CAMBRIDGE
UNIVERSITY PRESS

Research Article

Reduced-resolution beamforming: Lowering the computational cost for pulsar and technosignature surveys

D.C. Price[1,2]

[1]International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia and [2]SKA Observatory, Science Operations Centre, Kensington, WA 6151, Australia

Abstract

In radio astronomy, the science output of a telescope is often limited by computational resources. This is especially true for transient and technosignature surveys that need to search high-resolution data across a large parameter space. The tremendous data volumes produced by modern radio array telescopes exacerbate these processing challenges. Here, we introduce a 'reduced-resolution' beamforming approach to alleviate downstream processing requirements. Our approach, based on post-correlation beamforming, allows sensitivity to be traded against the number of beams needed to cover a given survey area. Using the MeerKAT and Murchison Widefield Array telescopes as examples, we show that survey speed can be vastly increased, and downstream signal processing requirements vastly decreased, if a moderate sacrifice to sensitivity is allowed. We show the reduced-resolution beamforming technique is intimately related to standard techniques used in synthesis imaging. We suggest that reduced-resolution beamforming should be considered to ease data processing challenges in current and planned searches; further, reduced-resolution beamforming may provide a path to computationally expensive search strategies previously considered infeasible.

Keywords: Radio telescopes; astronomical techniques; pulsars; technosignatures; search for extraterrestrial intelligence (SETI)

1. Introduction

Extracting science results from astronomy datasets can often be a computationally demanding process. Supercomputers have become vital tools in modern astronomy – unfortunately to the point that energy consumption of supercomputers and research infrastructure dominate astronomy's carbon footprint (Stevens et al. 2020; Portegies Zwart 2020). As well as the ecological impact, access to computing resources is limited by the technology and funding available. Observational astronomers must work within these boundaries to unravel the scientific mysteries and marvels hiding in their data.

In radio astronomy, one of the most computationally expensive exercises is searching for pulsars and other fast transients. In order to detect a new pulsar with unknown characteristics, one must search across a range of different periods, dedispersion trials, and pulse widths (see Chapter 6, Lorimer & Kramer 2004). Searches are particularly challenging when using radio array telescopes, as a large number of beams must be formed to cover a survey area, and the search process must be repeated on each beam. Nevertheless, there are active pulsar searches on several radio array telescopes (e.g. Sanidas et al. 2019; Chen et al. 2021; Singh et al. 2023; Bhat et al. 2023a). fast radio burst (FRB) searches (e.g. Law et al. 2015; Bailes et al. 2017; Bannister et al. 2017; Ng et al. 2017) must also search through multiple dispersion and

pulse width trials but are (somewhat) more manageable as they do not apply a periodicity search.

Another computationally demanding avenue is the search for extraterrestrial intelligence (SETI), which seeks to detect 'technosignatures' from extraterrestrial intelligence as a proxy for intelligent life. Indeed, the SETI@Home project (Anderson et al. 2002) was a pioneer in developing distributed computing in order to attain sufficient compute resources for data analysis. At one stage, the SETI@Home network – peaking at over 5.2 m volunteers – constituted the largest supercomputer on the planet.

Decades later, thanks to exponential 'Moore's law' scaling of compute capacity, the Breakthrough Listen project (Worden et al. 2017; Isaacson et al. 2017) is able to search petabytes of data for technosignatures with modest server clusters located at the observatory (MacMahon et al. 2018; Price et al. 2018; Lebofsky et al. 2019), making the distributed computing approach unnecessary. Even so, technosignature searches using radio arrays, such as the Breakthrough Listen programme on the MeerKAT telescope (Czech et al. 2021) and the recently announced COSMIC[a] instrument on the Karl G. Jansky Very Large Array, are orders of magnitude more challenging than previous single-dish searches, motivating new search techniques.

With a radio array, when a desired search strategy is limited by the available compute resources, it is often advantageous to sacrifice sensitivity and imaging fidelity against instantaneous field of view. As beam width scales with $\sim \lambda/D$, where D is the longest

[a]COSMIC: Commensal Open-Source Multimode Interferometer Cluster Search for Extraterrestrial Intelligence. https://science.nrao.edu/facilities/vla/observing/cosmic-seti.

baseline in the array and λ is wavelength, the number of beams required to cover a given survey area scales with D^2. As such, pulsar searches tend to only use the 'core' part of an antenna array, and longer baselines are dropped. For example, in the MeerKAT TRAPUM[b] survey, only the central 37–41 antennas from the 64-antenna array are regularly used, corresponding to 57–64% of the peak sensitivity (Chen et al. 2021). The LOTASS[c] survey used the dense central 'Superterp' core of LOFAR to balance field of view against sensitivity. Similarly, the Murchison Widefield Array (MWA) SMART[d] (Bhat et al. 2023a) uses the 'compact' MWA configuration of 128 tiles within ∼800 m.

The computational challenges for pulsar surveys are exacerbated at low frequencies, as the number of dispersion measrure (DM) trials required to correct for dispersive delays due to propagation through the interstellar medium scales with λ^2. The LOTAAS survey searched across 10 120 DM trials, which required 30M CPU-core hours on the Cartesius supercomputer to process data up to January 2019 (Sanidas et al. 2019). Based on first-pass benchmarks, the SMART survey will require ∼60M CPU-core hours to process ∼93 hours of observational data using the OzStar supercomputer (Bhat et al. 2023a). As the SMART team anticipate receiving ∼0.5–0.6M CPU-core hours per semester, they are in the process of optimising and accelerating their search pipeline to ensure processing is possible.

In this paper, we introduce a 'reduced-resolution' beamforming technique that allows for sensitivity to be traded off against sky coverage. Our reduced-resolution beamforming approach, applied after inter-antenna correlation, down-weights baselines longer than a specified distance. This per-baseline approach results in appreciably larger sensitivity than simply discarding all antennas outside of a core region. Reduced-resolution beamforming effectively covers a desired survey area with a smaller number of beams, for a moderate decrease in survey speed. By doing so, the downstream compute requirements for a given survey can be greatly reduced.

This paper is organised as follows. In Section 2, we give an overview of post-correlation beamforming, introduce a compact tensor notation, and highlight the relationship between a power beam and a pixel within a synthesis image. Reduced-resolution beamforming is introduced in Section 3, along with survey speed and computational cost metrics. In Section 4, we consider the application of reduced-resolution beamforming to the SMART survey; in Section 5, we argue that the survey speed for MeerKAT pulsar searches could be increased dramatically by reduced-resolution beamforming, without increasing the number of beams formed. We conclude with a short discussion in Section 6.

2. Post-correlation beamforming

Tied-array (coherent) beamforming is a common signal processing operation used to combine the voltage-level outputs from an array of N radio telescopes. A tied-array of N identical telescopes increases sensitivity by a factor of N over a single telescope, but this comes at the expense of a smaller field of view. Multiple tied-array beams can be formed to improve the instantaneous field of view of the array; however, storing and processing multiple beams can be challenging.

---

[b]TRAPUM: Transients and Pulsars with MeerKAT.
[c]LOTASS: LOFAR Tied-Array All-sky Survey.
[d]SMART: Southern-sky Rapid Two-Metre pulsar survey survey.

