HERA – Hydrogen Epoch of Reionization Arrays

Donald C. Backer[1], James Aguirre[2], Judd D. Bowman[3], Richard Bradley[4,5], Chris L. Carilli[4], Steven R. Furlanetto[6], Lincoln J. Greenhill[7], Jacqueline N. Hewitt[8], Colin Lonsdale[9], Sterphen M. Ord[7], Aaron Parsons[1], & Alan Whitney[9]

[1] *University of California, Berkeley; dbacker@astro.berkeley.edu; 510-NGC-5128*

[2] *University of Pennsylvania* [3] *California Institute of Technology*

[4] *National Radio Astronomy Observatory* [5] *University of Virginia*

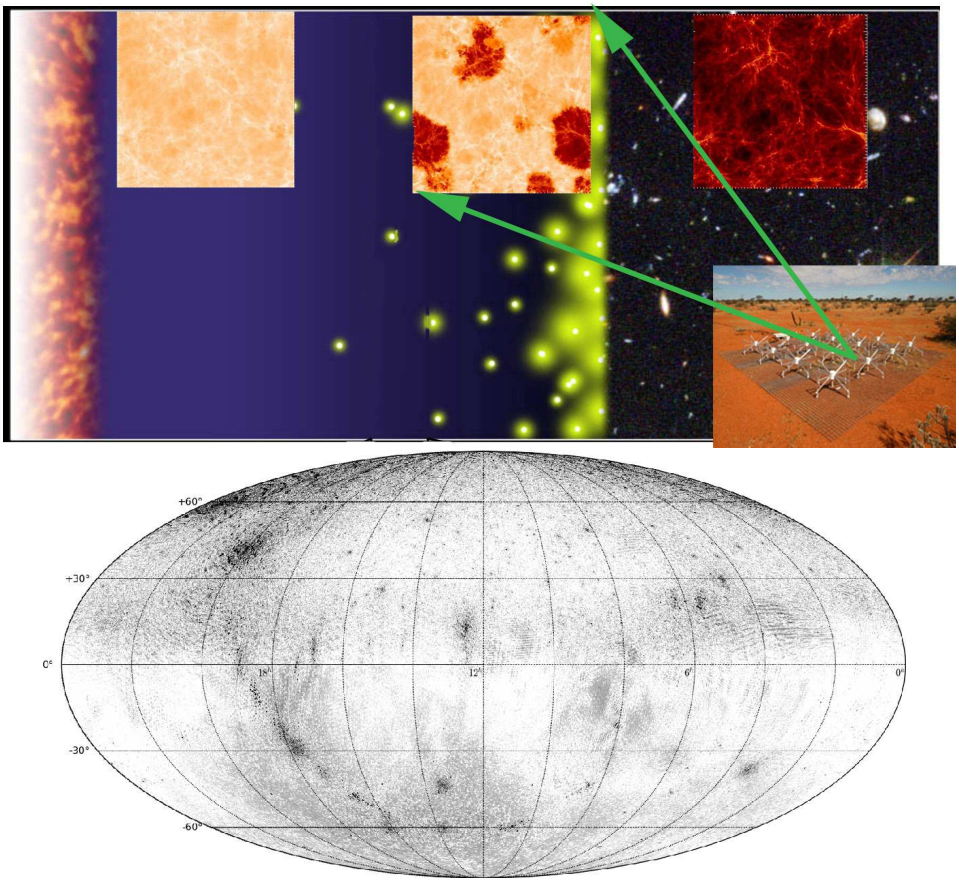
[6] *University of California, Los Angeles*

[7] *Harvard-Smithsonian Center for Astrophysics*

[8] *Massachusetts Institute of Technology* [9] *Haystack Observatory*

[11] *University of Washington* [12] *Curtin University of Technology*

RFI2: Submitted for consideration by the Astro2010 Decadal Survey Program Panel
RMS: Radio (Meter/Centimeter) and Millimeter/Submillimeter



Top: Simulated history of the Universe from recombination to the present, with example 21cm signals and a Murchison Widefield Array tile overlaid. Bottom: All-sky map at 150 MHz from Precision Array to Probe the Epoch of Reionization; scale log[Jy]: peak 10^4 Jy; min rms 100 mJy. Credits: G. Djorgovski, S. Furlanetto, J. Lazio, A. Parsons et al. (2009).

I. EXECUTIVE SUMMARY

US astronomers are leading the first exploration of large-scale structure in the high-redshift baryonic universe via the 21 cm line of hydrogen. We present a roadmap for the coming decade for the development of Hydrogen Epoch of Reionization Arrays (HERA) in order to achieve the key scientific goals of understanding the development of the first galaxies and their influences on the universe around them.

We repeat a statement about secondary science goals (radio sky catalog, pulsars, transient sources, solar and heliospheric physics) from our RFI1 submission, but do not develop the cost impact of this potential research that is not a simple and direct output of our main science program.

Our roadmap is divided into three phases:

1. 2010-2014—*HERA I: Detection* of the power spectrum of the 21 cm line emission from the epoch of reionization using the Murchison Widefield Array (MWA) and Precision Array to Probe the Epoch of Reionization (PAPER). These radio arrays are already under development (2008-2011; *HERA IA*). A second phase of activity during 2012-2015 has goals of completing and extending the arrays, campaigns to collect data, data analysis, and R&D leading up to a HERA II proposal in 2014. An expenditure of FY09 $25M is developed in this RFI2 (*HERA IB* with some costs through 2016).

2. 2015-2019—*HERA II: Characterization* of the 21 cm power spectrum and additional statistics yielding significant new galaxy formation astrophysics and cosmological physics by the end of the decade. A second-generation array with increased sensitivity will be constructed for this purpose. An expenditure of FY09 $62M is developed in this RFI2.

3. 2020 and beyond—*HERA III: Detailed imaging* of structures across a large fraction of cosmic time will be measured using a large third-generation radio array that could also serve as the core of the SKA-low.

In this HERA RFI2 whitepaper, we focus on the development and costing of the second-generation radio array capable of meeting the phase II science objectives. The reference design of the interferometer we propose is an array of 5000 dipole-based antennas with a total effective collecting area of $100,000$ m^2. The technical challenge of this array is dominated by the enormous data rate from so many antennas and the need for a very high precision calibration solution. We name our reference array "HERA-II". It will map 1000 deg^2 or more of the sky with 10 mK sensitivity over a wavelength range between 1.5 and 3 m (100 to 200 MHz).

Construction of HERA-II will commence in 2015 and the science program will begin in 2017. All of the core technology and analysis techniques required for the array will be fully demonstrated by the PAPER and MWA pathfinder arrays, both of which are scheduled to begin operations in 2010.

Significant sharing of the costs with international partners is possible. The costs already are reduced from total project costs owing to assumed infrastructure contributions from host country of the HERA. The costing estimates in this whitepaper are based on the real expenditures accrued by MWA and PAPER between 2005 and the present. With the high level of activity and progress of the two pathfinding efforts over the next few years, a HERA-II proposal mid-decade will be both firmly justified scientifically and accurately costed.

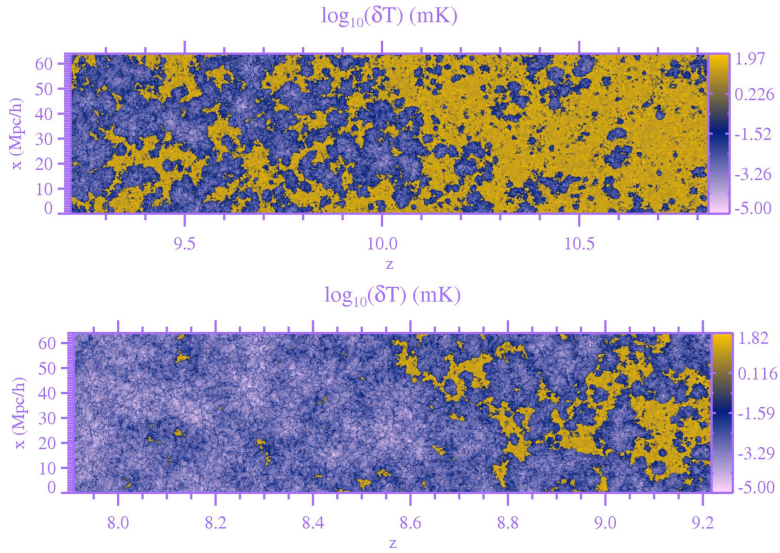


FIG. 1: Simulated maps of the 21cm background during the early and late stages of a particular reionization scenario (top and bottom panels). Purple regions are highly ionized; yellow regions are mostly neutral. In the vertical direction, the slice subtends ∼35' on the sky. From [23].

II. KEY SCIENCE GOALS OF HYDROGEN EPOCH OF REIONIZATION ARRAYS (HERA)

A. Primary Goal: Hydrogen at High Redshift

1. Introduction. Measurements of our Universe's fundamental parameters and its history have improved enormously over the past twenty years. One of the key remaining challenges is to explore (and exploit) the "high-redshift frontier" at $z > 7$ before and during the formation epochs of the first galaxies. Here we describe a new kind of telescope that can study the otherwise inaccessible cosmic "dark ages" (at $6 < z < 50$, during and before the "reionization" of intergalactic hydrogen) with the 21cm transition of neutral hydrogen: the Hydrogen Epoch of Reionization Array (HERA). HERA will address two sets of key questions: **what were the properties of high$-z$ galaxies, high-z galaxies and black holes, and how did they affect the Universe around them?**, and **does the standard cosmological model describe the Universe during the "dark ages?"** These are described more fully in the two science white papers, submitted to the Astro 2010 Decadal Survey, "Astrophysics from the Highly-Redshifted 21cm Line" and "Cosmology from the Highly-Redshifted 21cm Line."

Observations of HI 21cm emission from the neutral IGM offer a number of unique and powerful probes into the formation of the first galaxies and cosmic reionization [7]. The signal is rich in physical diagnostics, depending on four properties of the IGM: density, neutral fraction, "spin temperature" (i.e., the excitation temperature of the 21cm transition), and the local velocity field (sourced primarily by gravity). Because it specifically probes the neutral IGM, this provides the ideal complement to observations of the first galaxies and quasars in the near-IR using eg. JWST, and to CMB polarization studies, which probe the ionized IGM. Moreover, the 21cm line is the only direct probe of the preceding "dark ages," prior to the formation of the first stars and black holes.

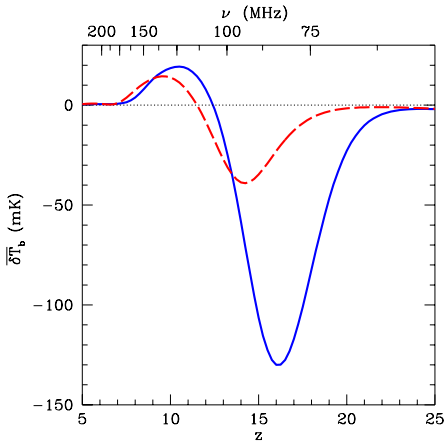


FIG. 2: Fiducial histories of the sky-averaged 21cm brightness temperature, $\delta\bar{T}_b$. The solid blue curve uses a typical Population II star formation history, while the dashed red curve uses only very massive Population III stars. Both fix reionization to end at $z_r \approx 7$. From [6].

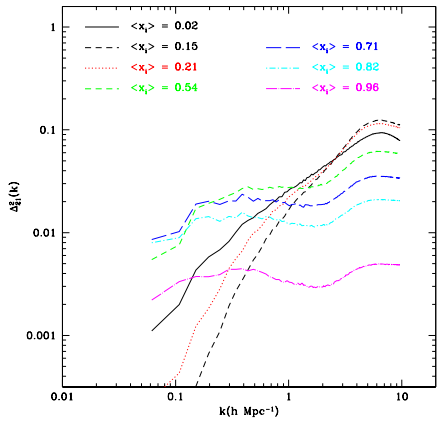


FIG. 3: Evolution of the spherically-averaged 21cm power spectrum during reionization, from a numerical simulation of that process. Here the fluctuations are presented in units of $\delta\bar{T}_b^2$ for a fully neutral universe; $\langle x_i \rangle$ is the ionized fraction during reionization. From [11].

2. Science and Required Measurements: Fig. 2 shows two example scenarios for the sky-averaged 21cm brightness temperature during the era of the first galaxies. The solid curve assumes that high-z star formation has similar properties to that at lower redshifts; the dashed curve assumes that these stars were all very massive metal-free objects. In either case, when the first galaxies form, they flood the Universe with ultraviolet photons that break the equilibrium between the CMB and 21cm line – rendering the gas visible in absorption against the CMB [14]. Somewhat later, X-rays – produced by the first supernovae or black holes – heat the IGM, turning this absorption into emission, which fades as ionizing photons from the maturing galaxies destroy the intergalactic HI. The large variations between these scenarios indicate that the 21cm background is quite sensitive to the basic parameters of high-z star formation.

Fig. 1 shows the most dramatic epoch to which the 21cm line is sensitive: the "reionization" epoch where ultraviolet photons from high-z galaxies ionized the entire IGM over a relatively short time interval [5]. This event marked the point when the small fraction of matter inside galaxies completely changed the landscape of the diffuse IGM gas. The 21cm background provides the *ideal* probe of reionization. Its weak oscillator strength (in comparison to Lyα) allows us to penetrate even extremely high redshifts. We can also image it across the entire sky – instead of only rare, isolated Lyα forest lines of sight. Moreover, unlike the CMB, it is a spectral line, and we can easily separate redshift slices to study the full history of the "dark ages." Finally, it directly samples the 95% (or more) of the baryons that reside in the IGM, requiring no difficult inferences about this material from the properties of the rare luminous galaxies.

From Fig. 1, the 21cm signal clearly fluctuates strongly as individual IGM regions grow through gravitational instability and are heated or ionized by luminous sources. These fluctuations can ideally be imaged, but over the short term statistical measurements are

likely to be more powerful, and these comprise the primary goal of HERA. Fig. 3 shows how these statistical fluctuations (here parameterized by the power spectrum) evolve throughout reionization. The fluctuations briefly fade as galaxies ionize their dense surroundings in the first stage of reionization, then increase as large ionized bubbles form, finally fading again as the gas is ionized. At higher redshifts, variations in the ultraviolet and X-ray backgrounds, in addition to the normal density fluctuations in the IGM, induce similar structure in the 21 cm background. These fluctuations will allow HERA to study two key questions, encompassing both galaxy formation and cosmology.

What were the properties of high−z galaxies, and how did they affect the Universe around them? The timing, character, and topology of reionization depend strongly on the properties of the first generations of luminous sources – exotic Population III stars, their more normal descendants, and even the first quasars. At the same time, the 21 cm background will provide a direct view of the emerging "cosmic web" that we see in later epochs as the Lyα forest, both statistically and possibly through absorption spectra of the first radio loud AGN. Understanding the interface between the luminous sources and their surroundings is key to unlocking the mysteries of galaxy formation.

Does the standard cosmological model describe the Universe during the "dark ages?" A second focus of this experiment is to open an entirely new cosmological era to precision tests. The prospects for "new" physics during this era are difficult to quantify but nevertheless exciting, and 21cm surveys can *potentially* dramatically improve cosmological constraints on parameters such as the inflationary power spectrum, the neutrino mass, and the curvature of the Universe [15]. A particularly promising avenue to isolate the cosmological information is through the *redshift-space distortions* of the 21cm background, in which peculiar velocities change the mapping from frequency to radial distance, amplifying fluctuations along the line of sight (but not in the plane of the sky). The peculiar velocities that source these distortions depend almost only on gravity and so provide a purer view of the underlying cosmological parameters. A key goal of HERA will be to measure these redshift-space distortions and to assess their potential utility for cosmological and astrophysical measurements.

