



**Sun HPC Consortium
Singapore
May 14-15, 2006**

Advances In Cluster Interconnects

**John Fragalla
Business Development Manager
High Performance Computing
Global Education and Research**

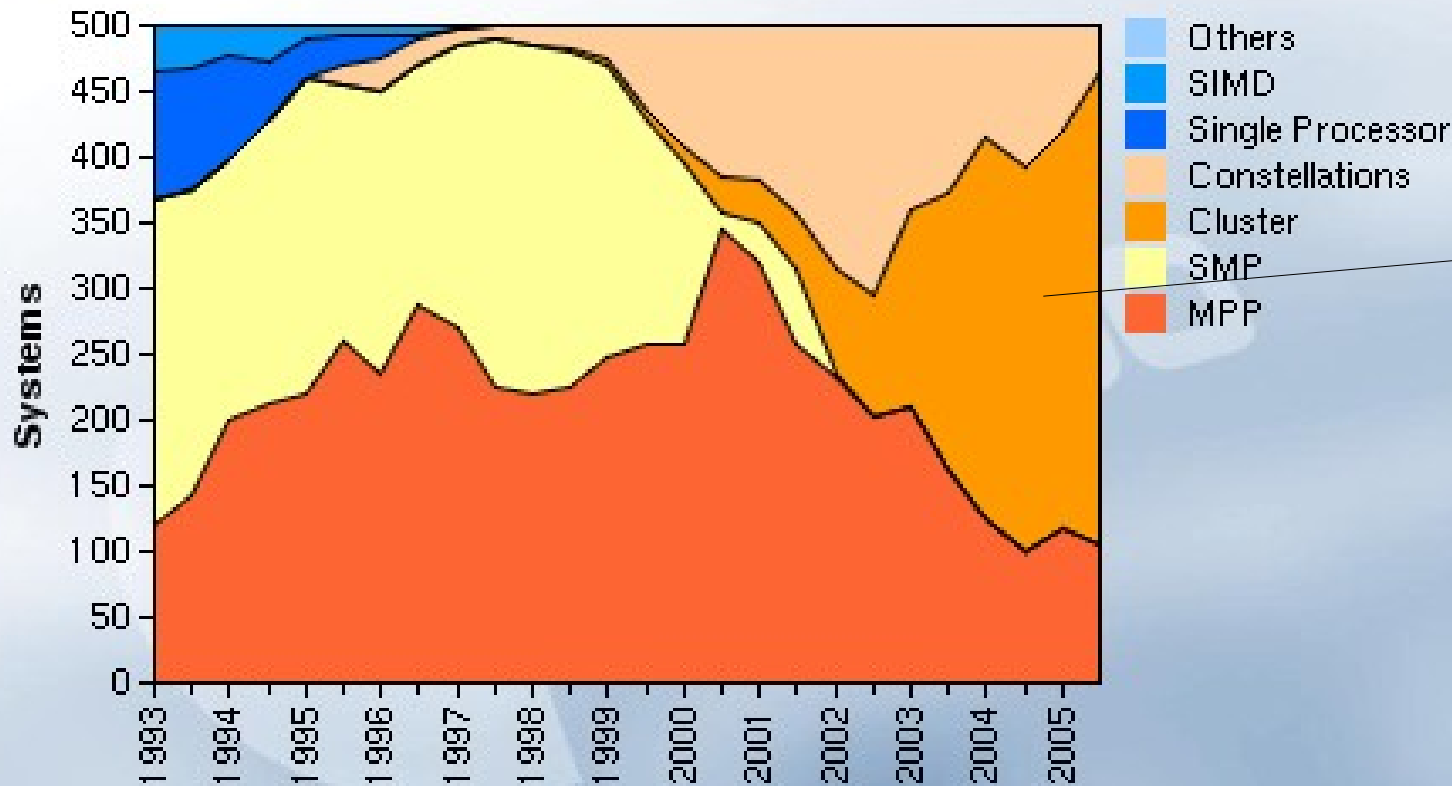
Agenda

- Top500 Trends
- I/O Buses
- Current Interconnects

System Types – Last 12 Years



Architectures / Systems

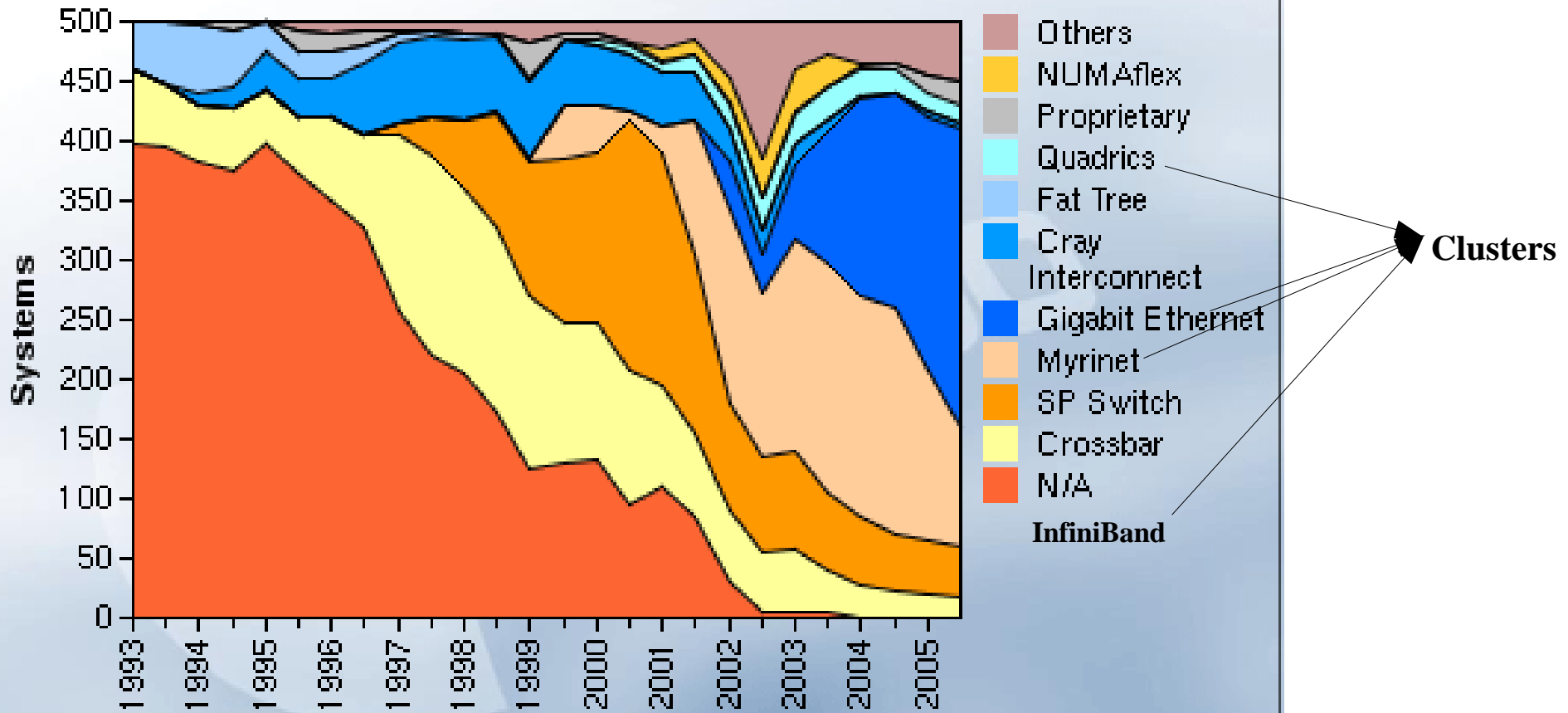


**Clusters:
> 70% of
Top500**

Top500 Interconnects – Last 12 Years



Interconnect Family / Systems



Top500 Interconnects – Last 4 Years

NOV 2005

	Count	Share %	Rmax Sum (GF)	Rpeak Sum (GF)	Procs Sum
Myrinet	101	20.20%	331075	504551	90980
Quadrics	14	2.80%	89107	116831	30052
Gigabit Ethernet	249	49.80%	610975	1175524	191364
Infiniband	27	5.20%	125700	186489	28016