An alternative approach, known as *incoherent beamforming*, sums the output power from each telescope. An incoherent beam retains the field of view, but the sensitivity only improves as $\sqrt{N}$. For computationally expensive algorithms – such as those required in pulsar, FRB, and technosignature searches – it is often infeasible to form tied-array beams across the instrument's entire field of view. On the other hand, the incoherent beam may not have enough sensitivity nor spatial resolution to be scientifically interesting.

A beamformed *voltage beam* from P antennas is given by the sum of weights $w(t) = Ae^{i\theta(t)}$ with a voltage stream $v(t)$:

$$b(t) = \sum_{p=1}^{P} w_p(t) v_p(t) \qquad (1)$$

The *power beam*, B, is the beam after squaring and averaging in time. Denoting time average with angled brackets ⟨⟩, and treating the weights as static across the time average interval, that is, $w_p(t) = w_p$, the power beam is given by:

$$B = \langle b(t) b^*(t) \rangle \qquad (2)$$

$$= \left\langle \left(\sum_{p=1}^{P} w_p v_p(t) \right) \left(\sum_{q=1}^{P} w_q v_q(t) \right)^* \right\rangle \qquad (3)$$

$$= \sum_{p=1}^{P} \sum_{q=1}^{P} w_p \langle v_p(t) v_q^*(t) \rangle w_q^* \qquad (4)$$

The quantity $\langle v_p(t) v_q^*(t) \rangle$ can be treated as a $(P \times P)$ matrix **V**. This is known as the visibility matrix and is the fundamental data product produced by a radio interferometer. Equation (4) can alternatively be written in terms of **V** and a weight vector **w**:

$$B = \mathbf{w} \, \mathbf{V} \, \mathbf{w}^H \qquad (5)$$

where $\mathbf{w}^H$ is the weight vector's Hermitian conjugate.

In the above, we treated weights as being static; in practice, weights must be updated as a source moves across the sky. The update cadence (i.e. maximum time average interval) required to avoid decorrelation loss depends on the longest inter-antenna baseline within the array; see Wijnholds et al. (2018) for further discussion.

There are therefore two paths towards generating a power beam, which we refer to as *pre-correlation* beamforming and *post-correlation* beamforming. These paths are shown diagramatically in Fig. 1. The figure also shows optional gridding steps, which allows multiple regularly spaced beams to be formed using a two-dimensional fast Fourier transform (2D FFT) if the input data are regularly gridded. The dashed lines in Fig. 1 represent hybrid architectures, for example, arrays like the MWA and LOFAR where beamforming is performed on a subset of antennas within a 'station', and then the station tied-array beams are correlated.

2.1 Direct imaging approaches

With the advent of large-N telescopes, there is increased interest in 'direct imaging' instruments which follow the pre-correlation gridded path of Fig. 1. In one approach, antennas are physically arranged on a grid, and then a spatial 2D FFT is applied to form images (Tegmark & Zaldarriaga 2009). This approach is extended by the Modular Optimal Frequency Fourier (MOFF) imaging technique. In the MOFF approach, antennas do not need to be physically placed on a grid; rather, voltages are gridded electronically based on the aperture illumation of consituent antennas
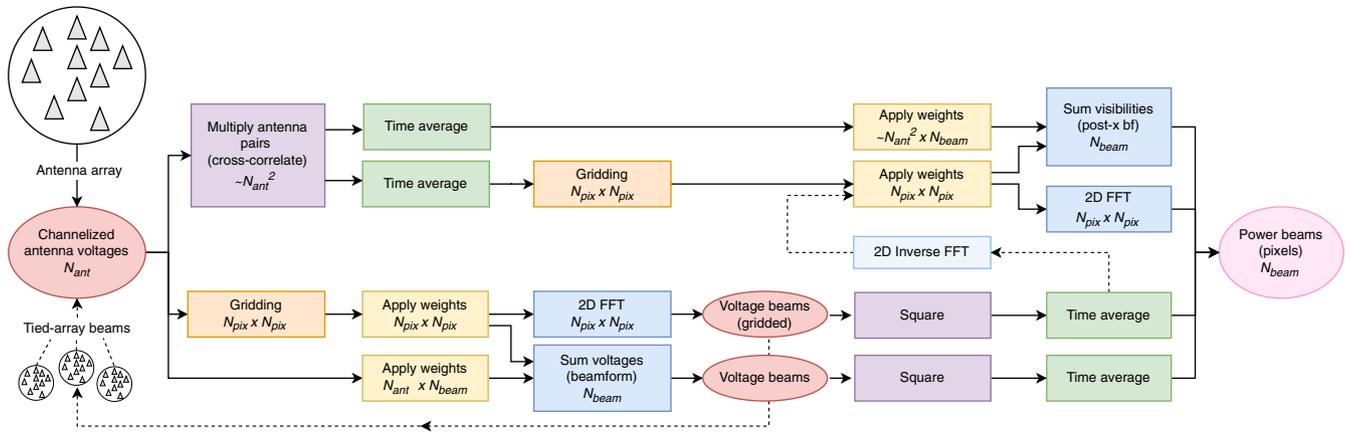
Figure 1. Block diagram showing (simplified) pathways to generate power beams – or equally, images – from antenna voltages (noting $N_{\text{beam}} \equiv N_{\text{pix}} \times N_{\text{pix}}$). Equivalent tasks are colour-coded and italic text corresponds to data output dimensions; dashed lines represent hybrid architectures. *Standard imaging* follows the top path (i.e. correlating antenna pairs first): correlations are time-averaged, gridded, weighted, and then a 2D FFT is applied to form the image. *Post-correlation beamforming* also follows this top path but sums visibilities without a gridding step to form a power beam (or multiple beams). *Standard tied-array beamforming* follows the bottom path (i.e. applying weights first): weights are applied to antenna voltages, which are summed to form a voltage beam (or multiple beams). This voltage beam can be squared and time-averaged to create a power beam. *Direct imaging correlators* follow the bottom path too but apply gridding at the start, so a 2D FFT can be used to form a grid of power beams.

(Morales [2011]; Thyagarajan et al. [2017]; Kent et al. [2019]). Note that a visibility matrix can be retrieved from a grid of beams via an inverse 2D FFT (Tegmark & Zaldarriaga [2009]; Foster et al. [2014]), which is shown in Fig. 1 as a dashed line between pathways.

Here, we focus on non-gridded methods, but note that equivalent trade-offs between resolution and sensitivity could be leveraged with these alternative architectures.

2.2 Tensor formalism

Equation (5) can be written in an equivalent compact form using summation notation:

$$B = \mathbf{w}\,\mathbf{V}\,\mathbf{w}^H \equiv \mathbf{w}^p\,\mathbf{V}_{pq}\,(\mathbf{w}^*)^q = \mathbf{W}^{pq}\mathbf{V}_{pq} \qquad (6)$$

In this notation, summation is implied over all indices that appear in an upper and lower index (* represents complex conjugation). So, we sum across indexes p and q (summation indices representing antennas) which returns a single value B. The quantity $\mathbf{W}_{pq} = \mathbf{w}_p\mathbf{w}_q^*$ is equivalent to the ($P \times P$) matrix formed from the outer product $\mathbf{w}\mathbf{w}^H$.

