Trip Report : DECworld, Boston

J.J.Bunn

DECCO Project

18th. May 1992

This trip report contains information obtained from DEC under non-disclosure rules. It is thus in our interest that the information is not further distributed. Please do not discuss the details of this report with anyone not on the DPG distribution list. Without being melodramatic, any leak of the information outside the DPG might seriously jeopardize our relations with DEC. Additionally, all figures mentioned in this report should be taken with a grain of salt, bearing in mind that they are not measured by us!

DECworld is an exhibition of existing and future products from Digital Equipment Corporation. It lasts for several week, during which time many thousands of "top-level" managers from DEC's customer base pass through the exhibition. Attendance is by invitation only. Selected customers are invited to a special area of the exhibition called the "sand-box" where are displayed the very latest DEC internal development equipment and software.

This DECworld was the biggest yet; around forty thousand participants in total attended the show, which was set up at the World Trade Centre on Boston's waterfront. The total number of chocolate chip cookies eaten during the coffee breaks apparently numbered 220,000! From CERN, the visiting contingent comprised Roy Billinge (Director for Informatics), Julian Bunn (DECCO Project Leader), Gottfried Kellner (Deputy Division Leader, ECP) and David Williams (Division Leader, CN). We were hosted by the DEC HEP office personnel (Michael Brown and Dominique Gillot), and by Antonio Zenobi (DEC Account Manager for CERN). (It is not known how many chocolate chip cookies they ate.)

CHAPTER 1

ALPHA

Of course, the word on everyone's lips at DECworld was "Alpha". There was much to see, and much ingenuity required in posing the right questions to winkle out the juicy details about the new chip and its associated products. Every DEC person had clearly been warned not to give anything away that had not been fully cleared by the marketing people, and even persistent pushing usually wouldn't budge them.

However, the following sections describe what was gleaned at the exhibits, by talking with the demonstrators, and also in discussion with people well placed in the Engineering groups at DEC.

1.1 The Alpha Chip

The Alpha chip operates at 3.3 volts and consumes about 40 Watts of power. Of the four-hundred odd pins on the chip, seventy are power lines and seventy ground. Contrary to rumours being spread by Hewlett-Packard, the chip does not require special cooling arrangements; the standard finned heatsink is not much larger than the chip itself, and does not necessarily need to be blown. The technology is 0.75 micron CMOS. There are already plans in motion to produce the chips at 0.5 microns, and DEC believe that they can probably go as far as 0.2 microns. At 0.5 microns, the expected SPECmark rating will be 400. Approximately 20% of the chips currently being manufactured are able to run at 200 MHz. The peak of the distribution is at 167 MHz (just missing the "supercomputer" ranking for export purposes!). Diagnostic runs on each chip as it comes "off the press" determine the number of blocked channels. Some channels can be re-routed, but others are show stoppers, and the chip must be junked (or handed round audiences at technology seminars). These diagnostics also determine the safe operating speed of each chip, and thus for which platform it is destined. For a volume purchase, DEC are selling Alpha chips at 1600 US dollars apiece.

DEC say that Alpha "in it's first incarnation" is a single-chip design. In discussion, Jesse Lipcon, who is in charge of the Alpha project in DEC, confirmed rumours that the architecture is specifically designed to allow running virtual machines (a la PR/SM on IBM mainframes). This means, of course, that one could partition an Alpha machine to, say, run NT at a certain percentage, and OSF/1 for the rest; particularly interesting for a workstation! For the VXCERN and DXCERN central services, this might be a particularly neat way of combining the two onto one platform, so consolidating system support, data backup, operational procedures and so on. In this case, one would partition between Open VMS and OSF/1.

