THE INFEASIBILITY OF HIGH-QUALITY IONOSPHERIC CALIBRATION OF SKA1-LOW

Document Number.. SKA-TEL-SDP-0000088

Document Type ....REP

Revision 2F

Author ....T.J. Cornwell

Date 2016-05-24

Document Classification..UNRESTRICTED

Status ....Released

Name	Designation	Affiliation	Signature	
Owned by:				
T.J. Cornwell	Author	T.J. Cornwell		
			Date: 24/06/2016	

The Infeasibility of High Quality Ionospheric Calibration of SKA1-LOW

Author: Tim Cornwell

Version: 2F

Executive summary

We present a theoretical framework for understanding and predicting the scientific performance of low frequency arrays such as LOFAR and SKA1-LOW. Specifically, we argue for a specific approach to calibration in which a smooth model for the ionosphere is determined. This allows calculation of the imaging dynamic range limit due to inevitably limited estimation of the ionospheric phase screen. We show that the imaging dynamic range scales roughly as the square root of the number of stations and the number of sources, compared to the number of stations for isoplanatic self-calibration. Since the number of sources suitable for calibration is normally much less than the number of stations, this is a very substantial obstacle to successful calibration of the ionosphere.

We use this quantitative model to investigate imaging the EOR with the current proposed SKA1-LOW. We find that, under this model, reaching the specified dynamic range (50dB) at any frequency is not possible within the lifetime of the telescope.

This means that the current proposed SKA1-LOW design (stations and the array) should be revisited with a view to improving performance for EOR imaging. Our framework provides a means to explore and evaluate possible design options. We give some recommendations for future work. Changing the configuration alone is not sufficient. Using much larger stations (e.g. diameter 105m) and modifying the configuration does offer some hope of a possible solution but introduces other obstacles due to the need to multi-beam at very high dynamic range.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Acronym	Definition
LOFAR	Low Frequency Array
LOW	Low frequency component of SKA1
SDP	Science Data Processing

1. Purpose of the document

The purpose of this of this memo is to investigate the calibration of the ionosphere above SKA1-LOW and the impact on imaging.

2. Scope of the document

This memo has estimation of the imaging performance of the EOR as the main scientific goal. Thus the problem addressed is the imaging of a smooth extended signal on a few degree scale seen through a turbulent and evolving ionosphere and bright foreground sources.

We do not address processing algorithms in detail, but address instead limits to the imaging performance originating in the calibration of the ionosphere. Although the analysis has implications for power spectrum analysis, we do not discuss those here.

3. References

[RD1]   L Koopmans, et al., "The Cosmic Dawn and Epoch of Reionisation with SKA.", Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14) 9 -13 June, n/a 2015 p. 1.

[RD2]   Mevius, M., 2014 http://www.astron.nl/lofarschool2014/Documents/Tuesday/Mevius.pdf

[RD3]   S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson, "Correcting direction-dependent gains in the deconvolution of radio interferometric images," *Astron. & Astrophys.*, vol. 487, pp. 419–429, 2008.

[RD4]   Noll R.J. "Zernike polynomials and atmospheric turbulence" J.Opt.Soc.Am. vol 66 No. 3 pp207 1976.

[RD5]   A. Glindemann, S. Hippler, T. Berkefeld, and W. Hackenberg, "Adaptive Optics on Large Telescopes," *Experimental Astronomy*, vol. 10, no. 1, pp. 5–47, Apr. 2000.

[RD6]   S J Wijnholds, J D Bregman, and A van Ardenne, "Calibratability and its impact on configuration design for the LOFAR and SKA phased array radio telescopes", Radio Science, 2011 vol. 46 (5) pp. RS0F07-n/a

[RD7]   R J van Weeren et al., The Astrophysical Journal Supplement Series 2016 vol. 223 (1) p. 2

[RD8]   Intema, H T and Van der Tol, S and Cotton, W D, "Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme-I. Method description and first results", A&A, 501, 1185–1205 (2009)

[RD9]   Intema, H. T. (2014). SPAM: Source Peeling and Atmospheric Modeling. Astrophysics Source Code Library

[RD10]   Dillon, J.~S. and Parsons, A.R. (2016), "Redundant Array Configurations for 21 cm Cosmology", ArXiv e-prints, 1602.06259

[RD11]   Martin, Poppy L and Bray, Justin D and Scaife, Anna M M, "Limits on the validity of the thin-layer model of the ionosphere for radio interferometric calibration", arXiv.org, 1604.03810v1

[RD12]   Loi, Shyeh Tjing et al., 2016, "Density duct formation in the wake of a travelling ionospheric disturbance: Murchison Widefield Array observations", Journal of Geophysical Research: Space Physics, 211, 1569-1586

[RD13]   Perley, R.A., (1999) "Synthesis Imaging in Radio Astronomy II" eds. G B Taylor, C L Carilli, and R A Perley, pp 275.

[RD14]   Sanaz Kazemi, Sarod Yatawatta, and Saleem Zaroubi (2013), "Clustered Calibration: An Improvement to Radio Interferometric Direction Dependent Self-Calibration", Monthly Notices of the Royal Astronomical Society, Volume 430, Issue 2, p.1457-1472

4. Overview

The latest proposed SKA1-Low configuration[1] is composed of 512 35-m diameter stations each having 256 log-periodic antennas susceptible to radiation from essentially the entire sky. There is a core of 166 stations spread over a diameter of 5km and a halo of 346 stations spread over the Boolardy station at radii up to 40km.

The prime science target for SKA1-Low is the Epoch of Reionization (see e.g. [RD1]) which is smooth and extended. The core of SKA1-LOW is designed for this purpose.

SKA1-Low observes through the ionosphere, an intensely time-variable medium about 300km above the surface of the Earth. The ionospheric phase functions obeys a Kolmogorov spectrum (see [RD2]) from fine scales to the large travelling ionospheric disturbances having wavelengths of hundreds of kilometres and speeds of 500km/h.
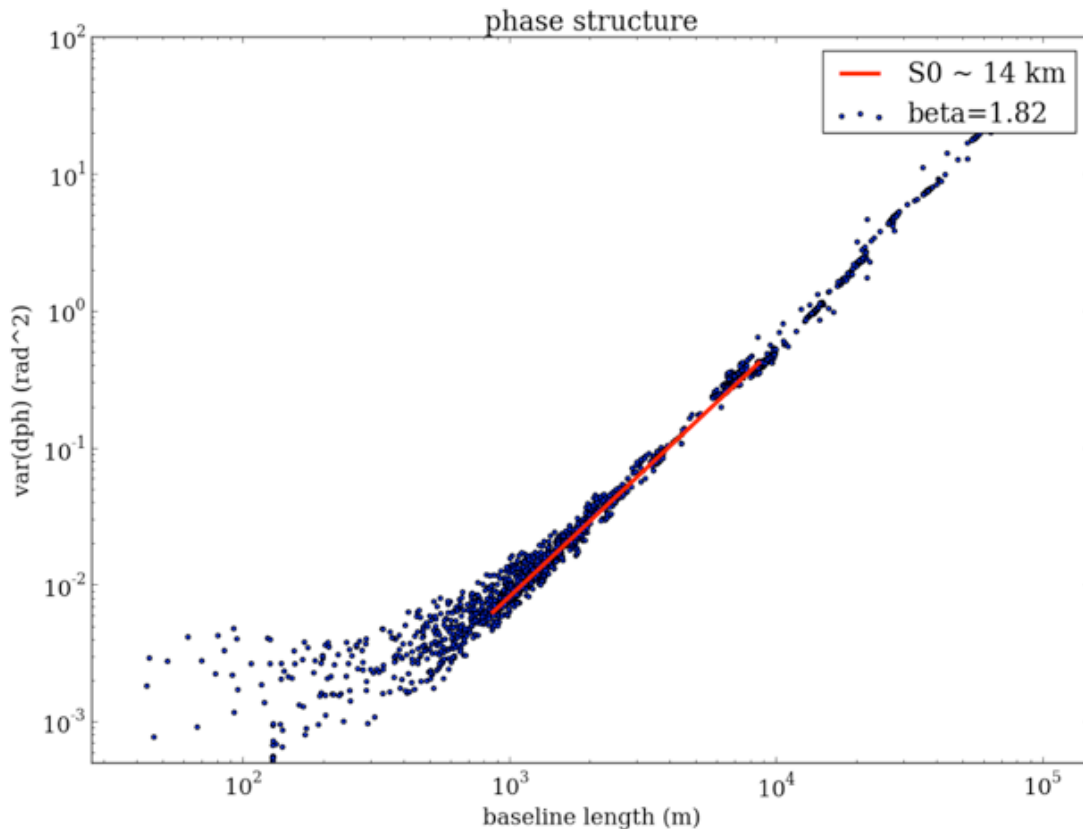
Figure 1 Ionospheric turbulence as a function of baseline.

The sky at the important frequencies (50 – 200 MHz) is composed of three principal components: first the radiation from the Sun second that from the Galaxy and then

---

[1] V4Drssl512red.kml kindly provided by Jeff Wagg.

extragalactic emission arising mostly from discrete radio sources. The Extragalactic sources are homogenous across the sky and obey relatively well-known source counts.

In our view the halo of SKA1-LOW has two purposes:

- Forming a broadband model of the sky at fine scale for removal from the field prior to imaging of the diffuse EOR signal
- Calibrating the ionosphere above the core so that the EOR imaging is possible.

It could be argued that the core can be used to estimate the phase screen. We will investigate this possibility.
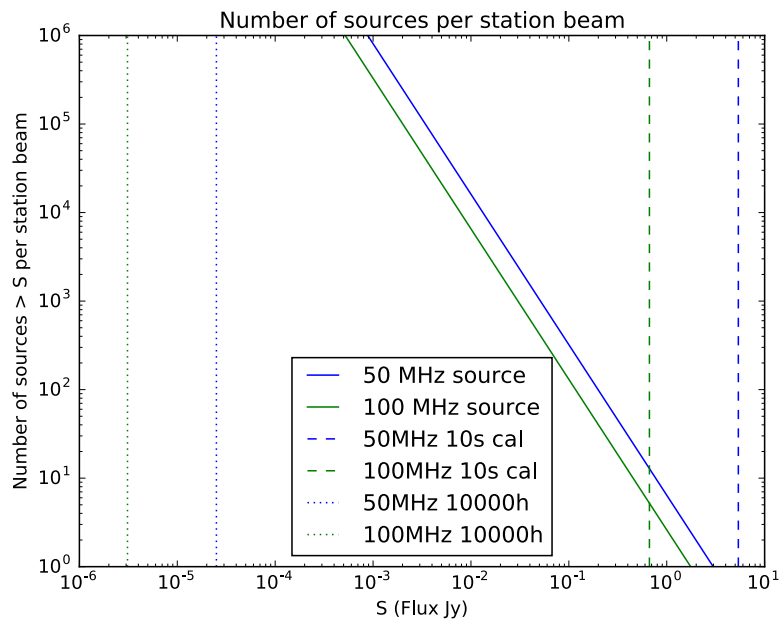
Figure 2 Number of sources per station beam for 50 and 100 MHz. Setting the threshold for calibration at 5 sigma in 10s, there are approximately 5.2 calibrators (>0.66 Jy) at 100MHz and none (>5.39 Jy) at 50 MHz.

Discrete sources can be cleaned and removed thus removing the associated sidelobes from an image of the field of view. This works well inside the station beam and in the first few sidelobes of the station beam. Beyond that the sources cannot be identified and removed and their synthesized beam sidelobes propagate into the station beam modulated by the station beam sidelobes.

The Fresnel scale is about 1km and at 100 MHz the entire station beam is about 55km at the height of the ionosphere so to constrain the entire ionospheric screen we would need about 55*55 pierce points = 32400 distributed evenly. For 40 stations in the halo, this requires about 70 detectable sources in 10s. We can see from Figure 2 that a more typical number is O(5) at 100MHz. Thus constraining the screen exactly is not possible in the roughly 10s available. Inevitably we will fall short. In this situation, we would like to be able to estimate the major modes of the phase screen first, and then be able to estimate the effect of truncated representation. This we discuss next.

5. Analysis of calibration

Calibration of non-isoplanatic radio systems telescopes has been extensively discussed in the literature over decades (recently: [RD6][RD7][RD8][RD9][RD14]). However, system-wide analysis of the calibration and imaging performance is rare. Simulations are a plausible route forward but we would prefer an analysis that is predictive and quantitative. Our focus will be on finding a way to estimate the limits imposed on imaging performance by the calibration problem. This follows on from similar work [RD6], adding more rigorous analysis to the estimation of the phase estimation. We will look at the calibrateability of the array in quantitative terms.

