

SMA Technical Memo # TBA

Subject: MAPPING SMA STORAGE FORMAT TO THE SOFTWARE PACKAGES CASA & MIRIAD -
Software limits for SMA pipelines and specific routines

Date: January 28, 2022

From: Jun-Hui Zhao (SAO)

Updated from the versions since June 12, 2020; (6/12/2020, 8/10/2020, 9/14-17/2020, 9/20-21/2020, 9/23/2020, 9/25/2020).

Updated version: <https://lweb.cfa.harvard.edu/~jzhao/warea/SMATechnemo/SOFTWlimits/SMAMemoSOFTWlimits.pdf>

Abstract

We describe the software limits that are implemented for mapping the data format (Caltech MIR¹) of SMA online storage to the offline software packages (CASA² and MIRIAD³) for data reduction. The CASA supported by NRAO is the major offline software package for reduction of the wSMA data. The reduction of old SMA archival data can be carried out with MIRIAD that was developed by the BIMA group and powered by the SMA for handling submillimeter interferometer data. In this memo, we documented the software limits that used in developments of the MIRIAD pipeline SMALOD; this specific MIRIAD routine for SMA converts the data from ASIC correlator and ASIC+SWARM hybrid correlator to the BIMA format. The software limits used in the CASA pipeline SWARM2CASA, converting SWARM data to measurementSet format via FITS format, are also described. The memo provides the documents that can be used in the maintenance of the SMA MIRIAD for the reduction of SMA archive data. These documents would be also useful for the further development of SMA pipelines, such as PYUVDATA, for reductions of wSMA data with CASA.

1 Introduction

The SMA[1][2] is a pioneer telescope in development of interferometer arrays at submillimeter wavelengths, such as the ALMA [3]. Mapping online storage formats of the SMA data into offline software packages is essential for scientific productions at the SMA. Over the past two decades, the SMA has been upgraded with the cutting-edge technology, exploring the architectures of correlators[4] for applications in interferometer array telescopes. For example, with the XF architecture, the SMA developed its first generation correlator with the application specific integrated circuits (ASIC)[1][5]. The first quadrant of the SMA ASIC correlator was accomplished and deployed to the telescope before 2008, processing a full 2-GHz bandwidth of data. In the following phase of development by implementing the second quadrant of ASIC correlator, the SMA was enhanced by doubling its capability with a full 4-GHz bandwidth. The ASIC correlator with the XF architecture has been adopted for the ALMA development[6].

In recent, along with the development for the Event Horizon Telescope (EHT)[7], the SMA wideband astronomical ROACH2 Machine (SWARM) has been developed for a 32 GHz bandwidth VLBI capable correlator and phase array[8]. The SWARM utilizes the FX architecture[8], and the

¹Millimeter Interferometer Reduction (MIR) is an IDL based software package, originally developed by Nick Scoville at Caltech for use with the OVRO millimeter array.[16]

²CASA, the Common Astronomy Software Applications package, is the primary data processing software for the Atacama Large Millimeter/submillimeter Array (ALMA) and NSF's Karl G. Jansky Very Large Array (VLA), and is frequently used also for other radio telescopes.[30]

³MIRIAD (Multichannel Image Reconstruction, Image Analysis and Display) is a toolbox, in the form of an environment with a large set of moderate-sized programs which perform individual tasks, involving calibration, mapping, deconvolution and image analysis of interferometric data. MIRIAD software was also used in parts of the Hat Creek (BIMA) telescope control software; data obtained from the telescopes is directly written into MIRIAD format with a MIRIAD user interface.[22][26]

system has been deployed at the SMA and EHT[7-9]. The 32-GHz bandwidth capacity with the 4 quadrants of the SWARM[9] are being used for the science operation; and the SMA has truly entered its wideband era (wSMA)[10]. In more recent development, two additional quadrants of the SWARM correlator have been depolyed at the telescope[42].