1.2 Workstations and "Mainframes"

One room at the exhibition was devoted to Alpha-based machines. This was called the "Alpha Showcase". This contained the following:

- A mixed architecture LAVC comprising an Alpha deskside (codename:Flamingo)machine, a VAXstation 4000 and a VAXstation 3100. Each machine was running a simulation of a game of pool. It was possible to use the normal system manager style commands such as "show cluster", "show memory" and so on: everything seemed nicely homogeneous. The operating system was Open VMS. In fact, all the machines in the exhibit that were running VMS were running Open VMS. It was quite clear from the "monitor cluster" display, that the Alpha machine was considerably faster than the other two (as expected). But the interest of this exhibit was the fact that a mixed architecture LAVC is already working.
- A comparison of the speeds of three machines: the Flamingo (running OSF/1), the Hewlett-Packard 9000-750 (running HP/UX) and the Sun SPARCstation 2 (running SunOS) on a floating point bound calculation (the Mandelbrot set). Each machine was equipped with 64 MBytes of memory. A separate VAXstation's task was to display the fractal generated from each of the three machines in a different window for each. In this way, the three machines could be reliably compared (none of them had the task of managing the display). One could judge the relative speeds by the time taken to fill each window with the relevant portion of the Mandelbrot set. (Note that the software was designed so that each machine calculated the same portion of the set; different parts of the set do not necessarily require the same number of calculations.) The observation was that the times taken were in the ratios 1:2:10 (Alpha:HP:Sun), i.e. the Flamingo is twice as fast as the HP on a floating point calculation.

  It is of note that the C code used for the calculation was compiled with optimization turned on on each machine. Interestingly, and much to the delight of everyone from DEC, the HP machine burned up (it runs very hot) on the day we were there, and had to be replaced by a model 720!
- An Alpha DataCentre machine, using 200 MHz chips. DEC's LINPACK results on the 1000x1000 problem indicate a speed of 97 MFLOPS for this machine. For the floating point part of the SPECmark suite, a figure of 220 SPECmarks was forthcoming. In Livermore Loops, with an average vector length of 167, the arithmetic mean yielded 28 MFLOPS, the geometric mean 22 MFLOPS and the harmonic mean 17.6 MFLOPS.

  The hardware itself looked very impressive and solid. The rather rotund gentleman demonstrating the machine spent most of his time standing right next to it with the access door wide open: the air cooling fan produced an extremely welcome draught in the hot room! (See later for details of the power requirements and cooling for an Alpha chip.) The cabinet is that of the VAX 6000 family.
- Oracle was being demonstrated on a deskside Alpha machine. Version 6.0 of RDBMS was being used. DECnet and MailBox drivers were also available. OracleCARD was not yet running on Alpha. SQLnet version 1.2 was running, although the released version will be 2.0, running with Oracle version 7.0. The demonstration was a trivial SQL example on a small database which looped over and over again.

- Matra Datavision's Euclid CAD/CAM package was being demonstrated on an Alpha workstation. Unfortunately, it was not an interactive version of Euclid, so we were unable to make the tests that Roland Messerli had asked us to. According to the demonstrator, the subjective speed-up on Alpha compared with the same demonstration run on a VAXstation 4000 was around a factor three. This is probably limited by the graphics hardware capabilities of the machines compared, in any case.
- Wollongong Group's TCP/IP for VMS was also running on Alpha. Again, the demonstration was very simplistic: an ftp of a file from a VAX machine to the host Alpha machine. The demonstrator pointed out that DEC sell Wollongong TCP/IP, but when asked why DEC would want to do this when they finally ship a fully-featured version of their "TCP/IP Services for VMS", she was hard pressed to find a reasonable reply.

1.3 Alpha PC

To complete the full suite of computers employing Alpha chips, DEC intend to bring to market an Alpha-based PC, which will be a competitor for 586 based models. In this device, much of the extra hardware (off-chip cache etc.) will be stripped away, and the clock speed reduced to 9 nanoseconds. The Alpha PC would run Microsoft's NT operating system. DEC believe they can make such machines very cheaply indeed; first announcements are expected towards the end of the year.

CHAPTER 2

STORAGE SYSTEMS AND SOFTWARE

Under this heading are grouped various bits of information gleaned both at the exhibition, and from the presentations made at the one day meeting held exclusively for the benefit of the European HEP customers.

