COMP2611: Computer Organization

The Processor: Multicycle Datapath & Control

Can We Adopt Variable Clock Implementation?

Motivation:

- Many instructions do not involve all five functions units

Question:

- Why don't vary the clock for instructions with shorter execution time?

Answer:

- A variable-speed clock is very difficult to implement in hardware

The better solution:

Consider a **shorter clock cycle** and allow different types of instructions to take **different numbers of clock cycles**

(clock cycle is derived from the basic functional unit delays)

Multicycle Implementation

What does it mean by **multicycle**?

- An instruction takes multiple cycles to execute
- Execution of each instruction is broken into a series of steps
  1. Fetch the instruction
  2. Decode instruction & read the registers
  3. Perform ALU operation
  4. Memory access (if necessary)
  5. Write back the result
- Each step takes equal amount of time (one cycle) to complete

Why Multicycle?

Key advantage:

- Different types of instructions can have different number of cycles
  - Short instructions can complete early
    - e.g. R-type instruction can execute in 4 cycles instead of 5
  - Better performance as compared to the single-cycle implementation

Other advantage:

- Better sharing of function units ⇨ less hardware ⇨ cheaper
  - A single functional unit can be used more than once per instruction (as long as it is used at different clock cycles)
    - e.g. a single memory unit is used for both instructions and data
    - e.g. need only a single ALU, rather than one ALU and two adders

Trade-off

As compared to single-cycle implementation, multicycle approach needs

a. More temporary storage to hold values between clocks

   i.e. one or more registers added after every major functional unit

   - Instruction registers (IR)
   - Memory data registers (MDR)
   - Registers A & B
   - ALUOut

   But, why? (will explain in next couple of slides)

a. More complex control

   - Multiplexors
   - Control signals

Question: Worth adding all these?

Example: Single-Cycle CPU vs. Multicycle CPU

The instruction mix of a program:

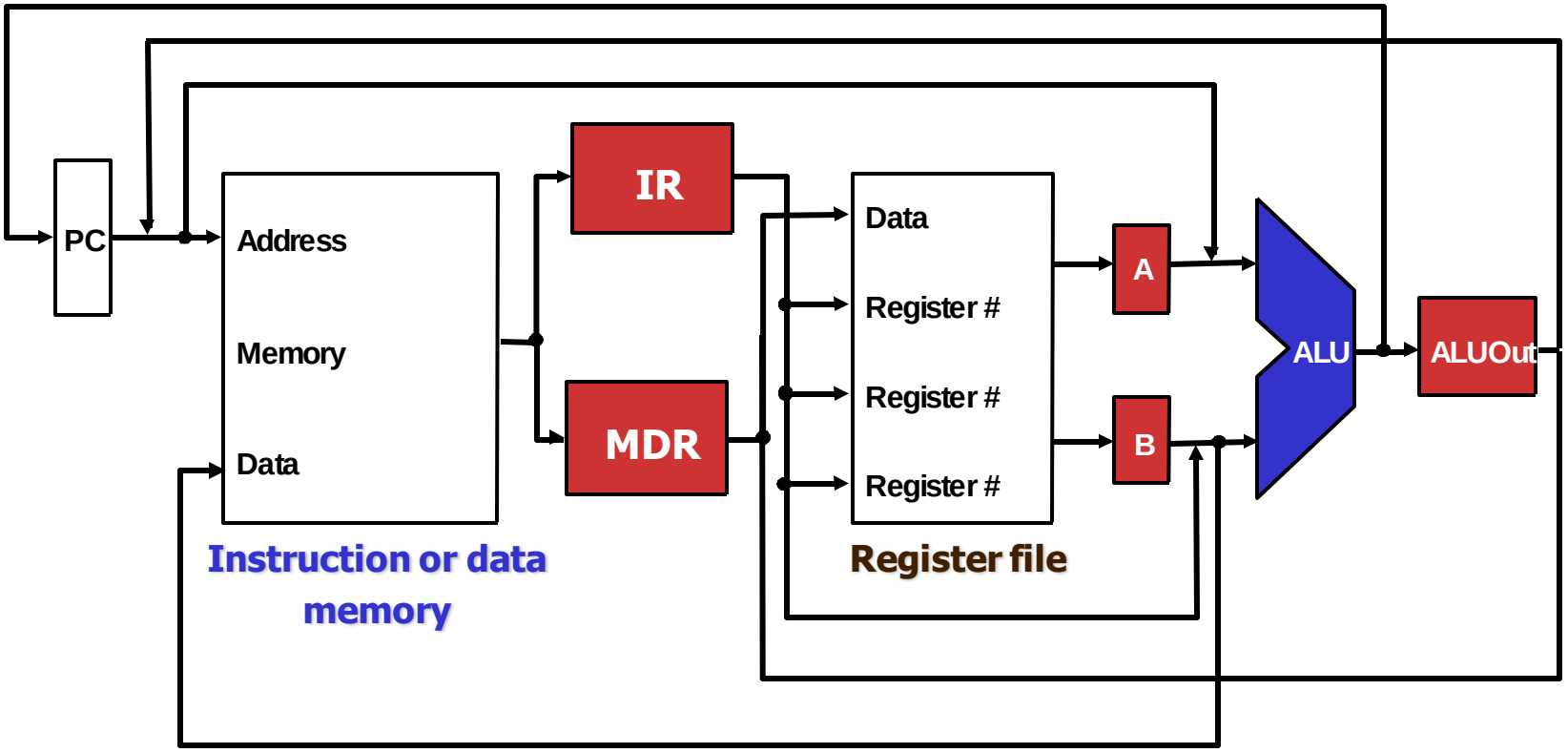
	Load	Store	Branch	Jump	ALU
# of instructions	25%	10%	11%	2%	52%
# of clocks/instr.	5	4	3	3	4

- Clock cycle of single-cycle CPU is **5x** of clock cycle of multicycle CPU
- For multicycle CPU (assuming the clock cycle is **R**)
  - CPI_{multi} = 0.25*5 + 0.10*4 + 0.11*3 + 0.02*3 + 0.52*4 = 4.12
  - $Time_{multi}$ = CPI_{multi} * $clock\ cycle_{multi}$ = 4.12 * **R** = **4.12R**
- For single-cycle CPU
  - CPI_{single} = 1.0
  - $Time_{single}$ = CPI_{single} * $clock\ cycle_{single}$ = 1.0 * **5R** = **5R**

5R / 4.12R = 1.21 ⇒ multicycle CPU speeds up by 21%

Overview of the Multicycle Datapath

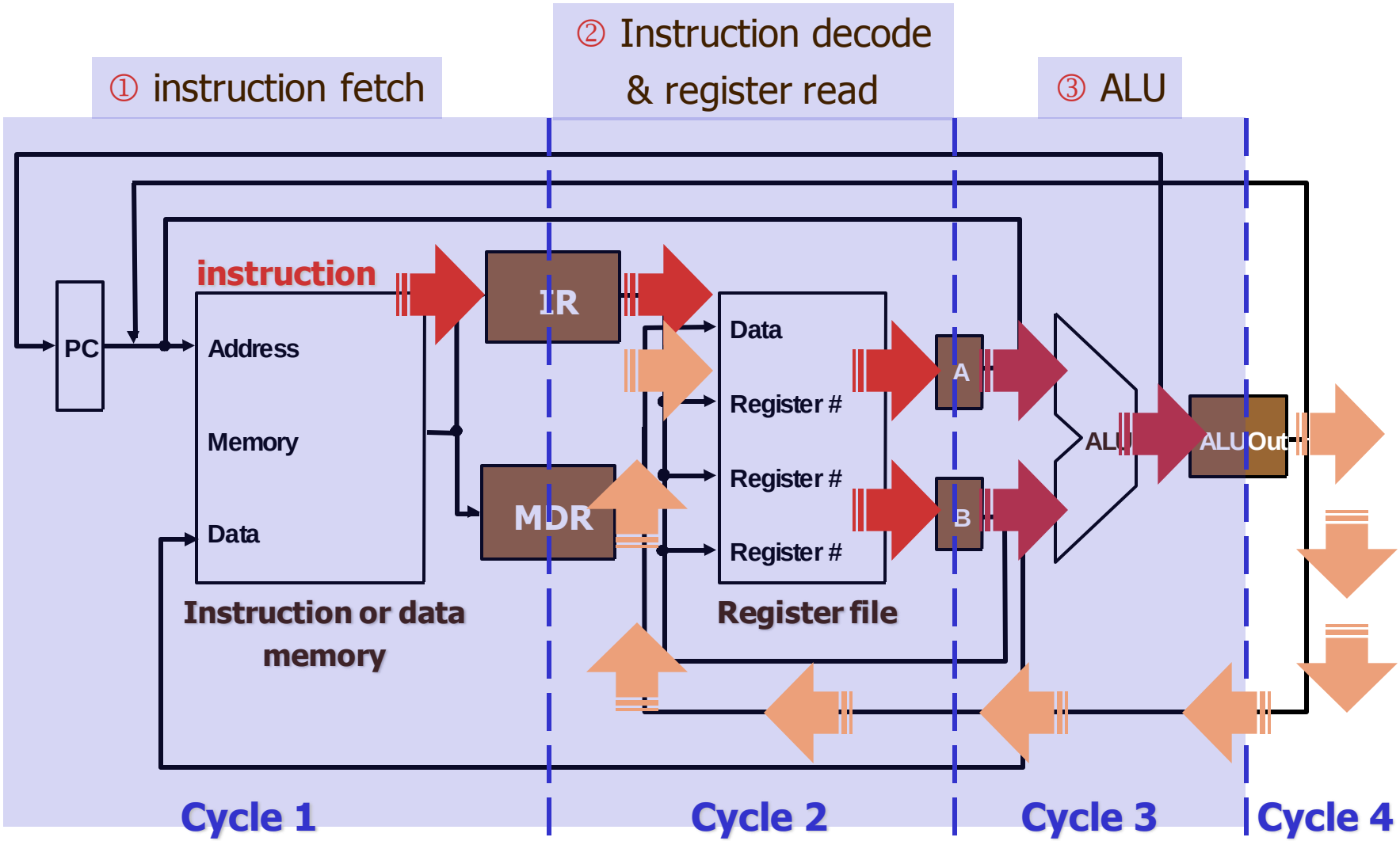
- **IR** ≡ Instruction Register; to hold instruction from memory
- **MDR** ≡ Memory Data Register; to hold data from memory
- **A, B**: registers to hold operand values from register file
- **ALUOut**: register to hold the output of the ALU



Blue: hardware being shared, Red: hardware added

Overview: How Do These Buffer Work?

