Cleaning radio interferometric images using a spherical wavelet decomposition

Chris J. Skipper, Anna. M. M. Scaife

Jodrell Bank Centre for Astrophysics, Alan Turin Building, The University of Manchester, Manchester, M13 9PL, UK

Jason D. McEwen

Mullard Space Science Laboratory (MSSL), University College London, Surrey, RH5 6NT

Abstract

The deconvolution, or cleaning, of radio interferometric images often involves computing model visibilities from a list of clean components, in order that the contribution from the model can be subtracted from the observed visibilities. This step is normally performed using a forward fast Fourier transform (FFT), followed by a 'degridding' step that interpolates over the uv plane to construct the model visibilities. An alternative approach is to calculate the model visibilities directly by summing over all the members of the clean component list, which is a more accurate method that can also be much slower. However, if the clean components are used to construct a model image on the surface of the celestial sphere then the model visibilities can be generated directly from the wavelet coefficients, and the sparsity of the model means that most of these coefficients are zero, and can be ignored. We have constructed a prototype imager that uses a spherical-wavelet representation of the model image to generate model visibilities during each major cycle, and find empirically that the execution time scales with the wavelet resolution level, J, as $O(1.07^J)$, and with the number of distinct clean components, N_C, as $O(N_C)$. The prototype organises the wavelet coefficients into a tree structure, and does not store or process the zero wavelet coefficients.

Keywords: techniques: interferometric

1. Introduction

The advent of the radio interferometer in astronomy opened a new high-resolution window on the universe, and the scale and ambition of astronomical interferometers continues to grow exponentially. The next generation of interferometers, such as the forthcoming Square Kilometre Array (SKA; Schilizzi et al. 2008; Dewdney et al. 2009), which is due to begin construction soon, and its precursors, MeerKAT (Jonas, 2009; Booth and Jonas, 2012) and ASKAP (Johnston et al., 2008), which are currently being brought online, will require new innovations, algorithms and hardware capabilities in order to meet the huge demand for compute capability that they will present.

A radio interferometer measures the interference pattern produced by two or more apertures, and therefore samples the sky in the Fourier domain. To recover images from these visibilities is an entirely computational problem, and one which scales (depending upon which step is being performed) in proportion to the number of baselines between pairs of antennas, the size of the image required, or the length of the observation. The greatest computational expense when constructing images from radio data is arguably the cleaning, or deconvolution, stage. Since, in interferometry, coverage of the Fourier plane is always incomplete it follows that the synthesised beam of an interferometer is imperfect compared to that of a filled-aperture detector. Correcting for these imperfections is a slow process.

To 'clean' radio images we would normally invoke an iterative process of subtracting instances of the synthesised beam from our uncleaned, or dirty, image, and instead build up a clean component list or model image (Högbom, 1974). The flowchart in Fig. 1 shows the construction of a component list from iterations of Högbom clean. Following this process of iterations we would then recalculate our visibilities based upon the model image, subtract these model visibilities from the observed visibilities, and generate a new dirty image which will be cleaned with a new set of deconvolution iterations.

Model visibilities can be calculated directly from the component list by summing the contribution from each component separately, or more commonly from the model image by using a forward fast Fourier transform (FFT), followed by an interpolation process known as degridding. The former scales in performance with the number of visibilities and the number of clean components, while the latter scales with the size of the image in the FFT stage, and the number of visibilities in the degridding stage. McEwen and Scaife (2008) developed a fast method to simulate visibilities which uses a wavelet representation over the surface of the celestial sphere. Their technique takes advantage of the sparse representation of the sky in the wavelet basis, and recognises that many of the coefficients are zero or nearly zero. Their algorithm uses hard thresholding to determine which coefficients to retain and which to reject, set by defining a fixed proportion of coefficients to retain, and alternatively, by using an annealing strategy. The authors report
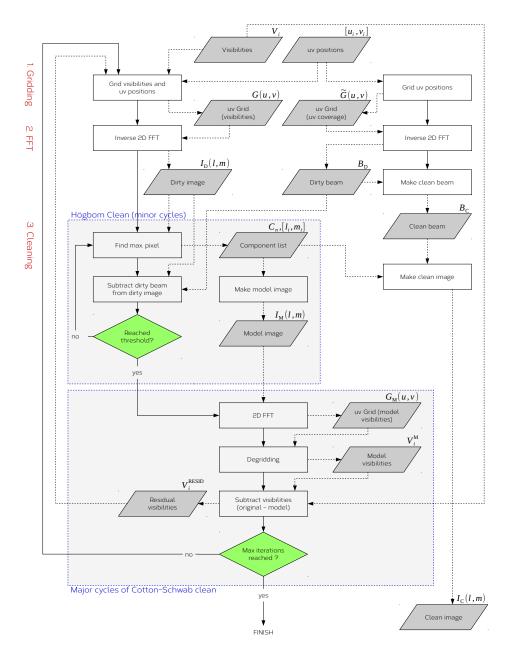
Figure 1: A typically radio-imaging pipeline, demonstrating the gridding, FFT and cleaning (deconvolution) stages. The minor cycle stage performs a Högbom clean, in which the dirty image is cleaned down to some threshold and a model image is built up from the identified components. In the major-cycle stage, a two-dimensional FFT is applied to the model image to transform to the uv domain, and an interpolation algorithm calculates each model visibility. The model image and clean image can be built up gradually with each iteration of Högbom clean.

that visibilities could be computed accurately when fewer than one per cent of the coefficients were retained.

When generating model visibilities from the model image using FFT and degridding steps we are using a tangent-plane approximation, which is only valid for small fields of view (FOV). As the FOV becomes larger, and the area of the sky within the image region encompasses more of the curvature of the celestial sphere, this approximation becomes less reliable. In contrast, a spherical representation of the model image, using wavelets or spherical harmonics, can be constructed over the whole of the sphere, and the visibilities can be generated with no reliance on a tangent-plane approximation, making a spherical representation ideal for reconstructing model visibilities for use in wide-field imaging deconvolution. The technique may therefore be of particular use to instruments designed for deep surveying of the sky, to which large FOVs are an essential requirement. McEwen and Scaife (2008) demonstrated that using a fast wavelet method is much more efficient at generating visibilities than using a spherical harmonic representation, so here we confine our discussion purely to the spherical wavelet method.

In this paper we have implemented the theory described by McEwen and Scaife (2008), and constructed a full imaging propotype which uses a spherical-wavelet representation to compute visibilities during the major cycles of the cleaning stage. We describe in detail the way in which cleaning is normally performed in radio astronomy, and how spherical wavelets fit in to this process.

