Meeting Notes
Correlator Face-to-Face Meeting
April 3rd & 4th
Socorro, New Mexico

path: /users/bsahr/EVLA/WIDAR_correlator/face-to-face_Apr2006
/f2fmeeting_notes.doc

General

The final version of these notes, and the notes for the Data Products Meeting of Wed, 4/5/2006, will be posted on the Correlator Face-to-Face web page:

http://www.aoc.nrao.edu/evla/geninfo/meetings/EVLA-Corr-FtoF.html

EVLA Overall Architecture (Butler)

There was a discussion of how best to support intermediate level users. This level of user is seen as somewhat problematic because neither templates nor hand-composed scripts are entirely appropriate for this level of expertise. Reference was made to an early document produced by Brent – NRC-EVLA Memo #012 Concept for an "Observation Builder" for Array and Correlator Configuration, 20Feb2001, B. Carlson available on the DRAO web site at: http://www.drao-ofr.hia-iha.nrc-cnrc.gc.ca/science/widar/private/Memos.html. The modification and validation of templates was discussed as a possible approach with

others responding that they hoped for something better. The need to support whatever approach is developed at the Proposal Preparation level was mentioned. In this context, the Correlator Resource Manager to be developed by Sonja was presented as a possible, partial solution. If written to be portable with respect to its runtime context an environment, the Correlator Resource Manager could be included in the observation preparation phase to serve as a tool to validate proposed correlator configurations and resource requirements. While not a complete solution, as there is no knowledge at this stage of observation preparation of downstream scheduling issues that might lead to resource conflicts, the ability to use the Correlator Resource Manager in this context might be very useful.

Barry raised the issue of the Science Data Model (SDM). Is the SDM an adequate and sufficient description of the correlator setup? A level of organization above that of VCI configuration commands is needed to describe correlator setup. Can the SDM be used to satisfy this need? Does it (the SDM) have what is needed to connect correlator setup & configuration to intention? Does the SDM support all of the possible WIDAR observing modes? Does it support it in a manner that is natural? Can the SDM translate to correlator observing modes and correlator setup?

Correlator Software Overview (Vrcic)

Reference was made to A25200N0001 Correlator Software Architecture [RFS] 17Feb2004 S. Vrcic/B. Rowen, on:
http://www.drao-ofr.hia-iha.nrc-cnrc.gc.ca/science/widar/private/Software.html
as once source for the material in the presentation.

The CPCC is conceived as a 1+1 system, high reliability, possibly with a hot rollover capability. Given the digital I/O needed by the CPCC, hot rollover may not only be quite expensive, but, perhaps, difficult or impossible to achieve. There seems to be an implication that the 1+1 capability will not duplicate the digital I/O. This point should be clarified. The MCCC is conceived of as a 1+1 system, with high reliability features. For the MCCC, a hot rollover capability is being considered. The CBE master node should also be a 1+1 system, with high reliability features and hot rollover capability if possible.

Data products, outstanding issues (see also the notes for the Data Products meeting of 4/5)

- CBE outputs & format not yet defined
- Radar mode output & format not yet defined
- There are a number of other Station board data products for which formats have not been defined

There was some discussion of correlator reconfiguration times w.r.t. the issue of switching between narrow band spectral line observing and wideband observations of a calibrator. The point was made that the CBE must also be reconfigured. It was concluded that this issue could be handled by sending the needed configurations ahead of time as switching between pre-loaded configurations should be very fast. However, care must be taken that the CMIB & MCCC software support the pre-loading of multiple configurations, and that

the MCC software be able to decompose multiple configurations and send out the needed configurations to the CMIBs.

A question that was raised and answered in the affirmative – there will be no phase changes when going from one correlator configuration to another, such as broadband to narrow band and back, or for different subband bandwidths.

Virtual Correlator Interface Overview (Vrcic)

The VCI is meant to be a machine-to-machine interface.