As the rapid progress in upgrade of the SMA correlator, the on-line storage formats were developed based on VLBA FITS-IDI[11-15] and Caltech MIR/IDL[16][17]. The SMA current online storage system adopts the Caltech MIR format[18]. Data pipelines were developed to support SMA data reduction on the software platforms, including AIPS[19][20], AIPS++[21][12], MIRIAD[22-26], MIR/IDL[16][27-28], GILDAS[29] and CASA[30-35]. A more general pipeline with Pythron PYUVDATA is being developed[34] to support reduction of wSMA data on CASA. In present, PYUVDATA has been successfully demonstrated to pipeline SMA storage data into CASA measurementSet format, handling the array data taken with the one receiver mode.

The rest of the memo describes the software limits that are used in development of pipeline software to convert the SMA storage data from the SMA correlators to MIRIAD and CASA. These limits were used in the development of the SMA pipelines for MIRIAD and CASA prior to the PYUVDATA development. Section 2 describes the software limits used in MIRIAD program SMALOD, the pipeline of converting SMA ASIC correlator data into BIMA format. Section 3 describes the software limits implemented in SMALOD to handle the ASIC and SWARM hybrid data. Section 4 describes the software limits used in SWARM2CASA[31], the pipeline to convert SWARM data to CASA format via FITS[35-41]. Section 5 is a short summary for this document.

Table 1: Software limits for handling ASIC correlator data

Variable name	Value	Brief description
MAXCHAN	6144 [†]	Max channels per BL, SB, RX, POL
MAXANT	8	Max number of antennas for SMA
MAXBAS	112	Max number of baselines = $\text{MAXANT} \times (\text{MAXANT} - 1) / 2 \times 2\text{RX} \times 2\text{SB}$
MAXINT	10,000	Max number of int scans = 13.9h for INT of 5s
MAXSCHAN	4096	Max channels per SPW or spectral chunk
MAXCHUNK	49	Max number of SPW per SB
MAXSRC	100	Max number of sources per track, 8.3 m per source in a 13.9-h track
MAXRX	2	Max number of RX per track
MAXPOL	4	Max POL states
MAXDATA	688,240	Max data per INT
MAXDATAT	6,882,400,000	Max data per track
MAXBYTE	55 GB	Max bytes per track

[†]MAXCHAN does not include the single channel of the continuum chunk.

2 Limits for ASIC correlator

The MIRIAD program SMALOD was developed to pipeline the SMA data produced from the ASIC correlator to MIRIAD via BIMA format. The max number spectral channels (MAXSCHAN)

produced from a chunk of ASIC correlator is 4,096 channels, covering a bandwidth of 104 MHz per spectral window (SPW). Overlapping the beginning and ending of the spectral channels between the adjacent chunks, a total of 48 spectral chunks produce a full 4-GHz bandwidth. The highest spectral resolution that can be achieved with the ASIC correlator is 25.4 kHz. However, there were a limited number of ASIC correlator chips. In order to cover the full 4-GHz bandwidth with two quadrants of the ASIC correlator, we have to compromise the spectral resolution. So the max channels per baseline (BL), per sideband (SB), per receiver (RX), and per polarization state (POL) is 6,144 for the standard SMA with 8 elements of antenna, giving max channel number of 128 per chunk or per SPW and a total of 48 spectral chunks. We note that the spectral resolution or the number channel for the ASIC correlator is configurable with the max number of channels 4096. For a standard operation with two sidebands and 128 channels per chunk, the spectral resolution corresponds to 0.8125 MHz.

The max number of baselines per integration (INT) is 122 with two receivers and two sidebands. The max number of visibility data points per integration (MAXDATA) equals to the $\text{MAXBAS} \times (\text{MAXCHAN} + 1) = 688,128$, where the continuum chunk, with a single channel computed by vector-averaging the spectral data, is included. Given the max number of integrations per track $\text{MAXINT} = 10,000$, the max number of visibility data points per observing track is 6,882,400,000. So the max data size per file is 55 GB.

Table 1 summarizes the software limits for handling ASIC correlator data. The limits were used in the development of the SMA pipeline SMALOD in MIRIAD.

