Applying Off-the-Shelf Technologies in eVLBI

Ari Mujunen, Ari.Mujunen@hut.fi
Metsähovi Radio Observatory
Helsinki University of Technology
Outline

- Overview of available off-the-shelf technologies
- Evaluation test results
- Current technology limitations
- Impact of technology limitations on system design and system scalability
- Near-future off-the-shelf trends
- A strategy for an uninterrupted 1Gbps eVLBI system
Overview of Available Off-the-Shelf Technologies

- COTS: means mainstream technology, used by millions of people
- Tremendous R&D investments by industry in general-purpose COTS disk, computer, tape, and networking technologies
- Thus industry does R&D for us faster than we can do on our own
- Only if we can directly use standard technologies without locking to them
COTS Disks and Computers

- Both are evolving at exponential “Moore rate” with no signs of slowdown

Sizes have tripled since the start of disk recorder projects!
COTS Tapes and Networking

- Faster high-capacity tapes “in the works”
- Sony S-AIT-1(,2,3,4), Quantum SDLT, HP/IBM/Seagate LTO2
  - 500GB, 320GB, 200GB / 240Mbps, 128Mbps, 240Mbps...
- Expensive drives (~$10k) and tapes; still slow; not at all a mainstream technology

- Networking migrates 100M->1G->10Gbps
  - ATM and others losing to Ethernet
  - 1Gbps Ethernet becoming mainstream
Evaluation Test Results

- Dell PE1600SC with ServerWorks GC-SL
- On paper, multi-Gbps
- In practice, 400-500Mbps with ~70% CPU load!
  - Two Promise IDE PCI ctrls
  - Single rd 600, wr 700Mbps
  - rd+wr: rd 400 + wr 550Mbps
  - VSIB rd+wr: the same
  - Sum of I/O always <1Gbps.
Evaluation Test Results

- MSI K7N2G-ILSR with nVidia nForce2
  - On paper only PCI32/33, <1Gbps
  - In practice 0.7Gbps from and 0.4Gbps to PCI board <- > on-board IDE with only ~45% CPU load
- Parallel ATA disks not suitable for >45cm cabling; limited COTS solutions to disk swaps
  - Longer cables force reducing speed from UDMA133/100 to UDMA66 or even UDMA33
    - UDMA33 allows only 120Mbps/disk (half of native perf.)
    -> 9 disks needed for 1Gbps
Current Technology Limitations

- Low-cost computers have just gained the capability to reach 0.5Gbps in all directions
- PCI(net/VSIB)<--memory<--disk
- The leap to >1Gbps will probably happen only after 1Gbps networking has replaced 100Mbps everywhere (2005?)
Impact of Limitations on System Design

- All subsystems achieve 0.5Gbps but 1Gbps is difficult/flaky
  - Input/Output (PCI)
  - Memory, CPU
  - Disks
  - 1Gbps Ethernet
- So let's make a scalable 0.5Gbps box!
  - Affordable enough, can buy as many as needed
  - Standard enough, can replace when needed
MVR Recorder with VSIB Board

VSI-H compatible
Chainable for:
  More speed
  More on-line storage

0.5Gbit/s for 1200 euros!
Near-Future Off-the-Shelf Trends

- Large Serial ATA (SATA) disks replace PATA
  - $1/GB, drops quickly (600GB for $300 in 2004?)
- 1Gbps Ethernet everywhere
- Backups migrate from tape -> disks
  - SATA native hot/warm swap connectors
- HyperTransport, PCI Express, ...
Motherboard Trends

- Multiple high-speed channels to memory
  - Intel E7501 2x Xeon, 3x HI2.0, 1x HI1.5
  - Intel E7505 (1x HI 2.0)
- AMD Opteron with 3x 6.4GB/s HyperTransport
- nVidia nForce 3 Pro chipset
- AMD-8000 chipset
A Strategy for an Uninterrupted 1Gbps eVLBI System

Mk4 Fmt

Mk4 Fmt

MMM

VSIC

VSIC

Set #1

“CoolSwap”

Set #2

VSIC

VSIC

VSIB

VSIB

VSIB

VSIB

1Gbps Switch

4x MVR (or MVR chains)
An Example

- 3.5h of 2 Gbps onto 16 200GB disks
- Or continuously 1Gbps onto 8 disks, swapping the other 8 every 3.5h
- Only 9477 euros
  - (plus disks)
  - Incl. 19in rack & 4x Gbps Ethernet /w 8-port switch
Metsähovi VSI Data System Evolution

- **Jun-2001** First wired prototype
- **Jul-2001** Tests of sustained Linux disk performance
- **Oct-2001** First prototype VHDL ready and running at 8bit@32MHz
- **Jan-2002** VSIB PCB board design ready
- **Feb-2002** VSIC PCB board design ready
- **Mar-2002** Second prototypes assembled and tested at 32bit@18MHz
- **Apr-2002** Last PCB changes for mass-production
- **Jun-2002** 256Mbps VSIB&VSIC playback tests at JIVE
- **12-Jul-2002** Fr! 256Mbps Mk4/5P Westerbork disk-Jodrell tape at 5GHz (fringes at JIVE)
- **Aug-2002** Total of 100 VSIB and VSIC boards produced
- **24-26-Sep-2002** 256Mbps iGRID e-VLBI demo (JB, WB, JIVE)
- **26-Sep-2002** 1Gbps ADS-1000 MH-Kashima 22GHz (RX problems)
- **2-Oct-2002** 1Gbps ADS-1000 MH-Kashima 22GHz (weather problems, Kashima typhoon)
- **16-Oct-2002** Fr! 1Gbps ADS-1000 MH-Kashima 22GHz (fringes found on W3OH at CRL)
- **22-Nov-2002** 1Gbps Mk4/5A MH-Jodrell 22GHz (RX problems)
- **27-Nov-2002** 1Gbps ADS-1000 MH-Kashima 22GHz (weather problems, MH snow)
- **5,7,13,14-Feb-2003** Fr! 1Gbps ADS-1000 MH-Kashima 22GHz (fringes found on 3C454.3 at CRL)
- **12-Mar-2003** Fr! 1Gbps Mk4/5A MH-Jodrell 22GHz (fringes on 3C84 at JIVE)