

SIGNAL PROCESSING CHALLENGES FOR RADIO ASTRONOMICAL ARRAYS

S. J. Wijnholds* A.-J. van der Veen† F. De Stefani◇ E. La Rosa◇ A. Farina◇

* Netherlands Institute for Radio Astronomy (ASTRON), Dwingeloo (The Netherlands)

† Delft University of Technology, Delft (The Netherlands)

◇ Selex ES, Rome (Italy)

ABSTRACT

Current and future radio telescopes, in particular the Square Kilometre Array (SKA), are envisaged to produce large images ($> 10^8$ pixels) with over 60 dB dynamic range. This poses a number of image reconstruction and technological challenges, which will require novel approaches to image reconstruction and design of data processing systems. In this paper, we sketch the limitations of current algorithms by extrapolating their computational requirements to future radio telescopes as well as by discussing their imaging limitations. We discuss a number of potential research directions to cope with these challenges.

Index Terms— Array signal processing, radio astronomy, imaging, image reconstruction, computational cost

1. INTRODUCTION

Current and future radio telescopes, in particular the Square Kilometre Array (SKA) [1], are envisaged to produce large images ($> 10^8$ pixels) with over 60 dB (up to 74 dB) dynamic range. The radio astronomical community is currently making detailed plans for the SKA [2], whose data deluge will require Exascale computing. In this paper, which is an introductory paper to the special session on array signal processing for radio astronomy at ICASSP 2014, we discuss the challenges faced by the radio astronomical community based on the current state-of-the-art in radio astronomical imaging techniques and propose several research directions to cope with these challenges.

Radio astronomical imaging arrays like the Low Frequency Array (LOFAR) [3] and the SKA are large arrays with receptors spread over an area of hundreds of kilometers. To keep data rates manageable, the LOFAR and some SKA subsystems exploit a hierarchical processing scheme. In the LOFAR, operating in the 10 – 240 MHz frequency range, the antennas are grouped in subarrays or *stations* with beamforming capability. The beamformed output signal of each station is fed into a correlator, where it is correlated with the beamformed output signals from other stations. The resulting covariances, referred to as *visibilities* in radio astronomy, are used to construct an image of the observed field to which

the stations are steered. The SKA is envisaged to consist of three subsystems, a Low Frequency Aperture Array (LFAA, operating in the 50 – 350 MHz frequency range), an array of reflector dishes with a single feed (SKA-dish) and an array of reflector dishes with a phased array feed (SKA-survey) [1]. Both dish arrays operate above 350 MHz. The LFAA has a similar architecture as LOFAR and hence follows the same beamforming and imaging hierarchy. The array of dishes with phased array feeds operates in a similar way, observing multiple directions within the range allowed by the optics of the reflector dishes. The receiving elements of such synthesis array are thus either stations or dishes (whose feeds can produce one or more beams). In either case, the signals from these receptors are correlated to produce the visibilities from which images can be reconstructed.

In the next section, we provide a brief introduction to image reconstruction with radio astronomical imaging arrays. In Sec. 3 we discuss the current state of the art in radio astronomical imaging to explain the challenges faced by the radio astronomical community with the advent of telescopes like LOFAR and SKA. The computational challenges are discussed in more detail in Sec. 4, which is valuable input to the discussion on trade-offs in the design of the computing platform in Sec. 5. We then sketch an image reconstruction process for the SKA, taking into account the limitations of the algorithms used. This requires accurate and unbiased estimation of many (several thousand) bright sources in the presence of a sea of millions of weaker sources. Our main conclusions are summarized in Sec. 7.

