K22 – Operating Systems: Design Principles and Internals

Fall 2025 @dit

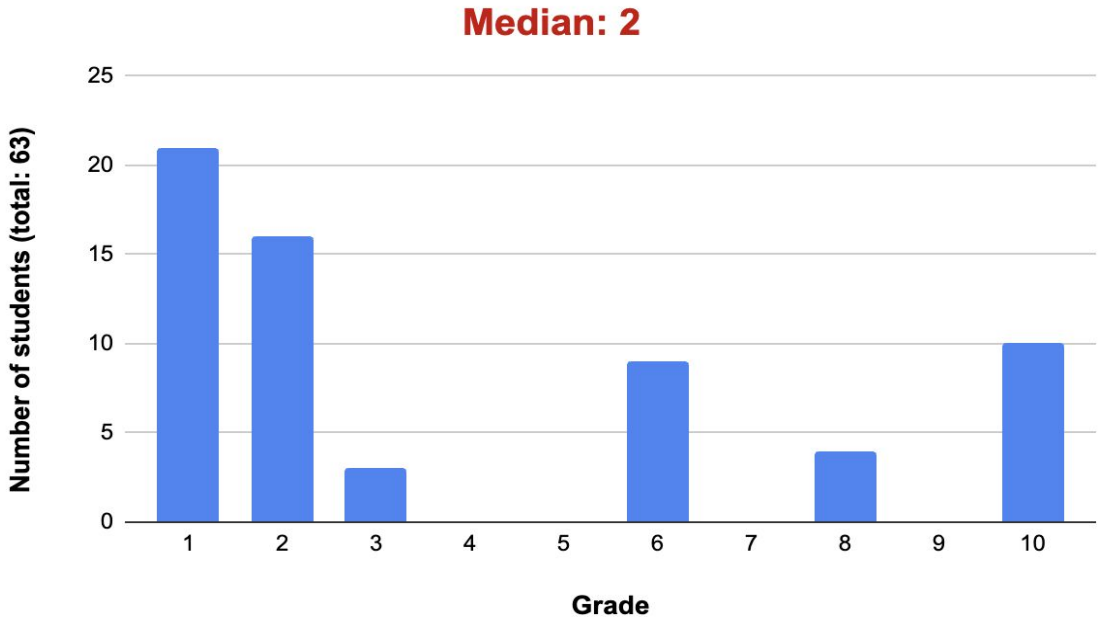
Vaggelis Atlidakis

Lecture 03

References: Similar OS courses @Columbia, @Stanford, @UC San Diego, @Brown, @di (previous years); and textbooks: Operating Systems: Three Easy Pieces, Operating Systems: Principles and Practice, Operating System Concepts, Linux Kernel Development, Understanding the Linux Kernel

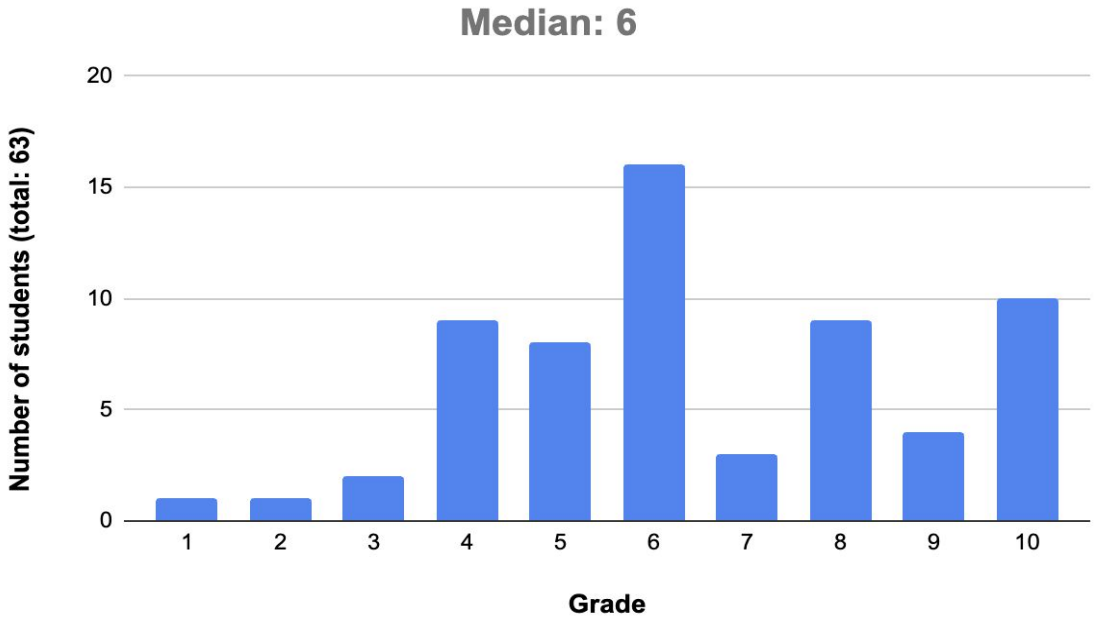
Quiz-01 (Disaster…)

- What was that **exactly**? Were you **here** last Monday?



Quiz-01 (Disaster…)

- Next time, be more conscious of the concept of negative points…

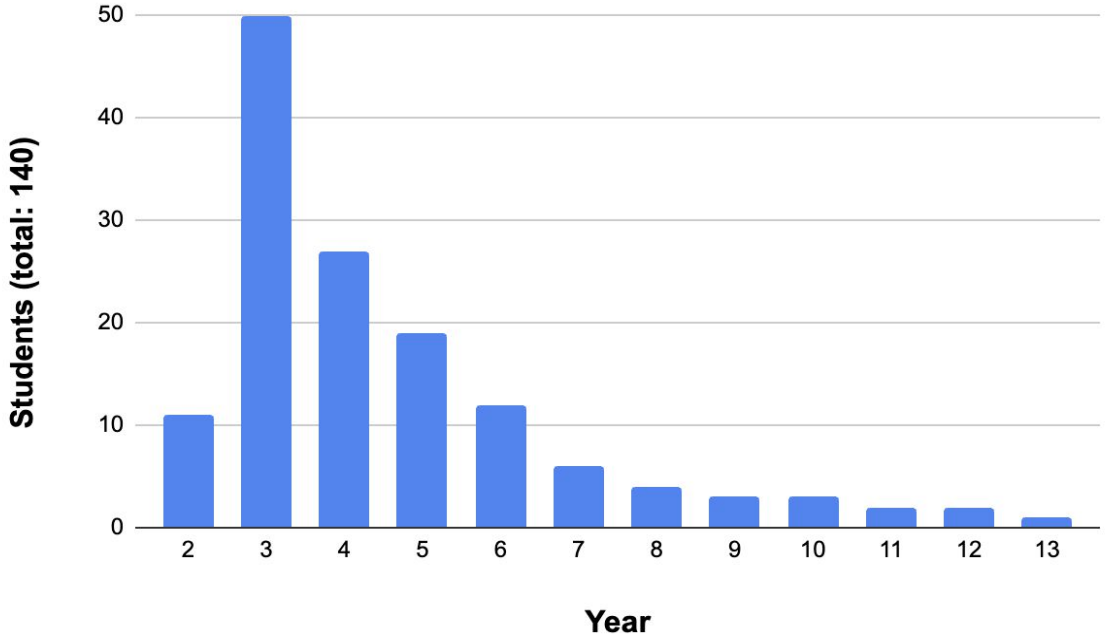


Quiz-01 (Disaster…)

