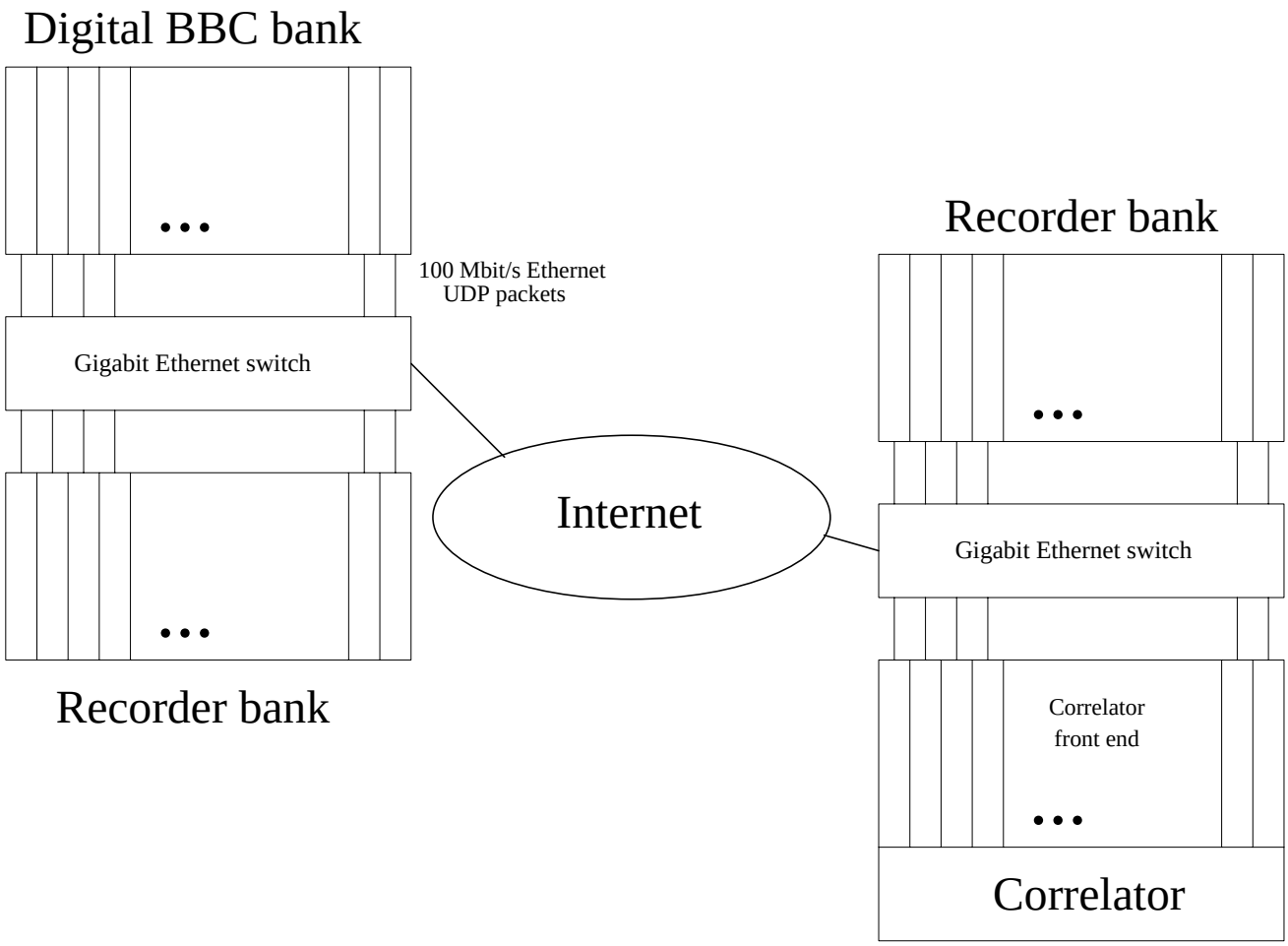
Concept for Next Generation VLBI

Jouko Ritakari, Metsähovi Radio Observatory.
October 31, 2000

Because of popular demand I will write down my thoughts on the future of VLBI technology, first presented in the EVN TOG meeting in Torun, Poland on 20th of October, 2000.

