Performance evaluation of baseline-dependent window functions with several weighing functions

A Thesis Submitted

in Fulfilment of the Requirements

for the Degree of

MASTERS OF SCIENCE

at

Rhodes University

by

Vanqa Kamvalethu

G16v3668

Supervised by

Prof Oleg Smirnov & Dr Marcellin Atemkeng

Department of Pure and Applied Mathematics

[February 2023]

DECLARATION

I, Vanqa Kamvalethu, with this declare that this thesis entitled "Performance evaluation of baseline-dependent window functions with several weighing functions ", submitted to Rhodes University towards the requirement of Masters of Science in Applied Mathematics is an original work carried out by me under the supervision of Dr M.T Atemkeng and Prof O.M Smirnov and has not formed the basis for the award of any degree or diploma, in this or any other university. I sincerely endorse the ethics and standards of academic research at Rhodes University. Whenever an existing idea, concept, external information, phrase, or result is used, I properly acknowledge and cite it.

Author **[Vanqa Kamvalethu]**

[February 2023]

ACKNOWLEDGEMENT

The existence and successful completion of this thesis necessitates that I acknowledge all the personages that contributed. I, therefore, extend my sincere heartfelt obligation toward my supervisors Dr Atemkeng and Prof Smirnov, for allowing me to do this project and providing all the required facilities, support, guidance and constant encouragement during my master's journey. I am genuinely grateful.

I am immensely grateful to the Rhodes university centre for Radio Astronomy Techniques and Technologies (RATT) group for their constant support and encouragement through our weekly student journals, shared interferometric resources and our quarterly busy weeks. I am also grateful to the Department of Physics and Electronics for providing me with a conducive working environment. My peers, and RATT members, thanks for the support, especially appreciation to Ntsikelelo Charles; thank you, for reading my thesis and sharing your experience.

During my two years of master's studies, I was financially supported by the South African Radio Astronomy Observatory (SARAO) in the form of travel grants to conferences, living allowance and working equipment. Without the SARAO financial assistance, I would not have reached this far; thank you very much SARAO.

Last but not least, I would like to express my deepest appreciation to our project administrator, Ms Zizipo Lusizi and the administrator from the Maths department Loyiso Nkayi. Thank you guys for helping me throughout this time by making sure accommodation and transportation were arranged for the conferences I attended.

ABSTRACT

Radio interferometric data volume is exponentially increasing with the potential to cause slow processing and data storage issues for radio observations recorded at high time and frequency resolutions. This necessitates that a sort of data compression is imposed. The conventional method to compress the data is averaging across time and frequency. However, this results in amplitude loss and source distortion at the edges of the field of view. To reduce amplitude loss and source distortion, baseline-dependent window functions (BDWFs) are proposed in the literature. BDWFs are visibility data compression methods using window functions to retain the signals within a field of interest (FoI) and to suppress signals outside this FoI. However, BDWFs are used with window functions as discussed in the signal processing field without any optimization. This thesis evaluates the performance of BDWFs and then proposes to use machine learning with gradient descent to optimize the window functions employed in BDWFs. Results show that the convergence of the objective function is limited due to the band-limited nature of the window functions in the Fourier space. BDWFs performance is also investigated and discussed using several weighting schemes. Results show that there exists an optimal parameter tuning (not necessarily unique) that suggests an optimal combination of BDWFs and density sampling. With this, ~ 4 % smearing is observed within the FoI, and ~ 80 % source suppression is achieved outside the FoI using the MeerKAT telescope at 1.4 GHz, sampled at 1 s and 184.3 kHz then averaged with BDWFs to achieve a compression factor of 4 in time and 3 in frequency.

Keywords: [Radio Interferometer, Smearing, BDWFs, Optimization, FoI shaping]

Contents

List of Figures

List of Tables

Abbreviations

ANN **A**rtificial **N**ueral **N**etwork

BDWF **B**aseline **D**ependent **W**indow **F**unction

DSP **D**igital **S**ignal **P**rocessing

FFT **F**ast **F**ourier **T**ransform

FoI **F**ield of **I**nterest

FoV **F**ield of **V**iew

FT **F**ourier **T**ransform

FWHM **F**ull **W**idth at **H**alf **M**aximum

IPR **I**mage **P**lane **R**esponse

JVLA **J**ansky **V**ery **L**arge **A**rray

LoFAR **L**ow **F**requency **A**rray

MLW **M**ain **L**obe **W**idth

MS **M**easurement **S**et

PSL **P**eak **S**idelobe **L**evel

RIME **R**adio **I**nterferometer **M**easurement **E**quation

SLR **S**idelobes **R**oll-off **R**ate

SLROR **S**idelobes **L**evel **R**oll-off **R**ate

SKA **S**quare **K**ilometre **A**rray

VLBI **V**ery **L**ong **B**aseline **I**nterferometer

WF **W**indow **F**unctions

Symbols

Description	*Variable*
angular resolution	θ
wavelength	λ
antenna indexes	pq
number of timeslots	N_t
integration time	Δt
number of channels	N_ν
bandwidth	$\Delta\nu$
frequency	ν
baseline-dependent decorrelation	D_{pq}
baseline-dependent decorrelation for BDWF	D^{α}_{pq}
noise penalty	$\mathcal{E}^{W}_{\mathcal{T}}$
pixel noise	σ_{pix}
visibility function for baseline pq	$\mathcal{V}_{pq}$
baseline vector	$\mathbf{u}_{pq}$
maximum baseline length	$\|\|\mathbf{u}_{max}\|\|$
boxcar window	Π
boxcar window in uv-plane	Π^{uv}_{pq}
robustness	R
sampling bin in the regular plane	$B^{\Delta t\Delta\nu}_{kl}$
weighting function for gridded visibility pq	$\mathcal{W}_{pq}$
shah or comb function	Ш
half phase difference in time and frequency	$\Delta\Phi, \Delta\psi$
activation function	φ
weights matrix	w_{km}
bias	b_k
cost function	E_f

CHAPTER 1

General Introduction

The essence of this thesis is to explore radio interferometric techniques such as radio interferometric visibility data compression together with signal processing techniques such as window functions. It is of paramount interest to find the best link in how the two disciplines can be connected and modelled to perform low loss data compression. Fundamentally this is essentially *optimization*. Optimization is exploring the most desirable possible solution to a problem with as minimal cost as possible. In this setting, we are concerned about optimizing certain features of data compression models by finding an optimal matched filter that will optimally re-sample the visibility data, which is a mathematical model that governs the nature of the interferometric measurements. Also, by optimization, we desire to find the best combination of Fourier space parameters, visibility sampling and weighting parameters. All this optimization is done to achieve a nearly lossless data compression scheme. This compression has the potential to allow for long-term storage of compressed visibilities and therefore allow an opportunity for later reprocessing with improved algorithms.

1.1 The origins of radio astronomy

Early in 1932, radio astronomy was established with an experiment conducted by Jansky (1933) at Bell Telephone Laboratories (Akhmanw & Khokblov, 1959). Jansky built the very first radio telescope, and this instrument was primarily designed to trace cosmic interference in radio communications. The experimental aim was to tackle the problem of sources of static interference in radio communications. In the process, he discovered far-off thunderstorm static as one of the causes of this interference. Jansky accidentally detected radio waves/radiation from an astronomical source in the experiment. The instrument was then also used to detect natural radio emissions from objects located light-years away from our planet. Jansky observed signals from a source that repeated at a rate of almost 24 hours, going through all compass points. After eliminating the sun as the source of the static, he deduced that the static source was not from our solar system. The signal was indeed propagating from our galaxy centre (Milky Way, Sullivan III (2009), Akhmanw & Khokblov (1959)). Jansky could not accurately measure the location, the flux density, and other properties of the radio source but only knew the existence of a radio source that dissipates radiations across the band at a specific frequency. To be able to measure such properties in detail, one needs an angular resolution, which is obtained by:

$$\theta \approx \frac{\lambda}{D}, \tag{1.1}$$

where λ represents the wavelength of a radio wave, and D is the diameter of a dish. By definition, Equation (1.1) measures the ability to discern fine detail in a radio source's structure, measured in radians. To obtain higher resolution, the diameter D of a dish must be large enough, meaning that a single-dish telescope has to have a large enough aperture/diameter. The Byrd Green Bank Telescope (Prestage et al., 2009) is the biggest movable radio telescope, of dimensions 100 by 110 meters. This is followed by the Effelsberg telescope, with an aperture of 100 meters, followed by the Jodrell Bank Lovell Telescope (Lovell, 1985), with 76 meter-diameter. Steerable telescopes of this size represent the limits of engineering, and they only attain limited resolution. In other words, at 21 cm wavelength, a telescope with a 100 meters dish size has an angular resolution of about seven arcminutes. Calls for big-sized-diameter telescopes (e.g., the Jodrell

Bank Mark IV and V telescopes at 305 and 122 meters, respectively) posed insurmountable design, manufacturing and maintenance challenges, and have also proven financially intractable. Hence, the approach to linking multiple telescopes through interferometry (Thompson (1999), Thompson et al. (1999)). As a result, at high frequency, interferometers such as Very Large arrays (JVLA, Napier et al. (1983), Thompson et al. (1980)) allow efficient imaging at arcseconds resolution, and large apertures can be utilized by combining the outputs of much smaller sizes, less than 20 meters in diameter. Hence, the precursor MeerKAT (Booth & Jonas, 2012)) and the future Square Kilometer Array (SKA, Dewdney et al. (2009)). The SKA will be one of the most gigantic and sensitive radio interferometers ever. SKA is expected to have thousands of single antenna dishes, achieving extremely high sensitivity and resolution (Krichbaum et al., 1999).

1.2 What is SKA

The SKA will be one of the biggest radio interferometers ever to be built in history, the beginning of what will become the most sophisticated intelligent network in existence. This will consist of two separate telescopes in its first development stage, covering the radio spectrum in the range of centimetres to meters of wavelengths. In the first phase, there is a SKA1-LOW telescope to be built at the Murchison Radio Observatory in Western Australia and will cover frequencies in the range of 50 MHz-350 MHz (L-band), perfectly suitable for measuring the redshifted 21 cm signal from the Epoch of Reionization (EoR). It will be comprised of linked antenna stations with a physical baseline of 65 km separating them. The SKA1-MID is the second component and will be built in the Karoo Astronomy Reserve of South Africa. It can support radio receivers that work at mostly high frequencies. The SKA1-MID will extend its antenna number to approximately 200, constructed to span baselines greater than 100 km. It is designed to address a number of fundamental science areas; these include but are not limited to the detection of gravitational waves through pulsar techniques and other fundamental science (Dewdney et al., 2009).

1.3 SKA Big data problem and data size estimates

Here, we shall briefly review the data size estimations for the future SKA, describe the fundamental procedure required for radio imaging, and highlight the challenges this project will face. We describe the data trajectories from the individual antenna stations (SKA1-LOW and SKA-MID). The per antenna measured voltages are transported from each antenna to the central signal processing facility. It is this data that radio astronomers work with for advanced scientific discoveries; in the SKA, the concern is data size resulting from the cross-correlation of a permutation of 200 antennas. This telescope is expected to produce approximately 0.5–1 TB of data per second. Since SKA observations are expected to be carried out more or less continuously, this results in thousands of megabits of data being recorded for any particular observation. This problem will be even more severe in terms of data processing since the data reduction techniques (imaging and calibration) for the SKA will be done in real-time by the Science Data Processor (SDP). To gain insight into the expected data rates, consider a preliminary case for MeerKAT data rates: A single visibility record is a 2×2 matrix of complex *correlations*. Assuming 32 bits per correlation, we have $32 \times 4 = 128$ bits per visibility record. Assuming n_a antennas, one has $n_a(n_a - 1)/2$ unique baselines. For a given baseline, T_{obs} is the total observing time. If Δt is the integration time, then N_t is the number of timeslots. In addition, if B is the total bandwidth, dividing it into frequency channels, giving a N_ν total channels of width ($\Delta\nu$), the correlator records a data size of the amount:

$$\mathrm{datasize} = \frac{n_a(n_a-1)}{2} \times \frac{T_{\mathrm{obs}}}{\Delta t} \times \frac{B}{\Delta v} \times 128 \text{ bits}, \tag{1.2}$$

where T_{obs} refers to the overall duration during which the observation was conducted, N_t represents the divisions of the total observing time into discrete time intervals, each of duration Δt. Mathematically, this relationship can be expressed as:

$$T_{obs} = N_t \times \Delta t. \tag{1.3}$$

Similarly for frequency, as described above, the relationship is given by

$$B = N_\nu \times \Delta\nu, \tag{1.4}$$

where B is channelised into n_ν channels of size $\delta\nu$ each.

In light of the modern big interferometers where observation is continuously taken over long periods and large bandwidths, this will result in storage issues and a computational burden for sub-processing. For example, the data size for MeerKAT, 64 antennas, 32768 channels, 8 hour observation with 8 s integration time will result in approximately 30 TB of data being recorded. Furthermore, for SKA1-mid, with an additional 136 antennas and significantly higher correlator dump time rates of 0.1 s, one is looking at approximately 2404 TB of data being recorded. The data sizes of this magnitude are considered a challenge, as they can not be stored long-term. Hence, the current SKA SDP plans to process the visibilities in real time to produce image cubes, and only store the image cubes. Typical image cubes have the spatial size of 2^{10} by 2^{16} pixels Scaife (2020). However, the risk with this is that the imaging algorithm used to process the cubes becomes fixed, and there is no opportunity to reprocess the data again with better algorithms, as they become developed in the future.

1.4 Problem statement

The compression of visibility data is an imperative option, as discussed. Traditionally, compression is done by simply averaging the visibility data across time and frequency. However, if the time and frequency compression factors are scaled above some limit, this leads to an undesirable effect known as *smearing or decorrelation*. This has a corresponding effect in uv-space since uv-space is made up of baseline samples which are functions of both time and frequency. The source distance from the field centre is equivalent to a phase gradient in the visibility plane, with a more significant gradient for sources at a large distance relative to the phase centre. To illustrate this concept, consider a simplified scenario of a single point source with a flux denoted as S. The point source is located in the sky, and its position is specified by the direction cosine elements, denoted by the vector $\mathbf{l} = (l, m, n)$. Here, l, m, and n are the directional cosine components that define the celestial position of the observed source in a coordinate system relative to the observational reference frame. The directional cosine elements represent the projection of the source's position vector onto the observational axes. Specifically, l represents the east-west

projection, m along the north-south direction, and n along the direction of the source. Similarly, consider a baseline denoted indexed pq. A baseline represents the separation vector between two antennas. Associated with this baseline is a vector $\mathbf{u}_{pq} = (u, v, w)$. These components, u, v, and w, quantify the projection of the baseline vector onto three orthogonal directions: u corresponds to the east-west direction, v to the north-south direction, and w to the vertical direction. For a particular baseline pq, the visibility of the single source on the fixed time-frequency bin is given by:

$$V_{pq} = Se^{-i\Phi(\mathbf{u}_{pq},\mathbf{l})}, \tag{1.5}$$

where $\Phi(\mathbf{u}_{pq}, \mathbf{l}) = \frac{2\pi}{\lambda}\mathbf{u}_{pq}\mathbf{l}^{\top}$. Note that Φ is a function of the baseline vector and direction cosine. It should be clear now that for any source positioned at an increasing distance away from the phase centre, for a given long baseline $\mathbf{u}_{pq}$ at sufficiently large time and frequency integrations, $\Phi(\mathbf{u}_{pq}, \mathbf{l})$ becomes large, which make the exponential term in Equation 1.5 to decrease S, resulting in significant signal attenuation.

1.5 Motivations

With the newer telescopes having longer baselines and wide fields of view, the decorrelation effect becomes more severe. Alternative methods for data compression and mitigation of the decorrelation effect were investigated in Atemkeng et al. (2016) which simply uses BDWFs to average visibility data. BDWFs result in a much more reduced amplitude loss within the FoI and less aliasing effect or distortions held to an acceptable level. BDWFs use window functions in the Fourier space to compress visibility data while preserving signal attenuation. The Window functions are not optimized and are used as known in signal processing theory. This thesis investigates the use of machine learning optimization via gradient descent to find optimal window functions that could be used with BDWFs. We note with the optimization that since window functions are band-limited in theory, and this has an urgent effect on the optimization results, which require a wide enough window function extension. To mitigate the latter, we study BDWFs

and their performance with several weighting schemes known in the literature.

1.6 Dissertation Structure

Chapter 1: I start by providing a holistic approach to this research and provide motivations.

Chapter 2: Here, I review some fundamental background literature on radio interferometer and aperture synthesis and optimization theory.

Chapter 3: Here, I provide the analytical development of data compression in a mathematical formalism, particularly for time and frequency averaging and the effect (signal loss), with major derivations from literature.

Chapter 4: In this chapter, I perform BDWFs simulations in a more context-based framework, using simulated MeerKAT data as they are studied and developed in Atemkeng et al. (2016).

Chapter 5: I revise filter theory in digital signal processing, then implement the gradient descent optimization method for filter design in radio interferometric data compression. I highlight the shortfalls of the current application of BDWFs and progress towards developing a solution.

Chapter 6: This is the main chapter where I present the findings of this research and justify their importance in data compression and interferometric data reduction.

Chapter 7: The general conclusion with the summary of the major results of this work and some possible suggestions of how this can benefit the radio community and implied future research.

CHAPTER 2

Fourier transform, filter theory and radio interferometry

2.1 Introduction

In this chapter, we discuss the Fourier transform concept, then proceed to review filter theory, fundamentals of radio interferometry and background literature on Artificial Neural Networks (ANNs, McCulloch & Pitts (1943)).

