Astronomy & Astrophysics

Efficient correction for both direction-dependent and baseline-dependent effects in interferometric imaging: An A-stacking framework

A. Young[1], S. J. Wijnholds[2], T. D. Carozzi[3], R. Maaskant[4], M. V. Ivashina[4], and D. B. Davidson[1]

[1] Department of Electrical and Electronic Engineering, Stellenbosch University, 7599 Stellenbosch, South Africa
e-mail: ayoung.at.sun@gmail.com
[2] Netherlands Institute for Radio Astronomy (ASTRON), 7990AA Dwingeloo, The Netherlands
[3] Department of Earth and Space Sciences, Chalmers University of Technology, 412 58 Gothenburg, Sweden
[4] Department of Signals and Systems, Chalmers University of Technology, 412 58 Gothenburg, Sweden

ABSTRACT

A general framework is presented for modeling direction-dependent effects that are also baseline-dependent, as part of the calibration and imaging process. Within this framework such effects are represented as a parametric linear model in which basis functions account for direction dependence, whereas expansion coefficients account for the baseline dependence. This separation enables the use of a multiple fast Fourier transform-based implementation of the forward calculation (sky to visibility) in a manner similar to the W-stacking solution for non-coplanar baselines, and offers a potential improvement in computational efficiency in scenarios where the gridding operation in a convolution-based approach to direction-dependent effects may be too costly. Two novel imaging approaches that are possible within this framework are also presented.

1. Introduction

Without properly correcting for various direction-dependent (DD) effects, such as the antenna radiation patterns, ionospheric phase delay, and non-coplanar baselines, the imaging performance of existing and future interferometer arrays may be limited (Smirnov 2011). One aspect of this problem concerns the determination of the various DD effects at the time of an observation, and in this context the characteristic basis function pattern (CBFP) method has been developed to provide an efficient parametrized model (i.e. high accuracy for very few parameters) with which unknown antenna radiation patterns may be solved (Maaskant et al. 2012). However, even when DD effects are known exactly, accurately correcting for them in a computationally efficient manner during the imaging process is difficult since these effects often vary over time as well as among the antenna elements in the array. The latter variation results in such effects also being baseline-dependent (BD), and in turn causes a breakdown of the Fourier transform relationship between the visibility data measured by an interferometer array and the sky brightness distribution (Offringa et al. 2014). This has a significant impact on the computational cost of estimating the sky (imaging), which is often performed iteratively (Bhatnagar et al. 2008; Tasse et al. 2013), and relies on the efficiency of the fast Fourier transform (FFT) to transform between the image and visibility planes.

One class of solutions to this problem accounts for DD effects, which enter as multiplicative distortions to the intensity distribution on the sky, in the visibility domain through utilization of the convolution theorem. Since the visibility sampling provided by a typical array is not regularly spaced on a rectangular grid, as is required by the use of the FFT, additional gridding and degridding steps in the form of a convolution are usually employed to relate visibilities on the FFT grid to those at the array sampling positions (Briggs et al. 1999). Exploiting this already-required convolution step to include DD effects is at the heart of these solutions, which include the *W-projection* (Cornwell et al. 2008) algorithm, which corrects for the non-coplanar baselines effect, and the *A-projection* (Bhatnagar et al. 2008) algorithm which corrects for more general DD effects.

Accurate implementation of these approaches, however, requires a visibility domain convolution kernel of which the support is dependent on the spatial frequency content of the DD effect accounted for, and may result in the computational cost of convolution significantly overshadowing that of the FFT. For non-coplanar baselines an alternative method, called *W-stacking* (Humphreys & Cornwell 2011; Offringa et al. 2014) has been developed to exploit the relatively cheap cost of the FFT by trading a slow convolution operation followed by a single FFT for a faster convolution and repeated FFTs. This method is based on grouping visibilities having similar w-terms together and then performing separate FFTs for each of these visibility groups.

In this paper a novel framework is presented that allows a similar approach to account for more general BD-DD effects. Owing to the similarity to W-stacking, and the extension from non-coplanar baselines to more general DD effects, our approach is called *A-stacking*. Within this framework the prevailing BD-DD effects are represented as a linear model in which the basis functions are DD but baseline-*independent*, while

the expansion coefficients are BD but direction-*independent*. Mathematically this separation of direction-dependence and baseline-dependence is convenient, as it results in the forward calculation (sky to visibility) assuming the form of a combination of separate Fourier transforms, and thus allows fast computation via repeated FFTs. Furthermore, the accuracy of the calculation is determined by the number of terms in the model, thus allowing a simple trade-off between computational cost and accuracy. In the next section the response of an interferometer array is reviewed, followed by Sect. 3 where the A-stacking formulation is presented. In Sect. 4 two alternative imaging approaches that are possible within this framework are developed, and Sect. 5 describes a procedure by which an appropriate linear model may be derived for prior characterized BD-DD effects via the singular value decomposition (SVD). Finally, in Sect. 6 some simulation results are presented to assess the performance of A-stacking based BD-DD models.

Although the framework presented here is generally applicable to any BD-DD effects, results will show that the method is most efficient when these effects are accurately described by low-order (relatively few terms) models. This typically applies to the BD-DD gain associated with the primary beams in an array comprising similar antennas, and the scope of this paper is limited to this particular effect. Furthermore, it is assumed throughout that the BD-DD antenna gains are known a priori; modeling and solving for unknown gains is the subject of ongoing work and will not be discussed herein.

2. Interferometer array response

The full-polarization response on one baseline of a narrow-band phase-tracking interferometer array is given by (see Hamaker et al. 1996; Thompson et al. 2004; Tasse et al. 2013)

$$\mathbf{V}_k = \int_{\Omega_{\mathrm{sky}}} \left(\overline{\mathbf{J}_1} \otimes \mathbf{J}_2 \right) \mathbf{I} e^{-j2\pi \mathbf{u}_k \cdot \boldsymbol{\ell}} \, d\boldsymbol{\ell}, \tag{1}$$

where 1 and 2 denote the antennas that comprise the interferometer formed on the kth baseline, $\otimes$ is the Kronecker product, and $\overline{x}$ indicates the complex conjugate of x. $\mathbf{J}_1$ is the 2×2 Jones-matrix which represents all instrumental effects introduced in the received signal[1] along channel 1. For a non-isotropic antenna the Jones-matrix term is DD and evaluating the Kronecker product in (1) yields a 4×4 direction-dependent matrix $\mathbf{A}_k$. The 4×1 vector $\mathbf{I}$ represents the sky coherency in the polarization coordinates of the antennas.

The vector $\mathbf{u}_k$ is the baseline between the antenna pair $(1, 2)$ expressed in the wavelength normalized Cartesian coordinates (u, v, w) with w pointing towards the center of the field of view (FoV) being tracked, and $\boldsymbol{\ell}$ is a vector representing a direction on the sky relative to the FoV center. Using direction cosines l and m relative to u and v, respectively, to express the direction on the sky yields

$$\mathbf{u}_k \cdot \boldsymbol{\ell} = u_k l + v_k m + w_k \left(\sqrt{1 - l^2 - m^2} - 1 \right) \tag{2a}$$

$$d\boldsymbol{\ell} = \frac{dl\,dm}{\sqrt{1 - l^2 - m^2}}. \tag{2b}$$

For the sake of simplifying notation, the denominator in (2b) is subsumed into the sky coherency. Furthermore, the w-term may be subsumed into the BD-DD gain $\mathbf{A}_k$ as a scalar factor

---

[1] Additive noise is omitted here.

and is omitted in the following. We consider for now only a single cross-correlation product, and assume the antennas have zero cross-polarization[2]. The measurement in (1) can be seen as sampling at the point (u_k, v_k) the visibility function $\mathcal{V}_k(u, v)$ which is related to an apparent sky

$$I_k(l, m) = A_k(l, m) I(l, m) \tag{3}$$

through

$$\mathcal{V}_k(u, v) = \int_{-\infty}^{\infty} A_k(l, m) I(l, m) e^{-j2\pi(ul+vm)} \, dl\,dm, \tag{4}$$

where the support of I_k spans only the visible region of the sky Ω_{sky}. Here the term $A_k(l, m)$ is used to indicate the entry in the first row and first column in $\mathbf{A}_k$ and its direction-dependence is explicitly stated. The subscript k in $\mathcal{V}_k(u, v)$ indicates the baseline-dependence of the sampled visibility function: for a fixed point in the uv-plane, $\mathcal{V}_k(u, v)$ may vary depending on which antenna pair is used to sample the visibility at that point.

