A Roadmap for Efficient Direct Imaging with Large Radio Interferometer Arrays

The E-field Parallel Imaging Correlator (EPIC) Collaboration: Nithyanandan Thyagarajan,1,* Adam P. Beardsley,2 Judd D. Bowman,2 Jayce Dowell,3 Greg B. Taylor,3 James Kent,4 and Daniel C. Jacobs2

1National Radio Astronomy Observatory, Socorro, New Mexico, USA
2School of Earth and Space Exploration, Arizona State University, Tempe, Arizona, USA
3Dept. of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico, USA
4Cavendish Laboratory, University of Cambridge, Cambridge, UK

ABSTRACT

Future radio arrays will have extremely demanding compute requirements. We present a technology development roadmap for direct imaging correlators which have the potential to alleviate harsh scaling. Overcoming technical challenges and generalizing algorithms will be essential to enable transformation science from next generation instruments. Submitted for consideration by the Astro2020 Decadal Survey Program Panel for Radio, Millimeter, and Submillimeter Observations from the Ground as a Technology Driver.
1. INTRODUCTION/BACKGROUND

Radio astronomy is entering an era in which interferometers consisting of hundreds to thousands of individual antennas are needed to achieve desired survey speeds. Nowhere is this more apparent than at radio frequencies below 1.4 GHz. The study of the history of hydrogen gas throughout the universe’s evolution is pushing technology development towards arrays of low-cost antennas with large fields of view and densely packed apertures. Similarly, the search for transient events and regular monitoring of the time-dependent sky is driving instruments in the same direction with the added requirement of fast read-outs. A number of new telescopes are under development around the world based on this new paradigm, several of which are illustrated in Fig. 1. In this white paper, we describe the growing need for alternative correlator architectures, and outline the technological barriers that must be overcome to enable cosmology and time-domain arrays envisioned for the 2020s.

In response to this Decadal Survey, many white papers have emphasized studying the Cosmic Dawn and the Epoch of Reionization (EoR) with observations of the highly redshifted 21 cm emission from intergalactic neutral hydrogen [1; 2; 3; 4; 5; 6; 7; 8]. The first generation of instruments designed to detect the large-scale spatial fluctuations in the 21 cm emission during reionization are currently performing science observations, including the Murchison Widefield Array (MWA) [9; 10; 11], the Donald C. Backer Precision Array to Probe the Epoch of Reionization (PAPER) [12], and the Low Frequency Array (LOFAR) [13]. A second generation instrument, the Hydrogen Epoch of Reionization Array (HERA) [14], was recently funded as a Mid-Scale Innovations Project, and will consist of 320 densely packed zenith pointed dishes plus 30 outriggers. Each of these arrays is highly optimized for the EoR measurement, yielding compact telescopes that are 1-2 km in diameter and consist of dipole antennas. In the case of HERA, nearly \( \sim 100\% \) of the core aperture will be filled.

Figure 1. New and planned radio arrays increasingly favor dense-packed configurations. Future arrays will require EPIC style backends in place of traditional correlators.

* t_nithyanandan@nrao.edu
for a total collecting area of $\geq 10,000 \text{m}^2$. The HERA roadmap\(^1\) calls for further development of technologies that will support significantly larger arrays.

After reionization, the neutral hydrogen is primarily confined to galaxies where the recombination rate is high, so the hydrogen traces the large scale structure. This opens the possibility of tracing the Baryon Acoustic Oscillations (BAOs) from a redshift of 1 to 6 from neutral hydrogen emission \(^{[15]}\). The Canadian Hydrogen Intensity Mapping Experiment (CHIME) \(^{[16]}\) and its southern hemisphere counterpart, the Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX) \(^{[17]}\), aim to directly measure the expansion history of the universe over a significant fraction of cosmic time. The ambitiously proposed Packed Ultra-wideband Mapping Array (PUMA\(^3\)) \(^{[18]}\) aims to use 32,000 elements spanning a diameter of 1.5 km at 50% filling factor.

