UC Berkeley
Teaching Professor
Dan Garcia

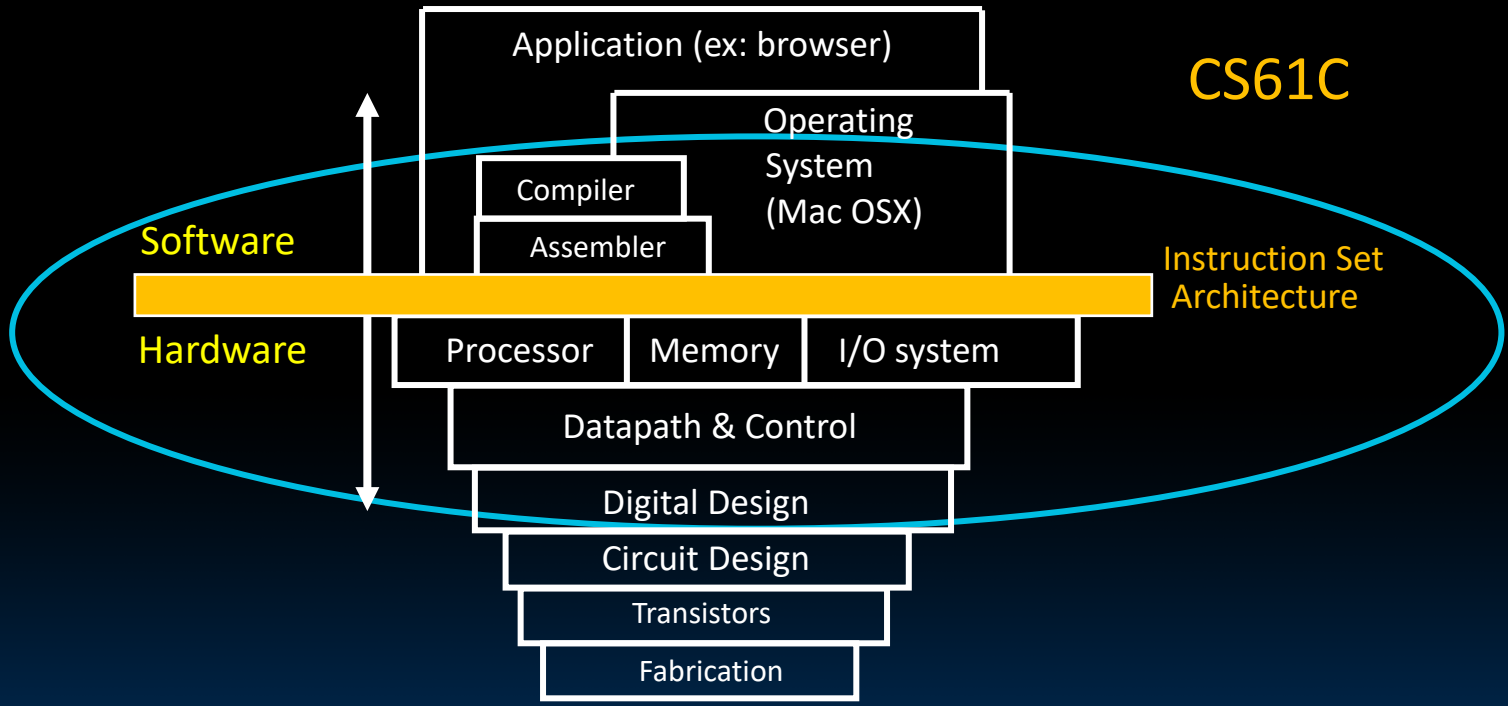
CS61C

Great Ideas in Computer Architecture (a.k.a. Machine Structures)

