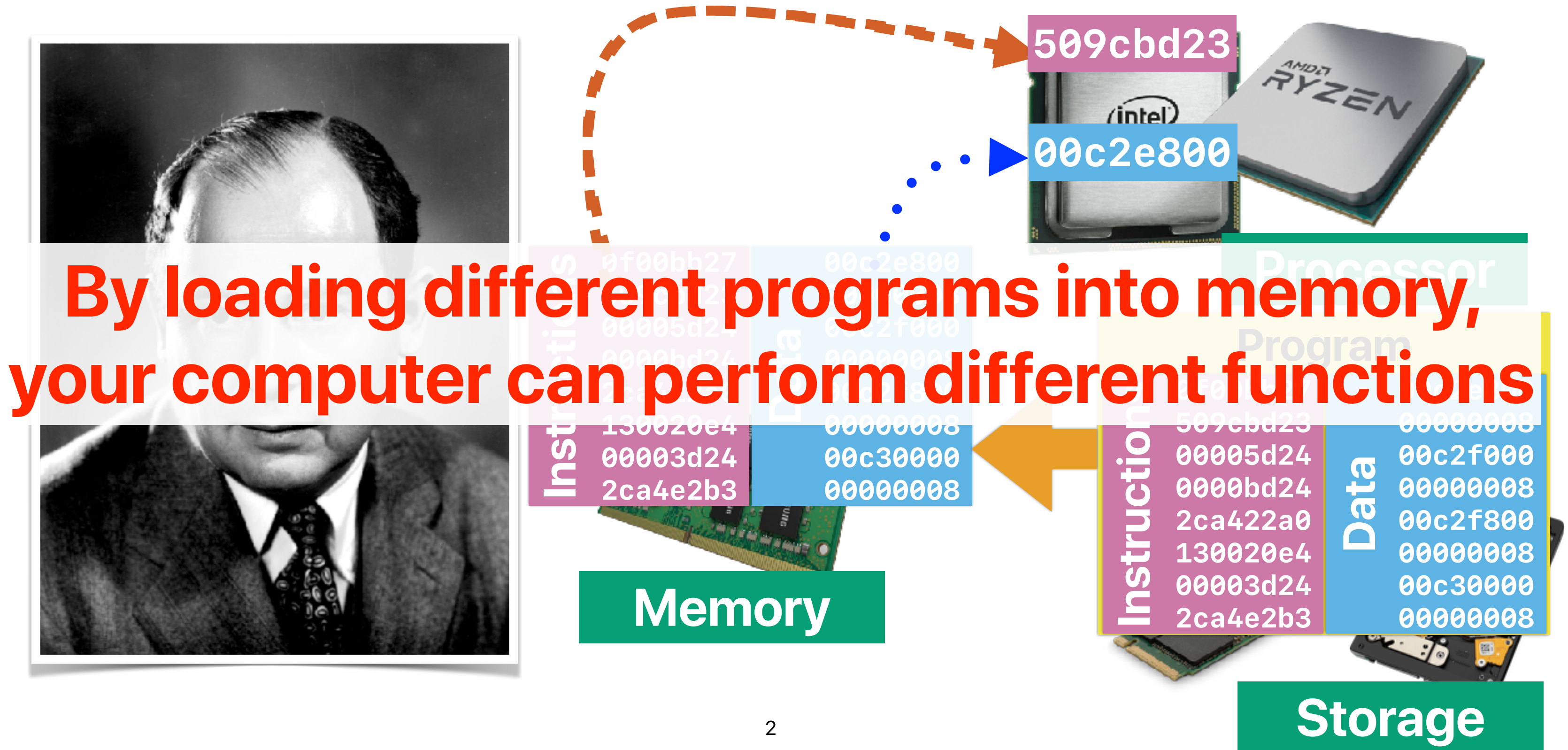


Performance (1): How Good Is "Good"?

