Astronomy & Astrophysics manuscript no. msmfs˙aa
January 22, 2011

A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry

U. Rau[1] and T.J. Cornwell[2]

[1] National Radio Astronomy Observatory, Socorro, NM, USA
e-mail: rurvashi@aoc.nrao.edu
[2] Australia Telescope National Facility, CSIRO, Sydney, AU
e-mail: tim.cornwell@atnf.csiro.au

Received XXX XX, 2009; accepted XXX XX, 2009

ABSTRACT

Context. The use of broad-band receivers in radio interferometery impacts standard continuum synthesis imaging in three ways. First, correlations measured at different frequencies sample the visibility function of the sky brightness at different spatial frequencies. Second, the sky brightness can change significantly across the large frequency range that these receivers are sensitive to. Third, direction-dependent instrumental effects such as the element response pattern also vary with frequency, and this affects the spectral characteristics of the measured sky brightness during wide-field continuum imaging. Existing algorithms designed to account for some of the above, can be shown to be insufficient to achieve target image fidelity and dynamic range over the wide bands that most new instruments are being built with.
Aims. This paper describes a multi-scale multi-frequency deconvolution algorithm for the minor-cycle of synthesis imaging that combines the information measured about the source spectrum with the additional spatial-frequency-coverage offered by multi-frequency-synthesis to reconstruct source structure.
Methods. The MS-MFS (multi-scale multi-frequency synthesis) algorithm discussed in this paper extends the MF-Clean algorithm to reconstruct spectral curvature in addition to spectral index and combines it with a multi-scale deconvolution approach similar to MS-Clean. We also discuss a simpler hybrid of spectral-line and continuum imaging methods that may suffice in certain situations.
Results. We show via simulations and application to wideband (E)VLA data (1 GHz to 2 GHz), that it is possible to reconstruct both spatial and spectral structure of compact and extended emission at the continuum sensitivity level and angular resolution allowed by the highest sampled frequency. We show imaging examples using simulations, multi-frequency VLA data and wideband EVLA data, to illustrate the capabilities of the MS-MFS algorithm and conditions under which these techniques are feasible and give accurate results.

Key words. multi-frequency-synthesis – multi-scale-deconvolution – spectral index – spectral curvature –

1. Introduction

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. The large instantaneous bandwidths offered by new front-end systems increases the raw continuum sensitivity of the instrument, as well as greatly enhances our ability to measure the detailed spectral structure of the incident radiation across large continuous bandwidths. One prominent reason for using broad-band receivers on an imaging interferometer has been to obtain a continuum image that makes use of the increased sensitivity and spatial frequency coverage offered by combining measurements from multiple frequencies. So far, the bandwidths used have been relatively narrow (< 25%), and effects due to spectral structure of the sky-brightness distribution have been considered only in the context of reducing errors in the continuum image.Any spectral information is obtained only as a by-product, and attention has not been paid to their accuracy. But now, the new bandwidths (100%) are large enough to allow the spectral structure of the sky brightness distribution to also be reconstructed to produce a meaningful astrophysical measurement. To do so, we need imaging algorithms that model and reconstruct both spatial and spectral structure simultaneously, and that are also sensitive to various effects of combining measurements from a large range of frequencies (namely varying ranges of sampled spatial scales and varying array-element response functions)

The simplest method of wide-band image reconstruction is to image each frequency channel separately and combine the results at the end. However, single-channel imaging is restricted to the narrow-band *uv*-coverage and sensitivity of the instrument, and source spectra can be studied only at the angular resolution allowed by the lowest frequency in the sampled range. For complicated extended emission, the single-frequency *uv*-coverage may not be sufficient to produce a consistent solution across frequency. While such imaging may suffice for many science goals, it does not take full advantage of what an instantaneously wide-band instrument provides, namely the sensitivity and spatial-frequency coverage obtained by combining measurements from multiple receiver frequencies.

Multi-Frequency-Synthesis (MFS) is the technique of combining measurements at multiple discrete receiver frequencies during synthesis imaging. MFS was initially done (REF) to increase the aperture-plane coverage of sparse arrays by using narrow-band receivers and switching frequencies during the observations. Wide bandwidth systems (10%) later presented the problem of bandwidth smearing, which was eliminated (REF) by splitting the wide band into narrow-band channels and mapping them onto their correct spatial frequencies during imaging. It was assumed that at the receiver sensitivities of the time, the measured 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, in particular, the first two terms of a Taylor series. A map of the average spectral index is derived from the coefficient maps. [Sault & Wieringa 1994] describe the MF-Clean algorithm which uses a formulation similar to double-deconvolution but calculate Taylor-coefficients via a least-squares solution. MF-Clean is implemented in the miriad software package. The MF-Clean formulation ignores the process of resampling multi-frequency visibilities onto a single grid of spatial frequencies, and for arrays with many overlapping spatial-frequency tracks, errors are incurred. More recently, Likhachev [2005] re-derive the least-squares method used in MF-Clean to include more than two series coefficients. [Bong et al. 2006] describe spatio-spectral MEM, an entropy based method in which single-channel imaging is done along with a smoothness constraint applied across frequency.

So far, these CLEAN-based multi-frequency deconvolution algorithms used point-source (zero-scale) flux components to model the sky emission. We show in this paper that with the MF-Clean approach, deconvolution errors that occur with a point-source model are enhanced in the spectral index image because of error propagation effects, and that the use of a multi-scale technique can minimize this. [Cornwell 2008] describe the MS-CLEAN algorithm which is a matched filtering technique using templates constructed from the instrument response to various large scale flux components. [Greisen et al. 2009] describe a method simular to MS-Clean, and [Bhatnagar & Cornwell 2004] describe the ASP-CLEAN algorithm that explicitly fits for the parameters of Gaussian flux components and uses scale size to aid the separation of signal from noise.

For high dynamic range imaging across wide fields-of-view, direction-dependent instrumental effects that need to be accounted for are the w-term and the antenna power pattern, both of which are frequency-dependent. [Bhatnagar et al. 2008]describe an algorithm for the correction of time-variable wide-field instrumental effects for narrow-band interferometric imaging.

In this paper, we describe MS-MFS (multi-scale multi-frequency synthesis) as an algorithm that combines variants of the MF-Clean and MS-Clean approaches to simultaneously reconstruct both spatial and spectral structure of the sky-brightness distribution. Frequency-dependent primary-beam correction is considered as a post-deconvolution correction step[1]. In section ZZ, we show imaging examples using simulations, multi-frequency VLA data and wideband EVLA data, to illustrate the capabilities and limits of the MS-MFS algorithm.

1.1. Wide-Band Imaging

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.

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-distance[2] 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 where the visibility function is sampled at all frequencies in the band, and there is sufficient information to reconstruct both spatial and spectral structure. 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.

For a flat-spectrum source, measurements at multiple frequencies sample the same sky brightnes distribution at different ranges of spatial scales, increasing the signal-to-noise of the measurements in regions of overlap, and providing better uv-coverage and angular resolution. Therefore standard deconvolution algorithms applied to measurements combined via MFS will suffice to reconstruct source structure across the full range of spatial scales measured across the band.

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.

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 by 100% in amplitude across the band would break the above assumption (band-limited signals). 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.

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.

[1] The integration of direction-dependent correction algorithms such as AW-Projection with MS-MFS will be discussed in a subsequent paper.

[2] 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.

Finally, for wide-field imaging, the frequency-dependence of the primary-beam can introduce artificial spectral effects that result in a 100% variation of the flux across the band. 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 before or during multi-frequency synthesis imaging.

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 structure of the sky brightness. By combining such models with the known frequency-dependence of the spatial-frequency coverage and element response function, it is possible to reconstruct the broad-band sky brightness distribution from incomplete spectral and spatial-frequency sampling.

2. Multi-scale Multi-frequency deconvolution

The MS-MFS algorithm described here is based on the iterative image-reconstruction framework[3] described in [Rau et al. 2009]. Sections xxx to yyy formulate the algorithm used in the minor-cycle of iterative deconvolution for MS-MFS. Differences between the multi-scale and multi-frequency parts of MS-MFS with the original MS-Clean and MF-Clean approaches are highlighted in sections XX and YY. The implementation of this algorithm in the CASA package is summarized in section ZZ.

