SDP Memo: Receive and Pre-process Visibility Data

Document number	SKA-TEL-SDP-0000028
Document type	REP
Revision	01C
Author	G. van Diepen, A.-J. Boonstra, A. Offringa, B. van der Tol, Á. Mika
Release date	2016-04-07
Document classification	Unrestricted
Status	Draft

Lead author:

Name	Designation	Affilitation
G. van Diepen	SDP.PIP.INP lead	ASTRON
Signature & Date:	Ger van Diepen Ger van Diepen (Apr 7, 2016) diepen@astron.nl	

Released by:

Name	Designation	Affilitation
R. Bolton	SDP Project Scientist	University of Cambridge
Signature & Date:	Rosie Bolton Rosie Bolton (Apr 7, 2016) rosie@mrao.cam.ac.uk	

Version	Date of issue	Prepared by	Comments
1.0	2015-02-06	A.-J. Boonstra	PDR submission
1.1	2015-04-10	A.-J. Boonstra	PDR panel comments addressed
01C	2016-04-07	G. van Diepen	dPDR submission

ORGANISATION DETAILS

Name	Science Data Processor Consortium

Table of Contents

List of abbreviations

AO flagger	André Offringa flagger
BDA	Baseline Dependent Averaging
CRAF	Committee on Radio Astronomy Frequencies
CSP	Central Signal Processor
ESF	European Science Foundation
FLOPS	Floating-point Operations per Second
FoV	Field of View
ICD	Interface Control Document
ITU	International Telecommunication Union
IUCAF	Scientific Committee on Frequency Allocations for Radio Astronomy and Space Science
LOFAR	The Low Frequency Array Radio telescope
MWA	Murchison Widefield Array
NSF	National Science Foundation
NRC	National Research Council
PDR	Preliminary Design Review
RA	Radio Astronomy
RFI	Radio Frequency Interference
SAGECal	Space Alternating Generalised Expectation Maximisation Calibration
SDP	Science Data Processor
SKA1	The Square Kilometre Array, phase 1
SKA1-Low	SKA1-Low telescope
SKA1-Mid	SKA1-Mid telescope
StEFCal	Statistically Efficient and Fast Calibration
TBD	To Be Decided
TM	Telescope Manager
VLA	Very Large Array
WSRT	Westerbork Synthesis Radio Telescope

List of symbols

$\mathbf{A}$	Re-phasing matrix
$\mathbf{C}$	Correction matrix
F_{s}	Correlator output rate in samples per second
$\mathbf{M}$	Mixing matrix
M_{bw}	Memory bandwidth (bytes s^{-1})
N_{A}	Number of interfering directions
N_{aps}	Number of accesses per visibility
N_{beams}	Number of beams
N_{bits}	Number of bits
N_{bl}	Number of baselines
N_{bpa}	Number of Bytes per access
N_{f}	Number of frequency bins
$N_{\mathrm{f,av}}$	Number of averaged frequency bins
N_{FLOPS}	Number of FLOPS
N_{fps}	Number of FLOP per sample
N_{pol}	Number of polarisations
N_{rt}	Number of telescopes
N_{t}	Number of time bins
$N_{\mathrm{t,av}}$	Number of averaged time bins
N_{vis}	Number of visibilities inc. autocorrelations
$\mathbf{R}$	Correlation matrix
$\mathbf{R}_{filt}$	Spatially filtered correlation matrix
$t - f$	Time – frequency
T_{s}	CSP visibility output integration time (correlator dump time) (s)
$(u, v, w)_{\mathrm{av}}$	(u, v, w) coordinates of averaged $t - f$ samples
V	Visibility
V_{corr}	Corrected visibility
V_{obs}	Observed visibility
$\mathbf{W}, \mathbf{W}_{filt}$	Spatial filter matrix
η	Fraction of a band occupied by RFI

List of Figures

Summary

This document provides the background and details of the receive and pre-processing components parametric model as summarised in PDR deliverable AD02. The document should also inform the PIP Design Documentation. The derived parametric equations are applied towards estimating the computational cost and memory bandwidth requirements of the various ingest pipeline components for each of the SKA1 telescopes, SKA1-Low and SKA1-Mid.

Applicable and reference documents

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.

Reference Number	Reference
AD01	SKA-TEL.SDP.SE-TEL.CSP.SE-ICD-001 Ratcliffe, S., Graser, F. and Carlson, B.: SKA1 Interface Control Document SDP to CSP, 2014.
AD02	SKA-TEL-SDP-0000040 Bolton, R., Cornwell, T. J. et al.: Parametric models of SDP compute requirements, 2016
AD03	SKA-TEL-SKO-0000040 Dewdney, P. and Tan, G. H.: SKA EMI/EMC standards and procedures, 2015
AD04	SKA-TEL-SDP-0000027 Nijboer, R et al.: SDP Pipelines Design, 2016

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.