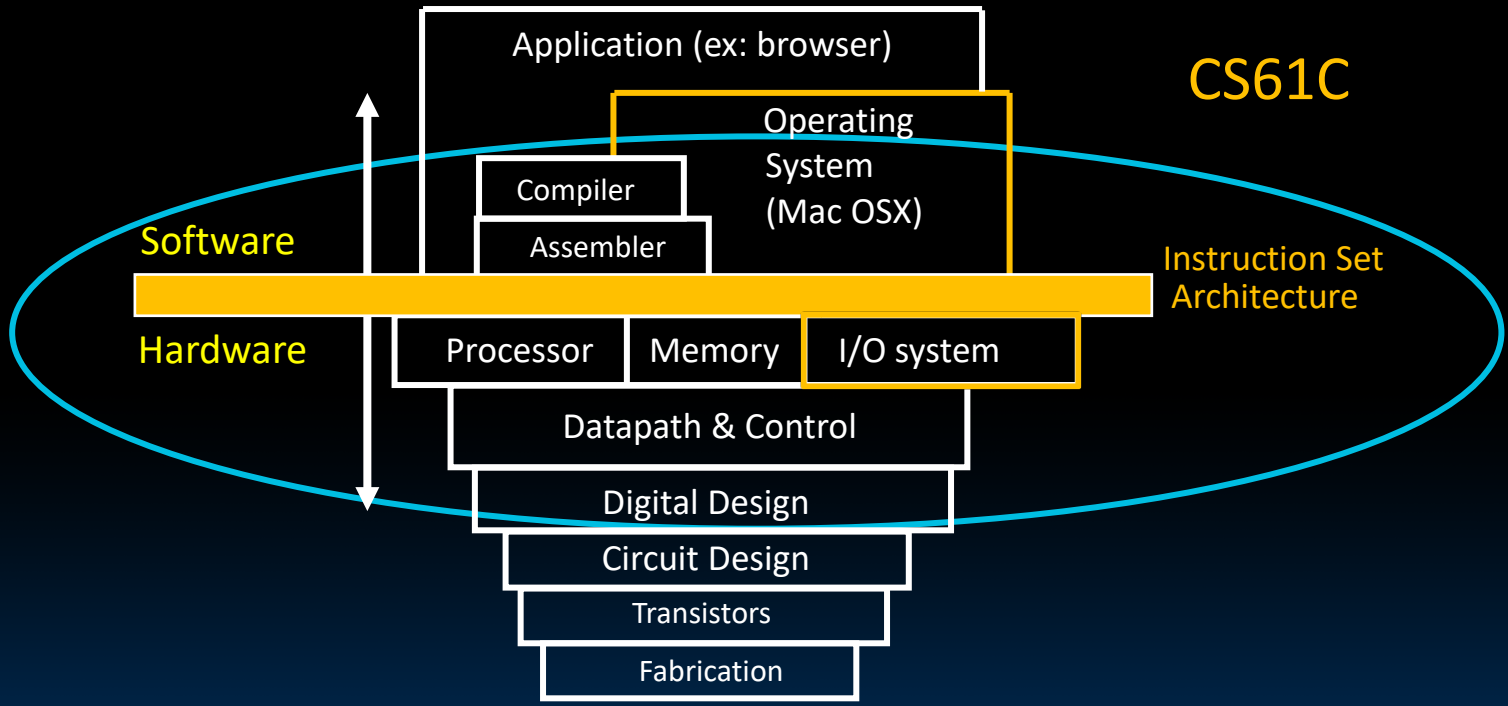
UC Berkeley
Professor
Bora Nikolić

Operating Systems and Virtual Memory

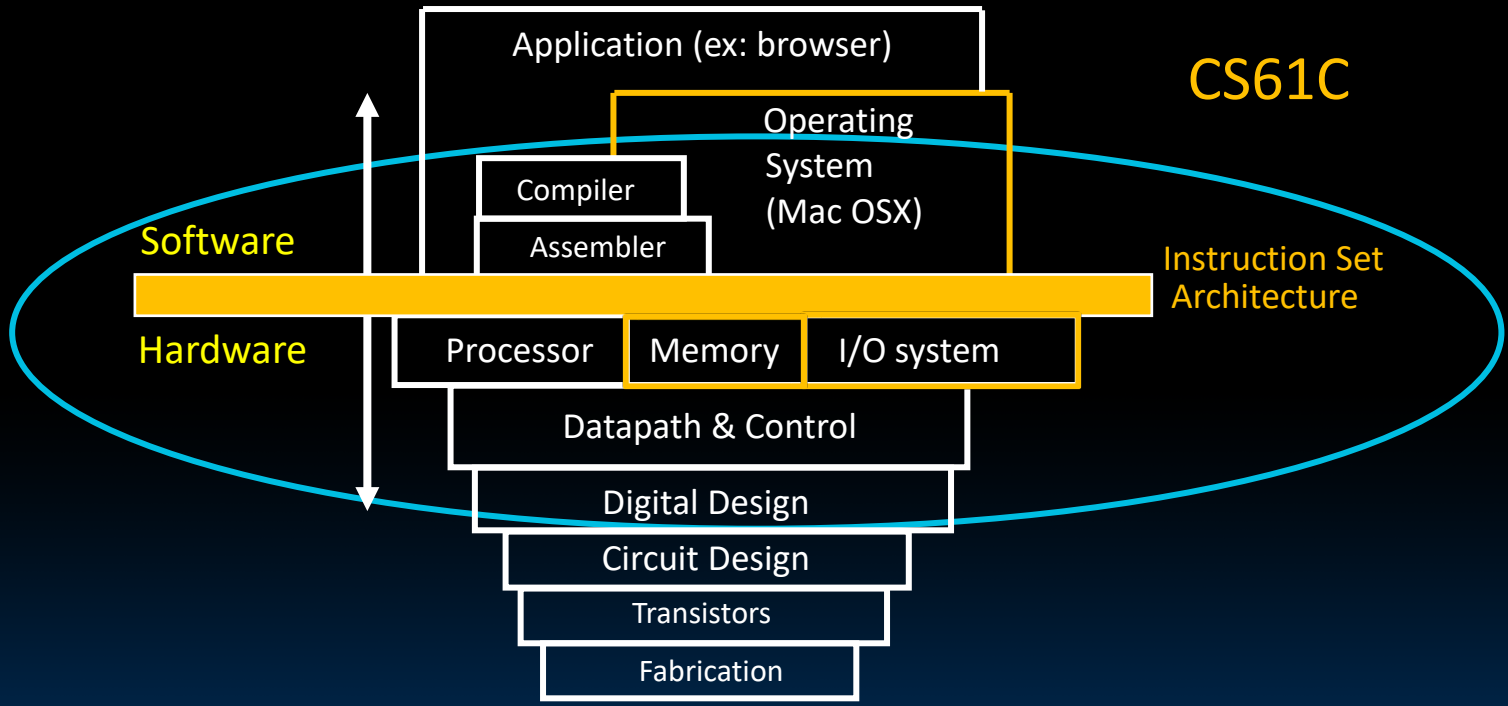
Machine Structures



Machine Structures



Machine Structures



New-School Machine Structures

Software

Parallel Requests
Assigned to computer
e.g., Search "Cats"

Parallel Threads
Assigned to core e.g., Lookup, Ads

Parallel Instructions
>1 instruction @ one time
e.g., 5 pipelined instructions

Parallel Data
>1 data item @ one time
e.g., Add of 4 pairs of words

Hardware descriptions
All gates work in parallel at same time

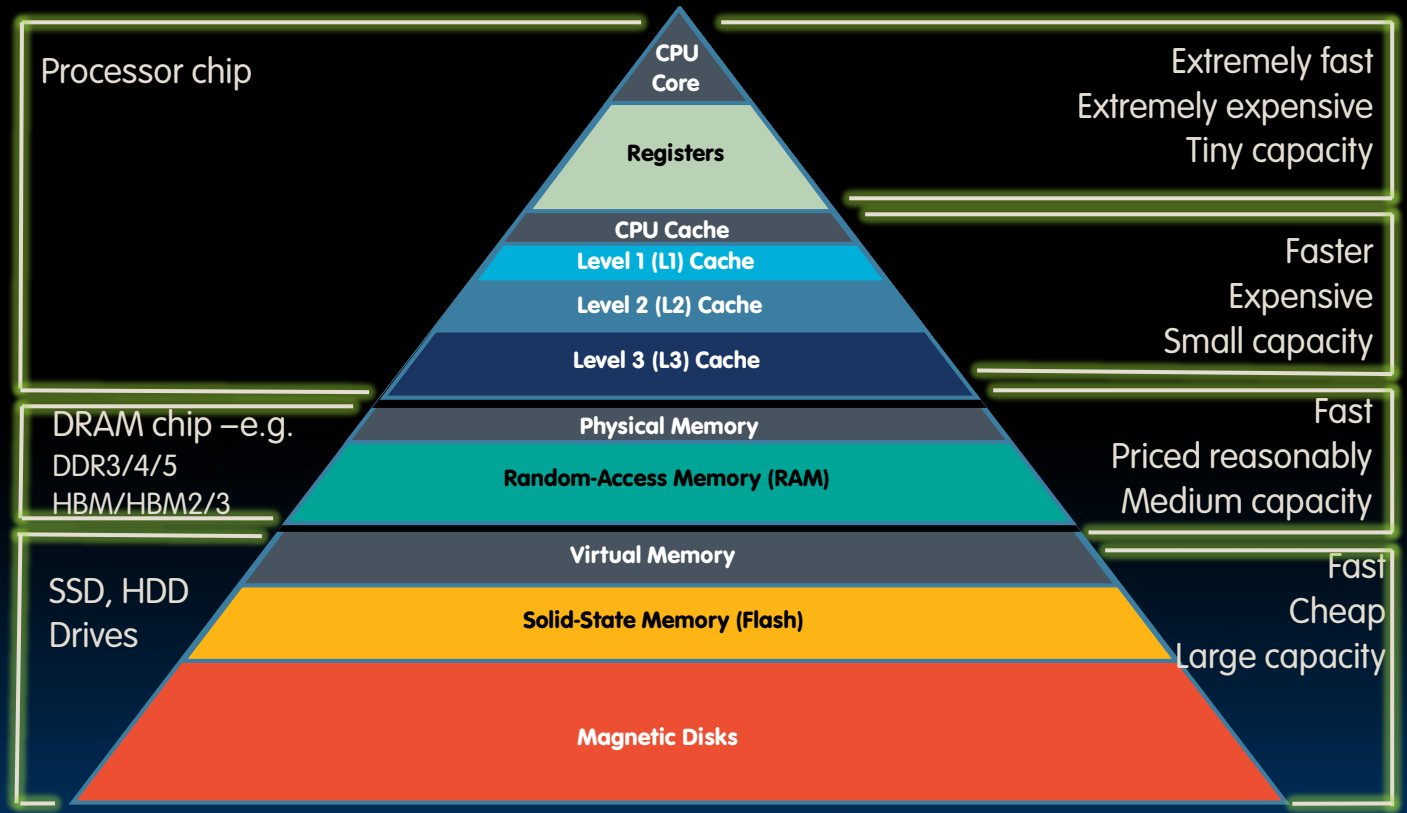
Harness Parallelism & Achieve High Performance

Hardware

Warehouse Scale Computer

Smart Phone

Great Idea #3: Principle of Locality / Memory Hierarchy



CS61C so far…

C Programs

```
#include <stdlib.h>

int fib(int n) {
  return
    fib(n-1) +
    fib(n-2);
}
```

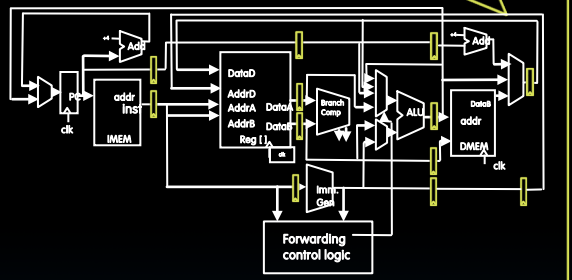
Project 1

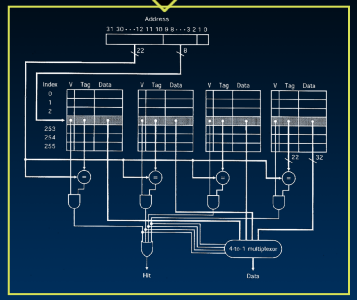
RISC-V Assembly

```
.foo
lw t0, 4(a0)
addi t1, t0, 3
beq t1, t2, foo
nop
```

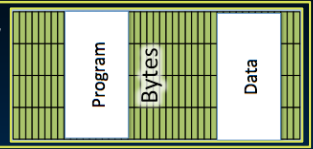
Project 2

Project 3





Memory



So How is a Laptop Any Different?

Adding I/O

C Programs

```
#include <stdlib.h>

int fib(int n) {
  return
    fib(n-1) +
    fib(n-2);
}
```

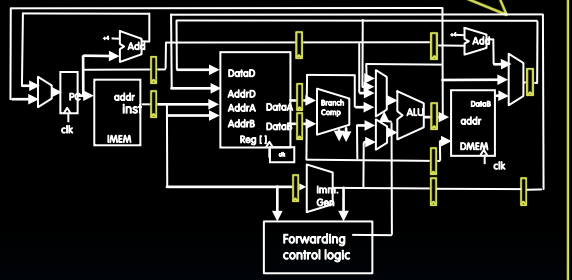
Project 1

RISC-V Assembly

```
.foo
lw t0, 4(a0)
addi t1, t0, 3
beq t1, t2, foo
nop
```

Project 2

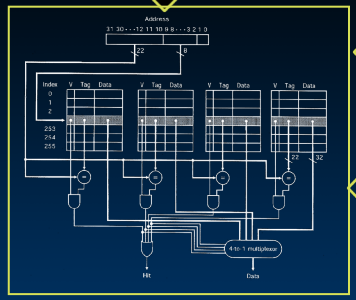
Project 3



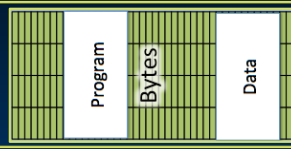
This module: Memory (DRAM) Storage (Disk) and I/O