2.2 Fourier transform

We begin this chapter by revisiting a fundamental concept in signal processing, the *Fourier transform* (FT). This transformation relates a signal from a representation in the time domain to its frequency equivalent. The FT of a signal $f(x)$ is given as:

$$F(y) = \mathcal{F}\{f(x)\} \tag{2.1}$$

$$= \int_{-\infty}^{\infty} f(x) \exp^{-2i\pi xy} \mathrm{d}x, \tag{2.2}$$

where x,$y \in \mathbb{R}$. In general, the transformation can be reversed to recover $f(x)$, and that is done via the inverse Fourier transform:

$$f(x) = \mathcal{F}^{-1}\{F(y)\} \tag{2.3}$$

$$= \int_{-\infty}^{\infty} F(y) \exp^{2i\pi xy} \, \mathrm{d}y. \tag{2.4}$$

Equations (2.2) and (2.4) tell us that $f(x)$ and $F(y)$ are unique Fourier pairs. Alternatively, the 2-D FT is given as:

$$F(u, v) = \mathcal{F}\{f(x, y)\} \tag{2.5}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp^{-2i\pi(xu+yv)} \, \mathrm{d}x\mathrm{d}y. \tag{2.6}$$

Its inverse is obtained analogous to the 1-D inverse FT, Equation (2.3) implies that corrugations of appropriate frequencies, phases, and amplitudes can be summed to produce the original two-dimensional waveform given by Equation (2.5). The above developments cater for physical signals and cannot be implemented in practice. In real life, one assumes that the 2-D function $f(x, y)$ has been spatially sampled along the spatial coordinates x with sample interval T_x and y sampled with interval F_y, respectively. The resulting sampled function is $f(kT_x, lF_y)$, where $k = 0, 1, ..., M-1$ and $l = 0, 1, ..., N-1$, and we refer to it as DFT defined as:

$$F\Big(\frac{m}{MT_x}, \frac{n}{NF_y}\Big) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} f\Big(\frac{k}{MT_x}, \frac{l}{NF_y}\Big) e^{-2i\pi(\frac{km}{M}+\frac{ln}{N})}. \tag{2.7}$$

This gives us the 2-D discrete FT along the data points defined by parameters k and l. To evaluate this, we compute N 1-D transforms: one for each l, where $l = 0, 1, ..., N-1$, along each data direction k and vise-versa. In applications such as image processing and filter design, a *Fast Fourier transform* (FFT, (Deeming, 1975), Cooley & Tukey (1965)) is used as an efficient algorithm to calculate DFT due to better algorithm scaling of $\mathcal{O}(n \log n)$ compared to direct DFT, which scales as $\mathcal{O}(n^2)$.

2.3 Filter theory

Digital filters are, systems of numerical calculations performed on a sampled, discrete-time-frequency signal to maximise or minimise specific components of that signal and improve the quality of a signal by eliminating additional noise, extracting information from the signal, or separating two or more previously combined signals (Alexander & Williams, 2016). In particular cases, filtering is designed by rejecting all frequencies outside the desired band. There are two types of digital filters that are widely used: Finite impulse response (FIR) and infinite impulse response (IIR). FIR filters have a finite-duration impulse response, which means that their output is only influenced by a finite number of past input samples. FIR filters are known for their stability, linearity, and straightforward design process, which involves designing a set of coefficients to determine the filter's behaviour. Due to these advantages, FIR filters are preferred from a design perspective. On the other hand, IIR filters have an infinite-duration impulse response, meaning that their output is influenced by an infinite number of past samples. In some cases, IIR filters can achieve similar filtering effects as RIF but will have fewer coefficients, making them more computationally efficient. However, designing IIR filters is generally more complex due to the need to satisfy stability and other design constraints. With that being said, FIR is favoured when a linear phase response and ease of design are essential. Based on the form of the frequency transfer function, the FIR digital filters are classified into; low-pass, band-pass, high-pass, and stop-band filter (Joaquim & Lucietto, 2011). A general formalization of an FIR filter is given as:

$$x[n] = \sum_{i=0}^{N} h[i]x[n-i], \tag{2.8}$$

where N is the filter order, $x[n]$ is input signal, and, $h[i]$ is FIR filter coefficients.

The DFT of an FIR filter is given as:

$$X(k) = \sum_{n=-\infty}^{\infty} x[n]e^{i2\pi kn}/N. \tag{2.9}$$

From Equation (2.9), one can find its equivalent inverse FT:

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{j2\pi nk/N}. \tag{2.10}$$

Now that we have laid the mathematical base for the FIR filter, the next step is to explore the properties which will lead us to the matched filter design. Consider a low-pass FIR filter with some cut off frequency ω_c, the impulse response $h(n)$ is given by :

$$h(n) = \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} e^{in\omega}\mathrm{d}\omega \tag{2.11}$$

$$= \frac{\omega_c}{\pi}\frac{\sin(n\omega_c)}{n\omega_c}. \tag{2.12}$$

As with the ideal filter, this response is non-causal and, therefore, can not practically be implemented. Hence, a causal response is required. A causal response can be obtained by, firstly, truncating the impulse to length $n = N$. Secondly, time delaying the impulse response to maintain the linear phase property, the symmetry of the response must be preserved (Alam & Hasan, 2010), and the required delay is denoted by $\alpha = (N-1)/2$. Hence, the shifted truncated impulse response is given by

$$h[n] = \begin{cases} \frac{\omega_c}{\pi}\frac{\sin(n-\alpha)\omega_c}{(n-\alpha)\omega_c} & n = 0 \text{ to } N-1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

The truncation of the impulse shows that the ideal frequency response can not be obtained. Truncating an infinite support impulse to a finite response gives rise to the effects of the pass and stop band ripples and finite transition band. As a remedy, we think of the truncated response $h_d(n)$ to be a result of multiplying the delay impulse response by some window function, $w(n)$;

$$h_d(n) = h(n-\alpha)w(n), \forall n, \tag{2.14}$$

where

$$w[n] = \begin{cases} 1 & n \leq 0 \leq N-1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

The window function's primary technique is to transform an infinite duration of the IIR to a finite duration FIR filter design (Alwahab et al., 2018). The window method for FIR design that is used in the DSP field is approximated approaches of Kumar & Purwar (2017), which are far from optimal results in the FIR filter design. Therefore, other FIR filter design strategies that are not based on the window method, such as gradient descent-based optimization, can be used.

2.4 Radio interferometry and aperture synthesis

Interferometry is a scientific practice using two or more radio telescope elements to observe radio source emissions. These elements, along with their electronics, are essentially used to synthesize signals by measuring Fourier components of the source brightness distribution. This instrument is referred to as an interferometer. The invention of an interferometer is motivated by the limitations of using a dish telescope, particularly the angular resolution given by Equation (1.1). To obtain an acceptable resolution for any given wavelength, the D in Equation (1.1) is now replaced by the maximum baseline length $||\mathbf{u}_{max}||$:

$$\theta \approx \frac{\lambda}{||\mathbf{u}_{max}||}, \tag{2.16}$$

where $\mathbf{u}_{max}$ is the largest separation between all pairs of antennas. Here, $||.||$ is the Euclidean norm. The limitation is the number of Fourier components that provide information about the observed source structure. To improve that, the radio signals are obtained from the relative orientation of both elements and the sources under observation, which change due to Earth's rotation. In astronomical terms, this is referred to as *Aperture synthesis* (Ryle & Hewish, 1960). Aperture synthesis techniques give an approximation angular resolution comparable to that of a single large dish telescope with a diameter D. If this is done repeatedly, each separation is repeated in different places, increasing the number of Fourier measurements.

2.4.1 Response of an interferometer and Measurement Equation

To demonstrate the principle of a radio interferometer, let us reduce the idea to that of a two-element interferometer A two-element radio interferometer measures the complex quantity $\mathcal{V}_{pq}(\mathbf{u})$, known as visibility data. Here, $\mathbf{u} = (u, v, w)$ is the vector that separates antennas p and q. The question becomes, how is this quantity measured? The fundamentals of modern radio interferometer measurements stem from the famous mathematical formalization known as the Radio Interferometric Measurement Equation (RIME, Hamaker et al. (1996), Smirnov (2011)). Here, we are interested in a RIME-based framework over some small sky regions (small angle approximation), not full-sky. We assume that the sky is composed of at least one discrete point source.

We define the visibility measurement from correlating antenna p and q as:

$$V_{pq} = \sum_{\mathbf{l}} (\mathbf{J}_{p\mathbf{l}} \mathbf{B}_{\mathbf{l}} \mathbf{J}_{q\mathbf{l}}^{H}), \tag{2.17}$$

where ^{H} is the hermitian operator, $\mathbf{J}_{p\mathbf{l}}$ and $\mathbf{J}_{q\mathbf{l}}^{H}$ are Jones matrix representing the signal propagation effects from the source to each antenna p and q over an image plane region at discretised sky $\mathbf{l} = (l, m, n-1)$. Here, l, m and n are directional cosines and the $n-1$ term ensures that the compensating delay phase centre vector is unity; otherwise, it is n. Note that, $\mathbf{B}_{\mathbf{l}}$ is the sky distribution matrix over an image plane region at $\mathbf{l}$. The formalism provided in Equation (2.17) is general concerning all the propagation effects of the Jones matrix. Let us now consider a subset of the propagation effects relevant to this work:

$$\mathbf{J}_{p\mathbf{l}} = \mathbf{G}_{p\mathbf{l}} K_{p\mathbf{l}} \tag{2.18}$$

$$\mathbf{J}_{q\mathbf{l}}^{H} = \mathbf{G}_{q\mathbf{l}}^{H} K_{q\mathbf{l}}^{H}. \tag{2.19}$$

The matrices $\mathbf{G}_{p\mathbf{l}}$ and $\mathbf{G}_{q\mathbf{l}}^{H}$ describe the direction-independent propagation effects in each path of antenna p and q and the scalar terms $K_{p\mathbf{l}}$ and $K_{q\mathbf{l}}^{H}$ describe the phase effects variations associated with antenna p and q at the sky position $\mathbf{l}$. For compactness, we can write $K_{pq\mathbf{l}} = K_{p\mathbf{l}} K_{q\mathbf{l}}^{H}$, which satisfies commutative property when applied to any matrix of the same dimensionality. This allows us to collapse the generic form given in Equation (2.17) and rewrite it as follows:

$$V_{pq} = \sum_{\mathbf{l}} (\mathbf{G}_{p\mathbf{l}} \mathbf{B}_{\mathbf{l}} \mathbf{G}_{q\mathbf{l}}^{H}) K_{pq\mathbf{l}}. \tag{2.20}$$

This allows one to define the Jones scalar term as the Fourier kernel that associates baseline vector $\mathbf{u}_{pq} = (u_{pq}, v_{pq}, w_{pq})$ and sky position vector $\mathbf{l}$;

$$K_{pq\mathbf{l}} = e^{-2\pi i \mathbf{u}_{pq} \mathbf{l}}. \tag{2.21}$$

2.5 Imaging

The RIME formalism, as defined in Equation (2.17), can also be presented in a functional notation, assuming no effects of uv-Jones term as:

$$\mathcal{V}_{pq}(u,v,w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{A(l,m)I(l,m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i \frac{\lambda}{\nu}(ul+vm+w(\sqrt{1-l^2-m^2}))} \mathrm{d}l\mathrm{d}m, \tag{2.22}$$

where $A(l,m)$ and $I_\nu(l,m)$ are the primary beam pattern and surface brightness, respectively. The antenna's reception pattern, $A(l,m)$, describes the sensitivity of the interferometer elements. Equation (2.22) tells us that the visibility as a function of baseline coordinates at a centre frequency ν is the product of the primary beam of each antenna multiplied by the phase term of each position on the sky integrated over the entire sky. According to Thompson (1999), Thompson et al. (2001), van Cittert-Zernike theorem Equation (2.22) is an equivalent representation of the Fourier component of the sky brightness, $I(l,m)$ together with the beam pattern $A(l,m)$. One may notice that Equation (2.22) is neither a 3-D nor 2-D Fourier transform and hence can not be efficiently implemented via FFTs. For a practical approximation, the following assumptions are imperative; the FoV must be small, i.e. $l^2+m^2 \approx 1$ or the interferometer is coplanar, meaning $w=0$, throughout. Having met these assumptions, and in addition, assuming the beam pattern is unity for all antennas, one can write:

$$\mathcal{V}_{pq}(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(l,m) e^{-2\pi i(ul+vm)} \mathrm{d}l\mathrm{d}m. \tag{2.23}$$

Equation (2.23) is the true visibility measurement. In practical situations, an interferometer measures the sampled version of Equation (2.23). Assuming that $S(u,v)$ is the sampling function defined as:

$$S(u,v) = \sum_{k}^{N} \delta(u-u_i, v-v_i), \tag{2.24}$$

where, $\delta(.,.)$ is a 2-D delta function. The sampled visibility function or measured sky brightness $S(u,v)\mathcal{V}_{pq}(u,v)$ output of the array data collection. To be precise, the interferometer is not necessarily sensitive to the sky but to the FT of the sky.

The measured sky brightness is the inverse FT of the sampled version of Equation (2.23):

$$I_{meas}(l,m) = \int\int S(u,v)\mathcal{V}(u,v) e^{2\pi i(ul+vm)} \mathrm{d}u\mathrm{d}v. \tag{2.25}$$

In practice, producing an image is done via discrete DFT. However, an FFT algorithm is used instead due to better computational efficiency. Using FFT results in image formation challenges

because FFT requires uniformly spaced samples in the spatial dimension, but in the uv-plane, the visibility data is irregularly sampled. To deal with this, gridding is adopted. Gridding is an algorithm that transforms the irregularly sampled uv domain into a regularly sampled one. This is known as gridding. The gridding mechanism uses a gridding kernel or convolutional kernel $C_a(u,v)$ to minimize aliasing effects. Mathematically, the gridded visibilities can be expressed as:

$$\mathcal{V}_g(u_k,v_k) = [[\mathcal{V}(u,v)S(u,v)] * C_a(u,v)]\text{Ш}(u_k,v_k), \tag{2.26}$$

where $\text{Ш}(u_k,v_k)$ is a sampling or comb function. Sampling by a Shah function means we regularly sample as a form of discretization data. Having obtained the gridded visibilities, also known as (resampled visibilities), the FFT is now valid, meaning we have arrived at the imaging process. Firstly, sampling with a Shah function introduces some aliasing as the collection of these delta functions often has large gaps in between discrete samples, causing sidelobes that affect or distort the shape of the Fourier window. Secondly, most arrays are designed such that most typical data points are in the inner region of the uv-plane and sparse in the outer, giving rise to shorter spacing being given more weights. This causes the visibility sampling to be different at each grid point, highlighting the difference in spatial features due to array configuration. These spatial features affect the resolution, sensitivity of the resulting image, and how the beam shape is controlled. To account for these, we introduce a *weighting function* $W(u,v)$ which down-weights certain data points and minimizes the effect of non-uniform density of visibility samples in the uv-plane. The weighting function shall be defined implicitly as a function of density weighting and tapering functions. Equation (2.26) can now be written as:

$$V_g(u_k,v_k) = [[V(u,v)S(u,v)W(u,v)] * C_a(u,v)]\text{Ш}(u_k,v_k). \tag{2.27}$$

We can now take the inverse FT of Equation (2.27) to get an image.

$$\mathcal{F}^{-1}\Big\{V_g(u_k,v_k)\Big\} = \mathcal{F}^{-1}\Big\{[\mathcal{V}(u,v)S(u,v)W(u,v)] * C_a(u,v)\Big\} * \mathcal{F}^{-1}\Big\{\text{Ш}(u_k,v_k)\Big\}, \tag{2.28}$$

$$I^D(l,m) = \mathcal{F}^{-1}\Big\{\mathcal{V}(u,v)\Big\} * \mathcal{F}^{-1}\Big\{S(u,v)W(u,v)\Big\}\mathcal{F}^{-1}\Big\{C_a(u,v)\Big\}\text{Ш}(l,m) \tag{2.29}$$

$$I^D(l,m) = \Big(I(l,m) * B(l,m)\Big)C'_a(l,m)\text{Ш}(l,m). \tag{2.30}$$

Where $I^D(l,m)$ is the resulting dirty image, which is true sky brightness $I(l,m)$ convolved with the point spread function $B(l,m)$ and the convolutional kernel $C'_a(l,m)$. One notes that under FT, the comb function is invariant, and multiplication by it results in aliasing, which manifests as a periodic function of l and m. We shall return to this in chapter 6 and elaborate on how it ties in with this thesis.

2.6 Conclusion

In this chapter, we have studied the principles of Fourier analysis, particularly the FT and how it is applied in image processing. Then moved to discuss filter theory and its use in FIR filter design. We also provided fundamental principles of radio interferometry, from the measurement function to imaging. Chapters 3 and 4 will implement these principles of the radio interferometer reviewed in this chapter through simulations.

CHAPTER 3

Simulation of Time-Frequency Decorrelation with the MeerKAT telescope

3.1 Introduction

In this chapter, we review the effect of time and bandwidth smearing caused by coarser time and frequency averaging. We start by deriving the average approximation form of the measurement equation or visibility averaging and convolution, then we show the effect of averaging from an empirical perspective, and then simulations are used to visualize the effect using MeerKAT data.

3.2 Averaging and Convolution

We introduce the concept of compression of the data, and the traditional method of compressing interferometric data is simply averaging the visibilities. We present the mathematical framework for visibility averaging at a single baseline pq at discrete time t_k and frequency ν_l instances. An actual interferometer can only measure the average visibility over some discrete finite time-

frequency bin $B_k^{[\Delta t]}$ and $B_l^{[\Delta \nu]}$ in time and frequency respectively:
$B_{kl}^{[\Delta t \Delta \nu]} = B_k^{[\Delta t]} \times B_l^{[\Delta \nu]}$.
This is the set of all ordered pairs or tuples (t_k, ν_l), where t_k is an element of $B_k^{[\Delta t]}$ and ν_l is an element of $B_l^{[\Delta \nu]}$

$$B_{kl}^{[\Delta t \Delta \nu]} = [t_k - \frac{\Delta t}{2}, t_k + \frac{\Delta t}{2}] \times [\nu_l - \frac{\Delta \nu}{2}, \nu_l + \frac{\Delta \nu}{2}]. \tag{3.1}$$

Where Δt and $\Delta \nu$ are the integrations in time and frequency respectively. The sampling bin in Equation (3.1) is finite and discrete over time and frequency. The motivation behind this is that in practice, the propagation effects in the Jones matrix **J** depend necessarily on the phase scalar matrix $K_{pq\mathbf{l}}$ and are time-frequency variable due to the baseline vector measured in units of wavelengths. This means the baseline vector is scaled by inverse wavelength, which then causes the phase matrix to depend on frequency. Again, assuming an Earth-based interferometer, Earth's rotation causes the baseline vector to rotate with time, making the scalar phase matrix time-dependent. Considering this, Equation (2.20) takes a new form of representation as being explicitly time and frequency dependent and we shall adopt that representation onwards.

