Tim Cornwell Consulting

Interferometry and Imaging

Implementation of the Rau-Cornwell MSMFS algorithm

Document number TCC-SDP-151123-2
Context . . . . (SDP.PIP)
Revision 3
Author . . . . T.J. Cornwell
Release date . . . . Wednesday 27th April, 2016 13:27
Document classification . . . . Unrestricted
Status . . . . Draft

Name	Designation	Affilitation
T.J. Cornwell	Author	TCC
Signature & Date:		

Name	Designation	Affilitation
T.J. Cornwell	Lead Consultant	TCC
Signature & Date:		

Version	Date of issue	Prepared by	Comments
0.3	27/04/2016	T.J. Cornwell	Ready for review by Anna and Steve

ORGANISATION DETAILS

Name	Tim Cornwell Consulting

Table of Contents

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27th April, 2016 13:27   Page 3 of 33

List of abbreviations

CDR   Critical Design Review

FFT   Fast Fourier Transform

FLOP   Floating Point Operations

FLOPS   Floating Point Operations per Second

FoV   Field of View

MS-MFS Multi-Scale Multi-Frequency Synthesis

PSF   Point Spread Function

SDP   Scientific Data Processor

SKA   Square Kilometre Array

SKAO   SKA Organisation

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27th April, 2016 13:27   Page 4 of 33

List of Figures

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27th April, 2016 13:27   Page 5 of 33

List of Tables

Summary

Parallel versions of both variants of MSMFS exist. These work by distributing in time (CASA) and frequency (ASKAPsoft). Parallelisation of the Minor Cycle is not straightforward and consequently both CASA and ASKAPsoft currently use single-threaded inner loops. This is unlikely to be successful for SKA, and instead the Minor Cycle has to be parallelised inside one CI and perhaps additionally distributed over multiple CI's.

In light of this memo, the following recommendations are made concerning the need for MFS:

1. Perform simulations to explore the need for MSMFS in SKA. Currently an L1 requirement directs SDP to implement MSMFS. However there is no justification given for this very computationally demanding processing. Indeed the promised excellent UV coverage of the SKA telescopes makes it at least questionable, although arguing from the work of Urvashi [RD??] does indicate that with MFS, the dynamic range would be limited to 10^4 for wide frequency span. Either CASA or ASKAPsoft could be used for the reconstructions.

2. Adding to the desire to see some more justification of the requirement, we note that the L1 requirements do not address the number of scales required, the number of Taylor terms, or the possibility of sub-banding (all of which appear in the SDP L2 requirements. We note three possible outcomes of the simulations: MFS can be ignored, MFS is required and MSMFS is sufficient, or MFS is needed and MFMFS is not sufficient. In the case of the latter outcome, investigation of other deconvolution techniques (e.g. [RD14]) would be required.

3. Assess the importance of the considerations in Section 8.

The following recommendations address the resource usage:

4. Implement both Image (CASA) and Visibility based (ASKAPsoft) calculation of the Taylor term images in the SDP Performance Model. The main reasons to choose one over the other are to do with the relative speed of gridding and FFTs. When it comes to implementation during construction, the costs of implementing both approaches should be minimal, and unimportant compared to the risk of implementing the wrong choice.

5. Update the SDP Performance Model to reflect the MSMFS memory usage:

$$M_{MSMFS} = (N_T(N_s + 1) + N_s)N_{pix}^2 \qquad (1)$$

6. Update the SDP Performance Model to reflect the MSMFS Flops:

$$Rflop_{MSMFSminor} = N_{minor}N_{pix}^2 N_s N_T \qquad (2)$$

7. Prototype a multi-threaded Minor Cycle on one Compute Island. GPUMultiscaleCLEAN might be a good starting point. Do MS and then MSMFS. This can be done without implementing a major cycle.

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27th April, 2016 13:27   Page 7 of 33

8. Model and then, if promising, prototype a distributed Minor Cycle on multiple CI's where synchronisation occurs every iteration. Do MS and then MSMFS. This can be done without implementing a major cycle.

The following recommendations address wider issues:

9. The SKAO Science Team should be consulted about the impact of seams in the images or spectral cubes. These may occur both spatially (if the facets are deconvolved separately, or in frequency, because of the disjoint use of MSMFS in the different sub-bands. It seems questionable that these can be tolerated.

10. We should note that MSMFS was developed in a serial platform where computation was expensive and memory access cheap. In the SDP context, the processing is distributed and parallel, and computation is cheap and memory access expensive. Although SDP has no remit to develop new algorithms, it could encourage researchers to address this new context.

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]	Dewdney, P. E. (2013). SKA1 System Baseline Design. SKA Office
[AD02]	McCool, R., Cornwell, T. (2013). Miscellaneous Corrections to the Baseline Design
[AD03]	SKA Phase 1 System (Level 1) Requirements Specification, T.Cornwell, SKA-OFF.SE.ARC-SKO-SRS-001_2
[AD04]	PDR.01 SDP.ARCH document
[AD05]	SDP Costing spreadsheet
[AD06]	Cornwell, T.J. (2015). SKA1 Telescope Calibration Framework. SKA Office (draft version)
[AD07]	CSP–SDP ICD

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27th April, 2016 13:27   Page 9 of 33

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-0000027, R. Nijboer: Pipelines Element Subsystem Design
[RD02]	SKA-TEL-SDP-0000038, R. Bolton: High Priority Science Objectives: Performance Analysis
[RD03]	SKA-TEL-SDP-0000029, S. Salvini: Pipelines: Calibration
[RD04]	SKA-TEL-SDP-0000030, A. Scaife: Imaging Pipeline
[RD05]	`http://www.astron.nl/casacore/trunk/casacore/doc/notes/229.html`
[RD06]	Imager.cc C++ source, CASA source code SVN revision 30821, `https://svn.cv.nrao.edu/svn/casa/trunk`
[RD07]	Parameterized deconvolution for wide-band radio synthesis imaging, Urvashi Rao Venkata, 2010, PhD thesis
[RD08]	S. Bhatnagar, T. J. (2008). Correcting direction-dependent gains in the deconvolution of radio interferometric images. A&A, 419-429
[RD09]	Rik Jongerius, S. W. (2014). An End-to-End Computing Model for the Square Kilometre Array. IEEE Computer, volume 47, number 9
[RD10]	SDP Element Concept. Paul Alexander, Chris Broekema, Simon Ratcliffe, Rosie Bolton, Bojan Nikolic, 2013, SDP-PROP-DR-001-1 (part of SKA SDP Consortium proposal)
[RD11]	Cornwell, T. J. (2008). Multi-Scale CLEAN deconvolution of radio synthesis images. IEEE Journal of Selected Topics in Sig. Proc., 2, 793801.
[RD12]	Conway, J. E., Cornwell, T. J., and Wilkinson, P. N. (1990). Multi-Frequency Synthesis - a New Technique in Radio Interferometric Imaging. Monthly Notices of the Royal Astronomical Society, 246, 490.
[RD13]	http://www.aanda.org/component/citedby/?task=crossref&doi=10.1051/0004-6361/201117104
[RD14]	Junklewitz, H., Bell, M. R., and Ensslin, T. (2015). A new approach to multi-frequency synthesis in radio interferometry. Astronomy & Astrophysics, 581, A59. http://doi.org/10.1051/0004-6361/201423465
[RD15]	Cornwell, T.J. (2016) Extension of the SDP performance model for cross data-island connectivity, TCC-SDP-151123-1-1
[RD16]	Rau, U., & Cornwell, T. J. (2011, June 14). A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry. arXiv.org. EDP Sciences. http://doi.org/10.1051/0004-6361/201117104
[RD17]	Bhatnagar, S., Rau, U., & Golap, K. (2013). Wide-field wide-band Interferometric Imaging: The WB A-Projection and Hybrid Algorithms. The Astrophysical Journal, 770(2), 91. http://doi.org/10.1088/0004-637X/770/2/91

1   Purpose of the document

The purpose of this document is to describe the Multi-Scale MultiFrequency Synthesis algorithm, the variant implementations in CASA and ASKAPsoft, and the scaling for insertion in the SDP Performance Model. We conclude with some recommendations.

2   Scope of the document

The scope includes only the published Multi-Scale, Multi-Frequency algorithm, existing implementations, and inclusion in the SDP processing pipelines. No other MFS algorithms are explored.

3   Background

The imaging performance of a radio synthesis telescope is related to the uv coverage. The uv coverage is dependent on the observing wavelength so diversity in uv can be obtained by observing over a range of frequencies [RD12]. This Multi-Frequency Synthesis (MFS) only works for radio sources in which broadband emission is present, giving a way to relate measurements at different frequencies. There have been multiple attempts at an MFS algorithm (e.g. [RD12]). MFS is usually most important for sources having extended emission. Hence multi-scale algorithms are favoured.

The Multi-Scale, Multi-Frequency Synthesis (MS-MFS) algorithm was developed by Urvashi [RD07], melding together the concepts from multi-scale clean [RD11] and multi-frequency synthesis e.g. [RD12], with the goal of improving the reconstruction of sky brightness from radio interferometric data. The explicit model used for the sky brightness is a collection of blobs of varying scale sizes and strengths, each with spectral behaviour described by a power law expanded into a Taylor series. It has been used in JVLA, ALMA, and ATCA observations [RD13].