2. SIGNAL PROCESSING FOR IMAGING ARRAYS

The signals from individual receptors, $x_p(t)$ can be stacked in the $P \times 1$ array signal vector $\mathbf{x}(t)$ that can, for a single channel and time slice, be described by

$$\mathbf{x}(t) = \sum_{q=1}^Q \mathbf{a}_q s_q(t) + \mathbf{n}(t) = \mathbf{A} \mathbf{s}(t) + \mathbf{n}(t). \quad (1)$$

In this data model, $s_q(t)$ is the signal from the q th source, $\mathbf{s}(t) = [s_1(t), \dots, s_Q(t)]^T$ and $\mathbf{n}(t)$ is the noise vector. It

is assumed that the source and noise signals are mutually uncorrelated i.i.d. Gaussian signals. Using the narrowband assumption and assuming that the integration time is sufficiently short to assume a fixed array and source geometry, the array response to the incident plane waves can be formulated as

$$\mathbf{a}_q = \left[e^{-jk_q \xi_1}, \dots, e^{-jk_q \xi_P} \right]^T, \quad (2)$$

where $\mathbf{k}_q = 2\pi \mathbf{l}_q / \lambda$ is the wave propagation vector towards the q th source located at \mathbf{l}_q (vector of direction cosines) and ξ_p is the position of the p th sensor. These array response vectors can be stacked in a $P \times Q$ matrix $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_Q]$ to arrive at (1). Please note that we have assumed a perfectly calibrated array in view of the scope of this paper. A detailed discussion on array calibration for radio astronomy can be found in [4]. For the same reason, we also ignore the fact that some additional processing may be needed to remove or otherwise deal with samples that are affected by radio frequency interference (RFI). An overview of RFI detection and mitigation for radio astronomy can be found in [5, 6].

The signals from the individual receptors are correlated to obtain the array covariance matrix

$$\mathbf{R} = \mathcal{E} \{ \mathbf{x}(t) \mathbf{x}^H(t) \} = \mathbf{A} \mathbf{\Sigma} \mathbf{A}^H + \mathbf{\Sigma}_n, \quad (3)$$

where $\mathbf{\Sigma} = \mathcal{E} \{ \mathbf{s}(t) \mathbf{s}^H(t) \}$ is a $Q \times Q$ diagonal matrix with the source powers $\boldsymbol{\sigma} = [\sigma_1, \dots, \sigma_Q]$ on its main diagonal and $\mathbf{\Sigma}_n$ is the diagonal noise covariance matrix. Ignoring the noise covariance matrix for simplicity (see [7] for a more complete treatment), (3) can be vectorized to produce

$$\mathbf{r} = \text{vec}(\mathbf{R}) = (\bar{\mathbf{A}} \circ \mathbf{A}) \boldsymbol{\sigma}, \quad (4)$$

where \circ denotes the Khatri-Rao or column-wise Kronecker product of two matrices. Since the source structure (the image $\boldsymbol{\sigma}$) is constant over time and frequency in many cases in radio astronomy, K short term integrations $\mathbf{r}_1, \dots, \mathbf{r}_K$ can be combined into a single measurement for the purpose of image reconstruction to give

$$\begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_K \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{A}}_1 \circ \mathbf{A}_1 \\ \vdots \\ \bar{\mathbf{A}}_K \circ \mathbf{A}_K \end{bmatrix} \boldsymbol{\sigma} = \mathbf{M} \boldsymbol{\sigma}. \quad (5)$$

In signal processing terms, imaging is the reconstruction of $\boldsymbol{\sigma}$ from \mathbf{r} , which essentially amounts to finding a left inverse of \mathbf{M} . Usually, this is done in two steps. The first is computation of the dirty image $\boldsymbol{\sigma}_d = \mathbf{M}^H \mathbf{r}$. The second is a deconvolution stage needed to obtain $\boldsymbol{\sigma}$ from $\boldsymbol{\sigma}_d$ [8, 7].

3. IMAGE RECONSTRUCTION CHALLENGES

Substitution of (2) in the above formulation for the dirty image quickly reveals a Fourier transform relationship between

the entries of the covariance matrix and the source fluxes (or pixel values). From a computational perspective, it is attractive to use the fast Fourier transform instead of the discrete Fourier transform. Unfortunately, typical array configurations are irregular, causing the visibilities to lie on an irregular grid. This is usually solved by gridding the measured visibilities. This is done by convolution with a gridding function. The size of this gridding function in terms of grid spacings, the kernel size N_{kernel} , determines the number of grid points each measured visibility contributes to. A larger kernel size allows for more elaborate spatial filtering and direction dependent gain compensation and leads to smaller errors in the image [9], but is computationally more demanding.