If we consider an integration bin in the correlator domain given by Equation (3.1), then the measured visibility function is simply an average of the visibility observed by the baseline pq. To obtain this we take the limit of Equation (2.20) as the sampling bin gets large:

$$\bar{V}_{pqkl} = \lim_{|B_{kl}^{\Delta t}|, |B_{kl}^{\Delta \nu}| \to (\infty, \infty)} V_{pq}(t_k, \nu_l) |B_{kl}^{\Delta t}||B_{kl}^{\Delta \nu}| \tag{3.2}$$

$$= \frac{1}{|B_{kl}^{\Delta t}||B_{kl}^{\Delta \nu}|} \int \int_{B_{kl}^{[\Delta t \Delta \nu]}} \mathcal{V}(u_{pq}(t, \nu) \mathrm{d}t \mathrm{d}\nu, \tag{3.3}$$

where $|B_{kl}^{\Delta t}|$ and $|B_{kl}^{\Delta \nu}|$ denotes the post averaged integrations in time and frequency. We have shown averaging in the regular time and frequency space. However, we can also show that averaging is equivalent to convolution by a boxcar window function in the irregular Fourier space (uv-plane). To develop this, we define a uv-space corresponding bin by B_{pqkl}^{uv} through a change of variables. Assuming a small enough sampling bin where the smearing effect is neglectable, the measured visibility representation in $t\nu$-plane is the same as in uv-plane. Allowing us to write:

$$\bar{V}_{pqkl} = \frac{1}{|B_{kl}^{\Delta t}||B_{kl}^{\Delta \nu}|} \int \int_{B_{pqkl}^{uv}} \mathcal{V}_{pq}(u, v) |\frac{\partial(u, v)}{\partial(t, \nu)}| \mathrm{d}u \mathrm{d}v. \tag{3.4}$$

Note that the integrand has only pq index, which indicates a baseline dependent, since the sampling bins in the uv space are non-regular or elliptical arcs and is a function of time and frequency. We can now introduce a boxcar window function:

$$\Pi^{t\nu}(t,\nu) = \begin{cases} \frac{1}{\Delta t \Delta \nu} & |t| \leq \Delta t/2, |\nu| \leq \Delta \nu/2 \\ 0 & otherwise \end{cases} \tag{3.5}$$

Equation (3.5) is a convolution kernel that is applied to get re-sampled visibilities; since the re-sampling function is defined in the uv-plane, it should be a per baseline function. When applying the boxcar window function to visibilities Equation (3.4) becomes equivalent to:

$$\bar{V}_{pqkl} = \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} \mathcal{V}_{pq}(t,\nu)\Pi^{t\nu}_{pq}(t-t_k,\nu-\nu_l)dtd\nu. \tag{3.6}$$

Similarly, Equation (3.6) can be written as a convolution of the true visibility by a boxcar function of a fixed sampling bin uv-space :

$$\bar{V}_{pqkl} = [\mathcal{V}_{pq}(t_k,\nu_l) * \Pi^{uv}_{pq}(t_k,\nu_l)], \tag{3.7}$$

where $*$ represents convolution.

3.3 Time and bandwidth Decorrelation/ smearing

Let us start this section by recalling the Fourier space irregular average in Equation (3.7). For a single baseline pq, through the inverse Fourier transform, we can obtain a dirty image which is the inverse Fourier transform of the collection of averaged visibility samples in Equation 3.7 at each kl:

$$\mathcal{I}^D_{pq} = \sum_{kl} \mathcal{F}^{-1}\Big\{\mathcal{V}_{pq}(t_k,\nu_l) * \Pi^{uv}_{pq}(t_k,\nu_l)\Big\} \tag{3.8}$$

$$= \sum_{kl} \mathcal{F}^{-1}\Big\{\Pi^{uv}_{pq}(t_k,\nu_l)\Big\}\mathcal{I}_{pqkl}, \tag{3.9}$$

where $\mathcal{I}_{pqkl} = \mathcal{F}^{-1}\Big\{\mathcal{V}_{pqkl}\Big\}$ is the apparent sky. The inverse Fourier transform of a boxcar-like function is a $sinc$-like function. Therefore, Equation (3.9) becomes:

$$\mathcal{I}^D_{pq} = \mathcal{D}_{pq}\mathcal{I}_{pq}, \tag{3.10}$$

where we assume that the sky $\mathcal{I}_{pq} \equiv \mathcal{I}_{pqkl}$ is constant across each sample at kl and

$$\mathcal{D}_{pq} = \sum_{kl} \text{sinc}(r_1)\text{sinc}(r_2), \tag{3.11}$$

where $r_1 = \frac{\Delta\phi}{2}$ and $r_2 = \frac{\Delta\psi}{2}$ are half of the phase difference in time and frequency defined as:

$$\Delta\phi = 2\pi\Delta\mathbf{u}_{pq}(t, \nu_l)\mathbf{l}, \tag{3.12}$$

$$\Delta\psi = 2\pi\Delta\mathbf{u}_{pq}(t_k, \nu)\mathbf{l}. \tag{3.13}$$

We can see from Equation (3.10) represents a dirty image seen by baseline pq as the apparent sky multiplied by the baseline-dependent $sinc$ responses. The baseline dependent response represents the overall effect of all the accumulated averaging effects emphasized by the extra kl subscripts. Each baseline will taper the sky differently depending on the time-frequency indices kl. For shorter baselines, the $sinc$ response appears wider, and for longer baselines, the response appears narrower, which signifies a strong effect on I_{pq}. Another way to look at this is that the per baseline window response acts as a factor determining amplitude variation as a function of integration over time-frequency intervals. Each re-sampled visibility is obtained in the Fourier space by scaling the measured visibility matrix with the factor. The scaling depends on the limits of $B_{kl}^{\Delta t}$, $B_{kl}^{\Delta\nu}$: if the data is initially sampled at very high resolution $(\Delta t, \Delta\nu \rightarrow 0)$; and $|B_{kl}^{\Delta t}|, |B_{kl}^{\Delta\nu}|$ are also small enough, then the boxcar function is approximately one, and that corresponds to no averaging, else if these are large, then the boxcar function will be very different from one, resulting in the magnitude of the visibility function being scaled down by dividing by large product which results in a significant amplitude loss.

Analogously we can also show the effect of regular averaging in the correlator domain, and one can adopt Equation (3.3) for this derivation; the measured visibility matrix is written as:

$$\bar{\mathcal{V}}_{pq}(t_k, \nu_l) = \int\int_{B_{kl}^{\Delta t \Delta\nu}} \Big[\sum_{\mathbf{l}} \Big(\mathbf{G}_{p\mathbf{l}}\mathbf{B}_{\mathbf{l}}\mathbf{G}_{q\mathbf{l}}^{H}\Big) K_{pqt\nu\mathbf{l}}\Big] \mathrm{d}\nu\mathrm{d}t. \tag{3.14}$$

Since both the integration and the summation are finite, one can always interchange the two. If we recall that we are dealing with net amplitude loss caused associated with only the Jones phase term (decorrelation in this case), the Jones scalar term $K_{pq\mathbf{l}}$ changes significantly across

the integration bin. We also assume that the jones matrix $\mathbf{G}_{p\mathbf{l}}$ and $\mathbf{G}_{q\mathbf{l}}^{H}$ together with the matrix brightness $\mathbf{B}_{\mathbf{l}}$ are constant across the integration intervals, then Equation (3.14) can be written as:

$$\bar{\mathcal{V}}_{pq}(t_k,\nu_l) = \sum_{\mathbf{l}} \mathbf{G}_{p\mathbf{l}}\mathbf{B}_{\mathbf{l}}\mathbf{G}_{q\mathbf{l}}^{H}\Big[\int\int_{B_{kl}^{\Delta t\Delta\nu}} K_{pq\mathbf{l}}\mathrm{d}\nu\mathrm{d}t\Big] \tag{3.15}$$

$$= \sum_{\mathbf{l}} \mathbf{G}_{p\mathbf{l}}\mathbf{B}_{\mathbf{l}}\mathbf{G}_{q\mathbf{l}}^{H} D_{pq} \tag{3.16}$$

$$= \sum_{\mathbf{l}} \mathbf{G}_{p\mathbf{l}}(D_{pq}\mathbf{B}_{\mathbf{l}})\mathbf{G}_{q\mathbf{l}}^{H}. \tag{3.17}$$

Since D_{pq} is a scalar factor, which commutes with any other Jones matrix; hence Equation (3.16) can be rearranged to Equation (3.17). Note that D_{pq} is defined as,

$$D_{pq} = \int\int_{B_{kl}^{\Delta t\Delta\nu}} K_{pq\mathbf{l}}\mathrm{d}\nu\mathrm{d}t. \tag{3.18}$$

It is now explicit that the true sky brightness matrix $\mathbf{B}_{\mathbf{l}}$ is tapered by D_{pq}, which is the time-frequency decorrelation term. One can still write D_{pq} as:

$$D_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \Pi^{t\nu}(t,\nu)K_{pq\mathbf{l}}d\nu dt \tag{3.19}$$

$$= \mathcal{F}\big\{\Pi^{t\nu}(t,\nu)\big\} \tag{3.20}$$

$$= \mathrm{sinc}(r_1)\mathrm{sinc}(r_2), \tag{3.21}$$

where $r_1 = \frac{\Delta\phi}{2}$ and $r_2 = \frac{\Delta\psi}{2}$ are half of the phase difference in time and frequency as defined in Equations (3.12) and (3.13) respectively.

Substituting D_{pq} back to Equation (3.17) one can now express the averaged visibility function as :

$$\bar{\mathcal{V}}_{pq}(t_k,\nu_l) = \sum_{\mathbf{l}} \mathbf{G}_{p\mathbf{l}}\Big(\mathrm{sinc}\big(2\pi\Delta\mathbf{u}_{pq}(t,\nu_l)\mathbf{l}\big)\mathrm{sinc}\big(2\pi\Delta\mathbf{u}_{pq}(t_k,\nu)\mathbf{l}\big)\mathbf{B}_{\mathbf{l}}\Big)\mathbf{G}_{q\mathbf{l}}^{H}. \tag{3.22}$$

Equation (3.22) tells that the measured visibility function experiences an amplitude modulation by a sinc function, which depends on the integration time, Δt and integration frequency, $\Delta\nu$ at a given sky position $\mathbf{l}$.

3.4 Simulations with MeerKAT data

The standard database structure used in this thesis to record the complex visibility data is the measurement set (MS, Kemball & Wieringa (2000)). The MS appears in software programs as a hierarchical directory in which the radio interferometric data is stored as tables and directories. The MS is created as a configuration file containing tunable parameters that determine the visibility data structure and size.

We made an MS from a MeerKAT at 1.4 GHz observation. The observation time is 400 s sampled at each 1 s integration time, each with a bandwidth of 30 MHz divided into channels of width 83.4 kHz each. Using the MeqTrees software (Noordam & Smirnov (2010)) a simulation is made, and the MS is populated with a 1 Jy source at a given position. The source visibilities are then resampled with a boxcar-like window function of sizes [100 s $\times$ 10 MHz], [50 s $\times$ 5 MHz] and [20 s $\times$ 2 MHz], where the notation [x s $\times$ y mHz] means x samples are averaged in time and y samples averaged in frequency i.e $|B_{kl}^{\Delta t}| = x$ and $|B_{kl}^{\Delta \nu}| = y$. The visibilities are then translated to the image using the csCLean imager (Voronkov & Wieringa (2004)) with natural weighting. We begin by individually showing the effect of time and frequency average: the smearing increases with increasing distance from the phase centre, and the effect worsens on large time/frequency intervals. The below Figures 3.1, 3.2 are reproduced from Atemkeng et al. (2016), with Figure 4.1 showing the effect of averaging in time (top-panel) and frequency (bottom-panel) and Figure 3.2 show averaging in both time and frequency over four different sampling bin: [1 s $\times$ 83.4 kHz], [20 s $\times$ 2 MHz], [50 s $\times$ 5 MHz] and [100 s $\times$ 10 MHz].

Figure 3.2 shows the effect of applying time and frequency averaging simultaneously: the smearing magnitude decreases following the same reason, but the smearing is a bit significant since it decreases proportionally to the product of time and frequency amplitudes.

Figure 3.3, provides simulations illustrating a 1 Jy point source observed at 1.4 GHz. The data is sampled at high time integration of 1 s and 83.4 kHz then resampled to 100s $\times$ 10 MHz. The simulation shows the 1 Jy point source at three distinct sky locations. The top panel of Figure 3.3 shows the point source at the phase centre of the observation; we note that there is no smearing. The middle shows a source at a specific sky position, where the source is located

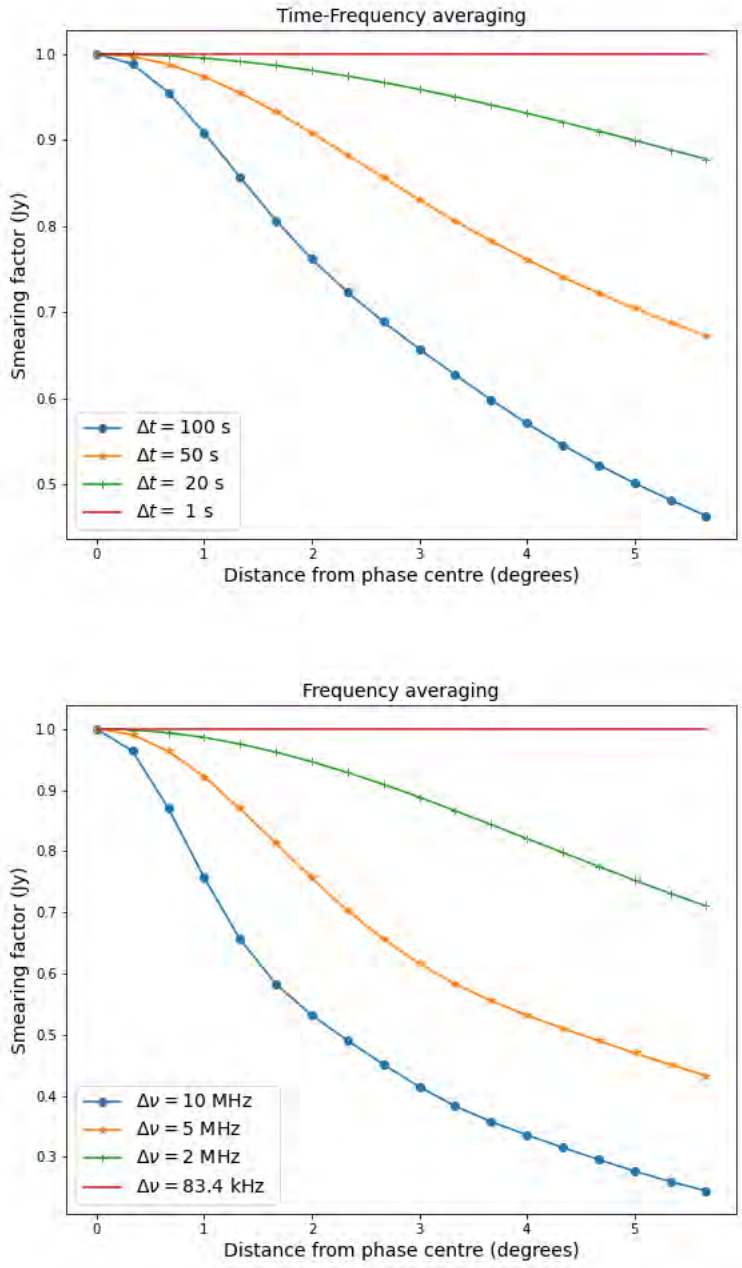

Figure 3.1: An example of a MeerKAT at 1.4 GHz simulation of a 1 Jy point source, showing effects of visibility averaging: Top-panel: Time smearing as a function of source distance from the phase centre. Bottom panel Frequency smearing as a function of distance from the phase centre.

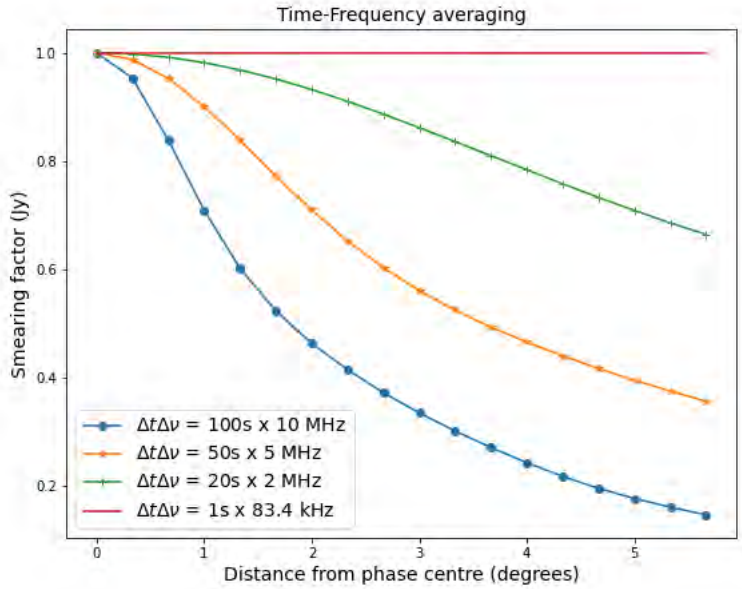

Figure 3.2: An example of a MeerKAT at 1.4 GHz simulation of a 1 Jy point source, showing the effect of visibility averaging at several time and frequency integrations.

at 1 deg relative to the phase centre, one can observe the distortion of the source's shape due to smearing. However, when compared to the bottom panel, in which the source is at 3 deg, the smearing is less pronounced.

3.5 Conclusion

In this Chapter, we have introduced, defined and derived the effect of smearing caused particularly by averaging in post-processing. We presented two forms in which smearing manifests; firstly, as a result of averaging in the regular plane and secondly, as a convolution with a boxcar window function in the irregular plane. We have also shown that smearing manifests as baseline size and distance from the phase centre. Lastly, the empirical results and synthesis imaging depicting smearing have been simulated.

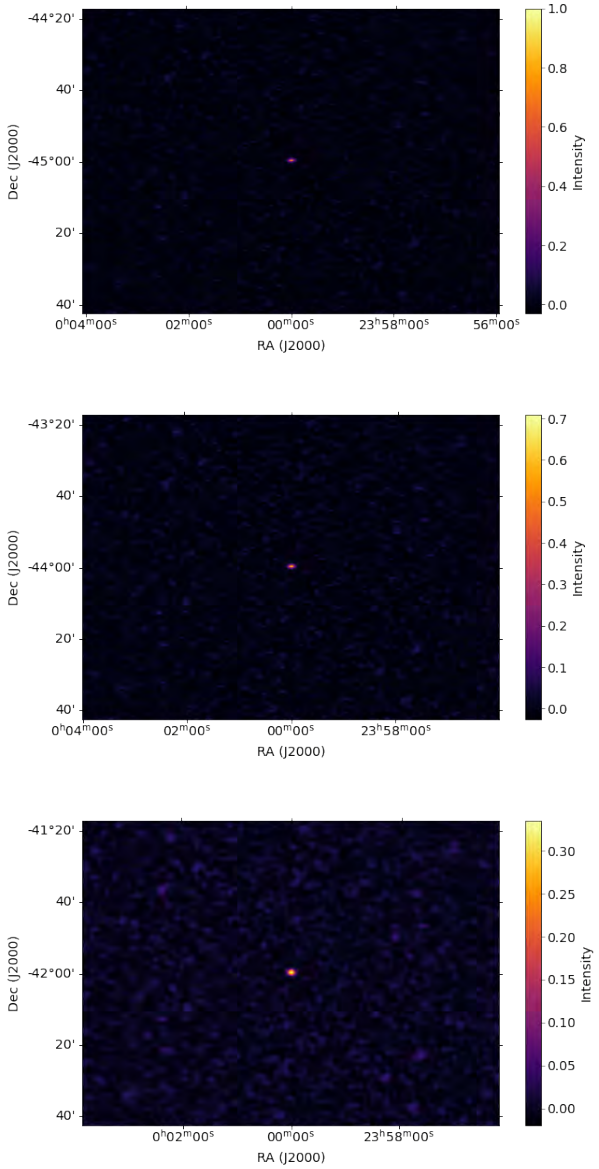

Figure 3.3: Simulations illustrating a 1 Jy point source observed at 1.4 GHz with visibility averaging using a sampling bin of size 100s × 10 MHz. The simulation shows three distinct sky locations: the top panel depicts the point source at the phase centre $(0°)$, the middle panel shows the point source positioned $(1°)$ offset from the phase centre, and the bottom panel features the point source located $(3°)$ offset from the phase centre.

