Application Performance on Dual Processor Cluster Nodes

by

Kent Milfeld

milfeld@tacc.utexas.edu

Avijit Purkayastha, Kent Milfeld, Chona Guiang, Jay Boisseau

Thanks

- Newisys (Austin, TX)
  AMD Opteron System
- Dell (Austin, TX) & Cray
  Intel Xeon System

OUTLINE

- HPC needs for Single- & Dual-processor Commodity system Nodes
- The architecture of
  Intel Xeon & AMD Opteron Systems
- Single & Dual Processor Xeon & Opteron Performance Comparison
  - Measured Memory Characteristics
  - Parallel vs Serial Execution of Codes on a Node
    - Kernels
    - Applications

Motivation

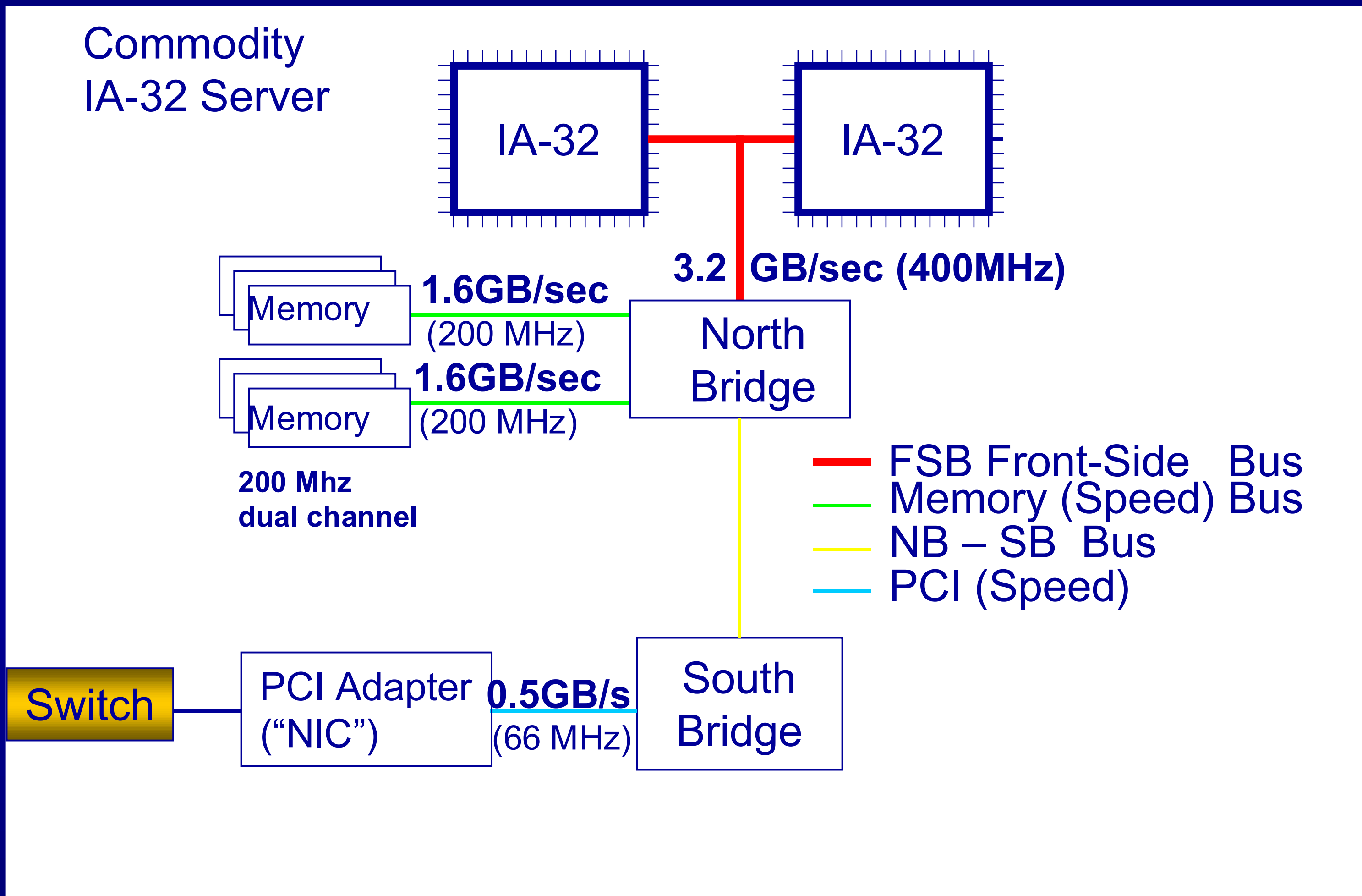
- 1995-2000 Commodity Massively Parallel Systems used uni-processor nodes:
  - Beowulf Systems
  - SP2(SC)
  - T3E
- Today the e-commerce market has driven the price of SMP servers down.
  - Dell, Gateway, HP/Compaq, … compete for this market.

