SDP Memo 049:
SDP limitations on wide-area mapping mode

Document Number .... SDP Memo 049
Document Type .... MEMO
Revision 2
Author .... Mark Ashdown
Release Date 2018-07-31
Document Classification .... Unrestricted
Status .... Draft

Lead Author	Designation	Affiliation
Mark Ashdown	SDP Project Scientist	University of Cambridge
Signature & Date: M. Ashdown M. Ashdown (Nov 19, 2018)		

Revision	Date of issue	Prepared by	Comments
1	2017-12-04	Mark Ashdown	Limitations from output data rate
2	2018-07-31	Mark Ashdown	Added section on computational load

SDP Memo Disclaimer

Table of Contents

1 Introduction

Wide-area mapping mode (WAMM) is a proposed mode of operation of SKA1-Mid in which the dishes are scanned continuously with respect to the sky. In this document we assess what the requirements on the SDP would be for running standard interferometric imaging pipelines on data acquired in this mode. From this we can determine which observing configurations the SDP can support given the resources that will be available. We do not consider other types of observation, such as intensity mapping, which would require non-standard pipelines.

In the following analysis we assume that the images from WAMM cover a strip of sky, with individual square images adjoining but not overlapping. We also assume the size of the images (and the width of the strip) the be equal to the first null of the primary beam of a dish with a $15\,\mathrm{m}$ aperture. We consider the following cases in the analysis:

- The six frequency bands shown in Table 1. The number of sub-bands, N_{sb}, is set by requiring that the ratio of the maximum to minimum frequency of each sub-band is not greater than 1.35 (equivalent to a fractional bandwidth ≈ 0.3). The resulting sub-bands are evenly spaced in the logarithm of the frequency.
- Three values of the maximum baseline, B_{max}, $150\,\mathrm{km}$, $15\,\mathrm{km}$ or $1.5\,\mathrm{km}$. The first value corresponds to the the full array.

- Three types of output, continuum images, and spectral images with 100 or 1000 channels.
- Three values of the angular scan speed, ω_{scan}. The first, $\omega_{\mathrm{scan}} = 4.18 \times 10^{-3}\,\mathrm{deg\,s^{-1}}$, is the sidereal rate, equivalent to making a drift scan observation at the celestial equator. The other two are $\omega_{\mathrm{scan}} = 0.1\,\mathrm{deg\,s^{-1}}$ and $\omega_{\mathrm{scan}} = 1.0\,\mathrm{deg\,s^{-1}}$.

Table 1: Frequency bands considered in this analysis. f_{min} and f_{max} are the minimum and maximum frequency, respectively, and N_{sb} is the number of sub-bands in the image.

Band	$f_{\mathrm{min}}/\mathrm{GHz}$	$f_{\mathrm{max}}/\mathrm{GHz}$	N_{sb}	Notes
1	0.35	1.05	4	
2	0.95	1.76	3	
5a(l)	4.6	7.1	2	lowest 2.5 GHz of band 5a
5a(h)	6.0	8.5	2	highest 2.5 GHz of band 5a
5b(l)	8.3	10.8	1	lowest 2.5 GHz of band 5b
5b(h)	12.9	15.4	1	highest 2.5 GHz of band 5b

We analyse the output data rate in Section 2 and the computational load in Section 3.

2 Output data rate

In this section, we calculate the output image data rates from the SDP for the various cases described in Section 1. We compare them to the other data rates in the system and consider them in relation to the capacity and performance of the SDP required for standard pointed imaging. Note all the data rates in this section are given in gigabytes per second ($\mathrm{GB\,s^{-1}}$).

The image data rate from the SDP can be compared to two other data rates in the system. The first is the input visibility data rate, R_{input} from the CSP. This rate depends on the number of baselines, and the values for the three maximum baselines considered in this analysis are shown in Table 2. The other relevant data rate is the capacity of the outgoing link from the SDP to the SRCs. This will be a $100\,\mathrm{Gbit\,s^{-1}}$ link, which we assume can sustain an output data rate $R_{\mathrm{output}} = 10\,\mathrm{GB\,s^{-1}}$.

Table 2: Input visibility data rate, R_{input}, for the values of the maximum baseline, B_{max}, considered in this analysis.

