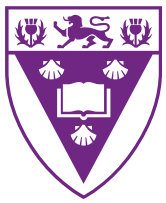


RHODES UNIVERSITY

Where leaders learn

Parametrised Gains for Direction-Dependent Calibration

Author:
Cyndie Russeeawon

Supervisor:
Prof. Oleg Smirnov

Co-supervisor:
Dr Landman Bester

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

in the

Centre for Radio Astronomy Techniques and Technologies
Department of Physics and Electronics

of

Rhodes University

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

February, 2021

Abstract

Calibration in radio interferometry describes the process of estimating and correcting for instrumental errors from data. Direction-Dependent (DD) calibration entails correcting for corruptions which vary across the sky. For small field of view observations, DD corruptions can be ignored but for wide field observations, it is crucial to account for them. Traditional maximum likelihood calibration is not necessarily efficient in low signal-to-noise ratio (SNR) scenarios and this can lead to overfitting. This can bias continuum subtraction and hence, restrict the spectral line studies. Since DD effects are expected to vary smoothly across the sky, the gains can be parametrised as a smooth function of the sky coordinates. Hence, we implement a solver where the atmosphere is modelled using a time-variant 2-dimensional phase screen with an arbitrary known frequency dependence. We assume arbitrary linear basis functions for the gains over the phase screen. The implemented solver is optimised using the diagonal approximation of the Hessian as shown in previous studies. We present a few simulations to illustrate the performance of the solver.

Declaration of Authorship

I, Cyndie Russeeawon, declare that this thesis titled, "Parametrised Gains for Direction-Dependent Calibration" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Cyndie Russeeawon

Date: February, 2021

Acknowledgements

I am very grateful for I received an opportunity to write this thesis. This has been an incredible journey for me. Here, I want to express my gratitude to all those who helped me get through this degree.

Starting from home, I am very blessed to have had the constant support of my parents and family. I appreciate all the love and patience my parents and siblings put into me. Mama, Papa, Didi, Girish and Gulshan are a treasure, I pray God gives to each and everyone. Then, I also want to thank and send some love to my grandmother, grandfather, cousins, uncles-aunties and especially Jaikishan.

Thank you Dr Nadeem Oozeer and Sydil for informing me about the RATT bursary. Dr Nadeem gave me a boost to consider entering the field. I would like to express my sincere gratitude to Dr D. Yannick Tangman, Dr Ravindra Boojhawon and Dr Nadeem for filling in for me as referees. I could not have gone through it without them and all my lecturers at the University of Mauritius, especially Prof. Vandna Jowaheer.

Thank you to Prof. Oleg Smirnov for having accepted me. I am very grateful to Ronel, Verushca and Zizipo for their assistance in clearing all my administrative bottlenecks. I am also indebted of Ridhima for she helped me with all my questions before coming here. I was lucky to have been admitted to Oakdene House. Special thank you to the warden Eric Ofei, the subbies Nadene and Nnamdi to have made me feel very comfortable. Coming to a new environment can be very intimidating. But, having friends like Lexy, Ulrich, Eric, Alex, James, Siyanda, Esther, Joel, Audrey, Robert, Gift, Noah, Josh, Kela, Marcel and Ianja made my life easier. I also need to add the Rock Family Church, Mama Deana, Tata Shaun, Mama Brenda, Mama Susan, Yolande and Papa Arthur to my Grahamstown family, for always giving me so much love. Special thank you to my friends back in Mauritius for always checking up on me, Neha, Chetan and Hoomeshwari.

They say it takes a village to raise a child. Here, I want to extend my gratitude to all my teachers. Without whom, the world would be still. Although it can sometimes be an unappreciated job, teachers always strive to give us the best. I am very blessed to have received the supervision of Prof. Oleg and Dr Landman Bester. They have always been very caring and patient to me. Whenever I did not understand something, they were always there to explain and make sure I am not stranded. From coding, write-up to presentations, there was no inconsistency in their presence and support. May God bless them in abundance! Thank you to my mentor, Assoc. Prof. Shazrene Mohammed, who continuously motivated me throughout this journey. Thank you to Prof. J. Jonas for inspiring me with the class of radio astronomy. Thank you to the RATT team for creating a supportive and fostering environment.

Thank you Ulrich for being there, whenever I had a bug in my code and I could not figure it out, you would take time out to help me find it. Thank you for proof-reading my thesis. Thank you for correcting me whenever I got any concept wrong. Thank you for everything. Also, thank you to Jonathan for proof-reading my thesis and assisting with all my questions. Last but not least, thank you to Lexy, my best friend, flatmate, colleague,

Room 48 shareholder and partner in crime. Thank you for your presence in my life and a friendship I shall cherish all my life.

Thank you God for helping me throughout this journey. Thank you to everyone who made this thesis possible: my supervisors, friends, family and my examiners for their valuable feedback. I hope this work adds value to the research being carried out in the field of radio interferometry. I also want to acknowledge the National Research Foundation for funding this project.

“Vakratunda Maha-Kaaya Surya-Kotti Samaprabha,
Nirvighnam Kuru Me Deva Sarva-Kaaryeshu Sarvadaa ||”

Contents

List of Figures

List of Tables

List of Recurring Abbreviations

EM	**E**lectro**m**agnetic
HI	Neutral Hydrogen
DDE	**D**irection-**D**ependent **E**ffect
DIE	**D**irection-**I**ndependent **E**ffect
SNR	**S**ignal-to-**N**oise **R**atio
FoV	**F**ield **o**f **V**iew
FWHM	**F**ull **W**idth at **H**alf-**M**aximum
PSF	**P**oint **S**pread **F**unction
RIME	**R**adio **I**nterferometer **M**easurement **E**quation
TEC	**T**otal **E**lectron **C**ontent
GN	**G**auss-**N**ewton
LM	**L**evenberg-**M**arquardt
NLLS	**N**on-**L**inear **L**east **S**quares
DoF	**D**egrees **o**f **F**reedom
KL	**K**arhunen-**L**oève
SPAM	**S**ource **P**eeling and **A**tmospheric **M**odelling
FBC	**F**ield-**B**ased **C**alibration
KAT-7	**K**aroo **A**rray **T**elescope
SKA	**S**quare **K**ilometre **A**rray
LOFAR	**Lo**w-**F**requency **Ar**ray
MWA	**M**urchison **W**idefield **A**rray
VLA	**V**ery **L**arge **A**rray
GMRT	**G**iant **M**etrewave **R**adio **T**elescope

List of Recurring Symbols

N_{A}	Number of antennas
N_{D}	Number of directions/ sources
N_{PAR}	Number of parameters per antenna
N^{t}_{INT}	Number of solution time intervals
N^{f}_{INT}	Number of solution frequency intervals
t_{INT}	Size of solution time interval
f_{INT}	Size of solution frequency interval
p, q, pq	Antenna and baseline indices
l_d, m_d	Spatial coordinates for source d
$V_{pqt\nu}$	Scalar visibility at time t and frequency ν
$\boldsymbol{X}_{dpq}$	Source coherency matrix
$\boldsymbol{G}_{p}$	DIE for antenna p
$\boldsymbol{E}_{dp}$	DDE for antenna p in the direction of source d
$\boldsymbol{\alpha}_{p}$	Parameter vector for antenna p
α_{pk}	k^{th} parameter for antenna p
h_{dk}	k^{th} polynomial term corresponding to source d
$\sum$	Summation
$\lvert(\cdot)\rvert$	Absolute value
$(\cdot)^{T}$	Transpose
$(\cdot)^{-1}$	Inverse
$(\cdot)^{H}$	Complex conjugate transpose
$\lvert\lvert\cdot\rvert\rvert_{F}$	Frobenius norm
$\boldsymbol{r}$	Residual vector
$\boldsymbol{J}$	Jacobian
$\mathrm{diag}(\cdot)$	Diagonal entries
$\mathbf{H}$	Hessian
δ	Dirac delta
$\mathrm{Re}(\cdot)$	Real component
$\mathrm{Im}(\cdot)$	Imaginary component
σ_{α}	Standard deviation for generation of $\boldsymbol{\alpha}$
σ_{N}	Standard deviation for noise realisation

Chapter 1

Introduction

Astronomy can be defined as the study of the objects that lie beyond our planet and the processes by which these objects interact with each other. Previously, many ancient civilisations around the world, including the Greeks and the Romans, tried to understand the lights in the night sky, that is, at optical wavelengths, and fit them into their view of the world. In 1932, a radio engineer named Karl Jansky working at Bell Laboratories in New Jersey made an accidental encounter with radio waves originating from our Milky Way galaxy itself (Jansky 1933). This discovery made it to the headlines of the New York Times. It motivated an astronomer called Grote Reber to map the Milky Way at 160 MHz using the parabolic radio antenna he had built in his backyard (Reber 1940). This led to the birth of radio astronomy, which can be described as the branch of astronomy dealing with cosmic radiation observable in the radio regime. The radio spectrum is wide, ranging from approximately 3 kHz to 1 THz. But, owing to the opacity of the atmosphere, ground-based radio astronomy is limited to a smaller range. The ionosphere absorbs or fully reflects the radio waves below 10 MHz, whereas the water vapour in the troposphere absorbs the radio radiation leading to the higher frequency cut-off at around 300 GHz (Thompson et al. 2017, § 1.2.1). Radio astronomy has an advantage over optical astronomy in the sense that radio waves are not easily scattered or absorbed by interstellar dust and gas as compared to the light waves.

This chapter aims to introduce the field of radio astronomy and the problem addressed in the rest of this work. In § 1.1, we list some of the celestial sources of radio radiation of great interest to astronomers. In order to characterise the radiation, a set of standard metrics is used by astronomers (some of which are found in § 1.2). We also introduce the associated measuring device and discuss how the device can be improved for better science towards the end of the same section. Following this, we introduce some basic concepts in radio interferometry and present the response of an ideal interferometer in § 1.3. The specific set of corruptions considered in this work is discussed in § 1.4. We further motivate our approach with the discussion in § 1.5 and finally present the layout for the rest of the thesis at the end of the chapter.

As for notations, we adhere to the following throughout this thesis. Upper and lower boldface cases denote matrices and vectors, respectively, or otherwise scalars. $(*)$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ stand for the convolution, transpose, complex conjugate transpose and inverse operators respectively.

1.1 Astrophysical sources of radio emissions

Radio waves consist of photons with the lowest energies, as compared to other forms of electromagnetic (EM) radiation. Reber's measurements showed that radio emission arises from a combination of thermal and non-thermal mechanisms. In addition, radio emission can be classified further into the continuum and line emissions. Continuum emission is the broadband radiation emitted over a very wide and continuous range of wavelengths in the EM spectrum by celestial objects. In contrast to continuum emission, line emissions occur at only one specific wavelength, and can therefore only be observed at this specific wavelength, or the redshifted version thereof. We further elucidate the distinction between line and continuum emissions in the sections below. This section is summarised mainly using the following: Fomalont and Perley (1999), Thompson et al. (2017), and Wilson et al. (2009).

1.1.1 Continuum emission

Continuum emission can be explained using thermal and non-thermal processes.

Thermal emission

Thermal emission, also known as black body radiation, is radiation emitted by black bodies over a wide area of the EM spectrum. A black body can be defined as a source that absorbs or emits "all" energy it receives. Based on Fig. 1.1, we may interpret that the intensity distribution of black body radiation is dependent on its frequency of emission, ν, and its temperature, T. This distribution can be described by the Planck's radiation law as,

$$I_\nu = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT}} - 1}, \tag{1.1}$$

where I_ν is defined as the spectral brightness of the black body (elaborated in § 1.2) (Condon and Ransom 2016). h, c and k denote the Planck's constant, the speed of light in vacuum and Boltzmann's constant respectively. Certain celestial objects (stars are a good example) can be considered as (slightly imperfect) black bodies. However, from Fig. 1.1, we see that these objects would have to be very hot to be observable in the radio regime, especially if they are outside our solar system. Thus, if thermal emission was the only mechanism by which celestial objects emit radiation, the radio sky would be largely empty. For these reasons, for a long time, astronomers did not realise the significant role that radio astronomy has to play in understanding the night sky. Jansky's accidental discovery indicated that non-thermal mechanisms must also be responsible for radio emission.

Non-thermal emission

We now know that synchrotron radiation is the main non-thermal emission mechanism in the radio regime. When charged particles move at very high speeds as close to that of light, and these experience an external magnetic force, synchrotron radiation is produced. The magnetic field exerts force in a direction perpendicular to their velocity which causes them

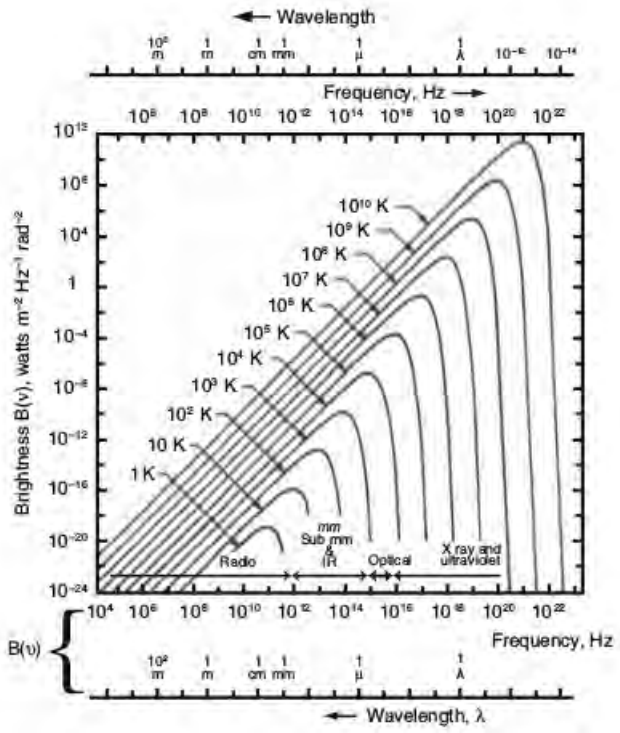


FIGURE 1.1: The radiation intensity distribution for a black body in equilibrium. The radiation intensity for such a body is particularly low at low frequencies for all temperatures. Credits: Wilson et al. (2009)

to move along a helical path (similar to a spiral staircase). Particles lose kinetic energy during this process and emit radiation to conserve energy. The energy loss depends on the associated absorption process. In order for the emitting electrons to have a well-defined kinetic temperature, synchrotron self-absorption occurs to keep the brightness temperature of the radiation below this kinetic temperature. Therefore, the brightness temperatures are not arbitrarily large at lower frequencies. As the velocity of the spiralling particle changes, so does the energy of the emitted radiation resulting in photons with a range of frequencies, and a peak at a critical frequency, ν_0. Its synchrotron spectrum possesses a logarithmic slope of $\frac{1}{3}$ at low frequencies, which eventually reaches a *broad* peak at around ν_0 and a steep negative slope towards higher frequencies (Mesinger 2019). It is given that ν_0 is directly proportional to the square of the energy of the charged particle and the magnitude of the perpendicular component of the magnetic field. So, as the particle travels around the magnetic field, it loses more and more energy. Additionally, the rate at which an electron loses energy to the synchrotron radiation is also proportional to the square of this energy, so electrons possessing higher energies are expected to deplete more rapidly. As a consequence, the particle follows a more compact spiral path, and the output ν_0 decreases as the synchrotron losses steepen source spectra at higher frequencies in the long run. The observed synchrotron spectrum is a result of the sum of the emission spectra of the individual particles. Since the energy distribution of the particles is a power distribution, the resulting spectrum appears to be one too. Besides synchrotron radiation, some other relevant non-thermal processes consist of Compton, inverse Compton scattering and thick-target Bremsstrahlung, which possess

different signature spectra (refer to Condon and Ransom 2016).

It is possible to distinguish between different radio continuum emission mechanisms. One way to do this is to make use of the frequency profile of the observed radiation. Many astrophysical sources' radiation scales (or is assumed to scale) as a power law in frequency i.e. $I_\nu \propto \nu^\alpha$ where α is known as the spectral index (Burke and Graham-Smith 2010; Thompson et al. 2017). It is known, for example, that we usually observe an almost flat ($\alpha \approx 0$) spectrum with thermal emission if it is optically thin[1], or a positive spectral index if it is optically thick. Hot interstellar ionised hydrogen and planetary nebulae are often sources of thermal emission. On the other hand, a negative spectral index ($\lesssim -0.5$) is typical for synchrotron emission (for example, Galactic supernova remnants with $\alpha \approx -0.5$). Thus, by measuring spectral indices, it is sometimes possible to establish the mechanism by which a source emits. This, in turn, aids our understanding of the source and hence the physical processes that are responsible for it.

While we will not delve further into the technicalities involved in determining accurate spectral indices, the above discussion motivates why it is important to try and preserve the true frequency profile of a source as much as possible. As we discuss later on, the calibration procedure can drastically alter the spectrum of a source if it is not done carefully, and this is particularly relevant to spectral line observations.

1.1.2 Spectral line observations

Spectral lines are thin and sharp emission or absorption features appearing in the spectra of gaseous and ionised sources. They occur when atoms or molecules transition from one energy state to another, often as a result of electrons moving between orbitals of different energies. For this to happen, the electron either has to absorb or emit a photon carrying energy equivalent to the difference in energy between the two states. We say that the EM radiation is quantised into photons whose energy, E is proportional to the frequency i.e. $E = \hbar\nu$. Some examples of radio spectral lines include recombination lines of ionised hydrogen, rotational lines of polar molecules such as carbon monoxide and the $\lambda = 21$ cm line of interstellar HI.

21 cm line of neutral hydrogen

Hydrogen is an element found in the highest concentration in the Universe. Most of it is present in the form of neutral hydrogen atoms, or HI, each of which consists of a single proton and an electron with no neutrons. Occasionally, the orbiting electron will undergo a spin-flip (hyperfine transition) and emit a photon with a wavelength of ≈ 21 cm. This results in a spectral line signature known as the hyperfine line. This scenario is depicted in Fig. 1.2. As shown there, we end up with two energy levels due to the magnetic interaction between the quantised electron and proton spins. When the relative spins switch directions from parallel to antiparallel, a photon with frequency $\nu = 1420.405$ MHz is emitted. While the transition rate is very low ($\approx 2.9 \times 10^{-15}$ per second), the sheer abundance of hydrogen

[1] Optical depth of a medium scales with the number of interactions (and scatterings) a photon will have before it leaves the medium.

in the Universe means that these lines will be present, albeit very faint (typically tens of mK). This makes them relatively difficult to detect. Nevertheless, spectral lines are narrow, so we know (almost) exactly at which frequencies to look for them. We do, however, have to account for the fact that the Universe is expanding.

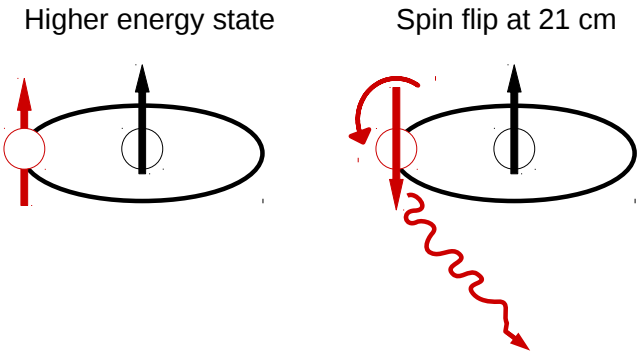


FIGURE 1.2: The 21 cm line naturally occurs due to the change in energy states in neutral hydrogen atoms.

The expansion of the Universe results in a Doppler shift, also known as the cosmological redshift, which changes the observed frequency of incoming radiation, as compared to when it was emitted. Redshift is a measure of how EM radiation from an object experiences a decrease in frequency as the object moves away from the observer. It is defined as

$$z = \frac{f_{emit} - f_{obs}}{f_{obs}}, \tag{1.2}$$

where f_{emit} is the frequency of the wave at the emitter, and f_{obs} is the observed frequency. Thus, the observed frequency scales with redshift as

$$f_{obs} = \frac{f_{emit}}{1+z}. \tag{1.3}$$

However, it is not just the expansion of the Universe that affects the observed frequency, and there can be a contribution from multiple kinematic effects. For example, when a galaxy rotates, and we are looking at it edge-on, the velocity of the HI on one side of the galaxy is different from the velocity on the other side. This results in the HI emission line being split into two peaks, one blueshifted and one redshifted (blueshift is the opposite of redshift). Thus, we can infer some kinematic information about the HI in the galaxy by measuring how the emission line has been distorted. This type of frequency profile is sometimes referred to as the double-horn profile. Fig. 1.3 is an example of the integrated HI spectrum of the spiral galaxy UGC 11707. The shape of this profile varies across different types of galaxies. The profile can be modelled using parameters such as the integral of the HI line (area under the curve) and line width, which give the total HI mass and projected circular velocity of the galaxy respectively. Then, with respect to Fig. 1.3, a generally wide and flat trough is observed with most disc galaxies and Gaussian profiles with dwarf galaxies. Westmeier et al.

(2013) (used in § 4.1) and Stewart et al. (2014) are examples of parametrised models used to fit a wide range of HI profiles.

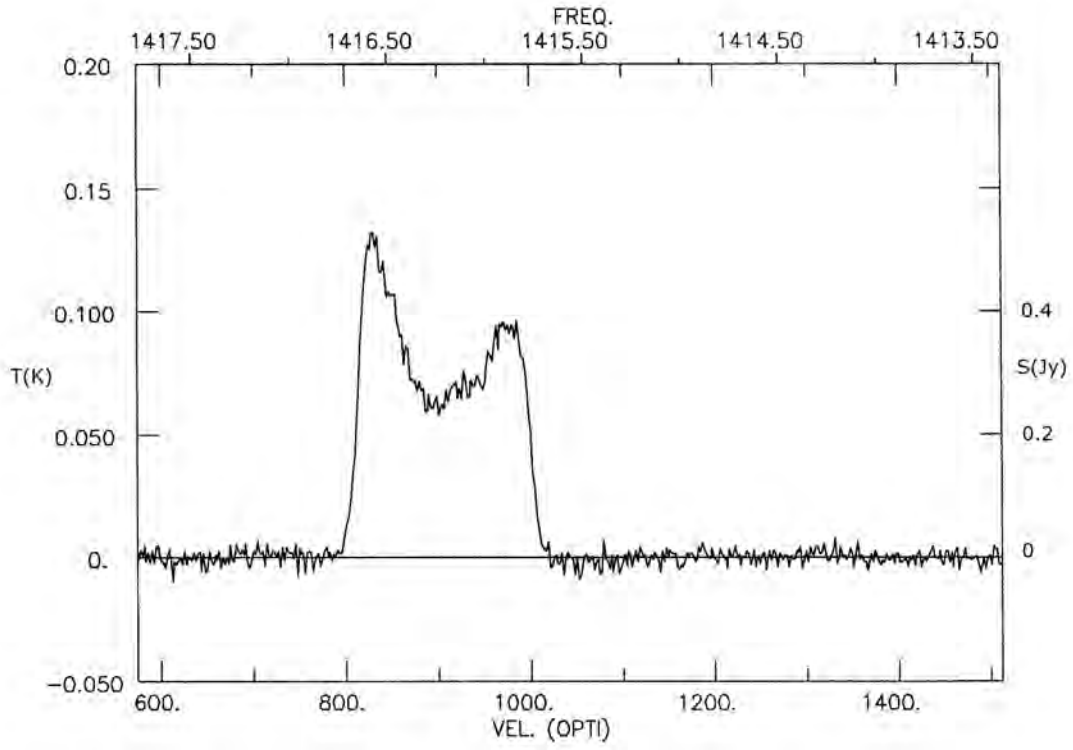


FIGURE 1.3: Spectral profile for the galaxy UGC 11707. The velocity and frequency are given in units of km s^{-1} and MHz respectively. The observed line centre frequency is 1416.2 MHz. The sharp flanks are typical of a spiral galaxy. Image from Haynes et al. (1998).

The 1420 MHz HI line is a useful tool in studying gas in galaxies, as well as tracking the broad distribution of galaxies in the Universe (known as *intensity mapping*). We will not pursue these topics further in this thesis. The important point to note is that it is crucial to try and preserve the frequency profiles of the sources during calibration. This is a topic that we will return to throughout the thesis. Next, we give a quick overview of another major science driver for radio astronomy viz. the Epoch of Reionisation (EoR).

Epoch of Reionisation

The EoR is the period in which matter in the intergalactic medium was changed from being neutral to ionised. Thus, the EoR is an essential period to astronomers as it contains a wealth of information about the structure formation of the early Universe. The prevalence of HI in the early Universe makes the redshifted 21 cm line a promising probe for the EoR (Heald et al. 2018; Thompson et al. 2017). The EoR is expected to have occurred between redshifts, $z = 6$ and $z = 12$. From Eqn. (1.2), we see that the observed frequencies for the 21 cm line corresponding to these redshifts are approximately 240 MHz and 120 MHz, respectively.

