CS 3410: Computer System Organization and Programming

Prof. Hakim Weatherspoon
CS 3410, Spring 2015
Computer Science
Cornell University
“Sometimes it is the people that no one imagines anything of who do the things that no one can imagine”

--quote from the movie The Imitation Game
“Can machines think?”

-- Alan Turing, 1950
Computing Machinery and Intelligence
The Bombe used by the Allies to break the German Enigma machine during World War II
Turing Machine
1936

Alan Turing
Course Objective

Bridge the gap between hardware and software
  • How a processor works
  • How a computer is organized

Establish a foundation for building higher-level applications
  • How to understand program performance
  • How to understand where the world is going
Where did it begin?

Electrical Switch
- On/Off
- Binary

Transistor

The first transistor on a workbench at AT&T Bell Labs in 1947
Moore’s Law

1965
• number of transistors that can be integrated on a die would double every 18 to 24 months (i.e., grow exponentially with time)

Amazingly visionary
• 2300 transistors, 1 MHz clock (Intel 4004) - 1971
• 16 Million transistors (Ultra Sparc III) – 2001
• 42 Million transistors, 2 GHz clock (Intel Xeon) – 2004
• 55 Million transistors, 3 GHz, 130nm technology, 250mm² die (Intel Pentium 4) – 2004
• 290+ Million transistors, 3 GHz (Intel Core 2 Duo) – 2007
• 721 Million transistors, 2 GHz (Nehalem) - 2009
• 1.4 Billion transistors, 3.4 GHz Intel Haswell (Quad core) – 2013
Microprocessor Transistor Counts 1971-2011 & Moore's Law

The graph shows transistor count doubling every two years.
Processor Performance Increase

![Graph showing processor performance increase over the years.](image-url)
Moore’s Law

1965
• number of transistors that can be integrated on a die would double every 18 to 24 months (i.e., grow exponentially with time)

Amazingly visionary
• 2300 transistors, 1 MHz clock (Intel 4004) - 1971
• 16 Million transistors (Ultra Sparc III) – 2001
• 42 Million transistors, 2 GHz clock (Intel Xeon) – 2001
• 55 Million transistors, 3 GHz, 130nm technology, 250mm$^2$ die (Intel Pentium 4) – 2004
• 290+ Million transistors, 3 GHz (Intel Core 2 Duo) – 2007
• 721 Million transistors, 2 GHz (Nehalem) - 2009
• 1.4 Billion transistors, 3.4 GHz Intel Haswell (Quad core) – 2013
Parallelism
Then and Now

**The first transistor**
- One workbench at AT&T Bell Labs
- 1947
- Bardeen, Brattain, and Shockley

**An Intel Haswell**
- 1.4 billion transistors
- 177 square millimeters
- Four processing cores

Then and Now

• **The first transistor**
  - One workbench at AT&T Bell Labs
  - 1947
  - Bardeen, Brattain, and Shockley

• **Galaxy Note 3**
  - 8 processing cores
Parallelism

Central Processing Unit (CPU)

Graphics Processing Unit (GPU)
GPU-type computation offers higher GFlops

(Source: Sam Naffziger, AMD)
GPUs: Faster than Moore’s Law

Graph courtesy of Professor John Poulton (from Eric Haines)
Supercomputers

• Petaflops \( (10^{15}) \)
  – GPUs/multicore/100s-1000s cores

China’s Tianhe-2 Supercomputer Maintains Top Spot on 42nd TOP500 List

2013-11-18 08:29:48+00:00

MANNHEIM, Germany; BERKELEY, Calif.; and KNOXVILLE, Tenn.—Tianhe-2, a supercomputer developed by China’s National University of Defense Technology, retained its position as the world’s No. 1 system with a performance of 33.86 petaflop/s (quadrillions of calculations per second) on the Linpack benchmark, according to the 42nd edition of the twice-yearly TOP500 list of the world’s most powerful supercomputers. The list was announced Nov. 18 at the SC13 conference in Denver, Colo.

