Computer Organization

Part 20 : Memory Organization Basics

UNIT – IV

Memories

- Used to **store information** within a computer, either programs or data.
  - Programs and data cannot be used directly from a disk or CD, but must first be moved in memory.

- **Main Memory & Cache Memory** refers as **internal memory** because it is placed at the main board. Communicates directly with CPU immediately.

- **Secondary & tertiary memory** refers as **external memory** (or auxiliary memory) because it is not located at the main board. Usually for back-up purpose.
<table>
<thead>
<tr>
<th>Tertiary</th>
<th>Secondary</th>
<th>Primary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow</td>
<td>Slow</td>
<td>Fast</td>
</tr>
<tr>
<td>Not directly connected to CPU</td>
<td>Not directly connected to CPU</td>
<td>Directly connected to CPU</td>
</tr>
<tr>
<td>Cheap</td>
<td>Large volume</td>
<td>Expensive</td>
</tr>
<tr>
<td>Large volume</td>
<td>Eg: Disk</td>
<td>Eg: Cache/RAM</td>
</tr>
</tbody>
</table>

### Memory Locations

- Each part of memory has a separate memory location, which can be referred to using a memory address.
Unit Terms

- Size of memory is measured in bytes (or multiples such as kilobytes (KB) or megabytes (MB))

<table>
<thead>
<tr>
<th>Unit</th>
<th>Exact Number of bytes</th>
<th>Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>kilobyte</td>
<td>2^10 bytes</td>
<td>10^3 bytes</td>
</tr>
<tr>
<td>megabyte</td>
<td>2^20 bytes</td>
<td>10^4 bytes</td>
</tr>
<tr>
<td>gigabyte</td>
<td>2^30 bytes</td>
<td>10^5 bytes</td>
</tr>
<tr>
<td>terabyte</td>
<td>2^40 bytes</td>
<td>10^6 bytes</td>
</tr>
<tr>
<td>petabyte</td>
<td>2^50 bytes</td>
<td>10^7 bytes</td>
</tr>
<tr>
<td>exabyte</td>
<td>2^60 bytes</td>
<td>10^8 bytes</td>
</tr>
</tbody>
</table>

- Number of bits for an address to uniquely access a memory location

\[
\text{Number of locations} = 2^{\frac{\log \text{memory capacity}}{\log 2}}
\]

Memory Characteristics

- Location
- Capacity
- Unit of transfer
- Access method
- Performance
- Physical type
- Physical characteristics
- Organisation
Memory Location

- CPU
- Internal
- External

Memory Capacity

- Word size
  - The natural unit of organisation
  - Common word size: 8, 16, 32 bits.
- Number of words
  - or Bytes
Memory Unit of Transfer

- **Internal**
  - The number of bits read-out of or written into memory at a time.
  - Usually governed by data bus width
- **External**
  - Usually a block which is much larger than a word
- **Addressable unit**
  - Smallest location which can be uniquely addressed
  - Word internally

Memory Access Methods (1)

- **Sequential**
  - Start at the beginning and read through in order
  - Access time depends on location of data and previous location
  - e.g. tape
- **Direct**
  - Individual blocks have unique address
  - Access is by jumping to vicinity plus sequential search
  - Access time depends on location and previous location
  - e.g. disk
Access Methods (2)

- **Random**
  - Individual addresses identify locations exactly
  - Access time is independent of location or previous access
  - e.g. RAM

- **Associative**
  - Data is located by a comparison with contents of a portion of the store
  - Access time is independent of location or previous access
  - e.g. cache

Memory Performance

- **Access time**
  - Time between presenting the address and getting the valid data

- **Memory Cycle time**
  - Time may be required for the memory to “recover” before next access
  - Cycle time is access + recovery

- **Transfer Rate**
  - Rate at which data can be moved
Memory Transfer Rate

- Transfer rate for random-access memory
  \( = \frac{1}{\text{cycle time}} \)

- Transfer rate for non-random-access memory
  \( = T_N = T_A + \frac{N}{R} \)

  Where
  - \( T_N \) = average time to read or write \( N \) bits
  - \( T_A \) = average access time
  - \( N \) = number of bits
  - \( R \) = transfer rate, in bits per second (bps)

Memory Physical Types

- Semiconductor
  - RAM & ROM
- Magnetic
  - Disk & Tape
- Optical
  - CD & DVD
- Others
  - Bubble
  - Hologram
Physical Characteristics

- Decay
- Volatility
- Erasable
- Power consumption

Summary...