Some of the most prominent low frequency radio astronomy facilities include the VLA (Very Large Array, Perley et al. 2011), GMRT (Giant Metrewave Radio Telescope, Swarup

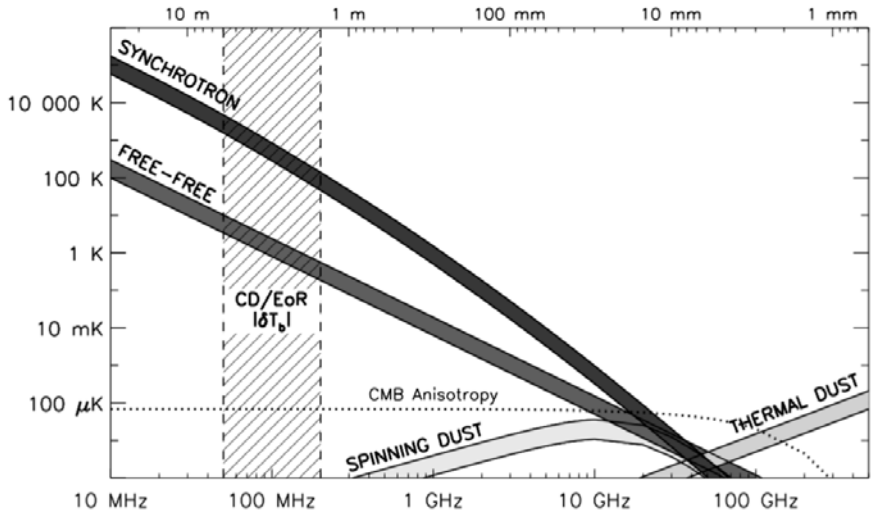


FIGURE 1.4: The main and bright Galactic foregrounds given as a function of frequency in total intensity (synchrotron emission, free-free emission from diffuse ionised gas and thermal dust emission). These results come from the EoR and the Cosmic Dawn (CD) experiments, where the CD is the era prior to the EoR (occurred at around $14 < z < 30$) describing the formation of the first galaxies. The faint shaded area corresponds to the range of redshifts where the EoR and the CD can be observed with the 21 cm line. Within this shaded box, the unshaded area ($|\delta \mathrm{T_b}|$) describes the range of brightness temperatures recorded in CD/ EoR experiments (about 10 mK $-$0.1 K). Therefore, the 21 cm line is intrinsically faint, and lies buried behind the much brighter (tens of K) foregrounds coming from Galactic continuum radio sources (and extragalactic ones, though not included here) (darker grey areas). The image is taken from Mesinger (2019), where the reader can also find further details about the rest of the plot.

1991), PAPER (Precision Array to Probe the Epoch of Reionisation, Parsons et al. 2010), HERA (Hydrogen Epoch of Reionisation, DeBoer et al. 2017), MWA (Murchison Widefield Array, Tingay et al. 2013) and LOFAR (Low-Frequency Array, van Haarlem et al. 2013). Some of these are multi-purpose telescopes, while others are specifically engineered to detect and characterise the EoR signal. As illustrated in Fig. 1.4, in terms of the total power and fluctuations, the Galactic synchrotron emission exceeds the faint cosmological 21 cm signal in orders of magnitude of $3-4$ and $2-3$ respectively. For example, as summarised from the redshifted 21 cm experiments shown in Fig. 6.1 in Mesinger (2019), the brightness temperatures at 150 MHz for the Galactic, extragalactic foregrounds and cosmological 21 cm signal are expected to be 1 K, 0.45 K and 10 mK respectively. Hence, successful detection of the EoR signal relies heavily on being able to remove the unwanted foregrounds (also sometimes referred to as continuum subtraction). We should keep in mind that the signal of the science target, being relatively faint compared to the foregrounds, is often left unmodelled during calibration. In what follows, we illustrate that careful calibration is required to retain the frequency profile of such a faint unmodelled signal of interest. First, however, we introduce the notation and terminology required to understand how such a signal is measured using an interferometer.

1.2 Fundamental quantities

An antenna is a device that translates an EM field into a voltage. With some effort, these voltages can be used to extract information about the radio sky. Of course, to achieve this, we need to know how they are related to the signal of interest. To this end, in this section, we define some important notions and quantities that are often encountered in radio astronomy. We will start with some of the fundamental observables and outline how they are actually detected by an antenna (refer to Condon and Ransom 2016 and Thompson et al. 2017 for a detailed description).

1.2.1 Brightness and flux

Consider the situation depicted below in Fig. 1.5. We say that radiation coming from a patch of sky with infinitesimal solid angle $\mathrm{d}\Omega$ intersects an arbitrary infinitesimal surface area $\mathrm{d}\sigma$ with angle θ if θ is the angle between the normal to $\mathrm{d}\sigma$ and the direction to the centre of $\mathrm{d}\Omega$. In this case, the infinitesimal power $\mathrm{d}P$ experienced at $\mathrm{d}\sigma$ can be expressed as

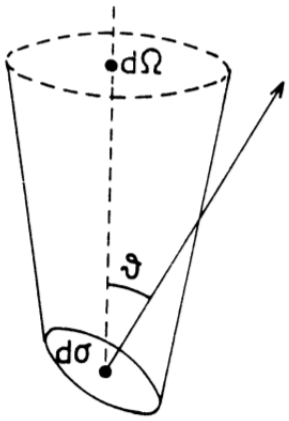


FIGURE 1.5: Relationship between the infinitesimal solid angle $\mathrm{d}\Omega$, infinitesimal surface $\mathrm{d}\sigma$, and θ, angle between the normal to $\mathrm{d}\sigma$ and centre of $\mathrm{d}\Omega$. Diagram from Wilson et al. (2009).

$$\mathrm{d}P = I_\nu \cos\theta \, \mathrm{d}\sigma \mathrm{d}\nu, \tag{1.4}$$

where $\mathrm{d}\nu$ is an infinitesimal bandwidth and I_ν is known as the specific intensity, brightness or spectral brightness (recall Eqn. (1.1)).

Note that, in the absence of distorting effects (e.g. diffraction or extinction), the brightness of a source is independent of the distance to the source. This is an intrinsic property of the source. On the other hand, the flux density, S_ν, passing through $\mathrm{d}\sigma$ coming from a source that occupies an area Ω_d can be obtained by integrating Eqn. (1.4) over Ω_d i.e.

$$S_\nu = \int_{\Omega_d} \frac{\mathrm{d}P}{\mathrm{d}\sigma \mathrm{d}\nu} \mathrm{d}\Omega = \int_{\Omega_d} I_\nu(\theta) \cos\theta \, \mathrm{d}\Omega, \tag{1.5}$$

which therefore has units of W $m^{-2}Hz^{-1}$. However, these units are not well adapted to astronomy because typical flux densities of astronomical sources are usually very small. For this reason, astronomers make use of the flux density unit called the Jansky (Jy) which is defined as

$$1\,\mathrm{Jy} = 10^{-26}\,\mathrm{W\ m^{-2}Hz^{-1}}. \tag{1.6}$$

Note that, since the angular size of the source depends on the distance from the receiver, flux density does have a dependency on distance. It is not an intrinsic property of the source.

1.2.2 Temperature and sensitivity

Brightness temperature is another important quantity that is often encountered in radio astronomy. It is defined as the temperature of a black body (in thermal equilibrium) that would result in the specific intensity at a particular frequency. From Eqn. (1.1), we can deduce that, in the Rayleigh-Jeans regime (i.e. $h\nu \ll kT$), brightness temperature T_{src} can be related to specific intensity via the relation

$$I_\nu = \frac{2kT\nu^2}{c^2}. \tag{1.7}$$

Even though the relation Eqn. (1.1) refers to thermal radiation, astronomers also use the term temperature to refer to non-thermal radiation. In this case, we can use Eqn. (1.7) to convert between brightness temperature and specific intensity. However, note that brightness temperature is only independent of frequency for a perfect black body. For this reason, brightness temperature is often quoted at a specific frequency.

This tendency to think about brightness in terms of temperature is also useful when considering how a source interacts with a receiver. Any real receiver contains electronic components that contribute thermal noise to the system. We say that the receiver has an associated temperature denoted T_{rec}. In addition to the thermal noise, each antenna also receives radiation from unwanted background sources (e.g. the sun, our own galaxy or nearby sources). Thus, we can also associate a temperature to the sky, denoted T_{sky}. The system temperature, T_{sys}, is defined as

$$T_{\mathrm{sys}} = T_{\mathrm{sky}} + T_{\mathrm{rec}}. \tag{1.8}$$

In practice, the system temperature is usually quite a bit larger than the brightness temperature of a source. Fortunately, many of the processes that contribute to the system temperature have a random nature, and hence the system temperature can be beaten down by averaging a large number of samples. Typically, the effective temperature decreases as the square root of the number of samples. Now, since the number of samples is proportional to the bandwidth, $\Delta\nu$, multiplied by the integration time, τ, we have an effective noise temperature of

$$T_{\mathrm{rms}} = \frac{T_{\mathrm{sys}}}{\sqrt{\Delta\nu\tau}}. \tag{1.9}$$

For a source to be detectable, its brightness temperature needs to be significantly larger than the noise temperature. In this sense, radio astronomers like to talk about the signal-to-noise

ratio (SNR) ($\frac{\mathrm{S}}{\mathrm{N}}$) which is defined by the ideal radiometer equation

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{T_{\mathrm{src}}}{T_{\mathrm{sys}}}\sqrt{\triangle\nu\tau}, \tag{1.10}$$

which in some sense defines the sensitivity of the antenna to astronomical sources of interest.

We can also think of antenna sensitivity in terms of an effective flux density by interpreting Eqn. (1.8) using Eqn. (1.7). Doing so we can define a quantity called the *system equivalent flux density* (SEFD) as the flux density equivalent of T_{sys}. The system temperature therefore gives rise to a root mean square of flux variations, $S_{\nu,\mathrm{rms}}$ according to

$$S_{\nu,\mathrm{rms}} = \frac{\mathrm{SEFD}}{\sqrt{\triangle\nu\tau}}. \tag{1.11}$$

From antenna theory, we also know that the SEFD is given by

$$\mathrm{SEFD} = \frac{2kT_{\mathrm{sys}}}{A_e}, \tag{1.12}$$

where A_e is the effective collecting area (or aperture) of the antenna. This shows that antenna sensitivity can be increased either by increasing the effective collecting area or by increasing the number of samples that we average over (i.e. integration time and bandwidth). However, sensitivity is not the only factor to consider when thinking about the quality of a telescope. The resolving power of the telescope is also a major consideration.

1.2.3 Resolution

The angular resolution of a telescope is defined as the minimum angular distance, θ, between distinguishable objects in the sky. The Rayleigh criterion states that two point sources are considered discernible when the centre of the diffraction pattern of one coincides with the first minimum of the diffraction pattern of the other. For a circular aperture of diameter D, the diffraction pattern is given by an Airy disk, and the Rayleigh criterion corresponds to the angle

$$\theta = 1.22\frac{\lambda}{D}, \tag{1.13}$$

where λ is the observing wavelength. Therefore, if the angular distance between 2 objects is less than θ, they will be observed as a single object, and thus, not resolved. Full Width at Half-Maximum (FWHM) is defined as the width of the beam at the point where the power received from a point source is half its peak value, and the FWHM of the main beam is often used to determine the resolution of the telescope. It is directly proportional to $\frac{\lambda}{D}$.

Astronomers often want θ to be as small as possible since that allows them to distinguish between different sources (or source features) more accurately. Optical astronomers have a bit of an advantage over radio astronomers in this regard since optical wavelengths are many orders of magnitude shorter than radio wavelengths. This implies that for the same size aperture, optical telescopes possess more resolving power than radio telescopes.

In the quest for higher resolution, radio astronomers have constructed larger and larger radio telescopes such as the Five-hundred-metre Aperture Spherical Telescope (FAST) in China (Di and Zhichen 2016). However, there is a practical limit imposed on the maximum size of single dish telescopes due to, for example, the strength of the materials used to build the telescope or the financial costs involved. At the extreme low radio frequency end, attaining sufficient sensitivity and resolution for certain science cases would require km-sized telescopes. Obviously, this poses serious engineering challenges. Fortunately, this problem can be solved using an interferometer.

1.3 The ideal radio interferometer

Interferometry represents a set of techniques where EM waves are superimposed to form interference fringes. In radio interferometry, we describe the contrast of these fringe patterns using the term *visibility*. Visibilities are complex quantities that are formed by correlating voltages measured between antenna pairs. In this section, we describe how the signal measured by an ideal interferometer essentially corresponds to the Fourier transform of the sky brightness distribution. As we discuss below, the relationship between the brightness distribution and the visibilities is only a Fourier transform under certain limiting conditions. Nevertheless, we can use this fact to develop some intuition for the data produced by an interferometer. For example, as with the Fourier transform, visibility amplitudes encode the observed source's shape while their phases encode the source position. We will start with the simplest possible radio interferometer viz. the two-element interferometer.

1.3.1 Two-element interferometer

A two-element interferometer consists of two antennas pointing in the same direction. Since, from the perspective of an Earth-bound observer, the night sky seems to rotate above us, the antennas are designed to be steerable. This allows them to track a fixed position in the sky which we call the phase direction or phase centre (the reason for which will be clarified shortly). Consider the scenario depicted in Fig. 1.6. Let $\vec{d}$ be a unit vector towards the phase centre and label the two antennas as p and q. Let us denote the position vectors of antenna p and q as $\vec{r}_p$ and $\vec{r}_q$, respectively. Then, we call the connecting vector between the two antennas a baseline vector and denote it as $\vec{b}$. Furthermore, it is assumed that sources of radiation are sufficiently far away so that, for all intents and purposes, the radiation we receive from them arrives as a plane wave. Clearly, wavefronts coming from direction $\vec{d}$ reach antenna q before antenna p so antenna q responds to the wavefronts before antenna p does. Then, from the geometry, the time delay (referred to as the geometric delay and denoted τ_g) is given by

$$\tau_g = \frac{\vec{b} \cdot \vec{d}}{c}, \tag{1.14}$$

where c is the speed of light in a vacuum. This basic setup lays the foundation for much of radio interferometry which, essentially, seeks to find a way to relate the voltages measured between antenna pairs to the intrinsic properties of the source.

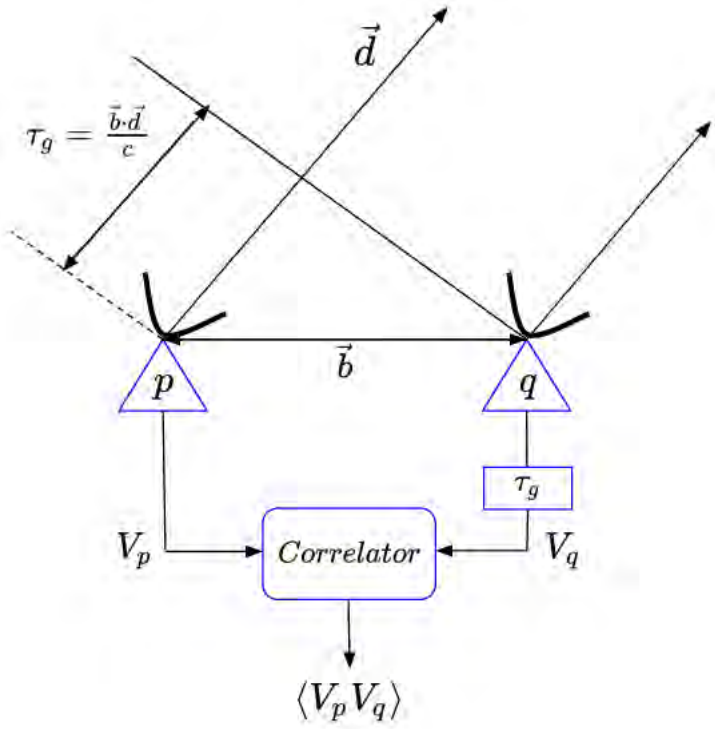


FIGURE 1.6: A schematic diagram of a simple two-element interferometer.

Since the signal we are interested in is an EM wave, and, for all intents and purposes, we are receiving it as a plane wave, its electric field obeys the wave equation for plane waves. Thus, if the electronic system can be assumed to be performing in the linear regime, the voltage at antenna p, v_p, can be parametrised as[2]

$$v_p(\vec{r_p}, \nu, t_p) = \epsilon_p(\vec{d}, \nu, t_p)e^{-iwt_p}, \tag{1.15}$$

where ϵ_p is the amplitude of the signal, t_p is the time at which the wavefront arrives at antenna p, and w stands for the angular frequency of the plane wave. By correlating voltages induced at antennas p and q, we obtain the visibility given by

$$v_{pq} = \langle v_p v_q^* \rangle = \langle \epsilon_p \epsilon_q \rangle e^{-iw\tau_g}, \tag{1.16}$$

where the angle brackets denote averaging over time, and $t_q = t_p - \tau_g$. Given the feeds are ideal (and no prior signal corruption along propagation path), and the source amplitude does not change with time, Eqn. (1.16) should result into

$$v_{pq} = \epsilon^2 e^{-iw\tau_g}, \tag{1.17}$$

with $\epsilon = \epsilon_p = \epsilon_q$. In order to establish a relationship between the visibility function and the intrinsic properties of the source, it is then convenient to make use of a coordinate system in which to express $\vec{b}$ and $\vec{d}$. Let us consider a 3D Cartesian coordinate system

[2]Here we consider only the highly idealised scalar problem. The more realistic polarised problem is addressed in § 2.

relative to the standard equatorial coordinates (H, δ) which denote hour angle and declination respectively (refer to § 4 and Fig. 4.1 of Thompson et al. 2017). Then, let $\vec{b} = (b_x, b_y, b_z)$ be the baseline vector in such a coordinate system, and the interferometer be pointing in the reference direction $\vec{d_0} = (H_0, \delta_0)$. This system can be written using (u, v, w) coordinates expressed as

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \sin H_0 & \cos H_0 & 0 \\ -\sin\delta_0 \cos H_0 & \sin\delta_0 \sin H_0 & \cos\delta_0 \\ \cos\delta_0 \cos H_0 & -\cos\delta_0 \sin H_0 & \sin\delta_0 \end{pmatrix} \begin{pmatrix} b_x/\lambda \\ b_y/\lambda \\ b_z/\lambda \end{pmatrix}, \tag{1.18}$$

where u and v span a plane orthogonal to the reference direction $\vec{d_0}$ with which w is aligned. Now, for an arbitrary position $\vec{d}$ on the celestial sphere relative to $\vec{d_0}$, we end up with the corresponding direction cosines $(l, m, n = \sqrt{1 - l^2 - m^2})$ of (u, v, w), $\vec{d} - \vec{d_0} = (l, m, (n-1))$. Then, by using the relative coordinates of some pair of antennas $\vec{b}_{pq} = (u_{pq}, v_{pq}, w_{pq})$, and a point in the sky, Eqn. (1.17) can be rewritten as,

$$v(u_{pq}, v_{pq}, w_{pq}) = Ie^{-2\pi i \frac{\nu}{c}(u_{pq}l + v_{pq}m + w_{pq}(n-1))}, \tag{1.19}$$

where $I = \epsilon^2$ stands for the total intensity of the wave. In the next section, we discuss Eqn. (1.19) in the context of multiple baselines and sources.

1.3.2 The van Cittert-Zernike theorem

The idealised discussion above pertains to radiation coming from a single direction. Loosely speaking, and under certain mild assumptions (see Thompson et al. 2017 for example), we can account for radiation coming from all directions simply by integrating over the entire sky. In this case, the visibilities measured between antenna pairs with relative positions $(u_{pq}, v_{pq}, w_{pq}) = (u_p - u_q, v_p - v_q, w_p - w_q)$ are given by

$$V(u, v, w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(l, m)I(l, m)e^{-2\pi i(ul + vm + w(n-1))}\frac{dldm}{n}, \tag{1.20}$$

where $I(l, m)$ is defined as the true sky brightness distribution, $A(l, m)$ is the combined antenna sensitivity pattern and we have dropped the subscripts p and q for notational simplicity. Now, the van Cittert-Zernike theorem states that interferometry samples the Fourier components of the sky brightness (Thompson et al. 2017). This is only really true for a coplanar array (i.e. $w = 0$) or if we are looking at very small area of the sky so that $n \approx 1$. In this regime, it is possible to get an image of the apparent sky (i.e. the sky attenuated by the antenna sensitivity pattern $\tilde{I}(l, m) = A(l, m)I(l, m)$) by applying a Fourier transform to the visibilities

$$\tilde{I}(l, m) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} V(u, v)e^{2\pi i(ul + vm)}dudv. \tag{1.21}$$

Unfortunately, since there are only a finite number of antennas, we never really have access to the continuous visibility function which means that an exact inversion is not possible.

The idea behind *aperture synthesis*, as developed by Sir Martin Ryle and colleagues ("The Synthesis of Large Radio Telescopes"; Ryle et al. 1960), is to use multiple small aperture antennas to sample the visibilities at enough (u, v) locations to synthesise a much larger aperture antenna of size (u_{max}, v_{max}) corresponding to the longest baseline coordinates. This obviously increases the achievable resolution, but it also introduces holes in the uv-coverage. Note that for N_{A} antennas, we get a maximum of $\frac{N_{\mathrm{A}}(N_{\mathrm{A}}-1)}{2}$ unique baselines. Thus, considering that N_{A} is necessarily finite, the measurable part of $V(u, v)$ is limited by N_{A} and the Earth-sky geometry. Nevertheless, we use the Earth rotation to fill in the (u, v) plane over time (called uv-tracks). Furthermore, since uv are in units of wavelengths, observations at multiple frequencies can be used to fill in the uv plane radially.

Defining a sampling function $S(u, v)$ such that $S(u, v) = 1$ where we have the measurements and zero otherwise, Eqn. (1.21) can be modified to

$$\begin{aligned} I_d(l, m) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(u, v)V(u, v)e^{2\pi i(ul+vm)}dudv, \\ &= \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(u, v)e^{2\pi i(ul+vm)}dudv\right) * \tilde{I}(l, m), \end{aligned} \tag{1.22}$$

where $I_d(l, m)$ is the reconstructed sky image. From Eqn. (1.22), the expression,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(u, v)e^{2\pi i(ul+vm)}dudv, \tag{1.23}$$

stands for the Point Spread Function (PSF) and it gives the interferometer response in the presence of a 1 Jy point source at the phase centre. The PSF and $I_d(l, m)$ are called the *dirty beam* and *dirty image* respectively. The dirty image is not representative of the true flux in the sky. To correct for artefacts introduced by incomplete sampling of the uv-plane, we need to use deconvolution algorithms. The most common such algorithm in radio astronomy is `CLEAN` (Högbom 1974). A detailed discussion of how `CLEAN` works is beyond the scope of this thesis. The basic idea behind the `CLEAN` algorithm is to locate the brightest peak in the residual image (which is equal to the dirty image at the outset), centre the PSF at this location and subtract some fraction of it pixel-wise from the image. At the same time, we keep track of the components that have been found by adding the same fraction of the peak to the model image, at the location of the maximum. This process is repeated until some predetermined convergence criteria are met (e.g. until the maximum in the residual image has decreased to some threshold or a maximum number of iterations has been reached). In our experiments in § 4, we will use the implementation of `CLEAN` in the `WSClean` software package (Offringa et al. 2014)[3].

The relation in Eqn. (1.20) is only valid for an ideal unpolarised interferometer. A more realistic model has to account for polarisation and the fact that the signal gets corrupted as it encounters certain obstacles along its path or propagation (for example free electrons, magnetic fields and imperfect receivers). These corruptions can be described using complex-valued 2×2 Jones matrices, and they need to be solved and corrected for or, in other words,

[3] `https://gitlab.com/aroffringa/wsclean`

the data need to be calibrated. More sophisticated versions of Eqn. (1.20) which can be used for calibration will be discussed in detail in § 2.1. For now, we will simply note that the propagation effects can be loosely divided into Direction-Independent Effects (DIEs) and Direction-Dependent Effects (DDEs). DIEs are effects which are constant across the sky and thus, DIEs are the same for each source. An example would be receiver gains. On the other hand, DDEs are effects which vary across the FoV and will therefore affect individual sources differently. The specific class of DDEs relevant to this work is discussed in the next section.

1.4 Atmospheric effects

Prior to detection at ground level, radio waves of cosmic origin travel through the Earth's atmosphere, where they are also contaminated. Fig. 1.7 illustrates the discrete layers that the atmosphere is made up of at the different altitudes and the corresponding temperatures the layers exist at. Atoms and molecules travel fast within a layer but very slowly between different layers. The layering is a result of the temperature variations of the gas molecules.

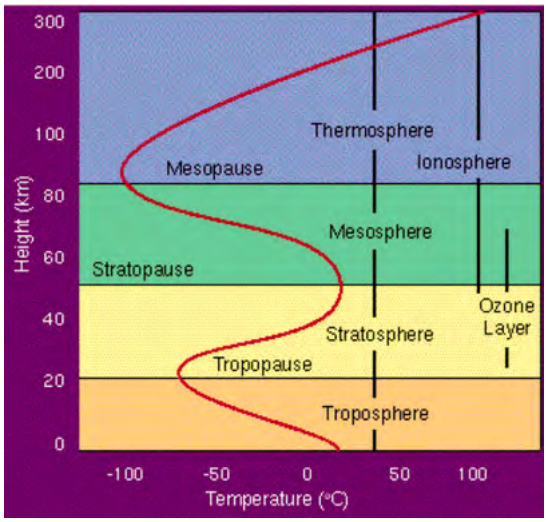


FIGURE 1.7: A schematic diagram of the atmospheric layers. Credits: Shapley (2011).

We can further classify the atmosphere into two media: neutral and ionised. The propagation of radio waves is mostly affected by the troposphere and ionosphere in the neutral and ionised media, respectively.