- All storage media to be produced by DEC will be designed to fit in STORage Modular Enclosures (STORME cabinets). These cabinets come in a variety of heights, the biggest holding around 140 GBytes of SCSI disk in a full configuration. The system is one of "shelves" of devices: a shelf might contain, for example, a RAID array of seven 1 GByte disks. Two or three shelves could be used to house an Alpha machine chassis. The cabinets are rather attractive, and look more like something from Cray than from DEC. The finish on the twin, curved front doors can be chosen from a selection of walnut, oak etc.! Additionally, the cabinets can be sited anywhere, not just in a climate-controlled environment. This makes them ideal as departmental or corridor file-servers, when equipped with an Alpha processor. Note that host interconnects for CI, DSSI, SCSI, FDDI to XMI, BI and TURBOchannel are foreseen in these cabinets. CI will be replaced eventually by GIGAswitch, in any case (see later under Network Systems and Software). Note also that disks and tapes may be mixed on the same shelf in a STORME cabinet. Indeed, DEC are working on tape "striping" (see below).
- Several SCSI disks were on display in the Sand Box, ranging from a 2.9 GByte (formatted) 5.25" (rated at 10 MBytes/sec) to a 220 MBytes (formatted) 2.5" device. DEC believe it is "too risky" having more than 3 GBytes on one spindle, so they are targetting smaller form factors rather than greater capacities.

  The buzzword in storage these days is RAID, and DEC were keen to show it has not escaped their attention! RAID-0 has been around for a while now as "Striping". RAID-1 even longer as "Shadowing". Now RAID-3 is available as a parallel transfer to n-1 disks with the nth. disk holding parity information. RAID-5 is much the same as RAID-0 except that the parity information is spread over the set. Unfortunately, at the moment, the host must take care of running the software which ensures the correct functioning of a RAID array, although this task will be moved to the local controller soon. A RAID-3 array on demonstration in the Sand Box was happily chuntering away. Removal of one disk in the set of five would have caused a four hour rebuild cycle to be initiated, so the demonstrator was loath to do it. The task can be scheduled immediately, for later, or as a background task. One in any case needs a hot spare to put in the place of the faulty (removed) disk.
- Under strict non-disclosure was a new technology magneto-optical drive for which patents were currently being filed. With a 5.25" form factor, 25 ms access time, and data rates from 2.5 to 4.2 MBytes/sec, this is codenamed "New Dawn", and initially will provide 16 GBytes of storage, followed later by 24 GBytes. The estimated MTBF is 300 khours, with effectively infinite numbers of write operations. It will be more than three times less expensive than standard (DEC) magnetic disk.

- Virtual disks were also in DEC's plans. They were showing the sophisticated ESP500, a SCSI-2 device in a 5.25" package. Internaly, this sports, as well as memory chips, a magnetic disk which holds a shadow copy of what's in RAM. The capacity will be 100 MBytes. Similarly, the ESE50 is a 600 MBytes virtual disk, rising to 1 GByte without local disk caching. An internal battery allows DEC to "guarantee" data integrity on power failure. They recommend this particularly for VAX 9000 machines.
- On the tape front, DEC were describing plans for yet another successor to the TK50, much to the incredulity of us all. However, 3480's are alive and well, and they intend announcing a 3480 robot device which will undercut STK at a price of around 100 kdollars for an entry silo (although David Williams believes that the robot will be a TK50-type device: this has to be clarified with DEC). They were interested that we were interested in optical tape, which they pointed out has very slow access times and bandwidth, but which can hold enormous quantities of data.
- On display elsewhere in the exhibition was a Network File Server. This is enclosed in a STORME cabinet, holding 58 GBytes of SCSI disk in RAID arrays, and using an Intel 486-based controller running Novell Netware, with connections for FDDI as well as Ethernet. The NINJA optical disk juke box was also on display, known now as the StorageServer 100.