CHAPTER 4

MeerKAT time and frequency smearing simulation with BDWFs applied

4.1 Introduction

In this chapter, we discuss the application of BDWFs. BDWFs are techniques used to compress the visibility data while preserving source amplitude loss at the edge of the FoI of an interferometer.

4.2 BDWFs and FoI shaping

In the context of BDWFs, the boxcar window function, as discussed in Equation (3.5), is now replaced by a window function that depends on the uv distance swept by the baseline pq:

$$\bar{V}_{pq}(\mathbf{u}(t_k,\nu_l)) = [\mathcal{V}_{pq}(\mathbf{u}(t,\nu)) * D^{\alpha}_{pq}(\mathbf{u}(t,\nu) - \mathbf{u}(t_k,\nu_l))], \tag{4.1}$$

where α and $D^{\alpha}_{pq}(t_k,\nu_l)$ are the chosen FoI and the BDWF respectively. The only difference, as discussed in Chapter 3 is that a $sinc$-like window rather than a boxcar is used to convolve the

visibility function. The extra subscript α means the window has an independent parameter α that could be tuned to a desired FoI. For simplicity, let us look at a 1-D case, in time, for a single baseline and sampling bin, D^{α}_{pq} is given as.

$$D^{\alpha}_{pq} = \mathrm{sinc}(\alpha\pi\Delta\mathbf{u}_{pq}(t)), \tag{4.2}$$

and we can write the Image Plane Response (IPR) as :

$$\mathcal{F}^{-1}\Big\{D^{\alpha}_{pq}\Big\} = \mathcal{F}^{-1}\Big\{\mathrm{sinc}(\alpha\pi\Delta\mathbf{u}_{pq}(t))\Big\} \tag{4.3}$$

$$= \frac{1}{\alpha}\Pi(\frac{r}{\alpha}), \tag{4.4}$$

where, $r = \frac{\Delta\phi}{2}$, the half phase difference in time defined in Equation (3.13). Figure 2.1 shows an example of D^{α}_{pq} and their corresponding inverse FT. Applying a window to different-sized baselines of the same sampling bin results in shorter baselines appearing as boxcar-like windows, medium baselines appearing as semi-arch-like windows and longer baselines appearing as $sinc$-like windows together with their corresponding IPRs. In the image plane, the physical interpretation is exactly contrary due to FT properties. In addition, the division by FoI parameter results in the shorter baselines corresponding to a wider $sinc$-like window (more like a boxcar) with a narrower IPR (still producing a $sinc$-like smearing response). In contrast, the longer baselines correspond to a narrower window with extended side lobes with a rectangular window-like IPR.

This chapter demonstrates through simulations the principle of averaging visibility data via BDWFs. We employ BDWF ($sinc$ window in this case) for the MeerKAT antenna station. Here we demonstrate the procedure to apply BDWFs to visibility data and how the quantities of interests (signal information) are measured. This chapter describes a series of simulations performed to inform the choice of optimal BDWFs to 1) average visibility data, 2) shape the FoI by conserving amplitude within a specific chosen FoI, and suppress any amplitude outside this field. Analytical approximations have been performed in Atemkeng et al. (2016) for JVLA-C, at various sampling rates, one particular interest was 100 s $\times$ 10 MHz sampling bin size post correlation at different FoI settings. In this chapter, we performed the exact simulations but now

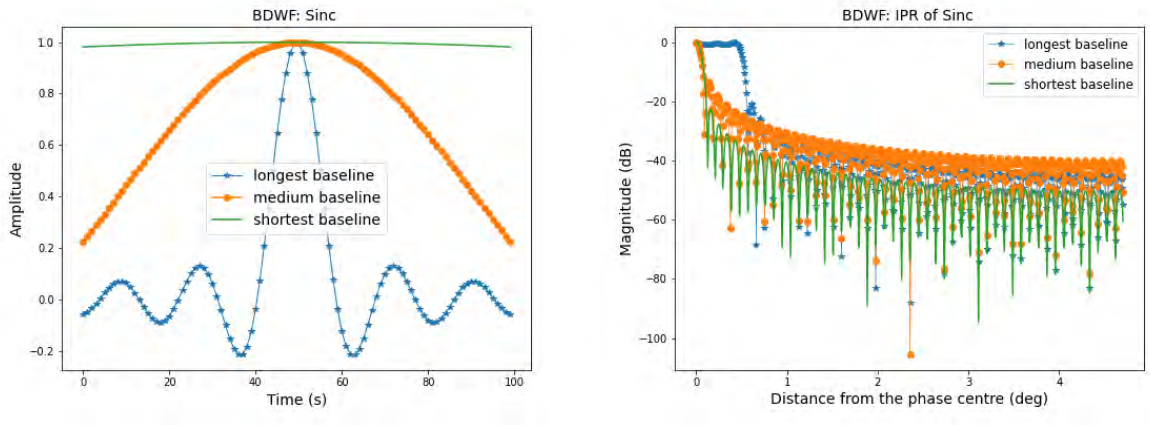

Figure 4.1: Left: Baseline sizes, with the longest baselines, looking $sinc$-like, and short baselines looking like boxcar-like window functions. Right: Corresponding IPR for each baseline

using the MeerKAT simulated data and will provide an equivalency of the JVLA-C sampling rate for MeerKAT and verify the analytical results.

4.3 Procedure and simulation

The effect of time and frequency decorrelation on MeerKAT data is to be studied here by performing simulations of time-frequency BDWFs for three FoI settings. A sky model containing point sources aligned regularly in the m-direction of the image plane and extending to approximately 5 degrees, is simulated. The Primary beam effects are neglected in the model so that the decorrelation with the BDWF period can be associated with only time and bandwidth averaging.

4.3.1 Creating datasets

We made an MS from MeerKAT at 1.4 GHz observation. The observation time is 400 s sampled at each 1 s integration time, each with a bandwidth of 30 MHz divided into channels of width 83.4 kHz each. Using the MeqTrees software, a simulation is made, and the MS is populated with a 1 Jy source at a given position. The source visibilities are then re-sampled with BDWFs of sizes [100 s $\times$ 10 MHz], [50 s $\times$ 5 MHz] and [20 s $\times$ 2 MHz], where the notation [x s $\times$ y

MHz] means x samples are averaged in time and y samples averaged in frequency i.e. $|B_{kl}^{\Delta t}| = x$ and $|B_{kl}^{\Delta \nu}| = y$. The visibilities are then translated to the image using the csCLean imager with natural weighting.

4.3.2 Imaging

We make a naturally density-weighted image of size 512×512 with a gridded cell size of $10.$ arcsec centred at each point source position in an array with 10 arcminute spacing. For each of the simulations, the procedure is as follows: The imaging is done with csCLean with zero iteration, meaning a dirty image was created. This is done to prevent additional effects that may vary between re-sampled datasets due to differences in source fluxes that might be caused by deconvolution due to CLEAN implementations.

4.3.3 Amplitude measurement of the sources

The amplitude is measured by sequentially extracting the flux of each simulated source in the dirty images using pyfits from astropy (Robitaille et al., 2013). This is done by accessing the first extension of the image data and taking its maximum for each source in the array. The flux response of the array sources should resemble the shape of a low-pass filter across the FoI. For example, given a FoI of 2 degree (equivalent to source at position 120 arcminutes in m-direction), the amplitude response should be preserved with as minimal decrease as possible within a 1-degree radius relative to the phase centre. It should ideally attenuate almost abruptly at exactly 1 degree. One can now see that this response shapes the FoI. To get better shaping means, the amplitude response should approximate a perfect low-pass. The desired response is an inverse FT of the convolution kernel we used to average the visibilities. The dirty image is now the true sky tapered by the inverse FT of the convolution kernel, which is a boxcar in the case of $sinc$. We can see this by recalling Equation (2.27) with additional parameter α.

The α parameter is the tunable FoI used to parameterize the spatial frequencies in the Fourier space. By definition of the scaling property of FT in the image plane, the parameter scales the window and then grids the pixels.

4.3.4 Simulation results

In this section, we show a typical performance of BDWFs used to re-sample the simulated visibility data, and we also observe or study other effects that play a role. The quantity we wish to measure in these simulations is the decorrelation of flux as a function of distance from the phase centre and increasing averaging bins in FoI shaping. Systematic effects on the source amplitude must be corrected before the measurement can be associated to decorrelation alone: array baseline distribution and density weighting schemes. Figures, 4.2, 4.3 and 4.4 depict the BDWF performance at different FoI settings.

- At 1 degree FoI, the BDWF's IPR is exactly the same as that of the traditional averaging.
- Increasing the FoI setting to 2 degrees, the IPR for BDWF is the same as that of traditional averaging except at a large sampling bin (e.g. 100 s $\times$ 10 MHz). At this sampling bin, the FoI induced by BDWF is broader than that of a traditional averaging.
- At 4 degrees FoI, the IPR induced by BDWF becomes broader as the sampling bin increases and it starts to shape the FoI as a function of increasing both the FoI setting and averaging bin.

We summarise the above results by showing how the BDWFs perform at different FoI settings and a specific sampling rate of 100 s $\times$ 10 MHz. To quantify the improvement, we find the ratio between the ideal lowpass, with the cutoff being half a given FoI, and the measured amplitude response corresponding to different sampling bins at different fields. We use Riemann sum to find the ratio of areas and then convert them into percentages. We start by defining two target functions which depend on the radial distance from the phase centre; let us define the total amplitude response area in the image plane as the sum of the within-FoI and out-of-FoI signal:

$$A(r) = A_{FoI}(r) + A_{\overline{FoI}}(r) \approx \begin{cases} 1 & r \leq \frac{FoI}{2} \\ 0 & r \geq \frac{FoI}{2} \end{cases} + \begin{cases} 1 & r \leq max(r) \\ 0 & otherwise. \end{cases} \tag{4.5}$$

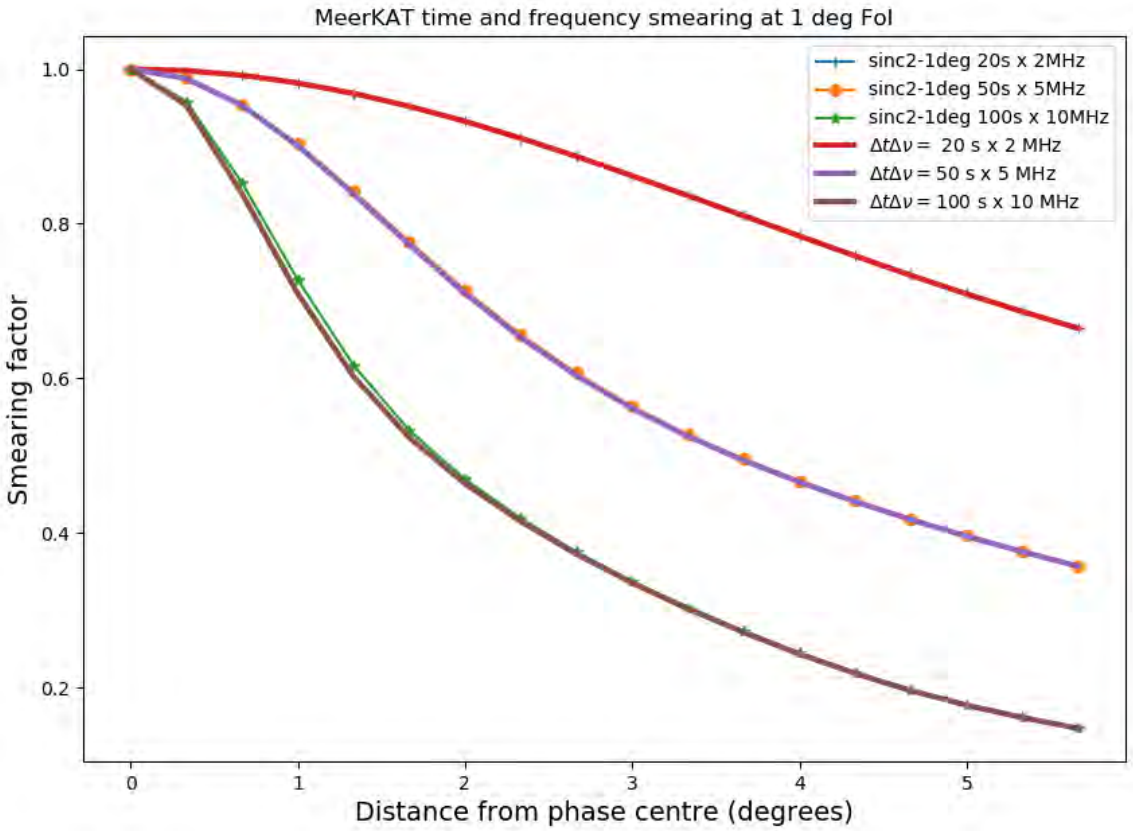

Figure 4.2: An example of BDWF simulation with MeerKAT 1.4 GHz smearing degree against source distance from the phase centre, for boxcar averaging at sampling intervals of 20 s $\times$ 2 MHz, 50 s $\times$ 5 MHz, and 100 s $\times$ 10 MHz bins, and for various BDWFs at 1 deg FoI

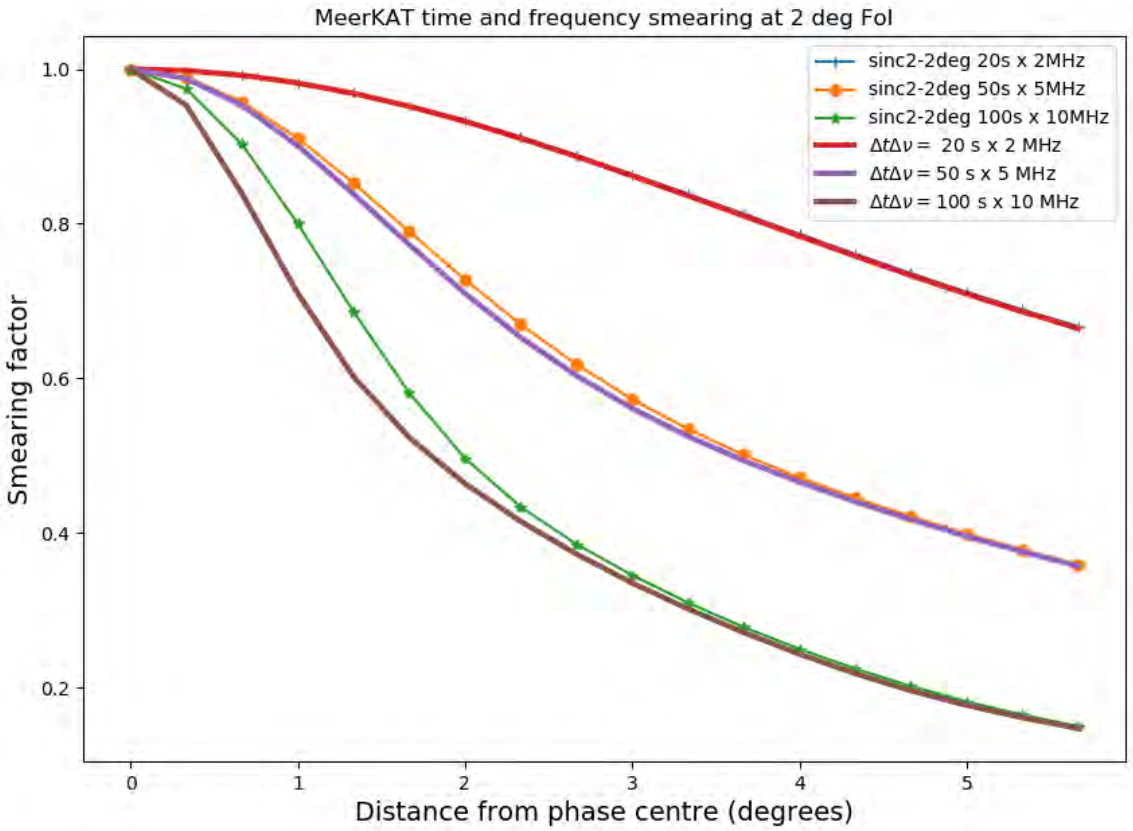

Figure 4.3: An example of BDWF simulation with MeerKAT 1.4 GHz smearing degree against source distance from the phase centre, for boxcar averaging at sampling intervals of 20 s $\times$ 2 MHz, 50 s $\times$ 5 MHz, and 100 s $\times$ 10 MHz bins, and for various BDWFs at 2 deg FoI