2. Radio imaging in astronomy

2.1. Overview

High-resolution radio imaging is generally obtained through the use of radio interferometers, which are arrays of radio antennas arranged in some configuration such that each pair discretely samples a unique Fourier component of the radio sky. Together, the correlated amplitude from each pair of antennae (the 'visibilities', V_i), and the u and v components of their baseline vector, form the two-dimensional Fourier transform of the sky image, $I(l, m)$, where l and m are directional cosines on the sky. The full baseline vector ($\boldsymbol{B}_i = [u_i, v_i, w_i]$, where $u_i, v_i, w_i \in \mathbb{R}$), includes the w term, and is orientated such that the w-axis points towards the phase position on the sky, and the u and v axes form the uv plane. The visibilities are related to the sky image by

$$V_j = \int \int \frac{I(l, m)}{\sqrt{1 - l^2 - m^2}} \exp\Big[i2\pi \Big(u_j l + v_j m + w_j \big(\sqrt{1 - l^2 - m^2} - 1 \big) \Big) \Big] \mathrm{d}l \, \mathrm{d}m, \tag{1}$$

and for small fields of view, where $1 - l^2 - m^2 \approx 1$, this equation reduces to

$$V_j = \int \int I(l, m) \exp\Big[i2\pi \Big(u_j l + v_j m \Big) \Big] \mathrm{d}l \, \mathrm{d}m, \tag{2}$$

i.e. a two-dimension Fourier transform of the sky image $I(l, m)$, while for larger fields of view the components of the baseline
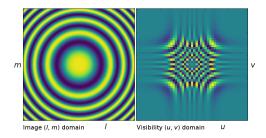

Figure 2: Example w-kernel for degridding, $K_{u,v}^j$, shown in the image domain (left panel) and uv domain (right panel).

vectors perpendicular to the uv plane (the w terms) require (see Cornwell et al. 2012; Pratley et al. 2018) that visibilities recovered through a Fourier transform are convolved with a 'w-kernel', $K_{u,v}^j$, given by (and shown graphically in Fig. 2)

$$K_{u,v}^j = \mathcal{F}\Big\{ \frac{1}{\sqrt{1 - l^2 - m^2}} \exp\Big[i2\pi w_j \big(\sqrt{1 - l^2 - m^2} - 1 \big) \Big] \Big\}. \tag{3}$$

As the visibility equation (Eqn. 1) cannot be used to analytically determine the sky image, $I(l, m)$, from a set of known visibilities it follows that fast Fourier transforms are almost always employed in the recovery of $I(l, m)$. The sky image is a convolution of the components present within the field of view and the synthesised beam pattern. In practise, the uv plane is only partially sampled, and the synthesised beam pattern (the 'dirty' beam, B_D) is imperfect. The process of recovering the clean sky image $I_C(l, m)$ traditionally follows a number of well-established steps:

1. Gridding: each visibility V_i is placed on a discrete grid, $G(u, v)$, at its associated uv position $[u_i, v_i]$. The visibilities are usually convolved with a 2D weighting function, such as a Gaussian or prolate spheroidal function, and the w-kernel function (if required). To generate the dirty beam, the visibilities are gridded again at the same positions, but this time their values are all set to unity; the resulting grid, $\hat{G}(u, v)$, is the uv coverage.

2. FFT: the 'dirty' image of the sky, $I_D(l, m)$, is recovered using a two-dimensional inverse FFT of the gridded visibilities, i.e. $\mathcal{F}^{-1}\{G(u, v)\}$, and the dirty beam, B_D, is similarly recovered using $\mathcal{F}^{-1}\{\hat{G}(u, v)\}$.

3. Cleaning/deconvolution: A cleaning algorithm is applied, that removes the convolution of the dirty beam by building a list of components directly from the dirty image. A clean beam, B_C, is constructed by fitting the dirty beam with a two-dimensional Gaussian function, rotated by some angle. The clean image, $I_C(l, m)$, is then recovered by the convolution of the component list with the clean beam.

This process is represented as a flow chart in Fig. 1.

2.2. The cleaning algorithm

The basis of deconvolving the dirty beam from the sky image is an iterative algorithm called the CLEAN algorithm (or simply Högbom clean; Högbom, 1974), and involves repeatedly identifying the pixel with the maximum value, C_i, at position $[l_i, m_i]$

3

(where $l_i, m_i \in \mathbb{Z}$) in the dirty image, and subtracting the dirty beam from that position, such that, for iteration n,

$$I_{\mathrm{D}}^{(n+1)}(l,m) = I_{\mathrm{D}}^{(n)}(l,m) - gC_n[B_{\mathrm{D}} \circledast \delta(l_n, m_n)], \qquad (4)$$

where g is the gain parameter, typically fixed to 0.1 or 0.2, and δ is the Dirac delta function; with each iteration the algorithm builds up a list of clean components and positions, C_n and $[l_n, m_n]$, a model image

$$I_{\mathrm{M}}(l,m) = \sum_i gC_i \, \delta(l_i, m_i), \qquad (5)$$

and a clean image

$$I_{\mathrm{C}}(l,m) = \sum_i gC_i[B_{\mathrm{C}} \circledast \delta(l_i, m_i)]. \qquad (6)$$

More advanced cleaning algorithms are built upon the Högbom clean, such as Clark clean (Clark, 1980), and Cotton-Schwab clean (Schwab, 1984). The latter introduces minor and major cycles, with the minor cycles provided by the iterations of the Högbom clean, and the major cycles, which are performed every time the minor-cycle stage reaches some cut-off threshold, described as follows:

1. FFT and 'degridding': A forward FFT is performed on the model image, $I_{\mathrm{M}}(l,m)$, to transform into the uv domain. An interpolation algorithm then calculates each 'model' visibility, V_i^{M}, at its known position $[u_i, v_i]$, in a process called degridding. The more accurate alternative is to calculate each visibility directly from Eqn. 1 (using the model image, $I_{\mathrm{M}}(l,m)$), but such an approach is only computationally efficient if the number of clean components is low.

2. Subtract visibilities: the model visibilities are subtracted from the observed visibilities to give the residual visibilities, such that $V_i^{\mathrm{RESID}} = V_i - V_i^{\mathrm{M}}$.

3. Gridding and FFT: the residual visibilities, V_i^{RESID}, are placed on the uv grid, and an inverse FFT is performed to generate a new dirty image.