Screen | Keyboard | Storage

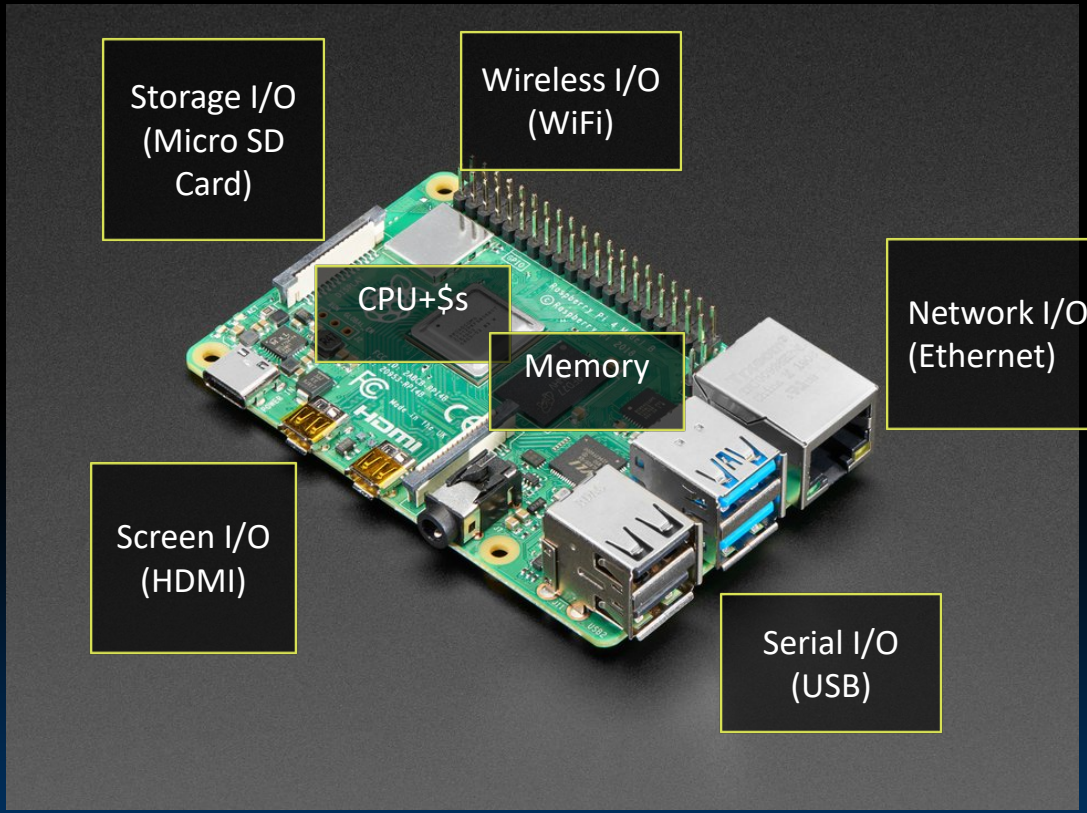
I/O (Input/Output)



Memory



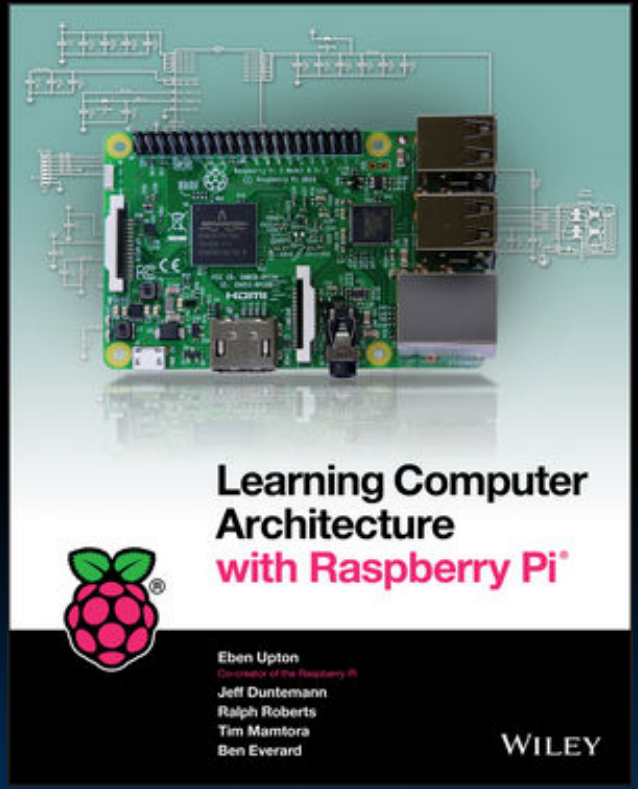
Raspberry Pi ($35)



It's a Real Computer!

CS61C with Raspberry PI?

- Lot's of concepts from 61C covered in a book, with Raspberry Pi exercises

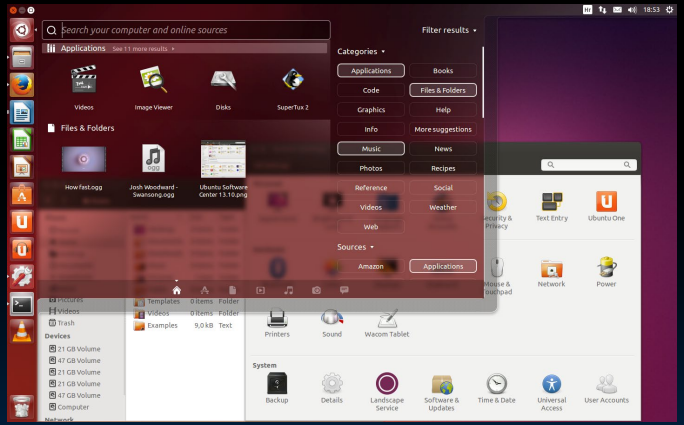


(and it is free to download if you are a Cal student:

https://onlinelibrary.wiley.com/doi/book/10.1002/9781119415534)

But Wait...

- That's not the same! When we run VENUS, it only executes one program and then stops.
- When I switch on my computer, I get this:

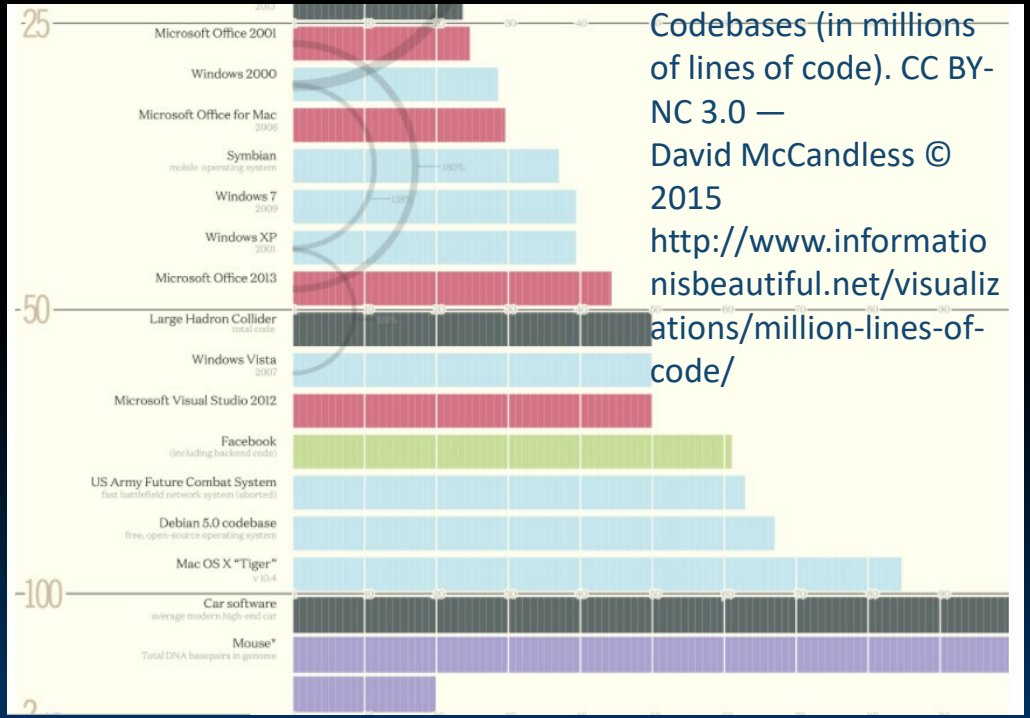


Yes, but that's *just* software! The Operating System (OS)

Operating System Basics

Well, "Just Software"