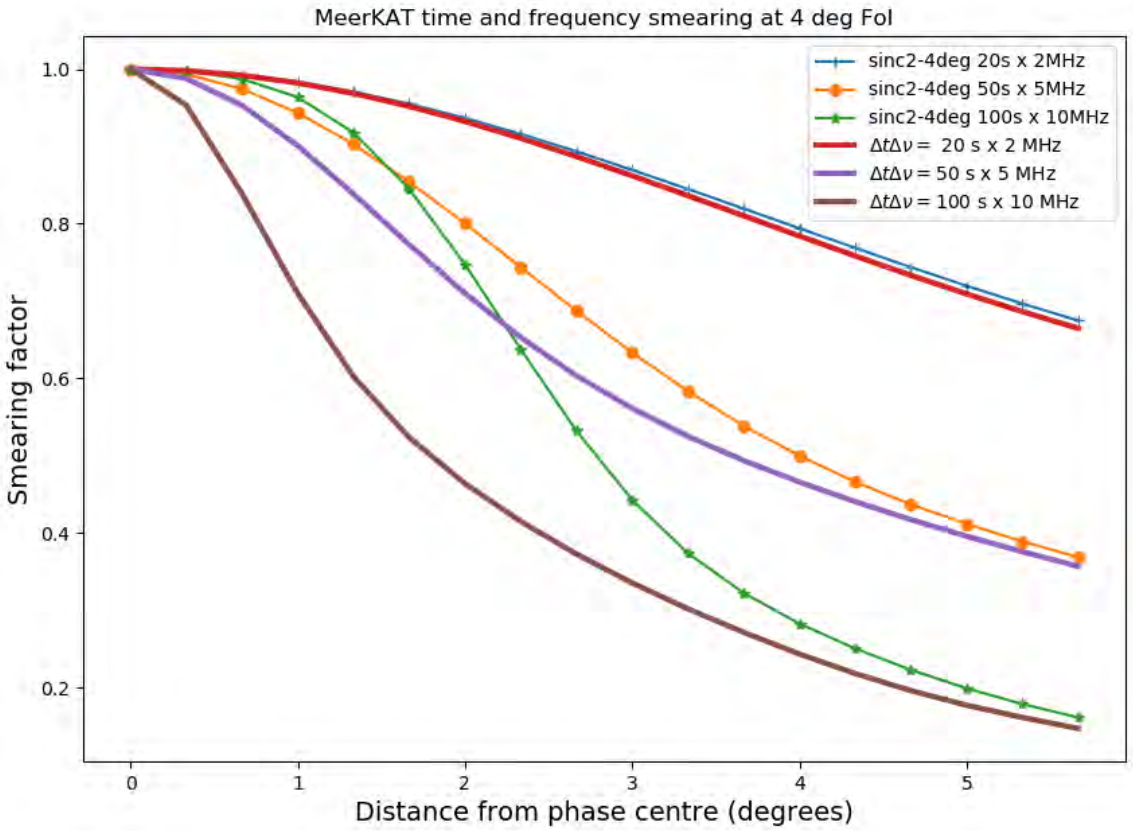

Figure 4.4: An example of BDWF simulation with MeerKAT 1.4 GHz smearing degree against source distance from the phase centre, for boxcar averaging at sampling intervals of 20 s $\times$ 2 MHz, 50 s $\times$ 5 MHz, and 100 s $\times$ 10 MHz bins, and for various BDWFs at 4 deg FoI

Now we define the Riemann finite sum over a closed interval $[r_0, r_n]$. The finite set of elements in the closed interval forms a partition. We define $f(j)$ as a function of an arbitrary factor $[r_i - r_{i-1}]$. Then the n^{th} Riemann sum of f concerning partition is defined as:

$$R(f_p) = \sum_{i=0}^{n} f_p(j_i)(r_i - r_{i-1}). \tag{4.6}$$

The performance of the BDWF is shown in Table 4.1. Table 4.1 shows the effect of visibility averaging in terms of the amount of signal and attenuated signal in percentages. We can observe that with conventional averaging, a significant amount of information is lost within the field of view, e.g., information of about 28 % and 12 % at 2 and 4 deg FoI, respectively, is lost through averaging visibilities using boxcar. We can compare these percentages with averaging using the BDWFs: at 2, 4 deg FoI about 8 % of the information is lost. We can note that a weighted average scheme conserves more signal within the FoI and compresses significantly outside as compared to the unweighted average scheme. Furthermore, it is worth noting that as the FoI size increases, the ability of a window function to shape the field is better due to the FT scaling property. A narrow window in the uv-plane corresponds to a wider IPR in the image domain. This is shown in Figure 4.1.

Table 4.1: Window function performance results

Field of views	Integrations/bin size	Conserved signal (%)	Suppressed signal (%)
1 deg	100 s $\times$ 10 MHz	90	63
1 deg (Normal Average)	100 s $\times$ 10 MHz	90	60
2 deg	100s $\times$ 10 MHz	92	63
2 deg (Normal Average)	100 s $\times$ 10 MHz	88	66
4 deg	100s $\times$ 10 MHz	92	64
4 deg (Normal Average)	100 s $\times$ 10 MHz	72	72
Ideal BDWF (1,2, 4 deg)	100 s $\times$ 10 MHz	100	100

4.4 Conclusion

I have discussed the technique of compressing the visibility data using BDWFs and how BDWFs are used to shape the FoI. An application of this is done via simulations on MeerKAT simulated data, after that we evaluated the BDWF's performance quantitatively relative to simple averaging.

CHAPTER 5

Revisiting BDWFs in the context of obtaining an optimal Field of Interest shaping via gradient descent optimization

5.1 Introduction

We introduce the context of optimization into the problem of FoI shaping. The chapter begins with a desire to use a gradient descent algorithm to optimize the FoI via BDWFs used to re-sample the visibility data. We re-examine what defines optimal FoI shaping in detail, and then proofs are provided as to whether optimizing only the window function is sufficient for FoI shaping. To that end, alternative optimization techniques are considered and explored to shape the FoI. This chapter starts with the BDWFs application's limitations and then proposes an optimization strategy based on gradient descent as a solution. From there, we explore efficient parameters for off-axis source suppression.

5.2 Using BDWFs for visibility data compression and FoI shaping has shown potential but is that the best we can achieve?

Beyond any reasonable doubt, replacing the unweighted average with a weighted average of visibility data brought a significant improvement in data compression and FoI shaping of a radio interferometer. To highlight a few, according to Atemkeng et al. (2016), the potential capabilities of tuned BDWFs have been theoretically and practically demonstrated that their use can both reduce smearing within the FoI regime (compared to normal "boxcar" averaging) and suppress unwanted signal outside FoI regime. At the same time, the data size is compressed to an acceptable level.

However, BDWFs (as applied currently) pose several negative effects. To mention a few is the unavoidable loss in sensitivity due to re-sampling using a weighted sum. Moreover, averaging over larger sampling bins, with larger tuned FoI where BDWFs appear as $sinc$-like weigh-down short uv-spacing, causing a decrease in nominal sensitivity. Another apparent downside of the current application of BDWFs is the poor off-axis attenuation compared to simple averaging. This poor off-axis attenuation implies far-field contamination, which then decreases the signal-to-noise in the FoI. However, this poor off-axis attenuation is mitigated by employing overlapping BDWFs as discussed in Atemkeng et al. (2016). Overlapping BDWFs gain their advantage from using visibility data points more than once in the application, approximating the infinite support of the $sinc$ window function. However, in the case of snapshot observations, the use of overlapping BDWFs is limited because the data available is small. While data compression is not needed in the snapshot regime, overlapping BDWFs could still be used to shape the FoI. Also, the computational complexity increases due to the duplicated visibility samples, making the overlapping BDWFs scale poorly. One would expect a method that uses non-overlapping BDWFs to result in an equivalent performance compared to overlapping BDWFs while few computation resources are in demand.

Another way to achieve optimal FoI shaping is using baseline-dependent averaging combined with BDWFs (Atemkeng et al., 2018). BDWFs ability to shape the FoI is, to some extent, limited

by the number of baselines of different sizes in the same fixed bin; by this, we mean very few long baselines sweeping out larger bins in the $u\nu$-space, with window function becoming $sinc$-like over them and the large number of short baselines sweeping out smaller bins in $u\nu$-space, with window function resembling boxcar-like shape over them. These downfalls motivated the study of this thesis.

5.3 The use of gradient descent-based optimization

This thesis attempts to provide the optimal solution to visibility data averaging by designing optimally matched filters for BDWFs. What we mean by this is that we want the IPR of the visibility domain filter to approximate a brick-wall boxcar window or an ideal lowpass filter response. We first propose using a gradient descent algorithm (Ruder, 2016) for this approximation. This method is designed in the Fourier space, which considers the ideal lowpass FIR filter as a target value, using a pre-existing ($sinc$ window) function to obtain the initial value and then using a modified backpropagation algorithm (Werbos, 1990). Various approaches were explored, including using a neural network for this purpose. However, after initial experimentation and evaluation, it became evident that a neural network-based solution was unnecessary and could be considered an overkill for the problem. The desired response could be effectively approximated using a simpler gradient descent algorithm in Fourier space, where a pre-existing sinc window function served as a suitable initial value, and a modified backpropagation algorithm was applied to fine-tune the filter parameters. The idea is that the change in the error value can decrease/flatten the stop-band and the pass-band ripple separately, depending on the desired FIR filter, and simultaneously decrease transition bandwidth. The main idea of using a gradient descent optimizer in linear phase FIR filter design is to find filter coefficients, by simply minimizing the mean squared error function. There exist some forms of error functions for the filter design. Let us visualize the output responses of ideal and actual filters. As discussed in Chapter 1, the actual response in real-life problems would not be perfectly rectangular. Hence, the ripples in the pass, stop band and finite transition band. The most intuitive way to make the actual response closer to the ideal response is to minimize the ripples in both pass and stop bands and

make the transition band infinite. To achieve this objective we consider a weighted error function of the form

$$E_f = \min(\max_f |w(f)[H_d(e^{if}) - H(e^{if})]|), \tag{5.1}$$

where $H(e^{if})$ is the actual frequency response of the input function ($sinc$ in this case) and $w(f)$ and $H_d(e^{if})$ are window functions and desired frequency response, respectively:

$$w(f) = \begin{cases} 1 & 0 \leq f \leq f_p \\ 0 & f_p \leq f \leq f_s \\ k_s & f_s \leq f \leq \pi \end{cases} \tag{5.2}$$

$$H_d(f) = \begin{cases} 1 & 0 \leq f \leq f_p \\ 0 & f_p \leq f \leq \pi, \end{cases} \tag{5.3}$$

where f_p is the cut-off pass-band frequency, f_s is the cut-off stop band frequency. Ideally, we would have only f_p, but due to the finite sampling problem, we have, in practice, f_s. Let E_1 and E_2 be the error representations on the pass-band and stop-band ripples, respectively. We have:

$$\mathrm{E}_1 = \max_{f \leq f_p}(|\mathrm{E_f}| - \delta_p) + \max_{f \geq f_s}(|\mathrm{E_f}| - \delta_s), \tag{5.4}$$

$$\mathrm{E}_2 = \Big(|H(f) - H_d(f)| - \delta_p\Big) + \Big(|H(f) - \delta_s|\Big). \tag{5.5}$$

For clarity, all the components in Equation (5.4) are illustrated in Figure 5.1, where δ_p and δ_s are small enough pass band and stop band components that govern the ripple size, respectively. The main error to be minimized is the sum of E_1 and E_2. The total squared error is defined as:

$$\mathrm{E} = \arg\min_{f_i}(\mathrm{E}_1 + \mathrm{E}_2) \tag{5.6}$$

$$= \sum_{i=1}^{N} \Big(|H_d(e^{if_i})| - |H(e^{if_i})|\Big)^2. \tag{5.7}$$

Equation (5.6) is non-linear; hence gradient descent optimization could be used to find an optimal filter by minimizing the error difference between the ideal window filter and an actual

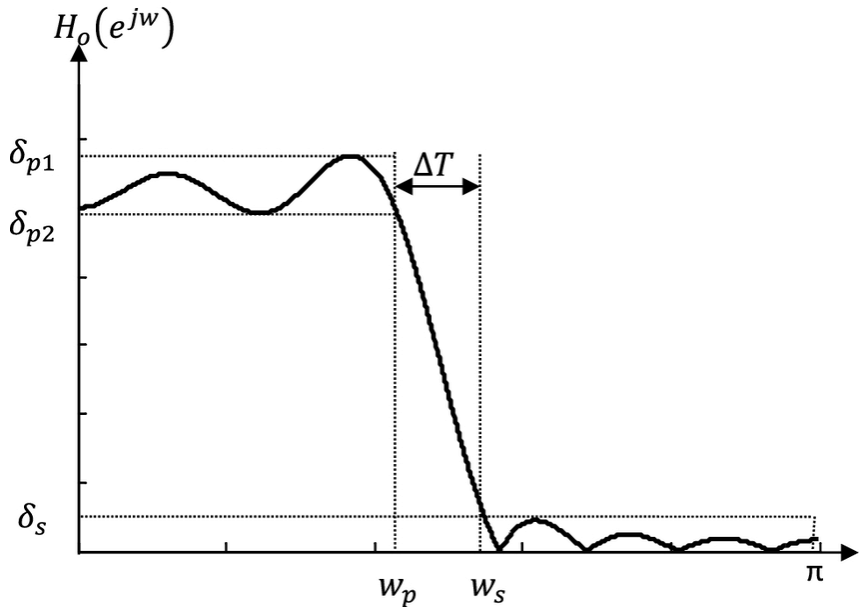

Figure 5.1: Window design: FIR ideal low-pass filter specifications vs a practical low-pass-like response (blue curve) obtained from a truncated $sinc$ transformation. Figure is taken from :Latifoğlu (2020)

FT of the $sinc$. Gradient descent-based algorithm considers a univariable or multi-variable function, $\mathrm{E}_f(\mathbf{w})$ which is defined and smooth (i.e. differentiable) in a neighbourhood of a point f. Also, $\mathrm{E}_f(\mathbf{w})$ descends fastest if one goes from the point $\mathbf{w}$ in the direction opposite to the gradient of $\mathrm{E}_f(\mathbf{w})$, i.e. at $(\mathbf{w}, -\nabla \mathrm{E}_f(\mathbf{w}))$. It follows that if we define the weight vector $\mathbf{w}(f) = (w(0), w(1), \cdots, w(N))$ to represent a set of parameters that the objective function depends on, in this case, the filter coefficients or the parameters that define the filter's response characteristics. Then, the objective function E(w) aim to minimize the difference between an ideal filter response and a realisable filter response parameterized by the vector $\mathbf{w}$. The goal is to find the optimal values for the filter coefficients that result in a filter response that closely approximates the ideal response. One then proceeds to modify the weight vector by taking the derivative of Equation (5.6) with respect to the parameter vector; this is done iteratively until the error in Equation (5.6) is minimal. In principle, this vector can have many components. If we denote each component as $\mathbf{w}_i$, then $\mathbf{w}_i$ becomes the i^{th} coordinate of the parameter of interest. Traditionally, $\mathbf{w}$ is randomly initialized. After this, a partial derivative of the cost function

is calculated with respect to $\mathbf{w}$ at the initial instance. The resulting vector of n input variables represents the positive direction of the function. Still, since the goal is to minimize the error, the derivative is multiplied by a negative sign which reverses its behaviour and causes it to descend to the minimum. However, reversing the gradient only concerns the direction to the minimum and does not say anything about how much. The magnitude by which the vector changes in the reverse direction is calculated by multiplying the value of the gradient at a point by some small enough constant λ. This constant is known as the learning rate, fixed along every direction of the steepest descent. The randomly initialized vector is then updated iteratively until a minimum is reached, and this process is known as backpropagation. The update of $\mathbf{w}$ follows:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \lambda \nabla \mathrm{E}_f(\mathbf{w}_n). \tag{5.8}$$

The pseudo-code of the error minimization is provided in Algorithm 1:

Algorithm 1 Gradient descent algorithm for filter design

procedure OPTIMIZATION OF THE FOURIER TRANSFORM OF BDWF
    Initialize $\mathrm{w}_1 = 0$
    calculate f_{max}
    if $\mathrm{E}_f(\mathbf{h}) \approx 0$ **then**
        stop
    else
        $-\nabla \mathrm{E}_f(\mathbf{w}_n) = -\frac{\partial}{\partial \mathbf{w}_n}(\mathrm{E}_f(\mathbf{w}_n))$
        for n from 1 to N **do**
            $\mathrm{w}_{n+1} = \mathbf{w}_n - \lambda \nabla \mathrm{E}_f(\mathbf{w}_n)$

Before we go any further with gradient descent, it is worth noting that the principle is valid for smooth functions, which means that the loss function is differentiable (Ruder, 2016) or at least partial in the search domain. However, an ideal low-pass filter is naturally non-smooth and exhibits a jump discontinuity at the cut-off (due to abrupt transition). In other words, the resulting rectangular pulse contains a high-peaked side lobes level because the abrupt time changes at its

edges increase its high-frequency content, resulting in higher and sharper side lobes. This leads to the high-frequency component being poorly suppressed, as shown in Figure 5.1. With that being said, we can conclude that: To mitigate this effect, we require that the transition boundary of the pulse be smooth, thereby resulting in much fewer high-frequency components and significant side lobes will be attenuated. The above premise leads us to the following remark: We either use smoothing windows or extend the window length in one domain to suppress the sidelobes. The latter is the overlap BDWFs.

To demonstrate further the ineffectiveness of realizing an optimal $sinc$ based on a perfect rectangular pulse. Having an infinity discontinuity would mean that most standard loss functions used in the optimization of the neural network will fail at the discontinuity depending on the particular loss function and the details of the optimization algorithm used. A possible solution around this comes from knowing precisely the matrices of the function we wish to optimize since it is an FT of a known, well-defined window function. Our matrices are pass-band and stop-band with a bit slow transition band; we then create a perfect low-pass filter by taking the values from the point in the transition band and collapsing them to pass and stop band classes creating an infinite slope transition. To better observe the pass-band shape, Figure 5.2 shows the response plotted logarithmically (i.e. decibels, dB):

Figure 5.2 shows explicitly that the peak side-lobe level of transformed $sinc$ (optimized-$sinc$ in the legend) is higher than that of the $sinc$, and the roll-off rate is slower than that of the $sinc$, which translates directly to much lower off-axis suppression. It is primarily due to the inherent characteristics of an ideal $sinc$ function. The FT of an ideal $sinc$ function has an infinite transition band, and truncating it to a finite length introduces undesirable side effects, such as the Gibbs' phenomenon. Figure 5.2 and 5.3 show the experimental results of this, from these figures, one can see that the resulting $sinc$ still exhibit higher sidelobes compared to the non optimized version of which this is the manifestation of the residual effects of the Gibbs phenomenon and inherent characteristics of inverse FT. In addition, even though one could also explore windowing functions technique which helps to smooth the shape transition and reduce sidelobes. However, this results in a more localized spectrum and in our case means less ability to shape the FoI, moreover, in the application of BDWFs, that would also lead to increasing the

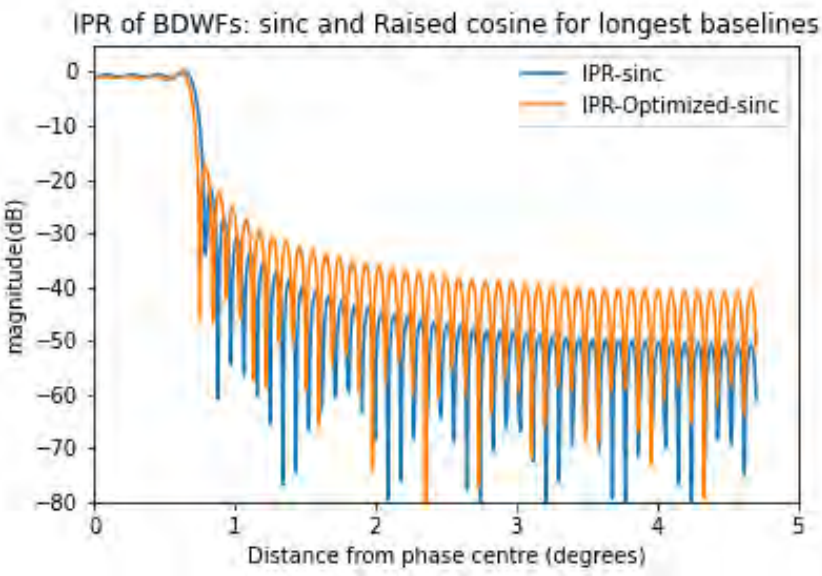

Figure 5.2: Window functions

IPR for a truncated $sinc$ and an optimized $sinc$

medium-sized baselines and short-sized baselines hence reducing number of longer baselines, resulting in poorer FoI shaping. Figure 5.3 shows the simulated results of applying the two $sinc$ windows with the IPR depicted in Figure 5.2.

