Complementary Resistive Switch based Neuromorphic Associative Capacitive Network

L. Nielen \(^1\), A. Siemon \(^1\), S. Tappertzhofen \(^1\), R. Waser \(^1\), S. Menzel \(^2\), E. Linn \(^1\)

\(^1\) Institut für Werkstoffe der Elektrotechnik II, RWTH Aachen University, Germany
\(^2\) JARA – Fundamentals of Future Information Technology
\(^3\) Department of Engineering, University of Cambridge, United Kingdom
\(^4\) Peter Grünberg Institut 7, Forschungszentrum Jülich GmbH, Jülich, Germany

21.01.2015
Outline

• Associative Capacitive Networks
  
a Neuronal application for non-volatile memories

• Complementary Resistive Switches (CRS)
  
a solution of the sneak path problem in passive ReRAM arrays

• Capacitive Readout of CRS cells

• CRS-based Associative Capacitive Network (ACN)

• Experimental ACN circuitry

• Simulative evaluation

• Summary
Associative Capacitive Network - Motivation

- Information stored via capacitance value
- Simultaneous activation function on all lines
- Summarized output

→ Content Addressable Memory

Problem with conventional ACNs:
Capacitances charge is volatile

CRS based solution:
non-volatile memory

Applications area:
- Pattern recognition
- Fast routing (reprogrammable lookup-tables)

Artificial neuron model

Content Addressable Memories (CAM)

- Rewritable memory with lookup-table functionality
- Pattern matching in a single cycle

Classical SRAM based CAM cell

- Conventional CAMs are implemented by SRAM core cells
- drawbacks:
  - static currents $\rightarrow$ high energy consumption
  - large area demand

Complementary Resistive Switches (CRS)

Overcoming the Sneak Path Obstacle in passive resistive arrays

- Two anti-serially connected elements
- Overall high cell resistance
- No pattern dependency
- Low static power losses
- No sneak paths

Capacitive Read-out

CRS cells offer a capacitive voltage divider property
- Both elements A and B are equal in terms of resistive switching
- But: capacitances differ (e.g. different areas)
- Capacitive read-out of the stored state

S. Tappertzhofen, E. Linn et al., Nanotechnology, 22, p. 395203 (2011)
CRS-based Associative Capacitive Network

Single ACN cell

\[ N \times 2N-ACN \text{ array} \]

ACN – 2-bit Example

Stored pattern: 0 1 and negated pattern 1 0
HD=2: maximum output \((x_1 = '1', x_2 = '0')\)
HD=1: equal output for \(x_1 = '0', x_2 = '0'\) and \(x_1 = '1', x_2 = '1'\)
HD=0: minimum output \((x_1 = '0', x_2 = '1')\)

Output voltage reflects Hamming Distance (HD) → Pattern matching

O. Kavehei, E. Linn et al., Nanoscale, 5, 5119 (2013)
Fabrication of ACN Cells

Fabrication of a μ-structure array

- Silicon wafer
- Platinum, Titanium and TiO₂ sputter deposition
- UV-lithography

Cell A: 24×10 µm²
Cell B: 20×20 µm²

Associative Capacitive Network

- $N \times 2M$ array of CRS devices
- green: stored template ($T$)
- yellow: stored negates of $T$ ($\overline{T}$)
- input: vector $X$ and $\overline{X}$

Die mounting in a 28 pin carrier
Contacting via wedge-wedge bonding with gold wires
Evaluation with an experimental circuitry
Experimental Setup for ACN Read-Out

Applying the search pattern

Generation of the search pattern by mechanical switches

all bit lines

L. Nielen; S. Tappertzhofen et al., IEEE SNW, p. 61 (2014)
Experimental Setup for ACN Read-Out

Impedance conversion on each word line
Experimental Setup for ACN Read-Out

Voltage-to-time conversion

→ HD detection via switching event
Experimental Results

Stored patterns:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

(a)

- Time $t$ [ns]
- Search pattern '00000000'
Experimental Results

Stored patterns:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

(b) Time $t$ [ns]

- Pulse
- $comp_0$
- $comp_1$
- $comp_2$
- $comp_3$
- $comp_4$
- $comp_5$
- $comp_6$
- $comp_7$

Search pattern: ‘11100000’
Memristive ECM model

\[ I = I_{\text{ion}}(V, w) + I_{\text{Tu}}(V, w) \]

\[ \dot{w} = C_1 \cdot I_{\text{ion}} \]

VerilogA core cell

(a)

(b)

$F = 40 \text{ nm}$

$C_{\text{seg}} = 2.76 \text{ aF}$  SiO$_2$ as interline material ($\varepsilon_r = 3.9$)