- Let's use R-type instruction as an example



More About Temporary Storage

Memory is connected to both IR and MDR

- Output of the memory goes to both of them at the same time
- It seems like
  - Instruction read will destroy the value stored in MDR, or
  - Data read will destroy the instruction stored in IR
- Does it matter? How to resolve these problems?

ALUOut

- Holds the ALU's output (which will serve as an input for next cycle)
- The ALU's output can be either
  - Data (to be written to register or memory), or
  - Branch or jump target address

Need for Additional Control Signals

- **All except IR** hold data only between a pair of adjacent clock cycles
  - Thus, **MDR**, **A**, **B**, **ALUOut** **do not need a write control signal**

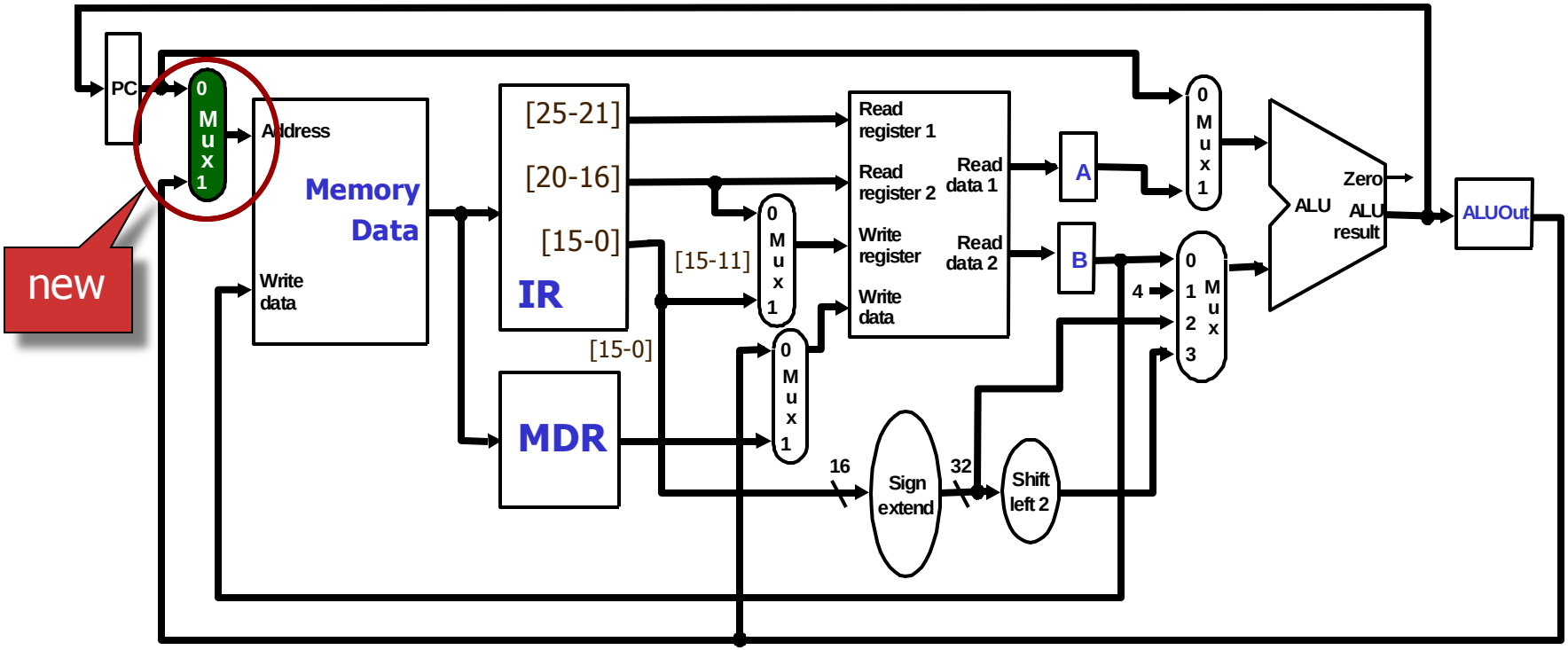
- **IR** needs to hold the instruction until the end of instruction execution
  - Thus, it **requires a write control signal**

- Since several functional units are shared for different purposes,
  - **Multiplexors** have to be added or expanded
  - Thus, **additional control signals are needed**
    - See next couple of slides for new multiplexors and control signals

Multiplexing Addresses to Read the Memory

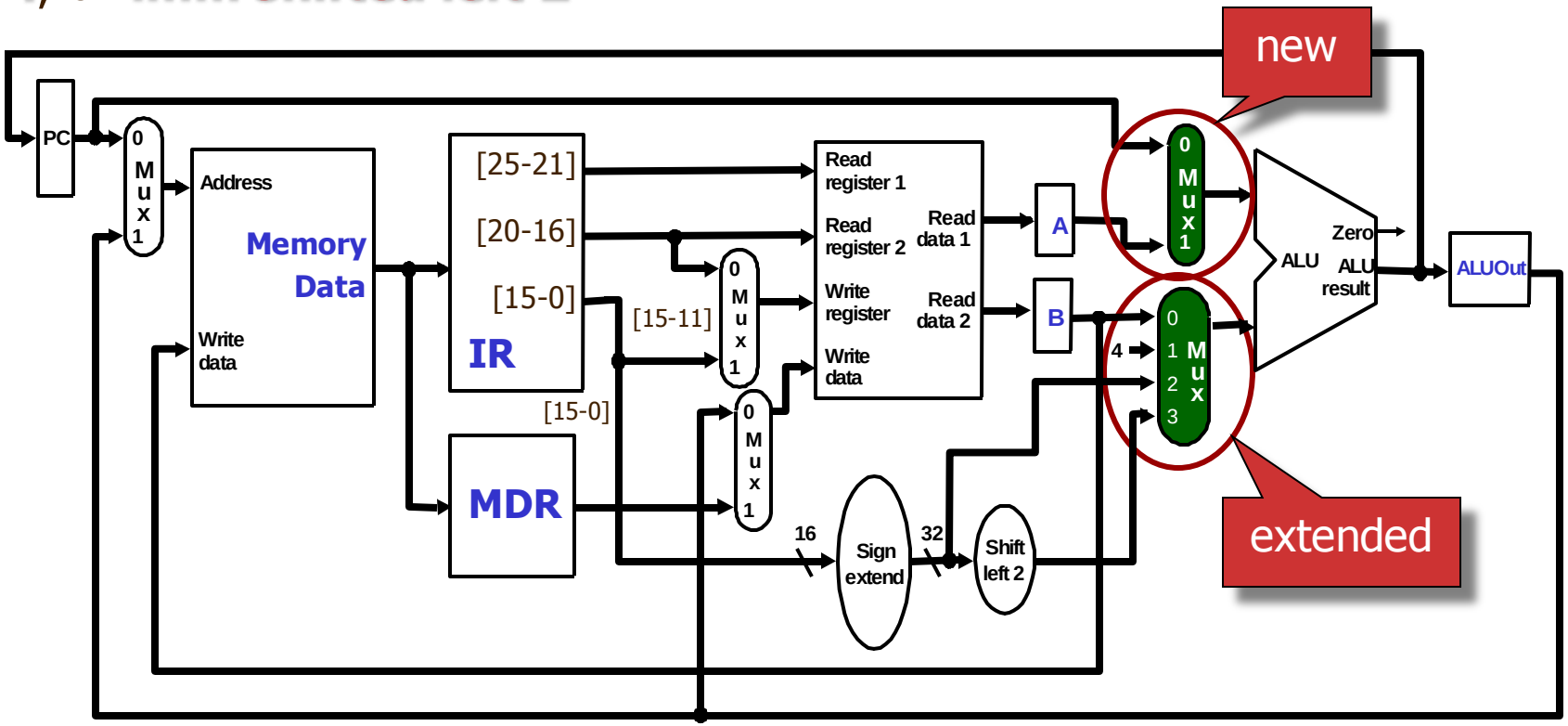
- Memory is shared by both instructions and data
  - The address to the memory module is no longer just **PC**
  - It can be either **instruction address** (PC) or **data address**

⇒ a **multiplexor** is needed to choose the source



Computation of Next PC Using the ALU

- The adders to compute next PC in single-cycle approach are removed
- So, generation of next PC and execution of operation share the ALU
  - The first source to the ALU can be either **A** or **PC**
  - The second source to the ALU can be either **B**, **sign-extended imm**, **4**, or **imm shifted left 2**



Constraints for Multicycle Datapath

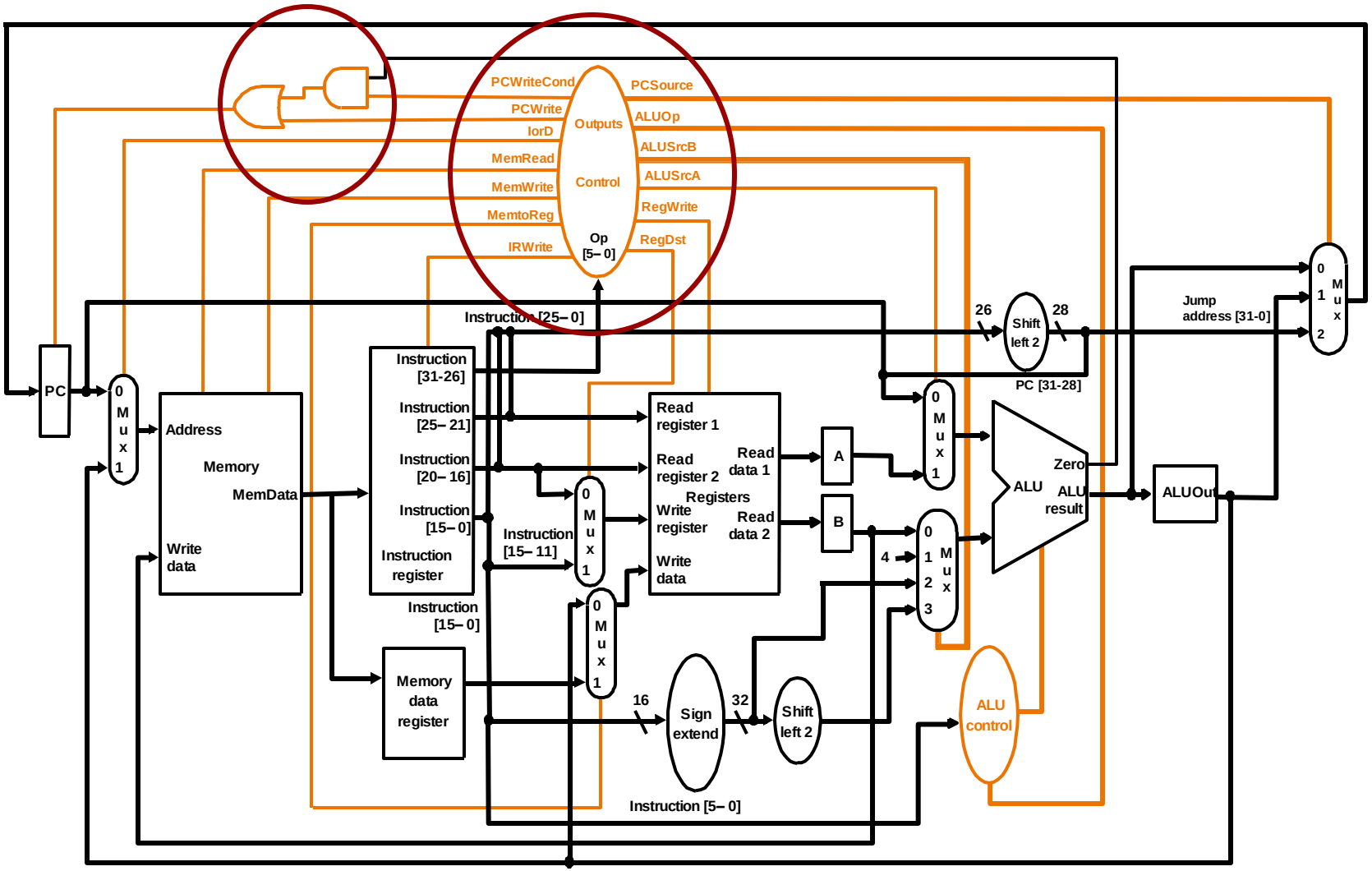
Only ONE ALU operation can be performed at each cycle

