#### HARDWARE SPECULATION

Mahdi Nazm Bojnordi

**Assistant Professor** 

School of Computing

University of Utah



#### Overview

- Announcement
  - Homework 3: Oct. 31st

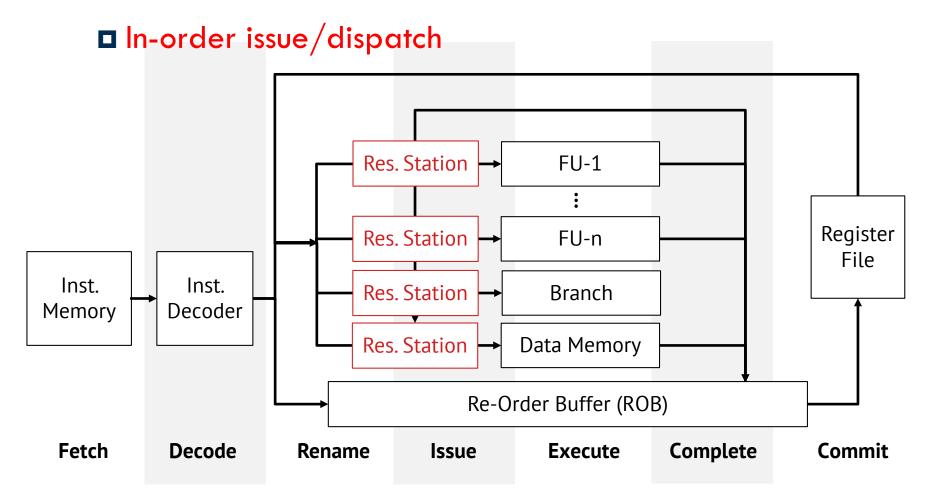
- □ This lecture
  - Out-of-order pipeline
    - Issue queue
    - Register renaming
    - Branch recovery
    - Speculated execution

#### Recall: Out-of-Order Execution

- Producer-consumer chains on the fly
  - Register renaming: remove anti-/output-dependences via register tags
  - Limited by the number of instructions in the instruction window (ROB)
- □ Out-of-order issue (dispatch)
  - Broadcast tags to waiting instructions
  - Wake up ready instructions and select among them
- Out-of-order execute/complete
- In-order fetch/decode and commit

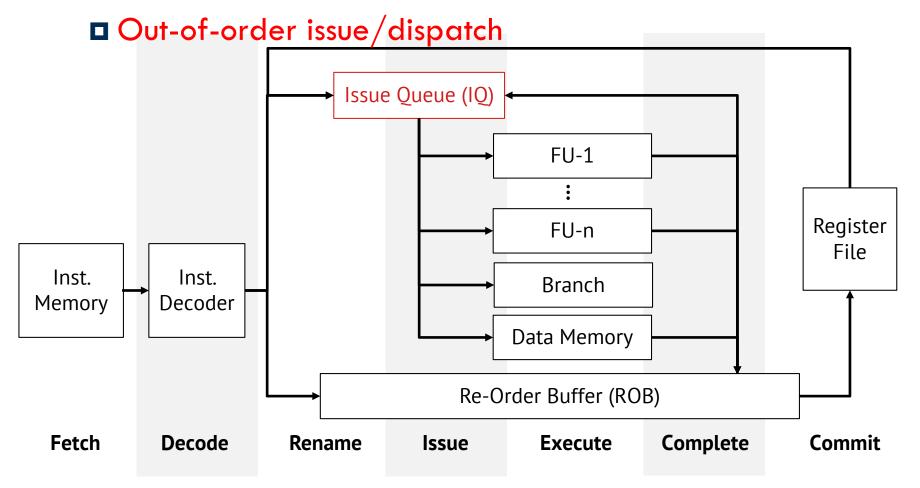
# Out-of-Order Pipelines

Distributed reservation stations



# Out-of-Order Pipelines

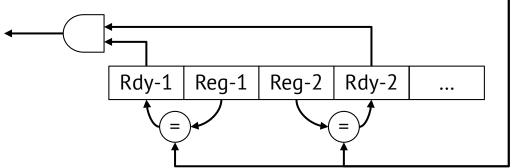
Out of order issue/dispatch to functional units



