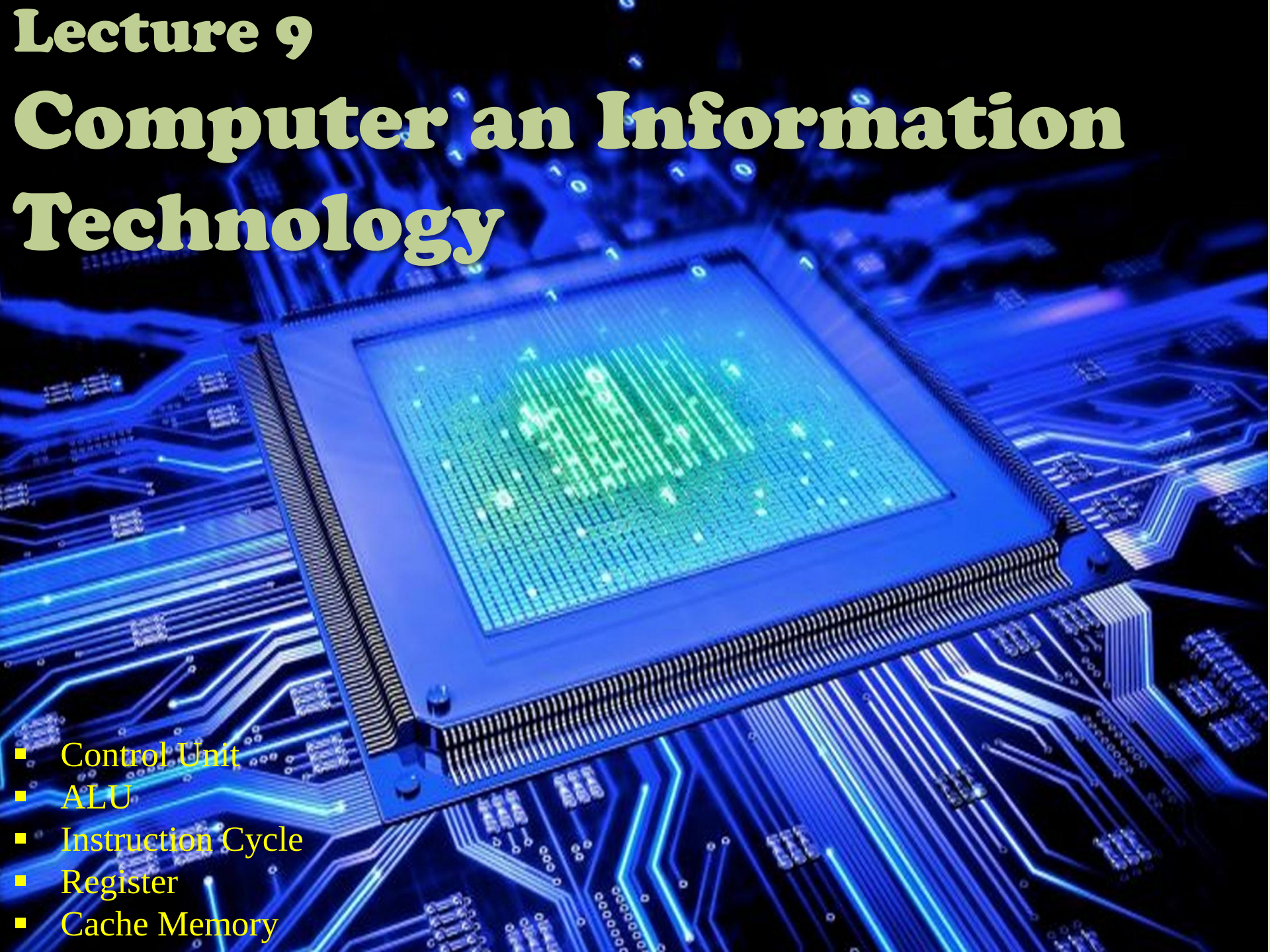
