Presented at the Asian Technology Information Program (ATIP) 3rd China HPC Workshop
SC07 Conference, Reno, Nevada, November 11, 2007

PBB: A Parallel Bioinformatics Benchmark Suite for Shared Memory Multiprocessors

Wenguang Chen,
Chuntao Hong

Tsinghua University

ATIP 3rd China HPC Workshop, November 11, 2007, Reno, NV

PBB: A Parallel Bioinformatics Benchmark Suite for Shared Memory Multiprocessors

CHEN Wenguang
HPC Inst., CS Dept., Tsinghua University

Outlines

- Motivation
- Benchmark selection & construction
- Benchmark characteristics
- Performance results
- Conclusions, Q&A

Motivation

- Wide use of bioinformatics applications
- The trend of multi-core

  => There should be a parallel bioinformatics benchmark
- SPEC CPU2000 may not match the characteristics of bioinformatics workloads well
- Existing bioinformatics benchmarks are not satisfactory

  => We need a new one

Existing Bioinformatics Benchmarks:

BioBench:
- does not cover some important domains
- no parallel program

BioPerf:
- includes only one parallel benchmark

BioParallel:

our previous work, includes 5 parallel applications

PBB benchmark suite:

Being more complete

7 applications covering 7 of the most important domains of bioinformatics

Keeping pace with the changing world

all the applications are parallelized

Benchmark Selection & Construction

1. Identify the most important application domains
2. Choose representative applications for each domain

   most popular, most advanced
3. Benchmark optimization & parallelization

The 7 applications:

1. Pairwise sequence alignment: BLAST-P
2. Global alignment: PLSA
3. Multiple sequences alignment: MUSCLE
4. Protein 3D structure prediction: Rosetta
5. Phylogenetic tree reconstruction: SEMPHY
6. Gene regulatory network learning: ModuleNet
7. Pattern study of Single Nucleotide Polymorphisms: SNP

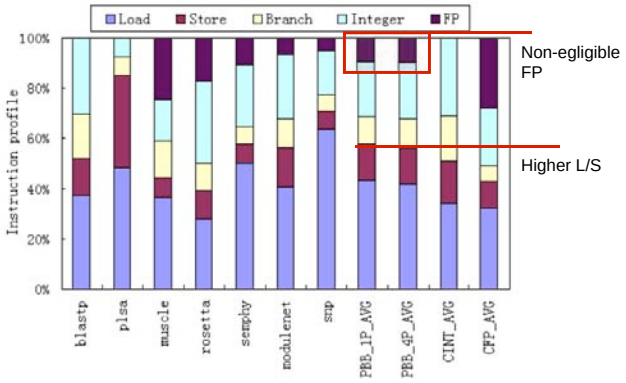
Benchmark Characteristics

Systems Used:

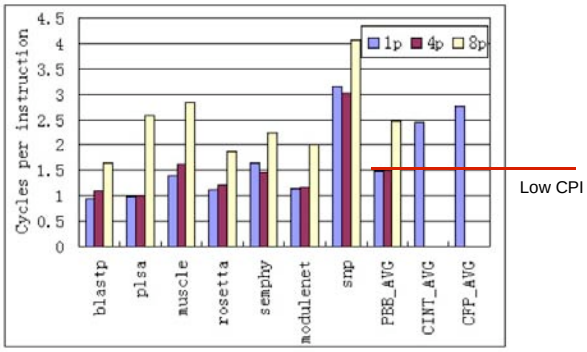
	RefSys	QP001	DC001	Hydra	Osprey	Unisys-Es700
CPU Name	PIII Xeon	XEON	Dual-Core Xeon	Itanium 2	Itanium 2	XEON
CPU Freq.	700MHz	2.8GHz	3.2GHz	1.3GHz	1.5GHz	3.0GHz
L1 D-cache	16KB 4-way	8KB 4-way	16KB 8-way	16KB	16KB	8KB
L2 Cache	1MB 8-way	512KB 8-way	4MB 8-way	256KB	256KB	512KB
L3 Cache	-	2MB 8-way	-	3MB	6MB	4MB
L4 Cache	-	-	-	-	-	32MB
# of Chips	4	4	2	4	4	16
HT Support	N	Y	Y	N	N	Y
Interconnect	FSB	FSB	FSB	FSB	FSB	FSB and Crossbar
Memory	1GB	4GB	4GB	4GB	16GB	8GB

