Computer Organization and Architecture

CH 02: COMPUTER EVOLUTION AND PERFORMANCE

Content

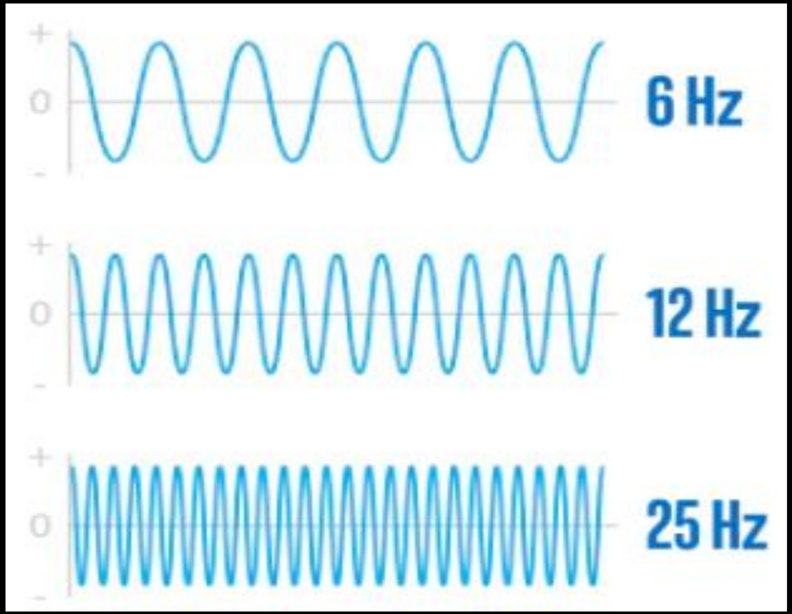
CH 02
A Brief History of Computers
Performance Assessment

Performance assessment

- Measures of processor speed.
    1. System clock
    2. Processor time to execute a program
    3. MIPS rate
    4. MFLOPS rate
    5. Average results of benchmarks
    6. Speed metric
    7. Rate metric

System Clock

- Operations performed by a processor are governed by a **system clock**.
- All operations begin with the pulse of the clock.
- The speed of a processor is dictated by the pulse frequency produced by the clock, measured in cycles per second, or Hertz (Hz).



Processor Time

- Processor is driven by a clock with a constant frequency f or, equivalently, a constant cycle time $\tau = 1/f$.
- The processor time T needed to execute a given program can be expressed as

$$T = I_c \times CPI \times \tau$$

  - I_c is the instruction count - the number of machine instructions executed for that program until it runs to completion or for some defined time interval.
  - CPI is the cycles per instruction for a program – number of cycles needed to execute an instruction.

Processor Time

- If all instructions required the same number of clock cycles, then CPI would be a constant value for a processor. <u>However,</u> on any give processor, the number of clock cycles required varies for different types of instructions, such as load, store, branch, and so on.

  - CPI_i is the number of cycles required for instruction type i.
  - I_i is the number of executed instructions of type i.

$$CPI = \frac{\sum_{i=1}^{n} (CPI_i \times I_i)}{I_c}$$

Processor Time

- During the execution of an instruction, part of the work is done by the processor, and part of the time a word is being transferred to or from memory.

$$T = I_c \times CPI \times \tau \quad \Rightarrow \quad T = I_c \times [p + (m \times k)] \times \tau$$

  - p is the number of processor cycles needed to decode and execute the instruction.
  - m is the number of memory references needed.
  - k is the ratio between memory cycle time and processor cycle time.

MIPS Rate

- MIPS rate is the rate at which instructions are executed, expressed as millions of instructions per second (MIPS).

$$\text{MIPS rate} = \frac{I_c}{T \times 10^6} = \frac{f}{CPI \times 10^6}$$

MFLOPS Rate

- Floating point performance is expressed as millions of floating-point operations per second (MFLOPS).
  - These are common in many scientific and game applications.

$$\text{MFLOPS rate} = \frac{\textit{Number of executed floating-point operations in a program}}{\textit{Execution time} \times 10^6}$$

Average results of benchmarks

- A benchmark suite is a collection of programs that together attempt to provide a representative test of a computer
- If we have m different benchmark program, and R_i is the instruction execution rate
  - **arithmetic mean**

$$R_A = \frac{1}{m}\sum_{i=1}^{m} R_i$$

  - **harmonic mean** (inverse of the average execution time)

$$R_H = \frac{m}{\sum_{i=1}^{m} \frac{1}{R_i}}$$

Speed Metric

- Measures the ability of a computer to complete a single task.
    - Defines a base runtime for each benchmark program using a reference machine.
- Results are reported as the **ratio** of the reference run time to the system run time.

$$r_i = \frac{Tref_i}{Tsut_i}$$

    - $Tref_i$ is the execution time of benchmark program i on the reference system and $Tsut_i$ is the execution time of benchmark program i on the system under test.
- The larger the ratio, the higher the speed.