- Will calculation of next PC conflict with other ALU operation?
  - If yes, what should we do then?
  - If no, why?

Only ONE register file access can be performed at each cycle

- Will read from **src** registers conflict with write to **dest** register?
  - If yes, what should we do then?
  - If no, why?

Datapath & Control for Multicycle Implementation



3. Multicycle Implementation

1. Instruction Fetch

Hardware operations:

```
IR = Memory[PC];
PC = PC + 4;
```

- Send the PC to the memory as address
- Read an instruction from memory
- Write the instruction into the IR
- Increment the PC by 4 without violating the ALU usage constraints
  - Hence, hide the time required to compute PC+4
  - But, it seems like we are reading & writing PC in the same cycle!
  - Would there be any race condition? How to resolve it?

2. Instruction Decode & Register Read

Hardware operations:

(Assume the existence of two registers and an offset field, no harm to do the computation early even if they do not exist)

```
A = Reg[IR[25-21]];
B = Reg[IR[20-16]];
ALUOut = PC + (sign-extend(IR[15-0]) << 2);
```

- Read **rs** & **rt** from register file and store them to registers **A** & **B**
- Compute the branch target address (sign extension and left shift)
- Store the branch target address in ALUOut
  - If the instruction is actually a branch, ALUOut will be ready to provide the branch target address during next cycle
  - Again, we hide the time (i.e. avoid extra cycle) required to find the branch target address

Hardware operations:

a) ALU operation, or

b) Memory address computation, or

c) Branch completion

We'll look at them one by one in subsequent slides

Hardware operations (for R-format):

```
ALUOut = A op B;
```

- Send both registers A and B to the ALU
- Perform ALU operation specified by the function code on A & B
- Store the result in ALUOut

Hardware operations (for lw and sw):

```
ALUOut = A + sign-extend(IR[15-0]);
```

- Sign-extend the 16-bit offset to a 32-bit value
- Send both register A and the 32-bit offset to the ALU
- Store the result in ALUOut

3c. Instruction Execution: Branch Completion

Hardware operations: (for condition branch)

```
if (A == B) PC = ALUOut;
```

- Send both registers A and B to the ALU for comparison
  - Comparison is done by subtraction
- Set the PC to the branch target address if A and B are equal
  - Look at the output Zero from ALU to tell if equal
- If Zero = 1, write ALUOut computed during instruction decode to PC

3c. Instruction Execution: Branch Completion (2)

Hardware operations: (for unconditional branch)

```
PC = PC[31-28] || (IR[25-0] << 2);
```

- Compute the jump address
  - Left-shift the 26-bit address field by 2 bits to give a 28-bit value
- Concatenate the 4 leftmost bits of the PC with the 28-bit value to form a 32-bit jump address
- Write the jump address to the PC

4. Memory Access or Completion

Hardware operations:

a) Memory access (load / store), or

b) Instruction completion (for all R-type and some I-type)

We'll look at them one by one in subsequent slides

Hardware operations:

```
MDR = Memory[ALUOut];     // for load instruction
   or
Memory[ALUOut] = B;       // for store instruction
```

- Address was computed and stored in ALUOut during previous step
- For a `load` instruction:
  - Retrieve a data word from memory with the specified address
  - Then, put the data word in `MDR`
- For a `store` instruction:
  - Write data word stored in `B` to memory at the specified address

Hardware operations: (for R-type instructions)

```
Reg[IR[15-11]] = ALUOut;
```

- Get the ALUOut value computed during previous step
- Write the value into destination register in the register file

5. Memory Read Completion

Hardware operations:

```
Reg[IR[20-16]] <= MDR;
```

- Get the data value stored in MDR during previous step
- Write the value into destination register in the register file

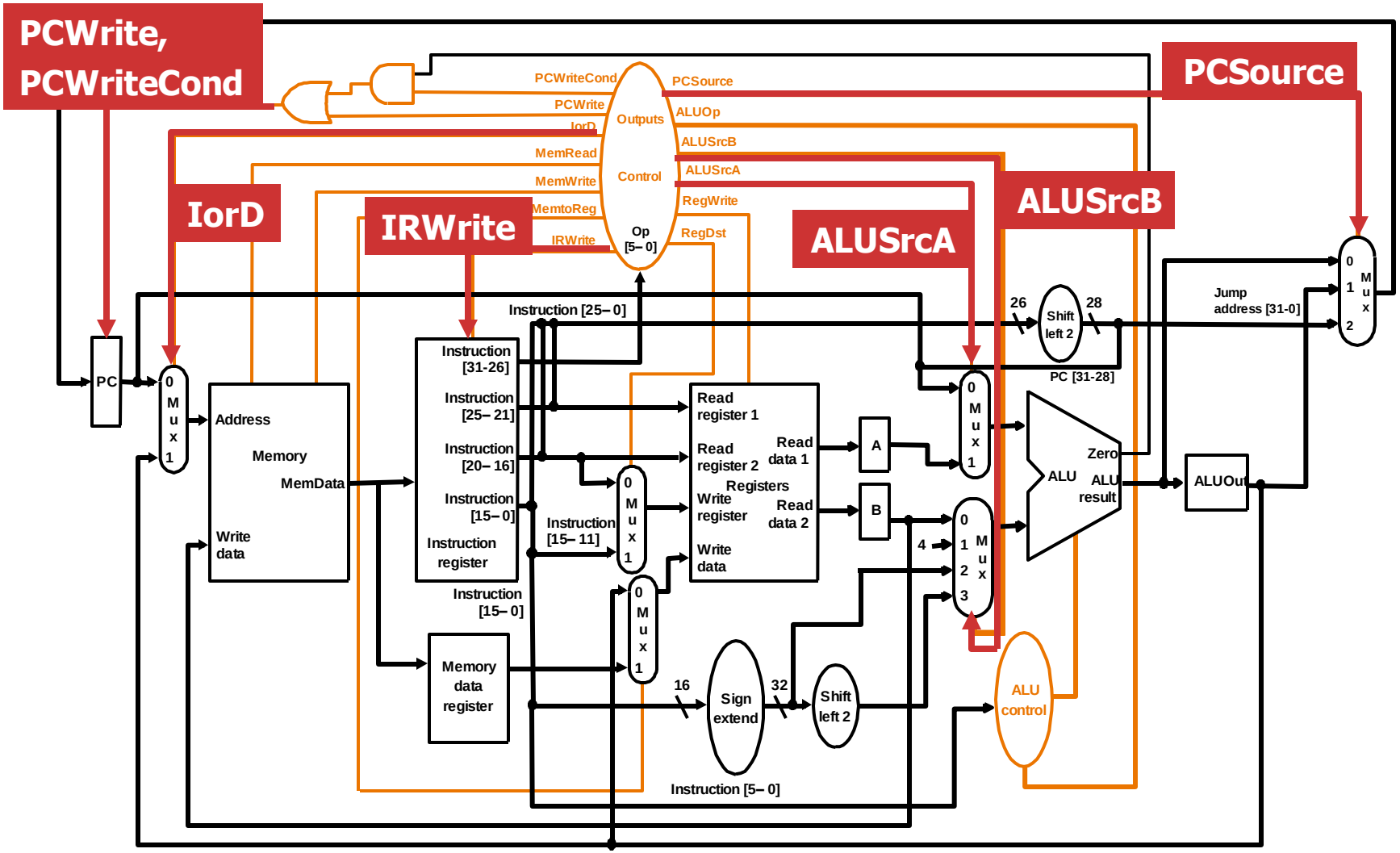
Summary of Execution Steps

	Actions for instructions			
Step name	**R-type**	**Memory references**	**Branches**	**Jumps**
Instruction fetch	IR = Memory[PC] PC = PC + 4			
Instruction decode / register fetch	A = Reg[IR[25-21]] B = Reg[IR[20-16]] ALUOut = PC + (sign-extend(IR[15-0]) << 2)			
Execution, addr comp., branch/jump completion	ALUOut = A op B	ALUOut = A + sign-extend(IR[15-0])	If (A == B) PC = ALUOut	PC = PC[31-28] || (IR[25-0] << 2)
Memory access, R-type completion	Reg[IR[15-11]] = ALUOut	Load: MDR=Memory[ALUOut]		
		Store: Memory[ALUOut] = B		
Memory read completion		Load: Reg[IR[20-16]] = MDR		
Total steps	**4**	**load: 5; store: 4**	**3**	**3**