Hung-Wei Tseng

Recap: von Neumann Architecture



Recap: Demo

```
if(option)
    std::sort(data, data + arraySize);  $O(n \log_2 n)$ 

for (unsigned c = 0; c < arraySize*1000; ++c) {
    if (data[c%arraySize] >= INT_MAX/2)
        sum ++;  $O(n)$ 
}
}
```

if option is set to 1: *$O(n \log_2 n)$*

otherwise, O(n): *$O(n)$*

3:00

**Where do you typically go for
lunch on campus? Why?**

**What do you want
your computer to be?**



What's your favorite restaurant on campus?

- Speed of service
- Taste of food
- Friendliness
- Cleanliness

Outline

- Definition of "Performance"
- What affects each factor in "Performance Equation"

Definition of “Performance”



CANONICAL

ubuntu® Enterprise ▾ Developer ▾ Community ▾ Download ▾

Downloads Overview Cloud IoT Raspberry Pi Server **Desktop** AMD-Xilinx Alternative downloads

Download Ubuntu Desktop

The open-source desktop operating system that powers millions of PCs and laptops around the world. Find out more about Ubuntu's features and how we support developers and organisations below.

[Ubuntu Desktop homepage](#) [Visit the Ubuntu Desktop blog »](#)

Peer instruction

- Before the lecture — You need to complete the required **reading**
- During the lecture — I'll bring in activities to ENGAGE you in exploring your understanding of the material
 - Popup questions
 - Individual **thinking** — use polls in Zoom to express your opinion
 - Group **discussion**
 - Breakout rooms based on your residential colleges!
 - Use polls in Zoom to express your group's opinion
 - Whole-classroom **discussion** — we would like to hear from you

Read

Think

Discuss

**Now, make sure you login to Poll
Everywhere (through the App or the
website) with UCRNetID**

**Now, you have at least 90 seconds
to answer the question!**

What do you care the most when?

- Consider the following performance metrics

1. Network Bandwidth (data/sec)

2. End-to-end Latency (ms)

3. Frame Rate (frames/sec)

4. Throughput (ops/sec)

Which option contains the best match of the most important performance metric for each application?

	Fortnite (Online gaming)	YouTube/Netflix	Download ISO images	Training an ML model
A	4	3	1	2
B	4	1	3	2
C	2	1	3	4
D	2	3	1	4
E	None of the above			

Now, it's time to discuss with your surroundings — and make sure you vote again after the discussion!

What do you care the most when?

- Consider the following performance metrics

1. Network Bandwidth (data/sec)

2. End-to-end Latency (ms)

3. Frame Rate (frames/sec)

4. Throughput (ops/sec)

Which option contains the best match of the most important performance metric for each application?

	Fortnite (Online gaming)	YouTube/Netflix	Download ISO images	Training an ML model
A	4	3	1	2
B	4	1	3	2
C	2	1	3	4
D	2	3	1	4
E	None of the above			

**Let's start with "end-to-end latency"
— how long it takes to execute a
program?**

CPU Performance Equation

$$Performance = \frac{1}{Execution\ Time}$$

$$Execution\ Time = \frac{Instructions}{Program} \times \frac{Cycles}{Instruction} \times \frac{Seconds}{Cycle}$$

