Performance & Power

Jin-Soo Kim (*jinsookim@skku.edu*)
Computer Systems Laboratory
Sungkyunkwan University
http://csl.skku.edu

Performance

Elements of CPU Performance

- Cycle time
- CPU pipeline
- Memory systems

Pipelining

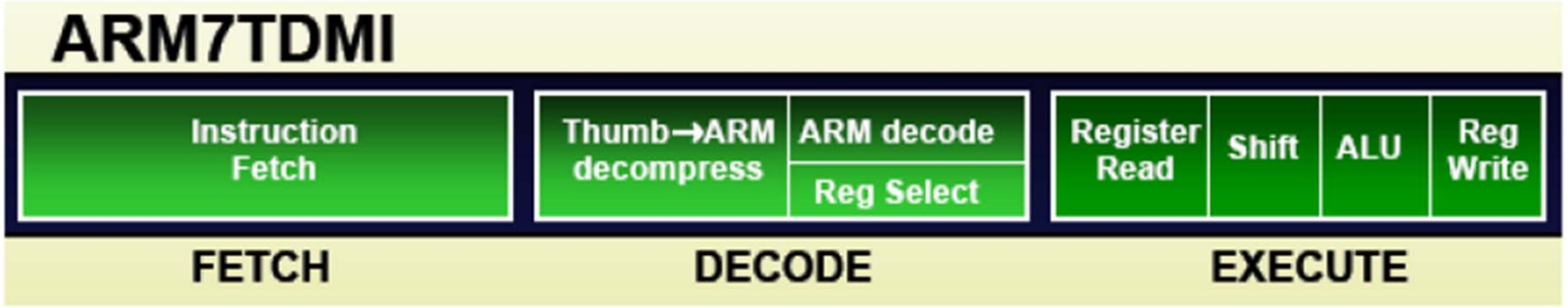
- Several instructions are executed simultaneously at different stages of completion
- Various conditions can cause pipeline bubbles that reduce utilization:
  - branches
  - memory system delays
  - dependencies among instructions

Performance Measures

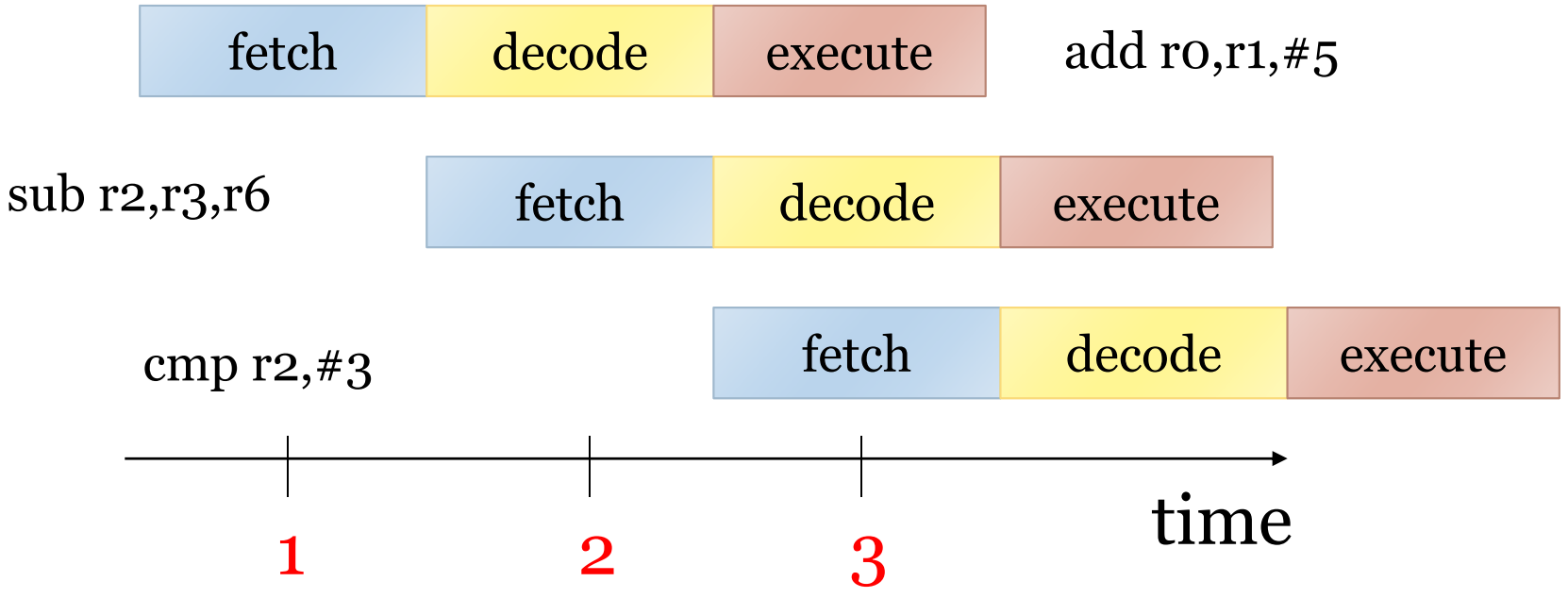
- Latency
  - Time it takes for an instruction to get through the pipeline
- Throughput
  - The number of instructions executed per time period
- Pipelining increases throughput without reducing latency

