EXPReS JRA1 FABRIC
Data Acquisition Requirements

Ari Mujunen, amn@kurp.hut.fi
Jouko Ritakari, jr@kurp.hut.fi
Metsähovi Radio Observatory

05-May-2006

Abstract

This document first describes the VLBI data acquisition process with the special focus on the desirable properties (requirements) which promote flexible development of novel ways to transfer and process VLBI data. The scope of the data acquisition work package within this JRA is defined and alternative implementation approaches are discussed. Finally, a recommended development strategy and suggestion to subsequent steps are presented.[1]

Contents

[1]This work has received financial support under the EU FP6 Integrated Infrastructure Initiative contract number 026642, EXPReS.

1 Background

During the approximate 30 year history of VLBI data acquisition system development efforts the traditional approach has been to create custom electronic solutions to RF input signal sampling, its time-tagging and formatting for subsequent magnetic tape recording, and centralized correlation processing at a dedicated central location. This process has been plagued by extended development cycles, in many cases more than five years, and the inability to take advantage of any emerging new technologies. Early in this history this style of development was well justified mainly because the data bandwidth and storage capacity requirements of capturing a VLBI RF signal far surpassed those which were readily available on the general technology market.

In the recent past, much of this has changed. The general Information and Communications Technology (ICT) is starting to meet and in some cases already surpass the VLBI requirements of multi-Gbps speed and multi-TB-storage facilities. Custom-designed electronics is quickly becoming a burden which prevents making full use of the latest ICT technology offerings.

At Metsähovi Radio Observatory (MRO, Helsinki University of Technology TKK) we believe that the main goal of EXPReS JRA1 FABRIC is to develop simple and minimalistic designs which allow us to transform special VLBI data into regular ICT data and make extensive use of the constantly-developing facilities offered by the commercial-off-the-shelf (COTS) ICT industry to process and transport that data. This will bring us a long-term sustainable and scalable solution in which its components can be easily renewed at the pace of ICT technology advances.

2 VLBI Data Sources

The very basic function of any VLBI data transport/storage system is to capture the radio frequency (RF) signals of a set of radio astronomical receivers situated across a wide geographic area and facilitate the comparison (or "correlation") of the simultaneously received signals. Traditionally this has been achieved by digitizing the receiver RF signal, supplementing it with time code bits originating at an atomic clock (an H maser), and by recording the digitized bit streams onto transportable media (typically magnetic tape). The media from all the participating locations has been physically shipped to a central "correlator" where the time code has been deciphered and the original sampled bit streams restored and compared/correlated with custom electronics.

In this chain, the reception of radio signals with receivers and digitizing the resulting RF signals happens outside ("before") the scope of this JRA. It is nevertheless worthwhile to make a few notes on the trends in the developments of receivers and digitizers.

Already for several years at higher sky frequencies (e.g. at approximately 8 GHz and higher) it has been relatively easy to achieve 500 MHz (or wider) RF bandwidth in an radio astronomical receiver. (At lower frequencies both RF technology and the international RF allocations limit the observable bandwidths.) With Nyquist sampling and the minimal number of sample bits, just one, this translates to 1 Gbps minimum. Often it is desired to observe simultaneously both left and right circular polarization (LCP, RCP) which are delivered as two RF signals, resulting in a continuous bit bandwidth requirement of 2 Gbps minimum. By acquiring two sample bits per sample (instead of just one) it is easy to increase the

bandwidth requirement to 2 Gbps and 4 Gbps, respectively. Since a RF bandwidth of 2000 MHz is fully attainable in a high-frequency receiver, generating 16 Gbps sample bit streams is not at all far away in the future.

The traditional way of digitizing a wide RF band has been to transform it into multiple adjacent narrow bands using analog devices called "baseband converters" (BBCs). Each of the narrow bands has been digitized at a lower sampling frequency, resulting in multiple lower-speed bit streams. The development of high-speed analog/digital converter chips (A/D converters, ADCs) has made it possible to abandon the complex analog separation of one wide RF band into several narrow ones, and instead directly digitize the wideband signal into one 1-bit or 2-bit stream of samples. Several developments (again, outside the scope of this JRA) are underway that combine high-speed A/D converters with FPGA chips to deliver multi-Gbps sample bit streams out of the receiver RF signal. The capabilities of these digitizers and A/D chips seem to match very well the 500..2000 MHz capabilities of receivers, so that technology-wise the reception chain is well-balanced this far.

This point in the reception chain where the RF signal has been digitized into bit streams is the input interface point to the "Data Acquisition Architecture" to be defined in the scope of this portion (WP1.1) of the FABRIC JRA. The acquisition solution must be capable of accepting the continuous bit streams and turning them into regular ICT data, to be handled with commodity processing and data communications equipment. For this, we will take a look at the various interconnect options that can be used in this input step.