**Both the HERA roadmap and the R&D plan for dark energy hydrogen intensity mapping\(^3\) call for development of FFT-based correlators to enable transformational science.**

Redshifted 21 cm experiments are not alone in driving the development of large, dense arrays. Radio transient searches are pushing into the same parameter space. Radio transients contain information about diverse astrophysical phenomena, ranging from the magnetic fields of exoplanets to exploding stars, compact object mergers, and ultra-relativistic flows. Thus, they provide invaluable insight into subjects as diverse as stellar evolution, relativistic astrophysics and cosmology \(^{[19]; [20]; [21]}\). In addition, follow-up observations of electromagnetic counterparts of gravitational wave events will provide a laboratory to study the extreme astrophysical events \(^{[22]}\). Several science white papers identified pulsar timing arrays as key instruments for probing the multi-messenger universe \(^{[23]; [24]; [25]; [26]; [27]; [28]}\). Arrays similar to the 21cm cosmology machines are ideal for radio transient searches because they combine high sensitivity with large fields of view, offering the possibility of real-time discovery, precise localization, and detailed physical insight \(^{[29]}\). However, these searches currently result in prohibitively high bandwidth requirements. For instance, the search for Fast Radio Bursts (FRBs) relies on monitoring tens to hundreds of square degrees at a cadence of $\lesssim 1 \text{ ms}$ and requires the ability to recover high temporal and spectral resolution images when an event is detected. Present-day digital backend architectures will be unable to meet these requirements when scaled to the specifications of next-generation radio arrays.

These important scientific objectives that will likely define the key astrophysics endeavors of the next decade require exceptional sensitivity. This can be most directly achieved via large collecting areas. In radio interferometer arrays, this translates to large numbers of collecting elements, often $> 1000$. For example, CHIME is currently operating with 2048 dipoles. Such massive instruments pose a leap in the amount of data to be stored and processed, and thus require data processing and transfer architectures. With scientific objectives relying on high cadence time domain information, there is an additional daunting challenge of processing the high data volumes in near real-time, writing them efficiently to storage, and triggering immediate follow-up observations.

The existence of a Fourier transform relation between the measured spatial correlation and the intensity distribution on the sky forms the basis for imaging with data from interferometer arrays. Most radio arrays to date have adopted a traditional correlator architecture that has served them well over the past decades. These correlators rely on measuring the spatial coherence of the incident

---

1 APC White Paper *A Roadmap for Astrophysics and Cosmology with High-Redshift 21 cm Intensity Mapping*, The HERA Collaboration
EM waves by correlating the voltages measured between each pair of antennas. The computing generally dominates the power budget, which scales as $N_a^2$, where $N_a$ is the number of antennas in the interferometer array. For large arrays, this implies a steep cost in computing and output data bandwidth. Ultimately, inability in affording this cost may limit the prospects of achieving the aforementioned key decadal science objectives. Resolving these challenges arising out of the data-intensive nature of next-generation interferometers requires novel and efficient architectures for processing, transfer, and storage.

Currently, alternate architectures are being explored to realize this Fourier transform relationship. They fundamentally rely on avoiding the spatial correlations which dominates the cost budget, and instead directly combine the antenna voltage measurements to obtain the intensity distribution in the image plane. By exploiting the efficiency of Fast Fourier Transforms (FFT), the computational cost can be potentially lowered from $\sim N_a^2$ to $\sim N_a \log_2 N_a$. This class of architectures can be broadly classified as Direct Imaging. There is a broad spectrum of applications for which different variants of direct imaging architectures are suited. One of the major recent advances in the direct imaging architecture has been the development and deployment of the E-field Parallel Imaging Correlator [EPIC; 30; 31; 32], which has consolidated existing direct imaging variants into a generic framework that can address the computational and output bandwidth needs of a wide range of antenna arrays that have large numbers of antennas and a dense layout. EPIC has extended the capability of simple direct imaging schemes that were only applicable to regularly gridded antenna layouts and identical arrays by bringing even arbitrary layouts consisting of non-identical behavior into the direct imaging umbrella. EPIC is now adoptable in time-domain imaging applications not only with telescope systems with gridded layouts such as HERA, CHIME, HIRAX, etc. but also in arbitrary layouts with known non-identical antenna behavior such as the Long Wavelength Array [LWA; 33] in Sevilleta (LWA-SV), OVRO-LWA, and other planned LWA stations, the proposed Deep Synoptic Array (DSA-2000\(^4\)), SKA1-low, etc.

The purpose of this white paper is to highlight the computing and time-domain imaging needs of next-generation arrays and provide directions for tackling the new challenges that arise with large next-generation arrays, specifically related to the real-time backend processing to meet unique scientific and instrumental needs.