The smearing factor represents source flux attenuation as a function of distance from the phase centre in the image domain. This flux results from the dirty image tapered by the top-hat function (inverse FT of $sinc$). The transformed $sinc$ shows that the off-axis suppression is lower, corresponding to high side-lobes and slow decaying side-lobes. This is evident that a top-hat function is impractical to realize as a filter with a frequency response of that shape is impractical.

5.4 Desired options for off-axis suppression

This section looks at possible off-axis suppression techniques, particularly avoiding under-sampling and using an optimal window function whose FT drops off rapidly outside the FoI. Here, we either look for a taper with the highest roll-off rate or find better criteria to sample such that a significant off-axis suppression is achieved. Conventional techniques usually control the win-

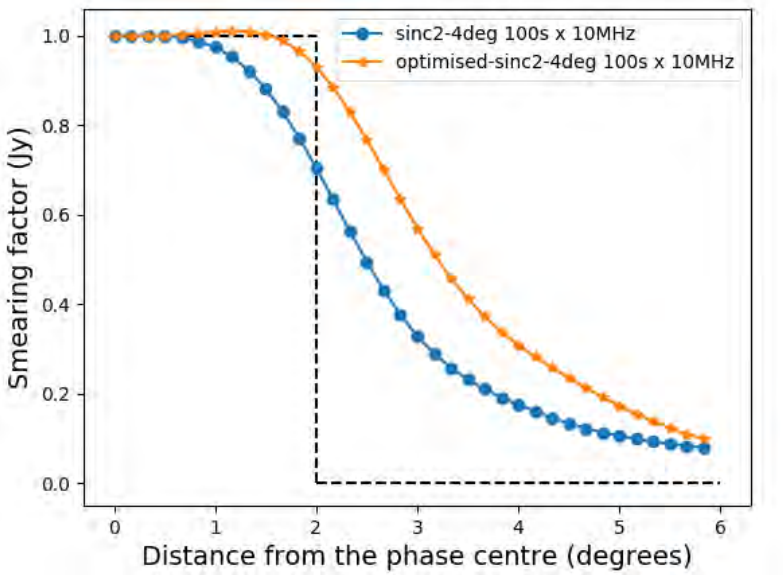

Figure 5.3: IPR induced by BDWF vs BDWF obtained from an ideal lowpass-like, with the abrupt transition. The window function used is the $sinc$

.

dow spectral characteristics by tuning several parameters of closed-form expressions given by a chosen window function. The commonly adjustable spectral characteristics of a $sinc$ window include the main lobe width (MLW), the peak sidelobe level (PSL), and the sidelobe roll-off rate (SLROR, Sun et al. (2019)), which are closely associated with the resolution and spectrum leakage or sensitivity (de Jesus Romero-Troncoso, 2016).

5.4.1 Raised cosine window

The waveform of the $sinc$ window with large bandwidth and infinity support makes it practically difficult to achieve as its frequency response has a perfect rectangular shape. We then explore raised cosine due to its adjustable waveform, controlled by a tunable roll-off parameter making it more practical and feasible. This window is an optimizable version of a $sinc$ because it is parameterized by a roll-off factor β (Alagha & Kabal, 1999). The raised cosine is defined as :

$$R_c(u,v) = \frac{sinc(u,v)cos(\beta u, \beta v)}{(1-2\beta u)(1-2\beta v)}, \tag{5.9}$$

where u,v and β are spartial frequencies and roll-off factor respectively. The good thing about raised cosine is that its shape varies smoothly on the closed interval [0,1]. Towards $\beta = 0$ raised cosine window becomes a $sinc$-like function, and towards $\beta = 1$, it becomes Gaussian-like. For these reasons, we propose that a raised cosine filter must be employed as it is a generalized $sinc$. The impact of the roll-off factor is subject to the user and the problem at hand. In this instance, one should not confuse the suggestion of generalizing the $sinc$ with calling raised cosine an optimal window function.

5.4.2 Window decomposition and recombination

A similar approach to parametrizing a $sinc$ with a roll-off factor is via (the decomposition and recombination) mechanism. Since the goal is to optimize a subset of chosen matrices of the window function, one considers taking a window function ($sinc$ in this case) and decomposing it into its Fourier modes (low spatial and high) or, equivalently, main lobe and sidelobes. The reason for doing this is that these two spatial frequency modes behave contrary; using a window

function to taper sidelobes causes the main lobe to broaden, and narrowing the main lobe causes high sidelobes, and that is so to preserve the properties of the window. Now that necessitates optimizing each metric independently and without interfering with the behaviour of the other. What is it that should be optimized? One notes the somewhat even distribution of values to the spatial frequency modes; a truncated $sinc$ tends to give more values to the lower and high spatial modes, causing wider main lobe and high sidelobes level, which increases the noise resulting in a low-resolution image and slower attenuation rate. This is mainly due to improper sampling and finite space representation. After decomposing the $sinc$ window, the function with a padded main lobe is then smoothed by applying a hamming function in order to minimise or taper the nearest side lobes. The resulting windowed function is then added to the smoothed high-frequency modes to reconstruct the $sinc$ function with only one difference which is minimal sidelobes:

Again, when applying this technique to the visibility data, we still observe no corresponding effect in the image plane regarding significant off-axis suppression. This is because BDWF being optimized is of a fixed baseline mapped into a fixed finite bin size. That particular baseline determines the average of these results in the finite-fixed bin. Hence no improvement in the shape of the IPR. This can be understood by considering the number of longer baselines optimized relative to the number of non-optimized short baselines. The shorter baselines will dominate the shape of the IPR.

5.4.3 Effect of density weighting and gridding convolution functions

This has motivated the slight change of angle in the proposed optimization scheme. This limitation suggests that perhaps one should look at the overall gridding process. Let us recall the expression for re-sampled visibilities:

$$\mathcal{V}_g(\mathbf{u}(t,\nu)) = \Big([\mathcal{V}_{pq}(\mathbf{u}(t,\nu))\underline{\mathcal{W}_{pq}(\mathbf{u}(t,\nu))}] * \underline{D^{\alpha}_{pq}(\mathbf{u}(t,\nu) - \mathbf{u}(t_k,\nu_l))} \Big) \underline{F(\mathbf{u}(t,\nu))}, \tag{5.10}$$

where $F(\mathbf{u}(t,\nu)) = \text{Ш}(\mathbf{u}(t,\nu)/\Delta\mathbf{u}(t,\nu))$ is the 2-D shah function. This says that the re-sampled visibility is a weighted measured visibility convolved with BDWF sampled on a chosen grid,

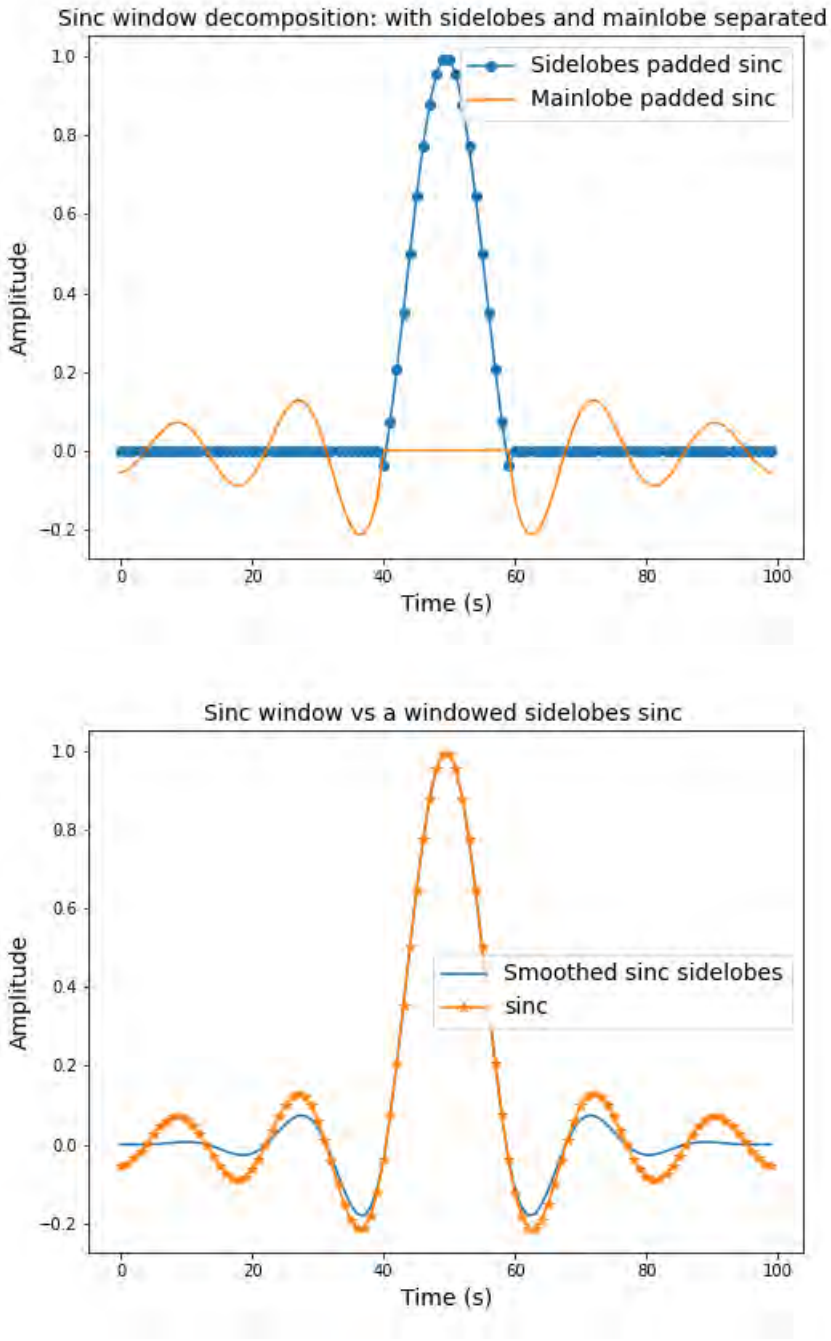

Figure 5.4: Top-panel: the $sinc$ is split into the main lobe and its associated sidelobes. Bottom panel: a hamming window is applied to only the sidelobes and compared with the original $sinc$, which we decomposed

where $\Delta\mathbf{u}$ represents grid spacing in the uv-space. $\mathcal{V}_g$ is normally not a true reconstruction due to Fourier effects of the function $\mathcal{W}_{pq}$, F and D^{α}_{pq}. The main effect in the image plane translates to, 1) multiplication by a truncated BDWF, 2) convolution by the PSF (Fourier transform of $\mathcal{W}_{pq}$), and 3) aliasing due to convolution with a FT of the comb function. For a proper reconstruction of $\mathcal{V}_q$, one has to correct the corruptions caused by the underlined functions. Let us look at these corruptions individually and examine them:

- An optimal BDWF is an infinite $sinc$ function, which is practically impossible. However, better approximations can be made by simply having more samples, in this context, a long enough baseline determined by a wider or larger sampling bin and larger FoI settings. The downside with wider sampling bins is 1): an increase in computation and 2): a decrease in sensitivity. With that being said, the obvious limitation to having an optimal BDWF is the FoI size and sampling intervals.

- This leaves us with two functions to alter $\mathcal{W}_{pq}$ and F:

- First, we examine $\mathcal{W}$, looking at the MeerKAT telescope, the density of the visibility samples in the uv-domain is highly non-uniform. This non-uniformity limits the ability of the re-sample convolution and, hence, the failure of achieving an optimal matched filter for FoI shaping. This strongly suggests that we must employ some compensation on the sample density, which means we are looking for an optimal weighting scheme. The solution to our density sampling problem is to modify the sampling function to a sum of weighted impulses. Such weighting function exists, i.e. uniform weighting and its derivatives.

- Lastly, corrections for F: one needs to use large grids on our uniform space to impose a strong bias in our chosen weighting density function. This spacing size ensures that samples in dense regions will be highly weighted, and samples in sparser regions will be over-weighted. This mechanism ensures equity in compensating for sample density. We shall apply these suggestions in chapter 6.

5.5 Conclusion

In summary, this chapter discussed how one should interpret to shape the FoI of an interferometer. For this, we justified the optimization through a different angle by simply redefining explicitly what to optimizE. This follows after we explored how do filter design theory can tune window functions to achieve the desired tapering response in both time and frequency spaces. We also showed the limitations of filter design to obtain an optimal matched filter under certain conditions, particularly the use of FFT to transform a time-frequency dependent signal to an inverse response in the image plane in designing a matched filter and using shorter filter orders or window sizes (this outlines a sampling problem). Most importantly, we established that finding only a matched filter is insufficient in achieving large FoI shaping due to the design nature of radio interferometers.

CHAPTER 6

Results and discussion

6.1 Introduction

This chapter reports the performance of BDWFs with several weighting schemes to improve FoI shaping. The JVLA in C-configuration and the MeerKAT telescopes are used for the evaluation.

6.2 MeerKAT and JVLA-C

The MeerKAT telescope consists of 64 antennas of a 13.5 meters diameter dish scatted over an 8 km radius. The 64 antennas are unevenly distributed such that 75% of them are centrally placed within a 1 km diameter "core region", and the 25% spread out over a 4 km radius, and approximately half of all MeerKAT baselines are centrally located. Depicted in Figure 6.1 is the number of baselines as a function of their length for both MeerKAT and JVLA stations.

The non-uniform distribution of the baselines in these two antennas manifests as follows in terms of FoI shaping: In Figure 6.2, the smearing behaviour at the same sampling rate for MeerKAT and JVLA-C differs more in the out-of-FoI suppression. The suppression rate in this

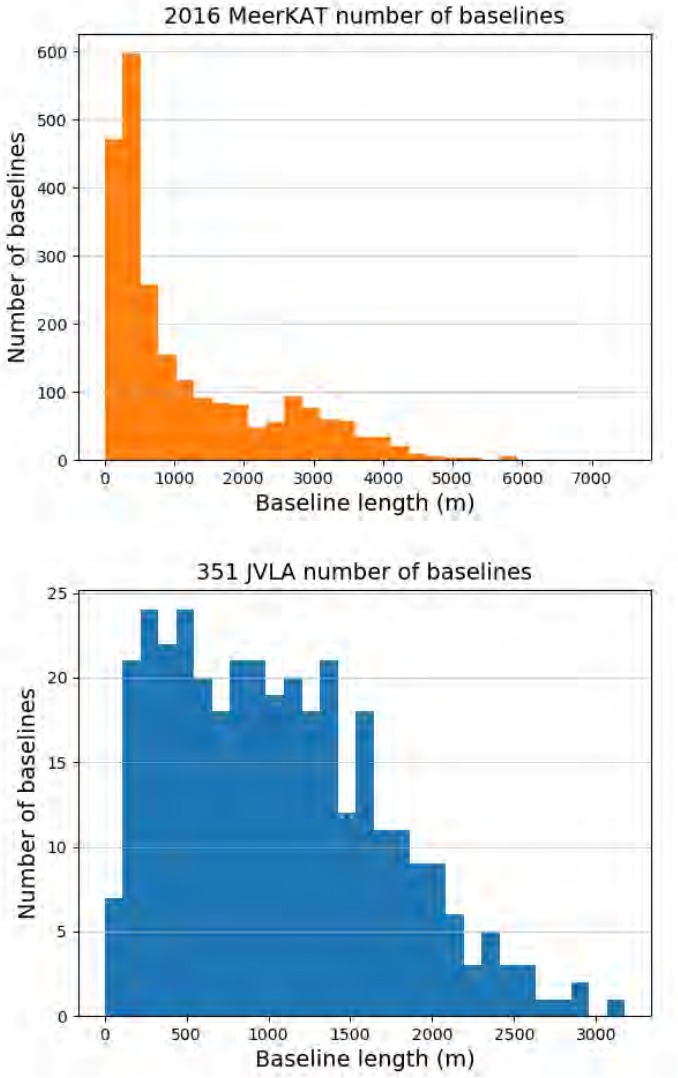

Figure 6.1: The number of baselines as a function of baseline lengths for MeerKAT and JVLA-C stations, baseline lengths range between 29 meters and 7474 meters for MeerKAT and from 45 meters to 3400 meters.

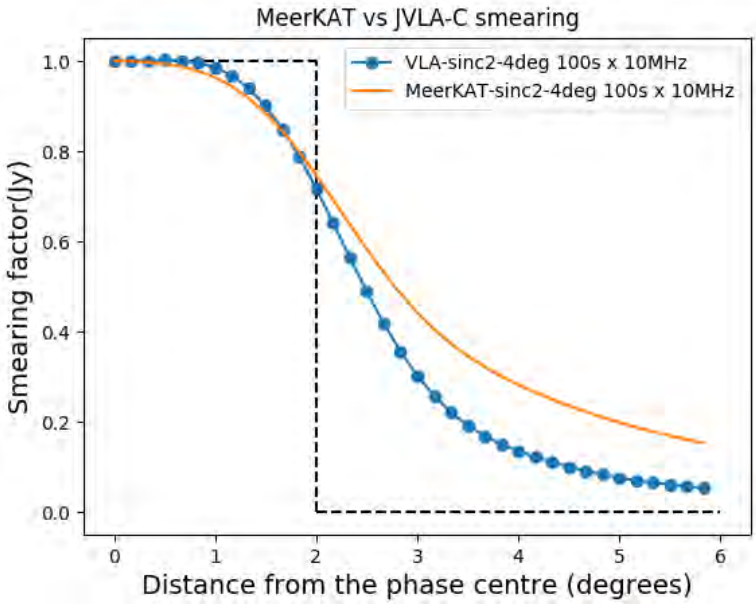

Figure 6.2: The smearing factor as a function of distance from the phase centre for MeerKAT and JVLA-C at the same sampling interval of 100s $\times$ 10MHz at 4 deg FoI.

regime (outside FoI) is no longer the function of all combined factors that cause smearing but is mostly the function of the array configuration. Similarly, the within-FoI regime is a function of shorter baselines. This implies that overall, for the same sampling rate (e.g., bin size), same FoI, and relative distance from the phase centre, the ability of a chosen BDWF to shape the FoI is dependent on the antenna distribution.

