JIVE UniBoard Correlator Memo 8

Some considerations regarding the design of the UniBoard correlator August 12, 2012

Arpad Szomoru

1. Introduction

This document summarises the current state of the UniBoard VLBI correlator being developed at JIVE and is primarily intended as a discussion piece. It attempts to show the limits of this correlator, imposed by the available resources on the FPGAs and the complexity of the design. Some trade-offs are discussed that could be made to reach higher spectral and/or time resolution, needed for wide-field VLBI. The resolutions are compared to what one might call "reasonable" resolutions, and the question is posed whether one would need to go beyond these reasonable values.

2. Current design

- 32 stations
- 64 MHz in 16 MHz bands
- 512 frequency points (will be 1024) per sub-band
- frequency resolution 31.5 kHz (will be 15.625 kHz)
- min integration time ~22 ms (based on all products, one 10GE output per node)
- max integration time 1 s (limited by resources used by corner turning on back node)
- 2 bit data (only important for data rates into correlator, internally 18 bits representation)

Increasing the number of frequency points from 512 to 1024 is being worked on right now. To prevent confusion, in the following I will refer to the design with 1024 frequency points and 15.6 kHz resolution as the current design.

Note that the minimum integration time is determined by the speed at which products can be sent out of the board, based on writing out *all* products. In practice we will of course not observe with 32 stations at any time soon. Right now one 10GE output is used per back node, but it would be possible to further decrease the minimum integration time by using all three available outputs.

The delay module is integrated in the design, separate bits are being tested, but not fully functional yet. Making it functional has been the highest priority of all three JIVE engineers during the past weeks.

Work on a pulsar gating module is starting about now as the work on the delay module is nearly completed. Luo will write up what resources pulsar gating will require on the front nodes.

VDIF packets at the input side have to have a specific size in this design. This limitation will be removed at a later time.

The exact way to continuously feed data and delay coefficients to the correlator and how to handle missing/invalid data has been extensively discussed. The current plan is to have an (external) control computer handle the synchronization between correlator and data/delay coefficient sources, through the polling of a "memory full / memory empty" condition stored in a FIFO on the UniBoard. This would allow the correlator to run at the fastest possible speed, limited only by the slowest data source (and the speed of the correlator itself of course). It would also solve any problems caused by invalid or missing data in the data streams.

Corner turning on the back node uses up a lot of the resources, and has to make use of DDR3 memory, which is (relatively) very slow which means it is very hard to meet timing. Fixes/modifications of the code require many iterations (because of this tight timing) to test different routings and lay-outs, which is very time consuming. This part of the design certainly seems to be very much on the bleeding edge. Ways are being considered to reduce the amount of memory used, which would ease some of the constraints.

J2ms2 will need quite some work still to be capable of handling UniBoard data.

3. Different personalities

Already at the onset of the project we considered having two separate personalities for spectral line and continuum work, or reserving one UniBoard as a spectral line correlator, with the same personality but limited input bandwidth.

4. Continuum only

It would be quite feasible and in a way attractive to make a low-resolution personality to exclusively handle continuum observations. Going from 1024 to 32/64 spectral points would actually remove the need for corner turning in DDR3 memory (in the back nodes), as all multiplying/accumulation could take place using on-chip memory only. With 32 spectral points, it might even be possible to fit two complete correlator engines on the FPGA, doubling the capacity of the correlator to 32 stations at 128 MHz. The 1-second maximum integration time limit would also disappear, enabling (nearly) arbitrarily long integration times. The 20-ms minimum integration time would go down by a factor of 16-32.

32 stations		
Input BW (MHz)	Δν (kHz)	Int. time (ms)
64 (maybe 128)	500.0	0.7
64	250.0	1.4

Table 1: Spectral resolution and minimum integration time for continuum-only personality

5. Spectral Line

The assumption here is that observers will want a continuum map covering as much bandwidth as available as well as a spectral line cube at the highest resolution. One option would be to send part of the frequency bins (after the FFT on the front node of the continuum correlator) back out via the front node 10GE connections and into a second UniBoard for additional filtering. Easier would be simply to send a copy of the sub-band of interest directly to a spectral line correlator.

This spectral line correlator could be implemented in the following ways:

1. A single sub-band (2-32 MHz) would be processed through a PFB, keeping the same number of FFT points as in the current design. Increasing the size of the FFT would be quite hard due to limited on-chip memory. Without any change to the FFT size, the frequency resolution would thus become 0.5 kHz (for a 2 MHz band) to 8 kHz (32 MHz band). After this one could just select the relevant frequency bins and send them to the back nodes for correlation or send all bins to the back nodes. The back node design will then depend on the number of bins transferred to it, either the current design using DDR3 memory for corner turning or the simpler continuum-only design. Transferring all bins to the back nodes while correlating in real time, the number of spectral bins to correlate per time unit will scale linearly with the total bandwidth processed, which means that the minimum integration time will be smaller for smaller bandwidths.
2. The full sub-band bandwidth could be filtered through a tunable FIR, followed by an FFT. This design could result in nearly arbitrarily high frequency resolution.
3. Spectral line processing could simply be offloaded to the SFXC

It is not yet quite clear which of the two first options would take up most resources, as a tunable FIR is fairly expensive; Hargreaves will make an estimate.

