Rhodes University

Data Compression, Field of Interest Shaping and Fast Algorithms for Direction-dependent Deconvolution in Radio Interferometry

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

Doctor of Philosophy

in the

Faculty of Science

by

Marcellin T. Atemkeng

GRAHAMSTOWN, SOUTH AFRICA

I have prepared my thesis specifically for the degree of Doctor of Philosophy, while under supervision in the Departement of Physics and Electronics, Rhodes University.

Supervisor:

Prof Oleg Smirnov (SKA SA and Rhodes University, South Africa)

Co-supervisors:

Dr Cyril Tasse (Observatoire de Paris, France)

Dr Griffin Foster (SKA SA Cape Town, South Africa)

Statement of Authenticity

I, the undersigned declare that the contents of this thesis represent my own work, and that the thesis has not previously been submitted for a qualification at any other university.

Marcellin T. Atemkeng

Publications

Publications directly resulting from the research presented in this thesis:

- M. T. Atemkeng, O. M. Smirnov, C. Tasse, G. Foster and J. Jonas, *Using baseline-dependent window functions for data compression and field-of-interest shaping in radio interferometry.* The Monthly Notices of the Royal Astronomical Society, volume 462, page 2542, 2016.
- M. T. Atemkeng, O. M. Smirnov, C. Tasse, G. Foster and J. Jonas, *Baseline-dependent averaging and baseline-dependent window functions for equal uv-distance data compression and field-of-interest shaping in radio interferometry.* The Monthly Notices of the Royal Astronomical Society (MNRAS), in preparation.
- M. T. Atemkeng, O. M. Smirnov, C. Tasse, G. Foster and J. Jonas, *Towards a position-dependent deconvolution scheme for radio interferometric imaging.* Astronomy & Astrophysics (A&A), in preparation.

Conference paper and technical reports resulting from my collaboration with the research outputs from this thesis:

- P. Zarka, M. Tagger, L. Denis, J. N. Girard, A. Konovalenko, et *al.*, *NenuFAR: instrument description and science case*, IEEE International Conference on Antenna Theory and Techniques (ICATT), Kharkiv, Ukraine, 2015.
- C. Tasse, M. Atemkeng, B. Cecconi, J. N. Girard, C. Dumez-Viou, O. Smirnov, P. Zarka, P. Larzabal, M. N. El Korso, R. Boyer, *NenuFAR in standalone mode: calibration, imaging, confusion, and decorrelation issues*; University of Paris Diderot, Paris Observatory, Volume 5, pp. 38-41, 10/10/2014.
- C. Tasse, B. Hugo, O. Smirnov, M. Atemkeng, E. Bester, E. Bonnassieux, J. Girard, S. Makhathini and S. Perkins, *Facetting for direction-dependent spectral deconvolution*, Astronomy & Astrophysics (A&A), in preparation.

Abstract

In radio interferometry, observed visibilities are intrinsically sampled at some interval in time and frequency. Modern interferometers are capable of producing data at very high time and frequency resolution; practical limits on storage and computation costs require that some form of data compression be imposed. The traditional form of compression is simple averaging of the visibilities over coarser time and frequency bins. This has an undesired side effect: the resulting averaged visibilities "decorrelate", and do so differently depending on the baseline length and averaging interval. This translates into a non-trivial signature in the image domain known as "smearing", which manifests itself as an attenuation in amplitude towards off-centre sources. With the increasing fields of view and/or longer baselines employed in modern and future instruments, the trade-off between data rate and smearing becomes increasingly unfavourable. Averaging also results in baseline length and a position-dependent point spread function (PSF). In this work, we investigate alternative approaches to low-loss data compression. We show that averaging of the visibility data can be understood as a form of convolution by a boxcar-like window function, and that by employing alternative baseline-dependent window functions a more optimal interferometer smearing response may be induced. Specifically, we can improve amplitude response over a chosen field of interest and attenuate sources outside the field of interest. The main cost of this technique is a reduction in nominal sensitivity; we investigate the smearing vs. sensitivity trade-off and show that in certain regimes a favourable compromise can be achieved. We show the application of this technique to simulated data from the Jansky Very Large Array and the European Very Long Baseline Interferometry Network. Furthermore, we show that the position-dependent PSF shape induced by averaging can be approximated using linear algebraic properties to effectively reduce the computational complexity for evaluating the PSF at each sky position. We conclude by implementing a position-dependent PSF deconvolution in an imaging and deconvolution framework. Using the Low-Frequency Array radio interferometer, we show that deconvolution with position-dependent PSFs results in higher image fidelity compared to a simple CLEAN algorithm and its derivatives.

Acknowledgements

A thesis? What a program! The scope becomes clear when one pays attention to those who contributed. I would like to express my deepest gratitude to my supervisors, Prof Oleg Smirnov, Dr Cyril Tasse and Dr Griffin Foster. It has been a privilege to work with them; their kindness, patience and devotion shaped me. It was an honour to work with them, especially in view of their human qualities and scientific competence. I appreciate all their contributions of time, ideas and funding to make my Ph.D. experience productive and stimulating. I would also like to thank our laboratory director, Prof Justin Jonas. I have benefited and learned enormously from this challenging research topic you once suggested to Prof Oleg Smirnov. It was really a relevant topic and one of the essentials for the success of the future Square Kilometre Array (SKA).

The European Very Long Baseline Interferometry Network (EVN) related research discussed in this thesis would not have been possible without having crossed paths with the Joint Institute for VLBI ERIC (JIVE) and the Netherlands Institute for Radio Astronomy (ASTRON). I appreciate their collaboration, particularly that of Dr Tammo Jan Dijkema, Dr Zsolt Paragi and Dr Aard Keimpema. I gratefully acknowledge the funding from the MIDPREP EU FP7 program that made my visits to ASTRON/JIVE possible.

I would like to express my special thanks of gratitude to Dr Tony Willis for having carefully read my thesis and suggested so many wonderful comments.

During the period of my studies, I benefited from funding from Rhodes University in the form of travel grants to conferences. Thank you, Rhodes! You are indeed *A Small University with a Big Heart.* I am and will always be a proud Rhodent. I am also grateful to the National Research Foundation of South Africa for funding this research.

To my colleagues, the members of the Centre for Radio Astronomy Techniques and Technologies, Rhodes University, and the members of the Radio Astronomy Research Group, SKA, Cape Town, South Africa; particularly that of Dr Kshitij Thorat, Dr Modhurita Mitra and Ridhima Nunhokee. I would also like to extend my deep gratitude for your brilliant presentations during the group meetings; I learned so much.

I would like to extend my deepest gratitude to our project administrator, Ronel Groenewald. She helped throughout this work by arranging accommodation, conferences and travel plans. She deserves gratitude for the invaluable contribution she has made. I would also like to thank the staff of the Department of Physics and Electronics, Rhodes University, for their support in various ways, in particular Prof Makaiko Chithambo, the Head of Department.

I would like to thank my uncles, Albert Atemkeng, Mekoudja Guimfac Rene and Norbert Atemkeng. Thank you for supporting me in everything; especially I cannot thank you enough for encouraging me throughout all my studies; you have been always present to support me in any way, from the financial to the social aspects. Thanks to Mrs Emmanuella Gwanmesia Atemkeng for all the attention, love and support you provided to me during my visit to the USA.

Lastly, I would like to thank my family for all their love and encouragement, especially my father, Robert Teufack, who raised me with a love of mathematics and supported me in all my pursuits.

Contents

Abbreviations

BDA	**B**aseline-**d**ependent **A**veraging
BDWF	**B**aseline-**d**ependent **W**indow **F**unctions
B	**B**lackman
DSP	**D**igital **S**ignal **P**rocessing
DDFacet	**D**irection-**d**ependent **F**acet
EVN	**E**uropean **V**LBI **N**etwork
FSCN	**F**ar **S**idelobe **C**onfusion **N**oise
FFT	**F**ast **F**ourier **T**ransform
FoI	**F**ield of **I**nterest
FoV	**F**ield of **V**iew
FAST	**F**ive **h**undred meter **A**perture **S**pherical **T**elescope
FT	**F**ourier **T**ransform
HM	**H**amming
HN	**H**an
IPR	**I**mage **P**lane **R**esponse
JVLA	**J**ansky **V**ery **L**arge **A**rray
JIVE	**J**oint **I**nstitute for **V**LBI in **E**urope
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
PB	**P**rimary **B**eam
PS0	**P**rolate **S**pheroidal wave function
RIME	**R**atio **I**nterferometer **M**easurement **E**quation
SLR	**S**idelobes **R**oll-off

SA	**S**outh **A**frica
SKA	**S**quare **K**ilometre **A**rray
VLBI	**V**ery **L**ong **B**aseline **I**nterferometry
WF	**W**indow **F**unctions

Symbols

$\mathcal{I}_{pq}$	apparent sky seen by baseline pq	Δt	integration time
$\mathcal{P}^{(dis)}$	approximated pseudo-PSF	$B_{\max}$	longest baseline length
$V_{pqkl}^{(\mathrm{dis})}$	approximated resampled visibility	ΔS	minimum detectable flux
$\mathcal{A}_p$ or $\mathcal{A}_q$	aperture illumination function	$\Xi_{X\mu}$	noise penalty
$\mathcal{V}_{pq}^{\mathrm{M}}$	averaged visibility value	$\sigma_{X\mu}$	noise for BDWF
D	antenna diameter	σ_{m}	noise on the simple averaging
$\mathcal{E}_p$, $\mathcal{E}_q$	antenna effective collecting area	σ_{pix}	noise in each pixel
p, q	antenna indexes	σ_{s}	noise across visibilities
θ	angular resolution	$\mathcal{V}^{\mathrm{nom}}$	nominal observed visibility
u, v, w	baseline coordinates	N_{pol}	number polarisations
$\mathsf{B}_{kl}^{[\Delta_{pq}t,\Delta_{pq}\nu]}$	baseline-dependent resampling bin	N_t	number of timeslots
X	baseline-dependent window function	N_{sub}	number of sub-bands
$\Delta_{pq}\nu$	baseline-dependent frequency sampling interval	n_a	number of antennas
$\mathcal{X}$	baseline-dependent window function	k	number of discrete points
$\Delta_{pq}t$	baseline-dependent time sampling interval	N_{pix}	number of pixels
B	baseline length	N_{pq}	number of baselines
$\mathbf{u}_{pq}^{\mathrm{m}}$	baseline vector in metre	N_{src}	number of sources
$\boldsymbol{u}$	baseline vector en wavelength	N_{facet}	number of facets
J_0^{Hm}	Bessel-Hamming window	$N_{\mathrm{pix,facet}}$	number of pixels per facet
J_0	Bessel window	N_t^{hires}	number high resolution points
$\mathsf{B}_{kl}^{[\Delta_{pq}t,\Delta_{pq}\nu]}$	bin indices of the resampling bin for pq	$\mathrm{Dim}\{\mathsf{B}_{kl}^{[\Delta t\Delta\nu]}\}$	number points in $\mathsf{B}_{kl}^{[\Delta t\Delta\nu]}$
k_B	Boltzmann constant	d_o	observed signal
$\Pi^{[k]}$ or Π	boxcar window	$\mathcal{R}$	phase gradient complex scalar
$\Pi^{[t\nu]}$	boxcar window in $t\nu$-space	d_r	physical signal
$\Pi_{pqkl}^{[uv]}$	boxcar window in uv-space	S	point source amplitude

BW	Butterworth window	$\mathcal{P}$	point spread function
$\Delta\nu$	channel width	PS_0	prolate spheroidal window
CF	compression factor	Δx	regular grid spacing
$\mathcal{K}$	cumulative distortion effect for all the visibilities	d_s	sampled signal
$\mathcal{V}_{pqkl}^{\mathrm{degrid}}$	degridded visibilities	$V_{pq}^{(\mathrm{s})}$	sampled visibilities
s	Dirac Comb	$\mathsf{B}_{kl}^{[\Delta t \Delta \nu]}$	sampling bin in $t\nu$-space
δ	Dirac delta function	$\mathsf{B}_{pqkl}^{[uv]}$	sampling bin in uv-space
l, m, n	direction cosines	V	single visibility value
$\boldsymbol{G}_p$, $\boldsymbol{G}_q$	direction-dependent effects	V_{pq}	single visibility value for pq
$\boldsymbol{E}_p$,$\boldsymbol{E}_q$	direction-independent effects	$\mathrm{sinc}^{\mathrm{Hm}}$	Sinc-Hamming window
$\mathcal{I}^{\mathrm{D}}$	dirty image	$\mathcal{I}_0$	sky distribution
$d_{s,k}$	discrete value	c	speed of light
$\mathcal{K}_{pqkl}$	distortion effect for a single visibility	T_{sys}	temperature
$\mathcal{F}$,$\mathcal{F}^{-1}$	Fourier transform and its inverse	t	time
ν	frequency	T_{obs}	total time for observation
$\mathcal{G}$	fringe induced by the w-term	B_{w}	total bandwidth
G	Gaussian window	$\mathcal{D}$	uv distances en metre
$\mathcal{V}_{pqkl}^{\mathrm{grid}}$	gridded visibilities	$\mathcal{V}_{pq}$or $\mathcal{V}$	visibility for baseline pq
$\Delta\Phi_{pq}$	half of the phase difference in frequency	λ	wavelength
$\Delta\Psi_{pq}$	half of the phase difference in time	$\mathcal{W}$	weighted-sampling function
$\mathcal{I}^{\mathrm{DI}}$	ideal dirty image	W	weight for a single data point
$\mathcal{T}_{pq}$	image plane response function	$\mathcal{W}_{pq}$	weighting function for baseline pq

In memory of my mother, maman Josephine Atemkeng Dongmo. Mom, as I walk through the journey of life, I remember how you helped me to grow, with love, truth, and honesty. Though it's been more than one and half decades since you were taken away, the memories are still strong and I wish you were here today.

Chapter 1

General Introduction

We all do optimisation. In our daily life, we seek to optimise our work time, our storage space, or the way we have to go to get us somewhere, etc. We all seek a better solution to problems that mark our existence. Generally, the optimisation is therefore intended to find the best solution: this is what the present work is all about. The research presented in this thesis focuses on the compression of measurements made by an assembly of radio antennas, connected in networks of observing elements, referred to as a *radio interferometer*. *Radio interferometry* is the science of studying such networks. Understanding the challenges related to signal compression in radio interferometry, where the signal is dominated by uncorrelated Gaussian noise from each station (radio antenna), requires that we understand some concepts. First, how the signal from each station is measured. Second, how the signal from different stations are correlated. Third, the resulting data structure, and its size and dimension.

Accordingly, one purpose of this introduction is to explain these concepts as simply as possible, with the aid of mathematically measurable quantities wherever a quantification is possible. I shall explain the necessity of compressing radio interferometric data, and the obstacles in handling volumes of data. I shall describe, as well, the data volume for some of the current radio interferometers and show how this volume will grow significantly for the future generation of radio interferometers, which will be used to image wide fields of view (FoVs) at high spectral and temporal resolution.

1.1 A brief history of radio astronomy

Although astronomy is one of the oldest sciences, the study of astronomical observations at radio frequencies started with the accidental discovery of radio waves emanating from the Milky Way [Akhmanw and Khokblov, 1959]. A radio wave is a type of electromagnetic radiation with frequencies in the ranges of 10^3 Hz to 3×10^{11} Hz [Brown, 1984]. In late 1936 Karl Jansky, who was working on long-distance radio communication, designed a rotating antenna receiver at a frequency of 20.5 MHz and discovered a signal (radio wave) coming from the centre of the Milky Way [Akhmanw and Khokblov, 1959]. After his discovery, it was only at the end of the Second World War that radio astronomy began as a consequence of the development of radar and the cheap availability of radar equipment [Arnold, 2014, Hey and Hey, 1973]. The promising results from these instruments provided answers to a number of old questions regarding the Milky Way structure. To observe the structure clearly, scientists began to build antenna receivers with a larger collecting area, capable of detecting radio waves at longer wavelengths. Besides the structure of the Milky Way, radio astronomy has played a significant role in our understanding of stars' formation, interaction between different celestial objects and the various epochs in the evolution of the universe [Koopmans et al., 2015].

By now, very big radio telescopes have been built, such as the 305-meter diameter Arecibo radio telescope [Goldsmith, 1996] and the 500-meter diameter FAST[1] radio telescope currently under construction in China [Nan, 2006, Nan et al., 2011]. These big telescopes have had their own limitations. Firstly, because of their size, it is impractical to build bigger dishes. Secondly, they are incapable of achieving high angular resolution.

The collecting area of the antenna dish is proportional to the amount of signal that can be detected; the larger the surface area, the weaker the detectable radio signal. The dish diameter thus drives the attainable flux sensitivity. Besides flux measurement, it is also useful to obtain a higher resolution map of astronomical objects. Because of the limitations of single-dish observation, techniques such as radio interferometry become imperative. A radio interferometer is formed by pairwise-correlating signals from a number of individual radio telescopes (either in real time, or off-line, using some sort of signal recording scheme). The separation of the antennas ultimately determines the maximum achievable resolution. At the same time, maintaining a coherent interferometer over array

[1] Five-Hundred-Meter Aperture Spherical Telescope

sizes in excess of tens of kilometers becomes increasingly challenging. The latter problem lies in the domain of techniques such as Very Long Baseline Interferometry (VLBI) [Felli and Spencer, 1989, 2012, Takahashi, 2000, Walker, 1999]. The aim of a radio interferometer is to detect astronomical radio sources and provide estimates of their properties, which include the source position, flux density among others. Radio interferometers such as the Jansky Very Large Array (JVLA) [Napier et al., 1983, Thompson et al., 1980] and the Low-Frequency Array (LOFAR) [van Haarlem et al., 2013a] use these techniques.

1.2 Limits of single-dish radio observations

For a single antenna with effective collecting area $\mathcal{E}_p$, and system temperature $T_{\rm sys}$, the minimum detectable flux ΔS, is directly proportional to $T_{\rm sys}/\mathcal{E}_p$, as:

$$\Delta S = \frac{2k_{\rm B}T_{\rm sys}}{\mathcal{E}_p\sqrt{\Delta\nu\Delta t}}, \tag{1.1}$$

where $k_{\rm B}$ is the Boltzmann constant [Johnson, 1927, 1928], Δt stands for the integration time and $\Delta\nu$ for the bandwidth. From the above equation, it can be seen that the ratio $T_{\rm sys}/\mathcal{E}_p$ is of critical importance for any single antenna: the larger the collecting area $\mathcal{E}_p$, the fainter ΔS. Also, the larger the bandwidth and/or the longer the integration time, ΔS decreases; we will come back to this later throughout this dissertation.

Coming now to the angular resolution θ, attainable by a single-dish telescope with diameter D, receiving a radio signal at a wavelength λ, we have the scaling:

$$\theta \sim \lambda/D. \tag{1.2}$$

For example, the large single-dish radio telescope of diameter 500 meters under construction in China will achieve a resolution of only ~ 2.9 arcmin at L-band, i.e. 1 GHz to 2 GHz [Nan et al., 2011]. The resolution and sensitivity limitation of single dishes pushed scientists to find novel techniques, hence the nascence of interferometry and aperture synthesis.

1.3 Interferometry and aperture synthesis

The concept of aperture synthesis developed by Martin Ryle [Ryle and Hewish, 1960] consists of synthesising a larger radio telescope from a set of smaller antennae. To arrive at this, the following procedure is applied:

(a) Correlate the voltages from two telescopes for a given length of time i.e. integration time. This gives one sampled point or data for one timeslot.

(b) For an interferometer array of n_a antennas, the number of samples for the given integration time will be $n_a(n_a-1)/2$, which is the instantaneous number of correlations during the integration time.

(c) Use the Earth's rotation to measure more samples with different orientations. The rotation of the Earth plays an important role in measuring data for different timeslots.

The most efficient way to fill a synthesised aperture is to use many small antenna telescopes and measure samples at shorter integration times over long periods, while making use of the Earth's rotation and frequency coverage and array layout (Figure 1.1 demonstrates this aspect).

The angular resolution achievable by the synthesised telescope is approximately $\lambda/B_{\max}$, where $B_{\max}$ is the longest baseline distance (a baseline is the distance between a pair of antennas). We note here that for compact interferometers observing at longer wavelengths, the constraint of not achieving better angular resolution (e.g. on the order of milliarcsecs) remains unchanged. This important aspect is worth noting from Table 1.1. In other words, *any observation at long wavelength (low frequency) requires an interferometer of long baselines to achieve an acceptable angular resolution.* A typical example is the LOFAR radio interferometer where the baselines are spread across a region of 400 km in diameter, observing at longer wavelengths in the range 1.2 meter−30 meter [Röttgering, 2003, Röttgering et al., 2006].

1.4 Methods of measurement of radio interferometric data

An interferometer array measures the quantity $\mathcal{V}(u, v, w)$, known as the visibility function. Here, the coordinates u, v and w are vector components in units of wavelength,

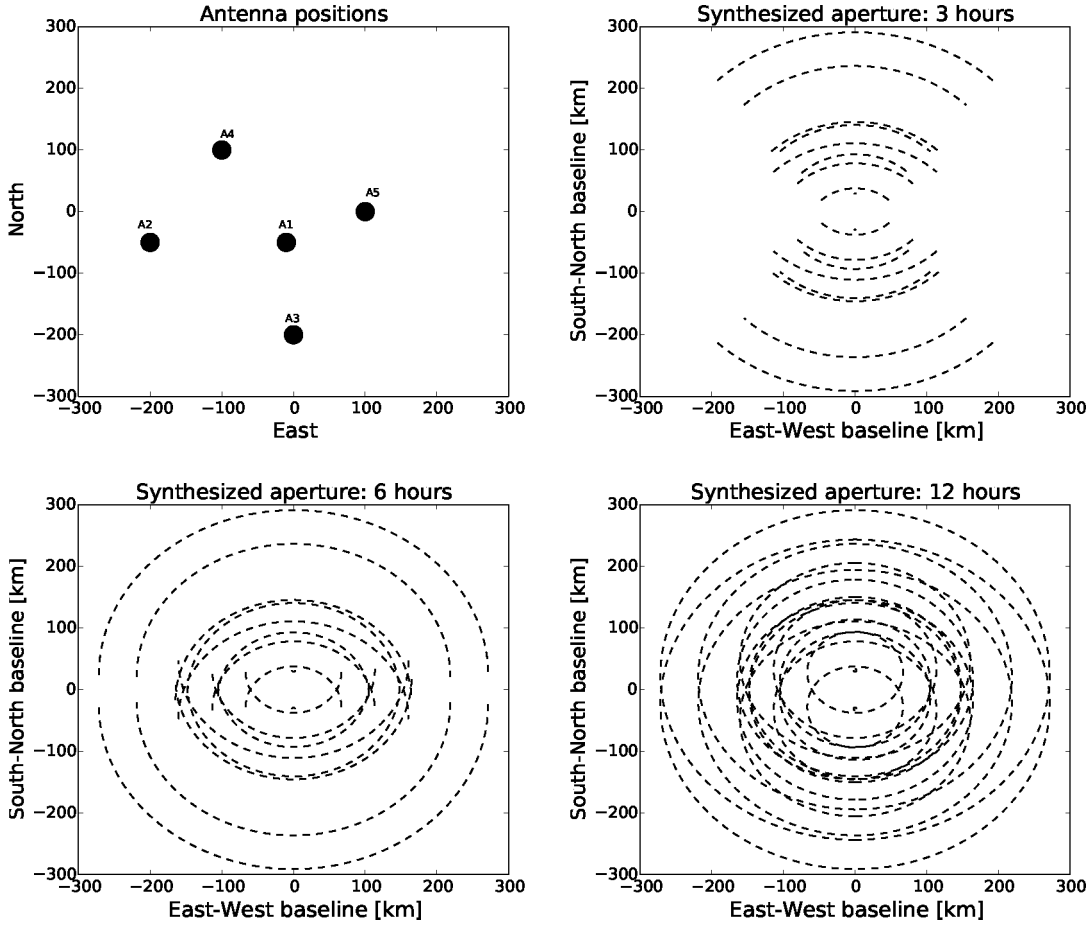

FIGURE 1.1: Antennas' position and the synthesised apertures. (Top left): With an array of five antennas and 10 correlations are produced at each integration time. In this case the array is pointing at the zenith. Each line of the synthesised aperture is the track drawn by the array baselines during the Earth's rotation. The sequence from the inner to the outer tracks are from short to long baselines. (Top right): Three hours synthesised aperture. (Bottom left): Six hours synthesised aperture. (Bottom right): Twelve hours synthesised aperture.

	observed frequency, ν			
$B_{\max}$	**100 MHz**	**1.4 GHz**	**10 GHz**	**100 GHz**
100 m	6188	442	62	6
8 km	77	6	0.8	0.08
36 km	17	1.22	0.17	0.017
1000 km	0.61	0.044	0.0061	0.00061

TABLE 1.1: Maximum angular resolution (in arcsec) achievable by various baselines lengths vs. frequency.

describing the distance between antennas p and q, called the *baseline*. The w axis is oriented towards the *phase centre* of the observed field, while u points East and v North. Given a sky distribution $\mathcal{I}_0(l, m)$, where l, m are the direction cosines, the nominal observed visibility is given by the van Cittert-Zernike theorem [Thompson, 1999, Thompson

et al., 2001] as:

$$\mathcal{V}^{\text{nom}}(u,v) = \iint_{lm} \frac{\mathcal{I}_0(l,m)}{\sqrt{1-l^2-m^2}} \, \mathrm{e}^{-2\pi \mathrm{i}\phi(u,v,w)} \mathrm{d}l\mathrm{d}m, \tag{1.3}$$

where $\phi(u,v,w) = ul+vm+w(n-1)$, and $n = \sqrt{1-l^2-m^2}$ (the $n-1$ term comes about when fringe stopping is in effect, i.e. when the correlator introduces a compensating delay to ensure $\phi = 0$ at the centre of the field, otherwise the term is simply n).

Given a pair of antennas p and q forming a baseline $\boldsymbol{u}_{pq} = (u_{pq}, v_{pq}, w_{pq})$, and taking into account the *primary beam* patterns $\mathcal{E}_p(l,m)$ and $\mathcal{E}_q(l,m)$ that define the directional sensitivity of the antennas as presented in Eq. (1.1), this becomes:

$$\mathcal{V}_{pq}(u,v) = \iint_{lm} \frac{\mathcal{E}_p \mathcal{I}_0 \mathcal{E}_q^H}{\sqrt{1-l^2-m^2}} \, \mathrm{e}^{-2\pi \mathrm{i}\phi(u,v,w)} \mathrm{d}l\mathrm{d}m, \tag{1.4}$$

where ^{H} represents the complex transpose. The first term being integrated is the *apparent sky seen by baseline pq*

$$\mathcal{I}_{pq} = \frac{\mathcal{E}_p \mathcal{I}_0 \mathcal{E}_q^H}{\sqrt{1-l^2-m^2}}, \tag{1.5}$$

which in general can be variable in time and frequency. For simplicity, let us assume that both the sky and the primary beam are constant, and that the primary beam is the same for all stations. All baselines will then see the same apparent sky throughout the measurement process. Let us designate this by $\mathcal{I}$, the apparent sky. Assuming a small FoV ($n \to 1$) and/or a co-planar array ($w = 0$), the above equation becomes a simple 2-D Fourier transform:

$$\mathcal{V}(u,v) = \iint_{lm} \mathcal{I} \, \mathrm{e}^{-2\pi \mathrm{i}(ul+vm)} \mathrm{d}l\mathrm{d}m, \tag{1.6}$$

or in functional form,

$$\mathcal{V} = \mathcal{F}\{\mathcal{I}\} \quad \text{and} \quad \mathcal{I} = \mathcal{F}^{-1}\{\mathcal{V}\}. \tag{1.7}$$

Here $\mathcal{F}$ is the Fourier transform and $\mathcal{F}^{-1}$ its inverse.

We will refer to $\mathcal{V}$ as the *ideal* visibility distribution; as opposed to the *measured* distribution, which is corrupted by averaging in the correlator, as we will explore in Chapter 3.

The effect of the primary beam can alternatively be expressed in terms of a convolution

with its Fourier transform, the *aperture illumination function* $\mathcal{A}_p(u,v)$. In functional form:

$$\mathcal{V}_{pq} = \mathcal{A}_p \circ \mathcal{V}_{pq}^{\text{nom}} \circ \mathcal{A}_q^H, \tag{1.8}$$

where $\mathcal{A}_p = \mathcal{F}\{\mathcal{E}_p\}$.

Eq. (1.6) can be also presented in the radio interferometric measurement equation formalism (RIME). For details on the RIME formalism see Hamaker et al. [1996], Smirnov [2011a,b]. The RIME describes in a compact and robust way all the effects (direction-dependent and direction-independent [Smirnov, 2011b]) that may occur when an interferometric measurement is in process. The 2-D Fourier transform full sky RIME, according to Smirnov [2011a,b], is given by:

$$\mathcal{V}_{pq} = \boldsymbol{G}_p\Big[\iint_{lm} \boldsymbol{E}_p\mathcal{I}\boldsymbol{E}_q^H \mathrm{e}^{-2\pi\mathrm{i}(ul+vm)}\mathrm{d}l\mathrm{d}m\Big]\boldsymbol{G}_q^H. \tag{1.9}$$

The formalism groups all the direction-independent effects of an antenna p into the matrix $\boldsymbol{G}_p$, and all its direction-dependent effects into the matrix $\boldsymbol{E}_p$. Note that the *primary beam* pattern, $\mathcal{E}_p$, is part of the direction-dependent effects.

1.5 Structure of radio interferometric data

I present a schematic description of how an interferometer array's overall data are mapped into a correlation matrix. Note that this mapping is an abstraction of how the correlation is saved into computer memory. I shall come back to this in Chapter 4. In level $L1$ of Figure 1.2, the vertical axis represents the data for baselines, and the horizontal axis are sub-bands (spectral windows). Each sub-band on its own is a sub-dataset projected in level $L2$, where the vertical axis is the number of integration times (timeslots) and the horizontal axis is the number of channels. Each cross-section (level $L3$) of times and channels is the visibility presented in Eq. (1.9). In general, Eq. (1.9) represents a complex scalar or 2×2 complex visibility. In terms of data storage, we restrict ourselves to 32 bits (complex data) per polarisation. If N_{pol} is the number polarisation then we have $32 \times N_{\text{pol}}$ bits per visibility. However, an interferometer of $n_a(n_a-1)/2$ baselines, N_{sub} sub-bands per baseline observing during a total time of T_{obs}, with integration time

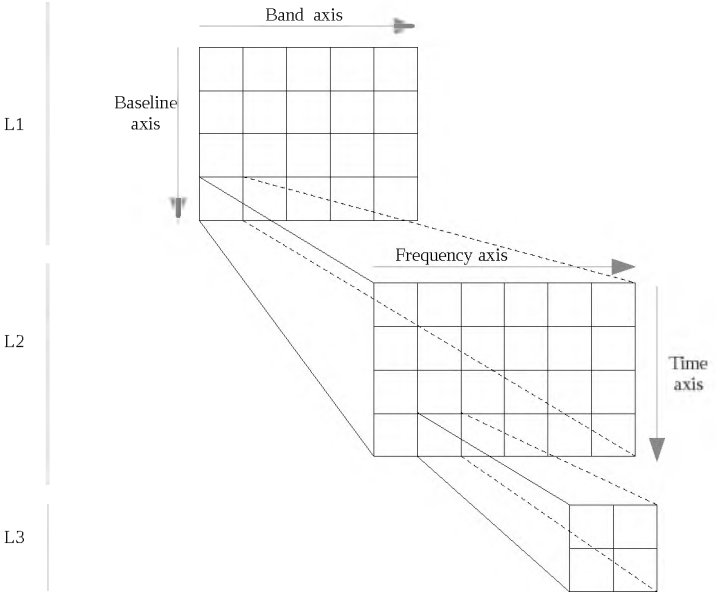

FIGURE 1.2: Interferometric data structure. A baseline is a set of sub-bands, each sub-band on it own is a set of data for timeslots and channels. The data shape for a baseline results in ($N_{\rm sub}$, N_t, N_ν, $N_{\rm pol}$).

Δt, and total bandwidth of $B_{\rm w}$, having $\Delta\nu$ of channel width, will record in memory

$$\text{data size} = \frac{n_a(n_a-1)}{2} \times N_{\rm sub} \times \frac{T_{\rm obs}}{\Delta t} \times \frac{B_{\rm w}}{\Delta\nu} \times N_{\rm pol} \times 32\ bits, \tag{1.10}$$

where, $T_{\rm obs}$ and Δt are in seconds, $B_{\rm w}$ and $\Delta\nu$ are measured in Hz. One can rewrite Eq. (1.10) as:

$$\text{data size} = \frac{n_a(n_a-1)}{2} \times N_{\rm sub} \times N_t \times N_\nu \times N_{\rm pol} \times 32\ bits, \tag{1.11}$$

where $N_t = T_{\rm obs}/\Delta t$ and $N_\nu = B_{\rm w}/\Delta\nu$ are the number of timeslots and channels respectively. Observing for long periods and over large bandwidths leads to storage issues for a big interferometer array, as well as substantial computational load for sub-processing. This is because $\{T_{\rm obs}, B_{\rm w}\} \propto \{N_t, N_\nu\}$ if Δt and $\Delta\nu$ must remain sufficiently small. Table 1.2 is obtained from Eq. (1.11) and shows clearly that any increase in the number of antennas, sub-bands, timeslots and channels results in increasing the data size and this poses significant computational challenges for processing. The case of 1000 antennas is close to what is expected from the Square Kilometre Array (SKA) [Krichbaum et al., 1999, Wright, 2002]. With these observing parameters, approximately 5709 TB data have to be recorded, which is already above the capacity of any modern hard disk [Cloudsim,

n_a	N_{sub}	T_{obs}[hour]	Δt[s]	B_{w}[MHz]	$\Delta\nu$[Hz]	$\sim$ **Memory**[Gb]
7	5	16	20	10^2	10^6	**0.4**
27	3	4	5	10^2	125000	**60**
64	5	4	10	10^2	10^6	**22**
1000	3	4	0.5	10^2	11000	**5846240**

TABLE 1.2: Interferometers observing parameters vs. data size

2015]. Since the number of baselines and sub-bands will remain constant for individual interferometers, the only possible way to reduce the data size is by increasing integration time and channel widths. Unfortunately, longer integration time and wider channel widths pose several challenges, as we will see in the section below.

1.6 Problem statement

A radio interferometer measures complex quantities called *visibilities*, which, following the van Cittert-Zernike relation [Thompson, 1999, Thompson et al., 2001], correspond to Fourier modes of the sky brightness distribution, corrupted by various instrumental and atmospheric effects. One particular effect, known as *time* and *bandwidth decorrelation* or its equivalent in the image plane referred to as *smearing*, occurs when the visibilities are averaged over a time and frequency bin of non-zero extent [Bridle and Schwab, 1999, 1989]. This unavoidably happens in the correlator (since the correlator output is, by definition, an average measurement over some interval), and also if data are further averaged in post-correlation for the purposes of compression and to reduce computational cost.

The effect of smearing is mainly a decrease in the amplitude of off-axis sources. This is easy to understand: the visibility contribution of a point source of flux S located in the direction given by the unit vector $\boldsymbol{\sigma}$ is given by:

$$V = S \exp\{\frac{2\pi i}{\lambda} \boldsymbol{u} \cdot (\boldsymbol{\sigma} - \boldsymbol{\sigma}_0)\}, \tag{1.12}$$

where $\boldsymbol{u}$ is the baseline vector, and $\boldsymbol{\sigma}_0$ is the phase centre (or fringe stopping centre) of the observation. The complex phase term above rotates as a function of frequency (due to the inverse scaling with λ) and time (due to the fact that $\boldsymbol{u}$ changes over time, at least in an Earth or orbit-based interferometer). Taking a vector average over a time/frequency

bin then results in a net loss of amplitude. The effect increases with baseline length and distance from the phase centre. Besides reducing apparent source flux, smearing also distorts the PSF, since different baselines (and thus different Fourier modes) are attenuated differently. The issue of time-frequency averaging has been addressed in the past by among others Thompson et al. [2008], where a Gaussian taper has been used to eliminate smearing at the edges of the FoV. However, the problem of eliminating smearing to about 5% or less within the FoV while compressing the data to an acceptable level has not been satisfactorily addressed before. On the other hand, smearing can be seen to be a useful side effect, since anything outside the desired FoV by definition is unwanted signal. The primary beam pattern of any real antenna features sidelobes and backlobes that extend across the entire sky, albeit at a relatively faint level. The faintness makes sidelobes useless for imaging any but the brightest sources. However, the sum total signal from all the sources in the primary beam sidelobes, modulated by their PSF sidelobes, contributes an unwanted global background called the *far sidelobe confusion noise* (FSCN). In very deep observations this may in principle become a bottleneck [Smirnov, O. M. and Frank, B. and Theron, I. P. and Heywood, I., 2012]. In other cases, individual extremely bright radio sources such as Cygnus A [Baars and Hartsuijker, 1972, Wilson et al., 2006] or Cassiopeia A [Baars and Hartsuijker, 1972] can contribute confusing signals from even the most distant sidelobes; the LOFAR telescope [van Haarlem et al., 2013b] has to deal with these so-called "A-team" sources on a routine basis. By suppressing distant off-axis sources, smearing somewhat alleviates both the FSCN and A-team problems.

Figures 1.3 and 1.4 are produced by simulating a series of high time-frequency resolution observations using MeqTrees [Noordam and Smirnov, 2010] and applying averaging. Figure 1.3 shows the attenuation of a 1 Jy source as a function of distance from the phase centre, for a set of different time and frequency intervals, while Figure 1.4 shows the distortion of the PSF for a source at four different sky locations (0°, 1°, 2°, 3°), for 100 s integration time and 5 MHz channel width. The simulations correspond to JVLA in the C configuration (0.035 km and 3.4 km for the minimum and maximum antenna separation respectively), with an observing frequency of 1.4 GHz. At this frequency, the first null of the primary beam is at $r \approx 36'$, and the half-power point is at $\sim 16'$, thus we can consider the "conventional" FoV (i.e. the half-power beam width) to be about 0.5° across. Note that the sensitivity of the upgraded JVLA, as well as improvements in calibration techniques [Perley, R., 2013], allow imaging to be done in the first primary

beam sidelobe as well (and in fact it may be necessary for deep pointings, if only to deconvolve and subtract sidelobe sources), so we could also consider an "extended" FoV out to the second null of the primary beam at $r \approx 1.25°$. Whatever definition of the FoV we adopt:

Figure 1.3 shows that to keep amplitude losses across the FoV to within some acceptable threshold, say 1%, the averaging interval cannot exceed some critical size, say 10 s and 1 MHz. Conversely, if we were to adopt an aggressive averaging strategy for the purposes of data compression, say 100 s and 5 MHz, the curves indicate that we would suffer substantial amplitude loss towards the edge of the FoV. Also, the PSF would change its shape significantly and differently at each sky position (Figure 1.4). Any imaging framework such as the CASA imager [Clark, 1980] and its derivatives [Cornwell, 2008, Offringa et al., 2014] will suffer substantially from this PSF distorted shape when trying to remove the PSF from the source during deconvolution.

The curves corresponding to acceptably low values of smearing across the FoV (i.e. up to 25 s and up to 1.25 MHz) have a very gentle slope, with very little suppression of sources *outside* the FoV.