Note: PC can be potentially modified at either step 1 or 3

4. Control Unit Design

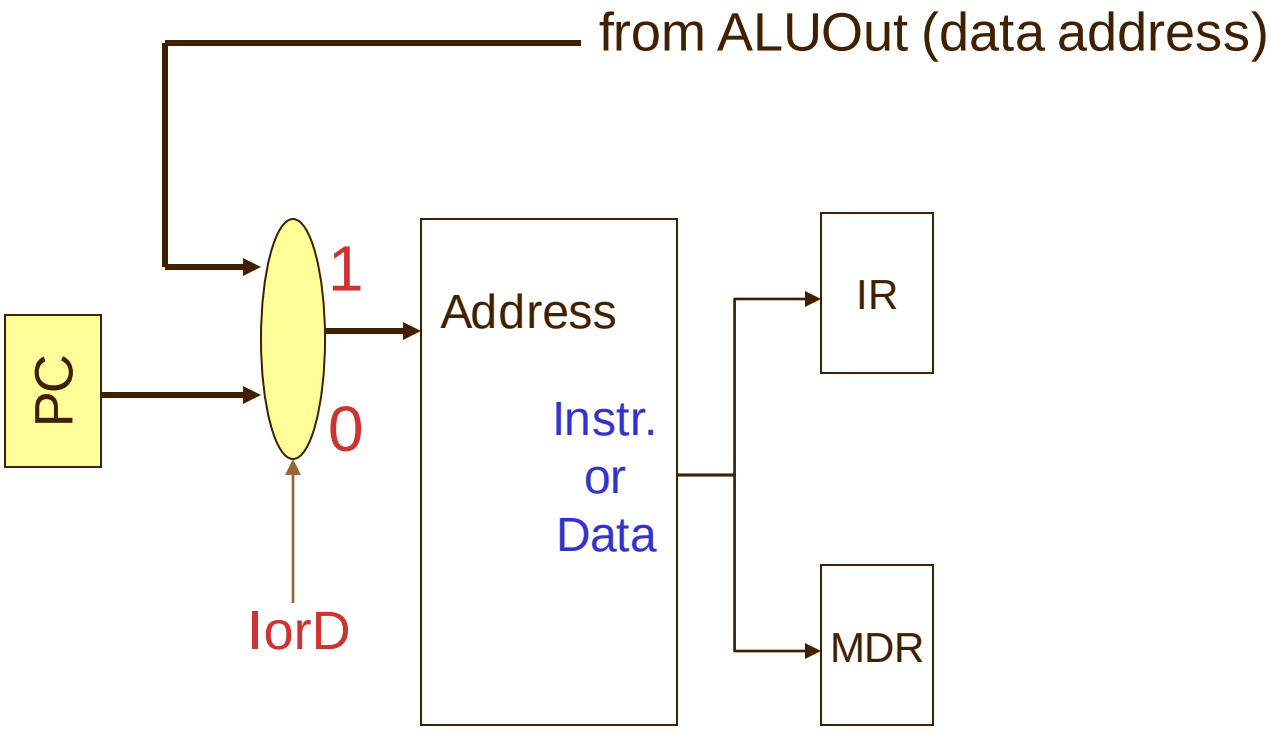
Control Signals Added in Multicycle Implementation



Control: IorD

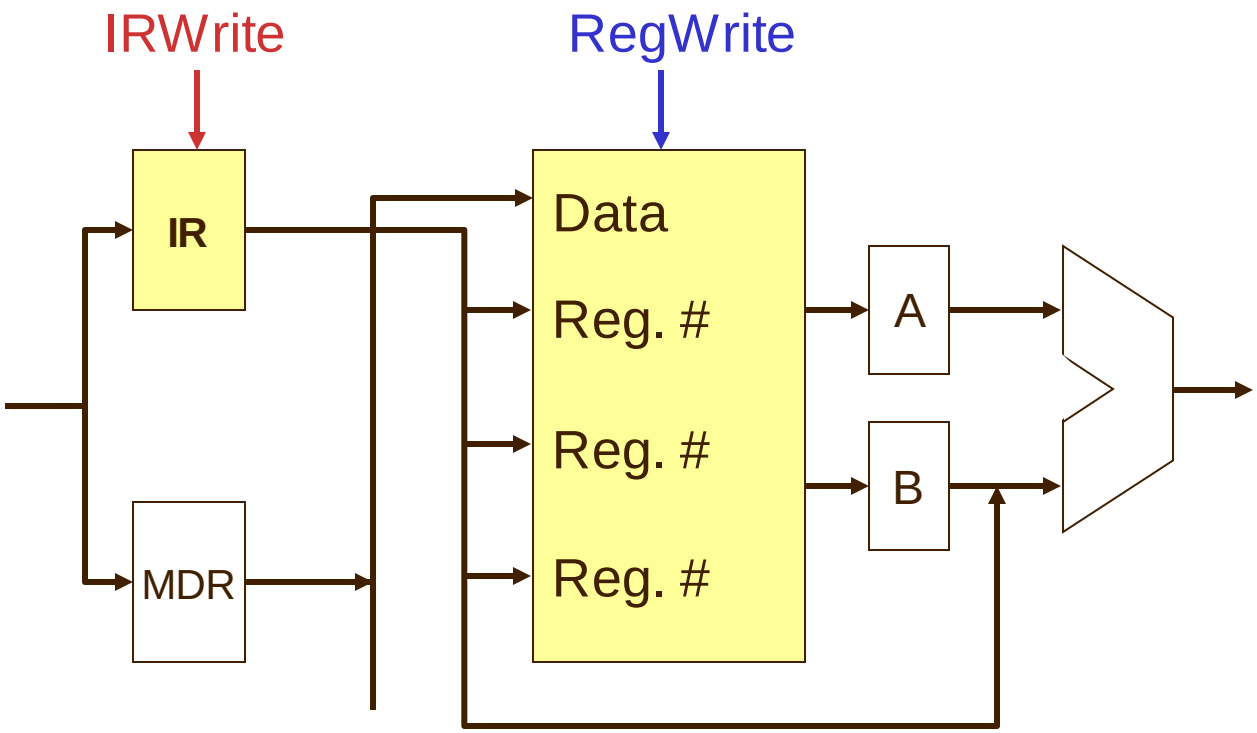
To select the source of address sent to the memory

- IorD = 0 ⇒ instruction, IorD = 1 ⇒ data



Control: IRWrite

- Enable write instruction register
  - IRWrite = 1 only upon instruction fetch; IRWrite = 0 otherwise
- Why no control for MDR?
  - Already have RegWrite to control when MDR being written to register



Control: ALUSrcA, ALUSrcB

To save hardware resources

- ALU operations and additions for branches/jumps share one ALU

What do we need for branches/jumps?

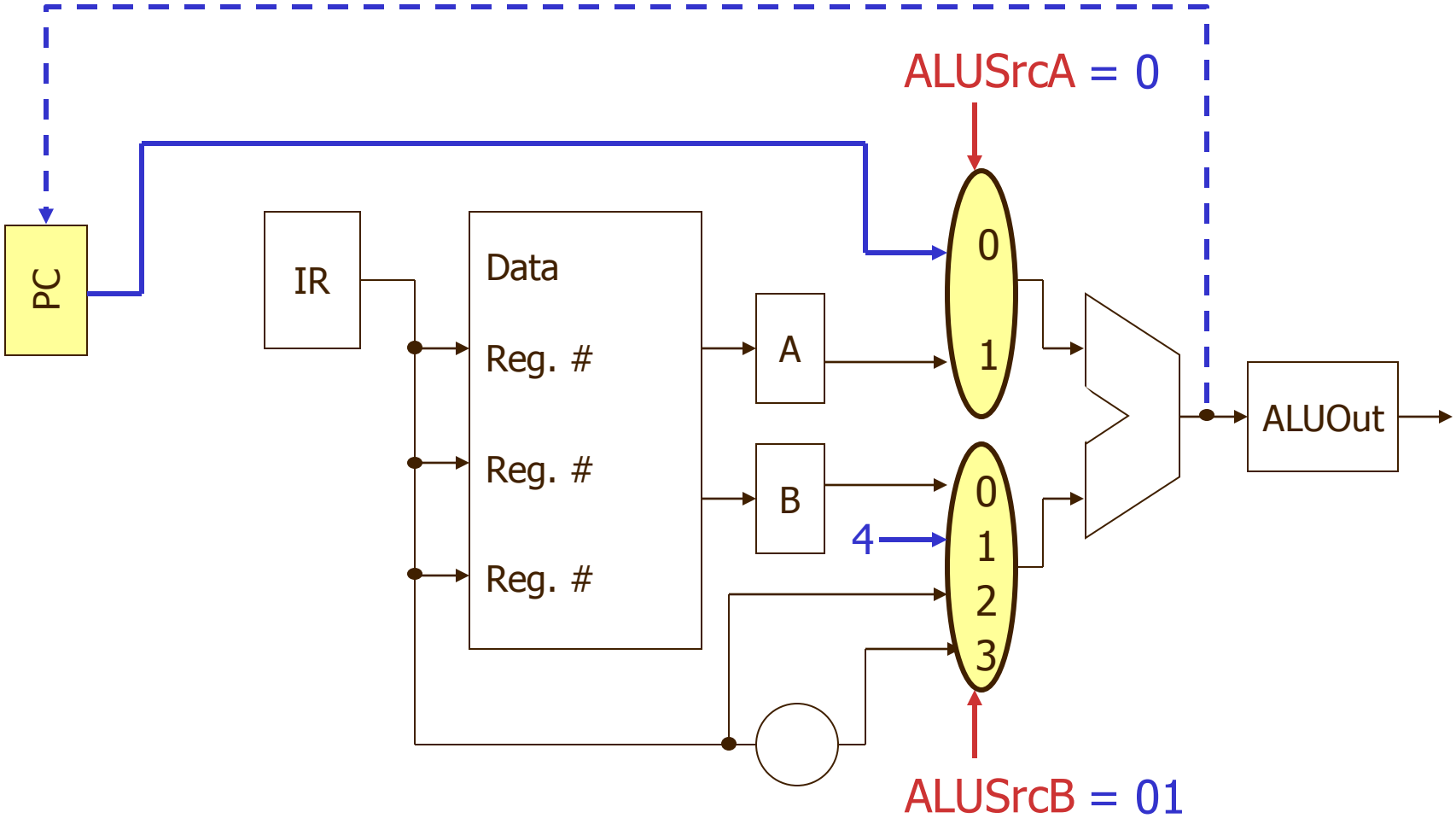
- (Taken) target, PC = PC + immediate << 2;
- (Non-taken) target, PC = PC + 4;
  - Also, the outcome of comparison (Zero?) for conditional branches
- (Unconditional) target, PC = PC[31-28] || (IR[25-0] << 2)

MIPS distributes all the above (target computations and comparison) to different cycles to fully utilize ALU without incurring unneeded cycle

