ATNF SKA Memo 12

On the calibration and imaging with eigenbeams

Maxim Voronkov, Tim Cornwell
CSIRO ATNF, PO BOX 76 Epping, NSW 1710, Australia

January 22, 2007

Abstract

We show that the majority of information is contained in the small number of linear combinations of the measured voltages. Therefore, a small number of synthetic beams, perhaps an order of magnitude smaller than the number of physical feeds, can be used for imaging and calibration. From the other side, only a small number of gains can be calibrated using the full-beam self-calibration approach. An algorithm to compute the beamformer weights for an optimal calibration has been described.

1 Introduction

Wide instantaneous field of view is one of the key science requirements for the Square Kilometre Array (SKA) as it directly affects the survey speed. A significant experience necessary to overcome the challenges of the wide field of view regime can be gained designing the SKA technology demonstrators, such as the xNTD/MIRA and KAT. These instruments will have an array of feeds mounted in the focal plane of a dish to achieve a better performance than conventional interferometers can provide (wider instantaneous field of view and better illumination of the aperture). There are two possible ways of working with the feed array. The first approach pursued by the KAT treats each feed independently. This is equivalent to mosaicing observations, where a number of adjacent pointings is observed simultaneously. The second approach represented by the xNTD employs a beamformer, which reduces the amount of correlated information by forming a certain number of the linear combinations from the signals received by each feed. Although the number of correlator inputs per antenna is less than the number of feeds, a freedom to choose weights of the linear combinations allow different optimization strategies. In this memo we consider a non-adaptive (weights are constant) beamforming in this second approach from the imaging and calibration perspective.

2 Effects of a finite aperture and beamforming

Real antennas have always a finite aperture size. Therefore an interferometer made of a pair of dishes of the diameter D with a separation between dish centres of B, samples a range of spatial frequencies from $B - D$ to $B + D$. A typical baseline length distribution is shown in Fig. 2. The measured visibility function is a linear combination of the radiation coherence function at all sampled scales with the weights corresponding to the reoccurrence of each baseline vector (e.g. Cornwell 1988; Subrahmanyan 2004):