Titan, a Cray XK7 system installed at the Department of Energy’s (DOE) Oak Ridge National Laboratory, remains the No. 2 system. It achieved 17.59 Pflop/s on the Linpack benchmark. Titan is one of the most energy efficient systems on the list consuming a total of 8.21 MW and delivering 2.143 gigaflops/W.

Sequoia, an IBM BlueGene/Q system installed at DOE’s Lawrence Livermore National Laboratory, is again the No. 3 system. It was first delivered in 2011 and achieved 17.17 Plop/s on the Linpack benchmark.
single-threaded free lunch

1975

1970s 1980s 1990s

2005 2011

20??

Welcome to the jungle

cloud-core

hetero-core

The free lunch is so over

multi-core

Exit Moore
Why?

- Parallelism
- Pipelining
Programmable Hardware

• Started in 1999
• Flexible, programmable
  – Vertex, Geometry, Fragment Shaders
• And much faster, of course
  • 1999 GeForce256: 0.35 Gigapixel peak fill rate
  • 2001 GeForce3: 0.8 Gigapixel peak fill rate
  • 2003 GeForceFX Ultra: 2.0 Gigapixel peak fill rate
  • ATI Radeon 9800 Pro: 3.0 Gigapixel peak fill rate
  • 2006 NV60: ... Gigapixel peak fill rate
  • 2009 GeForce GTX 285: 10 Gigapixel peak fill rate
  • 2011
    – GeForce GTC 590: 56 Gigapixel peak fill rate
    – Radeon HD 6990: 2x26.5
• 2012
  – GeForce GTC 690: 62 Gigapixel/s peak fill rate
Course Objective

Bridge the gap between hardware and software
  • How a processor works
  • How a computer is organized

Establish a foundation for building higher-level applications
  • How to understand program performance
  • How to understand where the world is going
How class is organized

Instructor: Hakim Weatherspoon (hweather@cs.cornell.edu)

Lecture:
• Tu/Th 1:25-2:40
• Statler Auditorium

Lab sections:
• Start next week
• Carpenter 104 (Blue room)
• Carpenter 235 (Red room)
• Upson B7

Required Textbooks

Suggested Textbook
Who am I?

Prof. Hakim Weatherspoon
  • (Hakim means Doctor, wise, or prof. in Arabic)

Career Path
  • Undergrad→grad→post-doc→professor
  • Washington → Berkeley → Cornell → Cornell
Who am I?

The promise of the Cloud

- ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

NIST Cloud Definition
Who am I?

The promise of the Cloud

• ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

NIST Cloud Definition
Who am I?

The promise of the Cloud

- ubiquitous, convenient, on-demand access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

NIST Cloud Definition

Requirements in distributed systems

- Networking
- Computation
- Storage
Who am I?

Cloud computing/storage

- Optimizing a global network of data centers
Course Staff

cs-3410-staff-l@cornell.edu

Lab/Homework TA’s
• Deniz Altinbuken <deniz@cs.cornell.edu> (PhD)
• Adam Campbell <atc89@cornell.edu> (PhD)
• Praveen Kumar <praveenk@cs.cornell.edu> (PhD)
• Vishal Shrivastav <vishal@cs.cornell.edu> (PhD)
• Rob Mcguinness <jrm369@cornell.edu> (MEng)
• Akshay Navalakha <adn47@cornell.edu> (MEng)
• Andrew Weymouth <asw75@cornell.edu> (MEng)
• Naman Agarwal <na298@cornell.edu>
• Maxwell Dergosits <mad293@cornell.edu>
• Antoine Pourchet <app63@cornell.edu>
• Gary Zibrat <gdz4@cornell.edu>
• Ari Karo <aak82@cornell.edu>
• Amy Chen <yc624@cornell.edu>
• Stephanie Guo <lg399@cornell.edu>
• Kylar Henderson <kdh59@cornell.edu>
• Megan Carpenter <mnc29@cornell.edu>
• Rebecca Stambler <rls462@cornell.edu>
• Yogisha Dixit <yad4@cornell.edu>
• Mahak Goel <mg785@cornell.edu>
• Spandan Agrawal <sga35@cornell.edu>
• Adwit Tumuluri <at627@cornell.edu>
• Daniel Liu <dl596@cornell.edu>
• Rishab Gupta <rsg246@cornell.edu>
• Oscar Pacheco <ofp3@cornell.edu>
• Lucas Derraugh <ldd49@cornell.edu>
• Brian Wang <bhw45@cornell.edu>
• Anthony Lin <al744@cornell.edu>
• Charles Lai <cjl223@cornell.edu>
• Jonathan Behrens <jkb229@cornell.edu>

