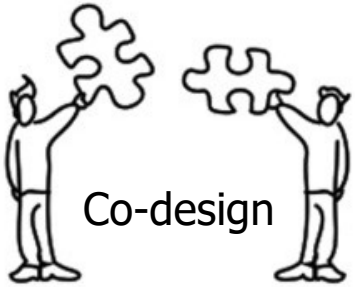
CSE 240A: Principles in Computer Architecture

Fall 2021

Jishen Zhao (https://cseweb.ucsd.edu/~jzhao/)

About me

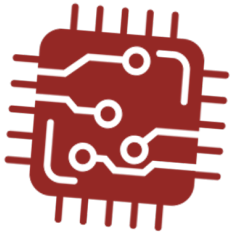
- Associate professor, CSE
- Research



Machine Learning

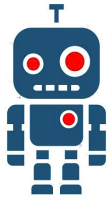
[ICCV'21, ICML'21, ICASSP'21, ICLR'20, MLSys'20, NeurIPS'19, IJCAI'19]

Computer Architecture



[USENIX ATC'21, IEEE Micro Top Picks'21, MICRO'13~16,18~20, ISCA'16/19, ASPLOS'19, HPCA'18/19 USENIX HotStorage'18]

Autonomous Driving Systems



[SSIV'21, IV'20, DAC'20, ICCD'20, ICMR '19, FSR'19, ICRA'19]

My research

- Machine learning for systems

Neural Decompilers
[Fu+, NeurIPS'19]

FACEBOOK AI Research Publications Pe

"N-Bref", https://github.com/facebookresearch/nbref

RESEARCH

Introducing N-Bref: a neural-based decompiler framework

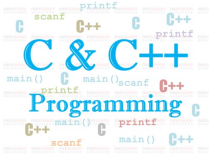
January 27, 2021

```
//i*15;
MOV R3, #15
STR R3, [R11, #-8]

//j*25;
MOV R3, #25
STR R3, [R11, #-12]

//i*i*j;
LDR R2, [R11, #-8]
LDR R3, [R11, #-12]
ADD R3, R2, R3
STR R3, [R11, #-8]
```

ASSEMBLY LANGUAGE



Machine Learning Methods

Deep Symbolic Superoptimization
[Shi+, ICLR'20]

https://github.com/shihui2010/symbolic_simplifier

Persistent Memory Programming Assistant
[Huang+, USENIX ATC'21]

"Ayudante", https://github.com/Hanxian97/Ayudante

My research

Safety-aware autonomous driving

"Autonomous vehicles are becoming mobile datacenters."
– H.J. Vögel, BMW Group Research, 2017

Our autonomous vehicle platforms:

UCSD Self-driving Vehicle

Wang+, AV software management with Linux containers, ICRA'19.

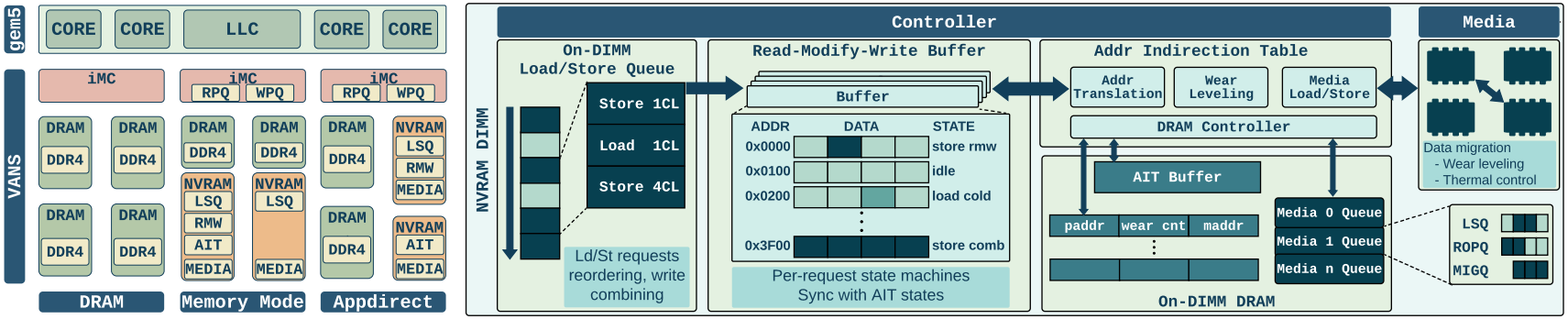
Industry autonomous vehicles and simulators

Zhao+, Safety model for AV computing systems, IV'20;
*Zhao+, AV perception latency model, ICCD'20 (**Best Paper in Track Award**)*

My research

Nonvolatile memory architecture

Characterizing and Modeling Non-Volatile Memory Systems



(a) System model. (b) NVRAM DIMM model.

