ANTICIPATED SKA1 HPC REQUIREMENTS

Document number SKA-TEL-SKO-0000941
Document Type .... REP
Revision 02
Author .... SKAO Science Team
Date 2019-03-04
Document Classification .... UNRESTRICTED
Status .... Released

Name	Designation	Affiliation	Signature	
Authored by:				
SKAO Science Team p.p. Robert Braun	Science Director	SKAO	Robert Braun	
			Date:	2019-03-05
Owned by:				
Robert Braun	Science Director	SKAO	Robert Braun	
			Date:	2019-03-05
Approved by:				
Robert Braun	Science Director	SKAO	Robert Braun	
			Date:	Mar 5, 2019
Released by:				
Joseph McMullin	Programme Director	SKAO	J P McMullin	
			Date:	2019-03-06

DOCUMENT HISTORY

Revision	Date Of Issue	Engineering Change Number	Comments
01	2018-05-21	-	First release
02	2019-03-04		Second Release

DOCUMENT SOFTWARE

	Package	Version	Filename
Word processor	MS Word	Word 2007	SKA-TEL-SKO-0000941-02_Anticipated_SKA1_HPC_Req
Block diagrams			
Other			

ORGANISATION DETAILS

Name	SKA Organisation
Registered Address	Jodrell Bank Observatory Lower Withington Macclesfield Cheshire SK11 9DL United Kingdom Registered in England & Wales Company Number: 07881918
Fax.	+44 (0)161 306 9600
Website	www.skatelescope.org

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIP....................Advanced Instrumentation Programme
AIPS ....Astronomical Image Processing System
ALMA................Atacama Large Millimetre Array
ATM..................Atmospheric Transmission of Microwaves
BD ....SKA1 Baseline Design
CASA................Common Astronomy Software Applications
CDR..................Critical Design Review
DD ....Direction Dependent (calibration)
DI......................Direction Independent (calibration)
DM....................Dispersion Measure
ECMWF............European Centre for Medium-range Weather Forecasts
EMI ....Electromagnetic Interference
EoR ....Epoch of Reionisation
FRB ....Fast Radio Burst
FWHM ....Full Width Half Maximum
GMRT...............Giant Metre-wave Radio Telescope
GPU..................Graphical Processing Unit
HPC..................High Performance Computing
HPSO ....High Priority Science Objective
ISW...................Integrated Sachs Wolfe effect
JVLA.................Jansky Very Large Array (or simply VLA)
LOFAR ....LOw Frequency ARray
NIP ....Non-Image Processing
PDR..................Preliminary Design Review
PFlops ....Peta Floating Point Operations per Second
PSF ....Point Spread Function
PWV ....Precipitable Water Vapour
RFI....................Radio Frequency Interference
RM....................Rotation Measure
RMS ....Root Mean Square
SED ....Spectral Energy Distribution
SEFD................System Equivalent Flux Density
SKA ....Square Kilometre Array
SKAO ....SKA Organisation
VLA...................Very Large Array
VLBI..................Very Long Baseline Interferometry

1 Introduction

1.1 Purpose of the Document

The purpose of this document is to provide estimates of the computing needs associated with a range of use cases for the SKA1 telescopes.

1.2 Scope of the Document

In this document the parametric model of compute costs developed by the Science Data Processor design consortium is used to estimate the HPC requirements of a range of science use cases. Rather than making use of default parameter settings within the model, the parameters are tuned to match the specific use cases. Once the needs of individual experiments are quantified, the mix of experiments that can be accommodated with some specific high performance computing deployment scenarios are investigated. **It must be noted that there are still serious uncertainties, by factors of many, regarding the HPC requirements as outlined in Section 4.2.**

2 References

2.1 Applicable Documents

The following documents are applicable to the extent stated herein. In the event of conflict between the contents of the applicable documents and this document, **the applicable documents** shall take precedence.

[AD1] Applicable Document 1
[AD2] Applicable Document 2

2.2 Reference Documents

The following documents are referenced in this document. In the event of conflict between the contents of the referenced documents and this document, **this document** shall take precedence.

[RD1] SKA-TEL-SDP-0000040, "Parametric Models of SDP Compute Requirements"
[RD2] Braun, R., 2013, A&A 551, A91, "Understanding synthesis imaging dynamic range"
[RD3] SKA-TEL-SKO-0000641, "SKA1 Error Budgets"
[RD4] SKA-TEL-SKO-0000818, "Anticipated SKA1 Science Performance"
[RD5] Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
[RD6] SKA-TEL-SDP-0000038, "SDP System Sizing"
[RD7] SKA-TEL-SKO-0000007, "SKA1 Level 0 Science Requirements"
[RD8] Bonaldi, A, et al. 2018, MNRAS, "The Tiered Radio Extragalactic Continuum Simulation", https://arxiv.org/abs/1805.05222

3 SKA1 Computational Needs

3.1 Science Data Processor Parametric Model

The SKA1 SDP design consortium have developed a parametric model that allows estimation of computational needs [RD1] to ingest (with the “Ingest” pipeline), calibrate (with “ICAL”) and generate both image and non-image data products (with one of the “DPrep” variants). It should be noted that the parametric model used here is only the visibility-based model. The NIP group have a parametric model for the time-domain processing within SDP, but since it is so much less demanding than the visibility processing it is not included in the discussion here. The calculation is tied to the specification of a number of key parameters. Many of the parameters follow from the basic telescope attributes such as station/dish size, number and configuration, while others rely on an understanding of the extragalactic source population.

The most important adjustable parameters of the basic telescope description are:

- B_{Max}, the maximum baseline that determines the angular resolution of data products.
- ν_{Min} and ν_{Max}, the minimum and maximum observing frequency.
- N_{PP}, the number of polarimetric data products (equals four for a full polarimetric calibration).
- T_{Point}, the total integration time that will ultimately be acquired for a given pointing.

The celestial source population must be considered because under essentially all circumstances it is necessary to utilise what is termed “self-calibration” in order to generate data products that achieve the theoretical thermal noise level rather than an artificial noise floor that is a consequence of imprecise instrumental calibration. For calibration purposes, the most important properties used in computational cost estimation are:

- N_{Ateam}, the number of sources that occur within the far side-lobes of the station/dishes within the 2π sr that are above the horizon at any time that must be “de-mixed” from the visibilities within the Ingest step of the processing. These are sometimes referred to as the “A-team” sources, since many have traditional names involving the constellation in which they occur followed by an alphabetical designator, with “A” typically being the brightest.
- N_{Source}, the number of source components that must be included within the ICAL self-calibration processing step.
- $N_{SelfCal}$, the number of self-calibration iterations needed to achieve the necessary quality of calibration within ICAL.
- N_{Major}, the number of major deconvolution cycles needed within each iteration of the ICAL process as well as during the production of image data products with DPrep.
- N_{fout}, the number of spectral channels for which calibration solutions are determined within ICAL as well as the number of final output spectral data products with DPrep.
- $N_{Ipatches}$, the number of Ionospheric patches for which complex gain solutions are determined.

3.2 Parametric Models of Celestial Source Populations

Attributes of the celestial source population are discussed at some length in [RD2]. The depth within the source population that must be modelled within the self-calibration process, so as not to unduly degrade the data product quality in deep integrations, where T_{Point} = 1000 hours, is documented in RD3.

It is vital to acknowledge from the outset that all relevant parameters for performance calculations will vary significantly with the observing frequency. As discussed in detail within RD4, the largest fractional bandwidth, $(\nu_{Max} - \nu_{Min})/\nu_C$, that can be usefully combined within a single image should in most cases be constrained to be less than about 0.3. From this perspective, the entire frequency coverage of SKA1 has been organised into a sequence of sub-bands defined in Tables 1 and 2 of RD4. We will refer to these as Low sb1 – sb6 and Mid sb1 – sb5 and sb9 – sb12 as shown in Table 1 below. In the case of SKA1-Low, the six sub-bands together cover the full accessible frequency range of 50 – 350 MHz. In the case of SKA1-Mid, sb1 – sb3 are covered with Band 1, sb4 – sb5 with Band 2, sb9 – sb10 with Band 5a and sb11 – sb12 with Band 5b as shown in Table 1.