6.3 Effect of array configuration on BDWFs

It is known that the efficiency of BDWFs in shaping the FoI is strongly affected by an array configuration. Modern interferometers such as the MeerKAT a core array with many short baselines compared to the number of long baselines. For small finite fixed bin sizes, the BDWFs become more truncated resulting in more boxcar-like windows in the uv-plane, as a consequence, the FoI becomes narrower. As a result, optimization becomes unpractical. However, we might still apply BDWFs on the current array design, but one would need to carefully treat the gridding process to

account for sample distribution and weighting. Figure 6.2 shows the smearing response from applying the same BDWF to different antenna stations. We see that the IPR induced by MeerKAT simulation rarely shapes the FoI as compared to the one obtained from the JVLA-C station. More suppression is due to long baselines being more sensitive to smearing, and JVLA-C has more of them than MeerKAT. This is in consistent agreement with Figure 6.1, which shows the number of baselines as a function of baseline length. Note that the discussion only applies to FoI shaping, and for a large amount of bins, BDWFs can shape the FoI of the MeerKAT telescope and achieve substantial data compression.

6.4 Optimization via fine parameter tuning

In this section, we reformulate our optimization problem to find the best combination of parameters that would yield a wider FoI shaping. This comes after a realization that optimization of the BDWF is not the unique function of FoI shaping. Alternatively, we explore a more generalized form of $sinc$, which is a raised cosine, as described in Section 6.5.1. Additionally, we extend the focus by looking into sampling, particularly gridding the uv-data in a grid of small cells using a shah function $\text{Ш}(\frac{u}{\Delta u}, \frac{v}{\Delta v})$, to minimize aliasing we chose $\frac{1}{\Delta u}$ to be large enough and this also ensure our scaling factor is small. We highlight a few parameters that improved the results demonstrated in Section 5 (changing the density weighting scheme from *natural* to *Briggs* (Briggs, 1995)) with 0.0 robustness, using grid cell size of 1.0 arcsec of an image of size 512×512 pixels. This modification considers the effect of array configuration by using a weighting scheme that brings balance or uniformizes to the uv-data points; the decrease in cell size minimizes artefacts, the result of the FFT.

6.4.1 Image parameters, density weighting, and tapering

The previous work on BDWFs uses natural weighting and $sinc$ BDWF with large cell size (10 arcsecs) of a small image of size 512×512. With natural weighting, the image parameters do not play a significant role. However, if one opts for either Briggs or uniform, then these parameters

must be treated carefully. Some form of parameter compatibility exists that ensures optimal visibility weighting and BDWF. We are now forced to employ a density compensating weighting that combines uniform and natural weighting features, balancing the resolution, sensitivity, and beam shape while balancing the conflicting requirements against each other reasonably associated with BDWF. Let us first tackle the image parameters (image size and grid cell size): These parameters are generally arbitrary; due to gridding problems in small images, we choose to decrease cell size per grid to ensure better sampling. Secondly, regarding density weighting selection criteria, for a heavily sampled array such as MeerKAT, a weighting scheme must be selected such that there is some equity in the weighting of uv-points. Fortunately, such a weighting scheme exists and is *uniform* weighting. This is in line with the statement by Briggs (1995), which states that for heavily dense arrays, it is usually best to use super-uniform weighting than the standard uniform and then balance the RMS noise with appropriate robustness. In summary, uniform weighting gives more weight to sparser regions, precisely the Fourier modes with more smearing than natural weighting does (this is because of the short period of longer baselines spent per uv cell than short baselines for a finite observation). Uniform weighting produces narrower beams since it fills the uv-plane more regularly. However, its disadvantage is due to the discontinuity of the weights at the grid edges for each sampled portion of the uv-plane, causing more pronounced sidelobes, which increases the confusion noise and compromises SNR. One way to solve this is to couple the uniform weighting with a BDWF. We should also notice that having weighted density and BDWF will affect the overall image sensitivity and increase noise. To counteract this, one needs to be considerate with weighting, which controls the degree of up-sampling of isolated points in the uv-plane. This brings us to the realization that we need a factor that ensures a careful up-sampling. Therefore, a robust parameter will do.

One can attempt to justify this mathematically. To initialize the process, we first acknowledge the work done by Briggs (1995) introducing *Robust weighting*. The critical element of the analysis is based on how the re-sampled visibilities should be weighted such that the applied BDWFs induce the desired IPR. For an efficient application of BDWFs, one should strongly consider the array distribution and observation length, and this is to find the best combination of the BDWF parameters and density weighting such that the IPR approximates an ideal low pass filter at a

given cut-off. Generally, when visibilities are transformed from Fourier to real space, the fundamental procedure normally starts with gridding, then applying density weighting and tapering. In gridding, the visibilities are transformed from irregular to regular representation form, the uv-plane is changed from elliptic to rectangular, and the rectangular plane is referred to as a regular grid with grid blocks being grid cells. This is all done by applying a shah or comb sampling function to the visibilities, the shah function is the function of the size of the gridding box in the Fourier domain, and here, we make this size large by setting the ratio of the maximum values of u and v representable on the gridded uv-plane and the maximum values of u and v present in the dataset to be large. This then translates to a much smaller amount $\Delta m = \Delta l = 1.0$ arcsec separating the pixels along the l and m direction for the image of size 512×512 (the image size is free and flexible).

Secondly, we weigh these gridded visibilities; in doing so, we are interested in the total gridded weight given by:

$$W_{pq}(u,v) = D^{\alpha}_{pq}(u,v) De_{pq}(u,v) w_{pq}(u,v), \tag{6.1}$$

where D^{α}_{pq} is a BDWF for gridded visibility pq , De_{pq} is a density weight for gridded visibility pq and w_{pq} signal-to-noise weight for gridded visibility pq. Here we shall start by showing why natural weighting rarely gives the desired shape. First case: if we naturally weight the visibilities, then De_{pq} is 1, and all samples are weighting the same in the uv cell. Now Equation (6.1) reduces to a product of BDWF with the signal-to-noise weight of each gridded visibility. The only additional variance is now added by the BDWF since it is a weighted approach as compared to a simple average and will overall result in relatively small noise. This means the impact of the BDWF is only significant on individual pq samples and their variance. The density of all the samples overcomes the ability of the BDWF, resulting in a partially shaped FoI. This is exactly what is shown in Atemkeng et al. (2016). From these results, it is clear that to shape the FoI somewhat fully, the visibilities must be tapered by a weighted average, e.g., a relatively sample-sensitive factor. This overcomes the power of sample density in our tapering function, creating an interdependency between density weighting and tapering, which improves FoI shaping. The derivations for Equations (6.3) and (6.4) for Briggs weighting are taken from (Briggs, 1995).

This necessitates us to base our attention on making the term $De_{pq} \neq 1$; now, this implies either uniform weighting or more flexibly robust weighting. As we have mentioned and justified, we shall consider robust density function for a gridded visibility pq as :

$$De_{pq} = \sum_{k \in c_{pq}} D_k \tag{6.2}$$

$$= \frac{constant}{\left(\sqrt{\frac{2}{w_{pq}}}\sqrt{w_k}\sigma_k\right)^2 \sum_{k \in c_{pq}} w_k + 2w_k\sigma_k^2}, \tag{6.3}$$

where σ_k is the variance associated with each visibility k, D_k is the density weight for visibility k, w_k is the signal-to-noise weight for visibility k. We note that Equation (6.3) holds whenever the sum of the signal-to-noise weight for each visibility k over a gridded cell pq is strictly positive; otherwise, De_{pq} varnishes. Equation (6.1) can be rewritten as :

$$W_{pq} = D_{pq}^{\alpha} \frac{constant}{\left(\sqrt{\frac{2}{w_{pq}}}\sqrt{w_k}\sigma_k\right)^2 \sum_{k \in c_{pq}} w_k + 2w_k\sigma_k^2} \Big(\sum_{k \in c_{pq}} \frac{1}{\sigma_k^2} / \sum_i \frac{1}{\sigma_i^2} \Big). \tag{6.4}$$

Equation (6.4) provides us with a more robust way to weigh the resampled visibilities; this shows that the weighting is now sensitive to the sample distribution. Now each visibility will be weighted proportional to its gridded cell pq with dependency in w_k and σ_k^2. This is the essence of robust weighting. Suppose $w_k\sigma_k^2$ is much larger than σ_{pq}. In that case, the dirty image we are minimizing looks like the dirty beam tapered by the BDWF meaning W_{pq} approaches a value different from 1, and we recover uniform weighting. If the converse, we essentially minimize over an image of random noise tapered by a raised cosine, and we recover natural weighting (Briggs, 1995). Equation (6.3) is mainly controlled by the squared first term in the denominator, reduces this term mean we are only left with some scaled signal-to-noise weight of visibility k resulting in either: 1) a constant density for all visibility if the numerator gets close to the proportionality constant in the numerator. 2) a larger density for gridded visibilities dominated by proportionality constant dependent on the signal-to-noise weight of each visibility. On the other hand, if we increase the squared first term in the denominator, then the summed-up signal-to-noise weights of each visibility will be improved, resulting in lowered density for gridded visibilities. This

implies that this term specifies the robustness and can be normalized to some range of values. Having specified this value, one proceeds to taper the optimal weighting visibilities, resulting in optimal gridded-weighted-tapered visibilities and a better FoI shaping.

6.5 Simulations

We begin this section by presenting how BDWFs are used to average visibility data. Furthermore, this is then extended by studying their effect on smearing and out-of-FoI source suppression. These simulations are performed based on interferometer configuration corresponding to MeerKAT, observing at 1.4 GHz. The procedure of the simulations is similar to that presented in section 5.2.4. We create the high-res MS, which is sampled at 1 s for a total observation time of 400 s, with 30 MHz total bandwidth sampled at a spectral resolution of 83.4 kHz, resulting in 360 frequency channels. A corresponding low-res dataset is then generated by simply re-sampling the hi-res with a BDWF of sizes 100 s $\times$ 10 MHz, 50 s $\times$ 5 MHz and 20 s $\times$ 2 MHz. The visibilities are then translated to the image using the csCLean imager with Briggs weighting of robustness 0.0. The low-res robust weighted visibility data is then transformed into dirty images, and the flux is measured.

Figures 6.3, 6.4, 6.5 and 6.6 show the amplitude response of several BDWFs at various sampling rates and FoI settings. There is a vertical line at 1 and 2 degrees FoI, which shows the ideal off-axis suppression. BDWFs' performance, is evaluated by comparing the amplitude response with these lines. The more the amplitude response approximates these lines, the better the performance of a given BDWF. The idea is the same in all the Figures, except that of Figures 6.3, which shows the BDWFs for natural and Briggs weightings for several uv sampling, the bold curves indicate Briggs plus BDWFs and the dotted curves indicate natural weighting plus BDWFs. Moreover, there are two grey curves that indicate normal averaging (with boxcar window) for natural and Briggs weighting. Figure 6.4 shows only Briggs weighted BDWFs, at 100s $\times$ 10 MHz and 50s $\times$ 5 MHz sampling bins at 1, 2 and 4 deg FoI settings. The bold curves represent BDWFs performance at 50s $\times$ 5 MHz sampling bin and the dotted curves represent BDWFs performance at 100s $\times$ 10 MHz sampling bin. Figures 6.5 and 6.6 show the performance at the

highest sampling rate and largest FoI setting. In Figure 6.5, the red region shows the conserved amplitude in terms of the area bounded by the ideal amplitude response and the one induced by a given BDWFs with a given weighting function. While the green shows the residual amplitude as a result of suppression. To indicate better response, both the green and red regions must be small in area, which is apparent in the top panel of Figure 6.5, and that is where BDWFs are applied with Briggs weighting. The bottom-panel of Figure 6.5 shows BDWFs applied with natural which rarely shapes the FoI. Nonetheless, BDWFs are good at conserving amplitude but have very poor performance at off-axis suppression. In all the Figures, one can notice that the FoI shaping through BDWF with Briggs weighting is much improved compared to that produced through natural weighting. We omitted uniform, and this is because it produces the same result as Briggs, for zero robustness. This is better demonstrated through Table 6.1, where FoI shaping is quantified by the area of conserved to the suppressed signal.

Table 6.1: Optimal window function performance results

Field of views	Integrations/ bin size	Conserved signal (%)	Suppressed signal (%)
1 deg	100 s $\times$ 10 MHz	86	80
1 deg	50 s $\times$ 5 MHz	88	70
2 deg	100 s $\times$ 10 MHz	92	91
2 deg	50 s $\times$ 5 MHz	86	88
4 deg	100 s $\times$ 10 MHz	97	83
4 deg	50 s $\times$ 5 MHz	94	82
Ideal BDWF (1, 2, 4 deg)	100 s $\times$ 10 MHz	100	100

6.5.1 Limitations of BDWFs applied with Briggs weighting

The drawback with applying BDWFs with various weighting scheme always result in a trade-off between a maximum achievable FoI shaping and sensitivity depending on a chosen FoI and sampling bin size. Briggs weighting is a reasonable compromise between FoI shaping and sensitivity

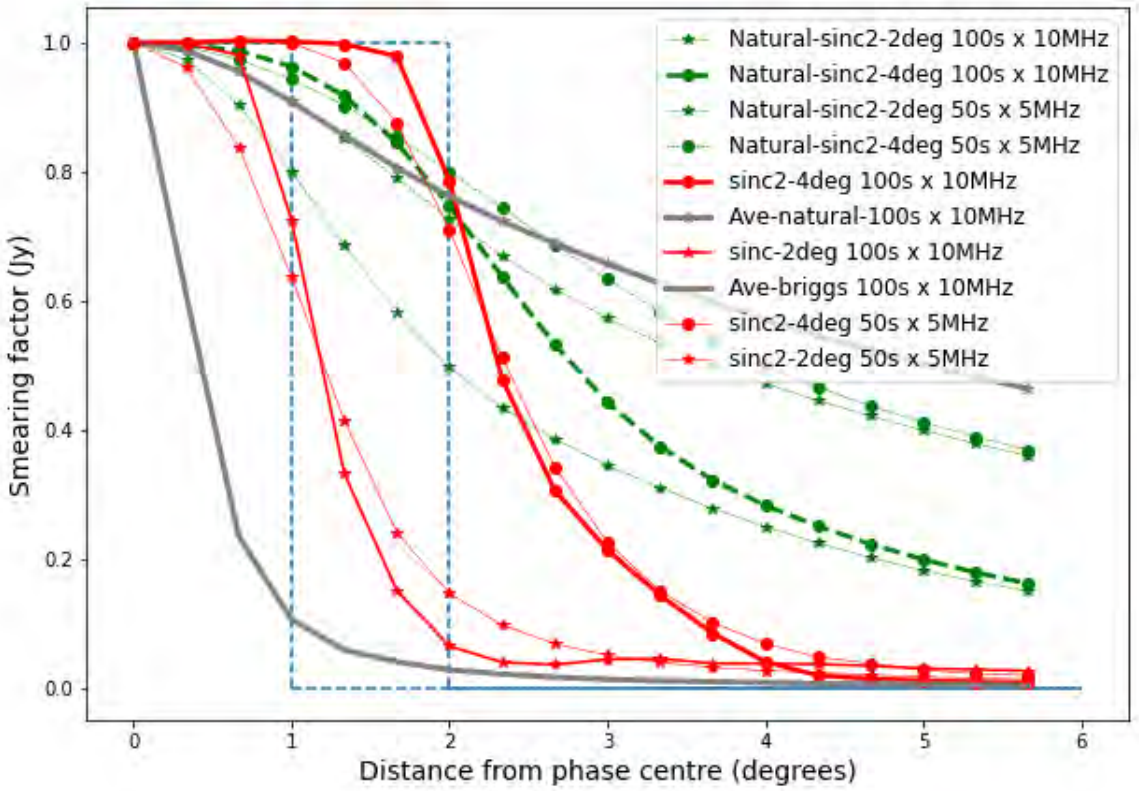

Figure 6.3: BDWF performance with MeerKAT at 1.4 GHz for various sampling intervals at 1, 2 and 4 degrees FoI, respectively; density weighting (robustness = 0.0) + BDWF