1.7 Motivations and contributions

With the growth of interferometric arrays in the number of elements, the number of frequency channels and long observations etc, where computation (and thus data size) becomes one of the main cost drivers, it is in principle desirable to average the data down as much as possible, without compromising the science goals. There are natural limits to this: firstly, we still need to sample the uv-plane critically; secondly, we need to retain sufficient spectral resolution; thirdly, we do not want to average (at least pre-calibration) beyond the natural variation of the calibration parameters, and fourthly, we want to keep smearing at acceptable levels in order not to lose too much signal. In this work, we concentrate specifically on the decorrelation/smearing problem. Typical observation cases are:

(a) In a compact interferometer, where the maximum usable FoV corresponds to the primary beam of the antennas. In most cases (but surveys especially) we want the

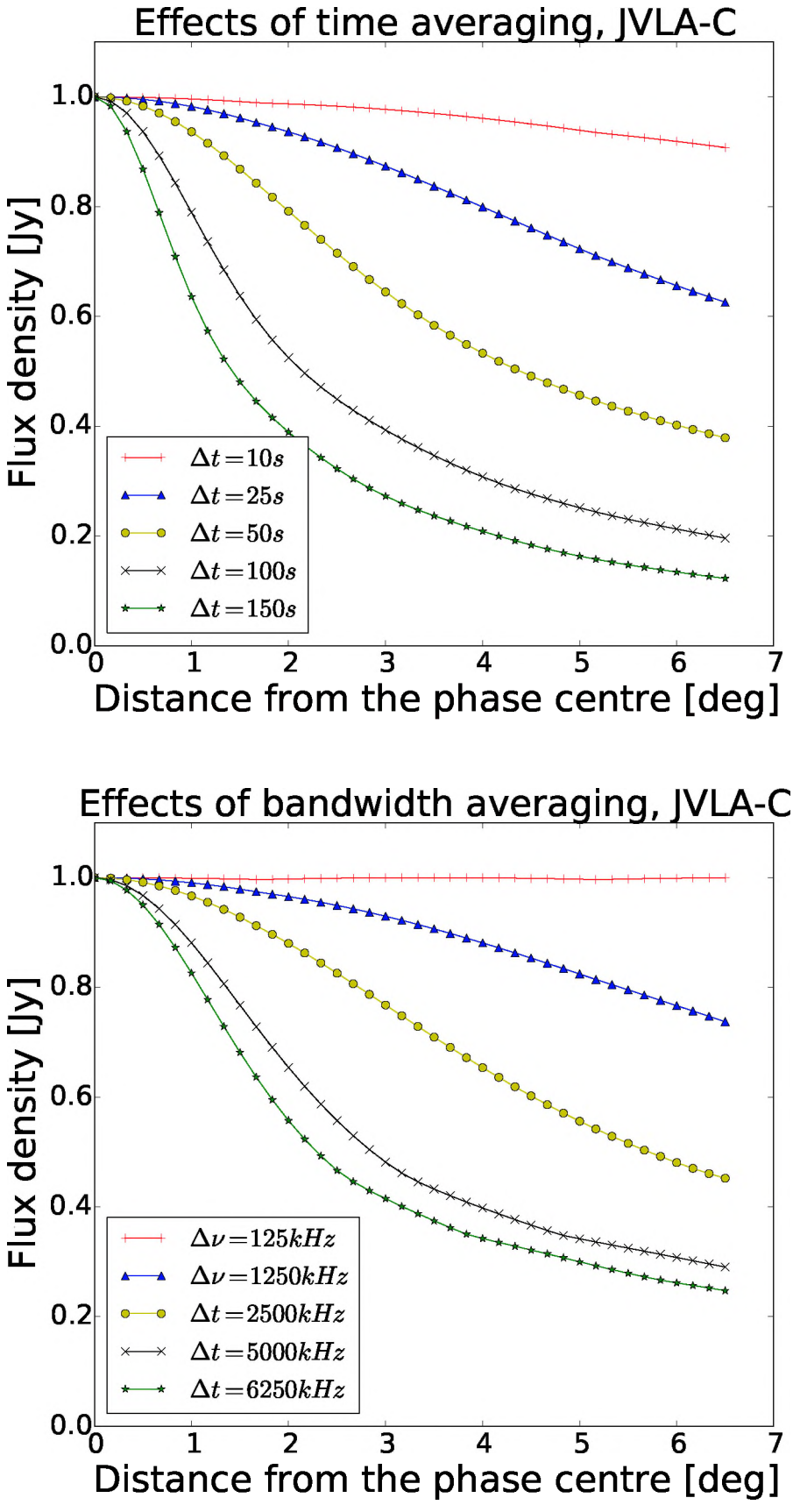

FIGURE 1.3: Effects of time and frequency averaging: the apparent intensity of a 1 Jy source, as seen by JVLA-C at 1.4 GHz, as a function of distance from the phase centre. (Top) Frequency interval fixed at 125 kHz, and time interval varies; (bottom) time interval fixed at 1 s, and frequency interval varies.

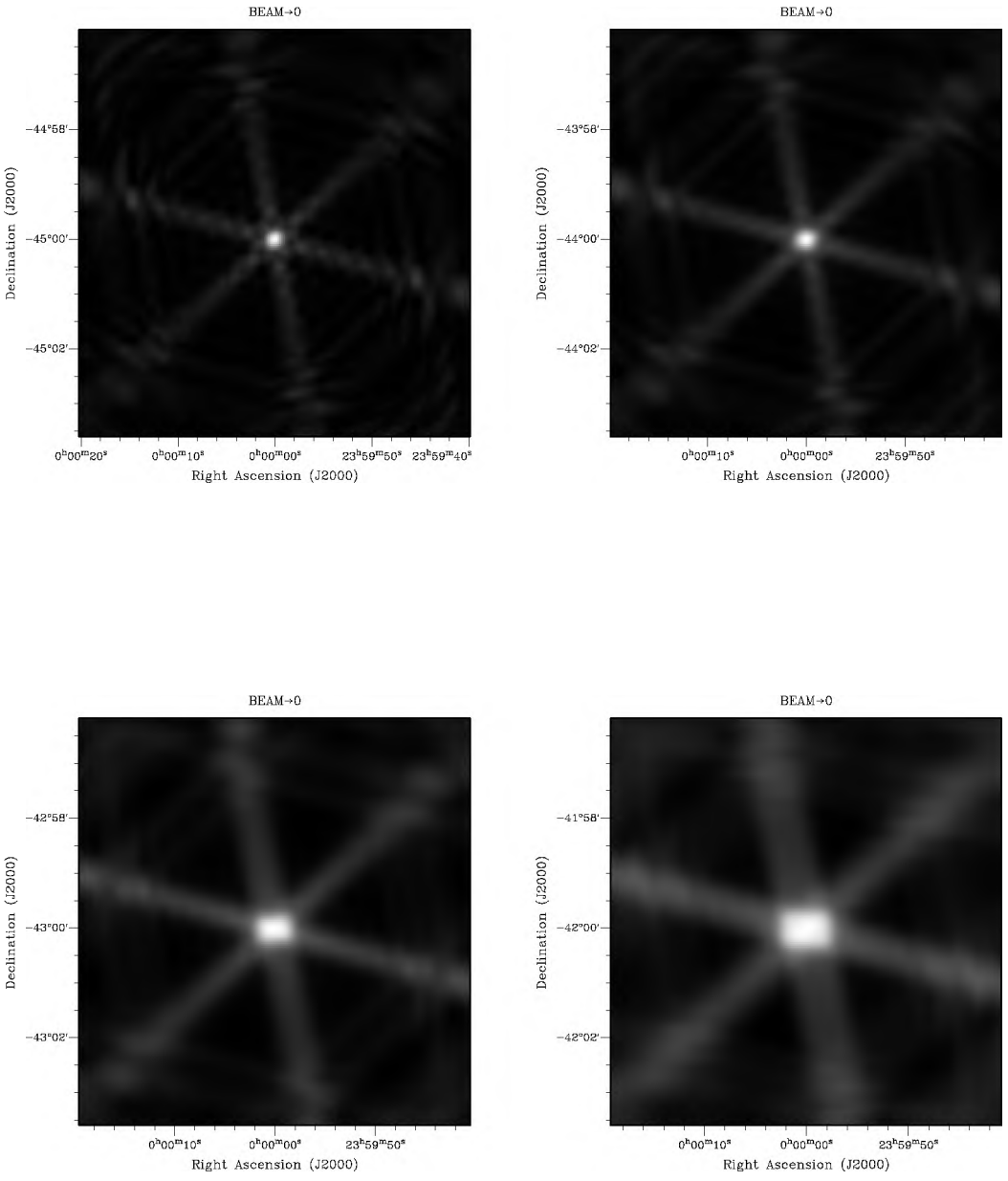

FIGURE 1.4: PSF at four sky locations, result of 100 s integration and 5 MHz bandwidth averaging. (Top left) PSF at the phase centre. (Top right) PSF at 1°. (Bottom left) PSF at 2°. (Bottom right) PSF at 3°.

effective FoV to reach this limit. This imposes an upper limit on the size of a time/frequency bin: it must be small enough to keep amplitude loss acceptably low across the entire primary beam FoV.

(b) For deep imaging it is useful to suppress bright sources in the primary beam sidelobes outside the FoV. Even with smearing of sources far from the phase centre, bright sources can contribute significant flux and PSF sidelobe artefacts to the image.

(c) In VLBI, where decorrelation is more severe, the effective FoV is determined by the smallest time/frequency bin size that a correlator can support and is normally much smaller than the primary beam [Keimpema et al., 2015a]. Modern VLBI correlators overcome this by employing a technique where the signal is correlated relative to multiple phase centres simultaneously, thus effectively "tiling" the primary beam by multiple FoVs. This has a computational cost that scales linearly with the number of phase centres.

When considering a short sequence of visibilities measured on one baseline, we can consider averaging as a convolution of the true visibility by a boxcar function corresponding to the uv-extent of the averaging bin, followed by sampling at the centre of each bin, as illustrated in Figure 1.5. Convolution in the visibility plane corresponds to multiplication of the image by an *image-plane response function* that is the Fourier transform of the convolution kernel, i.e. the window function; the Fourier transform of a boxcar is a "sinc-type taper". If we consider the entire uv-plane, averaging is only a "*pseudo-convolution*", since the different uv-bins (and thus their boxcars) will have different sizes and shapes determined by baseline length and orientation. Still, we can qualitatively view smearing as some kind of cumulative effect of an ensemble of image-plane tapers corresponding to all the different boxcars. We should note that this "smearing taper" is not the only tapering effect at work in interferometric imaging. Firstly, antennas have a non-zero physical extent: a measured visibility is already convolved by the aperture illumination functions of each pair of antennas (refer to Eq. (1.8)). The resulting image-plane taper is exactly what the primary beam is. Secondly, most imaging software employs convolutional gridding followed by the fast Fourier transform (FFT), which produces an additional taper that suppresses aliasing of sources from outside the imaged region.

If weighted averaging instead of simple averaging (whether in the correlator, or in post-processing) should be employed, this would correspond to a "pseudo-convolution" of the

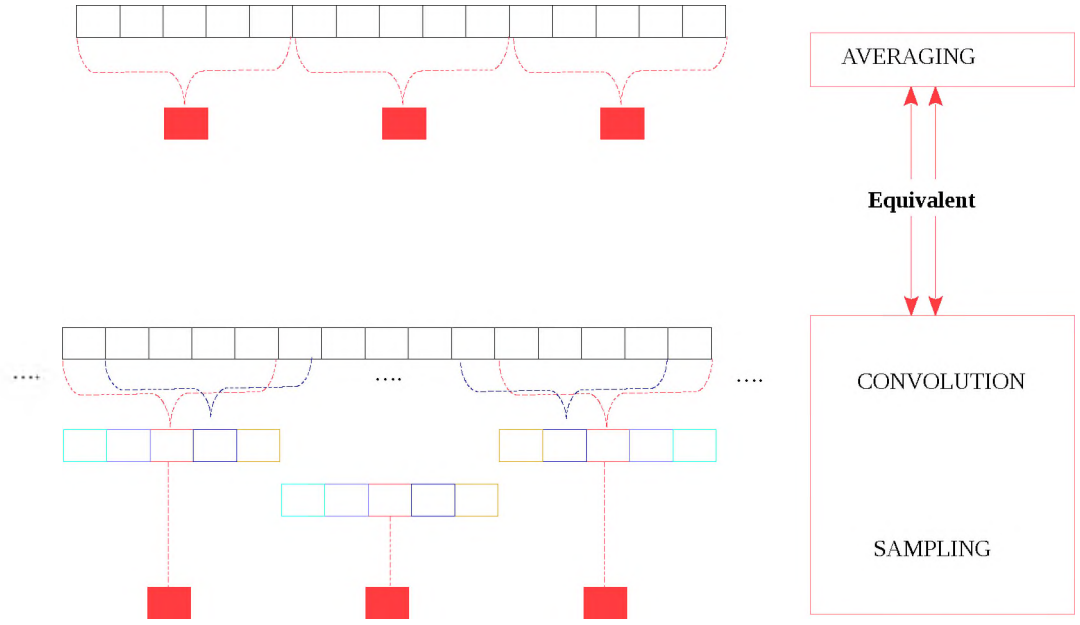

FIGURE 1.5: The graph shows the equivalent between averaging and convolution. The black grid presents 15 visibility measurements to average in three blocks. In the top level, each block is averaging. At the bottom level, each block is convolved with a boxcar window, the result is then sampled by a sampling function at the centre of each block: this process is equivalent to averaging.

uv-plane by some ensemble of *window functions*, different from boxcars, which would obviously yield different image-plane tapers and thus result in a different smearing response.

Throughout this work we use the term field of interest (FoI), which we differentiate from the FoV. In traditional interferometer usage, the FoV is related to the angular scale of the primary beam [Leisawitz et al., 2002, Morgan et al., 2011], whereas the FoI is a parameter of the scientific observation case, which may be related to the size of the primary beam, but this is not necessary. Thus, the FoI will be an adjustable parameter in the window functions we present.

Signal restoration is one of the applications of filter theory. Restoration needs to be performed when the signal has been distorted before being captured. Decorrelation/smearing is an example of signal distortion, hence filter theory becomes relevant (we refer the reader to the book of Smith et al. [1997] for an overview of filter theory). Filter theory suggests that a window function can be tuned to achieve some desired tapering response. An optimal taper would be one that was maximal across the desired FoI (**Regime 1** in

Figure 1.6), and minimal outside it (**Regime 2** and **Regime 3** in Figure 1.6). In this work:

(a) We apply filter theory to derive a set of *baseline-dependent window functions* (BDWFs) that approximate this more optimal smearing behaviour. The trade-off is an increase in thermal noise, since minimum noise can only be achieved with unweighted averaging. We show that this effect can be partially mitigated through the use of *overlapping* BDWFs. Offringa et al. [2012] have investigated a similar approach in the context of suppressing signals towards specific off-axis sources.

(b) We combine BDWFs with baseline-dependent averaging (BDA). The ability of BDWFs to shape the FoI is somewhat limited by the fact that shorter baselines sweep out a smaller bins in uv-space and over these small bins, BDWFs become similar to simple averaging. So, if BDA is employed to BDWFs, short baselines would be averaged over larger uv-bins, thus increasing the effectiveness of BDWFs.

(c) We note that the use of BDWFs results in a different position-dependent PSF than simple averaging. In others words, the smearing response of BDWFs results in a different smeared PSF shape. A fast scheme to approximate these different PSFs at all sky locations is proposed and implemented in DDFacet [Tasse et al., b, in preparation]: imaging software under development.

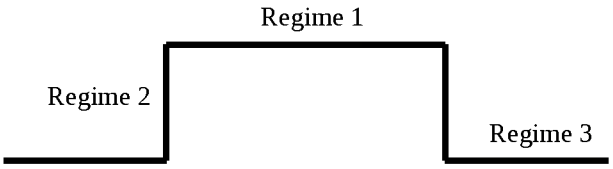

FIGURE 1.6: Optimal taper. In the image plane, this represents the ideal image plane tapering function.

In the era of the SKA and its pathfinders, where dealing with huge volumes of data is one of the main challenges, use of BDWFs potentially offers additional leverage in optimising radio observations. Decreased smearing across the FoI allows for more aggressive data averaging, thus reducing storage and computation costs. The trade-off is a loss of sensitivity, which pushes up observational time requirements. However, the decrease in smearing and noise from A-team sources could, conceivably, make up for some of the nominal sensitivity loss. In the VLBI case, use of BDWFs potentially offers an increase

in effective FoI at a given correlator dump rate, or equivalently, the ability to tile the primary beam with fewer phase centres, allowing for smaller correlators.

1.8 Thesis layout

This research work proposes a technique based on filter theory to compress radio interferometric data while keeping amplitude loss and distortion down to an acceptable level. This poses challenges to traditional compression approaches due to the Gaussian nature of the radio interferometric noise, which makes the data difficult to compress. We propose the use of BDWFs and overlapping BDWFs as suitable *uv*-space filters to deal with various aspects of data compression. Similarly to averaging, the use of BDWFs also distorts the PSF shape for sources out of the phase centre. Currently, there is no imaging algorithm that handles the PSF shape distortion at each sky location during deconvolution, since it is computationally expensive to do so. Hence, we have proposed a mathematical framework to approximate these PSFs at different locations in the sky. The approximation results in cheaper computation. Finally, we implement the position-dependent PSFs in an imaging framework. This thesis is organised as follows:

Chapter 2 starts with a brief review on sampling and aliasing in signal processing. Mapping between terms used in signal processing and aperture synthesis is proposed. A description for choosing a suitable window function for data compression and FoI shaping follows, and I conclude by choosing window functions used in the rest of the thesis.

In Chapter 3, I describe in detail the effect of averaging in the *uv*-space as well as in the image plane. This includes the mathematical relation between averaging and convolution and the analytical and empirical noise penalty for averaging. I then present BDWFs and overlapping BDWFs, including various figures of merit for some of these BDWFs, particularly the case of sinc and Bessel. I also discuss the performance of BDWFs applied to JVLA and compare the results to simple averaging. I end the chapter by showing the case of BDWFs for wide FoI imaging and data compression in VLBI.

Chapter 4 covers the applicability of BDA in radio interferometry, thereby describing the mathematics involved and the data size. I combine both BDA and BDWFs in spatial

filtering and discuss some results. I conclude by presenting some calibration challenges related to BDA.

Chapter 5 deals directly with the PSF in radio interferometry. I propose a mathematical framework that accurately approximates the PSF of off-axis sources, resulting in cheaper computation. This is followed by the implementation of position-dependent PSF into DDFacet. I present the results and discuss the computational cost compared to brute-force computation of the PSF.

In Chapter 6 I summarise the present work, draw conclusions and present future work.

Chapter 2

Digital Signal Processing: sampling, aliasing and window Functions

A *physical signal* is a continuous time sequence of real or complex numbers, and an *observed signal* is a discrete time sequence of real or complex numbers. These two definitions are fundamental to this chapter. Sampling consists of selecting points of the physical signal to observe so that one may estimate this physical signal from the observed signal [Saloma, 1993]. In other words, we want to infer a continuously varying function (physical signal) from a discrete time-varying function that is measured (observed signal). A physical signal is measured within a discrete time range, which is equivalent to sampling and truncation over a boxcar window [Vetterli et al., 2002]. Some questions that we will be answering in this chapter are on the best way for obtaining the observed signal and the best methods to study the physical signal from this observed signal. The chapter begins by presenting some sources of imaging artefacts. After this we look at sampling and properties of the FFT upon which signal analysis is based. I review window functions (finite impulse response filters) and establish the best window functions that would minimise artefacts and conserve useful information in radio interferometric data.

2.1 Imaging artefacts

Sampling during signal processing and naive calibration lead to loss of information and distortion of signals. Imaging artefacts manifest themselves in various ways, which can

be categorised as:

Ghost sources: These artefacts are very faint in nature and occur in response to the incompleteness of the calibration sky model used during self-calibration [Cornwell and Fomalont, 1989]. Ghost artefacts were initially discovered by Bos [1985] in a calibrated 610 MHz Westerbork observation. Their study was further extended by Grobler et al. [2014a], Nunhokee [2015] who showed that the ghost sources were formed in a peculiar pattern. This pattern is highly dependent on the linear geometry of the Westerbork Synthesis Radio Telescope [Hogbom and Brouw, 1974]. Further work by Grobler et al. [2014b] extends this to a 2-D interferometer layout.

Distant source sidelobes: These artefacts are generated by the sidelobes of bright objects extending inside the interferometer target field. These objects are sometimes found within the interferometer array FoV. A common example is the bright radio galaxy Cygnus A [Bach et al., 2002]. To overcome this problem with LOFAR observations, a catalogue of bright sources (*A-team* sources) that can contaminate observations has been established. The A-team sources are often subtracted iteratively from observations before analysis. This method has shown good results, but involves computation that scales as the number of A-team sources.

Sampling and the fast Fourier transform: These artefacts are of crucial importance in any field applying signal processing. The artefacts are due to the discontinuity, improper sampling and finite space representation. Physical signals are continuous and extend infinitely, while observed signals are digitised and band-limited. In signal processing, the discrete or the FFT is often applied to reconstruct the signal. In order to compute the FFT, signals are discretised and interpolated onto a regular grid, both in the time and frequency domains. The process of interpolation is often referred to as gridding. Gridding of the sampled signal consists of multiplying the signal with a convolution kernel, and since in practice the convolution kernel has finite support (is band-limited), it may lead to aliasing effects after applying the FFT to the gridded signal.

The FFT allows us to use mathematical operations such as convolution, which is computationally efficient when we analyse large data-sets. The FFT can be seen as a method for transforming a signal from the Fourier space to the real space with fewer computational resources compared to the discrete Fourier transform. For example, a data-set of dimension N will require N^2 complex multiplications and $N(N-1)$ additions to compute

its discrete Fourier transform, i.e. an arithmetic complexity of $\mathcal{O}(N^2)$, while the FFT will have a complexity of $\mathcal{O}(N \log_2 N)$ [Duhamel and Vetterli, 1990]. In this chapter we discuss the requirements imposed on sampling by the need to avoid sampling artefacts or aliasing.

2.2 Sampling and the fast Fourier transform

Below we discuss the requirements for sampling as described in Shannon et al. [1967]. Suppose that $d_r(x)$ is the physical signal and $d_s(x)$ its sampled form. According to Shannon et al. [1967], the sampled signal is expressed as:

$$d_s(x) = s(x)d_r(x), \tag{2.1}$$

where $s(x) = \sum_{k=-\infty}^{+\infty} \delta(x - k\Delta x)$, is an infinite series of Delta functions also known as a Dirac Comb at a regular spacing Δx and k is the number of discrete points on the sampling grid. This definition states that sampling is equivalent to multiplying the physical signal by the sampling function $s(x)$, as shown in Figure 2.1. Each discrete point on the sampling grid is:

$$d_{s,k} \equiv d_r(k\Delta x). \tag{2.2}$$

If one combines Eq. (2.1) and (2.2), the sampled signal becomes the sum of each discrete point on the sample grid multiplied by a Dirac delta function.

$$d_s(x) = \sum_{k=-\infty}^{+\infty} \delta(x - x_k)d_r(x_k), \quad x_k = k\Delta x. \tag{2.3}$$

Eq. (2.3) is depicted graphically at the bottom centre of Figure 2.1. Since we are interested in the Fourier and the time-domain signal, the Fourier transform is applied to Eq. (2.1). According to the convolution theorem, the Fourier transform of the product of two functions yields the convolution of their respective Fourier transform. This is written as:

$$\mathcal{F}\{d_s\}(y) = \Big(\mathcal{F}\{s\} \circ \mathcal{F}\{d_r\}\Big)(y), \tag{2.4}$$

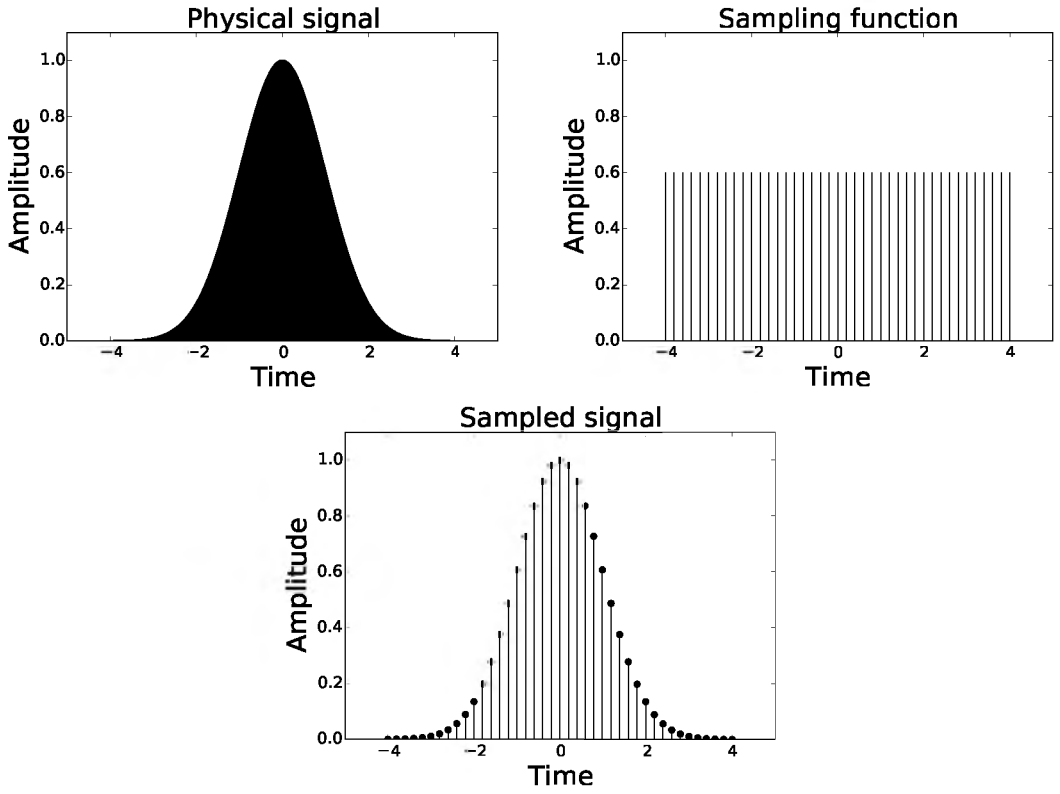

FIGURE 2.1: A Gaussian physical signal (top left) and its sampled form (bottom centre). The physical signal is multiplied by the Dirac Comb function (top right) and this results in the sampled signal (bottom centre).

where $\mathcal{F}$ denotes the Fourier transform and $\circ$ is the convolution operation. The Fourier transform of a Dirac Comb of regular spacing Δx is a Dirac Comb of regular spacing Δy

$$\mathcal{F}\{s\} = \Delta y \sum_{k=-\infty}^{+\infty} \delta(y - k\Delta y), \tag{2.5}$$

where $\Delta y = 2\pi/\Delta x$ is the sampling rate. Combining Eq. (2.5) and Eq. (2.4), we obtain the following:

$$\begin{aligned}\mathcal{F}\{d_s\}(y) &= \Big(\Delta y \sum_{k=-\infty}^{+\infty} \delta(y - k\Delta y)\Big) \circ \Big(\mathcal{F}\{d_r\}(y)\Big) \\ &= \Delta y \sum_{k=-\infty}^{+\infty} \mathcal{F}\{d_r\}(y - k\Delta y).\end{aligned} \tag{2.6}$$

See Appendix A for the detailed proof of Eq. (2.5) and Eq. (2.6).

For completeness, the spectrum from the point $k = 0$ in Eq. (2.6) can be separated:

$$\begin{aligned}\mathcal{F}\{d_s\}(y) &= \Delta y \mathcal{F}\{d_r\}(y) + \Delta y \sum_{k=-\infty, k\neq 0}^{+\infty} \mathcal{F}\{d_r\}(y - k\Delta y) \\ &\simeq \mathcal{F}\{d_r\}(y) + \sum_{k=-\infty, k\neq 0}^{+\infty} \mathcal{F}\{d_r\}(y - k\Delta y). \end{aligned} \tag{2.7}$$

It is shown in Eq. (2.7) that:

(a) The Fourier transform of the sampled signal duplicates the Fourier transform of the physical signal at each point in the Fourier transform of the Dirac Comb. This is illustrated in Figure 2.2.

(b) Sampling and the Fourier transform conserve the physical signal $\mathcal{F}\{d_r\}(y)$ at the point $k = 0$ and replicate it at all points $k \neq 0$ with regular spacing Δy.

(c) The copies for which $k \neq 0$ i.e. $\sum_{k=-\infty, k\neq 0}^{+\infty} \mathcal{F}\{d_r\}(y - k\Delta y)$ are aliased spectra.

In addition, it should be noted that any reconstruction of d_r from d_s requires the suppression of the aliased spectra i.e. all copies of $\mathcal{F}\{d_r\}$ for which $k \neq 0$ [Diakopoulos and Stephenson, 2005].

A method to suppress these aliased copies consists of multiplying $\mathcal{F}\{d_s\}$ by a boxcar window function, given that a boxcar window function is unity in the passband (area in which $k = 0$) and zero in the stopband (area in which $k \neq 0$).

Note that a complete reconstruction of d_r requires the absence of overlapping copies of the aliased spectra with the Fourier transform of the physical signal i.e. $\mathcal{F}\{d_r\}$ for which $k = 0$. Otherwise, d_r cannot be uniquely reconstructed. According to Diakopoulos and Stephenson [2005] the following conditions must hold for a unique recovery:

(a) The sampled signal must be band-limited in order to avoid that $\mathcal{F}\{d_r\}(y)$ at point $k = 0$ extends infinitely and overlaps with its aliasing copies.

(b) The sampling frequency Δy, must be greater than twice the highest frequency f_s contained in the signal ($\Delta y \geq 2f_s$). The quantity $2f_s$ is the minimum sampling frequency or the Nyquist rate required to avoid aliasing (see Figure 2.3). This condition makes Δy of critical importance to avoid aliasing [Shannon et al., 1967].

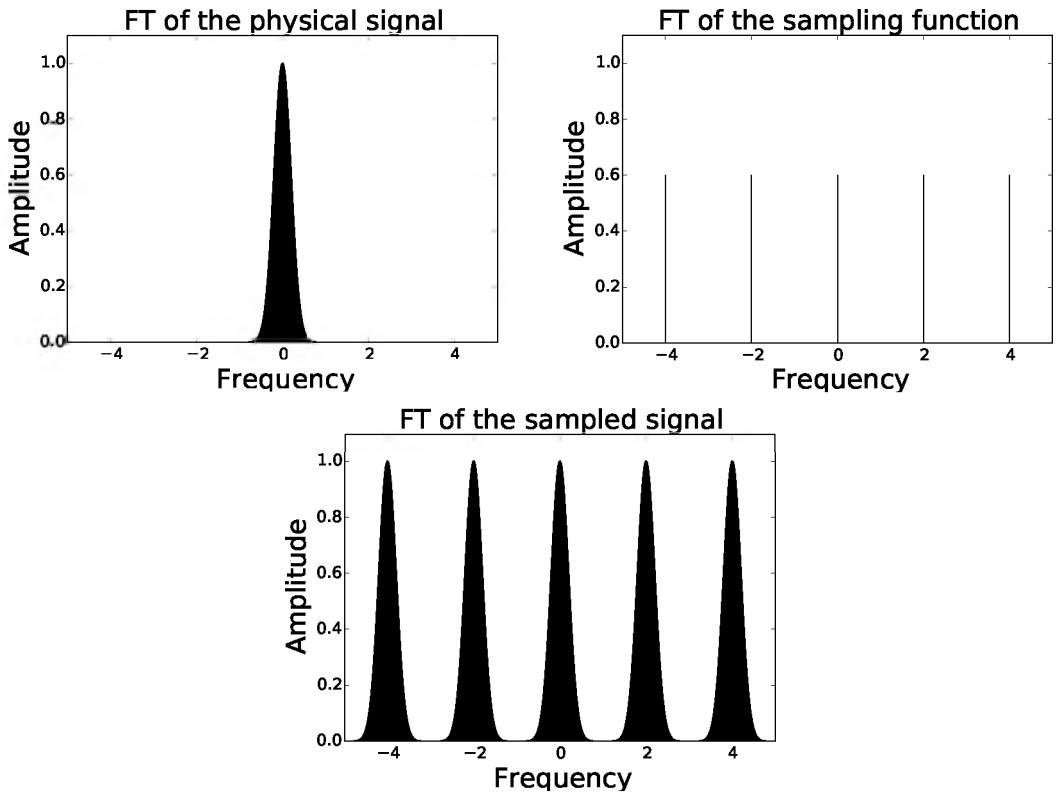

FIGURE 2.2: (Top left): The Fourier transform of the physical signal. (Top right): The Fourier transform of the Dirac Comb. (Bottom centre): The Fourier transform of the sampled signal. The plot shows that the Fourier transform of the sampled signal is a superposition of the Fourier transform of the physical signal at each point in the Dirac Comb.

Suppose that these conditions (a+b) are satisfied, and that $\Pi^{[k]}$ is a boxcar window function, given by:

$$\Pi^{[k]}(y) = \begin{cases} 1, & k = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

Let us now introduce the observed signal d_o; multiplying Eq. (2.7) by $\Pi^{[k]}$ we have:

$$\mathcal{F}\{d_o\}(y) = \Pi^{[k]}(y)\mathcal{F}\{d_s\}(y). \tag{2.9}$$

Applying the inverse Fourier transform to Eq. (2.9) and making use of the convolution theorem leads to:

$$d_o(x) = \left(\mathcal{F}^{-1}\{\Pi^{[k]}\} \circ d_s\right)(x). \tag{2.10}$$

So far, since we have assumed that conditions (a+b) are satisfied (i.e. no overlapping aliasing spectrum with $\mathcal{F}\{d_r\}(y)$ exists at the point $k = 0$), one can still write Eq. (2.9)

as follows:

$$\begin{aligned}\mathcal{F}\{d_o\}(y) &= \Pi^{[k]}(y)\mathcal{F}\{d_s\}(y) \\ &\simeq \Pi^{[k]}(y)\Big(\mathcal{F}\{d_r\}(y) + \sum_{k=-\infty,k\neq 0}^{+\infty} \mathcal{F}\{d_r\}(y-k\Delta y)\Big) \\ &\simeq \Pi^{[k]}(y)\mathcal{F}\{d_r\}(y). \end{aligned} \tag{2.11}$$

We note from this substitution that the term $\Pi^{[k]}(y)\sum_{k=-\infty,k\neq 0}^{+\infty}\mathcal{F}\{d_r\}(y-k\Delta y)$ is washed out because of the assumption that at $k=0$ no overlapping spectrum exists (i.e. $\Pi^{[k]}(y)\sum_{k=-\infty,k\neq 0}^{+\infty}\mathcal{F}\{d_r\}(y-k\Delta y)=0$). Likewise, after applying the Fourier transform to Eq. (2.11), we have the following convolution:

$$d_o(x) \simeq \Big(\mathcal{F}^{-1}\{\Pi^{[k]}\}\circ d_r\Big)(x). \tag{2.12}$$

Now one can see that Eq. (2.9) is approximately equal to Eq. (2.11) at $k\neq 0$:

$$\Pi^{[k]}(y)\mathcal{F}\{d_s\}(y) \simeq \Pi^{[k]}(y)\mathcal{F}\{d_r\}(y), \tag{2.13}$$

which leads to

$$d_s \simeq d_r \tag{2.14}$$

after simplifications.

So far, Eq. (2.14) demonstrates that the physical signal can be reconstructed from the sampled signal if conditions (a+b) are satisfied. However, ideal sampling cannot be applicable in practical problems. This is due to the convolution of the physical signal with a low pass taper in Eq. (2.12). The low pass taper employed in Eq. (2.12) is a "sinc-like" window function $\mathcal{F}^{-1}\{\Pi^{[k]}\}$, and thus cannot be represented with finite support.

2.3 BDWFs' relationship to digital signal processing

This work is a specific application of signal conditioning using digital signal processing (DSP). It is useful to relate this work to DSP concepts and terminology. Ideally, all spatial modes, up to the resolution of the longest baseline, are sampled in a 2-D continuous sky

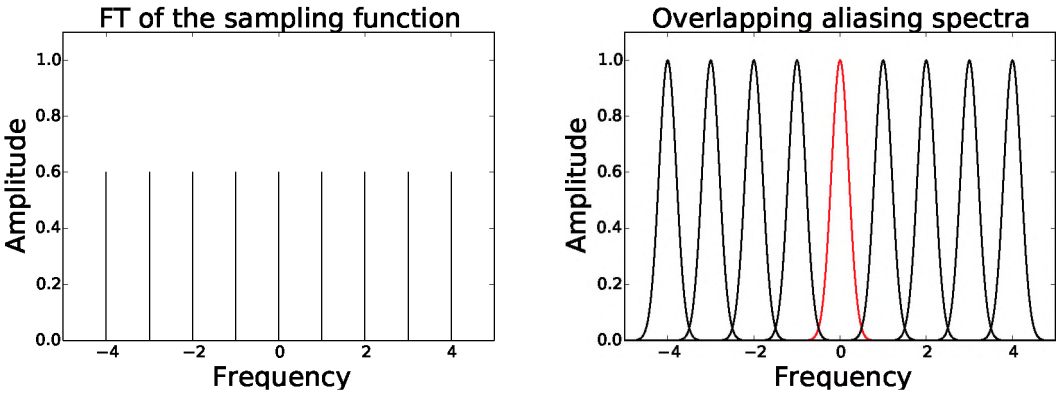

FIGURE 2.3: Sampling and aliasing: the signal is sampled at a frequency less than the minimum sampling rate. (Left): Dirac Com with spacing below the minimum sampling rate. (Right): Fourier transform of the sampled signal showing that the aliasing spectra are overlapping with the physical signal.

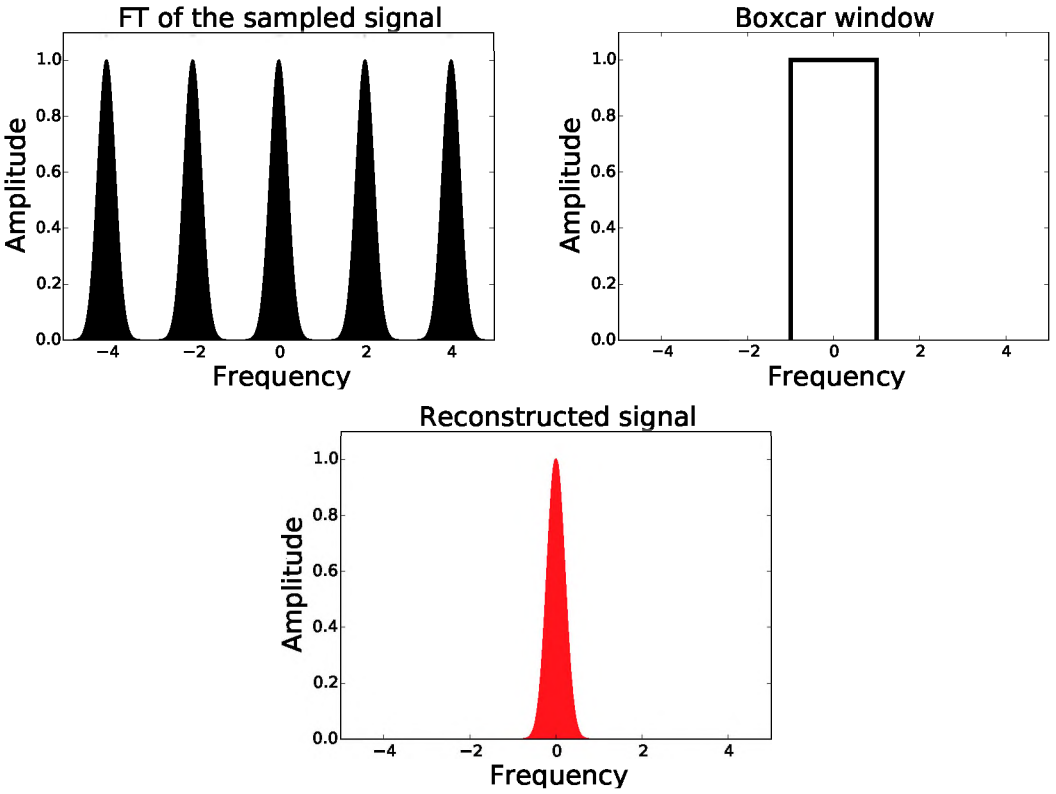

FIGURE 2.4: Ideal sampling showing that a multiplication of the Fourier transform of the sampled signal (top left) with a boxcar window (top right) suppresses all aliasing spectra and leaves only the Fourier transform of the physical signal (bottom centre).

image $\mathcal{I}$ (refer to Eq. (1.5)). This requires Nyquist sampling of the visibility domain out to the longest baseline spatial modes. This is rarely possible because of the unfilled holes in the *uv*-coverage during an observation, the lower spatial frequency cut-off due to physical element limitations, and sampling bias in the low spatial frequency region of the visibility plane compared to higher spatial frequencies due to baseline distribution. For a fixed time length, a long baseline will cover a longer track in visibility space compared to a shorter baseline, which results in the lower spatial modes being oversampled compared to higher spatial modes. On shorter baselines, the window functions act as filters with a larger noise equivalent bandwidth compared to longer baselines. This necessitates the use of BDWFs to optimise the image-plane response (IPR), which is the aggregate spectral response of the combined individual baseline filters.

A matched filter maximises the signal-to-noise ratio of a desired signal in the presence of noise. In the case of synthesis imaging, the desired signal is within a defined FoI in the sky. This is the passband region (i.e. **Regime 1** in Figure 1.6) of the IPR and any signal outside this FoI is in the stopband (i.e. **Regime 3** in Figure 1.6), which we would like to suppress. The choice of BDWF can depend on the requirement to limit the transition width (i.e. **Regime 2** in Figure 1.6) between the passband and stopband or limit the passband ripple, or minimise the sidelobe level or some other filter optimisation metric. An approximate matched filter will be an array, sky field, and observation dependent filter. The choice of window or filter size is a compromise that uses the largest possible window size that does not introduce significant smearing. The window size cannot easily be defined analytically, given that it depends on the FoI we want to achieve, the interferometer array layout, the observation frequency, and baselines' length and direction.

In signal processing, a finite impulse response window function is a mathematical function with limited support, i.e. zero outside some interval. Conventionally, a time series is convolved with a window to produce some desired response in the frequency domain. Applying this to our problem can lead to some confusion in terminology. Table 2.1 provides a mapping between the terms commonly used in signal processing, and their conceptual equivalent in aperture synthesis. Confusion in terminology can arise from the number of different signal domains used in this work. The visibility of a baseline is sampled constantly in the correlator time domain, but is irregularly sampled in the

Signal processing	**Aperture synthesis**
Frequency (freq) domain	Image plane
Time domain	Fourier plane or uv-plane
Spectral response or freq response	IPR
Time response	Fourier plane response
Cut-off time interval or time passband	uv-averaging bin
Cut-off freq interval or freq passband, or main lobe	FoI
Time stopband	Outside of the uv-bin
Freq stopband	Outside of the FoI
Octave	Doubling in size
Band-limited (applied to visibilities)	restricted FoI

TABLE 2.1: Mapping of terminology between DSP and aperture synthesis.