<table>
<thead>
<tr>
<th>Technology</th>
<th>Storage type</th>
<th>Access method</th>
<th>Alterability</th>
<th>Permanence</th>
<th>Typical access time $t_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semiconductor RAM</td>
<td>Electronic</td>
<td>Random</td>
<td>Read/Write</td>
<td>Volatile</td>
<td>10 ns</td>
</tr>
<tr>
<td>Semiconductor ROM</td>
<td>Electronic</td>
<td>Random</td>
<td>Read only</td>
<td>Non volatile</td>
<td>10 ns</td>
</tr>
<tr>
<td>Magnetic (Hard) disk</td>
<td>Magnetic</td>
<td>Semirandom</td>
<td>Read/Write</td>
<td>Non volatile</td>
<td>50 ns</td>
</tr>
<tr>
<td>Optical disk CD-ROM</td>
<td>Optical</td>
<td>Semirandom</td>
<td>Read only</td>
<td>Non volatile</td>
<td>100 ms</td>
</tr>
<tr>
<td>Erasable optical disk</td>
<td>Optical</td>
<td>Semirandom</td>
<td>Read/Write</td>
<td>Non volatile</td>
<td>100 ms</td>
</tr>
<tr>
<td>Magnetic tape</td>
<td>Magnetic</td>
<td>Serial</td>
<td>Read/Write</td>
<td>Non volatile</td>
<td>1 s Depends on access location</td>
</tr>
</tbody>
</table>
### Serial vs. Random Access

<table>
<thead>
<tr>
<th>Serial Access</th>
<th>Random Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serially data is accessed</td>
<td>Data is accessed randomly</td>
</tr>
<tr>
<td>Memory organized records are accessed sequentially</td>
<td>The unique address of memory is accessed independently</td>
</tr>
<tr>
<td>Memory access time depends upon position of storage location</td>
<td>Memory access time is independent on position of storage location</td>
</tr>
<tr>
<td>Fast memory access time</td>
<td>Memory access time is less</td>
</tr>
<tr>
<td>Cheaper</td>
<td>Comparatively costly</td>
</tr>
<tr>
<td>Non volatile</td>
<td>Volatile / non-volatile</td>
</tr>
<tr>
<td>e.g. Magnetic Tapes</td>
<td>e.g. Semiconductor Memories</td>
</tr>
</tbody>
</table>

### Memory Hierarchy

- Memory systems (a collection of various forms of memory) are constructed in a hierarchy
- **Why?**
  - Rule of thumb: the faster the memory the higher the cost in terms of price, making it very expensive to make all the memory out of the fastest memory devices.
  - Slower technologies are less expensive, making it more practical to make larger memories out of these devices.
- **Goal of a memory hierarchy**
  - to keep the data that is accessed most high up the hierarchy, so it can be accessed quickly
  - the least used at the bottom of the hierarchy.
Memory hierarchy ..
Memory Hierarchy ..

- At the top - registers
  - these are fast,
  - but only provide a very limited and temporary storage
  - usually part of the processor – in CPU
  - Expensive
- The next level - cache memory
  - expensive
  - fast access time (time taken to access the data stored).
  - The amount that can be stored (capacity) in cache is greater than is stored in the registers, but is slower
- The next level - main memory/internal memory
  - Greater capacity than cache, but is slower than cache to access.
  - RAM
- At the bottom of the hierarchy – secondary storage/external memory
  - the greatest capacity, but is the slowest to access
  - Backup/storage

Reference

- Computer Architecture and Organization
  - By A. P. Godse