2. DIRECT IMAGING: STATE OF THE FIELD

Direct imaging correlators have begun to be explored on deployed arrays including the Basic Element for SKA Training II (BEST-2) array [34], the MIT EoR (MITEoR) experiment [35] using the Omniscope [36; 37], an earlier incarnation at higher frequencies with the intent of pulsar timing [38; 39], and CHIME[16]. However, each of these examples make assumptions about the redundancy of the array layout, and require that the collecting elements are identical. On the other hand, the Modular Optimal Frequency Fourier [MOFF; 40] algorithm uses the antenna beam patterns to grid the electric field measurements to a regular grid in the software holography/A-transpose fashion [41; 42; 43] before performing the spatial FFT, achieving a scaling of $N_g \log_2 N_g$, where $N_g$ is the size of the grid in the FFT. This process has been shown to produce a data product mathematically
identical to images produced from traditional visibility-based techniques. We recently released the EPIC software\(^5\) [30] which implements the MOFF.

A challenge common to all direct imaging algorithms is the calibration of the antenna gains. With a traditional correlator, antenna-pair correlations are written to disk and used to calibrate offline before further processing such as imaging. However, a direct imaging correlator mixes the signals from all antennas before averaging and writing to disk, making calibration a requirement at the front end, before imaging and averaging. Previous solutions have involved applying calibration solutions generated from a parallel FX correlator [35; 34], or integrating the output of a dedicated FX correlator which periodically formed the full antenna-pair correlation matrix to solve for gains [44; 45]. While these solutions were sufficient to enable the exploration of FFT correlators and beamformers, they will not scale to future arrays with \(N_a \gtrsim 10^3\). Therefore, we recently devised the E-field Parallel Imaging Calibration (EPICal) algorithm [31] – a novel calibration scheme, which can be integrated into direct imaging correlators. It works directly with the antenna voltage streams and scales only as the number of antennas, \(O(N_a)\). An implementation of the algorithm is available with the EPIC software package.

Very recently, the EPIC architecture was successfully deployed on the LWA-Sevilleta (LWA-SV) station in New Mexico [32]. This has added significant new capability in increasing its best imaging cadence from \(\sim 1\) s to \(\sim 1\) ms at peak performance. This raises the prospects of undertaking more effective searches for astrophysically and cosmologically interesting transient events. EPIC is now available as an additional standard observing mode on LWA-SV telescope for its users interested in high time-resolution imaging. Further, EPIC has been tested (in offline mode) on data from other telescopes such as the LWA1, and OVRO-LWA.

When searching for high time resolution events, most existing radio telescopes operate in single-pixel (e.g. GBT, Arecibo) or beamforming modes (e.g. CHIME) with limited ability to localize the event on the sky on their own, though the Australian SKA Pathfinder (ASKAP) is showing first signs of a transformational ability to not only detect but also image and provide arcsecond localization of the transient events [46]. Providing modern instruments with the ability to search blindly and localize transient events will be a key ingredient to their success.

3. LANDSCAPE OF MODERN RADIO INTERFEROMETER ARRAYS

There are many current and upcoming large radio interferometers with a wide range of scientific applications. The diversity of their scientific goals and their unique technological challenges are leading to a variety of instrument designs. We identify below some of these broad design categories along with a few examples predominantly focused on cosmology and time-domain applications, and provide a roadmap for algorithmic, architectural, and technological advancements that will address the needs and challenges of these modern instruments.

3.1. Dense, Non-redundant and Hybrid Arrays

Interferometer arrays that rely on imaging will predominantly follow a non-redundant antenna placement. If study of large-scale structures is one of their key goals, the layout is expected to provide a high density of short antenna spacings. Examples include cores of the proposed DSA-2000

\(^5\) https://github.com/epic-astronomy/EPIC
and the OVRO-LWA, the LWA stations such as those that exist currently in New Mexico and the
ones planned, core stations of SKA1-low, and the core tiles of the MWA.

