces eigenvalues and eigenvectors sorted in order of decreasing power (or energy). Hence, one can easily get a low-rank approximation to the desired accuracy. Dimensionality reduction of the data can be produced by taking the first Nth dominant coefficients that approximate the data, where N is much less than the number of sampled data.

Zernike decomposition uses a 2D orthogonal basis built from ZP to capture various distortions and features in a signal. This basis by design can easily be used to decompose modes which are defined on circular apertures and allow for the analysis of the energy distribution over the various models that can be interpreted physically as optical aberrations. These two techniques were used to model the VLA and MeerKAT antenna PB beams. A comparison of the spatial distribution and the spectral behaviour of the EM and holography beams were also presented. These showed that the EM beams captured the average spatial power distribution of the beams particularly in the main-lobe but not in the nulls and side-lobes. The EM beams do not model the *ripple characteristics* which are present in the holography beams. These imply that more accurate PB models are required for the VLA antennas which will capture these features of the holography beams. A modified version of the *Cassebeam* software software was developed by Jagannathan *et al.* (2018), which attempts to absorb the *ripple characteristics* found in the

VLA holography beams in the optimised input parameters, such that the software will produce beam models that match the holography data.

We applied PCA to the holography data across the whole L-band to model the PBs and obtained decomposition coefficients while per frequency beams decomposition was done on the data using ZP decomposition. The PCA and ZP decompositions were used to model the PBs of the VLA telescopes from holography while only PCA was used to model the PBs of MeerKAT antennas. The number of decomposition coefficients used to model the beams are much less than the number of measured samples. This brought about data compression because a reduced amount of information is required to reconstruct the whole data. Some sub-band of the data was affected by RFI and was excluded from our analysis, so the coefficients (that are used for modelling the PBs) were further simultaneously compressed along the frequency axis (using DCT) and missing frequencies interpolated with Papoulis-Gerchberg algorithm for the DCT and Zernike coefficients. The MeerKAT holography data were also modelled using PCA and the decomposition coefficients were modelled and further compressed using DCT. The PCA reconstructed beams produced lower NRMSE[5] (below -15 dB on the average) compared to Zernike decomposition using the same number of coefficients.

Using the number of data points needed to store the data and the model as a measure of compression (i.e applying Equation (4.24a)), we computed the compressibility factor for the PCA technique. The PCA beam models generated (with 20 and 15 coefficients) for the VLA and MeerKAT antennas used only approximately 2% and 0.25% respectively of number of data point used by the holography data. This gives approximately 98% compression for the VLA antennas and 99.85% for the MeerKAT antennas.

Although the modified *Cassbeam* software can be used to produce approximate beam models for Cassegrain antennas through the optimised input parameters, this approach cannot easily be extended to model PBs for other antenna configurations, Gregorian antenna designs etc. We have shown how to model the PB of any type of radio telescope from holography measured data with PCA and Zernike decomposition using a small number of coefficients (much less than the number of sampled data).

If one has any form of (holography) data that is sampled in frequency which one wishes to model, the procedure outlined in this chapter can be used to model and compress the data. If there are missing data (or data that are affected by RFI or corrupted data), the recipes outlined here can also be used to interpolate and approximate them.

[5]NRMSE - normalised root mean squared errors

In this chapter, we have modelled the PB of radio antenna from holography data and we saw that the standing waves that exist in the VLA antenna, introduce ripple characteristics in the graph of its PBs FWHM and offset against frequency. We will investigate in Chapter 5 the impact of not accounting for these ripple effects in calibration and imaging in radio astronomy.

Chapter 5

Effects of beam spectral ripple on calibration and imaging

5.1 Introduction

In Chapter 2, we saw that the purpose of calibration is to recover the signal that has been corrupted by instrumental and propagation effects. Accurately accounting for the radio antenna PB is one of the important steps that must be taken during modern calibration of radio interferometric data from modern instruments. The framework that enables the implementation of the so-called direction-dependent (DD) calibration (Smirnov, 2011b,c,a) was presented in Chapter 2 using the RIME formalism. We also saw that the PB of a radio telescope is one of the DDEs that need to be accounted for during calibration. We have described in Chapter 4 two methods of modelling the PB of any radio telescope (in particular, the VLA and MeerKAT antenna arrays) using holography measurements with signal representation methods, such as PCA and Zernike decompositions. VLA and MeerKAT arrays have antennas of Cassegrain and Gregorian designs, respectively. One of the disadvantages of the parabolic reflector antenna telescope is the presence of standing-waves (see Chapter 3). This results from resonance effect that occur when signals bounce between the feed and sub-reflector resulting in constructive interferences at some frequencies and destructive interferences at others. Most existing PB models used in calibration do not incorporate this DDE introduced by the standing wave in them. The use of an inaccurate PB model during calibration results in incorrect estimation of source fluxes, creates artefacts in images and increases the noise level.

In this chapter, we will study the implications of not accounting (or accounting) for PBs' spectral distortion on the quality of calibration, with the specific example of the VLA L-band PBs. From the analysis of the holography data of the VLA primary beams (see Chapter 3), we saw that the standing wave in Cassegrain antennas induced a spectral distortion on the beams size and centre offset (graphs), which we called the *ripple characteristics*. Here, we present a model of the PB which incorporates the ripple characteristics. We simulated visibilities that include the effect of the ripple and show the cost of ignoring the ripple during calibration of continuum data. We will focus on the *ripple characteristics* of the beams. In all our experiments, we used simulated beams (i.e. beam models produced with *Cassbeam* not that which is derived from holography) to restrict the investigation to only one parameter (the ripple effect). Section 5.2 will present a brief discussion of the ripple in the holography beams and how we simulated beams with and without the associated spectral ripple features in Section 5.3. Section 5.4 deals with the continuum simulations using the latter, while Section 5.6 presents the effects of the ripple on the analysis of spectral line data. Section 5.7 presents the summary.

5.2 Holography and simulated beams

We mentioned earlier that the *Cassbeam* software has been used to produce PB models (EM beams) for the VLA antennas and presented some comparisons between the properties of the EM and the holography beams in chapter 4. Chapter 3 also discussed in details the characteristic spatial and spectral effects of the holography beams by the standing wave in these antennas. We will recall these comparisons to clarify the notations that will be used in this chapter.

5.2.1 Holography beams and ripples properties

Looking at the graphs of FWHM (beam size) and (normalised) centre offset (Normalised offset = $\frac{\text{offset}}{\text{FWHM}}$) vs frequency, for the *Cassbeam* and the holography beams presented in Figure 5.1, we identified three characteristic features:

1. Both the holography beams' FWHM and centre offset graphs are inversely proportional to frequency (and so are those of the *Cassbeam*). These we called the first-order characteristics; the graph plotted with blue colour - $\text{FWHM}_{\text{Cassbeam}}$ and $\text{offset}_{\text{Cassbeam}}$.

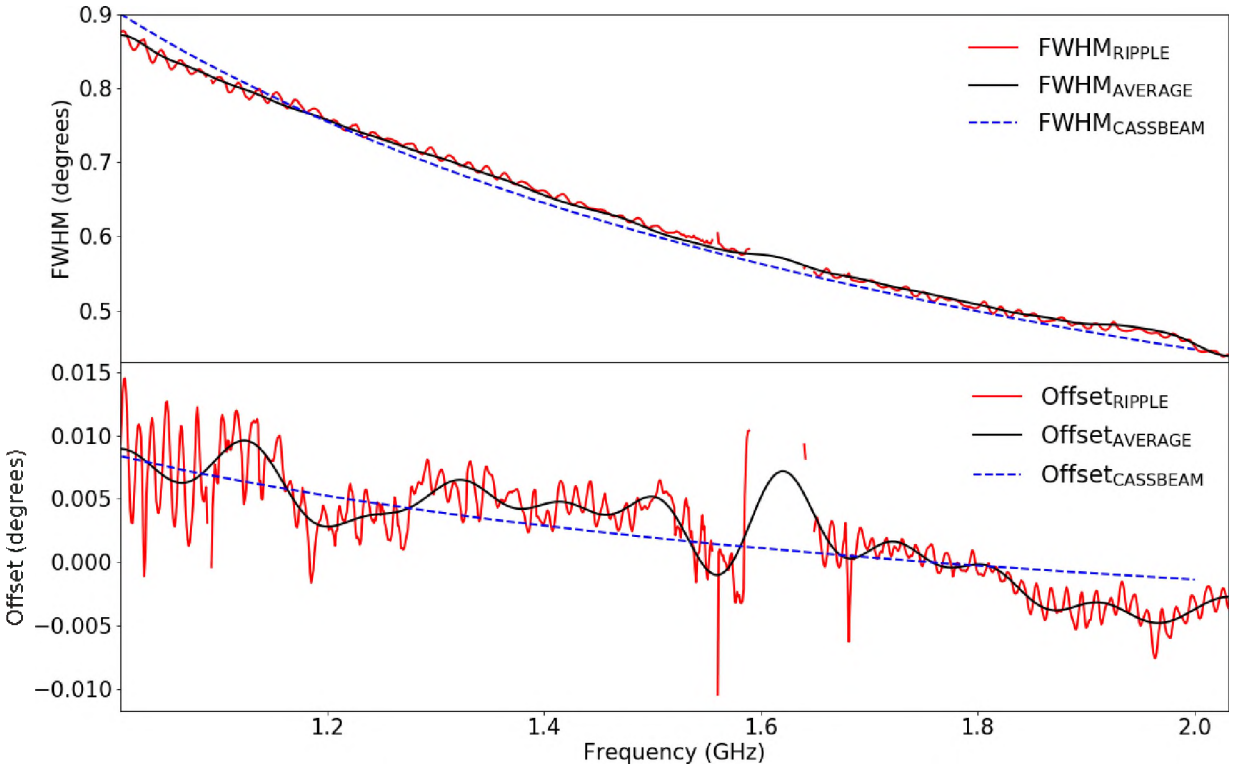

FIGURE 5.1: The graphs of the FWHM (top panel) and normalised offset (bottom panel) of the holography and simulated (*Cassbeam*) beams as a function of frequency. The x-offset was normalised by dividing it with the beams' size at the corresponding frequency.

2. The holography beams also have a fast oscillating trend - ripple behaviour with ν. This we called the third-order characteristic (the graph plotted with red colour in Figure 5.1 - $\text{FWHM}_{\text{RIPPLE}}$ and $\text{offset}_{\text{RIPPLE}}$); and

3. One can compute and plot a mean graph of these ripple distributions with ν for the FWHM (and centre offset) of the holography beams to obtain an average characteristics behaviour, which we called the second-order characteristic; (the graph plotted with black colour in Figure 5.1 - $\text{FWHM}_{\text{AVERAGE}}$ and $\text{offset}_{\text{AVERAGE}}$).

From Figure 5.1, we observe three characteristic features of the PB - the first, second and third-order characteristics.

5.3 Generating beams for the experiments

To proceed with the experiments presented in this section, we produced three PB cubes, each having one of the characteristic features we identified above. Each of these files

contain a beam at every 1 MHz step in frequency from 1.008 GHz to 2.008 GHz i.e. 1000 channels.

We generated three categories of PB models that were used in the experiments.

1. The first, are some theoretically simulated beams (with the *Cassbeam* software), whose FWHM and centre offset have first-order characteristics behaviour, called the *normal Cassbeam*, B_{CassN}.

2. Then, using B_{CassN}, we scaled and shifted these beams so that the FWHM and centre offset will have second-order characteristic behaviour with ν which we called the *average beams*, B_{CassA}.

3. Finally, using B_{CassN} we scaled and shifted these beams so that the FWHM and centre offset will have the third-order characteristics behaviour with ν which we termed the *ripple beam*, B_{CassR}.

These three beams will be used in the next section to perform continuum data simulation experiments.

5.4 Continuum simulations

Radio emissions received over a wide range of frequencies from celestial sources are generally referred to as radio continuum emissions. The brightness temperature (or intensity) of the sources basically varies slowly and smoothly with frequency. Radio continuum observations are observations done on *continuum sources* and the data obtained by so doing are called *continuum data*, while carrying out calibration on such data is referred to as *continuum calibration*.

5.4.1 Simulation details

In this section, we carry out a number of numerical experiments to investigate the effect of a PB model that incorporates distortions (such as the spectral ripples of the VLA beams) during calibration. Again, we use simulated data in the experiments so as to isolate and test only the contributions of the ripple in the VLA antennas PBs.

In these simulations, we estimate the effect of various factors involved in the process of self-calibration on the image dynamic-range and flux density recovery for the sources. These factors include the PB model, the solution intervals (i.e. time and/or frequency intervals between solution) and the spatial distribution of the sources in the sky with respect to extent of the beam. We start with a single-source simulation, then a two-source simulation and finally, random distributed multiple sources simulations. The general layout of these simulations are as follows:

- Generate a sky model with point source(s) with a flat spectrum. Simulate an empty measurement set (MS) using the *simms* software tool. Then, using B_{CassR} beam model, we simulate an observation using the sky model into a MS using the *MeqTrees* software (Noordam and Smirnov, 2010). The simulation parameters are listed in Table 5.1.

Description	values
Observation time	4 hours
Integration time	60 seconds
Field of view (FOV)	3 degrees (diameter)
Bandwidth	110 MHz (1.008 - 1.118 GHz)
SEFD (system equivalent flux density)	300 Jy

TABLE 5.1: The main parameters that were used to simulate the observations

- Then, DD Statistically Efficient and Fast Calibration (StEFCal) (Salvini and Wijnholds, 2014) calibration algorithm is applied to the observation with the *MeqTrees* software using the parameters listed in Table 5.2.

Description	values
PB model	B_{CassR}, B_{CassA} and B_{CassN}
gain solution interval in frequency	1, 9 and 18 MHz
Gain solution interval in time	1 minute
Bandwidth	110 MHz (1.008 - 1.118 GHz)

TABLE 5.2: The main parameters that are used to calibrate the observations

- We then image the calibrated visibilities (i.e. corrected data) using *WSCLEAN* (Offringa *et al.*, 2014) software using the parameters in Table 5.3.
- Run source finder and source extraction using *PyBDSF* (Mohan and Rafferty, 2015) software and analyse flux density of extracted sources.

Description	values
Software in use	*WSCLEAN*
Robust parameter	-2
cleaning threshold	0.5 × the noise in the image

TABLE 5.3: The main parameters that are used for imaging

The choice of the sub-band 1.008 - 1.118 MHz was inspired by the fact that it is the region where the ripple is the strongest in the L-band.

We present the results of these experiments and discuss them with plots of the calibration gain amplitude (versus channel bins for different gain solution intervals in frequency - 1, 9, 18 MHz), error on recovered flux densities, residual images after calibration and PB gain that is applied to (i.e. at the position of) each source (and its error) in each case.