3. Deconvolution using spherical wavelets

3.1. Background on wavelets

A wavelet series, such as the Haar wavelet (Haar, 1910), is a representation of a function, $F(t)$, in terms of an alternative basis or dictionary (hereafter we use the term basis for brevity). The basis is typically described in terms of a scaling function, ϕ, and a series of wavelet functions, ψ_j, at scales j. These functions, together with their coefficients, can be used to reconstruct the original function $F(t)$. In the explanation that follows, we restrict our discussion to the discrete wavelet transform, in which our original function, $F(t)$ is sampled discretely at regular time intervals, which we denote F_k.

The coefficients of the scaling and wavelet functions (denoted $\lambda_{j,k}$ and $\gamma_{j,k}$ respectively, for resolution j and index k) are calculated at multiple resolutions, such that at resolution j (where $j \geq 1$) there are 2^{j-1} scaling coefficients and 2^{j-1}

wavelet coefficients. At the highest resolution, $j = J$, the scaling coefficients of the Haar wavelet store the values of the discretely sampled function, i.e. $\lambda_{J,k} = F_k$, and at all resolutions $j < J$ the scaling coefficients are calculated using

$$\lambda_{j,k} = \frac{1}{2}\left(\lambda_{j+1,2k} + \lambda_{j+1,2k+1}\right). \qquad (7)$$

At the lowest resolution, $j = 1$, there is only one scaling coefficient, which stores the mean value of the function F_k. The wavelet coefficients are defined only for resolutions $j = 1$ to $j = J - 1$, and store the differences between the scaling coefficients of adjacent resolutions, i.e. $\gamma_{j,k} = \lambda_{j+1,2k} - \lambda_{j,k} = -\left(\lambda_{j+1,2k+1} - \lambda_{j,k}\right)$. The wavelet coefficients are more usually written

$$\gamma_{j,k} = \frac{1}{2}\left(\lambda_{j+1,2k} - \lambda_{j+1,2k+1}\right). \qquad (8)$$

The complete wavelet transform is described by the scaling coefficient at resolution $j = 1$, i.e. $\lambda_{1,0}$, and the wavelet coefficients at all levels, $\gamma_{j,k}$.

3.2. Wavelets on the sphere

The Haar wavelet transform can be applied to two-dimensional functions, such as images, as well as one-dimensional time series. Extensions of Haar wavelets to the surface of a sphere (Tenorio et al., 1999; Barreiro et al., 2000; McEwen and Scaife, 2008; McEwen et al., 2011) rely upon hierarchical pixelation schemes. Following McEwen and Scaife (2008) we base our pixelation upon the HEALPIX pixelation scheme [1] (Górski et al., 2005), but for the purposes of the research in this paper choose to develop our own software rather than implement any third-party code.

Like HEALPIX, our software divides the sphere into twelve equal-sized pixels at resolution $j = 1$, and at each higher resolution $j + 1$ we divide each pixel into four new, equal-sized pixels. This division into four new pixels with each increment in resolution, compared to two new pixels in Section 3.1, is simply a consequence of the Haar wavelets being extended from the one-dimensional case to the two-dimensional case. The number of pixels at resolution j, denoted N_j, is therefore given by $12 \times 4^{j-1}$. The Haar scaling function $\phi_{j,k}(\hat{s})$, and wavelets $\psi_{j,k}^m(\hat{s})$, are shown in Fig. 3.

To generate the scaling and wavelet coefficients we base our method upon the description in McEwen and Scaife (2008), where the scaling coefficients, $\lambda_{j,k}$, for resolution j and pixel k are defined as

$$\lambda_{j,k} = \frac{1}{4}\left(\lambda_{j+1,4k} + \lambda_{j+1,4k+1} + \lambda_{j+1,4k+2} + \lambda_{j+1,4k+3}\right), \qquad (9)$$

and the wavelet coefficients, $\gamma_{j,k}^m$, of type m are defined as

$$\gamma_{j,k}^0 = \frac{\sqrt{A_j}}{4}\left(\lambda_{j+1,4k} - \lambda_{j+1,4k+1} + \lambda_{j+1,4k+2} - \lambda_{j+1,4k+3}\right),$$

$$\gamma_{j,k}^1 = \frac{\sqrt{A_j}}{4}\left(\lambda_{j+1,4k} + \lambda_{j+1,4k+1} - \lambda_{j+1,4k+2} - \lambda_{j+1,4k+3}\right), \qquad (10)$$

$$\gamma_{j,k}^2 = \frac{\sqrt{A_j}}{4}\left(\lambda_{j+1,4k} - \lambda_{j+1,4k+1} - \lambda_{j+1,4k+2} + \lambda_{j+1,4k+3}\right),$$

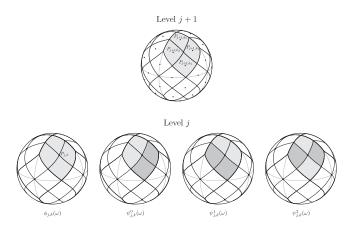[1] http://healpix.jpl.nasa.gov

Figure 3: Haar scaling function $\phi_{j,k}(\hat{s})$ and wavelets $\psi_{j,k}^m(\hat{s})$ (image from McEwen and Scaife 2008). Negative, zero, and positive constant values are indicated with dark shaded regions, unshaded regions, and light shaded regions respectively, and the scaling function and wavelets at resolution j and index k are non-zero only on pixel $P_{j,k}$. Each pixel at resolution j is divided into four pixels at level $j+1$.
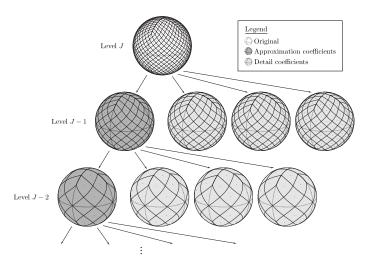

Figure 4: The recursive computation of scaling coefficients, $\lambda_{j,k}$, and wavelet coefficients, $\gamma_{j,k}^m$ (image from McEwen and Scaife 2008). The coefficients are first computed for resolution level J, and then recursively for levels $j = J - 1$, $j = J - 2$, etc. using Eqns. 9 and 10.

