

**PARAMETERIZED DECONVOLUTION FOR WIDE-BAND
RADIO SYNTHESIS IMAGING**

by

Urvashi Rao Venkata

Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Physics
with Dissertation in Physics

New Mexico Institute of Mining and Technology

Socorro, New Mexico

May, 2010

ABSTRACT

The introduction of broad-band receivers into radio interferometry has opened up new opportunities for the study of wide-band continuum emission from a vast range of astrophysical objects. To take full advantage of such instruments and achieve continuum sensitivities, we need image reconstruction algorithms that are sensitive to the frequency dependence of the instrument as well as the spectral structure of the sky brightness distribution. This dissertation project involved a study of existing methods to deal with wide-band effects during interferometric image reconstruction, followed by the development of a multi-scale, multi-frequency, synthesis-imaging algorithm (MS-MFS) that (a) takes advantage of the multi-frequency uv -coverage while reconstructing both spatial and spectral structure for compact, extended and moderately resolved sources, (b) constructs intensity, spectral-index and spectral-curvature maps at an angular resolution given by the highest frequency in the band, and (c) corrects for the frequency dependence of the antenna primary beam to enable wide-band imaging across wide fields of view. The MS-MFS algorithm has been implemented in the CASA and ASKAPsoft data-analysis packages, and validated through a series of feasibility tests. This algorithm was then applied to multi-frequency VLA observations of the M87 radio galaxy to derive a 1.1 - 1.8 GHz spectral-index map to complement existing high-angular-resolution low-frequency images. The resulting 75 MHz to 1.8 GHz spectra were compared with models predicted by two different spectral evolution models, and synchrotron lifetimes for various halo features were estimated and interpreted in the context of the dynamical evolution of structures in the M87 radio halo.

ACKNOWLEDGMENTS

I have been fortunate to have had the support of mentors, colleagues, friends and family, sustained over many years, as I worked my way through this project. Here is my attempt to express my gratitude to those who have shaped and shared this journey with me.

Tim Cornwell is a fantastic teacher, advisor and guide, and I feel very fortunate to have had him as a mentor and friend over the past several years. All along, he has gently guided my thinking in a way that has always showed me new ways of approaching and interpreting things. Despite moving to the opposite side of the planet within six months of me beginning my PhD, he has interacted and communicated with me in a way that this distance has hardly mattered. He has also periodically reminded me that a satisfying career really depends only on finding something that can sustain ones interest, because then all the effort required to make it work will always be worth it. Tim, thank you for everything, and I look forward to continuing to work with you.

I am very grateful to my advisory committee for permitting me to do a thesis project with a largely technical focus, something that has given me the opportunity to develop a skill-set tailored to algorithm development for radio interferometry at a time when at least eight new radio telescopes are being built worldwide. Jean Eilek, my academic advisor, has spent an incredible amount of time and effort towards making this dissertation a little less cryptic than what I may otherwise have allowed it to be, and has also, with great patience, taught me all the astrophysics I know. My interactions with her over the past couple of years have taught me a lot, including a few things about myself, and for all this, I thank her deeply. Many thanks to Frazer Owen, my pre-doctoral supervisor at the NRAO, for his guidance on observations and data analysis in this thesis project and for doing his best to ensure that I did not get distracted by all the active EVLA development happening at the NRAO during this final year of thesis writing. Many thanks to Dave Raymond and Dave Westpfahl for their roles in my thesis committee at various stages of evaluation.

I would like to thank the Physics department at New Mexico Tech for a wide range of graduate coursework in basic physics, radio astronomy, and introductory astrophysics, and also the Computer Science and Mathematics departments at the University of California, San Diego for graduate coursework that provided the basics for a lot of the technical work in this project. I would like to thank the National Radio Astronomy Observatory for being my primary source of financial support over the past six years, first through part-time employment as a programmer, and later through an NRAO pre-doctoral fellowship. I would also like to thank the Australia Telescope National Facility for financial and local support through their affiliated-student program during the two summers I spent in Sydney, NSW, Australia. I would like to thank both the NRAO and the ATNF for also providing travel support to numerous conferences to present my work.

Many many people at the NRAO have contributed to the progress of my thesis project, and I thank everyone for being very supportive throughout this process. I have walked into many offices with questions about algorithms, data, software, bugs, units and co-ordinate systems, and I would like to thank you all for always finding the time to think and respond to them. Many thanks to everyone on the CASA developer team, for providing and maintaining the software infrastructure for most of the work done in this project. I am very grateful for the experience I have gained while working both formally and informally with this team. In particular, I would like to thank Kumar Golap and Sanjay Bhatnagar for all the time and effort they have spent discussing intricate implementation details about the best way to fold my algorithms into the CASA software framework for imaging and primary-beam correction. I would also like to thank Maxim Voronkov for similar discussions regarding the ASKAPsoft software package, as well as many others at the ATNF for discussions of various kinds during my visits there. Thanks to all the NRAO coffee and Thors-day lunch folk, past and present, who have provided many (often much needed) breaks and conversation. Finally, I would like to give a big thank you to all the extremely helpful support staff with whom I have interacted over these past few years; you keep these places running and always seem to do it with a smile.

I did not get a chance to meet Dan Briggs, but his thesis was one of the first documents I was handed as an undergraduate on my first internship at the National Centre for Radio Astrophysics in Pune, India. I was asked to look at chapter 4 - algebraic deconvolution. This was my first introduction to the subject of image reconstruction in radio interferometry, and I find it interesting but perhaps not very surprising that I have ended up doing a thesis on precisely that, working with two of the same people who mentored him.

Rajaram Nityananda is another mentor to whom I have turned several times in the past decade, at almost every stage of career-related decision making. Each time, he has listened very patiently and offered advice, and I am very grateful for this. In addition to mentoring me during two internships at the NCRA, introducing me to the use of linear algebra in interferometric data analysis and showing me how enjoyable this field of research could be, he also introduced me to my future husband, and I thank him for all of the above.

Sanjay has been my biggest supporter in this long process. As a colleague, he has always asked the tough questions, and my approach to problem solving, both for my dissertation project and elsewhere, has benefited greatly from this. As my best friend and life partner, he has always been there for me to remind me about what is really important, and without him, I may not have completed this.

I owe a large fraction of my sanity to some very good friends, some here in Socorro, and others in many different locations and time-zones, who all have always had time for a quick chat, especially during this final year of thesis writing. I would like to say thank you to all of you who have shared this journey with me, either by being a part of it, through empathy derived from your own PhD experience, or by just watching, often with amusement.

Special thanks to my parents for all their advice and encouragement over the years, and for watching so patiently, while I continued on a seemingly never-ending course of study. Special thanks also to Sanjay’s family for their support and encouragement over the past few years. I wish Sanjay’s mother too could have seen me complete this.

I gratefully acknowledge Tim Cornwell for the initial suggestion of combining the CH-MSCLEAN and SW-MFCLEAN algorithms to incorporate multi-scale techniques into multi-frequency synthesis imaging and deconvolution, Frazer Owen for providing images of the M87 radio galaxy at 75 MHz, 327 MHz, 1.4 GHz, 4.8 GHz (some unpublished), Jean Eilek for running simulations to generate spectral models that I used for various spectral fits, and Rick Perley for setting up observe files for the first set of multi-frequency test observations (Cygnus A) used in this dissertation. Chapters 3 and 4 are the result of notes written for the article “Advances in Calibration and Imaging Techniques in Radio Interferometry”, Proceedings of the IEEE, Vol.97, No.8, p-1472, August 2009, with Sanjay Bhatnagar, Maxim Voronkov, and Tim Cornwell as co-authors. All software written for this project has made extensive use of several open-source development packages and infrastructure code within the CASA and ASKAPsoft packages.

This dissertation was typeset¹ with \LaTeX ² by the author.

¹Many thanks to Dave Green for some extremely useful \LaTeX tricks that have been used in the preparation of this document.

² \LaTeX document preparation system was developed by Leslie Lamport as a special version of Donald Knuth’s $\text{T}_{\text{E}}\text{X}$ program for computer typesetting. $\text{T}_{\text{E}}\text{X}$ is a trademark of the American Mathematical Society. The \LaTeX macro package for the New Mexico Institute of Mining and Technology dissertation format was adapted from Gerald Arnold’s modification of the \LaTeX macro package for The University of Texas at Austin by Khe-Sing The.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ALGORITHMS	xiii
LIST OF SYMBOLS	xiv
1. Introduction	1
1.1 Goals of this dissertation	1
1.2 Background	2
1.3 Chapter Outline	4
2. Synthesis Imaging and Radio Interferometry	7
2.1 Image Formation	7
2.1.1 Theory of Interferometric Imaging	9
2.2 Measurement Equation for Radio Interferometry	17
2.2.1 Signal Measurement	17
2.2.2 Measurement Equation for Synthesis Imaging	21
3. Standard Calibration and Imaging	26
3.1 Calibration	26
3.1.1 Gain solution and correction	27
3.1.2 Types of Calibration	28
3.2 Imaging	30
3.2.1 Writing and Solving the Imaging Equations	31

3.2.2	Gridding	40
3.2.3	Preconditioning	42
3.2.4	Deconvolution	45
4.	Imaging with Direction-Dependent Effects	50
4.1	Types of Direction Dependent effects	51
4.1.1	Antenna Primary Beam	51
4.1.2	Non-Instrumental Effects	52
4.2	Correction of direction-dependent effects	52
4.2.1	Image-domain corrections	53
4.2.2	Visibility-domain corrections	55
4.3	Wide-field Imaging with Generalized direction-dependent effects	59
4.3.1	Imaging Equations	59
4.3.2	Iterative Deconvolution	61
5.	Imaging with Frequency-Dependent Effects	67
5.1	Wide-Band Radio Interferometry	68
5.1.1	Multi-Frequency Measurements	68
5.1.2	Frequency Dependence of the Sky and Instrument	71
5.2	Comparison of Existing Wide-Band Imaging Methods	74
5.2.1	Existing and Hybrid Algorithms	75
5.2.2	Simulations and Results	77
5.2.3	Continuum imaging with dense <i>uv</i> -coverage	83
6.	Deconvolution with Images Parameterized as a Series Expansion	85
6.1	Multi-Scale Deconvolution	86
6.1.1	Multi-Scale Image model	86
6.1.2	Imaging Equations and Block Deconvolution	88

6.1.3	Differences with existing MS-CLEAN techniques	98
6.1.4	Example of the Multi-Scale Principal Solution	99
6.2	Multi-Frequency Synthesis Deconvolution	102
6.2.1	Multi-Frequency Image Model	103
6.2.2	Imaging Equations and Block Deconvolution	106
6.2.3	Difference with the Sault-Wieringa (SW-MFCLEAN) algorithm .	117
6.2.4	Accuracy of multi-frequency deconvolution	119
7.	Multi-Scale Multi-Frequency Synthesis Imaging	124
7.1	Multi-Scale Multi-Frequency Deconvolution	124
7.1.1	Multi-Scale Wide-Band Image model	126
7.1.2	Imaging Equations and Block Deconvolution	126
7.2	Correction of Frequency-Dependent Primary Beams	135
7.2.1	Multi-Frequency Primary-Beam Model	135
7.2.2	Imaging Equations and Block Deconvolution	145
8.	Wide-Band Imaging Results	155
8.1	Algorithm validation <i>via</i> simulated EVLA data	156
8.1.1	Narrow-field imaging of compact and extended emission	157
8.1.2	Wide-field imaging with Primary-Beam correction	163
8.2	Feasibility Study of MFS in various situations	166
8.2.1	Moderately Resolved Sources	166
8.2.2	Emission at very Large Spatial Scales	170
8.2.3	Foreground/Background Sources with Different Spectra	174
8.2.4	Band-limited signals	176
8.3	Wide-band imaging results with (E)VLA data	179
8.3.1	Wide-band imaging of Cygnus A	179
8.3.2	Wide-band imaging of M87	188

8.3.3	Wide-field wide band imaging of the 3C286 field	196
8.4	Points to remember while doing wide-band imaging	201
8.4.1	Using the MS-MFS algorithm	201
8.4.2	MS-MFS error estimation and feasibility	204
8.4.3	Multi-frequency synthesis <i>vs</i> single-channel imaging :	206
8.4.4	Future Work	207
9.	A High Angular-Resolution Study of the Broad-Band Spectrum of M87	210
9.1	The M87 cluster-center radio galaxy	211
9.1.1	Studying M87 evolution	213
9.2	Synchrotron spectra and their evolution	216
9.2.1	Synchrotron radiation - basic facts	216
9.2.2	Ageing of synchrotron spectra	217
9.3	Data, Spectral Fits and Synchrotron Ages	223
9.3.1	M87 Spectral data	223
9.3.2	Spectral Fitting	228
9.3.3	Calculating Synchrotron lifetimes	235
9.4	Interpretation	237
9.4.1	Do these ageing models fit ?	237
9.4.2	Conclusions and Future Work	240
10.	Conclusion	242
10.1	Wide-band image reconstruction	242
10.2	The spectral evolution of M87	244
A.	Imaging Sensitivity	246
B.	Linear Least Squares	247
Bibliography		252

LIST OF TABLES

5.1	Data Simulation Parameters for Wide-Band Imaging Tests	77
6.1	Multi-scale principal solution example	101
8.1	Parameters for Wide-Band EVLA Simulations	156
8.2	Measured errors with MS-MFS on Simulated Data	162
8.3	True, measured and corrected intensity and spectra for foreground sources .	175
8.4	Wide-band VLA observation parameters for Cygnus A	180
8.5	Wide-band VLA observation parameters for M87	188
8.6	Measured errors for I_{ν_0} , α and β in M87	194
8.7	Spectral Index of 3C286 field with and without primary-beam correction .	199
9.1	Minimum-energy B-fields in M87 - 1	227
9.2	Minimum-energy B-fields in M87 - 2	228
9.3	Synchrotron lifetimes across M87	236
9.4	Synchrotron lifetimes for M87 filaments	236

LIST OF FIGURES

2.1	Co-ordinate Systems for Radio Interferometry	16
3.1	Diagram : Sampling Weights and the Point Spread Function	33
3.2	Diagram : Normal Equations for Basic Imaging	34
4.1	Diagram : Normal equations with an image-domain primary beam	55
4.2	Diagram : Normal Equations for General Primary-Beam correction	62
4.3	Diagram : Modified Normal Equations for General Primary-Beam correction	62
5.1	Multi-Frequency <i>uv</i> -coverage of the EVLA at L-Band	69
5.2	Multi-Frequency EVLA Primary Beams (1.0, 1.5 and 2.0 GHz)	73
5.3	1D cuts through EVLA Primary Beams at 1.0, 1.5 and 2.0 GHz	73
5.4	Spectral Index of the EVLA Primary Beam	74
5.5	Standard Algorithms on Point Sources with non-Power-Law Spectra	80
5.6	Hybrid Algorithms on Point Sources with non-Power-Law Spectra	81
5.7	Standard Algorithms on Extended Emission with non-Power-Law Spectra .	82
5.8	Hybrid Algorithm applied to Cygnus-A simulation	84
6.1	Multi-Scale image representation	87
6.2	Diagram : Normal Equations for a Multi-Scale Sky Brightness Distribution	91
6.3	Diagram : Normal Equations for Multi-Scale Deconvolution	92
6.4	Example of the Multi-Scale Principal Solution	100
6.5	Diagram : Multi-frequency sampling weights and PSFs	107
6.6	Diagram : Normal Equations for Multi-Frequency Deconvolution	110
6.7	Error Estimates for Spectral Index	122
6.8	Peak Residuals and Errors for MFS with different values of N_t	123
7.1	Diagram : Multi-frequency primary beams	136
7.2	Diagram : Evaluating the multi-frequency primary beam model - 1	141

7.3	Diagram : Evaluating the multi-frequency primary beam model - 2	141
7.4	Diagram : Evaluating the multi-frequency primary beam model - 3	142
7.5	Diagram : Multi-frequency normal equations with the primary beam factored out	142
7.6	Average Primary Beam, Spectral Index and Curvature	143
7.7	Spectral Index and Curvature of the EVLA Primary Beam	144
7.8	Diagram : Normal Equations for MFS with Primary-Beam Effects	147
7.9	Diagram : Normal Equations for MFS and Primary-Beam Correction	147
8.1	Example : Simulated wide-band sky brightness distribution	158
8.2	Example : True Taylor coefficient images	158
8.3	Example : Reconstructed Taylor coefficient images	159
8.4	Example : Residual images	159
8.5	Example : MS-MFS final imaging data products	160
8.6	Example : MS-MFS with Primary Beam correction on simulated EVLA data	165
8.7	Moderately Resolved Sources : Single-Channel Images	167
8.8	Moderately Resolved Sources : <i>uv</i> -coverage and Visibility-Plot	168
8.9	Moderately Resolved Sources : MSMFS Images	169
8.10	Moderately Resolved Sources : MSMFS Images using first and last channels	169
8.11	Very Large Spatial Scales : Visibility plots	172
8.12	Very Large Spatial Scales : Intensity, Spectral Index, Residuals	173
8.13	Foreground and Background sources : Intensity and Spectral Index	175
8.14	Band-limited Signals : Multi-frequency images	177
8.15	Band-Limited Signals : Spectra across the source	178
8.16	VLA multi-frequency <i>uv</i> -coverage	181
8.17	Cygnus A : Intensity and residual images	186
8.18	Cygnus A : Spectral Index image	187
8.19	M87 halo : Intensity and Spectral Index	192
8.20	M87 halo : Residual Images	193
8.21	M87 core/jet/lobe : Intensity, Spectral index, Curvature	194
8.22	M87 core/jet/lobe : L-band spectrum	195

8.23 MFS with primary-beam correction : 3C286 field (C-configuration)	198
8.24 MFS with PB correction : 3C286 field (B-configuration)	200
8.25 MFS with w -projection : 3C286 field (B-configuration)	200
9.1 Radio/X-ray/Optical image of M87	212
9.2 Labeled image of M87	214
9.3 Spectral Ageing models	221
9.4 M87 - Stokes I images at 74 MHz, 327 MHz, 1.4 GHz, and Spectral Index between 1.1 and 1.8 GHz	224
9.5 Spectral Fits : Spectral index - all over the source	226
9.6 Spectral Fits : χ^2 as a function of s and ν	229
9.7 Spectral Fits : Initial Injection model	231
9.8 Spectral Fits : Ongoing Injection model	232
9.9 Spectral Fits : Initial Injection model all over the source	233
9.10 Spectral Fits : Ongoing Injection model all over the source	234

LIST OF ALGORITHMS

1	CLEAN with Cotton-Schwab major and minor cycles	49
2	CLEAN with Visibility-Domain Direction-Dependent Corrections	66
3	Multi-Scale CLEAN deconvolution	97
4	Multi-Frequency CLEAN	116
5	MS-MFS CLEAN : Set-up and major/minor cycle iterations	133
6	MF-MFS CLEAN : minor cycle steps	134
7	MF-MFS CLEAN with MF-PB correction : Major/minor cycles	153
8	MS-MFS with MF-PB correction : Pre-Deconvolution Setup	154

LIST OF SYMBOLS

General Symbols :

A_S	area of the source aperture	10
\vec{b}	baseline vector in u, v, w co-ordinates	12
b_{max}	maximum baseline length	70
B	magnetic field	211
B_{eq}	equipartition (or minimum energy) magnetic field	222
B_{dyn}	maximum magnetic field due to dynamic pressure-balance	213
c	speed of light	10
c_s	speed of sound	215
E, E^*	electric-field at the detector	10
E	energy	211
\dot{E}	rate of energy change	215
g	loop gain	39
g_i	complex antenna gain for one polarization for antenna i	20
$I(l, m)$	intensity as a function of position on the sky [$W\ m^{-2}Hz^{-1}Sr^{-1}$]	13
ij	subscript index for the baseline formed by antennas i and j	19
j_{syn}	observed synchrotron power	217
Jy	Jansky ($1\ Jy = 10^{-26}W\ m^{-2}Hz^{-1}$)	11
k_B	Boltzmann constant	211
m	number of parameters, number of image pixels	21
n	number of measurements, number of visibilities	21
n_x	number density of particles	211
N_s	number of scale basis functions	87
N_t	number of terms in the Taylor polynomial	105
N_a	number of antennas	14
N_c	number of channels	106
$N(\gamma)$	number density spectrum of particles with energies γ	217
l, m, n	direction cosines on the sky, about \vec{s}_0	13
\vec{R}	location of a source	9
\vec{r}	location of a detector	10
P_{amb}	ambient pressure	213
P_{syn}	synchrotron power	217
$Q(\gamma, t)$	particle energy source function	220
\hat{s}	unit vector in the direction \vec{R}	10

\hat{s}_0	unit vector towards the phase reference center	11
\vec{s}_σ	vector between \vec{s}_0 and \vec{s}	11
$S(u, v)$	uv -coverage	14
s	power-law index for the particle energy distribution	217
s, p	indices for scale basis functions $s, p \in \{0, N_s - 1\}$	89
t, q	indices for spectral basis functions $t, q \in \{0, N_s - 1\}$	108
t_{cool}	cluster cooling time	211
$t_{buoyant}$	rise time for a buoyant bubble	215
t_{driven}	expansion time for a driven bubble	235
t_{sound}	expansion time at the speed of sound	235
t_γ	synchrotron lifetime for a particle of energy γ	218
t_{syn}	synchrotron age	219
T	temperature	211
T_{sys}	system temperature	68
$V(u, v)$	visibility as a function of spatial frequency [$Wm^{-2}Hz^{-1}$]	13
V_{ij}	visibility measured by baseline ij	19
w_v^t	Taylor weight	105
w_{sum}	sum of weights	32
u_{min}	minimum spatial frequency	125
u_{max}	maximum spatial frequency	68
u, v, w	projected lengths of x, y, z along \vec{s}_0 [in units of λ], spatial frequency	12
x, y, z	lengths in the terrestrial co-ordinate system [m]	12
α	spectral index of the sky brightness	72
$\delta\alpha$	error on the spectral index	119
$\Delta\alpha$	change in spectral index across a frequency range	191
α_{jet}	spectral index of the M87 jet	237
α_{PL}	spectral index between P-band and L-band	191
α_{LC}	spectral index between L-band and C-band	191
α_{LL}	spectral index between the lower and upper ends of L-band	191
β	spectral curvature of the sky brightness	103
$\delta\beta$	error on the spectral curvature	161
β_{LL}	spectral curvature between the lower and upper ends of L-band	191
χ^2	un-normalized chi-square	27
δ -function	Kronecker delta function	14
γ	Lorentz factor (energy $E = \gamma mc^2$)	217
$\dot{\gamma}$	rate of change of γ	218
ξ, ξ^*	complex electric-field at the source	9
λ	wavelength	12
ν	observing frequency, subscript index for frequency channel	9
ν_0	reference frequency	70

ν_c	critical frequency	218
ν_{min}	minimum sampled frequency	70
ν_{max}	maximum sampled frequency	72
ν_{syn}	synchrotron characteristic frequency	217
$\Delta\nu$	bandwidth, channel width	11
σ	rms noise in the image	45
σ_{chan}	single-channel image rms	75
σ_{cont}	continuum image rms	75
σ_T	Thomson scattering cross section	217
τ	time delay between wavefront incident at two detectors	10
θ^{psf}	angular size of the PSF main lobe	33

Matrix and Vector Notation :

$\langle A \rangle$	time average or expectation value of quantity A	9
$[A]$, $[A_{m \times n}]$	matrix of quantity A with m rows and n columns	18
$\vec{A}, \vec{A}_{m \times 1}$	column vector with m elements	18
$[A^T]$	transpose of $[A]$	18
$[A^\dagger]$	conjugate transpose or adjoint of $[A]$	18
$[A^{-1}]$	inverse of $[A]$	22
$[A^+]$	pseudo inverse of $[A]$	27
$[\vec{A}], diag(\vec{A})$	diagonal matrix : elements of \vec{A} on the diagonal	24
$[AB], [A][B]$	product of two matrices $[A_{m \times n}]$ and $[B_{n \times k}]$	24
$[A]\vec{B}$	product of matrix $[A_{m \times n}]$ and vector $\vec{B}_{n \times 1}$	21
$[A \otimes B]$	outer product of two matrices $[A]$ and $[B]$	19
$\vec{A} \star \vec{B}$	convolution of two vectors \vec{A} and \vec{B}	15
$\vec{A} \cdot \vec{B}$	element-by-element multiplication, product of two diagonal matrices ..	12
\vec{A}/\vec{B}	element-by-element division	47
$\vec{A}^2, \sqrt{\vec{A}}$	element-by-element square and square root	44
\vec{A}^{-1}	element-by-element inverse or reciprocal	56
\vec{A}^*	element-by-element complex conjugate	56
$tr[A]$	trace of the matrix $[A]$, sum of diagonal elements	32
$peak(\vec{A})$	the maximum value of quantity A in the list	49
$mid(\vec{A})$	the middle value in the list of quantity A	97
A^{type}	quantity A of a particular <i>type</i>	15
A_{index}	$index^{th}$ element in a list of quantity A	18
$[A_{p,q}]$	matrix in row p and column q of a block matrix	89
$\begin{bmatrix} [A_{0,1}] & [A_{0,2}] \\ [A_{1,0}] & [A_{1,1}] \end{bmatrix}$	2×2 block matrix	89

Fourier Transforms :

$[F_{m \times m}]$ matrix operator : Discrete Fourier Transform 22

$[F]$ represents the forward transform (image-domain to uv -domain). $[F^\dagger]$ represents the reverse transform (uv -domain to image-domain). $[F^\dagger F] = m \mathbf{1}_m$ where $\mathbf{1}_m$ is the $m \times m$ Identity matrix. $[F^\dagger]$ gives an un-normalized Fourier inverse. $[F^{-1}] = \frac{1}{m}[F^\dagger]$ gives a normalized Fourier inverse.

Convolutions :

$[F^\dagger][\text{diag}(\vec{A})][F]$ image-domain convolution operator whose kernel is $[F^\dagger]\vec{A}\dots$ 33
 $(uv\text{-domain element-by-element multiplication with } \vec{A})$

$[F^\dagger][\text{diag}([F]\vec{A})][F]$ image-domain convolution operator whose kernel is $[F^\dagger]\vec{A}$... 33
 $(uv\text{-domain element-by-element multiplication with } [F]\vec{A})$

$[F][\text{diag}(\vec{A})][F^\dagger]$ image-domain convolution operator whose kernel is $[F^\dagger]\vec{A}\dots$ 33
 (image-domain element-by-element multiplication with \vec{A})

$[F][diag([F^\dagger]\vec{A})][F^\dagger]$ image-domain convolution operator whose kernel is $[F^\dagger]\vec{A}$ 33
 (image-domain element-by-element multiplication with $[F^\dagger]\vec{A}$)

The convolution of two vectors is equivalent to the product of their Fourier transforms and is given by $\vec{A} \star \vec{B} = [F^\dagger]([F]\vec{A}) \cdot ([F]\vec{B})$. A convolution operator constructed for a kernel \vec{A} is therefore given by $[F^\dagger][\text{diag}([F^\dagger]\vec{A})][F]$. A convolution operator is a circulant matrix with a shifted version of \vec{A} in each row, which when multiplied with \vec{B} implements the shift-multiply-add sequence of convolution. An image-domain convolution with \vec{A} as the kernel is equal to a uv -domain multiplication by $[F]\vec{A}$.

Measurement and Normal Equations :

$[A]\vec{x} = \vec{y}$	measurement equations	21
$[A^\dagger W A]\vec{x} = [A^\dagger W]\vec{y}$	normal equations	31

The measurement equation of an instrument describes the effect of the measurement process on the input signal (transfer function). It is given by $[A_{n \times m}] \vec{x}_{m \times 1} = \vec{y}_{n \times 1}$ where \vec{x} is a list of parameters and \vec{y} is a list of measurements (data). An estimate of \vec{x} can be recovered from the data by solving the measurement equation. When the matrix $[A]$ has no exact inverse, a solution can be obtained by the process of χ^2 -minimization. Setting $\nabla \chi^2 = 0$ to minimize $\chi^2 = ([A]\vec{x} - \vec{y})^\dagger W ([A]\vec{x} - \vec{y})$ gives another linear system of equations called the normal equations, given by $[A^\dagger W A] \vec{x} = [A^\dagger W] \vec{y}$.

Labelled matrices and vectors :

$\vec{1}$	a vector filled with ones	37
$[B]$	Beam matrix : $[B] = [F^\dagger][W^G][F]$	33
	(image-domain convolution operator with kernel given by \vec{I}^{psf}) (uv-domain element-by-element multiplication with \vec{W}^G)	
$[D_{ij}^{sky}(l, m)]$	full-polarization image-domain effects for baseline ij and direction l, m	20
$[D_{m \times m}^{sky}]$	multiplicative image-domain instrumental effect $[D^{sky}] = \text{diag}([F^\dagger]\vec{K}^{dd})$	24
$[G_{m \times m}]$	gridding convolution operator $[G] = [FXF^\dagger]$	41
	(uv-domain convolution operator with kernel given by $[F^\dagger]\vec{X}$) (image-domain element-by-element multiplication with \vec{X})	
$[G^{dd}]$	gridding convolution with K^{dd}	59
$[G^{pb}]$	gridding convolution with K^{pb}	59
$[G^{ps}]$	gridding convolution with the prolate spheroidal	41
$[H]$	Hessian matrix	31
$[H^{peak}]$	Hessian matrix for a single pixel	36
$[H^{ms}]$	Hessian for multi-scale (MS) imaging	89
$[H^{mfs}]$	Hessian for multi-frequency (MF) imaging	108
$[H^{mfs,\delta}]$	MF Hessian with a δ -function PSF	139
$[H^{mfs,\delta,pb}]$	MF Hessian with a δ -function PSF and primary beam	138
$[H^{mfs,pb^2}]$	MF Hessian with MF primary beams	146
$[H^{pb^2}]$	Hessian with primary beams	61
$[H_{s,p}]$	Hessian block for multi-scale imaging	89
$[H_{t,q}]$	Hessian block for multi-frequency imaging	108
$\left[H_{s,p} \right]_{t,q}$	Hessian block for multi-scale multi-frequency imaging	128
$\vec{I}_{m \times 1}$	column vector : list of m image pixel amplitudes	21
\vec{I}^{beam}	restoring beam	40
\vec{I}^{dirty}	dirty image	31
$\vec{I}^{dirty,pb}$	dirty image with primary beam	54
\vec{I}^{dirty,pb^2}	dirty image with primary beam and gridding with $[G^{pb}]$	60
$\vec{I}^{dirty,ms}$	multi-scale dirty images	89
\vec{I}_s^{dirty}	multi-scale dirty image for scale s	89
$\vec{I}^{dirty,mfs}$	dirty images for multi-frequency deconvolution	108
$\vec{I}^{dirty,mfs,pb}$	MF dirty images with MF primary beams	140
\vec{I}^{dirty,mfs,pb^2}	MF dirty images with MF primary beams and gridding with $[G^{pb}]$	146
\vec{I}_t^{dirty}	dirty image for t^{th} Taylor term in multi-frequency deconvolution	108
\vec{I}_s^{dirty}	dirty image for s^{th} scale and t^{th} Taylor term	126
$\vec{I}^{flat sky}$	flat sky image model	137

\vec{I}^{model}	model image, reconstructed estimate of \vec{I}^{sky}	27
$\vec{I}_p^{model,\delta}$	multi-scale model image for scale p	95
\vec{I}_q^{model}	multi-frequency model image for q^{th} Taylor series coefficient	106
\vec{I}_p^q	multi-scale multi-frequency model image for scale p and coefficient q ..	131
$\vec{I}^{obs, pb}$	observed multi-frequency primary beam	138
$\vec{I}^{pix,psol}$	principal solution for one pixel	36
$\vec{I}^{pix,dirty}$	dirty-image vector for one pixel	36
\vec{I}^{psf}	point spread function	31
$\vec{I}_{s,p}^{psf}$	convolution kernel for multi-scale deconvolution	89
$\vec{I}_{t,q}^{psf}$	convolution kernel for multi-frequency deconvolution	109
\vec{I}^{res}	residual image	39
\vec{I}_s^{shp}	scale basis function	86
\vec{I}^{sky}	sky brightness distribution	21
$\vec{I}^{sky,\delta}$	multi-scale sky brightness	89
$\vec{I}_p^{sky,\delta}$	multi-scale sky brightness for scale p	89
$\vec{I}^{sky,mfs}$	sky brightness for multi-frequency deconvolution	108
\vec{I}_q^{sky}	sky brightness for q^{th} Taylor coefficient for MF deconvolution	108
$\vec{I}^{sky}(l, m)$	full-polarization sky brightness distribution in direction l, m	20
\vec{I}^{wt}	weight image	35
\vec{I}^α	spectral index image	114
\vec{I}^β	spectral curvature image	114
$[J_{2 \times 2}]$	Jones matrix	18
\vec{J}	uv -plane aperture illumination function	56
$[J_{m \times m}]$	uv -plane aperture illumination function in matrix form $[J] = diag(\vec{J})$..	56
$[K_{4 \times 4}]$	outer product of two Jones matrices $([J_{2 \times 2}] \otimes [J_{2 \times 2}])$	20
$[K_{ij}(u, v)]$	$[K_{4 \times 4}]$ for baseline ij and 2-D spatial frequency u, v	20
\vec{K}	$\vec{K} = \vec{J} \star \vec{J}$ is a uv -plane convolution kernel	56
$[K_{m \times m}]$	uv -plane convolution kernel $[K] = diag(\vec{K})$	56
\vec{K}^{dd}	uv -domain convolution kernel for direction-dependent effects	56
\vec{K}^{mos}	convolution function for mosaicing	58
\vec{K}^{pb}	convolution of two aperture illumination functions	56
\vec{K}^{po}	convolution function for a pointing offset	57
\vec{K}^{wp}	convolution function for w-projection	58
\vec{P}_s	prolate spheroidal function	41
\vec{P}_b	$\vec{P}_b = [V_p^\dagger][V_p] = [F^\dagger]\vec{K}^{pb}$ is the antenna primary beam	56
$[P_b]$	antenna primary beam in matrix form $[P_b] = diag(\vec{P}_b)$	56
\vec{P}_{bq}	q^{th} -order coefficient of the primary-beam Taylor polynomial	135
\vec{P}_{ba}	spectral index of the primary beam	144

$\vec{P}_{b\beta}$	spectral curvature of the primary beam	144
\vec{P}_{bv}	primary beam at frequency ν	135
$\vec{P}_b^{mfs,mult}$	MF primary beam in terms of Taylor coefficients	139
$[P_b^{mfs,mult}]$	MF primary beam multiplier	140
$[R_{m_I \times m}]$	projection matrix, resampling operator m to m_I pixels	41
$[S]$	matrix operator : uv -coverage, sampling function, transfer function	23
$[S^{dd}]$	sampling function with baseline-based convolution	23
\vec{T}_s	uv -taper function for multi-scale deconvolution $[F]\vec{I}_s^{hp}$	88
\vec{T}_{uv}	uv -taper function	44
$\vec{V}_{n \times 1}$	column vector : list of n complex visibilities	21
\vec{V}^{corr}	corrected/calibrated visibilities	27
\vec{V}^{model}	model visibilities	27
\vec{V}^{obs}	observed visibilities	21
$\vec{V}_{ij}^{obs}(u, v)$	observed visibilities for baseline ij and spatial frequency u, v	20
\vec{V}^{res}	residual visibilities	39
\vec{V}_p	$\vec{V}_p = [F^\dagger]\vec{J}$ is the antenna voltage pattern	56
$[V_p]$	antenna coltage pattern in matrix form $[V_p] = diag(\vec{V}_p)$	56
$[W]$	diagonal matrix : measurement or visibility weights	31
$[W^G]$	gridded weights $[W^G] = [S^\dagger WS]$	43
$[W^{mfs}]$	multi-frequency Taylor weights	106
$[W^{pc}]$	preconditioning weights	43
$[W^{pc,G}]$	gridded preconditioning weights	43
$[W^{im}]$	imaging weights	43

Abbreviations :

AIPS	Astronomical Image Processing Software	86
ASKAP	Australian SKA Pathfinder	45
ASKAPsoft	Australian SKA Pathfinder Software	204
ASP	Adaptive-Scale Pixel	47
ASTRON	Netherlands Institute for Radio Astronomy	204
ATCA	Australia Telescope Compact Array	69
ATNF	Australia Telescope National Facility	204
CASA	Common Astronomy Software Applications	86
CASACore	Common Astronomy Software Applications - Core Libraries	204
CASAPY	Common Astronomy Software Applications - Python	204
CH – MSCLEAN	Cornwell-Holdaway Multi-Scale CLEAN	46
DFT	Discrete Fourier Transform	22
e – MERLIN	extended Multi-Element Radio Linked Interferometer Network ..	69
EVLA	Expanded Very Large Array	67

FFT	Fast Fourier Transform	22
FWHM	Full Width at Half Maximum	72
HPBW	Half Power Beam Width	70
IF	Intermediate Frequency	180
LOFAR	Low Frequency Array	204
MEM	Maximum Entropy Method	47
MF	Multi-Frequency	xviii
MF – CLEAN	Multi-Frequency CLEAN, MS-MFS with a point source model	47
MFS	Multi-Frequency Synthesis	67
MS	Multi-Scale	xviii
MS – CLEAN	Multi-Scale CLEAN	46
MS – MFS	Multi-Scale Multi-Frequency Synthesis	131
NNLS	Non Negative Least Squares	47
NRAO	National Radio Astronomy Observatory	204
PB	Primary Beam	64
PSF	Point Spread Function	15
RFI	Radio Frequency Interference	188
RMS	Root Mean Square	77
SNR	Signal to Noise Ratio	45
SPW	Spectral Window	180
STACK	Narrow-band Imaging and Stacking	75
SW – MFCLEAN	Sault-Wieringa Multi-Frequency CLEAN	46
VLA	Very Large Array	117
VLBA	Very Long Baseline Array	69

CHAPTER 1

INTRODUCTION

1.1 Goals of this dissertation

A new generation of broad-band radio interferometers is currently being designed and built to provide high-dynamic-range imaging capabilities superior to that of existing instruments. With large instantaneous bandwidths and high spectral resolutions, these instruments will provide increased imaging sensitivity and enable detailed measurements of the spectral structure of a variety of astrophysical sources, all with less telescope time than previously possible.

One desired data product from such instruments is a continuum image. A continuum image is a 2-D map of the sky-brightness distribution integrated over a range of frequencies, and the noise in such a map is inversely proportional to the square root of the total bandwidth used. However, the response of the interferometer varies with frequency. Also, continuum emission from most astrophysical radio sources shows significant spectral structure over the frequency ranges for which these new receivers are being optimized. Therefore, to make a continuum image at the desired sensitivity, it is essential to measure or reconstruct the spectral structure of the sky-brightness distribution before constructing an image of the integrated flux, and to do this while accounting for the frequency dependence of the instrument.

While the main goal of wide-band imaging is to obtain a high dynamic-range continuum image, the reconstructed spectral structure can also be a useful astrophysical measurement. This is especially true since wide-band spectra can now be measured across a continuous range of frequencies and not just a few widely-separated narrow frequency bands. For sources of broad-band continuum emission, this will enhance the ability to measure spectra and detect and localize frequencies at which spectral steepening, flattening or turnovers occur. For observations in which different frequencies probe source structure at different physical depths, these continuous measurements provide information about the 3-D structure of the emitting source. When both spectral-line and continuum emission is present, such instruments will allow the measurement of a more accurate broad-band model for background subtraction.

So far, wide-band image reconstruction techniques have focused on optimizing the accuracy and dynamic range achievable in the continuum image by suppressing deconvolution errors that arise when the spectral structure of the sky-brightness is neglected [Conway et al. 1990; Sault and Wieringa 1994]. However, the spectral models used in these

techniques are appropriate mainly for narrow bandwidths and give visible deconvolution errors when applied to the large bandwidths offered by new receivers. Also, any spectral information obtained is only a by-product of the continuum imaging process and attention is not paid to the accuracy of these spectral reconstructions as astrophysical measurements. Therefore, with the large instantaneous frequency ranges to which new instruments are sensitive, it becomes worthwhile to design algorithms that reconstruct both the spatial and spectral structure of the sky-brightness accurately enough for astrophysical use, while still producing the desired high dynamic-range continuum image.

Goals : The two main goals of this dissertation are listed below.

1. Evaluate the applicability of existing wide-band imaging techniques to data from new broad-band interferometers and identify areas that require algorithmic improvements. Develop and implement a multi-frequency image reconstruction algorithm that combines a multi-scale parameterization of the sky-brightness with a spectral model capable of representing arbitrary but smooth spectra. To enable wide-band imaging over wide fields of view, this algorithm must also correct for the frequency-dependence of the antenna primary beam.
2. Apply this algorithm to data from multi-frequency VLA observations (1 to 2 GHz) of the M87 cluster-center radio galaxy. Combine the obtained spectral information with existing images of M87 at lower frequencies, and compare broad-band spectra of various features in the M87 radio halo to spectra predicted by two different spectral evolution models. Estimate synchrotron lifetimes from both models and interpret the results in the context of the dynamical evolution of various features seen in the M87 radio halo.

1.2 Background

This section first summarizes the state of the art in multi-frequency, multi-scale and wide-field image reconstruction techniques for radio interferometry, and motivates the choices made for the algorithm developed as part of this dissertation. This is followed by a brief description of feedback processes due to an active galactic nucleus (AGN) as a possible source of energy that prevents the cooling flow in the hot core of the Virgo cluster, and discusses what new information a high-angular-resolution study of the broad-band spectra across the M87 radio halo can provide.

Wide-band Imaging Techniques : The simplest method of wide-band image reconstruction is to treat each frequency channel separately and combine the results at the end. However, single-channel imaging is restricted to the narrow-band sensitivity of the instrument and source spectra can be studied only at the angular resolution allowed by the lowest

frequency in the sampled range. While such imaging may suffice for some science goals, it does not take full advantage of what a wide-band instrument provides. The spatial-frequency coverage of the interferometer varies with observing frequency. This is a significant advantage from the point of view of image reconstruction because wide-band instruments sample a larger fraction of the spatial frequency plane than measurements at a single frequency. By combining measurements from multiple discrete receiver frequencies during imaging in a process called multi frequency synthesis (MFS), one can potentially increase the fidelity and sensitivity of the resulting image.

MFS was initially done to increase the spatial-frequency coverage of sparse arrays by using narrow-band receivers and switching frequencies during the observations. However, it was assumed that at the receiver sensitivities of the time, the sky-brightness was constant across the observed bandwidth. The next step was to consider a frequency-dependent sky-brightness distribution. [Conway et al. \[1990\]](#) describe a double-deconvolution algorithm based on the instrument's responses to a series of spectral basis functions. [Sault and Wieringa \[1994\]](#) describe a similar multi-frequency deconvolution algorithm (SW-MF-CLEAN) which models an image as a collection of point sources with linear spectra and uses the fitted slopes to derive an average spectral index for each source. For pure power-law spectra, both methods suggest using a linear spectral model in $\log I$ vs $\log \nu$ space instead of I vs ν space. These methods were developed for relatively narrow bandwidths, and these approximations can be shown to be insufficient to model typical spectral structure across the large frequency ranges that new wide-band receivers are sensitive to. Therefore, new algorithms need to work with a more flexible spectral model.

So far, these CLEAN-based MFS deconvolution algorithms used point-source flux components to model the sky emission. This choice is not well suited for extended emission, where deconvolution errors due to the use of a point-source flux model are enhanced in the spectral index image because of non-linear error propagation. Multiscale deconvolution techniques that model images using flux components of varying scale size are more accurate at deconvolving large-scale emission. [Cornwell \[2008\]](#) describes the CH-MS-CLEAN algorithm which performs matched filtering using templates constructed from the instrument response to various large-scale flux components. To improve the performance of multi-frequency deconvolution in the presence of extended emission, such multi-scale techniques need to be included.

Finally, none of the existing wide-band imaging methods address the frequency dependence of various direction-dependent instrumental effects. The dominant such effect is the changing size of the antenna primary beam across frequency. Wide-band imaging across wide fields of view therefore requires this frequency-dependence to be modeled and corrected for. If unaccounted for, the frequency-dependent attenuation of the incoming radiation will create spurious spectral structure in the reconstructed spectral structure. [Bhatnagar et al. \[2008\]](#) describe an algorithm for the correction of time-variable wide-field instrumental effects for narrow-band interferometric imaging, and this algorithm needs to be adapted to work for wide-band imaging as well.

Spectral evolution of the M87 radio halo : M87 is a large elliptical galaxy at the center of the Virgo cluster. It hosts an AGN with an active jet, and contains a 40kpc radio synchrotron halo. Measurements of the current jet power suggest that this AGN plays a significant role in reheating the intra-cluster medium (ICM) at the center of the Virgo cluster and preventing cooling below a certain temperature. However, the mechanism by which this feedback may be occurring and the relevant timescales and periodicity are not well understood. Ages of the observed radio halo estimated from models of buoyant or driven and expanding bubbles yield timescales an order of magnitude smaller than the expected cooling time, rendering the system incapable of reheating the cluster core on the required timescales by mechanical energy transport alone. However, observations of X-ray emission from the core of the Virgo cluster show possible correlations with some features in the observed M87 radio halo, suggesting that these are sites of possible energy transfer between the radio plasma and the thermal ICM and that energetic processes other than simple synchrotron ageing may be at play.

The goal of this project is to study high-angular-resolution broad-band synchrotron spectra of various features in the M87 radio halo to assess whether or not there is evidence for anything other than simple synchrotron ageing as the energetic particles travel outwards from the jet into the radio halo. High-resolution studies are required in order to separate bright filamentary structure from the apparently diffuse background and see whether any significant spectral differences appear. So far, high-angular-resolution images of the M87 halo have been made only at 74 MHz, 327 MHz and 1.4 GHz, and show spectra consistent with pure power-laws of very slightly varying index.

This project uses a spectral-index map constructed by applying the MS-MFS algorithm to multi-frequency VLA observations between 1.1 and 1.8 GHz to constrain the shape of the spectrum at the upper end of the sampled frequency range. The resulting wide-band spectra are then compared with those predicted from two different synchrotron evolution models, one representing simple synchrotron ageing after an initial injection of energetic particles, and the other representing synchrotron ageing with continuously injected (or re-energized) particles. Synchrotron lifetimes computed from these spectral fits are then analysed in terms of plausibility with respect to estimated dynamical ages of various features observed in the M87 radio halo.

1.3 Chapter Outline

Chapter 2 introduces the idea of image formation using a simple lens as well as an imaging interferometer, and then describes the measurement process of a radio interferometer as a system of linear equations that have to be solved in order to construct an image. The goal of this chapter is to present the relevant theory in a linear algebra framework, from which image reconstruction algorithms and their numerical implementations can be easily derived.

Chapter 3 covers established calibration, imaging and deconvolution techniques, and introduces the generic numerical optimization framework used by CLEAN-based iterative deconvolution algorithms. The basic theme emphasized in this chapter is the design of an image-reconstruction algorithm based on a linear-least-squares approach, and its adaptation to the inherently non-linear process of interferometric image reconstruction by splitting the process into major and minor cycles. This framework forms the basis of all the algorithms described in later chapters.

Chapter 4 describes recent advances in wide-field imaging algorithms. Algorithms that correct for time-varying and direction-dependent instrumental effects are described within the imaging framework introduced in Chapter 3 to show how such corrections are performed in practice as a part of the iterative image reconstruction process.

Chapter 5 introduces the problem of wide-band imaging, and discusses the major factors that affect the process of image reconstruction when wide-band receivers are used with an imaging interferometer. This is followed by a brief description of several existing wide-band imaging techniques and the results of a study done to test their suitability for continuum imaging with the EVLA telescope and identify areas of required improvement.

Chapters 6 and 7 are technical chapters that contain the main contributions of this dissertation to the existing literature on CLEAN-based deconvolution algorithms. The general theme of these chapters is the parameterization of the sky-brightness distribution as a linear combination of images and the use of this model within the iterative major and minor cycle framework introduced in Chapters 3 and 4. These chapters contain (a) formal derivations of a multi-scale and a multi-frequency deconvolution algorithm, (b) a comparison of the resulting algorithms with the existing CH-MS-CLEAN and SW-MF-CLEAN implementations with suggestions of ways to improve them, (c) the combination of these ideas into a practical multi-scale, multi-frequency deconvolution algorithm (MS-MFS), and (d) a multi-frequency parameterization of the antenna primary beam and an algorithm to model and correct for it during MS-MFS deconvolution.

Chapter 8 discusses a set of wide-band imaging examples that illustrate the capabilities and limits of the MS-MFS algorithm and wide-band primary-beam correction. The tests described in this chapter include sky-brightness distributions with structure at multiple spatial scales and arbitrary but smooth spectra, moderately resolved sources, emission at very large spatial scales, band-limited signals, overlapping sources with different spectra and emission across wide fields of view. These tests involve applying the MS-MFS algorithm implemented within the CASA package to simulated wide-band EVLA data as well as data from multi-frequency VLA observations of Cygnus A, M87 and the 3C286 field. This chapter concludes with a summary of various practical aspects of wide-band imaging and potential sources of error, and lists a set of ideas for an end user to keep in mind while using the MS-MFS algorithm.

Chapter 9 describes a study of the wide-band spectra of various features in the radio halo of the M87 galaxy. A 1.1 to 1.8 GHz spectral index map of the M87 radio halo

was constructed using the MS-MFS algorithm, and combined with existing high angular resolution images at 75 MHz, 327 MHz and 1.4 GHz to construct wide-band spectra with constraints on their slopes at the higher end of the sampled frequency range. These spectra are then analysed in the context of synchrotron evolution models and the dynamical evolution of structures observed in the M87 radio halo.

Chapter 10 contains a brief summary of the work done and results obtained, and lists some topics of future research in wide-band imaging techniques.

CHAPTER 2

SYNTHESIS IMAGING AND RADIO INTERFEROMETRY

This chapter introduces the theory of image formation and aperture synthesis and describes the working of a radio interferometer. Section 2.1 describes the process of image formation with a simple lens as well as with an imaging interferometer, with the goal of relating the formal theory of interferometric imaging with the familiar concept of a lens. Section 2.2 then describes the measurement process of a radio interferometer and expresses it as a system of linear equations that must be solved in order to construct an image. The goal of this section is to present the theory of synthesis imaging in a linear-algebra framework, from which image-reconstruction algorithms and their numerical implementations can be easily derived. The basic theory in this chapter follows that described in [Thompson et al. 1986; Taylor et al. 1999; Briggs 1995; Bhatnagar 2001; Cornwell 1995a;b; Hamaker et al. 1996; Sault et al. 1996].

2.1 Image Formation

An image of a distant object is formed when radiation from the object passes through an aperture of finite size and falls on a screen made up of some material capable of recording the intensity of the incident radiation. This is a natural process that can be explained with the basic concepts of wave interference and Fourier transforms. This section first describes the form of the far-field radiation pattern produced when a wavefront of electromagnetic radiation passes through an aperture, and then describes how an image of the resulting intensity distribution can be formed using a lens as well as an interferometer.

The simplest way to form an image of a distant object is with a convex lens that focuses parallel rays of light onto a screen placed at the focal plane of the lens. The size of the lens defines its aperture, the opening through which the incident light passes. The aperture of a one-dimensional lens can be described as an infinite collection of slits located within a given maximum distance from each other. When illuminated by a plane wave-front of electromagnetic radiation, each slit produces a diffracted wavefront that propagates out behind the aperture. Consider one pair of slits. The diffracted wavefronts from both slits are coherent, and will interfere with each other to produce a far-field wave-front whose amplitude varies sinusoidally with position on the wave-front. The resulting intensity pattern is called an interference fringe. When illuminated from a direction normal to the plane of the slits, the zeroth-order maximum of the fringe pattern is in line with the point directly between the slits, and the wavelength of the fringe is inversely proportional to the dis-

tance between the slits. When there are more than two slits, the observed intensity pattern is that formed from the superposition of the sinusoidal wavefronts created by each pair of slits. Therefore, for electromagnetic radiation incident on a lens aperture, the amplitude and phase of the resulting wavefront can be described as the superposition of an infinite number of sinusoidal wavefronts spanning a continuous but finite range of fringe wavelengths and phases. This is a Fourier series, and the complex coefficients of this series form the spatial Fourier transform of the incident radiation field at the aperture. The intensity of the resulting far-field radiation pattern is the image of the object as viewed through the aperture. Image formation is the process of capturing and recording this intensity distribution.

2.1.0.1 With a Lens

To form a real image of the intensity distribution behind the aperture, a lens is needed to focus the radiation onto an image plane. The curvature of the lens surface introduces differential path delays between the light passing through multiple slits. This alters the phase of the sinusoidal wavefront from each pair of slits, such that for a normally incident a plane wave front (from a distant point source), the wavefronts from all slit pairs add purely constructively only at a single point, creating an image of a point source on the focal plane. Radiation from a distant object of finite size (larger than a single point) can be described as the superposition of plane wavefronts incident from multiple directions. Within a certain angular distance from the lens axis, wavefronts from directions other than the normal will be focused at different locations on the focal plane, thus forming an image of the incident brightness distribution.

2.1.0.2 With an Interferometer

An interferometer forms an image of the intensity distribution behind the aperture by directly measuring the spatial Fourier coefficients that describe the far-field radiation pattern and then performing a Fourier inversion to form an image. A finite set of points (or slits) are defined on the aperture, and the amplitude and phase of the interference wavefront from each pair of slits is computed by measuring the electric fields (E-field) incident at the two aperture points and correlating them (taking the expectation of their product). By this process, each pair of slits measures the spatial Fourier transform of the radiation field incident at the aperture, at the spatial frequency given by the physical separation of the slits in units of wavelength (see next section). This is an indirect imaging technique called aperture synthesis where a finite collection of spatially separated detectors are used to construct a lens aperture of size given by the largest separation between any two pairs of slits. A synthesised aperture differs from the true aperture of a lens of the same size in that it is not continuous, but made up of a discrete and finite set of aperture points.

2.1.1 Theory of Interferometric Imaging

This section formally describes the process by which an interferometer measures the spatial Fourier transform of the sky brightness distribution, starting with the electromagnetic waves emanating from the source and ending with the formation of an image.

To make a 2-D image of a distant object that emits electromagnetic radiation, we need to measure the power of the radiation field produced by the object along a set of directions covering different parts of the source. To form such an image, the source needs to be spatially incoherent, where the radiation produced by one part of the source is not correlated with the radiation from any other part of the source. If this were not the case (spatially coherent source) then the radiation from different parts of the source will interfere with each other, and the observer will sample this interference pattern instead of the total power from each point on the source [Anantharamaiah et al. 1989].

Let $\xi(\vec{R}, t)$ represent the time-varying amplitude of the E-field component¹ of an electromagnetic wave (EM-wave) emanating from the direction \vec{R} . For a monochromatic EM-wave emanating from a time-invariant source of radiation, we can write $\xi(\vec{R}, t) = \text{Re}\{\xi_\nu(\vec{R})e^{-2\pi i\nu t}\}$, where ν is the frequency of the EM-wave and $\xi_\nu(\vec{R})$ is a complex function of position (also called the complex amplitude of the E-field [Goodman 2002]). The spatial coherence² of this radiation field between two points \vec{R}_1, \vec{R}_2 on the source is given by $\langle \xi_\nu(\vec{R}_1)\xi_\nu^*(\vec{R}_2) \rangle$ where $\langle \rangle$ denotes a time-average. For a spatially incoherent source, this function is non-zero only when $\vec{R}_1 = \vec{R}_2$ and it becomes $\langle \xi_\nu(\vec{R})\xi_\nu^*(\vec{R}) \rangle = \langle |\xi_\nu(\vec{R})|^2 \rangle$ which is proportional to the total power (brightness) emanating from the point \vec{R} on the source.

When the radiation travels from the source to the observer, the radiation incident on the observer is partially coherent. This is because for a source of finite angular size, as the distance from the source increases, the wave-fronts become planar and it becomes increasingly difficult to distinguish between radiation from slightly different points on the source. The van-Cittert-Zernike theorem of partially coherent light, states that the degree of spatial coherence of the radiation field from a distant spatially incoherent source is proportional to the spatial Fourier transform of the intensity distribution across the source [Thompson et al. 1986]. The process by which an interferometer measures this degree of spatial coherence, and the way it is related to the source intensity distribution, is described below.

¹ The instantaneous E-field component of a polarized EM-wave is usually described by a vector defined in the plane perpendicular to the direction of propagation of the EM-wave. This vector is described by two orthogonal polarization components X, Y corresponding to linear polarizations. For this analysis, let us consider only one component (say X) of the E-vector for a monochromatic EM-wave.

²The spatial coherence of a wavefront describes the amount by which two secondary wavefronts emanating from a pair of spatially separated points on the original wavefront will interfere, at a later time. It is defined as the cross-correlation of the radiation field at two spatially separated points, averaged over time.

2.1.1.1 Spatial Coherence of the incident E-field

Consider the E-field component of a quasi-monochromatic EM-wave emanating from a source located at \vec{R} and incident on a detector located at \vec{r} . The complex amplitude of the E-field incident at the detector $E_\nu(\vec{r})$ can be related to the strength of the EM-wave $\xi_\nu(\vec{R})$ ³ emanating from the direction \vec{R} via the Huygens propagator [Clark 1999].

$$E_\nu(\vec{r}) = \int_S \xi_\nu(\vec{R}) \frac{e^{2\pi i \nu(R-r)/c}}{|R - r|} dS \quad (2.1)$$

dS is a surface element on the celestial sphere and S represents the projected shape of the source on the celestial sphere.

Consider the E-fields emanating from locations \vec{R}_1 and \vec{R}_2 within the source aperture S . The degree of spatial coherence between the E-fields incident at two locations \vec{r}_1, \vec{r}_2 on the aperture of the imaging instrument is given as follows.

$$\langle E_\nu(\vec{r}_1) E_\nu^*(\vec{r}_2) \rangle = \left\langle \int_S \int_S \xi_\nu(\vec{R}_1) \xi_\nu^*(\vec{R}_2) \frac{e^{2\pi i \nu(R_2-r_1)/c}}{|R_2 - r_1|} \frac{e^{-2\pi i \nu(R_2-r_2)/c}}{|R_2 - r_2|} dS_1 dS_2 \right\rangle \quad (2.2)$$

where ν is the frequency of the incident EM-waves. Assuming that the radiation at the source is spatially incoherent, $\langle \xi_\nu(\vec{R}_1) \xi_\nu(\vec{R}_2) \rangle$ is non-zero only when $\vec{R}_1 = \vec{R}_2 \equiv \vec{R}$. Eqn. 2.2 can be re-written as follows.

$$\langle E_\nu(\vec{r}_1) E_\nu^*(\vec{r}_2) \rangle = \int_S \langle |\xi_\nu(\vec{R})|^2 \rangle \frac{e^{-2\pi i \nu(r_1-r_2)/c}}{\left(1 - \frac{r_1}{|\vec{R}|}\right)\left(1 - \frac{r_2}{|\vec{R}|}\right)} \frac{dS}{|\vec{R}|^2} \left[\int_S dS_1 \right] \quad (2.3)$$

The quantity $\int_S dS_1 \equiv A_S$ is the area across the source aperture (in units of m^2). Also, each surface element dS is related to the corresponding solid angle $d\Omega$ as $dS = |\vec{R}|^2 d\Omega$ and an integration over S can be replaced by an integration over the entire celestial sphere ($\xi_\nu(\vec{R})$ will be non-zero only within the aperture). Due to the large distance between the source and the detectors, we can assume that $\frac{r_1}{|\vec{R}|} \ll 1, \frac{r_2}{|\vec{R}|} \ll 1$. Also, let $\hat{s} = \frac{\vec{R}}{|\vec{R}|}$ denote the unit vector in the direction \vec{R} . Finally, let $\tau_{12} = (\vec{r}_1 - \vec{r}_2)/c$ represent the difference between the time taken for the EM-wave to propagate from \vec{R} to \vec{r}_1 and \vec{r}_2 . Eqn. 2.3 becomes

$$\langle E_\nu(\vec{r}_1) E_\nu^*(\vec{r}_2) \rangle = A_S \int \langle |\xi_\nu(\hat{s})|^2 \rangle e^{-2\pi i \nu \tau_{12}} d\Omega \quad (2.4)$$

³ The amplitude of the E-field component of a propagating EM-wave at a distance R from a source of amplitude A is given by $\frac{A}{R} e^{2\pi i \nu R/c}$. In Eqn. 2.1, $A = \int_S \xi dS$ and therefore, the quantity $\xi_\nu(\hat{s})$ has units of $V m^{-2}$ [Eilek, private communication]. The quantity $E_\nu(\vec{r})$ then represents the E-field (in units of V/m) incident at the detector due to the whole radiation source (approximated as a point source of total amplitude A located at $R \gg r$).

Let $I_\nu(\hat{s})$ denote the intensity or brightness distribution in units⁴ of $\text{Wm}^{-2}\text{Hz}^{-1}\text{Sr}^{-1}$. Then, we can write the power per unit area incident on the detector (due to radiation from the whole source) as

$$I_\nu(\hat{s}) dv d\Omega = A_S \frac{\langle |\xi_\nu(\hat{s})|^2 \rangle}{\mu_0 c} \quad (2.5)$$

where $\mu_0 c$ is the impedance of free space, and dv represents an infinitesimal bandwidth at the detector. Eqn. 2.4 becomes

$$\langle E_\nu(\vec{r}_1) E_\nu^*(\vec{r}_2) \rangle \propto \int I_\nu(\hat{s}) e^{-2\pi i \nu \tau_{12}} d\Omega \equiv V(\vec{r}_1 - \vec{r}_2, \nu) \quad (2.6)$$

The quantity $V(\vec{r}_1 - \vec{r}_2, \nu)$, a complex number, is a time-averaged correlation coefficient called a visibility and its value depends on the physical separation of the pair of detectors $\vec{r}_1 - \vec{r}_2$ but not on their absolute locations [Clark 1999]. An interferometer consists of an array of spatially separated detectors, and visibilities are measured for every pair of detectors. The length of time over which these correlations are averaged to form each visibility is called the integration time, and will be denoted by $\Delta\tau$.

2.1.1.2 Co-ordinate systems

Visibilities measured from a collection of detector pairs (at one frequency ν) are combined to form an image of the intensity distribution at that frequency. To describe this process, we need to define a set of co-ordinate systems that relate the sky brightness distribution with the aperture that is being synthesized as well as the physical locations of the detectors.

Figure 2.1 defines the three co-ordinate systems that are required to describe the measurement and imaging process for a radio interferometer located on the surface of the Earth. $\hat{X}\hat{Y}\hat{Z}$ represents a terrestrial co-ordinate system in which the physical locations of the antennas are defined. The point on the sky towards which the interferometer is to be steered is called the phase-reference center \hat{s}_0 , expressed in terms of source declination δ_0 and hour-angle H . The $\hat{l}\hat{m}\hat{n}$ co-ordinate system is used to describe the sky brightness distribution projected onto the celestial sphere which is written as $I(l, m, n) = I_\nu(\hat{s})$ where $l, m, n = \sqrt{1 - l^2 - m^2}$ are direction cosines describing a direction \hat{s} . The phase reference center is given by $\vec{s}_0(l = 0, m = 0, n = 1)$ and a point away from the phase center is given by $\hat{s} = \hat{s}_0 + \hat{s}_\sigma$. The final 2D image that is formed is a projection of this intensity distribution onto the tangent plane at \hat{s}_0 (defined by $\hat{l}\hat{m}$). The plane defined by $\hat{U}\hat{V}$ is the aperture plane of the array, defined as the plane perpendicular to the instantaneous direction \hat{s}_0 (also \hat{W}).

⁴ The power per unit area (at the detector) carried by an EM-wave from the whole source is $A_S |\xi_\nu(\hat{s})|^2 / \mu_0 c$ in units of W m^{-2} (note that $\mu_0 c$ is the impedance of free space). dv (Hz) and $d\Omega$ (Sr) represent infinitesimal bandwidth and solid angle respectively. Therefore, the intensity (or brightness) $I_\nu(\hat{s})$ has the units of $\text{W m}^{-2}\text{Hz}^{-1}\text{Sr}^{-1}$ or Jy Sr^{-1} [Kraus 1986] where the unit of Jansky is defined as $1 \text{ Jy} = 10^{-26} \text{ W m}^{-2}\text{Hz}^{-1}$.

The $\hat{U}\hat{V}\hat{W}$ system is related to $\hat{X}\hat{Y}\hat{Z}$ by a co-ordinate rotation defined by the two angles δ_0 and H .

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} \sin(H) & \cos(H) & 0 \\ -\sin(\delta_0)\cos(H) & \sin(\delta_0)\sin(H) & \cos(\delta_0) \\ \cos(\delta_0)\cos(H) & -\cos(\delta_0)\sin(H) & \sin(\delta_0) \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (2.7)$$

where x, y, z are physical distances measured in the $\hat{X}\hat{Y}\hat{Z}$ system in units of *metres*, and u, v, w are distances measured in the $\hat{U}\hat{V}\hat{W}$ system in units of signal wavelength $\lambda = \nu/c$. As the Earth rotates and the hour-angle H changes, the co-ordinates of each detector in the $\hat{U}\hat{V}\hat{W}$ system follow ellipses on the $\hat{U}\hat{V}$ plane. The vectors \vec{r}_1, \vec{r}_2 in Eqn. 2.6 are defined as $\vec{r}_1(u_1, v_1, w_1)$ and $\vec{r}_2(u_2, v_2, w_2)$ in units of wavelength, in the $\hat{U}\hat{V}\hat{W}$ system. A baseline is defined as the 3D vector between \vec{r}_1 and \vec{r}_2 and is given by $\vec{b}(u, v, w) = \vec{r}_1 - \vec{r}_2$ with $u = u_1 - u_2, v = v_1 - v_2$ and $w = w_1 - w_2$. Note that \vec{r}_1, \vec{r}_2 need not lie exactly on the aperture plane ($\vec{r}_1 \cdot \hat{s}_0 = w_1 \neq 0$ and $\vec{r}_2 \cdot \hat{s}_0 = w_2 \neq 0$). This means that at a given instant, the two detectors will not sample the same wavefront of the incident radiation. The time delay between the wavefront reaching the two detectors is given by $\tau = \vec{b} \cdot \hat{s}_0/\nu = (w_1 - w_2)/\nu$ and needs to be accounted for before the signals from each detector are correlated.

2.1.1.3 Delay Correction

The correlation coefficients measured *via* Eqn. 2.6 require that \vec{r}_1 and \vec{r}_2 lie on the aperture plane so that all detectors measure the same wavefront of radiation incident from direction \vec{s}_0 with no time delay between the measurements. However, for most synthesis arrays the detectors do not lie exactly in the aperture plane. Delay correction is the process of delaying the signals from each detector such that at any given instant, all detectors sample the wavefront incident at the aperture plane (and not at the physical locations of the detectors).

The delay applied to the detector at \vec{r}_1 is the signal travel time across a distance $\vec{r}_1 \cdot \vec{s}_0 = w_1$ (written here in units of wavelength). When $\delta_0 = 90^\circ, w_1 = w_2 = 0$ and the two detectors \vec{r}_1, \vec{r}_2 always sample the same incident wavefront. When $\delta_0 \neq 90^\circ, w_1$ and w_2 are usually non-equal and \vec{r}_1, \vec{r}_2 sample the incident wavefront at time delays given by $\tau_1^c = w_1/\nu$ and $\tau_2^c = w_2/\nu$ relative to the chosen origin of the terrestrial co-ordinate system. To correct these delays, the signals sent to the correlator are $E(\vec{r}_1, t - \tau_{c1})$ and $E(\vec{r}_2, t - \tau_{c2})$. These delays change as the Earth rotates, and continuously correcting them has the effect of pointing the aperture towards a fixed point on the sky \vec{s}_0 .

Now consider an EM-wave incident from a direction $\hat{s} = \hat{s}_0 + \hat{s}_\sigma$. The time delay between the wavefront at the two detectors after delay correction will be $\tau_{12} = \vec{b} \cdot (\hat{s} - \hat{s}_0)/\nu = (ul + vm + w(n - 1))/\nu$. This time delay is the same as the τ_{12} in Eqn. 2.6 which contributes to the phase of the measured complex visibility.

2.1.1.4 Spatial Fourier transform

We can now write Eqn. 2.6 in terms of the baseline components u, v, w and the direction cosines for various points on the sky l, m, n (see Fig. 2.1). For a source defined on the celestial sphere, $d\Omega = \frac{dldm}{n}$, and Eqn. 2.6 becomes

$$V(u, v, w) = \iint \frac{I(l, m, n)}{n} e^{-2\pi i(ul+vm+w(n-1))} dl dm \quad (2.8)$$

Here, l and m are 2D co-ordinates on the tangent plane at \vec{s}_0 . For a point on the sky given by $\hat{s} = \hat{s}_0 + \hat{s}_\sigma$, the term $(n - 1)$ describes the distance between the true curved sky and the tangent plane at \hat{s}_0 . The product $w(n - 1)$ is called the w -term and is proportional to the phase difference between the radiation reaching the two detectors forming the baseline \vec{b} , due to the curvature of the sky. $w(n-1) \neq 0$ implies that even after delay correction, the two detectors are not sampling the same phase front of incident radiation, and the term $e^{-2\pi i w(n-1)}$ is the Fresnel diffraction kernel that accounts for the propagation of a spherical wave across the distance $\lambda w(n - 1)$ for one detector so that both detectors in the baseline measure the same wave-front. If the region of the sky being imaged is close to the phase center ($n \approx 1$), the w -term goes to zero and Eqn. 2.8 describes a 2D spatial Fourier transform relation between the mutual coherence function and the source brightness.

$$V(u, v) = \iint I(l, m) e^{-2\pi i(ul+vm)} dl dm \quad (2.9)$$

Eqn. 2.9 is also called the van-Cittert-Zernike theorem. This 2D spatial Fourier transform of the source brightness is called the visibility function. Eqn. 2.6 describes the measurement of this continuous visibility function at one spatial frequency point. The values of $u = u_1 - u_2$ and $v = v_1 - v_2$ denote the spatial frequency measured by the pair of detectors at \vec{r}_1, \vec{r}_2 , and they are defined in units of $\lambda = c/v$ where v is the observing frequency⁵. The visibility function is defined across the spatial frequency plane (also called the uv -plane) whose axes \hat{u}, \hat{v} correspond to the \hat{U}, \hat{V} axes in Figure 2.1 when baseline vectors are anchored at the origin. Each baseline measures the complex-valued visibility function at one point on the uv -plane. The amplitude and phase at each measured spatial frequency describes the 2D interference fringe that is measured by the pair of detectors on the aperture plane. If the visibility function were to be sampled continuously at all spatial frequencies u, v , then Eqn. 2.9 can be inverted via the Fourier transform to yield an image of the brightness distribution of the source radiation.

$$I(l, m) = \iint V(u, v) e^{2\pi i(ul+vm)} du dv \quad (2.10)$$

An interferometer synthesizes an aperture using a finite set of discrete points. Therefore in practice, the visibility function is never sampled continuously on the spatial frequency plane. The next section discusses the consequences of this incomplete sampling.

⁵The frequency at which the EM-wave is measured will be referred to as the observing frequency or just frequency.

2.1.1.5 UV coverage

An interferometer measures the visibility function $V(u, v)$ at a discrete set of spatial frequencies. With N_a antennas, there are $N_a(N_a - 1)/2$ baselines that make simultaneous measurements at spatial frequencies given by the projections of the 3D baseline vectors $\vec{b}(u, v, w)$ onto the aperture plane. This sampling of the spatial frequency plane defines the instantaneous transfer function of the synthesis array and is called the uv -coverage. It can be represented by a collection of Kronecker δ -functions as

$$S(u, v) = \sum_k \delta(u - u_k) \delta(v - v_k) \quad (2.11)$$

where k is an index that represents a measurement from one baseline. The spatial frequency plane can be further sampled by varying the positions of the antennas with respect to the direction of the phase-reference center. For ground-based arrays, the Earth's rotation makes all projected baseline vectors $\vec{b} \cdot \vec{s}_0$ trace ellipses on the spatial frequency plane, slowly filling it up. This is called Earth Rotation Synthesis. Since the measured spatial frequencies are defined in units of the wavelength of the radiation, measurements at multiple observing frequencies can be used to increase the sampling of the spatial-frequency plane, and this is known as Multi-Frequency Synthesis. Since the spatial frequency measured by a baseline changes with time and observing frequency, measurements must be made at sufficiently high time and frequency resolution to prevent smearing (averaging of visibility data) on the spatial frequency plane. The result is generally a centrally dominated uv -plane sampling pattern with a hole in the middle and tapered outer edges. $S(u, v)$ now represents the total collection of sampled spatial frequencies (discretized as a function of baseline, time and frequency)⁶.

The sampling function or uv -coverage $S(u, v)$ defines the imaging properties of the synthesis array. The maximum measured spatial frequency defines the angular resolution of the instrument. The smallest measured spatial frequency defines the largest spatial scale that the instrument measures. The density of samples within the measured range defines the instruments natural sensitivity to different spatial scales.

2.1.1.6 Imaging Equation

For a synthesis array with a given uv -coverage, the image formed by Fourier inversion of the measured visibilities can be described as follows. The measurement process multiplies the true visibility function (of the sky brightness) by the uv -coverage of the

⁶Earth-rotation-synthesis and multi-frequency-synthesis require the assumption that the sky brightness distribution is invariant across the time and frequency range being sampled, so that measurements at different times and frequencies sample the same visibility function, but at different spatial frequencies. The large-scale brightness distribution from most astronomical sources remains constant over typical observation timescales, so the first assumption is, in general, satisfied. [Conway et al. \[1990\]](#) describe the effect of relaxing the flat-spectrum assumption for wide-bandwidth systems and algorithms to deal with the consequences.

instrument. The observed visibility function is $V^{obs}(u, v) = S(u, v)V(u, v)$ and the image formed by direct Fourier inversion of the measurements is given by

$$I^{obs}(l, m) = \iint S(u, v)V(u, v)e^{2\pi i(ul+vm)}du dv \quad (2.12)$$

The convolution theorem of Fourier transforms states that a point-wise multiplication of two functions in one domain is equal to a convolution in the other Fourier domain. The raw or dirty image $I^{dirty}(l, m)$ is therefore the result of a convolution of the true sky brightness $I(l, m)$ with the point spread function (PSF) of the instrument $I^{psf}(l, m)$ given by the Fourier transform of the uv -coverage.

$$I^{obs} = I \star I^{psf} \quad (2.13)$$

$$\text{where } I^{psf}(l, m) = \iint S(u, v)e^{2\pi i(ul+vm)}du dv \quad (2.14)$$

where ' \star ' denotes convolution. The point spread function describes the instrument's response to a point source ($V(u, v) = 1$ for a point source of unit brightness at the phase reference center). In other words, it is the image that the interferometer will produce when a plane monochromatic EM wave is incident on the aperture from only one direction on the sky. Since the observed image is a convolution of the sky brightness with a known instrumental point spread function, an estimate of the true sky brightness can be obtained *via* a deconvolution process (described in Chapter 3).

Eqn. 2.12 is the result of a theoretical analysis that defines the raw image that the interferometer will produce under ideal measurement conditions, and unpolarized electromagnetic radiation. The next section describes some practical aspects of measuring the E-field component of polarized electromagnetic radiation at radio frequencies, and folds it into the above analysis.

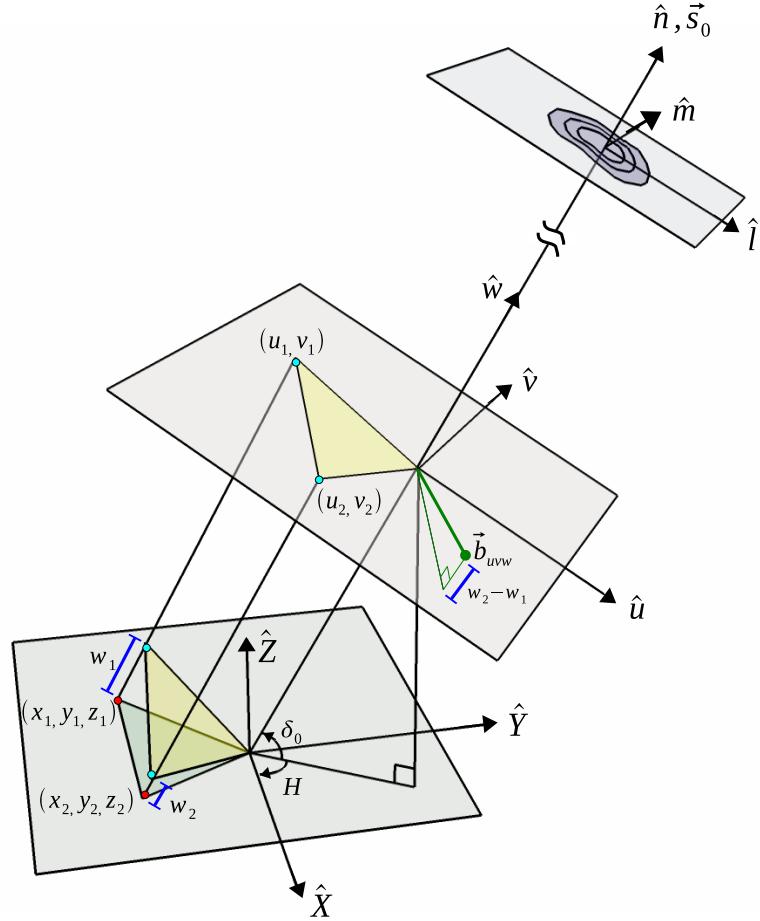


Figure 2.1: Co-ordinate Systems for Radio Interferometry : This diagram shows (a) the three co-ordinate systems involved in radio interferometric imaging, and (b) how a baseline vector is defined. $\hat{X}\hat{Y}\hat{Z}$ represents a terrestrial co-ordinate system in which \hat{Z} points toward the North celestial pole and the $\hat{X} - \hat{Y}$ plane is the equatorial plane of the Earth. \hat{X} is the intersection of the equatorial plane with the local meridian plane (defined as the plane through the poles of the Earth and the reference location of the array). \hat{Y} is towards the East (with respect to \hat{X}). \vec{s}_0 defines the direction to the point on the sky being imaged, expressed in terms of source declination δ_0 and hour-angle H . The $\hat{l}\hat{m}\hat{n}$ co-ordinate system is used to describe the 3D sky brightness distribution around \vec{s}_0 . The $\hat{u}\hat{v}\hat{w}$ plane is the aperture plane of the array, oriented perpendicular to the line of sight to the source \vec{s}_0 (also \hat{w}). The $\hat{u}\hat{v}\hat{w}$ system is related to $\hat{X}\hat{Y}\hat{Z}$ by a co-ordinate rotation defined by the two angles δ_0 and H . Let (x_1, y_1, z_1) and (x_2, y_2, z_2) represent the locations of two antennas in the terrestrial co-ordinate system (in units of metres), and (u_1, v_1, w_1) and (u_2, v_2, w_2) be the corresponding co-ordinates in the $\hat{u}\hat{v}\hat{w}$ system in units of wavelength λ . The distances w_1 and w_2 are proportional to the delays that have to be given to the signals from antennas 1 and 2 (relative to the chosen origin of the terrestrial co-ordinate system) to ensure that at any given instant, all antennas sample the same wave front of radiation incident from \vec{s}_0 . The 3D baseline vector between antennas 1 and 2 is given by as $\vec{b}_{uvw} = (u_2 - u_1)\hat{u} + (v_2 - v_1)\hat{v} + (w_2 - w_1)\hat{w}$. The 2D spatial frequency measured by this baseline is given by $(u_2 - u_1), (v_2 - v_1)$. As the Earth rotates, the hour-angle of the source changes, causing the projected antenna locations (and baseline vector) to trace ellipses on the now rotating $\hat{u}\hat{v}$ plane.

2.2 Measurement Equation for Radio Interferometry

The previous section described the theory of image formation and the working of an ideal interferometer. This section describes the process by which the electric field incident at a detector is measured, the effect of this measurement process on the input signal, and how the ideal imaging equations get modified when these effects are accounted for. This section introduces the concept of the measurement equation, a construct commonly used to describe the effect of the measurement process on the input signal. It is usually written in terms of the transfer function of the instrument, a function which describes the measurement process. The transfer function of an imaging interferometer includes its spatial-frequency sampling function as well as several factors that affect the incoming EM-wave before, during and after measurement. The process of image reconstruction (recovery of the input signal) is equivalent to solving the measurement equation *via* a process that may or may not involve the actual inversion of the transfer function.

Section 2.2.1 describes the E-field measured at each detector and the process of computing of a complex visibility from a pair of such measurements. It describes the practical implementation of the theory in section 2.1.1.1 for general polarized radiation. It uses a matrix notation commonly used in signal processing where orthogonal components of the E-field are listed as elements of a 2×1 vector and the effect of the instrument on such a signal is a 2×2 matrix operator. Section 2.2.2 describes the full measurement equation of the interferometer and introduces the matrix notation that will be used throughout the rest of this dissertation. The sky brightness distribution is represented by a list of m parameters and the instrument's transfer function (uv sampling function and the effect of signal measurement per antenna) is described as a $n \times m$ matrix operator. The product of these two matrices yields a list of n measurements. This matrix equation represents a system of linear equations which has to be solved in order to reconstruct an image of input sky brightness distribution. Chapter 3 describes this solution process in more detail.

2.2.1 Signal Measurement

The electric field components of the incoming electromagnetic radiation are measured at the locations of all antennas/detectors. The signals from each pair of antennas are then correlated (to evaluate Eqn. 2.6) to form a set of complex numbers that measure the source visibility function at the spatial frequencies given by the baseline vectors. This section follows the derivation and notation of [Hamaker et al. \[1996\]](#).

2.2.1.1 Electric Field at each Antenna

The electric field component of a polarized electromagnetic wave at a given instant is represented by a 2D vector lying in the plane perpendicular to its direction of

propagation. Let $\vec{E}_i = \begin{bmatrix} e^X \\ e^Y \end{bmatrix}_i = [e^X, e^Y]_i^T$ represent the two orthogonal components ⁷X, Y of this 2D vector⁸ for radiation measured at antenna i . Note that \vec{E}_i represents a continuous signal at one instant in time.

The radiation from an astrophysical source is modified when it propagates through the Earth's atmosphere and is measured by a an electronic receiver system. Jones matrices⁹ describe this modulation for the incident electric field as it passes through various elements of the measurement system. Such effects can be instrumental or non instrumental, and may or may not depend on the direction on the sky. A sequence of these effects are represented by a product of individual Jones matrices. Direction-independent effects for antenna i are usually described as $[J_i^{vis}] = [J_i^G][J_i^D][J_i^C]$, a 2×2 matrix product of complex antenna gains (J^G), polarization leakage between the nominally orthogonal dipoles (J^D) and feed configuration (J^C). Direction-dependent effects are described by $[J_i^{sky}] = [J_i^E][J_i^P][J_i^F]$, a product of antenna illumination patterns (J^E), parallactic angle effects (J^P) and tropospheric and ionospheric effects and Faraday rotation (J^F).

The two-component Jones vector measured at each antenna is

$$\vec{E}_i^{obs} = [J_i]\vec{E}_i \quad \text{where} \quad [J_i]_{2 \times 2} = [J_i^{vis}][J_i^{sky}] \quad (2.15)$$

Linear polarization components (X, Y) of the electric field are measured using a pair of dipoles positioned perpendicular to each other and orthogonal to the direction of propagation of the incident radiation. Circular polarization components (R, L) are measured using a pair of helical antennas, and signals can be electronically converted between linear and circular, if required. The measured E-field is in the form of a time-series of voltages for each polarization component. These signals are amplified and then sent to a backend system that applies delay corrections and computes visibilities. The signals can be digitized before or after delay correction or correlation.

⁷This discussion uses X, Y to denote the two orthogonal linear polarization components of an EM-wave. These derivations will hold if X, Y are replaced by R, L for right and left circular polarization states.

⁸Notation : Matrices are denoted by $[A]$. Vectors are denoted by \vec{A} or $[A]_{n \times 1}$ (for an n -element vector). The T superscript denotes a matrix transpose, and the \dagger superscript denotes conjugate transpose or operator adjoint.

⁹ The vector \vec{E}_i is a Jones vector; a commonly used notation to describe polarized light. A Jones matrix is a complex-valued 2×2 matrix operator that describes the effect of passing an EM-wave through a system that modifies it. It acts on an input Jones vector to produce an output Jones vector of modified EM-wave components. For example, for a measurement that uses a radio receiver, the diagonal elements of the Jones matrix correspond to instrumental gains that are applied to each component of \vec{E}_i and the off-diagonal elements describe the amount of leakage introduced between them during the measurement process.

2.2.1.2 Correlation for each Baseline

According to Eqn. 2.6, a visibility (or correlation coefficient) is measured as the time-averaged product of the complex amplitudes of the E-fields incident at each detector pair ($E_\nu(\vec{r}_1)$ and $E_\nu^*(\vec{r}_2)$). However, in practice, neither the incident radiation nor the measurement system is truly monochromatic. Also, the E-field component of the EM-wave $E(\vec{r}, t)$ incident at each detector varies with time. Therefore, the first step in the measurement process is to sample the incident $E(\vec{r}, t)$ at a finite time resolution. To represent the complete signal, the sampling time interval must be shorter than the reciprocal of twice the signal bandwidth (the Nyquist rate).

There are two ways of computing the correlation coefficient for each detector pair using these high time-resolution samples. In both cases, the result is obtained at a finite time-resolution $\Delta\tau$ (the desired integration time), a finite frequency-resolution $\Delta\nu$ (the desired channel width) and across a total bandwidth (controlled by the signal sampling rate). The first method is known as an FX correlation. In this method, we accumulate measurements of $E(\vec{r}, t)$ over a time interval t_{max} , compute its temporal Fourier transform to obtain $\xi_\nu(\vec{r})$ at a set of different discrete frequencies ν separated by $\Delta\nu = 1/t_{max}$, compute the product $E_\nu(\vec{r}_1)E_\nu^*(\vec{r}_2)$ for each ν and then average the results over the desired integration time $\Delta\tau$ (again, for each ν). The second method is known as an XF correlation. Here, we use the high time-resolution measurements of $E(\vec{r}, t)$ to compute the correlation product $E(\vec{r}_1, t)E^*(\vec{r}_2, t - \tau_{lag})$ for a series of time lags (τ_{lag}), and then compute the temporal Fourier transform of this product to obtain the power spectrum $E_\nu(\vec{r}_1)E_\nu^*(\vec{r}_2)$ at a frequency resolution of $\Delta\nu = 1/t_{lag}$, and finally average the results over the desired integration time $\Delta\tau$ (for each ν). The output from the correlator is a series of visibilities (discrete samples of the continuous visibility function).

$$\vec{V}_{ij}^{obs} = \langle \vec{E}_i^{obs} \otimes \vec{E}_j^{obs*} \rangle = \begin{bmatrix} \langle e_i^p e_j^{p*} \rangle \\ \langle e_i^p e_j^{q*} \rangle \\ \langle e_i^q e_j^{p*} \rangle \\ \langle e_i^q e_j^{q*} \rangle \end{bmatrix}^{obs} = \begin{bmatrix} V_{ij}^{pp} \\ V_{ij}^{pq} \\ V_{ij}^{qp} \\ V_{ij}^{qq} \end{bmatrix}^{obs} \quad (2.16)$$

where $\langle \rangle$ denotes a time-average and \otimes denotes an outer-product that generates four cross-correlation pairs (two cross-hand pq, qp and two parallel-hand XX, YY) per baseline¹⁰. e_i^p, e_i^q are the elements of \vec{E}_i^{obs} . The time average represents a discretization of the continuous signals at a sampling rate given by the integration time per visibility $\Delta\tau$. $[V_{ij}^{obs}]$ is a 4×1 coherency vector¹¹ for the baseline formed from antennas i and j and it can be written

¹⁰ The outer product (direct, tensor or Kronecker product) of two matrices $[A]$ and $[B]$ is given by a matrix where a_{ij} is replaced by $a_{ij}[B]$. Therefore, for two vectors $\vec{A} = [A]_{2 \times 1}$ and $\vec{B} = [B]_{2 \times 1}$ the outer product is a 4×1 vector given by $[a_1 b_1, a_1 b_2, a_2, b_1, a_2, b_2]^T$. For two 2×2 matrices the outer product is a 4×4 matrix where the i, j quadrant is given by $a_{ij}[B]_{2 \times 2}$. An important property of these outer products is $[A \otimes B][C \otimes D] = [AC] \otimes [BD]$.

¹¹The coherency vector is a 4×1 vector of cross-correlations formed from the four elements in the outer

in terms of the antenna-based Jones matrices as follows.

$$\vec{V}_{ij}^{obs} = \langle \vec{E}_i^{obs} \otimes \vec{E}_j^{obs*} \rangle = \langle [J_i] \vec{E}_i \otimes [J_j]^* \vec{E}_j^* \rangle = ([J_i] \otimes [J_j]^*) \langle \vec{E}_i \otimes \vec{E}_j^* \rangle = [K_{ij}] \vec{V}_{ij} \quad (2.17)$$

Therefore the measured coherence vector of visibilities is given by $\vec{V}_{ij}^{obs} = [K_{ij}] \vec{V}_{ij}$ where $[K_{ij}] = [J_i] \otimes [J_j]$ is a 4×4 matrix and \vec{V}_{ij} is the true visibility that Eqn. 2.6 measures. (If only one polarization component of the E-field is measured (say p), the Jones matrices and vectors become scalars (only one non-zero element) and Eqn. 2.17 simplifies to a single complex number per baseline $V_{ij}^{obs} = g_i g_j^* V_{ij}$, where $[J_i] = g_i$ represents a multiplicative complex gain for antenna i .)

2.2.1.3 Measurement Equation for one baseline

The ideal van Cittert Zernike theorem (Eqn. 2.9) can now be combined with the effect of the measurement process, to derive the full-polarization measurement equation. The visibility function sampled by baseline ij at one instant in time and at one frequency is given as follows¹².

$$\vec{V}_{ij}^{obs}(u, v) = [K_{ij}^{vis}] \iint [D_{ij}^{sky}(l, m)] \vec{I}^{sky}(l, m) e^{-2\pi i(u l + v m)} dl dm \quad (2.18)$$

Here, $\vec{I}^{sky}(l, m)$ is a 4×1 vector of the sky brightness distribution (in the direction l, m) corresponding to the four correlation pairs. u, v represents the spatial frequency sampled by baseline ij at one instant in time (given by the components of \vec{b}_{ij} in units of λ). $[K_{ij}^{vis}]$ is a 4×4 matrix that represents direction-independent instrumental effects that are constant across the field of view of each antenna (e.g. receiver gains). $[D_{ij}^{sky}(l, m)]$ is a 4×4 matrix that represents effects that vary with position on the sky (e.g. antenna primary beams, pointing offsets, ionospheric effects and the w -term).

The effect of $D_{ij}^{sky}(l, m)$ in Eqn. 2.18 is multiplicative in the image domain and can be represented as a convolution in the visibility domain. Let $K_{ij}^{dd}(u, v)$ represent the Fourier transform of $D_{ij}^{sky}(l, m)$ (for each of the four correlation pairs). Eqn. 2.18 can be re-written as follows.

$$\vec{V}_{ij}^{obs}(u, v) = [K_{ij}^{vis}] \left\{ [K_{ij}^{dd}(u, v)] \star \iint \vec{I}^{sky}(l, m) e^{-2\pi i(u l + v m)} dl dm \right\} \quad (2.19)$$

Here, \star represents convolution for each correlation product.

product of two 2×1 Jones vectors.

¹² In practice, each measurement is made over a finite bandwidth $\Delta\nu$ and time range $\Delta\tau$ and contains the integral of the visibility function over these time and frequency ranges. Section 2.2.2 elaborates on this discretization.

Eqns. 2.18 and 2.19 describe the measurement equation for one visibility. In practice, \vec{V}_{ij}^{obs} is measured for all $\frac{N_a(N_a-1)}{2}$ pairs of antennas ($i = 1 - N_a$ and $j = i - N_a$) for a series of integration timesteps and observing frequencies. All visibilities (baselines, timesteps and frequencies) for each correlation product XX, XY, YX, YY are then combined for imaging¹³.

The next section rewrites Eqn. 2.19 in a form where the sky brightness is no longer a continuous function of position l, m , but is described by a discrete set of parameters (e.g. pixels of an image of the sky). The true visibility function is also discretized and this allows us to represent the spatial frequency sampling function (uv -coverage) in the form of a matrix operator. The complete measurement equation can then be written as a matrix equation, or a system of linear equations that need to be solved in order to reconstruct an image of the input sky brightness distribution.

2.2.2 Measurement Equation for Synthesis Imaging

This section introduces the use of standard linear-algebra to describe the measurement process of an imaging interferometer. The sky brightness distribution is parameterized in some basis and the measured visibilities are expressed as functions of the sky parameters. The solution of the measurement equation can then be treated as a numerical optimization problem. This section introduces the linear-algebra notation that will be used in the rest of this dissertation to describe the measurement equation for various image parameterizations, instrumental effects and image reconstruction algorithms.

2.2.2.1 Generic measurement equation

Let the sky brightness distribution be described by m parameters listed in vector form as $\vec{I}_{m \times 1}^{sky}$, and let $\vec{V}_{n \times 1}^{obs}$ be a vector of n visibilities¹⁴. A generic measurement equation can be written as

$$\vec{V}_{n \times 1}^{obs} = [A_{n \times m}] \vec{I}_{m \times 1}^{sky} \quad (2.20)$$

where $[A]$ describes the process of making n measurements of the visibility function of the sky brightness distribution in terms of the m image parameters. $[A]$ is a generic label for a measurement matrix and the following chapters will discuss measurement equations using different specific forms of $[A]$.

The next few sections describe how various parts of Eqn. 2.19 are represented in this matrix notation and combined to construct the full measurement matrix $[A_{n \times m}]$.

¹³ The 4 correlations can either be imaged directly or after computing a Stokes vector I, Q, U, V of visibilities (via a linear 4×4 transform [Sault et al. 1996]).

