

• Interconnects status

– sources:

- HotInterconnects 03 “War of the Interconnects”
- Supercomputing 03 “Battle of the Network Stars”
- HotInterconnects 04
- OSU benchmarks (everyone quotes them, just not all of them...)

Top500 stats

	11/03	06/04	11/04
GigE	109 (-)	163 (0)	176 (0)
Myrinet	73 (2)	186 (2)	193 (2)
Quadrics	26 (4)	23 (3)	20 (2)
Infiniband	3 (1)	10 (0)	10 (1)
clusters	208 (7)	290 (6)	294 (5)

• HPLinpack benchmark -> Tflops

- application mix evaluates compute power at nodes, also bandwidth (aggregate and point-to-point), and ping-pong and collective comm. latency

Rank	Site Country/Year	Computer / Processors Manufacturer	70720 91750	20520 31363	19940 22938	13880 20480	12250 20240	11680 16384	10310 20019.2	9619 15300
1	United States/2004	BlueGene/L beta-System IBM	32768							
2	United States/2004	Columbia SGI	10160							
3	Japan/2002	NEC	5120							
4	Spain/2004	Manoelstrum IBM	354							
5	United States/2004	Thunder California Digital Corporation	40%							
6	United States/2002	ASCI Q HP	8192							
7	United States/2004	System X Self-made	2200							
8	United States/2004	IBM/ LNL	8192							
9	United States/2004	IBM	2944							
10	United States/2003	Tungsten Dell	2500							

year old status...

	10GigE	Infiniband	Myrinet	SGI	Quadrics
Network Environment	Any: LAN, SAN, MAN, WAN	SAN	SAN	SAN	SAN
Scalability	IP-routed + "Infinite" # of nodes		Source-routed + ?		Source-routed + ?
Cost Per Port	\$\$\$ (2004: \$1000000)	\$	\$	\$\$\$	\$\$\$
Performance MFC-to-MFC	336 MB/s 10.6 us (at socket level)	643 MB/s 6.2 us	250 MB/s 8.3 us	230 MB/s 9.7 us	308 MB/s 7.6 us
Protocols	Native TCP/IP TCP offload RDMA over TCP/IP	RDMA	RDMA	RDMA (or ESB: Private Shared Memory)	RDMA
Total Cost of Ownership	\$	\$\$\$	\$\$\$\$	\$\$\$	\$\$\$\$\$

performance – MPI user-to-user latency and bandwidth

- MPI/IBA gives **6.0** us latency and **843** MBytes/sec
- MPICH/GM **6.3** us latency and **250** MBytes/sec (**500** MBytes/sec with dual-port Ecards)
- MPI/Quadrics (Elan3) **4.0** us latency
- MPI/Quadrics (Elan4) around **2.4** us latency and around **908** MBytes/sec
- 10.0 GigE, at the sockets level (**10** us latency and around **916** MBytes/sec)

Cost

Adapters and Switch ports

- InfiniBand
 - around \$800/adapter and \$200-1000/port
- Myricom
 - \$795 for the latest E-cards, per port cost depends on the system configuration (~ \$1k)
- Quadrics
 - \$1000/Elan4 adapter, \$1.5-2K per port (small systems)
 - \$3-3.5K per port (larger systems)
- 10.0 GigE
 - around \$5.2K/adapter, \$10K per module (1-2 ports)

- Performance vs. Cost
(what is the metrics for performance!!!)

	High-Perf	Low-Perf
High Cost	Quadrics	10GigE
Low Cost	Myrinet	1GigE

- What should be the metrics for performance?

Possible metrics

- Latency and bandwidth always important, but...
- Also
 - overheads, overlap of communication and computation, CPU loads, reuse of buffers, scalability, manageability, reliability, network topology, communication patterns, efficient collective communication, standards compliance, virtualization...
 - what is your application's bottom line?
- Cost of ownership – not just \$ to set it up, but also to maintain and upgrade over time.