The algorithm is an elaboration of the Multiscale CLEAN [RD11]. It consists of two parts:

Major cycle The image residuals for the current model are calculated by Fourier transform of the images constituting the power law expansion and degridding to obtain the predicted visibility, subtraction from the observed visibility, then gridding, and Fourier transformation to the image plane. There are two ways of performing this cycle depending on whether the Taylor series is calculated in image or uv space. Typically 5 - 10 major cycles are required.

Minor cycle The minor cycle is a greedy algorithm which identifies the dominant candidate scale and removes that using the appropriate PSF centered on that component. The minor cycle repeats this until convergence. Typically hundreds or thousands of iterations are required.

A canonical major/minor cycle algorithm is shown in Figure 1. The two functions Predict and MakeImage are sub-pipelines. ASKAPSoft and CASA both follow this form but with different ways of performing Predict (Figures 2 and 3) and MakeImage (Figures 4 and 5)
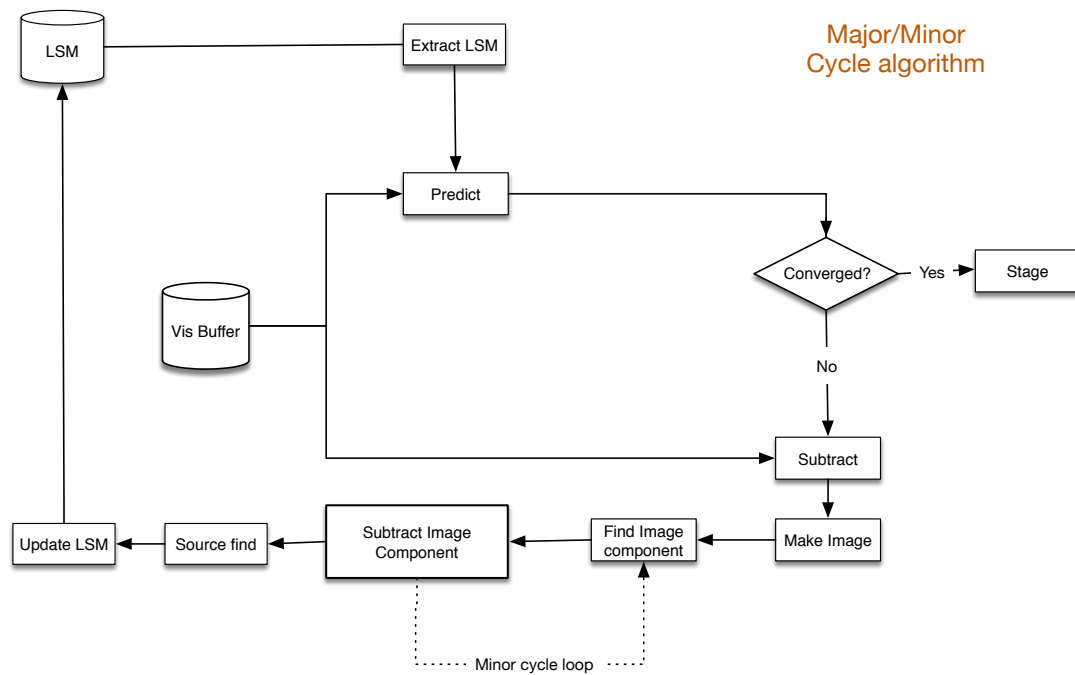
Figure 1: Structure of Major/Minor cycle algorithm

The major cycle returns to the visibility data each iteration, whereas the minor cycle is entirely image-based.

MSMFS is relatively high in complexity because the usual CLEAN process is coupled over both scale (MSClean) and frequency or Taylor terms (MFS). The minor cycle must conduct a search in scale and Taylor term for each peak search. The major cycle requires calculation of the residual Taylor terms, which implies gridding and degridding all relevant frequency information. Memory use is high since a large number of images must be kept and updated as the iteration proceeds.

Document no.: TCC-SDP-151123-2
Revision: 3
Release date: Wednesday 27th April, 2016 13:27

Unrestricted
Author: T.J. Cornwell
Page 12 of 33

4 Is MFS needed?

There is an L1 requirement [AD03], $SKA1 - SYS\_REQ - 2324$ specifying that MSMFS must be used for continuum imaging (see Table1).

ID	Requirement	Status	Parent Requirement	Verification
$SKA1 - SYS\_REQ - 2324$	Multi-frequency synthesis imaging. All imaging shall construct and make use of frequency dependent image models over the entire observed bandwidth.	Accepted	Reference	Demonstration

Table 1: Relevant L1 requirement

This is reflected in the L2 requirements shown in Table 2.

The last three L2 requirements, $SDP\_REQ - 676$, $SDP\_REQ - 677$, $SDP\_REQ - 678$, are an attempt to give some constraints to $SKA1 - SYS\_REQ - 2324$. However, they beg the same question as the original L1 requirement: is MSMFS needed to meet the scientific goals? If so, how many scales, Taylor terms, and sub-bands are required? In addition, there is also a potentially missing requirement concerning the connection between the models for different sub-bands - is there allowed to be a seam in the continuum-subtracted spectral line cube at the transitions between sub-bands?

Finally we could ask is there a superior algorithm already published? (see e.g. [RD14]). The degree of tuning required for MSMFS (number of scales and number of Taylor terms) is acceptable in a hands-on data processing scenario but may be not acceptable for a general purpose pipeline.

Some caution is advised in advocating that any form of MFS is not needed. Clearly ALMA and JVLA, which also have relatively good uv coverage, have benefited from the use of MSMFS. Furthermore, although the uv-coverage of SKA1-LOW and SKA1-MID is excellent after a long observation, there will inevitably be pressure to observe with shorter observation times. We should remember than the VLA was supposed to take 8 hours to image every field. With the advent of CLEAN and self-calibration, it was known even in the early eighties that 400 sources could be imaged in a day.

Urvashi has discussed the use of MSMFS for the JVLA (section 8.4.2.1 of [RD??]. With regard to just ignoring the effect of spectral index variations, she says:

> Dynamic-range limits when source spectra are ignored : If continuum imaging is done with only MFS gridding and source spectra are ignored, spectral structure will masquerade as spurious spatial structure. These errors will affect regions of the image both on-source and off-source and their magnitudes depend on the available

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27th April, 2016 13:27   Page 13 of 33

Function		Satisfies requirement		
Number	Name	Number	Name	Description
F1.9	Calibrate and Image	$SDP\_REQ - 390$	Multi-frequency synthesis imaging	The SDP shall construct and make use of frequency-dependent image models over the entire observed bandwidth.
F1.9	Calibrate and Image	$SDP\_REQ - 676$	Continuum imaging deconvolution scales	While performing continuum imaging the SDP shall deconvolve using a maximum number of 10 (TBC) scales.
F1.9	Calibrate and Image	$SDP\_REQ - 677$	Continuum imaging deconvolution Taylor terms	While performing continuum imaging the SDP shall deconvolve using a maximum number of 5 (TBC) Taylor terms.
F1.9	Calibrate and Image	$SDP\_REQ - 678$	Continuum imaging deconvolution by sub-band	While performing continuum imaging the SDP shall deconvolve by sub-band (TBC).

Table 2: Relevant L2 requirements

uv-coverage, the frequency range being covered, the choice of reference frequency, and the intensity and spectral index of the source. A rough rule of thumb for an EVLA-type uv-coverages (see section 6.2.4.2) is that for a point source of with spectral index = 1.0 measured between 1 and 2 GHz, the peak error obtained if the spectrum is ignored is at a dynamic range of $< 10^3$. Note that when all sources in the observed region of the sky have similar spectral indices, these errors can be reduced by dividing out an average spectral index (one single number over the entire sky) from the visibilities before imaging the

If we suppose that the SKA PSF is ten times better than that of the JVLA (an over estimate since the RMS sidelobe level goes only as the inverse of the number of antennas), the dynamic range limitation for SKA1-LOW and SKA1-MID still appears at about 10^4, which is clearly insufficient. Furthermore, the primary beam will vary across the frequency span, and this will lead to spurious structure if spectral effects are ignored (cf Urvashi). The introduction of sub-bands does ameliorate these problems to some degree but then the continuum structure model is only defined in pieces that are not necessarily consistent at the edges.

There is no mechanism to address these questions beyond simulations. Arguing from uv coverage metrics such as the UVGAP only helps in a relative sense.

5 Overview of MSMFS

5.1 Mathematical description

In the Multi-Scale CLEAN [RD11], the sky is modelled as a set of blobs. Performing this decomposition using a gradient search algorithm is very demanding of resources. Instead, in MSCLEAN, a greedy algorithm obtained by generalising the CLEAN algorithm is used. The greedy algorithm locates the amplitude, scale and location of the blob by a straightforward search in the set of residual images, each smoothed with the appropriate blob. For each minor cycle this produces an amplitude, location, and scale. The smoothed residual images and model images are updated, and iteration proceeds.

$$\vec{I}^{model} = \sum_{s=0}^{N_s-1} \vec{I}_s^{shp} \star \vec{I}_s^{sky,\delta} \tag{3}$$

where N_s is the number of spatial scales used to construct the image, and $\vec{I}_s^{sky,\delta}$ represents a collection of δ-functions that describe the locations and integrated amplitudes of flux components of scale s in the image. $\vec{I}_s^{shp}$ is a tapered truncated parabola of width proportional to s. The symbol $\star$ denotes convolution.