3. Technical Implementation & Performance Requirements: The ultimate goal of studying the 21cm background is to make detailed maps of the IGM throughout the "dark ages" and reionization, as in Fig. 1. The top axis of Fig. 2 shows the observed frequency range for these measurements: well within the low-frequency radio regime. Unfortunately, this is an extremely challenging band, because of terrestrial interference, ionospheric refraction, and (especially) other astrophysical sources (see [7]). In particular, the polarized Galactic synchrotron foreground has $T_{\rm sky} \sim 180(\nu/180\ {\rm MHz})^{-2.6}$ K, at least four orders of magnitude larger than the signal. For an interferometer, the noise per resolution element (with an angular diameter $\Delta\theta$ and spanning a bandwidth $\Delta\nu$) is then

$$\Delta T_{\rm noise} \sim 2\ {\rm mK}\ \left(\frac{10^5\ {\rm m}^2}{A_{\rm eff}}\right)\left(\frac{10'}{\Delta\theta}\right)^2\left(\frac{1+z}{10}\right)^{4.6}\left(\frac{\rm MHz}{\Delta\nu}\frac{100\ {\rm hr}}{t_{\rm int}}\right)^{1/2}, \qquad (1)$$

where $A_{\rm eff}$ is the effective collecting area and $t_{\rm int}$ is the integration time. These angular and frequency scales correspond to ∼30 Mpc.

Because only large scales are accessible to imaging, even with a large telescope, the current HERA program emphasizes statistical measurements with interferometers. Figure 4 shows some estimates for how well we can measure the 21cm power spectrum with a variety of experiments at two different redshifts of interest; HERA-II corresponds to the solid line

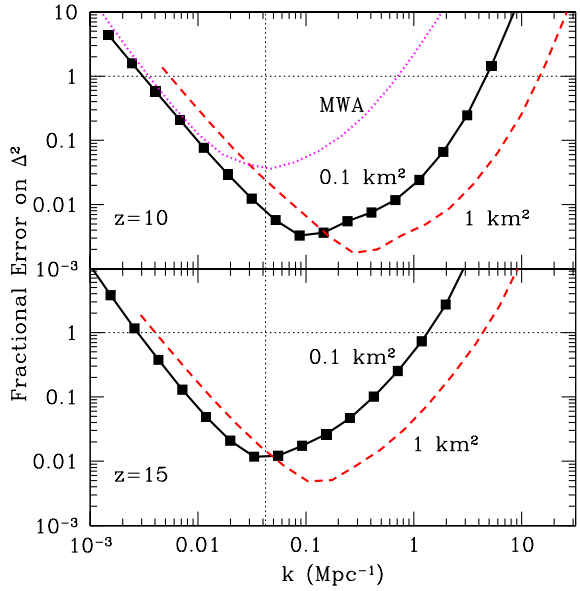


FIG. 4: Sensitivity of three fiducial arrays to the spherically-averaged power spectrum, at $z = 10$ and 15, expressed in units of $\delta\bar{T}_b^2$ for a fully neutral universe, as in Figure 3. The solid and dashed curves are for $A_{\rm eff} = 0.1$ and 1 km^2, respectively. Observing is split over four fields for 1000^h each, with a total bandwidth $B = 8$ MHz, $T_{\rm sky} = (471, 1247)$ K at $z = (10, 15)$, and $N = 5000$ stations centered on a filled core and with an envelope out to $R_{\rm max} = 3$ km. The black squares show the location of the independent k bins. The dotted curve shows the MWA sensitivity at $z = 10$. The vertical dotted line corresponds to the bandwidth; modes with smaller k are compromised by foreground removal.

in both panels. High-precision measurements of the 21 cm power spectrum over a range of physical scales requires $\sim 10^5$ m^2 in collecting area, baselines of a few km, and a wide bandwidth/field of view in order to build sufficient statistics.

4. Science Flowdown: The expected redshift range of reionization determines the frequency range of the instrument. Current constraints suggest the event spans z ~6–11, which translates to ~120–200 MHz and matches the peak sensitivity of the HERA-I arrays well. The central frequency of HERA-II will be determined based on the measurements of HERA-I in order to best sample the reionization era.

Statistical measurements require a high survey speed (etendue): the most relevant simple metric for a low-frequency radio telescope is the collecting area multiplied by the instantaneous survey volume (the latter set by the field of view and instantaneous bandwidth). The increased volume adds more modes to the sample, allowing a better measurement of the power spectrum and other statistics. This field of view requirement suggests a design with a large number of small antennae, which in turn requires a large, fast correlator and is one of the major development efforts of the existing HERA-I programs (see below).

For the low-surface brightness imaging relevant to high-redshift 21 cm emission, the large survey volume is much less relevant; the most important aspect is a nearly-filled aperture with a strong baseline concentration on the interesting scales. For HERA-II, a desire to image over > 20 Mpc scales suggests a compact core of a few hundred meters, and a total collecting area $\sim 10^5$ m^2. Failing to achieve this collecting area will compromise both the imaging and statistical constraints (see Fig. 4).

Theoretical predictions of the matter power spectrum predict that the characteristic scales of features during reionization are ~few–10 Mpc. At z ~6–10, 8 Mpc translates to an angular resolution ~4 arcmin (requiring baselines of just a few km) and a spectral resolution ~1 MHz. Generally, it is easier to improve frequency resolution than angular resolution, so smaller scales are probed primarily through the frequency dimension. However, some more advanced aspects of the 21 cm signal push toward better angular resolution, such as comparisons to near-IR galaxy surveys and measurements of redshift-space distortions

(which separate cosmological and astrophysical aspects of the signal). These will require a substantial fraction of baselines > 1 km.

In general, the most demanding requirements are set by practical considerations. Efficient point source removal during calibration requires some long baselines to achieve good angular resolution. Ionospheric calibration requires a large instantaneous bandwidth (which couples nicely to the survey speed requirement) and a reasonably good time resolution ($\sim$10 s). Foreground removal algorithms require large instantaneous bandwidth and dual polarization response with exquisite calibration. Terrestrial RFI excision requires excellent frequency resolution ($\ll$ MHz) and limited response to sources along the horizon.

B. Other Science Goals

While the focus of this submission is the exploration of intergalactic atomic hydrogen at high redshift, the proposed instruments will have significant capability to serve other scientific needs with modest incremental cost. A primary example is the transient radio sky that is a likely byproduct of image analysis for the deep integrations required to detect hydrogen structures. In A2010 Science WP 176, "The Dynamic Radio Sky: An Opportunity for Discovery," Lazio et al. point out that the time domain has been only sparsely explored. Recent discoveries indicate that there is much to be found on timescales from nanoseconds to years and at wavelengths from meters to millimeters. Unexpected phenomena such as rotating radio transients and coherent pulses from brown dwarfs have already been found, and new examples (and even new classes) of radio transients await discovery, including exotica such as orphan γ-ray burst afterglows, radio supernovae, tidally-disrupted stars, flare stars, and magnetars. Meter-wavelength transients will likely emphasize steep spectrum emission phenomena boosted by correlated particle motions.

In A2010 Science WP 150, Kasper et al. point out that low-frequency arrays are sensitive to both direct radio emission from coherent plasma processes in the solar corona and to the modification of radiation from background sources by the coronal and heliospheric plasma through Faraday Rotation (FR) and Interplanetary Scintillation (IPS). The greatly improved dynamic range, frequency coverage, and bandwidth of modern arrays will open a new window on the physics of magnetic reconnection and particle acceleration at shocks, with the potential to revolutionize our understanding of how CMEs evolve and to predict the severity of their impact at Earth. The leading science questions that the meter wavelength observations can contribute to are: How does the solar coronal magnetic field change with time? How does the coronal magnetic field extend into interplanetary space? How do Coronal Mass Ejections erupt into the heliosphere? How are particles accelerated by CMEs?

The instrumental properties required for 21cm studies, featuring wide field capability, large-N architecture, provision for high precision full polarimetric calibration, large fractional bandwidths, and high sensitivity, inevitably lead to excellent science potential for these and other non-21cm science investigations. Application-specific constraints on capability will lie largely in the digital domain, and the primary HERA implementations will employ, for example, time and spectral resolutions appropriate to 21cm science. It can nevertheless be appropriate for other science groups to contribute funded effort as warranted to augment the HERA digital hardware, firmware and software for their applications. In this document, we limit our focus to the 21cm reionization and dark ages science.

TABLE I: Technical Progression of Hydrogen Epoch of Reionization Arrays

Instrument	Time Line	$A_{\rm eff}$ (km^2)	Band-Width (MHz)	No. Elements	FoV (°)	Goal†
HERA I: PAPER	2010–14	0.003	80	128-512	60	Power Spectrum Detection;R&D
HERA I: MWA	2010–14	0.01	30	512	20-40	Power Spectrum Detection; R&D
HERA II	2015–19	0.1	:30	5,000	20-60*	Power Spectrum Characterization
HERA III†	2020–	$O(1)$	:100	O(50,000)	$O(40)$	Detailed Structure Imaging

*– Field of view depends on detailed trade-off study following HERA-I efforts.
†– HERA-III parameters will depend on experience gained by mid-decade.

III. TECHNICAL IMPLEMENTATION

The study of cosmic reionization using HI requires instruments with high sensitivity over a wide range of angular scales and redshifts, translating to wide field of view, wide bandwidth, and large collecting area. These instruments must further fully characterize the low-frequency celestial sky and model their own system response so that foregrounds can be effectively suppressed and the EoR signal revealed. This requires calibration of the array with high accuracy. Wide fields-of-view, large fractional bandwidths, radio frequency interference (RFI), and ionospheric variation all complicate the calibration of interferometric arrays. Obtaining an accurate sky model requires accurate models of primary beams, receiver passbands, gain variation, and array geometry. Attaining sufficient collecting area places demands on antenna design, correlator size, and data processing and handling.

These technical challenges are being addressed in phases summarized in Table I (above). The development of the technology is driven to advance the science from statistical detection of the EoR fluctuations to full three-dimensional imaging of the high-redshift universe.

HERA I: Power Spectrum Detection. The arrays now operating or under construction have $A_{\rm eff} \sim 10^{3-4}$ m^2 and so are limited to imaging only the most extreme ionized regions (such as those surrounding bright quasars). Nevertheless, these arrays have sufficiently large fields of view ($> 400^{(\circ)2}$) to make reasonably good statistical measurements [3, 16]. Fig. 4 shows the projected errors for the MWA at $z = 10$. The array will be able to detect fluctuations over a limited spatial dynamic range for $z < 12$, constraining the timing of reionization and some source physics.

HERA II: Power Spectrum Characterization. Fig. 4 shows that larger telescopes, with $A_{\rm eff} \sim 10^5$ m^2 (and large fields of view), will clearly be needed for precise measurements, and especially to identify distinctive features of the power spectrum (see Fig. 3). Instruments in this class will also be able to measure more advanced statistics, such as the redshift-space distortions induced by velocity fluctuations. These are extremely useful for breaking degeneracies in the signal [1] but lie beyond the reach of first-generation experiments. (For more information, see the white paper "Cosmology from the Highly-Redshifted 21cm Line".)

HERA III: Hydrogen Structures Imaging. For $A_{\rm eff} \sim 10^6$ m^2, imaging on moderate angular scales becomes possible, and statistical constraints become exquisite even at high redshifts (provided that the large field of view, not strictly necessary for imaging, is main-

tained; see Fig. 4). Plans for these later generations will evolve as we learn more about "dark age" physics and the experimental challenges ahead; for example, the Long Wavelength Array, MWA, and LOFAR will study the ionospheric calibration required to explore the high-z regime ($z > 12$, or $\nu < 110$ MHz) and help determine the relative utility of a terrestrial Square Kilometer Array or a far-side Lunar Radio Array (LRA). At the same time, we must explore whether telescope designs that are intended to be closely aligned with the observables, such as an FFT Telescope [24], are both practical and cost-effective.

The HERA roadmap, accompanied by efforts to improve theoretical modeling of the first galaxies and to enhance data analysis techniques (such as specialized statistical measures) will position the community to explore the major science questions concerning high-z galaxy formation, IGM evolution, precision cosmology.

A. HERA IA

The first-generation experiments must be extremely focused in order to attain the goal of a convincing detection of the EoR signal. We outline here the technical approaches of MWA and PAPER. These experiments focus on redshifts $z < 12$ and the coldest patches of sky. Both have the sensitivity to obtain statistical detection of the power spectrum of 21cm brightness temperature fluctuations and possibly to image the very rare, largest-scale structures formed at the end of the reionization epoch. They represent differing but in many ways complementary paths to addressing the key issue facing the first generation arrays: calibration. Various methods for estimating the effects of instrumental calibration on a statistical EoR detection have been analyzed [2, 17, 19], but achieving the requisite calibration quality and stability in early instruments remains an open problem [3, 8].

PAPER proceeds with an emphasis on hardware, controlling and understanding the individual antenna responses, so that the *a priori* estimates of the instrument response are of the highest possible quality, thereby minimizing the challenge to the calibration software. MWA emphasizes techniques enabled by High Performance Computing, new adaptable algorithms, high data volume, and a breadth of calibration information intended to solve for instrument response and ionospheric distortion on the fly. MWA also points, and picks out preferred regions of cold sky for EoR detection in order to minimize overall system noise. PAPER observes with a simpler transit strategy, and again emphasizes hardware stability and constancy to maximize quality of *a priori* calibration.

1. Murchison Widefield Array (MWA)

The MWA is an 80-300 MHz, large, synthesis array sited in a protected radio-quiet zone centered at Murchison Radio Observatory, Western Australia. The collecting area, field of view (FoV), and imaging characteristics of the MWA are tuned primarily to enable statistical detection of the power spectrum of 21cm brightness fluctuations for redshifts < 12. In addition, the MWA will study the Sun and heliosphere, including direct measurement of the magnetic field structure at different radii, and conduct a systematic sky survey for radio transient emission. See recent image in Fig. 6.

The array comprises 512 tiles of 16 dual polarization dipoles (Fig. 5; 4×4) that are steered electronically by an analog beamformer, achieving a primary beam width of $\sim 30^\circ$ at 200 MHz. This scale is well matched to the sizes of the coldest patches on the sky. Field of view is a primary design consideration of the array. Beam-forming has the added advantages of increased rejection of out-of-beam power from the complex polarized brightness

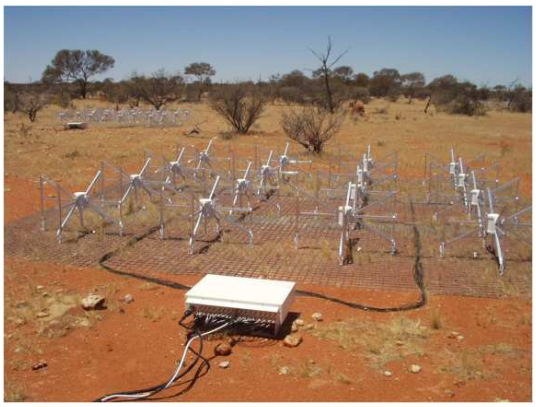


FIG. 5: MWA antenna tile (5m × 5m), consisting of 16 broadband, crossed dipoles with low-noise amplifiers in central hub. Signals are routed to the adjacent analog beamformer; power and coaxial cables to the digital receiver exit at the bottom [13].