#### Out-of-Order Issue Queue

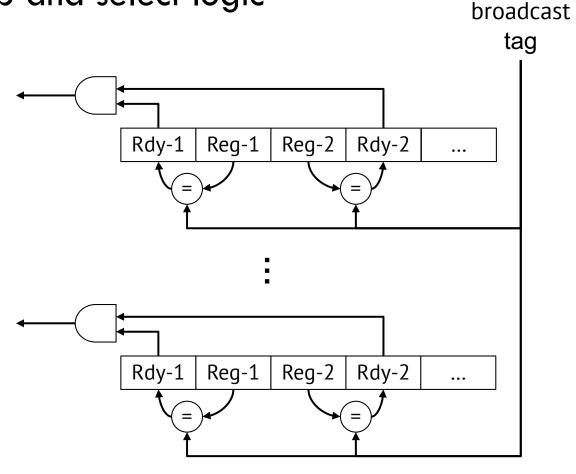
□ Two step wakeup and select logic

broadcast tag

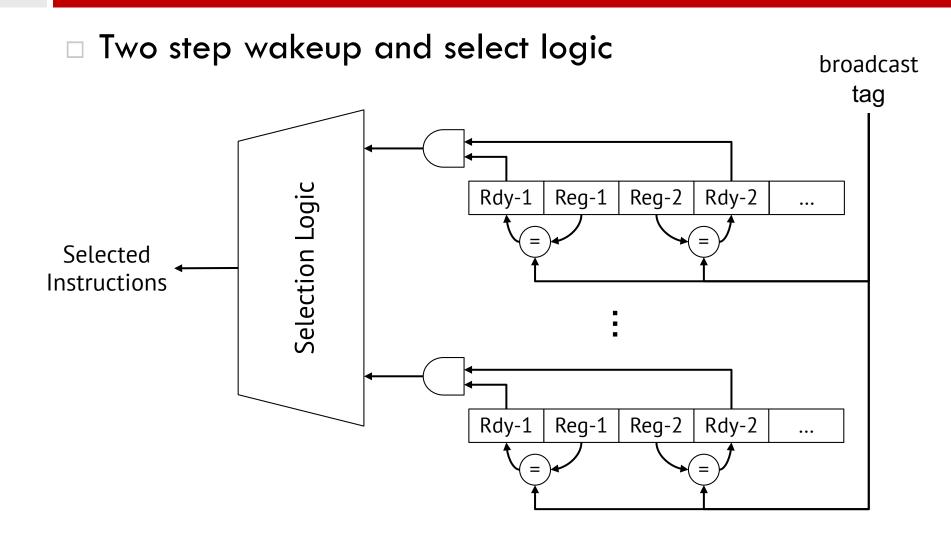


#### Out-of-Order Issue Queue

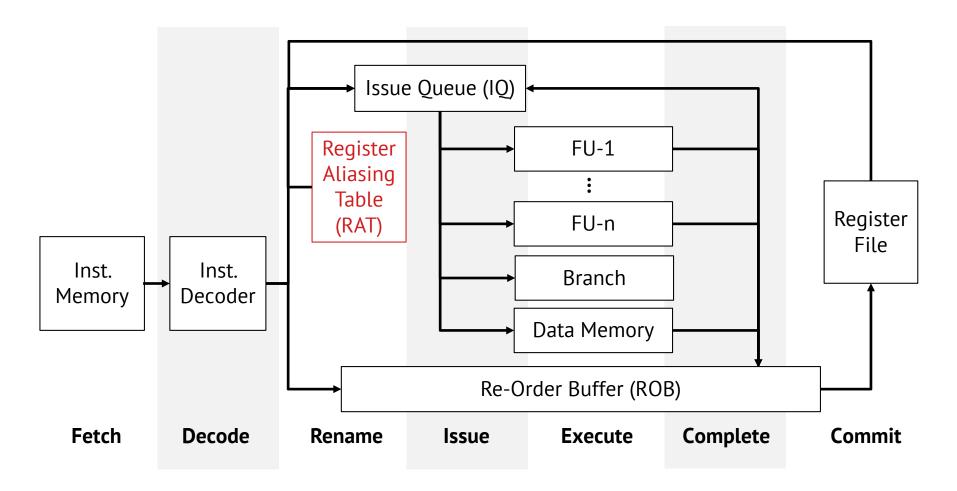
□ Two step wakeup and select logic



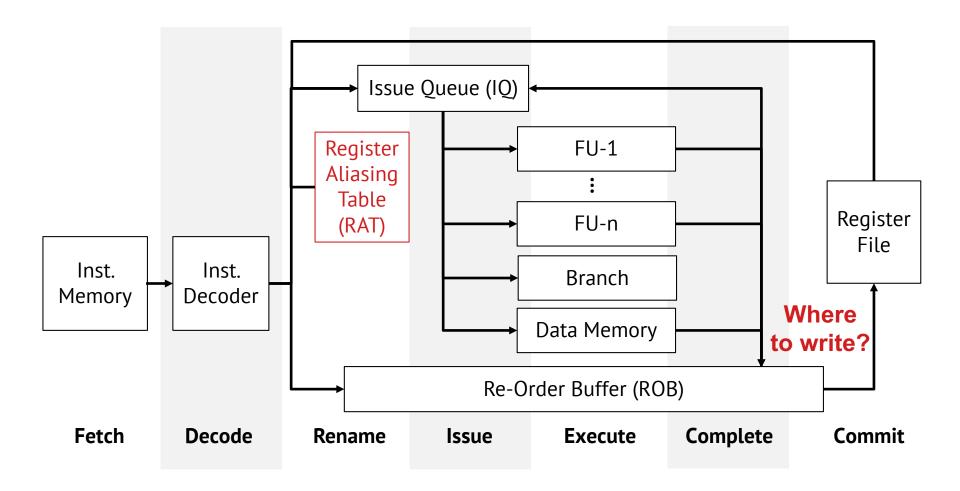
#### Out-of-Order Issue Queue



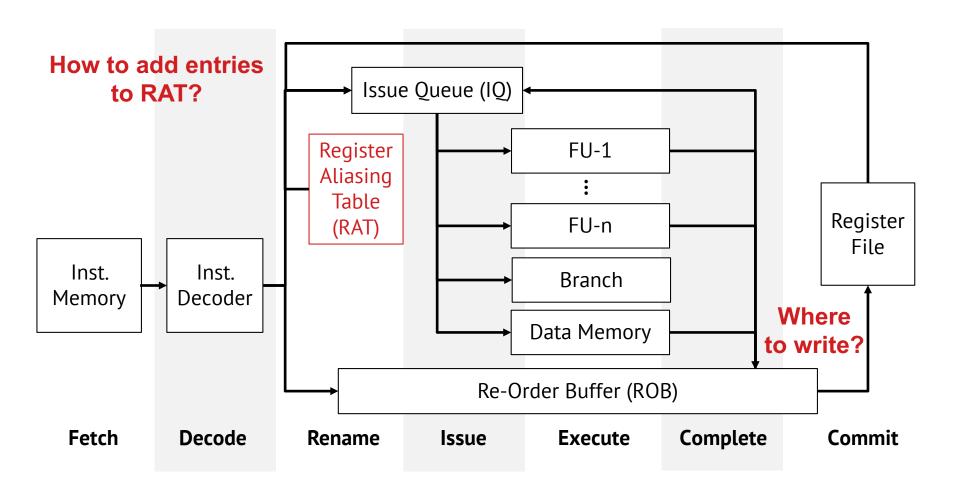
Register aliasing table for fast lookup



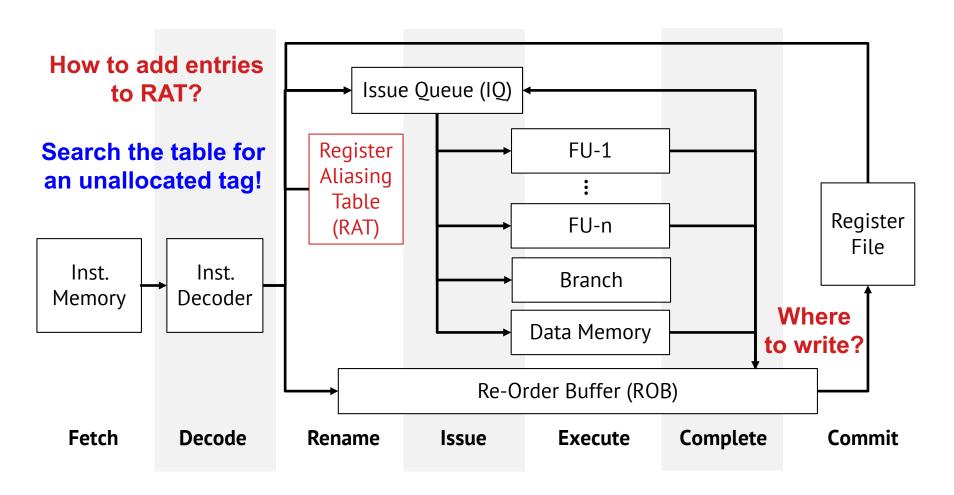
□ Register aliasing table for fast lookup