ARM7 Pipeline

- ARM7 has 3-stage pipeline:
  - Fetch instruction from memory
  - Decode opcode and operands
  - Execute



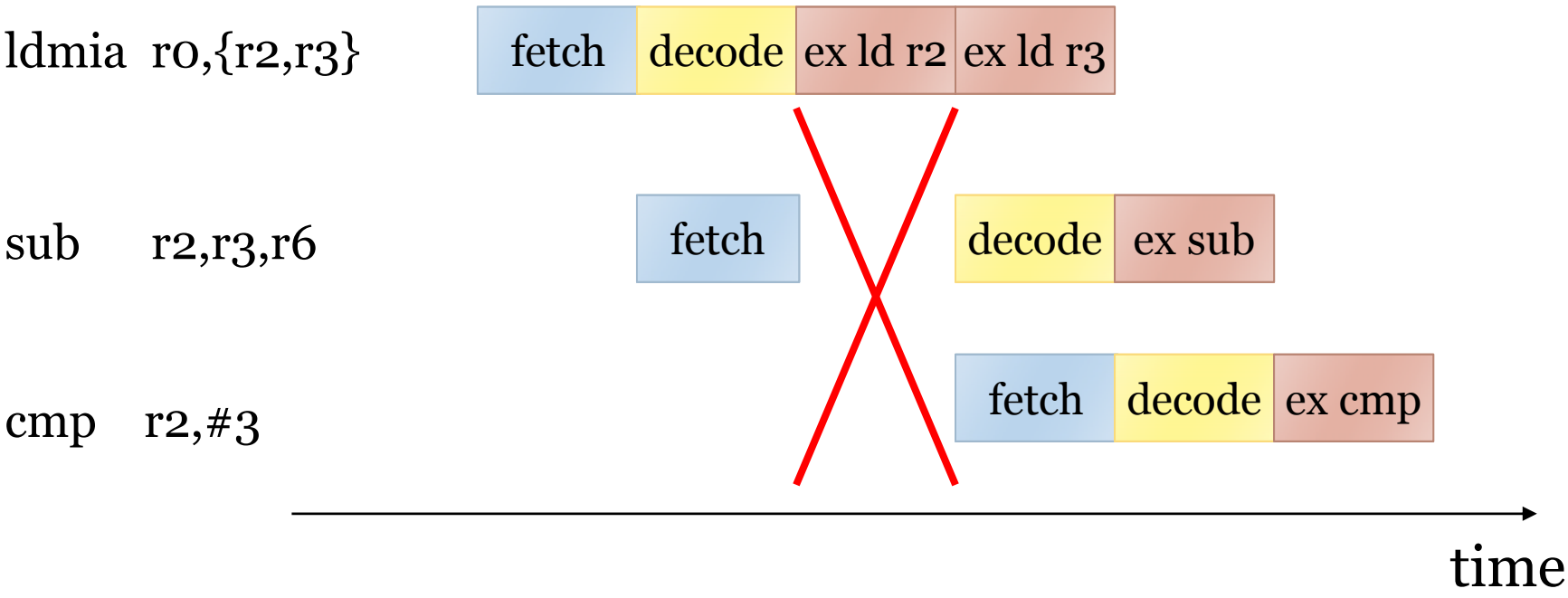
ARM7 Pipeline Execution



Pipeline Stalls

- If every step cannot be completed in the same amount of time, pipeline stalls.
- Bubbles introduced by stall increase latency, reduce throughput