¹⁴ Typically, $m = N_{pix}^2$ for an image of size $N_{pix} \times N_{pix}$, parameterized by its pixel amplitudes, and $n = \frac{N_a(N_a-1)}{2} \times N_{frequency\ channels} \times N_{timesteps} \times N_{correlation\ pairs}$

2.2.2.2 Discretization of the visibility and image domains

The uv -coverage described by Eqn. 2.11 is a set of δ -functions located on a continuous spatial frequency plane. However, in practice, visibility samples from each baseline are measured at finite time and frequency resolution ($\Delta\tau, \Delta\nu$). Note that $\Delta\tau, \Delta\nu$ always need to be smaller than the limits set by the temporal and spectral coherence of the incident radiation. When mapped to the spatial frequency plane, the shortest baseline will give the smallest $\Delta u, \Delta v$ that the interferometer measures¹⁵. Let us construct a spatial-frequency grid with cell sizes defined by $\min(\Delta u, \Delta v)$, such that all visibility measurements naturally map directly to pixels on this grid, and limits due to signal coherence are also satisfied. Let the number of uv -pixels be m , such that the largest measured spatial frequency is accounted for. A discrete Fourier transform (DFT) of this grid corresponds to an image of the sky extending across a field of view given by $\frac{1}{\Delta u}, \frac{1}{\Delta v}$ radians, and pixel size defined by the maximum spatial frequency covered by the uv -grid.

Let $\vec{I}_{m \times 1}^{sky}$ represent a one-dimensional pixelated image of the sky, over the entire field of view allowed by the measurements¹⁶. The complete but discretized visibility function for the sky brightness is then described as $\vec{V}_{m \times 1}^{sky} = [F_{m \times m}] \vec{I}_{m \times 1}^{sky}$, where $[F_{m \times m}]$ is the DFT operator¹⁷. This analysis can be directly generalized to two dimensions.

When all four correlation pairs $\{XX, XY, YX, YY\}$ are measured, we can write $\vec{I}_{4m \times 1}^{sky}$ and $\vec{V}_{4m \times 1}^{sky}$ as stacks of 4 vectors, each m pixels long and representing one polarization pair. The DFT operator becomes a 4×4 block diagonal matrix and will be denoted by $[F_{4m \times 4m}]$.

¹⁵This relation is derived from $\Delta u = \frac{\partial u(t, v)}{\partial t} \Delta t + \frac{\partial u(t, v)}{\partial v} \Delta v$ where $u(t, v)$ is given by Eqn. 2.7. The hour angle H is a function of time, $\lambda = c/v$ and x, y, z are the lengths of the shortest baseline.

¹⁶A pixel-based flux model is the most widely-used form of image parameterization, and is sufficient to describe all the main concepts related to image reconstruction *via* standard algorithms. The main focus of this dissertation is the use of advanced image parameterizations for multi-scale and multi-frequency image models. The models chosen for these algorithms can be described as linear combinations of pixelated images, and this formulation remains valid. (In this dissertation, non-pixel methods are discussed only when relevant.)

¹⁷The normalization convention used for all Fourier transforms described here is such that $[F^\dagger F] = m[\vec{I}_m]$, where $[\vec{I}_m]$ is an $m \times m$ identity matrix. The normalization is chosen as part of the reverse/inverse transform $[F]^{-1} = \frac{1}{m}[F^\dagger]$. Therefore, F is *not* a unitary operator. This choice is in accordance with the amplitude normalization convention used in radio interferometry. For a 5Jy point source at the phase center, calibrated visibilities are normalized to an amplitude 5Jy. While making an image, the amplitude of a point source at the phase center can be calculated as the vector average of n such visibilities (involving a normalization by n). In practice this is a weighted average, and a normalization by the sum of weights is done separately, only for the reverse (inverse) transform. For efficiency, the FFT algorithm [Cooley and Tukey 1965] is used to implement all Fourier transforms (unless otherwise stated). Note also that the FFT algorithm requires a regularly sampled set of data points, whereas a DFT explicitly evaluates the Fourier transform integral and can be computed for an irregularly sampled set of data points.

2.2.2.3 Spatial frequency coverage in matrix notation

The uv -coverage of a synthesis array (described in section 2.1.1.5) can be written as a sampling matrix $[S_{n \times m}]$ defined on this fine spatial-frequency grid. It operates on $\vec{V}_{m \times 1}^{\text{sky}}$, to yield n visibility measurements. $[S_{n \times m}]$ is a projection operator that maps elements from an $m \times 1$ list onto a list of n measurements, and contains only ones and zeros (the uv -coverage listed in Eqn. 2.11 consists of Kronecker δ -functions). Each row in $S_{n \times m}$ picks out one spatial frequency, and therefore can have only one non-zero entry. There can however be multiple measurements of the same spatial frequency, and columns of $[S_{n \times m}]$ can have more than one non-zero entry. Unmeasured spatial frequencies correspond to columns of $[S_{n \times m}]$ with no non-zero elements (the column rank of $[S_{n \times m}]$ is $< m$).

The same sampling function applies to all four correlation pairs $\{XX, XY, YX, YY\}$. Therefore, full-polarization sampling can be described by a $4n \times 4m$ block-diagonal matrix constructed from 4 instances of $[S_{n \times m}]$.

2.2.2.4 Direction-independent effects in matrix notation

Direction-independent instrumental effects (described in section 2.2.1.1, and denoted by K_{ij}^{vis} in Eqn. 2.19) can be written in matrix form for all n baselines and all 4 correlation pairs $\{XX, XY, YX, YY\}$. Let $[K_{4n \times 4n}^{\text{vis}}]$ be a 4×4 block matrix constructed from diagonal matrices of size $n \times n$ (when each element of $[K_{ij}^{\text{vis}}]_{4 \times 4}$ in Eqn. 2.18 is written out for all n baselines, it forms one $n \times n$ block with non-zero elements only on the diagonal). Non-zero off-diagonal blocks in these full-polarization matrices describe the coupling between different polarizations during the measurement process (*i.e.* off-diagonal terms of Eqn. 2.17).

2.2.2.5 Direction-dependent effects in matrix notation

Eqn. 2.19 shows that the visibilities measured by baseline ij are the result of a convolution of the true visibility function with a 2D function $K_{ij}^{dd}(u, v)$ that represents direction-dependent effects¹⁸. The visibility measured by baseline ij is no longer a sample of the visibility function at one spatial frequency, but the integral of the visibility function over a region defined by the shape of $K_{ij}^{dd}(u, v)$ around that one spatial frequency.

For each correlation pair, we can define a visibility-domain operator that convolves the true visibility function with \vec{K}_{ij}^{dd} before baseline ij samples it. Let $[S_{n \times m}^{dd}]$ represent a modified form of the sampling matrix $[S_{n \times m}]$ in which each row contains the vector \vec{K}_{ij}^{dd} centered at the spatial frequency measured by that baseline (given by the location of the corresponding δ -function in $[S_{n \times m}]$). The subscript ij indicates that these effects can be dif-

¹⁸ $K_{ij}^{dd}(u, v)$ is one element of the 4×4 matrix $[K_{ij}^{dd}(u, v)]$ used in Eqn. 2.19 and represents a uv -plane convolution function for one correlation pair.

ferent for different baselines and times. Therefore, \vec{K}_{ij}^{dd} can vary across the rows of $[S_{n \times m}^{dd}]$. The effect of multiplying $[S_{n \times m}^{dd}]$ with the true visibility function $\vec{V}_{m \times 1}^{sky}$ is a baseline-based convolution during the sampling process.

When all baselines have the same direction-dependent effects ($\vec{K}_{ij}^{dd} = \vec{K}^{dd}$) the sampling function can be separated from this baseline-based convolution. We can write $[S_{n \times m}^{dd}] = [S_{n \times m}][G^{dd}]$ where $[G^{dd}] = [FD^{sky}F^\dagger]$ is a convolution operator¹⁹ with \vec{K}^{dd} as the convolution kernel²⁰. $[D_{m \times m}^{sky}] = diag([F^\dagger]\vec{K}^{dd})$ is a diagonal matrix that represents the multiplicative image-domain effect of the visibility-domain convolution (compare with D^{sky} in Eqn. 2.18).

When all four correlation pairs are measured, the sampling matrix becomes a $4n \times 4m$ block matrix. Each $n \times m$ block contains $[S_{n \times m}^{dd}]$ constructed with a \vec{K}_{ij}^{dd} for the corresponding correlation pair. An important difference between $[S_{4n \times 4m}^{dd}]$ and $[S_{4n \times 4m}]$ is that $[S^{dd}]$ contains non-zero off-diagonal blocks that describe the coupling between the different polarizations.

2.2.2.6 Measurement equations in matrix form

The full measurement equation in block matrix form is given by writing Eqn. 2.19 for all baselines and combining it with the uv -coverage and other instrumental effects.

$$\vec{V}_{4n \times 1}^{obs} = [K_{4n \times 4n}^{vis}][S_{4n \times 4m}^{dd}][F_{4m \times 4m}]\vec{I}_{4m \times 1}^{sky} \quad (2.21)$$

where $\vec{V}_{4n \times 1}^{obs}$ consist of 4 segments of n visibilities each (one for each correlation pair). From this equations, we see that the measurement matrix $[A]$ in Eqn. 2.20 can be written as a product of a series of matrices as $[A_{4n \times 4m}] = [K_{4n \times 4n}^{vis}][S_{4n \times 4m}^{dd}][F_{4m \times 4m}]$. A solution of the complete measurement equation includes imaging and deconvolution along with the correction of direction-independent and dependent effects, both for all polarization components of the incident radiation and their correlations.

The algorithms described in this dissertation will focus on visibility data from only one correlation pair, assuming that the incident radiation is either unpolarized or has no linear polarization (when the X,Y components are measured and Q=0) or no circular polarization (when the R,L components are measured and V=0). In this case, the dimensions

¹⁹ The convolution of two vectors $\vec{a} \star \vec{b}$ is equivalent to the multiplication of their Fourier transforms. A 1-D convolution operator is constructed from \vec{a} and applied to \vec{b} as follows. Let $[A] = diag(\vec{a})$. Then, $\vec{a} \star \vec{b} = [F^\dagger diag([F]\vec{a})F]\vec{b} = [C]\vec{b}$. Here, $[F]$ is the Discrete Fourier Transform (DFT) operator. $[C]$ is a Toeplitz matrix, with each row containing a shifted version of \vec{a} . Multiplication of $[C]$ with \vec{b} implements the shift-multiply-add sequence required for the process of convolution.

²⁰The function with which a convolution is done is called the convolution kernel. It is the function that is shifted to all pixel locations during the shift-multiply-add sequence of convolution. For a convolution kernel \vec{a} , an image-domain convolution operator is constructed as $[F^\dagger diag([F]\vec{a})F]$, and a visibility-domain convolution operator is constructed as $[F diag([F]\vec{a})F^\dagger]$

of all the matrices in Eqn. 2.21 lose the factor of 4, and $[K_{n \times n}^{vis}]$ is a diagonal matrix. The measurement equations for observing unpolarized incident radiation and recording only one correlation pair are given below (matrix equivalent of Eqn. 2.19).

$$\vec{V}_{n \times 1}^{obs} = [K_{n \times n}^{vis}][S_{n \times m}^{dd}][F_{m \times m}]\vec{I}_{m \times 1}^{sky} \quad (2.22)$$

When instrumental effects are time-invariant and identical for all baselines, they can be factored out of the sampling matrix ($[S_{n \times m}^{dd}] = [S_{n \times m}][G_{m \times m}^{dd}]$) and written in the image-domain (matrix equivalent of Eqn. 2.18).

$$\vec{V}_{n \times 1}^{obs} = [K_{n \times n}^{vis}][S_{n \times m}][F_{m \times m}][D_{m \times m}^{sky}]\vec{I}_{m \times 1}^{sky} \quad (2.23)$$

In general, $\vec{V}_{n \times 1}^{obs}$, $[S]$ and $[F]$ are known and \vec{I}^{sky} , $[K^{vis}]$ and $[D^{sky}]$ are unknown. Estimates for $[K^{vis}]$ and $[D^{sky}]$ are obtained either by solution from the measured data or from existing measurements or models, leaving only \vec{I}^{sky} as the unknown variable to solve for.

The next two chapters describe the solution of the measurement equations shown in Eqns. 2.22 and 2.23. Standard synthesis imaging techniques address imaging and deconvolution with the correction of only direction independent effects. They solve Eqn. 2.23 by ignoring $[D^{sky}]$ and estimating $[K^{vis}]$ from separate observations of a source for which \vec{I}^{sky} is known. Chapter 3 describes these standard methods in detail. Techniques for correcting direction-dependent effects solve Eqn. 2.22 and use *a-priori* estimates for \vec{K}_{ij}^{dd} used to construct $[S^{dd}]$. These more recent techniques are described in Chapter 4. Chapters 6 and 7 describe and solve extensions of these measurement equations for broad-band radio interferometry in which the sky brightness distribution, the spatial frequency sampling pattern and instrumental effects vary with observing frequency.

CHAPTER 3

STANDARD CALIBRATION AND IMAGING

This chapter describes well-established calibration and imaging algorithms in the context of a linear-least-squares solution of the measurement equation. The algorithms described in this chapter follow the general ideas in [Taylor et al. \[1999\]](#) and [Briggs \[1995\]](#), and cover the calibration of direction-independent instrumental effects, and image reconstruction *via* an iterative deconvolution process.

To begin with, let us consider a simplified form of the measurement equation (given in Eqns.[2.21](#) and [2.23](#)) for only one correlation product and only direction-independent instrumental effects [K^{vis}].

$$\vec{V}_{n \times 1}^{obs} = [K_{n \times n}^{vis}][S_{n \times m}F_{m \times m}]\vec{I}_{m \times 1}^{sky} \quad (3.1)$$

The unknowns in Eqn. [3.1](#) are the sky brightness \vec{I}^{sky} and the elements of [K^{vis}]. Calibration (Section [3.1](#)) is the process of computing and applying an approximate inverse of [K^{vis}]. Imaging (Section [3.2](#)) is the process of reconstructing the sky brightness, \vec{I}^{sky} , by removing the effect of the instrument's incomplete spatial frequency sampling (extensions to the full polarization case are made within the discussions in Sections [3.1](#) and [3.2](#), and direction-dependent instrumental effects [D^{sky}] are discussed in Chapter [4](#)).

3.1 Calibration

To make an image that represents the true sky brightness distribution, the measured visibility data must first be calibrated to undo various instrumental effects that corrupt the incoming signals. Calibration is the process of first computing the elements of [K^{vis}] from visibility measurements of a source whose structure is known, and then using these solutions to remove the effect of direction-independent complex gains from the observed visibilities of the source of interest.

This section describes the basic procedure for calibrating visibility data, lists various types of calibration schemes, and briefly describes full-polarization calibration.

3.1.1 Gain solution and correction

The elements of $[K^{vis}]$ are computed by solving Eqn. 3.1, written in the following form.

$$\vec{V}_{n \times 1}^{obs} = [K_{n \times n}^{vis}] \vec{V}_{n \times 1}^{model} \quad (3.2)$$

where $\vec{V}_{n \times 1}^{model} = [S_{n \times m} F_{m \times m}] \vec{I}_{m \times 1}^{model}$ are visibilities that are computed from a known model of the source $\vec{I}_{m \times 1}^{model}$ by taking its spatial Fourier transform and sampling the result using S .

For the simple case of only one correlation pair, each element on the diagonal of $[K^{vis}]$ can be described as a product of two complex numbers. $K_{ij}^{vis} = g_i^p g_j^{*p}$ where g_i^p and g_j^p are multiplicative instrumental gains for antennas i and j . These complex gains are Jones matrix elements for the polarization components used to construct the correlation. The number of unknowns in this system is N_a , and $V_{n \times 1}^{model}$ provides $O(N_a^2)$ constraints to uniquely factor the baseline-based K_{ij}^{vis} into N_a antenna-based complex gains g_i^p . A weighted least-squares solution [Cornwell and Wilkinson 1981] of Eqn. 3.2 is found by minimizing

$$\chi^2 = \sum_{ij} w_{ij} |V_{ij}^{obs} - g_i g_j^{*} V_{ij}^{model}|^2 \quad (3.3)$$

and directly estimating antenna-based complex gains, where w_{ij} is a measured visibility weight, given by the inverse of the noise variance.

Gain corrections for all baselines (diagonal elements of $[K_{n \times n}^{vis}]^+$) are computed from the antenna-based gain solutions as $K_{ij}^{vis*} = 1/(g_i g_j^*)$ (for the element corresponding to baseline ij) and then applied to the observed visibilities to correct them¹.

$$\vec{V}_{n \times 1}^{corr} = [K_{n \times n}^{vis}]^+ \vec{V}_{n \times 1}^{obs} \quad (3.4)$$

An alternate formulation expresses the $N_a(N_a - 1)/2$ elements of $[K^{vis}]$ as an $N_a \times N_a$ correlation matrix with element K_{ij} in the i^{th} row and j^{th} column, and uses eigenvalue decompositions to solve for antenna-based complex gains. In cases where the measurements at each baseline contain random additive noise that cannot be factored into antenna-based terms (closure noise), baseline-based calibration is sometimes done to solve for the elements of $[K^{vis}]$ directly. However, this process is poorly constrained compared to standard antenna-based calibration, is not always a physically accurate approach, and must be used with caution.

¹ The + superscript denotes the pseudo-inverse of a matrix. A pseudo-inverse is an approximate inverse of a matrix. It is often used when an exact inversion is either impossible or intractable, either when the matrix being inverted is rank-deficient and has no inverse, or when the presence of noise in the data prevents an exact solution. A pseudo-inverse is often used to obtain a least-squares solution of a system of equations in the presence of noise. One way of computing the pseudo-inverse of a matrix $[A]$ is $[A^+] = [A^\dagger A]^{-1} [A^\dagger]$. This involves computing and inverting $[A^\dagger A]$ or some approximation of it, say, a diagonal approximation. Other methods use various matrix decompositions of $[A]$ to construct $[A^+]$.

3.1.2 Types of Calibration

Several commonly used calibration techniques are briefly summarized below.

3.1.2.1 Standard Calibration

For standard calibration, astronomical sources of known amplitude and/or structure are observed at regular intervals during an observation of a source of interest. The known true/model visibilities are used to compute antenna-based gain solutions for the time intervals over which the calibrator was observed. These gain solutions are interpolated across the time ranges where the source of interest is observed, and used to correct the observed visibilities (see [Fomalont and Perley \[1999\]](#); [Cornwell and Fomalont \[1999\]](#)).

The solution for antenna gains is often split into computing amplitudes and phases separately. Bright sources whose amplitudes are well known and do not vary with time are used as flux calibrators to compute gain amplitudes. Sources whose absolute positions are accurately known are used as phase calibrators to constrain gain phases. Ideal calibrators are extremely compact sources whose visibility functions are constant across the range of spatial frequencies measured by the synthesis array, but extended sources can also be used if their structure is also accurately known *a-priori*. Bandpass calibrators are flat-spectrum sources or those with a well-known spectral behaviour, and are used to compute the variation of instrumental gains as a function of frequency.

To increase the signal-to-noise ratio of correlations going into the algorithm that solves for the elements of $[K^{vis}]$, the visibility data are sometimes pre-averaged along data axes over which the solution is likely to remain stable. For example, bandpass calibration often uses time-averaged data because the bandpass shape is usually stable across certain time-intervals. Time-variable gain fluctuations are solved for during a second pass, where the now calibrated bandpasses are averaged across frequency to give a single measurement for each time-step.

3.1.2.2 Self Calibration

Since gain solutions for the target source are computed only by interpolating between calibrator scans, any gain fluctuations during the time when the target source is being observed will not be accounted for. Self-calibration is a process where a model of the target source itself is used to compute gain solutions during the time it is being observed. This model of the target source could be from *a-priori* information in the form of an existing image, or could be built up by a bootstrap method from the observed data. In general, self-calibration [[Schwab 1980](#); [Cornwell and Wilkinson 1981](#); [Thompson and Daddario 1982](#)] is an iterative combination of calibration and imaging. It is a two-stage χ^2 minimization process that iterates between the parameter subspaces of \vec{I}^{sky} and $[K^{vis}]$ and applies constraints appropriate to the different physics involved. During the computation

of $[K^{vis}]$ for calibration, the most current model of \vec{I}^{sky} is held constant and used in evaluating Eqn. 3.1 to compute model visibilities. Similarly, during imaging, the most current calibration solutions ($[K^{vis}]$) are applied and held constant.

If a high quality initial model of \vec{I}^{sky} is available, self-calibration often requires only one iteration. Depending on the availability of an external calibrator source, this calibration stage solves for either gain amplitudes or gain phases, or both. For example, a standard flux calibration *via* an external flux calibrator can be followed by a phase-only self-calibration step using a model whose structure is known to be the same as the target source. If an amplitude and phase calibration is required but the model and target differ in amplitude, the solution gain vector is scaled to unit norm to preserve the overall flux level of the target source.

When there is no *a-priori* information about the source or an external calibrator, the initial sky model is chosen as a point source of unit flux at the phase center and all antenna gains are unity. In this general case, several iterations of calibration and imaging are usually required before both the calibration solutions and the sky model converge to stable values. Also, the absolute position of the source (given by a common phase term across all antennas) and its absolute amplitude, are absorbed into the gain solutions, and are lost when the gain correction is applied. This iterative process is usually feasible only for sources with simple spatial structure.

3.1.2.3 Peeling

Peeling [Nijboer and Noordam 2007] is a technique where self-calibration is done one source at a time, with the calibration being undone after each source has been subtracted and replaced with a model. Peeling can either be done on all prominent sources one after another, or in combination with regular self-calibration in which it is applied only to sources whose calibration parameters differ significantly from a global solution. This method accounts for some directional dependence of the antenna gains, by calculating them separately along a few directions containing bright sources.

3.1.2.4 Full-polarization calibration

Full-polarization measurements contain correlations from all four polarization pairs. Each baseline measures the product of $[K_{ij}^{vis}] = [J_i^{vis}] \otimes [J_j^{vis}]^\dagger$ with the true coherence vector seen by that baseline. Eqn. 3.2 becomes

$$\vec{V}_{4n \times 1}^{obs} = [K_{4n \times 4n}^{vis}] \vec{V}_{4n \times 1}^{model} \quad (3.5)$$

and the elements of K_{ij}^{vis} are computed as described in section 3.1.1. For a source with known polarization characteristics, the true coherence vector is known (constant $\times [1, 0, 0, 1]$ for circular feeds and an unpolarised source) and one can form a system of linear equations

with the elements of $[K_{ij}^{vis}]$ as unknowns. For a single baseline, there are up to 10 degrees of freedom and 4 equations [Sault et al. 1996]. However, with an *a-priori* source model, measurements from all baselines provide enough constraints to uniquely factor the baseline-based $[K_{ij}^{vis}]$ matrices into antenna-based 2×2 Jones matrices ($4 \times N_a(N_a - 1)/2$ equations and $4 \times N_a$ unknowns). In its most general form, the elements of $[J_i^{vis}]$ can be computed by minimizing

$$\chi^2 = \sum_{ij} |\vec{V}_{ij}^{obs} - [J_i^{vis} \otimes J_j^{vis\dagger}] \vec{V}_{ij}^m|^2 \quad (3.6)$$

with respect to the antenna-based $[J_i^{vis}]$. Corrections can be applied by direct computation of $[K^{vis+}]$ from these solutions.

To simplify this solution process, polarization calibration is usually done in stages. First, only the diagonal elements of the Jones matrices are solved for, assuming zero leakage between the orthogonal feeds. Corrections are applied and a second stage solves only for the off-diagonal terms. Another method of simultaneously solving for antenna-based gains and leakages from only parallel-hand correlations XX, YY is described in Bhatnagar and Nityananda [2001].

3.2 Imaging

After calibration, the corrected visibilities \vec{V}_{nx1}^{corr} are ready to be converted into an image. The complex visibilities are mapped onto the spatial frequency grid via S_{mxn}^\dagger . An inverse Fourier transform of these gridded visibilities gives the raw or dirty image over the full field of view allowed by the time and frequency resolution of the visibility measurements². Full image reconstruction involves the removal of the effect of the instrument's known sampling function (*uv*-coverage). In interferometric imaging, there are some spatial frequencies that are actually not measured, so even if the instrument's transfer function (effect on the incoming signal) is completely known, the reconstruction of the sky brightness is a non-linear process. This is because it involves estimating the values of the visibility function at unmeasured regions of the spatial frequency plane. Various physical constraints are required to achieve this.

This section describes the process of interferometric image reconstruction in terms of the matrix equations being solved. Several linear-algebra concepts are introduced here to emphasize the relation between imaging techniques currently in use and the application of standard numerical optimization theory to solve inverse problems. Chapters 4 and 6 will later apply these same numerical optimization ideas to more complicated systems of equations, to derive imaging algorithms for multi-scale, multi-frequency image models along with wide-field instrumental effects.

²In practice, an image is usually made over a smaller field of view, and this is accomplished by resampling the visibilities onto a coarser spatial frequency grid before Fourier inversion. See Section 3.2.2 on gridding.

3.2.1 Writing and Solving the Imaging Equations

This section describes the imaging properties of the instrument, and introduces the standard algorithmic framework used by most radio interferometric imaging techniques. Sections 3.2.2, 3.2.3 and 3.2.4 later list details of the main computational steps involved in this image reconstruction process.

3.2.1.1 Measurement Equations

Using Eqns. 3.1 and 3.4, the measurement equation after calibration is given by

$$[S_{n \times m} F_{m \times m}] \vec{I}_{m \times 1}^{sky} = \vec{V}_{n \times 1}^{corr} \quad (3.7)$$

Here, \vec{I}^{sky} represents the sky brightness as a set of pixel amplitudes, and \vec{V}^{corr} is a list of measured visibilities. The measurement matrix ($[A]$ in Eqn. 2.20) is given by $[A] = [S][F]$.

3.2.1.2 Normal Equations

A weighted least-squares estimate of \vec{I}^{sky} is found by solving the normal equations³ constructed from the above measurement equation.

$$[F^\dagger S^\dagger W S F] \vec{I}_{m \times 1}^{sky} = [F^\dagger S^\dagger W] \vec{V}_{n \times 1}^{corr} \quad (3.8)$$

Here, $[W_{n \times n}]$ is a diagonal matrix of signal-to-noise-based measurement weights and $[S^\dagger]$ denotes the mapping of measured visibilities onto a regular grid of spatial frequencies⁴. The matrix on the LHS of Eqn. 3.8 is called the Hessian matrix $[H]$ and it describes the imaging properties of the instrument. The vector on the RHS is the dirty image \vec{I}^{dirty} defined as the image produced by direct Fourier inversion of the calibrated and gridded visibilities.

$$[H] = [F^\dagger S^\dagger W S F] \quad (3.9)$$

$$\vec{I}^{dirty} = [F^\dagger S^\dagger W] \vec{V}_{n \times 1}^{corr} \quad (3.10)$$

In the next two sections, we will describe the properties of $[H]$, define the point spread function \vec{I}^{psf} , and show that for standard interferometric imaging, Eqn. 3.8 describes the dirty image as the result of a convolution between the sky brightness and the point spread function (i.e. a discretized and 1-D version of Eqn. 2.13).

³ The weighted least-squares solution for a system of linear equations $[A]\vec{x} = \vec{b}$ is found by forming and solving the normal equations $[A^\dagger W A]\vec{x} = [A^\dagger W]\vec{b}$. Here, $[A] = [S][F]$ is the measurement matrix, $[W]$ is a diagonal matrix of weights and $[H] = [A^\dagger W A]$ is called the Hessian matrix. (See Appendix B for a derivation.)

⁴Note that the subscripts on the matrices in Eqn. 3.7 have been dropped in Eqn. 3.8. Hereafter, the shapes of individual matrices will be listed only when relevant to the point being made, and will default to their shapes as first defined.

3.2.1.3 Point Spread Function

The point spread function (PSF, \vec{I}^{psf}) is the impulse response function of the instrument. The PSF for a given direction on the sky is the image produced by a point source at that location. If the PSF is shift-invariant, it can be computed once, for a source at the phase center. Let us define \vec{I}^{psf} as the image produced by observing a point source of unit flux at the phase center. The PSF is the dirty image formed *via* the RHS of Eqn. 3.8 for a constant visibility function of unit amplitude (represented by a $n \times 1$ vector of ones) or the inverse Fourier transform of the gridded weights (weights accumulated onto an $m \times 1$ grid *via* the sampling matrix $[S]$).

$$\vec{I}_{m \times 1}^{psf} = [F^\dagger S^\dagger W] \vec{1}_{n \times 1} = [F^\dagger] \vec{W}_{m \times 1}^G \quad \text{where} \quad \vec{W}_{m \times 1}^G = [S^\dagger W] \vec{1}_{n \times 1} \quad (3.11)$$

\vec{W}^G is an $m \times 1$ vector containing a weighted average of the number of samples measured at each discrete spatial frequency. Since $[S]$ contains only ones and zeros, we can write $[W^G] = \text{diag}(\vec{W}^G) = [S^\dagger W S]$ as a diagonal matrix formed from the vector of gridded weights⁵. Note that \vec{I}^{psf} is the same as $I^{psf}(l, m)$ from Eqn. 2.14 but written with weights and in vector form.

Figure 3.1 shows a 1-D example of gridded weights and the PSF that is constructed from it. Note that if the sampling function were continuous (\vec{W}^G contains all ones and no zeros), the PSF would be a Kronecker δ -function.

The shape of the PSF is controlled by the uv -coverage $[S]$, and the visibility weights $[W]$. The minimum width of the main lobe of the PSF defines the angular resolution of the telescope and is controlled by the largest measured spatial frequency (given in units of radians as $\theta^{psf} = 1/u_{max}$ where u_{max} is the maximum baseline length in units of λ). The PSF has sidelobes (ripples with negative and positive amplitude) produced as a result of missing spatial frequencies. Also, an interferometer always has a central hole in its spatial-frequency coverage ranging from the origin of the uv -plane up to the shortest measured spatial frequency, and this gives a PSF with zero integrated area. The peak of the un-normalized PSF is given by the sum-of-weights $w_{sum} = \text{tr}[W^G]$ and represents the sensitivity of the instrument to a point source of unit amplitude.

3.2.1.4 Beam Matrix and Convolution

In this section, we show that the normal equations in Eqn. 3.8 describe the dirty image as a convolution of the sky brightness distribution with the PSF of the instrument.

Consider the Hessian matrix for standard imaging (Eqn. 3.9). By construction,

⁵Note that Eqn. 2.14 defines the PSF as the inverse Fourier transform of the uv sampling function without any measurement weights. Eqn. 3.11 is a discretized and practical version of this definition in which the samples are allowed to be weighted non-uniformly.

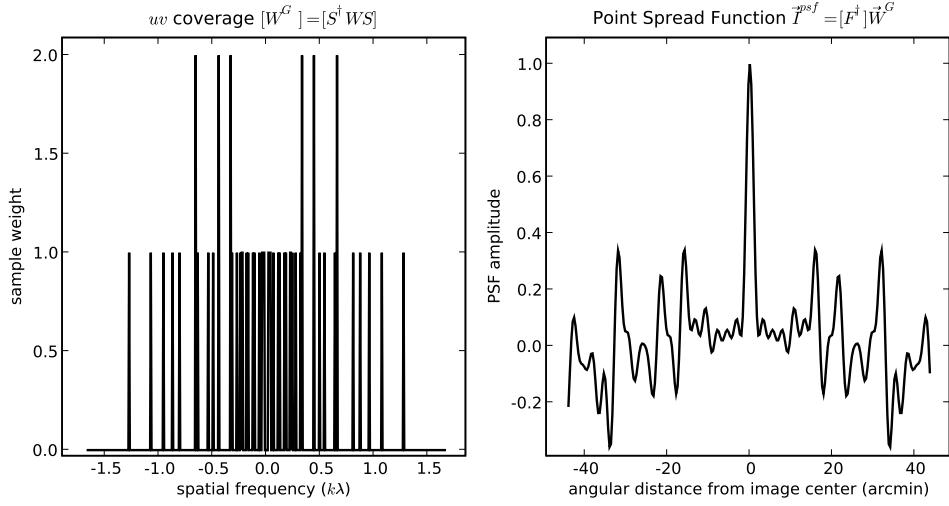


Figure 3.1: Sampling Weights and the Point Spread Function : This diagram shows a 1-dimensional example of the gridded weights and the point spread function that is constructed from it. The plot on the left shows the sample weights as a function of spatial frequency. The non-uniform amplitudes in this plot indicate a non-uniform sampling in which some of the measured spatial frequencies are sampled more than once. The plot on the right shows the point-spread-function (PSF) formed from the Fourier inverse of these gridded weights (Eqn. 3.11). The PSF has been normalized such that its peak value is unity. The width of the central lobe of the PSF defines the angular resolution of the interferometer. It is given by $\theta^{psf} = \frac{1}{u_{max}} \frac{180 \times 60}{\pi}$ arcmin, where u_{max} is the maximum spatial frequency in units of λ (in this example, $u_{max} = 1.3 k\lambda$ and $\theta^{psf} = 2.6'$ where ' denotes arc-minute). The lower-level structures seen on either side of the central peak are called sidelobes.

$[H] = [F^\dagger S^\dagger W S F]$ is a circulant convolution operator⁶ with \vec{I}^{psf} (given by $[F^\dagger] \vec{W}^G$ where \vec{W}^G is the diagonal of $[S^\dagger W S]$) as the convolution kernel⁷. This special form of $[H]$ in which each row contains a shifted version of the PSF (or instrument beam) is called the Beam matrix (denoted by $[B]$). The convolution equation of interferometric imaging is given as follows.

$$[B_{m \times m}] \vec{I}_{m \times 1}^{sky} = \vec{I}_{m \times 1}^{dirty} \quad \text{where} \quad [B] = [F^\dagger S^\dagger W S F] \quad (3.12)$$

⁶A circulant matrix is one that is diagonalized by the Fourier transform operator and its eigen-values are given by the Fourier transform of one of its rows. A convolution operator constructed as $[F^\dagger] diag([F]\vec{d})F$ (for \vec{d} as the convolution kernel) is a circulant matrix and $eig([C]) = diag([F]\vec{d})$ (see footnote 19 on page 24 for the definition of a convolution operator). For a two-dimensional convolution, $[F]$ is the outer product of two one-dimensional DFT operators, and $[C]$ is block-circulant with circulant blocks.

⁷In general, a matrix of the form $[F][diag(\vec{X})][F^\dagger]$ is a convolution operator with $[F^\dagger]\vec{X}$ as its kernel (the function that the operator applies the convolution with).

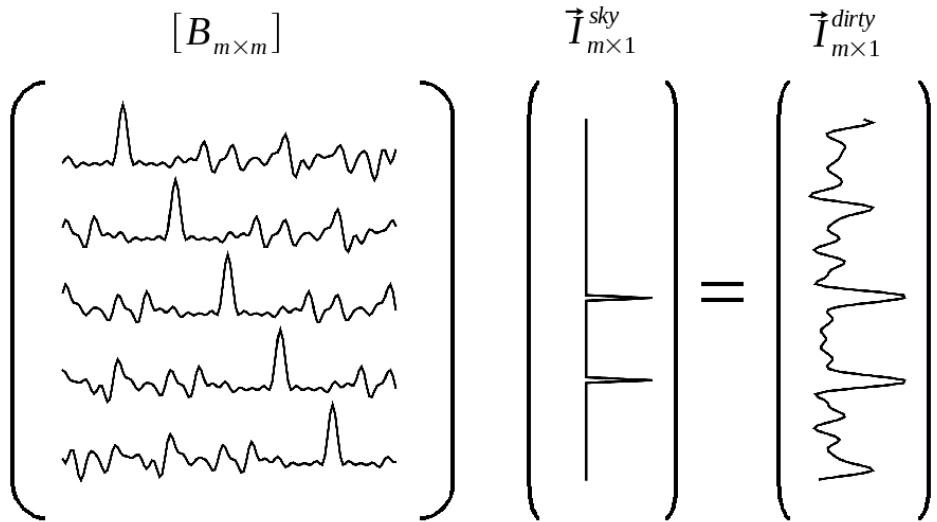


Figure 3.2: Normal Equations for Basic Imaging : This diagram represents the linear system of equations that describe the imaging process of an interferometer (Eqn. 3.12). The matrix on the left is the Beam matrix which consists of a shifted version of the PSF in each row (row i contains the PSF shifted to the location i on the sky). The column vector in the middle represents a one-dimensional empty sky with two point-sources and the vector on the right (\vec{I}^{dirty}) represents the dirty image. When the Beam matrix $[B]$ (on the left) is multiplied with the sky image \vec{I}^{sky} , it implements the shift-multiply-add sequence of a convolution. Therefore, this system of equations describes the dirty image as the result of a convolution of the sky with the PSF. This is the system of equations to be solved to reconstruct the image of the sky and the solution process represents a deconvolution of the PSF from the dirty image. The PSF used in this example is the same as that shown in Fig. 3.1.

Figure 3.2 is a pictorial representation of this convolution equation. The matrix on the left is the Beam matrix $[B]$, in which each row contains a shifted version of the PSF. The column vector in the middle represents a 1-D empty sky with two point-sources. When $[B]$ is multiplied by the sky image, it implements the shift-multiply-add sequence of a convolution, and the vector on the right represents the dirty image formed as a result of this convolution between the sky image and the PSF.

Eqn. 3.12 and Fig. 3.2 represent the system of equations that needs to be solved to obtain an estimate of the true sky brightness. This solution process is called a deconvolution, and the reconstructed estimate of \vec{I}^{sky} is called a model image (denoted as \vec{I}^{model}).

The diagram in Fig. 3.2 was constructed using 1-D (noise-free) numerical simulations of a simple sky brightness distribution and the PSF shown in Fig. 3.1 (with $m = 256$). The elements of the Hessian matrix were explicitly evaluated, and a matrix-vector product computed to obtain the RHS vectors. The diagram therefore represents a realistic result

and is not a toy illustration. The 1-D functions shown in the matrix on the LHS are from a selected subset of rows from the full matrix, chosen to illustrate the shape of the 1D functions in each row, and the locations of the peaks in each row correspond to the diagonal elements of the matrix. Several such diagrams are shown in later chapters of this dissertation, to illustrate the imaging equations in various situations (multi-scale, multi-frequency, and wide-field imaging). All these diagrams were produced using similar 1-D simulations that use the same sampling function and basic PSF as shown in Fig. 3.1.

3.2.1.5 Properties of the Hessian

A few properties of the Hessian are worth noting.

1. The elements on the diagonal of $[H]$ correspond to the peaks of the PSFs (given by the sum of weights) for each location in the image, and represent the sensitivity of the instrument to a point source of unit flux (in all directions). When $[H] = [B]$ the Hessian represents an imaging instrument in which the PSF is spatially invariant⁸ and all pixels in the weight image \vec{I}^{wt} are equal to w_{sum} .
2. A weight image \vec{I}^{wt} can be defined as an $m \times 1$ column vector constructed from these diagonal elements. When $[H] = [B]$, all elements (pixels) of the weight image contain the same number (w_{sum}). In the general case this is not true, and this weight image will be later used as a measure of the direction-dependent sensitivity of the instrument.
3. The eigen-values of $[H]$ are given by the diagonal matrix of gridded weights $[W^G] = [S^\dagger WS] = diag([F]I^{psf})$ (see Eqn. 3.11). When $[H] = [B]$, these are also the singular values⁹.

⁸ The rows of $[B]$ contain shifted versions of a single function, the PSF. This means that the instrument's impulse response function is identical for all directions on the sky. When direction-dependent instrumental effects are included in the measurement equations, the instrument's response changes with direction on the sky. The PSFs become spatially-variant, and the elements of \vec{I}^{wt} are different from each other and describe the direction-dependent sensitivity of the telescope.

⁹ The singular value decomposition of a matrix is given by $[A] = [U\Lambda_s V^\dagger]$ where $[U]$ and $[V]$ contain orthonormal columns and $[\Lambda_s]$ is a diagonal matrix of singular values. The eigen-value decomposition of a matrix is given by $[A] = [X\Lambda_e X^\dagger]$ where the columns of $[X]$ contain the eigen-vectors and $[\Lambda_e]$ is a diagonal matrix of eigen-values. When a matrix is Hermitian and symmetric, its singular values are related to its eigen-values as $[\Lambda_s] = abs([\Lambda_e])$. Therefore, for the Beam matrix that is by construction positive semi-definite, the eigen and singular-value decompositions are the same and $[U] = [V] = [X] = [F]$ and $[\Lambda_s] = [\Lambda_e] = [W^G]$. The singular value decomposition (SVD) of a matrix can be used to compute its pseudo inverse (an approximate inverse). The SVD is often used when the matrix to be inverted is rank-deficient. The SVD of the matrix can also be written as a sum of rank one matrices and their associated singular values $A = \sum_{i=0}^m U_i \lambda_i V_i^\dagger$. Its inverse is calculated by using only those singular values whose magnitude is larger than ϵ . Therefore, $A^+ = \sum_{\lambda_i > \epsilon}^m V_i \frac{1}{\lambda_i} U_i^\dagger$.

4. The diagonal elements of $[W^G]$ are positive for spatial-frequency grid cells that contain measurements, and zero for those that do not. Therefore, when the spatial-frequency plane sampling is incomplete, the inverse of $[W^G]$ and $[H]$ do not exist.

With this background, the next three sections will describe various ways of solving these normal equations to obtain an estimate of the sky brightness distribution.

3.2.1.6 Principal Solution

The principal solution (as defined in [Bracewell and Roberts \[1954\]](#) and used in [Cornwell et al. \[1999\]](#)) is a term specific to radio interferometry and represents the dirty image normalized by the sum of weights. It is the image formed purely from the measured data, with no contribution from the invisible distribution of images (unmeasured spatial frequencies). For isolated sources, the values measured at the peaks of the principal solution images are the true sky values as represented in the image model (in this case, a list of pixel amplitudes¹⁰). The principal solution is an approximate solution of the normal equations computed *via* a diagonal approximation of the Beam matrix $[B]$. In general, each diagonal element represents the sum of weights, w_{sum} and is equal to the value given by $mid\{\vec{I}^{psf}\}$ which for the PSF is also the location of its peak.

The advantage of using a diagonal approximation is that image pixels can be treated independently while computing the solution of the system. Further, for the Beam matrix (when $[H] = [B]$), all diagonal elements are equal and given by the peak value of the PSF. Therefore, the principal solution is computed by dividing all pixels in the dirty image by the peak of the PSF (whose value of w_{sum} can be picked from any diagonal element of $[B]$). To maintain consistency between definitions of the principal solution, and to introduce the notation that will be used in the later chapters, we will write the following equation to describe the operations that go into computing the principal solution one pixel at at time.

$$\vec{I}_{1 \times 1}^{pix,psol} = [H_{1 \times 1}^{peak}]^{-1} \vec{I}_{1 \times 1}^{pix,dirty} \quad (3.13)$$

where $[H_{1 \times 1}^{peak}]$ is (in this simple case) a one-element matrix containing the peak of the PSF (a diagonal element of $[B]$ from some row i), $\vec{I}_{1 \times 1}^{pix,dirty}$ is the value of the corresponding (i^{th}) pixel from the dirty image, and $\vec{I}_{1 \times 1}^{pix,psol}$ is the value of the principal solution at that i^{th} pixel. Note that the element in $[H^{peak}]$ is the sum of weights for the i^{th} pixel and is the i^{th} element of the weight image \vec{I}^{wt} .

Such a normalization by the Hessian diagonal is a combination of the DFT normalization of $\frac{1}{m}$ and a scaling by the sum of weights $w_{sum} = trace(W)$ that creates a PSF

¹⁰ This definition of the principal solution can be naturally extended to situations in which the image model is something other than a list of pixel amplitudes representing the intensity of the sky brightness distribution in all directions. The principal solutions for multi-scale and multi-frequency deconvolution are examples of such situations and are explained in sections [6.1.2.3](#) and [6.2.2.3](#).

of unit peak. This means that a point source of flux 1.0 Jy will give a peak value of 1.0 in the normalized dirty image. The values of the peaks in the normalized dirty image can now be interpreted physically in true flux units of Jy/beam. For an instrument with complete and uniform sampling where $[S]$, $[W]$ and $[H]$ are scalar multiples of identity matrices, \vec{I}^{psf} is a δ -function, the Hessian is purely diagonal, and this normalization gives the final reconstructed image.

For standard imaging, the principal solution is trivial to compute as an image-domain normalization by the sum of weights. However, as will be shown in later chapters for multi-scale and multi-frequency deconvolution, the principal solution can in general involve more than just a normalization.

3.2.1.7 Linear Deconvolution

Consider a filled-aperture telescope where there is complete but non-uniform sampling of the uv -plane. Let the distribution of samples follow a Gaussian function with maximum sensitivity at the centre of the uv -plane. This causes a blurring effect in the image. (A multiplication of the visibility function by a Gaussian in the spatial frequency domain is a convolution of the image with another Gaussian, resulting in blurring.)

This system can be solved *via* a linear deconvolution. If all spatial frequencies are measured at least once, $[S]$ has full column rank m , and the diagonal matrix of gridded weights $[W^G] = [S^\dagger WS]$ is positive definite, and therefore invertible. Let $[W^f]$ be an estimate for $[W^{G^{-1}}]$ such that $[W^f W^G] \approx [\vec{1}]$ (an $m \times m$ identity matrix). The deconvolution operator $[F^\dagger W^f F]$ can be applied to Eqn 3.8 to give $[F^\dagger W^f F][F^\dagger W^G F]\vec{I}^{sky} = m^2\vec{I}^{sky}$ which can then be normalized to recover I^{sky} . Ideally, $[W^f] = [W^{G^{-1}}]$ computed directly from $[W^G]$ will exactly invert the Hessian. However in the presence of noise, a direct computation of $[W^{G^{-1}}]$ will give artificially high weights to low signal-to-noise measurements, and this can introduce artifacts into the estimate of I^{sky} . In practice, $[W^f]$ is a Wiener filter which, in addition to inverting $[W^G]$, attenuates measurements at different spatial frequencies depending on their signal-to-noise ratios.

3.2.1.8 Non-Linear Deconvolution

A general interferometer samples the spatial frequency plane incompletely, with the associated sampling matrix $S_{n \times m}$ having a column rank $< m$. The $m \times m$ Hessian therefore has rank $< m$, making it a singular matrix with no exact inverse. Therefore, even though the convolution process described by the normal equations is linear, these equations have multiple solutions, and cannot be solved by a linear deconvolution.

An intuitive explanation of this non-uniqueness is that the data provide no constraints on what the unmeasured visibilities should be. Any choice of values at the unmeasured spatial frequencies will be indistinguishable from any other. More formally, the

dirty image on the RHS of Eqn. 3.8 lies in the range space of the Hessian matrix H . A rank-deficient Hessian implies that there is an entire range of images formed from spatial frequencies that fall in the null space of H , that if added to the sky model image I^{model} , will make no difference to the RHS of the normal equations. The solution to this system of equations is therefore non-unique. The set of images formed from the null-space of $[H]$ (unmeasured spatial frequencies) is called the invisible distribution [Bracewell and Roberts 1954]. A common way of filling in these unmeasured spatial frequencies is to use *a-priori* information about the typical structure of the sky to estimate the shape of the visibility function in between the measured spatial frequencies. This *a-priori* information is applied *via* a solution process that forces the model visibility function to agree with the data at all measured spatial frequencies.

3.2.1.9 Iterative CLEAN Deconvolution

This section describes the general framework used in most image reconstruction algorithms in radio interferometry. The steps given below follow the steepest-descent algorithm for χ^2 -minimization (described for radio interferometric imaging in Schwab and Cotton [1983]). All the algorithms in this dissertation are described within this framework.

In practice, the normal equations are solved *via* an iterative χ^2 -minimization process, not by explicitly evaluating the Hessian matrix and inverting it. This is because the Hessian matrix for interferometric imaging is usually singular with no exactly computable inverse, and is too large to handle numerically. Standard iterative deconvolution for interferometric imaging is based on a Newton-Raphson approach, and the following steps describe this process for radio interferometric image reconstruction. For an actual numerical implementation of these basic steps, several details need to be accounted for. Mainly, a preconditioning scheme is used to weight the visibility data (Section 3.2.3) while gridding them onto a regular grid of spatial frequencies (Section 3.2.2) and Fourier inverting to give the dirty image. Deconvolution is then a combination of successively building up an image of the sky by finding flux components and subtracting their effect from the dirty image (Section 3.2.4).

Pre-compute Hessian : Since the Hessian is a Toeplitz matrix (see footnote 19 on page 24) with a shifted PSF in each row, it suffices to compute and store only one instance of the PSF *via* Eqn. 3.21.

Initialization : Initialise the model image \vec{I}_0^m to zero or to a model that represents *a-priori* information about the true sky.

Major and minor cycles : There are two types of iterations, one nested within the other. The outer loop is called the major cycle and the inner loop is called the minor cycle. Steps

2 to **4** represent the minor cycle of iterations which operate in the image domain and search for flux components to form a model of the sky brightness. Steps **1** to **5** represent the major cycle in which the data and models are converted between the visibility and image domains so that χ^2 can be computed directly in the measurement domain.

1. **Compute RHS :** Compute an image from a set of visibilities. For the first iteration, this is the dirty image formed from the measured visibilities \vec{V}^{corr} . For subsequent iterations, it is called a residual image and is formed from the residual visibilities computed as $\vec{V}^{res} = \vec{V}^{corr} - \vec{V}^{model}$ where \vec{V}^{model} is the current best estimate of the true visibilities. In the first iteration, $\vec{V}^{model} = \vec{0}$, $\vec{V}^{res} = \vec{V}^{corr}$ and $\vec{I}^{res} = \vec{I}^{dirty}$. The residual image is normalized by the sum of weights.

$$\vec{I}^{res} = [F^\dagger S^\dagger W][\vec{V}^{res}] \quad (3.14)$$

This step is called the reverse transform¹¹.

2. **Find a Flux Component :** For iteration i , compute the update step by applying an operator T to the $\nabla\chi^2$ image.

$$\vec{I}_{(i)}^{model} = T(\vec{I}^{res}, \vec{I}^{psf}) \quad (3.15)$$

T represents a non-linear deconvolution of the PSF from I^{res} while filling-in unmeasured spatial frequencies (null space of the measurement matrix) to reconstruct an image of the sky brightness. This estimate of the sky brightness is called the model image \vec{I}^{model} . Section 3.2.4 describes T for several standard deconvolution algorithms¹².

3. **Update model :** Accumulate flux components from iteration i onto a model image.

$$\vec{I}^{model} = \vec{I}^{model} + g\vec{I}_{(i)}^{model} \quad (3.16)$$

g is called a loop-gain, takes on values between 0 and 1, and determines the step size for each iteration in the χ^2 minimization process.

4. **Update RHS :** The residual image is updated by subtracting out the contribution of the flux components found in iteration i , damped by the loop-gain.

$$\vec{I}^{res} = \vec{I}^{res} - g(\vec{I}^{psf} \star \vec{I}_{(i)}^{model}) \quad (3.17)$$

Repeat from Step 2 until some termination criterion is satisfied (usually, when T can no longer reliably extract any flux from I^{res}).

¹¹ When combined with the forward transform defined in step **5**, this residual image is equivalent to computing $\nabla\chi^2$ (see Appendix B for a derivation of an iterative Newton Raphson method).

¹² Following the standard calculation for the update step in a χ^2 minimization, $T(\vec{I}^{res}, \vec{I}^{psf}) = [F^\dagger S^\dagger W S F]^{-1} \vec{I}^{res}$. However, in our case since the Hessian is singular, this form of T is never explicitly computed.

5. **Predict :** Visibilities that would be measured for the current sky model \vec{I}^{model} are computed so that the model can be compared with the data \vec{I}^{corr} and new residual visibilities computed.

$$\vec{V}^{model} = [S F] \vec{I}^{model} \quad (3.18)$$

This is called the forward transform.

Repeat from Step 1 until convergence is achieved (usually, when \vec{V}^{res} and \vec{I}^{res} are noise-like).

Restoration : The final I^{model} is restored by first smoothing it to the maximum angular resolution of the instrument. This is done by convolving the final model image by a restoring beam \vec{I}^{beam} (a Gaussian whose width is chosen as the width of the central lobe of the PSF). This suppresses artifacts arising from unconstrained spatial frequencies beyond the measured range. Then, the final residual image I^{res} is added to the smoothed model image to account for any undeconvolved flux.

3.2.2 Gridding

The measured visibilities irregularly sample the continuous spatial frequency plane (for example, along elliptical tracks), and need to be binned onto a regular grid of spatial frequencies so that the FFT algorithm can be used for Fourier inversion. In section 2.2.2, a limiting spatial frequency grid was defined where the uv -pixel size is derived from the time and frequency resolution of the correlations, such that the visibility measurements naturally map to pixels on this grid. This spatial-frequency resolution corresponds to a very wide image field of view that is often impractical (due to very large image sizes) or unnecessary (due to a compact brightness distribution, or attenuation by antenna power patterns). To make an image over a more suitable (and smaller) field of view, the visibilities must map to uv -pixels on a coarser spatial-frequency grid.

Gridding can be described as an interpolation and resampling of the measurements taken on the fine spatial frequency grid, onto a coarser grid whose cell size is given by the smaller field of view over which an image is to be made. The sampling theorem states that if a function is band-limited, it can be completely represented by a set of samples spaced by the reciprocal of twice the bandwidth. In our case, the visibility function can be assumed to be band-limited because of the finite field of view within which the source of interest lies, and this defines a sampling interval on the spatial-frequency grid. Let m_I pixels on this coarse grid cover the same range of spatial frequencies as m did on the finer grid ($m_I < m$).

3.2.2.1 Visibility-domain convolution

Gridding is done as a convolutional resampling¹³, and can be described by the product of two operators. The first is $[G_{m \times m}]$, a convolution operator with a shifted version of the gridding-convolution function in each row. The second is a resampling matrix $[R_{m_I \times m}]$ with ones and zeros, whose columns define a Shah function that marks the grid onto which the function is resampled. Both operate on the fine grid, and $[R]$ reads off values at the locations of the coarse-grid cell centres.

$$\vec{V}_{m_I \times 1}^{gridded} = [R_{m_I \times m}] [G_{m \times m}] [S^{\dagger}]_{m \times n} [W_{n \times n}] \vec{V}_{n \times 1}^{obs} \quad (3.19)$$

where $[W]$ are visibility weights. $[S^{\dagger}]$ places the n visibility measurements onto the full-resolution spatial-frequency grid of size m . In practice, however, this full-resolution grid is not computed, and the result is directly evaluated on the coarse grid. In other words, the convolution and resampling are done as a single step for each visibility measurement.

A good choice for the gridding-convolution function is the prolate spheroidal function \vec{P}_s which has a small support-size (<10 grid cells) on the spatial-frequency plane and whose Fourier transform drops off rapidly beyond a certain distance from the center of the image (see [Briggs et al. \[1999\]](#)).

A visibility-domain convolution operator can be constructed with \vec{P}_s as the convolution kernel. This gridding-convolution operator in the spatial frequency domain is given by $G^{ps} = [FXF^{\dagger}]$ where $[X] = diag([F^{\dagger}] \vec{P}_s)$. This visibility-domain convolution is equivalent to multiplying the image domain by $\vec{I}^{ps} = [F^{\dagger}] \vec{P}_s$. Since \vec{P}_s has been introduced only for the purpose of interpolation, its effect needs to be removed from the image domain. To remove the multiplicative effect of the gridding-convolution function from the final $m_I \times 1$ image, a grid correction is done in the image domain using a truncated version of \vec{I}^{ps} , or by explicitly evaluating the Fourier transform of the prolate-spheroidal function used while gridding.

In Chapter 4, we will discuss the use of other convolution kernels that are used to implement direction-dependent corrections. In all cases the gridding-convolution operators are constructed as described above from the appropriate convolution kernels. The symbol $[G]$ is used for a generic convolution operator in the visibility domain.

¹³Convolutional resampling is a method that uses a convolution to interpolate between measurements to estimate the value of the observable at a set of locations that may be different from the actual measurements. For example, an interferometer measures the sky visibility function at an irregular set of points on the spatial frequency plane, but the use of the FFT algorithm for imaging requires that measurements lie on a regular grid. To achieve this, the measurements are first convolved with a smoothing kernel and the results are sampled at the new locations (regular grid).

3.2.2.2 Gridding and degridding

The following equations describe the reverse transform used in the deconvolution major cycle. The normalized dirty image and PSF are computed as follows by gridding a list of visibilities.

$$\vec{I}_{m \times 1}^{dirty} = w_{sum}^{-1} [I^{ps-1}] [F^\dagger R G^{ps} S^\dagger W] \vec{V}_{n \times 1}^{corr} \quad (3.20)$$

$$\vec{I}_{m \times 1}^{psf} = w_{sum}^{-1} [I^{ps-1}] [F^\dagger R G^{ps} S^\dagger W] \vec{I}_{n \times 1} \quad (3.21)$$

where the $[I^{ps-1}]$ is called the grid-correction step, and the division by $w_{sum} = \text{tr}[W]$ is a normalization that forces the peak of the PSF to represent one unit of flux. This normalized dirty image is also the principal solution of the normal equations (see section 3.2.1.6), where the normalization by w_{sum} represents the inversion of a diagonal approximation of the Hessian.

The model image \vec{I}_{model} obtained at the end of each minor cycle is used to predict visibilities that the interferometer would have measured for \vec{I}_{model} . The following equation describes the forward transform used in the deconvolution major cycle. The process of computing a list of visibilities from $[F]\vec{I}_{model}$ is called degridding. Here too, since \vec{P}_s is used only for interpolation on the uv -plane, its image-domain effect must be separately accounted for.

$$\vec{V}_{n \times 1}^{model} = [S G^{ps} R^\dagger F] [I^{ps}]^{-1} \vec{I}_{m \times 1}^{model} \quad (3.22)$$

where $[R^\dagger]$ maps the model visibility function from the coarse grid to the fine grid, before interpolating across the fine grid *via* a convolution, to evaluate the model visibilities at the sampled spatial frequencies.

The calculation of these transforms involves traversals of the entire set of visibility data and this is a computationally expensive operation. Therefore, deconvolution algorithms usually tailor the frequency of major and minor cycles to perform trade-offs between performance, accuracy and total number of iterations. Eqns. 3.20, 3.21 and 3.22 represent practical implementations of Eqns. 3.10, 3.11 and 3.18 respectively to compute the dirty image, point spread function and model visibilities. All references to \vec{I}^{dirty} , \vec{I}^{psf} and \vec{V}^{model} that appear in later chapters can refer to either form.

3.2.3 Preconditioning

In the solution of large systems of equations, preconditioning is a step that is designed to reduce the condition number¹⁴ of the system, making its solution more tractable

¹⁴ The condition number associated with a system of equations $[A]\vec{x} = \vec{b}$ gives a bound on how sensitive the solution vector \vec{x} will be to approximations in the solution process. A well-conditioned system has a low condition number, implying that the solution is equally sensitive to approximations involving any parameter. The condition number is given by the ratio of the largest to smallest eigen-values of matrix $[A]$. For a set

and stable. For the normal equations in Eqn. 3.8, the condition number depends on the distribution of gridded weights¹⁵, which controls the shape of the PSF. Traditionally, preconditioning has been done by computing and using imaging weights to alter the shape of the PSF according to whatever makes the normal equations easier to solve.

The imaging weights $[W_{n \times n}^{im}]$ are computed as the product of preconditioning weights $[W_{n \times n}^{pc}]$ and the measurement-noise-based weights $[W_{n \times n}]$, and are used instead of $[W]$ during gridding.

$$[W^{im}] = [W^{pc}][W] \quad (3.23)$$

The PSF that is formed using $[W] = [W^{im}]$ in Eqns. 3.11 and 3.21 and Figure 3.1 is the preconditioned PSF, and the Hessian (Beam matrix) becomes a convolution operator with this new PSF in each row. In practice, the preconditioning or weighting scheme is chosen to optimize either the shape of the PSF or the sensitivity of the instrument or some combination of both.

3.2.3.1 Types of Image Weighting

Several weighting schemes are described below (see [Briggs \[1995\]](#) for a complete and detailed description). For radio interferometric imaging, preconditioning weights are usually computed in gridded form as an $m \times 1$ list of uv -plane weights, and then de-gridded (or resampled *via* $[S_{n \times m}]$) to form an $n \times 1$ list of visibility weights ($\vec{W}_{n \times 1}^{pc} = [S_{n \times m}] \vec{W}_{m \times 1}^{pc,G}$). Eqn. 3.23 is then used with $[W_{n \times n}^{pc}] = diag(\vec{W}_{n \times 1}^{pc})$ to form imaging weights that are used during gridding. Note that this method often requires two passes through the data, one to construct the imaging weights, and one to apply them. Sometimes however, depending on numerical stability, the preconditioning weights can also be applied in gridded form. This method is advantageous in that it requires only one pass through the data.

Natural : The natural weighting scheme gives equal weight to all samples ($[W^{pc}] = [\vec{I}]$, an $n \times n$ identity matrix) and preserves the instrument's peak sensitivity, making it ideal for the detection of low signal-to-noise sources. However, since the gridded weights $[W^G] = [S^\dagger W S]$ are often proportional to sample density on the spatial frequency plane, an irregular sample density will give a high condition number and the PSF can have a wide main lobe and high sidelobes.

Uniform : Uniform weighting gives equal weight to each measured spatial frequency irrespective of sample density, thus theoretically reducing the condition number to 1. It is com-

of normal equations, the eigen-values of the Hessian are inversely related to the variance or uncertainty of each parameter in the system, and the condition number measures the maximum relative uncertainty between parameters.

¹⁵ In interferometric imaging, the presence of unmeasured spatial frequencies, will make some eigen-values exactly zero, giving a condition number of infinity. However, in practice, we can consider the condition number computed only from non-zero eigen-values, as the ratio between the largest and smallest non-zero gridded weights.

puted as a weight, based on sample density for each spatial frequency grid cell, and is equivalent to a pseudo-inverse of the gridded natural weights, computed *via* its singular-value decomposition (see footnote 9 on page 35). The $m \times 1$ list of gridded preconditioning weights ($\vec{W}^{pc,G}$) is constructed from the diagonal of $[W^G]$. The resulting PSF has a narrow main lobe and suppressed sidelobes across the entire image and is best suited for sources with high signal-to-noise ratios to minimize sidelobe contamination between sources. However, the peak sensitivity is much less than optimal, since data points in densely sampled regions have been weighted down to make the weights uniform. Also, isolated measurements can get artificially high relative weights and this may introduce further artifacts into the PSF. Note that in the case of complete sampling of the spatial frequency plane ($[S]$ is of full rank), uniform weighting implements a linear deconvolution (the uniformly-weighted PSF will be a δ -function, and the Hessian will be a trivially invertible diagonal matrix).

Super-Uniform : Super-uniform weighting is uniform weighting performed by counting visibility samples that fall not only within a single spatial frequency grid cell, but in a $N \times N$ block of cells around the grid cell of interest ($N = 3, 5, 7, etc$). This method tends to give higher weights to sparsely sampled regions of the uv -grid, as compared to densely sampled regions, and gives a PSF with inner sidelobes suppressed as in uniform weighting but far-out sidelobes closer to that with natural weights. The peak sensitivity is also closer to natural weighting.

uv -Taper : uv -tapering applies a multiplicative Gaussian taper \vec{T}_{uv} to the spatial frequency grid, to weight down high spatial-frequency measurements relative to the rest. This suppresses artifacts arising from poorly sampled regions near and beyond the maximum spatial frequency. This is important for deconvolution because visibilities estimated for these regions would have poor or no constraints from the data. Also, the natural PSF is smoothed by $[F^\dagger]\vec{T}_{uv}$ and this tunes the sensitivity of the instrument to scale sizes larger than the angular-resolution of the instrument by increasing the width of the main lobe. uv -tapering is usually applied in combination with one of the above weighting methods and is applied after gridding.

Briggs/Robust : Briggs or Robust weighting [Briggs 1995] creates a PSF that smoothly varies between natural and uniform weighting based on the signal-to-noise ratio of the measurements and a tunable parameter that defines a noise threshold. High signal-to-noise samples are weighted by sample density to optimize for PSF shape, and low signal-to-noise data are naturally weighted to optimize for sensitivity. The weights are derived by minimizing the $L2$ norm over an image of a point source of flux s as seen by the instrument. The weights are computed for all uv grid cells as

$$\vec{W}^{pc,G} = \frac{const}{s^2 \vec{W}^G + \Delta s^2} \quad (3.24)$$

where Δs^2 is a noise threshold and s is a tunable parameter. A large value of s makes the weighting nearly uniform, and a value of zero makes it tend to natural weighting. The

condition number of the system is reduced (compared to natural weighting), but higher than 1.0 because it weights low signal-to-noise measurements naturally.

Wiener Filter : Wiener filtering is a robust linear deconvolution that weights spatial frequencies with high signal-to-noise data uniformly, and suppresses spatial frequencies with low signal-to-noise ratios. The Wiener filter is constructed (pixel-by-pixel) as

$$\vec{W}^{pc,G} = \frac{\vec{W}^G^\dagger}{\left[\vec{W}^G^\dagger \cdot \vec{W}^G + \frac{\sigma^2}{\vec{P}_{signal}} \right]} \quad (3.25)$$

where σ^2 is the variance of the measurement noise and \vec{P}_{signal} is the expected power spectrum of the signal (square of the visibility function of the expected sky brightness distribution). The ratio $\sigma^2/\vec{P}_{signal} = 1/SNR$ is the inverse signal-to-noise ratio of the instrument across the visibility domain and can be a tunable parameter to bias the weighting towards uniform or natural weighting. The result is a nearly uniformly-weighted PSF that correctly represents the sensitivity of the instrument at various spatial frequencies¹⁶. A *uv*-taper can be built into the Wiener filter by setting $\vec{P}_{signal} = \vec{T}_{uv}^2 \equiv ([F]\vec{I}^{beam})^2$ to be a taper function that defines the angular scale to which the instruments sensitivity is tuned (\vec{I}^{beam} is a Gaussian of the chosen angular scale.). For a filled-aperture instrument, this form of preconditioning is a robust linear deconvolution that directly delivers the solution of the system.

3.2.4 Deconvolution

This section describes the minor cycle of iterative image reconstruction. For the minor cycle, \vec{I}^{dirty} is assumed to be a perfect convolution of the PSF with the true sky brightness, where \vec{I}^{dirty} , \vec{I}^{psf} are given by Eqns. 3.20 and 3.21. The operator T in Eqn. 3.15 constructs a model image \vec{I}^{model} via a deconvolution.

The CLEAN algorithm forms the basis for most deconvolution algorithms used in radio interferometry. The peak of the residual image gives the location and strength of a potential point source. The effect of the PSF is removed by subtracting a scaled \vec{I}^{psf} from \vec{I}^{res} at the location of each point source and updating \vec{I}^{model} (Eqn. 3.16). Many such iterations of finding peaks and subtracting PSFs form the minor cycle. Algorithm 1 lists pseudo-code for the basic CLEAN deconvolution algorithm that models the sky in a pixel basis. Basic CLEAN is best suited to isolated point sources whose amplitude is constant across the observing bandwidth. Deconvolution algorithms that produce multi-scale and multi-frequency source models are described in Chapter 6.

¹⁶ The use of a Wiener filter for post-gridding preconditioning was developed by T.J.Cornwell et al in 2008, for use with the ASKAP telescope. This work is unpublished.

3.2.4.1 Variants of CLEAN

Högbom : In Högbom CLEAN [Högbom 1974], the minor cycle subtracts a scaled and shifted version of the full PSF to update the residual image for each point source. After the initial dirty image is computed, only minor cycle iterations are done, making this a purely image-domain algorithm. It is computationally efficient but susceptible to errors due to inappropriate preconditioning that will not be corrected during the major cycle.

Clark : Clark CLEAN [Clark 1980] does a set of Högbom minor cycle iterations using a small patch of the PSF. This is an approximation that will introduce errors in the image model, but the minor cycle iterations are stopped when the brightest peak in the residual image is below the first sidelobe level of the brightest source in \vec{I}^{res} . The residual image is then re-computed as $\vec{I}^{\text{res}} = [F^\dagger]([F]\vec{I}^{\text{dirty}} - [F]\vec{I}^m)$ to eliminate aliasing errors.

Cotton-Schwab : Cotton-Schwab CLEAN [Schwab and Cotton 1983] is similar to the Clark algorithm, but computes the residual *via* a full major cycle as $\vec{I}^{\text{res}} = [F^\dagger S^\dagger W](\vec{V}^{\text{corr}} - [S F]\vec{I}^m)$. It is time consuming but relatively unaffected by preconditioning and gridding errors because it computes χ^2 directly in the measurement domain. It also allows highly accurate prediction of visibilities without pixelation errors.

Steer-Dewdney-Ito : The Steer-Dewdney-Ito CLEAN [Steer et al. 1984] minor cycle finds the locations of sources by setting an amplitude threshold to select pixels. The combined set of pixels is then convolved with the PSF and subtracted out *via* a Clark major cycle. This algorithm is more suited to deconvolving extended emission.

Multi-Resolution : Multi-Resolution CLEAN [Wakker and Schwarz 1988] performs a series of Högbom minor cycles on a smoothed version of the dirty image as well as on a difference image, to reconstruct large-scale structure that is inadequately sampled at the low spatial frequencies while retaining high resolution structure.

Multi-Scale : Cornwell-Holdaway Multi-Scale CLEAN (CH-MSCLEAN) [Cornwell 2008] is a scale-sensitive deconvolution algorithm designed for images with complicated spatial structure. It parameterizes the image into a collection of inverted tapered paraboloids. The minor cycle iterations use a matched-filtering technique to measure the location, amplitude and scale of the dominant flux component in each iteration, and take into account the non-orthogonality of the scale basis functions while performing updates. Section 6.1 contains a detailed description of a modified form of this algorithm (MS-CLEAN). A related method described in Greisen et al. [2009] uses *uv*-taper functions to create images at different spatial resolutions, and uses heuristics to choose a spatial scale at which to perform a set of Högbom minor cycle iterations.

Multi-Frequency : The Sault-Wieringa Multi-Frequency CLEAN (SW-MFCLEAN) [Sault and Wieringa 1994] is a wide-band deconvolution algorithm that models the sky brightness distribution as a collection of point sources with power-law spectra. The algorithm uses a matched-filtering technique based on spectral PSFs that describe the instrument's responses to point sources with spectra given by the first two terms in a Taylor series expansion.

Section 6.2 contains a detailed description of a generalized form of this algorithm (MF-CLEAN). Conway et al. [1990] and Likhachev [2005] describe similar multi-frequency imaging techniques.

3.2.4.2 Constrained Optimization Solvers

Maximum Entropy : The Maximum Entropy method (MEM) [Cornwell and Evans 1985; Narayan and Nityananda 1986], is a pixel-based deconvolution algorithm that performs a rigorous constrained optimization in a basis of pixel amplitudes. MEM uses the Bayesian formulation of χ^2 minimization, and applies a penalty function based on relative image entropy $-\vec{I}^m \ln(\vec{I}^m / \vec{I}^{prior})$. This choice of penalty function biases the estimate of the true sky brightness towards a known prior image \vec{I}^{prior} . If a flat image is chosen as the prior, the solution is biased towards being smooth, and produces a more realistic reconstruction of extended emission. A positivity constraint on the image pixels can be applied *via* a penalty function given by $\ln(\vec{I}^m)$, and an emptiness constraint can be applied *via* one given by $-\ln[\cosh(\vec{I}^m - \vec{I}^{prior})/\sigma]$ (σ is a noise threshold).

Non negative least-squares : The non-negative least-squares (NNLS) [Briggs 1995; Lawson and Hanson 1974] algorithm is another pixel-based method that solves a least-squares problem with linear-inequality range constraints for all its parameters. The main constraint that is applied is the positivity of all pixels in the model. This algorithm was shown to be well suited to moderately-resolved sources.

Adaptive Scale Pixel : The Adaptive Scale Pixel (ASP) [Bhatnagar and Cornwell 2004] deconvolution algorithm parameterizes the sky brightness distribution into a collection of Gaussians and does a formal, constrained optimization on their parameters. In the major cycle, visibilities are predicted analytically with high accuracy. In the minor cycle, the location of a flux component is chosen from the peak residual, and the parameters of the largest Gaussian that fits the image at that location are found.

Other Methods : Other methods include the Richardson-Lucy algorithm that computes a maximum-likelihood solution if the image noise follows Poisson statistics (appropriate mainly for filled aperture instruments), the Gerchberg-Saxton-Papoulis algorithm that iterates between the image and spatial frequency domains and applies support constraints in both domains, and the Singular Value Decomposition method [Briggs 1995] that combines a standard SVD inversion (uniform weighting) along with support constraints in the image domain. Other multi-scale methods (also listed in Cornwell [2008]) include Multi-Scale Maximum Entropy which performs MEM simultaneously on a set of images at different resolutions, and the use of Wavelets, Shapelets and Pixons to decompose the image in a suitable basis. Smear-Fitting [Reid 2003] models the source with a set of basis functions and then convolves each component with an elliptical Gaussian to account for the uncertainty in its shape and location. There are several Monte-Carlo based image reconstruction algorithms among which MC-Fit Rau and Cornwell [2005] models an image as

a set of overlapping 2D Gaussians and uses a Monte-Carlo method to derive the optimal set of flux components along with error bars for the best-fit parameters and [Sutton and Wandelt \[2006\]](#) uses customized sampling strategies to derive maximum-likelihood estimates. Spatio-Spectral MEM [[Bong et al. 2006](#)] is an entropy based method for wide-band imaging, where a smoothness constraint is applied across frequency. More recently, Compressed-Sensing techniques [[Wiaux et al. 2009](#)] have begun to be used for image reconstruction in radio interferometry.

3.2.4.3 Full-Stokes Imaging

The Stokes vector for polarised sky brightness $\vec{I}^{Stokes} = \{I, Q, U, V\}$ is related to the coherence vector of images $\vec{I}^{Corrs} = \{XX, XY, YX, YY\}$ corresponding to all four correlation pairs, *via* a linear transform (given by a 4×4 operator per image pixel [[Hamaker et al. 1996](#)]).

$$\vec{I}_{4m \times 1}^{Corrs} = [\mathcal{S}_{4m \times 4m}] \vec{I}_{4m \times 1}^{Stokes} \quad \text{where} \quad [\mathcal{S}_{4 \times 4}] = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & i \\ 0 & 0 & 1 & -i \\ 1 & -1 & 0 & 0 \end{pmatrix} \quad (3.26)$$

A full-Stokes deconvolution differs from standard methods in the computation of dirty images and the minor cycle. The Stokes vector of dirty images $\vec{I}^{dirty, Stokes}$ is computed by applying Eqn. 3.26 to the set of dirty images in the correlation basis $\vec{I}^{dirty, Corrs}$ given by Eqn. 3.20. The different Stokes parameters are considered to be linearly independent and deconvolution minor cycles are performed separately on each Stokes image. For compact sources, position constraints are sometimes applied across Stokes parameters based on the locations of peak residuals of the Stokes I image. [Holdaway and Wardle \[1990\]](#) describes an algorithm that applies the constraint of $I^2 \geq Q^2 + U^2 + V^2$ during deconvolution. [Sault et al. \[1999\]](#) describes another method where a steepest descent χ^2 minimization leads to the criteria of searching for peaks in $I^2 + Q^2 + U^2 + V^2$.

Algorithm 1: CLEAN with Cotton-Schwab major and minor cycles

Data: Calibrated visibilities : $\vec{V}_{n \times 1}^{corr}$
Data: *uv*-sampling function : $S_{n \times m}$
Data: Image noise threshold and loop gain σ_{thr}, g_s
Result: Model Image : $\vec{I}_{m \times 1}^m$

- 1 Compute the dirty image and the PSF : $\vec{I}^{dirty}, \vec{I}^{psf}$
- 2 Measure the PSF peak sidelobes : $f_{sidelobe} = max\{sidelobe(\vec{I}^{psf})\}$
- 3 Compute the restoring beam : $\vec{I}^{bm} =$ Gaussian fit to the main lobe of \vec{I}^{psf}
- 4 Initialize the model and residual images : $\vec{I}^m = 0; \vec{I}^{res} = \vec{I}^{dirty}$
- 5 **repeat** /* major cycle */
- 6 | Compute a flux limit : $f_{limit} = max\{\sigma_{thr}, f_{sidelobe} * max\{\vec{I}^{res}\}\}$
- 7 | **repeat** /* minor cycle */
- 8 | | Find the location and amplitude of the peak : $\vec{I}_i^m = peak(\vec{I}^{res})$
- 9 | | Update the model image : $\vec{I}^m = \vec{I}^m + g_s \vec{I}_i^m$
- 10 | | Update the residual image : $\vec{I}^{res} = \vec{I}^{res} - g_s [\vec{I}_i^m \star \vec{I}^{psf}]$
- 11 | **until** $Peak residual max\{\vec{I}^{res}\} < f_{limit}$
- 12 | Use current model image \vec{I}^m to predict model visibilities \vec{V}^m
- 13 | Compute a new residual image \vec{I}^{res} from $\vec{V}^{corr} - \vec{V}^m$
- 14 **until** $Peak residual max\{\vec{I}^{res}\} < \sigma_{thr}$
- 15 Restore the final model image : $\vec{I}^{restored} = \vec{I}^m \star \vec{I}^{bm} + \vec{I}^{res}$

CHAPTER 4

IMAGING WITH DIRECTION-DEPENDENT EFFECTS

This chapter introduces the concept of direction-dependent instrumental effects, discusses how they affect the measurement and imaging process, and describes ways to deal with them during image reconstruction. Direction dependent effects are those that cannot be described by a single antenna-based complex number and calibrated out *via* standard methods. They have to be dealt with during the imaging process. When these effects are included in the measurement equation, it results in a point spread function that varies with position and time, making the imaging equation no longer a pure convolution. Algorithms that correct for these effects usually use *a-priori* instrumental models, and iterate between an approximate minor cycle that assumes a position and time invariant point spread function, and an accurate major cycle that calculates model visibilities by predicting and applying antenna-based direction-dependent effects.

This chapter deals with the correction of direction-dependent effects during imaging, assuming that they are known *a-priori* (*via* measurements or models), and follows the discussions in [Bhatnagar et al. \[2008\]](#); [Cornwell et al. \[2008\]](#); [Uson and Cotton \[2008\]](#). This chapter does not discuss the processes involved in measuring or solving for direction-dependent antenna gains. Some methods that solve for these effects include the Peeling technique of calibration and imaging [[Nijboer and Noordam 2007](#)], the use of eigen-beams, an orthogonal set of basis functions to model the primary beam [[Voronkov and Cornwell 2007](#)], a self-calibration method for estimating and correcting antenna pointing offsets [[Bhatnagar et al. 2004](#)], and solving for parameters of a phase screen model for propagation effects [[Intema et al. 2009](#); [Yatawatta et al. 2008](#); [Cotton and Uson 2006](#)].

Section 4.1 briefly lists the different types of direction-dependent effects. Section 4.2.1 uses a simple example of full primary beam imaging with an array of identical time-invariant antennas, to introduce the theory of including direction-dependent effects in the measurement equations and algorithms to solve it. Section 4.2.2 generalizes this theory to include other direction-dependent instrumental effects, along with variability with time and baseline. This entire chapter assumes that the bandwidth being used is narrow enough that a single frequency analysis will suffice. Chapter 7.2 will later describe extensions that apply to wide-band receivers and multi-frequency-synthesis imaging.

4.1 Types of Direction Dependent effects

4.1.1 Antenna Primary Beam

Aperture Illumination Function : A commonly used array element for radio interferometry is a parabolic reflecting dish that collects the radiation incident on a large aperture and focuses it onto a feed. The E-field incident at each point on this aperture is modified by the radiation pattern of the feed as well as structures on the dish surface that make it deviate from an ideal parabola. Jones matrices (see Eqn. 2.15) describe this effect for each point on the aperture. The collection of Jones matrices for all locations across the aperture is called the aperture illumination function. Each antenna measures the integrated product of the E-field incident on its aperture and the corresponding aperture illumination function.

Voltage Pattern : Consider the aperture illumination function of an antenna for only one polarization component (say, only the first diagonal element of the Jones matrix). This function is related to the concept of a complex antenna gain (g_i) introduced in section 3.1.1. A single complex gain factor per antenna represents a constant multiplicative gain across the entire image field of view. Since there is a Fourier relation between the image plane and the aperture plane, such a constant image-domain gain maps to one aperture point. However, a real antenna and reflecting dish will cover a whole range of aperture points (a finite-sized aperture), and the Fourier transform of this aperture illumination function gives a direction-dependent complex gain per antenna, called a voltage pattern.

Primary Beam : Given a pair of finite antenna apertures, we can construct a set of baseline vectors defined between all possible points on the two apertures (and not just the baseline vector defined by the locations of the feeds). Each baseline of the interferometer is therefore sensitive to a range of spatial frequencies around the nominal value. The visibility measured by each antenna pair can be described as the result of a convolution of the source visibility function with a baseline aperture function evaluated at the nominal spatial frequency (see Eqn. 2.19). This baseline aperture function is the result of the convolution of the aperture illumination functions of both antennas and its Fourier transform is called the primary beam. A visibility-domain convolution with this aperture function is equivalent to a multiplication of the sky brightness distribution by this primary beam.

Varying primary beams : Various related instrumental effects can be modeled *via* the aperture illumination functions. An antenna pointing offset can be modeled by a phase gradient applied to the aperture illumination function of the antenna. For wide-field imaging *via* mosaicing, a collection of pointings can be described *via* phase gradients across the spatial frequency plane, and applied *via* the aperture illumination functions. If the antennas are on azimuth-elevation mounts, as they track a celestial source, the primary beams rotate on the sky and this effect can be modeled by a rotation of the aperture illumination functions. Aperture illumination functions are usually different for each set of feeds in the array (either due to different imperfections in the antenna structures, or the location of feeds on the focal-plane), and vary with time and frequency.

4.1.2 Non-Instrumental Effects

Refraction through the atmosphere : The incident radiation can be corrupted in various ways before it reaches the antennas and detectors. Refraction through various layers in the Earth's atmosphere can result in phase distortions that cause direction-dependent shifts and defocus in the image. This effect is usually described *via* a turbulent phase screen that distorts the radiation wavefront before it reaches the array elements. It can also be described as a direction-dependent complex gain that can vary across the field of view of each antenna. If these gains vary across the field of view, but are identical for all antennas in the array, there will be no effect on the image. If these gains are constant across the antenna fields of view, but different for each antenna, the image will be distorted but standard direction-independent calibration methods will remove this effect. If these gains vary across the field of view and are different for each antenna, it is a direction-dependent effect that can be accounted for only during imaging by an appropriate choice of aperture illumination function. These effects generally vary with time, frequency and polarisation.

The w -term : The w -term (described in section 2.1.1, Eqn. 2.8) results in another direction-dependent effect. The curvature of the sky causes the aperture plane of the interferometer to change with position on the sky, thus changing the effective uv -coverage and making the PSF vary across the region of sky being imaged. The magnitude of the w -term also varies with time as the phase reference center is being tracked, resulting in an apparent shift in the source position. If an image is created by approximating the curved sky with a single tangent plane, a source located far away from the phase reference center will be smeared out due to this time-dependent effect. In this situation, a single time and position invariant PSF cannot be used to deconvolve the entire image.

4.2 Correction of direction-dependent effects

This section describes how various direction-dependent effects are modeled and corrected during the imaging process. The measurement equations addressed here are simplified forms of Eqns. 2.22 and 2.23 containing only direction-dependent effects and ignoring $[K_{n \times n}^{vis}]$ (assuming all direction-independent effects have been calibrated out).

Section 4.2.1 discusses the solution of the system of equations shown in Eqn. 2.23 where a single image-domain function can be used to describe all direction-dependent effects (all antennas are identical and time-invariant). In this case, a common primary beam can be factored out of the deconvolved image and removed in a single image-domain correction step (from regions where it is above some amplitude threshold). Section 4.2.2 discusses the more general case where direction-dependent effects vary with time and are baseline dependent. These effects are modeled as amplitude and phase effects in the visibility domain and corrected using complex conjugates of these models. Some effects (*e.g.* pointing-offsets) can be described by functions of unit amplitude and a non-zero phase, and the use of the complex conjugate eliminates this effect completely. Other effects (*e.g.*

aperture illumination function and primary beam) are modeled by complex functions with non-unit amplitude and non-zero phase, and the use of the complex conjugate eliminates only phase effects, but introduces a second factor of amplitude, which has to be accounted for in the image-domain (for reasons related to numerical stability). For primary beam effects, the dirty image produced by this method contains a factor of primary beam squared. Section 4.3 describes the recently developed AW-projection [Bhatnagar et al. 2008] algorithm that solves the measurement equations shown in Eqn. 2.22 via a combination of image and visibility-domain operations during iterative deconvolution. This approach is illustrated using the example of primary beam correction in order to compare this method with the simple image-domain correction described in section 4.2.1.

4.2.1 Image-domain corrections

This section describes the imaging equations and their solutions for the special case where antenna primary beams are the only source of direction-dependent effects and they can be assumed to be (1) the same for all antennas, baselines and frequency channels and (2) constant in time (not rotating). In such a situation, the effect of a common antenna primary beam can be factored out of the imaging equations and a simple adaptation of standard imaging algorithms will suffice to solve the resulting system of equations. For the more general case (section 4.2.2) where aperture illumination functions are different for each antenna and vary with time, a common primary beam cannot be factored out of the imaging equations. However, the approach described in this section can always be used as an approximation.

4.2.1.1 Image Model

When the primary beams of all antennas are identical (or are assumed to be so), the brightness distribution measured by each antenna can be written as the product of the true sky brightness distribution and the primary beam. Let \vec{P}_b be an $m \times 1$ vector that represents the antenna primary beam. The observed sky brightness distribution can be modeled as follows.

$$\vec{I}^{model} = [P_b] \vec{I}^{sky} \quad (4.1)$$

4.2.1.2 Measurement Equations

The interferometer samples the product of the primary beam and the sky. The measurement equations for this system are given below. These equations are obtained by setting $D^{sky} = P_b$ in Eqn. 2.23.

$$[S][F] \left\{ [P_b] \vec{I}^{sky} \right\} = \vec{V}^{corr} \quad (4.2)$$

4.2.1.3 Normal Equations

The normal equations constructed for the above measurement equation are shown below (similar to Eqn. 3.8).

$$[F^\dagger S^\dagger W^{im} S F][P_b] \vec{I}^{sky} = [F^\dagger S^\dagger W^{im}] \vec{V}^{corr} \quad (4.3)$$

$$\text{or} \quad [B][P_b] \vec{I}^{sky} = \vec{I}^{dirty,pb} \quad (4.4)$$

$[B]$ is the Beam matrix consisting of shifted versions of \vec{I}^{psf} in each row and $\vec{I}^{dirty,pb}$ is the result of the convolution of $\vec{P}_b \cdot \vec{I}^{sky}$ with \vec{I}^{psf} (note that $\vec{I}^{dirty,pb}$, \vec{I}^{psf} are the dirty image and point spread function computed via Eqns. 3.20 and 3.21 and $[P_b] = \text{diag}(\vec{P}_b)$). The diagonal elements of $[B][P_b]$ form the weight image \vec{I}^{wt} (introduced for the standard imaging case at the end of section 3.2.1.4). For the case of wide-field imaging using standard imaging algorithms and identical primary beams the weight image is given by $\vec{I}^{wt} = w_{sum} \vec{P}_b$ and represents the relative sensitivity of the instrument at each location on the sky.

Figure 4.1 represents the normal equations given in Eqn. 4.4, for a sky image containing two point sources of equal amplitude. The matrix on the left of Figure 4.1 is the beam matrix $[B]$. The second matrix from the left represents the primary beam $[P_b]$. This system of equations can be viewed as the standard Beam matrix $[B]$ operating on the product of the sky and the primary beam. The effect of such a measurement on the RHS is that the off-center source and its response are attenuated by the value of the primary beam at the location of the source. Note that this is not the same as multiplying the dirty image (shown in Fig. 3.2) by the primary beam. A simple solution of these equations is to combine the primary beam with the image model and use standard imaging methods along with a post-deconvolution division of the model by the primary beam. Note however, that sources that fall within the null of the primary beam cannot be recovered.

4.2.1.4 Post-deconvolution primary beam correction

Standard deconvolution algorithms (described in Chapter 3) can be used to reconstruct a model image that represents the product of the sky and the primary beam ($\vec{I}^{model} = [P_b] \vec{I}^{sky}$). A post-deconvolution division of the model image by the primary beam (known *a-priori*) will give an image of the reconstructed sky brightness distribution.

$$\vec{I}^{model,corrected} = \vec{I}^{model} / \vec{P}_b \quad (4.5)$$

Note, however, that the field of view over which this correction is possible is limited to high signal-to-noise regions of the image. Such regions are usually determined *via* a cut-off threshold at a level of few percent of the primary beam peak.

Note that even when antenna aperture illumination functions are not identical and time-invariant, this approach can still be used as an approximation in the minor cycle and section 4.3 discusses how this is done.

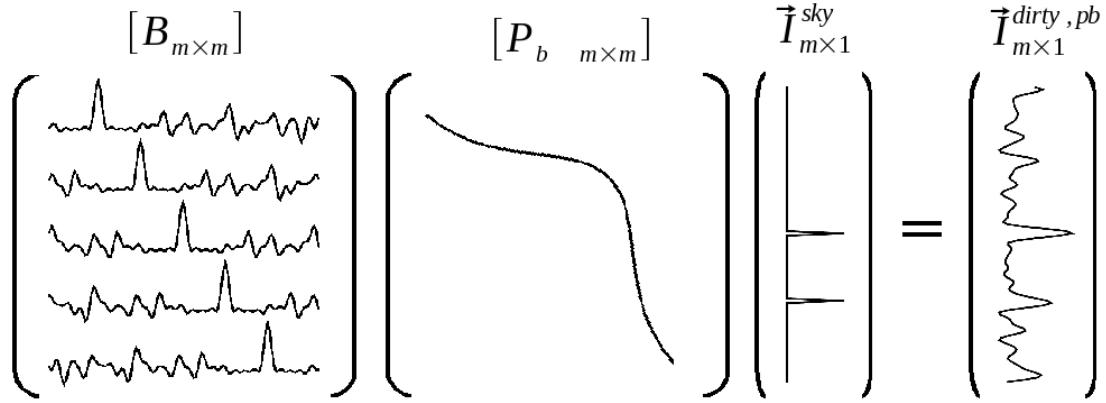


Figure 4.1: Normal equations with an image-domain primary beam: This diagram represents the imaging process of a radio interferometer when all antennas in the array have identical, time-invariant primary beams (Eqn. 4.4). In this diagram, $[B]$ is the Beam matrix, and $[P_b]$ represents a diagonal matrix with a Gaussian primary beam filling its diagonal elements. This system of equations can be viewed as the PSF being convolved with the product of the primary beam and the sky (here, with two point sources). The effect of the primary beam on the dirty image (RHS) is that the off-center source and its response are attenuated by the value of the primary beam at the location of the source (compare the peaks of $\vec{I}^{dirty,pb}$ with the RHS of Fig. 3.2 which uses the same Beam matrix $[B]$ and sky model \vec{I}^{sky} but has no primary beam). To remove this effect during image reconstruction (in non-zero regions of the primary beam), the primary beam can be combined with the sky model during deconvolution and the sky brightness reconstructed *via* a post-deconvolution division of the model image by the primary beam (*i.e.* solving this system of equations from left to right).

4.2.2 Visibility-domain corrections

In general, direction-dependent effects vary with time and frequency and are different for each antenna and baseline. This section describes a technique to model and correct these effects in the visibility domain where each measurement can be treated individually, and to do this within the standard deconvolution framework. These algorithms are described in detail in [Bhatnagar et al. \[2008\]](#); [Cornwell et al. \[2008\]](#).

4.2.2.1 Constructing a visibility-domain gridding-convolution function

The general idea of correcting direction-dependent effects in the visibility domain is the following. When there are direction-dependent effects, a measured visibility can be written as the result of convolving the true visibility function with another function \vec{K}_{ij}^{dd}

centered at the spatial frequency defined by the ij baseline¹. The effect of this convolution can be undone by placing the measured visibility on the spatial frequency plane and then convolving it with an estimate of $\vec{K}_{ij}^{dd^{-1}}$ before resampling it onto a coarse uv -grid². This calculation is similar to the convolutional resampling done for gridding (section 3.2.2) and the correction of direction-dependent effects can be combined with the gridding process as a sequence of convolutions on the spatial-frequency plane done before resampling³. A gridding-convolution function that corrects direction-dependent effects for baseline ij during gridding is $\vec{K}_{ij}^{dd^{-1}} \star \vec{P}_s$ where \vec{P}_s is the prolate spheroidal function. Several direction-dependent effects can be combined to form a single gridding-convolution function.

The image formed using these gridding-convolution functions will be devoid of all the direction-dependent effects that were included in the model (upto the accuracy allowed by FFT-related numerical errors and approximations and truncations used to compute these deconvolution kernel functions.).

Note that when the inverse of \vec{K}_{ij}^{dd} does not exist, its complex conjugate can be used to calculate an approximate inverse (element-by-element).

$$\vec{K}_{ij}^{dd^{-1}} \approx \vec{K}_{ij}^{dd^*} / (\vec{K}_{ij}^{dd^*} \cdot \vec{K}_{ij}^{dd}) \quad (4.6)$$

The following sections list various forms of \vec{K}_{ij}^{dd} .

4.2.2.2 Primary beam effects

Let $\vec{J}_{m \times 1}$ represent a single-polarization aperture illumination function⁴, discretized on the fine spatial frequency grid. Let $\vec{V}_p = [F^\dagger] \vec{J}$ be the corresponding antenna voltage pattern across the entire field of view allowed by the time and frequency resolutions of the measurements. The primary beam \vec{P}_{bij} affecting measurements from the baseline formed by antennas i and j is given by

$$\vec{P}_{bij} = \vec{V}_{pi} \cdot \vec{V}_{pj}^* = ([F^\dagger] \vec{J}_i) \cdot ([F^\dagger] \vec{J}_j^*) = [F^\dagger] (\vec{J}_i \star \vec{J}_j^*) = [F^\dagger] \vec{K}_{ij}^{pb} \quad (4.7)$$

where \star represents convolution⁵. During measurement, the sky brightness measured by baseline i, j is multiplied by this primary beam \vec{P}_{bij} . This is equivalent to a convolution on

¹ The elements of \vec{K}_{ij}^{pb} represent the convolution function for baseline ij , pixellated on a fine spatial frequency grid. The term $[K_{ij}^{dd}]$ in Eqn. 2.19 represents the same quantity, again for one baseline ij , but is a 4×4 matrix that describes the full-polarization response of the feeds, evaluated at *one* aperture point.

