Lecture 20

EE382a Summary

Christos Kozyrakis

http://ee382a.stanford.edu/

EE382a – Fall 2010 – Lecture 20
Announcements

- Project presentations: Thu 12/2\textsuperscript{nd}, 5-8pm, Gates 104
  - Signup for a 15min presentation slot

- Final exam on Wednesday 12/8\textsuperscript{th}, 12.15-3.15pm
  - Hewlett 101
  - Closed book exam, 1 page of notes, calculator
  - Local SCPD students must come to campus
Advice on Project Presentation

- Short presentation
  - 12 minutes on slides, 3 minutes on Q&A

- Rough outline
  - Title
  - Problem description
  - Thesis
  - Implementation overview (1-2 slides)
  - Methodology
  - 1-2 most interesting results
  - Conclusions + future work

- Ok to have material in back-up slides
Advice on Project Presentation

- Practice, practice, practice

- Focus on covering the most important points
  - The presentation is an invitation to read the paper
  - Subset of results, subsets of figures and bars

- Make it easy to read your slides
  - Highlight/point to interesting parts of graphs/figures
  - Provide main observation in written text
  - Don’t have too much text on slides
  - Avoid small fonts
Other Practical Issues

- Check presentation tips on webpage

- Ok to have just 1 speaker in the group

- While one group answers questions
  - The next group will setup their notebook

- Send me a copy of your slides
  - I can have them on my notebook to save time
Advice on Final Paper

- **Short paper**
  - 10-15 pages, single column, 10pt, 1.5x spacing
  - This is a 5-7 page paper in double column format

- **Proposed outline:**
  - Introduction
  - Problem statement & thesis
  - Methodology, experiments, results & lessons learned
  - Related work
  - Conclusions and future work
  - References
Advice on Final Paper

- Learn from the papers that you liked reading
- Focus your writing
  - Don’t provide a core dump on the topic
  - Iterate a few times on detailed outline first
  - Having a paragraph map to begin with helps a lot
- Use figures to explain difficult concepts
  - But make sure you explain figures
  - Also explain any equations, symbols, etc
- Check paper writing tips on webpage
Advice on Final Exam

- Review all reading material
  - All required readings + lecture notes
  - What would you put on one page of notes?
- Review homework + solutions
  - Exam problems will be shorter
- Check sample exam + solutions
  - To be posted soon

During the exam
- Solve easiest problems first
- Watch out for assumptions (state them clearly, avoid controversy)
- State major insight or symbolic formula first
EE382a in Review
EE382a in Review: Principles & Techniques

- Amdhal’s law
- Pipelining
- Parallel processing
- Out-of-order execution
- Critical path reduction
- Locality
- Memoization
- Prediction & speculation
- Amortization
EE382a in Review: Topics (1)

- **Deep pipelining**
  - Helps Freq but reduces IPC due to RAW and branch hazards

- **Superscalar processor**
  - Exploit ILP in sequential programs
  - Wide in-order fetch $\rightarrow$ wide out-of-order issue $\rightarrow$ wide in-order commit
  - Complexity of several structures grows by $O(n^2)$ where $n$ is issue width

- **Branch prediction**
  - Exploit local and global history to eliminate branch hazards
  - Checking confidence in prediction is important
  - Prediction accuracy is a fundamental limitation for superscalar processors

- **Register renaming**
  - Use large number of physical registers to avoid false dependencies
EE382a in Review: Topics (2)

- Memory disambiguation
  - Resolving difficult dependencies through memory
  - Must maintain uni-processor and multi-processor semantics
  - Careful speculation can help

- VLIW
  - Software-managed wide-issue processors
  - Simpler & lower power but cannot exploit dynamic info

- Dynamic binary translation
  - Dynamically instrument/optimize commonly executed code
  - Dynamic optimizations are becoming increasingly important
EE382a in Review: Topics (3)

- Multithreading
  - Helps tolerate latency and better utilize resources
  - With HW support, thread switching can be fast
  - Issues to watch out for: cache interference, OS issues (TLB etc)

- Multi-core processors
  - Motivated by power limitations of single-core designs
  - Shared view of physical memory
  - Challenges: coherence (snoop or directory based), consistency, support for synchronization (atomic instructions)
Multicore caching

- Basic options: shared vs private caches
  - Better utilization vs isolation
- In large-scale chips, caches are inherently non-uniform access structures
  - Optimizations: move data closer to requesting processors, selectively replicate, optimize read-write sharing etc
- Quality of service issues: enforcing isolation or controlled sharing through physical partitioning or by manipulating replacement policy
EE382a in Review: Topics (5)

- Vector processors
  - Most efficient organization for data-level parallelism
  - Multiple vector lanes for highly parallel execution
  - Data parallel designs need high memory bandwidth

- Graphics processors
  - Multi-cores where each core is a multi-threaded, data-parallel (vector-like) engine
  - HW tracking management of threads and conditional ops
  - Non-coherent caches + high-bandwidth memory system
  - Massive throughput performance
EE382a in Review: Topics (6)

- **Low power processors**
  - Focus on energy per instruction (EPI)
  - Higher ILP requires higher EPI (OOO overheads)
    - Single thread performance matters => heterogeneity
  - Programmability overheads => domain/custom accelerators

- **Complexity effective processors**
  - In-order commit through checkpointing
  - Enables speculation over longer periods
    - High latency misses, hardware faults, etc
    - Uses: runahead, continuous flow execution, ...

- **Memory technologies:** 3D, resistive mems, optics
Good Luck with the Project Wrap-up!

- Don’t forget the on-line class evaluation form