2-D visibility domain (for simplicity we are only analysing the uv-plane of the 3-D uvw-space). The imaging domain is the 2-D Fourier pair of the visibility space. In synthesis imaging we are making the assumption that the sky is a constant signal, but a time varied signal is measured because of the array rotation (due to the Earth' rotation) with respect to the observed field. This application is not a linear time-invariant system, which leads to complexity in the analysis of the signal conditioning. The window function filters are linear but are dependent on baseline length, which varies with time. It is not sufficient simply to analyse the filter impulse response. Instead, for our analysis we will use the aggregate IPR. This work does not attempt to find the optimal matched filter; we leave this to a future work. All the window functions presented are finite impulse response filters followed by decimation. That is, they are window-function-weighted, moving averages of the measured visibilities. In the simple case of non-overlapping windows the decimation rate is equal to the size of the window filter. The overlapping window filters presented in Section 3.2.1 can be seen as a type of poly-phase filter. We do not explore the use of infinite impulse response filters in this work, but this is a clear path for further study.

Broadly speaking we want the Fourier spectrum (i.e. IPR) of the visibility domain filter to approximate a boxcar window. Standard DSP filter terminology uses the "peak sidelobe level", the "main lobe width" and the "sidelobes roll-off" to describe the performance of window functions. In terms of the "ideal" IPR (see Figure 1.6), these correspond to the following desirable traits we want to achieve:

- Maximally conserve the signal within the FoI (**Regime 1** in the figure), and make the transition region (**Regime 2** in the figure) as sharp as possible: both of these correspond to a larger main lobe width.
- Attenuate sources outside the FoI (**Regime 3** in the figure): this corresponds to a lower peak sidelobe level and higher sidelobe roll-off.

Below we provide a review of some common 1-D window functions employed in DSP.

2.3.1 Boxcar window

The boxcar window for a cut-off time interval of $[-t_a, t_a]$ is defined as:

$$\Pi(t/t_a) = \begin{cases} 1 & -t_a \leq t \leq t_a \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

Figure 2.5 shows a plot of $\Pi(t/t_a)$ and its IPR. The thick grey and thin curves correspond to cut-off time intervals of $[-t_a, t_a]$ and $[-t_a/2, t_a/2]$ respectively. Note that when the cut-off time interval is larger, the main lobe width becomes narrower and the first sidelobe remains at the same level, while the other sidelobes are lower.

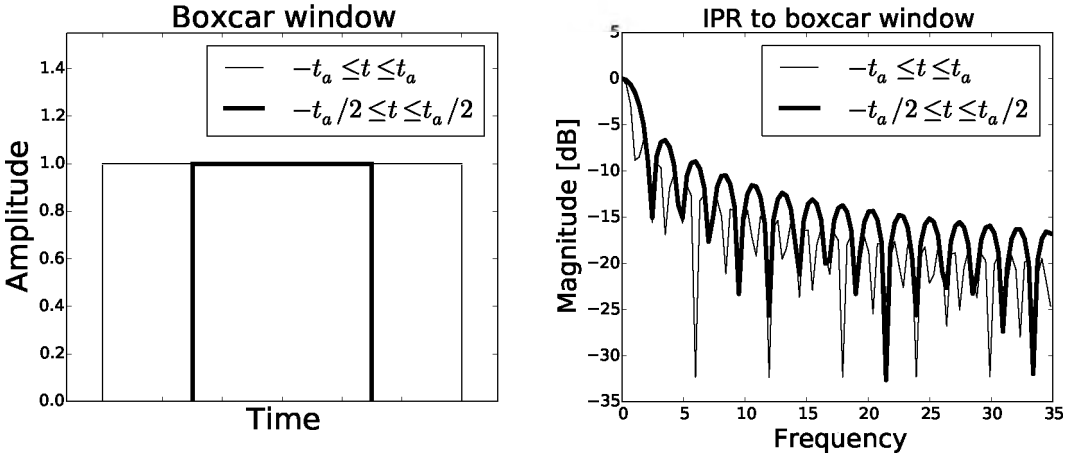

FIGURE 2.5: Boxcar and its IPR.

The other window functions given below are all multiplied with a boxcar to ensure a cut-off interval of $[-t_a, t_a]$, unless underwise mentioned.

2.3.2 Gaussian window

A Gaussian window function centred at zero with a standard deviation of σ_1 is given by:

$$\mathrm{G}(t) = \Pi(t/t_a)e^{-bt^2}, \tag{2.16}$$

where $b = (2\sigma_1^2)^{-1}$. The Fourier transform of the Gaussian term is given by $\mathcal{F}\{\mathrm{G}\} = \sqrt{\frac{b}{\pi}}e^{-cl^2}$, where $c = \pi^2/b$, i.e. it is also Gaussian with a standard deviation of $\sigma_2 = (2\pi\sigma_1)^{-1}$.

Figure 2.6 shows a plot of $\mathrm{G}(t)$ and its IPR. The window function is truncated at $[-t_a, t_a]$, with $b = 3$ for the thin curve and $b = 5$ for the thick grey curve. Its response is characterised by extremely low sidelobes, but a narrow main lobe.

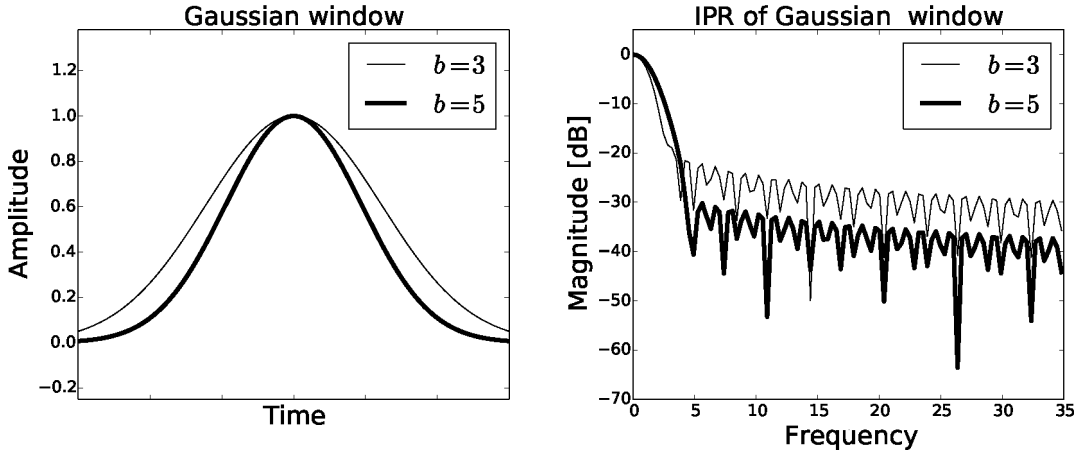

FIGURE 2.6: Gaussian and its IPR.

2.3.3 Butterworth window

A Butterworth window function is flat in the time passband, rolls off towards zero in the time stopband and is characterised by two independent parameters, the cut-off time $[-t_a, t_a]$ and the order p. The two parameters control the FoI and sidelobes attenuation. The Butterworth window function is given by:

$$\mathrm{BW}(t) = \Pi(t/t_a)\Big(1 + (t/t_a)^{2p}\Big)^{-1}. \tag{2.17}$$

Figure 2.7 shows Butterworth window functions for the same cut-off interval $[-t_a, t_a]$, with orders of p=1 and p=3. Note that increasing the order p conserves the main lobe width, and dramatically lowers distant sidelobes, at the cost of pushing up the near-in sidelobes.

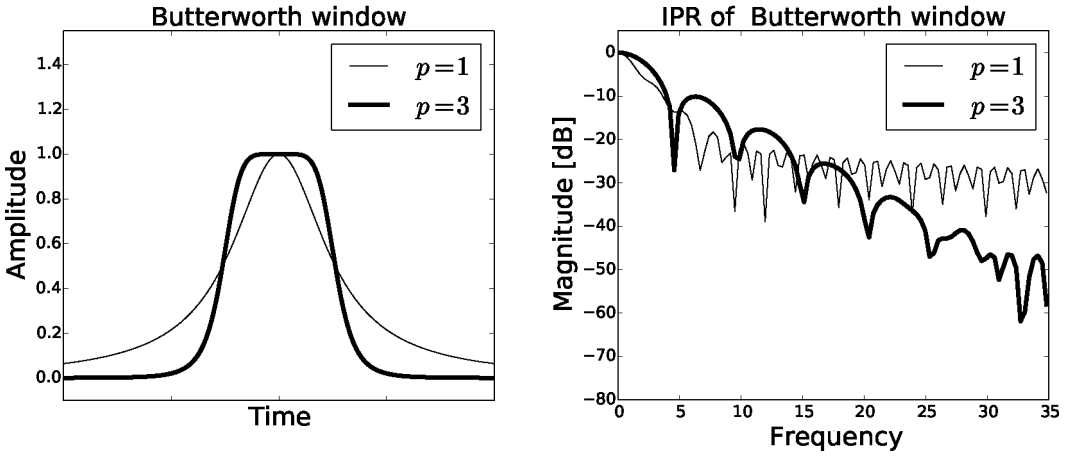

FIGURE 2.7: Butterworth and its IPR.

2.3.4 Prolate spheroidal window

This window function is given by a prolate spheroidal wave function of sequence zero ($n = 0$) characterized by two independent parameters, the cut-off time $[-t_a, t_a]$ and the order α [Delsarte et al., 1985, Walter and Soleski, 2006]. The two parameters control the FoI and sidelobes attenuation. The prolate spheroidal wave function PS_0 is the eigenfunction and solution of the integral:

$$\int_{-t_a}^{t_a} \mathrm{PS}_0(\xi) \frac{\sin(\frac{\alpha}{\pi}(t-\xi))}{\frac{\alpha}{\pi}(t-\xi)} d\xi = \lambda_{n=0,\alpha,t_a} \mathrm{PS}_0(t), \tag{2.18}$$

where $\lambda_{n=0,\alpha,t_a}$ is the corresponding eigenvalue. Figure 2.8 shows prolate spheroidal window functions for the same cut-off interval $[-t_a, t_a]$, with orders of $\alpha = 2\pi^3$ and $\alpha = 5\pi^3$. Note that increasing the order α increases the main lobe width, and dramatically lowers the sidelobes.

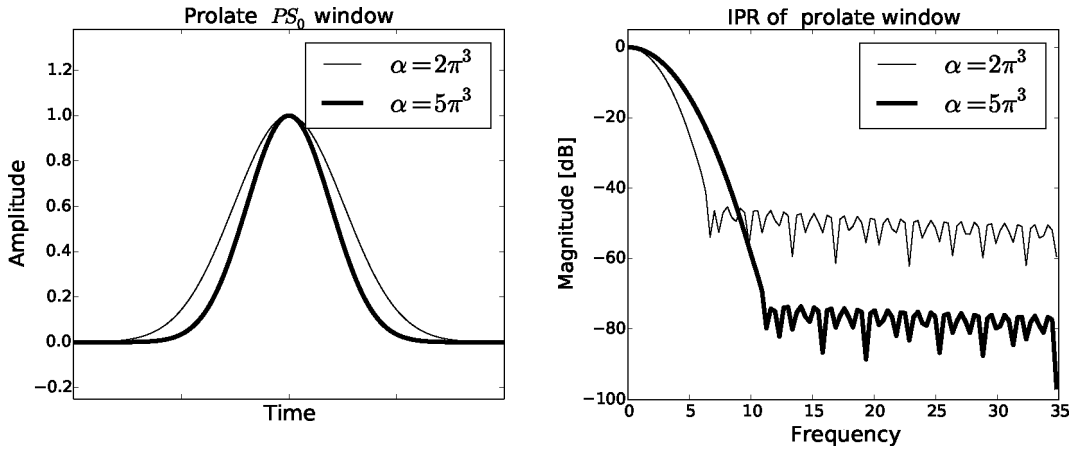

FIGURE 2.8: Prolate window (of order zeros) and its IPR.

2.3.5 Sinc window

The sinc window function is defined as:

$$\text{sinc}(t) = \Pi(t/t_a)\frac{\sin(\pi bt)}{\pi bt}. \tag{2.19}$$

Figure 2.9 shows sinc(t) for a fixed value of b, and cut-off intervals given by $[-t_a, t_a]$ and $[-t_a/2, t_a/2]$. Note that the IPR of a sinc window function is almost perfectly flat in the main lobe (more so for larger intervals). The sidelobe response is relatively poor, but better for larger intervals.

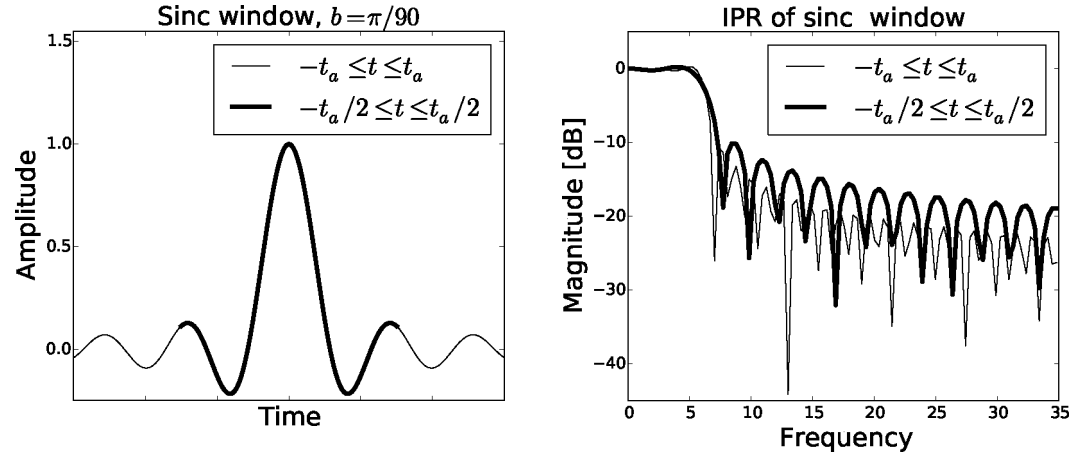

FIGURE 2.9: Sinc and its IPR.

2.3.6 Bessel J_0 window

This window function is given by a Bessel function of the first kind of order zero [Watson, 1995]. Using a power series expansion yields:

$$J_0(t) = \Pi(t/t_a) \sum_{k=0}^{\infty} \frac{(-1)^k (t/2)^{2k}}{(k!)^2}. \tag{2.20}$$

Figure 2.10 shows $J_0(t)$ and its IPR, with $J_0(t)$ truncated at time intervals $[-t_a, t_a]$ and $[-t_a/2, t_a/2]$. The performance of J_0 is somewhat worse than the sinc within the main lobe, and somewhat better in the sidelobes.

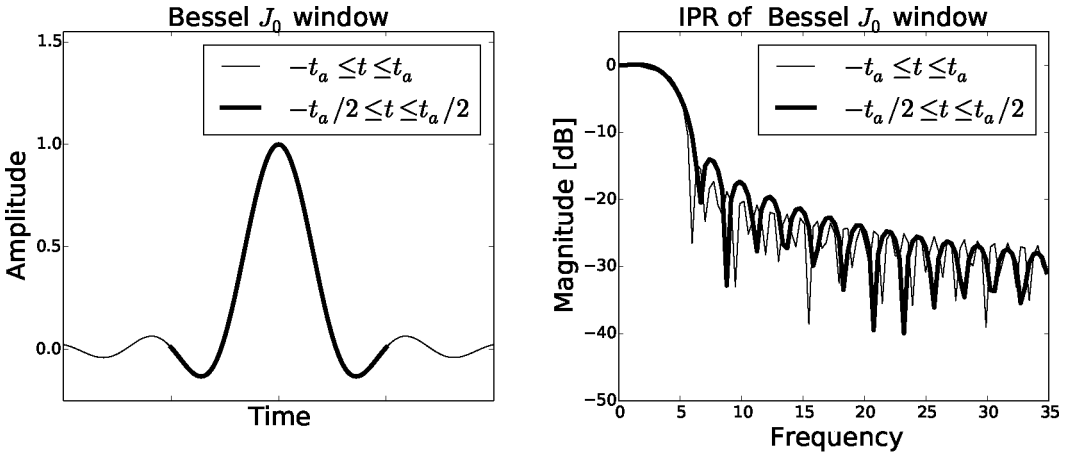

FIGURE 2.10: J_0 and its IRP.

2.3.7 Sinc-Hamming and J_0-Hamming windows

The Hamming, Han or Blackman filter is sometime multiplied by a J_0 or sinc to increase the passband and increase the stopband attenuation. The latter is defined as follows:

$$Y^{\mathrm{X}} = Y(t)X(t), \tag{2.21}$$

where $X(t)$ is the Hamming (Hm), Han (Hn) or Blackman (B) window functions, well known in signal processing literature [Nuttall and Carter, 1982, Podder et al., 2014] and Y is a $J_0(t)$ or sinc(t) window function. Figure 2.11 and 2.12 show the sinc$^{\mathrm{Hm}}$ and the J_0^{Hm} window functions and their IPRs. Compared to the sinc(t) and $J_0(t)$, they show lower peak sidelobe level and higher sidelobes roll-off.

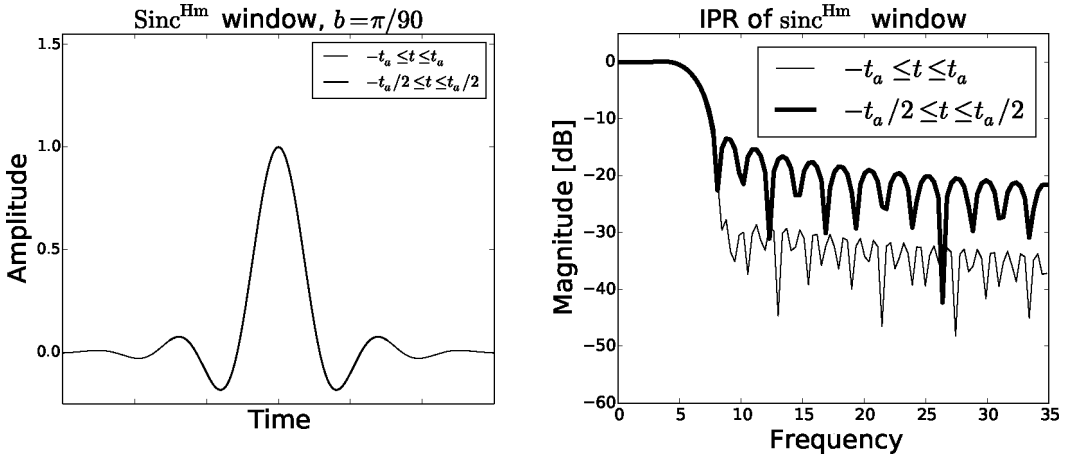

FIGURE 2.11: sinc-Hamming and its IPR.

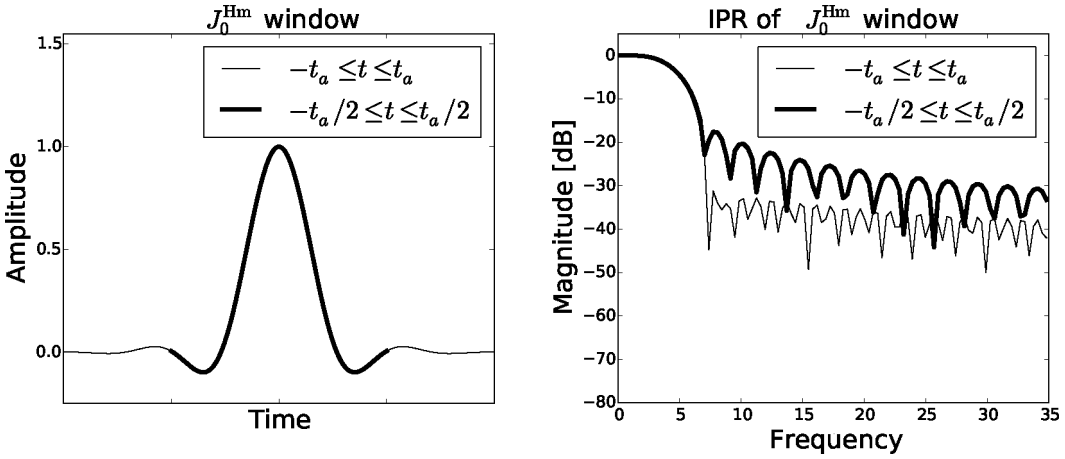

FIGURE 2.12: Bessel-Hamming and it IPR.

2.3.8 Overlapping windows and examples

This approach involves using the above windows within the resampling interval and allow their sidelobes to run into the neighborhood bands or adjacent resampling intervals. It implies that there could be more than one weighting function in a single resampling interval (e.g. see Figure 3.4). Also, Figure 2.5 to Figure 2.12 show that the longer the windows the better their IPRs. The main idea for this method is to lower the effect of truncation over smaller resampling intervals. This technique is used in several FFT convolution processors; overlapping multiple FFT windows together to average out the noise is a most obvious technique to smoothing the signal spectrum and to reduce the noise variance. One other application of overlapping window functions in signal processing is the weighted overlap add (WOLA) [Crochiere, 1980], that is used

particularly in audio compression to minimise blocking effects or audible discontinuities [Malvar and Staelin, 1989].

2.3.9 Relative performance

Table 2.2 summarises the performance of the different window functions. This table shows that the IPR of the sinc, the Bessel and all their derivatives with Hamming, Han and Blackman window functions have a larger main lobe, low peak sidelobe level and high sidelobe roll-off compared to others. Hence, they provide the most optimal IPRs. Note that Y^{X} requires two succesive visibility weightings, which may results in amplifying the thermal noise compared to the sinc or Bessel. This makes the sinc and the Bessel more suitable for this work. For this reason, we have chosen to use the sinc and the Bessel window functions to serve as the basis of BDWFs developed in the rest of this thesis. We use the following definitions to construct a 2-D sinc and Bessel from 1-D:

$$\mathrm{sinc}(u,v) = \mathrm{sinc}(\alpha u)\mathrm{sinc}(\alpha v) \tag{2.22}$$

$$J_0(u,v) = J_0(\alpha r),\ r = \sqrt{u^2+v^2}, \tag{2.23}$$

where the FoI is adjustable by the parameter α (in radians).

2.4 Conclusion

We have shown the origin of artefacts in signal processing and in radio interferometric measurements. Issues related to sampling and the FFT are discussed. We have seen that many of the artefacts are due to improper sampling that results from aliasing spectra overlapping with the physical signal. We also note the impossibility of an ideal sampling in practical problems because of finite space representation. Finally, we have reviewed some window functions and established the criteria to conserved signal within a FoI and suppression of an unwanted signal outside the FoI. We have chosen the sinc and the Bessel as the optimal windows for the design of BDWFs that we will be presenting in the next chapter.

WFs	∼MLW (at -3dB)		PSL (dB)	SLR (dB/oct)
$\Pi(t)$	$t \in \lvert t_a \rvert$	1.406	−6.663	−12.089
	$t \in \lvert t_a/2 \rvert$	2.812	−6.671	−11.065
$\mathrm{sinc}(t)$	$t \in \lvert t_a \rvert$	12.304	−10.889	−12.661
	$t \in \lvert t_a/2 \rvert$	12.304	−13.241	−11.447
$J_0(t)$	$t \in \lvert t_a \rvert$	9.140	−14.553	−12.011
	$t \in \lvert t_a/2 \rvert$	9.140	−13.614	−11.794
$\mathrm{G}(t)$	b=3	2.109	−21.535	−9.589
	b=5	2.812	−30.211	−9.091
$\mathrm{BW}(t)$	p=1	2.109	−13.718	−12.581
	p=3	4.218	−10.145	−27.330
$\mathrm{PS}_0(t)$	$\alpha = 2\pi^3$	3.515	−45.302	−7.424
	$\alpha = 5\pi^3$	4.218	−73.597	−6.375
$J_0^{\mathrm{Hm}}(t)$	$t \in \lvert t_a \rvert$	9.140	−35.724	−11.948
	$t \in \lvert t_a/2 \rvert$	9.140	−22.670	−19.527
$\mathrm{sinc}^{\mathrm{Hm}}(t)$	$t \in \lvert t_a \rvert$	12.656	−27.581	−13.817
	$t \in \lvert t_a/2 \rvert$	12.656	−13.469	−14.324

TABLE 2.2: Comparative performance of different window functions. WFs: window functions, MLW: main lobe width, PSL: peak sidelobe level, SLR: sidelobes roll-off. dB: decibel (ten times the logarithm to base ten of the window function response.)

Chapter 3

Use of Correlator Window Functions to Improve the Response Across the Field of Interest and for Data Compression

In Chapter 2 we presented a review on sampling and the Nyquist sampling conditions required to avoid aliasing. We studied some window functions used in DSP for restoration of a signal that has been distorted during sampling, and for the suppression of aliasing spectra. In this chapter, we deal with the problem of radio interferometric visibility averaging. We introduce the applicability aspects of window functions to interferometric visibilities, what we refer to as *BDWFs*, which involves two novel approaches: (i) a length scale BDWFs for interferometric data compression and FoI shaping; and (ii) a length scale overlapping BDWFs for data compression, and FoI shaping, and far field suppression. The proposed new compression schemes have been applied to simulated data from various radio interferometers such as MeerKAT (Karoo Array Telescope, precursor of the South African SKA) [Booth et al., 2009, Booth and Jonas, 2012], JVLA, NenuFAR [Tasse et al.,

This chapter draws extensively on: M. T. Atemkeng, O. M. Smirnov, C. Tasse, G. Foster and J. Jonas, *Using baseline-dependent window functions for data compression and field-of-interest shaping in radio interferometry.* MNRAS, volume 462, page 2542, 2016. It is acknowledged that some of the text in this chapter will therefore "match" that of the article. The reference in this footnote serves as a general reference for all such text.

a, Zarka et al., 2015] and the VLBI. In this chapter we present the results for the JVLA and VLBI.

3.1 Background and problem statement

Aperture synthesis tries to combine several elements of small telescopes in such a way that the combination will result in an effective synthesised instrument having larger aperture and higher resolution. Using the van Zittert-Cernike theorem and the rotation of the Earth, the synthesised instrument measures visibilities. The rest of this thesis deals with visibilities both as functions (i.e. entire distributions on the uv-plane), and single visibilities, i.e. values of those functions at a specific uv-point. To avoid confusion between functions in functional notation and their values, we will use $\mathcal{V}$ or $\mathcal{V}(u,v)$ to denote functions, and V to denote individual visibilities. Likewise, $\mathcal{I}(l,m)$ denotes a function on the lm-plane i.e. an image with l and m the angular distance from the phase centre. The symbol δ always denotes the Kronecker delta-function.

Whether considering polarisation or not, $\mathcal{V}$ can be taken to represent either scalar (complex) visibilities, or 2×2 complex visibility matrices as per the RIME formalism [Smirnov, 2011b]. Note, however, that the distinction between $\mathcal{V}$ as a complex scalar or a 2×2 complex matrix becomes important when we want to deal with each polarisation parameter separately, which is not the case in this work. Likewise, $\mathcal{I}$ can be treated as a scalar (total intensity) image, or a 2×2 brightness matrix distribution. The derivations below are valid in either case.

We shall use the symbols $\mathbf{u} = (u,v)$ or $\mathbf{u} = (u,v,w)$ to represent baseline coordinates in units of wavelength, and $\mathbf{u}^{\mathrm{m}}$ for units of metres, with $\mathbf{u} = \mathbf{u}^{\mathrm{m}}/\lambda = \mathbf{u}^{\mathrm{m}}\nu/c$.

3.1.1 Imaging, averaging and convolution

The rotation of the Earth causes the baseline to rotate in time, which we can denote by $\boldsymbol{u}^{\mathrm{m}}_{pq} = \boldsymbol{u}^{\mathrm{m}}_{pq}(t)$. The baseline in units of wavelength can be treated as a function of frequency and time (from this point onwards we shall assume that the sky is constant across the range of frequencies being observed):

$$\boldsymbol{u}_{pq}(t,\nu) = \boldsymbol{u}^{\mathrm{m}}_{pq}(t)\nu/c. \tag{3.1}$$

This, in turn, allows us to rewrite the visibility in Eq. (1.8) as a per-baseline function of t, ν:

$$V_{pq}(t,\nu) = \iint_{lm} \mathcal{I}\, \mathrm{e}^{-2\pi \mathrm{i}(u_{pq}(t,\nu)l + v_{pq}(t,\nu)m)} \mathrm{d}l\mathrm{d}m. \tag{3.2}$$

Synthesis imaging recovers the so-called "dirty image": the inverse Fourier transform of the measured visibility distribution $\mathcal{V}^{\mathrm{M}}$ sampled by a number of baselines pq at discrete time/frequency points. Inverting the Fourier transform produces the dirty image:

$$\mathcal{I}^{\mathrm{D}} = \mathcal{F}^{-1}\{\mathcal{W} \cdot \mathcal{V}^{\mathrm{M}}\}, \tag{3.3}$$

where $\mathcal{W}$ is the (weighted) sampling function over the uv-plane – a "bed-of-nails" function that is non-zero at points where we are sampling a visibility, and zero elsewhere. In the ideal case where $\mathcal{V}^{\mathrm{M}} = \mathcal{V}$ (the ideal visibility measurement, without any corruption effects, i.e. $\mathcal{F}^{-1}\{\mathcal{V}\} = \mathcal{I}$), the dirty image can also be expressed as a convolution of the apparent sky by the *point spread function* $\mathcal{P}$:

$$\mathcal{I}^{\mathrm{D}} = \mathcal{P} \circ \mathcal{I}. \tag{3.4}$$

The dirty image $\mathcal{I}^{\mathrm{D}}$ may be significantly different from the apparent sky $\mathcal{I}$ especially if the point spread function $\mathcal{P} = \mathcal{F}^{-1}\{\mathcal{W}\}$ has high sidelobe due to the coverage $\mathcal{W}$ being sparse. For this reason and throughout this chapter, we suppose an ideal sampling for the uv-distribution function, i.e. we assume a $\mathcal{P}$ that results in less imaging artefacts. Also, in this chapter, we do not treat the effect of averaging on the point spread function, leaving this to Chapter 5.

Designating each baseline as pq, and each time/frequency point as t_k, ν_l, we can represent $\mathcal{W}$ by a sum of "single-nail" functions $\mathcal{W}_{pqkl}$:

$$\mathcal{W} = \sum_{pqkl} \mathcal{W}_{pqkl} \tag{3.5}$$

$$= \sum_{pqkl} W_{pqkl} \delta_{pqkl}, \tag{3.6}$$

where δ_{pqkl} is a delta-function shifted to the uv-point being sampled:

$$\delta_{pqkl}(\boldsymbol{u}) = \delta(\boldsymbol{u} - \boldsymbol{u}_{pq}(t_k, \nu_l)) \tag{3.7}$$

and W_{pqkl} is the associated weight (refer to Briggs [1995], Yatawatta [2014] for an extensive discussion on several weighting schemes used in radio interferometric data). The Fourier transform being linear, we can rewrite Eq. (3.3) as:

$$\mathcal{I}^{\mathrm{D}} = \sum_{pqkl} W_{pqkl} \mathcal{F}^{H} \{\mathcal{V}^{\mathrm{M}}_{pqkl}\}, \tag{3.8}$$

where

$$\mathcal{V}^{\mathrm{M}}_{pqkl} = \delta_{pqkl} V^{\mathrm{M}}_{pq}(t_k, \nu_l) \tag{3.9}$$

i.e. the visibility distribution corresponding to the single visibility sample $pqkl$. We can further rewrite Eq. (3.3) again as:

$$\mathcal{I}^{\mathrm{D}} = \sum_{pqkl} W_{pqkl} \mathcal{F}^{-1} \{\mathcal{V}^{\mathrm{M}}_{pqkl}\}, \tag{3.10}$$

which shows that the dirty image $\mathcal{I}^{\mathrm{D}}$ can be seen as a weighted sum of images corresponding to the individual visibility samples $pqkl$ (each such image essentially being a single fringe pattern).

In the ideal case, we would be measuring instantaneous visibility samples, and (assuming no other instrumental corruptions), we would have $\mathcal{V}^{\mathrm{M}} \equiv \mathcal{V}$, with

$$V^{\mathrm{M}}_{pq}(t_k, \nu_l) = \mathcal{V}(\boldsymbol{u}_{pq}(t_k, \nu_l)), \tag{3.11}$$

and consequently,

$$\mathcal{V}^{\mathrm{M}}_{pqkl} = \delta_{pqkl} \mathcal{V}, \tag{3.12}$$

resulting in what we will call the *ideal* dirty image $\mathcal{I}^{\mathrm{DI}}$:

$$\mathcal{I}^{\mathrm{DI}} = \sum_{pqkl} W_{pqkl} \mathcal{P}_{pqkl} \circ \mathcal{I}, \quad \mathcal{P}_{pqkl} = \mathcal{F}^{-1}\{\delta_{pqkl}\}. \tag{3.13}$$

That is, in the ideal case, each term in the weighted sum is equal to the apparent sky $\mathcal{I}$ convolved with a PSF representing a single visibility sample, $\mathcal{P}_{pqkl}$.

However, an actual interferometer is necessarily non-ideal, in that it can only measure the average visibility over some time-frequency bin given by the *time* and *frequency sampling*

intervals $\Delta t, \Delta\nu$, which we will call the *sampling bin*:

$$\mathsf{B}_{kl}^{[\Delta t\Delta\nu]} = \left[t_k - \frac{\Delta t}{2}, t_k + \frac{\Delta t}{2}\right] \times \left[\nu_l - \frac{\Delta\nu}{2}, \nu_l + \frac{\Delta\nu}{2}\right]. \tag{3.14}$$

This measurement can be represented by an integration:

$$V_{pqkl}^{\mathrm{M}} = \frac{1}{\Delta t\Delta\nu} \iint\limits_{\mathsf{B}_{kl}^{[\Delta t\Delta\nu]}} \mathcal{V}(\boldsymbol{u}_{pq}(t,\nu))\mathrm{d}\nu\mathrm{d}t. \tag{3.15}$$

Inverting the relation of Eq. (3.1), we can change variables to express this as an integration over the corresponding bin $\mathsf{B}_{pqkl}^{[uv]}$ in uv-space:

$$V_{pqkl}^{\mathrm{M}} = \frac{1}{\Delta t\Delta\nu} \iint\limits_{\mathsf{B}_{pqkl}^{[uv]}} \mathcal{V}_{pq}(u,v)\left|\frac{\partial(t,\nu)}{\partial(u,v)}\right|\mathrm{d}u\mathrm{d}v, \tag{3.16}$$

where $\mathsf{B}_{pqkl}^{[uv]}$ is the corresponding bin in uv-space. Note that the sampling bins in $t\nu$-space are perfectly rectangular and do not depend on baseline (assuming baseline-independent averaging), while the sampling bins in uv-space are bounded by sections of elliptical arcs, and do depend on baseline (hence the extra pq index). Figure 3.1 illustrates these two aspects.

Assuming a bin small enough that the fringe rate $\partial\boldsymbol{u}/\partial t$ is approximately constant over the bin, we then have:

$$V_{pqkl}^{\mathrm{M}} \sim \iint\limits_{\mathsf{B}_{pqkl}^{[uv]}} \mathcal{V}(\boldsymbol{u})\mathrm{d}\boldsymbol{u}. \tag{3.17}$$

Now, let us introduce a *normalised boxcar window function*, $\Pi^{[t\nu]}$

$$\Pi^{[t\nu]}(t,\nu) = \begin{cases} \frac{1}{\Delta t\Delta\nu}, & |t| \leq \Delta t/2, \quad |\nu| \leq \Delta\nu/2 \\ 0, & \text{otherwise,} \end{cases} \tag{3.18}$$

using which we may re-write Eq. (3.15) as:

$$V_{pqkl}^{\mathrm{M}} = \iint\limits_{\infty} \mathcal{V}_{pq}(t,\nu)\Pi^{[t\nu]}(t-t_k,\nu-\nu_l)\mathrm{d}t\mathrm{d}\nu, \tag{3.19}$$

which can also be expressed as a convolution in $t\nu$-space:

$$V^{\mathrm{M}}_{pqkl} = [\mathcal{V}_{pq} \circ \Pi^{[t\nu]}](t_k, \nu_l). \tag{3.20}$$

Eq. (3.16) can also be rewritten as

$$V^{\mathrm{M}}_{pqkl} = \iint_{\infty} \mathcal{V}_{pq}\big(\boldsymbol{u}_{pq}(t,\nu)\big)\Pi^{[uv]}_{pqkl}\big(\boldsymbol{u}_{pq}(t,\nu) - \boldsymbol{u}_{pq}(t_k,\nu_l)\big)\mathrm{d}u\mathrm{d}v, \tag{3.21}$$

which is a convolution in uv-space:

$$V^{\mathrm{M}}_{pqkl} = [\mathcal{V}_{pq} \circ \Pi^{[uv]}_{pqkl}](\boldsymbol{u}_{pq}(t_k,\nu_l)), \tag{3.22}$$

where $\Pi^{[uv]}_{pqkl}$ is a normalised boxcar-like window function that corresponds to bin $\mathsf{B}^{[uv]}_{pqkl}$ in uv-space and also includes the determinant term of Eq. (3.16), i.e.

$$\Pi^{[uv]}_{pqkl}(\boldsymbol{u}_{pq}(t,\nu)) = \begin{cases} \frac{1}{\Delta \boldsymbol{u}_{pq}} \left|\frac{\partial(t,\nu)}{\partial(u,v)}\right|, & \boldsymbol{u}_{pq} \in \mathsf{B}^{[uv]}_{pqkl} \\ 0, & \text{otherwise.} \end{cases} \tag{3.23}$$

Eq. (3.20) and (3.22) make it explicit that each averaged visibility is drawn from a convolution of the underlying visibilities with a boxcar-like window function.

Note what Eq. (3.22) does and does not say. It does say that each individual averaged visibility corresponds to convolving the true visibilities by some window function. However, this window function is different for each baseline pq and time/frequency sample t_k, ν_l (which is emphasised by the subscripts to $\Pi^{[uv]}$ in the equations above). Averaging is thus not a "true" convolution, since the convolution kernel changes at every point in the uv-plane. We shall call this process a *pseudo-convolution*, and the kernel being convolved with ($\Pi^{[uv]}_{pqkl}$) an example of a BDWF. In subsequent sections we will explore alternative BDWFs.

In actual fact, a correlator (or an averaging operation in post-processing) deals with averages of discrete and noisy samples, rather than a continuous integration. Ignoring the complexities of correlator implementation, let us cast this process in terms of a simple averaging operation. That is, assume we have a set of *high-res* or *sampled visibilities* on a high-resolution time/frequency grid t_i, ν_j:

$$V^{(\mathrm{s})}_{pqij} = \mathcal{V}_{pq}(t_i, \nu_j) + \mathcal{N}[\sigma^{(\mathrm{s})}_{pqij}], \tag{3.24}$$

where $\mathcal{V}_{pq}$ is given by Eq. (3.2), and $\mathcal{N}$ represents the visibility noise term, which is a complex scalar or complex 2×2 matrix with the real and imaginary parts being independently drawn from a zero-mean normal distribution with the indicated r.m.s. [Wrobel and Walker, 1999]. The noise term is not correlated across samples. The *low-res* or *averaged* or *resampled* visibilities are then a discrete sum:

$$V^{\mathrm{M}}_{pqkl} = \frac{1}{n}\sum_{ij\in\mathsf{B}_{kl}} V^{(\mathrm{s})}_{pqij}, \tag{3.25}$$

where B_{kl} is the set of sample indices ij corresponding to the *resampling bin*, i.e.

$$\mathsf{B}_{kl} = \left\{ij:\ t_i\nu_j \in \mathsf{B}^{[\Delta t\Delta\nu]}_{kl}\right\}, \tag{3.26}$$

and $n = n_t \times n_\nu$ is the number of samples in the bin. Using the BDWF definitions above, this becomes a conventional discrete convolution (assuming a regular $t\nu$-grid):

$$V^{\mathrm{M}}_{pqkl} = \sum_{i,j=-\infty}^{\infty} V^{(\mathrm{s})}_{pqij}\Pi^{[t\nu]}(t_i - t_k, \nu_j - \nu_l). \tag{3.27}$$

In uv-space, this becomes a discrete convolution on an irregular grid:

$$V^{\mathrm{M}}_{pqkl} = \sum_{i,j=-\infty}^{\infty} V^{(\mathrm{s})}_{pqij}\Pi^{[uv]}_{pqkl}(\boldsymbol{u}_{ij} - \boldsymbol{u}_{kl}). \tag{3.28}$$

The $\boldsymbol{u}_{ij}$ grid being schematically illustrated by Figure 3.1 (left) shows a snapshot uv-coverage for a single integration time interval. We note here that for shorter baselines the uv-tracks are closer to the centre of rotation, while for longer baselines the tracks are further away from the rotation centre. Each dot mark represents a uv-bin. It is clearly visible on this figure that the space between two successive uv-bins on long baselines is wider compared to the one on short baselines.

