Lecture 1

- Today's topics:
  - Why computer organization is important
  - Logistics
  - Modern trends

Why Computer Organization

Image credits: uber, extremetech, anandtech

Why Computer Organization

- Embarrassing if you are a BS in CS/CE and can't make sense the following terms: DRAM, pipelining, cache hierarchies, I/O, virtual memory, …
- Embarrassing if you are a BS in CS/CE and can't decide whic processor to buy: 4.4 GHz Intel Core i9 or 4.7 GHz AMD Ryzen 9 (reason about performance/power)
- Obvious first step for chip designers, compiler/OS writers
- Will knowledge of the hardware help you write better and more secure programs?

Must a Programmer Care About Hardware?

- Must know how to reason about program performance and energy and security
- Memory management: if we understand how/where data is placed, we can help ensure that relevant data is nearby
- Thread management: if we understand how threads interact, we can write smarter multi-threaded programs

→ Why do we care about multi-threaded programs?

Example

200x speedup for matrix vector multiplication

- Data level parallelism: 3.8x
- Loop unrolling and out-of-order execution: 2.3x
- Cache blocking: 2.5x
- Thread level parallelism: 14x

Further, can use accelerators to get an additional 100x.

What This Course is About

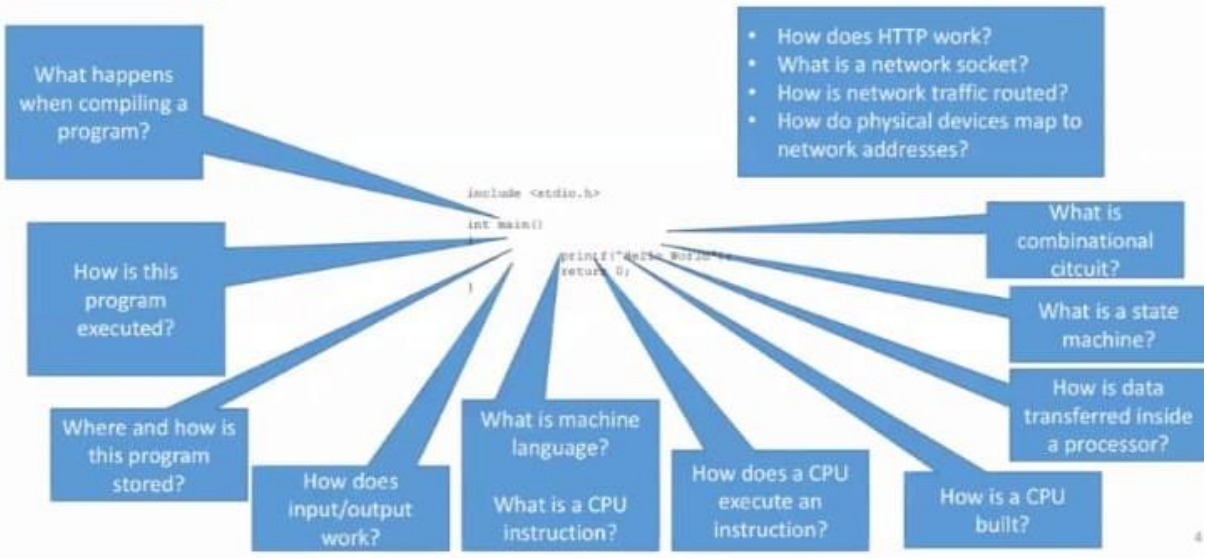
- How does a computer work?
- What does actually happen, if I compile and run this code?

```
include <stdio.h>

int main()
{
                printf("Hello
                World"); return 0;
}
```

- How do computers communicate?

