INVITED PAPER

Advances in Calibration and Imaging Techniques in Radio Interferometry

Several major calibration and image reconstruction algorithms used in radio interferometry are summarized using a common mathematical framework.

By Urvashi Rau, Sanjay Bhatnagar, Maxim A. Voronkov, and Tim J. Cornwell

ABSTRACT | This paper summarizes some of the major calibration and image reconstruction techniques used in radio interferometry and describes them in a common mathematical framework. The use of this framework has a number of benefits, ranging from clarification of the fundamentals, use of standard numerical optimization techniques, and generalization or specialization to new algorithms.

KEYWORDS | Algorithms; calibration; computing; imaging; radio interferometry

I. INTRODUCTION

The theory and practice of radio interferometry, including data processing, is well advanced and has been the subject of a graduate-level textbook [1].[1] In this paper, we aim to summarize recent advances in the theory and practice of calibration and imaging, arising from the work of several of the authors over the past ten years. We draw upon a number of our papers, placing the results in a common framework and nomenclature. We also present a number of new insights and algorithms arising in recent work.

The last decade has seen a substantial growth in the number and diversity of radio synthesis telescopes' being constructed. Examples include the Expanded Very Large Array (EVLA) [2], the Low Frequency Array (LOFAR) [3], the Square Kilometre Array (SKA) [4], the Australian Square Kilometre Array Pathfinder (ASKAP) [5], and the Karoo Array Telescope (MeerKAT) [6]. These telescopes bring both new science and new technical challenges. Prime amongst these challenges are:

- theory to describe new observing modalities and previously ignorable effects;
- algorithms to solve the resulting equations;
- a required increase in algorithmic performance in terms of sensitivity and dynamic range;
- a large increase (hundreds or thousands) in data volume;
- the need for algorithms adapted to high-performance computing, particularly the shift to highly parallel or concurrent processing.

The concept of a *measurement equation* is key to our work. Hamaker *et al.* [7] were particularly notable in emphasizing the importance of a single equation to describe a measurement process (as opposed to, say, a set of loosely related equations).

Section II describes the measurement equation in radio interferometry. Section III describes the solution of the measurement equation as an optimization problem and describes standard algorithms and methods used to solve it—calibration of direction independent instrumental effects and imaging using a point-source flux model. Section IV describes recent advances in algorithms that

Manuscript received November 10, 2008; revised January 8, 2009. First published June 23, 2009; current version published July 15, 2009.

U. Rau is with the Department of Physics, New Mexico Institute of Mining and Technology, Socorro, NM 87801 USA; the National Radio Astronomy Observatory, Socorro, NM 87801 USA; and the CSIRO Australia Telescope National Facility, Marsfield, NSW, Australia (e-mail: rurvashi@aoc.nrao.edu; urvashi.rau@gmail.com).

S. Bhatnagar is with the National Radio Astronomy Observatory, Socorro, NM 87801 USA (e-mail: bhatnagar.sanjay@gmail.com).

M. A. Voronkov and **T. J. Cornwell** are with the CSIRO Australia Telescope National Facility, Marsfield, NSW, Australia (e-mail: maxim.voronkov@csiro.au; tim.cornwell@csiro.au).

Digital Object Identifier: 10.1109/JPROC.2009.2014853

[1]This book is recommended for detailed descriptions of the fundamentals.

account for direction dependent instrumental effects during imaging. Section V describes recent advances in deconvolution algorithms.

II. MEASUREMENT EQUATION IN RADIO INTERFEROMETRY

Aperture synthesis is an indirect imaging technique where the spatial Fourier transform of an image is measured via its mutual coherence function. A radio interferometer [8] consists of a collection of spatially separated antennas. The aperture plane of the interferometer is the plane perpendicular to the instantaneous direction from the array to a reference point on the sky $\vec{s}_0$ called the phase-reference center. A *baseline* $\vec{b}_{ij}$ is defined as the vector between the three-dimensional locations of two antennas i and j, projected onto this aperture plane. The components of $\vec{b}_{ij}$ are measured in units of wavelength λ and denoted as u, v, w, where u, v are two-dimensional (2-D) spatial frequencies and w describes the height of an antenna relative to the plane of the array in the direction of $\vec{s}_0$. For electromagnetic radiation from a spatially incoherent brightness distribution, the mutual coherence function is defined as the time-averaged cross-correlation product of the total electric field measured at two aperture points (antennas) with a time delay between the measurements, and is given by

$$\Gamma(\vec{b}) = \int \left\langle E(\vec{s}, t) \cdot E^*(\vec{s}, t - \vec{b} \cdot \vec{s}/c) \right\rangle e^{-2\pi i \vec{b} \cdot \vec{s}/\lambda} d\Omega \qquad (1)$$

where $\vec{s} = \vec{s}_0 + \vec{\sigma}$ describes a point near the phase reference center, $E(\vec{s}, t)$ is the complex amplitude of the radiation emanating from a source in the direction $\vec{s}$, $\vec{b} \cdot \vec{s}/c$ is the time difference between the incoming radiation collected at two antennas separated by $\vec{b}$, and $d\Omega = d\vec{s}/R^2$, where R is the distance between the source and the aperture plane.

Signals from all antennas are delay corrected by a common factor given by $\vec{b} \cdot \vec{s}_0/c$ to steer the array towards $\vec{s}_0$. If the maximum remaining delay $\vec{b} \cdot \vec{\sigma}/c$ is smaller than the signal coherence time, the term in the angle brackets becomes the source autocorrelation function or the three-dimensional source brightness distribution $I(l, m, n)$, where $l, m, n = \sqrt{1 - l^2 - m^2}$ are direction cosines describing $\vec{\sigma}$. Equation (1) becomes

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

When the array is coplanar ($w \approx 0$), or the region of the sky being imaged may be assumed flat ($n \approx 1$), (2) describes a 2-D spatial Fourier transform relation between the mutual coherence function and the source brightness. This is the Van Cittert Zernike theorem [1] and forms the basis for interferometric imaging.

To measure polarized radiation [7], two nominally orthogonal components of the incident electric field $\vec{E}_i = [E^p \; E^q]_i^T$ are measured at each antenna i. Four cross-correlation pairs (two cross-hand and two parallel-hand) are formed per baseline as $\langle \vec{E}_i \otimes \vec{E}_j^* \rangle$. The resulting coherence vector is denoted as $\vec{V}_{ij} = [V^{pp} \; V^{pq} \; V^{qp} \; V^{qq}]_{ij}^T$. The vector of images corresponding to the four correlations is $\vec{I} = [I^{pp} \; I^{pq} \; I^{qp} \; I^{qq}]^T$ and is related to the standard Stokes vector of images by a linear transform.

