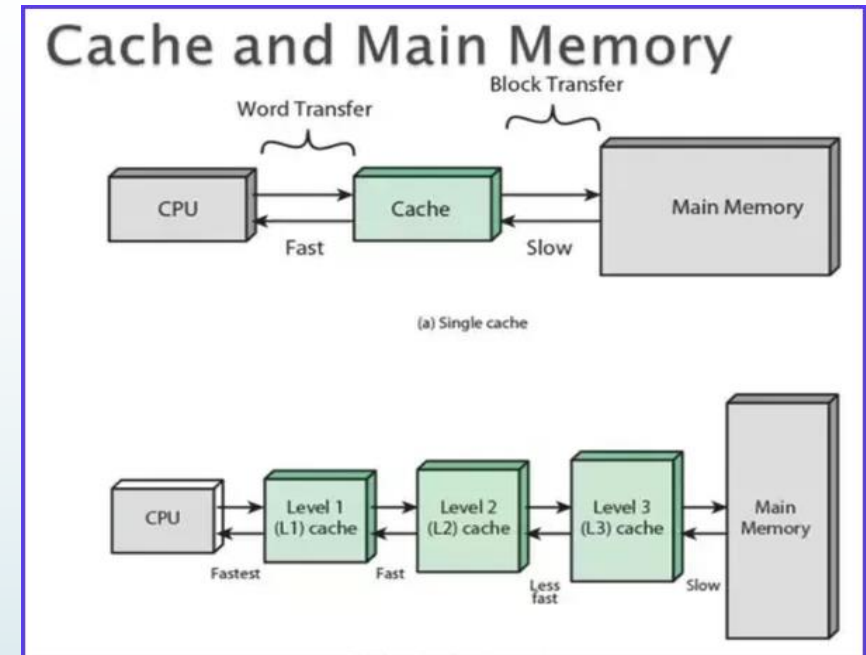



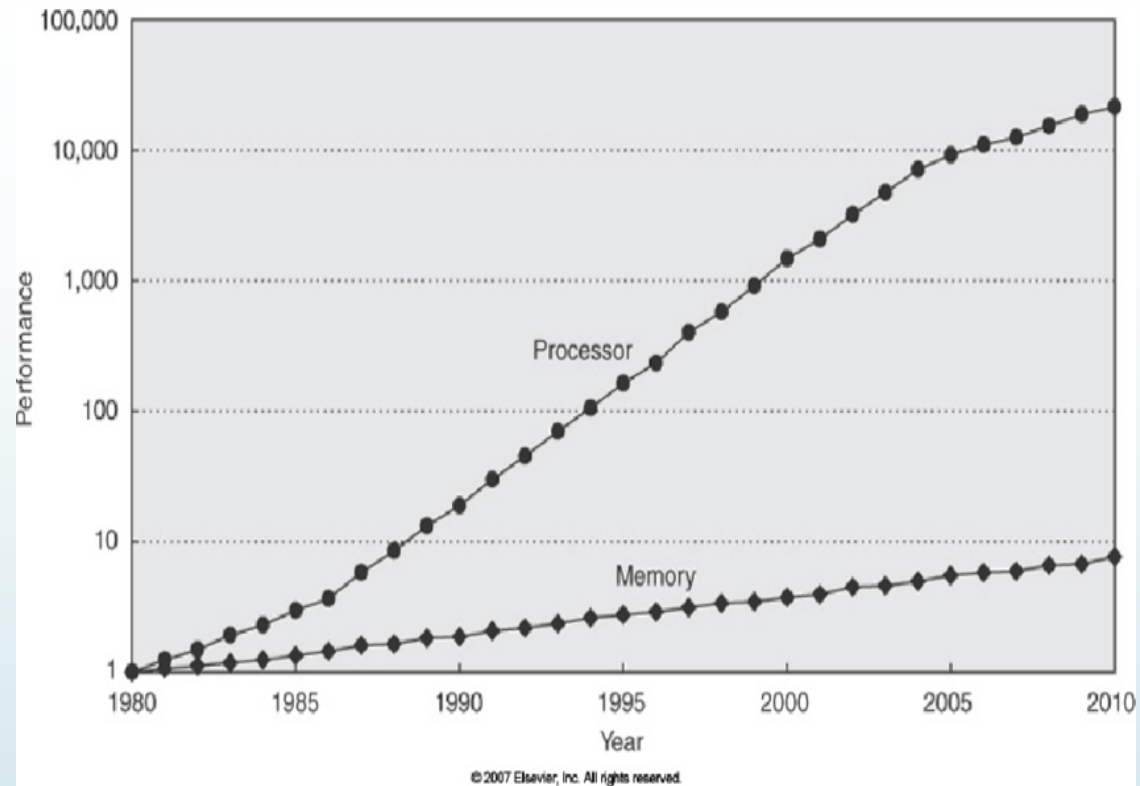
# Earliest cache Design

- The cache memory is a very high speed, expensive piece of memory, which is used to speed up the memory retrieval process.
- Due to their higher costs, the CPU comes with a relatively small amount of cache compared with the main memory.
- Example-**The Intel 80486** microprocessor, contains an 8K memory cache, and **the Pentium** has a 16K cache.