This procedure causes two problems that pose challenges for high dynamic range imaging:

- Images are produced on a grid of pixels, but sources do not naturally lie on grid points. This causes the dynamic range of gridded images to be limited to about 40 dB in practical scenarios [10, 11].
- Even after many hours of integration, visibility space may still not be fully sampled. As a result, the matrix \mathbf{M} in (5) may not have full column-rank.
- The true image will be convolved with the Fourier transform of the sampling function, which astronomers refer to as the *point spread function* (psf). The side-lobes of the psf of bright sources may be sufficiently high to drown weaker sources in the image, thus limiting the dynamic range of the image.

These issues are usually dealt with by modeling and subtraction of sources from the data. Traditionally, this is done by the CLEAN-algorithm originally proposed in [12]. This algorithm searches for the highest peak in the dirty image, estimates the exact source position and source flux of that source and subtracts a pre-specified fraction of the modeled source from the data. This cycle is repeated until the image is noise-like. This approach is related to sparse reconstruction using a greedy algorithm with relaxation. This technique produces very nice results in thermal noise limited scenarios, in which most pixels in the image contain only thermal noise. The increased sensitivity of LOFAR and SKA will push the thermal noise level to much lower values, thus revealing far more sources. This will likely make this approach prohibitively expensive.

Traditional telescopes consist of dishes spread over a relatively small geographical area (few kilometers) and provide a small field-of-view for observations. As a result, the space of visibility samples as well as the image can be approximated by a plane. The SKA and LOFAR (will) provide a much larger field-of-view and have a much larger geographical extent. This leads to another challenge:

- Current algorithms assume that the space of visibility samples and the image have a planar geometry, i.e., they do neither naturally handle the inherent spherical

geometry of large arrays on the Earth's surface nor the inherent spherical geometry of a large field-of-view on the celestial sphere.

The traditional solution to this problem is the W-projection algorithm [13], which projects the 3-D space of visibilities on a stack of "W-planes", effectively producing a 3-D volume of visibility samples that is Fourier transformed to a 2-D image. This projection can be incorporated in the gridding stage of the imaging process, but, as we will see in the next section, the number of required W-planes increases sharply with the size of the array and the size of the field-of-view. As a result, simply extending the current approach to SKA will again lead to an unacceptable compute load.

4. COMPUTATIONAL COSTS

The data model described by (3) assumes that the channel width Δf_{ch} is sufficiently small for the narrowband assumption to hold and that the integration time τ is sufficiently small to assume that the source and array geometry are fixed. If these assumptions break, averaging over Δf_{ch} and τ will cause decorrelation (loss of signal strength), especially at the highest observed spatial frequencies. In [14], it was found that $\Delta f_{\text{ch}} = \frac{1}{10} D_{\text{rec}} f / B_{\text{max}}$ and $\tau = 1200 D_{\text{rec}} / B_{\text{max}}$ are required to limit the worst case signal loss to 2%, where D_{rec} is the diameter of the station or dish, B_{max} is the longest baseline between two stations or dishes, which defines the size of the synthesized aperture, and f is the observing frequency.

Assuming Nyquist sampled time series from the individual receptors, the correlator input data rate $B_{\text{cor,in}}$ follows directly from the number of receptors P , the number of bits per sample N_{bit} and the total signal bandwidth W (may be divided in multiple frequency bands) as [15, 16]

$$B_{\text{cor,in}} = 4PN_{\text{bit}}W. \quad (6)$$

The factor four results from the fact that we need $2W$ real valued samples to satisfy the Nyquist criterion and from the assumption that each station or dishes measures both polarizations of the incoming radio waves.

The correlator processing power, expressed in operations per second, is simply

$$P_{\text{cor}} = 4(2P)^2 W, \quad (7)$$

where the factor 2 stems from the 2 polarizations and the factor 4 is caused by the fact that multiplication of two complex values requires four real valued multiply or add operations.

