

CSC 252: Computer Organization

Spring 2023: Lecture 14

Instructor: Yuhao Zhu

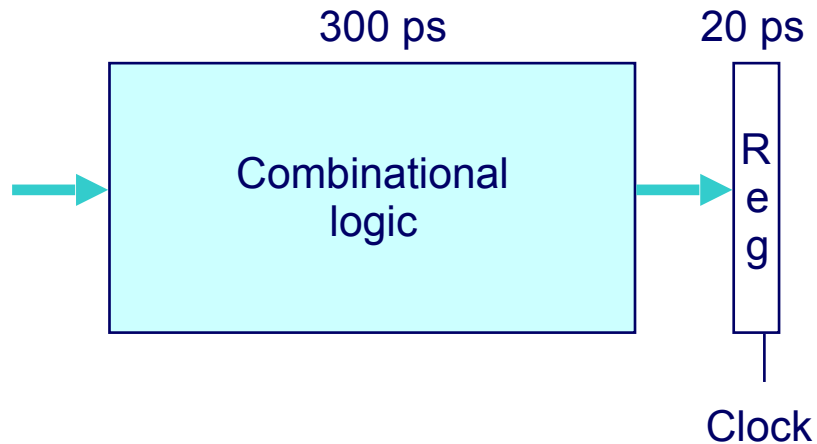
Department of Computer Science
University of Rochester

Announcement

- Programming assignment 3 out.
- Open-book midterm (anything on paper is OK; no electronics)

12	13	14	15	16	17	18
19	20	21	22	23	24 Today	25
26	27	28	Mar 1 Due	2	3 Mid-term	4

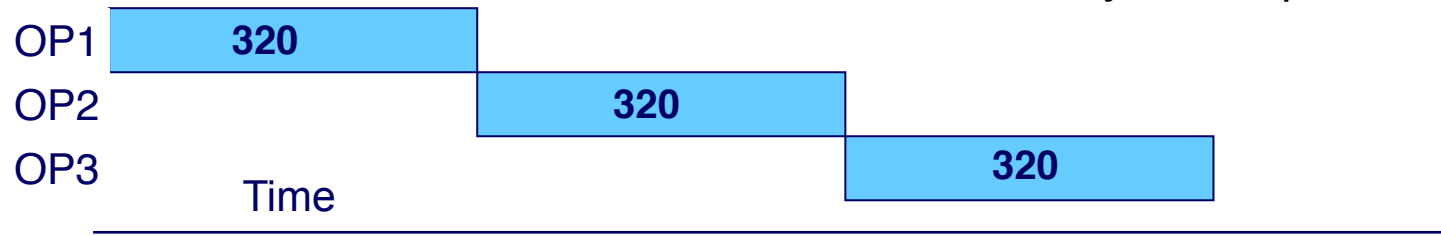
A Motivating Example



- Computation requires total of 300 picoseconds
- Additional 20 picoseconds to save result in register
- Must have clock cycle time of at least 320 ps

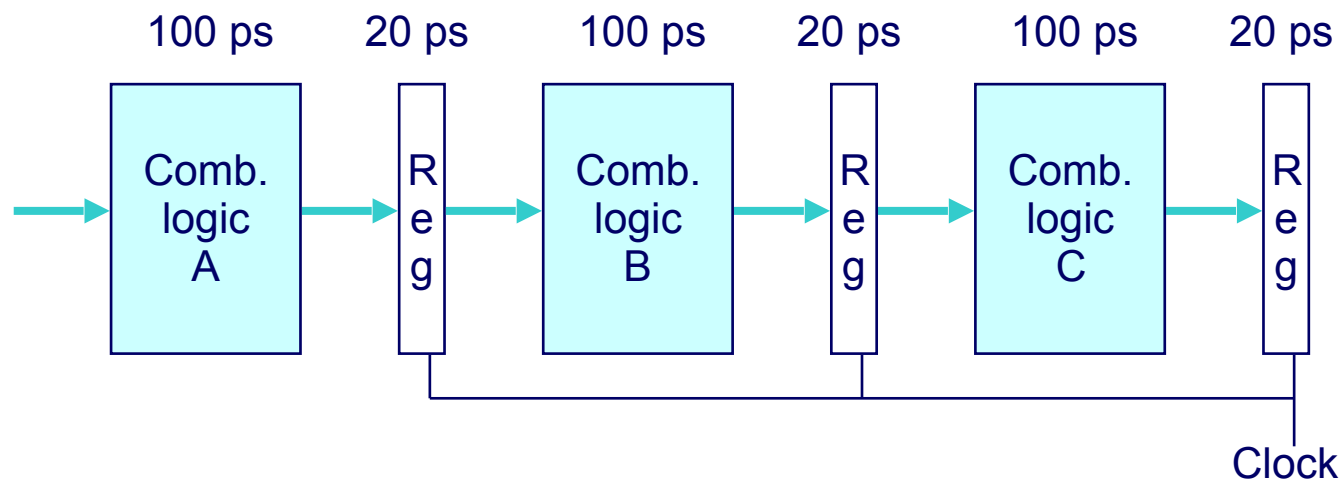
Pipeline Diagrams

- Time to finish 3 insts = 960 ps
- Each inst.'s latency is 320 ps



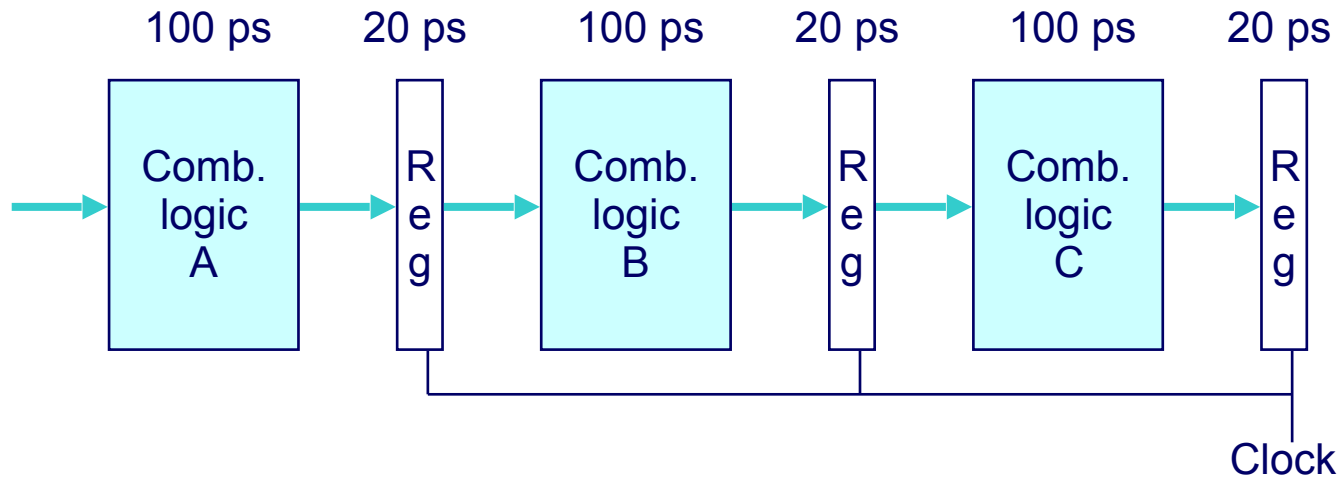
- 3 instructions will take 960 ps to finish
 - First cycle: Inst 1 takes 300 ps to compute new state, 20 ps to store the new states
 - Second cycle: Inst 2 starts; it takes 300 ps to compute new states, 20 ps to store new states
 - And so on...

3-Stage Pipelined Version

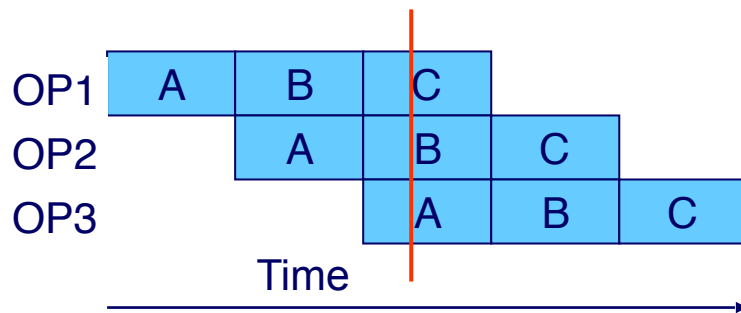


- Divide combinational logic into 3 stages of 100 ps each
- Insert registers between stages to store intermediate data between stages. These are called pipeline registers (ISA-invisible)
- Can begin a new instruction as soon as the previous one finishes stage A and has stored the intermediate data.
 - Begin new operation every **120 ps**
 - **Cycle time can be reduced to 120 ps**

3-Stage Pipelined Version

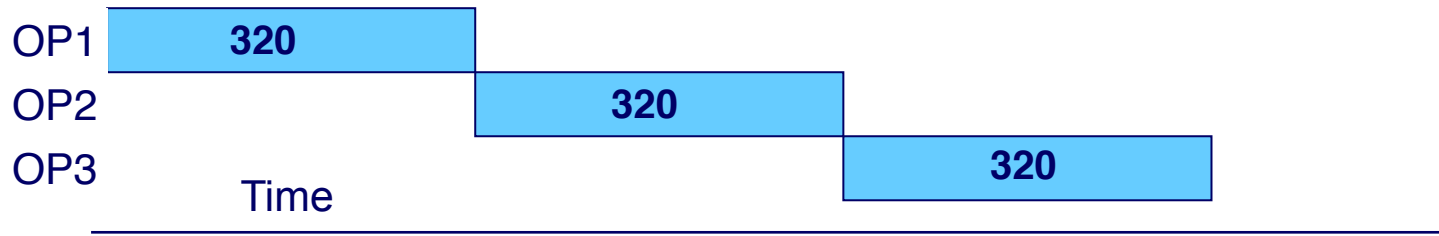


3-Stage Pipelined



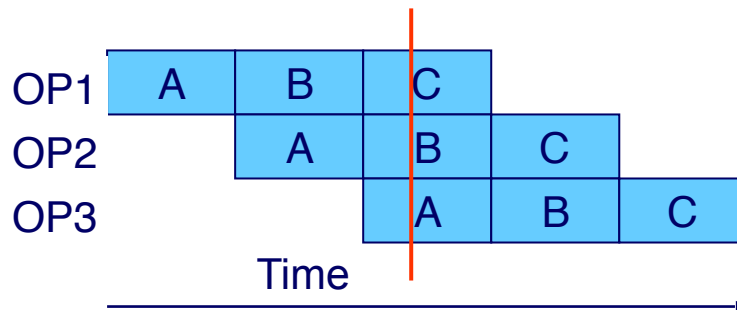
Comparison

Unpipelined



- Time to finish 3 insts = 960 ps
- Each inst.'s latency is 320 ps

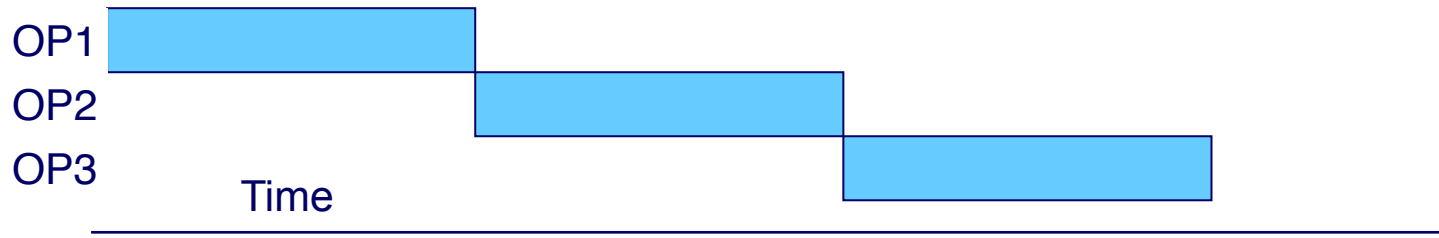
3-Stage Pipelined



- Time to finish 3 insts = $120 * 5 = 600$ ps
- But each inst.'s latency increases: $120 * 3 = 360$ ps

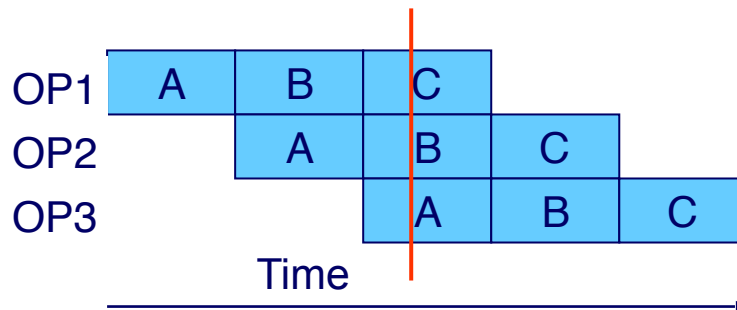
Benefits of Pipelining

- Time to finish 3 insts = 960 ps
- Each inst.'s latency is 320 ps



1. Reduce the cycle time from 320 ps to 120 ps

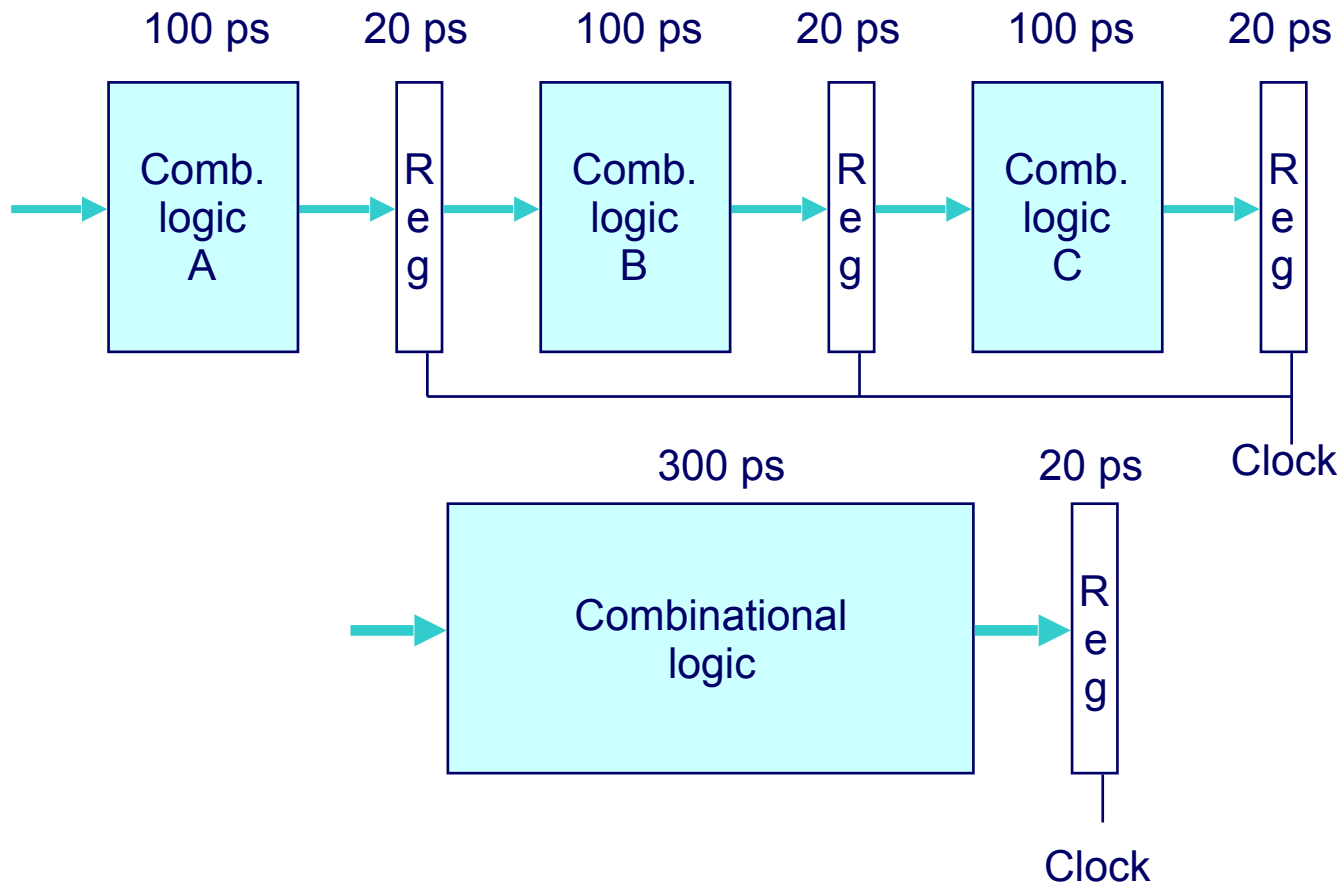
2. CPI reduces from 1 to 1/3 (i.e., executing 3 instruction in one cycle)



- Time to finish 3 insets = $120 * 5 = 600$ ps
- But each inst.'s latency increases: $120 * 3 = 360$ ps

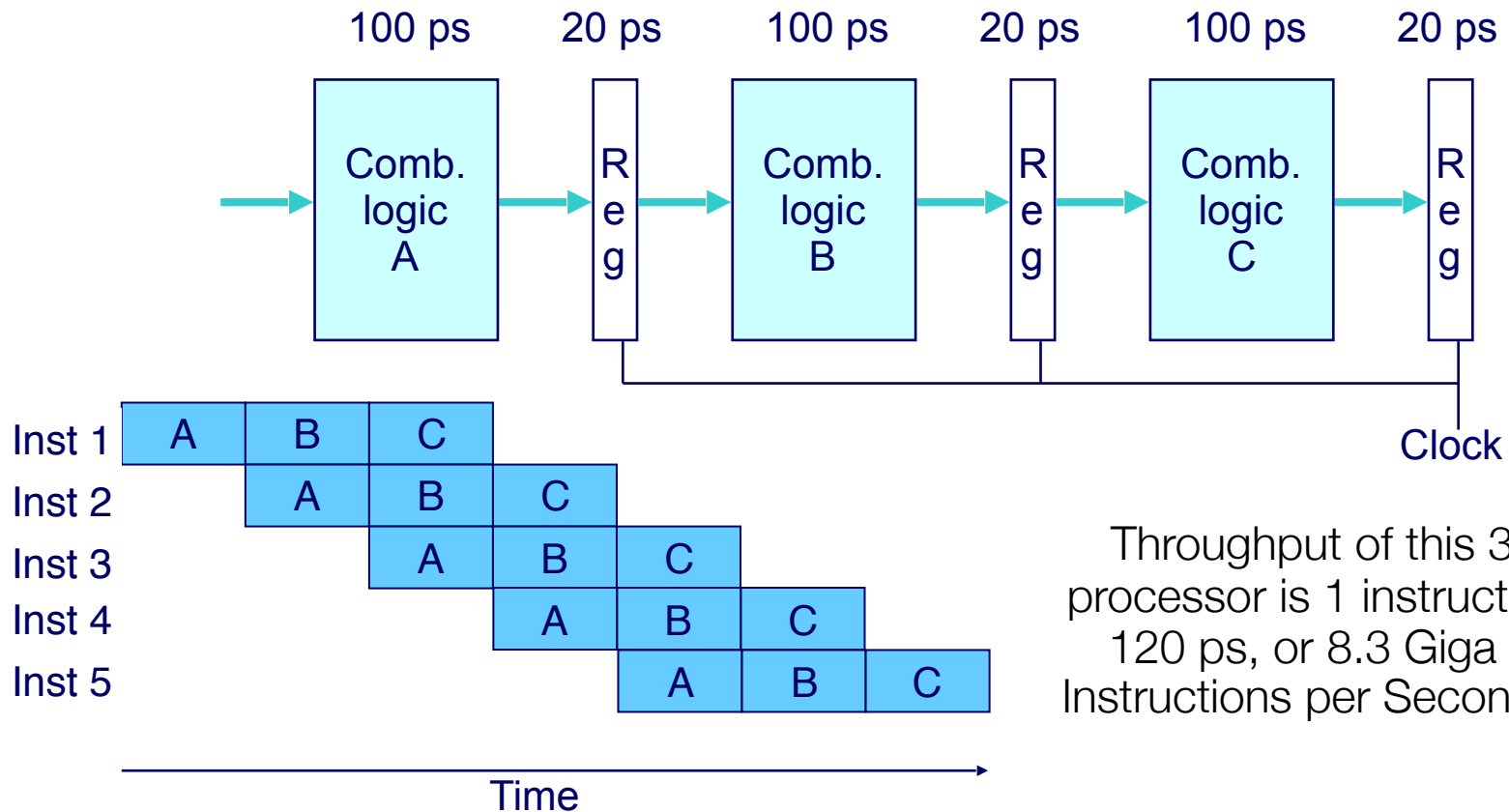
Pipeline Trade-offs

- Pros: Decrease the total execution time (Increase the “**throughput**”).
- Cons: Increase the latency of each instruction as new registers are needed between pipeline stages.



Throughput

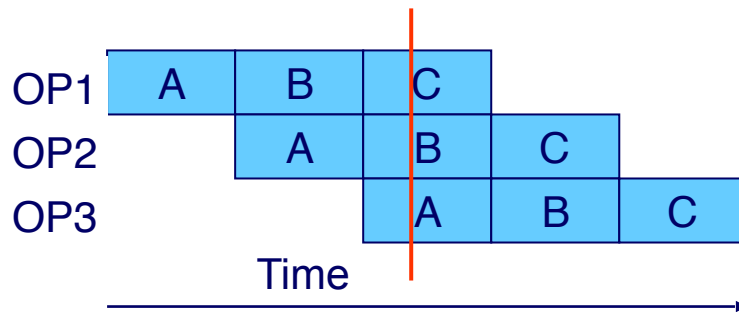
- The rate at which the processor can finish executing an instruction (at the steady state).



Throughput of this 3-stage processor is 1 instruction every 120 ps, or 8.3 Giga (billion) Instructions per Second (GIPS).