*[Wang+, MICRO 2020 (**IEEE Micro Top Picks 2020**)]*

LENS and VANS, open sourced at
https://github.com/TheNetAdmin/LENS-VANS

Today

- Course administrivia
- Introduction to this class
  - What is computer architecture?
  - Why study computer architecture?
  - What's different from CSE141?
- Using Zoom
  - If you want to ask a question
    - Post it in chat
    - Click raise hand

Course administrivia

- Class Meetings:
  - Tues & Thurs, 3:30-4:50pm
  - September 23 ~ December 2
  - Virtual lectures on Zoom (see class Google site for detailed info.)
- Instructor: Jishen Zhao (jzhao@ucsd.edu)
  - Office Hours: Fridays 11am – 12pm (the same Zoom link as lectures), starting from next week
- TAs: (Office hours TBD)
  - Vivian Lee (vyl003@ucsd.edu)
  - Sachin Deshpande (scdeshpa@ucsd.edu)
  - Link Lin (yul065@ucsd.edu)
  - Vaibhav Kant Tiwari (vktiwari@ucsd.edu)
- Should be able to enroll everyone currently on the waitlist

Course administrivia

- Instructor responsibility
  - Lectures
  - Create reading assignments and projects, help create HWs
  - Create exams
  - Office hours
  - Answer questions on Piazza

- TA responsibility
  - Post HWs
    - I'll propose most of HW questions
  - Grade assignments, exams, and projects
    - I'll go over the homework before posting grades
  - Office hours
  - Answer questions on Piazza

Disability accommodations

- If you need disability accommodation, please contact the Office for Students with Disability (OSD) to discuss appropriate accommodations right away

Course websites

- Google site + Gradescope + Piazza

  - Google site: https://sites.google.com/eng.ucsd.edu/cse240a-f21/
    Basic information and syllabus, calendar, slides, lecture notes, videos, homework questions and solutions, reading assignment, exam preparation materials, project information, etc.

  - Gradescope: https://www.gradescope.com/courses/316459
    For homework and exam submissions, project report submissions, and grades.

  - Piazza: https://piazza.com/ucsd/fall2021/cse240a
    For most communication -- announcements, Q&A, discussions, polls, etc.

urse Description

course provides a thorough and fundamental treatment of the art of computer architecture. Topics include concepts of von Neumann archit
uating CPU performance, instruction-set design and examples, pipelining, branch prediction, cache and memory hierarchy, main memory org
troller, virtual memory, multicore processors, cache coherence and memory consistency, and GPU architecture.

uired Knowledge: Basics of Instruction Set Architecture, cache design, and the design of single-cycle, multi-cycle, and pipeline processors. If
ropriate background, you should either 1) not take this class or 2) spend significant time reviewing the textbook and lecture notes from CSE1

rse outline and lecture materials are here.

rse calendar is here.

Agenda (subject to change)

CSE240A F21 | Home | Calendar | Lecture Materials | Assignments and Exams | Projects

Lecture Notes and Syllabus

Videos and notes for required materials will be posted before each class. Exercises and solutions will be posted after each class. These are the ones you should use for reviewing the class, HWs and exam.

Notes for live lecture sessions will be posted after each class. These are extended materials for those of you who wants to learn more about computer architecture, won't be video recorded, and will NOT appear in HWs or exam.

Lectures CSE240A F21 : Sheet1

Week	Date	Videos (before class, required for homework / exam)	Lecture notes/PDF (before class, required for homework / exam)	Class discussion notes (not required for homework / exam)	In-class Exercises	Assignments
Week 1	9/23/2021	01 Introduction (video)	01 Introduction			
Week 2	9/28/2021	02 Performance (video)	02 Performance	No class	No class	
	9/30/2021	03 Performance/ISA (video)	03 Performance/ISA	03ext Data center metrics: tail latency, energy efficiency, total cost of ownership		
Week 3	10/5/2021	04 ISA (video)	04 ISA	04ext ISA, programming model		
	10/7/2021	05 ISA/pipelining (video)	05 ISA/pipelining	05ext compiler for deep learning, machine learning		HW1 posted
Week 4	10/12/2021	06 Pipelining (video)	06 Pipelining	06ext		
	10/14/2021	07 Branch prediction (video)	07 Branch prediction	07ext Meltdown and Spectre attacks		
Week 5	10/19/2021	08 Branch prediction/hw multithrea	08 Branch prediction/hw multith	08ext Superscalar		
	10/21/2021	09 Out-of-order execution (video)	09 Out-of-order execution	09ext Google TPU		
Week 6	10/26/2021	10 Out-of-order execution (video)	10 Out-of-order execution	10ext Scoreboarding		HW1 due/HW2 posted
	10/28/2021	11 Memory hierarchy/cache (videc	11 Memory hierarchy/cache	11ext 3D/2.5D integrated architectures		
Week 7	11/2/2021	12 Cache (video)	12 Cache	12ext 3D memory architectures		
	11/4/2021	13 Main memory (video)	13 Main memory	13ext Near memory computing and machine learning acceleration		
Week 8	11/9/2021	14 Cache/memory review and adc	14 Cache/memory review and a	No class (may use the time to	No class (may use th	Midterm exam (take-home) p
	11/11/2021	No class (holiday)	No class (holiday)	No class (holiday)	No class (holiday)	Midterm exam (take-home) r