Models

It was during this presentation that it became clear that DRAO conceptualized models as NOT going through the VCI/MCCC software. I will not attempt to reproduce the discussion of this point. *Suffice it to say that it was concluded, in the end, that the VCI will be the path or channel of communication for models.* The point was made that if this additional traffic degrades VCI/MCCC responsiveness to an unacceptable level that the hardware platform running the VCI/MCCC software can be upgraded as needed – multiple processors, multiple core processors, or even multiple systems.

And, taking things a bit out of order, it was further agreed, w.r.t. to models that

- *Tone extractor models will be treated as a configuration (frequencies) item and not as data models that are to be updated in realtime*
- *It was thought that the same approach could and should be applied to pulsar models – treated as a configuration item rather than as a realtime model subject to periodic updates. At some point, Michael Rupen agreed to make further investigations of this issue, but the matter seemed to be settled during later discussion – pulsar models are not needed in realtime, they can be sent as configuration data, but the configuration must be updated on a timescale of once every "few" (< 10) minutes.*

Data Inputs & Outputs

Slide 4 of this presentation, which listed a number of input and output data items led to the statements that

- Where there is data to be archived, NRAO is responsible for archiving it
- That how the data is made available for archiving should fit into our (EVLA M&C) general scheme – multicast, compatible with an archive filler that needs to understand something about data format, but nothing about data content
- That a common schema should be established for antenna monitor data and correlator monitor data

The issue of Station Board and CBE data outputs is treated at greater length in the notes for the Data Products Meeting of Wed, 4/5/2006. One decision that came out of both the discussion during the presentations and during the Data Products meeting is that radar mode data is not suitable for the SDM and it may not be well suited to an extended monitor data format. It may require a special format and storage in flat files rather than the monitor data archive.

Alerts

The CMIBs send all alerts to a unix log server. Issues and questions:

- Will this log server be inside or outside the VCI ?
- Will there be a log server behind the VCI that the correlator uses to store its alerts, which are then passed along to a system wide log server?

In general, the issue seems to be one of compatibility between EVLA M&C alerts and correlator alerts, and the merging of all alerts into one system wide log.

State Counts to the CBE, Realtime Distribution of Data

It has been stated that the CBE needs state counts for one type of correction it must make (requantizer corrections?). While this point is known and acknowledged, nothing has been done to specify either how the state counts will get to the CBE, or the timeliness with which they are needed.

- The discussion of state counts raised a more general point – for information that is needed in quasi-realtime, such as Station Board WideBand Correlations, do we plan to get it from the archive or via some more immediate mechanism, such as multicast. The sense of the meeting seemed to be that 1) the archive is not timely enough, and 2) it would be sound engineering to design a system whose realtime functionality does not depend on the archive as a data source.

Having made these statements, the output items listed on slide 4, excepting visibility data, become candidates for multicast – Station Board Wideband Correlator Products, raw filter output (a.k.a radar mode data), meta-data of various types, alarms, logs, status reports. Some of these data items will be required, in realtime by multiple components of the overall system.

Transmission of Correlator Configurations

NRAO has the responsibility for defining the specific mechanism by which configurations will be transmitted to the correlator by the EVLA Monitor and Control System. This mechanism must allow for realtime correlation coexisting with VLBI playback correlation.

Schedule for the VCI, OTS testing - Models to the Prototype Correlator

It was agreed that a prototype version of the VCI should be ready in time for on-the-sky (OTS) testing of the prototype correlator. Further, it was agreed that this prototype version of the VCI should support transmission of models from the EVLA Monitor and Control System to the prototype correlator. The issue of exactly how models will be communicated to the prototype VCI is TBD. Will the models be cast in the form of XML messages and sent as datagrams (UDP)? A development plan that presents both a schedule and the state of the VCI at the time of OTS testing is needed.

