mMPU: Memristor Memory Processing Unit

Shahar Kvatsinsky
Viterbi Faculty of Electrical Engineering
Technion – Israel Institute of Technology

ICRI-CI Retreat May 2017
The External Memory Wall Problem
von Neumann (Architecture) Bottleneck

A bottleneck of both throughput and power!
### And an Energy Bottleneck

<table>
<thead>
<tr>
<th>Operation</th>
<th>Energy/Op (45 nm)</th>
<th>Cost (vs. Add)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add operation</td>
<td>0.18 pJ</td>
<td>1X</td>
</tr>
<tr>
<td>Load from on-chip SRAM</td>
<td>11 pJ</td>
<td>61X</td>
</tr>
<tr>
<td><strong>Send to off-chip DRAM</strong></td>
<td><strong>640 pJ</strong></td>
<td><strong>3,556X</strong></td>
</tr>
</tbody>
</table>

Attempts to Reduce Data Movement

Add Cache Memory

- Reduce the bottleneck (the cost of memory access)
- Requires locality
- Limited capacity
- Static energy
Processing “In-Memory” (PIM) 
Reducing Data Movement

Input → CPU → Output

Memory
Processing “In-Memory” (PIM)
Reducing Data Movement

Prior Art

90’s

Configuration
PIM machine

Active Pages

SA connected to SIMD pipeline

Recent

Automata Memory

90’s

Recent

Data transfer is still required to/from DRAM and PUs

Real Computing within the Memory
Beyond von Neumann Architecture

Input Device → CPU → Output Device

- Control Unit
- Arithmetic/Logic Unit

Memory Processing Unit (MPU)
mMPU: Solving the von Neumann Bottleneck

Moving from DRAM to memristive memory

mMPU: performing computation *USING* the memristive memory cells
Agenda

• The need for non-von Neumann architectures
• Memristive technologies
  • Memristive MPU (mMPU) architecture
  • mMPU potential
• Summary
Memristors
Emerging Nonvolatile Memory Technologies

Resistive RAM (RRAM)

Phase Change Memory (PCM)

STT MRAM
Memristor
The Missing Fourth Element?

Memristor – Memory Resistor
Resistor with Varying Resistance

Decrease resistance

Decrease resistance

Current
Voltage

Current
Agenda

• The need for non-von Neumann architectures
• Memristive technologies
• **Memristive MPU (mMPU) architecture**
• mMPU potential
• Summary
Logic within Memory

Logic Families

Akers array

MAGIC

IMPLY

MAGIC – Memristor Aided LoGIC

Example of MAGIC NOR

Initialize OUT to $R_{ON}$

$R_{ON} =$ Logic ‘1’

$R_{OFF} =$ Logic ‘0’

<table>
<thead>
<tr>
<th>$IN_1$</th>
<th>$IN_2$</th>
<th>$NOR$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Increase resistance

$R_{OFF} >> R_{ON}$

Real MAGIC

MAGIC NOR in a Crossbar
MAGIC NOR in a Crossbar
MAGIC NOR in a Memristive Memory

Hierarchy of Logical Functions

- Matrix multiplication
- Convolution
- MUL
- POW
- SQRT
- DIV
- ADD
- NOR
- AND
- SUB
- NOT
- XOR
- OR
- COPY
- NAND

Complete logic family

MAGIC - NOR
Logic Execution within Memristive MPU

\[ f: (A, B) \rightarrow C \]
Parallel Vector Operation within Memristive MPU

$$f^n : R^n \times R^n \rightarrow R^n$$

Latency of the vector operation is independent of the length of the vector
mMPU μArchitecture

mMPU μArchitecture

mMPU μArchitecture

mMPU μArchitecture

mMPU Systems

Accelerator or Main Memory?

CPU

Accelerators

Clock, Address, Data, and Controls

mMPU

TMS320C66x
KeyStone™
Multicore DSP

mMPU Systems

Memristive memory with processing capabilities

DRAM

DIMM

?
Issues Involved in mMPU Architecture

Memory Design

Periphery Design

mMPU Controller Design and Optimization

mMPU Architecture

CPU

mMPU Controller

Software

mMPU

Applications

Programming Model
Agenda

• The need for non-von Neumann architectures
• Memristive technologies
• Memristive MPU (mMPU) architecture
• mMPU potential
• Summary
Convolution with mMPU

Vs. Intel i5 Skylake

Vs. PMEM
(Clemons et al., MICRO 2016)
Agenda

• The need for non-von Neumann architectures
• Memristive technologies
• Memristive MPU (mMPU) architecture
• mMPU potential

• **Summary**
mMPU: Memory that Computes

Open Issues (partial list)

• DRAM/memristor system or memristor-only? Memory and/or accelerator?

• Controller optimization and automation

• Programming model

• Demonstration and evaluation

Work in Progress
mMPU – Huge Potential

- Memristors enable non-von Neumann machines to overcome the memory wall
- mMPU – real processing in memory
- Orders of magnitude better performance and energy