$R_{\text{seg}} = 0.86 \text{ } \Omega$

$D_A / D_B = 21 \text{ nm} / 14 \text{ nm} = 1.5 \ (C_A/C_B = 1/1.5)$

$D_A / D_B = 24 \text{ nm} / 12 \text{ nm} = 2 \ (C_A/C_B = 1/2)$

$D_A / D_B = 28 \text{ nm} / 7 \text{ nm} = 4 \ (C_A/C_B = 1/4)$

$D_A / D_B = 30 \text{ nm} / 5 \text{ nm} = 6 \ (C_A/C_B = 1/6)$

L. Nielen, A. Siemon et al., accepted Jetcas, (2015)
ACN Array Simulations

- Memristive ECM model implementation with VerilogA (SPICE)
- Array simulations performed by Cadence Spectre

Features
- Complete write and search operation feasible
- Implementation of coupling capacitances
- Different device capacitance ratios

L. Nielen, A. Siemon et al., accepted Jetcas, (2015)
Size Effects & Power Consumption

Minimum voltage margin $\Delta V$
- Increasing array size (i.e. pattern length)
  - $\rightarrow$ voltage interval corresponding to HD decreases

Search energy demand
Only caused by charging currents in:
- cells
- parasitic capacitances
  - $\rightarrow$ Power consumption scales with array size

<table>
<thead>
<tr>
<th>ACN ($C_A/C_B = 1.5$)</th>
<th>SRAM-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>32×64</td>
<td>256…1024×144</td>
</tr>
<tr>
<td>8 F² 40nm</td>
<td>120–1500 F² 32–65 nm</td>
</tr>
<tr>
<td>4.69 aJ/bit/search</td>
<td>&gt; 0.1 fJ/bit/search</td>
</tr>
</tbody>
</table>


L. Nielen, A. Siemon et al., accepted *Jetcas*, (2015)
Summary

- **Neuromorphic** application for Memristive Random Access Memories was demonstrated
- An **Associative Capacitive Network** was fabricated
- Development of an **Experimental Setup** for ACN Evaluation
- **Proof-of-Concept**: The ACN shows the predicted behavior
- Study of arrays promises **Low Power Consumption**
- Fully parallel search within the range of **Nanoseconds** was demonstrated using a simple measurement setup

Advantages of CRS based ACN concept:
- Hamming Distance detection (similarity)
- Fully passive 2-CRS cell implementation – small area demand
- Non-volatile → No Refresh → Low Power Consumption
- No reprogramming → fast read access
- No requirement of constant voltage supply
THANK YOU
FOR YOUR ATTENTION
Content Addressable Memories (CAM)

Content Addressable Memories (CAM)

Content Addressable Memories (CAM)

TABLE I
PERFORMANCE COMPARISON OF PRIOR WORKS

<table>
<thead>
<tr>
<th>Technology /Supply</th>
<th>This work</th>
<th>[25]</th>
<th>[10]</th>
<th>[3]</th>
<th>[19]</th>
<th>[20]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>65 nm /1.2V</td>
<td>130nm /1V</td>
<td>32nm/1V</td>
<td>65 nm /1V</td>
<td>65 nm /1V</td>
<td>130nm /1V</td>
</tr>
<tr>
<td>Search delay (ns)</td>
<td>1.07</td>
<td>0.9</td>
<td>0.145</td>
<td>1.92</td>
<td>0.6 (72 bits)</td>
<td>2.2 (240 bits)</td>
</tr>
<tr>
<td>FOM (fJ/bit/search)</td>
<td>0.77</td>
<td>1.827</td>
<td>1.07</td>
<td>1.98</td>
<td>0.99</td>
<td>1.3</td>
</tr>
<tr>
<td>Normalized FOM*</td>
<td>1</td>
<td>1.09</td>
<td>2.48</td>
<td>2.37</td>
<td>1.2</td>
<td>0.65</td>
</tr>
<tr>
<td>Frequency</td>
<td>500 MHz</td>
<td>250 MHz</td>
<td>N.A</td>
<td>250 MHz</td>
<td>450</td>
<td>N.A</td>
</tr>
<tr>
<td>Chip area(mm²)</td>
<td>0.125</td>
<td>1.4</td>
<td>N.A</td>
<td>99</td>
<td>0.078</td>
<td>N.A</td>
</tr>
<tr>
<td>Capacity</td>
<td>128×128</td>
<td>128×32</td>
<td>128×128</td>
<td>18 Mb</td>
<td>64×72</td>
<td>256×144</td>
</tr>
</tbody>
</table>

* Normalized FOM = FOM × (65 nm/technology node) × (1.2 V/VDD).