The visibility matrix can itself be written in summation notation, instead of using $\langle \rangle$ brackets, as:

$$\mathbf{V}_{pq} = \mathbf{v}_p^t(\mathbf{v}^*)_{qt} \qquad (7)$$

where t represents time.

These bold-face quantities are often referred to as *data tensors*. Tensors can simply be considered as multi-dimensional arrays, or generalisations of a matrix into higher dimensions. Here, we will follow the nomenclature used in the TensorFlow software package.[e] As we are using data tensors as a computational tool, we do not need to worry about other uses of tensors found in physics and mathematics, such as their interpretation as mappings between vector spaces, so we do not define a tensor metric here. We note that tensors have indeed been used previously in calibration and imaging (e.g. Smirnov [2011]; Price & Smirnov [2015]; Thekkeppattu et al. [2024]).

Whenever a pair of tensors have matching upper and lower indices, entries are summed across all matching indices; this is known as a *pairwise contraction*. The order of pairwise contractions for a chain of tensors is known as the *contraction path*. The computational cost to evaluate a tensor chain depends on the contraction path; finding the optimal path is an NP-hard problem that can quickly become intractable as the number of tensors increases.

2.3 Computing tensor contractions

Numerous software packages for performing tensor contractions are available. We highlight that the Python Numpy package (Harris et al. [2020]) includes a flexible 'one-liner' `einsum` method for computing contractions on Numpy arrays. An equivalent `einsum` method exists in the Cupy software package (Okuta et al. [2017]), which offloads computations to a graphics processor unit (GPU). If compiled against the NVIDIA cuTENSOR library,[f] highly optimised tensor cores on the GPU may be targeted; these offer orders-of-magnitude greater energy efficiency than regular GPU cores.

A highly optimised correlation code that uses tensor cores is detailed in Romein ([2021]), and a complex general matrix multiply (cGEMM) code has been developed for use for beamforming with a phased array feed.[g] Together, these two codes could be used to compute Equation (6), and the `einsum` method provides a straightforward path for prototyping and developing post-correlation beamforming techniques.

2.4 Multiple beams and frequency channels

We can extend Equation (6) with extra indices (subscripts l and m) to represent a grid of ($L \times M$) beams on the sky and across F frequency channels (subscript ν):

$$\mathbf{B}_{lm\nu} = \left(\mathbf{w}_{lmp\nu}\mathbf{v}_\nu^{pt}\right)\left(\mathbf{w}_{lmq\nu}^*(\mathbf{v}^*)_{\nu t}^q\right) \qquad (8)$$

That is, we sum across indexes p and q (summation indices) and output an N-dimensional array with indices (l, m, v). We can also write Equation (6) as:

$$\mathbf{B}_{lmv} = \mathbf{W}^{pq}_{lmv} \, \mathbf{V}_{pqv}. \tag{9}$$

If power beams are evaluated at a series of timesteps, we may write $\mathbf{B}_{lmv} = \mathbf{B}(t)_{lmv}$ as a function of time:

$$\mathbf{B}(t)_{lmv} = \mathbf{W}^{pq}_{lmv} \, \mathbf{V}(t)_{pqv}. \tag{10}$$

There are two key differences between Equations (8) and (11). The first is that the tensor $\mathbf{W}_{pqlmv}$ has ($P^2 \times L \times M \times F$) entries, but $\mathbf{w}_{plmv}$ has only ($P \times L \times M \times F$) entries. The second is that Equation (8) requires three pairwise contractions to compute, while Equation (11) is only one pairwise contraction – two contractions have already been performed to produce $\mathbf{W}_{lmpqv}$ and $\mathbf{V}_{pqv}$.

To extend to a polarisation-aware version, we may simply add a subscript x:

$$\mathbf{B}(t)_{lmvx} = \mathbf{W}^{pq}_{lmvx} \, \mathbf{V}(t)_{pqvx}. \tag{11}$$

where x represents a set of four polarisation coherency measurements (e.g. XX*, XY*, YX*, and YY* for a linearly polarised dual-polarisation antenna).

2.5 Comparison to interferometric imaging

In synthesis imaging, an image $I(l,m)$ is created by evaluating

$$I(l, m) = \frac{1}{M} \sum_{m=1}^{M} V(u_k, v_k) e^{2\pi(u_k l + v_k m)}, \tag{12}$$

where M is the number of baselines and (u_k, v_k) are coordinates relating to the baseline k between a (p, q) antenna pair (Equations 7–3, Briggs et al. 1999). This equation can written as a tensor contraction:

$$\mathbf{I}_{lm} = \mathbf{W}^k_{lm} \mathbf{V}_k \tag{13}$$

$$\mathbf{W}_{klm} = \frac{1}{M} exp\left(-i2\pi(u_k l + v_k m)\right), \tag{14}$$

which by splitting $k = pq$ can be rewritten as:

$$\mathbf{I}_{lm} = \mathbf{W}^{pq}_{lm} \mathbf{V}_{pq}, \tag{15}$$

$$\mathbf{W}_{pqlm} = \frac{1}{M} exp\left(2\pi i(u_{pq} l + v_{pq} m)\right). \tag{16}$$

Comparison of Equations (5) and (16) reveals that $\mathbf{I}_{pq} \equiv \mathbf{B}_{pq}$. That is, each pixel in an image is equivalent to a power beam, or alternatively, an image can be created out of a grid of power beams. Note that in synthesis imaging, autocorrelations are generally not included; that is, the weights tensor $\mathbf{W}_{pq} = 0$ if $p = q$. From a power beam interpretation, this is equivalent to subtracting the incoherently summed beam from the tied-array power beam (Roy et al. 2018).

2.6 Historical perspective

While the interpretation of image pixels as power beams is not commonly discussed, the link between imaging and beamforming is fundamental to radio astronomy and was leveraged by early interferometers, including the Mills Cross (Mills et al. 1958). However, for the most part, the historical development of pulsar search pipelines has run parallel to high-fidelity synthesis imaging systems, leading to disparate science communities and techniques.

Since their discovery in 1967 by Jocelyn Bell Burnell, a majority of pulsars have been discovered using single-dish instruments, such as the Parkes Murriyang, Arecibo, Lovell, and Green Bank telescopes (Manchester et al. 2005). As such, most observing techniques and knowledge within the pulsar community is derived from single-dish approaches. Pulsar search codes such as PRESTO[h] (Ransom 2011) were also designed for single-dish telescopes and are, in general, incompatible with interferometric data products.

In contrast, radio interferometers have predominantly been used for long-exposure synthesis imaging. Earth rotation synthesis, popularised by the Cambridge One-Mile Radio Telescope (Ryle 1962), combines data across many hours of observation to improve imaging fidelity. Instruments were not designed to output high time resolution data products; rather, they were designed to integrate data for as long as possible to minimise output data rates. Despite their lack of time resolution, pulsars have been detected in synthesis images via other characteristics; exceptionally, the first millisecond pulsar was detected in synthesis images as a source with an anomalous spectrum (Backer et al. 1982).

The quest to discover new pulsars has nonetheless led pulsar astronomers towards radio arrays. Modern interferometers have good sensitivity, a large field of view, and better localisation capability; on a technical level, tied-array beamforming is increasingly feasible. Fast imaging techniques – where each pixel is treated as a power beam – have been used for searches of millisecond transient pulses (Law et al. 2015, 2018), as have FFT-based beamforming approaches (Ng et al. 2017), but are yet to be eagerly adopted in pulsar periodicity searches.

