Emerging Memories: Are They Energy Efficient Enough?

H.-S. Philip Wong
Stanford University
Memory – Key Enabler for New Applications

256GB
8GB
64GB
16GB
Source: Intel
Energy, Environment, Sustainability

- **CPU consumes**
  - 27 – 57% of server power (2005 – 2007)
  - ~50% of CPU power consumed in devices
  - Exceeds total power from solar (US, 2007)

Compute and Memory Must be Balanced

- Amdahl’s system balance rules: “the MB/MIPS ratio ($\alpha$) in a balanced system is 1.”
  - Today’s system: $\alpha \approx 4$ (and rising) [1]

- Non-memory transistors increase only 3X in 10 years [2]
  - That’s all you can afford (power)

- Memory integration capacity will outpace logic > 10X

Demands from Above

- **Power breakdown of Intel IA-32, 48-Core Processor in 45 nm CMOS** – memory is a big piece

  **Full Power Breakdown**
  - Total: 125.3W
  - Cores: 69%
  - MC & DDR3-800: 19%
  - Routers & 2D-mesh: 10%
  - Clocking: 1.9W
  - Routers: 12.1W
  - Cores: 87.7W
  - MCs: 23.6W
  - Cores-1GHz, Mesh-2GHz, 1.14V, 50°C

  **Low Power Breakdown**
  - Total: 24.7W
  - Cores: 21%
  - MC & DDR3-800: 69%
  - Routers & 2D-mesh: 5%
  - Clocking: 1.2W
  - Routers: 1.2W
  - Cores: 5.1W
  - MCs: 17.2W
  - Cores-125MHz, Mesh-250MHz, 0.7V, 50°C

Nanoelectronics – Don’t Forget the Memory

- Computation
- Memory
- Communication
DRAM/HDD Bandwidth Gap

Disks becoming cheaper, but no faster

Semiconductor memories becoming faster and cheaper

Result: Speed gap is widening

J. Handy, HOTCHIPS 2010.
Opportunities For Memories

- White space has opened up in memory hierarchy
  - Speed/capacity gap
- Lots of GB needed for various applications
  - Need high density

Question is:

Will the memories be energy efficient enough?
Emerging Memory Candidates

- Phase change memory (PCM, PRAM, PCRAM)
- Spin torque transfer RAM (STTRAM)
- Resistive switching metal oxide RAM (RRAM)
- Conductive bridge RAM (CBRAM)
- ...and many others that fall into the category
  - I cannot make this into a good logic switch, but it has some hysteresis, so let’s call that a memory
Principles of Phase Change Memory (PCM)

Various phase-change materials:
- Ge$_2$Sb$_2$Te$_5$ (GST)
- AIST (AgInSbTe)
- GeSb, Sb$_2$Te and etc..

Amorphous
- High Resistance
- Melt-quenched

Crystalline
- Low Resistance

Annealing

- Resistance change memory: ~1000X difference in resistivity

PCM Programming – Joule heating

- Amorphization (RESET): Melt and quench ($T > T_{\text{Melt}}$)
- Crystallization (SET): Anneal ($T > T_{\text{crys}}$)
Phase Change Memory

PCM Status

Numonxy 1 Gb PCM array
45 nm generation (ISSCC, 2010)

512-Mbit PRAM die packaged with a Samsung 128Mbit UtRAM die in a multi-chip package in a Samsung mobile phone (Dec, 2010)
Scaling of Programming Current

Current density = 20 – 40 MA/cm²
(can be 10X lower with materials engineering)

STTRAM

Write current density = 2MA/cm²
Range: 3 – 5 MA/cm²

D. Halupka et al., “Negative-Resistance Read and Write Schemes for STT-MRAM in 0.13μm CMOS,” ISSCC, p. 256 (2010). (U. Toronto/Fujitsu)
RRAM is Hot!

Metal Oxide M-I-M Memory (RRAM)

- **Motivation:**
  - Low programming voltage (< 3V)
  - Material set compatible with conventional semiconductor processing (e.g. Ni, Hf, Al…)
  - Low temperature processing (BEOL-compatible)

- **Key issues:**
  - Physics of resistive switching
  - Device scaling properties
  - Device uniformity
  - A killer application
Basic I-V Characteristics

- **HRS (High Resistance State)**
- **LRS (Low Resistance State)**
- **RESET**
- **SET**
- **Compliance**

**Unipolar Mode**

**Bipolar Mode**

- **Voltage**
- **Current**
Prevailing Theory for Resistive Switching

Conductive Bridge Memory

1. Redox reaction
2. Ion migration (cation toward cathode)

\[ M \rightarrow M^+ + e^- \]
\[ M \leftarrow M^+ + e^- \]

Cu/Cu$_2$S

Ag-Ge-Se


CBRAM Filament Formation

Nanoconductive Bridge Typical I-V Characteristics

- Bipolar, asymmetric programming/erase
- Low programming/erase voltage
- Forming required

Ag/Ag₂S/Pt

An oxidizable electrode (Ag, Cu etc.) is needed to fulfill the forming and dissolving filamentary conduction paths.