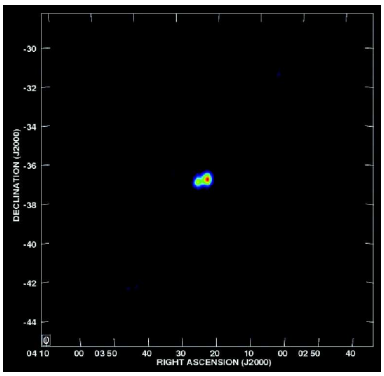


FIG. 6: MWA image of the double-lobed radio galaxy Fornax A at 159 MHz in a 5-minute exposure.

distribution of the low-frequency sky and rejection of RFI sources at the horizon. Digital, multi-bit, baseband sampling leverages recent advances in electronics, in particular Field Programmable Gate Arrays (FPGAs), to achieve uniform sensitivity, stability, and high-enough dynamic range to maintain linearity even in the presence of interference, as from satellites, aeronautical transmission, and reflections within the atmosphere of distant transmissions. Received signals are combined in a 1024-input FPGA-based correlator capable of $\sim 10^{13}\,\mathrm{s}^{-1}$ cross-multiply and accumulate operations. The native time and frequency resolutions of the output are 0.5 s and 10 kHz, respectively. Polyphase filtration enables spectral dynamic range of 10^5:1 and good performance even when faced with strong narrow-band interference. High time resolution enables tracking of ionospheric distortion of the sky images.

The array is centrally condensed within 1.5 km with a maximum extent of 3 km, to provide optimal sensitivity to the reionization signal (for the chosen A_{eff}) over a range of angular scales on the sky and over a range of redshifts. The synthesized beam is $\sim 2.4'$ at 200 MHz (half-power width). Dense sampling of spatial frequencies, enabled by the large number of apertures and full cross-correlation architecture, provides very good full-polarization point-source response, which is a necessity for characterizing and removing foreground contamination. The MWA is distinctive in its use of a once-through real-time calibration and imaging pipeline as opposed to iterative techniques such as self-calibration. The choice is motivated by the 160 gigabit s^{-1} data rate output by the correlator and practical limitations to storage. The pipeline achieves tile beam and ionospheric calibration using $O(10^2)$ simultaneous calibrator sources, subtracts strong sources from the data stream, constructs full-Stokes images, and and resamples to a storage frame in advance of time averaging every few to 10 minutes.

2. Precision Array to Probe the Epoch of Reionization (PAPER)

PAPER is a 100-200 MHz transit synthesis array sited at the Murchison Radio Observatory (MRO), with a prototype station in the NRAO radio-quiet zone in Green Bank, West

FIG. 7: PAPER antenna. The ground screen structure includes side reflectors that narrow the size of the primary beam to more closely match the size of colder patches in the synchrotron sky.

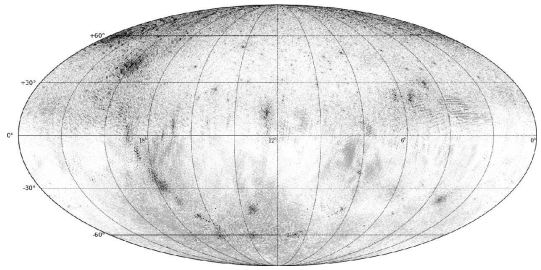


FIG. 8: All-sky map in a band between 138.8 MHz and 174.0 MHz. The northern hemisphere was imaged using PGB-8 data (Dec 38.5°) and the southern hemisphere with PWA-4 (Dec −26.7°) data. Scale log[Jy]: peak 10^4 Jy; min rms 100 mJy. [22].

Virigina. PAPER antenna elements are dual-polarization sleeved dipoles mounted above grounding structures with side reflectors (Fig. 7). These elements have been designed for smooth spatial and spectral responses to facilitate calibration. The primary beam has a FWHM of 60° at 150 MHz. The PAPER analog signal path flows from crossed dipole elements attached to a Pseudo-Differential Amplifier, through coaxial cable that runs above ground to a Receiver Card, which band limits signals to 130-185 MHz before transmitting them to the correlator.

A sequence of real-time, digital, FX correlators employing FPGA processors is addressing the growing digital signal processing (DSP) needs in progressively more ambitious PAPER deployments. These correlators are based on the flexible architecture described in [21], whereby DSP engines transmit packetized data through 10-Gbit Ethernet links to commercial switches that are responsible for routing data among boards. This architecture, analog-to-digital converters, modular FPGA-based DSP hardware, and software environment for programming, debugging, and execution were developed in collaboration with the Center for Astronomy Signal Processing and Electronics Research (CASPER) at the University of California, Berkeley. The hardware and firmware developed by CASPER emphasize modularity and scalability, and the flexibility of the correlator design shortens development time, allowing correlators of increasing scale to be developed in parallel to the rapid incremental build-out of the array.

PAPER has conducted a test field deployment at MRO (PWA-4) and maintains a site in Green Bank (PGB-4, 8, and currently 16). This allows prototyping to take place quickly under realistic field conditions. The PAPER approach to calibration has been to emphasize precision, with the goal of characterizing system components to within 1% in order to facilitate further model refinement via sky modeling and self-calibration. The PAPER antenna elements are easily re-deployed on their uniform length coaxial cables to test various array configurations.

B. HERA-IB

HERA-IB, represents that phase of HERA-II construction that bridges the tried technologies of PAPER and the MWA with those of HERA-II. Although a full HERA-II prototype is not envisaged, HERA-IB should be considered a testbed deployment of technologies that informs the design, scalings, priorities and schedule of HERA-II. This process is essentially a stage of HERA-II risk mitigation and will demonstrate every element of the HERA-II system individually, in the context of an operating 512 MWA or PAPER array. Tasks will include but are not limited to:

- *PAPER and MWA calibration hybridisation.* This tests the capability of the Real Time System (RTS) to calibrate PAPER data, initially post real-time, but eventually through the MWA pipeline. Conversely captured MWA data can be processed with the PAPER pipeline with a view to combining the knowledge gained into algorithms and methodologies applicable to HERA-II.
- *Real-time processing of PAPER data.* Using the RTS and Real Time Computer (RTC) assets of the MWA to calibrate and image PAPER data in real time. This will demonstrate the flexibility of the software calibration system and highlight any MWA specific algorithms, resulting in a more general RTS pipeline.
- *Hybrid MWA/PAPER-dipole operation with real-time calibration and imaging.* A full hybrid system using both MWA and PAPER technologies, DSP and RTC/RTS assets operating in concert, resulting in a general DSP, calibration and imaging pipeline.
- *Incorporation of HERA-II technologies.* Once developed the hybrid capability will permit the incorporation of new DSP technologies, RTC capability and antenna development. New assets can be dropped-in as they are tested and developed.

This platform will provide a testbed on the scale of a 1-10% HERA-II with minimal additional hardware costs. This represents a considerable hedge against HERA-II technology risk as we will develop all aspects of the HERA-II system within a framework of a fully functional existing telescope.

C. HERA II

The larger second-generation experiment (Table I) will be focused on more detailed statistical analyses of the reionization signal. Technical elements will build on experience with systems already in the field. Some will be production elements in use for scientific data acquisition; others will be development prototypes integrated into the arrays for testing and characterization. The central science of HERA II will be power spectrum characterization, the study of redshift-space distortions, pre-reionization cosmology, and early imaging of discrete structures.

Drawing upon lessons learned during HERA-I activity, HERA II will combine PAPER and MWA approaches and technologies (e.g., high correlator data rate will motivate real-time calibration and imaging as in MWA, but with some iterative elements to improve characterization of the instrument). Calibration purity and image dynamic range will be critical figures of merit (in addition to $A_{\rm eff}$ and FoV). Order of magnitude improvements will be sought to advance foreground subtraction over that achieved with HERA I.

Point-source subtraction will be fostered by likely incorporation of baselines up to the limit ($\sim$3 km) at which ionospheric distortions cannot be approximated by simple refractive shifts and perhaps minor second-order perturbation (focusing and de-focusing). The long baselines will suppress the confusion limit while denser sampling of baselines will reduce sidelobe contamination that also limits the detection floor. (At 150 MHz, there is $\sim$1 source per square degree with flux density greater than 1 Jy, growing more than linearly with the inverse of flux density.) The nature of the faint source population at low frequencies is currently unknown, beyond the scope of HERA I, and will constitute part of the HERA-II discovery space.

Accuracy of instrument polarization calibration will require careful engineering of antenna structures, a development activity to be conducted as part of HERA I operations. Post calibration purity will need to exceed that of HERA I in response to the need to improve subtraction of the intrinsically linearly polarized, diffuse Galactic synchrotron radiation, which has an angular spectrum falling as ℓ^{-3} in other wavebands. The polarization spectrum at low frequencies is at present not well constrained [10, 25], and so predictions of contamination from leakage are correspondingly uncertain. This too represents HERA-II discovery space. Achieving high image fidelity for this diffuse emission will require dense baseline coverage over the full range of baselines (in contrast to MWA where the longest baselines are supported by outrigger tiles) and provide constraints on missing polarized flux due to the minimum spacing between elements.

D. HERA III

At present, the technical outlines of HERA III may be defined at the level of published general specifications for the low-frequency component of the SKA. The technical elements of the system will be described in depth as a result of experience with operations and scientific campaigns using HERA II, as well as the development work that will be done using that array as a testbed. Several groups [4, 9, 12, 20] have shown that for correctly chosen antenna distributions and careful observing strategy, foreground removal even to the level of the tomography planned for HERA III is possible. HERA II will be suitable to implement and test these theoretical foreground subtraction schemes on very large volumes of data obtained with systematics that approximate the low level anticipated for HERA III.

E. Array Telescope

1. Telescope. The collecting area of the array for HERA II is on the order of 100,000 m^2 covering the spectral range from 1.5-3.0 meters (100-200 MHz) . Each element of the array consists of one or more antennas that provide the desired field of view (FoV) of approximately 30 degrees. A very large number of antennas will result from this basic requirement. [angular resolution ¿ max baseline length, distribution, planarity?] An important step in the development of HERA II is to engineer a cost-effective station concept where the element configuration is optimized for EoR science. This is a bound optimization problem, but one that has yet to be explored in detail. The antennas and corresponding station configuration would be a new design that is based solidly upon lessons learned from both PAPER and the MWA (and some ideas from LUNAR), and combines the best of both approaches to bound instrument systematics wherever possible to ease the calculation burden and improve calibration. The number of elements will range from 5000 MWA-like elements to 15,000 PAPER-like elements.

2. Lifetime. A likely scenario is that HERA II will evolve and make use of some of the infrastructure of a HERA I instrument. The lifetime of the HERA II facility is on the order of 5-7 years and would be retired when HERA III becomes operational. The array would be dismantled and the site returned to its native use at that time. HERA III, in turn, may serve as a core of SKA-Lo. In all phases there will be reuse of reuseable components from one generation to the next.

3. Technical maturity. The two areas of technology at the heart of HERA II are digital signal processing and the antenna configuration. The antenna configuration that forms the array element will consist of a phased array of wire antennas that are optimized for the FoV, power pattern smoothness over the operating band, dual polarization, and long-term stability. The rugged low noise amplifiers will be integrated together with the antenna structure for optimum sensitivity of the band. While ongoing significant development efforts are warranted, success of this array does not hinge on the development of advanced technologies.

4. Array element. The antennas array elements will be bent-wire structures that produce the required gain (1/FoV) over the operating band. These are very light-weight structures that reside over ground planes located near the Earth's surface. The materials used in the fabrications of these antennas consists of metals (aluminum, copper, brass) and common dielectrics (PVC, PS). The antennas can be easily reproduced with a modest investment in rudimentary construction jigs by the manufacturer. The are either zenith pointing or can be made to track over a limited range using controlled phasing techniques. The pointing accuracy is limited only by calibration requirements

The antennas have no moving parts. The beam of each array element would be steered by way of electrically delaying the signals from each antenna just prior to the summing junction that forms the beam.

6. Technical maturity areas. There are only two areas of moderate technical maturity: the correlator and the antenna element. In both cases, small N versions of these components either have already been developed (MWA and PAPER) or will be characterized as part of the HERA-II development plan.

7. Risks. The only significant risk to cost, schedule, and performance is the availability of a correlator capable of meeting the large-N requirement over the operating bandwidth.

8. Construction. The antenna construction is straightforward and will not require any special construction techniques.

9. Instrumentation. The instrumentation is rather straightforward radio frequency components and modern digital signal processing hardware. No special instrumentation is required.

10. Non-US participation. A central aspect of the implementation that requires non-US participation is the observing site having ultra-low radio frequency interference potential. Sites in Western Australia and South Africa meet this requirement. No site in the US has levels of radio interference low enough to meet this requirement. In our costing we make the assumption that the host country will provide significant resources for basic remote site infrastructure (power, internet, central lab facility).

F. Facility Construction

1. General description. The instrument requires a site of circular extent with a diameter of about 5 km. Altitude is not an issue, but the site must reside in a ultra-low radio

TABLE II: **HERA II Characteristics Table**

Antenna Array	Value	Units
Main and Effective Aperture Size	20 ($\propto \lambda^2$)	m^2
Size of Array Elements	25	m^2
Number of Array Elements	5000	—
Total Collecting Area	10^5	m^2
Angular Resolution	3 ($\propto \lambda$)	arc-minute
Field of View	30 ($\propto \lambda$)	degrees
Wavelength range	1.5 - 3.8 (200-80)	m (MHz)
Driving Wavelength for Accuracy	1.5 (200)	m (MHz)
Required Surface Accuracy (ground screen)	0.1	m
Number of Reflecting Surfaces	1	—
Total Moving Mass (Earth)	6×10^{24}	kg

interference environment [preserved over the lifetime of the instrument]. It should be level such that the array can be planarized to within a meter. The site should be accessible by truck. AC power can be provided by mains or portable units (solar, diesel, etc.) on the order of 1 MW. Reliable Internet access with speeds >1 Mbps is required for monitor & control (data will be transferred by removable hard drives). A central building of 500 sq ft with appropriate HVAC will be needed to house the back-end electronics. A laboratory building of about 500-1000 sq ft located a few km from the site would also be needed. On-site housing (about 6 persons) with restroom facilities should also be available. Basic communications (satellite phone or a terrestrial link) is required for site safety.

2. Electrical requirements. AC power on the order of 1 MW will be needed. No additional requirements are imposed during the construction period; actually, less power is needed during construction.

3. Infrastructure. The array elements will communicate with the central instrument hut via fiber optic cables. The antennas will be positioned directly on the ground – no concrete pad is needed.

4. No comment

5. Construction methods. No new or unique construction methods will be required. The PAPER group has been thinking about burial or stealthing of centrally located equipment hut to mitigate reflections off large metal surfaces.

6. Construction management. No details of the management plan can be provided at this time.

G. Observation Strategy

1. Observation Strategy. HERA observations will involve long observations of cold regions of the sky. Observations with HERA II will be restricted to the night time, and to times of the year when the center of the Galaxy is below the horizon at night. With these restrictions, 2000 hours of observations per year of the southern sky are possible. We anticipate that 1000 hours per year will be sufficiently high quality (discarding times of poor ionospheric conditions and interference) to use in the long integrations required to study the 21 cm signal. Theoretical forecasts (e.g. Bowman, Morales and Hewitt 2007; Mao et al. 2008; McQuinn et al. 2006) have shown that 1000 hours of integration on a single field will yield high precision 21 cm power spectrum measurements. Hence, the science goals may

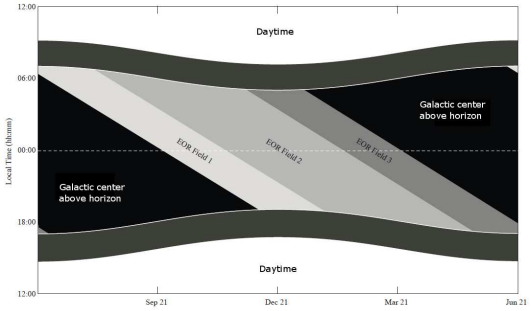


FIG. 9: Allowable observing schedule for HERA. Only the diagonal gray regions labeled as "EOR field 1-3" meet the requirements that both the sun and the Galactic center are below the horizon, providing a total of ∼2000 hours per year. Primary observing campaigns will commence in September and end in March each year. The three shades of gray illustrate how the observing time would be divided for three hypothetical target fields spaced by 6 hours in right ascension.

