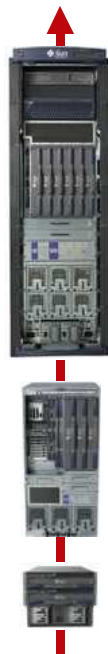


Alan Charlesworth
Sun Microsystems

The Sun Fireplane Interconnect in the Mid- Range Sun Fire Servers



Vertical & Horizontal Scaling



Many CPUs in one box

- Cache-coherent shared memory (SMP)
- Usually proprietary interconnect
- Can be dynamically partitioned
- High bandwidth & low latency
- Good performance on most parallel apps
- Can be more costly for bigger boxes



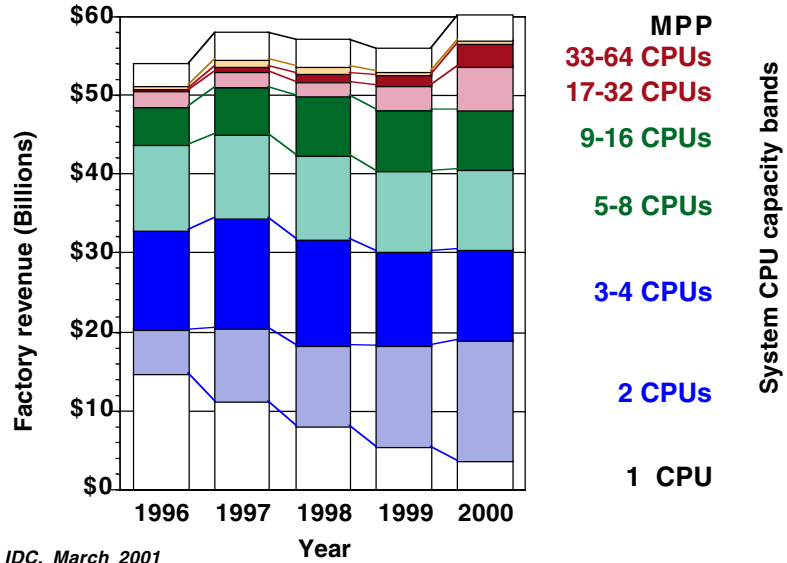
Many systems

- Separate systems, communicate by network APIs
- Usually commodity interconnect & boxes
- Needs partitionable parallel apps
- Databases typically can't be horizontally scaled

