INSTRUCTION PIPELINING (I)

1. The Instruction Cycle
2. Instruction Pipelining
3. Pipeline Hazards
4. Structural Hazards
5. Data Hazards
6. Control Hazards

The Instruction Cycle

- Fetch instruction (FI)
- Decode (DI)
- Fetch operand
- Execute instruction (EI)
- Write back operand (WO)

Instruction Pipelining

- Instruction execution is extremely complex and involves several operations which are executed successively (see slide 2). This implies a large amount of hardware, but only one part of this hardware works at a given moment.

- Pipelining is an implementation technique whereby multiple instructions are overlapped in execution. This is solved without additional hardware but only by letting different parts of the hardware work for different instructions at the same time.

- The pipeline organization of a CPU is similar to an assembly line: the work to be done in an instruction is broken into smaller steps (pieces), each of which takes a fraction of the time needed to complete the entire instruction. Each of these steps is a pipe stage (or a pipe segment).

- Pipe stages are connected to form a pipe:

  Stage 1 → Stage 2 → ... → Stage n

- The time required to execute a stage and move to the next is called a machine cycle (this is one or several clock cycles). The execution of one instruction takes several machine cycles as it passes through the pipeline.

Acceleration by Pipelining

Two stage pipeline:

- FI: fetch instruction
- EI: execute instruction

(machine) cycle → 1 2 3 4 5 6 7 8

Instr. i
Instr. i+1
Instr. i+2
Instr. i+3
Instr. i+4
Instr. i+5
Instr. i+6

We consider that each instruction takes execution time $T_{ex}$.

Execution time for the 7 instructions, with pipelining:

$(T_{ex}/2)\times 8 = 4\times T_{ex}$

- Acceleration: $7\times T_{ex} / 4\times T_{ex} = 7/4$
Acceleration by Pipelining (cont’d)

Six stage pipeline (see also slide 2):
- FI: fetch instruction
- FO: fetch operand
- DI: decode instruction
- EI: execute instruction
- CO: calculate operand address
- WO: write operand

Execution time for the 7 instructions, with pipelining:
\[(T_{ex}/6) \times 12 = 2 \times T_{ex}\]

- Acceleration: \(7 \times T_{ex} / 2 \times T_{ex} = 7/2\)
- After a certain time (N-1 cycles) all the N stages of the pipeline are working: the pipeline is filled. Now, theoretically, the pipeline works providing maximal parallelism (N instructions are active simultaneously).

\[\tau: \text{duration of one machine cycle} \]
\[n: \text{number of instructions to execute} \]
\[k: \text{number of pipeline stages} \]
\[T_{k,n}: \text{total time to execute } n \text{ instructions on a pipeline with } k \text{ stages} \]
\[S_{k,n}: \text{(theoretical) speedup produced by a pipeline with } k \text{ stages when executing } n \text{ instructions} \]

On a non-pipelined processor each instruction takes \(T_n = n \times k \times \tau\)

\[S_{k,n} = \frac{T_n}{T_{k,n}} = \frac{n \times k \times \tau}{(k + (n - 1)) \times \tau} = \frac{n \times k}{k + (n - 1)}\]

- Apply the formula to the examples on the previous two slides!
- For large number of instructions \((n \rightarrow \infty)\) the speedup approaches \(k\) (number of stages).

Pipeline Hazards

- Pipeline hazards are situations that prevent the next instruction in the instruction stream from executing during its designated clock cycle. The instruction is said to be stalled. When an instruction is stalled, all instructions later in the pipeline than the stalled instruction are also stalled. Instructions earlier than the stalled one can continue. No new instructions are fetched during the stall.

- Types of hazards:
  1. Structural hazards
  2. Data hazards
  3. Control hazards

80486 and Pentium: five-stage pipeline for integer instr.

PowerPC: eight-stage pipeline for FP instr.
four-stage pipeline for integer instr.

six-stage pipeline for FP instr.
### Structural Hazards

- Structural hazards occur when a certain resource (memory, functional unit) is requested by more than one instruction at the same time.

Instruction ADD R4,X fetches in the FO stage operand X from memory. The memory doesn’t accept another access during that cycle.

- **Penalty:** 1 cycle

- Certain resources are duplicated in order to avoid structural hazards. Functional units (ALU, FP unit) can be pipelined themselves in order to support several instructions at a time. A classical way to avoid hazards at memory access is by providing separate data and instruction caches.

### Data Hazards

- We have two instructions, I1 and I2. In a pipeline the execution of I2 can start before I1 has terminated. If in a certain stage of the pipeline, I2 needs the result produced by I1, but this result has not yet been generated, we have a data hazard.

I1: MUL R2,R3 \( R2 \leftarrow R2 \times R3 \)
I2: ADD R1,R2 \( R1 \leftarrow R1 + R2 \)

- **Penalty:** 2 cycles

### Data Hazards (cont’d)

- Some of the penalty produced by data hazards can be avoided using a technique called **forwarding** (bypassing).

- The ALU result is always fed back to the ALU input. If the hardware detects that the value needed for the current operation is the one produced by the previous operation (but which has not yet been written back) it selects the forwarded result as the ALU input, instead of the value read from register or memory.

- **Penalty:** 1 cycle

### Control Hazards

- Control hazards are produced by branch instructions.

  **Unconditional branch**
  
  - After the FO stage of the branch instruction the address of the target is known and it can be fetched

- **Penalty:** 3 cycles
Control Hazards (cont’d)

Conditional branch

ADD R1,R2  \[ R1 \leftarrow R1 + R2 \]
BEZ TARGET  \[ \text{branch if zero} \]

\[ \text{target} \]

Penalty: 3 cycles

Branch is taken

At this moment, both the condition (set by ADD) and the target address are known.

cycle → 1 2 3 4 5 6 7 8 9 10 11 12
ADD R1,R2
BEZ TARGET

Penalty: 2 cycles

Branch not taken

At this moment the condition is known and instr+1 can go on.

cycle → 1 2 3 4 5 6 7 8 9 10 11 12
ADD R1,R2
BEZ TARGET

Summary

- Instructions are executed by the CPU as a sequence of steps. Instruction execution can be substantially accelerated by instruction pipelining.

- A pipeline is organized as a succession of N stages. At a certain moment N instructions can be active inside the pipeline.

- Keeping a pipeline at its maximal rate is prevented by pipeline hazards. Structural hazards are due to resource conflicts. Data hazards are produced by data dependencies between instructions. Control hazards are produced as consequence of branch instructions.