

optimization (finish) / exceptions

last time

removing redundant operations

example: compiler often can't tell if function call produces same result each time

vector/SIMD instructions

extra wide registers can hold small (fixed-size) array of values

instructions that perform operation (e.g. add) on several pairs of values

typical implementation: extra wide ALU

typical implementation: extra wide data cache accesses

vector intrinsics

vector exercise (2)

```
long A[1024], B[1024];  
...  
for (int i = 0; i < 1024; i += 1)  
    for (int j = 0; j < 1024; j += 1)  
        A[i] += B[i] * B[j];
```

(casts omitted below to reduce clutter:)

```
for (int i = 0; i < 1024; i += 4) {  
    A_part = _mm256_loadu_si256(&A[i]);  
    Bi_part = _mm256_loadu_si256(&B[i]);  
    for (int j = 0; j < 1024; /* BLANK 1 */) {  
        Bj_part = _mm256_/* BLANK 2 */;  
        A_part = _mm256_add_epi64(A_part,  
                                   _mm256_mullo_epi64(Bi_part, Bj_part));  
    }  
    _mm256_storeu_si256(&A[i], A_part);  
}
```

What goes in BLANK 1 and BLANK 2?

- A. `j += 1, loadu_si256(&B[j])` B. `j += 4, loadu_si256(&B[j])`
C. `j += 1, set1_epi64(B[j])` D. `j += 4, set1_epi64(B[j])`

vector exercise 2 explanation

```
for (int i = 0; i < 1024; i += 1)
    for (int j = 0; j < 1024; j += 1)
        A[i] += B[i] * B[j];
/* -- transformed into -- */
for (int i = 0; i < 1024; i += 4)
    for (int j = 0; j < 1024; j += 1) {
        A[i+0] += B[i+0] * B[j];
        A[i+1] += B[i+1] * B[j];
        A[i+2] += B[i+2] * B[j];
        A[i+3] += B[i+3] * B[j];
    }
```

```
/* not the much harder to vectorize: */
for (int i = 0; i < 1024; i += 1)
    for (int j = 0; j < 1024; j += 4) {
        A[i] += B[i] * B[j+0];
        A[i] += B[i] * B[j+1];
        A[i] += B[i] * B[j+2];
        A[i] += B[i] * B[j+3];
    }
```

other vector instructions features

more flexible vector instruction features:

- invented in the 1990s

- often present in GPUs and being rediscovered by modern ISAs

reasonable conditional handling

better variable-length vectors

ability to load/store non-contiguous values

some of these features in AVX2/AVX512

alternate vector interfaces

intrinsics functions/assembly aren't the only way to write vector code

e.g. GCC vector extensions: more like normal C code

- types for each kind of vector

- write + instead of `_mm_add_epi32`

e.g. CUDA (GPUs): looks like writing multithreaded code, but each thread is vector “lane”

optimizing real programs

ask your compiler to try first

spend effort where **it matters**

e.g. 90% of program time spent reading files, but optimize computation?

e.g. 90% of program time spent in routine A, but optimize B?

profilers

first step — tool to determine where you spend time

tools exist to do this for programs

example on Linux: `perf`

example

Samples: 37K of event 'cycles', Event count (approx.): 3736755513				
Children	Self	Command	Shared Object	Symbol
+ 100.00%	0.00%	hclrs-with-debu	hclrs-with-debuginfo	[.] _start
+ 100.00%	0.00%	hclrs-with-debu	libc-2.31.so	[.] __libc_start_main
+ 100.00%	0.00%	hclrs-with-debu	hclrs-with-debuginfo	[.] main
+ 100.00%	0.00%	hclrs-with-debu	hclrs-with-debuginfo	[.] std::sys_common::backtrace::__rust_begin_short_backt
+ 100.00%	0.00%	hclrs-with-debu	hclrs-with-debuginfo	[.] hclrs::main
+ 99.99%	9.75%	hclrs-with-debu	hclrs-with-debuginfo	[.] hclrs::program::RunningProgram::run
+ 60.37%	31.67%	hclrs-with-debu	hclrs-with-debuginfo	[.] hclrs::ast::SpannedExpr::evaluate
+ 41.34%	23.29%	hclrs-with-debu	hclrs-with-debuginfo	[.] hashbrown::map::make_hash
+ 18.08%	18.07%	hclrs-with-debu	hclrs-with-debuginfo	[.] <std::collections::hash::map::DefaultHasher as core:
+ 16.33%	0.68%	hclrs-with-debu	hclrs-with-debuginfo	[.] hclrs::program::Program::process_register_banks
+ 9.54%	3.15%	hclrs-with-debu	hclrs-with-debuginfo	[.] std::collections::hash::map::HashMap<K,V,S>::get
+ 9.10%	9.09%	hclrs-with-debu	libc-2.31.so	[.] __memcmp_avx2_movbe
+ 6.11%	2.10%	hclrs-with-debu	hclrs-with-debuginfo	[.] hashbrown::map::HashMap<K,V,S>::get_mut
+ 2.32%	0.88%	hclrs-with-debu	hclrs-with-debuginfo	[.] std::collections::hash::map::HashMap<K,V,S>::get
+ 1.45%	0.52%	hclrs-with-debu	hclrs-with-debuginfo	[.] hashbrown::map::HashMap<K,V,S>::insert
0.37%	0.11%	hclrs-with-debu	hclrs-with-debuginfo	[.] <alloc::string::String as core::clone::Clone>::clone
0.19%	0.19%	hclrs-with-debu	libc-2.31.so	[.] malloc

an infinite loop

```
int main(void) {  
    while (1) {  
        /* waste CPU time */  
    }  
}
```

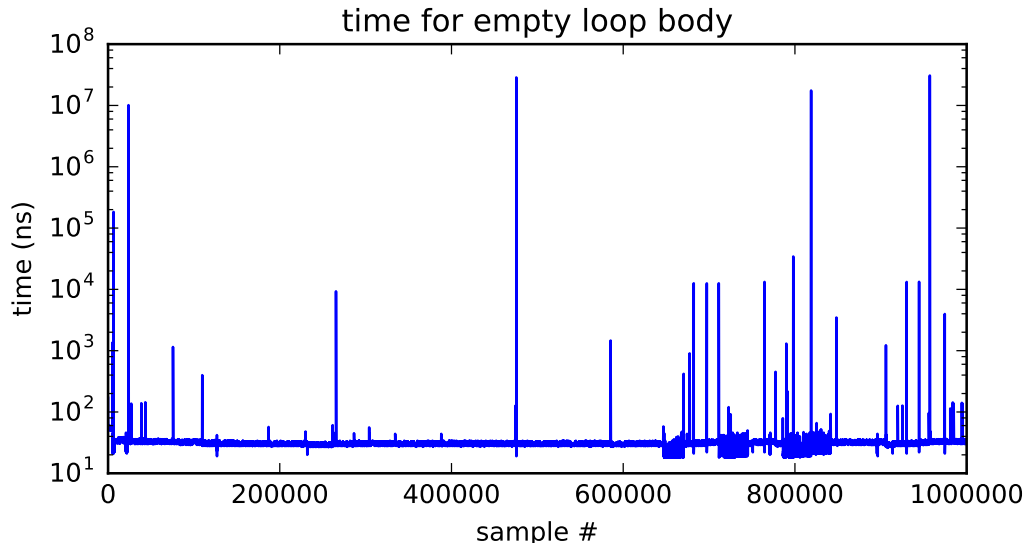
If I run this on a shared department machine, can you still use it?
...if the machine only has one core?

