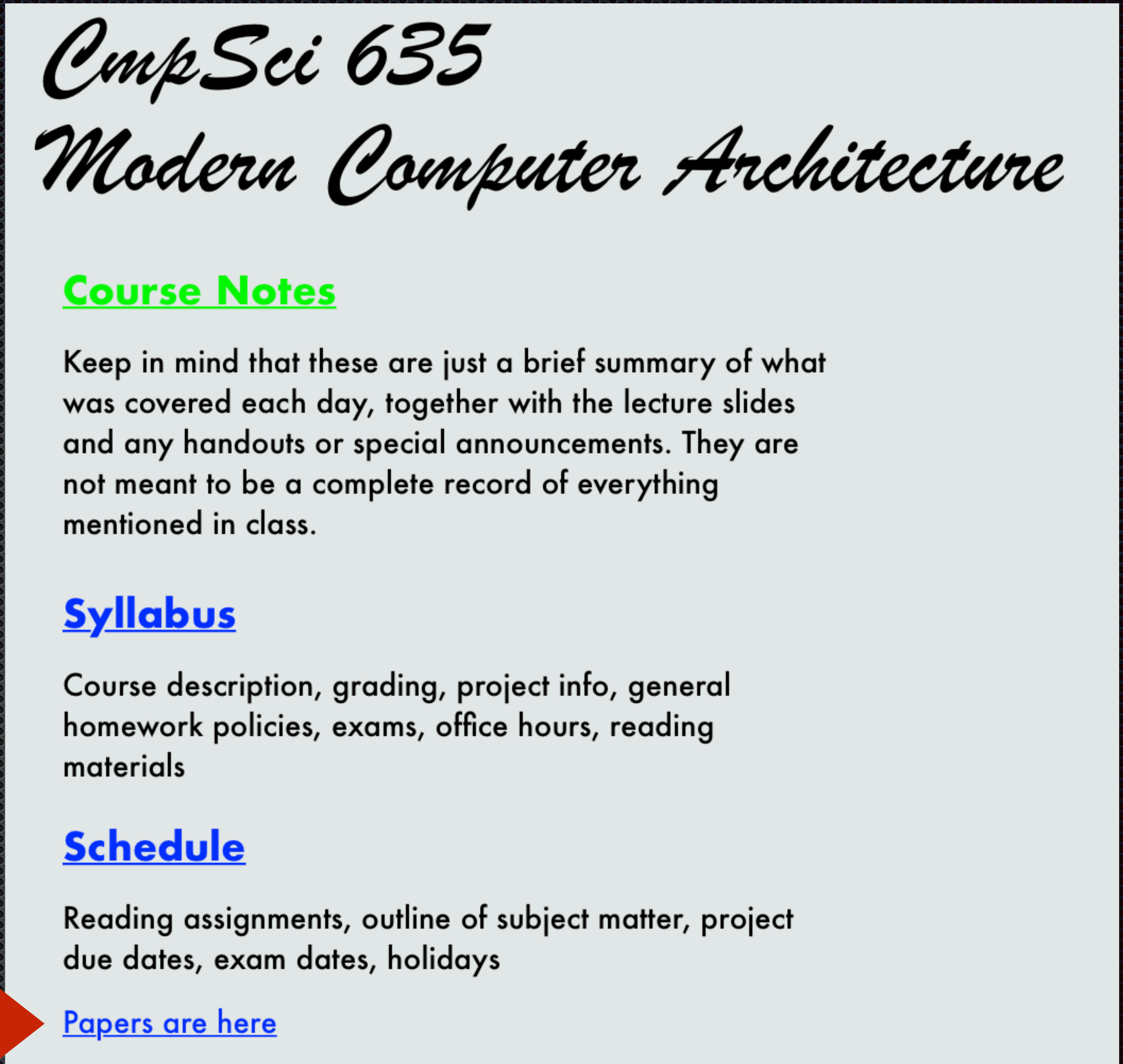
CmpSci 635

Modern Computer Architecture

Intro

- Chip Weems
- weems@cs.umass.edu
- http://www.cs.umass.edu/~weems
- CS 342
- Office hours: Th after class until 12:30, drop-ins, appointments

Web Page


CmpSci 635

Modern Computer Architecture

Course Notes

Keep in mind that these are just a brief summary of what was covered each day, together with the lecture slides and any handouts or special announcements. They are not meant to be a complete record of everything mentioned in class.