Flipped classroom

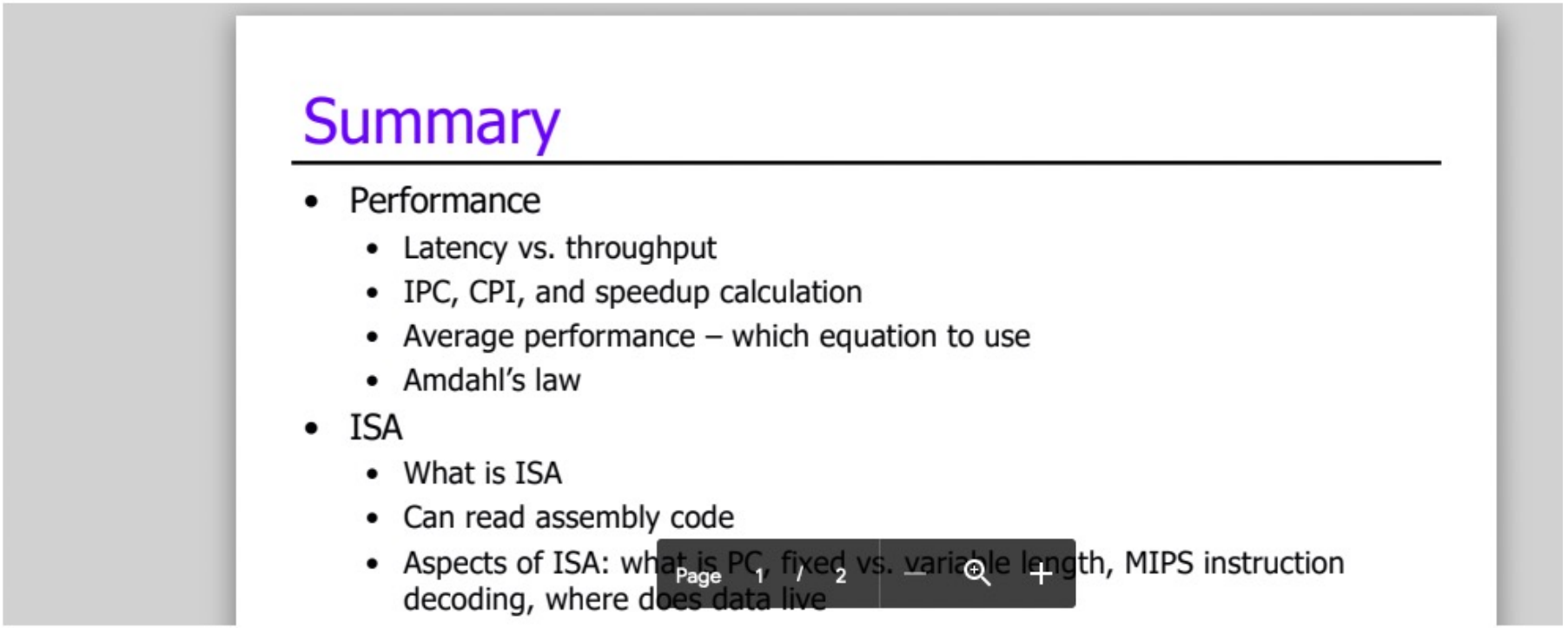
Week	Date	Videos (before class, required for homework / exam)	Lecture notes/PDF (before class, required for homework / exam)	Class discussion notes (not required for homework / exam)	In-class Exercises
Week 1	9/23/2021	01 Introduction (video)	01 Introduction		
Week 2	9/28/2021	02 Performance (video)	02 Performance	No class	No class
	9/30/2021	03 Performance/ISA (video)	03 Performance/ISA	03ext Data center metrics: tail latency, energy efficien	
Week 3	10/5/2021	04 ISA (video)	04 ISA	04ext ISA, programming model	