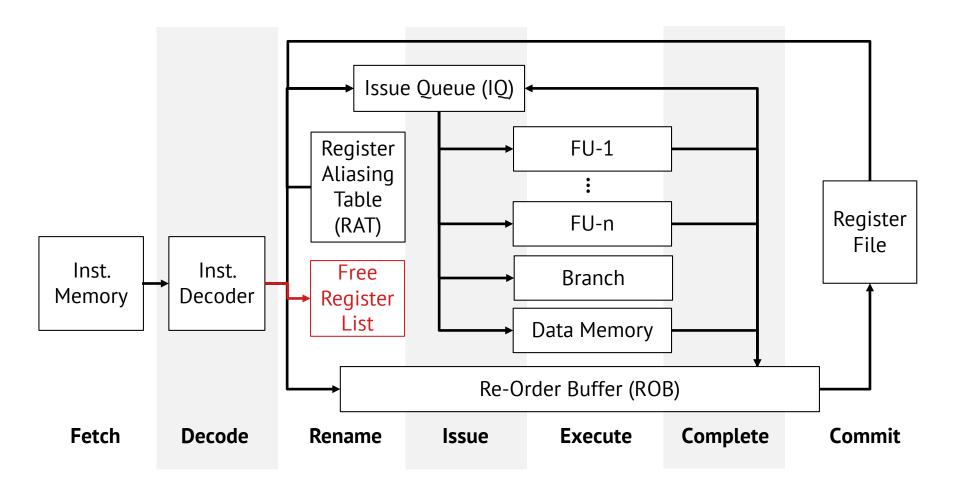
Register aliasing table for fast lookup



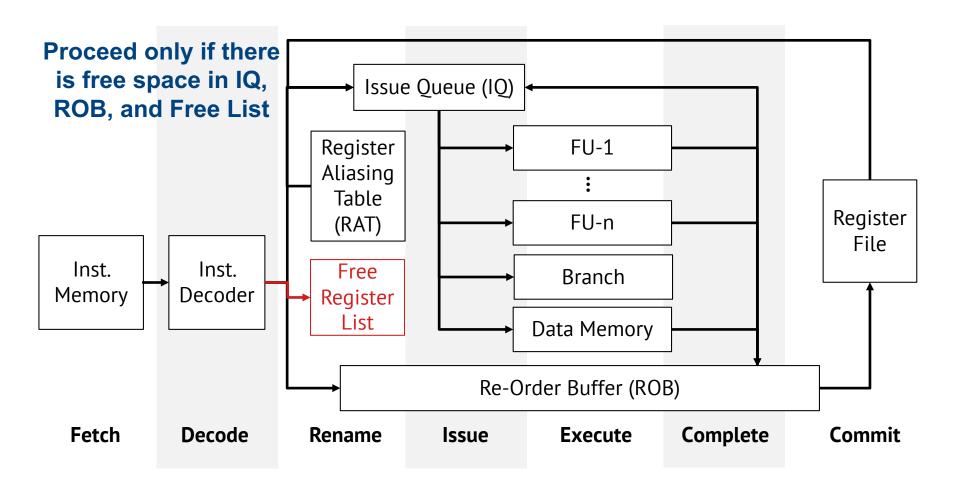
Register aliasing table for fast lookup



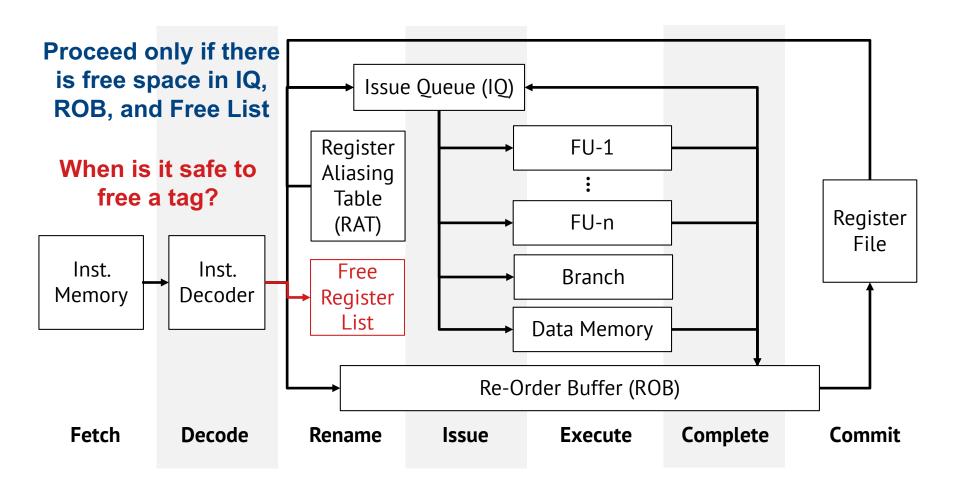
Free register list for fast register renaming



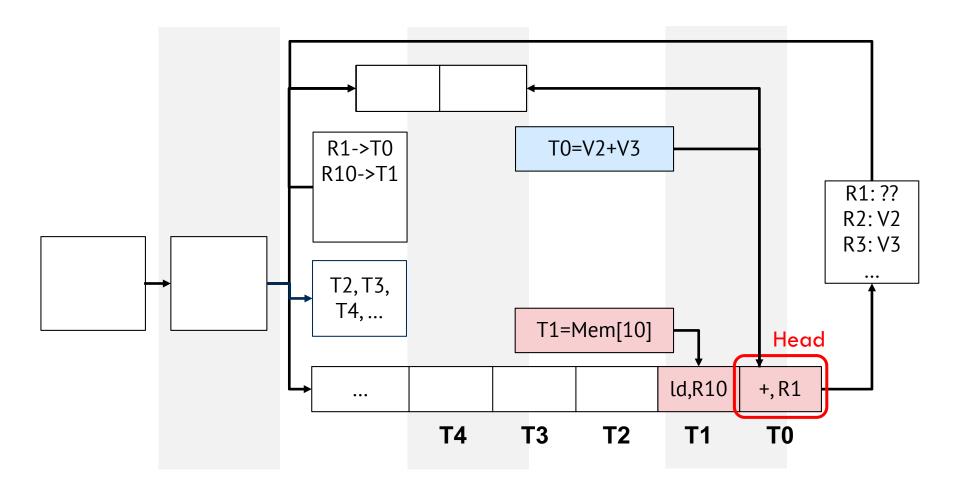
Free register list for fast register renaming



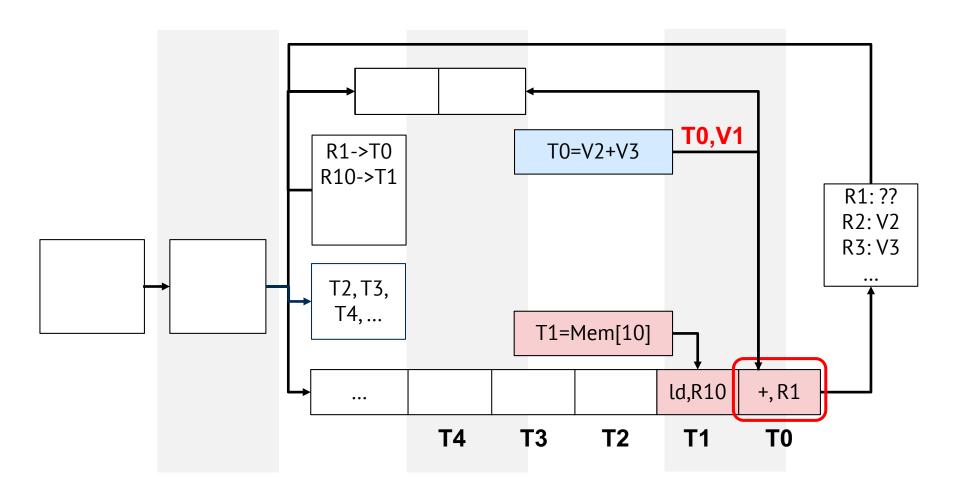
Free register list for fast register renaming



