VLBI2010: Commercial-off-the-Shelf Techonology Perspectives in 1996, 2003, and 2010

Ari Mujunen, Jouko Ritakari

Metsähovi Radio Observatory, Helsinki University of Technology
Contact author: Ari Mujunen, e-mail: `Ari.Mujunen@hut.fi`

Abstract

In the VLBI2010 session we are trying to predict the world a mere seven years in the future. In these seven years the infrastructure in computing and data communications will change more than anybody realizes now. We can get some insight about the rate of change by looking back seven years into year 1996. At that time microcomputers had 133MHz processors, 16MB of main memory and the largest hard disk was 2GB. Mainstream networking technology was 10Mbit/s Ethernet, the first 100Mbit/s equipment was starting to arrive. Global Internet connections were typically 155Mbit/s at their best. Now we have 3GHz processors (22 times improvement), 1GB main memory (62 times more), 200GB hard disks (100 times larger) and 1Gbit/s Ethernet networking (100 times faster). Internet backbone networks routinely use 10Gbit/s connections (66 times faster). If the trend continues, in 2010 we will have 66GHz processors with 60GB main memory and 20TB hard disks. Networks will work at 100Gbit/s speeds. Global connectivity will be available at 660Gbit/s. Of course these numbers are wrong in details, but it is better to use them than to think in terms of today's technology. It is imperative that we do our development work as fast as the computer industry and we must accept that at least one generation (possibly two) of VLBI equipment will become totally obsolete before we reach 2010.

1. Introduction

Historically, custom-designed VLBI data acquisition and data processing (correlation) equipment have enjoyed the status of bleeding-edge technology that no other data processing application area has matched in capacity nor in speed. The continuous advances in general-purpose data processing technology have, however, started to match and/or surpass the capabilities of specialized VLBI equipment.

It is not easy to recognize and comprehend the speed of mainstream technology development, mainly because it is far too easy to overestimate advances in the near future and far to easy to underestimate the development in a longer timescale. In this paper, an attempt is made to predict the level of impact of mainstream technology developments in the scale of seven years from today. The perspective is sought by seeing back into the past seven years from today. The scale of plus/minus seven years is apparently long enough to prevent "short-sightedness" of the true impact of continuing, exponential speed and capacity growth which can esily be seen in every major area of mainstream data processing technology development. Several key areas are being illustrated in Figure 1.

2. Disk Development

The backbone of VLBI data acquisition and processing, acquiring and storing RF data arriving from a sky source at real time, relies exclusively on high-rate, vast-capacity data media. Traditionally, this has been magnetic tape. However, the protected nature of encased magnetic disks has made it easier to enhance the data densities of magnetic disks much further than the data density of magnetic tape. At the same time, the market demand for high-capacity fast disks in the general PC market has increased the volume of hard disk shipments, making it economically viable to invest in developing faster and larger hard disks. As a result of this trend, the hard disks surpassed the general-purpose magnetic tapes (geared at data backup applications) years ago, and at this very moment they are quickly surpassing the capabilities of specialized VLBI magnetic tapes.

The largest standard hard disks of desktop PCs were 2GB in size in 1996. In 2003 we could easily buy 200GB hard disks for regular desktop use. Projecting this to 2010, we can expect standard form-factor hard disks to reach 20TB sizes, with the imminent increase in data transfer speed accompanying the capacity increase. The predictions of hard disk industry in [1] are actually even more optimistic, stating areal bit densities of 1Tb/in^2 by 2010.

What this means to VLBI is that by 2010, every general-purpose desktop PC is capable of recording and storing hours of multi-Gbps VLBI data onto its standard hard disk drive. (20TB is approximately 44 hours of 1Gbps data.) Intermediate buffering of acquired VLBI data will be a non-issue and it can be performed at will as required, not limited by hardware or economical boundary conditions.

3. Networking Development

Transporting VLBI data acquired at participating observing stations to be correlated has been—this far—the second major bottleneck in VLBI operations. Physical shipment of media has been the only viable option but this is being changed by global high-speed networks.