First, we need to set some conditions on what we will consider as an acceptable model of the calibration process. It must obey a number of conditions:

- It should have a clear description in terms of measurements, processing steps, algorithms, etc.
- It should be physically motivated, recognising the knowledge we have about the ionosphere, the telescope, and the sky.
- It should be stable, robust, repeatable, and free from (too many) free parameters
- It should provide quantitative insight into array design
- It should provide quantitative estimates of the calibration and imaging performance

With this in mind we have formulated the calibration problem for SKA1-LOW as follows:

1. The principal obstacle to calibration is a single thin layer in the ionosphere which suffers Kolmogorov turbulence.
2. The array core (up to a few km) must be well-designed in <u>Fourier space</u> to sample the EOR with necessary spacings and sensitivity thus allowing excellent imaging and $\sqrt{t_{sky}}$ improvement in SNR.
3. The core images through a 55 km diameter region in the ionosphere (at 100MHz).
4. The halo of the array (i.e. excluding the core) has two purposes:
   a. To find the fine scale structure of the foreground sources
   b. To estimate the ionosphere so it may be corrected in the core observations of the EOR.
5. The array halo (up to 55km) must be well-designed in <u>real space</u> to sample the ionospheric screen with uniformity and sensitivity (see also recommendation in [RD8]).
6. The array halo (up to 80km) must be well-designed in Fourier space for a realistic synthesis in order to determine the fine scale structure of the field sources. This is assured if point 5 is true but point 5 is not assured if point 6 is true.

This model is illustrated in Figure 5 and Figure 6, and the origin of the numbers is discussed in Appendix A.

This approach can be contrasted with, for example, the Facet Calibration approach [RD6] in which the phase screen is modelled by a patchwork of facets (e.g. Figure 3). FC is part of the

logical progression from fine-scale self-calibration schemes. However, piece-wise representation of the phase screen is in no sense smooth or representative of the Kolmogorov spectrum. Furthermore, the facets are comparable in angular scale to the EOR signal that we wish to recover. For context we show a typical spatial frequency required for EOR imaging in Figure 4. When imaging the EOR, the phase discontinuities at facet edges are likely to affect the very faint structure that is the target. The SPAM algorithm [RD8] does take account of the Kolmogorov spectrum in estimating the phase screen, by using a fitting process, but then only applies the phase correction on on each facet, reintroducing errors due to the finite faceting.

In this memo, we are not primarily concerned with the pierce-point phase-estimation algorithm but instead quantitative prediction of the limits to performance arising from incomplete modelling of the smooth phase screen. The incompleteness arises from two factors: the paucity of sufficiently strong calibrators to estimate the ionospheric phase screen in the ionospheric coherence time (about 10s at 100MHz), and the poor coverage of real space by the array configuration.
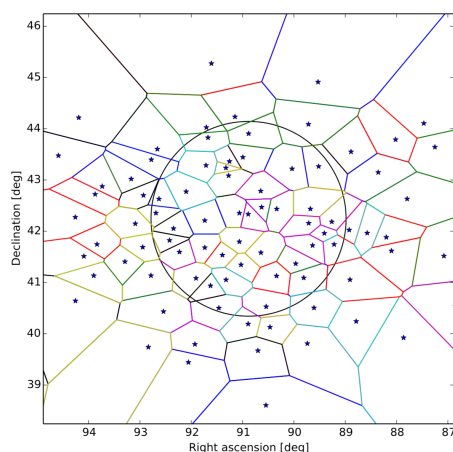
Figure 3 Facets used in an example from [RD6]. In Facet Calibration, each facet is calibrated in isolation, first removing the best estimate of the visibility contributed by the other facets [RD6]. Note that the dynamic range specified in SKA-SCI-0018 is to be assessed on 300" and larger.
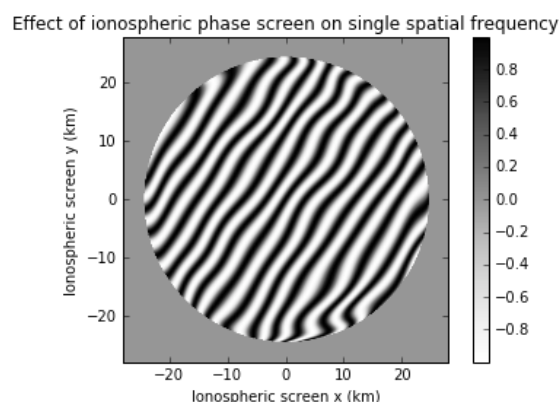
Figure 4 Example of a single spatial frequency distorted by a phase screen. The axes are lateral displacement at the height of the ionosphere,

Ionosphere

Core beam
λ/B

Fresnel zone $\sqrt{\lambda h_{iono}}$

Station beam
λ/D

h_{iono}

B
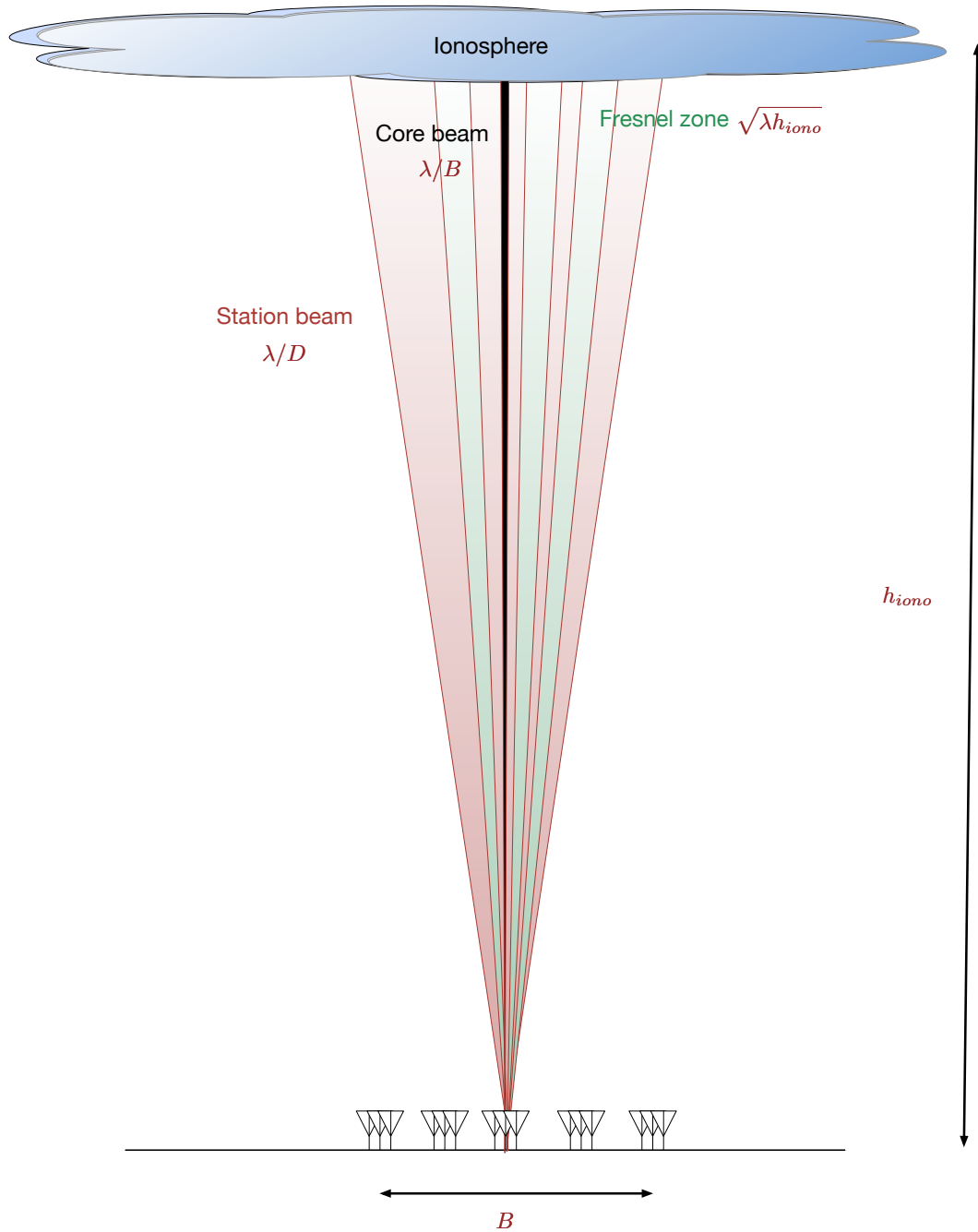
Figure 5 SKA1-Low in EOR imaging. To image the EOR only a compact core of a few km is required. Each station sees through a 55 km (at 100MHz) wide region of the ionosphere.
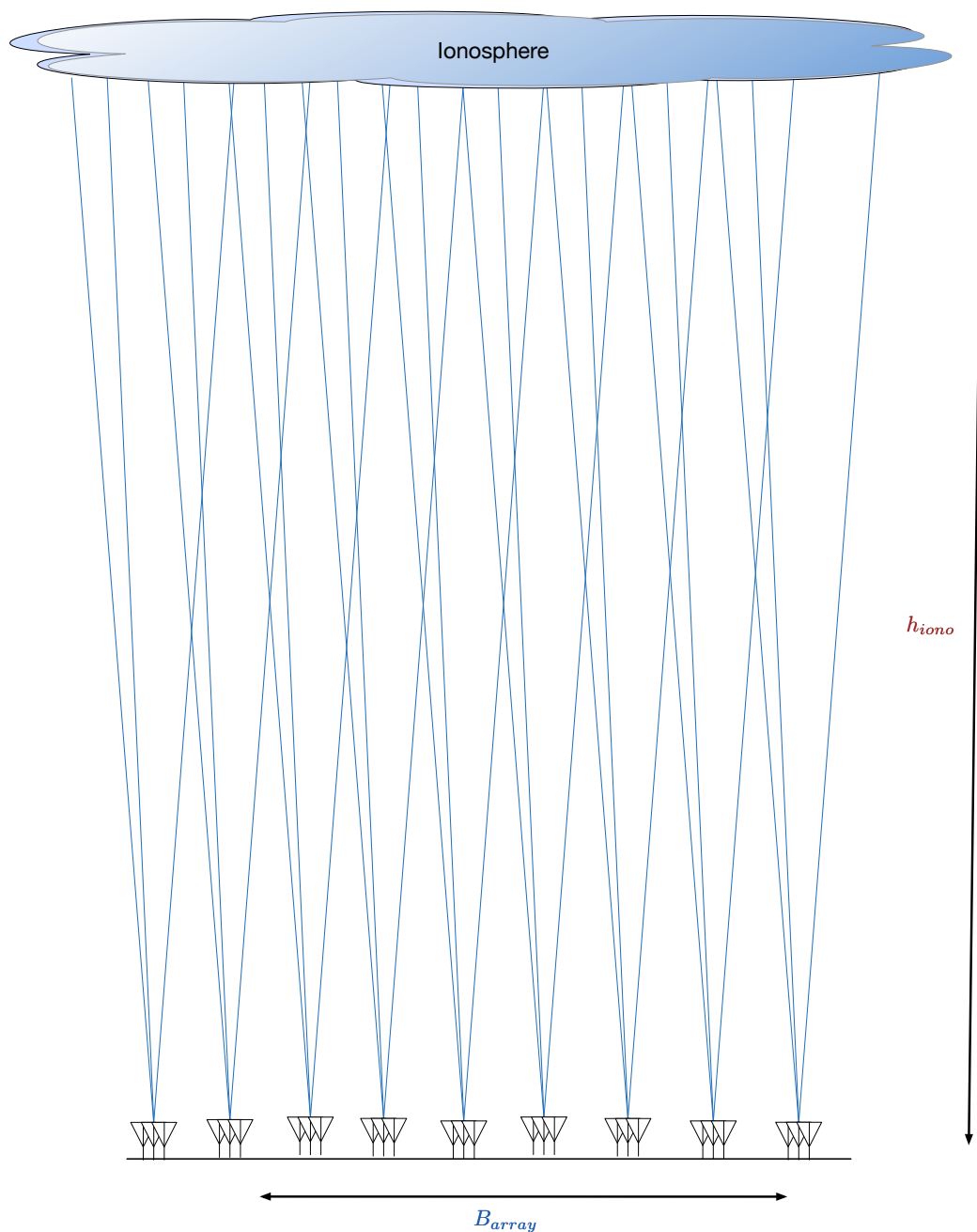
Figure 6 SKA1-Low in Extragalactic foreground imaging including ionospheric screen estimation. To image the full ionospheric phase screen that the EOR is seen through, the array must estimate and correct the phases in that 55km (at 100 MHz) screen requiring baselines up to 55km.

