SUPSI

Modern Computer Architecture

Massimo Coluzzi

Parallel and Concurrent Programming
Bachelor in Data Science

Hardware and Software

Modern programming languages provide a layer of abstraction over the underlying hardware.

Nowadays, people can write decent programs for simple sequential tasks without knowing anything about hardware architectures.

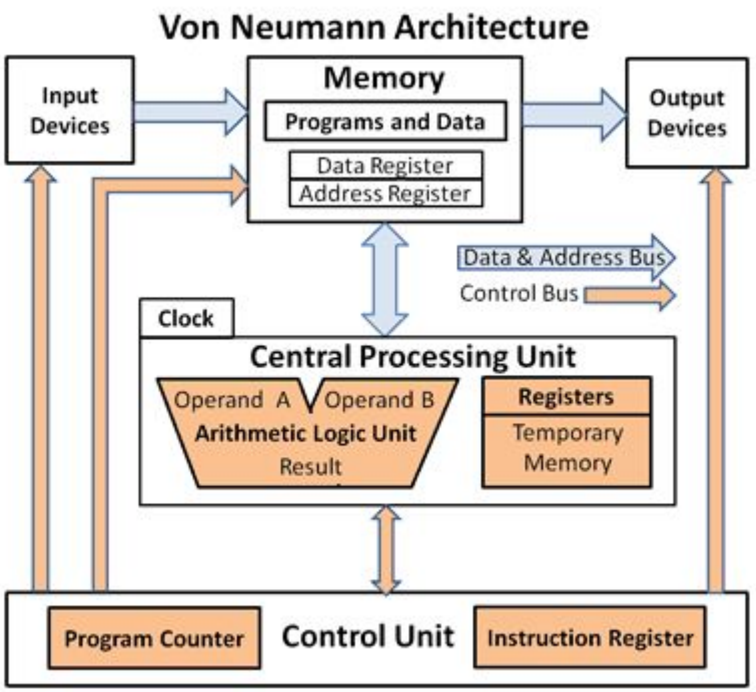
But knowing the exact relationship between hardware and software becomes crucial when dealing with parallel programming.

Therefore, we will summarize some basic concepts about modern computer architectures.