Equation (4) forms the basis of synthesis imaging: by measuring visibilities on a large number of unique baselines, the inversion of this relationship becomes more tractable. To this end, we consider the discrete form of (4). We let an $N_{\mathrm{p}} \times N_{\mathrm{p}}$ pixel image of the sky be represented by the $N_{\mathrm{p}}^2 \times 1$ vector $\boldsymbol{\sigma}$. Then the visibility measured on the kth baseline may be written as

$$\mathcal{V}_k(u_k, v_k) = \boldsymbol{\phi}_k^{\mathrm{T}} (\boldsymbol{b}_k \odot \boldsymbol{\sigma}) = \left(\boldsymbol{\phi}_k^{\mathrm{T}} \odot \boldsymbol{b}_k^{\mathrm{T}} \right) \boldsymbol{\sigma} = \boldsymbol{\phi}_k^{\mathrm{T}} \mathrm{diag}\,(\boldsymbol{b}_k)\,\boldsymbol{\sigma}, \tag{5}$$

where $\odot$ is the Hadamard or element-wise product, $\mathrm{diag}(\boldsymbol{x})$ forms a diagonal matrix by placing the elements in the vector $\boldsymbol{x}$ on the diagonal, $\boldsymbol{b}_k$ is the $N_{\mathrm{p}}^2 \times 1$ vector discretization of $A_k(l, m)$ over the image plane, and $\boldsymbol{\phi}_k$ is the $N_{\mathrm{p}}^2 \times 1$ vector in which the element associated with the nth image pixel is

$$\phi_k^{(n)} = e^{-j2\pi(u_k l_n + v_k m_n)}. \tag{6}$$

Grouping all visibilities measured on K unique baselines into a single $K \times 1$ vector yields

$$\mathbf{v} = (\boldsymbol{\Phi} \odot \mathbf{B})\,\boldsymbol{\sigma}, \tag{7}$$

where

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_K \end{bmatrix}^{\mathrm{T}} \tag{8a}$$

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_K \end{bmatrix}^{\mathrm{T}}. \tag{8b}$$

Even though the relation between the measured visibilities and the discretized sky in (7) is relatively simple, for long observations with large arrays the scale of this linear system precludes its direct inversion, and even its direct evaluation as part of an iterative solution may prove too costly (Tasse et al. 2013). An approach that is typically employed to circumvent this problem makes use of the Fourier transform relationship in (4) to enable more efficient calculations via the FFT (Briggs et al. 1999; Jackson et al. 1991). In order to illustrate this approach, we consider the visibility function $\mathcal{V}_0(u, v)$ that would be measured by a hypothetical interferometer array for which $A_k(l, m) = 1$,

$$\mathcal{V}_0(u, v) = \int_{\Omega_{\mathrm{sky}}} I(l, m) e^{-j2\pi(ul+vm)} \, dl\,dm. \tag{9}$$

---

[2] The scalar formulation is presented without loss of generality, and the impact of the full-polarization form of (1) is considered subsequently.
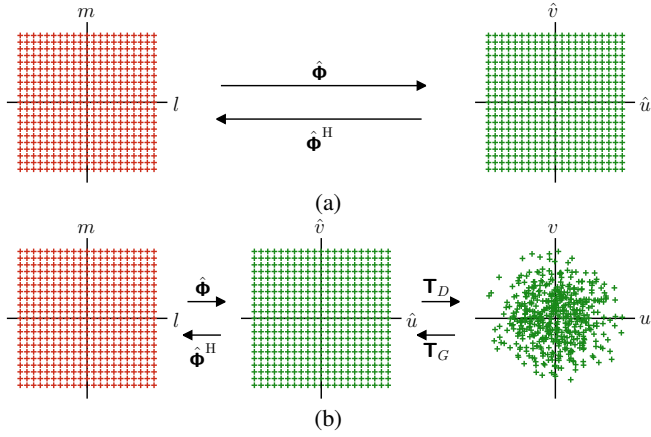
Fig. 1. Transformations between image and visibility domains **a)** for FFT compatible visibility plane sampling, and **b)** for irregular visibility plane sampling which requires an additional (de)gridding operation in order to utilize the FFT.

Since all baselines measure the same visibility function, the true sky $I(l, m)$ may be computed simply using the inverse Fourier transform. Furthermore, if the hypothetical array samples $\mathcal{V}_0$ on an $N_p \times N_p$ rectangular grid $\{(\hat{u}, \hat{v})\}$ which satisfies the following spacing requirements

$$\Delta \hat{u} = \frac{1}{N_p \Delta l}, \qquad \Delta \hat{v} = \frac{1}{N_p \Delta m} \qquad (10)$$

over the uv-plane, where Δl and Δm are the spacing between pixels in the image, then arranging the corresponding visibility samples in the $N_p^2 \times 1$ vector $\hat{\mathbf{v}}_0$ yields

$$\hat{\mathbf{v}}_0 = \hat{\boldsymbol{\Phi}} \boldsymbol{\sigma}, \qquad (11a)$$

where $\hat{\boldsymbol{\Phi}}$ is a discrete Fourier transform (DFT) matrix. Using the unitary property of $\hat{\boldsymbol{\Phi}}$ the inverse transform is easily computed as

$$\boldsymbol{\sigma} = \hat{\boldsymbol{\Phi}}^H \hat{\mathbf{v}}_0. \qquad (11b)$$

Moreover, owing to the structure of $\hat{\boldsymbol{\Phi}}$ the transforms in (11) may be computed efficiently using the FFT. This approach is depicted in Fig. 1a.

There are two caveats associated with the above outlined procedure. Firstly, the uv-sampling provided by any practical interferometer typically does not fall on a rectangular grid, and as such does not satisfy the requirements in (10). This means that, in order to utilize the efficiency of the FFT, it is necessary to relate the visibility samples on the rectangular grid $\{(\hat{u}, \hat{v})\}$, or *gridded* visibilities, to the irregularly sampled visibilities on $\{(u_k, v_k)\}$. That is, the following operators are required, for gridding

$$\hat{\mathbf{v}} = \mathbf{T}_G \mathbf{v} \qquad (12a)$$

and for degridding

$$\mathbf{v} = \mathbf{T}_D \hat{\mathbf{v}}. \qquad (12b)$$

This workaround is illustrated in Fig. 1b.

Secondly, the use of the FFT is based on the assumption that the same visibility function, which may be related to the same apparent sky, is sampled on all baselines. However, it is

well known that under certain conditions this assumption breaks down (Smirnov 2011), e.g. for non-coplanar baselines or non-identical primary beams. This problem is indicated by the dependence on k of both the left hand side of (4), as well as the integrand on its right hand side, as opposed to the form in (9) which is identical for all k. In such a case it is desirable to find some corrective transformation which can relate the baseline-dependent apparent skies (visibility functions) to a single apparent sky (visibility function) common to all baselines.

The A/W-projection algorithms present such a solution and are based on the observation that the image plane multiplicative distortions introduced by the BD-DD term A_k in (4) enter as convolutions in the visibility domain (Cornwell et al. 2008; Bhatnagar et al. 2008). This, combined with the fact that convolution is generally used to approximate the required gridding operations (Briggs et al. 1999; Jackson et al. 1991), means that BD-DD effects may be accounted for through a proper choice of gridding convolution kernel. For instance, given the visibility function $\mathcal{V}_0(u, v)$ uncorrupted by the effects of A_k, the irregularly sampled visibilities may be computed by evaluating

$$\mathcal{V}_k(u, v) = \mathcal{A}_k(u, v) * \mathcal{V}_0(u, v) \qquad (13)$$

at each point (u_k, v_k), and where $\mathcal{A}_k$ is the Fourier transform of A_k. Practical implementation of this approach requires the support of $\mathcal{A}_k$ to be limited to some finite region, which depends on the image plane spatial frequency content of A_k and the required accuracy of the calculation in (13). Consequently the cost of gridding may increase substantially in certain instances by correcting for BD-DD effects in this way.

Alternatively, provided that the computational bottleneck in the transformation between the image and visibility planes is not the FFT itself but the gridding step, a grouping together of baselines for which the direction-dependent effects are sufficiently similar and performing separate Fourier transforms for each group may provide a more efficient solution if it allows the computational cost of gridding to be reduced. This is the approach used in W-stacking to account for non-coplanar baselines (Humphreys & Cornwell 2011; Offringa et al. 2014). In the following it will be shown how the use of a linear model to represent a general BD-DD effect $A_k(l, m)$ results in a similar solution.

3. A-Stacking formulation: forward calculation

Suppose the factor $A_k(l, m)$ in (3) can be written in an exact form as the weighted combination of N_B basis functions[3],

$$A_k(l, m) = \sum_{i=1}^{N_B} a^{(i,k)} f_i(l, m), \qquad (14)$$

where each of the coefficients $a^{(i,k)}$ pertains to the specific baseline k, and the DD expansion functions $f_i(l, m)$ are common to all baselines. Using $\mathbf{f}_i$ to denote the $N_p^2 \times 1$ vector discretization of f_i on the image plane, the discrete form of (14) becomes

$$\boldsymbol{b}_k = \sum_{i=1}^{N_B} a^{(i,k)} \mathbf{f}_i. \qquad (15)$$

---

[3] Strictly speaking an exact expansion is not required for the continuous function, but only for the discretization over the image plane.

Substituting this expression into (5) gives

$$\mathcal{V}_k(u_k, v_k) = \boldsymbol{\phi}_k^{\mathrm{T}} \left[\left(\sum_{i=1}^{N_{\mathrm{B}}} a^{(i,k)} \mathbf{f}_i \right) \odot \boldsymbol{\sigma} \right] = \sum_{i=1}^{N_{\mathrm{B}}} a^{(i,k)} \boldsymbol{\phi}_k^{\mathrm{T}} (\mathbf{f}_i \odot \boldsymbol{\sigma})$$

$$= \sum_{i=1}^{N_{\mathrm{B}}} a^{(i,k)} \boldsymbol{\phi}_k^{\mathrm{T}} \operatorname{diag}(\mathbf{f}_i) \boldsymbol{\sigma}. \tag{16}$$

Using $\mathbf{a}_i = \begin{bmatrix} a^{(i,1)} & a^{(i,2)} & \cdots & a^{(i,K)} \end{bmatrix}^{\mathrm{T}}$ the result for all baselines can be written as

$$\mathbf{v} = \sum_{i=1}^{N_{\mathrm{B}}} \mathbf{a}_i \odot [\boldsymbol{\Phi}(\mathbf{f}_i \odot \boldsymbol{\sigma})] = \sum_{i=1}^{N_{\mathrm{B}}} \operatorname{diag}(\mathbf{a}_i) \boldsymbol{\Phi} \operatorname{diag}(\mathbf{f}_i) \boldsymbol{\sigma}, \tag{17}$$

which is the desired form relating the visibilities to the discretized sky.