This document describes the concepts needed in the design of the next-generation VLBI system.
In some cases I had to go to some detail to prove these things are feasible with the technology we can buy off-the–shelf now. If these are too tedious, please jump over the explanations and concentrate on the underlying philosophy.

I wrote this document from the VLBI point of view, but many of the concepts are valid also in connected arrays.



Background

A great deal of the complexity and high cost of VLBI systems is caused by old and unreliable recorder technology, which is not very well suited to the demands of VLBI. Maximum data rate for one track is 9 Mbit/s, so VLBI data needs to be divided into small 9 Mbit/s data streams.

Another cause for complexity is the need to shuffle these datastreams to the available tracks in the recording terminal and from the recorder tracks to the correlator channels in the playback terminal.

Both of these problems can be solved by using commercial-off-the-shelf (COTS) technology that is available now.

Guidelines for the next generation VLBI

- Commercial-off-the-shelf (COTS) components will be used wherever possible.
- Only commercially wide-spread technology should be used. The technology must have a clear upgrade path.
- The system must be easily mass produced and affordable.
- The system must be scalable. Baseband converters and recorder units can be easily added to the system if necessary.
- The system must be easily upgradable. For example, the recording technology needs to be upgraded several times during the lifetime of the system.
- The system must support near-realtime fringe checking using Internet.
- The system must support realtime VLBI, when bandwidth becomes affordable and VLBI migrates from tape recording to Internet.
- It is crucial that the baseband converters will be totally redesigned to be easily manufactured and reliable. The existing Mark III/IV VCs and VLBA BBCs are so obsolete that it is impossible to build a good system using them.

Things to be avoided

- Building non-standard recorders: the commercial tape drives already have better bit density and reliability than the custom recorders used in VLBI.
- Building non-standard data communications equipment: commercial equipment can be easily upgraded, custom interfaces are a pain. Part of this lesson has already been learned in the VSI interface project, VSI is a non-standard way of connecting non-standard equipment together.
- Separate control buses: MAT, MCB or CAN really have no advantage over standard Ethernet and make the system complicated. If Ethernet is used to transfer the data, the same connection can be used to control the equipment.
- Real-time operating systems: truly time-critical things should always be done in hardware.
- Centralized and bulky field system: there is no reason why the BBCs, recorder units etc. cannot obey the SNAP commands directly. Field system should be much simpler.

Subsystems of the Next Generation VLBI

Digital baseband converters

The most important part of the next generation VLBI system are the digital baseband converters.
Clearly we need baseband converters that are affordable, easy to manufacture and, what is most important, robust: they either work or are totally broken.

The concept has been discussed in the ALMA Memo #305: "A Digital BBC for the ALMA Interferometer", MMA Memo #204: "Digital Filtering in the MMA" and MMA Memo #248: "Computer Simulation of a FIR Filter for the MMA". http:///www.alma.nrao.edu/memos/.

The digital baseband converters will produce sampled data output using 100 Mbit/s Ethernet.
The BBCs will send the data using the UDP protocol (a protocol widely used in Internet).
The UDP packets will contain the unique address of the baseband converter in source address field and the address of the correlator channel in destination address field, as well as the time when the data was recorded. All further routing and processing will be based on these fields.
Formatting and sending of the UDP packets is done in hardware using the timing from hydrogen maser clock. Of course, the fast Ethernet controller should be part of the BBC ASIC.
Commercial VHDL implementations for 100 Mbit/s and gigabit Ethernet controllers are available.
A baseband converter will be controlled by a simple and low-cost computer with embedded Linux.

How many BBCs are needed is to be determined. Because the characteristics of the recording subsystem are hidden from the BBCs and the correlator, we can use a few wideband BBCs or many narrowband BBCs. Anyway, the BBCs should be low-cost modules that can be added to the system at will.