2.1. Parameterization of spatial structure

For a multi-scale model, an image can be written as a linear combination of images at different angular resolutions (Ref MS-CLEAN). A multiscale representation of an image is given by

$$\boldsymbol{I}^{model} = \sum_{s=0}^{N_s-1} \boldsymbol{I}_s^{shp} \star \boldsymbol{I}_s^{sky,\delta} \tag{1}$$

where N_s is the number of discrete spatial scales used to represent the image, and $\boldsymbol{I}_s^{sky,\delta}$ represents a collection of δ-functions that describe the locations and integrated amplitudes of flux components of scale s in the image. $\boldsymbol{I}_s^{shp}$ is a tapered truncated parabola of width proportional to s. The symbol $\star$ denotes convolution. In order to always allow for the modeling of unresolved sources, we choose the first scale function $\boldsymbol{I}_{s=0}^{shp}$ to be a δ-function.

2.2. Parameterization of spectral structure

The spectrum of each flux component is modeled by a polynomial in frequency (a Taylor series expansion about ν_0).

$$\boldsymbol{I}_\nu^{model} = \sum_{t=0}^{N_t-1} w_\nu^t \boldsymbol{I}_t^{sky} \quad \text{where} \quad w_\nu^t = \left(\frac{\nu-\nu_0}{\nu_0}\right)^t \tag{2}$$

where N_t is the order of the Taylor series expansion, and the I_t^m represent multi-scale Taylor coefficient images (moment maps ???). This decomposition is linear in the coefficients as well as in the basis functions.

[3] A steepest-descent chi-square minimization is done by iterating between two steps. A major cycle computes the RHS of the normal equations, and the minor cycle performs an approximate inverse of the Hessian matrix to generate an update direction (image model components).

These Taylor coefficients can be interpreted by choosing an astro-physically appropriate spectral model and performing a Taylor expansion to derive expressions that each coefficient maps to. One choice for a spectral model is a power law with a varying index, represented by a second-order polynomial in $log(I)\ vs\ log\left(\frac{\nu}{\nu_0}\right)$ space. The variation of the spectral index with frequency is described by an average spectral index α and a curvature term β.

$$I_\nu^{sky} = I_{\nu_0}^{sky}\left(\frac{\nu}{\nu_0}\right)^{\alpha+\beta log\left(\frac{\nu}{\nu_0}\right)} \tag{3}$$

The main reason behind this choice of interpretation is the fact that continuum synchrotron emission is usually modeled (and observed) as a power law distribution with frequency. Across the wide frequency ranges that new receivers are now sensitive to, spectral breaks, steepening and turnovers need to be factored into models, and the simplest way to include them and ensure smoothness, is spectral curvature. (FIGURE). (Wideband imaging algorithms describes in CCW/SW use a fixed spectral index across the band, and handle slight curvature by performing multiple rounds of imaging after removing the dominant/average α at each stage. They also suggest using higher order polynomials to handle spectral curvature.)

A Taylor expansion of Eqn 3 yields the following expressions for the first three coefficients ($t = 0, 1, 2$ in Eqn **??**) from which the spectral index α and curvature β images can be computed algebraically.

$$I_0^m = I_{\nu_0}^{sky} \ ; \ I_1^m = \alpha I_{\nu_0}^{sky} \ ; \ I_2^m = \left(\frac{\alpha(\alpha-1)}{2} + \beta\right) I_{\nu_0}^{sky} \tag{4}$$

Note that with this choice of parameterization, we are using a polynomial to model an exponential, and this has caveats, etc which are discussed in the section on errors. Also, note that there can be other interpretations of the Taylor coefficients and other expansions. A power-series expansion about α and β will yield a logarithmic expansion i.e. I vs log nu. (CHECK if/when this is better) It is however impractical to work directly in Log I and Log nu space because that involves taking logs and exp of image pixel amplitudes and this is highly unstable numerically.

Across wide fields of view, an approximate correction of the primary-beam and its frequency dependence can also be folded into this formulation. We can write

$$I_\nu^{sky} = P_\nu I_\nu^{true} = P_{\nu_0} I_{\nu_0}^{true}\left(\frac{\nu}{\nu_0}\right)^{[\alpha_{true}+\alpha_{PB}]+[\beta_{true}+\beta_{PB}]log\left(\frac{\nu}{\nu_0}\right)} \tag{5}$$

where P_{ν_0} is the primary beam at the reference frequency, and α_{PB} and β_{PB} are spectral index and curvature of the frequency dependence of the primary beam. An image reconstructed using such a formulation can be corrected in a post-deconvolution step, if the primary-beam and its frequency dependence is known *a priori*. This however corrects only for an average primary-beam and its frequency dependence, and a more accurate solution requires the AW-Projection algorithm.

2.3. Multi-scale multi-frequency model

For multi-scale and multi-frequency deconvolution, the sky-brightness distribution can be parameterized in a multi-scale basis (1), with the amplitudes of each component described by a polynomial across frequency (2). A region of emission in which the spectrum varies with position will then be modeled as a sum of wide-band flux components. The image flux model at each frequency can be written as a linear sum of coefficient images at different spatial scales.

$$\boldsymbol{I}_\nu^{model} = \sum_{t=0}^{N_t}\sum_{s=0}^{N_s} w_\nu^t \left[\boldsymbol{I}_s^{shp} \star \boldsymbol{I}_{\substack{s\\t}}^{sky}\right] \quad \text{where} \quad w_\nu^t = \left(\frac{\nu-\nu_0}{\nu_0}\right)^t \tag{6}$$

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. $\boldsymbol{I}_{\substack{s\\t}}^{sky}$ 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.

2.4. Measurement equations

The Measurement Equations[4] for a sky brightness distribution parameterized by Eqn.6 are

$$\boldsymbol{V}_\nu^{obs} \;=\; [\mathsf{S}_\nu][\mathsf{F}]\boldsymbol{I}_\nu^{model} = \sum_{t=0}^{N_t}\sum_{s=0}^{N_s} w_\nu^t [\mathsf{S}_\nu][T_s][\mathsf{F}]\boldsymbol{I}_{\substack{s\\t}}^{sky} \tag{7}$$

where w_ν are Taylor-weights, $[\mathsf{S}_\nu]$ represents the spatial-frequency sampling function for frequency ν, and the image-domain convolution with $\boldsymbol{I}_s^{shp}$ is written as a spatial-frequency taper function $[T_s]_{m\times m} = diag([\mathsf{F}]\boldsymbol{I}_s^{shp})$.

Combining measurements from all frequencies together, we can write

$$\boldsymbol{V}^{obs} \;=\; \sum_{t=0}^{N_t}\sum_{s=0}^{N_s}[\mathsf{W}_\mathsf{t}^{\mathsf{mfs}}][\mathsf{S}][T_s][\mathsf{F}]\boldsymbol{I}_{\substack{s\\t}}^{sky} \tag{8}$$

where $[\mathsf{W}_\mathsf{t}^{\mathsf{mfs}}]$ 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 (ν), containing w_ν^t. The multi-frequency uv-coverage of the synthesis array is represented by $[\mathsf{S}_{nN_c\times m}]$.

The full measurement matrix therefore has the shape $nN_c \times mN_sN_t$, which when multiplied by the set of N_sN_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.

$$\begin{bmatrix} [A_{\substack{0\\0}}] & [A_{\substack{0\\1}}] & [A_{\substack{0\\2}}] & [A_{\substack{1\\0}}] & [A_{\substack{1\\1}}] & [A_{\substack{1\\2}}] \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{\substack{0\\0}}^{sky} \\ \boldsymbol{I}_{\substack{0\\1}}^{sky} \\ \boldsymbol{I}_{\substack{0\\2}}^{sky} \\ \boldsymbol{I}_{\substack{1\\0}}^{sky} \\ \boldsymbol{I}_{\substack{1\\1}}^{sky} \\ \boldsymbol{I}_{\substack{1\\2}}^{sky} \end{bmatrix} = \boldsymbol{V}^{obs} \tag{9}$$

where $\left[A_{\substack{p\\q}}\right] = [\mathsf{W}_q^{mfs}][\mathsf{S}][T_p][\mathsf{F}]$

for $p \in \{0, N_s-1\}$ and $q \in \{0, N_t-1\}$

2.5. Normal equations

The Normal Equations[5] can be written in block matrix form, with each block-row (for scale size s, and Taylor term t) is given by

$$\sum_{p=0}^{N_s-1}\sum_{q=0}^{N_t-1}\left[\mathsf{H}_{\substack{s,p\\t,q}}\right]\boldsymbol{I}_{\substack{p\\q}}^{sky} \;=\; \boldsymbol{I}_{\substack{s\\t}}^{dirty} \quad \forall\; s\in[0,N_s-1], t\in[0,N_t-1] \tag{10}$$

where each $\left[\mathsf{H}_{\substack{s,p\\t,q}}\right]$ is an $m\times m$ block of the Hessian matrix, and $\boldsymbol{I}_{\substack{s\\t}}^{dirty}$ is one of N_sN_t dirty images.

