Performance

Learning Intentions

Describe the factors affecting computer system performance:

- Multi-core processors
- Width of data bus
- Cache memory

Make it smaller…

In 1965 the co-founder of Intel made a prediction that:

"The number of transistors incorporated in a chip will approximately double every 24 months."
Gordon Moore, Intel Co-founder

So far this 'law' has held true but there are physical limits that will be reached.
Current processors can be manufactured using 14 nanometre transistors

Video Example

https://www.youtube.com/watch?v=YIkMaQJSyP8

Multi-Core Processors

Some processors will be dual, quad, hex or even more **multi-core processors**

This means multiple processor units (cores) on the processor die

Each core is capable of executing its own program processes/tasks

Although code has to be written to take advantage of this capability

This means that multiple processes can be performed **concurrently**

Processor Clock Speed

Clock speed is the simplest measure of performance

This is the amount of operations that the computer can perform in one second

Modern home CPU's operate in the 2.4-3.2 GHz range

This means at least 2,400,000,000 **operations** per second

A **single operation** may require **multiple instructions**

Clock Speed as a measure of performance

This is most accurate as a measure of performance if the processors are of similar type etc. For example you cannot compare the clock speed of a Dual Core processor to that of a Quad core processor.

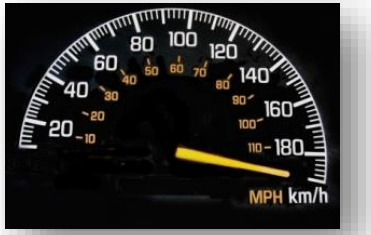
The fastest supercomputer only has a clock speed of 1.45Ghz (but it does have 40,960 of them)

Source (Jun 18):

https://www.top500.org/system/178764

https://www.top500.org/resources/top-systems/sunway-taihulight-national-supercomputing-center-i/

Speed isn't everything



Some processors have 2, four or six cores

2,3,4,6, 8 or 16 cores are the usual options

This means that there are 2, 4 or 6 smaller processors built into a single processor.

Each core can work on a **single instruction** at the **same time**

Top Supercomputer in the world (Jun 18) has access to 10,649,600 cores!

So which is better?
Dual core at 3.5Ghz or Quad core at 2.4Ghz?

Processor Configuration

- Does the system require a **multi core processor**?
  - Software will usually have to be coded and optimised for use in multi processor/multi thread environments

- If it is a particularly demanding application or a server application then it may even support or require **multiple CPU's**
  - A motherboard will need to support having multiple physical processors

Processor Type

You need to be specific in terms of the requirements of the CPU

This could be particularly limiting if developing for mobile devices

You may recommend an optimal specification as well as the ‘bare minimum’

The effect of Bus width changes

Changing the **width** of the **data bus** means that we can **move more or less data in a single operation**

- Hopefully leading to an increase in performance due to less operations required to move

Changing the **width** of the **address bus** means that we can **address more memory**

- Not necessarily leading to an increase in performance in all cases

How to measure performance?

The clock speed just measures the amount of operations that the processor can carry out.

However an instruction such as adding a number may take more than one operation.
So one appropriate measure is **MIPS (Millions of instructions per second)**

- This gives a more accurate measure of what the processor can actually achieve in a second.

Types of RAM

There are two main types of RAM

Static Ram (SRAM)
This RAM is VERY fast to access but is comparatively far more expensive than DRAM. This is usually reserved for cache memory.

Dynamic RAM (DRAM)
is the more widely used RAM and this is the type that you will use in your system
This RAM is volatile

Dynamic RAM

This is the main memory used in a computer system. Needs to be refreshed constantly

Can be ‘rated’ as PC10666 to PC14400

The PC number is its maximum *theoretical* bandwidth

Can also run at different clock speeds ranging from 1333 – 1800Mhz

Timings can also be written as 9-9-9-28

9 clock cycles to deliver data

The last number is the time for RAM to read or write from a different area in RAM

Static RAM

This is usually used for cache memory within processors or devices

Does not need to be refreshed as often as Dynamic RAM

A processor may have different levels of cache, Levels 1(Primary)-3.

The higher the level the longer the access times.

Access times can be in the area of 1-15 nanoseconds

Cache of a CPU

It guesses that if you are frequently using instructions then you may want to use them again

So it will store these in on-board cache memory.

Modern Intel processors have approx. 8MB of on-board cache (static) memory.

This saves the processor needing to utilise another fetch execute cycle, thus being more efficient.

As less memory reads required

Cache hits and misses

There is a finite amount of cache – it cannot store everything!

A **cache hit** is if a requested instruction is found in cache memory then this data is accessed in a matter of clock cycles

A **cache miss** occurs when data is not found in cache

This has to be fetched from RAM – comparatively much slower

But what gets put in the cache?