Reference Number	Reference
RD01	SKA-TEL-SDP-0000017 Wijnholds, S., Salvini, S. and Dodson, R.: Baseline-dependent averaging, 2015
RD02	van der Tol, S.: Weight computation in NDPPP, 2011 `http://www.lofar.org/operations/lib/exe/fetch.php?media=public:user_software:ndppp_weights.pdf`
RD03	Boonstra, A.-J., van der Veen, A.-J. and Raza, J.: IEEE International Conference on Acoustics, Speech, and Signal Processing, Spatial filtering of continuous interference in radio astronomy, volume 3, pages 2933-2936, 2002, doi 10.1109/ICASSP.2002.5745263
RD04	Boonstra, A.-J.: Radio frequency interference mitigation in radio astronomy, PhD Thesis at Technische Universiteit Delft, 2005
RD05	Cohen, J., et al.: CRAF Handbook for Radio Astronomy (Third Edition), 2005
RD06	Hellbourg, G., Weber, R., Capdessus, C. and Boonstra, A.-J.: Statistical Signal Processing Workshop, IEEE, pages 93-96 Oblique projection beamforming for RFI mitigation in radio astronomy, 2012, doi 10.1109/SSP.2012.6319860

Reference Number	Reference
RD07	ITU, Protection criteria used for radio astronomical measurements, Radio Astronomy Series, ITU-R RA.769-2, 2003
RD08	ITU, Levels of data loss to radio astronomy observations and percentage-of-time criteria resulting from degradation by interference for frequency bands allocated to the radio astronomy on a primary basis, Radio Astronomy Series, ITU-R RA.1513, 2003
RD09	ITU, Techniques for mitigation of radio frequency interference in radio astronomy, Radio Astronomy Series, ITU-R RA.2126-1, 2013
RD10	ITU, Handbook on Radio Astronomy, 2013
RD11	NRC, Spectrum Management for Science in the 21st Century, The National Academies Press, Washington, DC, 2010, ISBN 13: 978-0-309-14686-9
RD12	Offringa, A. R., de Bruyn, A. G., Zaroubi, S. and Biehl, M.: A LOFAR RFI detection pipeline and its first results, Proc. of Science, RFI2010
RD13	Offringa, A. R.: Algorithms for Radio Interference Detection and Removal, PHD Thesis at University of Groningen, The Netherlands, 2012
RD14	Offringa, A. R., J. J. van de Gronde and J. B. T. M. Roerdink, A morphological algorithm for improved radio-frequency interference detection, A&A volume 539, 2012
RD15	Offringa, A. R., Wayth, R. B., Hurley-Walker, N. and others The low-frequency environment of the Murchison Widefield Array: radio-frequency interference analysis and mitigation, PASA, 2015
RD16	RFI2010 RFI Mitigation Workshop Groningen, The Netherlands, Proceedings of Science (PoS), 2010
RD17	van der Tol, S., Jeffs, B.D. and van der Veen, A.-J.: Self-Calibration for the LOFAR Radio Astronomical Array, IEEE Transactions on Signal Processing, volume 55, pages 4497-4510, 2007, doi 10.1109/TSP.2007.896243
RD18	van der Tol, S.: Bayesian estimation for ionospheric calibration in radio astronomy, PHD Thesis at Technische Universiteit Delft, The Netherlands, 2009
RD19	Kazemi, S., S. Yatawatta, S. Zaroubi, P. Lampropoulos, de Bruyn, A. G., Koopmans, L. V. E. and J. Noordam: Radio interferometric calibration using the SAGE algorithm, Monthly Notice of the Royal Astronomical Society, volume 414, pages 1656-1666, 2011, doi 10.1111/j.1365-2966.2011.18506.x

1 Introduction

The ingest pipeline for the visibility data consists of a receive pipeline and two pre-processing pipelines: fast pre-processing and buffered pre-processing. Both pre-processing pipelines have the same functionality, but different latency requirements lead to different buffer and window sizes in their components.

The purpose of the receive pipeline is to receive data from the Central Signal Processor (CSP) and merge it with meta data from the Telescope Manager (TM). The purpose of a pre-processing pipeline is to apply data conditioning functions such as flagging and bright source removal prior to integration over time and frequency before sending it to the pipelines further downstream. The pulsar pipelines and the fast transients pipeline do not enter the SDP via the visibility ingest pipeline. The context of the ingest pipelines and their components in the Science Data Processor (SDP) pipelines design is available in document AD04.

This document describes the components of the ingest pipelines and a parametric model of them. The parametric equations are applied towards estimating the computational cost and memory bandwidth requirements.

2 The receive and pre-processing pipeline functions

The processing steps of the pipelines are schematically depicted in Figure 1 and 2.

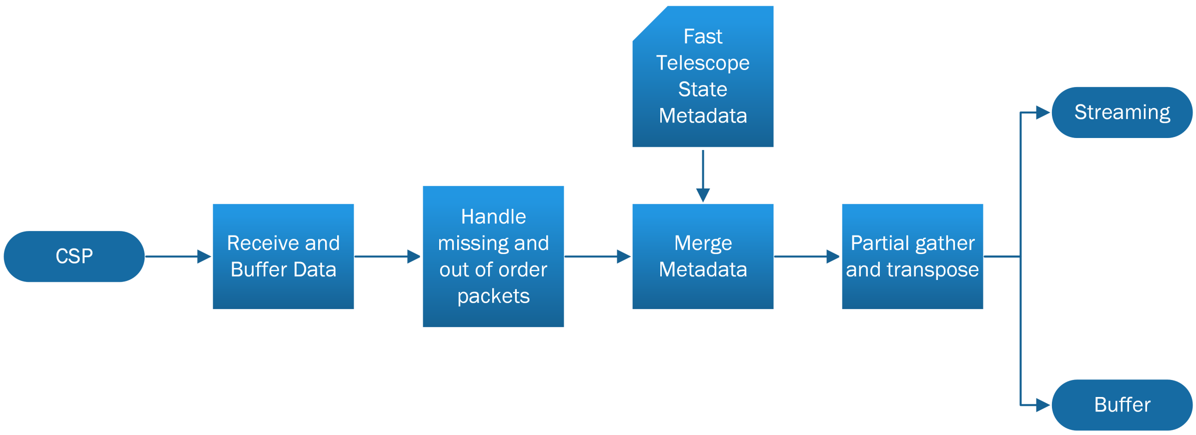

Figure 1: A functional and data flow breakdown of the F1.2 Receive Visibilities function. Visibilities are received from CSP, packaged into Data Drops and duplicated for further streaming and buffered processing.