The DSA-2000 is proposed to consist of 2000 \times 5 \text{ m} dishes spanning a 15 \text{ km} diameter area covering
the 0.7–2 GHz frequency band. It will be used for pulsar timing, and will also detect and localize
fast radio transients. The LWA stations in New Mexico (LWA1 and LWA-Sevilleta) consist of a
dense arrangement of 256 dipoles mostly placed inside a \sim 100 \text{ m} diameter. They operate at 10–
88 MHz. Their science goals include probing cosmic evolution and the high redshift Universe (e.g.,
cosmic dawn, EoR) and the transient Universe. The OVRO-LWA is focused on probing the cosmic
dawn via the redshifted 21 cm from H\text{\textsc{i}}, and detecting radio transients such as auroral emission from
exoplanets, coronal mass ejections from stars, compact object merger events, solar dynamic imaging,
etc. It currently operates in the frequency band 27–85 MHz with 288 dipoles with a majority of
them in a 200 m diameter area and the rest spread out inside a 1.5 km region. A future expansion
of OVRO-LWA will consist of 2000 dipole elements within 2.5 km with the core comprising of a few
hundreds of elements placed within a few hundred meters. EPIC has already been demonstrated on
data from these different LWA stations in software and is currently deployed on the LWA-Sevilleta
station [32; 30; 31].

The MWA consists of 4096 dual-polarization dipole antennas optimized for the 70-300 MHz fre-
quency range, arranged as “tiles”, each a 4\times 4 array of dipoles, the majority of which lie within a
\sim 1.5 km core region. EoR and transients are two key science drivers of the MWA. The array is a
technology and science precursor for the SKA. The first phase of SKA1-low will consist of \sim 130,000
dipole elements arranged into \sim 500 stations of \sim 40 m each consisting of 256 dipoles operating in
the 50–350 MHz band. It will consist of a core of stations with \sim 50% of collecting area packed into
\sim 1 \text{ km} diameter. Its key science goals include the study of the redshifted 21 cm line during the EoR
and Cosmic Dawn below redshifts of \( z < 27 \), and search for undiscovered fast transient phenomena.

3.2. Dense, Redundant Arrays

Redundantly spaced arrays are much more suited to detect spatial structures from the cosmic signal
on specific length scales or detect transient events on specific beamformed locations without spatial
structure information, rather than as a traditional imaging instrument. However, they still face the
challenge of computing and storing all the cross-correlated signals at cadences of \lesssim 1 \text{ s}.

The most useful scientific products in these cases are correlations, and not images. A direct imaging-
based approach will be still useful as it provides a big computational edge through the FFT algorithm.
For example, a spatial FFT followed by pixel-wise squaring and an inverse FFT will yield the required
correlations at FFT-like efficiency rather than as \( O(N^2) \). This is also referred to as an FFT correlator.
Modern arrays with redundant layouts are considering such an approach. Examples include HERA,
CHIME, HIRAX, and PUMA.

Current HERA will build out to 350 dishes of 14 m diameter each, with 320 of them in redundant
close-packed configuration. It will operate in the 50–250 MHz frequency range. Its primary scientific
objective is to probe the redshifted 21 cm signal from the EoR and Cosmic Dawn at \( z > 6 \).

CHIME is operating at 400–800 MHz with 4 close-packed parabolic cylinders each 100 m long and
20 m wide placed breadth-wise. Each cylinder consists of 256 dual-polarization dipole antennas with
a total of 1024 collecting elements. One of the key goals of CHIME is to produce intensity maps
of redshifted 21 cm from H\text{\textsc{i}} and detect the Baryon Acoustic Oscillations (BAO) during the Dark
Energy dominated era, \( 0.8 < z < 2.5 \).
Recently, CHIME has reached a major breakthrough by reporting 13 FRBs down to 400 MHz during a pre-commissioning phase [47]. The experiment is projected to detect hundreds of FRBs per year. HIRAX is a similar instrument in the southern hemisphere. It will consist of closely packed 1024 parabolic dishes in a gridded $32 \times 32$ layout each 6 m in diameter.

PUMA is a proposed ultra-wide band, high-sensitivity and low-resolution radio telescope array to operate at 200–1100 MHz, optimized for cosmology with 21 cm intensity mapping in the post-reionization era; and addressing other science goals amenable to such observations, including FRBs, pulsar monitoring and transients as a part of multi-messenger observations. Depending on the configuration (PUMA Petite or PUMA full), the number of antennas is proposed to range between 5000–32000 each with a diameter of 6 m, spread over a diameter of 600–1500 m placed in a close-packed hexagonal arrangement. With such a large number of antennas, it is essential to transform the paradigm of the data processing architecture, and indeed the white paper is proposing to deploy an FFT-correlator, which is a specific variant of the EPIC architecture concept.