timing nothing

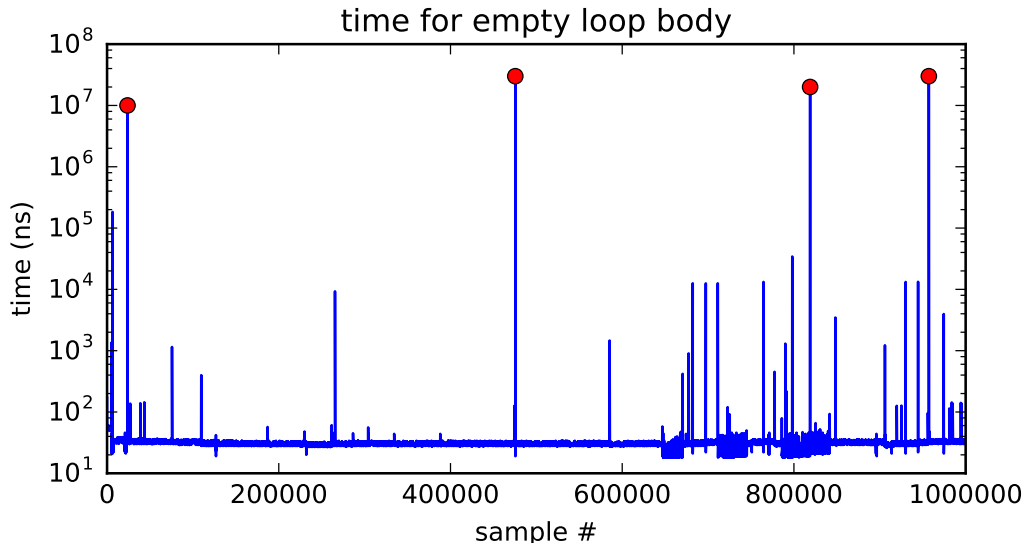
```
long times[NUM_TIMINGS];  
int main(void) {  
    for (int i = 0; i < N; ++i) {  
        long start, end;  
        start = get_time();  
        /* do nothing */  
        end = get_time();  
        times[i] = end - start;  
    }  
    output_timings(times);  
}
```

same instructions — same difference each time?

doing nothing on a busy system



doing nothing on a busy system



time multiplexing



time multiplexing



...

```
call get_time
```

```
// whatever get_time does
```

```
movq %rax, %rbp
```

———— million cycle delay ————

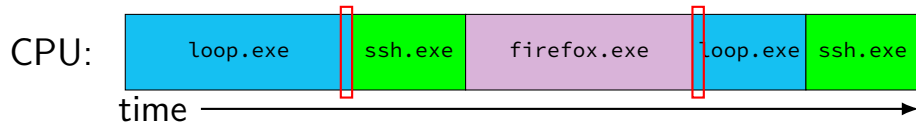
```
call get_time
```

```
// whatever get_time does
```

```
subq %rbp, %rax
```

...

time multiplexing



...

```
call get_time
```

```
// whatever get_time does
```

```
movq %rax, %rbp
```

———— million cycle delay ————

```
call get_time
```

```
// whatever get_time does
```

```
subq %rbp, %rax
```

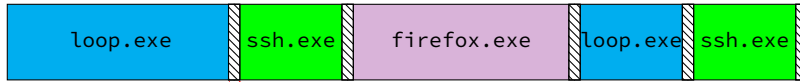
...

time multiplexing really



= operating system

time multiplexing really



= operating system

exception happens

return from exception

OS and time multiplexing

starts running instead of normal program

mechanism for this: **exceptions** (later)

saves old program counter, registers somewhere

sets new registers, jumps to new program counter

called **context switch**

saved information called **context**

context

all registers values

`%rax %rbx, ..., %rsp, ...`

condition codes

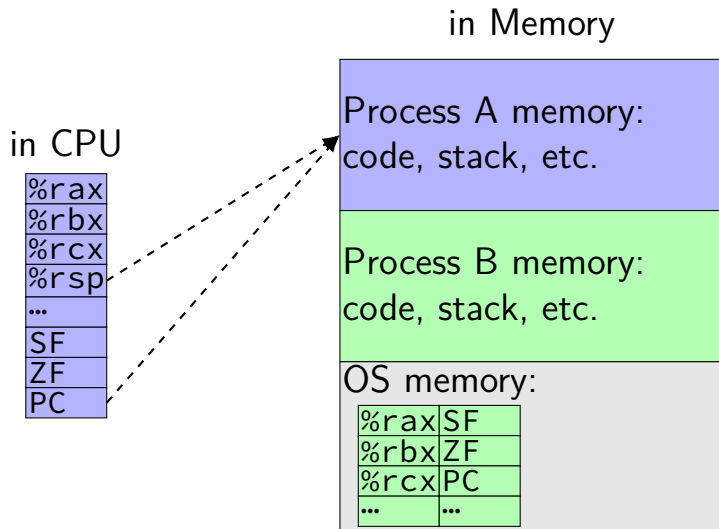
program counter

i.e. all visible state in your CPU except memory

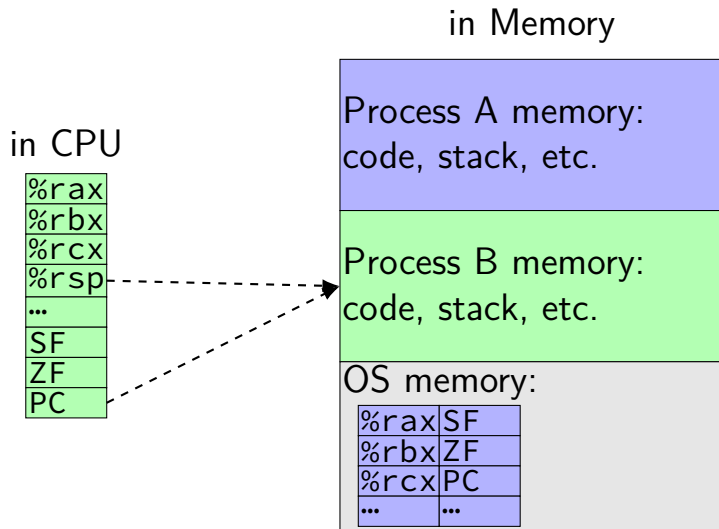
context switch pseudocode

```
context_switch(last, next):  
    copy_preexception_pc last->pc  
    mov rax, last->rax  
    mov rcx, last->rcx  
    mov rdx, last->rdx  
    ...  
    mov next->rdx, rdx  
    mov next->rcx, rcx  
    mov next->rax, rax  
    jmp next->pc
```

contexts (A running)



contexts (B running)



memory protection

reading from another program's memory?