$$V(\vec{u_0}) = \int C(\vec{u})W(\vec{u} - \vec{u_0})\, d\vec{u}, \qquad (1)$$

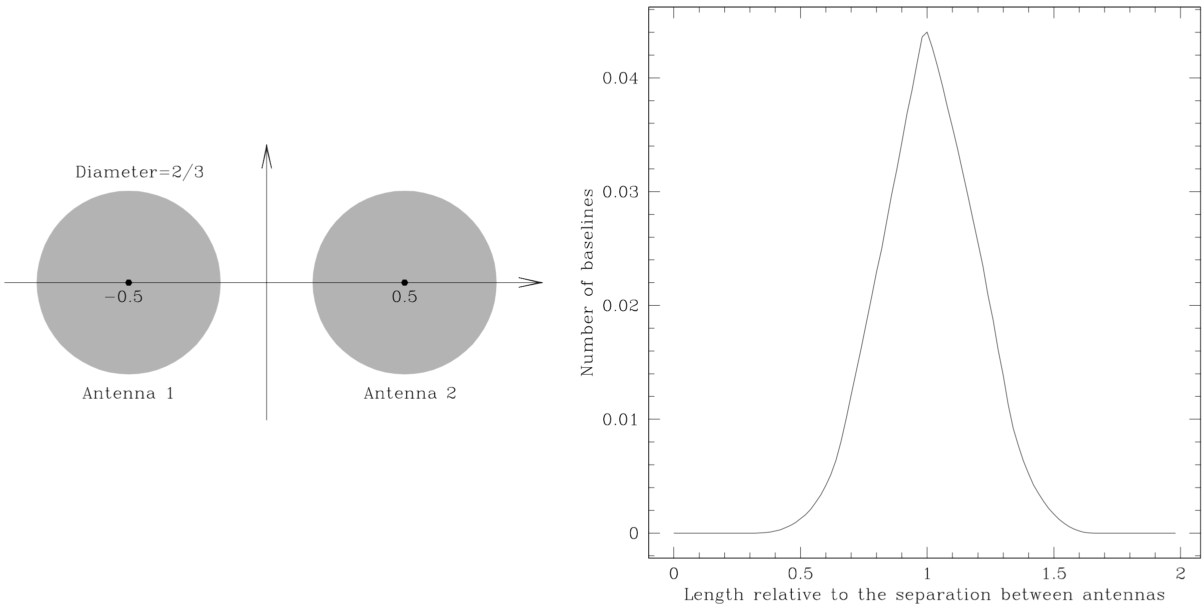

Figure 1: A single antenna pair (left) samples a number of spatial frequencies due to a finite aperture size. The baseline distribution shown in the right figure is calculated under assumption that the antenna diameter is 2/3 of the separation between antenna centres (all length units are given relative to this separation).

where $\vec{u}$ is a spatial frequency (baseline) vector expressed in the units of wavelength ($\vec{u_0}$ corresponds to the spacing between antenna centres), $C(\vec{u})$ is a coherence function which would have been measured by an interferometer pair with infinitesimal small apertures, and $W(\vec{u})$ is a weight. In the image domain this convolution relation corresponds to a multiplication of the true image to a Fourier transform of $W(\vec{u})$, i.e. to the primary beam $A(\vec{s}) = \mathcal{FT}\ W(\vec{u})$. Here $\vec{s}$ is a direction vector in the image domain (a conjugate variable for $\vec{u}$).

By observing the same area of the sky with a set of different $W(\vec{u})$ and, therefore, $A(\vec{s})$, it is possible to recover an estimate of the true image solving the following system

$$\left\{ \begin{array}{lcl} \tilde{I}_1(\vec{s}) & = & I(\vec{s})A_1(\vec{s}) \\ & \dots & \\ \tilde{I}_N(\vec{s}) & = & I(\vec{s})A_N(\vec{s}) \end{array} \right. \quad (2)$$

$$I(\vec{s}) = \frac{\sum\limits_{k=1}^{N} \tilde{I}_k(\vec{s})A_k(\vec{s})}{\sum\limits_{k=1}^{N} A_k(\vec{s})^2} \quad (3)$$

This is a well known linear mosaicing approach. The difference between individual $A_i(\vec{s})$ for conventional interferometers is typically a pointing centre $A_i(\vec{s}) = A(\vec{s} - \vec{s_i})$, $W_i(\vec{u})$ has a phase gradient in this case. The same is true (ignoring distortions) for a multi-feed interferometer where all feeds are treated independently (the pointing centre of an individual feed counts here rather than the dish pointing centre). However, the weights in the system with a beamformer give an additional degree of freedom. Given $A_1(\vec{s}),\ \dots A_N(\vec{s})$ corresponding to N weight sets, an estimate of the sky brightness can still be derived for those offsets, where there is no degeneracy (the denominator in (3) is not numerically close to zero). One can also expect that an appropriate choice of a subset of basis weight vectors (assuming that N is less than the number of feeds) can minimize the degeneracy within a given field of view.

The beamformer calculates a linear combination of input voltages. Therefore, a synthetic voltage pattern is

$$E(\vec{s}) = \sum_l w_l E_l(\vec{s}), \tag{4}$$

where $E_l(\vec{s})$ is a voltage pattern of the lth element and w_l is a corresponding complex weight. The power beam is a quadratic form

$$A(\vec{s}) = E(\vec{s})E^*(\vec{s}) = \sum_{l,m} w_l^* E_l^* w_m E_m = \vec{w}^H \mathcal{E}(\vec{s})\vec{w}, \tag{5}$$

where $\mathcal{E}(\vec{s})$ is a voltage pattern matrix $\mathcal{E}(\vec{s}) = \|E_l^*(\vec{s})E_m(\vec{s})\|_m^l$ for direction $\vec{s}$. Due to a quadratic dependence of $A_k(\vec{s})$ on weight, optimization of (3) becomes a non-trivial task. However, each individual $A_k(\vec{s})$ corresponding to the kth weight vector can be optimized for a certain property using powerful methods of the linear algebra (mainly the singular value decomposition and related subjects, Weisstein 2006). A generalization (by an appropriate weighting) of (3) to measurements with different noise level (e.g., Cornwell et al. 1993; Sault et al. 1996) ensures that the noise is optimized across the mosaic.