Techosignature searches have followed a similar trajectory. Frank Drake performed the first SETI search using the Tatel single-dish telescope in 1961 (Drake 1961); single-dish telescopes remained the primary technology used until the construction of the Allen Telescope Array (Welch et al. 2009). Many technosignature experiments have targeted nearby stars and have sought to maximise frequency coverage (e.g. Isaacson et al. 2017); hence, field of view has not historically been a science driver. But more ambitious surveys of nearby star targets (Czech et al. 2021) motivate faster survey speed, and there is a growing movement towards non-targeted widefield technosignature searches (Houston et al. 2021).

3. Reduced-resolution beamforming

One tactic used to improve sky coverage of a tied-array beamformer is to only use antennas within a 'core' region. Via Equation (8), selecting a core region is equivalent to setting weights $\mathbf{w}_{lmpv}$ to zero for any antenna p that is above a cut-off distance from a reference antenna (i.e. *per-antenna* filtering). To first order, the full-width at half-maximum (FWHM) of the resultant tied-array beam will scale with λ/d_{max}.

Post-correlation beamforming allows the user to set weights between any antenna pair ($\mathbf{W}_{pq}$), allowing *per-baseline* filtering (i.e. in fact, the standard approach in synthesis imaging). The per-baseline approach means that all baselines below a given threshold distance can be included in the power beam – including short baselines outside the core. It follows that a per-baseline filtering approach (only possible with post-correlation beamforming) will always yield a more sensitive power beam than a per-antenna approach.

[h]https://github.com/scottransom/presto.
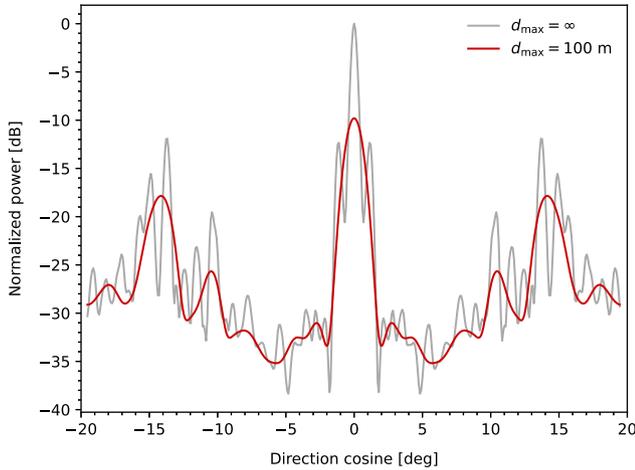
Figure 2. Reduced-resolution beam formed from the MWA compact configuration. The full-resolution beam is shown in grey; a reduced-resolution beam for a 100-m maximum baseline is shown in red.

We refer to post-correlation beamforming schemes that set weights to zero for any baseline above a cut-off distance as *reduced-resolution beamforming*. Controlling the cut-off distance essentially allows for sensitivity to be traded off against beam width. For simplicity, we will only consider weights with magnitude 0 (discard baseline) or 1 (keep baseline), but note there are myriad weighting schemes used in synthesis imaging that could be considered (see Chapter 10, Thompson et al. 2017).

Fig. 2 shows an example of reduced-resolution beamforming, as applied to MWA 128-tile Compact configuration (Wayth et al. 2018). The grey line shows the beam pattern (known as point spread function, or dirty beam, in synthesis imaging) for a zenith-pointed tied-array beam. The red line shows the resulting beam pattern using reduced-resolution beamforming, with a $d_{max} = 100$ m cutoff. The reduced-resolution beam is wider and smoother but has a lower power gain, which corresponds to lower sensitivity. Note Fig. 2 does not account for the tile beam pattern.

3.1 Sensitivity

From the radiometer equation, the expected thermal noise (in Jy units) of a beam formed via post-correlation beamforming can be written as:

$$\Delta S_{post-x} = \frac{k_B T_{sys}}{A_e \sqrt{\Delta \nu \tau N_{pol} (2N_{baselines} + N_{ant})}} \quad (17)$$

where we introduce the following quantities:

T_{sys}	System temperature
k_B	Boltzmann's constant
$\Delta \nu$	Bandwidth
A_{ant}	Effective area of an antenna element
τ	Integration time
N_{ant}	Number of antennas
N_{pol}	Number of polarisations
$N_{baselines}$	Number of baselines in array

Equation (17) differs slightly from the standard interferometer radiometer equation (Equation 9.26, Wilson et al. 2013) as it includes a N_{ant} term to account for autocorrelations. Noting that

$$N_{baselines} = N_{ant}(N_{ant} - 1)/2 \quad (18)$$

$$N_{ant}^2 = 2N_{baselines} + N_{ant} \quad (19)$$

one finds that Equation (17) reduces to the tied-array radiometer equation if all baselines are included:

$$\Delta S_{tied} = \frac{2k_B T_{sys}}{A_{ant} N_{ant} \sqrt{\Delta \nu \tau N_{pol}}}. \quad (20)$$

Additionally, if cross-correlation terms are not included (i.e. $N_{baselines} = 0$), then we retrieve the radiometer equation for an incoherent beam:

$$\Delta S_{incoherent} = \frac{2k_B T_{sys}}{A_{ant} \sqrt{\Delta \nu \tau N_{ant} N_{pol}}}. \quad (21)$$

If we choose a subset of baselines, based on maximum length, the resulting noise will lie between the ideal ΔS_{tied} thermal noise (all baselines included) and the incoherent $\Delta S_{incoherent}$ noise (only autocorrelations included). Put another way, reduced-resolution beamforming provides access to resolution and sensitivity regimes that lie between that of incoherent and coherent beamforming techniques.

3.2 Fractional sensitivity

Consider the case where a maximum baseline length d is imposed on an array with longest baseline D. For a tied array beamformer, the sensitivity can be treated as a function of baseline length, $\Delta S_{coherent}(d)$ (Equation (20)), with the number of antennas in the sub-array is $N'_{ant}(d)$. For a post-correlation beamformer (Equation (17)), the thermal noise level reached $\Delta S_{post-x}(d)$ depends on $N'_{baselines}(d)$.

It is useful to define a fractional sensitivity factor, f_S, that relates the sensitivity of the selected subarray to the full array:

$$\Delta S(d) = \Delta S(D)/f_S. \quad (22)$$

For a reduced-resolution beam with all autocorrelations, $N'_{ant} = N_{ant}$, and

$$f_S = \frac{\Delta S_{post-x}(D)}{\Delta S_{post-x}(d)} = \frac{\sqrt{2N'_{baselines}(d) + N_{ant}}}{N_{ant}} \quad (23)$$

whereas for a tied-array beam

$$f_S = \frac{\Delta S_{tied}(D)}{\Delta S_{tied}(d)} = \frac{N'_{ant}(d)}{N_{ant}} \quad (24)$$

and $N'_{ant} \leq N_{ant}$.

Fig. 3 shows the fractional sensitivity for a reduced-resolution beam and tied-array beam as a function of maximum baseline length, using the EDA2[i] array as an example (Wayth et al. 2021). The reduced-resolution beam always has a larger fractional sensitivity than a tied-array beam from a sub-selection of antennas. This fact is a key motivation for reduced-resolution beamforming.