5.4.2 Single-source simulations

To commence our investigation, we started with a single point source simulation. This case is of interest when calibration is carried out on a fairly bright source in the field (slightly offset), whose PSF side-lobes need to be correctly subtracted. It also offers us insight on how calibration of multiple sources might be affected in the absence of accurate PB models.

To do this, we simulate a single point-source at a distance of about 15 arcmin from the field centre and follow the procedure outlined above (in Section 5.4.1). We choose a slightly offset location within the main lobe. A single 2 Jy (intrinsic) flux density point source with a flat spectrum was simulated into an MS using the ripple beams, B_{CassR}. Some noise was also added based on the SEFD (system equivalent flux density) of the array. Then, applied DI StEFCal algorithm (Salvini and Wijnholds, 2014) (implemented in the *MeqTrees* software (Noordam and Smirnov, 2010)) on the data. These simulated data (using this sky model) were calibrated with the *ripple*, (B_{CassR}), *normal Cassbeam* (B_{CassN}) and *average* (B_{CassA}) beams using the frequency gain solution intervals of 1, 9, 18 MHz and maintained the gain solution time interval at 1 minute for all calibrations. The source had an apparent flux density of approximately 1.36 Jy.

Figure 5.2 presents plots of the gain amplitude against frequency channels. The gain amplitudes for RR on and LR off-diagonal elements are presented in the left and right

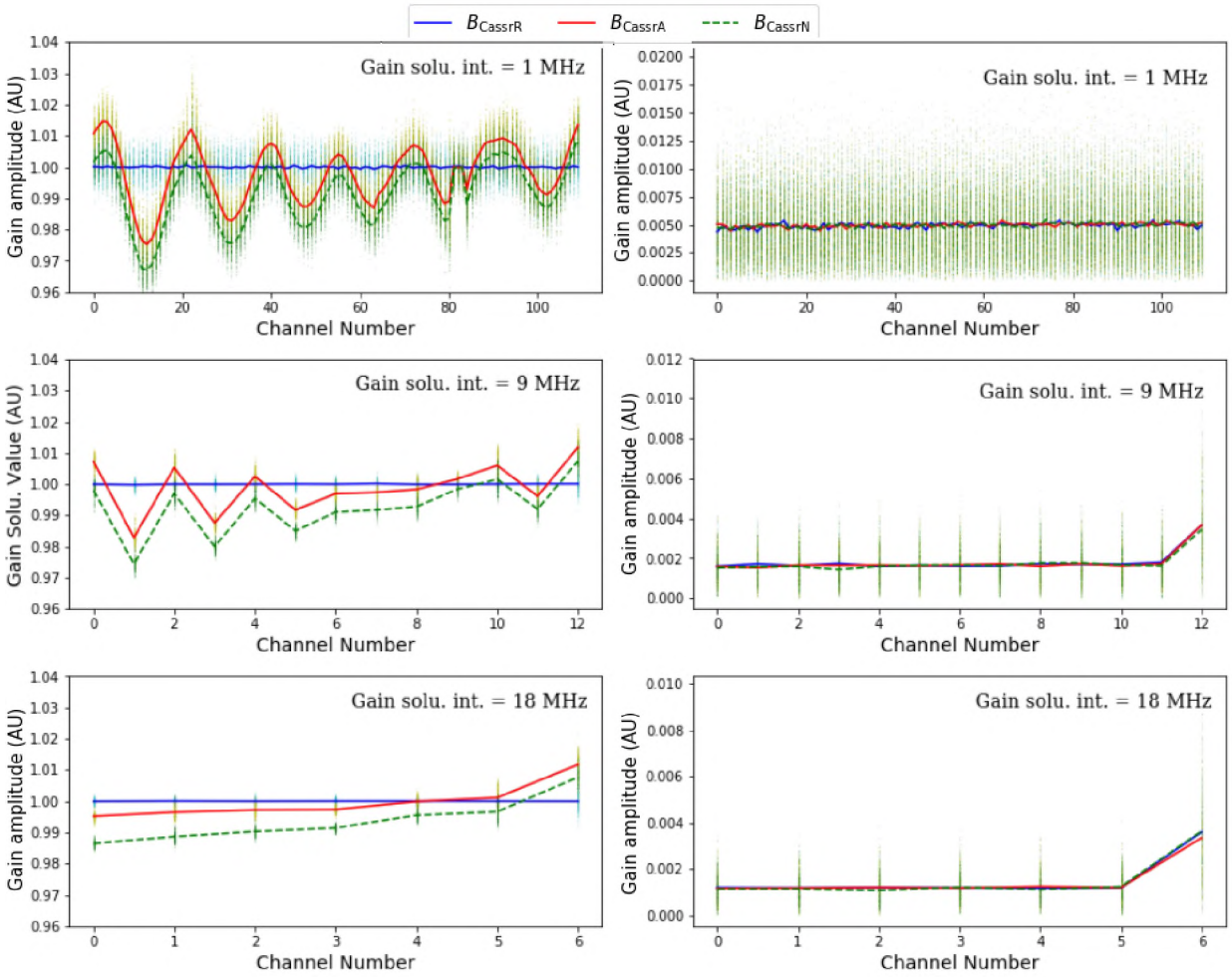

FIGURE 5.2: RR (first column) and LR (second column) gain solution plots for the single case study - for the calibration gain solution intervals in frequency of 1, 9 and 18 MHz. Shading is between the gains of the first and last frequencies, while the solid line represent the gain of the centre frequency in each case.

columns, respectively. The first row presents amplitude gain plots for calibration done with solution interval in frequency of 1 MHz, second row 9 MHz and the third row 18 MHz. Recall that in Section 4.5.3.2, we mentioned that the ripples in the graphs of the holography beams FWHM vs frequency have a frequency of approximately 17 MHz (Popping and Braun, 2008; Jagannathan *et al.*, 2018). Our choice of solution intervals is meant to expose how the calibration solution is affected by averaging over frequency. Note that we have only shown the graphs of RR on and LR off-diagonal elements, the other two are similar.

From these graphs we see that:

1. The calibration which was done with ripple (correct) beams produced the best calibration results and have a gain solution amplitude value of 1 (for diagonal terms)

which is expected in the absence of any other signal corruption effects.

2. There are manifestation of ripples in the gain amplitude plots of calibration done with the *normal Cassbeam* and *average* beam models, having values that vary around 1 (top left panel for diagonal terms).

3. As the solution interval (in frequency) increases, the gain values become closer to 1 for the calibration done with the wrong beams - *normal Cassbeam* and *average* beam models. The gain amplitude plots for these calibrations become almost flat (i.e. have no ripple) when the solution interval reaches 18 MHz, which is approximately the frequency of the ripple introduced by the standing wave.

4. The value of the gain amplitudes for the leakage term LR (right panel) are very small (approximately 0.0028).

5. Calibration done with the *average* beams produced a better overall result than that done with *normal Cassbeam* beams in terms of the calibration gain amplitude value. The calibration gains absorb the variations of the beam that are not accounted for by the incorrect models.

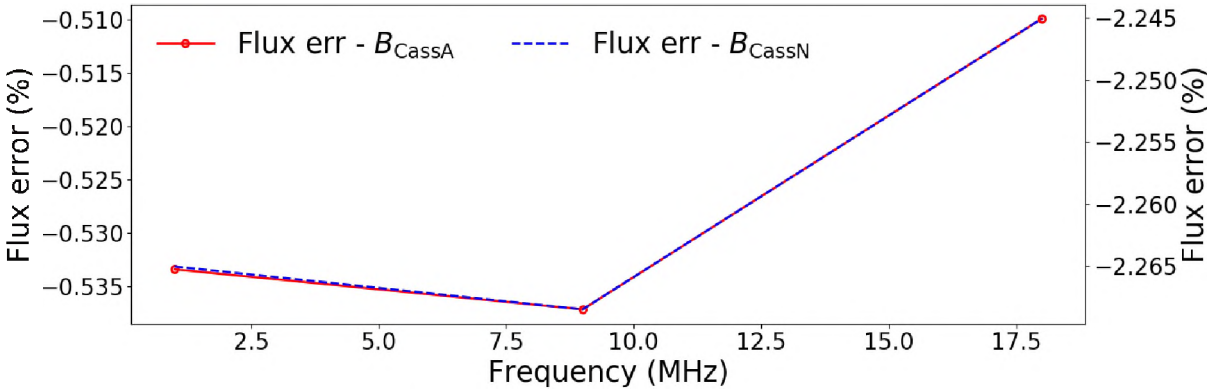

FIGURE 5.3: Graphs of the errors in recovered flux densities against gain solution interval in frequency - for calibration done with B_{CassA} and B_{CassN} beams, plotted on primary and secondary y-axis time respectively (for single-source simulations).

6. The errors in recovered flux densities for all the calibrations done with both beams (B_{CassA} and B_{CassN}) had overestimation of flux densities and had error value less than 2.5% for this source. The calibration done using gain solution interval of 9 MHz had worse flux density recovery than that done using 1 MHz but that which was done using gain solution interval in frequency of 18 MHz was better by 0.02%. Figure 5.3 shows this graphically. The errors in recovered flux densities of calibration done using B_{CassA} beam are plotted on the primary y-axis while that done using

B_{CassN} beam are plotted on the secondary y-axis. B_{CassA} beams had over all better flux density recovery than B_{CassN} for all calibrations. The slight increase in flux error at the intermediate solution interval (compared to the shortest interval) is a little puzzling. This could be due to how the solution interval samples the period of the frequency ripple.

5.4.3 Two-source simulations

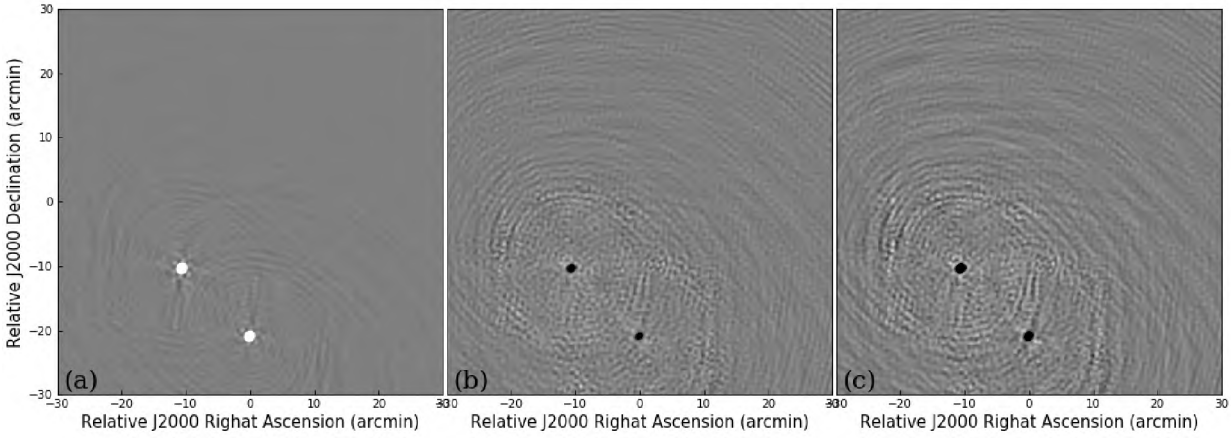

FIGURE 5.4: Two-sources simulation: (a) Stokes I image of the apparent sky = PB × intrinsic sky; residual images after calibration done with: (b) B_{CassA} and (c) B_{CassN} beams - using the solution intervals of 1 MHz and 1 minute in frequency and time respectively.

Continuing with our experiments, we perform a two-source simulation. Using the same configuration as in the single-source simulation case, we carried out a two-point source simulation (each of the sources having 2 Jy intrinsic flux density) following the procedure outlined in Section 5.4.1. One source was maintained at the same position (of 15 arcmins from the phase centre) as case 1, while the other was placed at a distance of approximately 21 arcmin from the phase centre but still within the main lobe.

The Stokes I images of the residual visibilities for the respective calibrations are shown in Figure 5.4 along with the apparent sky. Note that all calibrations were done with the same sky model that we used to simulate the observations.

The residual image for calibration done using B_{CassR} beam was not shown because is it completely noise-like (i.e. thermal noise limited). The second and third panels of Figure 5.4 show that the calibration done using B_{CassA} (second panel) and B_{CassN} (third panel) beam models overestimated the apparent fluxes of these sources. The PSF side lobes of these sources are poorly modelled and hence are not completely subtracted. So, in a case

where one has some faint sources that are within the side lobes, these sources could be hidden in the calibration noise.

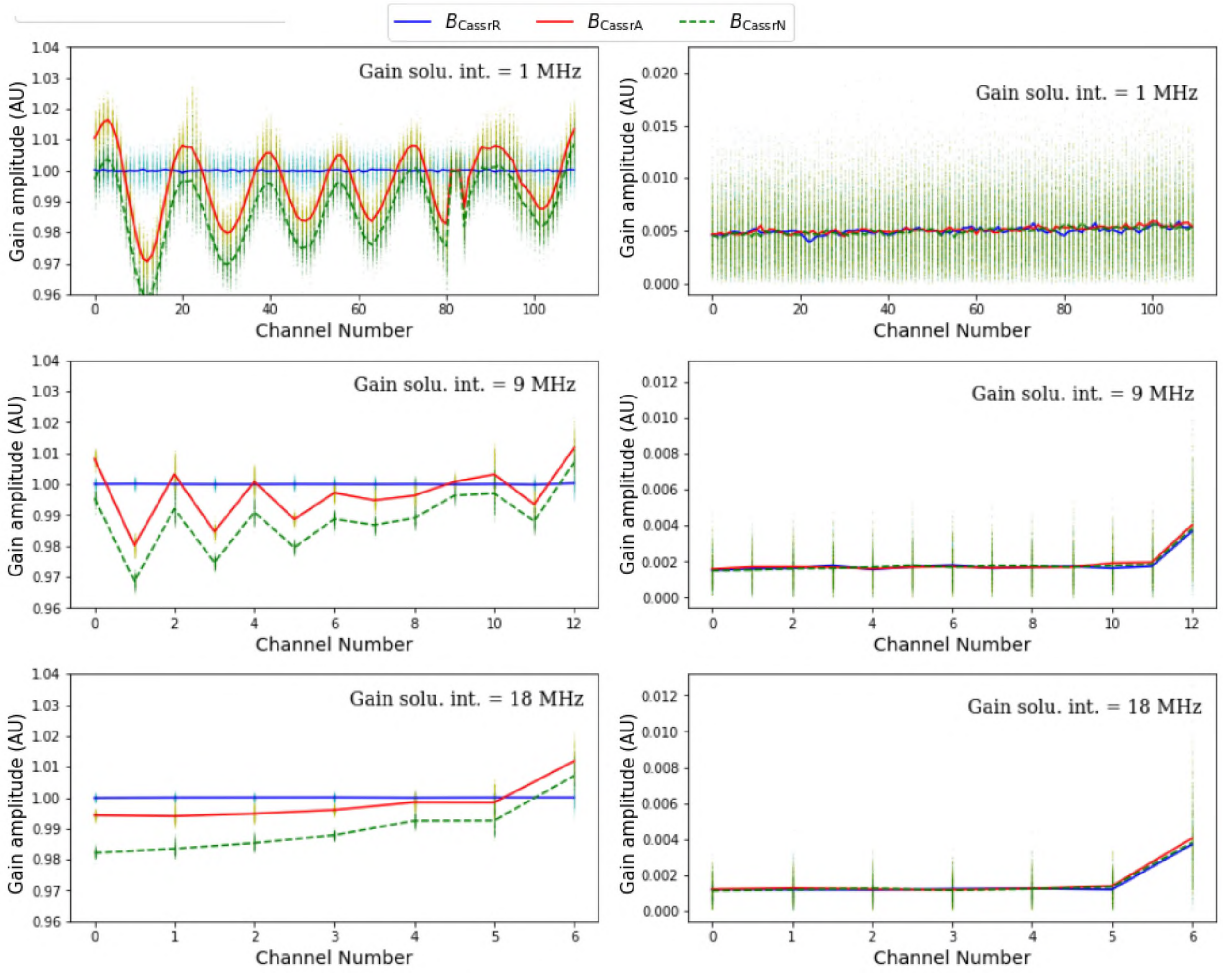

FIGURE 5.5: RR (first column) and LR (second column) gain solution plots for the case of two-sources simulation - for the gain solution intervals in the frequencies of 1, 9 and 18 MHz (for antenna 5). Shading is between the gains of the first and last frequencies while the solid line represent the gain of the centre frequency in each case.