The most straightforward approach to this optimization is to optimize an integral

$$F(\vec{w}) = \int A(\vec{w}, \vec{s})K(\vec{s})\, d\vec{s}, \tag{6}$$

where $K(\vec{s})$ is an arbitrary kernel, which essentially defines the desired optimization, and $A(\vec{w}, \vec{s})$ is a synthetic power beam. Substituting (5) to (6) gives

$$F(\vec{w}) = \vec{w}^H \left[\int \mathcal{E}(\vec{s})K(\vec{s})\, d\vec{s}\right] \vec{w} = \vec{w}^H \mathcal{E}\vec{w}, \tag{7}$$

where the matrix $\mathcal{E} = [\cdots]$ does not depend on the direction. The quadratic form (7) attains its maximum under condition of $\vec{w}^H\vec{w} = 1$, if $\vec{w}$ is an eigenvector of $\mathcal{E}$ corresponding to the largest eigenvalue. A synthetic beam corresponding to this weight eigenvector will be called *the first eigenbeam*. If the beamformer allows to calculate several linear combinations simultaneously, they can be configured to form eigenbeams corresponding to the largest eigenvalues of $\mathcal{E}$. The resulting image is given by (3) and such choice of weights ensures that $\sum\limits_{k=1}^{N} F(\vec{w}_k)$ is at maximum. It should be pointed out that this method produces orthogonal weights ($\vec{w}_k^H\vec{w}_l = 0$ for $k \neq l$ as they are eigenvectors of $\mathcal{E}$), and the power beams are not necessarily orthogonal. However, it follows from (4) that the synthetic voltage patterns corresponding to different eigenbeams are orthogonal in the sense that

$$\int E(\vec{w}_p, \vec{s})E(\vec{w}_q, \vec{s})K(\vec{s})\, d\vec{s} = 0 \quad \text{for } p \neq q, \tag{8}$$

where $E(\vec{w}_p, \vec{s})$ and $E(\vec{w}_q, \vec{s})$ are the voltage patterns corresponding to the p-th and q-th eigenbeams (p-th and q-th eigenvectors of $\mathcal{E}$, respectively).

The choise of the kernel $K(\vec{s})$ determines the optimization. One of the possible ways is to optimize with $K(\vec{s}) = I(\vec{s})$, the intensity of the model sky. In this case, the optimization has a clear physical meaning of maximizing the total collected power from the known sources. If the sky model is just a number of point sources, we can write