where A_j is the area of a single pixel at resolution j in steradians. At the most detailed resolution, $j = J$, the scaling coefficients are populated directly from the value of the spherical function at that position (see Fig. 4), i.e. $\lambda_{J,k} = F(\hat{s}_k)$, where $\hat{s}_k$ is a unit vector pointing from the origin to the position of pixel k, and the coefficients for all levels $j < J$ are computed recursively using Eqns. 9 and 10. Using these coefficients, full reconstruction of the original spherical function, $F(\hat{s})$, is possible using (from McEwen and Scaife 2008)

$$F(\hat{s}) = \sum_{k=0}^{N_1-1} \lambda_{1,k}\, \phi_{1,k}(\hat{s}) + \sum_{j=1}^{J-1} \sum_{k=0}^{N_j-1} \sum_{m=0}^{2} \gamma_{j,k}^m\, \psi_{j,k}^m(\hat{s}), \tag{11}$$

where N_j is the number of pixels at resolution level j, $\phi_{j,k}(\hat{s})$ is the scaling function for pixel k at resolution j, J is the most detailed resolution used, and $\psi_{j,k}^m$ is the wavelet of type m for pixel k at resolution j. However, for radio imaging as presented here inversion is not required as the original spherical function, $F(\hat{s})$, does not need to be recovered. Therefore, the scaling function, $\phi_{j,k}$, and wavelets, $\psi_{j,k}^m$, do not need to be computed explicitly.

3.3. Spherical wavelet measurement equation

During the major-cycle stage of Cotton-Schwab clean visibilities are normally recovered from the model image through the use of an FFT, followed by a de-gridding step, as described in Section 2.2. In doing so, some degree of accuracy is sacrificed while interpolating over the uv grid, when compared to calculating visibilities directly using Eqn. 1. One could generate visibilities directly from the component list without the need to correct for the w-terms, and without sacrificing accuracy, using a variation of Eqn. 1,

$$V_n = \sum_{m=0}^{N_C-1} C_m \exp[\, i2\pi \boldsymbol{B}_n \cdot \left(\hat{\boldsymbol{C}}_m^{uvw} - \hat{\boldsymbol{P}} \right)\,], \tag{12}$$

where N_C is the number of distinct components identified during cleaning, $B_n = [u_n, v_n, w_n]$ is the baseline vector for visibility V_n, $\hat{\boldsymbol{C}}_m^{uvw}$ is a unit vector pointing from the origin to the position of component C_m on the sphere (where $\hat{\boldsymbol{C}}_m^{uvw} = [c_m^u, c_m^v, c_m^w]$), and $\hat{\boldsymbol{P}}$ is a unit vector pointing from the origin to the phase position. The term 'distinct' is used in our definition of N_C as any clean components resolved to the same HEALPIX pixel at the highest resolution are summed together in the spherical model image, so the run time scales with the number of pixels that have one or more clean components assigned to them.

In practice, the coordinate system used by $\hat{\boldsymbol{C}}_m^{uvw}$ and $\hat{\boldsymbol{P}}$ will match that of $\boldsymbol{B}_n$ in orientation, and be aligned such that $\hat{\boldsymbol{P}} = [0, 0, 1]$, so the dot product in Eqn. 12 becomes $u_n c_m^u + v_n c_m^v + w_n \left(c_m^w - 1 \right)$.

A wavelet decomposition of Eqn. 12 is implemented based upon two spherical functions: the first represents the model image,

$$F^I(\hat{s}) = \sum_{i=0}^{N_C-1} C_i\, \delta(\hat{s} - \hat{\boldsymbol{C}}_i^{uvw}), \tag{13}$$

where δ is the Dirac delta function, and the second spherical function represents the plane wave,

$$F^P(\hat{s}, \boldsymbol{B}) = \exp[i2\pi \boldsymbol{B} \cdot \left(\hat{s} - \hat{\boldsymbol{P}} \right)], \tag{14}$$

which is shown graphically in Fig. 5 for the wavelet resolution levels $j = 1$ to $j = 6$. The visibility equation becomes an integration over the whole unit two-sphere S^2,

$$V_i = \int_{S^2} F^I(\hat{s})\, F^P(\hat{s}, \boldsymbol{B}_i)\, \mathrm{d}\Omega(\hat{s}), \tag{15}$$

where $\mathrm{d}\Omega(\hat{s}) = \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\phi$. This integration can be expressed in wavelet form as a summation involving only the non-zero scaling and wavelet coefficients.

The spherical equivalent of the model image, $F^I(\hat{s})$, is constructed using a SIN projection of the clean components from
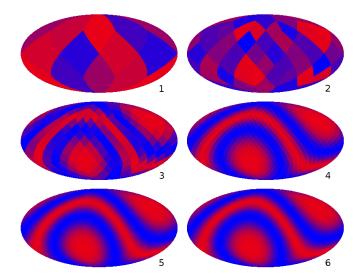
5

Figure 5: Scaling coefficients, $\eta_{j,k}(\boldsymbol{B})$, of the plane-wave function, $F^{\mathrm{P}}(\hat{\boldsymbol{s}}, \boldsymbol{B})$, for the first six resolution levels ($j = 1$ to $j = 6$), shown over the whole unit two-sphere in Mollweide projection.

the image plane to the surface of the sphere. For a clean component C_i at image-plane position $[l_i, m_i]$ the position vector is given by $\hat{\boldsymbol{C}}_i^{\mathrm{uvw}} = [l_i, m_i, \sqrt{(1 - l_i^2 - m_i^2)}]$. We hereafter use $\lambda_{j,k}$ and $\gamma_{j,k}^m$ to represent the scaling and wavelet coefficients of the model image, $F^{\mathrm{I}}(\hat{\boldsymbol{s}})$, and $\eta_{j,k}(\boldsymbol{B})$ and $\sigma_{j,k}^m(\boldsymbol{B})$ to represent the scaling and wavelet coefficients of the plane wave, $F^{\mathrm{P}}(\hat{\boldsymbol{s}}, \boldsymbol{B})$; McEwen and Scaife (2008) used $\delta_{j,k}^m$ to represent the plane-wave wavelet coefficients, but we choose to use $\sigma_{j,k}^m$ to distinguish this function from the Dirac delta function. The visibilities can be reconstructed directly from the resolution-J scaling coefficients, using

$$V_i = \sum_{k=0}^{N_J - 1} \lambda_{J,k} \, \eta_{J,k}(\boldsymbol{B}_i). \tag{16}$$