1.4.1 Tropospheric effects

The troposphere is the region of the atmosphere closest to the Earth, extending from the ground to an elevation of 7 to 10 km. It is also the area of all weather events. Tropospheric phase delay is caused by an excess path length due to refraction. This delay is roughly directly proportional to frequency ν. It is worth mentioning that due to the difference in zenith angle of the source at the antennas, the "horizontally" constructed troposphere induces a differential delay in an interferometer (Thompson et al. 2017). The troposphere has a varying refractive index, N due to variations in temperature (T), pressure and humidity. Given the medium is

made up of a dry and wet component, we can define N_{DRY} and N_{WET} indices depending on the partial pressure of dry air (P) and water vapour (e) respectively. We can therefore describe the atmospheric phase delay, ϕ_{ATM} of a two-way pathways between two points, $\vec{r_1}$ and $\vec{r_2}$ (namely the ground location of the receiver and the "imaginary" location of the place where the image is acquired respectively), proportional to the integration of the total refractivity along the line-of-sight. Using Eqn. (1) and Eqn. (2) from Zhongbo and Mallorqui (2019), ϕ_{ATM} can subsequently be written as,

$$\phi_{\mathrm{ATM}} = \frac{-4\pi}{\lambda} 10^{-6} \int_{\vec{r_1}}^{\vec{r_2}} N_{\mathrm{DRY}}(\vec{r}) + N_{\mathrm{WET}}(\vec{r}) d\vec{r}, \tag{1.24}$$

with

$$N(T, P, e) = N_{\mathrm{DRY}} + N_{\mathrm{WET}} = k_1 \frac{P}{T} + k_2 \frac{e}{T} + k_3 \frac{e}{T^2}, \tag{1.25}$$

where λ is the observing wavelength, and $k_1 = 0.776$ K Pa^{-1}, $k_2 = 0.716$ K Pa^{-1}, $k_3 = 3.75e^3$ K^2 Pa^{-1}. Besides, Fig. 1.8 illustrates that the phase fluctuations in radio interferometers at centimetre and millimetre wavelengths are predominantly generated by variations in the distribution of water vapour. This suggests that, in the atmospheric windows beyond 400 GHz, radio observations are only possible from very dry sites (for example, ALMA Brown et al. (2003)). Likewise, it becomes more difficult to predict excess path length when clouds are present. Water vapour is also not well mixed in the atmosphere, for its density approximately follows an exponential distribution with a scale height of 2 km. Therefore, it is difficult for it to be accurately predicted from surface meteorological records.

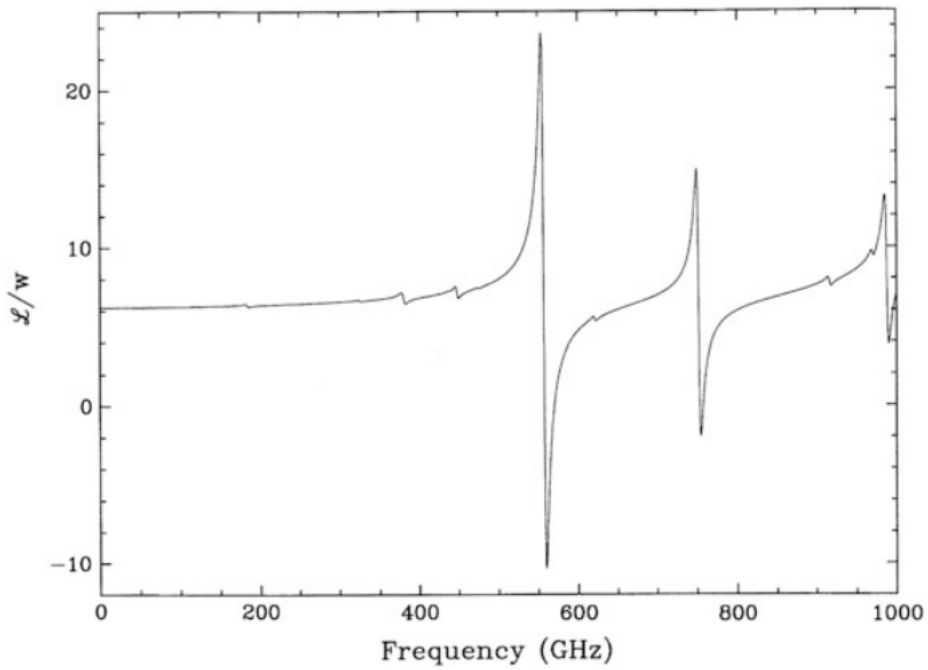


FIGURE 1.8: Plot of predicted phase delay due to water vapour per unit column density against frequency, using formulas from Liebe (1989). Credits: Thompson et al. (2017).

However, since most of the tropospheric phase delay occurs close to the ground, the phase

corruptions are usually decorrelated between antenna pairs but coherent across the FoV of each antenna (see Cotton et al. 2004; Smirnov 2011b). Therefore, this delay can be treated like a DIE and solved via an *antenna-based* calibration.

1.4.2 Ionospheric effects

The ionosphere is a partially ionised turbulent plasma in the upper atmosphere (from around 60 to 1000 km), with an electron density highly variable in time and space. The ionisation in the atmosphere is primarily triggered by solar radiation and is balanced by recombination at night[4]. Speaking of the electron density, large-scale variations are often recorded along the line-of-sight due to travelling ionospheric disturbances (TIDs). These are quasi-periodic wide-scale perturbations in the electron density. They can be caused by auroral heating, severe weather conditions, earthquakes or volcanic eruptions. TIDs can have periods of around minutes to hours.

Table 14.1 in Thompson et al. (2017) gives an idea of the magnitude of propagation effects as a function of the daytime and nighttime ionosphere. Most ionospheric effects scale inversely with frequency. As a result, ionospheric effects are not felt at higher frequencies, which again explains why the prominent atmospheric effects at higher frequencies are caused by the troposphere. At low frequencies (below 300 MHz), the dominant effects caused by the ionosphere are phase delay, Faraday rotation and scintillations. As discussed below, the phase and amplitude effects caused by the ionosphere can be very significant. We address each of the ionospheric effects in the following paragraphs. We finally close this section with a discussion on how the ionosphere affects the detection of the 21 cm line.

Ionospheric phase delay

The ionospheric refractive index n varies with time along the wave path, which implies time-dependent refraction. Similar to the troposphere, ionospheric refraction introduces a phase delay in the response; however, this delay scales with inverse frequency ν^{-1}. At low frequencies, this delay can easily reach $10^2 - 10^4$ rad. The thickness of the ionosphere is measured using Total Electron Content (TEC measured in TEC Units, TECU), which is equal to the size of the cloud of electrons encountered along the line-of-sight (1 TECU = 10^{16} electrons/m^2). Since an interferometer is only sensitive to phase differences, only integrated TEC difference (Δ TEC) between two signal paths to two stations makes up for the phase error due to the aforementioned delay. For this reason, the phase difference, $\Delta\phi$ between the antennas is can be expressed as,

$$\Delta\phi = -\frac{e^2}{4\pi\epsilon_0 c m_e}\frac{1}{\nu}\Delta\text{TEC}, \tag{1.26}$$

where ϵ_0 is the vacuum permittivity, and e and m_e are the electron charge and mass respectively (Loi et al. 2015, Eqn. (2)). The negative sign implies that the sources are expected to refract in the direction of the decreasing TEC values. Furthermore, the TEC values vary

[4]The ionosphere can be divided into several sub-layers, each of which is ionised by different components of the solar radiation. Albeit we understand that each sub-layer contributes to a specific effect, here, we consider the ionospheric effect as a whole.

very quickly (on a minute scale), and this can cause the phase errors to vary on relatively short time and small spatial scales (refer to Intema et al. 2009 for more details). During nighttime, the ionospheric TEC is reduced. The extent of these effects on the observations depends mostly on the FoV (V), array layout (A), and the scale of the ionospheric variations (S). Lonsdale (2005) provides for a detailed study on the 4 different regimes for ionospheric phase calibration involving the above parameters.

Consider Fig. 1.9. Regimes 1 and 2 involve a narrow FoV. In this case, V is usually

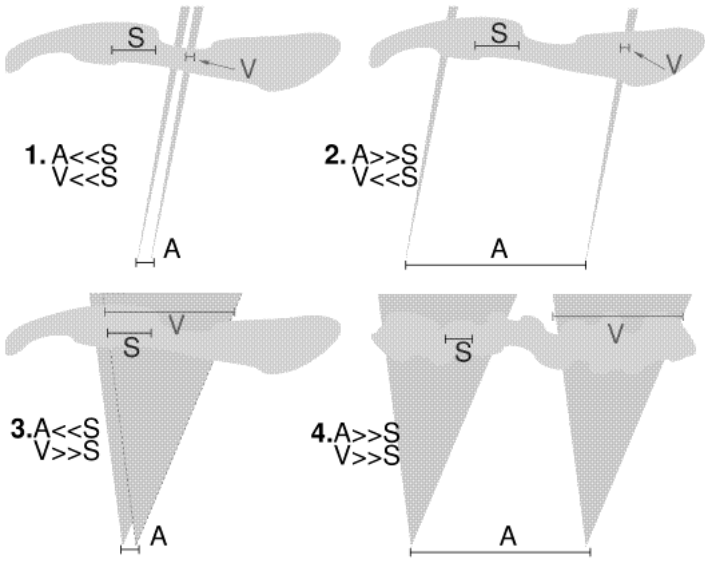


FIGURE 1.9: Schematic diagram describing different ionospheric calibration regimes. An interferometer is observing through some ionospheric electron density (grey clouds). The different regimes scale with the FoV, array layout and scale of the ionospheric variations. Diagram from Lonsdale (2005).

much smaller than the isoplanatic patch of TEC, and thus, each antenna in the array sees an approximately constant TEC across the FoV. Therefore, for narrow fields of view, the ionospheric phase can be treated as a DIE. With a very wide FoV, regimes 3 and 4 in the figure, we have that S $\ll$ V and the antennas now see variations in TEC across the FoV. Thus, while regimes 1 and 2 can be dealt with regular direction-independent self-cal, regimes 3 and 4 require direction-dependent calibration and are therefore the most challenging. In regime 3, the ionospheric phases cause the apparent source position to change with time and direction (refer to Fig. 1.10). However, regime 4 is even worse. Here, the ionospheric phases cause source position shifts as well as source deformations over time and direction. It is also important to mention that the more dynamic the ionosphere is in time, the higher the probability of blurring of sources, and an apparent loss in brightness due to smearing over a larger region of the image (Coker et al. 2009). Given their array configurations, the MWA and LOFAR are expected to operate in regimes 3 and 4 respectively. It is therefore crucial to develop efficient calibration strategies to mitigate these ionospheric effects. For example, if not corrected, position shifts can decrease the accuracy of cross-matching between different epochs of observations, moving sources around by over tens of arcseconds at a frequency of 100 MHz (Heald et al. 2018). But, more importantly, blurring caused by the ionospheric

effects is not necessarily identical between radio sources owing to the small-scale ionospheric structures. In this case, the average of ionospheric phases may differ across epochs.

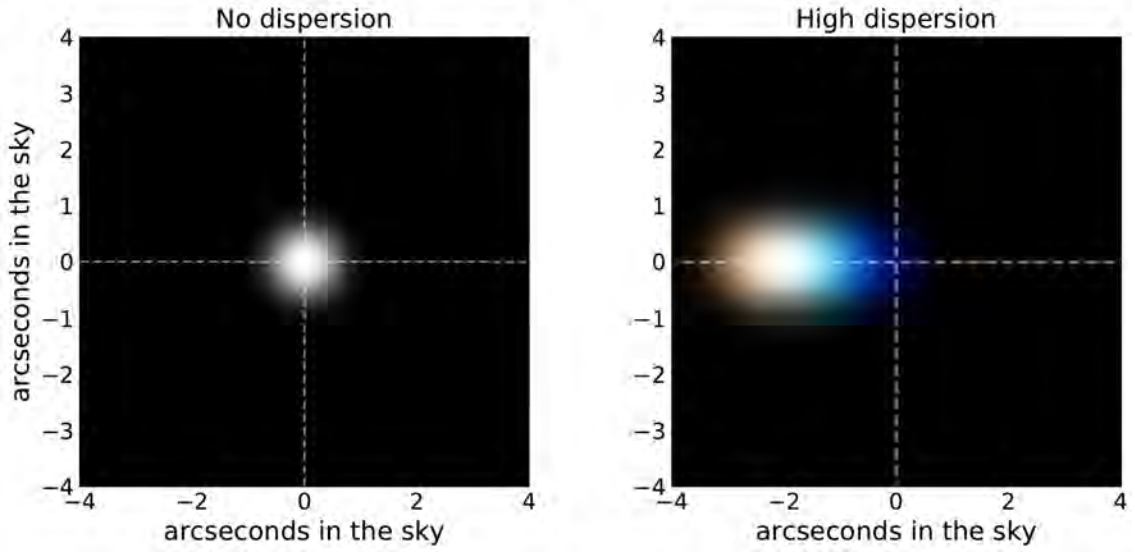


FIGURE 1.10: Diagram showing the effect of ionospheric phase delay on the position of a source. Consider a beam of light from a guiding camera. Note that each colour corresponds to a certain wavelength. When there is no dispersion (left image), all the spots from different wavelengths (visible region) fall on the same location (here, in the centre). As the camera is moved 60° to the left, the barycenter of the source moves along such that every spot represented by a wavelength moves by the corresponding amount of atmospheric dispersion (right image). And, as explained earlier, the apparent position shift (and phase delay) is frequency dependent, and thus, the spot defined by the blue wavelength is the least dispersed. Image from Wehbe et al. (2019).

Ionospheric Faraday rotation

Another effect that needs to be accounted for is ionospheric Faraday rotation. Faraday rotation occurs due to the interaction between the incident wave and the magnetic field of the ionised plasma. This effect is proportional to inverse frequency squared (ν^{-2}). Faraday rotation depends solely on absolute TEC values. In the presence of this effect, different stations see different polarisation rotation of a source due to the spatial structure of the ionosphere. Smirnov (2011b), for example, shows how differential Faraday rotation can make an unpolarised point source appear like a polarised one.

Scintillations

Scintillations are the rapid fluctuations in the phase and amplitude of the waves, as these scatter while propagating through the small-scale irregularities of the electron density of the medium, namely the ionosphere (see De Gasperin et al. 2018), causing a diffraction pattern to appear. These irregularities usually develop after sunset via various deionisation instability processes, and induce random variations in the refractive index of the medium. Thus, scintillation effects are found to be very dominant during nighttime, and can cause massive loss of data with these observations. Scintillations can even translate to image distortions (Herne

et al. 2013). Sometimes it can be difficult to distinguish between ionospheric[5], interstellar (associated with diffractive scattering) and interplanetary scintillations but, in general, they vary on a timescale of 30, 10 and 1 seconds, respectively (refer to § 14.3.2 and Table 14.2 from Thompson et al. 2017).

Consider the following expression of the received signal E,

$$E = (A_0 \delta A) e^{j(\phi_0 + \delta\phi)}, \tag{1.27}$$

where $E_0 = A_0 e^{j\phi_0}$ is the nominal received signal, and $\delta E = \delta A e^{j\delta\phi}$ is the scintillation signal with the amplitude and phase values δA and $\delta\phi$ respectively (refer to Eqn. (1) and Eqn. (2) in Skone et al. 2005). We can then define the scintillation parameter S_4 which stands for the ratio of the standard deviation of the intensity of the received signal to its mean. It is basically a normalised standard deviation of high-frequency "degeneralised" signal intensity minute-long observations. The values for S_4 can vary from 0.1 to 0.9 for weak and strong (occurring around the equatorial regions) scintillations respectively. Furthermore, we can quantify the phase scintillations using a Gaussian distribution with zero mean and standard deviation $\sigma_{\delta\phi}$. In the case of weak and strong scintillations, $\sigma_{\delta\phi}$ can be 0.05 radians and 1.0 radians respectively.

It is difficult to compare amplitude and phase scintillation effects. This depends on several factors, such as altitude, latitude, time, season, observing frequency, and so forth. For example, if we consider a Fresnel scale ($r_{\mathrm{F}} = \sqrt{\lambda z}$, z is the altitude of the ionosphere) which describes the size of the region on the ionosphere where the wavefronts are expected to be approximately in phase (see Loi et al. 2015), we find scintillations are more likely to occur at lower frequencies. However, we will not consider the scintillation effects in this work.

Ionosphere influencing detection of 21 cm line

We have noted that most ionospheric effects tend to be very severe at low frequencies, which coincide with the observation of many science targets such as the 21 cm line. Datta et al. (2016) examine the difficulties behind a 21 cm signal detection below 100 MHz in the presence of a highly dynamic ionosphere. Due to the combined ionospheric effect, from ionospheric refraction, absorption and thermal emission, additional non-smooth time-variable artefacts are created which can hinder foreground subtraction. Datta et al. (2016) also illustrate that "flicker" noise is present in the fluctuations of the ionospheric electron density. It is understood that flicker noise is a non-stationary process which can be used to model solar activity (Press 1978). Unlike Gaussian noise, flicker noise cannot be averaged down with longer observations. Additionally, the column densities in the different sub-layers introduce cumulative uncertainties in the ionospheric parameters. The authors explain that since all the sources in the FoV appear to move towards the zenith due to the ionospheric refraction, this results in a rise in the total power of the radiometer. We can, therefore, understand that ionospheric refraction induces an increase in the excess sky temperature. This discourages the possibility of any ground-based detection of the global 21 cm signal. In short, it is essential

[5] `https://www.spaceacademy.net.au/env/spwx/raiono.htm`

to solve for ionospheric corruptions for the successful detection of the 21 cm signal. In this work, we consider solving for ionospheric phase delays, particularly, for which we will provide some motivation in the next section.

1.5 Motivation

Modern interferometers are being built with very broad fields of view, massive fractional bandwidth as well as high sensitivity and resolution. These existing and future large arrays with especially low frequency capabilities, such as LOFAR, MWA and Square Kilometre Array (SKA), can attain their optimal performance given we use calibration algorithms that can model and remove atmospheric effects from the visibilities. As discussed in § 1.4.2, ionospheric effects tend to be more severe at low frequencies. Since many key science targets are only observable at low frequencies (and specifically the EoR studies), and their signals are usually much fainter than the foregrounds, it is essential that these algorithms are robust in low SNR scenarios and in the presence of incomplete sky models. Existing traditional calibration algorithms are not necessarily efficient in these regimes and can lead to overfitting (Sob et al. 2019). Overfitting can be thought of as the tendency for an unnecessarily complex model to explain the data too well (e.g. when there are too many free parameters and the model fits noise as well as signal). This is further discussed in § 3. In particular, overfitting can often lead to flux suppression in the presence of incomplete sky models. Patil et al. (2016) also show that faint unmodelled line emissions can be suppressed if calibration is not performed carefully and that the level of suppression depends on the relative flux of unmodelled sources. Thus, flux suppression can bias continuum subtraction and restrict spectral line studies.

In theory, the dominant synchrotron nature of the emission implies that the foregrounds should have a smooth spectral behaviour. Sardarabadi and Koopmans (2018) illustrate that if smoothness is not enforced across the gains, a big percentage of the 21 cm line can be severely suppressed. Hence, the model proposed in this work ensures smoothness of the gains along the spatial and spectral dimensions. In particular, we take advantage of the assumption that certain ionospheric effects (such as the ionospheric phase delay) can be adequately captured by a smoothly varying phase screen[6] with an approximately known scaling with frequency. Spatial smoothness of the phase screen is enforced by parametrising it as linear in the parameters model. As an application, we approximate an ionospheric "wedge" model (Wijnholds et al. 2010) by parametrising the per-antenna phases as planes. We then illustrate the advantages of our approach compared to traditional techniques by contrasting the flux suppression and spectral distortions that result when calibrating in the presence of a faint unmodelled spectral line source. Although there are many similar subtleties involved with the direct detection of the EoR signal, we limit the discussion to the 21 cm signal around late-time galaxies in this work. However, the mathematical framework used in this work should be beneficial for low frequency spectral line studies in general.

[6] We acknowledge the potential presence of other effects, particularly amplitude effects, but investigating them is beyond the scope of this work.

1.6 Outline

The rest of the thesis is broken into four chapters and organised as follows:

- Chapter 2 provides for a more sophisticated version of the van Cittert-Zernike theorem used for calibration. It also describes the different generations of calibration and outlines how the least-squares method can be used to calibrate interferometers.
- Chapter 3 defines the model used as well as its implementation.
- Chapter 4 gives an application of the model used, highlighting some advantages of the methods used.
- Chapter 5 summarises this work with some main conclusions and prospects.

Chapter 2

Measurement Model and Optimisation Strategy

The response of an interferometer can be influenced by many different corruptions, as mentioned in § 1.3.2. Since the birth of radio interferometry, astronomers have attempted to model and correct these effects using a variety of techniques. One such technique is hybrid mapping which makes use of closure quantities (Cornwell and Wilkinson 1981). However, these models were not appropriate for all types of instruments. Initially developed for radio polarimetry by Hamaker et al. (1996), the Radio Interferometer Measurement Equation (RIME) provides a solid foundation for radio interferometry. It is a rigorous mathematical framework, describing the measurement model of an interferometer. Smirnov (2011a,b,c) revisit the ideas of the original RIME papers. These attempt to incorporate DDEs while extending the formalism using a complex 2×2 Jones formalism, instead of using the 4×4 Mueller matrices of Hamaker et al. (1996).

This chapter begins with an introduction to the RIME in § 2.1. Furthermore, the chapter gives an outline of the different generations of calibration, grouped into three distinct categories in § 2.2. We then formulate calibration as a non-linear optimisation problem which can be solved using least-squares in § 2.3. Finally, we end the chapter with an overview of the existing approaches dealing with DDEs and ionospheric effects in § 2.4.

2.1 Radio Interferometric Measurement Equation

2.1.1 A full-sky RIME

The RIME generalises the van Cittert-Zernike theorem to account for polarisation and propagation effects. Considering a continuous sky brightness distribution described by a brightness matrix $\mathbf{B}(l,m)$, the RIME can be written as

$$\boldsymbol{V}_{pq} = \boldsymbol{G}_p \left(\int_l \int_m \frac{1}{n} \boldsymbol{E}_p \mathbf{B} \boldsymbol{E}_q^H e^{-2\pi i (u_{pq} l + v_{pq} m + w_{pq}(n-1))} dl dm \right) \boldsymbol{G}_q^H, \tag{2.1}$$

where the Jones matrices $\boldsymbol{G}_p$ and $\boldsymbol{E}_p$ denote the DIE and DDE respectively for antenna p, and $(\cdot)^H$ denotes the conjugate transpose operator. $\mathbf{B}(l,m)$ describes the state of polarisation of the EM radiation coming from all parts of the sky (Hamaker and Bregman 1996) and is

often expressed in terms of the Stokes parameters (Born and Wolf 2013). Therefore, the visibility $\boldsymbol{V}_{pq}$ is the response of the baseline pq to the entire sky in the presence of DI and DD gains. However, the modelled part of the sky consists of a finite number of sources and the brightness matrix of the entire sky is constructed by summing over the individual sources. This means that, in practice, we often work with a discretised form of the RIME, especially during calibration. Note that, unlike the brightness matrix which is assumed to be constant in time, the propagation effects are generally time-dependent. In what follows, we will refer to DDEs that differ between antennas and that can vary with time and frequency as non-trivial DDEs. The exception is the phase delay term $e^{-2\pi i(u_{pq}l+v_{pq}m+w_{pq}(n-1))}$ which is an intrinsic part of how an interferometer works. An example of a trivial DDE is the factor of $\frac{1}{n}$ in Eqn. (2.1).

2.1.2 Discrete RIME

Let us consider a sky consisting of a single point source situated in direction d with coordinates (l_d, m_d, n_d). By substituting a delta function model into the continuous RIME above, its measured visibility can be expressed as

$$\boldsymbol{V}_{pq} = \boldsymbol{J}_p \mathbf{B}_d \boldsymbol{J}_q^H, \tag{2.2}$$

where $\boldsymbol{J}_p$ is a Jones matrix describing the combined propagational effects experienced by the signal $\mathbf{B}_d$ as it travels from the source to where it is detected by antenna p. The total Jones matrix can be broken down into distinct effects which combine multiplicatively. For example, the phase delay term of Eqn. (1.14)) can be incorporated as a (scalar) Jones matrix i.e. $\boldsymbol{K}_{dp} = e^{-2\pi i(u_p l_d + v_p m_d + w_p(n_d-1))}$ where the subscript d indicates that it is a direction-dependent effect. There might also be an electric gain term, $\boldsymbol{G}_p$ say, associated with the antenna. In this case, the combined Jones matrix can be written as

$$\boldsymbol{J}_{dp} = \boldsymbol{G}_p \boldsymbol{K}_{dp}. \tag{2.3}$$

To incorporate multiple point sources into the model, we can simply sum over each source i.e.

$$\boldsymbol{V}_{pq} = \boldsymbol{G}_p \left(\sum_d \boldsymbol{K}_{dp} \mathbf{B}_d \boldsymbol{K}_{dq}^H \right) \boldsymbol{G}_q^H, \tag{2.4}$$

$$= \boldsymbol{G}_p \Big(\sum_d \boldsymbol{X}_{dpq} \Big) \boldsymbol{G}_q^H, \tag{2.5}$$

where $\mathbf{B}_d \in \mathbb{C}^{2\times 2}$, denotes the *brightness* matrix for direction d, and the source coherency, $\boldsymbol{X}_{dpq} \in \mathbb{C}^{2\times 2}$, stands for the visibility that the interferometer would otherwise measure in the absence of all corruptions in the direction of source d. In a nutshell, the aim of calibration is to solve for the unknown propagation effects in model such as Eqn. (2.5). When the propagation effect is a DIE, it is possible to apply the inverse of the estimated gains, denoted $\hat{\boldsymbol{G}}$, to the

data. This results in *corrected* visibilities which are computed using

$$\begin{aligned} \boldsymbol{V}_{pq}^{c} &= \hat{\boldsymbol{G}}_p^{-1} \boldsymbol{V}_{pq} \hat{\boldsymbol{G}}_q^{-H}, \\ &= \hat{\boldsymbol{G}}_p^{-1} \Big(\boldsymbol{G}_p \Big(\sum_d \boldsymbol{X}_{dpq} \Big) \boldsymbol{G}_q^H \Big) \hat{\boldsymbol{G}}_q^{-H}. \end{aligned} \tag{2.6}$$

Clearly, if $\hat{\boldsymbol{G}}_p$ is a good approximation of $\boldsymbol{G}_p$, we have effectively corrected for the corruptions present in the measured visibilities. However, since the observation is inevitably corrupted by an unknown noise realisation, this step has to be performed with care as it can lead to subtle artefacts (Grobler et al. 2014).