Note that our calibration analysis does not require any particular algorithm to be used though it may suggest some. We do, however, require that the algorithm apply the phase screen corrections without introducing significant errors. The formulation does help in determining upper bounds on the performance of any calibration scheme obeying this requirement.

In Table 1, we show a possible calibration algorithm meeting our requirements. We assume a component capable of determining the phase per point source per station. We then fit to the collection of pierce points to determine the screen. The screen is then propagated to each station for subsequent processing using the AW projection algorithm [RD3].

Table 1 Possible algorithm for calibration and imaging using a phase screen.

1. On all baselines every ~10 minutes
    a. Solve for gain per source per station
    b. Fit parametric DDE model (e.g. station primary beam)
    c. Use AW projection to image
2. On long baselines every 10s
    a. Solve for phase (or ΔTEC) per source per station avoiding use of EOR spacings if necessary
    b. Fit Zernike (or similar) model to obtain smooth phases
    c. Calculate the AW kernels for the halo stations
    d. Use AW projection to image
3. On short baselines every 10s
    a. Calculate the AW kernels for the core stations
    b. Use AW projection to image

To find constraints on the performance of any calibration scheme we must characterise how well the pierce points constrain a model of the phase screen. It is convenient to have a set of basis functions orthogonal on the disk. The Zernike polynomials (Figure 7) are a natural choice for modelling the screen (see also [RD8]). In addition Noll [RD1] shows that the Kolmogorov spectrum can be compactly represented in Zernike polynomials. Our approach is therefore as follows – postulate a method to calculate the phase at each source for each station and then calculate the design matrix A connecting these pierce points to the Zernike polynomial coefficients[2]. The conditioning may be calculated from the singular value decomposition of $A^T A$:

$$A_{J,i} = \sqrt{N_{station}} \frac{S_i}{\sigma_{vis}} Z_J(r_i, \phi_i)$$

(see Appendix B for more details). This approach does not require simulation of a Kolmogorov screen. Instead we need to simulate accurately the sources used for calibration, including position and strength (as modulated by the primary beam). In similar work, Dillon and Parsons [RD10] have used eigenvalue analysis of $A^T A$ to rank HERA configurations. We have applied the same approach to array design but our focus here is on the actual imaging performance in absolute quantitative terms.

---

[2] The SPAM method [RD8] is one existing possibility for the estimation of the pierce point phases.
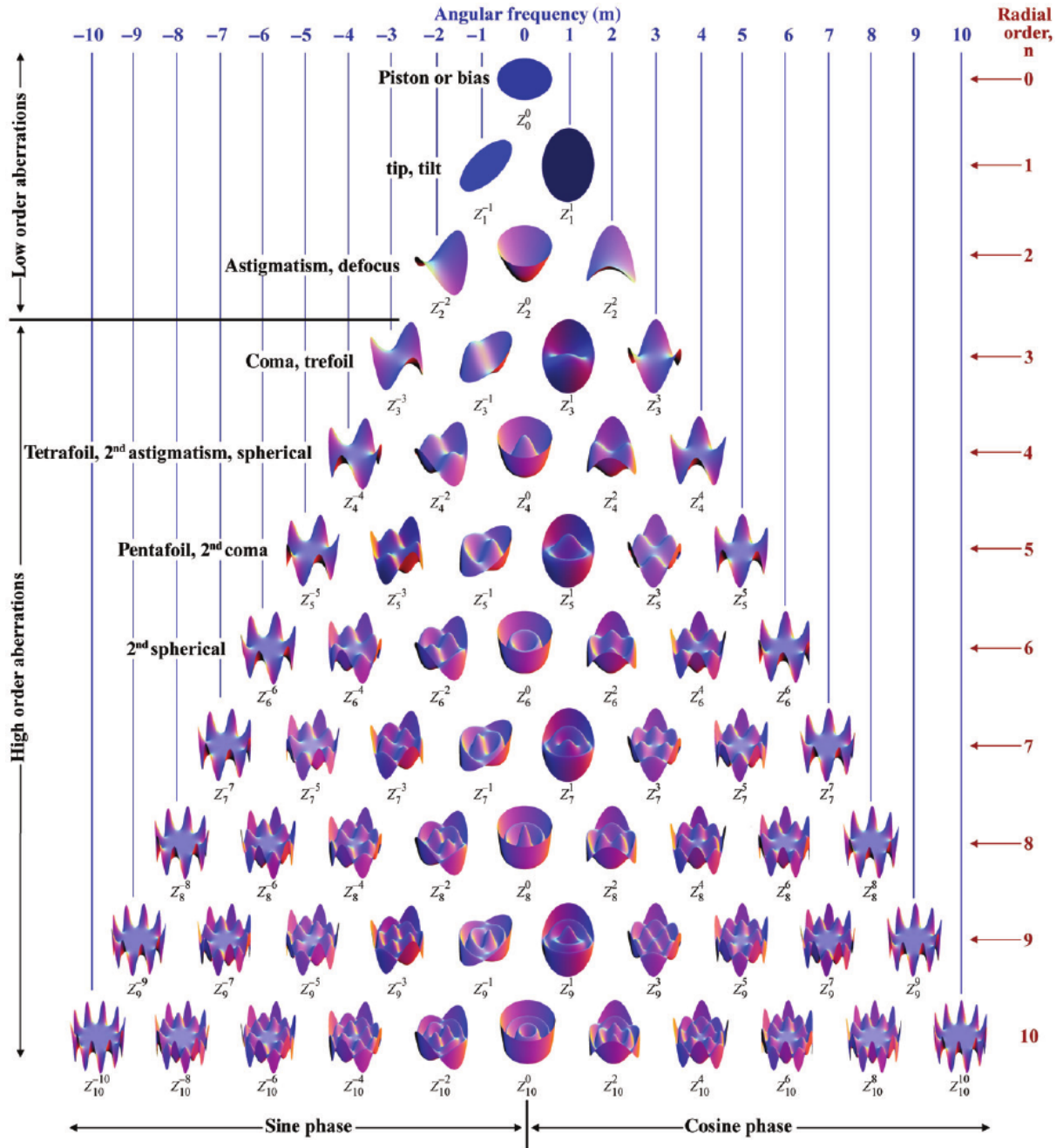
Figure 7 First Zernike polynomials on the disk.

The choice of the Zernike polynomials allows us to estimate the residual phase error[3]. Noll [RD1] shows that the asymptotic form for the residual phase variance across the screen after fitting the first J Zernike polynomials is:

Equation 1

$$\sigma_J^2 \sim 0.2944\, J^{-\frac{\sqrt{3}}{2}} \left(\frac{B}{r_0}\right)^{\frac{5}{3}} \qquad \text{rad}^2$$

---

[3] If we have simulated data, then the Karhunen-Loeve functions could be used, with the advantage that they give a better approximation [RD5]. However, there is no closed form and no equivalent of Equation 1.

where r_0 is the Fried parameter, defined so that the structure function is:

Equation 2

$$D(r) = 6.88 \left(\frac{r}{r_0}\right)^{5/3} \text{rad}$$

Note that the definition of the scale size r_0 in radio astronomy [RD1] differs from that of Fried [RD3]; we follow Fried. The implications of the Kolmogorov turbulence model for optical seeing are well described in a paper on "Adaptive Optics on Large Telescopes" [RD5]. Many of these results carry over directly to our context. The array of stations is analogous to the rubber mirror used in adaptive optics. Both correction schemes operate in the pupil place, and are non-isoplanatic.

We do not know the number of polynomials fit to any particular dataset but the number of non-zero singular values gives an upper bound.

The phase variance in Equation 1 is the mean square across the aperture. Estimating the dynamic range limit arising from this residual phase error requires care. The Perley [RD13] derivation of dynamic range for an array assumes that the station phases are independently drawn from the same distribution. However, in the case of fitting the ionospheric screen, the remaining phase errors are represented by summation of the remaining Zernike terms. Using the Perley [RD13] equations in this circumstance would overestimate the dynamic range. Another way of understanding this subtle point is that the estimation of the pierce points relies upon the independence of the errors on those pierce points.

On any one line of sight, each end of an interferometer sees a phase error of rms σ_J.

Equation 3

$$V(r_1, r_2) = \int e^{i(\phi(r_1) - \phi(r_2))} I(s) e^{2\pi js.(r_1 - r_2)} ds$$

Consider a point source of flux S at the phase centre. The error in the visibility is:

Equation 4

$$\Delta V(r_1, r_2) = S\left[e^{i(\phi(r_1) - \phi(r_2))} - 1\right]$$

The inverse transform of these visibility measurements will give the error in the dirty image:

Equation 5

$$\Delta I^D(s) = \frac{1}{M} \sum_{m=1}^{M} S\left[e^{i(\phi(r_1) - \phi(r_2))} - 1\right] e^{-2\pi js.(r_1 - r_2)}$$

The variance of this can be used to estimate the dynamic range. In the Perley derivations, the phase terms ϕ are drawn from independent distributions and so the terms in the variance average down. In our case, the cross terms must be drawn from Equation 1. For a single

instance of the ionospheric phase screen, the residual phase error will lead to a dynamic range limit of:

Equation 6

$$\Lambda_{iono} = \frac{1}{\sqrt{2}\sigma_J}$$

If we perform a solution every t_{iono} seconds and the total observing time is t_{sky} and the phase errors are uncorrelated and thus average down then the dynamic range limit is:

Equation 7

$$\Lambda_{iono} = 1.304 \, J^{\sqrt{3}/4} \left(\frac{B}{r_0}\right)^{-5/6} \left(\frac{t_{sky}}{t_{iono}}\right)^{1/2}$$

Or, we can estimate t_{sky} as:

Equation 8

$$t_{sky} = t_{iono} \left(\frac{\Lambda_{iono}}{1.304}\right)^{2} J^{-\sqrt{3}/2} \left(\frac{B}{r_0}\right)^{5/3}$$

Mevius [RD1] shows that both the slope and scale of the ionosphere vary substantially. For 100MHz, the Fried parameter $r_0 \sim 44$ km. For this equation, B should be the longest baseline within the primary beam projected on the ionosphere ($B \sim 55$ km). We note specifically that it is not the baselines required for the EOR imaging (typically hundreds of metres).

J is proportional to the number of pierce points i.e. the product of the number of antennas and the number of sources along with an efficiency term related to the distribution of pierce points. These compact forms allow some immediate conclusions can be drawn:

- In self-calibration of a phase screen directly above the array, we only need the value of the phase screen above each given station. As a consequence, the dynamic range increases as the number of stations. If the screen is some distance away, as is the case for the ionosphere, then it is necessary to form a model of the screen to correct lines of sight not probed by a calibrator. In this case, we have shown that the dynamic range only goes as roughly the square root[4] of the number of stations and the number of sources.
- Increasing the number of pierce points by adding more stations is expensive and ineffective – to increase the dynamic range by an order of magnitude the number of stations must increase by a factor of at least 520.
- Increasing the station diameter improves the dynamic range as $D^{5/6}$ since the uncorrected phase error across the ionospheric patch is smaller. Thus some improvement is possible by using larger stations which might necessitate multi-beaming to detect the EOR signal.

---

[4] Actually the $\sqrt{3}/4$ power.

In Figure 8, we show the observing time required to reach dynamic range of 40, 50, 60 dB as a function of highest J recovered.
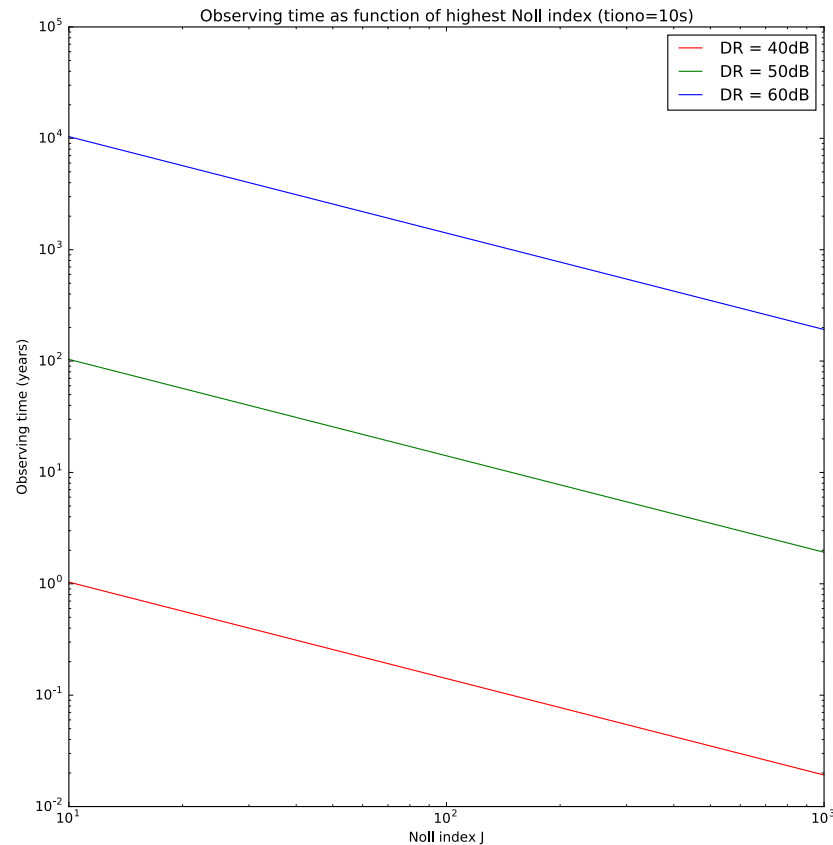
Figure 8 On-sky observing time required to reach specified dynamic range as a function of maximum Noll index recovered (at 100MHz).