Program A	Program B
<pre>0x10000: .word 42 // ... // do work // ... movq 0x10000, %rax</pre>	<pre><i>// while A is working:</i> movq \$99, %rax movq %rax, 0x10000 ...</pre>

memory protection

reading from another program's memory?

Program A	Program B
<pre>0x10000: .word 42 // ... // do work // ... movq 0x10000, %rax</pre>	<pre><i>// while A is working:</i> movq \$99, %rax movq %rax, 0x10000 ...</pre>

result: %rax is ...

- A. 42 B. 99 C. 0x10000
D. 42 or 99 (depending on timing/program layout/etc)
E. 42 or program might crash (depending on ...)
F. 99 or program might crash (depending on ...)
G. 42 or 99 or program might crash (depending on ...)
H. something else

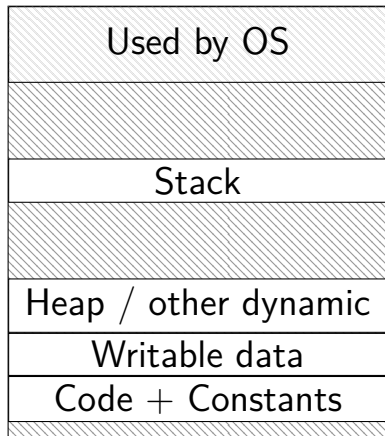
memory protection

reading from another program's memory?

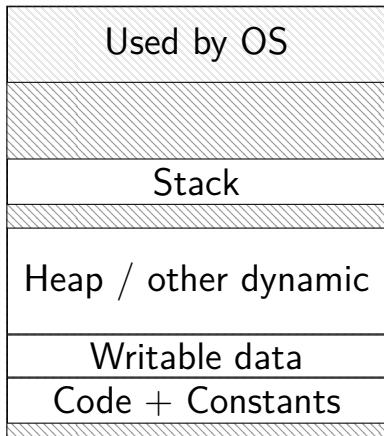
Program A	Program B
<pre>0x10000: .word 42 // ... // do work // ... movq 0x10000, %rax</pre>	<pre><i>// while A is working:</i> movq \$99, %rax movq %rax, 0x10000 ...</pre>
result: %rax is 42 (always)	result: <i>might crash</i>

program memory (two programs)

Program A



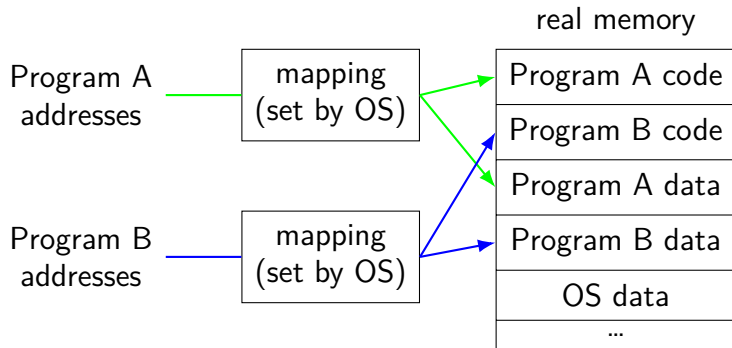
Program B



address space

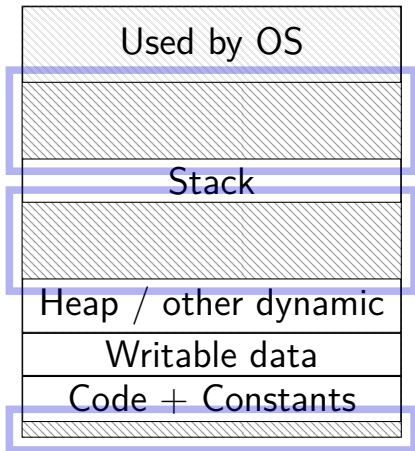
programs have **illusion of own memory**

called a program's **address space**

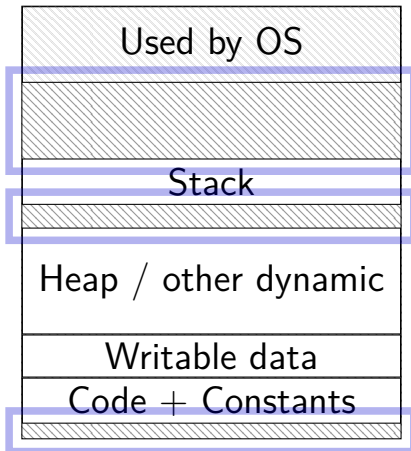


program memory (two programs)

Program A



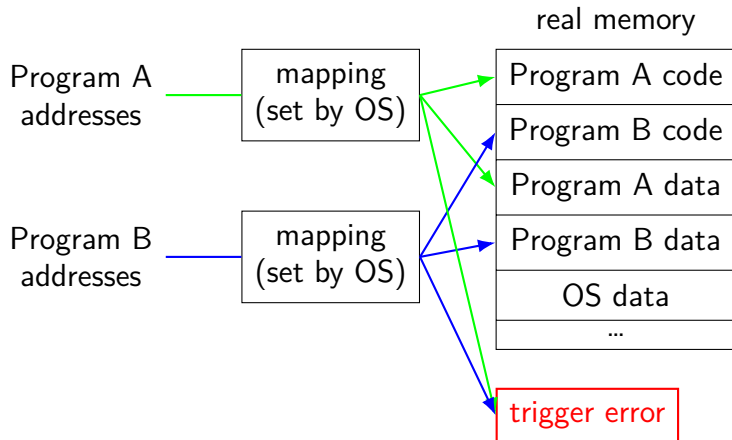
Program B



address space

programs have **illusion of own memory**

called a program's **address space**



address space mechanisms

topic after exceptions

called **virtual memory**

mapping called **page tables**

mapping part of what is changed in context switch

context

all registers values

`%rax %rbx, ..., %rsp, ...`

condition codes

program counter

~~i.e. all visible state in your CPU except memory~~

address space: map from program to real addresses

The Process

process = thread(s) + address space

illusion of **dedicated machine**:

thread = illusion of own CPU

address space = illusion of own memory

types of exceptions

interrupts — externally-triggered

timer — keep program from hogging CPU

I/O devices — key presses, hard drives, networks, ...