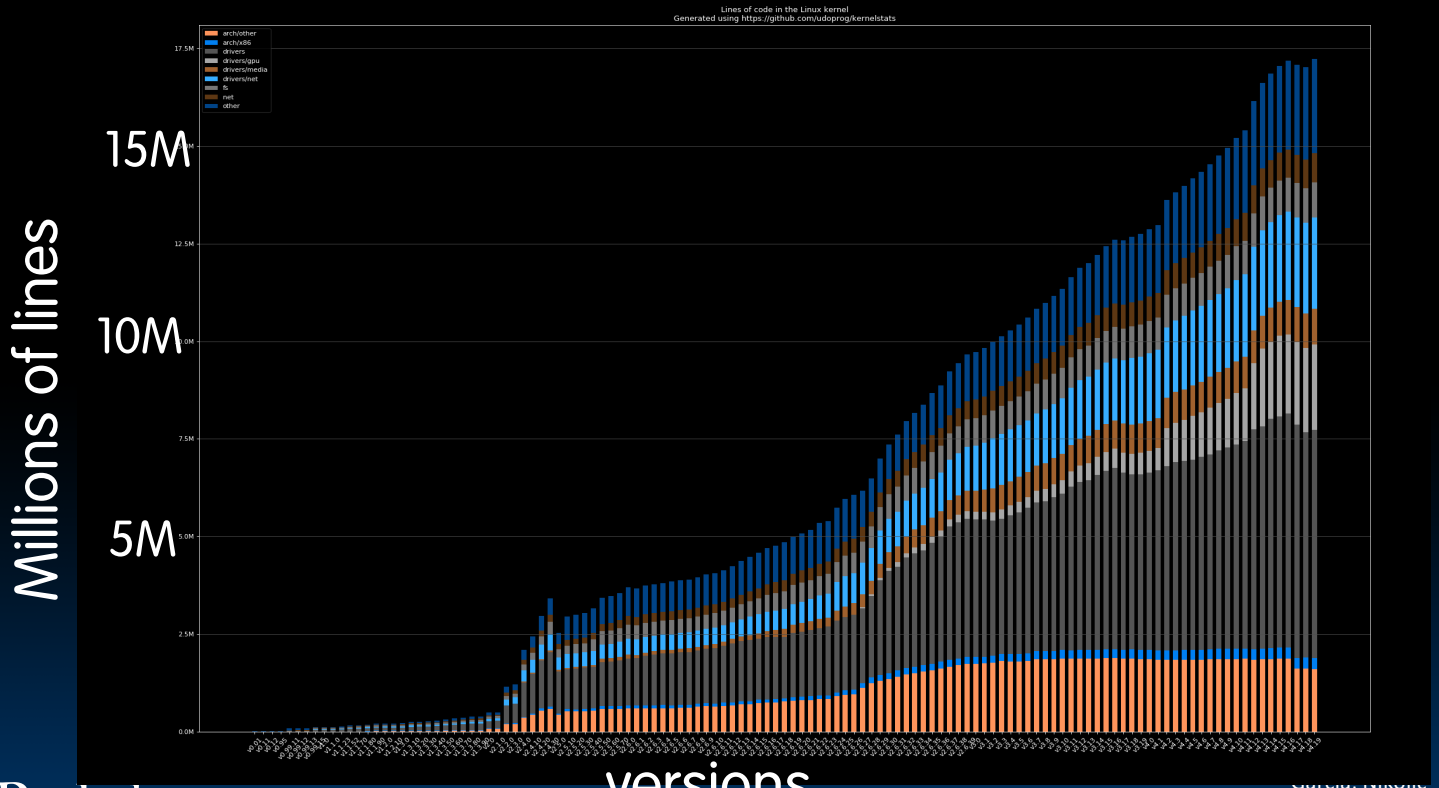
- The biggest piece of software on your machine?
- How many lines of code? These are guesstimates:



Codebases (in millions of lines of code). CC BY-NC 3.0 — David McCandless © 2015 http://www.informationisbeautiful.net/visualizations/million-lines-of-code/

Operating System

Lines of code in Linux kernel



What Does the OS do?

- OS is the (first) thing that runs when computer starts
- Finds and controls all devices in the machine in a general way
  - Relying on hardware specific "device drivers"
- Starts services (100+)
  - File system,
  - Network stack (Ethernet, WiFi, Bluetooth, …),
  - TTY (keyboard),
  - …
- Loads, runs and manages programs:
  - Multiple programs at the same time (time-sharing)
  - Isolate programs from each other (isolation)
  - Multiplex resources between applications (e.g., devices)

What Does the Core of the OS Do?

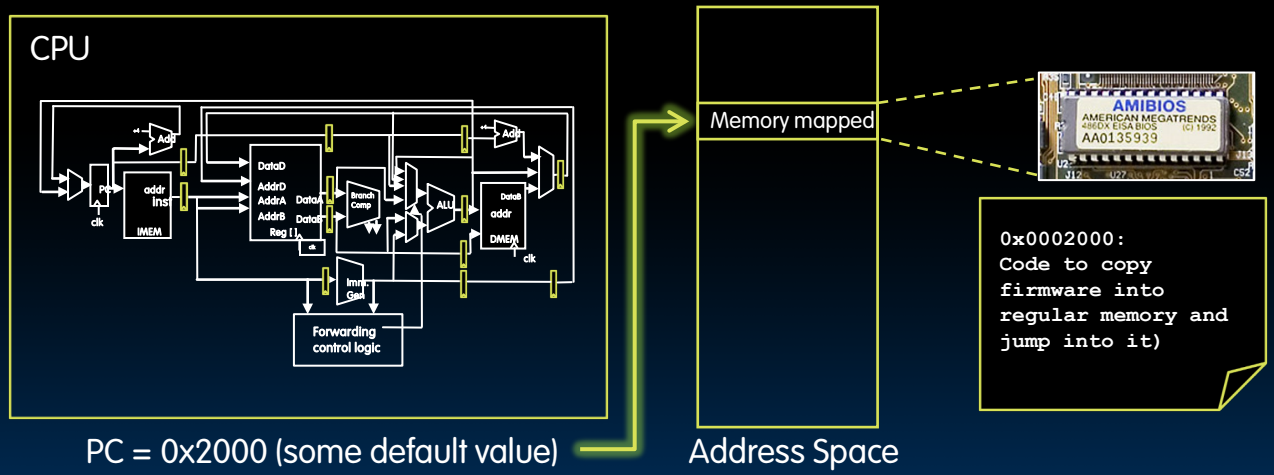
- Provide ***isolation*** between running processes
  - Each program runs in its own little world •
- Provide ***interaction*** with the outside world
  - Interact with "devices": Disk, display, network, etc... 11

What Does OS Need from Hardware?

- Memory translation
  - Each running process has a mapping from "virtual" to "physical" addresses that are different for each process
  - When you do a load or a store, the program issues a virtual address... But the actual memory accessed is a physical address
- Protection and privilege
  - Split the processor into at least two running modes: "User" and "Supervisor"
    - RISC-V also has "Machine" below "Supervisor"
  - Lesser privilege ***can not*** change its memory mapping
    - But "Supervisor" ***can*** change the mapping for any given program
    - And supervisor has its own set of mapping of virtual->physical
- Traps & Interrupts
  - A way of going into Supervisor mode on demand

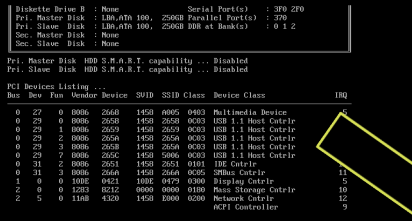
What Happens at Boot?

- When the computer switches on, it does the same as VENUS: the CPU executes instructions from some start address (stored in Flash ROM)

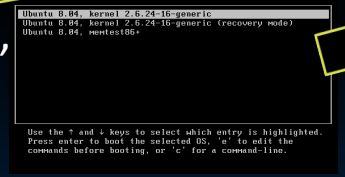


What Happens at Boot?

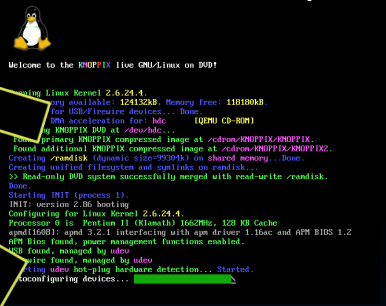
1. BIOS*: Find a storage device and load first sector (block of data)



2. Bootloader (stored on, e.g., disk): Load the OS *kernel* from disk into a location in memory and jump into it



3. OS Boot: Initialize services, drivers, etc.



*BIOS: Basic Input Output System

Operating System Functions

Launching Applications

- Applications are called “processes” in most OSs
  - Thread: shared memory
  - Process: separate memory
  - Both threads and processes run (pseudo) simultaneously
- Apps are started by another process (e.g., shell) calling an OS routine (using a “syscall”)
  - Depends on OS; Linux uses fork to create a new process, and execve (execute file command) to load application
- Loads executable file from disk (using the file system service) and puts instructions & data into memory (.text, .data sections), prepares stack and heap
- Set argc and argv, jump to start of main
- Shell waits for main to return (join)