Hardware and Software - It's all one Thing

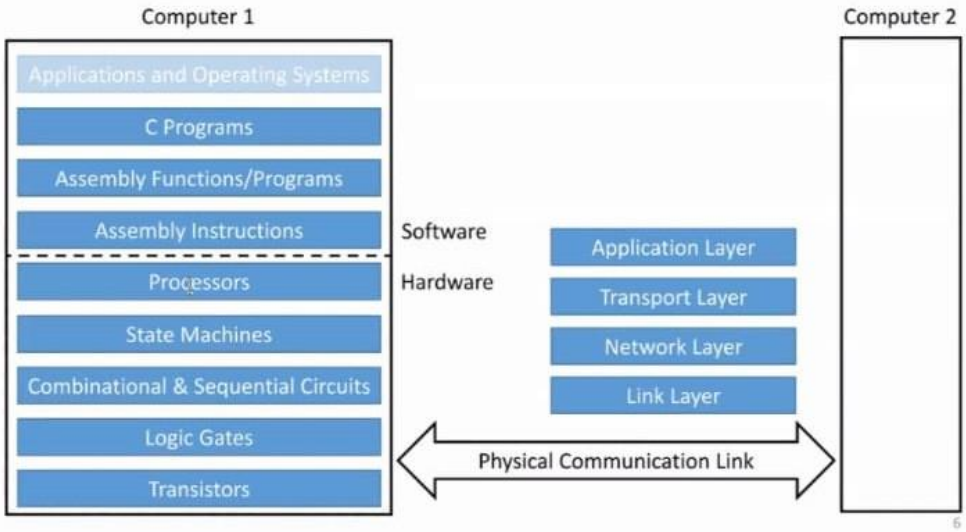


The Lecture follows a Bottom-up Approach

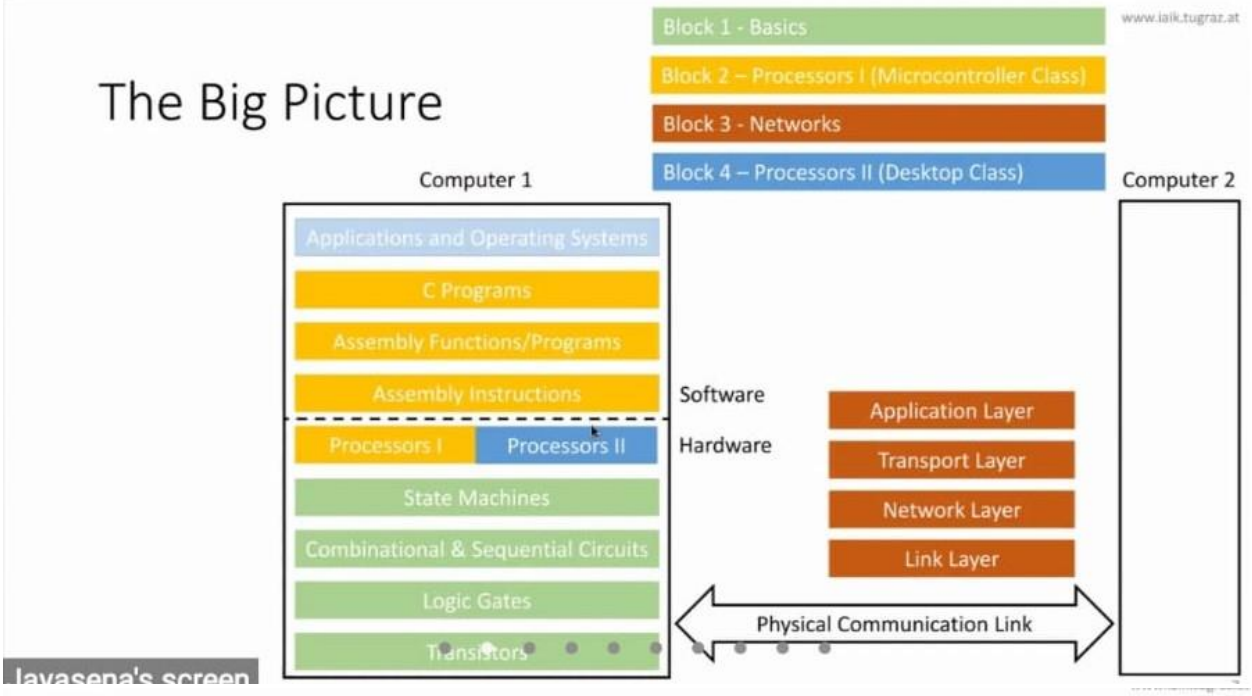
- Abstraction will be our most important tool
- We "play Lego" and we constantly build larger and more powerful