take it to the nth



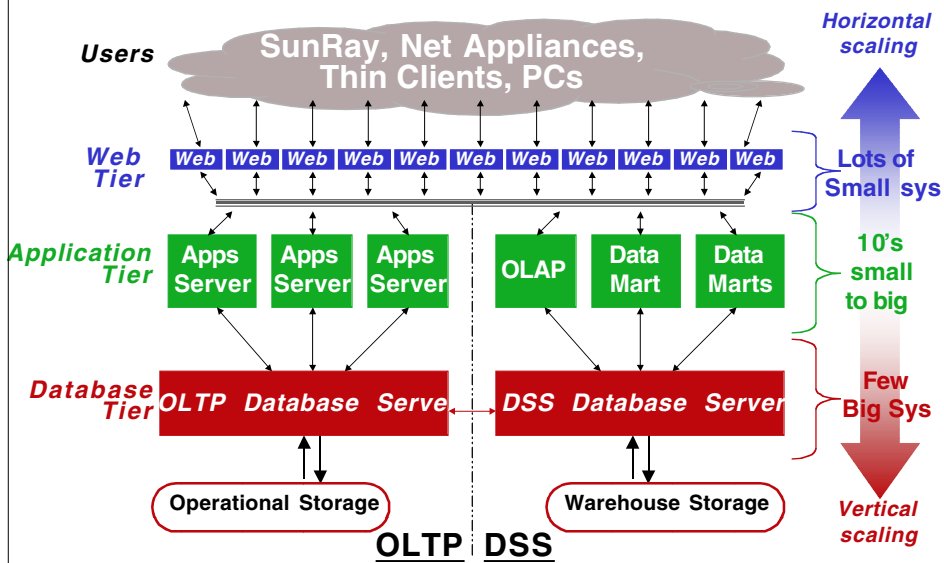
Shared-Memory Server Sales



take it to the nth



Datacenter Tiers

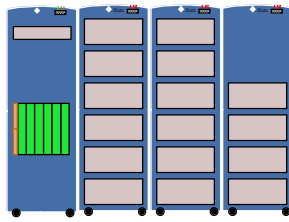


Big Database Examples

SF 6800: 1000 GB TPC-H

1 Server 3 Storage racks

24 CPUs 352 drives
80 GB mem 6 TB

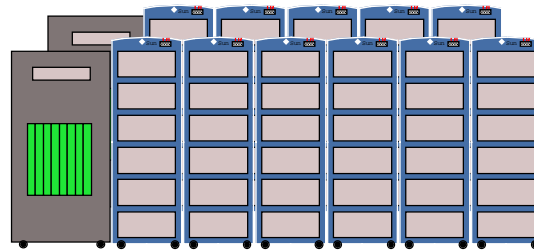


	5-yr Millions	KW	Sq ft
Server	\$1.1	5.4	8.8
Storage	\$0.6	7.5	24.4
Software	\$0.5		
HW Maint	\$0.9		
Total	\$3.1	12.9	33.2

E10K: 3000 GB TPC-H

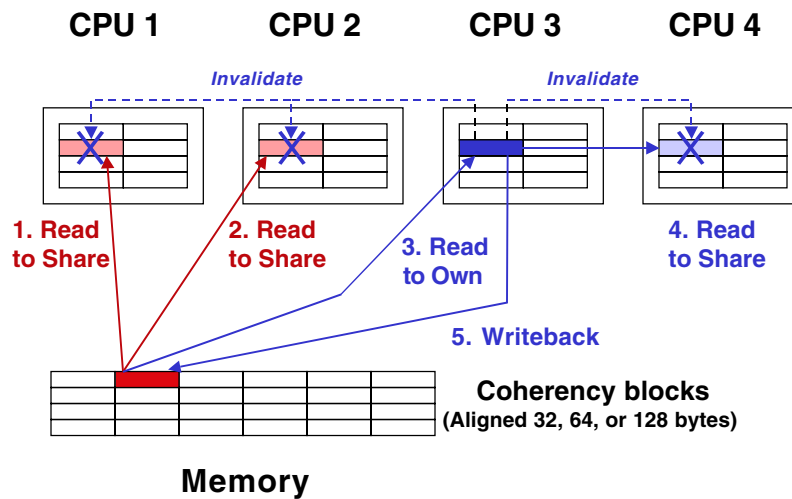
2 Servers 11 Storage racks

128 CPUs 1,392 drives
128 GB mem 24 TB



	5-yr Millions	KW	Sq ft
Server	\$4.6	13.5	27.1
Storage	\$2.5	31.8	89.4
Software	\$1.4		
HW Maint	\$4.6		
Total	\$13.2	45.3	116.5

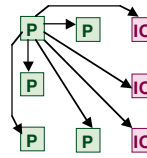
Cache Coherency



Broadcast & Point-to-Point

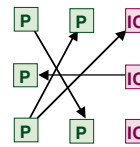
1. Broadcast (Snoopy)

- All addresses sent everywhere
- Snoop result computed in a few cycles
- Lowest possible latency, especially for cache-to-cache transfers
- Data bandwidth limited by snoop bandwidth