B_{max}/ km	R_{input}/ $\mathrm{GB\,s^{-1}}$
150	365
15	253
1.5	114

If the image data rate exceeds the input visibility data rate, this implies that the data is not being compressed by making images, and the resulting output is redundant. In this case, the SDP would not be performing its principal function in the system which is to compress the data. Indeed the SDP is expected to compress the data by at least a factor of ~ 40 on average, based on the ratio of the expected input and output data rates.

For standard pointed imaging, the SDP buffer will require a sustained write capacity of $\sim 400\,\mathrm{GB\,s^{-1}}$ to ingest the incoming visibilities into the buffer from the CSP, and a sustained read capacity of $\sim 4000\,\mathrm{GB\,s^{-1}}$, which is 10 times the write capacity, to support reading the data multiple times during iterative calibration and imaging. Therefore if the image data rate is even comparable to the visibility data rate, then it may exceed the write capacity of the buffer.

The capacity of the outgoing link to the SRCs gives a more stringent limit on the output data rate. It may be possible to exceed this limit for short periods, but this may require additional buffer capacity to cache the images which would have immediate cost implications for the SDP. It would also fill up the the outgoing link for a corresponding period of time. Over the long term, the average output data rate must be at or below the capacity of the outgoing link.

If we are making a continuous image of a strip of the sky as wide as the first null of the primary beam, the output data rate from the SDP would be

$$R_{\mathrm{image}} = 3.03 \times 10^{-4} \times \frac{\left(\omega_{\mathrm{scan}}/\mathrm{deg\,s^{-1}}\right)(f_{\max}/\mathrm{GHz})^2\,(B_{\max}/\mathrm{km})^2\,N_{\mathrm{f}} N_{\mathrm{p}} N_{\mathrm{t}}}{(f_{\min}/\mathrm{GHz})}\,\mathrm{GB\,s^{-1}} \quad (1)$$

where ω_{scan} is the angular speed of the scan, $f_{\min}$ the minimum frequency of the band, $f_{\max}$ the maximum frequency of the band, $B_{\max}$ the longest baseline used to make the image, N_{f} the number of output frequency channels, N_{p} the number of output polarisation channels, and N_{t} the number of Taylor terms (for continuum images). This assumes that the whole band is imaged together. If the band is split into sub-bands for imaging, the rate can be calculated by evaluating Eqn. 1 for each of the sub-bands and summing the results.

Note this calculation assumes that the image is the only data product being generated. In the standard imaging pipelines there are a number of ancillary data products of comparable size produced with the image.

R_{image} has been calculated for combinations of the parameter values described in Section 1. The resulting image data rates are shown in Table 3.

3 Computational load

In this section, we estimate the computational requirements on the SDP for processing data from WAMM. This estimate is based on the assumption that the standard imaging pipelines can be used to process the data. It is not known if the standard pipelines could be modified to work for WAMM, or substantially different pipelines would be needed. In either case, these non-standard pipelines would have to be analysed and incorporated into the model to get an accurate assessment of the computational requirements. This would be a substantial piece of work, and it is beyond the scope of this memo.

Since we are assuming that we are mapping a strip of sky as wide as the first null of the primary beam, the field-of-view quality parameter, Q_{fov}, has been set to 1 for all of the pipelines. This is appropriate for low dynamic range images that would be obtained from short observations. The length of the observation, T_{obs}, going into each image in the strip depends on the scan speed. We consider the three values of the scan speed described in Section 1, and to establish a point of comparison we also estimate the computational requirements for a 1-hour pointed observation, with all of the other parameters the same. The requirements are shown in units of $\mathrm{Pflop\,s^{-1}}$ (10^{12} floating point operations per second) and they are therefore the rate that would

Table 3: Imaging data rate, R_{image}, in $\mathrm{GB\,s^{-1}}$ for the cases considered in this study. 'Cont' is continuum imaging, 'Sp100' is spectral imaging with 100 channels, and 'Sp1000' is spectral imaging with 1000 channels. Values in red exceed the output data rate to the regional centres of $10\,\mathrm{GB\,s^{-1}}$.