Administrative Assistant:
• Jessica Depew <jd648@cs.cornell.edu>
Pre-requisites and scheduling

**CS 2110 is required** (Object-Oriented Programming and Data Structures)
- Must have satisfactorily completed CS 2110
- *Cannot take CS 2110 concurrently with CS 3410*

**CS 3420 (ECE 3140) (Embedded Systems)**
- Take either CS 3410 or CS 3420
  - both satisfy CS and ECE requirements
  - *However, Need ENGRD 2300 to take CS 3420*

**CS 3110 (Data Structures and Functional Programming)**
- Not advised to take CS 3110 and 3410 together
Pre-requisites and scheduling

CS 2043 (UNIX Tools and Scripting)
- 2-credit course will greatly help with CS 3410.
- Meets Mon, Wed, Fri at 11:15am-12:05pm in Hollister (HLS) B14
- Class started yesterday and ends March 5th

CS 2022 (Introduction to C) and CS 2024 (C++)
- 1 to 2-credit course will greatly help with CS 3410
- *Unfortunately, offered in the fall, not spring*
- Instead, we will offer a primer to C during lab sections and include some C questions in homeworks
# Schedule

(subject to change)

<table>
<thead>
<tr>
<th>Week</th>
<th>Date (Tue)</th>
<th>Lecture#</th>
<th>Lecture Topic</th>
<th>HW</th>
<th>Prelim Evening</th>
<th>Lab Topic</th>
<th>Lab/Proj</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22-Jan</td>
<td>1</td>
<td>Intro</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>27-Jan</td>
<td>2</td>
<td>Logic &amp; Gates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-Feb</td>
<td>3</td>
<td>Numbers &amp; Arithmetic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4-Feb</td>
<td>4</td>
<td>State &amp; FSMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7-Feb</td>
<td>5</td>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10-Feb</td>
<td>6</td>
<td>Simple CPU</td>
<td>HW1: Logic, Gates, Numbers, &amp; Arithmetic, FSM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>17-Feb</td>
<td>7</td>
<td>CPU Performance &amp; Pipelines</td>
<td>Logisim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>24-Feb</td>
<td>8</td>
<td>Pipelined MIPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3-Mar</td>
<td>9</td>
<td>Pipeline Hazards</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10-Mar</td>
<td>10</td>
<td>Control Hazards &amp; ISA Variations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>17-Mar</td>
<td>11</td>
<td>Linkers</td>
<td>Intro to UNIX/Linux</td>
<td>C Lab 2 (cont. IN-CLASS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>24-Mar</td>
<td>12</td>
<td>Caches 1</td>
<td>sh, gcc, How to tunnel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>31-Mar</td>
<td>13</td>
<td>Caches 2</td>
<td>VM &amp; Unix</td>
<td>VM &amp; Unix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>7-Apr</td>
<td>14</td>
<td>Caches 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14-Apr</td>
<td>15</td>
<td>Virtual Memory 1</td>
<td>Stack Smashing</td>
<td>Lab 3: Buffer Overflows handout</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21-Apr</td>
<td>16</td>
<td>Traps</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>28-Apr</td>
<td>17</td>
<td>Multi core Architectures &amp; GPUs</td>
<td>Linkers &amp; and more calling conventions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>5-May</td>
<td>18</td>
<td>Synchronization</td>
<td>HW2: Virtual memory, Caches, Virtual Memory</td>
<td>Lab 4: (IN-CLASS) Virtual Memory</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12-May</td>
<td>19</td>
<td>GPUS &amp; Prelim 2 Review</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>19-May</td>
<td>20</td>
<td>I/O</td>
<td>Prelim 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Grading