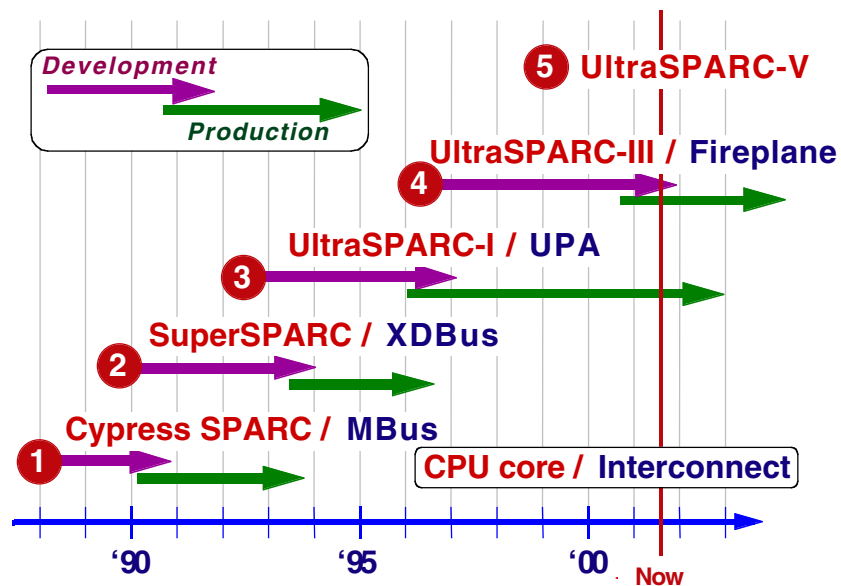


2. Point-to-point (Directory)

- Directory keeps track of who is “interested” in each block
- Addresses sent only to “interested” parties
- Latency usually longer
- Bandwidth can be much greater



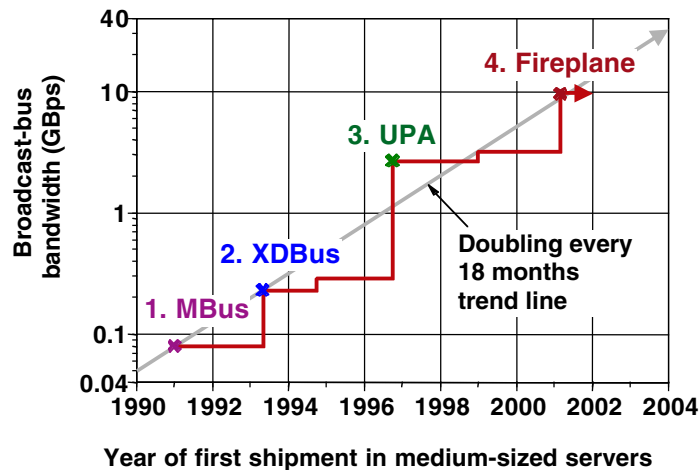
Sun Interconnect Timeline



Sun Interconnect Generations

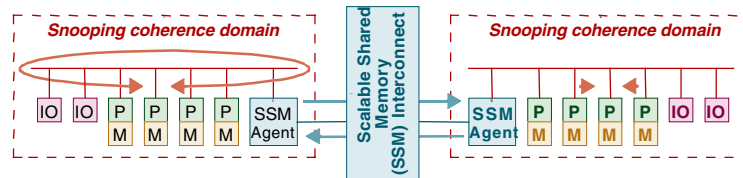
	①	②	③	④
	MBus	XDBus	UPA	Fireplane
Year (in mid-size servers)	1991	1993	1996	2001
System clock(MHz)	40	50-55	83-100	150
Coherency type	Broadcast			Broadcast point-to-point
Packet switching	Circuit	Packet switched		
Address & Data	Together		Separate	
Coherency block (bytes)	32	64		
Sys clocks/snoop	16	11	2	1
Address B/W (GBps)	0.08	0.3	3.0	9.6
# Address buses	1	4	4	>16
Max data B/W(GBps)	0.08	1.3	12.8	>150
Datapath width (bytes)	8		16	32
Wiring	Bused		Mid: Bused High: Switched	Switched

