Computer Memory and the Future of NVM

Caleb Froelich
Memory Market Overview

- **Volatile:**
  - (1964) – SRAM is used for limited, fast working memory – expensive and power-hungry.
  - (1967) – DRAM is used for high volumes of working memory – complex and slower than SRAM.

- **Non-Volatile:**
  - (1956) – PROM is used for storage of permanent data, usually low-level programs – data cannot be erased or changed.
  - (1972) – EEPROM can be erased and re-programmed. Still utilized today in modem, video cards and many electronic gadgets.
  - (1981) – **Flash memory is invented!** – A universal non-volatile memory type that is used in most computers as a storage medium.
    - Utilizes floating gate transistors.
    - MUCH, much faster.
    - Ability to clear and rewrite chunks of data.

"the number of transistors in a dense integrated circuit doubles about every two years"

What's the problem?
Emerging Developments in NVM

- Magnetoresistive Memory (MRAM):
  - Uses a magnetic layer to store a bit.
    - Denser than SRAM which results in a net less expensive chip.
    - Simpler to access than Flash memory and DRAM.
    - Faster and utilizes less power than other NVM.

- Ferroelectric Memory (FRAM):
  - Uses materials that exhibit magnetic-like properties:
    - Low power consumption – 1.5v compared to 10-14v for Flash.
    - Performs well across a wide spectrum of temperatures.
    - SPEED – write time is less than 50ns, 1000x faster than Flash/EEPROM.

- Phase-Change Memory (PCM)
  - PCM works by changing the phase of a special kind of glass (Chalcogenide glass) within the bit cell.
    - Faster than Flash – significantly better read and write times. No erase cycle required.
    - Lower power consumption.
    - Not susceptible to scaling difficulties.
    - Not yet commercially viable.

Emerging Developments in NVM
Future of NVM

- **3-D XPoint:**
  - Transistor-less with memory cells at the intersection of a massive amount of perpendicular wires.
  - A type of PCM.
  - Faster than NAND with greater endurance.
  - Twice as dense as DRAM.

- **Spin Transfer Torque RAM (STT-RAM):**
  - Uses electron spins to create a current that changes the magnetic orientation of a bit stored in a Magnetic Tunnel Junction (MTJ).
  - The BEST of all worlds!
  - Higher density, better scalability than CMOS technology.
  - Comparable performance to SRAM.
  - Low power consumption.

- **Goals for the future:**
  - NVM memory becomes embedded.
  - Reduce size below Flash’s scalability limit.
  - Replace levels 2 and 3 cache with NVM.
  - Reduce cost of emerging memory to increase competitiveness.

---

<table>
<thead>
<tr>
<th>Non-Volatile</th>
<th>SRAM</th>
<th>DRAM</th>
<th>Flash (NOR)</th>
<th>Flash (NAND)</th>
<th>FRAM</th>
<th>MRAM</th>
<th>PCM</th>
<th>STT-RAM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cell size ($F^2$)</td>
<td>50-120</td>
<td>6-10</td>
<td>10</td>
<td>5</td>
<td>15-34</td>
<td>16-40</td>
<td>6-12</td>
<td>6-20</td>
</tr>
<tr>
<td>Read time (ns)</td>
<td>1-100</td>
<td>30</td>
<td>10</td>
<td>50</td>
<td>20-80</td>
<td>3-20</td>
<td>20-50</td>
<td>2-20</td>
</tr>
<tr>
<td>Write/Erase time (ns)</td>
<td>1-100</td>
<td>50 / 50</td>
<td>10 ms / 100 ms</td>
<td>1 ms / 0.1 ms</td>
<td>50 / 50</td>
<td>3-20</td>
<td>50 / 120</td>
<td>2-20</td>
</tr>
<tr>
<td>Endurance</td>
<td>$10^{15}$</td>
<td>$10^{16}$</td>
<td>$10^5$</td>
<td>$10^{12}$</td>
<td>$10^9$</td>
<td>$&gt;10^{23}$</td>
<td>$10^{32}$</td>
<td>$&gt;10^{75}$</td>
</tr>
<tr>
<td>Voltage required</td>
<td>No</td>
<td>2 V</td>
<td>6-8 V</td>
<td>16-20 V</td>
<td>2-3 V</td>
<td>3 V</td>
<td>1.5-3.5 V</td>
<td>&lt;1.5 V</td>
</tr>
<tr>
<td>Cost</td>
<td>$$$$</td>
<td>$$</td>
<td>$</td>
<td>$</td>
<td>$$</td>
<td>$$</td>
<td>$$</td>
<td>?</td>
</tr>
</tbody>
</table>
Sources

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4182445/