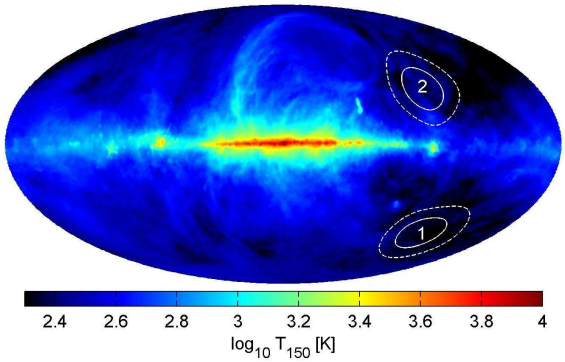


FIG. 10: The primary (1) and secondary (2) target fields planned for the MWA are centered at $\alpha, \delta = 60°, -30°$ $(\ell, b = 228°, -49°)$ and $\alpha, \delta = 155°, -10°$ $(\ell, b = 253°, 38°)$, respectively. The solid white curves indicate the 50% response power contour of the primary antenna tile beams and the dashed lines show the 10% power contour. Map adapted from Haslam et al. 1982.

be achieved in a single year of observations. The daylight hours and nighttime hours when the Galaxy is above the horizon would be available for secondary science (e.g. heliospheric and Galactic science, respectively). The HERA II EoR archive would also enable a broad variety of scientific studies. Secondary science may be pursued by externally funded guest collaborators.

Atmospheric conditions: Turbulence in the Earth's ionosphere refracts low-frequency radio waves. This acts much like atmospheric distortions at optical wavelengths, and must be corrected using techniques similar to wide-field adaptive optics. The relatively modest baselines (< 3 km) involved for HERA-II require fitting only a refractive shift model and **not** the more complicated 'global self-calibration' techniques that have been proposed to handle wide-field imaging on longer baselines. The appropriate time scales for changes in the ionosphere are of order tens of seconds, hence the HERA arrays will calculate new ionospheric distortion solutions on a cadence of about 10 seconds. The ionosphere is expected to be most stable during the late night and pre-dawn hours. Occasional traveling disturbances are expected to prevent calibration for short periods. The Western Australian and South African sites are at mid-latitudes in the southern hemisphere, where the ionosphere distortions are known to be relatively mild. HERA-I arrays will provide appropriate statistics to assess the impact of ionospheric conditions on the observing efficiency for HERA-II.

2. Engineering Activities: Diagnostic observations to monitor the instrument would be carried out during the day and during a small fraction (1%) of the nighttime hours. Maintenance and repair would require 2 FTEs (see operations section). Calibration is carried out while the science data are taken.

3. Observatory drawings: See Figure 11.

4. Software and Science Development: HERA-II will require real-time calibration and

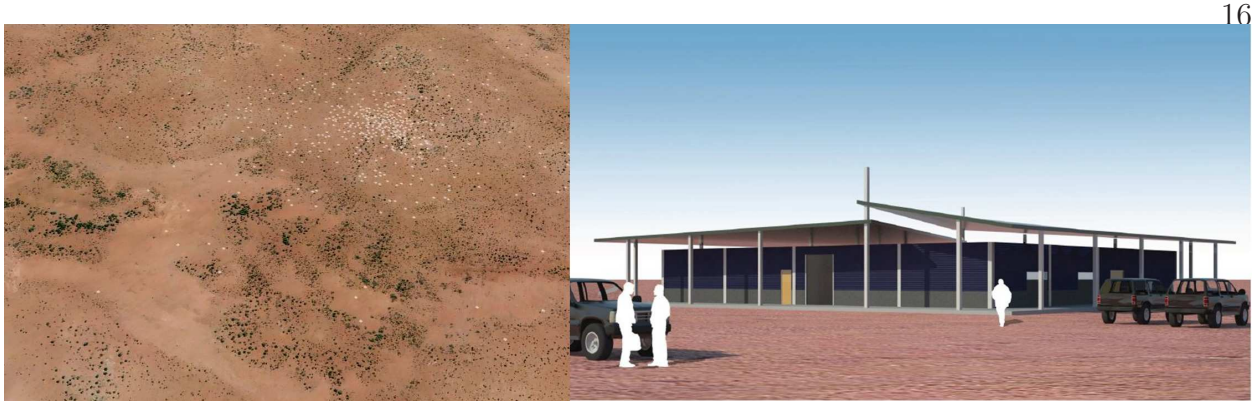


FIG. 11: *Left:* Artist's rendering of the 500 MWA antenna tiles (white dots) positioned at their planned locations at the MRO site. Each tile is 5x5 m and the total diameter of the array is 1.5 km. HERA-II will have 5000 antennas arranged over a diameter of 3 km. *Right:* Architectural rendering of the correlator and real-time computer support building that will be constructed at the MRO in 2010 in support of the MWA and ASKAP. MWA requires 3 rack cabinets of computer/electronics space in this facility. Similar infrastructure will be provided by the host site for HERA-II. *(Architectural rendering used without permission; credit CSIRO.)*

integration in order to reduce data volume. Sky maps and meta-data will be deposited into an archive for off-line analysis. The software required for this process and subsequent data product processing and foreground subtraction is extensive and constitutes a major investment. Software is required for correlation to accommodate the larger number of elements, for real-time data quality monitoring, for real-time calibration, for off-line data examination and selection, for mapping, for cosmological model fitting, and for detection of ionized regions:

- **Firmware**—The large-N correlator and beamformer electronics require custom firmware.
- **Monitor and control**—This subsystem is responsible for controlling the behavior of the end-to-end system from antennas to real-time data processing. It also receives, processes, and logs monitoring information of various kinds, including environmental data, subsystem performance metrics, calibration data and so on. Finally, it provides a user interface for controlling the instrument.
- **Testing and commissioning**—Non-scientific operations software is required to capture (nearly) raw visibility data so that we can refine calibration and mapping techniques. Due to the wide fields of view and high data rates, new engineering data analysis tools are being developed for HERA-I to support the testing, commissioning, and refinement of the instruments. Software for quick visualization of the state of the array is also needed given the hundreds to thousands of independent antenna/receiver paths that must be assessed.
- **Real-time data quality flagging**—The observations will vary in quality, depending on the state of the instrument, the state of the ionosphere, radio interference (RFI), etc. Automatic flagging of bad segments of data is critical so that they can be excluded from the ten-minute integrations stored in the archive since there will be no raw visibilities to recompute corrupted maps.

- **Real-time calibration and imaging**—The calibrator measurement loop (CML) measures apparent angular offsets induced by the ionosphere and the system gain toward known compact astronomical sources across the sky. These measurements are used to fit models of the ionosphere and instrument response, and support subtraction of strong sources that limit sidelobe contamination during calibration. Subtracting the contribution of each source is carried out sequentially, so that the stronger sources are removed before measurements of weaker sources are made. Estimation of calibration parameters is greatly over constrained due to the large number of antenna pairs. The wide-field nature of the instrument and real-time computing requirement pose challenges, however: direction-dependent gain and polarization response, confusion and sidelobe contamination (each field mapped by the will contain hundreds of relatively bright sources), the ionosphere, and real-time data reduction. Many of the promising techniques currently being investigated, such as iterative self-calibration and deconvolution algorithms, and some wide-field imaging algorithms, cannot currently be implemented in real time. They either cannot be used at all or need to be approximated.

- **Data archive interface and management**—Maintenance and SQL access to the multi-petabyte archive database. Support for sifting through the averaged data in the archive to select the best data. Current plans call for as much as two-thirds of the data to be discarded before analysis. This task requires a grading scheme for the data, and hunt through all of the data as it is taken to determine its quality, so cuts can be made later. The edited data sets will be added to the archive, along with a description of the cuts made.

- **Data product production**—It is unlikely that the data archive will be populated with fully calibrated Stokes maps. After a subset of the archived data is selected for an integrated map, some calibration, weighting, and coordinate conversion processes must be applied to produce a fully characterized scientific data product.

- **Foreground subtraction pipeline**—This off-line pipeline operates on the integrated scientific data cubes and associated meta-data and is divided into four principal stages:

  The first step is removal of bright points sources with flux greater than $S \gtrsim 100$ mJy. The real-time calibration system will 'peel' (self-calibrate and subtract) the brightest sources in the field from the raw visibilities. However, it is likely that only the $\sim$100 brightest sources (fewer than 10% of the total number sources above $\sim$100 mJy) can be peeled in real-time. The remaining point sources must be removed from gridded and integrated 'base maps.' A multi-frequency matched filter technique originally developed for use with CMB maps is under development for this purpose.

  The second step is removal of the diffuse continuum emission. The emission mechanisms of these foregrounds produce smooth spectra that follow power-law profiles, making them possible to separate from the redshifted 21 cm emission by their long spectral coherence scales compared to the small, $\sim$ 1MHz spectral coherence of the reionization signal. A variety of techniques have been investigated to accomplish this separation including polynomial fits along the spectral axis of each pixel in a base map or its uv-map conjugate, Minkowsky-functionals, and non-parametric techniques.

The third step of the foreground subtraction pipeline is polarized leakage subtraction. Gain and phase calibration errors and non-ideal feeds conspire to contaminate all Stokes parameters further. This leakage leads to some portion of the complex linear polarization signal finding its way into the Stokes I intensity maps that will be used for the 21 cm measurements. By exploiting the Fourier relationship between the polarized signal and the Faraday dispersion function, it is possible to identify components of polarized signal with specific Faraday depth. This process is known as rotation measure (RM) synthesis.

The final step of the pipeline is statistical template fitting. Many classes of errors and residuals remaining in the foreground-cleaned maps can be modeled given sufficient knowledge of the instrumental response. The statistical properties of these residuals can be exploited to provide an additional layer of foreground removal by fitting templates of the residual statistical structures during the 21 cm power spectrum parameter estimation. This technique is now commonly employed in CMB power spectrum analysis to account for the angular power of faint point sources and the effects of gravitational lensing.

- **Science analysis**—Common cosmological codes (e.g. CosmoMC) will be adapted to support parameter fitting and analysis of 21 cm power spectra. New analysis codes will be developed for 21 cm specific science, such as detection and characterization of ionized voids.

- **Instrument simulation pipeline**—Off-line high dynamic range source subtraction from gridded data products (as opposed to raw visibilities) requires reconstructing the integrated synthesized array beam in the direction of each point source. This process involves replaying the archived calibration meta-data of the entire observation through an off-line simulation of the instrument and the real-time calibration and measurement loop. The MIT Array Performance Simulator (MAPS) and the CML code produced by the HERA groups will be adapted to support this operation.

5. Pointing methods: Dipole antenna arrays have no moving parts, and large fields of view. The HERA-I pathfinder arrays are exploring both tracking and non-tracking observational approaches. HERA-II will adopt the more effective approach. If MWA-like tiles are chosen, azimuth-elevation coordinates will be calculated from commanded ra-dec coordinates, and delays for each antenna element will be computed and inserted in the signal processing path. This technique has been demonstrated with the existing MWA tiles. If a PAPER-like draft scan strategy is chosen, then it is not necessary to point the array.

IV. ENABLING TECHNOLOGIES

A. Antennas

REQUIREMENTS: The calibration requirements for HERA arrays are extremely stringent. At least one bright source with flux density $\gtrsim$100 Jy is expected in each field of view. Such a source produces an antenna temperature of about 150,000 K at 180 MHz in a synthesized beam of a few arc minutes, leading to a dynamic range requirement of better than 10^6. This dictates that each gridded *uv*-cell must have an uncertainty in its phase calibration of order 0.01° or smaller. For long integrations, each gridded *uv*-cell will be formed from the average of many visibility measurements, and the averaging of random phase errors relax the calibration requirement on a single visibility to more than 10°, which is not a challenge at these frequencies. *However, systematic antenna phase errors, that do not average down, have to be controlled at the level of the* 0.01° *requirement.*

SIGNIFICANCE: Critical. Foreground subtraction is *required* in order to enable 21 cm astrophysics and cosmology measurements.

ACTIVITIES TO DATE: Two antenna designs are being tested as part of the current phase of HERA-I activities: tiles of crossed 'bowtie' dipole arrays (for MWA) and individual crossed sleeve dipoles (for PAPER), each mounted above a ground plane structure. Prototype implementations of both designs have been constructed and tested in laboratory and field settings. The MWA has deployed 32 complete antennas for testing in Western Australia. PAPER has deployed a 16-dipole array at Green Bank. To date, both antenna designs have performed as expected (within the measurement limits) and both HERA-I arrays are progressing with plans to deploy full arrays based on the existing antennas. Nevertheless, it is very difficult to test low-frequency antennas to the level of precision required and final confirmation of the antenna calibration performance may not come until the completion of the HERA-I phase.

MATURATION PLAN: The technology development for HERA Phases II & III must include several years of study of the performance of the different antenna prototypes and array configurations as they are deployed in the field, observing wide fields dominated by confusing bright sources, and with the data processing carried through the full pipeline of calibration, imaging, and foreground subtraction. These studies started as part of detailed HERA-I pathfinder design, prototyping and simulations. They will continue as a vital part of Phase-I activities that require exact understanding of instrumental effects, and, as a result, make us ready for informed and accurate costing of future phase instruments.

A third class of antenna, units consisting of four phased helical antennas, are being prototyped as a part of design studies for the Lunar Radio Array (Lazio et al. science & program white papers). MWA and PAPER are complementary in their approach to control systematic errors in calibration; the Lunar Radio Array (LRA) approach is similar to that of PAPER in that the antenna element has been designed to have a single-lobed beam that varies smoothly with frequency and angle. Investigation of dipole and helical antennas addresses the open question of the effect of the polarization basis (linear for MWA and PAPER, circular for LRA) on achieving the required dynamic range in the presence of polarized foregrounds.

B. Array Configurations

REQUIREMENTS: Achieving the surface brightness sensitivity for precision characterization of the redshifted 21 cm power spectrum dictates a compact array of many antennas. Such configurations are susceptible, however, to correlated noise (cross-talk) and other systematic problems. In addition, calibration and foreground subtraction requirements compete with the 21 cm science requirements and tend to drive the array configuration to include a large number of long baselines ($\sim$3 km) in order to accurately localize point sources and reduce contamination from diffuse Galactic emission. Deploying thousands of antennas is a new frontier in radio astronomy.

SIGNIFICANCE: Important. Both HERA-I and HERA-II arrays depend on the sensitivity from a compact array configuration in order to reach their 21 cm science objectives.

ACTIVITIES TO DATE: Sensitivity studies for the MWA have shown that SNR in the 21 cm power spectrum can be improved over an order of magnitude by condensing most of the antennas into to the central core region. However, it has yet to be demonstrated in practice that a compact array configuration can be calibrated. PAPER is exploring uniform-coverage configurations that seek to minimize the number of redundant wavemodes sampled by an array in order to improve calibration and minimize sample variance in power spectrum measurements. For both experiments, progress is underway for characterizing galactic synchrotron and point-source foregrounds to understand how array configuration can be adjusted to reduce contamination.