One Requirement of Pipelining

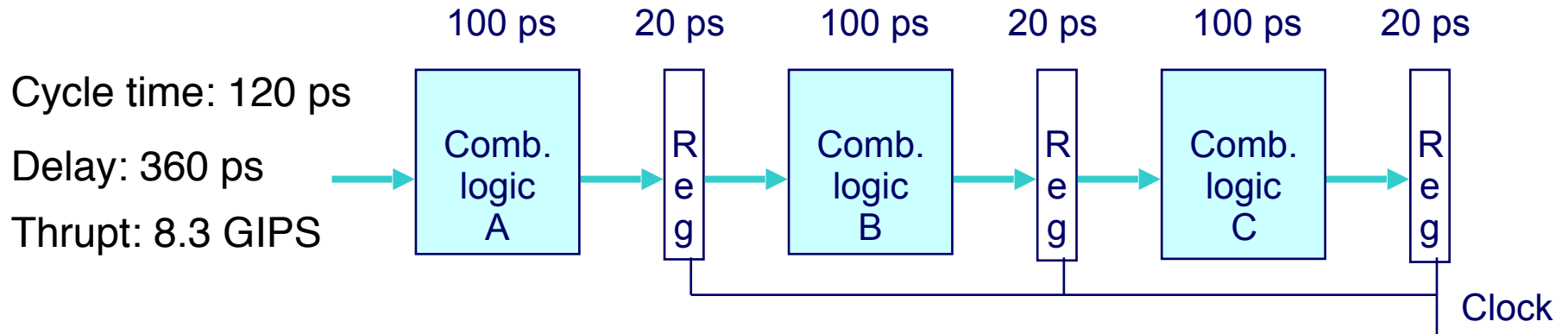
- The stages need to be using different hardware structures.
- That is, Stage A, Stage B, and Stage C need to exercise different parts of the combination logic.



- Time to finish 3 insets = $120 * 5 = 600$ ps
- But each inst.'s latency increases: $120 * 3 = 360$ ps

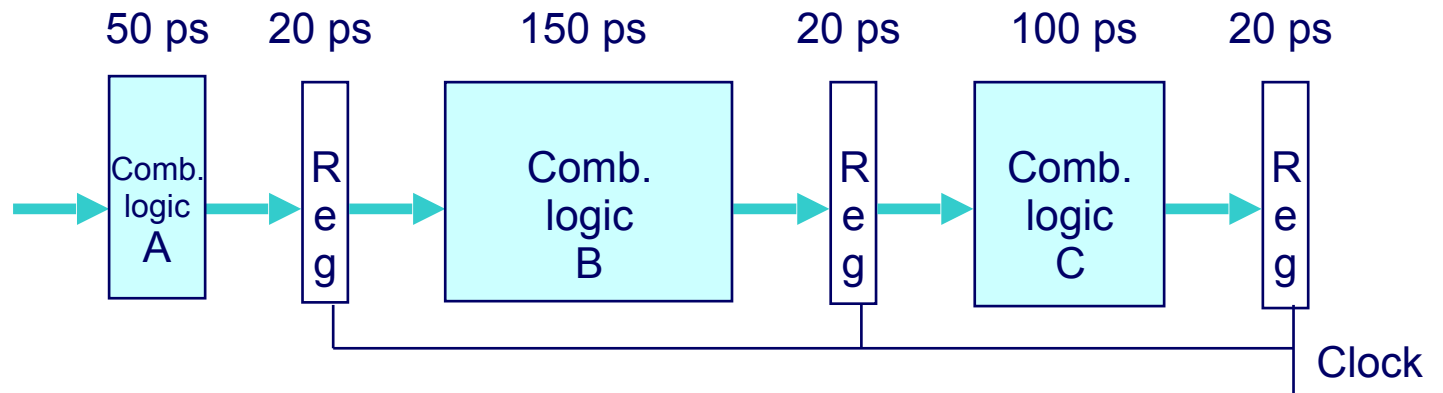
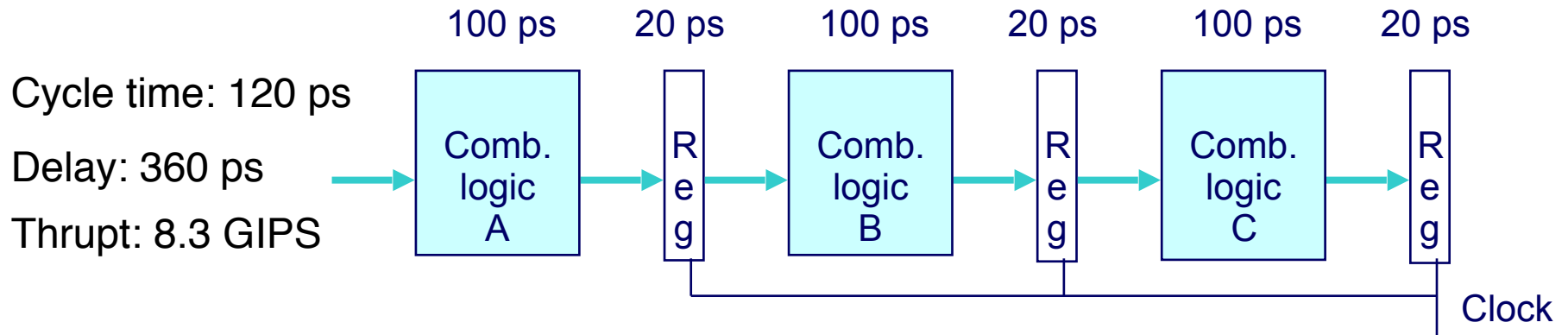
Aside: Unbalanced Pipeline

- A pipeline's delay is limited by the slowest stage. This limits the cycle time and the throughput



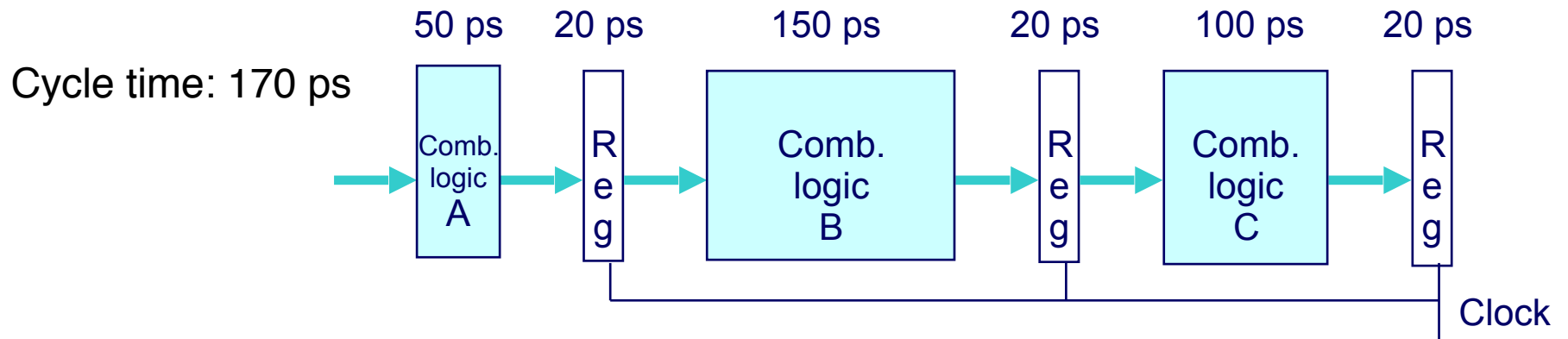
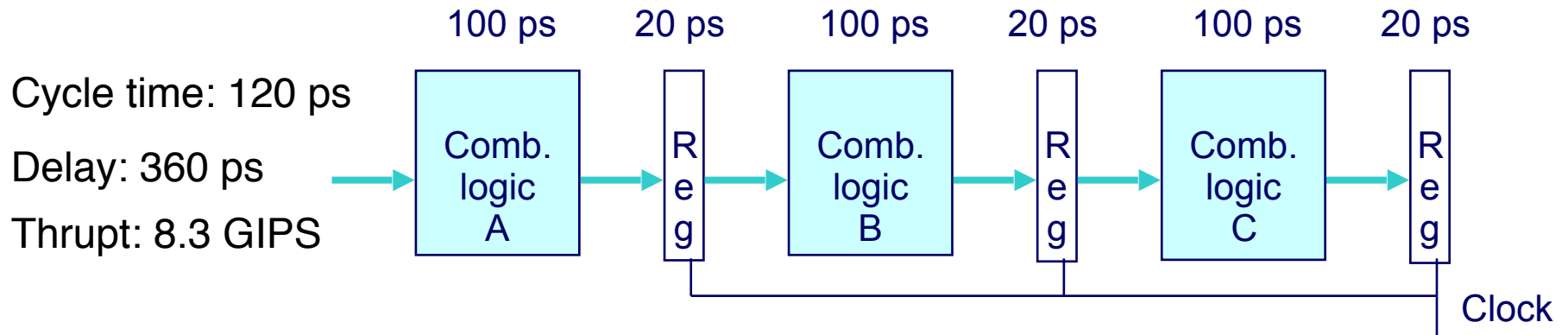
Aside: Unbalanced Pipeline

- A pipeline's delay is limited by the slowest stage. This limits the cycle time and the throughput



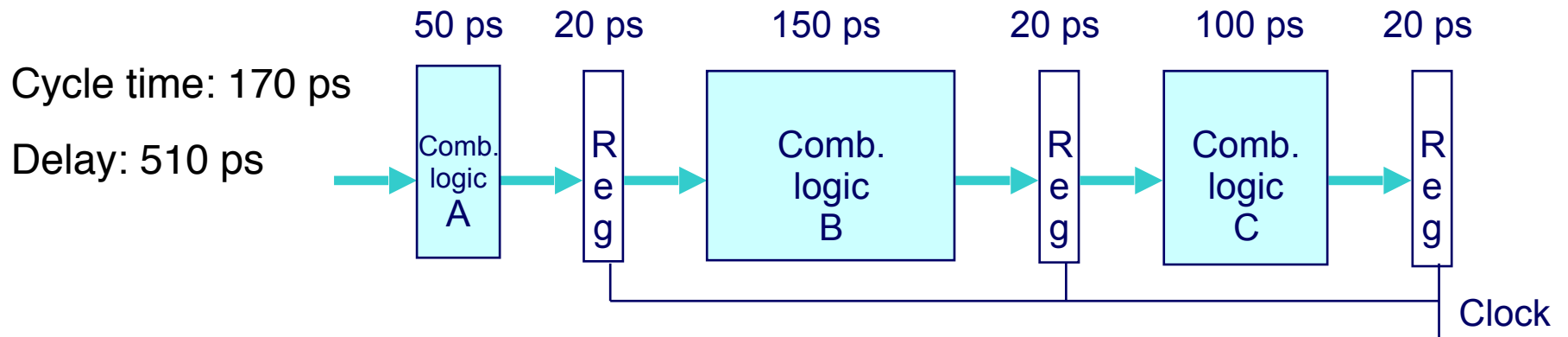
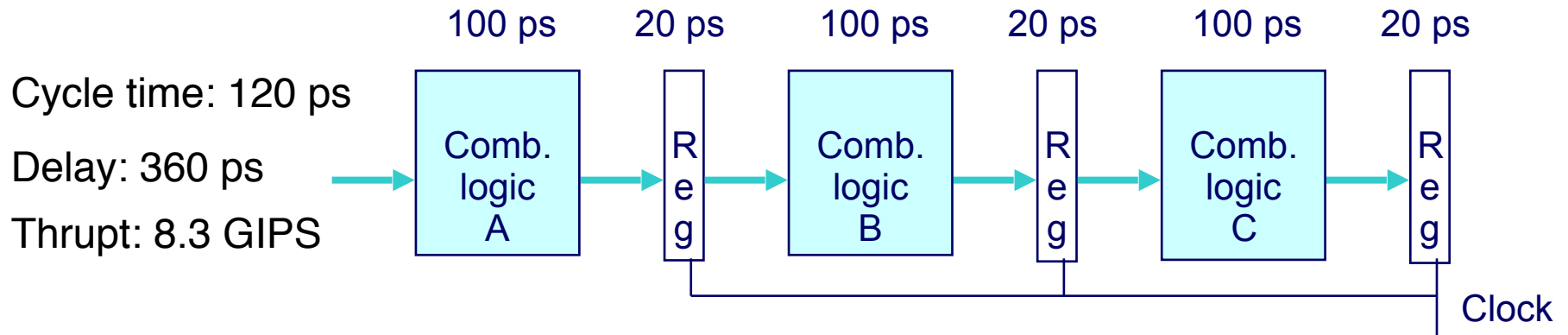
Aside: Unbalanced Pipeline

- A pipeline's delay is limited by the slowest stage. This limits the cycle time and the throughput



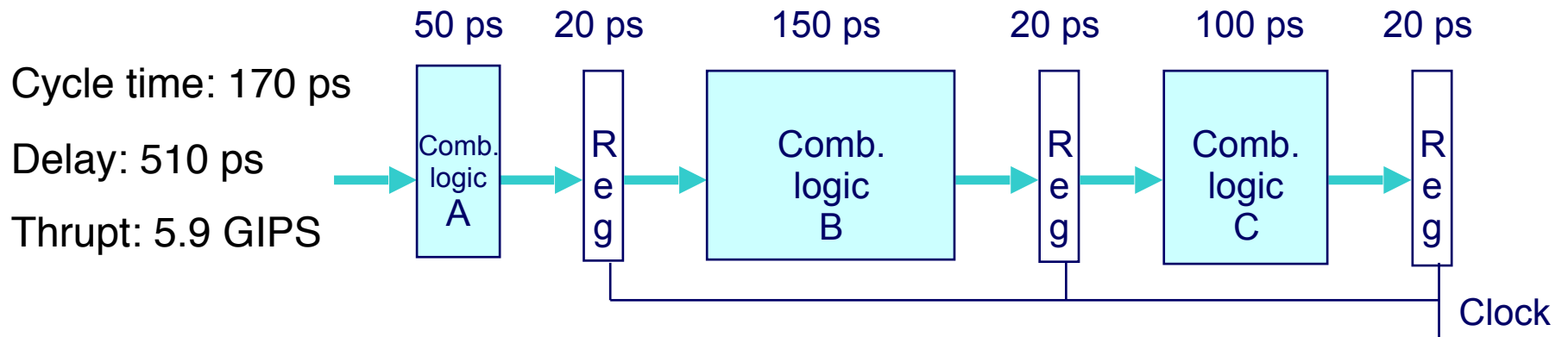
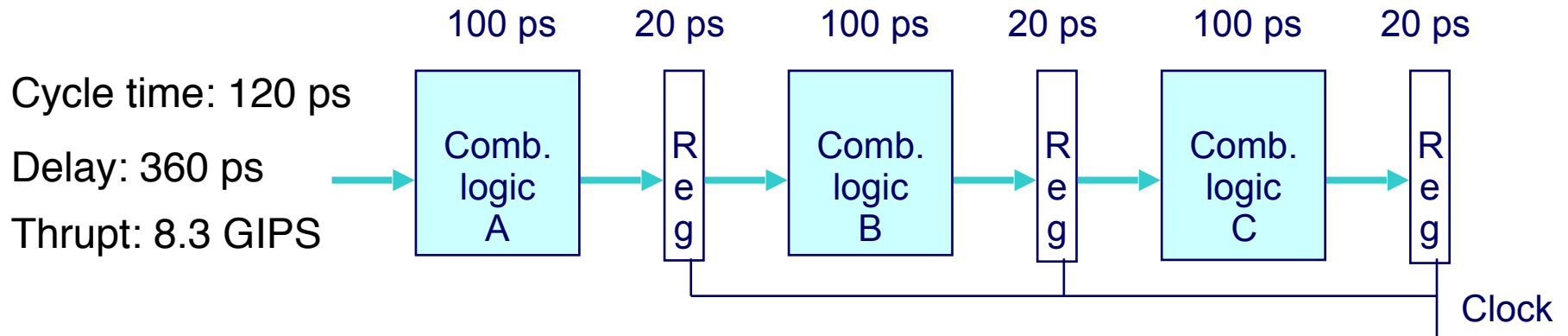
Aside: Unbalanced Pipeline

- A pipeline's delay is limited by the slowest stage. This limits the cycle time and the throughput



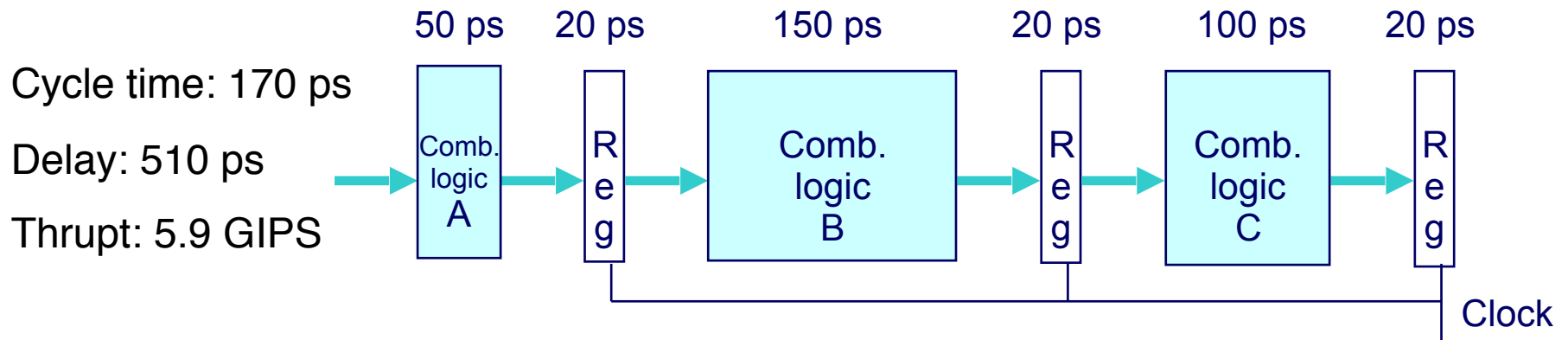
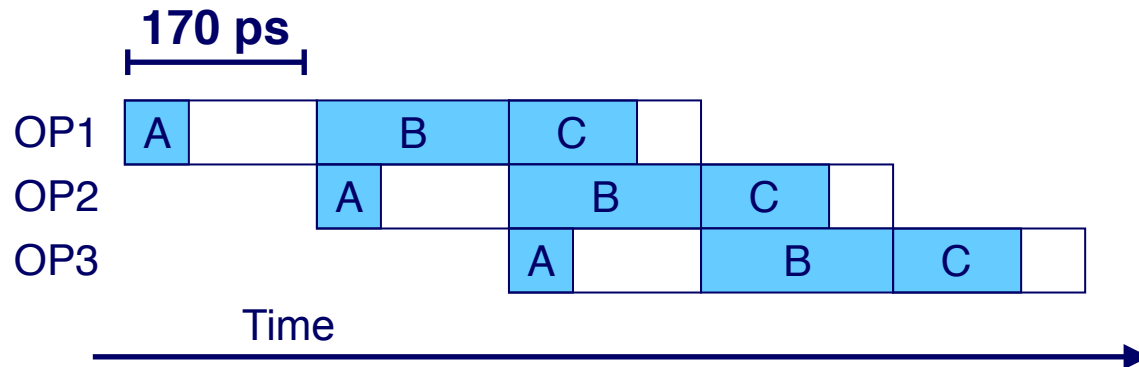
Aside: Unbalanced Pipeline

- A pipeline's delay is limited by the slowest stage. This limits the cycle time and the throughput



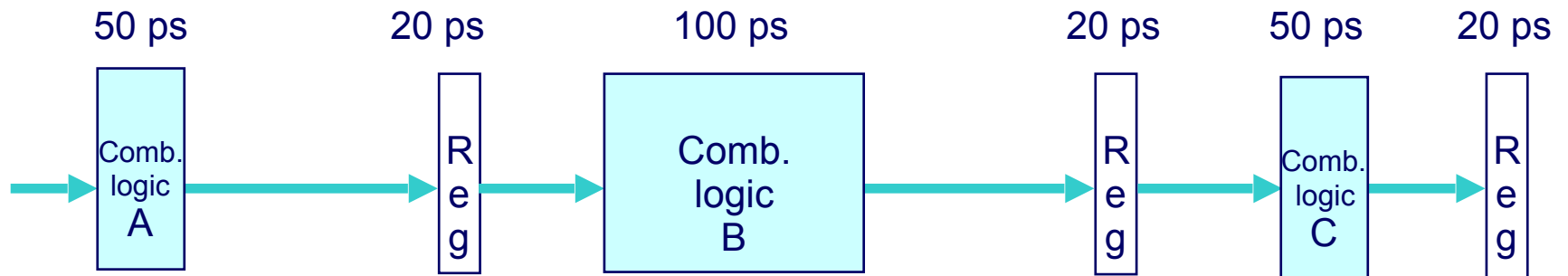
Aside: Unbalanced Pipeline

- A pipeline's delay is limited by the slowest stage. This limits the cycle time and the throughput



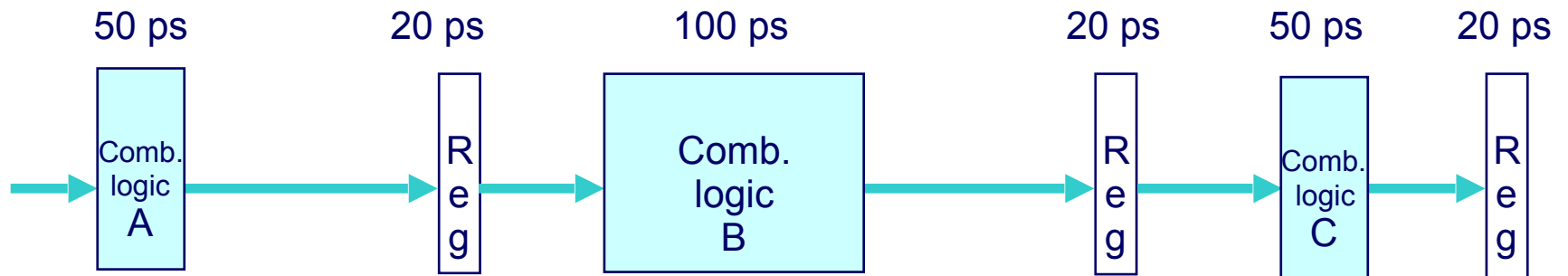
Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages



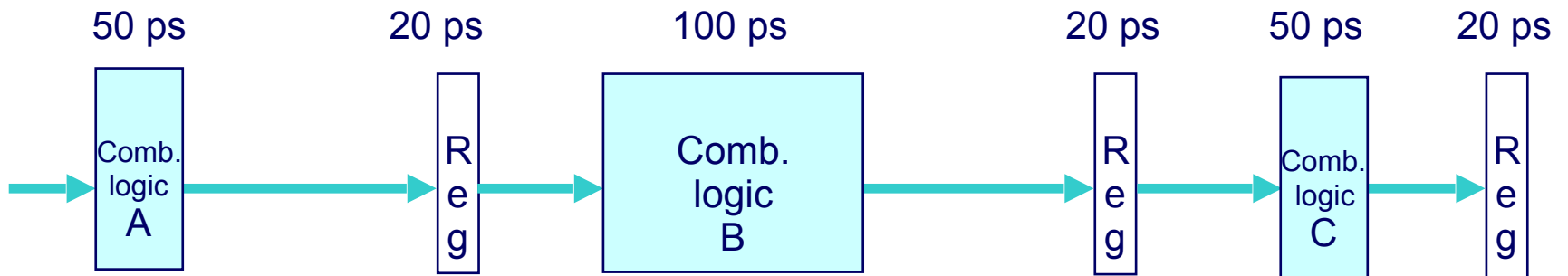
Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages
 - Not always possible. What to do if we can't further pipeline a stage?