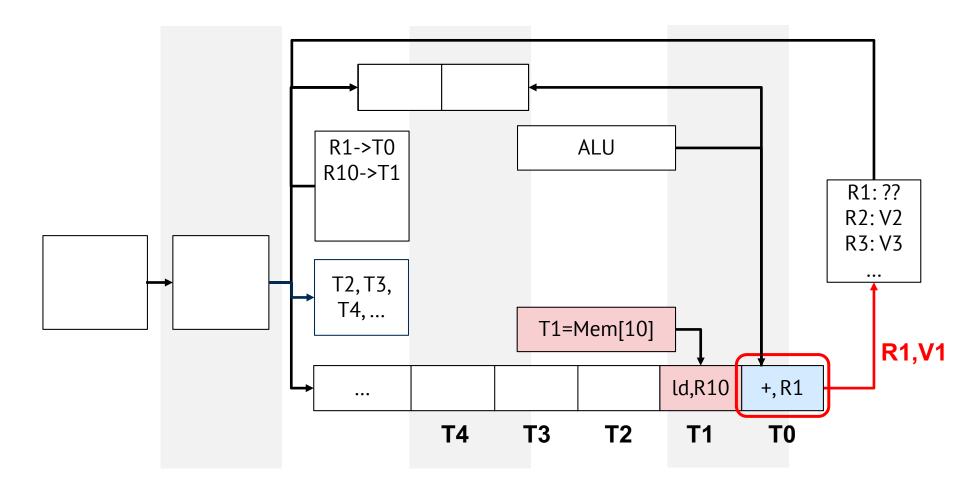
□ Update value in ROB



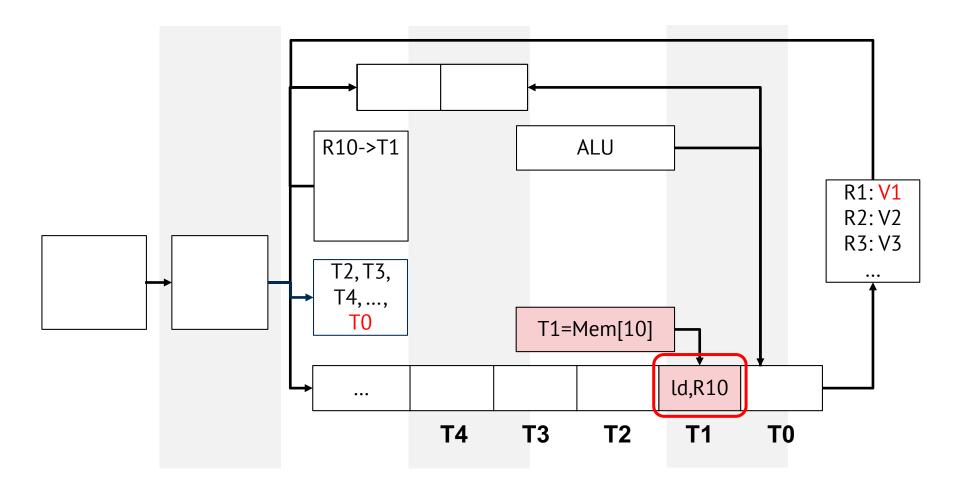
□ Update value in ROB



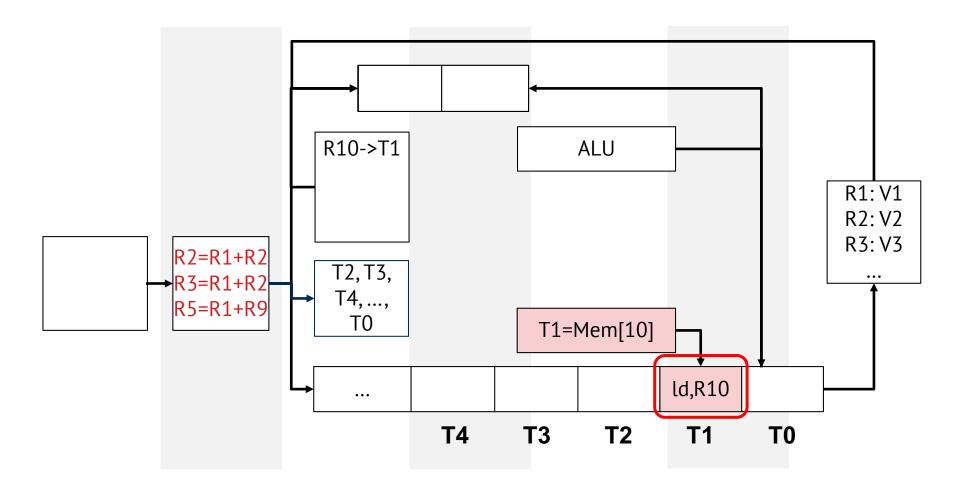
□ Register file write



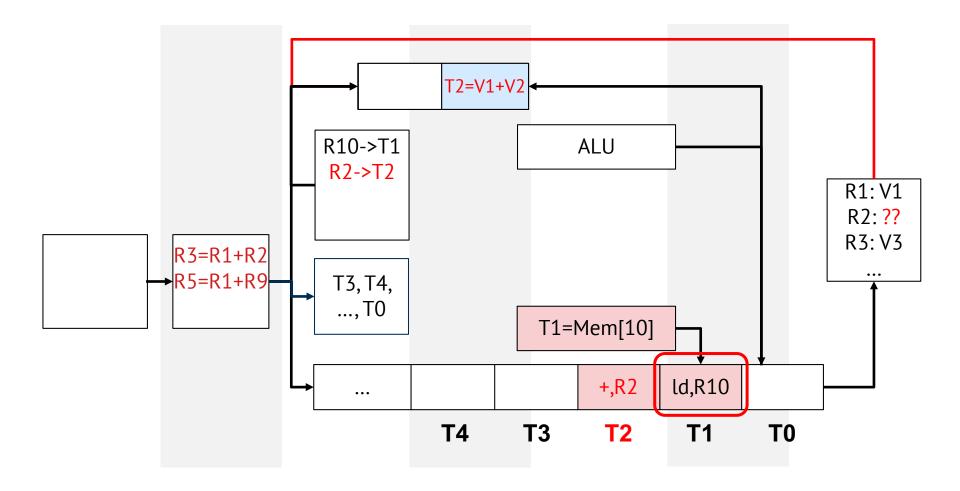
□ Update tables and ROB



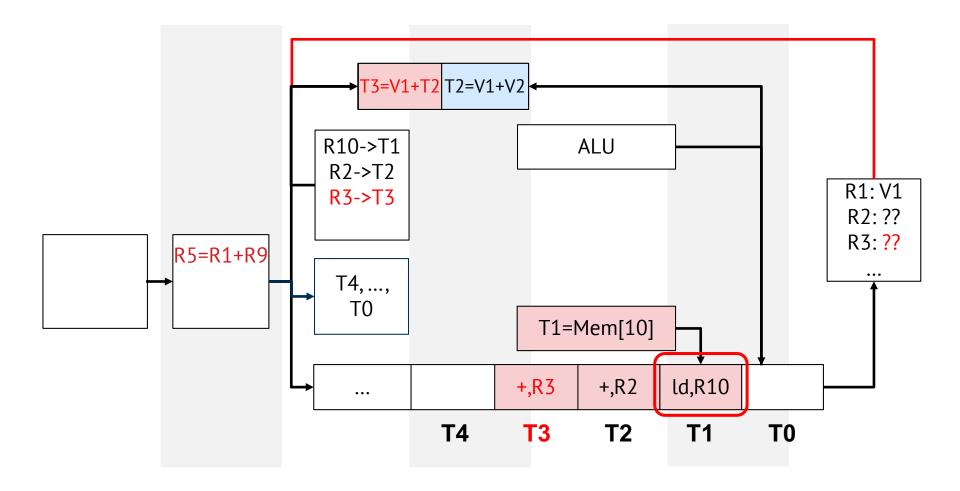
□ Allocate entries on ROB, IQ, and FL



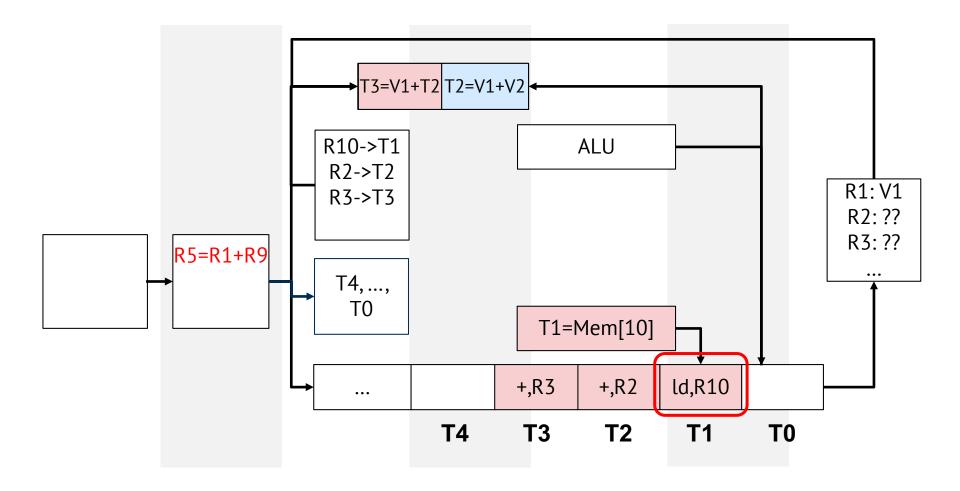
□ Allocate entries on ROB, IQ, and FL



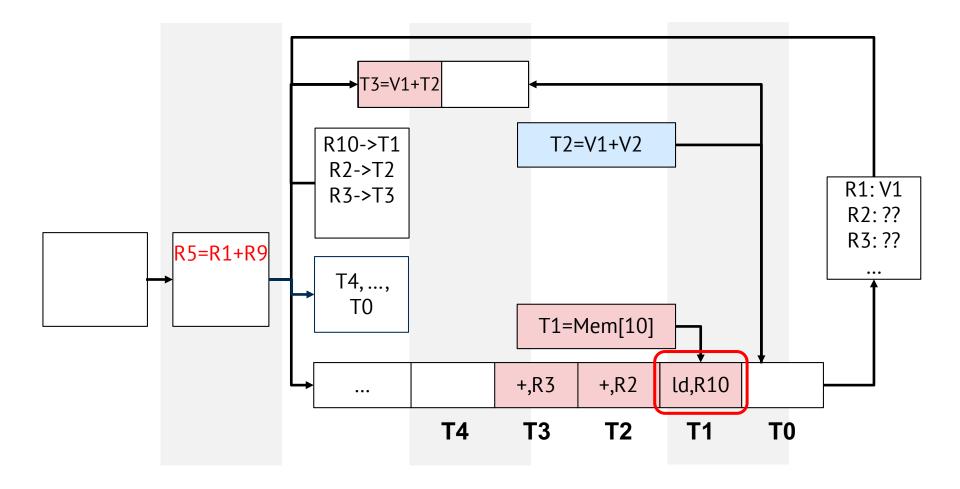
□ Allocate entries on ROB, IQ, and FL



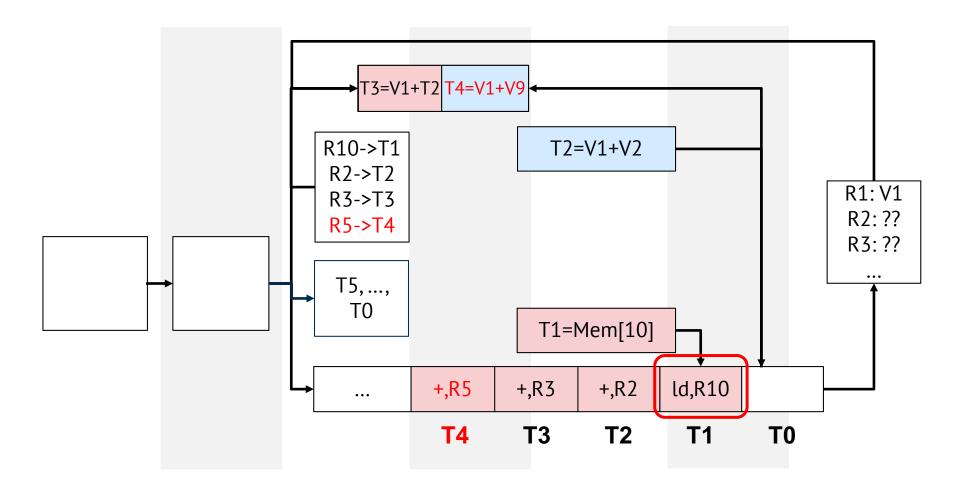
□ Instruction has to wait for free resources