ν_{min} (GHz)	**ν_c (GHz)**	**ν_{max} (GHz)**	**Sub-band**	**Band**	**σ_C (μJy/Bm)**	**θ'_{min} (“)**	**θ_{min} (“)**	**θ_{max} (“)**	**θ'_{max} (“)**
0.050	**0.060**	**0.069**	Low sb1		163	16.4	23.5	1175	3290
0.069	**0.082**	**0.096**	Low sb2		47	11.9	17.0	850	2379
0.096	**0.114**	**0.132**	Low sb3		26	8.6	12.3	614	1719
0.132	**0.158**	**0.183**	Low sb4		18	6.2	8.9	444	1244
0.183	**0.218**	**0.253**	Low sb5		14	4.5	6.4	321	899
0.253	**0.302**	**0.350**	Low sb6		11	3.3	4.6	232	650
0.35	**0.41**	**0.48**	Mid sb1	B1	16.8	1.015	2.031	270.8	541.6
0.48	**0.56**	**0.65**	Mid sb2	B1	8.1	0.745	1.489	198.6	397.2
0.65	**0.77**	**0.89**	Mid sb3	B1	4.4	0.546	1.092	145.6	291.2
0.89	**1.05**	**1.21**	Mid sb4	B2	2.7	0.400	0.801	106.8	213.5
1.21	**1.43**	**1.65**	Mid sb5	B2	2.0	0.294	0.587	78.3	156.6
1.65	**1.95**	**2.25**	Mid sb6		1.6	0.215	0.431	57.4	114.9
2.25	**2.66**	**3.07**	Mid sb7		1.4	0.158	0.316	42.1	84.2
3.07	**3.63**	**4.18**	Mid sb8		1.6	0.116	0.232	30.9	61.8
4.18	**4.94**	**5.70**	Mid sb9	B5a	1.4	0.085	0.170	22.7	45.3
5.70	**6.74**	**7.78**	Mid sb10	B5a	1.3	0.062	0.125	16.6	33.2
7.78	**9.19**	**10.61**	Mid sb11	B5b	1.2	0.046	0.091	12.2	24.4
10.61	**12.53**	**14.46**	Mid sb12	B5b	1.2	0.034	0.067	8.9	17.9

Table 1. Sub-band definitions and correspondence with Bands for SKA1-Low and SKA1-Mid. Image sensitivity within the indicated frequency bands for continuum observations (σ_C for $\Delta\nu/\nu_c \approx 0.3$) for an observation of $\Delta\tau$ = 1 hour (as explained in RD4). The range of Gaussian FWHM beam sizes for which the approximate sensitivity value applies is given by θ_{min} to θ_{max}. The Gaussian FWHM beam sizes at which a doubling of the image noise from this base level is realised are given by θ'_{min} and θ'_{max}.

We will first consider what might be a suitable estimate for N_{Ateam} (within the Ingest pipeline). From Figure 15 in RD3, the value required for the far side-lobe response factor, η_F, for SKA1-

Low is $\underline{\eta}_F = (10^{-3}, 3\ 10^{-3}, 0.1)$ at (50, 100, 350) MHz, while the anticipated instrumental value is $\eta_F = 0.5$. The ratio $\eta_F / \underline{\eta}_F$ is the depth relative to the brightest source on the sky that needs to be included in the all-sky de-mixing. Based on the source population that occurs within 2π sr of the NVSS survey [RD5], this implies N_{Ateam} = (20, 12, 8) at (50, 100, 350) MHz. The simple form below captures this frequency dependence.

$$N_{Ateam} = 7 + (\nu_C/350\text{MHz})^{-1.4} \quad \text{(for SKA1-Low, } 50 < \nu_C < 350 \text{ MHz)} \tag{1}$$

Similarly, for SKA1-Mid, Figure 16 of RD3 indicates required values of $\underline{\eta}_F = (0.02, 0.2, 1)$ at (350, 1000, 2000) MHz, while the anticipated instrumental value is $\eta_F = 0.2$. This implies N_{Ateam} = (12, 2, 0) at (350, 1000, 2000) MHz. The simple form below captures this frequency dependence.

$$\begin{aligned} N_{Ateam} &= -1 + (\nu_C/2000\text{MHz})^{-1.4} && \text{(for SKA1-Mid, } 350 < \nu_C < 2000 \text{ MHz)} \\ &= 0 && (\nu_C > 2000 \text{ MHz}) \end{aligned} \tag{2}$$

Next we will consider the number of source components, N_{Source}, that need to be incorporated into the self-calibration model (within the ICAL pipeline) of the primary beam and its near-in side-lobes. As discussed in RD3 (Sections 6.1 and 6.2), an effective peak side-lobe response of about $\underline{\varepsilon}_S = 10^{-4}$ must be achieved for each of SKA1-Low and SKA1-Mid, while the actual values are likely to be $\varepsilon_S \approx 10^{-2}$. The implication is that the brightest $\varepsilon_S/\underline{\varepsilon}_S$ = 2 dex of source components within the near-in side-lobes must be included within the calibration model. Within the main-lobe of each, the required source depth is about $1/\underline{\varepsilon}_S$ = 4 dex.

We make use of the recent "T-RECS" simulation [Bonaldi et al, RD8] to estimate actual source numbers, $N(>S)(S_{Min},\nu)$, per unit area and median sizes, $\theta(S_{Min},\nu)$, as function of frequency. There are three primary populations of radio continuum sources within this model: namely (1) Star Forming Galaxies (SFG), (2) Steep spectrum Radio Galaxies (RG) and (3) Flat spectrum AGN core emission (termed QSO). The number counts of the three source classes as well as their sum are illustrated by the symbols in Figure 1. The solid lines that are overlaid on the symbols are the empirical model that we have developed to provide a continuous representation with appropriate frequency scaling of the simulation. The individual population number counts are modelled with the product of a power-law and one minus the exponential of a second power-law,

$$N(>S)(S_{Min},\nu) = N_0\, (S_{Min}/S_T(\nu))\hat{}(-\beta_1)\ \{1 - \exp[\, - (S_{Min}/S_T(\nu))\ \hat{}(-\beta_2)\ ]\}, \tag{3}$$

while the transition flux density, S_T, between the two power law regimes is scaled with the observing frequency ratio raised to the effective spectral index of each object class,

$$S_T(\nu) = S_{T0}(\nu/\nu_0)^{\alpha}. \tag{4}$$

This functional form for N(>S) has the virtue of having limiting values of,

$$\begin{aligned} N(>S)(S_{Min}.\nu) &= N_0\, (S_{Min}/S_T(\nu))\hat{}(-\beta_1) && \text{(for } S \ll S_T) \\ &= N_0\, (S_{Min}/S_T(\nu))\hat{}(-\beta_1-\beta_2) && \text{(for } S \gg S_T). \end{aligned} \tag{5}$$

For the RG and QSO object classes, a fixed spectral index α_{RG} = -0.98 and α_{QSO} = -0.28 was found sufficient to permit frequency scaling between 0.15 and 15 GHz, while for the SFG object class it was found necessary to allow for an effective spectral index as function of frequency based on a thermal regime α_T = -0.1, a non-thermal regime α_{NT} = -0.95 and thermal fraction f_{Th0} = 0.3 at a reference frequency ν_0 = 1.4 GHz. So,

$$\alpha_{SFG}(\nu) = \log_{10}[f_{Th0}(\nu/\nu_0)^{\alpha T}] + (1- f_{Th0})\,(\nu/\nu_0)^{\alpha NT}]/\log_{10}(\nu/\nu_0). \qquad (6)$$