- Come pick up your quiz
  - sdi1600006
  - sdi1900010
  - sdi2100058
  - sdi2200178
  - sdi2200182
  - sdi2300064
  - sdi2300080
  - sdi2300134
  - sdi2300202

Administrivia

- Where we are today? ~140 ppl.



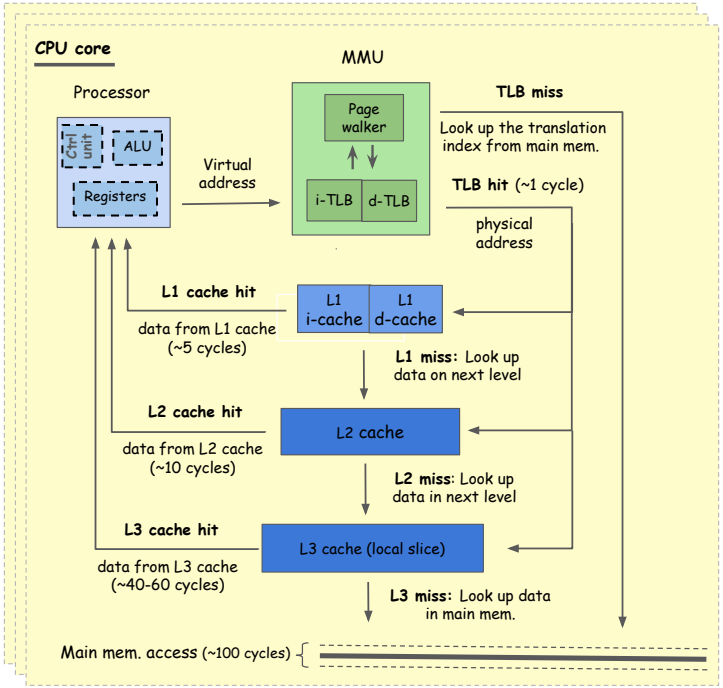
Administrivia

- Office hours
  - Every Thursday, 10.00-12.00 at A37
- Let us known of your **team of three**, using this form here
  - One submission per team, please!
  - The deadline is Sunday October 12, 23.59
  - This is a hard deadline
- Warm up instructions posted
- Anonymous feedback form

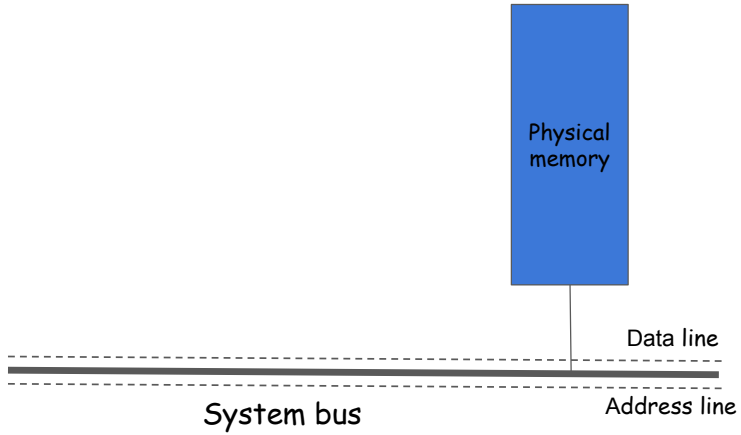
Summary of previous lecture(s)

- Introduction (10/10)
  - Q1: What is a OS?
  - Q2: Why do we need an OS?
  - Q3: Desirable properties for an OS?
  - Q4: Hardware model?
  - Q5: Design principles?
  - Q6: Hardware support to enforce design principles?
  - Q7: The basic hardware components so far?
  - Q8: How does the processor execute programs?
  - Q9: When does the processor accesses memory?
  - Q10: How does the system boot?

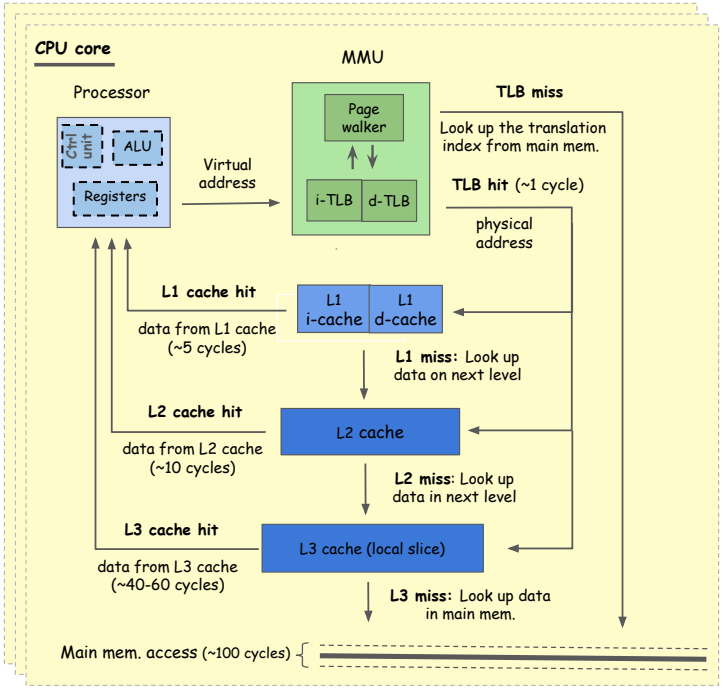
Handling I/O devices



- What are we missing in terms of hardware?