$$I(\vec{s}) = \sum_{i=1}^{N_{src}} F_i \delta(\vec{s} - \vec{s_i}), \tag{9}$$

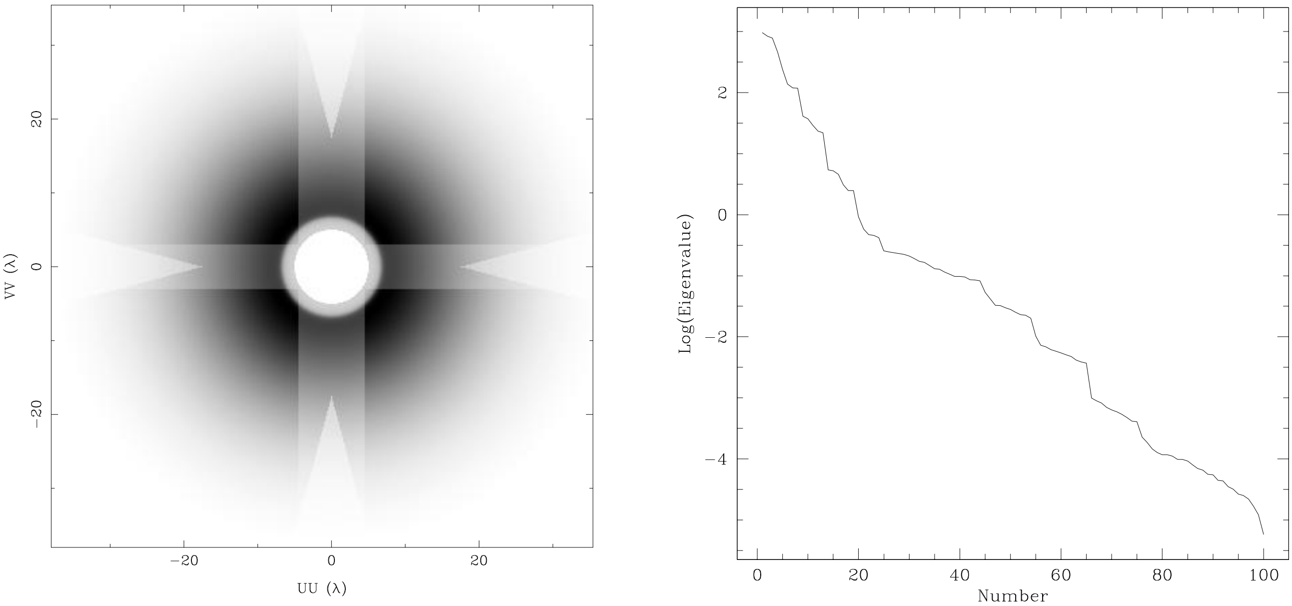

Figure 2: Left. Illumination pattern assumed for all individual elements. Right. The eigenvalue spectrum (in the logarithmic scale).

where F_i, $\vec{s_i}$ are the flux and position of the i-th source. Therefore, the total power captured by the synthetic beam is

$$F(\vec{w}) = \vec{w}^H \mathcal{E} \vec{w} = \vec{w}^H \left(\sum_{i=1}^{N_{src}} F_i \mathcal{E}(\vec{s}) \right) \vec{w}. \quad (10)$$

The first eigenbeam obtained with such a kernel can be used for calibration purposes as the location of the brightest sources serving as calibrators is very well known. However, the imaging would benefit from a uniform sensitivity across the field of view. Therefore, a constant $K(\vec{s})$ within a certain field of view or a tapered function may be a better alternative for imaging.

A similar approach can be used to optimize the ratio of two integral operators. Such problems arise for example if the spillover noise is considered and G/T is optimized (e.g. Brisken & Craeye). However, it requires a detailed knowledge (or model) of the voltage patterns for each feed. In the simulations presented below we assumed that the far field voltage pattern is completely determined by the illumination of the dish surface, i.e. the feed horn responses are confined within the solid angle subtained by the dish. Therefore, neither the synthetic voltage pattern, nor that for individual feeds have a spillover.

3 Results

We simulated a 10×10 array of feeds with the 20 arcmin separation at 1.4 GHz. The dish size was assumed to be 15 metres. The same tapered illumination pattern with a blockage due to the focal plane equipment and the struts supporting it (see Fig. 2) was assumed for all elements. The kernel $K(\vec{s})$ was a circular Gaussian with the scale length of 60 arcmin (FWHM of about 140 arcmin). The spectrum of eigenvalues is shown in Fig. 2 in the logarithmic scale. It falls rather steeply, which means that a small number of eigenbeams contains most of the information from any synthetic beam. Fig. 3 shows the first 15 eigenbeams. It follows from these plots that around 10 eigenbeams cover a 2×2 degree area without holes.

These results have an important implication for calibration. One of the major factors contributing to the gain variation with time is the LNA stability. As each feed has its own LNA, the gains have to be determined for each feed at least once. It can be done using the full-beam self-calibration because the location of brightest sources is very well known and there are plenty of them in the expected field of view at 1.4 GHz. However, the falling eigenvalue spectrum in Fig. 2 means that only a limited number of linear combinations of these gains can be determined with a single pointing of the telescope. This is because the combinations, which correspond to small eigenvalues contribute a negligible amount to the χ^2, which is minimized as a part of the calibration algorithm. However, if the relative gains can be tracked somehow (e.g. a noise source installed in the vertex of the dish illuminating the whole feed array) and both imaging and calibration is done with the same weights set, this small number of linear combinations of gains is all that has to be determined.