CHAPTER 3

NETWORKING SYSTEMS AND SOFTWARE

DEC have something rather exciting up their sleeves with GIGAswitch. They see LAN access methods evolving towards point to point links, away from shared broadcast channels. Backplane technologies will thus move from shared parallel devices to crossbar switches, and this is what GIGAswitch is. The advantage of point to point links is that they allow one to choose the speed of each connection according to what is required; why use a 1 Gigabit/sec link if only 10 Mbit/sec is required ?
The GIGAswitch is a crossbar switch which will operate at around 3.6 Gigabits/sec, handling packets at a rate of 6.25 million a second. Up to 34 FDDI rings may be attached to the switch, and most of the usual protocols will be supported. Logical sub-networks called DOMAINS may be defined in the switch, and several switches may be linked together. The networking people at DEC are very keen that CERN becomes a Field Test site for GIGAswitch, and this will happen in the Autumn this year. DEC firmly believe that GIGAswitch will remove I/O limits in the network itself: bottlenecks will only arise in the workstation or mainframe. Internally, they have run task to task applications over GIGAswitch at rates up to 60 MBytes/sec, at which point the internal machine buses saturate.

DEC are very active in the ATM Forum, but are worried about the unrealistic attitude towards the non-trivial problems in making a transition to the new technology. A second version of their Autonet intelligent switch project, called Autonet II, will be ATM based. There will also be ATM interfaces available for the GIGAswitch.

CHAPTER 4

OPERATING SYSTEMS AND SOFTWARE

The port of VMS to Alpha (or, more generally, Open VMS), will be fully complete by 1995. The vast majority of important components of VMS will be ported well before then. A major release of VMS (Version 6) is coming, which dispels the popular myth that "there will never be a version 6 of VMS"! This goes some way to confirming the notion that the advent of the Alpha chip has breathed new life into the VMS operating system.
However, DEC's stated long-term operating system strategy is OSF, and this to run on the Alpha platforms. They believe that DFS will be able in about one year from now. Apparently, OSF had just succeeded in getting a DFS client to work in OSF! They have developed their own log-based file system for OSF, which avoids the use of "fsck" (which is usually very slow). In their system, a log is written of every transaction on the file system, and this can be "re-played" in the case of a system crash, affording very fast recovery. Several vendors are apparently licensing this software from DEC: it's name is EPISODE.

DEC were advocating the use of Full Sail for distributde management of Unix systems (we have just received a Field Test kit at CERN). They now position DECathena as a "more centralised, hierarchical model", which is "not as flexible as Full Sail". This is in some contrast to their position this time last year (see F.Gagliardi and my trip report from then).

Most of us from HEP expressed our worries about how to easily migrate from existing MIPS-based solutions (DECstations etc.) to Alpha, in the fullness of time. Were DEC going to "leave us in the lurch" with a load of MIPS-based machines, while they concentrated all their efforts on the Alpha front ? Would they still support the MIPS platforms beyond the availability of the R4000 chips ? And was there really any point in paying to upgrade to these chips, rather than biting the bullet and moving straight to Alpha ? Although DEC made a good effort in trying to reassure us on these points, I think there is still some uncertainty, if not in DEC's minds, then certainly in our own.

CHAPTER 5

VARIOUS ITEMS

5.1 MPP

Some of us attended a seminar on DEC's plans for MIMD MPP systems. These will only run OSF/1 and/or Ultrix, and not VMS. They confirmed that they believe making MIMD MPP workable is largely a software problem. They have tools like a Fortran-77 to Fortran-90 converter, which is clever in making use of the array constructs in the new source form, and in unrolling DO-loops etc..
For SIMD MPP, they have around 100 DECmpp 12000 systems (SIMD) installed worldwide, and estimate MTBF of at least 1000 hours on these machines.

5.2 Multimedia

There were a lot of practical demonstrations of multimedia at the show. At various strategic points around the exhibition were touch-screen workstations which could be used to find out about the company, its products and staff. By touching a picture of one of the Vice Presidents, the picture would animate, and begin to speak. Sometimes the motion was a bit jerky, although in general the quality was very good. Of course, the subject matter of the speech was generally very tedious ....
Also around the exhibition were numerous 'phones which could be used to leave messages for other participants by means of DEC's DECvoice product. This is like an embellished multi-user answering machine, but does not, of course, use cassette tapes! Messages can be left for other people, ones own messages can be reviewed, forwarded to other people, annotated with one's own voice, and so on.