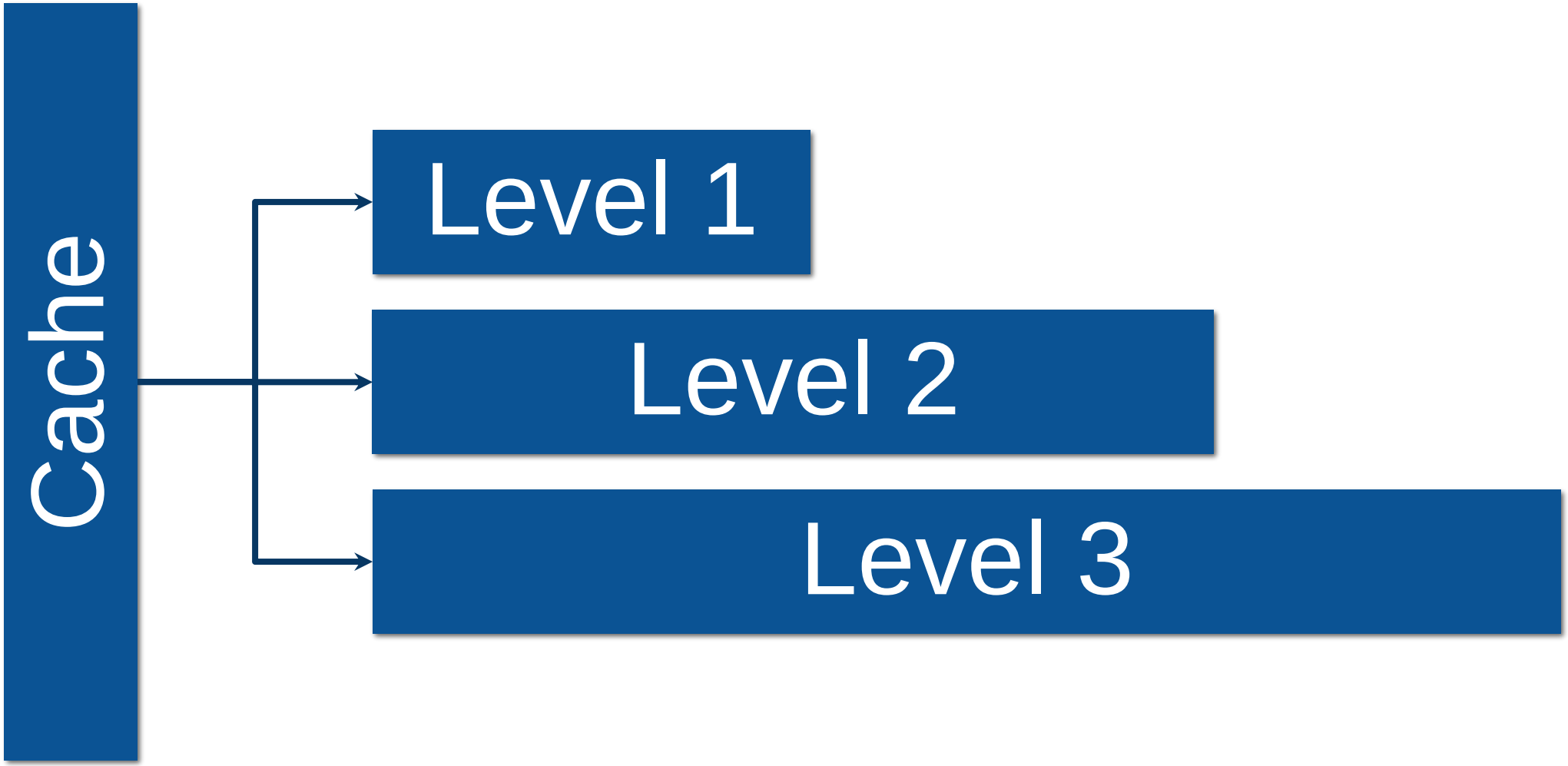
There are a few methods to optimise cache use :

- The CPU can fetch the next memory location to the one being asked for (typically 64 bytes)
- A hardware **prefetcher** can then also load a larger amount of data
  - Intel chips have a 256-byte prefetcher, so it caches the next 256 bytes after the line already loaded
- This prefetch can be triggered when successive **cache misses** occur (so **frequently used instructions** will be cached)

Modern CPU's can achieve cache hits of approx 80% using these and other techniques

More information on Intel CPU's and caching can be found here

Different Levels of Cache



Memory

Every system will require a pre-requisite amount of RAM in order to load and run

Again there may be a minimum as well as optimal amount