Workload analysis is performed on QP001

Instruction profile:

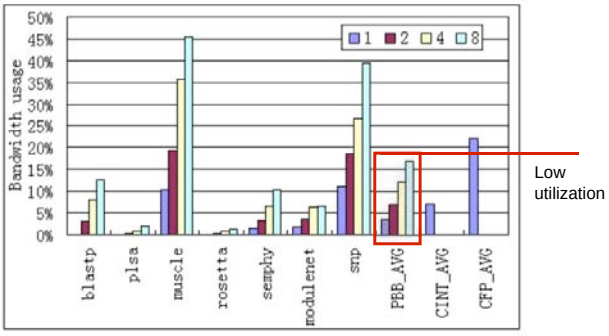


CPI



8P means QP001 with HT enabled

FSB bandwidth utilization



Performance Results

Benchmark scores:

$$PBB\_Score = \sqrt[6]{\prod_{1 \le i \le 6} \frac{r_i}{m_i}} \times 100$$

r_i : time used to run application i on the reference system

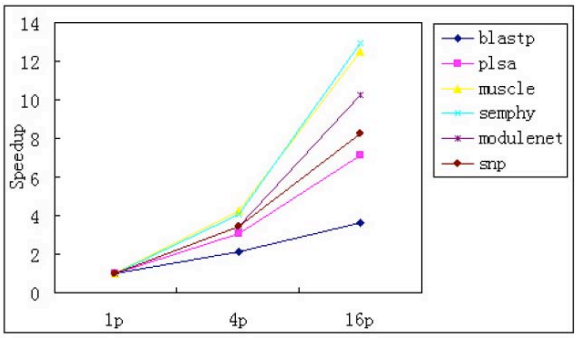
m_i : time used to run application i on the tested system

Rosetta is excluded for it produces random results

Scores of PBB

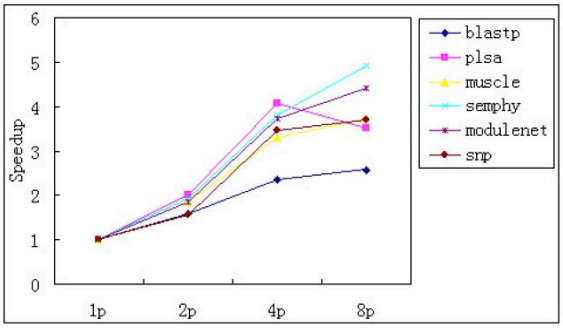
App.	Ref	QP001	DC001	Hydra	Osprey	Unisys-Es700
Plsa	114.9	45.4	32.0	55.6	48.0	19.3
Muscle	370.6	75.0	53.0	114	93.0	21.0
Semphy	114	25.2	21.5	49.8	32.2	22.3
Modulenet	68.1	18.2	15.1	62.0	54.8	7.58
Blast-p	308.3	80.0	75.0	87.5	76.0	54.7
SNP	108	27.9	14.7	77.9	68.8	15.7
PBB Scores	100	383	503	208.5	254	756

Parallel speedup



Tested on Unisys-ES700 with 16 Xeon

Hyper Threading Effects



Conclusions

Workload characteristics:

1. High percentage of load/store instructions
2. Non-negligible floating point instructions, but still significantly lower than SPECCPU 2000FP
3. Low CPI
4. Low memory bandwidth demand

Thanks

Any questions?