Gigabit Ethernet switches (record)

Gigabit Ethernet switches are commercially available. No modifications or programming are needed. The switches work at wire speed and are non-blocking. Data is switched to recorder units based on the destination address, the recorder units don’t receive frames that are going to other units.

If needed, the gigabit switch will multiplex data from several BBCs to one recorder unit or fan-out the data from one high-speed BBC to several recorder units. If fan-out is needed, the digital BBC must modify the destination field of the consecutive UDP packets so that the packets are switched to several recorder units.

Recorder / playback units

The function of the record/playback units is to hide the characteristics of the tape recorders, hard disks etc. from the rest of the system. If the recorder technology changes or the SCSI interface dies out, recorder banks can be changed without affecting the rest of the system. Even different recorder types may be used at the same time with no modifications in the system.

Recorder and playback units are based on COTS tape drives and embedded Linux computers. Recorder units have extremely limited functionality, the basic function is to receive data in UDP packets from Ethernet and store it to tape until the tape runs out. Playback units also have extremely limited functionality, the basic function is to read data from tape and send it to correlator when the correlator control requests it. The correlator control usually requests data frames that start at a certain time using a broadcast message. All playback units that have data send it to correlator channels.

High performance recorder / playback units (optional)

Optionally the recorder/playback units can contain a hard disk. This configuration is useful in near real time VLBI, where data can be stored and sent forward via Internet in the nighttime, when there is abundant bandwith. Hard disks may be useful in the playback units, because data can be transferred from the relatively slow tape storage to high-speed disk storage.

Gigabit Ethernet switches (playback)

The switches are commercially available, work at wire speed and are non-blocking.
The routing of data from playback units to correlator channels will be done automatically based on the destination address of the UDP packet.

At this moment the commercial gigabit Ethernet switches have maximum non-blocking bandwidth of 256 Gbps, enough for any playback system. The technology continues to be developed, the next generation switches will have 10 Gbit/s Ethernet ports.

Correlator front-end and control

Correlator front-end is built using general purpose computers with embedded Linux.
One computer is needed for each correlator channel.

When data is correlated, all data cassettes from the time to be correlated are placed in playback unit banks. The correlator control then sends a broadcast command to the playback units, requesting the units to send data from the first timeslot to the correlator front-ends.

The playback units then send the data, it is routed according to the address recorded in the data and arrives at the front-ends via gigabit Ethernet links.

The front-ends then feed the data to correlators. This process is repeated until the end of the data on the tapes.

Connected arrays

The same design can be used in connected arrays, the only change is that tape record/playback units are not needed. Digital BBC outputs are connected to the correlator using 100 Mbit/s or Gigabit Ethernet. The same Ethernet is used to control the radio telescopes, thus eliminating the need for a separate control network.

The ultimate example of connected array is of course the ALMA interferometer that will have 64 antennas with data streams of 8 Gbit/s per antenna. Even these speeds are well within the capabilities of currently available COTS Gigabit Ethernet switches. Some planning will be necessary, because the system will need several Gigabit Ethernet switches.

At this moment the maximum capacity of a single Gigabit Ethernet switch is 256 Gbps and the maximum rate in one fiber is 10 Gbps (using WDM technology). When ALMA becomes operational, 10 Gigabit Ethernet technology is widely used and inexpensive.

Appendix A: COTS tape drives

At this moment (October 2000) high-end tape drives have 200 GB capacity.

Street price for top-of-the-line Seagate 100/200GB LTO drives is 5953 dollars
The street price of the two-year-old technology, the 35/70 GB AIT drives has dropped to 1750 dollars.

We can deduce four things from these facts:

- Apparently the price per capacity is constant. We can use one high-end or three medium capacity drives to store the same amount of data. The price is roughly the same.
- If we start designing the system now, the price of the high-end drives has dropped to one third in two years when the design has been finished.
- If it takes two years to complete the design, the capacity of the high-end tape drives has tripled.
- The tape technology evolves so fast that it is necessary to change the tape drives several times during the expected lifetime of the whole system.

Tape drive technology has been discussed in greater detail in the Haystack memo
"Concept for an Affordable High-Data-Rate VLBI Recording and Playback System".