3.1.2 Effect of averaging on the image

As currently implemented in radio interferometric correlators, to reduce the effects of averaging the visibilities, the synthesis time and the frequency range must be broken up into a number of finite small averaging intervals. For each time and frequency interval, correlation is then treated as a distinct visibility, each with its own uv-bin. For example,

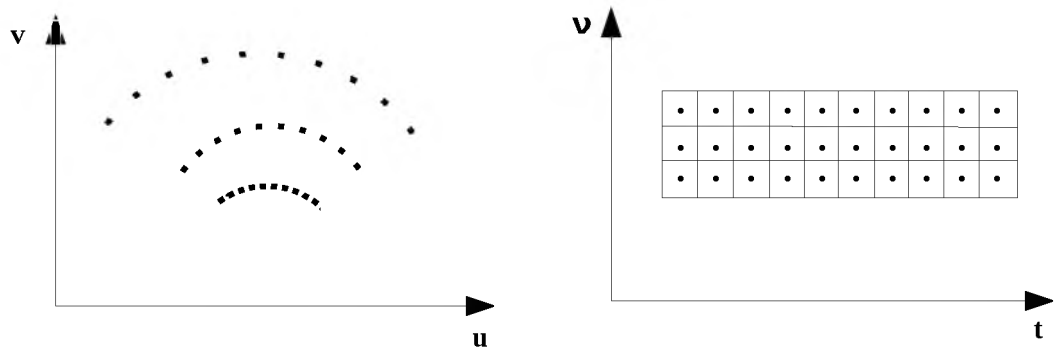

FIGURE 3.1: Schematic of uv-coverage for regularly spaced time-frequency samples, left: uv-space, right: $t\nu$-space. Baselines with a longer East-West component sweep out longer tracks between successive integrations.

if the synthesis time is T_{obs} and the frequency range is B_w, then $(T_{obs}/\Delta t)\times(B_w/\Delta\nu)$ will be the number of time-frequency correlations on each baseline. Under this for current correlators, $\{\Delta t, \Delta\nu\} \rightarrow \{0, 0\}$ shows that averaging becomes equivalent to sampling while on the other hand $\{T_{obs}, B_w\} \propto \{T_{obs}/\Delta t, B_w/\Delta\nu\}$ leads to a large data set and therefore storage issues. The Fourier phase component $2\pi\phi(u, v, w)$ is a function of frequency and time, with increasing variation over the averaging interval for sources far from the phase centre. The average of a complex quantity with a varying phase then effectively "washes out" amplitude, the effect being especially severe for wide FoIs [for an extensive discussion, see Bregman, 2012, Bridle and Schwab, 1999, 1989, Thompson et al., 2008]. In the uv-plane, this effect is often referred to as *time* and *bandwidth decorrelation*, and *smearing* in the image-plane.

The discussion above provides an alternative way to look at decorrelation/smearing. With averaging in effect, the relationship between the measured and the ideal visibility changes to (contrast this to Eq. (3.12)):

$$\mathcal{V}^{\mathrm{M}}_{pqkl} = \delta_{pqkl}(\mathcal{V} \circ \Pi^{[uv]}_{pqkl}). \tag{3.29}$$

Combining this with Eq. (3.10), and using the Fourier convolution theorem, we can see that the dirty image is formed as:

$$\mathcal{I}^{\mathrm{D}} = \sum_{pqkl} W_{pqkl}\mathcal{P}_{pqkl} \circ (\mathcal{I} \cdot \mathcal{T}_{pqkl}), \tag{3.30}$$

with the apparent sky $\mathcal{I}$ now tapered by the baseline-dependent *window response function* $\mathcal{T}_{pqkl}$, the latter being the inverse Fourier transform of the BDWF:

$$\mathcal{T}_{pqkl} = \mathcal{F}^{-1}\{\Pi^{[uv]}_{pqkl}\}. \tag{3.31}$$

In other words, the dirty image yielded by averaged visibilities (compare this to the ideal dirty image given by Eq. (3.13)) is a weighted average of per-visibility dirty images corresponding to a per-visibility tapered sky. The Fourier transform of a "boxcar-like" function is a "sinc-like" function, schematically illustrated in one dimension by Figure 3.2 (right). Time and bandwidth smearing represents the average effect of all these individual tapers or IPRs. Shorter baselines correspond to smaller boxcars and wider IPRs, longer baselines to larger boxcars and narrower IPRs and are thus more prone to smearing.

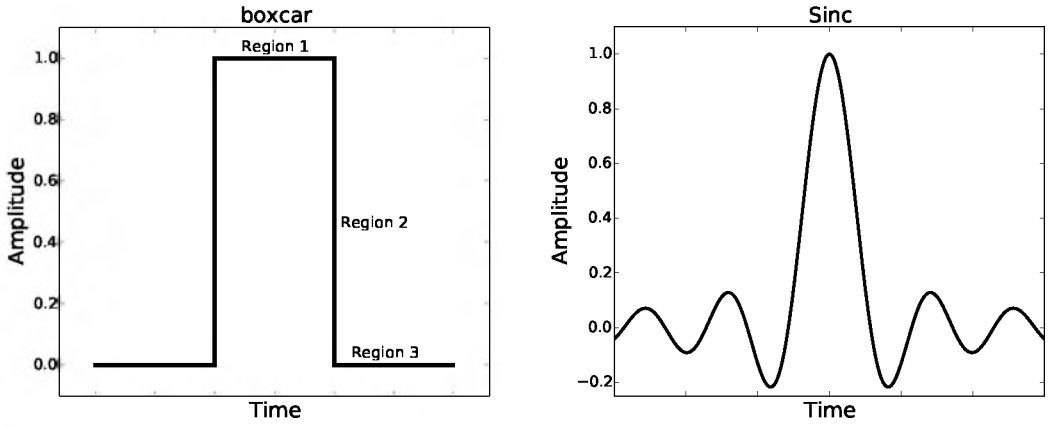

FIGURE 3.2: Left: Boxcar response. In the uv-plane, this represents the window function corresponding to simple averaging of visibilities. In the image-plane, this represents the ideal IPR function. Right: Sinc response. In the image-plane, this represents the window response function corresponding to a boxcar window function in the uv-plane. In the uv-plane, this represents the ideal window function.

3.1.3 The case for alternative BDWFs

The window response function induced by simple averaging (Figure 3.2, right) is far from ideal: it either suppresses too much within the FoI, or too little outside the FoI, or both. The optimal IPR would be boxcar-like, as in Figure 3.2 (left). The BDWF that would produce such a response is sinc-like, as in Figure 3.2 (right). The problem with a sinc is that it has infinite support; applying it over finite-sized bins necessarily means a *truncated* BDWF that results in a sub-optimal IPR. The problem of optimal filtering

has been well studied in signal processing (usually assuming a true convolution rather than the pseudo-convolution we deal with here), and we shall apply these lessons below.

The derivations above make it clear that using a different BDWF in place of the conventional boxcar-like $\Pi^{[uv]}$ could in principle yield a more optimal IPR. The obvious disadvantage is a loss in sensitivity. Each visibility sample is subject to an independent Gaussian noise term in the real and imaginary part; the noise of the average of a set of samples is minimised when the average is naturally weighted (or unweighted, if the noise is constant across visibilities). Thus, any deviation from a boxcar window function must necessarily increase the noise in the visibilities. Below we will study this effect both theoretically and via simulations, to establish whether this trade-off is sensible, and under which conditions.

3.2 Applying window functions to visibilities

While visibilities are (usually) regularly sampled in $t\nu$-space, in uv^{m}-space this is not so. In frequency, the sampling positions go as $\sim \lambda^{-1}$, while in time, baselines with a longer East-West component sweep out longer tracks between successive integrations (Figure 3.1). Applying a window function with a constant integration window in $t\nu$-space corresponds to different-sized windows in uv-space. In the case of simple averaging, this results in the boxcar-like window $\Pi^{[uv]}$ of Eq. (3.22) having a baseline-dependent scale. The scale of the IPR being inversely proportional to the scale of the window function, this results in more decorrelation (i.e. a narrower FoI) on longer baselines.

By defining our alternative window functions in uv-space (in units of wavelength), we can attempt to "even out" the decorrelation response across baselines. For a given BDWF $X(u, v)$, we have the following recipe for computing resampled visibilities (compare to Eq. (3.28)):

$$V^{\mathrm{M}}_{pqkl} = \frac{\sum\limits_{i,j \in \mathsf{B}_{kl}} V^{(\mathrm{s})}_{pqij} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}{\sum\limits_{i,j \in \mathsf{B}_{kl}} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}, \tag{3.32}$$

where $\boldsymbol{u}_{pqkl}$ is the midpoint of the resampling bin B_{kl} in uv-space. The main lobe of the window function then has the same scale across the entire uv-plane, while the resampling bins have different uv-sizes. Conversely, in $t\nu$-space the bins are regular,

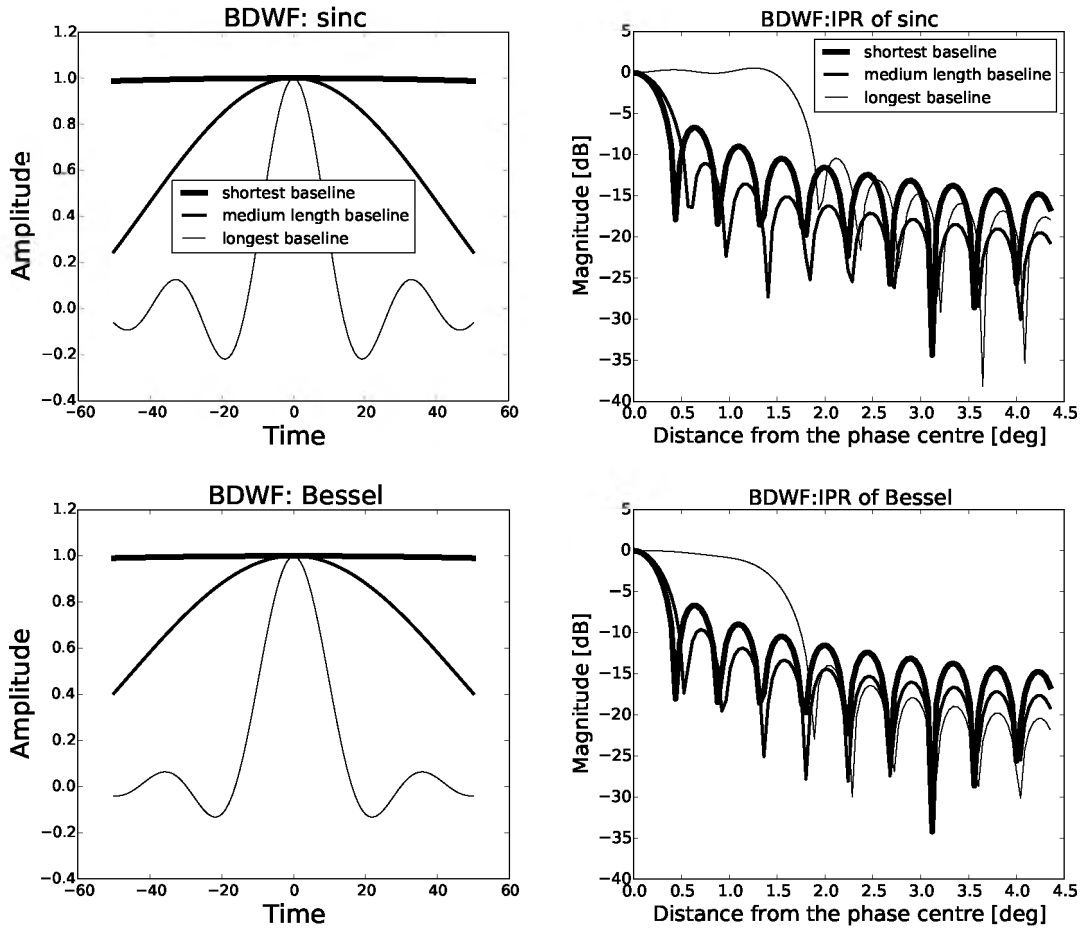

FIGURE 3.3: Cross-sections through two different BDWFs (top left: sinc, bottom left: Bessel) defined in uv-space, plotted along the time axis and their IPR (top right: IPR for sinc, bottom right: IPR for Bessel). This shows that the effective window function is a scaling and truncation of the underlying window function, with the shortest baselines reducing to a boxcar-like window function.

while the main lobe of the effective window function scales inversely with the baseline fringe rate. Furthermore, the window function is truncated at the edge of each bin; on the shortest baselines this truncation is extreme to the point of approaching the boxcar-like $\Pi^{[uv]}$ (Figure 3.3).

The downside of this simple approach is twofold. Firstly, while all the window functions above nominally exhibit far lower sidelobes than the boxcar (i.e. more suppression for out-of-FoI sources), they no longer perform that well under truncation, with extremely truncated window functions at the shorter baselines becoming boxcar-like. Secondly, taking a weighted sum in Eq. (3.32) increases the noise in comparison to simple averaging.

3.2.1 Overlapping BDWFs

A more sophisticated approach involves overlapping BDWFs. Simple averaging implicitly assumes that the resampling bins in Eq. (3.32) do not overlap for adjacent kl, since they represent adjacent averaging intervals. There is, however, no reason (apart from computational load) why we cannot take the sum in Eq. (3.32) over larger bins. Let us define the *window bin* for *overlap factors* of α, β as:

$$\mathsf{B}_{kl}^{[\alpha\beta]} = \left\{ij : \ t_i\nu_j \in \mathsf{B}_{kl}^{[\alpha\Delta t, \beta\Delta\nu]}\right\}, \tag{3.33}$$

i.e. as the set of sample indices corresponding to a bin of size $\alpha\Delta t \times \beta\Delta\nu$ in $t\nu$-space. Let us then compute the sums in Eq. (3.32) over the window bin. This becomes distinct from the resampling bin: while the latter represents the spacing of the resampled visibilities, the former represents the size of the window over which the convolution is computed. Only for $\alpha = \beta = 1$ do the two bins become the same.

In the overlapping regime, the baseline-dependent weight for a single visibility is not defined by a unique BDWF, but by the strength of the correlation between the overall overlapping BDWFs with the visibility. BDWFs in the overlapping regime are schematically illustrated in Figure 3.4. For simple averaging overlapping offers no benefit, since it only broadens $\Pi^{[uv]}$ and therefore increases smearing, but for a well-behaved BDWF, enlarging the window bin (while maintaining the same window function scale) means less truncation – thus lower sidelobes – and decreased noise, as more sampled visibilities are taken into account. On longer baselines, the IPR of a well-behaved overlapping BDWF means less smearing in the FoI and excellent out-of-FoI suppression (see Figure 3.5). On the shorter baselines, a well-behaved BDWF is equivalent to a boxcar and therefore enlarges the window size by overlapping results in decreasing the noise (see Section 3.2.2).

To distinguish overlapping BDWFs from non-overlapping ones, in the rest of this chapter we will designate the window functions employed as *WF*-$\alpha\times\beta$. For example, sinc-3×2, J_0-1×1 (i.e. no overlap), etc. If resampling is only done in one direction (only time or only frequency), we will indicate this as e.g. J_0-3×−.

As mentioned above, the computational load for overlapping window functions will increase linearly as the factors α and β increase. For a given resampling bin $\mathsf{B}_{kl}^{[\alpha\beta]}$ the

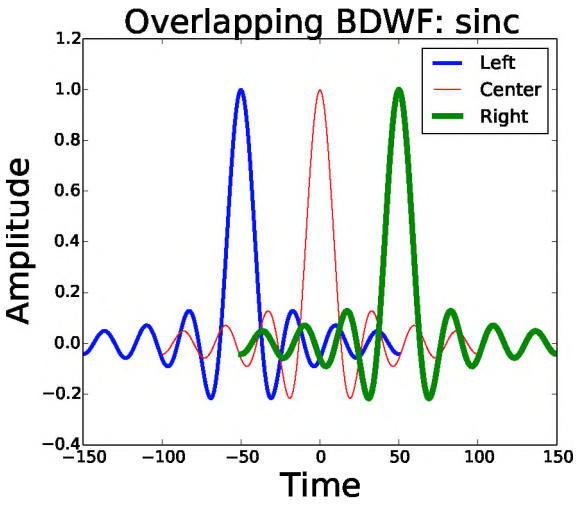

FIGURE 3.4: Overlapping BDWFs representing adjacent resampling bins (right resampling bin, centre resampling bin and left resampling bin). This corresponds to overlap factor $\alpha = 3$ in time.

computational complexity would scale as $\mathcal{O}(\alpha\beta n_t n_\nu)$ compared to $\mathcal{O}(n_t n_\nu)$, which is the complexity for applying a non-overlapping window function to visibilities within the resampling bin $\mathsf{B}^{[\Delta t \Delta \nu]}$. In this notation, the resampling bin for a non-overlapping window functions consists of $n_t \times n_\nu$ samples, which is the number of multiplications between the baseline-dependent weights and the visibilities.

Below we provide an alternative way to look at the overlaping BDWFs applied to visibilities. Let us reconsider that X is a BDWF with its appropriate resampling bin $\mathsf{B}^{[\Delta t \Delta \nu]}_{kl}$. Now, suppose that $\Delta^{\text{lft}}t$ and $\Delta^{\text{lft}}\nu$ are the overlap time-frequency sampling intervals associated with Δt and $\Delta \nu$ on their *left-hand side*. Similarly, $\Delta^{\text{rgt}}t$ and $\Delta^{\text{rgt}}\nu$ are the overlap time-frequency sampling intervals associated with Δt and $\Delta \nu$ on their *right-hand side*. The overlapping BDWFs sampling intervals are then given by:

$$\Delta^{\text{olp}}t = \Delta^{\text{lft}}t \cup \Delta t \cup \Delta^{\text{rgt}}t \tag{3.34}$$

$$\Delta^{\text{olp}}\nu = \Delta^{\text{lft}}\nu \cup \Delta \nu \cup \Delta^{\text{lft}}\nu. \tag{3.35}$$

The overlap sampling bin is now defined as:

$$\mathsf{B}^{[\Delta^{\text{olp}}t, \Delta^{\text{olp}}\nu]}_{kl} = \left[t_k - \frac{\Delta^{\text{olp}}t}{2}, t_k + \frac{\Delta^{\text{olp}}t}{2}\right] \times \left[\nu_l - \frac{\Delta^{\text{olp}}\nu}{2}, \nu_l + \frac{\Delta^{\text{olp}}\nu}{2}\right] \tag{3.36}$$

$$= \mathsf{B}^{\text{lft}}_{kl} \cup \mathsf{B}^{[\Delta t \Delta \nu]}_{lk} \cup \mathsf{B}^{\text{rgt}}_{kl}, \tag{3.37}$$

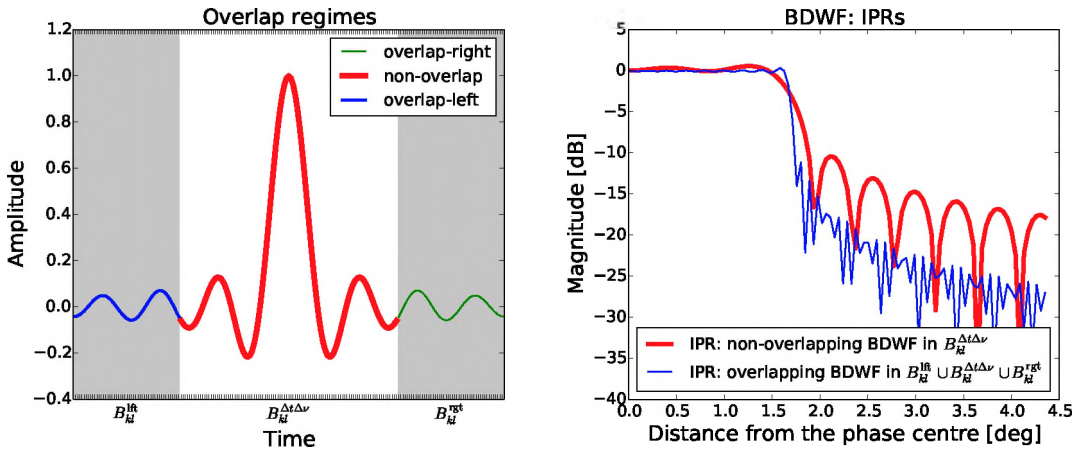

FIGURE 3.5: Overlapping BDWF. (Left) Blue curve: left-hand overlap function, green curve: right-hand overlap function, red curve: non-overlap function. (Right) The IPRs of an overllaping (blue) and non-overlapping (red) BDFW.

where $\mathsf{B}_{kl}^{\rm lft}$ and $\mathsf{B}_{kl}^{\rm rgt}$ are the set of time-frequency samples in the overlap regimes, and $\mathsf{B}_{kl}^{[\Delta t\Delta\nu]}$ is the resampling bin for a non-overlapping BDWFs defined in Eq. (3.14). Figure 3.5 (left) displays the time direction overlap regimes (i.e. $\mathsf{B}_{kl}^{\rm lft}$ and $\mathsf{B}_{kl}^{\rm rgt}$). We note that the window bin defined in Eq. (3.33) for overlap factors of α, β is equivalent to

$$\mathsf{B}_{kl}^{[\alpha\beta]} = \left\{ij : \ t_i\nu_j \in \mathsf{B}_{kl}^{[\Delta^{\rm olp}t, \Delta^{\rm olp}\nu]}\right\} \tag{3.38}$$

or in more detail we can write:

$$\mathsf{B}_{kl}^{[\alpha\beta]} = \left\{ij : \ t_i\nu_j \in \mathsf{B}_{kl}^{\rm lft}\right\} \cup \left\{ij : \ t_i\nu_j \in \mathsf{B}_{lk}^{[\Delta t\Delta\nu]}\right\} \cup \left\{ij : \ t_i\nu_j \in \mathsf{B}_{kl}^{\rm rgt}\right\} \tag{3.39}$$

$$= S_1 \cup \mathsf{B}_{kl} \cup S_2, \tag{3.40}$$

where $S_1 = \left\{ij : \ t_i\nu_j \in \mathsf{B}_{kl}^{\rm lft}\right\}$ and $S_2 = \left\{ij : \ t_i\nu_j \in \mathsf{B}_{kl}^{\rm rgt}\right\}$ are the number of time-frequency samples in the left-hand and right-hand overlap regimes respectively. The resampled visibilities for overlapping BDWFs (compare to Eq. (3.32)) are as follows:

$$V_{pqkl}^{\rm M} = \frac{\sum\limits_{i,j\in\mathsf{B}_{kl}^{[\alpha\beta]}} V_{pqij}^{(\rm s)} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}{\sum\limits_{i,j\in\mathsf{B}_{kl}^{[\alpha\beta]}} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}. \tag{3.41}$$

One may still decompose Eq. (3.41) as:

$$V^{\mathrm{M}}_{pqkl} = \frac{\sum\limits_{i,j\in S_1} V^{(\mathrm{s})}_{pqij} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl}) + \sum\limits_{i,j\in \mathsf{B}_{kl}} V^{(\mathrm{s})}_{pqij} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl}) + \sum\limits_{i,j\in S_2} V^{(\mathrm{s})}_{pqij} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}{\sum\limits_{i,j\in S_1} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl}) + \sum\limits_{i,j\in \mathsf{B}_{kl}} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl}) + \sum\limits_{i,j\in S_2} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}. \tag{3.42}$$

Figure 3.5 shows an overlapping BDWF $X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})$ with $ij \in S_1$ (or $t_i\nu_i \in \mathsf{B}^{\mathrm{lft}}_{kl}$), $X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})$ with $ij \in \mathsf{B}_{kl}$ (or $t_i\nu_i \in \mathsf{B}^{[\Delta t\Delta\nu]}_{lk}$) and $X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})$ with $ij \in S_2$ (or $t_i\nu_j \in \mathsf{B}^{\mathrm{rgt}}_{lk}$). We note that Eq. (3.42) becomes equivalent to the non-overlapping resampled visibilities in Eq. (3.32) when $S_1 = S_2 = \emptyset$. This is due to the sum over an empty set $\emptyset$:

$$\begin{cases} \sum\limits_{i,j\in S_1} V^{(\mathrm{s})}_{pqij} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl}) = 0, \\ \sum\limits_{i,j\in S_2} V^{(\mathrm{s})}_{pqij} X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl}) = 0. \end{cases} \tag{3.43}$$

3.2.2 Noise penalty estimates: analytic

Let us now work out analytically the *noise penalty* associated with replacing an unweighted average by a weighted sum. For simplicity, let us assume that the noise term has constant r.m.s. σ_{s} across all baselines and samples. If the resampling bin consists of $n_{\mathrm{avg}} = n_t \times n_\nu$ samples, and since the noise is not correlated between samples, the noise on the averaged visibilities in Eq. (3.25) will be given by:

$$\sigma^2_{\mathrm{m}} = \frac{1}{n^2_{\mathrm{avg}}} \sum_{i=1}^{n_{\mathrm{avg}}} \sigma^2_{\mathrm{s}} \tag{3.44}$$

$$= \frac{\sigma^2_{\mathrm{s}}}{n_{\mathrm{avg}}}. \tag{3.45}$$

Note that the noise is uncorrelated across averaged visibilities. We can therefore use the the imaging equation (i.e. Eq. (3.10)) to derive the following expression for the variance of the noise term in each pixel of the dirty image:

$$\sigma^2_{\mathrm{pix}} = \frac{(\sum_{pqkl} W^2_{pqkl}\sigma^2_{\mathrm{m}})}{(\sum_{pqkl} W_{pqkl})^2}, \tag{3.46}$$

which for natural image weighting ($W \sim \sigma_{\mathrm{m}}^{-1}$, i.e. $W \equiv 1$ in this case) is simply

$$\sigma_{\mathrm{pix}}^2 = \frac{1}{N}\frac{\sigma_{\mathrm{s}}^2}{n_{\mathrm{avg}}}, \tag{3.47}$$

where N is the total number of visibilities used for the synthesis.

To simplify further notation, let us replace $pqkl$ by a single index μ, enumerating all the low-res visibilities V_μ^{M}, with $\mu = 1 \ldots N$. If we now employ Eq. (3.32) to compute the low-res visibilities using some BDWF $X(u, v)$, the noise term becomes different for each visibility μ:

$$\sigma_{X\mu}^2 = \frac{\sum X^2(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}{\left[\sum X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})\right]^2}\,\sigma_{\mathrm{s}}^2, \tag{3.48}$$

where both sums are taken over the window bin, $i, j \in \mathsf{B}_{kl}$. Let us define the *visibility noise penalty* associated with BDWF X and visibility μ as the relative increase in noise over the unweighted average, i.e.

$$\Xi_{X\mu} = \frac{\sigma_{X\mu}}{\sigma_{\mathrm{m}}} = \frac{\sqrt{n_{\mathrm{avg}} \sum X^2(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}}{\sum X(\boldsymbol{u}_{pqij} - \boldsymbol{u}_{pqkl})}. \tag{3.49}$$

Note that in the case of overlapping BDWFs, the window bin in Eq. (3.48) is larger than the resampling bin, and contains n_X samples, with $n_X = \alpha\beta n_{\mathrm{avg}}$, where α and β are the overlap factors. For $\alpha = \beta = 1$, it is easy to see that $\Xi_{X\mu} \geq 1$, and only reaches 1 when $X \equiv 1$. In other words, non-overlapping BDWFs always result in a visibility noise penalty above 1, while overlapping BDWFs can actually *reduce* noise in the resampled visibilities.

While paradoxical at first glance, this reduction in noise does **not** result in a net gain in image sensitivity. The reason for this is that with overlap in effect, the noise terms become correlated across resampled visibilities kl (within the same baseline pq), with each high-res visibility sample contributing to multiple resampled visibilities, and the image noise term no longer follows Eq. (3.46).

If the resampled visibilities correspond to a single-channel snapshot, or if the BDWFs are non-overlapping, then the noise across visibilities remains uncorrelated, and we can

compute the *image noise penalty* associated with imaging weights W and BDWF X as

$$\Xi_X^W = \frac{\sigma_{\mathrm{pix},X}^2}{\sigma_{\mathrm{pix}}} \tag{3.50}$$

$$= \frac{n}{\sigma_{\mathrm{s}}} \frac{(\sum_\mu W_\mu^2 \sigma_\mu^2)}{(\sum_\mu W_\mu)^2} \tag{3.51}$$

$$= \frac{(\sum_\mu W_\mu^2 \Xi_{X\mu}^2)}{(\sum_\mu W_\mu)^2}. \tag{3.52}$$

In the case of natural weighting ($W_\mu = \sigma_\mu^{-1}$) this reduces to:

$$\Xi_X^{\mathrm{nat}} = \frac{N}{\sum_\mu \Xi_{X\mu}^{-1}}. \tag{3.53}$$

Alternatively one can formally express the noise penalty in a compact and robust matrix formalism as shown below. If we assume a non-polarised sky, a single channel timeslot visibility measured by a baseline pq is the complex Stokes value:

$$V_{pqij}^s = W_{pqij} V_{pqij}, \tag{3.54}$$

where V_{pqij} is an individual visibility (or a value sampled at ij of the ideal sky distribution Fourier component $\mathcal{V}_{pqij} = \delta_{pqij}\mathcal{V}$) and W_{pqij} accounts for weighting. Assuming a resampling square interval of size $n_t \times n_v$, we can package into a single matrix, $\boldsymbol{V}_{pq,d}^s$ of size $n_t \times n_\nu$ the sampled visibilities,

$$\boldsymbol{V}_{pq,d}^s = \begin{bmatrix} V_{pq11}^s & \cdots & V_{pq1n_\nu}^s \\ \vdots & \cdots & \vdots \\ V_{pqn_t1}^s & \cdots & V_{pqn_tn_\nu}^s \end{bmatrix} \tag{3.55}$$

where the index d indicates the d^{th} resampling bin. More intuitively, d will run from $1 \cdots N_t N_\nu$, where N_t and N_ν are the number of timeslots and channels of the low-res measurement respectively. This gives a total high-res timeslots of $n_t N_t$ and $n_\nu N_\nu$ channels.

On the other hand, the boxcar-like window function values for the sampled visibilities are grouped in the matrix, $\boldsymbol{\Pi}_{pqkl,d}^{[uv]}$ of size $n_t \times n_\nu$. We recall from the previous section that $kl \in B_{kl}$ and the entries of $\boldsymbol{\Pi}_{pqkl,d}^{[uv]}$ are the $(ij)^{th}$ value of the boxcar-like corresponding

to the $(ij)^{th}$ visibility point. This can be summarised explicitly as below:

$$\boldsymbol{\Pi}^{[uv]}_{pqkl,d} = \begin{bmatrix} \overbrace{\Pi^{[uv]}_{pqkl}}^{11} & \cdots & \overbrace{\Pi^{[uv]}_{pqkl}}^{1n_v} \\ \vdots & \cdots & \vdots \\ \overbrace{\Pi^{[uv]}_{pqkl}}^{n_t 1} & \cdots & \overbrace{\Pi^{[uv]}_{pqkl}}^{n_t n_v} \end{bmatrix} \tag{3.56}$$

We can write the average value in terms of series of linear transformations

$$V^{\mathrm{M}}_{pqkl} = \boldsymbol{e}^T \left[\boldsymbol{\Pi}^{[uv]}_{pqkl,d} \boldsymbol{V}^s_{pq,d}\right] \boldsymbol{e}, \tag{3.57}$$

where $\boldsymbol{e}$ is the column vector whose entries are all unity and ^{T} the transpose matrix operator.

Following a similar analogy, the total number of visibilities of the high-res measurement is packaged into a block-diagonal matrix $\boldsymbol{V}^s_\mu$ of size $N \times [n_t \times n_\nu] \times N$ as follows:

$$\boldsymbol{V}^s_\mu = \begin{bmatrix} \boldsymbol{V}^s_{pq,1} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \boldsymbol{V}^s_{pq,\frac{N}{2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & \boldsymbol{V}^s_{pq,N} \end{bmatrix} \tag{3.58}$$

where N is the total number of visibilities for the low-res measurement with $N = N_{pq} N_t N_\nu$, N_{pq} is the number of baselines, excluding autocorrelations ($p = 1 \cdots n_a$ and $q = p + 1 \cdots n_a$, with n_a the number of antennas).

Similarly, $\boldsymbol{\Pi}^{[uv]}_{pqkl,d}$ are regrouped into a block-diagonal matrix $\boldsymbol{\Pi}^{[uv]}_\mu$ of size $N \times [n_t \times n_\nu] \times N$. Therefore, we can write

$$\boldsymbol{V}^{\mathrm{M}}_\mu = \boldsymbol{e}_{\mathrm{lft}} \boldsymbol{\Pi}^{[uv]}_\mu \boldsymbol{V}^s_\mu \boldsymbol{e}_{\mathrm{rgh}}, \tag{3.59}$$

where $\boldsymbol{e}_{\mathrm{lft}}$ and $\boldsymbol{e}_{\mathrm{rgh}}$ are block-diagonal matrices with elements $\boldsymbol{e}^T$ and $\boldsymbol{e}$ respectively. Also, note that the low-res visibilities, $\boldsymbol{V}^{\mathrm{M}}_\mu$, is a diagonal matrix of size $N \times N$ with entries V^{M}_{pqkl} accordingly. The order of the multiplications is not important, since the matrices are diagonals and hence multiplication is commutative. Let us now spend some

time decomposing Eq. (3.59):

$$\boldsymbol{V}_{\mu}^{s} = \boldsymbol{W}_{\mu}\boldsymbol{V}_{\mu}. \tag{3.60}$$

The block diagonal matrix $\boldsymbol{W}_{\mu}$ of size $N\times[n_t\times n_\nu]\times N$ is constructed from W_{pqij} following a similar analogy as in Eq. (3.58); the same applies to $\boldsymbol{V}_{\mu}$, which is constructed from all the V_{pqij}.

If the number of pixels in the sky model is $N_{\text{pix}} \times N_{\text{pix}}$ and assuming that $\mathbf{F}$ is a block diagonal Fourier transform operator of size $N \times [N_{\text{pix}} \times N_{\text{pix}}] \times N$, where all its entries are elements of size $N_{\text{pix}} \times N_{\text{pix}}$, we have:

$$\begin{aligned}
\boldsymbol{V}_{\mu}^{\text{M}} &= \boldsymbol{e}_{\text{lft}}\boldsymbol{\Pi}_{\mu}^{[uv]}\boldsymbol{V}_{\mu}^{s}\boldsymbol{e}_{\text{rgh}} && (3.61)\\
&= \boldsymbol{e}_{\text{lft}}\boldsymbol{\Pi}_{\mu}^{[uv]}\boldsymbol{W}_{\mu}\boldsymbol{V}_{\mu}\boldsymbol{e}_{\text{rgh}} && (3.62)\\
&= \big(\boldsymbol{e}_{\text{lft}}\boldsymbol{\Pi}_{\mu}^{[uv]}\boldsymbol{W}_{\mu}\boldsymbol{e}_{\text{rgh}}\big)\big(\boldsymbol{F}\mathcal{I} + \epsilon_{\text{sky}}\big). && (3.63)
\end{aligned}$$

The error due to the noise then follows:

$$\begin{aligned}
E_{\text{err}} &= \big(\boldsymbol{e}_{\text{lft}}\boldsymbol{\Pi}_{\mu}^{[uv]}\boldsymbol{W}_{\mu}\boldsymbol{e}_{\text{rgh}}\big)\epsilon_{\text{sky}} && (3.64)\\
&= \boldsymbol{A}\epsilon_{\text{sky}}, && (3.65)
\end{aligned}$$

where $\boldsymbol{A} = \boldsymbol{e}_{\text{lft}}\boldsymbol{\Pi}_{\mu}^{[uv]}\boldsymbol{W}_{\mu}\boldsymbol{e}_{\text{rgh}}$. Let us assume that the vector ϵ_{sky} can take a finite set of directions, specifically the directions represented by $e_1, \ldots, e_{\text{N}_{\text{pix}}}$ with $\|\epsilon_{\text{sky}}\| = 1$, then let us look at the average of the squared error norm,

$$\begin{aligned}
\|E_{\text{err}}\|^2 = \sigma_{\text{pix}}^2 &= \frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{A}e_i\|_2^2 && (3.66)\\
&= \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{a_i}\|_2^2, && (3.67)
\end{aligned}$$

where $\mathbf{a_i}$ stands for the i^{th} column of $\boldsymbol{A}$ and $\|.\|_2$ is the Euclidean norm. The previous Eq. (3.67) can be written as:

$$\sigma_{\text{pix}}^2 = \frac{1}{N}\|\boldsymbol{A}\|_{\text{F}}^2, \tag{3.68}$$

where, $\|\boldsymbol{A}\|_{\mathrm{F}}$ is the Frobenius norm of $\boldsymbol{A}$ given by:

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \left(\sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{A}_{ij}^2\right)^{-\frac{1}{2}} \tag{3.69}$$

$$= \sqrt{\mathbf{Tr}\left(\boldsymbol{A}^T\boldsymbol{A}\right)}. \tag{3.70}$$

In this notation, **Tr** is the trace of the matrix.

3.2.3 Noise penalty estimates: empirical

In this section we employ simulations to empirically verify noise estimates computed using the derivation above. We generate a "high-res" JVLA-C measurement set (MS) corresponding to a 400 s synthesis with 1 s integration, with 30 MHz of bandwidth centred on 1.4 GHz, divided into 360 channels of 83.4 kHz each. The MS is filled with simulated thermal noise with $\sigma_{\mathrm{s}} = 1$ Jy. We then generate a "low-resolution" MS using 100 s integration, with a single frequency channel of 10 MHz. This MS receives the resampled visibilities. The size of the resampling bin is thus 100 s by 10 MHz, or 100×120 in terms of the number of high-res samples.

We then resample the high-res visibilities using a number of BDWFs, and store the results in the low-res MS:

- Standard averaging to 100 s and 10 MHz (using the middle 120 channels). This gives us the baseline noise estimate.
- Sinc and Bessel windows using the same bin, without an overlap, tuned to a FoI of 2°.
- The same windows with overlap factors of 4×3.

We then image the low-res MS and take the r.m.s. pixel noise across the image as an estimator of σ_{pix}, divide it by the baseline estimate produced with simple averaging, and compare the resulting noise penalty with that predicted by Eq. (3.53). Note that this estimator is not perfect, since image noise is correlated across pixels. Nonetheless, we obtain results that are broadly consistent with analytical predictions (Table 3.1).

BDWF	**Ξ analytic**	**Ξ sim**
sinc-1×1	1.247	1.22
sinc-4×3	1.242	1.20
Bessel-1×1	1.178	1.16
Bessel-4×3	1.109	1.12

TABLE 3.1: A comparison of image noise penalties associated with different BDWFs, computed analytically vs. simulations.

Figure 3.6 shows the predicted visibility noise penalty for the same BDWF, as a function of an East-West baseline component, which determines the baseline rotation speed. Each dot mark represents the noise penalty induced by a single baseline. Note that the noise penalty rises sharply towards longer East-West baselines. Note also that the penalty is well below 1 on shorter baselines, when overlapping is in effect. As mentioned in Section 3.2.1, this is because the BDWF becomes boxcar averaging (i.e. unweighted averaging or simple averaging) and can only decrease the noise when overlap is occurring. This is simple to understand analytically as follows: the noise term has constant r.m.s. $\sigma_{\rm s}$ across all baselines (including samples) and the resampling bin consists of $n_t \times n_\nu$ samples. Since the noise is not correlated between samples, the noise on the averaged (unweighted averaging) visibilities in Eq. (3.25) is given by:

$$\sigma_{\rm m}^2 = \frac{\sigma_{\rm s}^2}{n_t n_\nu}. \tag{3.71}$$

On the shorter baselines, the BDWF $X \equiv 1$ (i.e. unweighted averaging) and overlapping means enlarging the averaged resampling bin by a factor of $\alpha\beta$ with $\alpha > 1$ and/or $\beta > 1$. The noise on the averaged visibilities becomes:

$$\sigma_{X\mu}^2 = \frac{\sigma_{\rm s}^2}{\alpha\beta n_t n_\nu}, \tag{3.72}$$

which gives us a noise penalty of

$$\Xi_{X\mu} = \frac{\sigma_{X\mu}}{\sigma_{\rm m}} = \frac{1}{\sqrt{\alpha\beta}} < 1. \tag{3.73}$$

Note that as $\alpha\beta > 1$, it implies that $\sqrt{\alpha\beta} > 1$.

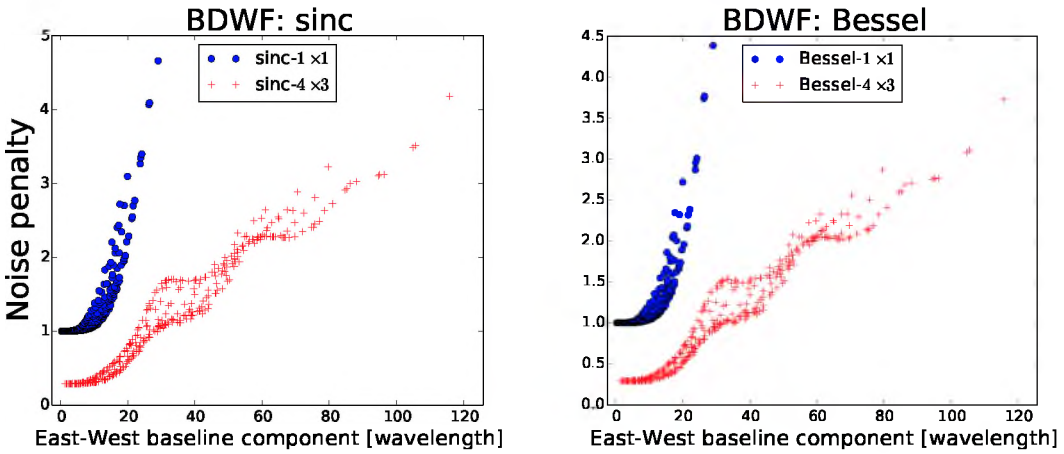

FIGURE 3.6: Noise penalty w.r.t. simple averaging as a function of baseline length for two different BDWFs (sinc and Bessel), with and without overlap. JVLA-C, 1.4 GHz, sampling intervals of 100 s and 10 MHz.