Lab (50% approx.)
- 5-6 Individual Labs
  - 2 out-of-class labs (5-10%)
  - 3-4 in-class labs (5-7.5%)
- 4 Group Projects (30-35%)
- Participation/Quizzes in lab (2.5%)

Lecture (50% approx.)
- 2 Prelims (35%)
  - Dates: March 3, April 30
- Homework (10%)
- Participation/Quizzes in lecture (5%)
Grading

Regrade policy
• Submit written request to lead TA, and lead TA will pick a different grader
• Submit another written request, lead TA will regrade directly
• Submit *yet* another written request for professor to regrade

Late Policy
• Each person has a total of *four* “slip days”
• Max of *two* slip days for any individual assignment
• For projects, slip days are deducted from all partners
• 25% deducted per day late after slip days are exhausted
Active Learning

iClicker: Bring to every Lecture

Put all devices into *Airplane Mode*
Fig. 1 Histogram of 270 physic student scores for the two sections: Experiment w/ quizzes and active learning. Control without.

L Deslauriers et al. Science 2011;332:862-864

Published by AAAS
Active Learning

Demo: What year are you in school?

a) Freshman
b) Sophomore
c) Junior
d) Senior
e) Other
Active Learning

Also, activity handouts will be available before class

In front of doors before you walk in
**Administrivia**

http://www.cs.cornell.edu/courses/cs3410/2015sp

- Office Hours / Consulting Hours
- Lecture slides, schedule, and Logisim
- CSUG lab access (esp. second half of course)

**Lab Sections (start *next week*)**

<table>
<thead>
<tr>
<th>Day</th>
<th>Time</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>2:55 – 4:10pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>W</td>
<td>8:40—9:55am</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>W</td>
<td>11:40am – 12:55pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>W</td>
<td>1:25—2:40pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>W</td>
<td>3:35 – 4:50pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>W</td>
<td>7:30—8:45pm</td>
<td>Carpenter Hall 235 (Blue Room)</td>
</tr>
<tr>
<td>R</td>
<td>8:40 – 9:55pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>R</td>
<td>11:40 – 12:55pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>R</td>
<td>2:55 – 4:10pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>F</td>
<td>8:40 – 9:55am</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>F</td>
<td>11:40am – 12:55pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>F</td>
<td>1:25 – 2:40pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
<tr>
<td>F</td>
<td>2:55 – 4:10pm</td>
<td>Carpenter Hall 104 (Blue Room)</td>
</tr>
</tbody>
</table>

- Labs are separate than lecture and homework
- Bring laptop to Labs
- *This week*: “hello world” lab: Intro to C and virtual machines
- *Next* week: Intro to logisim, logic circuits, and building an adder
Administrivia

http://www.cs.cornell.edu/courses/cs3410/2015sp
- Office Hours / Consulting Hours
- Lecture slides, schedule, and Logisim
- CSUG lab access (esp. second half of course)

Course Virtual Machine (VM)
- Identical to CSUG Linux machines
- Download and use for labs and projects
- https://confluence.cornell.edu/display/coecis/CSUG+Lab+VM+Information
Communication

Email
• cs-3410-staff-l@cornell.edu
• The email alias goes to me and the TAs, not to whole class

Assignments
• CMS: http://cms.csuglab.cornell.edu

Newsgroup
• http://www.piazza.com/cornell/spring2015/cs3410
• For students

iClicker
• http://atcsupport.cit.cornell.edu/pollsrvc/
Lab Sections, Projects, and Homworks

Lab Sections start *this* week
- This week: “hello world” lab: Intro to C and virtual machines
- Next week: Intro to logisim, logic circuits and building an adder

Labs Assignments
- Individual
- One week to finish (usually Monday to Monday)

Projects
- Two-person teams
- Find partner in same section

Homeworks
- One before each prelim
- Will be released a few weeks ahead of time
- Finish question after covered in lecture
Academic Integrity

All submitted work must be your own
  • OK to study together, but do not share soln’s
  • Cite your sources

Project groups submit joint work
  • Same rules apply to projects at the group level
  • Cannot use of someone else’s soln

Closed-book exams, no calculators

• Stressed? Tempted? Lost?
  • Come see us before due date!