MATURATION PLAN: An optimal configuration for 5000 antennas that balances calibration and science requirements will continue to be explored as we learn more from the HERA-I arrays. RF and electronic aspects of this development are coupled to the antenna and receiver hardware designs. Calibration and foreground solutions encompass the entire measurement pipeline and set this problem apart as a new and unique enabling technology. We anticipate HERA-II array configuration development will rely heavily on detailed instrument simulations based on measured properties of HERA-I arrays.

C. Large-N Digital Signal Processing (Correlator and Real-time System)

REQUIREMENTS: HERA-II will require a 10$\times$ increase in the number of antennas over the HERA-I arrays. The correlation schemes adopted by all existing large radio interferometers, including MWA and PAPER, have DSP costs that scale as N^2, for $N \gtrsim 100$. Developing technologies for building and operating large-N correlators at reasonable cost is critical. Calibrating and imaging the correlator output also poses a substantial computational burden at low frequencies. For HERA-II, data rates will prevent the traditional path of raw visibility storage followed by off-line post-processing. Real-time calibration and imaging will demand digital processing comparable in complexity to correlators.

SIGNIFICANCE: Critical in order to reach HERA-II cost objective.

ACTIVITIES TO DATE: Both MWA and PAPER have implemented prototype FPGA correlators that meet the cost requirements of the HERA-I phase. PAPER has deployed a 16-element correlator at Green Bank; a 64-element system is scheduled for 2010. The MWA prototype correlator is scheduled to be deployed with the 32-antenna system in September,

2009, with full 500-element correlator online in 2010. The MWA has developed preliminary real-time calibration and imaging software designed to run on a GPU-enabled computer cluster. This software will be tested along with the MWA correlator beginning in September. Significant research has been devoted to developing computing architectures that are scalable to the scale of a HERA-II array.

MATURATION PLAN: A plan for the development of large-scale DSP computers must be informed by the 2-year doubling in capabilities that digital electronics have sustained for ~ 50 years. With the budget allocated for correlation and calibration hardware growing linearly with N, HERA-II computing systems will have expanded in scale faster than Moore's Law is expected to lower the cost and density of computing hardware compared to HERA-I. Specific technical research and development to minimize the increase in cost will be addressed as part of the HERA-II development program (with benefits to other radio and physics applications):

Scalable computing architectures will be explored, as will be modular hardware components that can be quickly updated to employ the latest commercially available processors. HERA-I FPGA hardware will be updated to next-generation commercial processors. Developments in low-power electronics such as those being carried out for the space program may have implications for the design of large DSP processing systems for HERA-II. For correlators, each processor must receive data from every antenna.

For large numbers of antennas, providing cross-communication on this scale is a challenge. Switch-based correlator architectures appear promising [21] and may prove viable for large-N arrays should switch capacity continue to follow the growth of the internet. Real-time calibration algorithms must be further explored. Time-variable ionospheric conditions require that calibration solutions be computed over short time intervals before averaging. Research must address real-time calibration, wide-field imaging, and architectures for implementing these algorithms.

All of the above areas of technical development overlap with priorities for the larger Square Kilometer Array (SKA) program, albeit at small bandwidth per N. While this exploration is currently being led by low-frequency array developments, there will be significant potential for HERA-II to share research and engineering costs with SKA-mid projects, and with the open-source DSP community (e.g. CASPER). We also note that, separate from HERA-II development activities, alternative aperture synthesis architectures should be explored. Gridding electric field data before it is presented to the correlator so that FFT techniques can be employed to change the scaling of computational complexity in correlators from N^2 to $N \log N$ for compact configurations would be a massively enabling technology moving beyond HERA-II. The FFT Telescope [24] and the MOFF correlator [18] are two examples of architectures that merit further exploration.

Informed by the current development time of DSP computing hardware, processors will be selected for the HERA-II correlator and real-time calibration system during 2013. Hardware targeting these processors must be developed over the following 3 years, leading to the final HERA-II implementation ready for deployment beginning in 2016.

V. FACILITY & SCIENCE OPERATIONS

1. Brief description. The HERA telescopes are comprised of large numbers of fixed dipole antennas, with no moving parts that feed digital hardware subsystems both in the field and inside a central building. The array hardware is inherently massively parallel, both in antennas and frequency channels, and as a result is resilient against hardware failures of various kinds. Such failures cause gradual performance degradation, and allow for a scheduled maintenance strategy, rather than a reactive one.

The total MWA-512 power budget is of order 100 kW. For the same ($\sim$30 MHz) bandwidth as MWA, we can expect a per-tile power consumption of $\sim$50 Watts. The cross-correlation operation and subsequent calibration/imaging can, for the current purpose, be assumed to scale as the square of the number of antennas. The MWA, with 500-antenna tiles and 30-MHz bandwidth, performs full cross-correlation in a $\sim$10 kW power budget. (The computer receiving the correlator output is projected to consume up to 40 kW.) A HERA II design with $\sim$5000 antennas and the same 30-MHz bandwidth would need $\sim$1 MW for correlation using 2006 (Virtex 4) technology, but only 100-200 kW using projections for 2014 technology.

While HERA arrays are focused on EoR science, they will inevitably be powerful instruments for other science applications. We therefore assume that they will be operated continuously, even at times unfavorable for EoR observations such as during the the day, and when the galactic plane is high in the sky. For such continuous operation at the power levels projected for a 5000-antenna HERA II system, on- site diesel generation at current fuel prices would cost \$1-2M/year. Such costs would be lower by a factor of several if grid power were available.

3. Challenging operational constraints. None acknowledged. As mentioned above the array is static. There are concerns about fauna – cable chewing and antenna collisions have both been experienced with PAPER, MWA and LWA. While planned deployments are in near desert locations, there will be storms, flooding, lightning and even static charge buildup.

4. Science operations. The science operations team will work with the facility operations team and will execute the observations of the EoR fields. As described above, observations will be in campaign mode, collecting survey data for deposition in an archive. During the six-month EoR observing season, two people will be required to plan and conduct the night-time EoR observations; i.e., one FTE overaged over the year. In addition four FTE's will be required for data verification, preliminary processing, and maintenance of the archive. Therefore, a total of five FTE's are required for the science operations team. This team will be distributed among the partner institutions, maintaining close contact with the facility operations team via internet and telephone. Development and coding of software will take place at the partner institutions; the science operations team is responsible for implementation only. Maintenance and engineering will occur during the day and during the time of the year that is not set aside for EoR observing.

Other science may be carried out on a time available basis, with resources provided by the guest science teams.

5. Data archiving plan. The archive will provide limited visibility data, unprocessed time-averaged sky maps, and processed sky maps to the collaboration and, when the data are made public, to the astronomical community. To be developed:

- software that will provide off-line evaluations of data quality

- software to select visibility data according to criteria set by the user
- software to select averaged maps according to criteria set by the user
- software to carry out routine corrections for instrumental effects
- software to carry out routine foreground subtraction

Algorithms will be developed and coded by the science teams at the partner institutions. Archive personnel will ingest this software and incorporate it into the archive suite of software. We anticipate that four FTE's for four years will be required for development, dropping to two FTE's once operations begin. Methods from one generation will frequently serve as foundation for the next generation.

Deployment will require the purchase of a large storage system, which we estimate now at six Petabytes as discussed below (this will need to be re-evaluated closer to the time of purchase as prices are likely to change significantly between now and then). We anticipate that five FTE's will be required for deployment and maintenance of the archive beginning six months before operations begin and continuing through the lifetime of the facility.

6. Science data products. The data products will be:

- Limited *uv* data - during each observing seasons samples of uv data will be taken for use in algorithm development by the science teams. The amount will be set by the level of resources available for this purpose. We anticipate 10 TBytes per observing season, to be deleted at the beginning of the next observing season.
- Unprocessed time-averaged maps - the maps will have two spatial dimensions that depend on the resolution of the array chosen at the time of PDR. We anticipate 2048X2048 maps, four polarization components, and a weight for each pixel. If compression reduces data volume by a factor of two, each map will be 32 GBytes; 500 Terrabytes of storage for each year of observing will be required for map integration times of eight minutes.
- Metadata describing the state of the instrument and environment at the time of the observations, and statistical summaries of the data. Our experience with the MWA has taught us this will be equal in volume to the map data, or another 500 Terrabytes for each year of observing.
- Foreground-subtracted maps. These will be the same size as the unprocessed maps, or another 500 Terrabytes for one year of observing.

Therefore, each year of observing will require 1.5 Petabytes of data storage. For a four-year lifetime of the instrument, 6 Petabytes will be required in total.

7. Maintenance. We envisage no special constraints, special communications requirements or support requirements from other facilities, nor are there unusual or challenging operational constraints. Considering the physical scale of a HERA II installation, we expect the array to be relatively simple and inexpensive to operate, given the solid state design and lack of mechanical systems, remote operations being standard, and personnel support available within a few hours of the site (taking the Australian site as a model).

8. Data volume/day. For the EoR observing mode with eight-minute maps, each day will produce 12 hours X 7.5 maps/hour X 32 GBytes/map = 2.88 TBytes/day. This number must be tripled to account for the metadata and the processed maps, giving 8.6 TBytes/day.

8. Data volume/year. See above. 1.5 Petabytes.

10. Science/Operations Center. The science and operations center for the EoR science will consist of the five members of the science operations team and the five members of the archive team. Physical space will be provided by the partner institutions for housing storage facilities and computers. The teams will be located at institutions yet to be determined; the institutions will provide office space. Provision of space for HERA will be supported through the institutions by way of the usual overhead mechanism for recovery of costs.

The archived data will be made available to the community by way of web access.

A three-year program of first science by the project's science team will be funded directly by the activity. This program will support eight senior investigators (summer salaries), four research scientists, four postdocs, and eight graduate students for a total cost of approximately $2.8M per year.

There will not be a guest observing program. We recommend that an outside investigator program (for analysis of archival data) of $3M/year be funded and administered by the NSF.

VI. PROGRAMMATICS & SCHEDULE

A. Schedule

2009-2011—HERA IA: The decade begins with the completion of PAPER-128 and MWA-512 in 2010. Commissioning and operation of these pathfinder arrays in 2010-2011 will test the hardware and algorithm design concepts needed for 21 cm power spectrum characterization. This process will be conducted with current sources of funding and yearly follow-on NSF support pursued via AAG, ATI, and possibly URO programs, interdisciplinary initiatives such as CDI, and opportunities through CISE or ATM divisions.

2012-2014—HERA IB/II(Development): At the 2011 NSF opportunity, a new round of substantial instrumentation funding will be sought in order to conduct the design and development work for HERA-II between 2012 and 2015. During this period, MWA and PAPER will continue scientific operations and will also provide testbed facilities supporting development work on HERA-II. Priority instrumentation development areas will focus on new implementations of antenna elements for improved calibration, updated prototype correlator boards with next-generation Virtex 7 (or higher) FPGAs, and real-time computing solutions—all building on lessons from MWA and PAPER. Priority algorithm development areas will be astrophysical foreground characterization and subtraction, real-time processing techniques, and imaging dynamic range. Conceptual and preliminary design work on HERA-II will proceed as outlined in Tables III and IV. MWA and PAPER will be decommissioned at the end of this period.

2015-2021—HERA II (Construction and Operation): Funding to finalize the design and initiate construction of HERA-II will be pursued in 2014. With this funding beginning in 2015, the detailed design and prototyping process for the array will be completed by the end of 2015. Construction is expected to commence immediately thereafter and last for 24 months, finishing by the end of 2017. Commissioning of HERA-II is allocated 6 months, yielding science operations by 2018 Q3, in time for the 2018-2019 observing window. At least three annual observing campaigns are planned for the array, giving HERA-II an estimated lifetime lasting into 2021 or later. Tables III and IV list these milestones.

2020 and beyond—HERA III: Development to achieve 1 km^2 collection area over 100 to 200 MHz will be driven largely by reionization science, focusing on high-order statistics in HI power spectra and direct wide-field imaging. We anticipate this effort to commence following the successful operation of HERA-II for one observing season. Design and construction of this large instrument will likely last 5 years, hence its operation is expected to commence in the early- to mid-2020s. Once basic design and operation is secure, transformation over time into a more general-purpose, SKA-type facility is anticipated, with the addition of capability from other communities.

B. HERA-II Organization & Partnerships

PAPER is an NSF-funded collaboration among scientists at UC Berkeley, UVA, NRAO and UPenn in the US and at Curtin University in Perth.

TABLE III: HERA-II Key Phase Durations

Phase	Months
Phase A (HERA IB) - Conceptual Design	12
Phase B (HERA IB) - Preliminary Design	12
Phase C (HERA IB) - Detailed Design	12
Phase D (HERA II) - Construction	24
Phase E (HERA II) - Integration & Test	6
Phase F (HERA II) - Primary Operations	36
Start of Phase B to First Light/Signal Reception	48
Start of Phase B to Start of Operations	54
Project Total Funded Schedule Reserve	ND

TABLE IV: HERA-II Key Event Dates

Phase	Milestone Dates
HERA-II development funding commences	2012 Q3
Start of Phase A	2013 Q1
Start of Phase B	2014 Q1
Preliminary Design Review (PDR)	2014 Q4
HERA-II construction funding commences	2015 Q1
Critical Design Review (CDR)	2015 Q4
Date of End of Construction	2017 Q4
Date of First Light/Signal Reception	2018 Q1
Operational Readiness Date	2018 Q3
End of Operations	2021 Q4
Continued Data Analysis	2023 Q4

MWA is a collaboration among researchers at (i) Haystack Observatory, MIT, SAO, and Harvard University, each supported by NSF and institutional funds; (ii) a consortium of six Australian universities; and (iii) the Raman Research Institute. CSIRO provides logistical and select engineering support. Management is centered at the International Center for Radio Astronomical Research (ICAR) in Perth.

Murchison Radio-astronomy Observatory (MRO) is established and operated by CSIRO, in collaboration with Australian federal and state governments. Both PAPER and MWA operate at the MRO and receive support from CSIRO staff. We anticipate that HERA-II and HERA-III will seek to continue operating at MRO, although this is contingent on the findings of MWA and PAPER and on outcome of the SKA site selection.

HERA-II will be formed by joining the existing members of the PAPER and MWA collaborations in 2012 at the onset of HERA-II development activities. A board-governed consortium will include host-nation organizations and other overseas partner institutions. The Board will work under a memorandum of understanding and coordinate sponsor organizations. Near-industrial scale design, replication of hardware elements, and signal-processing performance will mandate formal project engineering and management. The organization structure that will be enacted is shown in Figure 12. Management responsibility will lie in a geographically distributed Project Office responding to the Board. Positions will include Director, Sponsor Liaison, Principal Engineers, Project Manager, Project Scientist, and Regional Project Managers and Engineers. Integration of regional positions will ensure communication, foster common purpose, and enact lines of spending authority and accountability that parallel funding lines from different sources.

C. HERA-II Risks

1. Radio Frequency Interference is present

Probability: Certain; Consequence: Minor

RFI may be generated within astronomical hardware or by external human activities. Good engineering practice mitigates the risk of internally generated RFI, informed by experience in design and deployment of quiet hardware at MRO during the 5 years prior to HERA-II construction. The MRO site is among the most radio quiet sites in the world, with legislated, regulatory control over transmitters, per ITU specifications, for distances up to $\sim$100 km.