4 Weight basis for calibration

Suppose now that we want to calibrate gains for all feeds, e.g. as a part of initial calibration procedure. The naive approach is to observe a strong source sequentially with each feed and set all weights for other feeds to zero

$$[\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_N] = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}, \tag{11}$$

where N is a number of feeds, and $[\ldots]$ represents a matrix with weight vectors as columns. The calibration procedure itself is not different from that for ordinary interferometers. However, the overheads are prohibitive because all feeds must observe the calibrator (Voronkov & Cornwell 2006). It seems more attractive to use all sources present in the field of view. If the beamformer forms more than just one linear combination, less number of dish pointings are required. However, as it follows from the previous section, if the number of feeds is much greater than 10, a robust gain solution cannot be obtained from the single pointing observations, even if the number of beamformer outputs were equal to the number of feeds. Suppose we have an appropriate number of observations with different pointings. In this section we attempt to determine the optimal set of weights for calibration.

Generalizing (10) we want to optimize a sum of N quadratic forms

$$F(\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_N) = \sum_{i=1}^{N} \vec{w}_i^H \mathcal{E}_i \vec{w}_i, \tag{12}$$

where $\mathcal{E}_i$ is a quadratic form matrix defined by (10), which is calculated for the i-th observation (and takes into account a possible pointing change and/or parallactic angle rotation between observations), and $\vec{w}_i$ is a weight vector for this observation. In addition, we want an orthonormal weight basis

$$\vec{w}_i^H \vec{w}_j = \delta_{ij}, \tag{13}$$

where δ_{ij} is the Kronecker delta. If all matrices $\mathcal{E}_i$ are the same (no pointing or parallactic anlge change), the problem reduces to the eigenproblem considered above. However, for different $\mathcal{E}_i$ it becomes more complicated as the eigenvectors corresponding to the largest eigenvalues of two different matrices are rarely orthogonal. Bolla et al. (1998) elaborated an algorithm to maximize (12) under constraints imposed by (13). Strictly speaking the algorithm was derived under assumption that all $\mathcal{E}_i$ matrices are real, symmetric and positive definite. The algorithm involves an iterative polar decomposition of a matrix formed from vectors $\mathcal{E}_i \vec{w}_i$

$$[\mathcal{E}_1 \vec{w}_1, \mathcal{E}_2 \vec{w}_2, \ldots, \mathcal{E}_N \vec{w}_N] = [\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_N]^{new} S, \tag{14}$$