3 Limits for ASIC+SWARM hybrid correlator

After successfully implementing the second quadrant of the ASIC correlator for 4-GHz bandwidth, the SMA was in an active phase of developing the SWARM correlator with FX architecture. Shortly, the SMA deployed the first two quadrants of the SWARM correlator to the telescope. In the 2011 fall, we received a data set consisting of 48 spectral chunks produced from the legacy ASIC correlator and 2 spectral chunks produced from the new SWARM correlator. Unlike the ASIC, the number of channels in a SWARM spectral chunk is fixed to 16,384, covering 2-GHz bandwidth. Thus, the spectral resolution in a raw SWARM data is high, 140 kHz. The pipeline software must be re-configure to compromise with the flexibility in the spectral configuration produced by the hybrid ASIC and SWARM correlators. The max number of channels now is the sum of the max number of channels for the ASIC correlator $\text{MAXCHAN}_{\text{ASIC}}$ and two times of the channel number in a SWARM chunk $\text{MAXCHAN}_{2\text{SWARM}}$; thus,

$$\text{MAXCHAN} = \text{MAXCHAN}_{\text{ASIC}} + \text{MAXCHAN}_{2\text{SWARM}} = 6,144 + 2 \times 16,384 = 38,912.$$

The max number of channels per spectral chunk is $\text{MAXSCHAN}=16,384$. The spectral configuration becomes the 48 ASIC chunks (S1-S48) and the 2 SWARM chunks (S49-S50) with a single-channel continuum chunk (C1). Total number of spectral chunk is $\text{MAXCHUNK} = 51$.

As the market price of RAM dropped, we also increased the max number of integrations to $\text{MAXINT}=12,960$ and max number of sources to $\text{MAXSRC} = 200$. Thus, the max number of visibility data points per integration becomes $\text{MAXDATA} = 4,358,256$. The max number of visibility data points per track becomes $\text{MAXDATAT} = 56,482,997,760$. The max size per file is 452 GB. Table 2 summarizes the software limits for supporting the hybrid spectral configurations produced by the ASIC and two quadrants of the SWARM correlators.

Table 2: Software limits for handling ASIC+SWARM hybrid correlator data

Variable name	Value	Brief description
MAXCHAN	38,912 [†]	Max channels per BL, SB, RX, POL
MAXANT	8	Max number of antennas for SMA
MAXBAS	112	Max number of baselines = MAXANT × (MAXANT − 1)/2 × 2RX × 2SB
MAXINT	12,960	Max number of INT scans = 18h for 5s integration
MAXSCHAN	16,384	Max channels per SPW or spectral chunk
MAXCHUNK	51	Max number of SPW per SB.
MAXSRC	200	Max number of sources per track, 10.8 min per source in a 18h track.
MAXRX	2	Max number of RX per track
MAXPOL	4	Max POL states
MAXDATA*	4,358,256	Max data points per INT
MAXDATAT*	56,482,997,760	Max data points per track
MAXBYTE	452 GB	Max bytes per track

[†]MAXCHAN does not include the single channel of the continuum chunk.

[‡]The max number of chunks (spw) included in the hybrid frequency configurations produced from ASIC and two quadrants of SWARM correlators. The SPWs of the ASIC+SWARM spectral configuration include a C1, 48 ASIC, and 2 SWARM chunks.
C1: A single channel, vector-averaged of the spectral data from S1-S50; S1-to-S48: The 48 spectral chunks produced from the ASIC correlator. S49-to-S50: The two SWARM chunks, 2 GHz BW each, produced from the SWARM correlator.

*MAXDATA = (1 + MAXCHAN) × MAXBAS.

*MAXDATAT = MAXDATA × MAXINT.

4 Limits for SWARM correlator

As the ASIC correlator retired from the science operation, the SWARM correlator becomes the correlator to process the interferometer data sampled by the array of the 8 elements. Now a total of 6 quadrants (Q1-Q6) of SWARM correlator have been deployed to the telescope. The capability of the SMA has been enhanced by increasing the processed bandwidth to 12 GHz per sideband per receiver, at 140 kHz resolution for all the SWARM spectra. The SMA can configure the telescope to provide a unique capability of continuous 44 GHz spectral coverage (with 4 GHz of ‘overlap’ coverage) within the common frequency regions of dual receiver pairs[42].