² A convolution with a function \vec{K} can be undone via a second convolution with $1/\vec{K}$ as the kernel.

³ A sequence of convolutions can be computed via a single convolution using a kernel constructed as the convolution of the individual kernels. This is possible because convolution is associative and commutative.

⁴ J in Eqn. 2.15 represents a 2×2 Jones matrix for *one* aperture point. \vec{J} is a vector formed from one element of the Jones matrix (the first diagonal element, one polarization), over a range of aperture points.

⁵ In Eqn. 2.17, $[K_{ij}] = [J_i] \otimes [J_j]$ is a 4×4 full-polarization matrix for *one* aperture point, but here, $\vec{K}_{ij}^{pb} = \vec{J}_i \star \vec{J}_j^*$ is the convolution of two aperture illumination functions, for one polarization pair. These two expressions become equal in the following situation. If the aperture illumination functions are δ -functions,

the spatial frequency plane with \vec{K}_{ij}^{pb} , a vector containing the convolution of the illumination patterns of antennas i and j .

For time-varying primary beams, $\vec{K}_{ij}^{pb} = \vec{J}_i \star \vec{J}_j^*$ is different for each timestep. When \vec{J}_i is different for different antennas, \vec{K}_{ij}^{pb} also varies with baseline. The A-projection algorithm [Bhatnagar et al. 2006; 2008] uses \vec{K}_{ij}^{pb*} as a time-varying baseline-based convolution function along with the prolate spheroidal function for gridding. Another approach for correcting for \vec{K}_{ij}^{pb} (or D_{ij}^{sky} for the primary beam) is the direct evaluation of the integral in Eqn. 2.18, separately for each baseline, during the forward and reverse transforms of iterative image deconvolution [Uson and Cotton 2008].

The use of \vec{K}_{ij}^{pb*} to correct primary beam effects is equal to using only the numerator of Eqn. 4.6. It cancels the phase of \vec{K}_{ij}^{pb} but squares its amplitude. The image formed by such a gridding-convolution function will contain no phase effects, but will be multiplied by an extra instance of the primary beam amplitude. To correct this, the image is divided by an average weight image $\vec{I}^{wt,pb} = [F^\dagger] \sum_{ij} \vec{K}_{ij}^{pb*} w_{ij} \vec{K}_{ij}^{pb}$ (where w_{ij} are the imaging weights). However, when \vec{K}_{ij}^{pb} varies with time, frequency and baseline, an image-domain normalization by the average weight image is an approximation of the denominator of Eqn. 4.6 and does not correctly cancel out the primary beams for the individual baselines and timesteps. The peak fluxes in the dirty image are therefore approximate, and the use of the correct \vec{K}_{ij}^{pb} for each baseline and timestep during the prediction of model visibilities is a necessary condition for the convergence of the image reconstruction process. Section 4.3 describes this process in detail.

4.2.2.3 Pointing Offsets

An antenna pointing offset can be described by a phase gradient across the aperture illumination function \vec{J}_i of that antenna. Let l_i, m_i and l_j, m_j be direction cosines that describe the pointing offsets of antennas i and j from the phase reference center (in the image domain). For a primary beam modeled as a Gaussian [Bhatnagar et al. 2004], the associated convolution operator for the baseline ij is $K_{ij}^{po} = e^{-[(l_i - l_j)/\sigma_l]^2 - [(m_i - m_j)/\sigma_m]^2} e^{i[(u - u_{ij})(l_i + l_j) + (v - v_{ij})(m_i + m_j)]}$ where σ_l, σ_m represent the width of the associated antenna primary beam. The phase term depends on the pointing offsets of the two antennas, and is a phase gradient across a range of spatial frequencies around the central spatial frequency (uv -track) measured by the baseline ij . For small offsets, $[K_{ij}^{po}]$ is purely a phase-effect and a convolution function constructed as \vec{K}_{ij}^{po*} will remove this offset.

then \vec{K}_{ij}^{pb} is also a δ -function, can be represented completely for baseline ij by a single scalar and becomes one element of the 4×4 matrix $[K_{ij}]$. This corresponds to a flat primary beam (Fourier transform of a δ -function). In other words, when there is only *one* aperture point, the voltage pattern is constant across the field of view and corresponds to a direction-independant complex antenna gain (section 3.1.1) that the process of calibration solves for. For a detector with a finite aperture (more than one aperture point), the antenna response (voltage pattern) is no longer a direction-independent effect.

4.2.2.4 The w -term

The effect of sky curvature across wide fields of view can be described as a baseline-based convolution of the source visibility function with $\vec{K}_{ij}^{wp} = [F][\vec{W}_{ij}^{proj}]$, where $W_{ij}^{proj} = e^{-2\pi i w_{ij}(\sqrt{l^2+m^2}-1)}$ (for each l, m) is the Fresnel propagator (see Eqn. 2.8) evaluated for every pixellated direction (l, m) in the image domain. The w -projection algorithm [Cornwell et al. 2003; 2008] uses \vec{K}_{ij}^{wp*} with the prolate spheroidal function during gridding-convolution. In this case, $[K^{wp}]$ is exactly unitary, and \vec{K}_{ij}^{wp*} cancels the w -term. This operation is equivalent to converting visibilities to that of a flat sky before making an image.

Other methods exploit the fact that the effect of the w -term is small close to the phase tracking center. Faceting algorithms divide the field of view into a number of facets. Images are made by either projecting the facet images onto the local tangent plane (image-plane faceting [Cornwell and Perley 1992]) and using the appropriate PSF for the deconvolution of individual facet images, or by projecting the (u, v) for each facet onto a single tangent plane in the gridding step required for an FFT-based reverse transform [Sault et al. 1999]. For very wide fields, a combination of faceting and w -projection is used.

4.2.2.5 Mosaicing

An image of a region of the sky much larger than the field of view of each antenna can be constructed from a mosaic of pointings [Cornwell 1988]. Mosaicing is an example where a visibility-domain convolution operator is used to introduce a direction-dependent effect and not correct it. The effect introduced is an intentional pointing offset. A mosaic observation usually consists of a series of pointings where the phase tracking center for each pointing coincides with its pointing center. Traditional instruments observe a series of pointings one after another, and focal plane arrays observe several pointings in parallel.

One approach to make a mosaic image is to deconvolve each pointing separately and then stitch together the final images, while accounting for any regions of overlap. This method is simple, but prone to errors if there are bright sources beyond the main lobe in any individual pointing, which is more than likely to be the case for a mosaicing observation.

Another approach is to combine the data from all pointings during the gridding stage, and create a single large image on which a single joint deconvolution can be done in the minor cycle. The gridding-convolution functions for each pointing need to have a phase gradient that describes the offset between the center of the image and that of each pointing center. The gridding-convolution operator is constructed from $\vec{K}_{ij}^{mos} = e^{i(u l_p + v m_p)}$ where l_p, m_p are direction cosines that describe the difference between the center of the mosaic image and the center of pointing p . The weight image will now represent the direction-dependent sensitivity across the entire field of view of the mosaic, due to different w_{sum} for the different pointings. The prediction step needs to apply the inverse of K^{mos} to compute model visibilities that can be compared with data from individual pointings.

4.3 Wide-field Imaging with Generalized direction-dependent effects

Section 4.3.1 lists the measurement and normal equations for wide-field imaging. Section 4.3.2 describes the AW-projection algorithm [Bhatnagar et al. 2008] that solves these equations *via* a combination of visibility-domain and image-domain operations. These equations and algorithms are listed in terms of a generic direction-dependent effect \vec{K}_{ij}^{dd} . However, to provide an intuitive feel for these equations, we also show the specific form they take when $\vec{K}_{ij}^{dd} = \vec{K}^{pb}$ (antennas with identical and time-invariant primary beams as discussed in section 4.2.1). The purpose of this exercise is to clarify the difference between the two methods (sections 4.2.1 and 4.3.2) and to show how a practical iterative algorithm combines approximate and exact calculations to converge to an appropriate solution.

4.3.1 Imaging Equations

This section describes the process of image reconstruction while correcting for direction-dependent effects within the iterative deconvolution framework introduced in Chapter 3.

4.3.1.1 Measurement equations

Eqn. 2.19 shows that the visibilities measured by baseline ij are the result of a convolution of the true visibility function with a function \vec{K}_{ij}^{dd} . Eqn. 2.22 shows the corresponding measurement equation in matrix form (and is re-written here).

$$\vec{V}^{corr} = [S^{dd}][F]\vec{I}^{sky} \quad (4.8)$$

\vec{I}^{sky} represents the sky brightness over the complete field of view allowed by the time and frequency resolution of the measurements, \vec{V}^{corr} are calibrated visibilities and $[S^{dd}]$ is a sampling matrix that includes a baseline-based convolution of the visibility function with \vec{K}_{ij}^{dd} .

When $\vec{K}_{ij}^{dd} = \vec{K}^{pb}$ (identical and time-invariant antennas), the primary beams of all antennas are given by $\vec{P}_b = [F^\dagger]\vec{K}^{pb}$ and we can construct a single convolution operator for all baselines (the entire spatial-frequency plane) as $[G_{m \times m}^{pb}] = [FP_bF^\dagger]$ (see section 2.2.2.5 for a derivation), and separate the steps of convolution and sampling. The measurement equations become

$$\vec{V}^{corr} = [S][G^{pb}][F]\vec{I}^{sky} = [S][F][P_b]\vec{I}^{sky} \quad (4.9)$$

where $[P_b] = diag(\vec{P}_b)$. This equation is the same as Eqn. 4.2 and shows the connection between an image-domain multiplication with the primary beam and a visibility-domain convolution with \vec{K}^{pb} .

4.3.1.2 Normal equations

The normal equations constructed from Eqn. 4.8 are given below.

$$[F^\dagger S^{dd^\dagger} W^{im} S^{dd} F] \vec{I}^{sky} = [F^\dagger S^{dd^\dagger} W^{im}] V^{corr} \quad (4.10)$$

The RHS of Eqn. 4.10 shows that $[S^{dd^\dagger}]$ convolves the measured visibility from each baseline (ij) with $\vec{K}_{ij}^{dd^\dagger}$ during gridding. The LHS shows that we now have two instances of the amplitude of $[S^{dd}]$ (their phases cancel out due to the complex conjugate in $[S^{dd^\dagger}]$, but their amplitudes multiply).

The weight image \vec{I}^{wt} is constructed as follows and describes the direction-dependent sensitivity of the imaging instrument (the direction-dependent equivalent of the weight image $\vec{I}^{wt} = w_{sum} \vec{I}$ discussed in section 3.2.1.5).

$$\vec{I}^{wt} = [F^\dagger] \sum_{ij} \vec{K}_{ij}^{dd^*} w_{ij} \vec{K}_{ij}^{dd} \quad \text{or} \quad [\vec{I}^{wt}] = [F^\dagger] [S^{dd^\dagger} W^{im} S^{dd}] \quad (4.11)$$

If the elements of \vec{K}_{ij}^{dd} contain only phase terms, the use of a complex conjugate during gridding will remove the direction-dependent effect completely. The resulting dirty image (RHS of Eqn. 4.10) will be devoid of direction-dependent effects and can be sent into a deconvolution minor cycle that knows nothing about direction-dependent effects (as described in section 3.2.1.9). If the elements of \vec{K}_{ij}^{dd} contain amplitude and phase terms (e.g. K^{pb}), the use of the complex conjugate will not correct it completely, and the dirty image (RHS of Eqn. 4.10) will not be free of direction-dependent effects. In this case, the weight image (Eqn. 4.11) is a measure of this multiplicative image-domain effect.

For example, when $\vec{K}_{ij}^{dd} = \vec{K}^{pb}$, we can write $[S^{dd}] = [S][G^{pb}] = [S][FP_b F^\dagger]$ and write the Hessian matrix as a product of two diagonal primary beam matrices and the Beam matrix $[B] = [F^\dagger S^\dagger W^{im} S F]$.

$$[P_b^\dagger][B][P_b] \vec{I}^{sky} = [P_b^\dagger][F^\dagger S^\dagger W^{im}] \vec{V}^{corr} \quad (4.12)$$

$$\text{or} \quad [H^{pb^2}] \vec{I}^{sky} = \vec{I}^{dirty,pb^2} \quad (4.13)$$

Note that the primary beam is a real function, and $[P_b^\dagger] = [P_b]$.

By analogy to Eqn. 4.11, \vec{I}^{wt} is computed by accumulating a weighted average of $\vec{K}_{ij}^{pb^*} \vec{K}_{ij}^{pb}$ (computed from aperture illumination functions that vary with antenna and time) onto the center of the spatial-frequency plane and then taking a Fourier inverse. This weight image now represents an average of the square of the primary beam amplitude (phase effects have been eliminated by this stage), and an average primary beam can be computed from it as follows.

$$\vec{I}^{wt} = w_{sum} \vec{P}_b^2 \quad \text{and} \quad \vec{P}_b = \sqrt{\vec{I}^{wt} / w_{sum}} \quad (4.14)$$

Note that if both sides of Eqn. 4.12 are divided by \vec{P}_b , we get back the normal equations shown in Eqn. 4.4, which can be solved using standard deconvolution algorithms for the minor cycle and interpreting the model image as a product of the sky and the primary beam.

Figures 4.2 and 4.3 are pictorial representations of the normal equations shown in Eqn. 4.12. Fig. 4.2 shows the Hessian matrix ($[H^{pb^2}] = [P_b][B][P_b]$) being multiplied with a sky brightness distribution containing two point sources. Note that $[H^{pb^2}]$ is not a convolution operator, and its diagonal represents the weight image $\vec{I}^{wt} = w_{sum}\vec{P}_b^2$. Figure 4.3 shows the same equation with the beam matrix and the primary beams separated out. This factorization shows that this system of equations can be solved (from left to right) by first dividing the RHS by \vec{P}_b , followed by a minor cycle deconvolution (to eliminate the effect of $[B]$) and finally a post-deconvolution division of the model image by the average primary beam \vec{P}_b . Section 4.3.2 describes this process.

4.3.1.3 Principal Solution

As defined in section 3.2.1.6, the principal solution is computed by dividing the dirty image (RHS of Eqn. 4.10) by the diagonal of the Hessian matrix (the weight image \vec{I}^{wt}). The peaks in the resulting image will measure the sky brightness distribution in units of Jy/beam. When $\vec{K}_{ij}^{dd} = \vec{K}^{pb}$ and $\vec{I}^{wt,pb} = w_{sum}\vec{P}_b^2$, the dirty image (RHS of Eqn. 4.12) contains two instances of the primary beam (one present in the data and one introduced by the gridding process). A division by the weight image will remove this factor of \vec{P}_b^2 and in the limit of complete spatial-frequency sampling, this principal solution will be the final reconstructed image.

4.3.2 Iterative Deconvolution

This section describes how direction-dependent corrections are applied within an iterative deconvolution framework. Direction-dependent effects that affect only the phase of the visibilities can be corrected during gridding, but it is not always possible to include amplitude effects in the gridding-convolution functions and image-domain operations are sometimes required. In the AW-Projection algorithm [Bhatnagar et al. 2008] discussed below, a combination of visibility-domain and image-domain corrections are used.

Pre-compute Hessian : The \vec{I}^{psf} is computed via Eqn. 3.21 (i.e. using only the prolate spheroidal \vec{P}_s as the gridding-convolution function and no direction-dependent effects).

The weight image \vec{I}^{wt} is computed via Eqn. 4.11. If all $[K_{ij}^{dd}]$ are unitary, the use of the complex conjugate will eliminate these effects and $\vec{I}^{wt} = w_{sum}\vec{1}$. Otherwise, the weight image is a measure of the direction-dependent sensitivity of the instrument.

Pre-compute Primary Beam : When the antenna primary beam is the dominant direction-dependent effect, an average primary beam is computed via Eqn. 4.14. In practice, primary beams vary with baseline and time and this average primary beam is only an approximation.

$$\left[H_{m \times m}^{pb^2} \right] \left[\vec{I}_{m \times 1}^{sky} \right] = \left[\vec{I}_{m \times 1}^{dirty, pb^2} \right]$$

Figure 4.2: Normal Equations for General Primary-Beam correction : This diagram represents the imaging process when the primary beam is the dominant direction-dependent effect and its correction is done *via* a visibility-domain convolution (Eqn. 4.13). The Hessian matrix is no longer a pure convolution operator, or in other words, the PSF is spatially variant (compare with $[B]$ in Figs. 3.2 and 4.1). The diagonal of the Hessian (Eqn. 4.14) represents the instrument's sensitivity as a function of direction. When this Hessian operates on the sky model, the peaks in the dirty image are attenuated by \vec{P}_b^2 and the sidelobes are attenuated by \vec{P}_b (compare with the RHS of Fig. 4.1 which uses the same sky model \vec{I}^{sky} but has only the one \vec{P}_b already present in the data). This system can also be written as shown in Fig. 4.3.

$$\left[P_b^T \right]_{m \times m} \left[B_{m \times m} \right] \left[P_b \right]_{m \times m} \left[\vec{I}_{m \times 1}^{sky} \right] = \left[\vec{I}_{m \times 1}^{dirty, pb^2} \right]$$

Figure 4.3: Modified Normal Equations for General Primary-Beam correction : This diagram is a factorized version of Fig.4.2 (Eqn. 4.12). The Hessian on the LHS is now written as a matrix product of two diagonal matrices containing the primary beam $[P_b]$ and a convolution operator (the Beam matrix $[B]$ from Fig. 3.2). This system can be solved from left to right by first multiplying both sides of the equation by $[P_b^{-1}]$ to convert the system into a pure convolution equation of the type shown in Fig.4.1. The remaining primary beam is combined with the sky model for a set of minor cycle deconvolution iterations, and divided out of the model image in a post-deconvolution step (before the prediction step re-introduces the effect).

Any image-domain operations that use it will not exactly correct the effect of time-variable beams. For sources in the main lobe of the primary beam, these errors may be negligible but for sources out in the sidelobes, beam asymmetries could result in 100% gain variations across the duration of the observation. However, the resulting errors are in the image domain and affect only the minor cycle and several iterations with accurate major cycle prediction are usually able to compensate for this.

Initialization : The model image \vec{I}^{model} is initialized to zero or to an *a-priori* model.

Major and minor cycles : Steps 1, 2, 6 and 7 represent the major cycle of iterative deconvolution, in which the data and models are converted between the visibility and image domains while accounting for direction-dependent effects. Steps 3 to 5 represent the minor cycle, and are identical to those described in section 3.2.1.9 for deconvolution without any direction-dependent corrections.

1. **Compute RHS :** The un-normalized residual image is computed with sampling matrix $[S^{dd}]$ constructed from a baseline-based $\vec{K}_{ij}^{dd^{-1}} \star \vec{P}_s$ as described in section 4.2.2.

$$\vec{I}^{res}_{m_l \times 1} = [F^\dagger R S^{dd^\dagger} W^{im}] \vec{V}^{res}_{n \times 1} \quad (4.15)$$

For the first iteration, $\vec{V}^{res} = \vec{V}^{corr}$ and $\vec{I}^{res} = \vec{I}^{dirty}$. All subsequent iterations use $\vec{V}^{res} = \vec{V}^{corr} - \vec{V}^{model}$.

2. **Normalization :** There are two ways of normalizing the residual image before beginning minor cycle iterations. One approach is to calculate the principal solution by dividing the RHS by \vec{I}^{wt} , and the other is to divide the RHS by an estimate of the average \vec{P}_b and the sum-of-weights. Both approaches are valid, so it is important to clarify the difference between the two.

- (a) **Flat-noise :** Normalization by $\vec{P}_b w_{sum}$ before deconvolution will result in a model image described by $\vec{I}^{model} = \vec{I}^{sky} \cdot \vec{P}_b$. (see Eqns. 4.12 and 4.4). When the primary beam is the dominant direction-dependent effect, this normalization is consistent with the measurement process (i.e. the interferometer samples the product of the sky and \vec{P}_b). The noise in the image is related directly to the measurement noise due to the interferometer, and is the same all across the image. The minor cycle can give equal weight to all flux components that it finds, but it needs to correct for this attenuation (assuming the form of attenuation is known) to generate a true model of the \vec{I}^{sky} . This form of normalization is useful when the primary beam is the dominant direction-dependent effect because the images going into the minor cycle satisfy a convolution equation. It is also more appropriate for single-pointing fields of view.

(b) **Flat-sky** : Normalization by the weight image $\vec{I}^{wt} = \vec{P}_b^2 w_{sum}$ before deconvolution will give a dirty image whose peaks are consistent with \vec{I}^{sky} and the model image will be free of the primary beam. However, the image going into the minor cycle will not satisfy a convolution equation and the noise in the dirty image will be higher in regions where \vec{I}_{pb}^{wt} is small. The minor cycle needs to account for this while searching for flux components (a signal-to-noise dependent CLEAN). This form of normalization is useful for mosaic imaging where the sky brightness can extend across many pointings. In this case a minor cycle that solves directly for \vec{I}^{sky} will incur fewer errors than one that solves for $\vec{I}^{sky}\vec{P}_b$. This is because mosaic observations are done for sources with spatial scales larger than the field of view of each antenna, and therefore not present in the data. Allowing the minor cycle to use flux components that span across beams of adjacent pointings will provide a better constraint on the reconstruction of these unmeasured spatial frequencies. A flat-sky dirty image is more likely to produce smoother large-scale emission.

3. **Find a flux component** : When the residual image is computed with a flat-noise normalization, this step is identical to step 2 for general iterative deconvolution (section 3.2.1.9). When the residual image has a flat-sky normalization, the process of finding valid flux components must take into account the varying noise and sidelobe levels across the field of view.
4. **Update model image** : Same as step 3 for general iterative deconvolution (section 3.2.1.9).
5. **Update RHS** : Same as step 4 for general iterative deconvolution (section 3.2.1.9).
Repeat from step 3 until the minor cycle flux limit is reached.
6. **Correct for PB** : Depending on the choice of normalization (step 2) the model image at the end of the minor cycle has to be further processed before model visibilities can be predicted from it.
 - (a) **Flat noise** : This step is equivalent to the post-deconvolution primary beam correction described in the previous section. A new model image is computed as $\vec{I}^{model}/\vec{P}_b$.
 - (b) **Flat sky** : No corrections are required because the flux model is already devoid of primary beam effects.
7. **Predict** : The prediction step of computing model visibilities from the current sky model needs to re-introduce all the direction-dependent effects that are being corrected for during gridding, before the model can be compared with the data for χ^2 computation.

$$\vec{V}_{n \times 1}^{model} = [S^{dd} R^\dagger F][I^{ps}]^{-1} \vec{I}_{m \times 1}^{model} \quad (4.16)$$

The reverse transform (steps 1 and 2) is usually approximate because the dirty image is normalized by a weight image computed as the average of all the time and baseline dependent effects, but this forward transform is accurate with model visibilities being predicted by applying time and baseline dependent effects to the individual visibilities.

Repeat from step 1 until the final convergence criterion is satisfied.

Restoration : The final model image is restored by smoothing it with a restoring beam, and adding back the residuals computed *via* Eqn. 4.15 and normalized by \vec{I}^{wt} (the principal solution).

Algorithm 2: CLEAN with Visibility-Domain Corrections for Direction-Dependent Effects.

Data: Calibrated visibilities : $\vec{V}_{n \times 1}^{corr}$
Data: Gridding convolution function : $\vec{K}_{ij}^{dd} \star \vec{P}_s$
Data: uv-sampling function : $S_{n \times m}$
Data: Image noise threshold and loop gain σ_{thr}, g_s
Result: Model Image : $\vec{I}_{m \times 1}^{model}$

- 1 Compute the PSF : \vec{I}^{psf}
- 2 Compute the weight image : \vec{I}^{wt} , using $[S^{dd}]$ and $[S^{dd^\dagger}]$
- 3 Compute the primary beam \vec{P}_b from \vec{I}^{wt}
- 4 Compute the dirty image : \vec{I}^{dirty} using $[S^{dd^\dagger}]$
- 5 Measure the peak PSF sidelobe : $f_{sidelobe} = \max\{\text{sidelobe}(\vec{I}^{psf})\}$
- 6 Initialize the model image and residual image : $\vec{I}^m = 0, \vec{I}^{res} = \vec{I}^{dirty}$
- 7 **repeat** /* major cycle */
- 8 Normalize the residual image : $\vec{I}^{res} = [\vec{P}_b]^{-1} \vec{I}^{dirty}$
- 9 Compute a flux limit : $f_{limit} = \max\{\sigma_{thr}, f_{sidelobe} * \max\{\vec{I}^{res}\}\}$
- 10 **repeat** /* minor cycle */
- 11 Find the location and amplitude of the peak : $\delta\vec{I}^m = \text{peak}(\vec{I}^{res})$
- 12 Update the model image : $\vec{I}^{model} = \vec{I}^{model} + g_s \delta\vec{I}^{model}$
- 13 Update the residual image : $\vec{I}^{res} = \vec{I}^{res} - g_s [\delta\vec{I}^{model} \star \vec{I}^{psf}]$
- 14 **until** $\max\{\vec{I}^{res}\} < f_{limit}$
- 15 Divide the model image by the primary beam : $\vec{I}^{model} = \vec{I}^{model} / \vec{P}_b$
- 16 Predict model visibilities \vec{V}^{model} from \vec{I}^{model} using $[S^{dd}]$
- 17 Compute a new residual image \vec{I}^{res} from $\vec{V}^{res} = \vec{V}^{corr} - \vec{V}^m$ using $[S^{dd}]$
- 18 **until** $\max\{\vec{I}^{res}\} < \sigma_{thr}$
- 19 Restore the final model image

CHAPTER 5

IMAGING WITH FREQUENCY-DEPENDENT EFFECTS

Broad-band receivers are being introduced into radio interferometry for two main reasons. The first is to increase the sensitivity of the instrument and provide high-dynamic-range continuum imaging capabilities superior to those of existing radio interferometers, and the second is to use the wide bandwidths for detailed measurements of the spectral structure of astrophysical sources. This chapter discusses the use of broad-band receivers in radio interferometry, describes its effects on the imaging process, and summarizes existing methods of multi-frequency image reconstruction.

The term multi-frequency synthesis (MFS) is commonly used to describe all methods of multi-frequency image reconstruction. In general, MFS imaging is the process of making a single continuum Stokes-I image by combining the measurements from all frequencies within the band. For the purpose of this dissertation, MFS refers to the process of creating an image by gridding all visibilities onto one single *uv*-grid and using standard deconvolution algorithms to construct an image of the continuum sky brightness. The basic assumption with this technique is that the sky brightness has a flat spectrum. Multi-frequency deconvolution refers to parameterizing the frequency dependence of the sky brightness and solving for these parameters during deconvolution. One approach is to calculate spectral PSFs that describe the instrument's response to different spectral basis functions (for example, Taylor series functions), and to then perform a joint deconvolution to simultaneously extract images of the sky brightness as well as its spectral structure. The Sault-Wieringa Multi-Frequency-CLEAN (SW-MFCLEAN) algorithm parameterizes each pixel in the image with a constant across frequency and a slope across frequency (first two Taylor functions). A more general description of this algorithm for an arbitrary number of terms in the Taylor polynomial will be referred to as MF-CLEAN (MF-CLEAN).

Section 5.1 first defines the problem of multi-frequency synthesis imaging by describing the main effects of using broad-band receivers for Stokes I imaging using radio interferometers. Section 5.2 compares several existing wide-band image reconstruction methods, explores their limits in the context of high-dynamic-range imaging with the EVLA, and identifies areas where improvements are required (multi-scale deconvolution, higher-order spectral model, and frequency-dependent primary-beam correction). Chapter 6 describes multi-scale and multi-frequency deconvolution algorithms in a common framework to demonstrate that they are based on a similar idea. Chapter 7 then describes new algorithms that build on the basic multi-scale and multi-frequency deconvolution methods and includes wide-field imaging concepts described in chapter 4.

5.1 Wide-Band Radio Interferometry

The use of broad-band receivers in an interferometer can increase the signal-to-noise ratio in the reconstructed continuum image by a factor equal to the square root of the increase in bandwidth (according to the ideal radiometer equation¹, the noise on a total-power measurement is proportional to $T_{\text{sys}} / \sqrt{\Delta\nu \cdot \tau}$ where T_{sys} is the instrumental system temperature, $\Delta\nu$ is the receiver bandwidth, and τ is the integration time.). However, to reconstruct a wide-band image of the sky and achieve the desired continuum sensitivities, imaging algorithms need to be sensitive to the effects of combining measurements from a large range of frequencies.

There are three main frequency-dependent effects that need to be accounted for when using broad-band receivers for Stokes-I imaging² using an interferometer. First, the uv -coverage of an interferometer is frequency-dependent and the related imaging properties of the array will vary across the sampled frequency range. Second, the incident radiation is usually neither exactly monochromatic nor constant across the wide-band, and the spectrum of the sky brightness also needs to be modeled and reconstructed. Finally, the size and shape of the aperture illumination function of each antenna depends on frequency and the antenna field of view and power response will vary across the band. Both these instrumental effects can be computed (or measured) and applied (or corrected) during imaging. Sections 5.1.1 and 5.1.2 describe these effects in detail, in the context of multi-frequency synthesis imaging.

5.1.1 Multi-Frequency Measurements

Multi-frequency measurements of the visibility function of the sky brightness distribution can improve the imaging properties of an interferometer. This section contains a brief description of how multi-frequency measurements are used in synthesis imaging.

5.1.1.1 Multi-Frequency Spatial-Frequency Coverage

Figure 5.1 shows an example of the multi-frequency uv -coverage of the EVLA telescope at three frequencies across L-Band. Each frequency channel ν measures a different range of spatial frequencies (given by $\frac{\nu}{c} \vec{b}_{\min}$ to $\frac{\nu}{c} \vec{b}_{\max}$ where $\vec{b}_{\min}, \vec{b}_{\max}$ are measured in meters) and the imaging properties of the telescope differ across the band. As the observing frequency increases, the width of the point-spread function (given by $\theta_\nu = 1/u_{\max}$ radians) and the sensitivity to large spatial scales (inversely proportional to the size of

¹ Appendix A lists the formula to calculate the image-domain noise level for a set of visibilities measured with a radio interferometer.

²For full-Stokes imaging, the polarization properties of the sky brightness and the instrument vary with frequency and need to be taken into account [Sault and Wieringa 1994]. However, all the discussions in this dissertation focus on only Stokes-I imaging and ignore all broad-band polarization effects.

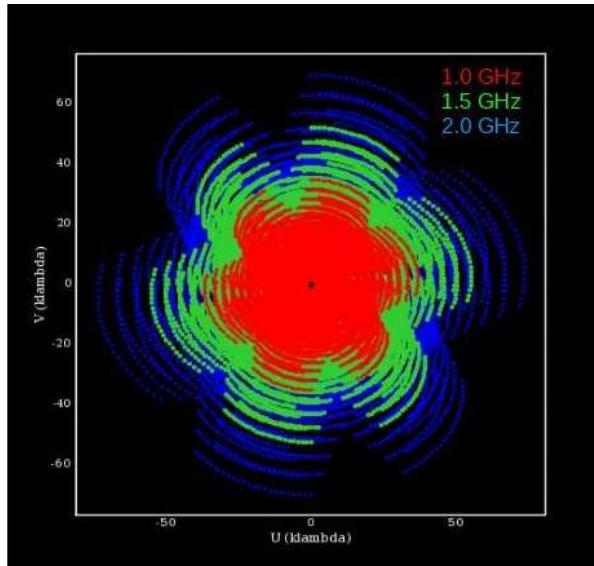


Figure 5.1: This plot shows the multi-Frequency uv -coverage of the EVLA at L-Band for a 4 hour synthesis run. The axes are in units of $k\lambda$ and the three colours (Red,Green,Blue) correspond to the uv -coverage at three frequencies 1.0, 1.5, 2.0 GHz. The angular resolution of the instrument for continuum imaging is given by the maximum spatial frequency (measured at 2.0 GHz). In regions of overlap between frequencies, redundant measurements increase the signal-to-noise ratio of the measured visibility function (for a flat-spectrum source). The combined uv -coverage also has fewer unmeasured spatial frequencies within the maximum range, and this increases the fidelity (accuracy) of the reconstructed image.

the zero-spacing hole) decreases. The phase delay due to the w -term is also frequency-dependent (in the sense that w is measured in units of wavelength), which means that the gridding convolution function used in w -projection must be chosen according to the observing frequency. For instruments like the EVLA that have a dense spatial frequency coverage at any given frequency, the main advantage of multi-frequency-synthesis is increased sensitivity and the measurement of the spectrum across the band. For instruments like the ATCA/e-MERLIN/VLBA with sparse uv -coverage, the main advantage of multi-frequency-synthesis is to fill in the unmeasured regions of the spatial frequency plane.

5.1.1.2 Frequency Resolution / Channel Width

The choice of frequency resolution (or channel width) at which visibilities must be measured for synthesis imaging is influenced by the observing frequency and the desired field of view. It is based on the concept of bandwidth smearing, a radial degradation in the resolution and sensitivity of the array due to the mapping of a wide range of spatial frequency co-ordinates (measured across a wide-band) onto that corresponding to the central

frequency of the band. A receiver with finite bandwidth that produces one single visibility per baseline and time measures the average of the visibility function over the range of spatial frequencies spanned by the receiver bandwidth. If this range of spatial frequencies spans more than one uv grid cell, there will be image-domain errors due to bandwidth smearing. To eliminate these errors, the frequency resolution of the visibility measurements is chosen such that the amount of smearing in the image domain is smaller than the angular resolution of the telescope within the antenna field of view³. This bandwidth-smearing limit is derived below.

Bandwidth-Smearing limit : Let u_0 represent the spatial frequency of the center of the band and u_ν be that of another frequency in the band, such that $u_0 = \frac{\nu_0}{\nu} u_\nu$. When a wide bandwidth receiver produces only one single visibility per baseline and time, the whole range of spatial frequencies it measures is mapped onto that given by the central frequency of the band. The process of mapping u_ν onto u_0 is equal to a scaling of the uv -co-ordinates by $\frac{\nu_0}{\nu}$. According to the similarity theorem of Fourier transforms a scaling of the co-ordinate system in one domain is equivalent to an inverse scaling in the other domain. Therefore,

$$F\left[V_\nu\left(\frac{\nu_0}{\nu}u_\nu, \frac{\nu_0}{\nu}\nu_\nu\right)\right] = \left(\frac{\nu}{\nu_0}\right)^2 I_\nu\left(\frac{\nu}{\nu_0}l, \frac{\nu}{\nu_0}m\right) \quad (5.1)$$

where $F[V_\nu(u_\nu, \nu_\nu)] = I_\nu(l, m)$. When a whole range of spatial frequencies are averaged in this way, the image domain effect is a radial smearing where different frequencies are shifted radially by different amounts⁴.

The following condition ensures that across a given field of view, the image-domain bandwidth smearing is smaller than the angular resolution of the instrument. This is of importance in deciding the frequency resolution (channel-width) with which to make the observation.

$$\frac{\Delta\nu}{\nu_0} < \frac{Resolution}{FoV} = \frac{\lambda/b_{max}}{\lambda/D} = \frac{D}{b_{max}} \Rightarrow \Delta\nu < \nu_0 \frac{D}{b_{max}} \quad (5.2)$$

The desired field of view of the image (that decides the spatial-frequency grid cell size) is given by the HPBW of the antenna primary beam. Therefore, all visibilities measured within this narrow frequency range will map to one single uv -grid cell. For broad-band receivers, this limit will change across the band, and the channel width should be chosen as the limit computed for ν_{min} .

³ Time-smearing is another effect similar to bandwidth smearing. It occurs when the integration timestep is long enough that the spatial frequency measured by the baseline changes by more than the size of one uv -pixel. A similar limit will give the coarsest time resolution with which the visibilities can be measured.

⁴ Frequency-Augmentation is an algorithm (developed by T.J.Cornwell (private communication)) that corrects the effect of bandwidth smearing in data where only a single visibility is measured across the entire bandwidth. In the major cycle of deconvolution, the model visibility function is degridded using a spatial-frequency-dependent convolution function (box) that averages the visibility function from a range of uv before comparing it with the measurements for χ^2 computation.

Channel Averaging : New correlators are capable of generating visibility measurements at very high frequency resolution (well beyond the bandwidth-smearing limit) and producing visibility measurements for tens of thousands of frequency channels. This results in extremely high data rates. However, once the data have been calibrated, chunks of channels can be averaged together up to the bandwidth-smearing limit to reduce the data-rate for the processing involved in continuum imaging. For example, the number of channels can be reduced from 16000 to about 500 for EVLA L-band continuum imaging over a single pointing field of view.

5.1.1.3 Multi-Frequency Synthesis Imaging

Standard multi-frequency synthesis (MFS) imaging involves just gridding together visibilities from multiple frequencies onto a single spatial-frequency grid, and assuming that all frequencies measure the same visibility function, just at different spatial frequencies. The dirty image and PSF are given by a modified form of Eqns. 3.20 and 3.21.

$$\vec{I}^{dirty} = w_{sum}^{-1} [I^{wt^{-1}}] [F^\dagger RG^{ps}] \left\{ \sum_\nu [S_\nu^\dagger W_\nu^{im}] \vec{V}_\nu^{corr} \right\} \quad (5.3)$$

$$\vec{I}^{psf} = w_{sum}^{-1} [I^{wt^{-1}}] [F^\dagger RG^{ps}] \left\{ \sum_\nu [S_\nu^\dagger W_\nu^{im}] \vec{1} \right\} \quad (5.4)$$

As long as the sky brightness is constant across the total measured bandwidth and concentrated within a small field of view, standard imaging and deconvolution algorithms can be used along with MFS to construct an accurate continuum image. For sources with spectral structure, this approach will convert any spectral variations of the visibility function into spurious spatial structure. Therefore, to reconstruct the broad-band sky brightness distribution correctly, a spectral model must be folded into the reconstruction process. Section 5.1.2 describes this problem in detail.

5.1.2 Frequency Dependence of the Sky and Instrument

Multi-frequency synthesis imaging with standard deconvolution algorithms is based on the assumption that the spectrum of the measured sky brightness distribution is flat and that multi-frequency measurements only contribute to additional samples of a single visibility function. In reality, however the sky brightness distribution and instrumental effects are usually frequency-dependent, and this can result in artifacts in the reconstructed image if they are not accounted for during the imaging process.

5.1.2.1 Broad-Band Radio Continuum Spectra

The sky brightness distribution of astronomical sources usually varies with frequency, either due to the spectral structure of the emitted radiation or if the spatial structure of the radiating object varies with frequency. Broad-band continuum emission from astrophysical sources is often best represented by a power-law. This power-law index can vary with frequency, as seen in the case of spectral breaks, steepening and turnovers that are visible across the wide frequency ranges that new receivers are now sensitive to. A power-law with varying index can be used to describe such a spectrum. I_{ν_0} is the sky brightness distribution at a reference frequency ν_0 , $\alpha(\nu)$ is a spectral index that varies with frequency, and I_ν is the flux seen at the observing frequency ν .

$$I_\nu = I_{\nu_0} \left(\frac{\nu}{\nu_0} \right)^{\alpha(\nu)} \quad (5.5)$$

For example, for synchrotron emission from radio galaxies, typical values of α range between 0.0 and -1.5 . A spectral index of -1.0 corresponds to a 50% change in flux across a bandwidth given by $\nu_{max} : \nu_{min} = 2 : 1$.

5.1.2.2 Frequency Dependent Instrumental Effects

The angular size of an antenna primary beam decreases as the observing frequency increases. This changes the field of view at each frequency (defined as the FWHM of the main lobe of the beam, given by λ/D where D is the diameter of the reflecting dish). As a result, a source located away from the pointing-center of these beams will be attenuated by different amounts across the frequency band. This will give rise to artificial spectral structure in the measurements which if uncorrected, will affect the imaging process in the same way as sky emission with intrinsic spectral structure would. Note that for EVLA antennas the feeds for both polarizations are offset from the antenna axis, leading to a polarization dependent pointing offset called beam squint, which is also frequency-dependent. Such offsets can be described by phase-ramps in the nominal aperture illumination pattern and are correctable during the imaging process.

Figure 5.2 shows the shape of the primary beam of an EVLA antenna at 1.0, 1.5 and 2.0 GHz. Figure 5.3 shows 1-dimensional cuts through these primary beams (chosen to pass through a peak in the first sidelobe) along with a frequency-averaged profile. From these plots, we can see that the primary beam introduces significant spectral structure into the measured brightness distribution even near the HPBW of the lowest frequency. Sources at higher angular distances from the pointing center will be completely attenuated (and not detected) at some range of frequencies within the band and this will result in a reduced continuum imaging sensitivity. However, such an instrument will still be sensitive to sources out to the full field of view of the lowest frequency (at the few % level) but the spectral structure due to the primary-beam sidelobes will not be monotonic and this will complicate the deconvolution of such sources from the continuum image.

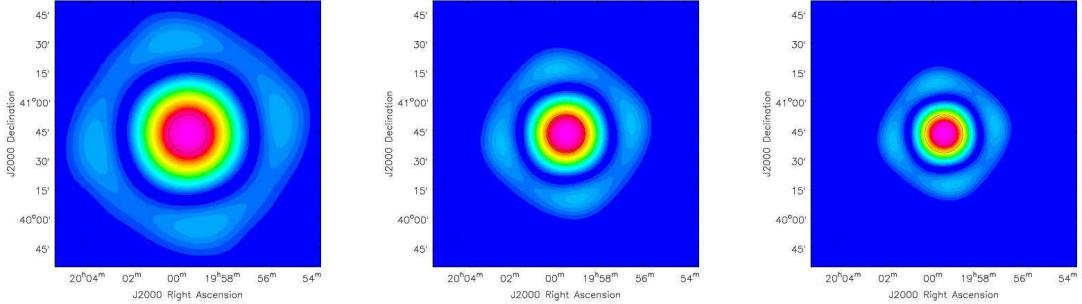


Figure 5.2: Simulated EVLA primary beams at 1.0, 1.5 and 2.0 GHz. These beams were computed from frequency-dependent models of antenna aperture illumination functions [Brisken 2003] and include the geometry of the various structural elements of an EVLA antenna (sub-reflector, feed position, feed-legs, etc.). The FWHMs at these three frequencies are $\lambda/D = 20, 27, 41$ arcmin and this defines the field of view at each frequency.

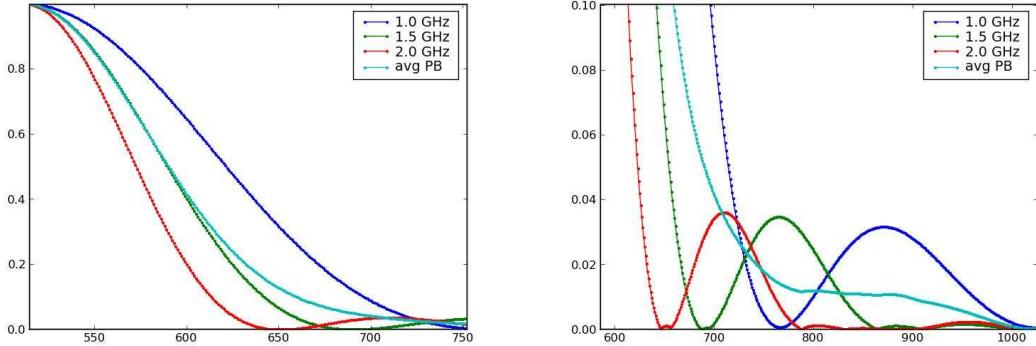


Figure 5.3: These plots show 1-dimensional cuts across the simulated EVLA primary beams shown in Fig.5.2 (chosen to pass through a sidelobe peak). The X axis is angular distance from the center of the beam (represented in units of image pixels, where each pixel represents 3 arcmin in this example, and the image center is at pixel 512), and the Y axis is the multiplicative gain associated with the primary beam at each angular distance. The blue, green and red lines represent the beam profiles at 1.0, 1.5 and 2.0 GHz, and the cyan line is a frequency-averaged profile along the same cut. The plot on the left covers the main lobe of the primary beams and shows that the half-power-beam-width at 1.0 GHz is near the null at 2.0 GHz. This indicates that the artificial spectral structure introduced by the primary beam can be very significant even before the HPBW at the lowest frequency in the beam. The plot on the right covers the first null and sidelobe at these three frequencies and shows that the first sidelobe at one frequency falls into the null at some other frequency. The average primary beam (shown in cyan) has an almost continuous sensitivity (at the few percent level) well out to the first sidelobe of the lowest frequency beam. This makes the instrument's sensitivity in the regions outside the main lobe a complicated function of frequency that cannot be ignored because the sensitivity in that region is high enough that artifacts from improperly deconvolved sources in that region will be visible above the noise level.

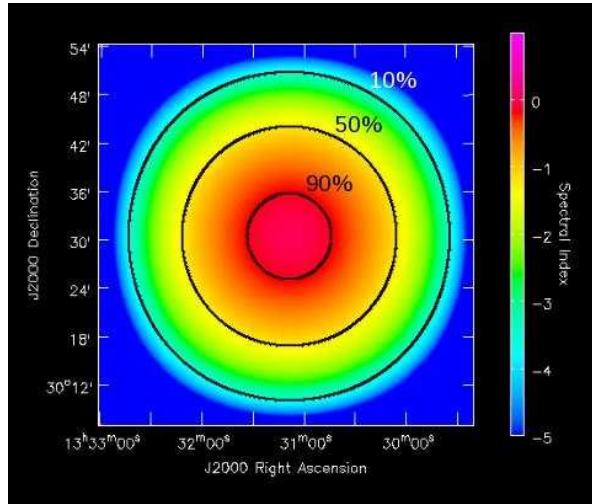


Figure 5.4: Spectral Index of the EVLA Primary Beam : This figure shows an overlay of the average spectral index (colour) and reference primary beam (contours) within the main lobe (*i.e.* it does not include sidelobes and therefore appears axisymmetric). The effective spectral index at the half-power point of the reference beam is about -1.4.

Spectral Index of the Primary Beam : The frequency dependence of the main lobe of the primary beam can be interpreted in terms of an instrumental spectral index. During measurement, the incident sky brightness at each observing frequency gets multiplied with the primary beam at that frequency. The measured spectral index therefore corresponds to the spectrum formed from the product of the sky and the primary beam. Figure 5.4 shows the spectral index and curvature due to the primary beam for an EVLA antenna (section 7.2.1 describes the relevant calculations). At the half-power point of the middle frequency, the effective spectral index computed for frequency-dependent EVLA primary beams is about -1.4. [Sault and Wieringa \[1994\]](#) analytically derive the same result by modeling the primary beam as a Gaussian whose width scales with frequency. A 1.0 Jy source of intrinsic spectral index of -1.0 will appear to have a flux of \sim 0.5 Jy and a spectral index of -2.4.

5.2 Comparison of Existing Wide-Band Imaging Methods

This section consists of a comparison between some existing and hybrid wide-band imaging methods applied to data simulated for the EVLA. The purpose of this study is to assess whether existing methods would suffice for the high-dynamic-range imaging requirements of the EVLA, and if not, to identify areas where improvements are required. The data product being evaluated is the continuum Stokes-I image and this study focuses on

dynamic-range limits due to errors in estimating spectral flux variations⁵. The techniques being compared are stacking of narrow-band images, MFS imaging, SW-MFCLEAN deconvolution and hybrids of single-channel and continuum imaging methods.

5.2.1 Existing and Hybrid Algorithms

5.2.1.1 Narrow-Band Imaging and Stacking (STACK)

A simple way to produce a continuum image is to construct a set of narrow-band images by deconvolving each frequency channel separately, and to then compute the sum of the single-channel images. This method does not depend on any spectral flux model and there will be no deconvolution errors due to spectral flux variation. However, the imaging fidelity and angular resolution of the image at each frequency is restricted to that given by the *uv*-coverage at that frequency. Non-linear deconvolutions on individual channel images can result in imaging artifacts that are neither coherent across the band (and are therefore not solvable) nor random (they do not average out), and this could introduce further artifacts when added together. Further, the imaging sensitivity is limited by the single-channel noise level (the continuum imaging sensitivity is related to the single-channel imaging sensitivity as $\sigma_{cont} = \sigma_{chan} / \sqrt{N_{chan}}$). Sources that are fainter than the single-channel noise may not be detected in the model images and will not be deconvolved from the final continuum image. The spectral structure of the source can be measured from the set of single channel maps, but the angular resolution of this computation is limited to that at the lowest frequency in the band because all images must be made at this lowest resolution before any spectral estimates can be derived.

5.2.1.2 Multi-Frequency-Synthesis Imaging (MFS)

A single continuum image is made by gridding visibilities from all frequency channels onto a single *uv*-grid and performing one deconvolution. This form of imaging benefits from the combined *uv*-coverage as well as the sensitivity gain of using data from all channels together. However, this algorithm assumes that all sources in the image have a flat spectrum across the frequency band. If the spectrum of a source is not flat, a direct combination of visibilities will generate spurious spatial structure at a level proportional to the magnitude of the spectral variation. However, if these errors are much lower than the continuum sensitivity limit, an accurate continuum reconstruction is still possible. One way to artificially improve this situation is to scale all channels according to an average spectral index (over all sources) before combined deconvolution in order to reduce visible deconvolution errors.

⁵ The contents of this section follow that in EVLA Memo 101 [[Urvashi R.V. et al. 2006](#)].

5.2.1.3 Sault-Wieringa Multi Frequency CLEAN (SW-MFCLEAN)

[Conway et al. \[1990\]](#) and [Sault and Wieringa \[1994\]](#) describe a multi-frequency deconvolution algorithm that models source spectra in terms of an amplitude and a slope, and performs a double or joint deconvolution to obtain a continuum map as well as an effective spectral index map. This algorithm is based on a matched-filtering idea and works for power-law spectra with a fixed spectral index across the band. Visible deconvolution errors result when the spectrum is not a pure power-law⁶.

5.2.1.4 Hybrids of Narrow-Band and Continuum Techniques

The basic idea of a hybrid wide-band method is to combine narrow-band channel imaging with one or more of the continuum methods in a multi-stage approach. The goal is to combine the advantages of single-channel imaging (simplicity and insensitivity to source spectra) with those of continuum imaging (deconvolution with full continuum sensitivity). However, if narrow-band single-channel imaging is one step in the process, there are a few restrictions that cannot be avoided. First, the angular resolution of the final images and spectral information will be restricted to that given by the lowest frequency. Also, these techniques will be suitable only for synthesis arrays with dense, single-channel uv -coverage where the collection of spatial frequencies measured at each channel is sufficient to reconstruct all the spatial structure there is at the angular resolution of that frequency. The hybrid methods that will be evaluated here are as follows.

1. **SW-MFCLEAN + STACK** : Apply the Sault-Wieringa MFClean algorithm to several successive subsets (chunks) of the full frequency range. The frequency or channel ranges to use for these subsets are based on *a priori* knowledge of the average spectral index of the sources being imaged and are chosen such that the spectrum within each chunk can be approximated by a linear spectrum.
2. **STACK + MFS with flattening** : Perform single-channel imaging to estimate the spectrum of all bright sources. Record this per-pixel spectrum, divide it out of the model image and predict flat-spectrum visibilities. Add in the residuals from the first stage, and perform MFS imaging on the now flat-spectrum data.
3. **STACK + MFS on residuals** : Perform single-channel imaging to deconvolve all bright sources stronger than the single-channel sensitivity limit. Remove the contribution of bright (spectrally varying) sources by subtracting out visibilities predicted from the model image cube. Perform MFS or SW-MFCLEAN on the continuum residuals.

⁶Section 6.2 describes a generalized version of this algorithm that accounts for variations of the spectral index across the band and lists differences with the Sault-Wieringa MF-CLEAN which works only with the first two terms of a Taylor expansion in ν .

Array	EVLA C array
Observing Band	LBand (1420MHz)
Total Bandwidth	320MHz (spread over 1280MHz)
Delta ν	40MHz
Frequency Resolution	10MHz
Frequency range	785MHz to 1985MHz
Reference Frequency	1420MHz
Number of channels	32
Cell size	2 arcsec
Image size	1024x1024 pixels
Image field of view	34 arcmin
Integration timestep	300 s
Total integration time	8 hours
Number of timesteps	8*3600/300=96
Noise per visibility	1.0 mJy
Theoretical RMS noise in per channel	$3.85 \times 10^{-6} \text{ Jy}$
Theoretical RMS noise in MFS image	$6.8 \times 10^{-7} \text{ Jy}$
Expected Dynamic Range	$0.1/6.8096 \times 10^{-7} = 1.468 \times 10^5$

Table 5.1: Data Simulation Parameters for Wide-Band Imaging Tests

5.2.2 Simulations and Results

Two wide-band datasets were simulated for these tests. The first was for a field of five point sources with amplitudes 100mJy, 10mJy, 1mJy, 100μJy and 10μJy. The single-channel noise was comparable to the flux of the weakest source. The 10mJy source was given a spectral index that varied between -0.5 and -1.5 between 1-2 GHz. The second dataset was simulated with continuum flux and spectral characteristics of a typical core,jet and hotspot. The brightest component (100mJy) had a flat spectrum, the 10mJy hotspot had a spectral index of -0.7, and the diffuse 'jet' had flux levels between 10μJy and 100μJy with spectral index varying between -0.1 and -0.5. Parameters used in the first two simulations are given in Table 5.2.2. They correspond to a T_{sys} (system temperature) of about 20K, and antenna and system efficiency (η_a and η_s) around 0.80⁷. For these simulations, the single-channel and continuum imaging sensitivities were calculated as described in Appendix A. The expected peak error (see section 6.2.4) due to a 10mJy source with spectral index α varying between -0.5 and -1.5 is $\sim 20 - 80 \mu\text{Jy}$ if the spectral structure is ignored, and $\sim 3 - 10 \mu\text{Jy}$ if the spectrum is modeled by a linear function.

⁷Only 32 channels were used because of data processing restrictions at the time of performing these tests. Any conclusions based on the difference between the single-channel and continuum sensitivity levels will still hold. Also, note that the target dynamic-range of $> 10^5$ has so far been achieved for a very small number of observed fields.

Figures 5.5 to 5.7 show the reconstructed image and the corresponding residual image for each algorithm, along with a set of measures to compare the relative accuracies of the algorithms. Separate estimates for on-source and off-source regions were computed using masks created by thresholding the known true image at a 2σ level. Only the inner quarter of each image was considered for CLEANing. The image fidelity was assessed by calculating the normalized χ^2 estimate between the known true image and the reconstruction. All results are based on automated runs of existing standard algorithms on simulated data. Carefully tuned deconvolution could in some cases result in better reconstructions.

Listed along with the results of each sample run are the following quantities.

1. Off source RMS : The achieved noise level in regions away from the true source.
2. Peak residual : The magnitude of the peak of the residual image. It represents the flux level of the minimum detectable/believable feature.
3. Dynamic Range (w.r.t. rms) : The ratio of the peak of the reconstructed image to the off-source rms. It represents the maximum dynamic-range achieved in the image.
4. Dynamic Range (w.r.t. peak residual) : The ratio of the peak of the reconstructed image to the peak residual. It represents the achieved dynamic-range w.r.t. believable features.

5.2.2.1 Conclusions from these tests

Existing and hybrid multi frequency synthesis algorithms were tested on simulated wide-band data, with the goal of determining how they perform against the requirement of $O(10^6)$ dynamic-range and $O(1\mu\text{Jy})$ image sensitivity. Tests were performed on data with point sources as well as extended flux components. The results were evaluated based on achieved rms levels as compared to the theoretical expected thermal noise, achieved dynamic-ranges as compared to those expected, the amount of large-scale deconvolution error, and image fidelity in terms of normalized χ^2 .

The main conclusions are :

1. Single-channel imaging and averaging is a simple algorithm that works independent of the form of spectral structure in the measurements, but often results in inaccurate reconstructions of extended emission, does not detect weak sources near the single-channel sensitivity limit and does not give noise-like continuum residuals. Also, all spectral information is limited to the angular resolution of the lowest frequency in the band.
2. Pure multi-frequency synthesis assuming a flat spectrum for all sources takes advantage of the combined *uv*-coverage and imaging sensitivity, but gives deconvolution artifacts around sources with a non-flat spectrum (roughly at the 10^3 dynamic-range for $\alpha = -1.0$ and 1GHz at L-Band).

3. The Sault-Wieringa Multi-Frequency CLEAN reaches target dynamic-ranges and image RMS levels for point sources with pure power-law spectra. Point sources with non-power-law spectra (α varies between -0.5 and -1.6 across 1GHz at L-Band) result in errors that are 10 times larger than the RMS noise level (denoted as 10σ). For extended sources with large-scale weak emission and non-power-law spectra, large-scale deconvolution artifacts appear at the 10σ level.
4. A hybrid method that combines single-channel imaging with a deconvolution on the continuum residuals is likely to produce accurate reconstructions and noise-like continuum residuals for synthesis arrays with dense *uv*-coverage per channel and well-behaved spectral noise characteristics. However, the second stage of combined deconvolution requires that the continuum residuals after spectral-line imaging, satisfy the convolution equation. Wide-band calibration errors, or deconvolution errors due to insufficient *uv*-coverage, can prevent the single-channel residuals from adding coherently to make the continuum residuals satisfy the convolution equation. Further, this method will not work for sparse synthesis arrays where the primary goal of wideband imaging is the increased *uv*-coverage. Also, the angular resolution of any spectral estimates is still restricted to that of the lowest frequency.

Therefore to improve these types of techniques, we need multi-scale methods that are able to model both spatial and spectral structure simultaneously. These methods must also use higher order terms in spectral series expansions to account for and accurately reconstruct non-power-law spectra. The frequency-dependence of the primary beam was not explicitly included in any of these tests, but the performance of these algorithms (and resulting dynamic-range limits) for wide-field imaging can be assessed by applying these results to the case where the sky spectrum is equal to that introduced by the instrument. Chapters 6 and 7 describe a multi-scale, multi-frequency deconvolution algorithm that reconstructs source spectral index and curvature in addition to total flux and also accounts for a frequency dependent primary beam.

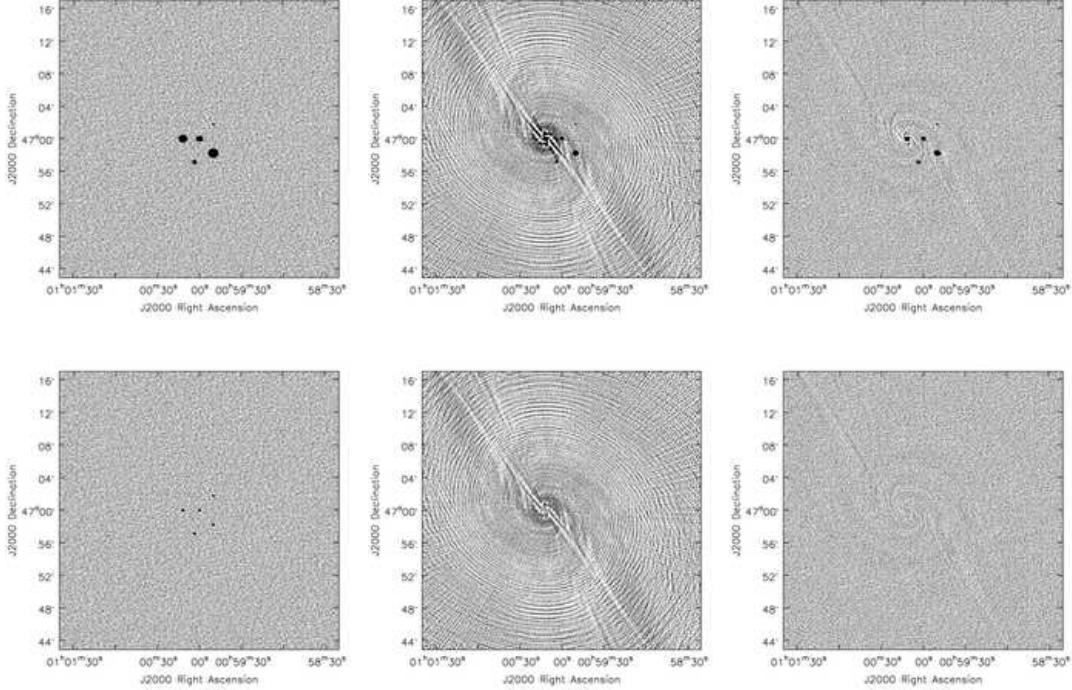


Figure 5.5: Standard Algorithms on Point Sources : This figure shows the restored images (top row) and residual images (bottom row) obtained by running the STACK (left), MFS (middle), SW-MFCLEAN (right) algorithms on a simulated dataset in which one point source has a spectral index that varies between 0.5 and 1.5 over the observing band. The STACK restored image shows relatively broadened components due to the varying spatial resolution for each channel. The residuals show traces of all sources, implying that the amplitudes and shapes of all the flux components have not been recovered well enough. There are no discernable deconvolution errors due to inaccurately modeled spectra, but the accuracy of the on-source flux is limited by the single-channel noise level. The MFS images show significant deconvolution errors around the source in question and a peak error of about $10\mu\text{Jy}$. The SW-MFCLEAN algorithm, which takes into account the first order beam, shows peak residuals at $\sim 8\mu\text{Jy}$, which are comparable to the level expected for the unaccounted-for second order beam, but higher because of the varying α across the band. The table below shows the RMS levels and dynamic-ranges achieved in these runs.

Point sources with spectral index varying between 0.5 and 1.6 for one source	Off-source RMS (Jy)	Peak residual (Jy)	Dynamic Range (w.r.t. rms)	Dynamic Range (w.r.t. peak residual)
Channel Averaging (STACK)	1.007e-06	2.164e-05	9.926e+04	4.621e+03
Bandwidth Synthesis (MFS)	1.849e-06	1.033e-05	5.408e+04	9.679e+03
Sault Algorithm (SW-MFCLEAN)	1.038e-06	9.607e-06	9.638e+04	1.041e+04

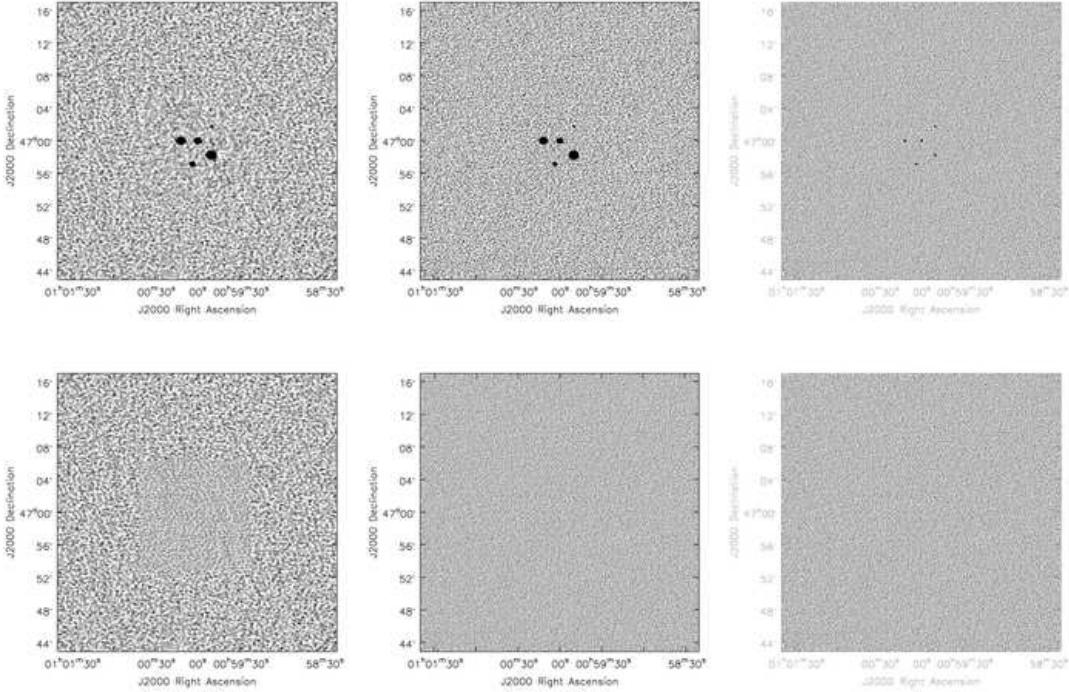


Figure 5.6: Hybrid Algorithms on Point Sources : This figure shows the restored images (top row) and residual images (bottom row) obtained by running three hybrid algorithms on the same dataset as in Fig.5.5. SW-MFCLEAN + STACK (left) : SW-MFCLEAN on 8-channel chunks followed by stacking resulted in high noise levels. STACK + MFS with flattening (middle) : Estimating spectrally varying flux from single-channel maps and flattening out the visibilities before doing a MFS left the weakest source un-detected. STACK + MFS on continuum residuals (right) : The top image shown here contains only the flux visible after the first stage of STACK imaging and subtraction (continuum residual image). The residual image shown below it is the result after the second step of MFS on the continuum SW-MFCLEAN was not required for the second stage (MFS sufficed) because at the end of the first stage, the peak flux was at the single-channel noise level of $\sim 4\mu\text{Jy}$, leading to a peak first-order beam sidelobe at $0.14\mu\text{Jy}$. This is lower than the theoretical continuum limit of $0.7\mu\text{Jy}$, and a flat-spectrum assumption would not lead to visible errors.

Point sources with spectral index varying between 0.5 and 1.6 for one source	Off-source RMS (Jy)	Peak residual (Jy)	Dynamic Range (w.r.t. rms)	Dynamic Range (w.r.t. peak residual)
SW-MFCLEAN + STACK	1.705e-06	8.607e-06	5.865e+04	1.161e+04
STACK + MFS with flattening	1.130e-06	5.733e-06	8.849e+04	1.744e+04
STACK + MFS on residuals	1.128e-06	5.829e-06	8.865e+04	1.715e+04

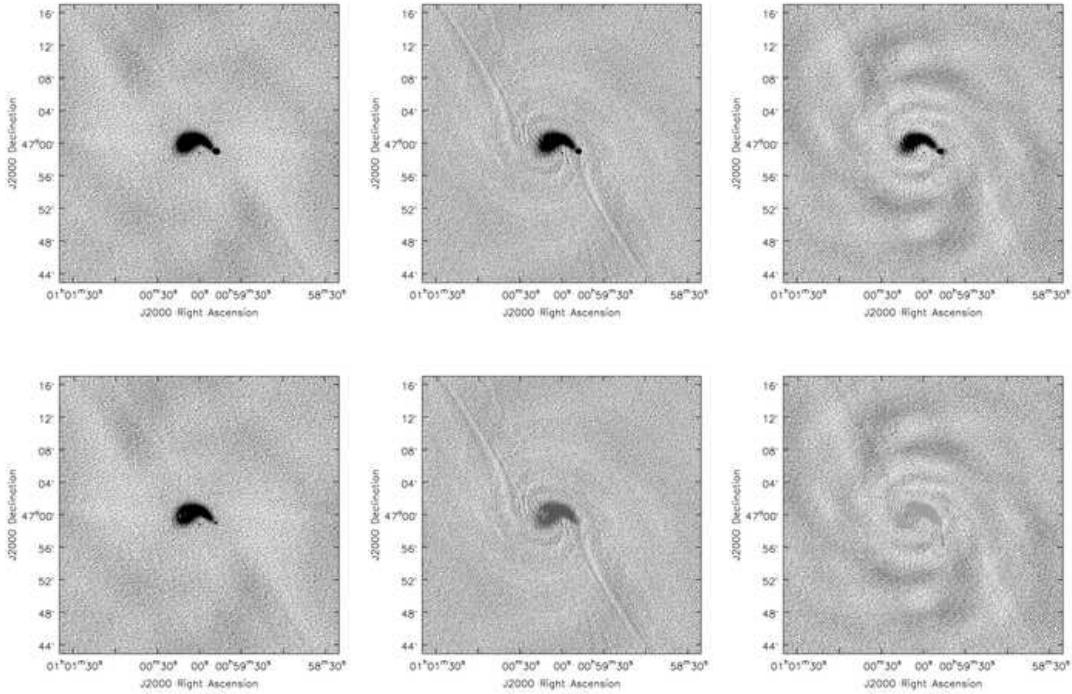


Figure 5.7: Standard Algorithms on Extended emission : This figure shows the restored images (top row) and residual images (bottom row) obtained by running the STACK (left), MFS (middle), SW-MFCLEAN (right) algorithms on a simulated dataset with an extended source whose spectrum varies across the source, for total frequency range of 640MHz. The STACK image shows low-level large-scale deconvolution errors arising from the limiting single channel sensitivity. The MFS algorithm produced more accurate on-source flux reconstruction with better large-scale deconvolution results. It shows errors primarily due to the spectrally varying flux in the hotspot. The SW-MFCLEAN algorithm was able to model a power-law component of the spectrally varying source, and reach a lower residual rms, but low-level large-scale deconvolution errors remain at the $10\mu\text{Jy}$ level. None of the algorithms reached the theoretical thermal noise.

Extended Core-Jet type source with spectral index between -0.1 and -0.7	Off-source RMS (Jy)	Peak residual (Jy)	Dynamic Range (w.r.t. rms)	Dynamic Range (w.r.t. peak residual)
Channel Averaging (STACK)	1.445e-06	2.142e-05	6.920e+04	4.6683e+03
Bandwidth Synthesis (MFS)	1.206e-06	6.041e-06	8.291e+04	1.655e+04
Sault Algorithm (SW-MFCLEAN)	1.233e-06	1.214e-05	8.110e+04	8.237e+03

5.2.3 Continuum imaging with dense uv -coverage

This section describes the application of the STACK+MFS algorithm to a simulated EVLA data set in which the single-frequency uv -coverage is sufficient to unambiguously reconstruct the spatial structure of the source. The goal of this test is to show that when the target science does not require spectral information at high angular resolution, wide-band imaging with data from synthesis arrays like the EVLA with very dense uv -coverage may require only a simple adaptation of existing and standard deconvolution algorithms. This test used a simulated data set and contained no calibration errors.

Simulation : Data were simulated for the EVLA C-configuration with 40 frequency channels spread between 1 and 4 GHz. The noise per visibility was 10 mJy, giving a theoretical point-source single-channel sensitivity of 50 μ Jy and continuum sensitivity of 8 μ Jy. The wide-band sky brightness distribution in this simulation was obtained by linear interpolation between 1.4 and 4.8 GHz maps of Cygnus-A [Carilli et al. 1991]. At each frequency, the source brightness was further modified to amplify the dynamic-range. The brightness at each pixel in the true-sky cube I was replaced by $I^{2.35}$ (*i.e.* the amplitude per pixel was raised to the power of 2.35) to increase the dynamic-range of the sky brightness distribution being simulated.

Imaging Results : These data were imaged using a hybrid method that combined single-channel imaging (STACK) with standard MFS. MS-CLEAN was used to deconvolve each channel separately down only to the single-channel sensitivity limit σ_{chan} . As a second step, standard MFS and MS-CLEAN was applied to the continuum residual image and iterations were terminated using a flux threshold given by $\sigma_{chan}/\sqrt{N_{chan}}$.

Figure 5.8 shows the imaging results. The image on the left shows that after only narrow-band imaging on all channels, there is undeconvolved emission that is undetected at the single-channel sensitivity level. The image on the right shows the final image after the deconvolution on the combined residuals and the deconvolution errors are markedly reduced. The achieved off-source noise levels were an order of magnitude higher than theoretical. A maximum dynamic-range of 530,000 was achieved (peak/off-source-rms) and the on-source dynamic-range was 40,000 (peak/on-source-rms).

When does this second stage work ? At the end of the first stage, the only undeconvolved flux comes from flat-spectrum residuals of all sources brighter than σ_{chan} as well as all sources weaker than σ_{chan} . Weak sources whose flux values lie below σ_{chan} but above $\sigma_{cont} = \sigma_{chan}/\sqrt{N_{chan}}$ are detected only in the combined deconvolution stage when MFS imaging is performed. If these weak sources have spectral structure, the flat-spectrum assumption of MFS imaging will lead to deconvolution errors a few orders of magnitude (*e.g.* a factor of 10^3 , see section 6.2.4.2) smaller than the current peak flux (σ_{chan}). Such errors

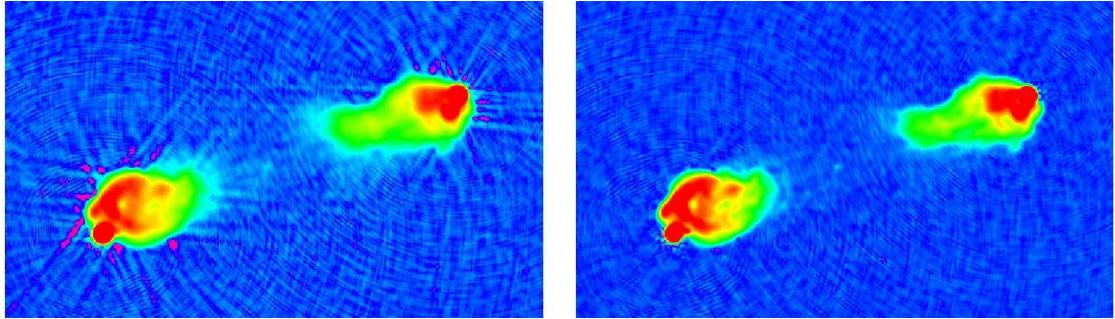


Figure 5.8: The hybrid of single-channel imaging (STACK) and MFS imaging on the continuum residuals was applied to data simulated for Cygnus-A (EVLA, C-array, 1-4 GHz). The left panel shows the result after the first stage (only spectral-line imaging) and shows that there is significant undeconvolved emission that was undetected at the single-channel sensitivity level. The image on the right shows the final image after the deconvolution on the combined residuals and shows significantly reduced deconvolution errors. Therefore, with sufficient uv -coverage, if the single-channel deconvolution is limited only by the single-channel sensitivity level, the residuals will add coherently such that the continuum residual image will satisfy a convolution equation (a sky model convolved with the PSF in Eqn. 5.4), and the second stage will be able to reach continuum sensitivity levels.

are likely to be below σ_{chan} because \sqrt{N} will almost always be less than 10^3 . This condition ($N_{chan} < 10^6$) can always be satisfied, because even though data may be observed at a very high spectral resolution (large N_{chan}) it can always be averaged down to the bandwidth smearing limit for the highest sampled frequency to reduce the number of channels during imaging.

This method can be used only to construct an image of the continuum flux. Only if there is sufficient single-channel uv -coverage to reconstruct an accurate model of the source structure (for example, fields of isolated point sources), spectral information may also be derived from such an approach. This idea has been tested on EVLA simulations with dense single-frequency uv -coverage as well as wide-band VLA data with relatively sparse uv -coverage at each frequency (see section 8.3.1), but it is yet to be verified on real EVLA wide-band data with real calibration errors.

CHAPTER 6

DECONVOLUTION WITH IMAGES PARAMETERIZED AS A SERIES EXPANSION

The general theme of this chapter is the description of the sky brightness distribution as a linear combination of images and using this model within an iterative CLEAN-based deconvolution framework. Most of the imaging methods described in Chapters 3 and 4 parameterize the sky brightness distribution as a single list of pixel amplitudes and assume that source structure and instrumental effects are constant across the entire bandwidth of data being imaged. This chapter relaxes these assumptions and describes how the added complexity and increased dimensionality of the parameter space can be folded into the standard measurement and imaging equations. In particular, Section 6.1 derives a multi-scale deconvolution method by describing an image as a linear combination of images at different spatial scales. Section 6.2 derives a multi-frequency deconvolution method by describing the spectral shape of the brightness distribution by a Taylor polynomial (a partial sum of a Taylor series). Pseudo-code listings of these algorithms (3 and 4) are shown at the end of each section. Chapter 7 later describes a multi-scale, multi-frequency deconvolution algorithm as a combination of the above ideas, and shows how a multi-frequency parameterization of the antenna primary beam can be folded into the same framework.

The algorithms described in this chapter follow the format used in Section 3.2.1. First, each pixel of an image model is defined as a linear combination of parameters and basis functions. Then, the imaging equations are derived by applying the interferometric measurement equation to each term in this linear series. The resulting normal equations are then described along with diagrams similar to Figure 3.2 to illustrate the image-domain effect of the measurement and modeling process, and to give a qualitative view of what is being solved (*i.e.* the form of the Hessian matrix to be inverted, and the vector to which this inverse is applied). The solution process is then described in two stages, the principal solution and iterative joint deconvolution. The principal solution involves only diagonal approximations of the matrices to be inverted and in the ideal case where the PSF is a δ -function this diagonal approximation will deliver the solution of the full system. The joint deconvolution is an iterative process similar to the CLEAN algorithm, but which simultaneously builds up solutions for all the coefficients in the linear series. Finally, these best-fit estimates for the pixel-based coefficients are converted into quantities that can be interpreted physically. A block matrix notation is used throughout this chapter. For reference, a generic description of weighted linear least-squares in block matrix form is described in Appendix B.

6.1 Multi-Scale Deconvolution

Images of astrophysical objects tend to show complex structure at different spatial scales. An image parameterization that works with independent pixels is ideal for the deconvolution of fields of isolated point-like sources that are smaller than the instrument’s angular resolution, but tends to break extended emission into a collection of compact sources. This often results in a physically inaccurate representation of the sky. However, such a reconstruction may be indistinguishable from the real sky because of the non-empty null space of the measurement matrix (unmeasured spatial frequencies) in which the model is unconstrained by the data. It therefore becomes important to provide *a priori* constraints on what the sky emission should look like. One way to naturally achieve this for emission with structure on multiple spatial scales is to parameterize the image in a scale-sensitive basis that spans the full range of scale sizes measured by the instrument. This forces pixel-to-pixel correlations during the reconstruction and provides a strong constraint on the reconstruction of visibilities in the null space of the measurement matrix. Also, when the peak amplitude of extended emission is close to the image noise level, spatial correlation length fundamentally separates signal from noise and scale-sensitive deconvolution algorithms generally give more noise-like residuals on large scales [Bhatnagar and Cornwell 2004].

Section 6.1.1 defines a multi-scale image model. Section 6.1.2 describes the normal equations that result from folding this model into the standard imaging equations and then describes an algorithm that reconstructs the sky brightness distribution at a range of spatial scales and combines the results to form the complete, multi-scale image. This discussion is a formal derivation of the CH-MSCLEAN (Cornwell-Holdaway MS-CLEAN) technique described in [Cornwell 2008], but describes a modified version that improves upon the existing algorithm. Section 6.1.3 lists the similarities and differences among this algorithm, the CH-MSCLEAN algorithm implemented in CASA and ASKAPsoft, and a matched filtering algorithm implemented in AIPS [Greisen et al. 2009]. Section 6.1.4 contains an example of the multi-scale series coefficients derived during deconvolution, emphasizes sources of uncertainty in this calculation and lists the algorithmic steps required to converge towards a stable solution.

6.1.1 Multi-Scale Image model

Let us represent the sky brightness distribution as a linear combination of images at different spatial scales. The image at each spatial scale s is written as a convolution between a set of δ -functions $\vec{I}_s^{sky,\delta}$ and a scale function \vec{I}_s^{shp} . The scale functions can be any set of 2D functions that represent structure at varying spatial scales, and Cornwell [2008] choose a set of tapered, truncated parabolas of different widths (proportional to s). The amplitude of each δ -function in $\vec{I}_s^{sky,\delta}$ represents the integrated amplitude of an extended flux component of scale size s , centered at the location of the δ -function. Figure 6.1 shows an example of this multi-scale representation.

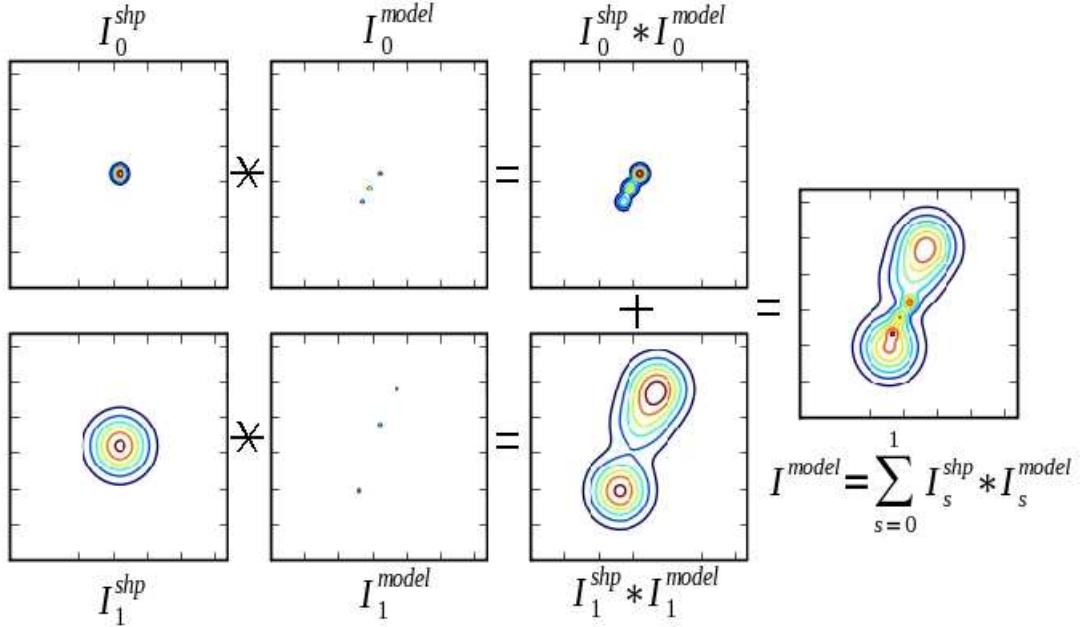


Figure 6.1: This figure shows the multi-scale representation of an image composed of two distinct spatial scales. The left column shows two scale basis functions \vec{I}_0^{shp} and \vec{I}_1^{shp} that represent symmetric flux components at two different spatial scales, normalized to unit area. The second column from the left shows model images $\vec{I}_0^{sky,\delta}$, $\vec{I}_1^{sky,\delta}$ with δ -functions that mark the total-flux and locations of flux components of corresponding spatial scale. The third column shows the resulting image at the two spatial scales, and the image on the right (\vec{I}^{sky}) shows the multiscale image formed from the sum of images at multiple spatial scales. The goal of a multi-scale deconvolution algorithm is to use the pre-defined set of scale basis functions shown in the first column, to extract the δ -function flux components shown in the second column, from visibilities measured for \vec{I}^{sky} .

For a finite set of N_s spatial scales, the multi-scale image model is written as follows.

$$\vec{I}^{model} = \sum_{s=0}^{N_s-1} \vec{I}_s^{shp} \star \vec{I}_s^{sky,\delta} \quad (6.1)$$

where $\vec{I}_s^{sky,\delta}$ are per pixel coefficients and \vec{I}_s^{shp} are the basis functions of this linear series. In order to always allow for the modeling of unresolved sources, we choose the first scale function $\vec{I}_{s=0}^{shp}$ to be a δ -function. Successive basis functions then correspond to inverted parabolas of larger widths (as s increases). Note that a choice of $N_s = 1$ reduces all the equations in this section to those in Chapter 3 where the image is parameterized using a set of δ -functions.

6.1.2 Imaging Equations and Block Deconvolution

This section contains a derivation of the normal equations for a multi-scale image model, followed by a description of the principal solution and its use in an iterative deconvolution process. The derivations in this section use a block-matrix notation (described in Appendix B) to represent the measurement and normal equations.

6.1.2.1 Measurement equations

An interferometer samples the visibility function of \vec{I}^{model} given by Eqn. 6.1. The measurement equation (similar to Eqn. 3.7) for a multi-scale representation of the sky brightness becomes

$$\vec{V}^{corr} = [SF]\vec{I}^{model} = \sum_{s=0}^{N_s-1} [SF](\vec{I}_s^{shp} \star \vec{I}_s^{sky,\delta}) = \sum_{s=0}^{N_s-1} [SF][F^\dagger T_s F]\vec{I}_s^{sky,\delta} = \sum_{s=0}^{N_s-1} [ST_s F]\vec{I}_s^{sky,\delta} \quad (6.2)$$

where $\vec{V}_{n \times 1}^{corr}$ is a list of n calibrated visibilities, $[S_{n \times m}]$ is the sampling matrix, $[F_{m \times m}]$ is the Fourier transform operator (image to spatial-frequency) and all images $\vec{I}_{m \times 1}$ are lists of m pixel amplitudes.

Each scale function is denoted by the subscript s and $[T_s] = diag(\vec{T}^s)$ is a diagonal matrix containing a spatial frequency taper function given by $\vec{T}^s = [F]\vec{I}_s^{shp}$. This taper $[T_s]$ is similar to a uv -taper described in Section 3.2.3.1. It gives lower spatial frequencies a higher weight compared to higher spatial frequencies and has the effect of tuning the sensitivity of the instrument to peak for a scale larger than the angular resolution of the telescope. The operator $[F^\dagger T_s F]$ is an image-domain convolution operator with \vec{I}_s^{shp} as its kernel (see footnote 19 on page 24 for the definition of a convolution operator).

When the sky brightness is written as the sum of images at multiple spatial scales, the full measurement matrix ($[A]$ in Eqn. 2.20) can be written in block matrix form with a horizontal stack of N_s blocks each of shape $n \times m$ each, and a vertical stack of N_s vectors of image pixels each of length m . This $n \times mN_s$ measurement matrix operates on the $mN_s \times 1$ column vector of image pixels to produce n visibilities.

An example for $N_s = 2$ is shown below.

$$\begin{bmatrix} [ST_0 F] & [ST_1 F] \end{bmatrix} \begin{bmatrix} \vec{I}_0^{sky,\delta} \\ \vec{I}_1^{sky,\delta} \end{bmatrix} = \vec{V}^{corr} \quad (6.3)$$

The column vector containing the image model is the equivalent of the second column of images (from the left) in Figure 6.1. The δ -functions in $\vec{I}_p^{sky,\delta}$ represent the total flux and location of flux components at the spatial scale denoted by p (see Eqn. 6.1).

6.1.2.2 Normal equations

A least-squares solution of Eqn. 6.2 can be obtained by forming and solving the following normal equations.

$$[H_{mN_s \times mN_s}^{ms}] \vec{I}_{mN_s \times 1}^{\delta sky, \delta} = \vec{I}_{mN_s \times 1}^{\text{dirty}, ms} \quad (6.4)$$

The Hessian $[H^{ms}]$ can be written in block-matrix form with $N_s \times N_s$ blocks of size $m \times m$ each, and the sky model $\vec{I}^{\delta sky, \delta}$ and dirty image $\vec{I}^{\text{dirty}, ms}$ as sets of N_s image vectors of size $m \times 1$ (Appendix B describes this block-matrix notation).

For example, the normal equations for $N_s = 2$ can be written in block-matrix form as

$$\begin{bmatrix} [H_{s=0, p=0}] & [H_{s=0, p=1}] \\ [H_{s=1, p=0}] & [H_{s=1, p=1}] \end{bmatrix} \begin{bmatrix} \vec{I}_{p=0}^{\delta sky, \delta} \\ \vec{I}_{p=1}^{\delta sky, \delta} \end{bmatrix} = \begin{bmatrix} \vec{I}_{s=0}^{\text{dirty}} \\ \vec{I}_{s=1}^{\text{dirty}} \end{bmatrix} \quad (6.5)$$

where the indices s, p vary from 0 to $N_s - 1$ and will henceforth denote block row and column indices for multi-scale equations.

Figures 6.2 and 6.3 are two pictorial representations of the normal equations for a multi-scale sky brightness distribution. Figure 6.2 shows the standard normal equations (similar to Figure 3.2), and Figure 6.3 depicts the normal equations shown in Eqn. 6.5 (in block matrix form), labeled as shown in Eqn. 6.4.

These block-matrix equations can be written row-by-row as follows.

$$\sum_{p=0}^{N_s-1} [H_{s,p}] \vec{I}_p^{\delta sky, \delta} = \vec{I}_s^{\text{dirty}} \quad \forall s \in \{0, \dots, N_s - 1\} \quad (6.6)$$

$$\text{where } [H_{s,p}] = [F^\dagger T_s F][B][F^\dagger T_p F] \quad (6.7)$$

$$\vec{I}_s^{\text{dirty}} = [F^\dagger T_s F] \vec{I}_s^{\text{dirty}} \quad (6.8)$$

$[B]$ is the Beam matrix ($[H]$ in Eqn. 3.9) and \vec{I}^{dirty} is the standard dirty image (Eqn. 3.10). $[B]$ is a convolution operator with the PSF \vec{I}^{psf} (Eqn. 3.11) as its kernel, and the operators $[F^\dagger T_s F]$ and $[F^\dagger T_p F]$ implement image-domain convolutions with scale functions $\vec{I}_s^{\delta hp}$ and $\vec{I}_p^{\delta hp}$. Therefore, each Hessian block $[H_{s,p}]$ is a new convolution operator¹ whose kernel will be denoted as $\vec{I}_{s,p}^{psf}$ ².

$$\vec{I}_{s,p}^{psf} = \vec{I}_s^{\delta hp} \star \vec{I}_p^{psf} \star \vec{I}_p^{\delta hp} \quad (6.9)$$

¹ Convolution is associative and commutative. A sequence of convolutions can be written as a single convolution with a kernel given by the convolution of all the individual kernel functions in the sequence.

² 1-D examples of the convolution kernels $\vec{I}_{s,p}^{psf}$ are shown in Figure 6.3 as the shifted rows in each Hessian block. The convolution kernels from the top row of blocks of the Hessian matrix (given by $\vec{I}_{s=0,p}^{psf} \forall p \in \{0, \dots, N_s - 1\}$) represent the instrument's responses to flux components of unit integrated flux and shape given by $\vec{I}_p^{\delta hp}$. These N_s functions are called scale-PSFs [Cornwell 2008] and represent the image-domain patterns being matched to the dirty image \vec{I}^{dirty} .

Similarly, each dirty image \vec{I}_s^{dirty} (in Eqn. 6.8) can be written as the result of convolving \vec{I}^{dirty} with the scale function \vec{I}_s^{shp} (smoothing the standard dirty image to various spatial scales).

$$\vec{I}_s^{dirty} = \vec{I}_s^{shp} \star \vec{I}^{dirty} \quad (6.10)$$

This is a matched-filtering³ operation that detects the best-matching spatial scales for every location in the image. For sources with equal total flux but different spatial scales, peaks in the smoothed dirty images correspond to the location of a source whose scale size best matches the spatial scale that it was smoothed with (Figure 6.3 demonstrates this).

The normal equations in Eqn. 6.6 can now be re-written in terms of the above convolution kernels and image vectors (Eqns. 6.9, 6.10).

$$\sum_{p=0}^{N_s-1} \vec{I}_{s,p}^{psf} \star \vec{I}_p^{sky,\delta} = \vec{I}_s^{shp} \star \vec{I}^{dirty} \quad \forall s \in \{0, \dots, N_s - 1\} \quad (6.11)$$

The purpose of this step is to show that the multi-scale dirty images are results of sums of convolutions, instead of one single convolution (Eqn. 3.12). The process of solving these normal equations is therefore referred to as a joint deconvolution that simultaneously estimates model images at all N_s spatial scales.

In order to compare this multi-scale representation with standard imaging as described in Section 3.2.1, Figure 6.2 shows a pictorial representation of the standard normal equations (similar to Figure 3.2) for a sky consisting of one δ -function (point source) and one Gaussian, both with equal total power but different spatial scales. The dirty image on the RHS is the standard dirty image and it peaks at the location of the point source, but the extended component is at the same level as the sidelobes and therefore hard to detect.

Figure 6.3 is a pictorial representation of the multi-scale normal equations for $N_s = 2$ (as written in Eqns. 6.4 and 6.5) for the same sky image used in Figure 6.2. The two scale basis functions used in this example are exactly matched to the point source and Gaussian present in the sky brightness distribution. The two model images contain one δ -function each, to mark the location and total flux for one point source and one Gaussian component. The Hessian is composed of a set of convolution operators and the dirty images on the RHS are smoothed versions of the \vec{I}^{dirty} . The first scale basis function I_0^{shp} is a δ -function, and therefore the first RHS vector \vec{I}_0^{dirty} is identical to the standard dirty image (RHS of Figure 6.2) and has a peak at the location of the point source. The second RHS

³Matched filtering is a technique used to detect the presence of a signal of some known form within a measured signal of arbitrary form. This is done by convolving the measured signal with known templates, and picking out the template that gives the highest value after convolution. This template is then said to be best matched to the data. In signal and image processing this convolution is usually implemented as a multiplication in the Fourier domain, or in other words, as a filter. In our case, matched-filtering with the scale function \vec{I}_s^{shp} in the image-domain is equivalent to using a uv -taper function $\vec{T}_s = [F]\vec{I}_s^{shp}$ in the spatial-frequency domain. This is equivalent to tuning the instrument's sensitivity to peak for the spatial scale s .

$$\begin{bmatrix} [B_{m \times m}] \\ \vdots \\ \text{[Three noisy signal traces]} \end{bmatrix} \left(\begin{bmatrix} \vec{I}_{m \times 1}^{sky} \\ \vdots \\ \text{[Two spikes]} \end{bmatrix} \right) = \left(\begin{bmatrix} \vec{I}_{m \times 1}^{dirty} \\ \vdots \\ \text{[Smoothed dirty image]} \end{bmatrix} \right)$$

Figure 6.2: Normal Equations for a Multi-Scale Sky Brightness Distribution : This diagram represents the standard process of image formation with an interferometer when the sky brightness distribution has structure at multiple spatial scales. In this example, the true sky model consists of two flux components of equal total power but different spatial scales (one δ -function and one Gaussian, each of unit integrated flux). This diagram represents Eqn. 3.12 and uses the same Beam matrix $[B]$ as in Figure 3.2 (displayed using fewer rows). The dirty image vector (on the RHS) shows the point source clearly, but the Gaussian component of equal total flux is almost masked by the sidelobes of the point-spread-function.

vector \vec{I}_1^{dirty} is the standard dirty image smoothed by I_1^{shp} . The peak in this smoothed dirty image is at the location of the flux component of matching scale (*i.e.* at the location of the Gaussian flux component). This is a demonstration of matched-filtering.