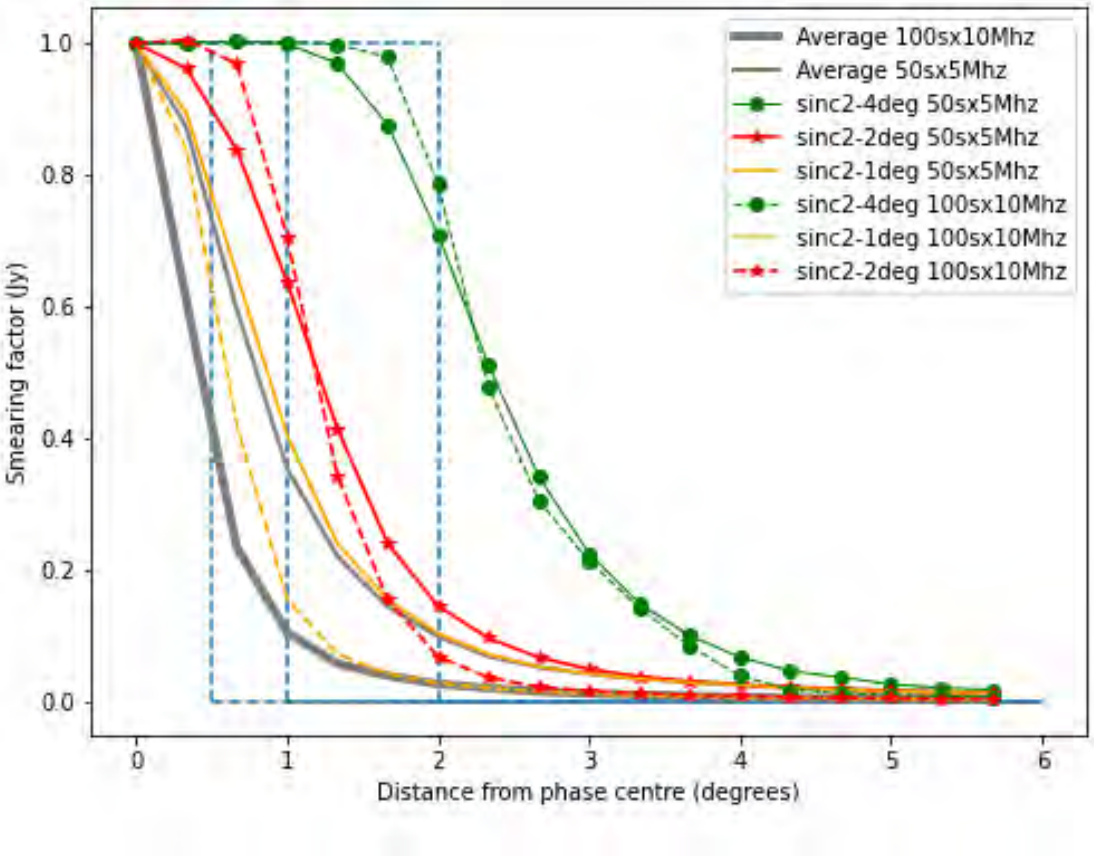

Figure 6.4: An example of BDWFs with Briggs weighting of robustness = 0.0

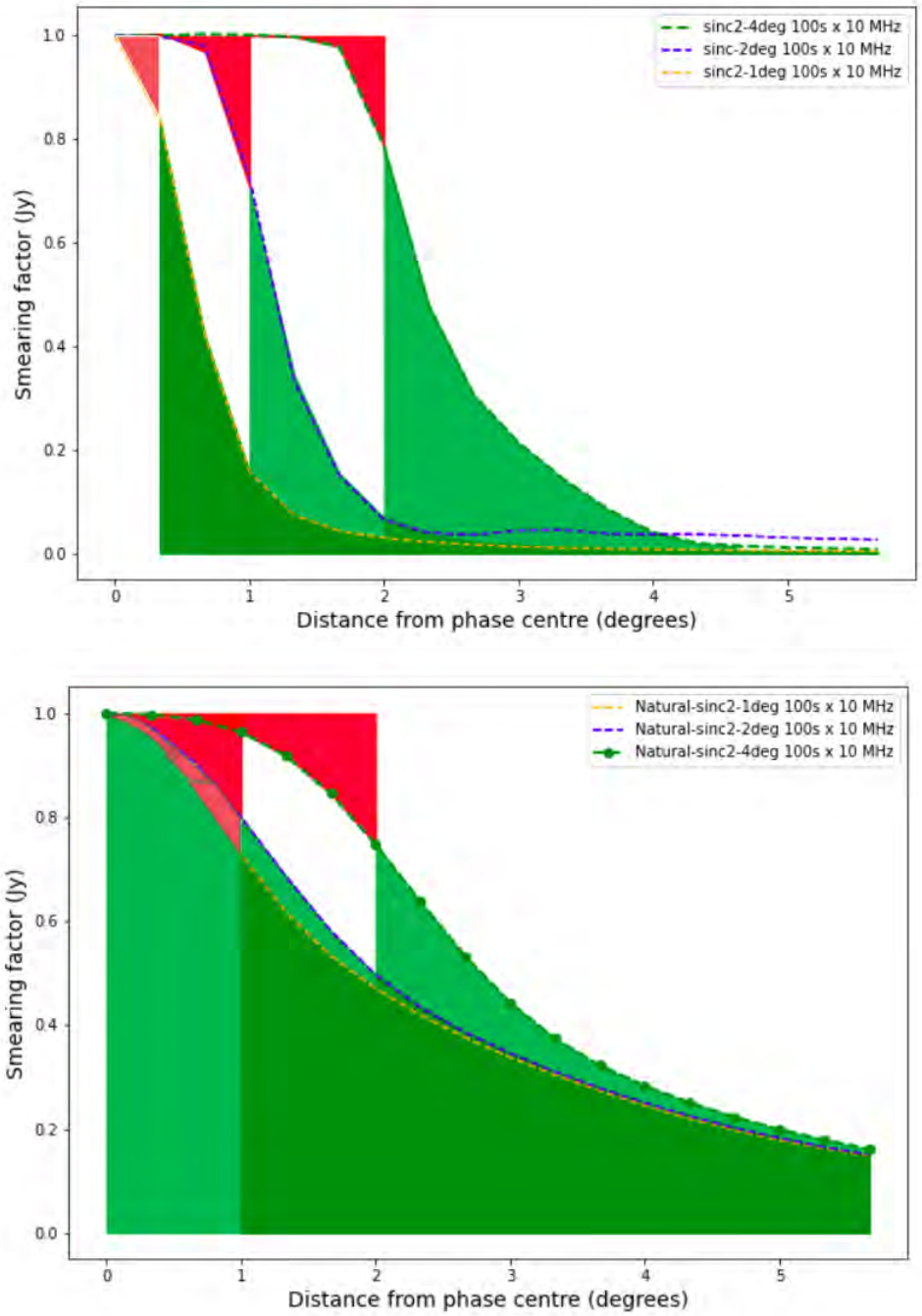

Figure 6.5: BDWF at 100 s $\times$ 10 MHz bin average at 1, 2 4 deg FoIs; top-panel: areas of interest showing improved performance of BDWFs application; bottom: sub-optimal BDWFs application

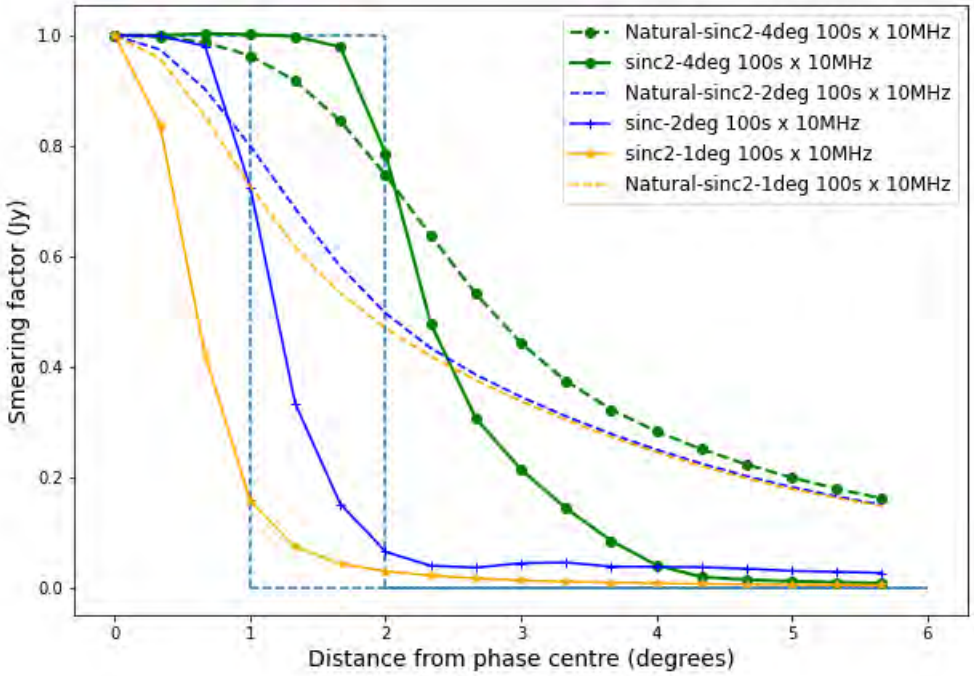

Figure 6.6: BDWF at 100 s $\times$ 10 MHz bin average at 1, 2 4 deg FoIs

when applying BDWFs with a density weighting scheme. The compromise comes about due to robustness (R). R is a sample weighting parameter that determines the magnitude on how much uv samples should be down-weighted or up-weighted. This parameter changes its behaviour based on the noise variance for averaged visibilities. Specifying the R-value gives a reasonable transition between natural and uniform weighting. R is generally taken from a closed interval $[-2, 2]$ where at the endpoints, we have: $R = +2$, which is close to natural weighting (most robust against the resolution) at a value of $R = -2$, which is close to uniform weighting (most robust against noise). For all $R \in (-2, 2)$, the useful regimes in robust weighting are usually obtained by moving inwards from the endpoints on either side. Firstly, moving in from the uniform weighting limit, one can often obtain a significant decrease in thermal RMS while giving up very little in resolution. This corresponds to a balanced trivial robustness of 0.0. Working downwards from the naturally weighted limit, the robustness of 1 results in a somewhat improved resolution with almost no less sensitivity (see Briggs (1995)). Applying robustness of 0.0 with BDWFs for various FoI settings and sampling bins results in an amplified noise compared to natural weighting with BDWDF for a particular FoI.

6.6 Conclusion

In this chapter, the results show that FoI shaping can be sufficiently attained. The justification of the parameter choice is given and explained in detail. We have also shown that there is a trade-off between FoI shaping and sensitivity for a chosen sampling interval. In other words, the size of the sampling bin is somewhat limited by the FoI as large FoIs result in BDWFs with noise scaled up. This emphasizes a threshold on the period of a BDWF used to average the visibilities. One can also establish that the choice of a BDWF sampling interval should depend on the scientific goal of the observation: by this, we mean if one is more interested in attending the best resolution, then a BDWF with a large period and wide FoI is suitable. If the goal is to maximize sensitivity, then the observation should use narrow sampling intervals with small FoIs which have compression limitations.

CHAPTER 7

General conclusion and future work

The current and future radio interferometers promise to reach their full scientific potential through their capability of sampling at high time and frequency resolutions. This generates far higher data rates, and these data rates are considered to be computationally prohibitive for current processing facilities. This is tackled traditionally by simply averaging the data across time and frequency. However, this leads to smearing/decorrelation, which is the decrease in source amplitude at the FoV and the distortion of the PSF. To reduce this amplitude loss, BDWFs are used as a two-fold solution. BDWFs prevent the amplitude loss on the achievable FoI compared to simply averaging and suppresses more signal at the edges of a FoI while the data is compressed at an acceptable level.

BDWFs are computationally expensive and require a lot of data in the Fourier space to achieve their compression capabilities. This means that for snapshot observations where data is limited, BDWFs could only be employed for FoI shaping and not for data compression. In this thesis, I studied the drawbacks of BDWFs in detail and investigated a machine-learning optimization technique using a gradient descent algorithm to optimize the window functions employed with BDWFs. The following were achieved in this thesis:

- I have Studied and discussed the effect of traditional averaging of visibility data.
- I discussed BDWFs and their limitations.
- I moved on to study several window functions in the signal processing field and established their relationship with BDWFs.
- The window functions employed with BDWFs are not optimized and are used as they are employed in signal processing. I proposed using gradient descent to optimize the window functions, which as a result, could shape the FoI to a greater extent compared to using traditional BDWFs without optimizing the window functions. The optimization, however, became unpractical due to the band-limited nature of the window functions in the Fourier space.
- I investigated BDWFs performance applied with several density weighting.

Using Briggs weighting, I have shown that a more enlarged FoI can still be attainable with even more significant suppression of unwanted signals outside the FoI. With this, ~ 4 % smearing is observed within the FoI, and ~ 80 % source suppression is achieved outside the FoI using the MeerKAT telescope at 1.4 GHz, sampled at 1 s and 184.3 kHz then averaged with BDWFs to achieve a compression factor of 4 in time and 3 in frequency. For an effective and enlarged FoI shaping, one must grid the visibilities such that an array geometry is compensated for. This work opens the door to several future works. It would be interesting to apply and evaluate the BDWFs' performance using real data while applying several weighting schemes. A broad investigation of approximation theories to solve the optimization of the FoI could also be a good avenue for future work since gradient descent is an iterative method.

Bibliography

Akgün, E., & Demir, M. 2018, International Journal of Assessment Tools in Education, 5, 491

Akhmanw, C., & Khokblov, R. 1959, Phys. Rev, 1, 1350

Alagha, N. S., & Kabal, P. 1999, IEEE transactions on Communications, 47, 989

Alam, S. S., & Hasan, M. T. 2010, International Journal of Electrical & Computer Sciences IJECS-IJENS, 10

Alexander, W., & Williams, C. M. 2016, Digital Signal Processing: Principles, Algorithms and System Design (Academic Press)

Alwahab, D. A., Zaghar, D. R., & Laki, S. 2018, in 2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP), IEEE, 1–4

Andrews, K., & Rajiv, B. 2007, Journal of Mathematical Analysis and Applications, 56, 237

Atemkeng, M., Smirnov, O., Tasse, C., Foster, G., & Jonas, J. 2016, Monthly Notices of the Royal Astronomical Society, 462, 2542

Atemkeng, M., Smirnov, O., Tasse, C., et al. 2018, Monthly Notices of the Royal Astronomical Society, 477, 4511

Balmer, L. 1997, Signals and systems: an introduction (Prentice Hall)

Bonnassieux, E., Edge, A., Morabito, L., & Bonafede, A. 2020, Astronomy & Astrophysics, 637, A51

Booth, R., & Jonas, J. 2012, African Skies, 16, 101

Bridle, A., & Schwab, F. 1999, in Synthesis Imaging in Radio Astronomy II, Vol. 180, 371

Bridle, A. H., & Schwab, F. R. 1989, in Synthesis Imaging in Radio Astronomy, Vol. 6, 247

Briggs, D. S. 1995, Ph. D. Thesis, New Mexico Institude of Mining and Technology

Burke, B. F., Graham-Smith, F., & Wilkinson, P. N. 2019, An introduction to radio astronomy (Cambridge University Press)

Callaghan, D. J. 2015, PhD thesis

Chang, C. C. 1958, Transactions of American Mathematical Society, 88, 467

Cooley, J. W., & Tukey, J. W. 1965, Mathematics of computation, 19, 297

Cornwell, T., & Wieringa, M. 1996, The Generic Instrument: V Design of Calibration and Imaging

de Jesus Romero-Troncoso, R. 2016, IEEE Transactions on industrial informatics, 13, 1291

Deeming, T. J. 1975, Astrophysics and Space Science, 36, 137

Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. 2009, Proceedings of the IEEE, 97, 1482

Dongare, A., Kharde, R., Kachare, A. D., et al. 2012, International Journal of Engineering and Innovative Technology (IJEIT), 2, 189

Felli, M., & Spencer, R. 1990, Space science reviews, vol. 54, no. 3/4, p. 453 (1990), 54, 453

Felli, M., & Spencer, R. E. 2012, Very Long Baseline Interferometry: Techniques and Applications, Vol. 283 (Springer Science & Business Media)

Gerla, B. 2004, in ISMVL '04: Proceedings of the 34th International Symposium on Multiple-Valued Logic (Washington, DC, USA: IEEE Computer Society), 49–54

Hamaker, J., Bregman, J., & Sault, R. 1996, Astronomy and Astrophysics Supplement Series, 117, 137

Hjellming, R. 1989, in Synthesis Imaging in Radio Astronomy, Vol. 6, 477

Hornik, K., Stinchcombe, M., & White, H. 1989, Neural networks, 2, 359

Jansky, K. G. 1933, Nature, 132, 66

Joaquim, M. B., & Lucietto, C. A. 2011, Digital signal processing, 21, 690

Karlik, B., & Olgac, A. V. 2011, International Journal of Artificial Intelligence and Expert Systems, 1, 111

Kemball, A., & Wieringa, M. 2000, URL: http://casa. nrao. edu/Memos/229. html

Knerr, S., Personnaz, L., & Dreyfus, G. 1990, in Neurocomputing (Springer), 41–50

Krichbaum, T., Witzel, A., & Zensus, J. 1999, SKA in VLBI. Technical report

Kucharavy, D., & De Guio, R. 2011, Procedia Engineering, 9, 402

Kumar, V., & Purwar, M. V. 2017, International Journal of Engineering and Technical Research, 7

Kurban, A. O. 2004, Industrial lubrication and tribology

Latifoğlu, F. 2020, Neural Computing and Applications, 32, 13323

Lovell, B. 1985, Oxford [Oxfordshire]; New York: Oxford University Press

McCulloch, W. S., & Pitts, W. 1943, The bulletin of mathematical biophysics, 5, 115

Mohamed, Z. E. 2019, Journal of the Egyptian Mathematical Society, 27, 1

Murphy, E. 2018, Science with a Next-generation Very Large Array, Astronomical Society of the Pacific conference series (Astronomical Society of the Pacific)

Nair, V., & Hinton, G. E. 2010, in Proceedings of the 27th international conference on machine learning (ICML-10), 807–814

Napier, P. J., Thompson, A. R., & Ekers, R. D. 1983, Proceedings of the IEEE, 71, 1295

Neves, A. C., González, I., Leander, J., & Karoumi, R. 2018, in Experimental Vibration Analysis for Civil Structures: Testing, Sensing, Monitoring, and Control 7, Springer, 73–84

Noordam, J. E., & Smirnov, O. M. 2010, Astronomy & Astrophysics, 524, A61

Offringa, A., McKinley, B., Hurley-Walker, N., et al. 2014, Monthly Notices of the Royal Astronomical Society, 444, 606

Prestage, R. M., Constantikes, K. T., Hunter, T. R., et al. 2009, Proceedings of the IEEE, 97, 1382

Robishaw, T., & Heiles, C. 2009, Publications of the Astronomical Society of the Pacific, 121, 272

Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, Astronomy & Astrophysics, 558, A33

Ruder, S. 2016, arXiv preprint arXiv:1609.04747

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, nature, 323, 533

Ryle, M., & Hewish, A. 1960, Monthly Notices of the Royal Astronomical Society, 120, 220

Scaife, A. 2020, Philosophical Transactions of the Royal Society A, 378, 20190060

Scaife, A., et al. 2019, Analysis of compute load, data transfer and data storage anticipated as required for SKA Key science

Shanks, J. L. 1967, Geophysics, 32, 33

Sharma, S., Sharma, S., & Athaiya, A. 2017, Towards Data Sci, 6, 310

Smirnov, O. M. 2011, Astronomy & Astrophysics, 527, A106

Sullivan III, W. T. 2009, Cosmic noise: A history of early radio astronomy

Sun, Y., Fan, H., Mao, E., Liu, Q., & Long, T. 2019, IEEE transactions on aerospace and electronic systems, 56, 2835

Taylor, G. B., Carilli, C. L., & Perley, R. A. 1999, Synthesis Imaging in Radio Astronomy II, 180

Thompson, A., Moran, J., & Swenson Jr, G. 2001

Thompson, A., Taylor, G., Carilli, C., & Perley, R. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, 11–36

Thompson, A. R. 1999, in Synthesis Imaging in Radio Astronomy II, Vol. 180, 11

Thompson, A. R., Clark, B., Wade, C., & Napier, P. J. 1980, The Astrophysical Journal Supplement Series, 44, 151

Voronkov, M., & Wieringa, M. 2004, Experimental Astronomy, 18, 13

Werbos, P. J. 1990, Proceedings of the IEEE, 78, 1550