### Content Addressable Memories (CAM)

#### TABLE I
**Features Summary and Comparisons**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>configuration</td>
<td>1024x144</td>
<td>256x128</td>
<td>512x144</td>
<td>256x128</td>
<td>1024x144</td>
<td>256x144</td>
</tr>
<tr>
<td>Technology</td>
<td>100 nm</td>
<td>0.18 μm</td>
<td>0.13 μm</td>
<td>0.18 μm</td>
<td>0.18 μm</td>
<td>65 nm</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>2.8x4.2 (chip)</td>
<td>1.21x0.56 (core)</td>
<td>1.5x1.7 (core)</td>
<td>0.84x0.92 (core)</td>
<td>3.67x0.98 (core)</td>
<td>1.01x0.43 (core)</td>
</tr>
<tr>
<td>Supply voltage (V)</td>
<td>1.2 V</td>
<td>1.8 V</td>
<td>1.2 V</td>
<td>1.8 V</td>
<td>1.8 V</td>
<td>1.0 V</td>
</tr>
<tr>
<td>Search time (ns)</td>
<td>2.20 ns</td>
<td>2.10 ns</td>
<td>4.80 ns</td>
<td>1.56 ns</td>
<td>100MHz</td>
<td>0.38ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400MHz</td>
</tr>
<tr>
<td>Energy metric (fl/bit/search)</td>
<td>0.700</td>
<td>2.330</td>
<td>0.590</td>
<td>1.420</td>
<td>6.300</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.165</td>
</tr>
<tr>
<td>Normalized Search time T* (ns)</td>
<td>1.716</td>
<td>1.365</td>
<td>2.880</td>
<td>1.014</td>
<td>N.A.</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N.A.</td>
</tr>
<tr>
<td>Normalized Energy metric E* (fl/bit/search)</td>
<td>0.316</td>
<td>0.260</td>
<td>0.205</td>
<td>0.158</td>
<td>0.702</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.165</td>
</tr>
</tbody>
</table>

Content Addressable Memories (CAM)

5T-2R NOR-type CAM

6T-2R NOR-type CAM

7T-2R NOR-type CAM

7T-2R NAND-type CAM

VerilogA core cell

\[ F = 40 \text{ nm} \]
\[ C_{\text{seg}} = 2.76 \text{ aF} \quad \text{SiO}_2 \text{ as interline material (r} = 3.9) \]
\[ R_{\text{seg}} = 0.86 \Omega \]
\[ D_A / D_B = 21 \text{ nm} / 14 \text{ nm} = 1.5 (C_A/C_B = 1/1.5) \]
\[ D_A / D_B = 24 \text{ nm} / 12 \text{ nm} = 2 (C_A/C_B = 1/2) \]
\[ D_A / D_B = 28 \text{ nm} / 7 \text{ nm} = 4 (C_A/C_B = 1/4) \]
\[ D_A / D_B = 30 \text{ nm} / 5 \text{ nm} = 6 (C_A/C_B = 1/6) \]
Monte Carlo Simulations

Probability–voltage distribution for seven outputs with HD=0 to 6 that correspond to ML₁ to ML₇. The solid blue line indicates VTH in our test-bench circuit. Outputs with a voltage amplitude above the threshold voltage are treated as ‘hit’ and below that are treated as ‘miss’.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (μ)</th>
<th>Relative 3σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply and input voltages</td>
<td>3 V</td>
<td>10%</td>
</tr>
<tr>
<td>Series resistors on each ML</td>
<td>100 Ω</td>
<td>10%</td>
</tr>
<tr>
<td>Device thickness</td>
<td>20 nm</td>
<td>10%</td>
</tr>
<tr>
<td>Top electrode width</td>
<td>5 μm</td>
<td>10%</td>
</tr>
<tr>
<td>Middle electrode width</td>
<td>10 μm</td>
<td>10%</td>
</tr>
<tr>
<td>Bottom electrode width</td>
<td>15 μm</td>
<td>10%</td>
</tr>
<tr>
<td>Load capacitor (C_ML)</td>
<td>22 pF</td>
<td>10%</td>
</tr>
<tr>
<td>RRAM ON resistance (R_ON)</td>
<td>1 kΩ</td>
<td>20%</td>
</tr>
<tr>
<td>RRAM OFF resistance (R_OFF)</td>
<td>1 MΩ</td>
<td>20%</td>
</tr>
</tbody>
</table>

Probability function (p) of seven top match-lines with minimum HDs versus $V_{ML}$. The detection probability can be interpreted as p. In order to detect HD=0, (ML1), with p=95% chance, $V_{TH}$ has to be around 1.5 V. Under these circumstances a successful detection over a range of outputs is achieved. If HD=0, the output is 1 with 95% probability. For HD=1, the output is 1 with 50% probability, while for HD=2 the probability for an output 1 is 10%. For HD=3, only a probability of 0.5% for observing an output 1 is given. The probability that output is 1 for HD > 3 tends to have negligible small values. Thus, HD < 4 are detectable with high probability.