$$\begin{aligned}
\left[\mathsf{H}_{\substack{s,p\\t,q}}\right] &= \left[A_{\substack{s\\t}}^{\dagger}\right][\mathsf{W}^{\mathsf{im}}]\left[A_{\substack{p\\q}}\right] &(11)\\
&= [\mathsf{F}^\dagger T_s \mathsf{S}^\dagger \mathsf{W}_t^{mfs}][\mathsf{W}^{\mathsf{im}\,mfs}][\mathsf{W}_q^{mfs}\mathsf{S}T_p\mathsf{F}] &(12)\\
&= [\mathsf{F}^\dagger T_s \mathsf{F}][\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}_t^{mfs}\mathsf{W}^{\mathsf{im}\,mfs}\mathsf{W}_q^{mfs}\mathsf{S}\mathsf{F}][\mathsf{F}^\dagger T_p\mathsf{F}] &(13)\\
&= [\mathsf{F}^\dagger T_s \mathsf{F}]\left\{\sum_\nu w_\nu^{t+q}[\mathsf{F}^\dagger\mathsf{S}_\nu^\dagger\mathsf{W}_\nu^{\mathsf{im}}\mathsf{S}_\nu\mathsf{F}]\right\}[\mathsf{F}^\dagger T_p\mathsf{F}] &(14)\\
&= [\mathsf{F}^\dagger T_s \mathsf{F}]\left\{\sum_\nu w_\nu^{t+q}[\mathsf{H}_\nu\}\right][\mathsf{F}^\dagger T_p\mathsf{F}] &(15)
\end{aligned}$$

$[\mathsf{H}_\nu] = [\mathsf{F}^\dagger\mathsf{S}_\nu^\dagger\mathsf{W}_\nu^{\mathsf{im}}\mathsf{S}_\nu\mathsf{F}]$ is the Hessian matrix formed using only one frequency channel, and is a convolution operator containing a shifted version of the single-frequency PSF $\boldsymbol{I}_\nu^{psf} = diag[\mathsf{F}^\dagger\mathsf{S}^\dagger W\mathsf{S}]$ in each row (see footnote 5). $[\mathsf{F}^\dagger\mathsf{T}_s\mathsf{F}]$ is also a convolution operators with I_s^{shp} as the kernel. The process of convolution is associative and commutative, and therefore, $\left[\mathsf{H}_{\substack{s,p\\t,q}}\right]$ is also a convolution operator whose kernel is given by

$$\boldsymbol{I}_{\substack{s,p\\t,q}}^{psf} = I_s^{shp} \star \left\{\sum_\nu w_\nu^{t+q}\boldsymbol{I}_\nu^{psf}\right\} \star I_p^{shp} \tag{16}$$

[4] The measurement equation of an imaging instrument describes its transfer function (the effect of the measurement process on the input signal). For an ideal interferometer (a perfect spatial frequency filter, with no instrumental gains), it can be written in matrix notation as follows. Let $I_{m\times1}^{sky}$ be a pixelated image of the sky and let $V_{n\times1}^{obs}$ be a vector of n visibilities. Let $S_{n\times m}$ be a projection operator that describes the uv-coverage as a mapping of m discrete spatial frequencies (pixels on a grid) to n visibility samples (usually $n > m$). Let $F_{m\times m}$ be the Fourier transform operator. Then, $[\mathsf{S}_{n\times m}][F_{m\times m}]\boldsymbol{I}_{m\times1}^{sky} = \boldsymbol{V}_{n\times1}^{obs}$

[5] The Normal Equations are the linear system of equations whose solution gives a weighted least-squares estimate of a set of parameters in a model (χ^2 minimization). For an ideal interferometer, it is given by $[\mathsf{F}^\dagger\mathsf{S}^\dagger W\mathsf{S}\mathsf{F}]I_{m\times1}^{sky} = [\mathsf{F}^\dagger\mathsf{S}^\dagger W]V_{n\times1}^{obs} = I_{m\times1}^{dirty}$ where $W_{n\times n}$ is a diagonal matrix of signal-to-noise based measurement weights and $\mathsf{S}^\dagger$ denotes the mapping of measured visibilities onto a spatial frequency grid. The Hessian (matrix on the LHS) is by construction a circulant convolution operator with a shifted version of $I_{m\times1}^{psf} = diag[\mathsf{F}^\dagger\mathsf{S}^\dagger W\mathsf{S}]$ in each row. The dirty image on the RHS (produced by direct Fourier inversion of weighted visibilities) is therefore the convolution of $I_{m\times1}^{sky}$ with I^{psf}, and these equations can be solved by a deconvolution.

Therefore, to compute the Hessian matrix, it suffices to compute one such kernel per Hessian block.

The dirty images on the RHS of Eqn.10 can be written as follows.

$$\boldsymbol{I}^{dirty}_{\substack{s\\t}} = [\mathsf{F}^\dagger T_s \mathsf{F}][\mathsf{F}^\dagger \mathsf{S}^\dagger {\mathsf{W}^{\mathsf{mfs}}_{\mathsf{t}}}^\dagger \mathsf{W}^{\mathsf{im}}]\boldsymbol{V}^{obs} \tag{17}$$

$$= [\mathsf{F}^\dagger T_s \mathsf{F}]\left\{\sum_\nu w^t_\nu [\mathsf{F}^\dagger \mathsf{S}^\dagger_\nu \mathsf{W}^{\mathsf{im}}_\nu]\boldsymbol{V}^{obs}_\nu\right\} \tag{18}$$

$$= I^{shp}_s \star \left\{\sum_\nu w^t_\nu \boldsymbol{I}^{dirty}_\nu\right\} \tag{19}$$

where $\boldsymbol{I}^{dirty}_\nu = [\mathsf{F}^\dagger \mathsf{S}^\dagger_\nu \mathsf{W}_\nu]\boldsymbol{V}^{obs}_\nu$ is the dirty image formed by direct Fourier inversion of weighted visibilities from one frequency channel.

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 N_t Taylor coefficient images each of size $m \times 1$, for all N_s spatial scales.

The normal equations in block matrix form for the example in Eqn.9 for $N_t = 3, N_s = 2$ is shown in Eqn.20. The Hessian matrix consists of $N_s \times N_s = 2\times 2$ blocks (the four quandrants of the matrix), each for one pair of spatial scale s, p. Within each quadrant, the $N_t \times N_t = 3\times 3$ 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.