The receive pipeline has the following components:

- Receive the data from the CSP and buffer them.
- Handle missing and out-of-order data packets and align them within the buffer. The buffer size is different for the data sent to the fast (streaming) and buffered pre-processing pipelines.
- Receive meta-information from TM (through the Fast Telescope State) relevant to the received CSP data, and merge them with the data stream.
- Flag data for misbehaving channels and antennae.

- Calculate data weights from the fraction of data flagged and the autocorrelation data.
- Calculate (u,v,w) coordinates (in metres) per antenna.
- Send the data to the fast pre-processing pipeline and the buffering step (to write the data into the buffer). This can take care of re-ordering the data as needed to make it fit with the input requirements of the pre-processing pipelines.

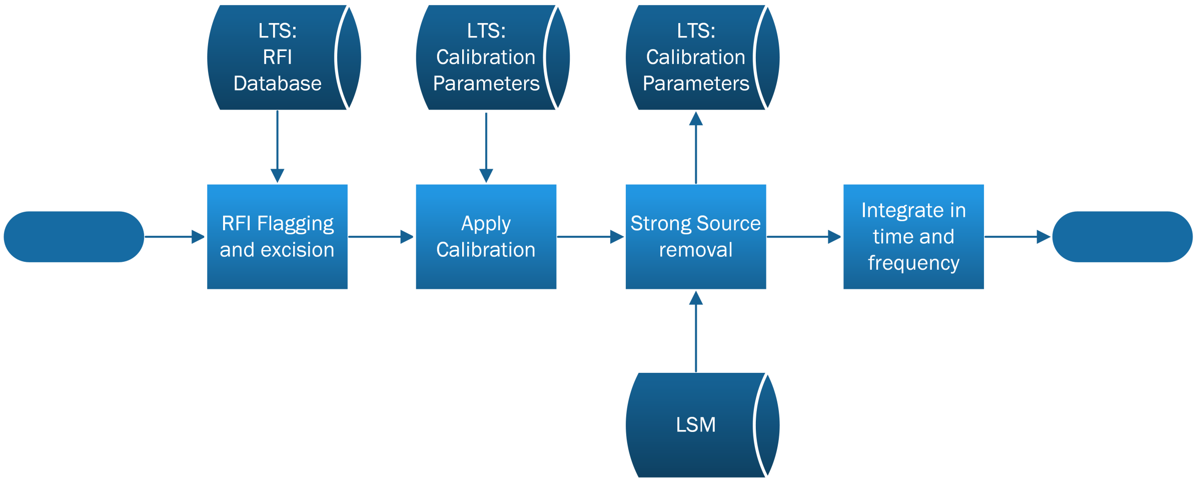

Figure 2: A functional and data flow breakdown of the F1.3 Pre-Process Fast and F1.8 Pre-Process Data functions. Processing steps include Flagging of RFI, Application of known Calibration solutions, Removal of Strong Source signals from outside the FoV, and an optional data compression step.

The pre-processing pipelines have the following components. It is expected that components such as phase rotation and data integration will rarely be used.

- Get the data. The fast pre-processing pipeline receives it from the receive pipeline; the buffered pre-processing pipeline reads it from the buffer.
- The RFI function includes flagging and possibly excision/spatial filtering, removing or attenuating man-made radio interference.
- Remove bright (three in the LOFAR case) sky sources outside the field of view with their spatial side-lobes.
- Apply gain solutions of previous calibration observations.
- Rotate the data and (u,v,w) coordinates to another phase centre.
- Integrate data in time and frequency to a level commensurate with the input requirements of the downstream pipelines.

The behaviour of some components depends on the pipeline they are used in. The Fast pre-processing Pipeline must have a low latency, hence the buffers used in data packet receive and RFI flagging will be much smaller (typically one integration time) than those in the buffered pre-processing Pipelines.

In the following sections, the more computationally intensive pipeline components are described in detail.

3 Merging of meta-data

3.1 Computation of the *(u,v,w)* coordinates

The CSP delivers to SDP visibility data and sufficient meta-data, either directly or via TM, to uniquely identify the received data and to associate it with relevant system components and system settings. The visibility data (two 32-bit single-precision floating point numbers) entering the SDP are associated with a time stamp, (u, v, w) coordinates and several ID's to uniquely identify them in terms of system components and configurations. The visibility meta data includes [AD01]:

- Time centroid: unit-less calculated centroid of the integration interval (8-bit integer).
- Flagging fraction: fraction of data (8-bit integer coded) that was correlated.

Flagging in the ingest pipeline is done on a per $t - f$ (time–frequency) sample basis. In the last processing block of the ingest pipeline (integrate in time and frequency) several time and frequency samples can be averaged. This will lead to smearing in the (u, v, w) space, but the data rate reduction advantage of $t - f$ averaging can outweigh (limited) smearing.

The $(u, v, w)_{\mathrm{av}}$ coordinates of averaged $t - f$ samples are the weighted sums of the individual (u, v, w) samples. The weight of an averaged sample is the sum of the weights of the individual samples. The averaging is done per baseline, per polarisation and per beam.