Step 1: The Von Neumann Architecture

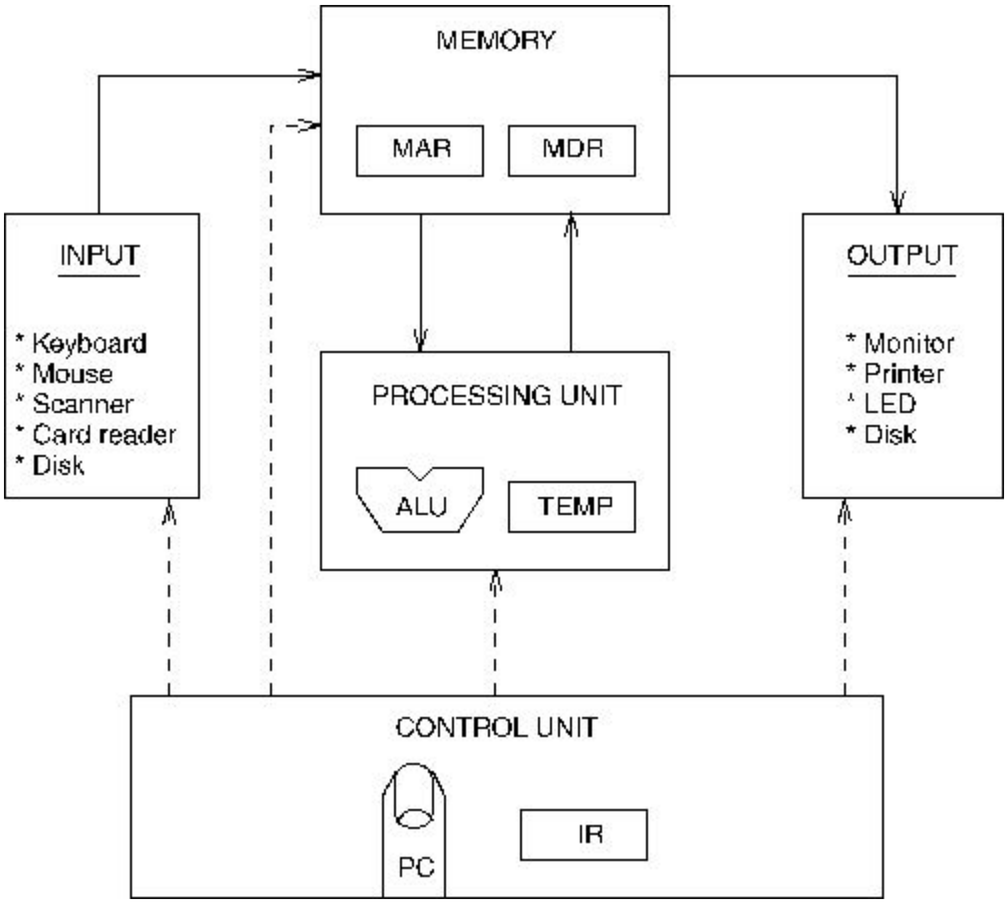
The Von Neumann Architecture

- Both programs and data are stored in memory (RAM).
- A program is executed by reading one instruction at a time from memory.
- The instruction to be executed is stored in the Instruction Registry.
- The location of the next instruction is stored in the Program Counter.



The Von Neumann Architecture

- It works pretty much the same when reading data from memory.
- The memory address to read is stored in the Memory Address Register (MAR).
- The content of the memory location is stored in the Memory Data Register (MDR).



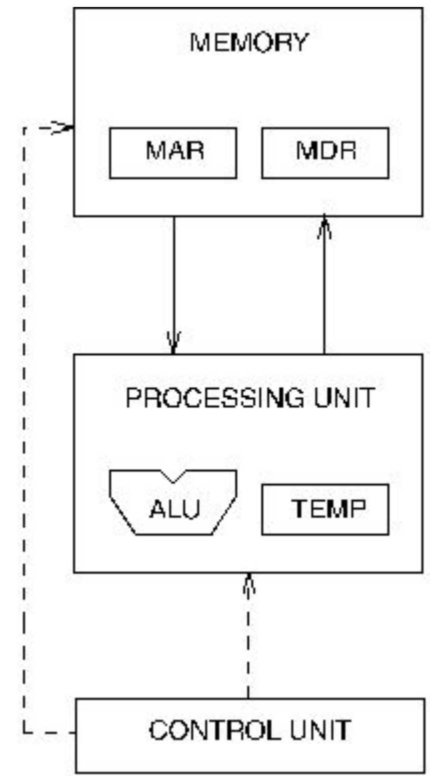
The Von Neumann Architecture

To LOAD a location (A):