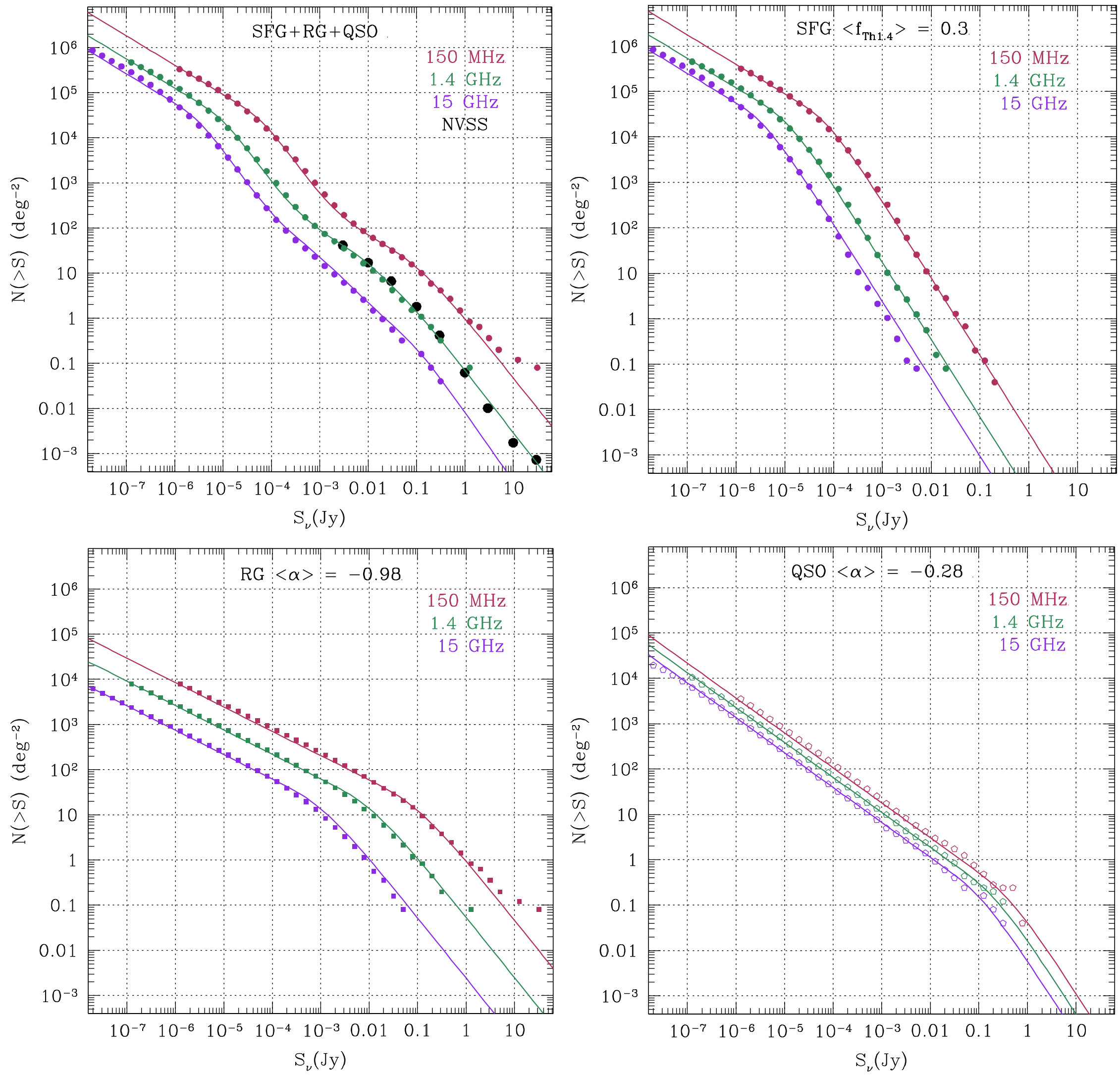


Figure 1. Number counts from the T-RECS simulations [RD8] at 0.15, 1.4 and 15 GHz in total (top left) and by source category; star-forming galaxies (top right), steep spectrum radio galaxies (bottom left) and flat spectrum Active Galactic Nuclei (bottom right). The empirical source count model for the source populations is overlaid as the solid lines in each panel. The NVSS number counts at 1.4 GHz [RD5] are also shown in the total distribution panel.

The brightest randomly occurring source of each object class within some total effective solid angle, Ω_{Tot}, is given by the approximate inversion of the N(>S)(S_{Min},ν) equation for that class,

$$S_{Max}(\Omega_{Tot},\nu) = S_T(\nu)(\Omega_{Tot}\, N_0\,)\text{^}[-1/\beta_1] \qquad (S_{Max} < S_T(\nu))$$
$$= S_T(\nu)(\Omega_{Tot}\, N_0\,)\text{^}[-1/(\beta_1+\beta_2)] \qquad (S_{Max} > S_T(\nu)) \qquad (7)$$

The total effective solid angle at zenith angle, ZA, is given approximately by,

$$\Omega_{Tot}(\nu) = \pi\,(1.22\,\lambda/D)^2\,/\cos(ZA)\ \ (sr) \qquad (SKA1\text{-}Low)$$
$$= \pi\,(1.22\,\lambda/D)^2\ \ (sr) \qquad (SKA1\text{-}Mid) \qquad (8)$$

The faintest source that must be included within the self-calibration model so as not to unduly degrade the thermal noise within very deep integrations (τ_H = 1000 hours) is implied by the effective *post-calibration* side-lobe parameter, $\underline{\varepsilon}_S$, of RD3. As shown in Figures 15 and 16 of RD3, this should be as low as about 10^{-4} near 120 MHz for SKA1-Low and also at 600 MHz for SKA1-Mid, and is relaxed toward both higher and lower frequencies. We model the required value as,

$$\underline{\varepsilon}_S(\nu,\tau_H) = 10\text{^}\{-4 - 0.5\,\log_{10}(\tau_H/1000) + 3.5[\log_{10}(\nu/0.12\text{ GHz})]^2\} \qquad (\text{for SKA1-Low})$$
$$= 10\text{^}\{-3.86 - 0.5\,\log_{10}(\tau_H/1000) + 0.5[\log_{10}(\nu/0.6\text{ GHz})]^2\}, \qquad (\text{for SKA1-Mid}) \qquad (9)$$

and include a $\tau_H^{-1/2}$ scaling for integrations less than 1000 hours, although in all cases we require $\underline{\varepsilon}_S(\nu) \leq 0.01$ as a minimum modelling depth.

The faintest source to consider during self-calibration of an observation is given by,

$$S_{Min}(\Omega_i,\nu,\tau_H) = \underline{\varepsilon}_S(\nu,\tau_H)\ S_{Max}(\Omega_{Tot},\nu)/A_i, \qquad (10)$$

where A_i is the station/dish beam attenuation that applies to the solid angle Ω_i.

The three different object classes are associated with different angular size distributions as shown in Figure 2. For the star forming galaxies, the size tabulated in RD8 is the scale length of an exponential disk, while for the active galactic nuclei it is the total projected size of the lobes or jets. Since we will assume that the flat spectrum AGN are core dominated, we have arbitrarily reduced their tabulated angular size by a factor of ten relative to RD8 to produce the anticipated core angular sizes of a few milliarcsec. Since the steep spectrum radio galaxies of RD8 saturate at 10 arcsec size, we have instead followed the NVSS size distribution at the highest flux densities for the RG population.

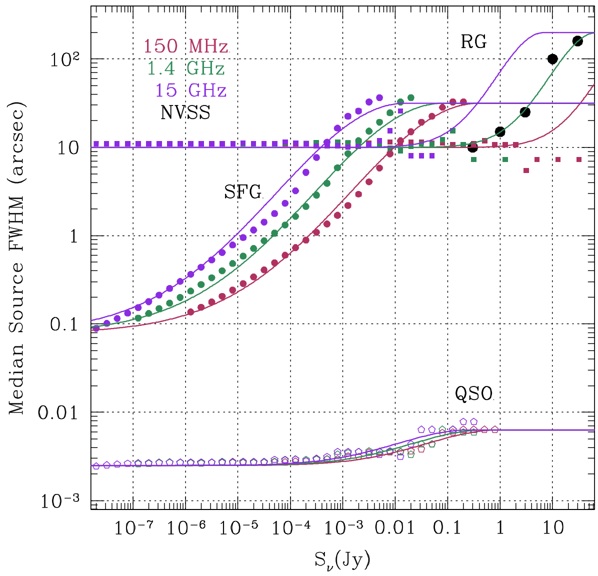


Figure 2. Median source size from the T-RECS simulations [RD8] at 0.15, 1.4 and 15 GHz by source category; star-forming galaxies (filled circles), steep spectrum radio galaxies (filled squares) and flat spectrum Active Galactic Nuclei (open pentagons). The empirical model for the source sizes is overlaid as the solid lines. The NVSS source sizes at 1.4 GHz [RD5] are also shown as the large filled circles at high flux densities.

The source size of each object class is modelled as,

$$\theta(S_{Min},\nu) = 10\hat{}(\log_{10}[\theta_0] + a_G \exp\{-[(\log_{10}(S_{Min}) - \log_{10}[S_G(\nu)])/\sigma_G]^2\}) \quad (\text{for } S_{Min} < S_G(\nu))$$
$$= 10\hat{}(\log_{10}[\theta_0] + a_G) \quad (\text{for } S_{Min} > S_G(\nu)), \qquad (11)$$

where,

$$S_G(\nu) = S_{G0}\,(\nu/\nu_0)^{\alpha}. \qquad (12)$$

In this case, a fixed effective spectral index of α_{SFG} = -0.7, α_{RG} = -0.98 and α_{QSO} = -0.28 was found sufficient for all three object classes.

Pop	$\log_{10}(N_0)$ (deg^{-2})	$\log_{10}(S_{T0})$ (Jy)	ν_0 (GHz)	α	β_1	β_2	θ_0 (“)	a_G	S_{G0} (Jy)	σ_G
SFG	4.30	-4.80	1.40	-0.70*	-0.65	-1.05	0.08	2.60	0.083	3.50
RG	1.15	-1.80	1.40	-0.98	-0.54	-0.80	10.	1.30	71	1.30
QSO	-0.88	-0.50	1.40	-0.28	-0.77	-0.90	0.003	0.40	0.54	2.00

Table 2. Parameters of the empirical model outlined above that is used to reproduce the T-RECs source densities and median angular sizes for the three object classes as function of observing frequency and limiting flux density. *Note that for the SFG object class, α is frequency dependant as described in the text.

We collect all of the parameters of our empirical model for both the source densities and sizes in Table 2.

The interferometer baseline that resolves a source of a particular angular size is given by,

$$B_R(\theta,\nu) = 5\ (\theta/10\ \mathrm{arcsec})^{-1}\ (\nu/1.4\ \mathrm{GHz})^{-1}\ \text{(in km)}. \tag{13}$$

We model the number of components, N_C, per RG source as,

$$\begin{aligned} N_{CRG}(\theta,\nu,B_{Max}) &= 1 && (\text{for } B_{Max} < B_R) \\ &= (B_{Max} / B_R)^{3/2} && (\text{for } B_R < B_{Max} < 2.92\ B_R) \\ &= 5 && (\text{for } B_{Max} > 2.92\ B_R), \end{aligned} \tag{14}$$

where we allow no more than 5 components per resolved steep spectrum source, to reflect the decreasing amplitude of any secondary source components.

