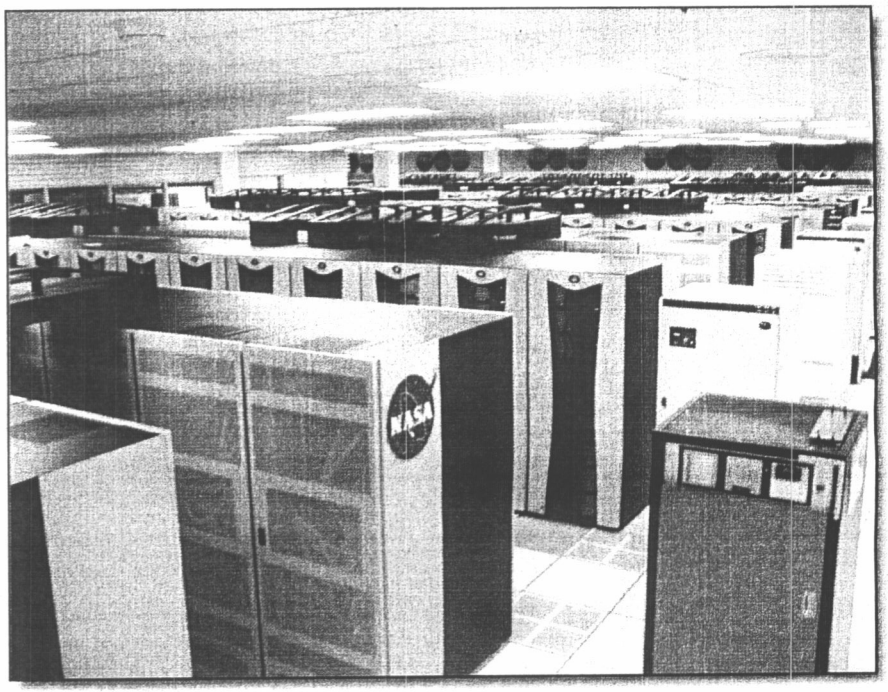
OpenMP Performance on the Columbia Supercomputer

Haoqiang Jin and Robert Hood
{hjin,rhood}@nas.nasa.gov
NASA Advanced Supercomputing Division
NASA Ames Research Center

NAS
NASA ADVANCED SUPERCOMPUTING

Columbia: World Class Supercomputing

- One of the world's fastest supercomputers providing 61 TFLOPs (10/20/04)
- Conceived, designed, built, and deployed in just 120 days
- A 20-node supercomputer built on proven 512-processor nodes
- Largest SGI system in the world with over 10,000 Intel Itanium 2 processors
- Provides the largest node size incorporating commodity parts (512) and the largest shared-memory environment (2048)
- 88% efficiency tops the scalar systems on the Top500 list



Systems: SGI Altix 3700 and 3700-BX2
Processors: 10,240 Intel Itanium 2
Global Shared Memory: 20 Terabytes

Front-End: SGI Altix 3700 (64 proc.)
Online Storage: 440 Terabytes RAID
Offline Storage: 6 Petabytes STK Silo

Internal Networks:
Internode Comm: Infiniband
Data Transfer: Gigabit Ethernet
Hi-Speed: 10 Gigabit Ethernet