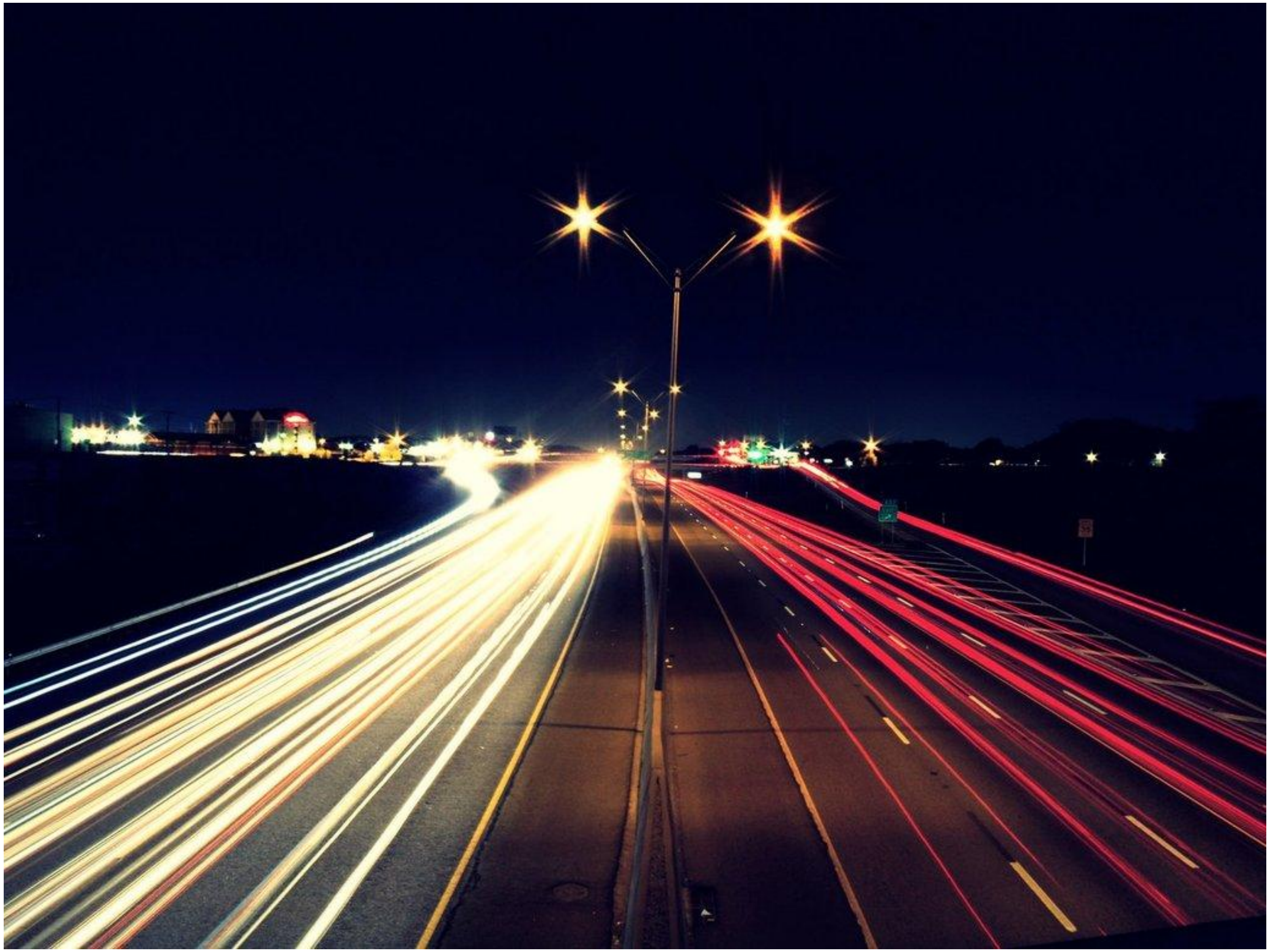
1. Write the address (A) into the MAR.
2. Send a “read” signal to the memory.
3. Read the data from MDR.

To STORE a value (X) to a location (A):

1. Write the data (X) to the MDR.
2. Write the address (A) into the MAR.
3. Send a “write” signal to the memory.



Step 2: The speed of signals



Question

How fast does electricity move through a conductor?

Answer

Electricity moves almost at the speed of light.

It means that, inside a circuit, signals move more or less at 300’000 Km/s.

Speed of signals

Modern chipsets have a base frequency of 3.8 GHz.

It means signals move inside a chip 3.8 billion times in a second.

Hence, a signal has 0.26ns time to reach its destination and return.

Question

Which distance can a signal cover in 0.26ns?

Answer

```
speed_of_signals = 300’000Km/s
available_time   = 0.26ns
distance         =
   speed_of_signals * available_time = 7.9cm
```