The gains in this case cannot fully absorb the beam mismatch in the source model, which is why we see errors in the case of these wrong beams. The gain plots for this two-sources simulation case are presented in Figure 5.5. The gain plots for all antennas are very similar, so we show only the gain plots for Antenna 5. Again, the effect of the beam spectral ripple is apparent, especially in the RR amplitude gain plots. The amplitude of the ripples also reduces as the solution interval increases in frequency. The graphs in Figure 5.5 are very similar to that of Figure 5.4, which suggest that the DDEs (in particular the E-Jones) that affected the brightest source (at 15 arcmins from the phase centre) dominated the calibration. The brightest source also had an intrinsic and apparent flux densities of 2 Jy and approximately 1.36 Jy, respectively while the other (positioned

at approximately 21 arcmins) had an intrinsic and apparent flux densities of 2 Jy and 0.83 Jy, respectively.

The errors in the recovered flux densities of the bright source are the same as the case of single source (see the top panel of Figure 5.6). The errors in the recovered flux densities of the offset source did not improve significantly with increasing value of the calibration gain solution interval in frequency for the calibration done with both beams (see the bottom panel of Figure 5.6). The errors in recovered flux densities of calibration done using B_{CassA} beam is plotted on the primary y-axis while that for B_{CassN} beam is plotted on the secondary y-axis. B_{CassA} beam recovered flux density better than B_{CassN} for all calibrations, the same as that of the single-source case. The artefacts around the source did not also improve significantly with increasing frequency gain solution interval. The slight increase in flux error at the intermediate solution interval (compared to the shortest interval in Figure 5.6) is a little puzzling. This could be due to how the solution interval samples the period of the frequency ripple

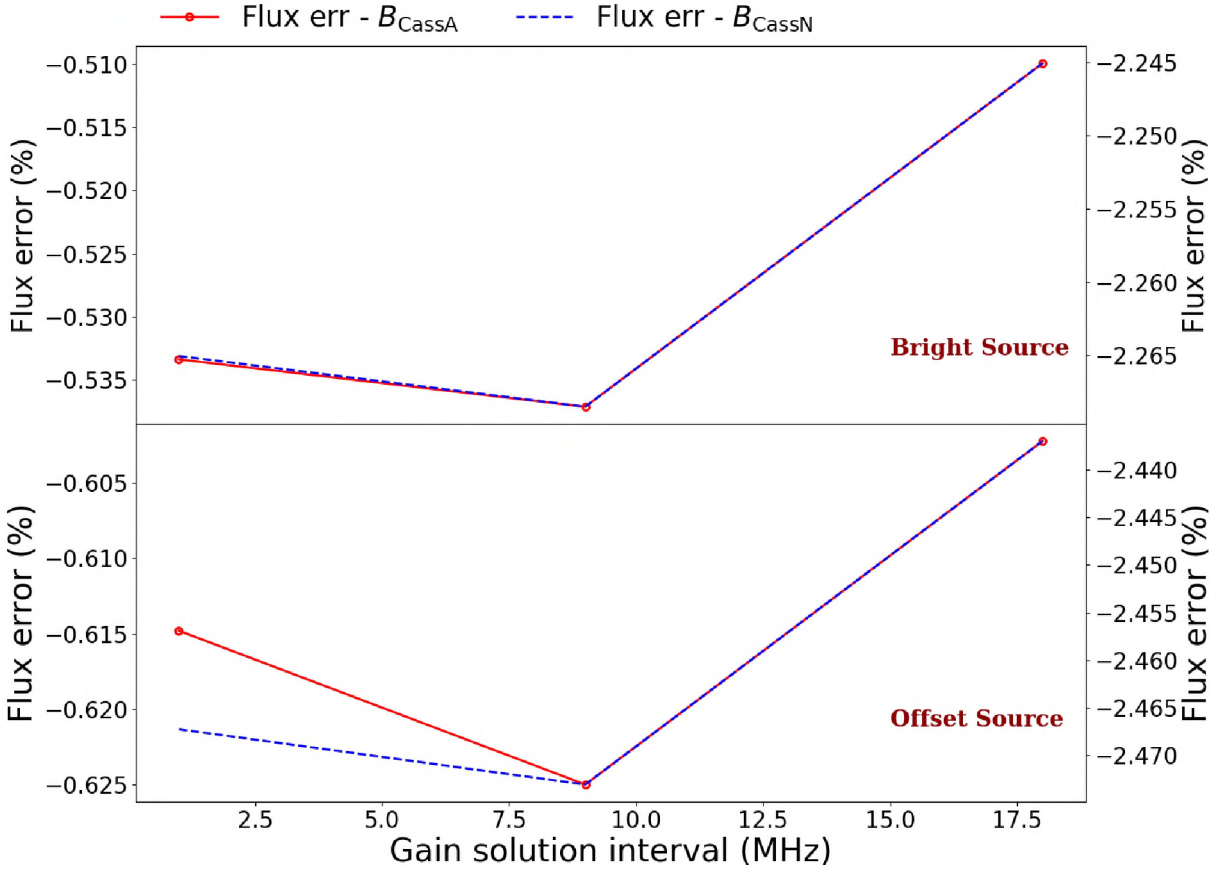

FIGURE 5.6: Errors in recovered flux densities against gain solution interval in frequency - for calibration done with B_{CassA} and B_{CassN} beams, plotted on primary and secondary y-axis time respectively. The top and bottom panels are for the bright and offset sources respectively - for two-source simulations.

5.4.4 Simulations with random distribution of sources

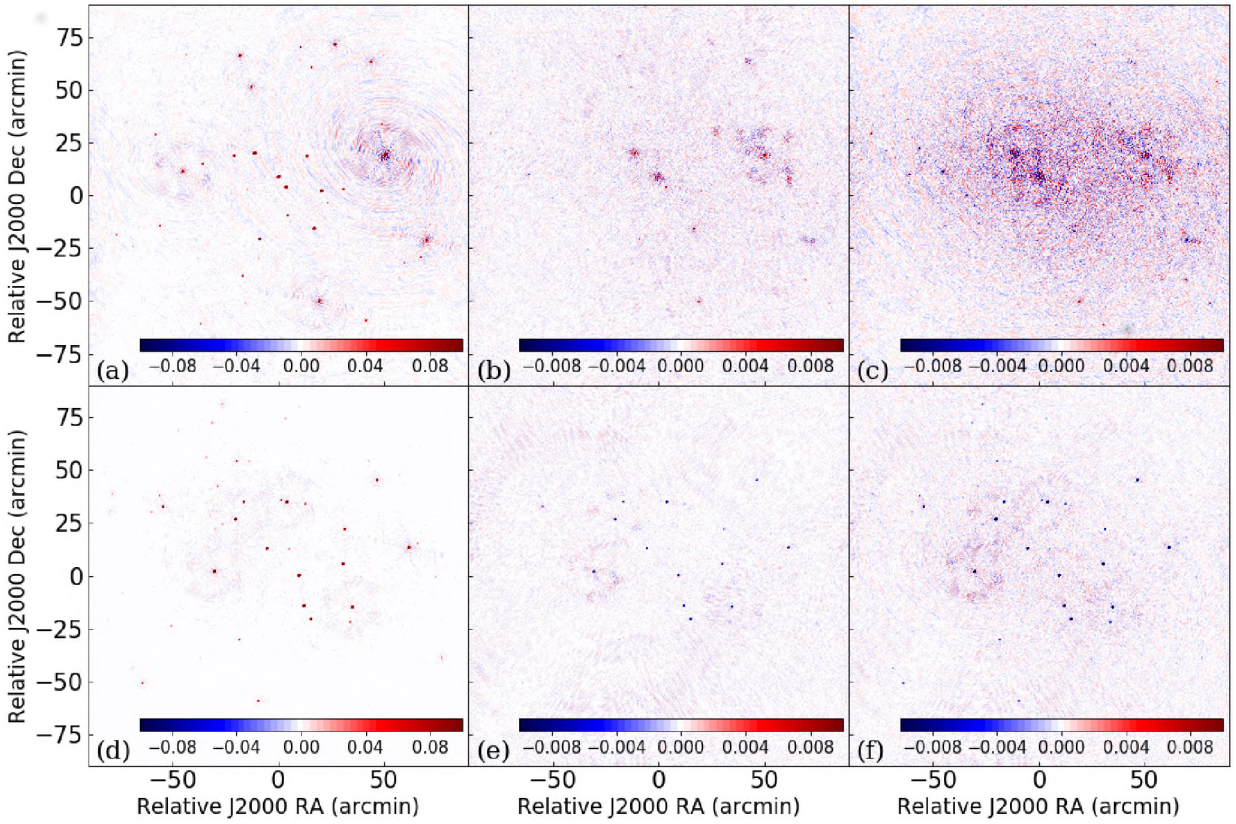

FIGURE 5.7: Two instances of the multi-sources simulation. The first column presents the Stokes I images of the apparent sky (= intrinsic sky × PB) in each case while the second and third columns present Stokes I images of the calibration residuals using B_{CassA} and B_{CassN} beam models, respectively, with a gain solution interval of 1 MHz. Each row is for a separate simulation. The colour bars have units of Jy

In the two experiments (cases 1 and 2) above, we considered two simple cases of sky models containing single- and two-point sources only, but in general, one expects multiple sources in a real observation of the sky. So, to test and ascertain how the ripple effects in the PB affect sources in a realistic case, we performed 100 simulations each having 100 point-sources with a flat spectrum distributed randomly within a FOV of 180 arcmins (using the ripple beams). Each of these simulations was done with the parameters and procedures outlined in Section 5.4.1. The simulated visibilities were calibrated with the same sky model that was used to simulate the data.

In Figure 5.7, we present the Stokes I images of two of randomly selected simulations. Each row represents an independent simulation. The first column presents images of the apparent sky while the second and third columns present the residual images for calibration done with B_{CassA} and B_{CassN} beams respectively. Again, the residual images of calibration done with the ripple beam, B_{CassR} are just uniform noisy maps (i.e. thermal

noise limited) because it is the very same beam used for simulating the observation. The first simulation (first row) has general source flux suppression[1] since the residual map still shows positive values at the location of the sources while the second selected simulation has a general overestimation of sources flux densities due to negative error values in source fluxes. The calibration done with the B_{CassA} beam had less artefacts and the sources are better subtracted (see Figure 5.7 (b) and (e)), than that of the calibration done using B_{CassN} beams (see Figure 5.7 (c) and (f)) for both instances. The calibration residuals in the first row are worse than that in the second by noting that sources are better recovered with less artefacts. It should be noted that the sky model in the first instance (first row) has its brightest source at 28 arcmins from the phase centre (in the first side-lobe close the first null) while the second has its brightest source at about 8 arcmins (still within the main lobe). This could be the origin of the observed difference.

As in Figure 5.5 (see above), Figure 5.8 presents the plots of the gain amplitude for the RR on and LR off-diagonal elements in the first and second columns, respectively for Antenna 5. The plots are similar to those of the single and two-sources simulations above (Figures 5.2 and 5.5, respectively). The graphs of the gain amplitudes, again manifest the characteristic behaviour of the beams spectral ripple effect (see top left panel of Figure 5.8) and the ripple in the gains flatten out as the solution interval increases.

We carried out further analyses on all the 100 multi-sources simulations by considering errors in source flux recovery, mean PB gain applied on each source, and mean gain amplitude. Values for the recovered fluxes were generated by executing source extraction on the images using the *PyBDSF* (Mohan and Rafferty, 2015) tool. The error in the flux density of the i^{th} source is calculated as:

$$f_{\text{err},i} = \frac{f_{B_{\text{CassR}},i} - f_{a,i}}{f_{B_{\text{CassR}},i}} \times 100, \tag{5.1}$$

where:
$f_{\text{err},i}$ is the error in the flux density of the i^{th} source in the image obtained by using the respective beams in the calibration,
$f_{a,i}$ is the ith source flux density for the calibration done with either B_{CassA} or B_{CassN} and $f_{B_{\text{CassR}},i}$ is the ith source flux density for the calibration done with B_{CassR} (the correct beam).

[1]Source flux suppression occurs when the recovered flux density for a certain source is less than the actual value of the source flux density.