Since SFG and QSO sources are centrally dominated we instead saturate the number of components per source within these object types at only 1.5, so that,

$$\begin{aligned} N_{CSFG/QSO}(\theta,\nu,B_{Max}) &= 1 && (\text{for } B_{Max} < B_R) \\ &= (B_{Max} / B_R)^{3/2} && (\text{for } B_R < B_{Max} < 1.31\ B_R) \\ &= 1.5 && (\text{for } B_{Max} > 1.31\ B_R). \end{aligned} \tag{15}$$

To obtain a count of source components within the calibration model it is necessary to sum,

$$N_{Obj}(\nu,\tau_H,B_{Max}) = \textstyle\sum_i \Omega_i(A_i)\ N_C(\theta,\nu,B_{Max})\ N(>S)(S_{Min},\nu,\tau_H)\ A_i \tag{16}$$

for each object class, where the sum is over the beam attenuation factor, A_i. The method of calculation makes use of the solid angle at a given beam attenuation, $\Omega_i(A_i)$, that we derive below. That beam attenuation is used to calculate the relevant S_{Min} from eqn. 10. This in turn allows calculation of both N(>S) and N_C from eqns. 3, 11, 13 – 15 to give N_{Obj} from eqn. 16. The total number of source components is given by the sum over the three object classes,

$$N_{Source}(\nu,\tau_H,B_{Max}) = \textstyle\sum_{Obj} N_{Obj}(\nu,\tau_H,B_{Max}). \tag{17}$$

The solid angle as function of beam attenuation is determined from a model of the station and dish beams. For the SKA1-Low stations we first consider a semi-random distribution of 256 antennas with a minimum centre to centre separation of 1.6m and maximum separation of D=38m. The fractional RMS amplitude error and phase error during station beam formation are assumed to be 2% and 2°, and a 2% random antenna failure rate is adopted. The station is assumed to be pointing at a representative zenith angle = 37°, and the product of two station voltage beams which are rotated relative to each other by 11° in azimuth is shown in Figure 3.

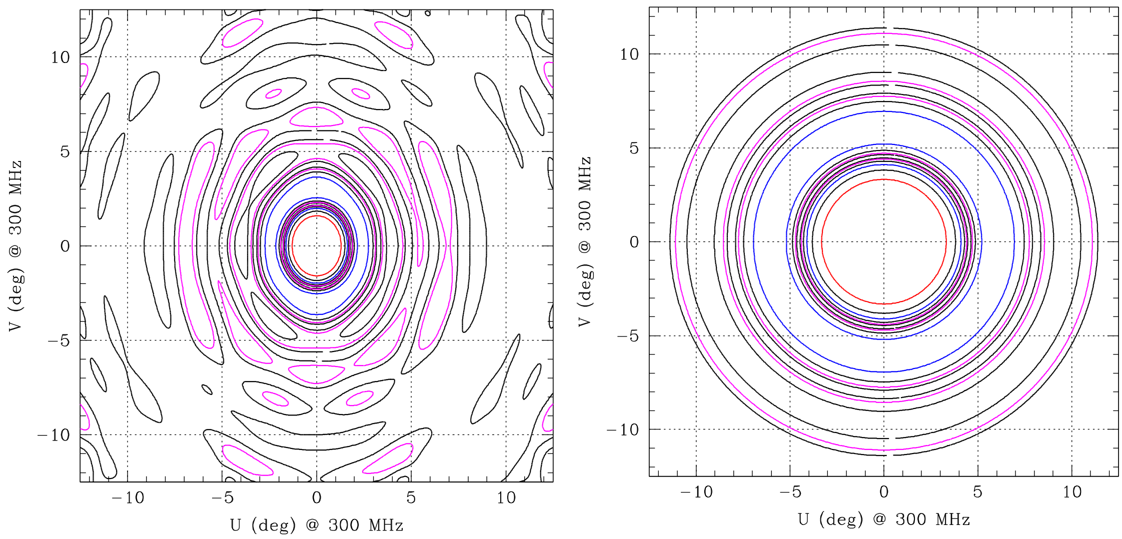


Figure 3. Beam model for SKA1-Low (left) and SKA1-Mid (right). The coloured contours are drawn at -10, -20 and -30 dB (red, blue, magenta), while intermediate black contours are drawn at -15, -25 and -35 dB.

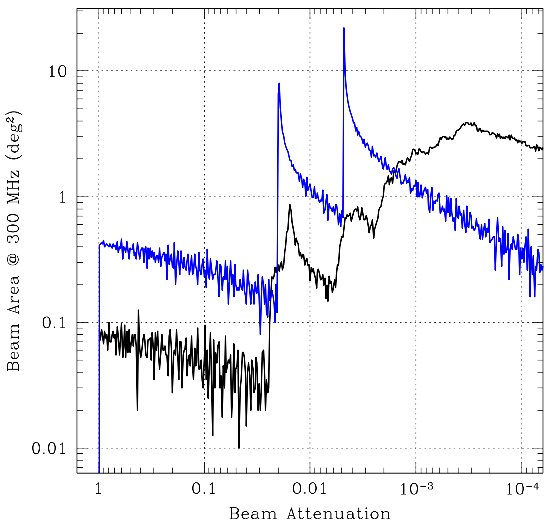


Figure 4. Beam area at a fixed reference frequency as function of attenuation for SKA1-Low (black) and SKA1-Mid (blue) prior to summation with the model for the far side-lobe response.

For the SKA1-Mid dishes we take as starting point an unobstructed, uniformly illuminated Airy pattern, $[J_1(r)/r]^2$, with r = D/λ and D = 15m. This is understood to be only an approximation, since the aperture illumination is actually tapered toward the edges and is non-axisymmetric. The basic beam patterns are shown as sky images at an arbitrary reference frequency in Figure 3 and accumulated into histograms of beam area as function of beam attenuation in Figure 4. The beam patterns can be linearly scaled to other frequencies, while the beam area will vary with the square of the relative frequency. To both of the patterns shown in Figures 3 and 4 we add an assumed frequency dependent far side-lobe contribution,

$$A_F = \eta_F (\lambda/D)^2, \quad (18)$$

where η_F = 0.5 for SKA1-Low and 0.2 for SKA1-Mid (see RD2) and renormalize the attenuations as,

$$A'_i = (A_i + A_F)/(1 + A_F). \tag{19}$$

We have evaluated $N_{Source}(\nu,\tau_H,B_{Max})$ for each of the frequency sub-bands defined in Table 1 in conjunction with four different total integration times that attempt to span the full range of possible experiment types, τ_H = 1000, 100, 3 and 0.1 hours as well as several different possible values of the desired angular resolution during calibration, corresponding to B_{Max} = 65 and 40km for SKA1-Low and B_{Max} = 150, 120, 100 and 50km for SKA1-Mid. We list total source component numbers, together with the relevant S_{Max} and S_{Min} (at the beam centre) for the case of the deepest integrations (τ_H = 1000 hr) at the largest maximum baseline in Table 3.

It is also necessary to estimate the number of self-calibration iterations that are required to achieve the desired level of calibration precision. Each Self-Cal iteration involves (a) the gridding, imaging and deconvolution (if needed) of the current data, (b) the refinement of the current Local Sky Model (LSM) of the field, (c) predicting the visibility data that is associated with the current LSM and (d) solving for updated calibration solutions to be applied for the next round of imaging.

From practical experience with the VLA imaging at 1.4 GHz, it is estimated that the minimum number of self-calibration iterations required for even the most rudimentary of sky models (with, say, N_{Source} = 100) is $N_{SelfCal}$ = 3. For more complex sky models, the required iteration number will increase. In the case of traditional self-calibration algorithms such as implemented in Classic *AIPS* and CASA, the required Self-Cal iteration number can increase quite dramatically with LSM complexity. The LOFAR EoR team [e.g. RD9] have developed an efficient self-calibration approach, based on the SageCal package of Yatawatta [RD10], that allows even complex LSMs, with $N_{Source} \approx 20000$, to be developed and calibrated with only $N_{SelfCal}$ = 5 major iterations per tracking observation. For the LOFAR EoR case, three independent data tracks were used to develop the North Celestial Pole LSM. With the greatly improved instantaneous visibility sampling and sensitivity of the SKA, there is the expectation that a single tracking observation will suffice for LSM generation. This experience suggests the functional form,

$$\begin{aligned} N_{SelfCal} &= 3 + \log_{10}(N_{Source}/100) && (N_{Source} > 100) \\ &= 3 && (N_{Source} < 100) \end{aligned} \tag{20}$$

With N_{Source} as high as 40,000, this would correspond to some 6 major self-cal iterations.

The number of so-called "major" cycles, N_{Major}, in which intermediate Local Sky Models are Fourier inverted and subtracted from the visibilities during deconvolution will also vary with the current complexity of the sky model. An indicative estimate of N_{Major} might be the current self-cal iteration number, so that over $N_{SelfCal}$ iterations there would be an average of N_{Major} = $N_{SelfCal}/2$.