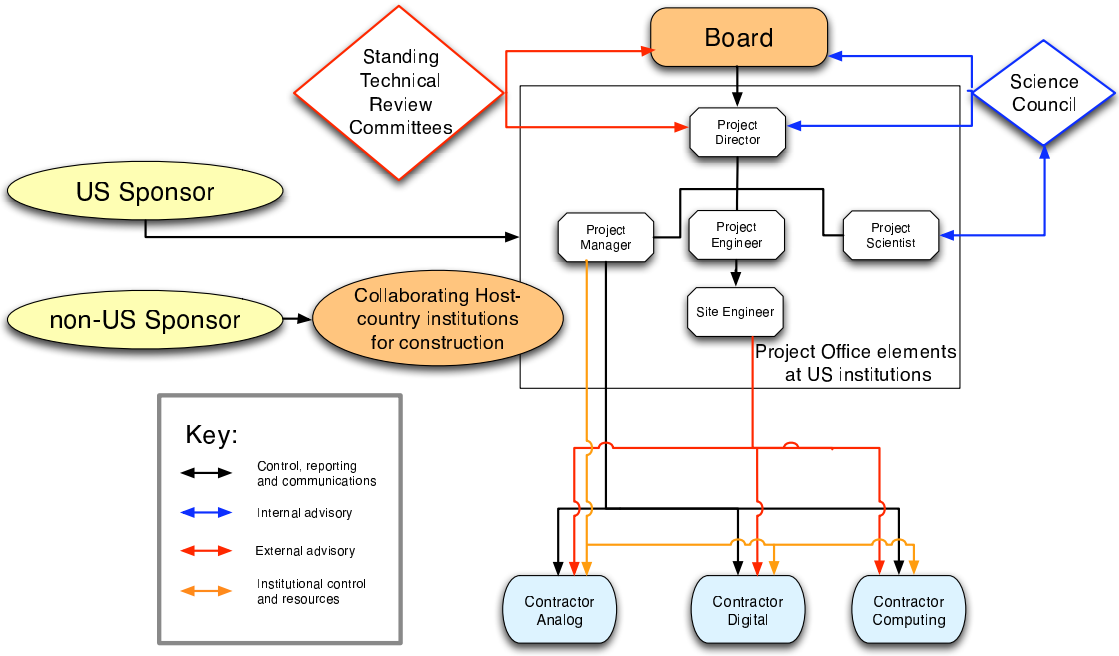


FIG. 12: HERA-II project management organizational chart.

TABLE V: Summary of HERA-II Risks

Risk	Likelihood (1-5)	Severity (1-5)	Score	Mitigation	Contingency
1 - RFI	5	2	10	Radio quiet site. Good engineering practice informed by 5 years experience at MRO	High dynamic range sampling, extensive flagging, filtering of severe narrow band RFI
2 - Limitied Antenna Performance	4	4	16	MWA & PAPER online in 2010 for high-precision test of antennas	Develop new response parameterization during 2012-2015.
3 - Expensive Large-N Correlation	3	4	12	Use late generation FPGAs. Existing Virtex 6 series already one order better than those used in MWA.	Replace $1 \times N$ element array by M^2 arrays of N/M elements.
4 - Expensive Real-Time Calibration and Imaging	1	5	5	Calibration task better constrained than MWA & PAPER. Enlarged compute capacity	Can process a subset of baselines to reduce load in calibration.
5 - Ionospheric Calibration Limits Dynamic Range	5	4	20	More collecting area enables more calibrators and better constrained correction.	Observations at most stable times. 5 years experience parameterizing/modeling ionosphere over site.
6 - Poor Repeatability of Faraday Rotation	3	3	9	Stable spatial sampling from almost full aperture should produce repeatable measurements of Q/U. Constrain ionospheric contribution via total electron content measurements.	Shaping of interferometer response to ensure stable spatial sampling.
7 - Incomplete Foreground Subtraction	3	4	12	Array has very full spatial sampling and a comparatively clean response (PSF). Task is easier than for MWA & PAPER.	Removal of point sources during calibration (peeling) or near real-time (added expense to real-time system).
8 - Instrument Response Poorly Determined	2	5	10	Better calibration, better elements, similar data volumes and computation to MWA & PAPER.	Full prototyping of this capability from MWA & PAPER will inform any contingency development.

Violations will be addressed via government channels, where appropriate. RFI contamination at lower levels, as a result of natural processes, or from satellites will be removed in data processing, starting with simple excision of known interference.

2. Limited Antenna Performance

Probability: Likely; Consequence: Major

For the numbers in question and broad bandwidth, array antennas may not be finely tuned. Mitigation will be provided first by experience gained in developing, manufacturing, and using three generations of hardware (MWA, PAPER, and HERA-1b). Key characteristics will include beam response stability, repeatability antenna-to-antenna, and science-driven limits on polarization leakage. The size of the HERA-II array will provide a second tier of mitigation. The ability of high-accuracy calibration techniques to "repair" data increases at least linearly with the number of antennas, due to signal-to-noise considerations and redundancy among baselines.

3. Large-N Correlation is Expensive/Impractical

Probability: Possible; Consequence: Major

Risk in correlation depends on the number of antennas, their organization (e.g., phased tiles), and instantaneous bandwidth. For 5K MWA-like tiles or 15K PAPER–like dipoles (as required to achieve similar collecting area), HERA-II correlation is conceptually 2-3 orders harder than for HERA-I, but the net challenge is softer. Moore's Law *has already* provided first tier risk mitigation. FPGAs now on the market (Virtex-6) are $\sim 10\times$ more capable than those used by MWA, and scalability of the correlator design would allow a practical HERA-II system to be built *today*. For a switch-based correlator (as used in PAPER), switch capacity is as critical as FPGA capacity.

4. Real-time Calibration/Imaging is Expensive/Impractical

Probability: Unlikely; Consequence: Major

Calibration algorithm performance increases quadratically with array size. The increase in array size to HERA II will massively over-determine the calibration problem. Unfortunately the calibration computing load also scales quadratically with the number of baselines. The MWA computing and calibration model is also very scalable. A single generation advance in GPU performance (by 2011), also splitting a single channel calibration load over 2 GPU (instead of performing 12 channels of load on a single GPU), combined with discarding the short baselines for calibration purposes and assigning the imaging task to auxiliary GPUs would meet the anticipated computing load of HERA II.

5. Ionosphere Calibration Limits Dynamic Range

Probability: Certain; Consequence: Major

Ionospheric mitigation schemes are similar the adaptive optics used to mitigate poor atmospheric seeing in the optical regime. In the low frequency radio case we apply a corrective distortion in the image plane constrained by the precise measurement of catalogue point sources across the field of view, the quality of the the correction is a function of the density and signal to noise of the catalogue sources. Mitigation therefore is provided by building a large array which provides sufficient calibration source density. We can further mitigate by only observing at times of stable ionosphere. The ionospheric calibration risk for HERA II is further mitigated by the fact that the performance of these methodologies will be extensively

TABLE VI: Technical Risk Matrix

Likelihood		Consequence				
		1 Insignificant	2 Minor	3 Moderate	4 Major	5 Severe
5	Certain		1		5	
4	Likely				2	
3	Possible			6	3	7
2	Unlikely				4	8
1	Rare					

Note— **(1)** RFI; **(2)** Antenna performance; **(3)** Large-N correlation; **(4)** Real-time calibration and imaging; **(5)** Ionospheric distortion of images; **(6)** Faraday Rotation; **(7)** Foureground subtraction; **(8)** Instrument deconvolution.

established by the operation of EOR and PAPER and HERA Ib.

6. Poor Repeatability of Faraday Rotation Measurements

Probability: Possible; Consequence: Moderate

Interferometers are not sensitive to the mean level of signal that they are exposed to, they are only sensitive to the spatial frequencies that are sampled by the baselines between antennas. This has no effect on point source detection as they are present on all scales, but it does effect diffuse emission. Different observations at different hour angles may produce different measurements of diffuse structure because different spatial frequencies are sampled. This results in different apparent levels of diffuse polarization, which effects the reliability of any Faraday rotation measurements. The HERA instruments mitigate this effect by ensuring that the spatial sampling is as full as possible, and by good characterization of the direction dependent polarized response of the elements.

7. Limited Foreground Subtraction

Probability: Possible; Consequence: Major

Extensive foreground subtraction, both of point sources and diffuse emission is required to reach the mK brightness temperatures of the 21 cm signal. The extent to which point sources can be removed from images is an extremely strong function of how well the instrument response function (PSF) is known. This risk is mitigated by increasing the filling factor of the array. HERA II with thousands of elements will have an unprecedented PSF, and it will characterize the direction dependence of the array elements via real-time calibration. Furthermore techniques for diffuse and point source subtraction will be developed as part of the MWA and PAPER experiments.

8. Difficulties Determining Instrument Response Function

Probability: Unlikely; Consequence: Severe

The HERA-II array has less demanding deconvolution and modeling challenges than MWA and PAPER. Its larger collecting area and number of elements results in better image quality and better calibration than the smaller first-generation arrays. The detection and characterization of the 21 cm power spectrum places very stringent requirements on a deconvolution and instrument modeling scheme. This risk is comprehensively mitigated by the 21 cm analysis required for MWA and PAPER.

VII. COST ESTIMATES

The RFI2 request for Cost Estimation is difficult for our project in its current state of evolution from current infancy through phases with critical feedback from one funding cycle (2-4 y) to the next. The 21cm reionization measurements will likely follow the phased development of Cosmic Microwave Background measurements from first detection which is a few years off to richer science and more technically demanding telescope arrays. The main goal for the decade ahead is clear: a second generation telescope array. We discuss in this document the staged approach to that array. We argue that cost estimation is reasonably accurate for the decade given the experience in hand already with the combination of relatively simple analog hardware and close tracking of sophisticated digital electronics. We don't have a formal HERA partnership in place, but have made commitments to do so in 2010; this decadal review process has stimulated close working relationships between the MWA and PAPER groups in the US. We will assume that international partnership will provide at least 25% of the projected HERA II project costs, but do not spell this out in tables.

All cost estimates in the following text are in 2009 dollars. In the following funding timeline tables we give costs both in FY09 and in real dollars using 3% inflation with summary in 2009 dollars. We also tabulate costs of construction in 2009 dollars.

In §VII A,VII B,VII C we first present 'ground truth' costing analysis of our current arrays, PAPER and MWA. In the past year the MWA group has developed a WBS for their expansion to 512 tiles. This can be made available to the panel upon request. The generic components of a low-frequency, cross-correlation synthesis-imaging array are:

- *Antenna* that includes the dipole, or dipole-like, element; replication of this element in the case of multi-element 'tile' reception; ground screen; and for tile receptors beamformer cabling and enclosure.
- *Cable transmission* that carries RF signals from the antenna to central processing equipment hut.
- *Analog electronics* comprised of: an 'active balun,' which provides the first stage of amplification and filtering; beamformer in the case of a tile receptor; receiver for further gain and filtering in the equipment hut.
- *Digital electronics* comprised of: Analog-to-Digital converter (ADC); spectrum creation, or 'F' operation; correlation and averaging, or 'X' operation; realtime calibration and imaging; data storage and transmission.
- *Analysis* that occurs subsequent to data collection and has multiple stages including: RFI excision; calibration; imaging; sky model subtraction; and, ultimately for HERA I,II phases, power spectrum estimation.

In §VII D we describe planned activity during 2012-14 that follows the initial MWA and PAPER deployments, data gathering and analysis during 2009-11. This time window will see both improved uses and extensions of MWA and PAPER as well as critical R&D that will inform the proposal for HERA II in 2014.

In §VII E we develop a strawman cost model based on current HERA I technologies and extrapolations, importantly, of the cost of signal processing. We cannot make a sound estimate of HERA III, but do put forward the goal of maintaining a strong R&D program in the late years of HERA II near the end of the decade, as we do for late years of HERA I, in §VII F.

A. HERA IA: PAPER Cost Analysis

Analog. PAPER costs for the subsystem from the dipole antenna through the analog electronics to the digitization stage is $3915 per antenna ($560/m^2). The estimate assumes US production and 32 units (as parts kits that can be containerized for shipping) and includes remote-site deployment costs (shipping, travel).

The bulk parts for the PAPER antenna and ground screen, which is shown in Figure 7, are currently fabricated under contract with US vendors. These parts kits costs are currently $386 and $1428, respectively. Assembly in the field can be achieved in 1 hour, for which we assume $35/hr rate using unskilled labor. Total: $1849/ant.

The balun component (first stage amplification) that snaps into each antenna has a current parts cost of $301 and requires 10 hours of technician time for assembly and testing at a $50/hour labor rate. Total: $801/ant.

The signals are routed through commercial TV coax that either lies on the ground, or is lightly buried to avoid chewing by fauna and also variable loss resulting from solar heating. The material cost for cable and connectors is $115 for 150m, an average distance from the hut for PAPER configurations currently considered. Cutting and connectorizing and testing these cables will add to this cost. Total $165/ant.

The second stage of amplification and filtering are achieved on 'receiver' boards. Parts cost for these boards is currently $275, and they take 5 hours of technician time to assemble and test. Total $525/ant.

There are a number of smaller items such as the EMC enclosure that shields the high-gain receivers from both the antennas themselves as well as the co-located digital electronics. Our current approach has been US based construction and then containerized shipment by sea to the remote site. A 32-antenna array can be shipped for approximately $5000 from Virginia through to a southern hemisphere site in about 1 month port to port. There are handling and (un)packing costs associated with this that might be reduced by more regional construction. PAPER's initial choice for this route has been driven both by desire for thorough testing at our Green Bank, WV engineering array site, and by lack of known vendors and experience with the remote site.

The current cost to procure and deploy 128 antennas will have a SUBTOTAL 1 of $490K. We expect unit quantity discounts approaching 20% for a 512-antenna array.

Digital. Digitization and *FX* correlation uses open-source, shared-development design led by the CASPER [21] and members of the South African Karoo Radio Telescope team.[1] A central CASPER goal is rapid development of correlator technology, maintaining hardware "on the Moore curve," with firmware that migrates between generations of FPGA devices. In 2009

[1] http://www.ska.ac.za/meerkat/index.php

dollars (k), the correlator hardware replication cost is $C_{\rm FX} = (4N + 0.04N^2) * (B/350)$, where N is the number of dual-polarization antennas, and B is the processed bandwidth (MHz). For the 2010 128 antenna array the correlator cost is $114k, assuming the fiducial 30-MHz bandwidth (below the current 100-MHz maximum of the PAPER correlator [21]), and fully-costed FPGAs. A full year of an engineer FTE ($200k) would be needed to replicate, test and deploy this unit assuming the shared-development model. The correlator rack associated cables, signal distribution, and other small items will raise the total by $50k. In conclusion, SUBTOTAL 2 is $362k, which is 74% of the antenna/electronics component.

Development and integration costs are approximate because of the staged growth of the instrument. Analog subsystem development costs have been distributed, and correlator development costs have been reduced considerably by CASPER. Thus far, additional FTE expenditure for system integration of a 32-antenna array is approximately equal to the cost of the analog subsystem; scaled to 128 antennas, this SUBTOTAL 3 is $400K. The estimated full cost of a 128-antenna PAPER instrument including development is then, $C_{\rm PWA128,full}$, is ~$2M.

B. PAPER IA: MWA Cost Analysis

Analog. MWA costing of antenna and associated electronics and cabling is well developed and based on extensive quotes for manufacture, testing and field installation of major and heavily replicated systems. Total costs per antenna element (4x4 tile, see below) is $5989/tile. At 150 MHz the effective area is 20 m^2 and therefore the MWA architecture provides a cost of $285/$m^2$, a factor of 2 lower than that of PAPER (previous section).