Supervisor Mode

- If something goes wrong in an application, it could crash the entire machine. And what about malware, etc.?
- The OS enforces resource constraints to applications (e.g., access to memory, devices)
- To help protect the OS from the application, CPUs have a supervisor mode (e.g., set by a status bit in a special register)
  - A process can only access a subset of instructions and (physical) memory when not in supervisor mode (user mode)
  - Process can change out of supervisor mode using a special instruction, but not into it directly – only using an interrupt
  - Supervisory mode is a bit like "superuser"
    - But used much more sparingly (most of OS code does *not* run in supervisory mode)
    - Errors in supervisory mode often catastrophic (blue "screen of death", or "I just corrupted your disk")

Syscalls

- What if we want to call an OS routine? E.g.,
  - to read a file,
  - launch a new process,
  - ask for more memory (malloc),
  - send data, etc.
- Need to perform a syscall:
  - Set up function arguments in registers,
  - Raise software interrupt (with special assembly instruction)
- OS will perform the operation and return to user mode
- This way, the OS can mediate access to all resources, and devices

Interrupts, Exceptions

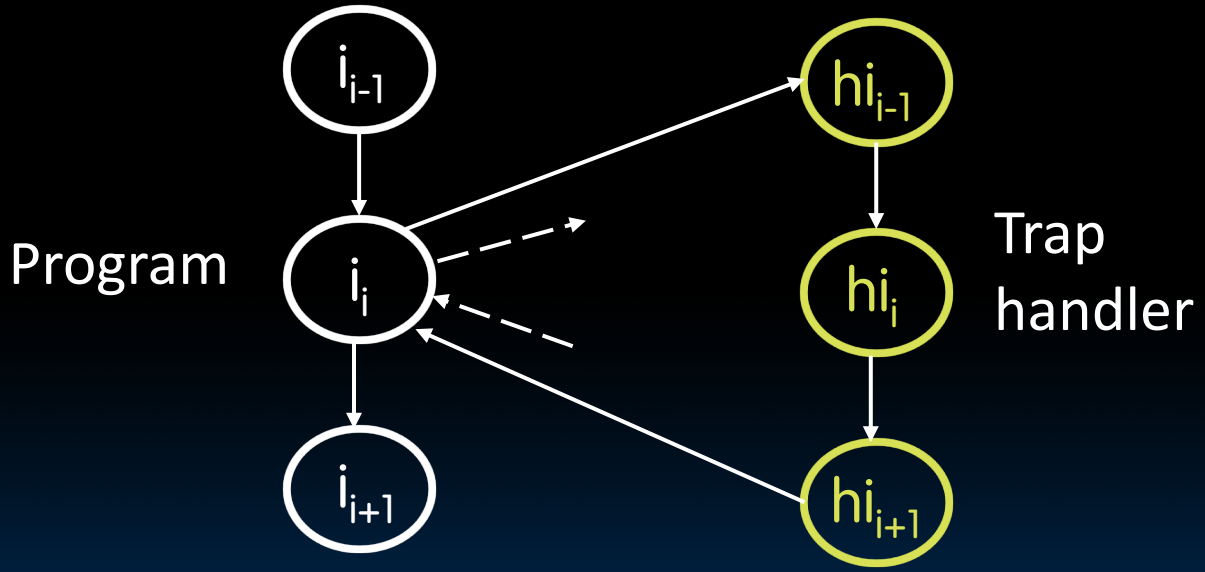
- We need to transition into Supervisor mode when "something" happens
- Interrupt: Something external to the running program
  - Something happens from the outside world
- Exception: Something done by the running program
  - Accessing memory it isn't "supposed" to, executing an illegal instruction, reading a csr not supposed at that privilege
- **ECALL**: Trigger an exception to the higher privilege
  - How you communicate with the operating system: Used to implement "syscalls"
- **EBREAK**: Trigger an exception within the current privilege

Terminology (In 61C)

- Interrupt – caused by an event *external* to current running program
  - E.g., key press, disk I/O
  - Asynchronous to current program
    - Can handle interrupt on any convenient instruction
    - "Whenever it's convenient, just don't wait too long"
- Exception – caused by some event *during* execution of one instruction of current running program
  - E.g., memory error, bus error, illegal instruction, raised exception
  - Synchronous
    - Must handle exception *precisely* on instruction that causes exception
    - "Drop whatever you are doing and act now"
- Trap – action of servicing interrupt or exception by hardware jump to "interrupt or trap handler" code

Trap Handling

- Altering the regular execution flow

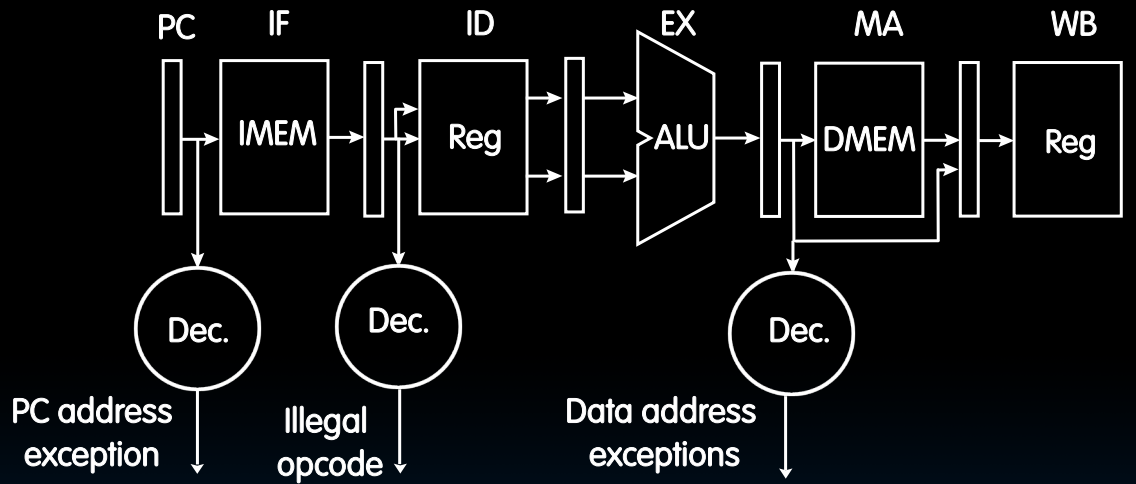


An *external or internal event* that needs to be processed - by another program; often handled by OS. The event is often unexpected from original program's point of view.

Precise Traps

- *Trap handler's view of machine state is that every instruction prior to the trapped one (e.g., memory error) has completed, and no instruction after the trap has executed.*
- Implies that handler can return from an interrupt by restoring user registers and jumping back to interrupted instruction
  - Interrupt handler software doesn't need to understand the pipeline of the machine, or what program was doing!
  - More complex to handle trap caused by an exception than interrupt
- Providing precise traps is tricky in a pipelined superscalar out-of-order processor!
  - But a requirement for things to actually work right!

Exceptions in a 5-Stage Pipeline



Trap Handling

- Exceptions are handled *like pipeline hazards*
  1) Complete execution of instructions before exception occurred
  2) Flush instructions currently in pipeline (i.e., convert to `nop`s or "bubbles")
  3) Optionally store exception cause in status register
     - Indicate type of exception
- 4) Transfer execution to trap handler
- 5) Optionally, return to original program and re-execute instruction

Multiprogramming

- The OS runs multiple applications at the same time
- But not really (unless you have a core per process)
- Switches between processes very quickly (on human time scale) – this is called a "context switch"
- When jumping into process, set timer (we will call this 'interrupt')
  - When it expires, store PC, registers, etc. (process state)
  - Pick a different process to run and load its state
  - Set timer, change to user mode, jump to the new PC
- Deciding what process to run is called scheduling