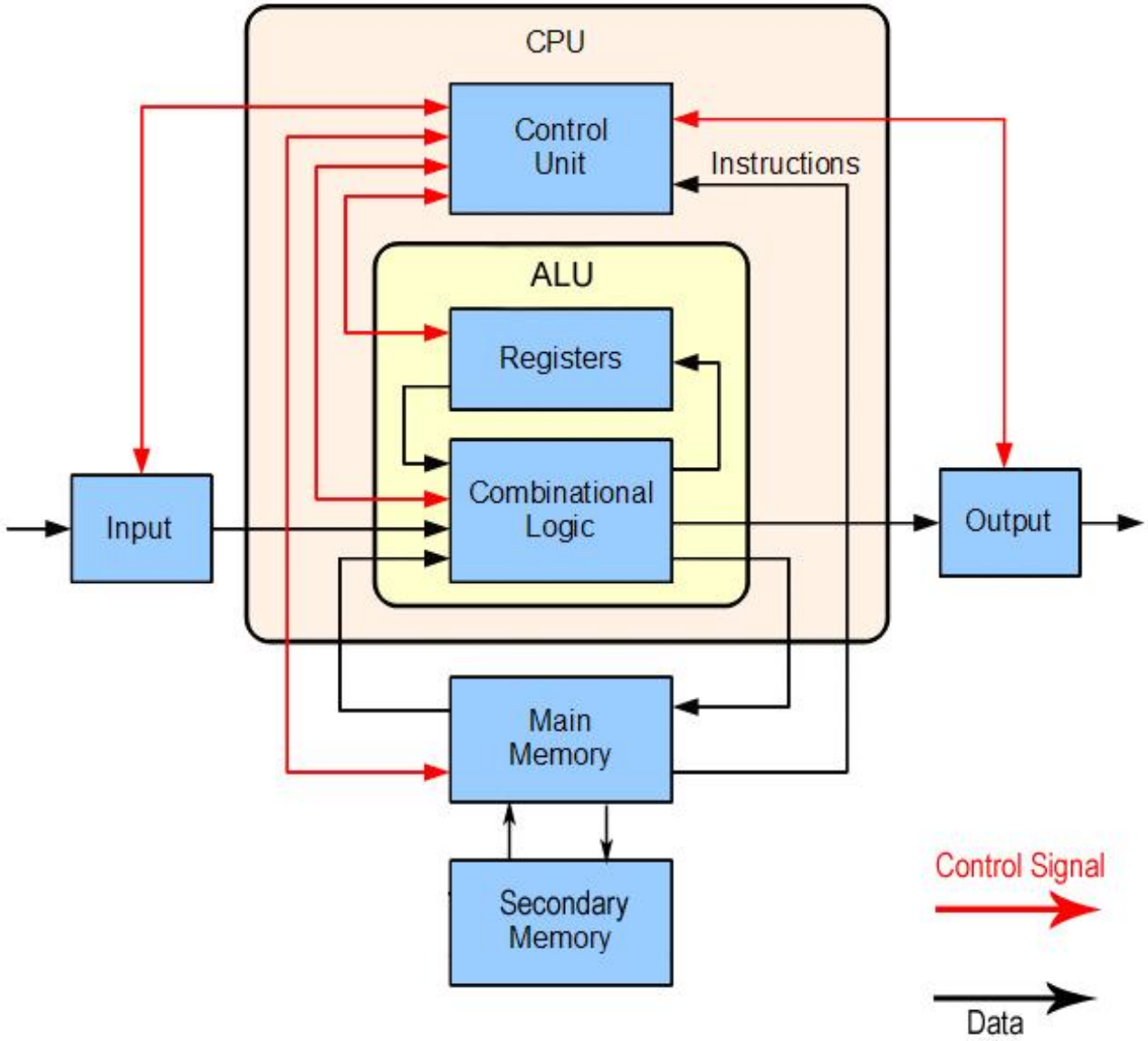
Lecture 9