In Figure 9 we show the Low configuration from the Baseline Design V2 (proposed)[5]. The pierce points are formed by a convolution of the sources on the sky and the stations on the ground (Figure 10). Note that the requirement on the uniformity of the pierce point distribution is less forgiving of the array configuration than Fourier plane coverage. To get uniform distribution of pierce points it is necessary that either there be many sources (typically >> 10) or the array is uniformly distributed in real space.

---

[5] The halo in the BDV2 proposed design is composed of superstations of 6 stations clustered together. We have assumed that each is separately correlated. The use of superstations has no advantages for ionospheric calibration (compared to dispersed stations) and could be harmful if beamforming is envisaged.
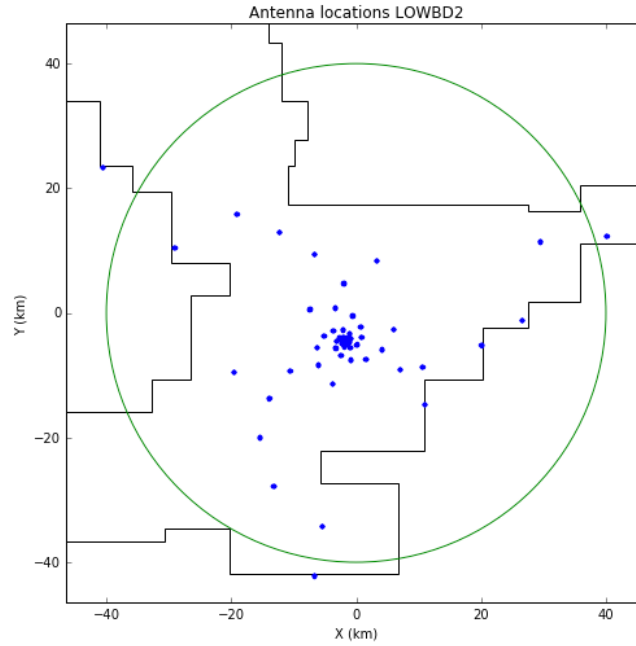
Figure 9 Station locations for Baseline Design configuration. The circle has diameter equal to the target maximum baseline. Each dot in the halo is composed of a "superstation" comprising stations in close proximity.
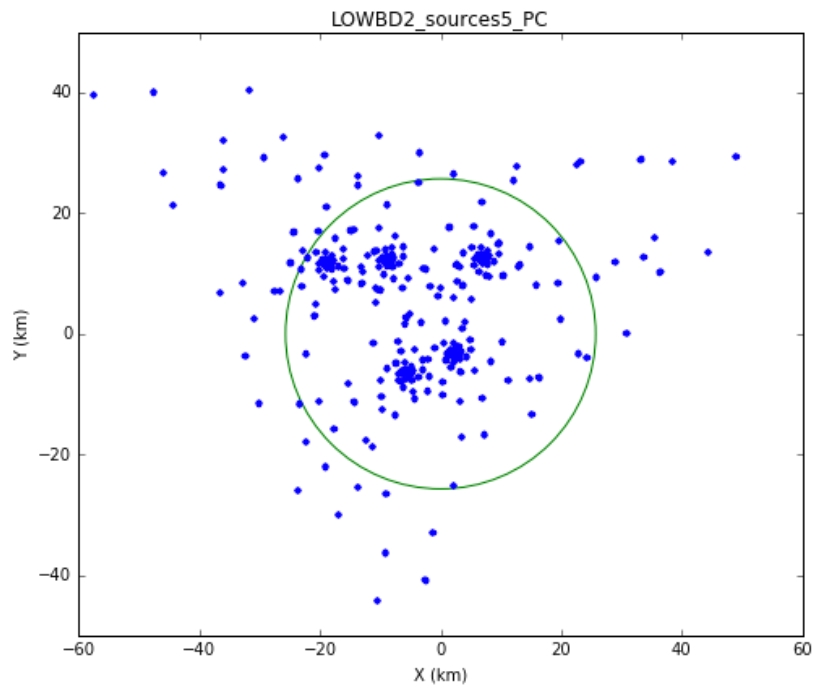
Figure 10 Pierce points for 5 sources and Baseline Design configuration. The ionosphere was assumed to be a thin layer at 300km altitude.

To assess the ability of a given collection of pierce points to constrain the phase screen we must calculate Z the coupling matrix between pierce points and the Zernike polynomials. Performing SVD of Z^TZ yields the singular value spectrum. The singular value spectrum quantifies the inverse variance of the estimate for each singular vector. Each corresponding singular vector is a linear combination of the basis functions (the Zernike polynomials). We

can derive an upper limit for J from the singular value spectrum specifically where the singular value drops below some threshold (e.g. the SNR for a particular singular vector is less than 5).

The steps in this reasoning are presented in Table 2.

Table 2 Logical steps in determining the dynamic range limitation due to incomplete ionosphere phase screen estimation.

Step	Yields	Comment
1. Model phase screen by Zernike polynomials	Orthogonal basis for phase screen	Could use a similar basis
2. Find pattern of pierce points for array design and random source distribution	Distribution of pierce points at the ionosphere	Assumes that we can estimate the phase of each pierce point separately
3. Calculate coupling matrix Z of pierce points to phase screen	Matrix connecting Zernike polynomial estimate to each pierce point	Neglects coupling between pierce point phase estimation
4. Calculate singular value spectrum of coupling matrix Z^TZ.	Ordered list of sensitivity to each singular vector	Shows utility of array and sources in determining phase screen
5. Find cut-off point J in singular value spectrum	Upper limit to highest Zernike polynomial that can be recovered	Shows limit to recovery of fine scales
6. Check if array is coherent	Phase error at cutoff	Must be < 1 for the array to be coherent
7. Find integrated ionospheric phase power not modelled	Residual phase power on unmodeled finer scales	Uses Noll formula
8. Calculate dynamic range limit in solution interval	Instantaneous dynamic range limit due to unmodeled finer scales	
9. Assume $\sqrt{t}$ scaling to obtain necessary integration time	Integration time required to average out residual phase screen errors	Assumes that the phase screen errors integrate down coherently

The fitting process is described in Appendix B. The calculations were been performed using a python notebook and supporting library file.[6]

---

[6] Available from http://github.com/timcornwell/low-calibration.

6. Implications for array configurations

The analysis involves a number of key parameters that can be estimated as function of frequency (see Table 3). These are critical in determining the feasibility of calibration.

Table 3 Key parameters as a function of frequency: Radius of ionospheric patch illuminated, ionospheric coherence time, image noise in ionospheric coherence time (with perfect calibration), visibility noise in ionospheric coherence time, pierce point weight for 1Jy source, pierce threshold, number of sources in ionospheric patch. Note that the jump between 70 and 80 MHz is due to the limited approximation to the sensitivity in the Baseline Design.

Frequency	HWZ	tiono	Image noise (tiono)	Vis noise (tiono)	Weight (tiono)	Pierce threshold (tiono)	Nsources (tiono)
MHz	km	s	Jy/beam	Jy	Jy^{-2}	Jy	
50.0	51.43	5.61	0.0636	32.548	0.5	7.192	0.4
60.0	42.86	6.53	0.0589	30.167	0.6	6.666	0.3
70.0	36.73	7.43	0.0553	28.290	0.6	6.251	0.2
80.0	32.14	8.30	0.0065	3.305	46.9	0.730	6.9
90.0	28.57	9.16	0.0061	3.147	51.7	0.695	6.0
100.0	25.71	10.00	0.0059	3.012	56.5	0.665	5.2
110.0	23.38	10.83	0.0057	2.894	61.1	0.640	4.6
120.0	21.43	11.64	0.0055	2.791	65.7	0.617	4.1
130.0	19.78	12.44	0.0053	2.700	70.3	0.597	3.7
140.0	18.37	13.24	0.0051	2.618	74.7	0.578	3.4
150.0	17.14	14.02	0.0050	2.543	79.1	0.562	3.1
160.0	16.07	14.79	0.0048	2.476	83.5	0.547	2.8
170.0	15.13	15.56	0.0047	2.414	87.9	0.533	2.6
180.0	14.29	16.32	0.0046	2.357	92.1	0.521	2.4
190.0	13.53	17.07	0.0045	2.305	96.4	0.509	2.3
200.0	12.86	17.82	0.0044	2.256	100.6	0.499	2.1

We will concentrate on the 100MHz analysis first. The array, MST solution, and pierce points for LOWBD2 observing 5 sources are given in Figure 11, and Figure 12, and Figure 13.
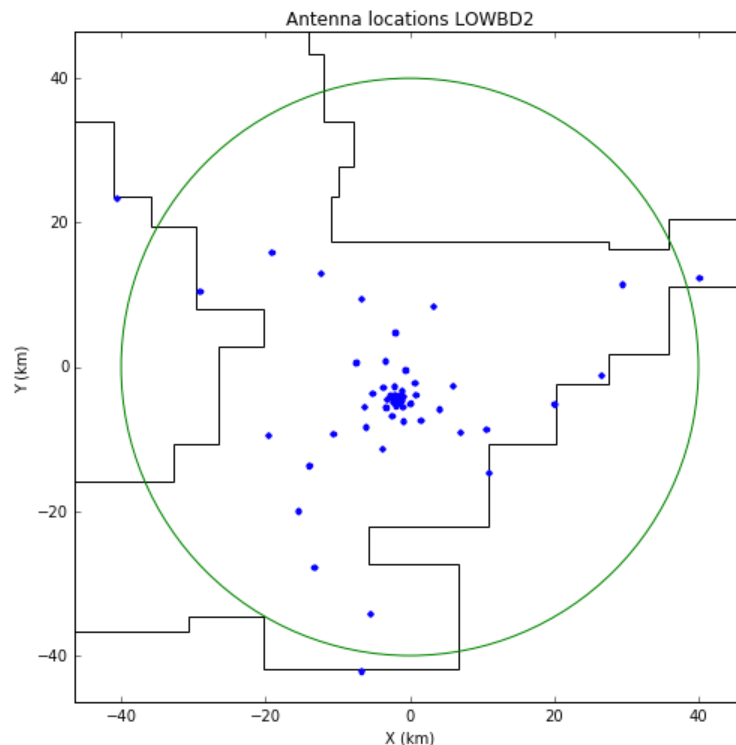
Figure 11 Station layout of current proposed SKA1-LOW configuration. In the halo, each dot consists of six stations arranged in a "superstation".

It is clear that the distribution of pierce points is far from uniform. There are two reasons: first, the use of spiral arms, and second, the use of superstations. The superstations provide islands of pierce points, increasing the signal to noise in that region but at the cost of any sampling elsewhere in pierce point space.