  bricks

The Big Picture



The Big Picture



Goal

- Get to know the machine you program → only this allows to write optimized code

- Understand the specifications of your device

ACM Turing Awards

- The Turing Award is the most prestigious award in computer science – it is the Noble Price of Computer Science

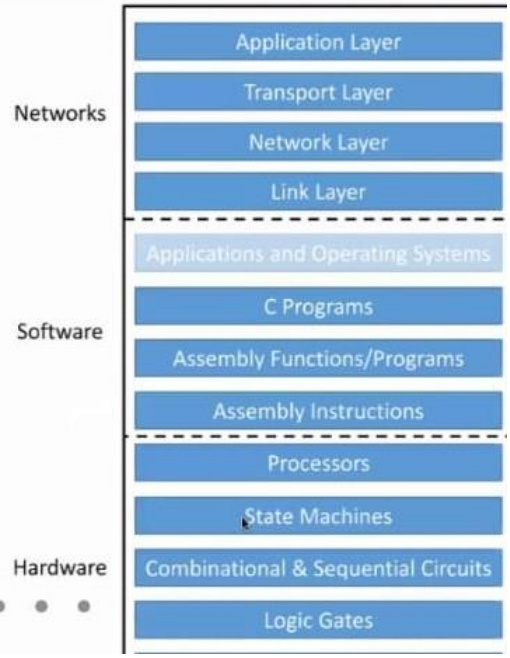
- David A. Patterson and John L. Hennessy received the Turing Award 2017 for their work on computer architectures and organization

Watch their Turing Lecture:

https://www.acm.org/hennessy-patterson-turing-lecture

Computer Organization and Networks

- In this course, we learn the basics to get the **big picture** → dig deeper in follow-up courses!