3.3 Simulations and results

In this section we use BDWFs to resample simulated visibility data, and study the effect on smearing and source suppression. Apart from a few examples documented separately, the basis interferometer configuration employed in the simulations corresponds to JVLA-C observing at 1.4 GHz. Similar to Section 3.2.3, we create a "high-res" measurement set corresponding to a 400 s synthesis at 1 s integration, with 30 MHz total bandwidth centred on 1.4 GHz, divided into 360 channels of 83.4 kHz each. The MS is populated by noise-free simulated visibilities corresponding to a single point source at a given distance from the phase centre. We then generate "low-res" MSs to receive the resampled visibilities, resample the latter using a range of BDWFs, convert the visibilities to dirty images (using natural weighting unless otherwise stated), and measure the peak source flux in each image. Since each dirty image corresponds to a single source, the peak flux gives us the degree of smearing or the smearing factor (i.e. the degree of amplitude decrease towards off-axis sources) associated with a given BDWF and sampling interval.

For the first set of simulations, the "low-res" MSs corresponds to a 100 s and 10 MHz synthesis. We employ three sampling rates, 25 s × 2.5 MHz, 50 s × 5 MHz and 100 s × 10 MHz (thus four timeslots and four channels, two timeslots and two channels, and a single-channel snapshot).

A typical performance comparison for the JVLA-C configuration at 1.4 GHz is given by Figure 3.7. This figure illustrates some of the principal achievements of the present

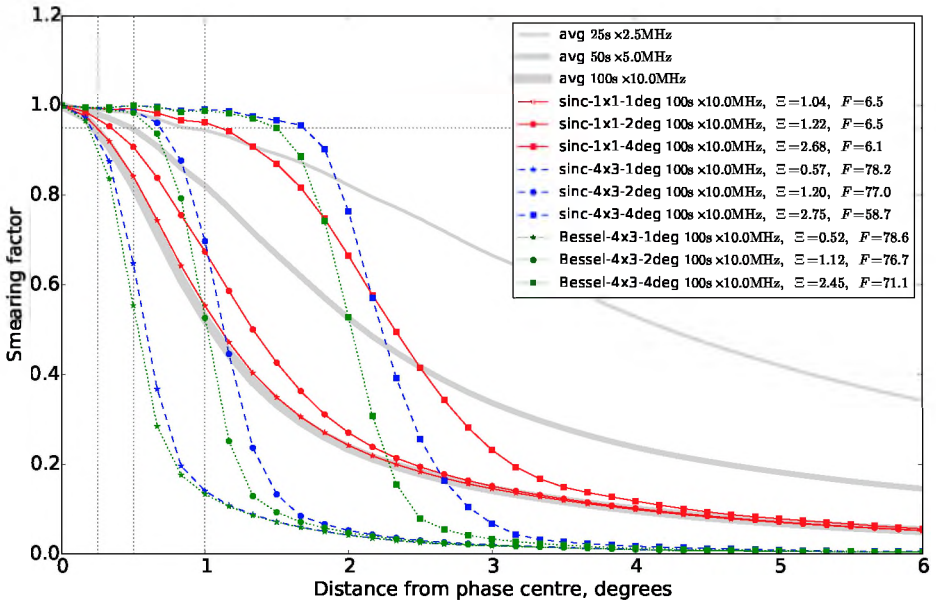

FIGURE 3.7: JVLA-C 1.4 GHz. Smearing as a function of distance from phase centre, for conventional averaging with 25 s × 2.5 MHz, 50 s × 5 MHz and 100 s × 10 MHz bins, and for several BDWFs with 100 s × 10 MHz bins. The noise penalty Ξ and the far-source suppression factor F are given relative to 25 s × 2.5 MHz averaging.

work, so it warrants detailed explanation. The horizontal axis represents the distance from phase centre, while the vertical axis represents the degree of smearing. Unity corresponds to no smearing; this is the case at the phase centre, thus all curves start at unity. The three thick gray curves correspond to simple averaging into 25 s × 2.5 MHz, 50 s × 5 MHz and 100 s × 10 MHz. We can (rather arbitrarily) define a series of "acceptable" smearing levels by specifying a FoI radius, and the maximum extent of smearing over that FoI. For the FoI radius, we may pick , for example, the half-power point of the primary beam, the main lobe of the primary beam, or the extent of the first sidelobe of the primary beam. For JVLA's 25-metre dishes at this frequency, these radii correspond to $\sim 0.25°$, $0.5°$, and $1°$, respectively; they are indicated by thin vertical lines in the figure. The thin horizontal line indicates our chosen smearing threshold of 0.95. Figure 3.8 is the results of Figure 3.7, where all the curves are normalised with respect to the 25 s × 2.5 MHz averaging curve.

For regular averaging, the three chosen bin sizes happen to correspond to chosen acceptable levels of smearing over the three chosen FoI values. The other curves show the performance of a few different BDWFs, all at 100 s × 10 MHz sampling. Three types of BDWFs are shown, indicated by line style and colour:

sinc-1×1 i.e. a non-overlapping sinc window (solid line, red)

sinc-4×3 i.e. an overlapping sinc window (dashed line, blue)

Bessel-4×3 i.e. an overlapping Bessel window (dotted line, green)

These are tuned to three different FoI settings, as indicated by the plot symbol: 1° (star), 2° (circle), 4° (square).

The plot is meant to show performance of BDWFs at 100 s × 10 MHz versus a "baseline case" of 25 s × 2.5 MHz averaging, the latter being an acceptable averaging setting for this particular frequency and telescope geometry. The legend next to the plot therefore indicates Ξ, the noise penalty associated with that particular BDWF, and F, the far source suppression factor. Both values are calculated w.r.t. the baseline case. Note the following salient features:

- All overlapping BDWFs provide outstanding far source suppression in this regime, with F in the $60 \sim 80$ range. The non-overlapping sinc (solid red lines) only achieves $F \sim 6$, which is similar to regular averaging at the same rate.
- Noise performance is excellent for the 1° BDWFs. There is a small noise penalty at 2°, and a larger (over a factor of 2) noise penalty at 4°. This can be easily understood by considering the shape of BDFWs as a function of FoI: smaller FoIs correspond to broader windows that become more "boxcar-like" over the sampling interval, and vice versa. This means that, in this particular configuration, BDWFs cannot achieve a FoI of 4° at 10 s × 100 MHz without a substantial sacrifice in sensitivity. We shall return to this issue below.
- If the desired FoI size is $r \sim 0.5 - 1°$, overlapping BDWFs (sinc-4x3-2deg and Bessel- 4x3-2deg) provide excellent performance at 100 s × 10 MHz. Compared to averaging at 25 s × 2.5 MHz, they achieve a factor of 16 data compression with minimal loss of sensitivity, with excellent tapering behaviour: the smearing performance across the FoI is equivalent to (or better) than that of simple averaging, and out-of-FoI source suppression is almost two orders of magnitude higher.

Figure 3.8 presents the same results in an alternative way. Here, the recovered flux is shown relative to the baseline case of 25 s × 2.5 MHz averaging. This clearly illustrates the excellent performance of overlapping BDWFs tuned to 2° FoI.

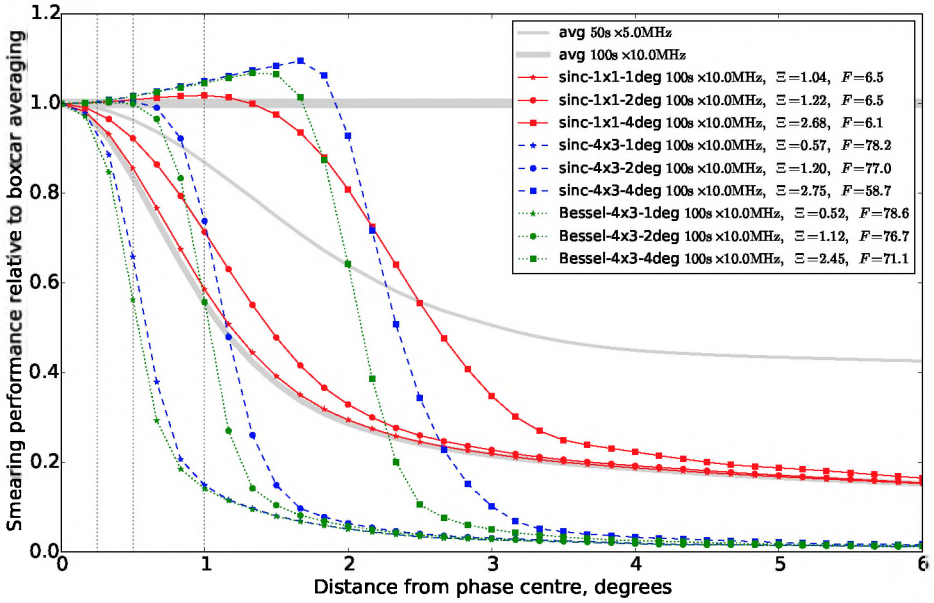

FIGURE 3.8: JVLA-C 1.4 GHz. Results of Figure 3.7 normalised to the 25 s × 2.5 MHz averaging curve.

3.3.1 Noise penalties and overlapping BDWFs

Values of $\Xi < 1$ above may be paradoxical at first, since one cannot theoretically exceed the noise performance of the unweighted average. This, however, is an artefact of the short simulation. Overlapping BDWFs are essentially averaging in "bonus signal" from regions of overlap extending outside the nominal time and frequency coverage. In this case, at 100 s × 10 MHz sampling, a BDWF with 4 × 3 overlap is actually adding up signal over a 400 s × 30 MHz bin, i.e. a bin that is a factor of 12 larger (though of course the bonus sensitivity thus gained is much less than the theoretically available $\sqrt{12}$, since the weights over the overlap regions correspond to the "wings" of the window function, and are thus small). This can easily result in lower per-visibility noise than that achieved by regular averaging over 100 s × 10 MHz, and correspondingly higher snapshot sensitivity.

In the more realistic case of a long, multiple-channel synthesis (what we will call a *full synthesis*), the effects of bonus sensitivity disappear. While the noise on individual visibilities remains nominally lower in a full synthesis thanks to the overlap, it becomes correlated across neighbouring *uv*-bins, so there is no net gain in image-plane sensitivity. Strictly speaking, at the "edge" of the synthesis, overlapping BDWFs are still pulling in

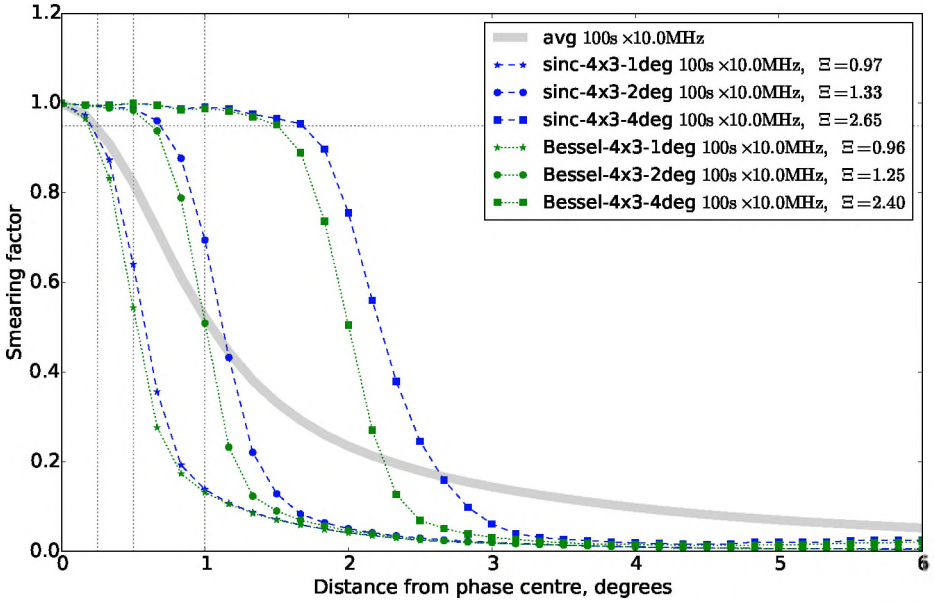

FIGURE 3.9: JVLA-C 1.4 GHz, long wideband synthesis. Smearing as a function of distance from phase centre, for conventional averaging with 100 s × 10 MHz bins, and for overlapping BDWFs with 100 s × 10 MHz bins. The noise penalty Ξ is given relative to 100 s × 10 MHz averaging.

some bonus signal from overlap regions extending beyond the synthesis coverage, but since the area of this overlap is negligible compared to the coverage of the full synthesis, so is the effect of the bonus signal.

In other words, simulating a snapshot observation results in underestimated noise penalties, compared to the real-life case of a full synthesis. We should expect the noise penalties to go up (and eventually exceed unity) as we increase the synthesis time and number of channels. Figure 3.9 presents the results of such a simulation. This shows a a 1800 s × 200 MHz synthesis, sampled at the same rates as above. The results should be compared to and contrasted with those of Figure 3.7. Note that the tapering response of BDWFs is nearly identical, while the noise penalties are indeed higher. With 4 × 3 overlap and 100 s × 10 MHz sampling, the total signal accessed by overlapping BDWFs corresponds to 2100 s × 220 MHz, which gives a theoretical maximum of a factor of ~ 1.13 in bonus sensitivity. In other words, the values of Ξ in Figure 3.9 are still underestimated, but by 13% at most (which explains $\Xi < 1$ for the 1° case). From this we may safely extrapolate that the noise penalty of BDWFs matched to $1 - 2°$ FoIs will remain reasonable even for a much longer and wider band synthesis.

3.3.2 FoIs and sampling rates

For BDWFs, a given FoI tuning represents a characteristic scale in the uv-plane, which is inversely proportional to the FoI parameter. On the other hand, the uv-bin sampled by any given visibility is proportional to the integration time, fractional bandwidth, and baseline length. Since the window function is truncated at the edge of the averaging bin (which can be larger than the sampling bin by a factor of several, if overlapping BDWFs are employed), there is, for any given baseline, some kind of optimal range of uv-bin sizes over which BDWFs tuned to a particular FoI setting are "efficient". Over smaller bins, BDWFs become equivalent to a boxcar averaging, over larger bins, BDWFs penalise too much sensitivity as they downweight more samples. Since this optimal bin size is proportional to baseline length, the overall optimum is dependent on the distribution of baselines in the array.

Furthermore, the sampling rate needs to be "balanced" in time and frequency for BDWFs to achieve efficient tapering response. If the uv-bins are elongated, the window function becomes truncated (i.e. more boxcar-like) across the bin, which reduces its ability to induce the desired taper. Since the orientation of the bins changes as the baseline rotates, the cumulative effect is an average degradation of the tapering response, in the sense that it becomes closer to that of boxcar averaging. In this sense, the optimal uv-bin shape is square-like. This occurs when the fractional bandwidth is equal to the arc section swept out by the baseline over one bin. For a polar observation (circular uv-tracks), we can express this as

$$\Delta\nu/\nu = 2\pi(B_x/B)(\Delta t/24\mathrm{h}), \tag{3.74}$$

where B is the baseline length, and B_x is its East-West component. Rewriting this in terms of more convenient units, we obtain:

$$\frac{\Delta\nu_{\mathrm{MHz}}}{\Delta t_{\mathrm{s}}} \approx \frac{\nu_{\mathrm{MHz}}}{14000} \cdot \frac{B_x}{B}, \tag{3.75}$$

leading to a simple rule-of-thumb: at 1.4 GHz, an East-West baseline sweeps out a square-like bin when the integration time in seconds is 10 times the channel width in MHz (hence the use of bin sizes such as 100 s $\times$ 10 MHz in the analysis here).

The interaction between uv-bin size and tapering response is illustrated in Figure 3.10. Here we compare the performance of two BDWFs tuned to a 4° FoI – a non-overlapping

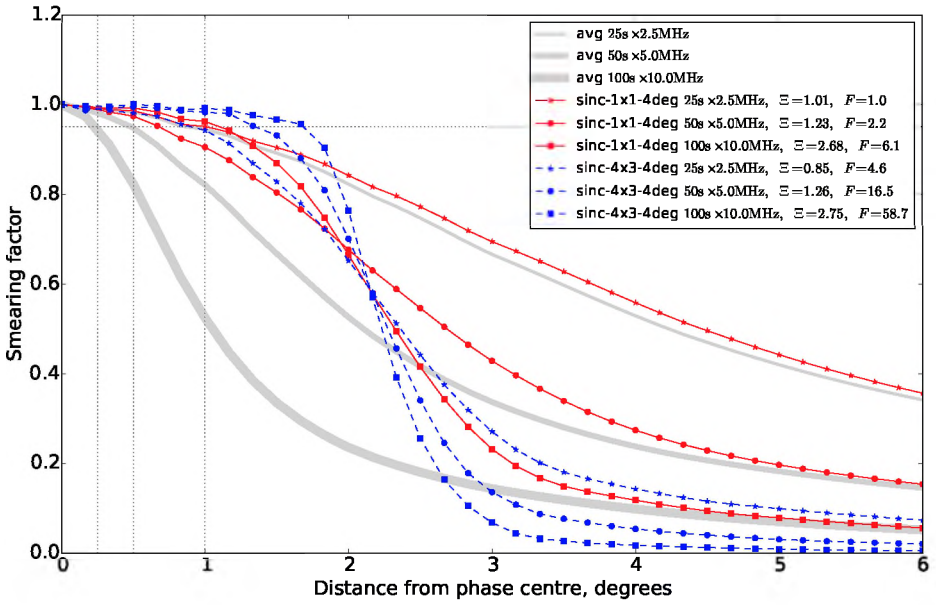

FIGURE 3.10: JVLA-C 1.4 GHz. Smearing as a function of distance from the phase centre, for conventional averaging with 25 s × 2.5 MHz, 50 s × 5 MHz and 100 s × 10 MHz bins, and for several BDWFs with 100 s × 10 MHz bins. The noise penalty Ξ and the far-source suppression factor F are given relative to 25 s × 2.5 MHz averaging.

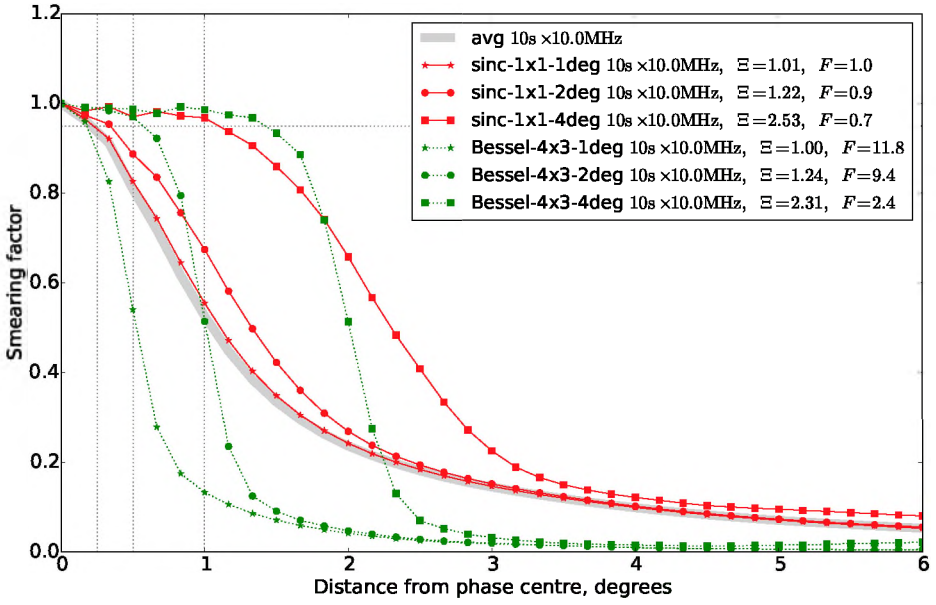

FIGURE 3.11: JVLA-C 14 GHz. Smearing as a function of distance from the phase centre, for conventional averaging with 2.5 s × 2.5 MHz, 5 s × 5 MHz and 10 s × 10 MHz bins, and for several BDWFs with 10 s × 10 MHz bins. The noise penalty Ξ and the far-source suppression factor F are given relative to 2.5 s × 2.5 MHz averaging.

sinc-1×1 filter (solid red lines) and an overlapping sinc-4×3 filter (dashed blue lines) – over three sampling bin sizes: 25 s × 2.5 MHz, 50 s × 5 MHz and 100 s × 10 MHz. For reference, the performance of boxcar averaging over the same bin sizes is indicated by the thick grey lines. Note how at the smaller bin size, the non-overlapping sinc is practically equivalent to a boxcar in terms of tapering response; at the larger bin size, it begins to shape the FoI. Introducing an overlap improves the response considerably. An overlapping filter at 25 s × 2.5 MHz achieves almost the same tapering response as a non-overlapping one at 100 s × 10 MHz (which is not surprising, if one considers that the effective averaging bin size in the former case is 100 s × 7.5 MHz). However, for all filters, at 100 s × 10 MHz the noise penalty goes up quite sharply.

This illustrates that 50 s × 5 MHz is an appropriate BDWF sampling rate for achieving a 4° FoI (for JVLA-C configuration at 1.4 Ghz), providing a reasonable trade-off between tapering response and noise penalty. At higher sampling rates, the tapering response is degraded, while at lower sampling rates, the noise penalty increases. In comparison (as we saw in the previous section), for FoIs of $1 - 2°$, BDWFs achieve a good trade-off at 100 s × 10 MHz sampling.

It is interesting to consider how optimal BDWF sampling changes as a function of array size. Figure 3.11 shows a simulation for JVLA-C at 14 GHz. (Since our results are completely determined by uv-plane geometry in wavelengths, this is equivalent to JVLA-C scaled up by a factor of 10 at an observing frequency of 1.4 GHz). From Eq. (3.75), we can see that square-like uv-bins correspond to sampling rate combinations such as 10 s × 10 MHz. The simulation presented here is for a 1800 s × 200 MHz synthesis, i.e. is closer to the full synthesis rather than a snapshot case. Comparing Figures 3.11 and 3.9, we find nearly identical BDWF performance (in terms of tapering response and noise penalty) at 14 GHz and 1.4 GHz, with only the optimal sampling rate being different.

3.3.3 BDWFs for wide-field VLBI

In the VLBI regime, it is usually a combination of smearing and data rates, rather than the primary beam, that effectively limits the FoI. For example, the current European VLBI Network (EVN) correlator [Keimpema et al., 2015a] operated by the Joint Institute for VLBI ERIC is capable of producing data at dump rates down to 10 ms, with 16 MHz of total bandwidth split into up into 8192 channels. The maximum available FoI of an EVN

experiment is restricted by the smallest primary beam, which is usually that of the 100 m Effelsberg telescope – about $10'$ in diameter at L-band. The EVN calculator[1] shows that a dump rate of 0.125 s and 1024 channels (16 kHz) is required to keep smearing to within 10% across this FoI. Due to the large computational and storage requirements, such data rates have only been employed in one-off experiments. For routine use, techniques such as multiple-phase centre correlation are more common. Typically, data are averaged into more modest sampling rates of 2 s and 32 channels. This restricts the effective (L-band) FoI to about $20''$, and thus limits the scientific usefulness of archival data to narrow-FoI experiments.

In this section we investigate whether the use of BDWFs can enable true wide-field VLBI. We simulate a 1.6 GHz EVN observation correlated to the JVLA in C configuration. The simulation employs seven EVN stations (Effelsberg, Hartebeesthoek, Jodrell Bank, Noto, Onsala, Torun, Westerbork, Shanghai), with a maximum baseline of 10161 km when the JVLA stations are included. Figure 3.12 compares the smearing response of simple averaging to that of two overlapping Bessel BDWFs, employing 0.5 s and 25 kHz sampling. At these data rates, it becomes almost practical to have a full-FoI EVN archive.

For comparison, we also show the performance of BDWFs for a hypothetical fast-transient archive application. In order to localise potential fast radio bursts, we would need to retain the native time resolution of 10 ms, with averaging only done in frequency, and that limited by the need to retain some spectra resolution for de-dispersion. In this regime, BDWFs are less efficient since the uv-bins are elongated. Figure 3.12 shows that this translates into less source suppression outside the FoI, but does not affect the ability to retain the FoI with 120 kHz bandwidth averaging using the Bessel BDWFs.

3.4 Conclusion

The goal of this work was to demonstrate the application of BDWFs to radio interferometry. We have demonstrated that BDWFs offer a number of interesting advantages over conventional averaging. The first of these is data compression, i.e. visibilities can be sampled at a lower rate, while retaining a large FoI. Compression by a factor of 16 with

[1] `http://www.evlbi.org/cgi-bin/EVNcalc`

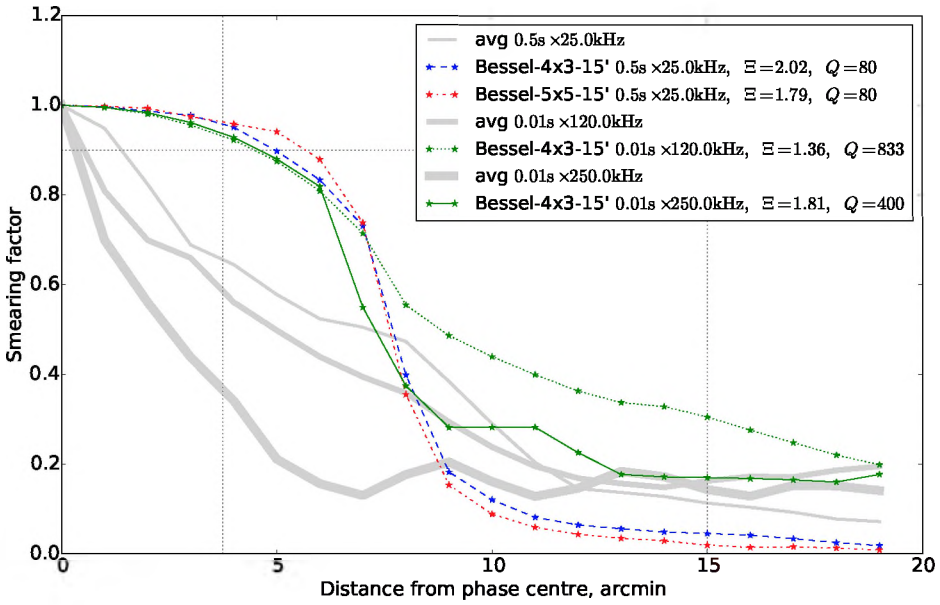

FIGURE 3.12: VLBI simulation (10161 km maximum baseline) at 1.6 GHz. Smearing as a function of distance from phase centre, for conventional averaging at 0.5 s × 25 kHz, 10 ms × 120 kHz and 10 ms × 250 kHz averaging, and overlapping Bessel BDWFs. The horizontal line corresponds to 10% smearing losses. The ratio Q shows the increase in data volume over conventional 2 s × 0.5 MHz averaging.

relatively little loss of sensitivity has been demonstrated. The second potential benefit of BDWFs is the increased suppression of unwanted signal from out-of-FoI sources. This reduces the overall level of far sidelobe confusion noise and lessens the impact of A-team sources in sidelobes. Thirdly, BDWFs can have an interesting impact in the VLBI case, as they allow the full primary beam FoI to be imaged using a single VLBI data-set. This opens the door to wide-field VLBI, which has previously been impractical.

As we saw above, the ability of BDWFs to shape the FoI is somewhat limited by the fact that shorter baselines sweep out smaller bins in uv-space, with window functions over them becoming boxcar-like. If BDA is employed, shorter baselines are averaged over larger uv-bins, thus increasing the effect of BDWFs. The next chapter will be devoted to combining BDWFs with BDA. Finally, we should note that the use of BDWFs results in a different position-dependent PSF than regular averaging. In other words, the smearing response of BDWFs results in a different smeared PSF shape. Chapter 5 will focus on methods of deriving this PSF shape, with a view to incorporating this into current imaging algorithms.

Chapter 4

Baseline-dependent Averaging and BDWFs Applied Across Equal uv-Distance for Data Compression

In the previous chapter, we looked at the various ways in which averaging larger visibility bins can result in an acceptable level of smearing (i.e. 5% or less decrease in source amplitude within the observation FoI). We showed, theoretically, that decorrelation/smearing increases on longer baselines compared to the shorter ones, and that decorrelation can even be avoided if the correlator performs the averaging procedure over shorter bins, which however results in high data rates. We made predictions pertaining to an equal uv-distance averaging across all baselines. In particular, averaging within sufficiently large bins for shorter baselines, while on the other hand, the longer baselines are averaged within shorter bins. Intuitively, this corresponds to the bins on all the baselines being averaged across an equal uv-distance. The question therefore arises whether such averaging technique (averaging over equal uv-distance on all baselines) will not only decrease smearing within the observation FoI, but importantly, also result in reducing the data size. Indeed, one expects that averaging for sufficiently large bin intervals on the

This chapter draws extensively on: M. T. Atemkeng, O. M. Smirnov, C. Tasse, G. Foster and J. Jonas, *Baseline-dependent averaging and baseline-dependent window functions for equal uv-distance data compression and field-of-interest shaping in radio interferometry.* Monthly Notices of the Royal Astronomical Society (MNRAS), in preparation. It is acknowledged that some of the text in this chapter will therefore "match" that of the article. The reference in this footnote serves as a general reference for all such text.

shorter baselines would be favourable for data compression, while averaging smaller intervals in uv-distance on the longer baselines will result in decreasing decorrelation. The second question pertains to calibration issues for BDA given that calibration is a complex visibilities correction process. The key point for the calibration properties of BDA will be to monitor the correlator differently at each averaging interval on different baselines. This implies that the calibration parameters will change differently with baselines and each of the frequency and/or time intervals.

In this chapter, it is shown that BDA results in decreasing smearing over the selected FoI, and applying BDA to BDWFs results in a subtantial decrease in smearing with less than 1% smearing over the FoI, and excellent source suppression out of the FoI. To obtain the results, we describe the mathematics behind BDA and the compression factor on each baseline. We identify the implementation criteria and describe three different algorithms for the implementation of BDA correlators. We implement BDA via simulations using the JVLA telescope and discuss the results.

4.1 Background and problem statement

The effect of time and bandwidth averaging is more severe on longer baselines than on shorter baselines. This is easy to understand as follows: the visibility from a baseline pq of a point source with brightness S is given by:

$$V_{pq} = S \exp\{i\phi\}, \ \phi = 2\pi \boldsymbol{u}_{pq} \cdot \boldsymbol{l}, \tag{4.1}$$

where ϕ is the phase of the source, $\boldsymbol{u}_{pq} = (u, v, w)$ the baseline vector in unit of wavelength and $\boldsymbol{l} = (l, m, n-1)$ the source location in the sky. This implies that the larger the norm $\|\boldsymbol{u}_{pq}\|$, i.e. the baseline length, the larger the phase and thus much more severe attenuation of the source brightness S occurs. Sources far away from the pointing direction have larger radii $\|\boldsymbol{l}\|$, thus a larger phase as well, i.e. the fringes of sources far from the pointing direction rotate rapidly in the uv domain.

In the rest of this chapter, we primarily refer to each baseline's East-West component. During time-frequency integration, the East-West component rotates, while the South-North does not. For the same length of time, two equal length baselines with different

orientations will sweep unequal *uv*-distances. This results in different degrees of decorrelation. In this sense, the East-West component determines the baseline rotation speed. This is illustrated in Figure 4.1.

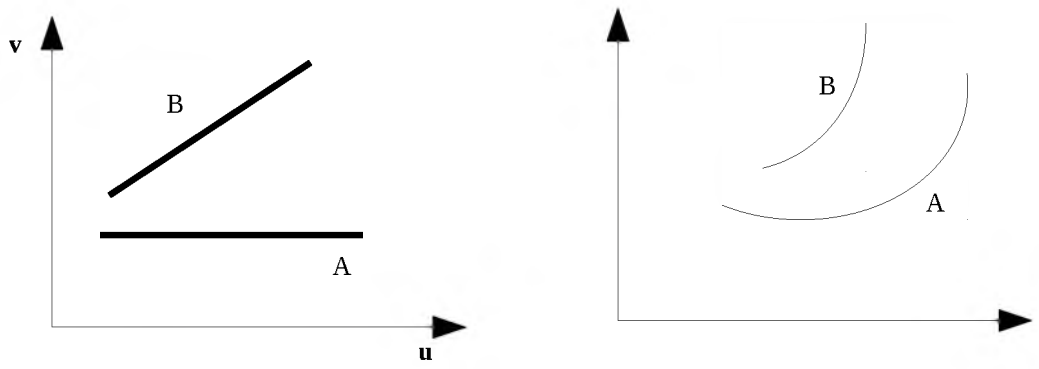

FIGURE 4.1: Two equal lengths baselines A and B with different orientation. Baseline A is an East-West baseline (the South-North component is zero) and B has non zero South-North and East-West components. The distances swept by these baselines East-West are different during the same time integration: baseline A rotates more rapidly than B.

Figure 4.2 is a simulated observation at 1.4 GHz of the JVLA in C configuration showing the amplitude of a source as seen by three baselines (longest, medium length and shortest baselines). In this figure, the amplitude of the source is plotted against the source distance from the phase centre. The top panel shows the effect of bandwidth averaging, while the bottom panel shows the effect of time averaging.

Figure 4.3 shows the amplitude loss as a function of East-West baseline length. We simulate the source at 30 arcmin away from phase centre and observe the amplitude loss with JVLA C. In this figure, the dot grey line shows the length of the shortest East-West baseline. In the top panel, the source is simulated then averaged using 100 s integration and 125 kHz channel width, while in the bottom panel the source is simulated then averaged using 1 s integration and 10 MHz channel width.

It is observed from Figure 4.2 and 4.3 that decorrelation/smearing is severe on longer baselines compared to shorter baselines. In addition, Figure 4.2 shows that smearing is also a function of source position in the sky.

Most of the data measured by any compact interferometer array comes from the shorter baselines. A good illustration is presented in Figure 4.4, which is a four-hour synthesis

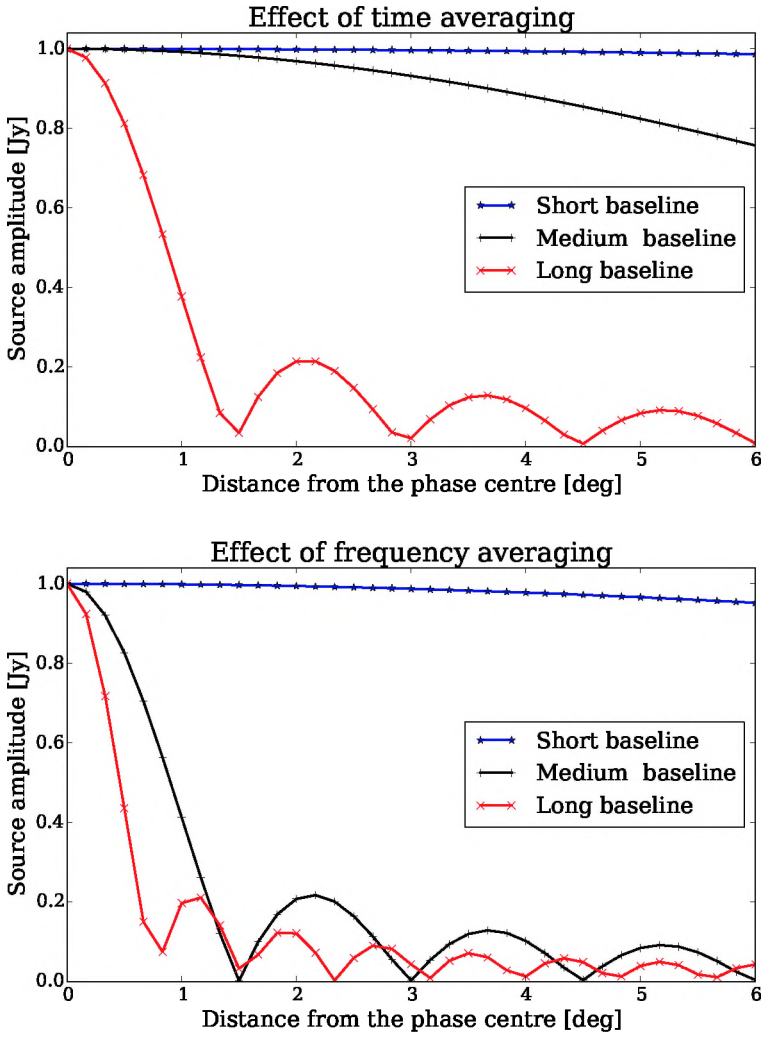

FIGURE 4.2: Amplitude loss: the apparent intensity of a 1 Jy source, as seen by JVLA-C at 1.4 GHz, as a function of distance from the phase centre. (Top): Time and frequency integrations fixed at 100 s and 125 kHz respectively; (Bottom): Time and frequency integrations fixed at 1 s and 10 MHz respectively.

uv-coverage of the MeerKAT observing at 1.4 GHz. The coverage shows that the data points are more populated in the inner core (red) than in the outer areas (blue). The data in the inner core are from the short baselines, while those in the outer core are from the long baselines. If more samples should be averaged in the inner core and very little at the outer, decorrelation would be avoided on the longer baselines and data compression would be carried out on the shorter baselines. This method, often referred to as *BDA*, was first proposed by Cotton [1999, 1989] as an approach for dealing with wide field imaging with little to no bandwidth and time averaging effects. The next section describes the mathematical details of BDA.

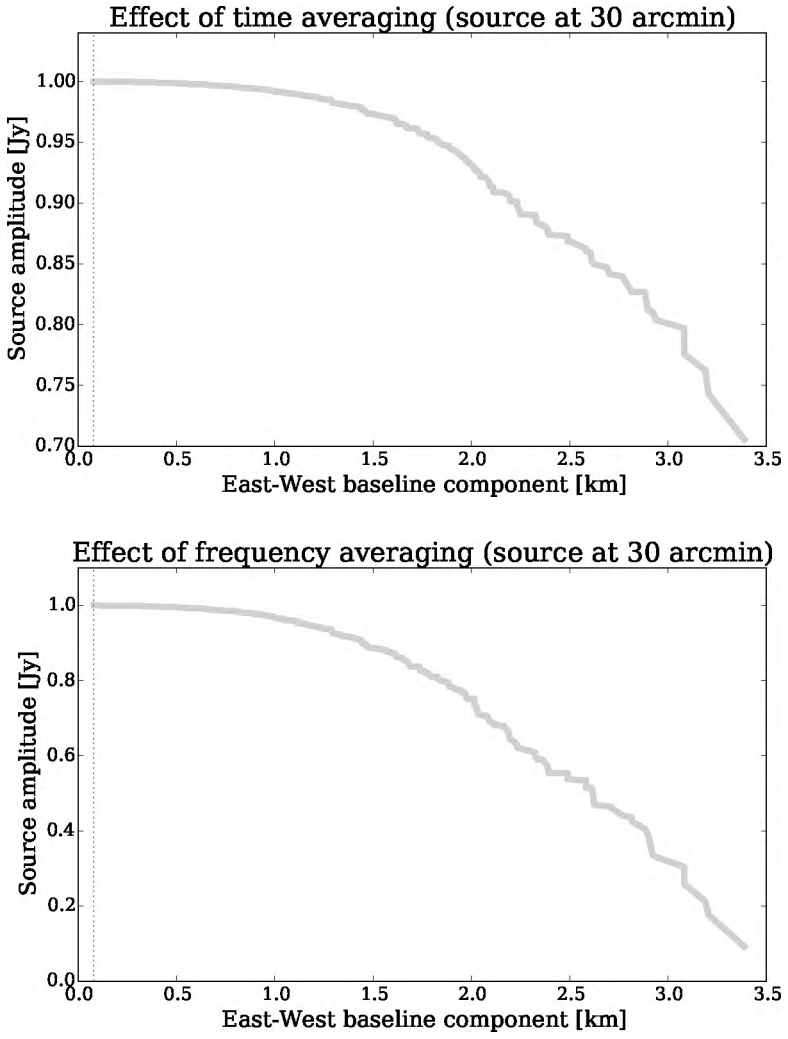

FIGURE 4.3: Amplitude loss: the apparent intensity of a 1 Jy source, as seen by JVLA-C at 1.4 GHz, as a function of East-West baseline components. (Top): Time and frequency integrations fixed at 100 s and 125 kHz respectively; (Bottom) Time and frequency integrations fixed at 1 s and 10 MHz respectively.