The slope of $\Delta S_{post-x}(d)$ depends on the antenna configuration. Fig. 4 considers the application of reduced-resolution beamforming to three radio arrays: the 64-antenna MeerKAT telescope

---

[i]EDA2: Engineering Development Array 2.
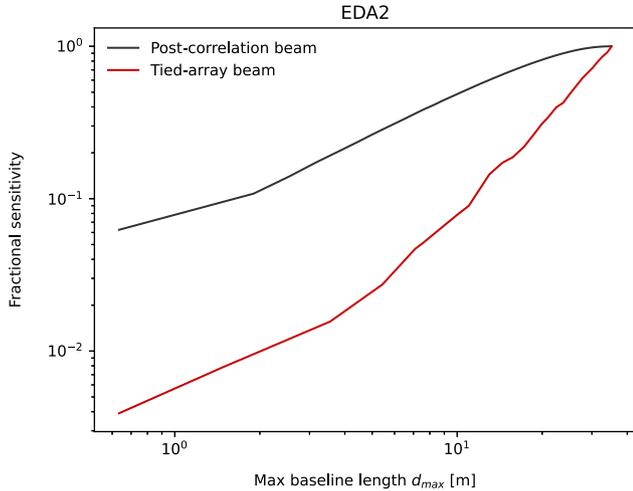
Figure 3. Comparison of EDA2 fractional sensitivity for a tied-array beam (red) and a post-x beamformed beam (black), as a function of maximum baseline length d_{max}. The post-x approach allows short baselines to be included, boosting sensitivity.

(Camilo 2018), the 128-tile MWA (in compact configuration, Wayth et al. 2018), and the EDA2 (Wayth et al. 2021). The top panels show the antenna layout for each array, and the middle panels show baseline distribution histograms as a function of baseline length. The lower panel shows the fractional sensitivity for each array as a function of maximum baseline length (solid black line), and the number of beams required to fill the primary field of view, assuming the beam width scales as λ/d (dashed red line).

A key takeaway from Fig. 4 is that all telescopes reach a reasonable fractional sensitivity ($\sim$0.5) with an order of magnitude fewer beams to fill their field of view than the full array. In the coming sections, we will explore how this finding could be applied on these telescopes.

3.3 Survey speed

Survey speed is a useful metric for optimising the performance of an array. The point-source survey speed figure of merit (PFoM) relates to the time taken to survey a field with solid angle Ω_{survey} to a required thermal noise level in Jy, $\Delta S_{\mathrm{survey}}$. Following Equation 3 of Cordes (2009),[j] we may define a PFoM modified in a similar fashion to Chen et al. (2021) to account for number of beams and maximum baseline length.

The fraction of the maximum PFoM (all antennas, beams covering instrumental FoV) reached with a reduced-resolution beamforming system is

$$f_{\mathrm{PFoM}} = \frac{\mathrm{PFoM}(d)}{\mathrm{PFoM}_{\mathrm{full}}} = f_S^2 \times f_{\mathrm{FoV}} \qquad (25)$$

where the fractional coverage of the instrument's field of view (Ω_{full}), depends on the number of beams formed:

$$f_{\mathrm{FoV}} = \frac{\Omega_{\mathrm{FoV}}(d)}{\Omega_{\mathrm{full}}} \approx \frac{N_{\mathrm{beam}}}{\Omega_{\mathrm{full}}} \left(\frac{\lambda}{d}\right)^2, \qquad (26)$$

assuming resolution follows the λ/d rule of thumb. However, the actual beam width depends on both the array configuration and the beam's declination in a complex fashion. Many radio arrays

---

have decreasing antenna density for longer baselines, for which λ/d underestimates the beam width. Nevertheless, if the number of beams is fixed, we expect the fractional field of view to decrease rapidly as d increases.

3.4 Computational cost

The approximate computational cost of a correlator and multi-pixel beamformer can be written in terms of complex-multiply accumulate operations per second (CMACs/s) as:

$$C_X = \big(N_{\mathrm{ant}}N_{\mathrm{pol}}(N_{\mathrm{ant}}N_{\mathrm{pol}} + 1)\big) \times \Delta\nu \qquad (27)$$

$$C_B = \big(N_{\mathrm{beam}} \, N_{\mathrm{ant}} \, N_{\mathrm{pol}}\big) \times \Delta\nu \qquad (28)$$

where $\Delta\nu$ is the bandwidth of data processed. To form post-correlation beams from the correlator output requires an additional step, with computational cost

$$C_{\mathrm{pxb}} = f_{\mathrm{baselines}} \times \left(N_{\mathrm{pxb}}N_{\mathrm{ant}}^2 N_{\mathrm{pol}}^2\right) \times \frac{\Delta\nu}{N_{\mathrm{int}}} \qquad (29)$$

where N_{int} is the number of time samples summed during integration, $f_{\mathrm{baselines}}$ is the fraction of baselines included, and N_{pxb} is the number of post-correlation beams. Note that the cost of FFT-based imaging scales proportionally to $N_{\mathrm{pxb}}\log_2 N_{\mathrm{pxb}}$, and for most arrays[k]

$$\log_2 N_{\mathrm{pxb}} \ll N_{\mathrm{ant}}^2, \qquad (30)$$

so we stress that the PXB approach is generally far more computationally expensive than the 2D FFT approach employed in imaging pipelines. That said, for fair comparison with computationally expensive searches – where only a few beams can be processed on any one compute server – we do not compare against FFT-based imaging any further in this article.

Based on Equations (28)–(29), tied-array beamforming is more computationally expensive if $N_{\mathrm{beams}} \lesssim N_{\mathrm{pxb}}$ and/or if the correlator does not integrate many time samples, that is, N_{int} is large.

This conclusion is consistent with Roy et al. (2018), which considers post-correlation beamforming techniques for pulsar studies with the Giant Metrewave Radio Telescope. Roy et al. (2018) also concludes that post-correlation beamforming is computationally cheaper for a large number of beams at low time resolution, and regular beamforming is cheaper for a small number of high time-resolution beams.

As with imaging, the cost of post-correlation beamforming (Equation (29)) can be reduced significantly if visibilities are gridded and a 2D FFT is used (Briggs et al. 1999). However, the FFT approach is not be suitable for a small number of beams and/or widefield imaging; further, when distributing processing tasks across a supercomputer, it may be advantageous to have only a small number of beams per node. As such, we present values for the non-gridded approach.

4. Application to MWA SMART

SMART (Bhat et al. 2023a) is a pulsar and fast transient search project on the MWA, using the 128-tile compact configuration (Wayth et al. 2018). SMART aims to conduct a pulsar search across the full sky below 30° declination, to a limiting sensitivity (10σ) of 2–3 mJy. Observations began in 2018 and are approximate 75%

---

[j]https://www.skatelescope.org/uploaded/26865_109_Memo_Cordes.pdf.

