Simulation and Modeling to Enhance Performance of Multicore Processor Systems

Pat Teller, Sarala Arunagiri, Mitesh Meswani
The University of Texas at El Paso

Jeanine Cook, Dustin Balise, Walleed Kohlani
New Mexico State University

Army High Performance Computing Research Center

High End Computing along the Rio Grande

Details of Project

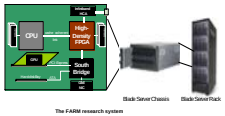
Collaboration among UTEP, NMSU, and Stanford University as a part of AHPCRC TA4, HPC Enabling Technologies and Advanced Algorithmic Development

Critical Army Challenges:

Given the new era of "pervasive parallelism", computing systems and advanced algorithm capabilities need to be developed

Army Mission Impact:

Efficient and reliable HPC libraries and computational tools will have a significant impact on the engineering and scientific communities of the Army and DoD.



Key TA4 Projects:

- Stream programming for High Performance Computing
- Simulation and Modeling to Enhance Performance of Multicore Processor Systems
- High Performance Optimization Library
- Flexible Architecture Machine
- Hybrid Optimization Schemes for Parameter Estimation Problems

Projects at The University of Texas-El Paso:

1. Simulation and Modeling to Enhance Performance of Multicore Processor Systems
2. Hybrid Optimization Schemes for Parameter Estimation Problems
3. Education and Outreach

Project Goal

Broad Goal Of Simulation Project

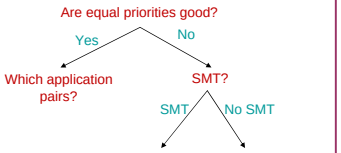
Propose code and/or operating system (scheduling) modifications to improve performance of Army applications executed on emerging technologies

PART 1: Performance of Simultaneous Multithreaded (SMT) Processor

Goal:

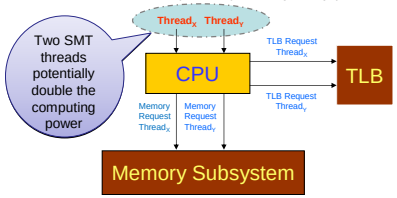
Improve Simultaneous Multithreaded (SMT) processor throughput using hardware thread priorities and SMT mode.

Questions investigated:



Simultaneous Multithreading (SMT)

Two or more threads simultaneously execute independent instruction streams, potentially doubling throughput.



SMT System with Two Hardware Threads

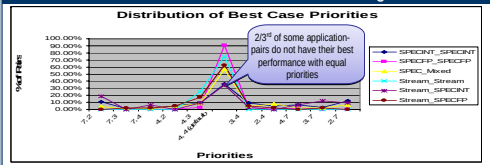
Problems due to Interference

Sharable resources:
- Microarchitecture/Pipeline
- Memory (competing working sets)
- TLB (virtual-to-real address translation)

Sharing characteristics:
- Interference depends on characteristics of applications
- Sharing at cycle-level granularity

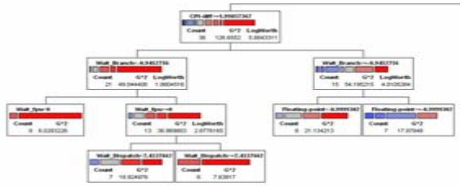
Performance effects:
- Reduced SMT throughput
- In extreme cases, throughput less than processor without SMT (Single-Threaded)

Results - Processor Utilization in SMT wrt Differing Thread Priorities



Application-Pair Class	% Minimum CPI Difference between Default and Best Case	% Maximum CPI Difference between Default and Best Case	% Minimum CPI Difference between Worst and Best Case	% Maximum CPI Difference between Worst and Best Case
SPECINT_SPECINT	0%	25.85%	0.51%	47.94%
SPECFP_SPECFP	0%	2.76%	7.41%	40.34%
SPEC_Mixed	0%	20.60%	3.76%	47.16%
Stream_Stream	0%	0.03%	2.47%	10.72%
Stream_SPECINT	0%	5.12%	0.33%	41.23%
Stream_SPECFP	0%	3.53%	7.05%	53.51%

Snapshot of a Decision Tree



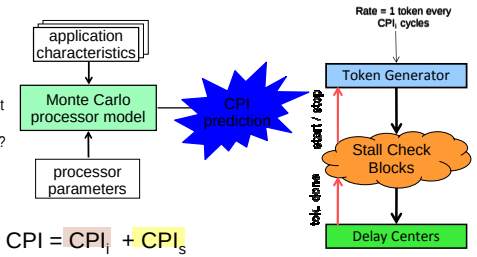
Results

SMT mode prediction: 97.5% accurate
SMT thread priority prediction: 83% accurate

PART 2: NMSU/LANL Monte Carlo Processor-Performance Model

Model Development Goal

- Fast Performance Prediction
- Fast Design Space Exploration
  - What if the double-precision FP unit is fully pipelined?
  - Should we improve the branch unit?
- Fast Bottleneck Analysis
  - Which processor component hurts performance the most?



$$CPI = CPI_i + CPI_s$$

Modeling Tools

- IBM Cell, Sun Niagara, and Intel Itanium Monte Carlo models (in-order processors)
  - Accurately predicts performance (to within 10%) on suite of scientific applications
- Tools for parameter collection on AMD Opteron platform

Current and Future Work

Work In Progress

- Extend Monte Carlo processor modeling methodology to model
  - (a) Out-of-order processors (targeting the Opteron)
  - (b) Communication between multicores
- Build Niagara 2 Monte Carlo model
- Investigate use of GPUs for general-purpose computation
- Gather application characteristics for DoD HPCMC FY-07 Benchmarks
- Find good analytical models of cache performance based on its parameters: associativity, cache size, and cache replacement policies
- Refine Monte Carlo simulation model to include cache and TLB behavior depending on cache parameters
- Use analytical modeling tools to investigate tradeoffs leading to improved performance or productivity in multicore systems

Future Work

We plan to extend our SMT research in the foillowing directions:
- Validate models using other benchmarks, in particular DoD benchmarks
- Improve model accuracy
- Extend the methodology to dynamic adaptation
- Extend the methodology for multicore processors with SMT

Plans for Monte Carlo simulation work
- Extend the model to SMT processors
- Include cache interference behavior into the Monte Carlo simulation model for SMT processors

Contact information

Patricia J. Teller
Computer Science Department
Email: pteller@utep.edu

Acknowledgements

ARL Grant No. W11NF-07-2-0027