Speed Metric

- Calculate the overall performance using the **geometric mean**

$$r_G = \left(\prod_{i=1}^{n} r_i \right)^{1/n}$$

  - where r_i is the ratio for the ith benchmark program, n is the number of benchmarks.

Rate Metric

- Measures the throughput or rate of a machine carrying out a number of tasks.
  - Multiple copies of the benchmarks are run simultaneously.
  - The number of copies is the same as the number of processors on the machine.
- The ratio is calculated as

$$r_i = \frac{N \times Tref_i}{Tsut_i}$$

  - $Tref_i$ is the reference execution time for benchmark i,
  - N is the number of copies of the program that are run simultaneously,
  - $Tsut_i$ is the elapsed time

Performance Assessment

- **Amdahl's law** calculates the speedup of a program using multiple processors compared to a single processor.

$$\text{Speedup} = \frac{\text{time to execute program on a single processor}}{\text{time to execute program on N parallel processors}}$$
$$= \frac{T(1-f) + Tf}{T(1-f) + \dfrac{Tf}{N}} = \frac{1}{(1-f) + \dfrac{f}{N}}$$

  - T is the total execution time of the program using a single processor.
  - N is the number of processors
  - $(1-f)$ is the execution time for a serial code
  - f is the execution time for a parallel code.

Performance Assessment

- Amdahl's law can be generalized to evaluate any design or technical improvement in a computer system.

$$\text{Speedup} = \frac{\text{Performance after enhancement}}{\text{Performance before enhancement}} = \frac{\text{Execution time before enhancement}}{\text{Execution time after enhancement}}$$

Exercises

BEFORE WE START…

There are 3 key steps to solve any problem?

Exercises

BEFORE WE START…

There are 3 key steps to solve any problem

1. **Memorizes the equations/formulas**
2. **Identify the given input**
3. **Identify the required output/results**

Exercises

Consider the execution of a program which results in the execution of 2 million instructions on a 400-MHz processor. The program consists of four major types of instructions. The instruction mix and the CPI for each instruction type are given below based on the result of a program trace experiment. Calculate the average CPI and MIPS rate.

Instruction Type	CPI	Instruction Mix
Arithmetic and logic	1	60%
Load/store with cache hit	2	18%
Branch	4	12%
Memory reference with cache miss	8	10%

Exercises

Instruction Type	CPI	Instruction Mix
Arithmetic and logic	1	60%
Load/store with cache hit	2	18%
Branch	4	12%
Memory reference with cache miss	8	10%

$$CPI = \frac{\sum_{i=1}^{n}(CPI_i \times I_i)}{I_c}$$

$$\text{MIPS rate} = \frac{I_c}{T \times 10^6} = \frac{f}{CPI \times 10^6}$$

We have CPIs = [1,2,4,8]; I_i = [60, 18, 12, 10], I_c = 100, $f = 400\ Mhz$

CPI = ?, MIPS rate = ?

CPI = $(0.6 * 1) + (2 * 0.18) + (4 * 0.12) + (8 * 0.1) = 2.24$

MIPS rate = $\frac{400*10^6}{2.24*10^6} \approx 178$

Exercises

A benchmark program is run on a 40 MHz processor. The executed program consists of 100,000 instruction executions, with the following instruction mix and clock cycle count:

Instruction Type	Instruction Count	Cycles per Instruction
Integer arithmetic	45000	1
Data transfer	32000	2
Floating point	15000	2
Control transfer	8000	2

Determine the effective CPI, MIPS rate, and execution time for this program.

Exercises

We have $I_i = [45000, 32000, 15000, 8000]$

CPI = $[1,2,2,2]$; $I_c = 100000,\ f = 40\ MHz$

Average CPI = ?, MIPS = ? , T = ?