The correlator output data rate is given by

$$B_{\text{cor,out}} = N_{\text{bit,out}} \frac{W}{\Delta f_{\text{ch}}} (2P)^2 / \tau, \quad (8)$$

where $N_{\text{bit,out}}$ is the sample size of the real valued output samples, e.g., 32 bit for single precision floating point format.

The input data rate to the imager is the same as the output data rate of the correlator, i.e., $B_{\text{im,in}} = B_{\text{cor,out}}$. The computational power required for forming a dirty image using W-projection assuming near real-time processing follows from [13]

$$P_{\text{im}} = N_{\text{op}} \frac{W}{\Delta f_{\text{ch}}} (2P)^2 \frac{1}{\tau} (R_{\text{f}}^2 + N_{\text{kernel}}^2), \quad (9)$$

where N_{op} is the number of operations required to put one visibility on one grid point, N_{kernel} is the size of the gridding kernel and $R_{\text{f}} = \lambda_{\text{max}} B_{\text{max}} / D_{\text{rec}}^2$ is the Fresnel number with λ_{max} the longest observing wavelength. The Fresnel number is a measure of the severity of the non-coplanarity of the array due to the curvature of the Earth and the non-coplanarity of the image due to the field-of-view of the receptors. The output data rate of the imager depends on the length of each observation and the number of frequency slices, but is assumed to be negligible compared to its input data rate.

5. COMPUTING ARCHITECTURES

To define a suitable computer architecture for correlation and imaging, compute intensity is a key metric. Compute intensity is defined as the ratio of the amount of processing and the amount of I/O and is usually expressed in units of flops per byte. Usually, a high compute intensity implies a good cache efficiency, while a low compute intensity requires a design in which the data can be streamed through the computing hardware. The compute intensity for the correlator is

$$I_{\text{cor}} = \frac{P_{\text{cor}}}{B_{\text{cor,in}} + B_{\text{cor,out}}} = \frac{32P\Delta f_{\text{ch}}\tau}{N_{\text{bit}}\Delta f_{\text{ch}}\tau + N_{\text{bit,out}}P} \quad (10)$$

The compute intensity for imaging is

$$I_{\text{im}} = \frac{P_{\text{im}}}{B_{\text{cor,out}}} = \frac{8N_{\text{op}}}{N_{\text{bit,out}}} (R_{\text{f}}^2 + N_{\text{kernel}}^2). \quad (11)$$

To get some typical figures for the SKA, we computed the I/O and processing requirements for the envisaged SKA-dish array [2]. The relevant specifications for the L-band receiving system of this array are summarized in Table 1. The corresponding values for the Dutch LOFAR low-band (10 – 90 MHz) system are added for comparison. The resulting computing requirements are summarized in Table 2.

Correlation algorithms can be handled with racks of computing boards. The choice of computing technology depends mainly on the cost of design and the cost of operation. GPUs and FPGAs have a much higher power consumption than ASICs, while ASICs have a higher design cost. Moreover boards based on GPUs and FPGAs need more powerful and energy consuming cooling systems than ASICs. The differences between FPGAs and GPUs are not crucial. Both technologies are compatible with the computational load needed to perform the correlation algorithms; they can be hosted

	SKA-dish	LOFAR
P	254	40
D_{rec} (m)	15	85
B_{max} (km)	200	120
λ_{max} (m)	0.33	30
R_f	293.3	498.3
W (MHz)	500	96
Δf_{ch} (kHz)	3.9	0.763
τ (s)	0.08	1.0
N_{bit}	8	16
$N_{\text{bit,out}}$	32	32
N_{op}	10	10
N_{kernel}	7	7

Table 1. System parameters for SKA-dish and LOFAR

	SKA-dish	LOFAR
$B_{\text{cor,in}}$ (Tbps)	4.1	0.25
$B_{\text{cor,out}}$ (Tbps)	13.2	0.026
P_{cor} (Tflops)	516	2.46
I_{cor} (flops / byte)	239	72.4
P_{im} (Pflops)	356	2.0
I_{im} (kflops / byte)	215	621

Table 2. I/O rates, processing requirements and compute intensities for SKA-dish and LOFAR

in scalable architectures, so there is not a feasibility issue. GPUs are easier to program than FPGAs and there are many boards available off the shelf. FPGAs have a better power consumption. Even in this case, time and costs are the main drivers, but the choice will depend on future technologies.