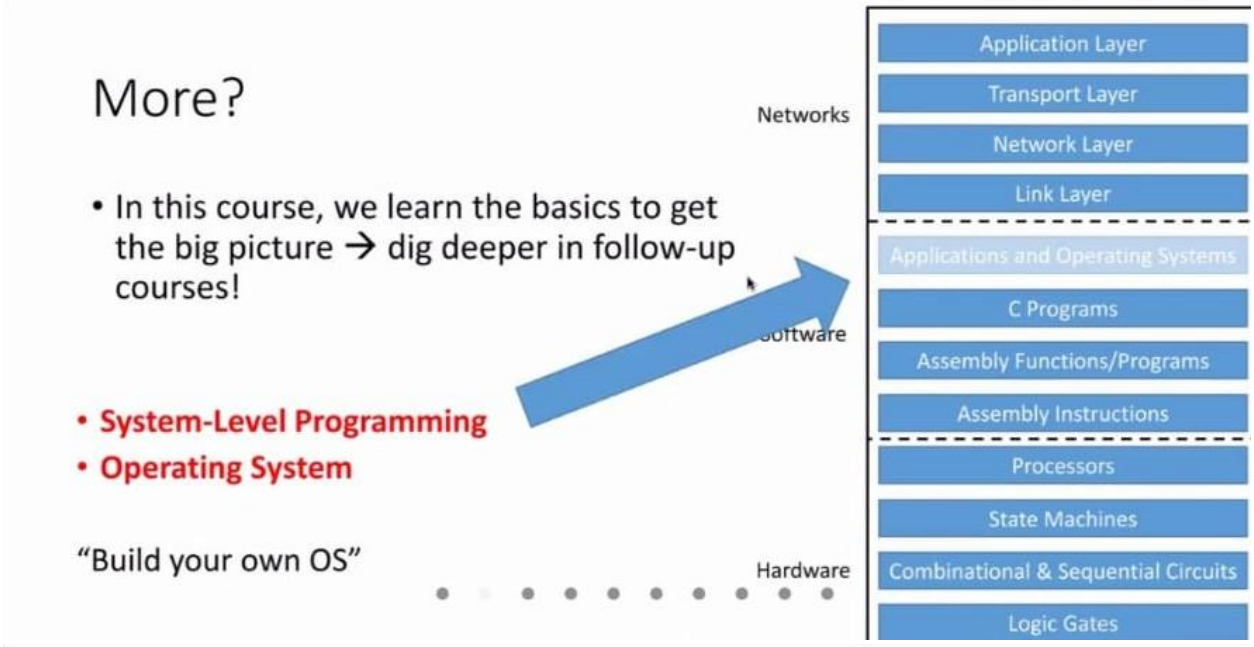


More?

- In this course, we learn the basics to get the big picture → dig deeper in follow-up courses!

- **System-Level Programming**
- **Operating System**

"Build your own OS"



More?

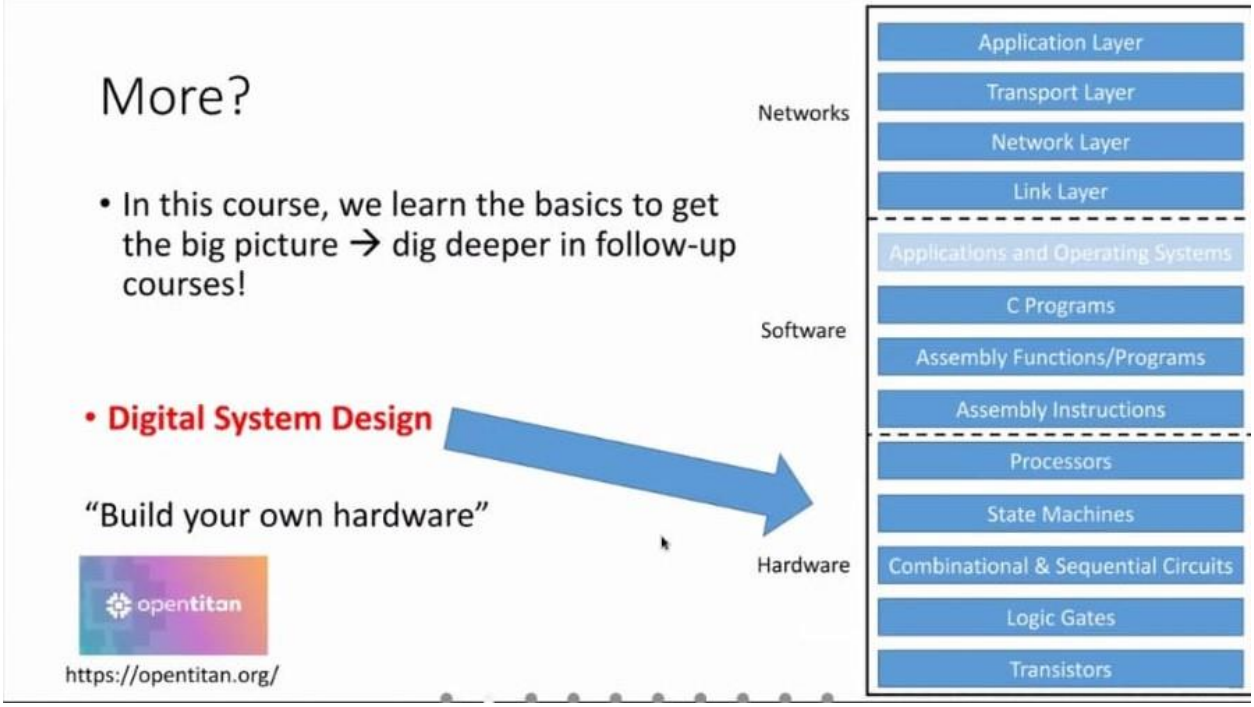
- In this course, we learn the basics to get the big picture → dig deeper in follow-up courses!

