Memristor Based Event Driven Neuromorphic Nano-CMOS Processor

by

Charanraj Mohan

Submitted to the Department of Electronics and Electromagnetism
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

UNIVERSITY OF SEVILLE

November 2020

© University of Seville 2020. All rights reserved.

Author . . . .

Charanraj Mohan
Department of Electronics and Electromagnetism
November 19, 2020

Certified by . . . .

Bernabé Linares-Barranco
Full Professor of Research
Thesis Supervisor

Certified by . . . .

Teresa Serrano Gotarredona
Tenured Scientist
Thesis Supervisor

Certified by . . . .

José Manuel de la Rosa
Full Professor
Thesis Supervisor

To my parents, sibling and wife.

Acknowledgments

I would like to thank my advisor Prof. Bernabé Linares-Barranco for being a great instructor and a supportive mentor, and for his patience and intuitive guidance throughout the thesis work. It is my pleasure to acknowledge and thank Prof. Teresa Serrano Gotarredona for her great advice and dedicated guidance in the completion of the thesis. My special thanks and gratitude goes to Prof. José Manuel de la Rosa for his invaluable guidance and motivation throughout the thesis. I owe a debt of gratitude for the knowledge, learnings, and skills acquired through timely discussions and feedback I received from them.

I am thankful to Prof. Francesca Campabadal Segura from Institut de Microelectrònica de Barcelona (IMB-CNM), who hosted me for the research stay. Under her guidance and support of her research team members (Marcos Maestro and Mireia Bargalló), I had a very informative and hands-on lab experience in wafer characterization.

I thank all members of the 'Neuromorphic Systems' group at Instituto de Microelectrónica de Sevilla (IMSE-CNM) for productive meetings and discussions. I thank and appreciate my colleagues- Amir, Laurentiu, Luis Camunas, Javad, Ajay, Ion Vornicu, Angela, and all the other colleagues from IMSE who have been supportive. I also like to thank Joaquin, Antonio, Miguel A., and Juan M. Repiso, who were instrumental in timely availing with the laboratory equipment and tools.

I would like to thank my parents- Mohan & Vasanthi who are always my inspiration. I am also thankful to my beloved wife- Abi for being very supportive and encouraging me during tough times.

The thesis work was supported by COGNET (63884), Lloyd's Register Foundation's ICON (G0086), EU H2020 grants- NeuRAM^3 (687299) , MeM-Scales (871371) and HERMES (824164).

Memristor Based Event Driven Neuromorphic Nano-CMOS Processor

by

Charanraj Mohan

Submitted to the Department of Electronics and Electromagnetism
on November 19, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Neuromorphic engineering is an emerging bio-inspired discipline that morphs the biological brain on custom silicon. Although memristors rose as a potential synapse to solve density challenge when monolithically integrated above the silicon structures, scalability remains an important bottleneck. Neuromorphic systems should be made more scalable to realize large networks. To contribute to this, we focus on significant challenges in memristor-based neuromorphic hardware related to– implementation of a low-power inference system that is used for learning or programming and the design of a new current attenuator that is used for efficient crossbar read-outs. The thesis also demonstrates the characterization of different memristors on various test-benches.

Thesis Supervisor: Bernabé Linares-Barranco
Title: Full Professor of Research

Thesis Supervisor: Teresa Serrano Gotarredona
Title: Tenured Scientist

Thesis Supervisor: José Manuel de la Rosa
Title: Full Professor

This doctoral thesis has been examined by a Committee as follows:

. . . .

. . . .

. . . .

Contents

List of Figures 13

List of Tables 25

List of Abbreviations and Acronyms 26

1 Introduction 31

1.1 Memristor based neuromorphic processors 31

1.1.1 Moore's Law and beyond Von Neumann architecture 31

1.1.2 Neuromorphic chips 32

1.1.3 Memristor- a favourable synapse 34

1.1.4 Memristive crossbar and sneak-path currents 37

1.1.5 STDP learning rule 38

1.2 Contributions in Memristor based crossbars 40

2 Charaterization of memristors 43

2.1 Need for memristor — characterization 43

2.2 Characterization of Neurobit memristors using ArC ONE® Platform . 44

2.3 Characterization of 1T1R OxRAM-based memristors using customized PCBs 48

2.3.1 Design of 4×4 and 8×8 1T1R crossbars and its packaging . . . 48

2.3.2 Design of circuits for the test-PCBs 53

2.3.3 Assembly and mounting of test-PCBs and customized boards . 56

2.3.4 Description and working of the experimental set-up 57

2.3.5 Experimental results of characterization of 1T1R based memristors 68
2.4 Characterization of MIM-based memristors using SPA 70
2.4.1 Experimental set-up 72
2.4.2 Characterization results of MIM-based memristors 75

3 Bulk-based three-stage DC offset Calibration Scheme for Memristive Crossbar 79
3.1 Need for offset calibration 79
3.2 Three-stage bulk-based DC offset calibration approach 80
3.3 Design of 1T1R crossbar with three-stage DC offset calibration scheme 81
3.3.1 Design of two-stage PMOS-based differential pair opamp . . . 83
3.3.2 Design of pulse-shaping digital blocks across wordlines of memristive crossbar 91
3.3.3 Design of body-input three-stage offset calibration scheme . . 93
3.3.4 Design of 4 × 4 1T1R crossbar 100
3.3.5 Design of I-pots 101
3.3.6 Design of D-flip flop based shift-register 103
3.4 Preparing an experimental set-up for calibration scheme 108
3.4.1 Packaging of chip 109
3.4.2 Design of circuits for the test-PCB 109
3.4.3 Assembly and mounting of test-PCB and auxiliary boards . . 113
3.4.4 FPGA SPARTAN®-6 driver board 115
3.5 Description and working of the experimental set-up of the calibration scheme 115
3.6 System-level simulation results for pattern recognition 117
3.6.1 Using Supervised Single-Shot Programming (SSSP) 120
3.6.2 Using STDP learning rule 123
3.7 Experimental results of memristive processor facilitated with bulk-based calibration scheme across wordlines 128

3.7.1 Preliminary test results 128
3.7.2 Three-stage bulk-based calibration scheme results of full input control-word 130
3.7.3 Characterization results of OxRAMs in 4×4 1T1R crossbar with on-chip DC offset calibration across wordlines 133
3.7.4 Template-matching results implemented on a calibrated 4×4 1T1R crossbar 137
3.7.5 Pattern recognition results using SSSP on calibrated crossbar 145
3.7.6 Pattern recognition results using STDP learning rule on calibrated crossbar 148

4 MCN Attenuator for Efficient Memristive Crossbar Read-Out 153
4.1 Need for a current attenuator 153
4.2 Design of Modified Current Normalizer (MCN attenuator) 153
4.3 Simulation Results 159

5 Conclusion and Future work 161

References 163

Appendix A PCB design details and guidelines 179

Appendix B MADII circuit design details 197

List of Figures

1-1 (a) 1T1R structure, (b) Layout preview of 1T1R in MAD200 PDK, (c) Microscopic view of the monolithically integrated hybrid CMOS and OxRAM [36]. 35

1-2 (a) A 4×4 memristive crossbar, (b) Read current and sneak-path current in a 4×4 memristive crossbar. 37

1-3 A synaptic junction that connects a pre-synaptic and a post-synaptic neuron. 39

2-1 A Neuro-Bit memristor connected to the active word-line and bit-line terminals of ArC ONE® platform. 45

2-2 (a) LRS results of a Neuro-Bit memristor for 50 'write' pulses for different pulse widths, (b) HRS results of a Neuro-Bit memristor for 50 'erase' pulses for different pulse-widths 46

2-3 (a) LRS results of a Neuro-Bit memristor for 50 'write' pulses for amplitudes, (b) HRS results of a Neuro-Bit memristor for 50 'erase' pulses for different amplitudes 47

2-4 Simulation results showing repetitive sequence of 'Erase'-'Read'-'Write'-'Read' operations carried out on 1T1R device after forming it- (a) Voltage biases applied at the terminals of the 1T1R device, (b) Drain currents of the selector MOSFET, (c) Resistance of the OxRAM, (d) A zoom preview of the resistance of the OxRAM after forming. . . . 49

2-5 (a) Schematic view of the 4 × 4 1T1R crossbar, (b) Layout view of the 4 × 4 1T1R crossbar. 50

2-6 (a) Schematic view of the 8 × 8 1T1R crossbar, (b) Layout view of the 8 × 8 1T1R crossbar. 50
2-7 Switching biases for active and default wordlines and bitlines for various operations performed in OxRAM when gate-lines are pulled bitline-wise. 51
2-8 Switching biases for active and default wordlines and bitlines for various operations performed in OxRAM when gate-lines are pulled wordline-wise. 51
2-9 Layout view of MAD200 chip showing layouts of 4 × 4 and 8 × 8 1T1R crossbars duly connected with their pads. 52
2-10 (a) Top view of a PLCC52 package (with its pin numbers) that is packaged with 4 × 4 1T1R crossbar of MAD200 chip, (b) Layout view of 4 × 4 1T1R crossbar duly labeled and numbered for packaging in PLCC52. 54
2-11 (a) Top view of a PGA100 package that is packaged with Outer-ring of MAD200 chip, (b) Bonding diagram of the Outer-ring using PGA100 package. 54
2-12 Functional block diagram of the test-circuit for testing 4 × 4 1T1R crossbar. 55
2-13 Functional block diagram of the test-circuit for testing 8 × 8 1T1R crossbar. 56
2-14 Experimental set-up for characterizing 4 × 4 1T1R crossbar. 58
2-15 Experimental set-up for characterizing 8 × 8 1T1R crossbar. 59
2-16 FSM used to characterize 1T1R crossbars. 60
2-17 Default and active biases (in the form of pulses) applied across rows, columns and gates of the 4 × 4 crossbar for the sequential OxRAM operations- 'Form_Global_Read'. 61
2-18 Default and active biases (in the form of pulses) applied across rows, columns and gates of the 4 × 4 crossbar for the sequential OxRAM operations- 'Erase_Global_Read'. 62

2-19 Default and active biases (in the form of pulses) applied across rows, columns and gates of the 4 × 4 crossbar for the sequential OxRAM operations- 'Write_Global_Read'. 62

2-20 Default and active biases (in the form of pulses) applied across rows, columns and gates of the 4 × 4 crossbar for the sequential OxRAM operations- 'Read'. 63

2-21 Default and active biases (in the form of pulses) applied across rows, columns and gates of the 8 × 8 crossbar for the sequential OxRAM operations- 'Form_Global_Read'. 63

2-22 Default and active biases (in the form of pulses) applied across rows, columns and gates of the 8 × 8 crossbar for the sequential OxRAM operations- 'Erase_Global_Read'. 64

2-23 Default and active biases (in the form of pulses) applied across rows, columns and gates of the 8 × 8 crossbar for the sequential OxRAM operations- 'Write_Global_Read'. 64

2-24 Default and active biases (in the form of pulses) applied across rows, columns and gates of the 8 × 8 crossbar for the sequential OxRAM operations- 'Read'. 65

2-25 Read operation in a single 1T1R device in the 4 × 4 crossbar: (a) Read scheme and biases showing a targeted 1T1R device in a 4 × 4 crossbar, (b) Observation of terminals of the 'Read' scheme in oscilloscope after a 'Write_Global_Read' operation performed on the targeted 1T1R device, (c) Observation of terminals of the read scheme in oscilloscope after an 'Erase_Global_Read' operation performed on the targeted 1T1R device. 65

2-26 Output of ADC Vs differential input of ADC for the whole resistance range thereby, highlighting both LRS and HRS of OxRAM. 66