The data volume generated by the SWARM correlator has drastically increased. The large volume of data has exceeded the MIRIAD’s capacity to simultaneously calibrate the wideband data with the original spectral resolution. However, the CASA has provided a capability of handling SMA SWARM data. Developments of SMA pipelines to CASA have been carried out. β -versions of pipeline SWARM2CASA installed on the SMA and the SAO Radio Telescope Data Center (RTDC) computing systems are being internally tested; and its data patches and some algorithms have been and will be implemented into online software and more general pipelines such as PYUVDATA. As attested, with the SWARM2CASA path, the CASA can process wSMA data by editing, calibrating, and imaging. A data patch for corrections of the online Doppler-errors occurred during 2011-2019[43] has been integrated into SWARM2CASA[44]; so the Doppler errors can be automatically

Table 3: Software limits for converting SWARM data to CASA

Variable name	Value	Brief description
NC	6	Number of SWARM quadrants
NCHANSPW	16,384	Max. channels per SWARM SPW or chunk
NCHANSWM ¹	98,304	Max. SWARM channels per BL, SB, RX, POL
NSB	2	Number of sideband
MAXRX	2	Max. number of RX per track
MAXSPW ²	7	Number of C1 & SWARM SPW per SP, per RX
NCHAN ³	98,305	Max. channels per BL, SB, RX, POL
MAXCHAN ⁴	393,220	Max. channels per BL, POL
MAXANT	8	Max. number of ANT's for the SMA
MAXPOL	4	Max. pol. states per ANT pair in dual RX
MAXBAS ⁵	224	Max. number of baselines
MAXINT	18,000 ⁶	Max. number of INT scans.
MAXDATA ⁷	22,020,320	Max. data per INT.
MAXSRC	200 ⁸	Max. number of sources per track.
MAXLENSRC	32 ⁹	Max. number of char. for source name string.
SBCHUNK ¹⁰	24	Max. number of SPW per SB.
MAXCHUNK ¹¹	48	Max. number of SPW per measurementSet.
MAXDATAC ¹²	22,020,096	Max. data per INT, per measurementSet.
MAXDATAT ¹³	396,361,728,000	Max. data per measurementSet.
MAXBYTE ¹⁴	3 TB	Max. byte size per measurementSet.

¹NCHANSWM=NC×NCHANSPW =98,304, does not include the single channel of the continuum chunk.

²MAXSPW=1+NC, included the single channels chunk C1: vector-averaged of the spectral data, or ch0.

³NCHAN= 1+NC×NCHANSPW, included the 1 channel of C1.

⁴MAXCHAN=MAXRX×NSB×(NC×NCHANSPW+1).

⁵MAXBAS=MAXPOL×NSB×MAXANT×(MAXANT-1)/2.

⁶The value corresponds to the number of int scans in an observing duration 15h for 3s int, or an observing duration 5h for 1s int.

⁷MAXDATA=(1+NC×NCHANSPW)×MAXBAS, included the 1 channel of C1.

⁸This value corresponds to a minimum int time of 4.5 min per source in a 15h observing duration.

⁹The last char. of the source name string is padded with " 0"; the actual number of char. is 31. However, a limit of 16 characters is imposed by UVFITS that is the intermediate format converting SMA online storage/archive data format into measurementSet of CASA.

¹⁰SBCHUNK=MAXPOL×NC.

¹¹MAXCHUNK=MAXPOL×NSB×NC, disclosed the C1 chunk, assuming that the frequency configurations produced from each of the two receivers are not identical.

¹²MAXDATAC=NC×NCHANSPW×MAXBAS.

¹³MAXDATAT=NC×NCHANSPW×MAXBAS×MAXINT.

¹⁴Only for float vis data, excluding the ~5% overhead for the accessory tables.

Note: The vector-averaged chunk C1 contained a single channel before Q5 and Q6 were deployed. For 6 quadrants, C1 contains 6 channels of data.

corrected via the SWARM2CASA path. Also, the CASA provides a flexible software platform via Python to allow users to develop their own routines or procedures for preparation of SMA observations and analysis of SMA images.