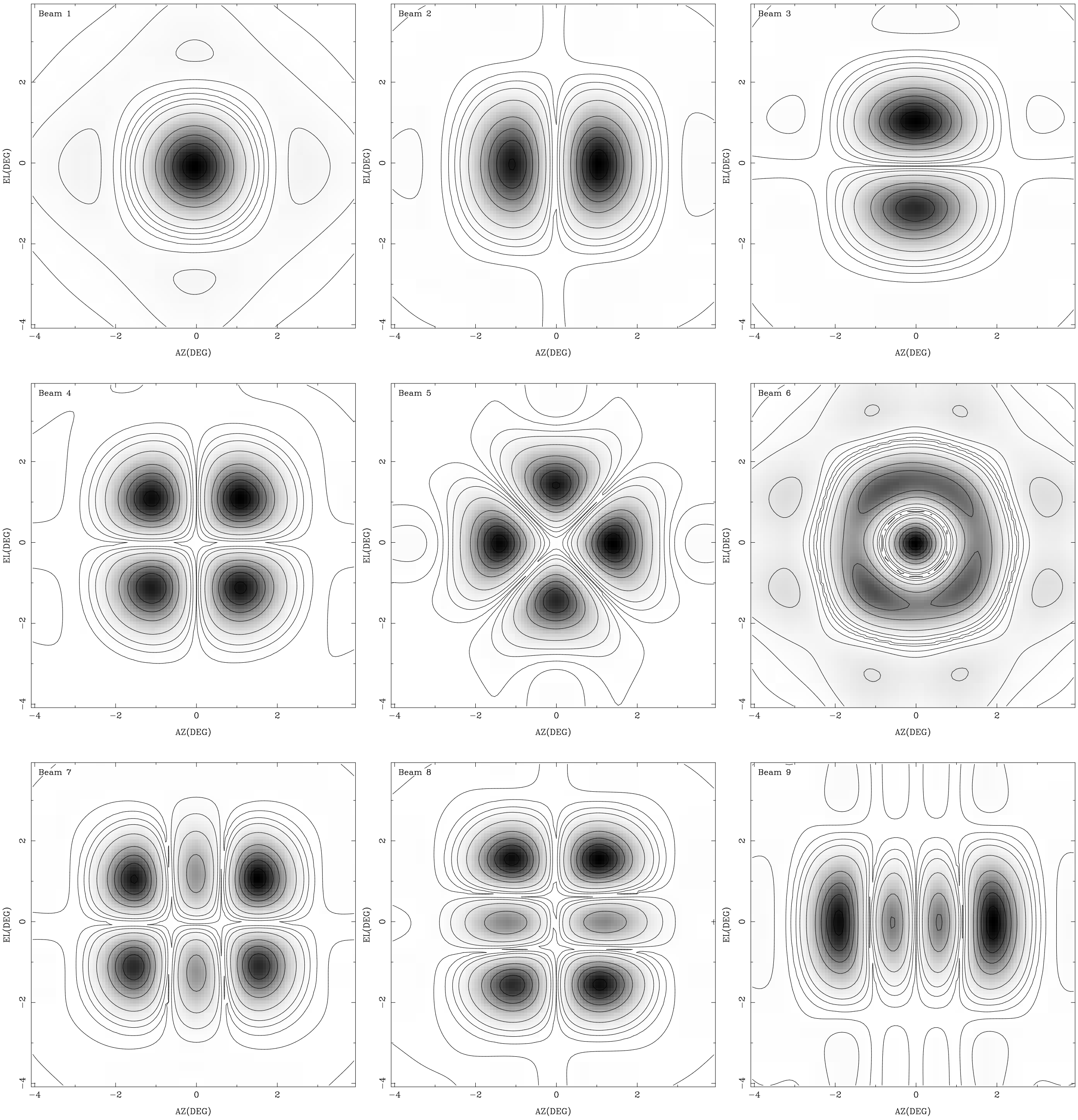

Figure 3: Eigenbeams corresponding to 15 largest eigenvalues. The simulations assumed observations at 1.4 GHz with a 15m dish and a 10×10 array of feeds with a 20 arcmin separation. Contours are at 0.1, 1, 3, 10, 30, 50, 70 and 90%.