NASA

NAS
NASA Advanced Supercomputing

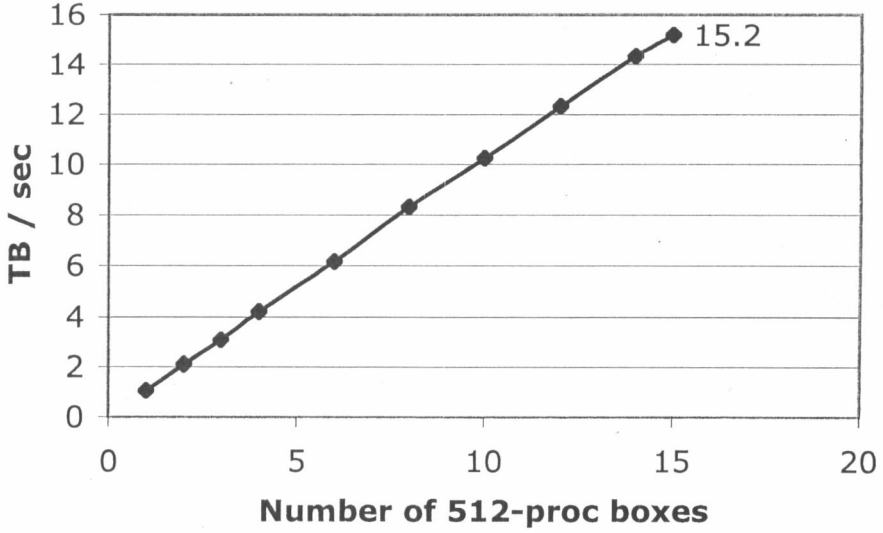
STREAM Benchmark

- Measures sustainable memory bandwidth
  - Using vector-like operations
- A hybrid version
  - MPI across Altix boxes, OpenMP within an Altix
- In November `03: achieved > 1 TB/sec on 512-processor Altix
  - Still best result submitted to date

- Columbia results
  - one 512-processor box:

Copy	0.90 TB/sec
Scale	0.90 TB/sec
Add	1.08 TB/sec
Triad	1.05 TB/sec

Triad Scaling (other measures similar):



NAS
NASA Advanced Supercomputing

CART3D OpenMP/MPI Scaling
(results for 4.6 and 6.4 million cell shuttle configuration)