### 3.3. Stations of multiple, dense, non-redundant arrays

The next-generation of large-N interferometers is going to include a hybrid version of aperture arrays, wherein there will be many “stations” each of which consists of dozens to hundreds of smaller antennas. Each station may be beamformed to act as a single larger aperture or the individual elements in the stations may each be used as an interferometric element. Examples of this model include the MWA, LOFAR, the LWA Swarm concept being proposed\(^6\), and the SKA1-Low.

These represent a multi-layered interferometer array suitable for multiple spatial scales. The inter-station spacings will achieve a fine angular resolution while the intra-station spacings can be used to capture large-scale structures in images with excellent surface brightness sensitivity. The intra-station images could be used for testing of coincidence rates and null hypotheses. When the data from all stations is combined while preserving the full interferometric information, they will represent a transformation in multi-scale hybrid imaging where the intra-station data adopt a direct imaging approach, while the traditional correlator approach can be applied to the inter-station data.

### 3.4. Multi-scale, Non-redundant Large Arrays

Modern broad-purpose arrays that serve a wide range of science are also emerging. A prime example will be the Next Generation Very Large Array (ngVLA\(^7\)). The ngVLA is proposed to operate in the frequency range 1.2–116 GHz with the array consisting of 244 18 m dishes. It consists of a core configuration made of 94 antennas in a semi-random distribution within a region of 1 km diameter, and a short-baseline array of 19 antennas of 6 m diameter in a very close packing. The key science goals include: 1) probing the origins of life in the Universe through unveiling the formation of solar system analogs, and probing the initial conditions for planetary systems and life, 2) studying the structure and evolution of galaxies from the early Universe to the present, 3) fundamental tests of gravity and extreme states of matter by studying pulsars in the Galactic center, and 4) formation and evolution of stellar-mass and super massive black holes. Particularly, the time-domain science applications include pulsar timing and gravitational waves, and FRBs.

A direct imaging approach may not be directly applicable yet to such generic arrays without further improvements. However, such approaches can be extremely useful as a diagnostic tool rather than a

---

\(^6\) APC White Paper: *The Swarm Development Concept for the LWA*, Taylor et al.

\(^7\) https://ngvla.nrao.edu/
Figure 2. Comparison of computational cost (left) and output bandwidth (right) with EPIC and FX-based approaches for a fast transient campaign (images at 0.5 ms cadence) using various interferometer arrays. The dotted line denotes where performances are equal. Points below the line indicate that an EPIC-based approach is more favorable. Blue symbols denote the full array layouts while orange symbols denote a core only. “SKA-Low Extreme Core” is a hypothetical configuration where all dipoles of the SKA1-Low core are correlated. “HERA-III” is an array which is filled with HERA’s core density out to the location of the outriggers. “MWA-III” consists of all 256 tiles currently deployed in two configurations for MWA-II.

4. PROSPECTS OF DIRECT IMAGING FOR LARGE, MODERN INTERFEROMETERS

The primary technical aspects that the general class of direct imaging approach proposed here aims to address are the computational cost and the output bandwidth, specifically for time-domain and cosmological applications. Figure 2 shows the computational cost (GFLOPS) and the output bandwidth required (in GB/s) for some of the current and planned telescopes with the standard correlator- and EPIC-based approaches for a processing cadence of 0.5 ms which is typical of fast time-domain applications. With EPIC, the output is calibrated images, while the output of the FX correlator is uncalibrated antenna-pair correlations. We assume application of the standard EPIC architecture as it exists at present. It is clearly seen that both aspects in many of the next-generation instruments are already efficiently addressed by the current EPIC architecture (the arrays that fall below the dotted line). Of those for which that’s not the case, most of them lie within an order of magnitude of the correlator-based approach in one or both aspects. This indicates that advancements in the direct imaging approach, some of which are listed below, will be able to provide additional efficiency to make it viable for a majority of the large, modern radio interferometers.