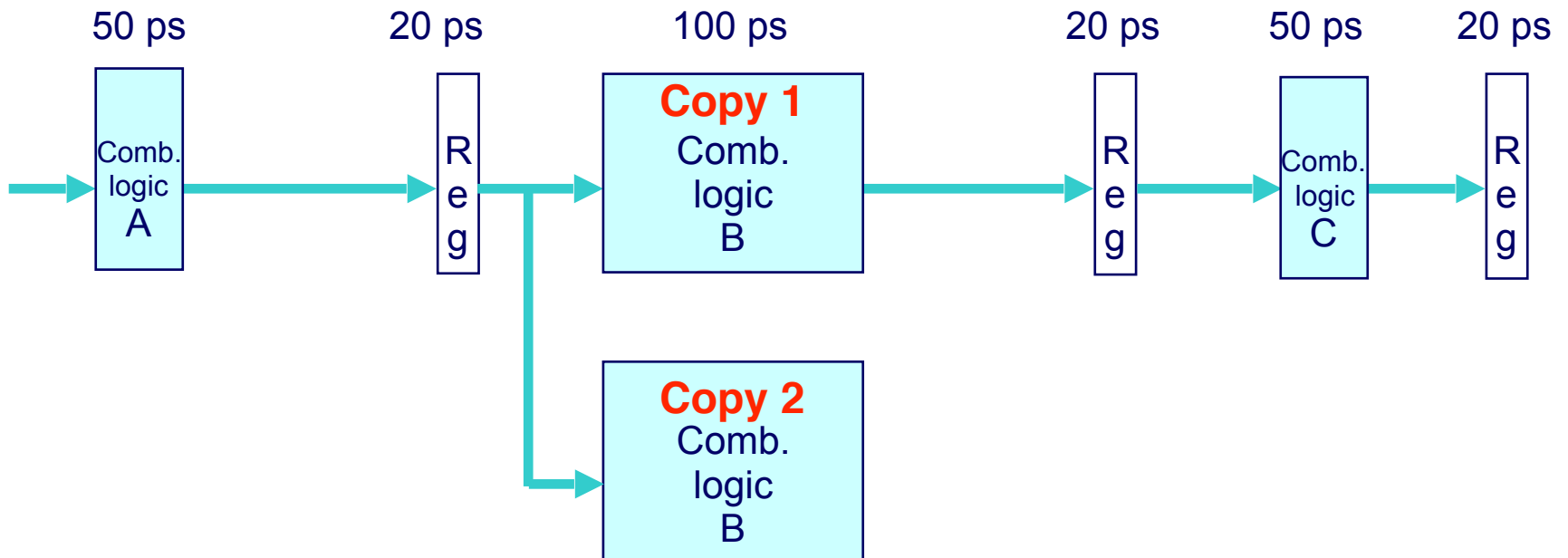
Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages
 - Not always possible. What to do if we can't further pipeline a stage?
- Solution 2: Use multiple copies of the slow component



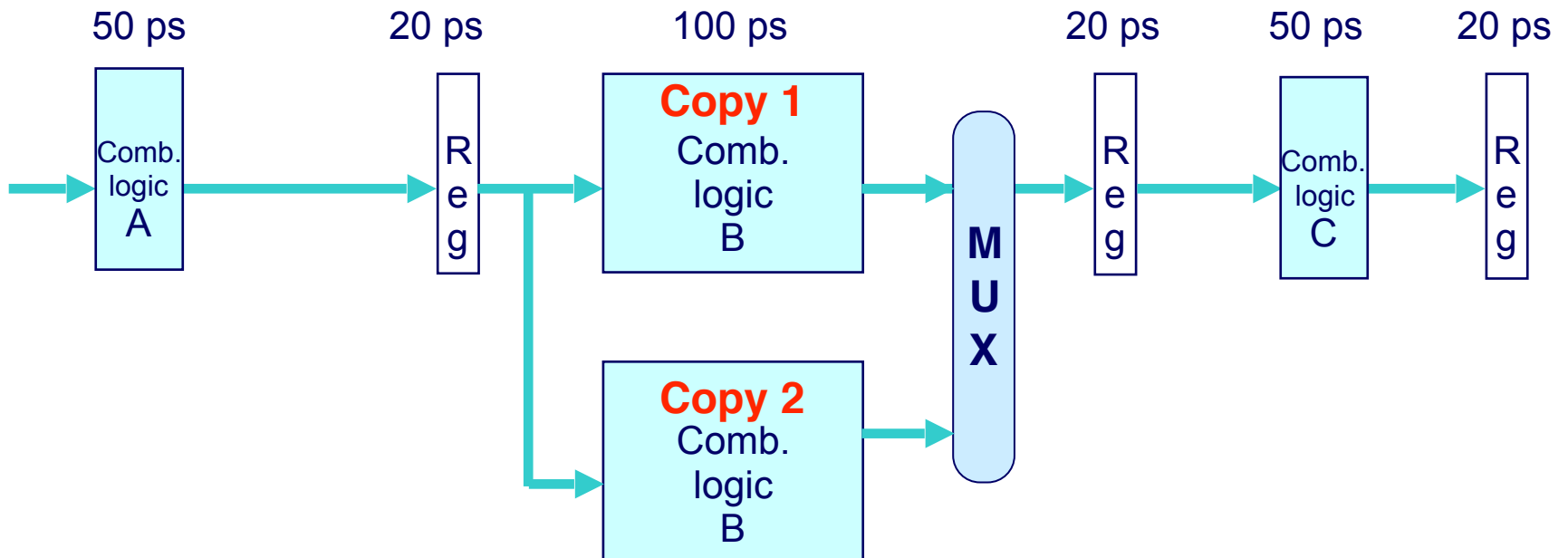
Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages
 - Not always possible. What to do if we can't further pipeline a stage?
- Solution 2: Use multiple copies of the slow component



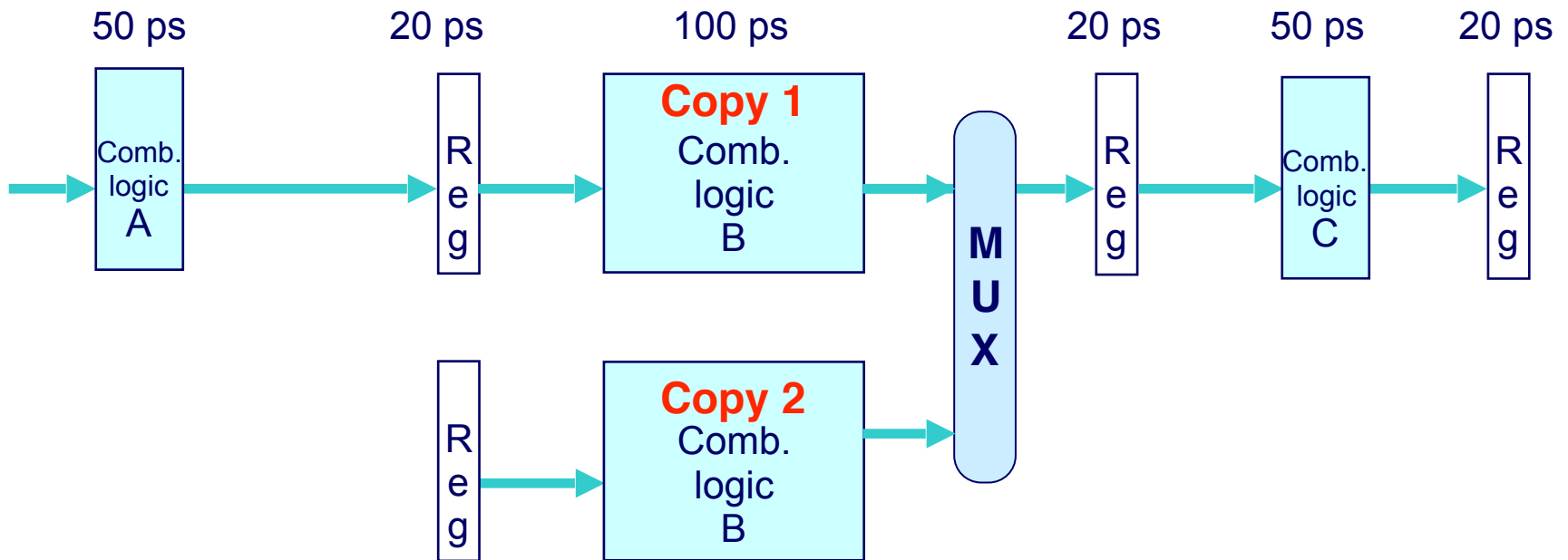
Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages
 - Not always possible. What to do if we can't further pipeline a stage?
- Solution 2: Use multiple copies of the slow component



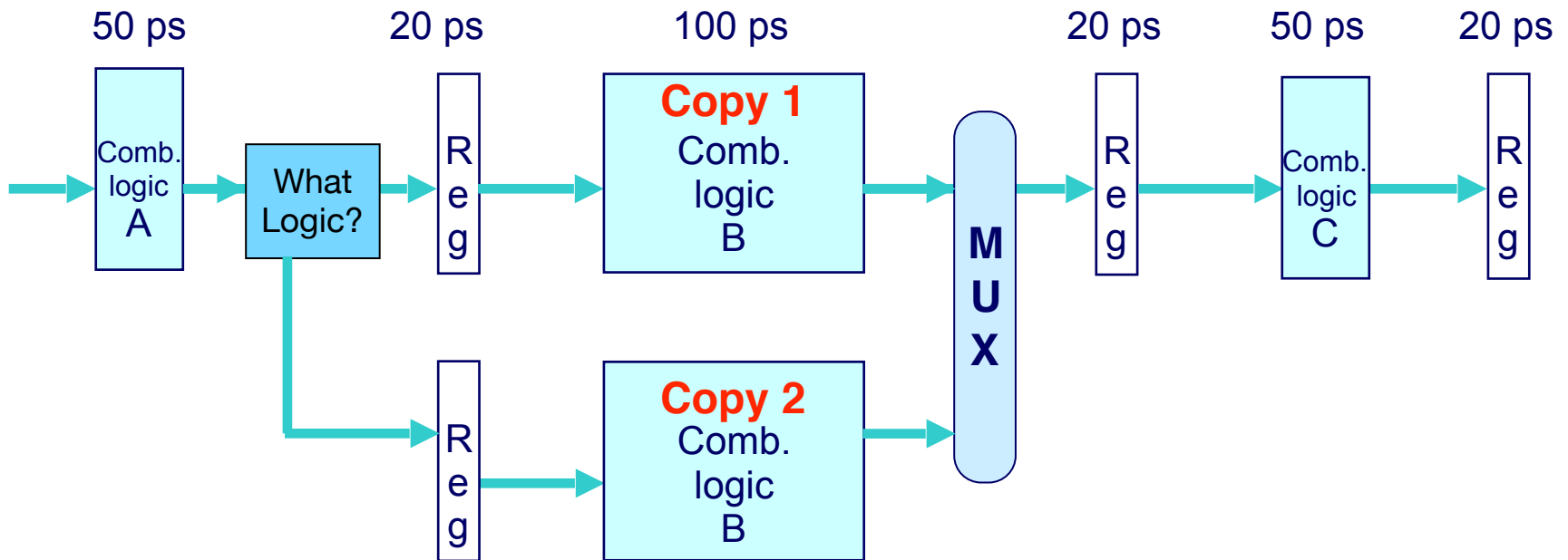
Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages
 - Not always possible. What to do if we can't further pipeline a stage?
- Solution 2: Use multiple copies of the slow component



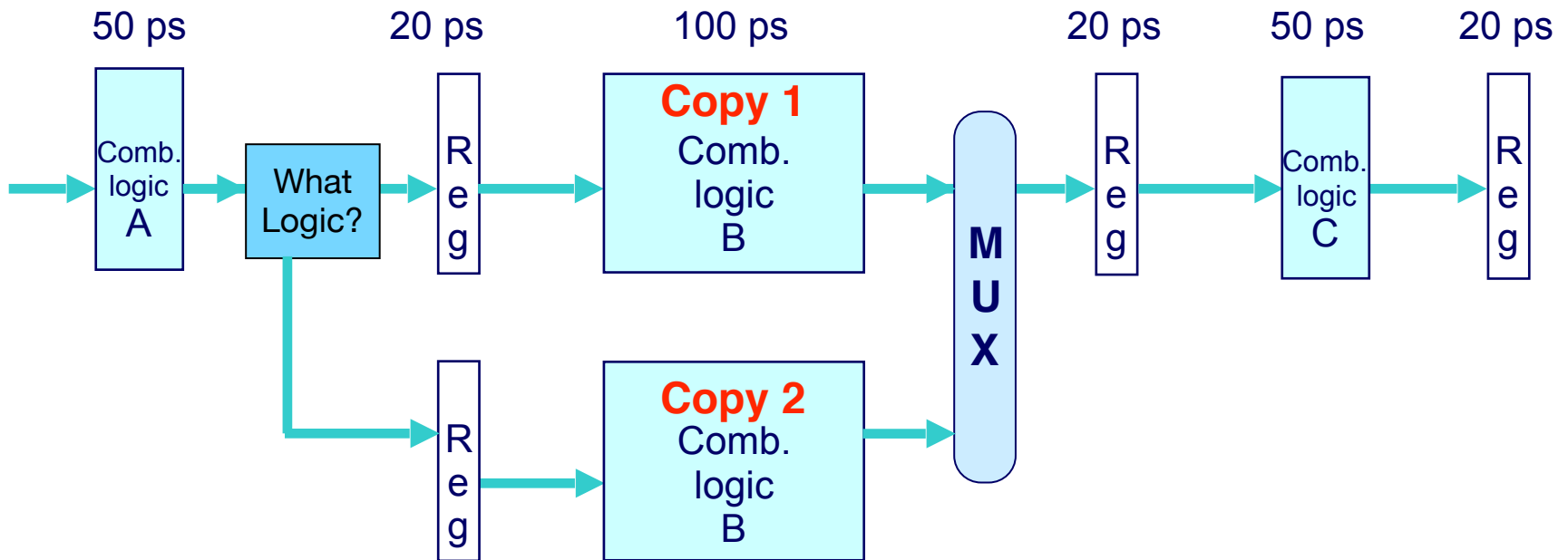
Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages
 - Not always possible. What to do if we can't further pipeline a stage?
- Solution 2: Use multiple copies of the slow component



Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages
 - Not always possible. What to do if we can't further pipeline a stage?
- Solution 2: Use multiple copies of the slow component

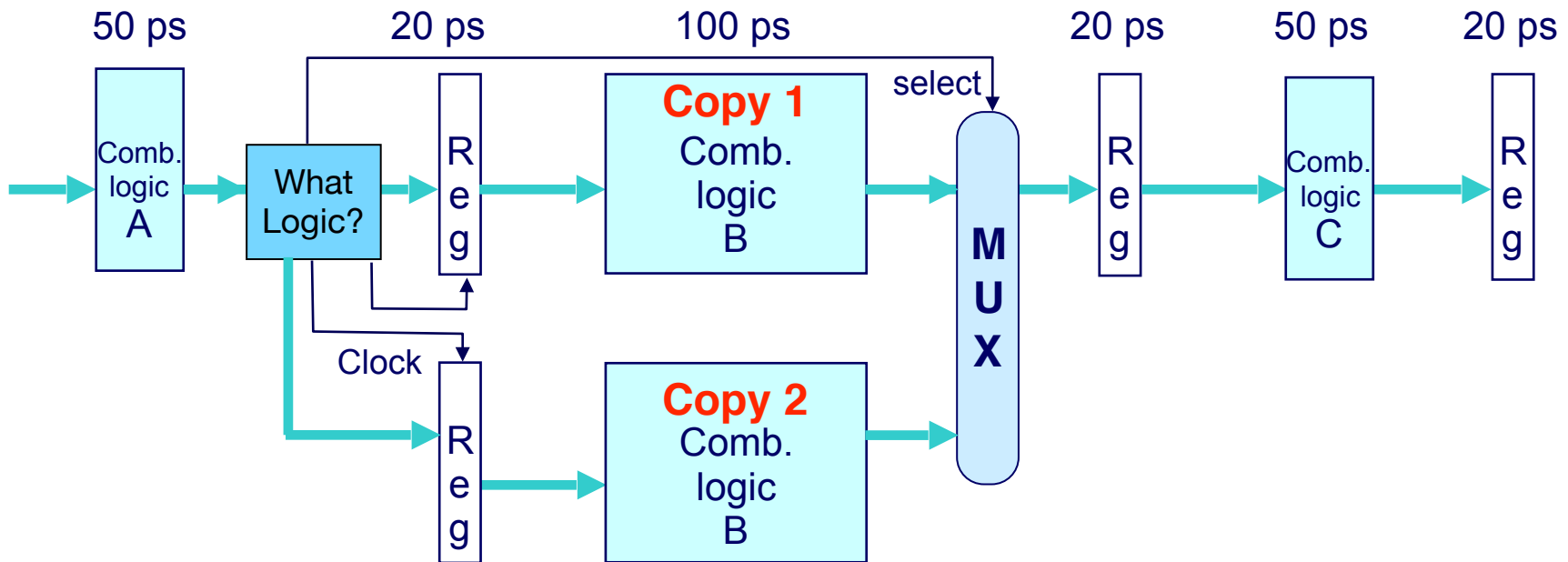


What logic do you need there?

Hint: it needs to control the clock signals of the two registers and the select signal of the MUX.

Aside: Mitigating Unbalanced Pipeline

- Solution 1: Further pipeline the slow stages
 - Not always possible. What to do if we can't further pipeline a stage?
- Solution 2: Use multiple copies of the slow component

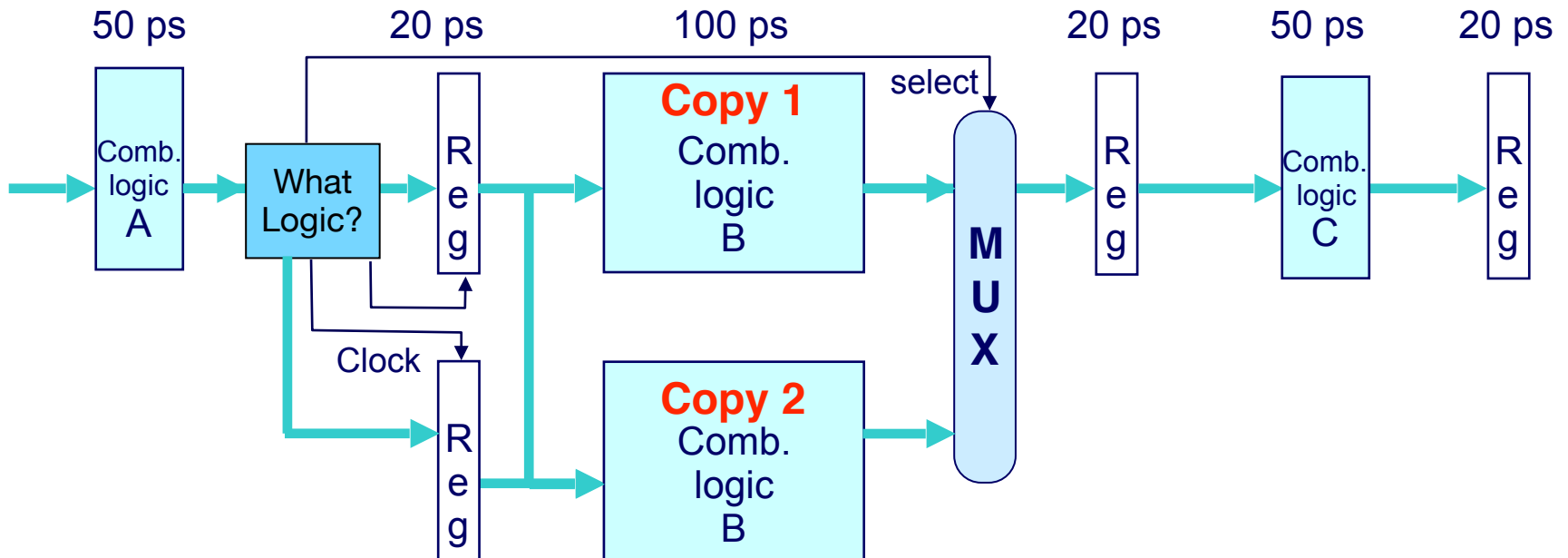


What logic do you need there?

Hint: it needs to control the clock signals of the two registers and the select signal of the MUX.