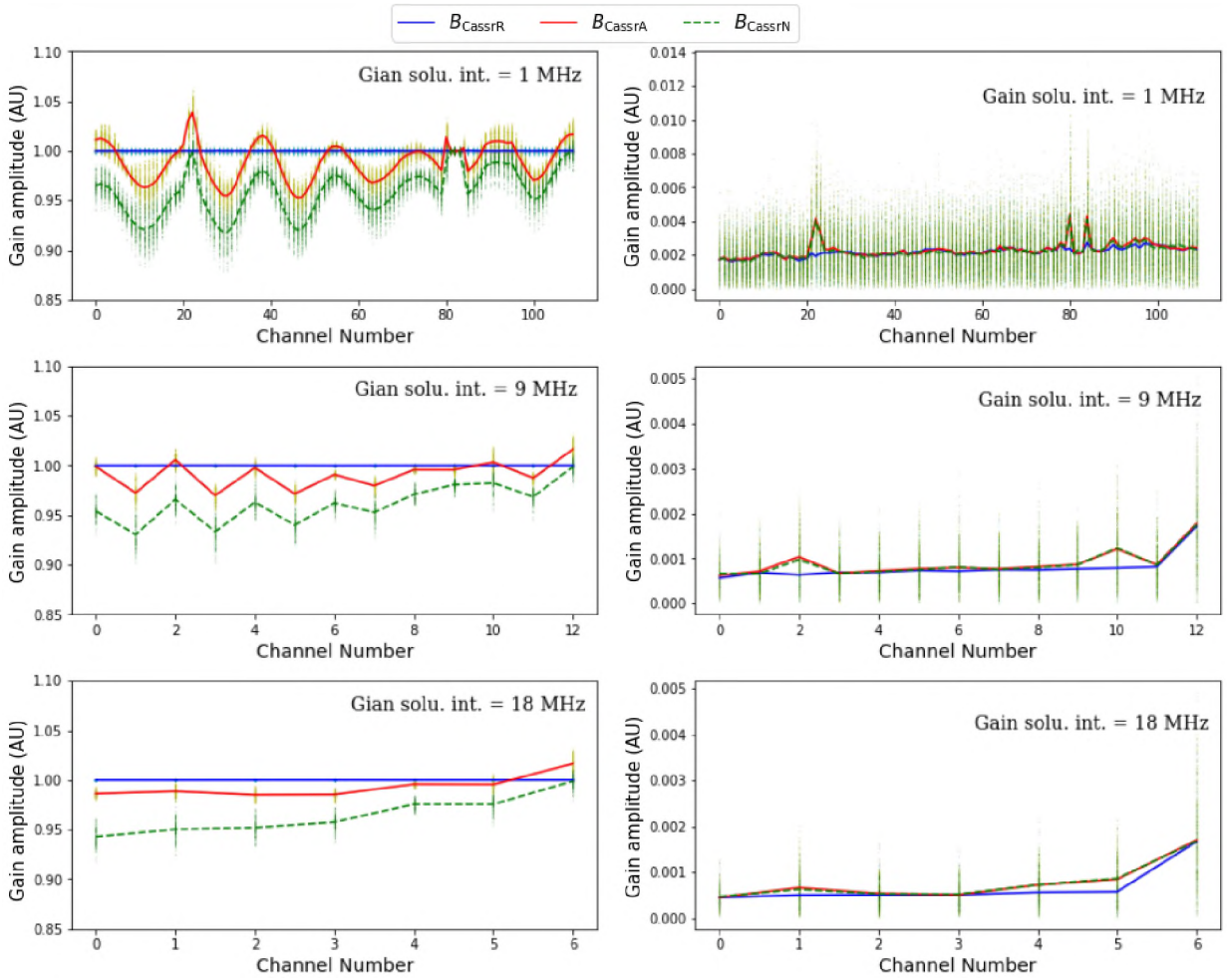

FIGURE 5.8: Graphs of the calibration gain amplitude for RR (first column) and LR (second column) for one of the sample simulation (second row of Figure 5.7) - for gain solution intervals in the frequencies of 1, 9 and 18 MHz as indicated in each panel (for Antenna 5). Shading is between the gains of the first and last frequencies while the solid line represent the gain of the centre frequency in each case.

In the same way, we calculated the mean calibration gain amplitude and the errors in them using:

$$g_{\text{err}} = \frac{g_{B_{\text{CassR}}} - g_a}{g_{B_{\text{CassR}}}} \times 100, \tag{5.2}$$

where:

g_{err} is the error in the desired mean calibration gain amplitude for calibration done with either B_{CassA} or B_{CassN},

g_a is the mean calibration gain amplitude for the calibration done with either B_{CassA} or B_{CassN} as the case may be and

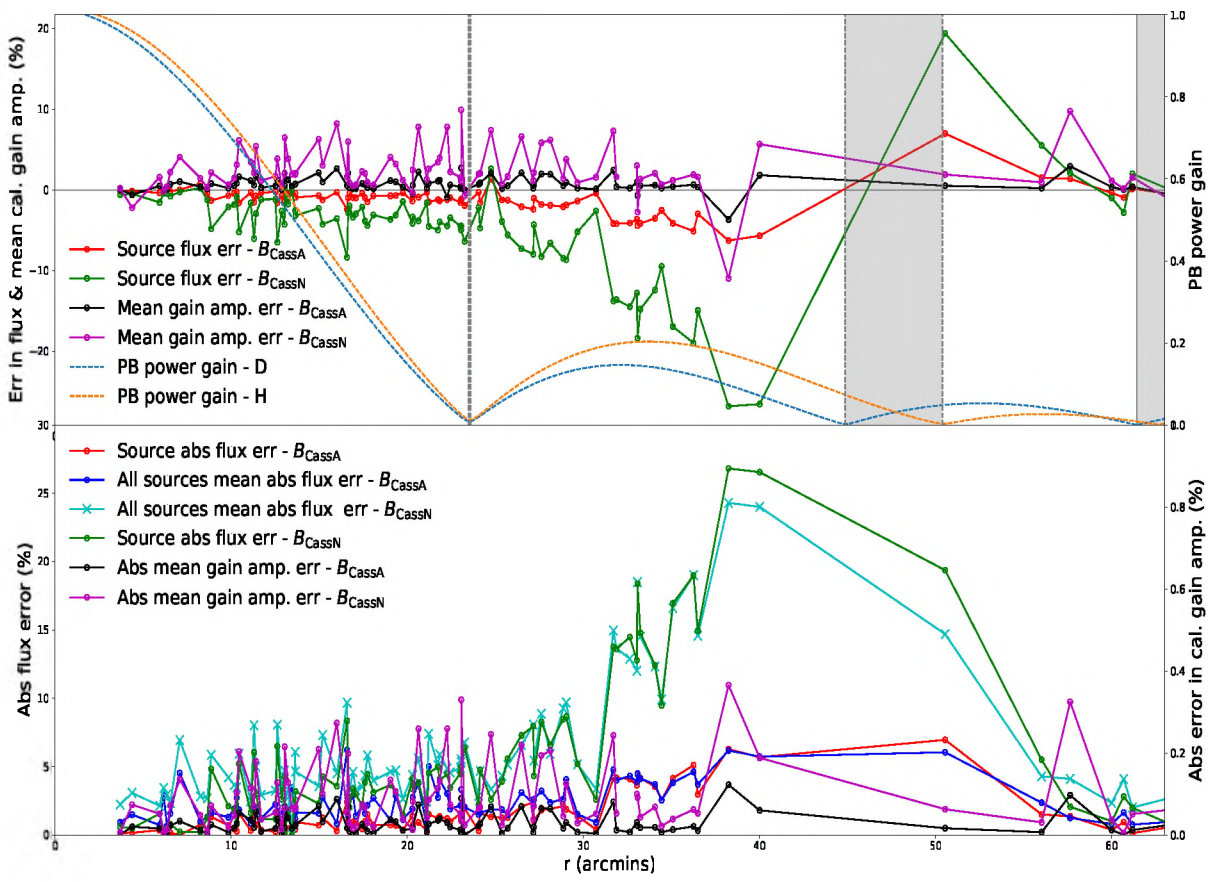

FIGURE 5.9: Graphs of the errors in recovered flux densities for the brightest source in each simulation (source flux err.) vs distance, r from the centre (top). The mean calibration gain amplitude errors (mean gain amp. err.) against brightest source r and the plots of 1D PBs power gain along the diagonal (D) and horizontal (H) (for B_{CassR}). (Bottom) graphs of the absolute value for the same data as in the top panel and the graphs of the mean of the absolute error in recovered flux densities for all sources at each simulation versus distance r of the brightest source (for calibration done with a frequency gain solution interval of 1 MHz).

$g_{B_{\mathrm{CassR}}}$ is mean calibration gain amplitude for the calibration done with B_{CassR} (the correct beam).

The PB power gain can be computed as:

$$P_{\mathrm{gain}}(\nu) = \frac{1}{2}\Big(\mathbf{Re}(\mathrm{RR}(\nu))^2 + \mathbf{Im}(\mathrm{RR}(\nu)^2 + \mathbf{Re}(\mathrm{LL}(\nu))^2 + \mathbf{Im}(\mathrm{LL}(\nu))^2\Big). \tag{5.3}$$

Finally, we computed the errors in mean PB gain that was applied on each source. Again, we used the mean PB gain of B_{CassR} applied to each source as a reference to compute the error in the mean PB gain for B_{CassA} and B_{CassN} with the expression:

$$g_{\text{applied,err}} = \frac{g_{\text{applied},B_{\text{CassR}}} - g_{\text{applied,a}}}{g_{\text{applied},B_{\text{CassR}}}} \times 100, \tag{5.4}$$

where: $g_{\text{applied,err}}$ is the error in the respective mean PB gain that was applied on each source for calibration done with either B_{CassA} or B_{CassN},
$g_{\text{applied,a}}$ is the mean PB gain that was applied on each source for the calibration done with B_{CassA} or B_{CassN} (as the case may be) and
$g_{\text{applied},B_{\text{CassR}}}$ is the mean PB gain that was applied on each source for the calibration done with B_{CassR} (the correct beam).

In the top panel of Figure 5.9, we have presented:

1. The plots of the errors in the recovered fluxes for the brightest source of each simulation against its radial distance from the centre for calibration done with either B_{CassA} and B_{CassN}.
2. The plots of the error in the mean calibration gain amplitude in each simulation for calibration done with CassA and CassN against the radial distance of the brightest source.
3. One-dimensional plots of the PB power gain along the x-axis and xy diagonal (for B_{CassR} computed using Equation (5.3)).

Our analysis here was restricted to all sources with radial positions lying within a radial distance less than or equal to approximately 60 arcmins from the phase centre, which corresponds to the radial distance ranging from the phase centre up to the third null. It can be concluded from the top panel of Figure 5.8:

1. That the errors in the mean calibration gain amplitude are mostly positive for both B_{CassA} and B_{CassN}.
2. That the calibration done with B_{CassA} produces lower errors than that done with B_{CassN} from the gain amplitude plots.
3. Positive errors in the mean gain amplitude produce overestimation of bright source flux densities while negative errors produced source flux suppression on average.

We wanted to see if there was any relationship between bright source flux density error and the mean absolute errors across all sources. The bottom panel of Figure 5.9 presents:

1. The plots of the absolute errors in the mean calibration gain amplitudes and the absolute errors in the brightest sources flux recovered against radial distances.

2. The plots of the mean absolute flux error for all sources against the radial distance of the brightest source in each simulation.

These plots show that the mean absolute error in flux density recovered for all sources is of comparable magnitude with the absolute errors in the recovered fluxes of the brightest source in each simulation. This meant that we could use the flux error analysis of the brightest source to approximately quantify how well flux densities of other sources in the FOV can be recovered. It also indicated that higher errors in mean calibration gain amplitude result in larger errors in sources flux recovery.

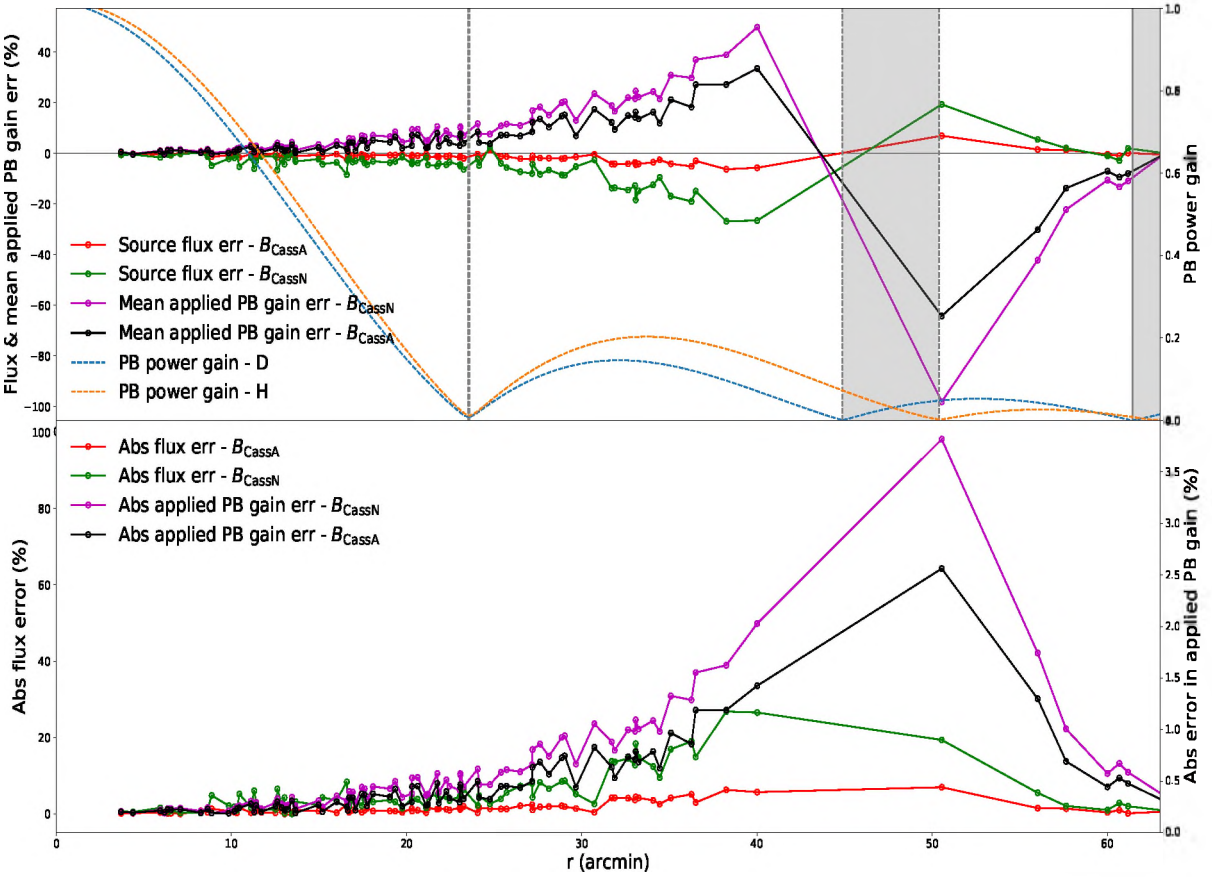

FIGURE 5.10: Graphs of the errors in the recovered flux densities for the brightest source (source flux err) of each simulation as a function of r (top), graphs of the error in the PB gain applied on each brightest source (applied PB gain err) vs r and the plots of 1D PBs power gain along D and H for B_{CassR}). The bottom panel presents the graphs of the absolute value of the same data in the top panel.