4.1.1 Baseline-dependent Averaging

As explained in Chapter 3, an interferometer measures the average visibility over a rectangular time-frequency bin given by the time and frequency sampling intervals Δt and $\Delta\nu$ respectively; what is called the resampling bin (see Chapter 3, Eq. 3.14)

$$\mathsf{B}_{kl}^{[\Delta t \Delta\nu]} = \left[t_k - \frac{\Delta t}{2}, t_k + \frac{\Delta t}{2}\right] \times \left[\nu_l - \frac{\Delta\nu}{2}, \nu_l + \frac{\Delta\nu}{2}\right], \tag{4.2}$$

where k and l are the indexes for the bin at the centre of Δt and $\Delta\nu$ respectively. The *resampling bin* $\mathsf{B}_{kl}^{[\Delta t \Delta\nu]}$ is a 2-D rectangular set, containing the time and frequency values for all bins measured during the sampling intervals Δt and $\Delta\nu$ (or a set of time and

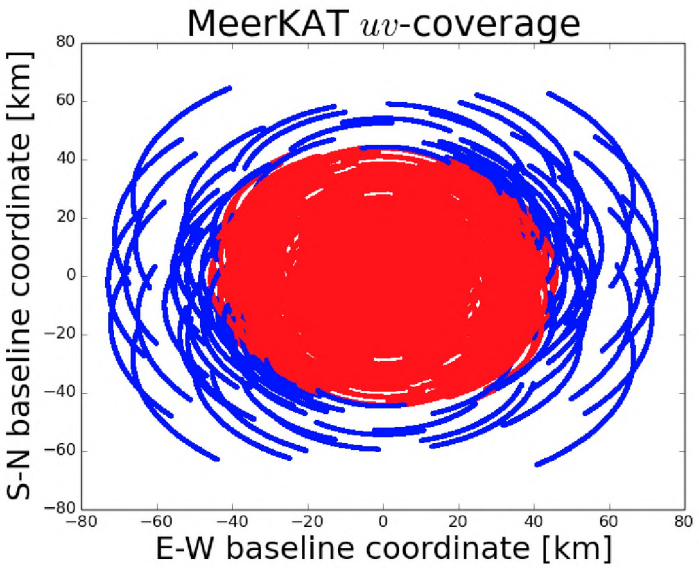

FIGURE 4.4: MeerKAT uv-coverage at 1.4 GHz, four hours' observation showing clearly that the data are most condensed at the centre core (red). The data at the centre core are from the shorter baselines, while the data at the outer (blue) are from the longer baselines.

frequency bins that are part of the uv-track draw by a baseline during the integrations Δt and $\Delta\nu$).

Note that for simple averaging the sampling intervals remain constant across all baselines while in BDA the sampling intervals are functions of baseline length and direction. This makes the sampling intervals vary across baselines. Let us denote by $\Delta_{pq}t$ and $\Delta_{pq}\nu$ the sampling intervals in time and frequency for the baseline pq. The resampling bin for the case of BDA is then given by:

$$\mathsf{B}_{kl}^{[\Delta_{pq}t,\Delta_{pq}\nu]} = \left[t_k - \frac{\Delta_{pq}t}{2}, t_k + \frac{\Delta_{pq}t}{2}\right] \times \left[\nu_l - \frac{\Delta_{pq}\nu}{2}, \nu_l + \frac{\Delta_{pq}\nu}{2}\right]. \tag{4.3}$$

Figure 4.5 shows a typical resampling bin for BDA and for simple averaging.

If we denote by $\mathcal{D}$ the function defining the distances between the centre time/frequency bin and an off-centre time/frequency bin, then we have:

$$\begin{aligned} \mathcal{D} : \mathsf{B}^{[\Delta_{pq}t,\Delta_{pq}\nu]} &\rightarrow \{\mathcal{R},\mathcal{R}\} \\ t,\nu &\mapsto \Big(\mathcal{D}(t-t_k,\nu), \mathcal{D}(t,\nu-\nu_l)\Big), \end{aligned}$$

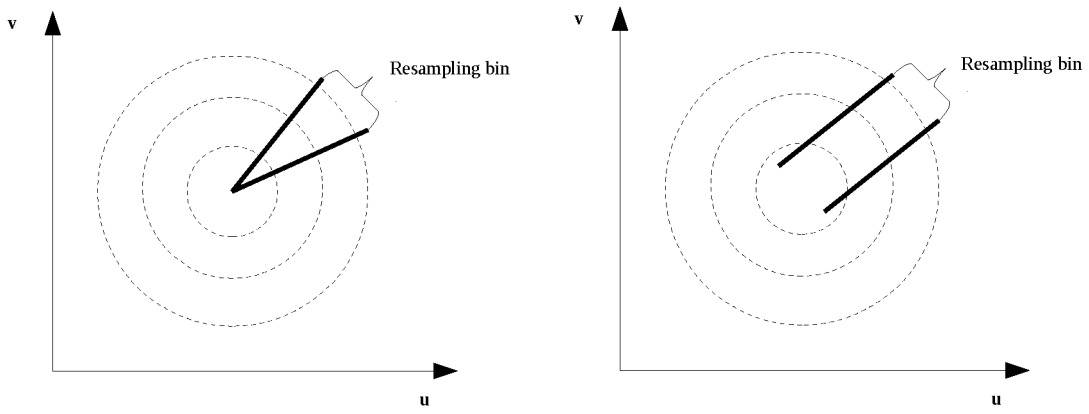

FIGURE 4.5: Applying BDA or simple averaging is equivalent to convolving the resampling bin by a boxcar windowing function. An East-West interferometer array, the BDA (right panel) corresponds to an equal convolution kernel and resampling bin across all baselines.

where $(t,\nu) \neq (t_k,\nu_l)$ and $\mathcal{R}$ is the set of real numbers. The distances $\mathcal{D}(t-t_k,\nu)$ and $\mathcal{D}(t,\nu-\nu_l)$ are defined as:

$$\mathcal{D}(t-t_k,\nu) = \|\boldsymbol{u}_{pq}(t-t_k,\nu)\| \tag{4.4}$$

$$\mathcal{D}(t,\nu-\nu_l) = \|\boldsymbol{u}_{pq}(t,\nu-\nu_l)\|\,. \tag{4.5}$$

The set B_{kl}^{pq} corresponding to the bin indices of the *resampling bin* for baseline pq is given as:

$$\mathsf{B}_{kl}^{pq} = \left\{ij:\ t_i\nu_j \in \mathsf{B}_{kl}^{[\Delta_{pq}t,\Delta_{pq}\nu]}\right\}. \tag{4.6}$$

Let us now denote by $\mathrm{Dim}\left\{\mathsf{B}_{kl}^{pq}\right\}$ the number of bin indices or samples in the resampling bin for the baseline pq. Suppose that $\mathsf{D}^{[\Delta_{pq}t]}$ and $\mathsf{D}^{[\Delta_{pq}\nu]}$ are the distances swept by the baseline pq for given baseline-dependent sampling intervals $\Delta_{pq}t$ and $\Delta_{pq}\nu$. We define these distances as follows:

$$\mathsf{D}^{[\Delta_{pq}t]} = \sum_{ij\in\mathsf{B}_{kl}^{pq}} \mathcal{D}(t_i-t_k,\nu), \tag{4.7}$$

and

$$\mathsf{D}^{[\Delta_{pq}\nu]} = \sum_{ij\in\mathsf{B}_{kl}^{pq}} \mathcal{D}(t,\nu_j-\nu_l). \tag{4.8}$$

The decorrelation is then defined by the product of decorrelation in time and decorrelation in frequency, i.e. by:

$$\mathsf{D}^{[\Delta_{pq}t,\Delta_{pq}\nu]} = \mathsf{D}^{[\Delta_{pq}t]} \times \mathsf{D}^{[\Delta_{pq}\nu]}. \tag{4.9}$$

The following constraints must be satisfied for averaging across equal uv-distance, i.e. for all East-West baselines $\alpha\beta \neq pq$ with $\|\boldsymbol{u}_{\alpha\beta}\| \neq \|\boldsymbol{u}_{pq}\|$:

$$\mathsf{D}^{[\Delta_{\alpha\beta}t,\Delta_{\alpha\beta}\nu]} = \mathsf{D}^{[\Delta_{pq}t,\Delta_{pq}\nu]} \tag{4.10}$$

$$\mathrm{Dim}\{\mathsf{B}_{kl}^{[\Delta_{\alpha\beta}t,\Delta_{\alpha\beta}\nu]}\} \neq \mathrm{Dim}\{\mathsf{B}_{kl}^{[\Delta_{pq}t,\Delta_{pq}\nu]}\}. \tag{4.11}$$

The condition in Eq. (4.10) is strictly equivalent to $\mathsf{D}^{[\Delta_{\alpha\beta}t]} = \mathsf{D}^{[\Delta_{pq}t]}$ and $\mathsf{D}^{[\Delta_{\alpha\beta}\nu]} = \mathsf{D}^{[\Delta_{pq}\nu]}$.

The constraints in Eq. (4.10) and (4.11) for BDA show that more samples are averaged on the shorter baselines than on the longer baselines. For example, if $p = 1$, $q = 2$ are the indexes for the longest baseline; $p = 2$, $q = 3$ the indexes for the medium length baseline and $p = 3$, $q = 4$ the indexes for the shortest baseline, we have:

$$\begin{aligned} \mathsf{D}^{[\Delta_{12}t,\Delta_{12}\nu]} &= \mathsf{D}^{[\Delta_{23}t,\Delta_{23}\nu]} \\ &= \mathsf{D}^{[\Delta_{34}t,\Delta_{34}\nu]} \\ \mathrm{Dim}\{\mathsf{B}_{kl}^{[\Delta_{12}t,\Delta_{12}\nu]}\} &< \mathrm{Dim}\{\mathsf{B}_{kl}^{[\Delta_{23}t,\Delta_{23}\nu]}\} \\ &< \mathrm{Dim}\{\mathsf{B}_{kl}^{[\Delta_{34}t,\Delta_{34}\nu]}\}. \end{aligned} \tag{4.12}$$

The averaged intervals are functions of baselines' length and direction, with longer baselines having shorter integration times and narrower channels, while shorter baselines are averaged over longer integration times and wider channels.

In Section 3.1.1 we defined the averaged or *resampled* visibilities for a baseline pq as the discrete sum:

$$V_{pqkl}^{\mathrm{M}} = \frac{1}{n_{pq}} \sum_{ij \in \mathsf{B}_{kl}^{pq}} V_{pqij}^{(\mathrm{s})}, \tag{4.13}$$

where $\mathrm{Dim}\{\mathsf{B}_{kl}^{pq}\} = n_{pq}$ is the number of samples that has been averaged. It was then shown in Section 3.1.2 that visibility averaging can be treated as a convolution with a

baseline-dependent boxcar window $\Pi^{[uv]}_{pqkl}$ at the centre of the resampling bin interval

$$\mathcal{V}^{\mathrm{M}}_{pqkl} = \delta_{pqkl}(\mathcal{V} \circ \Pi^{[uv]}_{pqkl}), \tag{4.14}$$

and this results in Eq. (4.15), when imaging the per-baseline visibilities:

$$\mathcal{I}^{\mathrm{D}} = \sum_{pqkl} W_{pqkl}\mathcal{P}_{pqkl} \circ (\mathcal{I} \cdot \mathcal{T}_{pqkl}), \tag{4.15}$$

with the apparent sky $\mathcal{I}$ tapered by the inverse Fourier transform of the baseline-dependent boxcar $\mathcal{T}_{pqkl}$:

$$\mathcal{T}_{pqkl} = \mathcal{F}^{-1}\{\Pi^{[uv]}_{pqkl}\}. \tag{4.16}$$

Now let us determine what Eq. (4.15) becomes for the case of BDA. For BDA, the uv-space boxcar windows, $\Pi^{[uv]}_{pqkl}$, are approximately equal across all baselines and therefore do not depend any more on baseline length and direction. In other words, for any baseline $\alpha\beta \neq pq$,

$$\Pi^{[uv]}_{\alpha\beta kl} \approx \Pi^{[uv]}_{pqkl}. \tag{4.17}$$

Eq. (4.17) is valid and correct only for the case of BDA. One fact to keep in mind is that while the size of all boxcar windows is equally fixed in uv-space, they are however sampled differently. For each baseline the East-West component rotates as a function of time, while the South-North does not. For the same length of time, two equal length baselines with different orientations will sweep unequal uv-distance and therefore result in lower spatial modes being oversampled compared to higher spatial modes (see Figure 4.6), this may result in a different IPR.

FIGURE 4.6: In uv-space the boxcar windows are sampled differently for a fixed length of uv-distance across baselines. (Left) sampling function for the boxcar window of the longest baseline. (Right) sampling function for the boxcar window of the shortest baseline.

However, if the pre-averaged visibilities are sampled at significantly high time and spectral resolution, then we can assume that all these boxcars at different baselines are sampled equally. Considering this assumption, we can write:

$$\mathcal{T}_{pqkl} \approx \mathcal{T}_{\alpha\beta kl}. \tag{4.18}$$

This shows that the time and bandwidth decorrelation represents the effect of a single taper in the image. With this condition, Eq. (4.15) becomes:

$$\mathcal{I}^{\mathrm{D}} \sim \sum_{pqkl} W_{pqkl} \mathcal{P}_{pqkl} \circ \mathcal{I} \cdot \mathcal{T}, \tag{4.19}$$

where $\mathcal{T} \approx \mathcal{T}_{pqkl} \approx \mathcal{T}_{\alpha\beta kl}$ is the smearing response, constant on all baselines: that is, the effect of a single taper on the image.

4.1.2 Compression factor

The compression factor is defined as the ratio between the pre-averaged data (high-res data) size and the averaged data (low-res data) size. In terms of the number of visibility points, the data size of the high-res MS is given by:

$$\text{data size hires} = \frac{n_a(n_a-1)}{2} \times N_{\mathrm{sub}} \times N_t^{\mathrm{hires}} \times N_\nu^{\mathrm{hires}} \times N_{\mathrm{pol}}, \tag{4.20}$$

where n_a is the number of antennas of the interferometer array, N_{sub} the number of sub-bands, N_{pol} is the number of polarisations, N_t^{hires} and N_ν^{hires} are the number of timeslots and channels of the high-res MS respectively. In the case of simple averaging, the data size of the averaged MS is given by:

$$\text{Data size avg} = \frac{n_a(n_a-1)}{2} \times N_{\mathrm{sub}} \times \frac{N_t^{\mathrm{hires}}}{n_t} \times \frac{N_\nu^{\mathrm{hires}}}{n_\nu} \times N_{\mathrm{pol}}, \tag{4.21}$$

where n_t represents the number time bins and n_ν the channels averaged on each baseline. Note that this formulation applies for equal compression across all baselines, therefore

the compression factor is defined by:

$$\text{Compression factor avg} = \frac{\text{Data size hires}}{\text{Data size avg}} \tag{4.22}$$

$$= n_t \times n_\nu. \tag{4.23}$$

The space savings, which are defined as the reduction in size relative to the high-res data size, follow as:

$$\text{space savings avg} = 1 - \frac{\text{Data size avg}}{\text{Data size hires}} \tag{4.24}$$

$$= 1 - \frac{1}{n_t \times n_\nu}. \tag{4.25}$$

Following the analogy for BDA resampling bins, the number of samples in the resampling bin for a baseline pq is defined as $\text{Dim}\{\mathsf{B}_{kl}^{[\Delta_{pq}t,\Delta_{pq}\nu]}\} = n_t^{pq} \times n_\nu^{pq}$, where n_t^{pq} and n_ν^{pq} are the baseline pq number of time and frequency samples respectively. The interferometer array data size in terms of number of visibility points for BDA then follows:

$$\text{Data size bda} = \sum_{pqkl} N_{\text{sub}} \times \frac{N_t^{\text{hires}} \times N_\nu^{\text{hires}}}{\text{Dim}\{\mathsf{B}_{kl}^{pq}\}} \times N_{\text{pol}} \tag{4.26}$$

$$= \sum_{pq} N_{\text{sub}} \times \frac{N_t^{\text{hires}}}{n_t^{pq}} \times \frac{N_\nu^{\text{hires}}}{n_\nu^{pq}} \times N_{\text{pol}}. \tag{4.27}$$

The interferometer array compression factor for BDA is then given by:

$$\text{Compression factor bda} = \frac{n_a(n_a-1)}{2} \times \left(\sum_{pq} \frac{1}{n_t^{pq} \times n_\nu^{pq}} \right)^{-1} \tag{4.28}$$

and the space savings by

$$\text{space savings bda} = 1 - \left(\frac{n_a(n_a-1)}{2} \right)^{-1} \times \sum_{pq} \frac{1}{n_t^{pq} \times n_\nu^{pq}}. \tag{4.29}$$

In the rest of this chapter, we refer to the compression factor as $CF = x_t \times y_\nu$, where x_t is the compression factor in time and y_ν is the compression factor in frequency. The notation $CF = x_t \times 1$ implies the data are compressed only in time by a factor of x_t, while $CF = 1 \times y_\nu$ implies that data are compressed only in frequency by a factor of y_ν. For BDA, the shorter baselines are compressed by much more than CF and the longer baselines by much less. This corresponds to CF for all the baselines with simple

averaging.

4.2 Implementation details

In practice, most existing software implementations assume that the correlation matrix is a time and frequency regular grid. Averaging entries in this correlation matrix over long times for short baselines and short times for long baselines results in an irregular grid. A better idea is to map this irregular grid onto a correlation matrix (i.e. regular grid) by either flagging out the supplementary points, or duplicating the averaged values onto these supplementary points. We will explain these processes in detail in the following sections.

4.2.1 Flagging

Most of the radio interferometric data reduction software has a flagging capability, through which bad data can be flagged and ignored. For BDA, we exploit this capability to force interferometric data reduction software to ignore some entries of the regularly gridded plane (e.g. the correlation matrix). In the flagging procedure, one has to make sure that the resampling bin contains an odd number of data points in time as well as in frequency:

$$\mathrm{Dim}\left\{\mathsf{B}_{kl}^{[\Delta_{\alpha\beta}t,\Delta_{\alpha\beta}\nu]}\right\} = (2k_t+1)(2k_\nu+1), \text{ where } k_t \text{ and } k_\nu \text{ are integers.} \tag{4.30}$$

Here, $2k_t+1$ is the number of visibilities to average in time and $2k_\nu+1$ the number of visibilities to average in frequency. This condition must be verified on all baselines otherwise the average baseline vector may not coincide with the mid time and frequency vector and this will lead to a phase shift. If this condition is satisfied, the average value is assigned to the mid point of the resampling bin interval, i.e. at the $k_t^{th}+1$ and $k_\nu^{th}+1$ visibility point. The other entries of the resampling bin are flagged. This flag will cause missing samples to be ignored during post-processing. Suppose that the data in Figure 4.7 represents 15 time-bins, for three different baselines (the longest, medium length and shortest baselines). Each bin is a 1 s integration sample and 15 s bins are averaged on the shortest baseline, 5 s on the medium length baseline and 3 s on the

longest baseline. The coloured pointers indicate the averaged bins and the black arrows are flagged points. Most of the flagging occurs on the shortest baseline.

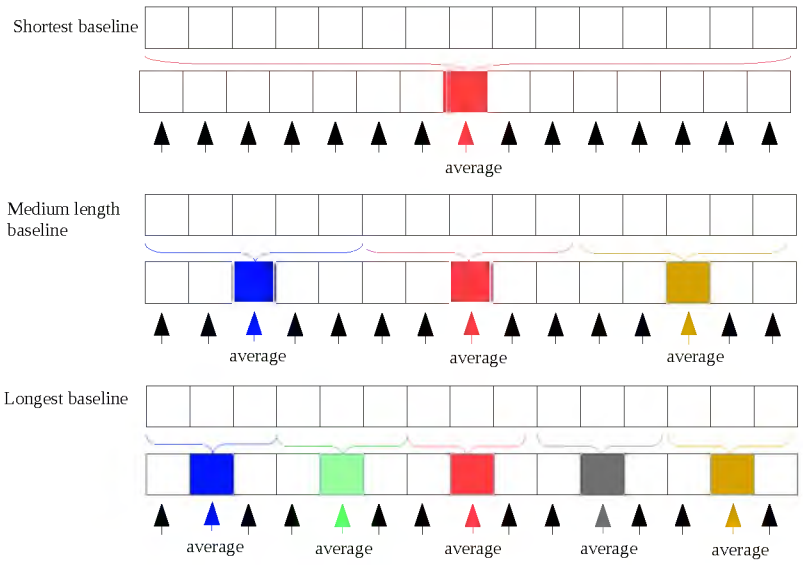

FIGURE 4.7: BDA with flagging: the bins are averaged and the average value is assigned at the centre of the resampling bin interval, while a flagging value is assigned to others points.

4.2.2 Duplication

This method consists of duplicating the average value at all entries of the resampling bins. While this process is easier to implement than the flagging method, it may not serve the purpose of data compression and/or quick computation for post-processing. It is easier to implement in the sense that one may not care or always verify that the number of visibility points in the resampling bin is an odd number. Furthermore, the data size of the resulting MS remains the same as the pre-averaged MS, since all values are duplicated along the pre-averaged MS. This method may be used in practice for cases where one does not want to estimate the averaged uv-coordinates from the pre-averaged MS. Similarly to the example in Figure 4.7, Figure 4.8 represents the duplicate method. Pointers with the same colour indicate the duplicate averaged values throughout the resampling bin intervals.

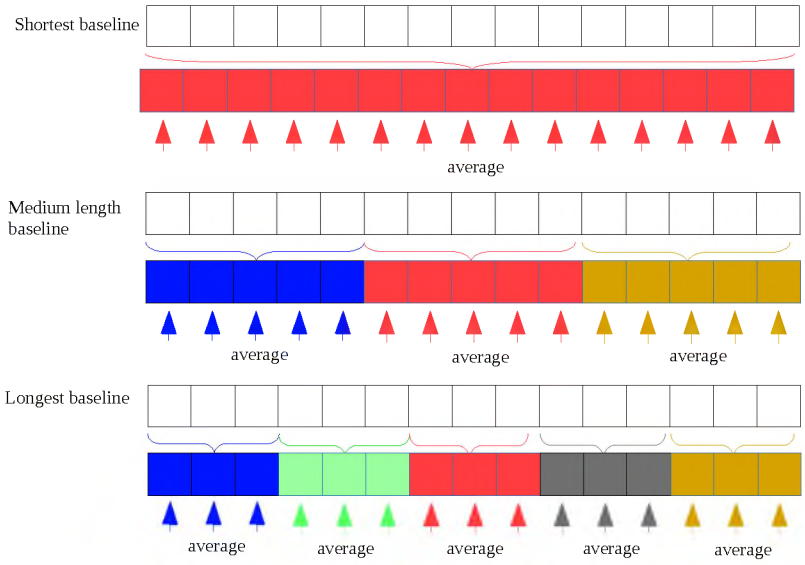

FIGURE 4.8: BDA with duplication: the bins are averaged and the average value is assigned at all points of the resampling bin interval.

4.2.3 Semi-duplication and flagging

This method consists of combining the flagging and the duplicate methods in order to benefit from their full advantages. In so doing, we seek both data compression and quick computation, while on the other hand the implementation is easier to handle. The idea is to duplicate the averaged bin along two central entries of the resampling bin if the total number of entries within this resampling bin is even, otherwise the averaged bin is assigned only to the central bin of the resampling bin. Any other entry is then flagged, as shown in Figure 4.9.

4.3 Simulations and results

Having explored the mathematics and difficulties behind the implementation of BDA, we now turn to the simulation aspects. We consider our tests on a sky model of 1 Jy point source at various sky positions, with no noise or other corruptions included. We evaluate the efficiency of a BDA correlator using two different procedures. Firstly, we simulate the source at a fixed sky position, apply BDA and measure the averaging effects separately on each baseline. Secondly, we simulate the point source at various angular distances from the phase centre and apply BDA and BDWFs across an equal

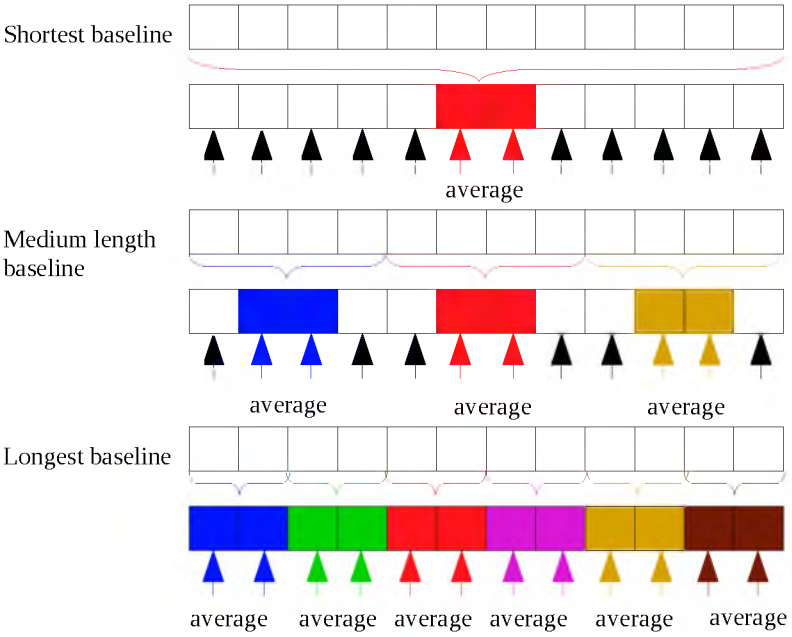

FIGURE 4.9: BDA with semi-duplication and flagging: the bins are averaged and the average value is duplicated only at the two central points of the resampling bin interval and the other bins are flagged.

uv-distance, thereby evaluating the interferometer array cumulative decorrelation effects on all baselines. Following the same procedure used in Section 3.3, we measure the source peak amplitude in each image after averaging. Since each dirty image corresponds to a single source, the peak gives us the degree of smearing associated with a given averaging method and compression factor.

4.3.1 Source amplitude vs. East-West baseline length

We simulate two high-res MSs each with a source at 30 arcmin from the phase centre of the observation. Furthermore, we generate two low-res MSs to receive the resampled visibilities. We present the amplitude loss as a function of East-West baseline length. The results of BDA and simple averaging are compared in Figure 4.10.

- The first MS consists of 10 frequency channels of 125 kHz width, and 2800 timeslots of 1 s integration time. In this test (top panel of Figure 4.10), the compression factor is fixed to $CF = 100 \times 1$ for both BDA and simple averaging.

- The second MS consists of 10 timeslots of 1 s integration, and 2400 frequency channels of 125 kHz. The compression factor is fixed to $CF = 1 \times 80$ both for BDA and simple averaging (bottom panel of Figure 4.10).

It is clearly noticeable in Figure 4.10 that on shorter baselines, the smearing rates of BDA and simple averaging are equivalent. The physical interpretation for this is that, at these compression factors, visibilities do not suffer from decorrelation on the shorter baselines. On the other hand, BDA reduces smearing on the longer baselines compared to simple averaging. This is because fewer samples are averaged on the longer baselines when applying BDA, causing a decrease in decorrelation.

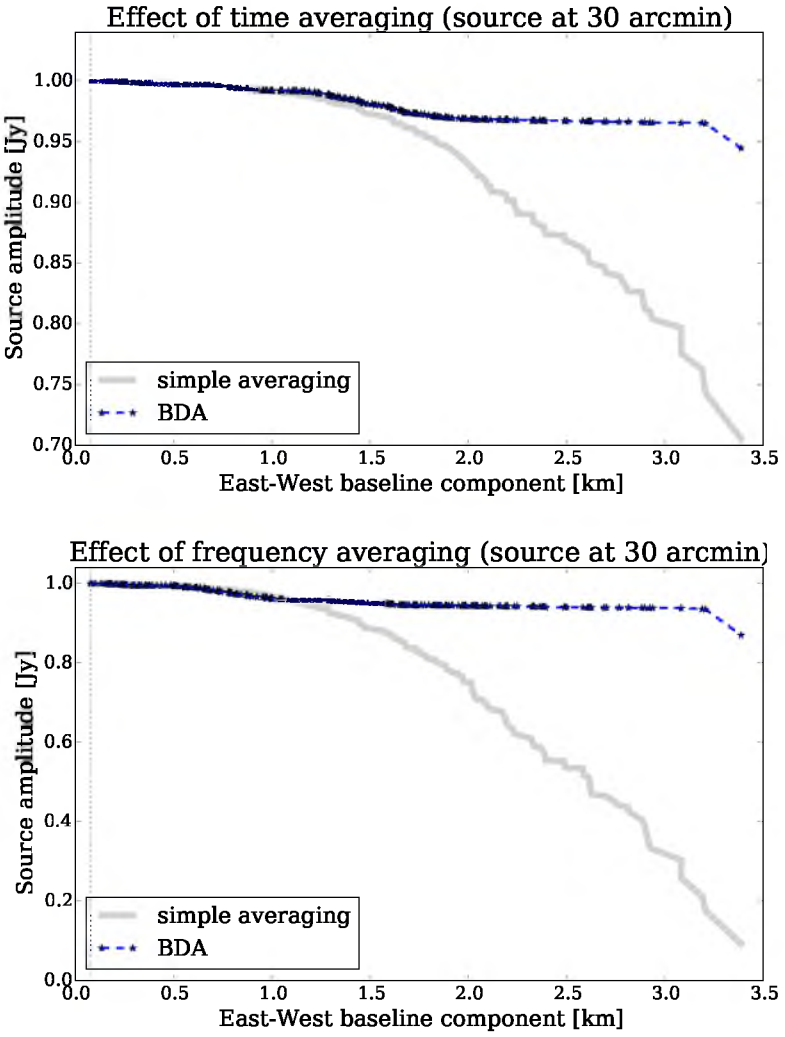

FIGURE 4.10: Amplitude loss: the apparent intensity of a 1 Jy source, as seen by JVLA-C at 1.4 GHz, as a function of East-West baseline components. (Top) Compression carried out only in time with compression factor fixed to 100; (Bottom) Compression is carried out only in frequency with compression factor fixed to 80.

4.3.2 Source amplitude vs. distance from the phase centre

In this section, we simulate a high-res MS corresponding to a 2800 s synthesis at 1 s integration, with 300 MHz total bandwidth centred at 1.4 GHz, divided into 2400 channels of 125 kHz each. The sky model is a single 1 Jy point source at a given distance from the phase centre. We then generate four MSs to receive the resampled visibilities:

(a) We prepare three MSs to receive the resampled visibilities for 25 s × 2.5 MHz, 50 s × 5 MHz and 100 s × 10 MHz yielding three compression factors of CF=25×20, CF=50×40 and CF=100×80 respectively.

(b) A fourth MS is prepared to receive the sampled visibilities for BDA. This MS is a copy of the high-res MS where one of the BDA implementation methods described in Section 4.2 is applied. Three compression factors are adopted for the BDA: CF=25×20, CF=50×40 and CF=100×80. It is important to note that, in the case of overlapping BDWFs, all the bins[1] reserved for the overlapping filters are flagged at the end of the resampled procedure, i.e. before the resampled visibilities are imaged.

(c) The synthesis time for these MSs is 2000 s with 200 MHz total bandwidth. It is quite remarkable that the synthesis time and the total bandwidth are less than the high-res MS. As stated in Section 3.2.1, for overlapping BDWFs a number of time and frequency samples must be allowed at the observation starting time and/or frequency and the observation ending time and/or frequency. These overlapping samples correspond here to 800 s and 100 MHz in time and frequency respectively.

As was discussed in Section 3.2, resampling with BDWFs also results in a different image-plane response. For example, with a sinc-1×1, longer baselines correspond to a sinc-like window function and boxcar taper in the image-plane. For shorter baselines, a sinc-1×1 corresponds to a boxcar and a sinc-like taper in the image-plane, and therefore increased smearing. Applying BDA to BDWFs may lead to a more optimal response, given that the short baselines and the long baselines are processed with similar visibility plane sinc-like filters. This implies an equal image-plane boxcar taper for each of the baselines.

To distinguish the case of BDA applied to overlapping BDWFs from non-overlapping ones, we will designate the overlapping BDWFs as bda-WF-α×β, i.e. for each baseline

[1]These bins are the samples at the starting time-frequency and ending time-frequency of the duplicated high-res MS prepared for resampling using overlapping BDWFs

pq the resampling bin is now of size $\alpha\Delta_{pq}t \times \beta\Delta_{pq}\nu$, where α and β are the overlap factors.

Figures 4.11, 4.12 and 4.13 depict the performance of simple averaging with resampling bin sizes of 25 s × 2.5 MHz, 50 s × 5 MHz and 100 s × 10 MHz compared to BDA:

(a) For CF=25×20 (Figure 4.11); two types of BDA are applied to BDWFs:

bda-sinc-4×4 i.e. an overlapping sinc window

bda-bessel-4×4 i.e. an overlapping Bessel window

(b) For CF=50×40 and CF=100×80 (Figures 4.12 and 4.13); two types of BDA are applied to BDWFs:

bda-sinc-2×2 i.e. an overlapping sinc window

bda-bessel-2×2 i.e. an overlapping Bessel window

These BDAs applied to BDWFs are turned to two different FoI settings, as indicated by the plots legend: 2 deg and 4 deg.

Based on the above discussion, we can now interpret some of our main results. These results can be alternatively appreciated by regarding the performance of BDA applied to BDWFs:

It is shown in Figure 4.11 that BDA with CF=25×20 provides good results in flux recovery, i.e. for 5% smearing this gives us a field with radius 1.4°, while simple averaging with CF=25×20 can only recover a field with radius 0.9°. We can also note that at this compression factor, BDA achieved a similar far field suppression performance compared to simple averaging.

Figures 4.12 and 4.13 show the resampling with compression factors of CF=50×40 and CF=100×80 respectively. All the BDAs provide better source recovery compared to simple averaging, and source suppression is similar to regular averaging at the same rate. Note how the far field pattern behaviour of BDAs at these regimes are closer to a sinc-like pattern. This is easily understood by looking at the IPR for BDA (refer to Section 4.18): the boxcar windows are similar for all baselines, larger compression factors correspond to larger boxcars that become more "sinc-like" in the image-plane.

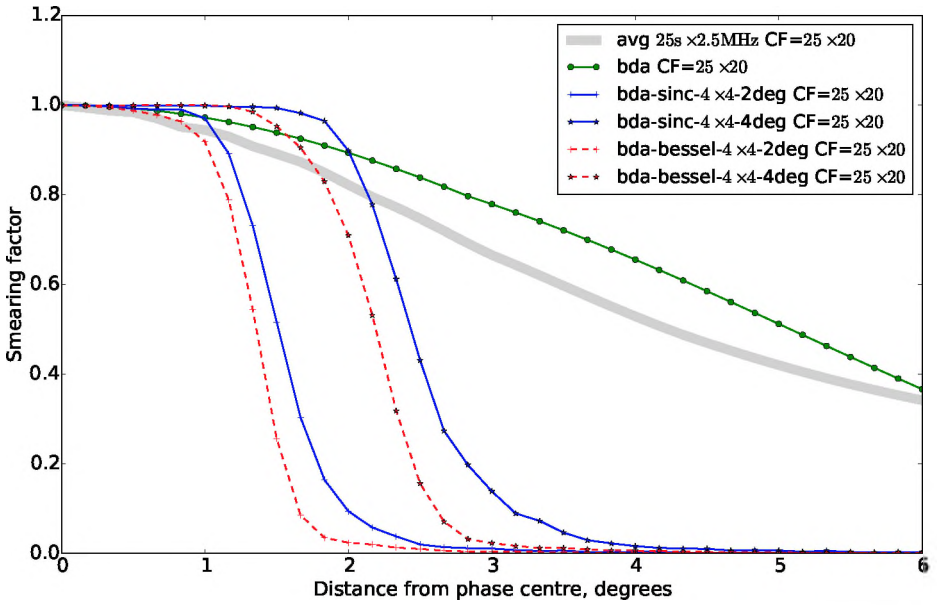

FIGURE 4.11: Amplitude loss: the apparent intensity of a 1 Jy source as seen by JVLA-C at 1.4 GHz as a function of distance from phase centre, for simple averaging with 25 s × 2.5 MHz bins, and for BDA, and BDA applied to BDWFs. The compression factor is fixed to $CF = 25 \times 20$ for all the resampling methods.

All the BDAs applied to overlapping BDWFs provide excellent performance at these compression factors compared to regular averaging or BDA: the smearing performance across the FoI is less than 1%, and out-of-FoI suppression is almost two orders of magnitude higher than simple averaging or BDA. Note the tapering behaviour for BDA applied to overlapping BDWFs at the different compression factors. As the compression factor increases, the response of BDA applied to overlapping BDWFs becomes flat: this clearly illustrates their excellent performance. The reason for this is that the sinc-like window function is applied across an equal *uv*-distance to all baselines. For larger compression factors the sinc-like window function becomes more proximate to the "sinc", which results in a more optimal "boxcar-like" in the image domain.

4.4 Conclusion

For a fixed time length, a long baseline will cover a longer track in visibility space compared to a shorter baseline, which results in lower spatial modes being oversampled compared to higher spatial modes. This necessitates the use of BDA to optimize the

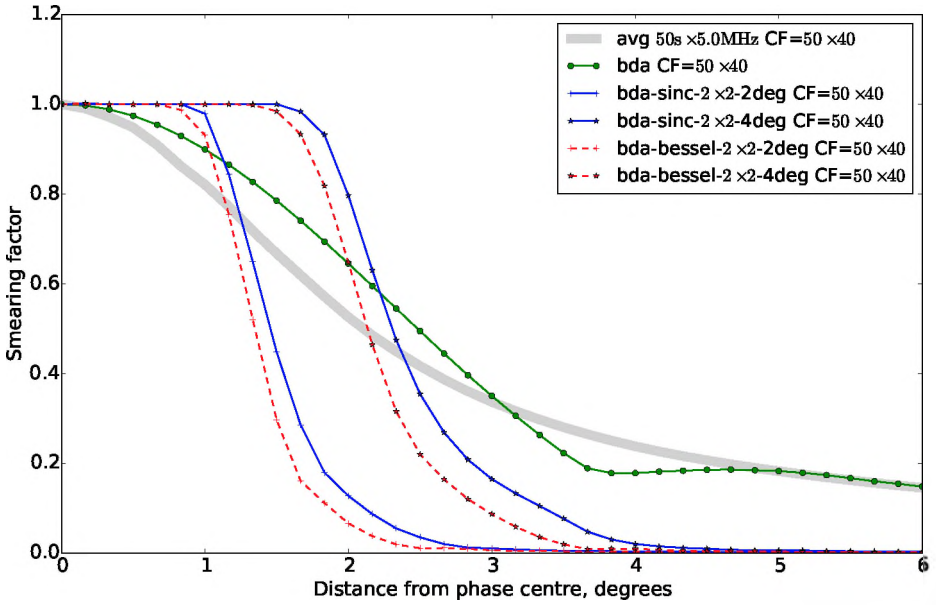

FIGURE 4.12: Amplitude loss: the apparent intensity of a 1 Jy source as seen by JVLA-C at 1.4 GHz as a function of distance from phase centre, for simple averaging with 50 s × 5 MHz bins, and for BDA, and BDA applied to BDWFs. The compression factor is fixed to $CF = 50 \times 40$ for all the resampling methods.

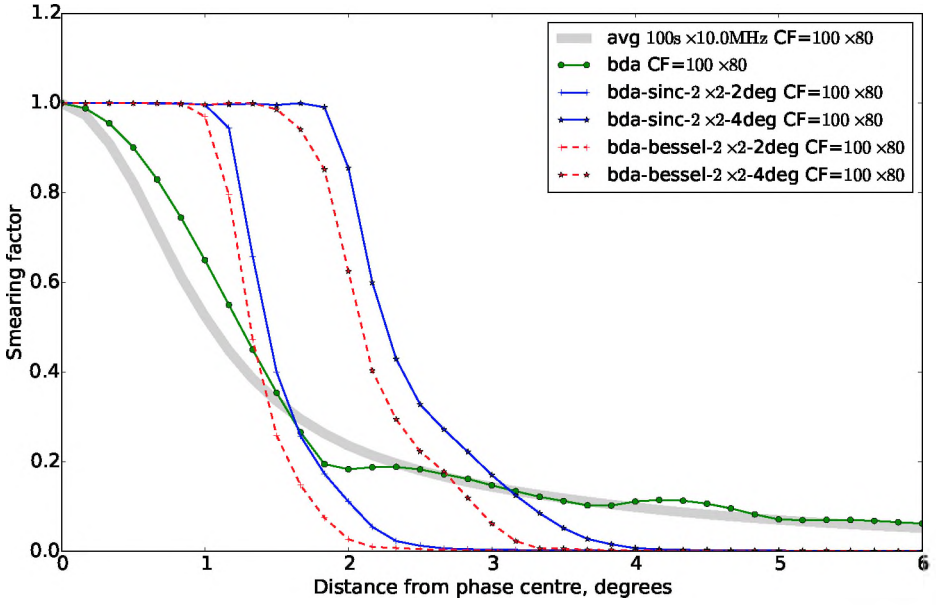

FIGURE 4.13: Amplitude loss: the apparent intensity of a 1 Jy source as seen by JVLA-C at 1.4 GHz as a function of distance from phase centre, for simple averaging with 100 s × 10 MHz bins, and for BDA, and BDA applied to BDWFs. The compression factor is fixed to $CF = 100 \times 80$ for all the resampling methods.

image-plane response for each baseline. The use of BDA and BDWFs for equal uv-space averaging has been investigated analytically and via simulations.

We have established that BDA by itself can only achieve data compression but not FoI shaping: BDA does decrease smearing over the FoI, while on the other hand, sources out-of-FoI are not suppressed compared to simple averaging. We have found that combining BDA with BDWFs results in excellent tapering behaviour, which can decrease smearing to about 1% or less over a selected FoI, with out-of-FoI source suppression almost two orders of magnitude higher than simple averaging, while the data are compressed at the same rate.

We should note that BDA also distorts the PSF, which reacts differently compared to simple averaging. The next chapter will be devoted to deriving this PSF at different sky positions for the case of BDA, simple averaging and BDWFs to integrate them into an existing deconvolution imaging algorithm.