Things get a little trickier when the observation is corrupted by non-trivial DDEs. In this case, the discretised form of Eqn. (2.1) can be written as[1]

$$\boldsymbol{V}_{pq} = \boldsymbol{G}_p \left(\sum_d \boldsymbol{E}_{dp} \boldsymbol{X}_{dpq} \boldsymbol{E}_{dq}^H \right) \boldsymbol{G}_q^H, \tag{2.7}$$

where $\boldsymbol{E}_{dp}$ denotes the DDE in the direction d. While it is in principle possible to incorporate known DDEs into the measurement model used during imaging, doing so is beyond the scope of this thesis. A far more common approach to mitigating the effects of DDEs is to calibrate for them in the direction of the worst affected sources and then to subtract these corrupted sources from the data. This technique is often referred to as *peeling* (Intema et al. 2009; Noordam 2004). It consists of firstly calibrating individual sources in order of decreasing brightness, such that we obtain a per-source set of time-dependent antenna-based phase corrections and a source model. This source model is subsequently subtracted from the visibility data while applying the DD gain solutions. However, these approaches are not completely optimal for a number of reasons. Firstly, in solving for a single DDE in the direction of each source, we have assumed that the DDE is constant across the extent of the source. Fortunately, most DDEs tend to be slowly varying in the spatial dimensions, so this is a reasonable assumption as long as the source is small compared to how quickly the DDE varies. Secondly, while peeling can in principle get rid of artefacts from sources surrounding the target, it cannot be used to correct for DDEs in the direction of the target. Thus, peeling does not provide a means to get high fidelity[2] images of the target when it experiences severe DDEs. However, the corrupting effects of DDEs are most troublesome when they affect compact sources that are brighter than the science target since the resulting sidelobes contaminate the rest of the field. Thus, it often suffices to subtract only the worst affected sources that are brighter than the science target. While this approach does not work in all cases, in this thesis, we will limit ourselves to this type of DDE mitigation.

The RIME offers a flexible model to handle an arbitrary number of propagation effects or Jones matrices. A purely Physics-based approach would attempt to model (or parametrise) each physical effect and incorporate an individual Jones matrix into the Jones chain for each effect and in the correct order. This, however, becomes computationally very complicated

[1] Detailed derivations for the continuous and discrete forms of the RIME can be found in Smirnov (2011a).
[2] Image fidelity can be defined as a measure of the accuracy of the reconstructed sky brightness distribution.

and is intractable in most cases. On the other hand, if we have a good enough model (i.e. $\boldsymbol{X}_{dpq}$) to begin with, we could incorporate the cumulative effect of all Jones matrices into a single phenomenological Jones matrix. This can reduce the number of terms to solve for but, without a good parametrisation, it can also result in overfitting and flux suppression (see Noordam and Smirnov 2010; Smirnov 2011b for example). In practice, the specific form of the RIME used depends on the objective of the study. In this work, we make use of a RIME of the form in Eqn. (2.5), where we only focus on solving for DD phase screens via calibration. As discussed later on, these screens are parametrised in a way that attempts to strike a balance between the purely Physics-driven approach and the more tractable phenomenological approach. They can also be incorporated as part of a larger chain in a solver (as is done in `CubiCal`, for example Kenyon et al. 2018). In the next section, we present a summary of the different generations of calibration involving radio interferometric data.

2.2 Evolution of calibration

Radio interferometric calibration is usually classified into three (more manageable) categories named as first, second and third generation calibration (denoted 1GC, 2GC and 3GC, respectively). These terms were originally coined by Noordam and Smirnov (2010). This section outlines these three phases of calibration.

First generation calibration

First generation calibration makes use of dedicated calibrator observations which are interspersed with observations of the target field. Each individual observation is referred to as a scan. Calibrator fields contain sources with known parameters such as flux, shape and spectrum. Using these known properties to derive a good source model then makes it possible to solve for the dominant Jones terms affecting the observation. This is done assuming that the Jones terms are direction-independent, which they are, for the most part. The solutions are then transferred onto the target field by interpolating between the calibrator scans and applying the interpolated solutions using Eqn. (2.6). This process roughly corrects the response of the interferometer and can be used to set the absolute flux, bandpass and correct for delay errors (Thompson et al. 2017). However, the interpolated gains are only a rough approximation of the true Jones terms affecting the target field. The solutions can and should be further refined in order to improve the dynamic range[3] of the observation. For this, we resort to second generation calibration.

Second generation calibration

Second generation calibration, which is often referred to as *self-cal*, is an iterative process which uses the target field to better calibrate and improve the observed data and dynamic range in the image (Cornwell and Wilkinson 1981). The early Eighties proved to be a golden era where the works of Schwab (1980) and Cornwell and Wilkinson (1981) pushed the upper

[3]The dynamic range is defined as the ratio of the peak brightness to an estimate of the noise in an image.

limits of self-cal methods. Prior to this, self-cal was known as hybrid mapping, which was based on the computation of per-baseline closure quantities (Cornwell and Wilkinson Cornwell and Wilkinson 1981). Self-cal can be performed using the following steps:

1. Create an initial sky model from 1GC calibrated data.
2. Get model visibilities by applying a Fourier transform on this model.
3. Refine the gain estimates by performing calibration with the improved (or more complete) set of model visibilities.
4. Apply an inverse of the gains on the observed visibilities to generate corrected visibilities (refer to Eqn. (2.6)).
5. Apply an inverse Fourier transform on the corrected visibilities to obtain a dirty image.
6. If the dynamic range of the image is good enough or the maximum number of set iterations is reached, stop. Else, deconvolve the dirty image to produce a restored image which updates the sky model. Go back to Step 2.

Residual amplitude and phase errors can be removed using self-cal provided there are sufficient SNR, enough baselines, and an accurate sky model. But the downside of this strategy is that the calibration is only as accurate as the model since model errors can get absorbed into the calibration solutions. This process also does not correct for DDEs because, as discussed above, there is no simple procedure to produce corrected visibilities in the presence of DDEs (refer to § 2.1.2). This is where the third generation calibration comes in.

Third generation calibration

Third generation calibration techniques refer to the correction of DDEs[4] in the observed data. Traditional self-cal works under the assumption that the DDEs are trivial and that these can be corrected for in the image plane after calibration and deconvolution have been performed. However, the effects of non-trivial DDEs become more apparent with increasing sensitivity, FoV and fractional bandwidth of the interferometric array. As a result, traditional self-cal is not sufficient for the kind of science targeted by modern interferometric arrays such as LOFAR, MWA, MeerKAT and eventually the SKA. Unfortunately, unmodelled DDEs can corrupt the observation and lead to image artefacts which make it difficult to study the fainter sources of interest (for example, see Smirnov 2011b,c). In general, the fainter the target, the more important it becomes to correct for DDEs. As discussed above in § 2.1.2, we only consider the case in which non-target sources affected by DDEs are subtracted (peeled) from the data. A set of time-variable phase corrections is estimated for the worst DDE corrupted source. This source model is removed from the visibility data, and the process is repeated for the next worse affected source. Once the unwanted corrupted sources have been subtracted from the data, we can correct for DDEs affecting the target by applying the inverse gain solutions in the direction of the target (obtained, for example, by interpolating the DD

[4]Smirnov (2011b) provides for a detailed analysis of many different sources of DDEs.

solutions or by solving for a gain term in the direction of the target if there is sufficient SNR). In the next section, we discuss how to approach calibration using the least-squares method. We will thereafter discuss some of the important additions to ionospheric DD calibration in § 2.4.

2.3 Least-squares method

Calibration can be defined as an optimisation problem. The first step in setting up an optimisation problem is identifying the objective function to optimise. In radio interferometry, the data are usually subject to Gaussian noise which naturally results in an objective function that can be dealt with using a non-linear least-squares (NLLS) method. Amongst the traditional NLLS methods, the Gauss-Newton (GN) (Madsen et al. 2004) algorithm is a commonly employed method with good convergence properties. We outline the derivation below.

Consider the following problem,

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{d} - \boldsymbol{v}(\boldsymbol{\alpha})||_F^2, \tag{2.8}$$

which describes a least-squares fit of the data $\boldsymbol{d} \in \mathbb{R}^m$ using a model $\boldsymbol{v} \in \mathbb{R}^m$ parametrised by (real) variables $\boldsymbol{\alpha}$ where $||\cdot||_F$ denotes the Frobenius norm. Note, in our case, although the data are complex, we will be optimising a real-valued function with respect to real-valued parameters. A similar derivation involving augmented vectors and complex-valued functions can be found in Amin (2012). If we let $\boldsymbol{r}(\boldsymbol{\alpha}) = \boldsymbol{d} - \boldsymbol{v}(\boldsymbol{\alpha})$ then, we want to minimise the χ^2 which is defined by the cost function $g(\boldsymbol{r}(\boldsymbol{\alpha})) = \boldsymbol{r}^H\boldsymbol{r}$. Now, for a small change in $\boldsymbol{\alpha}$, $\Delta\boldsymbol{\alpha}$ say, we have that $\boldsymbol{r}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha}) = \boldsymbol{d} - \boldsymbol{v}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})$. Using a Taylor expansion and neglecting higher-order terms,

$$\begin{aligned} \boldsymbol{v}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha}) &\approx \boldsymbol{v}(\boldsymbol{\alpha}) + \frac{\boldsymbol{v}'(\boldsymbol{\alpha})}{1!}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha} - \boldsymbol{\alpha}) + \mathcal{O}\Big(||\Delta\boldsymbol{\alpha}||^2\Big), \\ &= \boldsymbol{v}(\boldsymbol{\alpha}) + \boldsymbol{J}\Delta\boldsymbol{\alpha} + \mathcal{O}\Big(||\Delta\boldsymbol{\alpha}||^2\Big), \end{aligned} \tag{2.9}$$

where $\boldsymbol{J}$ is called the Jacobian matrix and it contains all the first-order partial derivatives of the visibilities, $\boldsymbol{v}$ with respect to $\boldsymbol{\alpha}$. If the cost function is now evaluated at $\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha}$,

$$\begin{aligned} g(\boldsymbol{r}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})) &= \left[\boldsymbol{r}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})\right]^H \left[\boldsymbol{r}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})\right], \\ &= \left[\boldsymbol{d} - \boldsymbol{v}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})\right]^H \left[\boldsymbol{d} - \boldsymbol{v}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha})\right], \\ &= \left[\boldsymbol{d} - \boldsymbol{v}(\boldsymbol{\alpha}) - \boldsymbol{J}\Delta\boldsymbol{\alpha}\right]^H \left[\boldsymbol{d} - \boldsymbol{v}(\boldsymbol{\alpha}) - \boldsymbol{J}\Delta\boldsymbol{\alpha}\right], \\ &= \left[\boldsymbol{r}(\boldsymbol{\alpha}) - \boldsymbol{J}\Delta\boldsymbol{\alpha}\right]^H \left[\boldsymbol{r}(\boldsymbol{\alpha}) - \boldsymbol{J}\Delta\boldsymbol{\alpha}\right], \\ &= \boldsymbol{r}^H(\boldsymbol{\alpha})\boldsymbol{r}(\boldsymbol{\alpha}) - \Big(\boldsymbol{J}\Delta\boldsymbol{\alpha}\Big)^H \boldsymbol{r}(\boldsymbol{\alpha}) - \boldsymbol{r}(\boldsymbol{\alpha})\boldsymbol{J}\Delta\boldsymbol{\alpha} + \Big(\boldsymbol{J}\Delta\boldsymbol{\alpha}\Big)^H \boldsymbol{J}\Delta\boldsymbol{\alpha}, \\ &= \boldsymbol{r}^H\boldsymbol{r} - 2\boldsymbol{r}^H\boldsymbol{J}\Delta\boldsymbol{\alpha} + \Big(\Delta\boldsymbol{\alpha}\Big)^H \boldsymbol{J}^H\boldsymbol{J}\Delta\boldsymbol{\alpha}. \end{aligned} \tag{2.10}$$

At the stationary point, we have that $\nabla_{\boldsymbol{\alpha}} g = 0$ leading to

$$\begin{aligned} -2\boldsymbol{r}^H \boldsymbol{J} + 2\boldsymbol{J}^H \boldsymbol{J} \Delta\boldsymbol{\alpha} &= 0, \\ \Delta\boldsymbol{\alpha} &= \left(\boldsymbol{J}^H \boldsymbol{J}\right)^{-1} \boldsymbol{J}^H \boldsymbol{r}. \end{aligned} \tag{2.11}$$

Eqn. (2.11) is called the update rule of the GN algorithm. The product $\boldsymbol{J}^H \boldsymbol{J}$ is the linearised approximation of the Hessian, also sometimes called the Fisher information. The full Hessian i.e. a matrix containing second-order derivatives of the objective function with respect to the parameter vector $\boldsymbol{\alpha}$, has an additional contribution to it. As discussed in Finsterle (2011), this additional contribution can become significant in the presence of large residuals or, in other words, when we are far from the minimum of the objective function or the data are very noisy. It usually becomes negligible when we are in the vicinity of a minimum of the objective function and can safely be neglected as long as a reasonable starting point is chosen for the optimisation. In addition, for $\Delta\boldsymbol{\alpha}$ to be a decent direction, the operator acting on $\boldsymbol{J}^H \boldsymbol{r}$ needs to be positive definite. Unlike the full Hessian, the approximate Hessian, $\boldsymbol{J}^H \boldsymbol{J}$, is guaranteed to be positive semi-definite as long as the number of data points is larger than the number of parameters we are solving for (see Nocedal and Wright 2000 for example). To avoid the possibly numerically unstable case in which $\boldsymbol{J}^H \boldsymbol{J}$ can become singular, the Levenberg-Marquardt (LM) algorithm (Levenberg 1944; Marquardt 1963) modifies the update rule to

$$\Delta\boldsymbol{\alpha} = \left(\boldsymbol{J}^H \boldsymbol{J} + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{J}^H \boldsymbol{r}, \tag{2.12}$$

where $\lambda > 0$ is tunable parameter that affects both the numerical stability and the rate of convergence of the algorithm. Usually it suffices to set λ to a value that approximately matches the desired numerical precision. Regardless of the exact form of the equation used, once $\boldsymbol{\Delta}\alpha$ has been computed, the solution at the k^{th} iteration is updated according to

$$\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_{k-1} + \Delta\boldsymbol{\alpha}_k. \tag{2.13}$$

As long as λ is chosen in some reasonable way, and given a sensible starting guess for the parameters, both the GN and LM algorithms have approximately second-order convergence rates. In addition, as with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, the procedure usually remains stable even when the linearised Hessian $\boldsymbol{J}^H \boldsymbol{J}$ is approximated further. Because of the typical data sizes in radio interferometry, this is an important aspect that we will return to later on.

2.4 Existing approaches to ionospheric calibration

Arrays operating at low frequencies such as the MWA and LOFAR experience DD ionospheric effects which limit the achievable dynamic range of the observations. Many methods have been developed to address this problem, and we briefly discuss some of them below.

SPAM

Source Peeling and Atmospheric Modelling (SPAM) is a calibration method which attempts to solve and correct for ionospheric effects while employing peeling (Intema et al. 2009). SPAM makes use of phase solutions from the peeled sources in the FoV to constrain an ionospheric phase screen. This screen is a time-varying two-dimensional "thin-layer" at a fixed height, corresponding to the small-scale structure of the ionosphere. The phase screen is further described using Karhunen-Loève (KL) transforms to obtain an optimal set of basis functions assuming some a priori known power spectrum for the phase screen. The method provides for more accurate results than existing self-cal.

FBC

In Field-based calibration (FBC), snapshot images of the bright sources are taken, whereby the position offsets are measured (Cotton et al. 2004). Then, an ionospheric model is estimated based on these measurements such that the ionospheric phase delays can be corrected for during imaging. However, such methods are found to be not suited for very long baseline arrays like LOFAR (Intema et al. 2009; van Weeren et al. 2016).

Faceted solutions

Tasse et al. (2018) present a wideband wide-field spectral deconvolution framework called `killMS` and `DDFacet`, based on image faceting, which can account for known DDEs during imaging and deconvolution. The method relies on separating the full FoV into polygonal facets and splitting the latter into smaller (usually square) facets. The `DDFacet` imager can apply known DDEs during imaging and deconvolution on a per-facet basis. The full RIME is then discretised by summing the contributions from each facet. Using the same form as the discretised RIME given in Eqn. (2.7), a solution is then obtained for each facet (using the `killMS` package), assuming that the solution is constant across the facet. The resulting per-facet solutions can then be propagated onto their constituent facets during imaging with `DDFacet`. This approach has been very successful with LOFAR data (Duncan et al. 2019; Shimwell et al. 2017, 2018; Williams et al. 2018).

SAGECal

Yatawatta et al. (2008) developed the SAGE (Space Alternating Generalised Expectation-maximisation) algorithm in the context of radio interferometric data. The algorithm is an iterative method for maximum likelihood estimation and shows a high speed of convergence. Kazemi et al. (2011) gives a complete description of the algorithm. The field is divided into multiple clusters of sources and then, solutions are estimated along each of the "cluster" directions. The method has the advantage of having taken into account all the sources across the field.

LOFAR Facet calibration

The LOFAR Facet calibration (van Weeren et al. 2016) is yet another calibration scheme to produce images of LOFAR-HBA (High Band Antenna, used to observed in the range $110-240$ MHz) Tier-1 survey depth and resolution. Similar to Faceted solutions, Facet calibration also divides the sky into a number of isoplanatic patches. However, its main goal is to produce fully corrected images of the radio sky. It builds itself using the peeling technique, and holds close resemblance to SPAM. Calibration solutions are obtained for one facet at a time, by solving for phases on short timescales (~ 10 s), and parallel hand gains on longer timescales (~ 10 minutes). The updated sky model represented by the facet is subtracted from the data. This process is repeated for the subsequent facets, which eventually leads to the corrected image of the full FoV.

Clustered calibration

Presented by Kazemi et al. (2013), clustered calibration is an attempt to modify traditional radio interferometric calibration by accounting for very faint sources. The idea is to calibrate faint sources nearby bright sources, as a source cluster, made up of faint sources, but sufficiently stronger than background noise. The solutions obtained for each cluster are assigned to the individual sources in that cluster. The approach shares similarities with the faceted approach of Tasse et al. (2018).

Various features from some of the mentioned methods are later compared in the Table 3.2.

2.5 Summary

In this chapter, we introduced the RIME as our measurement model for calibration and outlined the different generations of calibration. We also discussed how GN and LM can be applied to solve non-linear optimisation problems. Finally, we discussed some existing approaches used for DDE calibration. In the next chapter, we will introduce our proposed model for ionospheric calibration and cast it into the form of a least-squares problem which can be solved using the tools we have just discussed.

Chapter 3

Parametrised Phase Solver

Calibration of radio interferometric observations grows difficult at lower frequencies where fields of view naturally get wider, and the ionosphere becomes the dominant systematic DDE. Refraction of cosmic radio waves occurs in the ionosphere which results in phase errors in the measured signal. As we saw above, these phase errors can be solved for directly, using an approach such as peeling or facet-based solutions. DD solutions can require a large number of degrees of freedom (DoF) to capture the variation of phase with antenna, direction, frequency and time. This can be problematic for two reasons:

- A larger number of DoF requires higher SNR for the solutions to be stable;
- Excessive DoF in the calibration model can have undesirable side effects, such as flux suppression of unmodelled sources, and ghost source formation, which are faint source-like calibration artefacts, mostly visible (and noticeable) in residual maps (Grobler et al. 2014; Sob et al. 2019) (discussed in § 4).

Therefore, it could be hoped for that a physically-motivated parametrised solver could alleviate these problems, by capturing the ionospheric error behaviour with a smaller number of DoF. There are two obvious avenues to exploit. Firstly, we know that the frequency behaviour of the ionospheric phase delay is given by $\sim \mathrm{TEC}/\nu$[1], so we do not need to allocate explicit DoF to the variation in frequency. Secondly, we could attempt to construct a model for the ionospheric TEC that takes advantage of spatial and temporal continuity. Since the TEC is usually concentrated at a narrow range of altitudes and the Earth's radius is much bigger than the ionosphere's thickness, the ionosphere can be modelled as an infinitesimally thin spherical shell surrounding the Earth at a fixed height. This is also called the "thin-layer model" (for example see Martin et al. 2016). Thus, we can treat the ionosphere as a plane if we ignore the curvature of the Earth (see Fig. 3.1 below). Mathematically, this means that we can reduce the three-dimensional phase structure of the ionosphere to a two-dimensional one by integrating the free electron density along the line-of-sight (Intema et al. 2009). Two previous approaches to this problem have adopted the thin-layer model. In Field-Based Calibration (FBC), a linear "wedge" model (i.e. a linear gradient over the array) for the TEC

[1]Unfortunately, in real data, this relationship only stands when other effects/contaminants are removed from it. Among these effects, there exist higher-order ionospheric terms which introduce very significant errors at lower frequencies, (refer to § 2 and § 5 in Datta-Barua et al. 2008). These effects can be modelled by incorporating them as part of a larger Jones chain and solving for the individual terms (Kenyon et al. 2018). Nevertheless, the study and inclusion of these effects in our simulations are beyond the scope of this work.

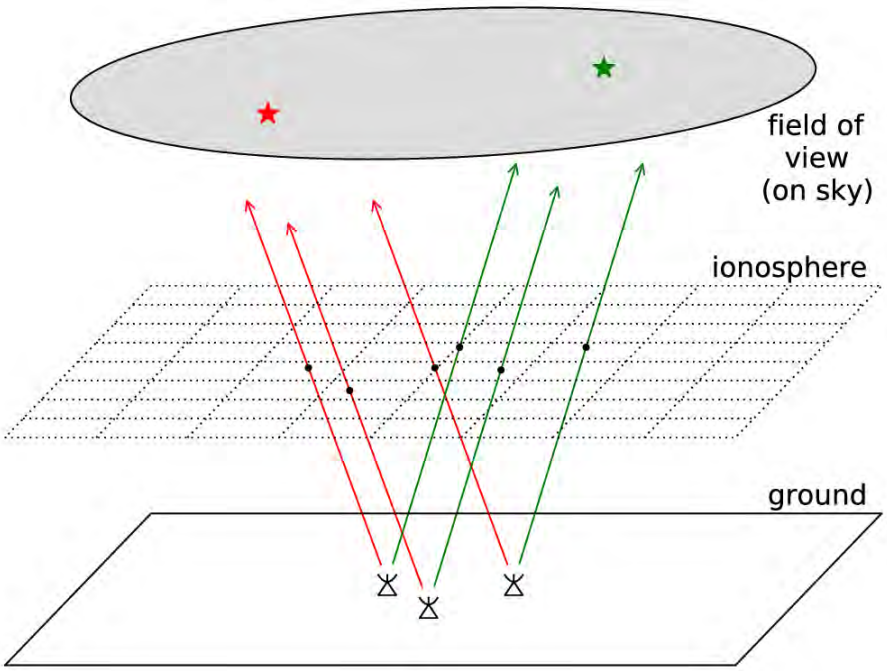


FIGURE 3.1: An illustration of the ionosphere as a plane at a fixed height. An array of 3 antennas is simultaneously observing 2 sources. The line-of-sight pierces the phase screen at the black dots, also called the *pierce points*. Image from Martin et al. (2016).

distribution is implicitly assumed. Images of the field are constructed at a specified time cadence, and positions of bright compact sources are measured. From the RIME, it can be seen that a phase gradient over the array can be equivalently expressed as an offset in the apparent source position. By inverting this relationship, FBC derives the wedge parameters from the measured source offsets. The reduction in DoF is considerable: instead of a per-source phase term, we have 3 parameters of the linear wedge, per time interval considered. The limitation of FBC is that, by design, only a linear TEC screen can be recovered, whose gradient varies from antenna to antenna. Higher-order terms in the TEC screen correspond to source distortions rather than simple positional shifts, and cannot be captured by FBC (and in fact, it can be expected that the performance of FBC even w.r.t. recovering the linear terms will degrade in the presence of TEC curvature, due to source shape distortion affecting the position measurements).

The SPAM method first uses a traditional peeling approach to derive DD phase solutions towards bright compact sources (introduced in § 2.4). SPAM then fits these solutions with a two-dimensional TEC screen composed of KL basis functions. The DoF are then given by the number of KL coefficients fitted per time interval. The advantage over FBC is that higher-order spatial distributions of the TEC can be represented. For both FBC and SPAM, the sky is tiled with small facets; for each facet, a model phase is estimated corresponding to the centre of the facet, and incorporated as part of the measurement model while imaging this facet, and eventually the whole apparent sky.