The MSMFS [RD07] algorithm decomposes the image into a set of N_s blobs each with N_t Taylor terms. The following description is drawn from [RD07]

$$\vec{I}_\nu^{model} = \sum_{t=0}^{N_t-1} w_\nu^t \vec{I}_t^{sky} \quad \text{where} \quad w_\nu^t = \left(\frac{\nu - \nu_0}{\nu_0} \right)^t \tag{4}$$

where N_t is the order of the Taylor series expansion, and the I_t^m represent multi-scale Taylor coefficient images.

The image flux model at each frequency can be written as a linear sum of coefficient images at different spatial scales.

$$\vec{I}_\nu^{model} = \sum_{t=0}^{N_t} \sum_{s=0}^{N_s} w_\nu^t \left[\vec{I}_s^{shp} \star \vec{I}_{\frac{s}{t}}^{sky} \right] \quad \text{where} \quad w_\nu^t = \left(\frac{\nu - \nu_0}{\nu_0} \right)^t \tag{5}$$

Here, N_s is the number of discrete spatial scales used to represent the image and N_t is the order of the series expansion of the spectrum. $\vec{I}_{\frac{s}{t}}^{sky}$ represents a collection of δ-functions that describe the locations and integrated amplitudes of flux components of scale s in the image of the tth series coefficient.

The visibility function corresponding to 5 is:

$$\vec{V}_\nu^{obs} = [S_\nu][F_s]\vec{I}_\nu^m = \sum_{t=0}^{N_t} \sum_{s=0}^{N_s} w_\nu^t [S_\nu][T_s][F_s]\vec{I}_t^{sky} \tag{6}$$

$\vec{V}_\nu^{obs}$ is a vector of $n \times 1$ visibilities measured at frequency ν. w_ν are Taylor-weights (shown in Eqn.**??**). $[S_\nu]$ is an $n \times m$ projection operator that represents the spatial-frequency sampling

Document no.: TCC-SDP-151123-2

Revision: 3

Release date: Wednesday 27th April, 2016 13:27

Unrestricted

Author: T.J. Cornwell

Page 16 of 33

function for frequency v. The image-domain convolution of model $\vec{I}_s^{\text{sky}}$ with $\vec{I}_s^{\text{shp}}$ is written as a spatial-frequency-domain multiplication. $\vec{I}_s^{\text{shp}} \star \vec{I}_s^{\text{sky}} = [\underaccent{\sim}{F}^{\dagger}][\underaccent{\sim}{T}_t][\underaccent{\sim}{F}]\vec{I}_s^{\text{sky}}$ where $[\underaccent{\sim}{T}_t]_{m \times m} = diag([\underaccent{\sim}{F}]\vec{I}_s^{\text{shp}})$ is a spatial-frequency taper function. All images are vectors of shape $m \times 1$.

The normal equations for the system described in Eqn. **??** can be written in block matrix form, with each block-row (for scale size s, and Taylor term t) given by

$$\sum_{p=0}^{N_s-1} \sum_{q=0}^{N_t-1} \left[\underaccent{\sim}{H}_{\substack{s,p \\ t,q}} \right] \vec{I}_{\substack{p \\ q}}^{\text{sky}} = \vec{I}_{\substack{s \\ t}}^{\text{dirty}} \quad \forall \; s \in [0, N_s - 1], t \in [0, N_t - 1] \tag{7}$$

Here, each $\left[\underaccent{\sim}{H}_{\substack{s,p \\ t,q}} \right]$ is an $m \times m$ block of the Hessian matrix, and $\vec{I}_{\substack{s \\ t}}^{\text{dirty}}$ is one of $N_s N_t$ dirty images.

$$\left[\underaccent{\sim}{H}_{\substack{s,p \\ t,q}} \right] = \left[\underaccent{\sim}{A}_{\substack{s \\ t}}^{\dagger} \right] [\underaccent{\sim}{W}^{\text{im}}] \left[\underaccent{\sim}{A}_{\substack{p \\ q}} \right] \tag{8}$$

$$= [\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_s \underaccent{\sim}{S}^{\dagger} \underaccent{\sim}{W}_t^{\text{mfs}\,\dagger}][\underaccent{\sim}{W}^{\text{im}}][\underaccent{\sim}{W}_q^{\text{mfs}} \underaccent{\sim}{S} \underaccent{\sim}{T}_p \underaccent{\sim}{F}] \tag{9}$$

$$= [\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_s \underaccent{\sim}{F}][\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{S}^{\dagger} \underaccent{\sim}{W}_t^{\text{mfs}\,\dagger} \underaccent{\sim}{W}^{\text{im}} \underaccent{\sim}{W}_q^{\text{mfs}} \underaccent{\sim}{S} \underaccent{\sim}{F}][\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_p \underaccent{\sim}{F}] \tag{10}$$

$$= [\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_s \underaccent{\sim}{F}] \left\{ \sum_{v} w_v^{t+q}[\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{S}^{\dagger} \underaccent{\sim}{W}_v^{\text{im}} \underaccent{\sim}{S} \underaccent{\sim}{F}] \right\} [\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_p \underaccent{\sim}{F}] \tag{11}$$

$$= [\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_s \underaccent{\sim}{F}] \left\{ \sum_{v} w_v^{t+q}[\underaccent{\sim}{H}_v] \right\} [\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_p \underaccent{\sim}{F}] \tag{12}$$

$[\underaccent{\sim}{W}^{\text{im}}]$ is a diagonal matrix of data-weights (and imaging-weights) and $[\underaccent{\sim}{W}_t^{\text{mfs}}]$ is a diagonal matrix containing Taylor-weights w_v^t. $[\underaccent{\sim}{H}_v] = [\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{S}^{\dagger} \underaccent{\sim}{W}_v^{\text{im}} \underaccent{\sim}{S} \underaccent{\sim}{F}]$ is the Hessian matrix formed using only one frequency channel, and is a convolution operator containing a shifted version of the single-frequency point-spread-function $\vec{I}_v^{\text{psf}} = diag[\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{S}^{\dagger} \underaccent{\sim}{W}_v^{\text{im}} \underaccent{\sim}{S}]$ in each row (see appendix **??** for details). $[\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_s \underaccent{\sim}{F}]$ and $[\underaccent{\sim}{F}^{\dagger} \underaccent{\sim}{T}_p \underaccent{\sim}{F}]$ are also convolution operators with $\vec{I}_s^{\text{shp}}$ and $\vec{I}_p^{\text{shp}}$ as their kernels. The process of convolution is associative and commutative, and therefore, $\left[\underaccent{\sim}{H}_{\substack{s,p \\ t,q}} \right]$ is also a convolution operator whose kernel is given by

$$\vec{I}_{\substack{s,p \\ t,q}}^{\text{psf}} = \vec{I}_s^{\text{shp}} \star \left\{ \sum_{v} w_v^{t+q} \vec{I}_v^{\text{psf}} \right\} \star \vec{I}_p^{\text{shp}} \tag{13}$$

Document no.: TCC-SDP-151123-2
Revision: 3
Release date: Wednesday 27$^{\text{th}}$ April, 2016 13:27

Unrestricted
Author: T.J. Cornwell
Page 17 of 33

The dirty images on the RHS of Eqn.7 can be written as follows.

$$\vec{I}_s^{\text{dirty}} \;\; = \;\; \left[\underset{t}{\underset{\sim}{A}}{}^{\dagger}_s\right][\underset{\sim}{W}^{\text{im}}]\vec{V}^{\text{obs}} \tag{14}$$

$$= \;\; [\underset{\sim}{F}{}^{\dagger}\underset{s}{\underset{\sim}{T}}\,\underset{s}{\underset{\sim}{S}}{}^{\dagger}\underset{t}{\underset{\sim}{W}}{}^{\text{mfs}\,\dagger}\underset{\sim}{W}^{\text{im}}]\vec{V}^{\text{obs}} \tag{15}$$

$$= \;\; [\underset{\sim}{F}{}^{\dagger}\underset{s}{\underset{\sim}{T}}\,\underset{s}{\underset{\sim}{F}}][\underset{\sim}{F}{}^{\dagger}\underset{\sim}{S}{}^{\dagger}\underset{t}{\underset{\sim}{W}}{}^{\text{mfs}\,\dagger}\underset{\sim}{W}^{\text{im}}]\vec{V}^{\text{obs}} \tag{16}$$

$$= \;\; [\underset{\sim}{F}{}^{\dagger}\underset{s}{\underset{\sim}{T}}\,\underset{s}{\underset{\sim}{F}}]\left\{\sum_{\nu} w_{\nu}^{t}[\underset{\sim}{F}{}^{\dagger}\underset{\underset{\nu}{\underset{\sim}{W}^{\text{im}}}}{\underset{\sim}{S}}{}^{\dagger}]\vec{V}_{\nu}^{\text{obs}}\right\} \tag{17}$$

$$= \;\; \vec{I}_s^{\text{shp}} \star \left\{\sum_{\nu} w_{\nu}^{t}\vec{I}_{\nu}^{\text{dirty}}\right\} \tag{18}$$

where $\vec{I}_{\nu}^{\text{dirty}} = [\underset{\sim}{F}{}^{\dagger}\underset{\underset{\sim}{W}^{\text{im}}}{\underset{\sim}{S}}{}^{\dagger}]\vec{V}_{\nu}^{\text{obs}}$ is the dirty image formed by direct Fourier inversion of weighted visibilities from one frequency channel.