The measured incoming radiation is modified by propagation effects and receiver electronics. Jones matrices describe this modulation for the electric field incident at each orthogonal pair of feeds $\vec{E}_i = [E^p \; E^q]_i^T$. Direction-independent effects for antenna i are described as $J_i^{\text{vis}} = [GDC]$, a 2×2 matrix product of complex antenna gains (G), polarization leakage (D), and feed configuration (C). Direction-dependent effects are described by $J_i^{\text{sky}} = [EPF]$, a product of antenna illumination patterns (E), parallactic angle effects (P), and tropospheric and ionospheric effects and Faraday rotation (F). The effect on each baseline ij is described by the outer product of these antenna-based Jones matrices given by $K_{ij}^{\{\text{vis,sky}\}} = [J_i \otimes J_j^\dagger]^{\{\text{vis,sky}\}}$, a 4×4 matrix. (In this paper, the $\dagger$ superscript is used to denote conjugate transpose or operator adjoint.)

The measurement equation [8] for one baseline (spatial frequency), one frequency channel, and one integration timestep is given by

$$\vec{V}_{ij}^{\text{obs}} = \left[K_{ij}^{\text{vis}}\right] \int \left[K_{ij}^{\text{sky}}\right] \vec{I}^{\text{sky}}(\vec{s}) e^{-2\pi i \vec{b} \cdot \vec{\sigma}/\lambda} d\Omega. \qquad (3)$$

All instrumental and propagation effects described by K_{ij} need to be corrected during image reconstruction.

So far, we have dealt with the signals measured at only one baseline. With n_{ant} antennas, there are $n_{\text{ant}}(n_{\text{ant}} - 1)/2$ baselines that make simultaneous measurements at multiple spatial frequencies. The spatial frequency plane can be further sampled by varying the positions of the antennas with respect to the direction of the phase-reference center. For ground-based arrays, the Earth's rotation makes all projected baseline vectors $\vec{b} \cdot \vec{s}_0$ trace ellipses on the spatial frequency plane, slowly filling it up. Measurements at multiple receiver frequencies also increase the sampling of the spatial-frequency plane. Measurements must be made at sufficiently high time and frequency resolution to prevent smearing (averaging of visibility data) on the spatial frequency plane. The result is generally a centrally dominated *uv*-plane sampling pattern with a hole in the middle and tapered outer edges. This is the transfer function of the synthesis array and is called the *uv-coverage* (see [8]).

The complete measurement equation can be written in matrix notation to include the effect of the uv-coverage. Let $I^{\rm sky}_{m\times 1}$ be a pixelated image of the sky and let $V^{\rm obs}_{n\times 1}$ 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 and c be the number of measured correlations (one, two, or all four of $\{pp, pq, qp, qq\}$). The measurement equation in block matrix form is

$$\vec{V}^{\rm obs}_{cn\times 1} = \left[K^{\rm vis}_{cn\times cn}\right]\left[S_{cn\times cm}\right]\left[F_{cm\times cm}\right]\left[K^{\rm sky}_{cm\times cm}\right]\vec{I}^{\,\rm sky}_{cm\times 1}. \quad (4)$$

Writing this completely in the spatial frequency domain

$$\vec{V}^{\rm obs}_{cn\times 1} = \left[K^{\rm vis}_{cn\times cn}\right]\left[S_{cn\times cm}\right]\left[G_{cm\times cm}\right]\vec{V}^{\rm sky}_{cm\times 1} \quad (5)$$

where $[G_{cm\times cm}] = [F_{cm\times cm}][K^{\rm sky}_{cm\times cm}][F^{\dagger}_{cm\times cm}]$ is a convolution operator[2] in the spatial frequency domain with $[FK^{\rm sky}]$ as the convolution filter.

All discussions that follow are of numerical algorithms, described within a mathematical framework amenable to implementation using standard optimization software.

III. STANDARD CALIBRATION AND IMAGING

This section describes the solution of the measurement equation as a numerical optimization problem. The measurement equation for a single correlation with no direction-dependent terms is given as

$$V^{\rm obs}_{n\times 1} = \left[K^{\rm vis}_{n\times n}\right]\left[S_{n\times m}F_{m\times m}\right]I^{\rm sky}_{m\times 1}. \quad (6)$$

Consider only the pp correlation product, and let the complex gains per antenna i be given by $[G_i] = g^p_i$. Then $K^{\rm vis}_{ij} = G_i \otimes G_j = g^p_i g^{*p}_j$ is a scalar and $[K^{\rm vis}_{n\times n}]$ is a diagonal matrix.

The unknowns in (6) are the sky brightness $I^{\rm sky}$ and the complex gain product for all visibilities $K^{\rm vis}$. A two-stage χ^2 minimization process iterates between these two parameter subspaces and applies constraints appropriate to the different physics involved. Calibration (Section III-A) is the process of computing and applying the inverse of $[K^{\rm vis}]$. Imaging (Section III-B) is the process of reconstructing the sky brightness $I^{\rm sky}$ by removing the effect of the instrument's incomplete spatial frequency sampling.

A. Calibration

The elements of $[K^{\rm vis}]$ are first estimated from observations of a source whose structure is known a priori ($V^{\rm model}_{n\times 1}$) by solving (5) in the form

$$V^{\rm obs}_{n\times 1} = \left[K^{\rm vis}_{n\times n}\right]V^{\rm model}_{n\times 1}. \quad (7)$$

A weighted least squares solution [9] of (7) is found by minimizing $\chi^2 = \sum_{ij} w_{ij}\left|V^{\rm obs}_{ij} - g_i g^*_j V^{\rm model}_{ij}\right|^2$, where w_{ij} is a measured visibility weight (inverse of noise variance) and $V^{\rm model}_{n\times 1}$ provides $O(n^2_{\rm ant})$ constraints to uniquely factor the baseline-based $K^{\rm vis}$ into $n_{\rm ant}$ antenna-based complex gains. In cases where the measurements at each baseline contain random additive noise that cannot be factored into antenna-based terms *(closure noise)*, a baseline-based calibration is sometimes done to solve for the elements of $[K^{\rm vis}]$ directly.

$[K^{\rm vis}_{n\times n}]^{-1}$ is reconstructed from these solutions and applied to the observed visibilities to *correct* them

$$V^{\rm corr}_{n\times 1} = \left[K^{\rm vis}_{n\times n}\right]^{-1}V^{\rm obs}_{n\times 1}. \quad (8)$$

To increase the signal-to-noise ratio of correlations going into the algorithm, the visibility data are sometimes preaveraged along data axes over which the solution is likely to remain stable.