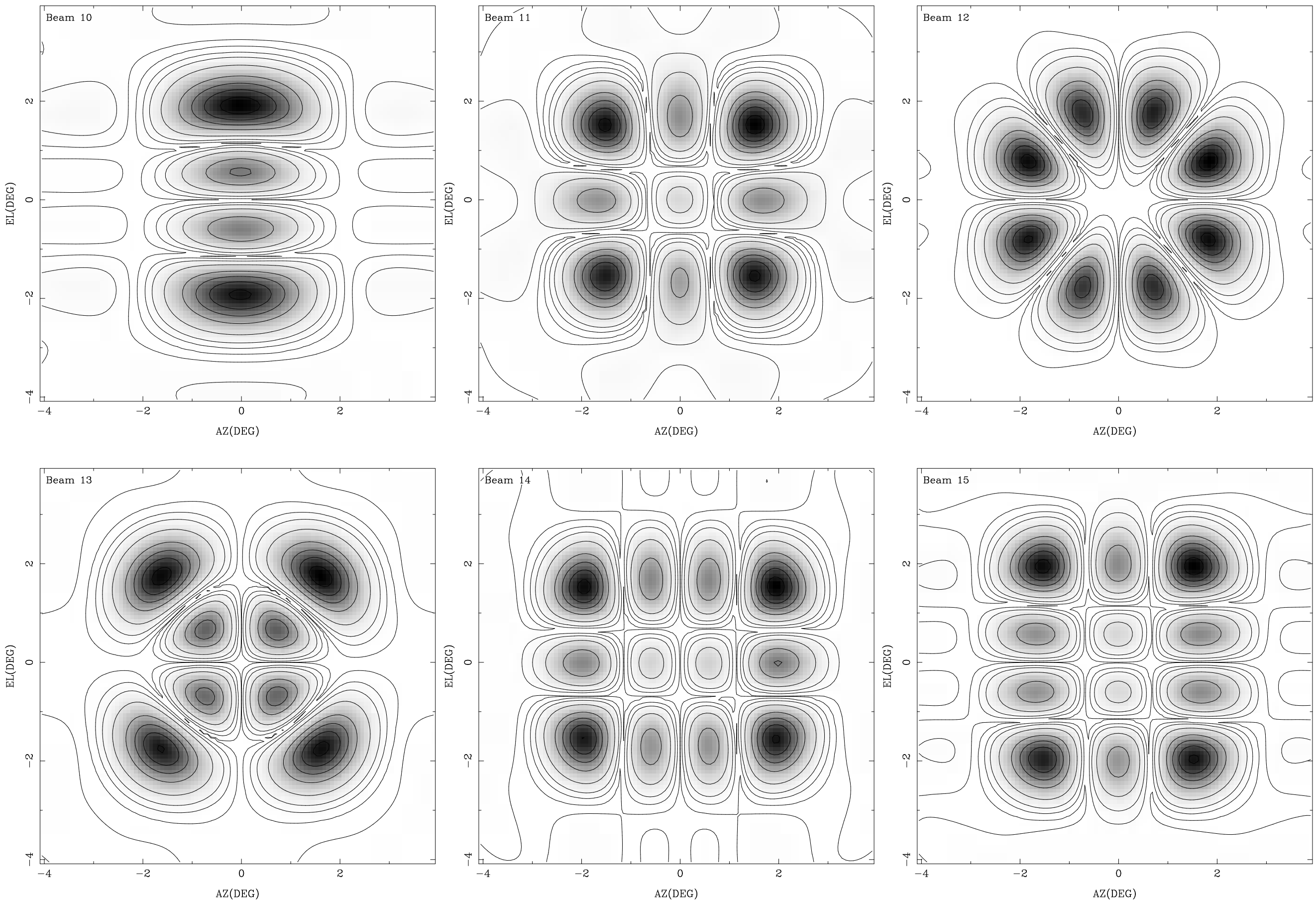

Figure 3: (continued)

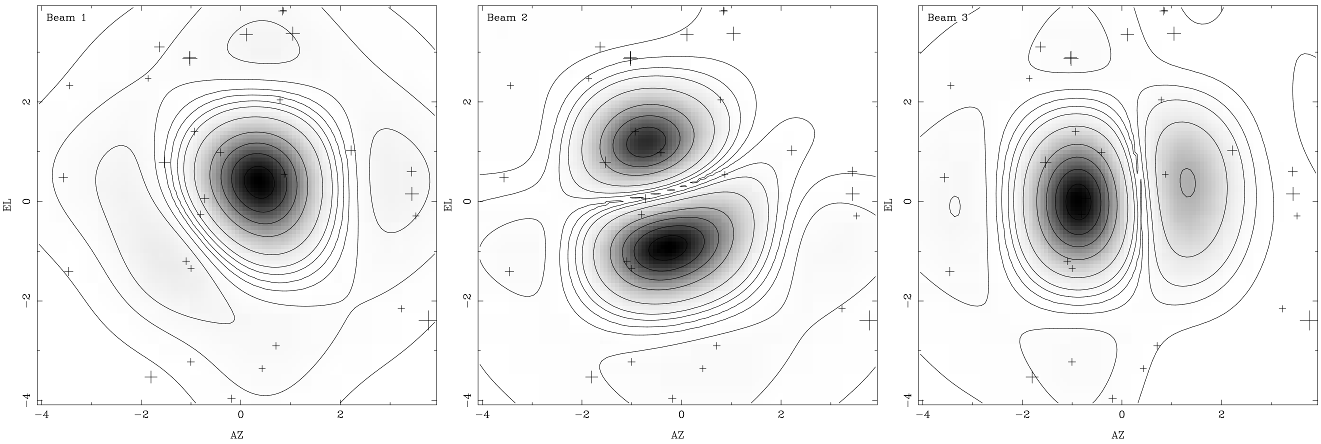

Figure 4: Beams corresponding to the weight vectors obtained by a 3-dimensional basis solution with an NVSS field taken as a sky model. The positions of known sources are marked by crosses (the cross size represent the flux).