We went further in the comparison of the errors by comparing the error in the mean PB gain applied on the brightest source in each simulation with the error in its flux density

recovery. Recall that these simulated observations were done across the frequency sub-band of 110 MHz so, the mean PB gain applied on the sources is the average across this band. The top panel of Figure 5.10 shows exactly the corresponding graphs as a function of radial distance for the calibration done with B_{CassA} and B_{CassN}. The 1D PB power gain profile along diagonal (D) and horizontal (H) is plotted to highlight the position of the nulls. The grey shaded regions indicate positions that may have sources located near a null. The graphs show that positive errors in the mean PB gain applied on the bright sources produced overestimation of the source flux densities while negative errors produced source flux suppression and vice versa. The bottom panel presents the graphs of absolute values for the quantities in the top panel. This bottom panel shows (in a better way) that large errors in the mean PB gain applied on each source produced larger errors in the bright source flux recovery. This is presented in another way in Figure 5.11, which presents the graph of the absolute errors in recovered flux densities of bright sources against the absolute mean error in the PB gain applied on each bright source. There is a clear upward trend.

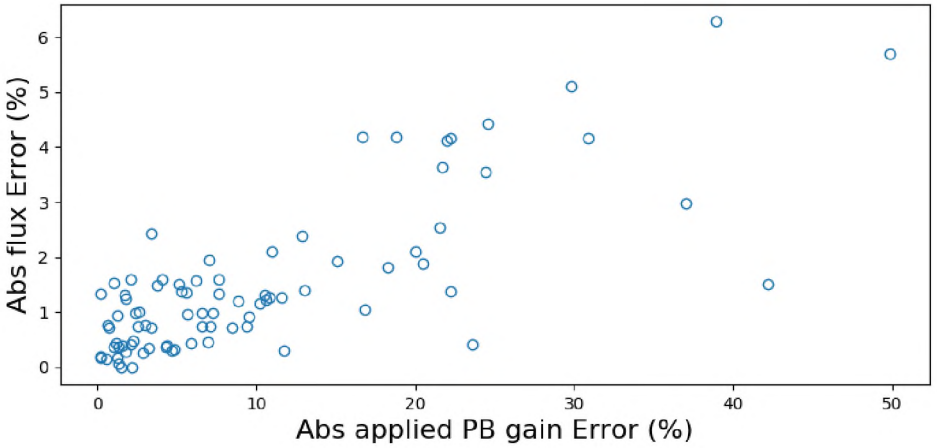

FIGURE 5.11: Graphs of the absolute errors in recovered flux densities of bright source vs the mean of the absolute error in the PB gain applied on each bright source (for calibration done with B_{CassA} using a gain solution interval of 1 MHz).

Figure 5.12 presents the graph of the brightest source absolute flux density errors against absolute calibration mean gain amplitude errors. There is a clear upward trend.

We can say from all these graphs, that the calibration done with B_{CassA} seems to bring better sources flux density recovery than that done with B_{CassN}.

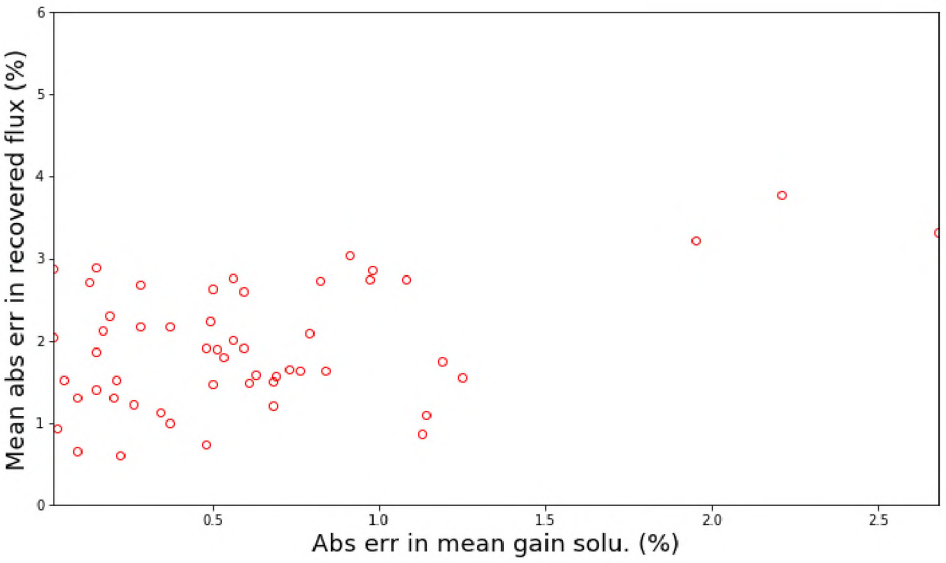

FIGURE 5.12: The graphs of the absolute errors in recovered flux densities of bright source (flux error) vs the mean error in the calibration gain amplitude (for calibration done with B_{CassA} using gain solution interval of 1 MHz).

5.5 Effects on dynamic range (DR)

Visibilities are generally corrupted by instrumental and environmental effects. Calibration is the process of removing (as much as possible these) corruptions from the data, to produce images which truly represent the observed sky. Even after calibration, the images produced generally still contain small errors. There is a challenge of determining the amount of errors left after calibration. This is not easy since one does not have *priori* knowledge of the sky distribution. DR is a metric commonly used to evaluate image quality.

There are many definitions of DR but we adopt the definition given in Taylor *et al.* (1999), which states that DR is defined as the ratio of the peak brightness, I_{peak} to the r.m.s noise, σ_I of a region in the image that has no emission given as:

$$\text{DR} = \frac{I_{\text{peak}}}{\sigma_I}. \tag{5.5}$$

We investigated how the PB size and centre offset ripples affect the DR of images after calibration with B_{CassA}, and B_{CassN} beam models.

We perform DR analysis on the results from the two selected multiple point-sources simulations of Section 5.4.4. Results of the analysis for the first simulation (the case

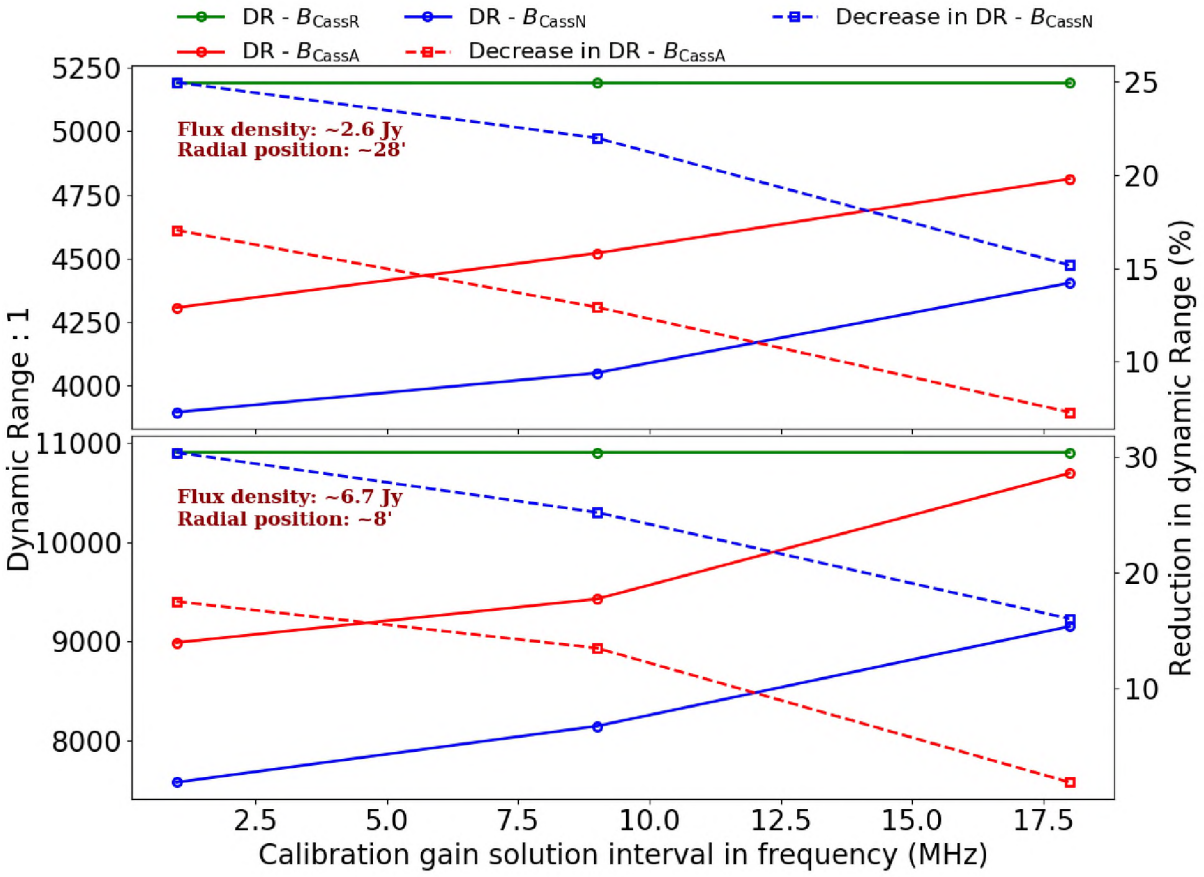

FIGURE 5.13: Graphs of the DR against the calibration gain solution intervals in frequency of the Stokes I images for the calibration done with B_{CassR}, B_{CassA} and B_{CassN} beams, and the percentage decrease in the DR. The case that has the bright source at approximately 28 arcmins is in the top panel while that which has the bright source at approximately 8 arcmins is in the bottom.

which has the bright source (2.6 Jy flux density) at about 28 arcmins is presented in the top panel while that for the second simulation (the bright source, having flux density of 6.7 Jy is at about 8 arcmins) is in the bottom panel of Figure 5.13. From this figure, we observe the following:

- The DR for the calibrations done with B_{CassR} is approximately constant while those of B_{CassA} and B_{CassN} beams increase with frequency solution interval.
- For the same gain solution interval, the B_{CassA} has higher DR compared to B_{CassN} (for the two cases) which implies better resultant images.
- The secondary y-axis shows that using a frequency gain solution interval of 18 MHz improved the DR by approximately 10% for both beams for the top panel and approximately 15% for the bottom panel.

- Using the approximate beam models for calibration (with frequency solution interval of 1 MHz) reduce the DR by approximately 17% and 25% for the B_{CassA} and B_{CassN} beams respectively for the top panel (and approximately 17% and 30% for the bottom panel).

These results show that ignoring the ripple characteristics in a PB model which is used for calibration indeed has impacts on the DR of the resulting image. In the cases under investigation, we see a reduction in DR to the tune of 30% because of the increased noise level in the images.

The next section considers the effect of the ripple characteristics on spectral data and HI science.

5.6 Spectral data simulations

Ignoring the"ripple" present in the PB of the VLA array beam model, not only has repercussions on the dynamic range and fidelity of the images of continuum emission but has implications when measuring the properties of spectral line sources as well.

A priori, we expect the ripple to affect spectral line parameter estimations through the combination of:
1) The difference between the PB model used for calibration and the "true" PB (i.e. B_{CassR}) - this depends on the position of the spectral line source;
2) The calibration gains derived and applied through the process of self-calibration – this depends on the parameters of observation and the field being observed.

To elaborate on the latter effect, we expect the gains of the calibration done using the incorrect PB models to be contaminated by the unaccounted-for PB ripples in the direction of the strongest source(s) in the FoV. The ripple in the PB amplitude would be absorbed into the estimated calibration gains and propagated via gain correction to the spectral line sources. In this section, we explore the magnitude of these effects and what it means for spectral line data measurement with VLA array. Note that this effect will only occur if self-calibration is performed with sufficiently small frequency solution intervals, which would only be necessary if the field contains bright sources. This is the scenario we examine here.

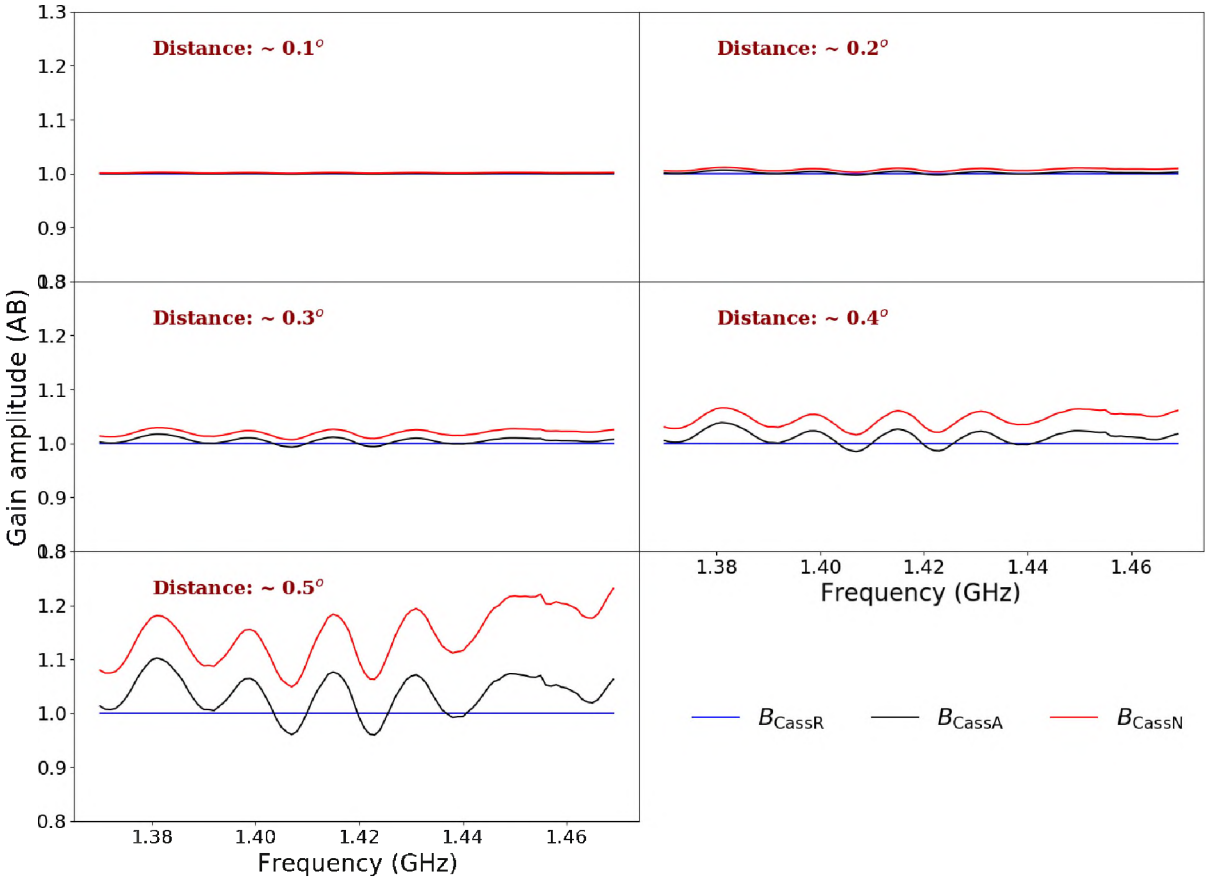

FIGURE 5.14: Self-calibration gain amplitude for calibration done with B_{CassR}, B_{CassA} and B_{CassN} beam models for the case when we have a single dominant offset source is placed at the distances of 0.1, 0.2, 0.3, 0.4 and 0.5 degrees from the phase centre (as indicated in the insert).

