

What Will Have the Greatest Impact in 2010: The Processor, the Memory, or the Interconnect?

José Duato

Dept. of Computer Engineering (DISCA)
Universidad Politécnica de Valencia, Spain

E-mail: jduato@gap.upv.es

WWW: <http://www.gap.upv.es>

Let's look at the relative evolution of processors, memories, and interconnects

My Desktop PC

Year	1995	2001
Processor	166MHz Pentium	>1.2GHz (cooling!)
Memory	16MB, 66MHz	512MB, 266MHz
Graphics	2MB, 2D	64MB, 2D, 3D, MPEG
Disk	2GB	40GB
I/O bus	32bit, 33MHz	32bit, 33MHz

My Desktop PC

Year	1995	2001
Processor	166MHz Pentium	>1.2GHz (cooling!)
Memory	16MB, 66MHz	512MB, 266MHz
Graphics	2MB, 2D	64MB, 2D, 3D, MPEG
Disk	2GB	40GB
I/O bus	32bit, 33MHz	32bit, 33MHz

- Highest speed improvement: Graphics card
- *Only the floppy disk drive and PCI bus remain the same!*

High-Speed Interconnects: Past

- High-speed interconnects only used in medium- and large-scale parallel computers.
- The interconnect was not the bottleneck.
- VLSI and packaging technology improvements were enough to scale interconnect performance.
- Most papers on interconnects from academia had no practical impact.

High-Speed Interconnects: Past

- High-speed interconnects only used in medium- and large-scale parallel computers.
- The interconnect was not the bottleneck.
- VLSI and packaging technology improvements were enough to scale interconnect performance.
- Most papers on interconnects from academia had no practical impact.

No interest from industry on high-speed interconnects architecture (e.g., HPCA-3 panel)

Present

Race for high-speed interconnects

- HyperTransport
- RapidIO
- InfiniBand
- 3GIO
- iSCSI

What is going on?

What is going on?

Shift from desktop computing
to Internet computing

What is going on?

Shift from desktop computing
to Internet computing

Focus on data centers and access through Internet

Data Centers

- Built around high-performance servers, storage subsystems, and STANs
- High-performance servers:
 - I/O buses are the main performance bottleneck.
 - InfiniBand, 3GIO, etc., will replace PCI buses with an scalable and reliable interconnect.
- Storage subsystems:
 - Need for faster, more scalable, reliable backplanes (and faster disks)

Data Centers

- Built around high-performance servers, storage subsystems, and STANs
- High-performance servers:
 - I/O buses are the main performance bottleneck.
 - InfiniBand, 3GIO, etc., will replace PCI buses with an scalable and reliable interconnect.
- Storage subsystems:
 - Need for faster, more scalable, reliable backplanes (and faster disks)

Need for high-speed interconnects in servers and storage

Internet

- Bandwidth demand in Internet doubles every six months.
- OC-192 optical links with DWDM increased fiber bandwidth by a factor of 160 to 320 with respect to OC-48.
- IP routers have become the main bottleneck in Internet.
- IP routers with thousands of ports will soon be needed.

Internet

- Bandwidth demand in Internet doubles every six months.
- OC-192 optical links with DWDM increased fiber bandwidth by a factor of 160 to 320 with respect to OC-48.
- IP routers have become the main bottleneck in Internet.
- IP routers with thousands of ports will soon be needed.

Need for high-speed interconnects in the IP router switch fabric

Other Areas

Industry is designing better interconnect architectures in other areas:

- Supercomputers:
 - Performance improvement driven by ASCI program.
 - Parallel computers with more than 64K processors being designed.
 - Interconnects are very likely to saturate.

Other Areas

Industry is designing better interconnect architectures in other areas:

- Supercomputers:
 - Performance improvement driven by ASCI program.
 - Parallel computers with more than 64K processors being designed.
 - Interconnects are very likely to saturate.
- Small- and medium-scale multiprocessors:
 - Alpha 21364 incorporates an on-chip router. Which will be the next one?

Future

Industry is very likely to require high-speed interconnects in more areas:

- On-chip networks:
 - As VLSI technology advances, transistors become faster but wires become slower.
 - Microprocessors will be designed as several subsystems linked through high-speed interconnects.

Future

Industry is very likely to require high-speed interconnects in more areas:

- On-chip networks:
 - As VLSI technology advances, transistors become faster but wires become slower.
 - Microprocessors will be designed as several subsystems linked through high-speed interconnects.
- Memory access:
 - Most of the memory latency is not memory access time but transmission delay.

Impact of Advances in VLSI

Larger integration scales will aggravate the interconnection problem:

- Wire delay (per unit length) increases while silicon delay decreases
- Most of the power will be used to drive wires
- Quantum technology will dramatically increase integration scale but interconnects between chips (or chips and disks) will not use the same technology

Impact of Advances in VLSI

Larger integration scales will aggravate the interconnection problem:

- Wire delay (per unit length) increases while silicon delay decreases
- Most of the power will be used to drive wires
- Quantum technology will dramatically increase integration scale but interconnects between chips (or chips and disks) will not use the same technology

*Interconnects will be the system bottleneck
and will consume most of the power*

Meanwhile in Academia...

Meanwhile in Academia...

There is a decreasing interest
on interconnects

Conclusions

- High-speed interconnects are gaining attention from industry very rapidly.

Conclusions

- High-speed interconnects are gaining attention from industry very rapidly.
- Academia follows industry trends on interconnect design with several years delay.

Conclusions

- High-speed interconnects are gaining attention from industry very rapidly.
- Academia follows industry trends on interconnect design with several years delay.
- Interconnects (and storage subsystems) will become much more important in the near future (2010?).

Conclusions

- High-speed interconnects are gaining attention from industry very rapidly.
- Academia follows industry trends on interconnect design with several years delay.
- Interconnects (and storage subsystems) will become much more important in the near future (2010?). Well, *infinitely* more important.

$$\frac{\textit{Interconnect Papers in HPCA}x \ (x > 8)}{\textit{Interconnect Papers in HPCA}8} \rightarrow \infty$$