3 Interconnects: Past, Present, and Future

Traditionally, the multiple low-rate bit streams generated by analog BBCs and low-speed samplers have been transported with custom-designed parallel signal buses, most often implemented with various differential electrical signalling schemes (RS422, differential ECL) and twisted-pair ribbon cables and connectors. The connector pinouts have been different in each variant of the same basic scheme of (typically 32) parallel bit streams. Different signalling and pinouts have effectively prevented connecting different sampler systems to different recording systems. Because of this, an international panel of VLBI users convened during 2000..2003 and created a standard called "VSI", "VLBI Standard Interface"[2] which we will explore further in the next section.

3.1 The VSI-H Interconnect

The VSI-H (H for "hardware") standardizes LVDS differential signalling, a common connector type (an 80-pin "MDR" type connector) and its pinout to support 32 parallel bit streams with one common clock signal and a "1pps" one-pulse-per-second synchronization signal, so that the various existing parallel-type VLBI interfaces could be coerced into one. This conversion can be done with relative ease and in fact MRO has developed in 2002 a simple "VSIC" converter board which converts the most popular existing legacy VLBI connectors into VSI-H. This is one of the strengths of VSI-H: a single VSI-H-based recording system can be connected to various existing legacy systems with simple converters.

[2] http://web.haystack.edu/vsi/index.html

There are some limitations in VSI-H, and most of these originate in its age (approximately five years). The connector type and the recommended cable types do not support the highest clock frequencies attainable with LVDS signalling (up to 622 MHz) but instead the standard defines 32 MHz only as the "base required level". This means 32*32 MHz == 1024 Mbps only, and although the standard defines 64 MHz as "double clock rate" and hints at 128 MHz as being achievable, the connectors and cables most certainly limit the maximum data rate of VSI-H to 4 Gbps. This is lower than what can be soon expected from receivers/digitizers and thus VSI-H should probably be treated as a potential bottleneck in data acquisition, at least in the longer-term future.

3.2 Alternative Interconnects

The VSI-H is certainly not the only option to connect to receivers/digitizers. The ICT industry is successfully replacing parallel interfaces with high-speed serial ones. These include e.g. PCI Express,[3] CameraLink, Infiniband,[4] Myrinet,[5] Serial ATA (SATA),[6] 1 Gbps Ethernet, and 10 Gbps Ethernet. Of these, PCI Express does not have cabling options and is only relevant inside a computer as an expansion connector. Infiniband, Myrinet, and SATA have their specialty uses but they can serve as examples of applicable cabling and connectors for high-speed serial interconnects. CameraLink and the 1 and 10 Gbps Ethernets deserve further investigation.

CameraLink is an LVDS-based industry standard geared towards transmitting high-speed video camera images over a limited-pin-count connector. PCI Express -based interface boards are already available commercially[7] which promise to deliver more than 5 Gbps data transmission speeds. These architectures should be carefully studied to detect if this existing standard could be adapted to VLBI uses.

1 Gbps Ethernet has already practically replaced 100 Mbps Ethernet as the commodity networking interface. 10 Gbps Ethernet is also progressing rapidly into the mainstream and it has the potential of replacing 1 Gbps Ethernet in many applications in the near future. Both the pervasive availability of high-speed Ethernets in ICT equipment and the applicability Ethernet in direct receiver/digitizer-to-processing connections over the (wide area) network warrant further exploration of these interconnect options.

A further advantage of 1 and 10 Gbps Ethernets is the availability of solutions (available cores) to directly implement these connections with the FPGA chips which are already present in the majority of new-generation digitizer designs. That is, since the new digitizers already use FPGAs to interface to high-speed A/D chips, these FPGAs could be used to pack sample bit streams into Ethernet packets, ready to be transmitted to ICT computing and/or networking equipment. There is a downside in this FPGA implementation, however: developing network protocols using FPGAs only (without a conventional computer/CPU) is currently significantly more difficult than with software-based CPU platforms.

[3] http://www.pcisig.com/specifications/pciexpress/
[4] http://www.infinibandta.org/
[5] http://www.myri.com/
[6] http://www.serialata.org/
[7] http://sine.ni.com/nips/cds/view/p/lang/en/nid/14518

3.3 Interconnect Recommendations

The already-available interconnects do not easily go beyond the 16 Gbps speeds that can possibly be attained with next-generation receivers/digitizers. The option to use multiple simultaneous connects can certainly help in this. On the other hand, even the 4 Gbps that can be offered even by the VSI-H connection does not necessarily become a real bottleneck in the foreseeable future, since the target commodity ICT equipment still needs several years to grow beyond 10 Gbps class, the fastest existing/emerging commodity communications technology. Thus the available high-speed digitizers have a lot of influence over which interconnects will become popular over the years. We envisage that initially (for a few years, up to approximately five) the 1..4 Gbps VSI-H will have a large share of the digitizer output market whereas the longer-term trend will be towards utilizing 10 Gbps Ethernet as the direct digitizer output connection.