Equation (17) states that the visibilities pertaining to a model sky may be calculated as follows, while fully accounting for the BD-DD effects contained in A_k:

1. Apply a per basis function image domain correction in the form of an element-wise multiplication to calculate a corresponding apparent sky,

$$\boldsymbol{\sigma}_i = \operatorname{diag}(\mathbf{f}_i) \boldsymbol{\sigma}. \tag{18a}$$

2. Fourier transform each apparent sky using the FFT, followed by a degridding step, to compute per basis function sets of visibilities,

$$\hat{\mathbf{v}}_i = \hat{\boldsymbol{\Phi}} \boldsymbol{\sigma}_i \tag{18b}$$
$$\mathbf{v}_i = \mathbf{T}_{\mathrm{D}} \hat{\mathbf{v}}_i. \tag{18c}$$

3. Apply a visibility domain correction in the form of an element-wise multiplication to each visibility set, and sum the resulting visibility sets,

$$\mathbf{v} = \sum_{i=1}^{N_{\mathrm{B}}} \operatorname{diag}(\mathbf{a}_i) \mathbf{v}_i. \tag{18d}$$

Extension of the above results to include the full-polarization response in the forward calculation is straightforward, and simply uses a separate model similar to (14) for each of the sixteen elements in $\mathbf{A}_k$.

3.1. Computational complexity

In order to identify conditions under which A-stacking may present an efficient alternative to a convolution based approach for the forward calculation, the computational complexity of the algorithm is compared here to that of A-projection. The algorithms are illustrated in Fig. 2.

Assuming a gridding convolution kernel of size $N_g \times N_g$ is required to meet dynamic range requirements (Duijndam & Schonewille 1999), the overall cost incurred by (18) can be shown to scale as

$$O(C_{\text{A-stack}}) = N_{\mathrm{B}}(N_{\mathrm{p}}^2 + N_{\mathrm{p}}^2 \log N_{\mathrm{p}}) + N_{\mathrm{B}} K N_g^2, \tag{19}$$

where the first term on the right-hand side accounts for the per basis function image plane correction and FFT, and second term

accounts for the gridding and visibility plane correction. In comparison the overall cost of an A-projection implementation, assuming a convolution kernel of size $N_{g_{\mathrm{A}}} \times N_{g_{\mathrm{A}}}$ is required to accurately account for the associated DD effects, scales as (Jongerius et al. 2014; Offringa et al. 2014)

$$O(C_{\text{A-proj}}) = N_{\mathrm{p}}^2 \log N_{\mathrm{p}} + K N_{g_{\mathrm{A}}}^2. \tag{20}$$

If the overall cost in both cases is dominated by the gridding step, that is

$$K N_{g_{\mathrm{A}}}^2, K N_g^2 \gg N_{\mathrm{p}}^2 \log N_{\mathrm{p}}, \tag{21}$$

then A-stacking presents a more efficient alternative to A-projection on condition that

$$N_{\mathrm{B}} < \left(\frac{N_{g_{\mathrm{A}}}}{N_g} \right)^2. \tag{22}$$

The total storage required to perform the A-stacking forward calculation in Fig. 2a completely in memory scales as

$$O(M_{\text{A-stack}}) = N_{\mathrm{B}} N_{\mathrm{p}}^2 + N_g^2 + N_{\mathrm{B}} K, \tag{23}$$

where the first term accounts for the per basis function image domain correction, the second accounts for the storage of the single gridding convolution kernel, and the last term accounts for the per basis function visibility domain corrections. In comparison, the total storage required to perform the A-projection forward calculation in Fig. 2b completely in memory scales as

$$O(M_{\text{A-proj}}) = N_{\mathrm{p}}^2 + K N_{g_{\mathrm{A}}}^2 + K. \tag{24}$$

Here storage of only a single image and visibility map are required, however a (potentially) different convolution kernel needs to be stored per baseline. The storage required or A-stacking is less than that for A-projection if

$$N_{\mathrm{B}} < \frac{K N_{g_{\mathrm{A}}}^2 + N_{\mathrm{p}}^2 - N_g^2}{N_{\mathrm{p}}^2 + K} \approx \left(\frac{K}{N_{\mathrm{p}}^2 + K} \right) N_{g_{\mathrm{A}}}^2, \tag{25}$$

where the approximation assumes that gridding dominates the computational cost as in (21) (which implies that $K N_{g_{\mathrm{A}}}^2 \gg N_{\mathrm{p}}^2$), and that $N_g^2 \ll K N_{g_{\mathrm{A}}}^2$.

Since $N_{g_{\mathrm{A}}}$ depends on the spatial frequency content of the primary beam patterns on the sky, and N_{B} depends on the inter-element variation among the primary beams, a clear distinction between cases where one algorithm should outperform the other (in the asymptotic limit) in terms of computing time and memory requirements is possible in principle. We also note that the cost of a convolution-based approach scales quadratically with an increase in $N_{g_{\mathrm{A}}}$, whereas that of the stacking approach in (18) scales linearly with an increase in N_{B}. This means that even with a moderately sized $N_{g_{\mathrm{A}}}$ a relatively large number of basis functions N_{B} may still render the proposed method more efficient.

Finally, since the efficiency of A-stacking depends on reducing the required size of the visibility plane convolution kernel, it may be necessary to avoid a convolution based approach to correct for non-coplanar baselines. As stated earlier, the w-term is easily included in the BD-DD modeling approach presented here, however, this may increase the number N_{B} of terms in (14) required to yield an accurate model. Alternatively, A-stacking may be combined with W-stacking (Offringa et al. 2014), which does not affect the cost of gridding, but does result in the number of image plane corrections and FFTs increasing by a factor N_w (number of w-layers). This means that the cost of repeated FFTs may become more expensive than gridding for fewer N_w than when combing W-stacking with A-projection in the so-called *hybrid w-stacking* (see Tasse et al. 2013).

(a) A-stacking
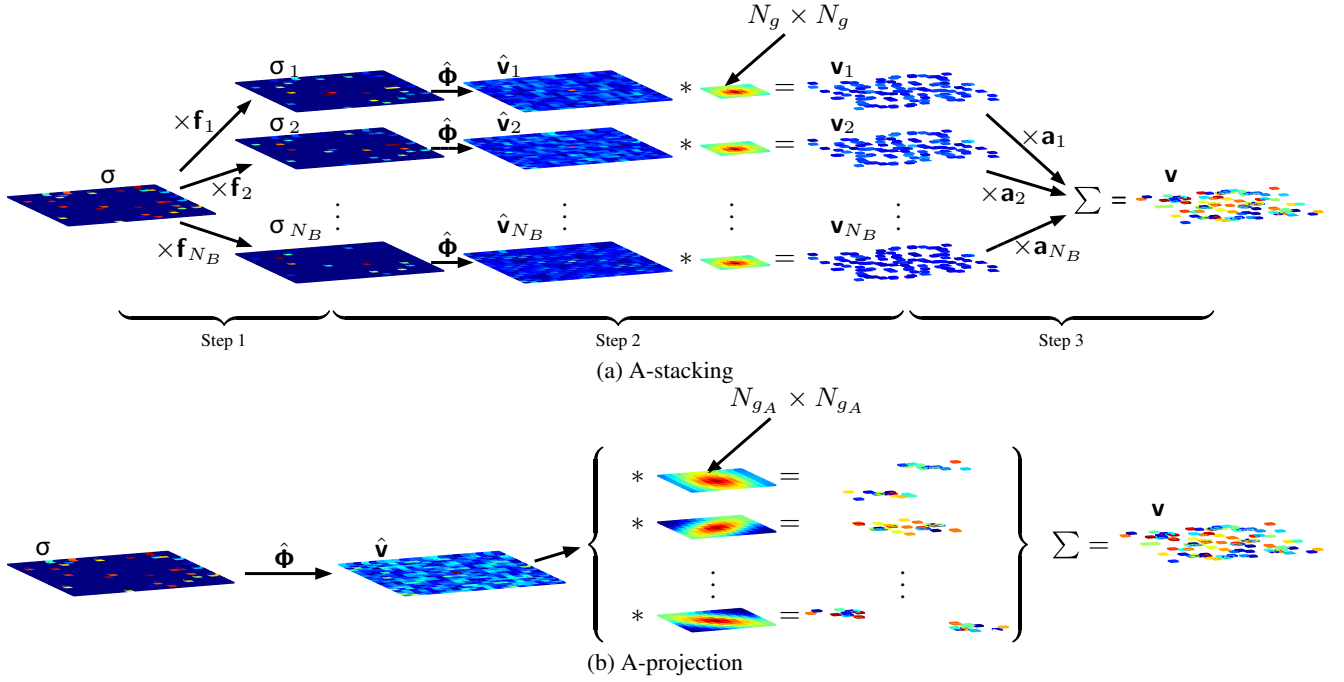
(b) A-projection

Fig. 2. a) Forward calculation in the A-stacking framework, subdivided into the algorithmic steps described by (18). **b)** Forward calculation using A-projection. The A-projection convolution kernel size is $N_{g_A} \times N_{g_A}$, as determined by the spatial frequency content of the direction-dependent effects on the sky, and can result in a relatively expensive gridding cost. A-stacking aims to reduce this cost by decreasing the size of the gridding convolution kernel to $N_g \times N_g$, as determined by the required image dynamic range; the penalty is that multiple FFTs and gridding operations are required.