B. Imaging

Using (6) and (8), the measurement equation after calibration is

$$\left[S_{n\times m}F_{m\times m}\right]I^{\rm sky}_{m\times 1} = V^{\rm corr}_{n\times 1}. \quad (9)$$

A weighted least squares estimate of $I^{\rm sky}$ is found by solving the *normal equations*

$$\left[F^{\dagger}S^{\dagger}WSF\right]I^{\rm sky}_{m\times 1} = \left[F^{\dagger}S^{\dagger}W\right]V^{\rm corr}_{n\times 1} \quad (10)$$

where $W_{n\times n}$ is a diagonal matrix of signal-to-noise based measurement weights and $S^{\dagger}$ denotes the mapping of measured visibilities onto a spatial frequency grid.

The Hessian $[F^{\dagger}S^{\dagger}WSF]$ on the left-hand side of (10) describes the imaging properties of the instrument, and

[2]A one-dimensional (1-D) convolution operator is constructed as follows. Consider $\vec{a}\star\vec{b}$. Let $[A]$ and $[B]$ be diagonal matrices constructed from vectors $\vec{a}$ and $\vec{b}$, respectively. Then $A\star B = F^{\dagger}(FA)(FB) = [F^{\dagger}A_F F]B = [C][B]$. Here, $diag(A_F)$ is the discrete Fourier transform (DFT) of $\vec{a}$ and $[C]$ is a Toeplitz matrix, with each row containing a shifted version of $\vec{a}$. Multiplication of $[C]$ with $\vec{b}$ implements the shift-multiply-add sequence required for the process of convolution. Since F is unitary, the singular value decomposition of $[C]$ is given by the Fourier transform, making it circulant. For a 2-D convolution, $[F]$ is the outer product of two 1-D DFT operators and $[C]$ is block-circulant with circulant blocks.

the right-hand side (RHS) describes the raw image produced by direct Fourier inversion of the calibrated visibilities. When $V^{\mathrm{corr}}_{n\times 1} = \vec{1}_{n\times 1}$, the RHS gives the *impulse response function* or *point spread function* of the instrument (I^{psf}), defined as the image produced when observing a point-source of unit brightness at the phase center. The Hessian is by construction a circulant convolution operator with a shifted version of I^{psf} in each row. Therefore, the *dirty image* produced by direct Fourier inversion of the measurements is the convolution of the true image I^{sky} with the point spread function (PSF) of the instrument, and the normal equations can be solved via a deconvolution.

Since S represents an incomplete sampling of spatial frequencies (column rank of $S_{n\times m}$ is $< m$), the Hessian is singular. Therefore, although this convolution is a linear operation, the Hessian cannot be directly inverted to create a linear deconvolution operator. Instead, an iterative Newton–Raphson approach is implemented as follows.

a) Initialize the model image I^m_0 to zero or to a model that represents a-priori information about the true sky.

b) *Major cycle:* Compute the $\nabla\chi^2$ (residual) image

$$I^{\mathrm{res}} = \left\{[F^\dagger S^\dagger W]\left[V^{\mathrm{corr}} - [SF]I^m_i)\right]\right\}. \tag{11}$$

The *forward* transform $V^m = [SF]I^m_i$ *predicts* visibilities that would be measured for the current sky model, and residuals are computed as $V^{\mathrm{res}} = V^{\mathrm{corr}} - V^m$.The *reverse* transform $I^{\mathrm{res}} = [F^\dagger S^\dagger W]V^{\mathrm{res}}$ computes an image from a set of visibilities. A *preconditioning* scheme decides how best to weight the visibility data (see Section III-B1) before *gridding* them onto a regular grid of spatial frequencies (see Section III-B2) and Fourier inverting to give I^{res}.

c) *Minor cycle:* Compute the update step by applying an operator T to the $\nabla\chi^2$ image. Update the model image.

$$I^m_{i+1} = I^m_i + T(I^{\mathrm{res}}, I^{\mathrm{psf}}) \tag{12}$$

T represents a nonlinear deconvolution of the PSF from I^{res} while filling in unmeasured spatial frequencies (null space of the measurement matrix) for a complete reconstruction of the image. Section III-B3 describes T for several standard deconvolution algorithms.

d) Repeat from b) until convergence is achieved (I^{res} is noise-like) or other termination criteria are satisfied (T can no longer reliably extract any flux from I^{res}).

e) The final I^m is *restored* by first smoothing it to the maximum angular resolution of the instrument to suppress artifacts arising from unconstrained spatial frequencies beyond the measured range and then adding in the final I^{res} to preserve any undeconvolved flux.

1) Preconditioning: The aim of preconditioning is to alter the shape of the PSF according to whatever makes the normal equations easier to solve. This is done by reweighting the data to tune the instrument's sensitivity to a particular type of source and signal-to-noise ratio [8].

The *natural weighting* scheme gives equal weight to all samples and preserves the instrument's peak sensitivity, making it ideal for the detection of low signal-to-noise sources. However, the nonuniform sample density on the uv-grid can give the PSF a wide main lobe and high sidelobes. *Uniform weighting* gives equal weight to each measured spatial frequency irrespective of sample density, and this lowers its peak sensitivity. The resulting PSF has a narrow main lobe and suppressed sidelobes across the entire image and is best suited for sources with high signal-to-noise ratios to minimize sidelobe contamination between sources. *Superuniform weighting* gives a PSF with inner sidelobes suppressed as in uniform weighting but far-out sidelobes closer to that with natural weights. The peak sensitivity is also closer to natural weighting. *UV tapering* suppresses high spatial frequencies and tunes the sensitivity of the instrument to peak for scale sizes larger than the resolution element. *Robust weighting* [10] creates a PSF that smoothly varies between natural and uniform weighting based on the signal-to-noise ratio of the measurements and a tunable parameter that defines a noise threshold.

The final *imaging weights* are given as $W^{\mathrm{im}} = W^{\mathrm{pc}}W$, where W^{pc} are preconditioning weights and W are measurement-noise based weights. The Hessian becomes a convolution operator with the preconditioned PSF in each row ($I^{\mathrm{psf}} = \mathrm{diag}[F^\dagger S^\dagger W^{\mathrm{im}}]$).

2) Gridding: The measured visibilities sample the spatial frequency plane along elliptical tracks and need to be binned onto a regular grid of spatial frequencies so that the fast Fourier transform (FFT) algorithm can be used for Fourier inversion. Gridding interpolation is done as a convolution [8]. Each weighted visibility is first multiplied with a prolate spheroidal function P_s centered at its true location. Then, values at the centers of all grid cells within a certain radius are read off. P_s acts as an antialiasing function. A *grid correction* is then done in the image domain to remove this multiplicative image-domain effect.

