Also AT.44.4/005

Australia Telescope Correlator Computer Upgrade

AT.39.3/014

David Loone
10 Sept, 1991

A few months ago, a group (Warwick Wilson, Mike Kesteven and David Loone) started looking for an upgrade path for the online computing systems. The main impetus for this was the difficulty the Correlator Control Computer was having with the data rate, although the observatory general computing is suffering also.

The online computing systems are divided into two parts, the observing system and the correlator control. The observing system at the moment consists of a VAX 8250 (NOEL) and the correlator control consists of a MicroVAX-II (SANCHO) with an array processor. The other 8250 (LEON) plays only a minor role in the online system and functions more as a general purpose offline computer and server for workstations.

1. The Correlator Control Computer.

The Correlator Control Computer performs two functions non-simultaneously. One is to perform the configuration of the correlator blocks by communicating over serial lines to the LSI-11/21 systems in each correlator block, a function which it is quite capable of handling. Its second function is to supervise the processing of the correlator data on the Sky Warrior array processor. The Sky Warrior uses dual ported memory on the QBus to access the data from the correlator. The data is then transmitted to one of the 8250's (currently NOEL) for archiving over a DMA link. There are two problems with this scheme. One is that the DMA line is inefficient (and difficult to filter where it exits the screened room) and the other is that the QBus is overloaded doing I/O to the array processor. The first of these can be fixed by using a different communications medium, the second by using a different processor bus. There are also another problem directly associated with the Sky Warrior in that Sky are showing an unwillingness to support it.

The newer VAX machines (*eg* in the MicroVAX 3400 and the VAX 4000 series) have different I/O architectures (they have the ethernet and disk interfaces directly on the CPU so that they don't use the QBus as heavily), but we have tested the array processor on these machines with no success. In fact Sky (and most other manufacturers numeric accelerator solutions) have abandoned VAX/QBus machines in favour of the VME bus as supported by most RISC/UNIX systems.

1.1 Short Term Solution.

We are currently looking at a short term solution to this problem. The options we are considering include:

- Using the existing ethernet or a separate ethernet segment to get the data to archive. SANCHO may have to get a larger disk (probably a SCSI disk) to buffer the correlator data before being sent over either Local Area VAXcluster or NFS to a disk or tape on LEON.
- Attaching a SCSI interface to SANCHO so that it can have the Exabyte tape drive attached locally. This would necessitate entry to the screened room to mount tapes for archiving, but wouldn't load the site ethernet.

Both these solutions get rid of the DMA interface, but may actually increase the amount of data to be transferred over the QBus. However, the present QBus DMA interface (DRV11) uses an inefficient method of transferring data on the QBus, and more modern interfaces (*eg* SCSI disk controllers) use much more efficient QBus transfer modes (block mode). We may be able to optimize disk accesses by, for example, using only contiguous files, minimizing disk latency. With two 1GByte disk drives, this is feasible. The cost of any of these should be not exceed $20k.

We are currently running on a 10-15 second integration cycle with the prospect of increased data rate as more channels come on line. The solutions described above would hopefully prevent further increases in the integration cycle time in the short term.

Warwick Wilson and David Loone are currently carrying out performance tests on SCSI disks and ethernet controllers to determine achievable data rates.

1.2 Long Term Solution.

The long term solution requiers a shift away from the MicroVAX/QBus architecture for the following reasons:

- The QBus does not have the **I/O bandwidth** to handle these types of I/O intensive applications.
- The VAX architecture does not have the **processing capacity** to manipulate the required amount data (even with a numeric accelerator).
- Developers of numeric accelerator solutions are **moving away from VAX platforms** to platforms more suited to numeric processing (*ie* RISC/UNIX/VME).

The main problem with moving away from the MicroVAX/QBus architecture is the software effort that needs to be put into the changeover, since the current code is heavily VMS specific.

Given the large number of manufacturers making numeric accelerators for VME machines (usually using the i860 processor and quoting upwards of 40MFlops and performing a 1K complex FFT in around 1msec) quite cheaply (around $25k or less) it would make sense to put a VME machine in the screened room. Such machines are available from many manufacturers, either using the VME as the primary system bus (*eg* many SUN systems) or on an adaptor to the CPU (*eg* a VME backplane can be attached to the SBus on a Sparc or the TurboChannel on a DECstation 5000). Such a computer would replace the MicroVAX-II and Sky Warrior combination in the screened room.

At present a number of ASCII terminals on the control desk are used to control the correlator and display subsets of the data. Another computer (a workstation) on the control desk could be used to perform several functions in an updated correlator:

- Act as "console" for the correlator, displaying windows similar to the terminals in use at the moment to display status and diagnostic information on the correlator and accept commands from the observer and the observing program (CACBS).

- Write data to archive media. This workstation would have attached the archive media (Exabyte, DAT, etc) and write the data to this media, either directly or staging it via a local disk. The transfer could be done with NFS or similar protocol.
- Act as display station. Because this processor has access to all the data from the correlator, processing associated with selection and display of the data can be offloaded from the machine in the screened room.

These two machines (the VME bus machine in the screened room and the workstation on the control desk) should be linked by FDDI. The absolute maximum data rate is projected to be around 250kBytes/sec. The maximum theoretical data rate on ethernet is about 1MByte/sec and may have problems handling this traffic when network overheads and other traffic are taken into account. FDDI interfaces are becoming more popular as a networking medium for high performance workstation applications such as this and thus, their prices can only drop (presently DEC have an FDDI interface for the DECstation 5000 for $12k).