Appendix B: COTS gigabit Ethernet switches

Commercially available gigabit Ethernet switches have enough bandwidth to be used for both VLBI recording terminals and correlators.

Several companies manufacture gigabit Ethernet switches, for example Extreme networks (www.extremenetworks.com), Juniper (www.juniper.net) and Hewlett-Packard (www.hp.com).

Some examples of the products available today:

Extreme Networks BlackDiamond 6816 (suitable for correlators):

> The BlackDiamond 6816 leads the industry in density with up to 192 Gigabit Ethernet ports in a single chassis.
> The BlackDiamond 6816 supports a maximum of 1,152 10/100BASE-TX ports when fully populated with the F96Ti. In addition, when fully populated with the F48Ti,the BlackDiamond chassis will accommodate a maximum of 576 10/100BASE-TX ports, or 448 100BASE-FX ports when fully populated with the F32F module.
> The BlackDiamond 6816 supports a maximum of 128 Gigabit Ethernet ports when fully populated with the G8Xi and G8Ti modules, or 192 Gigabit Ethernet ports when fully populated with the G12SXi modules.
> Non-blocking 256 Gbps backplane yields over 192 million packet per second throughput
> 10 Gbit/s Ethernet supported over single pair of fibers.

Extreme Networks Summit48

> With 48 switched 10/100 Mbps auto-negotiating Ethernet ports and two full-duplex GBIC-based 1000BASE-SX, LX or LX-70 Gigabit Ethernet ports, Summit48 has a 17.5 Gbps non-blocking switch fabric and a forwarding rate of 10.1 million packets per second. Summit48 comes with wire-speed Layer 2 and wire-speed basic Layer 3 switching using static routing or RIP V1/V2 routing protocols.

Extreme Networks Summit24

> With 24 switched 10/100 Mbps auto-negotiating Ethernet ports and one full-duplex GBIC-based 1000BASE-SX, LX or LX-70 Gigabit Ethernet port, Summit24 has an 8.5 Gbps non-blocking switch fabric and a forwarding rate of 5.1 million packets per second. Summit24 comes with wire-speed Layer 2 and wire-speed basic Layer 3 switching using static routing or RIP V1/V2 routing protocols.

Gigabit and 10 Gigabit trunk lines

> At this moment Gigabit Ethernet ports are widely used.
> 10 Gbit/s trunk lines (with WDM technology) have been available since Q1 of 2000.
> 10 Gbit/s Ethernet standard will be ready in March 2001.

These products are only examples of what is available today.
These products are relatively inexpensive because of the heavy competition on the gigabit ethernet switch market.

Appendix C: Embedded Linux controller

The tape record / playback units and the digital BBCs are controlled by relatively simple, low-cost computers with embedded Linux.

One (but not only) possibility for a controller is the Axis ETRAX 100 processor that is normally used in printer servers, storage servers etc.

One example of the capabilities of the processor is the developer board that costs 299 dollars.
Axis also promises to sell the processors only (in quantities).
Axis has apparently another type of processor board that supports SCSI and EIDE devices.

Axis Developer Board - Hardware

The board, which is 84 x 104 mm has the following features:

ETRAX 100, highly integrated 100MIPS 32bit RISC CPU
Ethernet 10/100 Mbps Twisted Pair
2 RS-232 serial ports, 9 pin male D-SUB with RX, TX, RTS, CTS, DTR, DSR, CD and RI signals.
1 RS-485/RS-422 serial port on a screw terminal block (1 combined RX/TX pair and 1 TX pair)
Power LED and Status LED
TEST button
RESET button
BOOT button - to enable network boot
Power: 9-24 V AC (or DC) on a standard connector and on screw terminal block
Possible to mount Logic Analyzer connector - making it possible to use Axis' logical analyzer for advanced low level debugging
2 parallel ports on pin headers (3.3V, one with buffers - one without)
FLASH: 2 MByte
RAM: 8 MByte DRAM
Shipped with Axis embedded Linux port

More info about the ETRAX 100 processor and the developer board can be found in www.axis.com and developer.axis.com