Let P_s be a diagonal matrix representing the prolate-spheroidal function. $G^{\mathrm{gc}} = [F(F^\dagger P_s)F^\dagger]$ is the corresponding gridding convolution operator in the spatial frequency domain, equivalent to multiplying the image

domain by $I^{wt}_{gc} = [F^{\dagger}P_s]_{m\times m}$. The normalized *dirty image* and PSF are computed as

$$I^{\{\mathrm{dirty},psf\}}_{m\times 1} = w^{-1}_{\mathrm{sum}}\left[I^{wt}_{gc}\right]^{-1}[F^{\dagger}G^{\mathrm{gc}}S^{\dagger}W^{\mathrm{im}}]V^{\{\mathrm{corr},1\}}_{n\times 1} \tag{13}$$

where division by $w_{\mathrm{sum}} = trace(W^{\mathrm{im}})$ normalizes the peak of the PSF to unity. Equation (13) describes the practical implementation of the *reverse transform* of the *major cycle* and I^{dirty} is the initial I^{res} used to start the iterations.

The model image I^{model} obtained at the end of each *minor cycle* is used in the *forward transform* as

$$V^{m}_{n\times 1} = [SG^{\mathrm{gc}}F]\left[I^{wt}_{gc}\right]^{-1}I^{m}_{m\times 1}. \tag{14}$$

The calculation of these transforms involves traversals of the entire set of visibility data, making it computationally expensive. Deconvolution algorithms usually tailor the frequency of *major* and *minor* cycles to perform tradeoffs among performance, accuracy, and total number of iterations.

3) Deconvolution: For the *minor cycle*, I^{dirty} is assumed to be a perfect convolution of the PSF with the true sky brightness, where $I^{\{\mathrm{dirty,psf}\}}$ are given by (13). The operator T in (12) constructs a model image I^{m} via a deconvolution.

The *CLEAN* algorithm forms the basis for most deconvolution algorithms used in radio interferometry. The peak of the residual image gives the location and strength of a potential point source. The effect of the PSF is removed by subtracting a scaled I^{psf} from I^{res} at the location of each point source and updating I^{m} [(12)]. Many such iterations of finding peaks and subtracting PSFs form the minor cycle.

The following deconvolution algorithms model the sky in a pixel basis and are best suited to isolated point sources whose amplitude is constant across the observing bandwidth. Deconvolution algorithms that produce multiscale and multifrequency source models are described in Section V.

In *Hogbom CLEAN* [11], the minor cycle subtracts a scaled and shifted version of the full PSF to update the residual image for each point source. Only one major cycle is done. It is computationally efficient but susceptible to errors due to inappropriate preconditioning. *Clark CLEAN* [12] does a set of Hogbom minor cycle iterations using a small patch of the PSF. A major cycle is performed when the brightest peak in the residual image is below the first sidelobe level of the brightest source in I^{res}. The residual image is then recomputed as $I^{\mathrm{res}} = [F^{\dagger}](FI^{\mathrm{dirty}} - FI^{m})$ to eliminate aliasing errors. *Cotton–Schwab CLEAN* [13] is similar to the Clark algorithm but computes the residual as $I^{\mathrm{res}} = [F^{\dagger}S^{\dagger}W](V^{\mathrm{corr}} - [SF]I^{m})$. It is time consuming but relatively unaffected by inappropriate preconditioning and gridding errors because it computes χ^2 directly in the measurement domain. It also allows highly accurate prediction of visibilities without pixelation errors. The *Steer–Dewdney–Ito CLEAN* minor cycle finds the locations of sources by setting an amplitude threshold to select pixels. The combined set of pixels is then convolved with the PSF and subtracted out via a Clark major cycle. This algorithm is more suited to deconvolving extended emission. *Maximum Entropy* (MEM) [14] methods and *Non-negative Least Squares* (NNLS) [10], [15] are pixel-based deconvolution algorithms that perform a rigorous constrained optimization in a basis of pixel amplitudes. MEM solves a least squares problem with a penalty function based on image entropy that biases the estimate of the true sky brightness towards a known prior image. NNLS deconvolution solves a least squares problem with linear inequality range constraints for all its parameters.

IV. CALIBRATION AND IMAGING WITH DIRECTION DEPENDENT INSTRUMENTAL EFFECTS

K^{sky}_{ij} in (3) represents the effects of direction-dependent (DD) gains in the measurement from a single interferometric baseline. These DD gains can result from a number of instrumental and atmospheric/ionospheric effects, are potentially different for each baseline, and can be a function of time, frequency, polarization, and direction. In the simplest form of this equation, these dependencies can be ignored, making K^{sky}_{ij} purely multiplicative in the image domain. Imaging can then proceed as described in Section III (correcting only for *direction-independent* terms), with the final image being divided by an estimate of K^{sky} to remove the multiplicative DD effects.

In its general form, (5) in the presence of DD effects for a telescope calibrated for $[K^{\mathrm{vis}}]$ can be written as

$$V^{\mathrm{obs}}_{n\times 1} = [S_{n\times m}]\left[G^{dd}_{m\times m}\right]V^{\mathrm{sky}}_{m\times 1}. \tag{15}$$

Each row of $[G^{dd}_{m\times m}]$ acts as a visibility-plane filter (see [1]) for the measurements from baseline ij and is given by $[G^{dd}_{ij}]_{1\times m} = \mathrm{diag}([FK^{\mathrm{sky}}_{ij}]_{m\times m})$, where K^{sky}_{ij} is assumed to be *known* from a-priori information. Note that K^{sky}_{ij} can also be separated into antenna based terms. We will exploit this property in Section IV-B3 to devise efficient solvers to solve for parametrized forms of $[G^{dd}]$ for *unknown* DD effects.

Equations (5) and (15) suggest the use of FFT-based forward and reverse transforms to account for DD effects using an appropriately constructed G^{dd} operator. Data prediction can incorporate DD effects by using G^{dd} as the

operator for resampling data from a regular grid (FFT of the model image) at points given by the operator S. The reverse transform can correct for DD effects by using the conjugate transpose of G^{dd} along with the standard antialiasing operator G^{gc} for gridding the data (see Section III-B2). For such a transform to efficiently correct for DD effects, the G_{ij}^{dd} filter must satisfy two properties: 1) it should have a finite support size (i.e., corresponding K^{sky} should be band-limited) and 2) it should be a unitary operator (or approximately so). Effects of the W-term and antenna primary beam patterns are two examples of DD effects, whose operators have these desirable properties.

A generalized version of (13) including the DD effects can be written as

$$I^{\{\mathrm{dirty,psf}\}} = \left[I_{dd}^{wt}\right]^{-1}\left[I_{gc}^{wt}\right]^{-1}\left[F^{\dagger}G^{\mathrm{gc}}G^{dd^{\dagger}}S^{\dagger}W^{\mathrm{im}}\right]V^{\{\mathrm{corr},1\}} \quad (16)$$

where

$$I_{dd}^{wt} = \left[F^{\dagger}G^{dd^{\dagger}}W^{\mathrm{im}}G^{dd}F\right] \quad (17)$$

$$I_{\mathrm{gc}}^{wt} = \left[F^{\dagger}P_s\right]. \quad (18)$$

In the absence of DD effects, G^{dd} is an identity matrix, $I_{dd}^{wt} = w_{\mathrm{sum}}[1_{m\times m}]$, and (16) reduces to (13). I_{gc}^{wt} is the same as the *grid correction* mentioned in Section III-B2 to correct for the image plane effects of the antialiasing operator P_s.