It can be shown (see McEwen and Scaife, 2008) that the orthogonality of the scaling and wavelet functions for spherical Haar wavelets leads to an expression for the visibilities in terms of the scaling and wavelet coefficients, given by

$$V_i = \sum_{k=0}^{N_1 - 1} \lambda_{1,k} \, \eta_{1,k}(\boldsymbol{B}_i) + \sum_{j=1}^{J-1} \sum_{k=0}^{N_j - 1} \sum_{m=0}^{2} \gamma_{j,k}^m \sigma_{j,k}^m(\boldsymbol{B}_i). \tag{17}$$

Starting with the component list generated by Högbom clean, C_i and $[l_i, m_i]$, the wavelet clean will construct the resolution-J scaling coefficients for the model image and the plane wave ($\lambda_{J,k}$ and $\eta_{J,k}(\boldsymbol{B})$ respectively), and recursively calculate the remaining coefficients $\lambda_{j,k}$, $\gamma_{j,k}^m$, $\eta_{j,k}(\boldsymbol{B})$ and $\sigma_{j,k}^m(\boldsymbol{B})$ for all resolutions j. Only one set of coefficients is calculated for the model image, but for the plane wave function, which is dependent upon the baseline vector, a separate set of coefficients must be generated for each visibility. A flow chart of the whole process is shown in Fig. 6.
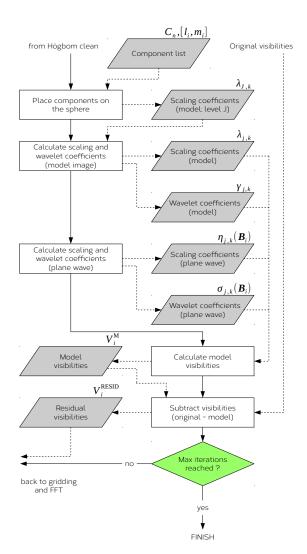

Figure 6: The major cycles of the wavelet clean transfer the clean components onto the sphere, and generate a set of scaling and wavelet coefficients from which the model visibilities can be calculated. The clean components directly set the model-image scaling coefficients at resolution J (the most detailed resolution used), and the remainder of the model-image scaling and wavelet coefficients are calculated recursively for levels $J - 1$, $J - 2$, etc. A new set of plane-wave coefficients must be generated for each visibility.
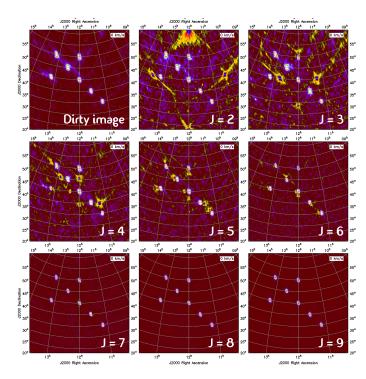
Figure 7: Simulated measurement set with an arrangement of seven compact sources centred at the phase position of $\alpha = 12h, \delta = 45°$, based upon radio antennae with Gaussian primary beams, arranged in the VLA-C configuration. The total observation time was 2 016 s, and all sources were set to a flux density of 1 Jy. The observation frequency was 10 MHz (unphysically low for the VLA, but chosen for a large field of view and low angular resolution), and the field of view 42°40′. The dirty image (top-left panel) was cleaned with 600 CLEAN iterations (minor cycles) over two major cycles, and the remaining panels show the clean images when the resolution parameter was increased from $J = 1$ to $J = 9$.

Table 1: Recommended wavelet resolution, J, for a variety of well-known radio telescopes and configurations. The wavelet resolution was chosen so that each pixel on the sphere oversamples the pixels in the 2-D model image by a factor of at least two.

Telescope	Angular resolution ($\Delta\theta$) [milliarcsec]	J	Wavelet resolution ($\Delta\theta_J$) [milliarcsec]
JVLA, L-band, A-config	1300[a]	20	400
JVLA, X-band, A-config	200[a]	23	50
e-MERLIN, L-band	150[b]	23	50
e-MERLIN, K-band	12[b]	27	3.1
EVN, 18 cm (L-band)	15[c]	26	6.3
EVN, 3.6 cm (X-band)	3[c]	29	0.79

NOTES:
[a] JVLA Observational Status Summary 2018B
(https://science.nrao.edu/facilities/vla/docs/manuals/oss/performance/resolution),
[b] Cycle-7 e-MERLIN observations (http://www.e-merlin.ac.uk/observe/call_cycle7.html),
[c] EVN user guide (http://www.evlbi.org/user_guide/res.html)

3.4. The coordinate system

The preferred coordinate system for radio interferometric data, which we call the uvw coordinate system, is aligned such that the w-axis points at the phase position, $\hat{\boldsymbol{P}}$, the u-axis points east when the phase position is at its zenith, and the v-axis completes the cross product. The other coordinate used by our software, which we call the xyz coordinate system, is aligned with the Earth's axis, such that the y-axis points to the north celestial pole (NCP; declination, $\delta = +90°$), and the x- and z-axes define the plane of the equator, with the x-axis pointed at right ascension (RA) $\alpha = 6$ hours and the z-axis pointed at RA $\alpha = 0$ hours. The xyz coordinate system is used as an intermediate step when calculating the RA and declination of a position on the sphere.

The Euler matrices, R_x, R_y and R_z, are used to perform rotations on the sphere, and the transformation from the uvw to the xyz coordinate system is given by

$$R_{\text{uvw}\to\text{xyz}}(\alpha_P, \delta_P) = R_y(\alpha_P)\, R_x(-\delta_P), \qquad (18)$$

where α_P and δ_P are the RA and declination of the phase position.

3.5. The algorithm

The recommended wavelet resolutions provided in Table 1 require that the software is able to distinguish very large numbers of pixels on the sphere. For example, in L-band, the recommended resolution, J, is 20 for the JVLA (in A-configuration), 23 for e-MERLIN, and 26 for the European Very-long Baseline Interferometer Network (EVN), corresponding to 3.3×10^{12}, 2.1×10^{14}, and 5.4×10^{16} unique pixels respectively.

However, the number of clean components is usually no greater than a few thousand, and the number of unique image pixels amongst those components no larger than a few hundred. Therefore, the vast majority of pixels on the sphere will have zero values, and can be ignored when constructing the plane wave function and the scaling and wavelet coefficients.