Syllabus

Course description, grading, project info, general homework policies, exams, office hours, reading materials

Schedule

Reading assignments, outline of subject matter, project due dates, exam dates, holidays

Papers are here

IMPORTANT

https://people.cs.umass.edu/~weems/homepage/courses/cmpsci-635.html

635 Course Goals

- Gain awareness of modern issues in architecture
- Understand architecture research methodology issues
- Historical perspective behind current trends
- Appreciate factors affecting system performance
- Synthesize new ideas for research

Course Notes

- New notes (blog of topics covered and all slides) will go up shortly after lectures in pdf
- Available via my web page
  - https://people.cs.umass.edu/~weems/homepage/635-Notes/index.html

635 Grading

- Midterm and Final, 20% each
- Reading homework, 60%

635 Emphasizes Reading

- Papers available as zip of pdfs (70MB) from web page
- I will present background material as needed and summarize each
- I will call on people randomly to offer discussion questions
- Homework for each reading is designed to prepare you for discussion

Reading Homework

- For each reading, please identify which paper you are reviewing (e.g., Diwan09)
  - Write two points of strength
  - Write two points of weakness
  - Write two discussion points
- One reading per class (24) = B level

Beyond the B

- Additional discussion points (up to 25% per paper for 3 extra)
- Additional papers (27 = B+, 30 = A-, 33 = A)
- Do additional research to address weakness in papers
- Volunteer to do presentation (= 3 readings — could involve a different paper)
- Do a semester project
- It is possible to earn enough extra credit to skip the final exam

Exams

- Open papers and notes
- Bring a calculator capable of doing exponents
- Best to read and highlight a hard copy that you can bring to exams

Extra Reading Research

- Given an identified point of weakness (e.g., an open question, something confusing or poorly explained, etc.)
- Find one or more additional sources that address the point, and briefly summarize what you find, including full references (use refereed sources)

Reading Schedule

- On the course site as a pdf
- Broken into sections by day
- Usually two or three papers per class
  - A few are doubles (e.g., manufacturer's white papers)
- Pick one you think is best for you

Next 635 Reading

- What do they explain well, and where do they need more depth or detail? What are some implications?
- Diwan points out some critical issues in experimental design for systems research
- This assignment is for “calibration” purposes - grade is just feedback to you (full credit for completion)

Survey

- Not a test -- just guidance for depth
- If you immediately know an answer, just write down enough to show that you do
- If you sort of remember it, then write something like “saw it before,” or “heard about it.”
- If it’s new to you, just put an X

Historical Perspective

How we got here and, where exactly is here?

Early Computing

- ☝
- Others:
  - Abacus, knots, stones

0th Tech Generation

- What did the earliest computing aids do for us?
- What effect did the next group have?
- What was the original definition of “computer”

1st Tech Generation

- What did electronic computers enable?
- What were the problems?

2nd Tech Generation

- Why switch to transistors?
- What are their advantages?

3rd Tech Generation

- Integrated circuits
- What are their main advantages?
- Note change in memory technology

4th Tech Generation

- Microprocessor
- Just more of the same -- why such a big deal?

5th Tech Generation

- Parallel
- Conceived at same time as first computers
- Why did it fail to take off?
- Why is it now taking over?

Functional Generations

- 1st: Memory aids -- increased accuracy, size of numbers
- 2nd: Automatic arithmetic -- greater accuracy, more complexity
- 3rd: Programmable -- extends accuracy to complex functions
- 4th: Reliable -- unlimited complexity, broader use, faster (to do more)
- 5th: Pervasive -- tolerate some failures

Microprocessors

- First were a huge step backward
- Grow in word size, add multiprocessing
- Then add pipeline parallelism (faster clock)
- Cache memory; one then multiple levels
- Superscalar (multiple pipelines)
- Multithreading

Hitting the Wall

- Faster clock needs more power
- Power becomes heat
- Heat slows circuits, increases power
- Greater complexity consumes more power for diminishing speedup
    - Longer relative distances for signals

Punctuated Equilibrium

- Gradual progress, then cross threshold
- Tend to reimplement earlier designs
- Then new designs appear
- Shakeout leads to stable, gradual progress

Parallelism

- Multiple cores don't require common clock
  - Easy utilization of chip area
- Keep clock constant, but increase performance
- Flexible power management
- Immediately available parallel workloads

More Parallelism

- Run out of easy workload distribution
- Shared memory becomes bottleneck
- Hard to find work to drive 16-cores
- One size doesn’t fit all -- heterogeneity
- Challenges dominant programming model
- Much more insidious bugs

The Future

- Low end eats profits, cuts R&D, potential for stagnation at high end
  - Likely to involve government support
- Greater diversity, more embedded, more heterogeneous
- Technology shifts in memory construction (could change everything)
- Increased emphasis on power/heat (dark silicon, warehouse scale)
- New programming models: confusion precedes convergence