4. Imaging with A-stacking: backward calculation

Imaging is concerned with the inversion of (7). As was already stated, the scale of this system of equations precludes the use of a direct method, so that a different approach to imaging is required. In this section two approaches possible within the A-stacking framework are derived.

4.1. Adaptation to the CLEAN algorithm

In theory, the Fourier relationship in (4) allows the apparent sky I_k to be determined from the visibility function $\mathcal{V}_k$ via the inverse Fourier transform. Given the discretization inherent to the limited sampling provided by any practical interferometer array and the image representation of the sky, the result of the practical equivalent of this is

$$\sigma_d = \frac{1}{K}\mathbf{\Phi}^H\mathbf{v} = \frac{1}{K}\mathbf{\Phi}^H(\mathbf{\Phi} \odot \mathbf{B})\sigma, \qquad (26)$$

which is the so-called dirty image[4]. Apart from the effects associated with $\mathbf{B}$, the limited sampling of the array also adds a distortion in the form of a convolution with the Fourier transform of the visibility plane sampling function, or point spread function (PSF) (Jackson et al. 1991). Not only does this limit the resolution of the obtained image to the scale of the PSF main lobe, but sidelobes in the PSF can also produce artifacts that may produce false sources, hide existing ones, or distort other sources in the image. Removing this corruption requires some deconvolution procedure. However, owing to the imperfect sampling of

---

[4] In practice this is usually calculated via gridding and applying the FFT, i.e. $\sigma_d = \frac{1}{K}\hat{\mathbf{\Phi}}\hat{\mathbf{v}}$. For the purposes of the present derivation calculation of the dirty image via the direct Fourier transform is used.

the array in general the solution is not unique, so that a non-linear deconvolution procedure is usually required (Cornwell et al. 1999). The algorithm perhaps most widely used for this purpose is CLEAN and its derivatives (Högbom 1974; Clark 1980; Schwab 1984). In general this algorithm is based on identifying a peak in the dirty image as the location of a point-like source, and then removing the effect of that source by subtracting an appropriately scaled and shifted PSF.

In order to demonstrate how this algorithm may be adapted within the A-stacking framework, we consider the dirty image produced by applying $\mathbf{\Phi}^H$ to the visibilities in (17)

$$\sigma_d = \frac{1}{K}\mathbf{\Phi}^H\mathbf{v} = \frac{1}{K}\mathbf{\Phi}^H\sum_{i=1}^{N_B}\mathrm{diag}\,(\mathbf{a}_i)\,\mathbf{\Phi}\,\mathrm{diag}\,(\mathbf{f}_i)\,\sigma$$

$$= \frac{1}{K}\sum_{i=1}^{N_B}\mathbf{\Phi}^H\mathrm{diag}\,(\mathbf{a}_i)\,\mathbf{\Phi}\,\mathrm{diag}\,(\mathbf{f}_i)\,\sigma. \qquad (27)$$

Replacing the true sky image vector σ with an image $\mathbf{e}_s$ corresponding to a single point source of unit intensity at the location (l_s, m_s) and an otherwise empty sky, yields the PSF

$$\mathbf{p}_s = \sum_{i=1}^{N_B} f_i(l_s, m_s)\left[\frac{1}{K}\mathbf{\Phi}^H\mathrm{diag}\,(\mathbf{a}_i)\,\mathbf{\Phi}\mathbf{e}_s\right]$$

$$= \sum_{i=1}^{N_B} f_i(l_s, m_s)\mathbf{q}_i(s). \qquad (28)$$

Herein $\mathbf{q}_i(s)$ is the PSF, centered at (l_s, m_s), and associated with applying the weights $\mathbf{a}_i$ to the visibilities. Although each $\mathbf{q}_i$ represents a shift-invariant PSF, that represented by $\mathbf{p}_s$ is not shift-invariant, because of the direction-dependent weighting $f_i(l_s, m_s)$

applied to each $\mathbf{q}_i$. Assuming a sky composed of a number N_s of point sources allows the following representation

$$\boldsymbol{\sigma} = \sum_{s=1}^{N_s} \sigma^{(s)} \mathbf{e}_s \qquad (29)$$

which, when substituted into (27) and using (28) gives the desired result,

$$\boldsymbol{\sigma}_{\mathrm{d}} = \sum_{i=1}^{N_B} \sum_{s=1}^{N_s} \sigma^{(s)} f_i(l_s, m_s) \mathbf{q}_i(s) = \sum_{s=1}^{N_s} \sigma^{(s)} \mathbf{p}_s. \qquad (30)$$

This expression presents an interesting decomposition of the dirty image: it is the superposition of a number of sub-images, each of which is a different apparent sky $\boldsymbol{\sigma}_i$ convolved with an associated PSF $\mathbf{q}_i$. More importantly, this interpretation can be used to adapt the PSF subtraction in the CLEAN algorithm to account more accurately for the BD-DD effects.

We let the N_B different PSFs $\{\mathbf{q}_i\}_{i=1}^{N_B}$ be pre-computed as part of initializing the CLEAN algorithm, and we let $\boldsymbol{\sigma}_r$ be the residual dirty image at the start of a PSF subtraction iteration. With a point source identified at location s in $\boldsymbol{\sigma}_r$ the PSF subtraction now proceeds as follows:

1. Weigh each PSF $\mathbf{q}_i(s)$ by the value of its corresponding DD basis function towards the direction of the source $f_i(l_s, m_s)$, and accumulate the result for all N_B basis functions to yield the total PSF $\mathbf{p}_s$ as in (28).
2. Scale $\mathbf{p}_s$ according to the intensity of the identified source and the loop gain parameter γ,

$$\mathbf{p}_s \leftarrow \gamma \sigma_r^{(s)} \mathbf{p}_s. \qquad (31a)$$

3. Update the residual image by subtracting the PSF,

$$\boldsymbol{\sigma}_r \leftarrow \boldsymbol{\sigma}_r - \mathbf{p}_s. \qquad (31b)$$

Combining the above procedure with the forward calculation in (18) provides an accurate method by which BD-DD effects may be accounted for in the imaging process.

4.2. Diagonal correction

Using the BD-DD effect model in (14) also produces a useful imaging approach for the case where deconvolution is not necessary, either because the image dynamic range is not high enough or because the spatial selectivity is very good, so that artifacts produced by the PSF sidelobe structure do not have a dominant effect on the image quality.

Suppose an overdetermined system in (7), that is, $K > N_p^2$. The well-known linear least squares (LLS) solution to such a system is

$$\tilde{\boldsymbol{\sigma}} = (\boldsymbol{\Phi} \odot \mathbf{B})^{\dagger} \mathbf{v} = \left[(\boldsymbol{\Phi} \odot \mathbf{B})^H (\boldsymbol{\Phi} \odot \mathbf{B}) \right]^{-1} (\boldsymbol{\Phi} \odot \mathbf{B})^H \mathbf{v}, \qquad (32)$$

where † denotes the Moore-Penrose pseudoinverse. Using the model in (15) and the result in (17), the LSS solution becomes

$$\tilde{\boldsymbol{\sigma}} = \mathbf{M}^{-1} \tilde{\boldsymbol{\sigma}}_{\mathrm{d}}, \qquad (33)$$

where the deconvolution matrix $\mathbf{M}^{-1}$ and the dirty image vector[5] $\tilde{\boldsymbol{\sigma}}_{\mathrm{d}}$ have been introduced,

$$\mathbf{M} = \left(\sum_{j=1}^{N_B} \mathrm{diag}(\mathbf{a}_j) \boldsymbol{\Phi} \, \mathrm{diag}(\mathbf{f}_j) \right)^H \left(\sum_{i=1}^{N_B} \mathrm{diag}(\mathbf{a}_i) \boldsymbol{\Phi} \, \mathrm{diag}(\mathbf{f}_i) \right)$$
$$= \sum_{i=1}^{N_B} \sum_{j=1}^{N_B} \mathrm{diag}(\mathbf{f}_j)^H \boldsymbol{\Phi}^H \mathrm{diag}(\mathbf{a}_j)^H \mathrm{diag}(\mathbf{a}_i) \boldsymbol{\Phi} \, \mathrm{diag}(\mathbf{f}_i) \qquad (34a)$$

$$\tilde{\boldsymbol{\sigma}}_{\mathrm{d}} = \left(\sum_{i=1}^{N_B} \mathrm{diag}(\mathbf{a}_i) \boldsymbol{\Phi} \, \mathrm{diag}(\mathbf{f}_i) \right)^H \mathbf{v}. \qquad (34b)$$

Because of the computational costs involved for typically encountered image sizes, direct evaluation of (33) may not be practicable. However, given the condition that deconvolution is unnecessary, the inversion of $\mathbf{M}$ becomes tractable in that it may be approximated as being diagonal[6]. This results from the fact that the off-diagonal entries in this matrix represent the flux leakage between different pixels in the image, the very effect deconvolution aims to correct.