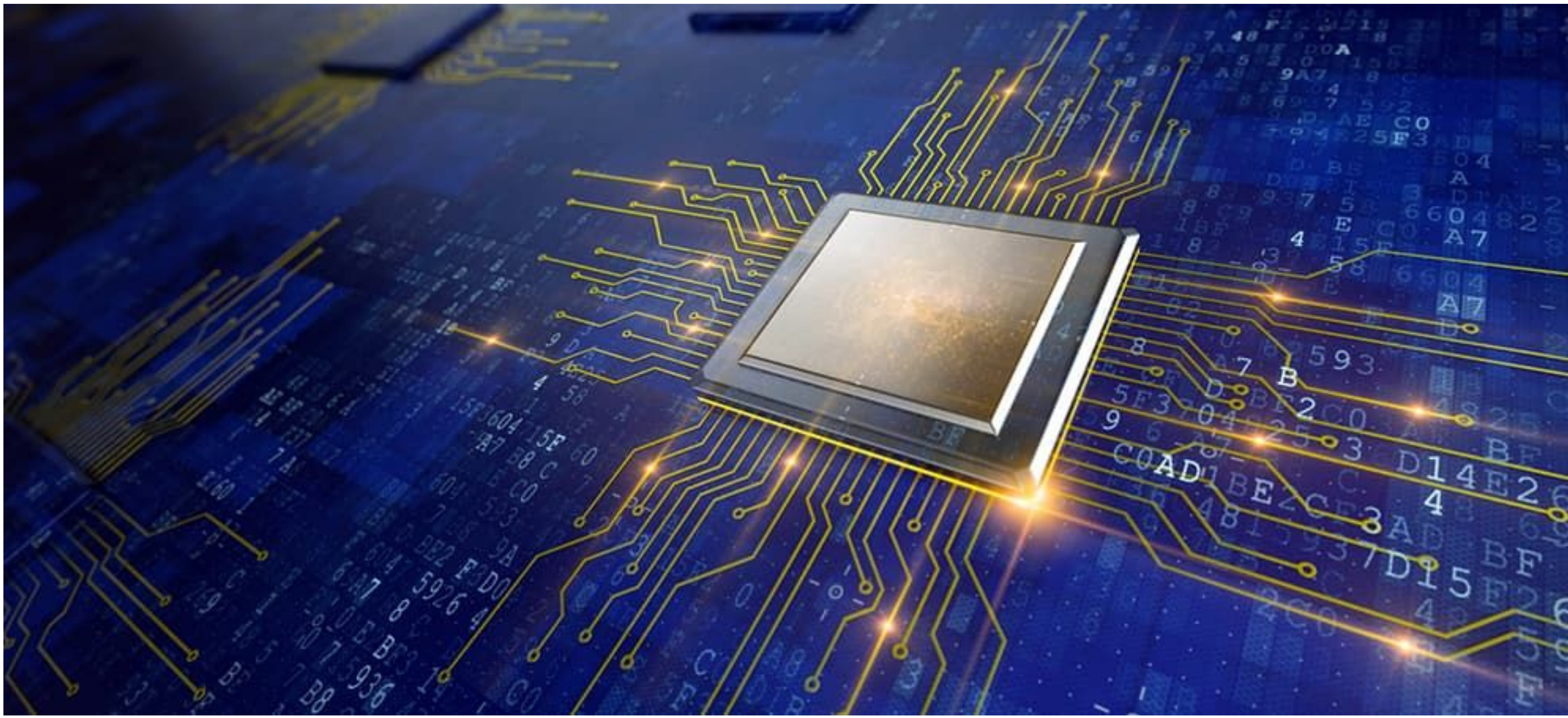
Since the signal needs to move forth and back the distance must be halved.

The overall answer is less than 4 cm.

Question

Computers and laptops are definitely bigger than 4 cm.
How can they work at 3.8 GHz?

Step 3: The Memory Architecture



The Memory Architecture

Every operation in a Von Neumann Architecture needs to access the memory.

- We need to read the next instruction to execute.
- We need to read variables and write results.

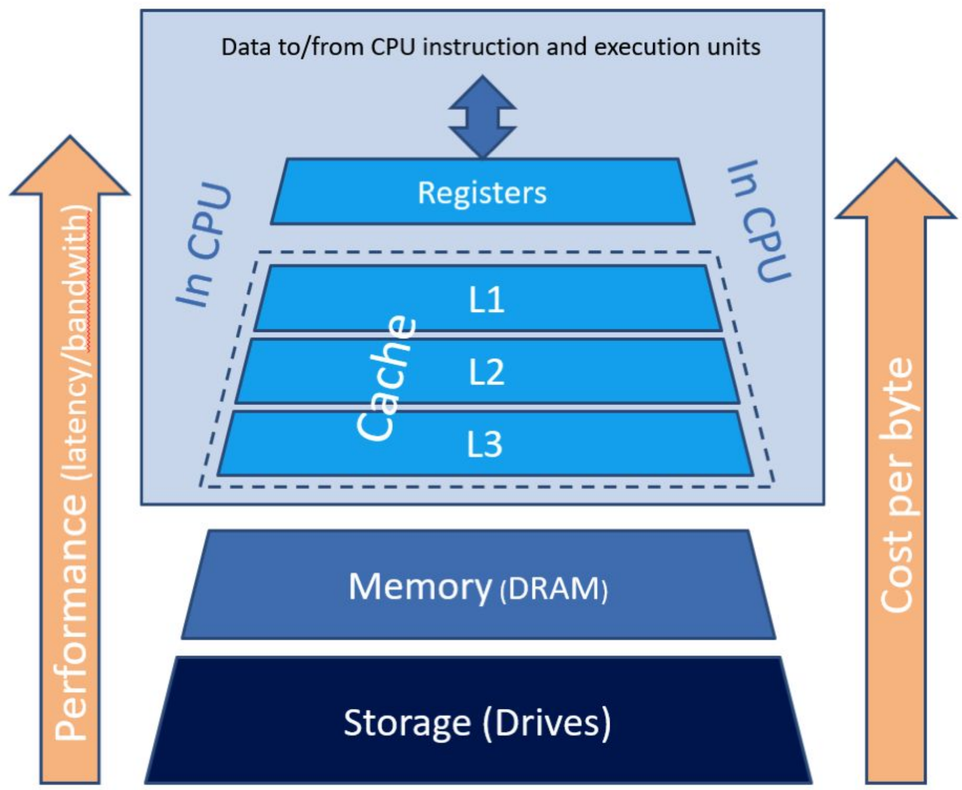
To allow performing 3.8 billions operations per second, memory must be at most 4 cm apart from the CPU.