Chapter 5

Fast Algorithms for Approximating the Offset PSF Response and Implementation in DDFacet

In Chapter 3 and 4 we showed that averaging distorts images of astronomical sources, in particular off-axis sources. It was shown in Section 1.6 that averaging also distorts the PSF differently at each sky position. In interferometric imaging, deconvolution is used to remove the PSF from the dirty image. To account for a position-dependent PSF, deconvolution algorithms would need to evaluate the PSF at each minor loop iteration in order to remove it from the dirty image in a correct manner. With a simple CLEAN algorithm [Clark, 1980] and its derivatives [Cornwell, 1983, Steer et al., 1984], deconvolution is regarded as an iterative approach, where the brightest pixel value is found step by step and subtracted from the image until the loop reaches a given threshold. During each step, the brightest pixel value is convolved by the PSF of the instrument before the result of the convolution is subtracted from the image. Note that in contrast to a simple CLEAN algorithm, which uses the same PSF to deconvolve all the sources, deconvolution with position-dependent PSFs must evaluate the distorted PSF at each successive pixel's position in oder to subtract accurately. This makes deconvolution

This chapter draws extensively on: M. T. Atemkeng, O. M. Smirnov, C. Tasse, G. Foster and J. Jonas, *Towards a position dependent deconvolution scheme for radio interferometric imaging.* Astronomy & Astrophysics (A&A), in preparation. It is acknowledged that some of the text in this chapter will therefore "match" that of the article. The reference in this footnote serves as a general reference for all such text.

with position-dependent PSFs a complex task to handle compared to a simple CLEAN algorithm, given that it is computationally expensive to account for the distorted PSF at each iteration. The brute-force computational effort or complexity to evaluate the position-dependent PSF during the CLEAN iterations scales as $\mathcal{O}(N_{\text{pix}}^4 \log_2 N_{\text{pix}})$, where N_{pix}^2 is the number of pixels or cells in the image (we refer the reader to Section 5.4 for an extensive discussion).

In this chapter, we use the phase gradient and the PSF of the instrument to establish a mathematical framework that approximates the position-dependent PSFs across the image. The proposed mathematical framework has lower computational complexity compared to the brute-force approach. We develop a position-dependent PSF deconvolution scheme for DDFacet [Tasse et al., b, in preparation]. This is a novel and unique approach in radio interferometric imaging to account for the PSF variation in a deconvolution algorithm. To begin with the approximation, we describe the position-dependent nature of the PSF and estimate the computational costs of a brute-force approach to computing it.

5.1 Motivation for position-dependent PSFs

The PSF is the response of an interferometer array towards a point source of unit flux; it is the Fourier transform of the *uv* sampling function. Radio interferometric arrays do not measure the true sky directly, but the true sky convolved with the array PSF. The PSF can be treated as the transfer function between the true sky and the observed *dirty* image. In many cases, the *uv*-coverage may be poorly sampled because of observational constraints, which lead to regions in *uv*-space where there are no measurements. For wide field arrays (refer to Section 5.2) the PSF can also vary substantially due to direction-dependent instrument effects across the sky even though each bin is sampled at high temporal and spectral resolution; this is ignored in this study.

Since the effect of the PSF is inherently present in the observed data, it needs to be removed in order for scientific analysis to be carried out. A good deconvolution algorithm must subtract the PSF sidelobes from the measured sky (i.e. dirty image) accurately, otherwise this will result in imaging artefacts. Therefore, any reconstruction of the true

sky from the dirty image must accurately remove the sidelobes for the near and the far field sources.

In this chapter, we adopt the following terminology: PSF is the PSF at the field centre of the observation and pseudo-PSF is the PSF as seen by an off-axis source.

Below, a mathematical model to understand the pseudo-PSFs is proposed. The mathematical formulations are well documented in radio interferometry literature [Woody, 2001a,b] but it is useful to present them for subsequent use. I start the formulation from the visibilities of the entire sky and then restrict this to the visibilities of a single point source.

Let us reconsider the van Cittert-Zernike theorem [Thompson, 1999, Thompson et al., 2001] for a two-element interferometer p and q in response to a source with spectral brightness distribution $\mathcal{I}$ as a function of the pointing direction $\boldsymbol{l}$. The visibility function $\mathcal{V}_{pq}$ is given by:

$$\mathcal{V}_{pq}(u,v) = \int \mathcal{I}(\boldsymbol{l}) \mathrm{e}^{-2\pi i \boldsymbol{u}_{pq} \boldsymbol{l}} \mathrm{d}\Omega, \tag{5.1}$$

where $\boldsymbol{l} = (l, m, n)$ is the vector that represents the position in the sky, and $d\Omega$ is the solid angle. Recall from Section 3.1.1 that $\boldsymbol{u}_{pq} = \boldsymbol{u}_{pq}(t,\nu) = (u_{pq}, v_{pq}, w_{pq})$ is the baseline vector in wavelength. Parameterising Eq. (5.1) in terms of sky coordinates results in:

$$\mathcal{V}_{pq}(u,v) = \int \mathcal{I}(\boldsymbol{l}) \mathrm{e}^{-2\pi i (u_{pq}(t,\nu) l + v_{pq}(t,\nu) m + w_{pq}(t,\nu) n)} \mathrm{d}\boldsymbol{l}, \tag{5.2}$$

where $n = \sqrt{1 - l^2 - m^2} - 1$ (refer to [Burke and Graham-Smith, 2010] for an extensive discussion for deriving Eq. (5.2)).

However, in reality an interferometer is non-ideal, in the sense that a measurement is the averaged visibility over some time-frequency bins denoted by the time and frequency *sampling intervals* $\Delta t, \Delta\nu$, known as the *sampling bin*

$$\mathsf{B}_{kl}^{[\Delta t \Delta \nu]} = \left[t_k - \frac{\Delta t}{2}, t_k + \frac{\Delta t}{2} \right] \times \left[\nu_l - \frac{\Delta \nu}{2}, \nu_l + \frac{\Delta \nu}{2} \right], \tag{5.3}$$

where k and l represent the indices of the centre time and frequency bins respectively.

The averaged measurement can be represented by an integration:

$$V^{\mathrm{M}}_{pqkl} = \frac{1}{\Delta t \Delta \nu} \iint\limits_{\mathsf{B}_{kl}^{[\Delta t \Delta \nu]}} \mathcal{V}_{pq}(\boldsymbol{u}(t,\nu))\mathrm{d}\nu\mathrm{d}t. \tag{5.4}$$

Eq. (5.4) is an essential step to derive the pseudo-PSF. If we assume that the baseline pq tracks a single point source located at position $\boldsymbol{l} = (l, m, n)$ with radius $||\boldsymbol{l}||$ from the phase centre and the source brightness is 1 Jy at position $\boldsymbol{l} = (l, m, n)$ then Eq. (5.2) simplifies to

$$\mathcal{V}_{pq}(u, v) = \mathrm{e}^{-2\pi i(u_{pq}(t,\nu)l + v_{pq}(t,\nu)m + w_{pq}(t,\nu)n)}. \tag{5.5}$$

Eq. (5.4) can thus be rewritten as an average of all the exponentials over the resampling bin $\mathsf{B}_{kl}^{[\Delta t \Delta \nu]}$

$$V^{\mathrm{M}}_{pqkl} = \frac{1}{\Delta t \Delta \nu} \iint\limits_{\mathsf{B}_{kl}^{[\Delta t \Delta \nu]}} \mathrm{e}^{-2\pi i(u_{pq}(t,\nu)l + v_{pq}(t,\nu)m + w_{pq}(t,\nu)n)}\mathrm{d}\nu\mathrm{d}t. \tag{5.6}$$

Eq. (5.6) reveals the following:

(a) If the source is at position $l = 0, m = 0$ (phase centre of the observation) then the inverse Fourier transform of Eq. (5.6) is equivalent to the inverse Fourier transform of the sampling function, i.e. the PSF at the phase centre as seen by the baseline pq.

(b) For sources at positions $(l, m) \neq (0, 0)$, the inverse Fourier transform of Eq. (5.6) gives a brute-force computation of the pseudo-PSF for the baseline pq.

(c) The inverse Fourier transform of the sum of V^{M}_{pqkl} over all baselines is referred to as *brute-force computation* for the pseudo-PSF of a source at (l, m) as seen by the array. The brute-force computation results in the exact pseudo-PSF, but the process is computationally costly. For example, with the FFT, $\mathcal{O}(N_{pq}N_tN_\nu \log_2 N_{\mathrm{pix}})$ is the complexity to evaluate a unique pseudo-PSF where N^2_{pix} stands for the number of pixels or cells in the image; N_{pq} is the number of baselines; N_t and N_ν are the number of time samples and frequency channels respectively. Assuming that N_{src} is the number of sources in the image that defines the number of pseudo-PSFs to evaluate, the total complexity for evaluating the brute-force pseudo-PSFs is $\mathcal{O}(N_{\mathrm{src}}N_{pq}N_tN_\nu \log_2 N_{\mathrm{pix}})$ and can even reach $\mathcal{O}(N^4_{\mathrm{pix}} \log_2 N_{\mathrm{pix}})$ in the worst case

where each pixel in the image is a source, i.e. $N_{\rm src} \sim N_{\rm pix}^2$ and the total number of visibility samples $N_{pq}N_tN_\nu \sim N_{\rm pix}^2$.

Figure 1.4 shows the inverse Fourier transform of the sum of $V_{pqkl}^{\rm M}$ over all baselines for a source at the phase centre and at 1°, 2°, 3° away from the phase centre. The data as observed by the JVLA in C-configuration was simulated with 1 s integration time for a period of 1 hour, with 100 kHz frequency channels for a total bandwidth of 10 MHz; further resampled to 100 s integration time and 5 MHz channel width. Note the smearing nature and the position-dependent smearing behaviour of the pseudo-PSFs.

In the next section, I present two algorithms that generate the pseudo-PSFs with a complexity of $\mathcal{O}\big(N_{\rm src}N_{pq}N_t^{\rm re}N_\nu^{\rm re}\log_2 N_{\rm pix}\big)$ and $\mathcal{O}\big(N_{pq}N_tN_\nu\log_2 N_{\rm pix}\big)+\mathcal{O}\big(\gamma\big)$, which can give much better scaling than $\mathcal{O}\big(N_{\rm src}N_{pq}N_tN_\nu\log_2 N_{\rm pix}\big)$, assuming $N_t^{\rm re}N_\nu^{\rm re} \ll N_tN_\nu$ and $\gamma \ll N_{pq}N_tN_\nu\log_2 N_{\rm pix}$.

5.2 Fast derivation of pseudo-PSFs

I now proceed with a step-by-step algorithm to approximate the pseudo-PSF. To begin with, the fringe induced by the w-term is denoted by $\mathcal{G}$

$$\mathcal{G}(t,\nu) = \mathrm{e}^{-2\pi i w_{pq}(t,\nu)n}. \tag{5.7}$$

We consider the w-term for both coplanar ($w_{pq}=0$) and non-coplanar ($w_{pq}\neq 0$) arrays, and for small FoIs ($\sqrt{1-l^2-m^2}\approx 1$) or wide FoIs ($\sqrt{1-l^2-m^2}\ll 1$).

Using Eq. (5.7), we can rewrite Eq. (5.6) as

$$V_{pqkl}^{\rm M} = \frac{1}{\Delta t\Delta\nu}\iint\limits_{\mathsf{B}_{kl}^{[\Delta t\Delta\nu]}} \mathcal{G}(t,\nu)\mathrm{e}^{-2\pi i(u_{pq}(t,\nu)l+v_{pq}(t,\nu)m)}\mathrm{d}t\mathrm{d}\nu. \tag{5.8}$$

Similar to Section 3.1.2, if Π is a normalised boxcar window function given by:

$$\Pi(t,\nu) = \begin{cases} \frac{1}{\Delta t\Delta\nu}, & |t|\leq \Delta t/2, \quad |\nu|\leq \Delta\nu/2 \\ 0, & \text{otherwise,} \end{cases} \tag{5.9}$$

and $\mathcal{W}$ is a weighted sampling function, then Eq. (5.8) can be reformulated as a 2-D Fourier transform

$$V^{\mathrm{M}}_{pqkl} = \iint_{\infty} \mathcal{W}(t-t_k, \nu-\nu_l)\Pi(t-t_k, \nu-\nu_l)\mathcal{G}(t,\nu)\mathrm{e}^{-2\pi i(u_{pq}(t,\nu)l+v_{pq}(t,\nu)m)}\mathrm{d}t\mathrm{d}\nu. \quad (5.10)$$

Using the convolution theorem Eq. (5.10) results in

$$V^{\mathrm{M}}_{pqkl} = \widetilde{\mathcal{G}}_{t\nu} \circ \left(\widetilde{\mathcal{W}}_{t-t_k,\nu-\nu_l}\right) \circ \left(\widetilde{\Pi}_{t-t_k,\nu-\nu_l}\right) \quad (5.11)$$

$$= \widetilde{\mathcal{G}}_{t\nu} \circ \left(\widetilde{\delta}_{t-t_k,\nu-\nu_l}\widetilde{\mathcal{W}}_{t\nu}\right) \circ \left(\widetilde{\delta}_{t-t_k,\nu-\nu_l}\widetilde{\Pi}_{t,\nu}\right), \quad (5.12)$$

where $\circ$ and $\widetilde{\ }$ denote the convolution operator and the Fourier transform respectively. The Dirac delta function δ comes from,

$$\mathcal{W}(t-t_k, \nu-\nu_l) = \mathcal{W}_{t-t_k,\nu-\nu_l} \quad (5.13)$$

$$= \delta_{t-t_k,\nu-\nu_l} \circ \mathcal{W}_{t,\nu} \quad (5.14)$$

and

$$\Pi(t-t_k, \nu-\nu_l) = \Pi_{t-t_k,\nu-\nu_l} \quad (5.15)$$

$$= \delta_{t-t_k,\nu-\nu_l} \circ \Pi_{t,\nu}. \quad (5.16)$$

Hence, using the convolution theorem we have $\widetilde{\mathcal{W}}_{t-t_k,\nu-\nu_l} = \widetilde{\delta}_{t-t_k,\nu-\nu_l}\widetilde{\mathcal{W}}_{t,\nu}$ and $\widetilde{\Pi}_{t-t_k,\nu-\nu_l} = \widetilde{\delta}_{t-t_k,\nu-\nu_l}\widetilde{\Pi}_{t,\nu}$. Let us denote the phase gradient $\widetilde{\delta}_{t-t_k,\nu-\nu_l} = \mathrm{e}^{-2\pi i(u_{pq}(t_k,\nu_l)l+v_{pq}(t_k,\nu_l)m)}$ by $\mathcal{R}_{t_k,\nu_l}$.

Likewise, Eq. (5.12) can be written as a convolution in uv-space:

$$V^{\mathrm{M}}_{pqkl} = \widetilde{\mathcal{G}}_{\boldsymbol{u}_{pq}(t,\nu)} \circ \left(\mathcal{R}_{\boldsymbol{u}_{pq}(t_k,\nu_l)}\widetilde{\mathcal{W}}_{\boldsymbol{u}_{pq}(t,\nu)}\right) \circ \left(\mathcal{R}_{\boldsymbol{u}_{pq}(t_k,\nu_l)}\widetilde{\Pi}_{\boldsymbol{u}_{pq}(t,\nu)}\right). \quad (5.17)$$

The inverse Fourier transform of Eq. (5.17) is the pseudo-PSF for a single visibility sample.

5.2.1 Visibility domain approximation

In the following, we describe the algorithm to approximate the visibilities of the pseudo-PSFs relative to the phase gradient. Since averaging and resampling with BDWFs distort the pseudo-PSFs differently, our algorithm is able to approximate the pseudo-PSFs in both cases, i.e. for simple averaging and resampling with BDWFs. To avoid confusion in terminology, V^{M} now denotes the resampling visibility with simple averaging or BDWFs and $V^{(\mathrm{dis})}$, the approximated version of V^{M}. If we consider $\mathcal{X}$ as a BDWF, then $\mathcal{X} \equiv \Pi$ gives the case of simple averaging. The averaged visibility in Eq. (5.17) can be approximated in terms of the phase changes in time and frequency as:

$$V_{pqkl}^{(\mathrm{dis})} = \widetilde{\mathcal{G}}_{\boldsymbol{u}_{pq}(t_k,\nu_l)} \circ \left(\mathcal{R}_{\boldsymbol{u}_{pq}(t_k,\nu_l)} \widetilde{\mathcal{W}}_{\boldsymbol{u}_{pq}(t_k,\nu_l)} \right) \circ \left(\mathcal{R}_{\boldsymbol{u}_{pq}(t_k,\nu_l)} \widetilde{\mathcal{X}}_{\Delta\Psi_{pq},\Delta\Phi_{pq}} \right), \tag{5.18}$$

where $\Delta\Psi_{pq}$ and $\Delta\Phi_{pq}$ are half of the phase difference in time and frequency respectively. I define these phases as

$$\begin{aligned} \Delta\Psi_{pq} =& \frac{2\pi}{2} \Big((u_{pq}(t_s,\nu_l) - u_{pq}(t_e,\nu_l) l + (v_{pq}(t_s,\nu_l) - v_{pq}(t_e,\nu_l)) m \Big) \\ =& \pi \Delta \boldsymbol{u}_{pq}(t,\nu_l) \boldsymbol{l} \end{aligned} \tag{5.19}$$

$$\begin{aligned} \Delta\Phi_{pq} =& \frac{2\pi}{2} \Big((u_{pq}(t_k,\nu_s) - u_{pq}(t_k,\nu_e)) l + (v_{pq}(t_k,\nu_s) - v_{pq}(t_k,\nu_e)) m \Big) \\ =& \pi \Delta \boldsymbol{u}_{pq}(t_k,\nu) \boldsymbol{l}, \end{aligned} \tag{5.20}$$

where $t_s = t_k - \frac{\Delta t}{2}$, $t_e = t_k + \frac{\Delta t}{2}$, $\nu_s = \nu_l - \frac{\Delta\nu}{2}$ and $\nu_e = \nu_l + \frac{\Delta\nu}{2}$ are the starting time, ending time, starting frequency and ending frequency of the sampling bin. Eq. (5.18) represents the approximated single visibility sample of the baseline pq during the integrations Δt and $\Delta\nu$.

From the approximation, if $\mathcal{X}$ is treated as a 2-D boxcar window function Π (simple averaging in time and frequency) then

$$\widetilde{\mathcal{X}}_{\Delta\Psi_{pq},\Delta\Phi_{pq}} = \widetilde{\Pi}_{\Delta\Psi_{pq},\Delta\Phi_{pq}} \tag{5.21}$$

$$= \mathrm{sinc}(\Delta\Psi_{pq}) \mathrm{sinc}(\Delta\Phi_{pq}), \tag{5.22}$$

where the 2-D sinc comes from the 2-D Fourier transform of a boxcar window function (see the proof in Appendix B).

In the brute-force computation of the pseudo-PSFs, each pre-averaged bin contributes equally to the computation. In our method, all the pre-averaged visibility bins at ($t_{i\neq k}$, $\nu_{j\neq l}$) are discarded during the approximation. Only the visibility bin at (t_k, ν_l) is used for the approximation, resulting in cheaper computation compared to the brute-force method (refer to Section 5.4 for the computing time estimate).

5.2.2 Image plane approximation

In the previous section, I presented an algorithm to approximate the pseudo-PSFs using the phase gradient. In this section, I describe an alternative algorithm to approximate the pseudo-PSFs relative to the PSF at the phase centre. The PSF is the inverse Fourier transform of the sampling function, sampled by a number of baselines pq at discrete time/frequency bins. Inverting the Fourier transform of the sum over all baselines of Eq. (5.17) and sampled at each kl results in the pseudo-PSF at (l, m)

$$\mathcal{P}^{(dis)} = \mathcal{F}^{-1}\{\sum_{pqkl} V^{\mathrm{M}}_{pqkl}\}. \tag{5.23}$$

Applying the convolution theorem to Eq. (5.23)

$$\mathcal{P}^{(dis)} = \sum_{pqlk} \mathcal{F}^{-1}\left\{\mathcal{R}_{\boldsymbol{u}_{pq}(t_k,\nu_l)}\widetilde{\mathcal{X}}_{\boldsymbol{u}_{pq}(t,\nu)}\right\}\mathcal{F}^{-1}\left\{\widetilde{\mathcal{G}}_{\boldsymbol{u}_{pq}(t,\nu)} \circ \left(\mathcal{R}_{\boldsymbol{u}_{pq}(t_k,\nu_l)}\widetilde{\mathcal{W}}_{\boldsymbol{u}_{pq}(t,\nu)}\right)\right\}, \tag{5.24}$$

where

$$\mathcal{P}_{pqkl} = \mathcal{F}^{-1}\left\{\widetilde{\mathcal{G}}_{\boldsymbol{u}_{pq}(t,\nu)} \circ \left(\mathcal{R}_{\boldsymbol{u}_{pq}(t_k,\nu_l)}\widetilde{\mathcal{W}}_{\boldsymbol{u}_{pq}(t,\nu)}\right)\right\}, \tag{5.25}$$

represents the PSF for a single visibility sample for the baseline pq, and the term

$$\mathcal{K}_{pqkl} = \mathcal{F}^{-1}\left\{\mathcal{R}_{\boldsymbol{u}_{pq}(t_k,\nu_l)}\widetilde{\mathcal{X}}_{\boldsymbol{u}_{pq}(t,\nu)}\right\}, \tag{5.26}$$

is the distortion function for this single visibility sample. We can write:

$$\mathcal{P}^{(dis)} = \sum_{pqkl} \mathcal{K}_{pqkl}\mathcal{P}_{pqkl}. \tag{5.27}$$

Let us denote by $\mathcal{P}$ the PSF at the phase centre of the array (i.e. $\mathcal{P} = \sum_{pqkl}\mathcal{P}_{pqkl}$). Let us further assume that one can find a function $\mathcal{K}$ representing the cumulative distortion

effects over all the visibility bins and treat Eq. (5.27) as a convolution

$$\mathcal{P}^{(dis)} = \mathcal{K} \circ \mathcal{P}. \tag{5.28}$$

The PSF at the phase centre of the array $\mathcal{P}$ is known while the function $\mathcal{K}$ is unknown. In the following paragraphs, I discuss an algorithm to find an approximation for $\mathcal{K}$.

In the time domain, assume the baselines trace out a perfect circle in the uv-domain as if they were East-West baselines (i.e. without a v offset in the ellipses) and observe a source at the zenith. This implies that a uv-track in time is a circle with radius $\sqrt{u_0^2 + v_0^2}$ and angular velocity of $2\pi/(3600 \times 24) = \pi/(432 \times 10^2)$, with the 3600×24 seconds indicating one day period rotation:

$$u(\theta) = \frac{\pi}{432 \times 10^2} \sqrt{u_0^2 + v_0^2} \cos(\theta) \tag{5.29}$$

$$v(\theta) = \frac{\pi}{432 \times 10^2} \sqrt{u_0^2 + v_0^2} \sin(\theta), \tag{5.30}$$

where $\theta = \arctan(u_0/v_0)$ is the angle of orientation. The baseline, which samples the bin at $(u - u_0, v - v_0)$, has coordinates (u_0, v_0). The uv-domain rotation speed in time is then given by the partial derivative of Eq. (5.29) and (5.30):

$$\frac{\partial u}{\partial \theta} = -\frac{\pi}{432 \times 10^2} \sqrt{u_0^2 + v_0^2} \sin(\theta) \tag{5.31}$$

$$\frac{\partial v}{\partial \theta} = \frac{\pi}{432 \times 10^2} \sqrt{u_0^2 + v_0^2} \cos(\theta). \tag{5.32}$$

We can now approximate half of the phase difference in time as:

$$\Delta\Psi \approx \pi \left(\frac{\partial u}{\partial \theta} l + \frac{\partial v}{\partial \theta} m\right) \Delta t. \tag{5.33}$$

When averaging is carried out on channels to produce a single channel it leads to decorrelation in the frequency domain. In the image domain, decorrelation can be characterised by the product of the fractional bandwidth $\Delta\nu/\nu$ with the source distance $\sqrt{l^2 + m^2}$ from the phase centre of the observation relative to the baseline length $\sqrt{u_0^2 + v_0^2}$. We can approximate half of the phase difference in frequency as:

$$\Delta\Phi \approx \pi \frac{\Delta\nu}{\nu} \sqrt{l^2 + m^2} \sqrt{u_0^2 + v_0^2}. \tag{5.34}$$

Hence, the time-frequency cumulative distortion effects represented by the function $\mathcal{K}$ are approximated as:

$$\mathcal{K} \approx \mathcal{F}^{-1}\{\mathcal{X}_{\Delta\Psi}\mathcal{X}_{\Delta\Phi}\} \tag{5.35}$$

and the approximated pseudo-PSF by

$$\mathcal{P}^{(dis)} = \left(\mathcal{F}^{-1}\{\mathcal{X}_{\Delta\Psi}\mathcal{X}_{\Delta\Phi}\}\right) \circ \mathcal{P}. \tag{5.36}$$

We note from this method that the pseudo-PSFs are approximated from the PSF at the phase centre of the array denoted here by $\mathcal{P}$. Evaluating $\mathcal{P}^{(dis)}$ at each given l and m is cheaper in computation, as we will analyse in Section 5.4. The above processes for approximating $\mathcal{K}$ is summarised as Algorithm 1, which computes a pixelwise approximation of $\boldsymbol{\mathcal{K}}$.

Algorithm 1. The uv-domain is a discretized measurement of dimension $M\Delta u \times L\Delta v$, where the discretized bins are separated by the amount of Δu and Δv in the u and v direction respectively. In the image domain, the pixels are separated by the amount of Δl and Δm in the l and m direction respectively.

procedure APPROXIMATION OF $\boldsymbol{\mathcal{K}}$.
  $\Delta u = \frac{1}{M\Delta m}$
  $\Delta v = \frac{1}{L\Delta l}$
  $step\_u = \frac{M-1}{M}\Delta u$
  $step\_v = \frac{L-1}{L}\Delta v$
  $cst = \frac{\pi}{432\times 10^2}$
  $u_0 = \frac{1-M}{2}\Delta u$
  for i from 1 to M **do**
    $v_0 = \frac{1-L}{2}\Delta v$
    for j from 1 to L **do**
      $\theta = \arctan(u_0/v_0)$
      $\partial u = -cst\sqrt{u_0^2+v_0^2}\sin\theta$
      $\partial v = cst\sqrt{u_0^2+v_0^2}\cos\theta$
      $\Delta\Psi = \pi(\partial u l_r + \partial v m_r)\Delta t$
      $\Delta\Phi = \pi\frac{\Delta\nu}{\nu}\sqrt{l_r^2+m_r^2}\sqrt{u_0^2+v_0^2}$
      $\boldsymbol{\mathcal{K}}_{ij} = \widetilde{\mathcal{X}}(\Delta\Psi)\widetilde{\mathcal{X}}(\Delta\Phi)$

17: $v_0 = v_0 + step\_v$
18: **end for**
19: $u_0 = u_0 + step\_u$
20: **end for**
21: **end procedure**

end

5.3 DDFacet and pseudo-PSFs: gridding-degridding

This section describes the implementation of pseudo-PSFs in the faceting DDFacet imager [Tasse et al., b, in preparation]. The main advantage of implementing pseudo-PSFs in a faceting imager is that we can deconvolve each facet with a unique pseudo-PSF, i.e. the pseudo-PSF at the centre of the facet. Pseudo-PSFs do not vary significantly across a small portion of the sky (see Figure 5.4). With the faceting approach, the sky map can be partitioned into subregions of grids or facets. To deconvolve each facet, the exact pseudo-PSF at the centre of the facet using Eq (5.23) is computed. The irregularly averaged visibilities for pseudo-PSFs computed from Eq. (5.17) are then transformed to interpolated gridded visibilities, which lie on a regular grid; this procedure is referred to as *gridding* and the reverse process that estimates the irregularly averaged visibilities in Eq. (5.17) from the gridded visibilities is referred to as *degridding*. Gridding and degridding can be represented as

$$\mathcal{V}^{\text{grid}}_{pqkl}\big(\boldsymbol{u}(t_c,\nu_c)\big) = \sum_{pqkl} \mathcal{C}_{pqkl}\big(\boldsymbol{u}(t_c,\nu_c) - \boldsymbol{u}(t_k,\nu_l)\big)\mathcal{V}^{(\text{m})}_{pqkl}\big(\boldsymbol{u}(t_k,\nu_l)\big) \tag{5.37}$$

$$\mathcal{V}^{\text{degrid}}_{pqkl}\big(\boldsymbol{u}(t_k,\nu_l)\big) = \sum_{pqcc} \mathcal{C}_{pqkl}\big(\boldsymbol{u}(t_k,\nu_l) - \boldsymbol{u}(t_c,\nu_c)\big)\mathcal{V}^{\text{grid}}_{pqkl}\big(\boldsymbol{u}(t_c,\nu_c)\big), \tag{5.38}$$

respectively. Eq. (5.37) shows that gridding is equivalent to a discrete convolution with a finite support convolution kernel $\mathcal{C}$. Visibility bins that are in the range of the convolution kernel are averaged and this averaged value is allocated to the corresponding point on the grid (i.e to point c).

5.4 Computational cost for exact and approximated pseudo-PSFs

The FFT computational complexity for a pseudo-PSF scales as:

$$\mathcal{O}\big(N_{pq}N_tN_\nu \log_2 N_{\text{pix}}\big), \tag{5.39}$$

where N_{pq} is the number of baselines; N_t and N_ν are the number of timeslots and channels respectively. The product $N_{pq}N_tN_\nu$ is the total number of visibilities and it predicts the time taken to evaluate the fringe induced by each baseline, multiplied by the source amplitude and followed by the summation over all the baselines visibilities. $N_{pq}N_tN_\nu$ typically scales as N^2_{pix} and we have the following computational complexity:

$$\mathcal{O}\big(N^2_{\text{pix}} \log_2 N_{\text{pix}}\big). \tag{5.40}$$

Assume that N_{src} is the number of sources in the sky map. An iterative computation of pseudo-PSFs leads to a complexity of

$$\mathcal{O}\big(N_{\text{src}}N_{pq}N_tN_\nu \log_2 N_{\text{pix}}\big). \tag{5.41}$$

In the worst case where each pixel is a source ($N_{\text{src}} \sim N^2_{\text{pix}}$), we get

$$\mathcal{O}\big(N^4_{\text{pix}} \log_2 N_{\text{pix}}\big). \tag{5.42}$$

The computational complexity $\mathcal{O}\big(N^4_{\text{pix}} \log_2 N_{\text{pix}}\big)$ is the predicted cost to evaluate the pseudo-PSFs using a brute-force method. This complexity tends to be in the order of N^4_{pix} and therefore scales very poorly for widefield imaging.

The uv-domain approximation for pseudo-PSFs uses the phase gradients (visibility at the centre of each resampling interval; $\mathcal{R}\boldsymbol{u}_{pq(t_k,\nu_l)}$) and discards other visibility bins for the approximation of a pseudo-PSF. Thus, if $N_{pq}N_tN_\nu$ is the total number of visibilities, then only a few visibility samples are used for the uv-domain approximation. In other words, if n_t and n_ν are the number of time and frequency bins to resample, then $N_{pq}N^{\text{re}}_tN^{\text{re}}_\nu$ will be the number of phase gradients used in the approximation. Note here that $N^{\text{re}}_t = N_t/n_t$ and $N^{\text{re}}_\nu = N_\nu/n_\nu$. The complexity to approximate a unique pseudo-PSF in uv-domain

scales as:

$$\mathcal{O}\big(N_{pq}N_t^{\text{re}}N_\nu^{\text{re}}\log_2 N_{\text{pix}}\big). \tag{5.43}$$

The phase gradient is different for each source. This is problematic, as it emphasises the iterative computation of the phase gradient for each source, which increases the complexity for all the pseudo-PSFs by a factor of N_{src}, i.e.

$$\mathcal{O}\big(N_{\text{src}}N_{pq}N_t^{\text{re}}N_\nu^{\text{re}}\log_2 N_{\text{pix}}\big), \tag{5.44}$$

which is much better than $\mathcal{O}\big(N_{\text{src}}N_{pq}N_tN_\nu\log_2 N_{\text{pix}}\big)$, given that $N_t^{\text{re}}N_\nu^{\text{re}} \ll N_tN_\nu$. Note here that the cost of evaluating the image plane response of the BDWFs $\mathcal{X}$ (i.e. $\widetilde{\mathcal{X}}$) is regarded as negligible, given that $\widetilde{\mathcal{X}}$ is a real function, and therefore does not involve any exponential functions (e.g. in Section 5.2.1, for $\mathcal{X} \equiv \Pi$, $\widetilde{\mathcal{X}}_{\Delta\Psi_{pq},\Delta\Phi_{pq}} = \text{sinc}\big(\Delta\Psi_{pq}\big)\text{sinc}\big(\Delta\Phi_{pq}\big)$, which is computationally negligible).

The image domain approximation uses the PSF at the phase centre to approximate all the pseudo-PSFs. The complexity of evaluating the exact PSF at the phase centre scales as $\mathcal{O}\big(N_{pq}N_tN_\nu\log_2 N_{\text{pix}}\big)$ (see Eq. (5.39)). Suppose that $\mathcal{O}\big(\gamma\big)$ is the computational complexity to evaluate the cumulative distortion effects $\mathcal{K}$ for all sources in the map. The image plane approximation for pseudo-PSFs shows a computation scaling of

$$\mathcal{O}\big(N_{pq}N_tN_\nu\log_2 N_{\text{pix}}\big) + \mathcal{O}\big(\gamma\big), \tag{5.45}$$

where the cost $\mathcal{O}\big(\gamma\big)$ can be regarded as negligible, given that the evaluation of $\mathcal{K}$ does not involve any exponential functions (see Section 5.2.2 for the definition of $\mathcal{K}$), which implies that $\gamma \ll N_{pq}N_tN_\nu\log_2 N_{\text{pix}}$ and shows that the complexity in Eq. (5.45) is lower compared to $\mathcal{O}\big(N_{\text{scr}}N_{pq}N_tN_\nu\log_2 N_{\text{pix}}\big)$.

In the case of facet imaging, where the sky is partitioned into a subset of pixels and each subset is deconvolved separately with a unique pseudo-PSF before the results of each clean subset are merged, the computational complexity of evaluating the exact pseudo-PSFs iteratively with N_{facet} facets is

$$\mathcal{O}\big(N_{\text{facet}}N_{\text{pix}}^2\log_2 N_{\text{pix}}\big), \tag{5.46}$$

where $N_{\text{facet}} \ll N_{\text{pix}}^2$ which is much better than $\mathcal{O}\big(N_{\text{pix}}^4 \log_2 N_{\text{pix}}\big)$.

If the far pseudo-PSF sidelobes are insignificant (below some given threshold), then the size of the pseudo-PSF of each facet can be restricted to the facet size. In this case the computational complexity now scales as:

$$\mathcal{O}\big(N_{\text{facet}} N_{\text{pix,facet}}^2 \log_2 N_{\text{pix,facet}}\big). \tag{5.47}$$

Here, $N_{\text{pix,facet}}^2$ is the number of pixels in each facet and $N_{\text{pix,facet}} \ll N_{\text{pix}}$. The complexity in Eq. (5.47) implies that deconvolution will run much faster compared to Eq. (5.46).

5.5 Simulations and results

This section starts with examples showing the accuracy of pseudo-PSFs approximation in the visibility and image domains. We present the approximation using two resampling schemes: simple averaging and BDWFs. The BDWF used in this test is the sinc-1×1 tuned to a 2° FoI. We show at the end of this section the results of deconvolution with pseudo-PSFs implemented in DDFacet.

5.5.1 Pseudo-PSFs: approximation accuracy

The accuracy of the pseudo-PSF approximation in the visibility and in the image plane can be measured and compared to the exact pseudo-PSF by means of simulations. Specifically, we have simulated the case for the JVLA in C configuration at 1.4 GHz.

The pseudo-PSFs' approximation in the visibility plane (shown in Figure 5.1) for the case of simple averaging and in Figure 5.2 for the case of sinc-1×1 turned out to be 2° FoI. The left panels of Figures 5.1 and 5.2 show the results of pseudo-PSFs' for a source at 1.5° while the right panels show the results for a source at 2.5°. The top panels of Figures 5.1 and 5.2 present the exact and the approximated pseudo-PSFs in dB, while the bottom panels are the residuals for the approximated pseudo-PSFs subtracted from the exact pseudo-PSFs.

The results in Figures 5.1 and 5.2 show the shape distortion and the accuracy of pseudo-PSFs approximation. The absolute residual error in Figure 5.1 is less than 0.08 and 0.05

for a pseudo-PSF at 1.5° and 2.5° respectively. Note that the pseudo-PSFs for sinc-1×1 turned out to be 2° FoI results in a different shape compared to simple averaging, with an absolute error less than 0.03 and 0.01 for pseudo-PSFs at 1.5° and 2.5° respectively.

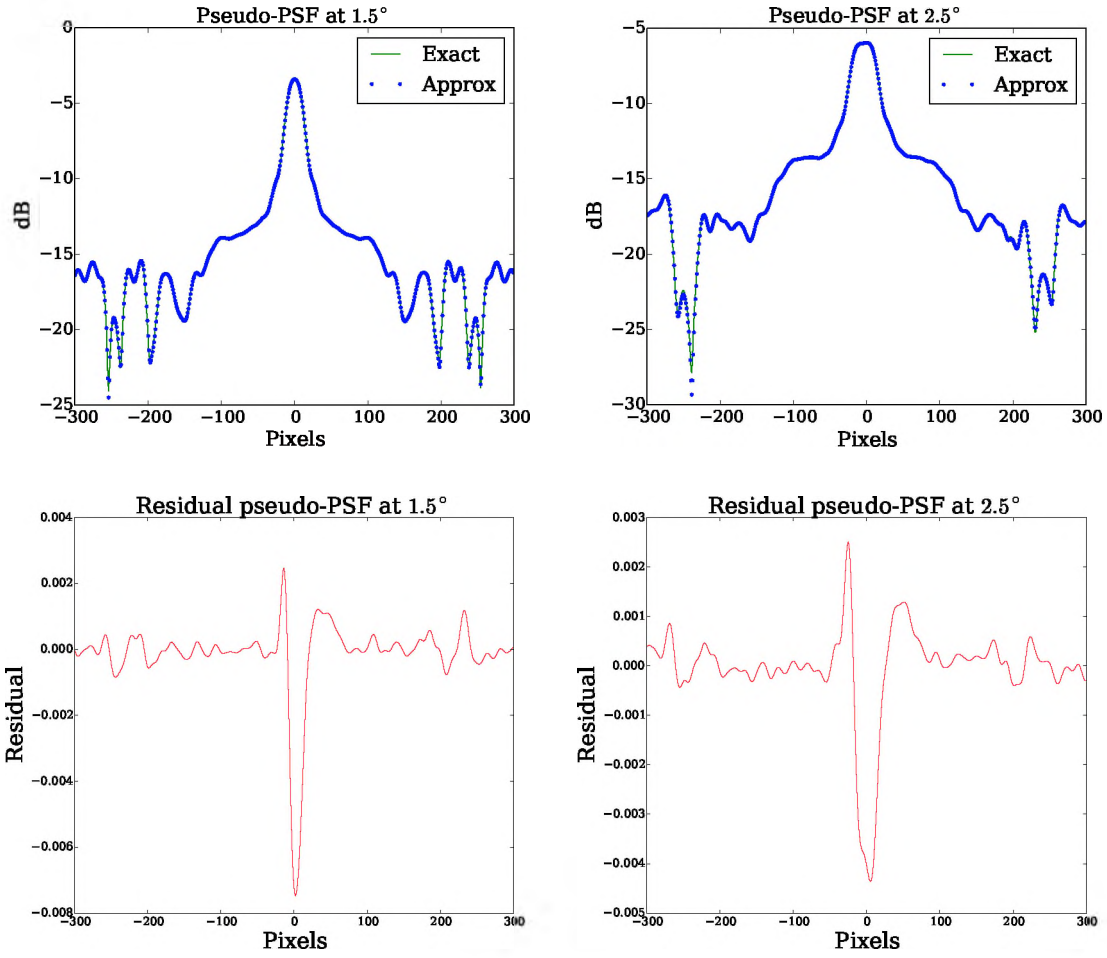

FIGURE 5.1: Simple averaging and pseudo-PSFs: exact PSF and approximation in the visibility domain. (Top panels) Exact and approximated pseudo-PSF for a source at 1.5° (left) and for a source at 2.5° (right) compared in decibel. (Bottom panels) Residuals for the approximated pseudo-PSFs subtracted from the exact pseudo-PSFs.

The image plane exact and approximated pseudo-PSF are shown in Figure 5.3. The top panel of this figure shows the exact pseudo-PSF, the middle panel shows the approximated pseudo-PSF, and the bottom panel shows the residual error. The results show that we achieve accuracy of approximation with an absolute residual less than 0.04 and 0.012 for a pseudo-PSF at 1.5° and 2.5° respectively.