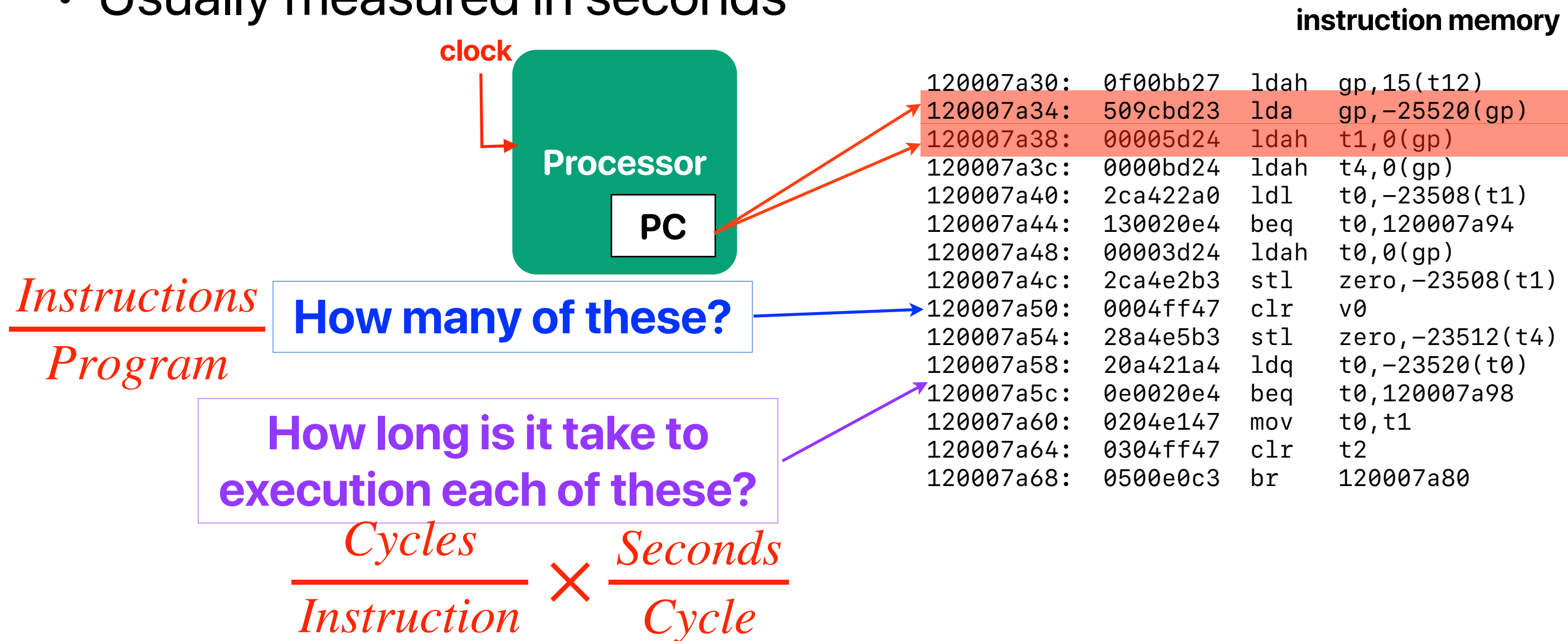
$$ET = IC \times CPI \times CT$$

$$1GHz = 10^9 Hz = \frac{1}{10^9} sec\ per\ cycle = 1\ ns\ per\ cycle$$

Frequency(i.e., clock rate)

Execution Time

- The simplest kind of performance
- Shorter execution time means better performance
- Usually measured in seconds



Performance Equation (X)

- Assume that we have an application composed with a total of **50000000000** instructions, in which **20%** of them are "Type-A" instructions with an average **CPI of 8** cycles, **20%** of them are "Type-B" instructions with an average **CPI of 4** cycles and **the rest** instructions are "Type-C" instructions with average **CPI of 1** cycle. If the processor runs at **3 GHz**, how long is the execution time?

A. 3.67 sec

B. 5 sec

C. 6.67 sec

D. 15 sec

E. 45 sec

$$ET = (5 \times 10^9) \times (20\% \times 8 + 20\% \times 4 + 60\% \times 1) \times \frac{1}{3 \times 10^{-9}} \text{sec} = 5$$

average CPI

$$ET = IC \times CPI \times CT$$

Speedup of Y over X

- Consider the same program on the following two machines, X and Y. By how much Y is faster than X?