[k]Noting $N_{\mathrm{pxb}} \equiv N_{\mathrm{pix}} \times N_{\mathrm{pix}}$, where N_{pix} is image width in pixels.
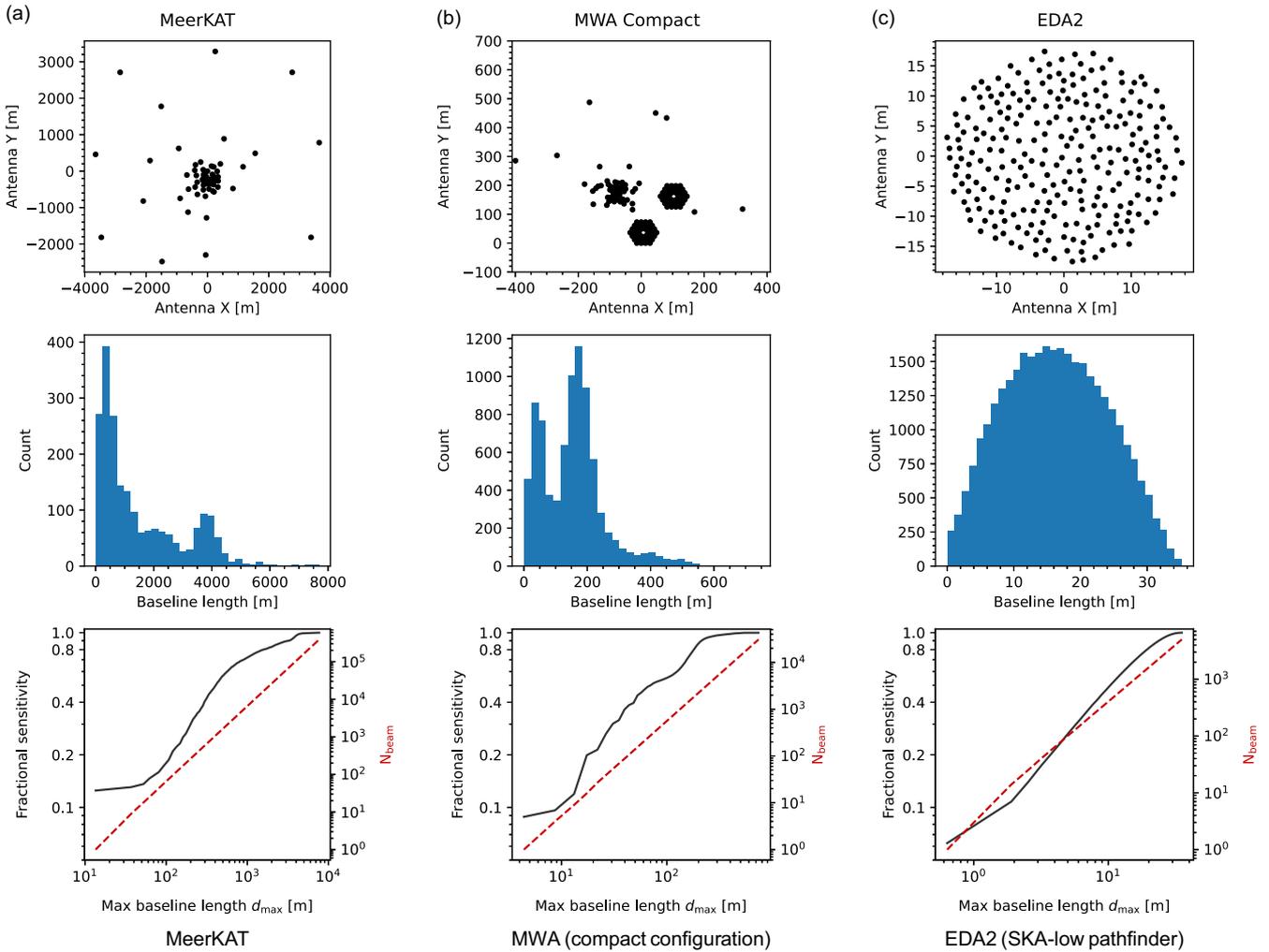
Figure 4. Sensitivity analysis for reduced resolution beamforming applied to MeerKAT, the MWA (compact configuration), and the EDA2. The top panels show antenna location, and the middle panels show histograms of baseline lengths for each telescope. The bottom panel shows the fractional sensitivity for reduced resolution beamforming, as a function of maximum baseline length (black lines). The minimum sensitivity is equivalent to incoherent beamforming, and maximum sensitivity is equivalent to standard beamforming. The number of beams required to cover each telescope's field of view scales with d_{max}^2 (dashed red lines).

complete (Bhat et al. 2023b). The SMART project stores voltage-level data products for each of 128 tiles within the MWA compact configuration, which allows beamforming and/or correlation to be performed offline. SMART consists of 70 pointings of 80-minute duration (93 hr total), which corresponds to ∼3 PB of data. These data are stored on the Pawsey Banksia object store.[l]

SMART uses a multi-pixel tied-array beamformer (Swainston et al. 2022), which is used to form up to 20 beams at once (per compute node). Each pointing requires ∼6 000 beams to cover the full field of view. The pulsar search is performed on power beam data (Stokes-I). On the OzSTAR supercomputer,[m] beamforming 10 minutes of data takes 2 kSU (service units), and postprocessing search tasks take 25 kSU (1 kSU is equivalent to 1 000 CPU-core hours, or 125 GPU-core hours). Extrapolating from these numbers, assuming processing scales linearly with observation time, processing all 70 pointings for the full 80 minutes

would require ∼1 120 kSU for beamforming, and ∼14 000 kSU for post-processing.

4.1 Trading sensitivity against compute requirements

The post-processing requirements for SMART scale proportionally with the number of beams searched. By applying reduced-resolution beamforming, the SMART survey volume can be searched with fewer beams, potentially improving computational requirements by an order of magnitude for a moderate drop in sensitivity. However, any improvement in downstream processing costs must be offset against any increases in beamforming costs. Using Equations (28)–(29) and the SMART survey parameters (Table 1), the computational costs are as follows:

$$C_X = 16.2 \text{ TCMAC/s} \qquad (31)$$

$$C_B = 767.6 \text{ TCMAC/s}. \qquad (32)$$

The comparable computational cost for reduced-resolution beamforming (C_{PXB}) as a function of fractional sensitivity f_s is shown in

Table 1. MWA SMART survey parameters for a single zenith pointing.

N_{ant}	128	Number of antennas
N_{pol}	2	Number of polarisations
N_{chan}	3 072	Number of channels
Ω_{FoV}	610	Field of view
N_{beam}	6 100	Number of beams to tile FoV
$\Delta\nu$	30.72	Bandwidth (MHz)
τ	100	Integration time (μs)
t_{obs}	80	Observation time (min)

Table 2. Computational requirements and corresponding sensitivity limits for MWA reduced-resolution beamforming, for increasing fractional sensitivity.

f_s	d_{max} (m)	N_{beam}	$N_{baseline}$	C_{PXB} (TCMAC/s)	ΔS (10σ) (mJy)
0.2	17.4	16	389	12	10.0–15.0
0.3	30.8	53	810	84	6.7–10.0
0.4	48.3	130	1 353	343	5.0–7.5
0.5	70.4	277	2 121	1 146	4.0–6.0
0.6	123.2	851	2 970	4 931	3.3–5.0
0.7	154.0	1 330	4 084	10 596	2.9–4.3
0.8	180.3	1 825	5 368	19 112	2.5–3.8
0.9	211.1	2 501	6 643	32 412	2.2–3.3

Table 2. The computational cost is considerable, as the integration time $\tau = 100\,\mu$s only allows $N_{int} = 2$. For such a large number of beams, computational cost could be reduced significantly by using a 2D FFT imaging approach. Regardless, Selecting $f_s = 0.5$ yields $C_{PXB} = 1\,146$ TMAC/s, that is, roughly $1.5\times$ the cost of tied-array beamforming, and the current SMART tiling approach could be maintained.