We let $\mathbf{M}_{\mathrm{diag}}$ be the matrix formed by setting all off-diagonal entries in $\mathbf{M}$ equal to zero. In order to determine the structure of this matrix, we consider the qth diagonal element in

$$\left[\boldsymbol{\Phi}^H \mathrm{diag}(\mathbf{a}_j)^H \mathrm{diag}(\mathbf{a}_i) \boldsymbol{\Phi} \right]^{(q,q)} = \sum_{k=1}^{K} \overline{\Phi^{(k,q)}} a^{(j,k)} a^{(i,k)} \Phi^{(k,q)}$$
$$= \mathbf{a}_j^H \mathbf{a}_i, \qquad (35)$$

where the relation $\overline{\Phi^{(k,q)}} \Phi^{(k,q)} = 1$ has been used. Since the result is independent of q, we can write

$$\mathbf{M}_{\mathrm{diag}} = \sum_{i=1}^{N_B} \sum_{j=1}^{N_B} \mathrm{diag}(\mathbf{f}_j)^H \left[\mathbf{a}_j^H \mathbf{a}_i \boldsymbol{I} \right] \mathrm{diag}(\mathbf{f}_i)$$
$$= \sum_{i=1}^{N_B} \sum_{j=1}^{N_B} \mathbf{a}_j^H \mathbf{a}_i \, \mathrm{diag}(\overline{\mathbf{f}_j} \odot \mathbf{f}_i), \qquad (36)$$

since $\mathbf{a}_j^H \mathbf{a}_i$ is scalar, and where $\boldsymbol{I}$ is an appropriately sized identity matrix.

This naturally leads to the following imaging procedure:

1. Compute the dirty image $\tilde{\boldsymbol{\sigma}}_{\mathrm{d},i}$ for each basis function by applying the weights $\overline{\mathbf{a}}_i$ to the visibilities, Fourier transforming to the image plane (via gridding and using the FFT), and then weighting the image values by $\overline{\mathbf{f}}_i$,

$$\tilde{\boldsymbol{\sigma}}_{\mathrm{d},i} = \mathrm{diag}(\mathbf{f}_i)^H \left[\hat{\boldsymbol{\Phi}}^H \mathbf{T}_G \left(\mathrm{diag}(\mathbf{a}_i)^H \mathbf{v} \right) \right]. \qquad (37a)$$

This is simply a practical implementation of each term in (34b).
2. Accumulate the result for the N_B basis functions to yield the dirty image,

$$\tilde{\boldsymbol{\sigma}}_{\mathrm{d}} = \sum_{i=1}^{N_B} \tilde{\boldsymbol{\sigma}}_{\mathrm{d},i}. \qquad (37b)$$

---

[5] This is not the same dirty image as in (26).
[6] In fact, the diagonal approximation may also be used to speed up deconvolution via an optimization procedure such as the Levenberg-Marquardt algorithm (Marquardt 1963). Such an approximation reduces the Jacobian (and hence the approximate Hessian) of the deconvolution problem $\tilde{\boldsymbol{\sigma}} = \mathbf{M}^{-1} \tilde{\boldsymbol{\sigma}}_{\mathrm{d}}$ to a diagonal matrix, which reduces the order of complexity of the algorithm from $O(N_p^6)$ to $O(N_p^4)$.

3. Compute the diagonal version of the deconvolution matrix $\mathbf{M}_{\text{diag}}$ using (36).

4. Compute the final result,

$$\tilde{\boldsymbol{\sigma}}_{\text{diag}} = \mathbf{M}_{\text{diag}}^{-1} \tilde{\boldsymbol{\sigma}}_{\text{d}}. \tag{37c}$$

5. Basis function construction

Although the above results are generally applicable for any exact expansion of A_k which has the form of (14), it is obviously desirable to obtain such an expansion which requires the least number of terms, since the cost of the above derived algorithms generally scale with N_{B}. In this section an approach aimed at producing such a model from a prior characterized A_k is presented.

For imaging purposes it is sufficient to obtain a model for the discretization of A_k on the image grid, that is $\boldsymbol{b}_k$ as in (15). From (8a) it is clear that finding a set of expansion functions $\{\mathbf{f}_i\}_{i=1}^{N_{\text{B}}}$ with which each $\boldsymbol{b}_k$ may be expressed as in (15) is equivalent to finding a basis for the row space of $\mathbf{B}$. One such basis may be obtained by computing the truncated Singular Value Decomposition (SVD),

$$\mathbf{B}^{\text{T}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{W}^{\text{H}}. \tag{38}$$

Selecting the columns in the $N_{\text{p}}^2 \times N_{\text{B}}$ matrix $\mathbf{U}$ (left-singular vectors) as the basis functions $\{\mathbf{f}_i\}$ in (15), it can be shown that each expansion coefficient $a^{(i,k)}$ is the entry on the ith row and kth column in the matrix[7]

$$\mathbf{a} = \mathbf{U}^{\text{H}}\mathbf{B}^{\text{T}} = \boldsymbol{\Sigma}\mathbf{W}^{\text{H}}. \tag{39}$$

This produces an exact expansion with $N_{\text{B}} \leq \min\left(N_{\text{p}}^2, K\right)$ number of terms, where the inequality may result from a linear dependence among the DD gains associated with each of the baselines[8].

Since the computational burden resulting from the use of the linear model in (15) scales linearly with the number of basis functions, it may be desirable to rather use $n_{\text{B}} < N_{\text{B}}$ number of terms, thus trading computational cost for accuracy. Furthermore, since the accuracy of transformations between the image and visibility planes is also affected by other factors (e.g. noise, gridding/degridding), the required accuracy may need to be only such that it does not limit the overall accuracy. With these considerations in mind, the aim is to produce a model which yields the highest precision for a given number of terms. For this reason, the model provided through the use of the SVD is especially useful in the case where $n_{\text{B}} < N_{\text{B}}$. Specifically, the sum of the squared distances between each of the rows in $\mathbf{B}$ and the vector space spanned by n_{B} left-singular vectors is minimized by choosing those left-singular vectors corresponding to the n_{B} largest singular values on the diagonal of $\boldsymbol{\Sigma}$ (Jolliffe 1986).

It should be noted that the computational cost incurred by evaluating (38) and (39) does not need to enter into the overall cost of the algorithms presented in the previous sections, since the result need only be determined once and can be stored

for repeated use over the course of one or more observations[9]. Nevertheless, one way to alleviate the cost of constructing the BD-DD model is to sample A_k over a sparser grid prior to computing the basis functions. The motivation for this approach is that the resolution obtainable with the entire array may be much higher than that required to accurately represent the radiation pattern of a single antenna element in such an array. We let $\boldsymbol{b}_k'$ be such a discretization of A_k over an $N_q \times N_q$ grid $\{(l', m')\}$ in the image plane, where $N_{\text{p}} = \alpha N_q$ with $\alpha > 1$, and

$$N_{\text{p}}\Delta l = N_q\Delta l', \qquad N_{\text{p}}\Delta m = N_q\Delta m'. \tag{40}$$

Using this discretization of the BD-DD effects the matrix $\mathbf{B}'$ is constructed, and the SVD is computed to yield $N_{\text{B}} \leq \min\left(N_q^2, K\right)$ left-singular vectors $\{\mathbf{f}_i'\}_{i=1}^{N_{\text{B}}}$. The model basis functions are now obtained by first interpolating each singular vector to extend its support onto the $N_{\text{p}} \times N_{\text{p}}$ grid,

$$\mathcal{T}_I : \{(l', m')\} \rightarrow \{(l, m)\}, \qquad \mathbf{g}_i = \mathcal{T}_I(\mathbf{f}_i'), \tag{41}$$

and then orthonormalizing the resulting set of vectors to yield $\{\mathbf{f}_i\}_{i=1}^{N_{\text{B}}}$. Each model coefficient $a^{(i,k)}$ is then computed by projecting the ith basis function onto the kth column of $\mathbf{B}^{\text{T}}$.

6. Results

In this section simulation results are presented in order to assess the performance of the A-stacking approach. First the impact of various factors on the accuracy of the forward calculation is considered, followed by a demonstration of the performance of the CLEAN algorithm when combined with the A-stacking approach.

Simulations pertain to a snapshot observation in a narrow frequency band around 50 MHz, using one polarization of the LOFAR Low Band Antenna (LBA) station at Onsala, Sweden as an interferometer array. A full-wave numerical model was analyzed in FEKO[10] to determine the radiation patterns of all the elements in the array (Young et al. 2014); see Fig. 3. Because of the effects of mutual coupling, which was fully accounted for in the numerical model, the primary beams exhibited a variation over the elements in the array so that a single primary beam approximation would not suffice. The phase reference for each pattern was located at the corresponding antenna position in the array, as used to determine the baseline uv-coordinates. Sky models were generated by randomly placing ten point sources of varying intensity on an image grid, so that the source statistics were in agreement with that reported in Bregman (2012) for frequencies below 1.4 GHz. The sky models were then used as input to various observation simulations. For each simulation the reference (exact) visibilities $\mathbf{v}^{\text{exact}}$ were calculated via direct evaluation of (7), which were then used as input for further analysis.