	Clock Rate	Instructions	Percentage of Type-A	CPI of Type-A	Percentage of Type-B	CPI of Type-B	Percentage of Type-C	CPI of Type-C
Machine X	3 GHz	5000000000	20%	8	20%	4	60%	1
Machine Y	5 GHz	5000000000	20%	13	20%	4	60%	1

- A. 0.2
- B. 0.25
- C. 0.8
- D. 1.25
- E. No changes

Speedup of Y over X

- Consider the same program on the following two machines, X and Y. By how much Y is faster than X?

	Clock Rate	Instructions	Percentage of Type-A	CPI of Type-A	Percentage of Type-B	CPI of Type-B	Percentage of Type-C	CPI of Type-C
Machine X	3 GHz	5000000000	20%	8	20%	4	60%	1
Machine Y	5 GHz	5000000000	20%	13	20%	4	60%	1

- A. 0.2
- B. 0.25
- C. 0.8
- D. 1.25
- E. No changes

Speedup

- The relative performance between two machines, X and Y. Y is n times faster than X

$$n = \frac{\textit{Execution Time}_X}{\textit{Execution Time}_Y}$$

- The speedup of Y over X

$$\textit{Speedup} = \frac{\textit{Execution Time}_X}{\textit{Execution Time}_Y}$$

Speedup of Y over X

- Consider the same program on the following two machines, X and Y. By how much Y is faster than X?

	Clock Rate	Instructions	Percentage of Type-A	CPI of Type-A	Percentage of Type-B	CPI of Type-B	Percentage of Type-C	CPI of Type-C
Machine X	3 GHz	5000000000	20%	8	20%	4	60%	1
Machine Y	5 GHz	5000000000	20%	13	20%	4	60%	1

A. 0.2 $ET_Y = (5 \times 10^9) \times (20\% \times 13 + 20\% \times 4 + 60\% \times 1) \times \frac{1}{5 \times 10^9} \text{sec} = 4$

B. 0.25 $Speedup = \frac{Execution\ Time_X}{Execution\ Time_Y}$

C. 0.8

D. 1.25

$$= \frac{5}{4} = 1.25$$

E. No changes

What Affects Each Factor in Performance Equation

How programmer affects performance?

- Performance equation consists of the following three factors
 - ① IC
 - ② CPI
 - ③ CT

How many can a **programmer** affect?

- A. 0
- B. 1
- C. 2
- D. 3

How programmer affects performance?

- Performance equation consists of the following three factors
 - ① IC
 - ② CPI
 - ③ CT

How many can a **programmer** affect?

- A. 0
- B. 1
- C. 2
- D. 3

Demo — programmer & performance

A

```
for(i = 0; i < ARRAY_SIZE; i++)
{
    for(j = 0; j < ARRAY_SIZE; j++)
    {
        c[i][j] = a[i][j]+b[i][j];
    }
}
```

B

```
for(j = 0; j < ARRAY_SIZE; j++)
{
    for(i = 0; i < ARRAY_SIZE; i++)
    {
        c[i][j] = a[i][j]+b[i][j];
    }
}
```

How many of the following make(s) the performance different between version A & version B?

- ① IC
- ② CPI
- ③ CT
- A. 0
- B. 1
- C. 2
- D. 3

Demo — programmer & performance

A

```
for(i = 0; i < ARRAY_SIZE; i++)  
{  
    for(j = 0; j < ARRAY_SIZE; j++)  
    {  
        c[i][j] = a[i][j]+b[i][j];  
    }  
}
```

B

```
for(j = 0; j < ARRAY_SIZE; j++)  
{  
    for(i = 0; i < ARRAY_SIZE; i++)  
    {  
        c[i][j] = a[i][j]+b[i][j];  
    }  
}
```

How many of the following make(s) the performance different between version A & version B?

