Interface Engineering of PCM for Improved Synaptic Performance in Neuromorphic Systems

M. Suri\textsuperscript{1}, O. Bichler\textsuperscript{2}, Q. Hubert\textsuperscript{1}, L. Perniola\textsuperscript{1}, V. Sousa\textsuperscript{1}, C. Jahan\textsuperscript{1}, D. Vuillaume\textsuperscript{3}, C. Gamrat\textsuperscript{2}, B. De Salvo\textsuperscript{1}

\textsuperscript{1}CEA-LETI, \textsuperscript{2}CEA-LIST, \textsuperscript{3}CNRS-IEMN (France)
What...?

Taking inspiration form computational neuroscience and learning mechanisms inside the brain

Why?

- Massive Parallelism
- Scalability (< 2L : the brain)
- Low Power (~ 20W: the brain)
- Immunity to Variability

Information Storage and Processing are not entirely different tasks

“Memory is Intelligent”
How? Neuron + Synapse + Learning Rule

Neuron
Source of Spikes or Action Potentials (electrical pulses)
Integrates all the incoming spikes and fires when a threshold is reached

Synapse
Communication channel or the medium between neurons.
Plasticity – can either Potentiate (LTP) or Depress (LTD)

Learning Rule
The protocol to change the synaptic weights (STDP)
Source of Inspiration…?

Region of the brain responsible for vision
Cerebral cortex, hippocampus, hypothalamus…

Main Issues?

(Very Large Numbers)
Neurons: $1.1 \times 10^{10}$
Synapses: $10^{14} – 10^{15}$

Need a Device which can change and remember simultaneously!! High transistor count to implement synapse circuit capable of learning !!

*Resistive memories for synapses…*
State-of-the-art
Different Strategies of Implementing STDP on PCM devices

D. Kuzum et. al. Nano Letters, 2011

The ‘2-PCM Synapse’

Advantages
1. Simplified Pulse Schemes
2. Very Low-Power
3. Based on Crystallization

Spiking Neural Network

M. Suri et. al. IEDM, 2011
In this Work

Improved System Efficiency compared to GST synapses

• Low Power at device Level
• Low Power at System Level

• Large Scale Demonstration (2 million synapses)
• Real World Learning Application (Pattern Extraction)

CMOS Neuron + PCM Synapse
Our Approach

Device Properties

Neural Network and Learning rule

Synapse parameters

Simplified STDP rule

Application

Pattern Extraction

Manan Suri, LETI Memory Workshop – 21st June, 2012
Resistance – Current (LTP/LTD)

Cell Resistance $R$ [ohms]

Current Density ($\times 10^8$ A/cm$^2$)

Potentiated State
Partially Crystallized
Strong Reset State

Manan Suri, LETI Memory Workshop – 21st June, 2012
Resistance – Current (LTP/LTD)

**Cell Resistance R [ohms]**

**Current Density (x 10^8 A/cm^2)**

- **LTP** (Long-Term Potentiation)
- **LTD** (Long-Term Depression)

**Potentiated State**
**Partially Crystallized**
**Strong Reset State**

**Set State**

**Crystalline**

**TOP ELECTRODE**
**Amorphous Region**
**TUNGSTEN PLUG**
**SiO2**
**SiO2**

**SET to RESET transition = Synaptic Depression** (LTD)
**RESET to SET transition = Synaptic Potentiation** (LTP)

Manan Suri, LETI Memory Workshop – 21st June, 2012
Forming Step – to break the HfO$_2$ Layer

$V_{bd} = 2.2V$ for 2nm thick
Improvement compared to GST devices

Reduction in Set and Reset Current (> 60%)

- Reduction in the effective contact area between the plug and the GST layer
- Less heat-loss (better thermal insulation)

Effective contact area = area of the plug

Small conducting Filament(s) decrease the effective contact area

Manan Suri, LETI Memory Workshop – 21st June, 2012
Programming Current reduction due to Interface layers (HfO$_2$ and TiO$_2$)

Less heat loss

Q. Hubert et al. ESSDERC, 2011
Potentiation (LTP) – GST + HfO$_2$

Gradual Crystallization

Conductance (Crystallization) increases with increase in pulse number and width

Manan Suri, LETI Memory Workshop – 21st June, 2012
Improvement compared to GST devices

Increased number of intermediate resistance points in the LTP curve (more than double of GST)
Possible impact of interface layer on crystallization

Capping layer can affect the Growth and the Nucleation rate

Based on the type of capping layer and the film thickness, the activation energy for crystal growth can either increase or decrease

R. Pandian, et al., JAP 100, 123511 \( \square 2006 \square \)
Our Approach

Device Properties

Neural Network and Learning rule

Synapse parameters

Simplified STDP rule

Application

Pattern Extraction
The ‘2-PCM Synapse’

LTP device: Current is Added, LTD device: Current is Subtracted
We use two devices to avoid the problem of abrupt LTD in PCM