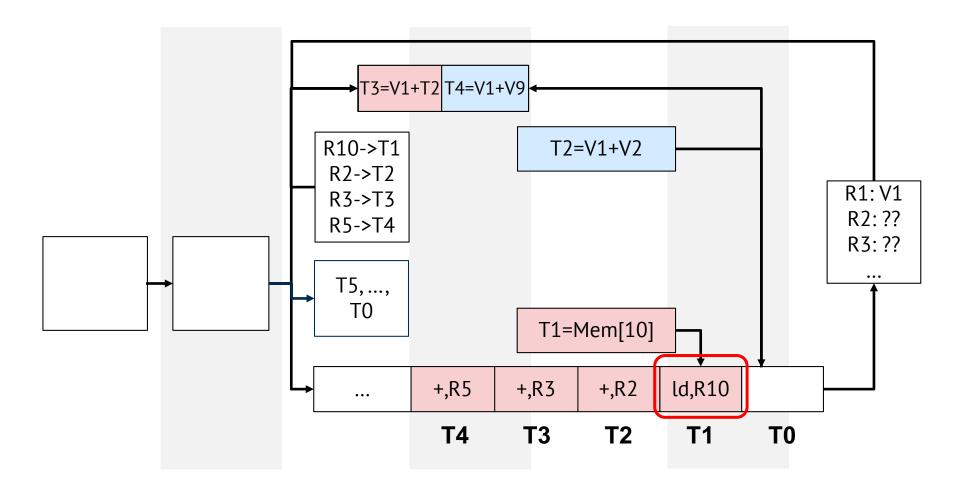
□ Issue ready instruction if free FU exists