Control: ALUSrcA, ALUSrcB by Example (1)

Base case: PC + 4

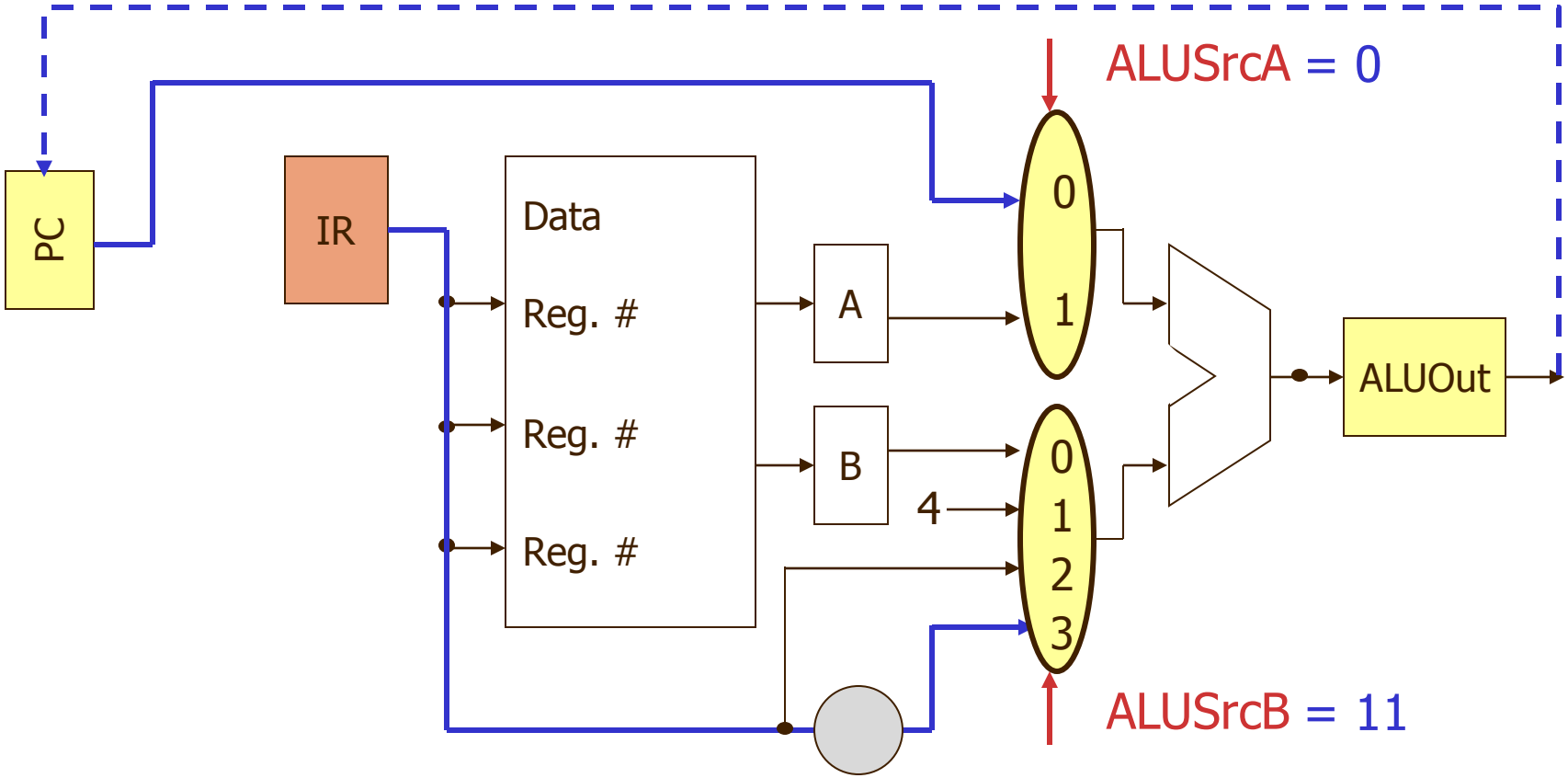
- Happens at step 1, while fetching instruction



Control: ALUSrcA, ALUSrcB by Example (2)

Conditional branch: e.g. beq $1, $2, target

- Calculation happens at step 2, while decoding instruction
- **ALUOut** stores the target address
  - Will be written to PC at step 3 if the comparison produces **Zero = 1**



Control: PCWriteCond, PCWrite, PCSource

Unconditional branch: e.g. j target

- Concatenation happens at step 3, does not involve ALU

Putting all together

- Writing to PC needs a MUX to select PC+4 or target address
  - Signals controlling the MUX come from PCWriteCond & PCWrite
    - PCWriteCond = 1, when the instruction is a conditional branch
    - PCWrite = 1, when the instruction is an unconditional branch
  - Signal = Zero • PCWriteCond + PCWrite
- PCSource selects the source of next PC
  - 00 ⇒ PC + 4, come from ALU
  - 01 ⇒ PC + offset, come from ALUOut
  - 10 ⇒ Jump address, come from concatenation

5b. Control Unit Design
(Finite state machine)

Control Unit Design for Multicycle Implementation

- Specifying control signals for single-cycle implementation is simple
  - Because everything happens within one cycle
  - Only require a set of truth tables based on instruction classes
- Specifying control signals for multicycle implementation is complex
  - Because an instruction is executed in a series of steps
  - Hence, have to specify signals for both the current and next steps

Two techniques for specifying the control:

1. **Finite state machine** representation
2. **Microprogramming** (will be explained if time permits)

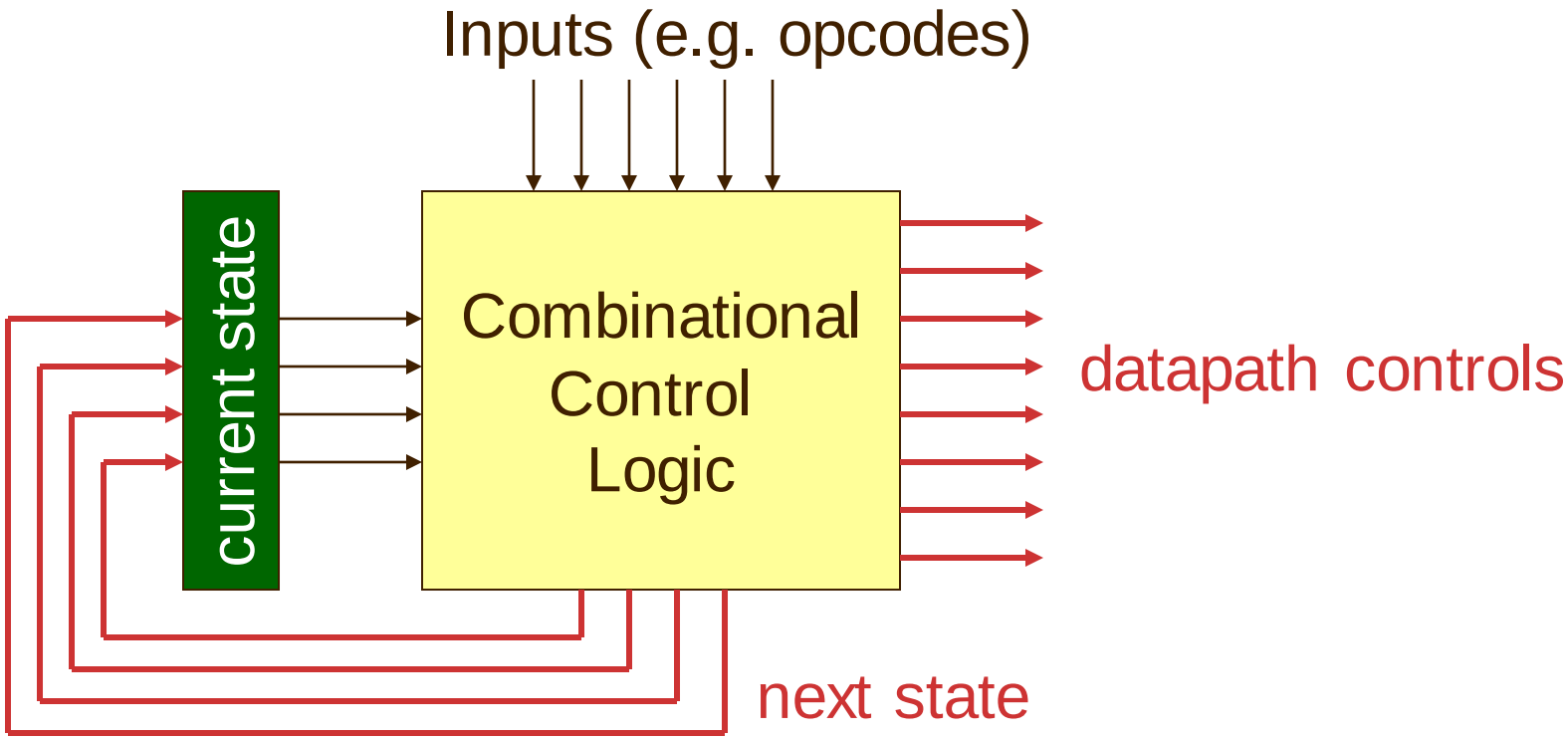
Finite State Machine (FSM)