aborts — hardware is broken

traps — intentionally triggered exceptions

system calls — ask OS to do something

faults — errors/events in programs

memory not in address space (“Segmentation fault”)

privileged instruction

divide by zero

invalid instruction

asynchronous

not triggered by
running program

synchronous

triggered by
current program

types of exceptions

interrupts — externally-triggered

timer — keep program from hogging CPU

I/O devices — key presses, hard drives, networks, ...

aborts — hardware is broken

traps — intentionally triggered exceptions

system calls — ask OS to do something

faults — errors/events in programs

memory not in address space (“Segmentation fault”)

privileged instruction

divide by zero

invalid instruction

asynchronous

not triggered by
running program

synchronous

triggered by
current program

types of exceptions

interrupts — externally-triggered

timer — keep program from hogging CPU

I/O devices — key presses, hard drives, networks, ...

aborts — hardware is broken

traps — intentionally triggered exceptions

system calls — ask OS to do something

faults — errors/events in programs

memory not in address space (“Segmentation fault”)

privileged instruction

divide by zero

invalid instruction

asynchronous

not triggered by
running program

synchronous

triggered by
current program

timer interrupt

(conceptually) external timer device
(usually on same chip as processor)

OS configures before starting program

sends signal to CPU after a fixed interval

types of exceptions

interrupts — externally-triggered

timer — keep program from hogging CPU

I/O devices — key presses, hard drives, networks, ...

aborts — hardware is broken

traps — intentionally triggered exceptions

system calls — ask OS to do something

faults — errors/events in programs

memory not in address space (“Segmentation fault”)

privileged instruction

divide by zero

invalid instruction

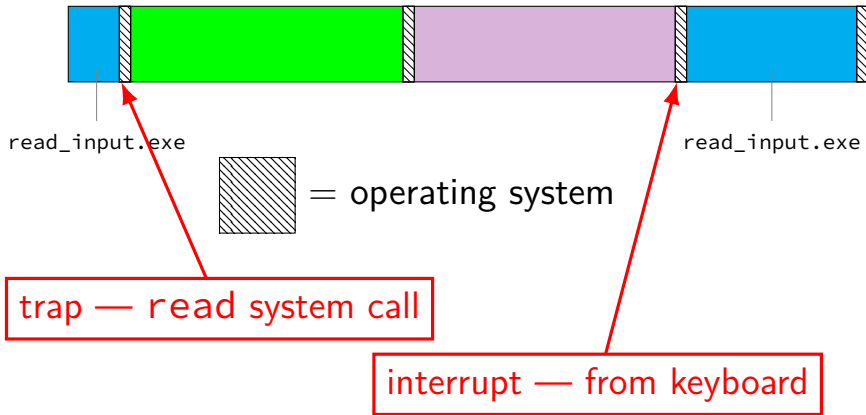
asynchronous

not triggered by
running program

synchronous

triggered by
current program

keyboard input timeline



types of exceptions

interrupts — externally-triggered

timer — keep program from hogging CPU

I/O devices — key presses, hard drives, networks, ...

aborts — hardware is broken

traps — intentionally triggered exceptions

system calls — ask OS to do something

faults — errors/events in programs

memory not in address space (“Segmentation fault”)

privileged instruction

divide by zero

invalid instruction

asynchronous

not triggered by
running program

synchronous

triggered by
current program

exception implementation

detect condition (program error or external event)

save current value of PC somewhere

jump to **exception handler** (part of OS)

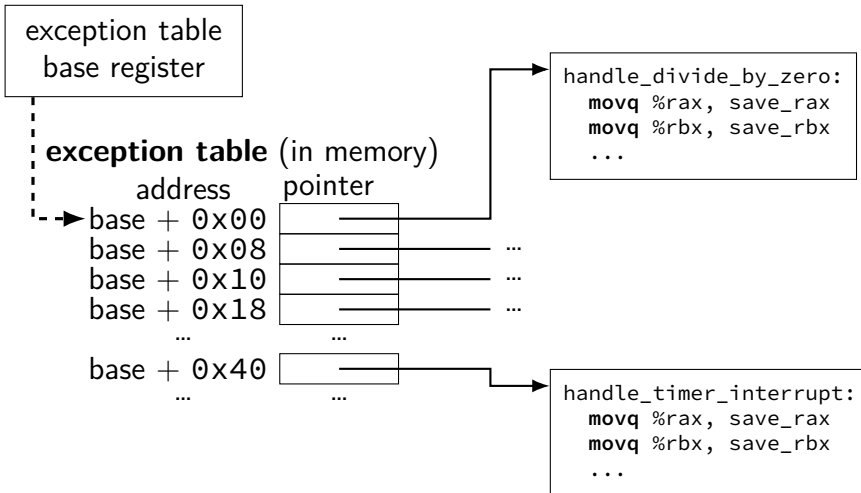
jump done without program instruction to do so

exception implementation: notes

I/textbook describe a **simplified** version

real x86/x86-64 is a bit more complicated
(mostly for historical reasons)

locating exception handlers



running the exception handler

hardware saves the **old program counter** (and maybe more)

identifies location of exception handler via table

then jumps to that location

OS code can save anything else it wants to , etc.

added to CPU for exceptions

new instruction: set exception table base

new logic: jump based on exception table

may need to cancel partially completed instructions before jumping

new logic: save the old PC (and maybe more)

to special register or to memory

new instruction: return from exception

i.e. jump to saved PC

added to CPU for exceptions

new instruction: set **exception table base**

new logic: **jump based on exception table**

may need to cancel partially completed instructions before jumping

new logic: save the old PC (and maybe more)

to special register or to memory

new instruction: return from exception

i.e. jump to saved PC

added to CPU for exceptions

new instruction: set exception table base

new logic: jump based on exception table

may need to cancel partially completed instructions before jumping

new logic: **save the old PC** (and maybe more)

to special register or to memory

new instruction: return from exception

i.e. jump to saved PC

added to CPU for exceptions

new instruction: set exception table base

new logic: jump based on exception table

may need to cancel partially completed instructions before jumping

new logic: save the old PC (and maybe more)

to special register or to memory

new instruction: **return from exception**

i.e. jump to saved PC

exception handler structure

1. save process's state somewhere
2. do work to handle exception
3. restore a process's state (maybe a different one)
4. jump back to program

handle_timer_interrupt:

`mov_from_saved_pc save_pc_loc`

`movq %rax, save_rax_loc`

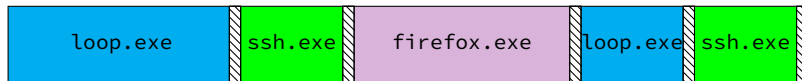
`... // choose new process to run here`

`movq new_rax_loc, %rax`

`mov_to_saved_pc new_pc`

`return_from_exception`

exceptions and time slicing



timer interrupt

exception table lookup

```
handle_timer_interrupt:
```

```
...
```

```
...
```

```
set_address_space ssh_address_space
```

```
mov_to_saved_pc saved_ssh_pc
```

```
return_from_exception
```

defeating time slices?

```
my_exception_table:
```

```
...
```

```
my_handle_timer_interrupt:
```

```
    // HA! Keep running me!
```

```
    return_from_exception
```

```
main:
```

```
    set_exception_table_base my_exception_table
```

```
loop:
```

```
    jmp loop
```

defeating time slices?