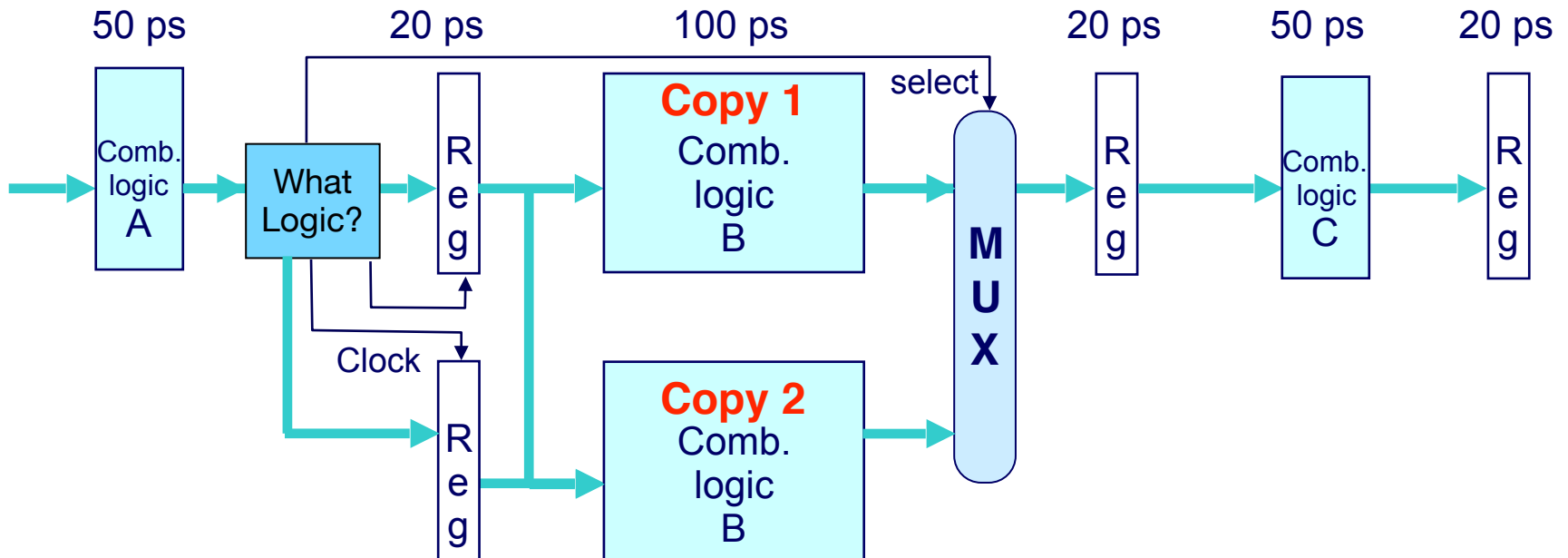
Aside: Mitigating Unbalanced Pipeline

- Data sent to copy 1 in odd cycles and to copy 2 in even cycles.



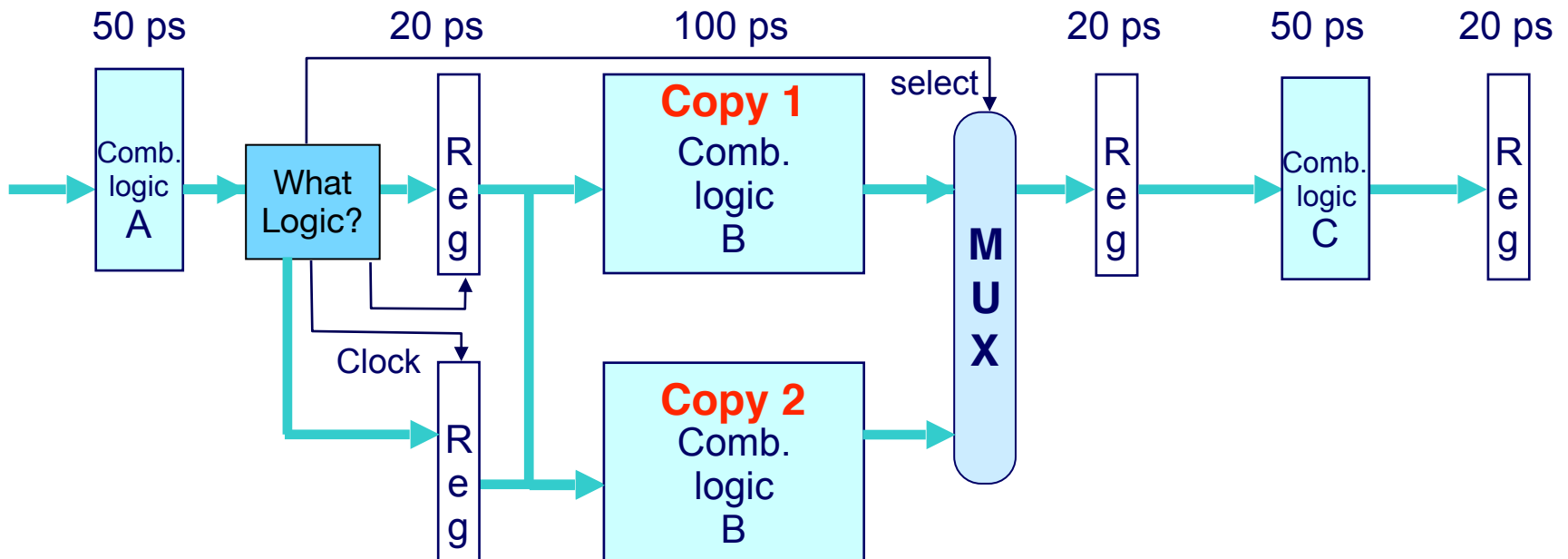
Aside: Mitigating Unbalanced Pipeline

- Data sent to copy 1 in odd cycles and to copy 2 in even cycles.
- This is called 2-way interleaving. Effectively the same as pipelining Comb. logic B into two sub-stages.



Aside: Mitigating Unbalanced Pipeline

- Data sent to copy 1 in odd cycles and to copy 2 in even cycles.
- This is called 2-way interleaving. Effectively the same as pipelining Comb. logic B into two sub-stages.
- The cycle time is reduced to 70 ps (as opposed to 120 ps) at the cost of extra hardware.



Another Way to Look At the Microarchitecture

Principles:

- Execute each instruction one at a time, one after another
- Express every instruction as series of **simple steps**
- Dedicated hardware structure for completing each step
- Follow same general flow for each instruction type

Fetch: Read instruction from instruction memory

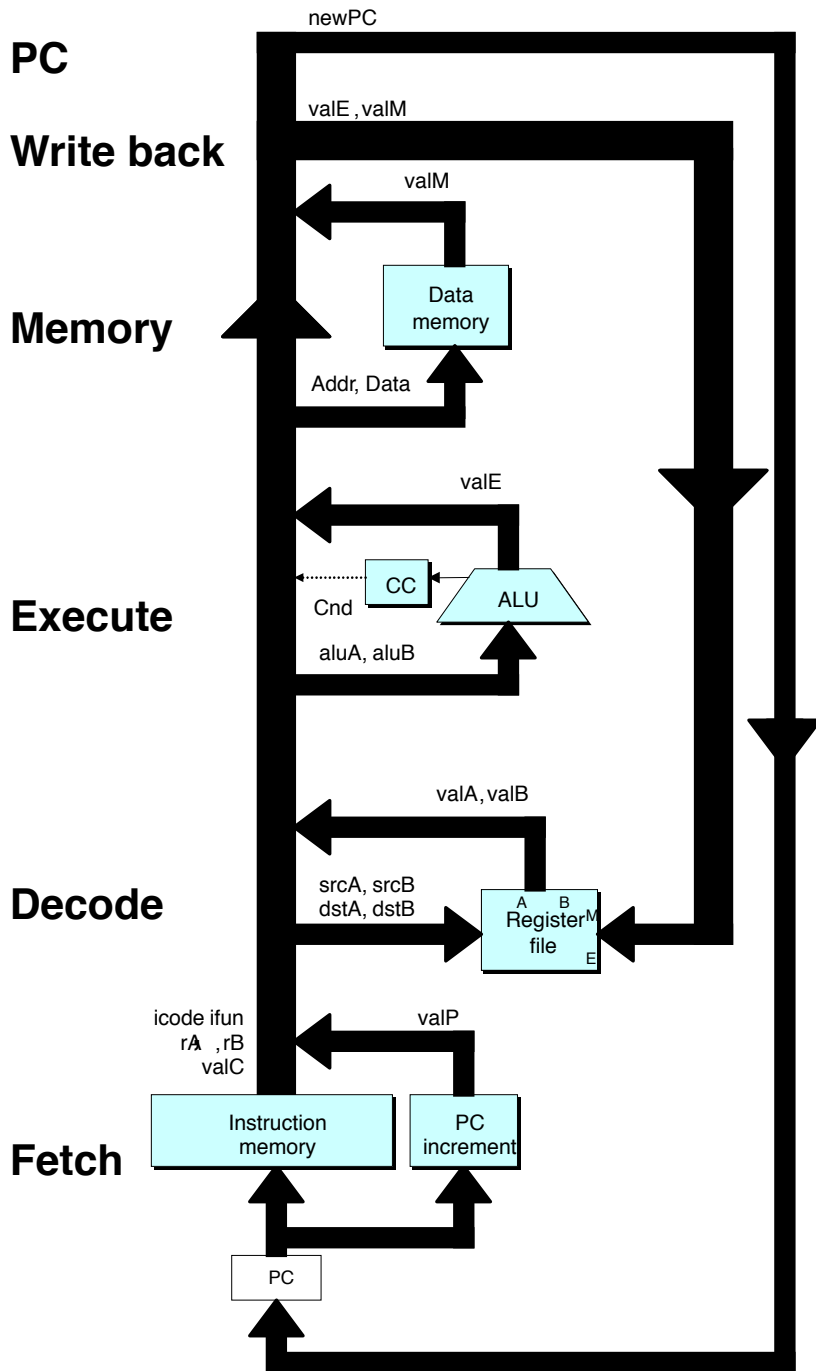
Decode: Read program registers

Execute: Compute value or address

Memory: Read or write data

Write Back: Write program registers

PC: Update program counter



Fetch

- Read instruction from instruction memory

Decode

- Read program registers

Execute

- Compute value or address

Memory

- Read or write data

Write Back

- Write program registers

PC

- Update program counter

Stage Computation: Arith/Log. Ops



Stage Computation: Arith/Log. Ops



	$OPq\ rA, rB$
Fetch	$icode:ifun \leftarrow M_1[PC]$
	$rA:rB \leftarrow M_1[PC+1]$
	$valP \leftarrow PC+2$

Read instruction byte

Read register byte

Compute next PC

Stage Computation: Arith/Log. Ops



	OPq rA, rB	
Fetch	icode:ifun \leftarrow M ₁ [PC]	Read instruction byte
	rA:rB \leftarrow M ₁ [PC+1]	Read register byte
	valP \leftarrow PC+2	Compute next PC
Decode	valA \leftarrow R[rA]	Read operand A
	valB \leftarrow R[rB]	Read operand B

Stage Computation: Arith/Log. Ops



	$OPq\ rA,\ rB$	
Fetch	$icode:ifun \leftarrow M_1[PC]$ $rA:rB \leftarrow M_1[PC+1]$	Read instruction byte Read register byte
	$valP \leftarrow PC+2$	Compute next PC
Decode	$valA \leftarrow R[rA]$ $valB \leftarrow R[rB]$	Read operand A Read operand B
	$valE \leftarrow valB\ OP\ valA$ Set CC	Perform ALU operation Set condition code register

Stage Computation: Arith/Log. Ops



	OPq rA, rB	
Fetch	icode:ifun $\leftarrow M_1[PC]$ rA:rB $\leftarrow M_1[PC+1]$	Read instruction byte Read register byte
	valP $\leftarrow PC+2$	Compute next PC
Decode	valA $\leftarrow R[rA]$ valB $\leftarrow R[rB]$	Read operand A Read operand B
	valE $\leftarrow valB \text{ OP } valA$ Set CC	Perform ALU operation Set condition code register
Memory		

Stage Computation: Arith/Log. Ops



	$OPq\ rA,\ rB$	
Fetch	$icode:ifun \leftarrow M_1[PC]$ $rA:rB \leftarrow M_1[PC+1]$	Read instruction byte Read register byte
	$valP \leftarrow PC+2$	Compute next PC
Decode	$valA \leftarrow R[rA]$	Read operand A
	$valB \leftarrow R[rB]$	Read operand B
Execute	$valE \leftarrow valB\ OP\ valA$	Perform ALU operation
	Set CC	Set condition code register
Memory		
Write back	$R[rB] \leftarrow valE$	Write back result

Stage Computation: Arith/Log. Ops



	OPq rA, rB	
Fetch	icode:ifun $\leftarrow M_1[PC]$ rA:rB $\leftarrow M_1[PC+1]$	Read instruction byte Read register byte
	valP $\leftarrow PC+2$	Compute next PC
Decode	valA $\leftarrow R[rA]$ valB $\leftarrow R[rB]$	Read operand A Read operand B
	valE $\leftarrow valB \text{ OP } valA$ Set CC	Perform ALU operation Set condition code register
Memory		
Write back	R[rB] $\leftarrow valE$	Write back result
PC update	PC $\leftarrow valP$	Update PC

Stage Computation: `rmmovq`

`rmmovq rA, D(rB)`

4	0	rA	rB
---	---	----	----

D

<code>rmmovq rA, D(rB)</code>

Stage Computation: `rmmovq`

`rmmovq rA, D(rB)`



	<code>rmmovq rA, D(rB)</code>
Fetch	<code>icode:ifun ← M₁[PC]</code> <code>rA:rB ← M₁[PC+1]</code> <code>valC ← M₈[PC+2]</code> <code>valP ← PC+10</code>

- Read instruction byte
- Read register byte
- Read displacement D
- Compute next PC

Stage Computation: `rmmovq`

`rmmovq rA, D(rB)`



<code>rmmovq rA, D(rB)</code>	
Fetch	<code>icode:ifun ← M₁[PC]</code>
	<code>rA:rB ← M₁[PC+1]</code>
	<code>valC ← M₈[PC+2]</code>
	<code>valP ← PC+10</code>
Decode	<code>valA ← R[rA]</code>
	<code>valB ← R[rB]</code>

Read instruction byte

Read register byte

Read displacement D

Compute next PC

Read operand A

Read operand B

Stage Computation: `rmmovq`

`rmmovq rA, D(rB)`



<code>rmmovq rA, D(rB)</code>	
Fetch	$icode:ifun \leftarrow M_1[PC]$ $rA:rB \leftarrow M_1[PC+1]$ $valC \leftarrow M_8[PC+2]$ $valP \leftarrow PC+10$
Decode	$valA \leftarrow R[rA]$ $valB \leftarrow R[rB]$
Execute	$valE \leftarrow valB + valC$

- Read instruction byte
- Read register byte
- Read displacement D
- Compute next PC
- Read operand A
- Read operand B
- Compute effective address

Stage Computation: `rmmovq`

`rmmovq rA, D(rB)`



<code>rmmovq rA, D(rB)</code>	
Fetch	$icode:ifun \leftarrow M_1[PC]$ $rA:rB \leftarrow M_1[PC+1]$ $valC \leftarrow M_8[PC+2]$ $valP \leftarrow PC+10$
Decode	$valA \leftarrow R[rA]$ $valB \leftarrow R[rB]$
Execute	$valE \leftarrow valB + valC$
Memory	$M_8[valE] \leftarrow valA$

- Read instruction byte
- Read register byte
- Read displacement D
- Compute next PC
- Read operand A
- Read operand B
- Compute effective address
- Write value to memory

Stage Computation: `rmmovq`

`rmmovq rA, D(rB)`



	<code>rmmovq rA, D(rB)</code>
Fetch	$icode:ifun \leftarrow M_1[PC]$ $rA:rB \leftarrow M_1[PC+1]$ $valC \leftarrow M_8[PC+2]$ $valP \leftarrow PC+10$
Decode	$valA \leftarrow R[rA]$ $valB \leftarrow R[rB]$
Execute	$valE \leftarrow valB + valC$
Memory	$M_8[valE] \leftarrow valA$
Write back	

- Read instruction byte
- Read register byte
- Read displacement D
- Compute next PC
- Read operand A
- Read operand B
- Compute effective address
- Write value to memory

Stage Computation: `rmmovq`

`rmmovq rA, D(rB)`



<code>rmmovq rA, D(rB)</code>	
Fetch	$icode:ifun \leftarrow M_1[PC]$ $rA:rB \leftarrow M_1[PC+1]$ $valC \leftarrow M_8[PC+2]$ $valP \leftarrow PC+10$
Decode	$valA \leftarrow R[rA]$ $valB \leftarrow R[rB]$
Execute	$valE \leftarrow valB + valC$
Memory	$M_8[valE] \leftarrow valA$
Write back	
PC update	$PC \leftarrow valP$

- Read instruction byte
- Read register byte
- Read displacement D
- Compute next PC
- Read operand A
- Read operand B
- Compute effective address
- Write value to memory
- Update PC

Stage Computation: Jumps

jXX Dest

- Compute both addresses
- Choose based on setting of condition codes and branch condition

Stage Computation: Jumps

	jXX Dest	
Fetch	$\text{icode:ifun} \leftarrow M_1[\text{PC}]$	Read instruction byte
	$\text{valC} \leftarrow M_8[\text{PC}+1]$	Read destination address
	$\text{valP} \leftarrow \text{PC}+9$	Fall through address

- Compute both addresses
- Choose based on setting of condition codes and branch condition

Stage Computation: Jumps

	jXX Dest	
Fetch	$\text{icode:ifun} \leftarrow M_1[\text{PC}]$	Read instruction byte
	$\text{valC} \leftarrow M_8[\text{PC}+1]$	Read destination address
	$\text{valP} \leftarrow \text{PC}+9$	Fall through address
Decode		

- Compute both addresses
- Choose based on setting of condition codes and branch condition

Stage Computation: Jumps

	jXX Dest	
Fetch	$\text{icode:ifun} \leftarrow M_1[\text{PC}]$	Read instruction byte
	$\text{valC} \leftarrow M_8[\text{PC}+1]$	Read destination address
	$\text{valP} \leftarrow \text{PC}+9$	Fall through address
Decode		
Execute	$\text{Cnd} \leftarrow \text{Cond}(\text{CC}, \text{ifun})$	Take branch?

- Compute both addresses
- Choose based on setting of condition codes and branch condition

Stage Computation: Jumps

	jXX Dest	
Fetch	$\text{icode:ifun} \leftarrow M_1[\text{PC}]$	Read instruction byte
	$\text{valC} \leftarrow M_8[\text{PC}+1]$	Read destination address
	$\text{valP} \leftarrow \text{PC}+9$	Fall through address
Decode		
Execute	$\text{Cnd} \leftarrow \text{Cond}(\text{CC}, \text{ifun})$	Take branch?
Memory		

- Compute both addresses
- Choose based on setting of condition codes and branch condition

Stage Computation: Jumps

	jXX Dest	
Fetch	$\text{icode:ifun} \leftarrow M_1[\text{PC}]$	Read instruction byte
	$\text{valC} \leftarrow M_8[\text{PC}+1]$	Read destination address
	$\text{valP} \leftarrow \text{PC}+9$	Fall through address
Decode		
Execute	$\text{Cnd} \leftarrow \text{Cond}(\text{CC}, \text{ifun})$	Take branch?
Memory		
Write back		

- Compute both addresses
- Choose based on setting of condition codes and branch condition

Stage Computation: Jumps

	jXX Dest	
Fetch	$\text{icode:ifun} \leftarrow M_1[\text{PC}]$	Read instruction byte
	$\text{valC} \leftarrow M_8[\text{PC}+1]$	Read destination address
	$\text{valP} \leftarrow \text{PC}+9$	Fall through address
Decode		
Execute	$\text{Cnd} \leftarrow \text{Cond}(\text{CC}, \text{ifun})$	Take branch?
Memory		
Write back		
PC update	$\text{PC} \leftarrow \text{Cnd} ? \text{valC} : \text{valP}$	Update PC

- Compute both addresses
- Choose based on setting of condition codes and branch condition