In his comments on these notes (email of 4/18/2006) Brent takes issue with the statement that "a prototype version of the VCI should be ready in time for OTS testing". He disagrees with this statement and goes on to suggest that perhaps the statement should be clarified by adding a sentence stating that the prototype VCI will ***only*** *support transmission of models from the EVLA M&C system to the prototype correlator. Ken Sowinski suggests a somewhat more positive approach – a statement that the VCI should* ***at least*** *support the transmission of models.*

The delivery of a prototype VCI with the prototype correlator was a point of contention at the Correlator face-to-face meeting and, apparently, it continues to be. It is my (Bill Sahr, the author of these notes) opinion that it was clearly stated during the discussion of this matter that NRAO considers the delivery of a prototype VCI with the prototype correlator to be essential. Speaking in my role as Monitor and Control Group Leader, I am strongly opposed to any effort to limit the delivered prototype VCI functionality to only the transmission of models. The EVLA Monitor and Control effort not only welcomes, but needs as much functionality as possible in the prototype VCI in order to advance its efforts to more fully integrate the correlator into the EVLA Monitor and Control System, and to test the software that will be used to configure, update, and otherwise operate the correlator. I would go much further that just delivery of a prototype VCI with the prototype correlator, arguing for agreement upon a schedule of staged delivery of increasingly mature versions of the VCI that NRAO can run on a nascent MCCC and test against the prototype correlator.

Output Data Products & Formats (Butler/Vrcic/McMullin/Pisano)

To begin at the end, with some of the agreed upon decisions and work assignments:

1. *Martin Pokorny & Brent agreed to undertake the task of determining the format for the visibility data to be written by the CBE for OTS testing – FITS or SDM. The issue of raw lags and lag sets written by the CBE for prototype board testing is addressed in the Data Products Meeting Notes and is also assigned to Martin Pokorny.*
   In his comments on these notes (email of 4/18/2006, Brent Carlson states: "I don't recall from the meeting my name being associated with this task. I think Martin has to take the lead on it, with me in a consultative roll if necessary. Deciding what data format is to be used for proto corr testing largely depends on the image processing software available to support that format. Right now, AIPS is available with UVFITS and the CBE can now write UVFITS, even though it currently does not support the full flexibility of the correlator. Whether ASDM and software to process it will be ready is certainly beyond my scope or knowledge."
2. *Michael Rupen & Martin Pokorny will determine if AIPS can read FITS as produced by the CBE.*
3. *Michael Rupen and Brent will examine the (A)SDM and make a determination of the degree to which it will support the correlator and the extensions that might be needed*
4. *Bryan Butler will work with Joe McMullin to determine the status of XML schemas for the SDM. Peter Dewdney adds that we need to be sure the needed schemas are produced, and we need to know the schedule for production of the schemas*
5. *Michael Rupen will confirm that it is possible to transform SDM formatted data to FITS formatted data via CASA. (I.e. – SDM into CASA, FITS out from CASA, FITS into AIPS.)*

To summarize points made & questions raised during the presentation and discussion:

- The byte order for the ALMA binary output data stream is a header item
- There is a hardwired axes order. Peter Dewdney opined that DRAO may need to change that order
- Brent presented a long list of items with the question of whether or not the SDM can handle them. The result was item 3 in the list given above.
- Michael Rupen asked if data was stored only after stitching the subbands. The answer was yes, but there is a mode for storing unstitched subbands.
- Joe McMullin made a clear statement that the ASDM is not adequate for all EVLA modes. Ed Fomalont & John Benson wrote a document describing the needed additions
- An analysis of CBE throughput that includes the requirement to reorganize the data before archiving it is needed.
- F. Villefond has a postdoc working on XML schemas for the ASDM
- Is UV FITS sufficient for prototype board testing and to validate the prototype correlator hardware? Brent's opinion was that for the most part yes, but the FITS format has no place for several items, including phase binning and subbands with different bandwidths. Eric Greisen volunteered that changes could be made. (See items 1, 2, & 5 above.)
- If the SDM were used for OTS testing, is CASA sufficiently well developed to offer the needed analysis capabilities?
- For ALMA, each subband will have the same integration time.
- Can we use the CASA software to read SDM formatted data, write it out in FITS format and then use AIPS to analyze the data (see item 5 above).