Although only limited support for Direction Dependent (DD) rather than Direction Independent (DI) calibration methods is currently implemented within the SDP parametric model [RD6], there is the ability to specify the number, $N_{Ipatches}$, of Ionospheric patches for which gain solutions will be determined as well as the associated timescale, t_{ICALI}. We will make use of this capacity in a more general sense to represent calculation of a variety of DD complex gain solutions, including but not restricted to the ionosphere itself. The most important additional category is characterisation of the time dependent off-axis gains due to non-axi-symmetry of the primary beam in each of its polarisation states and the near-in side-lobes. At the lowest frequencies, the number of directions for which solutions need be determined will be defined by the ionospheric isoplanatic patch size of about 1 deg^2 relative to the main-beam plus side-lobe field of view, while at higher frequencies this will decline to some minimum number that is needed to most efficiently characterise the beam shape properties. A simple functional form that captures the low frequency dependence (the summed FoV in the main-lobe and side-lobes in units of deg^2) is,

$$N_{Ipatches} = 1380\, d_m^{-2}\, \nu_{GHz}^{-2}. \tag{21}$$

The minimum value is assumed to be the smaller of 20, which might be sufficient to capture the beam shape properties, or $N_{Source}/3$, in the event that only a small number of sources in the Local Sky Model are being used to constrain the solution,

$$N_{Ipatches} = \min(20, N_{Source}/3). \tag{22}$$

We will assume t_{ICALI} = 10 sec throughout (but see below).

Finally, it is necessary to specify the major cycle number that would be associated with the final data product preparation (the DPrep pipelines). Assuming that the self-calibration has been undertaken to a depth (as outlined above) that is appropriate to the final depth of an observation, then the continuum model of the observed field that has been developed in the ICAL step is already appropriate for direct subtraction and optional restoration (with a matched Gaussian beam) to any desired data product. Beyond this, there should be very few circumstances under which further deconvolution of data products is necessary. As demonstrated in Section 7.1 and 7.2 of RD4, the dirty point spread function obtained with a "uniform" data weighting followed by Gaussian tapering during gridding, already provides an appropriately high image dynamic range, varying between about 15 and 50 dB for different combinations of observing track length and multi-frequency sampling, for the anticipated residual brightness distributions. We therefore assume N_{Maj} = 0 during the DPrep phase.

For reference, we collect the calibration parameters (N_{Ateam}, N_{Source}, $N_{SelfCal}$, N_{Major}, $N_{Ipatches}$) that follow from the simple model outlined in this section for the case of the deepest integrations (τ_H = 1000) and the largest available B_{Max} (= 65km for SKA1-Low and 150km for SKA1-Mid) in Table 3. As noted above, the value of N_{Major} is the average value that applies to each of the $N_{SelfCal}$ iterations, while for the case of data product preparation it is assumed that N_{Major} = 0. The calibration parameters are relaxed for more shallow integrations and reduced B_{Max}. Also shown in the Table are the model predictions for the brightest randomly occurring source within the field of view $S_{Max}(\Omega_{Tot},\nu)$ as well as the minimum apparent (as tapered by the primary beam) flux density of source, S_{Min}, that needs to be included in the LSM. The column headed "$S_{Min}/\sigma_{4/6h}$" in Table 3 gives the ratio of the faintest apparent flux density of source that is needed within the Local Sky Model to the RMS continuum noise level in a 4 (for Low) or 6 (for Mid) hour integration. This demonstrates that there should be sufficient signal-to-noise within a single-track observation within all of the frequency sub-bands of both SKA1-Low and SKA1-Mid to allow generation of even the most demanding LSM (suitable for the calibration of a net 1000^h integration).

ν_{min} (GHz)	ν_c (GHz)	ν_{max} (GHz)	Sub-band	Band	N_{ATm}	N_{Source}	S_{Max} (Jy)	S_{Min} (Jy)	$S_{Min}/\sigma_{4/6h}$	$N_{SelfCal}/N'_{SelfCal}$	N_{Maj}/N'_{Maj}	N_{Ipch}
0.050	**0.060**	**0.069**	Low sb1		19	36820	68	14m	172	6/1	3/1	336
0.069	**0.082**	**0.096**	Low sb2		15	35270	32	3.9m	166	6/1	3/1	180
0.096	**0.114**	**0.132**	Low sb3		12	28390	14	1.4m	107	5/1	3/1	93
0.132	**0.158**	**0.183**	Low sb4		10	24760	6.3	0.7m	78	5/1	3/1	48
0.183	**0.218**	**0.253**	Low sb5		9	17050	2.8	0.5m	71	5/1	3/1	25
0.253	**0.302**	**0.350**	Low sb6		8	9602	1.3	0.5m	91	5/1	2/1	20
0.35	**0.41**	**0.48**	Mid sb1	B1	8	29860	2.0	0.3m	44	6/1	3/1	36
0.48	**0.56**	**0.65**	Mid sb2	B1	5	25140	0.9	0.1m	30	6/1	3/1	20
0.65	**0.77**	**0.89**	Mid sb3	B1	3	21530	0.4	60μ	34	5/1	3/1	20
0.89	**1.05**	**1.21**	Mid sb4	B2	2	18770	0.2	20μ	18	5/1	3/1	20
1.21	**1.43**	**1.65**	Mid sb5	B2	1	16290	90m	15μ	18	5/1	3/1	20
1.65	**1.95**	**2.25**	Mid sb6		0	11430	50m	9μ	12	5/1	3/1	20
2.25	**2.66**	**3.07**	Mid sb7		0	6660	31m	7μ	12	5/1	3/1	20
3.07	**3.63**	**4.18**	Mid sb8		0	3770	20m	6μ	10	5/1	3/1	20
4.18	**4.94**	**5.70**	Mid sb9	B5a	0	2087	13m	5μ	9	5/1	2/1	20
5.70	**6.74**	**7.78**	Mid sb10	B5a	0	1117	8m	4μ	8	4/1	2/1	20
7.78	**9.19**	**10.61**	Mid sb11	B5b	0	582	5m	4μ	8	4/1	2/1	20
10.61	**12.53**	**14.46**	Mid sb12	B5b	0	293	3m	3μ	6	4/1	2/1	20

Table 3. Sub-bands and calibration parameters for SKA1-Low and SKA1-Mid. The parametric model parameters that correspond to the case of the deepest integrations using the maximum available baselines.

It is very important to note that the parameters specified above for self-calibration apply only to the first data track that is used to develop a Local Sky Model for a field. The self-calibration of all subsequent data tracks for that field will only require a single iteration, $N_{SelfCal} = 1$ and $N_{Major} = 1$, although using the same value of N_{Source} throughout. The implication is that observations that are comprised of N_{Track} coverages will have an effective $N'_{SelfCal}$ and N'_{Major} given by,

$$N'_{SelfCal} = (N_{SelfCal} + (N_{Track} - 1))/ N_{Track} \tag{23}$$

$$N'_{Major} = (N_{Major} + (N_{Track} - 1))/ N_{Track}. \tag{24}$$

As shown in Table 3, assuming a typical track length of 4^h for SKA1-Low and 6^h for SKA1-Mid, the effective iteration numbers for deep multi-track observations are essentially equal to unity. This greatly relaxes the total computational requirements for such experiments, although it does introduce an additional scheduling constraint for resourcing the development of the LSM for each field, which would typically have a 15 times greater computational load than the remaining tracks.

It is also important to specify the time and frequency sampling that is necessary for the calculation of predicted visibilities during the self-calibration process. Appropriate time (τ^*) and frequency $(\Delta\nu/\nu)^*$ sampling to avoid significant smearing effects at the edge of the field-of-view is tied to both the station/dish diameter and the maximum baseline that is specified for an observation as [RD2],

$$\tau^* = 1.38\ 10^4\ \eta_S\ D/B_{Max}\ (s) \quad (25)$$

$$(\Delta\nu/\nu)^* = \eta_S\ D/B_{Max} \quad (26)$$

in terms of an acceptable smearing factor, η_S, relative to the synthesized beam diameter. For $\eta_S = 0.1$, this yields $\tau^* = 0.8$s and $(\Delta\nu/\nu)^* = 6\ 10^{-5}$ for SKA1-Low (B_{Max} = 65km) and $\tau^* = 0.14$s and $(\Delta\nu/\nu)^* = 1\ 10^{-5}$ for SKA1-Mid (B_{Max} = 150km). However, achieving sufficient signal-to-noise per visibility to permit successful self-calibration to be undertaken will require data averaging to about $\tau_{Sol} = 10\tau^*$ and $(\Delta\nu/\nu)_{Sol} = 10(\Delta\nu/\nu)^*$ for both SKA1-Low (over its entire frequency range) and SKA1-Mid (in the range 0.35 to 1.5 GHz), as demonstrated in RD3. Further averaging in both time and frequency will be necessary for SKA1-Mid self-calibration above 1.5 GHz, increasing to about $\tau_{Sol} = 1000\tau^*$ and $(\Delta\nu/\nu)_{Sol} = 1000(\Delta\nu/\nu)^*$ at 15 GHz, which corresponds to about τ_{Sol} = 140s and $(\Delta\nu/\nu)_{Sol} = 0.01$. We model the frequency dependence of the solution intervals as,

$$\begin{aligned} \tau_{Sol}(B_{Max0}) &= 10\ \tau^* && (\nu_C < 1.5\ \text{GHz}) \\ &= 10\text{\^{}}[1 + 2\log_{10}(\nu_C/1.5)]\ \tau^* && (\nu_C > 1.5\ \text{GHz}) \end{aligned} \quad (27)$$

$$\begin{aligned} (\Delta\nu/\nu)_{Sol}(B_{Max0}) &= 10\text{\^{}}[1 - 0.5\log_{10}(\tau_H/1000)]\ (\Delta\nu/\nu)^* && (\nu_C < 1.5\ \text{GHz}) \\ &= 10\text{\^{}}[1 - 0.5\log_{10}(\tau_H/1000) + 2\log_{10}(\nu_C/1.5)]\ (\Delta\nu/\nu)^* && (\nu_C > 1.5\ \text{GHz}) \\ &\geq 0.01 \end{aligned} \quad (28)$$

In the case of the frequency solution interval we include a scaling as the square root of the total observation depth, while requiring a minimum solution interval in all cases of at least $(\Delta\nu/\nu)_{Sol} = 0.01$. For the LOFAR EoR processing it has been found necessary to predict visibilities during LSM generation at a time and frequency resolution of about τ_{Sol} = 10s and $(\Delta\nu/\nu)_{Sol} = 5\ 10^{-4}$, which is consistent with our estimate above. It should be noted that the signal-to-noise calculation of the solution intervals for SKA1-Low and SKA1-Mid is tied to the intrinsic, B_{Max0}, of the configurations and this remains true even if lower resolution data products (associated with a smaller B_{Max}) are being generated.