Motivation

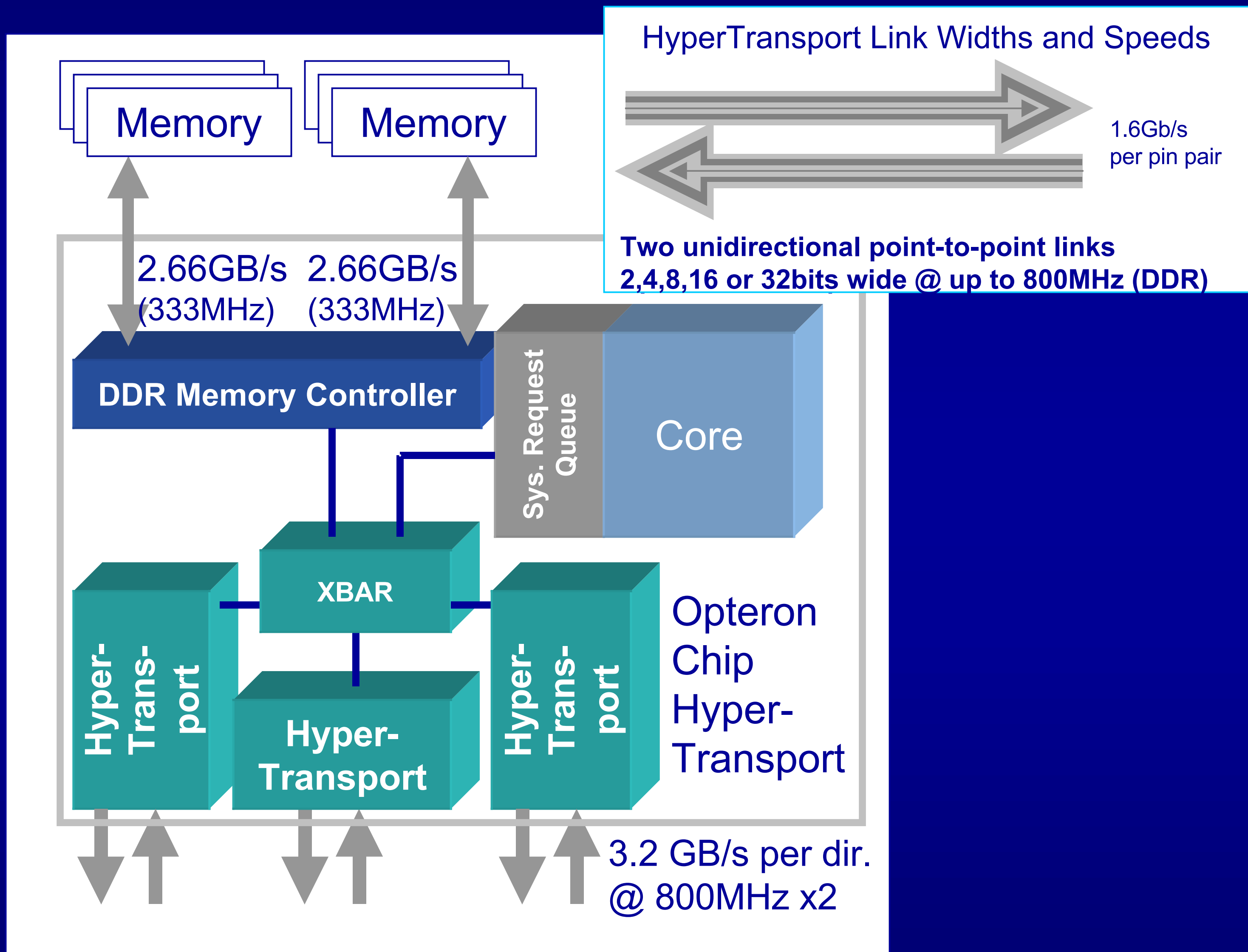
Dual processor scoreboard for HPC Applications:

Single	Dual	Item
☐	☒	Peak performance (TFLOP)
☐	☒	Cost Per Processor
		Memory Subsystem
☒	☐	No shared bus system
☒	☐	No Coherence in Caches (processor and “northbridge” & OS)
☒	☐	No False Sharing
☒	☐	Memory Size
		Message Passing
☒	☐	No Shared interconnect adapters
☐	☒	On-node MPI performance
☐	☐	I/O Performance
☒	☐	Local
☒	☐	Parallel