A FSM consists of

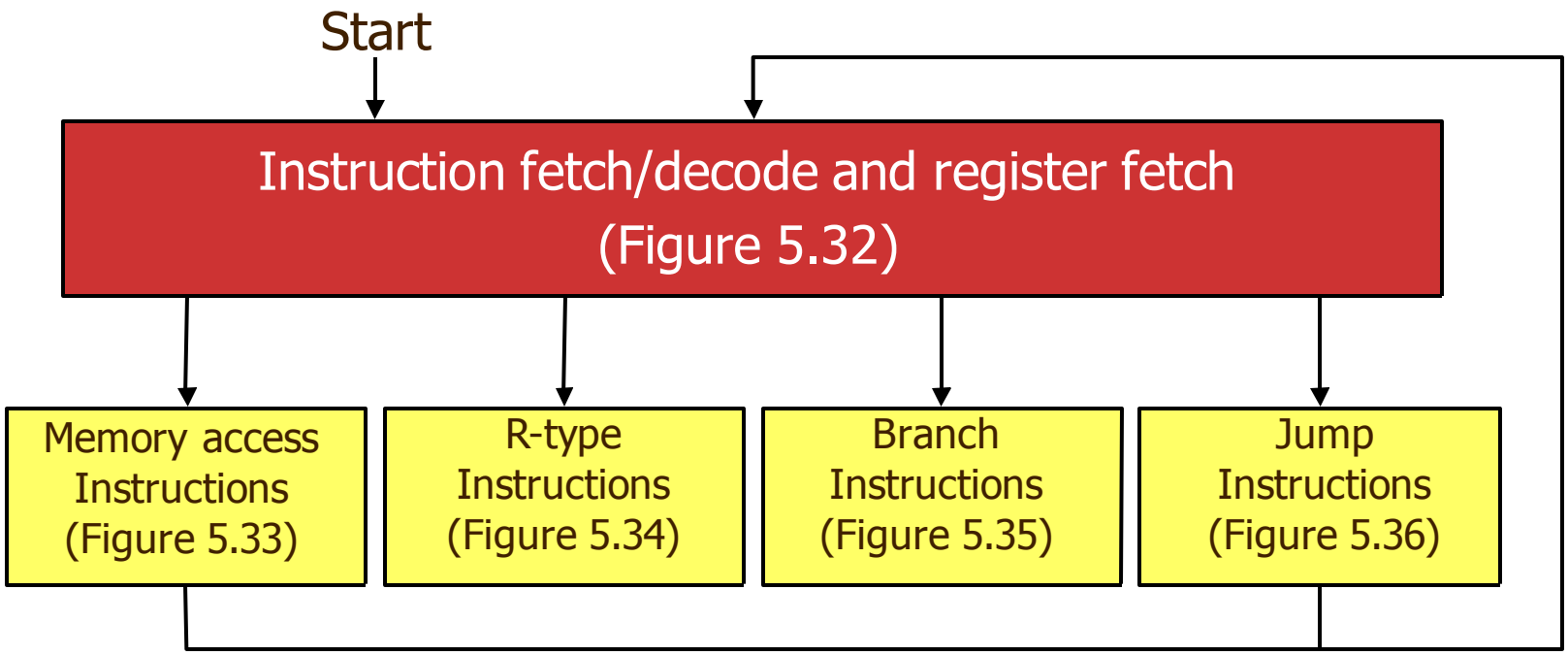
- A set of states & directions on how these states change

Directions

- Functions mapping (current state, inputs) to (new state)