Instruction Type	Instruction Count	Cycles per Instruction
Integer arithmetic	45000	1
Data transfer	32000	2
Floating point	15000	2
Control transfer	8000	2

CPI = $\frac{(1*45000)+(2*32000)+(2*15000)+(2*8000)}{100000} = 1.55$

$$CPI = \frac{\sum_{i=1}^{n}(CPI_i \times I_i)}{I_c}$$

MIPS rate = $\frac{40*10^6}{1.55*10^6} = 25.8$

$$\text{MIPS rate} = \frac{I_c}{T \times 10^6} = \frac{f}{CPI \times 10^6}$$

To calculate the execution time, use the equation

Then, $T = \frac{I_c}{MIPS\ rate\ *10^6} = \frac{100000}{25.8*10^6} = 0.00387\ s = 3.87\ ns$

Exercises

2.12. Early examples of CISC and RISC design are the VAX 11/780 and the IBM RS/6000, respectively. Using a typical benchmark program, the following machine characteristics result:

Processor	Clock Frequency	Performance	CPU Time
VAX 11/780	5 MHz	1 MIPS	12 x seconds
IBM RS/6000	25 MHz	18 MIPS	x seconds

The final column shows that the VAX required 12 times longer than the IBM measured in CPU time.

a. What is the relative size of the instruction count of the machine code for this benchmark program running on the two machines?

b. What are the CPI values for the two machines?

Exercises

We are given: VAX: $f = 5MHZ$, $MIPS = 1, T = 12x$

IBM: $f = 25MHZ, MIPS = 18, T = 1x$

$I_c = ?$, CPI = ?

$$\text{MIPS rate} = \frac{I_c}{T \times 10^6} = \frac{f}{CPI \times 10^6}$$

The instruction count $I_c = MIPS\ rate\ * (T * 10^6)$

For VAX, $I_c = 1 \ * 12x * 10^6$; For IBM, $I_c = 18 * 1x * 10^6$

The ratio between them = $\frac{18x*10^6}{12x*10^6} = 1.5$

CPI $= f/(MIPS\ rate * 10^6)$

For VAX, $CPI = \frac{5*10^6}{1*10^6} = 5$; For IBM, CPI $= \frac{25*10^6}{18*10^6} = 1.389$

Exercises

2.13. Four benchmark programs are executed on three computers with the following results:

	Computer A	Computer B	Computer C
Program 1	1	10	20
Program 2	1000	100	20
Program 3	500	1000	50
Program 4	100	800	100

The table shows the execution time in seconds, with 100,000,000 instructions executed in each of the four programs. Calculate the MIPS values for each computer for each program. Then calculate the arithmetic and harmonic means assuming equal weights for the four programs, and rank the computers based on arithmetic mean and harmonic mean.

Exercises

$$\text{MIPS rate} = \frac{I_c}{T \times 10^6}$$

$$R_A = \frac{1}{m}\sum_{i=1}^{m} R_i$$

$$R_H = \frac{m}{\sum_{i=1}^{m} \frac{1}{R_i}}$$

We are given: $I_c = 100\ 000\ 000$, T, m = 4 programs

MIPS rate = ? , arithmetic mean = ?, harmonic mean = ?

MIPS rate = $\frac{I_c}{T*10^6}$

	Computer A	Computer B	Computer C
Program 1	100	10	5
Program 2	0.1	1	5
Program 3	0.2	0.1	2
Program 4	1	0.125	1

To calculate the R_A and R_H, we have R_i for each computer

	Arithmetic mean	Rank	Harmonic mean	Rank
Computer A	25.325	1	0.25	2
Computer B	2.8	3	0.21	3
Computer C	3.26	2	2.1	1

Exercises

2.14. The following table, based on data reported in the literature [HEAT84], shows the execution times, in seconds, for five different benchmark programs on three machines.

a. Compute the speed metric for each processor for each benchmark, normalized to machine R. That is, the ratio values for R are all 1.0. Other ratios are calculated using Equation (2.5) with R treated as the reference system. Then compute the arithmetic mean value for each system using Equation (2.3).This is the approach taken in [HEAT84].

b. Repeat part (a) using M as the reference machine. This calculation was not tried in [HEAT84].

Benchmark	Processor		
	R	M	Z
E	417	244	134
F	83	70	70
H	66	153	135
I	39,449	35,527	66,000
K	772	368	369

Exercises

$$r_i = \frac{Tref_i}{Tsut_i}$$

$$R_A = \frac{1}{m}\sum_{i=1}^{m} R_i$$

We are given: T

Speed metric = ? (using R and M as reference machines) , arithmetic mean = ?

Normalized to R:

Benchmark	Processor		
	R	M	Z
E	1.00	1.71	3.11
F	1.00	1.19	1.19
H	1.00	0.43	0.49
I	1.00	1.11	0.60
K	1.00	2.10	2.09
Arithmetic mean	1.00	1.31	1.50

Normalized to M:

Benchmark	Processor		
	R	M	Z
E	0.59	1.00	1.82
F	0.84	1.00	1.00
H	2.32	1.00	1.13
I	0.90	1.00	0.54
K	0.48	1.00	1.00
Arithmetic mean	1.01	1.00	1.10

Exercises

2.15. To clarify the results of the preceding problem, we look at a simpler example.

Benchmark	Processor		
	X	Y	Z
1	20	10	40
2	40	80	20

a. Compute the arithmetic mean value for each system using X as the reference machine and then using Y as the reference machine. Argue that intuitively the three machines have roughly equivalent performance and that the arithmetic mean gives misleading results.