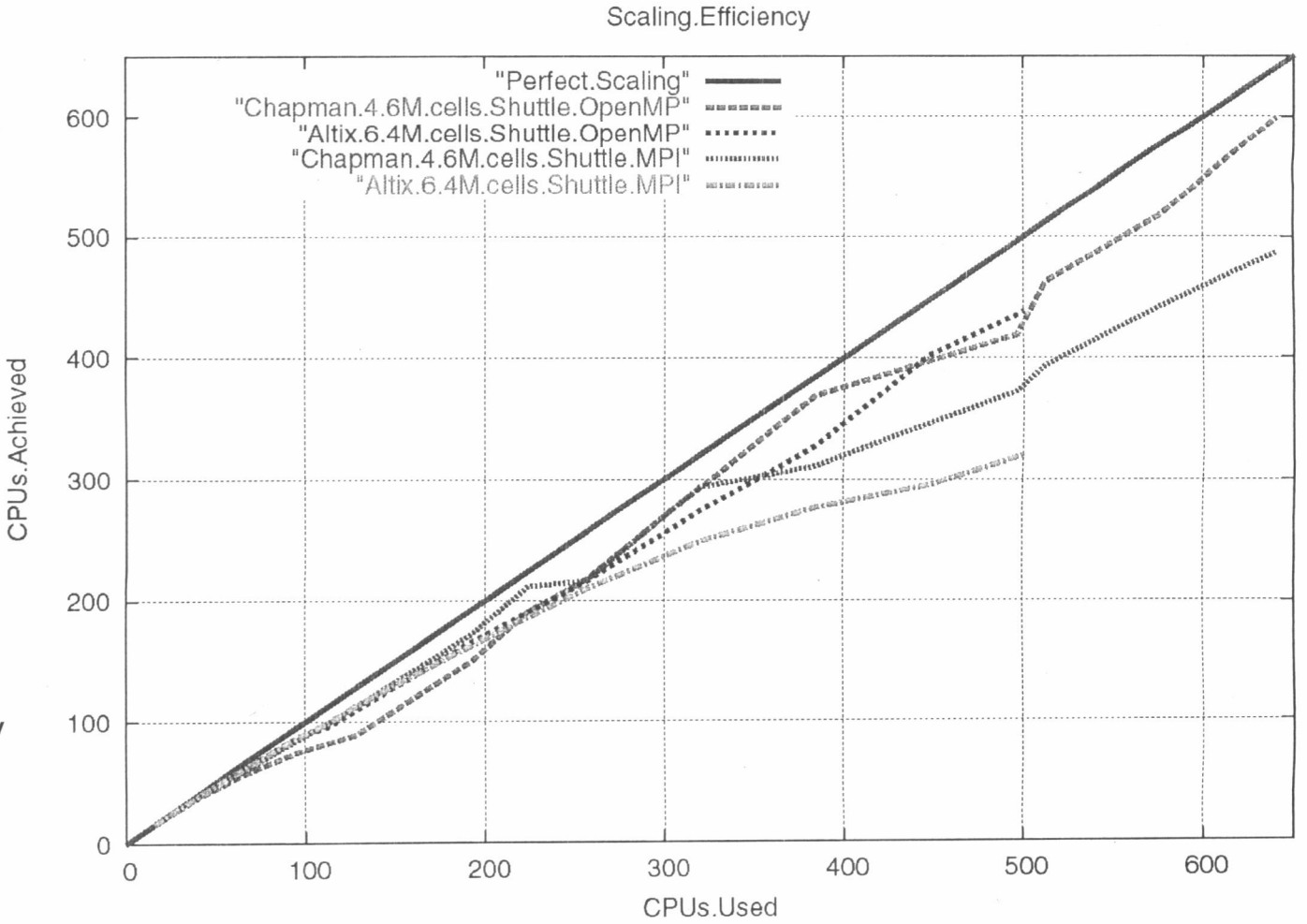
Scaling Efficiency

Altix-512p SSI (6.4 million cells)

	NCPUS	Efficiency
OpenMP	500	88 %
MPI	500	64 %

Origin-1024p SSI (4.6 million cells)