- Watch lecture videos that are posted before each class
- Live Zoom sessions (attendance is not mandatory)
  - Discuss extended topics (not required for homework/exam)
  - Q&A
  - In-class exercises
  - No video recording, but will post all the notes after each class
- Benefits
  - Learn the class materials on your own pace
  - Learn more than required materials
  - Customized Q&A, provide everyone with more opportunities to discuss with me on any questions
  - More exercises

Required materials vs. extended

- See required materials list on class Google site
  - Anything not on the list is not required

Required technical materials (for homework and exam)



Homework submission

- HW questions, reading assignments, and exam will be posted on the Google site
- Submission: Electronically through Gradescope
- NO hard copy
- Two HWs, one paper reading assignment, one exam
- Usually due at 11:59pm

Homework and Assignment Late Policy

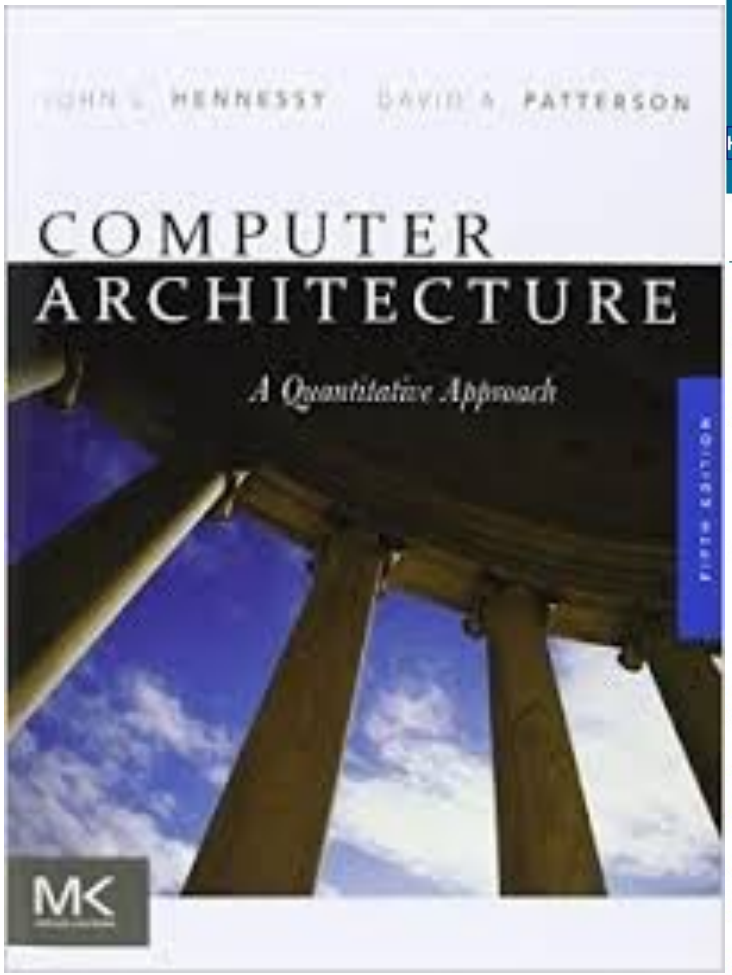
- Late penalty for each homework/reading assignment
  - One day (i.e., by 11:59pm the second day): 20% off
  - Two days: 40% off
  - Three days: 60% off
  - Four days: 80% off
  - Five days and more: 100% off

- Extension with an acceptable reason — send me an email to confirm
- No excuse
  - E.g., I submitted the wrong file, I fell asleep, I forgot to submit my homework, I don't know how to use GradeScope, my computer was broken down…

Regrading requests

- Concerns/questions/doubts about homework/reading/exam grading
  - Contact me within one week after the corresponding grade release
  - TAs and I will not respond to requests after that time period

Textbooks (not required)



MORGAN & CLAYPOOL PUBLISHERS

Hello. Sign in to personalize your visit. New user? Register now. University of California, San Diego

Home | Synthesis | Colloquium | Search | Profile | Author | Help | About | Bookstore

Quick search: Enter keyword within: This issue go

Synthesis Lectures on Computer Architecture

Lectures available online | Lectures under development | Order print copies

Editor
Margaret Martonosi, *Princeton University*

Founding Editor Emeritus: Mark D. Hill, University of Wisconsin, Madison

Synthesis Lectures on Computer Architecture publishes 50- to 100-page publications on topics pertaining to the science and art of designing, analyzing, selecting and interconnecting hardware components to create computers that meet functional, performance and cost goals. The scope will largely follow the purview of premier computer architecture conferences, such as ISCA, HPCA, MICRO, and ASPLOS.