2-27 HRS and LRS values of OxRAMs of all 64 1T1R devices in 8 × 8 crossbar. 68

2-28 Switching of resistance of OxRAM of a 1T1R device between HRS and LRS for 10^3 cycles. 69
2-29 Cycle-to-cycle variability results of LRS for 10^3 switching cycles. . . . 69
2-30 Cycle-to-cycle variability results of HRS for 10^3 switching cycles. . . . 70
2-31 Structure of a MIM based memristor. 71
2-32 Savannah 200 from Cambridge NanoTech Inc. [109]. 71
2-33 Wafer map showing a preview of a single chip with zoom previews of 5×5 µm^2 and 15×15 µm^2 sized MIM based memristors. 73
2-34 Experimental set-up for characterization of MIM based memristors. . 74
2-35 Experimental results of a 15×15 µm^2 sized MIM based memristor in wafer, W_1: (a) I-V characteristics during a breakdown attempt, (b) I-V characteristics during 19 full-cycles of switching between HRS and LRS, (c) Currents- I_{LRS} and I_{HRS} for a read voltage of V_{read} = -0.1 V during 19 full-cycles of switching between HRS and LRS. 75
2-36 Experimental results of a 15×15 µm^2 sized MIM based memristor in wafer, W_2: (a) I-V characteristics during a breakdown attempt, (b) I-V characteristics during 19 full-cycles of switching between HRS and LRS, (c) Currents- I_{LRS} and I_{HRS} for a read voltage of V_{read} = -0.1 V during 19 full-cycles of switching between HRS and LRS. 76
2-37 Experimental results of a 15×15 µm^2 sized MIM based memristor in wafer, W_3: (a) I-V characteristics during a breakdown attempt, (b) I-V characteristics during 19 full-cycles of switching between HRS and LRS, (c) Currents- I_{LRS} and I_{HRS} for a read voltage of V_{read} = -0.1 V during 19 full-cycles of switching between HRS and LRS. 76
2-38 Experimental results of a 15×15 µm^2 sized MIM based memristor in wafer, W_4: (a) I-V characteristics during a breakdown attempt, (b) I-V characteristics during 19 full-cycles of switching between HRS and LRS, (c) Currents- I_{LRS} and I_{HRS} for a read voltage of V_{read} = -0.1 V during 19 full-cycles of switching between HRS and LRS. 77

3-1 OxRAM currents for low read voltage pulses. 80

3-2 Conceptual diagram of the proposed three-stage calibration scheme in the 4 × 4 1T1R crossbar. 81

3-3 Scheme of a 4×4 1T1R crossbar with DC offset voltage calibration in each wordline. 82

3-4 Possible wells used in bulk CMOS process: (a) n-well process, (b) p-well process, (c) Twin-well process, (d) Triple-well process. 83

3-5 Schematic view of the two-stage differential opamp. 84

3-6 (a) Layout view of the two-stage differential opamp, (b) Parasitics extracted layout view of the two-stage differential opamp. 89

3-7 (a) Technology-process corner variation of DC transfer curve of opamp, (b) Monte Carlo variation of DC transfer curve of the opamp, (c) Monte Carlo distribution of DC offset voltage of opamp, (d) Comparison of nominal and layout-extracted simulated DC transfer curves of opamp. 90

3-8 Schematic view of the pre-synaptic driver across wordline, $w1$. 91

3-9 Layout view of a pre-synaptic driver. 92

3-10 Schematic view of the three-stage calibration scheme. 93

3-11 (a) Layout view of the three-stage calibration scheme with the decoders, (b) Parasitic extracted layout view of the three-stage calibration scheme with the decoders. 94

3-12 Technology-process corner variation results of the three-stage calibration scheme: (a) Technology-process corner variation results of signals - V_{ref} - V_d, $V_{ref} + V_d$, $Top1$ and $Bottom1$ (b) A zoom-preview of technology-process corner variation results showing signals - $Top2$, $Bottom2$ and Out_calib. 95

3-13 Monte Carlo variation results of the three-stage calibration scheme: (a) Monte Carlo variation results of signals - V_{ref} - V_d, $V_{ref} + V_d$, $Top1$ and $Bottom1$ (b) A zoom-preview of monte Carlo variation results showing signals - $Top2$, $Bottom2$ and Out_calib. 96

3-14 (a) Monte Carlo variation of DC offset voltage due to temperature with calibration at 27°C, (b) Sigma of DC offset voltage due to temperature variation with calibration at 27°C. 97

3-15 Comparison of nominal and layout-extracted simulated output, Out_calib of the calibration scheme. 98

3-16 (a) Simulation results during coarse (stage 1) calibration of DC offset voltage across wordline, $w1$ targeting the zero-crossing region, (b) Simulation results during fine (stage 2) calibration of DC offset voltage across wordline, $w1$ targeting the zero-crossing region, (c)Simulation results during finer (stage 3) calibration of DC offset voltage across wordline, $w1$ targeting the zero-crossing region. 99

3-17 (a) Schematic view of the 4 × 4 1T1R crossbar, (b) Layout view of the 4 × 4 1T1R crossbar. 100

3-18 Schematic view of the I-pot. 101

3-19 Simulated output currents of I-pot for 6 possible control bits of decade current splitter for the reference current, i_{ref} = 100 µA. 102

3-20 Scheme of n-bit D-flip flop based shift register. 103

3-21 (a) Schematic view of edge-triggered D-flip flop, (b) Schematic view of the latch. 104

3-22 Simulation results of a 3-bit shift register. 104

3-23 Layout view of I-pot with 14-bit shift-register. 105

3-24 Layout view of calibration scheme with its decoders and 12-bit shift-register. 105

3-25 Layout view of the three-stage calibration scheme implemented along the wordlines of a 4 × 4 1T1R crossbar with highlighted different sub-circuits in the outer-ring. 106

3-26 Layout view of the MAD200 chip highlighting the bulk-based calibration scheme implemented along the wordlines of a 4 × 4 1T1R crossbar with its pads duly labeled. 107

3-27 (a) Top view of the chip Packaged in PGA100 package, (b) Top view of the packaged chip after gently removing the top wrapper stuck above the package, (c) A zoom preview of the top view of the packaged chip, (d) An ultra-zoom preview of the top view of the chip captured using a microscope with the highlighted offset calibration circuits with its pads. 108

3-28 Functional block diagram of the test-circuit for testing the three-stage calibration scheme. 110

3-29 Scheme of the read circuitry implemented for the calibration scheme. 111

3-30 Comparison of ON/OFF pulses of different widths: (a) 5 ns ON/OFF pulse, (b) 10 ns ON/OFF pulse, (c) 15 ns ON/OFF pulse, (d) 20 ns ON/OFF pulse, (e) 25 ns ON/OFF pulse, (f) 30 ns ON/OFF pulse. . 114

3-31 Experimental set-up of the DC offset calibration scheme. 116

3-32 FSM used to calibrate DC offset voltage and perform OxRAM operations. 117

3-33 Patterns used for recognition-task using 4×4 crossbar: (a) Pattern-1, (b) Pattern-2, (c) Pattern-3, (d) Pattern-4. 118

3-34 Model based simulation environment implemented in Simulink environment. 119

3-35 Scheme of integrator and comparator implemented in Simulink environment. 120

3-36 Pre-synaptic pulses, reset, cycle and batch signals. 121

3-37 Integrator output voltage, reset, tposta and simulation time signals. . 121

3-38 Post-synaptic pulses- before and after programming. 122

3-39 Conceptual diagram showing patterns applied on a crossbar that has random weights. 124

3-40 Simulated Pre-synaptic pulses, reset signal and number of cycles showing different regions- A, B, C, D, E and F using STDP learning rule. 124

3-41 Simulated Post-synaptic pulses showing different regions- A, B, C, D, E and F using STDP learning rule. 125

3-42 Crossbar showing binary weights and its evolution- (a) Initial random weight, (b) 1st weight update, (c) 2nd weight update, (d) 3rd weight update, (e) 4th weight update, (f) Final weights. 125
3-43 Simulation results showing STDP weight updates for five different random initial weights. 127
3-44 Preview of output screen when testing shift-register. 129
3-45 Preview of output screen when calibrating DC offset across wordline, w_1. 130
3-46 Comparison of experimental and simulation results during stage 1 calibration of DC offset voltage across wordline, w_1. 131
3-47 Comparison of experimental and simulation results during stage 2 calibration of DC offset voltage across wordline, w_1 targeting the zero-crossing region. 132
3-48 Comparison of experimental and simulation results during stage 3 calibration of DC offset voltage across wordline, w_1 targeting the zero-crossing region. 132
3-49 Active row, column and gate biases applied in the form of pulses- showing a read operation after an erase operation. 133
3-50 Active row, column and gate biases applied in the form of pulses- showing a read operation after a write operation. 134
3-51 LRS and HRS values of OxRAMs of all 16 1T1R devices in the 4×4 crossbar where calibration of DC offset voltage is carried out. 135
3-52 LRS and HRS values for 10 switching cycles of an OxRAM of the 4×4 1T1R crossbar where calibration of DC offset voltage is carried out. . 136
3-53 Switching of resistance of OxRAM of a 1T1R device between HRS and LRS for 400 cycles. 137
3-54 Conceptual block diagram showing patterns fed as read pulses across the wordlines (or rows) of the calibrated crossbar, whose synaptic weights are switched to learned values. 138

3-55 Output of integrators and digital signals for template matching for read voltage of 0.13 V. 138
3-56 Output of integrators and digital signals for template-matching using read voltage of 0.13 V- with a zoom preview showing the integrated voltage. 139
3-57 Output of integrators and digital signals for template-matching using read voltage of 100 mV. 140
3-58 Output of integrators and digital signals for template-matching using read voltage of 70 mV. 141
3-59 Output of integrators and digital signals for template-matching using read voltage of 30 mV. 142
3-60 2.4 V signal from power supply directly observed on oscilloscope with 3 bit 'ERes' noise filter option. 143
3-61 Wordlines (or Rows) and digital signals for template-matching using read voltage of 30 mV. 143
3-62 Bitlines (or columns) and digital signals for template-matching using read voltage of 30 mV. 144
3-63 Gate biases and digital signals for template-matching using read voltage of 30 mV. 145
3-64 Flowchart for implementation of SSSP on calibrated crossbar by targeting- 'col1' to spike earlier for 'pattern-1', 'col2' to spike earlier for 'pattern-2', 'col3' to spike earlier for 'pattern-3' and 'col4' to spike earlier for 'pattern-4'. 146
3-65 Inference results by programming with the single shot variant of STDP targeting pattern1. 147
3-66 Flowchart for implementation of STDP learning rule on calibrated crossbar for pattern recognition. 149
3-67 Output of integrators and comparators with other digital signals during 1st weight-update. 150

3-68 Output of integrators and comparators with other digital signals during 2nd weight-update. 151
3-69 Output of integrators and comparators with other digital signals during 3rd weight-update. 152

4-1 Scheme of a 4×4 1T1R crossbar with pre-synaptic neurons, current attenuators and post-synaptic neurons. 154
4-2 Details of MCN circuit schematics. 155
4-3 Layout view of the MCN circuit. 157
4-4 (a) Minimum to maximum crossbar column inference current Vs Input current to neuron, (b) Comparison of average and standard deviation of output current for different attenuators: MCN circuit, MOS-ladder circuit, and WTA circuit considering process and mismatch variations with 100 Monte Carlo runs. 159
4-5 (a) Temperature variations of the output current of the MCN circuit, MOS-ladder and WTA circuit for different values of output currents, (b) Area and input-referred noise of the MOS circuit, MOS-ladder, and WTA circuit. 160