	NCPUS	Efficiency
OpenMP	640	94 %
MPI	640	76 %



OpenMP shows best scaling efficiency of 88% on 500 CPUs for Altix System

Source: mike aftosmis@nas.nasa.gov

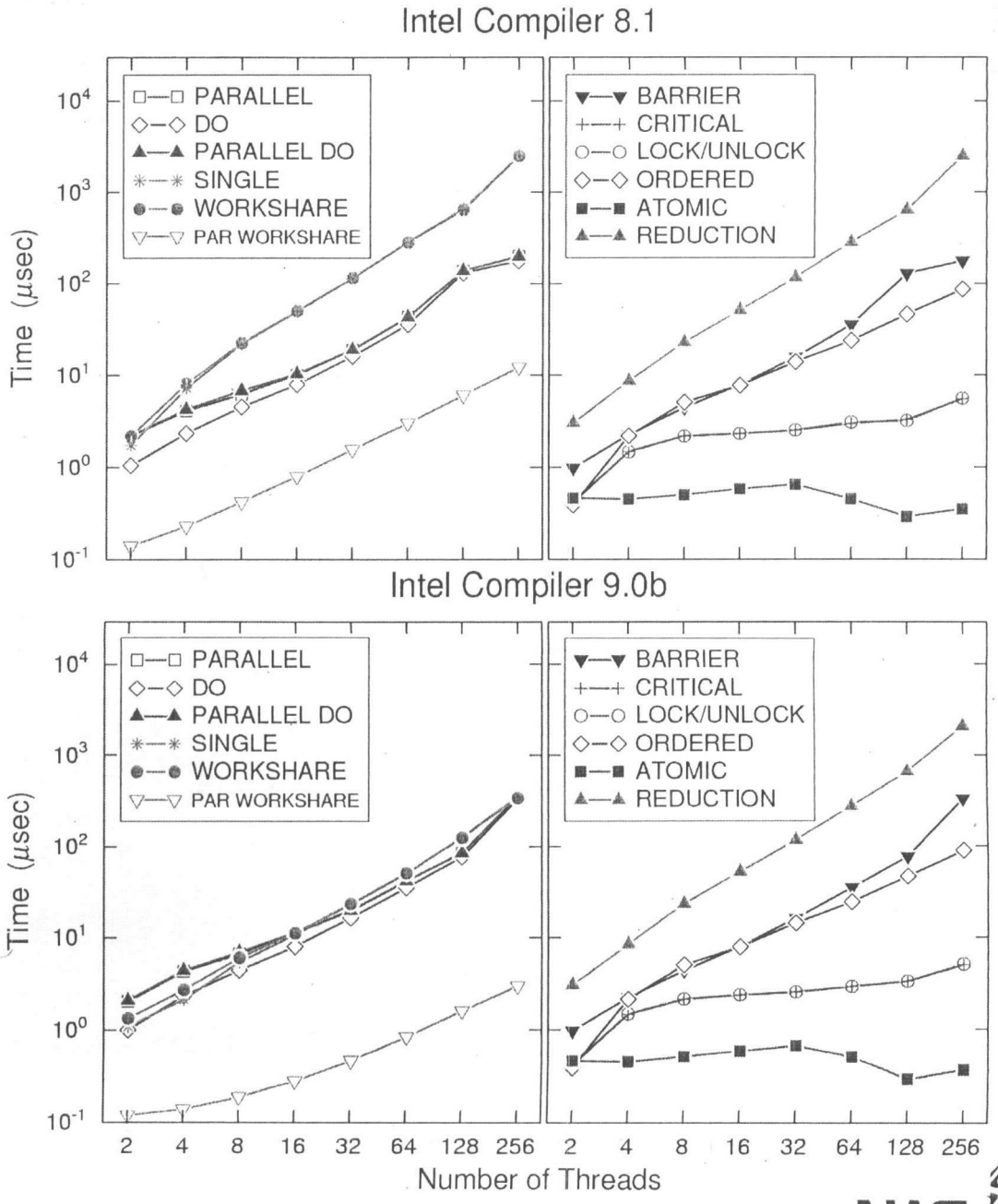
NASA

NAS
NASA Advanced Supercomputing

OpenMP Microbenchmarks

- From EPCC
  - Measure OpenMP overheads
- Two versions of the Intel compiler
  - 8.1 and 9.0beta