JUNE 2004

	Count	Share %	Rmax Sum (GF)	Rpeak Sum (GF)	Procs Sum
Myrinet	185	37	219314	348877	77772
Gigabit Ethernet	164	32.8	233300	488859	88954
Quadrics	23	4.6	89070	118533	33140
Infiniband	10	2	9772	13852	2684

JUNE 2003

	Count	Share %	Rmax Sum (GF)	Rpeak Sum (GF)	Procs Sum
Myrinet	179	35.8	83655	137790	43164
Gigabit Ethernet	63	12.6	30806	102370	20851
Quadrics	26	5.2	54264	74260	28080
Ethernet	14	2.8	4402	17160	15664
Infiniband	1	0.2	351	461	128

JUNE 2001

	Count	Share %	Rmax Sum (GF)	Rpeak Sum (GF)	Procs Sum
Myrinet	23	4.6	4293	8592	6668
Ethernet	15	3	1429	2975	4896
Quadrics	8	1.6	2171	2944	2176

Top500 Interconnect Trends

- Proprietary Interconnects
 - > Myrinet decreased 45%
 - > QsNet II decreased 46%
- Open Standards Interconnect
 - > Gigabit Ethernet increased 51%
 - > InfiniBand increased 170%

I/O Buses – AGP vs. PCI-X vs. PCI-E

Theoretical and Bi-Directional

- AGP (Advanced Graphic Port)
 - > 4x – 1.056 GB/sec
 - > 8x – 2.112 GB/sec
- PCI-X 133MHz Buses
 - > 1.024 GB/sec
- PCI-Express Buses (1 Lane = 2.0Gbps+2.0Gbps) 80% calculated
 - > 1x – 0.5 GB/sec
 - > 2x – 1.0 GB/sec
 - > 4x – 2.0 GB/sec
 - > 8x – 4.0 GB/sec
 - > 12x – 6.0 GB/sec
 - > 16x – 8.0 GB/sec
 - > 32x – 16.0 GB/sec

Interconnects - Current

Theoretical and Bi-Directional

- Gigabit Ethernet
 - > 0.128 GB/sec
 - > 60 us Latency
 - > 20 us Latency (On-Chip TCP Offload Engine)
 - > Level 5 : 10 us Latency
- Myricom Myrinet 2000
 - > Single Port Card - 0.5 GB/sec (PCI-X)
 - > Dual Port Card - 1.0 GB/sec (PCI-X)
 - > 2.6 us Latency
- Quadrics QsNet II
 - > 1.0 GB/sec (PCI-X)
 - > >= ~1.2 us Latency

Interconnects - Current

Theoretical and Bi-Directional

- InfiniBand 4x – Single Data Rate (SDR) 80% calculated
 - > 1.028 GB/sec (1x 4+4Gbps) with PCI-X
 - > 2.0 GB/sec (1x 8+8Gbps) with PCI-E 4x or 8x
 - > 3.5 us Latency
 - > InfiniHost III Mellanox Chipset
- InfiniBand 4x (8x) – Double Data Rate (DDR) 80% calculated
 - > 4.0 GB/sec (1x 16+16Gbps) with PCI-E 8x
 - > 3.5 us Latency
 - > InfiniHost III Mellanox Chipset
- Pathscale (Qlogic) HTX and PCI-E InfiniPath Adapter (SDR)
 - > 2.0 GB/s
 - > 1.29 – 1.7 us
 - > MemFree
 - > Proprietary Chipset

PCI-Express and InfiniBand

- PCI-E and IB SDR have the same:
 - > Signal Rate: 2.5 Gbps per Link/Lane
 - > Data Rate: 2.0 Gbps per Link/Lane (80% of Signal Rate)
- PCI-E Generation 2
 - > Double Data Rate (DDR)

John Fragalla

John.Fragalla@Sun.COM

Thank You