A-1 Schematic view of the test-circuit for testing 4×4 1T1R crossbar along with its ASIC. 180
A-2 Layout view of the PCB used for testing 4×4 1T1R crossbar. 181
A-3 3-D view of the designed PCB for testing 4×4 1T1R crossbar. 182
A-4 Assembled and mounted PCB for testing 4×4 1T1R crossbar. 183
A-5 Schematic view of the test-circuit for testing 8×8 1T1R crossbar along with its ASIC. 185
A-6 Schematic view of the test-circuit for testing calibration scheme implemented in 4×4 1T1R crossbar. 186
A-7 Schematic view of the test-circuit for testing opamp. 187
A-8 Layout view of the PCB used for testing different circuits in the outer-ring. 188

A-9 3-D view of the designed PCB for testing different circuits in the outer-ring. 189

A-10 Assembled and mounted PCB for testing different circuits in the outer-ring. 190

A-11 (a) Top view of the front-side of the packaged chip dully labelled with pin numbers, (b) Top view of the rear-side (mirrored) of the packaged chip dully labelled with pin numbers or addresses, (c) Top view of the front-side of the PGA ZIF 14 $\times$ 14 socket dully labelled with pin numbers or addresses and marked with the location where the packaged chip sits on it. 195

List of Tables

2.1 Control-bit for performing different operations for OxRAM in the targeted 1T1R device in the crossbar. 57

2.2 Bias conditions used for characterizing a single 1T1R device in the crossbars. 61

2.3 Different dielectrics used in ALD process 72

2.4 Comparison of V_{max}, V_{min}, V_{set} and V_{reset} for wafers- W_1, W_2, W_3 and W_4. 77

3.1 Design parameters of the two-stage PMOS-based differential-pair opamp. 90

3.2 Output of pre-synaptic driver of wordline, $w1$ for different combination of digital inputs. 92

3.3 MOSFET sizing and biasing parameters of the three-stage calibration scheme. 95

3.4 Values of the control signals for setting output current of I-pot to 40 µA.102

3.5 Control-bit for performing different operations for OxRAM in the targeted 1T1R device in the crossbar, whose wordlines are calibrated for DC offset voltage. 110

4.1 MOSFET-sizes and biases of proposed MCN circuit. 156

A.1 Pin addresses for packaged chip and PGA socket for different signals. 191

B.1 Design specifications of the two-stage opamp. 198

List of Abbreviations and Acronyms

1D1M 1 Diode 1 Memristor

1T1R 1 Transistor 1 ReRAM

AC Alternating Current

ADC Analog to Digital Converter

ALD Atomic Layer Deposition

ANN Artificial Neural Network

ARM Advanced RISC Machines

ASIC Application Specific Integrated Circuit

BOM Bill of Materials

CIS Component Information System

CMOS Complementary Metal Oxide Semiconductor

CMP Chemical Mechanical Planarization

CMP Circuit Multi Projects

CPLD Complex Programmable Gate Device

CPU Central Processing Unit

CVD Chemical Vapour Deposition

DC Direct Current

DDR Double Data Rate

DIP Dual In-line Package

DYNAPs Dynamic Neuromorphic Asynchronous Processors

EU European Union

EZ-USB Easy to use Universal Serial Bus

FEOL Front End of Line

FET Field Effect Transistor

FinFET Fin-shaped Field Effect Transistor

FPGA Field Programmable Gate Array

FSM Finite State Machine

GBW Gain Bandwidth Product

GDS Graphical Data System

GPIB General Purpose Interface Bus

GUI Graphical User Interface

HDL Hardware Description Language

HICANN High Input Count Analog Neural Network

HP Hewlett-Packard

HRS High Resistance State

IBM International Business Machines

IC Integrated Circuit

ICMR Input Common Mode Range

ISE Integrated Synthesis Environment

ITRS International Technology Roadmap for Semiconductors

JTAG Joint Test Action Group

LRS Low Resistance State

LS Level Shifter

LSI Large Scale Integration

LVR Linear Voltage Regulator

MCN Modified Current Normalizer

MESA Microsystems Engineering Science and Applications

MIM Metal Insulator Metal

MOM Metal Oxide Metal

MOS Metal Oxide Semiconductor

MOSFET Metal Oxide Semiconductor Field Effect Transistor

MPW Multi-Project Wafer

MSB Most Significant Bit
NDA Non Disclosure Agreement
NISO N Isolation
NMOS N-type Metal Oxide Semiconductor
NoC Network-on-Chip
OA Operational Amplifier
OCV On-Chip Variations
ODIN Online-learning Digital spiking Neuromorphic
OxRAM Oxide-based resistive Random Access Memory
PCB Printed Circuit Board
PDK Process Design Kit
PGA Pin Grid Array
PLCC Plastic Leaded Chip Carriers
PM Phase Margin
PMOS P-type Metal Oxide Semiconductor
PRS Pristine Resistance State
PVT Process Voltage and Temperature
RAM Random Access Memory
ReRAM Redox-based resistive Random Access Memory
RF Radio Frequency
ROLLS Reconfigurable On-line Learning Spiking
RSCE Reverse Short Channel Effect
RTL Register Transfer Language
SATA Serial Advanced Technology Attachment
SDRAM Synchronous Dynamic Random Access Memory
SMD Surface Mounted Device
SMU Source Measure Unit
SNN Spiking Neural Network
SPA Semiconductor Parameter Analyzer
SPDT Single Pole Double Throw

SPICE Simulation Program with Integrated Circuit Emphasis

SPST Single Pole Single Throw

SR Slew Rate

SRAM Static Random Access Memory

SSSP Supervised Single-Shot Programming

STDP Spike Time Dependent Plasticity

TEM Transmission Electron Microscope

TIA Transimpedance Amplifier

TSMC Taiwan Semiconductor Manufacturing Company

UMC United Microelectronics Corporation

USB Universal Serial Bus

VCM Voltage Common Mode

VHDL VHSIC Hardware Description Language

VLSI Very Large Scale Integration

WTA Winner Take All

ZIF Zero Insertion Force

Chapter 1

Introduction

1.1 Memristor based neuromorphic processors

1.1.1 Moore's Law and beyond Von Neumann architecture

In 1965, Gordon E.Moore postulated that the number of transistors packed into a given unit of space will double about two years [1]. However, the doubling of transistors occurred closer to every 18 months instead of every two years. Moore's law has been a macro trend and key indicator for transistor scaling in the semiconductor industry for many decades. This scaling brought the industry to a point where the transistor gate length is sub-20 nm [2]. The thinner dielectric in the gate caused gate leakage currents leading to high off currents for a transistor thereby, increasing the static power consumption. Despite these kinds of challenges, tech-giants like TSMC, Intel, etc. keep investing billions of dollars with the hope to keep Moore's law alive. Also, the ITRS came up with two areas to focus- 'More-Moore' and 'More-than-Moore' [3]. 'More-Moore' focuses on shrinking sizes of digital functionalities (logic and memory storage) to improve density and performance, while 'More-than-Moore' targets complementary techniques and novel architectures that integrate digital systems with non-digital systems to obtain performance improvement.

As the end of Moore's law seems closer than ever, there's an international effort underway to find alternatives to CMOS transistors. This lead to the concept-

'Beyond Moore', under which, devices like nanowire FET, SpinFET, tunneling transistors, atomic switch, memristors, molecular switch, etc. are been investigated. This further kept researchers to explore new approaches of computing that can move beyond the traditional von Neumann architecture [4]. Von Neumann architecture is the architecture that conventional computers and smart devices have today. It comprises the CPU, which operates sequentially on data fetched from memory.

'Neuromorphic computing' is one such approach that can be traced back to the 1980s when Caver Mead first proposed the concept of morphing biological neurons on custom silicon [5]. Like the biological brain, the main components in the neuromorphic computing system are the neurons interconnected by synapses. The main idea of silicon neuron is using the sub-threshold currents (in the order of nA) of transistors to mimic the biophysical properties that the neurons have. These brain-inspired neuromorphic computing systems have attracted research interest since they are alternate to classical von Neumann architecture because of the co-existence of memory and processing units. They are also called as non von Neumann architectures.

1.1.2 Neuromorphic chips

Neuromorphic chips are chips or ICs designed to focus on using a circuit layout that emphasize a high degree of parallelism, similar to neural net (in software code). The design can be sometimes baked as an ASIC or prototyped in an FPGA. Although early attempts to build neuromorphic chips date back to the late 1980s [6] the first large-scale implementations only came in the late 2000s [7]. The renowned neuromorphic chips in the last few decades are Neurogrid [7,8], BrainScaleS [9,10], TrueNorth [11], SpiNNaker [12] and Loihi [13]. Although they don't meet the complete architecture of biological neurons and synapses, it is always considered as a stepping stone towards it.

Neurogrid was one of the early successful neuromorphic platforms that can simulate 1 million neurons and 1 billion synapses [7]. It is also known for its fast neural simulation and energy efficiency. Neurogrid was designed in 2006 and first reported success in 2014 [8]. Its hardware platform consists of 16 neurocores, a cypress EZ-USB

FX2LP (for USB communication), a CPLD- to interface FX2 and neurocores and a daughterboard (for realizing primary axon-branching using FPGA and SRAMs). Each neurocore has a 256 × 256 silicon-neuron array, a transmitter, a receiver, a router, and two RAMs. A 12-bit control word, called the Neurogrid packet is used for neurocore-to-neurocore communication. The chip was fabricated in a 180 nm CMOS technology.

BrainscaleS' design was started in 2011 and can simulate about 196k neurons with 50 million synapses [9]. BrainscaleS was designed using wafer-scale integration, which uses the entire wafer (of diameter 20 cm, taped-out using 180 nm CMOS technology) as a single 'super-chip'. Horizontal and vertical connections on the wafer are used to pick one of the 44 reticles and each reticle has 8 HICANN, which are the integrated block of neuron circuits and its synapses. BrainscaleS 2 (the later version) was designed using 65 nm CMOS technology with built-in plasticity, which modifies both dendritic synaptic composition and synaptic weights to efficiently train large multi-compartment neurons [10].

TrueNorth, designed by IBM in 2014- has 4096 cores with 256 neurons each and each neuron has 256 synapses [11]. TrueNorth chip is designed using 28 nm CMOS technology and the chip architecture comprises a two-dimensional array of cores where long-range connections are implemented by sending spike events over a mesh routed network and local connections are established using a crossbar. Each core has a time-multiplexed neuron block, a memory (SRAM) block for storing data of neurons, a scheduler block for creating axon delays of incoming spike events, a router block for relaying spike events and an event-driven controller block. Comparatively, TrueNorth is an energy-efficient (26 pJ per synaptic event) neuromorphic platform.

SpiNNaker, built during 2009 by the University of Manchester- is a massively parallel multicore computing system. The chip is designed using UMC 130 nm CMOS technology. Physical hierarchy of the system in each node comprises two silicon dies- the SpiNNaker chip and the Mobile DDR SDRAM. The SDRAM is mounted on top of the SpiNNaker die and is wire-bonded. The 32 kB and 64 kB are inside the SpiNNaker chip as normal SRAM, like an L1 cache memory. Each processor

has 64 kB or tightly-coupled data memory and 32 kB of tightly-coupled instruction memory. Each SpiNNaker chip contains 18 identical ARM968 processors, a SDRAM controller, and a router [12]. The ARM cores communicate with each other and to outside through the router. SPIN-5 board has 48 SpiNNaker chips each with 18 ARM processors (864 ARM processors in total) and three SPARTAN6 FPGAs to communicate with other boards.