Prototype Board Testing (Ryan/Vrcic)

A CMIB server will be needed for OTS testing and for the limited and full versions of the WIDAR correlator. The CMIB root file systems reside on the CMIB server and are NFS mounted at boot time. Probably two to four rack-mounted Dell server class systems will suffice as the CMIB server.

Possible issues and questions

- It is assumed that the CMIB server will be located inside the correlator room and on the same correlator network that connects the CMIBs to the MCCC
- Has anyone really looked into the issue of 300 CMIBS trying to boot off the CMIB server, pretty much at the same time? Is this approach a prescription for a long period of network collisions and very slow boot times? Do we need to consider an alternative approach?

It was agreed to eliminate the VPN approach to NRAO <-> DRAO systems access, and, instead, to make the AOC & DRAO routers look internal to one another. (Assigned to James Robnett & Tony Hoffman)

Slide 4 of Sonja's portion of this presentation presents a proposal for a Correlator Top Level GUI.

Slide 8 presents a diagram of a proposed test configuration for OTS testing.

Slide 9 presents a high level GUI that is planned for use with the prototype correlator during OTS testing. During the discussion of this GUI, the statement was made that no VCI would be available for OTS testing. This decision was based on the incorrect impression that NRAO would not have an EVLA M&C System ready in time for OTS testing. As per statements made earlier, in the section of this document given over to the VCI presentation, it was finally agreed that

- *A prototype version of the VCI will be ready in time for on-the-sky (OTS) testing of the prototype correlator. Further, it was agreed that this prototype version of the VCI will support transmission of models from the EVLA Monitor and Control System to the prototype correlator. The issue of exactly how models will be communicated to the prototype VCI is TBD. Will the models be cast in the form of XML messages and sent as datagrams (UDP)? A development plan that presents both a schedule and the state of the VCI at the time of OTS testing is needed. DRAO is to supply this plan.*

Additionally,

- *It was recognized that DRAO will need some subset of the EVLA M&C infrastructure to develop and test the VCI. At some point we must supply DRAO with the necessary software, installation & configuration assistance, and training in the use, care, and feeding of the software. No mention was made of who will supply the computing platform(s) needed to run the software.*
- *Further, it is the responsibility of NRAO to supply the functionality implied by the model server shown in the test configuration diagram given in slide 8.*

And, at some point during the discussion it was noted that there will be three prototype correlators – one at the VLA site, one at Jodrell Bank, and some portion or subset of one at DRAO.

The Data Reorganizer or Fast Data Store (a.k.a. The Fast Formatter)

Bryan Butler presented a diagram that postulates an addition to the CBE for the purpose of reorganizing the data coming from the individual CBE nodes into a single binary data stream – something that serves the same purpose as the master node of the ALMA CDP. This issue has been raised before. These notes present a reproduction of Bryan's diagram, and add a diagram developed in August 2003 by Boyd Waters and Bill Sahr that places the notion of a binary data stream into the context of data distribution for realtime and quasi-realtime requirements.