Although these peaks mark the locations of flux components of matching scale, the flux values measured from the smoothed dirty images do not yet represent the total flux of the component (as would be desired to construct a set of model δ -functions $\vec{I}^{sky,\delta}$). The next section (6.1.2.3) describes how an approximate inversion of the Hessian can be used to calculate accurate total flux estimates for components at each spatial scale.

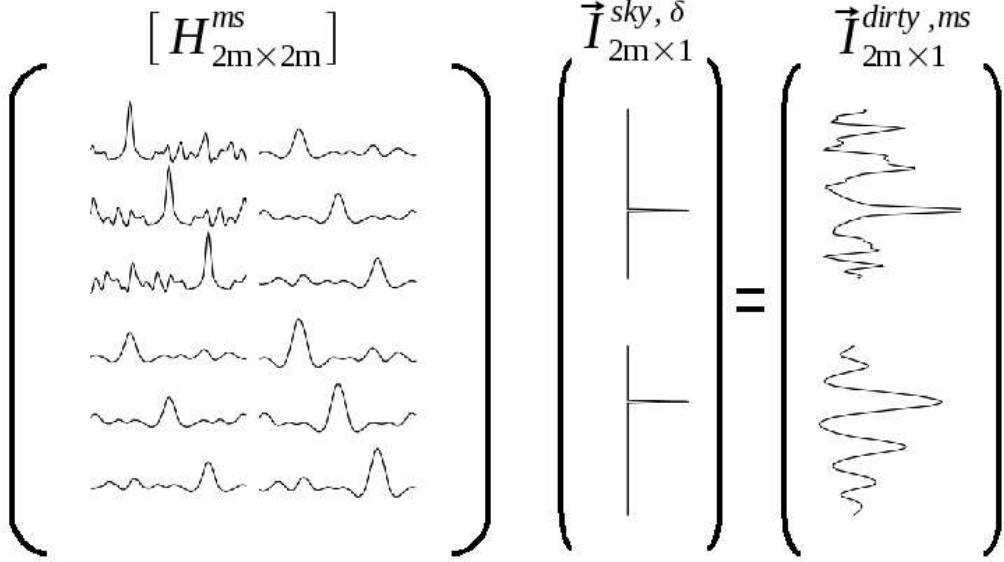


Figure 6.3: Normal Equations for Multi-Scale Deconvolution : This diagram is a pictorial representation of the normal equations formed when the sky brightness is described as the sum of images at multiple spatial scales (Eqns. 6.4 and 6.5 with $N_s = 2$). The sky model is a pair of image vectors containing δ -functions whose amplitudes represent the total flux of components centered at their locations (see Eqn. 6.1). The sky brightness in this example is the same as that in Figure 6.2, and consists of two flux components of equal total flux but different spatial scales. The basis functions used to represent the components are $\vec{I}_0^{shp} = \delta$ -function and $\vec{I}_1^{shp} = \text{Gaussian}$ whose scale matches the broad flux component in \vec{I}^{sky} (in Figure 6.2). The Hessian is a 2×2 block matrix, and each block (of size $m \times m$) is a convolution operator whose kernel is constructed from these basis functions (Eqn. 6.9). The RHS vectors are computed by smoothing the dirty image to different spatial scales (see Eqn. 6.11 and Fig. 6.1). The top RHS vector is the dirty image \vec{I}^{dirty} (unchanged by a convolution with a δ -function, and also the same as the RHS of Figure 6.2) and the bottom RHS vector is $\vec{I}^{dirty} \star \vec{I}_1^{shp}$. For sources with equal total flux (note the same height of the δ -functions in $\vec{I}_p^{sky, \delta}$), peaks in these smoothed dirty images correspond to the source whose scale-size best matches the spatial scale that it was smoothed with. This is a demonstration of matched filtering. Working in block-matrix form, the multi-scale dirty images (RHS) can be written as the result of linear combinations of convolutions. The multi-scale model images $\vec{I}^{sky, \delta}$ can be reconstructed via a combination of deconvolution and a block-inversion of the Hessian matrix.

(A few points can be noted about the Hessian matrix. Note that the top-left Hessian block $[H_{s=0,p=0}] = [B]$ is the Beam matrix, and the bottom-right block $[H_{s=1,p=1}]$ contains smoothed versions of the $[H_{s=0,p=0}]$ kernel. The off-diagonal blocks have smaller peaks than the diagonal blocks indicating that although the scale basis functions are coupled (non-orthogonal), the matrix of Hessian peaks $[H^{peak}]$ is well-conditioned.)

6.1.2.3 Principal Solution

When the flux model is a linear combination of images and the normal equations are written in block matrix form, we define the principal solution as the pseudo-inverse solution obtained using diagonal approximations of each Hessian block. This definition is a natural extension of the principal solution defined in Section 3.2.1.6 for standard imaging in which the principal solution is computed by inverting a diagonal approximation of the Beam matrix and the values measured at the peaks of the principal solution images (for isolated sources) are the true sky values as represented in the image model. For the system shown in Figure 6.2, the peaks represent the true sky brightness at each pixel. For the system shown in Figure 6.3, the peaks represent the total flux of a component of a certain spatial scale, centered at each pixel, and not the sky brightness measured at each pixel.

When each Hessian block is approximated by a diagonal matrix, all pixels can be treated independently and the principal solution can be computed one pixel at a time. Since each block $[H_{s,p}]$ is a convolution operator, all the diagonal elements per block are identical and given by $\text{mid}\{\vec{I}_{s,p}^{psf}\}$. Therefore, one single $N_s \times N_s$ element matrix (denoted as $[H_{N_s \times N_s}^{\text{peak}}]$) can be used to approximate the Hessian for all pixels. The N_s dirty images (RHS of Eqn. 6.6) can be written one pixel at a time by extracting one pixel from each image and forming a smaller vector (denoted by $I_{N_s \times 1}^{\text{pix,dirty}}$). The principal solution is obtained as follows, one pixel at a time (multi-scale equivalent of Eqn. 3.13).

$$I_{N_s \times 1}^{\text{pix,psol}} = [H_{N_s \times N_s}^{\text{peak}}]^{-1} I_{N_s \times 1}^{\text{pix,dirty}} \quad \text{for each pixel} \quad (6.12)$$

The values in $I_{N_s \times 1}^{\text{pix,psol}}$ are then filled back into the N_s model image vectors, also one pixel at a time. For an imaging instrument whose PSF is a δ -function, the principal solution gives the final image. When there is incomplete sampling, this inversion is valid only at the peaks of sources, and can be used only to measure the total flux of a flux component to be subtracted out during an iterative deconvolution. Section 6.1.4 and Figure 6.4 show the results of applying $[H^{\text{peak}}]^{-1}$ to $\vec{I}^{\text{pix,dirty}}$ for all pixels for a simulated example with three spatial scales ($N_s = 3$), and suggest heuristics to pick out only valid solutions.

In practice, this principal solution is used as follows. Since we cannot directly invert the Hessian to solve the normal equations, we separate this process into two steps. First, the principal solution is computed to get an estimate of the total flux per component, and then its contribution is subtracted out of all the RHS vectors.

6.1.2.4 Properties of $[H^{\text{peak}}]$

Some properties of $[H^{\text{peak}}]$ for multi-scale imaging and their implications are given below.

1. Each element of $[H^{\text{peak}}]$ represents the sum of uv -tapered gridded imaging weights and is given as follows.

$$H_{s,p}^{\text{peak}} = \text{mid}\{\vec{I}_{s,p}^{psf}\} = \text{tr}([T_s][S^\dagger WS][T_p]) \quad \forall s, p \in \{0 \dots N_s - 1\} \quad (6.13)$$

2. The elements on the diagonal of $[H^{peak}]$ correspond to $s = p$ and are a measure of the sensitivity of the instrument to a particular spatial scale. Note that the element $H_{0,0}^{peak}$ is the same as the peak of the PSF in the standard Beam matrix (assuming that I_0^{shp} is a δ -function). With uniform weighting, the spatial PSFs on the diagonal blocks are the autocorrelations of the regular PSFs at different spatial scales, and this measures the area under the main beam of the PSF for each spatial scale⁴.
3. The off-diagonal elements given by $s \neq p$ are a measure of the orthogonality⁵ of the basis set, for the given uv -coverage and weighting scheme. They measure the amount of overlap between basis functions in the measurement domain. Smaller values indicate a more orthogonal set of basis functions, and the instrument is better able to distinguish between the chosen spatial scales. For our multi-scale basis set, there will always be some overlap between the different uv -taper functions and this set will never be orthogonal. Therefore it becomes important to choose a suitable set of spatial scales, such that $[H^{peak}]$ is reliably invertible. The condition number of this per-pixel Hessian can be used as an estimate of how robust a solution will be, and can be used as a metric to select a suitable basis set of scale functions.
4. By choosing a set of spatial scales within the range the instrument is sensitive to, $[H^{peak}]$ will be a positive-definite symmetric matrix whose inverse can be easily computed *via* a Cholesky decomposition⁶. Also, the value of N_s is usually < 10 , making the inversion of $[H^{peak}]$ tractable.

6.1.2.5 Iterative Block Deconvolution (MSCLEAN algorithm)

This section describes the process of reconstructing a multi-scale image of the sky brightness using a CLEAN-based deconvolution algorithm. This description follows the same format as that of the CLEAN algorithm in Chapter 3 where the principal solution is used to produce solution estimates that then get refined *via* a steepest-descent optimization. Algorithm 3 lists the multi-scale deconvolution method described in this section.

⁴A diagonal approximation of $[H^{peak}]$ can be inverted and applied to the RHS dirty image vectors to normalize them by the area under the main beam of the PSF at different spatial scales. This is related to the scale-bias terms used in the multi-scale techniques described in Cornwell [2008] and Greisen et al. [2009] (see Section 6.1.3).

⁵The following definition of orthogonality is used here. Two vectors are orthogonal if their inner product is zero. The orthogonality of a pair of scale functions is measured by the integral of the product of their uv -taper functions. To account for uv -coverage, this integral is weighted by the sampling function (see Eqn. 6.13).

⁶A Cholesky decomposition is a decomposition of a symmetric positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose. It is used in the solution of system of equations $[A]\vec{x} = \vec{b}$ where $[A]$ is symmetric positive-definite. The normal equations of a linear least-squares problem are usually in this form. In our case, this linear least-squares problem corresponds to the representation of the sky brightness as a linear combination of basis functions [Press et al. 1988].

Pre-compute Hessian : The first step is to compute $\vec{I}_{s,p}^{psf}$ (Eqn. 6.9) for all possible pairs of scale basis functions. Since convolution is commutative, there will be $N_s + N_s(N_s - 1)/2$ distinct $\vec{I}_{s,p}^{psf}$ images (the diagonal and lower-diagonal terms of the $N_s \times N_s$ block symmetric Hessian matrix) to be computed and then stored.

$$\vec{I}_{s,p}^{psf} = [F^\dagger T_s T_p F] \vec{I}^{psf} \quad (6.14)$$

where \vec{I}^{psf} is the PSF (Eqn. 3.21), normalized to unit peak⁷. The matrix $[H^{peak}]$ is then constructed via Eqn. 6.13 and its inverse is computed and stored in $[H^{peak^{-1}}]$.

Initialization : The model image \vec{I}^{model} is initialized either to zero or to an *a priori* model.

Major and minor cycles : Iterations begin from step 1 and proceed through the following steps. Steps 2 to 4 form the minor cycle, and steps 1 and 5 form the major cycle. In the case of a non-empty initial model, the deconvolution process will begin from step 5.

1. **Compute RHS :** Residual images for each $s \in \{0, \dots, N_s\}$ are computed as

$$\vec{I}_s^{res} = [F^\dagger T_s F] \vec{I}^{res} \quad \text{or} \quad \vec{I}_s^{res} = \vec{I}_s^{shp} \star \vec{I}^{res} \quad (6.15)$$

where \vec{I}^{res} is the current residual image (Eqn. 3.14). For the first iteration $\vec{I}^{res} = \vec{I}^{dirty}$.

2. **Find a Flux Component :** The peak value in the dirty images across all scales is identified and the principal solution is computed at this location⁸. The $N_s \times 1$ solution vector (obtained via Eqn. 6.12) contains the total flux required for components at each spatial scale such that their combined contribution produces the measured flux value at that location in \vec{I}_0^{dirty} . The largest number in this solution vector is chosen as the total flux of a component at the scale to which this maximum corresponds⁹.

Let $\vec{I}_{p,(i)}^{model,\delta}$ represent the chosen flux component of scale size p (at iteration i). This model image contains a δ -function that marks the location of the center of this component and whose amplitude holds the estimated total flux for that component.

⁷ Note that the use of \vec{I}^{psf} with unit peak is equal to scaling both sides of the normal equations by a single scale-factor given by the sum of imaging weights. This is equivalent to defining the weight image \vec{I}^{wt} as the diagonal of the $[H_{0,0}]$ Hessian block, and normalizing all the RHS vectors by it.

⁸ Note that a solution computed via $[H^{peak^{-1}}]$ is valid only at the exact locations of the centers of each flux component. If this inverse is applied to all pixels before searching for peaks, PSF sidelobes are amplified and can mask weak sources even more than usual (see Section 6.1.3).

⁹ When there are exactly overlapping flux components that share the same center, then contributions at all scales are represented in the $N_s \times 1$ solution vector and can be simultaneously removed. However, it is impossible to distinguish this situation from the case of offset but overlapping components in which case a simultaneous solution will be inaccurate. Therefore it is safer to choose only one component at a time, the one corresponding to the largest number in the solution vector.

3. **Update model images** : A single multi-scale model image is accumulated with the chosen component at the p^{th} spatial scale as follows.

$$\vec{I}^{\text{model}} = \vec{I}^{\text{model}} + g(\vec{I}_{p,(i)}^{\text{model},\delta} \star \vec{I}_p^{\text{shp}}) \quad (6.16)$$

where g is a loop-gain that takes on values between 0 and 1 and determines the step size for each iteration in the χ^2 minimization process.

4. **Update RHS** : Each residual image vector is updated by subtracting the contribution of the selected flux component at the spatial scale p (given by $\vec{I}_{p,(i)}^{\text{model},\delta}$). This step is equivalent to evaluating the LHS of the normal equations (Eqn. 6.6) with a series of N_s model image vectors where only $\vec{I}_p^{\text{model},\delta}$ has non-zero elements, and then subtracting this result from the RHS image vectors. This update step can be implemented efficiently if the convolution kernels of each Hessian block $I_{s,t}^{\text{psf}}$ are pre-computed and stored (convolutions with δ -functions are shifted and scaled versions of $I_{s,t}^{\text{psf}}$).

$$\vec{I}_s^{\text{res}} = \vec{I}_s^{\text{res}} - g(\vec{I}_{s,p}^{\text{psf}} \star \vec{I}_{p,(i)}^{\text{model},\delta}) \quad (6.17)$$

This step can also be written in a perhaps more intuitive (but computationally expensive) way to compare it with the update step of standard CLEAN deconvolution (described in Section 3.2.1.9). The standard residual image \vec{I}^{res} (Eqn. 3.14) is updated by first convolving the model image $\vec{I}_{p,(i)}^{\text{model},\delta}$ with a scale PSF $\vec{I}_{s=0,p}^{\text{psf}}$ and then subtracting it out. The resulting residual image is then smoothed to different spatial scales to form the new set of RHS residual images (Eqn. 6.15).

$$\vec{I}^{\text{res}} = \vec{I}^{\text{res}} - g(\vec{I}_{s=0,p}^{\text{psf}} \star \vec{I}_{p,(i)}^{\text{model},\delta}) \quad \text{and then} \quad \vec{I}_s^{\text{res}} = \vec{I}_s^{\text{shp}} \star \vec{I}^{\text{res}} \quad (6.18)$$

This two-stage method (Eqn. 6.18) is possible only because $\vec{I}_{s,p}^{\text{psf}} = \vec{I}_{s=0,p}^{\text{psf}} \star \vec{I}_s^{\text{shp}}$ (according to Eqn. 6.9). Also, it is more computationally intensive than the first method (Eqn. 6.17) because of the extra convolutions that need to be done for every minor cycle iteration. The first method requires only a shift, scaling and subtraction for each flux component and makes use of pre-computed Hessian kernel functions.

Repeat from Step 2 until a flux limit is reached. This flux limit is usually chosen as the amplitude of the largest PSF sidelobe around the brightest source in \vec{I}_0^{res} .

5. **Predict** : Once the minor cycle flux limit is reached, the current best estimate of the multi-scale model image is used to predict model visibilities \vec{V}^{model} (using Eqn. 3.18).

Repeat from Step 1 until the residuals satisfy a stopping criterion usually based on an estimate of how noise-like they are.

Restoration : After convergence, the multi-scale model image is already in a form similar to that in the standard imaging case, where the only steps left are to smooth it with a restoring beam to suppress high spatial frequencies that fall beyond the range of the sampling function and to add in the standard residual image.

Algorithm 3: Multi-Scale Deconvolution as described in Section 6.1.2.5

Data: Calibrated visibilities : $\vec{V}_{n \times 1}^{corr}$
Data: uv-sampling function : $S_{n \times m}$
Data: Image noise threshold and loop gain σ_{thr}, g_s
Data: Scale basis functions : $\vec{I}_s^{shp} \forall s \in \{0, N_s\}$
Result: Model Image : $\vec{I}_{m \times 1}^{model}$

```

1 Compute the dirty image  $\vec{I}^{dirty}$  and psf  $\vec{I}^{psf}$ 
2 foreach scale  $s \in \{0, N_s\}, p \in \{s, N_s\}$  do
3   | Compute  $I_{sp}^{psf} = I_s^{shp} \star I^{psf} \star I_p^{shp}$ 
4 end
5 Construct  $[H^{peak}]$  and  $[H^{peak^{-1}}]$  with  $H_{s,p}^{peak} = mid(I_{sp}^{psf})$ 
6 Measure the peak psf sidelobe  $f_{sidelobe}$ 
7 Initialize the model  $\vec{I}^{model}$  and residual images  $I^{res}$ 
8 repeat                                     /* Major Cycle */
9   | foreach scale  $s \in \{0, N_s\}$  do
10  |   | Calculate smoothed residual images :  $I_s^{res} = I_s^{shp} \star I^{res}$ 
11  |   | Calculate a flux-limit for scale  $s$  :  $f_{limit,s}$ 
12  | end
13 repeat                                     /* Minor Cycle */
14   | foreach scale  $s \in \{0, N_s\}$  do
15   |   | Find the location and amplitude of the peak :  $p_s = peak(I_s^{res})$ 
16   | end
17   | Choose the location of the global peak  $max(p_s)$  for  $s \in \{0, N_s\}$ 
18   | Construct  $I^{pix,dirty}$ , an  $N_s \times 1$  vector from  $I^{res}$  over all  $s \in \{0, N_s\}$ 
19   | Compute principal solution  $I^{sol} = [H^{peak^{-1}}]I^{pix,dirty}$ 
20   | Construct a model  $I_p^{m,\delta}$  from the maximum amplitude entry in  $I^{sol}$ 
21   | Update the model image with a flux component of the chosen scale
22   | size  $p$ , location and amplitude :  $I^{model} = I^{model} + g \delta I_p^{m,\delta} \star I_p^{shp}$ 
23   | foreach scale  $s \in \{0, N_s\}$  do
24   |   | Update the residual image :  $I_s^{res} = I_s^{res} - g [\delta I_p^{m,\delta} \star I_{sp}^{psf}]$ 
25   | until Peak residual at any scale < Flux Limit at that scale
26   | Compute model visibilities  $V^{model}$  from the current model image  $I^{model}$ 
27   | Compute a new residual image  $I^{res}$  from residual visibilities
28   |  $V^{corr} - V^{model}$ 
28 until Peak residual at all scales < stopping threshold
29 Restore the final model image  $I^{model}$ 

```

6.1.3 Differences with existing MS-CLEAN techniques

There are two main differences between the multi-scale deconvolution algorithms described in [Cornwell \[2008\]](#), [Greisen et al. \[2009\]](#) and [Section 6.1.2.5](#). These differences are described below to emphasize the relation between these methods and show how the two existing methods and their implementations are approximations of the generic method described in this section ([6.1.2.5](#)).

- 1. Finding a flux component :** In the first two methods, the amplitude and scale of a flux component are chosen by searching for the peak in the list of dirty images after having applied a scale bias, an empirical term that de-emphasises large spatial scales. The scale bias $b_s = 1 - 0.6 s/s_{max}$ used in [Cornwell \[2008\]](#) (where s_{max} is the width of the largest scale basis function) is a linear approximation of how the inverse of the area under each scale function changes with scale size¹⁰. It is meant to be used to normalize residual images that have been smoothed with scale functions that have unit peak, before flux components are chosen. The algorithm described in [Greisen et al. \[2009\]](#) uses $b_s \approx 1.0/s^{2x}$ where $x \in \{0.2, 0.7\}$, to approximate a normalization by the area under a Gaussian, for the case when images are smoothed by applying a *uv*-taper that tends to unity for the zero spatial frequency. In the context of the algorithm described in [Section 6.1.2.5](#), the diagonal elements of $[H^{peak}]$ are a measure of the area under the main lobe of the PSF at each spatial scale, and both these normalization schemes are roughly equivalent to using a diagonal approximation of $[H^{peak}]$ and discarding all cross-terms when computing the principal solution before picking out flux components.

Once we have this understanding, we can see that the full Hessian $[H^{peak}]$ (and not just a diagonal approximation) can be inverted to get the normalization exactly right, especially for sources that contain overlapping flux components of different spatial scales. It can be shown that by applying the inverse of the full $[H^{peak}]$ to the RHS vectors before picking out a suitable amplitude and scale of a flux component, we are able to get a more accurate estimate of the total-flux of the flux component than by just reading off a peak from a series of dirty images biased by the MS-CLEAN b_s . This difference has been demonstrated on simulations ([Section 6.1.4](#)) where the inverse of $[H^{peak}]$ was applied to all pixels of a series of smoothed dirty-images, but the relative performance of this approach (compared to the existing methods) is yet to be analysed within the complete iterative deconvolution framework. It is likely that the technique described in [Section 6.1.2.5](#) would get more accurate minor cycle estimates and therefore converge in fewer iterations.

- 2. Minor cycle updates :** The update steps in [Cornwell \[2008\]](#) and [Section 6.1.2.5](#) evaluate the full LHS of the normal equations (to account for the non-orthogonality of the

¹⁰ When $s/s_{max} = 1.0$ the bias term is $1.0 - 0.6 = 0.4$ which is approximately equal to the inverse of the area under a Gaussian of unit peak and width, given by $1.0/\sqrt{2\pi} = 0.398$.

basis set) to update the smoothed residual images and subtract out flux components within the image domain. This allows each minor cycle iteration to search for the optimal flux component across all scales without having to recompute smoothed residual images in the visibility domain after each iteration. On the other hand, Greisen et al. [2009] ignores the cross-terms, performs a full set of minor cycle iterations on one scale at a time, and recomputes smoothed residual images *via* the visibility domain after every full set of minor cycle iterations¹¹.

A choice among these three methods (and other possible combinations) will depend on trade-offs between the accuracy within each minor cycle (for measured flux values as well as the update process), minimizing the computational cost per step, and optimizing global convergence patterns to control the total number of iterations. For an example of such a trade-off, see Section 7.1.2.3 (principal solution of the multi-scale multi-frequency normal equations).

6.1.4 Example of the Multi-Scale Principal Solution

This section contains an example of the principal solution computed by applying $[H^{\text{peak}}]^{-1}$ to all pixels in a set of smoothed dirty images (using a set of 2D Gaussians as the scale functions). The purpose of this example is to illustrate how this process is able to separate overlapping flux components of different spatial scales and give an accurate estimate of the total flux contained in each component, and to show when this gives a near optimal solution and when it will not.

Figure 6.4 shows a set of dirty images convolved with Gaussian scale functions (top row) and the result of Eqn. 6.12 (bottom row) over all pixels, for a simulated example of multi-scale imaging with $N_s = 3$. The simulated sky brightness distribution consists of flux components at two spatial scales given by Gaussians whose widths are 1 and 24 pixels. Three sources are constructed using these components. The point source on the top right has 0.1 Jy of flux. The source on the top left is a composite of a point source of flux 0.1 Jy and an extended source of total flux 1.0 Jy, centered on the same pixel. The source on the bottom is a similar composite in which the centers of the point and extended components are offset from each other. The three scale basis functions used for multi-scale imaging correspond to Gaussians of widths 1, 6 and 24 pixels. In this example, two basis functions exactly match the scales present in the sky model, and one does not¹².

¹¹Recomputing smoothed residual images by transforming between the image and visibility domains is a computationally expensive operation. Therefore, it is useful to either find a way to update them within the image domain or to reduce the frequency with which they are recomputed *via* transformations to and from the visibility domain.

¹²Note that in practice, it is usually impossible to find perfectly matching scale sizes for all flux components, and this principal solution will be an approximation. The example described here is only an illustration of what the principal solution means for multi-scale imaging.

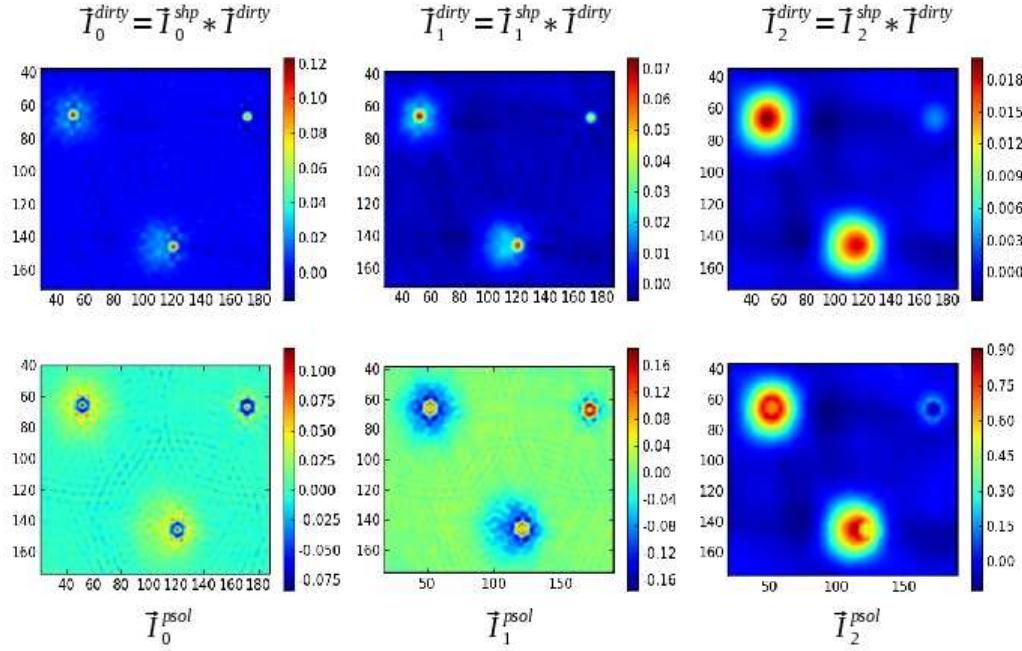


Figure 6.4: Example of the Multi-Scale Principal Solution : These images show a set of three dirty images (top row) and the corresponding principal solution images (bottom row). The purpose of this example is to demonstrate (a) the effect of smoothing the dirty images by \vec{I}^{shp} and what happens to the peak flux, and (b) the fact that the flux values in the principal solutions images are the true total flux values of each component. This result can be used in the minor cycle to get a good estimate of total flux in each flux component. The simulated sky in this example consists of a combination of point sources of flux 0.1 Jy and large flux components of total flux 1.0 Jy. One source is an isolated point source and two sources are composites of one point source and one extended source. Also, the scale basis functions \vec{I}_0^{shp} and \vec{I}_2^{shp} exactly match the point source and extended component respectively, but \vec{I}_1^{shp} matches neither. The top row of images are smoothed versions of the dirty image (Eqn. 6.10, or $\vec{I}^{\text{pix,dirty}}$ from Eqn. 6.12 with all pixels filled in). The image on the top left is the dirty image smoothed with a δ -function and shows the point sources clearly but the peak extended flux is relatively weak. The image on the top right is the dirty image convolved with a scale function matching the large-scale flux component, and shows a good match at the largest spatial scale, but the amplitude is wrong. These amplitudes can be corrected by computing the principal solution. The images in the bottom row are \vec{I}^{psol} ($\vec{I}^{\text{pix,psol}}$ for all pixels) the result of calculating the principal solution (via Eqn. 6.12) for all pixels. The values at the central locations of the sources in \vec{I}_0^{sol} and \vec{I}_2^{sol} are the correct total flux values for a source at the matching scale. The values at the locations of the sources in \vec{I}_1^{sol} are all zero, indicating that this spatial scale is not matched by any flux component. Table 6.1.4 shows the peak values in the top and bottom rows of images at the locations of the three sources.

Source	\vec{I}_0^{true}	\vec{I}_1^{true}	\vec{I}_2^{true}	$\vec{I}_0^{pix,dirty}$	$\vec{I}_1^{pix,dirty}$	$\vec{I}_2^{pix,dirty}$	\vec{I}_0^{sol}	\vec{I}_1^{sol}	\vec{I}_2^{sol}
Point (top right)	0.1	0.0	0.0	0.1	0.04	0.005	0.1	<0.01	<0.01
Extended (top left)	0.1	0.0	1.0	0.12	0.07	0.02	~0.1	<0.01	~0.9
Extended (bottom)	0.1	0.0	1.0	0.11	0.06	0.018	~0.1	<0.01	~0.95

Table 6.1: Multi-scale principal solution example

The top row of images shows the dirty images smoothed to these three spatial scales. The isolated point source peaks only in $\vec{I}_0^{pix,dirty}$ and is suppressed in the other two images. The extended sources peak in $\vec{I}_2^{pix,dirty}$ which is where their scale is best matched. However, the peak flux values in these images are far from the total flux values that the imaging process hopes to reconstruct. The bottom row of images show the principal solution images \vec{I}^{sol} for the three spatial scales. The peaks from \vec{I}^{sol} at the location of these sources are almost exactly equal to their total flux (as required for constructing the model image in the minor cycle). Note that at the central locations of all three sources, the value in \vec{I}_1^{sol} is zero, indicating that there is no flux component at that particular scale.

Table 6.1.4 shows the peak values measured for each source, in the smoothed dirty images as well as in the principal solution images. These numbers show that as desired, the peak values of the solution images give nearly correct total flux estimates for spatial scales that exactly match those in the data.

A few points to note are :

1. This solution is valid only at the locations of the centers of the flux components because it is derived from a diagonal approximation of each Hessian block. This solution will be valid across the entire image only if the PSF is a δ -function. With an interferometric PSF, the sidelobe structure is enhanced, and in some cases can produce artificial peaks that are higher than the true values at the center. In the case of overlapping components, the sidelobes of one component will cause errors in the estimate of the others.
2. The total flux estimates derived from the principal solution will give an exact solution only if the scale basis functions exactly match the scales present in the image. Otherwise it will still model it correctly, but may not be the optimal set of scale basis function and may use more flux components than required.

Therefore, the minor cycle of an iterative multi-scale deconvolution algorithm still needs heuristics to decide how to pick flux components. The principal solution gives a more accurate estimate than using a scale bias, but it is at the expense of higher sidelobe structure. The algorithm described in Section 6.1.2.5 suggests choosing the location of a source from the peak of $\vec{I}_0^{pix,dirty}$, calculating the principal solution only at that location, and subtracting out only the dominant flux component at that location.

6.2 Multi-Frequency Synthesis Deconvolution

The uv -coverage of a synthesis array can be greatly improved by using the fact that visibilities measured at different receiver frequencies correspond to different spatial frequencies. Multi-frequency synthesis (MFS) is the process of combining data from multiple spectral channels onto the same spatial-frequency grid during imaging to take advantage of the increased uv -coverage and imaging sensitivity. As long as the sky brightness does not vary across the total measured bandwidth, standard imaging and deconvolution algorithms can be used along with MFS. If the sky brightness varies across the observing bandwidth, the monochromaticity requirement of aperture synthesis breaks down and the 2D Fourier relation in the van Cittert Zernike theorem (Eqn. 2.9) does not hold. In other words, when data from multiple frequencies are gridded together, there is no way to tell if variations in the measurements across the spatial frequency plane are due to spatial structure, or spectral structure, or both. However, there is often enough information in the data (*via* the known frequency dependence of the sampling pattern) to separate the two, and both spatial and spectral structure can be derived simultaneously by choosing a physically appropriate model for the sky brightness distribution as well as its frequency dependence. If the spatial structure is known (*a priori* or *via* a physically appropriate flux model), any remaining structure on the spatial frequency plane can be attributed to spectral structure and separately modeled. Or, if the spectral structure is known, this information can be used to constrain the spatial structure.

Section 6.2.1 defines a broad-band flux model that approximates a power law with a polynomial in frequency. Section 6.2.2 describes a multi-frequency deconvolution algorithm that models the spatial structure by a collection of δ -functions, and the spectral structure as a smooth N^{th} order polynomial in frequency. The basic idea is to look at the spectra for individual locations on the sky, perform an N^{th} order polynomial fit *via* a least-squares approach, and produce $N + 1$ coefficient images, all within a deconvolution framework that takes advantage of the combined multi-frequency uv -coverage and optimizes the broad-band sensitivity *via* a weighting (preconditioning) scheme. This discussion is a formal derivation of the technique described in [Sault and Wieringa \[1994\]](#) and [Conway et al. \[1990\]](#) and describes a modified version of the Sault-Wieringa MF-CLEAN algorithm that improves upon its imaging fidelity (Section 6.2.3 describes this difference and its implications). Section 6.2.4.1 contains a brief discussion of the errors incurred by approximating a power-law with an N^{th} order polynomial. Chapter 7 later describes a similar algorithm that combines this approach with the multi-scale image model described in section 6.1. This entire section ignores all direction dependent instrumental effects and their frequency dependence. Section 7.2 deals with these effects by folding the antenna primary beams and their frequency dependence into the multi-frequency imaging equations.

6.2.1 Multi-Frequency Image Model

The sky brightness distribution of astronomical sources usually varies with frequency, either due to the spectral structure of the emitted radiation or if the spatial structure of the radiating object varies with frequency. In either case, an accurate reconstruction of the wide-band sky brightness distribution will require a frequency-dependent flux model to be folded into the measurement equation. Just as standard interferometric image reconstruction uses *a-priori* information about the spatial structure of the sky to estimate the visibility function in unmeasured regions of the uv -plane, multi-frequency imaging algorithms need to use *a-priori* information about the spectral structure of the sky brightness during reconstruction from data with incomplete spectral sampling. A wide-band flux model can provide these constraints on the sky spectrum during non-linear deconvolution.

A simple spectral model for the Stokes I components of broad-band continuum emission can be a polynomial in frequency. This functional form (a linear combination of basis functions) is desirable because it makes the numerical optimization process more tractable (can apply linear least-squares). However, broad-band continuum emission from astrophysical sources is often best represented by a power-law. Across the wide frequency ranges that new receivers are now sensitive to, spectral breaks, steepening and turnovers also need to be included in these models, and the simplest way to do so while ensuring smoothness is with a varying power-law index (spectral curvature).

A power law with a varying index can be represented by a second-order polynomial in $\log(I)$ vs $\log\left(\frac{\nu}{\nu_0}\right)$ space. The coefficients of the polynomial are the logarithm of the flux at a reference frequency $\log(I_{\nu_0})$, the average spectral index α and the curvature β .

$$\log(I_{\nu}) = \log(I_{\nu_0}) + \alpha \log\left(\frac{\nu}{\nu_0}\right) + \beta \log^2\left(\frac{\nu}{\nu_0}\right) \quad (6.19)$$

$$\Rightarrow I_{\nu,\alpha,\beta} = I_{\nu_0} \left(\frac{\nu}{\nu_0}\right)^{\alpha+\beta \log\left(\frac{\nu}{\nu_0}\right)} \quad (6.20)$$

Although Eqn. 6.19 describes a model that is parameterized as a polynomial, it is impractical to work directly in $\log(I)$ vs $\log(\nu)$ space because this involves the numerically unstable process of taking logarithms of image pixel amplitudes in the presence of noise. Two alternate polynomial models based on Taylor series expansions are described below (I vs ν and I vs $\log(\nu)$) followed by the definition of a wide-band flux model as a linear combination of a finite number of spectral basis functions.

6.2.1.1 Series expansion of $I_{\nu,\alpha,\beta}$ about $\nu = \nu_0$

Expanding Eqn. 6.20 about ν_0 yields a polynomial in I vs $\left(\frac{\nu-\nu_0}{\nu_0}\right)$ space. The physical parameters I_{ν_0} , α and β can be obtained from the first three terms. However, an accurate fit to the power law may require more terms in the series, and the error in the fit after an n^{th} order expansion is $O\left[\alpha\left(\frac{\nu-\nu_0}{\nu_0}\right)\right]^{n+1}$.

$$\begin{aligned}
I_{\nu,\alpha,\beta} &= I_{\nu_0} + \left[\frac{\partial I_{\nu,\alpha,\beta}}{\partial \nu} \right]_{\nu_0} (\nu - \nu_0) + \frac{1}{2} \left[\frac{\partial^2 I_{\nu,\alpha,\beta}}{\partial \nu^2} \right]_{\nu_0} (\nu - \nu_0)^2 \\
&\quad + \frac{1}{6} \left[\frac{\partial^3 I_{\nu,\alpha,\beta}}{\partial \nu^3} \right]_{\nu_0} (\nu - \nu_0)^3 + \dots
\end{aligned} \tag{6.21}$$

$$\begin{aligned}
I_{\nu,\alpha,\beta} &= \sum_{t=0}^N I_t \left(\frac{\nu - \nu_0}{\nu_0} \right)^t \quad \text{where } I_0 = I_{\nu_0} \\
I_1 &= I_{\nu_0} [\alpha] \\
I_2 &= I_{\nu_0} [\alpha(\alpha - 1)/2 + \beta] \\
I_3 &= I_{\nu_0} [\alpha(\alpha - 1)(\alpha - 2)/6 + \beta(\alpha - 1)] \\
&\dots \text{ and so on.}
\end{aligned} \tag{6.22}$$

6.2.1.2 Series expansion of $I_{\nu,\alpha,\beta}$ about $\alpha = 0, \beta = 0$

Another power-series expansion can be obtained via partial derivatives with respect to α and β . This expansion yields a polynomial in I vs $\log\left(\frac{\nu}{\nu_0}\right)$ space. The first few power-series co-efficients derived from a third-order expansion about $\alpha = 0, \beta = 0$ is given below. Here too, only the first three terms are needed to calculate $I_{\nu_0, \alpha}$ and β , and the error term after an n^{th} order expansion is $O\left[\alpha \log\left(\frac{\nu}{\nu_0}\right)\right]^{n+1}$.

$$\begin{aligned}
I_{\nu,\alpha,\beta} &= I_{\nu_0} + \alpha \left[\frac{\partial I_{\nu,\alpha,\beta}}{\partial \alpha} \right]_{0,0} + \beta \left[\frac{\partial I_{\nu,\alpha,\beta}}{\partial \beta} \right]_{0,0} + \alpha \beta \left[\frac{\partial^2 I_{\nu,\alpha,\beta}}{\partial \alpha \partial \beta} \right]_{0,0} \\
&\quad + \frac{\alpha^2}{2} \left[\frac{\partial^2 I_{\nu,\alpha,\beta}}{\partial \alpha^2} \right]_{0,0} + \frac{\beta^2}{2} \left[\frac{\partial^2 I_{\nu,\alpha,\beta}}{\partial \beta^2} \right]_{0,0} \\
&\quad + \frac{\alpha^3}{6} \left[\frac{\partial^3 I_{\nu,\alpha,\beta}}{\partial \alpha^3} \right]_{0,0} + \frac{\beta^3}{6} \left[\frac{\partial^3 I_{\nu,\alpha,\beta}}{\partial \beta^3} \right]_{0,0} \\
&\quad + \frac{\alpha^2 \beta}{2} \left[\frac{\partial^3 I_{\nu,\alpha,\beta}}{\partial \alpha^2 \partial \beta} \right]_{0,0} + \frac{\alpha \beta^2}{2} \left[\frac{\partial^3 I_{\nu,\alpha,\beta}}{\partial \alpha \partial \beta^2} \right]_{0,0} + \dots
\end{aligned} \tag{6.23}$$

$$\begin{aligned}
I_{\nu,\alpha,\beta} &= \sum_{t=0}^N I_t \left[\log\left(\frac{\nu}{\nu_0}\right) \right]^t \quad \text{where } I_0 = I_{\nu_0} \\
I_1 &= I_{\nu_0} [\alpha] \\
I_2 &= I_{\nu_0} [\alpha^2/2 + \beta] \\
I_3 &= I_{\nu_0} [\alpha^3/6 + \alpha \beta] \\
&\dots \text{ and so on.}
\end{aligned} \tag{6.24}$$

This expansion can be done about more realistic values of α and β (for example, $\alpha_0 = -0.7, \beta_0 = 0.0$). In this case the coefficients can be interpreted as $I_0 = I_{\nu, \alpha_0, \beta_0}$, $I_1 = I_{\nu, \alpha_0, \beta_0}[\alpha - \alpha_0]$, $I_2 = I_{\nu, \alpha_0, \beta_0}[(\alpha - \alpha_0)^2/2 + (\beta - \beta_0)]$ and so on, but the functional form of the spectral basis function will not change. To take advantage of a known average spectral index and curvature over all sources in the image, the data can be scaled by $(\frac{\nu}{\nu_0})^{\alpha_0 + \beta_0(\frac{\nu}{\nu_0})}$ prior to multi-frequency imaging. [Conway et al. \[1990\]](#) suggest this approach to reduce the magnitude of the higher-order terms in the series so that a solution with fewer terms in the Taylor series expansion (or even standard MFS) may suffice for an accurate deconvolution.

6.2.1.3 Image model for multi-frequency deconvolution

A sky brightness distribution that varies smoothly with observing frequency can be modeled as a linear combination of spectral basis functions and coefficient images (see Eqns [6.22](#) and [6.24](#) for two possible series expansions). The flux modeled at each frequency channel can be written as

$$\vec{I}_{\nu}^{model} = \sum_{t=0}^{N_t-1} w_{\nu}^t \vec{I}_t^{sky} \quad \text{where} \quad w_{\nu}^t = \left(\frac{\nu - \nu_0}{\nu_0} \right)^t \quad (6.25)$$

$$\text{or} \quad w_{\nu}^t = \left[\log \left(\frac{\nu}{\nu_0} \right) \right]^t \quad (6.26)$$

where N_t is the total number of terms in the series and \vec{I}_t^{sky} are the coefficient images (one set of coefficients per pixel). ν is the observing frequency and ν_0 is a chosen reference frequency and w_{ν}^t is the evaluated result of the t^{th} basis function for a given value of ν . w_{ν}^t is used as a weight during MFS gridding and will be called a Taylor-weight.

Eqns. [6.25](#) and [6.26](#) show two choices of spectral basis functions. In practice it is a simple matter to switch between the two, depending on the type of spectrum of the emission being imaged. The linear expansion (Eqn. [6.25](#)) was chosen for all the tests in this dissertation because it can be applied to arbitrary but smooth spectra. For pure power-law spectra associated with isolated point sources or extended sources with a constant spectral index (across the source), the logarithmic expansion is a better choice from the point of view of series convergence [[Conway et al. 1990](#)]. However, in general, the sum of two power-laws is not another power-law. Therefore, when the spectral index varies smoothly across extended emission, and the wide-band spatial structure is modeled using a collection of overlapping extended flux components with fixed spectral shapes (see Chapter [7](#) for a multi-scale multi-frequency deconvolution algorithm that uses this approach), a Taylor expansion in I vs ν space is a more general choice of parameterization. Section [6.2.4](#) discusses some of the errors associated with the use of such Taylor polynomials to fit pure power law spectra.

6.2.2 Imaging Equations and Block Deconvolution

This section contains a derivation of the normal equations for a multi-frequency image model. This is followed by a description of the principal solution and its use in an iterative joint deconvolution. Algorithm 4 lists the multi-frequency deconvolution method described in this section. The derivations in this section use a block-matrix notation (described in Appendix B) to represent the measurement and normal equations.

6.2.2.1 Measurement equations

We begin with an example of how the sampling function of the interferometer (and the PSF) changes with observing frequency. Figure. 6.5 shows a set of 1-D gridded sampling functions and corresponding PSFs for three different frequencies $\nu_1, \nu_2 = 2\nu_1, \nu_3 = 3\nu_2$ ($N_c = 3$). These plots show that each frequency measures a different range of spatial-frequencies, and the angular resolution of the instrument increases with frequency.

The goal of multi-frequency-synthesis is to use the combined uv -coverage from all measured frequencies and reconstruct the image at the angular resolution allowed by the highest frequency in the band. One way to accomplish this is to write separate measurement equations for each frequency, and then solve them simultaneously.

Let there be N_c observing frequencies with n measurements taken at each frequency ν . The visibility vector for each frequency (\vec{V}_ν^{corr}) has the shape $n \times 1$, and the sampling matrix $[S_\nu]$ has the shape $n \times m$ (Figure. 6.5 shows an example of how the sampling function and the corresponding PSFs change as a function of frequency). $[F_{m \times m}]$ is the Fourier transform operator (image to spatial-frequency) and all images \vec{I} are lists of m pixel amplitudes. The measurement equations for one frequency are given as follows.

$$\vec{V}_\nu^{corr} = [S_\nu][F]\vec{I}_\nu^{model} = \sum_{t=0}^{N_t-1} w_\nu^t [S_\nu][F]\vec{I}_t^{sky} \quad (6.27)$$

Note that the images $\vec{I}_t^{model} \forall t \in \{0, N_t - 1\}$ form the N_t coefficients of the series expansion. The order of each term in this Taylor polynomial is denoted by the subscript t .

A multi-frequency measurement equation can be written by combining measurements from all frequency channels. The full visibility vector \vec{V}^{corr} (note, no subscript ν) now contains all frequencies and has the shape $nN_c \times 1$.

$$\vec{V}^{corr} = \sum_{t=0}^{N_t-1} [W_t^{mfs}][S][F]\vec{I}_t^{sky} \quad (6.28)$$

The sampling operator $[S]$ (of shape $nN_c \times m$) is a vertical stack of $[S_\nu]$ matrices (of shape $n \times m$ each) over all N_c frequency channels. $[W_t^{mfs}]$ is a block diagonal matrix of shape $nN_c \times nN_c$, constructed from N_c diagonal matrices (each of shape $n \times n$, denoted by W_ν^{im} ,

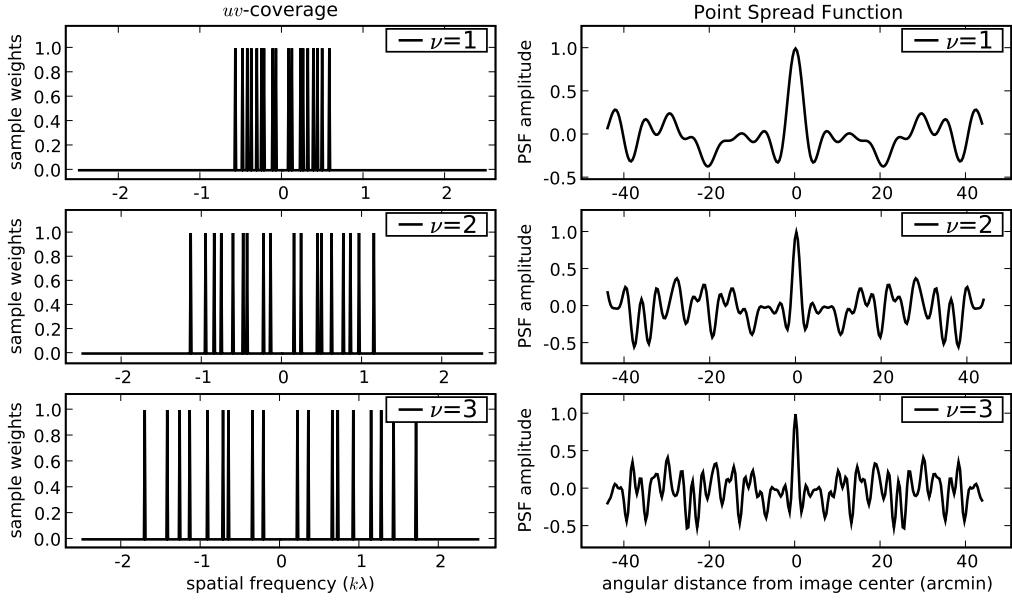


Figure 6.5: Multi-frequency sampling weights and PSFs : The plots on the left show 1-D gridded sampling functions $[S_\nu^\dagger W S_\nu]$ at three different frequencies ($\nu = 1, 2, 3$ from top to bottom). The plots on the right show the corresponding PSFs \vec{I}_ν^{psf} calculated via Eqn. 3.11. The range of sampled spatial frequencies scales linearly with frequency. This changes the shape of the PSF, and the width of the main lobe (angular resolution) decreases at higher frequencies ($\theta_{\nu=1}^{psf} = 5.7'$ with $u_{max} = 0.6 k\lambda$, $\theta_{\nu=2}^{psf} = 2.8'$ with $u_{max} = 1.2 k\lambda$, $\theta_{\nu=3}^{psf} = 1.9'$ with $u_{max} = 1.8 k\lambda$). These 1D plots are the multi-frequency equivalent of Figure 3.1 (with different spatial-frequency sampling functions chosen to illustrate the difference between frequencies).

and containing the weights w_ν^t). $[F_{m \times m}]$ is the same Fourier transform operator used in the single-frequency equations (Eqn. 6.27). The full measurement matrix ([A] in Eqn. 2.20) therefore has the shape $nN_c \times mN_t$ where each of the N_t block columns corresponds to one series coefficient. This $nN_c \times mN_t$ measurement matrix operates on the $mN_t \times 1$ column vector of image pixels to produce nN_c visibilities.

An example for $N_t = 3$ is shown below.

$$\left[\begin{bmatrix} W_0^{mfs} SF \end{bmatrix} \quad \begin{bmatrix} W_1^{mfs} SF \end{bmatrix} \quad \begin{bmatrix} W_2^{mfs} SF \end{bmatrix} \right] \left[\begin{array}{c} \vec{I}_0^{sky} \\ \vec{I}_1^{sky} \\ \vec{I}_2^{sky} \end{array} \right] = \vec{V}^{corr} \quad (6.29)$$

Each model image vector \vec{I}_q^{sky} is an image of the q^{th} coefficient of the Taylor polynomial used to represent the spectrum at each point on the sky (see Figure 8.1 for an example).

6.2.2.2 Normal equations

The least-squares solution of Eqn. 6.29 is computed by solving the normal equations (as shown in block-matrix form in Appendix B). With N_t terms in the model description, the Hessian is made up of $N_t \times N_t$ blocks, each of size $m \times m$. The sky model and dirty images are a set of N_t image vectors each of size $m \times 1$. Let us label the normal equations as follows.

$$[H_{mN_t \times mN_t}^{mfs}] \vec{I}_{mN_t \times 1}^{\text{sky},mfs} = \vec{I}_{mN_t \times 1}^{\text{dirty},mfs} \quad (6.30)$$

These normal equations can be written in block matrix form. As an example, consider the case where $N_t = 3$.

$$\begin{bmatrix} [H_{t=0,q=0}] & [H_{t=0,q=1}] & [H_{t=0,q=2}] \\ [H_{t=1,q=0}] & [H_{t=1,q=1}] & [H_{t=1,q=2}] \\ [H_{t=2,q=0}] & [H_{t=2,q=1}] & [H_{t=2,q=2}] \end{bmatrix} \begin{bmatrix} \vec{I}_{q=0}^{\text{sky}} \\ \vec{I}_{q=1}^{\text{sky}} \\ \vec{I}_{q=2}^{\text{sky}} \end{bmatrix} = \begin{bmatrix} \vec{I}_{t=0}^{\text{dirty}} \\ \vec{I}_{t=1}^{\text{dirty}} \\ \vec{I}_{t=2}^{\text{dirty}} \end{bmatrix} \quad (6.31)$$

Figure 6.6 is a pictorial representation of these normal equations (Eqn. 6.31) for $N_t = 3$, using the multi-frequency sampling functions shown in Fig. 6.5 (and labeled as shown in Eqn. 6.30). In the figures, the full Hessian matrix on the LHS of Eqn. 6.31 is denoted as $[H_{3m \times 3m}^{mfs}]$ and the model and dirty image vectors are denoted as stacks of $N_t = 3$ vectors each ($\vec{I}^{\text{sky},mfs}$, $\vec{I}^{\text{dirty},mfs}$).

These matrix equations can be written row-by-row as follows. The indices t,q vary from 0 to $N_t - 1$ and will henceforth denote block row and column indices for multi-frequency equations.

$$\sum_{q=0}^{N_t-1} [H_{t,q}] \vec{I}_q^{\text{sky}} = \vec{I}_t^{\text{dirty}} \quad \forall t \in \{0 \dots N_t - 1\} \quad (6.32)$$

There are two ways of writing and computing $[H_{t,q}]$ and \vec{I}_t^{dirty} . The following pairs of equations show that the Hessian blocks and RHS vectors can be computed either by gridding Taylor-weighted visibilities from all frequencies together (Eqns. 6.33, 6.34, Σ_v in the visibility domain), or by a Taylor-weighted sum of the Hessian and RHS vectors formed separately from each frequency (Eqns. 6.35, 6.36, Σ_v in the image domain).

1. Calculating the normal equations from the measurement equations as written in Eqn. 6.31 gives the following forms for the Hessian blocks and RHS vectors. In these equations, summations over frequency are implicit in the dimensions of the matrices that make up $[H_{t,q}]$.

$$\text{where } [H_{t,q}] = [F^\dagger S^\dagger W_t^{mfs\dagger} W_q^{im} W_q^{mfs} S F] \quad (6.33)$$

$$\vec{I}_t^{\text{dirty}} = [F^\dagger S^\dagger W_t^{mfs\dagger} W_q^{im}] \vec{V}^{\text{corr}} \quad (6.34)$$

$[W^{im}]$ is a block diagonal matrix formed from a set of N_c single-frequency diagonal weight matrices $[W_\nu^{im}]$, and \vec{V}^{corr} contains all nN_c visibilities.

2. Eqns. 6.33 and 6.34 can be re-written in a more intuitive form in terms of single-frequency sampling functions, weights and PSFs and explicit summations over frequency.

$$[H_{t,q}] = \sum_\nu w_\nu^{t+q} [F^\dagger S_\nu^\dagger W_\nu^{im} S_\nu F] = \sum_\nu w_\nu^{t+q} [B_\nu] \quad (6.35)$$

$$\vec{I}_t^{dirty} = \sum_\nu w_\nu^t [F^\dagger S_\nu^\dagger W_\nu^{im}] \vec{V}_\nu^{corr} = \sum_\nu w_\nu^t \vec{I}_\nu^{dirty} \quad (6.36)$$

Here, \sum_ν indicates a sum across N_c frequency channels, and $[B_\nu]$ and \vec{I}_ν^{dirty} are the Hessian (Beam matrix) and dirty image for frequency ν (given by Eqns. 3.9 and 3.10).

Consider each Hessian block $[H_{t,q}]$. Each term in the summation in Eqn. 6.35 is a Beam matrix (a convolution operator) and therefore each Hessian block $[H_{t,q}]$ is also a convolution operator (matrix multiplication is distributive). The kernel of each $[H_{t,q}]$ will be denoted as $\vec{I}_{t,q}^{psf}$ and is computed as a weighted sum of \vec{I}_ν^{psf} (Eqn. 3.11) computed at each frequency ν .

$$\vec{I}_{t,q}^{psf} = \sum_\nu w_\nu^{t+q} [F^\dagger S_\nu^\dagger W_\nu^{im}] \vec{I} = \sum_\nu w_\nu^{t+q} \vec{I}_\nu^{psf} \quad (6.37)$$

1-D examples of these convolution kernels are shown in Fig.6.6 as the shifted rows in each Hessian block. The kernels functions of the first row of Hessian blocks ($\vec{I}_{t=0,q}^{psf} \forall q \in \{0, N_t - 1\}$) represent the instruments response functions to a point source whose spectrum is given by the q^{th} spectral basis function (Eqn.6.25) and are called spectral PSFs [Sault and Wieringa 1994].

Fig. 6.6 represents the normal equations for a three-term series expansion ($N_t = 3$). The three segments of $I_{3m \times 1}^{sky,mfs}$ represent three coefficient images that make up a multi-frequency model for two point sources on an empty sky. The amplitudes of the δ -functions were chosen such that both point sources have unit total flux at the reference frequency, one has a positive slope in frequency while the other has a negative slope, and both have positive curvature. The Hessian matrix on the LHS is comprised of 3×3 blocks each of size $m \times m$. Each block is a convolution operator constructed from $\vec{I}_{t,q}^{psf}$. These equations show that the dirty image vectors on the RHS can be written as a linear combination of convolutions of spectral coefficient images with $\vec{I}_{t,q}^{psf}$. The Taylor-coefficient model images $I_{3m \times 1}^{sky,mfs}$ can be recovered via a combination of deconvolution and block inversion of the Hessian matrix.

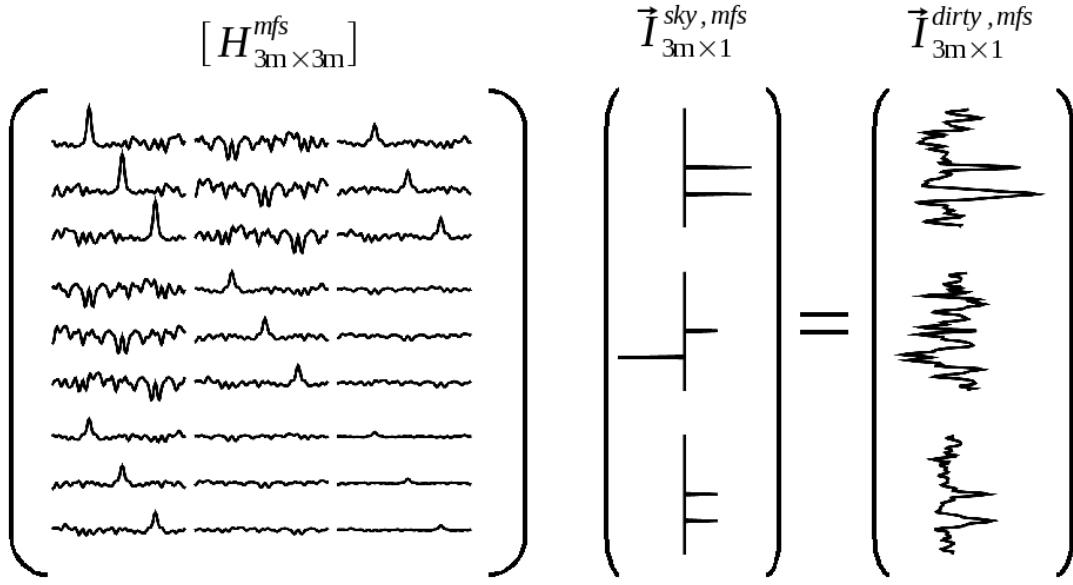


Figure 6.6: Normal Equations for Multi-Frequency Deconvolution : This diagram is pictorial representation of the normal equations formed when the sky spectrum is described as a 2^{nd} -order Taylor polynomial in frequency (Eqns. 6.30 and 6.31 with $N_t = 3$) and the spatial structure is described by a point-source flux model (no multi-scale). $I_{3m \times 1}^{sky, mfs}$ is a 1-D model of two point sources on an empty sky. The three segments correspond to the Taylor-coefficient images I_q^{sky} for $q = 0, 1, 2$ (see Eqn. 6.25) and represent the case where both point sources have unit total flux at the reference frequency (top vector, δ -function amplitudes are 1.0 and 1.0), the spectrum of one source has a positive slope in frequency while the other has a negative slope (middle vector, δ -function amplitudes are +0.5 and -1.0), and the spectra of both sources have positive curvature (bottom vector, δ -function amplitudes are +0.5 and +0.5). The Hessian matrix (LHS) consists of 3×3 blocks each of size $m \times m$, and each block is a convolution operator constructed from a pair of Taylor functions (Eqn. 6.35). The RHS vectors are computed as Taylor-weighted sums of the single-frequency dirty images (Eqn. 6.36). These equations show how the dirty image vectors (RHS) can be written as a linear combination of convolutions. The Taylor-coefficient model images $I_{3m \times 1}^{sky, mfs}$ can be recovered via a combination of deconvolution and block inversion of the Hessian matrix. (A few points can be noted about the Hessian matrix. The top-left block $[H_{t=0, q=0}]$ is the Beam matrix $[B]$ constructed from the sum of all the single-frequency PSFs (see Fig. 6.5 for the single-frequency PSFs used in this example). With only three frequencies, each row of the top middle block $[H_{t=0, q=1}]$ is the difference between the first and the last single-frequency PSF. The central values of the off-diagonal blocks are non-zero but smaller than the peaks on the diagonal, making $[H^{peak}]$ a well-conditioned matrix.)

6.2.2.3 Principal Solution

The principal solution is obtained by approximating each Hessian block by its diagonal and computing the solution independently for each pixel (similar to the principal solution for multi-scale imaging described in Section 6.1.2.3). The peak values measured from the resulting images are in true sky flux units, and these images contain no contribution from the invisible distribution of images.

The approximate Hessian is a set of $N_t \times N_t$ diagonal matrices. Since each block is a convolution operator with identical numbers on the diagonal, the entire Hessian approximation can be written as one single $N_t \times N_t$ element matrix denoted as $[H_{N_t \times N_t}^{peak}]$. The dirty image vectors are also written one pixel at a time as a $N_t \times 1$ vector and a per-pixel solution is obtained (similar to Eqn. 6.12) as follows.

$$I_{N_t \times 1}^{pix,psol} = [H_{N_t \times N_t}^{peak}]^{-1} I_{N_t \times 1}^{pix,dirty} \quad \text{for each pixel} \quad (6.38)$$

The values in $I_{N_t \times 1}^{pix,psol}$ are then filled back into the N_t model image vectors, also one pixel at a time. For an imaging instrument whose PSF is a δ -function, the principal solution gives the final reconstructed image. When there is incomplete sampling, this inversion is valid only at the peaks of sources and can be used only to find flux components during the minor cycle of deconvolution.

6.2.2.4 Properties of $[H^{peak}]$

Some properties of $[H^{peak}]$ for multi-frequency imaging are given below.

1. Each element of $[H^{peak}]$ is given by

$$H_{t,q}^{peak} = \text{mid}\{\vec{I}_{tq}^{psf}\} = \text{tr}\left[\sum_v w_v^{t+q} [S_v^\dagger W_v^{im} S_v]\right] \quad \forall t, q \in \{0 \dots N_t - 1\} \quad (6.39)$$

2. The elements on the diagonal of $[H^{peak}]$ correspond to $t = q$ and are a measure of the sensitivity of the instrument to a the t^{th} spectral basis function. Note that the element $H_{0,0}^{peak}$ is the same as the peak of the PSF in the standard Beam matrix (see Eqn. 3.11).
3. The off-diagonal elements ($t \neq q$) are a measure of the level of orthogonality of the basis set (see footnote 5 on page 94) for the given uv -coverage and weighting scheme. For example, in the special case of natural weighting and an equal and symmetric distribution of frequencies on either side of ν_0 , the value of $\text{mid}\{\vec{I}_{01}^{psf}\}$ will be exactly zero, reflecting the fact that the first two spectral basis functions are orthogonal¹³.

¹³Note that the series expansion of Eqn. 6.22 corresponds to a set of functions given by $1, x, x^2, x^3, \dots$,

4. $[H^{peak}]$ is a Vandermonde¹⁴ matrix whose inverse can be easily computed via a Cholesky decomposition. Also, the value of N_t is usually < 5 , making this inversion tractable.

6.2.2.5 Iterative Block Deconvolution

This section describes an iterative joint deconvolution process for multi-frequency imaging, similar to that described for multi-scale deconvolution. The data products in this case are a set of spectral coefficient images.

Pre-compute Hessian : $\vec{I}_{t,q}^{psf}$ (the kernels of each Hessian block) are first computed for all possible pairs of spectral basis functions. Since the multiplication of diagonal matrices is commutative, there will be $N_t + N_t(N_t - 1)/2$ distinct functions to be computed via Eqn. 6.37 in which the PSF for each frequency v is computed via Eqn. 3.11.

The peak of the un-normalized \vec{I}_{00}^{psf} is equal to $w_{sum} = \text{tr}[S^\dagger W^{im} S]$, and has no contribution from the spectral basis functions (since the zeroth order function is a vector of ones). All spectral PSFs are to be normalized by w_{sum} to get the peak of \vec{I}_{00}^{psf} to unity, but to retain the relative weights between spectral PSFs. This is equivalent to defining the weight image \vec{I}^{wt} as the diagonal of the $[H_{0,0}]$ Hessian block, and normalizing all the RHS vectors by it. The matrix $[H^{peak}]$ is constructed (via Eqn. 6.39) and its inverse is computed and stored in $[H^{peak^{-1}}]$.

Initialization : The N_t model images are first initialized to zero (or an *a-priori* model).

Major and minor cycles : The following steps describe the iterative deconvolution process for multi-frequency synthesis imaging. Steps 1 and 5 form the major cycle and steps 2 to 4 form the minor cycle.

1. **Compute RHS :** The residual images $I_t^{res}, \forall t \in [0, N_t - 1]$ for each term in the series are computed from the residual visibilities at each frequency (\vec{V}_v^{res}) via Eqn. 6.34 in

and a set formed from more than the first two basis functions functions, will not form an orthogonal basis. The Legendre series gives a polynomial model in which the basis functions are orthogonal. However, this orthogonality will manifest itself in the normal equations only in the case of natural weighting and symmetric frequency sampling, and therefore in practice, the non-orthogonality of the basis set is not significant.

¹⁴ A Vandermonde matrix is one in which the rows and columns contain a geometric progression of some kind. A useful property of such a matrix is that the diagonal elements contain even powers of the root function, and this guarantees positive-definiteness and therefore invertibility. A common situation in which such matrices arise is in computing least-squares solutions for the coefficients of a series expansion in which the basis functions form a geometric progression (say, $1, x, x^2, x^3, \dots$). With such a model, the rows and columns of the Hessian matrix will be a Vandermonde system and will therefore be invertible.

which the dirty images for each frequency ν are computed *via* Eqn. 3.14. For the first iteration, $\vec{V}_\nu^{\text{res}} = \vec{V}_\nu^{\text{corr}}$ and $\vec{I}_t^{\text{res}} = \vec{I}_t^{\text{dirty}}$.

2. **Find a Flux Component** : The principal solution (obtained *via* Eqn. 6.38) is computed for all pixels, one at a time. A solution set comprised of δ -functions for each spectral coefficient image is constructed from the solutions at the location of the largest $q = 0$ component.

Alternatively, the principal solution can be computed only at the location of the peak of the 0^{th} -order dirty image (as described in Section 6.1.2.5 for multi-scale deconvolution). This may not be the most optimal flux component to be subtracted out at this iteration, but since this direct solution is strictly valid only at the locations of source peaks, this approach may be useful at low signal-to-noise levels. In practice, iterations can switch back and forth between the two forms of finding flux components, as a trade-off between convergence speed and stability.

Either way, the result of this step is $I_{q,(i)}^{\text{model}} \forall q \in \{0, N_t - 1\}$, a set of N_t model images, each containing a δ -function that marks the location of the source (the index i is an iteration counter). The amplitudes of these δ -functions are the coefficients of the Taylor polynomial that has been fitted to the source spectrum at that one location.

3. **Update model images** : Model images for each spectral coefficient are updated as

$$I_q^{\text{model}} = I_q^{\text{model}} + g I_{q,(i)}^{\text{model}} \quad \forall q \in [0, N_t - 1] \quad (6.40)$$

g is a loop-gain that takes on values between 0 and 1 and controls the step size of each iteration of the χ^2 minimization process.

4. **Update RHS** : The residual images on the RHS are updated by evaluating and subtracting out the entire LHS of the normal equations for the model image vectors ($I_{q,i}^{\text{model}}$, $\forall q$) obtained in iteration i .

$$I_t^{\text{res}} = I_t^{\text{res}} - g \left(\sum_{q=0}^{N_t-1} I_{t,q}^{\text{psf}} \star I_{q,(i)}^{\text{model}} \right) \quad (6.41)$$

This update step can be implemented efficiently if the convolution kernels of each Hessian block $I_{t,q}^{\text{psf}}$ are pre-computed and stored (convolutions with δ -functions are shifted and scaled versions of $I_{t,q}^{\text{psf}}$).

There are two differences¹⁵ between this update step and that for multi-scale decon-

¹⁵ Note that the update step for multi-frequency deconvolution (Eqn. 6.41) is derived using exactly the same mathematical idea used for standard deconvolution (Eqn. 3.17 in Section 3.2.1.9) and multi-scale deconvolution (Eqn. 6.18 in Section 6.1.2.5). All these algorithms evaluate the LHS of the normal equations using the model images for iteration (i) , and subtract it out from the current RHS vectors. The differences between these algorithms arise purely from the different flux models (standard CLEAN uses a single set of δ -functions, multi-scale CLEAN associates a spatial scale with each set of δ -functions and multi-frequency CLEAN uses sets of δ -functions for each coefficient of the Taylor polynomial).

volution (Eqn. 6.18). First, each dirty image I_t^{res} is updated by subtracting out the contribution of all N_t coefficients (and hence the summation in Eqn. 6.41). Second, this update step cannot be written in two stages as was done for multi-scale deconvolution (Eqn. 6.18) because the convolution kernels of each Hessian block ($I_{t,q}^{psf}$) cannot be written as a sequence of convolutions (see Section 6.2.3 for a discussion of what this implies).

Repeat from Step 2 until a flux limit is reached. At any stage in the minor cycle, additional constraints can be placed on the model image by calculating the spectral-index and/or curvature images and discarding unrealistic values.

5. **Predict** : Once the minor cycle flux limit is reached, the current best estimates of the spectral coefficient images are used to predict multi-frequency model visibilities.

$$V_v^{model} = \sum_{t=0}^{N_t-1} [W_t^{mfs}] [SG^{pc}F][I_t^{pc}]^{-1} I_t^{model} \quad (6.42)$$

Residual visibilities are computed for each frequency as $\vec{V}_v^{res} = \vec{V}_v^{corr} - \vec{V}_v^{model}$ and then processed as in Step 1 to construct the next residual image.

Repeat from Step 2 until the residuals satisfy a statistically derived stopping criterion.

Restoration : After convergence, the model spectral coefficient images can be interpreted in different ways. If applicable, the final image products can then be smoothed with the restoring beam and the residuals are added back in. Some forms applicable for radio astronomy are described below as additional operations that need to be performed on the model images.

1. The most obvious data products are the spectral-coefficient images themselves, which can be directly smoothed by the restoring beam. The residual images that are added back in should be the principal solution computed from the final residuals, to ensure that any undeconvolved flux has the right flux values.
2. For the study of broad-band radio emission, the spectral coefficients can be interpreted in terms of a power law in frequency with varying index (as described in Section 6.2.1). The data products are images of the reference-frequency flux $\vec{I}_{\nu_0}^{sky}$, the spectral-index \vec{I}^α and the spectral curvature \vec{I}^β .

$$\vec{I}_{\nu_0}^{sky} = \vec{I}_0^{model} \quad (6.43)$$

$$\vec{I}^\alpha = \frac{\vec{I}_1^{model}}{\vec{I}_0^{model}} \quad (6.44)$$

$$\vec{I}^B = \frac{\vec{I}_2^{model}}{\vec{I}_0^{model}} - \frac{\vec{I}^\alpha(\vec{I}^\alpha - 1)}{2} \quad (6.45)$$

Spectral index and curvature images can be calculated only in regions where the values in \vec{I}_0^{model} are above a chosen threshold. In this case, it is appropriate to smooth the final $\vec{I}_{v_0}^{sky}$ image with a restoring beam, but not the spectral index or curvature images.

3. An image cube can be constructed by evaluating the spectral polynomial *via* Eqn. 6.25 for each frequency. This form of data product is useful for sources whose emission is not well modeled by a power law, but is a smooth polynomial in frequency. Band-limited signals that taper off smoothly in frequency are one example.
4. An image of the continuum flux can be constructed by evaluating and summing up the flux at all frequencies. Note that this continuum image is different from the reference-frequency image which represents the flux measured at only one frequency.

Algorithm 4: MF-CLEAN with Cotton-Schwab Major/Minor Cycles

Data: Calibrated visibilities : $\vec{V}_v^{corr} \forall v$
Data: uv-sampling functions : $S_v \forall v$
Data: Image noise threshold and loop gain σ_{thr}, g_s
Result: Model Coefficient Images : $\vec{I}_q^{model} \forall q \in \{0, N_t - 1\}$

```

1 foreach  $t \in \{0, N_t - 1\}, q \in \{t, N_t - 1\}$  do
2   | Compute the spectral PSF  $\vec{I}_{tq}^{psf}$ 
3 end
4 Construct  $[H^{peak}]$  and  $[H^{peak^{-1}}]$  with  $H_{t,q}^{peak} = mid(\vec{I}_{t,q}^{psf})$ 
5 Measure the peak  $\vec{I}_{00}^{psf}$  sidelobe  $f_{sidelobe}$ 
6 Initialize the model  $\vec{I}_t^{model}$  for all  $t \in \{0, N_t - 1\}$ 
7 repeat                                     /* Major Cycle */
8   foreach  $t \in \{0, N_t\}$  do
9     | Compute the residual image  $\vec{I}_t^{res}$ 
10    end
11   Calculate a Flux-Limit for from  $\vec{I}_0^{res}$  :  $f_{limit}$ 
12   repeat                                     /* Minor Cycle */
13     if Peak of  $\vec{I}_0^{res} > 10 \sigma_{thr}$  then
14       foreach pixel do
15         | Construct  $\vec{I}^{pix,dirty}$ , an  $N_t \times 1$  vector from  $\vec{I}_t^{res} \forall t \in \{0, N_t - 1\}$ 
16         | Compute principal solution  $\vec{I}^{sol} = [H^{peak^{-1}}] \vec{I}^{pix,dirty}$ 
17       end
18       Choose the solution vector at the location of the peak of  $\vec{I}_0^{sol}$ 
19     else
20       Find the location of the peak of  $\vec{I}_0^{res}$ 
21       Construct  $\vec{I}^{pix,dirty}$  from  $\vec{I}_t^{res} \forall t \in \{0, N_t - 1\}$ , at this location
22       Compute  $\vec{I}^{sol} = [H^{peak^{-1}}] \vec{I}^{pix,dirty}$  at this location
23     end
24     foreach  $t \in \{0, N_t - 1\}$  do
25       | Update the model image :  $\vec{I}_t^{model} = \vec{I}_t^{model} + g_s \vec{I}_t^{sol}$ 
26       | Update the residual image :  $\vec{I}_t^{res} = \vec{I}_t^{res} - g \sum_{q=0}^{N_t-1} [\vec{I}_{tq}^{psf} \star \vec{I}_t^{sol}]$ 
27     end
28   until Peak residual in  $\vec{I}_0^{res} < f_{limit}$ 
29   Compute model visibilities  $\vec{V}_v^{model}$  from  $\vec{I}_t^{model} \forall t \in \{0, N_t - 1\}$ 
30   Compute a new residual image  $\vec{I}^{res}$  from  $\vec{V}_v^{corr} - \vec{V}_v^{model}$ 
31 until Peak residual in  $\vec{I}_0^{res} < \sigma_{thr}$ 
32 Calculate spectral index and curvature images from  $\vec{I}_q^{model}$ , and restore them

```

6.2.3 Difference with the Sault-Wieringa (SW-MFCLEAN) algorithm

The SW-MFCLEAN algorithm described in [Sault and Wieringa \[1994\]](#) follows the theory in the previous section for $N_t = 2$, but its implementation follows a matched-filtering approach, using spectral PSFs $\vec{I}_{t=0,q}^{psf} \forall q \in \{0, N_t - 1\}$ as the template functions. Formally, the matched filtering approach is exactly equal to the calculations shown in Eqns.[6.32](#) to [6.34](#) only under the conditions that there is no overlap on the spatial frequency plane between measurements from different observing frequencies, and all measurements are weighted equally across the spatial-frequency plane (uniform weighting). In other cases (when there is overlap between frequencies on the *uv*-plane), it can be shown to lead to errors in estimating the spectral coefficients, especially when there is extended emission in the image.

The following is a simple way to state the problem. Consider a spatial-frequency grid cell onto which measurements from two different baselines and frequencies map. Let V_1, V_2 be the measured visibilities at two frequencies 1, 2 and let w_1, w_2 be their Taylor-weights. A matched-filtering approach calculates $(w_1 + w_2)(V_1 + V_2)$, whereas Eqns.[6.32](#) to [6.34](#) require the computation of $(w_1 V_1) + (w_2 V_2)$. The two are equivalent only for flat spectrum sources where $V_1 = V_2$ or when there is no such overlap between measurements from different observing frequencies (V_1 and V_2 map to different spatial frequencies).

The SW-MFCLEAN algorithm was initially developed for the ATCA telescope, an East-West array of antennas with circular *uv*-coverage patterns and minimal spatial-frequency overlap across channels. This matched filtering approach therefore worked well. However, when applied to data from the VLA (where *uv*-tracks intersect each other and there is considerable spatial-frequency overlap), numerical instabilities limited the fidelity of the final image, especially with extended emission. Changing the computations to those described for MFS deconvolution in Section [6.2.2.5](#) eliminated this instability (determined using simulated VLA data).

6.2.3.1 Differences

In the SW-MFCLEAN algorithm, the Hessian block kernels and dirty images are computed *via* FFT-based convolutions in which gridded Taylor-weights are multiplied with gridded visibilities : $(w_1 + w_2)(V_1 + V_2)$.

$$\vec{I}_{t,q}^{psf,sw} = \vec{I}_t^{psf} \star \vec{I}_q^{psf} \quad \text{where} \quad \vec{I}_x^{psf} = [F^\dagger S^\dagger W_x^{mfs} W_x^{im}] \vec{1} \quad \text{for } x = t, q \quad (6.46)$$

$$\vec{I}_t^{dirty,sw} = \vec{I}_t^{psf} \star \vec{I}_t^{dirty} \quad \text{where} \quad \vec{I}_t^{dirty} = [F^\dagger S^\dagger W_t^{im}] \vec{V}^{corr} \quad (6.47)$$

According to Eqns.[6.32](#) to [6.34](#) (MFCLEAN algorithm), the Hessian block kernels and dirty images are to be computed by multiplying the visibility measurements with the Taylor-weights before gridding the result : $(w_1 V_1) + (w_2 V_2)$.

$$\vec{I}_{t,q}^{psf} = [F^\dagger S^\dagger W_t^{mfs} W_q^{mfs} W_q^{im}] \vec{1} \quad (6.48)$$

$$\vec{I}_t^{\text{dirty}} = [F^\dagger S^\dagger W_t^{mfs} W_p^{im}] \vec{V}^{\text{corr}} \quad (6.49)$$

6.2.3.2 Conditions for equality

The two methods listed above (Eqns. 6.46, 6.47 and Eqns. 6.48, 6.49) are equivalent only under certain conditions. Consider Eqn. 6.33 for $[H_{t,q}]$. Iff $[S^\dagger] = [S^\dagger S S^\dagger]$ and $[S] = [S S^\dagger S]$, then $[S^\dagger]$ can be replaced by $[S^\dagger S S^\dagger]$. Further, $[S S^\dagger]$ and $[W_t^{mfs\dagger}]$ are both diagonal matrices of size $nN_c \times nN_c$ and therefore commute. In this case, $[H_{t,q}]$ becomes

$$[H_{t,q}] = [F^\dagger S^\dagger W_t^{mfs\dagger} W_p^{im} W_p^{mfs} S F] \quad (6.50)$$

$$\Rightarrow \vec{I}_{tq}^{\text{psf}} = I_t^{\text{psf}} \star I^{\text{psf}} \star I_q^{\text{psf}} \quad (6.51)$$

$$\text{where } I^{\text{psf}} = [F^\dagger S^\dagger W_p^{im}] \vec{1} \quad \text{and} \quad I_t^{\text{psf}} = [F^\dagger S^\dagger W_t^{mfs}] \vec{1}$$

This is still not the same as Eqn. 6.46 which has two instances of $[W_p^{im}]$. Therefore, only when $[W_p^{im}]$ is an identity matrix (equally weighted visibilities) will the kernel functions from both methods be identical $\vec{I}_{t,q}^{\text{psf}} = \vec{I}_{t,q}^{\text{psf}, \text{sw}}$. A similar argument holds for the dirty images. The restriction of $[S^\dagger] = [S^\dagger S S^\dagger]$ and $[S] = [S S^\dagger S]$ implies that each row and column in $[S]$ has only one 1, with the rest being 0. Since $[S]$ has dimensions $nN_c \times m$, the maximum number of non-zero elements must be m . Therefore, any of the m discrete spatial frequencies cannot be measured at more than one baseline or frequency channel. However, consider the $m \times m$ diagonal matrix of gridded imaging weights $[W_v^G] = [S_v^\dagger W_v^{im} S_v]$ per frequency channel. A projection operator $[S_v^G]$ of shape $(m \times m)$ can be constructed for each frequency channel, with each diagonal element corresponding to one spatial frequency grid cell. Measurements from multiple baselines that map onto the same spatial frequency grid cell are treated as a single measurement in $[S_v^G]$, with an increased weight in $[W_v^G]$. The use of uniform weighting will flatten out $[W_v^G]$ as required for equality with Eqn. 6.46. Written this way, with multiple frequencies, $[S_v^G]$ has dimensions $mN_c \times m$, and the restriction of $[S^\dagger] = [S^\dagger S S^\dagger]$ and $[S] = [S S^\dagger S]$ means that any spatial frequency must not be measured in more than one frequency channel. Therefore, a pure matched-filtering approach is strictly valid only for uniform weighting and when all filled spatial frequency grid cells contain measurements from only one frequency channel.

6.2.4 Accuracy of multi-frequency deconvolution

This section illustrates some of the errors that arise when an N_t -term Taylor polynomial is used to model a power-law spectrum during multi-frequency synthesis imaging. The metrics used to evaluate these errors are the magnitude of the residuals (or remainder) and the absolute errors on the values of physical quantities derived from the fitted coefficients (intensity at a reference frequency, spectral index and curvature). Section 6.2.4.1 compares the accuracy of the derived value of α when different functional forms are used to model a power-law spectrum. These errors were measured for a fixed total bandwidth and for different signal-to-noise ratios. Section 6.2.4.2 shows how the peak residuals and errors on I_{v_0} , α and β vary with the order of the Taylor polynomial used with multi-frequency deconvolution. These errors were measured for different total bandwidths and a fixed signal-to-noise ratio.

These trends are meant to be used as guidelines when choosing parameters during multi-frequency deconvolution. However, note that this section only shows measured errors for a few simple examples and makes no attempt to estimate or predict these errors for a generic data set or type of spectrum¹⁶.

6.2.4.1 Accuracy of power-law parameters derived from a polynomial

The errors on the polynomial coefficients and quantities derived from them will depend on the number of measurements of the spectrum, the signal-to-noise ratio of the measurements, and their distribution across a frequency range. They will also depend on the order of the polynomial used in the approximation. Although the physical parameters I_{v_0} , α and β can be obtained from the first three coefficients of a Taylor expansion of a power-law with varying index (Eqns.6.22 and 6.24), a higher order polynomial may be required during the fitting process to improve the accuracy of the first three coefficients¹⁷. In the case of very noisy spectra, errors can also arise from attempting to use too many terms in the polynomial fit.

Figure 6.7 illustrates the above trends for the value of α derived from a polynomial fit to a spectrum constructed via Eqn. 6.20 ($I_0^{true} = 10.0$, $\alpha^{true} = -1.5$, $\beta^{true} = -0.5$, $v_0 = 2.4\text{GHz}$) and evaluated between 1-4 GHz. Gaussian random noise was added to give measurement signal-to-noise ratios of 100, 10 and 1 for three such spectra. These spectra were fitted using a linear least-squares method on two series expansions (Eqns.6.22,6.24) for different numbers of terms in the series $N = 2, 3, 4, 5$, and also by a non-linear least-squares method to fit α and β directly. The plots show the error on the derived spectral index $\delta\alpha = \alpha^{fitted} - \alpha^{true}$ for each case.

¹⁶This dissertation does not contain a formal error analysis of the multi-frequency synthesis deconvolution algorithm for different basis functions, bandwidths and signal-to-noise ratios. Such an analysis will ultimately be required in order to prescribe a set of rules for a generic observation, and is work in progress.

¹⁷Conway et al. [1990] comment on a bias that occurs with a 2-term Taylor expansion, due to the use of a polynomial of insufficient order to model an exponential.

Noticeable trends (based on $\delta\alpha$) are listed below.

1. For high SNR, higher order fits give better results. For low SNR, higher order fits give larger errors.
2. In most cases, a Taylor expansion of a power law about $\alpha = 0, \beta = 0$ is a better choice than a Taylor expansion about $\nu = \nu_0$.
3. For spectra between 1 and 4 GHz with $\alpha \approx -1.5 \pm 0.4$, a 3rd or 4th order Taylor expansion (either form) is most appropriate.

These trends can be used to choose the spectral basis function and number of terms N_t to be used in the multi-frequency deconvolution algorithm (Section 6.2.2.5), based on *a priori* knowledge of the average spectral index and the signal-to-noise ratio of the measurements. When there are both high and low signal-to-noise sources, a multi-stage approach using different values of N_t might be required. For example, deconvolution runs can begin with $N_t > 3$ but once the peak residual reaches 10σ , a switch to $N_t = 2$ might be beneficial (note that this situation has not yet been tested).

6.2.4.2 Peak Residuals

This section shows an example of the errors obtained when the order of the polynomial chosen for imaging is not sufficient to model the power-law spectrum of the source. EVLA datasets (8 hour synthesis) were simulated for 5 different frequency ranges around 2.0 GHz. The sky brightness distribution used for the simulation was one point source whose flux is 1.0 Jy and spectral index is -1.0 with no spectral curvature. The bandwidth ratios¹⁸ for these 5 datasets were 100%(3:1), 66%(2:1), 50%(1.67:1), 25%(1.28:1), 10%(1.1:1).

Figure 6.8 shows the measured peak residuals and absolute measured errors on I_{ν_0}, α, β when these datasets were imaged using multi-frequency deconvolution with $N_t = 1$ to $N_t = 7$ and a linear spectral basis (Eqn. 6.25). All these datasets were imaged using a maximum of 10 iterations, a loop-gain of 1.0, natural weighting and a flux threshold of 1.0 μ Jy. No noise was added to these simulations (in order to isolate and measure numerical errors due to the spectral fits). Peak residuals were measured over the entire 0th order residual image, and errors on I_{ν_0}, α, β were computed at the location of the point source by taking differences with the ideal values of $I_{\nu_0} = 1.0, \alpha = -1.0, \beta = 0.0$.

Noticeable trends from these plots are listed below.

¹⁸There are two definitions of bandwidth ratio that are used in radio interferometry. One is the ratio of the highest to the lowest frequency in the band, and is denoted as $\nu_{high} : \nu_{low}$. Another definition is the ratio of the total bandwidth to the central frequency $(\nu_{high} - \nu_{low})/\nu_{mid}$ expressed as a percentage. For example, the bandwidth ratio for $\nu_{low} = 1.0$ GHz, $\nu_{high} = 2.0$ GHz is 2 : 1 and 66%.

1. All errors appear to decrease exponentially (linearly in log-space) as a function of increasing order of the polynomial, and as a function of decreasing total bandwidth. For very narrow bandwidths, the use of high-order polynomials increases the error.
2. The peak residuals are much smaller than the error incurred on the peak source flux at the reference frequency I_{ν_0} and the errors on α and β .
3. As an example, for a 2:1 bandwidth ratio, a source with spectral index = -1.0, and $N_t = 4$, the achievable dynamic range (measured as the ratio of the peak flux to the off-source peak residual) is about 10^5 , the error on the peak flux at the reference frequency is 1 part in 10^3 , and the absolute errors on α and β are 10^{-2} and 10^{-1} respectively.

Note that these trends are based on one simple example, and further analysis is required to understand the source of these errors and assess how they vary as a function of α and β . [Conway et al. \[1990\]](#) suggest that for an $(N_t - 1)$ -order polynomial, the peak residuals proportional to the product of α and the peak sidelobe level of the next higher order N_t^{th} spectral PSF. However, the results of the above tests do not follow this rule for all bandwidth ratios. Further work is required to (a) understand these errors in terms of signal-to-noise and in the presence of deconvolution errors and (b) be able to predict limiting dynamic ranges and error-bars on α and β .

Note that all the code implementations for this dissertation use the linear expansion given by Eqn. 6.25 (a polynomial in I vs $(\nu - \nu_0)/\nu_0$ space) to model an arbitrary spectrum. However, in the case of a power-law, a logarithmic expansion given by Eqn. 6.26 (a polynomial in I vs $\log(\nu/\nu_0)$ space) might need fewer terms than the linear expansion to model a power-law spectrum and yield better results. [Conway et al. \[1990\]](#) state that the logarithmic expansion has better convergence properties than the linear expansion when $\alpha \ll 1$, but this is yet to be tested for arbitrary values of α . Further, for given values of α and β , the radius of convergence of each series expansion defines a maximum bandwidth that it can be used with. Further work is required to do a formal comparison between these two sets of spectral basis functions and their convergence properties when applied to arbitrary spectral shapes.

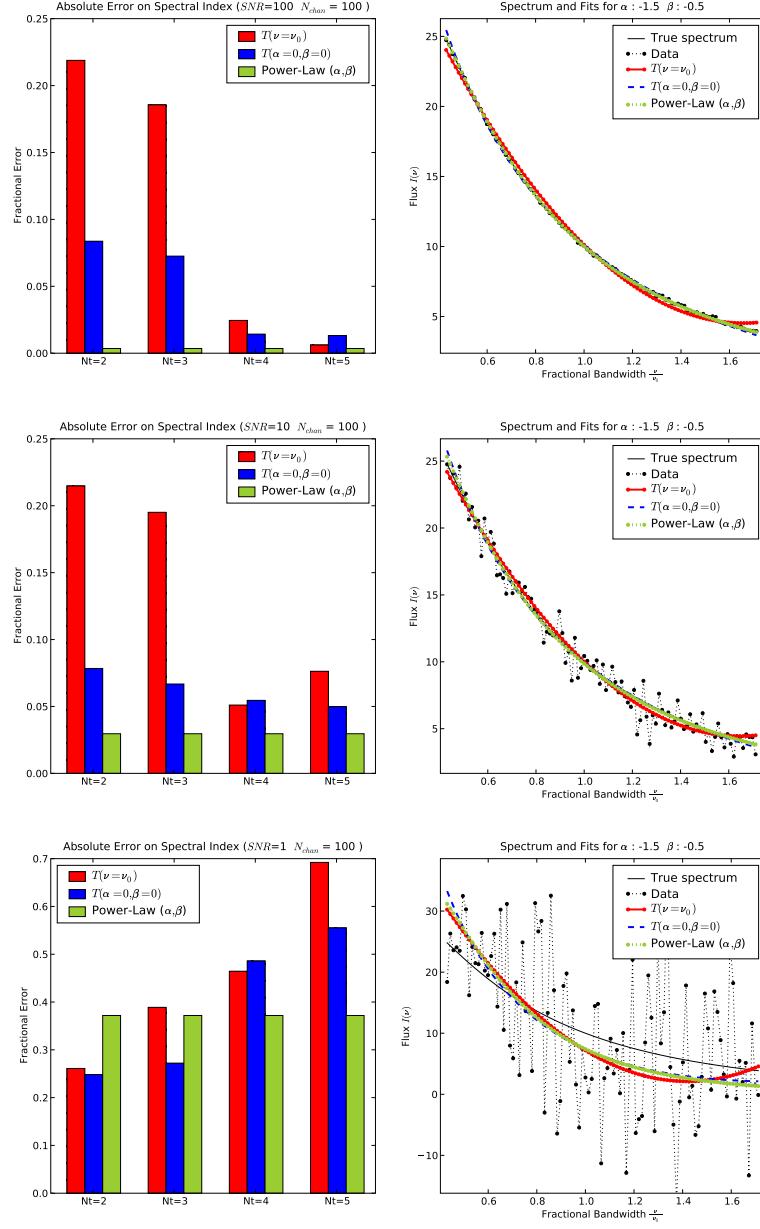


Figure 6.7: These plots show the average error on the fitted spectral index ($\delta\alpha = \alpha^{fitted} - \alpha^{true}$) from 100 noisy measurements of a power-law spectrum defined by $I_0^{true} = 10.0$, $\alpha^{true} = -1.5$, $\beta^{true} = -0.5$. The rows represent different signal-to-noise ratios (Top : 100, Middle : 10, Bottom : 1). The left column shows the average $\delta\alpha$ with $N_t = 2, 3, 4, 5$ terms in the series, for three different functional forms (Red/Left : $T(\nu = \nu_0)$: Taylor expansion of I_ν about ν_0 , Blue/Middle : $T(\alpha = 0, \beta = 0)$: Taylor expansion of I_ν about $\alpha = 0, \beta = 0$, Green/Right : Power Law with varying index). The right column shows the corresponding spectra for $N_t = 3$. Noticeable trends are (a) For high SNR, higher order fits give better results. (b) For low SNR, higher order fits give larger errors. (c) In most cases, a Taylor expansion about $\alpha = 0, \beta = 0$ is a better choice than an expansion about ν_0 . (d) For spectra between 1 and 4 GHz with $\alpha \approx -1.5 \pm 0.4$, a 3rd or 4th order Taylor expansion is most appropriate.

