EE 382M COMPUTER PERFORMANCE EVALUATION AND BENCHMARKING

Fall 2014 T-Th 11:00 am – 12:30 pm Unique 17395 PHR 2.116

Instructor: Dr. Lizy Kurian John
Office: POB 3.114
Office Phone: 232-1455
Office Hours: T Th 9:00-10:30am
e-mail: ljohn at ece dot utexas dot edu
TA/Grader: Jee Ho Ryoo (jr45842 at utexas dot edu)
TA Office: POB 3.142
TA Office Hours: MWF 9:00-10:30am

Course home page will be on Canvas (go to courses.utexas.edu)

Course Description: Evaluating computer architectures have become extremely difficult due to the complexity of the systems/processors and the complexity of the applications that run on the computers. This course will focus on techniques to evaluate performance and power/energy consumption of microprocessors and computer systems. Course notes and several papers from the computer architecture, performance evaluation and workload characterization related conferences will be used as course material.

Prerequisites:

EE 360N/460N/382N.1 - Computer architecture. If you did not take this course at UT, please see me with info on the course you took (book, assignments, exams)

Good programming skills (C and Unix) and at least one assembly language

Text Book:

No required text book, but I'll use material from several sources including

Performance Evaluation and Benchmarking, Edited by Lizy John and Lieven Eeckhout, CRC Press, Taylor & Francis, (Optional)

Measuring Computer Performance: A Practitioner's Guide, by David Lilja, Cambridge University Press, 2000 (Optional)

A Collection of Papers from conferences and journals. Will provide list on course web page

Other References:

Computer Organization and Design, Patterson and Hennessy

Computer Architecture: A Quantitative Approach, Hennessey & Patterson

ISCA, HPCA, Micro, ASPLOS, IISWC and ISPASS Proceedings.

Grading Policy:

HW Assignments, Paper Critiques, Lit Survey, Class Participation	30%
Test1	20%
Test2	20%
Project and Project Presentation	30%

A = 92%, A- = 89%, B+= 85%, B=80%, B- = 78%, C= 70%, D=60%, F= Below 60%
HW assignments will include some paper and pencil assignments, some programming assignments, paper reading/critiquing, etc. Class participation will include participating in paper discussions, discussion leading when it is your turn, being on-time to class, etc

Course contents:

- Issues in Evaluating Performance and Power/Energy of Computers
- Measurement Tools and techniques, Trace Driven and Execution Driven Simulation
- Benchmarks, CPU-intensive, Commercial and database, web server
- Workload Characterization (Quantitative and Analytical)
- Characterization of Emerging Applications
- Statistical techniques for Performance Evaluation
- Trace Generation and Validation, Synthetic Traces, Verification of Simulators
- Design of Experiments
- Analytical Modeling of Processors, Statistical modeling, Hybrid Techniques- Application of queuing theory, Markov models and probabilistic models for computer system evaluation

Course Evaluation: There will be a formal course evaluation towards the end of the semester. I will also be doing several informal intermediate evaluations. I am interested in tailoring the course to result in maximum benefit for you. Please feel free to offer comments.

Academic Dishonesty: Faculty in the ECE Department are committed to detecting and responding to all instances of scholastic dishonesty and will pursue cases of scholastic dishonesty in accordance with university policy. All parties in our community -- faculty, staff, and students -- are responsible for creating an environment that educates outstanding engineers, and this goal entails excellence in technical skills, self-giving citizenry, and ethical integrity. Industry wants engineers who are competent and fully trustworthy, and both qualities must be developed day by day throughout an entire lifetime. Scholastic dishonesty includes, but is not limited to, cheating, plagiarism, collusion, falsifying academic records, or any act designed to give an unfair academic advantage to the student. Penalties for scholastic dishonesty are severe and can include, but are not limited to a record in your academic folder, a zero on the assignment/exam, re-taking the exam in question, an F in the course, or expulsion from the University. Don't jeopardize your career by an act of scholastic dishonesty. You can find UT Honor Code at http://registrar.utexas.edu/catalogs/gi09-10/ch01/index.html).

Drop Policy: An engineering student must have the Dean's approval to add or drop a course after the fourth class day of the semester. Adds and drops are not approved after the fourth class day except for good cause. "Good cause" is interpreted to be documented evidence of an extenuating nonacademic circumstance (such as health or personal problems) that did not exist on or before the fourth class day. Applications for approval to drop a course after the fourth class day should be made in the Office of Student Affairs, Ernest Cockrell, Jr. Hall 2.200

Students with disabilities may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471-6259, http://www.utexas.edu/diversity/ddce/ssd/.