To design a configuration specifically for ionospheric calibration we moved the 346 stations outside the core onto a slightly dithered raster over the Boolardy station ensuring that all stations lay within 40km of the core. The array and one instance of the pierce points are shown in Figure 16 and Figure 17. We show the design matrix A and the singular vectors U in Figure 14 and Figure 15. The singular vectors are the rows of U in this display.
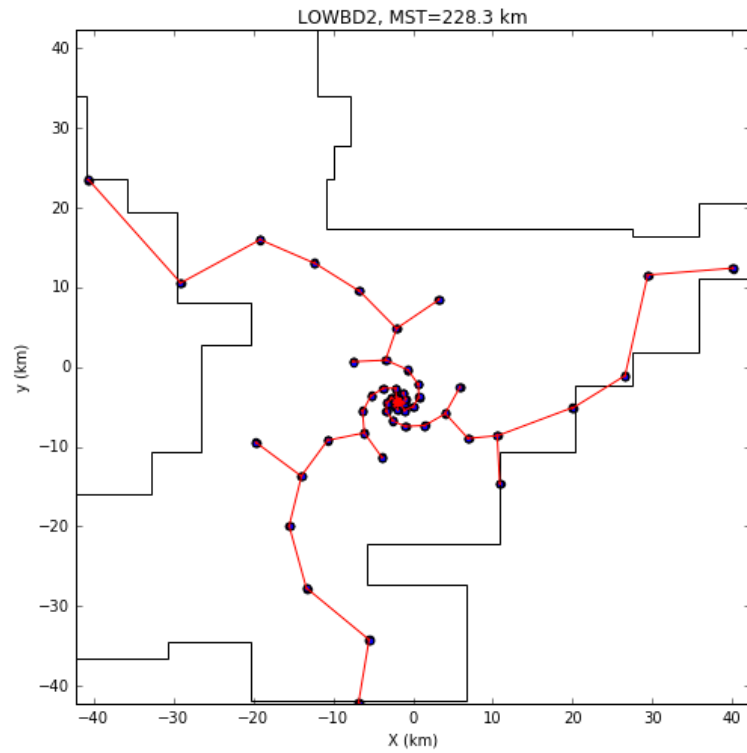
Figure 12 MST solution for LOWBD2
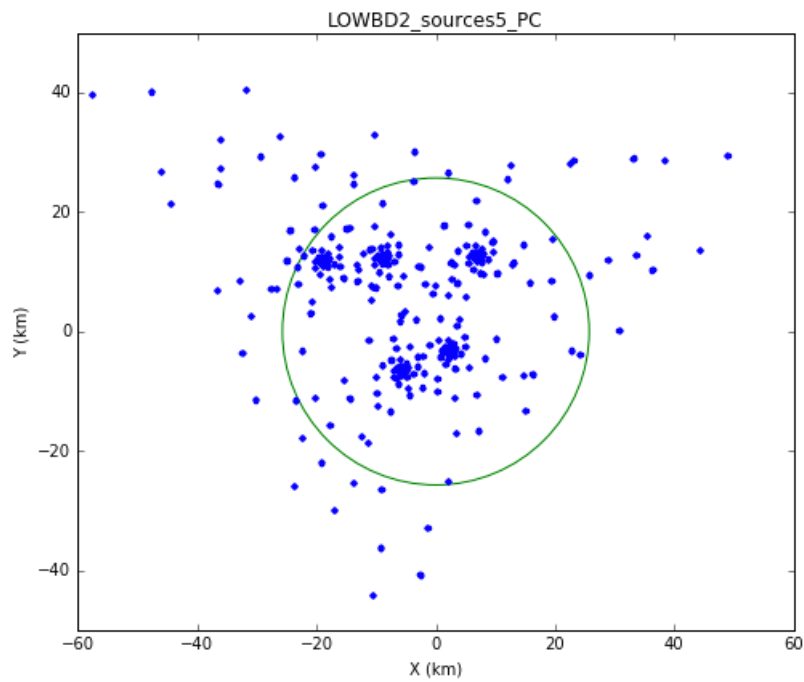
Figure 13 Pierce point distribution for LOWDB2 observing 5 sources.
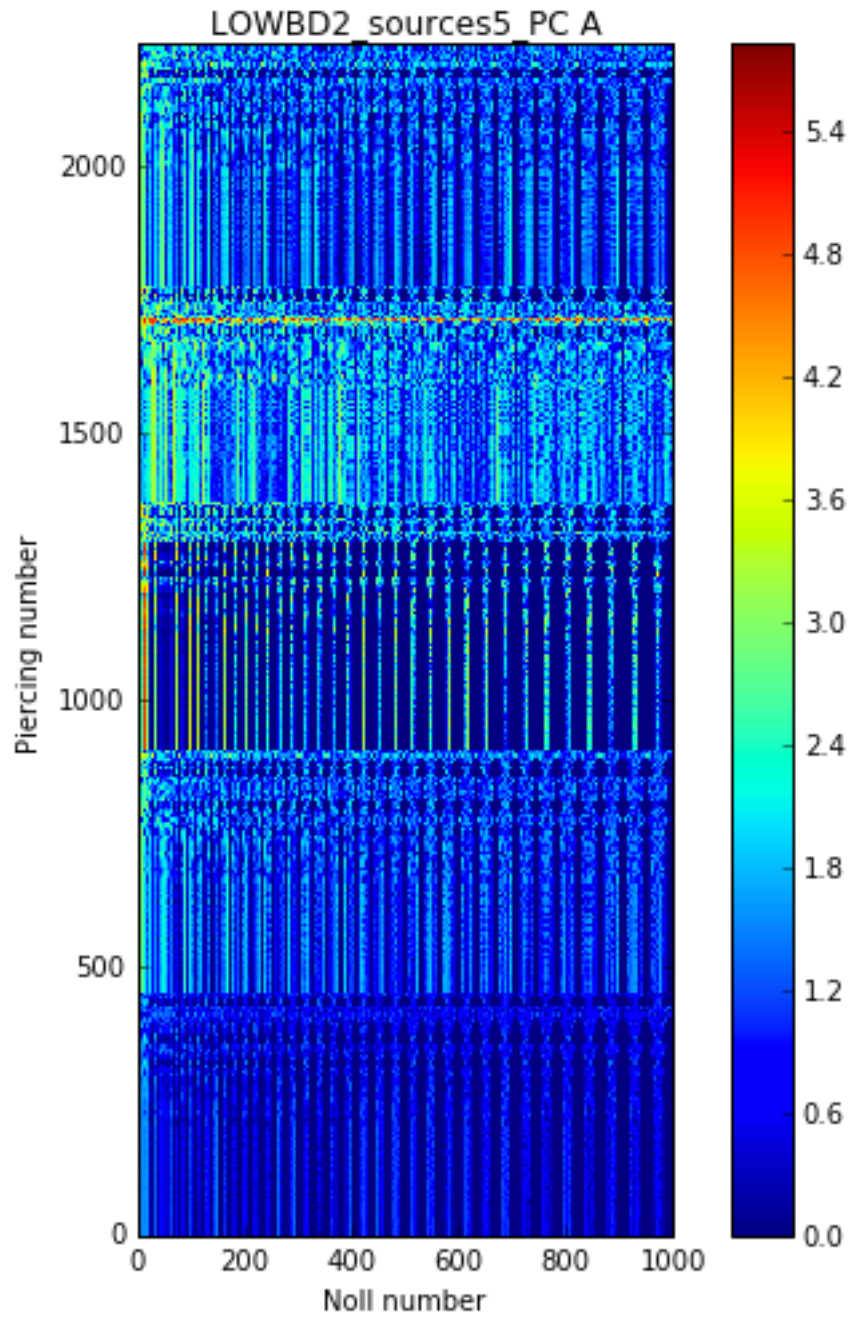
Figure 14 Design matrix A for LOWBD2 observing one realisation of 5 sources at 100 MHz. For clarity the square root is shown.
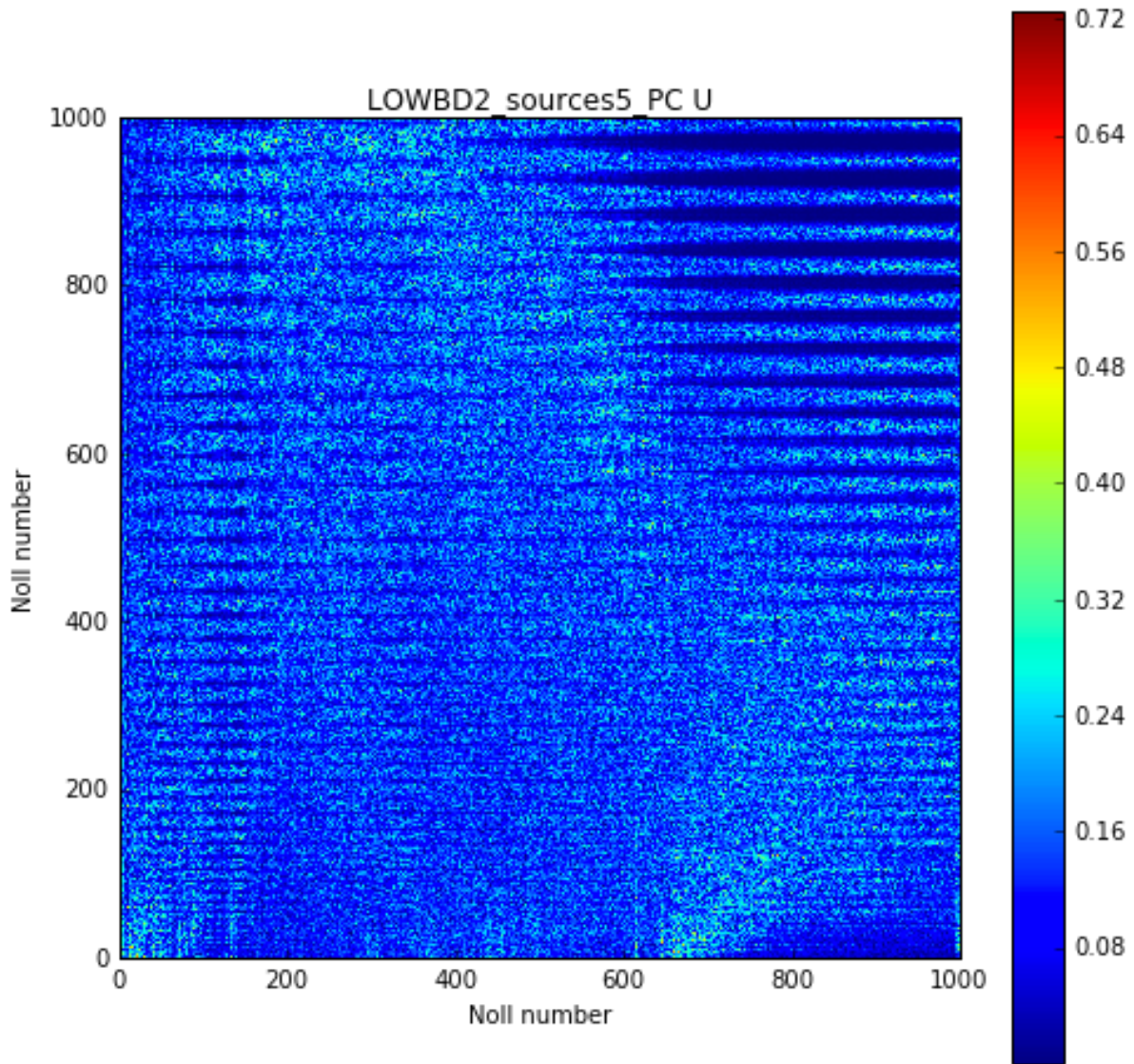
Figure 15 Singular vectors U for LOWBD2 observing one realisation of 5 sources at 100 MHz. The singular vectors are the rows. For clarity the square root is shown.
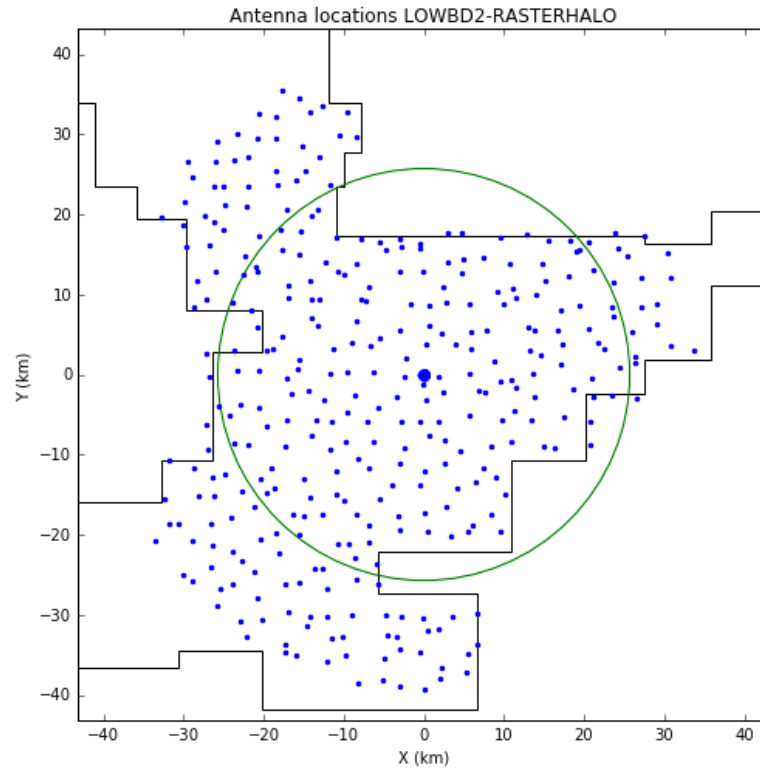
Figure 16 SKA1-LOW-RASTERHALO designed to provide a halo with uniform sampling in real space and a core with good sampling in u,v space. This is not a proposal for a configuration but an excellent configuration for ionospheric calibration.
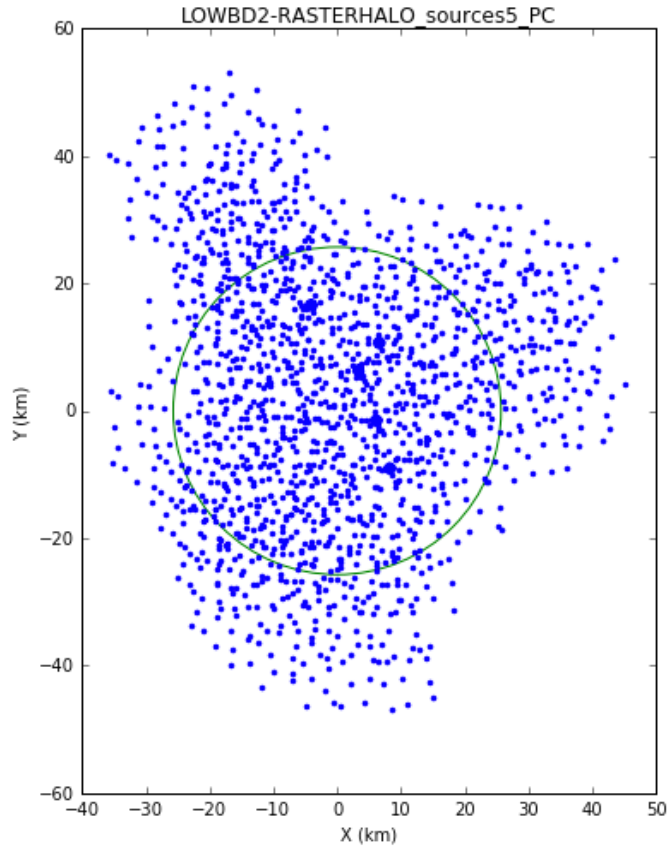
Figure 17 Typical set of ionospheric pierce points for LOWBD2-RASTERHALO. 5 sources were chosen at random locations. The green circle shows the region through which each station sees the sky. This is for one realisation. The results below averaged over 10 realisations. Pierce points outside the circle do not constrain the phase screen estimate.