$$\begin{bmatrix}
\left[H_{\substack{0,0\\0,0}}\right] & \left[H_{\substack{0,0\\0,1}}\right] & \left[H_{\substack{0,0\\0,2}}\right] & \left[H_{\substack{0,1\\0,0}}\right] & \left[H_{\substack{0,1\\0,1}}\right] & \left[H_{\substack{0,1\\0,2}}\right] \\
\left[H_{\substack{0,0\\1,0}}\right] & \left[H_{\substack{0,0\\1,1}}\right] & \left[H_{\substack{0,0\\1,2}}\right] & \left[H_{\substack{0,1\\1,0}}\right] & \left[H_{\substack{0,1\\1,1}}\right] & \left[H_{\substack{0,1\\1,2}}\right] \\
\left[H_{\substack{0,0\\2,0}}\right] & \left[H_{\substack{0,0\\2,1}}\right] & \left[H_{\substack{0,0\\2,2}}\right] & \left[H_{\substack{0,1\\2,0}}\right] & \left[H_{\substack{0,1\\2,1}}\right] & \left[H_{\substack{0,1\\2,2}}\right] \\
\left[H_{\substack{1,0\\0,0}}\right] & \left[H_{\substack{1,0\\0,1}}\right] & \left[H_{\substack{1,0\\0,2}}\right] & \left[H_{\substack{1,1\\0,0}}\right] & \left[H_{\substack{1,1\\0,1}}\right] & \left[H_{\substack{1,1\\0,2}}\right] \\
\left[H_{\substack{1,0\\1,0}}\right] & \left[H_{\substack{1,0\\1,1}}\right] & \left[H_{\substack{1,0\\1,2}}\right] & \left[H_{\substack{1,1\\1,0}}\right] & \left[H_{\substack{1,1\\1,1}}\right] & \left[H_{\substack{1,1\\1,2}}\right] \\
\left[H_{\substack{1,0\\2,0}}\right] & \left[H_{\substack{1,0\\2,1}}\right] & \left[H_{\substack{1,0\\2,2}}\right] & \left[H_{\substack{1,1\\2,0}}\right] & \left[H_{\substack{1,1\\2,1}}\right] & \left[H_{\substack{1,1\\2,2}}\right]
\end{bmatrix}
\begin{bmatrix}
\boldsymbol{I}^{sky}_{\substack{0\\0}} \\ \boldsymbol{I}^{sky}_{\substack{0\\1}} \\ \boldsymbol{I}^{sky}_{\substack{0\\2}} \\ \boldsymbol{I}^{sky}_{\substack{1\\0}} \\ \boldsymbol{I}^{sky}_{\substack{1\\1}} \\ \boldsymbol{I}^{sky}_{\substack{1\\2}}
\end{bmatrix}
=
\begin{bmatrix}
\boldsymbol{I}^{dirty}_{\substack{0\\0}} \\ \boldsymbol{I}^{dirty}_{\substack{0\\1}} \\ \boldsymbol{I}^{dirty}_{\substack{0\\2}} \\ \boldsymbol{I}^{dirty}_{\substack{1\\0}} \\ \boldsymbol{I}^{dirty}_{\substack{1\\1}} \\ \boldsymbol{I}^{dirty}_{\substack{1\\2}}
\end{bmatrix} \tag{20}$$

This is the system of equations to be solved to obtain estimates of the model parameters $I^{sky}_{\substack{p\\q}}$. The spatial-frequency sampling of a real interfetometer is always incomplete ([S] is rank-deficient). Therefore, each Hessian block, and the entire Hessian matrix is singular, an exact inverse does not exist. An accurate reconstruction can be obtained only via successive approximation (iterative numerical optimization).

2.6. Principal Solution

An approximate solution of the normal equations can be computed via diagonal approximations of all Hessian blocks. Each Hessian block is a convolution operator with a shifted version of $I^{psf}_{\substack{s,p\\t,q}}$ in each row, and elements on the diagonal within each Hessian block are the same. The Hessian matrix with a diagonal approximation for each block $\left[\mathsf{H}_{\substack{s,p\\t,q}}\right]$ can be written as one $N_t N_s \times N_t N_s$ element matrix ([H^{peak}]), that applied to all pixels independently. An approximate solution of the normal equations can be obtained by inverting [H^{peak}] once, and applying it to all pixels of the dirty images, one pixel at a time.

This process of doing a pixel-by-pixel inversion of the diagonal-approximate Hessian matrix gives the principal solution[6] of the system. Such a solution will be correct only at the locations of the centers of flux components (source peaks), and must be augmented with an iterative optimization approach to ensure accuracy. In the case of perfect sampling (where the Hessian blocks are truly diagonal), the principal solution will directly give images of series-coefficients.

2.6.1. Properties of [H^{peak}]

1. 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.

$$H^{peak}_{\substack{s,p\\t,q}} = mid\left\{\boldsymbol{I}^{psf}_{\substack{s,p\\t,q}}\right\} = tr\left[\sum_\nu w^{t+q}_\nu [T_s \mathsf{S}^\dagger_\nu \mathsf{W}^{\mathsf{im}}_\nu \mathsf{S}_\nu T_p]\right] \tag{21}$$
$$\forall\ s, p \in \{0...N_s - 1\}\ ,\ \ t, q \in \{0...N_t - 1\}$$

2. The off-diagonal elements measure the orthogonality[7] between the various basis functions, 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.
3. The condition number of this matrix (or of blocks within this matrix) will indicate if the chosen set of basis functions and spatial-frequency coverage provide enough constraints to provide a stable solution, and can be used as a metric to choose a suitable basis set. For a simple example, if a 3-term solution is attempted with data from only two distinct frequencies, [H^{peak}] will be singular. Or, for some choice of multi-frequency *uv*-coverage, the visibilities measured by the instrument for two different spatial scales may become hard to distinguish. Then, the cross-term element of [H^{peak}] corresponding to this combination could have a higher value, indicating that the two parameters are highly coupled, and there is insufficient information in the data and sampling pattern to distinguish between the scales. A similar situation can arise to create ambiguity between spatial or spectral structure (an extreme example is multi-frequency measurements from only one baseline.
4. In general, [H^{peak}] will be a positive-definite symmetric matrix whose inverse can be easily computed *via* a Cholesky

[6] The principal solution (as defined in Bracewell & 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). It is also an approximate solution of the normal equations, calculated using a diagonal approximation of the Hessian (elements on the diagonal are the sum of weights). 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.

[7] The 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. 37).

decomposition[8]. Also, the value of N_s is usually < 10, making the inversion of $[H^{peak}]$ tractable.

5. Several approximations can be made about the structure of $[H^{peak}]$ to simplify its inversion, and it is important to understand the numerical implications of these trade-offs. One is a block-diagonal approximation of the full Hessian (*i.e.* using only those blocks of the Hessian in Eqn. 20 for which $s = p$; top-left and bottom-right quadrants). This approximation ignores the cross-terms between spatial scales and assumes that the scale basis functions are orthogonal. This is never true for a set of tapered truncated paraboloids, but this approximation works because of the iterative χ^2-minimization process. Now, a multi-frequency principal solution can be done separately on each remaining $N_t \times N_t$ block, one spatial scale at a time. This automatically does a normalization across scales that corresponds to a diagonal approximation of the multi-scale Hessian (see section **??** for alternate ways of computing the multi-scale solution). The update step of the iterative deconvolution still needs to evaluate the full LHS of the normal equations while subtracting out a flux component

2.7. MS-MFS algorithm

This section describes an iterative joint deconvolution process that solves the normal equations (Eqn.10) and produces a set of N_t series-coefficient images at N_s different spatial scales. The algorithm presented in this section is listed with more details in Algorithm 1 on the next page and Algorithm **??** on page ??)

Pre-compute Hessian : Convolution kernels for all distinct blocks in the $N_sN_t \times N_sN_t$ Hessian are evaluated *via* Eqn. 16. All kernels are normalized by w_{sum} such that the peak of $\boldsymbol{I}^{psf}_{\substack{0,0\\0,0}}$ is unity, and the relative weights between Hessian blocks is preserved. A set of N_s matrices each of shape $N_t \times N_t$ and denoted as $[H^{peak}_s]$ are constructed from the diagonal blocks of the full Hessian (blocks for which $s = p$ in Eqn. 20). Their inverses are computed and stored in $[H_s^{peak^{-1}}]$.

Initialization : All N_sN_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 residual images :** The RHS vectors (residual or dirty images) $\boldsymbol{I}^{dirty}_{\substack{s\\t}}\ \forall\ t \in \{0, N_t\text{-}1\}$ of the normal equations are computed *via* Eqn. 17 by first computing the multi-frequency dirty images and then convolving them by the scale basis functions.
2. **Find a Flux Component :** The principal solution is computed for all pixels, one scale at a time.

$$I_s^{pix,psol} = [H_s^{peak^{-1}}]\, I_s^{pix,dirty} \qquad \text{for each pixel, and scale } s \tag{22}$$

Here, $[H^{peak}_s]$ is the s^{th} block (of size $N_t \times N_t$) in the list of diagonal-blocks of $[H^{peak}]$, and $I_s^{pix,dirty}$ is the $N_t \times 1$ vector constructed from $\boldsymbol{I}^{dirty}_{\substack{s\\t}}\ \forall\ t \in \{0, N_t\text{-}1\}$.
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. Other heuristics may also be employed to make this choice, for example, pick the set of components that makes the largest impact on the value of χ^2. Let the scale size of this chosen subset 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 $\boldsymbol{I}^{shp}_{p,(i)}$. 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 $\left\{\boldsymbol{I}^{model}_{\substack{p\\q},(i)}\right\}; q \in [0, N_t]$.
3. **Update model images :** A single multi-scale model image is accumulated for each Taylor coefficient.

$$\boldsymbol{I}^{model}_q = \boldsymbol{I}^{model}_q + g\left(\boldsymbol{I}^{model}_{\substack{p\\q},(i)} \star \boldsymbol{I}^{shp}_p\right) \qquad \forall q \in [0, N_t] \tag{23}$$

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.

$$\boldsymbol{I}^{res}_{\substack{s\\t}} = \boldsymbol{I}^{res}_{\substack{s\\t}} - g\left(\sum_{q_i=0}^{N_t-1}\left[\boldsymbol{I}^{psf}_{\substack{s,p\\t,q}} \star \boldsymbol{I}^{model}_{\substack{p\\q},(i)}\right]\right) \tag{24}$$

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. **??** for multi-frequency imaging. Residual visibilities are computed as $\boldsymbol{V}^{res}_\nu = \boldsymbol{V}^{corr}_\nu - \boldsymbol{V}^{model}_\nu$.
Repeat from Step 1 until a global convergence criterion is satisfied.

