CS1101: Lecture 15

Computer Systems Organization: Primary Memory

Dr. Barry O'Sullivan
b.osullivan@cs.ucc.ie

Course Homepage
http://www.cs.ucc.ie/~osullb/cs1101

Lecture Outline

- Bits
- Memory Addresses
- Organising Memory
- Words
- Error-Correcting Codes
- Cache Memory
- Memory Packaging and Types
- Memory Configurations
- **Reading**: Tanenbaum, Chapter 2 Section 2

Bits

- The basic unit of memory is the binary digit, called a **bit**.
- A bit may contain a 0 or a 1.
- The binary number system requires only two values to be distinguished.
- Consequently, it is the most reliable method for encoding digital information.

Memory Addresses

- Memories consist of a number of **cells** (or **locations**) each of which can store a piece of information.
- Each cell has a number, called its **address**, by which programs can refer to it.
- If a memory has n cells, they will have addresses 0 to $n-1$.
- All cells in a memory contain the same number of bits.
- If a cell consists of k bits, it can hold any one of 2^k different bit combinations.

Organising Memory

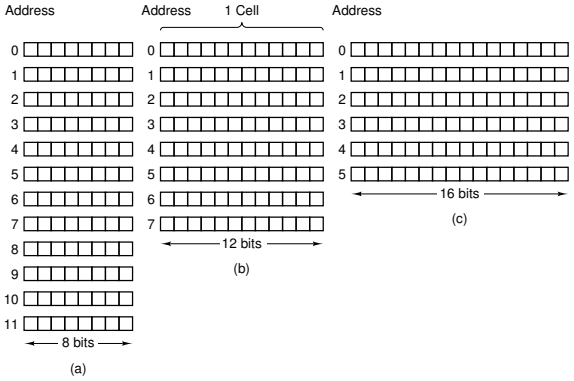

Figure 2.9 Three ways of organising a 96-bit memory.

Organising Memory

- Adjacent cells have consecutive addresses (by definition).
- Computers that use the binary number system (including octal and hexadecimal notation for binary numbers) express memory addresses as binary numbers.
- If an address has m bits, the maximum number of cells addressable is 2^m.
- For example, an address used to reference the memory of Fig. 2-9 (a) needs at least 4 bits in order to express all the numbers from 0 to 11.
- A 3-bit address is sufficient for Fig. 2-9 (b) and (c), however.

Organising Memory (continued)

- The number of bits in the address determines the maximum number of directly addressable cells in the memory and is independent of the number of bits per cell.
- For example, a memory with 2^{12} cells of 8 bits each and a memory with 2^{12} cells of 64 bits each need 12-bit addresses.

Words

- The significance of the cell is that it is the smallest addressable unit.
- Nearly all manufacturers have standardised on an 8-bit cell, which is called a **byte**
- Bytes are grouped into **words**
- A computer with a 32-bit word has 4 bytes/word, whereas a computer with a 64-bit words has 8 bytes/word.
- Most instructions work on words
- For example, a 32-bit machine will have 32-bit registers and instructions for manipulating 32-bit words – such as adding two words together.

Error-Correcting Codes

- Computer memories can make errors occasionally due to voltage spikes on the power line or other causes.
- To guard against such errors, some memories use error-detecting or error-correcting codes.
- When these codes are used, extra bits are added to each memory word in a special way.
- When a word is read out of memory, the extra bits are checked to see if an error has occurred.

Error-Correcting Codes

- Suppose that a memory word consists of m data bits to which we will add r redundant, or check bits.
- Let the total length be n (i.e., $n = m + r$).
- An n-bit unit containing data and r check bits is often referred to as an n-bit **codeword**.
- The number of bit positions in which two codewords differ is called the **Hamming distance**.
- The error-detecting and error-correcting properties of a code depend on its Hamming distance.

Cache Memory

- Techniques are known for combining a small amount of fast memory with a large amount of slow memory to get the speed of the fast memory (almost) and the capacity of the large memory at a moderate price.
- The small, fast memory is called a **cache** (pronounced ”cash”).
- The basic idea behind a cache is simple:
  - The most heavily used memory words are kept in the cache.
  - When the CPU needs a word, it first looks in the cache.
  - Only if the word is not there, does it go to main memory.
- If a substantial fraction of the words are in the cache, the average access time can be greatly reduced.

Cache Memory

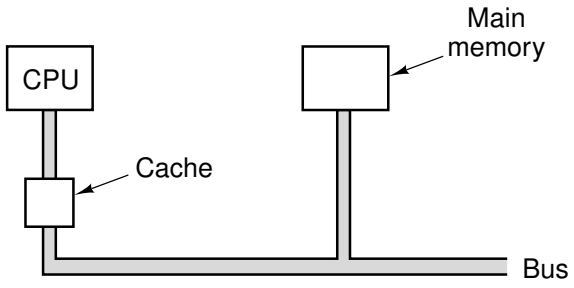

Figure 2-16. The cache is logically between the CPU and main memory. Physically, there are several possible places it could be located.

Memory Packaging and Types

- Until the early 1990s, memory was manufactured, bought, and installed as single chips.
- Chip densities went from 1k bits to 1M bits and beyond, but each chip was a separate unit.
- Today, a different arrangement is often used.
- A group of chips, typically 8 or 16, is mounted on a tiny printed circuit board and sold as a unit.
- This unit is called a SIMM (Single Inline Memory Module) or a DIMM (Dual Inline Memory Module), depending on whether it has a row of connectors on one side or both sides of the board.

Memory Packaging and Types

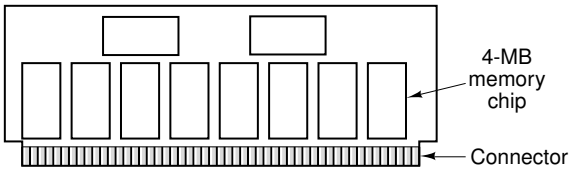

Figure 2-17. A single inline memory module (SIMM) holding 32 MB. Two of the chips control the SIMM.

Memory Configurations

- A typical SIMM configuration might have eight chips with 32 megabits (4 MB) each on the SIMM.
- The entire module would then hold 32 MB.
- Many computers have room for four modules, giving a total capacity of 128 MB when using 32-MB SIMMS.
- Often these SIMMs can later be replaced by 64-MB or larger SIMMs as needed.