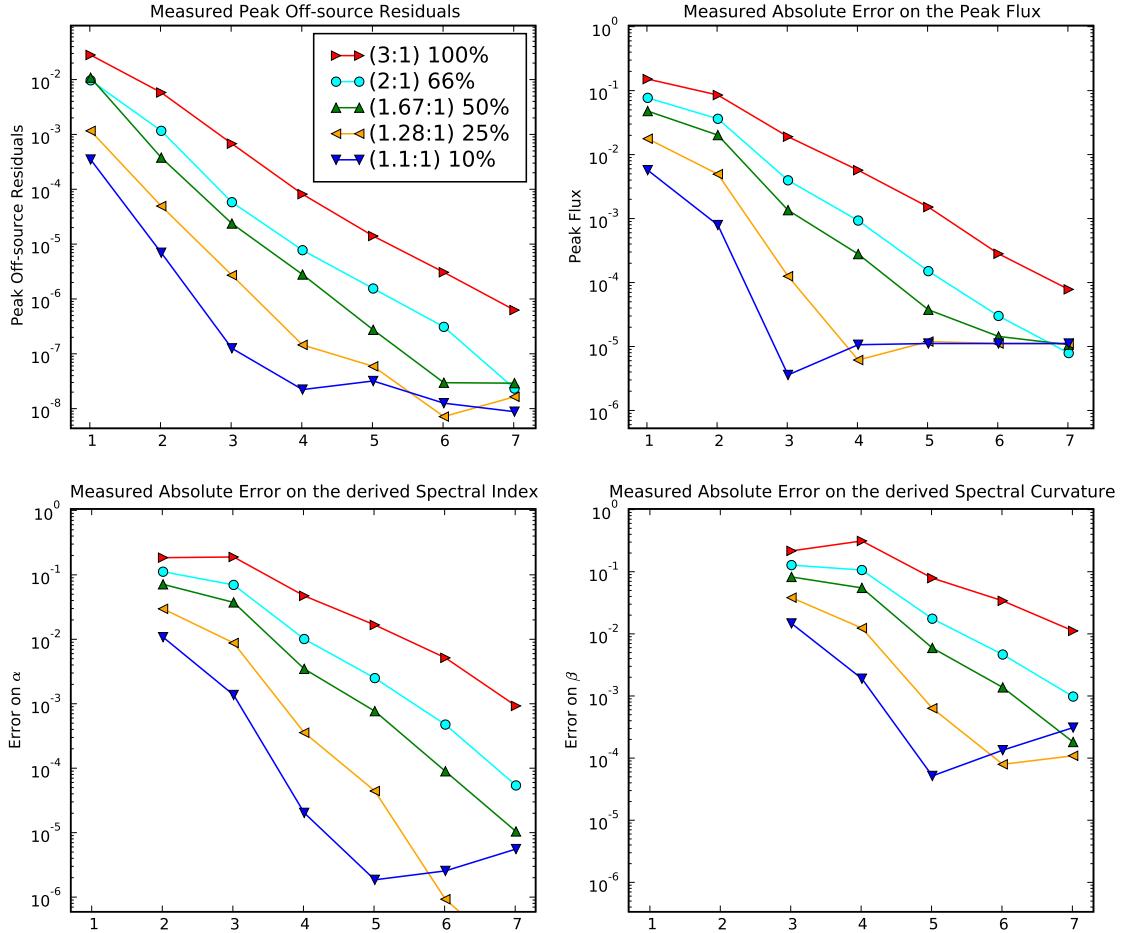


Figure 6.8: Peak Residuals and Errors for MFS with different values of N_t : These plots show the measured peak residuals (top left) and the errors on I_{ν_0} (top right), α (bottom left), and β (bottom right) when a point-source of flux 1.0 Jy and $\alpha=-1.0$ was imaged using Taylor polynomials of different orders ($N_t = 1 - 7$) and a linear spectral basis (Eqn. 6.26). This simulation was done with EVLA *uv*-coverages (for an 8 hour synthesis run) and 100%, 66%, 50%, 25% and 10% fractional bandwidths, with a reference frequency of 2.0 GHz. No noise was added to these simulations. All runs used a loop-gain of 1.0, used natural weighting, and were terminated after either 10 iterations or a flux threshold of $1\mu\text{Jy}$. The x-axis of all these plots show the value of N_t used for the simulation. Plots for α and β begin from $N_t = 2$ and $N_t = 3$ respectively because at least that many terms are required to calculate these derived quantities. Noticeable trends from these plots are (a) The peak residuals decrease by about a factor of 8 with each increase of 1 more polynomial coefficient. (b) The peak residuals are larger for larger fractional bandwidths. (c) The errors on I_{ν_0} , α , β are larger than the peak residuals, but they too decrease with increasing N_t . For very narrow bandwidths, the use of a very high-order polynomial increases the error. For a 2:1 bandwidth ratio, a spectral index of -1.0 and very high signal-to-noise, a 5th or 6th order Taylor expansion is most appropriate (when a linear spectral basis is used).

CHAPTER 7

MULTI-SCALE MULTI-FREQUENCY SYNTHESIS IMAGING

This chapter deals with the combination of the multi-scale, multi-frequency and wide-field imaging algorithms described in chapters 6 and 4 to derive a method that forms a multi-scale reconstruction of the broad-band sky brightness distribution while accounting for the frequency-dependence of the antenna field-of-view. Section 7.1 first motivates the need of using a multi-scale image model along with multi-frequency synthesis, and then describes a combined multi-scale multi-frequency deconvolution algorithm. Section 7.2 describes an extension of this algorithm for wide-field imaging in which the frequency dependence of the primary beam is included and corrected for during image reconstruction. Chapter 8 later shows imaging results using these algorithms, contains a discussion about error estimation and shows a set of examples that test the feasibility of multi-scale multi-frequency image reconstruction for moderately resolved sources, very large spatial scales, overlapping flux components with different spectra, and band-limited signals.

7.1 Multi-Scale Multi-Frequency Deconvolution

We begin with a discussion of how well we can reconstruct both spatial and spectral information from an incomplete set of visibility samples at multiple observing frequencies and describe how our choice of a flux model influences the image reconstruction process when each observing frequency measures a different set of spatial frequencies.

The spatial frequencies sampled at each observing frequency ν are between $u_{min} = \frac{\nu}{c} b_{min}$ and $u_{max} = \frac{\nu}{c} b_{max}$, where u is used here as a generic label for the uv -distance¹ and b represents the length of the baseline vector (in units of meters) projected onto the plane perpendicular to the direction of the source. The range of spatial frequencies between u_{min} at ν_{max} and u_{max} at ν_{min} represents the region that is sampled at all frequencies in the band. Within this region, both spatial and spectral information is measured in comparable detail and there is sufficient information to reconstruct them both. The spatial frequencies outside this region are sampled only by a fraction of the band and the accuracy of a broad-band reconstruction depends on how well the spectral and spatial structure are constrained by an appropriate choice of a flux model.

A few examples are used to illustrate the importance of an appropriate flux model.

¹ The uv -distance is defined as $\sqrt{u^2 + v^2}$ and is the radial distance of the spatial frequency measured by the baseline from the origin of the uv -plane, in units of wavelength λ .

1. A compact, unresolved source with spectral structure is measured as a point source at all frequencies, and u_{max} at ν_{max} gives the maximum angular resolution at which this source can be imaged. Since the visibility function of a point source is flat across the entire spatial frequency plane, its spectrum is adequately sampled by the multi-frequency measurements. Using a flux model in which each source is a δ -function with a smooth polynomial spectrum, it is possible to reconstruct the spectral structure of the source at the maximum possible angular resolution.
2. For resolved sources with spectral structure, the accuracy of the reconstruction across all spatial scales between u_{min} at ν_{min} and u_{max} at ν_{max} depends on an appropriate choice of flux model, and the constraints that it provides. For example, a source emitting broad-band synchrotron radiation can be described by a fixed brightness distribution at one frequency with a power-law spectrum associated with each location. Images can be made at the maximum angular resolution (given by u_{max} at ν_{max}) with the assumption that different observing frequencies probe the same spatial structure but measure different amplitudes (usually a valid assumption). This constraint is strong enough to correctly reconstruct even moderately resolved sources that are completely unresolved at the low end of the band but resolved at the higher end. On the other hand, a source whose structure itself changes across the band would break the above assumption. One example is with multi-frequency observations of solar magnetic loops where the different frequencies probe different layers in the upper chromosphere and can have very different structures. In this case, a complete reconstruction would be possible only in the region of overlapping spatial frequencies (between u_{min} at ν_{min} and u_{max} at ν_{max}), unless the flux model includes constraints that bias the solution towards one appropriate for such sources.
3. The lower end of the spatial frequency range presents a different problem. The size of the central hole in the uv -coverage increases with frequency. Spectra are not measured adequately for emission whose visibility function is non-zero only below u_{min} at ν_{max} and a flat-spectrum large-scale source can be indistinguishable from a relatively smaller source with a steep spectrum. Additional constraints in the form of total-flux values for each frequency may be required for an accurate reconstruction.

To summarize, just as standard interferometric image reconstruction uses *a priori* information about the spatial structure of the sky to estimate the visibility function in unmeasured regions of the uv -plane, multi-frequency image reconstruction algorithms need to use *a priori* information about the spectral as well as spatial structure of the sky brightness. By combining such models with the known frequency-dependence of the spatial-frequency coverage it is possible to reconstruct the broad-band sky brightness distribution from incomplete spectral and spatial-frequency sampling.

7.1.1 Multi-Scale Wide-Band Image model

The multi-frequency synthesis algorithm described in section 6.2 models the sky brightness distribution as a collection of δ -functions with power-law spectra. This parameterization provides strong enough constraints on the image reconstruction process when applied to fields of isolated point sources. However, when it is applied to fields containing extended emission it leads to errors in the reconstruction, similar to the single-frequency case where large scale emission is broken into a collection of compact flux components of the size of the telescope angular resolution. With multi-frequency synthesis, these errors are enhanced mainly due to error propagation during the calculation of derived quantities such as spectral index and curvature as the ratios of two noisy images each containing deconvolution errors. It therefore becomes important to use a flux model and image reconstruction algorithm that can ensure smoothness in the reconstruction and improve the fidelity of the coefficient images used to calculate these derived quantities. One option is to parameterize the sky brightness distribution in the multi-scale basis described in section 6.1 and associate a polynomial spectrum with each flux component. A region of emission in which the spectrum varies with position will be modeled as a sum of wide-band flux components and the reconstruction algorithm would simultaneously reconstruct the spatial and spectral structure of the source in terms of these parameters.

For multi-scale and multi-frequency deconvolution, the image flux model at each frequency can be written as a linear sum of coefficient images at different spatial scales. This is a combination of the multi-scale and multi-frequency image models described in sections 6.1.1 and 6.2.1.

$$\vec{I}_\nu^{model} = \sum_{t=0}^{N_t} \sum_{s=0}^{N_s} w_\nu^t \left[\vec{I}_s^{shp} \star \vec{I}_s^{sky} \right] \quad \text{where} \quad w_\nu^t = \left(\frac{\nu - \nu_0}{\nu_0} \right)^t \quad (7.1)$$

Here, N_s is the number of discrete spatial scales used to represent the image and N_t is the order of the series expansion of the spectrum. \vec{I}_s^{sky} represents a collection of δ -functions that describe the locations and integrated amplitudes of flux components of scale s in the image of the t^{th} series coefficient. \vec{I}_s^{shp} is a tapered truncated parabola whose width is given by s (introduced in section 6.1.2 for multi-scale deconvolution).

7.1.2 Imaging Equations and Block Deconvolution

This section combines the multi-scale multi-frequency image model with the standard measurement equations and then explores the structure of the normal equations. This is followed by a description of the principal solution and an iterative block deconvolution algorithm.

7.1.2.1 Measurement Equations

Visibilities are measured at a set of N_c observing frequencies, and calibrated using a model derived from a source with known structure and spectrum. There are n visibilities measured at each frequency. The single frequency and multi-frequency measurement equations are given below.

$$\vec{V}_v^{obs} = \sum_{t=0}^{N_t} \sum_{s=0}^{N_s} w_v^t [S_v][T_s][F] \vec{I}_t^{sky} \quad (7.2)$$

$$\vec{V}^{obs} = \sum_{t=0}^{N_t} \sum_{s=0}^{N_s} [W_t^{mfs}][S][T_s][F] \vec{I}_t^{sky} \quad (7.3)$$

where $[W_t^{mfs}]$ is a diagonal $nN_c \times nN_c$ matrix of weights, comprised of N_c blocks each of size $n \times n$ for each frequency channel (v). The multi-frequency uv -coverage of the synthesis array is represented by $[S_{nN_c \times m}]$. The image-domain convolution with \vec{I}_s^{shp} is written as a spatial-frequency taper function $[T_s]_{m \times m} = \text{diag}([F]\vec{I}_s^{shp})$.

The full measurement matrix ([A] in Eqn. 2.20) therefore has the shape $nN_c \times mN_s N_t$, which when multiplied by the set of $N_s N_t$ model sky vectors each of shape $m \times 1$, produces nN_c visibilities.

For $N_t = 3, N_s = 2$ the measurement equations can be written as follows, in block matrix form. The subscript p denotes the p^{th} spatial scale and the subscript q denotes the q^{th} Taylor coefficient of the spectrum polynomial.

$$\left[\begin{array}{cccccc} \left[A_{p=0} \right] & \left[A_{p=0} \right] & \left[A_{p=0} \right] & \left[A_{p=1} \right] & \left[A_{p=1} \right] & \left[A_{p=1} \right] \end{array} \right] \left[\begin{array}{c} \vec{I}_{p=0}^{ky} \\ \vec{I}_{p=0}^{ky} \\ \vec{I}_{p=0}^{ky} \\ \vec{I}_{p=1}^{ky} \\ \vec{I}_{p=1}^{ky} \\ \vec{I}_{p=1}^{ky} \end{array} \right] = \vec{V}^{obs} \quad (7.4)$$

where $\left[A_p \right] = [W_q^{mfs}][S][T_p][F]$
for $p \in \{0, N_s - 1\}$ and $q \in \{0, N_t - 1\}$

7.1.2.2 Normal equations

The normal equations in block matrix form for the same example ($N_t = 3, N_s = 2$) become

$$\left[\begin{array}{ccccccc} \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[\vec{I}_{p=0,q=0}^{sky} \right] \\ \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[\vec{I}_{p=0,q=1}^{sky} \right] \\ \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[\vec{I}_{p=0,q=2}^{sky} \right] \\ \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[\vec{I}_{p=1,q=0}^{dirty} \right] \\ \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[\vec{I}_{s=0,t=0}^{dirty} \right] \\ \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[\vec{I}_{s=0,t=1}^{dirty} \right] \\ \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=0} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[H_{s=0,p=1} \right] & \left[\vec{I}_{s=0,t=2}^{dirty} \right] \\ \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[\vec{I}_{p=1,q=0}^{sky} \right] \\ \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[\vec{I}_{p=1,q=1}^{dirty} \right] \\ \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[\vec{I}_{s=1,t=0}^{dirty} \right] \\ \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[\vec{I}_{s=1,t=1}^{dirty} \right] \\ \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=0} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[H_{s=1,p=1} \right] & \left[\vec{I}_{s=1,t=2}^{dirty} \right] \end{array} \right] = \left[\begin{array}{c} \vec{I}_{p=0,q=0}^{sky} \\ \vec{I}_{p=0,q=1}^{sky} \\ \vec{I}_{p=0,q=2}^{sky} \\ \vec{I}_{p=1,q=0}^{sky} \\ \vec{I}_{p=1,q=1}^{sky} \\ \vec{I}_{p=1,q=2}^{sky} \\ \vec{I}_{s=0,t=0}^{dirty} \\ \vec{I}_{s=0,t=1}^{dirty} \\ \vec{I}_{s=0,t=2}^{dirty} \\ \vec{I}_{s=1,t=0}^{dirty} \\ \vec{I}_{s=1,t=1}^{dirty} \\ \vec{I}_{s=1,t=2}^{dirty} \end{array} \right] \quad (7.5)$$

When all scales and Taylor terms are combined, the full Hessian matrix contains $N_t N_s \times N_t N_s$ blocks each of size $m \times m$ and containing information from all frequency channels, and N_t Taylor coefficient images each of size $m \times 1$, for all N_s spatial scales. The indices s, p correspond to row and column indices for the multi-scale blocks, and the indices t, q correspond to row and column indices for the multi-frequency blocks.

The ordering of the rows and columns in Eqn. 7.5 was chosen such that the Hessian consists of $N_s \times N_s = 2 \times 2 = 4$ blocks (the four quadrants of the matrix). Each quadrant corresponds to one pair of spatial scales s, p . Within each quadrant, the $N_t \times N_t = 3 \times 3 = 9$ matrices correspond to various pairs of t, q (Taylor coefficient indices). This layout shows how the multi-scale and multi-frequency aspects of this imaging problem are combined and illustrates the dependencies between the spatial and spectral basis functions. Note that the 3×3 block in the top left quadrant corresponds to the entire Hessian matrix in Eqn. 6.31 (since \vec{I}_0^{shp} is a δ -function).

These equations can be written out row-by-row as follows.

$$\sum_{p=0}^{N_s-1} \sum_{q=0}^{N_t-1} \left[H_{s,p} \right] \vec{I}_q^{sky} = \vec{I}_t^{dirty} \quad \forall s \in \{0 \dots N_s - 1\}, t \in \{0 \dots N_t - 1\} \quad (7.6)$$

$$\left[H_{s,p} \right] = [F^\dagger T_s F][H_{t,q}][F^\dagger T_p F] \quad (7.7)$$

$$\vec{I}_t^{dirty} = [F^\dagger T_s F] \vec{I}_t^{dirty} = I_s^{shp} \star \vec{I}_t^{dirty} \quad (7.8)$$

Here, $[H_{t,q}]$ and \vec{I}_t^{dirty} are the multi-frequency Hessian blocks and dirty images as defined in Eqns. 6.33 and 6.34 (section 6.2.2) for the multi-frequency normal equations, $I_s^{shp} = [F^\dagger] \vec{T}_s$ are the scale basis functions as defined in section 6.1.2 for multi-scale imaging, and all convolutions shown here are implemented via Fourier transforms.

Each Hessian block $[H_{t,q}^{s,p}]$ is a convolution operator ($[H_{t,q}]$ is a convolution operator and convolution is associative and commutative). Its kernel $\vec{I}_{s,p}^{psf}$ is constructed by convolving each multi-frequency convolution kernels $I_{t,q}^{psf}$ (Eqn. 6.37) with a pair of scale basis functions (similar to Eqn. 6.9).

$$\vec{I}_{s,p}^{psf} = I_s^{shp} \star I_{t,q}^{psf} \star I_p^{shp} \quad (7.9)$$

These convolution kernels from the first row of Hessian blocks ($s = 0, t = 0, p \in \{0, N_s - 1\}, q \in \{0, N_t - 1\}$) represent the instrument's response functions to a flux component of unit total flux whose shape is given by the p^{th} scale basis function and whose spectrum is given by the q^{th} Taylor function.

7.1.2.3 Principal Solution

As described in the section 6.1.2.3 for multi-scale imaging and in section 6.2.2.3 for multi-frequency imaging, the principal solution here too, is found by using a diagonal approximation of each Hessian block to create $[H^{\text{peak}}]$ as an $N_t N_s \times N_t N_s$ element matrix, inverting it and applying it to all pixels of the dirty images, one pixel at a time.

When the principal solution is to be used within an iterative joint deconvolution, a few simplifying assumptions may be needed to trim computational costs. For a source with complicated spatial structure the number of distinct spatial scale basis functions is typically $N_s \approx 10$, and for power-law spectra with indices around -1.0, $N_t = 4$ or $N_t = 5$ terms in the series are required to accurately model the power law with a polynomial (across a 2:1 bandwidth). Therefore typically, $N_s N_t \approx 50$. Although the inversion of $[H^{\text{peak}}]$ may be tractable, the computational cost of a 50×50 matrix multiplication applied per pixel to a set of 10^6 pixels over a large number of iterations may be prohibitive in comparison to the numerical accuracy that this exact inversion provides. Several approximations can be made about the structure of $[H^{\text{peak}}]$ to simplify its inversion, and it is important to understand the numerical implications of these trade-offs.

One possible simplification is a block-diagonal approximation of the full Hessian (*i.e.* using only those blocks of the Hessian in Eqn. 7.5 for which $s = p$). This approximation ignores the cross-terms between spatial scales and assumes that the scale basis functions are orthogonal. Now, a multi-frequency principal solution (as described in section 6.2.2.3) can be done separately on each remaining $N_t \times N_t$ block, one spatial scale at a time ($\forall s \in \{0, N_s - 1\}$). The MFS principal solution for each scale s is given below (same as Eqn. 6.38 for each spatial scale s).

$$I_s^{\text{pix,psol}} = [H_s^{\text{peak}}]^{-1} I_s^{\text{pix,dirty}} \quad \text{for each pixel, and scale } s \quad (7.10)$$

Here, $[H_s^{\text{peak}}]$ is the s^{th} block (of size $N_t \times N_t$ on the diagonal of $[H^{\text{peak}}]$, and $I_s^{\text{pix,dirty}}$ is the $N_t \times 1$ vector constructed from $\vec{I}_s^{\text{dirty}} \forall t \in \{0, N_t - 1\}$. Note that the process of solving

the multi-frequency $[H^{peak}]$ for each scale automatically does a normalization across scales that corresponds to a diagonal approximation of the multi-scale H^{peak} (see sections 6.1.2.5 and 6.1.3 for alternate ways of computing the multi-scale solution).

The main result of using such an approximation while computing solutions for each pixel location and scale size is that the per-pixel matrix multiplications are much smaller. However, this approximation is never accurate because a set of tapered truncated paraboloids cannot form an orthogonal basis set. This inaccuracy is not a major problem while finding flux components because in the context of an iterative optimization the main penalty of taking slightly inaccurate steps is slower convergence and the resulting computational cost is usually offset by the smaller per-pixel operations to make this a useful trade-off. In other words, this approximation works not because the orthogonality assumption is valid, but because an iterative χ^2 -minimization process tolerates inaccurate steps during each iteration. The update step of the iterative deconvolution still needs to evaluate the full LHS of the normal equations while subtracting out a flux component.

7.1.2.4 Properties of $[H^{peak}]$

Some properties of $[H^{peak}]$ for multi-scale multi-frequency imaging are given below.

1. Each element of $[H^{peak}]$ is given by

$$H_{s,p}^{peak} = \text{mid} \left\{ \vec{I}_{t,q}^{psf} \right\} = \text{tr} \left[\sum_v w_v^{t+q} [T_s S_v^\dagger W_v^{im} S_v T_p] \right] \quad (7.11)$$

$$\forall s, p \in \{0 \dots N_s - 1\}, t, q \in \{0 \dots N_t - 1\}$$

2. The elements on the diagonal of $[H^{peak}]$ are a measure of the instrument's sensitivity to a flux component of unit total flux whose shape and spectrum is given by each of the $N_s N_t$ possible pairs of spatial and spectral basis functions.
3. The off-diagonal elements measure the orthogonality between the various basis functions, for the given uv -coverage and weighting scheme (compare with footnote 5 on page 94). For some choices of uv -coverage, frequency coverage, and scale size, the visibilities measured by the instrument for two different spatial scales can become hard to distinguish. The element of $[H^{peak}]$ corresponding to this combination could have a higher value, indicating that there is no information in the data and sampling pattern to distinguish between spatial or spectral structure while modeling the visibility function. The condition number of this matrix can be used as a metric to choose a suitable basis set.

7.1.2.5 Iterative Block Deconvolution (MS-MFS algorithm)

This section describes an iterative joint deconvolution process that produces a set of N_t Taylor coefficient images at N_s different spatial scales. The algorithm presented here (listed in two parts as Algorithm 5 on page 133 and Algorithm 6 on page 134) follows a set of steps similar to those for multi-scale and multi-frequency deconvolution (Chapter 6).

Pre-compute Hessian : Convolution kernels for all distinct blocks in the $N_s N_t \times N_s N_t$ Hessian are evaluated *via* Eqns. 7.9 and 6.37. All kernels are normalized by w_{sum} such that the peak of $\vec{I}_{0,0}^{psf}$ is unity, and the relative weights between Hessian blocks is preserved. This is equivalent to defining the weight image \vec{I}^{wt} as the diagonal of the $[H_{00}]$ Hessian block, and normalizing all the RHS vectors by it. A set of N_s matrices each of shape $N_t \times N_t$ and denoted as $[H_s^{peak}]$ are constructed from the diagonal blocks of the full Hessian (blocks for which $s = p$ in Eqn. 7.5). Their inverses are computed and stored in $[H_s^{peak^{-1}}]$.

Initialization : All $N_s N_t$ model images are initialized to zero (or an *a priori* model).

Major and minor cycles : The normal equations are solved iteratively by repeating steps 1 to 5 until some termination criterion is reached. Steps 1 and 5 form one major cycle, and repetitions of Steps 2 to 4 form the minor cycle.

1. **Compute RHS :** The dirty images for the RHS of the normal equations are computed *via* Eqn. 7.8 by first computing the multi-frequency dirty images and then smoothing them by the scale basis functions.
2. **Find a Flux Component :** The principal solution (as described in section 7.1.2.3) is computed for all pixels, one scale at a time *via* Eqn. 7.10. The principal solution consists of N_s sets of N_t Taylor-coefficient images. For iteration i , the N_t element solution set with the dominant $q = 0$ component across all scales and pixel locations is chosen the current flux component. Let the scale size for this set be p .

The result of this step is a set of N_t model images, each containing one δ -function that marks the location of the center of a flux component of shape $\vec{I}_{p,(i)}^{shp}$. The amplitudes of these N_t δ -functions are the Taylor coefficients that model the spectrum of the total flux of this component. Let these model images be denoted as $\left\{ \vec{I}_{q,(i)}^{model} \right\}; q \in [0, N_t]$.

3. **Update model images :** A single multi-scale model image is accumulated for each Taylor coefficient.

$$\vec{I}_q^{model} = \vec{I}_q^{model} + g \left(\vec{I}_{q,(i)}^{model} \star \vec{I}_{p,(i)}^{shp} \right) \quad \forall q \in [0, N_t] \quad (7.12)$$

where g is a loop-gain that takes on values between 0 and 1 and controls the step size for each iteration in the χ^2 -minimization process.

4. **Update RHS :** The RHS residual images are updated by evaluating and subtracting out the entire LHS of the normal equations. Since the chosen flux component corresponds to just one scale, the evaluation of the LHS is a summation over only Taylor terms.

$$\vec{I}_s^{res} = \vec{I}_s^{res} - g \left(\sum_{q_i=0}^{N_t-1} \left[\vec{I}_{s,p}^{psf} \star \vec{I}_{q,(i)}^{model} \right] \right) \quad (7.13)$$

Repeat from Step 2 until the minor-cycle flux limit is reached.

5. **Predict :** Model visibilities are computed from each Taylor-coefficient image, in the same way as in Eqn. 6.42 for multi-frequency imaging. Residual visibilities are computed as $\vec{V}_v^{res} = \vec{V}_v^{corr} - \vec{V}_v^{model}$.

Repeat from Step 1 until a global convergence criterion is satisfied.

Restoration : The final Taylor coefficient images are restored and interpreted in the same way as described in standard multi-frequency restoration (described at the end of section 6.2.2.5).

Algorithm 5: MS-MFS CLEAN : Set-up and major/minor cycle iterations

Data: calibrated visibilities : $\vec{V}_v^{corr} \forall v$
Data: uv-sampling function : $[S_v]$
Data: image noise threshold and loop gain σ_{thr}, g_s
Data: scale basis functions : $\vec{I}_s^{shp} \forall s \in \{0, N_s - 1\}$
Result: model coefficient images : $\vec{I}_q^m \forall q \in \{0, N_t - 1\}$
Result: spectral index and curvature : $\vec{I}_\alpha^m, \vec{I}_\beta^m$

```

1 for  $t \in \{0, N_t - 1\}, q \in \{t, N_t - 1\}$  do
2   | Compute the spectral PSF  $\vec{I}_{tq}^{psf}$ 
3   | for  $s \in \{0, N_s - 1\}, p \in \{s, N_s - 1\}$  do
4     |   | Compute the scale-spectral PSF  $\vec{I}_{sp}^{psf} = \vec{I}_s^{shp} \star \vec{I}_p^{shp} \star \vec{I}_{tq}^{psf}$ 
5   | end
6 end
7 for  $s \in \{0, N_s - 1\}$  do
8   | Construct  $[H_s^{peak}]$  from  $mid(\vec{I}_{s,s}^{psf})$  and compute  $[H_s^{peak^{-1}}]$ 
9 end
10 Initialize the model  $\vec{I}_t^m$  for all  $t \in \{0, N_t - 1\}$  and compute  $f_{sidelobe}$ 
11 repeat /* Major Cycle */
12   | for  $t \in \{0, N_t - 1\}$  do
13     |   | Compute the residual image  $\vec{I}_t^{res}$ 
14     |   | for  $s \in \{0, N_s - 1\}$  do
15       |     |   | Compute  $\vec{I}_{s,t}^{res} = \vec{I}_s^{shp} \star \vec{I}_t^{res}$ 
16     |   | end
17   | end
18 Calculate  $f_{limit}$  from  $\vec{I}_{0,0}^{res}$ 
19 repeat /* Minor Cycle */
20   | Compute  $I_q^m \forall q \in \{0, N_t - 1\}$  and update  $\vec{I}_{s,t}^{res} \forall s, t$  (Algorithm 6 on
      | the following page)
21 until Peak residual in  $\vec{I}_{0,0}^{res} < f_{limit}$ 
22 Compute model visibilities  $V_v^m$  from  $I_t^m \forall t \in \{0, N_t - 1\}$ 
23 Compute a new residual image  $I^{res}$  from residual visibilities  $V_v^{corr} - V_v^m$ 
24 until Peak residual in  $\vec{I}_{0,0}^{res} < \sigma_{thr}$ 
25 Calculate  $\vec{I}_{v_0}^m, \vec{I}_\alpha^m, \vec{I}_\beta^m$  from  $I_t^m \forall t \in \{0, N_t - 1\}$  and restore the results

```

Algorithm 6: MF-MFS CLEAN : minor cycle steps

Data: residual images : $\vec{I}_{s,t}^{res}$
Data: scale basis functions : \vec{I}_s^{shp}
Data: scale-Spectral PSFs : $\vec{I}_{sp}^{psf}_{tq} \quad \forall s \in \{0, N_s - 1\}, p \in \{s, N_s - 1\}$
Data: Hessian for each scale : $[H_s^{peak}] \quad \forall s \in \{0, N_s - 1\}$
Result: model coefficient images : $I_q^m \quad \forall q \in \{0, N_t - 1\}$
Result: updated residual images : $I_{s,t}^{res} \quad \forall s \in \{0, N_s - 1\}, t \in \{0, N_t - 1\}$

```

1 for  $s \in \{0, N_s - 1\}$  do
2   if Peak of  $\vec{I}_{s,0}^{res} > 10 \sigma_{thr}$  then
3     foreach pixel do
4       Construct  $I_s^{rhs}$ , an  $N_t \times 1$  vector from  $I_{s,t}^{res} \quad \forall t \in \{0, N_t - 1\}$ 
5       Compute principal solution  $I_s^{sol} = [H_s^{peak}]^{-1} I_s^{rhs}$ 
6     end
7     Choose  $I^{sol} = \max\{I_{t=0}^{sol}, \forall s \in \{0, N_s - 1\}\}$ 
8   else
9     Find the location of the peak in  $\vec{I}_{s,0}^{res}, \forall s \in \{0, N_s - 1\}$ 
10    Construct  $I_s^{rhs}$ , from  $I_{s,t}^{res}$  for the chosen  $s$ , at this location
11    Compute  $I^{sol} = [H_s^{peak}]^{-1} I_s^{rhs}$  at this location
12  end
13 end
14 for  $t \in \{0, N_t - 1\}$  do
15   Update the model image :  $I_t^m = I_t^m + g_s I_{s_i}^{shp} \star I_t^{sol}$ 
16   for  $s \in \{0, N_s - 1\}$  do
17     Update the residual image :  $I_{s,t}^{res} = I_{s,t}^{res} - g \sum_{p=0}^{N_s-1} \sum_{q=0}^{N_t-1} [I_{sp}^{psf}_{tq} \star I_q^{sol}]$ 
18   end
19 end

```

7.2 Correction of Frequency-Dependent Primary Beams

This section describes the combination of the multi-scale multi-frequency deconvolution algorithm described in the previous section (7.1) with methods to correct for direction-dependent instrumental effects (chapter 4); in particular, the antenna primary beam and its frequency dependence.

The angular size of the primary beam of the antenna decreases with an increase in observing frequency (see Fig.5.2, section 5.1.2). Sources away from the pointing center of the beam are attenuated by different amounts across the frequency band and this introduces artificial spectral structure into the measurements. To recover both spatial and spectral structure of the sky brightness across a large field of view, the frequency dependence of the primary beam must be modeled and removed during multi-frequency synthesis imaging.

Section 7.2.1 describes the multi-frequency primary beam as a polynomial in frequency (for each direction on the sky) and describes how the coefficients of this polynomial are computed. Section 7.2.2 then describes how this model is used within the multi-scale multi-frequency synthesis imaging and deconvolution framework. Algorithms 7 on page 153 and 8 on page 154 describe the complete wide-field multi-scale multi-frequency deconvolution algorithm. Chapter 8 later shows wide-field imaging results derived from applying this algorithm to simulated and real data.

7.2.1 Multi-Frequency Primary-Beam Model

Let us assume that the primary beam at each frequency ν (denoted as $\vec{P}_{b\nu}$) is known either from a theoretical model or *via* measurements. The spectrum of the multi-frequency primary beam can be described by a polynomial for every direction on the sky.

$$\vec{P}_{b\nu} = \sum_{q=0}^{N_t-1} w_\nu^q \vec{P}_{bq} \quad (7.14)$$

Here, \vec{P}_{bq} is the q^{th} coefficient of the polynomial representing the frequency dependent primary beam and w_ν^q are the corresponding basis functions (Taylor-weights). Given a set of single-frequency primary beams the coefficients of this $(N_t - 1)^{\text{th}}$ order polynomial can be computed as a least-squares solution by solving the following normal equations.

$$\sum_{q=0}^{N_t-1} \left\{ \sum_{\nu} w_\nu^{t+q} \right\} \vec{P}_{bq} = \sum_{\nu} w_\nu^t \vec{P}_{b\nu} \quad \forall t \in \{0, N_t - 1\} \quad (7.15)$$

The unknowns in this system are \vec{P}_{bq} . The weights are known and the RHS can be computed from the known single-frequency primary beams. A set of \vec{P}_{bq} can then be computed by solving this system of equations for every direction on the sky.

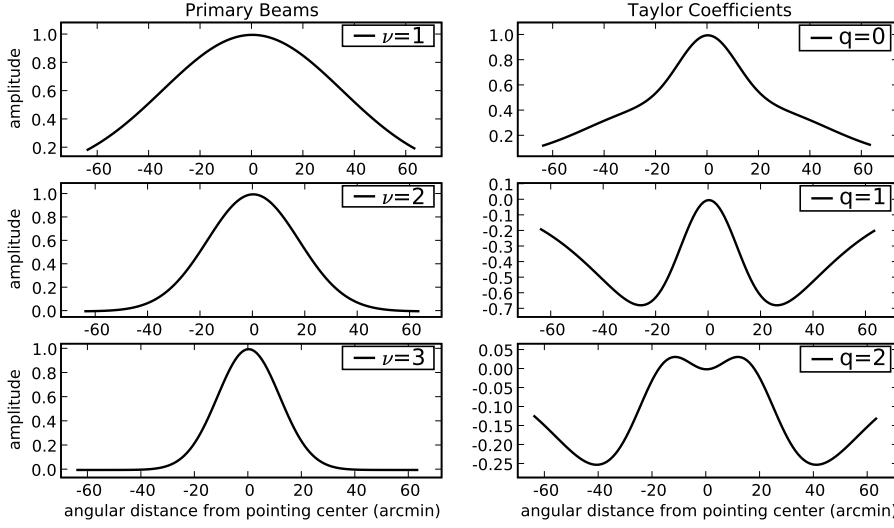


Figure 7.1: Multi-frequency primary beams : The plots on the left show a set of 1-D Gaussians that represent primary beams at three different frequencies (Eqn. 7.16 with $\nu = \{\nu_0/2, \nu_0, 3\nu_0/2\}$ from top to bottom). The plots on the right are the first three coefficients of the Taylor-polynomial required to represent the frequency dependence of the primary beam. These coefficients were evaluated numerically by solving Eqn. 7.15 (compare with the functions shown in Eqns. 7.17 to 7.19).

7.2.1.1 Coefficients of the primary beam polynomial

Figure 7.1 shows a 1-D example of a multi-frequency primary beam and the first three Taylor polynomial coefficients that represent it. For simplicity, consider a Gaussian primary beam whose width scales inversely with frequency (x represents the angular distance from the center of the primary beam).

$$P_b(x, \nu) = e^{-\frac{x^2}{2} \left(\frac{\nu}{\nu_0}\right)^2} \quad (7.16)$$

The first three coefficients of a Taylor polynomial (Eqn. 6.25) fitted to the spectrum of the primary beam at each location x are given as follows.

$$P_{b0} = P_b(x, \nu_0) \quad (7.17)$$

$$P_{b1} = -x^2 P_{b0} \quad (7.18)$$

$$P_{b2} = -x^2(1 - x^2)P_{b0} \quad (7.19)$$

The plots on the right of Fig. 7.1 show the numerical estimates of P_{b0}, P_{b1}, P_{b2} computed using the three primary beams shown in the left column of plots. In this simple example with three widely-spaced frequencies, a 2nd-order Taylor-polynomial is insufficient to model the primary beam spectrum very accurately and in practice, higher order polynomials are used.

7.2.1.2 Evaluating the Multi-Frequency Primary-Beam model

Eqn. 7.15 shows the system of equations to be solved to compute the coefficients of the multi-frequency primary beam model. However in practice, w_v^t and w_v^{t+q} are not the only weights that are used. In a real observation, in addition to the frequency-dependent shape of the primary beam, the sensitivity of the instrument tends to vary across frequency due to different measurement and imaging weights, and across time due to the rotation of azimuthally asymmetric beams. Therefore we need to compute an average primary beam by following the exact measurement process present in a particular dataset and modeling the resulting average beam and its frequency dependence.

The following is a thought experiment to describe how such a calculation of primary-beam coefficients fits into the MFS imaging framework. The purpose of such a description is to illustrate how the effect of the average primary beam can be both measured and removed during MFS imaging. It uses the idea that when a flat sky (constant unit amplitude) is imaged using an interferometer with complete uv -plane sampling (δ -function PSF) and a direction-dependent gain (primary beam), the observed image will be the primary beam². The normal equations for such a system can be re-written to separate the primary beam from the instrument, creating another system of equations that can be easily solved *via* the MFS principal solution (section 6.2.2.3). The result of this process is a set of polynomial coefficients of the primary beam that include information about the time-variability of the primary beam and the measurement and imaging weights.

Sky model : Let a flat sky model (with no frequency dependence) be denoted by \vec{I}_q^{model} , an interferometer with full sampling described with $[S_v] = diag(\vec{I})$ and a frequency-dependent primary beam given by $[P_{bv}] = diag(\vec{P}_{bv})$.

Measurement equations : The measurement equations for this system are given by

$$\vec{V}_v^{obs} = \sum_{q=0}^{N_t-1} w_v^q [F][P_{bv}]\vec{I}_q^{flat\ sky} \quad (7.20)$$

$$\text{where } \vec{I}_{q=0}^{flat\ sky} = \vec{1} \text{ and } \vec{I}_{q>0}^{flat\ sky} = \vec{0} \quad (7.21)$$

Our goal is to solve the normal equations for this system and evaluate the set of polynomial coefficients \vec{P}_{bq} (introduced in Eqn. 7.14).

We will begin by writing expressions for the dirty images formed with this instrument and sky brightness, and then show how they fold into the normal equations.

² Note that this is only an alternate interpretation of the effect of the primary beam, and in practice, an interferometer with full uv -sampling or a flat sky are not required to compute the primary beam coefficients.

Observed image : The t^{th} observed³(dirty) image can be written as follows.

$$\vec{I}_t^{obs\ pb} = \sum_{\nu} w_{\nu}^t [F^\dagger W_{\nu}] \vec{V}_{\nu}^{obs} \quad (7.22)$$

$$= \sum_{q=0}^{N_t-1} \left\{ \sum_{\nu} w_{\nu}^{t+q} [F^\dagger W_{\nu} F] [P_{b\nu}] \right\} \vec{I}_q^{model} \quad \text{using Eqn. 7.20 for } \vec{V}_{\nu}^{obs} \quad (7.23)$$

$$= \sum_{\nu} w_{\nu}^t [F^\dagger W_{\nu} F] \vec{P}_{b\nu} \quad \text{using Eqn. 7.21 for } \vec{I}_q^{model} \quad (7.24)$$

$$= \sum_{q=0}^{N_t-1} \left\{ \sum_{\nu} w_{\nu}^{t+q} [F^\dagger W_{\nu} F] \right\} \vec{P}_{bq} \quad \text{using Eqn. 7.14 for } \vec{P}_{b\nu} \quad (7.25)$$

Eqns. 7.22 to 7.25 show four different ways of expressing the same observed (dirty) image. We can arrange these equations into LHS and RHS pairs to form systems of equations with known quantities on the RHS and unknowns on the LHS. We can then solve the resulting system of equations to calculate the unknowns. Here, the unknowns are the coefficients of a Taylor polynomial that describes the primary beam, and the known quantities are the single-frequency primary beams⁴.

Normal Equations : Let us denote the normal equations constructed from Eqn. 7.20 as

$$[H^{mfs,\delta,pb}] \vec{I}^{flat\ sky} = \vec{I}^{obs\ pb} \quad (7.26)$$

where $\vec{I}^{flat\ sky}$ and $\vec{I}^{obs\ pb}$ are vertical stacks of $\vec{I}_q^{flat\ sky}$ and $\vec{I}_t^{obs\ pb}$ respectively (from Eqns. 7.21 and 7.22). The matrix $[H^{mfs,\delta,pb}]$ can be described by writing each block-row (of Eqn. 7.26) as a system of equations formed with Eqn. 7.23 as the LHS and Eqn. 7.24 as the RHS.

$$\sum_{q=0}^{N_t-1} \left\{ \sum_{\nu} w_{\nu}^{t+q} [F^\dagger W_{\nu} F] [P_{b\nu}] \right\} \vec{I}_q^{flat\ sky} = \sum_{\nu} w_{\nu}^t [F^\dagger W_{\nu} F] \vec{P}_{b\nu} \quad (7.27)$$

Each block of $[H^{mfs,\delta,pb}]$ is given by the expression within curly braces on the LHS of Eqn. 7.27. This system represents the use of an interferometer with complete uv -sampling and frequency-dependent primary beams to observe a flat-spectrum flat sky and produce images given by Taylor-weighted sums of $\vec{P}_{b\nu}$. Figure 7.2 is a pictorial representation of Eqn. 7.26 computed using Eqn. 7.27.

³In this example with complete uv -sampling, the dirty image will be called the observed image.

⁴Note that this process is numerically identical to fitting Taylor polynomials to the primary beam spectrum one pixel at a time. It is described in this manner only to convey the connection between this process and multi-frequency deconvolution, and show that primary beam coefficients can be computed from the same weight images that the MS-MFS algorithm already computes and uses.

Now, our goal is to solve for the series coefficients of the primary beam (\vec{P}_{bq}). To do this, we can re-write the LHS of Eqn. 7.27 using Eqn. 7.25.

$$\sum_{q=0}^{N_t-1} \left\{ \sum_{\nu} w_{\nu}^{t+q} [F^{\dagger} W_{\nu} F] \right\} \vec{P}_{bq} = \sum_{\nu} w_{\nu}^t [F^{\dagger} W_{\nu} F] \vec{P}_{b\nu}^{\text{obs}} \quad (7.28)$$

We can now write a new set of normal equations as follows.

$$[H^{\text{mfs},\delta}] \vec{P}_b^{\text{mfs}} = \vec{I}^{\text{obs pb}} \quad (7.29)$$

Here \vec{P}_b^{mfs} is a vertical stack of primary-beam coefficients \vec{P}_{bq} as defined in Eqn. 7.14, $[H^{\text{mfs},\delta}]$ is the MFS Hessian matrix computed with full uv sampling and without any primary beams⁵. Fig. 7.3 is a pictorial representation of Eqn. 7.29 computed using Eqn. 7.28.

Therefore, given a set of single-frequency primary beams, the RHS of Eqn. 7.29 can be easily evaluated and Taylor coefficients of the primary beam can be computed one pixel at a time *via* the the MFS principal solution (Eqn. 6.38). This analysis can be taken further to show how these primary beam coefficients can be separated from the true sky brightness distribution.

7.2.1.3 Separating the primary beam from the sky

This section shows how the wide-band primary beam coefficients can be used for image-domain primary beam corrections (basically, a division of two polynomials). Let $\vec{I}_q^{\text{sky,pb}} \forall q \in \{0, N_t - 1\}$ be a set of Taylor polynomial coefficients that represent the product of the sky \vec{I}_q^{model} and the primary beam \vec{P}_{bq} . The first three terms of this polynomial product can be written as follows.

$$\vec{I}_0^{\text{sky,pb}} = \vec{P}_{b0} \vec{I}_0^{\text{model}} \quad (7.30)$$

$$\vec{I}_1^{\text{sky,pb}} = \vec{P}_{b1} \vec{I}_0^{\text{model}} + \vec{P}_{b0} \vec{I}_1^{\text{model}} \quad (7.31)$$

$$\vec{I}_2^{\text{sky,pb}} = \vec{P}_{b2} \vec{I}_0^{\text{model}} + \vec{P}_{b1} \vec{I}_1^{\text{model}} + \vec{P}_{b0} \vec{I}_2^{\text{model}} \quad (7.32)$$

This polynomial product can be written as the product of a block lower triangular matrix (which we shall call $[P_b^{\text{mfs,mult}}]$) and a stack of image vectors (\vec{I}^{model}).

If we re-write \vec{P}_b^{mfs} in Eqn. 7.29 as the product of the primary beam and a flat sky, we get the following system of equations (depicted pictorially in Fig. 7.4).

$$[H^{\text{mfs},\delta}] [P_b^{\text{mfs,mult}}] \vec{I}^{\text{flat sky}} = \vec{I}^{\text{obs pb}} \quad (7.33)$$

⁵Compare the expression within curly braces in Eqn. 7.28, with the expression for a MFS Hessian block given in Eqn. 6.35 and set $[S]$ to an identity matrix to emulate full uv sampling. Also, the magnitudes of the diagonal elements of $[H^{\text{mfs},\delta}]$ (shown as δ -functions in Fig. 7.3) are equal to the elements of $[H_{\text{peak}}]$ for MFS imaging (Eqn. 6.39).

Note that the RHS vectors in Figs. 7.2, 7.3 and 7.4 are identical. This suggests that although the true measurement process requires that the primary beam be treated as part of the instrument, we can separate the primary beam from the instrument to compute its Taylor-coefficients and also separate it from the sky brightness distribution to correct the reconstructed image.

Finally, this factorization can be used with a real interferometer (incomplete uv sampling) to give the following normal equations (derived by analogy with Eqn. 7.33).

$$[H^{mfs}][P_b^{mfs,mult}]\vec{I}^{sky,mfs} = \vec{I}^{dirty,mfs,pb} \quad (7.34)$$

Fig. 7.5 is a pictorial representation of these factorized normal equations⁶. This factorization of \vec{P}_{bq} out of both the MFS Hessian and the sky model allows the use of standard MFS deconvolution techniques, followed by a post-deconvolution image-domain correction (*via* polynomial division) to separate the primary beam from the sky brightness. Note that this process is equivalent to solving this system of equations by inverting each matrix on the LHS from left to right and applying these inverses in the same order to the RHS vectors (see Fig. 4.1 for the single-frequency equivalent of Fig. 7.5).

⁶If incomplete sampling is included in Eqns. 7.22 to 7.24, $[F^\dagger W_\nu F]$ becomes $[F^\dagger S_\nu^\dagger W_\nu S_\nu F] = [B_\nu]$ (the Beam matrix for frequency ν). However, the factorization of \vec{P}_{bq} out of the frequency summation still holds.

$$\left[H_{3m \times 3m}^{mfs, \delta, pb} \right] \quad \vec{I}_{3m \times 1}^{flat sky} = \vec{I}_{3m \times 1}^{obs pb}$$

Figure 7.2: Evaluating the multi-frequency primary beam model - 1: This diagram is a pictorial representation of Eqn. 7.26 and describes an observation of a flat spectrum flat sky with a filled-aperture interferometer (δ -function PSFs) and frequency-dependent primary beams. The sky model $\vec{I}^{flat sky}$ is given by Eqn. 7.21, the matrix on the LHS is the MFS Hessian matrix with full uv -sampling and primary beams. The symmetric taper across the diagonal of each block in the LHS matrix is a Taylor-weighted primary beam. The RHS vectors are Taylor-weighted averages of the single-frequency beams. The frequency-dependent primary beams used in this example are the same as those shown in Fig. 7.1.

$$\left[H_{3m \times 3m}^{mfs, \delta} \right] \quad \vec{P}_b^{mfs}_{3m \times 1} = \vec{I}_{3m \times 1}^{obs pb}$$

Figure 7.3: Evaluating the multi-frequency primary beam model - 2 : This diagram is a pictorial representation of Eqn. 7.29 and is another way of obtaining the same RHS as Eqn. 7.26 (compare with Fig. 7.2). The coefficients of the primary-beam polynomial form the sky brightness distribution (compare \vec{P}_b^{mfs} with the right column of plots in Fig. 7.1), and the LHS matrix is the MFS Hessian matrix with full uv -sampling (similar to the Hessian in Fig. 6.6 but with δ -functions) but no primary beams (all diagonal elements per block are equal). This is the system of equations to be solved to compute polynomial coefficients for the primary beam spectrum from weighted averages of single-frequency primary beams, by using $[H^{peak}]$ constructed for the multi-frequency principal solution (section 6.2.2.3).

$$\left[H_{3m \times 3m}^{mfs, \delta} \right] \left[P_b^{mfs, mult} \right] \left[I_{3m \times 1}^{\text{flat sky}} \right] = \left[I_{3m \times 1}^{\text{obs pb}} \right]$$

Figure 7.4: Evaluating the multi-frequency primary beam model - 3 : This diagram is another representation of the imaging equations shown in Figs. 7.2 and 7.3. It shows how the effect of the primary beam can be separated from the MFS Hessian as well as from the sky model, but still give the same RHS. It follows Eqn. 7.33 and shows the matrix-vector product of the frequency-dependent primary beam and the flat sky. The block lower-triangular form of the primary beam matrix implements a polynomial multiplication between the primary beam and sky brightness, in terms of their polynomial coefficients (Eqns. 7.30 to 7.32).

$$\left[H_{3m \times 3m}^{mfs} \right] \left[P_b^{mfs, mult} \right] \left[I_{3m \times 1}^{\text{sky, mfs}} \right] = \left[I_{3m \times 1}^{\text{dirty, mfs, pb}} \right]$$

Figure 7.5: Multi-frequency normal equations with the primary beam factored out : This diagram represents an observation of a field of two point sources with non-flat spectral structure, using an interferometer with incomplete *uv*-sampling and frequency-dependent primary beams. It follows Eqn. 7.34 in which the Hessian matrix (on the left) and the sky model $I_{3m \times 1}^{\text{sky, mfs}}$ are the same as shown in Fig. 6.6 (for multi-frequency imaging with no primary beams). The multiplicative effect of the average primary beam is shown via the matrix $[P_b^{mfs, mult}]$ (similar to Fig. 7.4). This system can be solved from left to right. MFS deconvolution applied directly to the RHS vectors produces a set of coefficients that represent the product of the primary beam and the sky brightness. The sky brightness and its spectrum can be recovered in a second step by computing and applying the inverse of the lower-triangular polynomial-multiplier matrix $[P_b^{mfs, mult}]$. (see Fig. 4.1 for the single-frequency version of this system).

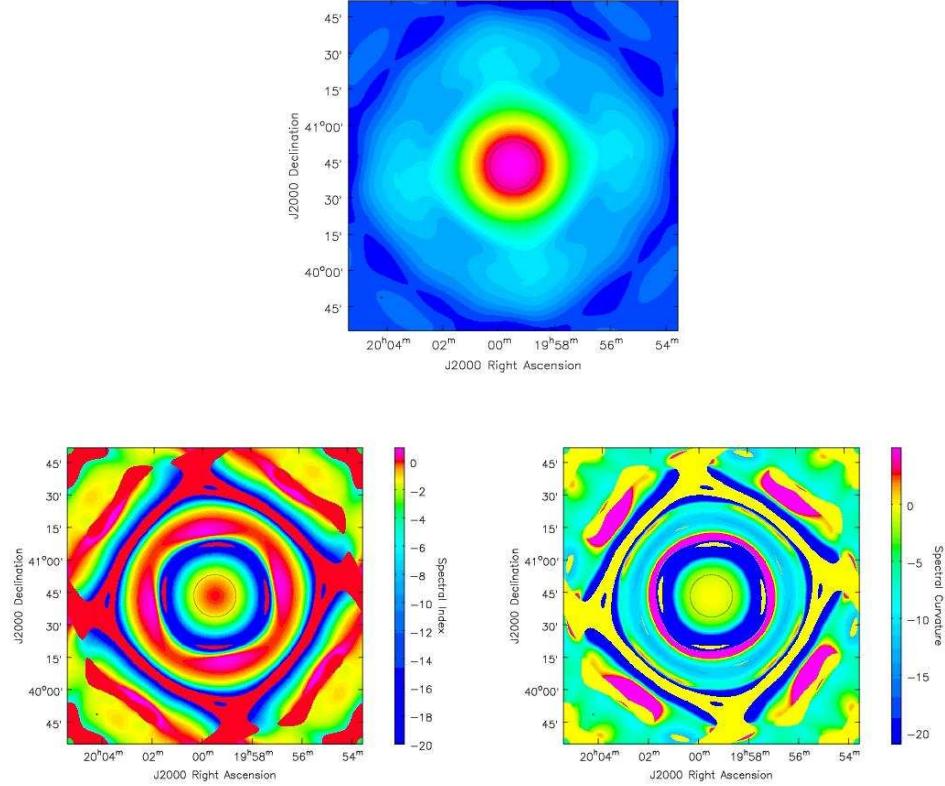


Figure 7.6: Average Primary Beam, Spectral Index and Curvature : These images show the frequency dependence of the primary beam and its variation with direction on the sky. They are the result of computing primary beam coefficients (section 7.2.1.2) from the set of single-frequency primary beams shown in Fig.5.2 and calculating the effective spectral index and curvature from these coefficients (*via* Eqn. 6.43 to 6.45). The average primary beam (also first Taylor coefficient) (top), its spectral index (left) and curvature (right) are shown over a field of view extending past the second sidelobe at the reference frequency. Note that the reference primary beam $\vec{P}_{b\nu_0}$ has a smooth extended sidelobe at the few % level, but $\vec{P}_{b\alpha}$ and $\vec{P}_{b\beta}$ are smooth functions only within the main lobe where the frequency variation is monotonic. The spectral index of the main lobe of the primary beam (*i.e.* no sidelobes) is shown in more detail in Fig.5.4.

7.2.1.4 Spectral index and curvature of the Primary Beam

This section briefly discusses a power-law interpretation of the frequency dependence of the primary beam in order to illustrate its effect on the MFS imaging process. The spectral index and curvature due to the primary beam can be computed from the first three primary-beam polynomial coefficients (as shown in Eqn. 6.43 to 6.45).

Figure 7.6 shows images of the reference-frequency primary beam and the spectral index and curvature associated with the average beams for an EVLA antenna. These beams were computed from the single-frequency primary beams shown in Fig.5.2.

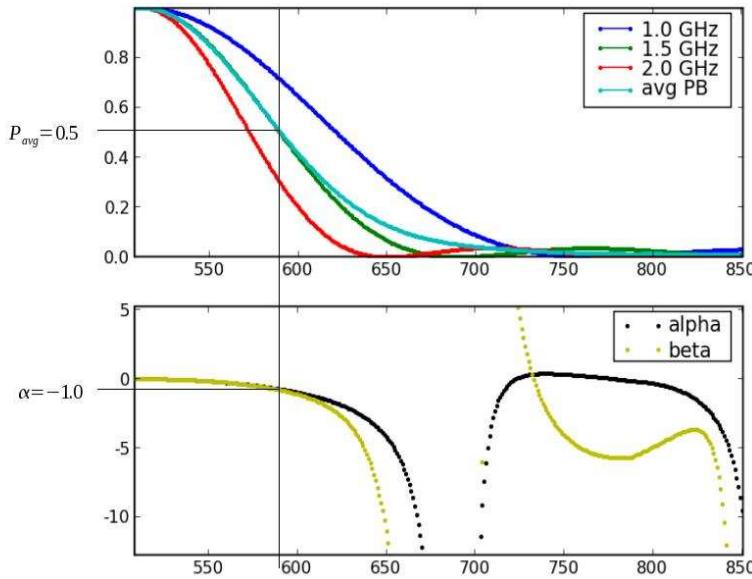


Figure 7.7: Spectral Index and Curvature of the EVLA Primary Beam : The top plot shows 1D cuts through the EVLA primary beam at three L-band frequencies 1.0, 1.5 and 2.0 GHz. The cut was chosen to pass through a peak in the first sidelobe, and shows only one half of the beam (the X-axes is in units of image pixels with the pointing center at 512). The bottom plot shows the corresponding average beam, spectral index and curvature (from Fig.7.6), also as a function of angular distance from the pointing center in units of image pixels. Note that at the locations of the nulls of the reference frequency (1.5 GHz, green line) α and β diverge, but have stable values within the main lobe as well as in a significant part of the first sidelobe.

The \vec{P}_b^α and \vec{P}_b^β images, show that outside the main lobe of the reference beam, the spectral index and curvature taken on high values and vary rapidly. This indicates that a power law model with varying index is valid only in regions where the primary beam spectrum is monotonic (regions in or near the null are blanked out and appear with value zero). Figure 7.7 further illustrates this point. The top plot shows a one-dimensional cut through the multi-frequency primary beams shown in Fig.5.2 (the cut was chosen to pass through a sidelobe peak in the individual beams).The bottom plot shows the values of spectral index \vec{P}_b^α and curvature \vec{P}_b^β as a function of angular distance form the pointing center. Note that at the null of the reference-frequency primary beam, \vec{P}_b^α and \vec{P}_b^β are unstable, but again take on stable values in the first side-lobe. This behaviour decides the regions of the image for which the frequency-dependent primary beam can or cannot be corrected for using a spectral index/curvature representation (in the general case, the polynomial coefficients can be used directly for the image-domain corrections.).

7.2.2 Imaging Equations and Block Deconvolution

This section describes an algorithm for multi-scale multi-frequency synthesis imaging with the correction of direction-dependent effects that vary with antenna, baseline, time and frequency. It uses the multi-frequency primary beam model derived in section 7.2.1 and folds it into the iterative deconvolution process described in section 7.1 for multi-scale multi-frequency imaging. The correction of direction-dependent effects is done *via* the methods described in chapter 4, in particular the primary-beam correction algorithm described in section 4.3.2 to remove the effect of the antenna primary beam *via* a combination of visibility-domain and image domain operations⁷.

For visibility-domain operations, the frequency-dependence of the primary beam can be naturally accounted for by scaling the size of the aperture illumination function and using it to construct a gridding convolution function for each observing frequency⁸. For image-domain operations, the frequency-dependence of the primary beam has to be absorbed into the image model by describing the observed sky as the product of the true source spectrum and a frequency dependent primary beam (Eqn. 7.35). By independently measuring and modeling the spectral behaviour of the antenna primary beam, we can separate the two and recover the true sky brightness and spectrum.

7.2.2.1 Wide-Band Image Model with the Primary Beam

For image-domain operations, we will define a flux model for the observed brightness distribution as the product between the true sky brightness and the primary beam, both represented as power-laws with varying spectral indices (see sections 6.2.1 and 7.2.1.4).

$$I_v^{obs} = I_{\nu_0} P_{b\nu_0} \left(\frac{\nu}{\nu_0} \right)^{[\alpha + \alpha_{PB}] + [\beta + \beta_{PB}] \log\left(\frac{\nu}{\nu_0}\right)} \quad (7.35)$$

The image model used in the measurement equations is a multi-scale multi-frequency representation (defined in Eqn. 7.1) of this observed sky brightness \vec{I}_v^{obs} . The model image $\vec{I}_{t,s}^m$ for each spatial scale s and term t in the series now represents a collection of δ -functions that describe the location and total flux of flux components that describe the product of the sky brightness and the primary beam (and not just the sky brightness, as was the case in section 7.1.1, Eqn. 7.1).

⁷Note also that the flat-sky *vs* flat-noise discussion in step 2 of the algorithm described in section 4.3.2 for single-frequency primary-beam correction applies to multi-frequency primary-beam correction as well.

⁸This is sufficient for direction-dependent effects such as the w -term and beam squint which are also frequency-dependent. For a given baseline, the value of w changes with frequency and the appropriate w -projection kernel must be chosen to construct the gridding convolution function for each frequency (see section 4.2.2.4). The beam squint (a polarization-dependent pointing offset for reflecting dishes with off-axis feeds) is a fixed fraction of the primary-beam width and therefore varies with frequency and requires a frequency-dependent phase to be applied to the aperture illumination function used to construct gridding convolution functions.

7.2.2.2 Measurement and Normal Equations

This section describes the MFS normal equations formed when the image model shown in Eqn. 7.35 is used with methods for correcting direction-dependent effects. Let us combine the normal equations for MFS imaging (shown in Eqns. 6.35 and 6.36) with those for single-frequency primary-beam correction (shown in Eqn. 4.12 for the measurement equation shown in Eqn. 4.9). Let the resulting normal equations be denoted as follows.

$$[H^{mfs,pb^2}] \vec{I}^{sky,mfs} = \vec{I}^{dirty,mfs,pb^2} \quad (7.36)$$

Each block row of this system is written as follows.

$$\sum_{q=0}^{N_t-1} \left\{ \sum_{\nu} w_{\nu}^{t+q} [P_{b\nu}^{\dagger}] [B_{\nu}] [P_{b\nu}] \right\} \vec{I}_q^{sky} = \sum_{\nu} w_{\nu}^t [P_{b\nu}] \vec{I}_{\nu}^{dirty} \quad (7.37)$$

Here, \vec{I}_{ν}^{dirty} is the dirty image computed for frequency ν via Eqn. 3.10 and $[B_{\nu}]$ is the Hessian (Beam) matrix constructed from Eqn. 3.12 for each ν . Note that there are no multi-scale terms in this equation. This is because we are analysing image-domain effect of the primary beam on the full multi-scale image (the result of a linear combination of basis functions (Eqn. 6.1)) and not the multi-scale model image which consists of a set of δ -functions.

By analogy with Eqns. 7.22 to 7.24 and Eqn. 7.34 for MFS normal equations with primary beams, we factor the primary beams out of the summation over ν and re-cast Eqn. 7.36 in terms of a pre-multiplication and a post-multiplication of the MFS Hessian with the primary beam.

$$[P_b^{mfs,mult^T}] [H^{mfs}] [P_b^{mfs,mult}] \vec{I}^{sky,mfs} = \vec{I}^{dirty,mfs,pb^2} \quad (7.38)$$

Note that the factorization of the wide-band primary beam on the left of the MFS Hessian is an approximation, and is shown here only in analogy with Eqn. 4.12 for the single-frequency case.

Fig. 7.9 shows this factorization (Eqn. 7.38) for an example in which $N_t = 3$, $N_s = 1$ and the sky is composed of two point sources with different spectral characteristics (the same sky brightness distribution, and MFS Hessian as shown in Fig. 6.6).

This factorization shows that when gridding convolution functions are constructed from aperture illumination functions and used during gridding, the system of normal equations contains two instances of the primary beam and its frequency dependence⁹. This set of equations is solved from left to right as a pre-deconvolution normalization by the primary beam, MFS deconvolution, and a post-deconvolution correction of the primary beam to separate it from the sky brightness distribution.

⁹The presence of two instances of the primary beam in the normal equations when aperture-illumination-based gridding convolution functions are used is similar to single-frequency case described in section 4.3. Also, when standard gridding is used (similar to section 4.2.1 for the single frequency case), there will be only one instance of the primary beam and its frequency dependence.

$$\left[H_{3m \times 3m}^{mf, pb^2} \right] \quad \vec{I}_{3m \times 1}^{sky, mfs} \quad \vec{I}_{3m \times 1}^{dirty, mfs, pb^2}$$

Figure 7.8: Normal Equations for MFS with Primary-Beam Correction : This diagram represents the normal equations shown in Eqn. 7.36 in which the antenna primary beams are included in the measurement equation, and the dirty images on the RHS are formed by applying visibility-domain corrections for direction-dependent effects *via* gridding convolution functions. The multi-frequency PSFs used for this example are shown in Fig. 6.5, the primary beams are shown in Fig. 7.1 and the sky model is the same as in Fig. 6.6. Note that the diagonals of all Hessian blocks are scaled by $\vec{I}^{wt} \propto \vec{P}_b^2$ (compare with Fig. 4.2 for the single-frequency case).

$$\left[P_b^{mf, mult} {}^T \right] \quad \left[H_{3m \times 3m}^{mf} \right] \quad \left[P_b^{mf, mult} \right] \quad \vec{I}_{3m \times 1}^{sky, mfs} \quad \vec{I}_{3m \times 1}^{dirty, mfs, pb^2}$$

Figure 7.9: Normal Equations for MFS and Primary-Beam Correction : This diagram represents the same system as shown in Fig. 7.8 but with the wide-band primary beam factored out of the MFS Hessian (Eqn. 7.38). This matrix product is similar to that shown in Fig. 4.3 for the single-frequency case, but here, the factorization of the left-most primary beam matrix out of the Hessian is an approximation (and the RHS vectors are not identical to those in Fig. 7.8). Note that the matrix in the middle is the MFS Hessian that contains no direction-dependent effects. The solution of this system of equations proceeds from left to right. The primary beam matrix on the left of the MFS Hessian is eliminated first by dividing the RHS by the multi-frequency polynomial. The MFS Hessian is then eliminated *via* a multi-frequency deconvolution. The second polynomial multiplier matrix is eliminated as a post-deconvolution step. All these steps are done with the minor cycle of an iterative image reconstruction process (when MS-MFS is combined with primary-beam correction).

7.2.2.3 Iterative Block Deconvolution with Primary-Beam correction

Each step of an iterative block deconvolution will now be described in detail. This is a multi-scale multi-frequency deconvolution algorithm that follows the methods described in section 4.3.2 for the correction of direction-dependent effects (algorithm 7 on page 153 lists the relevant steps).

Pre-compute Hessian : All $N_s N_t \times N_s N_t$ terms in the MS-MFS Hessian ($t, q \in [0, N_t]; s, p \in [0, N_s]$) are evaluated by computing all unique multi-scale multi-frequency convolutional kernels $\vec{I}_{s,p}^{psf}$ (Eqn. 7.9). This step is the same as that described in section 6.2.2.5 for MS-MFS deconvolution.

$$\vec{I}_{s,p}^{psf} = \vec{I}_s^{shp} \star \left\{ \sum_v w_v^{t+q} \vec{I}_v^{psf} \right\} \star \vec{I}_q^{shp} \quad (7.39)$$

where \vec{I}_v^{psf} is the PSF computed via Eqn. 3.11. All the convolution kernels are normalized by w_{sum} such that the peak of the zeroth order function is unity, and the relative weights between all $\vec{I}_{s,p}^{psf}$ are preserved. A set of N_s [H^{peak}] matrices is constructed for each spatial scale s . Each is a matrix of size $N_t \times N_t$, representing a block diagonal approximation of the MS-MFS Hessian (shown in Eqn. 7.5).

Pre-compute Weight Images : A set of weight images \vec{I}_t^{wt} are constructed from weighted sums of the aperture illumination functions used to form the dirty image vectors.

$$[\vec{I}_t^{wt}] = \sum_v w_v^t [F^\dagger] [S_v^{dd^\dagger} W_v^{im} S_v^{dd}] \approx \sum_v w_v^t (tr[W_v^{im}]) [\vec{P}_{bv}^2] \quad (7.40)$$

where $[\vec{I}_t^{wt}] = diag(\vec{I}_t^{wt})$ and $[\vec{P}_{bv}^2] = diag(\vec{P}_{bv}^2)$. $[S_v^{dd}] \approx [S_v][G_v] = [S_v][FP_{bv}F^\dagger]$ is a convolution operator on the spatial frequency plane when all primary beams are assumed to be identical. In practice this implies the use of an approximate average primary beam for image-domain corrections, but it is understood that all visibility-domain corrections are still done using baseline and time-dependent functions. If $[G_v]$ is unitary, then $[G_v^\dagger G_v]$ is the Identity matrix, and the weight image will contain only the sum of the measurement and imaging weights. For the primary beam, $[G_v]$ is non-unitary and the process of gridding does not correct it completely and the weight image will show evidence of this (compare with Eqn. 4.14 for the single-frequency case).

Pre-compute Primary Beam : Polynomial coefficients for the average primary beam are obtained by applying the inverse of $[H_{s=0}^{peak}]$ to the above set of weight images \vec{I}_t^{wt} . This is similar to the process described in section 7.2.1 to calculate the primary beam polynomial. Compare the expression for the weight image in Eqn. 7.40 with the RHS of the system

of equations used to compute the primary beam coefficients (Eqn. 7.24). The form of the RHS vector is the same, but it is now a weighted sum of \vec{P}_{bv}^2 (and not a weighted sum of \vec{P}_{bv}). The principal solution computed for each pixel now represents the coefficients of a polynomial formed from the square of the primary beam. The series coefficients of the primary beam (\vec{P}_{bq}) are extracted from these coefficients by calculating the square root of the resulting polynomial in terms of its coefficients (again, compare with Eqn. 4.14 for the single-frequency case).

A polynomial square root is equivalent to a division by two in log-space. For $N_t = 3$, the calculation of the polynomial square root is numerically identical to converting these coefficients into power-law parameters¹⁰ (using Eqns. 6.43 to 6.45), computing a reference-frequency image, a spectral index image and a spectral curvature image, and then taking the square-root of the reference-frequency beam, dividing the spectral index and curvature images by two, and recomputing Taylor-coefficients \vec{P}_{bq} for a polynomial representation of this new power law (using Eqn. 6.25). Let us denote the parameters of this new power law as \vec{P}_{bv_0} for the primary beam at the reference frequency, $\vec{P}_{b\alpha}$ as the spectral index due to the primary beam and $\vec{P}_{b\beta}$ for spectral curvature (see Figs. 7.6 and 5.4 for 2D images). These primary beam parameters represent the frequency dependence of the weighted average of the individual primary beams.

Initialization : Iterations begin by initializing the set of model images for each Taylor-term $\vec{I}_q^{model} \forall q \in \{0, N_t - 1\}$ to zero or to an *a priori* model.

Major and minor cycles : The normal equations for MS-MFS imaging (shown in Eqn. 7.5) are solved iteratively by repeating steps 1 to 7 until some termination criterion is reached. Steps 1,2 and 7 form the major cycle, and steps 3, 4 and 6 form the minor cycle. Direction dependent effects are corrected during the major cycle and the effect of these corrections on the minor cycle depends on the type of normalization chosen for the residual images.

1. **Compute RHS :** Residual images for all pairs of spatial scales ($s \in [0, N_s]$ and Taylor terms ($t \in [0, N_t]$)) are computed as follows.

$$\vec{I}_s^{res} = \vec{I}_s^{shp} \star \vec{I}_t^{res} \quad (7.41)$$

$$\text{where } \vec{I}_t^{res} = \sum_v w_v^t [F^\dagger R S_v^{dd} W_v^{im}] \vec{V}_v^{res} \approx \sum_v w_v^t [P_{bv}] \vec{I}_v^{res} \quad (7.42)$$

¹⁰ Note however, that $N_t = 3$ may not be sufficient to accurately model the primary beam spectrum in regions outside the main lobe. In this case, an explicit N_t order polynomial square root needs to be computed.

Since we intend to use the MS-MFS deconvolution algorithm for the minor cycle, all direction dependent effects need to be removed by the time the RHS vectors are computed *via* Eqn. 7.41. If not, the minor cycle has to be interpreted differently.

These dirty images need to be normalized before beginning the minor cycle. Based on the form of the normal equations when visibility-domain corrections for direction-dependent effects are done, there are several types of normalizations. They are described below as parts of steps 2 and 6.

2. **Normalization :** There are two ways in which the RHS dirty images can be pre-processed before beginning the minor cycle iterations (see section 2 for the single-frequency version of this discussion).

- (a) **Flat noise :** The first approach is to divide the RHS vectors by one instance of the multi-frequency primary beam. This is equivalent to eliminating the upper-triangular block matrix on the left of $[H_{mfs}]$ in Fig.7.9 and can also be done by computing the dirty image separately for each frequency and dividing it by \vec{P}_{bv} .

We are left with one instance of the primary beam, and the system of equations being solved are equivalent to those shown in Fig.7.5 (the case where standard gridding is used to compute the dirty images and no visibility-domain corrections of direction-dependent effects are applied). The noise in the image is the same across the entire field of view, but the flux is modulated by the instrument primary beams. The primary beam on the right of $[H^{mfs}]$ is treated as part of the sky model and taken out of the final result of each minor cycle.

(An alternate approach is to divide all the RHS images by \vec{P}_{b0} (i.e. not a polynomial division). This corrects for one instance of an average primary beam and creates flat-noise RHS images, but leaves in the frequency-dependence of the beam. The minor cycle model image will contain one factor of \vec{P}_{bv_0} and two factors of $\vec{P}_{b\alpha}$ and $\vec{P}_{b\beta}$.)

- (b) **Flat sky :** The second approach is to divide the RHS image by the weight images \vec{I}_t^{wt} or to apply a polynomial division with the coefficients of the square of the primary beam (see section 7.2.2.3). This is equivalent to dividing the dirty image at each frequency by the square of the primary beam at that frequency \vec{P}_{bv}^2 .

These operations produce RHS images that represent the peak sky brightness not modulated by the primary beams, but the noise is not uniform across the images and this has to be accounted for while searching for flux components. Further, the normal equations in Eqn. 7.37 and Fig.7.9 show the two primary beams on either side of the MFS Hessian, and any operation that uses a single-step image-domain correction by the square of the primary beam will be an approximation.

(An alternate approach is to divide the RHS images by $\vec{P}_{b_{\nu_0}}^2$. This will give a flat-sky intensity image, but the minor cycle model image will still pick up and two factors of \vec{P}_{ba} and $\vec{P}_{b\beta}$.)

It is important to note that both types of normalization must be done to \vec{I}_t^{dirty} in Eqn. 7.41 before smoothing it to different spatial scales via Eqn. 7.41. This is because the image model is a multi-scale representation of the sky multiplied by the primary beam ($\sum_s \vec{I}_s^{shp} \star \vec{I}_s^m = \vec{P}_b \vec{I}^{sky}$) and multiplication and convolution do not commute.

3. **Find a Flux Component :** A flux component is chosen in the same as described in step 2 of the MS-MFS deconvolution algorithm and principal solution (described in sections 7.1.2.5 and 7.1.2.3) The principal solution is computed for all pixels, one scale at a time via Eqn. 7.10. The N_t element solution set of Taylor coefficients with the dominant $q = 0$ component across all scales and pixel locations, is chosen the current flux component. Let the scale size for this set be p_i . The chosen solution set for iteration i is given by $\{\vec{I}_{q,(i)}^{model}\}; q_i \in [0, N_t]$.
4. **Update model images :** Multi-scale model images are accumulated for each Taylor coefficient (same as Eqn. 7.12).

$$\vec{I}_p^{model} = \vec{I}_p^{model} + g \left(\vec{I}_{q,(i)}^{model} \star I_{p_i}^{shp} \right) \quad \forall q \in \{0, N_t - 1\} \quad (7.43)$$

where g is a loop-gain that takes on values between 0 and 1 and controls the step size for each iteration in the χ^2 -minimization process.

5. **Update RHS :** The RHS residual images for each Taylor term are updated as follows (same as Eqn. 7.13).

$$I_t^{res} = I_t^{res} - g \left(\sum_{q_i=0}^{N_t-1} \left[I_{i,q}^{psf} \star I_{q,(i)}^{model} \right] \right) \quad (7.44)$$

Repeat from Step 3 until a pre-computed flux-limit is reached.

6. **Correct for PB :** Depending on the choice of normalization (step 2), the multi-frequency model at the end of the minor cycle needs to be corrected for the primary beam and its frequency dependence.
 - (a) **Flat noise :** This step is equivalent to a polynomial division that eliminates the primary-beam matrix on the right of the MFS Hessian in Fig. 7.9. In the case of flat-noise normalization, the model image contains one instance of the primary beam and its frequency dependence. The model images for each Taylor coefficient are corrected using their power-law interpretation¹¹. First the reference-

¹¹Note that the numerical steps involved in primary-beam correction via a power-law model are exactly equivalent to a polynomial division (when the sky and the primary beam are both modeled by polynomials in frequency). For $N_t = 3$, the choice of a power-law instead of a polynomial to represent the frequency-dependence of the primary beam is irrelevant from the point of view of an image-domain correction.

frequency image (\vec{I}_{model,v_0}) spectral index and curvature images ($\vec{I}_{model,\alpha}, \vec{I}_{model,\beta}$) are calculated from the model coefficients *via* Eqns. 6.43 to 6.45 and then the primary beam is removed as $\vec{I}_{model,v_0} / \vec{P}_{b_{v_0}}, \vec{I}_{model,\alpha} - \vec{P}_{b_\alpha}, \vec{I}_{model,\beta} - \vec{P}_{b_\beta}$. Polynomial coefficients for this new power law are then recomputed *via* Eqn. 6.22 and filled into $\vec{I}_t^{model} \forall t \in \{0, N_t - 1\}$ (to be used during prediction).

- (b) **Flat Sky :** No corrections are required because the flux model is already devoid of primary-beam effects (in regions away from the nulls).

In both the above cases, the alternate forms of normalization described in step 2 require different multiples of $\vec{P}_{b_{v_0}}, \vec{P}_{b_\alpha}$ and \vec{P}_{b_β} to be removed from the model image.

The advantage of using the power-law model to separate the primary beam from the sky brightness is that fidelity constraints can be applied on the resulting α and β images before converting them back to Taylor-coefficients. The disadvantage of using this power-law model is that it will be accurate only for parts of the primary beam that are well represented by a power law and $N_t = 3$ suffices to model it (within the main lobe). Out in the sidelobes, $N_t > 3$ terms are usually required to describe the primary-beam polynomial and is more accurate to do the above correction *via* an explicit polynomial division in terms of its coefficients.

7. **Predict :** Model visibilities are computed from each Taylor-coefficient image in the same way as in Eqn. 6.42 for multi-frequency imaging.

$$\vec{V}_v^{model} = \sum_{t=0}^{N_t-1} [W_t^{mfs}] [S^{dd^\dagger} G^{pc} R^\dagger F] [\vec{I}_t^{ps}]^{-1} \vec{I}_t^{model} \quad (7.45)$$

The use of $[S^{dd^\dagger}]$ during de-gridding re-introduces all the direction dependent effects so that the model visibilities can be compared with the data for χ^2 computation (compare with Eqn. 4.16 for the single-frequency case). Since these direction-dependent effects are re-introduced in the visibility domain, it is done separately for each baseline, timestep and frequency, and takes into account any variability. Therefore, even if the minor cycle uses approximate average primary beams, the prediction step and the major cycle are always computed accurately and this is necessary for the iterations to eventually converge.

Residual visibilities are computed as $\vec{V}_v^{res} = \vec{V}_v^{corr} - \vec{V}_v^{model}$.

Repeat from Step 1 until a convergence criterion is reached.

Restoration : The final Taylor coefficient images are restored and interpreted in the same way as described in standard multi-frequency restoration (section 6.2.2.5).

Algorithm 7: MF-MFS CLEAN with MF-PB correction : Major/minor cycles

Data: calibrated visibilities : $\vec{V}_v^{corr} \forall v$
Data: primary beams : $\vec{P}_{bv} \forall v$
Data: uv-sampling function : $[S_v]$
Data: image noise threshold and loop gain σ_{thr}, g_s
Data: scale basis functions : $\vec{I}_s^{shp} \forall s \in \{0, N_s - 1\}$
Result: model coefficient images : $\vec{I}_q^{model} \forall q \in \{0, N_t - 1\}$
Result: spectral index and curvature : $\vec{I}_\alpha^{model}, \vec{I}_\beta^{model}$

- 1 Use Algorithm 8 on the following page to pre-compute
 $\vec{I}_{sp}^{psf}, [H_s^{peak}], \vec{P}_{bv_0}, \vec{P}_{b\alpha}, \vec{P}_{b\beta}$
- 2 Initialize the model \vec{I}_t^{model} for all $t \in \{0, N_t - 1\}$ and compute $f_{sidelobe}$
- 3 **repeat** /* Major Cycle */
- 4 **for** $t \in \{0, N_t - 1\}$ **do**
- 5 Compute the residual image \vec{I}_t^{res}
- 6 Normalize \vec{I}_t^{res} by \vec{P}_{bv_0}
- 7 **for** $s \in \{0, N_s - 1\}$ **do**
- 8 Compute $\vec{I}_{s,t}^{res} = \vec{I}_s^{shp} \star \vec{I}_t^{res}$
- 9 **end**
- 10 **end**
- 11 Calculate f_{limit} from $\vec{I}_{0,0}^{res}$
- 12 **repeat** /* Minor Cycle */
- 13 Compute $I_q^{model} \forall q \in \{0, N_t - 1\}$ and update $\vec{I}_{s,t}^{res} \forall s, t$ (Algorithm 6 on page 134)
- 14 **until** Peak residual in $\vec{I}_{0,0}^{res} < f_{limit}$
- 15 Calculate power-law parameters : $\vec{I}_{v_0}^m, \vec{I}_\alpha^m, \vec{I}_\beta^m$ from $I_q^{model} \forall q$
- 16 Remove primary beam :
- 17 $\vec{I}_{v_0}^{new} = \vec{I}_{v_0}^m / \vec{P}_{bv_0}, \vec{I}_\alpha^{new} = \vec{I}_\alpha^m - 2\vec{P}_{b\alpha}, \vec{I}_\beta^{new} = \vec{I}_\beta^m - 2\vec{P}_{b\beta}$
- 18 Re-compute Taylor coefficients $\vec{I}_q^{new} \forall q$ from $\vec{I}_{v_0}^{new}, \vec{I}_\alpha^{new}, \vec{I}_\beta^{new}$
- 19 Compute model visibilities V_v^{model} from $I_q^{new} \forall q \in \{0, N_t - 1\}$
- 20 Compute a new residual image I^{res} from residual visibilities
 $V_v^{corr} - V_v^{model}$
- 21 **until** Peak residual in $\vec{I}_0^{res} < \sigma_{thr}$
- 22 Calculate spectral index and curvature images, and restore the results

Algorithm 8: MS-MFS with MF-PB correction : Pre-Deconvolution Setup

Data: primary beams : $\vec{P}_{b\nu} \forall \nu$
Data: uv-sampling function : $[S_\nu]$
Data: scale basis functions : $\vec{I}_s^{shp} \forall s \in \{0, N_s - 1\}$
Result: scale-spectral PSFs : $\vec{I}_{tq}^{psf}, [H_s^{peak}]$
Result: primary beam model : $\vec{P}_{b\nu_0}, \vec{P}_{b\alpha}, \vec{P}_{b\beta}$

```

1 for  $t \in \{0, N_t - 1\}, q \in \{t, N_t - 1\}$  do
2   | Compute the spectral PSF  $\vec{I}_{tq}^{psf}$ 
3   | for  $s \in \{0, N_s - 1\}, p \in \{s, N_s - 1\}$  do
4     |   | Compute the scale-spectral PSF  $\vec{I}_{s,p}^{psf} = \vec{I}_s^{shp} \star \vec{I}_p^{shp} \star \vec{I}_{tq}^{psf}$ 
5   | end
6 end
7 for  $s \in \{0, N_s - 1\}$  do
8   | Construct  $[H_s^{peak}]$  from  $mid(\vec{I}_{s,s}^{psf})$  and compute  $[H_s^{peak^{-1}}]$ 
9 end
10 for  $t \in \{0, N_t - 1\}$  do
11   | Compute the weight image  $\vec{I}_t^{wt} = \sum_\nu w_\nu^t(tr[W_\nu^{im}])[\vec{P}_{b\nu}]^2$ 
12 end
13 foreach pixel do
14   | Construct  $I^{rhs}$ , from  $\vec{I}_t^{wt}$  at this location
15   | Compute the primary beam Taylor coefficients  $\vec{P}_b^{sol} = [H_0^{peak^{-1}}]\vec{I}^{rhs}$ 
16 end
17 Compute power-law parameters  $\vec{P}_{b\nu_0}^{sol}, \vec{P}_{b\alpha}^{sol}, \vec{P}_{b\beta}^{sol}$  from  $\vec{P}_{b,t}^{sol} \forall t \in \{0, N_t - 1\}$ 
18 Compute primary-beam parameters  $\vec{P}_{b\nu_0} = \sqrt{\vec{P}_{b\nu_0}^{sol}}, \vec{P}_{b\alpha} = \frac{\vec{P}_{b\alpha}^{sol}}{2}, \vec{P}_{b\beta} = \frac{\vec{P}_{b\beta}^{sol}}{2}$ 

```

CHAPTER 8

WIDE-BAND IMAGING RESULTS

This chapter presents a set of wide-band imaging results to illustrate the capabilities of the multi-scale, multi-frequency deconvolution algorithms described in chapter 7. The examples presented here focus on the EVLA at L-band (1 to 2 GHz) but the results are generic enough to be transferred to other arrays and frequencies. The description of each example emphasizes the accuracy with which spatial and spectral structure can be recovered for a particular type of source and signal-to-noise ratio, and discusses how the choice of image model and algorithm affected the imaging process. Error estimates, dynamic-ranges and performance metrics are presented and discussed wherever relevant in order to convey an idea of what to expect when one uses these methods for spatio-spectral image reconstruction.

Section 8.1 describes imaging results based on simulated EVLA data to demonstrate the capabilities of the MS-MFS algorithm for narrow and wide-field wide-band imaging. Section 8.2 demonstrates the applicability of this algorithm to situations with incomplete spectral sampling where *a priori* information in the form of an image model is used to bias the solution towards a physically appropriate description of the sky brightness. Section 8.3 shows the imaging results from a set of wide-band VLA observations of Cygnus A, M87 and the 3C286 field. Section 8.4 summarizes several practical aspects of wide-band imaging and lists the main factors to keep in mind while using the MS-MFS algorithm for spatio-spectral imaging.

The MS-MFS algorithms described in the chapter 7 were implemented using the CASA libraries (version 2.4), validated using data simulated for the EVLA and applied to wide-band VLA observations taken as a series of snapshots at multiple frequencies. The multi-scale, wide-band flux model used for all the imaging runs in this chapter is given by Eqn. 7.1. Spatial structure is modeled with a collection of multi-scale flux components, and the position-dependent spectrum of the sky brightness distribution is written as a Taylor polynomial in frequency (*i.e.* a polynomial in I vs ν space, and not in $\log(I)$ vs $\log(\nu)$ space). The simulations used for these tests represented an 8 hour synthesis run with the EVLA in D configuration at L-band with an instantaneous bandwidth of 1 GHz. Wide-band data were obtained from the VLA *via* a series of short observations that cycled through a list of frequencies between 1 and 2 GHz. The end result of such an observation was a series of 10 to 20 VLA snapshots at 10 to 16 discrete frequencies within the range of the new EVLA receivers at L-band for those antennas that had them and within the range of the VLA receivers for the rest.

Telescope	EVLA (D configuration)
Observing Band	Lband (1-2 GHz)
Phase reference center	19:59:28.5 +40.44.01.5 J2000
Angular resolution	60, 40, 30 arcsec at 1.0,1.5,2.0 GHz
Cell size	8 arcsec
Image size	1024×1024 pixels (34 arcmin)
Number of channels	20
Channel Width	10 MHz
Spacing between channels	50 MHz
Instantaneous bandwidth	200 MHz (spread across 1 GHz)
Reference Frequency	1.5 GHz
Total integration time	8 hours
Integration time per visibility	200 s
System temperature T_{sys}	35K
Noise per visibility	7.2 mJy
Single-channel point-source sensitivity	22.8 μ Jy (theoretical)
Continuum point-source sensitivity	5.1 μ Jy (theoretical)
Expected dynamic range	8000
Achieved continuum RMS (off source)	8 μ Jy/beam
Achieved dynamic range	4000
Number of spectral series coefficients	$N_t = 5$
Set of spatial scales	0,6,10 pixels

Table 8.1: Parameters for Wide-Band EVLA Simulations : These simulations were designed to minimize the size of the simulated dataset and consist of a set of 20 frequency channels spread across the full 1 GHz instantaneous bandwidth with visibility samples being measured once every 3.3 minutes. A very low noise level was used in order to test and validate the algorithm.

8.1 Algorithm validation via simulated EVLA data

The multi-scale multi-frequency deconvolution algorithms described in chapter 7 were validated using datasets simulated for the EVLA. Section 8.1.1 presents narrow-field imaging results and section 8.1.2 illustrates the effect of a frequency-dependent primary beam and shows imaging results with and without primary-beam correction.

The simulations used wide-band flux components constructed as 2D Gaussians whose amplitudes follow a power-law with frequency. Extended emission was modeled by a sum of these flux components. Overlapping flux components with different power-law spectra we used to construct sources whose spectra were not pure power laws and also varied smoothly across the source. For wide-field imaging tests, antenna primary beams were included in the simulations by using visibility domain convolution functions that were constructed from frequency-dependent aperture illumination functions that rotate with time and have phase variations that model the EVLA beam squint. The parameters

of the simulated EVLA observation are listed in Table 8.1. The data products that were evaluated were the sky brightness distribution at a reference frequency along with maps of spectral index and spectral curvature.

8.1.1 Narrow-field imaging of compact and extended emission

Objective : The goals of this test are to assess the ability of the MS-MFS algorithm to reconstruct both spatial and spectral information about a source in terms of a linear combination of compact and extended flux components with polynomial spectra (flux model described in section 7.1.1) as well as to test how appropriate this flux model is when the true sky brightness is a complex extended source whose spectral characteristics vary smoothly across its surface.

Sky brightness : Wide-band EVLA observations were simulated for a sky brightness distribution consisting of one point source with spectral index of -2.0 and two overlapping Gaussians with spectral indices of -1.0 and $+1.0$. Fig.8.1 shows the reference frequency image of this simulated source, plots of the spectrum at different locations on the source, and the resulting spectral index and curvature maps. The spectral index across the resulting extended source varies smoothly between -1.0 and $+1.0$, with a spectral turnover in the central region corresponding to a spectral curvature of approximately 0.5 . Fig.8.2 shows the first three Taylor coefficient maps that describe this source.

MS-MFS Imaging : Two wide-band imaging runs were done using the MS-MFS algorithm and the results compared. The first used a multi-scale flux model (section 7.1.1) in which $N_t = 3$ and $N_s = 4$ with scale sizes defined by widths of $0, 6, 18, 24$ pixels and the second used a point-source flux model in which $N_t = 3$ and $N_s = 1$ with one scale function given by the δ -function (to emulate the MF-CLEAN algorithm described in section 6.2.1). A 5σ flux threshold of about $20\mu\text{Jy}$ was used as the termination criterion.

Results : The results from these imaging runs are shown in Fig. 8.3 (three Taylor coefficients), Figure 8.4 shows residual images over a larger region of the sky, and Fig. 8.5 shows the intensity at the reference frequency, spectral index and spectral curvature. All figures show the results with both MS-MFS and MF-CLEAN.

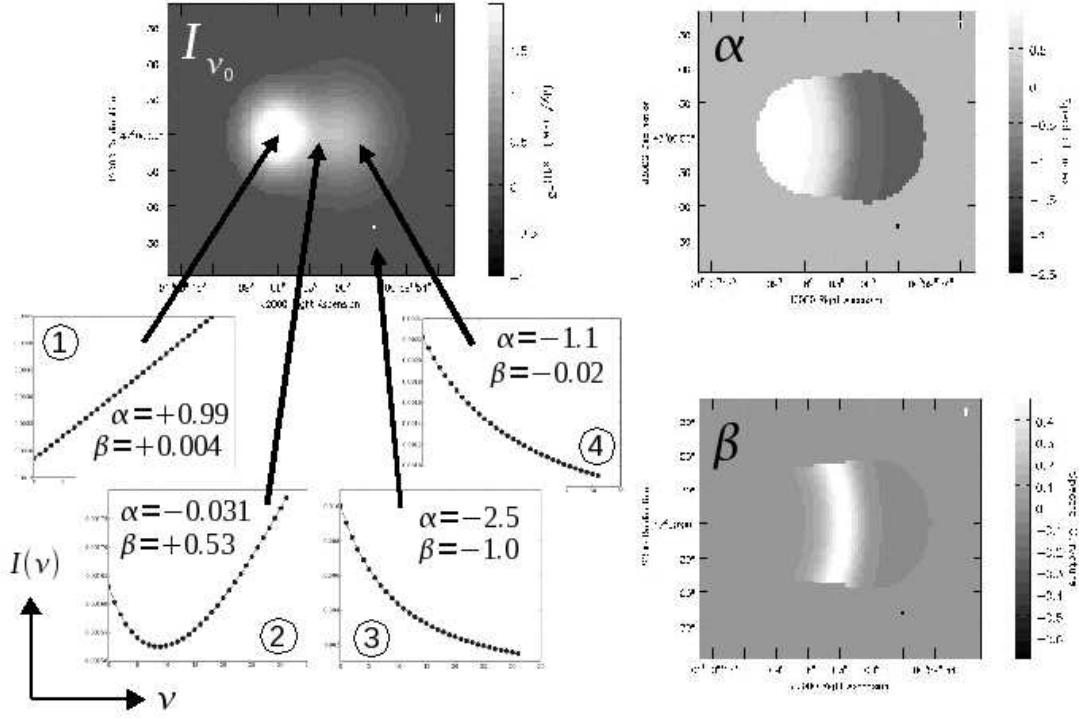


Figure 8.1: Simulated wide-band sky brightness distribution : These images represent the wide-band sky brightness distribution that was used to simulate EVLA data to test the MS-MFS algorithm. The image on the top left shows the total intensity image of the source at the reference frequency \vec{I}_{ν_0} . The plots on the bottom left show spectra (and their power law parameters) at 4 different locations. The spectral index varies smoothly between about +1 and −1 across the extended source and is −2.5 for the point source. The spectral curvature has significant values only in the central region of the extended source where the spectrum turns over within the sampled range. The images on the right show these trends in the form of spectral index (top) and spectral curvature (bottom) maps.