Restoration : After convergence, the model spectral coefficient images can be interpreted in different ways. If required, the final image products can be smoothed with the restoring beam and the residuals are added back in.

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 correct 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 2.2). The

[8] A 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 $[A]\boldsymbol{x} = \boldsymbol{b}$ 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].

data products are images of the reference-frequency flux $\boldsymbol{I}^{sky}_{\nu_0}$, the spectral-index $\boldsymbol{I}^{\alpha}$ and the spectral curvature $\boldsymbol{I}^{\beta}$.

$$\boldsymbol{I}^{sky}_{\nu_0} = \boldsymbol{I}^{model}_0 \tag{25}$$
$$\boldsymbol{I}^{\alpha} = \boldsymbol{I}^{model}_1/\boldsymbol{I}^{model}_0 \tag{26}$$
$$\boldsymbol{I}^{\beta} = \left[\boldsymbol{I}^{model}_2/\boldsymbol{I}^{model}_0\right] - [\boldsymbol{I}^{\alpha}(\boldsymbol{I}^{\alpha}-1)/2] \tag{27}$$

Spectral index and curvature images can be calculated only in regions where the values in $\boldsymbol{I}^{model}_0$ are above a chosen threshold.

3. An image cube can be constructed by evaluating the spectral polynomial *via* Eqn. **??** 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 adding up the flux at all frequencies. This continuum image is different from the reference-frequency image which represents the flux measured at only one frequency.

Primary-beam correction : A correction for the average primary-beam and its frequency dependence can be done as a post-deconvolution step. The primary-beams are first evaluated or measured as a function of frequency, and the frequency-dependence per pixel modeled by a power-law or a polynomial (perferably the same spectral polynomial used for the image reconstruction). Primary-beam correction can be done as follows.

$$\boldsymbol{I}^{new}_{\nu_0} = \boldsymbol{I}^{m}_{\nu_0}/\boldsymbol{P_b}_{\nu_0} \tag{28}$$
$$\boldsymbol{I}^{new}_{\alpha} = \boldsymbol{I}^{m}_{\alpha} - \boldsymbol{P_b}_{\alpha} \tag{29}$$
$$\boldsymbol{I}^{new}_{\beta} = \boldsymbol{I}^{m}_{\beta} - \boldsymbol{P_b}_{\beta} \tag{30}$$

Note that if a polynomial is fit for the frequency-dependence of the primary beam, and $\boldsymbol{P_b}_{\nu_0}, \boldsymbol{P_b}_{\alpha}, \boldsymbol{P_b}_{\beta}$ computed from it, the above operation is numerically identical to doing a polynomial division in terms of two sets of coefficients (for $N_t <= 3$). A brute-force polynomial-division using more series coefficients will yield a more accurate solution. Note however, that such a correction will not be accurate if there are time-dependent variations in the primary-beam, and will require integration with the AW-Projection algorithm discussed in [**?**].

2.7.1. Software Implementation :

The MS-MFS algorithm described in section 2.7 has been implemented and released *via* the CASA[9] software package (version 2.4 onwards). The data products are N_t spectral-coefficient images, a spectral index image, and a curvature images (if $N_t > 2$). The parameters that control the algorithm are (a) ν_0 : a reference frequency chosen near the middle of the sampled frequency range, about which the Taylor expansion is performed, (b) N_t : the number of coefficients of the Taylor polynomial to solve for, and (c) N_s and $\boldsymbol{I}^{shp}_s$: a set of scale sizes in units of image pixels to use for the multi-scale representation of the image. The resulting wide-band image model can then be used within a standard self-calibration loop.

[9] http://casa.nrao.edu Common Astronomy Software Applications is being developed at the http://www.nrao.eduNational Radio Astronomy Observatory

Algorithm 1: MS-MFS Algorithm :

Data: calibrated visibilities : $\boldsymbol{V}^{corr}_{\nu}\ \forall \nu$
Data: *uv*-sampling function : $[\mathsf{S}_\nu]$
Data: image noise threshold and loop gain σ_{thr}, g_s
Data: scale basis functions : $\boldsymbol{I}^{shp}_s\ \forall s \in \{0, N_s-1\}$
Result: model coefficient images : $\boldsymbol{I}^m_q\ \forall q \in \{0, N_t-1\}$
Result: spectral index and curvature : $\boldsymbol{I}^m_\alpha, \boldsymbol{I}^m_\beta$

for $t \in \{0, N_t-1\}, q \in \{t, N_t-1\}$ **do**
    Compute the spectral PSF $\boldsymbol{I}^{psf}_{tq}$
    for $s \in \{0, N_s-1\}, p \in \{s, N_s-1\}$ **do**
        Compute the scale-spectral PSF $\boldsymbol{I}^{psf}_{\substack{sp\\tq}} = \boldsymbol{I}^{shp}_s \star \boldsymbol{I}^{shp}_p \star \boldsymbol{I}^{psf}_{tq}$
    end
end
for $s \in \{0, N_s-1\}$ **do**
    Construct $[H^{peak}_s]$ from $mid(I^{psf}_{\substack{s,s\\t,q}})$ and compute $[H^{peak^{-1}}_s]$
end
Initialize the model $\boldsymbol{I}^m_t$ for all $t \in \{0, N_t-1\}$ and compute $f_{sidelobe}$
repeat `/* Major Cycle */`
    for $t \in \{0, N_t\text{-}1\}$ **do**
        Compute the residual image $\boldsymbol{I}^{res}_t$
        for $s \in \{0, N_s\text{-}1\}$ **do**
            Compute $\boldsymbol{I}^{res}_{s,t} = \boldsymbol{I}^{shp}_s \star \boldsymbol{I}^{res}_t$
        end
    end
    Calculate f_{limit} from $\boldsymbol{I}^{res}_{0,0}$
    repeat `/* Minor Cycle */`
        for $s \in \{0, N_s\text{-}1\}$ **do**
            if *Peak of* $\boldsymbol{I}^{res}_{s,0} > 10\ \sigma_{thr}$ **then**
                foreach *pixel* **do**
                    Construct I^{rhs}_s, an $N_t \times 1$ vector from $I^{res}_{s,t}\ \forall\ t \in \{0, N_t\text{-}1\}$
                    Compute principal solution $I^{sol}_s = [H^{peak^{-1}}_s] I^{rhs}_s$
                end
                Choose $I^{sol} = max\{I^{sol}_{t=0},\ \forall\ s \in \{0, N_s\text{-}1\}\}$
            else
                Find the location of the peak in $\boldsymbol{I}^{res}_{s,0},\ \forall\ s \in \{0, N_s\text{-}1\}$
                Construct I^{rhs}_s, from $I^{res}_{s,t}$ for the chosen s, at this location
                Compute $I^{sol} = [H^{peak^{-1}}_s] I^{rhs}_s$ at this location
            end
        end
        for $t \in \{0, N_t-1\}$ **do**
            Update the model image : $I^m_t = I^m_t + g_s\, I^{shp}_{s_i} \star I^{sol}_t$
            for $s \in \{0, N_s\text{-}1\}$ **do**
                Update the residual image : $I^{res}_{s,t} = I^{res}_{s,t} - g \sum_{p=0}^{N_s-1} \sum_{q=0}^{N_t-1} [I^{psf}_{\substack{sp\\tq}} \star I^{sol}_q]$
            end
        end
    until *Peak residual in* $\boldsymbol{I}^{res}_{0,0} < f_{limit}$
    Compute model visibilities V^m_ν from $I^m_t\ \forall t \in \{0.N_t-1\}$
    Compute a new residual image I^{res} from residual visibilities $V^{corr}_\nu - V^m_\nu$
until *Peak residual in* $\boldsymbol{I}^{res}_0 < \sigma_{thr}$
Calculate $\boldsymbol{I}^m_{\nu_0}, \boldsymbol{I}^\alpha, \boldsymbol{I}^\beta$ from $I^m_t\ \forall t \in \{0.N_t-1\}$ and restore the results
If required, remove average primary beam : $\boldsymbol{I}^{new}_{\nu_0} = \boldsymbol{I}^m_{\nu_0}/\boldsymbol{P_b}_{\nu_0}, \boldsymbol{I}^{new}_\alpha = \boldsymbol{I}^m_\alpha - \boldsymbol{P_b}_\alpha, \boldsymbol{I}^{new}_\beta = \boldsymbol{I}^m_\beta - \boldsymbol{P_b}_\beta$