Loihi, launched by Intel in 2018- is first fully integrated asynchronous SNN chip [13]. The chip is designed using Intel's 14 nm FinFET process. It has a total of 130k neurons and 130 million synapses. The chip is a manycore mesh that comprises 128 neuromorphic cores, 2 embedded x86 processor cores, and off-chip communication interfaces. An asynchronous NoC facilitates communication between cores in the form of packetized messages with write, read-requests, read-response messages for core management, x86-to-x86 messaging, spike messages for SNN computation and barrier messages for time synchronization between cores. Each core contains 1024 spiking neural units grouped into tree-like structures.

Other neuromorphic chips include Darwin [14], ROLLS [15], ODIN [16], DYNAPs [17] etc. whose comparison based on specifications like technology, feature size, number of transistors, number of neurons, number of synapses, energy, etc. has been done in literature [18,19]. There are also low-cost user-friendly boards like NeuroShield that feature NM500 neuromorphic chip, which can be driven by raspberry pi or arduino [20].

1.1.3 Memristor- a favourable synapse

Memristors emerged as promising circuit elements for neuromorphic computing circuits. It relates electric charge and flux non-linearly. After Chua coined the word 'Memristor' and later when HP labs showed its physical existence, many memristor models were implemented to characterize, study and explore its potential applications [21–26]. Memristors are a good choice of a candidate when used as artificial synapses in CMOS-based neuromorphic computing circuits due to its property of non-volatility, analog behavior, and continuously distributed resistive states. Differ-

ent switching mechanisms such as redox-based, phase-change, the magnetic junction and ferroelectric, and different physical models such as conductive filament, Schottky barrier, charge trapping and electrochemical migration of point defect have been investigated to better understand the switching phenomenon [27–35].

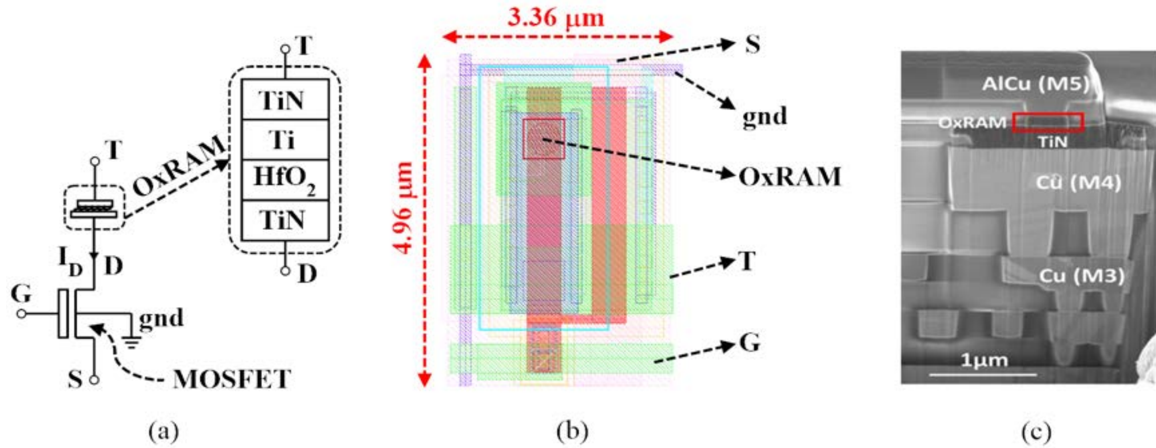


Figure 1-1: (a) 1T1R structure, (b) Layout preview of 1T1R in MAD200 PDK, (c) Microscopic view of the monolithically integrated hybrid CMOS and OxRAM [36].

Among the different resistive switching memristors, ReRAM memristor devices, that operate by a conductive filament switching mechanism, emerged to be a very promising artificial synaptic device for high-density down-scaled synaptic crossbars. ReRAM memristor technology combines the features of high-speed performance of present SRAMs with the non-volatile property of flash memory, which can be realized at low power consumption. ReRAM memristive devices are also known for their robustness and integration capability [37].

Metal oxide-based ReRAM often referred to as OxRAM, comprises a transition metal oxide layer sandwiched between two-terminal metal electrodes so that it exhibits change in resistance when voltage pulses are applied to the electrodes. The switching-behavior of these ReRAM devices depends on the transition material (also called as dielectric) and metal electrodes. A variety of such transition materials– HfO_2, NiO, Al_2O_3, Nb_2O_5, $SrTiO_3$, $Pr_{0.7}Ca_{0.3}MnO_3$, CuO_2, Ag_2S and AgGeSe– have been experimented and their switching characteristics have been studied in the literature [38–42].One such metal oxide-based ReRAM is HfO_2-based memristive OxRAM

device, which is operated in binary mode, so that, the resistance of the embedded TiN-Ti-HfO_2-TiN structure can be switched between two different resistance states, namely: LRS– typically in the order of kΩ– and HRS– in the order of hundreds of kΩ to MΩ [43, 44].

HfO_2-based OxRAMs are known for their low-switching energy, high switching speed, and high endurance when compared to other oxide-based ReRAMs [45]. The two resistance states are dynamically selected by a control voltage applied to a series-connected MOSFET (or selector MOSFET) transistor, leading to the so-called 1T1R structure, as illustrated in Fig. 1-1(a). With other collaborators in the project-NeuRAM3[1] we have the 1T1R synaptic structure that can be used in a PDK to design our circuits and this synaptic device is used in the thesis. The layout view of the 1T1R synapse in the MAD200 PDK is shown in Fig. 1-1(b). The hybrid process involves- 1) Deposition of 4 Cu Metal layers- M_1, M_2, M_3 and M_4, 2) Deposition of TiN bottom electrode and CMP[2] touch, 3) Memory stack (HfO_2 10nm/Ti 10nm/TiN) deposition by CEA-Leti[3], 4) ϕ 300 nm MESA patterning. 5) Encapsulation and CMP and 6) Placing of vias and M_5 layer deposition. Fig. 1-1(c) shows the microscopic view of the monolithically integrated hybrid CMOS and OxRAM.

NMOS is used as selector MOSFET with size- W = 6.7 µm and L = 0.5 µm, as recommended by Leti. The filament of the OxRAM is formed by applying a bias V_{TS}= 4 V, 10 µs pulse and gate bias V_{GS}= 1 V, with a recommended compliance forming current of about 1 µA. For a RESET operation, a bias of V_{ST}= 3 V, 100 ns pulse is applied by keeping the gate fully ON (V_{GT}= VDD). For a SET operation, a bias V_{TS}= 2.4 V, 100 ns pulse is applied along with the gate bias, V_{GS}= 1.5 V. For

[1]NeuRAM3 (Neural-computing architectures in advanced monolithic 3D VLSI technologies) project was an EU project that worked towards the development of a monolithically integrated 3D technology in CMOS at design rules with integrated ReRAM synaptic elements and implement on-chip learning on a scalable platform using adaptive characteristics of electronic synaptic elements. Webpage: www.neuram3.eu/

[2]Circuits Multi-Projets® (CMP) is a Multi-Project Wafer (MPW) service organization providing support for cost-effective prototyping and low volume production. Circuits are fabricated on mature process lines for academics and industrial. CMP distributes Design-Kits, that contains technology files, simulation models, design rules, and standard cell libraries. Requested design kits are sent to the customer after a non-disclosure agreement (NDA) with CMP.

[3]CEA-Leti is one of the participants of the NeuRAM3 project.

a read operation, a read voltage of V_{TS} or V_{Read}= 0.3 V is applied with a gate bias, V_{GS}= 3.8 V.

1.1.4 Memristive crossbar and sneak-path currents

The term 'crossbar switch' has its origin in 1913, when J. N. Reynolds from Western Electric thought of using a crosspoint or a coordinate array to operate a large number of relay contacts by using a small number of magnets [46]. Many crossbar based architectures using two-terminal devices for memory, logic, and neuromorphic applications have been suggested during early 2000 [47–52]. Crossbars based neuromorphic circuits shown in Fig.1-2 (a) comprise of two layers of parallel electrodes that are perpendicular to each other. The two layers of parallel electrodes act as the word-lines ($W_{1,2,3,4}$) and bit-lines ($B_{1,2,3,4}$). They are arranged in a two-dimensional array, which has a synaptic element at each intersection or crosspoint. The synaptic elements can be programmed to 'LRS' or 'HRS' that represent logic- '1' or '0' respectively when appropriate voltages are applied to the word-lines and bit-lines.

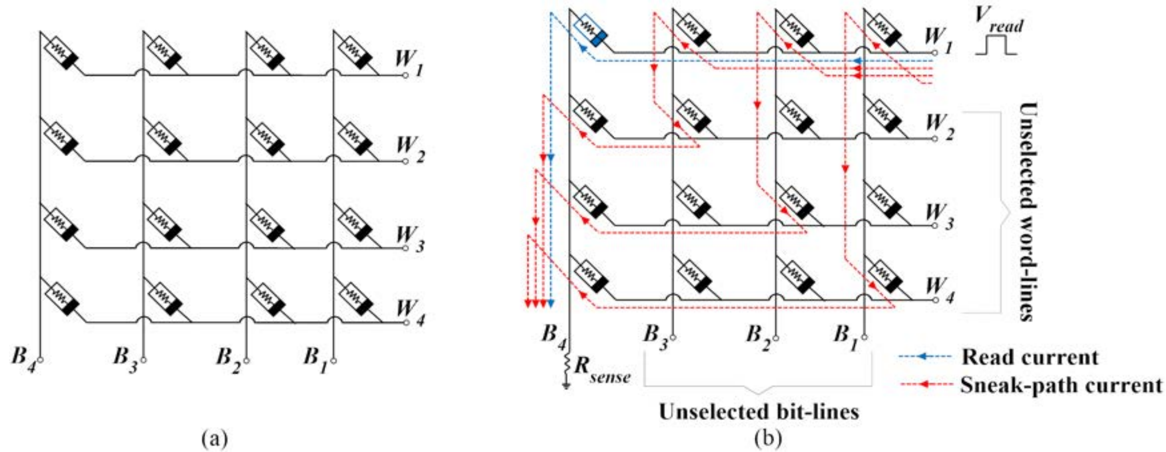


Figure 1-2: (a) A 4×4 memristive crossbar, (b) Read current and sneak-path current in a 4×4 memristive crossbar.

A big setback for the memristive crossbars is the sneak-path current. Sneak-path currents are the currents that pass through the unselected path of word-lines and bit-lines, which can aggravate read and write operation performance of the crossbar thereby limiting the scalability of the crossbar [53]. When a particular synaptic device

is targeted in a crossbar and its current is read by applying a read voltage- V_{read}, along with the desired read current, sneak-path currents also appear across the inference bit-line, as shown in Fig.1-2 (b). Sneak paths can cause the state of the synaptic device to be misread or changed unintentionally. Sneak paths are relevant when using crossbars as conventional digital memories (select bitline, wordline, and read value). However, when using the crossbar for analog vector-matrix multiplications, sneak paths are not critical.

Several architectures, synaptic devices and read/write procedures have been proposed to mitigate the effect of sneak path currents in crossbars. Synaptic devices like two anti-serial memristors [54] and architectures like unfolded crossbar with 1D1M devices [55] have been investigated to mitigate sneak-path currents. Read techniques like- multiport read-out system with mathematical cancellation of sneak-path currents [56], threshold-based read-out system [57], a two-step read process based on “open-column" semantics [58], a three-step read (or multistage read) process for determination of state of the memristor in the presence of sneak paths [59] have been investigated. Write bias schemes- V/2 and V/3 have proven low write energy and high read margins thereby, minimizing the effect of sneak-path currents [60–62]. One approach to avoid sneak-path currents is to use 1T1R synapse [63], but this can limit the scalability. However, a greater crossbar density and smaller area overhead can be achieved by fabricating the memristor fabric on top of the CMOS layer using sub-CMOS feature size nanowires using different fabrication process. This can be achieved using spatially-distributed interface pins to connect the top-level CMOS metal layer to the nanowire crossbars [64, 65]. Another interesting approach to eliminate sneak-path currents is using demultiplexer circuits based on encoded nanowire doping [66].