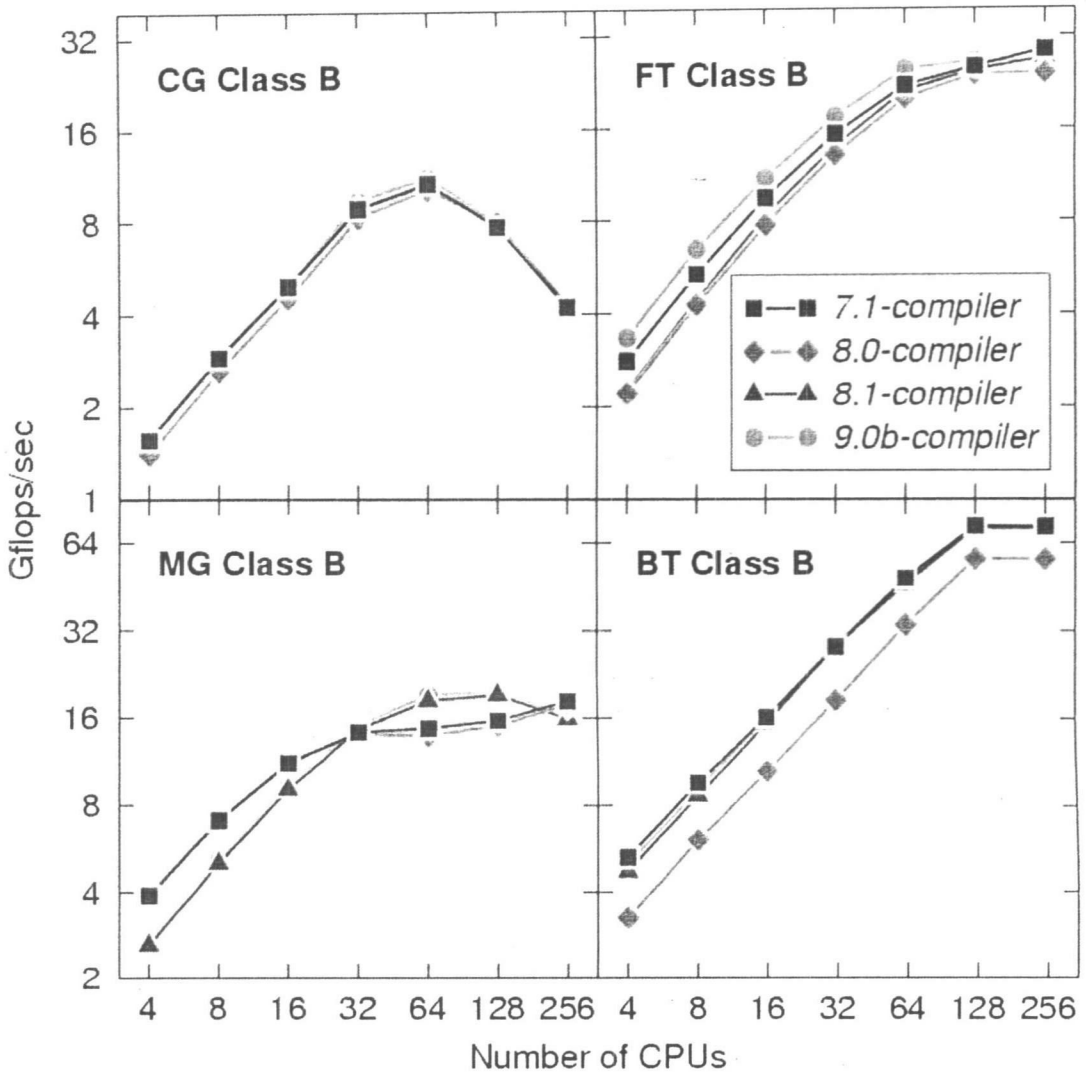
- 9.0beta improves scaling of SINGLE and WORKSHARE over 8.1
- REDUCTION could be improved



NAS Parallel Benchmarks

- Mimic computation and data movement in computational fluid dynamics (CFD)
- Single zone and multi-zone versions (NPB and NPB-MZ)
- Programming paradigms
  - MPI, OpenMP, hybrid