2.8. Relation to MF-Clean

A point-source multi-frequency deconvolution algorithm can be derived by setting $N_s = 1$ and using $I_0^{shp} = \delta$-function in the derivations in sections 2.4 and 2.5. This section discusses the difference between the resulting multi-frequency Hessian and RHS vectors and those described in the MF-Clean algorithm [Sault & Wieringa 1994], and shows that the MF-Clean approach will incur errors for arrays with dense and irregular spatial-frequency coverage where tracks from different baselines intersect.
The normal equations can be written in block matrix form (for example, for $N_t = 3$).

$$\begin{bmatrix} [H_{0,0}] & [H_{0,1}] & [H_{0,2}] \\ [H_{1,0}] & [H_{1,1}] & [H_{1,2}] \\ [H_{2,0}] & [H_{2,1}] & [H_{2,2}] \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_0^{sky} \\ \boldsymbol{I}_1^{sky} \\ \boldsymbol{I}_2^{sky} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_0^{dirty} \\ \boldsymbol{I}_1^{dirty} \\ \boldsymbol{I}_2^{dirty} \end{bmatrix} \tag{31}$$

where each block $[\mathsf{H}_{t,q}]$ is a convolution operator with $\boldsymbol{I}_{t,q}^{psf}$ as its kernel.

$$\boldsymbol{I}_{t,q}^{psf} = \sum_\nu w_\nu^{t+q} \boldsymbol{I}_\nu^{psf} \tag{32}$$

$$\boldsymbol{I}_t^{dirty} = \sum_\nu w_\nu^t \boldsymbol{I}_\nu^{dirty} \tag{33}$$

The MF-Clean algorithm described in [Sault & Wieringa 1994] follows a matched-filtering approach using functions called spectral-psfs, which are equivalent to the convolution kernels from the first row of Hessian blocks ($q = 0$) in Eqn.32. Hessian elements and RHS vectors are calculated by convolving spectral-psfs with themselves and the residual images.

$$\boldsymbol{I}_{t,q}^{psf} = \left\{ \sum_\nu w_\nu^t \boldsymbol{I}_\nu^{psf} \right\} \star \left\{ \sum_\nu w_\nu^q \boldsymbol{I}_\nu^{psf} \right\} \tag{34}$$

$$\boldsymbol{I}_t^{dirty} = \left\{ \sum_\nu w_\nu^t \boldsymbol{I}_\nu^{psf} \right\} \star \left\{ \sum_\nu \boldsymbol{I}_\nu^{dirty} \right\} \tag{35}$$

Formally, this matched filtering approach is exactly equal to the calculations shown in Eqns.32 and 33 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). Appendix **??** contains a derivation that demonstrates this.

In practice this difference manifests itself as follows. 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.32 and 33 require the computation of $(w_1V_1) + (w_2V_2)$. The two are equivalent only for flat spectrum sources where $V_1 = V_2$ or when there is no overlap between the spatial frequency grid cells measured from different observing frequencies (V_1 and V_2 map to different spatial frequencies).

The MF-Clean 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 in Eqns.32 and 33 eliminated this instability (determined using simulated VLA data).

2.9. Relation to MS-Clean

A narrow-band (or flat-spectrum) multi-scale deconvolution algorithm can be derived by setting $N_t = 1$ in the derivations in sections 2.4 and 2.5.
The normal equations can be written in block matrix form (for example, for $N_s = 2$).

$$\begin{bmatrix} [H_{0,0}] & [H_{0,1}] \\ [H_{1,0}] & [H_{1,1}] \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_0^{sky,\delta} \\ \boldsymbol{I}_1^{sky,\delta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_0^{dirty} \\ \boldsymbol{I}_1^{dirty} \end{bmatrix} \tag{36}$$

The elements of $[\mathsf{H}^{peak}]$, the $N_s \times N_s$ matrix formed from the multi-scale Hessian blocks are given by

$$H_{s,p}^{peak} = mid\left\{\boldsymbol{I}_{s,p}^{psf}\right\} = tr([T_s][\mathsf{S}^\dagger \mathsf{W} \mathsf{S}][T_p]) \quad \forall\, s,p \in \{0...N_s{-}1\} \tag{37}$$

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. With uniform weighting, the kernels $\boldsymbol{I}_{s,s}^{psf}$ on the diagonal blocks are the autocorrelations of the different scale functions $\boldsymbol{I}_s^{shp}$, as measured by the interferometer, and this gives the area under the main beam of the PSF for each spatial scale. The principal solution can be computed by inverting $[\mathsf{H}^{peak}]$ and applying it to the RHS vectors.

There are two differences between this approach, and those described in Cornwell [2008] and Greisen et al. [2009].

1. **Finding a flux component :** In both forms of MS-Clean, 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 size[10]. 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. Both these normalization schemes are approximations of using a diagonal approximation of $[H^{peak}]$ 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, we are able to get a more accurate estimate of the total-flux of the component at each scale than by just reading off a peak from a series of dirty images biased by the MS-Clean b_s. However, this solution gives correct values only at the locations of the centers

[10] g 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 $1.0/\sqrt{2\pi} = 0.398$.

of the flux-components, and introduces large errors in the psf sidelobes. Therefore, for reasons of stability, a diagonal approximation is a more appropriate choice (demonstrated on simulated VLA data).

2. **Minor cycle updates :** The update steps in Cornwell [2008] and Section **??** evaluate the full LHS of the normal equations 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 by going to the visibility domain in each iteration. 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 iterations.

A choice among these three methods will depend on trade-offs between the accuracy within each minor cycle iteration, the computational cost per step, and optimized global convergence patterns to control the total number of iterations.

3. Hybrids of Narrow-Band and Continuum Techniques

The preceeding sections discussed multi-frequency solutions that used data from all measured frequencies together, so as to take advantage of the combined spatial-frequency coverage. However, there are some situations where single-channel methods used in combination with multi-frequency-synthesis (and no built-in spectral model) will be able to deliver scientifically useful wide-band reconstructions at the continuum sensitivity.

The basic idea of a hybrid wide-band method 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).

1. Deconvolve each channel separately upto the single-channel sensitivity limit. At this stage, only sources brighter than σ_{chan} will be detected and deconvolved.
2. Remove the contribution of bright (spectrally varying) sources by subtracting out visibilities predicted from the model image cube. At this stage, the peak residual brightness is at the level of the single-channel noise limit σ_{chan}.
3. Perform regular multi-frequency-synthesis imaging (flat-spectrum assumption) on the continuum residuals to extract flux that lies between σ_{chan} and σ_{cont}. Accordiing to [**?**], and as shown in section XX, errors due to this flat-spectrum assumption become visible only above a dynamic range of 1000 (for $\alpha = -1.0$ and a 2:1 bandwidth ratio). Therefore, as long as the sensitivity improvement between a single-channel the full band is less than 1000, this second step will incur no errors even if the remaining flux has spectral structure. This requirement translates to $N_{chan} < 10^6$, which is usually satisfied[11]
4. Add model images from both steps, and restore the results. For unresolved sources, it may be appropriate to use a clean-beam fitted for the highest frequency, but in general, to not bias spectral information, all channels should be restored using a clean beam fitted to the lowest frequency in the range.