Sun Snooping Progress



Fireplane Cache Coherency

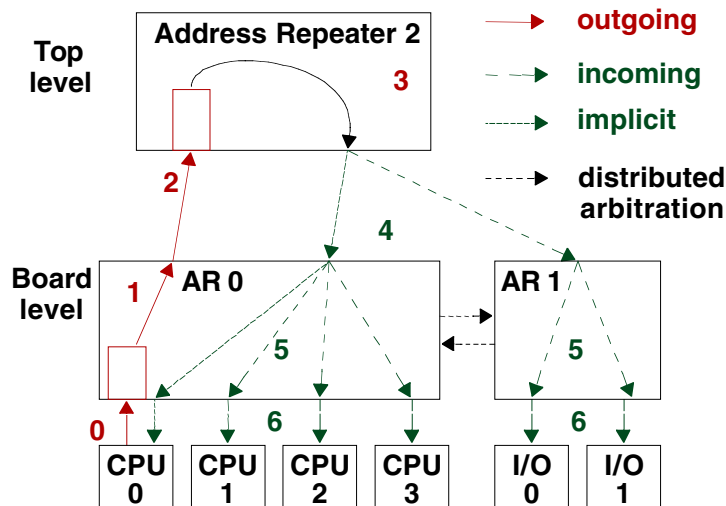
1. Broadcast (snoopy) coherency *inside* a snooping coherency domain



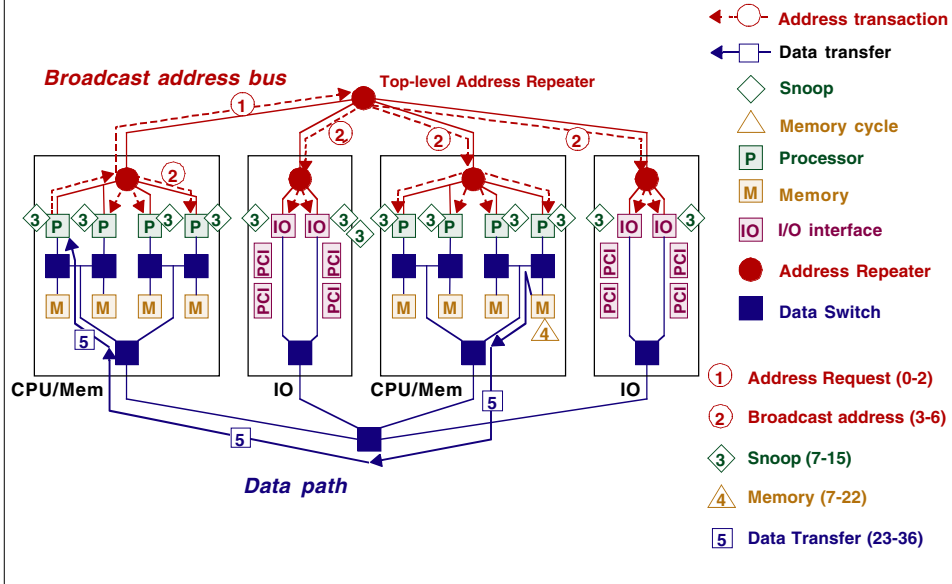
- P Processor
- M Memory
- IO I/O controller