By UT Austin policy, you must notify me of your pending absence at least fourteen days prior to the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.

EE 382M **More Info on Grading**

Paper and Pencil HW Assignments + Programming assignments (Approx 500 pts)

News presentation – Each student should present an interesting news item on a modern processor with 1-2 slides to class (no more than 2 mnts + 3 mnt discussion) (40 pts)

Paper Critiques, Questions based on papers – read paper and write critique before the class the paper is being discussed (10 pts per writeup = approx 150 pts)

Scribing – One student takes notes during lecture, refines it and uploads it for everybody within 36 hours after class. (40 pts each time)

Class Participation – 5 pts each time.

Be on-time
Participate in discussions
No Disruptive behavior

If you are late you lose some points. 2 of those grades will be dropped. (20 * 5 = 100). If you are absent, you get 0.

Literature Survey – 50 pts

All of the above – expect it to be approximately 900 pts. And that will constitute 30% of the course grade.

Preliminary project Interest Document (to help find matching partners) - Sep 13

Literature Survey - Sep 27

Project proposal due - Oct 2

Project Interim Report 1 - Nov 1
Project Interim Report 2 - Nov 25

Project Presentations – Dec 1 and 3

Test 1 (20%) - October 9
Test 2 (20%) - Nov 20

Literature Survey and Project Proposal

Suggested group size: 2 students (if a project justifies another group size, talk to me and we can decide based on the specific scenario)

There are several types of projects:

- Workload Characterization on Desktops/Servers
- Workload Characterization on Embedded Systems (Java enabled boards, mobile systems)
- Experiments with Emerging Workloads (Workload/Benchmark Characterization)
- Improving Speed of Simulation by Simpoint, Sampling, etc.
- Power Measurements/Simulation
- Analytical Models
- Create a new Cloud benchmark
- Benchmark Similarity Studies

Reproducing results from a published paper from IISWC, ISPASS, ISCA, ASPLOS, HPCA, or MICRO will be acceptable as a project. If you make an extension to what has been published, that will be excellent.

The **project proposal** should address

Objectives - What are you trying to find out? What's the problem you are trying to solve? What is the interesting question you are trying to answer?

Background and Motivation - What have others done in this area? Why do you think it is important to do more work? What is the significance of this work? It is important to relate what you are doing to what others have done before.

Research Plan - How you plan to do it? Any existing simulators or tools or are you planning to build your own tools? If developing a simulator, the level of the details. What experiments do you plan to perform?

Expected Outcome - The results of the project. What would be the outcome from the project once it is completed.

Significance or impact of the work/study

Literature Survey and Project Proposal - 5 to 10 pages (Single spaced IEEE format)
This survey is your preliminary reading to identify a suitable project. You choose an area of interest and read 3-10 papers in that area. Must contain summary of at least 3 non-www references. ISCA, Micro, HPCA, ASPLOS, ISPASS, IISWC, PACT, IEEE-TC are all potential sources of references. You may use white papers or www sources as additional references. If you are not finding enough references talk to me early in the semester, way before the project proposal is due.

Project Proposal (2-3 pages) – 2.5% of course grade
Interim Reports (1-3 pages) – 2.5% of course grade (1% and 1.5%)
Project Presentation – 5% of course grade
Final Project report – 20% of course grade

Suggested Project Topics:

Performance/Power Characterization of Cloud Workloads/Benchmarks (eg: BigDataBench, CloudSuite)

Performance/Power Characterization of Virtualization Workloads/Benchmarks(eg: SPEC Virt)

Performance/Power Characterization of Analytics Workloads/Benchmarks (eg: Graph 500, TPCH)

Performance/Power Characterization of Embedded Workloads/Benchmarks (Android phone apps, tablet apps, sensor networks)

Performance/Power Characterization of Web Server Workloads/Benchmarks

Performance/Power Characterization of GPGPU workloads

Performance/Power Characterization of Java benchmark suites'

Synthetic proxies for supercomputing applications and comparison against miniapps, dwarves, etc.

A synthetic proxy benchmark for Java applications

A synthetic proxy benchmark for database applications

A synthetic proxy for GPGPU applications

Synthetic Proxy for Cloud (Hadoop Perf Eval without Hadoop)

A power virus (max power benchmark) for GPGPU

Do an FFT or MATMUL accelerator on a system like the Convey and compare performance of a pure software implementation against the accelerated version

Similarity/Dissimilarity of XXX workloads (within XXX suite, how similar are the benchmarks) (or, how similar is XXX suite with YYY suite)

Similarity/Dissimilarity of HINT and CPU 2006

If you are interested in any of the above projects, I can suggest some papers for literature survey.