Recent work by Albert et al. (2020a,b) further extends this approach. By taking the prior of the 3D correlation structure of the refractive index of the ionosphere, Albert et al. (2020a)

use Gaussian Process Regression (GPR) to infer a 3D model of the ionosphere from DD gain solutions. In Albert et al. (2020b), they apply this to real data from LOFAR-HBA (using DD solutions derived by the `killMS` package[2] in the LoTSS pipeline). For computational reasons, they fall back onto a thin-shell ionospheric model and perform GPR in 2D. They show substantial improvement in LOFAR-HBA images using this technique, whereby the resulting images have fewer DDEs between bright calibrators, and the root-mean-squared residuals around bright sources are reduced by 37% on average, with respect to the LoTSS (second data release) images.

One common feature of the methods described above is that they can be considered a type of *a posteriori* regularisation. These methods rely on fitting TEC models to intermediate (i.e. derived) quantities, rather than the raw data. In the case of FBC, these are the measured source positions; for the other two approaches, these are the DD phase solutions. Implementing such a two-step inference has been necessary for computational and operational reasons, but this does make the methods sensitive to parameters of the calibration process (imaging intervals, peeling solution intervals, and so forth) in ways that are not mathematically obvious, and makes the statistical treatment of uncertainties difficult. It also means that any problems associated with DD solutions (and the preceding DI calibration) are already "baked in" before ionospheric inference commences. Examples of such problems are flux suppression, ghost sources, and other imaging artefacts. Albert et al. (2020b) describe the regularisation steps taken with DD solutions in order to alleviate such artefacts. Note that these steps have been developed fully empirically, through a process of trial and error, and still do not entirely eliminate all the artefacts. Furthermore, "outlier" DD solutions are liable to cause instability in the subsequent inference steps, and Albert et al. (2020b) go to considerable efforts to develop a method to flag them.

In principle, these problems can be sidestepped entirely by inferring ionospheric parameters (and DI calibration parameters) directly from the measured visibilities, given we compromise on the degree of the complexity of the considered ionosphere. This method has the advantage of minimising DoF from the outset (thus, potentially, minimising imaging artefacts), as well as working in a data space where the noise statistics are simple and well-defined. Such a treatment, however, constitutes a huge and computationally expensive inference problem. In this work, we only take the first step towards it, by formulating and solving a minimisation problem that infers ionospheric parameters directly from the raw visibility data.

Following in the footsteps of previous approaches, we approximate a "thin-layer" ionosphere by a 2D TEC screen composed of polynomial basis functions. We then use the TEC/ν relationship to derive phase terms for the RIME and use this as the forward modelling step in our optimisation problem. The model is referred to as `pphase` throughout this work. Note that we make one further simplification (compared to FBC and SPAM), in that we model an independent TEC screen above each station, rather than a single "global" TEC screen over the array. Doing so simplifies the mathematics of the optimisation problem, at the cost of increasing the number of DoF in comparison to a single-screen model. An extension of our approach to a global screen solver will form the basis of future work. Alternatively, one could

[2] `https://github.com/saopicc/killMS`

consider a single screen for the array core and independent screens per group of adjacent remote stations, such that there are fewer number of DoF without generalising, for example, extremes of ionospheric phase errors across an array. Having said that, in terms of the discussion in § 2.1.2, our solver can be thought of as a combined physical-phenomenological approach.

We present a mathematical framework for the aforementioned model in § 3.1. The corresponding derivatives are presented in § 3.2. Finally, we test the solver and present a summary of its implementation in § 3.3.

3.1 Mathematical framework

We consider the following NLLS problem,

$$\min_{\{\boldsymbol{\alpha}\}} \sum_{pq} ||\boldsymbol{R}_{pq}||_F^2, \quad \boldsymbol{R}_{pq} = \boldsymbol{D}_{pq} - \boldsymbol{V}_{pq}, \tag{3.1}$$

where,

$$\boldsymbol{V}_{pq} = \sum_d \boldsymbol{E}_{dp} \boldsymbol{X}_{dpq} \boldsymbol{E}_{dq}^H, \tag{3.2}$$

and $\boldsymbol{D}_{pq}$ and $\boldsymbol{V}_{pq}$ stand for the observed and modelled visibilities, respectively. We can express the gains as,

$$\begin{aligned} \boldsymbol{E}_{dp} &= \begin{pmatrix} \mathcal{E}_{dp}^{XX} & 0 \\ 0 & \mathcal{E}_{dp}^{YY} \end{pmatrix}; \\ \mathcal{E}_{dp}^{XX} &= e^{i\phi_{dp}^{XX}}; \\ \phi_{dp}^{XX}(l_d, m_d, \nu) &= \left(\sum_{k=0}^{N_{\text{PAR}}-1} h_k(l_d, m_d) \alpha_{pk}^{XX} \right) f(\nu). \end{aligned} \tag{3.3}$$

Instead of solving for a phase as it is with phase-only gains, Eqn. (3.3) parametrises a DDE $\boldsymbol{E}_{dp}$ as a function of frequency ν and source positions $(l_s,\ m_s)$ to fit for a per-antenna 2D screen. $f(\nu)$ is the expected frequency scaling (ν^{-1} for a TEC screen), $h_k(l_d, m_d)$ (or h_{dk}) is an arbitrary (ideally orthogonal) basis and $\alpha_{pk} \in \mathbb{R}$ is the k^{th} coefficient for antenna p. We refer to $\boldsymbol{\alpha}_p$ as the parameter vector for antenna p. This vector is of size N_{PAR} (number of parameters). In this work, we restrict our derivation to the case where both the gains and data are diagonal. For this reason, the gains above are expressed with only on-diagonal correlations for the case of linear feeds (although the same expressions are applicable for circular feeds). Equivalent expressions for the YY terms can be obtained by replacing XX with YY. It is important to note that physically, we expect the visibilities corresponding to the parallel-hand correlations to be the same (ignoring the noise contribution to the feeds). But, for the purpose of the mathematics here, we assume that they are independent. This also ensures a more generic model, which can be later applied to other different problems.

This work assumes that the effect of the ionosphere can be adequately captured by a per-station TEC screen, and solves for the parameters of the screen. The method is only

partially Physics-based: the frequency scaling of the phase is physical, but the fact that the screen is constructed independently per-antenna is not. We can think of this method as a mix between the Physics-based and purely phenomenological approaches, which seeks robustness by restricting the number of free parameters in the model (to be revisited in the § 4.3).

3.1.1 Example

A p^{th}-antenna linear "wedge" TEC screen corresponds to $N_{\rm PAR} = 3$ and $f(\nu) = \frac{1}{\nu}$. The phase can be written as

$$\phi_{dp}^{XX} = \frac{1}{\nu}(\alpha_{p0}^{XX} + \alpha_{p1}^{XX} l_d + \alpha_{p2}^{XX} m_d), \tag{3.4}$$

which represents a plane at an arbitrary height. It is not essential to express the plane's height in the above equation since the phase is presented as infinitely thin. Hence, although we solve for different values of α_{pk} at different heights, we end up with the same phase effect. Furthermore, the frequency scaling follows from the expected frequency dependence of the propagation effects mentioned in § 1.4.2. Here, the polynomial basis is bivariate and is of degree $\frac{1}{2}(N_{\rm PAR} - 1)$. We omit cross terms. Through the polynomial, we can enforce the smoothness of the phase screen across the sources. With respect to the number of DoF, say we are to calibrate 5 independent frequency channels, in 10 directions, for which we are using 2×2 diagonal Jones, then for each calibration interval, how many free parameters are we solving for, traditionally and with `pphase`? This is illustrated with different array cases in Table 3.1. The number of free parameters per interval is calculated by multiplying the number of channels × antennas × Jones into number of directions for traditional algorithms, and into $N_{\rm PAR}$ (= 3 here) for `pphase`. The reduction in DoF is given by $1 - \frac{N_{\rm PAR}}{N_{\rm D}}$, and is equal

Array	Traditional	`pphase`
VLA (27)	2700	810
LOFAR (52)	5200	1560
MeerKAT (64)	6400	1920

TABLE 3.1: Number of free parameters to calibrate per calibration interval, given 5 frequency channels, in 10 directions and diagonal Jones. We quantify the decrease in number of DoF when calibrating using `pphase` instead of traditional methods.

to 70% in this scenario.

3.1.2 Solution intervals

In radio interferometry, we often perform calibration with the idea of estimating gains which are assumed to be constant over some interval, rather than at the full time and/or frequency resolution. This premise is almost always necessary to achieve sufficient SNR on each baseline (usually > 3, see Sob et al. 2019, for an analysis), and keep the resulting calibration solutions from becoming too noisy. Thus, longer solution intervals are often employed to avoid ill-conditioned regimes (especially when working in low SNR regimes). However, if the solution interval is too long, small scale (in both time and frequency) variations cannot be captured by

the calibration solutions. Thus, optimal solution intervals are a trade-off between sufficient SNR and the ability to track the corresponding fluctuations (refer to Fig. 5.4 in Heald et al. 2018 for an illustration of solution time interval).

Within our minimisation framework, it is possible to impose solution intervals on the screen parameters, and therefore on the derived gains. Since the frequency variation of the phase is explicitly captured by the $1/\nu$ scaling, this naturally promotes large solution intervals in frequency, possibly as large as the entire observed band. For wideband instruments, this makes it much easier to achieve sufficient SNR for a stable solution. The solution time interval can then be chosen in accordance with the time variability of the ionosphere. It is worth pointing out, however, that in practice one may end up solving over smaller intervals anyway because of memory constraints and limited processing power (discussed later in § 3.3.1). Smaller solution intervals in frequency may also be needed if there are other systematic phase effects in the data not captured by the $1/\nu$ scaling.

Since we construct an independent minimisation problem within each solution interval, all the following formulations are given for a single interval only. The minimisation problem can be recast over an interval by adding a sum over time and frequency samples to Eqn. (3.1). This can be derived by following the formulation of § 2.3 of Smirnov and Tasse (2015).

To sum most of the features of `pphase`, we present Table 3.2. It compares some of the mentioned approaches in § 2.4 with the `pphase` model.

	Facet Calibration	SPAM	SAGECal	`pphase`
Main purpose	Corrected images of the sky	Corrected images of the sky	Source subtraction	Corrected images of the sky
Solving for scalar phases	Y	Y	N	Y
Solving for parallel hand gains	Y	Y	Y	Y
Solving for cross hand gains	N	N	Y	N
Correction (i) and imaging (ii) of visibility data with DDE solutions	Y	Y	N	Y (i), N (ii)
Explicit removal of instrumental effects (clocks)	Y	Y	N	N
Global phase screen modelling	N	Y	N	N
Optimised solver	N	N	Y	N
Solutions obtained for all directions instantaneously	N	N	Y	Y
Works on LOFAR HBA data	Y	N	Y	Y
Solution intervals can vary per direction	Y	Y	Y	N
Solving for amplitude (i) and phases (ii) on different timescales	Y	Y	N	N (i), Y (ii)

TABLE 3.2: Comparison between Facet calibration, SPAM, SAGECal and `pphase`. The first four columns are copied from Table 2, van Weeren et al. (2016).

3.2 Derivatives

The NLLS problem presented in Eqn. (3.1) can be solved using the LM algorithm described in § 2.3. We proceed by deriving the first-order partial derivatives with respect to the parameter $\boldsymbol{\alpha}$, by employing the chain rule,

$$\begin{aligned}
\frac{\partial V^{XX}_{pqt\nu}}{\partial \alpha^{XX}_{pk}} &= \frac{\partial V^{XX}_{pqt\nu}}{\partial \mathcal{E}^{XX}_{dp}} \quad \times \quad \frac{\partial \mathcal{E}^{XX}_{dp}}{\partial \alpha^{XX}_{pk}}, \\
&= \sum_{d=1}^{N_\mathrm{D}} X_{dpq} \left(\mathcal{E}^{XX}_{dq}\right)^H \times i h_{dk} f(\nu) \mathcal{E}^{XX}_{dp}, \\
&= \sum_{d=1}^{N_\mathrm{D}} i h_{dk} f(\nu) \mathcal{E}^{XX}_{dp} X_{dpq} \left(\mathcal{E}^{XX}_{dq}\right)^H; \\
\frac{\partial V^{XX}_{pqt\nu}}{\partial \alpha^{XX}_{qk}} &= \frac{\partial V^{XX}_{pqt\nu}}{\partial \left(\mathcal{E}^{XX}_{dq}\right)^H} \quad \times \quad \frac{\partial \left(\mathcal{E}^{XX}_{dq}\right)^H}{\partial \alpha^{XX}_{qk}}, \\
&= \sum_{d=1}^{N_\mathrm{D}} -i h_{dk} f(\nu) \mathcal{E}^{XX}_{dp} X_{dpq} \left(\mathcal{E}^{XX}_{dq}\right)^H,
\end{aligned} \tag{3.5}$$

for an observed visibility at time t and frequency ν. To simplify further derivations, let $a_{pqk} = \sum_{d=1}^{N_\mathrm{D}} h_{dk} f(\nu) \mathcal{E}_{dp} X_{dpq} \left(\mathcal{E}_{dq}\right)^H$. An equivalent expression for YY can be obtained by following the steps above. Since V_{pq} is Hermitian, we are free to use only the upper or lower half of $\boldsymbol{J}$. Note that the auto-correlations $(pp,\, qq)$ are not used. Within this framework, $\boldsymbol{J}$ is an array of shape (number of visibility data points $\times$ $2N_\mathrm{PAR} N_\mathrm{A} N^t_\mathrm{INT} N^f_\mathrm{INT}$), where N^t_INT and N^f_INT stand for the number of solution time and frequency intervals processed simultaneously. Using Eqn. (3.5), we can rewrite the entries to $\boldsymbol{J}$,

$$\boldsymbol{J}_{[pq][jk]} = \begin{cases} i a_{pqk}, & j = p, j \neq q, \\ -i a_{pqk}, & j = q, j \neq p, \\ 0, & \text{otherwise}, \end{cases} \tag{3.6}$$

for a single correlation, where each entry gives the partial derivatives of the visibilities per baseline $[pq]$ across a given solution interval s with respect to the parameter α_{jk}. In this sense, we are using p, q and j, k to designate the rows and columns respectively, within the s block which will in turn iterate over $N^{tf} = N^t_\mathrm{INT} + N^f_\mathrm{INT}$. As previously discussed in § 2.3, it is common to approximate the Hessian $\mathbf{H}$ of the objective function as $\boldsymbol{J}^H \boldsymbol{J}$. Conforming to a similar formalism from § 2 in Smirnov and Tasse (2015), we present an algorithm below to compute $\boldsymbol{J}^H \boldsymbol{J}$ for a single correlation.

Algorithm 1 Calculate $\boldsymbol{J}^H\boldsymbol{J}$

Require: $N_{\text{PAR}} \neq 0$
Ensure: $\boldsymbol{J}^H\boldsymbol{J} \neq$ singular
  $\boldsymbol{J}^H\boldsymbol{J}_{s,[j_1k_1][j_2k_2]} \leftarrow 0$
  for $s \leftarrow$ 1 **to** N^{tf} **do**
    for $j_1 \leftarrow$ 1 **to** N_{A} **do**
      for $j_2 \leftarrow$ 1 **to** N_{A} **do**
        for $k_1 \leftarrow$ 1 **to** N_{PAR} **do**
          for $k_2 \leftarrow$ 1 **to** N_{PAR} **do**
            if $j_1 = j_2$ **then**
              while $p = j_1$ or $q = j_1$ **do**
                $\boldsymbol{J}^H\boldsymbol{J}_{s,[j_1k_1][j_2k_2]} += a_{pqsk_1}a_{pqsk_2}$
              end while
            else if $p = j_1$ and $q = j_2$ **then**
              $\boldsymbol{J}^H\boldsymbol{J}_{s,[j_1k_1][j_2k_2]} -= a_{pqsk_1}a_{pqsk_2}$
            else if $q = j_1$ and $p = j_2$ **then**
              $\boldsymbol{J}^H\boldsymbol{J}_{s,[j_1k_1][j_2k_2]} -= a_{pqsk_1}a_{pqsk_2}$
            end if
          end for
        end for
      end for
    end for
  end for

In § 3.3.2, we consider a further approximation of this $\boldsymbol{J}^H\boldsymbol{J}$.

3.3 Implementation and verification

`CubiCal` is a radio interferometric calibration suite implemented as a Python package. The suite is accelerated with Numba, which is a just-in-time compiler (Lam et al. 2015). Numba makes it possible to compile Python functions into code executable at speeds comparable to C or Fortran simply by adding a decorator. CubiCal involves several specialised solvers with the main aim of exploiting complex optimisation. The exact details can be found in Kenyon et al. (2018). The `pphase` solver has been implemented in Python and adapted to work in `CubiCal`.

We verify our implementation on a set of simulated data. This has the advantage of ground truth being known, so that an objective evaluation of the solver's performance can be done. To verify if the `pphase` tool works successfully in calibrating radio interferometric data, we consider a case of a known gains' model with $N_{\text{PAR}} = 3$ (as given by Eqn. (3.4)) and $N_{\text{D}} = 5$ randomly generated point sources (each 1 Jy) within a 1° FoV.

The first step in each simulation is creating a measurement set (MS) to store visibilities for the simulation. We create MSs using a Python-based software called `simms`[3]. `simms` is a Python wrapper around the `casacore` tool `makems`[4] used for generating measurements. Similar to `makems`, `simms` takes an antenna table and configuration parameters for our simulated

[3] `https://github.com/ratt-ru/simms`
[4] `https://github.com/ska-sa/makems`

observation. For this simulation, we use `simms` to create a Karoo Array Telescope (KAT-7) observation (Booth et al. 2009). We consider a noise-free MS containing 8 channels starting from 1.4 GHz, of channel width 20 MHz with a synthesis time of 0.5 hours, integrated at every 30 seconds, producing a total of 60 time stamps. We compute the visibilities using `montblanc` (Perkins et al. 2015). The observed visibility data set is constructed (from a custom script

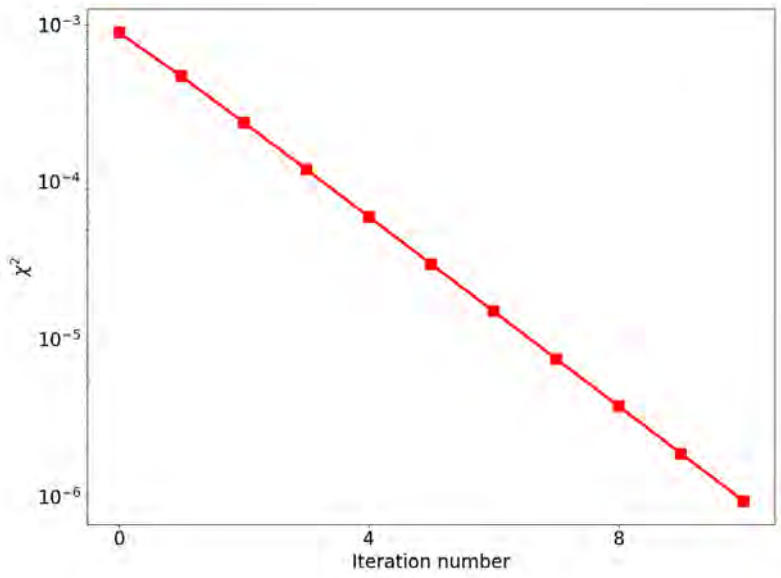


FIGURE 3.2: Plot of χ^2 values obtained at each iteration of the run.

in `codex-africanus`[5], with the same gain model `pphase` uses) using random normal values for $\boldsymbol{\alpha}$ with mean 0 and standard deviation $\sigma_\alpha = 0.2$. This means that the phases are being randomly sampled at each time and antenna. We initialise all initial phase values to zero. We proceed with calibration at full time and frequency resolutions.

Fig. 3.2 shows chi-square as a function of iteration. This confirms that the `pphase` solver has very regular convergence properties, with a monotonically decreasing difference in the squared residual visibilities. We note again that we are using a "full" approximate Hessian as $\boldsymbol{J}^H\boldsymbol{J}$ at this point.

In Fig. 3.3, we can see that the phase differences recovered by the `pphase` solver are almost identical to the ground truth[6]. The differences in the two phase differences are estimated to vary with a standard deviation of about 0.017° and 0.016° corresponding to XX and YY auto-correlations respectively. This is expected for it is a noise-free simulation, and we are calibrating the data with a complete (known) sky model. Note that we only show one source across all the baselines in the figure, but the plots for the remaining sources exhibit the same behaviour.

[5] `https://github.com/landmanbester/codex-africanus/blob/phase_screen_sim/africanus/calibration/utils/examples/apply_phase_screen_to_ms.py`

[6] There is no point in comparing the absolute phases per-antenna since an interferometer is only sensitive to the differences in phases.

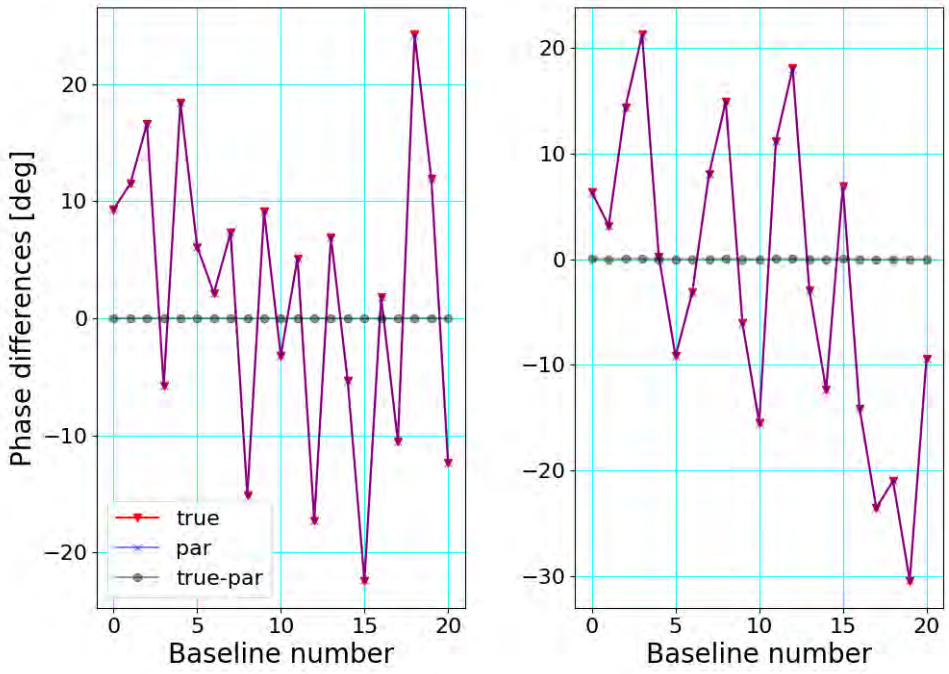


FIGURE 3.3: Plots of phase differences recovered by the `pphase` solver (X) against the true phase differences (▼) for a single source and $\boldsymbol{\alpha}$ values corresponding to XX (left) and YY (right) correlations. The difference between the true and recovered values is given by the black circles.

3.3.1 Numerical and computational considerations

Some computational aspects need to be accounted for in the implementation.

- Given the shape of $\boldsymbol{J}$, a full computation of it can produce considerable memory pressure, and restrict the size of the data chunk that can be loaded into memory at a time. `CubiCal` implements a chunking strategy that iterates through the data in smaller parts.

- As with many such algorithms, issues of numerical conditioning and precision need to be taken into account. In particular, since the l, m coordinates can be small in magnitude, the coefficients representing higher-order polynomial terms in the Jacobian and Hessian can end up differing from those corresponding to the constant phase term by several orders of magnitude. This can lead to very "unbalanced" entries in the Jacobian and Hessian, leading to loss of numerical precision when working with these matrices. This can be prevented by using a simple linear re-scaling of the variables, so that the (l, m) coordinates are normalised to an interval $[0, 1]$, while preserving the phase centre (using a custom function). This does not change the underlying optimisation problem but provides a numerically more balanced matrix.

- Given the 2π ambiguity in phase (as well as an unconstrained overall phase offset inherent to interferometric measurements), there is an infinite number of $\boldsymbol{\alpha}_{\text{SOL}}$ solutions yielding the same model in the visibility space. This means the optimisation problem has no global minimum. Traditional phase-only solvers work around this by, e.g., constraining phase solutions to a $[0, 2\pi)$ interval, and by designating a reference antenna with zero phase. For a parametrised phase solver such as `pphase`, such uniqueness constraints are difficult to propagate from gain-phase space back onto the parameter space. We, therefore, do not attempt to set any constraints on the parameters. Rather, we evaluate the performance of the solver by looking at the derived phase differences.

The numerical approximations to the matrices above are discussed in the next section.