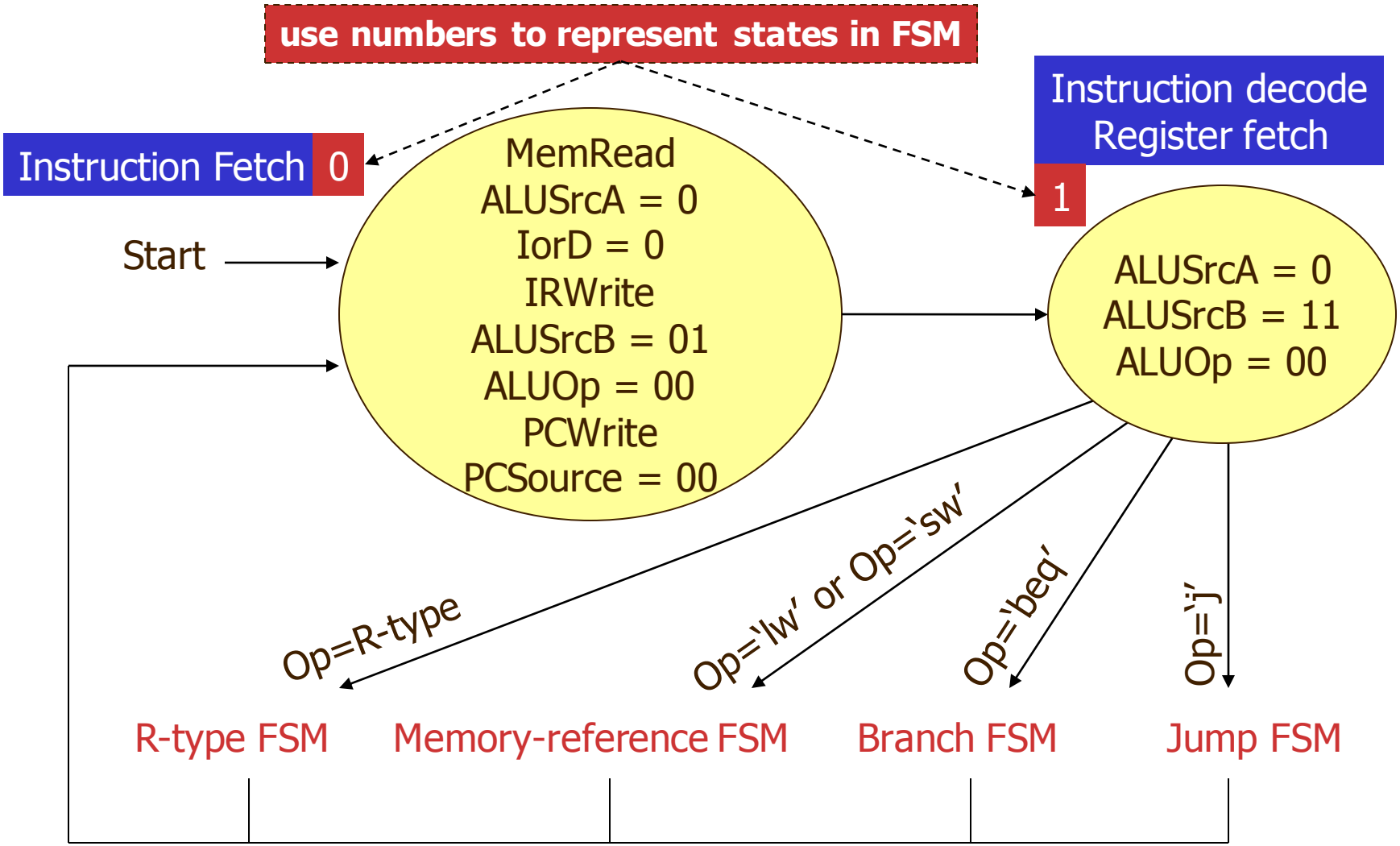


High-Level View of Finite State Machine Control

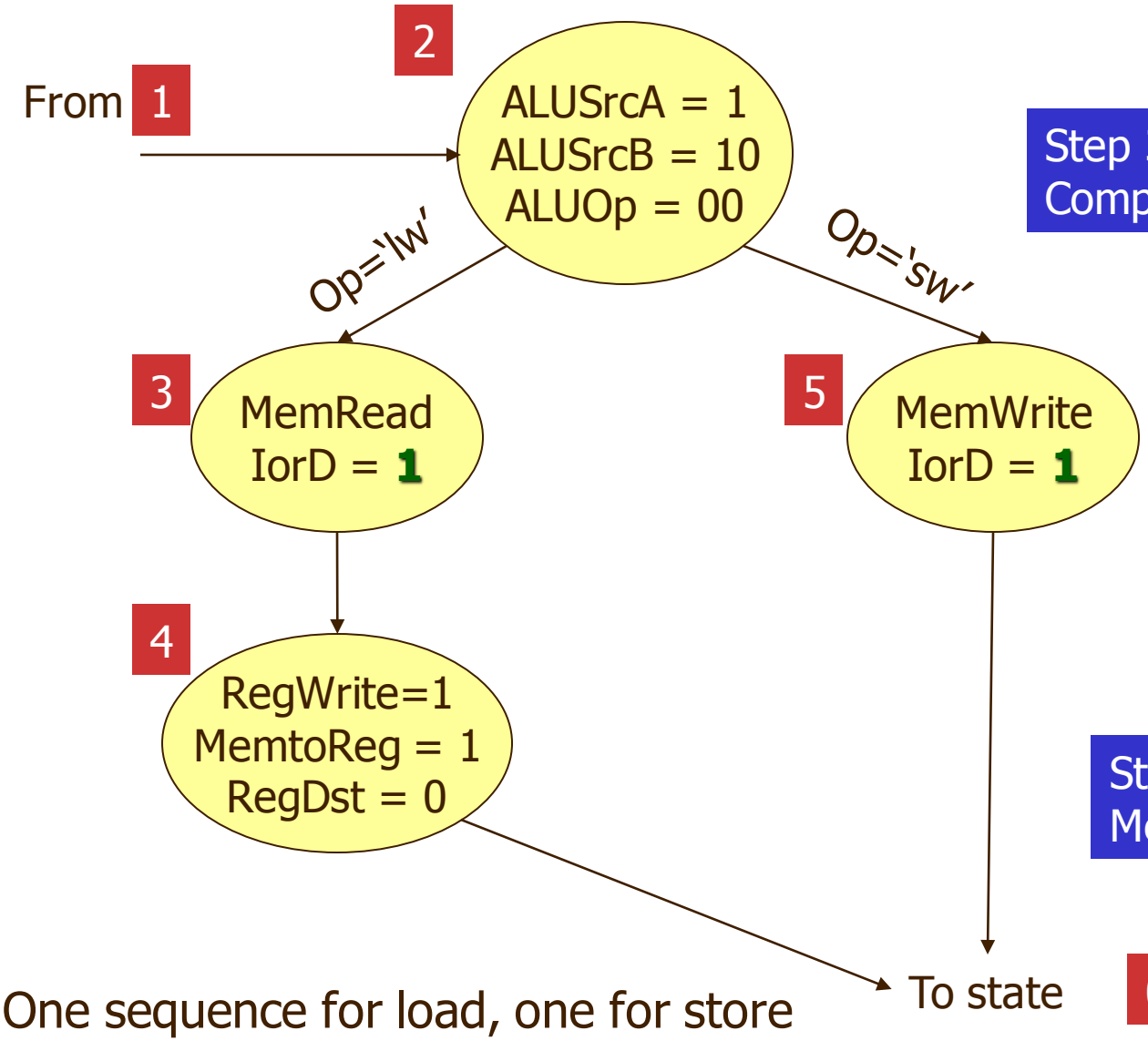


Instruction Fetch & Decode

- This part is common to all instructions



Memory Reference FSM



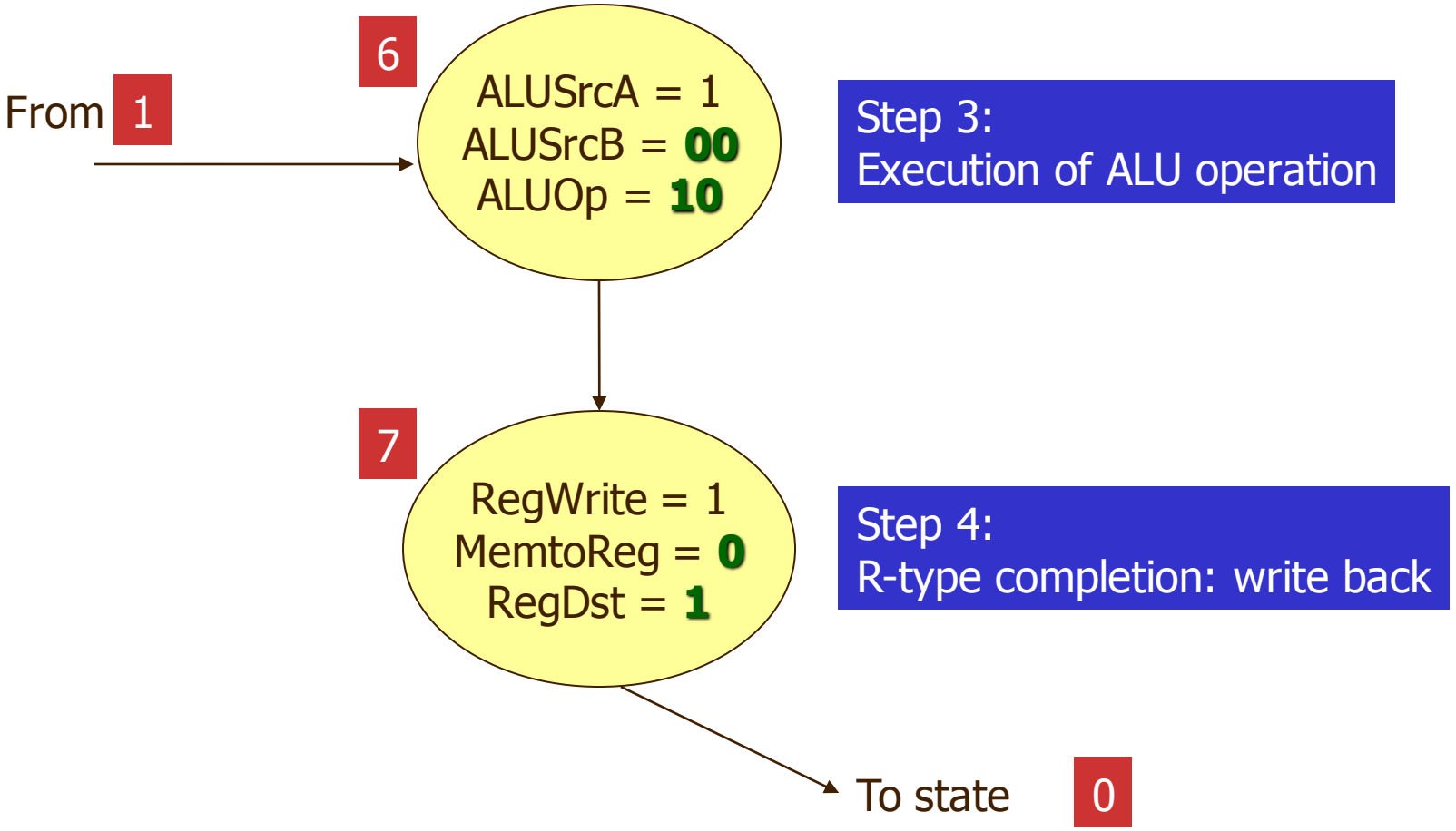
Step 3:
Computing memory address

Step 4:
Access memory

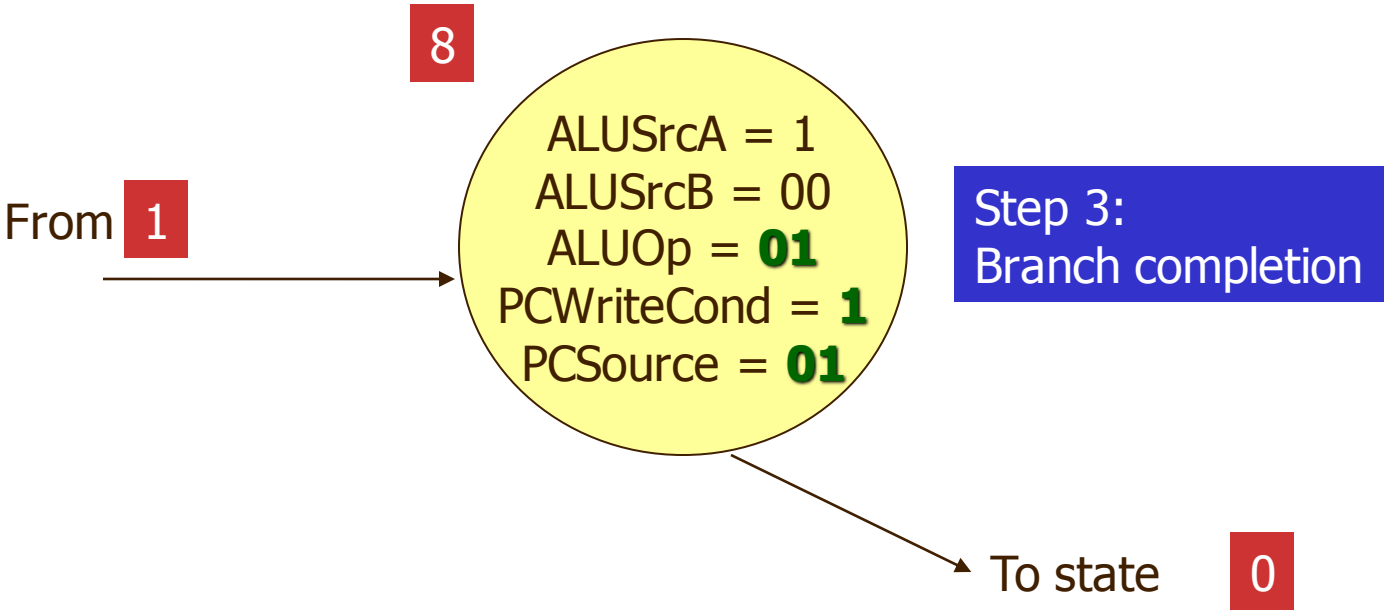
Step 5:
Memory read completion

- One sequence for load, one for store

R-Type FSM

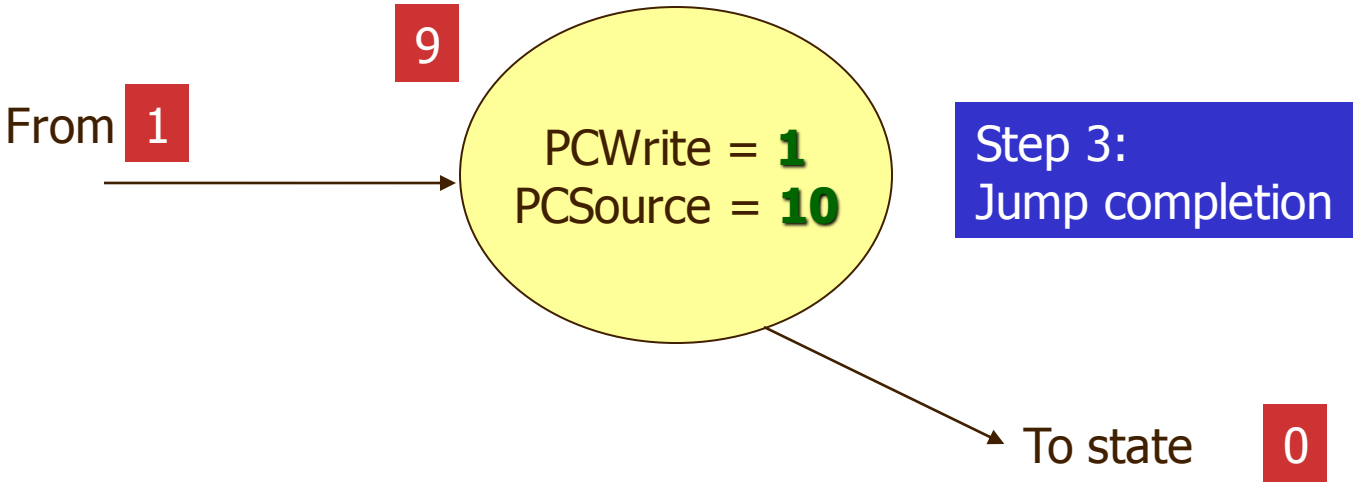

From 1
6
ALUSrcA = 1
ALUSrcB = 00
ALUOp = 10
Step 3:
Execution of ALU operation
7
RegWrite = 1
MemtoReg = 0
RegDst = 1
Step 4:
R-type completion: write back
To state 0

Branch FSM



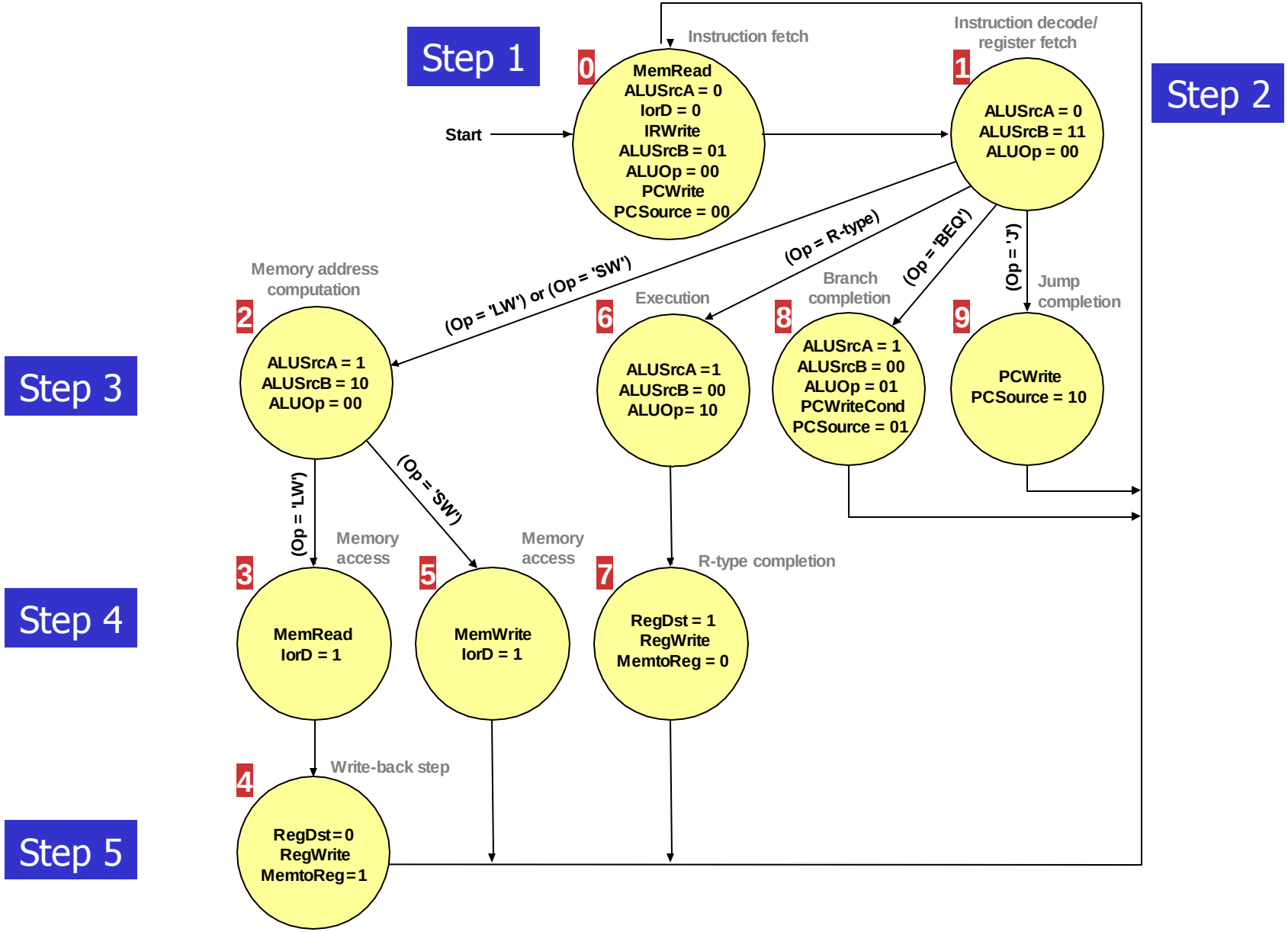
- ALUSrcA, ALUSrcB, ALUOp cause ALU to compare the registers
- PCSource, PCWriteCond enable change of PC if branch condition is true
  - Zero output of the ALU gives the branch condition