wrote a program that tries to set the exception table:

```
my_exception_table:
```

```
...
```

```
main:
```

```
    // "Load Interrupt
```

```
    //  Descriptor Table"
```

```
    // x86 instruction to set exception table
```

```
    lidt my_exception_table
```

```
    ret
```

result: **Segmentation fault** (exception!)

types of exceptions

interrupts — externally-triggered

timer — keep program from hogging CPU

I/O devices — key presses, hard drives, networks, ...

aborts — hardware is broken

traps — intentionally triggered exceptions

system calls — ask OS to do something

faults — errors/events in programs

memory not in address space (“Segmentation fault”)

privileged instruction

divide by zero

invalid instruction

asynchronous

not triggered by
running program

synchronous

triggered by
current program

privileged instructions

can't let **any program** run some instructions

allows machines to be shared between users (e.g. lab servers)

examples:

- set exception table

- set address space

- talk to I/O device (hard drive, keyboard, display, ...)

- ...

processor has two modes:

- kernel mode — privileged instructions work

- user mode — privileged instructions cause exception instead

kernel mode

extra one-bit register: “are we in kernel mode”

exceptions **enter kernel mode**

return from exception instruction **leaves kernel mode**

types of exceptions

interrupts — externally-triggered

timer — keep program from hogging CPU

I/O devices — key presses, hard drives, networks, ...

aborts — hardware is broken

traps — intentionally triggered exceptions

system calls — ask OS to do something

faults — errors/events in programs

memory not in address space (“Segmentation fault”)

privileged instruction

divide by zero

invalid instruction

asynchronous

not triggered by
running program

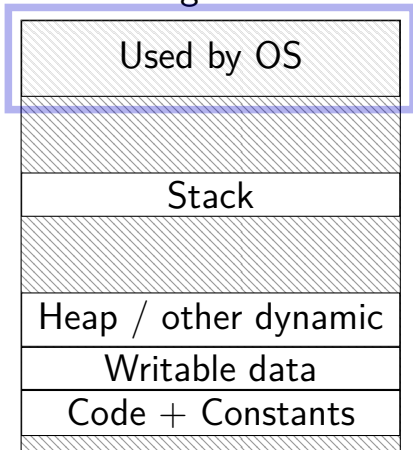
synchronous

triggered by
current program

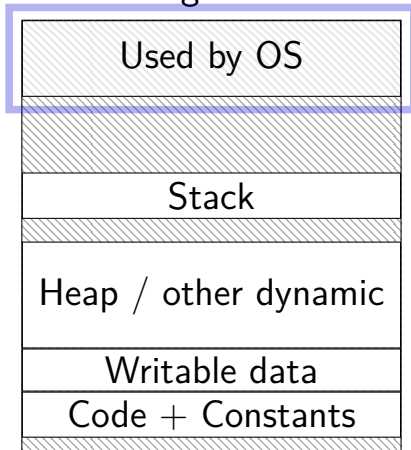
what about editing exception table?

program memory (two programs)

Program A



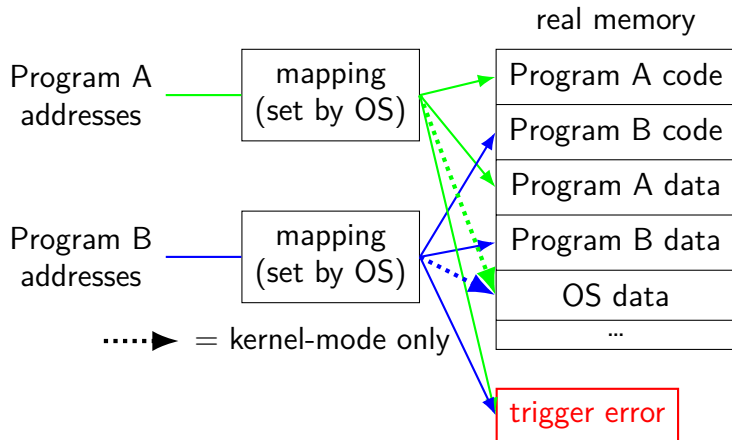
Program B



address space

programs have **illusion of own memory**

called a program's **address space**



protection fault

when program tries to access memory it doesn't own

e.g. trying to write to OS address

when program tries to do other things that are not allowed

e.g. accessing I/O devices directly

e.g. changing exception table base register

OS gets control — can crash the program
or more interesting things

types of exceptions

interrupts — externally-triggered

timer — keep program from hogging CPU

I/O devices — key presses, hard drives, networks, ...

aborts — hardware is broken

traps — intentionally triggered exceptions

system calls — ask OS to do something

faults — errors/events in programs

memory not in address space (“Segmentation fault”)

privileged instruction

divide by zero

invalid instruction

asynchronous

not triggered by
running program

synchronous

triggered by
current program

which requires kernel mode?

which operations are likely to fail (trigger an exception to run the OS instead) if attempted in user mode?

- A. reading data on disk by running special instructions that communicate with the hard disk device
- B. changing a program's address space to allocate it more memory
- C. returning from a standard library function
- D. incrementing the stack pointer

kernel services

allocating memory? (change address space)

reading/writing to file? (communicate with hard drive)

read input? (communicate with keyboard)

all need privileged instructions!

need to run code in kernel mode

Linux x86-64 system calls

special instruction: `syscall`

triggers **trap** (deliberate exception)

Linux syscall calling convention

before `syscall`:

`%rax` — system call number

`%rdi`, `%rsi`, `%rdx`, `%r10`, `%r8`, `%r9` — args

after `syscall`:

`%rax` — return value

on error: `%rax` contains -1 times “error number”

almost the same as normal function calls

Linux x86-64 hello world

```
.globl _start
.data
hello_str: .asciz "Hello, World!\n"
.text
_start:
    movq $1, %rax # 1 = "write"
    movq $1, %rdi # file descriptor 1 = stdout
    movq $hello_str, %rsi
    movq $15, %rdx # 15 = strlen("Hello, World!\n")
    syscall

    movq $60, %rax # 60 = exit
    movq $0, %rdi
    syscall
```

approx. system call handler

```
sys_call_table:
```

```
    .quad handle_read_syscall  
    .quad handle_write_syscall  
    // ...
```

```
handle_syscall:
```

```
    ... // save old PC, etc.  
    pushq %rcx // save registers  
    pushq %rdi  
    ...  
    call *sys_call_table(,%rax,8)  
    ...  
    popq %rdi  
    popq %rcx  
    return_from_exception
```

Linux system call examples

`mmap`, `brk` — allocate memory

`fork` — create new process

`execve` — run a program in the current process

`_exit` — terminate a process

`open`, `read`, `write` — access files
terminals, etc. count as files, too

system call wrappers

can't write C code to generate syscall instruction

solution: call “wrapper” function written in assembly

which of these require exceptions? context switches?