We represent our coefficients in a tree structure, as shown in Fig. 8, where each node contains the scaling and wavelet

It is essential when reconstructing visibilities from the model image that the resolution of our wavelet representation is sufficiently high that no significant loss of astrometric precision occurs. Fig. 7 demonstrates the effect of setting the resolution parameter J too low, using a simulated measurement set based upon the Karl Jansky Very Large Array (JVLA) in 'C' configuration. Whereas the clean image constructed using resolution $J = 9$ shows a deep contrast between the seven point sources and the background noise, the clean images constructed using lower values of J show higher levels of background noise, and contain patches of negative flux density at the positions where the clean components have been placed in the spherical representation of the model image. These dark patches are seen to migrate towards the correct component positions as J is increased.

The angular size of each pixel on the sphere (given by $\Delta\theta_j \sim \sqrt{A_j} = \sqrt{4\pi/N_j} = 2^{1-j}\sqrt{\pi/3}$) should be no larger than the angular size of each pixel in the dirty image, and preferably the spherical representation of the model image should oversample the 2-D representation by a factor of two. Table 1 shows the best angular resolution that can be achieved with a variety of well-known radio telescopes and configurations, and the recommended wavelet resolutions.
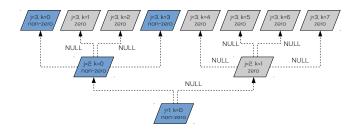
Figure 8: The scaling and wavelet coefficients are stored in a tree structure, with each node containing the coefficients for the model image and plane wave at resolution j, pixel k, as well as a pointer to up to four child nodes at resolution $j + 1$. Most of the pixels contain no clean components, so the plane wave and wavelet coefficients are not calculated for these branches.

Figure 9: The execution time of wavelet clean, in microseconds per visibility. Left: the resolution, J, is varied between 12 and 25, and the performance is shown for fixed numbers of dinstinct clean components, $\widetilde{N}_C$, which are indicated on the right-hand side of the panel. Right: the number of distinct clean components is varied between 10 and 140, and the performance is shown for $J = 12$, $J = 15$, and $J = 25$, also indicated to the right of the panel.

coefficients of pixel k at resolution j, and a pointer to up to four child nodes at resolution $j + 1$. A null pointer indicates that the child node, and all further nodes along this branch, are zero valued and can be ignored. The algorithms that work on these data, such as those that generate the plane wave and those that calculate the scaling and wavelet coefficients, work recursively, and starting from resolution $j = 1$ traverse the tree to find all non-zero nodes.

Before cleaning begins a cross-reference table is constructed that maps each image pixel $[l, m]$ to a pixel k_J on the sphere at the highest resolution level J. From the image pixel $[l, m]$ we determine the position vector on the unit two-sphere, $[l, m, \sqrt{(1 - l^2 - m^2)}]$, which is then transformed from the uvw coordinate system to the xyz coordinate system using the Euler rotation $R_{uvw \to xyz}$. It is then trivial to calculate the right ascension and declination, $[\alpha_{l,m}, \delta_{l,m}]$. The mapping is performed recursively, starting by identifying in which of the twelve pixels at level $j = 1$ we find the spherical coordinate $[\alpha_{l,m}, \delta_{l,m}]$, and then recursively moving down the tree by looking at the child nodes.

When a clean component, C_i, is identified at position $[l_i, m_i]$, the tree structure is traversed from $j = 1$ upwards. The node ID at level J (denoted k_J) is known from the cross-reference table, and the node ID at all lower levels is given by $k_j = k_J >> 2(J - j)$, where $>>$ denotes a bitwise shift to the right, i.e. an integer division by $2^{2(J-j)}$. The nodes k_1 to k_J are created, if they don't already exist, and the relevant pointers from parent to child nodes are set. When a new node is created at level J the plane-wave function at this position on the sphere is calculated immediately. We do not calculate the scaling or wavelet coefficients at levels $j < J$ until all clean components have been found, and the full size of the tree is known.

3.6. Performance

3.6.1. Run-time scaling

The performance cost of the wavelet clean algorithm can be broken down into the steps of building the cross reference between the image pixels and the sphere pixels, generating the scaling and wavelet coefficients of the model image, generating the plane-wave function, generating the scaling and wavelet coefficients of the plane-wave function, and reconstructing the
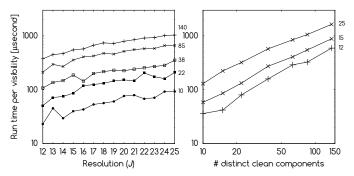
visibilities from the coefficients. The first two steps are performed only once per imaging run, and once per major cycle respectively, and therefore the major computational expense will be provided by the remaining steps, which must be performed once for every visibility.

Following Högbom clean, the N_C distinct clean components, with values C_i and image positions $[l_i, m_i]$, are cross referenced to positions on the sphere at resolution J, resulting in $\widetilde{N}_C$ distinct clean components. The process of traversing the tree to generate the plane wave, computing the scaling and wavelet coefficients, and reconstructing the visibilities all scale as $O(\widetilde{N}_C)$, for each baseline.

In Fig. 9 we show the time taken, per major cycle, by the wavelet clean prototype to process each visibility in the same simulated measurement set used in Fig. 7. In the left panel, the resolution, J, was varied between 12 and 25, and the algorithm performance was found empirically to scale exponentially as $O(1.07^J)$, such that an increase in resolution from $J = 12$ to $J = 25$ results in only a ~2.4-times increase in execution time, and delivers a ~8 200-times improvement in angular resolution. In the right panel, the number of minor cycles was adjusted such that the number of distinct clean components, $\widetilde{N}_C$, varied between 10 and 140, and the performance was found to scale as $O(\widetilde{N}_C)$.

Although generating the plane-wave function required using sine and cosine functions, this stage was responsible for only 5 per cent of the compute time demanded by each visibility. The dominant computational expense was the generation of the scaling and wavelet coefficients at resolution levels $j = 1, ..., J - 1$, which took 60 per cent of processing time. The process of reconstructing the visibilities accounts for the remaining 35 per cent.

A one-hour observation using the JVLA, with a dump time of one second, generates a total of 1.26 million visibilities. A suitable resolution for an L-band, A-configuration observation would be $J = 20$, and generating wavelet coefficients, and reconstructing visibilities, from such a data set would take 1 000 s, per major cycle, if 140 distinct clean components were found, or 146 s if only 15 distinct clean components were

found.