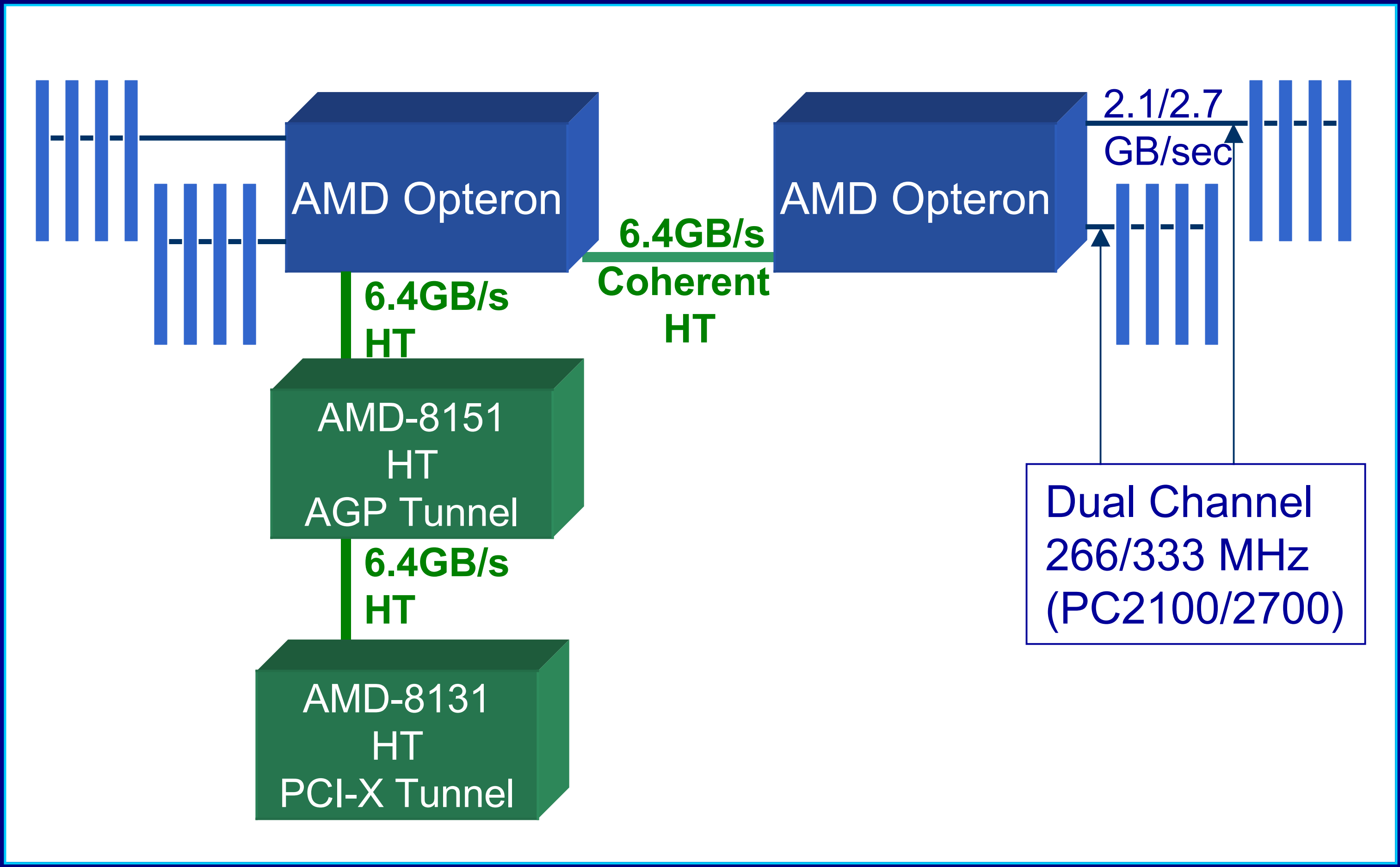
Intel Architecture



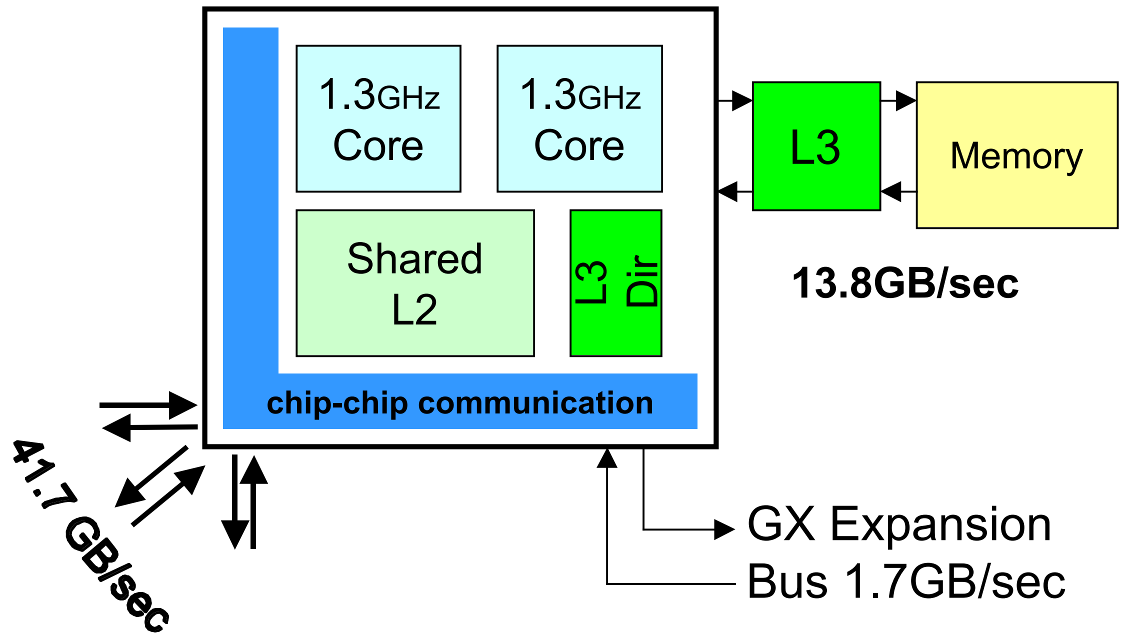
Intel Architecture



AMD Architecture



IBM Power4



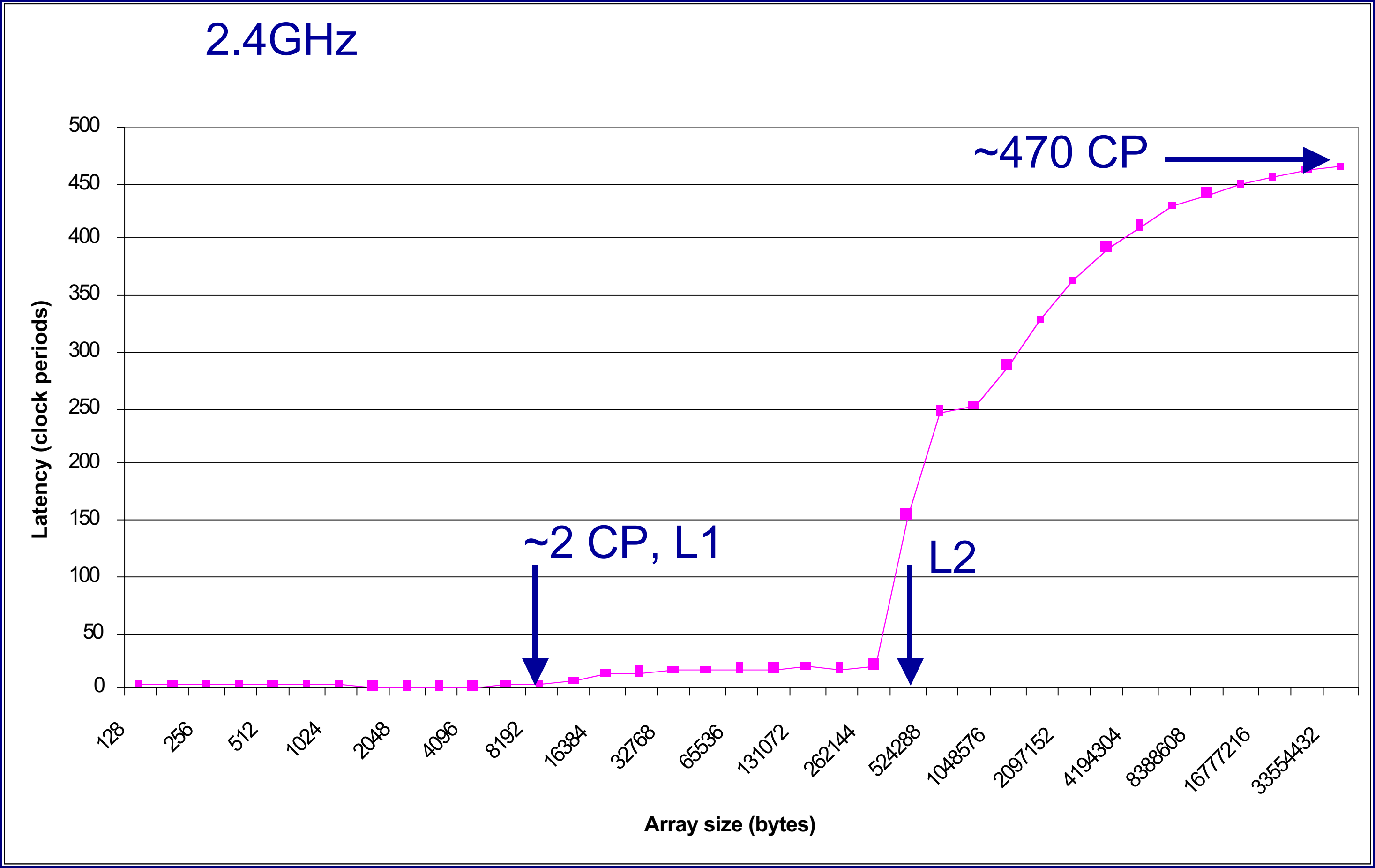
1.3GHz Core
1.3GHz Core
Shared L2
L3 Dir
chip-chip communication
L3
Memory
13.8GB/sec
41.7 GB/sec
GX Expansion
Bus 1.7GB/sec