Three special cases are discussed in the following sections.

1) When $G^{dd^{\dagger}}G^{dd}$ is an identity matrix, I_{dd}^{wt} is still w_{sum} and (16) can be used to generate $I^{\{\mathrm{dirty}\}}$ free of the relevant DD effects. The effect of the W-term discussed in Section IV-A corresponds to this case.
2) When $G_{ij}^{dd^{\dagger}}G_{ij}^{dd}$ is a time-dependent function, $I_{dd}^{wt} = w_{\mathrm{sum}}[K^{sky^{\dagger}}K^{\mathrm{sky}}]$. DD effects due to time-varying antenna primary beams represent an example of this case, as is discussed in Section IV-B.
3) Mosaic imaging or single pointing imaging with heterogeneous antenna arrays corresponds to the case where $G_{ij}^{dd^{\dagger}}G_{ij}^{dd}$ is not the same for all i and j. This is discussed in Section IV-C.

A. Correction for the W-Term

The W-term is related to the fact that (1) holds for coherence between the E-field measured at two points on a common constant phase front of the incident radiation [16]. This is true only when the array is coplanar, and the source being tracked is at the local zenith [17]. Therefore, in general, the image and visibility planes are not related by a 2-D Fourier transform. The use of the 2-D FFT for imaging wide-fields results in a PSF that is no longer shift-invariant, making standard deconvolution algorithms unsuitable. However, if a Fresnel diffraction kernel is used as a propagator [18] to compute the E-field measured at one of the antennas of each baseline, the 2-D Fourier relation can be preserved. This propagator is equal to the Fourier transform of the W-term in (2) $\left(e^{\iota w\sqrt{1-l^2-m^2}}\right)$. Two algorithms commonly used to correct for the effects of the W-term are described below.

1) Faceting Algorithms: The effect of the W-term is small close to the phase tracking center. This property is exploited by algorithms that divide the field of view into a number of facets. Images are made by either projecting the facet images onto the local tangent plane (image-plane faceting [19]) and using the appropriate PSF for the deconvolution of individual facet images, or by projecting the (u, v) for each facet onto a single tangent plane in the gridding step required for an FFT-based reverse transform [20]. This latter method produces a single flat image and has several run-time and imaging performance benefits [21].

2) W-Projection Algorithm: In (15), the operator $[G^{dd}]$ can be used to account for the W-term by choosing $K_{ij}^{\mathrm{Sky}} = e^{w_{ij}(\sqrt{1-l^2-m^2}-1)}$. This W-term operator G^{dd} is strictly unitary (by construction) and has a finite support (due to the antialiasing operator G^{gc}). It will therefore correct for the W-term during image deconvolution [21], [22]. Conservatively speaking, the *W-projection* algorithm is about an order of magnitude faster than faceting, and for the same amount of computing time can deliver higher dynamic range images [21].

B. Correction for Primary Beam

With the increased instantaneous sensitivities of next-generation telescopes and long integrations required for high dynamic range imaging, antennas can be considered neither identical nor stable as a function of time. Therefore, next-generation imaging algorithms need to include corrections for the effects of time-varying antenna primary beams [23], [24]. Algorithms to correct for these effects can be broadly classified into two categories, namely, corrections in the image plane versus corrections in the Fourier plane.

1) Image Plane Correction: When K_{ij}^{sky} is different for each baseline, one approach for correcting DD effects is the direct evaluation of the integral in (3) for the forward and reverse transforms during iterative image deconvolution [25]. The resulting run-time load for realistic data sizes can, however, be prohibitive. To reduce the compute load somewhat, an FFT-based reverse transform (Section III-B2) is used, but this requires making assumptions about the variability of either the sky emission or the antenna power pattern.

2) Fourier Plane Correction—The A-Projection Algorithm: The visibility-plane filter describing the effects of the

antenna primary beams is the autocorrelation of the antenna aperture illumination function. For a finite-sized antenna, this clearly has a finite support in the Fourier domain. However, the resulting *effective* operator $(FG^{dd}/\sqrt{I_{dd}^{wt}})$ is only approximately unitary [24]. The *A-projection* algorithm uses accurate forward and approximate reverse transforms based on the primary beam operator to correct for time-variable primary beam effects (see [24] for details and an example of its application to full-beam imaging with the VLA). Apart from the initial setup time required to compute the antenna aperture function, the run-time performance of this algorithm, when imaging the entire field of view up to the first side lobe of the antenna power pattern, is equivalent to that of standard image deconvolution algorithms using a gridding convolution function with a support size ~30% larger in linear extent.

3) Pointing Self-Calibration: Antenna pointing errors make K_{ij}^{sky} (and the resulting G_{ij}^{dd}) different for each baseline. When K^{sky} represents effects of antenna primary beams, K_{ij}^{sky} can be decomposed into two antenna-based terms as $J_i^{\mathrm{sky}} \otimes J_j^{\mathrm{sky}\dagger}$, each parametrized for pointing errors, which can be recovered by solving the resulting parametrized measurement equation. However, iterative solvers using K_{ij}^{sky} to represent pointing errors necessarily require evaluation of the integral in (3) in each iteration and have proved to be impractically slow.

An alternate approach is to solve for antenna pointing errors in the visibility domain, where it is efficient to compute G_{ij}^{dd} parametrized by pointing errors. Given a model for the sky, the *Pointing SelfCal* algorithm [26] iteratively solves for these pointing errors. This algorithm can be efficiently implemented using the forward and reverse transforms described in Section IV-B2. The effects of pointing errors can also be corrected along with other direction-dependent effects as part of an iterative image deconvolution.