Band	Scan Speed	$B_{\max} = 150\,\mathrm{km}$ Cont	$B_{\max} = 150\,\mathrm{km}$ Sp100	$B_{\max} = 150\,\mathrm{km}$ Sp1000	$B_{\max} = 15\,\mathrm{km}$ Cont	$B_{\max} = 15\,\mathrm{km}$ Sp100	$B_{\max} = 15\,\mathrm{km}$ Sp1000	$B_{\max} = 1.5\,\mathrm{km}$ Cont	$B_{\max} = 1.5\,\mathrm{km}$ Sp100	$B_{\max} = 1.5\,\mathrm{km}$ Sp1000
1	Sidereal	11.9	11.9	119	0.119	0.119	1.19	0.00119	0.00119	0.0119
	$0.1\,\mathrm{deg\,s^{-1}}$	285	285	2850	2.85	2.85	28.5	0.0285	0.0285	0.285
	$1.0\,\mathrm{deg\,s^{-1}}$	2850	2850	28500	28.5	28.5	285	0.285	0.285	2.85
2	Sidereal	20.9	20.9	209	0.209	0.209	2.09	0.00209	0.00209	0.0209
	$0.1\,\mathrm{deg\,s^{-1}}$	500	500	5000	5	5	50	0.05	0.05	0.5
	$1.0\,\mathrm{deg\,s^{-1}}$	5000	5000	50000	50	50	500	0.5	0.5	5
5a(l)	Sidereal	91.7	91.7	917	0.917	0.917	9.17	0.00917	0.00917	0.0917
	$0.1\,\mathrm{deg\,s^{-1}}$	2200	2200	22000	22	22	220	0.22	0.22	2.2
	$1.0\,\mathrm{deg\,s^{-1}}$	22000	22000	220000	220	220	2200	2.2	2.2	22
5a(h)	Sidereal	107	107	1070	1.07	1.07	10.7	0.0107	0.0107	0.107
	$0.1\,\mathrm{deg\,s^{-1}}$	2560	2560	25600	25.6	25.6	256	0.256	0.256	2.56
	$1.0\,\mathrm{deg\,s^{-1}}$	25600	25600	256000	256	256	2560	2.56	2.56	25.6
5b(l)	Sidereal	160	160	1600	1.6	1.6	16	0.016	0.016	0.16
	$0.1\,\mathrm{deg\,s^{-1}}$	3830	3830	38300	38.3	38.3	383	0.383	0.383	3.83
	$1.0\,\mathrm{deg\,s^{-1}}$	38300	38300	383000	383	383	3830	3.83	3.83	38.3
5b(h)	Sidereal	209	209	2090	2.09	2.09	20.9	0.0209	0.0209	0.209
	$0.1\,\mathrm{deg\,s^{-1}}$	5010	5010	50100	50.1	50.1	501	0.501	0.501	5.01
	$1.0\,\mathrm{deg\,s^{-1}}$	50100	50100	501000	501	501	5010	5.01	5.01	50.1

need to be sustained for the duration of the observation if the processing were done in real time. We break down the requirements into the following items:

- Real-time processing consisting of Ingest, Real-time calibration (RCAL) and Fast imaging.
- Iterative self-calibration (ICAL)
- Continuum imaging (DPrepA)
- Spectral imaging with 100 channels (DPrepB)
- Spectral imaging with 1000 channels (DPrepC)

The rates for the individual items must be summed to get the total rate required for the processing, although not all of the imaging pipelines (DPrep{A,B,C}) need to be run for a particular observation.

Note the computational requirements in this section are the effective rates required for the processing, which do not take into account the efficiency of the hardware. The standard computational efficiency adopted by the SDP consortium is $10\,\%$, therefore the results must be multiplied by 10 to convert to the theoretical computational requirement for the hardware.

The effective requirements are shown for the three maximum baselines in Tables 4, 5, and 6. They should be compared to the effective computational rate of the Mid SDP in the design baseline, which is $\sim 12.5\,\mathrm{Pflop\,s^{-1}}$ (this assumes half of a total of $250\,\mathrm{Pflop\,s^{-1}}$ theoretical compute for Phase 1, with an computational efficiency of $10\,\%$.) The computational load for an observation can exceed this value, but it comes at the cost of taking longer to run the batch

Table 4: Effective computational rate in $\mathrm{Pflop\,s^{-1}}$ for $B_{\max} = 150\,\mathrm{km}$. 'Cont' is continuum imaging, 'Sp100' is spectral imaging with 100 channels, and 'Sp1000' is spectral imaging with 1000 channels.