M. Suri et. al. IEDM, 2011

Manan Suri, LETI Memory Workshop – 21st June, 2012
Feed Forward Neural Network: Simulation

2- Layer Spiking Neural Network
Total Neurons: 70 (60 + 10)
2 Synapses per neuron (Inhibiting/Exhibiting)
Total Synapse ~ 2 million => 4 million PCM Devices
LTP characteristics used in simulations

20% Standard Dev. Dispersion added to model robustness to variability of the neural network
Biological and simplified STDP learning rule

Pre-Neuron spikes before Post-Neuron: Potentiation (LTP)

Post-Neuron spikes before Pre-Neuron: Depression (LTD)

% Change in Synaptic Weight is constant
Extended LTD window
Our Approach

- Device Properties
  - Synapse parameters
- Neural Network and Learning rule
  - Simplified STDP rule
- Application
  - Pattern Extraction

Manan Suri, LETI Memory Workshop – 21st June, 2012
Data for our Neural Network

DVS Silicon Retina-Generates AER data
128 x 128 pixels
(Address Event Representation)


Recorded Data – Cars passing on a Freeway

Relative Change in the luminous intensity at each pixel is an event

AER = Pixel Address + Time Stamp + Event

Manan Suri, LETI Memory Workshop – 21st June, 2012
Feed Forward Neural Network: Simulation

2- Layer Spiking Neural Network
Total Neurons: 70 (60 + 10)
2 Synapses per neuron (Inhibiting/Exhibiting)
Total Synapse ~ 2 million => 4 million PCM Devices
Evolution of Neuron Sensitivity during Learning

Recorded Stimuli | Neuron - 1 | Neuron - 2

Neuron-3 | Neuron - 4 | Neuron - 5

Manan Suri, LETI Memory Workshop – 21st June, 2012
Learning Results
(Detection of Cars in individual lanes)

Total Learning duration ~ 10 minutes
Avg. Lane Detection Rate > 90 %

Final Sensitivity Map for 10 Neurons
Performance and comparison to GST devices

<table>
<thead>
<tr>
<th>Pulse Type</th>
<th>/device</th>
<th>/device/sec</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GST Devices</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read</td>
<td>1265</td>
<td>1.9</td>
<td>$4.97 \times 10^9$</td>
</tr>
<tr>
<td>Set</td>
<td>106</td>
<td>0.16</td>
<td>$4.16 \times 10^8$</td>
</tr>
<tr>
<td>Reset</td>
<td>4.2</td>
<td>0.0062</td>
<td>$1.65 \times 10^7$</td>
</tr>
<tr>
<td><strong>Total System Learning Power = 112µW</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GST + HfO₂ Devices</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read</td>
<td>1265</td>
<td>1.9</td>
<td>$4.97 \times 10^9$</td>
</tr>
<tr>
<td>Set</td>
<td>144</td>
<td>0.21</td>
<td>$5.64 \times 10^8$</td>
</tr>
<tr>
<td>Reset</td>
<td>2.6</td>
<td>0.0038</td>
<td>$1.00 \times 10^7$</td>
</tr>
<tr>
<td><strong>Total System Learning Power = 60µW</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Performance and comparison to GST devices

<table>
<thead>
<tr>
<th>Pulse Type</th>
<th>/device</th>
<th>/device/sec</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GST Devices</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read</td>
<td>1265</td>
<td>1.9</td>
<td>$4.97 \times 10^9$</td>
</tr>
<tr>
<td>Set</td>
<td>106</td>
<td>0.16</td>
<td>$4.16 \times 10^8$</td>
</tr>
<tr>
<td>Reset</td>
<td>4.2</td>
<td>0.0062</td>
<td>$1.65 \times 10^7$</td>
</tr>
<tr>
<td><strong>Total System Learning Power</strong></td>
<td></td>
<td></td>
<td><strong>112μW</strong></td>
</tr>
</tbody>
</table>

| **GST + HfO2 Devices** | | | |
| Read       | 1265    | 1.9         | $4.97 \times 10^9$ |
| Set        | 144     | 0.21        | $5.64 \times 10^8$ |
| Reset      | 2.6     | 0.0038      | $1.00 \times 10^7$ |
| **Total System Learning Power** | | | **60μW** |
Conclusions

- We propose the ‘2-PCM Synapse’, a unique programming methodology and a novel feed forward Neural Network based on PCM synapses.

- We show a real world application using our neuromorphic system based on PCM synapses – complex visual pattern extraction

- Addition of HfO$_2$ interface layer to GST devices increases the system power/energy efficiency by-

  a) Decreasing the Iset and Ireset
  b) Increasing the number of intermediate resistance states in the LTP curve

- Total synaptic programming power is as low as 60µW with interface layer PCM synapses.
Questions??

Thank you for your Attention!!

(manansuri@cea.fr)