3.3 HPC Parameter Space Exploration

Having established relevant values for some of the key model parameters it is useful to explore the computing requirements associated with different use cases. We consider four different final depths of observation, τ_H = 0.1, 3, 100 and 1000 hours, several alternate values of B_{Max} as well as a generic continuum observation (with N_{PP} = 4 and N_{fout} = 30 spectral channels across a sub-band) and a generic spectral observation (with N_{PP} = 2 and N_{fout} = 3000 spectral channels across a sub-band). During the ICAL step, N_{fout} is tied to the solution interval, $(\Delta\nu/\nu)_{Sol}$, defined above for both the continuum and spectral line cases, since this will be assumed to be an appropriate spectral resolution to track intrinsic and apparent spectral structure in the continuum model of each field. For final depths of 100 and 1000 hours, the computational load is based on the average over all tracks. It should be noted that the first track, when LSM generation is undertaken, will have a computational load that is typically 15 times larger.

The computational requirements for the spectral use cases are shown in Figure 5, while those for the continuum use cases are in Figure 6. In all cases, an overall computational efficiency of 10% has been assumed (but see Section 4.2) in deriving the HPC needs.

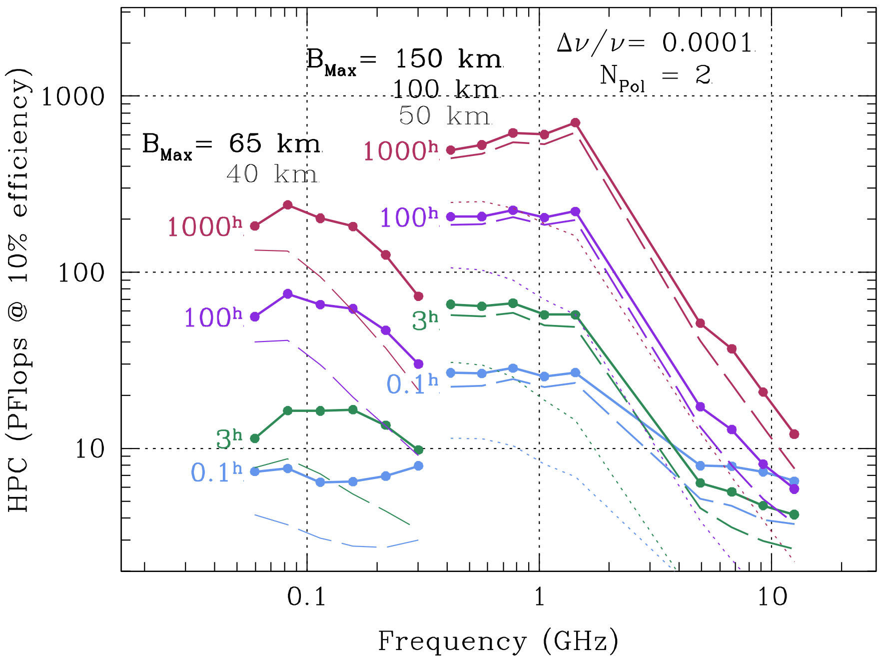


Figure 5. Computational needs for a range of SKA1 generic spectral line experiment use cases. The filled circles and solid lines pertain to the largest values of B_{Max}. The four different coloured versions pertain to the four different values of the final depth per pointing direction of an experiment as labelled. The dashed and dotted curves of the same colour illustrate the consequence of reductions to B_{Max} to the values indicated.

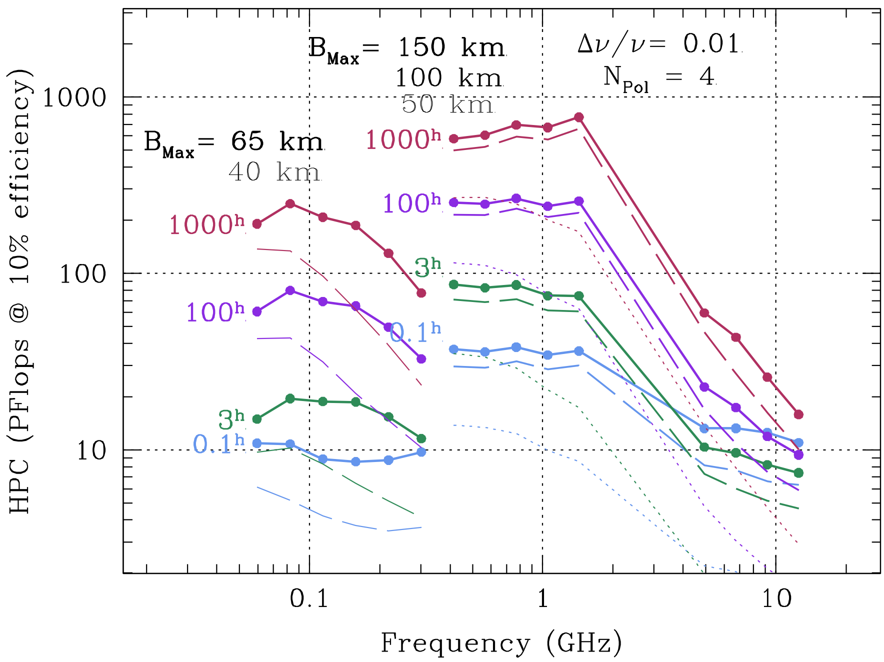


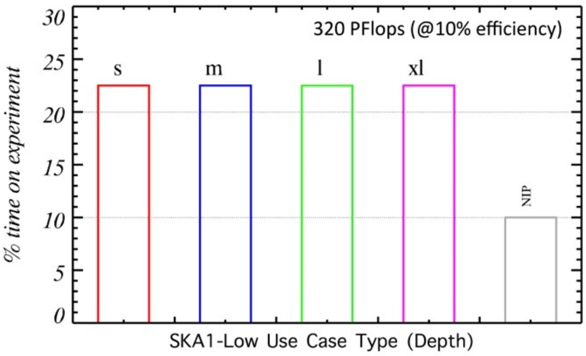
Figure 6. Computational needs for a range of SKA1 generic continuum experiment use cases. The filled circles and solid lines pertain to the largest values of B_{Max}. The four different coloured versions pertain to the four different values of the final depth per pointing direction of an experiment as labelled. The dashed and dotted curves of the same colour illustrate the consequence of reductions to B_{Max} to the values indicated.

What is apparent from the plots is the strong dependence of computational cost on the observing frequency, the final depth of integration as well as the required angular resolution. It should be noted that the number of dishes contributing to SKA1-Mid at frequencies greater than 4 GHz is assumed to decline from 197 to 133, since the 64 MeerKAT dishes are not assumed to be equipped with such receivers. What is not shown in the plots is the breakdown in cost among the three processing steps (Ingest, ICAL, DPrep) that have been summed to give the total. It is found that the ICAL step completely dominates the HPC cost, representing between 94 and 98% of the total in all cases considered. Once the calibration of an observation has been completed, a wide range of data products could be generated at minimal incremental cost, almost independent of the spectral and angular resolution that is required. Another important point to note is that the differences in HPC costs of the continuum and spectral line use cases are essentially due to the difference in the specified polarisation product numbers, N_{PP} = 4 versus 2 during the ICAL processing step. It is likely that a reasonable polarisation calibration will be essential under most circumstances, so from this point onward we will consider only the N_{PP} = 4 use cases to provide a conservative estimate.

Use cases that do not require visibility-based data products, which we will designate with the generic abbreviation, NIP (non-image processing), would instead make use of the processing steps: Ingest, followed by RCAL, the real-time calibration pipeline. The time-averaged HPC requirements in this mode correspond to less than about 1 PFlops for both SKA1-Low or SKA1-Mid, so we will adopt a typical HPC load of 1 PFlop for NIP use cases.