Memory Latency

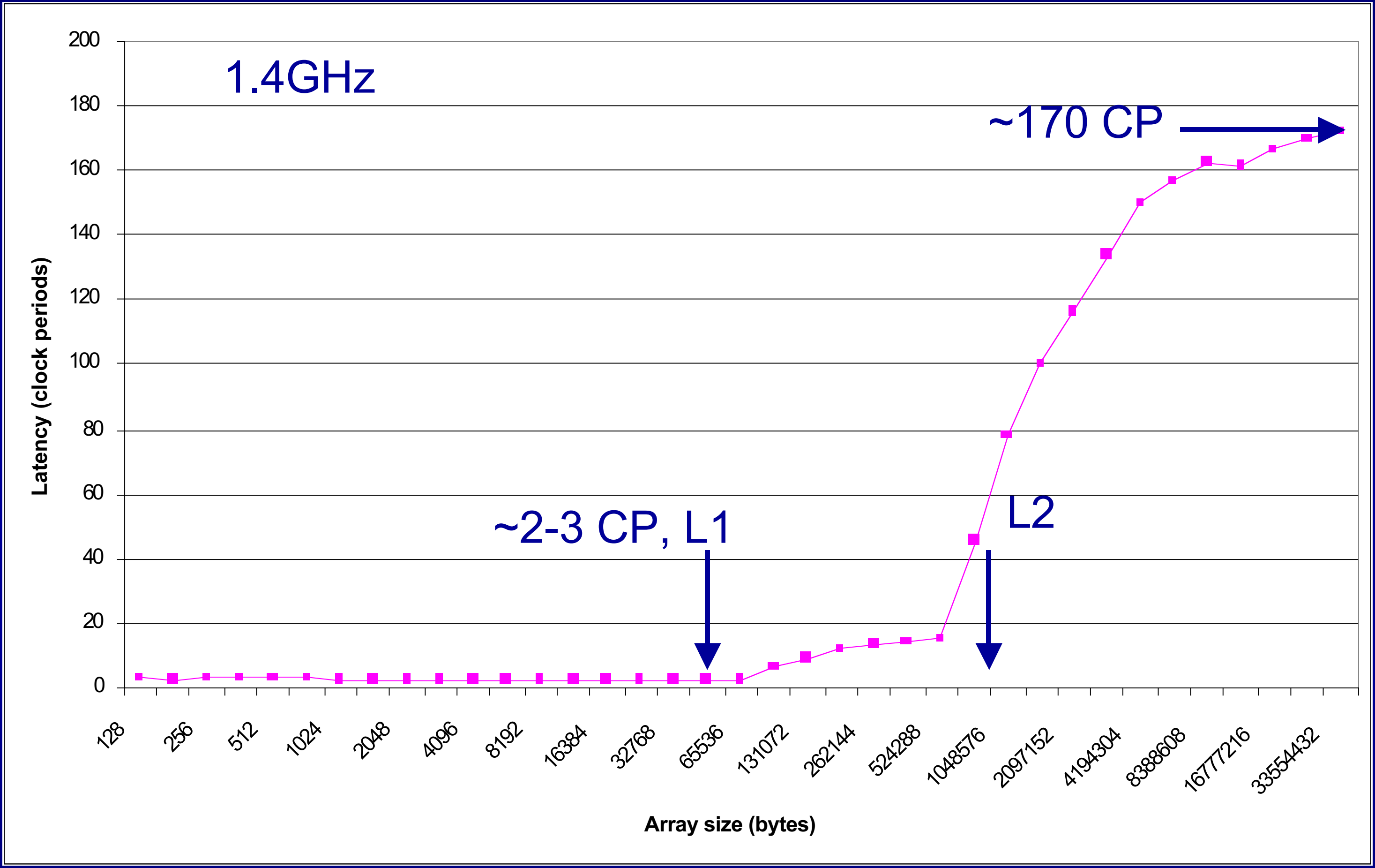
```
I1  =  IA(1)
DO  I  =  2,N
    I2  =  IA(I1)
    I1  =  I2
END  DO
```

1.) Load IA with sequence 1→N.
2.) Randomize IA entries.
3.) Measure Clock Periods of loop.
    (CPs/N = single memory access time = latency)
4.) Loop does not optimizes: no prefetching or streams