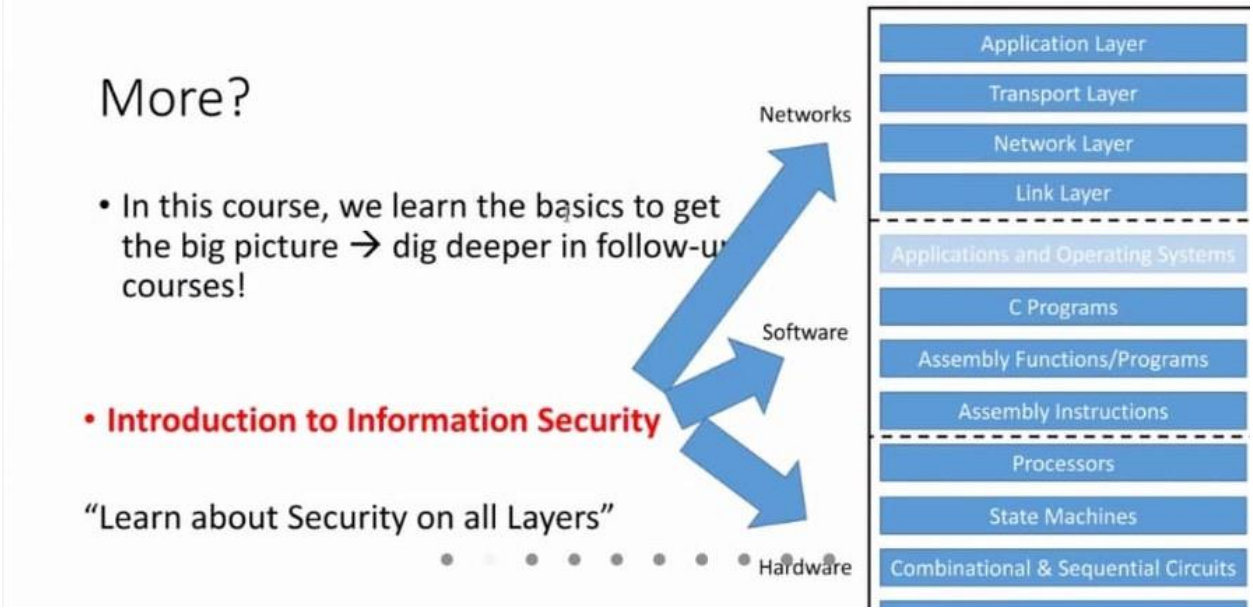
- **Digital System Design**

"Build your own hardware"

opentitan

https://opentitan.org/





Activity 2

Instruction : Answer all questions.

I/O Module	Programming I/O	Address Bus	Data Bus
Interrupt Driven I/O	Direct Memory Access	Control Bus	Tags

i. _____ carries the information being transmitted.

ii. _____ identifies where the information is being sent.

iii. _____ describes aspects of how the information is being sent, and in what manner.

iv. With _____, data are exchanged between the processor and the I/O module.

v. With _____, the processor issues an I/O command, continues to execute other instruction, and is interrupted by the I/O module when the latter has completed its work.

vi. A _____ device can transfer data directly to and from memory rather than using the CPU as an intermediary.

vii. Any program, operation or device that transfers data to or from a computer and to or from peripheral device is called _____.

viii. _____ are used identify where cached data originated.

Summary

- Two roadblocks: power and ideas
- Fixed power budget because of cooling constraints; implies th frequency can't be increased; discourages complex ideas
- End of voltage (Dennard) scaling in early 2010s
- Has led to dark silicon and dim silicon (occasional turbo)

Important Trends

- Running out of ideas to improve single thread performance
- Power wall makes it harder to add complex features
- Power wall makes it harder to increase frequency
- Additional performance provided by: more cores, occasional spikes in frequency, accelerators