Past Class Projects that became Papers:

1. M. Clark and L. K. John, "Performance Evaluation of Configurable Hardware Features on the AMD-K5", In Proceedings of the IEEE International Conference on Computer Design (ICCD 99), Oct 1999, pp. 102-107. (Acceptance rate: 71 accepted/220 submissions = 32%)

2. R. Radhakrishnan, J. Rubio and L. John, “Characterization of Java Applications at ByteCode and UltraSPARC Machine Code Levels”, ICCD 1999, Oct 1999, pp. 281-284. (Acceptance rate: 71 accepted/220 submissions = 32%)
3. G. E. Allen, B. L. Evans, and L. K. John, “Real-Time High-Throughput Sonar Beamforming Kernels Using Native Signal Processing and Memory Latency Hiding Techniques”, Proc. IEEE Asilomar Conf on Signals, Systems and Computers, Pacific Grove, CA, Oct 24-27, 1999, pp. 137-141.
4. H. Nguyen and L. John, “Exploiting SIMD Parallelism in DSP and Multimedia Algorithms Using the AltiVec Technology”, Proceedings of the ACM International Conference on Supercomputing (ICS 99), Greece, June 1999, pp. 11-20. (Acceptance rate: 57 accepted/180 submissions = 32%)
5. R. Radhakrishnan and L. John, “A Performance Study of Modern Web Server Applications”, Euro-Par 1999, Lecture Notes in Computer Science, Springer, pp. 239-247. (Acceptance rate: 188 acccepted/343 submissions = 55%)
6. S. Banerjee, H. R. Sheikh, L. K. John, B. L. Evans, and A. C. Bovik, “VLIW DSP vs. Superscalar Implementation of a Baseline H.263 Video Encoder”, Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers, vol. 2, Pacific Grove, CA, Oct 29-Nov 1, 2000, pp. 1665-1669.
7. W. Lloyd Bircher, M. Valluri, J. Law and L. John, “Runtime Identification of Microprocessor Energy Saving Opportunities”, International Symposium on Low Power Electronics and Design (ISLPED), Aug 2005, pp. 275-280.
8. Jian Chen, Nidhi Nayyar, and Lizy K. John, Mapping of Applications to Heterogeneous Multi-cores Based on Micro-architecture Independent Characteristics, Third Workshop on Unique Chips and Systems (UCAS), Held in conjunction with IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), April 2007.

TENTATIVE READING LIST:

(The first half of the papers in this list are likely to be used in class this semester, but I have used many of the papers from the second half in some years. The second half are also papers you can try to read for Literature survey and project ideas.)

1. Chapter 2 of John & Eeckhout’s book – Computer Performance Evaluation and Benchmarking.
2. D. Bhandarkar and J. Ding, "Performance Characterization of the Pentium Pro Processor", Proceedings of the 3rd High Performance Computer Architecture Symposium, pp. 288-297, 1997
3. P. Bose and T. M. Conte, "Performance Analysis and Its Impact on Design", IEEE Computer, pp. 41-49, May 1998.
4. Aashish Phansalkar, Ajay Joshi, and Lizy K. John, “Analysis of Redundancy and Application Balance in the SPEC CPU2006 Benchmark Suite”, ISCA 2007

5. Reinhold P. Weicker, "An Overview of Common Benchmarks", IEEE Computer, pp. 65-75, December 1990.

6. Henning, J. - SPEC CPU2006 analysis: SPEC CPU suite growth: an historical perspective

7. John L. Henning, "SPEC CPU2006 Memory Footprint", Computer Architecture News, Vol. 35, No.1 - March 2007

8. L. K. John. "More on finding a single number to indicate overall performance of a benchmark suite", ACM SIGARCH Computer Architecture News, Volume 32 , Issue 1 (March 2004), pp. 3-8.

5. Darryl Gove, "CPU2006 Working Set Size", Computer Architecture News, Vol. 35, No.1 - March 2007

6. Wendy Korn, Moon S. Chang, "SPEC CPU2006 sensitivity to memory page sizes", Computer Architecture News, Vol. 35, No.1 - March 2007

7. David Brooks, Vivek Tiwari, Margaret Martonosi, "Wattch: a framework for architectural-level power analysis and optimizations", ISCA, pp. 83 - 94, 2000.

8. T. Sherwood and E. Perelman and G. Hamerly and B. Calder, "Automatically characterizing large scale program behavior", ASPLOS 2002

9. J. R. Mashey. "War of the benchmark means: time for a truce", ACM SIGARCH Computer Architecture News, Volume 32 , Issue 4 (September 2004), pp. 1-14.