The advantages of this approach are its simplicity, and that it can handle wide-band reconstructions with band-limited signals and spectral-lines. The disadvantages are that the angular resolution of the images and spectral information will be restricted to that given by the lowest frequency (a factor of two worse than what is possible), and high-fidelity deconvolution of resolved structure requires sufficient single-frequency spatial-frequency coverage to unambiguously reconstruct all the spatial structure of interest, at all frequencies.

[11] Even if visibilities are measured at a very high frequency resolution, they can be averaged across frequency ranges upto the bandwidth-smearing limit for the desired image field-of-view.

4. Imaging results

4.1. Simulation

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 **??**) 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.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.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 **??**) 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 **??**). A 5σ flux threshold of about 20μJy was used as the termination criterion.

Results : The results from these imaging runs are shown in Fig. 3 (three Taylor coefficients), Figure 4 shows residual images over a larger region of the sky, and Fig. 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.

Fig. 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 $\boldsymbol{I}_{\nu_0}$. 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.

Fig. 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. 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. **??**). All images are displayed at the same flux scale.

Fig. 3. Reconstructed Taylor coefficient images : These images show the first three Taylor coefficients (similar to Fig. 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.

Fig. 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.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 μJy and off source is 5 μJy. 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 μJy. (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.

Fig. 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.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 **??** 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 **??** 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 $I(\nu_0)$ 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 5μJy (close to the theoretical RMS of 3 μJy as listed in Table **??**). 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 200μJy despite an apparent decrease in the residuals during the minor cycle iterations. The off source RMS in the result is about 50 μJy.

4.2. Multi-frequency VLA observations 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 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 & 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.

Observations : Wide-band data were taken as described in Table **??** 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.6. The data were inspected visually and visibilities that were affected by strong radio frequency interference were flagged (masked).

Fig. 6. 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.

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 differently[12]. This caused errors in the correlator in-

[12] 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

put 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 idenfity 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 **??** 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.

1. **MS-MFS :** The MS-MFS algorithm was run with a 2^{nd}-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 **??**). 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 **??** 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 7 shows the reconstructed total-intensity images (top row) and the residual images (bottom row) obtained from these two methods. Figure 8 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 **??** 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 **??** (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

EVLA antennas when the source being observed was bright enough to contribute to increasing the overall system temperature.

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.

4.3. 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. 1996; Owen et al. 2000]. This central region is surrounded by a large diffuse radio halo (7 to 14 arcmin) with many bright narrow filaments ($\approx 10'' \times 3'$).

Multi-frequency VLA data were taken in a similar way as described for Cygnus-A, with a series of 10 snapshots at 16 different frequencies within the sensitivity range of the EVLA L-band receivers.

Fig.9 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.10 shows a plot of the spectrum formed from the integrated flux in the central bright region.

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.

Fig. 7. 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 **??**. 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.

Fig. 8. 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 **??**. 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.

The spectral index map[13] of the bright central region (at 3 arcsec resolution) shows a near flat-spectrum core with $\alpha_{LL} = -0.25$, a jet with $\alpha_{LL} = -0.5$ and lobes with $-0.6 > \alpha_{LL} > -0.7$.

[13] The 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

Fig. 9. 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 $\triangle\alpha = 0.2$ 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).

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.

This bright central region had sufficient (>100) signal-to-noise to be able to detect spectral curvature. The third panel in Fig.9 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 $\triangle\alpha = \beta\frac{\triangle\nu}{\nu_0} \approx -0.2$.

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 $\triangle\alpha \approx 0.2$) are consistent with these independent calculations.

The points in Fig.10 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 $\triangle\alpha \approx 0.2$ 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.

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.

Fig. 10. 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 $\triangle\alpha \approx 0.2$ 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)].

5. MS-MFS error estimation and feasibility

5.1. Dynamic-range vs N_t

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 **??**) 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 them[14].

5.1.1. 3C286 field (EVLA)

- frequency dependence of the primary-beam of the EVLA..... - -1.4 at the HPBW (from SW paper and from EVLA sims).

- spectral indices of surrounding sources : xx yy zz..... corrected to xxxxx. Confirmed by a second test observation pointed at those background sources.

[14] Note 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.

5.1.2. Error as a function of bandwidth ratio

ADD FIGURE

ALSO REFER TO CCW formula. Magnitude of diff Taylor terms : Are they visible in image as errors (vs) SNR needed to solve.

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 ratios[15]for these 5 datasets were 100%(3:1), 66%(2:1), 50%(1.67:1), 25%(1.28:1), 10%(1.1:1).

Figure.11 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. **??**). All these datasets were imaged using a maximum of 10 iterations, a loop-gain of 1.0, natural weighting and a flux threshold of $1.0\mu Jy$. 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 0^{th} 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.

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 anre 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. **??** (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. **??** (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 $\alpha << 1$, 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.

5.2. Error on Spectral Index

5.2.1. Propagation of deconvolution errors

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 **??** 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 curvature by an order of magnitude.

5.2.2. Overlapping sources

5.2.3. Errors as a function of SNR

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 **??** discusses empirically derived error bars for the spectral index based on these factors. These results show that to measure $\alpha = -1.0$ with less than 100% error bars, we need a single-channel signal-to-noise ratio greater than about 6, and to measure β with less than 100% error, we need a single-channel signal-to-noise of greater than about 100. Note that these numbers dictate the required accuracy of wide-band flux calibration.

REFER TO CCW BIAS : Error due to using too few Taylor terms to fit the exp with a poly. CCW comment on a bias that occurs with a 2-term T-exp. This is just the use of insufficient terms of a polynomial to model an exponential.

The errors on the polynomial coefficients and quantities derived from them will depend on the number of measurements

[15] There 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 $(\nu_{high} - \nu_{low})/\nu_{mid}$ espressed as a percentage. For example, the bandwidth ratio for $\nu_{low} = 1.0$ GHz, $\nu_{high} = 2.0$ GHz is $2 : 1$ and 66%.

Fig. 11. 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 α=-1.0 was imaged using Taylor polynomials of different orders ($N_t = 1-7$) and a linear spectral basis (Eqn. **??**). 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 $1\mu Jy$. 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 5^{th} or 6^{th} order Taylor expansion is most appropriate (when a linear spectral basis is used).

Fig. 12. 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 3^{rd} or 4^{th} order Taylor expansion is most appropriate.

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_{ν_0}, α and β can be obtained from the first three coefficients of a Taylor expansion of a power-law with varying index (Eqns.**??** and **??**), a higher order polynomial may be required during the fitting process to improve the accuracy of the first three coefficients[16]. In the case of very noisy spectra, errors can also arise from attempting to use too many terms in the polynomial fit.

Figure 12 illustrates the above trends for the value of α derived from a polynomial fit to a spectrum constructed *via* Eqn. 3 ($I_0^{true} = 10.0, \alpha^{true} = -1.5, \beta^{true} = -0.5, \nu_0 = 2.4$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.**??**,**??**) 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.

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 **??**), 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).

[16] Conway 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.

Fig. 13. 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 decreaseing 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.

5.3. 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.

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 13 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 apparant only in the higher frequencies where the top source (with a positive spectral index) is brighter.

Fig. 14. 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.

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 14 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.**??** 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.

5.4. Emission at large spatial scales

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 unconstrained 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 kλ or 10.3 arcmin at 1.0 GHz, and 1.3 kλ 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 kλ at the reference frequency of 2.5 GHz), and one steep spectrum point-source (α=-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. 15 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. 16 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.

5.5. Band-limited signals

spectral lines, continuum subtraction....

5.6. 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.**??** and **??**). 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.

Fig. 15. 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 <25m in addition to b >100m 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 16 shows the images that resulted from these tests.

Fig. 16. 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.