Pipeline Stages

Fetch

- Select current PC
- Read instruction
- Compute incremented PC

Decode

- Read program registers

Execute

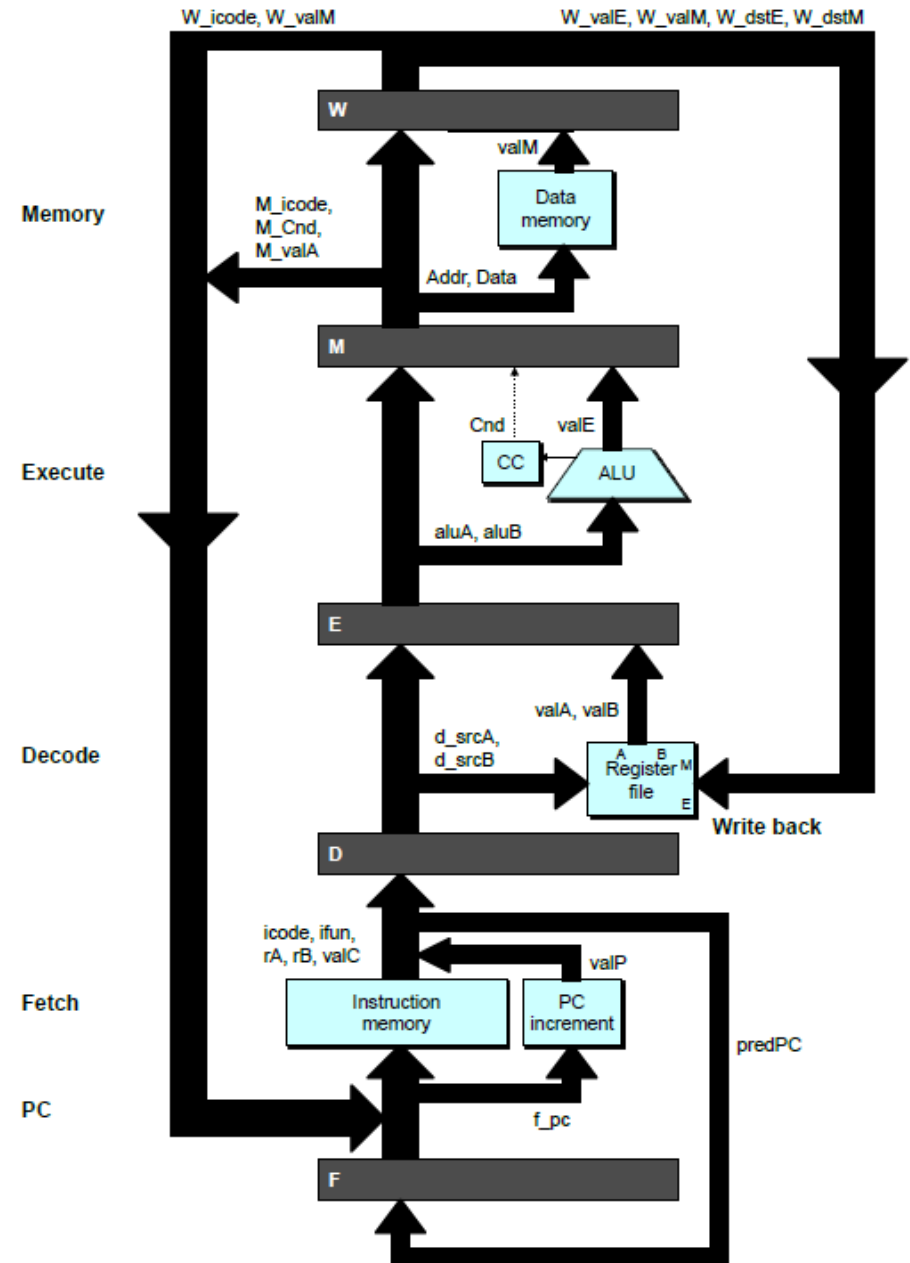
- Operate ALU

Memory

- Read or write data memory

Write Back

- Update register file



Real-World Pipelines: Car Washes

Real-World Pipelines: Car Washes

Sequential



Real-World Pipelines: Car Washes

Sequential



Pipelined



Real-World Pipelines: Car Washes

Sequential



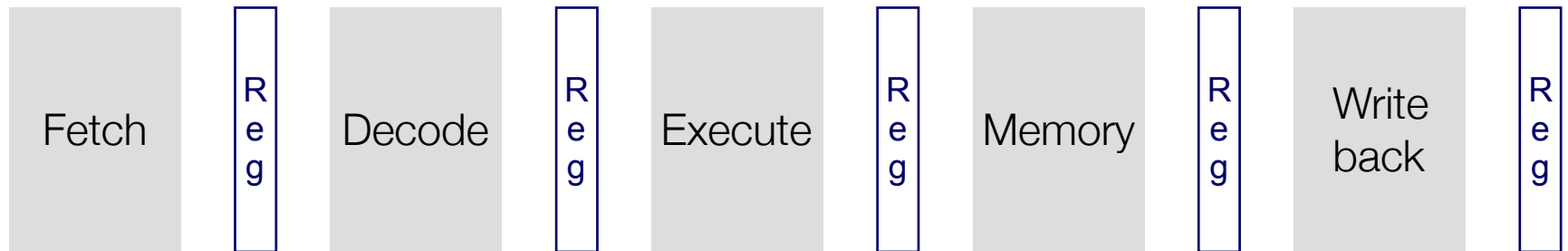
Pipelined



Idea

- Divide process into independent stages
- Move objects through stages in sequence
- At any given times, multiple objects being processed

Pipeline Illustration

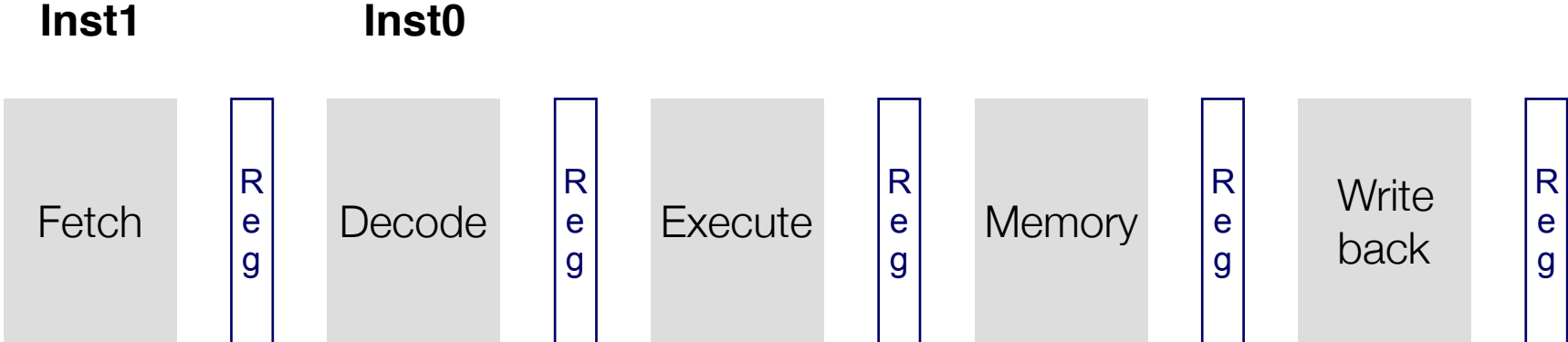


Pipeline Illustration

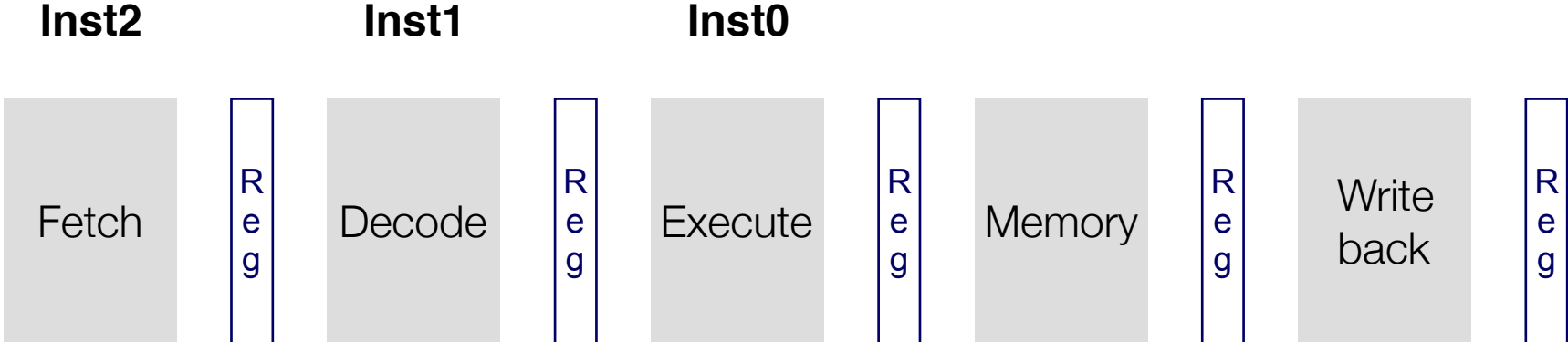
Inst0



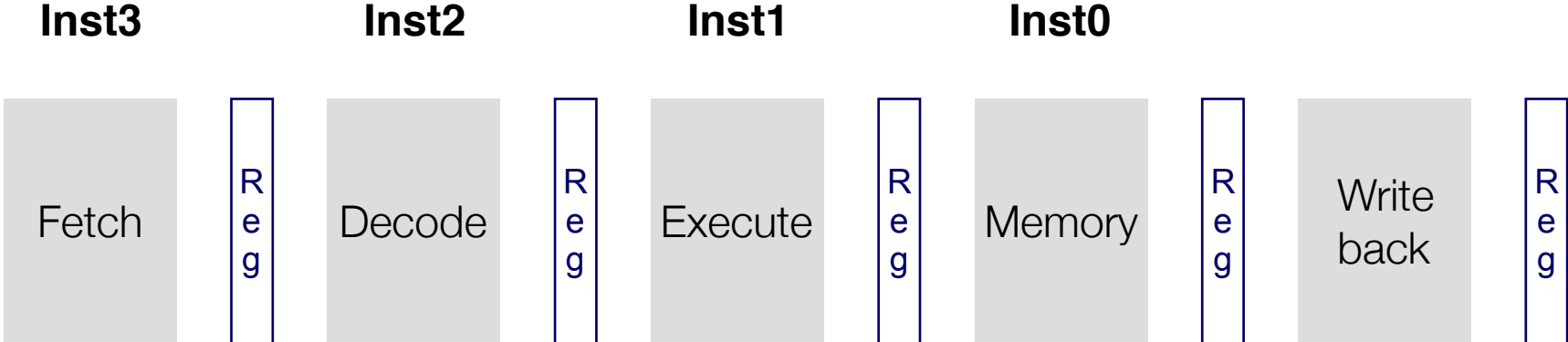
Pipeline Illustration



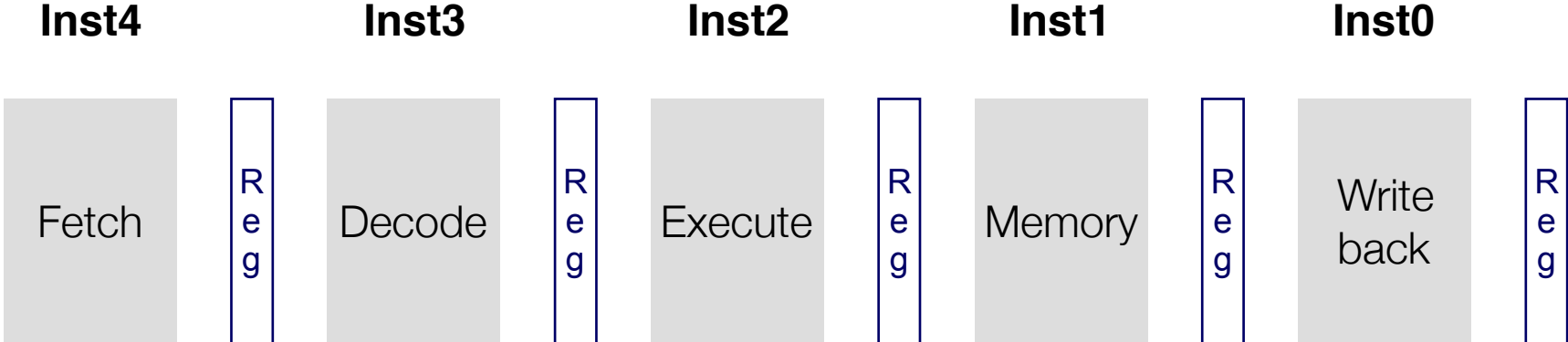
Pipeline Illustration



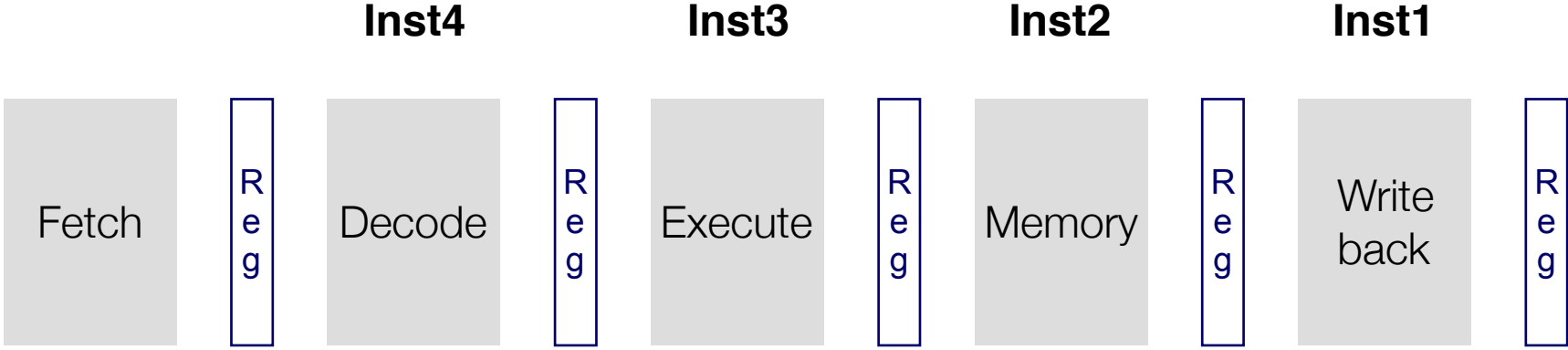
Pipeline Illustration



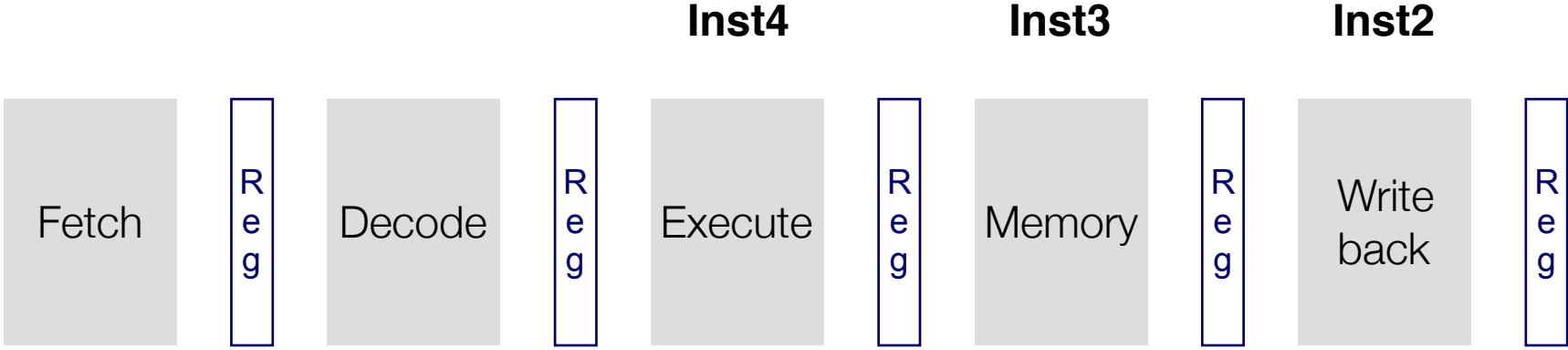
Pipeline Illustration



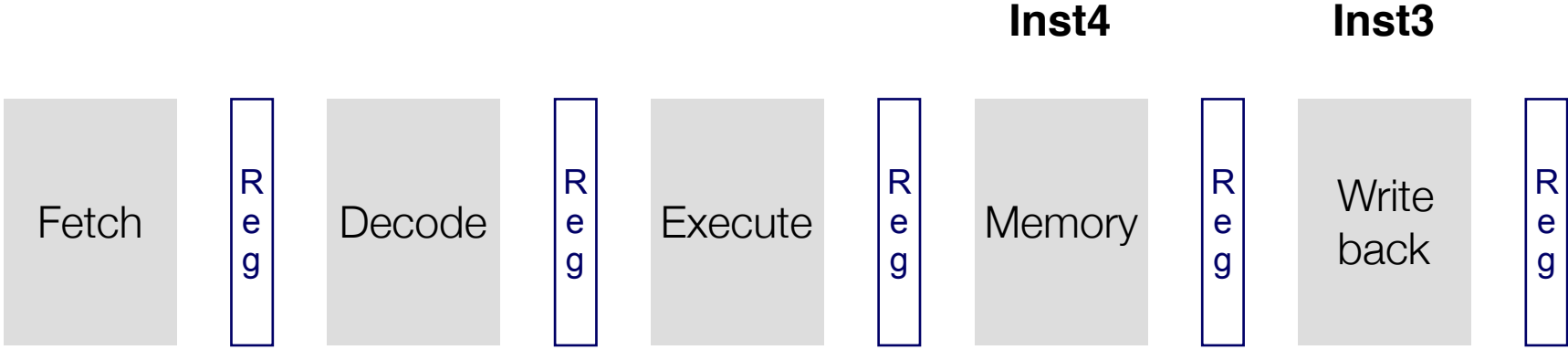
Pipeline Illustration



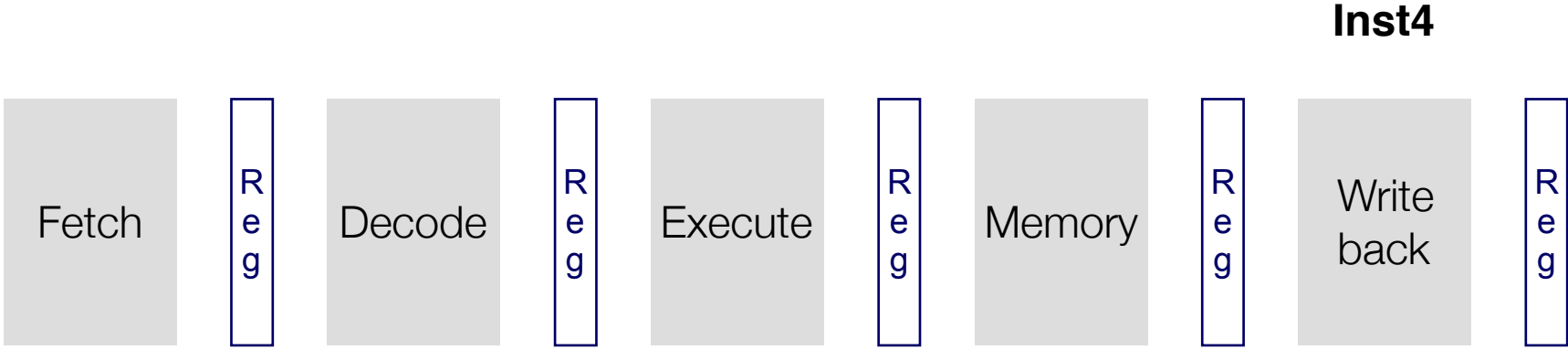
Pipeline Illustration



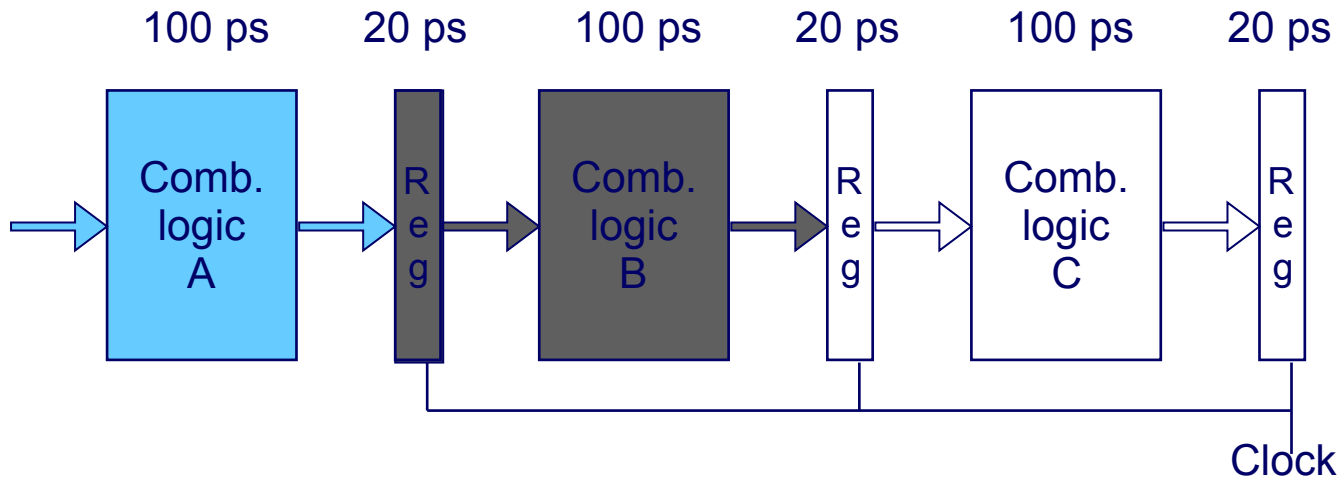
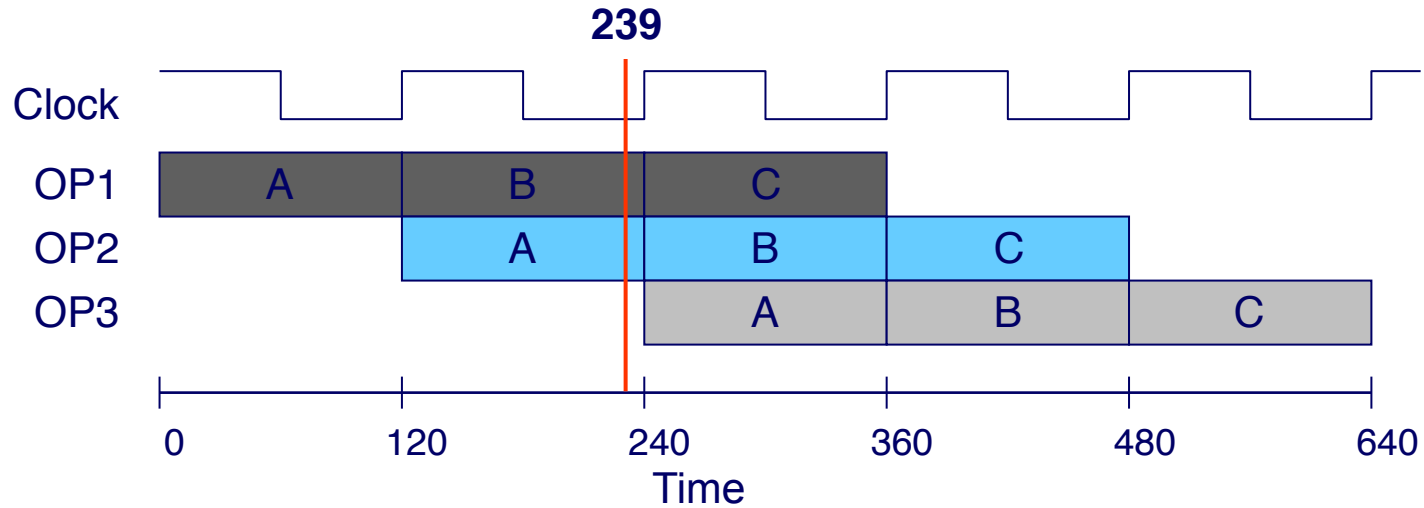
Pipeline Illustration