2. Point-to-point (directory) coherency *between* snooping coherency domains

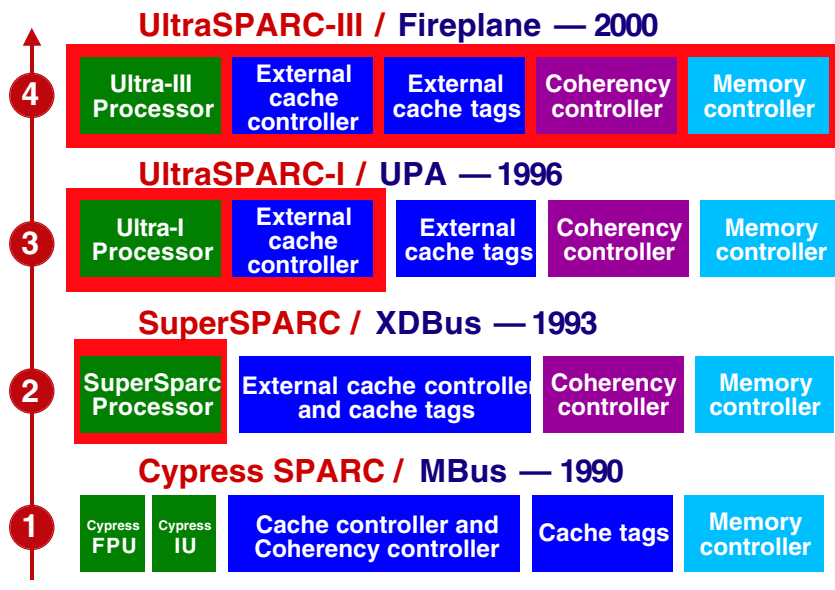
Address Bus Implementation



Snoopy Coherence Domain



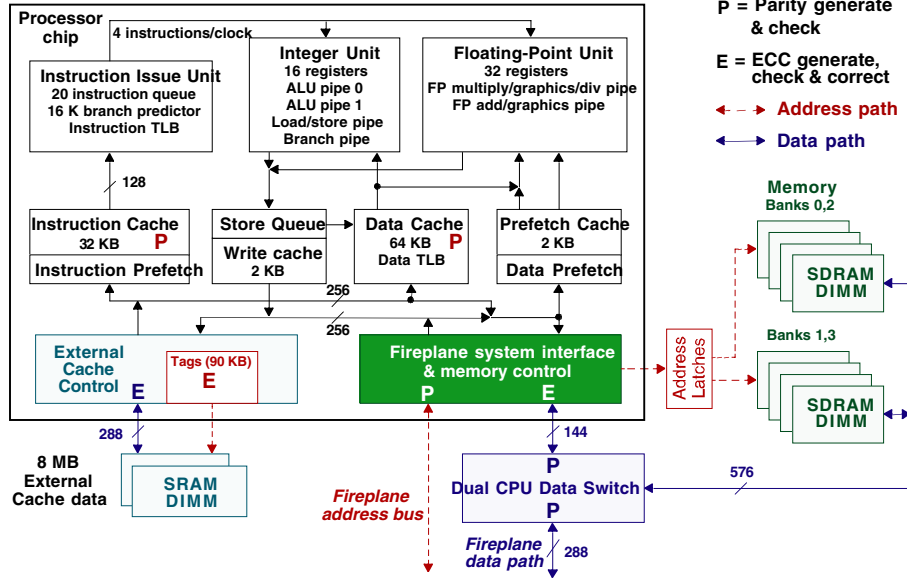
Increasing CPU Integration



take it to the nth



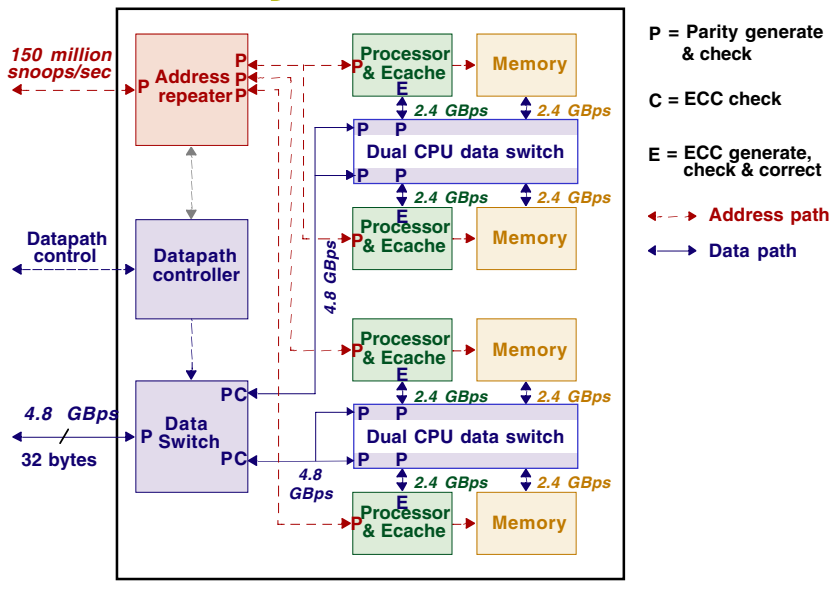
UltraSPARC-III Processor



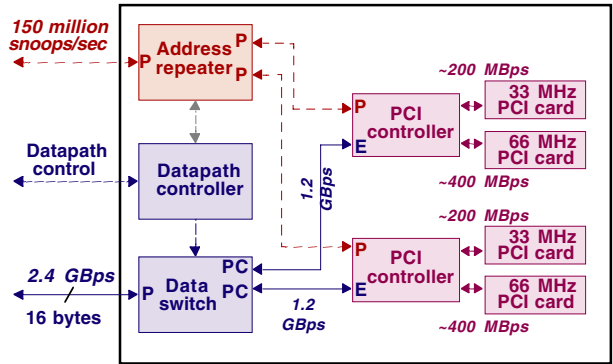
take it to the nth



Sun Fire System Board



Sun Fire IO Assembly



P = Parity generate & check