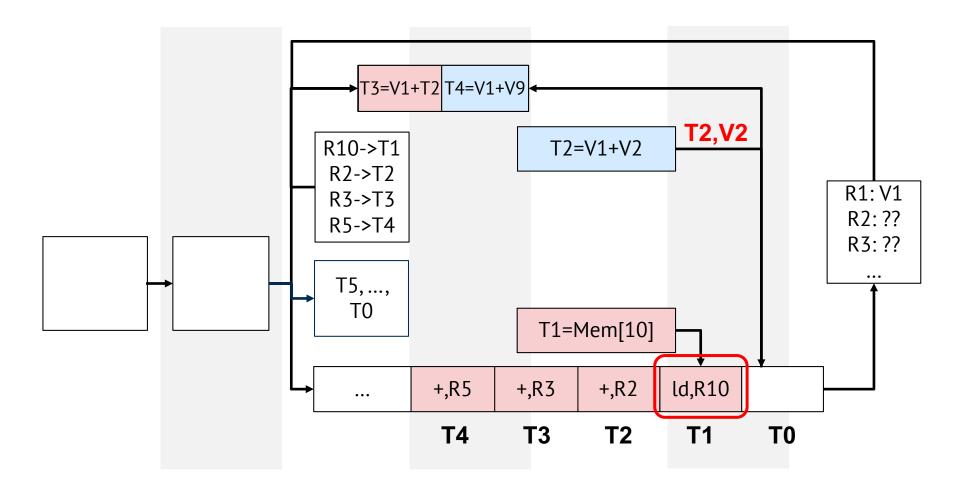
□ Out-of-order issue is now possible



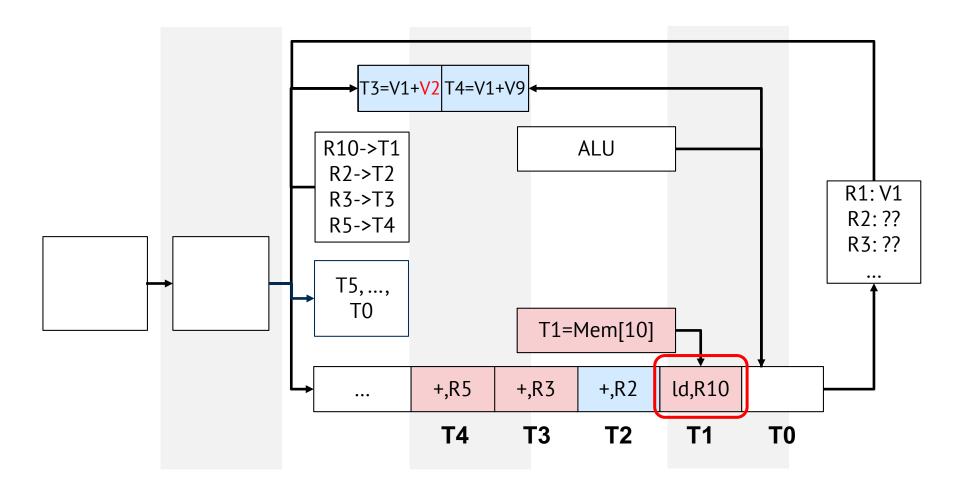
□ Out-of-order issue is now possible



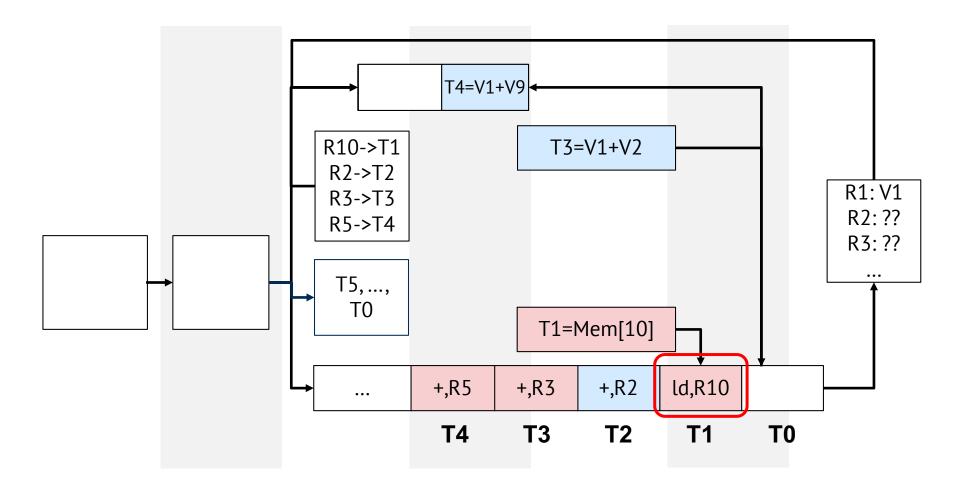
□ Out-of-order issue is now possible



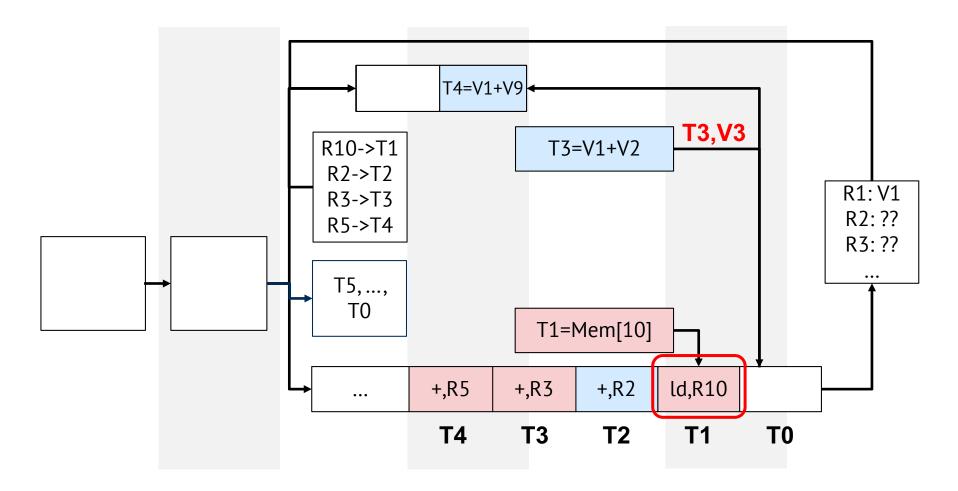
□ Wakeup and select



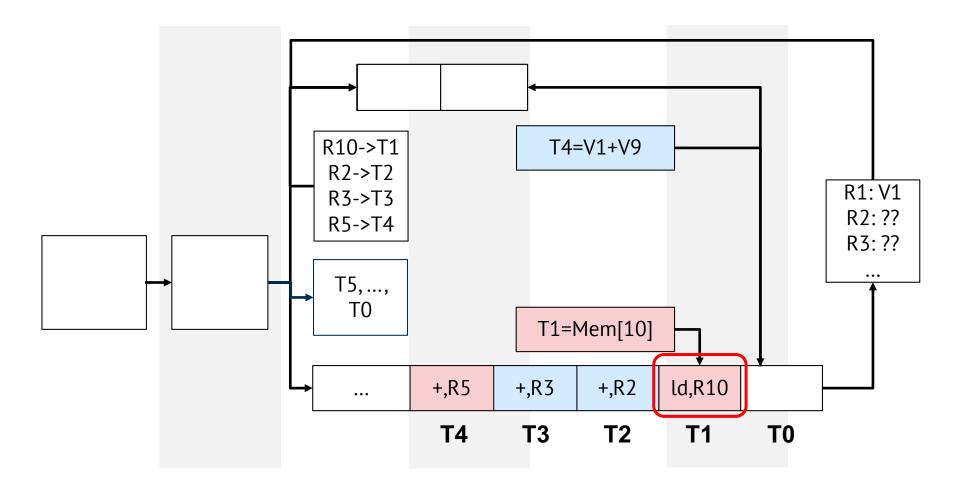
Keep the program order to avoid starvation



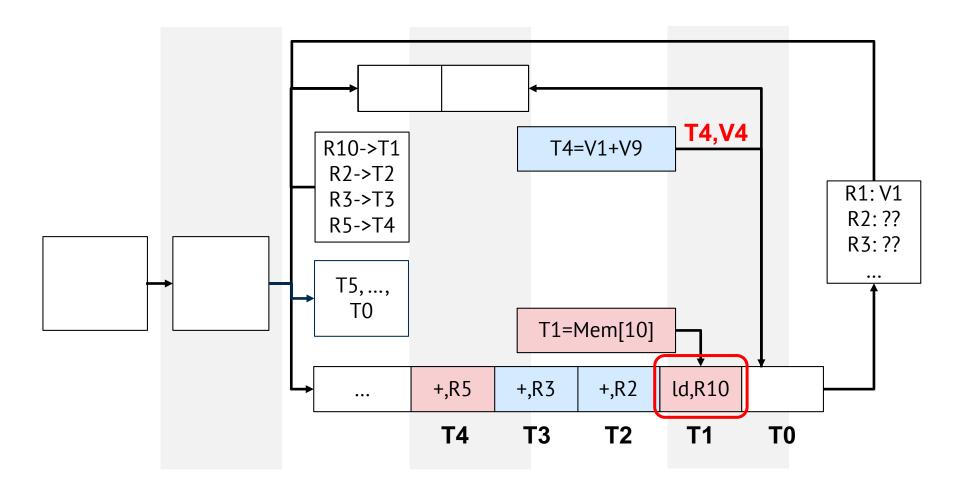
Keep the program order to avoid starvation



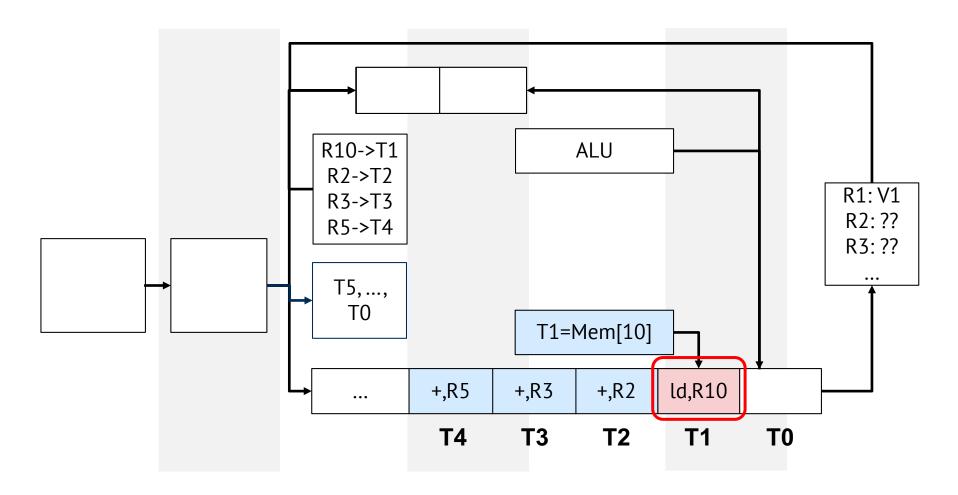
□ Issue ready instructions



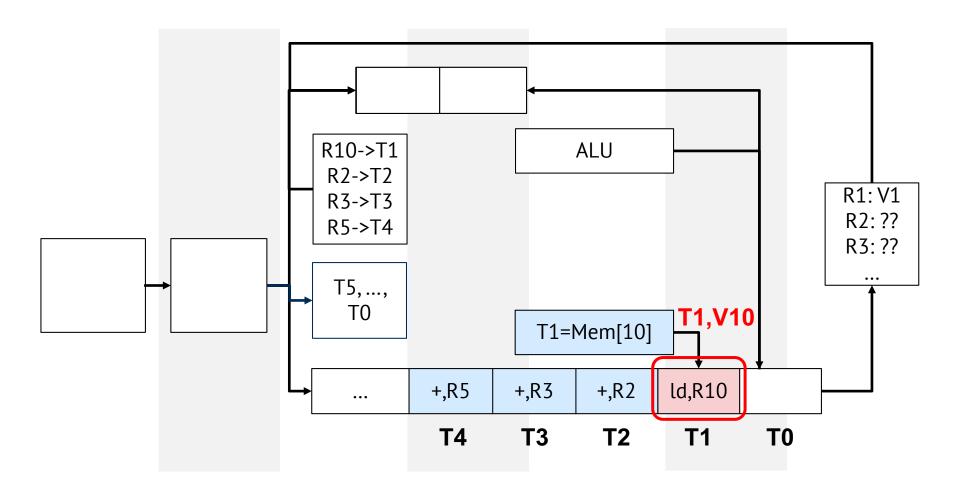
□ Issue ready instructions



#### □ Update ROB

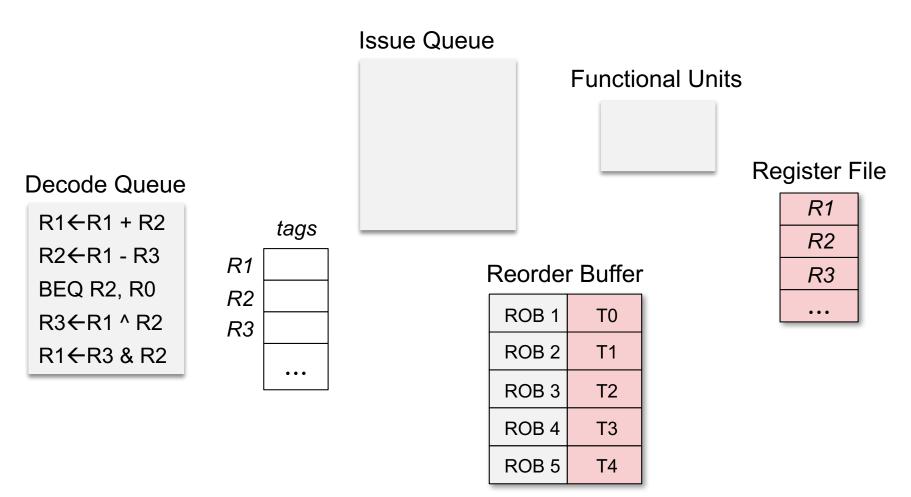


#### □ Update ROB



# Register Renaming Example

■ Where values are stored?