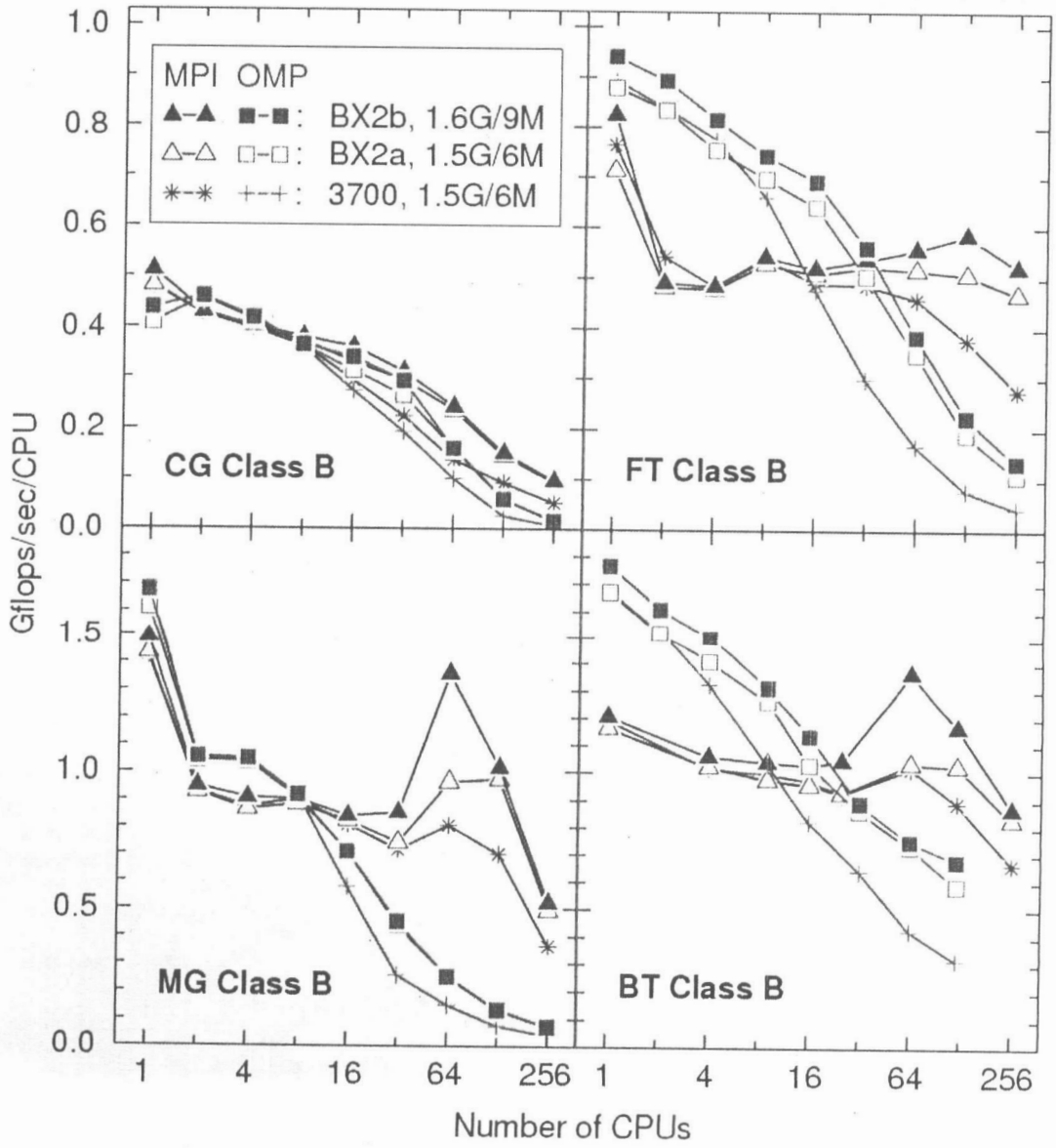
- Four OpenMP benchmarks from NPB3.2-OMP
  - CG, MG, FT, BT
- Four different versions of the Intel compiler
  - 7.1, 8.0, 8.1, 9.0beta
- Performance on one Altix node