When all scales and Taylor terms are combined, the full Hessian matrix contains $N_tN_s \times N_tN_s$ blocks each of size $m \times m$, and N_t Taylor coefficient images each of size $m \times 1$, for all N_s spatial scales.

The normal equations in block matrix form for the example in Eqn.**??** for $N_t = 3, N_s = 2$ is shown in Eqn.19. The Hessian matrix consists of $N_s \times N_s = 2 \times 2$ blocks (the four quandrants of the matrix), each for one pair of spatial scale s, p (the upper indices). Within each quadrant, the $N_t \times N_t = 3 \times 3$ matrices correspond to various pairs of t, q (Taylor coefficient indices; the lower indices). This layout shows how the multi-scale and multi-frequency aspects of this imaging problem are combined and illustrates the dependencies between the spatial and spectral basis functions.

$$\begin{bmatrix} \left[\underset{\sim}{H}{}^{0,0}_{0,0}\right] & \left[\underset{\sim}{H}{}^{0,0}_{0,1}\right] & \left[\underset{\sim}{H}{}^{0,0}_{0,2}\right] & \left[\underset{\sim}{H}{}^{0,1}_{0,0}\right] & \left[\underset{\sim}{H}{}^{0,1}_{0,1}\right] & \left[\underset{\sim}{H}{}^{0,1}_{0,2}\right] \\[6pt] \left[\underset{\sim}{H}{}^{0,0}_{1,0}\right] & \left[\underset{\sim}{H}{}^{0,0}_{1,1}\right] & \left[\underset{\sim}{H}{}^{0,0}_{1,2}\right] & \left[\underset{\sim}{H}{}^{0,1}_{1,0}\right] & \left[\underset{\sim}{H}{}^{0,1}_{1,1}\right] & \left[\underset{\sim}{H}{}^{0,1}_{1,2}\right] \\[6pt] \left[\underset{\sim}{H}{}^{0,0}_{2,0}\right] & \left[\underset{\sim}{H}{}^{0,0}_{2,1}\right] & \left[\underset{\sim}{H}{}^{0,0}_{2,2}\right] & \left[\underset{\sim}{H}{}^{0,1}_{2,0}\right] & \left[\underset{\sim}{H}{}^{0,1}_{2,1}\right] & \left[\underset{\sim}{H}{}^{0,1}_{2,2}\right] \\[6pt] \left[\underset{\sim}{H}{}^{1,0}_{0,0}\right] & \left[\underset{\sim}{H}{}^{1,0}_{0,1}\right] & \left[\underset{\sim}{H}{}^{1,0}_{0,2}\right] & \left[\underset{\sim}{H}{}^{1,1}_{0,0}\right] & \left[\underset{\sim}{H}{}^{1,1}_{0,1}\right] & \left[\underset{\sim}{H}{}^{1,1}_{0,2}\right] \\[6pt] \left[\underset{\sim}{H}{}^{1,0}_{1,0}\right] & \left[\underset{\sim}{H}{}^{1,0}_{1,1}\right] & \left[\underset{\sim}{H}{}^{1,0}_{1,2}\right] & \left[\underset{\sim}{H}{}^{1,1}_{1,0}\right] & \left[\underset{\sim}{H}{}^{1,1}_{1,1}\right] & \left[\underset{\sim}{H}{}^{1,1}_{1,2}\right] \\[6pt] \left[\underset{\sim}{H}{}^{1,0}_{2,0}\right] & \left[\underset{\sim}{H}{}^{1,0}_{2,1}\right] & \left[\underset{\sim}{H}{}^{1,0}_{2,2}\right] & \left[\underset{\sim}{H}{}^{1,1}_{2,0}\right] & \left[\underset{\sim}{H}{}^{1,1}_{2,1}\right] & \left[\underset{\sim}{H}{}^{1,1}_{2,2}\right] \end{bmatrix} \begin{bmatrix} \vec{I}^{\text{sky}\,0}_0 \\[6pt] \vec{I}^{\text{sky}\,0}_1 \\[6pt] \vec{I}^{\text{sky}\,0}_2 \\[6pt] \vec{I}^{\text{sky}\,1}_0 \\[6pt] \vec{I}^{\text{sky}\,1}_1 \\[6pt] \vec{I}^{\text{sky}\,1}_2 \end{bmatrix} = \begin{bmatrix} \vec{I}^{\text{dirty}\,0}_0 \\[6pt] \vec{I}^{\text{dirty}\,0}_1 \\[6pt] \vec{I}^{\text{dirty}\,0}_2 \\[6pt] \vec{I}^{\text{dirty}\,1}_0 \\[6pt] \vec{I}^{\text{dirty}\,1}_1 \\[6pt] \vec{I}^{\text{dirty}\,1}_2 \end{bmatrix} \tag{19}$$

This is the system of equations to be solved. The spatial-frequency sampling of a real interferometer is always incomplete ($[\underset{\sim}{S}]$ is rank-deficient). Therefore, each Hessian block, and the

entire Hessian matrix is singular, and an exact inverse does not exist[1]. An accurate reconstruction can be obtained only via successive approximation (iterative numerical optimization).

However, the Hessian is still useful since some of the diagonal terms can used to decouple the pixels in Taylor term space as well as possible. In practice, the Hessian (for each scale) for the peak of the PSF in Taylor term space is of modest size and can be inverted and used. The greedy algorithm then locates the scale and location in the reference-frequency image (as convolved with the set of scales). For each minor cycle this produces a location, scale, and values for all Taylor terms. The residual images and model images are updated, and iteration proceeds. The actual algorithm is shown in Algorithm 1 (CASA form) and Algorithm 2 (ASKAP form).

In summary, MSMFS decouples Taylor terms using the Hessian in Taylor-space evaluated at the peak of the PSF, and uses a greedy CLEAN-like algorithm to decouple in scale and sky position. The scales can be decoupled by an orthogonalisation process prior to deconvolution but the resulting images tend to have non-physical regions of negative brightness.

---

[1]Even if $[\underline{H}]$ were invertible, it is impractical to evaluate and invert the full Hessian (each row of each Hessian block represents an image).

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27$^{\text{th}}$ April, 2016 13:27   Page 19 of 33

6 CASA and ASKAPsoft implementations

The MS-MFS algorithm as implemented in CASA shown in Algorithm 1. The ASKAPsoft version is shown in Algorithm 2 with the variant lines in red.

We can relate these back to the graphical form Figure 1 by introducing variant sub pipelines for Predict and MakeImage. These are graphically shown in Figures 2 to 5:
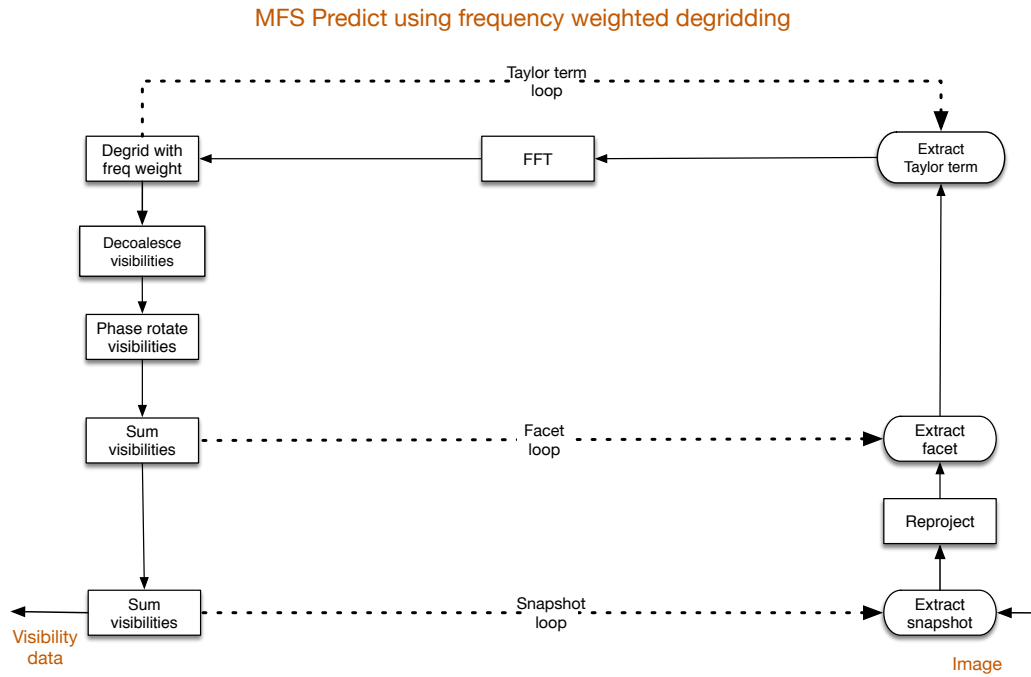

Figure 2: Structure of UV-weighting Predict, as used in ASKAPsoft.