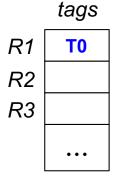
# Register Renaming Example

#### ■ Where values are stored?



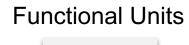
Decode Queue

R1←R1 + R2 R2←R1 - R3 BEQ R2, R0 R3←R1 ^ R2 R1←R3 & R2





Issue Queue





R1	1
R2	2
R3	3
•••	

_		D 66	
≺ео	rder	Buffe	er

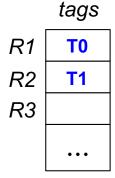
ROB 1	T0	R
ROB 2	T1	ı
ROB 3	T2	ı
ROB 4	Т3	ı
ROB 5	T4	ı

### ■ Where values are stored?

#### Issue Queue

T0←R1 + R2 T1←T0 - R3 **Functional Units** 

#### **Decode Queue**



### Reorder Buffer

ROB 1	T0	R1
ROB 2	T1	R2
ROB 3	T2	
ROB 4	Т3	
ROB 5	T4	

### Register File

5.4
R1
R2
R3
• • •

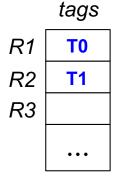
### ■ Where values are stored?

#### Issue Queue

T0←R1 + R2 T1←T0 - R3 BEQ T1, R0 **Functional Units** 

### **Decode Queue**

R1←R1 + R2 R2←R1 - R3 BEQ R2, R0 R3←R1 ^ R2 R1←R3 & R2



### Reorder Buffer

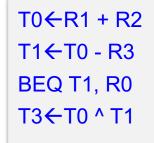
ROB 1	T0	R
ROB 2	T1	R
ROB 3	T2	
ROB 4	Т3	ı
ROB 5	T4	ı

### Register File

R1	
R2	
R3	
•••	

### ■ Where values are stored?

#### Issue Queue







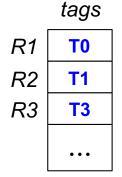
R2

R3

Reorde	r Buffe	r	
DOD 4	TO	D4	

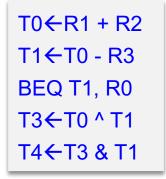
ROB 1	T0	R1
ROB 2	T1	R2
ROB 3	T2	
ROB 4	Т3	R3
ROB 5	T4	

### **Decode Queue**

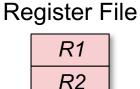


### □ Where values are stored?

### Issue Queue



### **Functional Units**



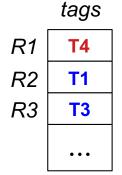
R3

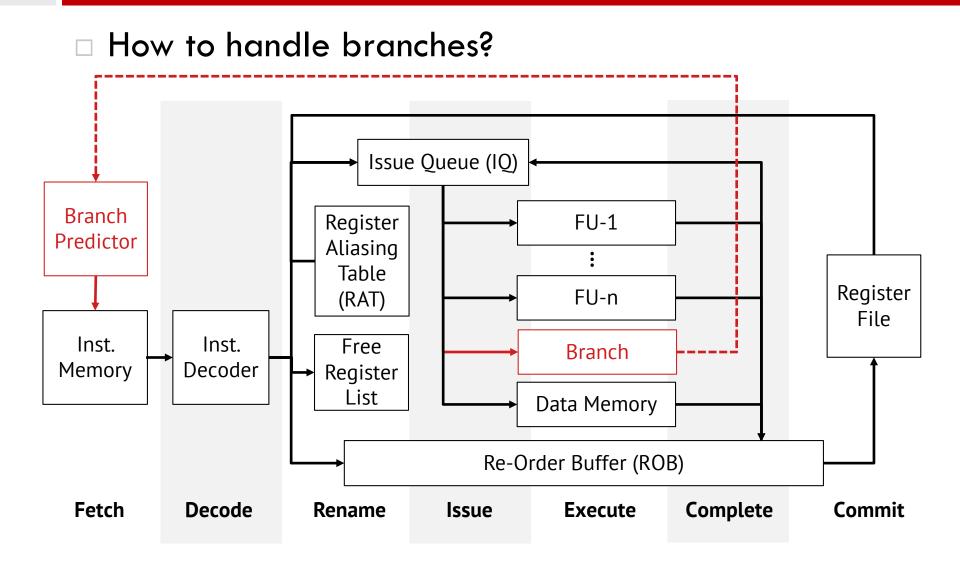
• • •

## Reorder Buffer

ROB 1	T0	R1
ROB 2	T1	R2
ROB 3	T2	
ROB 4	Т3	R3
ROB 5	T4	R1

### **Decode Queue**





### Revisit Branch Prediction

□ Problem: find the average number of stall cycles caused by branches in a pipeline, where branch misprediction penalty is 20 cycles, branch predictor accuracy is 90%, and branch target buffer hit rate is 80%. Every fifth instruction is a branch; 30% of branches are actually taken.

### Revisit Branch Prediction

□ Problem: find the average number of stall cycles caused by branches in a pipeline, where branch misprediction penalty is 20 cycles, branch predictor accuracy is 90%, and branch target buffer hit rate is 80%. Every fifth instruction is a branch; 30% of branches are actually taken.

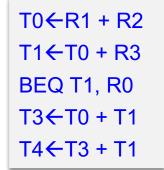
- $\blacksquare$  Average misses = 1- (0.3x0.9x0.8 + 0.7x0.9) = 0.151
- Average stalls = 20x0.2x0.151 = 0.6

## Speculated Execution

- Problem: branch may significantly limit performance
  - consumer of a load or long latency instructions
- Solution: speculative instruction execution
  - Fetch and decode instructions speculatively
  - Issue and execute speculative instructions
  - Branch resolution
    - Nullify the impact of speculative instructions if mispredicted
    - Commit speculative instructions (writes to register file/memory) only if prediction was correct

### Squash all mispredicted entries

#### Issue Queue





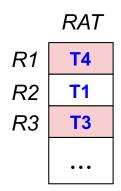


### Register File

R1
R2
R3
• • •

### **Decode Queue**

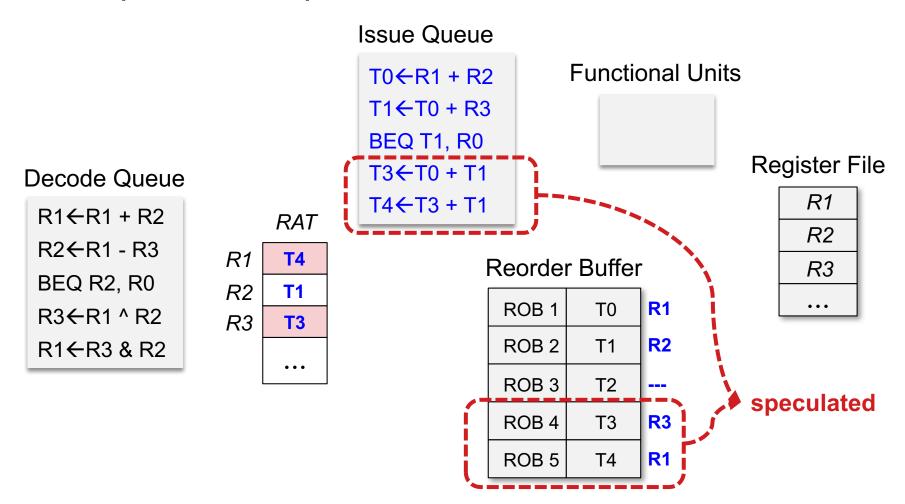
R1←R1 + R2
R2←R1 - R3
BEQ R2, R0
R3←R1 ^ R2
R1←R3 & R2