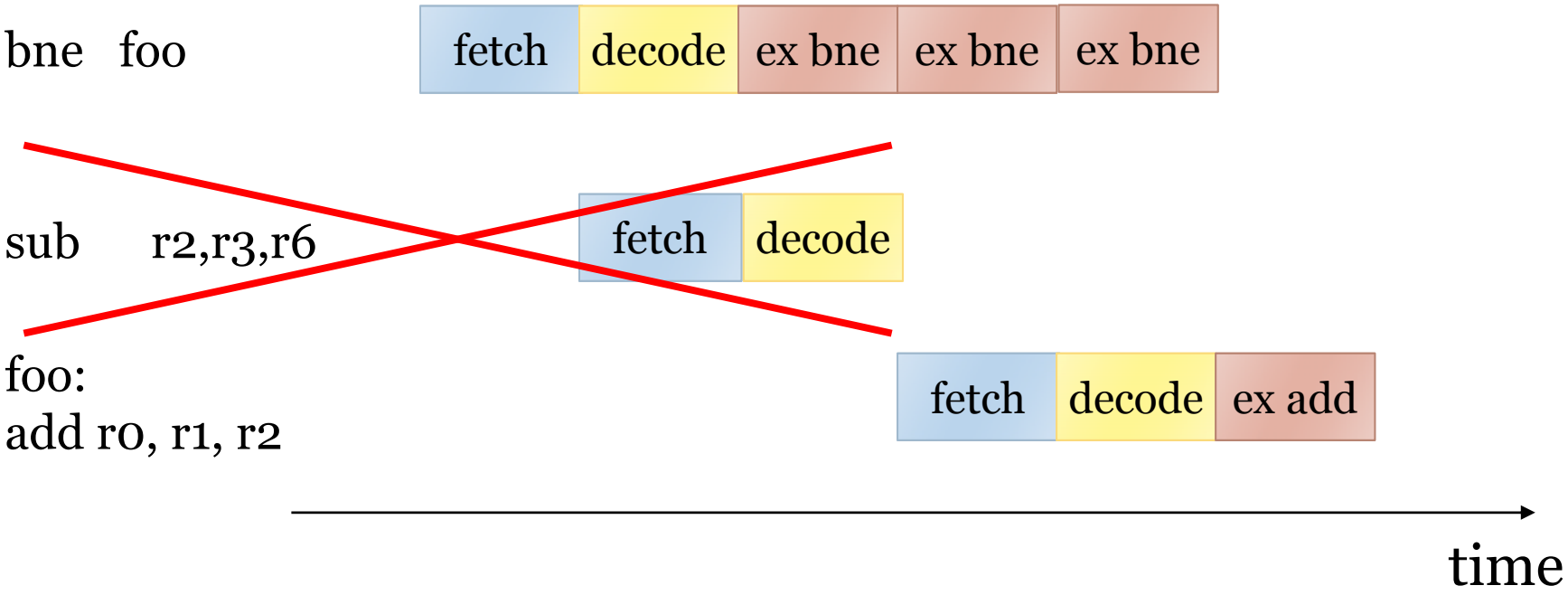
ARM Multi-cycle LDMIA



Control Stalls

- Branches often introduce stalls (branch penalty)
  - Stall time may depend on whether branch is taken
- May have to squash instructions that already started executing
- Don't know what to fetch until condition is evaluated

ARM Pipelined Branch



Delayed Branch

- To increase pipeline efficiency, delayed branch mechanism requires *n* instructions after branch always executed whether branch is executed or not.

- Use NOP (No Operation) instruction if there are not enough instructions to fill the delayed branch window.

Example: FIR Filter (1)

- Execution time of FIR filter?
- Only branch in loop test may take more than one cycle
  - BLT loop takes 1 cycle best case, 3 cycles worst case

```
for (i = 0, f = 0; i < N; i++)
    f = f + c[i] * x[i];
```

```
loop
    LDR     R4, [R3, R8]    ; R4 = c[i]
    LDR     R6, [R5, R8]    ; R6 = x[i]
    MUL     R4, R4, R6      ; R4 = c[i]*x[i]
    ADD     R2, R2, R4      ; f += c[i]*x[i]
    ADD     R8, R8, #4      ; R8 += 4
    ADD     R0, R0, #1      ; i += 1
    CMP     R0, R1          ; if (i < N)
    BLT     loop            ; Continue
```

Example: FIR Filter (2)

Block	Variable	# instructions	# cycles
Initialization	t_{init}	7	7
Body	t_{body}	4	4
Update	t_{update}	2	2
Test	t_{test}	2	[2, 4]

$$t_{loop} = t_{init} + N(t_{body} + t_{update}) + (N\text{-}1)\, t_{test,worst} + t_{test,best}$$

Loop test succeeds is worst case

Loop test fails is best case

Memory System Performance

- Caches introduce indeterminacy in execution time
- Cache performance depends on order of execution
- Cache miss penalty: added time due to a cache miss

Type of Cache Misses

- Compulsory miss
  - Location has not been referenced before
- Conflict miss
  - Two locations are fighting for the same block
- Capacity miss
  - Working set is too large

Scratch pad memories

- Alternative to cache:
  - Software determines what is stored in scratch pad.
- Provides predictable behavior at the cost of software control.
- In some CPUs/DSPs, cache can be configured as scratch pad memory.

Power Consumption

CPU Power Consumption

- Most modern CPUs are designed with power consumption in mind to some degree.

- Energy = Power x Time
  - Heat depends on power consumption
  - Battery life depends on energy consumption

CMOS Power Consumption

- Voltage drops
  - Power consumption proportional to V^2
- Toggling
  - More activity means more power
- Leakage
  - Basic circuit characteristics
  - Can be eliminated by disconnecting power

CPU Power-saving Strategies

- Reduce power supply voltage
- Run at lower clock frequency
- Disable function units with control signals when not in use
- Disconnect parts from power supply when not in use

Power Management Styles

- Static power management
  - Does not depend on CPU activity
  - e.g., user-activated power-down mode
- Dynamic power management
  - Based on CPU activity
  - e.g., disabling off function units

StrongARM SA-1100

- Processor takes two supplies:
  - VDD is main 3.3V supply
  - VDDX is 1.5V
- Three power modes
  - Run: normal operation
  - Idle: stops CPU clock, with logic still powered
  - Sleep: shuts off most of chip activity; 3 steps, each about 30µs; wakeup takes > 10 ms

SA-1100 Power State Machine