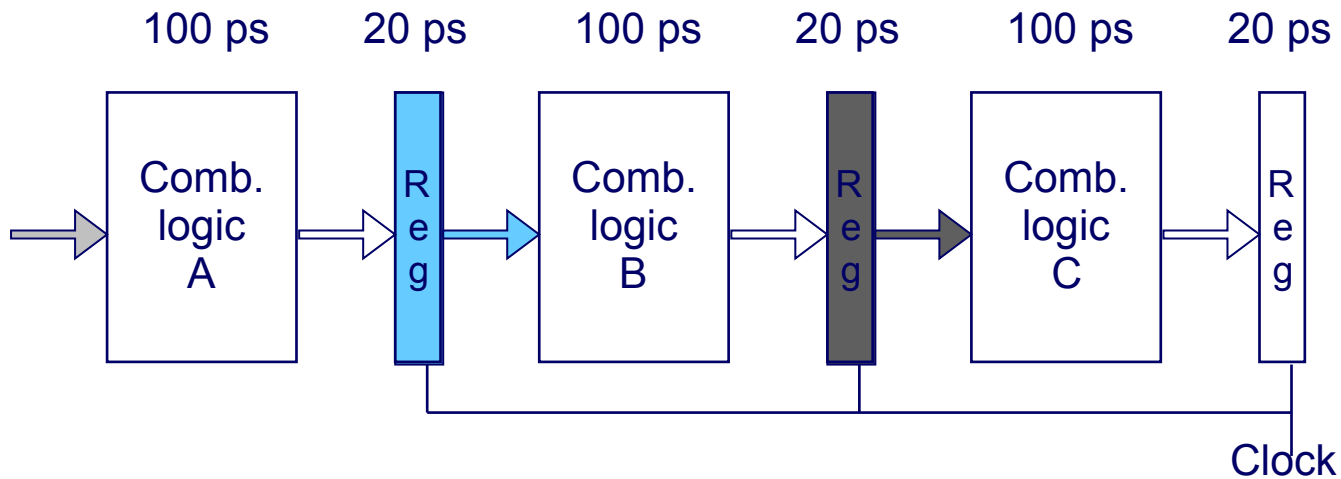
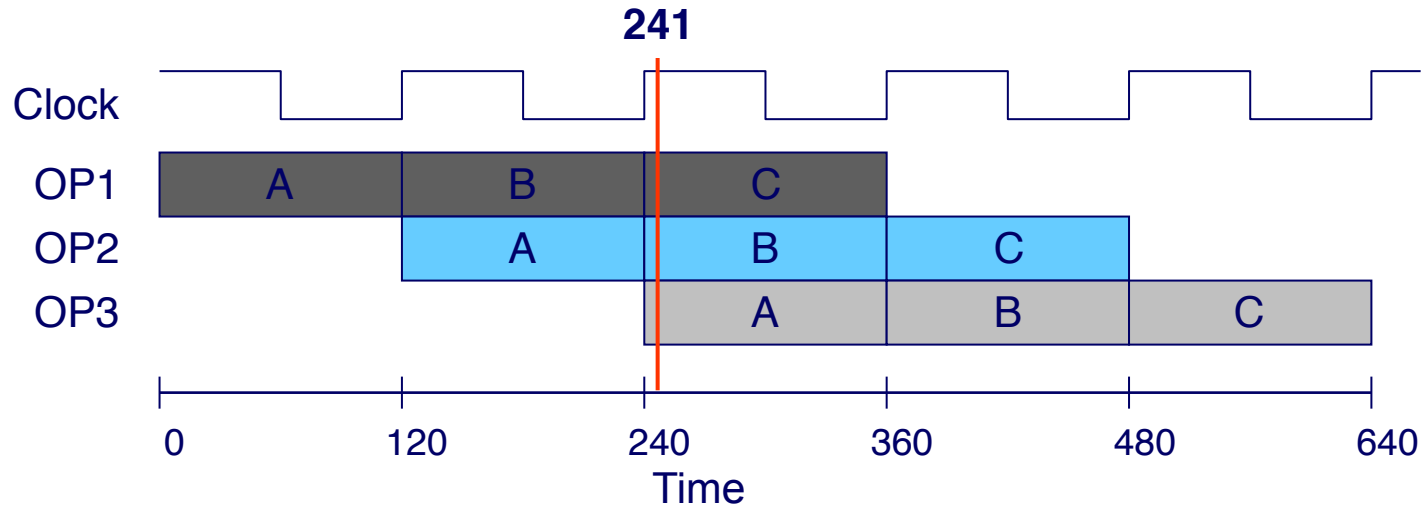
Pipeline Illustration



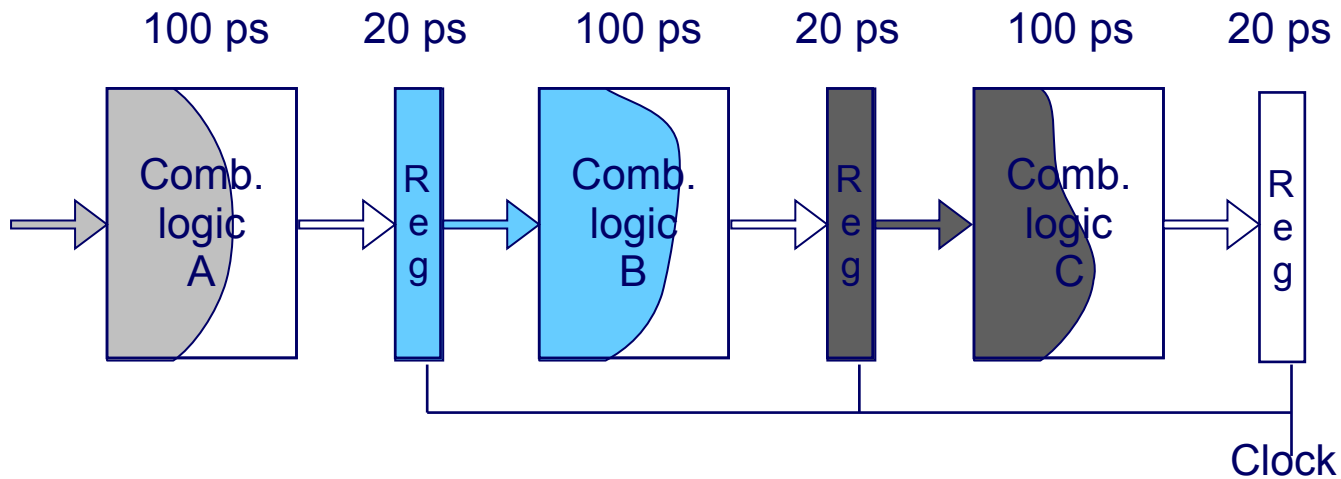
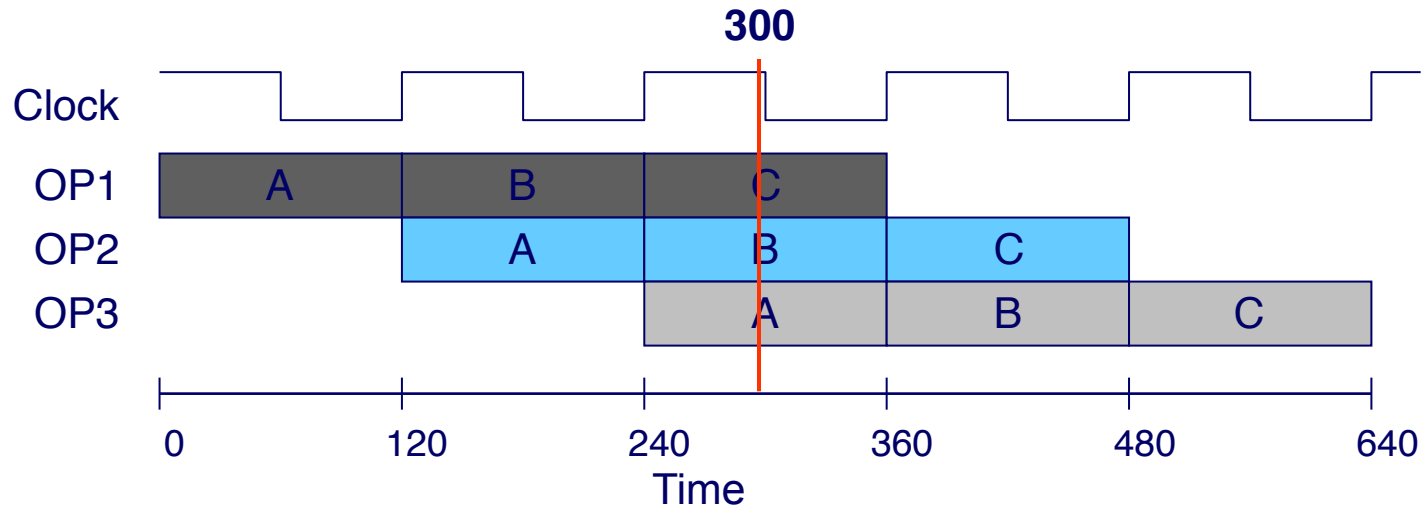
Another Illustration



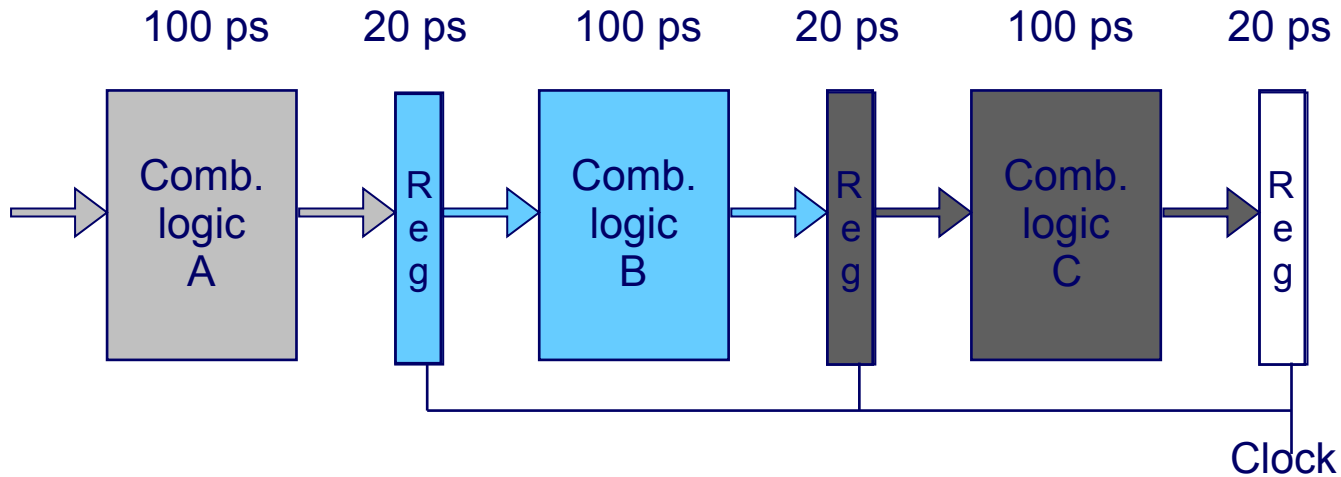
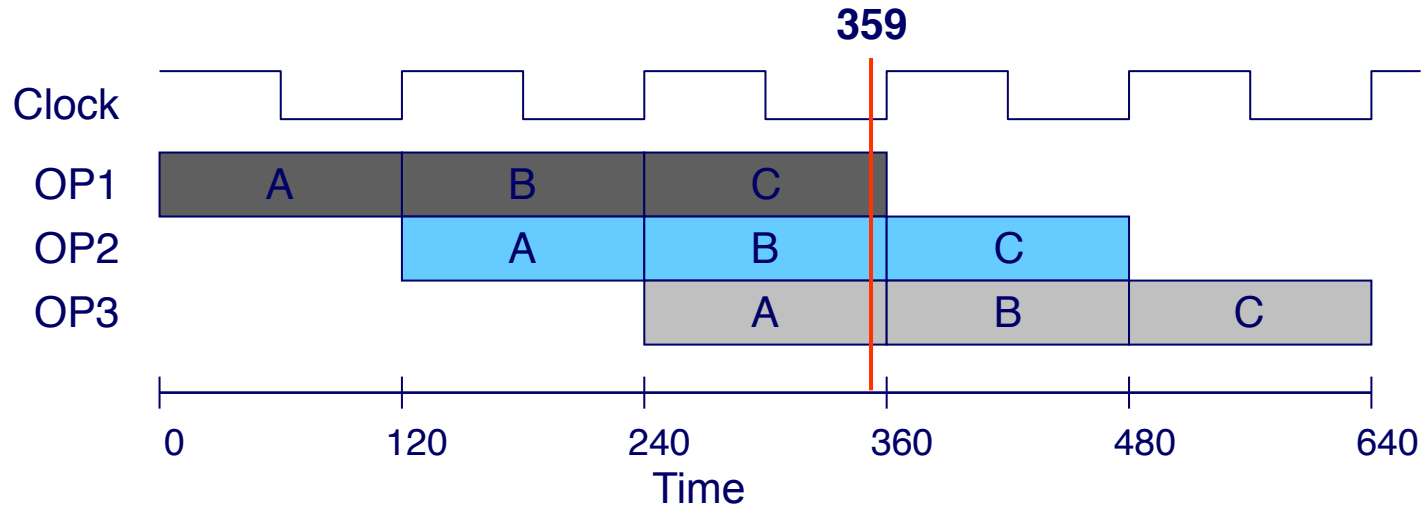
Another Illustration



Another Illustration



Another Illustration



Making the Pipeline Really Work

- Control Dependencies
 - What is it?
 - Software mitigation: Inserting Nops
 - Software mitigation: Delay Slots
- Data Dependencies
 - What is it?
 - Software mitigation: Inserting Nops

Control Dependency

- **Definition:** Outcome of instruction A determines whether or not instruction B should be executed or not.
- Jump instruction example below:
 - `jne L1` determines whether `irmovq $1, %rax` should be executed
 - But `jne` doesn't know its outcome until after its Execute stage

```
xorg %rax, %rax
jne L1           # Not taken
irmovq $1, %rax # Fall Through
L1  irmovq $4, %rcx # Target
    irmovq $3, %rax # Target + 1
```

Control Dependency

- **Definition:** Outcome of instruction A determines whether or not instruction B should be executed or not.
- Jump instruction example below:
 - `jne L1` determines whether `irmovq $1, %rax` should be executed
 - But `jne` doesn't know its outcome until after its Execute stage

```
xorg %rax, %rax
jne L1 # Not taken
irmovq $1, %rax # Fall Through
L1 irmovq $4, %rcx # Target
irmovq $3, %rax # Target + 1
```

1

F

Control Dependency

- **Definition:** Outcome of instruction A determines whether or not instruction B should be executed or not.
- Jump instruction example below:
 - `jne L1` determines whether `irmovq $1, %rax` should be executed
 - But `jne` doesn't know its outcome until after its Execute stage

```
xorg %rax, %rax
jne L1          # Not taken
irmovq $1, %rax # Fall Through
L1: irmovq $4, %rcx # Target
    irmovq $3, %rax # Target + 1
```

	1	2
	F	D
		F

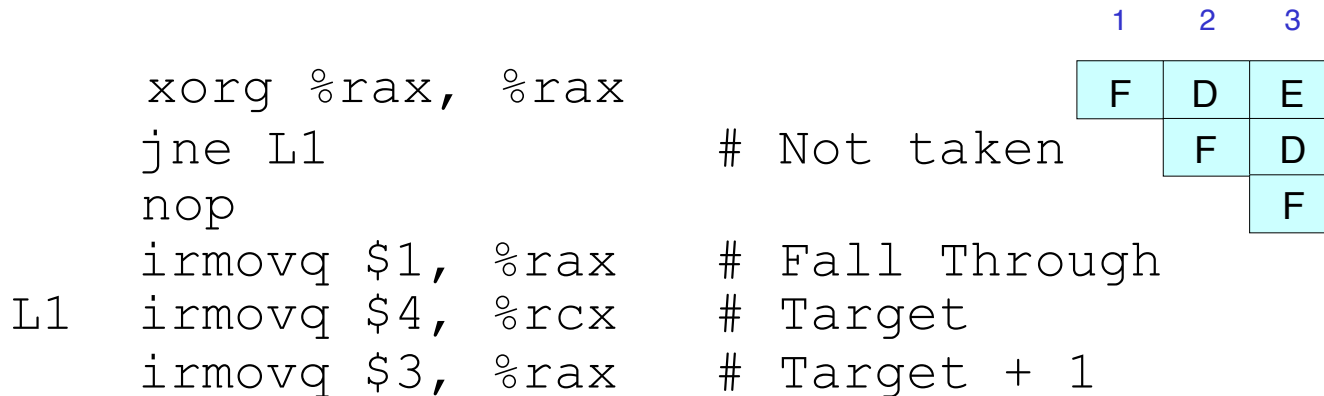
Control Dependency

- **Definition:** Outcome of instruction A determines whether or not instruction B should be executed or not.
- Jump instruction example below:
 - `jne L1` determines whether `irmovq $1, %rax` should be executed
 - But `jne` doesn't know its outcome until after its Execute stage

			1	2	3
	<code>xorg %rax, %rax</code>		F	D	E
	<code>jne L1</code>	# Not taken		F	D
	<code>irmovq \$1, %rax</code>	# Fall Through			
L1	<code>irmovq \$4, %rcx</code>	# Target			
	<code>irmovq \$3, %rax</code>	# Target + 1			

Control Dependency

- **Definition:** Outcome of instruction A determines whether or not instruction B should be executed or not.
- Jump instruction example below:
 - `jne L1` determines whether `irmovq $1, %rax` should be executed
 - But `jne` doesn't know its outcome until after its Execute stage



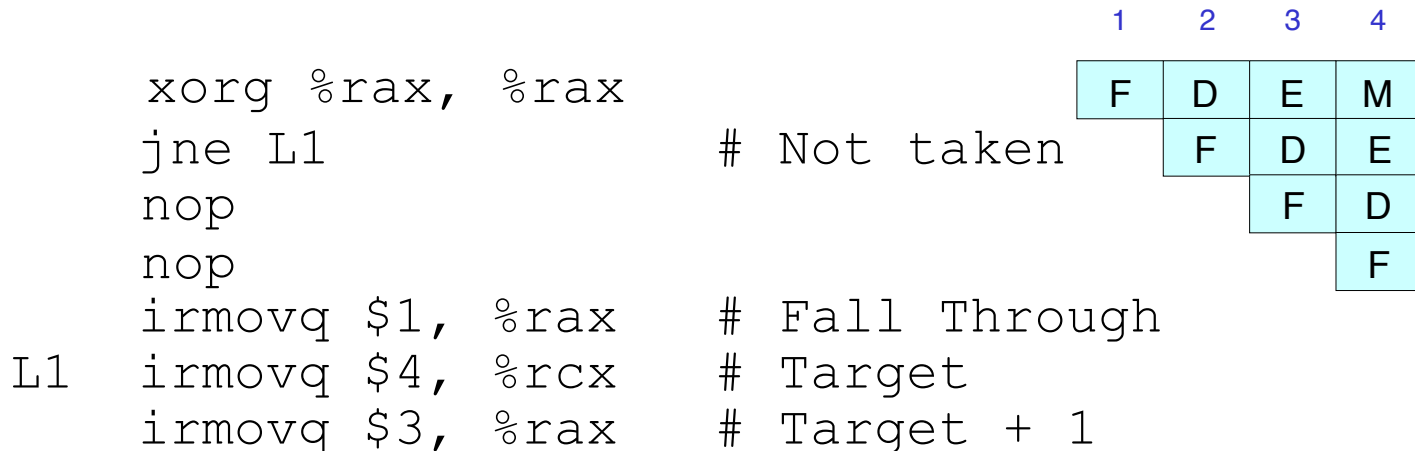
Control Dependency

- **Definition:** Outcome of instruction A determines whether or not instruction B should be executed or not.
- Jump instruction example below:
 - `jne L1` determines whether `irmovq $1, %rax` should be executed
 - But `jne` doesn't know its outcome until after its Execute stage

			1	2	3	4
	<code>xorg %rax, %rax</code>		F	D	E	M
	<code>jne L1</code>	# Not taken		F	D	E
	<code>nop</code>				F	D
	<code>irmovq \$1, %rax</code>	# Fall Through				
L1	<code>irmovq \$4, %rcx</code>	# Target				
	<code>irmovq \$3, %rax</code>	# Target + 1				

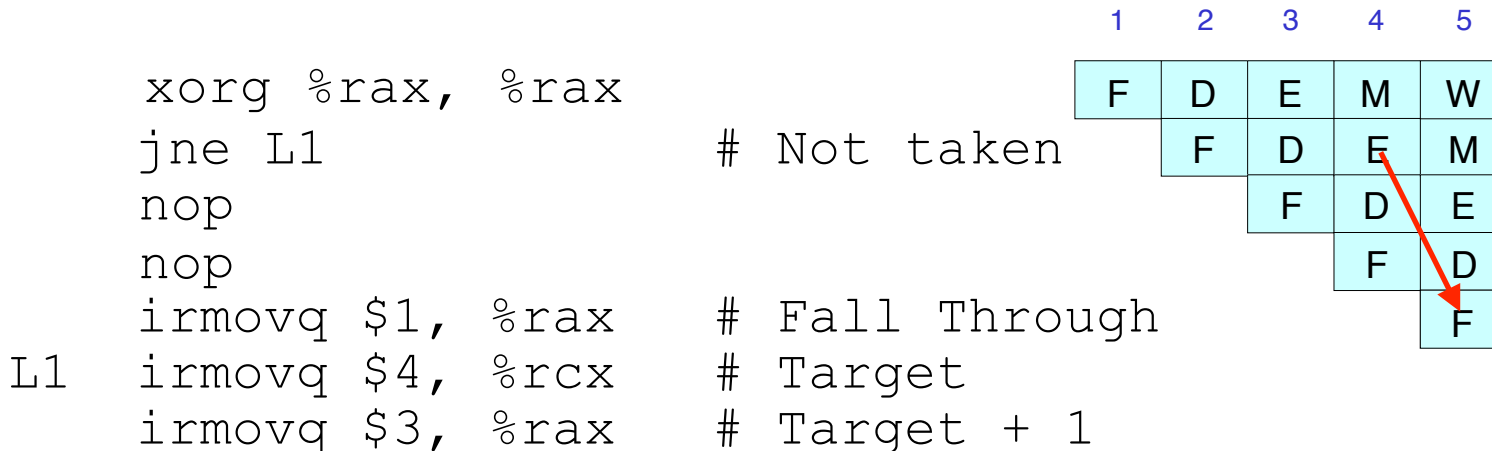
Control Dependency

- **Definition:** Outcome of instruction A determines whether or not instruction B should be executed or not.
- Jump instruction example below:
 - `jne L1` determines whether `irmovq $1, %rax` should be executed
 - But `jne` doesn't know its outcome until after its Execute stage



Control Dependency

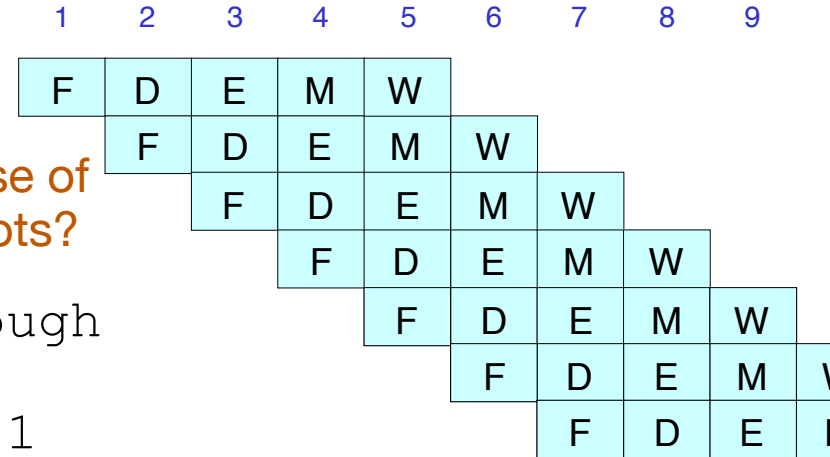
- **Definition:** Outcome of instruction A determines whether or not instruction B should be executed or not.
- Jump instruction example below:
 - `jne L1` determines whether `irmovq $1, %rax` should be executed
 - But `jne` doesn't know its outcome until after its Execute stage



Delay Slots

```
xorg %rax, %rax
jne L1
nop
nop
L1: irmovq $1, %rax    # Fall Through
    irmovq $4, %rcx  # Target
    irmovq $3, %rax  # Target + 1
```

Can we make use of the 2 wasted slots?