3.6.2. Image quality

NGC 628 (M74) is a face-on spiral galaxy located at a distance of ~7.3 Mpc (Sharina et al., 1996). In Fig. 10 we reduce a 20 kilosecond (ks), 2.5 GHz S-band JVLA observation (in D-configuration) of NGC 628, using a cell size of 3.7 arcsec, and a total of 3 000 minor cycles (see Mulcahy et al. (2017) for a comprehensive discussion of this source and data set). The wavelet resolution was set to $J = 18$, giving 1.6 arcsec resolution on the sphere. The panels show the central 256 x 256 pixel regions of the dirty image (panel a), and clean images (using 20 000 iterations; panel b is cleaned with a conventional CLEAN algorithm, including degridding in the major-cycle stage, and panel c is cleaned with wavelet clean), along with the noise rms as measured in a 400 x 400 pixel region of the images that is free from bright sources.

For these images Wavelet clean produced a lower rms (24.41 μJy) compared to the algorithm using conventional cleaning (31.39 μJy), representing a 22 per cent reduction in noise. Since wavelet clean does not rely upon a flat-plane approximation of the sky when generating model visibilities it follows that the benefits to using this approach are expected to depend heavily upon the FOV of the image.

We have therefore constructed a simulated measurement set (see also Fig. 7) based upon the JVLA in 'C' configuration. The data is a single channel and single polarisation observation of seven compact sources centred on the phase position $\alpha = 12h, \delta = 45°$. We have used an unphysically low observation frequency of 10 MHz, and no primary beam pattern, in order that a large FOV can be tested (~40°).

Fig. 11 compares the dirty image (panel a), constructed over a 1024 x 1024 pixel grid using natural weighting and 64 w-planes, with the clean images produced by cleaning using a conventional CLEAN algorithm (panels b and c; the latter includes the w-kernel in the degridding stage, whereas the former has just the anti-aliasing kernel), and the clean image produced by wavelet clean (panel d, using $J = 10$). The root-mean-square (rms) noise, which is computed for each image from a region of 400 by 400 pixels at the bottom-left corner, was found to be 1.00 mJy for the wavelet-cleaned image (at $J = 10$), a factor of nearly 5x improvement over that of the image obtained using degridding (4.82 mJy).

3.7. Comparison with CASA

The Common Astronomy Software Application (CASA) [2] (McMullin et al., 2007) is a well-established tool for generating cleaned images from radio interferometric data. Combining gridding, fft, and cleaning into a single task, CASA is a benchmark against which the performance of wavelet clean can be assessed.

We hereafter use the term 'CASA clean' to represent the 'clean' method of the imager tool in CASA. We first use the 'defineimage' method to set the image size and cell size, and to

use the multifrequency synthesis (mfs) mode. If w-projection is used then the parameters are set using the 'setoptions' method. Finally, cleaning is performed using the 'clean' method with the Cotton-Schwab algorithm, and a loop gain of 0.1.

In Fig. 12 we plot the rms noise of images from two datasets, each cleaned using CASA clean and wavelet clean, as a function of the number of minor cycles. The top panel shows the rms of the images of NGC 628 (see Fig. 10), and the bottom panel shows the rms of the simulated wide-field data set from Fig. 11.

3.7.1. NGC 628

For NGC 628 we used natural weighting and a wavelet resolution of $J = 18$ (1.6 arcsec angular resolution). The image sizes are 1024 by 1024 pixels in size, and the cell size is 3.7 arcsec per pixel. The rms was calculated from a 400 by 400 pixel region of the clean image in which no bright sources are found. The wavelet clean algorithm and CASA clean tasks were found to clean to a comparable level of noise (the rms is 23.5 μJy and 27.2 μJy respectively). We do not at this stage compare the respective run times of wavelet clean and CASA clean, except to note that the execution time of our current implementation of wavelet clean is considerably longer than that of CASA clean.

3.7.2. Simulated wide-field data

For the simulated wide-field data we constructed clean images that are 1024 by 1024 pixels in size, have angular sizes of 42 deg, 40 minutes along each axis, and a 150 arcsecond cell size. W-projection was used for both CASA clean and wavelet clean (gridding only), and the number of w-planes used is indicated on the plot next to each curve.

As the number of w-planes is increased the CASA cleaned images show a diminishing return in the improvement of the rms noise, ultimately levelling out at a similar noise level to that found in the wavelet cleaned images that used eight w-planes during gridding. When 64 w-planes were used for both wavelet clean and CASA clean we find the rms noise of wavelet clean images to be a factor of more than nine times lower than those of CASA clean images (359 μJy and 3.29 mJy respectively.)

4. Conclusions

Generating model visibilities directly from a component list is a more accurate, but potentially computationally expensive, means of cleaning radio images, when compared to the more traditional method of performing a forward FFT and degridding. Forthcoming radio interferometers, such as the Square Kilometer Array (SKA), will not only offer large fields of view, and require large images (thousands of megapixels), but also generate extremely large numbers of visibilities (a few million per second). Although generating model visibilities directly eliminates the need for a forward FFT during each major cycle, we instead must sum over all items in the component list, for each visibility. Consequently, this method may be more efficient in cases where the number of clean components is small, and the size of the image is large. In this paper we have described our prototype which builds a full-sky model image from
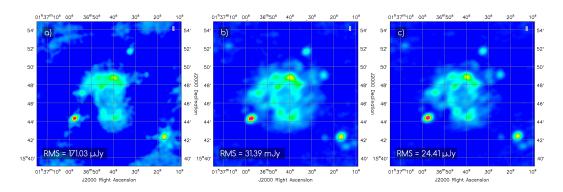

Figure 10: a) Dirty image constructed from a JVLA (D-config) observation of NGC 628 (from 13 March 2013, project 13A-014), generated using natural weighting. b) Image cleaned with 20 000 CLEAN iterations using a conventional CLEAN algorithm that includes degridding. c) Image cleaned with 20 000 iterations using wavelet clean.
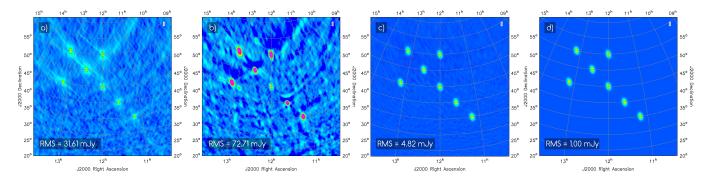

Figure 11: a) Dirty image from a wide-field simulated measurement set (See Fig. 7 for description), generated using 64 w-planes. b) Image cleaned with 700 CLEAN iterations over four major cycles using a conventional CLEAN algorithm that includes a forward FFT and degridding. For degridding we used only the anti-aliasing kernel, whereas both the anti-aliasing kernel and w-kernel (64 w-planes) were used during gridding. c) Image cleaned with 700 CLEAN iterations over four major cycles using a conventional degridding algorithm to calculate the model visibilities. We have used w-projection, with 64 w-planes, in both the forward and inverse directions. d) The same measurement set, cleaned with 580 CLEAN iterations over four major cycles, and using 64 w-planes while gridding and the wavelet clean algorithm in place of degridding.
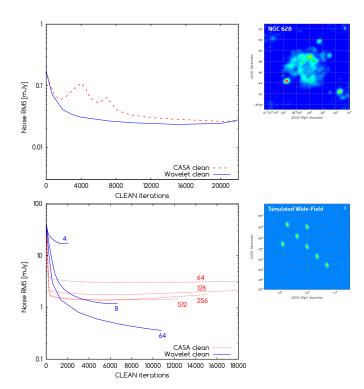
10

the component list, and decomposes this image into spherical wavelet coefficients.