Computer an Information Technology



- Control Unit
- ALU
- Instruction Cycle
- Register
- Cache Memory

Computer Architecture



The working of a CPU can simply be shown using this architecture.

Here we can see, the CPU is made of **control unit**, **arithmetic logic unit**, various **registers** and all of them communicate with one another with help from various **buses**.

Good to Know

CPU Clock Cycle

A clock cycle, or simply a "cycle," is a single electronic pulse of a CPU. During each cycle, a CPU can perform a basic operation such as fetching an instruction, accessing memory, or writing data. Since only simple commands can be performed during each cycle, most CPU processes require multiple clock cycles.

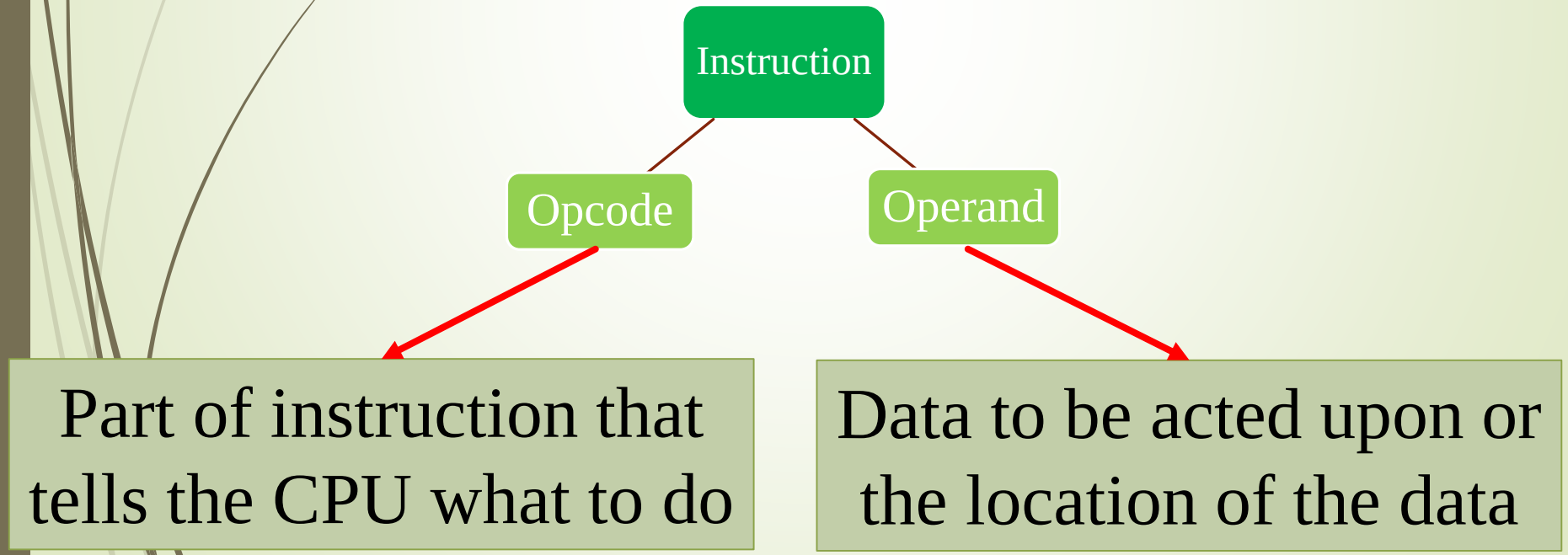
CPU clock along with some other factors, determines the speed of instruction execution speed of a CPU. Clock speed is measured in Hz (Hertz) which is in fact the unit for frequency and hence clock cycle or clock speed is also known as clock frequency.

A 3.5GHz (pronounced as “Giga Hart-j) CPU clock generates 3.5 billion pulses in one second and each of these pulses performs at least one task in any part of a CPU.

Good to Know

Instruction

From start to shutdown, a computer CPU processes instructions. When the users see their application program working, the CPU in fact processing quadrillions of instructions for the OS. To put this in simple, we breakdown these instruction in following form:



Instruction Cycle

For processing any instruction, the CPU goes through some simpler steps. Number of steps varies with CPU architecture but for simplicity we will break down an instruction into 4 steps.

Fetch

Bringing an instruction from memory.