Table 3 summarizes the software limits used in the CASA pipeline SWARM2CASA. We introduced a variable NC to account for the number of SWARM quadrants deployed to the telescope over past a few years in the development of the SWARM correlator. During 2017 and early 2018, NC=3 was used for the first three SWARM quadrants, then NC = 4 for the four SWARM quadrants, and now NC = 6 for the regular operations with six quadrants of the SWARM correlator. We note that the SMA online system has not provided measurements of spectrum-based Tsys yet. In the near future, algorithms for Tsys corrections and visibility weight calculation are needed to implement the SMA data pipelines to CASA. Therefore, Table 3 may need to be updated as well.

5 Summary

The software limits described in this memo cover the three development periods for the SMA offline software. In the first two periods, involving the SMA data produced from the legacy ASIC correlator, the hybrid ASIC and two quadrants of the SWARM correlators, we developed a pipeline SMALOD and numerous SMA specific routines and sub-routines to support SMA data reduction with Miriad. The SMA now is in the wideband generation, equipped with the SWARM correlator with a number of quadrants upto 6. Over the past decade, the volume of SMA data per astronomical track has increased by 2 orders in magnitude. The CASA appears to be the data reduction software for wSMA. Various pipeline paths to CASA have been tried out for SMA data reduction over the past a few years. The documents written in this memo can be used in maintenance of SMA MIRIAD software for reduction of the old but valuable SMA data in the RTDC archive, and would be also useful for the developments of pipelines for reduction of wSMA data with CASA.

References -

- [1] P.T.P. Ho, J.M. Moran, K.Y. Lo 2004, ApJL, 616, L1
- [2] R. Blundell 2004, In *Fifteenth International Symposium on Space Terahertz Technology*, Northhampton, MA
- [3] Atacama Large Millimeter/submillimeter Array (ALMA) - <https://public.nrao.edu/telescopes/alma/>
- [4] Thompson, A. R., Moran, J. M. & Swenson Jr., G. W. 2017, *Interferometry and Synthesis in Radio Astronomy*, Third Edition, Springer, Switzerland
- [5] Peck, A., Schinckel, A., SMA Team 2007, in *Exploring the Cosmic Frontier: Astrophysical Instruments for the 21st Century*, ESO Astrophysics Symposia, European Southern Observatory series. Edited by A. P. Lobanov, J. A. Zensus, C. Cesarsky & P. J. Diamond; Springer-Verlag, Berlin and Heidelberg, Germany, p. 49
- [6] Escoffier, R. P., Comoretto, G., Webber, J. C., et al. 2007, A&A, 462, 801
- [7] Event Horizon Telescope (EHT) - <https://eventhorizontelescope.org>
- [8] Primiani, R. A., Young, K. H., Young, A., et al. 2016, JAI, Vol. 5, No. 4, id 1641006-810
- [9] SMA SWARM - <https://www.cfa.harvard.edu/sma/Projects/SWARM/>
- [10] Wideband SMA (wSMA) Specification - <http://events.asiaa.sinica.edu.tw/workshop/20161027/specification.php>
- [11] Zhao, J.-H., Mailhot, P. & Tsutsumi, T. 2000, *The SMA On Line Data Archive & Storage System: Software Development and Hardware Prospects*, SMA Technical Memo #138
- [12] Zhao, J.-H. & Tsutsumi, T. 2001, *Astronomical Data Analysis Software and Systems X, ASP Conference Proceedings*, Vol. 238. Edited by F. R. Harnden, Jr., Francis A. Primini, and Harry E. Payne. San Francisco: Astronomical Society of the Pacific, p164