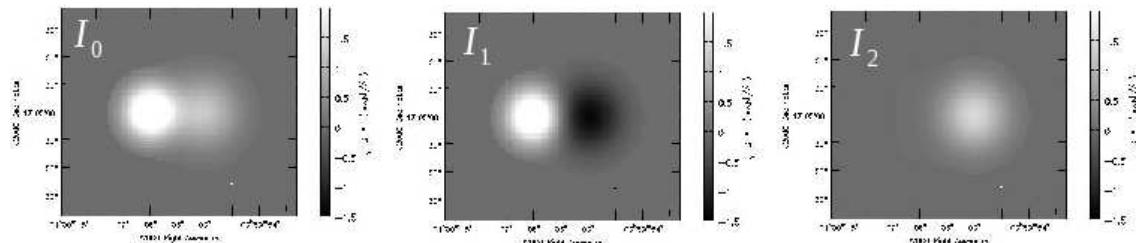


Figure 8.2: True Taylor coefficient images : These images show the first three Taylor coefficients for the polynomial expansion of the wide-band flux distribution shown in Fig. 8.1. These images are the (left) intensity at the reference frequency $I_0 = I_{\nu_0}$, (middle) first-order Taylor-coefficient $I_1 = \alpha I_{\nu_0}$ and (right) second-order Taylor-coefficient $I_2 = (\alpha(\alpha - 1)/2 + \beta) I_{\nu_0}$ (see Eqn. 6.22). All images are displayed at the same flux scale.

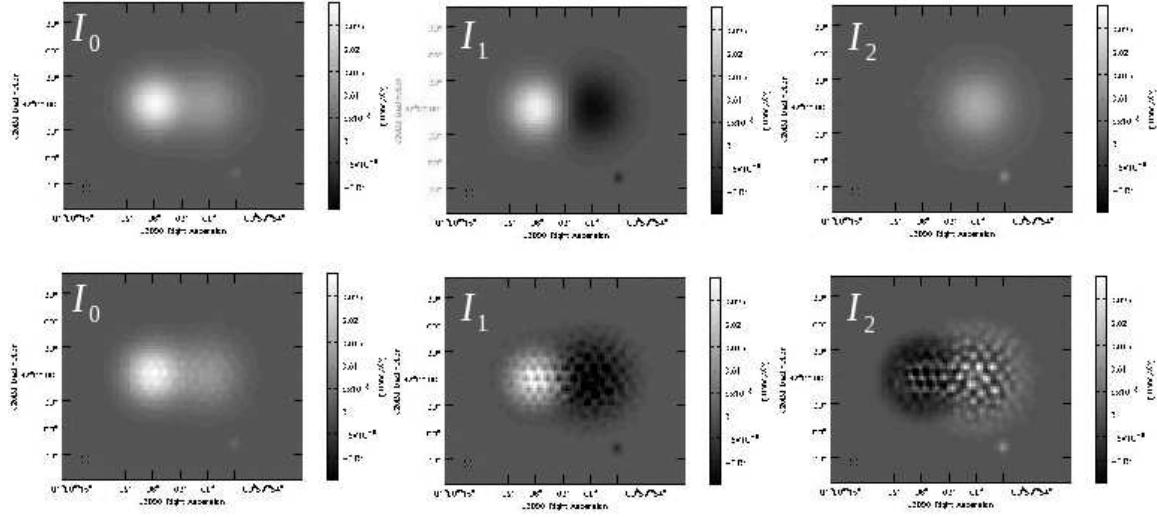


Figure 8.3: Reconstructed Taylor coefficient images : These images show the first three Taylor coefficients (similar to Fig. 8.2) obtained using two different wide-band flux models. The top row shows the results of using a multi-scale wide-band flux model (MS-MFS) and the bottom row shows the results of using a point-source wide-band flux model (MF-CLEAN, or MS-MFS with only one spatial scale given by a δ -function). All images are displayed at the same flux scale.

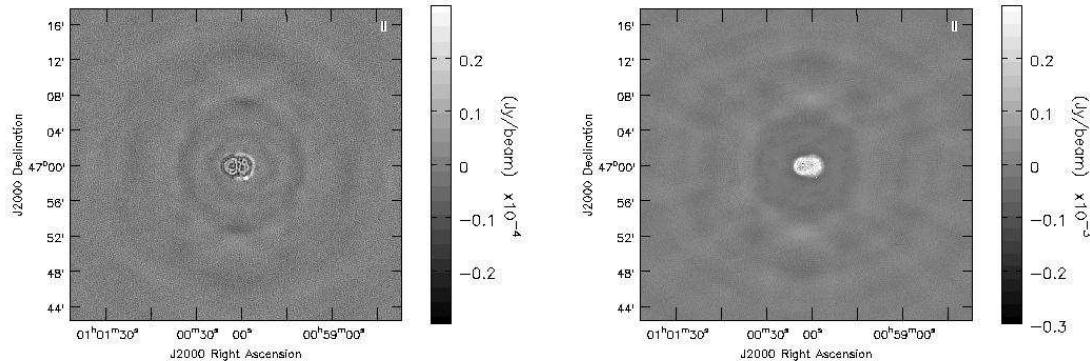


Figure 8.4: Residual images : This figure shows the residual images obtained after applying MS-MFS to wide-band EVLA data simulated for the sky brightness distribution shown in Fig. 8.5. The residual image on the left is obtained when a multi-scale flux model was used (MS-MFS). The RMS noise on source is about $20 \mu\text{Jy}$ and off source is $5 \mu\text{Jy}$. Compare this with the residual image on the right from a point-source deconvolution (MF-CLEAN) where the on source RMS is about 0.2 mJy and off source is $50 \mu\text{Jy}$. (Note that the displayed data ranges are different for these two images. The flux scale for the image on the left is $\pm 0.3 \times 10^{-4}$ and for the right is $\pm 0.3 \times 10^{-3}$.) This clearly demonstrates the advantage of using a multi-scale flux model.

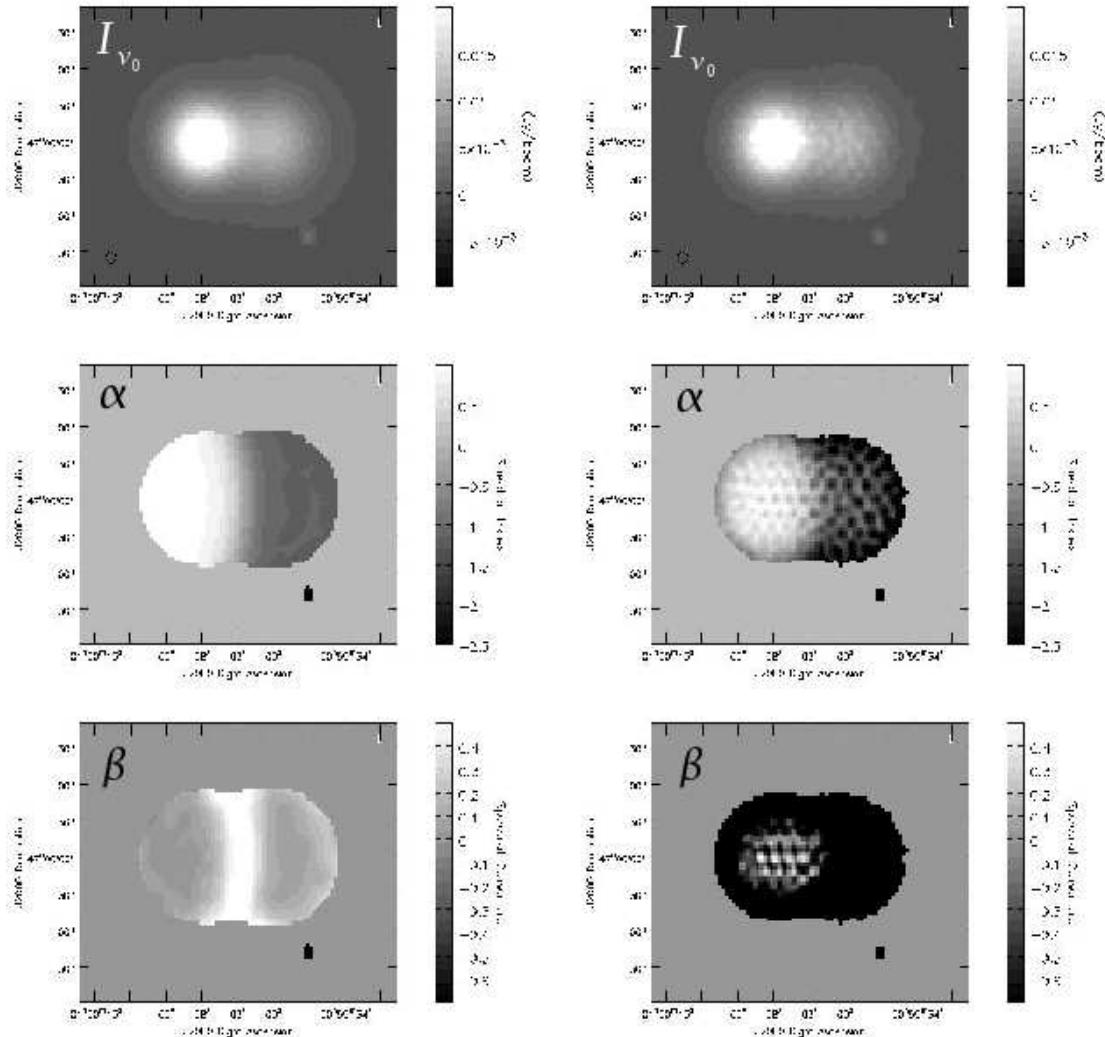


Figure 8.5: MS-MFS final imaging data products : These images show the results of applying MS-MFS to wide-band EVLA data simulated for the sky brightness distribution described in Fig.8.1. The left column shows the results of using a multi-scale wide-band flux model (MS-MFS) and the right column shows the results of using a point-source wide-band flux model (MF-CLEAN, or MS-MFS with only one spatial scale given by a δ -function). The top, middle and bottom rows correspond to the intensity image at the reference frequency I_{ν_0} , the spectral index α and spectral curvature β maps respectively. The flux scale for each left/right pair of images is the same, and the sharp source boundaries in the spectral index and curvature maps are because of a flux threshold used to compute them. With a multi-scale flux model (MS-MFS, left), the reconstructions of α and β are accurate to within 0.1 in high signal-to-noise regions. With a point-source flux model (MF-CLEAN, right), deconvolution errors break extended emission into flux components of the size of the resolution element and these errors transfer non-linearly to the spectral index and curvature maps. Table 8.2 compares the true and reconstructed values of I_{ν_0}, α, β for three regions of this sky brightness distribution.

The main points to note from these images are listed below.

1. With a multi-scale multi-frequency flux model (MS-MFS) the spectral index across the extended source was reconstructed to an accuracy of $\delta\alpha < 0.05$ with the maximum error being in the central region where the spectral index goes to zero and $N_t = 3$ is too high for an accurate fit (section 6.2.4 describes how the choice of N_t affects the solution process). The spectral curvature across the extended source was estimated to an accuracy of $\delta\beta < 0.1$ in the central region with the maximum error of $\delta\beta \approx 0.2$ in the regions where the curvature signal goes to zero and the source surface brightness is also minimum (the outer edges of the source).
2. With a multi-frequency point-source model (MF-CLEAN) the accuracy of the spectral index and curvature maps was limited to $\delta\alpha \approx 0.1, \delta\beta \approx 0.5$. This is because the use of a point source model will break any extended emission into components the size of the resolution element and this leads to deconvolution errors well above the off-source noise level (note the difference between the intensity images $I(v_0)$ produced with MS-MFS vs MF-CLEAN). Error propagation during the computation of spectral index and curvature as ratios of these noisy reconstructed images leads to high error levels in the result.
3. The imaging run that used a multi-scale image model was terminated at a 5σ noise threshold. The peak residual is about 20 μJy and the off-source RMS is 5 μJy (close to the theoretical RMS of 3 μJy as listed in Table 8.1). The imaging run that used a point-source model was terminated after at least four successive major cycles failed to reduce the peak residual below 200 μJy despite an apparent decrease in the residuals during the minor cycle iterations. The off source RMS in the result is about 50 μJy .

Error Estimates : The errors on the reconstructed intensity map at the reference frequency, spectral index and curvature were estimated based on a comparison with smoothed versions of the corresponding true images. Table 8.2 shows these numbers for three regions on the simulated sky brightness distribution (labelled as 1,2 and 3). One general point to note from these results is that MS-MFS tends to give more accurate results than MF-CLEAN because the errors on the reconstructed α and β depend strongly on the magnitude of the deconvolution error in the coefficient images. MF-CLEAN has larger deconvolution errors in the coefficient images, and since it is unlikely that these errors preserve the ratios between the coefficient images, the errors in the spectral index and curvature maps increase. With sufficient signal to noise ($\text{SNR} \approx O(10)$ ¹ for spectral index and $\text{SNR} \approx O(100)$ for spectral curvature), it is possible to reconstruct the spectral index and curvature across the source to accuracies of within 0.1.

¹The expression $O(n)$ represents 'of the order of n'.

Observed Errors with MS-MFS	Region 1	Region 2	Region 3
Peak brightness I_0 (Jy/beam)	0.0292	0.0128	0.0032
On-source residual I_{on}^{res}	1×10^{-5}	1×10^{-5}	2×10^{-5}
Off-source residual I_{off}^{res}	3×10^{-6}	3×10^{-6}	3×10^{-6}
$\delta I = I_0 - I^{true} $	1×10^{-5}	4×10^{-5}	1×10^{-4}
$SNR = I_0 / \max(I_{on}^{res}, \delta I)$	1800	320	32
Measured $\alpha \pm \delta\alpha$	0.99 ± 0.005	-0.13 ± 0.11	-2.45
Measured $\beta \pm \delta\beta$	0.016 ± 0.01	0.61 ± 0.05	-1.12

Observed Errors with MF-CLEAN	Region 1	Region 2	Region 3
Peak brightness I_0 (Jy/beam)	0.0309	0.0129	0.0031
On-source residual I_{on}^{res}	2×10^{-4}	4×10^{-4}	2×10^{-4}
Off-source residual I_{off}^{res}	1.2×10^{-5}	1.2×10^{-5}	1.2×10^{-5}
$\delta I = I_0 - I^{true} $	1×10^{-4}	1×10^{-4}	1×10^{-4}
$SNR = I_0 / \max(I_{on}^{res}, \delta I)$	190	43	31
Measured $\alpha \pm \delta\alpha$	0.7 ± 0.17	-0.17 ± 0.26	-2.58
Measured $\beta \pm \delta\beta$	-0.5 ± 0.3	-0.5 ± 0.35	-1.19

Table 8.2: Measured errors with MS-MFS on Simulated Data : These tables compare the true and measured values of the peak flux, spectral index and spectral curvature for three regions of the simulated sky brightness distribution (labelled as 1,2 and 3 in Fig. 8.1) and two algorithms (top) MS-MFS and (bottom) MF-CLEAN (see Fig. 8.5 for the corresponding images). The purpose of this comparison is to (a) show that when there is sufficient SNR, MS-MFS is more accurate than MF-CLEAN and (b) give examples of how the error bars on α and β vary as a function of SNR. In region 1, the spectrum is close to a pure power law with no curvature ($\alpha = 0.99, \beta = 0.0$). In region 2, there is a strong spectral turnover but the average spectral index is very small ($\alpha = 0.031, \beta = 0.535$). Region 3 is the point source located at the edge of the extended emission ($\alpha = -2.5, \beta = -1.0$). The measured errors $\delta\alpha, \delta\beta$ were obtained by constructing error images from the difference between the true and reconstructed spectral index and curvature images, and then calculating the standard-deviation of all points within a finite region of these difference maps (they are approximate). Region 3 contains no error-bars on α, β because the above calculation cannot be done with one pixel.

8.1.2 Wide-field imaging with Primary-Beam correction

Objective : The goal of this simulation is to test the MS-MFS algorithm with primary beam correction to reconstruct both compact and extended emission whose spectral structure is modified by the frequency dependence of the primary beam. The primary beams are simulated with time-variability arising from their rotation with time as well as beam squint. This is to test for any difference in performance and imaging fidelity when direction-dependent corrections are applied as a single post-deconvolution image-domain correction versus a combination of visibility-domain and image-domain operations.

Sky brightness and primary beams : Wide-field wide-band EVLA observations were simulated for a sky brightness distribution consisting of one large 2D Gaussian (about 10 arcmin in diameter) with a constant spectral index of -1.0 across its entire surface and two point sources with spectral indices of 0.5 and 0.0. The Gaussian is centered at the 80% point of the reference frequency primary beam and the spectral index due to the primary beam ranges between 0 and -0.5 across its surface. The two point sources are located near the 70% point of the reference-frequency primary beam where the spectral index of the beam is about -0.5. EVLA primary beams were simulated from numerically derived aperture illumination functions [Brisken 2003] and applied *via* time-varying visibility-domain convolution functions during the simulation (as shown in Eqn. 4.8).

MS-MFS Imaging with Primary-beam correction : The MS-MFS algorithm was run with $N_t = 5$, and $N_s = 3$ with the scale-widths in pixels are [0,6,20]. A 5σ convergence threshold was used as the termination criterion. Wide-band primary-beam correction was done in two different ways and their results compared. The first method used a single post-deconvolution image-domain correction that divided out a polynomial model of the time-averaged primary beam (as described in the caption of Fig.7.5). The second method used a combination of visibility-domain and image domain corrections that accounted for the time-variability of the antennas (rotation with time) and the effect of beam squint (a polarization dependent pointing offset arising from the location of the feeds on EVLA antennas).

Results : Figure 8.6 shows the results of these simulations. The image on the top left shows the reference frequency intensity image after correction for the primary beam. The image on the top right shows the spectral index map without primary beam correction and the bottom row of images are the corrected spectral index maps obtained *via* the two methods described above.

The main points to note from these results are as follows.

1. From the un-corrected spectral index image we can see that the spectral indices of the point sources are the sum of that of the source and of the primary beam at that location. The spectral index of the extended source is tilted with the numbers ranging between -1.0 and -1.5 from one edge of the source to the other. Both point sources have taken on an additional spectral index of -0.5 .
2. From the bottom two images we can see that both methods will give the same qualitative reconstruction of the true spectral index of the source, but the second method (right) has much better noise properties. This is only because it accounts for the variability of the primary beam and is not restricted to the use of a time-averaged primary beam.
3. The accuracy to which the spectral indices of the point sources were reconstructed was about $\delta\alpha = 0.01$. For the extended source, the errors are dominated by the residual multi-scale deconvolution errors that prevent a smooth reconstruction even in the (top right) image of the uncorrected spectral index (wide-band versions of the MEM and ASP-CLEAN algorithms might be required to reduce these errors). The accuracy with which the spectral index was computed across the extended source was about $\delta\alpha \approx 0.2$.

These results show that for a field of view within the HPBW of the primary beam at the reference frequency, it is possible to model the frequency dependence of the beam by a power law with varying index, and use this model to do image-domain corrections of the beam. The largest field-of-view over which this model has been shown to work is down to the few-percent point of the beam at the highest frequency (near the first null at the highest frequency and close to the HPBW at the lowest frequency; see Fig.5.3). Beyond this field-of-view, the power-law model breaks down, and explicit polynomial division will be required to correct for the primary beam.

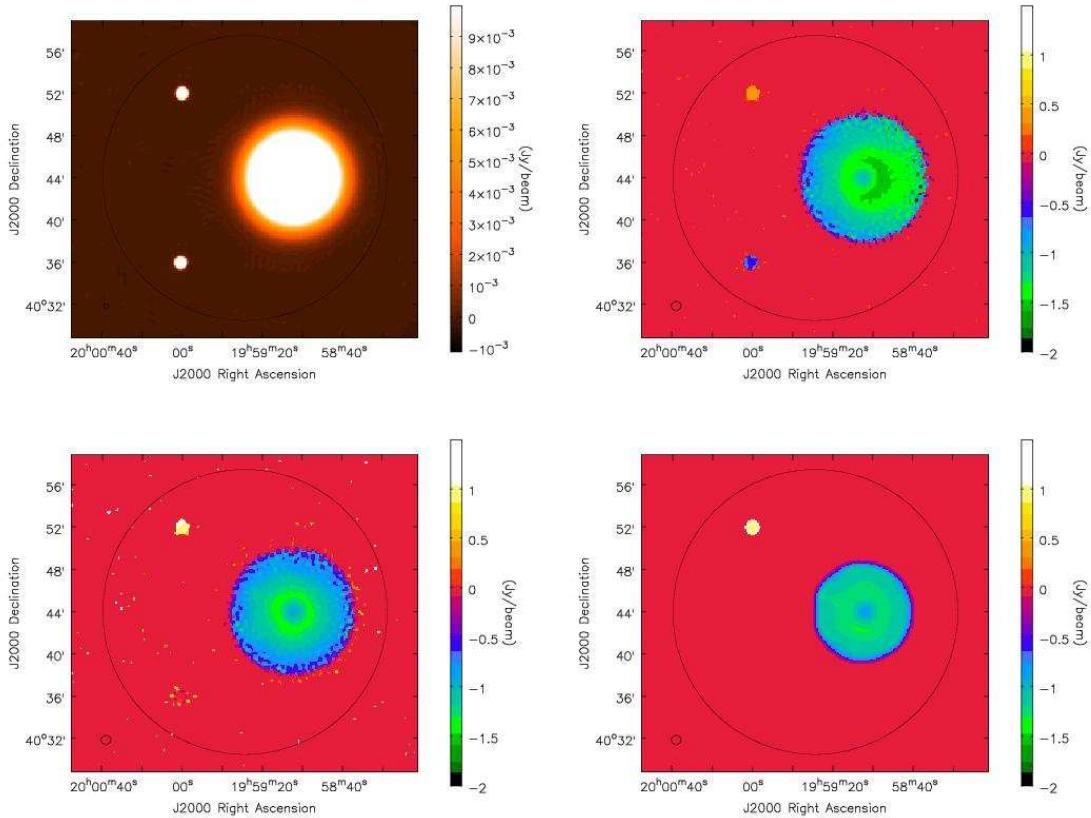


Figure 8.6: MS-MFS with wide-band primary beam correction on simulated EVLA data : The image on the top left is the intensity image at the reference frequency and shows two point sources (spectral index of +1.0 (top) and 0.0(bottom)) and one extended source with a constant spectral index of -1.0. The image on the top right shows the spectral index map constructed by using MS-MFS without any primary beam correction. The apparent spectral indices of the point sources are +0.5 (top) and -0.5 (bottom) and range from -1.0 to -1.5 for the extended source (left to right). The second row of images shows the spectral index maps after primary-beam correction *via* a single post-deconvolution image-domain correction with an average primary beam and its spectrum (left, section 4.2.1) and a combination of visibility and image domain corrections that takes into account the time-variability or rotation of the beam and the effect of beam squint (right, section 4.2.2).

8.2 Feasibility Study of MFS in various situations

This section consists of a set of imaging examples that illustrate the feasibility of wide-band synthesis imaging mainly when the uv sampling is insufficient to directly measure all the spatial and spectral structure within the full range of spatial frequencies allowed by the broad-band receivers. These examples were chosen to emphasize the role of an appropriate flux model in an image reconstruction algorithm and how it can often provide physically realistic *a priori* information to the solution process (see the first two pages of chapter 7 for an introductory discussion about the choice of an appropriate flux model).

Section 8.2.1 describes the reconstruction of source spectra at spatial scales that are unresolved at the low-frequency end of the band but resolved at the high-frequency end. This example shows that for broad-band synchrotron emission it is possible to reconstruct the source spectrum at the angular resolution allowed by the highest frequency in the band. Section 8.2.2 describes the reconstruction of spectra at very large spatial scales for which the visibility function falls within the central hole in the uv -coverage for the upper half of the frequency range. This example illustrates an ambiguity between spatial and spectral structure that can arise from such measurements and shows that the use of *a priori* total-flux constraints can solve this problem. Section 8.2.3 shows how the multi-scale wide-band flux model used in the MS-MFS algorithm naturally separates the contributions from overlapping sources that differ in spatial and spectral structure. Section 8.2.4 demonstrates how the MS-MFS algorithm performs when the spectrum of the radio emission is not a smooth low-order polynomial. This example tests the applicability of the wide-band model to band-limited emission which can be represented with a 4th or higher order polynomial (and not just power-law spectra).

8.2.1 Moderately Resolved Sources

Objective : Traditionally, spectral structure has been measured from wide-band interferometry data only after making a set of narrow-band images and smoothing them to the angular resolution of the lowest frequency in the band. For the 2:1 frequency ranges now becoming available, the angular resolution changes by a factor of two across the band, and smoothing the images to the lowest resolution results in a considerable loss of information. The goal of this test is to demonstrate how a flux model that accurately describes the type of emission being observed can influence the wide-band imaging process to reconstruct the spectral structure of the incoming radio emission at the angular resolution of the highest frequency in the sampled range, even though the lower frequency data measure the sky brightness at lower angular resolutions.

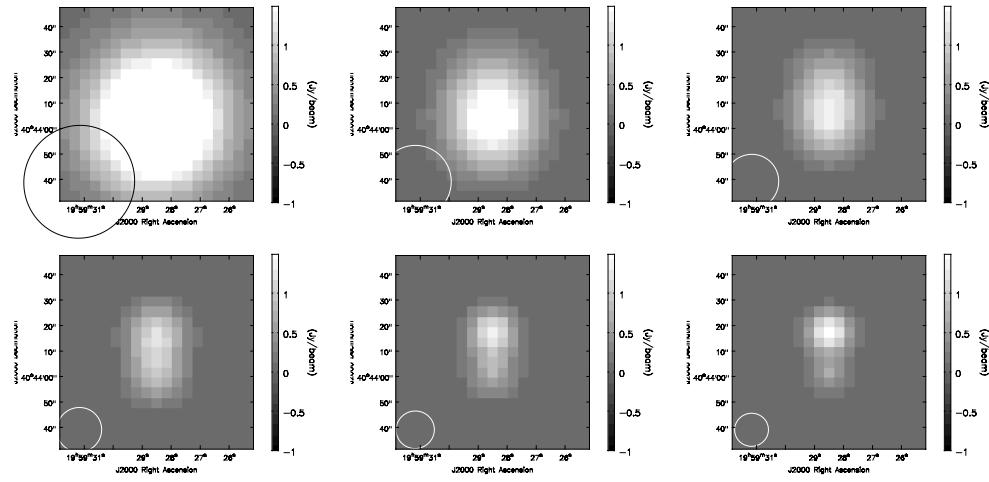


Figure 8.7: Moderately Resolved Sources – Single-Channel Images : These figures show the 6 single-channel images generated from simulated EVLA data between 1 and 4 GHz in the EVLA D-configuration. The angular resolution at 1 GHz is 60 arcsec, and at 4 GHz is 15 arcsec and the white circles in the lower left corner shows the resolution element decreasing in size as frequency increases. The sky brightness consists of two point sources, each of flux 1.0 Jy at a reference frequency of 2.5 GHz and separated by 18 arcsec. The pixel size used in these images is 4.0 arcsec. From these single-channel images we can see that the sources begin to be resolved only at the higher end of this frequency range, and at the lower end of the band is barely distinguishable from a single point source centered on the bottom point source. The top point source has a spectral index of +1.0 and the bottom one has a spectral index of -1.0.

EVLA Simulation : Wide-band EVLA data were simulated for the D-configuration across a frequency range of 3.0 GHz with 6 frequency channels between 1 and 4 GHz (600 MHz apart). This wide frequency range was chosen to emphasize the difference in angular resolution at the two ends of the band (60 arcsec at 1 GHz, and 15 arcsec at 4.0 GHz). The sky brightness chosen for this test consists of a pair of point sources separated by a distance of 18 arcsec (about one resolution element at the highest frequency), making this a moderately resolved source. These point sources were given different spectral indices (+1.0 for the top source and -1.0 for the bottom one). Figure 8.7 shows the 6 single-channel images of this source. At the low frequency end, the source is almost indistinguishable from a single flux component centered at the location of the bottom source whose flux peaks at the low-frequency end. The source structure becomes apparent only in the higher frequencies where the top source (with a positive spectral index) is brighter. Figure 8.8 shows the multi-frequency *uv*-coverage and the sampled visibilities in this simulated dataset. These plots show that the double-source structure becomes apparent only beyond the first few frequencies in the range, making this a suitable dataset to use to test the MS-MFS algorithm on sources that are unresolved at one end of the band and resolved at the other.

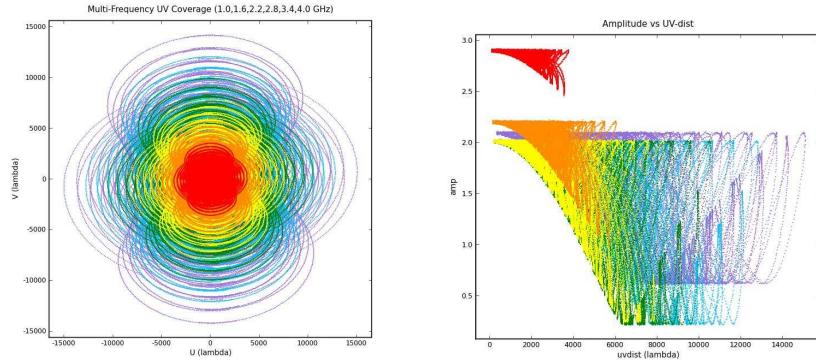


Figure 8.8: Moderately Resolved Sources - uv -coverage and Visibility-Plot : These plots show the multi-frequency uv -coverage (left) and the sampled visibilities (right). The colours indicate frequency, going from red to violet as frequency increases. The visibility plot shows that at the lowest frequency, the interferometer sees the sky as a single point source whose flux is the sum of both point sources (~ 2.9 Jy) at 1 GHz. As the frequency increases, the double-source structure becomes apparent in the form of visibility-domain fringes.

MS-MFS Imaging Results :

1. These data were imaged using the MS-MFS algorithm with $N_t = 3$ and $N_s = 1$ with only one spatial scale (a δ -function). Figure 8.9 shows the results of this imaging run. The intensity distribution, spectral index and curvature of this source were recovered at the angular resolution allowed by the 3.6 GHz samples (18 arcsec). These results show that for a source that can be modeled as a set of flux components (in this case point-sources) with polynomial spectra, even partial spectral measurements at the highest angular resolution are sufficient to reconstruct the full spectral structure.
2. A second imaging run was performed using only the first and last channels (1.0 GHz and 4.0 GHz). The source is almost completely unresolved at 1 GHz (point sources separated by 18 arcsec within a 60 arcsec resolution element), and just resolved at 4 GHz (with an 15 arcsec resolution element). The goal of this exercise was to test the limits of this algorithm and the ability of the flux model to constrain the solution when the data provide insufficient constraints. The MS-MFS algorithm was run with $N_t = 2$ and $N_s = 1$ and used the same number of iterations as the previous example. Fig.8.10 contains the resulting intensity image and spectral index map and shows that it is still possible to resolve the source and measure its spectral index at the resolution of the highest frequency. However, the deconvolution errors are considerably higher. The obtained peak residual of 5 mJy is not much larger than the 3 mJy level obtained when all 6 channels were used while imaging, indicating that this reconstruction is not well constrained by the data and the model plays a very significant role.

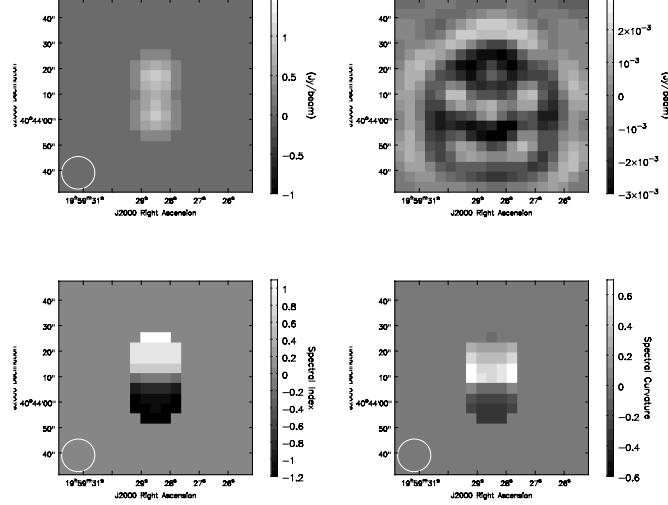


Figure 8.9: Moderately Resolved Sources – MSMFS Images : These images show the results of running MS-MFS on EVLA data that was simulated to test the algorithm on moderately resolved sources. The test sky brightness distribution consists of two point sources with spectral indices +1.0 (North) and -1.0 (South) separated by one resolution element at the highest frequency. The four images shown here are the intensity at 2.5 GHz (top left), the residual image with a peak residual of 3 mJy (top right), the spectral index showing a gradient between -1 and +1 (bottom left) and the spectral curvature which peaks between the two sources and falls off on either side (bottom right). These results demonstrate that an appropriate flux model will constrain the solution to a physically realistic one even when the spectral measurements are incomplete at the highest resolution.

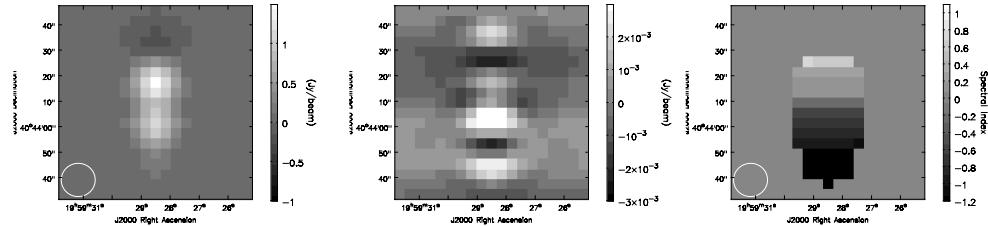


Figure 8.10: Moderately Resolved Sources – MSMFS Images using first and last channels : These images show the result of MS-MFS on two channels of data with very different angular resolutions (60 arcsec at 1 GHz, and 15 arcsec at 4 GHz). The intensity image (left) and the spectral index image (right) show that the intensity and spectrum have been reconstructed at the 15 arcsec resolution. However although the peak residual (middle) of about 5 mJy is not much higher than in Fig.8.9, there are visible deconvolution errors that lead to errors in peak intensity and spectral index.

8.2.2 Emission at very Large Spatial Scales

Objective : This section demonstrates an ambiguity between spatial and spectral structure that can arise when multi-frequency measurements are made of very large-scale emission. The goal of this exercise is to show the effect of this ambiguity in the images and spectra of very large scale emission that are reconstructed by the MS-MFS algorithm and to suggest a possible remedy.

Consider a very large (extended) flat-spectrum source whose visibility function falls mainly within the central hole in the *uv*-coverage at the highest observing frequency. With multi-frequency measurements, the size of the central hole in the *uv*-coverage increases with observing frequency, and for this source the minimum spatial frequency sampled per channel will measure a decreasing peak flux level as frequency increases. Since the reconstruction below the minimum spatial frequency involves an extrapolation of the measurements and is un-constrained by the data, these decreasing peak visibility levels can be mistakenly interpreted as the result of a source whose amplitude itself is decreasing with frequency (a less-extended source with a steep spectrum). Usually, a physically realistic flux model is used to apply constraints in these unsampled regions of the *uv*-plane and MS-MFS models the sky brightness with polynomial spectra associated with a set of extended 2D symmetric flux components. However, with this model a large flat-spectrum source and a smaller steep-spectrum source are both allowed and considered equally probable. This creates an ambiguity between the reconstructed scale and spectrum that cannot always be resolved directly from the data, and requires additional information (perhaps a low-frequency narrow-band image to constrain the spatial structure, low-resolution spectral information, or total-flux constraints).

EVLA Simulation : Wide-band EVLA data were simulated for the D-configuration across a frequency range of 3.0 GHz centred at 2.5 GHz. (6 frequency channels located 600 MHz apart between 1.0 and 4.0 GHz). The size of the central hole in the *uv*-coverage was increased by flagging all baselines shorter than 100 m and the wide frequency range was chosen to emphasize the difference between the largest spatial scale measured at each frequency. (0.3 k λ or 10.3 arcmin at 1.0 GHz, and 1.3 k λ or 2.5 arcmin at 4.0 GHz).

The sky brightness chosen for this test consists of one large flat-spectrum 2D Gaussian whose FWHM is 2.0 arcmin (corresponding to 1.6 k λ at the reference frequency of 2.5 GHz), and one steep spectrum point-source ($\alpha=-1.0$) located on top of this extended source at 30 arcsec away from its peak.

MS-MFS Imaging Results : These data were imaged using the MS-MFS algorithm with $N_t = 3$ and $N_s = 3$ with scale sizes given by [0,10,30] pixels. Two imaging runs were performed with these parameters and both were terminated after 100 iterations in order to be able to compare their performance in terms of the peak residuals.

Fig. 8.11 shows the visibility amplitudes present in the simulated data (left column) as well as in the reconstructed model (right column) at each of the 6 frequencies for these two imaging runs (top, bottom). Fig. 8.12 shows images of the intensity, spectral index and residuals for these runs and compares them to the true sky brightness reconstructed when all frequencies sample at least 95% of the total flux of the source.

1. The first imaging run applied the MS-MFS algorithm to the simulated data after flagging all baselines below 200m. No additional constraints were used on the reconstruction. The visibility plots and imaging results show that from these data it is not possible to distinguish large flat-spectrum source from a slightly less-extended steep spectrum source. This occurs because the visibility function is unconstrained by the data within the central uv hole and given the MS-MFS flux model, both source structures are equally probable. Note that the spectrum of the point-source was correctly estimated as -1.0 . This run was repeated a few times with slightly different input scale sizes, and the results changed between a flat-spectrum source and a source with a steep spectrum. If a scale size corresponding to the exact size of the source was present in the set, the algorithm was able to reconstruct the correct flux and spectrum.
2. A second imaging run was performed on the same dataset, but this time with additional information in the form of total-flux constraints at each observing frequency. These constraints were added in by retaining a small number of very short-baseline measurements at each frequency in order to approximate the presence of total-flux (or integrated flux) estimates (only baselines between 25 m and 100 m were flagged from the original EVLA D-configuration simulated data). In practice, these constraints could be provided by single-dish measurements or estimates from existing low-resolution information about the structure and spectrum of the source. The visibility plots and imaging results with this dataset show that the short-spacing flux estimates were sufficient to bias the solution towards the correct solution in which the large extended source has a flat spectrum and the point source has a spectral index of -1.0 . Note that the residuals are at the same level as in the previous run. This demonstrates that without the additional information about total-flux per frequency, both flux models are equally poorly constrained by the data themselves.

These results show that in the central unsampled region of the uv -plane where there are no constraints from the data, the MS-MFS flux model can produce ambiguous results and additional information about the flux at low spatial-frequencies is required (perhaps in the form of total-flux constraints per frequency). For complex spatial structure on these very large scales, the additional constraints may need to come from existing low-resolution images of this field and the associated spectra. One way to avoid this problem altogether (but lose some information) is to flag all spatial-frequencies smaller than u_{min} at ν_{max} and not attempt to reconstruct any spatial scales larger than what ν_{max} allows.

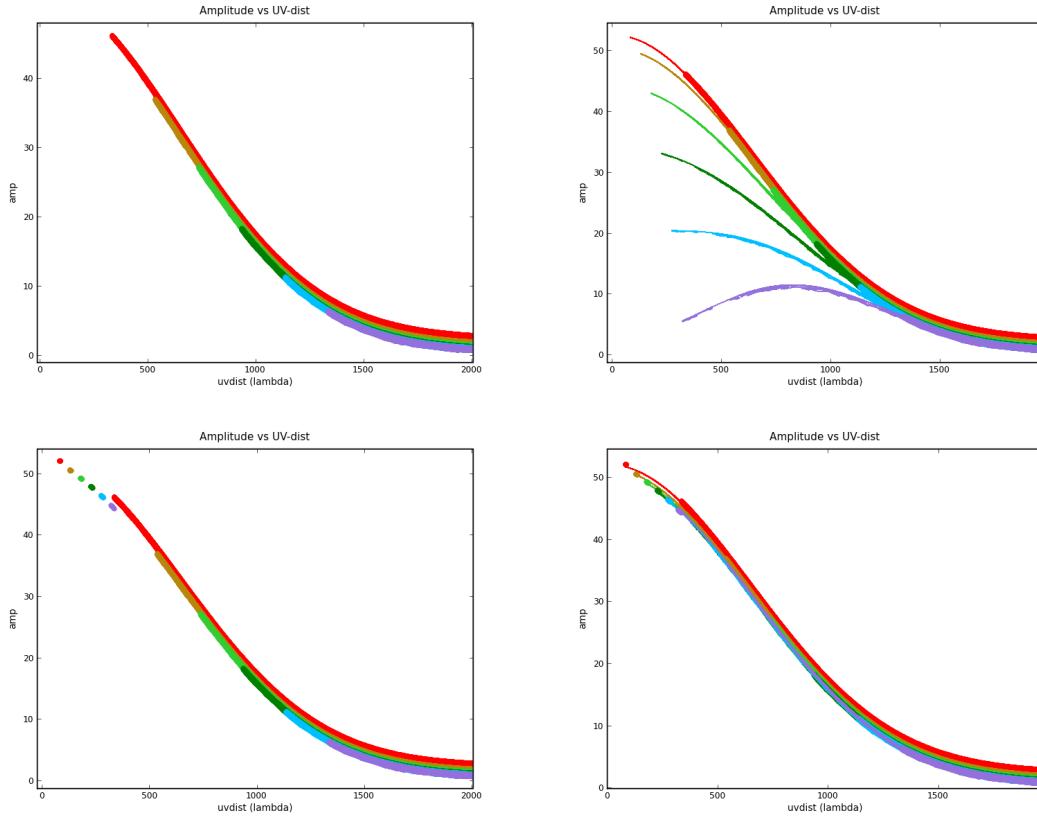


Figure 8.11: Very Large Spatial Scales - Visibility plots : These plots show the observed (left) and reconstructed (right) visibility functions for a simulation in which a large extended flat-spectrum source is observed with an interferometer with a large central hole in its uv -coverage. The different colours/shades in these plot represent 6 frequency channels spread between 1 and 4 GHz. These data were imaged in two runs. The first imaging run (top row) used only baselines $b > 100$ m to emphasize the changing size of the central hole in the uv -coverage across the broad frequency range. The plot on the top left shows how the different frequencies measure very different fractions of the integrated flux of the large flat-spectrum source. The plot on the right shows that these data can be mistakenly fit using a less-extended source with a steep spectrum (instead of the large single source with a flat spectrum). This is possible because within the central uv hole the spectrum is un-constrained by the data and given the MS-MFS flux model, both source structures are equally probable. The second imaging run (bottom row) used baselines $b < 25$ m in addition to $b > 100$ m to approximate the addition of nearly total-flux measurements to the first dataset to attempt to constrain the solution. The plot on the bottom right shows that this additional information in the form of short-spacing constraints (or very low-spatial frequency measurements) is sufficient to be able to reconstruct the correct sky brightness distribution. Figure 8.12 shows the images that resulted from these tests.

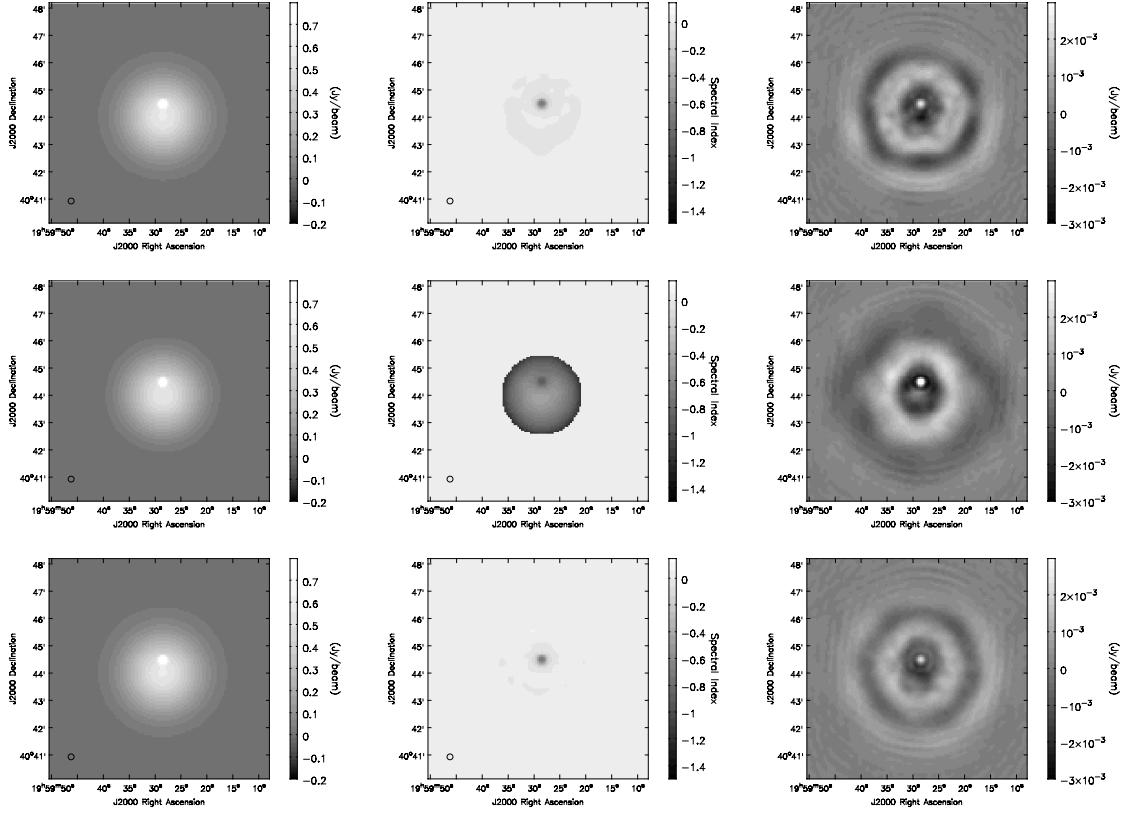


Figure 8.12: Very Large Spatial Scales - Intensity, Spectral Index, Residuals : These images show the intensity distribution (left), spectral index (middle) and the residuals (right) for three different imaging runs that applied the MS-MFS algorithm to the simulated EVLA D-configuration data described in this section (note that the flux scale used for the residual images in the right column is 3 orders of magnitude smaller than the scale used for the intensity image in the left column). The true sky flux consists of one large flat-spectrum symmetric flux component and one steep-spectrum ($\alpha = -1.0$) point source.

Top Row : When all baselines are used for imaging, each frequency samples more than 95% of the integrated flux. This is sufficient to reconstruct the true brightness distribution and spectrum.

Middle Row : When the central *uv*-hole is increased in size by using only baseline $b > 100m$, the reconstructed model is a slightly smaller flux component (compare the left column of images) with a steep spectrum (compare the middle column of images).

Bottom Row : When very short spacing (approximately total-flux) estimates are included during imaging (using spacings $b < 25m$ and $b > 100m$), the true sky brightness distribution is again recovered. Note that the large-scale residuals in all three runs are at the same level (2 mJy). These results show that the spectra are unconstrained by the data for very large spatial scales whose visibility functions fall within the central *uv*-hole at the highest frequency in the band, and additional information is required.

8.2.3 Foreground/Background Sources with Different Spectra

Objective : This section contains a simple example of wide-band imaging with background subtraction for the case where a compact foreground source of emission lies on top of a more extended source with a different spectrum. When there are overlapping sources with different spectral structure, the result of wide-band imaging represents the combined flux and spectrum. Similar to standard imaging, if the flux and spectrum of the background are available, the flux and the spectrum of the foreground source can be separated from the background *via* a simple polynomial subtraction (using the polynomial-coefficients).

The use of a multi-scale flux model has an additional advantage when it comes to background subtraction. When the spatial scales of the foreground and background flux are very different, the MS-MFS algorithm naturally separates the two and models the integrated flux as a sum of compact and extended flux components with different spectra. Note that this is true for any multi-scale image flux model, irrespective of spectrum. In the ASP-CLEAN algorithm where the final data product is constructed from a list of flux components, this separation is done naturally and components can be picked out from the results.

EVLA Simulation : Data were simulated for the EVLA D-configuration with 6 frequency channels spread between 1 and 2 GHz. The sky brightness consists of one large 2D Gaussian of integrated flux of 100 Jy over a 4 arcmin radius (peak flux of about 1 Jy/beam at 30 arcsec resolution (EVLA-D at 2.0 GHz)) and $\alpha=1.0$, two 1 Jy point sources on top of this extended source with spectral indices given by $\alpha=+0.5$, -0.5, and one isolated 1 Jy point source with $\alpha=-0.5$.

MS-MFS Imaging Results : The MS-MFS algorithm was applied to this dataset using $N_t = 5$ and $N_s = 3$ with the set of scales sizes given by [0, 10, 30] pixels. Iterations were terminated using a 1 mJy threshold. Figure 8.13 shows the resulting images of the first two polynomial coefficients and the spectral index. Background subtraction is done as a polynomial subtraction. The first two polynomial coefficients are given as follows.

$$I_0^{total} = I_0^{back} + I_0^{front} \quad (8.1)$$

$$I_1^{total} = I_0^{back} \alpha^{back} + I_0^{front} \alpha^{front} = I_0^{total} \alpha^{total} \quad (8.2)$$

The values of I_0^{total} , I_0^{total} , I_0^{back} and α^{back} are measured from the images. The background flux and spectrum are estimated from a region near the foreground source. The measured and corrected flux and spectral indices of the two foreground sources are listed in Table 8.3. These results show how background subtraction can be performed using the polynomial coefficient images before constructing the spectral index maps. Alternatively, if only intensity and spectral maps exist, polynomial coefficients can be constructed *via* Eqns. 8.1 and 8.2 before subtracting them.

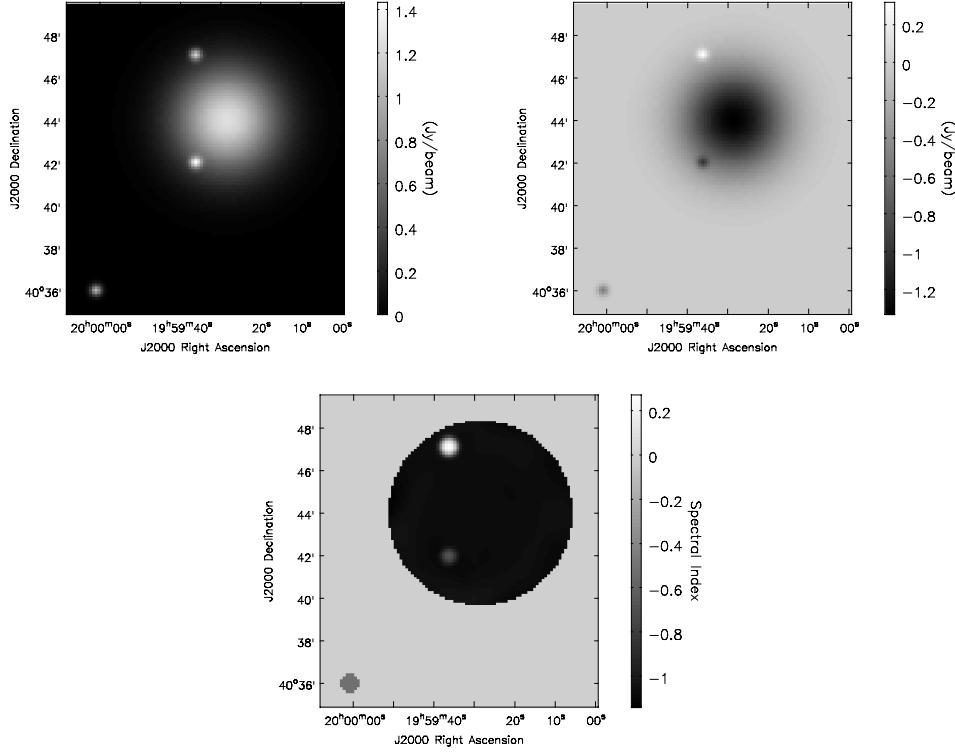


Figure 8.13: Intensity and Spectral Index : These images show the results of applying the MS-MFS algorithm to a simulated dataset in which the sky flux has a pair of foreground point sources on top of an extended background. The top two images show the first two polynomial coefficients (0^{th} -order coefficient or intensity I_0^{total} : top left, 1^{st} -order coefficient I_1^{total} : top right) and the bottom image is the spectral index map computed as the ratio of the coefficient images. The flux and spectral index of the extended source and isolated point source are recovered correctly, but the two point sources located within the extended source have the wrong values. Table 8.3 shows how the flux and spectral index of the two foreground sources can be recovered *via* a polynomial subtraction.

Foreground Source	I_0^{total}	I_1^{total}	α^{total}	I_0^{back}	I_1^{back}	α^{back}	I_0^{front}	I_1^{front}	α^{front}
top	1.172	0.321	+0.27	0.185	-0.196	-1.05	0.987	0.517	+0.52
bottom	1.434	-0.979	-0.68	0.429	-0.466	-1.08	1.005	-0.513	-0.51

Table 8.3: True, measured and corrected intensity and spectra for foreground sources : This table lists the first two polynomial coefficients and the spectral index for the two foreground point sources on the extended background ('top' refers to the topmost source, and 'bottom' refers to the point source in the middle of the image). The true flux values are $I_0=1$ Jy/beam, $\alpha=+0.5$ for the top point source and $I_0=1$ Jy/beam, $\alpha=-0.5$ for the bottom point source. These two sources are on top of a background source with $\alpha = -1.0$. The corrected intensity is given by $I_0^{front} = I_0^{total} - I_0^{back}$, and the corrected spectral index is given by $\alpha^{front} = (I_1^{total} - I_1^{back}) / (I_0^{total} - I_0^{back})$.

8.2.4 Band-limited signals

Objective : The goal of this test is to evaluate how well the MS-MFS algorithm is able to reconstruct the wide-band structure of a source when the emission is detected in only part of the sampled frequency range; in other words a band-limited signal. Since the MS-MFS algorithm uses a polynomial to model the spectrum of the source (and is not restricted to a power-law spectrum) it should be able to reconstruct such structure as long as it varies smoothly. It should however be noted that for a band-limited signal, the angular resolution at which the structure can be mapped will be limited to the resolution of the highest frequency at which the signal is detected (and not the highest resolution allowed by the measurements).

One type of band-limited radiation is synchrotron emission from solar prominences where different frequencies probe different depths in the solar atmosphere. The structures are generally arch-like with lower frequencies sampling the top of the loop and higher frequencies sampling the legs. So far, multi-frequency observations of such sources have been made by a set of simultaneous narrow-band measurements. It may be advantageous to use the combined *uv*-coverage offered by multi-frequency synthesis during imaging, especially since solar prominences are highly time-variable and long synthesis runs to accumulate single-frequency *uv*-coverage are not possible.

EVLA Simulation : Data were simulated for the EVLA D-configuration with 20 channels spread between 1 and 3 GHz (each channel is 100 MHz apart). The wide-band sky was constructed to follow a loop structure as seen from vertically above it. The lower frequencies show the structure of the connected part of the loop and the higher frequencies (that represent deeper layers) show the two legs of the loop. A point source was also added to one of the legs to test the angular resolution to which the reconstruction was possible.

MS-MFS Imaging Results: The MS-MFS algorithm was run on these simulated data, using $N_t = 5$ to fit a 4th-order polynomial to the source spectrum (to accomodate its nearly band-limited nature) and $N_s = 3$ with scales given by [0, 10, 30] pixels. Iterations were terminated after 200 iterations. Figs. 8.14 and 8.15 show a comparison of the true and reconstructed structure at 5 different frequencies between 1 and 3 GHz. These images show that except at the ends of the frequency range where the sky brightness is at its minimum, the reconstruction is quite close to the true sky flux. A second run was performed using only one timestep of data to simulate a snapshot observation. The results were similar between 1.2 and 2.2 GHz but were worse at the ends of the sampled range. Tests with more realistic wide-band sky brightness distributions are required. These results show that it is possible to reconstruct the structure of band-limited structure as long as the flux varies smoothly with frequency and N_t is chosen appropriately.

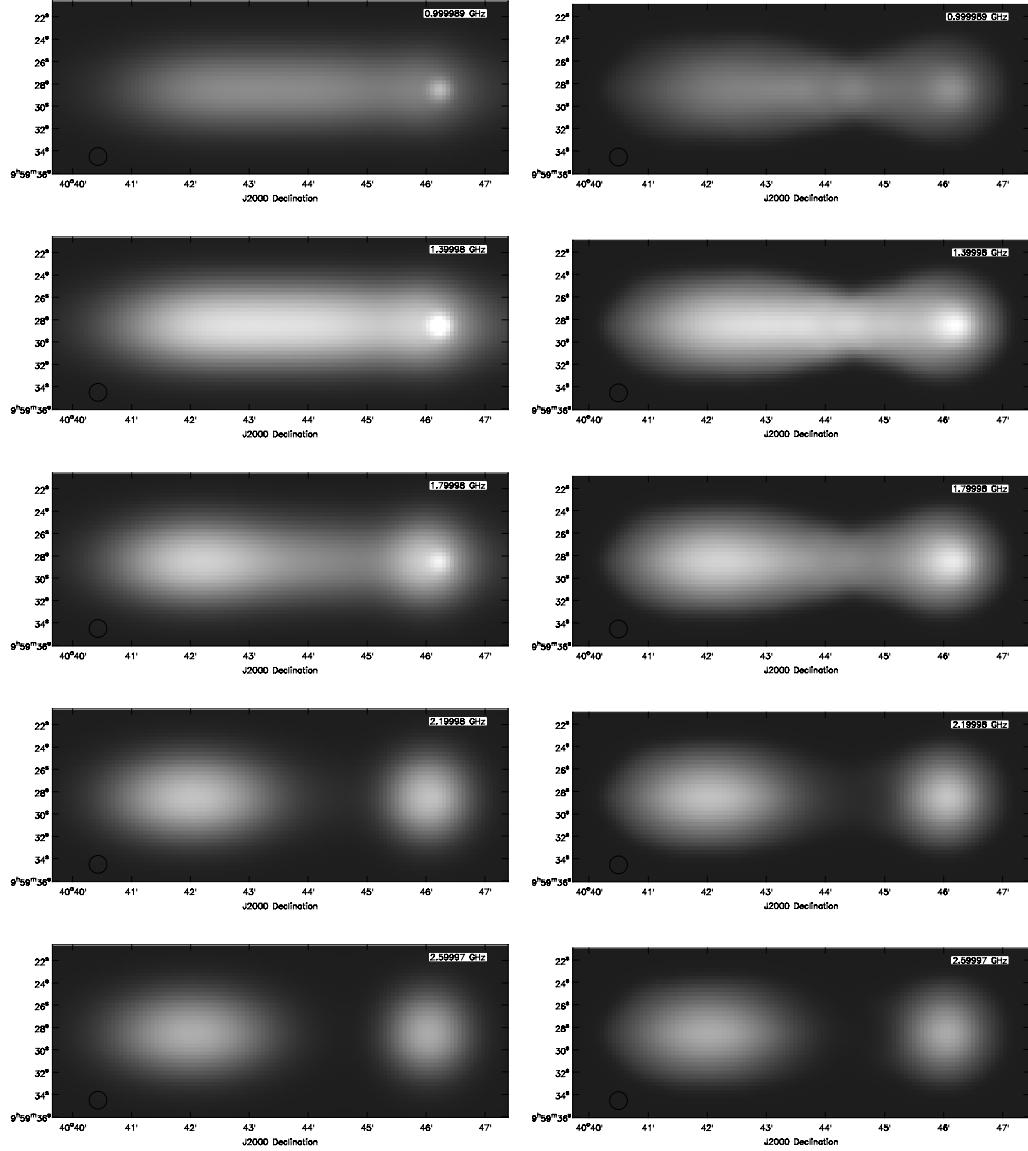


Figure 8.14: Band-limited Signals - Multi-frequency images : These images show a comparison between the true sky brightness (left column) and the brightness reconstructed using the MS-MFS algorithm (right column) at a set of five frequencies (1.0, 1.4, 1.8, 2.2 and 2.6 GHz on rows 1 through 5). All images are at the angular resolution allowed by the highest frequency in the band. This structure represents the arch-like structure of a solar prominence viewed from above, with higher frequencies probing deeper into the solar atmosphere. The images on the right show that most of this structure is recovered with the largest errors being in the central region where the signal spans the shortest bandwidth. Also, the point source on the right was reconstructed at an angular resolution slightly larger than that of the highest sampled frequency and corresponds to the highest frequency at which this spot is brighter than the background emission.

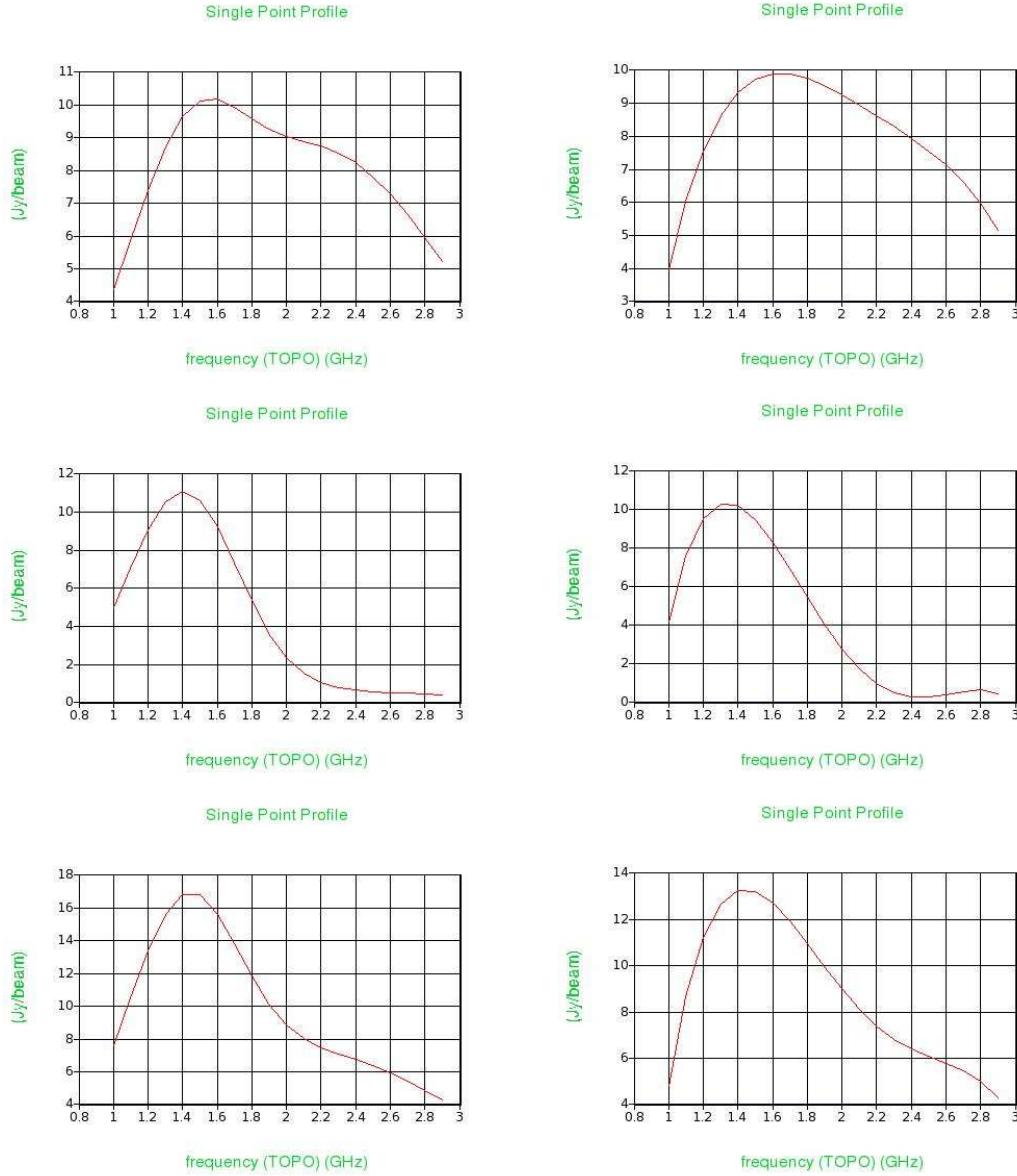


Figure 8.15: Band-Limited Signals - Spectra across the source : These plots show the true (left column) and reconstructed (right column) spectra at different locations for the example discussed in this section (shown in Fig.8.14). The spectra in the top row correspond to the left end of the loop at the location of the leg and shows smooth structure stretching almost all across the band. The spectra in the middle row correspond to the middle of the source where the only structure in the line-of-sight is the upper part of the loop. At this location, there is emission only within a small fraction of the band. The bottom row shows spectra for a point on the right end of the loop at the location of the point source. Here, there is broad-band emission (due to the leg) with relatively narrow-band emission on top of it. From these plots we can see that except for the ends of the frequency range, the reconstruction is close to the true sky brightness.

8.3 Wide-band imaging results with (E)VLA data

This section describes imaging results using wide-band VLA data to test the MS-MFS algorithm along with wide-band calibration techniques. At the time these tests were performed, 11 VLA antennas had been fitted with interim EVLA receivers (1–2 GHz, new wide-band L-band feeds but with VLA polarizers), and the remaining antennas had the old VLA L-band feeds and receivers (1.2 to 1.8 GHz). The VLA correlator had a maximum instantaneous bandwidth of 50 MHz and wide-band data had to be taken as a series of narrow-band snapshot observations that cycled through a set of discrete frequencies spanning the full frequency range allowed by the receivers. Similar snapshot observations of the VLA primary calibrator source 3C286 were interlaced with these frequency cycles in order to derive the flux scale for all the observations. Tables 8.4 and 8.5 list the observation parameters that were used to acquire data for the Cygnus A and M87 fields, and Fig. 8.16 shows an example of the single-frequency and multi-frequency *uv*-coverage that resulted from these observations.

Section 8.3.1 describes how these data were used to test the ability of the MS-MFS algorithm to reconstruct spatial and spectral structure for a complex extended source from a set of incomplete single-frequency measurements. Similar observations were made for M87 to test the algorithm on a source with very extended low signal-to-noise spatial structure and a total angular size extending out to the 75% point of the primary beam (section 8.3.2). The resulting spectral index map was then used to study the broad-band spectra of features across the M87 halo (described in detail in chapter 9). The flux calibrator for both these observations was 3C286, a field containing several bright (50 mJy) background point sources spread out to the 70% point of the primary beam at 1.4 GHz. These calibrator data were used independent of Cygnus A and M87 to test the MS-MFS algorithm with wide-band primary beam correction (section 8.3.3).

Note that all the wide-band data used for the tests in this section came from an interferometer that produced only narrow-band output (< 50MHz). Wide-band data were taken by cycling through frequencies during the observation and there were no simultaneous full-bandwidth measurements. These were the only type of wide-band data available at the time the MS-MFS algorithm was being developed and implemented.

8.3.1 Wide-band imaging of Cygnus A

Objective : Wide-band VLA observations of the bright radio galaxy Cygnus A were used to test the MS-MFS algorithm on real data as well as to test standard calibration methods on wide-band data. Most of the images so far made of Cygnus A and its spectral structure have been from large amounts of multi-configuration narrow-band VLA data [Carilli et al. 1991] designed so as to measure the spatial structure as completely as possible at two widely separated frequencies. The goal of this test was to use multi-frequency snapshot observations of Cygnus A to evaluate how well the MS-MFS algorithm is able to

Telescope	VLA (B configuration)
Observing Band	800 MHz at Lband (1.3 - 2.1 GHz)
Target Source	Cygnus A (19:59:28.3560 +40.44.02.0750)
Calibrator Source	3C286 (13:31:08.314, +30.30.31.156)
Angular resolution	4.1, 3.2, 2.6 arcsec at 1.3, 1.7, 2.1 GHz
Cell size	0.7 arcsec
Image size	1024×1024 pixels (11.9 arcmin)
VLA correlator mode ² (4 IF)	RR/LL (6.25 MHz = 32×0.195 MHz)
Number of spectral windows (SPWs)	9 (out of 18)
Number of channels per SPW	19 (out of 32)
Channel width	0.195 MHz
Instantaneous bandwidth	3.7 MHz (out of 6.25 MHz)
Reference Frequency	1.7 GHz
Total integration time per SPW	30 min
Integration time per visibility	3.0 sec
Total time on source	~ 5 hours
System temperature T_{sys}	~ 250 K for Cygnus A
Noise per visibility	3.0 Jy theoretical
Single-SPW point-source sensitivity	1.1 mJy
Continuum point-source sensitivity	0.3 mJy
Expected dynamic range	240000

Table 8.4: Wide-band VLA observation parameters for Cygnus A : Wide-band data were taken using the VLA by cycling through a set of 9 frequency tunings and taking narrow-band snapshot observations at each tuning. This cycle was repeated 20 times to give a total of about 30 minutes per frequency tuning. Figure 8.16 shows the single and multi-frequency *uv*-coverage for these observations.

simultaneously reconstruct its spatial and spectral structure from measurements in which the single-frequency *uv*-coverage was insufficient to accurately reconstruct all the spatial structure at that frequency.

Cygnus A Cygnus A an extremely bright (1000 Jy) radio galaxy with a pair of bright compact hotspots about 1 arcmin away from each other on either side of a very compact core, and extended radio lobes associated with the hotspots that have broad-band synchrotron emission at multiple spatial scales. From many existing measurements [Carilli and Barthel 1996], this radio source is known to have a spatially varying spectral index ranging from near zero at the core, -0.5 at the bright hotspots and up to -1.0 or more in the radio lobes.

²IF represents intermediate frequency, a label used at the VLA to denote frequency ranges that are sent into the correlator simultaneously. Another label for these frequency ranges is spectral window (SPW).

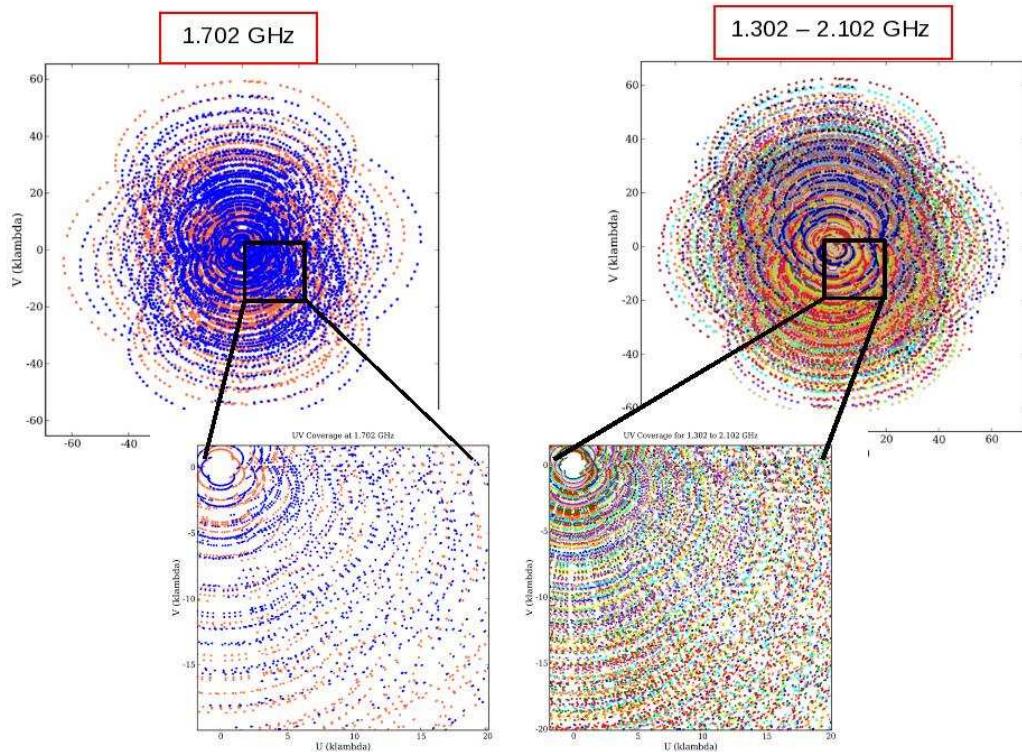


Figure 8.16: VLA multi-frequency uv -coverage : This figure shows the multi-frequency uv -coverage of VLA observations of Cygnus A, taken as a series of narrow-band snapshot observations. The plots on the left show the uv -coverage from one frequency channel (20 snapshots at 1.7 GHz). By zooming into the central region (bottom left) and comparing the spacing between the measurements to the size of the uv grid cells being used for imaging we can show that the single-frequency measurements are incomplete. The plot on the right shows the multi-frequency uv -coverage using nine frequency tunings. A zoom-in of the same central region (bottom right) shows that for the chosen uv grid cell size (or image field of view over which the image is to be reconstructed) the combined sampling leaves no unmeasured grid cells. The imaging results from these observations will test our ability to reconstruct both spatial and spectral information from incomplete spatial frequency samples at a discrete set of frequencies.

Observations : Wide-band data were taken as described in Table 8.4 using the VLA 4-IF mode which allowed four simultaneous data streams containing RR and LL correlations at two independent frequency tunings. A set of 18 frequencies were chosen such that they spanned the entire frequency range allowed by the new EVLA receivers (1–2 GHz). Visibilities that used antennas with the older receivers were flagged for regions of the band not covered by the receivers (below 1.2 GHz and above 1.8 GHz). The uv -coverage for this

dataset for the RR correlations is shown in Fig.8.16. The data were inspected visually and visibilities that were affected by strong radio frequency interference were flagged (masked).

Calibration : Standard techniques were used to calibrate these data. Flux calibration at each frequency was done *via* observations of 3C286. Phase calibration was done using an existing narrow-band image of Cygnus A at 1.4 GHz [[Carilli et al. 1991](#)] as a model.

At the time of these observations, the VLA correlator was getting inputs from a combination of VLA and EVLA antennas. A gain control system that was temporarily put in place to accomodate the use of new EVLA antennas with the VLA correlator treated the two independent frequency tunings in the 4-IF mode differently³. This caused errors in the correlator input for very strong sources (Cygnus A) that increased the input power level beyond the linear power range of the VLA correlator. Observations of the calibrator source 3C286 were not affected by this problem. We were therefore able to calibrate all the frequency tunings for Cygnus A and use the resulting wide-band spectrum along with the known integrated flux and spectral index of Cygnus A to idenitify which of the frequency tunings of Cygnus A were affected. It was found that every alternate frequency (the second of each pair of simultaneous frequency tunings (B/D) in the VLA 4-IF mode) was affected. Therefore to safely eliminate the effect of this problem for our tests, one of the two simultaneous frequency tunings were flagged from the recorded visibilities reducing the number of spectral windows from 18 to 9. The final dataset used for imaging consisted of nine spectral windows each of a width of about 4 MHz and separated by about 100 MHz.

Imaging : These data were imaged using two methods, the MS-MFS algorithm and a hybrid method consisting of STACK + MFS on residuals (see section [5.2.1.4](#) for a description of this method). Their results were compared to evaluate the merits of the MS-MFS algorithm over the much simpler hybrid method that used a combination of existing standard methods. The data products evaluated were the total-intensity image, the continuum residual image and the spectral index map. The effect of the primary beam was ignored in these imaging runs because the angular size of Cygnus A is about 2 arcmin, which at L-band is within a few percent of the HPBW of the primary beam, a region where the antenna primary beam and its spectral effects can be ignored.

³ To allow the use to new EVLA antennas with the old VLA correlator, an automatic gain control had to be used at each EVLA antenna to mimic the old VLA antennas and ensure that the input power levels to the VLA correlator were within the range over which it has a linear response. The type of gain control was being done differently for the two simultaneous frequency tunings in the VLA 4-IF mode. The A/C IF stream used an automatic gain controller based on power levels measured in 1 second and the B/D IF stream used a static look-up table to decide attenuation levels. This resulted in a difference in power levels for the A/C and B/D data streams for all baselines that involved EVLA antennas when the source being observed was bright enough to contribute to increasing the overall system temperature.

1. **MS-MFS :** The MS-MFS algorithm was run with a 2nd-order polynomial to model the source spectrum and a set of 10 scale basis functions of different spatial scales to model the spatial structure ($N_t = 3, N_s = 10$). Iterations were terminated using a 30 mJy stopping threshold. A theoretical continuum point-source sensitivity of 0.38 mJy was calculated for this dataset using an increased system temperature of $T_{sys} = 250$ (due to the high total power of Cygnus A).
2. **Hybrid :** The second approach was a hybrid algorithm in which the MS-CLEAN algorithm was run separately on the data from each spectral window and then a single MS-CLEAN run was performed on the continuum residuals (the STACK + MFS on residuals hybrid algorithm described in 5.2.1.4). The total intensity image was constructed as an average of the single channel image plus the result of the second stage on the continuum residuals. This method is the same as that used in section 5.2.3 to test the hybrid algorithm for the case of dense single-frequency *uv*-coverage. Note however that the observations being described in this section do not have dense single-frequency *uv*-coverage, and the purpose of applying this hybrid method is to emphasize the errors that can occur if this method is used inappropriately.

Results : Figure 8.17 shows the reconstructed total-intensity images (top row) and the residual images (bottom row) obtained from these two methods. Figure 8.18 shows the spectral maps constructed *via* the two methods described above as well as from existing images at 1.4 and 4.8 GHz.

1. **Intensity and Residuals :** Both methods gave a peak brightness of 77 Jy/beam at the hotspots and a peak brightness of about 400 mJy/beam for the fainter extended parts of the halo. The residual images for both methods showed correlated residuals due to the use of a multi-scale flux model composed of a discrete set of scales (small-scale correlated structure within the area covered by the source, but no visible large-scale deconvolution errors due to missing large-scale flux).

The off-source noise level achieved in the continuum image with MS-MFS was about 25 mJy, giving a maximum dynamic range of about 3000. The peak on-source residuals were at the level of 30 mJy. Further iterations did not reduce these residuals, and the use of a higher-order polynomial $N_t > 3$ introduced more errors in the spectral index map (see section 6.2.4.1 for a discussion about errors on the spectral index as a function of N_t and the SNR of the measurements). The off source RMS reached by the hybrid method was about 30 mJy, with the peak residuals in the region of the source of 50 mJy. Deeper imaging in either stage did not reduce these residuals.

Note also that both methods were almost two orders of magnitude above the theoretical point-source sensitivity shown in Table 8.4 (calculated for an equivalent wide-band observation). However, the achieved RMS levels were consistent with the best

RMS levels previously achieved with the VLA at 1.4 GHz for this particular source at L-band (~ 20 mJy, [Perley, R. (private communication)]).

2. **Spectral Index :** The image on the top left is the result of the MS-MFS algorithm and shows spectral structure at multiple scales across the source. For comparison, the image at the bottom is a spectral-index map constructed from existing narrow-band images at 1.4 and 4.8 GHz, each constructed from a combination of VLA A, B, C and D configuration data [Carilli et al. 1991]. These two images (top-left and bottom) show a very similar spatial distribution of spectral structure. This shows that despite having a comparatively small amount of data (20 VLA snapshots at 9 frequencies) the use of an algorithm that models the sky brightness distribution appropriately is able to extract the same information from the data as standard methods applied to large amounts of data. The estimated errors on the spectral index map are < 0.1 for the brighter regions of the source (near the hotspots) and ≥ 0.2 for the fainter parts of the lobes and the core.

The image on the top right shows the spectral index map constructed from a spectral cube (a set of 9 single-channel images) containing the results of running the MS-CLEAN algorithm separately on each frequency and then smoothing the results down to the angular resolution at the lowest frequency in the range. Note that the single-frequency observations consisted of 20 snapshots of Cygnus A. This *uv*-coverage is too sparse to have measured all the spatial structure present in the source, and the non-uniqueness of the single-frequency reconstructions caused the images at the different frequencies to differ from each other enough to adversely affect the spectra derived from these images.

3. **Spectral Curvature :** Note that although Cygnus A itself has more than sufficient signal-to-noise to measure any spectral curvature, very low level deconvolution errors (3 orders of magnitude below the bright 77 Jy/beam hotspot) dominate the region around the very bright hotspots and this is sufficient to destroy the spectral curvature images. That is, the signal-to-error ratio of the higher-order coefficient images is too low to measure a physically plausible curvature term (corresponding to a change in α of < 0.2 across 700 MHz at 1.4 GHz).

Wide-band Self Calibration : A few tests were done to test whether a self-calibration process that used wide-band flux models would yield any improvement on the gain solutions or imaging results.

Two sets of calibration solutions were computed and compared. For the first set of solutions, several rounds of amplitude and phase self-calibration were run, beginning with a point-source model and using the MS-MFS algorithm to iteratively build up a wide-band flux model. Self-calibration was terminated after new gain solutions were indistinguishable from that of the previous run. The second set of solutions was found by using

a single 1.4 GHz model for amplitude and phase self-calibration (with gain amplitudes normalized to unity to preserve the source spectrum). No significant difference was found and the second set of solutions were chosen for imaging.

As an additional test, the final wide-band flux model generated *via* the MS-MFS algorithm was used to predict model visibilities for a wide-band self-calibration step (amplitude and phase) to test if this process yielded any different gain solutions. Again, on these data, there was no noticeable improvement in the continuum residuals or on the stability of the spectral-index solution in low signal-to-noise regions.

This suggests that either the use of a common 1.4 GHz model image for all individual frequencies did not introduce much error, or that the residual errors are dominated by the effects of multi-scale wide-band deconvolution and the flux model assumed by the MS-MFS algorithm. Further tests are required with much simpler sky brightness distributions and real wide-band data, in order to clearly ascertain when wide-band self-calibration will be required for high-dynamic range imaging.

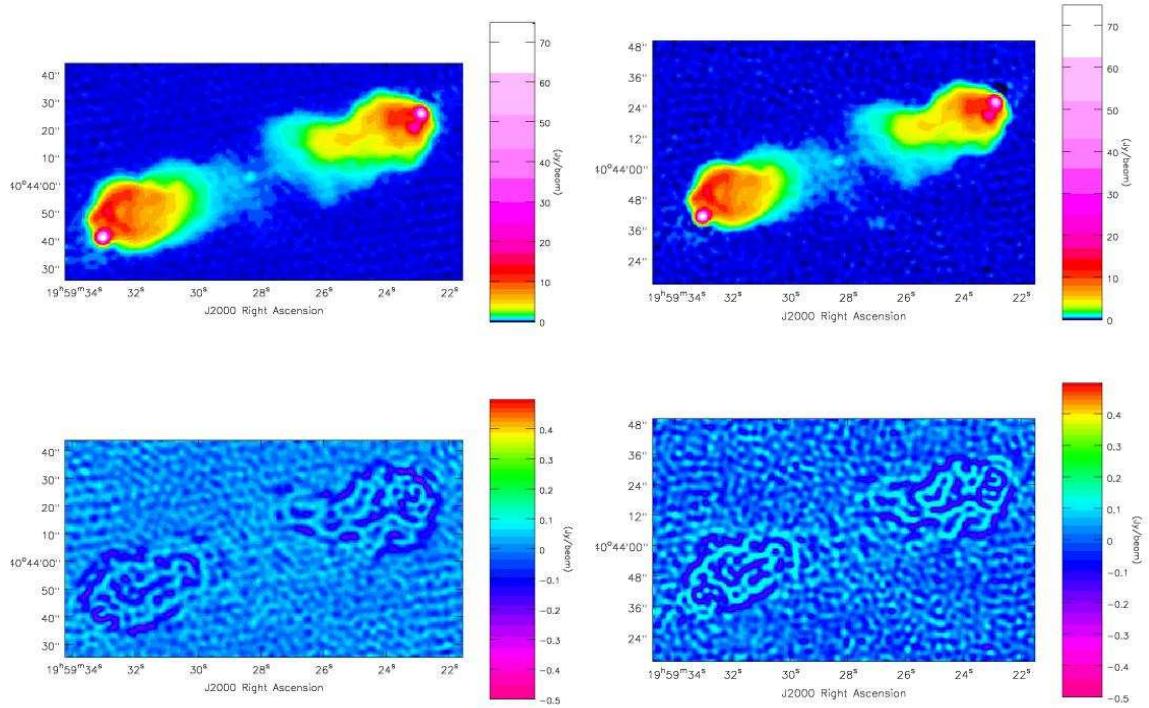


Figure 8.17: Cygnus A : Intensity and residual images : These images show the total intensity (top row) and residual images (bottom row) obtained by applying two wide-band imaging methods to Cygnus A data taken as described in Table 8.4. The images on the left are the result of the MS-MFS algorithm and those on the right are with the STACK + MFS hybrid in which MS-CLEAN was used for all the deconvolutions (single-channel deconvolutions followed by second deconvolution on the continuum residuals. The total intensity images show no significant differences. Both residual images show correlated residuals of the type expected for the MS-CLEAN algorithm that uses a discrete set of scale sizes (the error pattern obtained by choosing a nearby but not exact spatial scale for a flux component will be a ridge running along the edge of each flux component). The peak and off source residuals for the MS-MFS algorithm are 30 mJy and 25 mJy and with the hybrid algorithm are 50 mJy and 30 mJy respectively, showing a very mild improvement in continuum sensitivity with the MS-MFS algorithm.

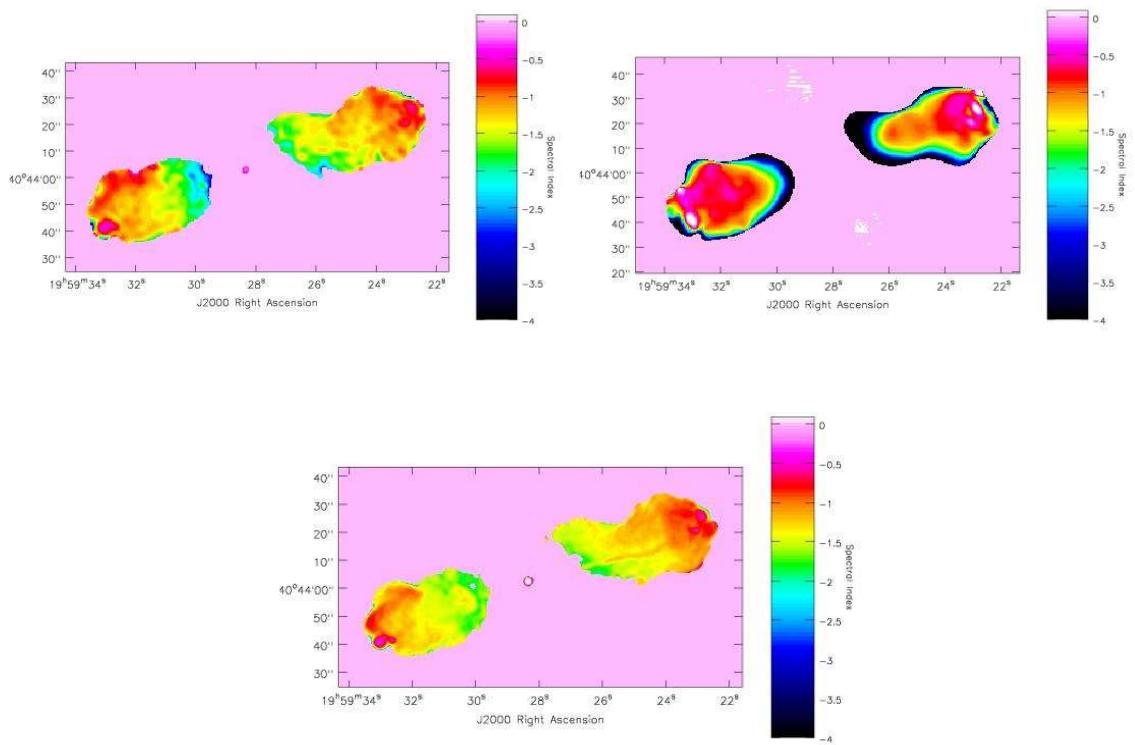


Figure 8.18: Cygnus A : Spectral Index image : These images show spectral index maps of Cygnus A constructed *via* the MS-MFS algorithm (top left) and the hybrid algorithm (top right) applied to the data described in Table 8.4. The image at the bottom is a spectral index map constructed from two narrow-band images at 1.4 and 4.8 GHz obtained from VLA A,B,C and D configuration data at these two frequencies [Carilli et al. 1991]. The spatial structure seen in the MS-MFS spectral index image is very similar to that seen in the bottom image. For comparison, the spectral index map on the top-right clearly shows errors arising due to non-unique solutions at each separate frequency as well as smoothing to the angular resolution at the lowest frequency.

Telescope	VLA (C)
Observing Band	800 MHz at L-band (1.1 – 1.8 GHz)
Target Source	M87 (12:30:49.600 +12.23.19.078)
Calibrator Source	3C286 (13:31:08.314, +30.30.31.156)
Angular resolution (C)	16.5, 12.5, 10.11 arcsec at 1.1, 1.45, 1.8 GHz
Cell size	3.0 arcsec
Image size	1024×1024 pixels (51.2 arcmin)
Correlator mode (2 IF)	RR/LL (12.5 MHz = 16×0.781 MHz)
Number of spectral windows (IFs)	16 (out of 20, due to RFI)
Number of channels per SPW (IF)	10 (out of 16, eliminating end channels)
Channel width	0.781 MHz
Instantaneous bandwidth	7.8 MHz (out of 12.5 MHz)
Reference Frequency	1.45 GHz
Total integration time per SPW	20 min
Integration time per visibility	5.0 s
Total time on source	~ 5.5 hours
System temperature T_{sys}	~ 50 K
Noise per visibility	0.2 Jy theoretical
Single-SPW point-source sensitivity	0.6 mJy
Continuum point-source sensitivity	0.05 mJy
Expected dynamic range	300000

Table 8.5: Wide-band VLA observation parameters for M87: Wide-band observations of M87 were done using the VLA in C and B configurations and cycling through a set of 16 frequency tunings with narrow-band snapshots at each frequency. All frequencies were cycled through 10 times, to generate about 20 minutes of data per frequency tuning. This table shows the parameters for the C-configuration observation. Two similar observations were carried out in the B-configuration and the data later combined.

8.3.2 Wide-band imaging of M87

Objective : Wide-band VLA observations of the M87 cluster-center radio galaxy were taken in order to make a high angular resolution image of the spectral index along various features within its radio halo. The goal of this project was to combine the spectral index information obtained from these data with existing spectral index information below L-band in order to study spectral evolution models for different parts of the M87 halo. This study is presented in detail in chapter 9. Also, this source consists of a bright compact region of emission on top of a relatively faint diffuse background. This structure is useful to test the dynamic range capabilities of the MS-MFS algorithm and the effect of low-level deconvolution errors on the reconstructed spectral index (even when the signal-to-noise ratio on the background emission is sufficient to be able to measure α).

M87 : M87 is a bright (200 Jy) radio galaxy located at the center of the Virgo cluster. The spatial distribution of broad-band synchrotron emission from this source consists of a bright central region (spanning a few arcmin) containing a flat-spectrum core, a jet (with known spectral index of -0.55) and two radio lobes with steeper spectra ($-0.5 > \alpha > -0.8$) [Rottmann et al. 1996a; Owen et al. 2000]. This central region is surrounded by a large diffuse radio halo (7 to 14 arcmin) with many bright narrow filaments ($\approx 10'' \times 3'$). Further, the bright central region is roughly two orders of magnitude brighter than the brightest filaments in the surrounding extended halo.

Observations : Wide-band VLA observations of M87 were carried out in both C and B configurations (10 hours in C and 20 hours in B). The observation parameters for the C-configuration are shown in Table 8.5. The observations consisted of a series of snapshots at 16 different frequencies within the sensitivity range of the EVLA L-band receivers. Note that the minimum spatial frequency required to detect the largest spatial-scale (about 7 arcmin) present in the M87 emission is $0.102k\lambda$. At 1.4 GHz, the minimum spatial frequency measured in the C-configuration is $0.175k\lambda$ and in the B-configuration is $1.05k\lambda$. Therefore, the B-configuration data could measure only the relatively compact emission (bright central region and filaments in the halo) and was included to increase the angular resolution of those measurements. Data affected by radio frequency interference were flagged after visual inspection.

Calibration : Standard calibration techniques were used to calibrate these data. Flux calibration at each frequency was done *via* observations of 3C286 and phase calibration was done using an existing narrow-band image of M87 at 1.4 GHz [Owen, F. (private communication)] as a model. This calibration was done separately for the C and B configuration data which were then combined for imaging.

Imaging :

1. **MS-MFS :** The MS-MFS algorithm was applied to these data to make images of the reference-frequency intensity and the spectral index. The parameters used for this run were $N_t = 3$, $N_s = 11$ with a set of spatial scales given by scale basis functions of widths 0, 3, 9, 12, 16, 20, 25, 30, 60, 80, 140 pixels. Iterations were terminated at a threshold of 10 mJy because the spectral solutions began to get unstable below this threshold (see section 6.2.4.1 for a discussion on how the errors on the spectral index vary with N_t and the SNR of the data). This threshold was an order of magnitude above the theoretical point-source sensitivity.
2. **Primary-beam correction :** The $7' \times 14'$ radio halo extends out to the 85% level of the EVLA primary beam at 1.4 GHz where the intensity is attenuated by 15%.

The effective spectral index at this angular distance from the pointing center is about -0.3 , but for most of the halo and regions of bright filaments this spectral index is < -0.05 . Primary beam correction was done *via* a post-deconvolution image-domain correction by dividing the intensity image by an image of the main lobe of the primary beam at the reference frequency and subtracting the image of the primary-beam spectral index from the uncorrected M87 spectral index map (step 6 on page 151 in section 7.2.2.3 describes this image-domain correction).