Switching at (±0.6 V, 1 MHz)

Top: As-formed switched-on state
Middle: Switched-off state
Bottom: Switched-on state after the initial switching-off process (bottom).
So, Let’s Talk About Energy

- **Reading a bit**
  - Determined by the resistance of the lowest resistance state (LRS)
  - Typical: read voltage=200mV, read time = 10 ns
  - LRS: RRAM = 10kΩ – 100MΩ, CBRAM = kΩ – GΩ, STTRAM = few kΩ, PCM = kΩ – MΩ
  - Read energy ~ 0.4pJ – 4aJ

- **Writing a bit**
  - Depends on physics and materials
Write Energy – PCM and STTRAM

- **Write voltage roughly constant**
  - PCM ~ 1 – 2V
  - STTRAM ~ 0.5 – 1V

- **Write current density**
  - PCM: 20 – 40 MA/cm²
  - STTRAM: 3 – 5 MA/cm²

- **Write speed**
  - PCM: 10 – 30 ns
  - STTRAM: 1 – 10 ns

- **Write energy per bit (currently, scales with device size)**
  - PCM: 3 pJ/bit (0.3 pJ/bit experimentally demonstrated)
  - STTRAM: ~0.3 pJ/bit
Write Energy – RRAM and CBRAM

- **Write voltage roughly constant**
  - RRAM ~ 1.5 – 2V (V\(_{\text{RESET}}\) > V\(_{\text{SET}}\))
  - CBRAM ~ 1.5 – 3V (V\(_{\text{SET}}\) > V\(_{\text{RESET}}\))

- **Write current**
  - Almost a free variable, depends on the resistance value of the low resistance state (LRS)
  - LRS can be “freely adjusted” by controlling the programming current for the LRS
  - LRS: RRAM = 10kΩ – 100MΩ, CBRAM = kΩ – GΩ
  - RRAM ~ µA (or even 10 nA), CBRAM ~ 100 µA
  - Tradeoff between lower current (energy) and read current (time) and retention

- **Write speed**
  - Write time decreases exponentially with write voltage
  - Energy = (V\(^2\)/R\(_{\text{LRS}}\)) × t
  - So, use large voltage amplitude to decrease time to reduce energy

- **Write energy per bit (filamentary, does NOT scale with device size)**
  - RRAM: 60 fJ/bit – 3 pJ/bit
  - CBRAM: 1 – 5 pJ/bit
Cross-Point Memory Array with Selection Device

Bitline (B/L)

Wires

Memory element

Selection device

Wordline (W/L)

Top electrode

Selection device plus memory element

Bottom electrode
Let’s talk about wires
Bitline/Wordline Resistance

J. Liang, Y. Wu, H.-S. P. Wong, ACM JETC (submitted)
Energy Consumed in Wires

Charging wires = $CV^2$

Static energy = $I \times V \times t$

$C_{tot} = \frac{l}{2} \times C_{20} + l \times 2C_{21}$ (F)

Charging Energy ≈ Static Energy

Energy consumed in wires ~ 100X smaller than devices
(not including readout circuits)

J. Liang (2011)
Energy Per Switch – Logic vs Memory

Write energy per bit
PCM
STTRAM
CBRAM
RRAM

Wire energy (1kb × 1kb)

Read energy per bit

“Smallest FET”: $4.2 \times 10^{-20}$ J

“Smallest Relay”: $10^{-21} – 2 \times 10^{-20}$ J

Next up: figure out how many logic transitions per memory write
Students and Post-Docs

NEM Relay
Daesung Lee

NEM Relay
J Provine

Nanotube device
Lucky Liyanagi

Nanotube device
Henry Chen

Nanotube / graphene interconnect
Helen Chen

Chip in Cell
Kokab Parizi

NEM relay
Xiaoying Shen

NEM relay
Soogine Chong

Self-assembly
He Yi

Self-assembly
Marissa Caldwell

Carbon Nanotube
Jason Parker

Phase change memory
Ethan Ahn

Memory
Jiale Liang

Phase Change memory
Rakesh Jeyasingh

Metal oxide memory
Shimeng Yu

CMOS device/circuit
Jieying (Ivy) Luo

Metal oxide memory
Yi Wu

III-V, Ge, and CMOS
Crystal Kenney

III-V, Ge, and CMOS
Jenny Hu
Sponsors and Collaborators

Stanford INMP
(Toshiba, Intel, TI, IBM, AMD, TEL, AMAT, COSAR, Synopsis)

Stanford NMTRI
(Toshiba, Intel, COSAR (Samsung, Hynix), Micron, SanDisk, Intermolecular)
Non-Volatile Memory Technology Research Initiative (NMTRI) at Stanford University
Technical Collaborators