5.6.1 Simulations

To ascertain and quantify the above effects, we carried out simulations using the following process:

- First, we simulate an empty VLA D-configuration MS. This corresponds to a 4 hours synthesis time, 60 s integration time (to keep the data size small), and a frequency range of 1.37 to 1.47 GHz, with 1 MHz channel size. The frequency range was chosen to mimic the most common of radio spectral line observations, that of HI (neutral Hydrogen) 21 cm line produced through hyperfine splitting.

- Then, generate a sky model which has two faint point sources – one at the phase-centre and other at (a randomly selected distance of) approximately 0.35 degrees from the centre (both having intrinsic flux densities of 1 mJy and having a flat spectrum). These are the spectral (line) sources, and we call this sky model the

faint source sky model or faint sky. This model will be used to investigate how the ripple effects affect the spectral properties of the faint, diffuse and HI sources.

- Next, generate another sky model with a point source (also having a flat spectrum) at an offset position of 0.1 degrees (from the phase-centre) having apparent flux density of approximately 1 Jy. This is our continuum foreground sky model, and this source is referred to as the bright source.

- Then, simulate visibilities (with no noise) using B_{CassR} beam for the faint sky model and the continuum (foreground) sky model separately over the frequency band of the MS.

- Add the visibilities of the faint sources to that of the continuum foreground source to get our full simulated data.

- Calibrate the full data using the continuum foreground sky model with B_{CassR}, B_{CassA} and B_{CassN} beam models (using gain solution frequency interval of 1 MHz and time of 1 minute).

- Repeat the simulations for the offset continuum bright source placed at distances of 0.2, 0.3, 0.4 and 0.5 degrees from the phase centre and maintaining the apparent flux densities of this source at approximately 1 Jy (for all the simulations). The apparent flux density of this source was maintained at approximately 1 Jy so that we probe the PB influence on the estimated gains, while not being affected by relative apparent flux densities of the source at different distances.

- Finally, analyse the residual images of the calibrated data after continuum subtraction of the bright source.

The graph of the self-calibration gains as a function of frequency (for RR element and one antenna) are presented in Figure 5.14. From these graphs, we see that:

- The average gain for the self-calibration done with B_{CassR} is approximately 1 (having no ripple) for all cases as expected. For the self-calibration done using B_{CassA} and B_{CassN} the gains have absorbed the spectral line ripples.

- The amplitude of these ripples increases at the distance of the foreground source increase from the centre.

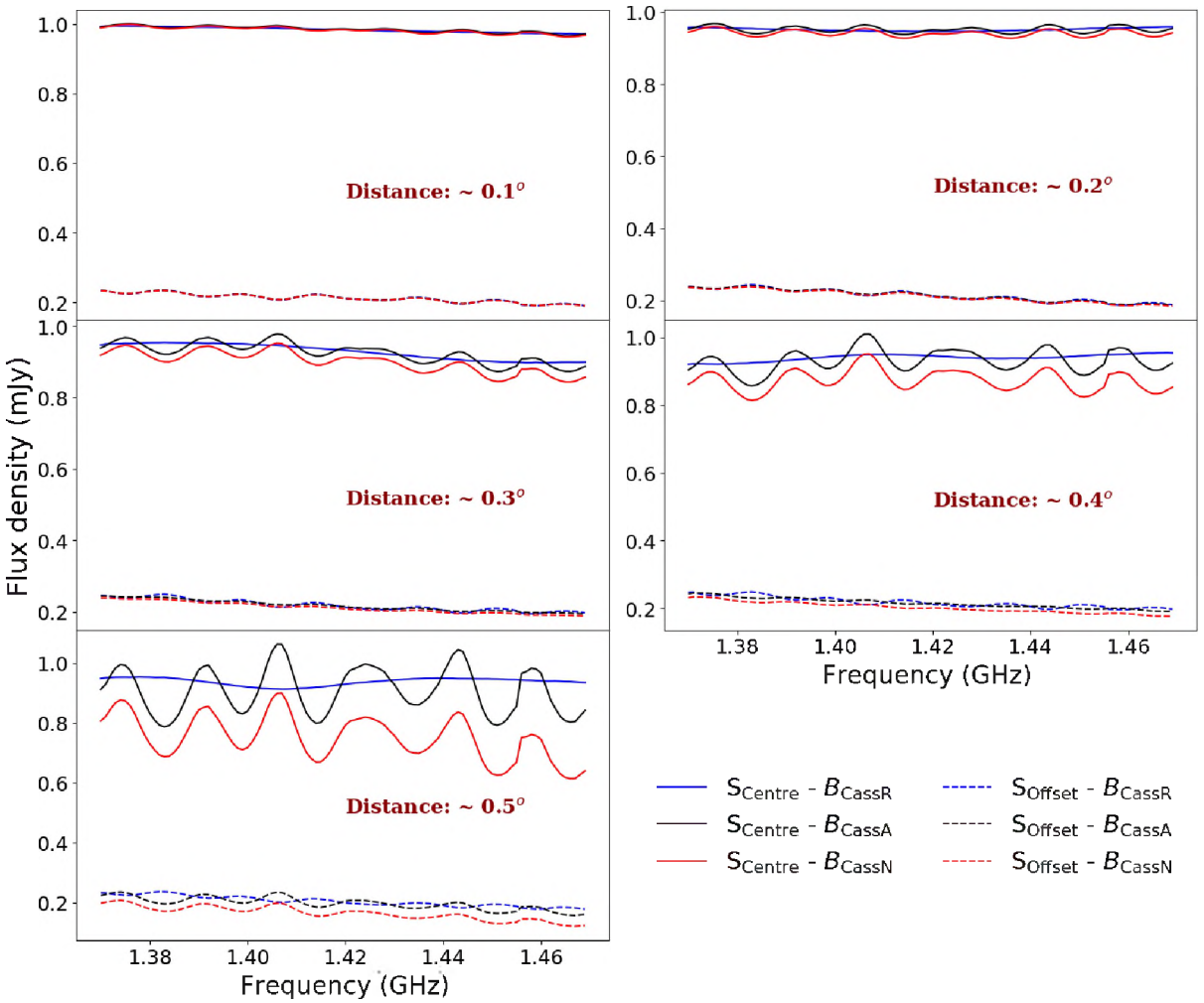

FIGURE 5.15: Recovered apparent flux densities of the faint sources for calibration done with B_{CassR}, B_{CassA} and B_{CassN} beam models. For the case when the bright dominant continuum offset (or bright) source is placed at the distances as indicated in the insert. $\mathrm{S}_{\mathrm{Centre}}$ and $\mathrm{S}_{\mathrm{Offset}}$ represent the centre and offset sources, respectively.

- The average gains for calibration done using B_{CassA} had smaller ripple amplitude than that done using B_{CassN} for all cases. That is, the gain amplitude for B_{CassA} better approximates the gain value of 1.

The recovered flux densities of the two faint point sources (after continuum foreground subtraction) as a function of frequency are shown in Figure 5.15. From this figure, we can see that:

- The B_{CassR} beams spectral ripple characteristics were transferred to the spectral behaviour of the recovered flux densities of faint sources (for calibration done using B_{CassA} and B_{CassN} beam models).

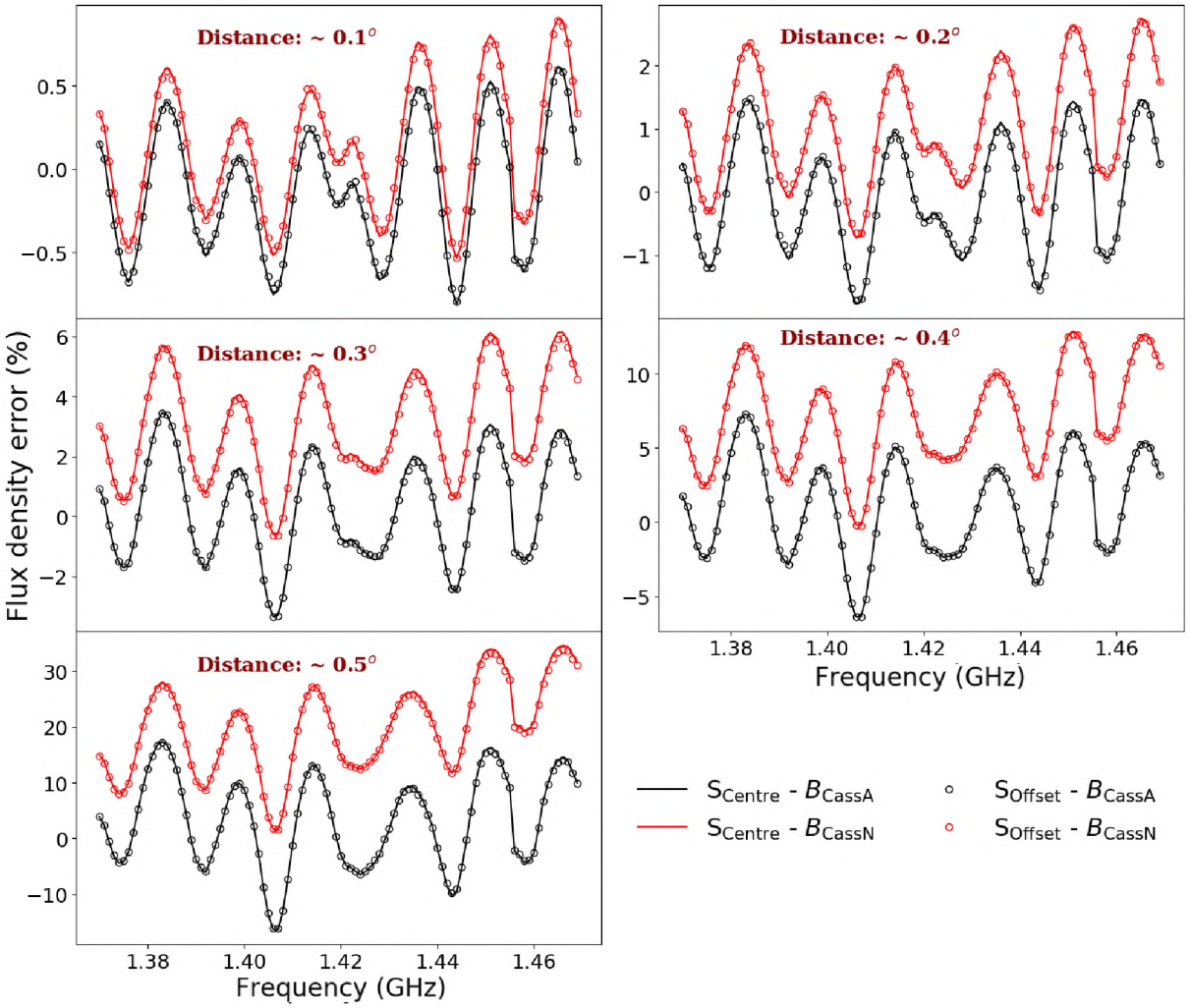

FIGURE 5.16: Errors in recovered flux densities of the faint sources for calibration done using B_{CassA} and B_{CassN} beam models for the case when we have a single dominant offset source placed at the distances as indicated in the insert.

- Note the behaviour of the ripple on the (flux density of the) faint offset source (bottom curves). There is a general trend in increasing ripple amplitude as the dominant source is moved further off-centre, but at 0.3 (and 0.4) degrees there is actually a slight decrease.

Overall, the B_{CassA} model has a better flux recovery for the two faint sources for all simulations. To see this better, let us consider Figure 5.16. In Figure 5.16, we present graphs of the errors in recovered flux densities of the two faint sources (after continuum subtraction of the bright source). From these graphs, we see that:

- The ripple characteristics affect how the flux densities of the sources are recovered.

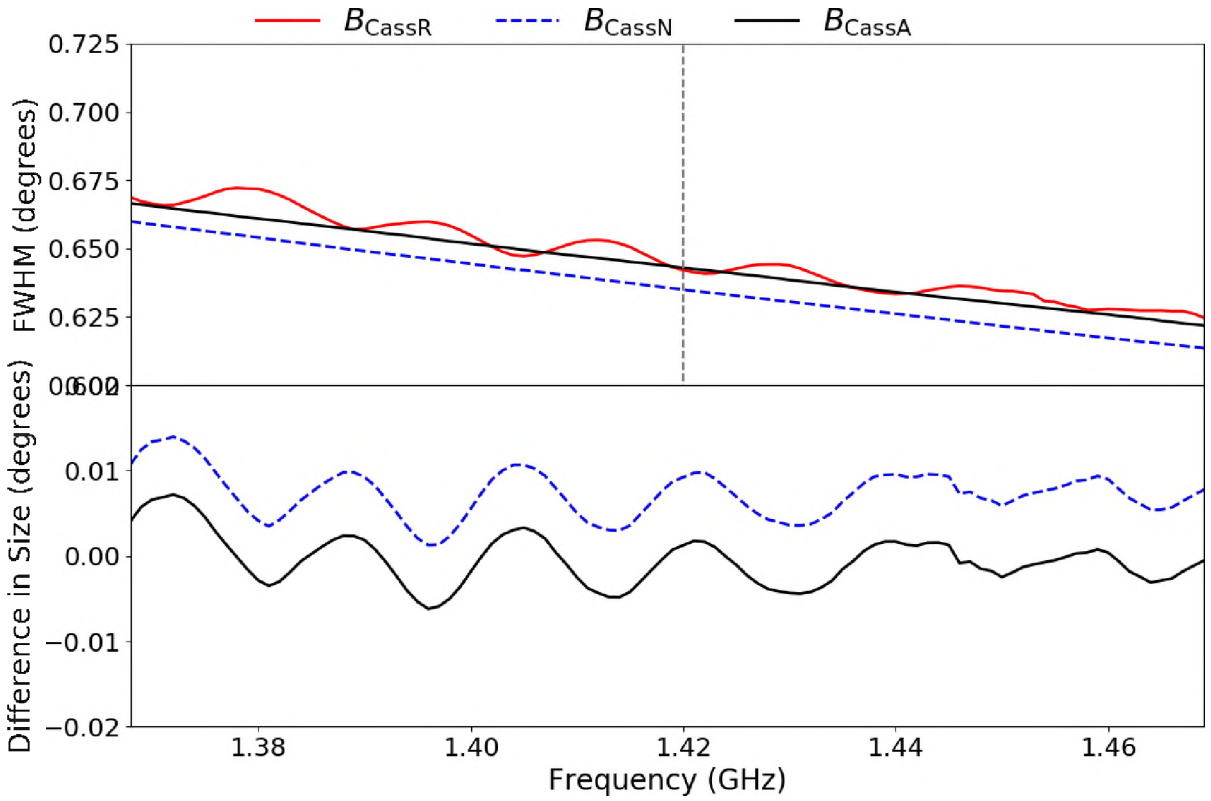

FIGURE 5.17: The top panel is the zoom-in plots of the graphs of FWHM of B_{CassR}, B_{CassA} and B_{CassN} beam models. In the bottom panel, is presented difference in FWHM ($B_{\mathrm{CassR}} - B_{\mathrm{CassA}}$ and $_{\mathrm{CassR}} - B_{\mathrm{CassN}}$). The vertical dash line indicates the position of the centre frequency for this band.