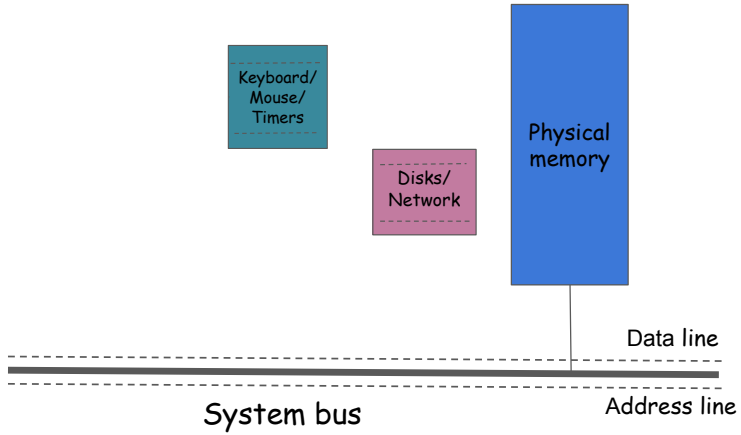


Handling I/O devices

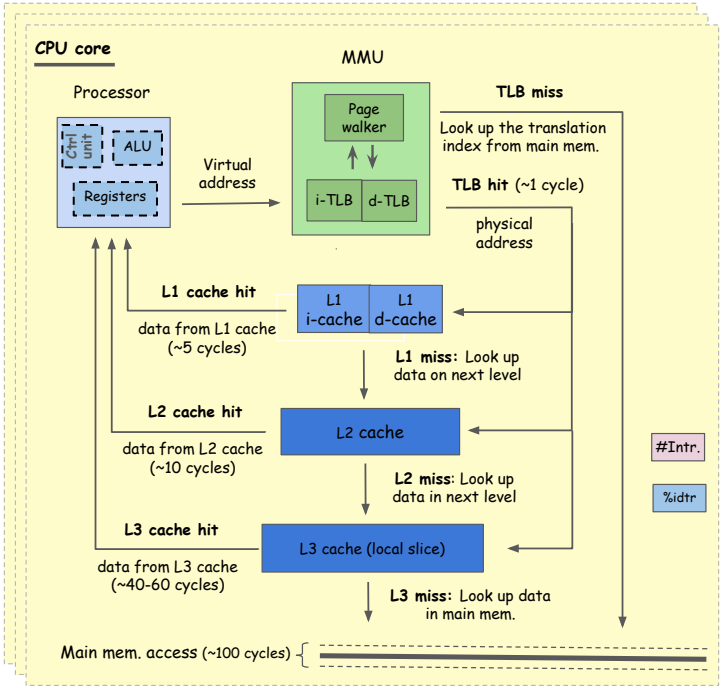


- What are we missing in terms of hardware?
  - Keyboard? Mouse?
  - Disk? Network?

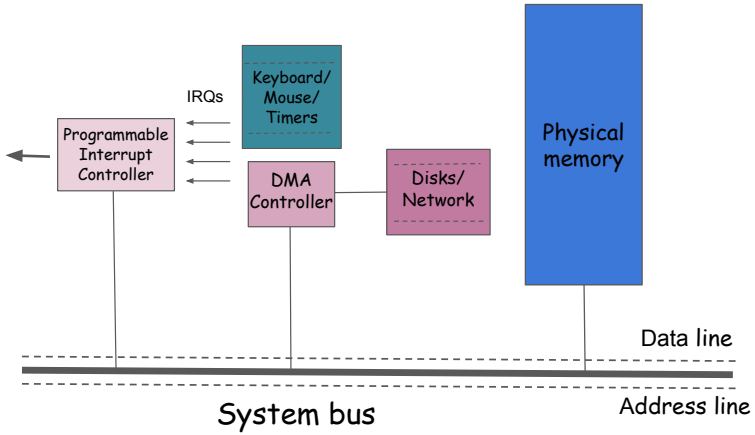
  } I/O devices



Handling I/O devices



- I/O devices
  - Need a way to get processor's attention
  - Interrupts, Interrupt Requests (IRQs), Direct Memory Access (DMA) controler...