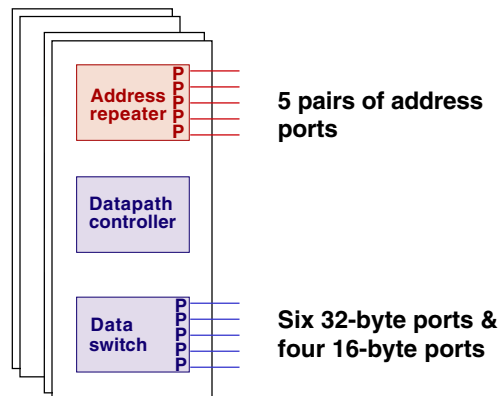
C = ECC check

E = ECC generate, check & correct

← - - → Address path

← - - - - - → Data path

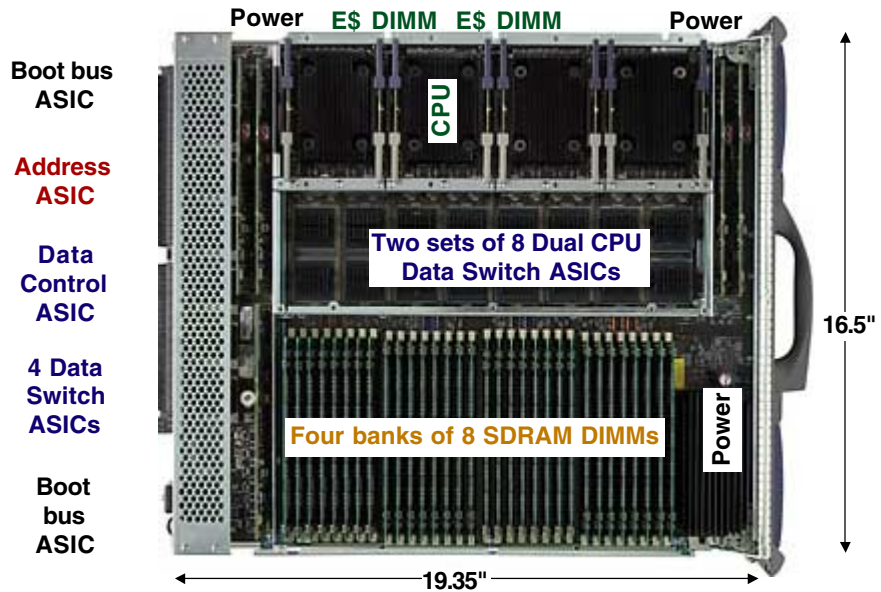
Fireplane Switch Boards



take it to the nth



System Board Picture



take it to the nth



I/O Assembly Pictures

6 slot cPCI



8 slot PCI



take it to the nth



Cabinet Pictures

3 x Sun Fire 3800

Sun Fire 4800/4810

Sun Fire 6800



12 CPUs + 16 PCI slots



3 x (8 CPUs + 12cPCI slots)