3. **Single spectral-window imaging :** To verify the MS-MFS reconstruction of the wide-band spectrum, data from the 16 individual spectral windows were also imaged independently and the spectrum of the integrated flux within the bright compact central region was compared between the two methods. This comparison was possible only for the bright compact central region for which the single-frequency snapshot *uv*-coverage sufficed.

Results : Fig.8.19 shows the resulting intensity (top left) and spectral index maps for M87 at an angular resolution of 12 arcsec (C-configuration). Fig.8.20 shows the on-source and off-source residuals. Fig.8.21 shows the intensity, spectral index and spectral curvature maps of the bright central region at an angular resolution of 3 arcsec (C+B-configuration). Fig.8.22 shows a plot of the spectrum formed from the integrated flux in the central bright region.

1. Intensity and Residuals :

The peak brightness at the center of the final restored intensity image was 15 Jy with an off-source RMS of 1.8 mJy and an on-source RMS of about between 3 and 10 mJy. The residual images show low-level correlated residuals at the location of the source but deconvolution errors are almost absent from the rest of the image, indicating that the best off-source RMS noise level for these data has almost been reached. The maximum dynamic range (ratio of peak brightness to off-source RMS) is about 8000, with the on-source dynamic range (ratio of peak brightness to on-source RMS) of about 1000. The peak brightness in the bright filaments is about 50 to 70 mJy (on-source SNR of about 10), and the peak brightness in the faint diffuse halo is 10 to 20 mJy (on-source SNR of a few).

2. **Spectral Index :** The spectral index map⁴ of the bright central region (at 3 arcsec resolution) shows a near flat-spectrum core with $\alpha_{LL} = -0.25$, a jet with $\alpha_{LL} = -0.5$

⁴The spectral index between two frequency bands *A* and *B* will be denoted as α_{AB} . For example, the symbol α_{PL} corresponds to the frequency range between P-band (327 MHz) and L-band (1.4 GHz), and α_{LL} corresponds to two frequencies within L-band (here, 1.1 and 1.8 GHz). A similar convention will be used for spectral curvature β .

and lobes with $-0.6 > \alpha_{LL} > -0.7$. The bright halo filaments show a steeper spectral index of $\alpha_{LL} \approx -0.8 \pm 0.1$ and the diffuse halo emission shows $\alpha_{LL} \approx -1.1 \pm 0.1$.

The signal-to-noise ratios at various parts of the source can be used to compute error bars on the spectral index and curvature (presented in Table 8.6). These numbers show that in the bright central region and in the halo there is sufficient signal-to-noise to measure the spectral index but any realistic spectral curvature (for broad-band synchrotron emission) is detectable only within the central bright region. Further, the region immediately surrounding the central region is affected by very low level deconvolution errors that are much stronger than the on-source residuals. The effective signal-to-error ratio in this region is about 5.0 which corresponds to an error of >0.3 on a spectral index of -1.0 . The errors on the spectral index map are a very strong function of deconvolution errors (as can be seen from artifacts around the bright central region) which as demonstrated by this example is a significant problem for high-dynamic-range imaging of extended emission.

3. Spectral Curvature :

This bright central region had sufficient (>100) signal-to-noise to be able to detect spectral curvature. The third panel in Fig.8.21 shows the spectral curvature measured within this region. Note that the error bars on the spectral curvature are at the same level as the measurement itself. Therefore, a reliable estimate can only be obtained as an average over this entire bright region. The average curvature is measured to be $\beta_{LL} = -0.5$ which corresponds to a change in α across L-band by $\Delta\alpha = \beta_{v_0}^{\Delta\nu} \approx -0.2$.

These numbers were compared with two-point spectral indices computed between 327 MHz (P-band), 1.4 GHz (L-band), and 4.8 GHz (C-band) from existing images [Owen et al. 2000],[Owen, F. (private communication)]. Across the bright central region, $-0.36 > \alpha_{PL} > -0.45$ and $-0.5 > \alpha_{LC} > -0.7$. The measured values ($-0.5 > \alpha_{LL} > -0.7$ and $\Delta\alpha \approx 0.2$) are consistent with these independent calculations.

4. Comparison with single-frequency maps :

The points in Fig.8.22 shows the integrated flux over the central bright region of M87 (shown in $\log(I)$ vs $\log(\nu/\nu_0)$ space) from the 16 single-spectral-window images. The curved line passing through these points is the average spectrum that the MS-MFS algorithm automatically fit for this region. It corresponds to $\alpha \approx -0.52$ and $\Delta\alpha \approx 0.2$ across the source. The straight dashed lines correspond to constant spectral indices of -0.42 and -0.62 and show that the change in α across the band is approximately 0.2 (as also calculated from $\beta_{LL} = -0.5$ that the MS-MFS algorithm produced). Note that the scatter seen on the points in the plot is at the 1% level of the values of the points (signal-to-noise of 100). Also evident from the plot is the fact that the curvature signal is at a signal-to-noise ratio of 1. These results show that a signal-to-noise of > 100 is required to measure a change in spectral index of 0.2 across 700 MHz at 1.4 GHz.

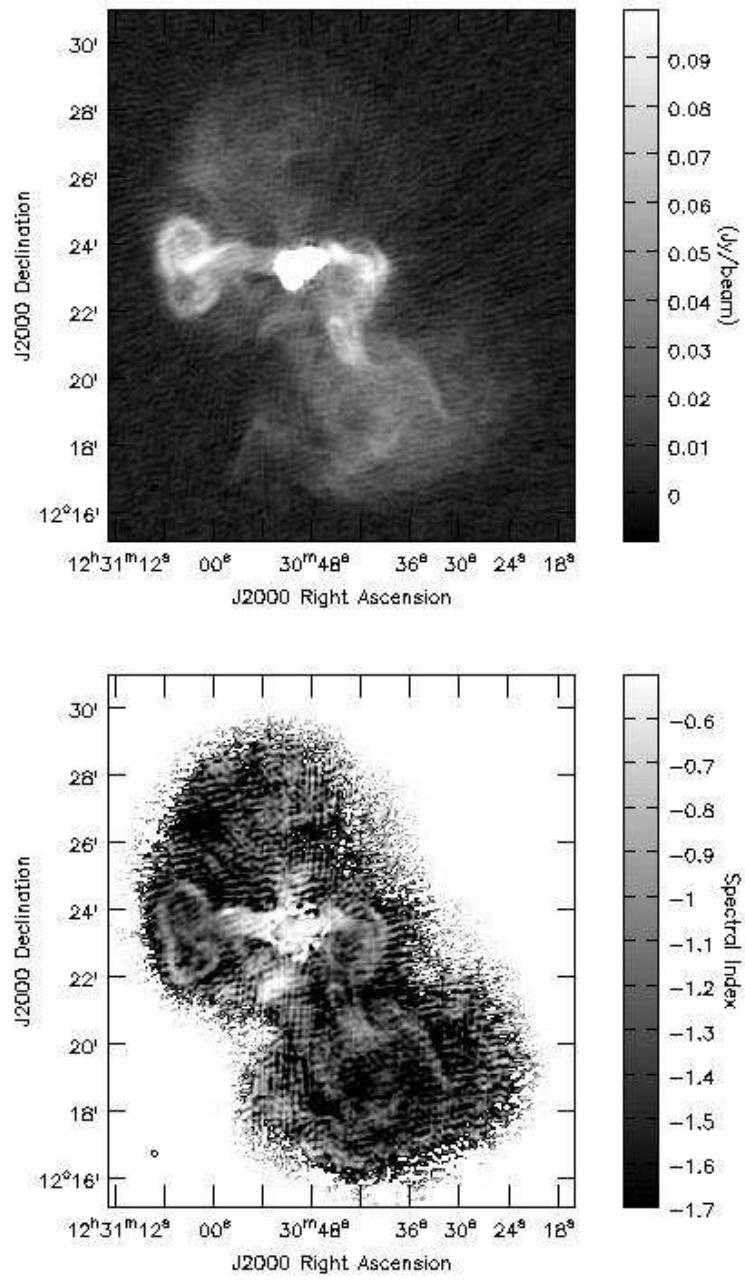


Figure 8.19: M87 halo : Intensity and Spectral Index : These images show the results of applying the MS-MFS algorithm to wide-band VLA data taken as described in Table 8.5. The images are at 12 arcsec resolution, and show the intensity distribution for M87 at 1.5 GHz (top), and the corresponding spectral index (bottom) and Figure 8.21 shows the bright central region at a higher angular resolution and Table 8.6 lists flux values, spectral indices and error-bars for different parts of the source.

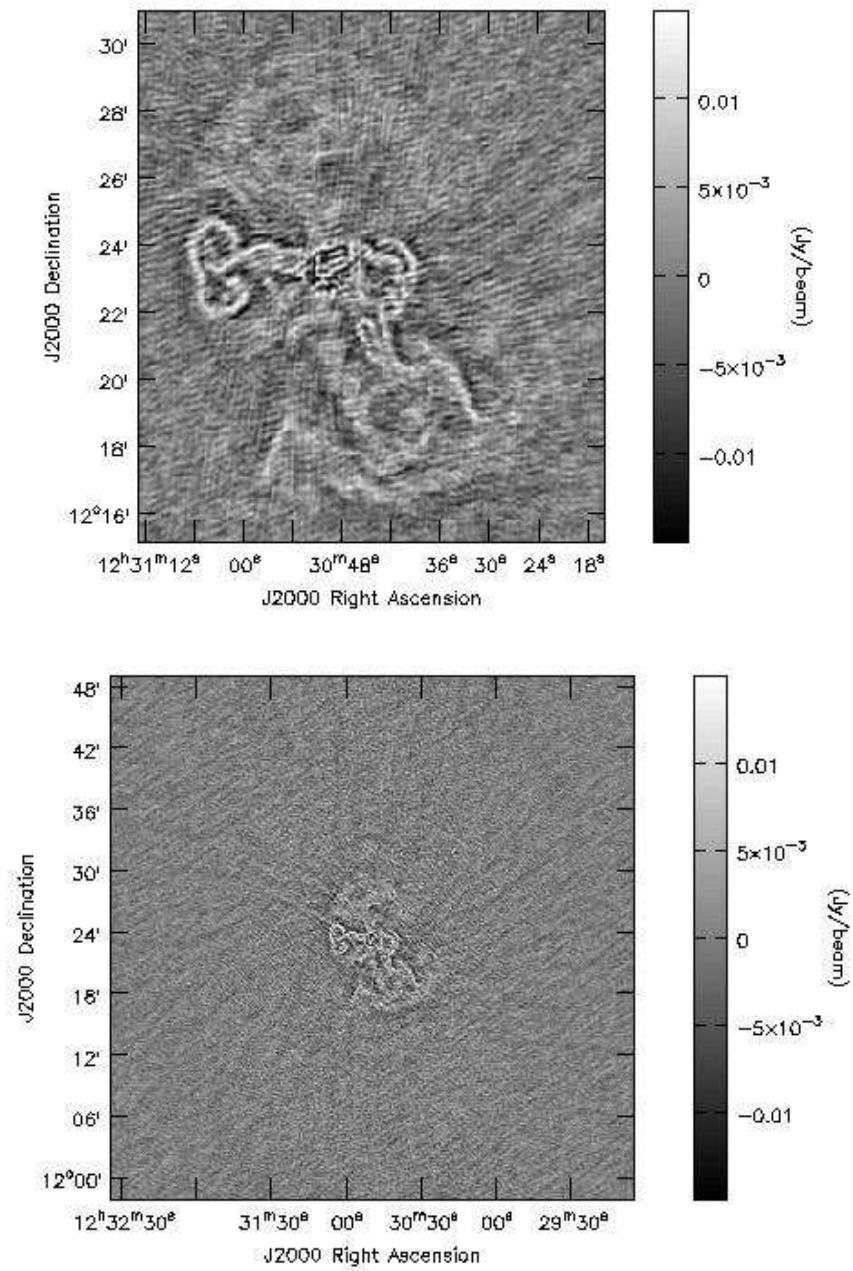


Figure 8.20: M87 halo : Residual Images : These images show the residual image at two different fields-of-view. On-source residuals are shown on the top and off-source residuals at the bottom. These residuals are displayed using a flux scale 10 times smaller than that used in the intensity image in Fig. 8.19. The peak on-source residual is at the level of 10 mJy, but the off-source residuals show no clearly visible trace of large-scale deconvolution errors.

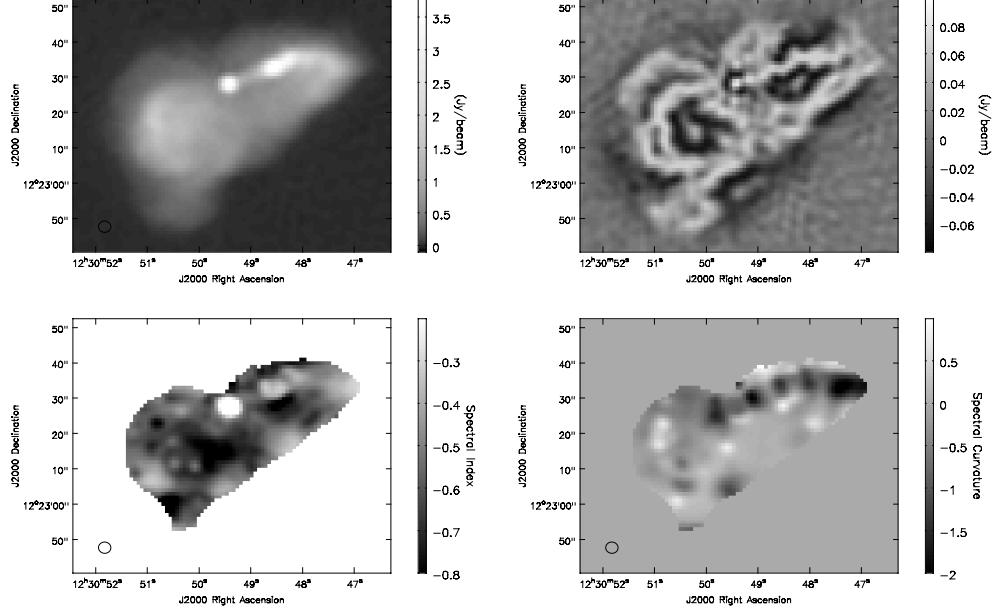


Figure 8.21: M87 core/jet/lobe : Intensity, Spectral index, Curvature : These images show 3-arcsec resolution maps of the central bright region of M87 (core+jet and inner lobes), where the signal-to-noise was sufficient for the MS-MFS algorithm to detect spectral curvature. The quantities displayed are the intensity at 1.5 GHz (top left), the residual image (top right), the spectral index (bottom left) and the spectral curvature (bottom right). The spectral index is near zero at the core, varies between -0.36 and -0.6 along the jet and out into the lobes. The spectral curvature is on average 0.5 which translates to $\Delta\alpha = 0.2$ across L-band. The peak of the source is 4.6 Jy, the on-source RMS is 40 mJy/beam and this gives an on-source signal-to-error ratio of about 100 . Note that the flux scale on the residual image (top right) is about 2 orders of magnitude lower than the total-intensity image (top left).

	core	jet	lobes	filaments	diffuse halo
I_0 (Jy/beam)	4.5	4.6	1.7	0.09	0.03
Residual I_{on}^{res}	0.04	0.04	0.04	0.015	0.01
$SNR=I_0/I_{on}^{res}$	112	115	42	6	3
$\alpha \pm \delta\alpha$	0.005 ± 0.05	-0.36 ± 0.02	-0.63 ± 0.06	-0.95 ± 0.1	-1.5 ± 0.3
$\beta \pm \delta\beta$	-0.8 ± 0.3	-0.9 ± 0.7	-0.2 ± 0.2	—	—

Table 8.6: Measured errors for I_{v_0} , α and β in M87 : This table shows the signal-to-noise ratio for different features of M87, and the observed values for α and β for those features. The fluxes are in units of Jy/beam and the errors $\delta\alpha$ and $\delta\beta$ are estimates based on the measured variations across different pixels within each feature.

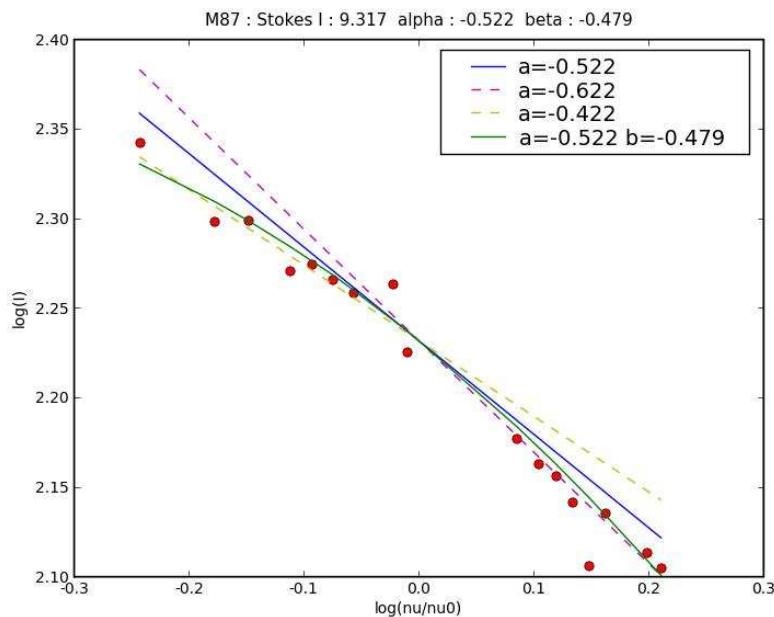


Figure 8.22: M87 core/jet/lobe : L-band spectrum : This plot shows the spectrum formed from the integrated flux within the central bright region between 1.1 and 1.8 GHz. The points are the integrated flux measured from single-spectral-window model images, the curved line is the average spectrum that the MS-MFS algorithm automatically fit to these data in this region. This spectrum corresponds to an average $\alpha_{LL} = -0.52$ and a change of $\Delta\alpha \approx 0.2$ across the band (1.1 to 1.8 GHz). The straight dashed lines represent pure power-law spectra with indices -0.42 and -0.62 and are another way of showing that the change in α across the band is about 0.2. These numbers are consistent with two-point spectral indices computed between 327 MHz (P-band), 1.4 GHz (L-band), and 4.8 GHz (C-band) ($-0.36 > \alpha_{PL} > -0.45$ and $-0.5 > \alpha_{LC} > -0.7$) from existing images [Owen et al. 2000],[Owen, F. (private communication)].

8.3.3 Wide-field wide band imaging of the 3C286 field

Objective : The goal of this observation is to verify the accuracy of wide-band primary beam correction in combination with MS-MFS using a simple field of widely separated point sources. The corrected spectral indices of sources away from the pointing center are then verified by direct measurements (by pointing directly at one of these background sources). The accuracy of the primary-beam model being used in the correction is also verified *via* measurements of the primary beam at multiple frequencies.

3C286 : The 3C286 field consists of a bright 14 Jy compact synchrotron radio source surrounded by an almost perfect grid of about six compact background objects ranging in brightness from 20 mJy to 300 mJy. These background sources are located about 8 to 12 arcmin away from 3C286. The EVLA antenna primary beam at L-band (1.4 GHz) is 28 arcmin across and these background sources are roughly at the 60% to 70% level of the primary beam where the spectral index due to the primary beam is between -0.5 and -0.7 .

Observations : Both the observations described in the previous sections (Cygnus A and M87) used 3C286 as a flux calibrator so no new observations were required to obtain wide-band data for this field. To verify the corrected spectral indices of the background sources, two additional test observations were done. The first was a set of holography⁵ runs at two frequencies (1.185 and 1.285 GHz) from which the amplitude of the antenna primary beam was measured and a two-point spectral index computed as a function of angular distance. At the half-power point, the measured spectral index was about -1.4 , which matches the values obtained from the theoretical models used in the imaging algorithms. The second test was to make a direct measurement of the spectral index of one of the background sources 8 arcmin away from 3C286 by pointing directly at it and eliminating any spectral effects due to the primary beam. This observation also places 3C286 at a distance of 8 arcmin from the pointing center, giving another independent pair of measurements of source spectral index (one direct and one indirect) to test the accuracy of the indirect measurement.

Calibration : Since 3C286 was the calibrator chosen for observations of Cygnus A and M87, gain solutions were found by using an *a priori* model for its spectrum, a pure power law with spectral index of -0.476 [[Perley and Taylor 2003](#)] across L-band. The data with a background source at the pointing center were calibrated using scans taken during the same observation run with 3C286 at the pointing center.

⁵One meaning of the term ‘holography’ is the process of measuring the primary beam and the aperture illumination pattern of a reflecting dish and antenna system. Holography observations were used for this test to measure the actual primary beam and its frequency dependence in order to compare them with the model primary beams that are used in the image reconstruction process. The purpose of this test was to ensure that the true instrumental primary beam and the models used in the image reconstruction software to correct their effect are nearly identical to each other.

MS-MFS Imaging : The 3C286 calibrator data (taken in the VLA C-configuration during observations of M87) were imaged using the MS-MFS algorithm with $N_t = 3$ and $N_s = 1$, first without any primary-beam correction and then with wide-band primary-beam correction taking into account the time-variability and beam squint.

The later test observations were taken when the VLA was in the B-configuration. At this higher angular resolution, 3C286 is slightly resolved, and at 8 arcmin away from the phase center the effect of the w -term becomes significant enough for its effects to be visible in the image. At the time of these observations, the MS-MFS algorithm could work either with primary-beam correction or with multi-frequency w -projection (section 4.2.2.4 and Cornwell et al. [2008]), but not both together⁶. Therefore, these data were imaged in two runs and the results compared. The first run used primary-beam correction methods that use a combination of visibility-domain and image-domain operations to derive corrected intensities and spectral indices (section 4.3.2 describes the algorithm used here). The second run used only w -projection in the visibility domain and implemented primary-beam correction as a post-deconvolution image-domain correction (section 4.2.1). The corrected spectral indices obtained by these two methods were then compared to the values measured by direct measurement (with the source at the pointing center).

Imaging Results : Figure 8.23 shows the imaging results (intensity and spectral index) for the C-configuration data and Fig.8.24 shows the intensity images for the test observation taken in the B-configuration. Fig.8.25 shows 3C286 imaged without and with multi-frequency w -projection⁷.

1. **Intensity and Residuals :** The peak fluxes measured from the intensity image from the C-configuration data were verified with flux values from the corresponding field within the NVSS catalog [Condon et al. 1998]. The peak of 3C286 was 14 Jy/beam, and the background sources range between 20 mJy/beam and 400 mJy/beam. The off-source RMS was measured as 0.5 mJy, close to the theoretical point-source continuum sensitivity for the calibrator data.
2. **Spectral Index of the sky :** The spectral index of 3C286 (at the pointing center) was measured as -0.476 (the spectral index for which the data were calibrated). When the primary beam was ignored, the background sources show spectral indices ranging between -1.1 and -1.4 . With primary beam correction, they reduce to roughly -0.5

⁶Note that the algorithm described in section 4.3 to correct for direction-dependent effects can include a combination of direction-dependent effects and is not restricted to correcting only one of them at a time. However, the software implementation of the primary-beam correction algorithm in CASA at the time these data were analysed did not include the w -term and therefore it had to be done separately.

⁷Multi-frequency w -projection refers to the use of w -projection during multi-frequency synthesis imaging (*i.e.* the gridding convolution functions are different for each frequency because the value of w changes across frequency).

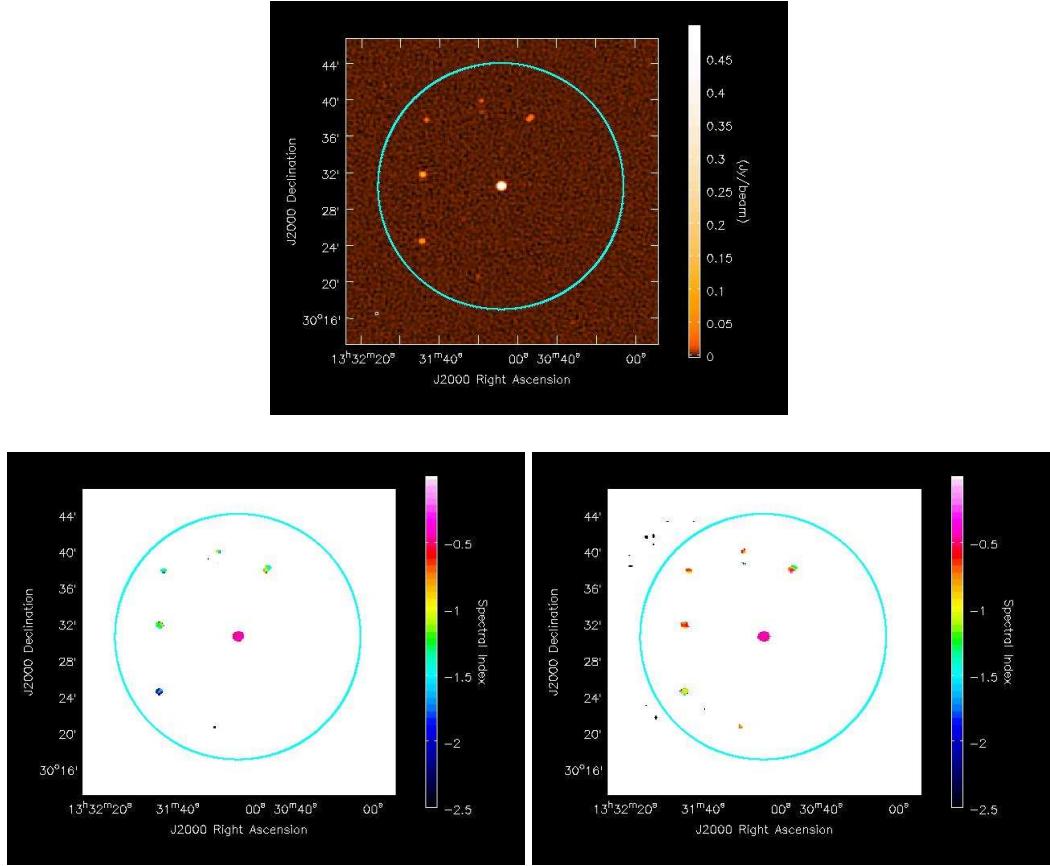


Figure 8.23: MFS with wide-band primary-beam correction : 3C286 field (C-configuration) : These images show the results of applying MS-MFS with primary-beam correction on the C-configuration calibrator data (3C286 field) taken during observations of M87. Shown here are the intensity map (top) and two spectral index maps; one without any primary beam correction (bottom left) and with wide-band primary-beam correction (bottom right). The large circle represents the FWHM of the reference primary beam (1.5 GHz). In the un-corrected spectral index map, the off-center sources show spectral indices between -1.1 and -1.4 which become -0.5 to -0.7 in the corrected map.

to -0.7 . The measured and corrected spectral indices of 3C286 and one of the background sources (due East of 3C286) are shown in Table 8.7. These numbers show that for a field of isolated point sources, it is possible to correct for the frequency dependence of the primary beam to an accuracy of < 0.1 at least within the FWHM at the reference frequency.

3. **Spectral index of the primary beam :** A pair of 1-D primary beam profiles were obtained from a holography scan that measured the beam in 11 directions within the main lobe. The measured beams and two-point spectral indices computed from them

Source Location	3C286 Center	3C286 West of Center	Background East of Center	Background Center
Peak brightness I_0	14 Jy	14 Jy	200 mJy	200 mJy
Off-source RMS I^{res}	1 mJy	10 mJy	1 mJy	10 mJy
SNR = I_0/I^{res}	14000	1400	200	20
α_{MFS+PB}	-0.476	—	-0.602	—
α_{MFS+WP}	-0.476	-0.994	-0.976	-0.577
$\alpha_{MFS+WP+PB}$	-0.476	-0.442	-0.475	-0.577

Table 8.7: Spectral Index of 3C286 field with and without primary-beam correction : This table shows the spectral index of 3C286 and one background source measured directly as well as with primary-beam correction and w -projection. The first and third columns represent the observation in which 3C286 was at the pointing center (all calibrator observations for M87 in the B-configuration). The second and fourth columns represent the short test observation (and hence high RMS) in which the background source due East of 3C286 was placed at the pointing center (w -projection was required for this imaging run to eliminate errors around 3C286). These numbers show the difference between the values of α measured directly with the source at the pointing center and indirectly via an explicit primary-beam correction. For 3C286 (first two columns), this difference is 0.034. For the background source (last two columns) this difference is about 0.1. These numbers suggest that with a SNR of at least 20, and a field of isolated point sources, it is possible to remove the effect of the primary beam on the sky spectral index to an accuracy of equal to or better than 0.1 on α (within the FWHM at the reference frequency).

match those obtained from the theoretical model used in the imaging algorithms. For the locations of interest in this test, the primary-beam profiles from the holography data showed a spectral index of ≈ -0.6 (at the 70% point of the beam).

4. **Multi-frequency w -projection :** The images of 3C286 produced from VLA B-configuration data in which the phase center is 8 arcmin away from the source show expected differences when MS-MFS is used without and with multi-frequency w -projection. The peak off-source residuals reduce from 260 mJy to 110 mJy with the use of w -projection. Note that multi-frequency w -projection is automatically accomplished by the regular w -projection algorithm that chooses the gridding convolution kernel based on the value of w for each baseline and frequency channel.

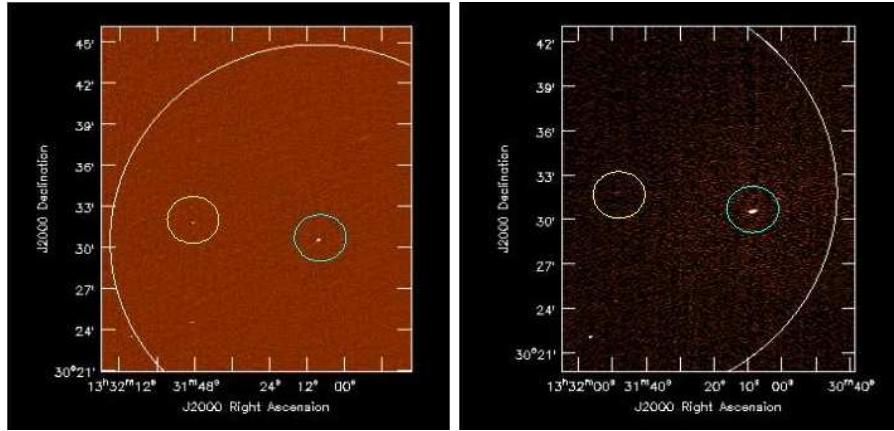


Figure 8.24: MFS with PB correction : 3C286 field (B-configuration) : These images show the intensity maps for the test observation of the 3C286 field (VLA B-configuration). The circle represents the HPBW of the reference-frequency primary beam. The image on the left shows 3C286 at the pointing center. It was made using all the calibrator data from the B-configuration observations of M87 and the RMS achieved was 1 mJy. The image on the right shows one background source at the pointing center and 3C286 located 8 arcmin away. It was made using test observation data (at five frequencies across L-band) and reached an RMS of 10 mJy. The spectral indices measured for these sources are listed in Table 8.7.

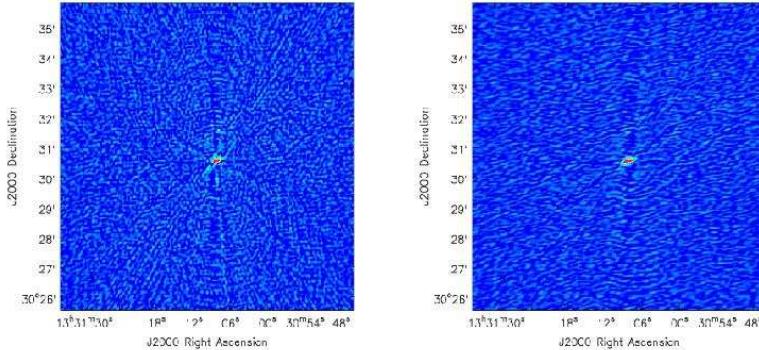


Figure 8.25: MFS with w -projection : 3C286 field (B-configuration) : These images show the region around 3C286 made from VLA B-configuration data in which 3C286 was located 8 arcmin away from the phase and pointing center. The image on the left is the result of MFS without w -projection and the peak off-source residual is 260 mJy. The image on the right is with multi-frequency w -projection and has a peak off-source residual of 110 mJy. The off-source RMS (away from 3C286) for both runs was about 10 mJy. No primary-beam correction was done in these runs and the measured spectral indices included the frequency dependence of the primary beam. The time-variability of the primary beam (due to rotation and squint) was not accounted for, and might explain the high peak residual compared to the off-source RMS.

8.4 Points to remember while doing wide-band imaging

This section briefly summarizes several practical aspects of wide-band imaging. The goals of this section are (a) to list out the key points that are needed to make effective use of the MS-MFS wide-band imaging algorithm, (b) to understand sources of error as well as the implications of various choices of parameters for a given type of broad-band sky brightness distribution, and (c) to recognize when the use of such methods will provide a significant advantage over much simpler single-channel methods and when they will not. Note that all the wide-band data used for the tests in this dissertation were either simulated, or came from interferometers with narrow-band receivers with which wide-band data were taken by cycling through frequencies (*i.e.* no simultaneous full-bandwidth measurements). Therefore, in addition to the current results, several tests with real wide-band data will be required in order to establish a robust data analysis path for wide-band imaging.

Section 8.4.1 discusses the MS-MFS algorithm and explains the meaning of four main parameters that control it. Section 8.4.2 discusses dynamic-range limits when various spectral effects are ignored, lists various sources of error that affect the accuracy of the spectral reconstructions, and summarizes the ability of the MS-MFS algorithm to reconstruct an accurate wide-band model of the sky brightness distribution when additional information about the source is required. Section 8.4.3 compares single-channel methods of wide-band imaging with those that use multi-frequency synthesis, and discusses the image fidelity, dynamic range and computational complexity associated with both types of methods. Section 8.4.4 lists topics for related future work (additional tests and algorithmic improvements).

8.4.1 Using the MS-MFS algorithm

Algorithm : The MS-MFS algorithm models the spatial sky brightness distribution as a sum of 2D Gaussian-like functions (with equal major and minor axes). The spectrum is modeled by allowing the amplitude of each flux component to follow an N^{th} -order polynomial in frequency. Extended emission with spectral structure that varies across the source is modeled by the sum of multiple flux components with different spectra. The MS-MFS algorithm combines multi-scale deconvolution with multi-frequency-synthesis and performs a linear least-squares optimization to solve for the polynomial coefficients for each chosen flux component.

Data Products : The basic products of the MS-MFS algorithm are a set of $N + 1$ multi-scale coefficient images that describe the spectrum of the sky brightness at each pixel (coefficients of an N^{th} -order polynomial). The 0^{th} -order coefficient image is the Stokes I intensity image at the reference frequency (not the continuum image defined as the integrated flux across the full sampled bandwidth). To create the continuum image, the polynomial has to be evaluated and summed over all frequency channels. Derived quantities such as the spectral index and spectral curvature are computed from the coefficient images (see

Eqns. 6.43 to 6.45).

User-controlled parameters : There are four main parameters that control the operation of the MS-MFS algorithm.

1. **Reference Frequency ν_0** : A reference frequency is chosen near the middle of the sampled frequency range. It is the frequency about which a Taylor-expansion of the power-law spectrum is done while forming the polynomial coefficients.
2. **Number of polynomial coefficients N_t** : The user must specify the appropriate number of polynomial coefficients to use to describe the source spectrum. In general, we are using a truncated Taylor series to model a power-law and the number of terms to use will depend on the (expected) spectral index of the sky brightness (note that the only power-laws that can be exactly fitted with a finite Taylor series are those whose indices are positive integers).
 - (a) If the source spectrum can be represented by a straight line in I vs ν space, or if MS-MFS is being done using only two sets of narrow-band data, choose $N_t = 2$. In this case the only data products are maps of the reference-frequency intensity and the spectral index⁸.
 - (b) Sources with negative spectral indices of about -0.5 across a 2:1 bandwidth will require $N_t = 3$. This is an empirically derived estimate based on the imaging runs described in this chapter and section 6.2.4.2. Note that although images of intensity, spectral index and curvature can be computed from the first three coefficient images, it is often necessary to use $N_t > 3$ for spectral indices stronger than -1.0 in order to fit the spectrum better and hence improve the accuracy of the estimates of the first three coefficients. Some prior knowledge of the source spectrum and the signal-to-noise ratio of the measurements is required in order to make an appropriate choice of N_t .
 - (c) For extended emission, deconvolution errors will contribute to the error in the spectral index and curvature maps. This is because it cannot be guaranteed that deconvolution artifacts will preserve the ratios between coefficient images. $N_t = 3$ to $N_t = 5$ have given the best results so far for the types of observations and simulations described in this chapter.
 - (d) The signal-to-noise ratio of the data should also be taken into account to avoid trying to fit a high-order polynomial to a very noisy spectrum. Section 6.2.4.1 gives empirically derived suggestions for N_t for different signal-to-noise ratios.

⁸Note that a straight line in I vs ν space does not represent a power-law. However, since the spectral index of a power law can be obtained from the first two coefficients of the Taylor expansion of a power-law (see Eqn. 6.22), a straight-line fit to the spectrum in I vs ν space can be used to estimate the spectral index of the power law.

3. **Set of spatial scales N_s** : The user must specify a set of scales sizes (in units of pixels) to use for the multi-scale representation of the image [Cornwell 2008]. For a field of isolated point sources a scale vector of [0] ($N_s = 1$) will run a point-source version of MS-MFS. For an extended source with structure on multiple spatial scales, this scale vector must be chosen such that the most obvious scale sizes present in the image are represented. This choice is therefore highly dependent on the structure in the image itself. If the source structure is partially known (from previous imaging runs) then the vector of scale sizes can be chosen by counting pixels across various features in the image (e.g. [0, 6, 20] for an imaging run in which extended features in the image are roughly 6 and 20 pixels across). Overall, there is no well-established method of choosing an appropriate set of spatial scales.
4. **Stopping threshold** : A user-specified flux threshold is used on the 0^{th} -order residual image to control when iterations are terminated. For fields with isolated point sources, this threshold can be chosen to be comparable to the theoretical continuum noise level. However, for complex extended emission, a very deep deconvolution can increase the on-source errors in the higher-order coefficient images (by adding flux that is not well-constrained by the data and is therefore incoherent across the different coefficient images). These errors then propagate non-linearly into the spectral index maps. Therefore, for complex extended sources, it is recommended that the iterations be terminated once off-source residuals become noise-like, irrespective of there being on-source residuals at or slightly above the off-source noise level⁹.

Wide-band self-calibration : The broad-band flux model generated by the MS-MFS algorithm can be used within a self-calibration loop in exactly the same manner as standard self-calibration. The purpose of such a self-calibration would be to improve the accuracy of the calibration.

Software Implementation : The MS-MFS algorithm described in section 7.1 has been implemented and released *via* the CASA¹⁰ software package (version 2.4 onwards). Wide-band primary-beam correction (section 7.2) has been implemented and tested within the CASA system, but is yet to be formally released. These algorithms were implemented in C++ within the existing major/minor cycle code framework of CASAPY and can be accessed *via* the `clean` task and the `imager` tool. The minor cycle of the MS-MFS algorithm was implemented as part of the CASACore set of libraries¹¹. Wide-band self-calibration

⁹Note that this description applies only to the MS-MFS algorithm which does not yet have built-in constraints based on a astrophysically-plausible range of values that all the higher-order spectral coefficients are allowed to take on.

¹⁰Common Astronomy Software Applications is used by the National Radio Astronomy Observatory

¹¹CASACore is a set of libraries that implement basic functionalities required for radio interferometric data analysis and is currently being shared by the NRAO for the EVLA and ALMA, the ATNF for ASKAP,

and the STACK+MFS hybrid method were implemented *via* CASAPY (python) scripts using the tool interface, and these scripts are not part of any formal release. The MS-MFS algorithm was also implemented in the ASKAPsoft¹² software package for use with the ASKAP telescope and was tested within the ASKAPsoft parallelization framework.

8.4.2 MS-MFS error estimation and feasibility

Section 8.4.2.1 describes various sources of error that can arise when the MS-MFS algorithm is used for wide-band imaging. Section 8.4.2.2 then describes how the algorithm is expected to perform in situations where additional information about the source is usually required.

8.4.2.1 Error Estimation

Dynamic-range limits when source spectra are ignored : If continuum imaging is done with only MFS gridding and source spectra are ignored, spectral structure will masquerade as spurious spatial structure. These errors will affect regions of the image both on-source and off-source and their magnitudes depend on the available *uv*-coverage, the frequency range being covered, the choice of reference frequency, and the intensity and spectral index of the source. A rough rule of thumb for an EVLA-type *uv*-coverages (see section 6.2.4.2) is that for a point source of with spectral index $\alpha = -1.0$ measured between 1 and 2 GHz, the peak error obtained if the spectrum is ignored is at a dynamic range of $< 10^3$. Note that when all sources in the observed region of the sky have similar spectral indices, these errors can be reduced by dividing out an average spectral index (one single number over the entire sky) from the visibilities before imaging them¹³.

Factors affecting the accuracy of the measured spectral index : Deconvolution errors contribute to the on-source error in the Taylor coefficient images, and these errors propagate to the spectral index map which is computed as a ratio of two coefficient images. Table 8.2 lists the estimated and observed errors in spectral index and curvature for a simulated example and shows that the deconvolution errors that result when a point-source flux model is used to deconvolve extended emission, can increase the error bars on the spectral index and

and ASTRON for the LOFAR telescope.

¹²Australian SKA Pathfinder software is being developed at the Australia Telescope National Facility

¹³Note that such a division will reduce the signal-to-noise ratio of the higher-order terms of the series (for the remaining spectral structure). Therefore, although the removal of an average spectral index could reduce the level of imaging artifacts obtained when source spectra are ignored, the lower signal-to-noise ratio of the spectral signature could increase the error on the derived spectral index when MS-MFS is used. Note also that this point is not specific to the MS-MFS algorithm, but is a general statement about how the accuracy of a fit depends on the SNR of the signal being fitted.

curvature by an order of magnitude. The accuracy to which α and β can be determined also depends on the noise per spectral data point, the number of sampled frequencies, the total frequency range of the samples, and the number of spectral parameters N_t in the fit. Section 6.2.4.1 discusses empirically derived error bars for the spectral index based on these factors.

Effect of the frequency-dependence of the Primary beam : When wide-band imaging is done across wide fields-of-view, sources away from the pointing center will be attenuated by the value of the primary beam at each frequency. Wide-band imaging results from such data ignoring the primary beam will contain spurious spectral structure. For the EVLA primary beams between 1 and 2 GHz, this extra spectral index at the half-power point is about -1.4 and about -0.6 at the 70% point (see Figs.5.4 and 7.7). Note that even if the source has a flat spectrum, this artificial spectral index can cause errors at the levels described for ignoring source spectra in the restored intensity image.

Accuracy to which the primary beam spectrum can be removed : Tests on simulated and real data show that up to the 70% point of the primary beam (at the reference frequency), the spectral index can be corrected to within 0.05 for point sources with signal-to-noise ratios of greater than 100, and to within 0.1 for point sources with signal-to-noise ratios of about 10. For extended emission, the errors are dominated by the effects of multi-scale deconvolution errors and not primary-beam correction. On high signal-to-noise simulations ($\text{SNR}>100$) with extended sources located at the 60% point of the primary beam at the reference frequency, the spectral index was recovered to within an error of 0.2.

8.4.2.2 Feasibility of wide-band imaging

Unresolved and Moderately resolved sources : Consider a source with broad-band continuum emission and spatial structure that is either unresolved at all sampled frequencies or unresolved at the low-frequency end of the band and resolved at the high-frequency end. The intensity distribution as well as the spectral index of such emission can be imaged at the angular resolution allowed by the highest frequency in the band. This is because compact emission has a signature all across the spatial frequency plane and its spectrum is well sampled by the measurements. The highest frequencies constrain the spatial structure and the flux model (in which a spectrum is associated with each flux component) naturally fits a spectrum at the angular resolution at which the spatial structure is modeled. Note that such a reconstruction is model-dependent and may require extra information in order to distinguish between sources whose observed spectra are due to genuine changes in the shape of the source with frequency and those with broad-band (power-law) emission emanating from each location on the source.

Very large sources : At the lower end of the sampled spatial-frequency range, the size of

the central uv -hole increases with observing frequency. For very large spatial scales whose visibility functions are adequately sampled (more than 80% of the integrated flux) only at the lower end of the frequency range, an ambiguity between spatial scale and spectrum can arise during the reconstruction. This is because the spectrum of this source is not well-sampled by the measurements. A flat-spectrum extended source can be mistaken for a steep spectrum less extended source, and *vice-versa*. This problem can be avoided by providing short-spacing flux constraints (from single-dish observations) to bias the solution, or by flagging all spatial frequencies below u_{min} at ν_{max} (the smallest spatial frequency sampled by the highest observed frequency) to filter out these large spatial scales.

Overlapping sources : When overlapping sources have different spectral structure, the result of wide-band imaging is the combined intensity and per-pixel spectrum. However, when the foreground and background structure has emission at very different spatial scales, a flux model that associates a spectrum with each flux component naturally separates the overlapping sources and represents the source as a sum of overlapping sources with different spectra. The intensity and spectrum of foreground sources can be recovered from the final output coefficient images by performing a polynomial subtraction (total spectrum - background spectrum), before computing the spectral index and curvature of foreground. Note that this is a simple extension of standard background subtraction.

Sources with band-limited emission : The observed spectrum of a source whose structure itself changes with frequency cannot be described using a power-law spectral model, but it can sometimes be described by a high-order polynomial ($N_t > 4$). The MS-MFS model with a high-order polynomial ($N_t > 4$) can be used to model these 'spectra' as long as the emission varies smoothly across frequency. In this case, images of spectral index and curvature have no meaning, and the final reconstructed images must be interpreted in terms of polynomial coefficients or by evaluating a spectral cube from these coefficients. Note however, that the highest angular resolution at which structure can be imaged is controlled by the highest observing frequency at which the emission is detected.

8.4.3 Multi-frequency synthesis *vs* single-channel imaging :

Image Fidelity and Dynamic Range : The main advantage of multi-frequency synthesis over single-channel imaging (for continuum imaging) is the increased image fidelity and dynamic range allowed by the use of the combined uv -coverage and broad-band sensitivity during image reconstruction.

Spatial resolution : The angular resolution of the continuum emission is at the resolution allowed by the highest frequency in the band. Further, MFS with a suitable flux model can reconstruct the spectral structure of the source also at the angular resolution allowed by the higher end of the sampled frequency range. Note that with single-channel imaging, the spectral structure can be recovered only at the angular resolution of the lowest frequency in the band.

Spectral Structure : The signal-to-noise ratio required to measure spectral structure is the same for both single-channel and MFS methods. Spectra can be measured accurately only for sources that are several times (~ 10) brighter than the single-channel noise level. The smallest spectral index that can be measured corresponds to a flux variation across the band that is comparable to the single-channel noise level. Note that these single-channel noise levels include errors in wide-band calibration.

Channel Averaging : Even if data are measured with a very high frequency resolution ($N_{chan} > 10000$) the process of imaging almost never requires it. Given a desired image field-of-view, one can calculate the bandwidth-smearing limit and average multi-channel data up to that limit. This will reduce the computational overhead for gridding and degridding. Note also that this is possible only for imaging. Calibration (and self-calibration loops) will still require the full frequency resolution.

Computation Cost : In general, MFS imaging is less expensive than single-channel imaging methods. However, single-channel methods are embarrassingly parallel¹⁴ and therefore very easy to distribute over a set of compute-nodes. The minor cycle of deconvolution MS-MFS imaging is hard to parallelize but the major cycle is easy to parallelize and significant speed-ups are still possible (this has been demonstrated *via* the ASKAPsoft implementation of MS-MFS).

Hybrid Methods : When wide-band measurements have very dense *uv*-coverage per frequency, wide-band calibration errors are minimal, and the target science does not require a very high angular resolution for spectral reconstructions, then a simple hybrid of single-channel imaging followed by a second stage of MFS imaging on the continuum residuals might suffice for high-fidelity and high dynamic-range continuum imaging. Also, if all sources of emission in the field of view have similar spectral indices, a common average spectral index can be removed from the calibrated data before continuum MFS imaging, to reduce the level at which errors due to unaccounted for spectral variations occur.

8.4.4 Future Work

Tests with real wide-band data : The imaging results presented in this chapter used either simulated wide-band EVLA data or multi-frequency data formed from a set of narrow-band VLA observations. This is because real wide-band EVLA data were not available at the time these algorithms were being developed (*i.e.* during the transition between the VLA and EVLA telescopes and before the EVLA wide-band correlator became available).

1. Data from the multi-frequency VLA observations demonstrated the ability of the MS-MFS algorithm to reconstruct spatial and spectral structure over wide-fields of

¹⁴Using parallel computing terminology, embarrassingly parallel problems are those that can be easily split into several smaller problems that can be operated upon independently and require minimal amounts of communication between compute nodes.

view, but did not test its high-dynamic-range capabilities. The MS-MFS imaging algorithm as well as the STACK+MFS hybrid need to be tested on real wide-band EVLA data to ascertain their high dynamic range imaging capabilities.

2. The effects of removing an average spectral index from a wide-band dataset before imaging need to be evaluated in terms of high-dynamic-range capability as well as the accuracy of the reconstructed spectrum.
3. Wide-band self-calibration needs to be tested to evaluate whether existing methods will suffice for high dynamic-range imaging.
4. Also, MS-MFS has been used only on extremely bright and extended objects (Cygnus A and M87) and a field of point sources (3C286) and tests on more typical sources are required before the conclusions described in this section can be applied generically.

Algorithm Improvements : There are several aspects of the MS-MFS algorithm for which improvements are possible.

1. One aspect of the MS-MFS algorithm that needs more work is how to determine appropriate values for N_t and N_s and to select a set of spatial scales. These parameters depend on the wide-band spatial structure of the sky brightness, the multi-frequency uv -coverage of the interferometer, the weighting scheme used, and the signal-to-noise of both spatial and spectral structure.
2. The MS-MFS algorithm uses a polynomial in I vs ν space to model the sky spectrum, even though broad-band radio emission usually follows power-laws. This is because the chosen flux model describes the wide-band sky brightness as a sum of overlapping extended flux components with fixed spectra, and a power-law sky spectrum cannot always be written as a sum of more power-laws. However, for sources with pure power-law spectra (*i.e.* isolated point sources, or extended emission with a constant spectral index across the source) a polynomial in $\log I$ vs $\log \nu$ space may be more appropriate in terms of the accuracy of the reconstructed values of α and β . This point needs to be tested, preferably on a field of compact sources.
3. The stability of the MS-MFS algorithm in the low signal-to-noise regime is yet to be understood. Non-linear constrained optimization techniques might have to be used instead of the simple linear least-squares methods described in the previous chapters in order to constrain solutions to astrophysically plausible values when the constraints from the data itself are insufficient.

Additional methods : All the algorithms described in this dissertation ignore source polarization and the ability to do full-polarization imaging using wide-band data is required for high sensitivity polarization measurements. This involves the use of an appropriate flux

model that accounts for the polarization signature of the source as a function of frequency during MFS imaging, and this is still a subject of research.

Another important application of wide-band imaging is the construction of a broad-band model of the continuum flux for the purpose of continuum subtraction. The standard practice has been to image the continuum emission using only those channels with no known spectral lines in them, and then subtract it out of the entire dataset. The same approach can in principle be used along with the MS-MFS algorithm to model and remove the broad-band continuum, as long as channels with spectral lines in them can be identified (or marked *via a-priori* information) before wide-band imaging. This needs to be tested with real wide-band data. As of now, continuum subtraction in the presence of a large number of unknown spectral lines remains a research problem.

CHAPTER 9

A HIGH-ANGULAR-RESOLUTION STUDY OF THE BROAD-BAND SPECTRUM OF M87

Cores of the densest galaxy clusters are expected to have cooling flows that trace radiative losses from the intra-cluster medium (ICM) and have cooling times shorter than a Hubble time. However, the hot cores of many clusters show no evidence of cooling below a temperature of roughly a third of the measured temperature in the inner regions of the cluster. One way of reconciling this cooling-flow problem is heating *via* accretion powered outflows from an active galactic nucleus (AGN) at its core. Observations of cluster-center radio galaxies (CCRGs) that host these AGN have suggested a feedback model that might be responsible for balancing the cooling flow. One aspect of this process that is not well understood is the mechanism by which energy from AGN outflows could be transported out into the thermal ICM and the timescales on which this happens. So far, most calculations of the lifetimes of features seen within the radio haloes (of sources like M87) have been based on source expansion models. Synchrotron spectra provide another way of studying the energetics and lifetimes of features in the halo. Observed wide-band spectra can be compared to those predicted by various evolution models to explain how they formed.

In this project, wide-band spectra of several regions of the M87 radio halo were constructed from existing high angular-resolution images at 74 MHz (4-band), 327 MHz (P-band), and 1.4 GHz (L-band) and a spectral index map between 1.1 and 1.8 GHz (constructed from wide-band VLA L-band observations). These spectra were compared with model spectra derived from two spectral evolution models (initial-injection and ongoing-injection). Preliminary results suggest that spectra in the inner few kpc (the inner radio lobes) are consistent with an ongoing injection of particles with the energy distribution as seen in the jet. For features in the halo, timescales consistent with expansion and buoyancy timescales can be obtained *via* the initial-injection model, but the data constrain the power-law index of the initial electron energy distribution to be steeper than that observed at the jet. These features can also be modeled *via* the ongoing-injection model for a wide range of initial energy distribution indices and give timescales that range from twice the expansion timescales for steep injected spectra to a few times smaller than the expected cooling time when the energy injection index is the same as that observed in the jet. Note, that the large error-bars on the current L-band spectral index estimates render all the wide-band spectra used in this analysis consistent with pure power laws and this introduces a high degree of uncertainty on any conclusions derived from estimates of break frequencies beyond the measured range.

Section 9.1 briefly describes the cooling-flow problem and the idea of AGN feedback and summarizes the relevant existing information about M87. Section 9.2 contains the basics of synchrotron spectra and their evolution *via* two different models and lists the calculations used for estimating B-fields and source lifetimes. Section 9.3 shows the results of spectral fits to these theoretical models for M87. Section 9.4 interprets the results in terms of plausible evolution models, ages and injected electron energy distributions.

9.1 The M87 cluster-center radio galaxy

M87 is a large elliptical radio galaxy located at the center of the Virgo cluster. Galaxy clusters usually show evidence of hot cores with strong X-ray thermal emission from the ICM near the center of the cluster. The total energy content of the hot ICM is given by its temperature as $E = \frac{3}{2}n_x k_B T$, radiative loss rates are proportional to density squared ($L \propto n_x^2$), and a cooling time can be computed as their ratio. This cooling time is inversely proportional to the density $t_{cool} \propto \frac{1}{n_x}$ or in other words, high-density regions cool faster. This means that the center of the cluster cools first, followed by outer regions, and this is called a cooling flow.

The expected cooling time can be calculated by measuring the density and temperature of the thermal ICM (from X-ray measurements of the bremsstrahlung spectrum). For M87, the cooling time estimated from X-ray measurements is $t_{cool} \approx 1\text{Gyr}$. and the cooling radius is not much larger than the observed size of the radio halo. The first problem one encounters is that t_{cool} is often much less than the Hubble time, suggesting that the cluster cores ought to have cooled by now and not still show high temperatures. The second problem is the lack of X-ray emission lines from the cooling gas below a third of the measured temperature of the cluster core. The frequency and amplitude of X-ray emission lines from the ICM gas (on top of bremsstrahlung emission spectrum) can be predicted for different temperatures. The observed lines can be matched to these predictions for a range of temperatures that it passes through as it cools, with the maximum being the background temperature. For sources like M87, these predicted lines are present down to a temperature of $3.5 \times 10^7\text{K}$ [Peterson et al. 2003]. This means that the gas is losing energy, but also not cooling below this point. These observations and calculations suggest there must be some internal source of energy, possibly correlated with the observed radio halo, that balances the cooling below that temperature, and keeps the cluster core hot.

One possible source of energy input capable of balancing the cooling flow is an accretion powered outflow from an AGN containing a super-massive black hole (SMBH) at the center of the cluster. The M87 galaxy hosts an AGN with an observed jet outflow, making it an ideal candidate for the study of AGN feedback as a possible explanation of the cooling-flow problem. Calculations of the jet power in M87 have been shown to roughly balance the energy loss due to thermal radiation in the ICM [Owen et al. 2000]. The mechanism of this energy transfer is thought to be a feedback loop in which the cooling

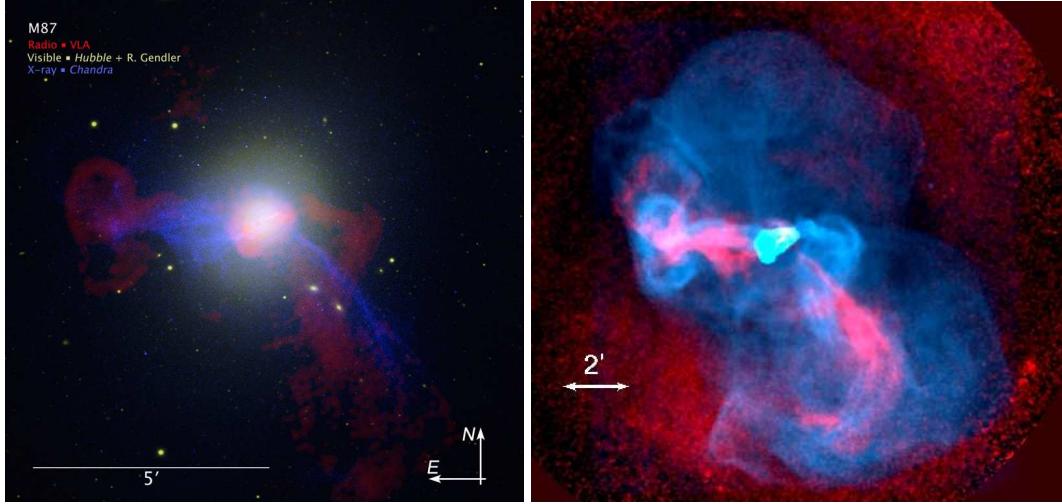


Figure 9.1: Radio/X-ray/Optical images of M87 : The image on the left is a composite of optical, radio and X-ray images of the elliptical galaxy M87 (Credits:X-ray: NASA/CXC/CfA/W. Forman et al.; Radio: NRAO/AUI/NSF/W. Cotton; Optical: NASA/ESA/Hubble Heritage Team (STScI/AURA), and R. Gendler). The image on the right is a composite of radio and X-ray images that shows the structures in the M87 halo and a strong correlation between X-ray and radio emission (Credits: Radio : NRAO/AUI/NSF/F.N.Owen; X-ray: NASA/CXC/Cfa/W.Forman et al.).

ICM gas sinks to the bottom of the gravitational potential well of the cluster and feeds the AGN *via* accretion so that the AGN pumps out a corresponding amount of energy through jet outflows. This energy is then transported out to the ICM to heat it up again.

The least-understood step in this loop is the mechanism by which the jet power is transferred across very large distances to heat up the ICM in all directions. In some galaxy clusters, there is evidence of bubbles rising buoyantly and in some cases, these bubbles are seen to displace the thermal ICM plasma and form cavities in the X-ray loud thermal ICM (evident as bounded regions of low X-ray luminosity compared to the surrounding, and often coincident with regions of high radio synchrotron emission). The inner lobes of the M87 radio emission coincide with one such X-ray cavity, but structures outside this region in the M87 halo do not (no observed X-ray cavity on large scales). Instead, the radio halo shows evidence of buoyant bubbles of plasma rising up from the AGN, and features seen in X-ray emission correlate roughly with some features in the radio halo, suggesting possible mixing of the radio plasma and the ICM. Another way of transporting energy to the ICM is through sound-waves and observations of the Perseus and Virgo clusters show ripples that look like propagating sound waves. Figure 9.1 shows two images of M87 to illustrate the relation between its observed optical, X-ray and radio emission.

9.1.1 Studying M87 evolution

The next step towards explaining the cooling flow problem for M87 *via* AGN feedback is to study the evolution of various structures seen in the radio and X-ray images and to understand the energetic processes present within them. Feedback processes and the timescales at which they may be occurring for the Virgo cluster can be studied by modeling the formation and evolution of various features in the M87 radio halo. One way of studying this is to use direct dynamics to model the source as a buoyant or driven and expanding bubble that physically transports energy between the AGN and the thermal ICM. Synchrotron spectra are another way of studying how various features in the M87 halo evolve spectrally as they carry energy away from the AGN. Synchrotron ages are independent of direct dynamics (bubble rise/expansion timescales, sound speed, etc), but are highly dependent on B-field estimates and the chosen model for the evolution of the energy distribution of the ensemble of radiating particles. Some dynamical age estimates derived from bubble expansion and buoyancy timescales are listed below, along with existing information about B-fields in the halo and information derived from low-resolution synchrotron spectra.

Fig. 9.2 shows an image of M87 at 327 MHz in which various features are labeled. Radio emission from M87 shows an energetic 2kpc jet and a pair of bright \sim 5kpc inner radio lobes. Outside this bright central region is a pair of \sim 20kpc East-West structures that appear to be connected to the bright central region and are labeled as the ear-lobe (East) and ear-canal (West). To the North and the South of the inner lobes are a pair of large \sim 40kpc diffuse structures (labeled as halos) with well-defined outer boundaries. All structures outside the inner radio lobes are comprised of narrow-extended bright features (labeled as filaments) with low-brightness diffuse emission in between (labeled as background).

Magnetic fields : If the M87 halo is an expanding lobe modeled by a fluid flow in pressure equilibrium with the ICM outside the bubble, the ambient pressure P_{amb} of the surrounding ICM can be used to calculate an upper limit on the average internal B-field B_{dyn} (*via* the expression $P_{amb} = B^2/8\pi$). From X-ray measurements of the ICM temperature, we get $P_{amb} = n_x k_B T = 1.2 \sim 4 \times 10^{-11} \text{ dyn/cm}^2$ where $T = 9 \sim 28 \times 10^6 \text{ K}$ and $n = 0.01$ [Owen et al. 2000; Shibata et al. 2001; Molendi 2002]. The B-field calculated from P_{amb} is $B_{dyn} = 17 \sim 31 \mu\text{G}$.

B_{dyn} is an upper limit only on the average internal B-field (over the entire halo), and turbulent flows and shock compressions on much smaller scales can enhance the B-fields in localised regions in the halo. The total energy density is given by $\rho v^2 + P_{amb} = B^2/8\pi$ where v is the local turbulent flow velocity and ρ is the density. The amount by which the B-field is enhanced due to turbulence will depend on v and its relation to the sound speed in the ICM ($c_s = \sqrt{kT/m}$). For example, in the case of a supersonic turbulent flow ($\rho v^2 > P_{amb}$), the additional magnetic energy density scales as the square of the Mach number ($\mathcal{M} = v/c_s$).

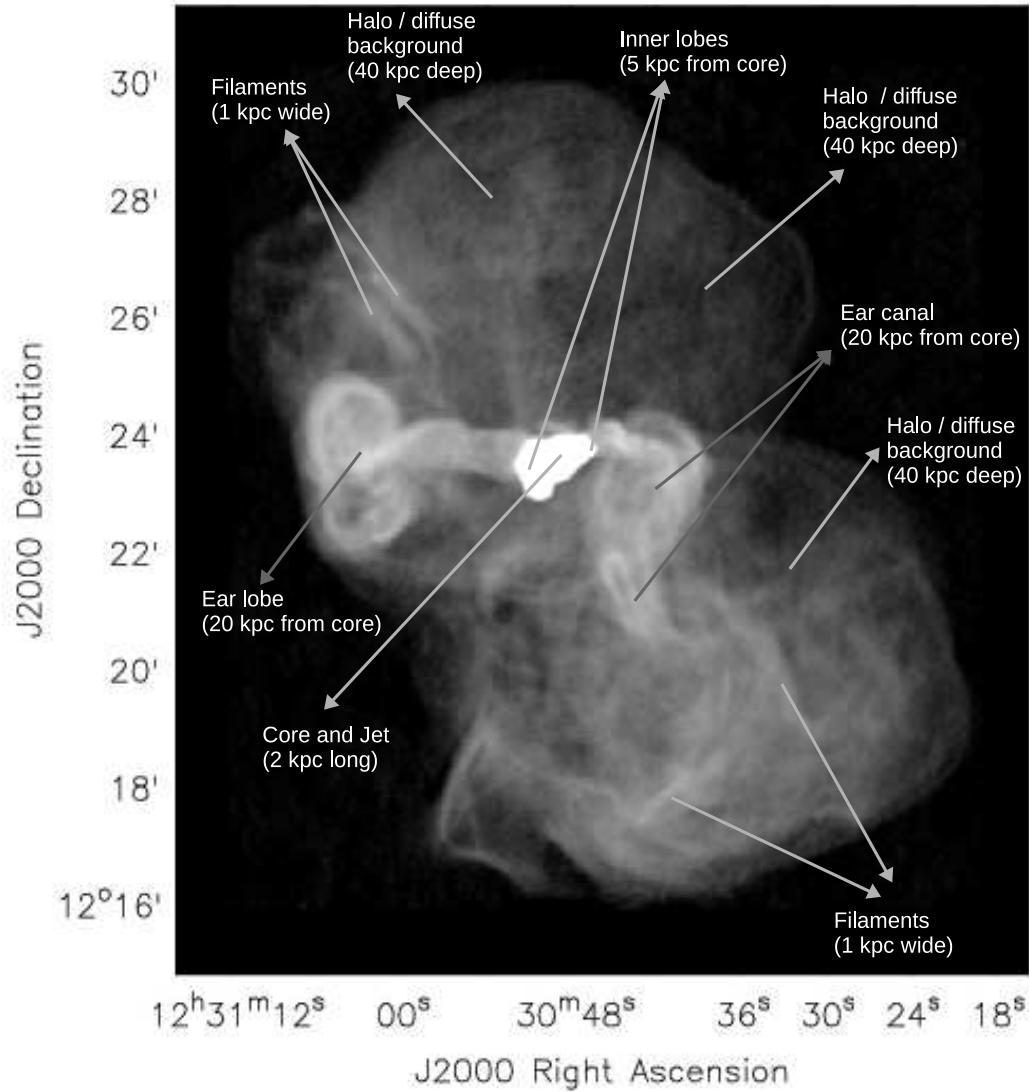


Figure 9.2: Labeled image of M87 : This is a 327 MHz image of the M87 radio galaxy, made using the VLA [Owen et al. 2000]. It shows a bright central region with a 2kpc radio jet and 5kpc inner radio lobes, a pair of ~20kpc structures to the East (ear-lobe) and West (ear-canal) of the bright central region, and two ~40kpc halos to the North and South. Narrow extended filamentary structure is seen throughout the ear lobe/canal structures and the halo, with low-level diffuse background emission in between.

Faraday-rotation measurements [Owen et al. 1990] estimate the B-fields around each inner radio lobe to be between 20 and 40 μG . Owen et al. [2000] show that the B-field consistent with the minimum pressure in various parts of the outer halo lie between 7 and 10 μG .

Dynamical Age (driven bubble) : For an expanding lobe powered by a constant energy source at the center and overpressured with respect to its surroundings, the age of the source can be estimated as the time taken for the outer edge of the lobe to expand to a certain size (volume). The volume of the lobe $V(t)$ is related to the input power, external number-density and lifetime as follows.

$$V(t) = c_v \left(\frac{\dot{E}}{m_p n_x} \right)^{\frac{3}{5}} t^{\frac{9}{5}} \Rightarrow t_{dyn} = \left[V \left(\frac{\dot{E}}{m_p n_x} \right)^{-\frac{3}{5}} \right]^{\frac{5}{9}} \quad (9.1)$$

where c_v is an order unity constant [Eilek 1996]. Using this expression with $n_x = 0.01$, volume $V(kpc^3) = \frac{4}{3}\pi 40^3$ we get $t_{dyn} \approx 120\text{Myr}$ for $\dot{E} \approx 10^{44}\text{erg/sec}$ ($n_x \approx 0.01$ and $\dot{E} \approx 10^{44}\text{erg/sec}$ were obtained from Owen et al. [2000]).

Dynamical Age (passive buoyant bubble) : For a buoyant bubble rising up through an atmosphere of hot plasma, models suggest $t_{buoyant} \approx 40 \sim 60\text{Myr}$ for a distance of about 40 kpc [Churazov et al. 2001]. An upper limit on the speed of such a bubble is given by the sound speed $c_s = \sqrt{kT/m_p}$. For example, for $T = 2 \times 10^7\text{K}$, $\Rightarrow c_s = 4.0 \times 10^7\text{cm/sec}$ and the sound travel time for a distance of 20 kpc is 50 Myr (and 100 Myr for 40 kpc).

Synchrotron spectra (jet) High angular-resolution studies of the M87 jet have shown that its spectral index at radio wavelengths is $\alpha_{jet} \approx -0.5$ (Owen, private commn.) and Bicknell and Begelman [1996] reconstruct this result via models of the jet outflow. Perlman and Wilson [2005] also show that the broad-band spectrum of the jet between radio and X-rays is consistent with a continuous injection of energetic particles ($\alpha_{jet} \approx -0.6$), or in other words, an active jet.

Synchrotron spectra (lobes and halo) Low-resolution synchrotron spectra [Rottmann et al. 1996a] show that the P-L spectral index (between 327 MHz and 1.4 GHz) in the halo is $\alpha_{PL} = -0.7 \sim -1.3$ and the C-X spectral index (between 5 GHz and 10 GHz) is $\alpha_{CX} = -2.0 \sim -2.8$. Rottmann et al. [1996b] analyse images at 333 MHz, 1.4 GHz and 10.55 GHz and suggest a spectral break between 5 and 11 GHz and timescales of 30~40 Myr for the ear lobe/canal regions (but they do not quote B-fields). These ages are roughly consistent with the timescales calculated from direct dynamics [Churazov et al. 2001] with $B=6.5\mu\text{G}$. However, the angular resolution in these images is insufficient to study the spectral variations across different features of M87 (lobes, bubbles, halo filaments and background).

High resolution images of the 40 kpc-scale structure in the M87 halo [[Owen et al. 2000](#)] have so far been made only at frequencies of 1.4 GHz and below, and probe only the edge of the expected region of spectral turnover (1 to 10 GHz). Across this range (75 MHz to 1.4 GHz), there is no clear sign of a spectral break or cut-off.

Goal of this project (test evolution models for M87) : In order to study the broadband spectra of various isolated features in the M87 halo, model their spectral evolution as a function of distance from the jet, and constrain their synchrotron ages, we need high angular-resolution observations that directly measure the spectrum between 1 and 10 GHz. The project described in this chapter is the first step and involved a wide-band observation of M87 between 1.1 and 1.8 GHz and the use of the wide-band imaging algorithms described in chapter [7](#) to construct a spectral-index map across L-band. This 1.1 to 1.8 GHz spectral index map was used along with existing images at 74 MHz, 327 MHz and 1.4 GHz to constrain the slope of the spectrum at the high-frequency end of the measured range (spectral slope at 1.4 GHz). Two types of synchrotron evolution models were tested by fitting these wide-band spectra to numerical models of spectra that were evolved over the approximate lifetime of the source, starting with different electron energy distributions.

9.2 Synchrotron spectra and their evolution

Section [9.2.1](#) summarizes the basics of synchrotron spectra [[Pacholczyk 1970](#)]. Section [9.2.2](#) describes the concept of spectral ageing, two models of synchrotron ageing based on an initial or a continuous injection of particles with a power-law energy distribution and shows the difference between the observed spectrum for these two cases. Calculations of the minimum-energy B-fields used in the synchrotron age estimates are also described here. Section [9.3](#) later describes the spectra obtained from multi-frequency images of M87, the process of fitting models to the data to obtain best-fit estimates of the critical frequency, and using them to calculate the ages of various features in the M87 halo.

9.2.1 Synchrotron radiation - basic facts

A charged particle moving in a magnetic field gyrates around magnetic lines of force, feels an acceleration towards the axis of its helical orbit, and radiates with a dipole power pattern around the direction of acceleration. When the charged particle moves at relativistic speeds, this is called synchrotron radiation.

For relativistic particles, the radiation pattern for each particle is no longer a symmetric dipole pattern and the power is boosted along the direction of motion of the particle (synchrotron beaming). To a distant observer, this radiation appears pulsed because as each particle moves around its orbit, its radiation beam intersects the observers line of sight only for a small fraction of its total orbit. The observed duration of these beamed

synchrotron pulses gives rise to a characteristic frequency ν_{syn} of the observed radiation.

$$\nu_{syn} = \frac{3}{4\pi} \frac{e}{mc} B \gamma^2 \sin \theta \quad (9.2)$$

where γ denotes particle energy $\gamma = E/mc^2$ and θ is the pitch angle. The shape of the synchrotron spectrum from a particle of energy γ is given by a modified Bessel function.

$$P_{syn}(\nu, \gamma) = \frac{\sqrt{3}e^3 B \sin \theta}{mc^2} \left(\frac{\nu}{\nu_{syn}} \right) \int_{\nu/\nu_{syn}}^{\infty} K_{\frac{5}{3}}(\eta) d\eta \quad (9.3)$$

The low-frequency end of this spectrum follows a power law of the form $\nu^{\frac{1}{3}}$, the high-frequency end shows an exponential decay $e^{-\nu/\nu_{max}}$ and the spectrum peaks at $0.29\nu_{syn}$. The total radiated power averaged over an ensemble of particles with energy given by γ and an isotropic distribution of pitch angles is given by

$$\langle P_{syn} \rangle = \frac{4}{9} \frac{e^4}{m^2 c^3} B^2 \gamma^2 = \frac{c \sigma_T}{6\pi} \gamma^2 B^2 \quad (9.4)$$

Here, σ_T is the Thomson scattering cross section. Astrophysical sources contain charged particles with a wide range of energies. From the observed power spectrum of cosmic rays, we choose a power-law distribution of particle energies $N(\gamma) = N_0 \gamma^{-s}$ (s is the spectral index of the power law for an energy range $\gamma_{min} < \gamma < \gamma_{max}$ where $\gamma_{max} \gg \gamma_{min}$). The total synchrotron spectrum is given by a convolution of the single-energy spectrum and $N(\gamma)$.

$$j_{syn}(\nu) = \int N(\gamma) P_{syn}(\nu, \gamma) d\gamma \propto B^{\frac{s+1}{2}} \left(\frac{\nu}{c_1} \right)^{\alpha} \quad \text{where } \alpha = -\frac{s-1}{2} \quad (9.5)$$

where $c_1 = 6.3 \times 10^{18} Hz$. The result (in the above energy range) is another power law with a spectral index α . The spectral shape at the low and high frequency ends of this spectrum follow that of the single electron energy spectrum.

9.2.2 Ageing of synchrotron spectra

Section 9.2.2.1 describes the computation of synchrotron age from a measured break frequency, and section 9.2.2.2 describes two evolution models that produce different spectral shapes on either side of the observed break.

9.2.2.1 Break frequency and synchrotron age

For an ensemble of particles with the same initial energy γ , there is a characteristic timescale associated with the lifetime of these radiating particles. This is known as the synchrotron age, and is estimated from the ratio of the total energy ($E = \gamma mc^2$) to the rate

of energy loss due to synchrotron radiation ($\dot{E} \propto -\gamma^2 B^2$). Therefore, $t_\gamma = E/\dot{E} \propto 1/B^2\gamma$ and particles with higher energies (or located in regions of higher B-field) radiate faster and have a shorter lifetime.

In an ensemble of particles spanning a wide range of initial energies, the higher-energy particles radiate and deplete faster. After a time t , all particles at energies high enough such that $t_\gamma < t$ would no longer be radiating. This creates a break in the electron energy distribution at γ_c such that $t_{\gamma_c} = t$. When $N(\gamma) \propto \gamma^{-s}$, a break in the energy distribution at γ_c causes a break in the power-law of the observed synchrotron spectrum at a critical frequency ν_c (related to γ_c via Eqn. 9.2). The shape of the spectrum on either side of the observed break will depend on the initial value of s and the time dependence of $N(\gamma)$. As time progresses, this break will move to lower frequencies but the shape of the spectrum on either side of the break will not change.

If the B-field is known, the age of a population of relativistic particles can be estimated from measurements of the critical or break frequency ν_c . Let B denote the local B-field with which the observed synchrotron emission is associated. Let $B_{rad} = \sqrt{8\pi U_{rad}}$ (where $U_{rad} \propto T^4$ and $T=2.7$ K) denote the equivalent B-field due to inverse-Compton losses (the minimum $B_{rad} = 3\mu G$ and corresponds to energy lost when CMB (cosmic microwave background) photons scatter off the relativistic particles and gain energy). The energy loss rate due to synchrotron radiation (Eqn. 9.4) is given by

$$\frac{d\gamma}{dt} = -k\gamma^2(B^2 + B_{rad}^2) \quad \text{where } k = \frac{\sigma_T}{6\pi mc} \quad (9.6)$$

Eqn. 9.6 can be solved to obtain an expression for a critical energy γ_c . This critical energy represents the maximum particle energy present in the ensemble after a time $t_{syn} = \int dt$ (t_{syn} is called the synchrotron lifetime).

$$\gamma_c = \frac{1}{k \int [B^2(t) + B_{rad}^2] dt} \quad (9.7)$$

$B(t)$ represents a time-varying B-field as encountered by the particle. The critical energy γ_c can be related via Eqn. 9.2 to a critical frequency ν_c . This critical frequency is a measured quantity, and is the observed break frequency of the synchrotron spectrum. Given ν_c , a synchrotron age t_{syn} can be computed from Eqns. 9.7 and 9.2 for two different situations, as follows (note that all t_{syn} calculations in this chapter use electron masses $m = m_e$).

Homogeneous B-field : If the particles have seen a constant B-field over their entire lifetime, either by moving through a homogeneous B-field or by not moving very far in an inhomogeneous B-field ($B(t) = B$), we can calculate the synchrotron lifetime t_{syn} as follows.

$$t_{syn} = \left[\frac{27\pi emc}{\sigma_T^2} \right]^{\frac{1}{2}} \left[\frac{B}{[B^2 + B_{rad}^2]^2} \right]^{\frac{1}{2}} \nu_c^{-\frac{1}{2}} \quad (9.8)$$

Here, t_{syn} is in seconds, ν_c is in Hz and B is in Gauss. Eqn. 9.8 can be written as $t_{syn} = 1.6 \times 10^9 B^{-\frac{3}{2}} \nu_c^{-\frac{1}{2}}$ years, where B is in μG , ν is in GHz and B_{rad} is neglected.

Inhomogeneous B-field : If the particles have encountered varying B-fields during their lifetimes, a modified calculation of t_{syn} is required. A particular measured critical frequency $\nu_c \propto B\gamma_c^2$ can be obtained either from particles at a high energy and low B-field, or by lower-energy particles in a higher B-field. Therefore, if the particles have spent a large fraction of their lifetime in a low field region before moving to a high-field region from where they are currently radiating, a synchrotron lifetime calculated from the observed ν_c via Eqn. 9.8 will give ages that are shorter than the true lifetime of the particles (Eqn. 9.8 assumes that the particle has spent its entire lifetime in the (higher) B-field that it is currently encountering).

To account for this discrepancy, we can re-write Eqn. 9.8 in terms of past and present B-fields. Let B_{now} represent the B-field from which the particles are currently radiating. Let $B(t) = \langle B \rangle$ represent an average B-field that the particle has encountered through most of its lifetime. We can calculate a synchrotron age t_{syn} as follows.

$$t_{syn} = \left[\frac{27\pi emc}{\sigma_T^2} \right]^{\frac{1}{2}} \left[\frac{B_{now}}{[\langle B \rangle^2 + B_{rad}^2]^{\frac{1}{2}}} \right]^{\frac{1}{2}} \nu_c^{-\frac{1}{2}} \quad (9.9)$$

If a particle spends most of its lifetime in a low B-field region but is currently radiating from a high B-field region, using $B_{now} > \langle B \rangle$ will give a larger and perhaps more accurate t_{syn} . This calculation can be used to interpret the observed synchrotron spectra in regions that appear to have localized high B-fields compared to their surroundings (for example, narrow magnetically confined filaments located within a large region of diffuse radio emission). This model may be useful in situations where Eqn. 9.8 gives lifetimes that are much shorter than any physically plausible dynamical model of particle transport across large distances (*i.e.* from the source of energetic particles to the locations where they are currently radiating from), especially if there is additional evidence to suggest localized high B-field regions or sites of local particle re-acceleration.

9.2.2.2 Ageing models and spectral shapes

The age of an ensemble of radiating particles is related to the observed break frequency ν_c , but the shape of the observed spectrum on either side of this break depends on the initial particle energy distribution and how this energy distribution evolves with time. As the source ages, ν_c decreases, but the shape of the spectrum below and above ν_c does not change.

Let $N(\gamma, t)$ describe the electron energy distribution function in terms of energy γ and time t . As particles age and lose energy the change in the shape of $N(\gamma, t)$ can be written in terms of a continuity equation for the number density of radiating particles in a one-dimensional energy space (see section 6.3 of [Pacholczyk \[1970\]](#)).

$$\frac{\partial N(\gamma, t)}{\partial t} + \frac{\partial}{\partial \gamma} \left[N(\gamma, t) \frac{d\gamma}{dt} \right] = Q(\gamma, t) \quad (9.10)$$

where $N(\gamma, t) \frac{d\gamma}{dt}$ is the flux of electrons with energies passing through the value γ in one unit of time as the result of losses and gains of energy by the electrons. The source function $Q(\gamma, t)$ gives the number of electrons at each energy that are injected into the radiating region at unit time and per unit energy interval.

- 1. Initial Injection model :** An initial power law distribution of particles is allowed to age without further replenishment. This is modeled using $Q(\gamma, t) = \delta(t - t_0)\gamma^{-s}$. As the particles age, a critical energy γ_c forms, beyond which all particles have stopped radiating and this gives a spectral break at ν_c (related to γ_c via Eqn. 9.2).

Within the energy range over which this initial synchrotron spectral power-law holds ($\gamma_{min} < \gamma < \gamma_{max}$), the spectral index on the low-frequency side of this break is $\alpha = -\frac{s-1}{2}$ where s is the power-law index of the initial energy distribution, and the spectrum on the high-frequency side shows exponential decay (from the single-energy power-spectrum at the highest surviving energy).

- 2. Ongoing Injection model :** A set of particles with a power-law distribution of energies is continually injected into the system. Particles at all energies are therefore aging as well as being replenished. However, since the high energy particles age faster, there will still be a break in the spectrum at ν_c , but this break is not as sharp as for the initial-injection model.

This form of ageing is modeled by choosing $Q(\gamma, t) = \gamma^{-s}$ to represent a constant input of particles with the same energy distribution, and setting $\frac{\partial N(\gamma, t)}{\partial t} = 0$ to calculate a steady-state solution above the break frequency. This solution is given by $N(\gamma) \propto \gamma^{-(s+1)}$. With this $N(\gamma)$ in Eqn. 9.5, the resulting spectrum has a spectral index of $\alpha=0.5$ where $\alpha = -\frac{s-1}{2}$. Therefore, the observed spectrum below ν_c is a power law derived from the initial power law distribution of electron energies and it steepens by $\Delta\alpha = -0.5$ across the break frequency.

Spectral models representing the two above cases can be obtained by numerically solving Eqn. 9.10. For M87, a set of spectra were generated using electron energy distributions whose power-law indices s range from 1.8 to 2.8 in steps of 0.1 and evolving them over 60 Myr. These solutions were obtained for the initial injection as well as ongoing injection models described above¹ and the only difference between the two models is the form of the source term $Q(\gamma, t)$. The resulting spectra are given in terms of ν/ν_c where ν_c represents a critical frequency at which a spectral break occurs. Figure 9.3 shows an example of the predicted wide-band spectra for $N(\gamma) \propto \gamma^{-2.0}$ resulting from no ageing (initial conditions), and ageing via the initial and ongoing injection models.

¹These numerical solutions were computed by J.A.Eilek and all the spectral fits described in this chapter used the resulting model spectra.

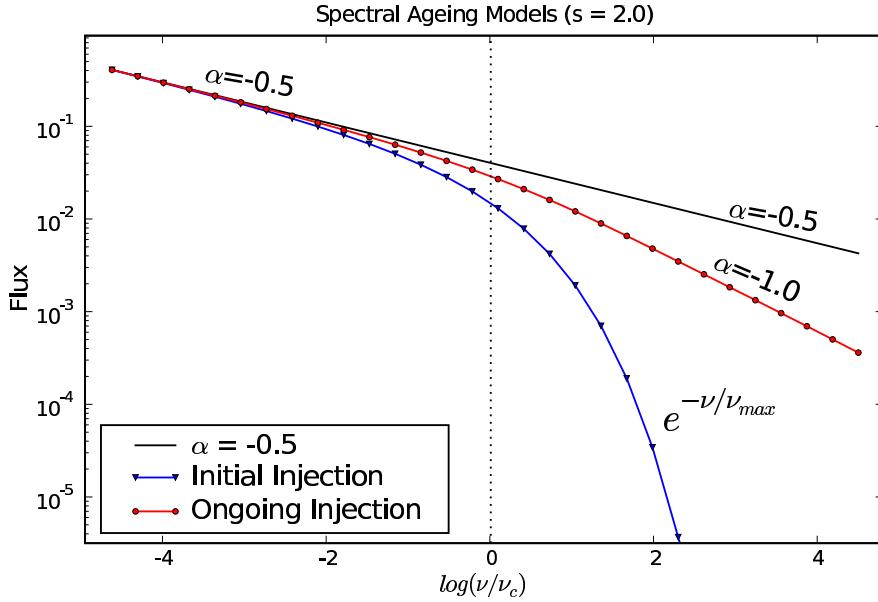


Figure 9.3: Spectral Ageing models : This plot shows examples of the predicted wide-band spectra resulting from no ageing (initial conditiona), and ageing *via* the initial and ongoing injection models. These spectra are plotted as functions of ν/ν_c to show the steady-state solutions. A particle energy distribution of $N(\gamma) \propto \gamma^{-2.0}$ was chosen, giving rise to an initial power-law spectrum with $\alpha = -0.5$. Ageing *via* the initial injection model shows an exponential decay beyond ν_c . Ageing *via* the ongoing injection model shows a steepening of the spectral index by 0.5 across the break.

Other models : There are several other theoretical models for the evolution of synchrotron spectra that are based on non-uniform or time-variable B-fields and turbulence. Eilek et al. [2003] discuss how local MHD turbulence could energize particles throughout the halo, replenish the high-energy particles, and prevent the observed spectrum from steepening. *In-situ* particle acceleration can also occur in regions with varying B-field strengths due to particles scattering off turbulent Alfvén waves [Eilek et al. 1997]. However, there are no established methods of predicting the electron synchrotron spectra resulting from this form of *in situ* acceleration [Eilek et al. 2003]. Power-law synchrotron spectra with spectral breaks can also result from power-law distributions of B-field strengths [Eilek and Arendt 1996].

9.2.2.3 Computing Equipartition B-fields

Calculations of the synchrotron age of a source *via* Eqn. 9.9 require that the B-field be known. In the absence of measurements that directly probe the B-field strength, equipartition provides a commonly used estimate.

The observed synchrotron luminosity L_{syn} depends on the magnetic field B , as

well as the total electron energy U_{el} , both of which are unknown. The total energy of a synchrotron source is the sum of the energy in the magnetic fields and from relativistic particles $U_{tot} = U_B + U_{el}$. Minimizing the total energy U_{tot} with respect to B results in a relation of approximate equality between U_{el} and U_B (equipartition).

$$U_B = \frac{3}{4}(1+k)U_{el} \Rightarrow U_{tot(min)} = \frac{7}{4}(1+k)U_{el} = \frac{7}{3}U_B \quad (9.11)$$

where $kU_{el} = U_{pr}$ is the energy contribution from protons. $U_{tot(min)}$ is then considered as the minimum total energy required to make a synchrotron source, and can be related directly to L_{syn} and the volume of the source. The total minimum energy density u_{min} and the minimum-energy B-field B_{eq} (often referred to as the equipartition B-field²) can be computed as follows.

$$u_{min} = \frac{U_{tot(min)}}{\Phi V} = c_{13} \left(\frac{3}{4\pi} \right)^{\frac{3}{7}} (1+k)^{\frac{4}{7}} \Phi^{-\frac{4}{7}} V^{-\frac{4}{7}} L_{syn}^{\frac{4}{7}} \quad (9.12)$$

where L_{syn} is the source luminosity, V is the source volume, Φ is a fraction of the source volume occupied by the magnetic field, and c_{13} is a constant that depends on the spectral index and frequency range over which this calculation is being performed (tabulated in [Pacholczyk \[1970\]](#)). The minimum-energy B-field can then be computed as follows.

$$B_{eq} = \left[\frac{24\pi}{7} u_{min} \right]^{\frac{1}{2}} \quad (9.13)$$

[Govoni and Feretti \[2004\]](#) rewrite Eqn. 9.12 in terms of measured quantities (I_0 [$mJy/asec^2$] at a frequency ν_0 [MHz], spectral index α between two frequencies ν_1, ν_2 and source depth D [kpc]). u_{min} can be written in units of [$ergs/cm^3$] as

$$u_{min} = \zeta(\alpha, \nu_1, \nu_2) (1+k)^{\frac{4}{7}} \nu_0^{\frac{4\alpha}{7}} (1+z)^{\frac{12+4\alpha}{7}} I_0^{\frac{4}{7}} D^{-\frac{4}{7}} \quad (9.14)$$

where z is the source redshift, and $\zeta(\alpha, \nu_1, \nu_2) = \left(\frac{2\alpha-2}{2\alpha-1} \right) \frac{\nu_1^{(1-2\alpha)/2} - \nu_2^{(1-2\alpha)/2}}{\nu_1^{(1-\alpha)} - \nu_2^{(1-\alpha)}}$. Tabulated values of ζ are presented for $\nu_1 = 10MHz$, $\nu_2 = 10GHz$, for α between 0.0 and 2.0 in increments of 0.1. Note that these values contain the assumption that α changes by less than 0.1 between 10 MHz and 10 GHz. When spectral curvature ($\delta\alpha > 0.1$) is measured, a piecewise linear approximation of the log spectrum may be more appropriate. However, for the calculations in this chapter, we used $k=1$, and the listed values of $\zeta(\alpha, 10MHz, 10GHz)$, specifically $\zeta = 6.77 \times 10^{-13}$ for $\alpha = 0.9$.

For a constant homogeneous B-field filling the entire volume of the source, the source depth D is estimated from the spatial extent of the observed emission. A bright

²The minimum-energy B-field is derived by minimizing $U_{tot} = U_{el} + U_B$, the minimum-pressure B-field is derived by minimizing $P_{tot} = P_{el} + P_B$, and the equipartition B-field is derived from setting $U_{el} = U_B$. All three methods give similar B-fields, and are often used interchangeably.

filament atop an extended background may be considered as a region of high B-field, compared to the background. Therefore B-fields can be computed separately for foreground and background features, with a source depth corresponding to the diameter of a filament for the foreground calculation.

9.3 Data, Spectral Fits and Synchrotron Ages

Section 9.3.1 describes the multi-frequency images of M87 that were used for this project and shows the measured spectra and calculated equipartition B-fields for different regions of the source. Section 9.3.2 describes the spectral-fitting process used and the results obtained (best-fit critical frequencies for different electron energy distributions, for two evolution models). Section 9.3.3 lists the synchrotron ages calculated using the best-fit critical frequencies.

9.3.1 M87 Spectral data

Intensity Images : The intensity images used for this analysis were existing VLA images of M87 at 4, P, and L bands for the halo (see Fig.9.4) and 4, P, L and C bands for the inner bright region (core, jet and inner lobes)³. Images at each of these frequencies were smoothed to 25 arcsec resolution to match the angular resolution of the 74 MHz image (the measured flux values are in units of $Jy/beam = Jy/(25arcsec)^2$).

1.1 to 1.8 GHz spectral Index map : A spectral index map across L-band (1.1 to 1.8 GHz) was obtained *via* the wide-band observations discussed in section 8.3.2. This spectral index map was used along with the existing 1.4 GHz intensity map to estimate the total intensity at 1.1 and 1.8 GHz. The L-band intensity and spectral index maps were corrected for the VLA primary beam and its frequency dependence *via* a post-deconvolution correction.

Error-bars : The data used for spectral fits were at 74 MHz, 327 MHz, 1.1 GHz, 1.4 GHz and 1.8 GHz (including 4.8 GHz for the bright central region). Error-bars for the data points were computed as $\sqrt{\sigma_{fluxscale}^2 + \sigma_{rms}^2}$, where $\sigma_{fluxscale}$ is a 3% error due to absolute amplitude calibration. σ_{rms} is an image-based rms error, derived from the off-source rms and averaged by the number of pixels in the flux calculation ($\frac{\sigma}{\sqrt{N}}$). For the 1.1 and 1.8 GHz points, errors were computed *via* error-propagation using the errors on the 1.4 GHz image and the L-band spectral-index map.

³The VLA images of M87 at all four bands were obtained from F.N.Owen and then regridded and smoothed to match their angular resolutions.