Document no.: TCC-SDP-151123-2
Revision: 3
Release date: Wednesday 27th April, 2016 13:27

Unrestricted
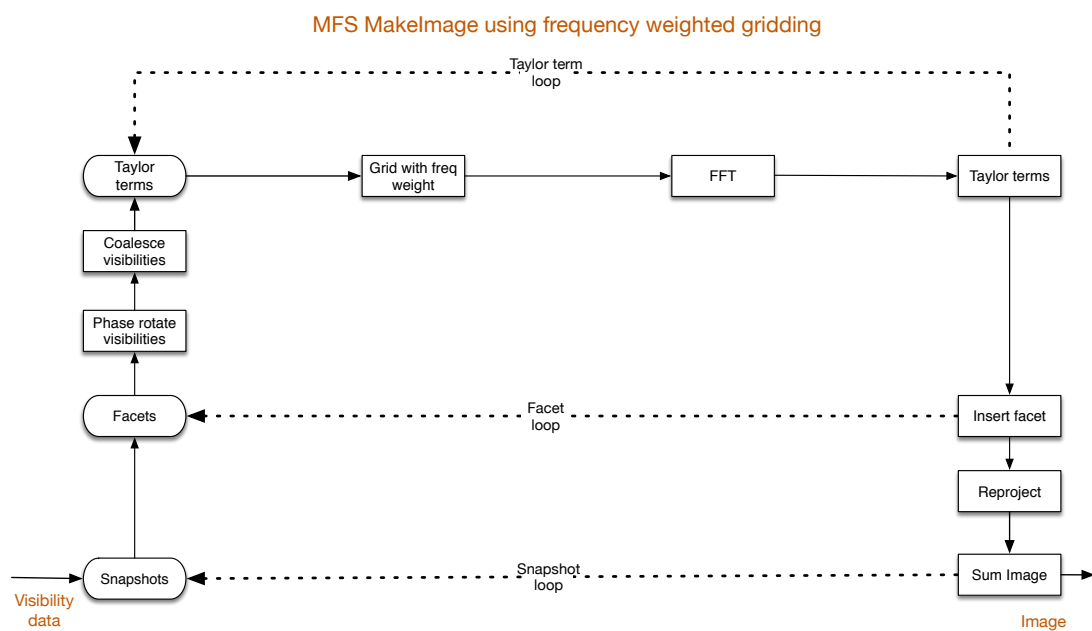Author: T.J. Cornwell
Page 20 of 33

Figure 3: Structure of UV-weighting MakeImage, as used in ASKAPsoft. Note that the three loops (denoted by the rectangle with curved ends) can be permutated at will.

Document no.: TCC-SDP-151123-2

Revision: 3

Release date: Wednesday 27th April, 2016 13:27

Unrestricted

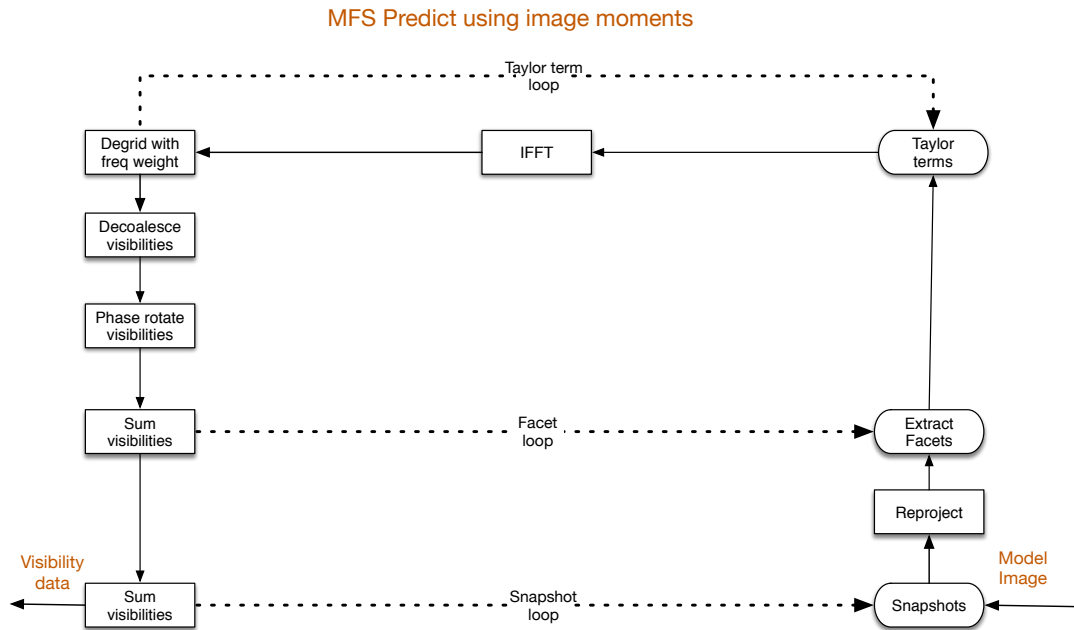Author: T.J. Cornwell

Page 21 of 33

Figure 4: Structure of Image plane-weighting Predict, as used in CASA. Note that the three loops (denoted by the rectangle with curved ends) can be permutated at will.

The CASA and ASKAPsoft approaches should be equivalent apart from numerical differences such as due to the convolution function. Hence either can be used according to the most favourable scaling. Compared to CASA, ASKAPsoft does approximately two orders of magnitude less FFTs and 5 times as much gridding. Both forms should therefore be available in the SDP Performance Model and used as the context dictates.

The Minor Cycle is identical in CASA and ASKAPsoft. The Minor cycle is depicted graphically in Figure 6.

Algorithm 1: MS-MFS, as implemented in CASA

Data: calibrated visibilities : $\vec{V}_{\nu}^{\text{obs}}$ $\forall \nu$

Data: uv-sampling function and weights : $[\underline{S}_{\nu}], [\underline{W}_{\nu}^{\text{im}}]$

Data: input : number of Taylor-terms N_{t}, number of scales N_{s}

Data: input : image noise threshold, σ_{thr}, loop gain g

Data: input : scale basis functions : $\vec{I}_{s}^{\text{shp}}$ $\forall s \in \{0, N_{\text{s}} - 1\}$

Data: input : reference frequency ν_0 to compute $w_{\nu} = \left(\frac{\nu - \nu_0}{\nu_0} \right)$

Result: model coefficient images : $\vec{I}_{t}^{\text{m}}$ $\forall t \in \{0, N_{\text{t}} - 1\}$

Result: intensity, spectral index and curvature : $\vec{I}_{\nu_0}^{\text{m}}, \vec{I}_{\alpha}^{\text{m}}, \vec{I}_{\beta}^{\text{m}}$

1 **for** $t \in \{0, N_{\text{t}} - 1\}, q \in \{t, N_{\text{t}} - 1\}$ **do**

2   Compute the spectral Hessian kernel $\vec{I}_{tq}^{\text{psf}} = \sum_{\nu} w_{\nu}^{t+q} \vec{I}_{\nu}^{\text{psf}}$

3   **for** $s \in \{0, N_{\text{s}} - 1\}, p \in \{s, N_{\text{s}} - 1\}$ **do**

4   Compute scale-spectral kernels $\vec{I}_{sp \atop tq}^{\text{psf}} = \vec{I}_{s}^{\text{shp}} \star \vec{I}_{p}^{\text{shp}} \star \vec{I}_{tq}^{\text{psf}}$

5   **end**

6 **end**

7 **for** $s \in \{0, N_{\text{s}} - 1\}$ **do**

8   Construct scale-dependent Taylor space Hessian $[\underline{H}_{s}^{\text{peak}}]$ from $mid(\vec{I}_{s,s \atop t,q}^{\text{psf}})$ and compute $[\underline{H}_{s}^{\text{peak}^{-1}}]$

9 **end**

10 Initialize the model $\vec{I}_{t}^{\text{m}}$ for all $t \in \{0, N_{\text{t}} - 1\}$

11 **repeat**   /* Major Cycle */

12   **for** $t \in \{0, N_{\text{t}}\text{-}1\}$ **do**

13   Compute $\vec{I}_{t}^{\text{res}} = \sum_{\nu} w_{\nu}^{t} \vec{I}_{\nu}^{\text{res}}$ from frequency-weighted residual images $\vec{I}_{\nu}^{\text{res}}$

14   **for** $s \in \{0, N_{\text{s}}\text{-}1\}$ **do**

15   Convolve with s^{th} scale-function $\vec{I}_{s \atop t}^{\text{res}} = \vec{I}_{s}^{\text{shp}} \star \vec{I}_{t}^{\text{res}}$

16   **end**

17   **end**

18   Calculate minor-cycle threshold f_{limit} from $\vec{I}_{0 \atop 0}^{\text{res}}$

19   **repeat**   /* Minor Cycle */

20   **for** $s \in \{0, N_{\text{s}}\text{-}1\}$ **do**

21   **foreach** *pixel* **do**

22   Construct $\vec{I}_{s}^{\text{rhs}}$, an $N_{\text{t}} \times 1$ vector from $\vec{I}_{s \atop t}^{\text{res}}$ $\forall t \in \{0, N_{\text{t}}\text{-}1\}$

23   Compute principal solution $\vec{I}_{s}^{\text{sol}} = [\underline{H}_{s}^{\text{peak}^{-1}}]\vec{I}_{s}^{\text{rhs}}$

24   Fill solution $\vec{I}_{s}^{\text{sol}}$ into model images $\forall t : \vec{I}_{s \atop t}^{\text{m,sol}}$

25   **end**

26   **end**

27   Choose $\vec{I}_{p \atop t}^{\text{m}} = max\{\vec{I}_{s \atop t=0}^{\text{m,sol}}, \forall s \in \{0, N_{\text{s}}\text{-}1\}\}$

28   **for** $t \in \{0, N_{\text{t}} - 1\}$ **do**

29   Update the model image : $\vec{I}_{t}^{\text{m}} = I_{t}^{\text{m}} + g\,[\vec{I}_{p}^{\text{shp}} \star \vec{I}_{t}^{\text{m}}]$