3.2 Computation of weights

Defining weighting schemes in relation to imaging is not part of the ingest pipeline. Ingest does however calculate a weight for each visibility received from CSP. A simple weighting scheme is to define the weight of a visibility by the ratio of observed time (disregarding the flagged time) and the integration time. This ratio is defined by the flagging fraction received from the CSP. A more complicated weighting scheme (as used by LOFAR) also uses the autocorrelation data to estimate the weight as explained in [RD02].

To set the weights the variance of the visibilities has to be known. The variance of visibility $v_{\mathrm{k,l}}$ over baseline kl is given by

$$\sigma_{\mathrm{kl}}^2 = \sigma_{\mathrm{k}}^2 \sigma_{\mathrm{l}}^2 / N_{\mathrm{int}}$$

where σ_{l}^2 and σ_{k}^2 are the autocorrelations of stations k, l respectively and N_{int} are the number of samples in an integration interval. The autocorrelations can be obtained from the data. An estimate of the variance is given by

$$\sigma_{\mathrm{l}}^2 \approx v_{\mathrm{ll}}$$

Note that the estimate of the variance will be more accurate when based on a longer integration period, at least as long as the noise is stationary over that period. On the other hand, estimation of the weights is not very critical. The information is in the visibilities, the weights only determine how efficiently this information is being used. For practical reasons we assume the weights are estimated from the same integration period as the visibilities.

4 RFI mitigation

The co-existence of active spectrum users and passive users, such as radio astronomy, is guided by rules and regulations set by national and international bodies such as the International Telecommunication Union, ITU. These regulations include agreements on spectrum allocation, spectrum use, protection of services, radio quiet zones, etcetera. Overviews of spectrum management in relation to radio astronomy can for example be found in documents by ESF-CRAF [RD05] and NRC [RD11]. Further information can also be found oon the web sites of ITU (`www.itu.int`), CRAF (`www.craf.eu`), ESF (`www.esf.org`), and IUCAF (`www.iucaf.org`) and NSF (`www.nsf.gov`).

As the protection of passive services in many cases is not perfect, Radio Frequency Interference (RFI) mitigation measures must be taken in order to minimise the detrimental effect on the astronomical end-products. RFI mitigation measures (apart from spectrum regulation) include the establishment of radio quiet zones, shielding of equipment, taking measures to keep receive systems linear, and including active RFI suppression techniques in the radio telescope signal processing chain [AD03].

Lists of selected RFI mitigation related papers, reports and theses on theoretical and experimental RFI mitigation techniques can be found for example in ITU RA 2126 document [RD09] and in the RFI Mitigation overview presented at the 2010 Groningen RFI Mitigation workshop [RD16]. RFI mitigation considerations in relation to radio astronomy can also be found for example in documents by ITU [RD07, RD08 and RD10].

4.1 Definitions

In the following, a few common definitions are listed relating to RFI mitigation.

RFI mitigation: this is the suppression or removal of interfering signals impinging on a telescope or telescope array. This removal can cause secondary effects such as induced non-linearities.

Flagging: this is tagging (flagging) radio astronomical data samples which are known to be affected by RFI. This is usually done by using a real-time detection scheme, but in principle, this could also be done by using a priori or a posteriori knowledge of the transmission scheme. Flagging does not alter the data samples themselves, it only "flags" them. In nearly all cases, flagging is done on time-frequency channels, although it could also be applied to spatial channels.

Excision: this is the same as flagging except that the data samples are actually removed or are replaced by zeros or other numbers. This "cutting-out" can be applied in the time, frequency and spatial domains.

Subtraction or waveform subtraction, or (adaptive noise) cancellation: this means estimating (some of the) properties of the interfering signals, and subtracting the reconstructed RFI signal so as to remove it or at least reduce its amplitude. Adaptive cancellation using a reference antenna can also be considered to be a spatial technique.

Spatial filtering: this is similar to applying a spectral null in a frequency spectrum, but in this case applied in the spatial domain. It can be applied in phased-array stations, but also on correlated data sets. A spatial filter can be applied in a beamformer or at pre-correlation level by using a multiplicative filter-weighting scheme, or in post correlation data by filter-matrix multiplications.

The RFI mitigation measures chosen for the pre-processing pipelines are flagging and excision/spatial filtering. The flagger is expected to operate in most of the imaging observations, the excision/spatial filtering is expected to operate only in selected frequency bands. The CSP flagger operates on sub-integration timescales (correlator output integrated samples); the SDP flagger operates on CSP output integration timescales and longer.

4.2 Flagging

Flagging is the process of detecting and marking samples that are contaminated by radio-frequency interference. An example of a well-developed flagger is the AO flagger described in RD13, and which is used for example in LOFAR and MWA. The AO flagger strategy has been also used for L-band data from e.g., the WSRT and VLA, and the processing requirements for these flaggers provide therefore inputs for estimating the flagging requirements for the different SKA1 telescopes.

4.2.1 Flagging computational costs

Table 3.1 in RD13 shows that the AO flagger has a computational cost of $N_{\rm fps} \sim 278$ FLOP/sample and an estimated total processing load of $N_{\rm FLOPS} = 0.1$ TFLOPS for LOFAR.