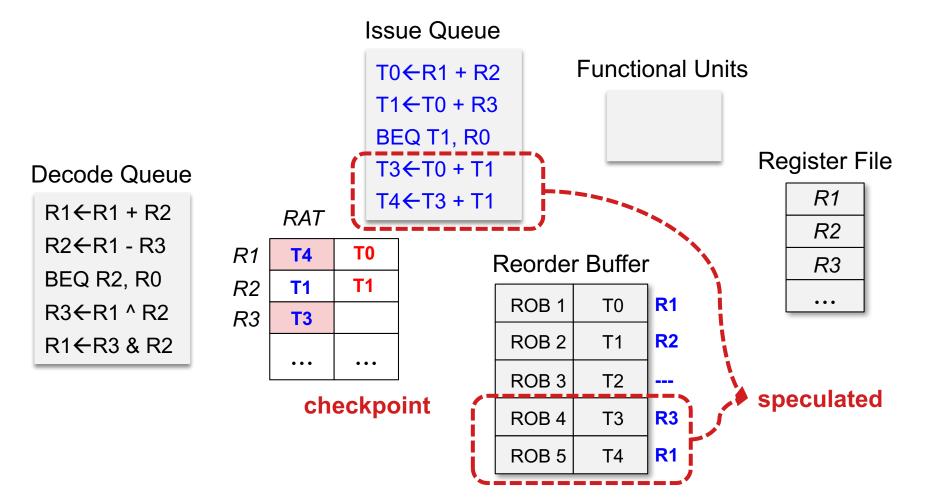
### Reorder Buffer

ROB 1	T0	R1
ROB 2	T1	R2
ROB 3	T2	
ROB 4	Т3	R3
ROB 5	T4	R1

Squash all mispredicted entries

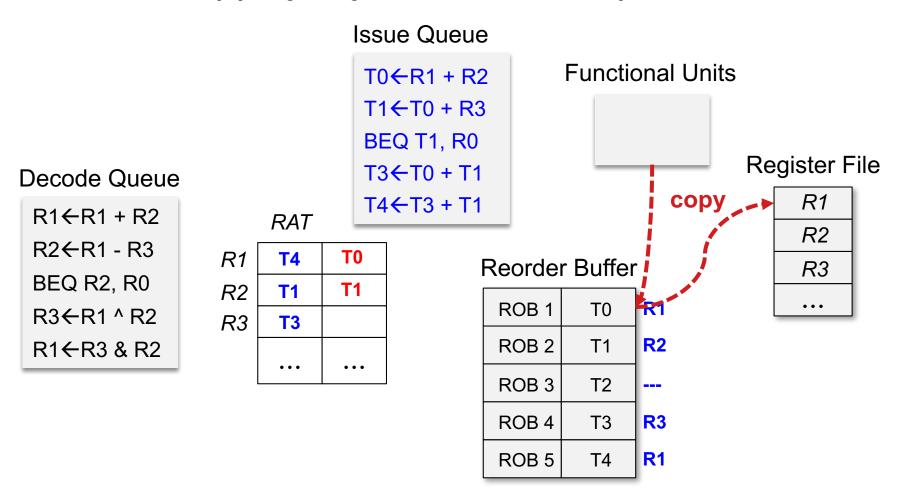


Squash all mispredicted entries



## Physical Register File

Avoid copying register values multiple times



## Physical Register File

Front RAT

**P4** 

### Avoid copying register values multiple times

Note1: only a subset of the Phy. Reg. file is committed at any time.

### Decode Queue

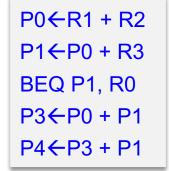


R2 P1 R3 P3 R3 R2 R2 ...

R1

Note2: no need for storing values in ROB or IQ

#### Issue Queue



**Functional Units** 



P0

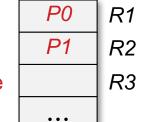
P1

P2 P3

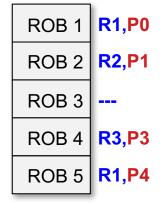
**P4** 

P5

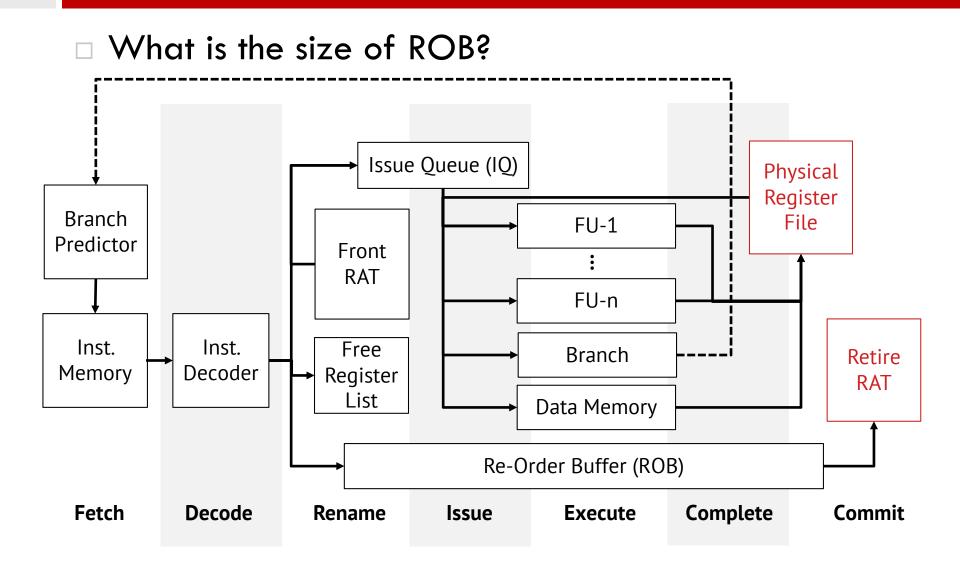
Retire RAT



Reorder Buffer Phy. Reg. File

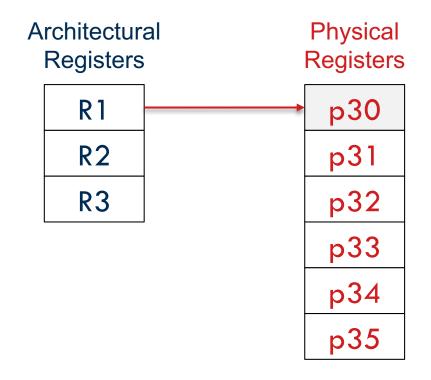


## Double RAT Architecture



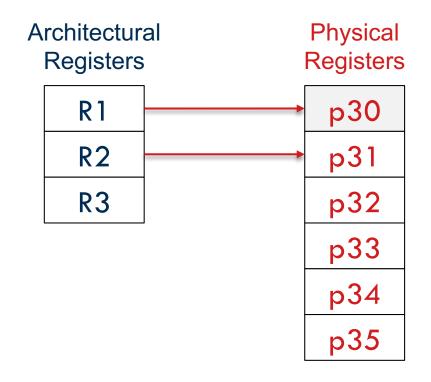
□ Example: when is it safe to free p30 (R1)?

ADD R1, R2, R3
SUB R2, R1, R3
...
ADD R3, R1, R2
...
SUB R1, R3, R2
ADD R2, R1, R3



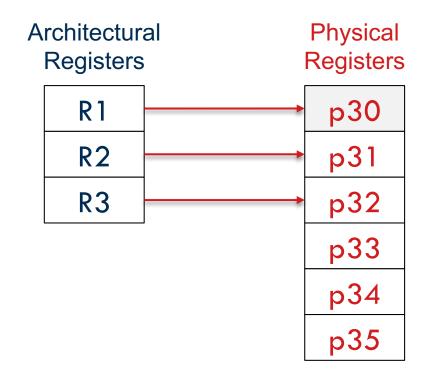
□ Example: when is it safe to free p30 (R1)?

ADD R1, R2, R3
SUB R2, R1, R3
...
ADD R3, R1, R2
...
SUB R1, R3, R2
ADD R2, R1, R3



□ Example: when is it safe to free p30 (R1)?

ADD R1, R2, R3
SUB R2, R1, R3
...
ADD R3, R1, R2
...
SUB R1, R3, R2
ADD R2, R1, R3



□ Example: when is it safe to free p30 (R1)?

