CS/EE 6810: Computer Architecture

• Background: CS 3810 or equivalent, based on Hennessy and Patterson’s Computer Organization and Design

• Text for CS/EE 6810: Hennessy and Patterson’s Computer Architecture, A Quantitative Approach, 5th or 6th Edition

• Topics
  ➢ Measuring performance/cost/power
  ➢ Instruction level parallelism, dynamic and static
  ➢ Memory hierarchy
  ➢ Multiprocessors
  ➢ Accelerators, security
  ➢ Storage systems and networks
Lectures

• Class format:
  ▪ Most lectures pre-recorded and posted on YouTube
  ▪ Regular lectures every Mon/Wed (will be recorded)
  ▪ Allocate time every week to do video review (either watch the compact YouTube videos or the class lecture)
  ▪ Available for office hours right after every lecture (or email me to set up a time)
Organizational Issues

- TAs: Sarabjeet Singh, office hours on Tue 1-2, Thu 11-12, TA queue details to be posted shortly

- Canvas for hw submissions, announcements, grades

- Special accommodations, add/drop policies (see class webpage)

- Class web-page, slides, notes, and videos at http://www.cs.utah.edu/~rajeev/cs6810
Grading

• Midterm (25%), Final exam (25%), Homeworks (50%)

• We will drop your lowest homework score

• No tolerance for cheating
Lecture 1: Computing Trends, Metrics

- Topics:
  - Technology trends
  - Metrics (performance, energy, reliability)
Historical Microprocessor Performance

Source: H&P textbook
Microprocessor Performance

![42 Years of Microprocessor Trend Data](image)

Source: karlrupp.net
Points to Note

• The 52% growth per year is because of faster clock speeds and architectural innovations (led to 25x higher speed)

• Clock speed increases have dropped to 1% per year in recent years

• The 22% growth includes the parallelization from multiple cores

• Moore’s Law: transistors on a chip double every 18-24 months
Power Impact

- Increasing frequency led to power wall in early 2000s
- Frequency has stagnated since then
- End of voltage (Dennard) scaling in early 2010s
- Has led to dark silicon and dim silicon (occasional turbo)
Performance Stagnation

• Running out of ideas to improve single thread performance
• Power wall makes it harder to add complex features
• Power wall makes it harder to increase frequency
• Additional performance provided by: more cores, occasional spikes in frequency, accelerators
Clock Speed Increases

Source: H&P textbook
Recent Microprocessor Trends

- Transistors: 1.43x/year
- Cores: 1.2 - 1.4x
- Performance: 1.15x
- Frequency: 1.05x
- Power: 1.04x

Source: Micron University Symp.
Processor Technology Trends

- Transistor density increases by 35% per year and die size increases by 10-20% per year... more functionality
- Transistor speed improves linearly with size (complex equation involving voltages, resistances, capacitances)
- Wire delays do not scale down at the same rate as logic delays
- The power wall: it is not possible to consistently run at higher frequencies without hitting power/thermal limits
What Helps Performance?

• In a clock cycle, can do more work -- since transistors are faster, transistors are more energy-efficient, and there’s more of them

• Better architectures: finding more parallelism in one thread, better branch prediction, better cache policies, better memory organizations, more thread-level parallelism, moving computations to memory, accelerating some kernels, ...
Where Are We Headed?

Modern trends:

- Clock speed improvements are slowing (power constraints)
- Difficult to further optimize a single core for performance
- Multi-cores: each new processor generation will accommodate more cores
- Need better programming models and efficient execution for multi-threaded applications
- Need better memory hierarchies
- Need greater energy efficiency
- Dark silicon, accelerators
- Reduced data movement
- Emergence of new metrics: security, reliability
- Emergence of new workloads: ML, graphs, genomics
More Diverse Platforms
New Design Concerns