Information for prospective authors

Series ISSNs: 1935-3235 (print) and 1935-3243 (electronic)

XML What is RSS?

The Datacenter as a Computer: Designing Warehouse-Scale Machines, Third Edition
Luiz André Barroso, Urs Hölzle, Parthasarathy Ranganathan
October 2018
Abstract | PDF (9029 KB) | PDF Plus (5167 KB)

Principles of Secure Processor Architecture Design
Jakub Szefer
October 2018
Abstract | PDF (3387 KB) | PDF Plus (3269 KB)

General-Purpose Graphics Processor Architectures
Tor M. Aamodt, Wilson Wai Lun Fung, Timothy G. Rogers
May 2018

Reading technical papers

- As graduate students, you are now researchers. Most of the information you need will be in technical papers
  - You will learn a lot about doing good systems research by reading others' papers
  - The ability to rapidly scan and understand technical papers is key to your success
- So, you will read a paper
  - Paper review assignment (review template will be provided)
  - Important supplement to lectures and book

Project (Pick one of four options)

- 1. Recreate results from a recent research paper to see
  - If they are reproducible ***ISCA, MICRO, ASPLOS, HPCA, PACT, …***
  - If they still hold (recent tech advances & new app requirements)
- 2. Implementation projects (will be posted on the Google site)
- 3. Propose your own -- in/related to computer architecture
  - Need my approval by submitting a project proposal before by end of Week2
- 4. Write a literature survey paper on a specific topic
  - Have a limited cap score, at most two survey papers per topic
- Note:
  - 1 – 3: Maximum team of two people (so keep an eye out for a potential partner). But each need to write a separate report.
  - 4: Must be solo project
- Project reports due on December 14
- One report per person
  - If working as a team, each team member submit a report –
    - Can have the same content, but need to describe contributions by individual (e.g., particular experimental results, …)

Grading

- Homework: 15% (two homework assignments)
- Reading assignment: 8% (one paper reading assignment)
- Midterm exam: 45%
  - Can count for MS comprehensive exam
- Project: 32%
  - What to submit: project report
  - 27 points cap for survey projects

Smiling: 0%

Integrity policy

- Not considered as cheating
  - Discussing homework in groups (with the writeup done separately, later)
- Considered as cheating
  - Discussing homework with someone who has already completed the problem, or looking at their completed write-up
  - Using hw solutions from the web, previous versions of the class, or anywhere else
  - Receiving, providing, or soliciting assistance from another student during a test
- Penalty
  - Will receive an F for the class and will not be allowed to drop
  - Will be reported to their college dean

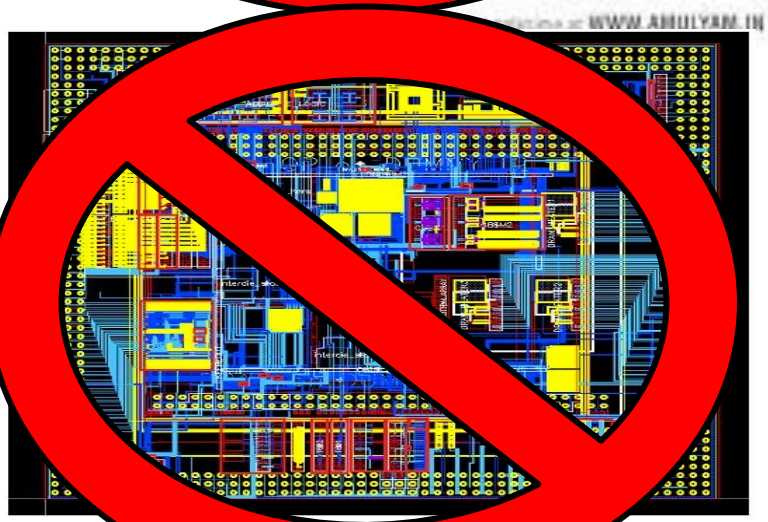
What's different from CSE141

- Technical materials
  - Some may look similar to what you learned in CSE141 or equivalent courses
  - Something new to you: e.g., sophisticated caching and cache coherence schemes, memory scheduling policies, main memory organization, memory consistency, more GPU architectures, …
- Reading papers and paper reviews
- The project
  - Deeper understanding of one research area
  - Hands on experience of architecture design with significant implementation efforts
- What it's like to do computer architecture research
  - Critical thinking!

Questions?