3.4 "Unconstrained" HPC Requirements

Having quantified the HPC needs of a representative range of use cases, it is possible to calculate the total requirements that would be needed to accommodate some distribution of those experiments. For the purpose of this calculation we will assume that 10% of observing time is used for NIP applications, although this is simply to enable calculation of a total HPC load and should not be interpreted as an indication of the likely time allocations to any project types. Further, we will assume that the maximum possible value of B_{Max} and all simultaneously accessible sub-bands will be processed. In practise this implies B_{Max} = 65 km and all six sub-bands for SKA1-Low, while in the case of SKA1-Mid this implies B_{Max} = 150 km and the full bandwidth of either Band 1, 2, 5a or 5b. We then assume an equal fraction of observing in each of the four depth categories τ_H = 0.1, 3, 100 and 1000 hours (given the short-hand designations of s, m, l, and xl) and in the case of SKA1-Mid, in each of the available frequency Bands. The total requirements that emerge for this uniform distribution of use cases is between 320 and 307 PFlops for each of SKA1-Low and SKA1-Mid and is illustrated graphically in Figure 7. This represents about 2.5 times the HPC capacity that was estimated for the Baseline Design SDP HPC deployment of about 260 PFlops (however, see Section 3.7 below).



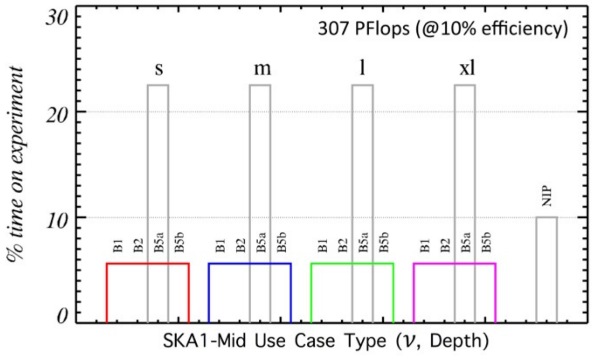


Figure 7. Fractional scheduling of different use cases with no further constraint on the total HPC capacity for SKA1-Low (left) and SKA1-Mid (right). The corresponding HPC capacity is about 320 PFlops for SKA1-Low and 307 PFlops for SKA1-Mid. The four different observing depths are designated (s, m, l, xl) in the plots. The individual feed systems of SKA1-Mid (B1, B2, B5a, B5b) are indicated.

3.5 Constrained HPC Capabilities (about 400 PFlops)

It is also possible to consider scenarios that are subject to some constraint on the available HPC resources. We first consider the case where a total of about 400 PFlops is available for the combination of SKA1-Low and SKA1-Mid.

For the case of SKA1-Low, we keep B_{Max} fixed at its maximum value of 65km to allow the highest angular resolution to be retained, and continue to assume an NIP observation fraction of 10%. Based on the outcome of the unconstrained case discussed above, it is clear that it will not be possible to achieve a uniform distribution of the four observing depth categories while processing all six of the SKA1-Low sub-bands simultaneously. As an illustrative example, we arbitrarily divide the total accessible band-width into two components: Low sub-bands 1

– 3 and sub-bands 4 – 6. With this adjustment, it is possible to achieve a uniform distribution over all four observing depth categories and the two sub-bands groupings, as illustrated in Figure 8.

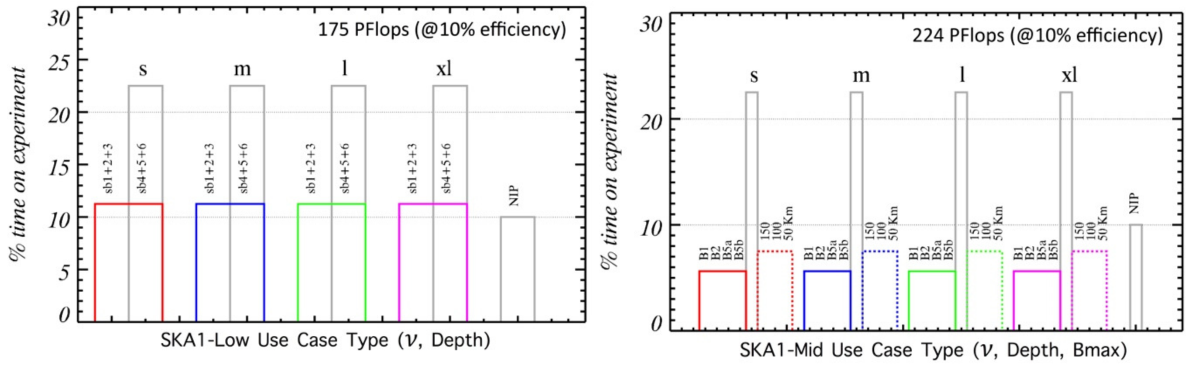


Figure 8. Fractional scheduling of different use cases that respect an HPC constraint of about 400 PFlops for the combination of SKA1-Low (left) and SKA1-Mid (right). The corresponding HPC capacity is about 175 PFlops for SKA1-Low and 224 PFlops for SKA1-Mid. The four different observing depths are designated (s, m, l, xl) in the plots, and the grey histograms with this labelling represent the sum of experiments at that depth. The SKA1-Low bandwidth is divided into two frequency ranges. The individual feed systems of SKA1-Mid (B1, B2, B5a, B5b) are indicated, as well as three different cases of B_{Max}.

For the case of SKA1-Mid, we also assume an NIP fraction of 10%, keep the full bandwidth of either Band 1, 2, 5a or 5b, but allow B_{Max} to vary between 50, 100 and 150km. A uniform distribution of observing fractions over frequency Bands and B_{Max} categories is possible in this case as demonstrated in Figure 8.

3.6 Highly Constrained HPC Capabilities (about 85 PFlops)

In a more constrained variant of the case just considered, we assume a total HPC capacity of about 80 PFlops that is distributed in some way between SKA1-Low (35) and SKA1-Mid (46). With this lower capacity, we make use of a reduced B_{Max} of 40 km for SKA1-Low and 120km for SKA1-Mid. For SKA1-Mid, we also include the cases of B_{Max} = 100 and 50km and consider each of the frequency sub-bands in isolation, rather than being tied to what is simultaneously accessible. And finally, we have deliberately excluded the extremely deep integrations (τ_H = 1000 hours, designated xl) from consideration.

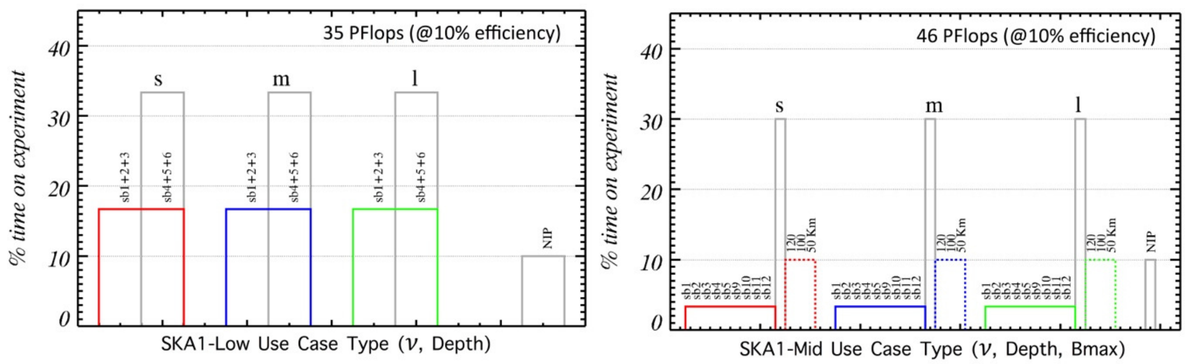


Figure 9. Fractional scheduling of different use cases that respect an average HPC constraint of about 80 PFlops for the combination of SKA1-Low (left) and SKA1-Mid (right). The three different observing depths are designated (s, m, l) in the plots, and the grey histograms with this labelling represent the sum of experiments at that depth. The SKA1-Low bandwidth is divided into two groups of sub-bands. The SKA1-Mid frequency coverage is divided into the nine sub-bands (each of 32% width) defined in Table 1.

With the adjustments to the use case mix outlined above, we achieve the uniform fractional time allocations shown in Figure 9 with a deployment of 35 PFlops for SKA1-Low and 46 PFlops for SKA1-Mid.

3.7 Relation to the previously defined High Priority Science Objectives

When developing a time-averaged sizing of HPC needs within RD6, a suite of experiments termed the "High Priority Science Objectives", as listed in Table 2 of the Level 0 Science Requirements [RD7], was used. We repeat the most relevant columns of that Table 4 below, together with additional columns which place each of these notional experiments into the context developed here. The relevant values of B_{Max}, the frequency sub-bands and the depth category are listed, followed by the corresponding instantaneous compute capacity. The "Fraction" column shows what fraction of an observing schedule might be filled with the relevant experiment assuming that this is in proportion to the total integration time requested divided by the sum of time requested for all experiments on that telescope within this list, while the final column gives the product of observing fraction with instantaneous load. The weighted sum of compute load calculated in this way is 113 PFlops for SKA1-Mid and 165/84 PFlops for SKA1-Low depending on whether B_{Max} is fixed at 65 or 40km. This yields a total HPC requirement of about 280 PFlops, which is essentially what is specified in the SKA1 Design Baseline.