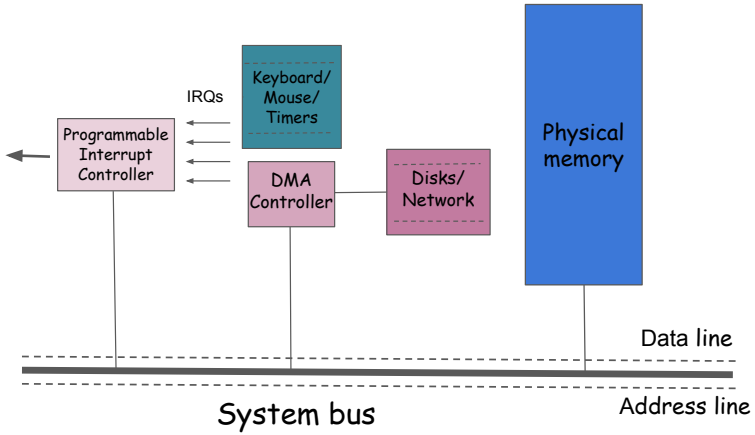


Handling I/O devices

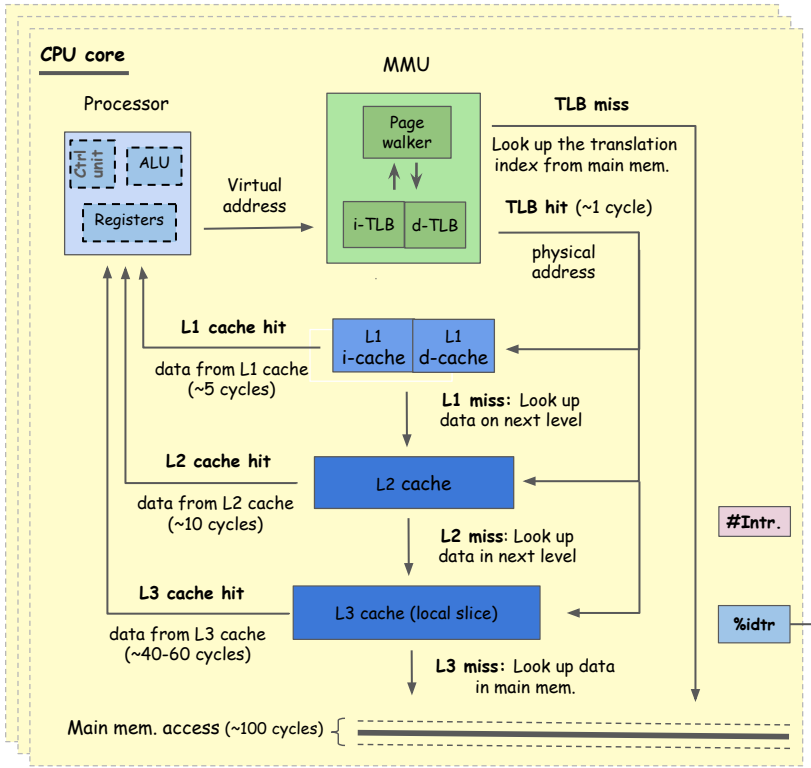
Execution pipeline

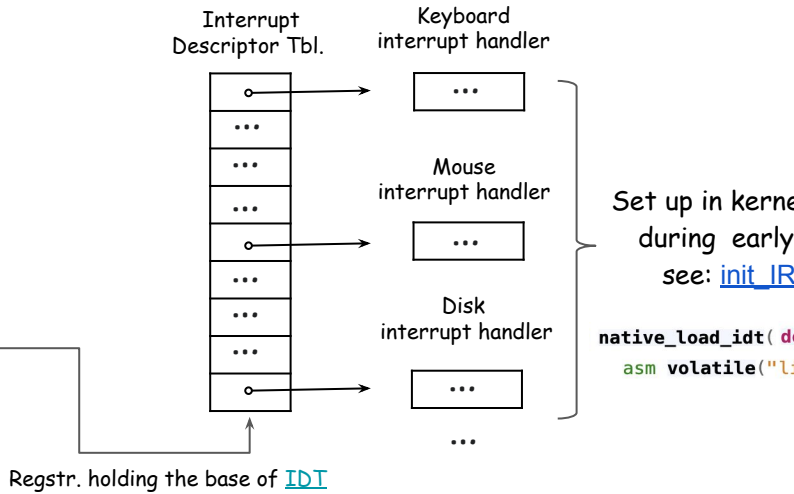
```
while {
    IF - Instruction Fetch
    ID - Instruction Decode
    EXE - Execute Instruction
    MEM - Memory Access
    WB - Write back
    ip → ip + 1                          Instruction
    if (Interrupt) {                     boundary
        What??
    }
}
```

- I/O devices
  - Need a way to get processor's attention
  - Interrupts, Interrupt Requests (IRQs), Direct Memory Access (DMA) controler…



Interrupt handling

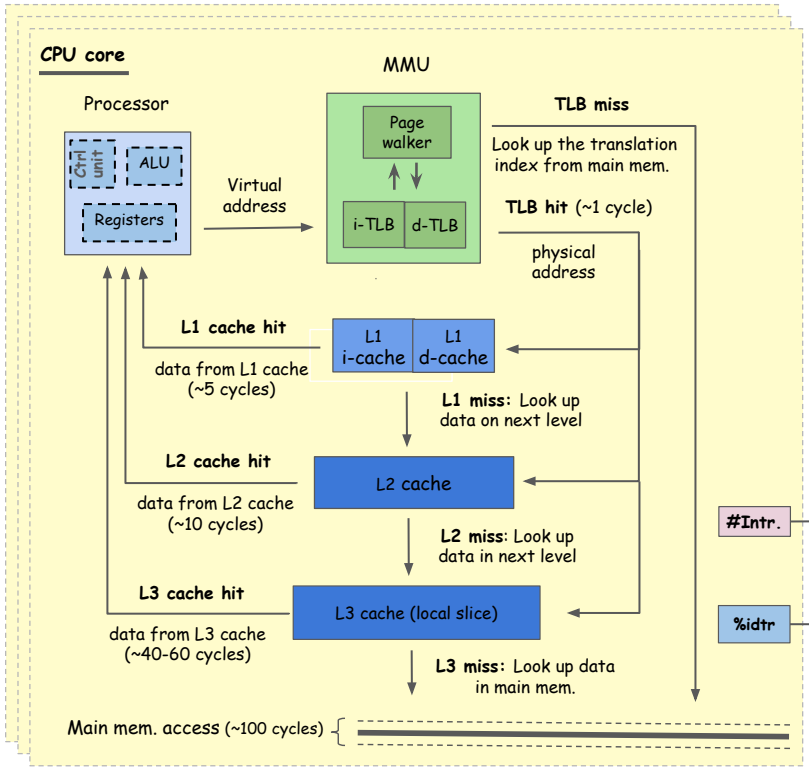


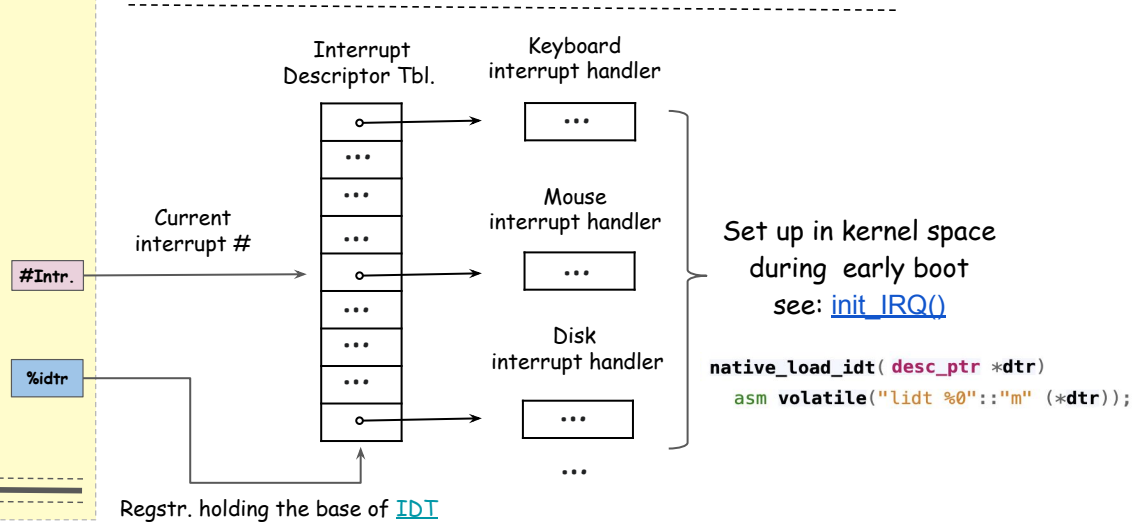


Set up in kernel space during early boot see: init_IRQ()

```
native_load_idt( desc_ptr *dtr)
  asm volatile("lidt %0"::"m" (*dtr));
```