6.1. Accuracy of the forward calculation

Model visibilities were calculated from the input sky using various approaches, and then compared to the exact visibilities. The sky model extended over the region $|l|, |m| \leq 0.5$ and was

---

[7] The rows in $\mathbf{a}$ are mutually orthogonal owing to the unitary property of $\mathbf{W}$, so that $\mathbf{a}_j^{\text{H}}\mathbf{a}_i = 0$ for $i \neq j$. This result may be used to reduce the complexity of evaluating (36) from $O(N_{\text{B}}^2)$ to $O(N_{\text{B}})$.

[8] For instance, this may apply to earth rotation synthesis where multiple visibility measurements are obtained between the same antenna pair, resulting in identical rows in $\mathbf{B}$. Even when the primary beams are varying in time, e.g. rotating primary beams on the sky for alt-az mount reflector antennas, a linear dependence will still be present if such variation is negligible over time scales much larger than the integration time.

[9] The caveat here is that, where the BD-DD effects are unknown and also need solution, the solvable model may need to be in an appropriate form to allow precomputing (38) and (39). This is not within the scope of the present contribution and is the focus of ongoing work.

[10] https://www.feko.info/
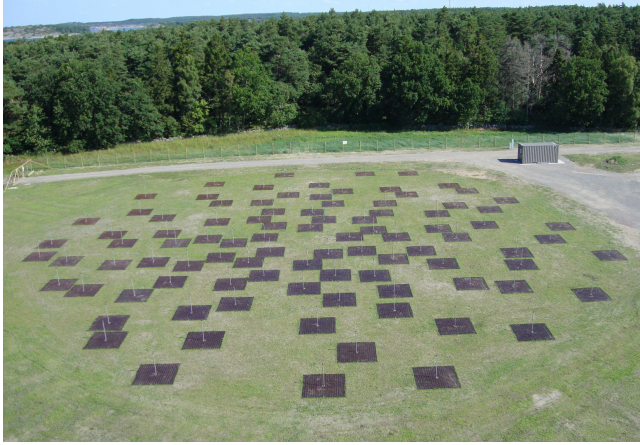
(a)
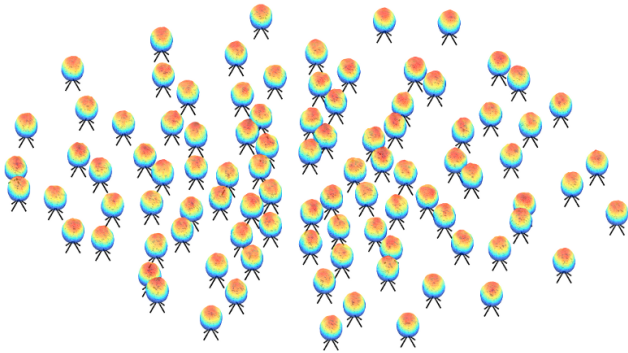
(b)

Fig. 3. a) The LOFAR LBA station at Onsala Space Observatory, Sweden. Photograph courtesy of Leif Helldner. **b)** FEKO model of the 96 element LBA station showing the radiation patterns (magnitude) of each antenna in the array. The array comprises dual-polarized inverted-V antennas above a ground plane (not shown in model). Generally a larger degree of inter-element variability is observed among the patterns of the antennas that are closely spaced than in the patterns of those that are more isolated.

discretized with $N_p = 64$ pixels along each dimension (i.e. 64×64 pixel image). The following model visibilities were computed:

1. $\mathbf{v}^{\text{avg}}$ – Visibilities are calculated using (7), but the DD gain on all baselines is assumed identical and equal to $\boldsymbol{b}_0$, for which the average power pattern over all antennas in the array is used.
2. $\mathbf{v}^{\text{avg-fft}}$ – Again an identical $\boldsymbol{b}_0$ is used for all baselines, but the visibilities are calculated via the FFT, followed by a degridding step which is implemented as cubic interpolation.
3. $\mathbf{v}^{\text{stack}}_{n_B}$ – Visibilities are calculated using (17), but truncating the summation over i after $n_B < N_B$ terms and using basis functions computed via the SVD in (38).
4. $\mathbf{v}^{\text{stack-fft}}_{n_B}$ – Using again only n_B model terms, visibilities are calculated via the procedure outlined in (18), that is, using the FFT and degridding.
5. $\mathbf{v}^{\text{stack-interp}}_{n_B,\alpha}$ – Here $\boldsymbol{b}'_k$ was calculated on an $N_q \times N_q$ grid with $N_p = \alpha N_q$ in the lm-plane prior to computing the SVD in (38). Cubic interpolation was used to extend the support of the basis functions onto the image grid, and the Gram-Schmidt method was then used to orthonormalize the basis functions. Finally, visibilities were computed using (17) and truncating the summation over i after $n_B < N_B$ terms.
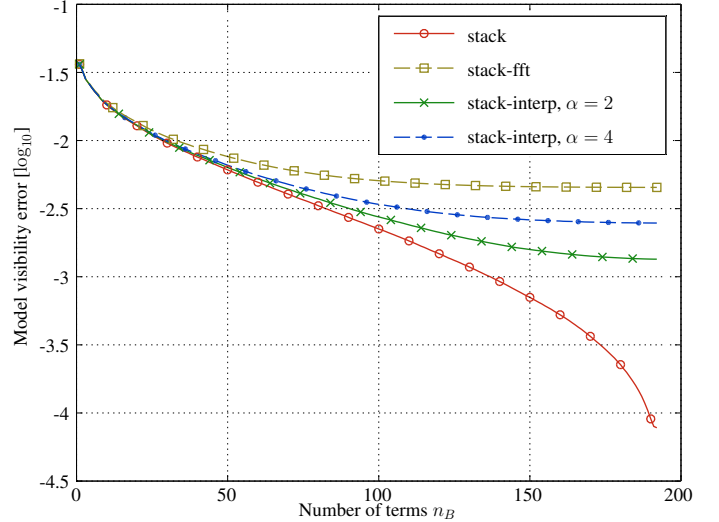
Fig. 4. Error in model visibilities as a function of number of BD-DD gain model terms. The error was computed using the exact visibilities as reference.

In all cases the error between a reference visibility vector and model visibility vector was computed as

$$\epsilon(\text{model}, \text{reference}) = \left(\frac{\|\mathbf{v}^{\text{model}} - \mathbf{v}^{\text{ref}}\|_2}{\|\mathbf{v}^{\text{ref}}\|_2} \right). \qquad (42)$$

Figure 4 shows the accuracy of the forward calculation obtained with various truncated A-stacking based BD-DD gain models. The result shown is the average of $\epsilon(\text{model}, \text{exact})$ over 1000 different generated sky models. As can be expected, in all cases the accuracy improves with an increase in the number n_B of model terms. The highest accuracy is achieved by using the direct Fourier transform in (17), combined with basis functions derived directly from a discretization of A_k on the full image grid (stack model). Since the visibilities computed in this manner converge to the exact visibilities for $n_B = N_B$ this result can be used to estimate the highest accuracy that may be obtained with a model of a given number of terms. The error decays rapidly as n_B is increased for the first few terms ($n_B \lesssim 10$), and thereafter at a somewhat slower rate, until $n_B \approx 192$ where a sharp drop in the error occurs. This sudden decrease in error is related to a similar sudden decrease in the Singular Value (SV) spectrum of $\mathbf{B}^{\text{T}}$ at twice the number of antennas in the array, $2 \times 96 = 192$. Using a sparser discretization of A_k to expedite the construction of the basis functions via the SVD (stack-interp models) is seen to introduce an error between about $10^{-2.6}$ and $10^{-2.9}$, depending on the sparsity of the grid, and only represents a significant loss in accuracy for a model with more than about $n_B = 60$ terms. Similarly, utilizing the FFT-based calculation (stack-fft model), an error of around $10^{-2.3}$ and associated with the degridding operation is introduced which only has a significant impact on the accuracy for a model with more than roughly $n_B = 40$ terms.

In order to put the results in Fig. 4 in perspective, we consider the accuracy for various models reported in Table 1. The error incurred by using a baseline-independent DD gain to compute the visibilities is around -1.404. Using $n_B = 10$ terms of the A-stacking model and computing visibilities via the FFT already reduces the error by more than 50%, and using $n_B = 40$ the error is reduced by more than 75%. (With $n_B = 40$ and using the direct Fourier transform reduces the error slightly more, by about 80%.) This indicates that a large improvement over using a baseline-independent DD gain model is possible with just

Table 1. Various model visibility errors computed using (42).

Model	Reference	$\epsilon(\text{model, reference})$
avg	exact	−1.404
avg-fft	avg	−2.344
Stack, $n_{\text{B}} = 10$	exact	−1.739
Stack, $n_{\text{B}} = 20$	exact	−1.892
Stack, $n_{\text{B}} = 40$	exact	−2.120
Stack-fft, $n_{\text{B}} = 10$	exact	−1.725
Stack-fft, $n_{\text{B}} = 20$	exact	−1.865
Stack-fft, $n_{\text{B}} = 40$	exact	−2.053

Notes. Errors are shown in log-scale.