The MWA antenna is a tile consisting of a 4 by 4 array of 'bowtie' elements snapped onto a stiff ground screen laying on the ground as shown in Figure 5. In quantities of 100s the tile costs run $1600/tile and $300/tile for the two components, resp., with all costs included through to full deployment.

The tile is operated as a small phased array to boost gain toward a particular field of view during a given night of observing. The beamformer electronics, which use analog delay line technique, cost $300/tile.

The tiles are geographically clustered in sets of 8 to form nodes. Coaxial cabling from beamformers into a temperature-stabilized node enclosure costs $70/tile. The node electronics consist of amplification, filtering, direct sampling, digital filtering to 1-MHz resolution, band selection and conversion to fiber optic modulation. This electronics and enclosure cost $3000/tile, about half of the per tile costs. Each node receives a clock signal for the digital signal processing distributed from a central source; this costs $144/tile.

Fiber optic communication along a typical path costs $125/tile.

Digital. The FPGA-based, FX correlator for MWA-512 array cost is now accurately estimated at $784k. In this total are approximately $125k for 'F' operation that scales with number of tiles (N); $541k for 'X' operation that scales with N^2; and $120k for infrastructure (racks, cables, controller). At N of 512, correlator costs become dominated by the N^2 factor.

The MWA-512 will require Real Time Calibration (RTC) system that rapidly processes correlator samples, compares to a sky model and adjusts data – adaptive optics at meter wavelengths – to allow longer integrations that reduce data storage needs to practical level.

The planned RTC hardware cost for MWA-512 RTC is $385k, and consist of gpu-loaded computers.

Total MWA-512 construction cost is now well controlled at the 10% level, $4.2M. Like PAPER, MWA development has been staged, with substantive preparatory work prior to initiation of the project contributing major design elements. This makes the full cost difficult to estimate. A plausible forecast at completion of the 512-element array is $11M for development across the contributing partners and sponsors.

C. HERA IA: Common Costs Analysis

General Infrastructure Costs. For work at MRO, these costs are covered by CSIRO in its site operations model. We assume that infrastructure will continue to be an in-kind contribution intended to enable astronomy at the remote MRO or comparable foreign site. As a result, central power generators, buildings, internet connectivity and the like have *not* been included in the calculations.

Operations Costs. Because the MWA model includes full-time (24×365) operation, we use it for estimation of related costs. (PAPER is a campaign instrument for which conditions are optimal 3 months per year.)

Fixed dipole arrays have intrinsically low operating costs and follow a different model than traditional radio (or other) observatories (e.g., no allocations are needed for mechanical elements, and cryogenics.) The dominant operations costs are for the stable power supply that is required by digital signal processing and High Performance Computing (HPC) systems, and modular, transportable data storage that is anticipated to service on the order of 50-500 TB per year depending on science application, where periodic site visits are required to change out disk systems. Otherwise, most operation will be controlled remotely, via an existing satellite or future hard link.

An early estimate for diesel fuel required for a budget $O(100)$ kW (c. 2007) was $350K/y for full-time observing ($24^h \times 365^d$). An independent estimate comes from CARMA, which has been running under diesel generator power at a high altitude site. Their load is 260 kW and, while fuel prices have been going through gyrations, they report $0.1555/kWH over past year. The cost per kWH for equipment amortization and maintenance costs adds $0.0487/kWH. The annual full cost at these rates for 100 kW would be $178K, about a factor of 2 less although much of that can result from price volatility. We proceed with estimate of $250K/y.

For maintenance of MWA, 2 FTEs based within a few hundred km of the site will be required; $170K/y. The mean time between failure for hardware elements exposed to ambient conditions and for HPC elements are under study; we assign a hopefully generous $100K/y for supplies and repairs and minor upgrades. We assume that engineering anticipates serious failure modes of parts. Site visits and administrative details will add an additional $50K/y. *Total operations cost is then $570K/y.*

Research & Development. In §IV we describe important R&D efforts that will be ongoing. Once MWA and PAPER arrays are deployed, technical staff including students can be working on future deployments with improved performance and critical components for future phases. *We schedule R&D at $500K/y during HERA IA.*

Science and Related Costs. Expenses related to reduction and analysis is proportional to data volume and inversely proportional to the time available to complete work. Two labor-intensive activities for both PAPER and MWA will be all-sky surveys to characterize foreground emission and long integrations of selected cold patches on the sky. The latter will support iterative extraction of foreground emission models and statistical analyses to search for a reionization signature. For either array, survey and reionization studies each can readily generate hundreds of TB of raw data per year, with complete analysis desired on an annual cycle. Prior to the start of science operations, cost estimates are necessarily uncertain. MWA planning in the US foresees $O(10)$ FTEs working directly with these data (crudely, a few TB per person per month), with $O(\$500\text{K})$/y in off-site computing hardware and storage costs. *The total for science and related is about $2M/y.*

D. HERA IB (2012-14) Cost Estimation

There are two major activities during 2012-2014: science observations and data analysis from the MWA-512 and PAPER phase 2, and the design of the HERA II array. The efficient use of resources thus leads to a bifurcation in effort, with scientists and developers taking different paths. This is driven by the need to 'freeze' development on the arrays during this period to obtain consistent science data, and to allow the science team to concentrate on issues specific to the data at hand. Consequently, it is necessary to budget manpower for the parallel operation of both groups without the assumption that their effort will necessarily be multi-use. That said, the flow-down of requirements and practical feedback from the science team based on the ongoing analysis will be of crucial importance in prioritizing the work of the development team.

During this period, the MWA-512 will be mature and conducting regular observations. The base budget for facility operations and maintenance is ∼$0.5M/year, with an additional ∼$1.2M/year devoted specifically to science and associated data processing needs. A major contribution to this effort will come from US NSF funds, while it is expected that 25% will come from international collaborators and, e.g., independent postdoc appointments. The plans for the PAPER team are more fluid but will focus on improved scientific results relative to 128-antenna array currently planned, either via expansion or by fielding a hybrid array with the MWA. Further, the team will work on a 10% prototype HERA II technology demonstrator. *Exclusive of the HERA II prototype construction, the costs for operations and science for both groups are $7.8M.*

The key elements of the research and development in this period, as outlined earlier in our Enabling Technologies section, §IV, are: the specification of the HERA II array element and associated balun and beamforming electronics ($0.7M); the choice of antenna configuration(s) ($0.3M); the several components of digital signal processing ($2M); and end-to-end software ($1.0M). Overall system design, coordination and contingency will take an additional $1M. If the basic element is a tile, then the analog or digital beamforming must be specified. Whether or not to use analog (coax) or digital (fiber) signal distribution between the antennae or antenna nodes and the correlator, which is partially determined by array configuration, has a major impact on cost and the organization of the central correlator. The correlator is a major challenge and we expect HERA group to be at the international forefront with tackling the large-N, scalable technology albeit at narrow bandwidth per signal. These

critical efforts, which will inform and benefit from parallel SKA activities, are critical to the preparation of the HERA II instrument in 2014. *Total R&D costs for the period are $5M.*

There are elements of scaling up to HERA II which exhibit a 'threshold effect.' That is, it is not possible to fully test the efficacy of the approach until the size of the array reaches a critical size. This is particularly true of precision calibration schemes. This drives the need for a HERA II prototype that is a significant (10%) fraction of the final array in order to fully test and demonstrate the approach. The exact split between R&D and prototype will be further assessed in 2011 when this HERA IB work will be proposed to the NSF and other agencies. The field research for this stage can of course be done using HERA IA deployments at the southern hemisphere and Green Bank sites with equipment that is separate from the HERA IA science arrays. The prototype array can potentially provide new science results during construction of the full HERA II telescope. Scaling from the projected costs of the full project that are in many areas well defined based on current HERA I arrays, *the prototype array will cost $6M.*

Summary of Phase IB 2012-14 (USD FY09; two activities): Operations $3M; Science $4.8M; R&D $5M; Prototype $6M; Reserve $0.5M; TOTAL ∼$20M.

The use of HERA I telescopes and HERA II prototype will provide useful science. And science analysis of HERA I data will continue throught ot 2016. We extend then this effort.

Summary of Phase IB 2015-16 (USD FY09): Operations $2M; Science $2M; R&D $0.5M; TOTAL ∼$4.5M.

E. HERA II (2015-19) Cost Estimation

Costs for HERA II are dominated by construction and operation. Organization of construction and operation was discussed in §VI B. Figure 12 provides the overall management structure. Cost uncertainty is dominated by signal processing: both the FPGA-based correlation and the subsequent real time analysis to calibrate ('adaptive optics' correction of ionosphere as well as solution for dynamic instrumental gain terms), and then to average as a means of handling the raw data rates.

The bulk of HERA II development work will be accomplished during the earlier MWA and PAPER Phase IB programs up to 2014. We adopt a fiducial array of 5000 phased tiles each consisting of 4x4 subarrays of antennas. This scale is driven simply by desire for collecting array to meet sensitivity goals discussed in our science case. We assume an instantaneous bandwidth of 30 MHz that, again, is matched to our science requirements. Experience with MWA and PAPER and development work within those programs will be used to justify deviations from these parameters, such as a larger number of individual antennas, a somewhat smaller number of larger tiles, or a mixture. Similarly experience with Phase I might lead to adjustments in the center frequency.

We also assume for this cost estimation that HERA II will be located at the MRO site in Australia. Alternate siting of HERA II is possible if cost, scientific and/or technical drivers (array design, RFI, etc) present a compelling case for major changes. In that case, we reasonably assume infrastructure costs for US collaboration will continue to be an in-kind contribution from the international partner. Importantly, we are not in this cost estimation including power infrastructure. We will further assume that the power consumption charge will be provided by the host country as part of an international collaboration. Implicitly, a

US site is not considered at this time owing to strict limit on RFI.

Antenna and Related. For antenna and digital receiver hardware, we adopt the MWA cost model discussed above. Though HERA-II units will be potentially more sophisticated, most cost will be in development (e.g., propagation modeling, firmware programming), rather than to pay for manufacturing and as stated above we assume these costs are incurred during Phase IB.

FX Correlator. engineer, with the For correlation we adopt the current cost of the MWA unit scaled to a larger number of inputs, with a time-dependent cost reduction applied to DSP processing hardware reflecting gains in FPGA processing efficiency based on products (e.g. Xilinx Virtex 6 FPGAs) close to market today. This costing model may be expressed as:

$$C(N, B, t) = 5.1N + 7.6B + .011NB + 8.3e - 5N^2B \left(\frac{1}{2}\right)^{\frac{t-t_L-t_0}{t_m}} \quad (2)$$

where C is cost in \$1k in 2009, B is bandwidth in MHz, N is number of antennas, t is the year in which a correlator system is deployed, t_L is the lead time in years required between selecting a commercially available processing element and deploying a correlator based on that processor (nominally, 4 years), t_0 is the expected 2010 deployment date of the 512-antenna MWA system on which this costing model is based, and t_m is the time constant for doubling according to Moore's Law as applied to FPGA processors (trends in the last 5 years indicate $t_m \approx 1.5 years$; we conservatively estimate t_m to be 2 years).

Figure 14 illustrates the scaling trends of hardware cost with antenna number for fiducial deployment dates according to the above model. With $O(N^2)$ components of cost decreasing exponentially with time, it is undesirable to have them dominate hardware costs, since a moderate delay in schedule could substantially reduce the cost of the project. We propose that $O(N^2)$ components be one-third of the cost of the $O(N)$ array costs, reflecting a balance between lowering the cost of DSP hardware by delaying the deployment schedule of a correlator system, the decreasing marginal utility of delaying once $O(N^2)$ components are no longer the leading cost, and the need to set a competitive schedule for acheiving reionization science goals. The intersections between $O(N^2)$ cost components (colored solid lines in Figure 14) with $\frac{1}{3}O(N)$ cost components (dotted line) suggest an optimal number of antennas for a given

Targeting an optimal array size, our best hardware cost estimate is \$3M for a 5000-antenna, 30-MHz array brought online in 2016. In addition to this direct cost we add 9 FTE-years of effort by a digital engineer, a software programmer and a staff scientist; \$1350k. Associated computing, disk storage, cabling, racking, clock generation and distribution add an additional \$750k. The total cost is \$5.1M.

Real-Time Computer. For the Real-Time Computer (RTC), we scale from the stream processing system to be deployed for MWA starting in Q4 2009, rebalancing loads and data rates, and extrapolating computing speed to 2013. We adopt a heterogeneous computing architecture, comprising a multi-core CPU cluster with multiple high-performance graphics processing units (GPUs) as co-processors (Figure 13). The RTC processes correlator output, delivered via 10 g-E lines and generates images for each spectral channel, covering the entire field of view. The output will be dumped over 1 and 10-gE lines to a mass archive (co-located with the RTC at a local support base or via long-haul fiber optic cable if computing is performed on site.)

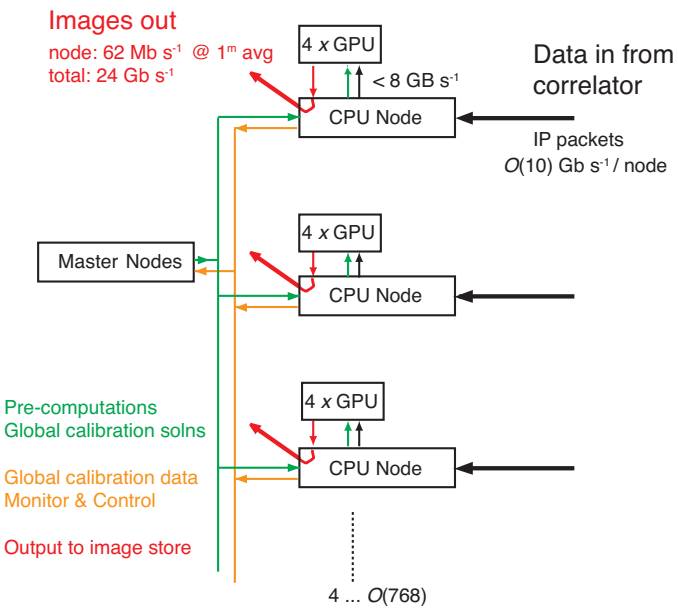


FIG. 13: Schematic architecture of the Real-Time Computer, shown as a scaled version of the MWA system. Calibration and imaging pipelines run in parallel on each node, processed entirely within the local GPUs prior to transfer of images back to node memory at the end for accumulation High-performance CPUs are required for the accumulation, data flagging, I/O management, and execution of distributed-computing operations not suited to GPU architectures, e.g., global solutions for the ionosphere and coordinate transforms. Characteristic data rates are indicated, assuming each node processes 2 spectral channels.

Operation is parallel by frequency interval with a fixed set of channels assigned to each computing node (Figure 13). Communications among RTC nodes, when required, is low-bandwidth scatter-gather transfer of calibration data over a separate dedicated 1-gE network. Images for each frequency are accumulated in local CPU-host memory before archiving. Shorter rather than longer integration times are preferable from the standpoint of EOR science analysis.

Off-the-shelf technology will be crucial to achieve a cost effective real-time computing system, an approach being applied to the MWA. The number of nodes is determined by their capacity, load per frequency channel (crudely, Flops s^{-1}), and number of channels. Data rates for input and output are shown in Figure 13. Simple scaling of the MWA real-time computing system, and scalings therefrom (sub-system to sub-system), would yield a *minimum* 5 PFlop s^{-1} total load.