While the beamforming cost is higher, the number of beams to search drops from 6 100 to 277. It follows that post-processing requirements will drop by a factor of $22\times$, while sensitivity drops only by a factor of $2\times$. We extrapolate that the post-processing requirements on OzSTAR would drop from $\sim$14 000 to $\sim$636 kSU. The SMART team anticipates securing 500–600 kSU per annum (Bhat et al. 2023a), meaning it is plausible that the entire survey could be processed within a year if reduced-resolution beamforming is adopted (instead of an implausible 22 years).

The processing requirements for SMART may also be alleviated by planned improvements to the performance of their pulsar search pipeline. The SMART pipeline currently uses PRESTO, a CPU-only code written primarily in ANSIC (Ransom 2011). GPU search codes, such as astroacclerate (Armour et al. 2020) and peasoup (Barr 2020), are hoped to offer an order-of-magnitude improvement. For example, the astroacclerate Fourier domain acceleration search is up to $8\times$ faster than its PRESTO equivalent (Adámek et al. 2020); recent work to exploit half-precision data types may offer an additional $1.6\times$ speedup (White et al. 2023). Alleviating I/O bottlenecks, such as disc read speed, may also improve efficiency. Regardless, any speedup from pipeline optimisation is complementary to the reduced-resolution beamforming approach.

4.2 Application to technosignature searches

A narrowband technosignature search of SMART data has been proposed, which would be the first all-sky technosignature search in the Southern Hemisphere. As such, a SMART technosignature search would place some of the most stringent constraints on the prevalence of putative engineered transmitters in the Galaxy. It is argued that all-sky SETI searches at low frequency are one of the most compelling methods to detect evidence of technologically capable life beyond Earth (Garrett et al. 2017; Houston et al. 2021). Narrowband technosignature searches do not require high time resolution, which decreases the computational cost of post-correlation beamforming. For example, decreasing the time resolution from 100 μs to 1 s decreases C_{PXB} (Equation (29)) by a factor of 10 000.

However, narrowband technosignature searches require high-frequency resolution, meaning that large data volumes must be stored in memory on which a large FFT can be performed. This will limit the number of beams that can be processed at any time. As with the pulsar search, the post-processing requirements for narrowband technosignatures scale linearly with the number of beams. Similar trade-offs between sensitivity and number of beams are thus well motivated for technosignature surveys.

5. Application to MeerKAT: Optimising survey speed

The TRAPUM programme (Chen et al. 2021) is a tied-array pulsar search programme on the MeerKAT 64-antenna array (Camilo 2018). Pulsar searches are conducted using up to $\sim$1 000 power beams, formed from a tied-array beamforming approach (Barr 2018). An incoherent beam may also be formed, and the maximum number of beams can increase to 4096 if a smaller subset of antennas is beamformed.

TRAPUM uses a PFoM modified for MeerKAT to optimise survey speed (Equation 4 Chen et al. 2021), from which the optimal number of antennas is found to be between 37 and 41, depending on the altitude observed; this corresponds to a $\sim$1 km maximum baseline and 58–64% fractional sensitivity. If reduced-resolution beamforming was applied, the corresponding fractional sensitivity would be 71–75% for the same baseline cut-off. The extra baselines would also slightly improve the beam sidelobe response by filling in the uv-plane.

Based on $N_{ant} = 64$, and time integration used in the TRAPUM search mode ($\tau = 76\,\mu$s, $N_{int} = 32$), the computational cost of reduced-resolution beamforming yields $C_{PXB} = 4C_B$. Memory requirements may also be higher, which could limit the number of reduced-resolution beams that could be formed.

Fig. 5 shows a comparison of survey speed against maximum baseline length for the MeerKAT array. If tied-array beamforming is used, the maximum baseline length corresponds to the size of the 'core' region of antennas used. For reduced-resolution beamforming, the maximum baseline length corresponds to the subset of baselines from the entire array. Note the PFoM for the tied-array beam is comparable to Figure 6 Chen et al. (2021). The PFoM is consistently higher for reduced-resolution beamforming. For $N_{beam} = 1\,024$, the maximum PFoM peaks $3.9\times$ higher for the reduced-resolution approach.

However, taking into account the instrument's capabilities, we conclude that post-correlation beamforming is unlikely to yield a higher PFoM. TRAPUM processes data streams in real time, and the beamformer is computationally-bound. So, the
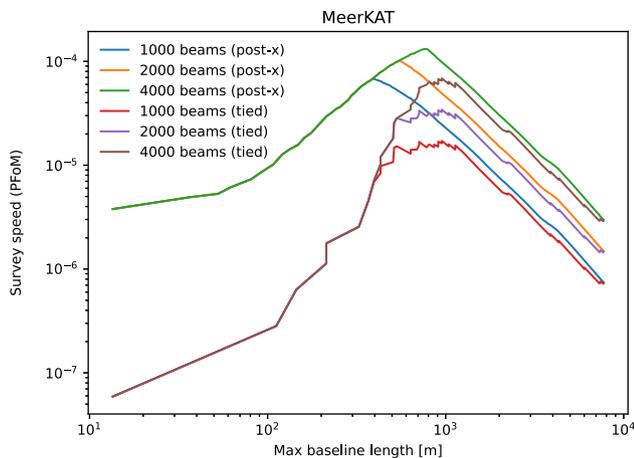
Figure 5. Comparison of MeerKAT survey speed with reduced resolution beamforming and tied-array beamforming approaches.

increase in computational cost ($C_{PXB} = 4C_B$) outweighs the benefit of the $3.9\times$ PFoM increase: $4\times$ fewer beams could be formed. Nevertheless, the approach may be useful for other projects on MeerKAT where the downstream search algorithms are computationally expensive.

6. Conclusions

This article has introduced reduced-resolution beamforming: a post-correlation beamforming approach that allows sensitivity to be traded against the number of beams needed to cover a survey area. Reduced-resolution beamforming offers an alternative to tied-array beamforming of a core region of antennas. The sensitivity of a reduced-resolution beam is always higher than tied-array beamforming of a core antenna region.

There are two pathways towards forming power beams: pre-correlation beamforming, in which voltages are summed, squared then averaged, and post-correlation beamforming, in which antenna pairs are cross-correlated before summation and averaging (Fig. 1). Our approach relies on the latter, in which per-baseline weights can be applied in lieu of per-antenna weights.

In Section 2, we introduced a tensor formalism to highlight the difference between pathways of Fig. 1 and showed that a pixel within an interferometric image is equivalent to a power beam. Tensors have become standard data structures within commonly used software packages such as Numpy and offer compact notation; given these advantages, we encourage the use of tensor formalisms for further research on beamforming approaches.