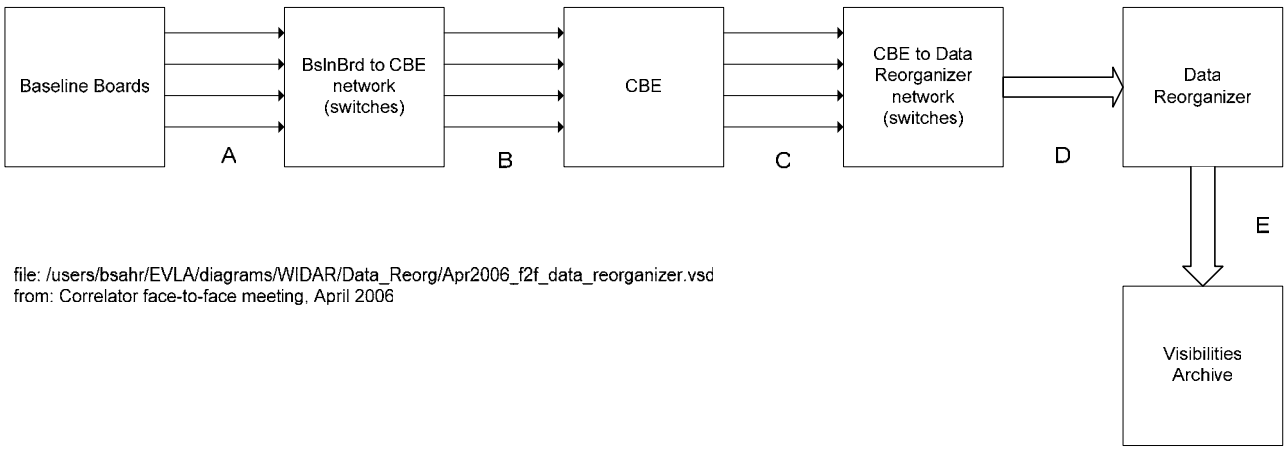

Figure 1 Data Reorganizer

Lags flow from the WIDAR Correlator Baseline Boards to the CBE via a set of network switches (points A & B). Binary data streams flow from the CBE nodes to yet another set of switches (C) and then converge upon the Data Reorganizer (D). The links at A and B are 1 Gbit/sec, at C and D they are currently unspecified. Finally, data flows from the Data Reorganizer to the Archive. The specification for data into the archive is a peak rate of 25 Mbytes/sec, with the expectation that the typical data rate will be lower.

There is an obvious, possible bottleneck at D. There are two factors to consider w.r.t. this possible bottleneck. First, less data flows out of the CBE (point C) than into it, and the Data Reorganizer need not be a single system. Some sort of binary cascade as per figure 2 could be configured, if it were needed. Of course, the complexity and cost represented by the Cascaded Data Reorganizer is to be avoided if at all possible.

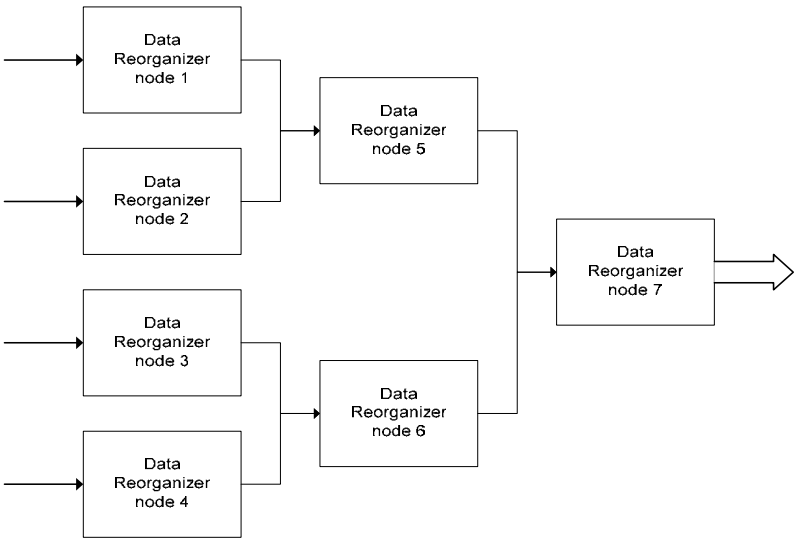

Figure 2 Data Reorganizer Cascade

Figure 3, Fast Data Store and Formatter, was an earlier attempt both to show the need for processing power between the CBE and the archive for the purpose of reorganizing the

CBE data stream, and to begin to address the issue of quasi-realtime distribution of the data coming from the CBE. At an early stage in the design of the EVLA software, the question was asked – are the timing requirements associated with the Real-Time Calibrator Analysis Tool (RTCAT, now renamed to Telescope Calibration or TelCal) and some of the post processing elements such that visibility data can be obtained from the archive, or must the data be made available in a more timely fashion? Figure 3 does not really answer that question, but is a step in the direction of doing so. TelCal's (RTCAT) needs are the most stringent. Figure 3 posits a shared memory connection – multiple processes in a single or dual CPU box, or, perhaps, multiple processor boards sharing a backplane? The specific nature of the connection is not really specified. Also left