Each computing node requires 6 TFlop s^{-1} to process one frequency channel. The total count for an instantaneous bandwidth of ∼30 MHz and 40 kHz channels is 768 channels. For GPU performance of 1 TFlop s^{-1} in mid-2008, we anticipate cards capable (conservatively) of 8 TFlop s^{-1} in 2013. This corresponds to an NVIDIA product cycle of ∼18 months for a doubling of capacity, as has been demonstrated at least from first introduction of the G8800 family. Out of conservatism, we extrapolate only three generations; the early lock provides a stable real-time code-development platform during the HERA-1b prototype phase, and 3 years in advance of first deployments. We therefore assume GPU performance of 6 TFlop s^{-1},

well within these estimates.

Each node will host four GPUs ($1.5K each), which will service two spectral channels, the GPU load will be broken into calibration and imaging for each channel (one GPU each). This division of processing enables critical elaboration on the MWA calibration system, enabling iterative subtraction of extended sources as well as point sources and full correction for array non-coplanarity. These are straightforward but are not now part of the MWA system. Peeling dominates the computing load, and a simple two-times iteration of this operation from scratch would raise the computing budget to 11 PFlop s^{-1}. For two channels per node, real-time computing requires 1536 GPUs and 384 nodes, which provide a computing budget of 12 PFlop s^{-1}.

Generic high-performance CPU hosts will house the GPUs. Were construction undertaken today, these would be built around high-throughput dual quad-core systems with Nehalem processors, parallel Tylersberg chipsets, and O(24 GB) of RAM ($5.0K each). The total cost for computing nodes would be $4.2M (not including local storage).

Aggregate output data rate from the nodes to storage will throttle array performance. For 1^m averaging of images, the output data rate will be 62 Mb s^{-1} node^{-1} or will be $\sim$ 24 Gb s^{-1} total (Figure 13). EOR data analysis benefits from short accumulation times, which improves foreground subtraction accuracy and eases the task of forward modeling the sky. Here we assume 1^m drift scans that track the sidereal motion of designated EOR fields, but the optimal number will be based on knowledge of the sky accumulated with MWA and PAPER, as well as HERA-1b prototype instrument performance.

Individual nodes will accept data directly from the correlator via 10-gE channels, without external network infrastructure. Node output will be managed by inexpensive 1-gE protocol. The real-time processing does not require high bandwidth all-to-all communication which allows a low-cost networking topology consisting of a small number of medium capacity 1-gE switches (16x24 port) with 2 $\times$ 8-port 10-gE switch serving as a backplane (total $\sim$ $52K).

Power consumption will be dominated by the computing nodes. If we assume no growth in power consumption for top-of-the-line GPUs (200 W) and CPU nodes (150 W, including e.g., memory, storage, interfaces, cooling), the total for 384 nodes is 373 kW. The 18 switches will draw $\sim$ 10 kW for basic systems that afford only limited configuration management.

Storage requirements are difficult to estimate in advance of optimized EOR analysis techniques to be developed during the HERA-1 phase. For 10^m time-averaging of images, one year of images (1700^h) would require $\sim$ 1.5 PB. This constitutes a minimum storage budget. For 1^m averaging (i.e., no savings in science analysis inside 1 year), the total would be $\sim$ 20 PB in a RAID6 configuration.

Assuming a doubling of storage capacity per disk every two years, a baseline of 2 TB per disk in 2010, 10 W per disk, and a 2015 lock on storage procurement, power consumption for storage would be up to 18 kW. Assuming a media cost of $\sim$ $0.1 Gb^{-1}, the corresponding budget impact would be $2.0M in what could be considered to be a worst case computational scenario vis-a-vis the mechanics of EOR data analysis.

In the event the RTC is housed on-site, the archive can be held off-site if limited local working storage is available. As the EOR observing has only a 50% duty cycle it is possible to store a minimal amount of data on site, whereby the data can be removed via a fiber connection during the off-duty hours. We budget two 2 TB disks per node, which permits 140^h of observing. The minimum useful fiber bandwidth will be 24 Gb s^{-1} (assuming 50 %

duty-cycle).

The footprint for the RTC and storage will depend closely on cooling technologies available either on site or at the local support base. For simply air cooling, we assume 3U per RTC node and 2/3 loading of each rack to make room for ventilation, and network and distributed UPS/line conditioning hardware, in the event a monolithic UPS is not provided by the host institution. This footprint is 14 racks. Storage based on successors to today's 2 TB disks is anticipated to provide ~ 500 TB in a 4U profile. If this density is met, then storage will require ~ 1.3 racks. The total footprint for real-time computing will be ~16 racks.

The personnel budget related to the RTC during HERA-II supports specification, engineering, procurement, assembly, and management of the system, which would be $\sim 10\times$ as large as the Roadrunner system that tops the Top500 list of supercomputers in 2009. Staffing will depend sensitively on MTBF statistics established during the HERA- 1a and HERA-1b phases. Statistics are limited for GPUs with continuous peak load conditions. At this time, we estimate a staff of 4 FTEs for the duration of HERA-II, based on a division of labor between maintenance of the storage facility, the network, and computing nodes, individually. One additional staff memory provides coverage so as to avoid single-point failures and backup.

The personnel budget for the RTS during HERA-II will support reinforcement of development code taken from HERA-1b for production processing. Basic tasks will include optimization, maintenance, and firmware and operating system patches, More advanced work within HERA-II will include algorithm development that targets optimization of the facility once it is online and performance can be analyzed in detail. HERA-II staffing of the RTS will require 7 FTEs with responsibilities divided among interferometric algorithm research and development (3), multi-core CPU programming (1), GPU programming (1), parallel programming and networks, optimization (1), and build-integrity and test suites (1). The manpower budget is formulated based on experience gained by the real-time systems group for MWA.

The real-time computing budgets are \$6.3M, 400 kW, and 16 racks.

Operations & Maintenance. Operations and maintenance are anticipated to be a small fraction of capital costs for HERA II, after the model being applied to PAPER and MWA. We anticipate a 5% levy on resources, although this may be conservative if locally generated renewable energy is available on site in place of diesel-powered generation. Maintenance is low as the telescope as no moving parts. The conservative estimate is \$2M per year.

Science and Related Costs. Expenses related to reduction and analysis is proportional to data volume and inversely proportional to the time available to complete work. Two labor-intensive activities for HERA II will be all-sky surveys to continue to characterize foreground emission and long integrations of selected cold patches on the sky. The latter will support iterative extraction of foreground emission models and statistical analyses to search for a reionization signature. Estimates of science budget will be informed by efforts in the first half of the decade. We foresee $O(10)$ FTEs working directly with these data with $O(\$500\text{K})$/y in off-site computing hardware and storage costs. *The total for science and related is about \$2M/y.*

Summary of HERA II 2015-19 (USD FY09): Construction ~\$40M; Project Management \$1.5M; 3y Operations \$6M; 3y R&D toward HERA III \$5.5M; 3y Science \$7.5M; Reserve \$2M; TOTAL ~\$62.5M

HERA II array use and data analysis will extend beyond the decade.

F. HERA III

HERA II core engineering and technical teams will continue to design and develop for HERA III once production construction is launched for HERA II and assuming scientific priority of the effort remains high. We outline our expected areas of ongoing R&D in §IV. New ones will certainly arise in this demanding technology and calibration project. Base support at the level of $2M/year by 2018 will support this component of our activity. As with other aspects of our project, there is strong coupling between our goals and those of the international SKA program.

HERA II: 5000 4x4 Tile Cost Estimate -- 2009 July

Subsystem	HERA II #units	Unit Cost USD 2009	Total Cost USD 2009
4x4 Tile w Balun	5,000	1,600	8,000,000
Ground Screen	5,000	300	1,500,000
Beamformer	5,000	750	3,750,000
Cabling – BF to Receiver	5,000	70	350,000
Clock Distribution to Node	625	1,150	718,750
Rcvr w/FPGA & Enclosure at Node	625	24,000	15,000,000
FO Cabling – Rcvr to Correlator	625	1,000	625,000
FX Correlator	1	5,000,000	5,000,000
Real Time Computing	1	5,000,000	5,000,000
TOTAL			**39,943,750**

Antenna Costs per m^2

MWA Unit Cost per Tile USD	PAPER Cost per Dipole, USD	Note
1600	1202	1
300	1446	
750	0	
70	0	
144	0	
3000	525	2
125	165	
	625	3

Total:		5989	3963	4
Cost per m^2:		285	566	

Notes:
1 - PAPER entry includes antenna & balun
2 - PAPER entry is receiver after cabling
3 - PAPER entry covers shipping & deployment, which is inclusive in MWA estimates
4 - PAPER costs would come down 20% for volume purchase.

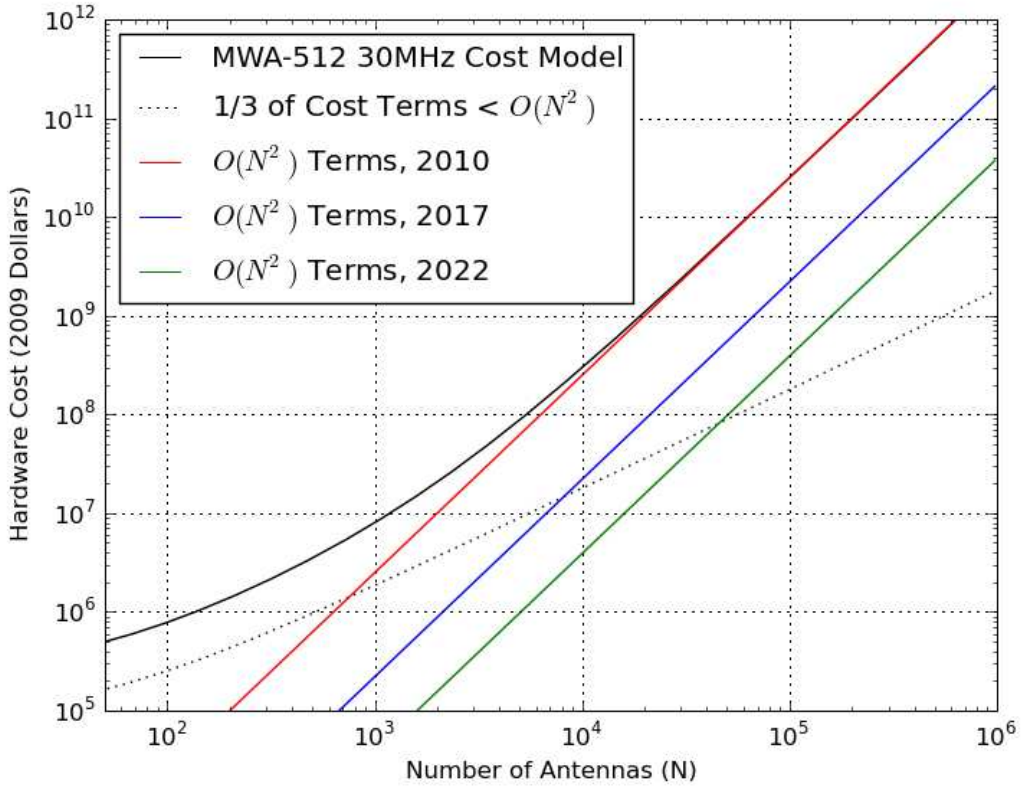


FIG. 14: FX Correlator cost scaling with array size N and development/deployment date

Hydrogen Epoch of Reionization Arrays

	HERA IB			*HERA II*						
Item	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**TOTAL (REAL)**	**TOTAL (FY09)**
Inflation	1	1	1	1	1	1	1	1		
Research & Development, H-II	1000	2000	2000	500					5500	5500
Research & Development, H-III					500	1000	2000	2000	5500	5500
Proj Manage/Sys Engr, H-II		200	300	400	400	300	200	200	2000	2000
HERA IB, prototype HERA II	1000	4000	1000						6000	6000
HERA II Construction				8000	16000	6000			30000	30000
Data System, Software Dev, H-II				2000	4000	4000			10000	10000
Operations & Maintenance, H-I	1000	1000	1000	1000	1000				5000	5000
Operations & Maintenance, H-II						2000	2000	2000	6000	6000
Science, H-I	1600	1600	1600	1000	1000				6800	6800
Science, H-II						1500	3000	3000	7500	7500
Reserve	100	200	200	400	800	400	200	200	2500	2500
TOTAL COST	4700	9000	6100	13300	23700	15200	7400	7400	**86,800**	**86,800**

Hydrogen Epoch of Reionization Arrays

	HERA IB			*HERA II*						
Item	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**TOTAL (REAL)**	**TOTAL (FY09)**
Inflation	1.09	1.13	1.16	1.19	1.23	1.27	1.3	1.34		
Research & Development, H-II	1093	2251	2319	597					6259	5500
Research & Development, H-III					615	1267	2610	2688	7179	5500
Proj Manage/Sys Engr, H-II		225	348	478	492	380	261	269	2452	2000
HERA IB, prototype HERA II	1093	4502	1159						6754	6000
HERA II Construction				9552	19678	7601			36831	30000
Data System, Software Dev, H-II				2388	4919	5067			12375	10000
Operations & Maintenance, H-I	1093	1126	1159	1194	1230				5801	5000
Operations & Maintenance, H-II						2534	2610	2688	7831	6000
Science, H-I	1748	1801	1855	1194	1230				7828	6800
Science, H-II						1900	3914	4032	9846	7500
Reserve	109	225	232	478	984	507	261	269	3064	2500
TOTAL COST	5136	10130	7072	15881	29148	19255	9655	9945	**106,221**	**86,800**

[1] Barkana, R. & Loeb, A. 2005, ApJ, 624, L65
[2] Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, ApJ, 638, 20
[3] —. 2007, ApJ, 661, 1
[4] —. 2008, arXiv:astro-ph/0807.3956
[5] Fan, X., Carilli, C. L., & Keating, B. 2006, ARAA, 44, 415
[6] Furlanetto, S. R. 2006, MNRAS, 371, 867
[7] Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, Phys. Rep., 433, 181
[8] Gunst, A. 2007, LOFAR/ASTRON, 52
[9] Jelić, V. & 12 coauthors. 2008, MNRAS, 389, 1319
[10] La Porta, L. & Burigana, C. 2006, A&A, 457, 1
[11] Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2008, ApJ, 680, 962
[12] Liu, A., Tegmark, M., & Zaldarriaga, M. 2008, arXiv:astro-ph/0807.3952
[13] Lonsdale, C. J., e. a. 2009, arXiv:astro-ph/0604040
[14] Madau, P., Meiksin, A., & Rees, M. J. 1997, ApJ, 475, 429
[15] Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, Phys. Rev. D , 78, 023529
[16] McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, ApJ, 653, 815
[17] Morales, M. F. 2005, ApJ, 619, 678
[18] —. 2008, arXiv:astro-ph/0812.3669
[19] Morales, M. F., Bowman, J. D., Cappallo, R., Hewitt, J. N., & Lonsdale, C. J. 2006, New Astronomy Review, 50, 173
[20] Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006, ApJ, 648, 767
[21] Parsons, A., Backer, D., Siemion, A., Chen, H., Werthimer, D., Droz, P., Filiba, T., Manley, J., McMahon, P., Parsa, A., MacMahon, D., & Wright, M. 2008, PASP, 120, 1207
[22] Parsons, A., Backer, D. C., & 13 co-authors. 2009, arXiv:astro-ph/0904.2334v1
[23] Shapiro, P., Iliev, I., Mellema, G., Pen, U.-L., & Merz, H. 2008, in AIP CS, Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian, 68
[24] Tegmark, M. & Zaldarriaga, M. 2008, arXiv:astro-ph/0805.4414
[25] Tucci, M., Carretti, E., Cecchini, S., Fabbri, R., Orsini, M., & Pierpaoli, E. 2000, New Astronomy, 5, 181