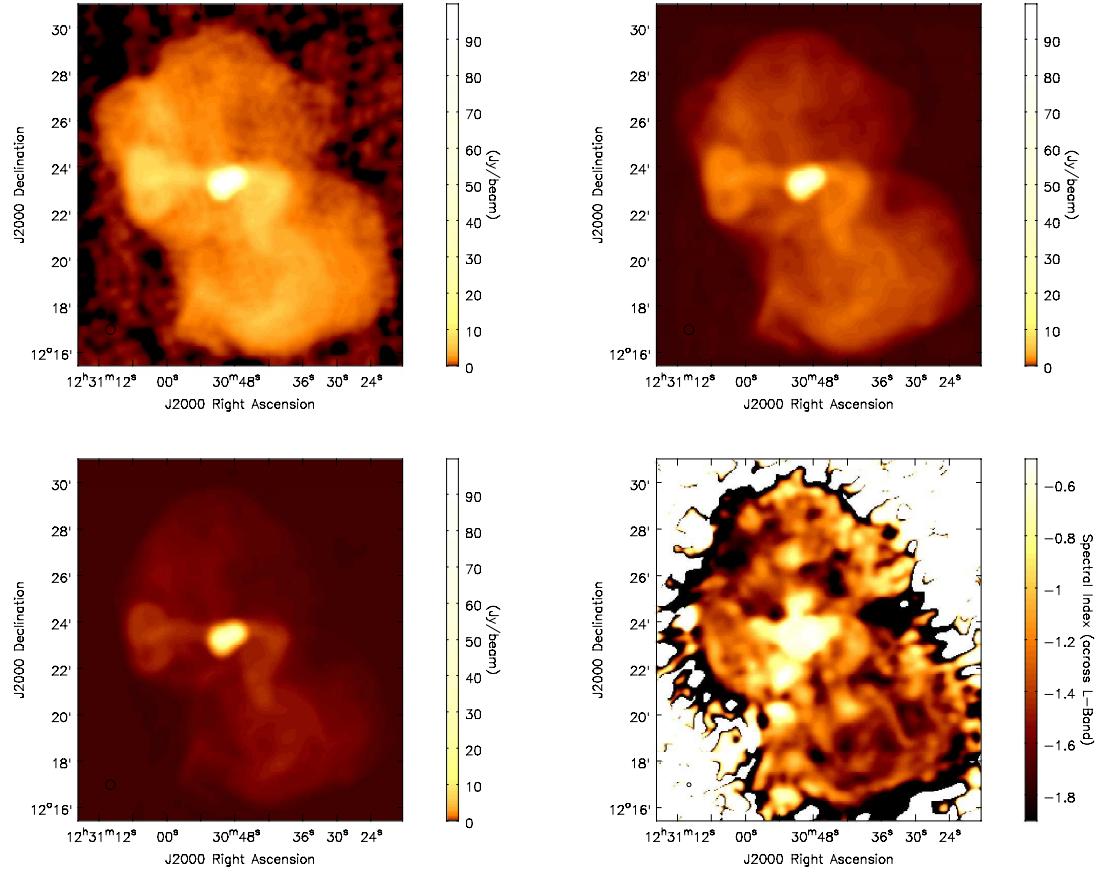


Figure 9.4: M87 : Stokes I images at 74 MHz (top left), 327 MHz (top right) and 1.4 GHz (bottom left), and the spectral-index map between 1.1 and 1.8 GHz (bottom right). All images at 25 arcsec resolution and the total-intensity images are displayed with the same flux-scale. The spectral index map was constructed from smoothed versions of the first two coefficient images produced by the MS-MFS algorithm.

Average spectral index across the source : Figure 9.5 shows spectra derived from these data for 11 regions across M87, along with the result of fitting a pure power law (single spectral index across the entire frequency range) to them. The regions were chosen as follows. L and M are measured in the core/jet and inner lobes, A, B and C are in filamentary regions in the bright 'ear-lobe' and 'ear-canal' regions, D,E and F are in fainter filamentary structure in the outer halo and G,H and I are meant to represent the diffuse halo background.

1. The first point to note from the fitted spectral indices is that the central bright region shows an average spectral index consistent with that measured from high angular resolution images of the M87 jet and inner lobes.

2. Second, the fitted values of α outside the bright central region show a slight gradient in the spectral index (a steepening of about 0.1) between the 'ear' structures and the rest of the outer halo. However, the uncertainty on the fitted value of α is itself about 0.05 (estimated from the spectral variations within each box), making the results consistent with no spectral gradient.
3. Finally, all the spectra for regions outside the central bright region show only a slight hint of steepening at 1.4 GHz. Using the current VLA L-band spectral index map the single-pixel error bars are large enough that this steepening is consistent with no steepening, but when the image RMS is averaged over the regions marked by the boxes, the error bars become comparable or less than the amount of steepening.

Overall, these wide-band spectra are consistent with pure power-laws. There are hints of spectral steepening across L-band, which is consistent with existing low-resolution measurements that show a significant steepening somewhere between 1 GHz and 10 GHz. However, additional measurements are required to confirm this. In particular, since the current L-band spectral index map was constructed from 10 VLA snapshots at 16 frequencies between 1.1 and 1.8 GHz, a real wide-band EVLA D-configuration observation at L-Band is expected to improve the deconvolution results and therefore reduce the error-bars on the L-band data points (a D-configuration observation will also better constrain the spectrum of the low-level extended halo emission). Further, high angular-resolution observations between 2 GHz and 10 GHz are also required to confirm if this steepening suggested by the L-band spectral index maps is real or not and to assess if there are significant differences between different parts of the halo. Note that at these higher frequencies with the EVLA, wide-band mosaicing observations will be required.

Section 9.3.2 describes a series of spectral fits that were done with the existing data and the L-band steepening it suggests, to estimate synchrotron ages for the initial injection and ongoing injection models of spectral evolution.

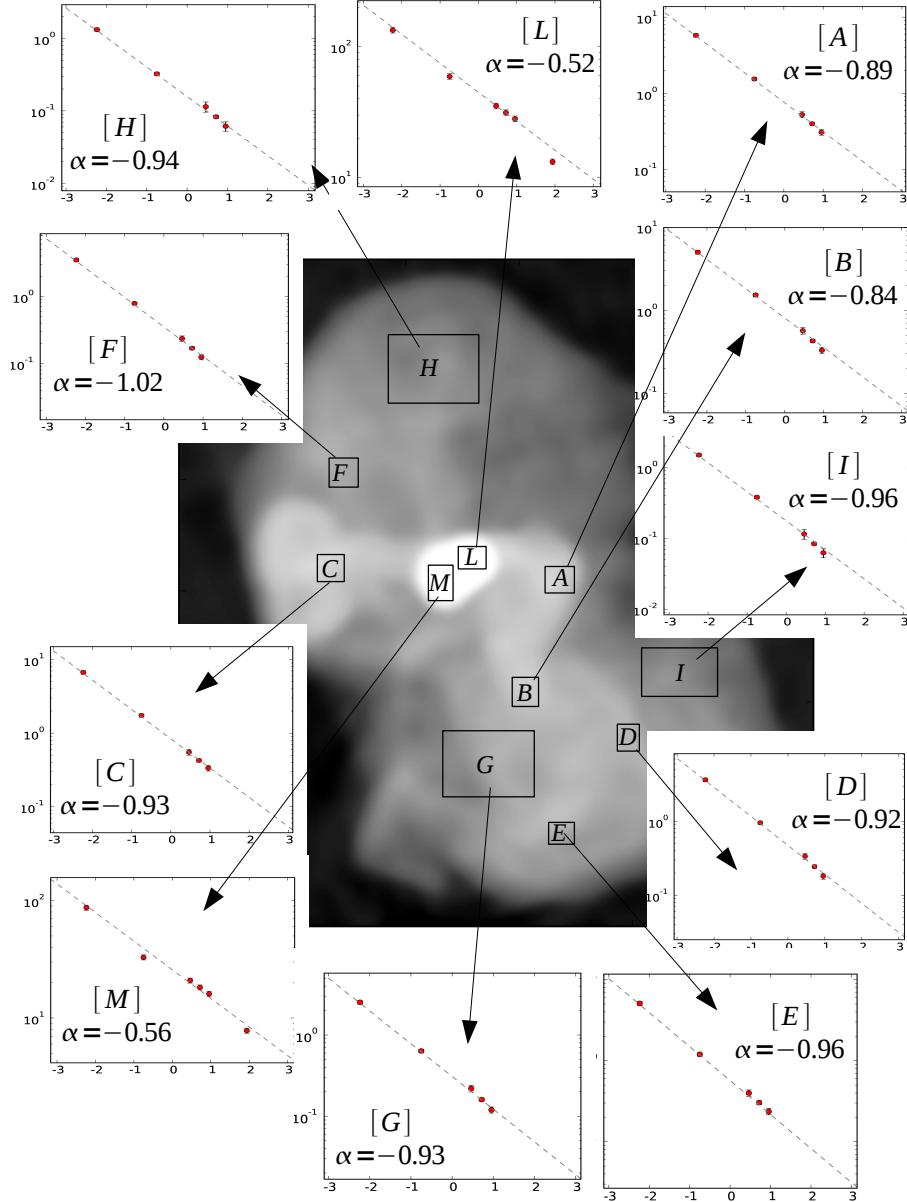


Figure 9.5: Spectral index - all over the source : This figure shows the measured intensities at 74 MHz, 327 MHz, 1.1 GHz, 1.4GHz, 1.8 GHz (and 4.8 GHz for the central region) for 11 regions across M87, along with the result of fitting a pure power law (single spectral index across the entire frequency range) to them. A few trends to note from these plots are (a) α in the central bright region is consistent with the known α of the M87 jet. (b) there is a slight gradient ($\Delta\alpha \lesssim 0.1$) between inner and outer regions of the halo (A,B,C vs G,H,I), but this variation within the error bar of the fit ($\delta\alpha \approx 0.05$) and (c) most regions show a slight steepening of the spectrum at 1.4 GHz, but this steepening is significant with respect to the error-bars only when averaged over several image pixels.

Region	D [kpc]	I_0 [mJy/asec 2]	α	B_{eq} [μG]
L	5 (10)	54.251	-0.50	33.2 (27.2)
M	5 (10)	43.196	-0.53	32.9 (27.0)
A	20 (40)	0.845	-0.89	10.0 (8.2)
B	20 (40)	0.764	-0.84	8.6 (7.03)
C	20 (40)	0.950	-0.93	9.8 (8.07)
D	40	0.437	-0.92	6.2
E	40	0.573	-0.96	7.4
F	40	0.359	-1.02	6.7
G	40	0.302	-0.93	5.8
H	40	0.132	-0.94	4.6
I	40	0.120	-1.01	4.7

Table 9.1: Minimum-energy B-fields in M87 : This table shows minimum-energy/equipartition B-fields computed for several regions across M87. The intensities I_0 were picked from the 1.4 GHz image (at 25arcsec resolution and scaled to compute I_0 in units of mJy/asec 2), and spectral indices α were from single power-law fits, for the regions labeled in Fig. 9.5. Eqn. 9.14 was used to compute the B-fields, for the listed values of distances D . These B-field values were used in Eqn. 9.9 to compute the synchrotron ages listed in Table 9.3 using the assumption of $B_{now} = \langle B \rangle$.

9.3.1.1 Calculating B-fields

Minimum energy B-fields were computed for several regions of M87 (as labeled in Figure 9.5). The values of I_0 were taken from the L-band ($\nu_0 = 1.4$ GHz) image, and α is the best-fit single α across the full sampled frequency range. Minimum energy B-fields were computed *via* Eqn.9.14 with $z = 0.02$ and $k = 1$. The following tables list the minimum energy B-field computed for each region along with the chosen source depth, the observed intensity and average spectral index. Table 9.1 shows the B-fields computed using the observed intensities. Table 9.2 shows B-fields computed by treating the observed filaments as foreground sources on a diffuse background. Filament intensities and spectral indices were computed by subtracting the average flux measured in two regions and recomputing the spectral index. The source depth used for the foreground B-field calculation was estimated from the observed width of the filaments (≈ 1 kpc).

The B-fields listed in Table 9.1 for regions D through I roughly agree with minimum-pressure estimates listed in Owen et al. [2000] as well as Owen et al. [1990] which derive B-fields from Faraday-rotation measurements around the inner radio lobes (regions L and M). The numbers also show that B-fields in regions A,B and C are stronger than elsewhere in the halo (even when the same source depth of 40 kpc is used for all regions). The central bright region shows a significantly higher B-field (with a source depth of 5 to 10 kpc), as do the filament B-fields computed with source depths of 1 kpc. For comparison, the maximum average B-field computed using pressure-balance arguments from the energy density of the external ICM thermal gas (measured *via* its temperature) ranges between $B_{dyn} = 18 \sim 31 \mu G$ for the observed range of temperatures.

Filament	I_{fil} [mJy/asec 2]	α_{fil}	B_{eq} [μG]
A - I	0.724	-0.87	21.5
A - H	0.711	-0.89	22.2
B - I	0.643	-0.78	19.6
B - H	0.629	-0.81	18.0
C - I	0.846	-0.91	21.3
C - H	0.828	-0.93	22.1
D - I	0.317	-0.86	16.5
D - H	0.300	-0.91	15.7
E - I	0.459	-0.94	19.1
E - H	0.441	-0.96	19.4

Table 9.2: Minimum-energy B-fields for M87 filaments : This table shows minimum-energy/equipartition B-fields computed for several filamentary regions across M87. These regions are spatially compact but long and are treated as being separate from the diffuse background. The filament intensities and spectral indices were computed using the difference between the intensities measured on a filament and the diffuse background. A source size of $D = 1.0$ kpc was used for all these calculations, to represent the filament thickness as seen from high resolution images. These filament B-fields are later be used to compute synchrotron lifetimes (listed in Table 9.4) via Eqn. 9.9 where $B_{now} = B_{eq}$ for the filaments, and $\langle B \rangle$ as B_{eq} for the background (from Table 9.1).

9.3.2 Spectral Fitting

This section describes the process used to fit the measured wide-band spectra to the initial injection and ongoing injection spectral evolution models, and the results obtained for different parts of the source. Model spectra were obtained as described in section 9.2.2.2 for 11 values of s ranging from 1.8 to 2.8 ($N(\gamma) \propto \gamma^{-s}$) and evaluated for 30 frequencies ranging from 10 MHz to 10 GHz. The data consist of 5 (or 6) flux measurements between 75 MHz and 1.8 GHz (or 4.8 GHz).

Goal : For each value of s , find a ν_c that gives the best fit of the data to the model. Obtain best-fit solutions for both the initial injection and ongoing injection models.

Method : The two variable parameters are ν_c and an amplitude scaling factor. The model spectra are described in terms of ν/ν_c . Therefore, for the process of fitting, ν_c is a free parameter that decides how the data points shift along the x-axis (defined by ν/ν_c). The amplitudes of the models are arbitrarily scaled. To compare them with the data, they need to be scaled to match the data at one frequency. (The choice here was 74 MHz.) For each model, χ^2 was computed⁴ for a range of possible values for ν_c , and value corresponding to the minimum χ^2 was chosen as the best-fit ν_c .

⁴Reduced χ^2 values were computed using these 5 data points, 3 degrees of freedom (since ν_c is the only parameter being fit for each s), and an estimate of the data variance obtained as a few percent of the flux at L-band. However, such an estimate made from 5 irregularly spaced data points with non-Gaussian errors is not a robust measure of the goodness of fit that can be compared to the ideal value of 1.0. The χ^2

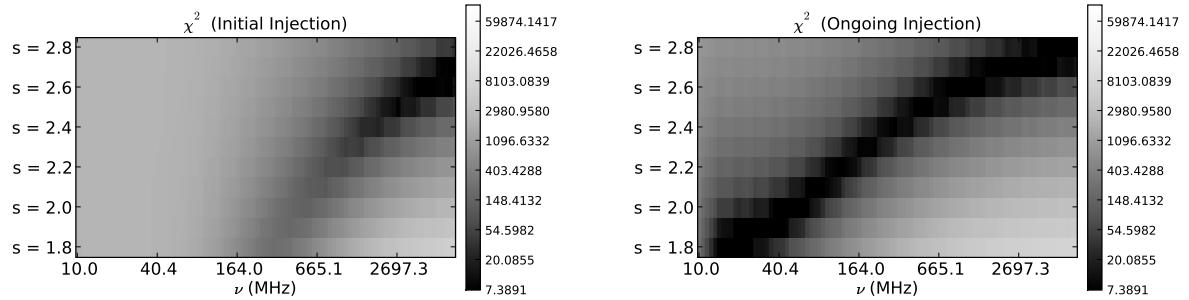


Figure 9.6: Spectral Fits : χ^2 as a function of s and ν for region I : (Left): Initial, (Right) Ongoing. Darker regions correspond to lower values of χ^2 . These plots show that for the initial-injection model, better fits are obtained for higher values of s and give ν_c values greater than 3.0 GHz. For the ongoing-injection model, all values of s between 1.8 and 2.8 give good fits with ν_c ranging between 10 MHz and 4 GHz. Note that below $s = 2.1$ and above $s = 2.5$ there is a higher uncertainty on ν_c (the widths of the darker regions increase for these values of s). This is because we are fitting the asymptotes by a spectrum consistent with a single power-law, and ν_c is almost unconstrained there.

Output : The results of these spectral fits is a value of ν_c for each value of s , for different features across the source. This is the critical frequency to be used to calculate the synchrotron age. Values of ν_c vs s were computed for the two ageing models described in section 9.2.2.2.

Error bars : The uncertainty on the best-fit value of ν_c was estimated *via* a Gaussian fit to the 1D χ^2 function (evaluated for several ν_c) in the neighbourhood of the minimum. For these data points and models, the average uncertainty on the best-fit ν_c was $\pm 30\%$.

Results : Figure 9.6 shows the χ^2 surface as a function of two variables s and ν_c for a subset of the region labeled as I in Fig. 9.5 and Figures 9.7 and 9.8 show the corresponding model spectra and data points.

1. Initial Injection model : The left panel of Fig. 9.6 shows χ^2 for the initial injection model and Fig. 9.7 shows the corresponding spectra plotted using the best-fit values of ν_c for ($s=2.0, 2.2, 2.4$ and 2.6). Both these figures show that lower values of χ^2 (< 10) are obtained only for $s > 2.3$ and give ν_c values between 1 and 8 GHz. This is because the five sampled frequencies do not show steepening consistent with

distribution for 3 degrees of freedom shows that there is a 50% probability of the reduced χ^2 being less than 0.8, a 10% probability of it being less than 0.2 and a 1% chance of it being greater than 3. Further, the true number of degrees of freedom for this problem lies between 1 and 3 since the three L-band data points are not independent (the 1.1 GHz and 1.8 GHz data points are derived from the 1.4 GHz values and the L-band spectral index). Therefore, these χ^2 values were used only to measure how the goodness of fit varies with s and ν_c . These trends were verified by doing a Kolmogorov-Smirnov test designed for a small sample set and this showed the same trends as χ^2 .

an exponential drop-off and therefore must all be below or near ν_c . Therefore it is not unexpected that better fits are obtained only when $\nu_c > 2\text{GHz}$ and all data points fall in the single-power-law region of the synchrotron spectrum where the average spectral index of -0.9 constrains the value of s to be about 2.8. Note however, that when ν_c does not lie within the sampled frequency range, any ν_c fits are based on extrapolated spectra and are more uncertain.

2. Ongoing Injection model : The right panel of Fig. 9.6 shows the χ^2 surface for the ongoing injection model and Fig. 9.8 shows the corresponding spectra plotted with best-fit ν_c values for $s=2.0, 2.2, 2.4$ and 2.6 . In this case, low values of χ^2 are obtained for all sampled values of s , suggesting that the spectral steepening is too gradual for these data points to constrain the model. However note that these fits show a basic trend of particles with a steeper particle energy distribution having higher best-fit ν_c values and hence shorter lifetimes (the particles require a shorter amount of time to steepen to the currently observed spectrum).

Figures 9.9 and 9.10 show χ^2 plots similar to Fig. 9.6 for 11 regions of the M87 halo. They show that for each value of s , steeper average spectra give lower best-fit ν_c values and the darker regions of these plots moves towards the top-left. Also, brighter regions have sharper χ^2 minima indicating slightly smaller error-bars on the best-fit values of ν_c . All plots show that steeper electron energy distributions take shorter amounts of time (higher best-fit ν_c) to reach the observed steepened spectra.

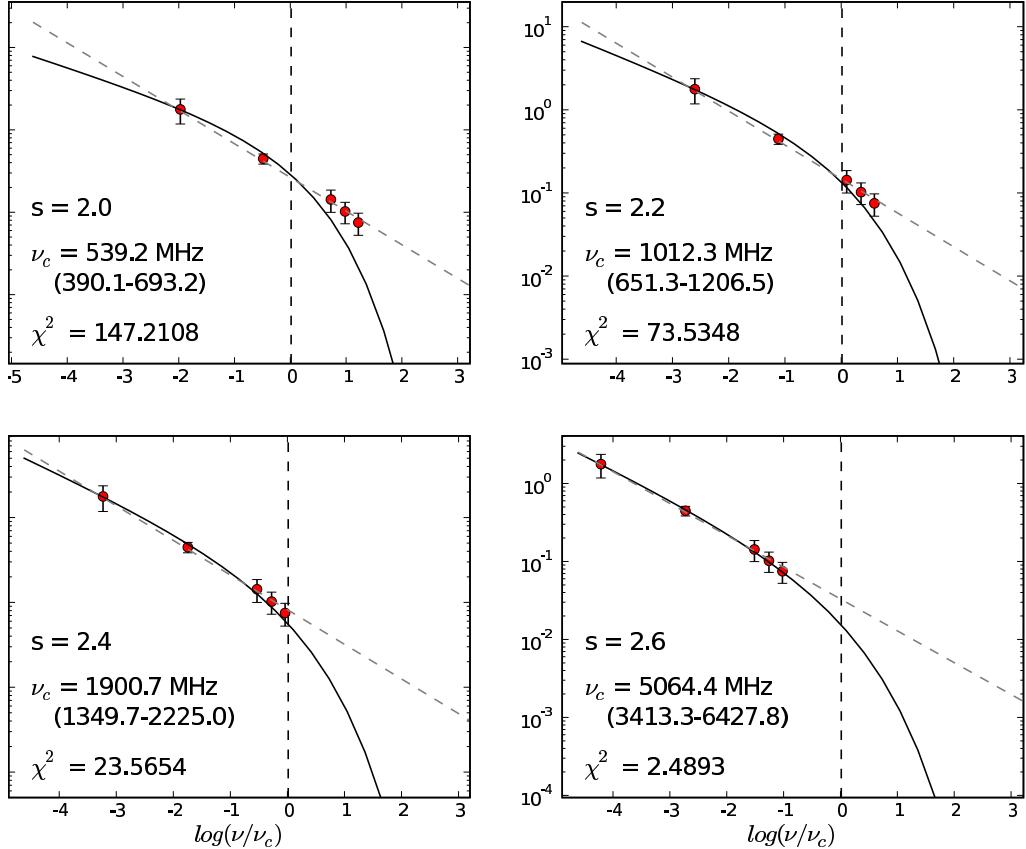


Figure 9.7: Spectral Fits - Initial Injection model : This plots shows the 5 data points (cicles) overlaid on four model spectra (solid lines) derived for four different values of $s = 2.0, 2.2, 2.4, 2.6$. The slanting dashed lines passing through the data points represent a single spectral-index fitted to all 5 data points ($\alpha = -0.93$). The vertical dashed line indicates the critical frequency ν_c , and all spectra have been shifted such that ν_c for all the fits are aligned. These data points were obtained from a subset of the region marked I in Figs.9.5 and 9.9. Values of χ^2 for these fits are shown in the left image in Figure.9.6 and show that higher values of s have better fits. This is because these data points are consistent with a power-law (single- α) and can only correspond to the below- ν_c regions of the synchrotron spectrum. The slight steepening seen in the three L-band points provides a strong constraint on ν_c (which also makes any ν_c fits highly dependent on the error in the measured L-band spectral index).

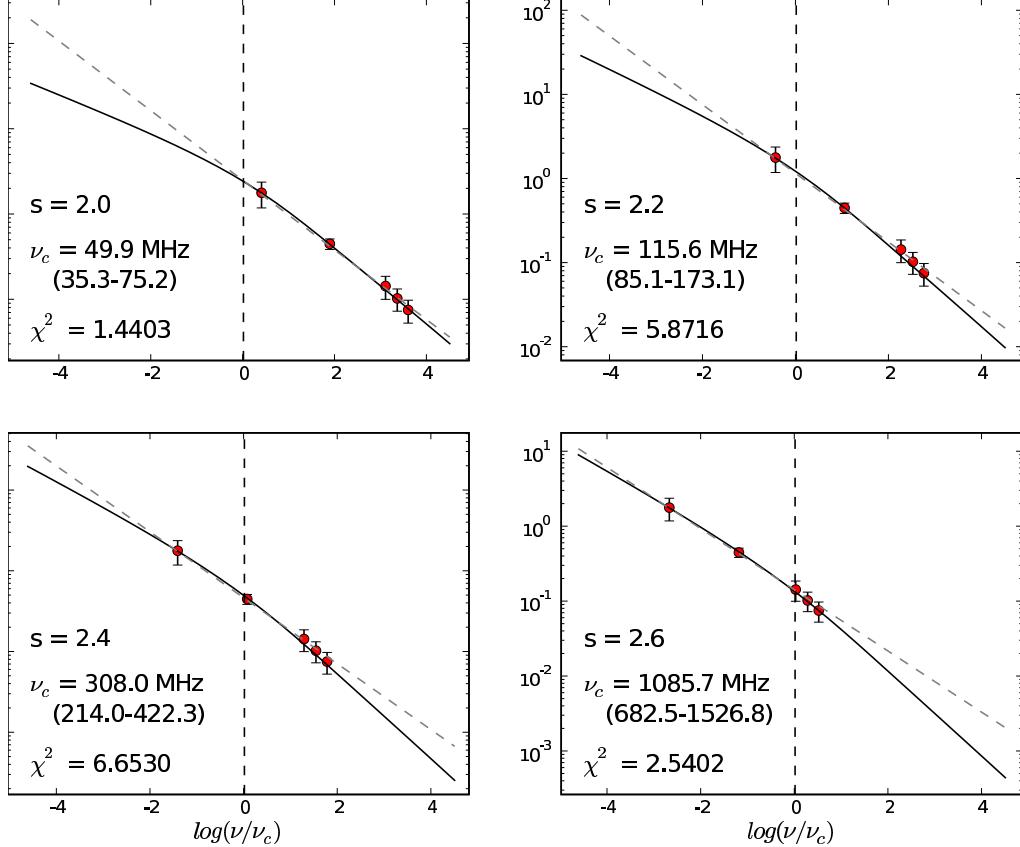


Figure 9.8: Spectral Fits - Ongoing Injection model : This plots shows the 5 data points (circles) overlaid on several model spectra (solid lines) derived for four different values of $s = 2.0, 2.2, 2.4, 2.6$. The dashed lines passing through each set of data points represent a single spectral-index fitted to all 5 data points ($\alpha = -0.93$). The vertical dashed line indicates the critical frequency ν_c , and all spectra have been shifted such that ν_c for all the fits are aligned. These data points were obtained from a subset of the region marked I in Figs.9.5 and 9.10. Values of χ^2 for these fits are shown in the right image in Figure.9.6. These fits show that these data do not constrain the value of ν_c or s for the ongoing injection model. This is because the models show a very slow steepening of the spectrum around ν_c and the data points are also consistent with a power-law (single- α) and show only a slight steepening across L-band.

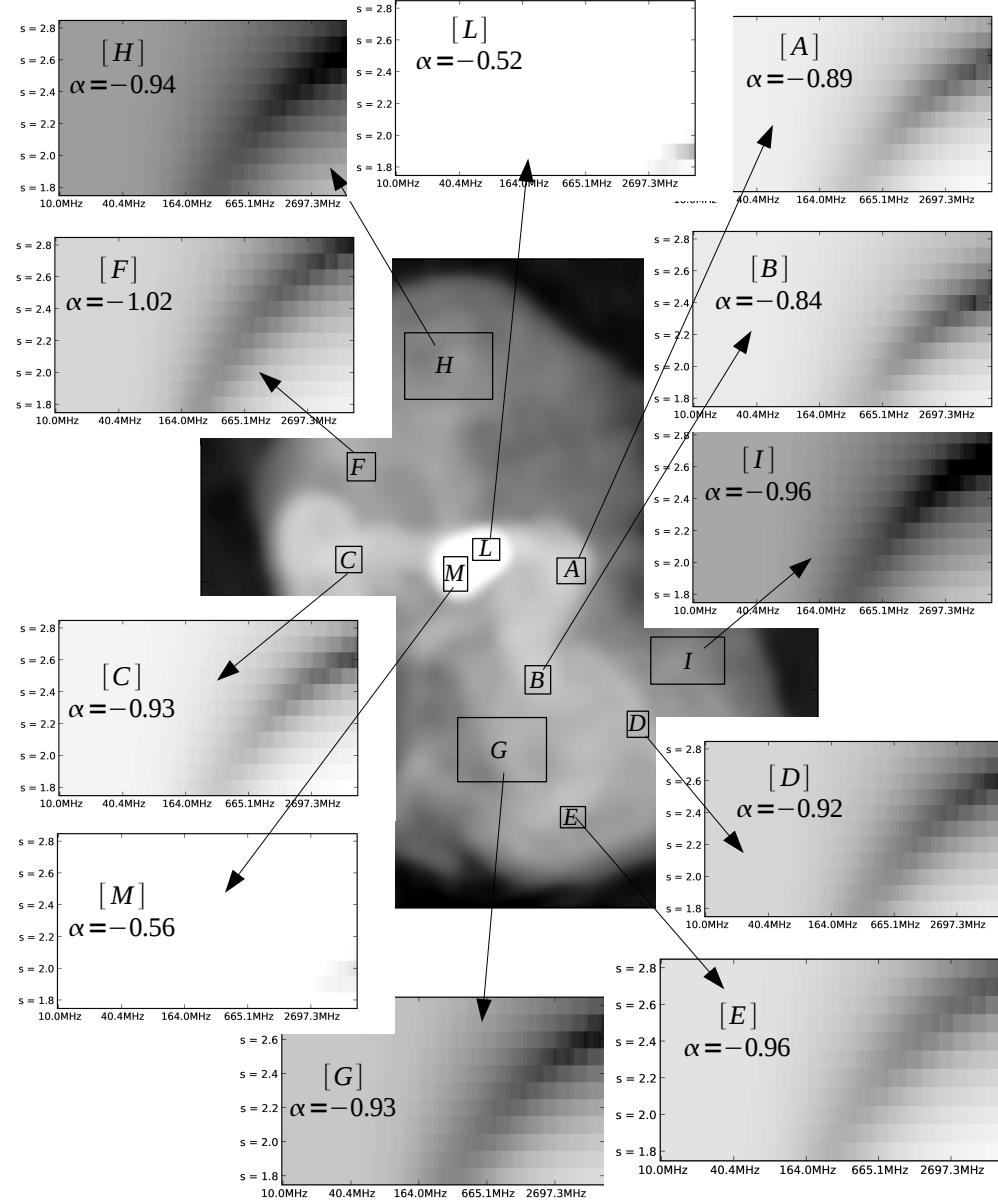


Figure 9.9: Initial Injection model - all over the source : These plots show the values of χ^2 as a function of s and ν_c . Darker regions correspond to lower χ^2 values. The central bright region does not fit the initial-injection model for any s between 1.8 and 2.8. For the rest of the halo, these data appear to rule out the initial-injection model for $s < 2.4$. These plots show that the initial injection model gives relatively good fits only for values of $s > 2.3$, and the corresponding best-fit critical frequencies lie above 2 GHz (consistent with low-resolution measurements that suggest steepening between 1 and 10 GHz). Regions with steeper spectra show a slight shift of the χ^2 minima towards higher ν_c values and steeper initial particle energy power laws.

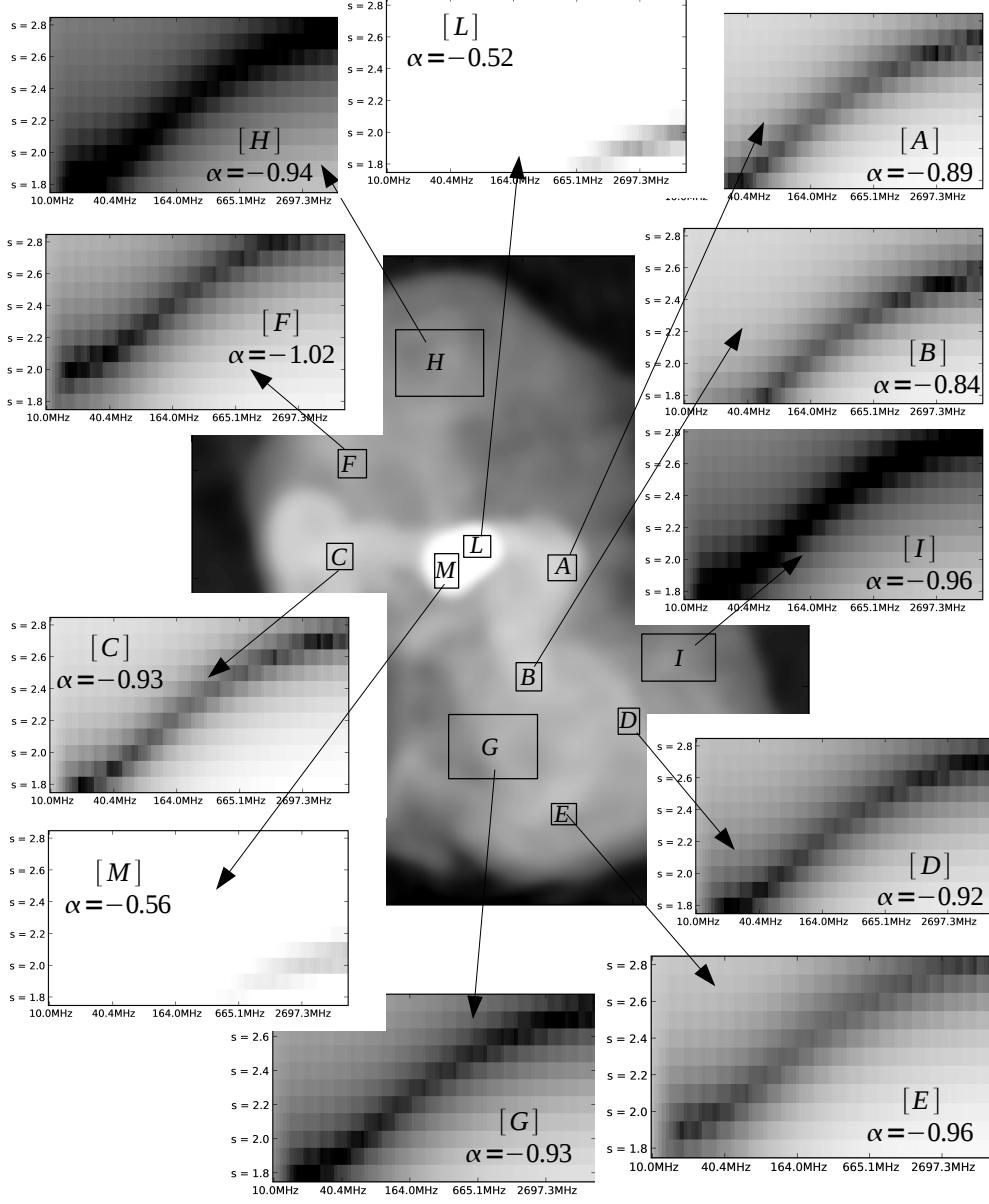


Figure 9.10: Ongoing Injection model - all over the source These plots show the values of χ^2 as a function of s and ν_c . Darker regions correspond to lower χ^2 values. The central bright region shows relatively good fits for $s = 2.0, 2.1$ and $\nu_c > 2\text{GHz}$, a result consistent with the idea of radio lobes being continuously fed by a jet with an injection index of 2.1 (and measured spectral index of -0.55). In the rest of the halo, all values of s between 1.8 and 2.8 give best-fit ν_c values with comparable absolute χ^2 values. This shows that with the current data, the ongoing injection model cannot be ruled out. Regions with steeper observed spectra show a slight shift of the χ^2 minima towards lower ν_c values (more ageing) and steeper injected spectra. However, the steepening across the spectral break as well as the measured spectrum are too gradual to be able to constrain both s and ν_c simultaneously.

9.3.3 Calculating Synchrotron lifetimes

Synchrotron ages of different features across the source were calculated using best-fit values of ν_c for spectral models with $s = 2.0$ and $s = 2.5$ for ongoing-injection and $s = 2.5$ for initial-injection. $s = 2.0$ was chosen because high angular-resolution wide-band observations of the M87 jet have shown a constant spectral index of -0.5 (corresponding to an injection index of $s = 2.0$). $s = 2.5$ was chosen for the the rest of the calculations, as it gave best-fit solutions for most regions in the halo for the initial-injection model (no good fits were obtained for $s < 2.4$ with initial-injection). The ongoing injection model gave valid fits for all tested values of s . Here, $s = 2.0$ and $s = 2.5$ are representative of the best-fit values in regions of the spectrum where we are fitting asymptotes, and they bracket the range of best-fit ν_c values allowed by this model.

Synchrotron ages were computed using both equipartition B-fields shown in Tables. 9.1 and 9.2 and the maximum average B-field ($B_{dyn} = 27\mu G$) given by the ambient pressure. Table 9.3 lists the synchrotron ages calculated using Eqn. 9.9 with $B_{now} = \langle B \rangle = B_{eq}$ to represent a homogeneous B-field seen by the particle throughout its lifetime. Table 9.4 lists synchrotron ages of filamentary structures treated separately from the diffuse background. Two sets of calculations were done using $\langle B \rangle = B_{eq}$ from Table 9.1 as background B-fields. The first used $B_{now} = B_{eq}$ from Table 9.2 for filament B-fields and and the second used $B_{now} = B_{dyn}$.

The main trends shown by these numbers are

1. The inner radio lobes (regions L,M) give $t_{syn} = 3 \sim 5$ Myr for ongoing injection with $s = 2.0$ (with both B_{eq} and B_{dyn}).
2. With equipartition B-fields, the ear lobe/canal (regions A,B,C) give $t_{syn} \approx 20$ Myr for initial injection and $s = 2.5$, and $t_{syn} = 30 \sim 200$ Myr for ongoing injection ($2.0 \leq s \leq 2.5$). With $B_{dyn} = 27\mu G$ these ages are ~ 5 times smaller.
3. With equipartition B-fields, the halo (regions D through I) give $t_{syn} = 40 \sim 70$ Myr for initial injection and $s = 2.5$, and $t_{syn} = 90 \sim 800$ Myr for ongoing injection ($2.0 \leq s \leq 2.5$). With $B_{dyn} = 27\mu G$ these ages are ~ 8 times smaller.
4. For the filaments, we get $t_{syn} \approx 100$ Myr for initial injection and $s = 2.5$ and $t_{syn} = 100 \sim 1000$ Myr for ongoing injection.

For comparison, timescales obtained from direct dynamics (for 40 kpc) include $t_{buoyant} \approx 60$ Myr from a buoyant bubble model [Churazov et al. 2001], $t_{driven} = 50 \sim 120$ Myr from a driven expanding bubble model with $\dot{E} = 10^{44} \sim 10^{45}$ erg/sec and $n_x = 0.01$ [Owen et al. 2000] and $t_{sound} \approx 100$ Myr from the local sound speed. Timescales from low-resolution wide-band spectra [Rottmann et al. 1996b] are 30 to 40 Myr for the ear lobe/canal regions (regions A,B,C, 20 kpc scale).

	D [kpc]	ν_c [MHz] Ongoing s=2.0	$t_{syn}^{B_{eq}}$ [Myr]	$t_{syn}^{B_{dyn}}$ [Myr]	ν_c [MHz] Ongoing s=2.5	$t_{syn}^{B_{eq}}$ [Myr]	$t_{syn}^{B_{dyn}}$ [Myr]	ν_c [MHz] Initial s=2.5	$t_{syn}^{B_{eq}}$ [Myr]	$t_{syn}^{B_{dyn}}$ [Myr]
L	5 (10)	7100	3.2 (4.3)	4.3						
M	5 (10)	5000	3.9 (5.2)	5.1						
A	20 (40)	100	150 (190)	35	1300	42 (54)	10	6200	19 (25)	4.6
B	20 (40)	190	140 (175)	26	2900	35 (44)	7	7100	22 (28)	4.3
C	20 (40)	66	190 (250)	44	820	54 (70)	13	4400	23 (30)	5.5
D	40	81	300	40	940	90	11	4700	40	5.3
E	40	57	300	48	710	85	13	3600	37	6.1
F	40	17	600	87	310	140	20	2300	52	7.5
G	40	70	360	43	820	100	12	4100	46	5.7
H	40	66	460	44	820	130	12	4100	58	5.7
I	40	21	800	78	440	180	17	2700	71	7.0

Table 9.3: Synchrotron lifetimes : This table lists the synchrotron lifetimes calculated using the best-fit critical frequencies for s=2.0 and s=2.5 for the ongoing-injection model and for s=2.5 for the initial-injection model. The uncertainty on the fitted ν_c values is about $\pm 30\%$ which gives an uncertainty of $\pm 15\%$ on the synchrotron lifetime.

	ν_c [MHz] Ongoing s=2.0	$t_{syn}^{B_{eq}}$ [Myr]	$t_{syn}^{B_{dyn}}$ [Myr]	ν_c [MHz] Ongoing s=2.5	$t_{syn}^{B_{eq}}$ [Myr]	$t_{syn}^{B_{dyn}}$ [Myr]	ν_c [MHz] Initial s=2.5	$t_{syn}^{B_{eq}}$ [Myr]	$t_{syn}^{B_{dyn}}$ [Myr]
A-I	140	630	700	1800	180	200	7200	88	100
A-H	110	760	850	1400	210	240	6300	100	110
B-I	280	420	500	7200	84	99	7200	88	99
B-H	230	480	600	4700	100	130	7200	87	100
C-I	87	790	900	1000	230	260	5000	100	120
C-H	66	1000	1100	820	290	320	4400	120	140
D-I	130	570	730	1800	160	200	7200	78	100
D-H	87	740	970	1000	210	280	4700	100	130
E-I	75	800	970	820	250	290	4400	100	139
E-H	61	970	1200	760	280	330	3800	120	150

Table 9.4: Synchrotron lifetimes for filaments : This table lists the synchrotron lifetimes calculated using the best-fit critical frequencies for s=2.0 and s=2.5 for the ongoing-injection model and for s=2.5 for the initial-injection model. The spectral data used for these fits were computed as the difference between the filament and background intensities. Two sets of calculations were done using the the equipartition field calculated for the background (regions I and H) as $\langle B \rangle$ in Eqn. 9.9. The first used filament B-fields from Table 9.2 and the second used B_{dyn} as the filament B-field. Here too, the uncertainty on the fitted ν_c values is about $\pm 30\%$ which gives an uncertainty of $\pm 15\%$ on the synchrotron lifetime.

9.4 Interpretation

This section discusses whether or not any of the synchrotron ageing models fit the data, whether or not the synchrotron ages are consistent with other age estimates, and what these results (and better measurements) could tell us about the synchrotron processes at play within the M87 radio halo.

9.4.1 Do these ageing models fit ?

9.4.1.1 Core / Jet / Inner lobes

For the bright central region (labeled as L and M) consisting of the core, the 2 kpc jet and inner radio lobes <5 kpc from the core, the ongoing-injection model fits well for $s \approx 2.0$, giving a best-fit $\nu_c > 5\text{GHz}$ and an age of < 5 Myr for the inner-lobes. No valid fits were obtained for the initial injection model with s between 1.8 and 2.8, or for the ongoing injection model with $s > 2.2$. Note that a synchrotron age of ≈ 5 Myr is smaller than the timescale of 17 Myr derived from the sound-speed across 5 kpc (using $T=10^7\text{K}$, derived from $P = 14.5 \times 10^{-11}\text{dyn/cm}^2$ at a distance of $\sim 5\text{kpc}$ from the core [[Owen et al. 2000](#)]), but is consistent with a $2 \sim 4$ Myr dynamic expansion time calculated for a driven bubble (Eqn. 9.1) over a distance of 5 kpc with $\dot{E} \approx 10^{44}\text{ergs/sec}$. Also, within this region, the equipartition B-fields are similar to the equivalent B-field that balances the external pressure and gives similar timescales.

An injection index of $s = 2.0$ for the M87 jet is consistent with $\alpha_{jet} \approx -0.5$ as known from high resolution observations of the M87 jet ([Owen, private commn.](#)). Also, [Perlman and Wilson \[2005\]](#) show that the broad-band spectrum of the M87 jet (radio to X-rays) is consistent with a continuous injection index of $s = 2.2$, and the critical frequency estimated from measurements of the jet spectrum between radio, optical and X-ray bands is at about 100 THz (infrared).

9.4.1.2 Halo : Initial Injection model

The simplest spectral evolution model for regions outside the bright central region is the initial-injection model in which energetic particles are produced in the jet and the travel outwards in the form of buoyant or expanding bubbles and age *via* synchrotron radiation with no additional sources of energy.

Outside the central bright region, the data and spectral fits rule out all values of $s < 2.4$ for the initial-injection model. The model spectra predicted for $s \leq 2.4$ have below- ν_c spectral indices of -0.7 (and less) that are flatter than the average spectral index of -0.9 measured between 75 MHz and 1.8 GHz. Also, the initial-injection model predicts significant curvature even in the sampled frequency range (75 MHz to 1.8 GHz), and the lack of such curvature is a strong indicator even without measurements between 1 and 10

GHz. This means that if $s=2.0$ is the only possible source, something is preventing the higher-energy electrons from cooling and steepening the spectrum and this system cannot follow the initial-injection model.

Better spectral fits were obtained for $s = 2.5$ and above, leading to a ν_c of between 1 and 10 GHz. These numbers are consistent with low-resolution measurements from [Rottmann et al. \[1996a\]](#) that show significant steepening between 1 and 10 GHz, and numerical models from [Churazov et al. \[2001\]](#) that predict an average spectral index of -1.0 below 1 GHz and a drop off beyond 5 GHz.

These critical frequencies give synchrotron ages or 20 to 30 Myr for regions A,B,C and about 35 to 70 Myr for regions in the outer halo. These ages are computed from equipartition B-fields (timescales of 5 to 7 Myr are obtained using the maximum B-field derived from arguments of pressure balance with the ICM). For comparison, sound speed calculations (from $T \sim 10^7 K$, [\[Shibata et al. 2001\]](#)) give timescales of 70 Myr and 140 Myr for 20 kpc and 40 kpc respectively. Also, expansion timescales for a driven bubble are 16 Myr and 53 Myr for 20 kpc and 40 kpc respectively, with $\dot{E} \sim 10^{44}$ ergs/sec and $n_x = 0.01$. The buoyant bubble simulations of [Churazov et al. \[2001\]](#) suggest that a distance of 40 kpc can be reached in 67 Myr.

These timescales match within their uncertainties, but the biggest discrepancy in these results is that the jet has an observed injection index of -0.5, corresponding to $s = 2.0$, but outside the central bright region it is clearly not possible to fit the data with $s = 2.0$ and the initial injection model. However, if we consider the 'ear-lobe/canal' and structures in the outer halo to have formed from a previous cycle of AGN activity, there is no reason for the previous injection spectrum to have been $s = 2.0$. If it had a steeper injected spectrum and a low B-field ($\sim 7\mu G$, similar to the computed equipartition fields), the initial-injection model gives plausible ages. Further, an age difference of ~ 100 Myr between the inner radio lobes and the outer halo could further suggest a 100 Myr duty cycle of AGN activity.

Finally, note that the observed spectra are nearly consistent with a pure power-law and only the L-band spectrum shows slight steepening (comparable to the size of the per-pixel error-bars). Therefore, all these spectral fits are constrained largely by the current L-band spectral index map (which contains the effect of deconvolution errors and low signal-to-noise of the halo emission). Also, these fits work *only* for ν_c greater than any observed frequency. Therefore one can only obtain a lower limit on ν_c , and therefore, a upper-limit on the associated synchrotron lifetimes.

However, these data do suggest a ν_c of a few GHz, and further observations at C-band (4.8 GHz) and higher are required to see whether the observed power law continues, or a turn-over followed by an exponential drop-off is observed.

9.4.1.3 Halo : Ongoing Injection model

The ongoing-injection model applies only to regions that are continuously fed by an energy source, or to regions where there is some local form of particle injection. Out in the halo, a continuous particle injection is an unlikely scenario, but particles may be locally re-energised by scattering off turbulent Alfvén waves in an inhomogeneous B-field (for example).

Outside the bright central region, all values of $s = 1.8 \sim 2.8$ give good fits with the ongoing-injection model with v_c ranging all the way from 30 MHz to 6 GHz and give synchrotron lifetimes ranging from 90 Myr to 800 Myr. These timescales range from bubble expansion and buoyancy timescales, to values comparable with the expected cooling time $t_{cool} \approx 1$ Gyr.

One interpretation of having such a wide range of valid solutions is that the synchrotron evolution model does not follow a continuous particle injection model with a fixed injection index, and other processes such as B-field inhomogeneities may be at play. However, the most likely reason for these multiple solutions is that all the spectra are consistent with pure power-laws and these fits have a high degree of uncertainty. Values of $s < 2.1$ and > 2.3 give better fits because the below- v_c and above- v_c power-laws match the observed power-law spectra ($\alpha \approx -0.9$ matches the spectrum for $s = 2.8$ below v_c and for $s = 1.8$ above v_c). Also, since in these regions we are fitting asymptotes, v_c is not well constrained, and only upper and lower limits can be obtained. Further, the predicted curvature across the break is very gradual, and spectra that are consistent with a pure power-law (within error-bars) give reasonable spectral fits even across the region of curvature although these fits have higher χ^2 values than fits to the asymptotes. However, note that in general, the absolute χ^2 values obtained with the ongoing-injection model were consistently lower than those obtained with the initial-injection model (most likely the result of large error-bars).

9.4.1.4 Filaments

The apparent correlation between structures seen in the radio and X-ray in the 'ear-lobe' and 'ear-canal' regions suggests some form of local activity that might contribute to the transfer of energy between the radio plasma and the surrounding thermal ICM. Also, the compact filamentary structure seen throughout the halo suggests regions of high B-fields and possible sites of local particle re-energizing. To check if either of these models apply, we need to isolate the filaments from the diffuse background and analyse them separately.

Ages derived using Eqn. 9.9 for ongoing-injection in filaments give timescales of 0.5 to 1.0 Gyr for $s = 2.0$ (again, comparable to $t_{cool} \approx 1$ Gyr) and 0.1 to 0.3 Gyr for $s = 2.5$. The timescales calculated for the filaments are consistently larger than those computed with the total observed intensity, an effect expected for particles moving from lower B-field regions to higher B-fields regions from where they are currently radiating (Eqn. 9.9 for

inhomogeneous B-fields). This model and the obtained timescales may imply the presence of structures (with high B-fields) that are perhaps persistant across cycles of AGN activity and produce high frequency synchrotron radiation when particles move into them. Instead of (or in addition to) increased B-fields, these regions could also be sites of *in-situ* particle re-acceleration where the fraction of high-energy particles is increased (note that in this case, the spectral shape is likely to differ from the ongoing injection model). Alternatively, these large t_{syn} values could be the result of over-estimating the B-fields in the filaments or under-estimating the average background B-field (*i.e.* if equipartition does not hold). Therefore, these data do not rule out the possibility of these filaments being isolated sites of activity (possibly with high B-fields) other than simple ageing of particles with an initial energy spectrum. Also, timescales obtained with the initial-injection model and $s = 2.5$ are ~ 100 Myr, which is still comparable to the dynamic age of the outer halo. This suggests that these filaments are also consistent with spatially compact regions with high B-fields compared to the surrounding, passively moving through the halo as it expands.

To probe these ideas further and ascertain whether there is any significant difference between the filaments and their surroundings, we need to isolate filament and background spectra more accurately, especially in the frequency range of 1 to 10 GHz where there should be a measureable difference if these filaments do represents local sites of particle re-energising. If a significant difference in the spectral shape is measured between structures in the halo and regions in the ear lobe/canal where increased X-ray emission is present, it may give evidence for the ear lobe/canal regions to be sites of local energetic activity and energy transfer between the radio plasma and the ICM.

9.4.2 Conclusions and Future Work

Spectra in the inner few kpc (the lobes immediately around the jet) are consistent with an ongoing injection of particles with the energy distribution as seen in the jet $s \approx 2.0$, and a synchrotron age of ≈ 5 Myr which is also consistent with dynamical estimates. For features in the halo (filaments, background and large-area averages), ages consistent with expansion and buoyancy timescales (~ 20 Myr for regions A,B and C, and $40 \sim 70$ Myr for the halo) can be obtained with the initial injection model of synchrotron ageing with $s \approx 2.5$. These data appear to reject all initial-injection fits for $s < 2.4$, suggesting that if this model were to apply, the radiating particles need to have originated from perhaps a previous cycle of AGN activity in which the injected energy distribution had a steeper $N(\gamma)$. There is also a slight hint of spectral steepening from the inner regions to the outer halo, but these variations are within the calculated uncertainties and need better measurements and imaging (across L-band) to confirm. Outside the inner radio lobes, the ongoing injection model gives plausible solutions for a wide range of s (1.8 to 2.8), showing that the spectral data used for these fits are unable to constrain the model. However, this model cannot be ruled out, and more sensitive observations are required in order to ascertain whether the predicted shallow steepening is present or not.

The above results can be combined to suggest that the inner radio lobes and the 40 kpc halo may have originated from two different cycles of AGN activity (one with $s = 2.5$ and one with $s = 2.0$) and possibly separated by ~ 100 Myr. The inner radio lobes are continuously being fed by particles from the jet, whereas the much larger structures are the result of passively ageing particles. The only parts of the halo where something other than passive synchrotron aging may be happening are the bright filaments. Timescales of $100 \sim 200$ Myr are obtained with the ongoing-injection model ($s = 2.5$) for bright filaments in regions A,B and C (where the X-ray emission appears to be correlated with the radio). These timescales are up to a factor of 2 larger than dynamical estimates, and correspond to particles radiating from high B-field regions. These B-fields are comparable to the maximum possible field derived from pressure-balance with the surrounding ICM, and could signal regions with inhomogeneous B-fields and local energetic activity that may contribute to the transfer of energy between the radio halo and the ICM.

9.4.2.1 Future observations

To take the ideas discussed above to their logical conclusions, further observations are required to (a) probe the high-angular-resolution structure of the halo at frequencies above 2 GHz and (b) produce high dynamic-range spectral information to treat filaments separately from the diffuse background.

With real EVLA data at L-Band it is expected that spatio-spectral deconvolution errors will reduce, making the L-band spectral index map more reliable. The EVLA D-configuration *uv*-coverage is required for sensitivity to large spatial scales (diffuse halo), and C and B configurations will provide the required angular resolution to isolate filaments from the background.

Measurements at 4.8 GHz and higher are required to test whether the slight steepening suggested by the current L-band data is real or not, and if it is, whether there is a sharp drop-off in flux between 1 and 10 GHz at small spatial scales (similar to that observed from low-resolution images), or whether the entire halo or parts of it show flatter spectra. Such observations with the EVLA C-band (in D-configuration) will require a mosaic observation with wide-band primary-beam correction, and perhaps single-dish observations to fill in the short spacing flux.

CHAPTER 10

CONCLUSION

In accordance with the goals of this dissertation outlined in chapter 1, a general purpose multi-scale multi-frequency deconvolution algorithm (MS-MFS) was developed for use in broad-band radio interferometry, and then applied to multi-frequency VLA observations of the M87 radio galaxy to study the observed broad-band spectra of various features in its radio halo. Section 10.1 summarizes the work done to develop the MS-MFS algorithm with its current capabilities, points out the requirement for tests using real wide-band data, and lists a few topics of future research in wide-band image reconstruction. Section 10.2 summarizes the results obtained from a high angular resolution study of the broad-band spectrum of the M87 radio halo and suggests future observations required to take the next step.

10.1 Wide-band image reconstruction

Summary : The first step of this project was to evaluate the applicability of existing wide-band image reconstruction methods to data from broad-band interferometers and identify areas that required algorithmic improvements. Tests on simulated EVLA data showed that the existing multi-frequency synthesis methods are adequate for narrow-field imaging of isolated point sources with pure power-law spectra, but inadequate for sources with extended emission or spectra that are not pure power-laws. These tests also showed that when the single-frequency *uv*-coverage of the interferometer is sufficient to unambiguously reconstruct the spatial structure of the source, a simple hybrid of single-channel imaging and multi-frequency synthesis could potentially deliver required image dynamic ranges on the continuum image. However, spectral information would still be a by-product and available only at the angular resolution of the lowest frequency in the band.

Based on the results from the above tests, the next step was to develop a new multi-frequency synthesis algorithm that combined multi-scale deconvolution techniques along with a spectral model capable of representing arbitrary but smooth spectral shapes. For wide-field imaging, methods to model the frequency dependence of the primary beam and correct for it during multi-frequency synthesis and deconvolution were also developed.

In order to understand the details involved in formulating and implementing such algorithms, it became necessary to work out and describe the basic numerical optimization framework used in most established calibration and imaging algorithms in radio interferometry. Recently developed algorithms that correct for direction-dependent instrumental

effects, perform multi-scale deconvolution and multi-frequency synthesis imaging were also described in this framework in order to clarify the connections between all these methods and show how they could be extended individually and also combined into a practical implementation. An analysis of the existing multi-scale and multi-frequency deconvolution algorithms in this framework led to ideas for demonstrable improvements in both the algorithms.

The resulting MS-MFS algorithm parameterizes the 2-D sky brightness distribution using a multi-scale basis and describes the spectrum per pixel as a polynomial. The data products are a set of coefficient images describing this polynomial for each pixel, and images of the continuum emission, spectral index and spectral curvature can be derived from them. The MS-MFS algorithm improves upon existing wide-band imaging methods in the following ways (a) a multi-scale parameterization suited to both compact and extended emission, (b) a flexible spectral model to allow arbitrary spectral shapes including partially band-limited signals (c) the use of *a-priori* information about synchrotron spectra to reconstruct spectral structure at the angular resolution allowed by the highest frequency in the band, and (d) a method to model the frequency-dependence of the antenna primary beam and to evaluate and use this model within the image-reconstruction process. The MS-MFS algorithm was implemented within the CASA and ASKAPsoft data analysis packages.

Since the MS-MFS algorithm was developed and implemented before real wide-band data from the EVLA was available, all algorithm validation tests were performed either on simulated wide-band EVLA data or data from multi-frequency VLA observations between 1 and 2 GHz (taken as a series of narrow-band snapshot observations). The algorithm was tested on sources with spectral structure on multiple spatial scales, moderately-resolved sources with power-law spectra, overlapping sources with different spectra, sources with band-limited emission and sources with broad-band emission over wide fields-of-view. These tests have shown satisfactory results in terms of dynamic range and accuracy. Further tests of both the MS-MFS and the simpler hybrid algorithm using real wide-band EVLA data would help in order to quantify errors and establish a general-use data analysis path.

Future work : This new generation of broad-band interferometers has opened up a wide range of astrophysical opportunities that will require further algorithm research and development. For example, the use of wide-band data for full-polarization high dynamic-range imaging will have to take into account the effects of frequency-dependent source and instrumental polarization, and it is not clear whether the spatial and spectral flux models used in the MS-MFS algorithm are appropriate for wide-band Stokes Q,U and V imaging. Then, the possibility of combining recently developed rotation-measure synthesis with wide-band imaging is also worth exploring from the point of view of simultaneously obtaining accurate spatial and spectral reconstructions and therefore increasing the fidelity of the results.

Even for Stokes I imaging, other algorithms must be explored to address areas where the MS-MFS formalism may not be the best choice. High dynamic-range wide-band imaging simulations have shown that the algorithm is currently limited by its choice of multi-scale image parameterization. Therefore, wide-band extensions of algorithms like ASP-CLEAN are worth exploring in combination with more advanced numerical optimization techniques. An initial investigation into such an approach has shown very promising results (not included as part of this dissertation) and must be taken to its logical conclusion. Wide-band primary-beam correction with the MS-MFS algorithm has shown good results only within the main lobe of the primary beam at the highest frequency (about the HPBW at the lower end of a 2:1 bandwidth). A careful evaluation of the involved errors must be carried out for fields-of-view beyond this limit, at least in the context of accurate model prediction for wide-band mosaicing applications.

Finally, the benefits of using broad-band receivers are the greatest when the narrow-band spatial-frequency coverage of the imaging interferometer is too sparse to be useful on its own, or if the source of emission is time-variable and synthesis observations cannot be spread out in time. VLBI imaging is one such area where a wide-band imaging algorithm that reconstructs both spatial and spectral structure simultaneously from incomplete measurements could yield significant improvements over conventional techniques. Wide-band image reconstruction applied to sources whose time-varying spatial and spectral structure is of astrophysical interest is another area which could benefit from such algorithms.

10.2 The spectral evolution of M87

Summary : The MS-MFS algorithm developed in the first part of this dissertation project was applied to data from multi-frequency VLA observations of the M87 cluster-center radio galaxy between 1.1 and 1.8 GHz in order to complement existing low-frequency measurements of the broad-band spectrum of various features in its 40 kpc halo. The resulting spectra were compared with a set of model spectra derived from two different spectral evolution models. Best-fit break frequencies were estimated and synchrotron ages were calculated and interpreted in the context of dynamical evolution models and their timescales for various features observed in the M87 radio halo.

A spectral index map constructed from multi-frequency L-band observations of the M87 radio halo was combined with existing images at 75 MHz, 327 MHz and 1.4 GHz in order to constrain the slope of the broad-band spectrum at the upper end of the sampled range. These wide-band spectra were then compared with spectra obtained from two different synchrotron evolution models, one representing the passive ageing of a set of energetic particles with an initial power-law distribution of energies, and the other representing a continuous injection of energy either by a continuous flow or some reheating mechanism. A series of spectral fits were performed to estimate break frequencies and synchrotron ages

for both spectral models and various features across the radio halo.

The main results of this study are as follows. Spectra in the central bright region corresponding to the active 2 kpc jet and inner radio lobes ($<5\text{kpc}$ from the core) are consistent with a continuous ongoing injection of particles with an energy power-law index of $s \approx 2.0$ and give ages of $< 5\text{MYr}$, all of which are consistent with existing high angular resolution measurements of the jet spectrum. Elsewhere in the halo, the data fit the initial injection model only for $s \geq 2.4$ and give synchrotron ages consistent with dynamic estimates. These results suggest that the inner lobes are currently being powered by a jet outflow with an injection index of $s \approx 2.0$, and the outer halo is the remnant of a previous cycle of AGN activity in which the injection index of the particles was $s \geq 2.4$ and not the currently observed $s = 2.0$. The spectra of a few bright filamentary structures were then analysed separately from the diffuse background, as being regions of either high B-fields or local energetic activity in which the particles have spent only a small fraction of their lifetime. This analysis yielded filament B-fields stronger than the surroundings, and synchrotron ages longer than dynamical estimates of the halo age, suggesting that these high B-field regions are either sites of local energetic activity (other than ongoing-injection) or high B-field regions persistent across cycles of AGN activity. Finally, outside the inner radio lobes, all spectra yield acceptable fits for the ongoing-injection model for $1.8 < s < 2.8$, suggesting that this model also works but these data are unable to adequately constrain it.

Future Work : Further observations are required to confirm these results because most of the spectral fits gave break frequencies above 2GHz where there were no data points. Also, the spectral fits were strongly influenced by the measured slope at L-band, which given the error bars were also consistent with no steepening. A true EVLA wide-band observation of the M87 radio halo at L-band (1-2 GHz) is expected to reduce deconvolution errors and therefore improve the quality of the slope constraints.

Also, high angular-resolution observations at C-band (4-8 GHz, D-configuration) are required in order to confirm whether or not the halo shows an exponential drop-off between 2 and 10 GHz as suggested by these data and shown by low-resolution observations, and to also see if there is any difference between the spectra of the bright filaments and the background between 2 and 8 GHz. At C-band, the field-of-view of an EVLA antenna is a few times smaller than the angular extent of the M87 radio halo and this measurement would require a mosaic observation. The MS-MFS algorithm with wide-band primary-beam correction can in principle be used for wide-band mosaicing, but this has not been demonstrated yet. Therefore, a wide-band mosaic of M87 at C-band would (a) probe the halo spectra at high resolution and help confirm or reject the initial-injection model of spectral evolution and (b) serve as a good target field on which to test the MS-MFS algorithm for mosaicing and establish a data analysis path for future such observations.

APPENDIX A

IMAGING SENSITIVITY

The sensitivity (lowest detectable flux above the noise) for the output from a single baseline (or, the noise per visibility value) is given by

$$\Delta V_{ij} = \frac{1}{\eta_s} \frac{SEFD}{\sqrt{2\Delta\nu\tau_{acc}}} \quad (\text{A.1})$$

where $SEFD = 10^{23}T_{sys}(2k_B/\eta_a A)$ Jy is the *Source-Equivalent Flux Density*, defined as the source flux density that effectively doubles the T_{sys} of the receivers. τ_{acc} is the per visibility integration time in *seconds*, $\Delta\nu$ is the channel bandwidth in *Hertz*, A is the collecting area of an antenna in cm^2 , η_a is the antenna efficiency, η_s is the system efficiency and T_{sys} is the antenna system temperature in K .

The image sensitivity for a single channel stokes I image (using RR and LL data) is given as

$$\Delta I_m = \frac{1}{\sqrt{2}\eta_s} \frac{SEFD}{\sqrt{N(N-1)\Delta\nu\tau_{int}}} \quad (\text{A.2})$$

where τ_{int} is the total integration time in *seconds*. The factor of $\sqrt{2}$ in the denominator is for the two independant data channels (RR and LL). The image sensitivity for a multi-frequency image over N_{ch} channels is given as $\Delta I_m / \sqrt{N_{ch}}$.

In terms of eqn A.1, this is equal to the numerical estimate based on the number of data points as given by

$$\Delta I_m = \frac{\Delta V_{ij}}{\sqrt{\frac{N(N-1)}{2} N_{ch} N_t N_{pol}}} \quad (\text{A.3})$$

where $N_t = \tau_{int}/\tau_{acc}$ is the number of timesteps and $N_{pol} = 2$ for a stokes I image that uses the RR and LL polarizations.

For example, for $T_{sys} = 35K, \eta_a = 0.55, \eta_s = 0.78, N = 27, A = \pi(1250^2)cm^2, \tau_{int} = 8hr, \tau_{acc} = 300sec, \Delta\nu = 10MHz, N_{ch} = 128$ and $N_{pol} = 2$.

```

SEFD : 357.803  Jy
Total effective bandwidth : 1280  MHz
Noise per visibility : 5.92e-03  Jy
Image Sensitivity : 2.01e-06  Jy

```

APPENDIX B

LINEAR LEAST SQUARES

Measurement Equations : An instrument that measures a physical quantity can often be characterized as a system of linear equations. Consider a set of n measurements of a physical quantity that is described by a list of m parameters. Let $[A]$ be a measurement matrix of shape $n \times m$, \vec{x} a vector of parameters of shape $m \times 1$ describing the physical quantity, and \vec{b} the data vector of shape $n \times 1$.

$$[A]\vec{x} = \vec{b} \quad (\text{B.1})$$

The matrix $[A]$ describes how the physical quantity \vec{x} gets modified by the instrument during the measurement process, and is called the *transfer function* of the instrument. The effect of this transfer function has to be removed from the measured values \vec{b} to recover a true estimate of \vec{x} . This corresponds to finding $[A^{-1}]$ and computing $\vec{x} = [A^{-1}]\vec{b}$.

If the matrix $[A]$ is square, positive definite and of full rank and there is no noise in the measurement, then an exact $[A^{-1}]$ exists and \vec{x} can be exactly computed. In a real system, there is usually noise associated with the measurement and $[A]$ may not be positive definite by construction. In both these cases, an exact $[A^{-1}]$ will not exist, but a pseudo inverse $[A^+]$ can be calculated *via* χ^2 minimization designed to give a (weighted) linear least-squares estimate for \vec{x} .

Normal Equations : Let $[W]$ be an $n \times n$ diagonal weight matrix associated with the measurement noise in $\vec{b}_{n \times 1}$.

$$\chi^2 = ([A]\vec{x} - \vec{b})^\dagger [W] ([A]\vec{x} - \vec{b}) \quad (\text{B.2})$$

$$= \vec{x}^\dagger [A^\dagger W A] \vec{x} + \vec{b}^\dagger [W] \vec{b} - \vec{x}^\dagger [A^\dagger W] \vec{b} - \vec{b}^\dagger [W A] \vec{x} \quad (\text{B.3})$$

To minimize χ^2 , take its derivative w.r.to the parameters \vec{x}^\dagger and solve $\nabla \chi^2 = 0$. Note that $\left[\nabla_{\vec{x}^\dagger}\right] \vec{x} = 0$.

$$\nabla \chi^2 = [A^\dagger W A] \vec{x} - [A^\dagger W] \vec{b} \equiv 0 \quad (\text{B.4})$$

$$[A^\dagger W A] \vec{x} = [A^\dagger W] \vec{b} \quad (\text{B.5})$$

These are called the Normal Equations, and the matrix on the left-hand-side is given by $[A^\dagger W A] = \frac{1}{2} \nabla^2 \chi^2$ and is called the Hessian matrix.

Pseudo-Inverse : The solution is given by

$$\vec{x} = [A^\dagger WA]^{-1}[A^\dagger W]\vec{b} \quad (\text{B.6})$$

$[A^+] \equiv [A^\dagger WA]^{-1}[A^\dagger W]$ is the pseudo inverse of the system. If $[A]$ is unitary, then $[A^+] = [A^{-1}]$.

Eqn B.6 can be evaluated exactly *via* LU or Cholesky decomposition if $[A^\dagger WA]$ is invertible. When a direct computation of the solution is not feasible or $[A^\dagger WA]$ is singular, an iterative approach is required to solve $\nabla\chi^2 = 0$.

Iterative Solution : The solution of $\nabla\chi^2 = 0$ is a root-finding problem. Consider the Taylor expansion of a function $f(x)$.

$$f(x) = f(a) + f'(x)|_a(x - a) \quad (\text{B.7})$$

Setting $f(x) = 0$, gives

$$x = a - \frac{1}{f'(x)|_a}f(a) \quad (\text{B.8})$$

Therefore, $\nabla\chi^2 = 0$ can be solved iteratively from an initial guess as follows.

$$\vec{x}_{i+1} = \vec{x}_i + \alpha[A^\dagger WA]^{-1}\left([A^\dagger W]\vec{b} - [A^\dagger WA]\vec{x}_i\right) \quad (\text{B.9})$$

where $\alpha \in \{0, 1\}$ is a step-size and i indicates iteration number. Various approximations of $[A^\dagger WA]^{-1}$ can be used in this iterative process. In a first-order optimization, $[A^\dagger WA]^{-1}$ is approximated by the inverse of the main diagonal of the Hessian and a step size $\alpha < 1$ is used to dampen the effect of this inaccuracy (steepest descent, levenberg-macquart). Higher order methods provide better approximations of $[A^\dagger WA]^{-1}$ (BFGS, etc..)

Linear least squares in Interferometric Imaging : For the problem of interferometric imaging, the measurement matrix is usually singular and we need to work with the normal equations. For standard imaging, the Hessian matrix $[A^\dagger WA]$ is a Toeplitz matrix with a shifted version of a single function in each row. When applied to a column vector of image pixels, it implements the shift-multiply-add sequence of a convolution. The associated convolution kernel (elements of the middle (unshifted) row) is called the point-spread-function. The RHS of the normal equations is therefore the result of a convolution between a vector of image pixels representing the sky brightness distribution and another vector of image pixels containing the point-spread-function. Therefore, in the context of interferometric imaging, the iterative process described above is called a deconvolution. A diagonal approximation of the Hessian matrix and its inverse reduces to a pixel-by-pixel normalization of the RHS by the peak of the point-spread-function. The pixel amplitudes at the locations of the source peaks measure the source flux in physical units, and provide enough information to construct a model to add to \vec{x}_i .

General Block Linear Least Squares : Another way to write the measurement equations are to express the linear system of equations as a sum of several such systems. The Hessian matrix and normal equations are derived in the same way, but are now written in block matrix form. In certain situations, such as when the parameters are themselves modeled as a linear combination of unknown coefficients and known basis functions, such a form allows a simple but effective approximation of the Hessian and its inverse.

Let $[A_p]$ be an $n \times m$ measurement matrix for the quantity \vec{x}_p written as an $m \times 1$ list. A set of N_p measurements are added together to form an $n \times 1$ list of data (\vec{b}) . The block-matrix equivalent of Eqn.B.1 is given as follows.

$$\sum_{p=0}^{N_p} [A_p] \vec{x}_p = \vec{b} \quad (\text{B.10})$$

The measurement and normal equations can be written in standard block matrix form, and an iterative solution computed similar to regular linear least squares. For simplicity, let $N_p = 3$, and let there be m parameters in each subset \vec{x}_p .

Measurement Equations in Block form :

$$\begin{bmatrix} [A_0] & [A_1] & [A_2] \end{bmatrix} \begin{bmatrix} \vec{x}_0 \\ \vec{x}_1 \\ \vec{x}_2 \end{bmatrix} = \vec{b} \quad (\text{B.11})$$

All N_p measurement matrices of shape $n \times m$ are placed side-by-side to form a larger measurement matrix. The list of parameters becomes a vertical stack of N_p vectors each of shape $m \times 1$. The new measurement matrix of shape $n \times mN_p$ operates on an $mN_p \times 1$ list of parameters to form an $n \times 1$ list of measurements.

Normal Equations in Block form : The normal equations can also be written in block-matrix form. Note that matrix products can be written block-by-block to follow the element-by-element multiplications implicit in Eqns. B.2 to B.5 (which show how normal equations are constructed from measurement equations). The following equation shows Eqn. B.5 in block matrix form, before multiplying out the matrix blocks.

$$\begin{bmatrix} [A^\dagger_0] \\ [A^\dagger_1] \\ [A^\dagger_2] \end{bmatrix} [W] \begin{bmatrix} [A_0] & [A_1] & [A_2] \end{bmatrix} \begin{bmatrix} \vec{x}_0 \\ \vec{x}_1 \\ \vec{x}_2 \end{bmatrix} = \begin{bmatrix} [A^\dagger_0] \\ [A^\dagger_1] \\ [A^\dagger_2] \end{bmatrix} [W] \vec{b} \quad (\text{B.12})$$

When these block matrices are multiplied out, the normal equations become a system of equations in which the Hessian has the shape $mN_p \times mN_p$ and the list of parameters and the RHS vector contain $mN_p \times 1$ elements each.

$$\begin{bmatrix} [A^\dagger_0 WA_0] & [A^\dagger_0 WA_1] & [A^\dagger_0 WA_2] \\ [A^\dagger_1 WA_0] & [A^\dagger_1 WA_1] & [A^\dagger_1 WA_2] \\ [A^\dagger_2 WA_0] & [A^\dagger_2 WA_1] & [A^\dagger_2 WA_2] \end{bmatrix} \begin{bmatrix} \vec{x}_0 \\ \vec{x}_1 \\ \vec{x}_2 \end{bmatrix} = \begin{bmatrix} [A^\dagger_0 W]\vec{b} \\ [A^\dagger_1 W]\vec{b} \\ [A^\dagger_2 W]\vec{b} \end{bmatrix} \quad (\text{B.13})$$

The full Hessian consists of $N_p \times N_p$ blocks, each of size $m \times m$. the list of parameters is a stack of N_p column vectors, and the RHS vector is a set of N_p weighted inversions of the data vector.

Iterative Solution in Block form : An iterative solution to this system is obtained by following Eqn. B.9 with block matrices.

$$\begin{bmatrix} \vec{x}_0 \\ \vec{x}_1 \\ \vec{x}_2 \end{bmatrix}_{i+1} = \begin{bmatrix} \vec{x}_0 \\ \vec{x}_1 \\ \vec{x}_2 \end{bmatrix}_i + \alpha \begin{bmatrix} [A^\dagger_0 WA_0] & [A^\dagger_0 WA_1] & [A^\dagger_0 WA_2] \\ [A^\dagger_1 WA_0] & [A^\dagger_1 WA_1] & [A^\dagger_1 WA_2] \\ [A^\dagger_2 WA_0] & [A^\dagger_2 WA_1] & [A^\dagger_2 WA_2] \end{bmatrix}^{-1} \begin{bmatrix} [A^\dagger_0 W]\vec{b} - \sum_{p=0}^2 [A^\dagger_0 WA_p]\vec{x}_p \\ [A^\dagger_1 W]\vec{b} - \sum_{p=0}^2 [A^\dagger_1 WA_p]\vec{x}_p \\ [A^\dagger_2 W]\vec{b} - \sum_{p=0}^2 [A^\dagger_2 WA_p]\vec{x}_p \end{bmatrix} \quad (\text{B.14})$$

A few points to note about this solution process are :

1. The computation of the RHS of the normal equations via $[A^\dagger_p W]\vec{b}$ is equivalent to a matched-filtering process, where the p^{th} RHS vector is a measure of how close the data match a template (contained in $[A_p]$).
2. Off diagonal Hessian blocks are a measure of the non-orthogonality of the p basis functions. A Hessian with non-zero blocks only on the diagonal implies a perfectly orthogonal basis set, and then each parameter can be solved independent of the others. If off-diagonal blocks contain non-zero elements, it implies that the parameters are coupled and need to be treated together during the solution process.
3. The condition number of the block Hessian matrix gives a measure of how robust an inversion would be. For example, if the basis functions are linearly dependent, the block Hessian will be rank-deficient, there will be some eigen-values equal to zero, and the condition number rises sharply, indicating that the solution process is highly sensitive to variations in its parameters.
4. If the p basis functions form a geometric progression, then this block Hessian is a Vandermonde system with a geometric progression in each row and column. A useful property of such a system is that the diagonal blocks are the result of even powers of basis functions and the system is guaranteed to be block positive definite and block-invertible or block-diagonalizable.

Block Linear least squares in Interferometric Imaging : For interferometric imaging in which the model is represented as a linear combination of coefficients and basis functions, each Hessian block is a Toeplitz convolution operator. The RHS vector is therefore a sum of convolutions of different image pixel vectors and convolution kernels, and the process of solving such a system is called a block or joint deconvolution. In this particular form, the Hessian can be approximated by a matrix of diagonal blocks where the elements on the diagonal of each block is the peak of the point-spread-function represented by that block. This provides a better estimate of its inverse than just a diagonal approximation of the full Hessian, and forms an $N_p \times N_p$ matrix to be inverted per pixel. When $N_p = 1$, this process is equivalent to a pixel-by-pixel normalization by the peak of the point-spread-function. When the basis functions involve a geometric progression, this per-pixel $N_p \times N_p$ matrix is exactly invertible, and simultaneously produces estimates for all the coefficients involved in describing the amplitude of that pixel.

Bibliography

- Anantharamaiah, K. R., Cornwell, T. J., and Narayan, R. (1989). Synthesis Imaging of Spatially Coherent Objects. In Perley, R. A., Schwab, F. R., and Bridle, A. H., editors, *Synthesis Imaging in Radio Astronomy*, volume 6 of *Astronomical Society of the Pacific Conference Series*, page 415. 9
- Bhatnagar, S. (2001). *Radio Study of Galactic Supernova Remnants and the Interstellar Medium*. PhD thesis, University of Pune; National Centre for Radio Astrophysics; Tata Institute of Fundamental Research, Pune, India. 7
- Bhatnagar, S. and Cornwell, T. J. (2004). Scale sensitive deconvolution of interferometric images. I. Adaptive Scale Pixel (Asp) decomposition. *Astron. & Astrophys.*, 426:747–754. 47, 86
- Bhatnagar, S., Cornwell, T. J., and Golap, K. (2004). Solving for the antenna based pointing errors. Technical report, EVLA Memo 84. 50, 57
- Bhatnagar, S., Cornwell, T. J., and Golap, K. (2006). Corrections of errors due to antenna power patterns during imaging. Technical report, EVLA Memo 100. 57
- Bhatnagar, S., Cornwell, T. J., Golap, K., and Uson, J. M. (2008). Correcting direction-dependent gains in the deconvolution of radio interferometric images. *Astron. & Astrophys.*, 487:419–429. 3, 50, 53, 55, 57, 59, 61
- Bhatnagar, S. and Nityananda, R. (2001). Solving for closure errors due to polarization leakage in radio interferometry of unpolarized sources. *Astron. & Astrophys.*, 375:344–350. 30
- Bicknell, G. V. and Begelman, M. C. (1996). Understanding the Kiloparsec-Scale Structure of M87. *ApJ*, 467:597. 215
- Bong, S.-C., Lee, J., Gary, D. E., and Yun, H. S. (2006). Spatio-spectral Maximum Entropy Method. I. Formulation and Test. *ApJ*, 636:1159–1165. 48
- Bracewell, R. N. and Roberts, J. A. (1954). Aerial Smoothing in Radio Astronomy. *Australian Journal of Physics*, 7:615. 36, 38
- Briggs, D. S. (1995). *High fidelity deconvolution of moderately resolved sources*. PhD thesis, The New Mexico Institute of Mining and Technology, Socorro, New Mexico, USA. 7, 26, 43, 44, 47

- Briggs, D. S., Schwab, F. R., and Sramek, R. A. (1999). Imaging. In G. B. Taylor, C. L. Carilli, & R. A. Perley, editor, *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, page 127. [41](#)
- Brisken, W. (2003). Using grasp8 to study the vla beam. Technical report, EVLA Memo 58. [73](#), [163](#)
- Carilli, C. L. and Barthel, P. D. (1996). Cygnus A. *Astron. & Astrophys. Review*, 7:1–54. [180](#)
- Carilli, C. L., Perley, R. A., Dreher, J. W., and Leahy, J. P. (1991). Multifrequency radio observations of Cygnus A - Spectral aging in powerful radio galaxies. *ApJ*, 383:554–573. [83](#), [179](#), [182](#), [184](#), [187](#)
- Churazov, E., Brüggen, M., Kaiser, C. R., Böhringer, H., and Forman, W. (2001). Evolution of Buoyant Bubbles in M87. *ApJ*, 554:261–273. [215](#), [235](#), [238](#)
- Clark, B. G. (1980). An efficient implementation of the algorithm 'clean'. *Astron. & Astrophys.*, 89:377. [46](#)
- Clark, B. G. (1999). Coherence in radio astronomy. In *Astron. Soc. Pac. Conf. Ser. 180: Synthesis Imaging in Radio Astronomy II*, page 1. [10](#), [11](#)
- Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., and Broderick, J. J. (1998). The NRAO VLA Sky Survey. *Astron. J.*, 115:1693–1716. [197](#)
- Conway, J. E., Cornwell, T. J., and Wilkinson, P. N. (1990). Multi-Frequency Synthesis - a New Technique in Radio Interferometric Imaging. *mnras*, 246:490. [1](#), [3](#), [14](#), [47](#), [76](#), [102](#), [105](#), [119](#), [121](#)
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Math. of Comp.*, 19:297. [22](#)
- Cornwell, T., Braun, R., and Briggs, D. S. (1999). Deconvolution. In G. B. Taylor, C. L. Carilli, & R. A. Perley, editor, *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, page 151. [36](#)
- Cornwell, T. and Fomalont, E. B. (1999). Self-Calibration. In G. B. Taylor, C. L. Carilli, & R. A. Perley, editor, *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, page 187. [28](#)
- Cornwell, T. J. (1988). Radio-interferometric imaging of very large objects. *Astron. & Astrophys.*, 202:316–321. [58](#)

- Cornwell, T. J. (1995a). The generic interferometer: I overview of calibration and imaging. Technical report, AIPS++ Note 183. [7](#)
- Cornwell, T. J. (1995b). The generic interferometer: II image solvers. Technical report, AIPS++ Note 184. [7](#)
- Cornwell, T. J. (2008). Multi-Scale CLEAN deconvolution of radio synthesis images. *IEEE Journal of Selected Topics in Sig. Proc.*, 2:793–801. [3](#), [46](#), [47](#), [86](#), [89](#), [94](#), [98](#), [203](#)
- Cornwell, T. J. and Evans, K. J. (1985). A simple maximum entropy deconvolution algorithm. *Astron. & Astrophys.*, 143:77–83. [47](#)
- Cornwell, T. J., Golap, K., and Bhatnagar, S. (2003). W-projection: A new algorithm for non-coplanar baselines. Technical report, EVLA Memo 67. [58](#)
- Cornwell, T. J., Golap, K., and Bhatnagar, S. (2008). The non-coplanar baselines effect in radio interferometry: The w-projection algorithm. *IEEE Journal of Selected Topics in Sig. Proc.*, 2:647–657. [50](#), [55](#), [58](#), [197](#)
- Cornwell, T. J. and Perley, R. A. (1992). Radio-interferometric imaging of very large fields - the problem of non-coplanar arrays. *Astron. & Astrophys.*, 261:353–364. [58](#)
- Cornwell, T. J. and Wilkinson, P. N. (1981). A new method for making maps with unstable radio interferometers. *Mon. Not. Roy. Astron. Soc.*, 196:1067–1086. [27](#), [28](#)
- Cotton, Jr., W. D. and Uson, J. (2006). Ionospheric Phase Errors and Corrections at 1 m Wavelength. In *Bulletin of the American Astronomical Society*, volume 38 of *Bulletin of the American Astronomical Society*, page 1015. [50](#)
- Eilek, J., Hardee, P., and Lobanov, A. (2003). Particle acceleration in the M87 jet. *New Astronomy Review*, 47:505–507. [221](#)
- Eilek, J. A. (1996). How Radio Sources Stay Young: Spectral Aging Revisited. In P. E. Hardee, A. H. Bridle, & J. A. Zensus, editor, *Energy Transport in Radio Galaxies and Quasars*, volume 100 of *Astronomical Society of the Pacific Conference Series*, page 281. [215](#)
- Eilek, J. A. and Arendt, P. N. (1996). The Synchrotron Spectrum of Diffuse Radio Sources: Effects of Particle and Field Distributions. *ApJ*, 457:150. [221](#)
- Eilek, J. A., Melrose, D. B., and Walker, M. A. (1997). Synchrotron Aging in Filamented Magnetic Fields. *ApJ*, 483:282. [221](#)
- Fomalont, E. B. and Perley, R. A. (1999). Calibration and Editing. In G. B. Taylor, C. L. Carilli, & R. A. Perley, editor, *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, page 79. [28](#)

- Goodman, J. W. (2002). *Introduction to Fourier Optics*. McGraw-Hill. 9
- Govoni, F. and Feretti, L. (2004). Magnetic Fields in Clusters of Galaxies. *International Journal of Modern Physics D*, 13:1549–1594. 222
- Greisen, E. W., Spekkens, K., and van Moorsel, G. A. (2009). Aperture synthesis observations of the nearby spiral ngc 6503: Modeling the thin and thick h i disks. *Astron. J.*, 137(6):4718–4733. 46, 86, 94, 98, 99
- Hamaker, J. P., Bregman, J. D., and Sault, R. J. (1996). Understanding radio polarimetry. I. Mathematical foundations. *Astron. & Astrophys. Suppl. Ser.*, 117:137–147. 7, 17, 48
- Högbom, J. A. (1974). Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron. & Astrophys. Suppl. Ser.*, 15:417–426. 46
- Holdaway, M. A. and Wardle, J. F. C. (1990). Maximum entropy imaging of polarization in very long baseline interferometry. In Gmitro, A. F., Idell, P. S., and Lahaie, I. J., editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 1351, pages 714–724. 48
- Intema, H. T., van der Tol, S., Cotton, W. D., Cohen, A. S., van Bemmel, I. M., and Röttgering, H. J. A. (2009). Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results. *aap*, 501:1185–1205. 50
- Kraus, J. D. (1986). *Radio astronomy*. Powell, Ohio: Cygnus-Quasar Books, 1986. 11
- Lawson, C. L. and Hanson, R. J. (1974). *Solving Least Squares Problems*. Prentice-Hall Series in Automatic Computation. Prentice-Hall. 47
- Likhachev, S. (2005). Multi-Frequency Imaging for VLBI. In J. Romney & M. Reid, editor, *Future Directions in High Resolution Astronomy*, volume 340 of *Astronomical Society of the Pacific Conference Series*, page 608. 47
- Molendi, S. (2002). On the Temperature Structure of M87. *ApJ*, 580:815–823. 213
- Narayan, R. and Nityananda, R. (1986). Maximum entropy image restoration in astronomy. *Ann. Rev. Astron. Astrophys.*, 24:127–170. 47
- Nijboer, R. J. and Noordam, J. E. (2007). LOFAR Calibration. In Shaw, R. A., Hill, F., and Bell, D. J., editors, *Astronomical Data Analysis Software and Systems XVI*, volume 376 of *Astronomical Society of the Pacific Conference Series*, page 237. 29, 50
- Owen, F. N., Eilek, J. A., and Kassim, N. E. (2000). M87 at 90 Centimeters: A Different Picture. *ApJ*, 543:611–619. 189, 191, 195, 211, 213, 214, 215, 216, 227, 235, 237

- Owen, F. N., Eilek, J. A., and Keel, W. C. (1990). Detection of large Faraday rotation in the inner 2 kiloparsecs of M87. *ApJ*, 362:449–454. [215](#), [227](#)
- Pacholczyk, A. G. (1970). *Radio Astrophysics : Nonthermal Processes in Galactic and Extragalactic Sources*. W.H.Freeman and Company. [216](#), [219](#), [222](#)
- Perley, R. A. and Taylor, G. B. (2003). The vla calibrator manual. Technical report, NRAO. [196](#)
- Perlman, E. S. and Wilson, A. S. (2005). The X-Ray Emissions from the M87 Jet: Diagnostics and Physical Interpretation. *ApJ*, 627:140–155. [215](#), [237](#)
- Peterson, J. R., Kahn, S. M., Paerels, F. B. S., Kaastra, J. S., Tamura, T., Bleeker, J. A. M., Ferrigno, C., and Jernigan, J. G. (2003). High-Resolution X-Ray Spectroscopic Constraints on Cooling-Flow Models for Clusters of Galaxies. *ApJ*, 590:207–224. [211](#)
- Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1988). *Numerical Recipes in C*. Press Syndicate of the University of Cambridge, Cambridge. [94](#)
- Rau, U. and Cornwell, T. J. (2005). Monte Carlo Image Analysis in Radio Interferometry MC-FIT: A Bayesian Approach to Object Detection. In P. Shopbell, M. Britton, & R. Ebert, editor, *Astronomical Data Analysis Software and Systems XIV*, volume 347 of *Astronomical Society of the Pacific Conference Series*, page 168. [47](#)
- Reid, R. (2003). *Smear Fitting: A New Method of Obtaining Images From Interferometric Data*. PhD thesis, Department of Astronomy and Astrophysics, University of Toronto. [47](#)
- Rottmann, H., Mack, K., Klein, U., and Wielebinski, R. (1996a). The radio lobes of Virgo A at 2.8cm wavelength. *Astron. & Astrophys.*, 309:L19–L22. [189](#), [215](#), [238](#)
- Rottmann, H., Mack, K. H., Klein, U., Wielebinski, R., Kassim, N., and Perley, R. (1996b). The Radio Lobes of Virgo A. In R. D. Ekers, C. Fanti, & L. Padrielli, editor, *Extragalactic Radio Sources*, volume 175 of *IAU Symposium*, page 343. [215](#), [235](#)
- Sault, R. J., Bock, D. J., and Duncan, A. (1999). Polarimetric imaging of large fields in radio astronomy. *Astron. & Astrophys.*, 139:387. [48](#), [58](#)
- Sault, R. J., Hamaker, J. P., and Bregman, J. D. (1996). Understanding radio polarimetry. II. Instrumental calibration of an interferometer array. *Astron. & Astrophys. Suppl. Ser.*, 117:149–159. [7](#), [21](#), [30](#)
- Sault, R. J. and Wieringa, M. H. (1994). Multi-frequency synthesis techniques in radio interferometric imaging. *Astron. & Astrophys. Suppl. Ser.*, 108:585–594. [1](#), [3](#), [46](#), [68](#), [74](#), [76](#), [102](#), [109](#), [117](#)

- Schwab, F. R. (1980). Adaptive calibration of radio interferometer data. In W. T. Rhodes, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 231 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 18–25. [28](#)
- Schwab, F. R. and Cotton, W. D. (1983). Global fringe search techniques for VLBI. *Astron. J.*, 88:688–694. [38](#), [46](#)
- Shibata, R., Matsushita, K., Yamasaki, N. Y., Ohashi, T., Ishida, M., Kikuchi, K., Böhringer, H., and Matsumoto, H. (2001). Temperature Map of the Virgo Cluster of Galaxies Observed with ASCA. *ApJ*, 549:228–243. [213](#), [238](#)
- Steer, D. G., Dewdney, P. E., and Ito, M. R. (1984). Enhancements to the deconvolution algorithm 'CLEAN'. *aap*, 137:159–165. [46](#)
- Sutton, E. C. and Wandelt, B. D. (2006). Optimal Image Reconstruction in Radio Interferometry. *AJ Supp.*, 162:401–416. [48](#)
- Taylor, G. B., Carilli, C. L., and Perley, R. A., editors (1999). *Astron. Soc. Pac. Conf. Ser. 180: Synthesis Imaging in Radio Astronomy II*. [7](#), [26](#)
- Thompson, A. R. and Daddario, L. R. (1982). Frequency response of a synthesis array - Performance limitations and design tolerances. *Radio Science*, 17:357–369. [28](#)
- Thompson, A. R., Moran, J. M., and Swenson, Jr., G. W. (1986). *Interferometry and Synthesis in Radio Astronomy*. John Wiley & Sons, Inc. [7](#), [9](#)
- Urvashi R.V., Cornwell, T. J., and Myers, S. (2006). Multi frequency synthesis imaging for the evla : An initial investigation. Technical report, EVLA Memo 101. [75](#)
- Uson, J. M. and Cotton, W. D. (2008). Beam squint and Stokes V with off-axis feeds. *Astron. & Astrophys.*, 486:647–654. [50](#), [57](#)
- Voronkov, M. and Cornwell, T. (2007). Calibration and Imaging with eigenbeams. In *From Planets to Dark Energy: the Modern Radio Universe. October 1-5 2007, The University of Manchester, UK. Published online at SISSA, Proceedings of Science*, p.139. [50](#)
- Wakker, B. P. and Schwarz, U. J. (1988). The Multi-Resolution CLEAN and its application to the short-spacing problem in interferometry. *Astron. & Astrophys.*, 200:312–322. [46](#)
- Wiaux, Y., Puy, G., Boursier, Y., and Vandergheynst, P. (2009). Compressed sensing for radio interferometry: spread spectrum imaging techniques. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7446 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*. [48](#)
- Yatawatta, S., Zaroubi, S., de Bruyn, G., Koopmans, L., and Noordam, J. (2008). Radio Interferometric Calibration Using The SAGE Algorithm. *ArXiv e-prints*. [50](#)