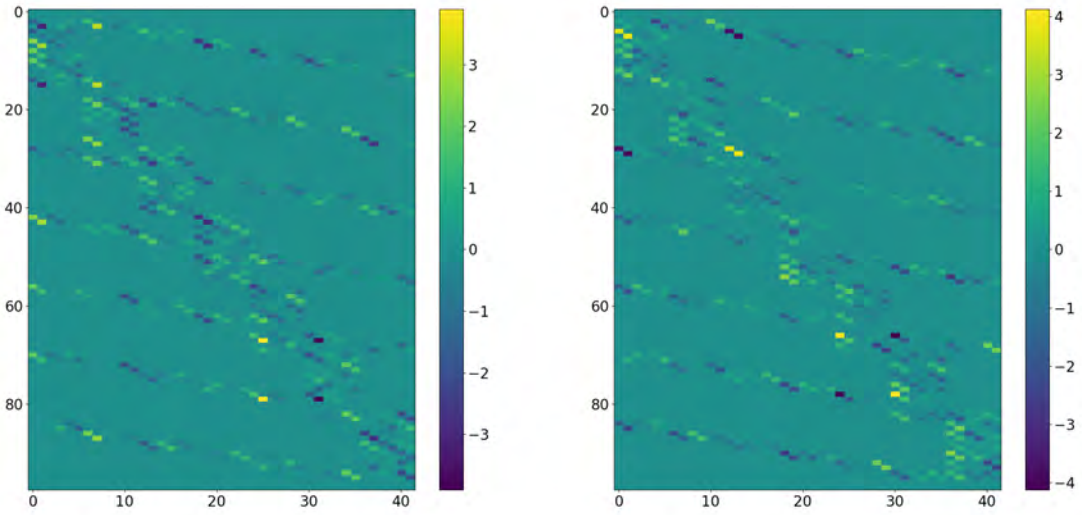
3.3.2 Jacobian, Hessian and approximations

In non-linear optimisation, it is often important to compute large Jacobian and Hessian matrices. Practical considerations such as computational complexity and stability often necessitate certain approximations (mentioned in § 2.3). For example, as already noted, we avoid computing second-order derivatives by approximating the full Hessian with its convex approximation viz. $\boldsymbol{J}^H\boldsymbol{J}$. This is by no means a new idea; there is a wealth of optimisation theory that utilises this approximation for both efficiency and improved convergence properties of the algorithm. However, because of the large problem sizes that are typical in radio interferometry, the $\boldsymbol{J}^H\boldsymbol{J}$ matrix can be very large. Thus, the numerical inversion required for the LM updates is not always feasible and necessitates additional approximations. In this work[7], by analogy with Smirnov and Tasse (2015), we consider approximating the full $\boldsymbol{J}^H\boldsymbol{J}$

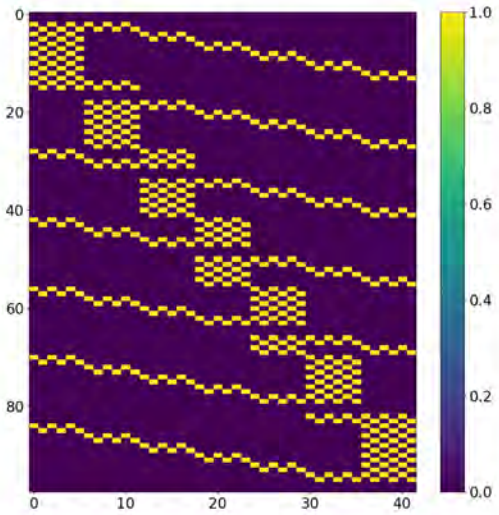
[7]Note that, iterative numerical inversion schemes for symmetric matrices such as the conjugate gradient can also be used. These can be very effective, especially if the sparse nature of the Jacobian is exploited. This is an avenue we intend to explore in future research.

with a block-diagonal matrix where the individual blocks correspond to the covariance between the parameters of a single antenna. This avoids the need for inverting the full Hessian and can result in substantial computational savings. However, we also need to understand the impact this has on the accuracy and validity of the solutions.

Fig. 3.4 shows the Jacobian's layout for the simulation described in § 3.3. The Jacobian possesses relatively few non-zero entries. This is firstly because derivatives corresponding to auto-correlation terms are zero (or not considered). Secondly, by construction in Eqn. (3.6), the derivatives are non-zero only when either of the row antenna indices matches with the column antenna index. Thus, the Jacobian is not very dense, specifically depicted by Fig. 3.4c. This suggests that we can also plan a more efficient implementation of the Jacobian in the future. Fig. 3.5 shows a real diagonally dominant Hessian, which possesses a negligible imaginary component.

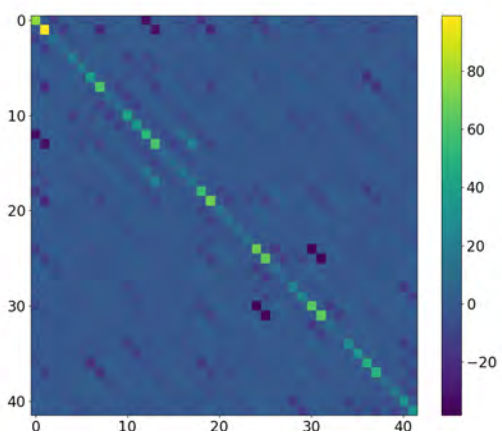


(a) Re($\boldsymbol{J}$). (b) Im($\boldsymbol{J}$).

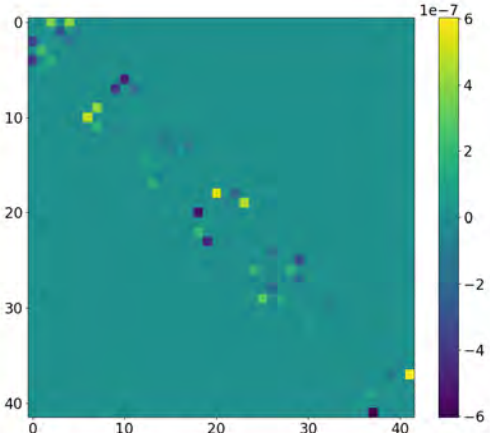


(c) $|\mathrm{Re}(\boldsymbol{J}) > 0|$.

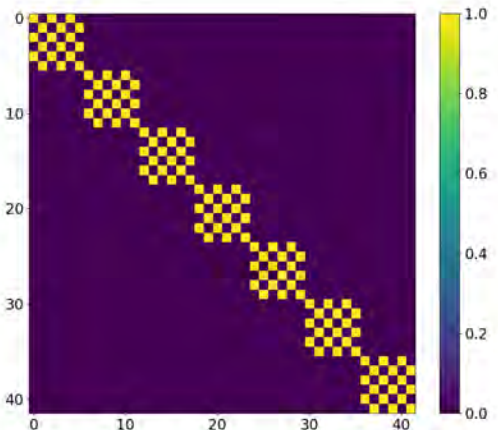
FIGURE 3.4: Plots demonstrating the structure of the Jacobian at one time and frequency stamp and solution interval for the simulation described in § 3.3. (c) is a mask image showing the non-zero entries of the Jacobian. This emphasises the nearly-block-diagonal structure.



(a) $\mathrm{Re}(\boldsymbol{J}^H\boldsymbol{J})$.



(b) $\mathrm{Im}(\boldsymbol{J}^H\boldsymbol{J})$.



(c) $|\mathrm{Re}(\boldsymbol{J}^H\boldsymbol{J}) > 0|$.

FIGURE 3.5: Plots describing the Hessians at one solution interval for simulation in § 3.3. (b) shows that the Hessian has a negligible imaginary part. (c) is a mask image showing the block-diagonal structure being used for the approximation.

Our parametrisation induces an approximately block-diagonal structure, as shown in Fig. 3.5. This indicates high covariance between the screen parameters corresponding to one station, and relatively low covariance between different stations. This is in direct analogy to the CohJones algorithm proposed by Smirnov and Tasse (2015), where the Hessian is approximated by a block-diagonal matrix, with each block corresponding to one antenna. This separates the problem into different subsets of parameters corresponding to different antennae. Given that the cost of the full matrix inversion scales cubically with the size of the matrix, we then invert each block on the diagonal of the Hessian, reducing the size of the problem from $2N_{\mathrm{A}}N_{\mathrm{PAR}} \times 2N_{\mathrm{A}}N_{\mathrm{PAR}}$ for the full matrix to $2N_{\mathrm{PAR}} \times 2N_{\mathrm{PAR}}$ per block/antenna. Another important component in the simulation is the $\boldsymbol{J}^H\mathbf{r}$ component (refer to § 2.3). Where it may be a straightforward dot product in the "full" $\boldsymbol{J}^H\boldsymbol{J}$ case, $\boldsymbol{J}^H\mathbf{r}$ entries are sorted and allocated for the multiplication to the corresponding block of the $\boldsymbol{J}^H\boldsymbol{J}$ in the block-diagonal case.

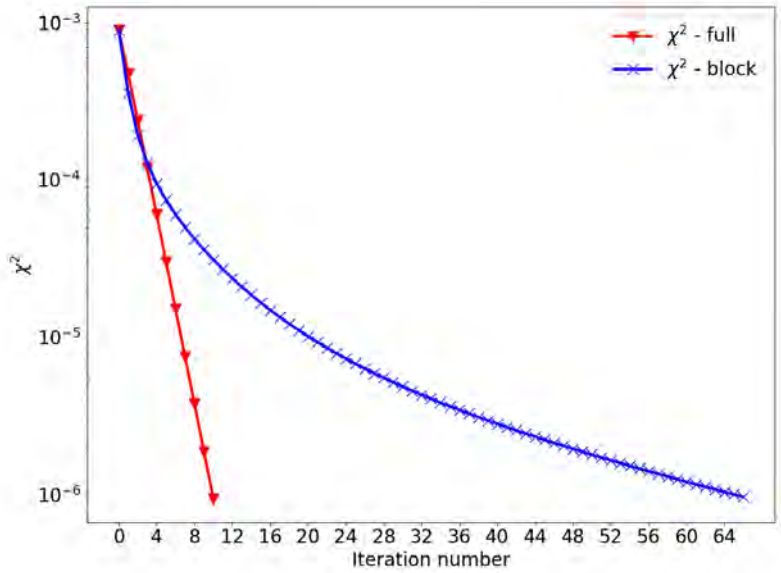


FIGURE 3.6: Plot illustrating the convergence of the GN algorithm when using both full (▼) and block-diagonal (X) Hessians.

Smirnov and Tasse (2015) observe a slower convergence rate when using the block-diagonal approximation of the Hessian, which, in their case, is more than compensated for by the computational savings of inverting a simpler matrix. Here, we would expect to see a similar behaviour. The convergence properties are shown in Fig. 3.6. Curiously, and contrary to Smirnov and Tasse (2015), we observe *better* initial convergence with the block-diagonal approximation (compared to the full $\boldsymbol{J}^H\boldsymbol{J}$) over the first few iterations. After iteration 5, the full implementation starts outperforming the approximation as expected. Note that by this stage, the χ^2 has already improved by at least one order of magnitude. The solutions produced by the full and block-diagonal Hessians have χ^2 values of 9.43×10^{-7} and 9.70×10^{-7}, and can therefore, be considered essentially identical. However, it is not clear if this in a general trend or if it only applies to certain types of problems; for example, in this case where it is a well-understood ideal simulation with known parameters.

3.4 Summary

In this chapter, we introduced the parametrised phase model, `pphase`, and the corresponding solver. We verified that the solver can recover accurate ionospheric phases on simple simulated data. We also discussed the use of approximations and their essential repercussions. We demonstrated how this approximation suits our parametrisation scheme and how, by taking advantage of the approximation, we can reduce our computational load. Finally, we acknowledge that the implementation of `pphase` could be further optimised by utilising iterative numerical inversion schemes and by further exploiting the sparse nature of the Jacobian. This is left as an avenue for future research.

Chapter 4

Application

Sky models are often incomplete (or incorrect) because the sky or calibration parameters are unknown. Incomplete sky models result in bias during traditional radio interferometric gain calibration. Joseph et al. (2018) provide for a theoretical estimate for this bias. These incomplete sky models usually comprise foregrounds that are brighter than faint unmodelled sources, such as the 21 cm signal around star-forming galaxies (recall § 1.1.2). Signals such as the extended diffuse Galactic foreground (even if known) or the 21 cm line are very often not included in the sky model, either because they are too faint or they are too complex to be expressed using a rather limited parametrisation scheme. Grobler et al. (2014) show that *ghost* or spurious sources and loss in real emission usually manifest due to incompleteness of the model. These are usually best observed in the residual images, where we expect noise-like structure, but instead, record some significant structure. This mechanism can be understood as calibration trying to move some of the "real" flux from the modelled sources to compensate for the missing flux of the unmodelled source/s. Flux suppression is another result of the same process. It has been extensively studied because, with the unprecedented sensitivity of the upcoming telescopes, fainter calibration artefacts can no longer be ignored (see Grobler et al. 2014, 2016; Kazemi and Yatawatta 2013; Sob et al. 2019; Wijnholds et al. 2016). Thus, we require enhanced calibration techniques to tackle these issues.

As we noted in the previous chapter, one of the motivations for the `pphase` solver is to make the calibration process more robust to flux suppression compared to a traditional direction-dependent solver. In this chapter, we make use of simulations to demonstrate the superiority of the parametrised solver in terms of mitigating the flux suppression of unmodelled sources.

In practice, any given observation will contain a large number of unmodelled sources. However, following Grobler et al. (2014), we set up an idealised example containing only a single unmodelled source, and we use the same model throughout the chapter. We intend to investigate how ionospheric calibration impacts the morphology and spectral profile of potential HI sources. Thus, we choose to give the unmodelled source both spatial extent and a spectral signature. Since the ionosphere typically has to be treated as a DDE, we include a number of relatively bright point sources into the calibration model. We can find an example of a similar study in Sardarabadi and Koopmans (2018) where the suppression of the 21 cm signal power spectrum is examined.

The chapter is structured as follows. General details for all the simulations are listed in §

4.1. Each of the three remaining sections focusses on a specific aspect of the study. In § 4.2, we examine the suppression of the unmodelled source after calibration. We then repeat this simulation at various solution intervals and SNR regimes in § 4.3. We conclude the chapter in § 4.4 by looking at the spectrum of the unmodelled source after calibration.

4.1 Simulation methodology

In this section, we explain the different methods employed in setting up the simulations. To ensure a fair comparison, we perform all of the simulations using the same set of methods described below.

Creating Measurement Sets

For all the simulations in this chapter, we simulate observations with the VLA in its compact D configuration. The simulations are performed using 10 frequency channels with a starting frequency of 100 MHz, a channel width of 5 MHz, an integration time of 30 seconds and a total synthesis time of 0.5 hours. In all cases, we simulate a field centred at RA 0 and declination +30°. The restoring beam sizes (i.e. a Gaussian matching the FWHM of the main lobe of the PSF) are $9.7' \times 8.7'$ at the lowest frequency and $6.6' \times 6.0'$ at the highest frequency.

Note that this simulation does not match any actual VLA observing mode. The choice of array layout and observing frequency for our simulations is somewhat notional and was driven by purely practical considerations. We wanted a compact array layout (so that images could be reconstructed with a relatively small number of pixels) with enough baselines to constrain the DoF for DD solutions, yet not so many baselines as to make the computational experiments unwieldy. VLA-D is a familiar array layout providing a suitable number of baselines. Observing frequency and bandwidth were chosen to show substantial ionospheric effects.

Sky models

For all simulations, we use a 2° by 2° FoV with a number of modelled sources which are randomly generated point sources (using a custom script). We assume that the sky is unpolarised. Here, we strictly use unpolarised sources because polarised sources can introduce artefacts in the total intensity spectra and studying polarised sources is beyond the scope of this thesis. We also assume that the sources have a flat spectrum, and their total flux density is scaled to sum to 5 Jy. Both the source positions and their brightness follow uniform distributions. It is also assumed that our sources are non-varying (that is, not changing in brightness and structure) in time during the observations. The number of modelled sources differs for simulations 4.2 and 4.3 and is specified within each section.

We represent the unmodelled source by a diffuse Gaussian blob with a specific spectral signature. We place the diffuse Gaussian component at the centre of the FoV, as shown in Fig. 4.2c. We make the "blob" fainter than the modelled sources (0.2 of the total modelled

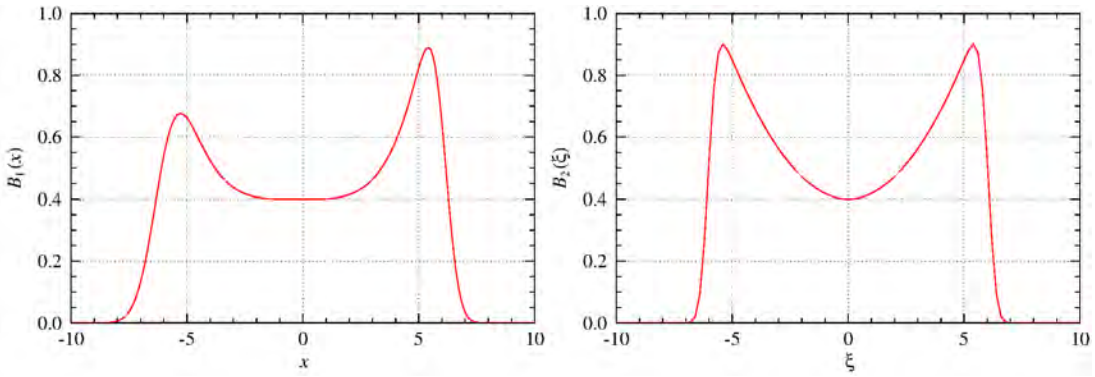


FIGURE 4.1: Examples of an asymmetric generalised busy function and a simplified busy function on the left and right of the image respectively. This image is taken from Westmeier et al. (2013), where the associated parameters are also listed. The busy function discussed is flexible and useful in fitting the spectra of neutral hydrogen of observed galaxies. In this work, we use it to generate a specific and recognisable spectral signature.

flux). We use a busy function similar to Westmeier et al. (2013) for the spectrum (Fig. 4.1). We set the FWHM of the Gaussian component to 0.1 of the FoV. Note that the source is not meant to be astrophysically accurate in any way (its spectrum is reminiscent of an HI galaxy but is much wider, and it is at frequencies where we would rather expect cosmological HI), but we merely seek to make a notional example of extended emission with a spectral signature.

When generating the distribution of modelled sources, we ensure that they are positioned sufficiently far from the blob. This is because the strongest calibration artefacts tend to be generated at or near the positions of the brightest sources. We seek to measure the effect of unmodelled flux suppression in isolation; thus, we do not want to confuse it with bright artefacts associated with modelled sources. The easiest way to avoid such confusion is to ensure the modelled sources and the unmodelled source are spatially separated. Moreover, the effect of the location of the modelled directions on flux suppression can be investigated in future work.

Adding corruptions

We proceed to add ionospheric corruptions using the `pphase` model (from Eqn. (3.3)), with $N_{\mathrm{PAR}} = 3$ (given in Eqn. (3.4)). We compute the visibilities using `montblanc`. The phase screen coefficients, $\boldsymbol{\alpha}$, are drawn from a random normal distribution with mean 0 and standard deviation $\sigma_\alpha = 0.5$. Finally, the visibilities are corrupted with different noise realisations drawn from a random normal distribution with mean 0 and standard deviation σ_{N}. These fairly idealised corruptions are meant for testing and verification purposes, and more realistic versions will be investigated in future work.

Calibration

After corrupting the visibilities, we perform DD calibration using both the `pphase`[1] and direction-dependent phase-only (denoted by `dd-phase`[2]) solvers in `CubiCal`. The maximum number of iterations is set to 50. The convergence criterion is fixed to 10^{-8}. We use unity gains as an initial guess for the `pphase` solver. We ideally expect a monotonic convergence rate. Here, the solvers can terminate for the following reasons: 100% divergence, 100% stalling (no significant changes in the residuals) or the maximum number of iterations reached.

We use `CubiCal` to generate residuals, i.e. visibilities where the modelled sources have been subtracted, using the best-available DD-solutions, with the remaining data not being DD corrected. In the ideal case, these residual visibilities contain only the contribution of the unmodelled source.

Imaging

In order to evaluate the effect of DD solutions on the unmodelled source, it is then necessary to generate images. This complicates an objective evaluation: dirty images are corrupted by convolution with the PSF, while the result of deconvolution is highly dependent on the cleaning settings, particularly since the source is spatially extended. Ideally, we would like to be sure that we really are measuring flux suppression and calibration artefacts, rather than deconvolution artefacts. In real life (and even in our simple simulations), making a clean separation between the two is impossible, since calibration artefacts often limit deconvolution. The best we can do is ensure that the same deconvolution settings are applied throughout so that when we compare the results of different simulations, we have a high degree of confidence that the observed differences are truly due to the calibration.

Having noted that caveat, our imaging is done as follows. We use `WSClean` single-scale mode to image and deconvolve the residuals. The images have a size of 512×512, with pixel size of $100''$. The image size is significantly larger than the extent of the sky model, this is to ensure a cleaning border during deconvolution, as well as to inspect the rest of the FoV for possible ghosts.

For the reasons given below, we employ both natural and Briggs' $r = 0$ weightings in the sections to follow. The former is expected to maximise the recovery of extended emission, while the latter improves the PSF, and thus the deconvolution of any residuals from the compact sources. To be more specific, we present multi-frequency synthesised *model* as well as dirty images in § 4.2 produced using Briggs' $r = 0$ weighting. Here, this is a good choice of weighting as it stands for an equal trade-off between resolution and sensitivity to a weak source like the blob. In § 4.4, we make use of per-frequency channel *model* as well as restored images produced from both Briggs' $r = 0$ and natural weightings. We find it useful to use both weightings for these reasons. We are aware that natural weighting method leads to a

[1] Calibration is carried out using a block-diagonal Hessian (refer to § 3.3.2).

[2] We run the *complex* solver in a *phase-diag* mode at the outset which makes it an equivalent non-parametric phase-only solver. We also use the notation "com" in multiple plots, with reference to the complex solver.

higher SNR optimum image as compared to Briggs' $r = 0$. On the other hand, the single-scale `CLEAN` favours better preservation of the morphology of extended sources when utilising Briggs' $r < 2$ as opposed to natural weighting. Thus, both weightings are important.

To extract further information from the images, we estimate the total flux of the blob by summing all the pixel values within a cross-section region of the images which completely overlaps the blob. The cross-section region is selected using the `tigger` viewer[3] and it is the same for all the images. This process is repeated for the images with blobs recovered by both tools. These sums are discussed accordingly in their sections.

4.2 Suppression of extended emission

The simulation in this section is a proof-of-concept aimed at demonstrating the difference in flux suppression by the `dd-phase` and `pphase` solvers in the presence of an incomplete sky model. The details of the 3 modelled sources used are given in Table 4.1. This simulation is noise-free since we do not want to confuse the effects of varying SNR with model incompleteness in this section. This ensures that our only source of calibration error is the diffuse blob not being present in the sky model. The PSF and a dirty image of the corrupted visibilities are displayed in Fig. 4.2a and Fig. 4.2b respectively. Fig. 4.2c shows a dirty image of the visibilities containing the Gaussian blob only, which we use as the "ground truth" in subsequent comparisons. Note that the blob is present in Fig. 4.2b, but is completely obscured by the PSF sidelobes of the brighter sources.

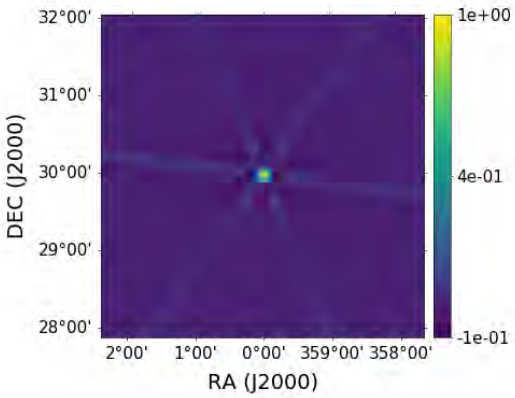
Name	RA [°]	DEC [°]	I [Jy]
A0	−1.74	−0.45	1.95
A1	1.44	0.93	1.61
A2	−1.13	1.35	1.44

TABLE 4.1: List of the modelled sources for simulation 4.2 with respect to the phase centre (given to 2 d.p).

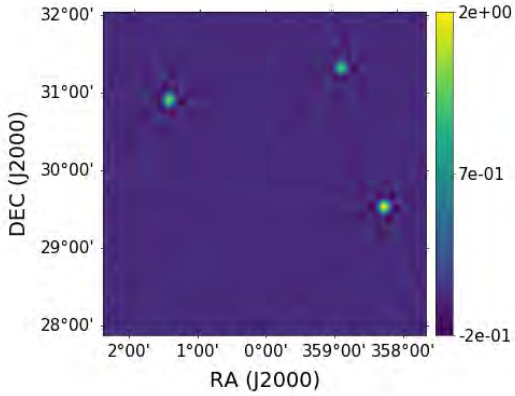
We proceed to calibration. Since we do not want to confuse the effects of solution intervals with incomplete sky models at this moment, and we have sufficiently high SNR, we calibrate at full time and frequency resolutions. Hence, we use solution time and frequency intervals of $t_{\text{INT}} = 1$ (30 s) and $f_{\text{INT}} = 1$ (5 MHz) respectively. For this run, the calibration terminates at the 50^{th} iteration with χ^2 values of 0.01263 and 0.01592 obtained by the `dd-phase` and `pphase` solvers respectively. Since in this case (and in this case only) both solvers have the same number of DoF, the χ^2 values are expected to be comparable.