Table 2. Comparison between memory usage of A-stacking and A-projection forward calculations based on (23) and (24), respectively, and using $K = 9121$ unique baselines.

Method	$N_{\text{p}} = 64$	$N_{\text{p}} = 1024$
A-stack, $n_{\text{B}} = 10$	2.2 MB	177.5 MB
A-stack, $n_{\text{B}} = 20$	4.2 MB	338.9 MB
A-stack, $n_{\text{B}} = 40$	8.3 MB	661.7 MB
A-stack, $n_{\text{B}} = 96$	19.6 MB	1565.5 MB
A-stack, $n_{\text{B}} = 192$	38.9 MB	3114.9 MB
A-proj, $N_{g\text{A}} = 25$	87.0 MB	103.0 MB
A-proj, $N_{g\text{A}} = 50$	348.1 MB	364.1 MB
A-proj, $N_{g\text{A}} = 100$	1391.8 MB	1407.8 MB

**Notes.** Results are shown for small and moderate size images, and for various levels of accuracy in modeling BD-DD gains (number of terms n_{B} for A-stacking, and convolution kernel size $N_{g\text{A}}$ for A-projection). Double precision computation is assumed.

the first few terms of the BD-DD model, even when the calculation is performed using the FFT followed by a degridding step. Finally, the error in the avg-fft model when using avg model as the reference confirms the error introduced by the degridding operation to be around the $10^{-2.3}$ level.

Memory usage statistics for the A-stacking forward calculations based on (23) are listed in Table 2. Results show how the memory requirement increases with the number of terms n_{B} in the BD-DD model for the $N_{\text{p}} = 64$ images used here, and also for $N_{\text{p}} = 1024$ to indicate the requirements for larger images. Since the overall cost scales linearly with the number of terms in the model a simple trade-off between computational cost and calculation accuracy is available. For comparison, the memory requirements for A-stacking are also shown for different convolution kernel sizes. We note that number of baselines relative to the image sizes considered means that the memory requirements for A-projection is almost independent of image size.

6.2. CLEAN performance

In order to demonstrate how A-stacking may impact on the imaging performance, two simple CLEAN algorithms were implemented to use BD-DD gain models of various levels of accuracy. Here the direct transforms in (17) and (27) were used for the forward and backward calculations, respectively, so that the impact of the accuracy of the BD-DD model on the result could be isolated. The sky models used here extended over the same region as before, but used a coarser grid with $N_{\text{p}} = 16$ pixels along each dimension (i.e. 16×16 pixel image).
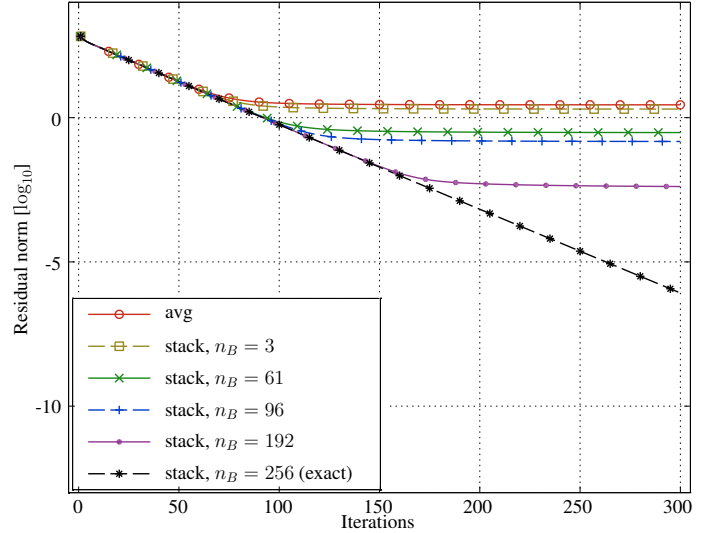
Fig. 5. Convergence of VDSS-CLEAN implementation for DD gain models of various levels of accuracy. Figure shows how the residual image norm is reduced over iterations of source subtraction in the visibility plane.

Visibility domain source subtraction: VDSS-CLEAN algorithm

The first CLEAN algorithm is based on using the accurate A-stacking forward calculation, and proceeds as follows:

1. Initialize the model sky $\boldsymbol{\sigma}_m$ and residual image $\boldsymbol{\sigma}_r$ as

$$\boldsymbol{\sigma}_m = \mathbf{0}, \qquad \boldsymbol{\sigma}_r = \boldsymbol{\sigma}_{\text{d}} \oslash \boldsymbol{b}_0, \tag{43}$$

where $\oslash$ indicates element-wise division, $\mathbf{0}$ is the zero vector, $\boldsymbol{b}_0$ is the average power pattern over all antennas in the array, and $\boldsymbol{\sigma}_{\text{d}}$ is obtained by computing (27) from the visibilities $\mathbf{v}^{\text{exact}}$.

2. The residual norm is calculated as $r = \log_{10}(\|\boldsymbol{\sigma}_r\|_2)$.

3. The peak in $\boldsymbol{\sigma}_r$ is identified, and its location (l_s, m_s) and intensity $\sigma_r^{(s)}$ used to update the model sky

$$\boldsymbol{\sigma}_m \leftarrow \boldsymbol{\sigma}_m + \gamma \sigma_r^{(s)} \mathbf{e}_s, \tag{44}$$

where γ is the loop gain.

4. Model visibilities $\mathbf{v}^{\text{model}}$ are calculated using the model sky $\boldsymbol{\sigma}_m$ in either (5) with the baseline-*independent* gain $\boldsymbol{b}_0$, or in (17) with the A-stacking model and truncating the summation over i after $n_{\text{B}} \leq N_{\text{B}}$ basis functions. This is used to compute residual visibilities,

$$\mathbf{v}^{\text{res}} = \mathbf{v}^{\text{exact}} - \mathbf{v}^{\text{model}}. \tag{45}$$

5. The residual image is updated by substituting $\mathbf{v}^{\text{res}}$ for the visibilities in (27), and applying a DD correction in the form of element-wise division by $\boldsymbol{b}_0$ to the result.

6. Steps 2 through 5 are repeated for a fixed number of iterations.

Figure 5 shows how the residual norm is reduced over the number of iterations of the VDSS-CLEAN algorithm, and the result displayed is the mean of r over 1000 different generated input skies. Assuming that at each iteration a source is correctly identified, and its intensity underestimated, the energy in the residual image should ideally decrease monotonically with each subtraction. A comparison is shown between DD models of various levels of accuracy, ranging from the same average antenna power

Table 3. Mean residual norms (log-scale) after 300 iterations of each CLEAN algorithm for various BD-DD models.

Model	VDSS-CLEAN	IDSS-CLEAN
avg	0.445	0.721
stack, $n_B = 3$	0.298	0.569
stack, $n_B = 61$	−0.516	−0.129
stack, $n_B = 96$	−0.825	−0.422
stack, $n_B = 192$	−2.387	−1.915
stack, $n_B = 256$	−6.076	−4.610

Table 4. Image dynamic range after 300 iterations of each CLEAN algorithm for various BD-DD models.

Model	VDSS-CLEAN	IDSS-CLEAN
avg	32.8 dB	24.0 dB
stack, $n_B = 3$	34.3 dB	25.5 dB
stack, $n_B = 61$	42.4 dB	32.5 dB
stack, $n_B = 96$	45.5 dB	35.5 dB
stack, $n_B = 192$	61.1 dB	50.4 dB
stack, $n_B = 256$	98.0 dB	77.3 dB

Fig. 6. Convergence of IDSS-CLEAN implementation for DD gain models of various levels of accuracy. Figure shows how the residual image norm is reduced over iterations of source subtraction in the image plane.

pattern over all baselines (avg) to the exact A-stacking model (stack, $n_B = N_B = 256$). The intermediate models correspond to keeping only those terms corresponding to Singular Values (SVs) in (38) that are above 1.0% ($n_B = 3$) and 0.1% ($n_B = 61$) relative to the maximum; $n_B = 96$ uses as many terms in the model as antennas in the array, and $n_B = 192$ uses twice as many terms. The latter model is of interest since the SV spectrum exhibits a sharp drop after the 192th SV. In general the algorithm is seen to converge at ever lower values of the residual norm as the model accuracy is increased. The results for each model after the final iteration are summarized in Table 3. Compared to the result for using the average DD gain, using A-stacking with just 3 terms the residual is reduced by about 29%, and with 61 terms by 89%.

The distributions of residual norms after 300 iterations and for all 1000 simulations are shown in Fig. 7 for four different DD models. In 95% of the simulations the average beam model yielded $r < 0.55$, and the A-stack model with $n_B = 3$ and $n_B = 61$ yielded $r < 0.45$, and $r < −0.35$, respectively. Using the exact model resulted in $r < −5.65$ for the same percentage of simulations. For a small fraction of the simulations the algorithm converged to a relatively large residual irrespective of the DD gain model used.

Image domain source subtraction: IDSS-CLEAN algorithm

The second CLEAN implementation is based on the procedure outlined in (31) from Sect. 4.1. Using the measured visibilities $\mathbf{v}^{\text{exact}}$ the dirty image is computed using (27), and the residual image σ_r is initialized to this dirty image without any DD correction as was done for the VDSS-CLEAN algorithm. The algorithm then proceeds by repeatedly performing PSF subtraction via (31) to update σ_r for a fixed number of iterations, and computing the residual norm $r = \log_{10}(\|\sigma_r\|)$ at each iteration.