The direct imaging approach combined with the technology in modern arrays aims to push the envelope of time-domain capabilities. Figure 3 shows the computational tradeoff as the temporal resolution required for time-domain applications get more demanding than 0.5 ms cadence using the a hypothetical MWA-III configuration as an example. A correlator-based approach is bound to perform worse as temporal resolution required is \( \lesssim 0.2 \) ms. This is because the ability to average correlations over longer integration times before imaging is significantly reduced as the timescale of interest becomes shorter. The cost of imaging required at faster cadence starts to dominate.
direct imaging approach is unaffected because the imaging step is designed to operate at the highest cadence regardless of the output cadence.

This has a significant implication as modern arrays further seek to expand the time-domain capabilities. The temporal resolution required in the output plays a significant role in the efficiency achievable with a correlator-based approach. It worsens with faster cadence. This example predicts an important trend for all the arrays shown in Figure 2. As the computations and bandwidth become inefficient in the correlator-based approach while remaining unaffected in the EPIC-based approach, a faster cadence will see all the telescopes moving horizontally to the right in the parameter space of Figure 2. This will make the EPIC-based approach even more favorable for many modern arrays.

Modern arrays promise to open to new windows to studying the Universe. But it requires addressing the computational and data rate challenges. The EPIC-based approach has shown promise with its initial successful deployment. In the next decade, we need to prioritize advancing the current EPIC-based approach to include the different arrays designs. Applying an EPIC-based approach also requires devising a number of algorithmic and technological solutions to achieve the scientific goals with high efficiency in computations and bandwidth and can thus drive new technologies and algorithms.

5. DRIVER OF NEW TECHNOLOGIES AND ALGORITHMS

To realize the promise of direct imaging correlators, and unlock transformational science cases in cosmology and time-domain astrophysics, many technical challenges must be overcome. Some of these issues will be shared with any effort to move to real-time processing (e.g. calibration), while others are unique to the direct imaging paradigm. Here we enumerate what we see as the most pressing issues that must be addressed to support the large-N arrays envisioned for the next decade.

5.1. Calibration

Traditional interferometric calibration has involved modeling antenna-pair correlations and solving for complex gains for each antenna and frequency before further processing the data. This task must be completed before forming images, which combines signals from many antennas and destroys the independence of the measurements. Any real-time imaging system will need to lock in the calibration before imaging, the precision of which is paramount for cosmology experiments which rely on decoupling bright foregrounds from the dim signal via spectral structure \[48; 49\].

Direct imaging systems in particular will need to find novel ways to calibrate because they never form the correlations which have been the basis of traditional calibration. As noted in Sec. 2, several solutions have been proposed and early demonstrations have been performed. Any method adopted for calibrating direct imaging arrays must be developed to production level. This will include demonstrating robustness against data contaminants such as RFI, direction-dependent calibration for wide field of view instruments, and valid solutions in the face of real-world complex skies.

![Figure 3](image.png)
5.2. Optimization

Direct imaging to date has been mostly performed on available hardware which can be repurposed relatively easy for experimental demonstrations (usually GPUs). As the field moves toward real applications on large-N arrays, it will be necessary to highly optimize the computing to fully leverage the $N \log N$ scaling. This will likely involve porting to more specialized hardware such as field programmable gate arrays (FPGAs) or even application-specific integrated circuits (ASICs). While the experimental nature of early demonstrations and the high cost to entry have previously kept scientists fixed to flexible GPUs, circuit design experts will need to be engaged in the near future.

Beyond moving to more efficient hardware, optimization on these chips will require novel developments in signal processing. For example, [32] showed that while the operations of the EPIC correlator scale well and are in principle parallel, memory bandwidth can easily become a bottleneck. Unlike an FX correlator where the operations are very simple, direct imaging involves shuffling data significantly and developments in on-chip memory management will likely lead to early breakthroughs. While the EPIC algorithm is in principle general, some of these optimizations may be instrument-specific and substantial development will be needed for each application.

5.3. Advanced Imaging Algorithms

Many of the advances in radio synthesis imaging will need to be reinvented in the framework of direct imaging. For example, $w$–projection [50] can be used to account for non-coplanarity of antenna elements, and anti-alias gridding can remove artificial processing artifacts. While these algorithms can easily be extended to E-field measurements, they remain to be demonstrated in practice. Deconvolution (e.g. CLEAN [51]) will likely be necessary for precision cosmology experiments which need to overcome foregrounds orders of magnitude brighter than the target signal. The Fast Holographic Deconvolution (FHD [52]) was built on the same holographic/A-transpose framework as the MOFF algorithm, and so will be a good candidate to generalize. Algorithmic advances along these lines are likely to be general to the diverse applications described in Sec. 3.