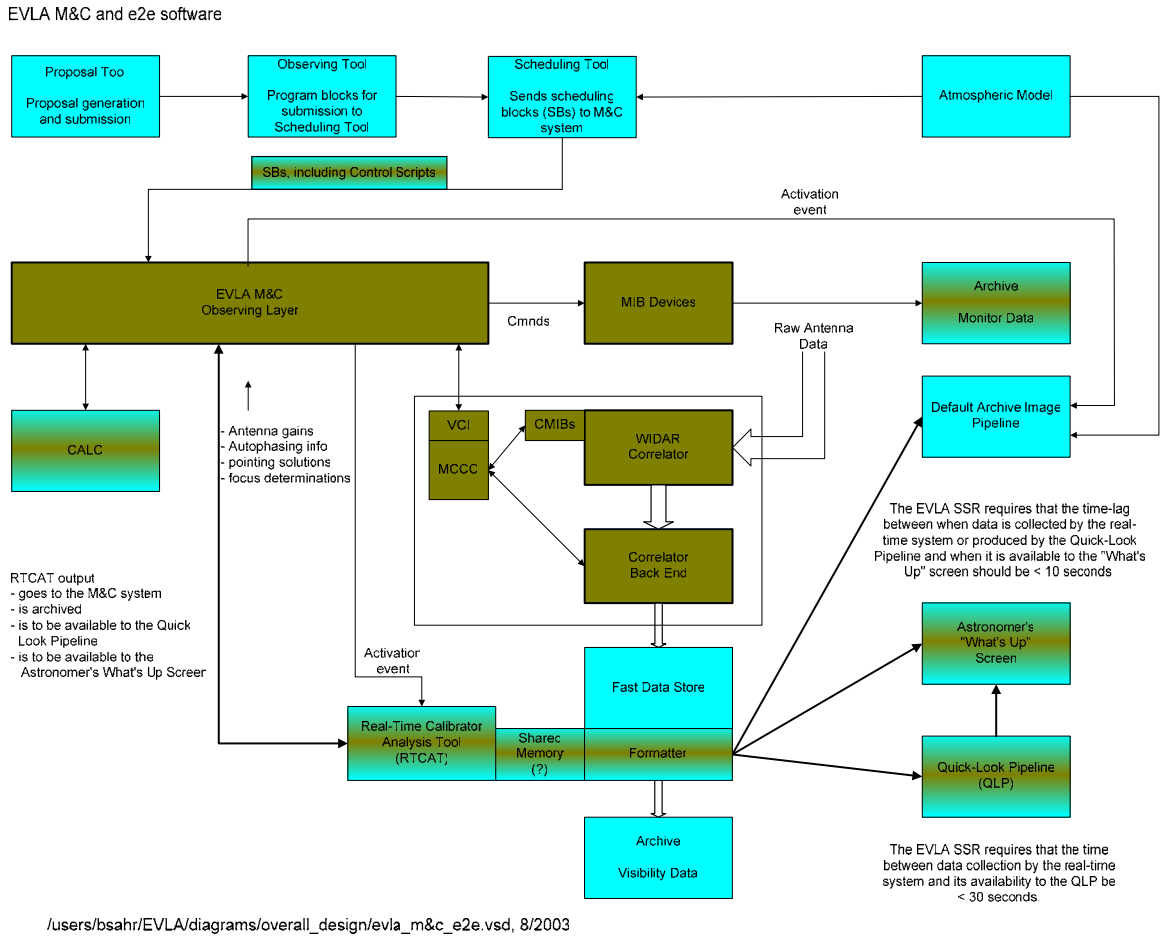

Figure 3 Fast Data Store & Formatter

unspecified are the connections from the Formatter to the Quick-Look Pipeline, the "What's Up" screen (now renamed to the Observation Monitor) and the Default Archive Image Pipeline - perhaps a network connection, possibly implying another switch? As with the Cascaded Data Reorganizer, the extra cost and complexity represented by figure 3 is to be avoided if at all possible. TelCal may require an intimate and high speed connection to a Data Reorganizer or Fast Store/Formatter. It may be possible for the

Quick-Look Pipeline, the "What's Up" screen, and the Default Archive Image Pipeline to obtain their data from the Archive. All of these issues are currently unresolved.