30   **for** $s \in \{0, N_{\text{s}}\text{-}1\}$ **do**

31   Update the residual image : $\vec{I}_{s \atop t}^{\text{res}} = \vec{I}_{s \atop t}^{\text{res}} - g\,\sum_{q=0}^{N_{\text{t}}-1} [\vec{I}_{sp \atop tq}^{\text{psf}} \star \vec{I}_{t}^{\text{m}}]$

32   **end**

33   **end**

34   **until** *Peak residual in $\vec{I}_{0 \atop 0}^{\text{res}} < f_{\text{limit}}$*

35   Compute model visibilities $\vec{V}_{\nu}^{\text{m}}$ from $\vec{I}_{t}^{\text{m}}$ $\forall t \in \{0. N_{\text{t}} - 1\}$

36   Compute residual visibilities $\vec{V}_{\nu}^{\text{res}} = \vec{V}_{\nu}^{\text{obs}} - \vec{V}_{\nu}^{\text{m}}$

37 **until** *Peak residual in $\vec{I}_{0}^{\text{res}} < \sigma_{\text{thr}}$*

38 Restore the model Taylor-coefficients $\vec{I}_{t}^{\text{m}}$ $\forall t \in \{0. N_{\text{t}} - 1\}$

39 Calculate $\vec{I}_{\nu_0}^{\text{m}}, \vec{I}_{\alpha}^{\text{m}}, \vec{I}_{\beta}^{\text{m}}$ from $\vec{I}_{t}^{\text{m}}$ $\forall t \in \{0. N_{\text{t}} - 1\}$

40 If required, remove average primary beam : $\vec{I}_{\nu_0}^{\text{new}} = \vec{I}_{\nu_0}^{\text{m}}/\vec{P}_{\nu_0}; \vec{I}_{\alpha}^{\text{new}} = \vec{I}_{\alpha}^{\text{m}} - \vec{P}_{\alpha}; \vec{I}_{\beta}^{\text{new}} = \vec{I}_{\beta}^{\text{m}} - \vec{P}_{\beta}$

Algorithm 2: MS-MFS, as implemented in ASKAPsoft: red denotes difference from CASA

Data: calibrated visibilities : $\vec{V}_\nu^{\text{obs}} \; \forall \nu$

Data: uv-sampling function and weights : $[S_\nu], [W_\nu^{\text{im}}]$

Data: input : number of Taylor-terms N_t, number of scales N_s

Data: input : image noise threshold, σ_{thr}, loop gain g

Data: input : scale basis functions : $\vec{I}_s^{\text{shp}} \; \forall s \in \{0, N_s - 1\}$

Data: input : reference frequency ν_0 to compute $w_\nu = \left(\frac{\nu - \nu_0}{\nu_0}\right)$

Result: model coefficient images : $\vec{I}_t^{\text{m}} \; \forall t \in \{0, N_t - 1\}$

Result: intensity, spectral index and curvature : $\vec{I}_{\nu_0}^{\text{m}}, \vec{I}_\alpha^{\text{m}}, \vec{I}_\beta^{\text{m}}$

1   **for** $t \in \{0, N_t - 1\}, q \in \{t, N_t - 1\}$ **do**

2   Compute the spectral Hessian kernel $\vec{I}_{tq}^{\text{psf}}$ from transform of frequency weight sampling $\sum_\nu w_\nu^{t+q} S_\nu$

3   **for** $s \in \{0, N_s - 1\}, p \in \{s, N_s - 1\}$ **do**

4   Compute scale-spectral kernels $\vec{I}_{sp \atop tq}^{\text{psf}} = \vec{I}_s^{\text{shp}} \star \vec{I}_p^{\text{shp}} \star \vec{I}_{tq}^{\text{psf}}$

5   **end**

6 **end**

7   **for** $s \in \{0, N_s - 1\}$ **do**

8   Construct scale-dependent Taylor space Hessian $[H_s^{\text{peak}}]$ from $mid(\vec{I}_{s,s \atop t,q}^{\text{psf}})$ and compute $[H_s^{\text{peak}^{-1}}]$

9 **end**

10 Initialize the model $\vec{I}_t^{\text{m}}$ for all $t \in \{0, N_t - 1\}$

11 **repeat**   /* Major Cycle */

12   **for** $t \in \{0, N_t\text{-}1\}$ **do**

13   Compute $\vec{I}_t^{\text{res}} = \sum_\nu w_\nu^t \vec{I}_\nu^{\text{res}}$ from frequency-weighted residual visibilities $\vec{V}_\nu^{\text{res}}$

14   **for** $s \in \{0, N_s\text{-}1\}$ **do**

15   Convolve with s^{th} scale-function $\vec{I}_{s \atop t}^{\text{res}} = \vec{I}_s^{\text{shp}} \star \vec{I}_t^{\text{res}}$

16   **end**

17   **end**

18   Calculate minor-cycle threshold f_{limit} from $\vec{I}_0^{\text{res}}$

19   **repeat**   /* Minor Cycle */

20   **for** $s \in \{0, N_s\text{-}1\}$ **do**

21   **foreach** *pixel* **do**

22   Construct $\vec{I}_s^{\text{rhs}}$, an $N_t \times 1$ vector from $\vec{I}_{s \atop t}^{\text{res}} \; \forall t \in \{0, N_t\text{-}1\}$

23   Compute principal solution $\vec{I}_s^{\text{sol}} = [H_s^{\text{peak}^{-1}}] \vec{I}_s^{\text{rhs}}$

24   Fill solution $\vec{I}_s^{\text{sol}}$ into model images $\forall t : \vec{I}_{s \atop t}^{\text{m,sol}}$

25   **end**

26   **end**

27   Choose $\vec{I}_{p \atop t}^{\text{m}} = max\{\vec{I}_{s \atop t=0}^{\text{m,sol}}, \; \forall s \in \{0, N_s\text{-}1\}\}$

28   **for** $t \in \{0, N_t - 1\}$ **do**

29   Update the model image : $\vec{I}_t^{\text{m}} = I_t^{\text{m}} + g \, [\vec{I}_p^{\text{shp}} \star \vec{I}_{p \atop t}^{\text{m}}]$

30   **for** $s \in \{0, N_s\text{-}1\}$ **do**

31   Update the residual image : $\vec{I}_{s \atop t}^{\text{res}} = \vec{I}_{s \atop t}^{\text{res}} - g \, \sum_{q=0}^{N_t-1} [\vec{I}_{sp \atop tq}^{\text{psf}} \star \vec{I}_{p \atop t}^{\text{m}}]$

32   **end**

33   **end**

34   **until** *Peak residual in $\vec{I}_{0 \atop 0}^{\text{res}} < f_{\text{limit}}$*

35   Compute model visibilities $\vec{V}_\nu^{\text{m}}$ from $\vec{V}_t^{\text{m}} \; \forall t \in \{0.N_t - 1\}$

36   Compute residual visibilities $\vec{V}_\nu^{\text{res}} = \vec{V}_\nu^{\text{obs}} - \vec{V}_\nu^{\text{m}}$

37 **until** *Peak residual in $\vec{I}_0^{\text{res}} < \sigma_{\text{thr}}$*

38 Restore the model Taylor-coefficients $\vec{I}_t^{\text{m}} \; \forall t \in \{0.N_t - 1\}$

39 Calculate $\vec{I}_{\nu_0}^{\text{m}}, \vec{I}_\alpha^{\text{m}}, \vec{I}_\beta^{\text{m}}$ from $\vec{I}_t^{\text{m}} \; \forall t \in \{0.N_t - 1\}$

40 If required, remove average primary beam : $\vec{I}_{\nu_0}^{\text{new}} = \vec{I}_{\nu_0}^{\text{m}} / \vec{P}_{\nu_0}; \, \vec{I}_\alpha^{\text{new}} = \vec{I}_\alpha^{\text{m}} - \vec{P}_\alpha; \, \vec{I}_\beta^{\text{new}} = \vec{I}_\beta^{\text{m}} - \vec{P}_\beta$

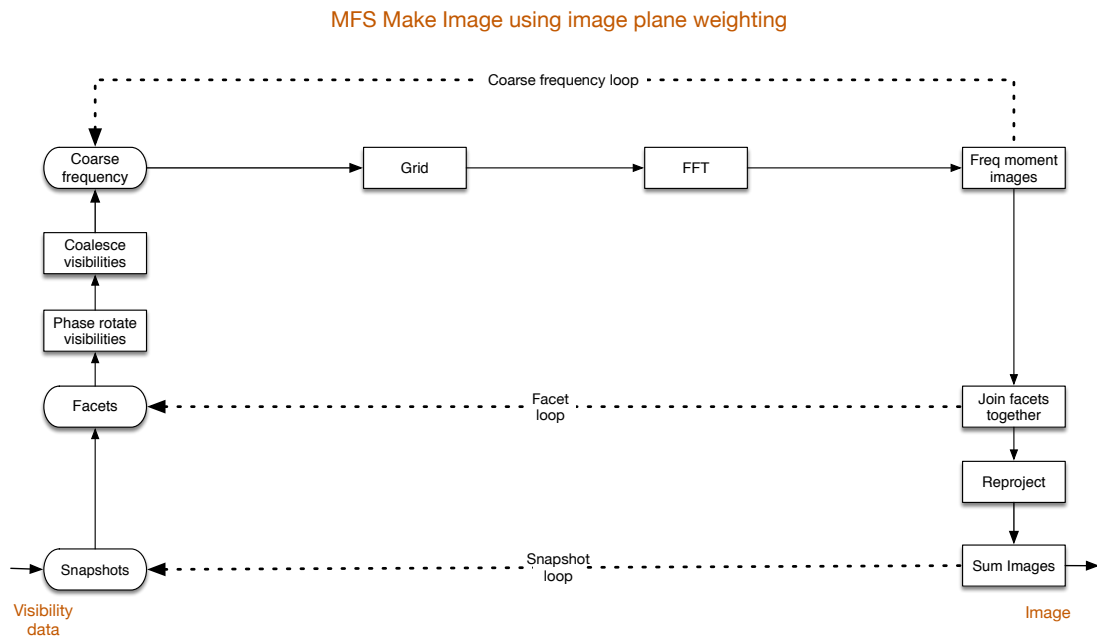**MFS Make Image using image plane weighting**