There are two main advantages of reduced-resolution beamforming. First, by decreasing the number of beams needed to cover a survey area, the downstream processing requirements to search the beams is lower: in Section 4, we show by applying reduced-resolution beamforming to the SMART survey, the same sky volume could be covered with $22\times$ fewer beams, while retaining 50% fractional sensitivity. Second, for a fixed number of beams, the survey speed of the telescope may be increased: in Section 5, we argue that the PFoM for the TRAPUM search mode on the MeerKAT telescope could be increased by $3.9\times$ by if reduced-resolution beamforming was adopted, albeit at the expense of increased computational requirements for beamforming.

While reduced-resolution beamforming demands a sensitivity trade-off, it could be offset by improved downstream processing.

For example, in a pulsar search, more DM or acceleration trials may become feasible if fewer beams need to be searched, improving the final signal-to-noise of candidates. One could also consider using reduced-resolution beamforming as a first-pass approach to find candidates above a less stringent threshold (e.g. 5σ), then form tied-array beams to follow-up candidates, setting a more stringent threshold (e.g. 10σ) for validation or rejection.

Nevertheless, there are limitations of the approach. Reduced-resolution beamforming produces power beams, so it cannot be applied in cases where voltage beams are needed (e.g. coherent dedispersion). For science cases where confusion noise is a relevant concern, such as continuum imaging, reduced-resolution beamforming is not appropriate. Another disadvantage is that the decrease in angular resolution means that any transient object found will have a comparatively poor localisation, although this could be improved using intra-beam methods (Obrocka et al. 2015). Finer localisation of candidate objects could also be done by reprocessing/reobserving at full resolution and then using known parameters (e.g. DM, period, pulse width) to narrow the search space. Finally, the computational requirements for post-correlation beamforming may become unfeasibly large for arrays with a large number of antennas (Equation (29)). Memory requirements for storing the correlation matrix and beamforming weights may also limit the number of beams that a single compute node could process.

All things considered, we conclude that reduced-resolution beamforming is best suited to science cases where downstream processing requirements dominate the processing budget, such as searches for pulsar, technosignatures, and fast transients. We encourage experimental application and verification of the approach to ease the data processing challenges faced by current and planned searches.

References

Adámek, K., Dimoudi, S., Giles, M., & Armour, W. 2020, in Astronomical Society of the Pacific Conference Series, Vol. 522, Astronomical Data Analysis Software and Systems XXVII, ed. P. Ballester, J. Ibsen, M. Solar, & K. Shortridge, 477

Anderson, D. P., Cobb, J., Korpela, E., Lebofsky, M., & Werthimer, D. 2002, CoACM, 45, 56

Armour, W., et al. 2020, AstroAccelerate, Zenodo

Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M., & Goss, W. M. 1982, Natur, 300, 615

Bailes, M., et al. 2017, PASA, 34, e045

Bannister, K. W., et al. 2017, ApJ, 841, L12

Barr, E. 2020, Peasoup: C++/CUDA GPU pulsar searching library, Astrophysics Source Code Library, record ascl:2001.014, ascl:2001.014

Barr, E. D. 2018, in Pulsar Astrophysics the Next Fifty Years, Vol. 337, ed. P. Weltevrede, B. B. P. Perera, L. L. Preston, & S. Sanidas, 175

Bhat, N. D. R., et al. 2023a, arXiv e-prints, arXiv:2302.11911

Bhat, N. D. R., et al. 2023b, arXiv e-prints, arXiv:2302.11920

Briggs, D. S., Schwab, F. R., & Sramek, R. A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 127

Camilo, F. 2018, NatAs, 2, 594

Chen, W., Barr, E., Karuppusamy, R., Kramer, M., & Stappers, B. 2021, JAI, 10, 2150013

Czech, D., et al. 2021, PASP, 133, 064502

Drake, F. D. 1961, PhT, 14, 40

Foster, G., Hickish, J., Magro, A., Price, D., & Zarb Adami, K. 2014, MNRAS, 439, 3180

Garrett, M., Siemion, A., & van Cappellen, W. 2017, arXiv e-prints, arXiv:1709.01338

Harris, C. R., et al. 2020, Natur, 585, 357

Houston, K., Siemion, A., & Croft, S. 2021, AJ, 162, 151

Hunter, J. D. 2007, CSE, 9, 90

Isaacson, H., et al. 2017, PASP, 129, 054501

Kent, J., et al. 2019, MNRAS, 486, 5052

Law, C. J., et al. 2015, ApJ, 807, 16

Law, C. J., et al. 2018, ApJS, 236, 8

Lebofsky, M., et al. 2019, PASP, 131, 124505

Lorimer, D. R., & Kramer, M. 2004, Handbook of Pulsar Astronomy (Vol. 4)

MacMahon, D. H. E., et al. 2018, PASP, 130, 044502

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993

Mills, B. Y., Little, A. G., Sheridan, K. V., & Slee, O. B. 1958, Proc. IRE, 46, 67

Morales, M. F. 2011, PASP, 123, 1265

Ng, C., et al. 2017, in XXXII International Union of Radio Science General Assembly & Scientific Symposium (URSI GASS) 2017, 4

Obrocka, M., Stappers, B., & Wilkinson, P. 2015, A&A, 579, A69

Okuta, R., Unno, Y., Nishino, D., Hido, S., & Loomis, C. 2017, in Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)

Portegies Zwart, S. 2020, NatAs, 4, 819

Price, D. C., & Smirnov, O. M. 2015, MNRAS, 449, 107

Price, D. C., et al. 2018, PASA, 35, e041

Ransom, S. 2011, PRESTO: PulsaR Exploration and Search TOolkit, Astrophysics Source Code Library, record ascl:1107.017, ascl:1107.017

Romein, J. W. 2021, A&A, 656, A52

Roy, J., Chengalur, J. N., & Pen, U.-L. 2018, ApJ, 864, 160

Ryle, M. 1962, Natur, 194, 517

Sanidas, S., et al. 2019, A&A, 626, A104

Singh, S., et al. 2023, ApJ, 944, 54

Smirnov, O. M. 2011, A&A, 531, A159

Stevens, A. R. H., Bellstedt, S., Elahi, P. J., & Murphy, M. T. 2020, NatAs, 4, 843

Swainston, N. A., et al. 2022, PASA, 39, e020

Tegmark, M., & Zaldarriaga, M. 2009, PhRvD, 79, 083530

Thekkeppattu, J. N., Wayth, R. B., & Sokolowski, M. 2024, arXiv e-prints, arXiv:2401.08039

Thompson, A. R., Moran, J. M., & Swenson, George W., J. 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition, doi: 10.1007/978-3-319-44431-4

Thyagarajan, N., Beardsley, A. P., Bowman, J. D., & Morales, M. F. 2017, MNRAS, 467, 715

Wayth, R., et al. 2021, JATIS, 8, 1

Wayth, R. B., et al. 2018, PASA, 35, e033

Welch, J., et al. 2009, IEEE Proc., 97, 1438

Wes McKinney. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt, & J. Millman, 56

White, J., et al. 2023, ApJS, 265, 13

Wijnholds, S. J., Willis, A. G., & Salvini, S. 2018, MNRAS, 476, 2029

Wilson, T. L., Rohlfs, K., & Hüttemeister, S. 2013, Tools of Radio Astronomy, doi: 10.1007/978-3-642-39950-3

Worden, S. P., et al. 2017, A&A, 139, 98