Band	Scan speed	Real-time	ICAL	Cont	Sp100	Sp1000
1	Pointed	2.002	6.369	7.967	7.708	9.521
	Sidereal	2.002	6.371	7.967	7.714	9.618
	$0.1\,\mathrm{deg\,s^{-1}}$	2.002	6.466	7.978	8.029	12.699
	$1.0\,\mathrm{deg\,s^{-1}}$	2.002	6.980	8.080	9.272	24.862
2	Pointed	1.586	3.857	4.802	4.681	8.286
	Sidereal	1.586	3.863	4.803	4.711	8.617
	$0.1\,\mathrm{deg\,s^{-1}}$	1.586	3.985	4.824	5.074	12.395
	$1.0\,\mathrm{deg\,s^{-1}}$	1.586	4.689	4.959	8.313	44.794
5a(l)	Pointed	1.306	2.638	3.277	3.200	6.210
	Sidereal	1.306	2.648	3.277	3.270	6.990
	$0.1\,\mathrm{deg\,s^{-1}}$	1.306	2.890	3.299	4.945	23.743
	$1.0\,\mathrm{deg\,s^{-1}}$	1.306	5.163	3.505	20.680	181.091
5a(h)	Pointed	1.161	2.055	2.541	2.494	4.886
	Sidereal	1.161	2.068	2.542	2.586	5.828
	$0.1\,\mathrm{deg\,s^{-1}}$	1.161	2.383	2.570	4.747	27.445
	$1.0\,\mathrm{deg\,s^{-1}}$	1.161	5.344	2.833	25.051	230.482
5b(l)	Pointed	1.068	1.638	2.044	1.993	3.887
	Sidereal	1.068	1.653	2.045	2.125	5.192
	$0.1\,\mathrm{deg\,s^{-1}}$	1.068	1.872	2.065	5.186	35.802
	$1.0\,\mathrm{deg\,s^{-1}}$	1.068	3.926	2.254	33.936	323.303
5b(h)	Pointed	0.913	1.038	1.296	1.263	2.474
	Sidereal	0.913	1.053	1.297	1.463	4.473
	$0.1\,\mathrm{deg\,s^{-1}}$	0.913	1.392	1.327	6.114	50.984
	$1.0\,\mathrm{deg\,s^{-1}}$	0.913	4.575	1.611	49.799	487.830

pipelines and also has an impact on the buffer because the visibilities need to be stored for longer. The scheduling of batch processing inside the SDP enables it to cope with the varying computational requirements from different observations, but the average load cannot exceed that supported by the hardware. The timescale over which this average load condition must be satisfied is determined by the size of the buffer, and it is likely to be a few days.

4 Conclusion

The output data rates from shown in Section 2 constrain the observing configurations that can be supported by the SDP. Imaging using the full array is extremely challenging because all combinations of frequency band and scan speed produce data rates that exceed the $10\,\mathrm{GB\,s^{-1}}$ capacity of the outgoing link, sometimes vastly so. It might be possible to do continuum imaging or coarse spectral imaging in bands 1 and 2 for drift scans, if there is sufficient capacity in the SDP buffer, since their data rates are not too much greater than that of the outgoing link. Restricting the imaging to a small sub-array makes the problem more tractable, since the

Table 5: Effective computational rate in $\mathrm{Pflop\,s^{-1}}$ for $B_{\max} = 15\,\mathrm{km}$. 'Cont' is continuum imaging, 'Sp100' is spectral imaging with 100 channels, and 'Sp1000' is spectral imaging with 1000 channels.