- The errors in the recovered flux densities of the two faint sources are approximately the same for the calibration done with the same beam in each simulation. This confirms that the errors in recovered spectral properties of the faint sources are dominated by the errors in recovered properties of the bright continuum source.

- The errors increase with an increase in the position of the bright source from the phase centre.

- Although the graphs of the recovered flux densities in Figure 5.15 appear to exhibited relatively flat spectral characteristics for the faint offset source, for the case of 0.3 and 0.4 degrees (which could be mistaken to imply better flux recovery). We see that in Figure 5.16, it did not improve the faint source flux recovery. From Figure 5.16, we also see that the error in recovered flux densities for the centre and offset faint sources are the same. How well the sources are recovered is strongly dominated by the error at the position of the continuum bright source.

- There is a systematic shift between the graphs of errors in recovered flux densities for calibration done using B_{CassA} and B_{CassN} beam models in each simulation. Indeed,

these shift can be noticed in all the other graphs in Figures 5.14 and 5.15. This observed effect can be explained if we consider Figure 5.17.

In Figure 5.17, we present the zoom-in graphs of the FWHM of B_{CassR}, B_{CassA} and B_{CassN} beam models within the band (1.37 to 1.47 GHz) under consideration. These graphs (in the top panel) show a systematic difference (shift) between the graphs of the FWHM of the B_{CassA} and B_{CassN} beam models. In the bottom panel of the same, we present the difference between the size of the beams (B_{CassA} and B_{CassN}) and that of the ripple beam (B_{CassR}). These graphs explain the reason why we have these observed systematic offset in the results of the simulations done using B_{CassA} and B_{CassN}, which we have presented in this section. The trend in the graphs of the calibration gains and recovered flux densities (of the faint sources) closely follow the trend of the graphs in Figure 5.17 (for B_{CassA} and B_{CassN}).

5.6.2 Implications for HI mass measurements

From the above investigation, we see that the absorption of the PB ripples into the self-calibration gains would result in several errors. The first of these errors would be an incomplete continuum subtraction of the foreground sources. This would subsequently result in continuum artefacts, which would, in turn, affect the estimation of spectral line measurement, unless the false variation introduced into the gains (and thus flux density of sources) are modelled by using a function of appropriate complexity (e.g. a high degree polynomial while carrying out uv-plane subtraction).

The mass of a HI source is directly proportional to its measured flux density. Apart from this aforementioned rather obvious effects, a more subtle effect would be its effect on HI mass measurement. One can obtain the flux density variation as a function of reshift by computing redshift from the frequency. We may take the variation of these flux densities as percentage error in measuring the "true" HI mass. The errors in the recovered flux density are less than 10% in our simulations when foreground continuum source is within the main lobe (see Figure 5.16). For the case of 0.5 degrees, we see errors up to approximately 35% in flux density recovery (for B_{CassN}). One of the reasons why these errors are small within the main lobe (less than 10%) is because the VLA PB ripple (peak-to-peak) amplitude decreases with frequency and is small at this band. At lower frequencies (where we have large ripple amplitude), the error in the HI flux density would be more prominent. What makes this effect important is the fact that this is a

systematic effect which would affect the distribution of the properties of the HI sources (with redshift) rather than a single HI source. This is because the typical width of a HI source might be approximately 2 MHz, while the ripple has a frequency which is several times larger than that. An accurate estimation of this effect would require, for example, convolving the error found in flux density/mass with the distribution of HI sources with redshift, while also considering more realistic simulations. This is beyond the scope of this investigation but would form an important direction for future work.

5.7 Summary

A number of experiments were performed in this chapter to test the impact of not incorporating the ripple characteristics found in the VLA antenna PB sizes and offsets in PB models used for calibration and imaging. We started with a single point-source simulation, followed by two-source simulations, and finished with 100 different multiple sources simulations (having a random distribution of 100 point-sources each). All the sources had flat spectra. Calibration was done on all these simulations using the correct beam and two approximate beams for different frequency gain solution intervals and a time interval of 1 minute.

The results show that calibration done with the approximate beam models absorb the ripple effects into the calibration solutions, thereby producing either flux suppression or overestimation. Artefacts are also produced around bright sources because their PSF side-lobes are not properly subtracted. The recovered flux density errors were less than 10% on the average. Increasing the frequency gain solution intervals from 1 to 18 MHz had a negligible improvement in the recovered flux density of the sources for calibration done with the approximate beams.

Calculating the DR of the images produced after calibration for beams that completely ignored the ripple characteristics (*Cassbeam* model) reduced the DR by a factor of up to approximately 30%, while that done with a beam that takes the average characteristics into account (average beam model) had a reduction of up to approximately 17%. Increasing the frequency gain solution intervals improved the DR and using a frequency gain solution interval of 18 MHz in calibration improved the image quality produced by up to 15%. We used the frequency gain solution intervals of 1, 9 and 18 MHz for the calibrations because the frequency of the ripples is about approximately 17 MHz (for the VLA antennas (Popping and Braun, 2008; Jagannathan *et al.*, 2018)). We wanted to check

if setting the solution interval to some value larger than the ripple period is sufficient to ignore the ripple characteristics in the PB models for the VLA in the L-band. This indeed improves the DR (i.e. image quality) but is not sufficient for one to completely ignore ripple characteristics in the PB model for VLA antennas, particularly in the case where the goal is to produce a high DR image.

Finally, using PB models which ignore the ripple effects could lead to errors of up 35% (and 20% for the average beam) in the recovered flux densities of faint emissions after continuum subtraction of the bright source. This could also result to errors of up to 35% when measuring HI mass and other spectral properties for the case when the bright continuum foreground source lies at the side-lobes for a beam model which completed ignores the ripple effects.

Chapter 6

Conclusions

The major focus of this study was to develop an accurate and efficient modelling technique based on holography measured data.

The PB of the radio antennas have been modelled with a $\cos^3$ function (Popping and Braun, 2008; Braun *et al.*, 2007), Jacobi-Bessel patterns (Galindo-Israel and Mittra, 1977; Rahmat-Samii and Galindo-Israel, 1980), CBFP (Young *et al.*, 2013; Mutonkole and de Villiers, 2015) and *Cassbeam* software (Brisken, 2003) for Cassegrain antennas. The latter uses the technique of optical ray-tracing with antenna parameters to model the PB of Cassegrain antennas. These techniques produce models of PB from simulations which do not fully capture the real spectral and spatial behaviours of antennas PB. Jagannathan *et al.* (2018) improved on the *Cassbeam* software by developing an algorithm to optimise the input parameters of this software to absorb the *ripple effect* observed in the VLA holography measured beams, e.g. parameters like the size of the sub-reflector (aperture blocking). These methods require prior knowledge of the design of the antenna and cannot easily be applied, for example, to produce PB models for non-aperture blocking offset Cassegrain or Gregorian designs. In our approach, we use a set of techniques that are antenna design-independent to generate PB models for the VLA and MeerKAT antennas from holography measured data.

We used two orthonormal basis methods to model the PB from holography measurement - PCA and ZP. The PCA PB model for VLA had a reconstruction error of approximately 4% across all antennas while that for the MeerKAT had a reconstruction error of approximately 3% (with 20 and 15 coefficients respectively). These produced approximately 98% and 99.85% compression (compared with measured data) for the VLA and MeerKAT antennas respectively. The output VLA and MeerKAT antennas beams are

available as described in Appendix A. In this study, we outlined the steps which can be used to produce accurate PB models for any telescope (which is independent of its design) based on holography measured data as well as interpolate across frequencies. Applying the PCA technique to produce PB models for Cassegrain off-axis (VLA) and Gregorian offset (MeerKAT) antenna designs demonstrate the potential of using it to model PB for antennas of any design from its holography measured beams.

The technique detailed in this study can be applied to model and compress any holography data that are sampled in frequency. If it has some sub-bands which were corrupted (by RFI or missing), one can use the recipes outlined in this study to interpolate and approximate them.

The impact of using PB models that did not incorporate the ripple effect was investigated via simulation. These show that, although the calibration done with these incorrect PB models manifested some errors in the gain amplitude, the Stokes I images of the results had less than 10% average errors in sources flux density recovery. The errors in the fluxes' density recovery of the brightest source dominated how well other sources were recovered.

From the experiments, if 10% error in the result is sufficient for the science of interest, one can continue with a beam model that ignores the ripple effects for calibration. Finally, the dynamic range analysis shows that this will produce degradation of up to 30% in the output images for PB models that ignore completely the spectral beam sizes and centre offset ripples (the default beams produced by the *Cassbeam* software), while the models that account for the average spectral ripples characteristics (average beam) will produce a degradation of up to 17% in the image. Increasing the frequency gain solution interval used for continuum calibration, improved the image quality by a factor of 10% for the normal *Cassbeam* model and 15% for the average beam model. Hence, using high frequency solution intervals is not sufficient to ignore the ripples in the PB model for the VLA, particularly for the case where the science of interest requires high dynamic range.

Another case where the ripple effects are important is measurement of spectral line emission in radio astronomy. In this work, we have considered the case of neutral Hydrogen 21 cm line. We find that ignoring the PB ripples results in errors in measurement of line emission properties. These errors are caused by both the trivial image plane PB correction and the process of continuum subtraction. We see that these errors can translate into up to 30% variation in HI mass. We also see that these effects are highly dependent on the solution interval of the self-calibration. Longer solution interval in frequency tends to

"average out" these effects. However, we note that these will be important for up coming surveys especially those done with high frequency resolution.

6.1 Future studies

The ripple effects that were observed in the VLA beams at L-band are also present in the S-band but with smaller amplitude. Future work would involve extending this study in time and frequency and for all frequency bands. Future holography measurements will obtain data at different elevations and directions; these procedures can be applied to the datasets to derive models that incorporate variations in the PB due to elevation. Machine/deep learning could also be used to extract the features of the beam in frequency and model it. Another approach one can take is to model an average antenna, and derive the correction coefficients for each antenna from this average beam model.

The impact of ripple characteristics in the PB can be tested by having an unmodelled transient source in the data, and testing how it can be recovered depending on its location on the beam. It can also be tested if one can use holography method to get information about the ionosphere and how the location of the ionosphere would affect the PB during observations. One can also test if the holography methods could be used to get information on the local ionosphere during the observation.

Appendix A

Data products

A number of data products have been produced as part of this work and should be considered an integral part of the thesis. These primarily consist of "beam cubes" implementing the different VLA beam models derived during this work. The beam cubes are available for download via `https://www.ratt-ru.org/projects-beams-jvla.html`.

Each beam model is represented by eight FITS cubes: one cube per the real and imaginary part of each of the four Jones matrix elements. This is designated by suffixes on the file names, i.e. `*rr_re.fits` corresponds to the RR correlation (i.e. E_{11} matrix element), the real part. For reference purposes, the FITS header of each cube looks as follows:

```
SIMPLE  =                    T / conforms to FITS standard
BITPIX  =                  -64 / array data type
NAXIS   =                    3 / number of array dimensions
NAXIS1  =                  513
NAXIS2  =                  513
NAXIS3  =                  992
DATE    = '2019-05-22 12:32:12.401224'
DATE-OBS= '2019-05-22 12:32:12.401224'
ORIGIN  = 'KIHEANETU'
TELESCOP= 'VLA     '
OBJECT  = 'beam    '
EQUINOX =               2000.0
CTYPE1  = 'X       '           / points right on the sky
CUNIT1  = 'DEG     '
```

```
CDELT1  =             0.011082 / degrees
CRPIX1  =                  257 / reference pixel (one relative)
CRVAL1  =                  0.0
CTYPE2  = 'Y       '           / points up on the sky
CUNIT2  = 'DEG     '
CDELT2  =             0.011082 / degrees
CRPIX2  =                  257 / reference pixel (one relative)
CRVAL2  =                  0.0
CTYPE3  = 'FREQ    '
CDELT3  =            1000000.0 / frequency step in Hz
CRPIX3  =                    1 / reference frequency postion
CRVAL3  =         1008000000.0 / reference frequency
```

Each cube thus has "spatial" dimensions of 513×513, and 1000 frequency channels of 1 MHz each, starting with 1008 MHz. The X/Y axis refers to the telescope frame: X points right, and Y points up for an observer standing behind the dish and looking in the same direction. For a parallactic angle of 0, the X-axis thus points West on the sky, and the Y-axis points North[1].

In all cases, the beam patterns have been shifted and interpolated so that the peak of the effective Stokes I beam is at the centre (reference) pixel, which implies that the peaks of the RR and LL beams are offset from the centre due to the "squint".

The following models are available for download:

cassNormal: This is the beam model produced by the *Cassbeam* package using its default VLA settings, with the peak then shifted to the centre pixel as discussed above. This model is referred to as *normal Cassbeam* (B_{CassN}) in Chapter 5.

cassAverage: This is the "cassNormal" model, rescaled so that the beam size matches the average beam size as a function of frequency, as measured on the holography data. This model is referred to as *average* beam (B_{CassA}) in Chapter 5.

cassRipple: This is the "cassNormal" model, rescaled so that the beam size matches the average beam size and the ripple as a function of frequency, as measured on the holography data. This model is referred to as *ripple* beam (B_{CassR}) in Chapter 5.

[1]This scheme of encoding the beam as FITS files is natively supported by the `MeqTrees`, `DDFacet` and `killMS` software packages

PCA: This is the model derived from holography using the PCA approach, as discussed in Chapter 4.

ZP: This is the model derived by fitting Zernike polynomials and keeping the 15 highest coefficients, as discussed in Chapter 4.