Delay Slots

```

    xorg %rax, %rax
    jne L1
    nop
    nop
    irmovq $1, %rax
L1:  irmovq $4, %rcx
    irmovq $3, %rax

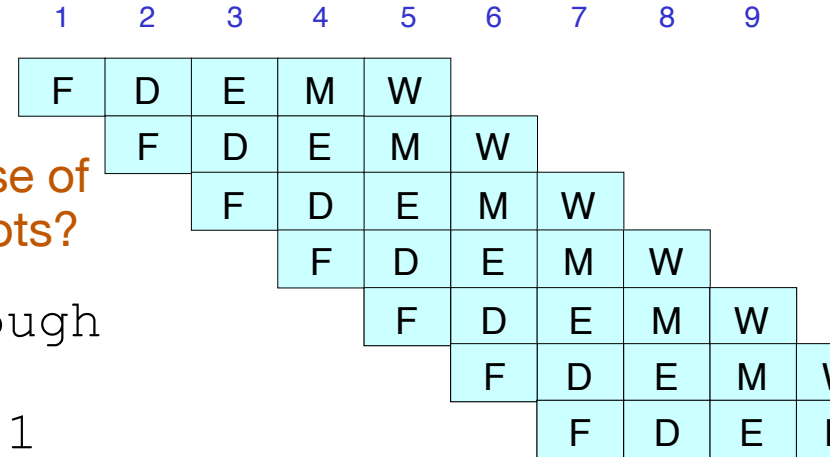
```

Can we make use of the 2 wasted slots?

```

# Fall Through
# Target
# Target + 1

```



```

if (cond) {
    do_A();
} else {
    do_B();
}
do_C();

```

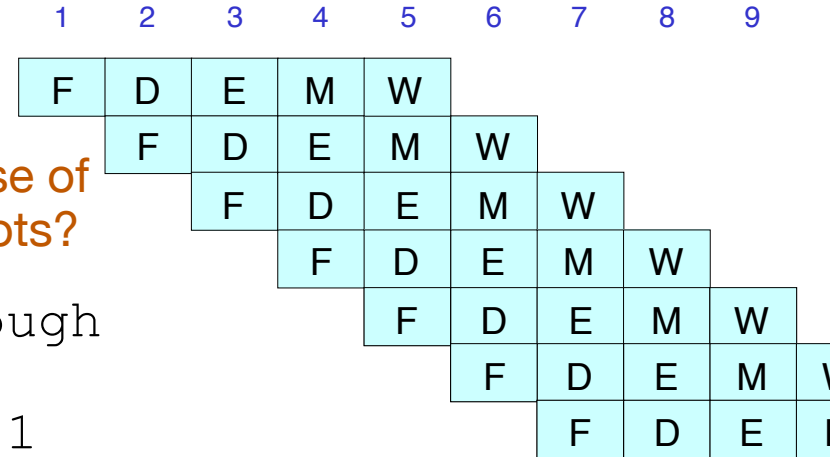
Delay Slots

```

    xorg %rax, %rax
    jne L1
    nop
    nop
    irmovq $1, %rax    # Fall Through
L1:  irmovq $4, %rcx    # Target
     irmovq $3, %rax    # Target + 1

```

Can we make use of the 2 wasted slots?



Have to make sure `do_C` doesn't depend on `do_A` and `do_B`!!!

```

if (cond) {
    do_A();
} else {
    do_B();
}

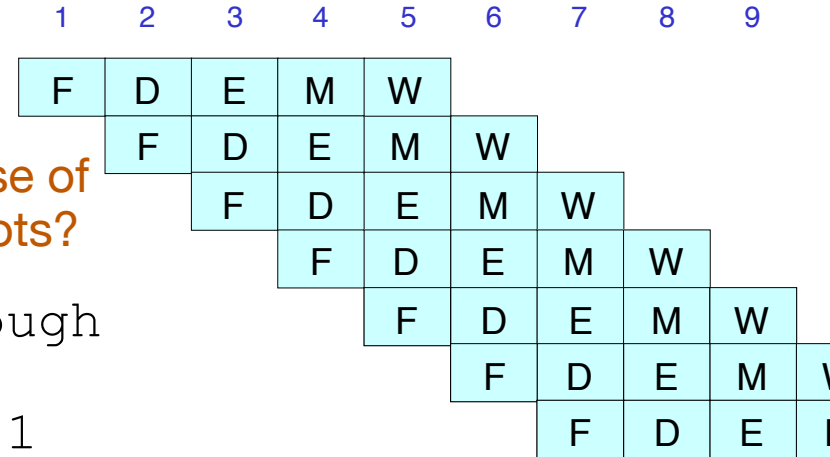
do_C();

```

Delay Slots

```
xorg %rax, %rax
jne L1
nop
nop
L1: irmovq $1, %rax # Fall Through
    irmovq $4, %rcx # Target
    irmovq $3, %rax # Target + 1
```

Can we make use of the 2 wasted slots?



A less obvious example

```
do_C();
if (cond) {
    do_A();
} else {
    do_B();
}
```

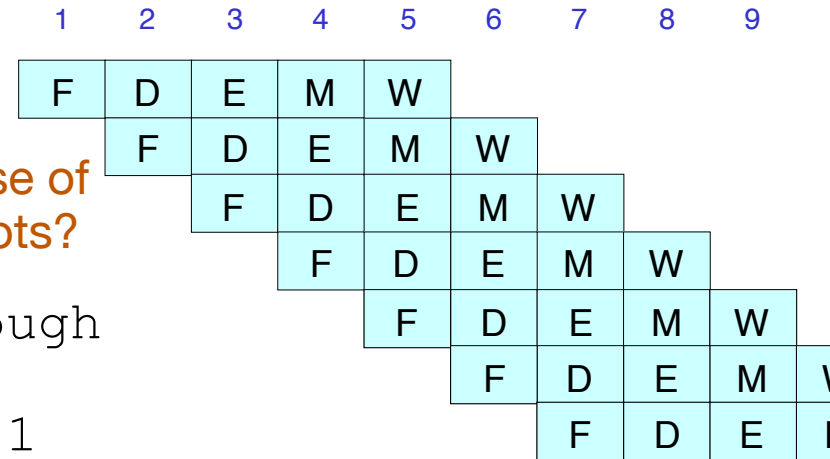
Delay Slots

```

    xorg %rax, %rax
    jne L1
    nop
    nop
    irmovq $1, %rax    # Fall Through
L1:  irmovq $4, %rcx    # Target
     irmovq $3, %rax   # Target + 1

```

Can we make use of the 2 wasted slots?



A less obvious example

```

do_C();
if (cond) {
    do_A();
} else {
    do_B();
}

```

```

add A, B
or C, D
sub E, F
jle 0x200
add A, C

```

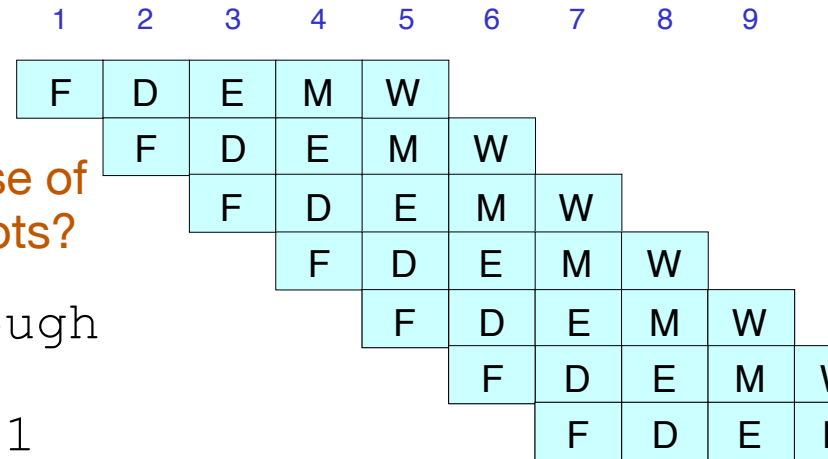
Delay Slots

```

    xorg %rax, %rax
    jne L1
    nop
    nop
    irmovq $1, %rax    # Fall Through
L1:  irmovq $4, %rcx    # Target
     irmovq $3, %rax    # Target + 1

```

Can we make use of the 2 wasted slots?



A less obvious example

```

do_C();
if (cond) {
    do_A();
} else {
    do_B();
}

```

```

add A, B
or C, D
sub E, F
jle 0x200
add A, C

```

```

add A, B
sub E, F
jle 0x200
or C, D
add A, C

```

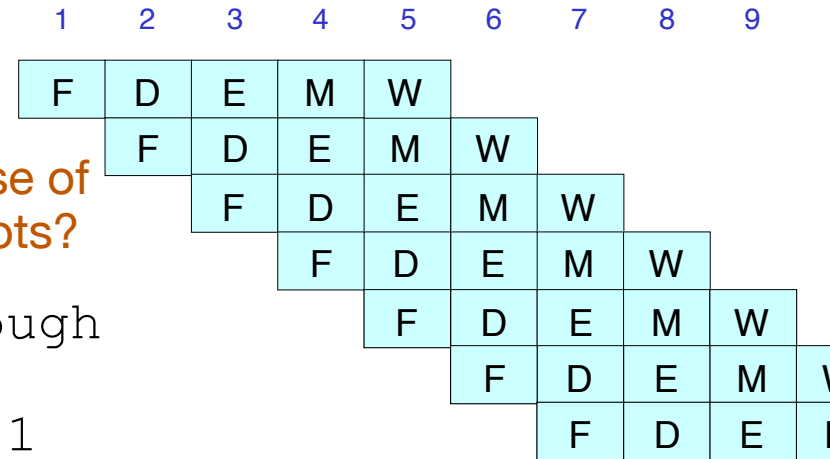
A green arrow points from the **or C, D** instruction in the second code block to the **or C, D** instruction in the first code block, indicating that the branch delay slot is filled with the branch target instruction.

Delay Slots

```

xorg %rax, %rax
jne L1
nop
nop
L1: irmovq $1, %rax      # Fall Through
    irmovq $4, %rcx    # Target
    irmovq $3, %rax    # Target + 1

```



Can we make use of the 2 wasted slots?

A less obvious example

```

do_C();
if (cond) {
    do_A();
} else {
    do_B();
}

```

```

add A, B
or C, D
sub E, F
jle 0x200
add A, C

```

```

add A, B
sub E, F
jle 0x200
or C, D
add A, C

```

Why don't we move the sub instruction?

Resolving Control Dependencies

- **Software Mechanisms**

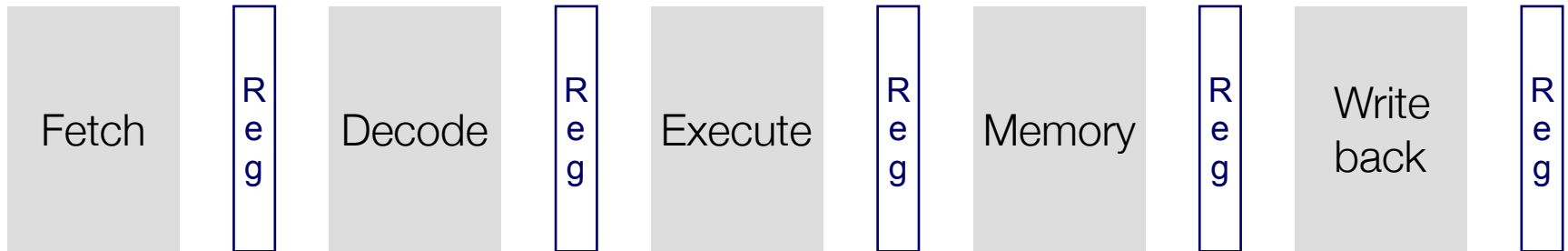
- Adding NOPs: requires compiler to insert nops, which also take memory space — not a good idea
- Delay slot: insert instructions that do not depend on the effect of the preceding instruction. These instructions will execute even if the preceding branch is taken — old RISC approach

- **Hardware mechanisms**

- Stalling (Think of it as hardware automatically inserting nops)
- Branch Prediction
- Return Address Stack

Hardware Generated Nops (Bubble and Stalling)

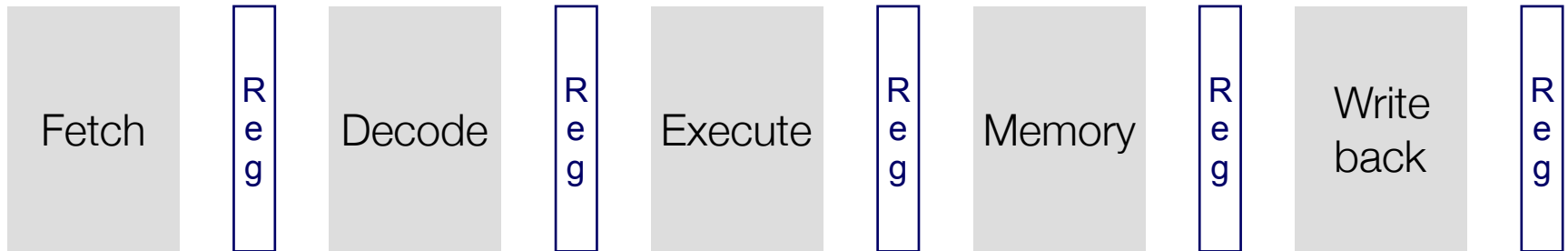
- **Stall**: the pipeline register shouldn't be written
- **Bubble**: signals correspond to a nop
- Why is it good for the hardware to do so anyways?



Hardware Generated Nops (Bubble and Stalling)

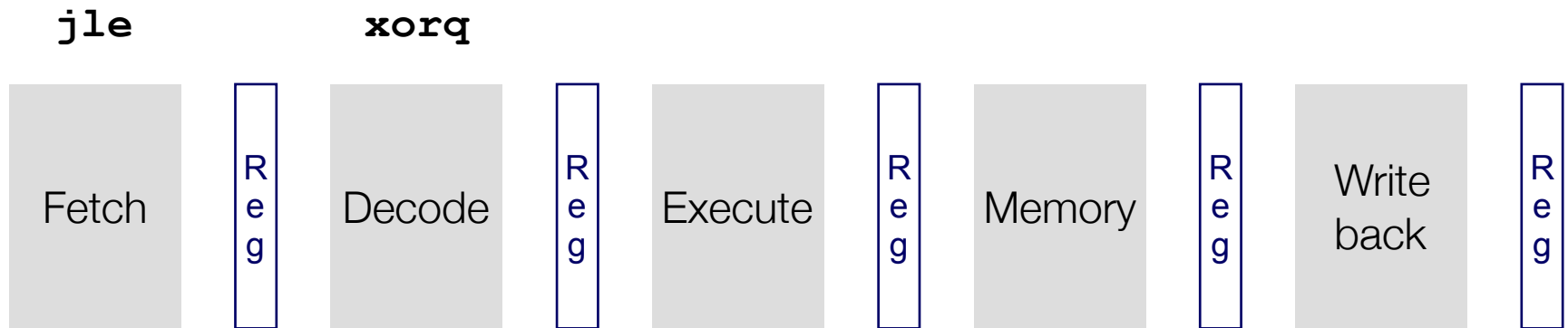
- **Stall:** the pipeline register shouldn't be written
- **Bubble:** signals correspond to a nop
- Why is it good for the hardware to do so anyways?

`xorq`



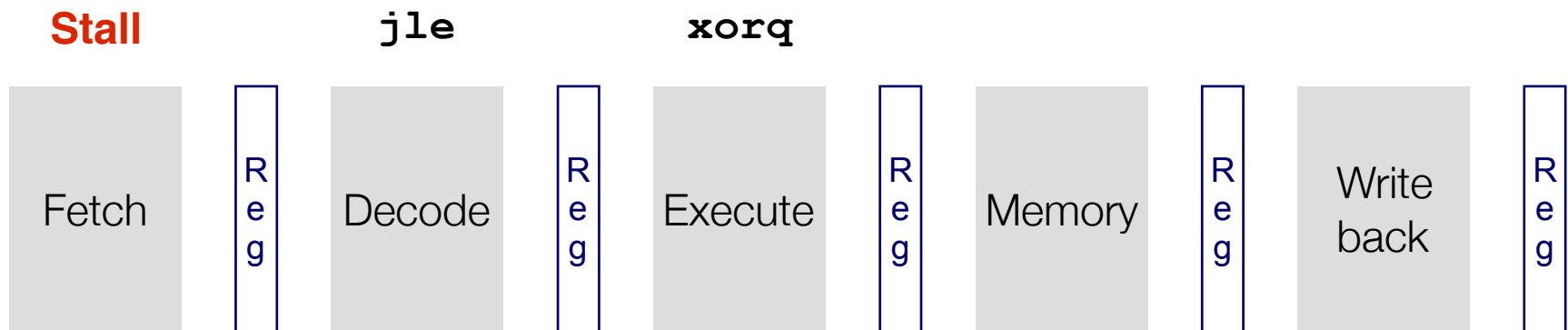
Hardware Generated Nops (Bubble and Stalling)

- **Stall:** the pipeline register shouldn't be written
- **Bubble:** signals correspond to a nop
- Why is it good for the hardware to do so anyways?



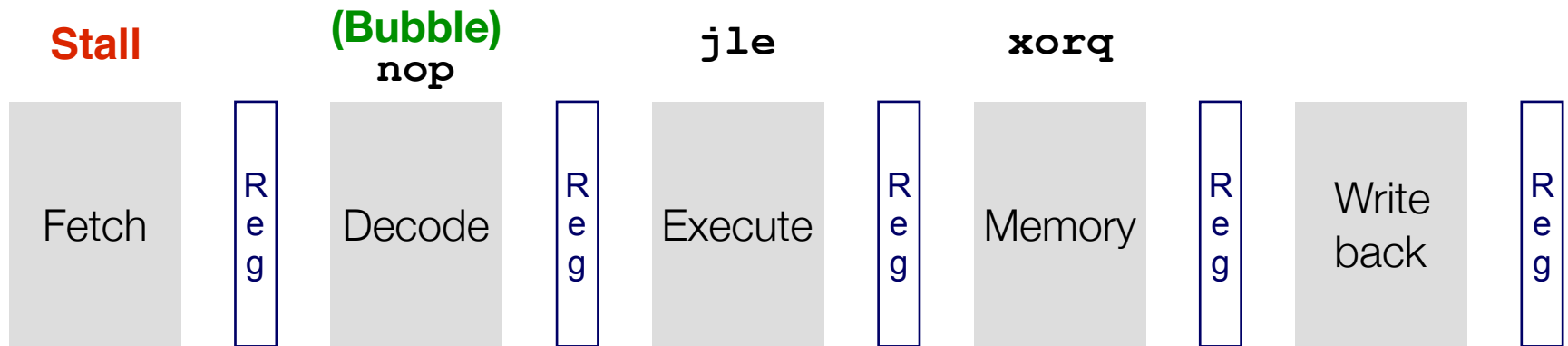
Hardware Generated Nops (Bubble and Stalling)

- **Stall**: the pipeline register shouldn't be written
- **Bubble**: signals correspond to a nop
- Why is it good for the hardware to do so anyways?