Decode

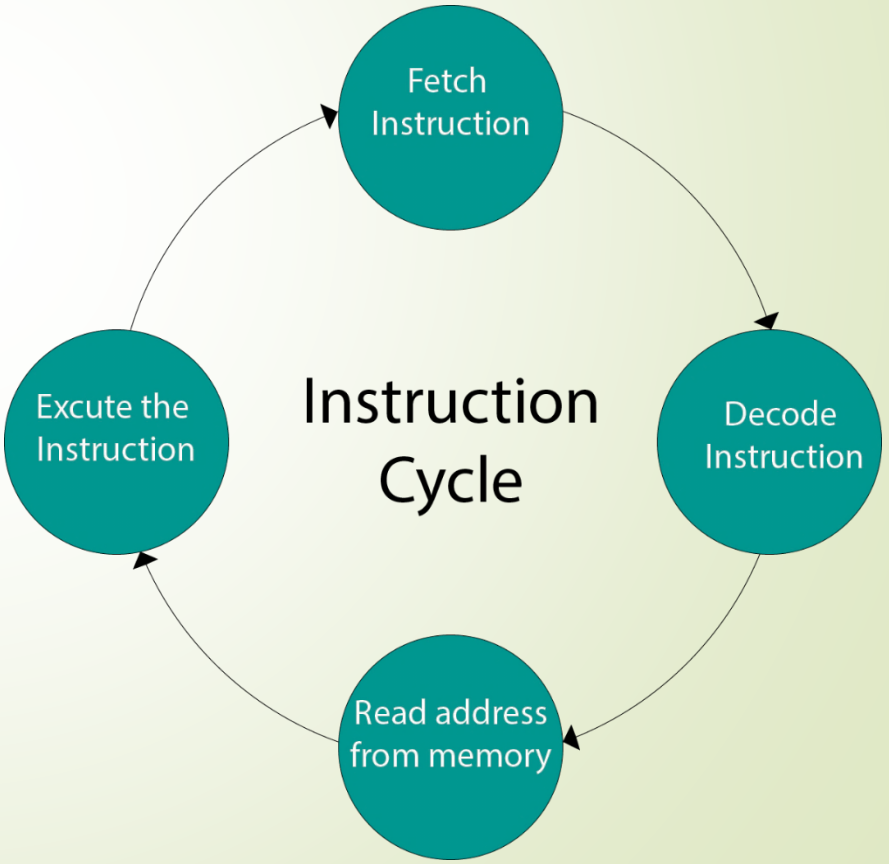
Retrieving the operation code or Opcode from the instruction.

Read Address from Memory

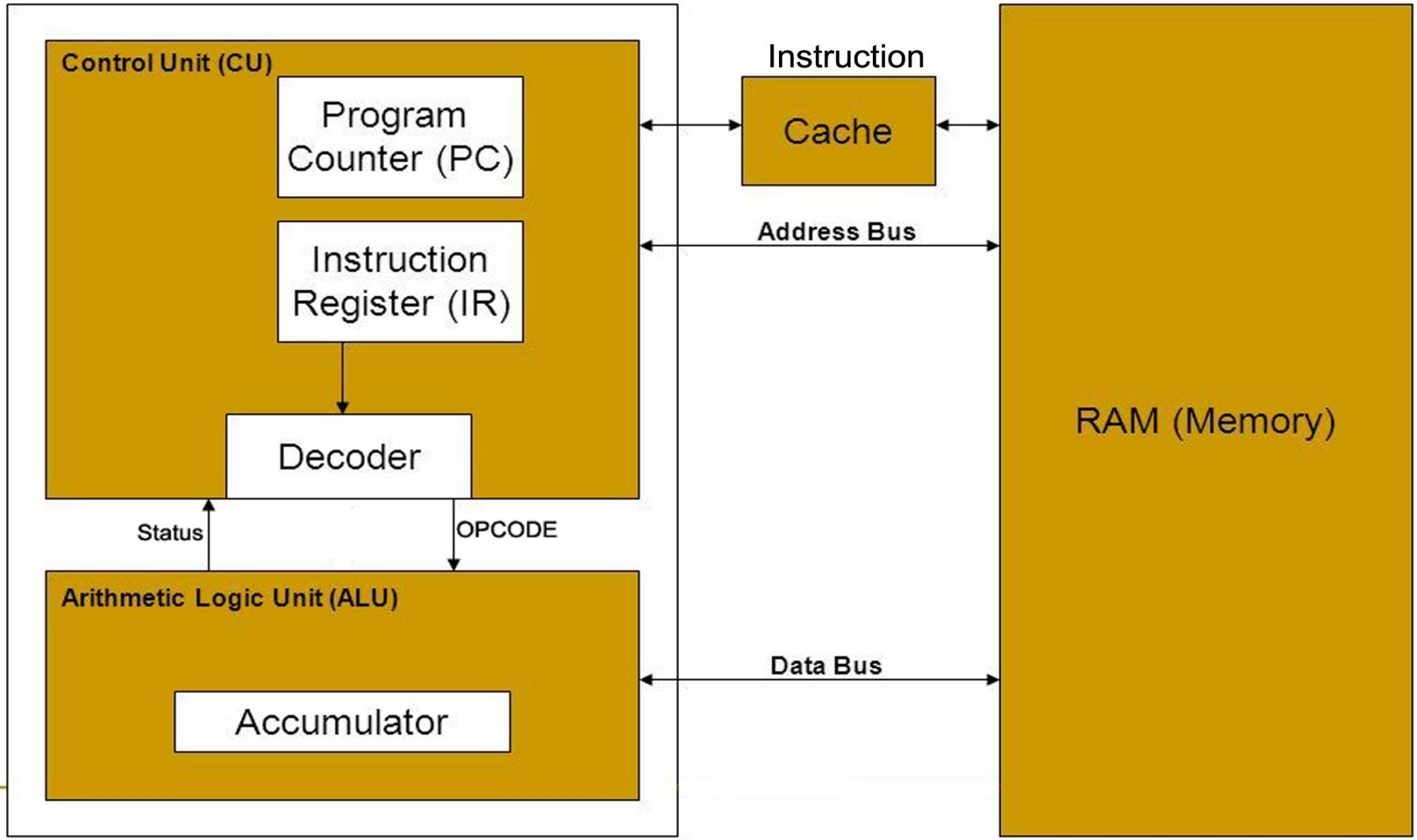
Decoding also generates operand and then the data is located from memory or register.