- [13] Diamond, P. J., Benson, J., Cotton, W., D., Wells, D. C., et al. 1998, *FITS Interferometry Data Interchange*, VLBA correlator memo No. 108
- [14] Flatters, C. 1998, *Information About the FITS Interferometry Data Interchange Format*, <https://www.cv.nrao.edu/fits/documents/drafts/idi-format.html>
- [15] Greisen, E. W. 2011, *The FITS Interferometry Data Interchange Convention — Revised*, in AIPS Memo 114r
- [16] Keto, E., MIR/IDL data path - <https://www.cfa.harvard.edu/sma/mir>
- [17] SMA MIR storage formats for ASIC correlator - <https://www.cfa.harvard.edu/sma/dataFormat/index.html>
- [18] Young, K. (Taco) & A. Kovacs 2020, SMA Wiki: *Current SMA data file format (for SWARM correlator)* updated on 2020-07-02 & updated by M. Gurwell on 2020-07-30
- [19] NRAO AIPS - <http://www.aips.nrao.edu/index.shtml>
- [20] Zhao, J.-H., Young, K. H., Herrnstein, R. M., et al. 2003, ApJL, 586, L29
- [21] NRAO AIPS++ - https://casa.nrao.edu/aips2_docs/aips++.html
- [22] Sault, R. J., Teuben, P. J., & Wright, M. C. H. 1995, in *Astronomical Data Analysis Software and Systems IV*, edited by R. A. Shaw, H. E. Payne, & J. J. E. Hayes, ASP Conf. Series, 77, 433
- [23] Zhao, J.-H. 2013, in *Astronomical Data Analysis Software and Systems XXII*, Proceedings of a Conference held at University of Illinois, Champaign, Illinois, USA 4-8 November 2012. San Francisco: Astronomical Society of the Pacific, 2013, p.181
- [24] MIRIAD for SMA - <https://www.cfa.harvard.edu/sma/miriad/>
- [25] MIRIAD for ATCA - <https://www.atnf.csiro.au/computing/software/miriad/>
- [26] MIRIAD for BIMA & CARMA - <https://www.astro.umd.edu/~teuben/miriad/>
- [27] Qi, C., The MIR Cookbook - <https://www.cfa.harvard.edu/~cqi/mircook.html>
- [28] SAO RTDC, MIR for SMA - <https://www.cfa.harvard.edu/rtdc/SMAdata/process/mir/>
- [29] GILDAS for IRAM 30m and NOEMA - <http://www.iram.fr/IRAMFR/GILDAS/>
- [30] NRAO CASA - <https://casa.nrao.edu>
- [31] Zhao, J.-H., SMA pipeline to CASA via FITS - <https://www.cfa.harvard.edu/~jzhao/SMA-CASA/>
- [32] Young, K. H., SMA pipeline to CASA via FITS-IDI - <https://www.cfa.harvard.edu/rtdc/SMAdata/process/casa/convertcasa/sma2casa/>
- [33] SAO RTDC, CASA for SMA - <https://www.cfa.harvard.edu/rtdc/SMAdata/process/casa/>
- [34] Discussion on *PYUVDATA*, in the weekly SMA users' meeting organized by Garrett Keating, 2020
- [35] Wells, D. C., Greisen, E. W., & Harten, R. H. 1981, A&AS, 44, 363
- [36] Greisen, E. W. & Harten, R. H. 1981, A&AS, 44, 371
- [37] Grosbøl, P., Harten, R. H., Greisen, E. W., and Wells, D. C. 1988, A&AS, 73, 359
- [38] Harten, R. H., Grosbøl, P., Greisen, E. W., and Wells, D. C. 1988, A&AS, 73, 365
- [39] Cotton, W. D., Tody, D. B., and Pence, W. D. 1995, A&AS, 113, 159
- [40] Greisen, E. W. 2019, *AIPS FITS File Format*, in AIPS Memo 117 of NRAO
- [41] NASA/GSFC: *The FITS Support Office* - <https://fits.gsfc.nasa.gov>
- [42] SMA Observer Center: *Array Status and Technical Information* - <http://sma1.sma.hawaii.edu/status.html>
- [43] https://www.cfa.harvard.edu/sma/Newsletters/pdfFiles/SMA_NewsJuly2019.pdf, p22
- [44] Zhao, J.-H. 2020, *ALGORITHMS - Corrections for Doppler tracking errors*, SMA Technical Memo #TBA
- [45] Zhao, J.-H. 2020 updated, *SWARM2CASA PIPELINE - Status and Software Architecture*, SMA Technical Memo #TBA