The LOWBD2-RASTERHALO array is not a viable proposal for construction, mainly because the infrastructure costs are high. In Figure 18, we show the MST solution. The value of the MST, 853 km, is nearly four times that of LOWBD2 (228 km).
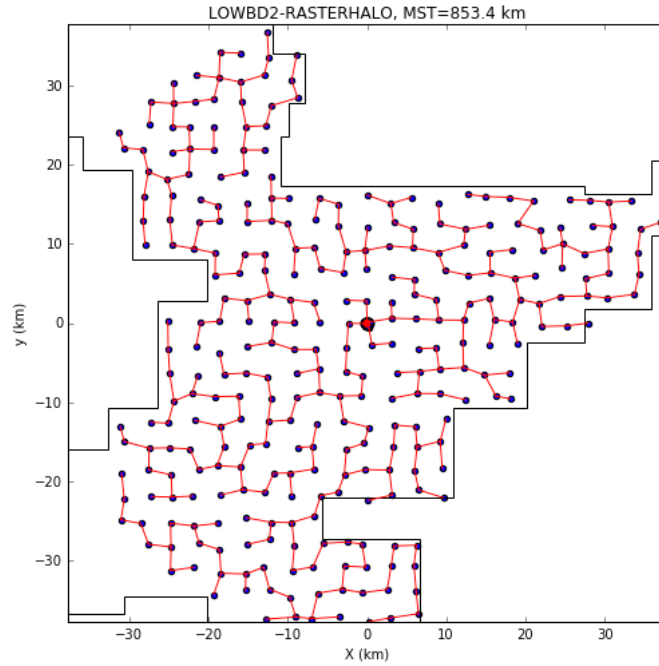
Figure 18 Minimum Spanning Tree for LOWBD2-RASTERHALO. For comparison, LOWBD2 has MST=228 km.

Both the MST and the efficiency of the array at collecting pierce points can be improved by reducing the maximum baseline to the size of diameter of the primary beam projected on the ionosphere (twice the parameter HWZ in Table 3: roughly 55 km at 100MHz for 35m station). The resulting array LOWBD2-RASTERHALO100MHz is shown in Figure 19 and the pierce points are shown in Figure 20. The MST is 630 km, thus much reduced from that of the slightly larger array.
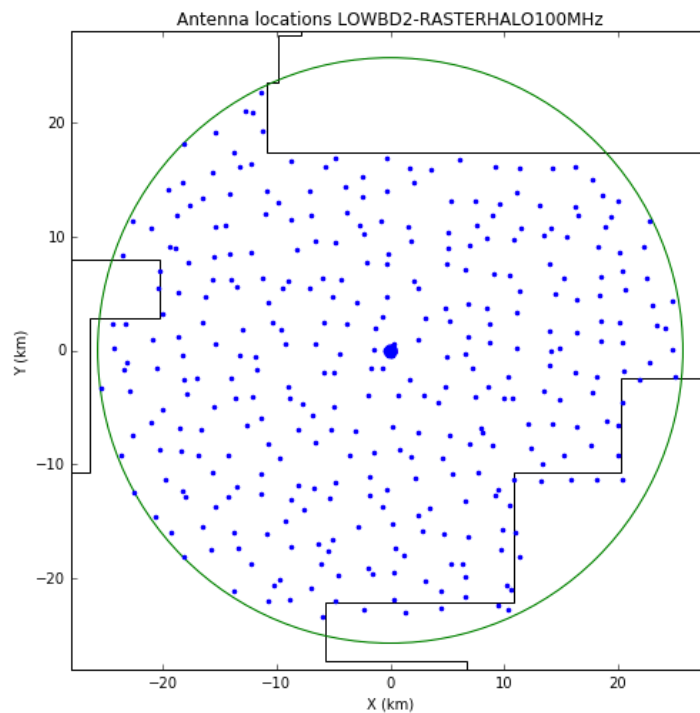
Figure 19 Station layout for LOWBD2-RASTERHALO100MHZ, formed by using a slightly dither raster scaled to match the 100MHz station beam at the ionosphere.
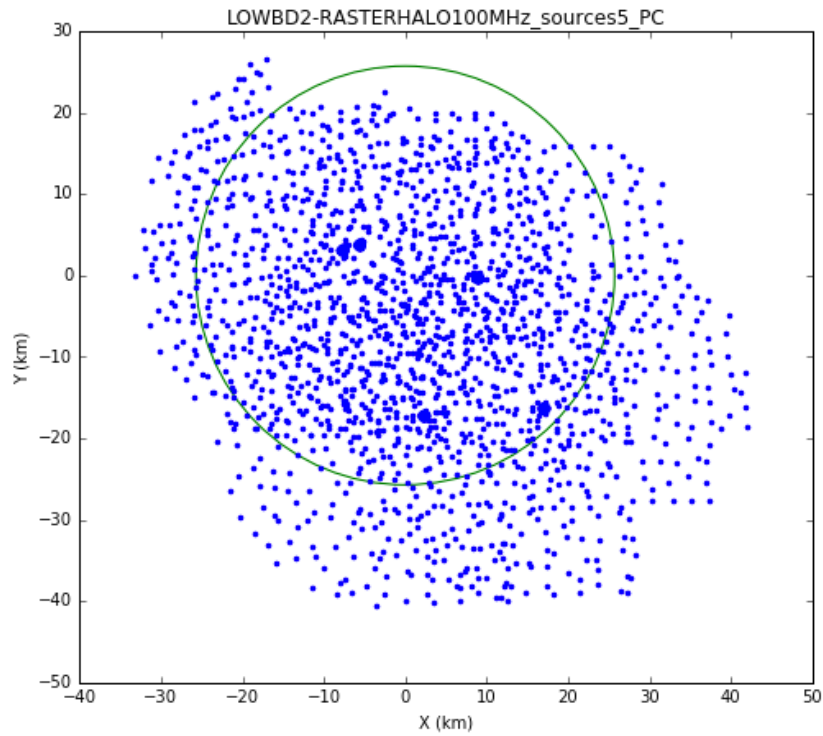
Figure 20 Pierce points for LOWBD2-RASTERHALO100MHz for one realisation of five sources.

We have analysed all these arrays, as listed in Table 4. Figure 21 shows the singular value spectrum for LOWBD2 in the case of 5, 10, 20 sources, averaging over 10 realisations for each. Figure 22, Figure 23, and Figure 24 show the singular value spectra (scaled to SNR) for each array in the case of 5, 10 and 20 sources in the field of view, averaged over ten realisations of the sources.

Table 4 For 5, 10, and 20 sources suitable for calibration at 100MHz: peak SNR of phase screen fitting, maximum J recovered, predicted phase error for 10s integration, DR for 10000h integration, Time on sky for 50dB dynamic range (years), and length of Minimum Spanning Tree (km). To obtain real time, the time on sky has to be divided by the achieved efficiency of observing (not more than 0.3). The typical case in 5 sources (shown with grey background)

Configuration	Nsources per field	Peak	J_{max}	Phase error (10s)	DR (10000h)	Tsky (50dB)	MST
		(SNR)		(rad)	(dB)	(y)	(km)
LOWBD2	5	50.4	59	0.11	41.0	70.7	
	10	79.8	88	0.09	41.8	50.0	228.3
	20	94.8	175	0.07	43.0	27.6	
LOWBD2-CORE	5	53.6	7	0.27	37.0	448.0	
	10	57.9	14	0.20	38.3	245.8	7.9
	20	56.7	26	0.15	39.5	143.8	
LOWBD2-RASTERHALO	5	50.8	63	0.10	41.1	66.8	
	10	53.7	110	0.08	42.2	41.2	775.0
	20	73.1	251	0.06	43.7	20.2	

LOWBD2-RASTERHALO100MHz	5	41.1	36	0.13	40.1	108.5	491.2
	10	65.1	151	0.07	42.8	31.3	
	20	69.0	266	0.06	43.8	19.2	

Some immediate conclusions:

- For all of the arrays and all numbers of source (5, 10, 20), the configuration is always coherent (i.e. the phase error across the array achieved by fitting the phase screen is always less than a radian).
- With 10000h (1.14 years) on-sky integration and a very favourable 20 sources available, LOWBD2 provides dynamic range of 44 dB. Integrating longer to reach the required 50dB is possible but would require 276 years (assuming observing efficiency 0.3). Thus the current configuration cannot image EOR within the lifetime of the telescope.
- Replacing the halo of LOWBD2 with a 40km radius dithered raster helps but at a large increment in infrastructure because of the increased MST.
- The outer parts of the configurations lead to pierce points that are outside the main lobe of the primary beams and therefore cannot be used. Shrinking the halo down to 25.7 km radius improves the efficiency of the array at observing useful pierce points, nearly halves the infrastructure costs, and only increases the integration time slightly. The best-case elapsed time to 50dB dynamic range is 64 years. The disadvantage is that the long baselines needed for structure determination are lost. In addition, the scaling only works at 100 MHz, and for 50 MHz, a larger array is needed.
- The importance of having many sources in the field for calibration is now clear (see e.g. Figure 21 and Table 4). Unfortunately, this will make the effects of other calibration errors more prominent.
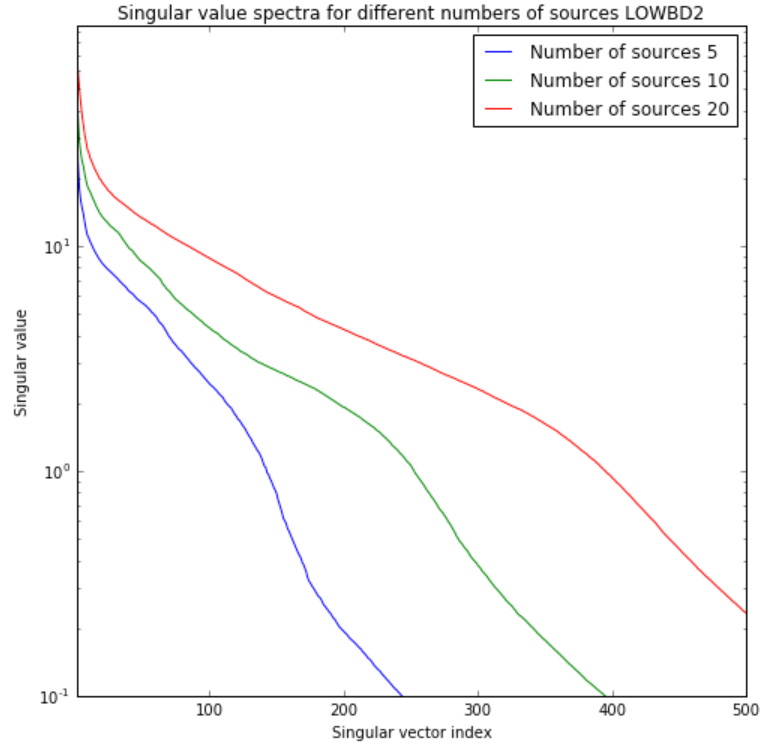
Figure 21 Singular value spectra for array LOWBD2 (the current proposed array) with different number of sources in the field.
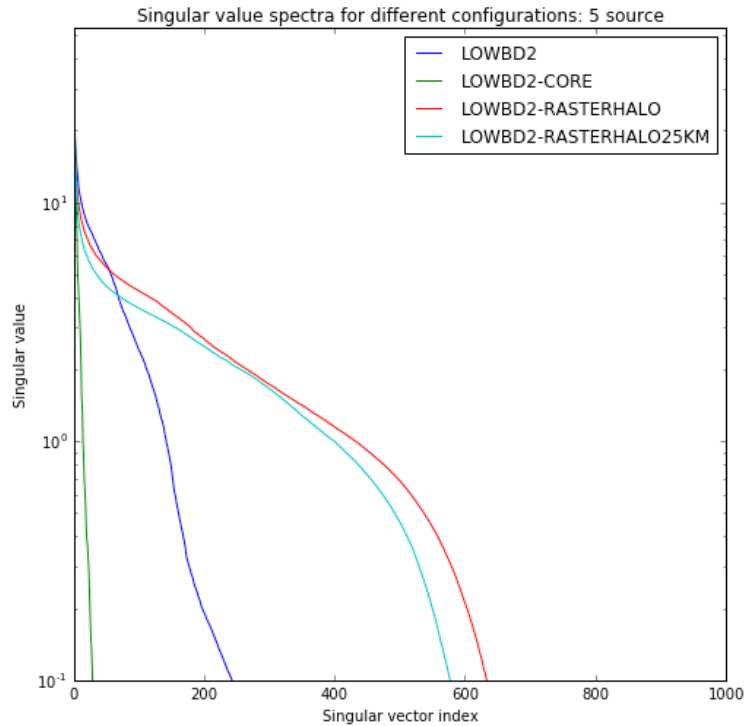
Figure 22 Singular value vs singular value index. This is a typical case at 100MHz: 5 sources were chosen at random locations and with flux obeying logN/logS, and a simple model for the primary beam applied. The results were obtained by averaging over 10 realisations. The singular values are scaled to be approximate SNR. Setting a threshold of SNR > 5 leads to the estimate of J_{max}.
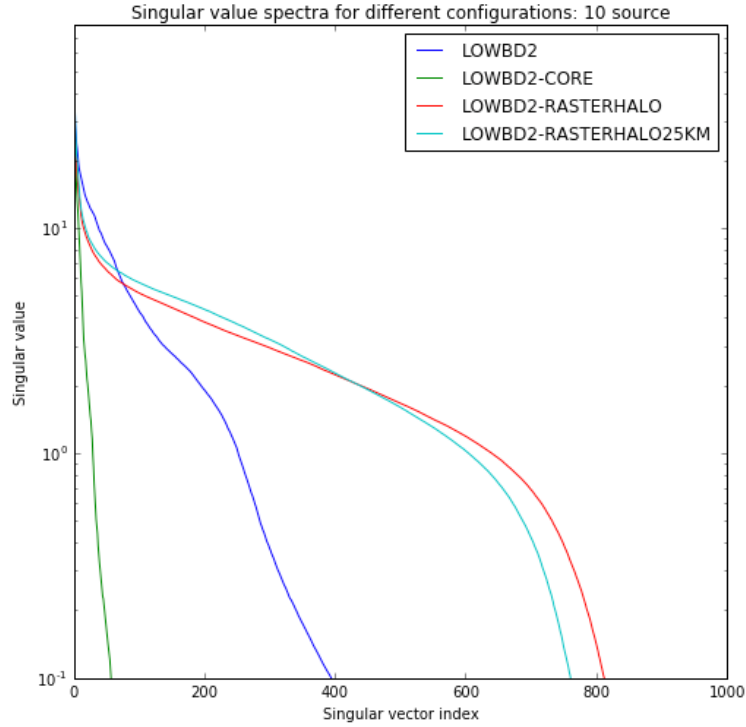
Figure 23 Singular value vs singular value index. This is slightly optimistic: at 100MHz 10 sources were chosen at random locations and with flux obeying logN/logS, and a simple model for the primary beam applied. The results were obtained by averaging over 10 realisations. The singular values are scaled to be approximate SNR. Setting a threshold of SNR > 5 leads to the estimate of J_{max}.
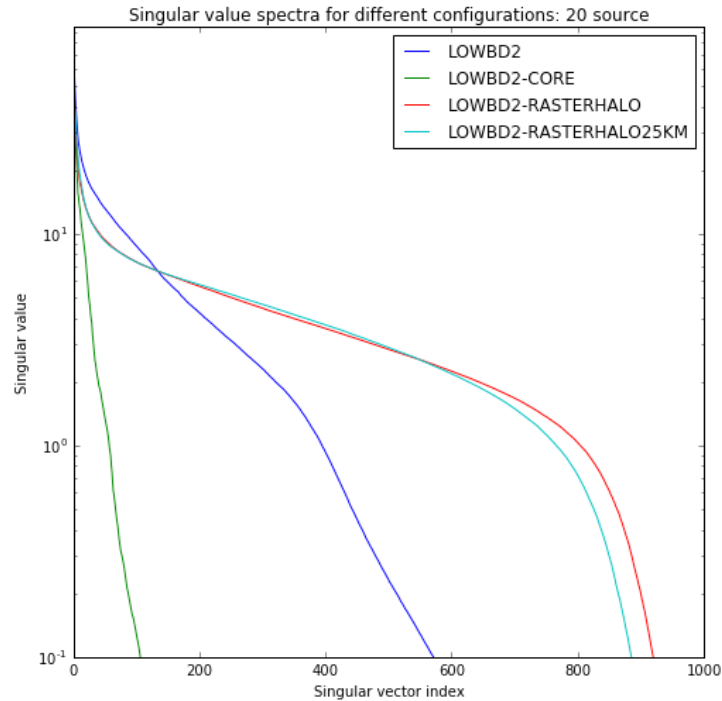
Figure 24 Singular value vs singular value index. This is very optimistic: 20 sources were chosen at random locations and with flux obeying logN/logS, and a simple model for the primary beam applied. The results were obtained by averaging over 10 realisations. The singular values are scaled to be approximate SNR. Setting a threshold of SNR > 5 leads to the estimate of J_{max}.

We have looked at some important cases:

- Doubling the station size to 70m, while fixing the collecting area in the core and halo, brings a substantial improvement but the 5 source case is still totally impossible.

Increasing the station diameter to 105m brings more improvement (see

- Table 5). However, it would be necessary to multi-beam, and then the phasing problem becomes that of reconciling the phase and amplitude frames of the different beams.
- At 50 MHz, there are usually no calibrator sources strong enough to allow estimation of the phase screen in the coherence time (Table 3). Even if we can find a field with one or two sources, the peak SNR does not exceed the threshold 5. Hence, calibration of SKA1-LOW to allow EOR imaging will not be possible.
- At 200 MHz, there are typically only 2 sources per field of view (Table 3). As a consequence, phase screen estimation is not possible, except for an optimised array configuration and a field with an unusual number of sources.

If the size of the primary beam is decreased either by increasing the station size or the observing frequency, it would be necessary to multi-beam, and then the phasing problem becomes that of reconciling the phase frames of the different beams. This should be automatic given the use of a common local sky model.

Table 5 Station diameter increased to 105m, with total collecting area fixed. For 5, 10, and 20 sources suitable for calibration, peak SNR of phase screen fitting, maximum J recovered, predicted phase error for 10s integration, DR for 10000h integration, Time on sky for 50dB dynamic range (years), and length of Minimum Spanning Tree (km). To obtain real time, the time on sky has to be divided by the achieved efficiency of observing (not more than 0.3). The grey shading denotes the typical number of sources.

Configuration	Nsources per field	Peak	J_{max}	Phase error (10s)	DR (10000h)	Tsky (50dB)	MST
		(SNR)		(rad)	(dB)	(y)	(km)
RASTERHALO70M26KM	5	49.0	81	0.05	44.1	16.9	117.9
	10	73.6	163	0.04	45.4	9.2	
	20	82.3	265	0.03	46.3	6.1	
RASTERHALO105M19KM	5	31.0	74	0.04	45.4	9.3	117.9
	10	36.5	101	0.03	46.0	7.1	
	20	51.1	204	0.02	47.3	3.9	

Table 6 200 MHz analysis. Typically, 2.1 source per field. For 2 and 4 sources suitable for calibration, peak SNR of phase screen fitting, maximum J recovered, predicted phase error for 10s integration, DR for 10000h integration, Time on sky for 50dB dynamic range (years), and length of Minimum Spanning Tree (km). To obtain real time, the time on sky has to be divided by the achieved efficiency of observing (not more than 0.3). The grey shading denotes the typical number of sources.

Configuration	Nsources per field	Peak	J_{max}	Phase error (10s)	DR (10000h)	Tsky (50dB)	MST
		(SNR)		(rad)	(dB)	(y)	(km)
LOWBD2	2	61.1	35	0.08	42.5	35.0	228.3
LOWBD2	4	64.8	61	0.06	43.6	21.7	228.3
LOWBD2-CORE	2	22.2	4	0.19	38.4	229.3	7.9
LOWBD2-CORE	4	48.9	10	0.13	40.2	103.7	7.9
LOWBD2-RASTERHALO	2	39.2	16	0.11	41.1	69.0	827.1
LOWBD2-RASTERHALO	4	46.8	29	0.08	42.2	41.2	827.1
LOWBD2-RASTERHALO200MHz	2	37.3	24	0.09	41.8	48.6	343.4
LOWBD2-RASTERHALO200MHz	4	49.8	108	0.05	44.6	13.2	343.4

7. Summary

We have argued that for calibration of the ionosphere over SKA1-Low the phase screen above each station should be modelled in a physically realistic way, and we have investigated the consequences of this view. A minimal model would be smooth and without artificial discontinuities (see also [RD8] for further justification of this view). A Zernike representation is convenient for this purpose, and also has the advantage of a compact, known representation of the Kolmogorov spectrum of turbulence in the ionosphere. There are many reasons why the turbulence model is over-optimistic. There are non-Kolmogorov structures in the ionosphere (i.e. Travelling Ionospheric Disturbances (TIDs), multiple layers [RD9], and vertical tubes [RD12]) that may swamp the turbulence spectrum instantaneously although statistically they drive the turbulence. In that sense, our results are the best case and reality may be worse.

Our analysis follows the spirit of that advanced by Wijnholds, Bregman, and van Ardenne [RD6], with the addition of a more rigorous calculation of the conditioning. We have shown how to evaluate the conditioning of the fitting of Zernike polynomials to the ionospheric phase or TEC screen. We have proposed an approach to estimating an upper bound to the dynamic range achievable. The dynamic range upper bound is a simple function of the uppermost Zernike polynomial measureable given the likely distribution of calibrator sources in the field and the known array configuration. In general, we find previous analysis [RD6] to be significantly over-optimistic in that the time to a certain dynamic range is substantially larger than previously expected.