$N_{\rm FLOPS}$ is computed using $N_{\rm FLOPS} = N_{\rm fps} N_{\rm pol} N_{\rm f} N_{\rm bl} N_{\rm beams}/T_{\rm s}$, where $N_{\rm pol}$ is the number of visibility polarisations (XX*, YY*, XY*, YX*); $N_{\rm bl} = 1/2 N_{\rm rt}(N_{\rm rt} + 1)$ is the number of visibilities (baselines) including autocorrelations and $N_{\rm rt}$ is the number of telescopes; $N_{\rm f}$ is the number of frequency bins; $T_{\rm s}$ is the correlator dump time in seconds, and $N_{\rm beams}$ is the number of beams. This leads to:

$$N_{\rm FLOPS} \sim 0.5 N_{\rm fps} N_{\rm pol} N_{\rm f} N_{\rm rt}(N_{\rm rt} + 1) N_{\rm beams} T_{\rm s}^{-1} \qquad (1)$$

For LOFAR, assuming $N_{\rm rt} = 38, N_{\rm f} = 256 \times 248, N_{\rm pol} = 4, T_{\rm s} = 1, N_{\rm beams} = 1$ leads to $N_{\rm FLOPS} = 52$ GFLOPS. The LOFAR papers RD12 and RD13 mention 0.1 TFLOPS, but this assumes $N_{\rm rt} = 50$. Naively scaling this to SKA1 (disregarding data access) would yield the following (assuming the SKA1 numbers given in the SKA L1 requirements):

- $N_{\rm f} = 65536$, for both SKA concepts
- $N_{\rm rt} = 512, 197$ for respectively SKA1 Low and Mid
- $N_{\rm beams} = 1, 1$ for respectively SKA1 Low and Mid
- $T_{\rm s} = 0.9, 0.14$ s (correlator dump time) for SKA1 Low and Mid, respectively

Using $N_{\rm fps} = 278$, $N_{\rm pol} = 4$, this leads to:

- SKA1-Low: $N_{\rm FLOPS} \sim 10.7$ TFLOPS
- SKA1-Mid: $N_{\rm FLOPS} \sim 10.2$ TFLOPS

In the case of applying the AO flagger to LOFAR, recent analysis (email comm. Oct. 2014, with A. Offringa) shows that an improved algorithm for the Scale Invariant Rank (SIR) operation leads to a reduction of required resources, reducing $N_{\rm fps} \sim 278$ FLOP/sample to ~ 150 FLOP/sample. The SIR algorithm is a technique to search for RFI-contaminated samples by looking at the density of already established flags. The reduction in computational requirements is the result of a new algorithm for this operation, which is described in RD14. For the MWA, RD15 measures a ~ 25 FLOP/sample. This discrepancy can be explained by different measuring methods and the smaller flagging window size of the flagger employed for the MWA, which results in slightly lower accuracy.

4.2.2 Flagging memory bandwidth

No memory-bandwidth requirements have been published for the AO flagger used by LOFAR, but these can be calculated from the algorithm description. The AO flagger processes the data baseline-by-baseline, which requires a transpose of the visibilities. The flagger iterates over the transposed data and accesses the data a few more times. Because these are local accesses, part of these accesses are cached. It can therefore be assumed that a total of $N_{\rm aps} \approx 4 - 8$ accesses per visibility of $N_{\rm bpa} = 8$ bytes per access are required, which results in a total memory bandwidth of:

$$M_{\rm bw} = 0.5 N_{\rm bpa} N_{\rm aps} N_{\rm pol} N_{\rm f} N_{\rm rt} (N_{\rm rt} + 1) N_{\rm beams} T_{\rm s}^{-1} \quad ({\rm bytes\ s}^{-1}) \tag{2}$$

With the same values as before, this results in:

- SKA1-Low: $M_{\rm bw} \sim 1225$ to 2450 GByte/s
- SKA1-Mid: $M_{\rm bw} \sim 1170$ to 2340 GByte/s

Transposing the data in order to change the ordering from per time step to per baseline is difficult without further partitioning. Such partitioning is therefore performed for both the LOFAR and MWA instruments, where the data are split into 200-kHz sub-bands and 2-min snapshots, respectively.

4.3 Spatial filtering (null steering)

Spatial filtering or null steering is a version of RFI excision which makes use of the side lobes and nulls in the reception pattern of phased arrays. This technique is **NOT** part of the SDP Pipeline Design, but is discussed here for completeness.

By adjusting the beam-former weights, nulls can be positioned in the direction of known RFI sources (both fixed and mobile). However, fast variations in the side lobe gains can confound the calibration algorithm. Therefore, these variations need to be known and often only fixed or slowly varying nulls are allowed [RD09]. In terms of (u, v) data, a spatial filter can be applied by pre- and post-multiplying the correlation matrix $\mathbf{R}$ by a spatial filtering matrix $\mathbf{W}$ (e.g., a projection matrix, see RD06).

The null-forming algorithms applied in radio astronomy fall in the class of techniques called "subspace projections". The associated projection matrices can be derived from a subspace analysis of observed corelation matrices, or they can be constructed using the telescope array geometry and RFI source direction (and source power) if a priori known. The algorithms aim at suppressing RFI entering the system from a particular direction while preserving the characteristics of the main lobe.

Null-steering is an effective way of mitigating RFI from fixed transmitters but also from satellites. A crucial factor for moving transmitters is the update rate of the direction estimate and the spatial filter derived from it. Propagation effects (scattering, multi-path) may reduce the effectiveness of spatial filtering due e.g., to the fact that the apparent direction of incidence spreads out over a significant angular range. The effectiveness of null-forming techniques degrades in the presence of angle spread and with decreasing interference-to-noise ratio.

