Our fifth view of computer organization: The modern digital computer has three major functional hardware units: CPU, Main Memory and Input/Output (I/O) Units.

Processor - Memory Gap
- Performance gap: CPU (55% each year) vs. DRAM (7% each year)
  - Processor operations take of the order of 1 ns
  - Memory access requires 10s or even 100s of ns
  - Each instruction executed involves at least one memory access
Memory Organization

- Memory is organized and accessed in ways to hide this gap

The fastest memories are expensive and thus not very large

<table>
<thead>
<tr>
<th>Capacity</th>
<th>Access Time</th>
<th>Cost (per GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100s B</td>
<td>ns</td>
<td>$Millions</td>
</tr>
<tr>
<td>10s KB</td>
<td>few ns</td>
<td>$100s Ks</td>
</tr>
<tr>
<td>MBs</td>
<td>10s ns</td>
<td>$10s Ks</td>
</tr>
<tr>
<td>100s MB</td>
<td>100s ns</td>
<td>$100s</td>
</tr>
<tr>
<td>10s GB</td>
<td>10s ms</td>
<td>$10s</td>
</tr>
</tbody>
</table>
Computer Organization v5

- Our sixth view of computer organization
- A fast memory can help bridge the CPU-memory gap
- The fastest memories are expensive and thus not very large

Memory Technology

- Random Access Memory (RAM)
  - Any byte of memory can be accessed without touching the preceding bytes
  - RAM is the most common type of memory found in computers and other digital devices
  - There are two main types of RAM
    - DRAM (Dynamic Random Access Memory)
      - Needs to be "refreshed" regularly (~ every 8 ms)
      - 1% to 2% of the active cycles of the DRAM
      - Used for Main Memory
    - SRAM (Static Random Access Memory)
      - Content will last until power turned off
      - Low density (6 transistor cells), high power, expensive, fast
      - Used for caches
Memory Technology

- Single-transistor DRAM cell is considerably simpler than SRAM cell
- This leads to dense, high-capacity DRAM memory chips

![Diagram of DRAM and SRAM cells]

RAM Organization

- One memory row holds a block of data, so the column address selects the requested bit or word from that block
- RAS or Row Access Strobe triggering row decoder
- CAS or Column Access Strobe triggering column selector

![Diagram of RAM organization]

Each intersection represents a 6-T SRAM cell or a 1-T DRAM cell.
RAM Organization

- **Latency**: Time to access one word
  - **Access time**: time between the request and when the data is available (or written)
  - **Cycle time**: time between requests
  - Usually cycle time > access time
- **Bandwidth**: How much data from the memory can be supplied to the processor per unit time
  - Width of the data channel * The rate at which it can be used

DRAM Packaging

- **DIMM (Dual Inline Memory Module)** contains multiple chips arranged in “ranks”
  - Each rank has clock/control/address signals connected in parallel (sometimes need buffers to drive signals to all chips), and data pins work together to return wide word
  - A modern DIMM usually has one or two ranks (occasionally 4 if high capacity)
**Disk Memory Basics**

- Hard disk drive (HDD), hard disk, hard drive is the dominant secondary storage device in computer systems.

1. Head movement from current position to desired cylinder: **Seek time** (0 - 10 ms)

2. Disk rotation until the desired sector arrives under the head: **Rotational latency** (0 - 50 ms)

3. Disk rotation until sector has passed under the head: **Data transfer time** (< 1 ms)

**Solid-State Drive (SSD)** uses integrated circuits and has no moving mechanical components:

- Low latency
- Low power
- Solid state reliability
- Widening application range
  - Embedded devices
  - Desktop and laptop PC
  - Server and supercomputer
Intel Haswell

4th Generation Intel® Core™ Processor Die Map
22nm Haswell 14-Gate 3-D Transistors

Next Class

- Exam II  Lectures L09 – L16