In 1996 the mainstream local area networking technology was 10Mbit/s Ethernet; correspondingly, global Internet connectivity was available at the 155Mbit/s ATM level. In 2003, LANs are quickly migrating to 1Gbit/s Ethernet and the global Internet commonly has 10Gbit/s backbone connections available. This 100x development predicts 100Gbit/s LANs and 660Gbit/s backbone connections in 2010 which greatly surpasses the VLBI transport demands and effectively eradicates the need for physical media shipments. (Transferring one 20TB hard disk (44 hours of 1Gbps data) over a 100Gbit/s connection takes less than half an hour.)

4. Processor Development

Perhaps the most astonishing advances have been made in semiconductor technology and especially in the area of computer CPUs. The desktop processors of 1996 performed at the level of 53 million instructions per second and with 133 MHz clock rate. [2] Today in 2003, 3 GHz processors routinely crunch 6 billion instructions per second, effectively 113 times faster. Combined with main memory and bandwidth developments this has made it feasible to perform VLBI correlation entirely in software of general-purpose computers. The processors of 2010 will have the processing power of "66 GHz" CPUs with 678 billion instructions per second.

Currently, software correlators can process in the range of 64–128 megasamples per second of a few baseline data in real time. (The exact estimation formulas are shown in [3].) A network of 10 observing stations could correlate the 45 baselines at 1Gbps in real-time using approximately 8 PCs per station. This takes maximum advantage of time-sliced FX architecture where delay tracking, fringe rotation, cross-correlation, and integration consume less Gflops than the initial FFTs—which only need to be calculated per-station and can be "re-used" for all baselines, thus avoiding "baseline explosion".

The architecture where an experiment is divided in time equally to all participating stations to be correlated results in well-balanced full-duplex network traffic. The large main memories make it easy to keep data of a given time slot from all stations in main memory. As a whole, the stations only need $(n-1)/n$ times the disk buffering capacity that they need for their own data to be able to correlate all stations of $1/n$ time of the experiment.

5. VLBI-Specific Hardware Development

Historically, the VLBI data acquisition systems have been constrained by the recording system, thus less emphasis has been put into sampling bandwidth or receiver RF bandwidth. VLBI receivers have had less than 500 MHz of RF bandwidth in each frequency channel and this has been largely unchanged since 1978. The last time DAR sampling bandwidth has increased was in 1996 when the Mark IV formatter was introduced with 2048 Mbps sampling capability. The previous doubling occurred in 1989 when the VLBA DAR featured 1024 Mbps samplers.

Advances in semiconductor technology, especially in high-speed A/D converters, cellular base station RF components, and digital FPGAs all contribute to the fact that it will be increasingly easy to increase the sampling bandwidth using direct IF sampling and digital signal processing and filtering techniques. Thus we will quickly collide into the RF bandwidth limitations of VLBI receivers—2Gbps takes care of the whole 500 MHz at two bits per sample.

6. Conclusions

Using short timescales (1–2 years) to estimate technology capabilities can lead to the false impression that "technology at first advances just a little bit and then all development stops." When designing VLBI2010 equipment, falling to this trap would lead to designs which cannot keep in pace with the advances in mainstream technology.

In the areas of data storage, networking, and processing speed it is likely that mainstream technology will easily satisfy and surpass VLBI requirements before 2010. It is unclear if VLBI-specific technology (samplers, receivers) can be developed quickly enough to match this capacity increase of COTS devices.

References

[1] Maxtor Corporation, Areal Density Chart, `http://www.maxtor.com/en/images/technologies/add.gif`

[2] van Dorst, Wim, WvD@clifton.nl, `http://www.linux.org/docs/ldp/howto/BogoMips.html`, v34, 2003-08-07, jubilee edition, 2003.

[3] Moritaka Kimura, Junichi Nakajima, Hiroshi Takeuchi, and Tetsuro Kondo, 2-Gbps PC Architecture and Gbps data processing in K5/PC-VSI, In: Technology Development Center News, Communications Research Laboratory (CRL), Serial No. 23, Kashima Space Research Center, Tetsuro Kondo and Yasuhiro Koyama (eds.), 12–13, November 2003.