Figure 6 shows how the residual norm is reduced over the number of iterations for the IDSS-CLEAN algorithm. The result shown is the mean of r over 1000 different generated sky models and a comparison is shown for various DD gain models. Initially the residual decreases at a steady rate for all DD gain models; beyond a certain number of iterations this decrease slows down significantly, and the residual level at which this occurs
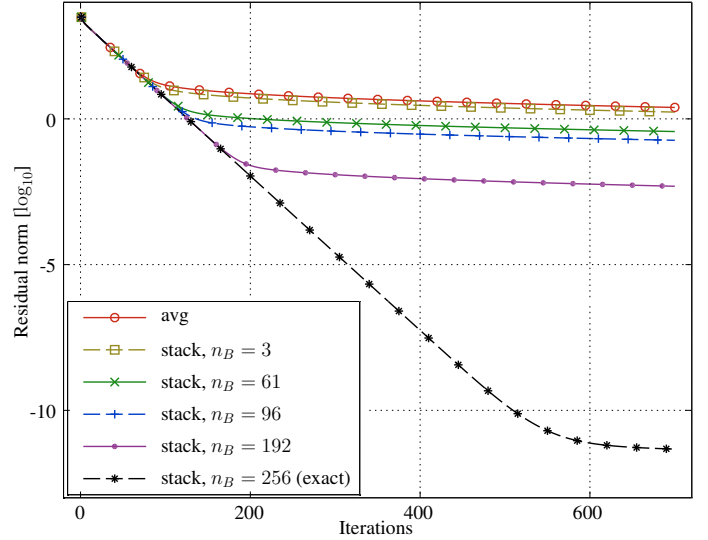
depends once again on the accuracy of the DD gain model. For comparison, the results after 300 iterations of IDSS-CLEAN for the different gain models are also shown in Table 3. Apart from somewhat higher residuals after the same number of iterations as compared to that in VDSS-CLEAN, the decrease in the residual with the increase in the number of terms in the gain model is similar for both algorithms. To help put these results into perspective, the dynamic range of the images obtained after 300 iterations of either CLEAN algorithm was calculated and listed in Table 4.

Assuming that a distinct source (in a distinct location) is identified in each iteration of the IDSS-CLEAN algorithm, a differently weighted combination of the PSFs associated with each of the DD basis functions may be required at each iteration. This combination step requires up to $N_B N_p^2$ complex multiplications and is the computational bottleneck for this image domain deconvolution approach. As can be expected the cost of a single iteration of this algorithm is much cheaper than that for the VDSS-CLEAN algorithm, which includes an image domain correction that also scales as $N_B N_p^2$, as well as the costs associated with transforming between the image and visibility planes. This holds even when utilizing the efficiency of an FFT-based implementation, see (19).

7. Conclusion

A novel framework for modeling baseline-dependent direction-dependent effects was presented. The approach is based on the expansion of BD-DD effects in the form of a weighted sum of basis functions, where the basis functions are direction-dependent, and the coefficients account for the baseline-dependence. Related to the W-stacking method which accounts for non-coplanar baselines, the present approach, called A-stacking, offers an alternative method to the convolution based algorithm A-projection. As such it offers a potential improvement in computational efficiency in scenarios where A-projection results in a significant increase in the gridding cost.

Using the proposed modeling technique the calculation from sky to visibilities is achieved by combining the result from a number of separate Fourier transforms, which may be
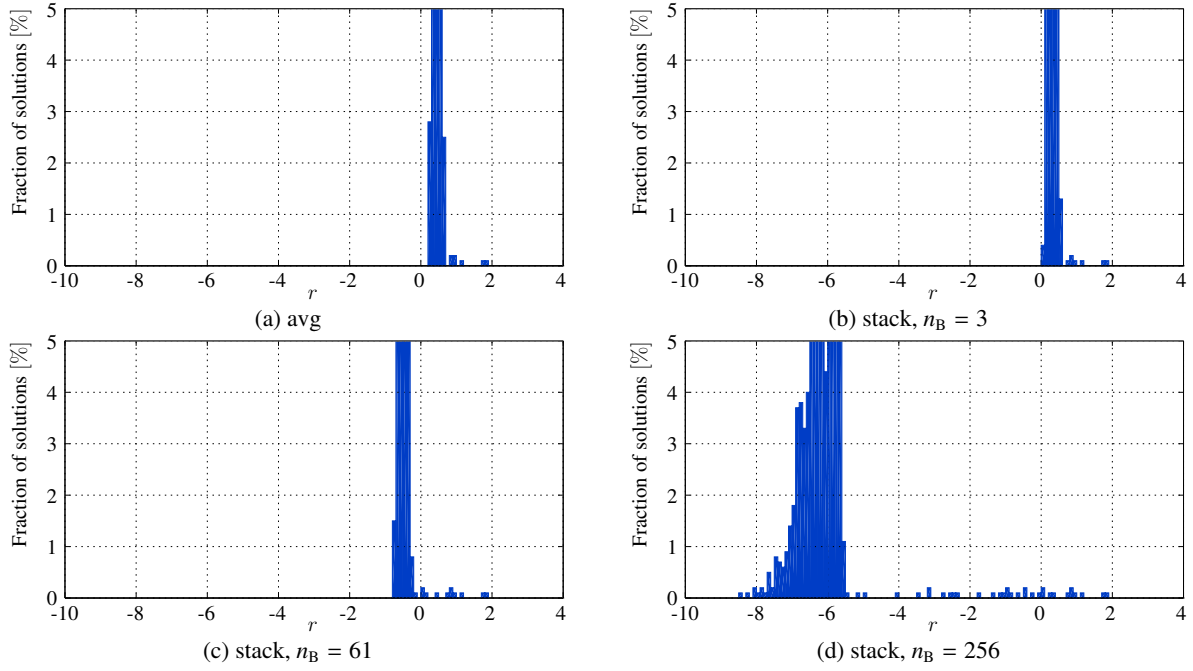
Fig. 7. Distribution of the residual norms after 300 VDSS-CLEAN iterations for 1000 different generated skies. Results are shown for different DD models. We note that the upper limit of the vertical axis has been reduced to improve the clarity of results at lower densities.

implemented to utilize the efficiency of the FFT. The accuracy of this calculation is directly controlled by the number of terms retained in the model, and yields a simple trade-off between accuracy and computational cost. Furthermore, results have shown that good performance may be achieved with relatively few terms, given that an appropriate basis is chosen for the linear model. A method to obtain such a basis for a prior characterized BD-DD effect was presented, and is based on the use of the SVD.

Within this framework, two different imaging strategies were also derived. One strategy takes the form of an adaptation of the PSF subtraction cycle in a typical CLEAN deconvolution process, while the other presents an imaging approach where deconvolution is deemed unnecessary. The use of A-stacking model in two different CLEAN algorithms was also used to demonstrate how this modeling approach may affect image quality. In either case the image residual after a fixed number of iterations was seen to decrease steadily as the model accuracy was improved.

Since the model relies on an accurate characterization of the BD-DD effects, further work focuses on the development of a solvable BD-DD model which is compatible with the A-stacking approach.

References

Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A&A, 487, 419
Bregman, J. D. 2012, Ph.D. Thesis, University of Groningen
Briggs, D. S., Schwab, F. R., & Sramek, R. A. 1999, in Synthesis Imaging in Radio Astronomy II, ASP Conf. Ser., 180, 127
Clark, B. G. 1980, A&A, 89, 377
Cornwell, T., Braun, R., & Briggs, D. S. 1999, in Synthesis Imaging in Radio Astronomy II, ASP Conf. Ser., 180, 151
Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, IEEE J. Select. Topics Signal Process., 2, 647
Duijndam, A. J. W., & Schonewille, M. A. 1999, Geophysics, 64, 539
Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&AS, 117, 137
Högbom, J. A. 1974, A&AS, 15, 417
Humphreys, B., & Cornwell, T. J. 2011, Analysis of convolutional resampling algorithm performance, Tech. Rep. 132, SKA Memo
Jackson, J. I., Meyer, C. H., Nishimura, D. G., & Macovski, A. 1991, IEEE Trans. Medical Imaging, 10, 473
Jolliffe, I. 1986, Principal Component Analysis (Springer)
Jongerius, R., Wijnholds, S., Nijboer, R., & Corporaal, H. 2014, Computer, 47, 48
Maaskant, R., Ivashina, M. V., Wijnholds, S. J., & Warnick, K. F. 2012, IEEE Trans. Antennas Propag., 60, 3614
Marquardt, D. W. 1963, J. Soc. Industr. Appl. Math., 11, 431
Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606
Schwab, F. R. 1984, AJ, 89, 1076
Smirnov, O. M. 2011, A&A, 527, A107
Tasse, C., van der Tol, S., van Zwieten, J., van Diepen, G., & Bhatnagar, S. 2013, A&A, 553, A105
Thompson, A. R., Moran, J. M., & Swenson, G. 2004, Interferometry and Synthesis in Radio Astronomy, 2nd edn. (Wiley)
Young, A., Carozzi, T., Maaskant, R., Ivashina, M. V., & Davidson, D. B. 2014, in Int. Conf. Electromagnetics in Advanced Applications (ICEAA), 462