Band	Scan speed	Real-time	ICAL	Cont	Sp100	Sp1000
1	Pointed	0.126	0.137	0.165	0.167	0.575
	Sidereal	0.126	0.137	0.165	0.170	0.598
	$0.1\,\mathrm{deg\,s^{-1}}$	0.126	0.174	0.168	0.281	1.708
	$1.0\,\mathrm{deg\,s^{-1}}$	0.126	0.514	0.193	1.324	12.137
2	Pointed	0.104	0.071	0.086	0.088	0.311
	Sidereal	0.104	0.074	0.086	0.099	0.417
	$0.1\,\mathrm{deg\,s^{-1}}$	0.104	0.148	0.091	0.400	3.430
	$1.0\,\mathrm{deg\,s^{-1}}$	0.104	0.841	0.142	3.230	31.729
5a(l)	Pointed	0.099	0.050	0.060	0.063	0.226
	Sidereal	0.099	0.060	0.061	0.124	0.836
	$0.1\,\mathrm{deg\,s^{-1}}$	0.099	0.298	0.079	1.583	15.426
	$1.0\,\mathrm{deg\,s^{-1}}$	0.099	2.538	0.243	15.286	152.458
5a(h)	Pointed	0.094	0.039	0.046	0.049	0.179
	Sidereal	0.094	0.052	0.047	0.129	0.983
	$0.1\,\mathrm{deg\,s^{-1}}$	0.094	0.363	0.070	2.032	20.011
	$1.0\,\mathrm{deg\,s^{-1}}$	0.094	3.283	0.285	19.904	198.726
5b(l)	Pointed	0.096	0.032	0.038	0.041	0.153
	Sidereal	0.096	0.041	0.039	0.153	1.275
	$0.1\,\mathrm{deg\,s^{-1}}$	0.096	0.256	0.055	2.786	27.603
	$1.0\,\mathrm{deg\,s^{-1}}$	0.096	2.276	0.203	27.515	274.893
5b(h)	Pointed	0.089	0.020	0.024	0.026	0.105
	Sidereal	0.089	0.034	0.025	0.202	1.863
	$0.1\,\mathrm{deg\,s^{-1}}$	0.089	0.368	0.049	4.293	42.773
	$1.0\,\mathrm{deg\,s^{-1}}$	0.089	3.508	0.280	42.717	427.014

Table 6: Effective computational rate in $\mathrm{Pflop\,s^{-1}}$ for $B_{\mathrm{max}} = 1.5\,\mathrm{km}$. ‘Cont’ is continuum imaging, ‘Sp100’ is spectral imaging with 100 channels, and ‘Sp1000’ is spectral imaging with 1000 channels.

Band	Scan speed	Real-time	ICAL	Cont	Sp100	Sp1000
1	Pointed	0.003	0.002	0.002	0.004	0.031
	Sidereal	0.003	0.003	0.002	0.007	0.054
	$0.1\,\mathrm{deg\,s^{-1}}$	0.003	0.039	0.005	0.117	1.162
	$1.0\,\mathrm{deg\,s^{-1}}$	0.003	0.380	0.030	1.159	11.573
2	Pointed	0.002	0.001	0.001	0.004	0.030
	Sidereal	0.002	0.004	0.001	0.014	0.135
	$0.1\,\mathrm{deg\,s^{-1}}$	0.002	0.078	0.007	0.315	3.144
	$1.0\,\mathrm{deg\,s^{-1}}$	0.002	0.771	0.057	3.141	31.402
5a(l)	Pointed	0.002	0.001	0.001	0.003	0.030
	Sidereal	0.002	0.011	0.001	0.064	0.639
	$0.1\,\mathrm{deg\,s^{-1}}$	0.002	0.249	0.019	1.521	15.207
	$1.0\,\mathrm{deg\,s^{-1}}$	0.002	2.488	0.183	15.204	152.036
5a(h)	Pointed	0.002	0.001	0.001	0.003	0.030
	Sidereal	0.002	0.014	0.002	0.083	0.832
	$0.1\,\mathrm{deg\,s^{-1}}$	0.002	0.325	0.024	1.984	19.834
	$1.0\,\mathrm{deg\,s^{-1}}$	0.002	3.245	0.238	19.831	198.306
5b(l)	Pointed	0.002	0.001	0.000	0.003	0.030
	Sidereal	0.002	0.010	0.001	0.115	1.150
	$0.1\,\mathrm{deg\,s^{-1}}$	0.002	0.225	0.017	2.744	27.436
	$1.0\,\mathrm{deg\,s^{-1}}$	0.002	2.245	0.165	27.432	274.323
5b(h)	Pointed	0.002	0.000	0.000	0.003	0.030
	Sidereal	0.002	0.015	0.001	0.178	1.785
	$0.1\,\mathrm{deg\,s^{-1}}$	0.002	0.349	0.026	4.264	42.639
	$1.0\,\mathrm{deg\,s^{-1}}$	0.002	3.488	0.256	42.635	426.354

imaging data rate scales as $B_{\max}^2$.

The computational requirements shown in Section 3 provide additional constraints on the observing configurations. The requirements of the batch pipelines increase with scan speed, but for ICAL and the continuum imaging pipelines the increase is modest and can be accommodated by the SDP. The requirements of the spectral imaging pipelines increase much faster with scan speed, and the increase is larger for the higher frequencies. This rules out doing spectral imaging at the fastest scan speeds, even with a very short maximum baseline.