Interrupt handling

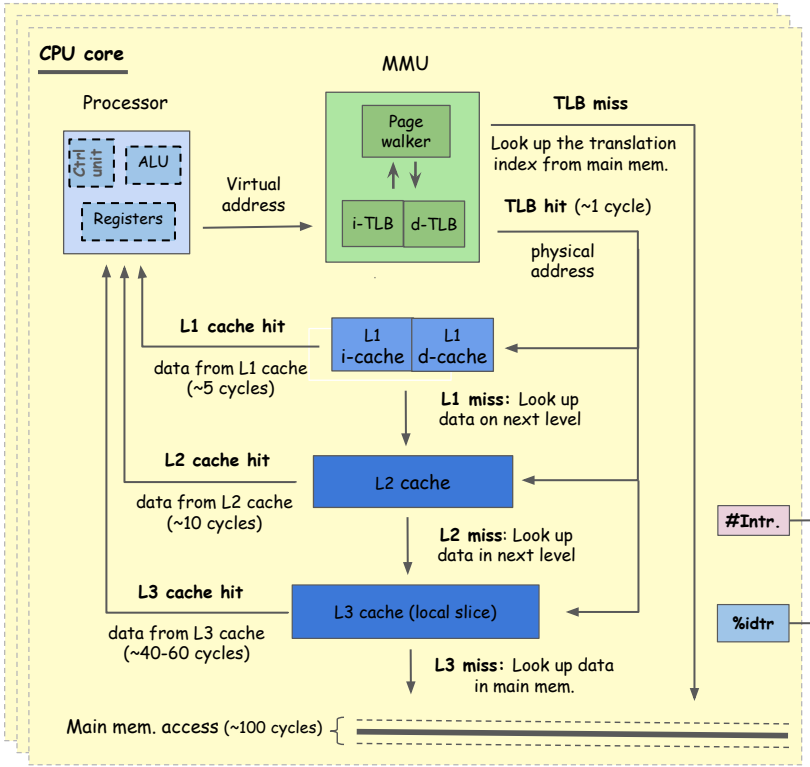




Set up in kernel space during early boot
see: init_IRQ()

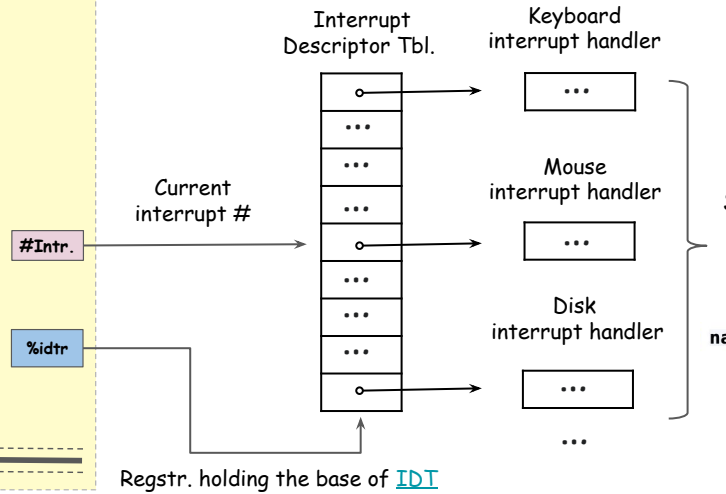
```
native_load_idt( desc_ptr *dtr)
  asm volatile("lidt %0"::"m" (*dtr));
```

Interrupt handling



do_handle_interrupt()

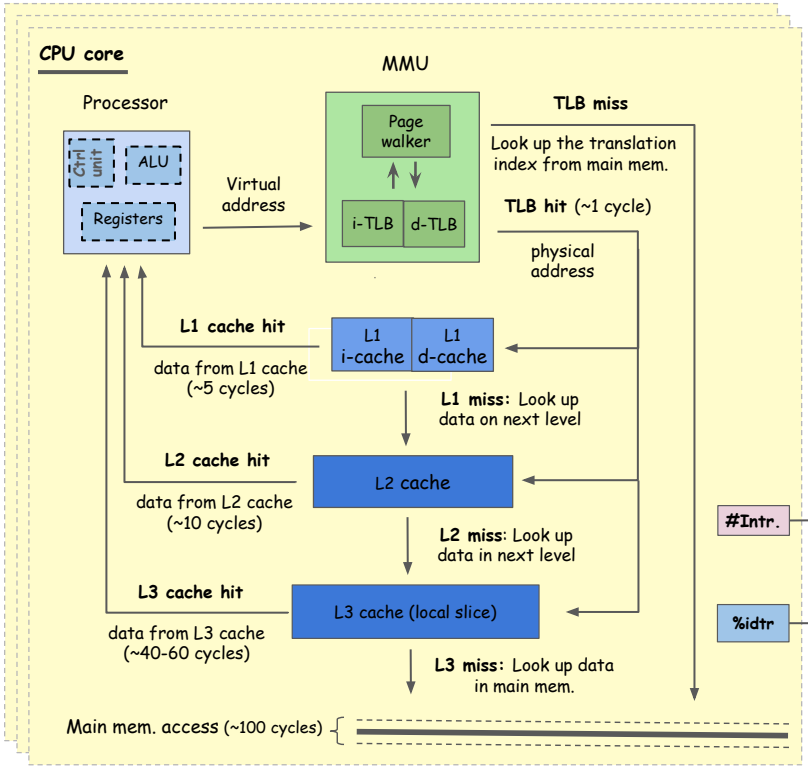
- Save CPU registers and interrupt error code (if any)
- Use #intr. and %idtr to execute the appropriate handler
- Restore CPU context



Set up in kernel space during early boot see: init_IRQ()

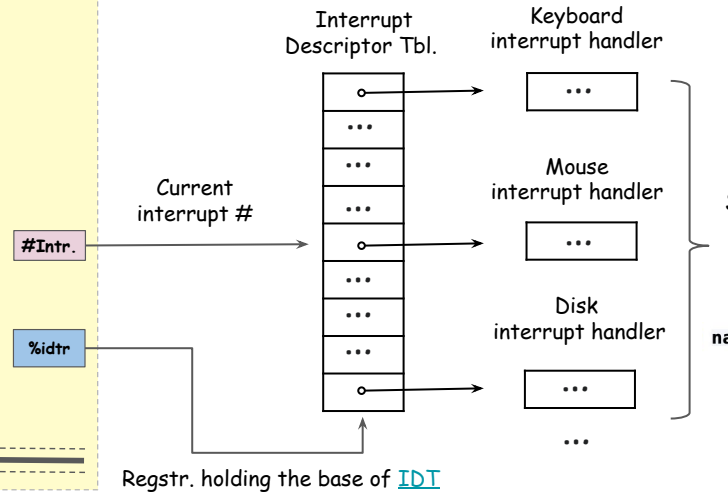
```
native_load_idt( desc_ptr *dtr)
  asm volatile("lidt %0"::"m" (*dtr));
```

Interrupt handling



do_handle_interrupt()