Execute

Logic circuit in ALU performs the operation onto the data.



Control Unit & Arithmetic Logic Unit Functioning During an Instruction Cycle

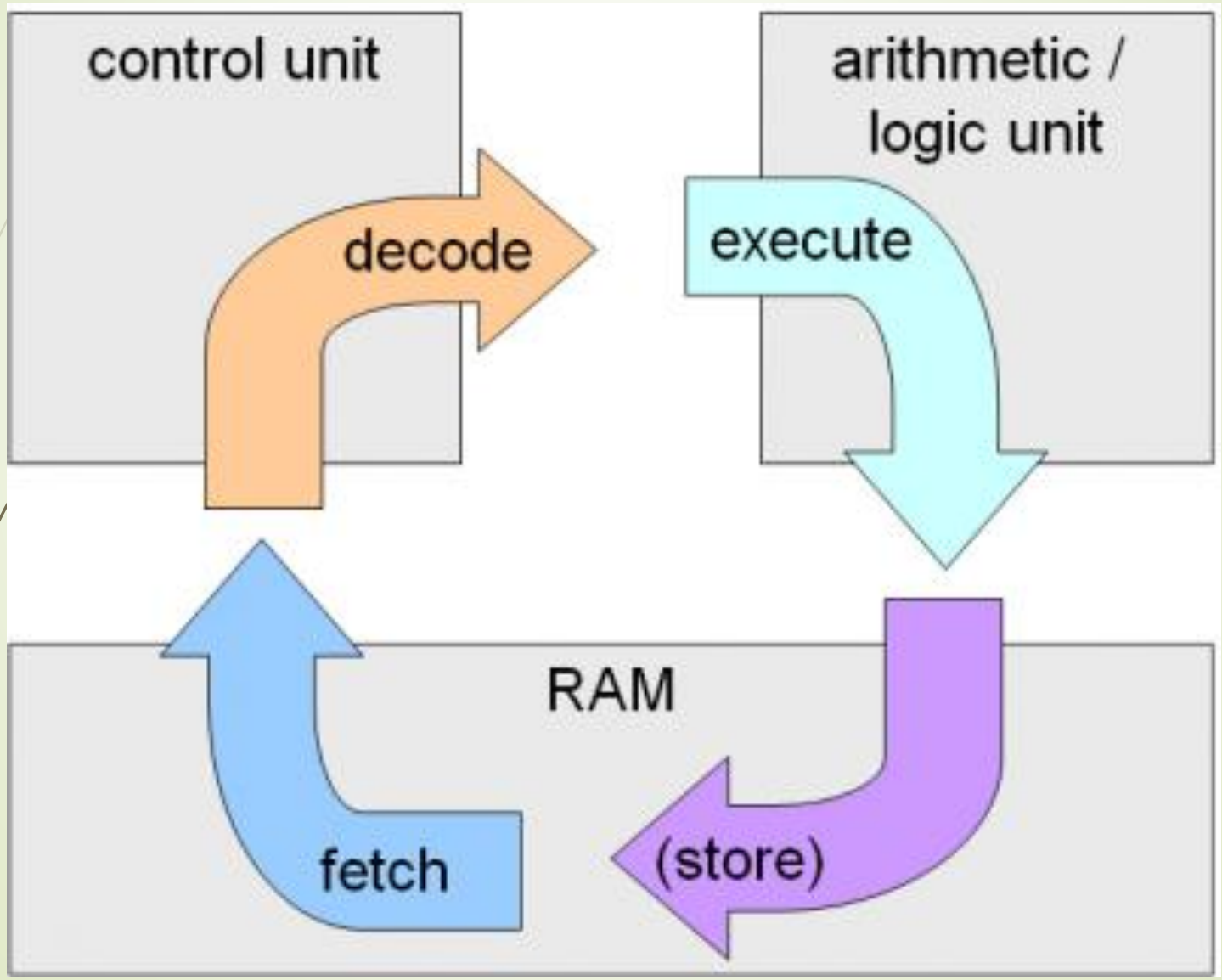


Control Unit & Arithmetic Logic Unit Functioning During an Instruction Cycle

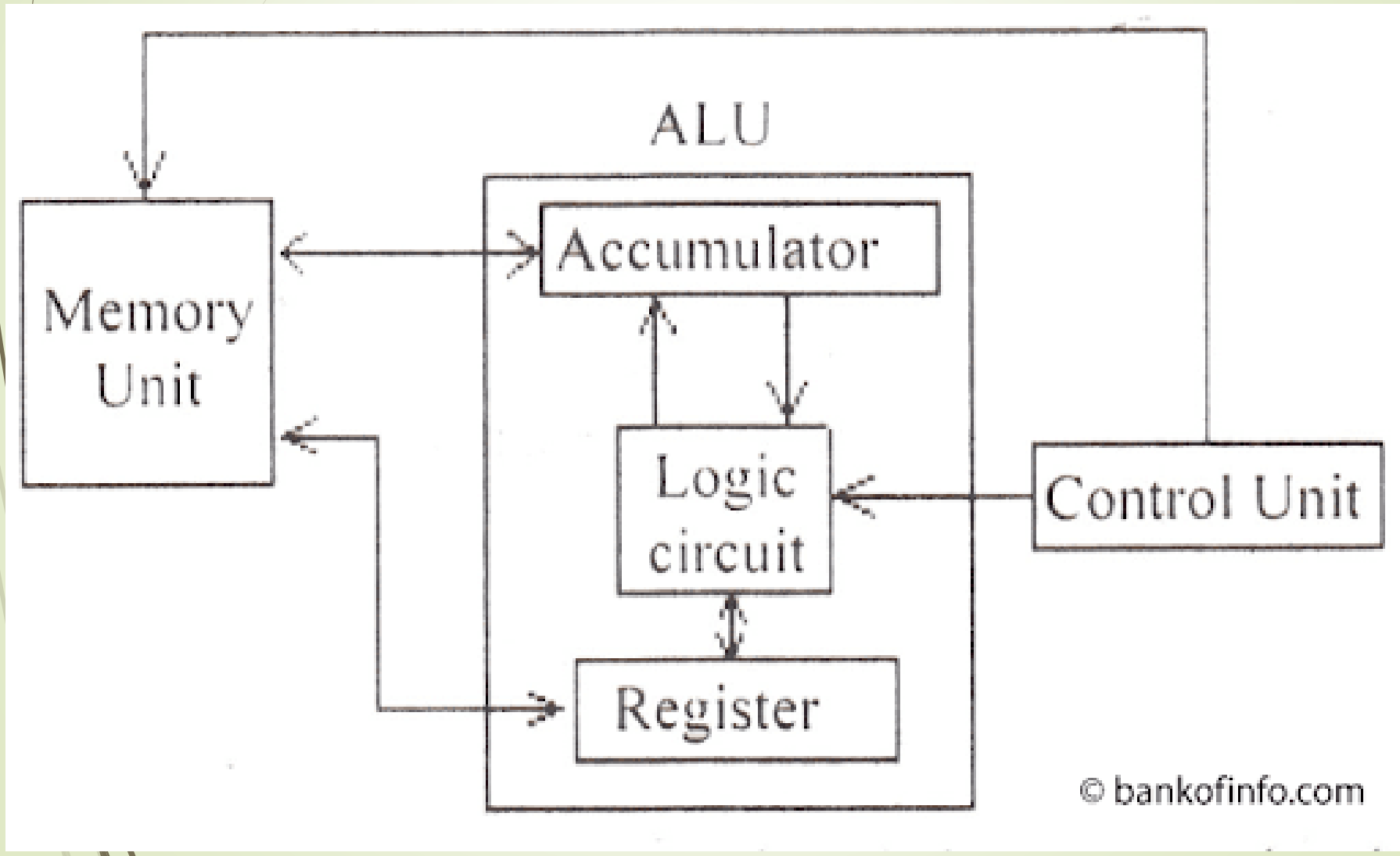
Control Unit Functions

- Address of an instruction is collected from a register.
- The instruction is then brought from the memory/cache and stored in a register.
- The instruction is then decoded to retrieve the operation code or Opcode and the operand (data or location of the data).
- If the ALU status is not busy then the Opcode is passed to the ALU.
