MWAX Visibility File Format

The MWA Correlator v2 (aka MWAX Correlator) is still under development. This article summarises the significant differences between the legacy (Ord /Current/v1) correlator and the new MWAX correlator.

> New Correlator is in development, so the info on this page may change without notice!

- Introduction
  - Overview of Changes to Assumptions in the new MWAX correlator
- Correlator File Naming Convention
- FITS Format
  - FITS Primary HDU
  - FITS Extension Visibility ImgHDU
    - Visibility Header
    - Visibility Data
  - FITS Extension Weights ImgHDU
    - Weights Header
    - Weights Data

Introduction

Each coarse channel is processed separately on a correlator server (traditionally called a "gpubox"). When operating at full capacity there are 24 coarse channels processed (although the '24' number is arbitrary and can change per observation). Thus there will be multiples of 24 gpubox files produced per observation if all gpuboxes are working. In order to prevent any individual gpubox file / coarse channel file becoming too big, the files may be split into multiple parts. However this split always occurs so as to never break up a sub-observation (8 secs). This splitting is taken care of by the data capture code.

Overview of Changes to Assumptions in the new MWAX correlator

- The associated metafits file for each observation should be considered the source of truth for everything, except things that vary per coarse channel, which will be in the primary HDU of each coarse channel FITS file.
  - Only the most basic information such as the obsid, projectid, time, correlator mode are repeated in the FITS Primary HDU (purely for convenience).
- The number of antenna can change on observation boundaries, so do not assume 128. The new MWAX correlator is designed to support 256+ tiles, but will likely start at 128 and grow as next generation receivers are added to the array. E.g. the next increment might be 144T.
- Within the FITS image HDUs for visibilities, NAXIS1 represents the fine channels * polarisations * 2 (real/imaginary). NAXIS2 represents the baselines (antennas x (antennas+1))/2
- Antennae are ordered by the "antenna" value as per the ‘BINTABLE’ in the associated metafits file called ‘TILEDATA’. See MWAX Antenna Ordering for more details.
- Baselines are antenna vs antenna, not input vs input (e.g. ant0 vs ant2, not ant0x vs ant2y)
- Baselines are in lower regular triangular order. 0-0..0-n, then 1-1..1-n, then 2-2..2-n, etc.
- Coarse channel numbers (CORRCHAN) are always in sky frequency ascending order. (There is no longer a receiver channel number ordering flip when receiver channel number is >128)
- Polarisations are in xx,xy,yx,yy order
- Real and imaginary data values are 32 bit floats
- There are significantly more correlator modes to support. Rely on FINECHAN and INT_TIME for the correct mode information from the metafits. The current working list of supported modes is shown the below figure "MWAX Correlator 256T modes" (supported modes in green; unsupported modes in red).
- The number of coarse channels per observation could change (once we deploy replacement receivers, or if we allow astronomers to choose LESS than 24 course channels with the existing receivers, or if an MWAX server is offline), so do not assume 24.
- Instantaneous bandwidth / coarse channel width could change (once we deploy replacement receivers), so do not assume 30.72 MHz (24 x 1.28 MHz).
- Weights are provided in the gpubox files (after each visibility HDU containing one timestep/integration). The weights and how they are determined is discussed below:
  - Each visibility has a multiplicative weight applied, based on a data occupancy metric that takes account of any input data blocks that are missing due to lost UDP packets or RFI excision (a potential future enhancement). The centre (DC) ultrafine channel is excluded when averaging and the centre output channel values are re-scaled accordingly. Note that only 250 Hz of bandwidth is lost in this process, rather than a complete output channel.
  - As part of the M&C system, the application of weights can be turned on or off per observation based on the science case/needs. With weights not applied, the data will be averaged in the correlator without taking into account the weights. Either way the weights are supplied in each alternate ImgHDU for your information.
- The MWAX correlator will provide options on a per observation to apply geometric and cable delays. When these are on, they do not need to be done by downstream tools such as Cotter or the RTS.
- **Existing visibilities from the legacy/Ord correlator still need to be supported by downstream tools** (e.g. Cotter / RTS).
- The MWAX FITS files contain a keyword "CORR_VER" which represents the correlator version number. If this keyword is missing, assuming this is a FITS file from the Ord/Legacy correlator. "2" is the value for the MWAX correlator.