Introduction to Computer Architecture

Course focus: Computer Architecture



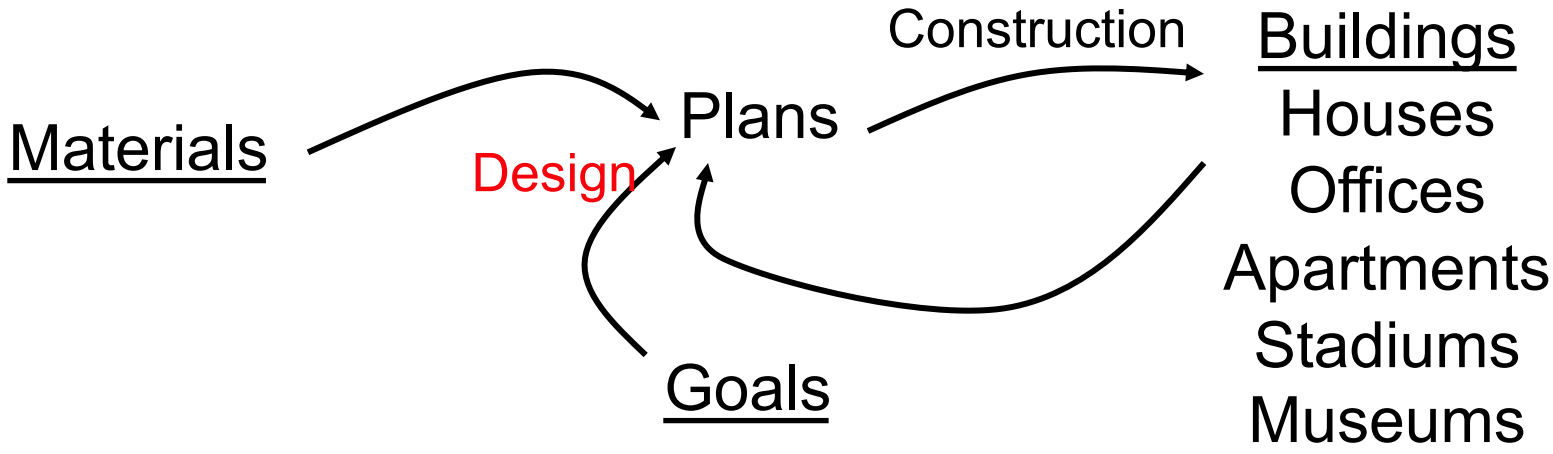
Course focus:

Basic ideas about Computer Organization and Hardware/software interface

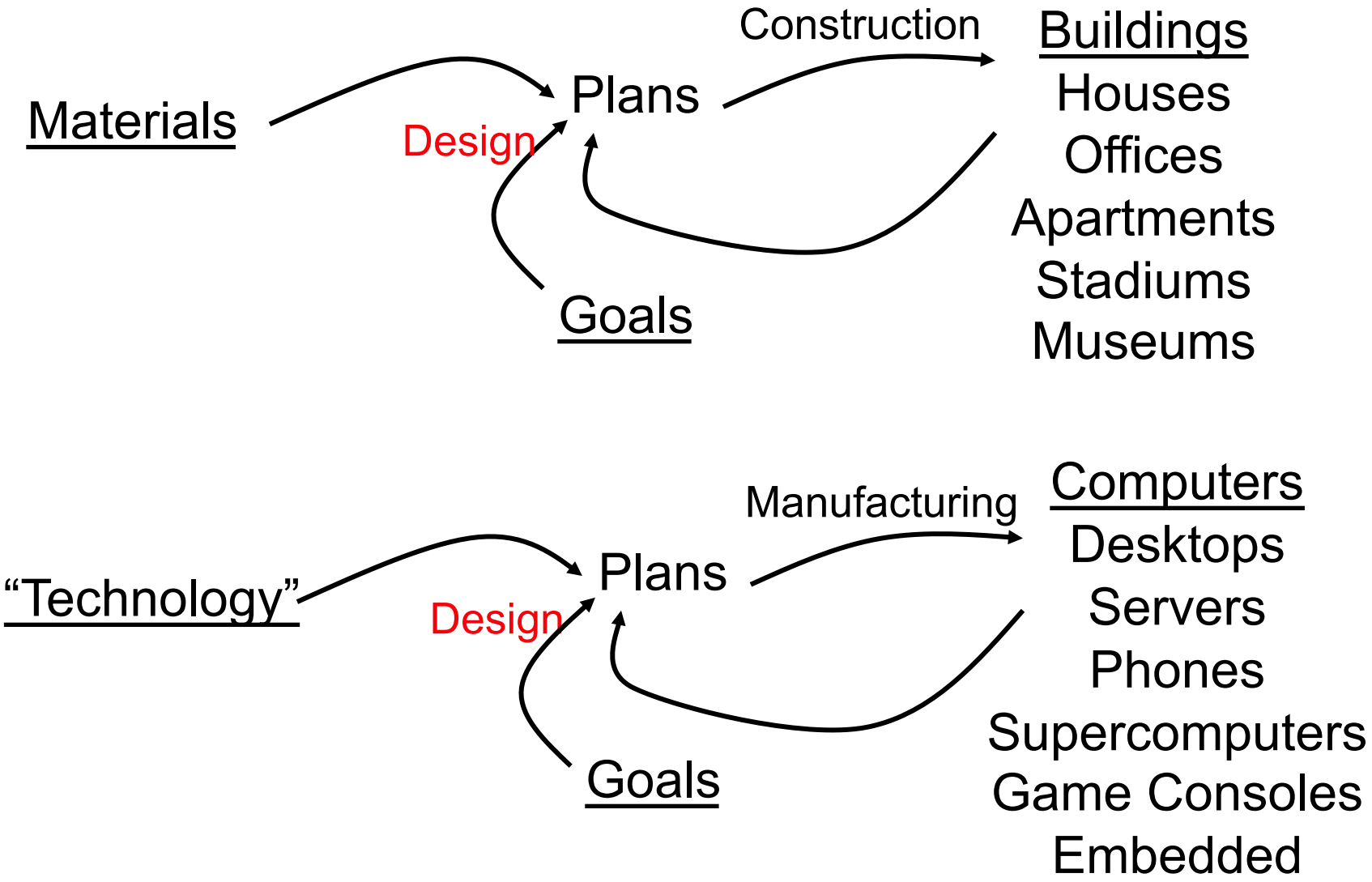
What is Computer Architecture?