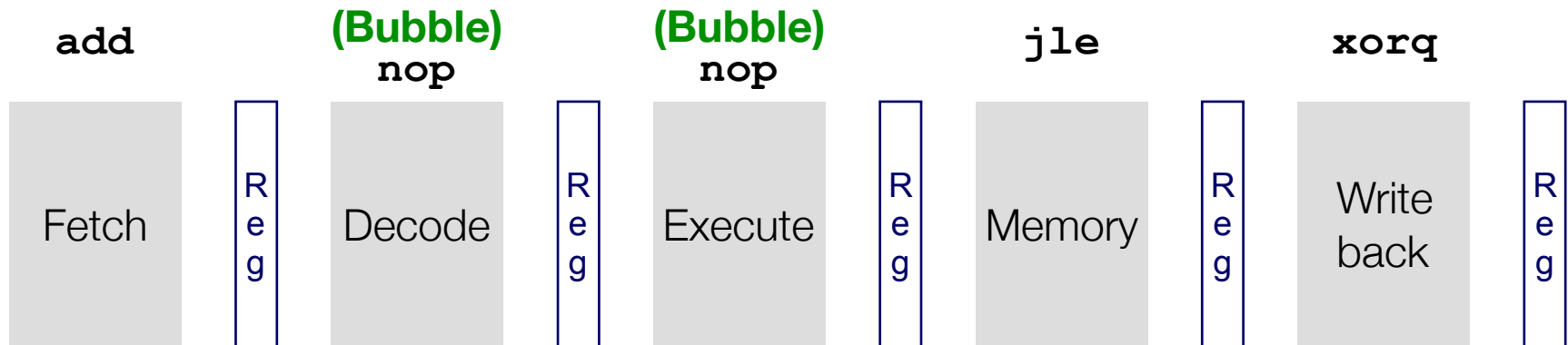
Hardware Generated Nops (Bubble and Stalling)

- **Stall**: the pipeline register shouldn't be written
- **Bubble**: signals correspond to a nop
- Why is it good for the hardware to do so anyways?



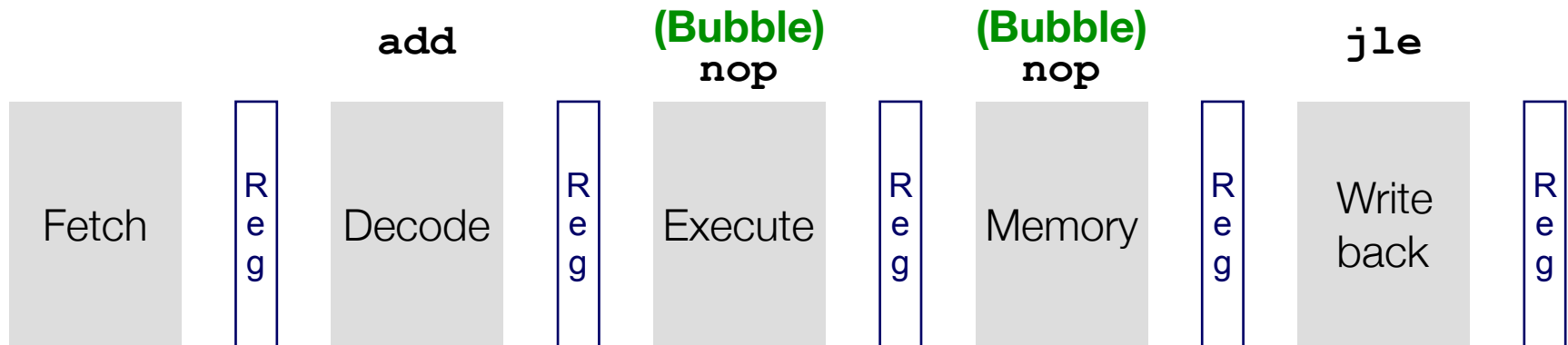
Hardware Generated Nops (Bubble and Stalling)

- **Stall:** the pipeline register shouldn't be written
- **Bubble:** signals correspond to a nop
- Why is it good for the hardware to do so anyways?



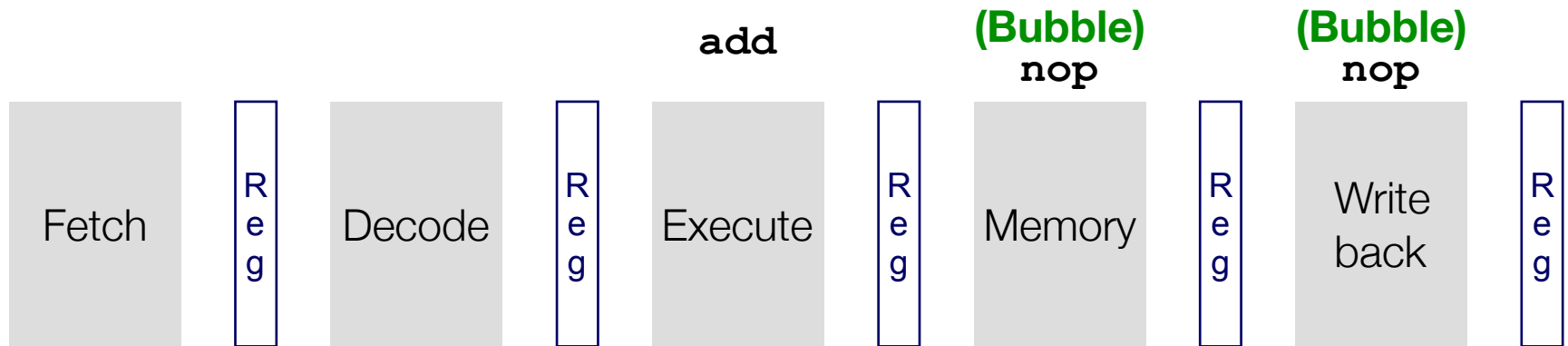
Hardware Generated Nops (Bubble and Stalling)

- **Stall:** the pipeline register shouldn't be written
- **Bubble:** signals correspond to a nop
- Why is it good for the hardware to do so anyways?



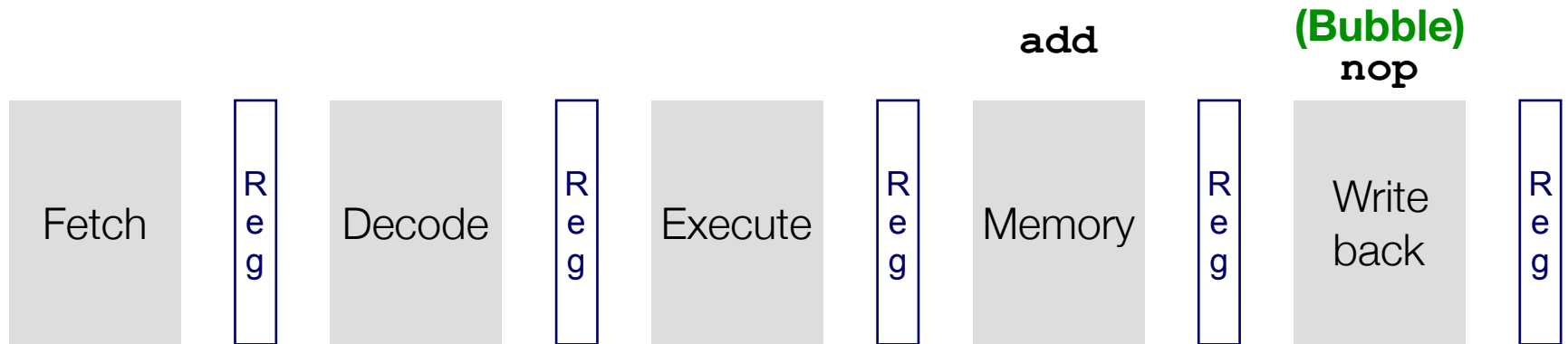
Hardware Generated Nops (Bubble and Stalling)

- **Stall:** the pipeline register shouldn't be written
- **Bubble:** signals correspond to a nop
- Why is it good for the hardware to do so anyways?



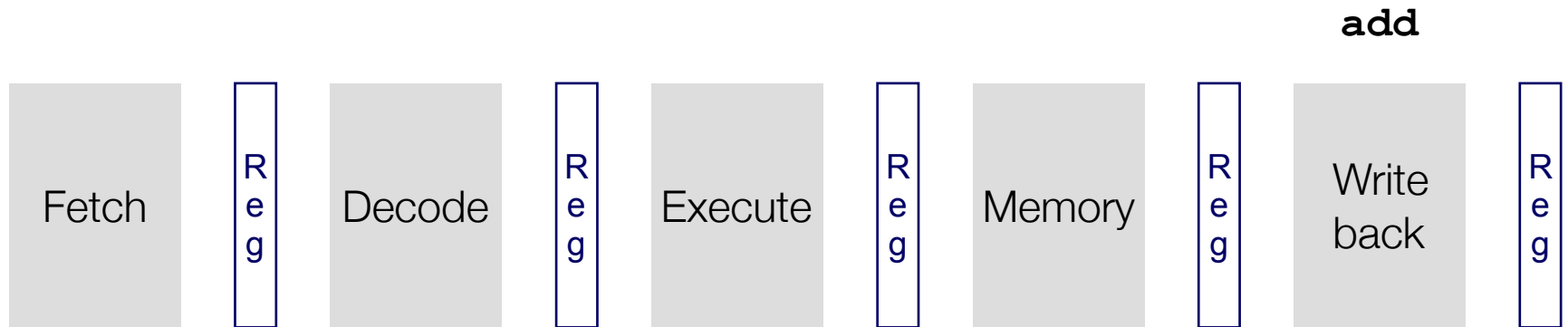
Hardware Generated Nops (Bubble and Stalling)

- **Stall:** the pipeline register shouldn't be written
- **Bubble:** signals correspond to a nop
- Why is it good for the hardware to do so anyways?



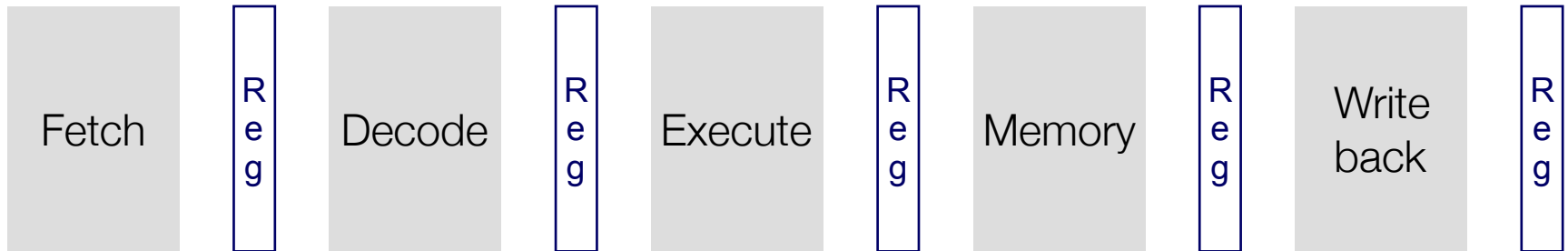
Hardware Generated Nops (Bubble and Stalling)

- **Stall:** the pipeline register shouldn't be written
- **Bubble:** signals correspond to a nop
- Why is it good for the hardware to do so anyways?



Hardware Generated Nops (Bubble and Stalling)

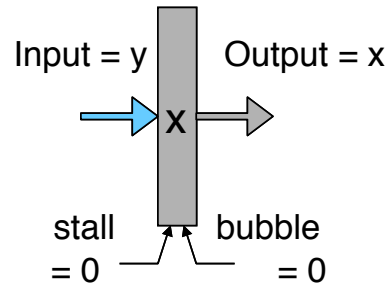
- **Stall**: the pipeline register shouldn't be written
- **Bubble**: signals correspond to a nop
- Why is it good for the hardware to do so anyways?



How are Stall and Bubble Implemented in Hardware?

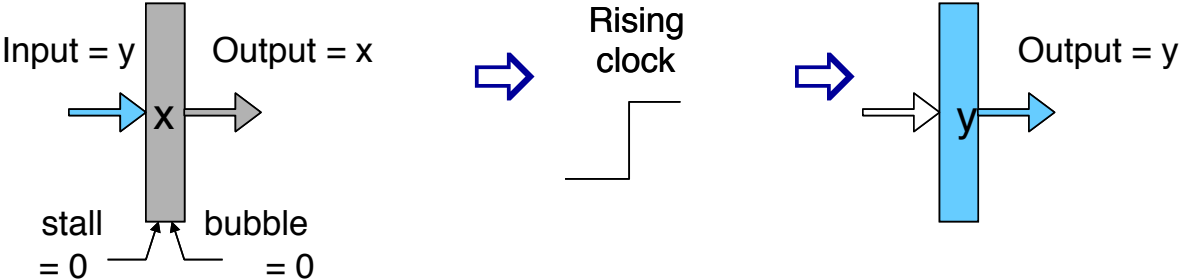
How are Stall and Bubble Implemented in Hardware?

Normal



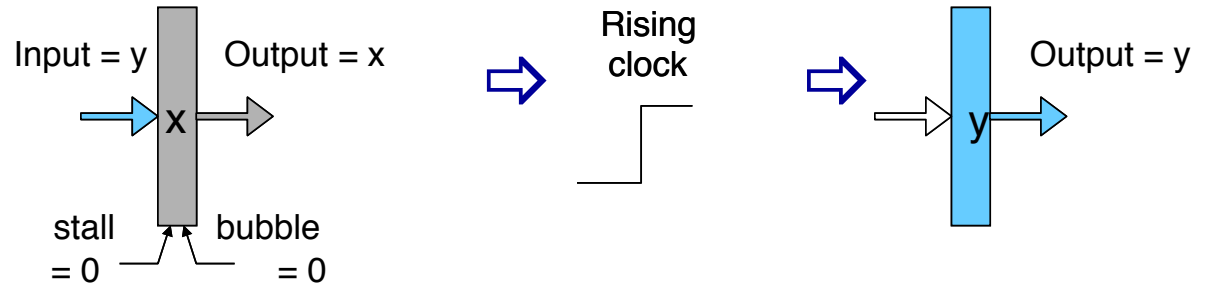
How are Stall and Bubble Implemented in Hardware?

Normal

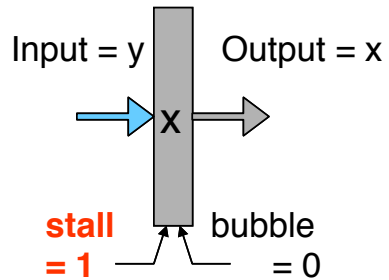


How are Stall and Bubble Implemented in Hardware?

Normal

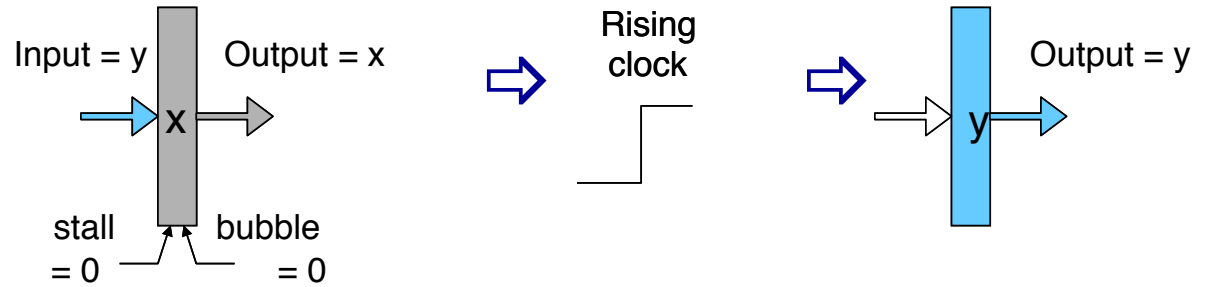


Stall

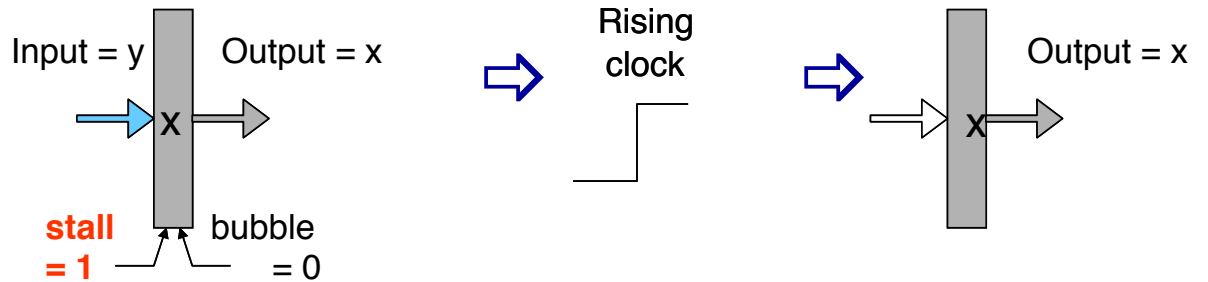


How are Stall and Bubble Implemented in Hardware?

Normal

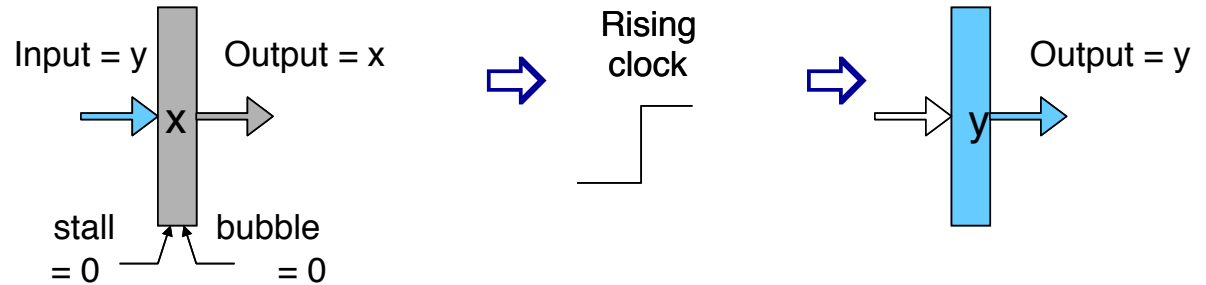


Stall

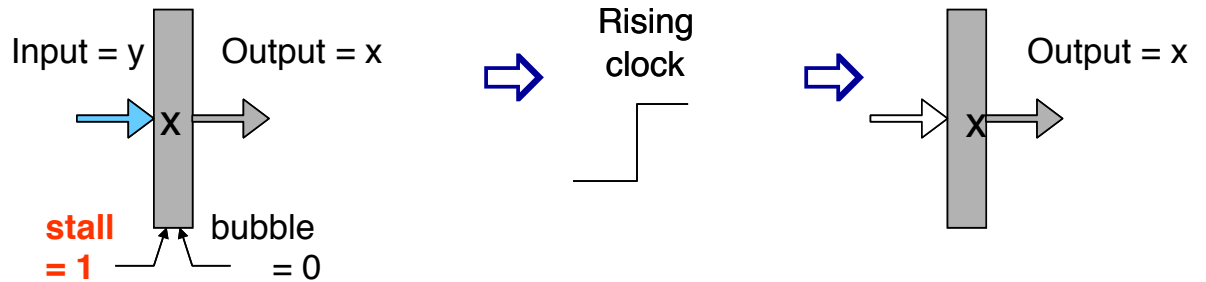


How are Stall and Bubble Implemented in Hardware?

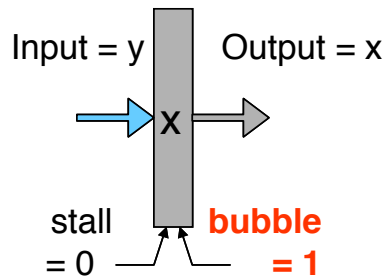
Normal



Stall

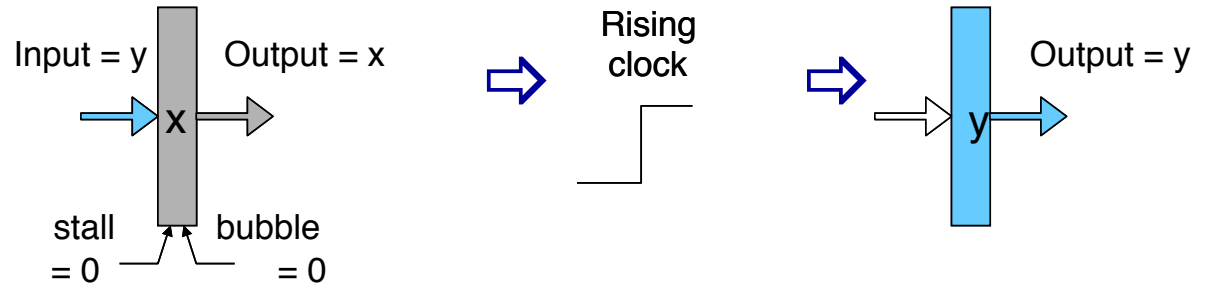


Bubble

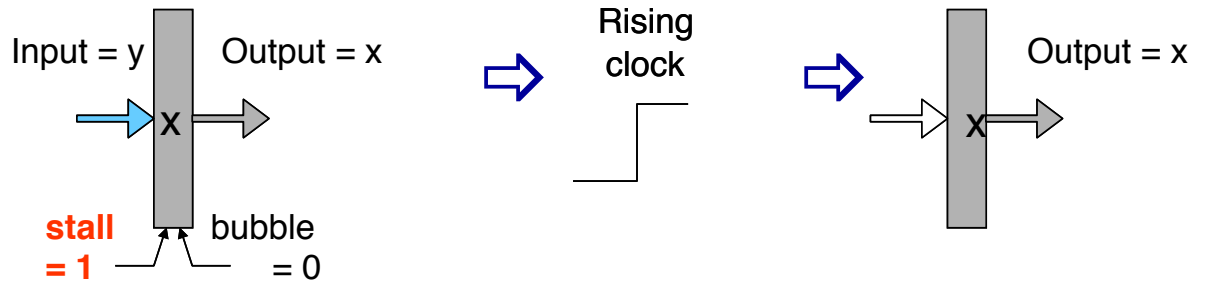


How are Stall and Bubble Implemented in Hardware?

Normal



Stall



Bubble