For the sake of simplicity, these notes will hereafter refer to the Data Reorganizer or the Fast Data Store as the "Fast Formatter". The term seems to capture the essence of what it must do.

The form of the Fast Formatter should not be taken as "set". The matter requires investigation. It was suggested that peer-to-peer software may provide a model for portions of a solution.

Other issues raised w.r.t. the Fast Formatter:

- *Who will work on the software? Likely answer – Martin Pokorny.*
- *Who will pay for the cost of the network connecting the CBE nodes to the Fast Formatter, and who will pay for the Fast Formatter hardware? To be negotiated.*

Real-Time Data Display (Del Rizzo)

There are no issues to discuss here that are not covered in more detail and depth in the notes for the Data Products meeting of Wed, 4/5/2006. Those notes have been posted on the Correlator Face-to-Face web page:

http://www.aoc.nrao.edu/evla/geninfo/meetings/EVLA-Corr-FtoF.html

Baseline Board to CBE Connection (Robnett/Rowen)

The maximum output rate of a Baseline Board was given as 100K lag frames per second, which is taken to be 1 Gbits/second. This works out to an integration time of approximately 11 milliseconds.

The network as currently conceived consists of ten 24 port switches, 16 baseline boards per switch, which allows upto 8 CBE nodes per switch, although the DRAO budget for the CBE is currently based on the assumption of a total of 50 nodes for the CBE.

Software will be needed, perhaps running on the MCCC, that tells the baseline boards where to send the lags.

The possibility of setting the MTU of the Baseline Board to CBE nodes network high enough to pack > 1 lag frame per packet was discussed. This configuration change could be made in the Baseline Board FPGA, or, perhaps, even in the switches & the CBE NICs with no change to the Baseline Board FPGA.

That 50 nodes will be sufficient for the CBE is not really known. Neither has much thought been given to the issue of how swapping in a spare node for a failed node will be handled.

Prototype Correlator Testing, Including On-The-Sky (OTS) Testing (Carlson)

Delivery of the prototype correlator chips may slip to early June because of packaging issues.

Six (6) months is being allowed for prototype chip and board testing in Penticton.

Six (6) months is being allocated to get the prototype correlator ready for OTS testing.

Stage 3 production = the "Limited" Edition correlator, consisting of 16 station boards and 16 baseline boards – 2 full racks.

To where (what archive) will the prototype correlator OTS test records be written, and how will this data be moved from the VLA site to the AOC? And, as a related issue, Michael Rupen is to do an estimate of the output to be expected from the prototype correlator.

Brent Carlson and Michael Rupen are to develop a list of the software that will be needed for the prototype correlator, with Sonja to be kept closely informed. Brent and Michael are good candidates for evaluating the appropriateness of the SDM as a format for prototype correlator output.

In his comments on these notes (email of 4/18/2006), Brent stated that he will upgrade the OTS test plan to include a section on software requirements and, at Sonja's suggestion, top-level GUI screen shots.

A side issue

What is the relationship of the VCI software to the CBE software? How will monitor and control for the CBE be handled? Will the CBE communicate with the world beyond the correlator via the VCI? There is a need to coordinate the schedule for VCI development with the schedule for CBE software development.

Station Board Testing (Fort)

No issues raised or decisions taken that seem appropriate to these notes.

Programmer's Guide to Timing, Synchronization, etc (Carlson)

Slide 49 contains what might be one of the better lists of Station board data products.

Station Board Real-Time Software (Rowen)

An RFS document for the realtime software for the Station Board and a manual for the Module Access Handlers are needed. Bruce Rowen is to write these documents.

A period of testing of the prototype test software, prior to the start of the actual prototype tests is needed. This requirement is not really called out in any of the written planning or schedules.

Miscellaneous Items & Issues

Another software face-to-face meeting is being planned – tentatively set for September 2006 in Penticton.