Since the blob is not modelled, it is expected to be present in the residual images after calibration. We can, therefore, assess the performance of the `pphase` and `dd-phase` solvers in terms of how well they recover the blob's flux, shape and spectrum. Fig. 4.3 shows the *model* images recovered by deconvolution, with (a) obtained by deconvolving the "ground truth" visibilities containing only the uncorrupted blob. The discrete pattern is typical for

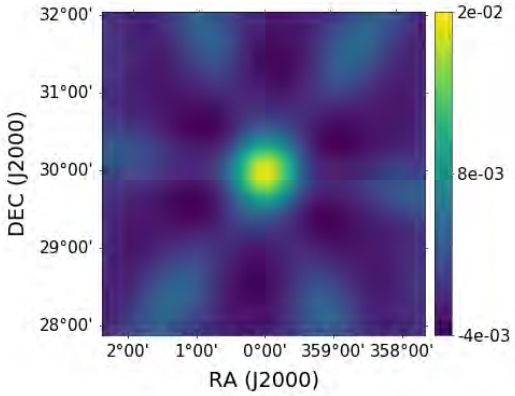
[3] `https://github.com/ska-sa/tigger`



(a) PSF.



(b) Corrupted visibilities.



(c) Gaussian blob.

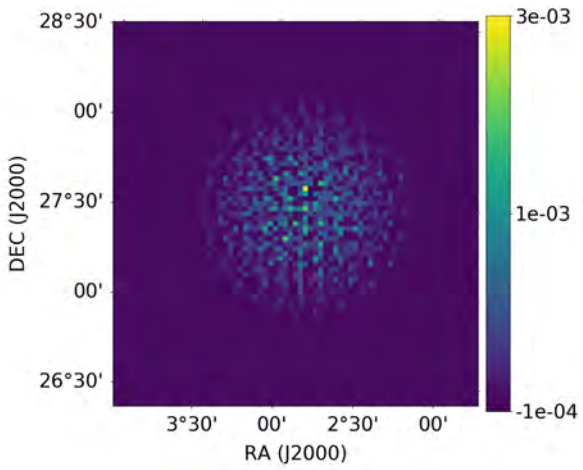
FIGURE 4.2: Following the simulation setup in § 4.2, the PSF (a) and dirty images of the corrupted data (b) and Gaussian blob (c) are shown. The blob is present in (b), but obscured by the sidelobes of the brighter sources.

models recovered by single-scale `CLEAN`[4]. Crucially, we see that the `pphase` solver recovers a blob that is quite similar to the ground truth, while the blob from the `dd-phase` solver is distorted by artefacts.

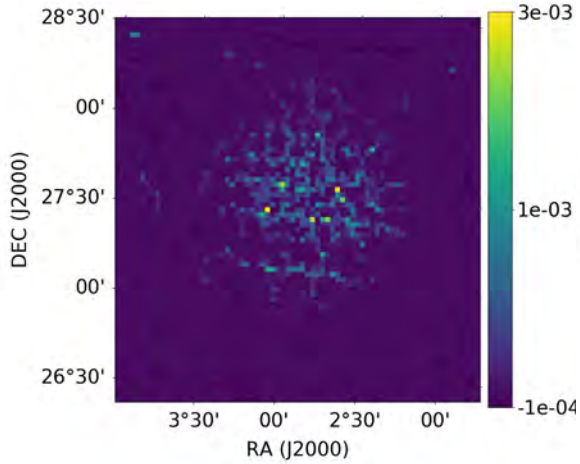
We can put a quantitative measure on how well the blob is recovered by measuring its total flux. We do this by summing the model pixel values. When comparing the % total fluxes to the ground truth, the `dd-phase` solver (Fig. 4.3b) suppresses about 42.7% of the Gaussian blob against 18.9% by the `pphase` solver (Fig. 4.3c). As discussed below, the large degree of suppression most likely stems from the fact that we are solving for individual phase screens at each time. In reality, atmospheric phase screens are expected to vary smoothly with time so the degree of flux suppression can be reduced by using time solution intervals.

One reason for the `pphase` solver outperforming the `dd-phase` solver on flux recovery is

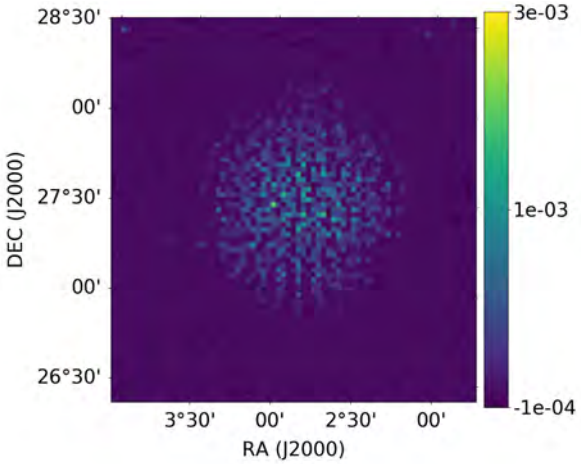
[4]For reasons explained above, we did not experiment with multi-scale `CLEAN` for deconvolution, since that would have introduced more unrelated parameters yet into our experiment.



(a) True blob.



(b) Recovered blob using the `dd-phase` solver.



(c) Recovered blob using the `pphase` solver.

FIGURE 4.3: Residual images after calibration show that `dd-phase` solver (b) tends to cause more flux suppression of the blob as compared to the `pphase` solver (c), as evidenced by the darker areas overlapping the blob. All images share the same colour scale.

doubtless due to the fact that we impose a smooth behaviour of the gains by using parametrisation (refer to Eqn. (3.3)). Here, it is important to acknowledge the limitation of this simulation setup viz. we are solving for a model using the same parametrisation with which we corrupted the data. Since the gains are restricted to follow the prescribed parametrisation, they cannot be arbitrary in the case of the `pphase` solver, unlike the `dd-phase` solver. This discourages the `pphase` gains from absorbing the blob's flux. Likewise, the `dd-phase` solver has a lower χ^2 value than `pphase` solver, as the former overfits the residual visibilities by absorbing any unmodelled fluxes into the gains. This highlights the fact that lower χ^2 values do not necessarily correspond to better calibration. Besides, the smaller but still large degree of % total flux suppression by the `pphase` solver, can be explained by the independent phase screen per antenna at each time. In the next section, it is shown that it is possible to reduce the flux suppression by the `pphase` solver with fewer DoF. For example, this can be done by using solution time intervals, because in reality, the coefficients of the screen vary smoothly with time.

The streak-like artefacts crossing the blob in Fig. 4.3 are reminiscent of PSF sidelobes associated with the model sources. A simplistic explanation for them would be that the model sources have been over- or under-subtracted due to inaccuracies in the gain solutions (which are in turn caused by the unmodelled flux). If these are PSF sidelobes, they would be cleaned away down to some certain threshold! However, artefacts caused by antenna-dependent errors will not deconvolve away. Instead, we are probably seeing the equivalent of primary suppression ghosts (Wijnholds et al. 2016, § 7). Note that even in the simplest case, ghost sources are convolved by their own *Ghost Spread Function* (GSF) (Grobler et al. 2014), which is similar to, but not exactly equal to, the PSF. It is these GSF sidelobes that dominate the dirty images, and they cannot be effectively removed through a deconvolution process. Because the unmodelled flux in our case is extended, the ghost situation becomes a lot more complicated than that considered by Grobler et al. (2014) and Wijnholds et al. (2016), so we do not explore it analytically.

The streaks are quite prominent in Fig. 4.3b, and can be barely seen in Fig. 4.3c. This indicates that the `pphase` solver is a lot less prone to ghost formation in this scenario.

4.3 Varying number of modelled sources

In § 4.2, we performed a high SNR simulation, without adding noise to the visibilities. In this section, we extend the setup of § 4.2 to include noise as an additional corruption. In doing so, we will examine the effects of factors like solution intervals, SNR and model concentration on the performances of both solvers.

Solution intervals are intervals over which we assume that the gains are constant and attempt to provide for a single solution for that specific interval (discussed in § 3.1.2). In low SNR settings, there is tension and trade-off in choice of intervals. On one hand, we require them to be sufficiently long to provide sufficient SNR. On the other hand, we also need solution intervals to be short enough to capture rapid time and frequency variations, for example, in the ionosphere (refer to § 1.4.2). Hence, we investigate how the two solvers behave with different solution intervals, which can be defined over time, frequency or both.

We first discuss solution time intervals. In principle, the integration time is the highest resolution one can seek solutions at. But, it is equally important to note that the integration time is usually much shorter than the rate at which the atmospheric effects vary (for example, minute scales for ionospheric TEC variations or under a minute for scintillation effects (see § 1.4.2)). So, while for low frequency observations the ionospheric phase offsets can vary very rapidly in frequency, with respect to the integration time, the coefficients to the phase screen can vary slower. Since the overall trend in frequency variations can be captured with the frequency scaling, $f(\nu)$, the `pphase` solver can use long intervals both in frequency and in time. In our simulation, however, the coefficients are (identically and independently distributed) randomly being drawn at each time slot (= 30 s) (refer to § 4.1), and we have enough SNR in one integration by construction, so longer time intervals would be neither useful nor helpful. However, with real data, one would still prefer choosing a solution time interval smaller than a minute especially with a highly dynamic ionosphere. Moreover, since

both `dd-phase` and `pphase` solvers treat the time axis the same way, we expect to obtain similar performance at the same solution time interval. Therefore, the solution time interval is fixed at $t_{\text{INT}} = 1$. Following the motivation for the incorporation of frequency scaling in § 3.1.2, we decide to test it by using solution frequency intervals of sizes of 1, 2, 5 and 10 channels corresponding to bandwidths of 5, 10, 25 and 50 MHz respectively. Note that an interval of 10 corresponds to the full bandwidth.

Because of the sensitivity of the modern interferometer, there are usually a large number of sources picked up in any given field. Sob et al. (2019) show that flux suppression of the `dd-phase` solver is a function of the distribution of flux across the field. In particular, the quality of calibration solutions depends not only on the SNR of the sky model but also on the spatial flux distribution. To explore this effect, we create two more sky models of 5 and 7 point sources using the same methods as in § 4.1.

In order to investigate the above, we run different simulations with the same MS but different sky models (3^5, 5 and 7 point sources, plus the blob in each case) and different noise realisations of $\sigma_{\text{N}} = 0, 0.1, 0.5$ (referred as no-noise, high SNR and low SNR simulations respectively). In the image space, these translate to the blob having a peak flux of 2 mJy/beam as compared to the image plane noise having a standard deviation of approximately 0.2 and 1.0 mJy/beam in the cases of high and low SNR respectively. We include a linear ionospheric screen (i.e. three parameters per-antenna as discussed in § 4.1) in the simulations as per the previous section. We then calibrate the simulated data against the point source sky model using the `pphase` and `dd-phase` solvers (with a gain per source in the latter case), using different solution frequency intervals. We then examine the residual visibilities and compare them to the "ground truth" visibilities corresponding to an uncorrupted blob. Furthermore, in the case of noisy simulations, we compare the residual visibilities to the visibilities corresponding to the total of the added noise and uncorrupted blob (let us call it the *noisy blob*).

In order to conduct a systematic comparison across the different scenarios, we would like to use a quantitative metric that is not influenced by the imaging and deconvolution process. Consider that after subtracting the modelled sources (with their respective calibration solutions applied) the residual visibilities should ideally contain the visibilities of the blob plus a realisation of the Gaussian noise (when $\sigma_{\text{N}} > 0$). Thus, as a metric for comparison, we choose to look at the root mean squared error (RMSE) of the visibilities obtained after subtracting the modelled sources with respect to the ground truth visibilities of the blob with the noise realisation added back in. From the simulations and discussion of the previous section, we know that the significant effect on the unmodelled blob is flux suppression, so the RMSE can be expected to be a good proxy for this. We summarise the resulting RMSE metrics for the different combinations of simulations and solution intervals in Fig. 4.4. In the discussion that follows, when we refer to solver "performance", we mean the RMSE, with better performance corresponding to lower RMSE and lower flux suppression.

From the plots, we can make the following observations:

- RMSE is quite similar for the no-noise (red curve) and high SNR (blue curve) scenarios, for both solvers, across all simulations. This suggests that the dominant source of

[5] We are using the same sky model from § 4.2.

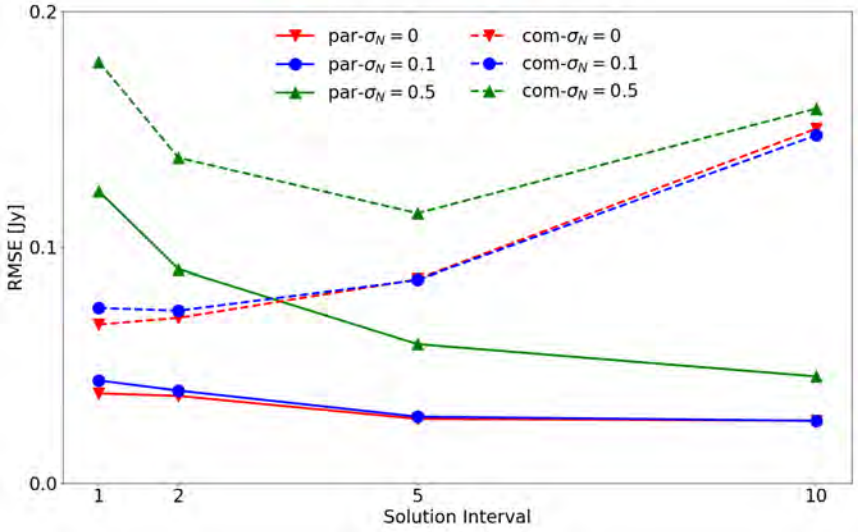
error in these regimes is the contamination of calibration solutions due to unmodelled flux. In the low SNR regime (green curve), thermal noise begins to make a significant contribution to the error budget.

- The performance of the `pphase` solver is continually consistent across all three sky models. This suggests that the parametrised solver has very little sensitivity to spatial flux distribution.

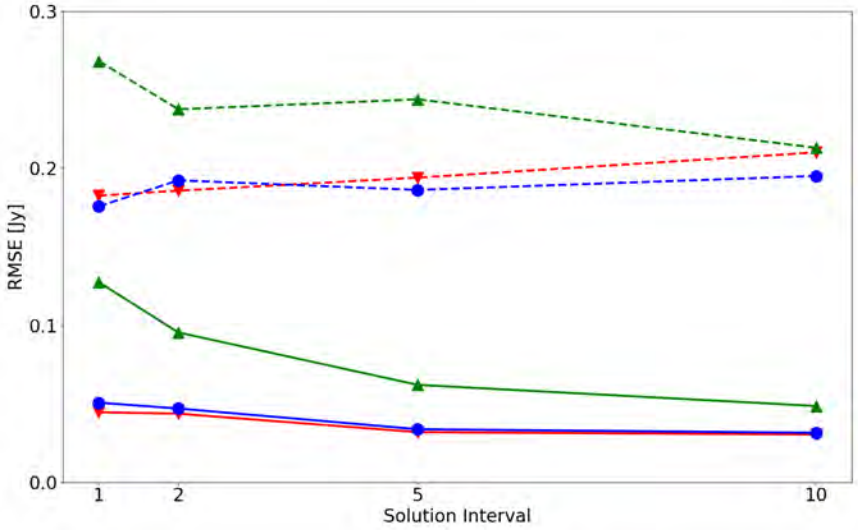
- By contrast, the performance of the `dd-phase` solver varies considerably across the sky models, supporting the findings of Sob et al. (2019). The 5-source scenario generally produces higher RMSE than the 3-source and 7-source scenarios. The trends with respect to solution frequency intervals are also different. The 3-source case nicely illustrates the tension between short and long intervals mentioned above: in the high SNR regime, RMSE consistently increases towards longer intervals (as the solver is less able to capture the frequency variation), while in the low SNR regime, there is a clear "sweet spot" is at $f_{\text{INT}} = 5$, with presumably thermal noise coming to dominate at shorter intervals.

- The performance of `pphase` flattens out asymptotically as the solution interval approaches the full band. This is not surprising, as the $1/\nu$ parametrisation can capture the phase behaviour across the band. With shorter solution intervals, the `pphase` solver has an excess of DoF, and we can expect overfitting and more flux suppression. What is interesting is that in the no-noise and high SNR regimes (red and blue curves), this effect is very mild, and only becomes pronounced in the low SNR regime (green curve). Our interpretation is that the distribution of sources across the FoV is able to constrain the phase screen and keep the solutions stable, even after solving across shorter frequency intervals. As noted earlier, shorter solution frequency intervals may be required in practice to capture any additional phase systematics, so this stability is a useful feature.

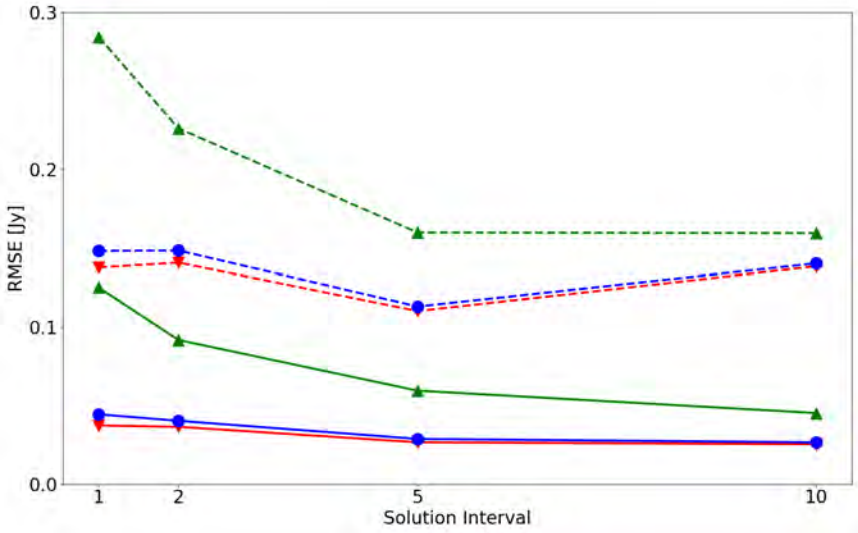
The crucial finding of this section is not just that the `pphase` solver outperforms the `dd-phase` solver in terms of flux suppression (this was expected by construction, and also from the results of the previous section), but also that its suppression behaviour is extremely consistent and predictable, particularly at longer solution intervals (and higher SNR).



(a) Sky model of 3 point sources.



(b) Sky model of 5 point sources.



(c) Sky model of 7 point sources.

FIGURE 4.4: Plots of the RMSE of the difference between the true/noisy and recovered blobs by each solver at different noise realisations (▼ $\sigma_N = 0$, ● $\sigma_N = 0.1$, ▲ $\sigma_N = 0.5$), $f_{INT} = 1, 2, 5, 10$ and sky models of 3, 5 and 7 point sources. The solid and dashed lines correspond to the `pphase` ("par") and `dd-phase` ("com") solvers, respectively.

4.4 Recovered spectrum

In this simulation, we examine the spectra of the blobs as recovered by the two solvers. We focus on one simulation from the previous section: that of 7 modelled sources, with a noise level of $\sigma_{\mathrm{N}} = 0.5$. From Fig. 4.4c, it follows that $f_{\mathrm{INT}} = 5$ is the optimal solution interval for the `dd-phase` solver, as it shows the lowest RMSE, so this is what we use for calibration. Since we want to compare the performance of the `pphase` solver in the same regime, we use the same solution interval.

Following calibration, we image the residual visibilities, and deconvolve them separately per channel, using natural and Briggs' $r = 0$ weightings. We extract spectra in two different ways: (a) peak flux, given by the peak pixel of the blob in the restored image (this is the central pixel of the image), and (b) total flux, given by the sum of the pixels in the model image. The results are presented in Fig. 4.5 and Fig. 4.6.

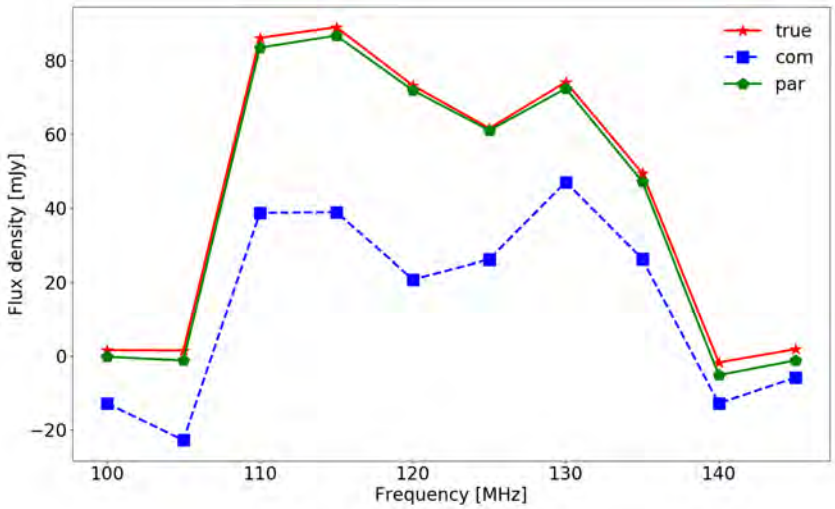
From the results of the previous two sections, we can already anticipate the `pphase` solver doing better, but nonetheless the difference in the spectra is quite striking. The recovered spectrum is extremely close to ground truth, whereas the spectrum resulting from `dd-phase`, while reproducing the overall spectral feature, underestimates it by a very substantial amount.

It is interesting to note that the total flux in the first two and last two channels, which is zero in the simulation by construction (red curve), becomes negative in the spectrum recovered with the `dd-phase` solver (blue curve). This underscores the fact that flux suppression is a bias effect, and not a multiplicative one. We interpret this as follows. Wijnholds et al. (2016) show that flux suppression is a result of gain solutions transferring flux from modelled sources into negative ghosts (also referred to as secondary suppression ghosts) on top of unmodelled sources. The net unmodelled flux across the solution interval ($f_{\mathrm{INT}} = 5$, i.e. half the band) is positive, so a negative ghost is formed, with the same spectral shape as the modelled source, which in our case is spectrally flat. This results in negative net flux in the channels where we had zero flux to begin with.

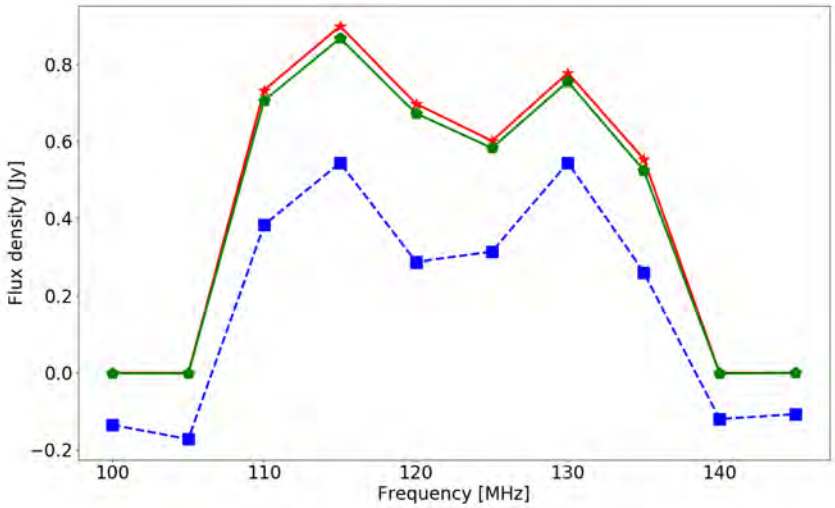
We also note that the suppression effect is seemingly more pronounced with Briggs' weighting. As expected, natural weighting recovers more extended emission in the first place, which we see when comparing the red curves in Fig. 4.5 and Fig. 4.6. However, we can also see even by eye that the blue curve "loses" more flux in the Briggs' weighting case. Recall also that flux suppression across the blob took the form of streaks, i.e. occurred on higher spatial scales (as demonstrated in Fig. 4.3). We can speculate that suppression ghosts tend to follow spatial scales corresponding to the modelled sources (since Grobler et al. (2014) and Wijnholds et al. (2016) only considered point sources, they did not examine this issue). Given that in our simulation the modelled sources were compact, it is to be expected that ghost formation would be more "efficient" at weightings that favour the recovery of compact emission.

A curious question that emerges from this is whether spectral features from the *modelled* sources can get imprinted on the unmodelled sources via this mechanism. Consider a bright model source with a spectral feature in emission: the ghost formation mechanism should then transfer this feature onto the secondary suppression ghost, creating a feature in absorption in the spectra of unmodelled sources. Such an observational scenario would be very interesting

to explore in future work.

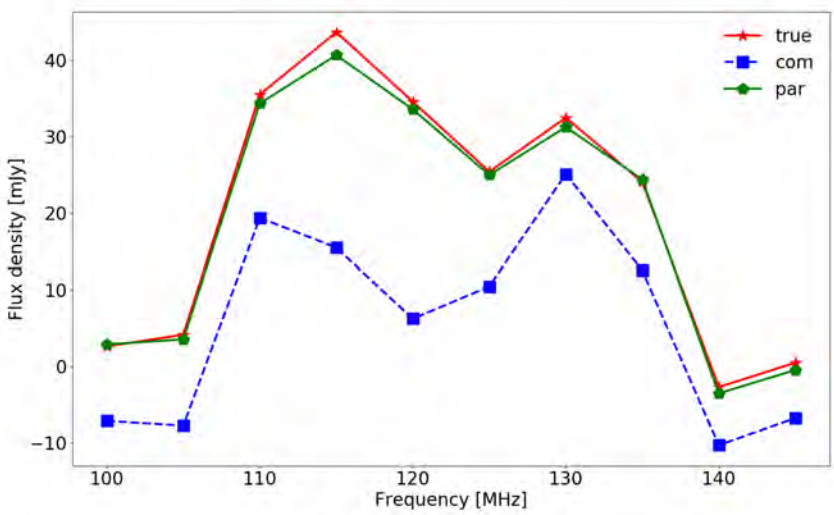


(a) Peak flux density (Jy/beam) as measured from the restored image.