Memory Latency Xeon



Memory Latency AMD

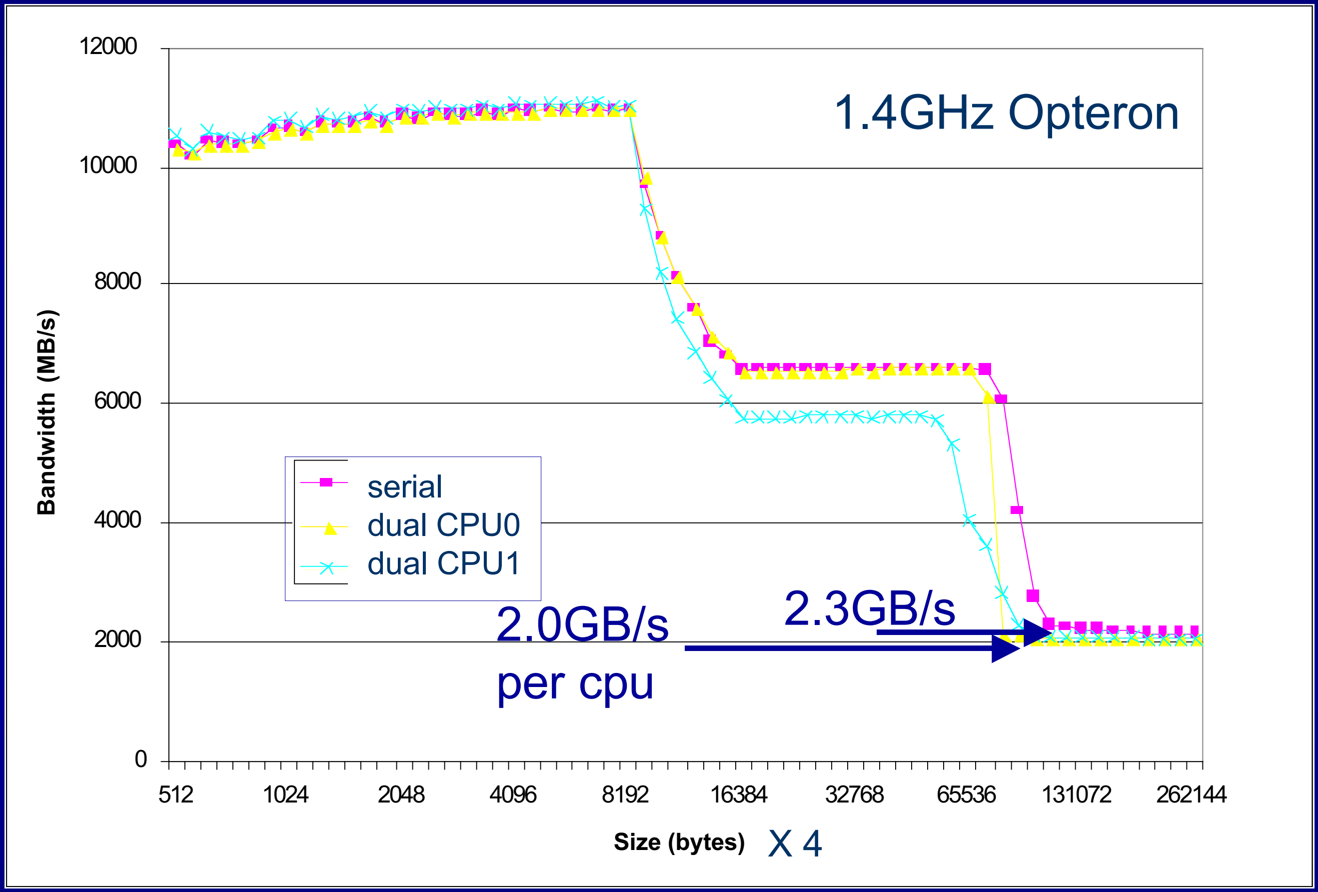


Memory Bandwidth

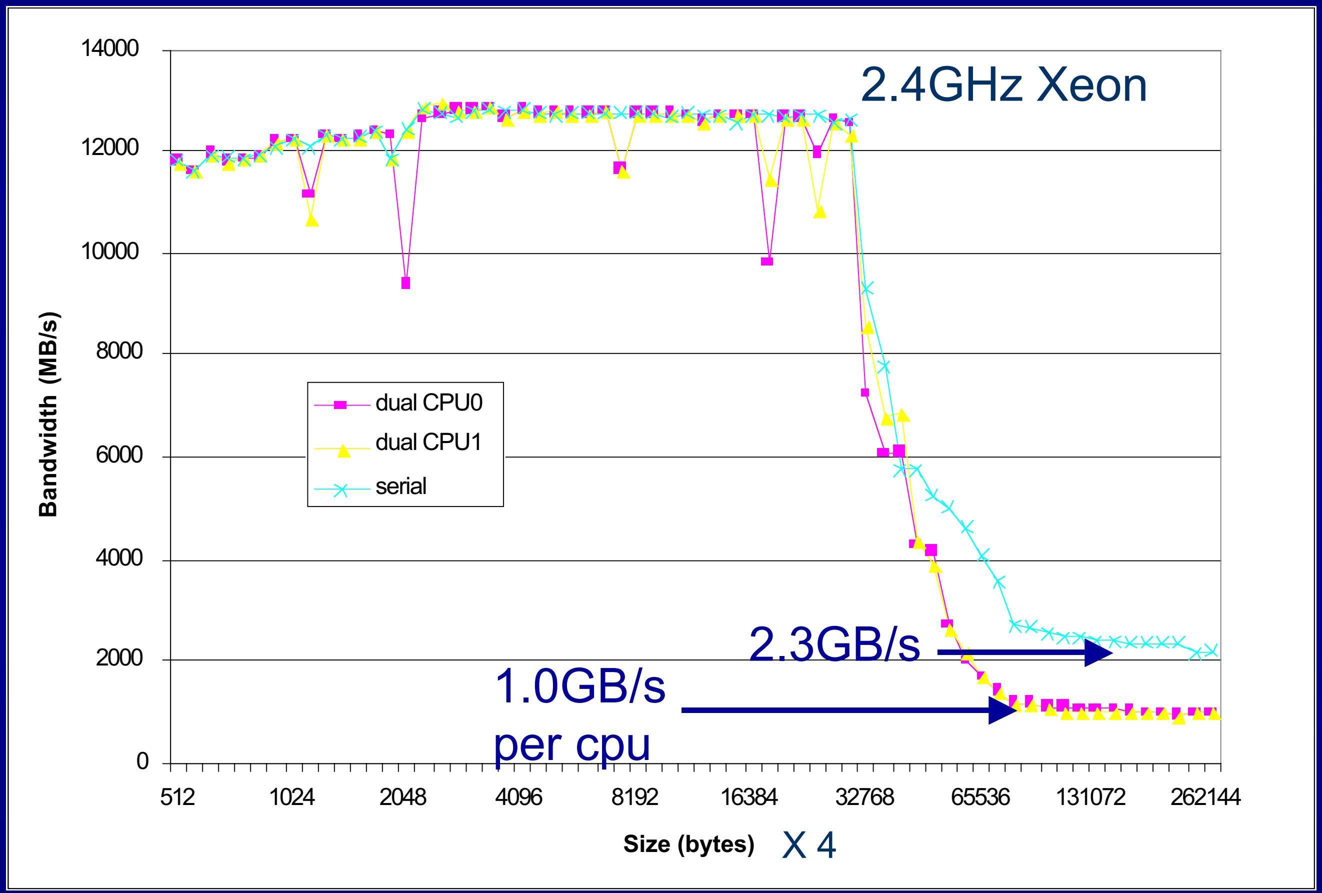
```
DO I = 1,N
   S = S + A(I)
   T = T + B(I)
END DO
```

1.) -O3, unrolling = 2
2.) Two streams—
    gives high, reasonable bandwidths expected across memory & caches

AMD SP/DP memory bandwidth



Xeon SP/DP memory bandwidth



STREAM Results

Kernel	Intel Xeon	AMD Opteron
Copy	1213	2162
Scale	1206	2093
Add	1381	2341
Triad	1375	2411

Serial Execution, (MB/sec).