Table 2 summarizes the first option.

32 stations		
Input (MHz)	Δv (kHz)	Int. time (ms)
64	15.6	22
32	7.8	11
16	3.9	5.5
2	0.1	0.7

Table 2: Spectral resolution and minimum integration time (option 1)

6. The tricky bit

Unfortunately, the processing of multiple phase centers requires both high frequency and time resolution, as well as the full continuum bandwidth. For this we may need to trade off resolution and number of stations and bandwidth per board.

Reducing the input bandwidth will increase the frequency resolution but obviously increase the number of UniBoards needed to process the data. Time resolution will improve as explained in the previous section.

Reducing the number of stations by half will give an additional factor of 2 in frequency resolution. This will result in a doubling of the number of bins to correlate, and as a consequence a doubling of the minimum integration time. Fortunately this is offset by the fact that there will be 4 times fewer correlation products to calculate.

Note that it is not likely that there will be sufficient resources on the back nodes to do the phase shifting. One solution would be to dump the samples at very high frequency and time resolution onto another UniBoard, with yet another personality, or do the final processing in software, for example on the SFXC.

In table 3 the resolutions are listed that can be reached. For comparison, the resolutions needed to get to less than 10% decrease of response at the FWHM of a 100m dish (Wrobels formulae) in L and C band are listed in the table 4. I also list the same numbers divided by 10, which are currently used as a limit at JIVE in order to completely rule out any contribution by smearing. In table 5 the distance from the center of the field to the FWHM (of a 100m dish) is listed.

Input (MHz)	Δν (kHz)	Int. time (ms)	# boards needed (1Gbps/station)
32 stations			
64	15.6	22	2
32	7.8	11	4
16	3.9	5.5	8
16 stations			
64	7.8	11	2
32	3.9	5.5	4
16	2.0	2.8	8

Table 3: Spectral resolution and minimum integration time for multiple phase center modes

Baseline (km)	lambda (cm)	Int. Time (ms)	Δν (kHz)	Int. Time (ms)	Δν (kHz)
		10%	10%	1%	1%
2000	18	900	133	90	13
2000	6	900	399	90	40
8000	18	225	33	22.5	3
8000	6	225	100	22.5	10

Table 4: resolutions needed for no more than 10% decrease in response at FWHM of 100 m dish, with a (crude) extrapolation to 1% decrease

lambda (cm)	18	6
FWHM/2 (")	240	80

Table 5: FWHM/2 of a 100m dish, in seconds of arc

7. Output data rates

All these minimum integration times are based on a 10Gbps output data rate per back node. This means that for example for multiple phase centre correlation at the highest time resolution the output of one board (assuming 32 stations, 8 hours observation) would amount to 136 TiB.

Under the same assumptions, in continuum mode with a "normal" integration time of 1 sec, one board will produce ~3 TiB output data. Assuming 16 stations, same bandwidth and resolution, lowers this to 0.75 TiB.

8. Conclusion, way forward

The advantages of making a low-resolution continuum correlator personality are that the design will become simpler and that it may be possible to process a full 128MHz bandwidth from 32 stations on one UniBoard. However, it would be quite a big change from the current design and what is more, we still need the current design for multiple phase center work.

The spectral line personality with tunable FIR (option 2) would give most flexibility but would also imply a fairly drastic change in design. Option 1, the current design, provides less flexibility but enough resolution for most purposes. The SFXC correlator could handle more extreme resolutions. As spectral line observations will always be narrow band anyway, maybe we should consider simply doing all spectral line work on the SFXC.

The multiple phase center personality, basically the current design, seems to provide enough resolution for "reasonable" wide-field work. By using more boards and/or fewer stations unreasonable demands can be dealt with. Considering the amount of data produced at the highest time resolution, dumping the intermediate products to disk and doing the phase shifting and integration completely off-line does not seem feasible. Note that after the application of the phase corrections the data sets are integrated in time and down-sampled in frequency, which brings about a tremendous reduction of data volume; the highest spectral- and time resolutions are only needed up until the application of the phase corrections. With this in mind, the optimal solution would obviously be to do all processing including the down sampling on the back node.

Note that I have calculated the resolutions needed to have no more than a 10% and 1% decrease at the FWHM of the Ef beam, which presumably dominates the primary beam shape of the EVN. No doubt, some people will want to use phase centers even further removed. One might wonder if 1% is not complete overkill, considering how poorly the primary beam shape is known anyway. One should also beware of over-complicating the correlator design; an FPGA-based correlator will never be as flexible as a software correlator.

In conclusion:

1. After the summer holidays, a detailed work plan will be made listing what parts of the firmware, the control and the offline software still are incomplete, with estimates of effort involved and a time line leading to full operability at the end of this year
2. We will stick to the current correlator design
3. Spectral observations will be offloaded onto the (expanded) SFXC
4. Pulsar gating will be implemented by Luo during his remaining months at JIVE
5. Only then, work will start on the multiple-phase center capability. This will involve (in order of preference)
    a. finding a way to make phase shifting on the back node possible after all by making the current design more efficient
    b. creating a completely new, but fairly simple, personality for the UniBoard
    c. writing new software (basically a stripped-down version of SFXC)
6. At the lowest priority, a spectral line personality with tunable FIR could be developed