1.1.5 STDP learning rule

STDP is a family of learning mechanisms in computational neuroscience that exploits spike-based computation. STDP date back to 1993, when Gerstner first reported it [67]. Although the experimental existence of STDP in the biological brain has been

observed by many neuroscience researchers [68–75] the molecular and electrochemical principles behind it are still a debate [76].

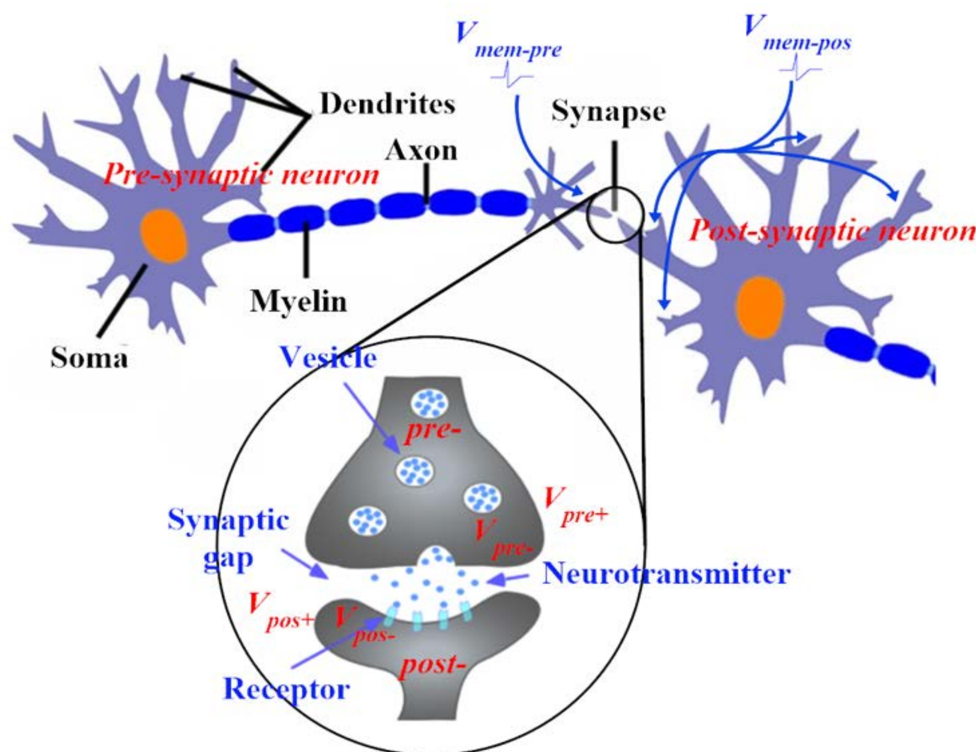


Figure 1-3: A synaptic junction that connects a pre-synaptic and a post-synaptic neuron.

Fig.1-3 illustrates a preview of the synaptic junction, where the pre-synaptic and post-synaptic neurons connect. The pre-synaptic neuron sends an action potential— $V_{mem-pre}$ (= V_{pre+} - V_{pre-}) to the synapse, which cumulatively generates a post-synaptic action potential— $V_{mem-pos}$ (= V_{pos+} - V_{pos-}) at the membrane of the post-synaptic neuron. Neurotransmitters are released into the synaptic cleft due to the pre-synaptic action potential. Each synapse or synaptic junction is characterized by synaptic weight (or strength)— w, which determines the efficacy of the pre-synaptic spike in contributing the cumulative action at post-synaptic neurons. By STDP, the change in synaptic weight, $\triangle$w is a function of the time difference between the pre-synaptic spike, $t_{\rm pre}$ and post-synaptic spike, $t_{\rm pos}$. Hence, change in synaptic weight, $\triangle \mathrm{w} = \xi(\triangle \mathrm{T})$, where $\triangle \mathrm{T} = \mathrm{t_{pos}} - \mathrm{t_{pre}}$. For positive $\triangle$T a potentiation of synaptic weight happens i.e. $\triangle \mathrm{w} > 0$ and for negative $\triangle$T a depression of synaptic weight

happens i.e. $\triangle w < 0$. Unlike Hebbian learning [77], which considers the mean firing rate of pre and post synaptic spikes, STDP takes into account the spikes' relative time. The machine learning and computational neuroscience community have been using STDP for applications like pattern learning, object recognition since the early 2000 [78–86].

1.2 Contributions in Memristor based crossbars

Trade-offs such as reduced chip area and energy efficiency are potential challenges that the chip designers of the neuromorphic system have. The thesis mainly focuses to address these challenges on a memristor-based neuromorphic chip which are explained here in brief.

The first challenge was to make the neuromorphic system low power or energy-efficient during inference. For this, we need to investigate how small read pulses can be sent. Applying such small read pulses go vain when the DC offset of the system is decisively more to ruin the system. To overcome this, we proposed a bulk-based three-stage DC offset calibration scheme across the wordlines of the memristive crossbar via a PMOS-based 2-stage differential opamp. Chapter 3 describes the proposed calibration scheme. It gives an insight view of the design of the proposed calibration scheme in the chip, chip packaging, design, and assembly of the test-PCBs, experimental set-up, and results. The results certainly include- preliminary tests of the chip, characterization of synapses, results of the calibration scheme, template matching results for pattern recognition, pattern recognition using programming and learning algorithms.

The second challenge was related to the reduced chip area. During inference operation in a memristor-based fully-connected neural network, the LRS read currents are higher. Due to this, an extremely large integrating capacitor (larger than nF) is needed, which can easily increase the chip area and limit scalability for large networks. To overcome this, we proposed a new current attenuator circuit that can scale down the inference currents by a factor of about 10^4. The results are compared with other

current attenuators. Chapter 4 describes the design of the new current attenuator and its results.

The thesis also throws light on characterizing memristors using three different experimental set-ups, which are described in Chapter 2. One experimental set-up was using a commercially available memristor and characterizing it in ArC ONE platform. The second experimental set-up is based on a full custom silicon design that has memristors on top of it by a hybrid nano-CMOS process. Here, the characterization of memristors is done using custom-designed PCBs and FPGA driver. The third experimental set-up is based on wafer-level measurments of memristors, which are characterized using a semiconductor parameter analyzer directly or by a computer using GPIB. These characterization experiments emphasize skill-based learning on different platforms and give a first hands-on experience towards establishing a memristor-based learning system.

Chapter 2

Charaterization of memristors

2.1 Need for memristor — characterization

Characterization of memristors is vital not only to understand the underlying working mechanism and electrochemical reactions happening in the device but also to optimize the performance of future devices. Resistance switching in memristors can be characterized or studied using spectroscopic techniques, scanning probes, in situ TEM observations, scanning tunneling microscopy investigations, using electronic circuits, and other electrical approaches [87–91]. Specifications— such as ON/OFF ratio, switching speed, retention[1], endurance[2] and variation (device-to-device, cycle-to-cycle, variations due to fab. process, etc.) determine the qualitative behavior of the memristor. The fabricated memristors to be characterized can be either at the wafer level or can be cut into tiny dies for encapsulation or bonding, which depend on the experimental set-up for characterization. Here, in our work, we have characterized the resistance states (LRS and HRS) of three different memristors using various experimental set-ups. Some memristor foundries will have a wafer-level test set-up to test quickly, get results, and alter the materials in the memristive device so that, it saves time. On the other hand, in some, it takes some time to pack and plan test-boards to test the devices. The reason to characterize different memristors is

[1]Retention- measuring the resistance of the device periodically for a fixed overall duration
[2]Endurance- switching a device between LRS and HRS for several cycles

to mainly investigate their resistance states, functionality, and working so that they can be used to build system-level approaches such as programming STDP or even implementing learning.

2.2 Characterization of Neurobit memristors using ArC ONE® Platform

ArC ONE® platform is used to characterize single or array of selector-less memristor devices either directly on wafers or in packed samples [92–94]. The ArC ONE® board mainly comprises an mBED microcontroller driver, bias generating opamps, sense resistor bank, read/write feedback buffers, programmable current source for current pulsing, TIA read opamps, PLCC68 DIP socket, header pins, power management block, resistor banks and digital components like decoders, multiplexers and switches. The working principle of ArC ONE® platform is to pick the active and default wordlines and bitlines and apply DC voltage pulses for the needed operation. The user can perform operations such as read, write and erase in a sequence or a closed-loop to form the device, to plot I-V characteristics, to do endurance and retention tests. Users can choose between V/2 and V/3 write bias schemes for mitigating sneak-path currents [60–62] and the maximum crossbar size that can be characterized is 32×32. The digital lines of the platform are controlled using mBed microcontroller, which receives the command from a bespoke software control interface (ArC ONE Control®), programmed in MATLAB and Python, via a USB link. Memristor devices on the wafer are characterized using header pins via a custom probe-card[3], whereas memristor samples are packed in the PLCC68 package and are placed in the DIP socket.

Neuro-Bit— world's first commercially available memristor is an Ag-chalcogenide based on a two-terminal device [95–99]. The device structure of Neuro-Bit comprises a layer of Ge_2Se_3 (30 nm), Ag_2Se (50 nm), and Ag (50 nm)- sandwiched between the

[3]Probe card is a jig docked to a wafer prober to serve as a connector between the LSI chip electrodes and an LSI tester such as SPA. Probe card provides an electrical path between the tester and the circuits on the wafer.

top and bottom electrodes [100]. The 44-pin PLCC breakout board of Neuro-Bit has 20 bonded memristor devices [101]. The pre-programming resistance of the device is 50 MΩ. A positive voltage sweep from 0 to 1 V with compliance current values between 100 nA and 30 µA can cause the device to switch to LRS (also called as 'write' and typical LRS value = 8 kΩ) at a certain voltage (called V_{set}) by making the Ag^+ ions to migrate into the active chalcogenide, Ge_2Se_3 layer and create a low resistance path through the insulator. Similarly, a negative voltage sweep from 0 to -1 V with compliance current values between 100 µA and 10 mA can cause the device to switch to HRS (also called as 'erase' and typical HRS value = 13 MΩ) at a certain voltage (called V_{reset}) by removing Ag^+ ions back from the active chalcogenide layer. The device is read by applying a read voltage, V_{read} = 50 mV.