Kernel	Intel Xeon	AMD Opteron
Copy	1167	3934
Scale	1162	4087
Add	1273	4561
Triad	1281	4529

Parallel Execution, two threads (MB/sec).

MPI On-Node Bandwidth

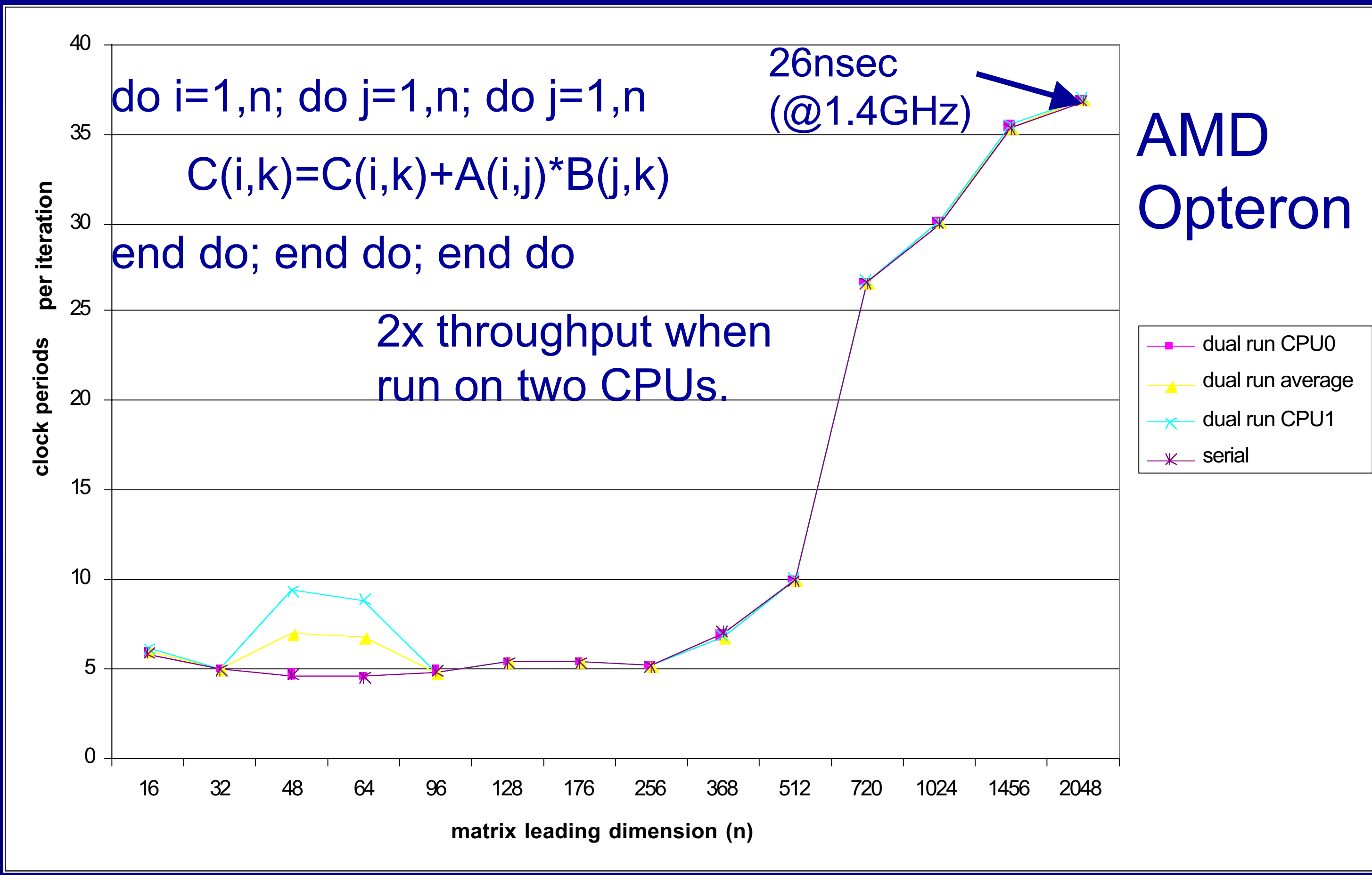
It should be faster than node-to-node.

	(MB/sec)
DELL 2650	295 @ 2MB
Opteron Suse-64 ch_p4	172 @ 2MB
Opteron Suse-64 ch_shmem	404 @ 2MB
IBM P690 HPC	1324 @ 2MB
IBM P690 Turbo	1398 @ 2MB
IBM P655 HPC	1684 @ 2MB