Figure 5: Structure of Image plane-weighting MakeImage, as used in CASA. Note that the three loops (denoted by the rectangle with curved ends) can be permuted at will.
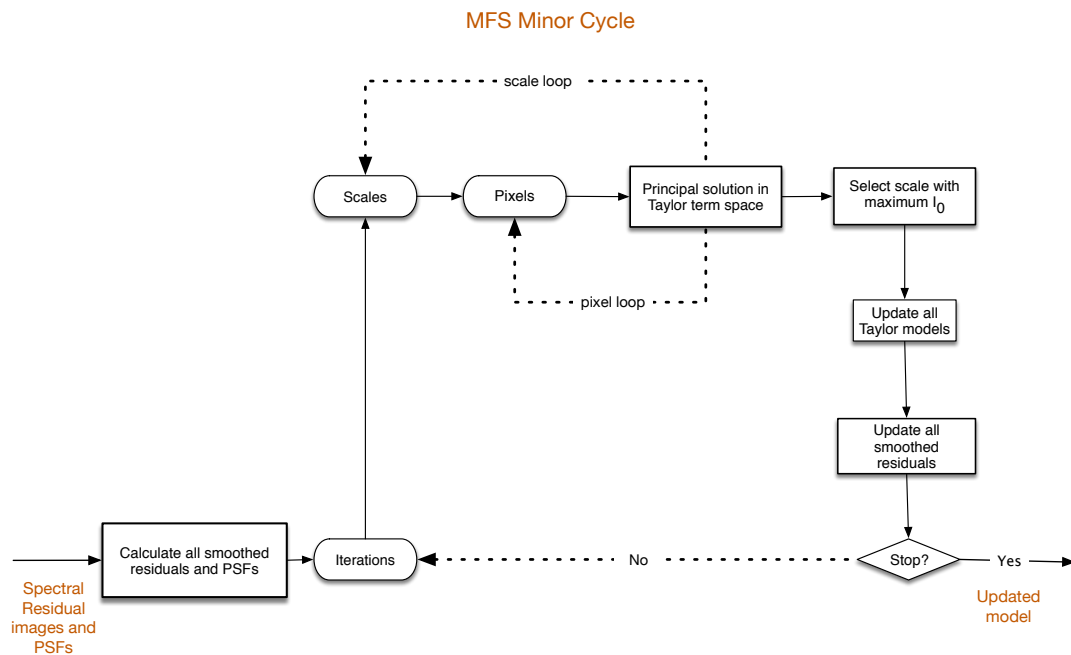

MFS Minor Cycle

Figure 6: Structure of Minor cycle for MSMFS as used in both CASA and ASKAPsoft.

Document no.: TCC-SDP-151123-2
Revision: 3
Release date: Wednesday 27th April, 2016 13:27

Unrestricted
Author: T.J. Cornwell
Page 25 of 33

7   Parallel processing

When moving the MSMFS algorithm to a distributed (loosely coupled) and parallel (tightly coupled) architecture such as the SDP architecture, there are a large number of important factors that must be considered.

Optimum partitioning

Coupling across partitions   e.g. is the same pixel addressed in two different partitions? If so, is there a satisfactory approach to reconciling the two values? This is a generic question that should be considered for all distributed and parallel processing.

Synchronisation points   e.g. is a global synchronisation point required so that the deconvolutions can be made consistent? Is there an acceptable algorithm for reconciling separate facets?

Amount of CU memory

Loading of CU memory backplane   The intrinsic operations in MSMFS are mostly simple SAXBY type operations. Hence the memory backplane will be overloaded, and the CPU data-starved.

Access to visibility store   What access to the visibility store is required? A straightforward implementation of the Major/Minor cycle algorithm loads the observed data every major cycle. However, as discussed in [RD15], this is actually only true of the initialisation and any self-calibration cycles. For other cycles, it is necessary only to load the uvw coordinates and flagging information.

7.1   Partitioning for Predict and MakeImage

The SDP pipeline framework [RD04], allows partitioning across multiple axes. In this case, the natural partitions and typical values are:

Polarisation   Only Stokes I can be modelled using MSMFS. Modelling the Faraday Rotation is possible but there is no existing algorithm.

Frequency   O(500) coarse channels are required to avoid bandwidth smearing/ There is no advantage to finer channels (see e.g. section 8.4 in [RD??]).

Sub-bands   O(1-3) partitions of the coarse channels are required to ensure that the power law approximation is adequate.

Time   O(43200)   O(1s) correlator samples are required to avoid smearing on the longest baselines

Facet   O(100) facets to limit the size of the AW projection kernel

Scales   O(5) scales are typically required to model extended structure (see section 8.4 in [RD??])

Taylor terms O(5) terms are required to represent the spectral behaviour for moderately complex spectrum sources (see section 8.4 in [RD??])

According to the logic of the MSMFS algorithm, sub-bands can be defined as the limit of the power law approximation for brightness so consequently can always be used as the coarsest partition. For Predict and MakeImage the remaining choices are as shown in Figure 3 (most rapid first):

Table 3: Possible ordering of partitions UV weighting version of Predict and MakeImage

Frequency	Time	Facet	Sub-Band
Frequency	Facet	Time	Sub-Band
Time	Facet	Frequency	Sub-Band
Time	Frequency	Facet	Sub-Band
Facet	Time	Frequency	Sub-Band
Facet	Frequency	Time	Sub-Band

The choice is between the two orderings of Time and Facet. Since deconvolution is non-linear, we prefer that all times be represented in a single facet rather than the other way around. Hence according to this level of analysis, the best ordering is [Frequency, Time, Facet, Sub-Band]. Parallelisation over Compute Nodes starts from the left to the right, Distribution over Compute Islands works from the right to the left (see Figure 7):
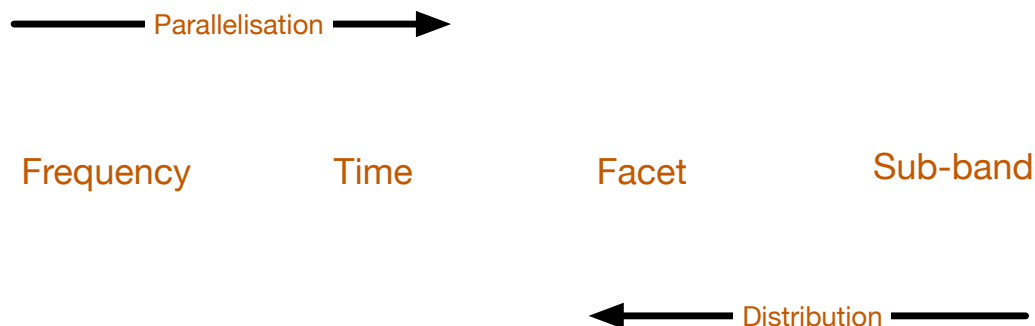
⟶ Parallelisation ⟶

Frequency   Time   Facet   Sub-band

⟵ Distribution ⟵

Figure 7: Preferred order of parallelisation and distribution for MakeImage and Predict sub-pipelines.

In this approach, the facets will almost certainly be processed on different compute islands. Note we have a dilemma on how to deal with the deconvolution of the facets.

- Construct the facets with some padding, deconvolve each separately, and reconcile after deconvolution by facet to facet broadcast.

- Construct the facets without padding, send all facets to one compute island, perform deconvolution, distribute models to facets.

- Each CI holds a distinct facet, which do not need to overlap. Synchronise in each minor cycle after peak finding. After a peak is found, all islands receive all peaks and choose the same global peak and update accordingly. Since the limit of coupling is assumed to be the CLEAN patch size, all peaks sufficiently far apart can be processed simultaneously.

The first approach will inevitably introduce edge effects in the final image revealing the facet grid, while the second will block processing while deconvolution proceeds on the different sub-band compute nodes. We would then benefit from two different types of nodes:

Predict and MakeImage Requires the visibility data but not many scratch images. Stalls during MinorCycle.

MinorCycle Requires a substantial number of scratch images, but not the visibility data. Stalls during Predict and MakeImage.

The third approach to deconvolution requires low latency communications. It should be investigated.

7.2 Partitioning for Minor Cycle

In the minor cycle, the first calculation is [Pixel, Taylor Term], and then the Scale convolutions are calculated. Hence the natural order overall is [Pixel, Taylor Term, Scale]. This is used in both CASA and ASKAPsoft. Both use a single threaded Minor Cycle. For SDP, a multi-threaded but not-distributed Minor Cycle should be prototyped.

