



Presentation to 2004 Symposium on High Performance Interconnects Panel:

Future Trends in HPC Interconnects

Where we should be going

Thomas Sterling

California Institute of Technology
and

NASA Jet Propulsion Laboratory

August 25, 2004

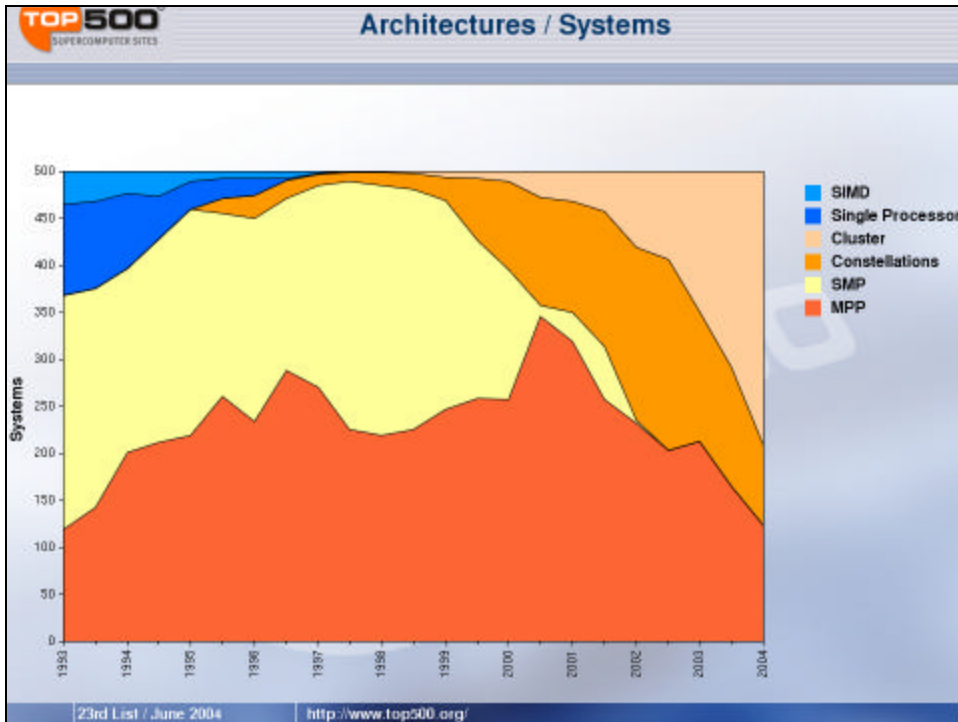




The Two Realities in HPC



- ◆ Commodity Clusters
 - Clusters -NOW
 - CD LLNL Thunder
 - Constellations
 - HP ASCI Q, ARC Columbia
- ◆ Custom Supercomputers
 - Distributed memory MPPs
 - Cray Red Storm
 - Distributed shared memory MPPs
 - SGI Origin
 - PVPs
 - NEC Earth Simulator, Cray X1





Networks for Clusters

- ◆ Networks limited by interfaces to commodity nodes
 - Throughput: PCI, PCI-X, PCI express, IBA
 - Latency > 1 microsecond barrier
- ◆ Cost dominates value to customer
 - Must be less than the nodes to which they connect
- ◆ Performance impact varies dramatically
 - Application
 - Programming model
- ◆ Other factors:
 - RAS
 - Scalability
- ◆ Major train wreck coming
 - New generation of processing chips capable of 100 Gflops
 - Enormous disparity between flops and bps
 - Utilizations will be low, but we will have to give up on that as a metric

August 25, 2004 Thomas Sterling - Caltech & JPL 4



Network for MPPs

- ◆ The situation is not much better
- ◆ On the positive side:
 - Higher density packaging reduces distance
 - Mesh interconnects make sense in some contexts
 - Higher price ceiling because everything costs more
 - May justify:
 - High degree switches
 - High cost physical transport technology
- ◆ On the negative side:
 - Same train wreck from realization that ALUs are cheap
 - Latencies are still thousands of cycles
 - But hey, memories suck too.