With a P_{im} of 356 Pflops imaging is computationally intensive and there is currently no High Performance Computer (HPC) powerful enough to perform it in real-time. We thus have to wait for more powerful HPC systems that will be available in a few years. However, we can do work on system design. A good design could start from a scalable architecture made by a computing network that distributes data to different HPCs at 13.2 Tbps. This design is based on the well known Data Parallel skeleton. Each HPC performs imaging on a set of frequencies. If P_{HPC} is the computational power of a single HPC, the number of HPCs N_{HPC} is equal to $\alpha P_{\text{im}}/P_{\text{HPC}}$ where α is larger than one and depends on the efficiency of HPC and the efficiency of the algorithm parallelization. Inside each HPC many parallel patterns can be used. A design based on the data parallel skeleton allows to build large, scalable and easily upgradable systems.

6. NEW RESEARCH DIRECTIONS

In Sec. 3 a number of limitations of current algorithms were identified. In this section we discuss how these issues opens

new research directions.

The dynamic range limitation due to gridding of the image implies that bright sources with $\text{SNR} > 40$ dB need individual modeling and subtraction. This can involve even more sources, say all sources with $\text{SNR} > 30$ dB, in case of a poor psf. Given the success of the CLEAN algorithm and the fact that only a small fraction of the pixels will contain a bright source, sparse reconstruction techniques appear to be a natural choice to handle the bright sources. Several sparse reconstruction methods have therefore been considered, including use of convex relaxation as for example in the recently proposed Sparsity Averaging Reweighted Analysis (SARA) approach [17].

Since such techniques are computationally intensive and the image will become increasingly less sparse when approaching its thermal noise, techniques based on the sparsity assumption should only be used to remove the bright sources after which an image of all the remaining weaker sources can be produced using the computationally more efficient W-projection algorithm. This implies that the bright sources (several thousand) need to be estimated in the presence of noise and a sea of millions of weak sources. This provides a natural criterion to assess the performance of the proposed sparse reconstruction techniques for the radio astronomical performance: how robust is the estimation of bright sources to the presence of the sea of weak sources? This can be measured quantitatively by the level of residual structures left after subtraction of the bright sources from the visibility data.

For arrays like LOFAR and SKA the computational cost of W-projection is mainly driven by the R_f^2 -term in (9). Recently, the W-snapshots algorithm was proposed reducing the R_f^2 -scaling to $R_f^{2/3}$ -scaling [18]. However, this algorithm still projects the inherent spherical geometry of the array on a stack of planes. It would therefore be interesting to explore signal processing techniques that exploit the spherical geometry of the array and the spherical geometry of images with a large field of view. Such techniques have been considered [19, 20], but need to become more mature to handle a demanding application like the SKA.

7. CONCLUSIONS

We have formulated the radio astronomical image reconstruction problem in signal processing terms and discussed the currently used approach. This approach has dynamic range limitations due to the introduction of a grid of pixels and incomplete sampling, it does not exploit the spherical geometry inherent to the problem and its computational costs do not scale well to future telescopes. We have suggested to use sparse reconstruction techniques to remove the brightest sources using the level of residual source structures after subtraction to compare different algorithms and to focus more on algorithms exploiting the spherical geometry of the problem.