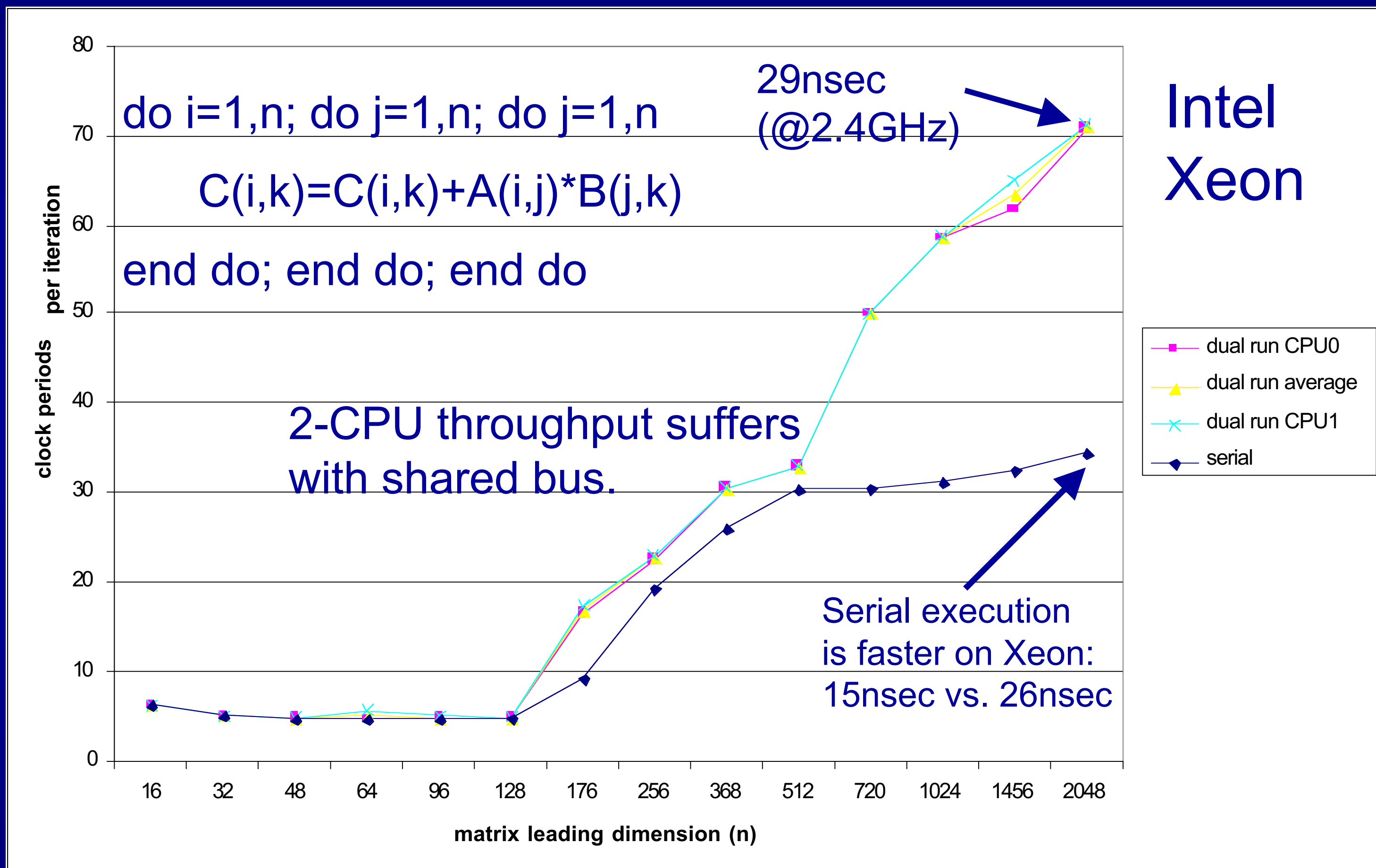
Different implementations of MPI will vary with On-Node Performance.

Hand Coded Matrix-Matrix Multiply

Accesses Memory with 1 stream and 1 strided pattern.
(Don't do this at home in your optimized code ☺.)



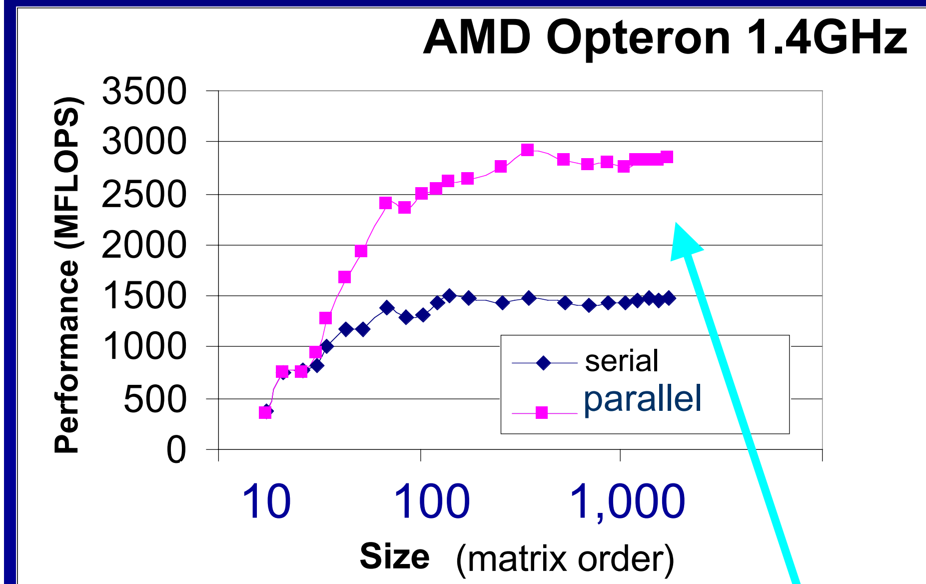
Hand Coded Matrix-Matrix Multiply

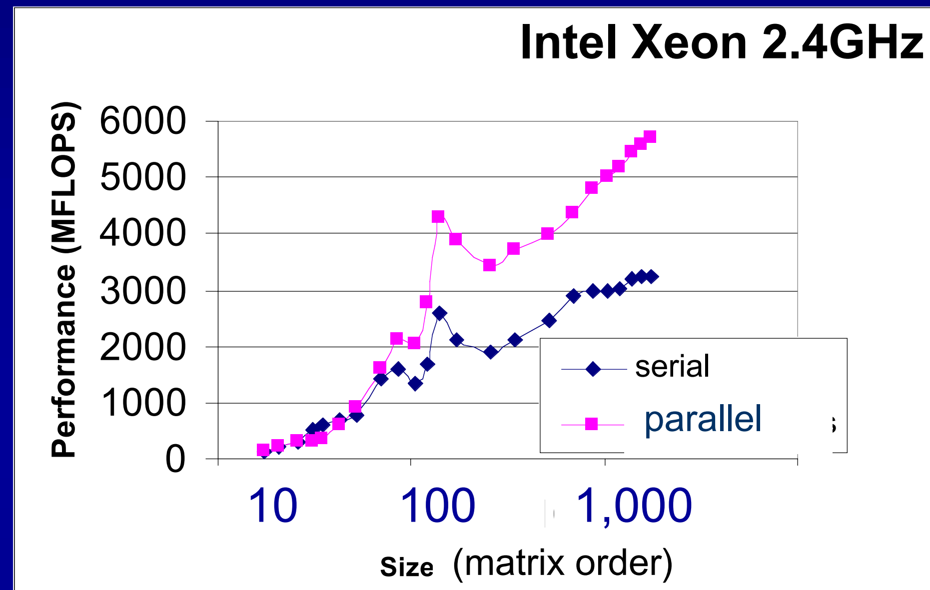


2-CPU throughput suffers
with shared bus.