4 Computer-Centric Architectures

The processing and data communications power of a commodity PC computer has grown dramatically over the years of its existence. In 2002 we realized this at MRO and we developed a simple VSI-H interface board called the "VSIB" that can be inserted into a regular Linux PC using a PCI (32 bit, 33 MHz) expansion slot. Initial explorations on PC disk subsystems in 2001[8] let us expect that we probably could transfer 256 Mbps only from the VSI-H connector into PC memory and from there to the disks. In 2003 when the board design was completed the PC industry had, however, improved the performance level so that 512 Mbps could be transferred with ease.

The power of this computer-centric data acquisition approach lies in the simplicity of the VLBI-specific part of the equipment and design needed. All the other parts of the recording and data transmission system are exactly the same as in any regular ICT application of a Linux PC. When improved disk subsystem hardware or novel networking software becomes available, a computer-centric system can quickly be adapted to using the latest technology.

A downside of using regular computers is that the performance sometimes lies below of what could be attained with custom-built specialty electronics. This is demonstrated by the "VSIB" board so that it cannot handle the full 1 Gbps bandwidth of the VSI-H interconnect. However, by proper design, the majority of these solutions can be made scalable so that multiple PCs can be used to increase the total bandwidth. Since combining a simple and thus low-cost board with a low-cost commodity PC results in exceptionally low total per-unit cost, the total system cost remains lower than with custom electronics, even when multiple units are required to achieve the desired total system performance.

4.1 Computer Subsystem Balance

In a computer-centric data acquisition system the acquisition process is utilizing all parts of the computer: the expansion bus, the main memory, the disk subsystem, and the network interface. None of these alone can be allowed to form a bottleneck, and each of these should be as independent from another so that data

[8] ftp://web.haystack.edu/pub/mark5/004.pdf

paths inside a PC can be freely selected in software according to the VLBI data transmission/processing direction and needs.

The existing "VSIB" 32-bit, 33 MHz PCI board is limited by the PCI standard to a theoretical maximum of 1056 Mbps (without any addressing overheads). In practice this drops down to 600..700 Mbps and it still requires that no other competing devices are actively demanding the shared traditional PCI bus. Thus the VSIB limits the maximum VLBI rate to 512 Mbps. A 3..4 disk subsystem is limited to approximately the same 600..700 Mbps, and so is 1 Gbps Ethernet. Thus, regardless the data direction (in/out/disk/net), the 512 Mbps VLBI data rate can be comfortably achieved, but not much more.

The situation is gradually changing. Whereas in 2002 during VSIB development the main memory bandwidth available in commodity PCs was quite close to 1200..1400 Mbps, it is today closer to 20 Gbps. The shared 32-bit PCI bus is being replaced by point-to-point serial PCI Express bus segments, each of which can transmit and receive simultaneously a theoretical 2000 Mbps per direction and per the so-called "lane" which can be combined in configurations x1, x2, x4, x8, and x16, for theoretical bandwidths up to over 30 Gbps per expansion board. The true attainable net bandwidths are naturally lower, and depend on main memory bandwidth and the effectiveness of the PCI Express central controller chip. With the introduction of on-board 10 Gbps Ethernet controllers the PC is quickly transforming into a "10 Gbps-class" universal data processing/transmission engine.

The transformation does not happen overnight in the sense that the first PCI Express -based PCs would be fully "10 Gbps-class" in every data direction immediately after their introduction. However, the development goal is clear and in this scheme the existing VSIB becomes and remains the main bottleneck. A new PCI Express (x2..x4) interface board would rectify this imbalance.

5 FPGA-Oriented Architectures

In the section on interconnects we already mentioned that new-generation digitizers routinely employ FPGA chips to interface to high-speed A/D chips and that these FPGAs could be used to realize e.g. 10 Gbps Ethernet connections directly in digitizers. One could envision a scheme where digitizers were connected directly to 10 Gbps Ethernet-equipped PCs without any additional expansion boards. It could be possible to develop even more complex network protocols on the FPGAs directly and connect digitizers directly onto wide-area networks so that they could natively transmit their data over the network directly to the processing equipment.

The benefit of this approach would be that if/when 10 Gbps Ethernet-equipped PCs were available, ther would be no need for separate expansion boards to connect the PC to the digitizer. The development burden for the VLBI-to-ICT data conversion would be transferred to the digitizer design.