Data Rate of Visibility Output of MWAX servers combined @ 256T

Based on 256T correlated and 30.72 MHz instantaneous bandwidth

Fine channels (per 1.28 MHz coarse channel)	fine channel			Time Resolution (s)					
	fscrunch	count	width kHz	0.25	0.5	1	2	4	8
	1	6400	0.2	5,164.03	2,582.01	1,291.01	645.50	322.75	161.38
	2	3200	0.4	2,582.01	1,291.01	645.50	322.75	161.38	80.69
	4	1600	0.8	1,291.01	645.50	322.75	161.38	80.69	40.34
	5	1280	1.0	1,032.81	516.40	258.20	129.10	64.55	32.28
	8	800	1.6	645.50	322.75	161.38	80.69	40.34	20.17
	10	640	2.0	516.40	258.20	129.10	64.55	32.28	16.14
	16	400	3.2	322.75	161.38	80.69	40.34	20.17	10.09
	20	320	4.0	258.20	129.10	64.55	32.28	16.14	8.07
	25	256	5.0	206.56	103.28	51.64	25.82	12.91	6.46
	32	200	6.4	161.38	80.69	40.34	20.17	10.09	5.04
	40	160	8.0	129.10	64.55	32.28	16.14	8.07	4.03
	50	128	10.0	103.28	51.64	25.82	12.91	6.46	3.23
	64	100	12.8	80.69	40.34	20.17	10.09	5.04	2.52
	80	80	16.0	64.55	32.28	16.14	8.07	4.03	2.02
	100	64	20.0	51.64	25.82	12.91	6.46	3.23	1.61
	128	50	25.6	40.34	20.17	10.09	5.04	2.52	1.26
	160	40	32.0	32.28	16.14	8.07	4.03	2.02	1.01
	200	32	40.0	25.82	12.91	6.46	3.23	1.61	0.81
	256	25	51.2	20.17	10.09	5.04	2.52	1.26	0.63
	320	20	64.0	16.14	8.07	4.03	2.02	1.01	0.50
	400	16	80.0	12.91	6.46	3.23	1.61	0.81	0.40
	640	10	128.0	8.07	4.03	2.02	1.01	0.50	0.25
	800	8	160.0	6.46	3.23	1.61	0.81	0.40	0.20
	1280	5	256.0	4.03	2.02	1.01	0.50	0.25	0.13
	1600	4	320.0	3.23	1.61	0.81	0.40	0.20	0.10
	3200	2	640.0	1.61	0.81	0.40	0.20	0.10	0.05
	6400	1	1,280.0	0.81	0.40	0.20	0.10	0.05	0.03

Legend:

n.nn Gbps	Gbps combined of all 24 mwax servers visibilities in this mode.
	Existing legacy correlator mode

Correlator File Naming Convention

Example: 1224396656_20181024T061040_ch201_001.fits

OBSID_YYYYMMDDThhnnss_chCCC_NNN.fits

where:

- OBSID = observation ID
- Start date / time of subobservation (UTC) - the date/time is provided by metafits file when the correlator is building the FITS file.
  - YYYY = 4 digit year
  - MM = 2 digit month (zero left padded)
  - DD = 2 digit day (zero left padded)
  - T = date/time separator
  - hh = hour (24 hr zero left padded)
  - nn = minute (zero left padded)
  - ss = second (zero left padded)
- CCC = Channel number (0-255)
- NNN = File number (0...n)

In the example above, the file represents the second file of the coarse channel number 201 for obs_id 1224396656.