However, in the minor cycle, we have a modelling problem that is not going to be wished away. The science demands seamless imaging in multiple axes, whereas computing would like to partition the image cube in those same axes. There is a tendency to assert that one axis or another does not have to be seamless but that has not be conclusively argued. Distributing the minor cycle across Compute Islands may be possible in some cases (e.g. spectral line after continuum estimation and subtraction), but we should also be considering how best to parallelise the Minor Cycle on a single, appropriately specified Compute Island.

7.3 Implementation in CASA

The CASA tclean program supports parallel execution using python to run copies on different nodes. This mode of parallel execution fits the CASA execution model. The visibility data are partitioned in time across the available processors. Gridding and transforms (i.e. Predict and MakeImage) are done separately and in parallel per partition. Then the Images (and weights) are gathered and then normalised. A serial minor cycle is run, blocking all other processing, and then the model image is scattered to all processes. The Major cycle is done in parallel per partition.

Thus in summary, the Major Cycles are parallelised while the deconvolution runs serially.

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27th April, 2016 13:27   Page 28 of 33

7.4 Implementation in ASKAPsoft

The ASKAPsoft code is mirrored in the SDP Github repository. The ASKAPsoft cimager is MPI-enabled, written as parallel program from the beginning/ Distribution over MPI is via frequency, rather than time as in CASA. Predict and MakeImage are distributed using MPI (via a thin bookkeeping layer). The MinorCycle is performed on the master, and thus blocks until the deconvolution steps have completed.

The minor cycle algorithm is in C++ Template DeconvolverMultiTermBasisFunction.tcc. The basis function is abstracted and can be any class having the interface DeconvolverMulti-TermBasisFunction. There are two forms present: one for point sources, and one for the same blobs used in CASA - a truncated upside down parabola. The basis functions may optionally be orthogonalised using the Gramm-Schmidt algorithm though there is little relevant experience.

The optimum blob location and scale is found by using one of two criteria:

MAXBASE0 The peak of term 0 across after decoupling in term.

MAXCHISQ The peak of chisquared across scale after decoupling in term.

The second criterion is expensive to compute and usually MAXBASE0 is used. Thus for each scale and for each pixel the pixel value at the reference frequency has to be determined by pre-multiplying by the inverse Hessian (in term-term space). This means that the cost of the search is:

$$Rflop_{MSMFSminor} = N_{minor}N_{pix}^2 N_s N_T \qquad (20)$$

To save on calculation, a number of cached images are used. For ASKAPsoft, these are shown in Table 4.

Table 4: ASKAPsoft cimager use of memory during minor cycles

Item	Axes	Size (words)	Needed?
Basis function FFT	scales	$N_{pix}^2 N_S$	During initialisation
PSF FFT	2 * Taylor terms	$2N_{pix}^2 N_T$	During initialisation
Spectral Dirty image	Taylor terms, scales	$N_{pix}^2 N_T N_s$	During iteration
Spectral Dirty PSF	Taylor terms, scales	$N_{patch}^2 N_T N_s$	During iteration
Spectral Model image	Taylor terms	$N_{pix}^2 N_T$	During iteration
Basis functions	scales	$N_{pix}^2 N_s$	During iteration
Inverse of Hessian	Taylor terms, Taylor terms, scales	$N_T^2 N_s$	During Iteration

where N_{pix} is the number of pixels on an axis of the image, N_{patch} is the number of pixels in the patch of the Dirty beam used, N_T is the number of Taylor terms, and N_S is the number of scales/basis functions.

Thus, keeping only the dominant terms, the total memory used during iteration is:

$$M_{MSMFS} = (N_T(N_s + 1) + N_s)N_{pix}^2 \qquad (21)$$

Document no.: TCC-SDP-151123-2
Revision: 3
Release date: Wednesday 27th April, 2016 13:27

Unrestricted
Author: T.J. Cornwell
Page 29 of 33

Using L2 assumptions values of $N_t = 5$ and $N_s = 10$, we find that the memory use is about $65 N_{pix}$. For a 200,000 by 200,000 pixel image, this requires 10.4 TB total. Spread over O(100) Compute Islands (partitioning by facet and sub-band), this requires 104 GB per CI. The overhead for performing the MSMFS minor cycle is a factor of fifty, and for the major cycle a factor of 5 (if ASKAP) in gridding, and a factor of O(500) in FFT (if CASA).

8 Other considerations

Ignoring the behaviour of the MSMFS algorithm, there are other considerations arising from the use of any MFS technique (this discussion drawn from section 8.4 of [RD07]).

Calibration The use of MSMFS or any technique taking advantage of simple models of broadband behaviour requires that the calibration across the observing band be well understood. Otherwise errors in bandpass calibration will couple into derived spatial structure. Hence there is a derived system requirement on bandpass accuracy, even just for continuum imaging. This should be evaluated via simulations and analysis.

Anomalous spectral index objects such as some pulsars, GHz-peaked sources, very extended radio galaxies, may limit dynamic range.

Primary beam model If the true spectrum is to be recovered, the primary beam must be modelled accurately.

UV coverage effects The fine structure is constrained by the high frequency measurements and the extended emission is constrained by the low frequency measurements. In principle, these biases can be avoided by limiting the range of uv coordinates to a common set at all frequencies, but this does somewhat defeat the purpose of MFS. It is not clear how serious this effect will be for SKA.

In addition, Bhatnagar et al. [RD18] have drawn attention to the difficulties of performing primary beam correction over a wide bandwidth. They note some difficulties with the straightforward application of what they call the Narrow Band MSMFS algorithm (which we call simply MSMFS) to large bandwidths. Specifically the apparent Noise to Signal Ratio becomes very large outside the primary beam. They argue that analysis of the images scientifically relies upon understanding the noise level outside the primary beam sampled area. This is undoubtedly correct, and MIRIAD and ASKAPsoft address this by designating a primary beam cutoff beyond which the noise alone is used without normalisation. In their approach, Bhatnagar et al. advocate correcting the primary beam to a desired effective primary beam. This is quite sensible. Some evaluation of this approach should be performed using simulations, and the scientific analysis implications should be considered.

9 Recommendations

Parallel versions of both variants of MSMFS exist. These work by distributing in time (CASA) and frequency (ASKAPsoft). Parallelisation of the Minor Cycle is not straightforward and consequently both CASA and ASKAPsoft currently use single-threaded inner loops. This is unlikely to be successful for SKA, and instead the Minor Cycle has to be parallelised inside one CI and perhaps additionally distributed over multiple CI's.

In light of this memo, the following recommendations are made concerning the need for MFS:

1. Perform simulations to explore the need for MSMFS in SKA. Currently an L1 requirement directs SDP to implement MSMFS. However there is no justification given for this very computationally demanding processing. Indeed the promised excellent UV coverage of the SKA telescopes makes it at least questionable, although arguing from the work of Urvashi [RD??] does indicate that with MFS, the dynamic range would be limited to 10^4 for wide frequency span. Either CASA or ASKAPsoft could be used for the reconstructions.

2. Adding to the desire to see some more justification of the requirement, we note that the L1 requirements do not address the number of scales required, the number of Taylor terms, or the possibility of sub-banding (all of which appear in the SDP L2 requirements. We note three possible outcomes of the simulations: MFS can be ignored, MFS is required and MSMFS is sufficient, or MFS is needed and MFMFS is not sufficient. In the case of the latter outcome, investigation of other deconvolution techniques (e.g. [RD14]) would be required.

3. Assess the importance of the considerations in Section 8.

The following recommendations address the resource usage:

4. Implement both Image (CASA) and Visibility based (ASKAPsoft) calculation of the Taylor term images in the SDP Performance Model. The main reasons to choose one over the other are to do with the relative speed of gridding and FFTs. When it comes to implementation during construction, the costs of implementing both approaches should be minimal, and unimportant compared to the risk of implementing the wrong choice.

5. Update the SDP Performance Model to reflect the MSMFS memory usage:

$$M_{MSMFS} = (N_T(N_s + 1) + N_s)N_{pix}^2 \tag{22}$$

6. Update the SDP Performance Model to reflect the MSMFS Flops:

$$Rflop_{MSMFSminor} = N_{minor}N_{pix}^2 N_s N_T \tag{23}$$

7. Prototype a multi-threaded Minor Cycle on one Compute Island. GPUMultiscaleCLEAN might be a good starting point. Do MS and then MSMFS. This can be done without implementing a major cycle.

8. Model and then, if promising, prototype a distributed Minor Cycle on multiple CI's where synchronisation occurs every iteration. Do MS and then MSMFS. This can be done without implementing a major cycle.

The following recommendations address wider issues:

9. The SKAO Science Team should be consulted about the impact of seams in the images or spectral cubes. These may occur both spatially (if the facets are deconvolved separately, or in frequency, because of the disjoint use of MSMFS in the different sub-bands. It seems questionable that these can be tolerated.

10. We should note that MSMFS was developed in a serial platform where computation was expensive and memory access cheap. In the SDP context, the processing is distributed and parallel, and computation is cheap and memory access expensive. Although SDP has no remit to develop new algorithms, it could encourage researchers to address this new context.

Document no.: TCC-SDP-151123-2   Unrestricted
Revision: 3   Author: T.J. Cornwell
Release date: Wednesday 27th April, 2016 13:27   Page 33 of 33