where S is a symmetric matrix. The polar decomposition, and therefore a new approximation to the weight basis $[\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_N]$ can be obtained via the Singular Value Decomposition (SVD), for

$$[\mathcal{E}_1\vec{w}_1, \mathcal{E}_2\vec{w}_2, \ldots, \mathcal{E}_N\vec{w}_N] = UWV^H, \tag{15}$$

the new weight set is

$$[\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_N]^{new} = UV^H. \tag{16}$$

Therefore, the algorithm can be written as follows:

1. Form an initial guess to the weight basis, e.g. using (11)
2. Compute a matrix which has $\mathcal{E}_i\vec{w}_i$ vectors in its columns
3. Compute the singular value decomposition of this matrix and form a new approximation to the weights using (16)
4. If the matrix norm of $[\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_N]^{new} - [\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_N]$ is larger than a desired convergence threshold, proceed to step 2.

An example of beams corresponding to a 3-dimensional basis obtained using NVSS catalogue as a sky model is shown in Fig. 4. This algorithm may be hard to extend for optimizing a sum of Rayleigh quotients (ratios of quadratic forms). Therefore, using weights calculated by this algorithm may result in less than ideal signal to noise ratio as the spillover is not minimized. Various heuristics may help (e.g. an additional weighting function reducing the influence of the sources located at the edge), but they require a further study involving a more accurate models of element voltage patterns.

5 Conclusions

1. A relatively small number of synthetic beams (linear combinations in the beamformer) contains most of the information.
2. Only a small number of linear combinations of gains can be determined in the full-beam self-calibration procedure (using all known sources in the field of view). Therefore, one needs to track the relative gains somehow (e.g. by a noise source in the dish vertex illuminating the whole focal plane array installation).

3. For calibration purposes, eigenbeams can be formed by optimizing a total flux detected by each antenna, given the locations of the known sources. A similar optimization problem generates eigenbeams suitable for imaging. These problems can be extended to minimize the noise as well, if one has a good model of the spillover.

4. Imaging with eigenbeams is possible and, moreover, it is not significantly different from the ordinary linear mosaicing.

5. There exists an algorithm to compute an orthonormal weight basis, which provides the largest sum of fluxes attained for each weight vector. This could be an ideal solution if the calibration of all gains is required.

References

Brisken W., Craeye C., 2004, *Focal Plane Array Beam-forming and Spill-over Cancellation using Vivaldi Antennas*, *EVLA Memo*, 69; *http://www.aoc.nrao.edu/evla/geninfo/memoseries/evlamemo69.pdf*

Bolla M., Michaletzky G., Tusnady G., Ziermann M., 1998, *Extrema of Sums of Heterogeneous Quadratic Forms*, *Linear Algebra and its Applications*, 269, 331

Cornwell T.J., 1988, *Radio-interferometric imaging of very large objects*, *A&A*, **202**, 316

Cornwell T.J., Holdaway M.A., Uson J.M., 1993, *Radio-interferometric imaging of very large objects: implications for array design*, *A&A*, **271**, 697

Sault R.J., Staveley-Smith L., Brouw W.N., 1996, *An approach to interferometric mosaicing*, *A&AS*, **120**, 375

Subrahmanyan R., 2004, *Radio interferometers with wide bandwidths*, *MNRAS*, **348**, 1208

Voronkov M.A., Cornwell T.J., 2006, *Notes on the calibration and simulation strategy*, *ATNF SKA Memo 11*

Weisstein E.W., 2006, *Singular value decomposition*, From MathWorld – A Wolfram Web Resource, *http://mathworld.wolfram.com/SingularValueDecomposition.html*