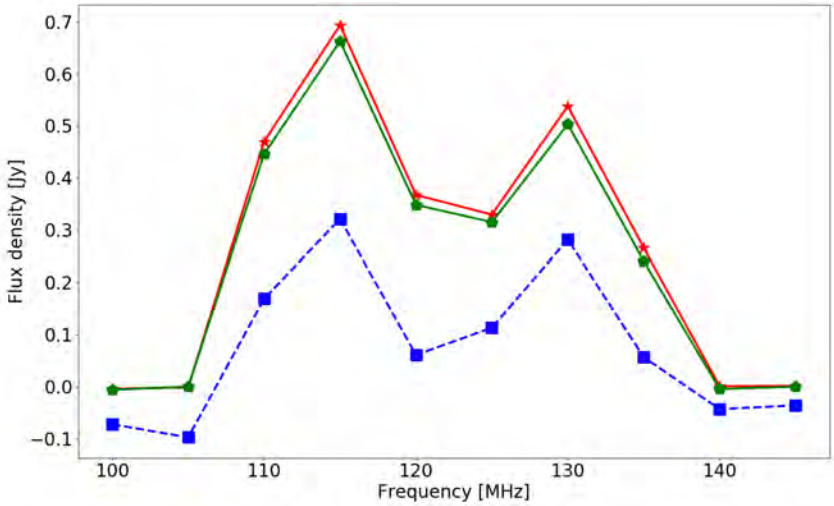


(b) Total flux across the blob.

FIGURE 4.5: Spectra recovered for the Gaussian blob using the `dd-phase` ("com") (■) and `pphase` ("par") (⬟) solvers, compared to ground truth (★). Deconvolution using natural weighting.



(a) Peak flux density (Jy/beam) as measured from the restored image.



(b) Total flux across the blob.

FIGURE 4.6: Spectra recovered for the Gaussian blob using the `dd-phase` ("com") (■) and `pphase` ("par") (⬟) solvers, compared to ground truth (★). Deconvolution using Briggs' $r = 0$ weighting.

Fig. 4.7 illustrates the true and recovered blobs by the solvers in units of visibility amplitude. Given that the recovered blobs in image space give an idea of the locality of the flux suppression, we find the recovered spectra in the visibility space providing a general picture of the suppression, irrespective of the deconvolution artefacts and imperfect subtraction. We observe a similar trend in the recovered spectra, whereby the `pphase` solver outperforms the `dd-phase` solver by a significant amount. Note that the first and last two channels are not empty here, due to the noise realisation.

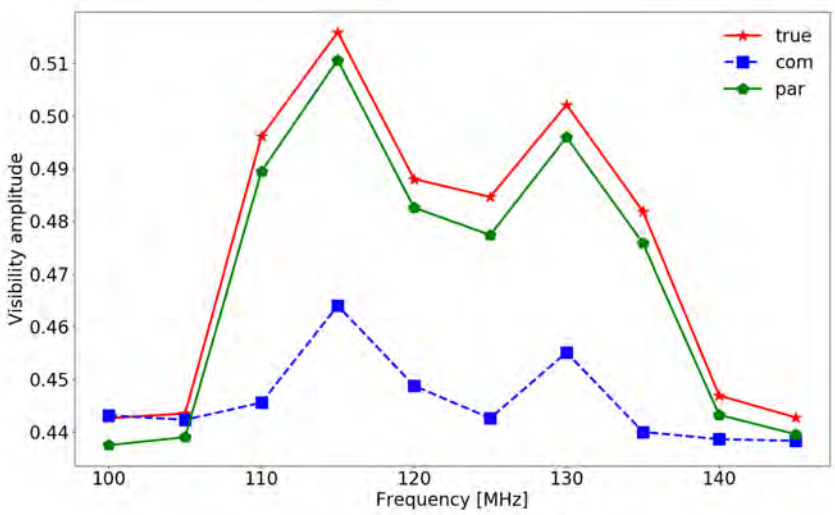


FIGURE 4.7: Spectra recovered for the Gaussian blob, estimated as the mean visibility amplitude value.

4.5 Summary

This chapter presented some simple simulations that demonstrated how some undesirable effects of direction-dependent solvers, such as flux suppression and ghost formation, get propagated onto visibility data. We also showed that a parametrised solver mitigates these effects quite considerably. Although simplistic, these simulations serve as a proof-of-concept to demonstrate the robustness of using parametrised solvers over traditional non-parametric solvers, in addressing flux suppression. The results show that the incorporation of frequency scaling in the gains helps in catching the overall trend in frequency variations. By parametrising the gain phase screen as a smooth function of the spatial coordinates, we are enforcing smoothness of the DDE and also potentially reducing the number of free parameters that need to be solved for.

Of course, given that the parametrisation that we use to simulate the data is the same as the one employed by the `pphase` solver, the result may seem somewhat preordained. However, we should keep in mind that neither solver "knows" about the unmodelled flux in the simulation. What we have demonstrated in this work is that our parametrised solver is considerably better at conserving this unmodelled flux, while fitting the underlying ionospheric effects.

Chapter 5

Conclusions and Future Work

Calibration of radio interferometric data becomes an increasingly difficult task towards lower frequencies (mostly $\leq$ 300 MHz) (discussed in § 1.4 and § 1.5). This is mostly owing to the DDEs introduced by the ionosphere. In this work, we attempted to solve for these ionospheric effects by taking advantage of the *a priori* knowledge that the ionosphere tends to introduce a phase screen (delay) in our response. This screen was parametrised as a function of sky and frequency, mimicking a unique (i.e. not global) screen experienced by each antenna (refer to § 3.1). We verified the implementation of the `pphase` solver in § 3.3 using a simulation with known gains and visibilities for a KAT-7 MS. We also used this fairly small simulation to investigate certain approximations that are required to make our approach tractable for more realistic problems. In particular, we discussed the assumption that the Hessian can be well approximated as a block-diagonal matrix and verified that the approximation does not significantly affect the accuracy of the solutions (see § 3.3.2).

The primary reason for investigating calibration of low frequency observations was to align the project with key science projects that require accurate calibration in low SNR regimes. Here, flux from unmodelled sources often gets absorbed in calibration solutions with unparametrised solvers. In § 4, we use a simulation based on the VLA-D array layout to show that we can decrease flux suppression by carefully parametrising the phase screen. Moreover, the addition of frequency scaling in our scheme made the solver adept in capturing the general frequency trend. It should be noted that, since the simulations carried out in § 4 involved visibilities corrupted with the same gain model that the `pphase` solver uses, it is perhaps not surprising to find that it performs much better than the `dd-phase` solver. Additional testing is definitely required to verify that this holds in the more general case, for example with real data, along a much larger number of directions. However, we expect this result to hold as long as a parametrisation is used that is well adapted to the problem at hand.

The parametrisation of the phase screen considered in this work is not necessarily meant to be realistic. However, since the numerical formalism we have developed is suitable for modelling any linear parameters in the phase screen with an a priori known frequency scaling, it is possible to incorporate more realistic models. For example, following Intema et al. (2009), we could impose a known power spectrum (or equivalently two-point correlation function) for the phase variations. This can be achieved by performing a discrete Karhunen-Loève transform on the covariance matrix that results from evaluating the two-point correlation

function at the pierce points for the calibration problem (refer to the Appendix of Intema et al. 2009 for further details). This results in a well adapted set of orthogonal basis functions which can then be used within our formalism. Additionally, as an extension of the current formalism, we might want to consider a 3D phase screen in order to account for the thickness of the layer as well as the curvature of the Earth.

In addition to considering more physical phase screen models, there are a number of improvements that our implementation of the solver could benefit from. In particular, we could optimise both speed and memory use by using sparse representations of the Jacobians required. Memory usage can be further optimised by representing the Jacobian as an implicit operator which does not have to be loaded into memory all at once. We can explore iterative numerical inversion schemes, which could prove to be more computationally efficient and numerically stable least-squares scenario. We also plan to investigate the use of just-in-time compilation frameworks such as Numba to implement our solver in a flexible and efficient manner. Furthermore, the `pphase` solver as implemented in `CubiCal` is limited to diagonal data and gains. Future work will investigate the possibility of extending this framework to allow for off-diagonal terms in the data, and possibly even cross-correlation terms in the gains. Finally, it is essential to reiterate on the importance of making high fidelity images of science targets, for example the 21 cm signal, where we plan on continuing to focus on peeling sources with DDEs, and applying DD solutions to all regions of the FoV.

Bibliography

Albert, J. G. et al. (Jan. 2020a). "A probabilistic approach to direction-dependent ionospheric calibration". In: *A&A* 633.A77. DOI: https://doi.org/10.1051/0004-6361/201935668.

Albert, J. G. et al. (Mar. 2020b). "Probabilistic direction-dependent ionospheric calibration for LOFAR-HBA". In: *A&A* 635.A147. DOI: https://doi.org/10.1051/0004-6361/201937424.

Amin, F. M. (2012). "Complex-Valued Neural Networks: Learning Algorithms and Applications". PhD thesis. University of Fukui.

Booth, R. S. et al. (Oct. 2009). "MeerKAT Key Project Science, Specifications and Proposals". In: arXiv: 0910.2935.

Born, M. and E. Wolf (2013). *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier.

Brown, R. L. et al. (Nov. 2003). "ALMA - the Atacama Large Millimetre Array". In: *Elsevier*. DOI: 10.1016/j.asr.2003.03.028.

Burke, B. F. and F. Graham-Smith (2010). *An Introduction to Radio Astronomy, 3rd Edition*. Cambridge University Press.

Coker, C. et al. (Feb. 2009). "Simultaneous radio interferometer and optical observations of ionospheric structure at the Very Large Array". In: *Radio Science - RADIO SCI* 44. DOI: 10.1029/2008RS004079.

Condon, J. J. and S. M. Ransom (2016). *Essential Radio Astronomy*. Princeton University Press. URL: https://www.cv.nrao.edu/course/astr534/PDFnew.shtml.

Cornwell, T. J. and P. N. Wilkinson (Sept. 1981). "A new method for making maps with unstable radio interferometers". In: *MNRAS* 196, pp. 1067–1086. DOI: 10.1093/mnras/196.4.1067.

Cotton, W. et al. (2004). "Beyond the Isoplanatic Patch in the VLA Low Frequency Sky Survey". In: *Proceedings of SPIE* 5489, pp. 180–189. DOI: 10.1117/12.551298.

Datta, A. et al. (Oct. 2016). "Effects of the ionosphere on ground-based detection of the global 21 cm signal from the cosmic dawn and the dark ages". In: *ApJS* 831.6. DOI: 10.3847/0004-637X/831/1/6. URL: https://doi.org/10.3847/0004-637X/831/1/6.

Datta-Barua, S. et al. (Oct. 2008). "Bounding higher-order ionosphere errors for the dual-frequency GPS user". In: *Radio Science* 43. DOI: 10.1029/2007RS003772.

De Gasperin, F. et al. (2018). "The effect of the ionosphere on ultra-low frequency radio-interferometric observations". In: *Astronomy & Astrophysics* 615, A179. DOI: 10.1051/0004-6361/201833012. URL: https://doi.org/10.1051/0004-6361/201833012.

DeBoer, D. R. et al. (Mar. 2017). "Hydrogen Epoch of Reionization Array (HERA)". In: *The Astronomical Society of the Pacific* 129.974. DOI: 10.1088/1538-3873/129/974/045001. URL: https://iopscience.iop.org/article/10.1088/1538-3873/129/974/045001.

Di, L. and P. Zhichen (July 2016). "The Five-hundred-metre Aperture Spherical Radio Telescope project". In: *Radio Science* 51, pp. 1060–1064. DOI: 10.1002/2015RS005877.

Duncan, K. J. et al. (Feb. 2019). "The LOFAR Two-metre Sky Survey. IV. First data release: Photometric redshifts and rest-frame magnitudes". In: *A&A* 622.A3. DOI: https://doi.org/10.1051/0004-6361/201833562.

Finsterle, S. (Feb. 2011). "A truncated Levenberg-Marquardt algorithm for the calibration of highly parameterized nonlinear models". In: URL: https://escholarship.org/uc/item/9xv077kc.

Fomalont, E. B. and R. A. Perley (1999). "Calibration and Editing". In: *Synthesis Imaging in Radio Astronomy II*. Ed. by G. B. Taylor et al. Vol. 180. Astronomical Society of the Pacific Conference Series, p. 79.

Grobler, T. L. et al. (Apr. 2014). "Calibration artefacts in radio interferometry - I. Ghost sources in Westerbork Synthesis Radio Telescope data". In: *MNRAS* 439, pp. 4030–4047. DOI: 10.1093/mnras/stu268. arXiv: 1402.1373 [astro-ph.IM].

Grobler, T. L. et al. (Sept. 2016). "Calibration artefacts in radio interferometry - III. Phase-only calibration and primary beam correction". In: *MNRAS* 461, pp. 2975–2992. DOI: 10.1093/mnras/stw1437. arXiv: 1606.06320 [astro-ph.IM].

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

Hamaker, J. P. et al. (May 1996). "Understanding radio polarimetry. I. Mathematical foundations." In: *A&AS* 117, pp. 137–147.

Haynes, M. P. et al. (Jan. 1998). "Asymmetry in High-Precision Global HI Profiles of Isolated Spiral Galaxies". In: *The Astronomical Journal* 115.1, pp. 62–79. DOI: 10.1086/300166.

Heald, G. et al. (2018). *Low Frequency Radio Astronomy and the LOFAR Observatory*. Springer.

Herne, D. et al. (2013). "Ionospheric Phenomena and Low-Frequency Radio Astronomy". In: ed. by W. Short and I. Cairns, pp. 129–140.

Högbom, J. A. (June 1974). "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines". In: *A&AS* 15, p. 417.

Intema, H. T. et al. (July 2009). "Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results". In: *A&A* 501, pp. 1185–1205. DOI: 10.1051/0004-6361/200811094. arXiv: 0904.3975 [astro-ph.IM].

Jansky, K. G. (1933). "Electrical disturbances apparently of extraterrestrial origin". In: *Proceedings of the Institute of Radio Engineers* 21.10, pp. 1387–1398.

Joseph, R. C. et al. (Nov. 2018). "The bias and uncertainty of redundant and sky-based calibration under realistic sky and telescope conditions". In: *The Astronomical Journal* 156.6, p. 285. DOI: 10.3847/1538-3881/aaec0b. URL: https://iopscience.iop.org/article/10.3847/1538-3881/aaec0b.

Kazemi, S. and S. Yatawatta (Oct. 2013). "Robust radio interferometric calibration using the t-distribution". In: *MNRAS* 435, pp. 597–605. DOI: 10.1093/mnras/stt1347. arXiv: 1307.5040 [astro-ph.IM].

Kazemi, S. et al. (June 2011). "Radio interferometric calibration using the SAGE algorithm". In: *MNRAS* 414, pp. 1656–1666. DOI: 10.1111/j.1365-2966.2011.18506.x. arXiv: 1012.1722 [astro-ph.IM].

Kazemi, S. et al. (2013). "Clustered calibration: an improvement to radio interferometric direction-dependent self-calibration". In: *MNRAS* 430, pp. 1457–1472.

Kenyon, J. S. et al. (Aug. 2018). "CubiCal - fast radio interferometric calibration suite exploiting complex optimization". In: *MNRAS* 478, pp. 2399–2415. DOI: 10.1093/mnras/sty1221. arXiv: 1805.03410 [astro-ph.IM].

Lam, S. K. et al. (2015). "Numba: A LLVM-based Python JIT Compiler". In: 9781450340052. DOI: 10.1145/2833157.2833162.

Levenberg, K. (1944). "A method for the solution of certain non-linear problems in least squares". In: *Quarterly of applied mathematics* 2.2, pp. 164–168.

Liebe, H. J. (1989). "An Atmospheric Millimetre-Wave Propagation Model". In: *International Journal of Infrared, Millimetre and Terahertz Waves* 10, pp. 631–650.

Loi, S. T. et al. (Nov. 2015). "Quantifying ionospheric effects on time-domain astrophysics with the Murchison Widefield Array". In: *Monthly Notices of the Royal Astronomical Society* 453 (3), pp. 2731–2746. DOI: 10.1093/mnras/stv1808.

Lonsdale, C. J. (Dec. 2005). "Configuration Considerations for Low Frequency Arrays". In: *From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science*. Vol. 345, p. 399.

Madsen, K. et al. (2004). *Methods for Non-Linear Least Squares Problems (2nd ed.)* eng.

Marquardt, D. W. (1963). "An algorithm for least-squares estimation of nonlinear parameters". In: *Journal of the society for Industrial and Applied Mathematics* 11.2, pp. 431–441.

Martin, P. L. et al. (2016). "Limits on the validity of the thin-layer model of the ionosphere for radio interferometric calibration". In: 459, pp. 3525 –3531. ISSN: 4. DOI: https://doi.org/10.1093/mnras/stw853.

Mesinger, A., ed. (2019). *The Cosmic 21-cm Revolution*. 2514-3433. IOP Publishing. ISBN: 978-0-7503-2236-2. DOI: 10.1088/2514-3433/ab4a73. URL: http://dx.doi.org/10.1088/2514-3433/ab4a73.

Nocedal, J. and S. J. Wright (2000). *Numerical Optimisation*. Springer.

Noordam, J. E. (Oct. 2004). "LOFAR calibration challenges". In: *Ground-based Telescopes*. Ed. by J. M. Oschmann Jr. Vol. 5489. Proc. SPIE, pp. 817–825. DOI: 10.1117/12.544262.

Noordam, J. E. and O. M. Smirnov (Dec. 2010). "The MeqTrees software system and its use for third-generation calibration of radio interferometers". In: *A&A* 524, A61. DOI: 10.1051/0004-6361/201015013. arXiv: 1101.1745 [astro-ph.IM].

Offringa, A. R. et al. (Oct. 2014). "WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy". In: *MNRAS* 444, pp. 606–619. DOI: 10.1093/mnras/stu1368.

Parsons, A. R. et al. (Apr. 2010). "The Precision Array for Probing the Epoch of Reionisation: Eight Station Results". In: *The Astronomical Journal* 139 (4), pp. 1468–1480. DOI: 10.1088/0004-6256/139/4/1468.

Patil, A. H. et al. (Dec. 2016). "Systematic biases in low frequency radio interferometric data due to calibration: the LOFAR EoR case". In: *MNRAS* 463, pp. 4317–4330. URL: https://doi.org/10.1093/mnras/stw2277.

Perkins, S. J. et al. (Sept. 2015). "Montblanc: GPU accelerated radio interferometer measurement equations in support of Bayesian inference for radio observations". In: *Astronomy and Computing* 12, pp. 73–85. DOI: 10.1016/j.ascom.2015.06.003. arXiv: 1501.07719 [cs.DC].

Perley, R. A. et al. (Aug. 2011). "The Expanded Very Large Array - A new telescope for new science". In: *The Astrophysical Journal* 739.1. DOI: 10.1088/2041-8205/739/1/l1. URL: https://iopscience.iop.org/article/10.1088/2041-8205/739/1/L1.

Press, W. H. (Jan. 1978). "Flicker noises in astronomy and elsewhere". In: *Comments on Astrophysics* 7.4, pp. 103–119.

Reber, Grote (1940). "Cosmic static". In: *Proceedings of the IRE* 28.2, pp. 68–70.

Ryle, M. and A. Hewish. "The Synthesis of Large Radio Telescopes". In: *MNRAS* 120 (3), pp. 220–230.

Ryle, M. et al. (Jan. 1960). "The Synthesis of Large Radio Telescopes by the Use of Radio Interferometers". In: *Antennas and Propagation, IRE Transactions on* 7, pp. 120–124. DOI: 10.1109/TAP.1959.1144745.

Sardarabadi, A. M. and L. V. E. Koopmans (Dec. 2018). "Quantifying suppression of the cosmological 21-cm signal due to direction-dependent gain calibration in radio interferometers". In: *MNRAS* 483, pp. 5480–5490. URL: https://doi.org/10.1093/mnras/sty3444.

Schwab, F. R. (Jan. 1980). "Adaptive calibration of radio interferometer data". In: *1980 International Optical Computing Conference I*. Ed. by W. T. Rhodes. Vol. 231. Proc. SPIE, pp. 18–25. DOI: 10.1117/12.958828.

Shapley, P. (2011). *Composition of the Atmosphere*. URL: http://butane.chem.uiuc.edu/pshapley/GenChem1/L9/1.html.

Shimwell, T. et al. (Feb. 2017). "The LOFAR Two-metre Sky Survey. I. Survey description and preliminary data release". In: *A&A* 598.A104. DOI: https://doi.org/10.1051/0004-6361/201629313.

Shimwell, T. et al. (Sept. 2018). "The LOFAR Two-metre Sky Survey. II. First data release". In: *A&A* 622.A1. DOI: https://doi.org/10.1051/0004-6361/201833559.

Skone, S. et al. (Sept. 2005). "Investigating the impact of ionospheric scintillation using a GPS software receiver". In: *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005)*, pp. 1126–1137. URL: https://www.ion.org/publications/abstract.cfm?articleID=6308.

Smirnov, O. M. (Mar. 2011a). "Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism". In: *A&A* 527, A106. DOI: 10.1051/0004-6361/201016082. arXiv: 1101.1764 [astro-ph.IM].

— (Mar. 2011b). "Revisiting the radio interferometer measurement equation. II. Calibration and direction-dependent effects". In: *A&A* 527, A107. DOI: 10.1051/0004-6361/201116434. arXiv: 1101.1765 [astro-ph.IM].

— (Mar. 2011c). "Revisiting the radio interferometer measurement equation. III. Addressing direction-dependent effects in 21 cm WSRT observations of 3C 147". In: *A&A* 527, A108. DOI: 10.1051/0004-6361/201116435. arXiv: 1101.1768 [astro-ph.IM].

Smirnov, O. M. and C. Tasse (May 2015). "Radio interferometric gain calibration as a complex optimization problem". In: *MNRAS* 449, pp. 2668–2684. DOI: 10.1093/mnras/stv418. arXiv: 1502.06974 [astro-ph.IM].

Sob, U. M. et al. (Oct. 2019). "Radio interferometric calibration using a complex Student's t-distribution and Wirtinger derivatives". In: *Monthly Notices of the Royal Astronomical Society* 491.1, pp. 1026–1042. DOI: 10.1093/mnras/stz3037. URL: https://doi.org/10.1093/mnras/stz3037.

Stewart, I. M. et al. (July 2014). "A simple model for global HI profiles of galaxies". In: *A&A* 567.A61. DOI: 10.1051/0004-6361/201423602.

Swarup, G. (1991). "Giant Metrewave Radio Telescope (GMRT)". In: *International Astronomical Union Colloquium* 131, pp. 376–380. DOI: 10.1017/S0252921100013671.

Tasse, C. et al. (Apr. 2018). "Faceting for direction-dependent spectral deconvolution". In: *A&A* 611, A87. DOI: 10.1051/0004-6361/201731474. arXiv: 1712.02078 [astro-ph.IM].

Thompson, A. R. et al. (2017). *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*. DOI: 10.1007/978-3-319-44431-4.

Tingay, S. J. et al. (Jan. 2013). "The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies". In: *Publ. Astron. Soc. Australia* 30. DOI: https://doi.org/10.1017/pasa.2012.007.

van Haarlem, M. P. et al. (Aug. 2013). "LOFAR: The LOw-Frequency ARray". In: *A&A* 556.A2. DOI: https://doi.org/10.1051/0004-6361/201220873.

van Weeren, R. J. et al. (Mar. 2016). "LOFAR Facet Calibration". In: *ApJS* 223, p. 2. DOI: 10.3847/0067-0049/223/1/2. arXiv: 1601.05422 [astro-ph.IM].

Wehbe, B. et al. (Nov. 2019). "The impact of atmospheric dispersion in the performance of high-resolution spectrographs". In: *MNRAS*. arXiv: 1911.08391.

Westmeier, T. et al. (Nov. 2013). "The busy function: a new analytic function for describing the integrated 21-cm spectral profile of galaxies". In: *MNRAS* 438.2, pp. 1176–1190. DOI: 10.1051/0004-6361/201423503. arXiv: 1311.5308 [astro-ph.IM].

Wijnholds, S. et al. (Apr. 2010). "Calibration challenges for the next generation of radio telescopes". In: *IEEE Signal Processing Magazine* 27.9. arXiv: 1004.0156.

Wijnholds, S. J. et al. (Apr. 2016). "Calibration artefacts in radio interferometry - II. Ghost patterns for irregular arrays". In: *MNRAS* 457, pp. 2331–2354. DOI: 10.1093/mnras/stw118.

Williams, W. L. et al. (Nov. 2018). "The LOFAR Two-metre Sky Survey. III. First data release: optical/IR identifications and value-added catalogue". In: *A&A* 622.A1. DOI: https://doi.org/10.1051/0004-6361/201833564.

Wilson, T. L. et al. (2009). *Tools of radio astronomy*. Vol. 5. Springer.

Yatawatta, S. et al. (Oct. 2008). "Radio Interferometric Calibration Using The SAGE Algorithm". In: *ArXiv e-prints*. arXiv: 0810.5751.

Zhongbo, H. and J. J. Mallorqui (Oct. 2019). "An accurate method to correct for atmospheric phase delay for InSAR with the ERA5 Global Atmospheric Model". In: *Remote Sensing* 11.17 (2072-4292). DOI: 10.3390/rs11171969.