C. Mosaicing

Mosaicing observations consist of a number of independent pointings covering a large field of view with an adequate sampling. Instruments with focal plane arrays can be considered to observe a number of mosaic pointings in parallel, while traditional instruments observe only one pointing at a time. Mosaicing observations can be treated in a natural way using the formalism of (16)–(18). Every pointing of the mosaic corresponds to a separate G^{dd} and I_{dd}^{wt}. The difference may be as little as the pointing direction (i.e., a translation of I_{dd}^{wt} and phase gradient for G^{dd}), although more substantial changes are possible (e.g., for inhomogeneous arrays). Also, in the presence of noise, (16) does not adequately constrain the dirty image in those parts of the sky where the weight I_{dd}^{wt} is low. The solution is a generalization of (16) where the product $[I_{dd}^{wt}]_k I^{\mathrm{dirty}}$ for every pointing k is combined to form a linear system of equations. This is known as linear mosaicing.

$$I^{\{\mathrm{dirty},\ \mathrm{psf}\}} = \left[I_{dd}^{wt}\right]^{-1} \sum_k \left[I_{gc}^{wt}\right]^{-1} \left[F^\dagger G^{\mathrm{gc}} G_k^{dd\dagger} S^\dagger W_k^{\mathrm{im}}\right] V_k^{\{\mathrm{corr},1\}} \quad (19)$$

where the weight is given by a similar generalization of (17)

$$I_{dd}^{wt} = \sum_k \left[F^\dagger G_k^{dd\dagger} W_k^{\mathrm{im}} G_k^{dd} F\right]. \quad (20)$$

Strictly speaking, the PSF calculated as a response to a point source located at the center of the mosaic (or any other location; but the same for all pointings) is valid only for one particular location. For any other direction in the field of view, the contributions of individual pointings are different, causing a different response. Therefore, the deconvolution performed in the minor cycle is always an approximate operation, and a number of major cycles is usually required. However, this fact allows one to optimize the PSF calculation by taking into account only one pointing that contributes the most to (19) (e.g., the closest pointing to the center of the mosaic). Another way is to use a representative pointing and apply a phase shift to the convolution operator G to center the primary beam (i.e., to remove the offset of this pointing with respect to the mosaic center).

This approach to mosaicing is a form of a joint deconvolution because the data from all pointings are combined before the deconvolution takes place. It was shown to be superior to independent deconvolution where the final image is computed as a weighted sum of deconvolved subimages corresponding to individual pointings of the mosaic [27].

V. IMAGING ALGORITHMS WITH ADVANCED IMAGE PARAMETERIZATIONS

So far, the discussions in this paper have focused on the calibration and imaging of visibilities from one polarization pair, the use of a pixel basis to parameterize the sky brightness distribution, and the assumption that source structure is constant across the entire bandwidth of data being imaged. In this section, we relax these assumptions and describe how standard methods can be augmented to handle the added complexity of the increased dimensionality of the parameter space.

A. Multiscale CLEAN Deconvolution

Images of astrophysical objects tend to show complex structure at different spatial scales. The use of a pixel-basis for deconvolution is ideal for fields of isolated point-like sources that are smaller than the instrument's angular

resolution but tends to break extended emission into a collection of compact sources, which is often inaccurate. A better choice is to parameterize the image in a scale-sensitive basis that spans the full range of scale sizes measured by the instrument. This provides a strong constraint on the reconstruction of visibilities in the null space of the measurement matrix. Also, since spatial correlation length fundamentally separates signal from noise, scale-sensitive deconvolution algorithms generally give more noise-like residuals [28].

The minor cycle of the *Multiscale CLEAN* algorithm [29] parameterizes the image into a collection of inverted tapered paraboloids (h_k, $k = 1 : n_{\text{scales}}$) whose widths are chosen from a predefined list. PSFs and *dirty* images corresponding to each spatial scale are calculated by smoothing $I^{\{\text{dirty,psf}\}}$ from (13) by each h_k. Each iteration i of the minor cycle follows a matched-filtering technique where the location, amplitude, and scale of each new component is chosen from $\max\{I^{\text{res}} \star h_k\}$ ($\star$ denotes convolution) and the update step accounts for the nonorthogonality of the different basis functions h_k. *MS-CLEAN* works very well for complicated spatial structure, but its performance is limited by working with a discrete set of scale sizes and the fact that if an inappropriate component is chosen, it takes the addition of many more components to correct it. Typically, $n_{\text{scales}} \approx 8$ for a source with complex spatial structure. *Multiresolution CLEAN* [30] performs a series of Hogbom minor cycles at different angular resolutions beginning at the lowest resolution to collect all extended emission and progressing to higher resolutions. PSFs and residual images at different resolutions are made by varying the image pixel sizes during gridding. Its limitations are similar to *MS-CLEAN*, in that there is no way to undo a component selection in case a better option becomes available later in the iterations, and is less robust since it searches for components one scale size at a time. The *ASP CLEAN* [28] algorithm parameterizes the sky brightness distribution into a collection of Gaussians and does a formal constrained optimization on their parameters. In the major cycle, visibilities are predicted analytically with high accuracy. In the minor cycle, the location of a flux component is chosen from the peak residual, and the parameters of the largest Gaussian that fits the image at that location are found. The minimization proceeds over subspaces consisting of sets of localized Gaussians whose parameters are varied together. This prevents errors due to inappropriate fits from propagating very far into the iterations. The computing costs and run-times of each minor cycle iteration of *MS-CLEAN* and *ASP-CLEAN* are a few times worse than *Hogbom-CLEAN*. However, they parameterize the sky brightness more physically, and convergence is achieved in far fewer iterations.

B. Multifrequency Synthesis Imaging

The uv-coverage of a synthesis array can be greatly improved by using the fact that visibilities measured at different receiver frequencies correspond to different spatial frequencies. *Multifrequency synthesis* (MFS) is the process of combining data from multiple spectral channels onto the same spatial-frequency grid during imaging to take advantage of the increased uv-coverage and imaging sensitivity. As long as the sky brightness is the same across the total measured bandwidth, standard imaging and deconvolution algorithms can be used along with MFS. If the sky brightness varies across the observing bandwidth, the narrow-band (or monochromaticity) requirement of aperture synthesis breaks down and the Fourier relation in the Van Cittert Zernike theorem does not hold. The following algorithms fold a frequency dependence of the image sky model into the measurement equation to handle this problem in the *minor cycle*.

MF-CLEAN [31] is a matched filtering technique based on *spectral PSFs* that describe the instrument's responses to point sources with spectra given by Taylor series functions [see (21) and (22)]. Source spectra ($I(\nu)$) are modeled as a power law, and a first-order Taylor expansion of $I(\nu)$ is combined with the regular imaging equation to describe the dirty image as a sum of convolutions given by $I^{\text{dirty}} = \sum_t I_t^{\text{dirty}} = \sum_t C_t I_t^m \star I_t^{\text{psf}}$, where I_t^{psf} are the *spectral PSFs* for $t = \{0, 1\}$. The deconvolution minor cycle simultaneously solves for C_0 and C_1 for each component added to the model image. This algorithm uses a pixel basis and is most suited for point sources with pure power-law spectra with a weak frequency dependence. *MS-MF-CLEAN* [32] is a multiscale multifrequency deconvolution algorithm that extends *MF-CLEAN* to work with the instrument's response to a polynomial spectrum (nth-order Taylor series) at multiple spatial scales. This algorithm is suited for extended emission and features with nonlinear spectra described by a power law of varying index across the observing band.