The sparsity of the model images generated during cleaning makes a wavelet method of simulating visibilities a feasible option, whereas computing those visibilities using the spherical harmonic method is slower, and scales poorly (McEwen and Scaife, 2008). Our prototype exploits this sparsity by only storing the non-zero coefficients, and navigating these components efficiently using a tree structure that links parent nodes directly to their child nodes.

We find that the image fidelity of wavelet-clean images is better than that of images cleaned with CASA, and the improvement in quality is particularly noticeable when the image FOV is large. For our wide-field (∼42° FOV) simulated dataset we observe a 8x improvement in the rms noise of the clean image (using the same number of w-planes whilst gridding), and a much smaller improvement of ∼9.5 per cent in an S-band JVLA observation of NGC 628 (without w-projection).

Acknowledgements

References

Barreiro, R.B., Hobson, M.P., Lasenby, A.N., Banday, A.J., Górski, K.M., Hinshaw, G., 2000. Testing the Gaussianity of the COBE DMR data with spherical wavelets. MNRAS 318, 475–481. astro-ph/0004202.

Booth, R.S., Jonas, J.L., 2012. An Overview of the MeerKAT Project. African Skies 16, 101.

Clark, B.G., 1980. An efficient implementation of the algorithm 'CLEAN'. A&A 89, 377.

Cornwell, T.J., Voronkov, M.A., Humphreys, B., 2012. Wide field imaging for the square kilometre array, in: Image Reconstruction from Incomplete Data VII, p. 85000L. 1207.5861.

Dewdney, P.E., Hall, P.J., Schilizzi, R.T., Lazio, T.J.L.W., 2009. The Square Kilometre Array. IEEE Proceedings 97, 1482–1496.

Górski, K.M., Hivon, E., Banday, A.J., Wandelt, B.D., Hansen, F.K., Reinecke, M., Bartelmann, M., 2005. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ 622, 759–771. astro-ph/0409513.

Haar, A., 1910. Zur theorie der orthogonalen funktionensysteme. Mathematische Annalen 69, 331–371.

Högbom, J.A., 1974. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. A&AS 15, 417.

Johnston, S., Taylor, R., Bailes, M., Bartel, N., Baugh, C., Bietenholz, M., Blake, C., Braun, R., Brown, J., Chatterjee, S., Darling, J., Deller, A., Dodson, R., Edwards, P., Ekers, R., Ellingsen, S., Feain, I., Gaensler, B., Haverkorn, M., Hobbs, G., Hopkins, A., Jackson, C., James, C., Joncas, G., Kaspi, V., Kilborn, V., Koribalski, B., Kothes, R., Landecker, T., Lenc, E., Lovell, J., Macquart, J.P., Manchester, R., Matthews, D., McClure-Griffiths, N., Norris, R., Pen, U.L., Phillips, C., Power, C., Protheroe, R., Sadler, E., Schmidt, B., Stairs, I., Staveley-Smith, L., Stil, J., Tingay, S., Tzioumis, A., Walker, M., Wall, J., Wolleben, M., 2008. Science with ASKAP. The Australian square-kilometre-array pathfinder. ExA 22, 151–273. 0810.5187.

Jonas, J.L., 2009. MeerKAT - The South African Array With Composite Dishes and Wide-Band Single Pixel Feeds. IEEE Proceedings 97, 1522–1530.

McEwen, J.D., Scaife, A.M.M., 2008. Simulating full-sky interferometric observations. MNRAS 389, 1163–1178. 0803.2165.

McEwen, J.D., Wiaux, Y., Eyers, D.M., 2011. Data compression on the sphere. A&A 531, A98. 1108.3900.

McMullin, J.P., Waters, B., Schiebel, D., Young, W., Golap, K., 2007. CASA Architecture and Applications, in: Shaw, R.A., Hill, F., Bell, D.J. (Eds.), Astronomical Data Analysis Software and Systems XVI, p. 127.

Figure 12: Top: The rms of wavelet clean and CASA clean images, based upon the data for NGC 628, as a function of the number of CLEAN iterations. We used natural weighting, a cell size of 3.7 arcsec, and a wavelet resolution of $J = 18$. The dashed red lines represent CASA clean, and the solid blue lines wavelet clean. Bottom: The rms of wavelet clean and CASA clean images, based upon the same simulated wide-field data set shown in Fig. 11. We used natural weighting, a cell size of 150 arcsec, and a wavelet resolution of $J = 10$. The number of w-planes used is indicated on the plot (for wavelet clean, w-projection was used in gridding only).

Mulcahy, D.D., Beck, R., Heald, G.H., 2017. Resolved magnetic structures in the disk-halo interface of NGC 628. A&A 600, A6. 1701.04829.

Pratley, L., Johnston-Hollitt, M., McEwen, J.D., 2018. A fast and exact *w*-stacking and *w*-projection hybrid algorithm for wide-field interferometric imaging. ArXiv e-prints 1807.09239.

Schilizzi, R.T., Dewdney, P.E.F., Lazio, T.J.W., 2008. The Square Kilometre Array, in: Ground-based and Airborne Telescopes II, p. 70121I.

Schwab, F.R., 1984. Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry. AJ 89, 1076–1081.

Sharina, M.E., Karachentsev, I.D., Tikhonov, N.A., 1996. Photometric distances to NGC 628 and its four companions. A&AS 119, 499–507.

Tenorio, L., Jaffe, A.H., Hanany, S., Lineweaver, C.H., 1999. Applications of wavelets to the analysis of cosmic microwave background maps. MNRAS 310, 823–834. astro-ph/9903206.