FITS Format

The FITS files generated by the new correlator have the following structure:

- 1 Primary HDU
- Extension HDU (Visibilities for timestep/integration 0)
- Extension HDU (Weights for timestep/integration 0)
- Extension HDU (Visibilities for timestep/integration 1)
- Extension HDU (Weights for timestep/integration 1)
- ...

FITS Primary HDU

Keyword	(MWA) Valid Values	Change from v1 format?	Notes
SIMPLE	T		conforms to FITS standard
BITPIX	8		array data type
NAXIS	0		number of array dimensions
EXTEND	T		
COMMENT	FITS (Flexible Image Transport System) format is defined in 'Astronomy		
COMMENT	and Astrophysics', volume 376, page 359; bibcode: 2001A&A...376..359H		
CORR_VER	2	New	Correlator format version: <missing>= v1 (Ord / Legacy Correlator), 2 = v2 (MWAX Correlator)
COMMENT	Visibilities: 1 integration per HDU: [baseline][finechan][pol][r,i]	New	Description of data format for visibility HDUs
COMMENT	Weights: 1 integration per HDU: [baseline][pol][weight]	New	Description of data format for weight HDUs
MARKER	e.g. 0,1,2... n (where n is the final integration/timestep)		Data offset marker of first HDU (all channels should match). The marker increments by 1 per integration /timestep
TIME	e.g. 1531077040		Unix start time of the data in this file
MILLITIM	0-999		Milliseconds component of TIME
PROJID	e.g. G0000		MWA Project ID
OBSID	e.g. 1215112256		MWA Observation ID (GPS start time of observation)
FINECHAN	0.2, 0.4, 0.8, 1, 1.6, 2, 3.2, 4, 5, 6.4, 8, 10, 12.8, 16, 20, 25.6, 32, 40, 51.2, 64, 80, 128, 160, 256, 320, 640, 1280	New	Correlator mode: Fine channel width in kHz
NFINECHS	1-6400 e.g. 128 == 10 kHz fine channels (1280 kHz / 10 kHz)	New	Number of fine channels in each coarse channel
INTTIME	0.25, 0.5, 1.0, 2.0, 4.0, 8.0	Changed (was only 0.5, 1, 2)	Correlator mode: Integration time (s)

NINPUTS	(2-n) in increments of 16 (due to xGPU limitation) 256 == 128T 288 == 144T 512 == 256T	New	Number of inputs into the correlation products
CORRHOST	e.g. gpubox27	New (pseudo-hostname used to be in the filename)	Hostname of the correlator server which processed this coarse channel
CORRCHAN	0-N e.g. 0-23	New (was in filename as part of the gpubox number). Thus number is now in sky frequency order. e.g. CORRCHAN 20 > CORRCHAN 19 when mapped to corresponding sky frequency - e.g 172.8 MHz > 170 MHz	Coarse channel number selected for correlation.
MC_IP	v.x.y.z	New	Multicast IP the data was addressed to
MC_PORT	nnnnnn	New	Multicast port the data was addressed to

FITS Extension Visibility ImgHDU

Visibility Header

This HDU represents the visibilities for one coarse channel for one time integration / time step.

Keyword	(MWA) Valid Values	Change from v1 format?	Notes
XTENSION	IMAGE		Image Extension created by MWA DataCapture
BITPIX	-32	Changed: v1 Pre Oct 2014 = -32 (32 bit float) v1 Post Oct 2014= 32 (32 bit integer) v2 = -32 (32 bit float)	number of bits per data pixel (negative is floating point)
NAXIS	2		number of array dimensions
NAXIS1	(coarse channel width / fine_chan_width) * pols(xx,xy,yx,yy) * 2 For 10 kHz fine chan: (1280 / 10) * (4) * 2 == 1024	Changed: was: baselines * pols * 2	Fine Channels * Polarisations * 2 (real/imag)
NAXIS2	For 128T = (128*129)/2 = 8256 For 256T = (256*257)/2 = 32896	Changed: was: fine channels	Baselines
PCOUNT	0		number of group parameters
GCOUNT	0		number of groups
COMMENT	FITS (Flexible Image Transport System) format is defined in 'Astronomy		
COMMENT	and Astrophysics', volume 376, page 359; bibcode: 2001A&A...376..359H		
TIME	e.g. 1531077040		Unix start time of this HDU
MILLITIM	0-999		Milliseconds component of TIME
MARKER	e.g. 0,1,2... n (where n is the final time step)		Data offset marker of HDU (all channels should match). The marker increments by 1 per time step