Some direction-dependent effects in K_{ij}^{sky} (e.g., effect of the primary beam) are also frequency dependent. Therefore, the *spectral PSFs* and *dirty images* used in the minor cycle can be computed as another generalization of (16).

$$I_t^{\{\text{dirty,psf}\}} = \left[I_{dd}^{wt}\right]^{-1}\left[I_{gc}^{wt}\right]^{-1}\sum_{\nu}\left[F^{\dagger}G^{\text{gc}}G_{\nu}^{dd^{\dagger}}S^{\dagger}W_{\nu,t}^{\text{im}}\right]V_{\nu}^{\{\text{corr},1\}} \tag{21}$$

where

$$W_{\nu,t}^{\text{im}} = W^{\text{im}}((\nu - \nu_0)/\nu_0)^t. \tag{22}$$

The *weight* image describes the noise variation across the image due to imaging weights and frequency-dependent K_{ij}^{sky} and is given by

$$I_{dd}^{wt} = \sum_{\nu}\left[F^{\dagger}G_{\nu}^{dd^{\dagger}}W_{\nu,t=0}^{\text{im}}G_{\nu}^{dd}F\right]. \tag{23}$$

C. Full Polarization Calibration and Imaging

The preceding sections have dealt with the calibration and imaging of only one correlation pair *pp* from a single feed. This section deals with the full-polarization calibration of a pair of potentially imperfect orthogonal feeds and the imaging of all four Stokes parameters.

1) Full-Stokes Calibration: Each baseline measures the product of $K_{ij}^{\text{vis}} = J_i^{\text{vis}} \otimes J_j^{\text{vis}*}$ with the true coherence vector seen by that baseline. Equation (5) becomes

$$\vec{V}_{4n\times 1}^{\text{obs}} = \left[K_{4n\times 4n}^{\text{vis}}\right]\vec{V}_{4n\times 1}^{\text{model}} \tag{24}$$

and the elements of K_{ij}^{vis} are computed as described in Section III-A. For a source with known polarization characteristics, the true coherence vector is known (constant $\times$ [1,0,0,1] for circular feeds and an unpolarized source), and one can form a system of linear equations with the elements of K_{ij}^{vis} as unknowns. For a single baseline, there are up to ten degrees of freedom and four equations [33]. However, with an a-priori source model, measurements from all baselines provide enough constraints to uniquely factor the baseline-based K_{ij}^{vis} matrices into antenna-based Jones matrices ($4 \times n_{\text{ant}}(n_{\text{ant}} - 1)/2$ equations and $4 \times n_{\text{ant}}$ unknowns). In its most general form, the elements of J_i^{vis} can be computed by minimizing $\chi^2 = \sum_{ij} |\vec{V}_{ij}^{\text{obs}} - [J_i^{\text{vis}} \otimes J_j^{vis*}]\vec{V}_{ij}^{m}|^2$ with respect to the antenna-based J_i^{vis}.

In existing software packages, polarization calibration is usually done in stages. First, only the diagonal elements of the Jones matrices are solved for, assuming zero leakage between the orthogonal feeds. Solutions are then applied and only off-diagonal terms are solved for. Another method of solving for antenna-based gains and leakages from only parallel-hand correlations pp, qq is described in [34]. The effects of depolarization cannot be factored into Jones matrices, and a baseline-based calibration is sometimes carried out by artificially imposing constraints between the elements of K_{ij}^{vis}.

2) Full-Stokes Imaging: The Stokes vector for polarized sky brightness $\vec{I}^{\text{stokes}} = \{I, Q, U, V\}$ is related to the vector of images corresponding to the correlations $\{pp, pq, qp, qq\}$ as

$$\vec{I}_{4m\times 1}^{\text{sky}} = [\mathcal{S}_p]_{4m\times 4m}\,\vec{I}_{4m\times 1}^{\text{stokes}} \tag{25}$$

where $\mathcal{S}_p$ holds a 4×4 linear operator per image pixel [7]. A full-Stokes deconvolution differs from standard methods in the computation of *dirty* images and the *minor cycle*. The Stokes vector of dirty images $\vec{I}^{\text{dirty,Stokes}}$ is computed by applying (25) to the set of dirty images in the correlation basis $\vec{I}^{\{\text{dirty,corr}\}}$ given by (13) or (16). The different Stokes parameters are considered to be linearly independent and deconvolution minor cycles are performed separately on each Stokes image. For compact sources, position constraints are sometimes applied across Stokes parameters based on the locations of peak residuals of the Stokes I image. [35] describes an algorithm that applies the constraint of $I^2 \geq Q^2 + U^2 + V^2$ during deconvolution.

VI. CONCLUSION

We have presented a complete mathematical framework for describing many of the major calibration and imaging algorithms used in radio interferometry. This framework can be used for three purposes: a) elucidating the fundamental assumptions and details of algorithms, b) isolating the mathematical structure so that standard libraries can be used, and c) allowing both generalization and specialization to generate new algorithms.

The computing and software issues connected with the use of this framework are substantial, especially given the large data volumes and processing loads being contemplated for new radio telescopes. We will discuss these issues further in a subsequent paper. We note that this framework can also be used to address other algorithms not discussed here. These include the *peeling* technique for direction-dependent calibration, the problem of *ionospheric calibration* as a direction-dependent effect, and the excision of radio-frequency interference from measured visibility data. ■

Acknowledgment

The authors would like to thank K. Golap and R. Nityananda for helpful discussions. They would also like to thank R. Thompson and the referees for their careful reading of this paper, and helpful comments and suggestions.

The authors also thank the many people who have contributed to two principal software packages in which their work resides: the widely available CASA project (previously AIPS++) from NRAO and the software being developed for ASKAP.

REFERENCES

[1] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy*. New York: Wiley, 1986.

[2] R. A. Perley, P. J. Napier, and B. J. Butler, "The expanded very large array: Goals, progress, and plans," in *Proc. SPIE Conf. Ser.*, vol. 5489, J. M. Oschmann, Jr., Ed., Oct. 2004, vol. 5489, pp. 784–795.

[3] H. Röttgering, "LOFAR, a new low frequency radio telescope," *New Astron. Rev.*, vol. 47, pp. 405–409, Sep. 2003.

[4] R. T. Schilizzi, "The square kilometre array *EAS Pub. Series*, vol. 15, pp. 445–463, 2005.