Jump FSM



- Concatenation is executed all the times
- Concatenation = `PC[31-28] || (IR[25-0] << 2)`
  - But, the result of concatenation is used only when PCSource = 10
  - Otherwise, dropped

Putting All Together



Optional
5c. Control Unit Design
(Microprogramming)

Simplifying Control Design by Microprogramming

- Use a FSM to represent the MIPS control is relatively easy & feasible

However,

- In case of a large instruction set and a large number of complex addressing modes (e.g., Intel x86/Pentium instruction set), there will be thousands of states in the FSM, which becomes too large and cumbersome to handle

As programs become large, we need a good structuring techniques to keep the programs comprehensible

The solution is called **Microprogramming**

- It basically treats the control design as a programming problem
  i.e. use microinstructions to '*program controls we need*' in each step
- Note that:
  - Datapath remains the same
  - Only control is specified differently

Microprogramming

- The set of controls to be asserted during a cycle are considered as an "control instruction"
- To distinguish such low-level control instructions from MIPS' instructions, we call them **microinstructions or microcodes**
- Executing a microinstruction has the effect of asserting the control signals specified by the microinstruction

- Values asserted on the control lines are represented symbolically in a **microprogram**, a representation of the microinstructions

In brief, you can say that each instruction

- Consists of a sequence of microinstructions
- # of microinstructions for the instruction =
  # of cycles to execute the instruction

Microinstructions

Microinstructions contain fields that specify

- **ALU Control, SRC1 and SRC2**
  - ALUop and the sources
- **Register control**
  - Read or write to the register file, and
  - Select the source of the value to write
- **Memory**
  - Read or write, and
- **PCWrite Control**
  - Specify the writing on the PC
- **Sequencing**
  - How to choose the next microinstruction

Example of Microinstruction

Label	ALU Control	SRC1	SRC2	Register Control	Memory	PCWrite Control	Sequencing
FETCH	ADD	PC	4		Read PC	ALU	Seq
	ADD	PC	Extshift	READ			DISPATCH 1
Mem1	ADD	A	Extend				DISPATCH 2
LW2					Read ALU		Seq
				Write MDR			FETCH
SW2					Write ALU		FETCH
Rformat1	Func Code	A	B				Seq
				Write ALU			FETCH
BEQ1	Subt	A	B			ALUOut cond	FETCH
JUMP1						Jump Addr	FETCH

- We can think of the **dispatch operation** as a ***case*** or ***switch*** statement with the opcode field and the dispatch table **T** used to select one of ***n*** different microinstruction sequences with one of ***n*** different labels (all ending in "**T**")

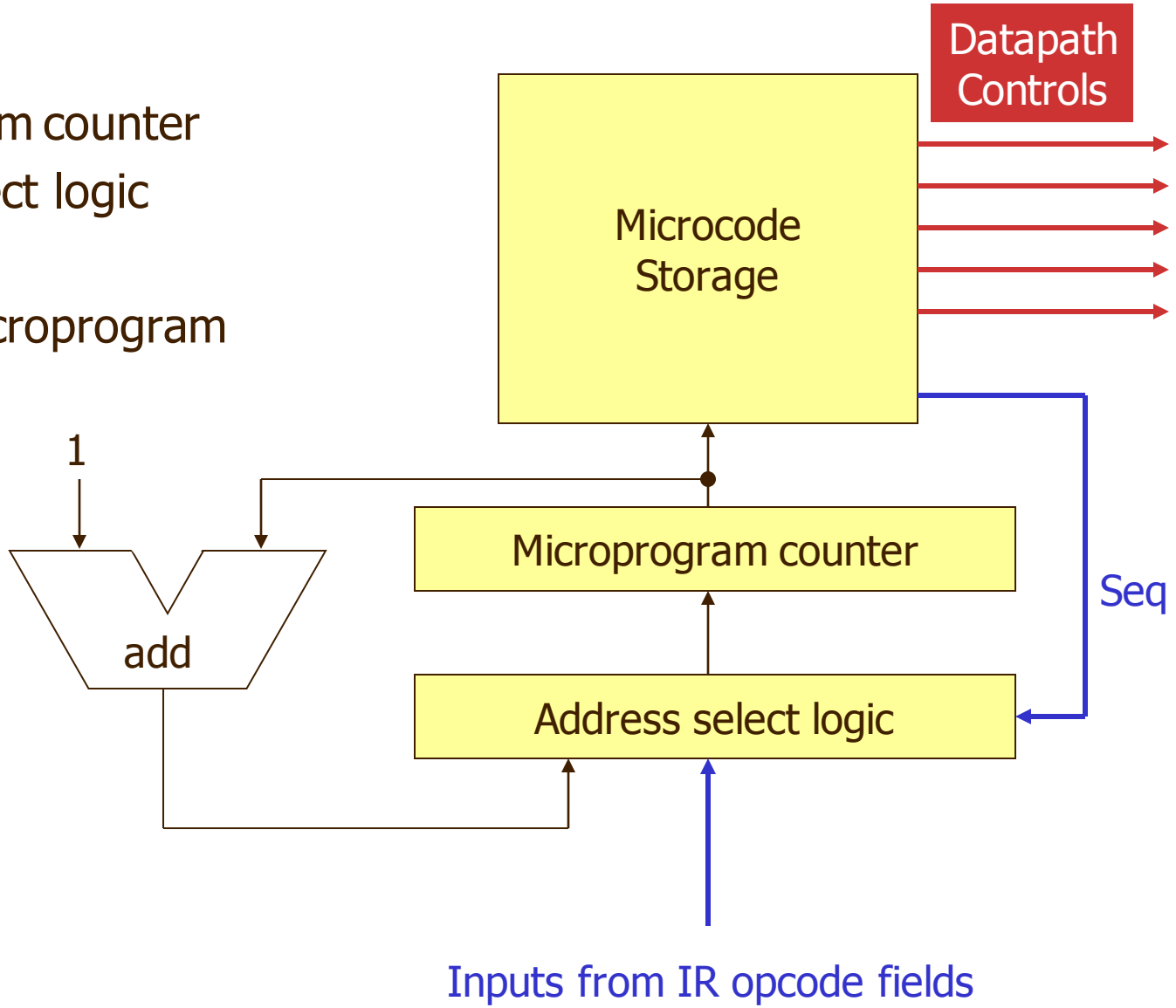
Implementing Microprogramming

We need

- Microprogram counter
- Address select logic

Storage for microprogram

- ROMs
- PLAs



Microinstructions and their fields

Field name	Value	Signals active	Comment
ALU control	Add	ALUOp = 00	Cause the ALU to add.
	Subt	ALUOp = 01	Cause the ALU to subtract; this implements the compare for branches.
	Func code	ALUOp = 10	Use the instruction's function code to determine ALU control.
SRC1	PC	ALUSrcA = 0	Use the PC as the first ALU input.
	A	ALUSrcA = 1	Register A is the first ALU input.
SRC2	B	ALUSrcB = 00	Register B is the second ALU input.
	4	ALUSrcB = 01	Use 4 as the second ALU input.
	Extend	ALUSrcB = 10	Use output of the sign extension unit as the second ALU input.
	Extshft	ALUSrcB = 11	Use the output of the shift-by-two unit as the second ALU input.
Register control	Read		Read two registers using the rs and rt fields of the IR as the register numbers and putting the data into registers A and B.
	Write ALU	RegWrite, RegDst = 1, MemtoReg = 0	Write a register using the rd field of the IR as the register number and the contents of the ALUOut as the data.
	Write MDR	RegWrite, RegDst = 0, MemtoReg = 1	Write a register using the rt field of the IR as the register number and the contents of the MDR as the data.
Memory	Read PC	MemRead, IorD = 0	Read memory using the PC as address; write result into IR (and the MDR).
	Read ALU	MemRead, IorD = 1	Read memory using the ALUOut as address; write result into MDR.
	Write ALU	MemWrite, IorD = 1	Write memory using the ALUOut as address, contents of B as the data.
PC write control	ALU	PCSource = 00 PCWrite	Write the output of the ALU into the PC.
	ALUOut-cond	PCSource = 01, PCWriteCond	If the Zero output of the ALU is active, write the PC with the contents of the register ALUOut.
	jump address	PCSource = 10, PCWrite	Write the PC with the jump address from the instruction.
Sequencing	Seq	AddrCtl = 11	Choose the next microinstruction sequentially.
	Fetch	AddrCtl = 00	Go to the first microinstruction to begin a new instruction.
	Dispatch 1	AddrCtl = 01	Dispatch using the ROM 1.
	Dispatch 2	AddrCtl = 10	Dispatch using the ROM 2.