The Memory Architecture

We cannot make all memory to fit int 16 cm^2.

Therefore, the solution is to divide memory into multiple caching layers.

The portion of memory that will be used in the next future will be moved to L3 cache.

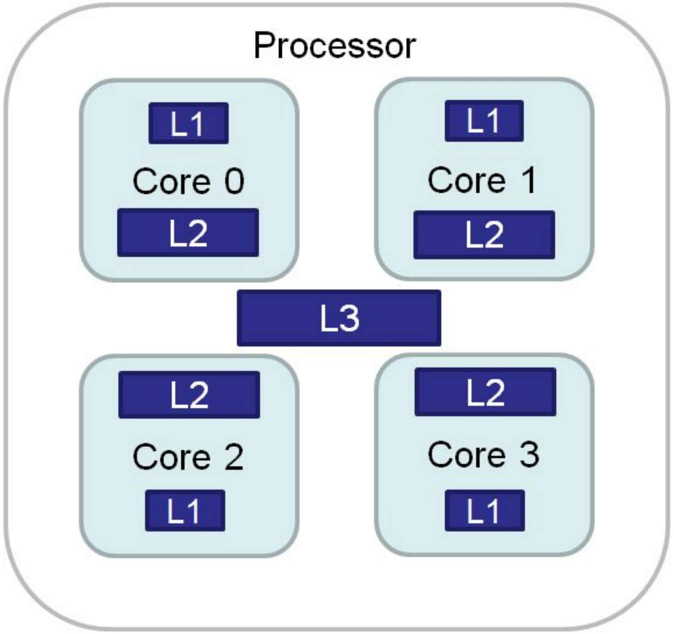


The Memory Architecture

L3 cache is common to all the CPU cores.

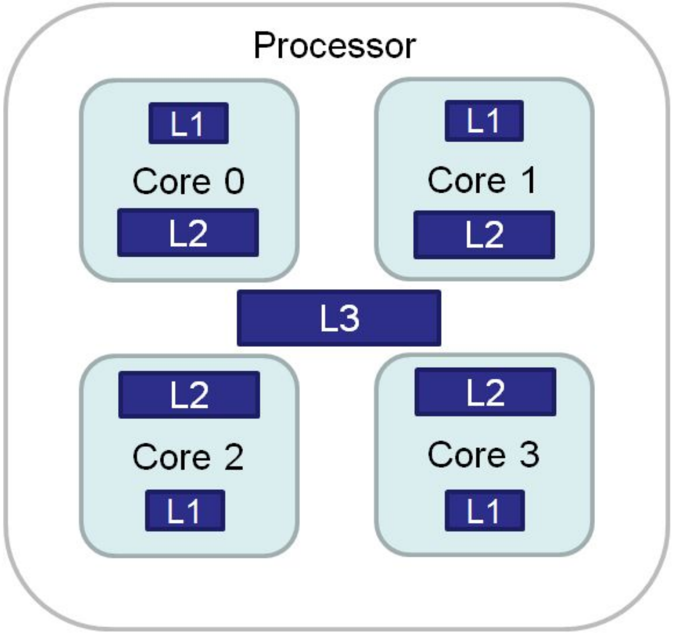
The same approach is then iterated moving to L2 cache the portion of L3 cache that will be used by each core in the next future.

Finally, the portion of memory actually used is moved to the L1 cache for best performance.



Question

What happens if all the cores read and write the same data?



Answer

The aim of this course is mainly to answer this question.

To better understand the content of this course, you should have clear the components of the Von Neumann architecture and how they interact.