Some further explanation of the entries within Table 4 is warranted. Inspection of the list of experiments demonstrates that while the range of experiments defined for SKA1-Mid is quite broad and is varied in terms of computational cost, the experiment list for SKA1-Low contains only the extremely HPC challenging Epoch of Reionisation entries on the one hand and the NIP experiments on the other. Actual observing schedules for SKA1-Low are likely to be much more varied.

Science Objective	SWG	High Priority Science Objective Number	Mode	Frequency	Observing Area		Integration		HPC					
				Range Low - High	Total Area	Angular Resolution Min:Max	Total (hr)	Per Pointing	B_Max (km)	sub-bands	Depth	PFlops	Fraction'	PF*Frac'
EoR - Imaging AASKA14:001	CD/EoR	1	Imaging	50 - 200 MHz	100 deg2	10:1000 arcsec	5000	2000 hr	65/40	Low 1-4	xl	648/ 336	0.156	101/52
EoR - Power Spectra AASKA14:001	CD/EoR	2	imaging/Power Spectrum	50 - 200 MHz	1000 deg2	10:1000 arcsec	5000	200 hr	65/40	Low 1-4	l	231/113	0.156	36/18
			imaging/Power Spectrum	50 - 200 MHz	10000 deg2	10:1000 arcsec	5000	20 hr	65/40	Low 1-4	m/l	175/88	0.156	28/14
Pulsar Searching AASKA14:040	Pulsars	4	Non-Imaging	150 - 350 MHz	30000 deg2	320 arcsec	12750	40 mn	1	Low 4-6	s/m	1	0.398	0.40
			Non-Imaging	650 - 950 MHz	2400 deg2	105 arcsec	800	10 mn	1	Mid 3	s	1	0.009	0.01
			Non-Imaging	1250 - 1550 MHz	2400 deg2	60 arcsec	2400	10 mn	1	Mid 5	s	1	0.028	0.03
Pulsar Timing AASKA14:037	Pulsars	5	Non-Imaging	150 - 350 MHz	0.9 arcmin2	8 arcsec	4300	40 mn	1	Low 4-6	s/m	1	0.134	0.13
			Non-Imaging	950 - 1760 MHz	0.7 arcmin2	7 arcsec	1600	15 mn	1	Mid 4-5	s	1	0.019	0.02
HI - High z AASKA14:128	HI	13	Imaging	790 - 950 MHz	5.4 deg2	3:5 arcsec	5000	1000 hr	50	Mid 3	xl	247	0.058	14.29
HI - Low z AASKA14:129	HI	14	Imaging	1300 - 1400 MHz	3.8 deg2	3:5 arcsec	2000	200 hr	50	Mid 5	l	63	0.023	1.46
HI - Galaxy AASKA14:130	HI	15	Imaging	1415 - 1425 MHz	1080 deg2	5:60 arcsec	12600	4.4 hr	50	Mid 5	m	17	0.146	2.48
Transients - FRB AASKA14:055	Transients	18	Non-imaging/ Commensal	650 - 950 MHz	30000 deg2	105 arcsec	10000	2 msec	1/150	Mid 3	s	36	0.116	4.17
CoL - Planet formation AASKA14:117	Cradle of Life	22	Imaging	8 - 12 GHz	0.05 deg2	0.04:1 arcsec	6000	600 hr	150	Mid 11-12	xl	42	0.069	2.92
Magnetism - RM-grid AASKA14:092	Magnetism	27	Imaging	1000 - 1700 MHz	31000 deg2	2 arcsec	10000	7.4 mn	100	Mid 4 -5	s	59	0.116	6.83
Cosmology - High z IM AASKA14:019	Cosmology	32	Auto-Correl/ Commensal	350 - 1050 MHz	30000 deg2	1.7 deg	10000	2.2 hr @ 190 Dishes	1/50	Mid 1-3	m	98	0.116	11.34
Cosmology - ISW, Dipole AASKA14:018, 032	Cosmology	33	Imaging	1000 - 1700 MHz	31000 deg2	2 arcsec	10000	7.4 mn	100	Mid 4-5	s	59	0.116	6.83
Continuum - SFR(z) AASKA14:067	Continuum	37 + 38	Imaging	1000 - 1700 MHz	1000 deg2	0.5:1 arcsec	10000	3.8 hr	150	Mid 4-5	m	149	0.116	17.25
			Imaging	1000 - 1700 MHz	7.8 deg2	0.5:1 arcsec	2000	95 hr	150	Mid 4-5	l	496	0.023	11.48
			Imaging	1000 - 1700 MHz	0.38 deg2	0.5:1 arcsec	2000	2000 hr	150	Mid 4-5	xl	1438	0.023	33.29
			Imaging	7 - 11 GHz	0.5 deg2	0.05:1 arcsec	1000	16.4 hr	150	Mid 11-12	m/l	18	0.012	0.21
			Imaging	7 - 11 GHz	30 arcmin2	0.05:1 arcsec	1000	1000 hr	150	Mid 11-12	xl	42	0.012	0.49

Table 4. Key parameters of the High Priority Science Objectives and their computing requirements. Under the "HPC" heading are columns that place each experiment into the current context of B_{Max}, sub-band coverage and depth category, together with the associated instantaneous and time averaged computing needs assuming fractional allocations proportional to the time requested.

4 Conclusions

4.1 Implications of Staged HPC Deployment

It has been recognised from its inception, that the SKA will rely on a substantial compute capacity to deliver its science-ready data products. For this reason, the "Science Data Processor" design element has been charged with developing a software architecture that in the procurement phase will allow suitable high throughput pipelines to be developed and deployed on suitable high performance hardware platforms. An outcome of the recent Cost Control Process, as described at the link below,

24th SKA Board Meeting – July 2017

is the likelihood that the initial SKA1 HPC deployment may be smaller than previously anticipated. While unfortunate, it is also of relevance that the resources for an HPC refresh back up to the full Design scope are already planned within the Operations Budget. As such, a highly constrained HPC environment is only likely to persist for the first few years. A very pertinent question to answer is: what might the scientific implications be of a reduced initial deployment?

Based on the analysis presented in the previous Sections, the science implications of various HPC deployments are beginning to be better understood. With an initial HPC deployment of 50 – 85 PFlops, it is likely that some forms of commensality in particular will be reduced in the first instance. While a relatively broad range of experiments can be successfully processed with this capacity, they may be constrained to only one or a few sub-bands (of 32% fractional band-width) at a time within the lower frequency ranges of both SKA1-Low and SKA1-Mid. Since shallow integrations are easiest to support, there may well be utility in an early focus on shallow, wide-field surveys rather than the deep pointed observations that are most computationally demanding. Once an HPC refresh to the level of about 260 PFlops has been realised, most of these constraints would be removed. The full range of anticipated use cases could be supported with such an in-house HPC capacity with a relatively flat distribution of time fractions over use cases.

It should also be noted that there may be significant external HPC capacity within the network of SKA Regional Centres (SRCs). It has already been acknowledged that for some specific use cases, most notably the Epoch of Reionisation processing, a well-defined hand-over point may be utilised, of partially calibrated averaged visibilities rather than fully calibrated image cubes. It may be the case more generally, that Key Science Project Teams organise HPC resourcing for the most demanding experiments within the SRC context. If this were the case, it would reduce pressure on the mix of experiments that could be supported by the centralised Observatory HPC capacity. On the other hand, the most computationally demanding aspects of calibration are best undertaken while there is access to the visibilities with full time and frequency resolution and it is not foreseen that there will be either enough archive capacity to store these indefinitely or enough data transport capacity to distribute them to remote processing sites. It will be important that the centralised HPC resources are adequate.

4.2 Caveats

There are many uncertainties in the estimates presented here that should be noted. Some of the most serious are:

1. The computational efficiency that can be achieved in a real HPC deployment will depend critically on the combination of software and hardware design. Within this document we have assumed a 10% efficiency throughout. This could be viewed as pessimistic by factors of many, which would have major implications for the associated HPC sizing.
2. The ICAL component of the parametric model (which completely dominates the HPC cost projections) has only a limited degree of support for Direction Dependant (DD) calibration methods to calibrate the ionosphere as well as to track time-variable off-axis instrumental gain. More comprehensive support for such methods within the model would improve its predictive value.
3. The computational cost within ICAL in the estimates presented here is completely dominated (at the 95% level) by the "Predict via DFT" step that scales linearly with the number of source components being modelled. It is conceivable that under some circumstances it may be sufficient to make use of an FFT, or faceted set of FFTs, for this purpose. If this were the case, it could result in HPC savings by factors of several relative to what is reported here. Alternately, as suggested by M. Ashdown (SDP Project Scientist), processing bottlenecks such as the DFT may be well-matched to highly efficient implementation on specialised hardware platforms such as Graphical Processing Units (GPUs). This could lead to large savings relative to what is reported here. A case in point is the LOFAR EoR calibration and imaging pipeline that is currently demonstrating greater than 80% computational efficiency with existing code and GPU hardware. **If comparable efficiencies were achieved with SKA-scale data rates, it would reduce all of the HPC estimates made in this document by more than a factor of eight.**