10. J. L. Gustafson and Q. O. Snell, "HINT: A New Way to Measure Computer Performance", Hawaii International Conference on System Sciences, Vol II, pp. 392-401, 1995.

11. Joshua Yi, David Lilja, and Douglas Hawkins, " A Statistically Rigorous Approach for Improving Simulation Methodology". HPCA 2003.

12. Mark Oskin, Frederic T. Chong, Matthew Farrens, "HLS: combining statistical and symbolic simulation to guide microprocessor designs", ISCA, pp. 71 - 82, 2000.

13. A. Joshi, L. Eeckhout, L. K. John, and C. Isen, "Automated Microprocessor Stressmark Generation", Proceedings of the IEEE International High Performance Computer Architecture (HPCA) Symposium, 2008

14. K. Ganesan, Jungho Jo, W. Lloyd Bircher, D. Kaseridis, Zhibin Yu, and Lizy K. John, SYMPO: A Systematic Approach for Escalating System-Level Power Consumption using Synthetic Benchmarks", Proceedings of the 19th International Conference on Parallel Architectures and Compilation Techniques (PACT), Vienna, Austria, September 11-15, 2010.

15. Ajay Joshi, Lieven Eeckhout, Robert H. Bell Jr. and L. K. John, Distilling the Essence of Proprietary Workloads into Miniature Benchmarks. ACM

Transactions on Architecture and Code Optimization (TACO), Vol. 5, Issue 2, August 2008, pp.10:1-10:33

16. T. S. Karkhanis and J. E. Smith, “A First-Order Superscalar Processor model”, ISCA 2004
17. Mark Who, Sangwon Seo, Scott Mahlke, Trevor Mudge, Chaitali Chakrabarti, and Krisztian Flautner, AnySP: Anytime Anywhere Anyway Signal Processing, ISCA 2009
18. Sheng Li, Jung Ho Ahn, Richard D. Strong, Jay B. Brockman, Dean M. Tullsen, Norman P. Jouppi, McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures, Micro 2009
19. Victor W. Lee,Changkyu Kim,Jatin Chhugani,Michael Deisher,Daehyun Kim,Anthony D. Nguyen,Nadathur Satish,Mikhail Smelyanskiy,Srinivas Chennupaty,Per Hammarlund,Ronak . Debunking the 100X GPU vs. CPU myth: an evaluation of throughput computing on CPU and GPU, ISCA 2010, pp. 451 – 460.
20. Dam Sunwoo ; Wang, W. ; Ghosh, M. ; Sudanthi, C. Blake, G. ;Emmons, C.D. ; Paver, N.C. , A structured approach to the simulation, analysis and characterization of smartphone applications , IISWC 2013
21. Hadi Esmaeilzadeh, Emily Blem, Renée St. Amant, Karthikeyan Sankaralingam, Doug Burger, Dark Silicon and the End of Multicore Scaling, ISCA 2011
22. Lim, K. ; Univ. of Michigan, Ann Arbor, MI ; Ranganathan, P. ; Jichuan Chang ; Patel, C. Mudge, T.; Reinhardt, S., Understanding and Designing New Server Architectures for Emerging Warehouse-Computing Environments, ISCA 2008
23. Tony Nowatzki Jaikrishnan Menon Chen-Han Ho Karthikeyan Sankaralingam, gem5, GPGPUSim, McPAT, GPUWattch, "Your favorite simulator here" Considered Harmful, WDDD Workshop
24. Chapter 3 of John & Eeckhout’s book – Computer Performance Evaluation and Benchmarking.
25. Bhandarkar, D and Clark D.W., Performance from architecture: comparing a RISC and a CISC with similar hardware organization, ASPLOS 91
26. Luiz Andr Barroso, Kourosh Gharachorloo and Edouard Bugnion, "Memory System Characterization of Commercial Workloads",Proceedings of ISCA-25, Barcelona, Spain, June 1998.
27. T. M. Conte and W-M. Hwu, "Benchmark Characterization for Experimental System Evaluation", Proceedings of the 1990 Hawaii International Conference on System Sciences (HICSS), Vol. I, Architecture Track, pp. 6-18, 1990.
28. L. John, P. Vasudevan and J. Sabarinathan, "Workload Characterization: Motivation, Goals and methodology", pages 3 to 12 from "Workload Characterization: Methodology and Case Studies", IEEE Computer Society, 1999. (this came out from a class project for this class in 1998)

29. SMARTS: Accelerating Microarchitecture Simulation via Rigorous Statistical Sampling, ISCA 03,

30. V. S. Iyengar, L. H. Trevillyan, P. Bose, "Representative Traces for Processor Models with Infinite Cache", Proceedings of HPCA-2, 1996.