Protection, Translation, Paging

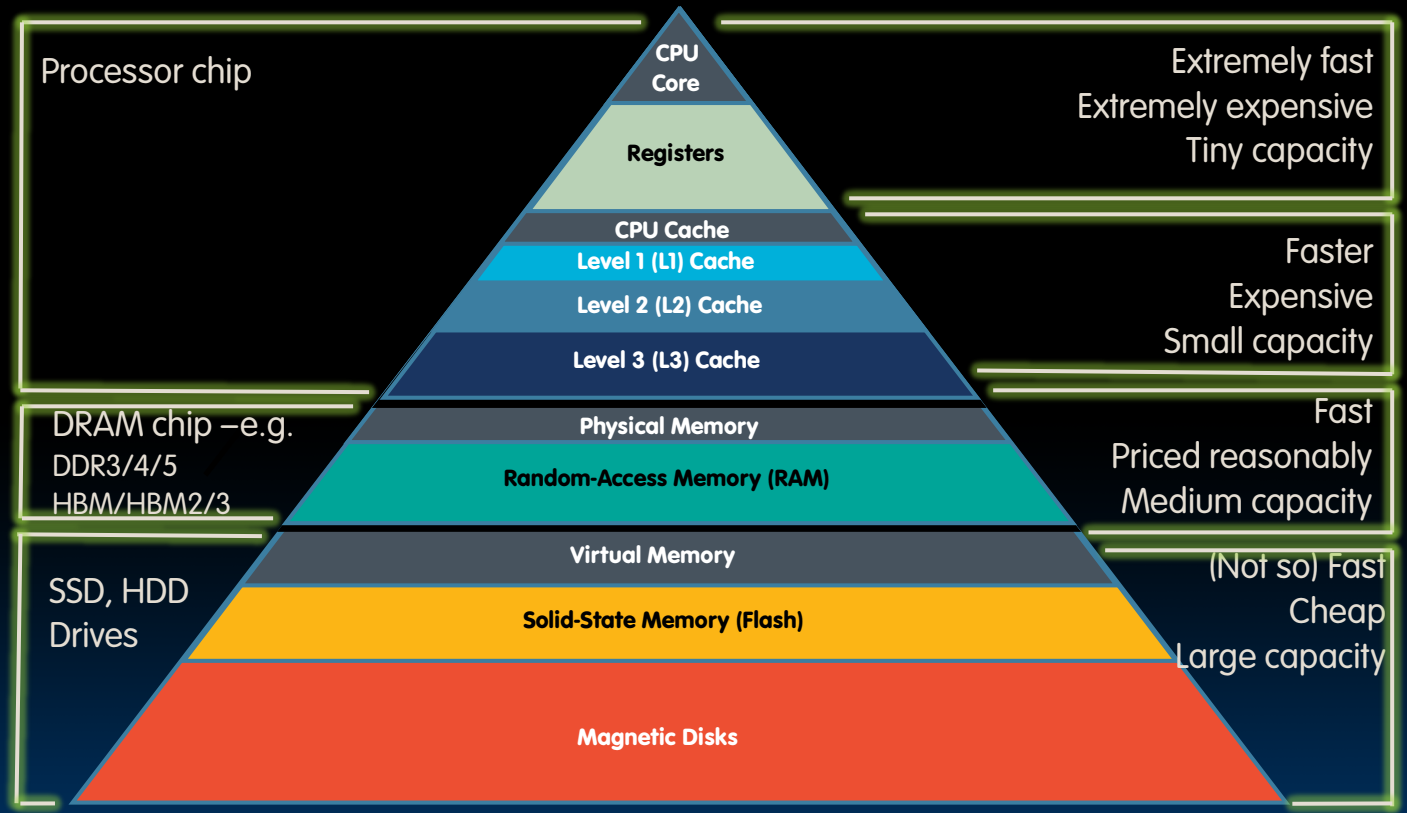
- Supervisor mode alone is not sufficient to fully isolate applications from each other or from the OS
  - Application could overwrite another application's memory.
  - Typically programs start at some fixed address, e.g. 0x8FFFFFFF
    - How can 100's of programs share memory at location 0x8FFFFFFF?
  - Also, may want to address more memory than we actually have (e.g., for sparse data structures)
- Solution: Virtual Memory
  - Gives each process the *illusion* of a full memory address space that it has completely for itself

Virtual Memory Concepts

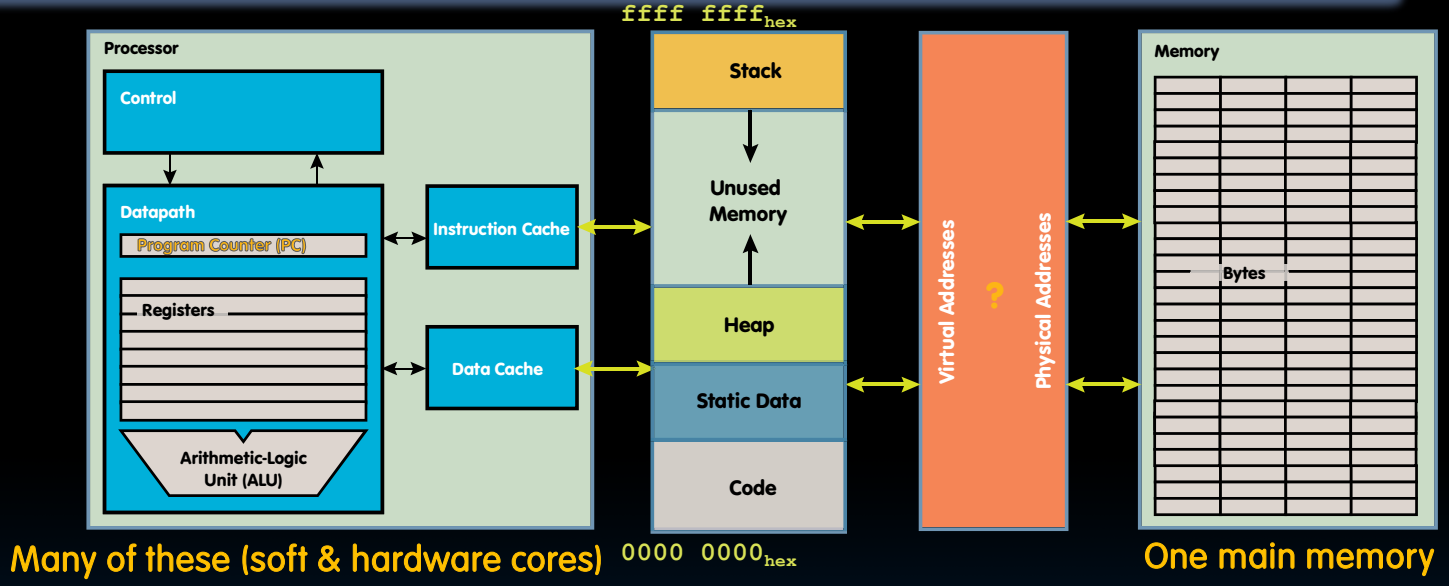
Virtual Memory

- Virtual memory - Next level in the memory hierarchy:
  - Provides program with illusion of a very large main memory: Working set of “pages” reside in main memory - others are on disk
  - Demand paging: Provides the ability to run programs larger than the primary memory (DRAM)
  - Hides differences between machine configurations
- Also allows OS to share memory, protect programs from each other
- Today, more important for protection than just another level of memory hierarchy
- Each process thinks it has all the memory to itself
- (Historically, it predates caches)

Great Idea #3: Principle of Locality / Memory Hierarchy



Virtual vs. Physical Addresses



- Processes use virtual addresses, e.g., 0 … 0xffff,ffff
  - Many processes, all using same (conflicting) addresses
- Memory uses physical addresses (also, e.g., 0 … 0xffff,ffff)
  - *Memory manager maps virtual to physical addresses*

Address Spaces

- Address space = set of addresses for all available memory locations
- Now, two kinds of memory addresses:
  - **Virtual Address Space**
    - Set of addresses that the user program knows about
  - **Physical Address Space**
    - Set of addresses that map to actual physical locations in memory
    - Hidden from user applications
- Memory manager maps (‘translates’) between these two address spaces

Bora's Laptop

```
bora@DESKTOP-QAB1DCR:~$ lscpu
Architecture:                    x86_64
CPU op-mode(s):                  32-bit, 64-bit
Byte Order:                      Little Endian
Address sizes:                   39 bits physical, 48 bits virtual
CPU(s):                          8
On-line CPU(s) list:             0-7
Thread(s) per core:              2
Core(s) per socket:              4
Socket(s):                       1
Vendor ID:                       GenuineIntel
CPU family:                      6
Model:                           126
Model name:                      Intel(R) Core(TM) i7-1065G7 CPU @ 1.30GHz
Stepping:                        5
CPU MHz:                         1497.604
BogoMIPS:                        2995.20
Hypervisor vendor:               Microsoft
Virtualization type:             full
L1d cache:                       192 KiB
L1i cache:                       128 KiB
L2 cache:                        2 MiB
L3 cache:                        8 MiB
```

Analogy

- Book title like virtual address
- Library of Congress call number like physical address
- Card catalogue like page table, mapping from book title to call #
- On card for book, in local library vs. in another branch like valid bit indicating in main memory vs. on disk (storage)
- On card, available for 2-hour in library use (vs. 2-week checkout) like access rights

Memory Hierarchy Requirements

- Allow multiple processes to simultaneously occupy memory and provide protection – don't let one program read/write memory from another

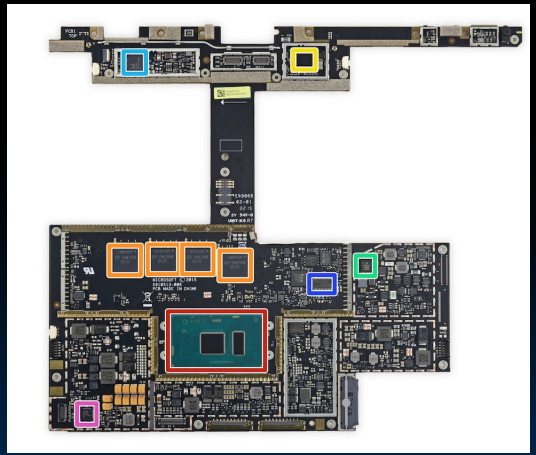
- Address space – give each program the illusion that it has its own private memory
  - Suppose code starts at address 0x40000000. But different processes have different code, both residing at the same address. So each program has a different view of memory.