- ① IC
- ② CPI
- ③ CT
- A. 0
- B. 1
- C. 2
- D. 3

Demo — programmer & performance

A

```
for(i = 0; i < ARRAY_SIZE; i++)  
{  
    for(j = 0; j < ARRAY_SIZE; j++)  
    {  
        c[i][j] = a[i][j]+b[i][j];  
    }  
}
```

B

```
for(j = 0; j < ARRAY_SIZE; j++)  
{  
    for(i = 0; i < ARRAY_SIZE; i++)  
    {  
        c[i][j] = a[i][j]+b[i][j];  
    }  
}
```

$O(n^2)$

Complexity

$O(n^2)$

Same

Instruction Count?

Same

Same

Clock Rate

Same

???

CPI

???

Use “performance counters” to figure out!

- Modern processors provides performance counters
 - instruction counts
 - cache accesses/misses
 - branch instructions/mis-predictions
- How to get their values?
 - You may use “perf stat” in linux
 - You may use Instruments —> Time Profiler on a Mac
 - Intel’s vtune — only works on Windows w/ intel processors
 - You can also create your own functions to obtain counter values

Demo — programmer & performance

A

```
for(i = 0; i < ARRAY_SIZE; i++)  
{  
    for(j = 0; j < ARRAY_SIZE; j++)  
    {  
        c[i][j] = a[i][j]+b[i][j];  
    }  
}
```

B

```
for(j = 0; j < ARRAY_SIZE; j++)  
{  
    for(i = 0; i < ARRAY_SIZE; i++)  
    {  
        c[i][j] = a[i][j]+b[i][j];  
    }  
}
```

$O(n^2)$

Complexity

$O(n^2)$

Same

Instruction Count?

Same

Same

Clock Rate

Same

Better

CPI

Worse

Demo — programmer & performance

A

```
for(i = 0; i < ARRAY_SIZE; i++)
{
    for(j = 0; j < ARRAY_SIZE; j++)
    {
        c[i][j] = a[i][j]+b[i][j];
    }
}
```

B

```
for(j = 0; j < ARRAY_SIZE; j++)
{
    for(i = 0; i < ARRAY_SIZE; i++)
    {
        c[i][j] = a[i][j]+b[i][j];
    }
}
```

How many of the following make(s) the performance different between version A & version B?

① IC

☒ ② CPI

③ CT

A. 0

B. 1

C. 2

D. 3

Programmer's impact

- By adding the "sort" in the following code snippet, what the programmer changes in the performance equation to achieve **better** performance?

```
std::sort(data, data + arraySize);
```

```
for (unsigned c = 0; c < arraySize*1000; ++c) {  
    if (data[c%arraySize] >= INT_MAX/2)  
        sum ++;  
}
```

- A. CPI
- B. IC
- C. CT
- D. IC & CPI
- E. CPI & CT

Programmer's impact

- By adding the "sort" in the following code snippet, what the programmer changes in the performance equation to achieve **better** performance?

```
std::sort(data, data + arraySize);
```

```
for (unsigned c = 0; c < arraySize*1000; ++c) {  
    if (data[c%arraySize] >= INT_MAX/2)  
        sum ++;  
}
```

- A. CPI
- B. IC
- C. CT
- D. IC & CPI
- E. CPI & CT

Programmer's impact

- By adding the "sort" in the following code snippet, what the programmer changes in the performance equation to achieve **better** performance?

```
std::sort(data, data + arraySize);
```

```
for (unsigned c = 0; c < arraySize*1000; ++c) {  
    if (data[c%arraySize] >= INT_MAX/2)  
        sum ++;  
}
```

A. CPI

B. IC

C. CT

D. IC & CPI

E. CPI & CT



programmer changes IC as well, but
not in the positive direction

Programmers can also set the cycle time

<https://software.intel.com/sites/default/files/comment/1716807/how-to-change-frequency-on-linux-pub.txt>

```
=====
Subject: setting CPU speed on running linux system
```

If the OS is Linux, you can manually control the CPU speed by reading and writing some virtual files in the "/proc"

1.) Is the system capable of software CPU speed control?