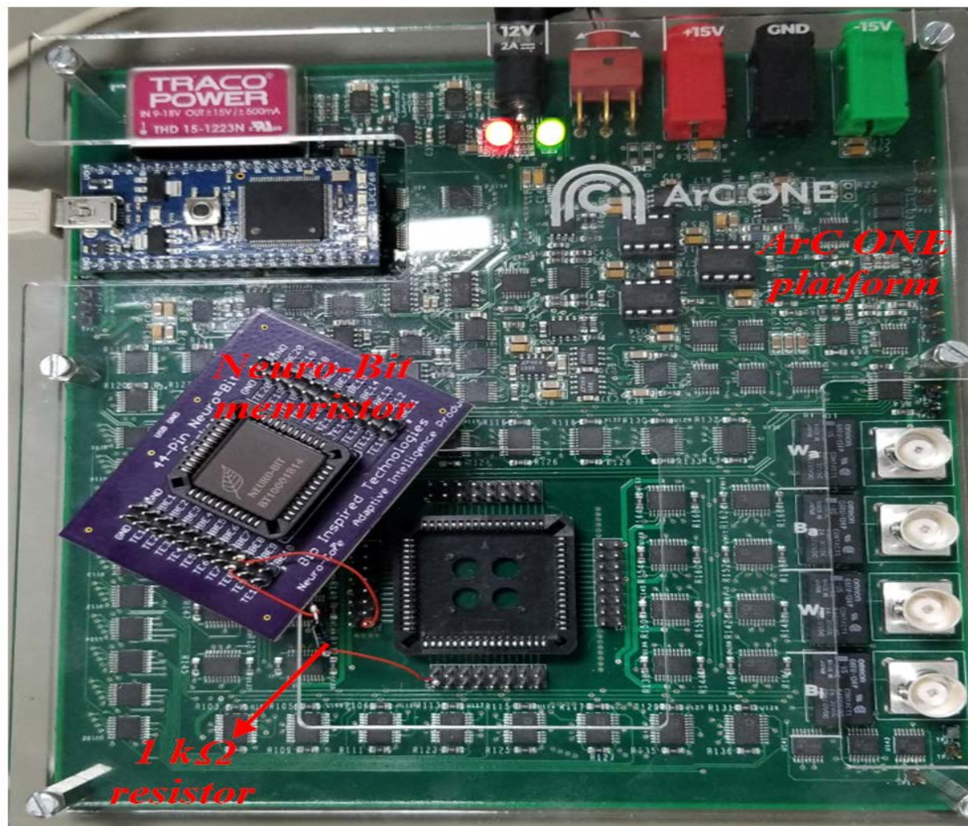


Figure 2-1: A Neuro-Bit memristor connected to the active word-line and bit-line terminals of ArC ONE® platform.

The main challenges in operating Neuro-Bit devices are the intrinsic variability and sensitivity of the device. No two devices switch at same V_{set} or V_{reset} or some-

times even at same compliance current and similar response may not be observed in two or more cycles of 'write' or 'erase' for the same device. These variations attribute to the stochasticity of the synaptic device, which results in the escape of local minima during learning and inference [102]. Although the stochastic synapses raise reliability concerns in ANN, there are many embracing approaches such as- increasing the effective resolution of synaptic weights [11, 103, 104], using stochastic neurons [105, 106], etc. Neuro-Bit devices are also so sensitive so that, when carelessly biased above the needed V_{set} or biased below the needed V_{reset} or without the required compliance current can easily short the device or sometimes permanently hang at LRS or can result in irreparable damage when transients exceed few mV (also includes environmental noise, if any). Hence, it is always recommended to carefully set-up the experimental test-bench and to attempt 'write' and 'erase' operations with very conservative values for each Neuro-Bit memristor until the user is comfortable with its performance.

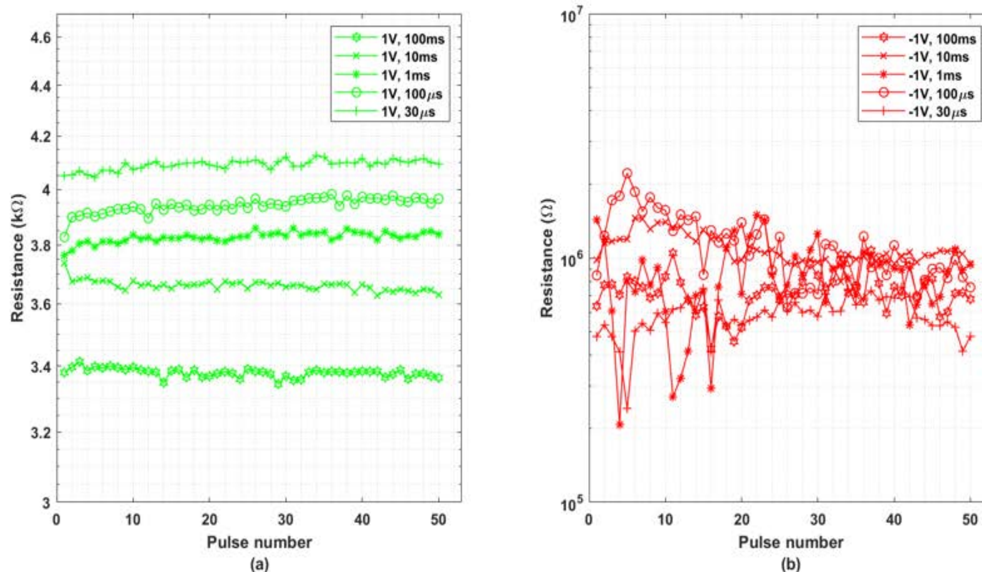


Figure 2-2: (a) LRS results of a Neuro-Bit memristor for 50 'write' pulses for different pulse widths, (b) HRS results of a Neuro-Bit memristor for 50 'erase' pulses for different pulse-widths

In our experiment, we want to have a first-look of resistance switching of Neuro-Bit memristor by using 'write' and 'erase' DC pulses. For this, we initially tried to verify the I-V characteristic of Neuro-Bit memristors using HP4145 SPA, whose observation

of DC voltage sweep measurements are documented in the user manual [101]. We observed the high indeterministic nature of the Neuro-Bit memristors during the DC voltage sweep experiments on several devices. After observing some successful repetitive switching in one Neuro-Bit device in HP4145 SPA, we connected it to the active wordline and bitline of ArC ONE® platform through a 1 kΩ resistor (as a safety precaution to limit current during LRS) to characterize its switching using DC voltage pulses, as shown in Fig.2-1. We varied both amplitude and pulse-width of the 'write' and 'erase' DC pulses using the ArC ONE Control ® GUI interface. Fig.2-2 shows the LRS and HRS results for 50 'write' and 'erase' pulses for different pulse-widths. Fig.2-3 shows the LRS and HRS results for 50 'write' and 'erase' pulses for the different amplitudes of the pulse. A high variability at HRS is observed in these results.

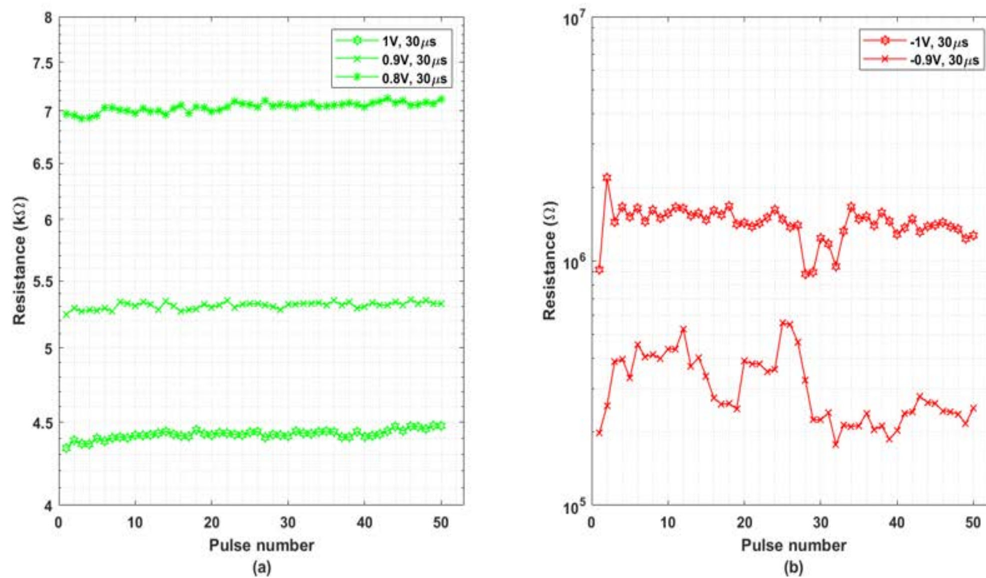


Figure 2-3: (a) LRS results of a Neuro-Bit memristor for 50 'write' pulses for amplitudes, (b) HRS results of a Neuro-Bit memristor for 50 'erase' pulses for different amplitudes

2.3 Characterization of 1T1R OxRAM-based memristors using customized PCBs

Here, we made a full-customized experimental set-up for characterizing 1T1R OxRAM-based memristor, whose design-flow mainly involved— designing 1T1R crossbars using MAD200[4] PDK, packaging of the chip, planning circuits for the test-boards in order to test the packed-chip, designing the test-PCBs, assembling of components on PCB followed-by PCB mounting, making auxiliary boards, setting-up the experimental test-bench and programming drivers to get characterization results.

2.3.1 Design of 4×4 and 8×8 1T1R crossbars and its packaging

Before designing the crossbars, a single 1T1R device (shown in Fig. 1-1(a)) is nominally simulated in eldoD[5] simulator using MAD200 PDK design tool to know the range of resistance values.

After forming the device, a sequence of voltage pulses are applied across the terminals of 1T1R device, such that a repetitive sequence of 'Erase'-'Read'-'Write'-'Read' operations are performed after forming the device. Voltage biases, drain current of the selector MOSFET, and resistance of the device are shown in Fig. 2-4. The nominal simulation results in LRS = 13.65 kΩ and HRS = 836.4 kΩ for a 1T1R device. Since these are the preliminary analyses, a reasonable amplitude and pulse widths are used for switching. Later, in the experimental set-up, we have used the Leti[6]-recommended biases and pulse-widths for different operations. Fig. 2-5 shows the schematic and layout view of the 4×4 1T1R crossbar and Fig. 2-6 shows the schematic and layout view of the 8×8 1T1R crossbar. The principle behind operating the crossbar is—

[4]MAD200 (or Memory Advanced Demonstrator 200mm) uses monolithic integration of OxRAMs above the four-metal layered 130 nm CMOS technology. Leti collaborates to grow OxRAMs above these metal layers, which is followed by the deposition of the metal layer, M5.

[5]EldoD is used as the simulator, as the OxRAM is modeled in SPICE netlist and is used as an add-on in the PDK.

[6]CEA-Leti is a foundry and one of the participants of the NeuRAM3 project. Although Leti has recommended and shared the promising bias conditions (amplitude, pulse-width, and compliance current) for the 1T1R device, we did try various amplitudes and pulse-widths with our test-benches to find the optimal values.