The processing nodes would not necessarily have to be computers or PCs. An example of this style of FPGA-oriented architecture can be found in the "BEE2/iBOB" design.[9] The benefits of this approach would include higher processing power for a limited number of processing tasks which would then have to be implemented as FPGA firmware instead of regular Linux PC software. The attractiveness of this tradeoff depends greatly on the sophistication level of the FPGA firmware development tools available, as complex tasks developed today as software by default would then have to be developed more in hardware fashion.

[9] http://bee2.eecs.berkeley.edu/wiki

6 Conclusions and Guidelines for Next Steps

To optimally support the transition from 1 Gbps (or less) legacy custom VLBI interconnects to multi-Gbps network-capable commodity interconnects, we feel that supporting 2 or 4 Gbps enhanced clock rate VSI-H interface would be a very attractive "transition option" for the timescale of 3–5 years into the future. The basic 32 MHz clock 1 Gbps VSI-H already has widespread use in the VLBI community, with further migration efforts planned and underway. The process of turning VLBI data in VSI into regular 10 Gbps Ethernet communications data can be more readily experimented and developed using the commodity PC as the prototyping and development platform than what would be possible with FPGA firmware solutions. This guideline is based on the assumption that 10 Gbps Ethernet will gradually gain widespread acceptance in both short-range interconnect and wide-area networking applications and thus it will appear as standard in commodity PCs.

From practical application point of view, a 2–4 Gbps VSI-H interface board can be used not only with new next-generation digitizers but with existing and shortly upcoming digitizers such as the "dBBC". Furthermore, the bandwidth development of commodity PCs happens gradually and it will take a few years before the full 4 Gbps capability of the enhanced VSI-H can be utilized in the commodity PCI Express architecture. (The corresponding development has already occurred with the legacy PCI bus where it took several years for the PC architecture implementations to actually reach the theoretical maximum bandwidths.) It is easy to see that in a few years, a PCI Express-equipped commodity PC with a native 10 Gbps Ethernet connection can easily sustain streaming 4 Gbps of VSI-H originating data in all data directions.

Such "4 Gbps sustained VSI/Ethernet" PC units will pave the way to migrating the fully-developed networking protocol software implementations into FPGA firmware, thus easing considerably the development of high-speed A/D–FPGA-based digitizers with "native" 10 Gbps Ethernet connectivity. The opposite is also true: FPGA-based Ethernet development can be verified by operating them against the "4 Gbps VSI PC".

6.1 Requirements Summary

For the preferred data acquisition solution of a 2–4 Gbps VSI-H-to-PCI Express interface board several key requirements are listed together with some implementation boundary conditions.

The board must be able to transfer input data via PCI Express into PC main memory at its full 4 Gbps bandwidth (up to the maximum supported by the target). The opposite direction, from main memory to VSI-H output is desired but not absolutely required, since the main focus of the JRA WP2 is distributed correlation where the VLBI data, once captured as regular ICT data will also be processed as such.

The data transfer process must incur minimal overheads both in the PCI Express hardware connection and in the Linux device driver part. Quite simply, the board must be capable of using PCI Express Direct Memory Addressing (DMA) so that input data transfers leave the rest of the PC (the processor, other data paths) open to simultaneous other uses.

Although the PCI Express single lane theoretical bandwidth is 2000 Mbps full-duplex there is significant amount of addressing and "handshaking" overhead

present in PCI Express lane connections. A PCI Express x2 board can theoretically support 4000 Mbps which is actually not sufficient for a 4096 Mbps VLBI data stream. Thus a minimum of four PCI Express lanes must be used in the design, or, with two lanes only, it must be accepted that 2 Gbps may prove to be the attainable maximum.

Two basic implementation solutions for PCI Express are either dedicated chips such as those manufactured by PLX Technology[10] or FPGA cores, available from both FPGA chip vendors and from independent suppliers.[11] Further investigation is needed in the design stage to select the optimal solution. Dedicated chips currently do not seem to offer more than PCI Express x1 bandwidths, and FPGA cores quite often have high non-recurring license costs which are prohibitive for a R&D project such as this JRA.

Low manufacturing cost, low part count, and design simplicity are of course desirable properties for this board. The capability to output (a possibly slightly modified copy) of the VSI-H data inputted can be useful in scaling the system to multiple PCs, although its usefulness must be seriously considered, since in the end, with sufficiently powerful PCs, the aim is to capture the maximum 4 Gbps allowed by the VSI-H interface standard into a single PC unit.

[10] http://www.plxtech.com/products/pci_express/default.asp
[11] http://www.xilinx.com/prs_rls/ip/0622_pcie_x8core.htm, http://www.hitechglobal.com/ipcores/pciexpress.htm