

Lecture # 2

Department of Electrical Engineering

The Role of Performance

Chapter 2

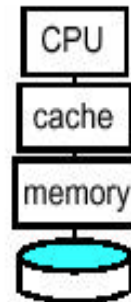
saeed_zafar@comsats.edu.pk

Comsats Institute of Information Technology

1

Performance Measurement

- The computer performance can't be defined by single equation.
- Cache bandwidth.
- Main memory bandwidth.
- I/O performance
- Also depends on Application.
- Performance is multidimensional
 - CPU execution time
 - by instructions or sequence(floating point, integer, branch performance)
 - **Response time/execution time** :The Time between start & completion of task.
 - **Throughput** : The total amount of work done in a given time.



2

Which of these airplanes has the best performance?

<u>Airplane</u>	<u>Passengers</u>	<u>Range (mi)</u>	<u>Speed (mph)</u>	<u>Pasng throughput</u> <u>Pasng x mph</u>
Boeing 737-100	101	630	598	60,398
Boeing 747	470	4150	610	286,700
BAC/Sud Concorde	132	4000	1350	178,200
Douglas DC-8-50	146	8720	544	79,424

- How much faster is the Concorde compared to the 747?

3

Performance

- To maximize the Performance, minimize the response time or execution time for some task.

- For some program running on machine X,

$$\text{Performance}_X = 1 / \text{Execution time}_X$$

- “For two m/c X & Y, performance of X is n times faster than Y”.

$$\text{Performance}_X > \text{Performance}_Y$$

$$\text{Performance}_X / \text{Performance}_Y = n$$

4

Performance

- Problem:
 - machine A runs a program in 20 seconds
 - machine B runs the same program in 25 seconds.

How much faster A than B?

5

Performance measurement

- Computer which takes least time to perform same amount of work is fastest.
Time is Different
- **Clock cycle** : Discrete time interval determining when events take place in HW.
- **CPU execution time or CPU time** : The time CPU spends computing for tasks (Executing waits for I/O and running of other programs.)
- **User CPU time** : CPU time spent in the program.
- **System CPU Time** : CPU time spent by OS to perform tasks on behalf of program.
- **Response time or time elapsed**: Total time required to complete the task

6

For CPU execution time formula.

- CPU execution time = **CPU cc for program x CC time** for program

Alternatively, clk rate and cc time are inverses,

- CPU execution time = $\frac{\text{CPU clk cycle for program}}{\text{Clock rate}}$ for program

7

Continued...

- CPU clock cycles = **Instructions for a program X Avg clk cycles per instruction.**
- Clk cycles per instruction, which Avg number of clock cycles per instruction.
- CPI is the average of all instructions executed in the program.

8

Example

Our favorite program runs in 10 seconds on computer A, which has a 400 Mhz. clock. We are trying to help a computer designer build a new machine B, that will run this program in 6 seconds. The designer can use new (or perhaps more expensive) technology to substantially increase the clock rate, but has informed us that this increase will affect the rest of the CPU design, causing machine B to require 1.2 times as many clock cycles as machine A for the same program. **What clock rate should we tell the designer to target?"**

9

CPI Example

- Suppose we have two implementations of the same instruction set Architecture (ISA).

For some program,

Machine A has a clock cycle time of 1 ns. and a CPI of 2.0

Machine B has a clock cycle time of 2 ns. and a CPI of 1.2

What machine is faster for this program, and by how much?

10

Performance Role

$$\text{Time} = \frac{\text{Instructions}}{\text{Program}} \times \frac{\text{Clock cycles}}{\text{Instruction}} \times \frac{\text{Seconds}}{\text{Clock cycle}}$$

Components of performance	Units of measure
CPU execution time for a program	Seconds for the program
Instruction count	Instructions executed for the program
Clock cycles per instruction (CPI)	Average number of clock cycles per instruction
Clock cycle time	Seconds per clock cycle

$$\text{CPU clock cycles} = \sum_{i=1}^n (\text{CPI}_i \times C_i)$$

$$\text{CPI} = \frac{\text{CPU clock cycles}}{\text{Instruction count}}$$

C_i is the count of the number of instructions of class i executed. CPI_i is the average number of cycles per instruction for that instruction class, and n is the number of instruction classes.

11

Amdahl's Law

- **Performance improvement to be gained by using faster mode is limited by the fraction of the time the faster mode can be used.**
- **Speed up is the measure of how machine performs after some enhancement relative to how it performed previously.**
- **Speedup = $\frac{\text{Performance for entire task using the enhancement when possible}}{\text{Performance for entire task without using the enhancement when possible}}$**
- **Fraction_{Enhanced} Always <1**
- **Speedup_{Enhanced} The improvement gained by enhanced execution mode**

$$\text{Execution time}_{\text{new}} = \text{Execution time}_{\text{old}} \times \left((1 - \text{Fraction}_{\text{enhanced}}) + \frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}} \right)$$

$$\text{The overall speedup: } \text{Speedup}_{\text{overall}} = \frac{\text{Execution time}_{\text{old}}}{\text{Execution time}_{\text{new}}}$$

12

MIPS

- **MIPS is Accurate Measure for Comparing Performance among Computers.**
- **MIPS is Dependent on the Instruction set.**
- **It varies b/w Programs on same Computer.**
- **MIPS can vary inversely to performance.**

$$\text{MIPS (Performance?)} = \frac{\text{Instruction count}}{\text{Execution time} \times 10^6} = \frac{\text{clock rate}}{\text{CPI} \times 10^6}$$

13

Number of Instructions Example

- A compiler designer is trying to decide between two code sequences for a particular machine. Based on the hardware implementation, there are three different classes of instructions: Class A, Class B, and Class C, and they require one, two, and three cycles (respectively).

The first code sequence has 5 instructions: 2 of A, 1 of B, and 2 of C

The second sequence has 6 instructions: 4 of A, 1 of B, and 1 of C.

Which sequence will be faster? How much?
What is the CPI for each sequence?

14

MIPS example

- Two different compilers are being tested for a 500 MHz machine with three different classes of instructions: Class A, Class B, and Class C, which require one, two, and three cycles (respectively). Both compilers are used to produce code for a large piece of software.

The first compiler's code uses 5 million Class A instructions, 1 million Class B instructions, and 1 million Class C instructions.

The second compiler's code uses 10 million Class A instructions, 1 million Class B instructions, and 1 million Class C instructions.

- Which sequence will be faster according to MIPS?
- Which sequence will be faster according to execution time?