Physical Memory and Storage

Memory

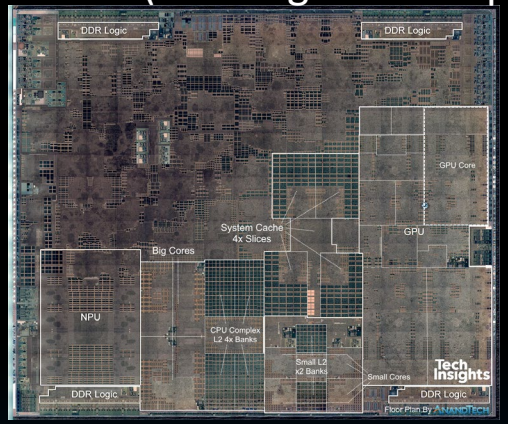
- Memory (DRAM)

Desktop/server



MS Surface Book

Apple A12 Bionic
(DRAM goes on top)



Volatile

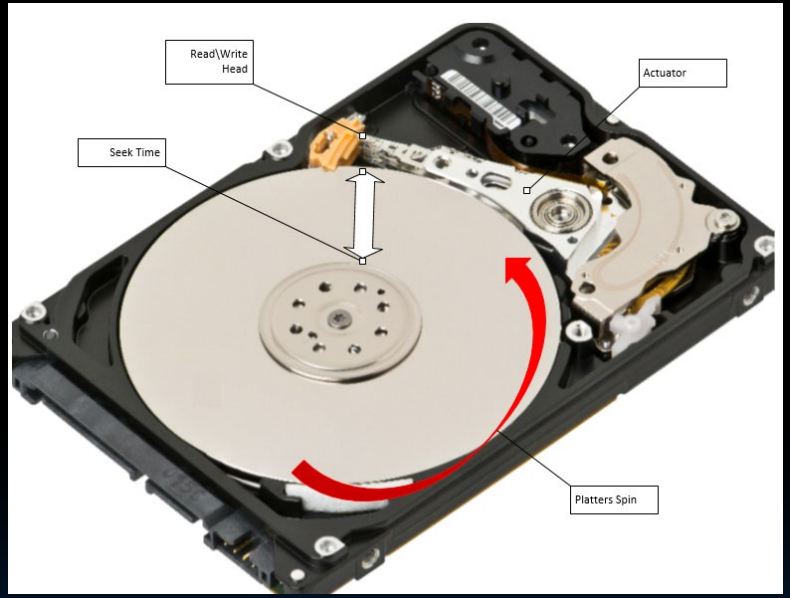
Latency to access first word: ~10ns
(~30-40 processor cycles)
Each successive (0.5ns – 1ns)
Each access brings 64 bits
Supports 'bursts'

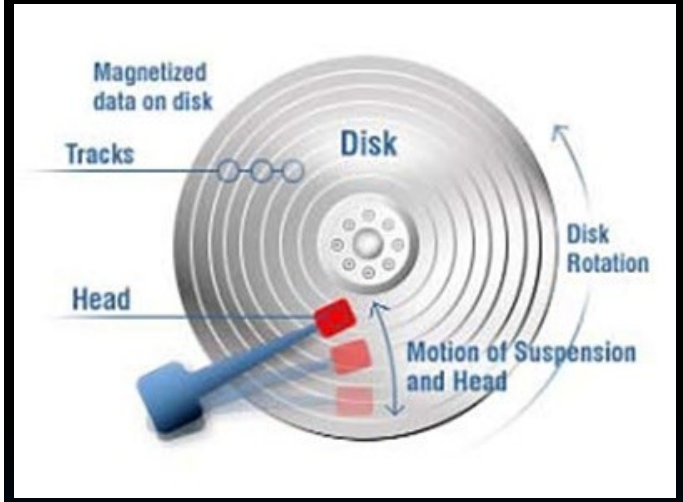
Storage – "Disk"

Attached as a peripheral I/O device; Non-volatile

- SSD
  - Access: 40-100µs (~100k proc. cycles)
  - $0.05-0.5/GB
- HDD
  - Access: <5-10ms (10-20M proc. cycles)
  - $0.01-0.1/GB

Aside … Why are Disks So Slow?





- 10,000 rpm (revolutions per minute)
- 6 ms per revolution
- Average random access time: 3 ms (~10^7 processor cycles)

How Hard Drives Work?

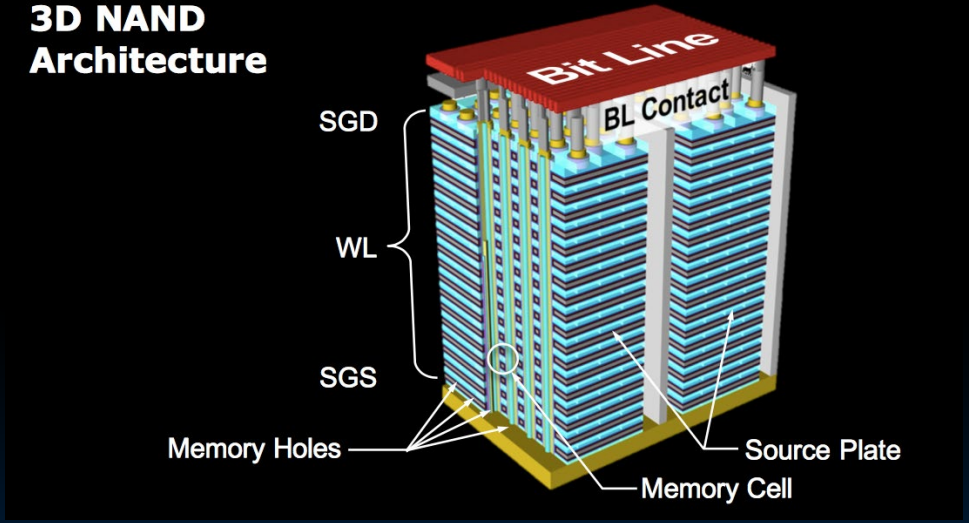
- Nick Parlante's https://cs.stanford.edu/people/nick/how-hard-drive-works/
  - Several YouTube videos as well

What About SSD?

- Made with transistors
- Nothing mechanical that turns
- Like “Ginormous” register file
  - Does not ”forget” when power is off (non-volatile)
- Fast access to all locations, regardless of address
- Still much slower than register, DRAM
  - Read/write blocks, not bytes
  - Potential reliability issues
- Some unusual requirements:
  - Can’t erase single bits – only entire blocks

Flash Memory

- 3D array of bit cells (up to 256 layers!)

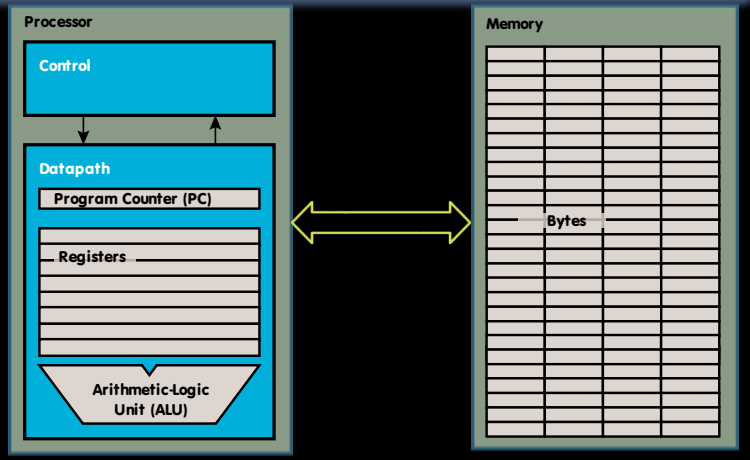


Western Digital/Semiengineering

Memory Manager

Virtual Memory

- In a ‘bare metal’ system (w/o OS), addresses issued with loads/stores are real physical addresses

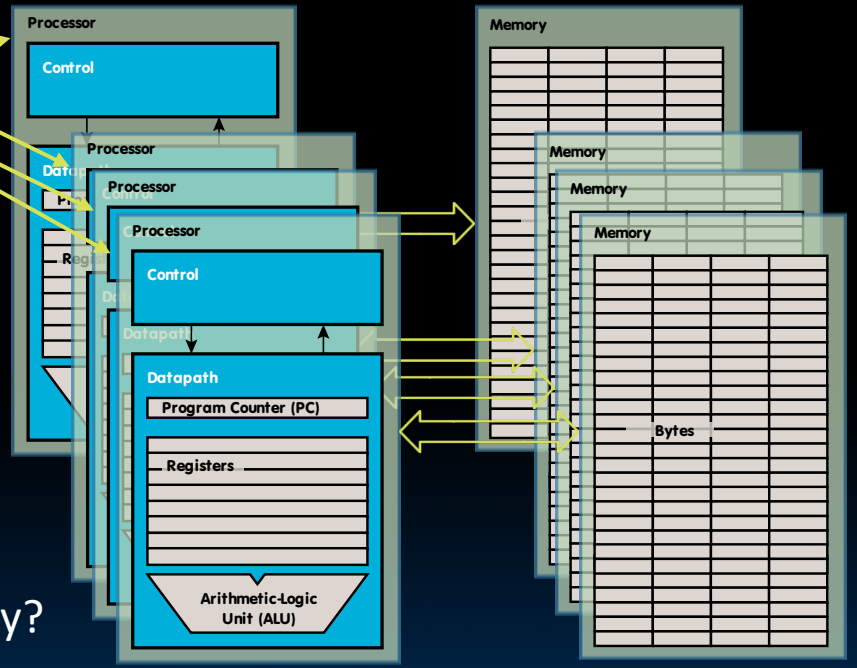


  - In this mode, any process can issue any address, therefore can access any part of memory, even areas which it doesn’t own
    - Ex: The OS data structures
  - We should send all addresses through a mechanism that the OS controls, before they make it out to DRAM - a translation mechanism
    - Check that process has permission to access a particular part of memory