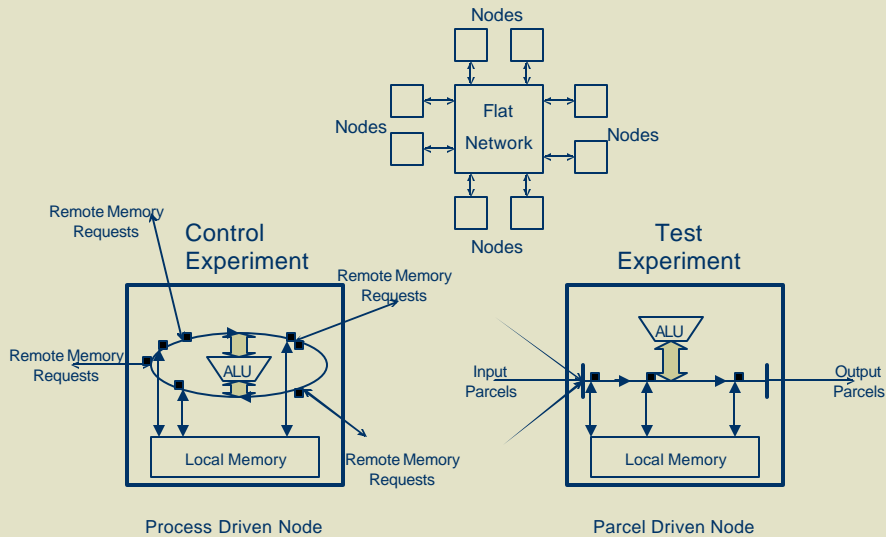


Answer: It's the processor, stupid

- ◆ Interconnects can provide high bandwidth at moderate latencies
- ◆ But only the processors can keep the network busy
- ◆ Custom processors for HPC using high capacity networks must:
 - Manage many (thousands) of inflight accesses simultaneously
 - Operate without blocking on pending communications
- ◆ One possible solution is custom processors
 - Message driven computation
 - Split-transaction processing



Parcel Simulation Latency Hiding Experiment



August 25, 2004

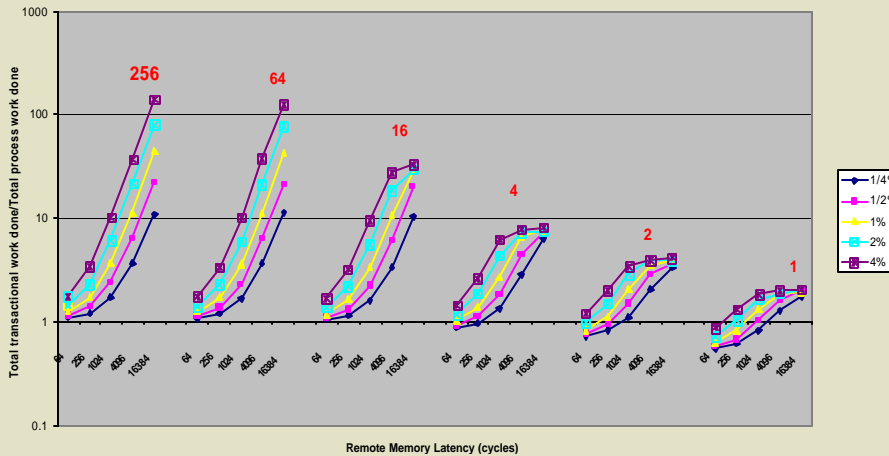
Thomas Sterling - Caltech & JPL

7



Latency Hiding with Parcels with respect to System Diameter in cycles

Sensitivity to Remote Latency and Remote Access Fraction
16 Nodes
deg_parallelism in RED (pending parcels @ t=0 per node)



August 25, 2004

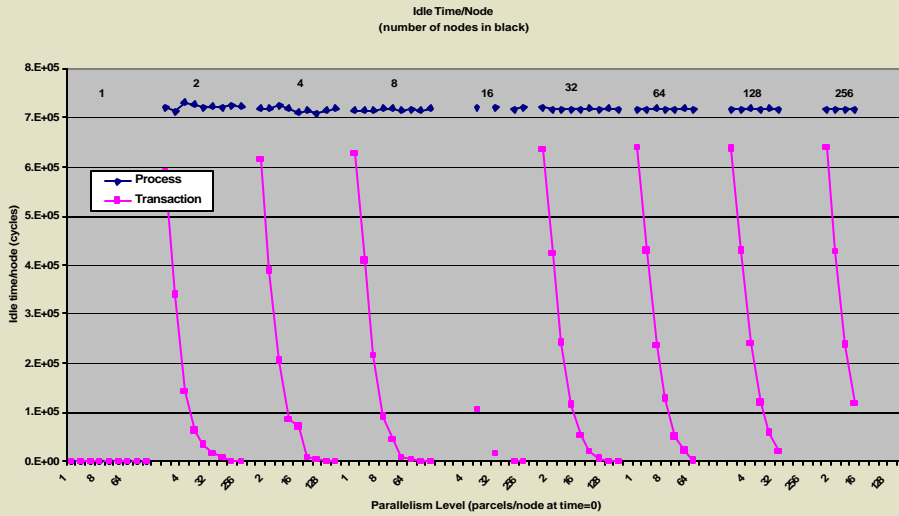
Thomas Sterling - Caltech & JPL

8



Latency Hiding with Parcels

Idle Time with respect to Degree of Parallelism



August 25, 2004

Thomas Sterling - Caltech & JPL

9