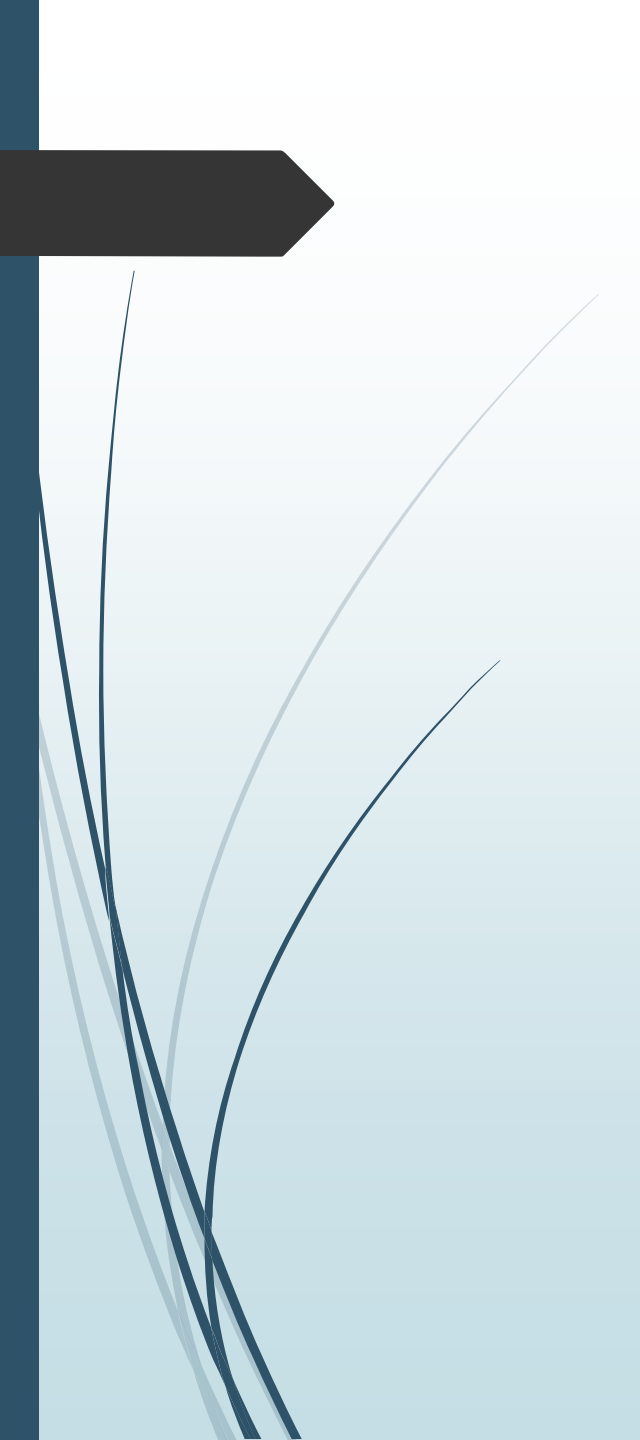


- Memory cache was first used on PCs at the **386DX** timeframe. **The Intel 80386**, also known as **i386** or just **386**, is a 32-bit microprocessor introduced in 1985.
- The amount of available memory cache varied as well depending on the motherboard model and typical values for that time were 64 KB and 128 KB. At this time the memory cache controller used an architecture known as “write-through,” i.e., when the CPU wants to store data in memory – the memory cache controller updates the RAM memory immediately.

- 
- In 1989 when Intel first put memory on the CPU with the 80486, a small amount (8 KB) of memory cache inside the CPU. This internal memory cache was called L1 (level 1) or “internal,” while the external memory cache was called L2 (level 2) or “external.” Typical amounts for that time were 128 KB and 256 KB.
  - Later 486 models added the “write back” cache architecture, which is used until today, where for write operations the RAM memory isn’t updated immediately, the CPU stores the data on the cache memory and the memory controller updates the RAM memory only when a cache miss occurs.
  - With the first Pentium processor (**Pentium-200**), Intel created two separated internal memory caches, one for instructions and another for data (at the time with 8 KB each). This architecture is still used to date, and that is why you sometimes see the L1 memory cache being referred as 64 KB + 64 KB, for example – this is because there are one 64 KB instruction L1 cache and one 64 KB data L1 cache.
  - The problem with the L2 memory cache being external is that it is accessed with a lower clock rate, because since 486DX2 the CPU internal clock rate is different from the CPU external clock rate. While a Pentium-200 worked internally at 200 MHz, it accessed its L2 memory cache at 66 MHz, for example. Then with P6 architecture Intel moved the memory cache from the motherboard to inside the CPU – what allowed the CPU to access it with its internal clock rate .
  - This same architecture is used until today: both L1 and L2 memory caches are located inside the CPU running at the CPU internal clock rate.



**Moore's Law** states that we can expect the speed and capability of our computers to increase every couple of years, and we will pay less for them.



Thank You...