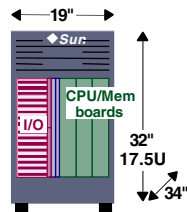
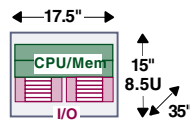
24 CPUs + 32 PCI slots

take it to the nth



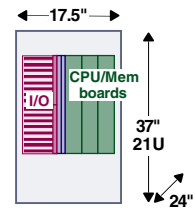
Sun Fire Server Cabinets

Sun Fire 3800
Rack mount
8 processors
2 CPU/Memory boards
2 I/O assemblies
2 domains
9.6 GBps peak BW



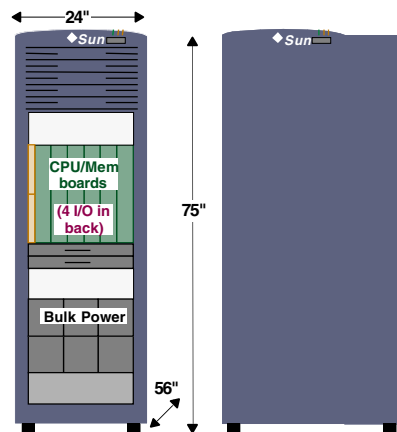
Sun Fire 4800

Deskside or rack mount
12 processors
3 CPU/Memory boards
2 I/O assemblies
2 domains
9.6 GBps bisection BW



Sun Fire 4810

Rack mount
12 processors
3 CPU/Memory boards
2 I/O assemblies
2 domains
9.6 GBps bisection BW



Sun Fire 6800

Cabinet
24 processors
6 CPU/Memory boards
4 I/O assemblies
4 domains
9.6 GBps bisection BW

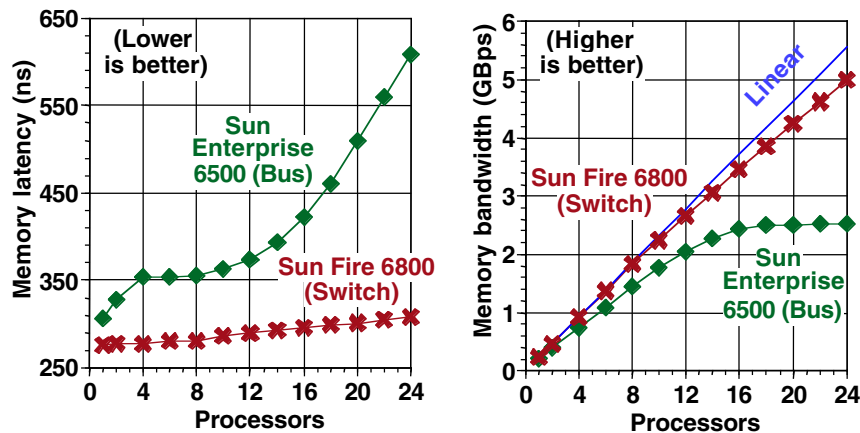


Large Server Cabinet

>64 processors

A Micro Benchmark

Parallel pointer-chasing



Benchmark Record

- **SpecWeb99**
 - 12 CPUs: Web serving
- **SpecJBB**
 - 24 CPUs: 32-bit OLTP app-tier perf
- **TPC-H 1 TB Decision Support**
 - 24 CPUs: Price/perf & perf/CPU
- **Oracle Apps**
 - 24 CPUs: OLTP performance
- **PeopleSoft**
 - 24 CPUs: General Ledger
 - 24 CPUs: Financials