Plagiarism in any form will not be tolerated
Why do CS Students Need Transistors?
Why do CS Students Need Transistors?

Functionality and Performance
Why do CS Students Need Transistors?

To be better Computer Scientists and Engineers

- Abstraction: simplifying complexity
- How is a computer system organized? How do I build it?
- How do I program it? How do I change it?
- How does its design/organization effect performance?
Computer System Organization

![Diagram of computer system organization with labeled parts such as Processor, Hard drive, Spot for memory DIMMs, Motherboard, Fan, DVD drive, and I/O bus slots.]
Computer System Organization

Computer System = 
Input +
Output +
Memory +
Datapath +
Control

Registers
CPU

Video

Memory

Network

Disk

USB
Serial

Keyboard
Mouse

Audio

Serial

Network

Disk

USB

Serial

Keyboard
Mouse

Audio
Compilers & Assemblers

C

```
int x = 10;
x = 2 * x + 15;
```

MIPS

```
addi r5, r0, 10
muli r5, r5, 2
addi r5, r5, 15
```

r0 = 0

MIPS

```

00100000000001010000000000001010
00000000000000101001010100001000000
00100000101010101000000000001111
```

```
op = addi   r0   r5  10
```

```
op = addi   r5   r5  15
```

assembler
Instruction Set Architecture

ISA

• abstract interface between hardware and the lowest level software

• user portion of the instruction set plus the operating system interfaces used by application programmers
Basic Computer System

A processor executes instructions
  • Processor has some internal state in storage elements (registers)

A memory holds instructions and data
  • von Neumann architecture: combined inst and data

A bus connects the two
How to Design a Simple Processor

Memory

Inst

32

2

00

pc

new pc calculation

Register file

r5

Control

5 5 5

alu

10

0

10

00: addi r5, r0, 10

04: muli r5, r5, 2

08: addi r5, r5, 15
Inside the Processor

AMD Barcelona: 4 processor cores

Figure from Patterson & Hennessy, Computer Organization and Design, 4th Edition
How to Program the Processor: MIPS R3000 ISA

Instruction Categories

• Load/Store
• Computational
• Jump and Branch
  – coprocessor
• Memory Management

<table>
<thead>
<tr>
<th>OP</th>
<th>rs</th>
<th>rt</th>
<th>rd</th>
<th>sa</th>
<th>funct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>jump</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>target</td>
</tr>
</tbody>
</table>

Registers

- R0 - R31
- PC
- HI
- LO
Overview

Instruction Set Architecture

- Application
- Operating System
  - Compiler
  - Firmware
- Memory system
- I/O system
- Datapath & Control
- Digital Design
- Circuit Design
Applications

Everything these days!

• Phones, cars, televisions, games, computers,…
Applications

- Cell Phones
- PCs
- TVs


- 1997: 1193 millions
- 1999: 295 millions
- 2001: 114 millions
- 2003: 405 millions
- 2005: 502 millions
- 2007: 1182 millions

- Cell Phones: 1193, 295, 114, 405, 502, 1182
- PCs: 135, 136, 202, 265
- TVs: 200, 265, 189, 200

Categories: Cell Phones, PCs, TVs

- Cloud Computing
- NVidia GPU
- Xilinx FPGA
- Berkeley mote
- Cell Phone
- Cars
Covered in this course

Application
  Operating System
    Compiler   Firmware
Memory system  Instr. Set Proc.  I/O system
  Datapath & Control
    Digital Design
    Circuit Design

Instruction Set Architecture
Reflect

Why take this course?

- Basic knowledge needed for *all* other areas of CS:
  - operating systems, compilers, ...
- Levels are not independent
  - hardware design ↔ software design ↔ performance
- Crossing boundaries is hard but important
  - device drivers
- Good design techniques
  - abstraction, layering, pipelining, parallel vs. serial, ...
- Understand where the world is going