EMERGING MEMORY SYSTEMS

Mahdi Nazm Bojnordi
Assistant Professor
School of Computing
University of Utah
Overview

- Upcoming deadline
  - March 29th: sign up for student paper presentation

- This lecture
  - DRAM technology scaling issues
  - Charge vs. phase based memory
  - Phase change memory
DRAM Cell Structure

- One-transistor, one-capacitor
  - Realizing the capacitor is challenging

- 1T-1C DRAM
- Charge based sensing
- Volatile
Memory Scaling in Jeopardy

Scaling of semiconductor memories greatly challenged beyond 20nm

The World’s Tallest Buildings

A skyscraper being built in Jeddah aims for a record-breaking height that’s expected to exceed 3,281 feet.

<table>
<thead>
<tr>
<th>Name</th>
<th>Height (feet)</th>
<th>Completed</th>
<th>Est. Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeddah Tower (under construction)</td>
<td>&gt; 3,281</td>
<td>2018</td>
<td></td>
</tr>
<tr>
<td>Burj Khalifa (world’s tallest building), Dubai</td>
<td>2,717</td>
<td>2010</td>
<td></td>
</tr>
<tr>
<td>Shanghai Tower, Shanghai</td>
<td>2,073</td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>Makkah Royal Clock Tower, Mecca</td>
<td>1,972</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>One World Trade Center, New York</td>
<td>1,776</td>
<td>2014</td>
<td></td>
</tr>
<tr>
<td>Taipei 101, Taipei</td>
<td>1,667</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>Chrysler Building (61st tallest), New York</td>
<td>1,046</td>
<td>1930</td>
<td></td>
</tr>
</tbody>
</table>

Note: Includes antennas  
Source: The Council on Tall Buildings and Urban Habitat

A/R=10
Addressing DRAM Issues

- Overcome DRAM shortcomings with
  - System-DRAM co-design
  - Novel DRAM architectures, interface, functions
  - Better waste management (efficient utilization)

- Key issues to tackle
  - Reduce refresh energy
  - Improve bandwidth and latency
  - Reduce waste
  - Enable reliability at low cost
Alternative to DRAM

Key concept: replace DRAM cell capacitor with a programmable resistor

- 1T-1C DRAM
- Charge based sensing
- Volatile

- 1T-1R STT-MRAM, PCM, RRAM
- Resistance based sensing
- Non-volatile
Charge vs. Phase

- **Charge Memory (e.g., DRAM, Flash)**
  - Write data by capturing charge $Q$
  - Read data by detecting voltage $V$

- **Resistive Memory (e.g., PCM, STT-MRAM, memristors)**
  - Write data by pulsing current $dQ/dt$
  - Read data by detecting resistance $R$
Limits of Charge Based Memory

- Difficult charge placement and control
  - Flash: floating gate charge
  - DRAM: capacitor charge, transistor leakage
- Reliable sensing becomes difficult as charge storage unit size reduces

[slide ref: Mutlu]
Leading Contenders

**STT-MRAM**
- Limited to single-level cell
- 3D un-stackable
+ High endurance ($\sim10^{15}$)
+ $\sim4$ns switching time
+ $\sim50$uW switching power

**PCM-RAM**
+ Multi-level cell capable
+ $4F^2$ 3D-stackable cell
- Endurance: $\sim10^9$ writes
- $\sim100$ns switching time
- $\sim300$uW switching power

**R-RAM**
+ Multi-level cell capable
+ $4F^2$ 3D-stackable cell
- Endurance: $10^6$~$10^{12}$ writes
+ $\sim5$ns switching time
+ $\sim50$uW switching power

[Halupka, et al. ISSCC’10]
[Pronin. EETime’13]
[Henderson. InfoTracks’11]

[ITRS’13]
Positioning of New Memories

- SRAM
- DRAM
- FLASH
- HDD
- STT
- PCM
- RRAM

Higher Speed
Lower Cost
Higher Endurance
Capacity
Phase Change Memory

Phase change material (chalcogenide glass) exists in two states:

- **Amorphous**: Low optical reflectivity, high electrical resistivity
- **Crystalline**: High optical reflectivity, low electrical resistivity

Amorphous chalcogenide materials
Phase Change Memory

- **Write**: change phase via current injection
  - **SET**: sustained current to heat cell above $T_{\text{cryst}}$
  - **RESET**: cell heated above $T_{\text{melt}}$ and quenched

- **Read**: detect phase via material resistance
  - amorphous/crystalline

[slide ref: Mutlu]
Some emerging resistive memory technologies seem more scalable than DRAM (and they are non-volatile)
- Example: Phase Change Memory
- Expected to scale to 9nm (2022 [ITRS])
- Expected to be denser than DRAM: can store multiple bits/cell

But, emerging technologies have shortcomings as well
- Can they be enabled to replace/augment/surpass DRAM?
Hybrid Memory Systems

Hardware/software manage data allocation and movement to achieve the best of multiple technologies
PCM Latency

- Latency comparable to, but slower than DRAM
  - Read Latency: 50ns (4x DRAM, 10-3x NAND Flash)
  - Write Latency: 150ns (12x DRAM)
  - Write Bandwidth: 5-10 MB/s (0.1x DRAM, 1x NAND Flash)

[slide ref: Mutlu]