- A. program calls a function in the standard library
- B. program writes a file to disk
- C. program A goes to sleep, letting program B run
- D. program exits
- E. program returns from one function to another function
- F. program pops a value from the stack

a note on terminology (1)

real world: inconsistent terms for exceptions

we will follow textbook's terms in this course

the real world won't

you might see:

- 'interrupt' meaning what we call 'exception' (x86)

- 'exception' meaning what we call 'fault'

- 'hard fault' meaning what we call 'abort'

- 'trap' meaning what we call 'fault'

- ... and more

a note on terminology (2)

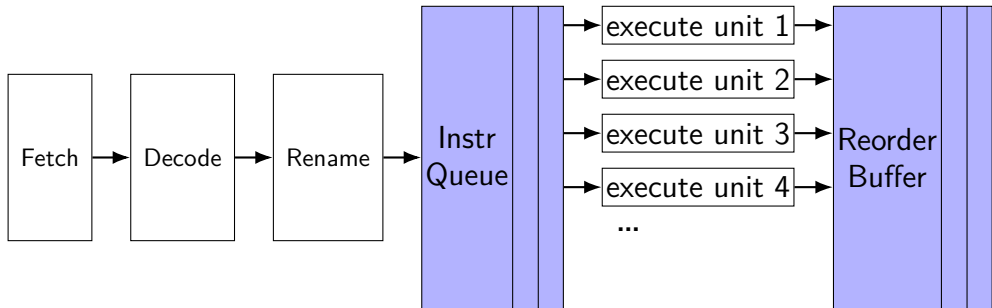
we use the term “kernel mode”

some additional terms:

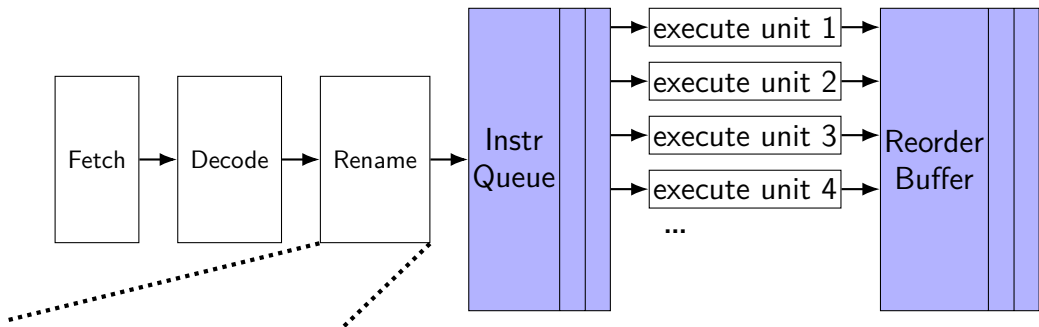
- supervisor mode
- privileged mode
- ring 0

some systems have **multiple levels** of privilege
different sets of privileged operations work

exceptions and OOO (one strategy)



exceptions and OOO (one strategy)

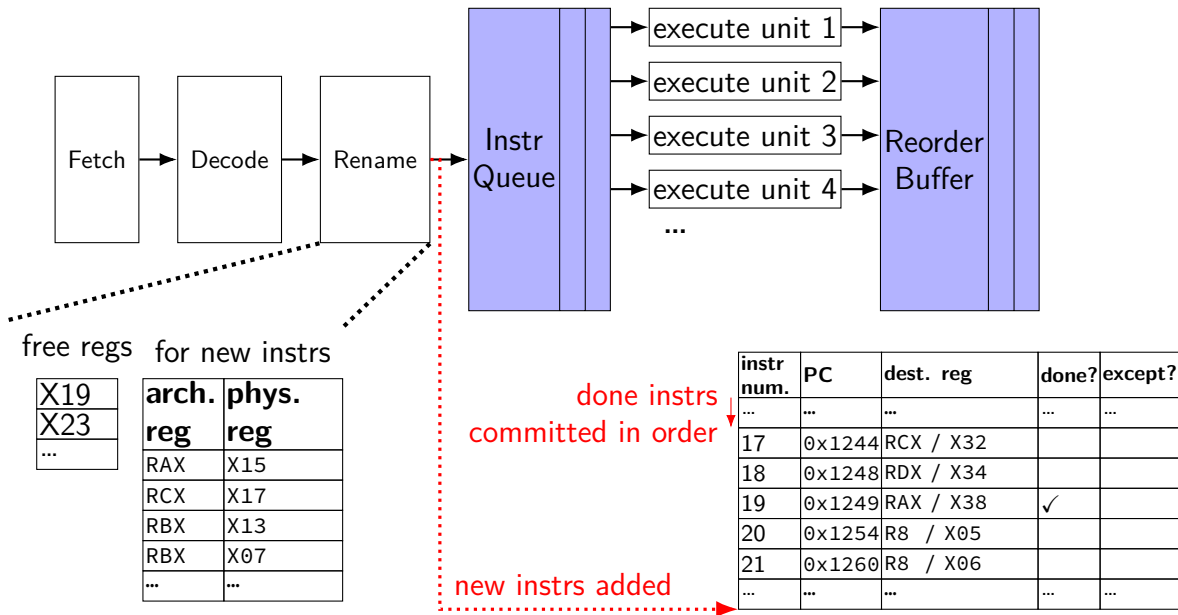


free regs for new instrs

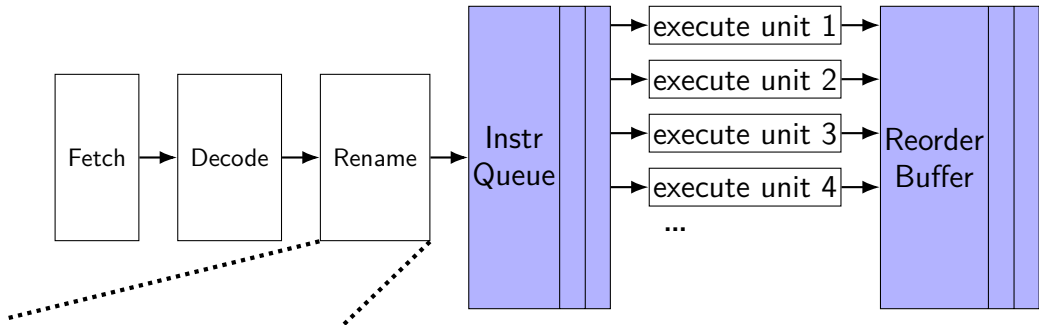
X19
X23
...

arch. reg	phys. reg
RAX	X15
RCX	X17
RBX	X13
RBX	X07
...	...