[5] S. Johnston, M. Bailes, N. Bartel, C. Baugh, M. Bietenholz, C. Blake, R. Braun, J. Brown, S. Chatterjee, J. Darling, A. Deller, R. Dodson, P. G. Edwards, R. Ekers, S. Ellingsen, I. Feain, B. M. Gaensler, M. Haverkorn, G. Hobbs, A. Hopkins, C. Jackson, C. James, G. Joncas, V. Kaspi, V. Kilborn, B. Koribalski, R. Kothes, T. L. Landecker, E. Lenc, J. Lovell, J.-P. Macquart, R. Manchester, D. Matthews, N. M. McClure-Griffiths, R. Norris, U.-L. Pen, C. Phillips, C. Power, R. Protheroe, E. Sadler, B. Schmidt, I. Stairs, L. Staveley-Smith, J. Stil, R. Taylor, S. Tingay, A. Tzioumis, M. Walker, J. Wall, and M. Wolleben, "Science with the Australian square kilometre array

pathfinder," *Pub. Astron. Soc. Australia*, vol. 24, pp. 174–188, Dec. 2007.

[6] J. Jonas, "MeerKAT as an SKA pathfinder," Rhodes Univ., Tech. Rep. [Online]. Available: http://www.ska.ac.za/meerkat/index.shtml

[7] J. P. Hamaker, J. D. Bregman, and R. J. Sault, "Understanding radio polarimetry: I. Mathematical foundations," *Astron. Astrophys. Suppl. Ser.*, vol. 117, pp. 137–147, May 1996.

[8] G. B. Taylor, C. L. Carilli, and R. A. Perley, Eds., Astron. Soc. Pacific Conf. Ser. 180: Synth. Imag. Radio Astron. II., 1999.

[9] T. J. Cornwell and P. N. Wilkinson, "A new method for making maps with unstable radio interferometers," *Mon. Not. Roy. Astron. Soc.*, vol. 196, pp. 1067–1086, Sep. 1981.

[10] D. S. Briggs, "High fidelity deconvolution of moderately resolved sources," Ph.D. dissertation, New Mexico Inst. Mining Technol., Socorro, NM, Mar. 1995.

[11] J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl. Ser.*, vol. 15, pp. 417–426, 1974.

[12] B. G. Clark, "An efficient implementation of the algorithm 'clean'," *Astron. Astrophys.*, vol. 89, p. 377, Sep. 1980.

[13] F. R. Schwab, "Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry," *Astron. J.*, vol. 89, pp. 1076–1081, Jul. 1984.

[14] T. J. Cornwell and K. J. Evans, "A simple maximum entropy deconvolution algorithm," *Astron. Astrophys.*, vol. 143, pp. 77–83, 1985.

[15] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, ser. Prentice-all Automatic Computation. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[16] M. Born and E. Wolf, *Principles of Optics*. Oxford, U.K.: Pergamon, 1959.

[17] T. J. Cornwell and R. A. Perley, "Synthesis imaging in radio astronomy ii," in *Astron. Soc. Pac. Conf. Ser. 180: Synth. Imag. Radio Astron. II*, G. B. Taylor, C. L. Carilli, and R. A. Perley, Eds., 1999.

[18] J. W. Goodman, *Introduction to Fourier Optics*. New York: McGraw-Hill, 2002.

[19] T. J. Cornwell and R. A. Perley, "Radio-interferometric imaging of very large fields—The problem of non-coplanar arrays," *Astron. Astrophys.*, vol. 261, pp. 353–364, Jul. 1992.

[20] R. J. Sault, D. J. Bock, and A. Duncan, "Polarimetric imaging of large fields in radio astronomy," *Astron. Astrophys.*, vol. 139, p. 387, 1999.

[21] T. J. Cornwell, K. Golap, and S. Bhatnagar, "The non-coplanar baselines effect in radio interferometry: The w-projection algorithm," *IEEE J. Sel. Topics Signal Process.*, vol. 2, pp. 647–657, Oct. 2008.

[22] T. J. Cornwell, K. Golap, and S. Bhatnagar, "W-projection: A new algorithm for non-coplanar baselines," EVLA Memo 67, Tech. Rep., Dec. 2003.

[23] S. Bhatnagar, T. J. Cornwell, and K. Golap, "Corrections of errors due to antenna power patterns during imaging," EVLA Memo 100, Tech. Rep., Aug. 2006.

[24] S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson, "Correcting direction-dependent gains in the deconvolution of radio interferometric images," *Astron. Astrophys.*, vol. 487, pp. 419–429, Aug. 2008.

[25] J. M. Uson and W. D. Cotton, "Beam squint and Stokes V with off-axis feeds," *Astron. Astrophys.*, vol. 486, pp. 647–654, Aug. 2008.

[26] S. Bhatnagar, T. J. Cornwell, and K. Golap, "Solving for the antenna based pointing errors," EVLA Memo 84, Tech. Rep., Aug. 2004.

[27] T. J. Cornwell, "Radio-interferometric imaging of very large objects," *Astron. Astrophys.*, vol. 202, pp. 316–321, Aug. 1988.

[28] S. Bhatnagar and T. J. Cornwell, "Scale sensitive deconvolution of interferometric images: I. Adaptive Scale Pixel (ASP) decomposition," *Astron. Astrophys.*, vol. 426, pp. 747–754, Nov. 2004.

[29] T. J. Cornwell, "Multi-scale CLEAN deconvolution of radio synthesis images," *IEEE J. Sel. Topics Sig. Process.*, vol. 2, pp. 793–801, Oct. 2008.

[30] B. P. Wakker and U. J. Schwarz, "The multi-resolution CLEAN and its application to the short-spacing problem in interferometry," *Astron. Astrophys.*, vol. 200, pp. 312–322, Jul. 1988.

[31] R. J. Sault and M. H. Wieringa, "Multi-frequency synthesis techniques in radio interferometric imaging," *Astron. Astrophys. Suppl. Ser.*, vol. 108, pp. 585–594, Dec. 1994.

[32] U. Rau and T. J. Cornwell, "Multi-scale multi-frequency synthesis imaging in radio interferometry," unpublished.

[33] R. J. Sault, J. P. Hamaker, and J. D. Bregman, "Understanding radio polarimetry: II. Instrumental calibration of an interferometer array," *Astron. Astrophys. Suppl. Ser.*, vol. 117, pp. 149–159, May 1996.

[34] S. Bhatnagar and R. Nityananda, "Solving for closure errors due to polarization leakage in radio interferometry of unpolarized sources," *Astron. Astrophys.*, vol. 375, pp. 344–350, Aug. 2001.

[35] M. A. Holdaway and J. F. C. Wardle, "Maximum entropy imaging of polarization in very long baseline interferometry," in *Proc. SPIE Conf. Ser.*, vol. 1351, A. F. Gmitro, P. S. Idell, and I. J. Lahaie, Eds., Nov. 1990, vol. 1351, pp. 714–724.

[36] T. Cornwell and M. Wieringa, "The design and implementation of synthesis calibration and imaging in AIPS++," in *Astron. Soc. Pacific Conf. Ser. 125: Astron. Data Anal. Software Syst. VI*, 1997, pp. 10–17.