The hardware required to implement the above architecture is available from several vendors, notably DEC and Sun. The advantages of this architecture are:

- We **separate the bulk data processing stream from the array operations** in a sensible manner. The current system attempts this, but the processor and I/O balance between the two streams are grossly incorrect. RISC architectures are well suited to manipulating large amounts of raw data efficiently (such as feeding a numeric processor) and VAX architectures are well suited to performing more general purpose functions (such as operating the array and collecting monitor data). A small, well defined set of data needs to be communicated between the two systems (similar to the current LEON/SANCHO communications). This can be achieved using normal communications protocols over the site ethernet.
- The **on-line display** of the data subsets on the control desk workstation can be computed independently of the data processing in the screened room, thus minimizing the interaction between these two machines. Furthermore, even the humblest of RISC workstations has much more computing horsepower than the present Correlator Control Computer. Consequently, much more **powerful display algorithms** can be developed.
- The site ethernet is not asked to carry any of the bulk data load.
- There are **no I/O bottlenecks** as there are at present. Modern workstations have I/O channels capable of at least 50MBytes/sec (easily enough to get data to and from a numeric accelerator) and the FDDI link has a theoretical maximum bandwidth of 10MBytes/sec.
- If more processing power is added to the control desk workstation by using an i860 based processor or some special purpose processor, **new array capabilities and operational modes** may be possible (*eg* on-line data calibration).
- There are no filtering problems associated with getting FDDI into the screened room.
- Not all the software on the Correlator Control Computer would need to be rewritten. The sections which deal with configuring the correlator could remain (either temporarily or permanently) running on the MicroVAX-II, under control of either the observing software or the data processing machine.

1.3 Cost.

As noted above, an i860 processor board would cost less than $25k. A suitably configured DECsystem 5000 with VME bus would be around $60k including the FDDI interface and some development software. A similar machine for the control desk workstation including the FDDI interface and development and graphics software would be around $60k. A hardware interface needs to be developed. This would be largely based on a firmware modification to the current MicroVAX-II interface ($10k). Thus, using DEC components, the project would have an **upper limit** on the budget of around **$170k.** I have less knowledge of the Sun product range, but it would **certainly cost less rather than more.**

These prices are **constantly falling** and no hardware would need to be purchased until at least a few months after the project had begun, because initial software development could be done on existing or low end workstations (*eg* for c++, a long planning period is required before any code is written). In fact, components like the FDDI interfaces (and even the numeric accelerator itself) could wait until near the end of the project to be purchased, as these types of components will have the most dynamic price drops and performance increases. In addition, some of the cost may be reduced by reusing some existing hardware (*eg* if disks are purchased for the current MicroVAX-II for the short term solution, they can be reused on a workstation when the MicroVAX-II is retired as data processing supervisor).

The manpower estimate for the project is between **2 to 3 man years.** The project would be subject to **peer reviews** building on the experience gained by the ACC upgrade.

1.4 Conclusion.

All our efforts to find an upgrade path for the Correlator Control Computer involving VAX/QBus architectures have failed. The short-term solution proposed above can be considered **as a band aid only** and **cannot** be relied on in the long term if for no other reason than **lack of support** by Sky for our array processor. The situation with other array processor vendors is much the same. We could not guarantee the current system (even with the band aid described above) to keep working satisfactorily for more than about three years. The fact that we have had to double and then triple the originally specified 5 second integration cycle (and the situation will only get worse as we begin using recirculation and more spectral line modes) only **adds to the urgency.**

It is also clear that if one were designing a correlator processor system today that the VAX architecture would not be considered. Unfortunately, the correlator was designed when the RISC revolution was in its infancy and the VAX solution was the only practical one at the time. Because of this, we will need to bear the cost of the software effort needed to make the correlator **reliable** and able to perform to its **original specification.**

2. The Observing and Offline Systems.

Because the Correlator computing is diverging from the VAX/VMS world, the upgrade paths of the two also diverge to some extent. As mentioned above, the VAX/VMS architecture, while not at all applicable to numeric data processing, is quite applicable to the more general purpose roles required in running the array. Indeed, the array is being run quite happily at present on only the modest processing power of one VAX 8250 with plenty to spare. In addition, the requirements for driving the array are not predicted to increase as are the requirements for the Correlator Control Computer.

One of the aims in the network strategies for the observatory is to make the ACCs less reliant on a particular host computer by putting the ACC communications on terminal servers. When this has been achieved, we have much more flexibility about how we organise our observatory computing resources, to the extent that the observing program will be able to run on a VAXstation. The eventual aim is the replacement of the pair of VAX 8250 machines which are currently the focus of the site computing (though by no means the bulk of the processing power) and becoming obsolete. Once we have reduced our reliance on terminal lines directly attached to a particular machine, we are able to use cheaper, busless workstation-type architectures. To put this in perspective, 5 years is generally acknowledged to be the useful life span of a computer system, and our VAX 8250 pair are at that age now. This is illustrated by the fact that every product in the current VAX range has at least twice the processing power of a VAX 8250 and even the low end have ample I/O capability to support that processing power (the VAX 8250 is blessed with much more I/O bandwidth than the processor can ever use).

Next financial year, we will be looking at replacing our current VAX 8250 pair with this type of computer. Additional benefits of this are greatly reduced maintenance charges and the freeing up of a large section of the building which can be used to provide better visitor accommodation