NOTE: there is no BSCALE in the v2 correlator format.

Visibility Data

To simplify downstream processing, the correlator will reorder the data produced by xGPU so that the time step / integration is the slowest moving dimension, then within each time step: baseline (in a lower regular triangular order by antenna order from the metafits file) then fine channel (in ascending sky frequency) and polarisation of x*x, x*y, y*x, y*y, and finally for each polarisation there will be a real and imaginary floating point value. The ordering of data thus matches closely to the uvfits ordering scheme (citation needed!).

```
time | baseline | freq | pol | r,i
```

Below is an ***example*** using **4 antennas** (ant0, ant1, ant2, ant3) example, with **2 fine channels** (0,1) and **4 polarisations** (xx,xy,yx,yy) in **one time step** / integration (we only have time step per HDU). The ImgHDU data would like the following:

antA	antB	ch0.xx	ch0.xy	ch0.yx	ch0.yy	ch1.xx	ch1.xy	ch1.yx	ch1.yy
0	**0**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
0	**1**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
0	**2**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
0	**3**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
1	**1**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
1	**2**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
1	**3**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
2	**2**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
2	**3**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i
3	**3**	r, i	r, i	r, i	r, i	r, i	r, i	r, i	r, i

FITS Extension Weights ImgHDU

Weights Header

This HDU represents the weights for one coarse channel for one time integration / time step.

Keyword	(MWA) Valid Values	Notes
`XTENSION`	IMAGE	Image Extension created by MWA DataCapture
`BITPIX`	-32	number of bits per data pixel (negative is floating point)
`NAXIS`	2	number of array dimensions
`NAXIS1`	pols(xx,xy,yx,yy) == 4	Polarisations
`NAXIS2`	`For 128T = (128*129)/2 = 8256` `For 256T = (256*257)/2 = 32896`	Baselines
`PCOUNT`	`0`	number of group parameters
`GCOUNT`	`0`	number of groups
`COMMENT`	`FITS (Flexible Image Transport System) format is defined in 'Astronomy`	
`COMMENT`	`and Astrophysics', volume 376, page 359; bibcode: 2001A&A...376..359H`	
`TIME`	`e.g. 1531077040`	Unix start time of this HDU
`MILLITIM`	0-999	Milliseconds since TIME
`MARKER`	e.g. 0,1,2... n (where n is the final time step)	Data offset marker of HDU (all channels should match). The marker increments by 1 per time step

Weights Data

As of writing the weights HDU is filled with zeros. In the near future, proper weight values will be supplied from the MWAX correlator.

Below is an example using 4 antenna (ant0, ant1, ant2, ant3) example and 4 polarisations (xx,xy,yx,yy) in one time step / integration (we only have time step per HDU). The ImgHDU data for the weights would like the following:

	weight.xx	weight.xy	weight.yx	weight.yy
ant0,ant0	w	w	w	w
ant0,ant1	w	w	w	w
ant0,ant2	w	w	w	w
ant0,ant3	w	w	w	w
ant1,ant1	w	w	w	w
ant1,ant2	w	w	w	w
ant1,ant3	w	w	w	w
ant2,ant2	w	w	w	w
ant2,ant3	w	w	w	w
ant3,ant3	w	w	w	w