Library Matrix-Matrix Multiply (DGEMM)

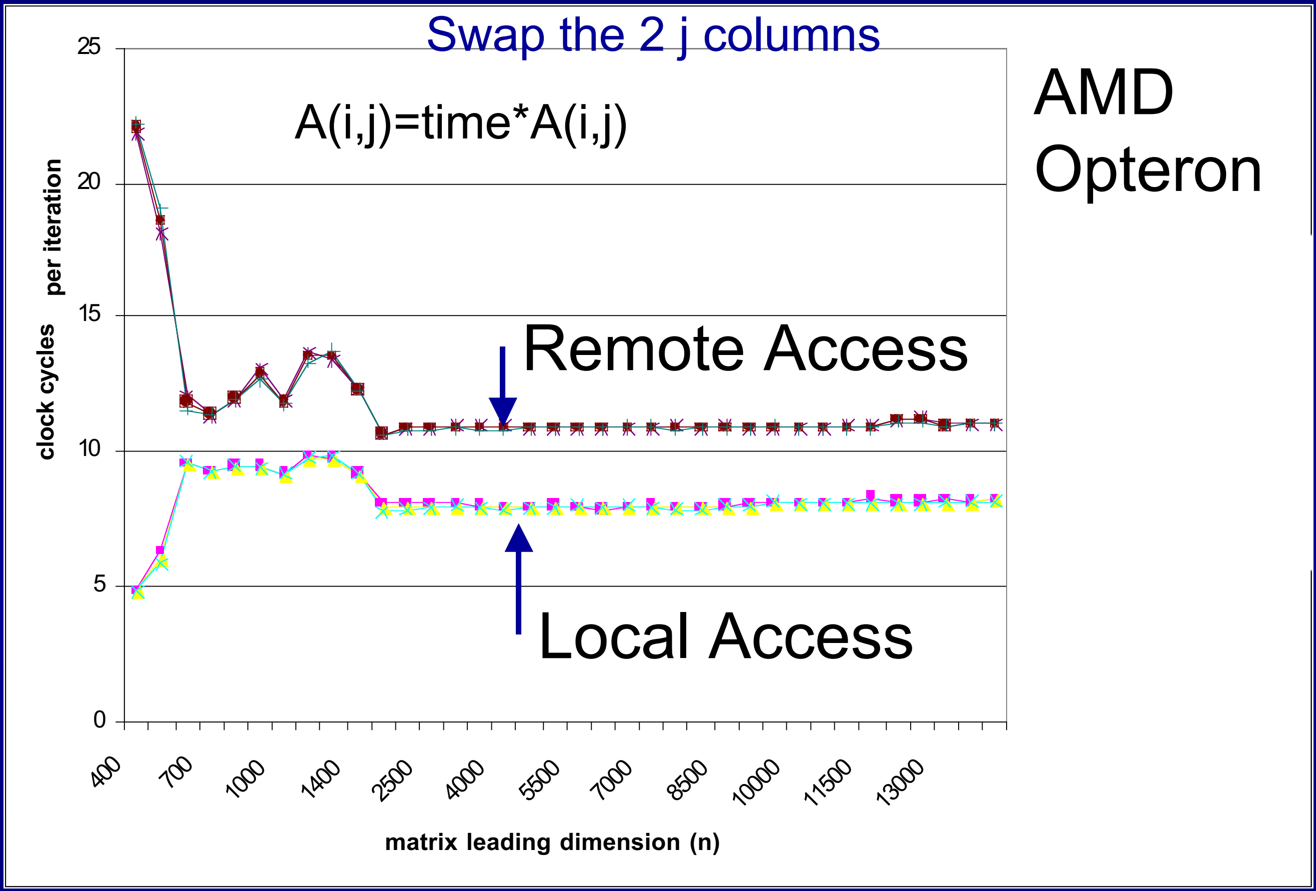
MKL 5.1 Library





May be much higher with Opteron-optimized Libs – (e.g., NAG Lib.)

Remote & Local Memory Read/Write



1.) Each processor writes a column to local memory.
2.) Each processor reads/writes to same column. (Local Access)
3.) Each processor "swaps" column index and reads/writes to remote memory. (Remote Access)

Applications

- SM: Stommel model of ocean "circulation" ; solves 2-D partial differential equation.
  - Uses Finite Difference approx for derivatives on discretized domain, (timed for a constant number of Jacobi iterations).

Memory Intensive

Platform	Serial SM (sec)	Parallel SM (sec)
AMD Opteron 2P	68.0	43.4
Intel Xeon 2P	57.5	63.4

- MD: Molecular Dynamics of argon lattice.
  - Uses Verlet algorithm for propagation (displacement & velocities).

Compute Intensive

Platform	Serial MD (sec)	Parallel MD (sec)
AMD Opteron 2P	9.48	5.3
Intel Xeon 2P	7.14	5.4

Summary

	SERIAL		Parallel	
	Opteron	Xeon	Opteron	Xeon
Latency	Low	High	Overlapped	Overlapped
Band-width	~2GB/s	~2GB/s	2x	1x
MxM (per CP)	Lower	Higher	2x mem	1x mem
MXM (time)	Higher	Lower	slightly lower	slightly higher
DGEMM	Low	High	Scale: 1.9x (MKL 5.1 not optimized for AMD)	Scale:1.8x 2x Opteron performance

Summary

- Performance of dual-processor systems varies with memory architecture and processor speed.
  - AMD memory bandwidth scales by 2x when second processor is used– (using “local” memory).
  - Xeon memory bandwidth is shared by second processor.
  - Xeon outperforms Opteron on serial compute-intensive codes (due to speed: 2.4GHz Xeon vs. 1.4GHz Opteron); but lead can be eliminated with dual-processor execution of (parallel) programs when memory bandwidths & synchronizations are involved.