The filtering process may consist of the following steps (see RD04):

1. Form short-term (sub-second) covariance estimates.
2. If the spatial signature of the interferer is unknown, estimate it by doing an eigen-analysis of the covariance matrix.

3. Form the projection matrix (see RD04 and RD06) using the estimates on the interferer spatial signature vectors and interferer power.
4. Filter the covariance matrix by applying the projection matrix. (Note that during this step the astronomical data will be modified as well. A correction has to be applied later.)
5. Next, the modified short-term covariance matrices are averaged to yield the long-term (order of 10 s) estimate.
6. In order to recover the unbiased estimate of the covariance matrix, a correction needs to be applied to the long-term estimate. This implies a multiplication by the inverse of the correction matrix (constructed from the long-term average of the Kronecker products of the transposed and unchanged filter matrices).

The most computationally expensive parts of the algorithm are the projection (step no. 4, see above) and the post-filtering correction (step no. 6, above). The projection step scales with the third power of the number of antennas. The correction step involves a matrix inversion and scales with the fourth power of the number of antennas, but it can be applied to integrated data.

The computational costs can be greatly reduced if the properties of the interferer are known a priori. In this case the scaling is linear with the number of antennas. If only the direction of the interferer is known then the scaling is quadratic with the number of antennas.

An alternative to spatial filtering is the application of de-mixing to the RFI sources (see below). In this case the interferer is considered to be an astronomical source, which is removed from the (u, v) snapshot just as ordinary sky sources would be treated. There exist various alternatives (of similar complexity) to the spatial filtering algorithm mentioned above.

The spatial filter, unlike flagging, will only be applied in selected frequency bands where time-continuous signals are expected, and where they may block spectral lines or where they may cover a relative large fraction of a (sub-)band.

4.3.1 Spatial filtering computational cost

Let $N_{\rm rt}$ be the number of telescopes, $N_{\rm beams}$ the number of beams, $N_{\rm f}$ the number of frequency channels and $T_{\rm s}$ the correlator dump-time. Then the number of covariance matrices $\mathbf{R}$ per dump time is $N_{\rm f}$; the matrix $\mathbf{R}$ is an $N_{\rm rt} \times N_{\rm rt}$ matrix. Suppose that 10% of the frequency band needs a spatial filter, denote this fraction by η. Assuming the SKA1 numbers given in the SKA L1 requirements:

- $N_{\rm f} = 65536$, for both SKA concepts
- $N_{\rm rt} = 512, 197$ for respectively SKA1 Low and Mid
- $N_{\rm beams} = 1, 1$ for respectively SKA1 Low and Mid
- $T_{\rm s} = 0.9, 0.14$ s for respectively SKA1 Low and Mid

A spatial filter operation can be implemented by multiplying the covariance matrix $\mathbf{R}$ with a spatial filter matrix (projection matrix) $\mathbf{W}_{\rm filt}$ of size $N_{\rm rt} \times N_{\rm rt}$: $\mathbf{R}_{\rm filt} = \mathbf{W}_{\rm filt}\mathbf{R}\mathbf{W}_{\rm filt}$. This would require $\sim N_{\rm rt}^3$ operations for each matrix multiplication. The total compute load would then be:

$$N_{\rm FLOPS} \sim N_{\rm beams} N_{\rm pol} N_{\rm f} \eta N_{\rm rt}^3 T_{\rm s}^{-1}. \quad (3)$$

This yields for:

- SKA1-Low: $N_{\text{FLOPS}} \sim 4$ TFLOPS
- SKA1-Mid: $N_{\text{FLOPS}} \sim 1.5$ TFLOPS

4.3.2 Spatial filter memory bandwidth

A spatial filter applied in the ingest is applied on the integration timescale and once on data averaged to the order of tens of seconds at the end of the ingest pipeline. The short-term spatial filter is applied per frequency bin, per polarisation and per beam, but with all baselines combined. The data needed per spatial filter operation is the number of baselines $N_{\text{bl}} = 0.5N_{\text{rt}}(N_{\text{rt}} + 1)$ times the 2×4 bytes (N_{bpa}) for the complex visibilities. It also needs N_{rt}^2 filter coefficients for each channel and integration time. Assuming a spatial filter is applied to $\eta = 10\%$ of the total band, this leads to a memory bandwidth M_{bw} of:

$$M_{\text{bw}} = (0.5N_{\text{rt}}(N_{\text{rt}} + 1) + N_{\text{rt}}^2)T_{\text{s}}^{-1}\eta N_{\text{f}} N_{\text{pol}} N_{\text{bpa}} \qquad (\text{bytes s}^{-1}) \qquad (4)$$

for all polarisations and all beams.

This yields for:

- SKA1-Low: $M_{\text{bw}} \sim 92$ GByte/s
- SKA1-Mid: $M_{\text{bw}} \sim 87$ GByte/s

4.3.3 Estimating the spatial structure of an interferer

The subspace of an interferer can be estimated "fully", requiring order N_{rt}^3 operations. Estimating the dominant subspace in the case of strong RFI would scale with N_{rt}^2. In the first case, the required FLOPS would be similar to applying the filter itself. In the second case it would be a factor N_{rt} cheaper.

The correction matrix $\mathbf{C}$ mentioned in RD03 and RD04 is expensive as it scales with $(N_{\text{rt}}^2)^3$ (inverse of a Kronecker product of the matrix $\mathbf{R}$). This correction matrix needs to be applied only at timescales of the order of minutes. There also exist alternatives which require less compute power, but which effectiveness needs to be confirmed. One such alternative is applying an oblique projector [RD06] instead of an "ordinary" projection matrix. This oblique projector reduces unwanted distortions of the side-lobe pattern caused by the filter.