8. REFERENCES

- [1] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio, "The Square Kilometre Array," *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1482–1496, Aug. 2009.
- [2] P. E. Dewdney et al., "SKA1 System Baseline Design," Tech. Rep. SKA-TEL-SKO-DD-001, SKA Project Office, Manchester (UK), 12 Mar. 2013.
- [3] M. P. van Haarlem et al., "LOFAR: The LOw-Frequency ARray," *Astronomy & Astrophysics*, vol. 556, no. A2, pp. 1–53, Aug. 2013.
- [4] S. J. Wijnholds, S. van der Tol, R. Nijboer, and A.-J. van der Veen, "Calibration Challenges for Future Radio Telescopes," *IEEE Signal Processing Magazine*, vol. 27, no. 1, pp. 30–42, Jan. 2010.
- [5] A. J. Boonstra, *Radio Frequency Interference Mitigation in Radio Astronomy*, Ph.D. thesis, Delft University of Technology, Delft (The Netherlands), 14 June 2005.
- [6] A. R. Offringa, *Algorithms for Radio Interference Detection and Removal*, Ph.D. thesis, University of Groningen, Groningen (The Netherlands), 22 June 2012.
- [7] A.-J. van der Veen and S. J. Wijnholds, "Signal Processing Tools for Radio Astronomy," in *Handbook of Signal Processing Systems*, S. S. Bhattacharyya et al., Eds., pp. 421–464. Springer, second edition, 2013.
- [8] R. Levanda and A. Leshem, "Synthetic Aperture Radio Telescopes," *IEEE Signal Processing Magazine*, vol. 27, no. 1, pp. 14–29, Jan. 2010.
- [9] A. J. W. Duijndam and M. A. Schonewille, "Nonuniform fast Fourier transform," *Geophysics*, vol. 64, no. 2, pp. 539–551, Mar.-Apr. 1999.
- [10] W. D. Cotton and J. M. Uson, "Pixelization and Dynamic Range in Radio Interferometry," *Astronomy & Astrophysics*, vol. 490, no. 1, pp. 455–460, Oct. 2008.
- [11] S. Yatawatta, "Fundamental limitations of pixel based image deconvolution in radio astronomy," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jerusalem (Israel), 4-7 Oct. 2010, pp. 69–72.
- [12] J. A. Högbom, "Aperture Synthesis with a Non-regular Distribution of Interferometer Baselines," *Astronomy & Astrophysics Supplement*, vol. 15, pp. 417–426, 1974.
- [13] T. J. Cornwell, K. Golap, and S. Bhatnagar, "The Non-coplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 647–657, Oct. 2008.
- [14] P. Alexander, A. J. Faulkner, K. Grainge, and J. Lazio, "Software and Computing CoDR - Analysis of Requirements Derived from the DRM," Tech. Rep. WP2-050.020.010-RR-001, SKA Project Office, Manchester (UK), 17 Jan 2012.
- [15] S. J. Wijnholds and R. Jongerius, "Computing Cost of Sensitivity and Survey Speed for Aperture Array and Phased Array Feed Systems," in *IEEE AfriCon, URSI BEJ meeting*, Pointe Aux Piments (Mauritius), 9-12 Sept. 2013.
- [16] R. Jongerius, S. Wijnholds, R. Nijboer, and H. Corporaal, "End-to-end compute model of the Square Kilometre Array," *IEEE Computer*, 2014, accepted for publication.
- [17] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 21 Oct. 2012.
- [18] T. J. Cornwell, M. A. Voronkov, and B. Humphreys, "Wide field imaging for the Square Kilometre Array," in *Image Reconstruction for Incomplete Data VII*, Oct. 2012, vol. 8500 of *SPIE Proceedings*.
- [19] R. A. Kennedy, P. Sadeghi, and Z. Khalid, "Optimal signal processing on the 2-sphere: A general operator approach to signal recovery," in *Biomedical and Astronomical Signal Processing (BASP) Frontiers Workshop*, Villars-sur-Ollon (Switzerland), 27 Jan.- 1 Feb. 2013.
- [20] B. Leistedt and J. D. McEwen, "Flaglets: Exact Wavelets on the Ball," in *Biomedical and Astronomical Signal Processing (BASP) Frontiers Workshop*, Villars-sur-Ollon (Switzerland), 27 Jan.- 1 Feb. 2013.