exceptions and OOO (one strategy)



exceptions and OOO (one strategy)



free regs for new instrs for complete instrs

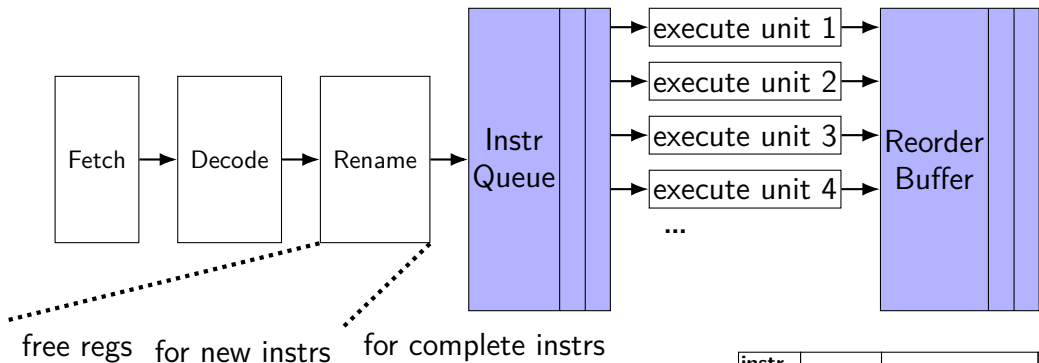
X19
X23
...

arch. reg	phys. reg
RAX	X15
RCX	X17
RBX	X13
RBX	X07
...	...

arch. reg	phys. reg
RAX	X21
RCX	X2 X32
RBX	X48
RDX	X37
...	...

instr num.	PC	dest. reg	done?	except?
...
17	0x1244	RCX / X32	✓	
18	0x1248	RDX / X34		
19	0x1249	RAX / X38	✓	
20	0x1254	R8 / X05		
21	0x1260	R8 / X06		
...

exceptions and OOO (one strategy)



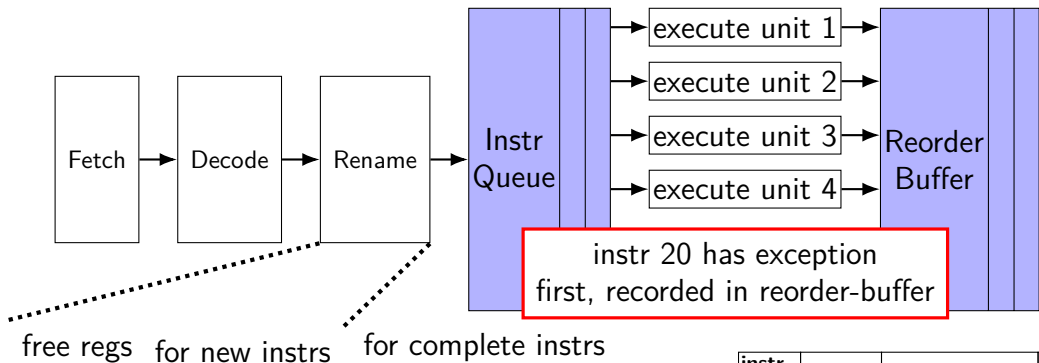
X19
X23
...

arch. reg	phys. reg
RAX	X15
RCX	X17
RBX	X13
RBX	X07
...	...

arch. reg	phys. reg
RAX	X21
RCX	X2 X32
RBX	X48
RDX	X37
...	...

instr num.	PC	dest. reg	done?	except?
...
17	0x1244	RCX / X32	✓	
18	0x1248	RDX / X34		
19	0x1249	RAX / X38	✓	
20	0x1254	R8 / X05		
21	0x1260	R8 / X06		
...

exceptions and OOO (one strategy)



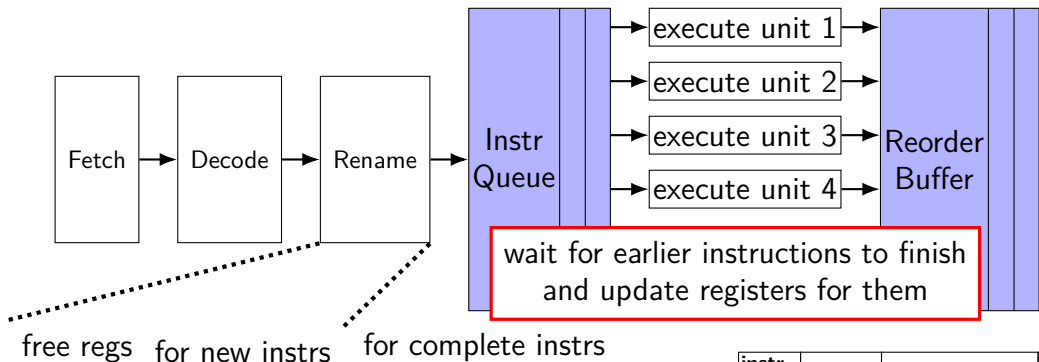
X19
X23
...

arch. reg	phys. reg
RAX	X15
RCX	X17
RBX	X13
RBX	X07
...	...

arch. reg	phys. reg
RAX	X21
RCX	X2 X32
RBX	X48
RDX	X37
...	...

instr num.	PC	dest. reg	done?	except?
...
17	0x1244	RCX / X32	✓	
18	0x1248	RDX / X34		
19	0x1249	RAX / X38	✓	
20	0x1254	R8 / X05	✓	✓
21	0x1260	R8 / X06		
...

exceptions and OOO (one strategy)



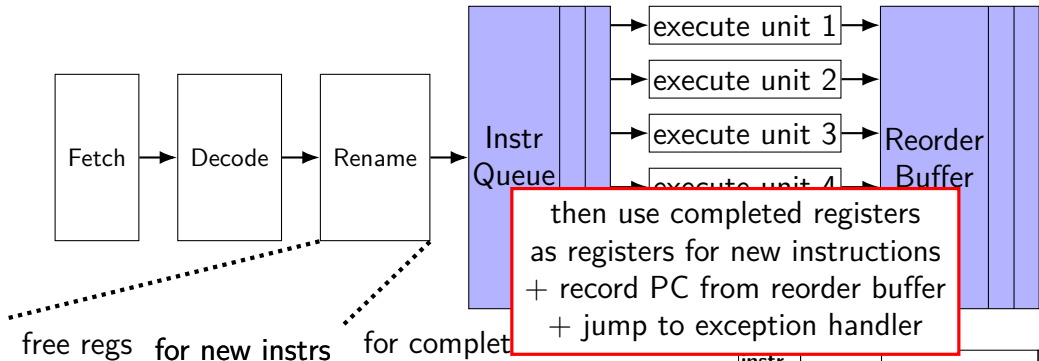
X19
X23
...

arch. reg	phys. reg
RAX	X15
RCX	X17
RBX	X13
RBX	X07
...	...

arch. reg	phys. reg
RAX	X21 X38
RCX	X2 X32
RBX	X48
RDX	X37 X34
...	...