31. Ramesh Radhakrishnan, N. Vijaykrishnan, Lizy John, A. Sivasubramaniam, Juan Rubio, and Jyotsna Sabarinathan, "Java Runtime Systems: Characterization and Architectural Implications", IEEE Transactions on Computers, pp. 131-146, vol. 50, issue 2, February, 2001.

26. E. Berg, E. Hagersten, "StatCache: a probabilistic approach to efficient and accurate data locality analysis", Performance Analysis of Systems and Software, 2004 IEEE International Symposium on – ISPASS 2004, Pg 20-27.

27. Gokul B. Kandiraju, Anand Sivasubramaniam, "Characterizing the d-TLB Behavior of SPEC CPU2000 Benchmarks", SIGMETRICS, 2002.

28. Singhal, R.; Venkatraman, K.S.; Cohn, E.; Holm, J.G.; Koufaty, D; Lin, M-.J.; Madhav, M.; Mattwandel, M.; Nidhi, N.; Pearce, J.; Seshadri, M. "Performance Analysis and Validation of the Intel® Pentium® 4 Processor on 90nm Technology." vol 8 no 1, February 2004.

29. Sebastien Nussbaum and James Smith, "Modeling Superscalar Processors via Statistical Simulation", PACT, 2001.

30. B. Black, A. S. Huang, M. H. Lipasti, and J. P. Shen, "Can Trace-Driven Simulators Accurately Predict Superscalar Performance", Proceedings of the Intl Conference on Computer Design (ICCD), pp. 478-485, October 1996.

31. Lieven Eeckhout and Koen DeBosschere, "Hybrid Analytical-Statistical Modeling for Efficiently Exploring Architecture and Workload Design Spaces", PACT, 2001.

32. Lieven Eeckhout, Hans Vandierendonck, and Koen De Bosschere, "Workload Design: Selecting Representative Program-Input Pairs", PACT, 2002.

33. Sherwood, Erez Perelman and Brad Calder, "Basic Block Distribution Analysis to Find Periodic Behavior and Simulation Points in Applications", PACT, 2001.

34. David Ofelt, John L. Hennessy, "Efficient performance prediction for modern microprocessors", SIGMETRICS, pp. 229 - 239, 2000.

35. B. Black and J. P. Shen, "Calibration of Microprocessor Performance Models", IEEE Computer, pp. 59-65, May 1998.

36. Ramesh Radhakrishnan, Ravi Bhargava and Lizy John, "Improving Java Performance Using Hardware Translation", In Proceedings of 15th ACM international Conference on Supercomputing, pages 427-439, 2001.

37. P. K. Dubey, G. B. Adams, and M. J. Flynn, "Instruction Window Size Trade-Offs and Characterization of Program Parallelism", IEEE Transactions on Computers, Vol. 43, No. 4, pp. 431-442, April 1994.

38. E. N. Elnozahy, "Address trace compression through loop detection and reduction", SIGMETRICS, pp. 214 - 215, 1999.

39. D. B. Papworth, "Tuning the Pentium Pro Microarchitecture", IEEE Micro, Vol. 16, No. 2, pg 8, April 1996.

40. J. Dean, J. E. Hicks, C. A. Waldspurger, W. E. Weihl, and G. Chrysos, "Profile Me: Hardware Support for Instruction Level Profiling on Out of Order Processors", MICRO-30, pp. 292-302, 1997.

41. D. B. Noonburg and J. P. Shen, "A Framework for Statistical Modeling of Superscalar Processor Performance", HPCA-3, pp. 298-309, 1997.

42. Tao Li, Lizy John, Anand Sivasubramaniam, Narayanan Vijaykrishnan and Juan Rubio, "Understanding and Improving Operating System Effects in Control Flow Prediction", In Proceedings of the Tenth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-X), 2002.

43. Eric Schnarr, James R. Larus, "Fast out-of-order processor simulation using memoization", ASPLOS, pp. 283 - 294, 1998.

44. Carl J Mauer, Mark D. Hill, David A. Wood, "Full System Timing-First Simulation", SIGMETRICS, 2002.

45. Sohum Sohoni, Rui Min, Zhiyong Xu, Yiming Hu, "A study of memory system performance of multimedia applications", SIGMETRICS, pp. 206 - 215, 2001.

46. Larry Brisson, Mariko Sakamot, Akira Katsuno, Aiichiro Inoue, Yasunori Kimura, "Reverse Tracer: A Software Tool for Generating Realistic Performance Test Programs", HPCA, 2002.