Bibliography

K M B Asad, K Iheanetu, J Girarad, M de Villiers, T Ansah-Narh, O Smirnov, M Santos, K Thorat, and S Makhathini. Primary beam effects of radio astronomy antennas: Ii. meerkat l-band primary beam models. *Sumbitted to Monthly Notices of the Royal Astronomical Society*, 2019.

RMA Azzam and NM Bashara. Ellipsometry and polarized light, 1977.

JWM Baars, R Lucas, JG Mangum, and JA Lopez-Perez. Near-field radio holography of large reflector antennas. *arXiv preprint arXiv:0710.4244*, 2007.

Constantin Balanis. Antenna theory analysis and design 3rd edition, a john wiley & sons. *Inc., Publication*, 2005.

N Barry, B Hazelton, I Sullivan, MF Morales, and JC Pober. Calibration requirements for detecting the 21 cm epoch of reionization power spectrum and implications for the ska. *Monthly Notices of the Royal Astronomical Society*, page stw1380, 2016.

Michael Bass, Casimer DeCusatis, Jay M Enoch, Vasudevan Lakshminarayanan, Guifang Li, Carolyn MacDonald, Virendra N Mahajan, and Eric Van Stryland. Handbook of optics, volume i: Geometrical and physical optics, polarized light, components and instruments (set). *The McGraw-Hill Companies*, 2009.

S Bhatnagar, TJ Cornwell, K Golap, and Socorro NRAO. Evla memo 84. solving for the antenna based pointing errors. Technical report, Tech. rep., NRAO, 2004.

AJ Boonstra and AJ Van Der Veen. Gain decomposition methods for radio telescope arrays. In *Statistical Signal Processing, 2001. Proceedings of the 11th IEEE Signal Processing Workshop on*, pages 365–368. IEEE, 2001.

RS Booth, WJG De Blok, JL Jonas, and B Fanaroff. Meerkat key project science, specifications, and proposals. *arXiv preprint arXiv:0910.2935*, 2009.

Max Born and Emil Wolf. Principles of optics electromagnetic theory of propagation. *Principles of Optics Electromagnetic Theory of Propagation, INterference and Diffraction of Light 2nd edition by Max Born, Emil Wolf New York, NY: Pergamon Press, 1964*, 1, 1964.

Max Born and Emil Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffaction of Light*. Pergamon Press, 1980.

Max Born and Emil Wolf. Principles of optics, 7th (expanded) ed. *Cambridge U. Press, Cambridge, UK*, 890, 1999.

Lee C Bradley and Jan Herrmann. Phase compensation for thermal blooming. *Applied Optics*, 13(2):331–334, 1974.

Robert Braun, TA Oosterloo, Raffaella Morganti, Uli Klein, and Rainer Beck. The westerbork sings survey-i. overview and image atlas. *Astronomy & Astrophysics*, 461 (2):455–470, 2007.

FH Briggs, W van Driel, RC Kraan-Korteweg, and E Sorar. Driftscan surveys in the 21 cm line with the arecibo and nançay telescopes. *Publications of the Astronomical Society of Australia*, 14(1):37–44, 1997.

Walter Brisken. Evla memo 58. using grasp8 to study the vla beam, 2003.

WN Charman. Wavefront technology: past, present and future. *Contact Lens and Anterior Eye*, 28(2):75–92, 2005.

TJ Cornwell and PN Wilkinson. A new method for making maps with unstable radio interferometers. *Monthly Notices of the Royal Astronomical Society*, 196(4):1067–1086, 1981.

Dirk de Villiers. Characteristic basis function pattern modeling of wideband reflector antenna beam pattern frequency variations. In *2018 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, pages 220–223. IEEE, 2018.

Dirk IL De Villiers. Prediction of aperture efficiency ripple in clear aperture offset gregorian antennas. *IEEE Transactions on Antennas and Propagation*, 61(5):2457–2465, 2013.

Chelo Ferreira, José L López, Rafael Navarro, and Ester Perez Sinusia. Orthogonal systems of zernike type in polygons and polygonal facets. *arXiv preprint arXiv:1506.07396*, 2015.

Victor Galindo-Israel and R Mittra. A new series representation for the radiation integral with application to reflector antennas. *IEEE Transactions on Antennas and Propagation*, 25(5):631–641, 1977.

TL Grobler, CD Nunhokee, OM Smirnov, AJ Van Zyl, and AG De Bruyn. Calibration artefacts in radio interferometry–i. ghost sources in westerbork synthesis radio telescope data. *Monthly Notices of the Royal Astronomical Society*, 439(4):4030–4047, 2014.

Christine Guillemot and Olivier Le Meur. Image inpainting: Overview and recent advances. *IEEE signal processing magazine*, 31(1):127–144, 2014.

Johan P Hamaker. Understanding radio polarimetry-iv. the full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity. *Astronomy and Astrophysics supplement series*, 143(3):515–534, 2000.

JP Hamaker, JD Bregman, and RJ Sault. Understanding radio polarimetry. i. mathematical foundations. *Astronomy and Astrophysics Supplement Series*, 117(1):137–147, 1996.

JA Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 15:417, 1974.

K Iheanetu, J Girarad, O Smirnov, K M B Asad, M de Villiers, K Thorat, S Makhathini, and R Perley. Primary beam effects of radio astronomy antennas: I modelling the karl jansky. vla l-band beam using holography. *Monthly Notices of the Royal Astronomical Society*, 2019.

HT Intema, S Van der Tol, WD Cotton, AS Cohen, IM Van Bemmel, and HJA Röttgering. Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme-i. method description and first results. *Astronomy & Astrophysics*, 501 (3):1185–1205, 2009.

D Robert Iskander, Michael J Collins, and Brett Davis. Optimal modeling of corneal surfaces with zernike polynomials. *IEEE Transactions on biomedical engineering*, 48 (1):87–95, 2001.

Preshanth Jagannathan, S Bhatnagar, W Brisken, and AR Taylor. Direction-dependent corrections in polarimetric radio imaging. ii. a-solver methodology: A low-order solver for the a-term of the a-projection algorithm. *The Astronomical Journal*, 155(1):3, 2018.

Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A*, 374(2065):20150202, 2016.

R Clark Jones. A new calculus for the treatment of optical systems. vii. properties of the n-matrices. *Josa*, 38(8):671–685, 1948.

S Kazemi and S Yatawatta. Robust radio interferometric calibration using the t-distribution. *Monthly Notices of the Royal Astronomical Society*, 435(1):597–605, 2013.

S Kazemi, S Yatawatta, S Zaroubi, P Lampropoulos, AG de Bruyn, LVE Koopmans, and J Noordam. Radio interferometric calibration using the sage algorithm. *Monthly Notices of the Royal Astronomical Society*, 414(2):1656–1666, 2011.

Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

RP Linfield. A modified hybrid mapping technique for vlbi data. *The Astronomical Journal*, 92:213–218, 1986.

Kaj Madsen, Hans Bruun Nielsen, and Ole Tingleff. Methods for non-linear least squares problems. *IMM DTU*, 2004.

Virendra N Mahajan. Zernike circle polynomials and optical aberrations of systems with circular pupils. *Applied optics*, 33(34):8121–8124, 1994.

John Makhoul. A fast cosine transform in one and two dimensions. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(1):27–34, 1980.

Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

Iván Martí-Vidal and JM Marcaide. Spurious source generation in mapping from noisy phase-self-calibrated data. *Astronomy & Astrophysics*, 480(1):289–295, 2008.

Modhurita Mitra, Sphesihle Makhathini, Griffin Foster, Oleg Smirnov, and Richard Perley. Incorporation of antenna primary beam patterns in radio-interferometric data reduction to produce wide-field, high-dynamic-range images. In *Electromagnetics in Advanced Applications (ICEAA), 2015 International Conference on*, pages 494–497. IEEE, 2015.

Niruj Mohan and David Rafferty. Pybdsf: Python blob detection and source finder. *Astrophysics Source Code Library*, 2015.

D Morris. Chromatism in radio telescopes due to blocking and feed scattering. *Astronomy and Astrophysics*, 67:221–228, 1978.

David D Mulcahy. *Nearby spiral galaxies at low frequencies.* PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, Cork, Ireland, 2014.

Ngoy Mutonkole and Dirk IL de Villiers. Characteristic basis function patterns method for reflector antenna calibration: An extension to multiple frequencies. In *Antennas and Propagation (EuCAP), 2015 9th European Conference on*, pages 1–5. IEEE, 2015.

P. J. Napier. The vla upgrade project. memo series 5. a larger subreflector for the vla, 1996.

Robert J Noll. Zernike polynomials and atmospheric turbulence. *JOsA*, 66(3):207–211, 1976.

Jan E Noordam. Lofar calibration challenges. In *SPIE Astronomical Telescopes+ Instrumentation*, pages 817–825. International Society for Optics and Photonics, 2004.

Jan E Noordam and Oleg M Smirnov. The meqtrees software system and its use for third-generation calibration of radio interferometers. *Astronomy & Astrophysics*, 524: A61, 2010.

AR Offringa, Benjamin McKinley, Natasha Hurley-Walker, FH Briggs, RB Wayth, DL Kaplan, ME Bell, Lu Feng, AR Neben, JD Hughes, *et al.* Wsclean: an implementation of a fast, generic wide-field imager for radio astronomy. *Monthly Notices of the Royal Astronomical Society*, 444(1):606–619, 2014.

Edward L O'Neill. *Introduction to statistical optics.* Courier Corporation, 2003.

Athanasios Papoulis. A new algorithm in spectral analysis and band-limited extrapolation. *IEEE Transactions on Circuits and systems*, 22(9):735–742, 1975.

Nathan Grier Parke. *Matrix optics.* PhD thesis, Massachusetts Institute of Technology, 1948.

Ajinkya H Patil, Sarod Yatawatta, Saleem Zaroubi, Léon VE Koopmans, AG de Bruyn, Vibor Jelić, Benedetta Ciardi, Ilian T Iliev, Maaijke Mevius, Vishambhar N Pandey, *et al.* Systematic biases in low-frequency radio interferometric data due to calibration: the lofar-eor case. *Monthly Notices of the Royal Astronomical Society*, 463(4):4317–4330, 2016.

R Perley. Jansky very large array primary beam characteristics. *EVLA Memo*, 195, 2016.

Richard A Perley, Frederic R Schwab, and Alan H Bridle. Synthesis imaging in radio astronomy. In *Synthesis Imaging in Radio Astronomy*, volume 6, 1989.

Attila Popping and Robert Braun. The standing wave phenomenon in radio telescopes-frequency modulation of the wsrt primary beam. *Astronomy & Astrophysics*, 479(3): 903–913, 2008.

Yahya Rahmat-Samii and Victor Galindo-Israel. Shaped reflector antenna analysis using the jacobi-bessel series. *IEEE Transactions on Antennas and propagation*, 28(4):425–435, 1980.

GROTE Reber. Cosmic static. *Proceedings of the IRE*, 36(10):1215–1218, 1948.

Alan W Rudge, K Milne, and A David Olver. *The handbook of antenna design*, volume 16. IET, 1982.

M Ryle, A Hewish, and J Shakeshaft. The synthesis of large radio telescopes by the use of radio interferometers. *IRE Transactions on Antennas and Propagation*, 7(5):120–124, 1959.

Stefano Salvini and Stefan J Wijnholds. Stefcal—an alternating direction implicit method for fast full polarization array calibration. In *2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS)*, pages 1–4. IEEE, 2014.

Jim Schwiegerling, John E Greivenkamp, and Joseph M Miller. Representation of videokeratoscopic height data with zernike polynomials. *JOSA A*, 12(10):2105–2113, 1995.

PF Scott, M Ryle, and A Hewish. First results of radio star observations using the method of aperture synthesis. *Monthly Notices of the Royal Astronomical Society*, 122 (1):95–111, 1961.

Oleg M. Smirnov. Revisiting the radio interferometer measurement equation. III. Addressing direction-dependent effects in 21 cm WSRT observations of 3C 147. *A&A*, 527:A108, March 2011a. doi: 10.1051/0004-6361/201116435.

Oleg M Smirnov. Revisiting the radio interferometer measurement equation-i. a full-sky jones formalism. *Astronomy & Astrophysics*, 527:A106, 2011b.

Oleg M Smirnov. Revisiting the radio interferometer measurement equation-iii. addressing direction-dependent effects in 21 cm wsrt observations of 3c 147. *Astronomy & Astrophysics*, 527:A108, 2011c.

C Tasse, B Hugo, M Mirmont, O Smirnov, M Atemkeng, L Bester, MJ Hardcastle, R Lakhoo, S Perkins, and T Shimwell. Faceting for direction-dependent spectral deconvolution. *Astronomy & Astrophysics*, 611:A87, 2018.

G. B. Taylor, C. L. Carilli, and R. A. Perley, editors. *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, 1999.

A Richard Thompson, James M Moran, and George W Swenson Jr. *Interferometry and synthesis in radio astronomy*. John Wiley & Sons, 2008.

Anthony Richard Thompson, James M Moran, George Warner Swenson, *et al. Interferometry and synthesis in radio astronomy*. Wiley New York et al., 1986.

M áP van Haarlem, MW Wise, AW Gunst, George Heald, JP McKean, JWT Hessels, AG De Bruyn, Ronald Nijboer, John Swinbank, Richard Fallows, *et al.* Lofar: The low-frequency array. *Astronomy & Astrophysics*, 556:A2, 2013.

SJ Wijnholds, TL Grobler, and OM Smirnov. Calibration artefacts in radio interferometry–ii. ghost patterns for irregular arrays. *Monthly Notices of the Royal Astronomical Society*, 457(3):2331–2354, 2016.

PN Wilkinson, J Conway, and J Biretta. Developments in continuum imaging. In *The Impact of VLBI on Astrophysics and Geophysics*, volume 129, pages 509–518, 1988.

Thomas L Wilson, Kristen Rohlfs, and Susanne Hüttemeister. *Tools of radio astronomy*, volume 5. Springer, 2009.

James C Wyant and Katherine Creath. Basic wavefront aberration theory for optical metrology. *Applied optics and optical engineering*, 11(s 29):2, 1992.

Sarod Yatawatta. Radio interferometric calibration using a riemannian manifold. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 3866–3870. IEEE, 2013.

André Young. *Improving the direction-dependent gain calibration of reflector antenna radio telescopes*. PhD thesis, Stellenbosch: Stellenbosch University, 2013.

André Young, Rob Maaskant, Marianna V Ivashina, Dirk IL de Villiers, and David Bruce Davidson. Accurate beam prediction through characteristic basis function patterns for the meerkat/ska radio telescope antenna. *IEEE Transactions on Antennas and Propagation*, 61(5):2466–2473, 2013.

Frits Zernike. Diffraction theory of the knife-edge test and its improved form, the phase-contrast method. *Monthly Notices of the Royal Astronomical Society*, 94:377–384, 1934.