NAS
NASA ADVANCED SUPERCOMPUTING

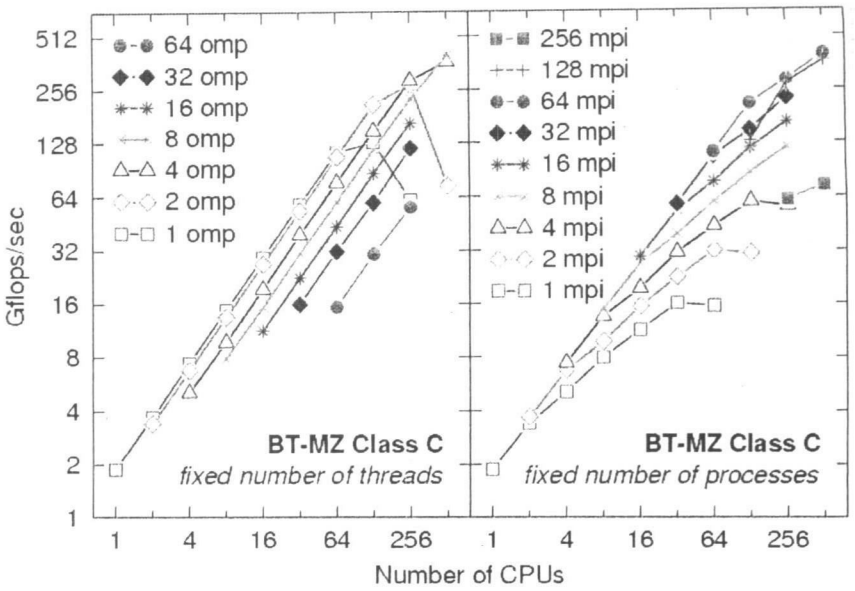
Altix 3700 Versus 3700-BX2

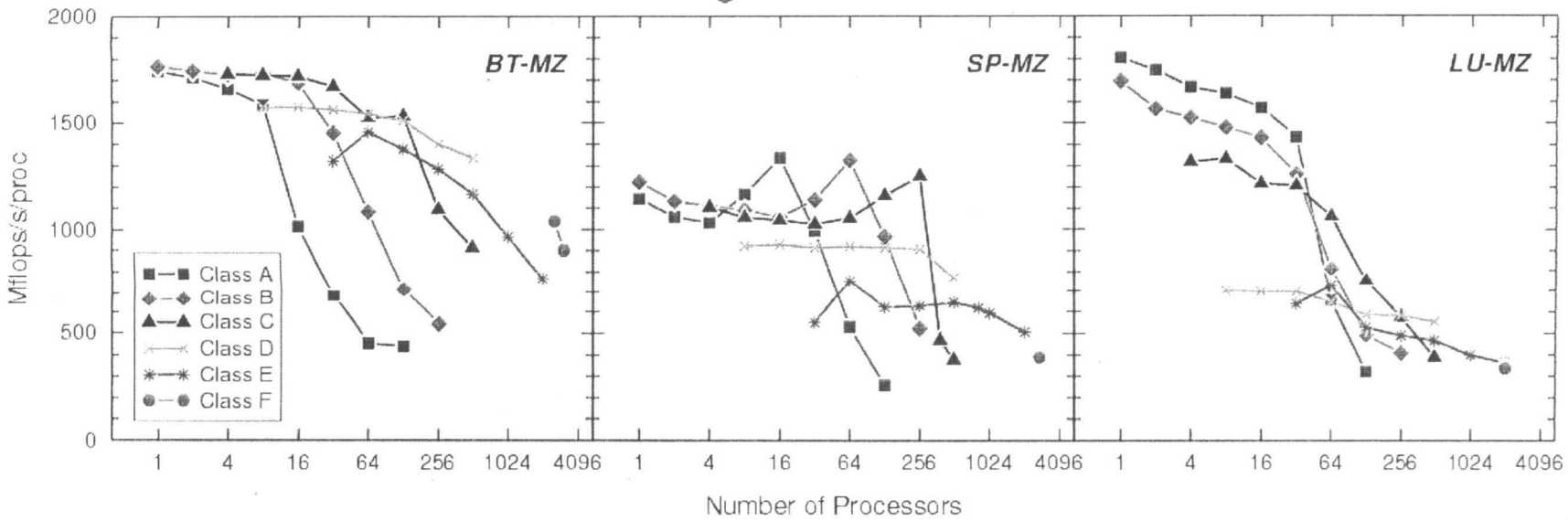
- Three different Altix nodes on Columbia
  - 3700: 1.5GHz, 6M L3
  - BX2a: 1.5GHz, 6M L3
  - BX2b: 1.6GHz, 9M L3
- Four benchmarks from NPB-MPI and NPB-OMP
  - CG, MG, FT, BT
- Double bandwidth of the BX2 nodes has large performance impact on OpenMP
- OpenMP performed better for small number of CPUs but MPI scaled better
- Larger cache resulted in better MPI performance around 64 CPUs, but less visible in OpenMP