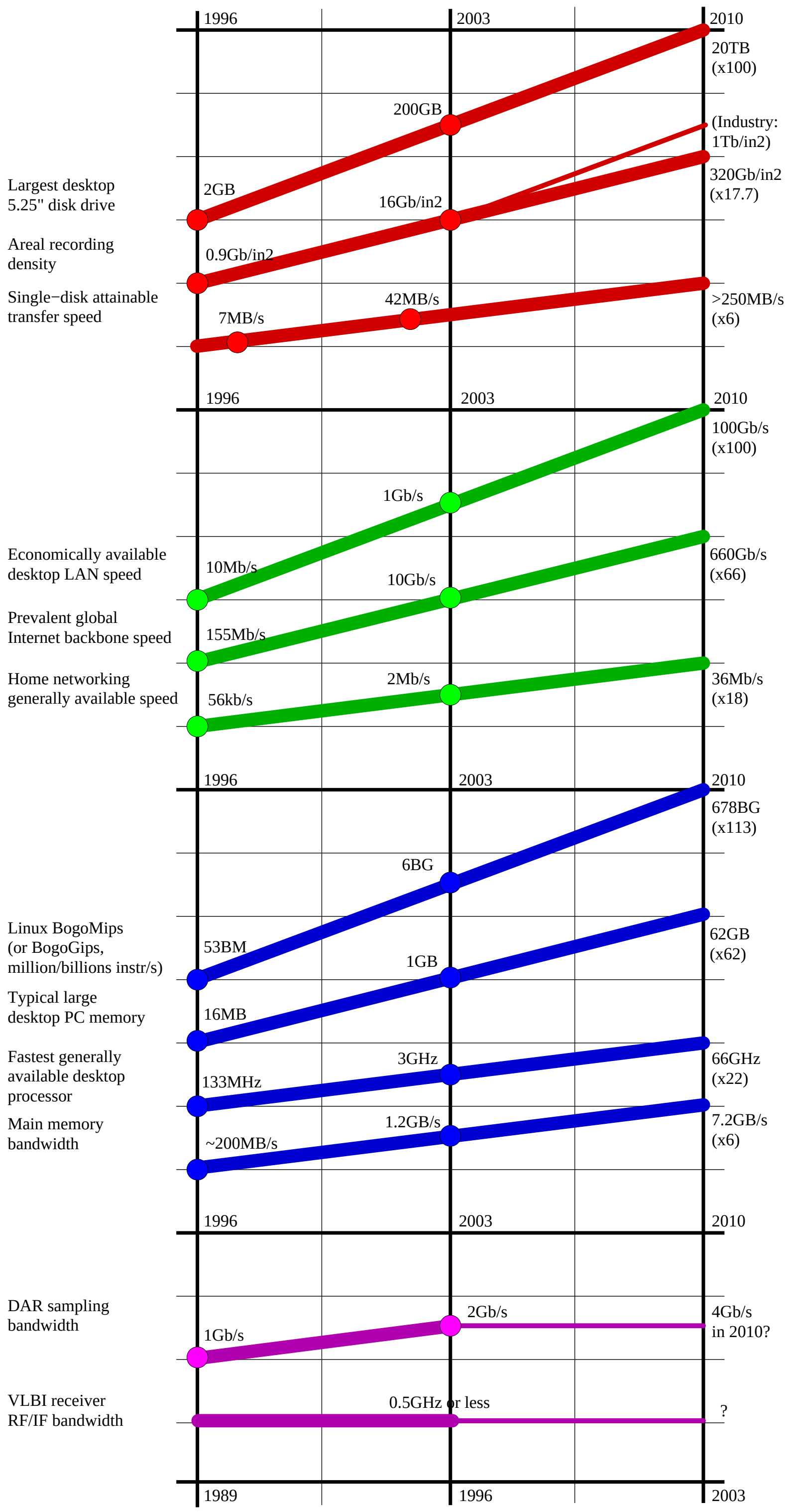

Figure 1. **Several key COTS trends in 1996..2010.**