- Control unit commands the memory to send the data to ALU.
- With both the Opcode and data in hand, ALU performs the logical/arithmetic operation onto those data.
- Control unit commands the memory to store the result which is coming out from ALU.
- Control unit starts processing a new instruction and the whole cycle restarts.

Control Unit & Arithmetic Logic Unit Functioning During an Instruction Cycle



Arithmetic Logic Unit Performing Arithmetic/Logical Operation



Arithmetic Logic Unit Performing Arithmetic/Logical Operation

ALU Function

- ALU receives the Opcode from control unit and the logic circuits are set according to the Opcode.
- Data from main memory arrives at ALU and are stored in accumulator and in register.
- Logic circuit senses the data stored in accumulator and the register and performs the operation according to the Opcode.
- The result of the operation is stored in accumulator and is sent back to the memory as per control unit's command.

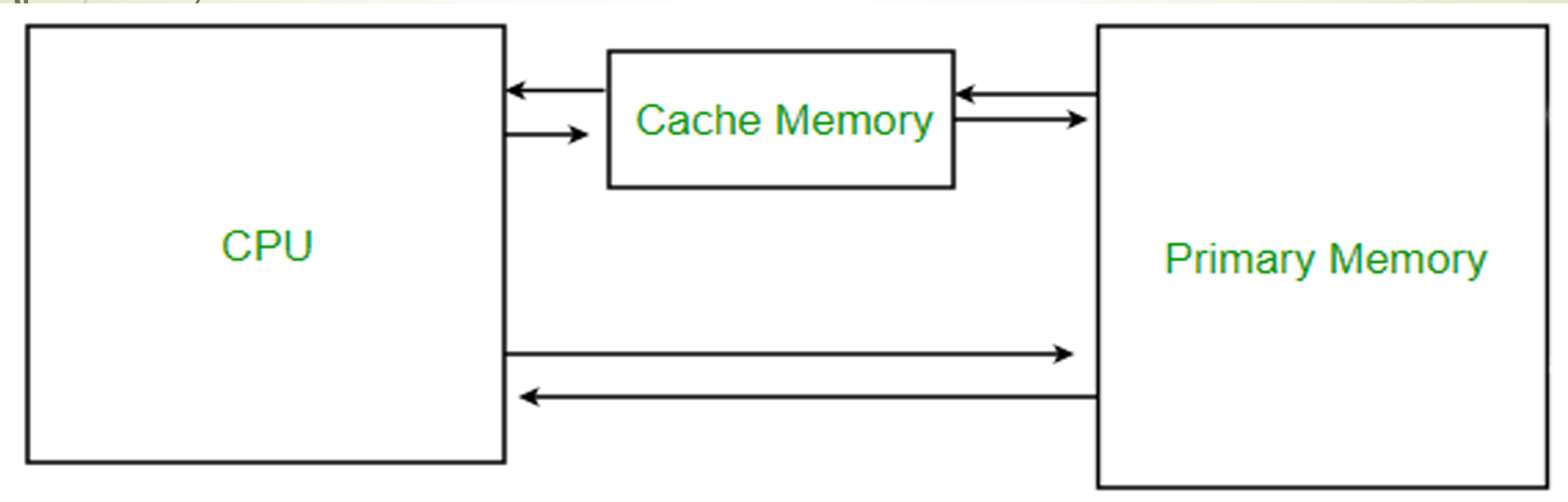
Good to Know

Register: Registers are quickly accessible memories of very short length found inside the processor. Their capacity is measured in bits and the length of the binary data that a specific register can hold defines its capacity or length.
Usually, a register can be accessed, read or written with data in a single clock pulse and no memory in a computer is faster than a register.