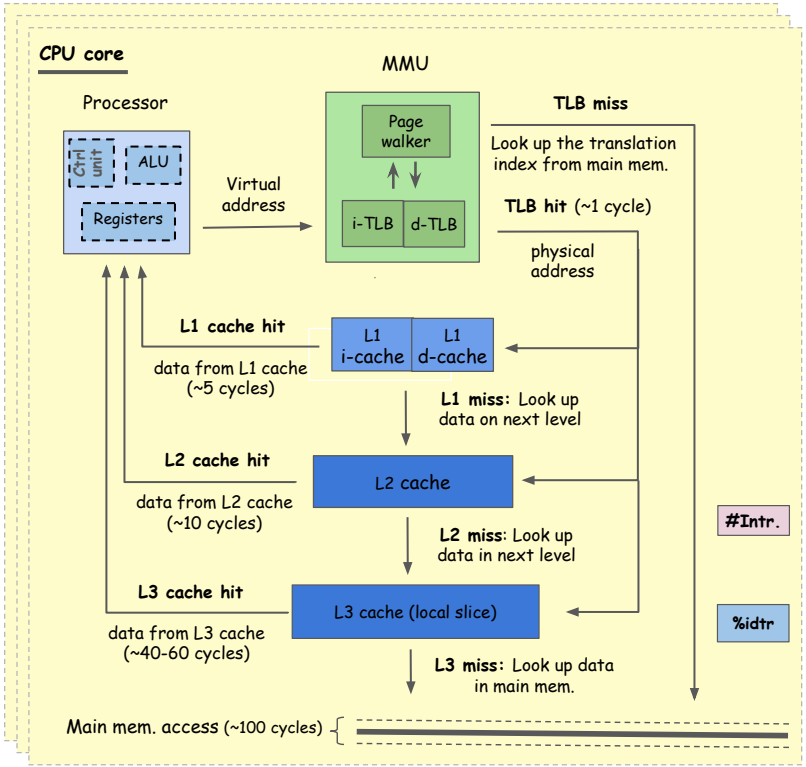
- **Switch to kernel stack, if in user mode**
- Save CPU registers and interrupt error code (if any)
- Use #intr. and %idtr to execute the appropriate handler
- Restore CPU context



Set up in kernel space during early boot see: init_IRQ()

```
native_load_idt( desc_ptr *dtr)
  asm volatile("lidt %0"::"m" (*dtr));
```

Interrupt handling (revised for clarity)



do_handle_interrupt()

- Switch stack
  - Kernel stack, if in user mode
  - Dedicated stack for critical exceptions
- Save CPU registers
- Save interrupt error code (if any)
- Use #intr. and %idtr to execute the appropriate interrupt handler
- Restore CPU context

Overview of what's coming…

- **Introduction (10/10)**
- **Events**
  - Q1: When does the OS run?
  - Q2: What asynchronous events invoke the OS?
  - Q3: What synchronous events invoke the OS?
  - Q4: What is POSIX?
  - Q5: Why do we need POSIX?
  - Q6: What is Linux?

When does the OS run?

The OS is a giant handler of events, which runs in response on two types of events

When does the OS run?

The OS is a giant handler of events, which runs in response on two types of events

- ***Asynchronous events***: Events that occur due to reasons external to the program instructions that the processor was currently executing ⇒ Interrupts
- ***Synchronous events***: Events that occur synchronously as a result of the execution of a program's instructions

What synchronous events invoke the OS? (revised for clarity)

Exceptions invoking the OS

- ***Faults***: Exceptions that allow the program to be restarted without loss of continuity
  - Does not break program continuity
  - Next instruction is the faulting instruction (restart)
  - The hope is that the OS will be able to "revert the mess" that caused it
  - Example: Page fault
- ***Aborts***: Exceptions used to report severe errors
  - Breaks program continuity
  - The processor may be unable report the precise instruction that caused it
  - Example: Double fault (invalid mem. access in a fault handler), machine check error
- ***Traps***: Exceptions reported immediately after the execution of a trapping instruction
  - Does not break program continuity
  - Next instruction is the one following the trapping instruction
  - The old way of making system calls!

X86 exceptions vector (addition)

Vector NR	Exception/Interrupt Name	Type
0	Divide Error	Fault
1	Debug Exception	Fault/Trap
2	Non-Maskable Interrupt (NMI)	Interrupt
3	Breakpoint Exception	Trap
4	Overflow Exception	Trap
5	Bound Range Exceeded	Fault
6	Invalid Opcode	Fault
7	Device Not Available (No Math Coprocessor)	Fault
8	Double Fault	Abort
9	Coprocessor Segment Overrun (Legacy)	Fault
10	Invalid TSS	Fault
11	Segment Not Present	Fault
12	Stack-Segment Fault	Fault
13	General Protection Fault (GPF)	Fault
14	Page Fault	Fault
15	Reserved	-
...		
20	Virtualization Exception	Fault
21-31	Reserved	-
32-255	User-Defined Interrupts	Interrupt

- From Intel's Software Developer's Manual, p. 3268

- Check also Tab.-7-4/5: Conditions for a Double Fault

- And see kernel/idt.c, for where the page fault exception handler is one of the first to be added

What synchronous events invoke the OS? (addition)

Exceptions invoking the OS

- ***Faults***: Exceptions that allow the program to be restarted without loss of continuity
- ***Aborts***: Exceptions used to report severe errors
- ***Traps***: Exceptions reported immediately after the execution of a trapping instruction

System calls: The interface between user programs and the OS

- The defacto mechanism for user programs to request services from the OS
- Linux x86_64 defines ~500 syscalls and aarch64 ~460:

x86/entry/syscalls/syscall_64.tbl

/ arch / x86 / entry / syscalls / syscall_64.tbl

```
60      common  exit            sys_exit
61      common  wait4           sys_wait4
62      common  kill            sys_kill
        ...
```

arm64/tools/syscall_64.tbl

/ arch / arm64 / tools

```
93  common  exit            sys_exit
94  common  exit_group      sys_exit_group
95  common  waitid          sys_waitid
        ...
```

Syscall argument passing (aarch64 example) (addition)

DESCRIPTION

_exit(int *status*): Terminates the calling process…

The value *status & 0xFF* is returned … as the exit status…

```
SYSCALL_DEFINE1(exit, int, error_code)
{
        // keep only the last byte, and shift it one byte left
        do_exit((error_code&0xff)<<8);
}

int main(void) {
   __asm__ __volatile__ ("mov x0, #123");   // Syscall arg0 to reg. %x0, and so on
   __asm__ __volatile__ ("mov x8, #93");    // Syscall NR to reg. %x8 (NR 93 is exit)
   __asm__ __volatile__ ("svc  #0");        // Execute the syscall
}
```

```
$ ./svc_test

$ echo $?

123
```

Syscall argument passing (x86_64 example) (addition)

DESCRIPTION

_exit(int *status*): Terminates the calling process…

The value *status & 0xFF* is returned … as the exit status…

```
SYSCALL_DEFINE1(exit, int, error_code)
{
        // keep only the last byte, and shift it one byte left
        do_exit((error_code&0xff)<<8);
}

int main(void) {
   __asm__ __volatile__ ("mov $123, %rdi");  // Syscall arg0 to reg. %rdi, and so on
   __asm__ __volatile__ ("mov $60, %rax");   // Syscall NR to reg. %rax (NR 60 is exit)
   __asm__ __volatile__ ("syscall");         // Execute the syscall
}
```

```
$ ./syscall_test

$ echo $?

123
```

How do user programs know what services an OS offers?

