CS/EE 6810: Computer Architecture

• Class format:
  ▪ Most lectures on YouTube *BEFORE* class
  ▪ Use class time for lecturing, discussions, clarifications, problem-solving, assignments
Introduction

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


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

• Office hours, MEB 3414, by appointment

• TAs: Padmashree TS, Zinnia Mukherjee, see class webpage for office hrs

• 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: (Sections 1.1 - 1.5, 1.8 - 1.10)

  ➢ Technology trends
  ➢ Metrics (performance, energy, reliability)
Historical Microprocessor Performance

Source: H&P textbook
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
- End of Dennard scaling
- Moore’s Law: transistors on a chip double every 18-24 months
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)… can lead to clock speed improvements!

- 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 (Turbo Mode can cause occasional frequency boosts)
What Helps Performance?

• Note: no increase in clock speed

• 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, etc.
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
  ➢ In some domains, wimpy cores are attractive
  ➢ Dark silicon, accelerators
  ➢ Reduced data movement
Power Consumption Trends

• Dyn power $\propto$ activity x capacitance x voltage$^2$ x frequency

• Capacitance per transistor and voltage are decreasing, but number of transistors is increasing at a faster rate; hence clock frequency must be kept steady

• Leakage power is also rising; is a function of transistor count, leakage current, and supply voltage

• Power consumption is already between 100-150W in high-performance processors today

• Energy = power x time = (dynpower + lkgpower) x time
Power Vs. Energy

• Energy is the ultimate metric: it tells us the true “cost” of performing a fixed task

• Power (energy/time) poses constraints; can only work fast enough to max out the power delivery or cooling solution

• If processor A consumes 1.2x the power of processor B, but finishes the task in 30% less time, its relative energy is $1.2 \times 0.7 = 0.84$; Proc-A is better, assuming that 1.2x power can be supported by the system
Title

• Bullet