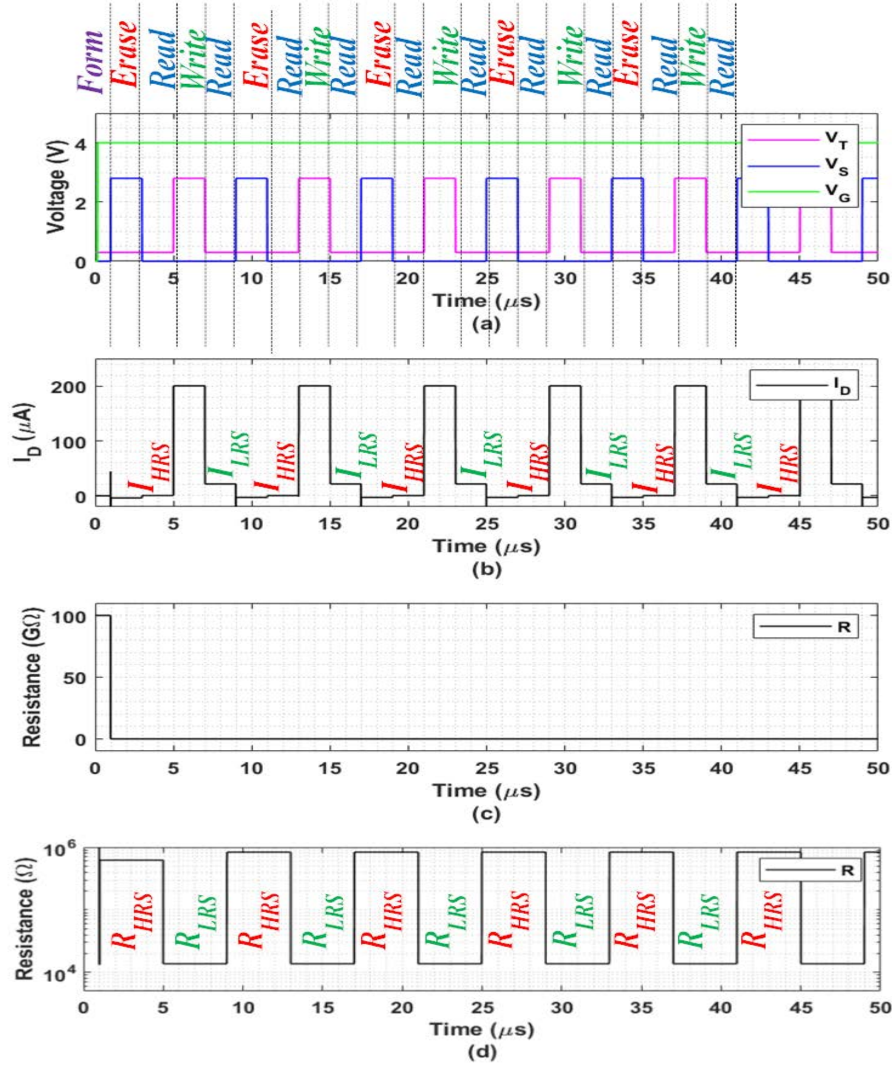


Figure 2-4: Simulation results showing repetitive sequence of 'Erase'-'Read'-'Write'-'Read' operations carried out on 1T1R device after forming it- (a) Voltage biases applied at the terminals of the 1T1R device, (b) Drain currents of the selector MOSFET, (c) Resistance of the OxRAM, (d) A zoom preview of the resistance of the OxRAM after forming.

to choose the active wordline (or row), bitline (or column) and gate-bias and perform the needed operation such as 'form', 'erase', 'write' and 'read'. The remaining default rows and columns are biased with default values so that their corresponding 1T1R devices are not disturbed. The active and default biases for OxRAM operations slightly vary depending on the configuration of the gate-lines. In 4 × 4 1T1R crossbar the gate-lines are pulled bitline-wise whereas, in 8 × 8 1T1R crossbar, the gate-lines are pulled wordline-wise.

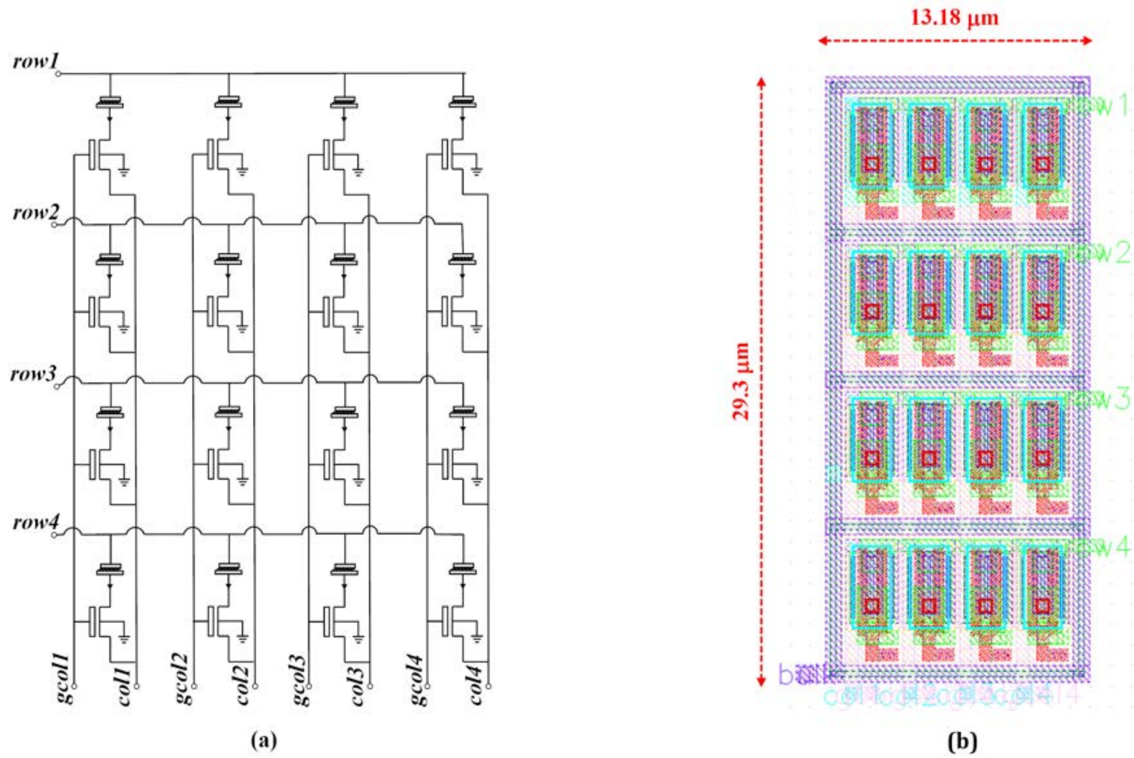


Figure 2-5: (a) Schematic view of the 4 × 4 1T1R crossbar, (b) Layout view of the 4 × 4 1T1R crossbar.

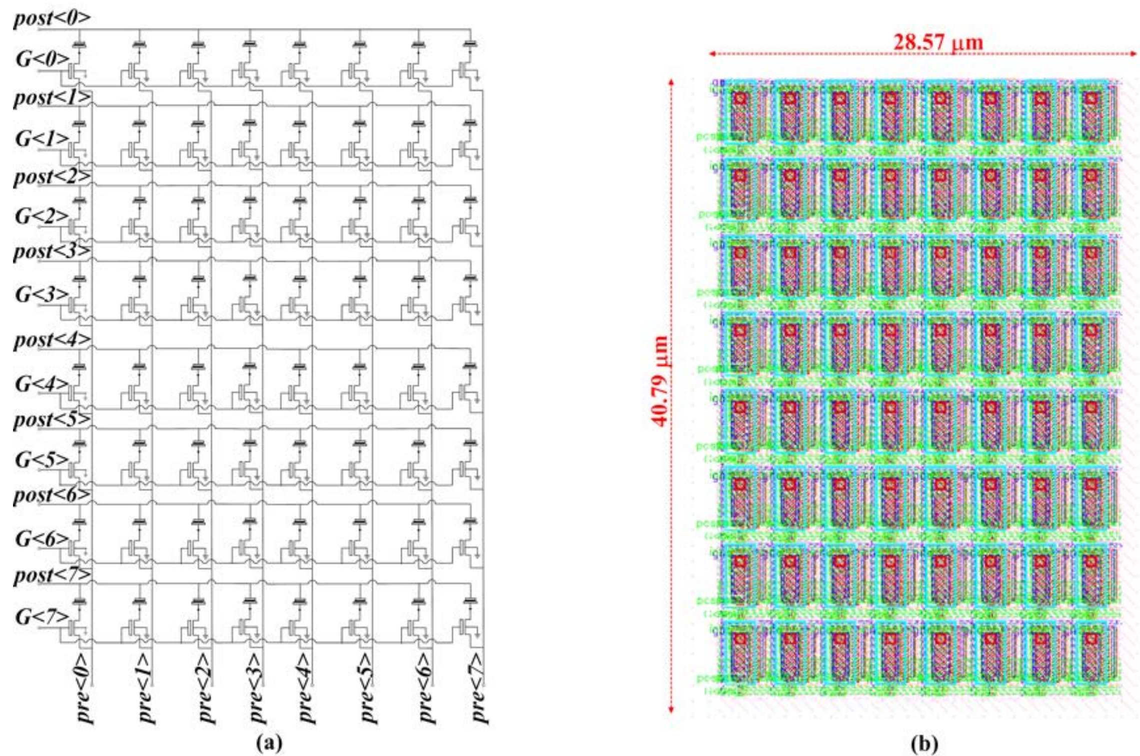


Figure 2-6: (a) Schematic view of the 8 × 8 1T1R crossbar, (b) Layout view of the 8 × 8 1T1R crossbar.

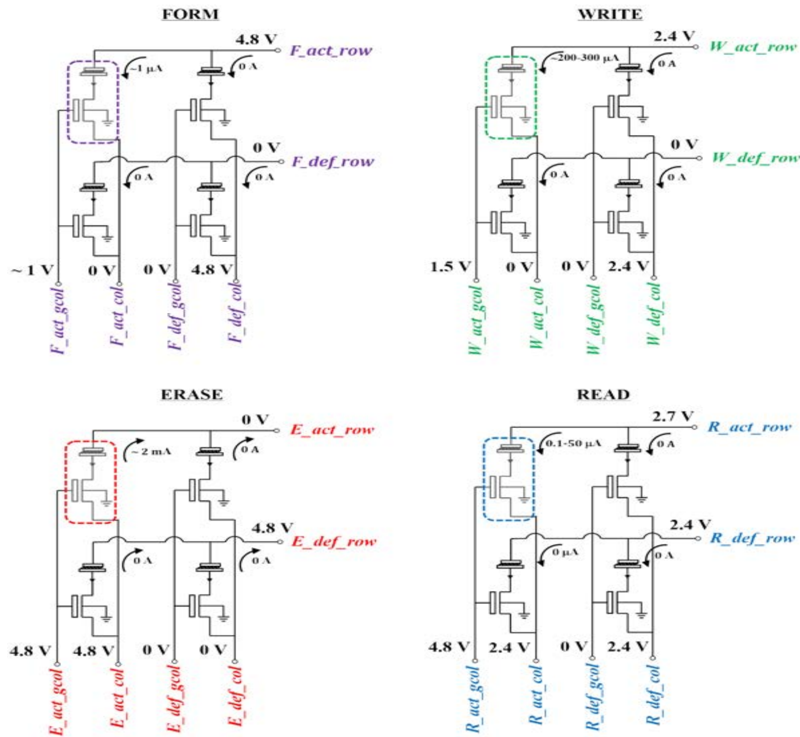


Figure 2-7: Switching biases for active and default wordlines and bitlines for various operations performed in OxRAM when gate-lines are pulled bitline-wise.

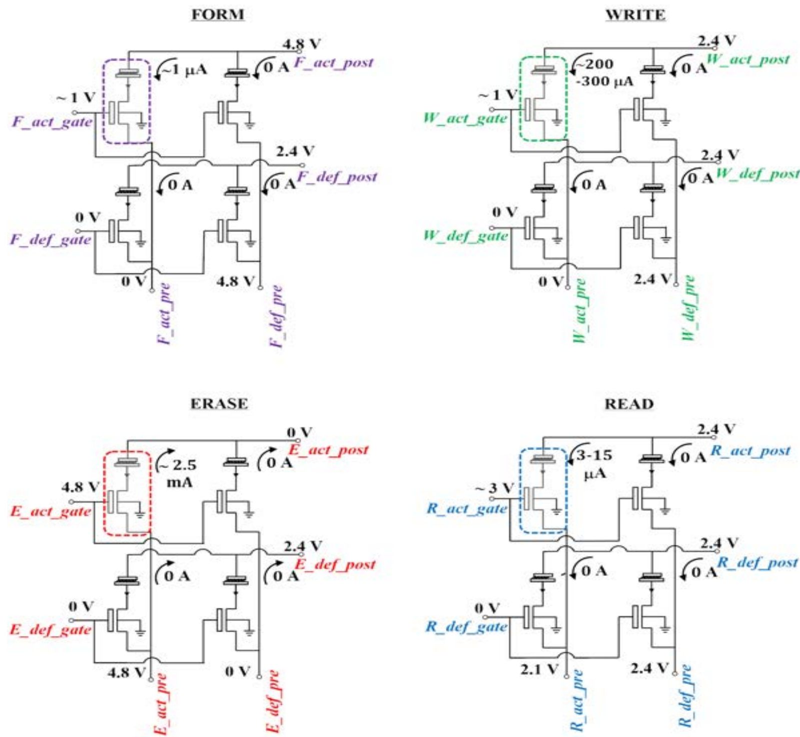


Figure 2-8: Switching biases for active and default wordlines and bitlines for various operations performed in OxRAM when gate-lines are pulled wordline-wise.