brick wall
mortar
firebrick
perforated brick
hollow brick
partition tile
solid brick

What is ~~Computer~~ *building* Architecture?



What is Computer Architecture?



Details

- A house has a few rooms
  - Living room
  - Bedroom
  - Restroom
  - …

They are interconnected

- A computer has a few components
  - Processor core (CPU core)
  - Caches
  - Main memory
  - Disk
  - …

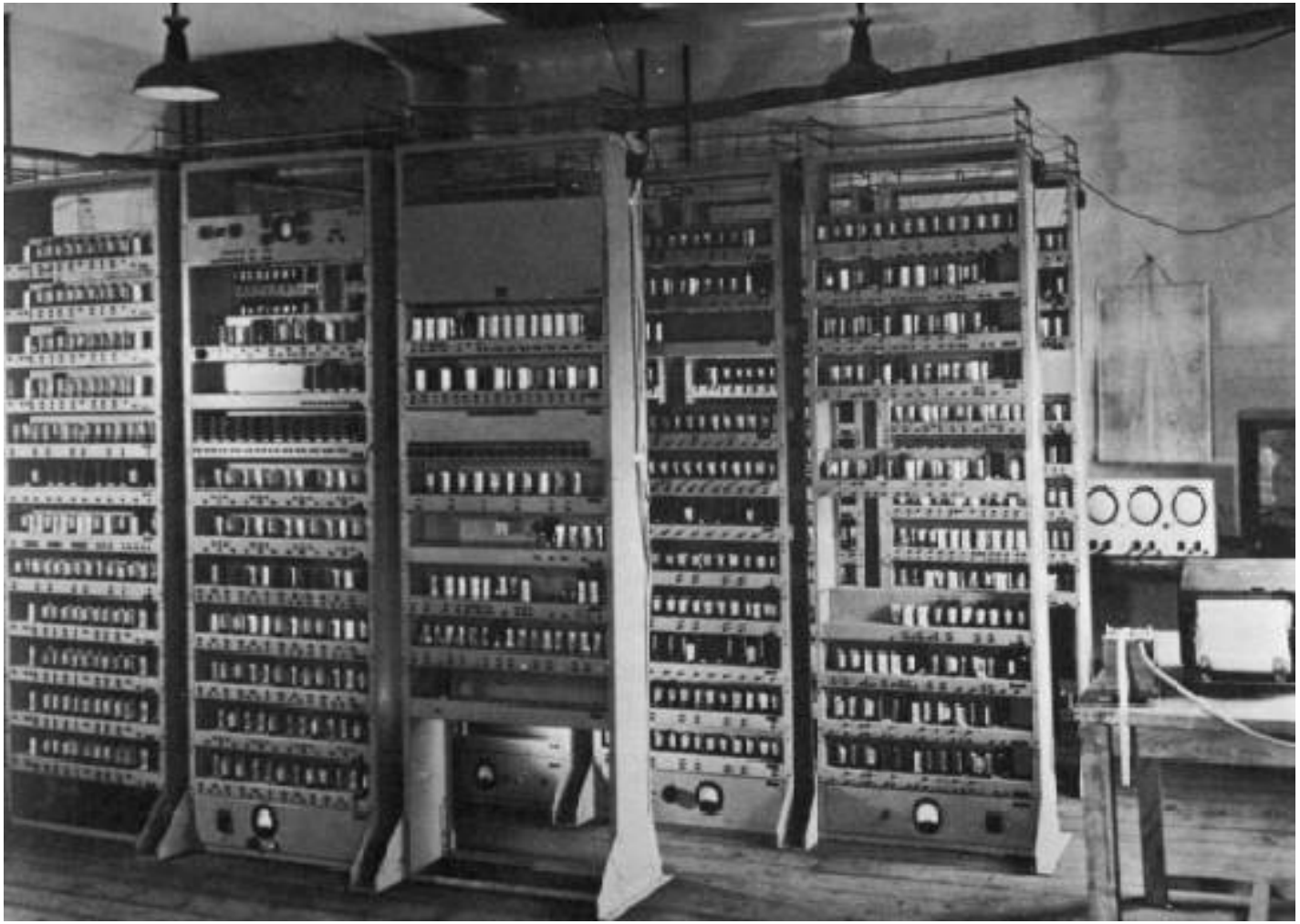
They are interconnected too

Why study computer architecture?

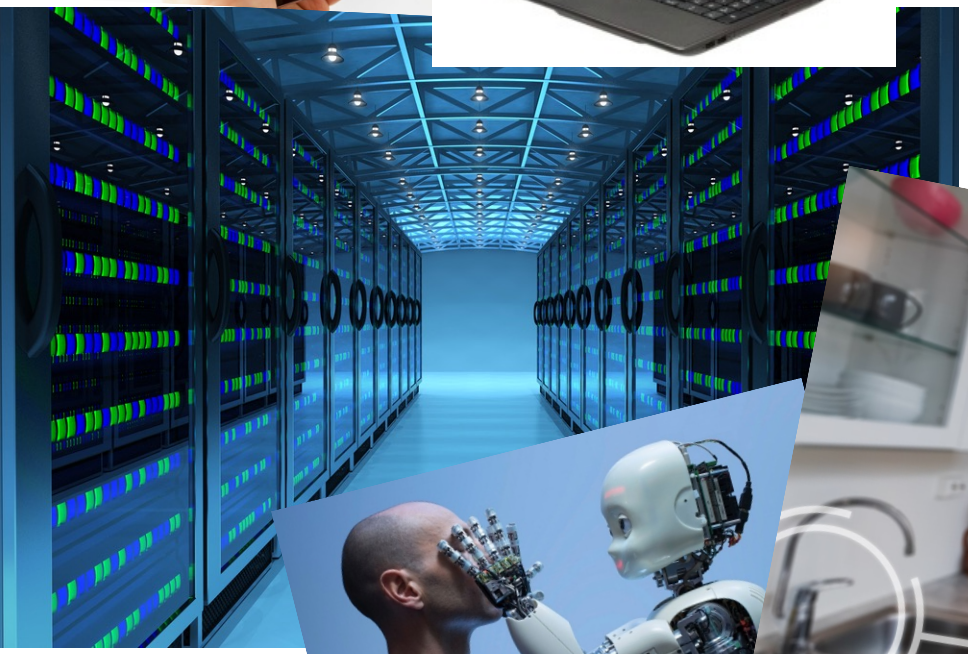
- **Understand where computers are going**
  - Future capabilities drive the (computing) world
  - Real world-impact: no computer architecture → no computers!
- **Understand high-level design concepts**
  - The best architects understand all the levels
    - Devices, circuits, architecture, compiler, OS, applications
- **Understand computer performance**
  - learn valid experimental methodologies
- **Write better software**
  - The best software designers also understand hardware
  - Need to understand hardware to write fast software
- **Design hardware**
  - At Intel, AMD, IBM, ARM, Qualcomm, Oracle, NVIDIA, Samsung

History and Trend

Computers then…



Computers now



Heads up

- No class on Tuesday next week
  - The first extended topic needs background that is in the first two lecture videos

- Keep an eye on course calendar updates and assignments on the course Google site

- Make sure you can access all the course websites: Google site, Gradescope, and Piazza. And Canvas (as backup site)