If the "directory" /sys/devices/system/cpu/cpu0/cpufreq exists, speed is controllable.

-- If it does not exist, you may need to go to the BIOS and turn on EIST and any other C and P state control and vi

2.) What speed is the box set to now?

Do the following:

```
$ cd /sys/devices/system/cpu
```

```
$ cat ./cpu0/cpufreq/cpuinfo_max_freq
```

```
3193000
```

```
$ cat ./cpu0/cpufreq/cpuinfo_min_freq
```

```
1596000
```

3.) What speeds can I set to?

Do

```
$ cat /sys/devices/system/cpu/cpu0/cpufreq/scaling_available_frequencies
```

It will list highest settable to lowest; example from my NHM "Smackover" DX58SO HEDT board, I see:

```
3193000 3192000 3059000 2926000 2793000 2660000 2527000 2394000 2261000 2128000 1995000 1862000 1729000 159600
```

You can choose from among those numbers to set the "high water" mark and "low water" mark for speed. If you set "h:

4.) Show me how to set all to highest settable speed!

Use the following little sh/ksh/bash script:

```
$ cd /sys/devices/system/cpu # a virtual directory made visible by device drivers
```

```
$ newSpeedTop=`awk '{print $1}' ./cpu0/cpufreq/scaling_available_frequencies`
```

```
$ newSpeedLow=$newSpeedTop # make them the same in this example
```

```
$ for c in ./cpu[0-9]* ; do
```

```
> echo $newSpeedTop >${c}/cpufreq/scaling_max_freq
```

```
> echo $newSpeedLow >${c}/cpufreq/scaling_min_freq
```

```
> done
```

```
$
```

5.) How do I return to the default - i.e. allow machine to vary from highest to lowest?

Edit line # 3 of the script above, and re-run it. Change the line:

```
$ newSpeedLow=$newSpeedTop # make them the same in this example
```

```
To read
```

How programmer affects performance?

- Performance equation consists of the following three factors

① ✓ IC

② ✓ CPI

③ ✓ CT

How many can a **programmer** affect?

A. 0

B. 1

C. 2

D. 3

How programming languages affect performance

- Performance equation consists of the following three factors
 - ① IC
 - ② CPI
 - ③ CT

How many can the **programming language** affect?

- A. 0
- B. 1
- C. 2
- D. 3



Programming languages

- Which of the following programming language needs to highest instruction count to print "Hello, world!" on screen?
 - A. C
 - B. C++
 - C. Java
 - D. Perl
 - E. Python



Programming languages

- Which of the following programming language needs to highest instruction count to print "Hello, world!" on screen?
 - A. C
 - B. C++
 - C. Java
 - D. Perl
 - E. Python



Programming languages

- How many instructions are there in "Hello, world!"

	Instruction count	LOC	Ranking
C	600k	6	1
C++	3M	6	2
Java	~145M	8	5
Perl	~12M	4	3
Python	~33M	1	4
GO (Interpreter)	~1200M	1	6
GO (Compiled)	~1.7M	1	1.5

Programming languages

- Which of the following programming language needs to highest instruction count to print "Hello, world!" on screen?
 - A. C
 - B. C++
 - C. Java
 - D. Perl
 - E. Python

Announcement

- Cancellation
 - Hung-Wei's office hour today
 - CS203 lecture next Monday
- Reading quiz due next Wednesday before the lecture
 - We will drop two of your least performing reading quizzes
 - You have two shots, both unlimited time
- Assignment #1 released
 - We typically give you two weeks to work on an assignment
 - Due on 10/14
- Assignment #0 due on 10/7
- Check our website for slides, eLearn for quizzes/assignments, piazza for discussions
- Youtube channel for lecture recordings:
<https://www.youtube.com/c/ProfUsagi/playlists>

Computer Science & Engineering

203

つづく