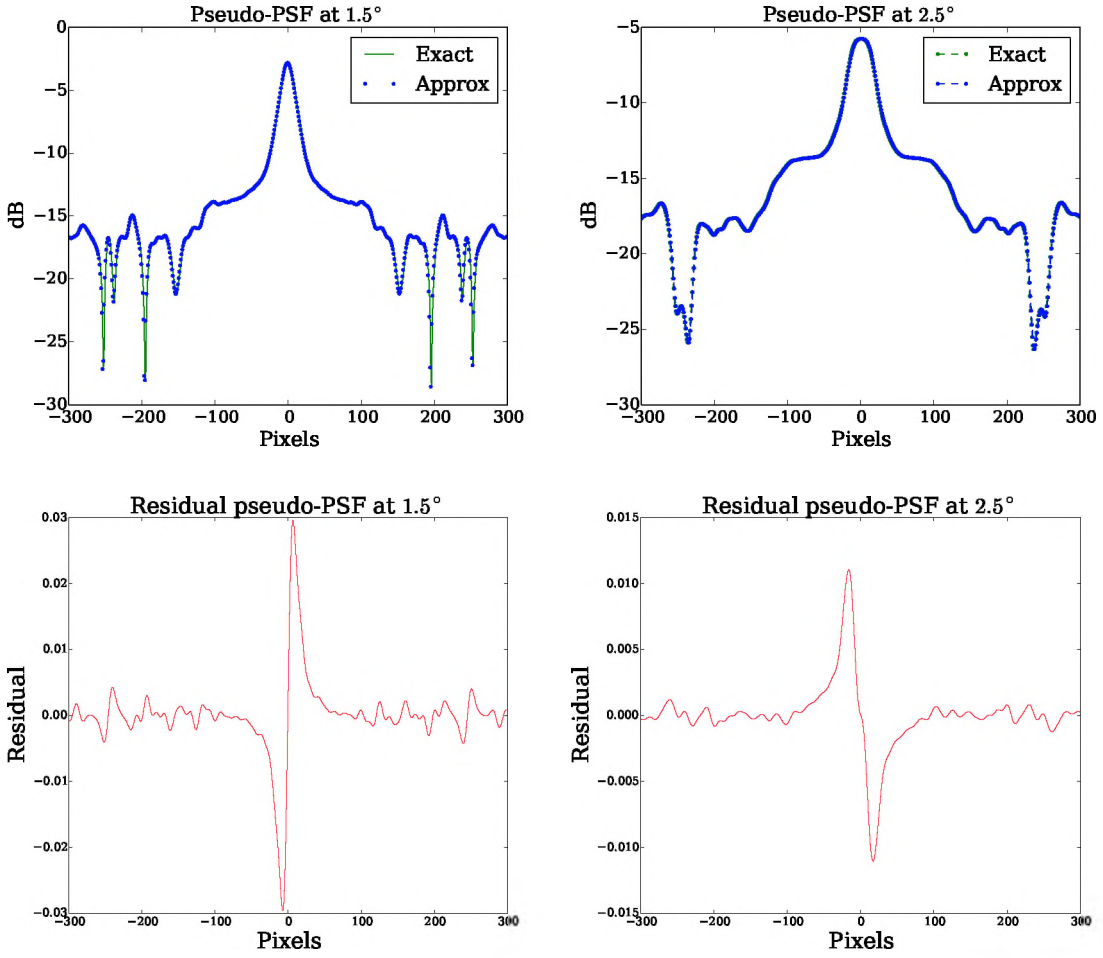

FIGURE 5.2: sinc-1×1 and pseudo-PSFs: exact PSF and approximation in the visibility domain. (Top panels) Exact and approximated pseudo-PSF for a source at 1.5° (left) and for a source at 2.5° (right) compared in log scales. (Bottom panels) Residuals for the approximated pseudo-PSFs subtracted from the exact pseudo-PSFs.

5.5.2 DDFacet and pseudo-PSFs

In this section the effects when pseudo-PSFs are taken into account during deconvolution are studied. We simulate visibility data for the LOFAR radio telescope. The LOFAR configuration employed in the simulation corresponds to a 1 hour synthesis at 1 s integration with 10 MHz total bandwidth centred at 200 MHz, channelised into 100 channels of 100 kHz each. The measurement set is populated by noise-free simulated visibilities with a random sky model. We then generate a low-res measurement set to receive the resampled visibilities for 50 channels averaged together and convert the visibilities to dirty image, then deconvolve using DDFacet in its original form and in its modified implementation accounting for pseudo-PSFs. Example of output images are shown in Figure 5.4 and 5.5. The images are of size 80 arcmin by 80 arcmin and divided into 21

facets by 21 facets. For deconvolution with pseudo-PSFs, each facet is deconvolved with the pseudo-PSF at the centre of the facet. The size of each facet determines the degree of variability of pseudo-PSFs within the facet; the larger a facet, the higher the degree of variability. This means that there is an advantage in using smaller facets rather than larger facets for pseudo-PSFs' deconvolution. However, the computational constraints for a faceting imager using smaller facets during deconvolution remain unchanged, since each facet is imaged separately, then deconvolved.

Figure 5.4 shows the variation of pseudo-PSFs across the image. Note the severe distortion of pseudo-PSFs away from the phase centre. Also, note that the pseudo-PSFs do not vary significantly within a small region in the image. This allows some tolerances for implementing a unique pseudo-PSF to deconvolve each of the sources within a small region in the image. The analysis shows that pseudo-PSFs are suitable for a faceting deconvolution framework, since the pseudo-PSFs are not evaluated for each source in the image, but instead for each facet, which results in a decrease in the computational cost of the pseudo-PSFs.

The top panel of Figure 5.5 shows the deconvolved image produced by DDFacet without using pseudo-PSFs. It illustrates the degraded image fidelity; the sources are severely distorted with the far-field sources smeared, compared to near-field sources. We cannot recover the structures of the original sources because the sources are deconvolved using the PSF at the centre of the image and neglect the fact that the pseudo-PSFs are distorted across the image.

The bottom panel of Figure 5.5 shows the result from the modified DDFacet, accounting for pseudo-PSFs. We see that we can recover the true shape of the original sources because the effects of the pseudo-PSFs have been subtracted properly. The image shows that the use of pseudo-PSFs is effective in deconvolution. In fact, the shape fidelity is expected to be exact for all the sources in the image because each of these sources is deconvolved with its exact PSF or a PSF relatively close to the source exact PSF.

5.6 Conclusion

In this chapter, I have presented a novel approach to radio interferometric imaging: the use of position-dependent PSFs in deconvolution. I have shown that boxcar averaging

and BDWFs applied to visibilities result in a distortion of the PSF which depends on sky position. I have briefly described the radio interferometric PSF and proposed mathematical frameworks to approximate a position-dependent PSF in both the visibility and image plane. These frameworks are cheaper in computation compared to the iterative brute-force computation of position-dependent PSFs. The advantage of this approximation is that the PSF at different sky positions can be used to deconvolve images at a lower computational cost. A pseudo-PSF deconvolution algorithm was then implemented in DDFacet. Simulation yielded high fidelity reconstruction of sources in the sky with pseudo-PSFs deconvolution. In addition to the image fidelity, pseudo-PSF deconvolution allows one to deconvolve with fewer cleaning iterations. Since the smeared PSF are subtracted from the source, the fainter pixels affected by smearing are discarded for the next CLEAN iteration. The potential of the algorithms proposed in this chapter represent a step towards the development of advanced deconvolution techniques that use fewer computational resources and are capable of providing higher image fidelity.

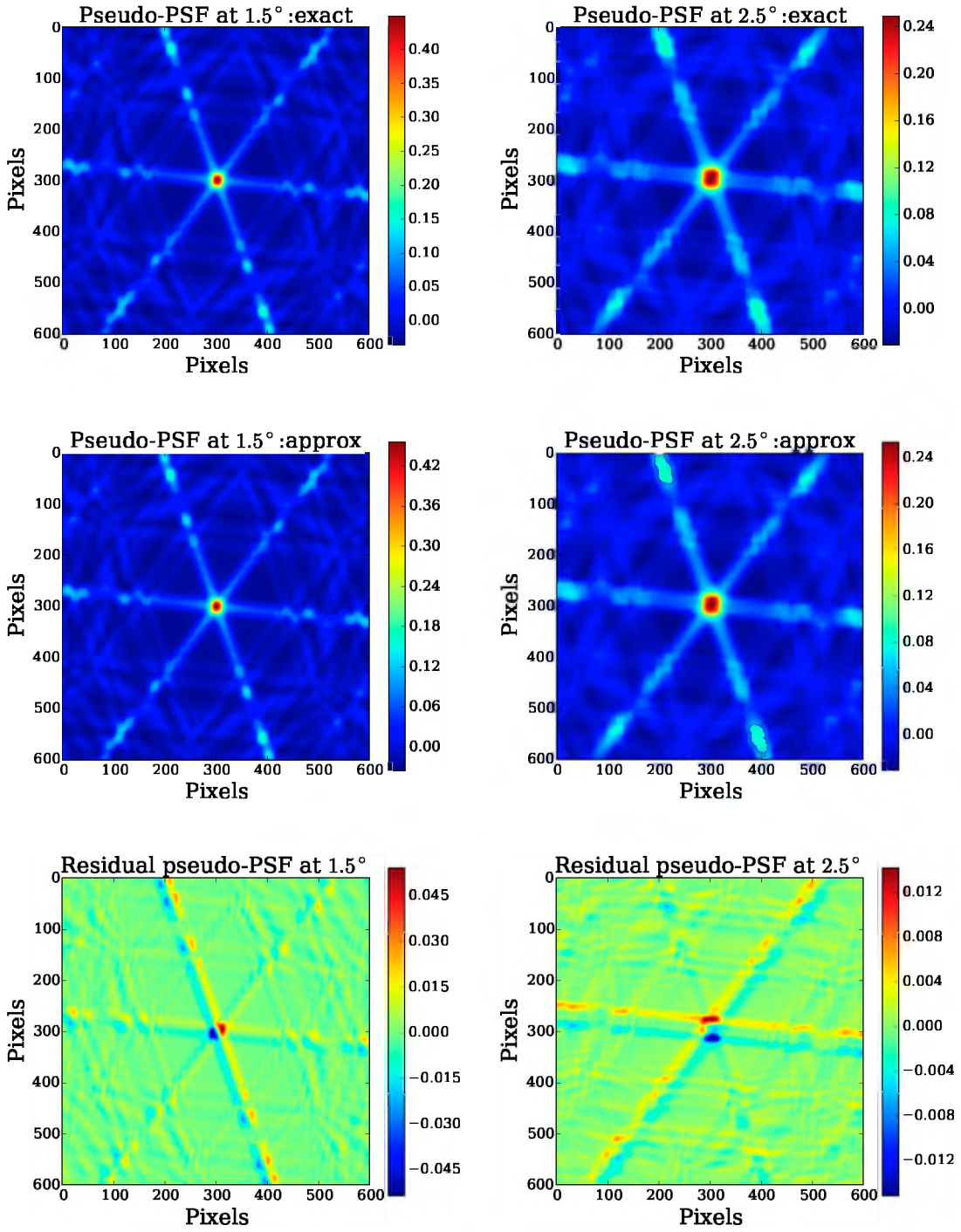

FIGURE 5.3: Simple averaging pseudo-PSFs: exact PSF and approximation in the image plane. (Top panels) Exact pseudo-PSF for source at 1.5° (left) and source at 2.5° (right). (Centre panels) Approximated pseudo-PSFs for a source at 1.5° (left) and for a source at 2.5° (right). (Bottom panels) Residuals for the approximated pseudo-PSFs subtracted from the exact pseudo-PSFs.

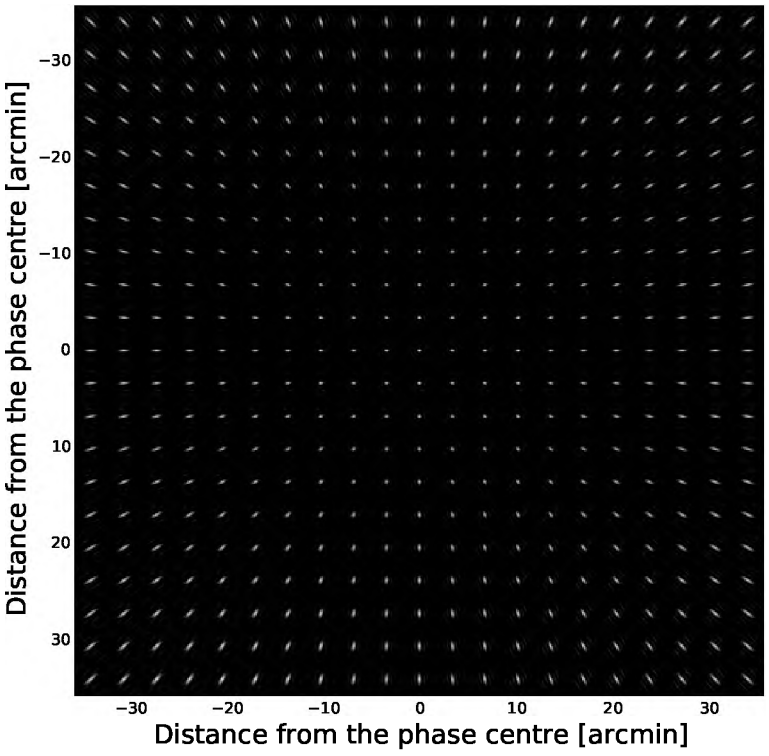

FIGURE 5.4: Pseudo-PSFs for 1 s integration time, 5 MHz compression intervals for LOFAR telescope observing at 200 MHz. Note the smeared and position-dependent nature of pseudo-PSFs.

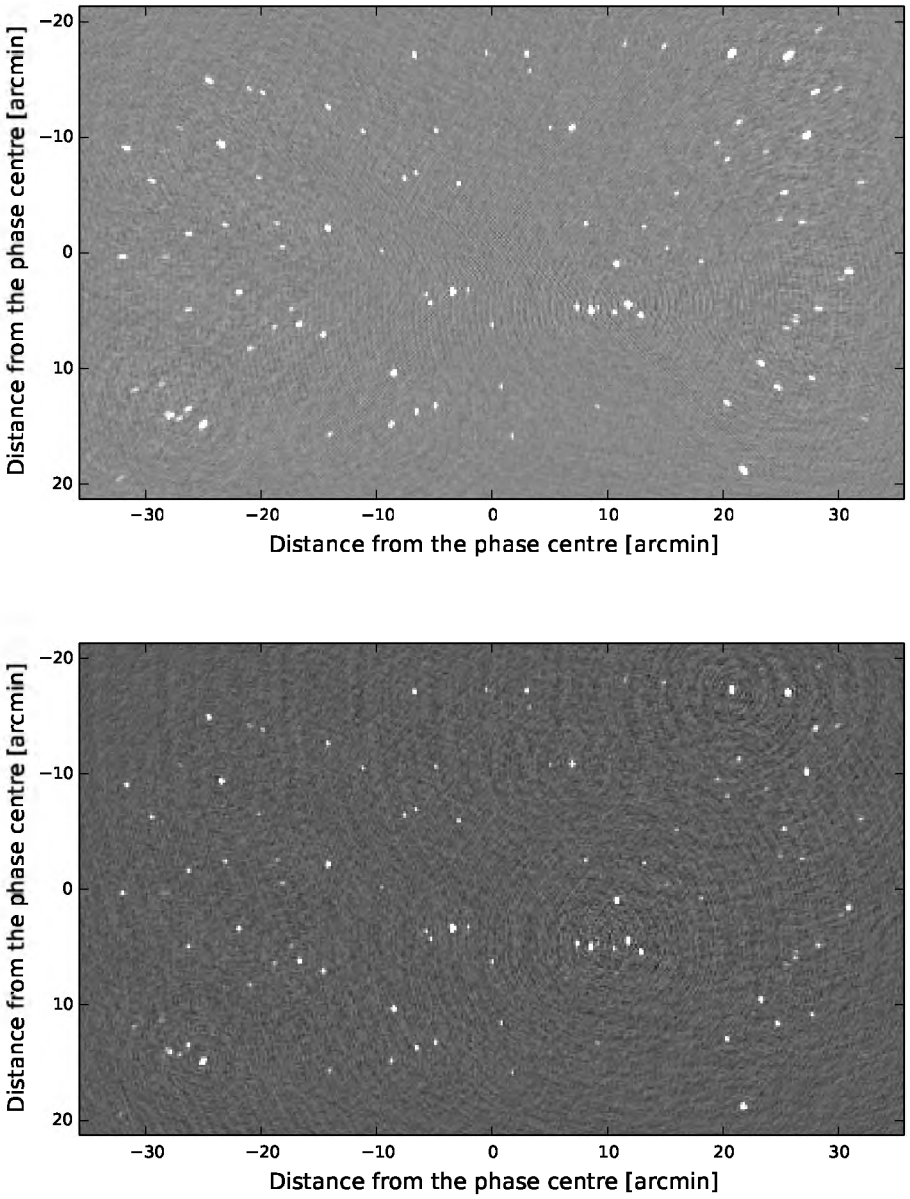

FIGURE 5.5: (Top): DDFacet without pseudo-PSFs; each source is deconvolved with the PSF at the centre of the image, i.e. the inverse Fourier transform of the uv-coverage. Note the effects of smearing in the image, with a significant shape distortion for far field sources. (Bottom): DDFacet with pseudo-PSFs; each source is deconvolved with the pseudo-PSF at the centre of the facet to which it belongs. The result shows excellent ability of pseudo-PSFs to recover the image fidelity. The shape of each source in the image is exact.

Chapter 6

General Conclusion and Perspectives

6.1 Problem and objectives

A radio interferometric correlator produces samples of the true *uv*-distribution by averaging the signal of each baseline over discrete time/frequency bins. The bins may be further enlarged by averaging in the post-processing stage. This averaging results in a baseline-length-dependent loss of signal amplitude and phase coherence, which is dependent on distance from the image phase centre. In general, the effect of averaging causes the visibilities to decorrelate and creates smearing artefacts in the image domain. For surveys with the new generation of radio telescopes such as MeerKAT and in the future the SKA, with wide FoIs and large bandwidth, the bin size, i.e. correlator dump time and channel width, must be kept very small to keep decorrelation down to acceptable levels, which leads to very high data rates. In the VLBI regime, the baselines are so long that smearing is the main bottleneck on the achievable FoI. Averaging also distorts the PSF differently at each sky position, since different baselines are attenuated differently. The goal of this thesis was threefold, namely;

(a) Develop a theoretical model to understand decorrelation/smearing and propose a novel approach to eliminate it while the visibilities are averaged at higher time/frequency compression factors without any compromise of the science goals.

(b) Develop an analytical model to understand decorrelation/smearing with a BDA correlator and study the possibility of decreasing smearing at a higher compression rate with BDA.

(c) Study the distortion of the PSF induced by averaging, and propose a novel algorithm to evaluate these PSFs at lower computational costs across each pixel in the sky map. Integrate a multi-PSFs deconvolution scheme in an imaging algorithm.

6.2 Methodology and results

I started the work by studying the origins of imaging artefacts as well as smearing effects in radio interferometry from a general signal processing point of view. Some window functions that are well-known in DSP were examined for the feasibility of their use in interferometric visibility averaging. Similarly, mathematical tools relevant to the problem of decorrelation and smearing were also studied. The main results and conclusions from these studies are:

(a) I have shown that visibility averaging (as currently implemented in correlators) is similar to convolution at the centre of the averaging intervals with a "boxcar" window function. Further, I have demonstrated, both theoretically and with simulations, that the use of suitably tuned window functions (referred to as BDWFs in this thesis) can both reduce smearing over the selected FoI (compared to simple "boxcar" averaging) and suppress unwanted signal from regions outside the FoI while the data are compressed to an acceptable level. The penalty for this is a loss in nominal sensitivity (since an unweighted average represents theoretically maximum sensitivity) at the centre of the FoI, but I have shown that this is offset by a decrease in smearing over the rest of the FoI. The study was further extended to the use of overlapping BDWFs. The following was established: The use of overlapping BDWFs can reduce smearing to under 5% and recover this nominal sensitivity loss, with excellent out-of-FoI source suppression. In principle, this should allow surveys to reach the same depth across the FoI, while reducing data sizes (by allowing for larger time/frequency bins). In the VLBI regime, larger FoIs become possible.

(b) I have studied the mathematical and implementation aspects for BDA. I have studied the use of BDA for decreasing smearing and for data compression. It was shown that

BDA results in eliminating decorrelation on the longer baselines, while the data are significantly compressed on the shorter baselines. It was also shown that applying BDA to overlapping BDWFs results in less than 1% smearing over a selected FoI with excellent performance in far field suppression.

(c) I have also explored the PSF behaviour of off-axis sources (referred to as pseudo-PSFs). Using linear algebra, I have proposed two analytical methods to understand and approximate these pseudo-PSFs. The first method uses the phase gradient to approximate the *uv*-domain decorrelation effects for pseudo-PSFs. The second method uses the PSF at the phase centre of the observation to approximate the smeared behaviour of pseudo-PSFs from the image plane. The proposed methods are cheaper in computation compared to an iterative brute force computation of pseudo-PSFs. The approximation allows us to implement pseudo-PSFs in DDFacet imager, which is a novel and unique approach in radio interferometric imaging. It was shown that pseudo-PSFs deconvolution results in higher image fidelity compared to standard deconvolution algorithms.

6.3 Perspectives and future work

The work presented in this thesis opens up several possibilities for future work. Firstly, designing an optimally matched filter for a BDWF is an interesting avenue of further research. In practical situations, the image-plane response of a sinc-like lowpass filter is far from ideal in the sense that a sinc-like filter is band-limited (zero outside some intervals) and discretised. Filter design theory for lowpass filters could therefore be used to explore an ideal image plane response, by using an approximation to define the ideal response parameters, such as the passband, the transition band and the stopband.

The second avenue involves evaluating the degree of source suppression as a function of array layout and BDFWs' parameters, i.e. the passband, transition band, stopband and the size of the filter.

The third avenue of exploration consists of investigating and exploring calibration with BDA and BDWFs applied across equal *uv*-distance. Currently, BDA and BDWFs can only be used post-calibration. Exploring the calibration parameters for BDA and BDWFs

could open a new research avenue in radio interferometry, in view of the effective use of BDA and BDWFs.

Finally, this thesis was restricted to simulations. The next step will be to implement each of the techniques presented in this work in practical research scenarios, e.g. applying the filters to real interferometric data. Currently, BDWFs are being implemented in the EVN software correlator (SFXC) at JIVE [Keimpema et al., 2015b] and the results are already promising. It would be beneficial to include BDWFs in other interferometric arrays as well.

Appendix A

The Fourier Transform of the Sampled Signal

The Fourier transform of a function f is defined as follows:

$$\mathcal{F}\{f\}(y) = \int_{-\infty}^{+\infty} f(x)e^{-iyx}\mathrm{d}x,$$

and its inverse

$$f(x) = (2\pi)^{-1} \int_{-\infty}^{+\infty} \mathcal{F}\{f\}(y)e^{iyx}\mathrm{d}y.$$

Now, it can be shown that the Fourier transform of a Dirac Comb is a Dirac Comb: A Dirac comb s is defined as below:

$$s = \sum_{k=-\infty}^{+\infty} \delta(x - k\Delta x) \quad \text{where k is an integer,}$$

and its Fourier transform is

$$\mathcal{F}\{s\}(y) = \int_{-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \delta(x - k\Delta x)e^{-iyx}\mathrm{d}x.$$

Since the Fourier transform is a linear operation, the previous equation can be rewritten as:

$$\mathcal{F}\{s\}(y) = \sum_{k=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x - k\Delta x) e^{-iyx} \mathrm{d}x.$$

For $\mu \equiv x - k\Delta x$, then $x = \mu + k\Delta x$, given that k and Δx are constant, $\mathrm{d}x = \mathrm{d}\mu$, we have

$$\begin{aligned}
\mathcal{F}\{s\}(y) &= \sum_{k=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x - k\Delta x) e^{-iyx} \mathrm{d}x \\
&= \sum_{k=-\infty}^{+\infty} e^{-iyk\Delta x} \int_{-\infty}^{+\infty} \delta(\mu) e^{-iy\mu} \mathrm{d}\mu \\
&= \sum_{k=-\infty}^{+\infty} e^{-iyk\Delta x}.
\end{aligned}$$

A Dirac Comb with period Δm is presented in Fourier series as:

$$\sum_{k=-\infty}^{+\infty} \delta(t - k\Delta m) = \frac{1}{\Delta m} \sum_{k=-\infty}^{+\infty} e^{-i\frac{2\pi}{\Delta m}kt}.$$

Let $\Delta x = 2\pi/\Delta y$; we can write

$$\begin{aligned}
\mathcal{F}\{s\}(y) &= \sum_{k=-\infty}^{+\infty} e^{-i\frac{2\pi}{\Delta y}ky} \\
&= \Delta y \sum_{k=-\infty}^{+\infty} \frac{1}{\Delta y} e^{-i\frac{2\pi}{\Delta y}ky} \\
&= \Delta y \sum_{k=-\infty}^{+\infty} \delta(y - k\Delta y).
\end{aligned}$$

The proof then follows:

$$\begin{aligned}
\mathcal{F}\{d_s\}(y) &= \Big(\mathcal{F}\{s\} \circ \mathcal{F}\{d_r\}\Big)(y) \\
&= \int_{-\infty}^{+\infty} \mathcal{F}\{s\}(x)\mathcal{F}\{d_r\}(y-x)\mathrm{d}x \\
&= \Delta y \int_{-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \delta(x-k\Delta y)\mathcal{F}\{d_r\}(y-x)\mathrm{d}x \\
&= \Delta y \sum_{k=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x-k\Delta y)\mathcal{F}\{d_r\}(y-x)\mathrm{d}x \\
&= \Delta y \sum_{k=-\infty}^{+\infty} \mathcal{F}\{d_r\}(y-k\Delta y).
\end{aligned}$$

Appendix B

The Fourier Transform of 2-D Boxcar Window

The Fourier transform of 2-D boxcar window Π defined as:

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

yields

$$\begin{aligned} \mathcal{F}\{\Pi(t,\nu)\} &= \frac{1}{\Delta t \Delta \nu} \mathcal{F}\{\Pi(t)\}\mathcal{F}\{\Pi(\nu)\} && \text{(B.2)} \\ &= \frac{1}{\Delta t \Delta \nu} \left(\int_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} e^{-2\pi j u t} \, \mathrm{d}t \right) \left(\int_{-\frac{\Delta \nu}{2}}^{\frac{\Delta \nu}{2}} e^{-2\pi j u \nu} \, \mathrm{d}\nu \right) && \text{(B.3)} \\ &= \frac{1}{\Delta t \Delta \nu} \left(\left[\frac{-1}{2\pi j u} e^{-2\pi j u t} \right]_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} \right) \left(\left[\frac{-1}{2\pi j u} e^{-2\pi j u \nu} \right]_{-\frac{\Delta \nu}{2}}^{\frac{\Delta \nu}{2}} \right) && \text{(B.4)} \\ &= \left(\frac{\sin(\pi \Delta t u)}{\pi \Delta t u} \right) \left(\frac{\sin(\pi \Delta \nu u)}{\pi \Delta \nu u} \right) && \text{(B.5)} \\ &= \mathrm{sinc}(\pi \Delta t u) \mathrm{sinc}(\pi \Delta \nu u). && \text{(B.6)} \end{aligned}$$

Bibliography

Akhmanw, C. and Khokblov, R. (1959). Karl Jansky: His Career at Bell Telephone Laboratories. *Phys. Rev*, 1:1350.

Arnold, S. (2014). The History of Radio Astronomy. In *Getting Started in Radio Astronomy*, pages 1–32. Springer.

Baars, J. and Hartsuijker, A. (1972). The decrease of flux density of Cassiopeia A and the absolute spectra of Cassiopeia A, Cygnus A and Taurus A. *Astronomy & Astrophysics*, 17:172.

Bach, U., Krichbaum, T., Alef, W., Witzel, A., and Zensus, J. (2002). Proper motion in Cygnus A. *arXiv preprint astro-ph/0207082*.

Booth, R., De Blok, W., Jonas, J., and Fanaroff, B. (2009). MeerKAT key project science, specifications, and proposals. *arXiv preprint arXiv:0910.2935*.

Booth, R. and Jonas, J. (2012). An overview of the MeerKAT Project. *African Skies*, 16:101.

Bos, A. (1985). *On instrumental effects in spectral line synthesis observations*. PhD thesis, Rijksuniversiteit te Leiden.

Bregman, J. D. (2012). *System Design and Wide-field Imaging Aspects of Synthesis Arrays with Phased Array Stations*. PhD thesis, University of Groningen, Groningen, The Netherlands, pp81-82.

Bridle, A. and Schwab, F. (1999). Bandwidth and time-average smearing. In *Synthesis Imaging in Radio Astronomy II*, volume 180, page 371.

Bridle, A. H. and Schwab, F. R. (1989). Wide field imaging I: Bandwidth and time-average smearing. In *Synthesis Imaging in Radio Astronomy*, volume 6, page 247.

Briggs, D. (1995). High fidelity interferometric imaging: robust weighting and NNLS deconvolution. In *Bulletin of the American Astronomical Society*, volume 27, page 1444.

Brown, W. C. (1984). The History of Power Transmission by Radio Waves. *IEEE Transactions on Microwave Theory and Techniques*, pages 1230–1242.

Burke, B. F. and Graham-Smith, F. (2010). *An introduction to radio astronomy*. Cambridge University Press.

Clark, B. (1980). An efficient implementation of the algorithm "CLEAN". *Astronomy & Astrophysics*, 89:377.

Cloudsim (2015). *PERvasive Computing & COMmunications for sustainable development*. PhD thesis, University of Technology, School of Industrial Engineering and Management, Industrial Engineering and Management.

Cornwell, T. (1983). A method of stabilizing the clean algorithm. *Astronomy and Astrophysics*, 121:281–285.

Cornwell, T. and Fomalont, E. B. (1989). Self-calibration. In *Synthesis Imaging in Radio Astronomy*, volume 6, page 185.

Cornwell, T. J. (2008). Multiscale CLEAN deconvolution of radio synthesis images. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):793–801.

Cotton, W. (1999). Special problems in imaging. In *Synthesis Imaging in Radio Astronomy II*, volume 180, page 357.

Cotton, W. D. (1989). Special problems in imaging. In *Synthesis Imaging in Radio Astronomy*, volume 6, page 233.

Crochiere, R. (1980). A weighted overlap-add method of short-time Fourier analysis/synthesis. *IEEE Transactions on Acoustics, Speech, and Signal Proceeding*, 28(1):99–102.

Delsarte, P., Janssen, A., and Vries, L. (1985). Discrete prolate spheroidal wave functions and interpolation. *SIAM Journal on Applied Mathematics*, 45(4):641–650.

Diakopoulos, N. A. and Stephenson, P. D. (2005). Anti-Aliased Lines Using Run-Masks. In *Computer Graphics Forum*, volume 24, pages 165–172. Wiley Online Library.

Duhamel, P. and Vetterli, M. (1990). Fast Fourier transforms: a tutorial review and a state of the art. *Signal Processing*, 19(4):259–299.

Felli, M. and Spencer, R. E. (1989). *Very Long Baseline Interferometry*. Number 283. Springer Science & Business Media.

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

Goldsmith, P. F. (1996). The second Arecibo upgrade. *Potentials, IEEE*, 15(3):38–43.

Grobler, T., Nunhokee, C., Smirnov, O., van Zyl, A., and de Bruyn, A. (2014a). Calibration artefacts in radio interferometry–I. Ghost sources in Westerbork Synthesis Radio Telescope data. *Monthly Notices of the Royal Astronomical Society*, 439(4):4030–4047.

Grobler, T., Smirnov, O., and Nunhokee, C. (2014b). Calibration artefacts in KAT-7 data. In *General Assembly and Scientific Symposium (URSI GASS), 2014 XXXIth URSI*, pages 1–4. IEEE.

Hamaker, J., Bregman, J., and Sault, R. (1996). Understanding radio polarimetry. I. Mathematical foundations. *Astronomy and Astrophysics Supplement Series*, 117(1):137–147.

Hey, J. S. and Hey, J. (1973). *The Evolution of Radio Astronomy*, volume 1. Science History Publications New York.

Hogbom, J. and Brouw, W. (1974). The synthesis radio telescope at Westerbork. Principles of operation, performance and data reduction. *Astronomy and Astrophysics*, 33:289.

Johnson, J. B. (1927). Thermal Agitation of Electricity in Conductors. *Nature*, 119:50–51.

Johnson, J. B. (1928). Thermal Agitation of Electricity in Conductors. *Physical Review*, 32(1):97.

Keimpema, A., Kettenis, M., Pogrebenko, S., Campbell, R., Cimó, G., Duev, D., Eldering, B., Kruithof, N., van Langevelde, H., Marchal, D., et al. (2015a). The SFXC software correlator for very long baseline interferometry: algorithms and implementation. *Experimental Astronomy*, pages 1–21.

Keimpema, A., Kettenis, M., Pogrebenko, S., Campbell, R., Cimó, G., Duev, D., Eldering, B., Kruithof, N., van Langevelde, H., Marchal, D., et al. (2015b). The SFXC software correlator for very long baseline interferometry: algorithms and implementation. *Experimental Astronomy*, pages 1–21.

Koopmans, L., Pritchard, J., Mellema, G., Abdalla, F., Aguirre, J., Ahn, K., Barkana, R., van Bemmel, I., Bernardi, G., Bonaldi, A., et al. (2015). The Cosmic Dawn and Epoch of Reionization with the Square Kilometre Array. *arXiv preprint arXiv:1505.07568*.

Krichbaum, T., Witzel, A., and Zensus, J. (1999). SKA in VLBI. Technical report.

Leisawitz, D., Frey, B. J., Leviton, D. B., Martino, A. J., Maynard, W. L., Mundy, L. G., Rinehart, S. A., Teng, S. H., and Zhang, X. (2002). The Wide-Field Imaging Interferometry Testbed I: purpose, testbed design, data, and synthesis algorithms. *arXiv preprint astro-ph/0210005*.

Malvar, H. S. and Staelin, D. H. (1989). The LOT: transforms coding without blocking effects. *IEEE Transactions on Acoustics, Speech, and Signal Proceeding*, 37(4):553–559.

Morgan, J. S., Mantovani, F., Deller, A., Brisken, W., Alef, W., Middelberg, E., Nanni, M., and Tingay, S. (2011). VLBI imaging throughout the primary beam using accurate UV shifting. *Astronomy & Astrophysics*, 526:A140.

Nan, R. (2006). Five Hundred Meter Aperture Spherical Radio Telescope (FAST). *Science in China series G*, 49(2):129–148.

Nan, R., Li, D., Jin, C., Wang, Q., Zhu, L., Zhu, W., Zhang, H., Yue, Y., and Qian, L. (2011). The Five-Hundred-Meter Aperture Spherical Radio Telescope (FAST) Project. *International Journal of Modern Physics D*, 20(06):989–1024.

Napier, P. J., Thompson, A. R., and Ekers, R. D. (1983). The very large array: Design and performance of a modern synthesis radio telescope. *Proceedings of the IEEE*, 71(11):1295–1320.

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

Nunhokee, C. D. (2015). Link Between Ghost Artefacts, Source Suppression and Incomplete Calibration Sky Models. Master's thesis, Rhodes University.

Nuttall, A. H. and Carter, G. C. (1982). Spectral estimation using combined time and lag weighting. *Proceedings of the IEEE*, 70(9):1115–1125.

Offringa, A., McKinley, B., Hurley-Walker, N., Briggs, F., Wayth, R., Kaplan, D., Bell, M., Feng, L., Neben, A., Hughes, J., et al. (2014). WSClean: Widefield interferometric imager. *Astrophysics Source Code Library*, 1:08023.

Offringa, A. R., de Bruyn, A. G., and Zaroubi, S. (2012). Post-correlation filtering techniques for off-axis source and RFI removal. , 422:563–580.

Perley, R. (2013). High Dynamic Range Imaging. presentation at "The Radio Universe @ Ger's (wave)-length" conference (Groningen, November 2013), http://www.astron.nl/gerfeest/presentations/perley.pdf.

Podder, P., Khan, T. Z., Khan, M. H., and Rahman, M. M. (2014). Comparative Performance Analysis of Hamming, Hanning and Blackman Window. *International Journal of Computer Applications*, 96(18).

Röttgering, H. (2003). LOFAR, a new Low Frequency Radio Telescope. *New Astronomy Reviews*, 47(4):405–409.

Röttgering, H., Braun, R., Barthel, P., van Haarlem, M., Miley, G., Morganti, R., Snellen, I., Falcke, H., de Bruyn, A., Stappers, R., et al. (2006). LOFAR Opening up a new window on the Universe. *arXiv preprint astro-ph/0610596*.

Ryle, M. and Hewish, A. (1960). The synthesis of large radio telescopes. *Monthly Notices of the Royal Astronomical Society*, 120(3):220–230.

Saloma, C. (1993). Computational complexity and the observation of physical signals. *Journal of Applied Physics*, 74(9):5314–5319.

Shannon, C. E., Gallager, R. G., and Berlekamp, E. R. (1967). Lower bounds to error probability for coding on discrete memoryless channels. I. *Information and Control*, 10(1):65–103.

Smirnov, O. (2011a). Revisiting the radio interferometer measurement equation-II. Calibration and direction-dependent effects. *Astronomy & Astrophysics*, 527:A107.

Smirnov, O. M. (2011b). Revisiting the radio interferometer measurement equation-I. A full-sky Jones formalism. *Astronomy & Astrophysics*, 527:A106.

Smirnov, O. M. and Frank, B. and Theron, I. P. and Heywood, I. (2012). Understanding the impact of beamshapes on radio interferometer imaging performance. In *Int. Conf. on Electromagnetics in Advanced Applications (ICEAA 2012). 2-7 September. Cape Town, South Africa.*

Smith, S. W. et al. (1997). The scientist and engineer's guide to digital signal processing.

Steer, D., Dewdney, P., and Ito, M. (1984). Enhancements to the deconvolution algorithm'CLEAN'. *Astronomy and Astrophysics*, 137:159–165.

Takahashi, F. (2000). *Very Long Baseline Interferometer.* IOS press.

Tasse, C., Atemkeng, M., Cecconi, B., Girard, J., Dumez-Viou, C., Smirnov, O., Zarka, P., Larzabal, P., El Korso, M., and Boyer, R. NenuFAR in standalone mode: calibration, imaging, confusion, and decorrelation issues. *NenuFAR: instrument description and science case*, page 38.

Tasse, C., Hugo, B., Smirnov, O., Atemkeng, M., Bester, E., Bonnassieux, E., Girard, J., Makhathini, S., and Perkins, S. Facetting for direction-dependent spectral deconvolution.

Thompson, A. R. (1999). Fundamentals of radio interferometry. In *Synthesis Imaging in Radio Astronomy II*, volume 180, page 11.

Thompson, A. R., Clark, B., Wade, C., and Napier, P. J. (1980). The Very Large Array. *The Astrophysical Journal Supplement Series*, 44:151–167.

Thompson, A. R., Moran, J. M., and Swenson, Jr., G. W. (2001). *Interferometry and Synthesis in Radio Astronomy.* Wiley, New York, 2 edition.

Thompson, A. R., Moran, J. M., and Swenson Jr, G. W. (2008). *Interferometry and Synthesis in Radio Astronomy.* John Wiley & Sons.

van Haarlem, M., Wise, M., Gunst, A., Heald, G., McKean, J., Hessels, J., De Bruyn, A., Nijboer, R., Swinbank, J., Fallows, R., et al. (2013a). LOFAR: The LOw-Frequency ARray. *Astronomy & Astrophysics*, 556:A2.

van Haarlem, M., Wise, M., Gunst, A., Heald, G., McKean, J., Hessels, J., de Bruyn, A., Nijboer, R., Swinbank, J., Fallows, R., et al. (2013b). LOFAR: The Low-Frequency Array. *Astronomy & Astrophysics*, 556:A2.

Vetterli, M., Marziliano, P., and Blu, T. (2002). Sampling signals with finite rate of innovation. *Signal Processing, IEEE Transactions*, 50(6):1417–1428.

Walker, R. (1999). Very Long Baseline Interferometry. In *Synthesis Imaging in Radio Astronomy II*, volume 180, page 433.

Walter, G. and Soleski, T. (2006). Prolate spheroidal wavelet sampling in computerized tomography. *Sampling Theory in Signal and Image Processing*, 5(1):21.

Watson, G. N. (1995). *A treatise on the theory of Bessel functions*. Cambridge University Press.

Wilson, A., Smith, D. A., and Young, A. (2006). The cavity of Cygnus A. *The Astrophysical Journal Letters*, 644(1):L9.

Woody, D. (2001a). Radio Interferometer Array Point Spread Functions I. Theory and Statistics. Technical report, ALMA Memo 389, National Radio Astronomy Observatory.

Woody, D. (2001b). Radio Interferometer Array Point Spread Functions II. Evaluation and Optimization.

Wright, M. (2002). A model for the SKA. Technical report, SKA memo 16.

Wrobel, J. and Walker, R. (1999). Sensitivity. In *Synthesis Imaging in Radio Astronomy II*, volume 180, page 171.

Yatawatta, S. (2014). Adaptive weighting in radio interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 444(1):790–796.

Zarka, P., Tagger, M., Denis, L., Girard, J., Konovalenko, A., Atemkeng, M., Arnaud, M., Azarian, S., Barsuglia, M., Bonafede, A., et al. (2015). NenUFAR: Instrument description and science case. In *2015 International Conference on Antenna Theory and Techniques (ICATT)*, pages 1–6. IEEE.