Exercises

$$r_i = \frac{Tref_i}{Tsut_i}$$

$$R_A = \frac{1}{m}\sum_{i=1}^{m} R_i$$

We are given: T

arithmetic mean = ? (using X and Y as reference machines)

Normalized to X:

Benchmark	Processor		
	X	Y	Z
1	1	2.0	0.5
2	1	0.5	2.0
Arithmetic mean	1	1.25	1.25

Normalized to Y:

Benchmark	Processor		
	X	Y	Z
1	0.5	1	0.25
2	2.0	1	4.0
Arithmetic mean	1.25	1	2.125

Exercises

Normalized to X:

Benchmark	Processor		
	X	Y	Z
1	1	2.0	0.5
2	1	0.5	2.0
Arithmetic mean	1	1.25	1.25

Normalized to Y:

Benchmark	Processor		
	X	Y	Z
1	0.5	1	0.25
2	2.0	1	4.0
Arithmetic mean	1.25	1	2.125

Machine Y is 2 times faster than machine X for benchmark 1. But for benchmark 2, machine X is 2 times faster than machine Y. Similarly, for benchmark 1, X is 2 times faster than Z, but for benchmark 2, Z is 2 times faster than X. However, when normalizing to X, we see that Y and Z are 25% faster than X. When normalize to Y, we find that X is 25% faster than Y, and Z is faster than Y by 1.125.

Exercises

2.16. Consider the example in Section 2.5 for the calculation of average CPI and MIPS rate, which yielded the result of CPI 2.24 and MIPS rate 178. Now assume that the program can be executed in eight parallel tasks or threads with roughly equal number of instructions executed in each task. Execution is on an 8-core system with each core (processor) having the same performance as the single processor originally used. Coordination and synchronization between the parts adds an extra 25,000 instruction executions to each task. Assume the same instruction mix as in the example for each task, but increase the CPI for memory reference with cache miss to 12 cycles due to contention for memory.

a. Determine the average CPI.

b. Determine the corresponding MIPS rate.

c. Calculate the speedup factor.

Instruction Type	CPI	Instruction Mix
Arithmetic and logic	1	60%
Load/store with cache hit	2	18%
Branch	4	12%
Memory reference with cache miss	8	10%

Exercises

We are given: “Assume the same instruction mix as in the example for each task, but increase the CPI for memory reference with cache miss to 12 cycles due to contention for memory.”

Instruction Type	CPI	Instruction Mix
Arithmetic and logic	1	60%
Load/store with cache hit	2	18%
Branch	4	12%
Memory reference with cache miss	~~8~~ 12	10%

$$CPI = \frac{\sum_{i=1}^{n}(CPI_i \times I_i)}{I_c}$$

The average CPI = (1× 0.6) + (2 × 0.18) + (4 × 0.12) + (12 × 0.1) = 2.64. Therefore, the CPI has been increased since the time for memory access is also increased.

Exercises

We are given: $f = 400\ MHz$ in the original problem

$$\text{MIPS rate} = \frac{I_c}{T \times 10^6} = \frac{f}{CPI \times 10^6}$$

MIPS = 400/2.64 = 152. There is a corresponding drop in the MIPS rate.

Exercises

$$\text{MIPS rate} = \frac{I_c}{T \times 10^6} = \frac{f}{CPI \times 10^6}$$

We are given: $I_c = 2000\,000$ instructions, MIPS = 178 in the original problem

The speedup factor equals to the ratio of the execution times. The execution time is calculated as the following: $T = \frac{I_c}{MIPS \times 10^6}$.

For the one processor, $T_1 = \frac{2000000}{178 \times 10^6} = 11\ ms$.

For the 8 processors, each processor executes 1/8 of the 2 million instructions plus the 25,000 (Coordination and synchronization overhead)

$$T_8 = \frac{\frac{2 \times 10^6}{8} + 0.025 \times 10^6}{152 \times 10^6} = 1.8 \text{ ms}$$

Therefore we have

$$\text{Speedup} = \frac{\text{time to execute program on a single processor}}{\text{time to execute program on } N \text{ parallel processors}} = \frac{11}{1.8} = 6.11$$

TASK

CH 02
2.2
2.3
2.11
2.14 (c, d)
2.15 (b)