Our assumptions inevitably err on the side of optimism for calibration. We have neglected other important, mostly damaging, effects:

- The relationship between phase and TEC
- Incomplete knowledge of the sky and antenna performance [RD7]
- Non-linear coupling and bias in the solution for phases of pierce points
- Non-linear coupling and bias in the peeling of bright sources
- Off-zenith primary beam effects
- The effect of sources outside the primary beam
- Other non-ideal behavior of the telescope (see e.g. [RD6])
- The strong variability seen in ionospheric behavior (see e.g. [RD1])
- Vertical structure in the ionosphere (i.e. multiple layers [RD9] and tubes [RD12])
- Algorithms to estimate and apply the phase screens, such as clustered calibration [RD8][RD14]
- Noise arising from the intra-station calibration
- Fresnel effects

Our conclusions are:

- We find that the current proposed SKA1-LOW design (station size and array configuration) leads to observing times longer than the lifetime of the telescope (50 years).
- The pertinent L0 requirement is:
  - *SCI-REQ-18: SKA1-LOW shall provide 50 dB brightness dynamic range at 300 arcsec spatial and 1 MHz spectral resolution to enable EoR imaging and power spectra generation at 50 – 200 MHz (HPSO 1 and 2).*
- With the current proposed SKA1-LOW design and a "golden" case of O(10) sources useful for calibration in 10s, and the efficiency of observing 0.3, meeting requirement SCI-REQ-18 would take 276 years elapsed time.
- Our analysis is for the full resolution of the SKA1-LOW core: about 1 degree. This is an order of magnitude higher than specified in SC-REQ-18. However, Equation 6 says that the dynamic range is independent of the resolution. However, this is a subtle point and one worth further exploration, perhaps by simulations.
- To illustrate the limits of full field calibration, we analyse an extreme array configuration, LOWBD2-RASTERHALO, where the 346 SKA1-Low outer stations are re-deployed in a raster across the Boolardy station. We have chosen this approach since the outer stations should uniformly sample real space rather than u,v space as is normal (and as the SKA1-LOW current proposed configuration seems to have been designed). This provides an upper bound of the performance, should the configuration be decided in this way. However, none of the configurations analysed would permit imaging of the EOR within the lifetime of the telescope.
- Matching the array size to the footprint of the station beam on the sky, brings improvement but it does not fix the problem. LOWBD2-RASTERHALO100MHz observing 20 sources, takes 90 years for efficiency 0.3! For a more realistic number of sources, 5, the time taken is 464 years.

Larger stations should be investigated (

- Table 5). Larger stations (e.g. 70m and 105m) do help but would require multi-beaming and then the problem is shifted to establishing a common phase frame for all beams, and reconciling the primary beams.
- Our phase screen fitting approach will not work at 50 MHz since the density of sufficiently strong calibrators is too low. This means that imaging the entire field at 50 MHz is not possible. Imaging of limited size regions will be possible.
- Our analysis is highly sensitive to the frequency and the constraints on station locations imposed by the Boolardy station. For lower frequencies, the pierce points project down to the ground considerably outside the boundary of Boolardy station.

Our conclusions are perhaps surprising in the context of high quality images from LOFAR. However, in this analysis we have explicitly constrained the phase model to be continuous over the entire field. On smaller angular scales where a smooth, continuous model is not sought, the dynamic range limits do not apply. Hence the success of particular partial-beam calibration techniques (such as Facet Calibration [RD6]) does not give any assurance that the full-field calibration required for EOR imaging will be successful.

8. Next steps

We have presented a theoretical framework for understanding and predicting the scientific performance of low frequency arrays such as LOFAR and SKA1-LOW. While this model undoubtedly could be (and should be) improved, our use of it has allowed one main concrete conclusion: imaging the EOR with the current proposed SKA1-LOW will take decades. This means that the SKA1-LOW design (station and array configuration) should be revisited with a view to improving performance for EOR imaging.

However, the trade-off space for design is limited.

- The best array configurations solutions for different frequencies differ.
- The configuration could be split into three separate parts for three different purposes:
  - Imaging EOR (requiring 1km baselines at 100MHz),
  - Imaging the calibrators accurately by synthesis (by LOFAR experience requiring at least 80km baselines), and
  - Imaging the ionospheric phase screen within the ionospheric coherence time (requiring 55km baselines at 100MHz, see Appendix A).
- Increasing the number of stations in the halo by an order of magnitude does bring the observing time down to less than a decade.
- Larger station size (such as 70m or 105m) would ease the estimation of the phase screen but at the cost of requiring multi-beaming and phase connection between the beams to measure the EOR signal. The EOR signal is still spread over the same extent in the ionosphere and this region must still be made phase-stable for imaging to be possible. In addition, the station beams must be made consistent enough to allow the requisite dynamic range. Hence, it is inevitable that to reach the requisite dynamic range we must solve both ionospheric phase calibration and station primary beam calibration.
- Changing the size and number of the stations has consequences throughout the system. For example, the backplane in the station may be insufficient, and the station correlator will become substantially larger. Also, larger stations will require multiple beams to be processed in the array correlator and SDP. Hence a system-wide analysis is needed.

We have not yet performed a formal sensitivity analysis to determine the result of changing some key assumptions (such as the ionospheric coherence time, the number of sources suitable for pierce point phase estimation, or the achieved receiver performance at 100MHz). This is especially important given the large degree of variability in the ionosphere [RD2]. Selecting data based on large r_0 may be more efficient than averaging many smaller r_0 cases.

The framework presented here can and should be tested via simulations and via appropriate analysis of LOFAR observations. For example, the effect of a discontinuous phase screen model on EOR signal should be evaluated.

Introducing the phase screen via the pierce points is straightforward but artificial. A more natural approach would be to develop an A-kernel that was parametrized appropriately and work with that directly. This algorithm would be conceptually simpler and free of the coupling effects undoubtedly present in the pierce point route. However, and probably inevitably, the computational load will increase.

The risk of designing SKA1-LOW without performing the types of analysis described in this memo is that the telescope will not achieve one of the two key science goals within an acceptable timeframe. Obviously, the science goals should be examined carefully to check the dynamic range specification. However, even if the dynamic range goal is lowered to, say, 45 dB, thus decreasing the observing time by an order of magnitude, we should not lose sight of the difficulty of imaging EOR.

The bottom line is that imaging a very weak, diffuse signal seen with a foreground of bright sources and through a turbulent, rapidly variable ionosphere is very difficult. This should not surprise us. The limited number of sources available for phase estimation in the ionospheric coherence time precludes sufficiently accurate estimation of the phase screen, as a consequence the integration to high dynamic range takes an unacceptably long time (longer than the designed lifetime of the telescope). Our results suggest that the pathways to success are limited and require a fresh look at the telescope design, including station size and array configuration.

Acknowledgements

Appendix A Geometric constraints on array properties

A station with diameter D sees radiation that has passed through a region of radius approximately $h_{iono}\frac{2\lambda}{D}$ (FWZ) in the ionosphere. The radius of piercing points is then:

Equation 9

$$r_{piercecore} = h_{iono}\frac{2\lambda}{D} + r_{core}$$

Equation 10

$$r_{piercehalo} = h_{iono}\frac{2\lambda}{D} + r_{halo}$$

If the station diameter is 35m and the height of the ionosphere is 300km then the radius of the footprint of the station in the ionosphere is $r_{piercecore}$ =100km (6m) 50km (3m).

In order for the long baselines to be measuring any sources projected into the same ionospheric region as the core the radius of the stations from the core must be not greater than the twice the radius of piercing points (see Figure 25). To ensure sufficient overlap:

Equation 11

$$r_{halo} < h_{iono}\frac{2\lambda}{D}$$

Alternatively if r_{halo} is fixed by some argument the relationship could be used to constrain the station diameter.

Equation 12

$$r_{halo}D > 2\lambda h_{iono}$$

To ensure complete overlap $r_{halo} = 0$ but longer baselines are needed to model the compact sources. There are multiple possible ways of calculating the maximum baseline required.

- Experience with LOFAR (M. Brentjens) says that an order of magnitude increase in baseline over the core is required. Since the core is roughly 3km in radius the halo should be about 30km in radius. In this case $D > 102m$ (6m) $D > 50m$ (3m).
- If the station diameter is fixed at 35m (as in the Baseline Design) then $r_{halo} > 51km$ (6m) $r_{halo} > 25km$ (3m).
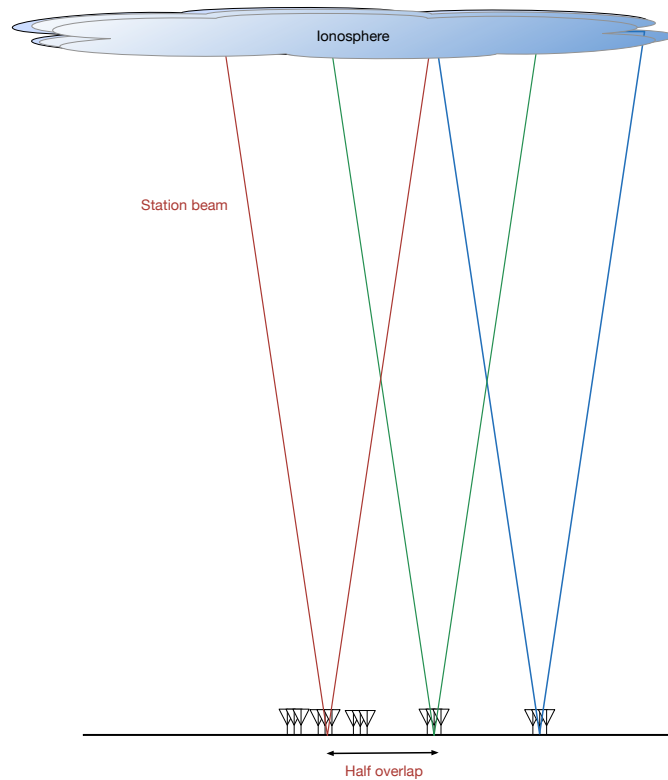
Figure 25 Overlap of halo station beam with core station beam. The radius of the halo must be less than the radius of the piercing region. If Red is measuring EOR then Green provides useful information on the ionosphere that Red sees but Blue does not. The Green stations are only needed to (a) provide a high resolution sky model (b) aid in the estimation of the ionosphere.
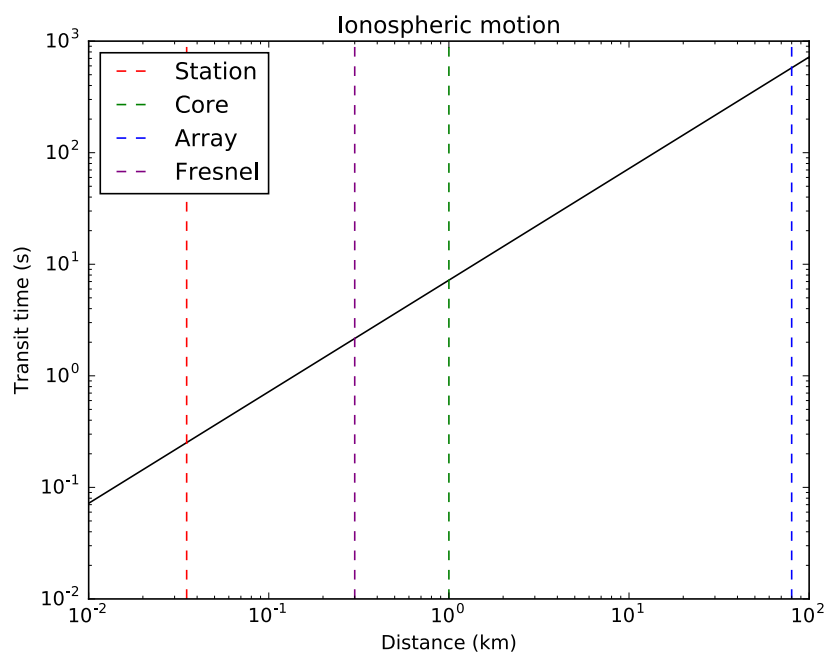

Figure 26 Timescale for ionosphere at 100 MHz moving at 500km/h to translate a given distance.

Appendix B Zernike polynomial analysis

The Zernike polynomial estimate is a straightforward application of linear algebra.

Equation 13

$$AX = Y$$

The unknown phase screen in the Zernike polynomial basis is X, A is the measurement matrix, and Y is the measurements.

The elements of the design matrix A are:

Equation 14

$$A_{J,i} = Z_J(r_i, \phi_i)$$

where (r, ϕ) are the coordinates of the pierce point in polar coordinates.

To account for noise, we introduce the diagonal array W which has diagonal elements equal to inverse variance of the input phase errors. If the source strength is S, the visibility noise in the integration time is σ_{vis}, the number of stations is $N_{station}$, then:

Equation 15

$$W_{i,i} = N_{station}\left(\frac{S_i}{\sigma_{vis}}\right)^2$$

In reality, the phase error estimates are correlated since some degrees of freedom have been absorbed in the pierce point phase estimation. For the large number of stations considered here the effect is small.

We define a matrix:

Equation 16

$$w = \sqrt{W}$$

We then absorb w into A:

Equation 17

$$((wA)^T wA)X = (wA)^T Y$$

So finally, we have that the design matrix A has elements:

Equation 18

$$A'_{J,i} = \sqrt{N_{station}}\frac{S_i}{\sigma_{vis}}Z_J(r_i, \phi_i)$$

Applying straightforward SVD to $A'^T A$ yields $U\Sigma V^h$. We analyse the singular value spectrum Σ to understand the conditioning of the weighted least squares problem.