Hybrid MPI+OpenMP

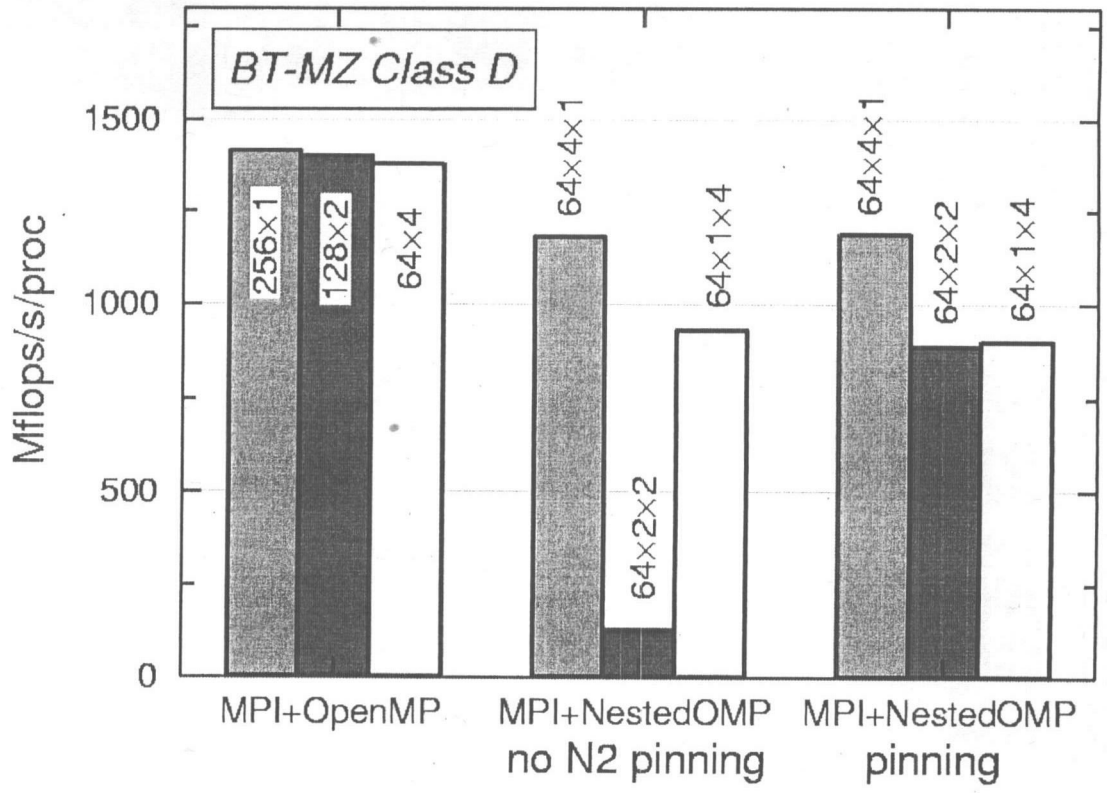
- Multi-zone versions of NPB
  - LU-MZ: 4x4 zones, equal size
  - SP-MZ: #zones grows, equal size
  - BT-MZ: #zones grows, varying size
- Hybrid parallelization
  - MPI across zones, OpenMP in zone
- Single-node performance
  - MPI versus OpenMP
- Multiple-node performance
  - best achieved





Nested OpenMP for Multi-level Parallelism

- MPI+Nested-OpenMP in BT-MZ
  - MPI for inter-zone parallelism
  - Nested OpenMP
    - Outer level (N1) for inter-zone parallelism
    - Inner level (N2) for loop-level parallelism within each zone
- Compare with the standard hybrid version
  - MPI across zones, OpenMP within a zone
  - Intel 8.1 compiler



- Nested OpenMP introduced additional overhead
- Thread-to-processor binding (pinning)
  - required for hybrid codes for better performance
  - very important for nested OpenMP (more than a factor of 6 improvement)

Notation: N_{mpi} x N_{N1_omp} x N_{N2_omp}