instr num.	PC	dest. reg	done?	except?
...
17	0x1244	RCX / X32	✓	
18	0x1248	RDX / X34	✓	
19	0x1249	RAX / X38	✓	
20	0x1254	R8 / X05	✓	✓
21	0x1260	R8 / X06		
...

exceptions and OOO (one strategy)



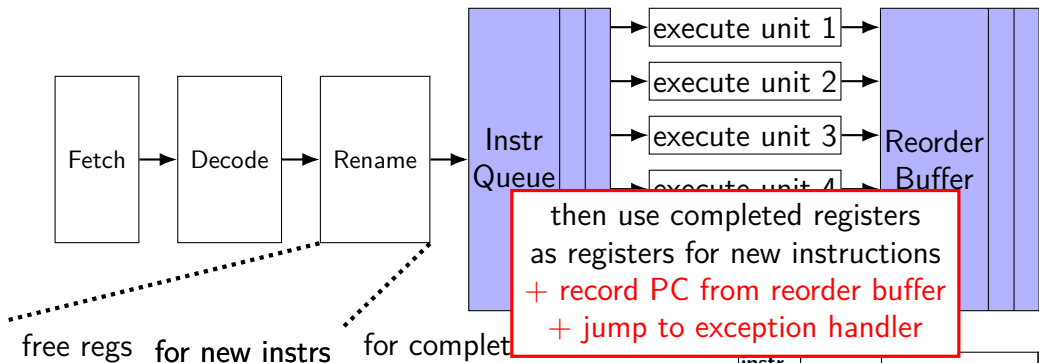
X19
X23
...

arch. reg	phys. reg
RAX	X38
RCX	X32
RBX	X48
RBX	X34
...	...

arch. reg	phys. reg
RAX	X21 X38
RCX	X2 X32
RBX	X48
RDX	X37 X34
...	...

instr num.	PC	dest. reg	done?	except?
...
17	0x1244	RCX / X32	✓	
18	0x1248	RDX / X34	✓	
19	0x1249	RAX / X38	✓	
20	0x1254	R8 / X05	✓	✓
21	0x1260	R8 / X06		
...

exceptions and OOO (one strategy)



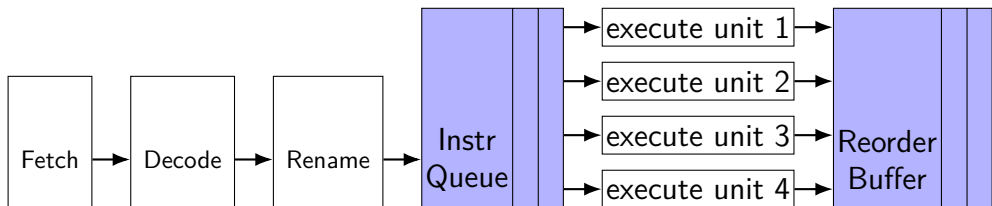
X19
X23
...

arch. reg	phys. reg
RAX	X38
RCX	X32
RBX	X48
RBX	X34
...	...

arch. reg	phys. reg
RAX	X21 X38
RCX	X2 X32
RBX	X48
RDX	X37 X34
...	...

instr num.	PC	dest. reg	done?	except?
...
17	0x1244	RCX / X32	✓	
18	0x1248	RDX / X34	✓	
19	0x1249	RAX / X38	✓	
20	0x1254	R8 / X05	✓	✓
21	0x1260	R8 / X06		
...

exceptions and OOO (one strategy)



variation: could store architectural reg. values instead of mapping for completed instrs. (and copy values instead of mapping on exception)

free regs for new instrs for complete instrs

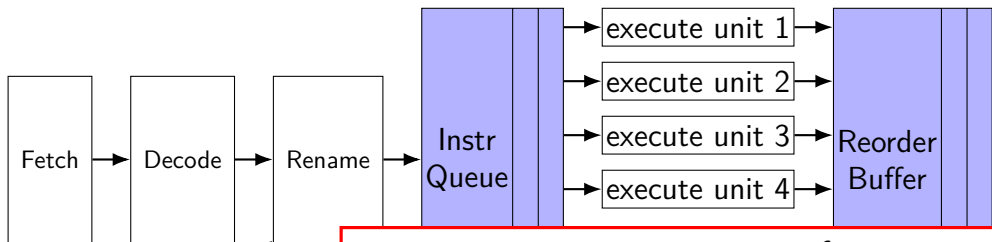
X19
X23
...

arch. reg	phys. reg
RAX	X15
RCX	X17
RBX	X13
RBX	X07
...	...

arch. reg	value
RAX	0x12343
RCX	0x234543
RBX	0x56782
RDX	0xF83A4
...	...

instr num.	PC	dest. reg	done?	except?
...
17	0x1244	RCX / X32	✓	
18	0x1248	RDX / X34	✓	
19	0x1249	RAX / X38	✓	
20	0x1254	R8 / X05	✓	✓
21	0x1260	R8 / X06		
...

exceptions and OOO (one strategy)



stopping instructions in progress for exception
similar to how 'squashing' mispredicted instructions

free regs for new instrs for complete instrs

X19
X23
...

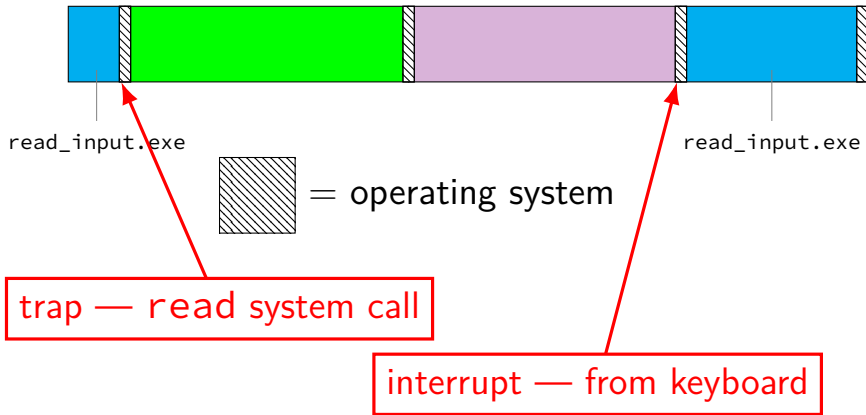
arch. reg	phys. reg
RAX	X15
RCX	X17
RBX	X13
RBX	X07
...	...

arch. reg	phys. reg
RAX	X21 X38
RCX	X2 X32
RBX	X48
RDX	X37 X34
...	...

instr num.	PC	dest. reg	done?	except?
...
17	0x1244	RCX / X32	✓	
18	0x1248	RDX / X34	✓	
19	0x1249	RAX / X38	✓	
20	0x1254	R8 / X05	✓	✓
21	0x1260	R8 / X06		
...

backup slides

keyboard input timeline



backup slides