Virtual Memory

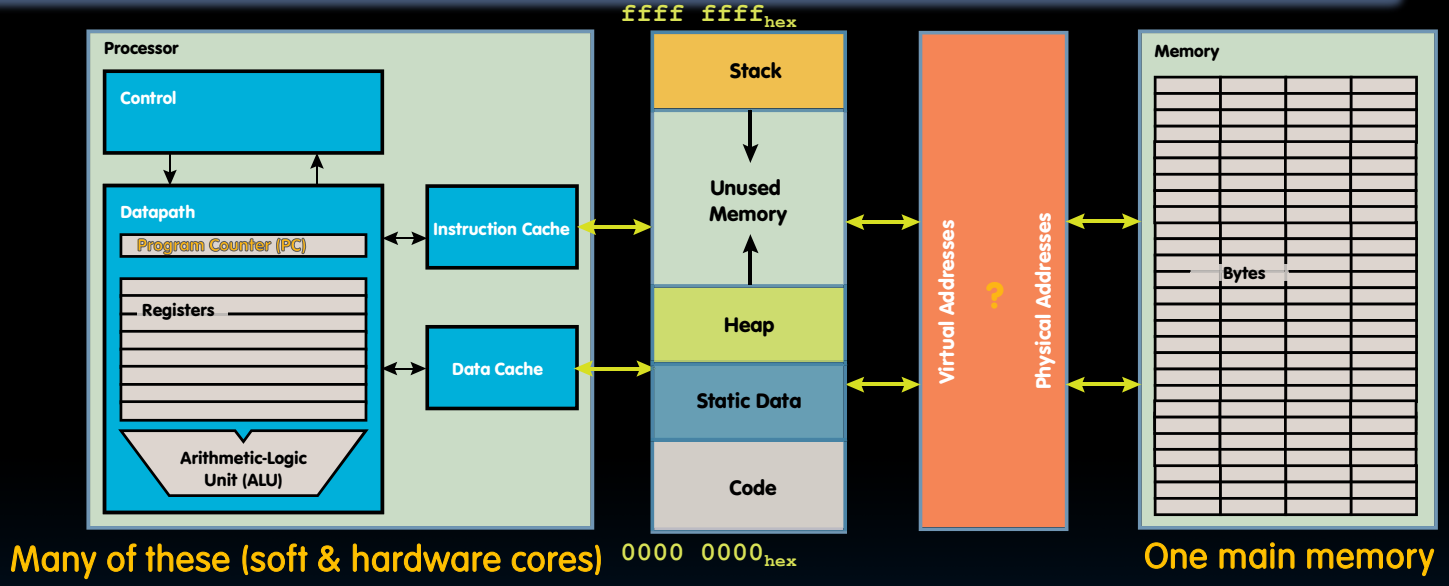
100+ Processes, managed by OS

```
0:04.34 /usr/libexec/UserEventAgent (Aqua)
0:10.60 /usr/sbin/distnoted agent
0:09.11 /usr/sbin/cfprefsd agent
0:04.71 /usr/sbin/usernoted
0:02.35 /usr/libexec/nsurlsessiond
0:28.68 /System/Library/PrivateFrameworks/Calend
0:04.36 /System/Library/PrivateFrameworks/GameCe
0:01.90 /System/Library/CoreServices/cloudphotos
0:49.72 /usr/libexec/secinitd
0:01.66 /System/Library/PrivateFrameworks/TCC.fr
0:12.68 /System/Library/Frameworks/Accounts.fram
0:09.56 /usr/libexec/SafariCloudHistoryPushAgent
0:00.27 /System/Library/PrivateFrameworks/CallHi
0:00.74 /System/Library/CoreServices/mapspushd
0:00.79 /usr/libexec/fmfd
```



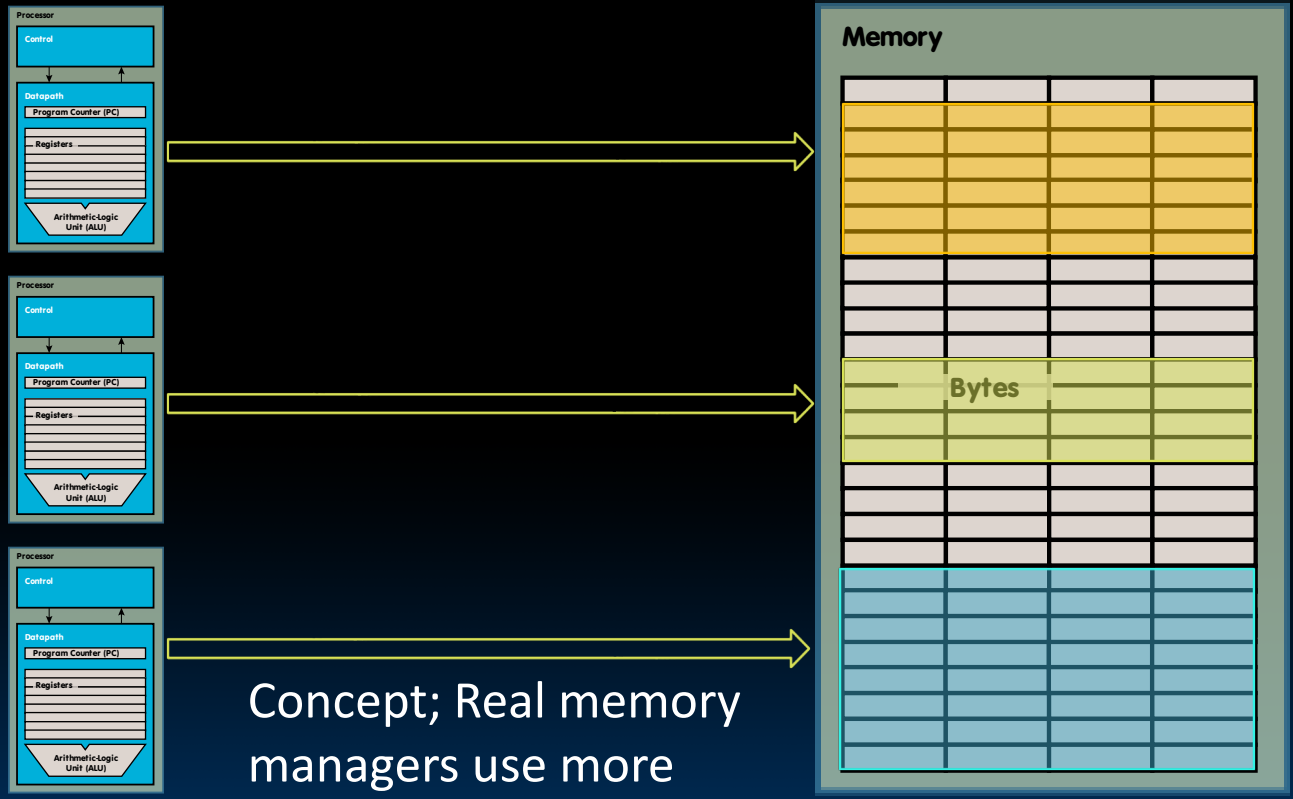
- 100's of processes
  - OS multiplexes these over available cores
- But what about memory?
  - There is only one!
  - We cannot just "save" its contents in a context switch …

Review: Virtual vs. Physical Addresses



- Processes use virtual addresses, e.g., 0 … 0xffff,ffff
  - Many processes, all using same (conflicting) addresses
- Memory uses physical addresses (also, e.g., 0 … 0xffff,ffff)
  - *Memory manager maps virtual to physical addresses*

Conceptual Memory Manager



Concept; Real memory managers use more complex mappings

Responsibilities of Memory Manager

1) Map virtual to physical addresses
2) Protection:
   - Isolate memory between processes
   - Each process gets dedicate ”private” memory
   - Errors in one program won’t corrupt memory of other program
   - Prevent user programs from messing with OS’s memory
3) Swap memory to disk
   - Give illusion of larger memory by storing some content on disk
   - Disk is usually much larger and slower than DRAM
     - Use “clever” caching strategies