- We need a standard for that
  - Portable Operating System Interface for UNIX (**POSIX**) is the IEEE standard for portable UNIX-based OS syscall interfaces

How do user programs know what services an OS offers?

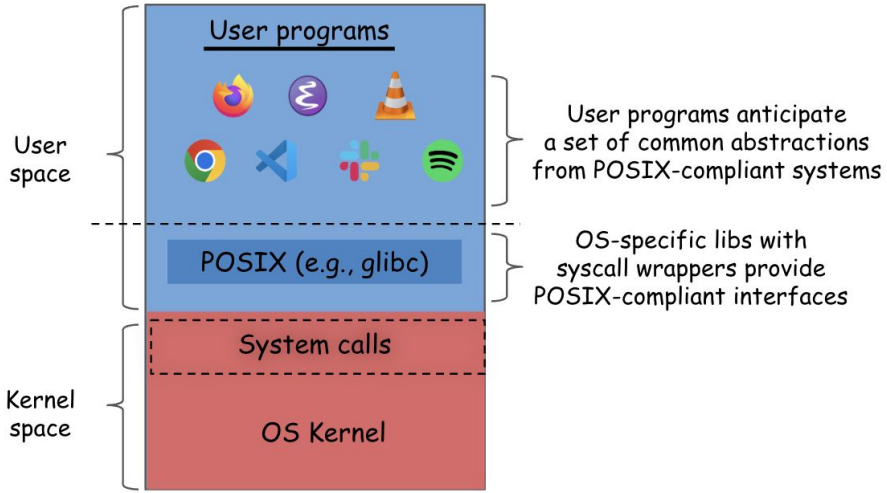
- We need a standard for that
  - Portable Operating System Interface for UNIX (**POSIX**) is the IEEE standard for portable UNIX-based OS syscall interfaces
  - ***Describes a set of fundamental abstractions needed for portable application development***

How do user programs know what services an OS offers?

- We need a standard for that
  - Portable Operating System Interface for UNIX (**POSIX**) is the IEEE standard for portable UNIX-based OS syscall interfaces
  - ***Describes a set of fundamental abstractions needed for portable application development***
  - User programs can directly invoke system calls, but going through POSIX ensures portability across POSIX-compliant OSes.
    - User programs must not break on kernel updates
    - OK to recompile code, if I move to another OS
    - But…don't make me rewrite it

How do user programs know what services an OS offers?

- We need a standard for that
  - Portable Operating System Interface for UNIX (**POSIX**) is the IEEE standard for portable UNIX-based OS syscall interfaces



What does POSIX define?

POSIX defines ~1,200 interfaces, around the following abstractions

What does POSIX define?

POSIX defines ~1,200 interfaces, around the following abstractions

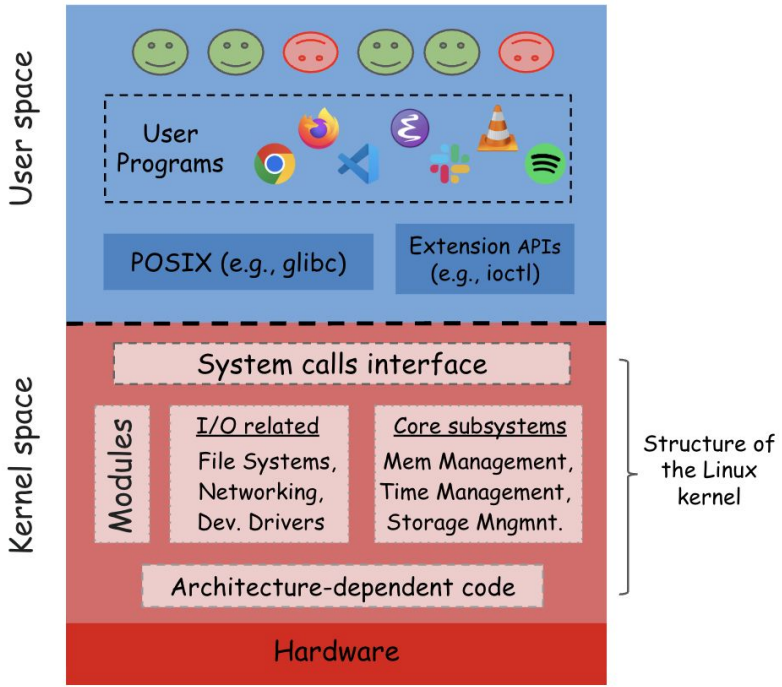
- ***Processes***: A process is an address space with one or more threads executing within that and the required system resources for those threads
- ***Threads***: A thread is a single flow of control within a process, with its own system resources required to support a flow of control
- ***Files***: A file is an object that can be written to, read from, or both

What does POSIX define?

POSIX defines ~1,200 interfaces, around the following abstractions

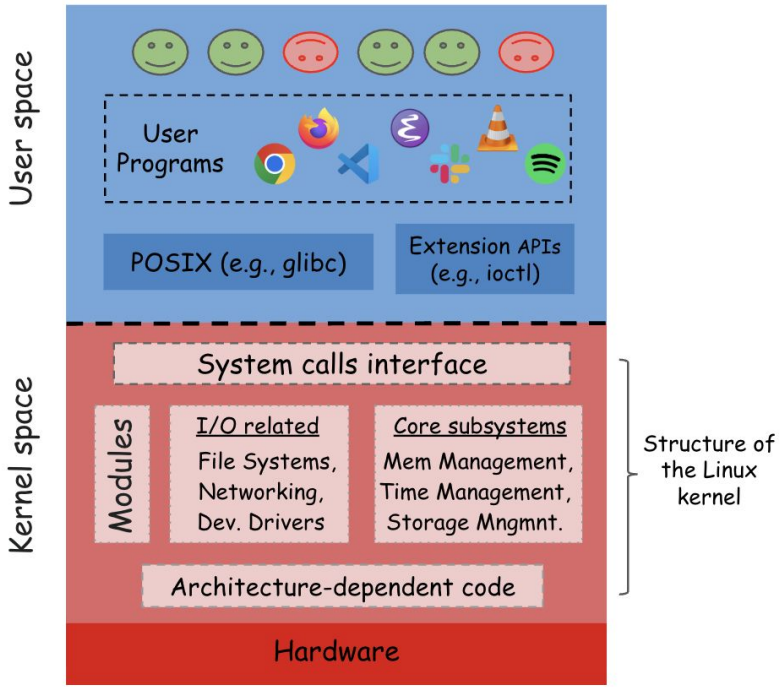
- ***Processes***: A process is an address space with one or more threads executing within that and the required system resources for those threads
- ***Threads***: A thread is a single flow of control within a process, with its own system resources required to support a flow of control
- ***Files***: A file is an object that can be written to, read from, or both
- The complete spec with all definitions is here

What is Linux?