5.4. Generalization

The concept of direct imaging is not new [53; 38], but the the earliest versions strictly required rectilinear layouts. [37] relaxed this constraint by generalizing it to incorporate shears and rotations on hierarchical grids. Even then, the point spread function (PSF) of such an array suffers from severe grating lobes due to the regularity. The MOFF algorithm took another approach to generalize by introducing a gridding step and completely removing the antenna regularity requirement [40]. While allowing freedom of antenna placement, this method is most efficient for very compact arrays. We see in Fig. 2 that some very large arrays still do not benefit from this paradigm due to relative sparsity. Further development is needed to generalize to arbitrary large-N arrays while still achieving a tolerable scaling. There are several potential ways to extend the EPIC framework:

1. The sparse FFT [54] can be used to avoid unnecessary computations when the data grid is largely zeros. Optimizing this algorithm on ASICs will be technically challenging. However, because the array footprint is fixed, it may be highly optimizable.
2. FFT-Log is an analogue to the standard FFT which produces the exact Fourier transform for logarithmically spaced samples [55]. This can be used in conjunction with arrays like the DSA-2000 which have compact cores with decreasing density at larger radius.
3. A potential hybrid scheme could use the antenna gridding of EPIC combined with the hierarchical gridding of Omniscope to account for relatively sparse antennas while removing grating in the PSF by dithering antenna positions.

4. In the case of arrays with dense cores and only a few distant outriggers, a direct imaging algorithm could be used for the core while traditional FX correlations are formed between the outriggers and the outrigger-core baselines. The correlations would scale as $N_{\text{core}} \times N_{\text{outriggers}} \ll N_s^2$, and can be incorporated into the image in post processing.

5. For non-redundant, non-compact arrays, a direct Fourier transform method can be used to efficiently beamform the antenna signals to a large number of sky pixels (Kent et al. in prep.). It is likely that combinations of the above solutions will be needed to accommodate future arrays, and each instrument will likely need to be uniquely optimized.

5.5. **Transient backends**

Extremely high time resolution imagers produce enormous output data volumes quickly. Efficient and reliable transient search backends will be required to compress the data and/or trigger data dumps. The Fast Dispersion Measure Transform (FDMT) [56] was developed to search many DM and arrival time trials on radio data. However, when performing blind searches on many resolution elements, the FDMT will require computing at least as large as the correlator itself. To keep up with this data rate, highly optimized implementations will need to be produced. Specific science cases may also find ways to cut down on parameter space based on a priori knowledge or assumptions. For example, searches for prompt radio emission following gravitational waves from neutron star mergers may benefit from searching a more narrow DM space based on localization or sensitivity information.

6. **CONCLUSION AND RECOMMENDATIONS**

Large, modern interferometer arrays are emerging with enormous promise of expanding the scientific horizons. Study of large-scale structures in the early Universe and explosive time-domain events are two key frontiers. These topics have not only been deemed as significant goals worthy of pursuit in the last decadal survey but also continue to remain extremely relevant for the next decade. Traditional correlator-based data processing techniques may fall short of achieving these objectives if the computational costs and output bandwidths are not kept economical. We have commenced efforts in the last decade to address these challenges using an alternate and generic direct-imaging approach. Following this promise, it is evident that an extension of such an architecture to a wide range of new arrays will contribute to realizing their full potential, especially in the time-domain and studies on cosmological scales. Indeed, a number of new, powerful interferometer arrays are being proposed – the PUMA, DSA-2000, individual LWA stations and the LWA Swarm, and the ngVLA – to name a few. Adopting a direct imaging approach for these arrays both require and drive new technologies and algorithms built around this approach. We have identified new generalizations in algorithms for calibration, data processing and imaging, custom optimizations for efficient processing and bandwidth management, and transient search-and-monitor backends as the key priorities for the next decade in this direction. This roadmap applies to many classes of large interferometer arrays that will be proposed or operational in the next decade rather than focus on a specific instrument. While this technology development will support larger projects, the activities described within this paper qualify as a small (< $20M) ground-based project.
REFERENCES


