

Primary memory

Prof Rajeev Barua
E-A 001 -- Slide set 6



The primary memory is the memory that is directly addressable by the CPU.

It is stored in the main memory, and a part of it is stored in the cache memory for speed.

- Main memory is built of DRAM (Dynamic Random Access Memory)
 - Semiconductor memory built using capacitances, not gates!
 - Advantage: small area per byte \Rightarrow low cost \Rightarrow larger memory
- Cache memory is built of SRAM (Static Random Access Memory)
 - Semiconductor memory built using logic gates (like AND, OR, NOT)
 - Each bit is stored in one Flip-Flop, which is the fundamental unit of storage for boolean logic.
 - Gates are also used to build all of the CPU.
 - Advantage: Faster access, but more expensive.

Memory is composed of collections of "bits" (binary digits)

- Groups of 8 bits = byte (usually the smallest addressable unit)

Most accesses on modern computers are in words:

- The length of word is machine dependent:
 - On 32-bit machines = 1 word = 32 bits
 - On 64-bit machines = 1 word = 64 bits
- Could be smaller on embedded processors.

An address of k bits can access 2^k bytes of memory, since most memory is byte addressed.

- Hence a memory of M bytes requires ceiling ($\log_2 M$) bits to address it.

Byte ordering

Byte ordering refers to the order of bytes in a word.

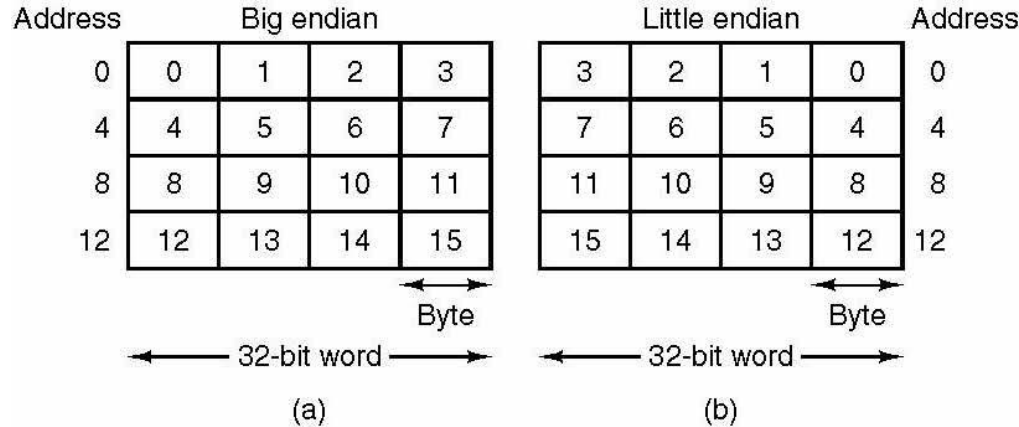


Figure 2-11. (a) Big endian memory. (b) Little endian memory.

- Numbering above implies low number is high-significance bits within a word.
- Endianness only refers to the order of bytes in memory, not registers!!
- Different ISAs have different endianness. Look up which one is which.
- Having machines with different endianness can lead to bugs in software when file on one machine is transmitted to another if transferred naively.
 - Solution: File transfer software such as FTP must change order of bytes within words when transferring across machines with different endianness.