5 Removal of very bright sources

There exist several methods to suppress the contribution of off-axis sources (bright sources in the far side lobes) during either pre-processing or calibration. Such techniques are usually used to subtract the so-called A-team sources but they can also be applied for RFI mitigation. Below, we describe one such technique, called de-mixing. Note that de-mixing falls under the category of spatial filtering-type methods.

Another way to remove the bright sources is a direct approach by doing a directional gain calibration and source subtraction as used in packages such as SAGECal [RD19].

5.1 De-mixing

De-mixing is a technique to subtract bright sources outside the field of view. It is used by LOFAR to remove the A-team sources such as CasA and CygA. The signal of these sources will be strongly suppressed by phase-rotating the data to the sources and averaging the data in time and frequency. It uses the following assumptions:

- The data will be averaged to a resolution for which the time and frequency smearing for the sources in the FoV can be considered small.
- The number of very bright interfering sources is small. That is, small compared to the number of samples that will be summed up during the averaging step. Other sources with lower flux densities will be removed in the calibration and imaging pipelines.
- Direction-dependent calibration at the original resolution is prohibitively expensive.
- The time-frequency smearing suppresses the signal from the interfering sources to some extent, but does not completely remove it.
- Calibrating the gains in the direction of the interfering sources on the averaged data is sub-optimal, because part of the signal is already lost due to time-frequency smearing.

However, with new techniques such as SAGECal and StEFCal, the third assumption does not hold anymore and a direction-dependent calibration at the original resolution is feasible as discussed in section 5.2.

5.1.1 De-mixing procedure

The de-mixing algorithm consists of the following steps:

1. Create multiple compressed data sets, phase-shifted and averaged for the different directions of interest, i.e., the FoV plus a number N_{A} of interfering directions (typically the directions of the "A-team" sources) . Time and frequency averaging is done over respectively $N_{\mathrm{t,av}}$ and $N_{\mathrm{f,av}}$ bins.
2. Compute a mixing matrix per baseline, which describes how much of the signal from each of the A-team source directions ends up in each of the averaged data sets.
3. Multiply the averages with the inverse of the mixing matrix to obtain pure averages. This yields the correlations with the bright A-team source contributions subtracted. This procedure does not always work because the mixing matrix can be ill-conditioned. This situation occurs in particular for small domains, short baselines or if the separation between sources is small. There exists a solution for this (see RD18), however it would require too much detail to include the description here.

Given the issues mentioned above, it can be beneficial to incorporate de-mixing in the calibration; this is the approach used in LOFAR [RD17]. Since de-mixing works per baseline but calibration uses all baselines simultaneously, the latter is more robust with respect to baseline-related problems. Here, the idea is to calibrate for all directions simultaneously on the joined set of compressed data sets. The following steps are taken:

1. Predict the visibilities for each direction at the resolution of the averaged data.

2. Multiply with the mixing matrix.

3. Minimise the squared error, where the squared errors are summed over the averaged data sets. (In LOFAR, the Levenberg-Marquard algorithm is used for this purpose.)

5.1.2 De-mixing computational cost

The various steps in the de-mixing algorithm scale as follows. Please note that the compute cost of one operation is assumed to be one FLOP. Let $N_{\rm bl} = \frac{1}{2}N_{\rm rt}(N_{\rm rt}-1)$ be the number of baselines.

1. For each baseline create the $(N_{\rm f,av} \times N_{\rm t,av}, N_{\rm A}+1)$ re-phasing matrix $\mathbf{A}$. Throughout the frequency band there are $N_{\rm f}/N_{\rm f,av}$ of such matrices needed. Creating these matrices requires a phase shift and averaging at the full resolution (each taking 20 FLOP), also for the target field. Thus in total for all baselines

$$N_{\rm FLOP} \sim 20 N_{\rm f} N_{\rm t} (N_{\rm A}+1) N_{\rm bl} N_{\rm pol}.$$

2. Compute, for each baseline, the $(N_{\rm A}+1, N_{\rm A}+1)$ mixing matrix $\mathbf{M}$ from the re-phasing matrix $\mathbf{A}$, which takes 4 FLOP for each direction: $\mathbf{M} = \mathbf{A}^H\mathbf{A}$. This requires in total for all baselines:

$$N_{\rm FLOP} \sim 4\frac{N_{\rm t}}{N_{\rm t,av}}\frac{N_{\rm f}}{N_{\rm f,av}}\left(\frac{1}{2}(N_{\rm A}+1)N_{\rm A}\right)N_{\rm bl}N_{\rm pol}$$

3. Solve the X and Y gains per telescope in the direction of the sources using an $O(N^2)$ solver with N_{it} iterations. Using StEFCal a 2x2 Jones matrix requires per averaged time/channel:

$$N_{\rm FLOP}^{\rm predict} \sim 64 N_{\rm rt}^2 N_{\rm A} + 242 N_{\rm rt} N_{\rm A} + 128 N_{\rm rt}^2$$
$$N_{\rm FLOP}^{\rm solve} \sim 48 N_{\rm rt}^2 N_{\rm A} N_{\rm it}$$

   Thus the total cost is

$$N_{\rm FLOP} \sim \frac{N_{\rm t}}{N_{\rm t,av}}\frac{N_{\rm f}}{N_{\rm f,av}}((64+48N_{\rm it})N_{\rm rt}^2 N_{\rm A} + 242 N_{\rm rt} N_{\rm A} + 128 N_{\rm rt}^2) \quad (5)$$

   Note that the $O(N^3)$ Levenberg-Marquardt solver, as used in original LOFAR implementation, is very expensive.