Computer bus: Computer bus is a group of wires that carries a signal and according to type of its payload, there are three types of buses:
Address bus - Carries memory addresses from the processor to other components such as primary storage and input/output devices.
Data bus - Carries the data between the processor and other components.
Control bus - Carries control signals from the processor to other components.

Cache Memory

Cache memory is used to store frequently used data that can be retrieved way more faster instead of collecting the data from its original storage. In computing, this cache is utilized in various field. For example, internet browsers keep cache so that they don't have to load everything from the internet. However, here we are going to concentrate on CPU cache where the CPU stores copies of frequently used data that are originally stored in main memory or RAM.



How CPU Cache Improves Performance

The CPU cache is a very high speed but expensive piece of memory. To minimize cost, the CPU comes with a relatively small amount of cache compared to the main memory.

In a system with no cache memory, every time the CPU requests for data, it would send the request to the main memory which would then provide the data to the CPU. Because the modern processors are way faster than the main memory, the data from the memory will cost several CPU cycles to reach the CPU and during that period the CPU will be idle.

As cache stores frequently used data and can provide them to CPU faster, the system performance will improve. And it is very common that a result generated from past instruction processings are required as data for later instructions.

How CPU Cache Improves Performance

Levels of Cache: Cache memory is categorized into three levels based on its ease and speed of access. Modern processors uses specially designed algorithm to guess the data that can be required in future operations.

Level 1(L1) Cache: Closest and fastest of all. It only stores the copy of most used data. An old processor like Intel Pentium 2117U has a total of 64KB of L1 cache for each of its 2 cores.

Level 2(L2) Cache: Anything that fails to get in L1 cache, are stored in L2 cache. L2 cache has greater capacity but lacks the speed of L1. CPU looks in L1 cache before L2. Intel Pentium 2117U has a total of 256KB of L2 cache for each of its 2 cores.

Level 3(L3) Cache: L3 is the largest and slowest of all. It still out performance the main memory. It stores the copy of all those data that failed to get a place in L1 and L2 cache. 2MB of single L3 cache is built inside the afore mentioned processor.

How CPU Cache Improves Performance

Hit and Miss

A cache **hit** occurs when the requested data can be found in a cache, while a cache **miss** occurs when the requested data is not in any of the cache. Cache hits are served by reading data from the cache, which we by this time know is faster than re-computing a result or reading from main memory. The more requests that can be served from the cache, the faster the system performs.

To do that-

We have to store more data in cache.

To store more data in cache-

We have to buy processor with larger cache.

To buy processor with large cache-

We need to spend more cash.

So, the more **CASH** you spend the more **CACHE MEMORY** you get.

Some other clever CPU architecture to improve performance

Multicore Processor: Most modern processors are multicore processors. These cores have their own cache and can process their own instructions simultaneously.

Hyper-threading: This technique allows a core to switch to instructions from another program or thread when there is a wait due to latency of memory access. Most cheap processors now can handle twice the number of threads for a given number of cores. Because of this, the effective number of cores (or we can say processors) are doubled from the view point of Operating System.

Some other clever CPU architecture to improve performance

Basic five-stage pipeline

Clock cycle / Instr. No.	1	2	3	4	5	6	7
1	IF	ID	EX	MEM	WB		
2		IF	ID	EX	MEM	WB	
3			IF	ID	EX	MEM	WB
4				IF	ID	EX	MEM
5					IF	ID	EX

(IF = Instruction Fetch, ID = Instruction Decode, EX = Execute, MEM = Memory access, WB = Register write back).

In the fourth clock cycle (the green column), the earliest instruction is in MEM stage, and the latest instruction has not yet entered the pipeline.

Pipelining: Pipelining is also known as instruction pipelining and in this technique, the stages of a single instruction is divided into more stages, as many as possible. During a clock pulse, various parts of CPU are kept busy working with various stages of instruction from multiple instructions.

Stay Home, Stay Safe
Always put on a mask when you are in public!