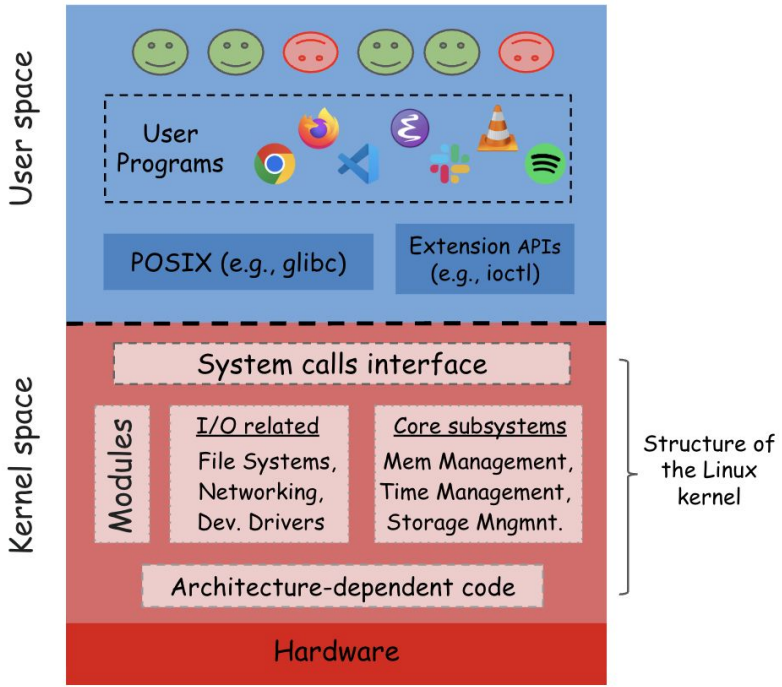
- A modern, open-source, POSIX-compliant OS kernel.

What is Linux?



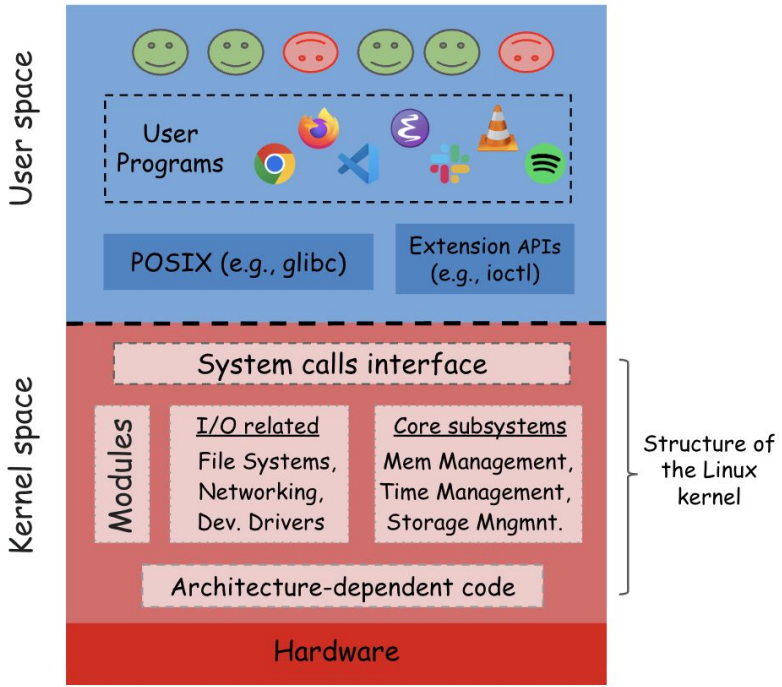
- A modern, open-source, POSIX-compliant OS kernel.
- **Written in '91 by Linus Torvalds from scratch ~100 KLoC (Jan '25: > 40 MLoC).**

What is Linux?



- A modern, open-source, POSIX-compliant OS kernel.
- **Written in '91 by Linus Torvalds from scratch ~100 KLoC (Jan '25: > 40 MLoC).**
- ***The most popular Unix-like OS today***
  - Servers: Over 96% of the world's top 1 million websites run on Linux servers
  - Cloud: AWS, Google Cloud, and Azure rely heavily on Linux
  - Supercomputers: Almost 100% of the top 500 supercomputers run Linux

What is Linux?



- A modern, open-source, POSIX-compliant OS kernel.
- **Written in '91 by Linus Torvalds from scratch ~100 KLoC (Jan '25: > 40 MLoC).**
- ***The most popular Unix-like OS today***
  - Servers: Over 96% of the world's top 1 million websites run on Linux servers.
  - Cloud: AWS, Google Cloud, and Azure rely heavily on Linux
  - Supercomputers: Almost 100% of the top 500 supercomputers run Linux
  - Android devices: >3 billion devices

Recap

- Introduction (10/10)
- Events
  - Q1: When does the OS run?
  - Q2: What asynchronous events invoke the OS?
  - Q3: What synchronous events invoke the OS?
  - Q4: What is POSIX?
  - Q5: Why do we need POSIX?
  - Q6: What is Linux?

Overview

- We'll start from hardware and follow a question-oriented approach
  - ~~Intro [Q: What is an OS?]~~
  - ~~Events [Q: When does the OS run?]~~
  - Runtime [Q: How does a program look like in memory?]
  - Processes [Q: What is a process?]
  - IPC [Q: How do processes communicate?]
  - Threads [Q: What is a thread?]
  - Synchronization [Q: What goes wrong w/o synchronization?]
  - Time Management [Q: What is scheduling?]
  - Memory Management [Q: What is virtual memory?]
  - Files [Q: What is a file descriptor?]
  - Storage Management [Q: How do we allocate disk space to files?]