4. Correct the observed visibilities V by subtracting the sources from the data at full resolution. The cost of subtraction is:

$$N_{\rm FLOP} \sim 2 N_{\rm t} N_{\rm f} N_{\rm A} N_{\rm bl} N_{\rm pol}$$

   However, this has to be done at full resolution, so the sources have to be predicted at full resolution. The costs of the prediction are shown above.

These numbers have to be multiplied with the number of beams and divided by the correlator dump time. Using $N_{\rm pol} = 4$ and $N_{\rm bl} = N_{\rm rt}^2/2$, the FLOP rate comes to:

$$N_{\rm FLOPS} \sim N_{\rm f} N_{\rm bl} N_{\rm pol}((154+121N_{\rm rt}^{-1})N_{\rm A} + 84 + \frac{N_{\rm A}^2 + (33+24N_{\rm it}+121N_{\rm rt}^{-1})N_{\rm A}+64}{N_{\rm t,av}N_{\rm f,av}})N_{\rm beams}T_{\rm s}^{-1} \quad (6)$$

The amount of averaging determines the cost of the de-mixing operation. Due to latency requirements no time averaging can be done in the fast pre-processing pipeline. For the LOFAR standard pipeline typical averaging factors are $N_{\mathrm{t,av}} = 5$ and $N_{\mathrm{f,av}} = 32$.

For SKA1-Mid no demixing needs to be done. For the SKA1-Low buffered pre-processing pipeline, assuming:

- $N_{\mathrm{A}} = 4$, the number of A-teams sources to subtract
- $N_{\mathrm{f}} = 65536$
- $N_{\mathrm{t,av}} = 5, N_{\mathrm{f,av}} = 32$
- $N_{\mathrm{rt}} = 512$
- $N_{\mathrm{beams}} = 1$
- $N_{\mathrm{it}} = 50$ solve iterations
- $T_{\mathrm{s}} = 0.9$ s

This yields for the buffered pre-processing pipeline of SKA1-Low: $N_{\mathrm{FLOPS}} \sim 27.9$ TFLOPS.
For the fast pre-processing pipeline $N_{\mathrm{t,av}} = 1$, yielding $N_{\mathrm{FLOPS}} \sim 32.7$ TFLOPS.
Using $N_{\mathrm{A}} = 10$ in the standard pipeline yields $N_{\mathrm{FLOPS}} \sim 65.0$ TFLOPS.

5.1.3 Memory bandwidth of de-mixing

In the de-mixing process, memory access is dominated by the visibility data, which are accessed twice for each source direction.

$$M_{\mathrm{bw}} = 2(N_{\mathrm{A}} + 1)T_{\mathrm{s}}^{-1} N_{\mathrm{bl}} N_{\mathrm{f}} N_{\mathrm{pol}} N_{\mathrm{bpa}} \,\mathrm{bytes/s} \tag{7}$$

For the SKA1-Low this yields $M_{\mathrm{bw}} \sim 3$ TBytes/s.

5.2 Direct approach

Instead of doing a phase shift and average step to smear out the signals as used in de-mixing, it is possible to directly solve for the X and Y gains per station in the directions of the bright sources. A non-linear $O(N^2)$ solver such as StEFCal or SAGECal can be used to obtain the solutions.

It requires the prediction of the N_A sources and an iterative solve with N_{it} iterations. The cost equations for the predict, solve, and subtract are the same as in the previous section, but these steps have to be done at full resolution. Note that the subtract does not need an extra predict. In total it comes to:

$$N_{\mathrm{FLOPS}} \sim N_{\mathrm{f}} N_{\mathrm{bl}} N_{\mathrm{pol}} ((34 + 24N_{\mathrm{it}} + 121N_{\mathrm{rt}}^{-1})N_{\mathrm{A}} + 64) N_{\mathrm{beams}} T_{\mathrm{s}}^{-1} \tag{8}$$

Using the parameter values from the previous section yields $N_{\mathrm{FLOPS}} = 47.6$ TFLOPS for 4 sources in the buffered pipeline. For 10 sources it yields 118.2 TFLOPS. Because no averaging is done, the costs are the same for the fast pre-processing pipeline.

Note that LOFAR did not use this approach (and moved to de-mixing) because an $O(N^3)$ Levenberg-Marquardt solver was used making it too expensive.

6 Baseline-dependent averaging

The ingest rate can be greatly reduced by applying baseline-dependent averaging (BDA). The criterion for this is that the de-correlation due to integration in time and frequency is a (small) fraction of the L1 2% de-correlation requirement.

The theoretical analysis described in the baseline-dependent averaging PDR supporting document [RD01] shows that baseline-dependent averaging is indeed a feasible approach for the SDP. The details of implementation in the SDP will be worked out after the SDP PDR.

However, in the current SDP design [AD04] BDA will only be used in the imaging pipelines.

7 Risks

- The Receive component needs to receive the meta data in time, otherwise the latency for the fast processing pipelines becomes too high. This means that TM and the Fast Telescope State Producer have to be capable of delivering the meta data in time.
- The compute rate for the demixing step scales with the square of the number of bright sources. If this number becomes high, the compute rate gets very high. The direct approach seems to scale better.