Advanced Topics in Computer Architecture (aka Energy-Efficient Computer Architecture)

Instructor: Josep Torrellas

CS 534

Presentations

- Presenter role:
  - Prepare own slides (not use slides from the web)
  - Focus mostly on the assigned paper
  - You are expected to also read other papers on the issue, to explain better and answer questions (Start with the references of the paper)
  - Assume the audience has read the paper in advance
  - Focus the presentation on the key ideas, and results of the paper

  - Plan for about 30-45 minutes of presentation. Discussion should take the rest of the time (during and after your presentation)

  - Send slides to instructor before midnight the day before

Hints on the Presentations

- Discussion topics:
  - New ideas and techniques learned
  - Comparison to other papers read
  - Strengths of the paper
  - Shortcomings of the paper
  - How would you do things differently
  - Open questions, new research directions suggested by the paper

Hints on the Presentations

- Do not leave the presentation for the last minute
  - Following up papers and preparing the discussion topics takes time
- Practice your talk before class
  - This is a requirement
- You will be graded for:
  - Clarity of slides
  - Clarity of presentation
  - Technical command on the topic
  - Quality of the class discussion
  - Time management

Presentations. Audience Role

- Audience role:
  - You have to read the paper in advance
  - You need to send me before midnight the day before, the paper critique: it is your analysis of the paper → this is your homework.
  - During the discussion, raise issue(s).
  - Reading the paper carefully will help you participate in the discussion

Papers

Please respond with only 2-3 sentences per question.

- 1. Give a brief summary of the paper
- 2. What are the paper's main strengths?
- 3. What are the paper's main weaknesses?
- 4. How would you improve this work?
- 5. Write a question/issue to raise in class

Next Class

- Process variation paper
- Volunteer for the presentation needed asap
- From now until Friday 5pm: everyone use the newsgroup to sign up for N papers (there are 27 papers)

Energy-Efficient Computing

- State of the Art:

Performance: 11 PF

Power: 6-11 MW (idle to loaded)

10MW = $10M per year electricity

University of Illinois Blue Waters Supercomputer

- Extreme Scale computing: 100x more capable for the same power consumption and physical footprint

Recap: How Did We Get Here?

- Ideal Scaling (or Dennard Scaling): Every semicond. generation:
  - Dimension: 0.7
  - Area of transistor: 0.7x0.7 = 0.49
  - Supply Voltage V_{dd}, C: 0.7
  - Frequency: 1/0.7 = 1.4

$$P_{dyn} \propto CV_{dd}^2 f$$

Constant dynamic power density

Recap: How Did We Get Here?

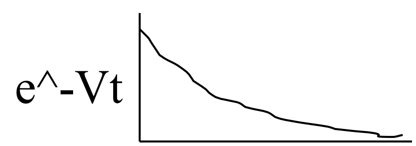
$$P_{dyn} \propto CV_{dd}^2 f$$

Constant dynamic power density

- Real Scaling: V_{dd} does not decrease much
  - If too close to threshold voltage (V_{th}) → slow transistor

$$T_g \propto \frac{L_{\text{eff}} V}{\mu (V - V_{\text{th}})^{\alpha}}$$

  - Dynamic power density increases with smaller tech
  - Additionally: There is the static power



$$P_{sta} \propto V_{dd} T^2 e^{-qV_t/kT}$$

Power density increases rapidly

Low Voltage Operation

- V_{dd} reduction is the best lever for energy efficiency:

- Advantages:
  - Reduces energy per operation quickly
- Drawbacks:
  - Lower speed
  - Higher variation in gate delay and power consumption