6. Discussion

1. Summary : Can do this'n'that (image and spectrum reconstruction + astrophysical parameters) ...at the angular resolution defined at the highest measured frequency.
2. Single-Channel Imaging Hybrid vs MS-MFS
3. Astrophysics : A variety of astrophysical observations could gain from these new instruments due to the increased sensitivity and large instantaneous bandwidth. Synchrotron spectra can be measured within the instantaneous bandwidth, to detect/measure turnovers/breaks. Snapshot imaging of highly variable sources (sunspots, supernovae, AGN) can be improved with the increased uv-coverage of mfs while deriving the instantaneous spectra at multiple spatial scales. Weak supernova remnants could be detected against a continuum background by their spectral signature. Spectral information can be obtained at the highest angular resolution. This is especially significant for moderately resolved sources. The atmospheres of stars have an angular size of xxx and are resolvable at xxx GHz in the middle of the EVLA frequency coverage. MFS with data between xx and yy GHz could yield wideband spectra at the angular resolution defined by yy GHz.
   Sunspot flares - magnetic loops - frequency probes different depths - band-limited signals.
   VLBI position offsets as a function of frequency - can be resolved (defined) via MFS.
   Multiple arrays can be combined by matching spatial frequency coverage across wide frequency ranges. For example, EVLA-D between 5 and 50 GHz would match with EVLA-C in L and S bands... (as has always been done, but it can be imaged together - not sure what final advantage this will bring though.. Snapshot imaging with EVLA subarrays could also achieve this.
4. Software (CASA and ASKAPSOFT), Cost and Parallelization (Hybrid vs MSMFS)
5. Limitations - moderately resolved sources will need simultaneous deconvolution of multipel components (ASP) - unresolved at lower freq, resolved at higher - 1D simulations show that it is possible to reconstruct, if at least XXX fraction of the bw has resolved info.
6. Future Directions Out into the null of the PB (alternate parameterisations). Full-Stokes imaging (ref SW)... parameterize differently in lm and freq but same idea.

Acknowledgements. The authors would like to thank the NRAO and NMT for support during the PhD thesis project that resulted in this algorithm and implementation in a usable software package. Many thanks to XX, YY, ZZ,... for useful discussions.

References

Bhatnagar, S. & Cornwell, T. J. 2004, Astron. & Astrophys., 426, 747

Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, Astron. & Astrophys., 487, 419

Bong, S.-C., Lee, J., Gary, D. E., & Yun, H. S. 2006, ApJ, 636, 1159

Bracewell, R. N. & Roberts, J. A. 1954, Australian Journal of Physics, 7, 615

Carilli, C. L. & Barthel, P. D. 1996, Astron. & Astrophys. Review, 7, 1

Carilli, C. L., Perley, R. A., Dreher, J. W., & Leahy, J. P. 1991, ApJ, 383, 554

Conway, J. E., Cornwell, T. J., & Wilkinson, P. N. 1990, mnras, 246, 490

Cornwell, T., Braun, R., & Briggs, D. S. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 151

Cornwell, T. J. 2008, IEEE Journal of Selected Topics in Sig. Proc., 2, 793

Greisen, E. W., Spekkens, K., & van Moorsel, G. A. 2009, Astron. J., 137, 4718

Likhachev, S. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 340, Future Directions in High Resolution Astronomy, ed. J. Romney & M. Reid, 608

Owen, F. N., Eilek, J. A., & Kassim, N. E. 2000, ApJ, 543, 611

Press, W., Flannery, B., Teukolsky, S., & Vetterling, W. 1988, Numerical Recipes in C. (Cambridge: Press Syndicate of the University of Cambridge)

Rau, U., Bhatnagar, S., Voronkov, M. A., & Cornwell, T. J. 2009, IEEE Proceedings, 97, 1472

Rottmann, H., Mack, K., Klein, U., & Wielebinski, R. 1996, Astron. & Astrophys., 309, L19

Sault, R. J. & Wieringa, M. H. 1994, Astron. & Astrophys. Suppl. Ser., 108, 585

.0.1. Difference with SW-MF-Clean

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 : $(w_1 + w_2)(V_1 + V_2)$.

$$\boldsymbol{I}^{psf,sw}_{t,q} = \boldsymbol{I}^{psf}_{t} \star \boldsymbol{I}^{psf}_{q} \quad \text{where} \quad \boldsymbol{I}^{psf}_{x} = [\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{mfs}_{x}\mathsf{W}^{\mathsf{im}}]\mathbf{1} \;\text{ for }\; x \in t,q \tag{.38}$$

$$\boldsymbol{I}^{dirty,sw}_{t} = \boldsymbol{I}^{psf}_{t} \star \boldsymbol{I}^{dirty} \quad \text{where} \quad \boldsymbol{I}^{dirty} = [\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{im}}]\boldsymbol{V}^{corr} \tag{.39}$$

According to Eqns.**??** to **??** (MFCLEAN algorithm), the Hessian block kernels and dirty images are to be computed by multiplying the visibility measurments with the Taylor-weights before gridding the result : $(w_1V_1) + (w_2V_2)$.

$$\boldsymbol{I}^{psf}_{t,q} = [\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{mfs}}_{\mathsf{t}}\mathsf{W}^{\mathsf{mfs}}_{\mathsf{q}}\mathsf{W}^{\mathsf{im}}]\mathbf{1} \tag{.40}$$

$$\boldsymbol{I}^{dirty}_{t} = [\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{mfs}}_{\mathsf{t}}\mathsf{W}^{\mathsf{im}}]\boldsymbol{V}^{corr} \tag{.41}$$

Conditions for equality The two methods listed above (Eqns..38,.39 and Eqns..40,.41) are equivalent only under certain conditions. Consider Eqn. **??** for $[H_{t,q}]$. Iff $[\mathsf{S}^\dagger] = [\mathsf{S}^\dagger\mathsf{S}\mathsf{S}^\dagger]$ and $[\mathsf{S}] = [\mathsf{S}\mathsf{S}^\dagger\mathsf{S}]$, then $[\mathsf{S}^\dagger]$ can be replaced by $[\mathsf{S}^\dagger\mathsf{S}\mathsf{S}^\dagger]$. Further, $[\mathsf{S}\mathsf{S}^\dagger]$ and $[\mathsf{W}^{\mathsf{mfs}^\dagger}_{\mathsf{t}}]$ are both diagonal matrices of size $nN_c \times nN_c$ and therefore commute. In this case, $[H_{t,q}]$ becomes

$$\begin{aligned}
[H_{t,q}] &= [\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{mfs}^\dagger}_{\mathsf{t}}\mathsf{W}^{\mathsf{im}}\mathsf{W}^{\mathsf{mfs}}_{\mathsf{p}}\mathsf{S}\mathsf{F}] \qquad (.42)\\
&= [\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{mfs}^\dagger}_{\mathsf{t}}\mathsf{S}\mathsf{F}][\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{im}}\mathsf{S}\mathsf{F}][\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{mfs}}_{\mathsf{p}}\mathsf{S}\mathsf{F}]\\
\Rightarrow \quad \boldsymbol{I}^{psf}_{tq} &= I^{psf}_{t} \star I^{psf} \star I^{psf}_{q} \qquad (.43)\\
&\text{where} \quad I^{psf} = [\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{im}}]\mathbf{1} \quad \text{and} \quad I^{psf}_{t} = [\mathsf{F}^\dagger\mathsf{S}^\dagger\mathsf{W}^{\mathsf{mfs}}_{\mathsf{t}}]\mathbf{1}
\end{aligned}$$

This is still not the same as Eqn. .38 which has two instances of $[\mathsf{W}^{\mathsf{im}}]$. Therefore, only when $[\mathsf{W}^{\mathsf{im}}]$ is an identity matrix (equally weighted visibilities) will the kernel functions from both methods be identical $\boldsymbol{I}^{psf}_{t,q} = \boldsymbol{I}^{psf,sw}_{t,q}$. A similar argument holds for the dirty images. The restriction of $[\mathsf{S}^\dagger] = [\mathsf{S}^\dagger\mathsf{S}\mathsf{S}^\dagger]$ and $[\mathsf{S}] = [\mathsf{S}\mathsf{S}^\dagger\mathsf{S}]$ implies that each row and column in $[\mathsf{S}]$ has only one 1, with the rest being 0. Since $[\mathsf{S}]$ 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 $[W^G_\nu] = [\mathsf{S}^\dagger_\nu\mathsf{W}^{\mathsf{im}}_\nu\mathsf{S}_\nu]$ per frequency channel. A projection operator $[\mathsf{S}_\nu{}^G]$ 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 $[\mathsf{S}_\nu{}^G]$, with an increased weight in $[W^G_\nu]$. The use of uniform weighting will flatten out $[W^G_\nu]$ as required for equality with Eqn. .38. Written this way, with multiple frequencies, $[\mathsf{S}_\nu{}^G]$ has dimensions $mN_c \times m$, and the restriction of $[\mathsf{S}^\dagger] = [\mathsf{S}^\dagger\mathsf{S}\mathsf{S}^\dagger]$ and $[\mathsf{S}] = [\mathsf{S}\mathsf{S}^\dagger\mathsf{S}]$ 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.