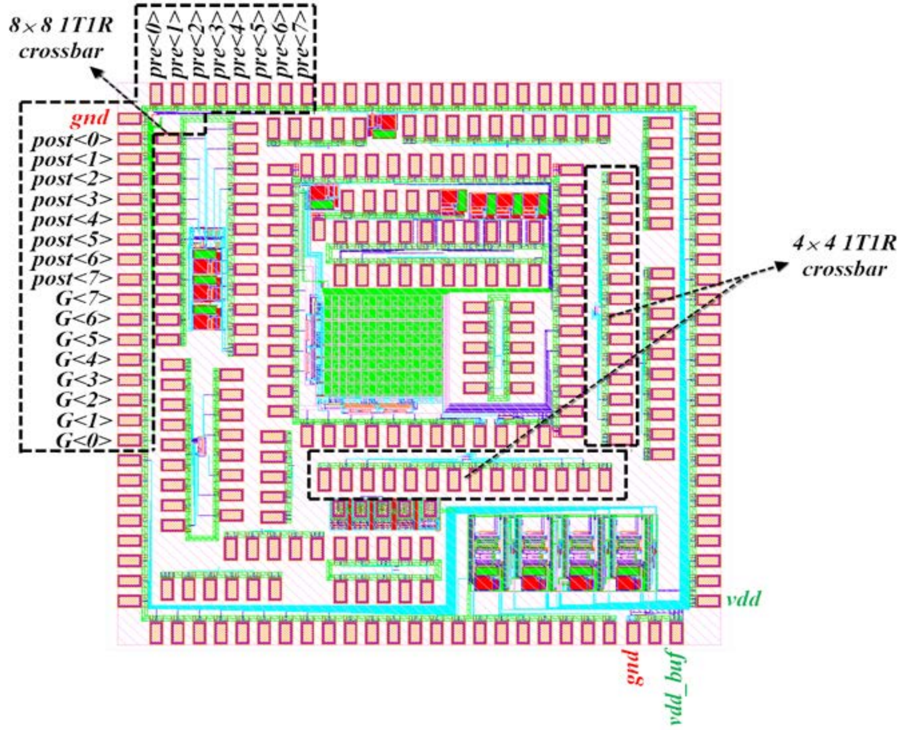


Figure 2-9: Layout view of MAD200 chip showing layouts of 4 × 4 and 8 × 8 1T1R crossbars duly connected with their pads.

The change in orientation of the gate-lines is made between the two crossbars to differentiate and realize how the active and default biases of bitlines and gate-lines vary in both cases. Nevertheless, in both crossbars the principle idea– to target a 1T1R device and perform the desired OxRAM operation is the same. Fig. 2-7 shows the active and default biases for various operations performed in OxRAM when the gate-lines are pulled bitline-wise in a crossbar whereas, fig. 2-8 shows the active and default biases for various operations performed in OxRAM when the gate-lines are pulled wordline-wise. In figures- fig. 2-7 and fig. 2-8 the nomenclature used for the terminals is 'X_YYYY_ZZZZ', where 'X' denotes the first-letter of the operations such as, 'F' (for 'form' operation), 'W' (for 'write' operation), 'E' (for 'erase' operation) and 'R' (for 'read' operation). 'YYYY' denotes the type of bias, which can be 'act' (for active) or 'def' (for default). 'ZZZZ' denotes the type of the terminal, which can be either wordline (denoted as 'row' or 'post') or bitline (denoted as 'col' or 'pre') or gate-terminal (denoted as 'gcol' or 'gate').

Both the crossbars are part of the circuits that are designed and taped out in the MAD200[7] run. MAD200 chip comprises three categories of layouts and they are— layouts of circuits whose pads form the outer-ring[8], layouts of circuits whose pads form the Inner-ring (which has other circuits submitted for MAD200 run), and the layouts of circuits that are present between the two rings. The layout of 8 × 8 1T1R crossbar along with its pads is part of the outer-ring, whereas the layout of 4 × 4 1T1R crossbar is present between the two rings. The layout view of MAD200 chip highlighting both 4 × 4 and 8 × 8 1T1R crossbars along with their pads are shown in Fig. 2-9. The outer-ring (which has 8 × 8 1T1R crossbar) is packaged in PGA100 package and the 4 × 4 1T1R crossbar is packaged in the PLCC52 package. Fig. 2-10(a) shows the top view of the packaged chip that has a layout of 4 × 4 1T1R crossbar. Fig. 2-10(b) shows the layout view of 4 × 4 1T1R crossbar with its pads numbered for packaging. Fig. 2-11(a) shows the top view of the packaged chip that has outer-ring and Fig. 2-11(b) shows the bonding diagram of the outer-ring using the PGA100 package.

2.3.2 Design of circuits for the test-PCBs

Test-circuits are designed on PCBs to characterize the crossbars. Fig. 2-12 and Fig. 2-13 show the functional block diagram for testing 4 × 4 and 8 × 8 1T1R crossbars respectively. The main functional blocks in these test-circuits are the opamps (used for analog biases), switches, level-shifters (used for bi-directional conversion between 4.8 and 3.3 V), decoders, linear voltage regulators (used for ensuring different supply voltages), ADC (used for reading the current) and SPARTAN®-6 FPGA board (used to program and digitally control the PCB and crossbar).

The nomenclatures used for the functional blocks in Fig. 2-12 and Fig. 2-13 are— 'S' for switches, 'ADC' for Analog-to-Digital converter, 'OA' for Operational Am-

[7] MAD200 is run for tape-out of the microchip in the NeuRAM3 project. The GDS file of the chip was submitted on February 2017 and the chip (also referred to as 'MAD200 chip') was received after 18 months due to the complex hybrid fab. procedure.

[8] Outer-ring of MAD200 chip comprises layouts of three circuits along with its pads. Layout of 8 × 8 1T1R crossbar is one of the three circuits.

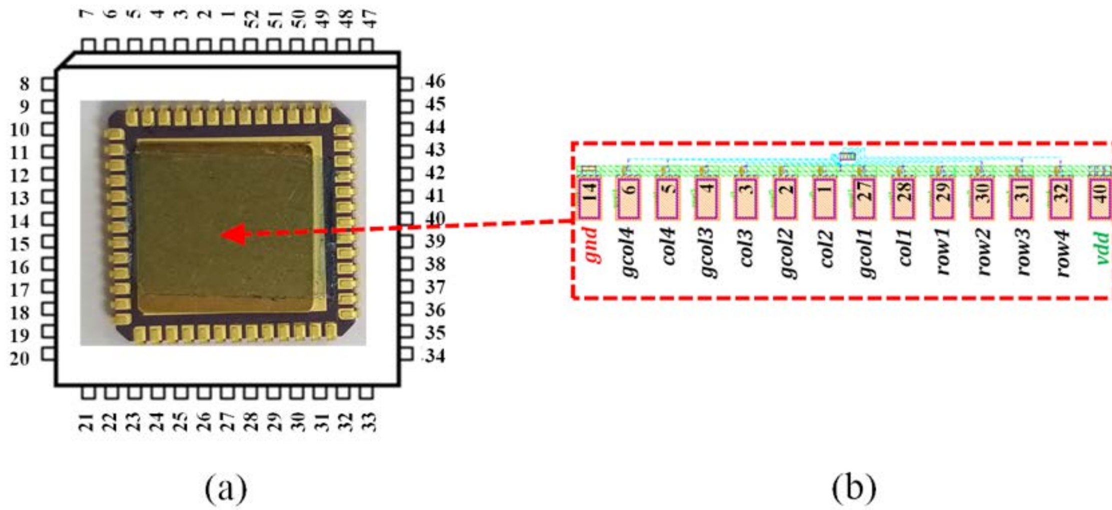


Figure 2-10: (a) Top view of a PLCC52 package (with its pin numbers) that is packaged with 4 × 4 1T1R crossbar of MAD200 chip, (b) Layout view of 4 × 4 1T1R crossbar duly labeled and numbered for packaging in PLCC52.

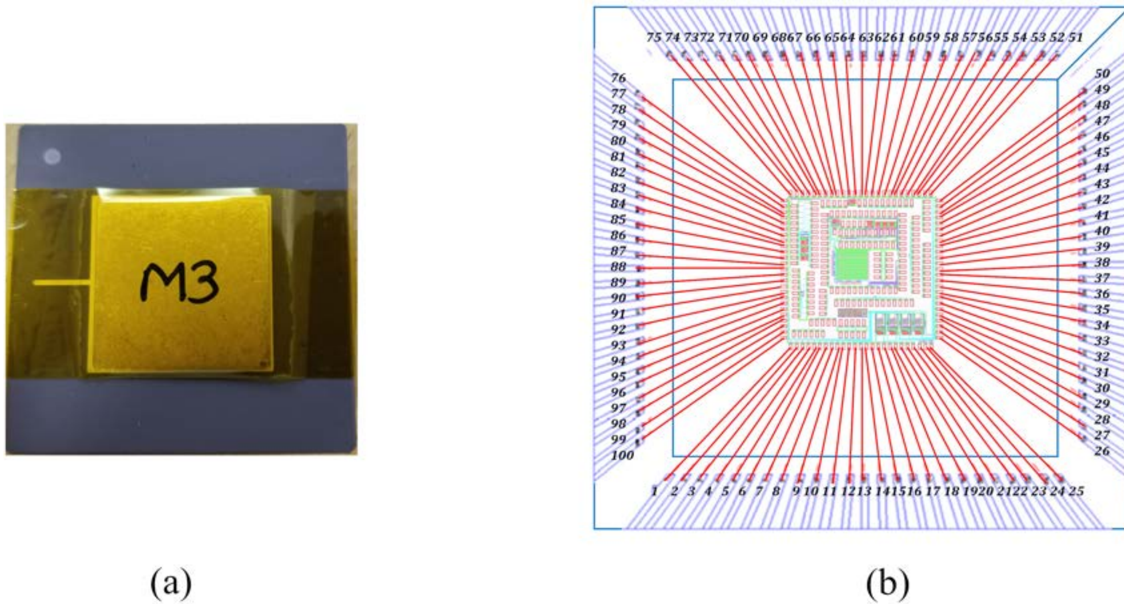


Figure 2-11: (a) Top view of a PGA100 package that is packaged with Outer-ring of MAD200 chip, (b) Bonding diagram of the Outer-ring using PGA100 package.

plifiers', 'LS' for Level-shifters, and 'LVR' for Linear Voltage Regulator. Functional block diagrams also show the control-bits needed to target a device in 4 × 4 (6-bit) and 8 × 8 (9-bit) crossbars. These control-bits are marked in color in each device location of the crossbar array. A detailed schematic of the test-circuits for testing 4 × 4 and 8 × 8 1T1R crossbars are shown in Fig. A-1 and Fig. A-5 respectively in

appendix A.

Test-circuits for crossbars facilitate targeting a 1T1R device in the crossbar and performing the needed operation without disturbing other devices. The active and default- wordlines, bitlines, and gates are chosen using decoders. The opamps are used to set optimal values of biases that are needed for the targeted 1T1R device to 'form', 'erase', 'write', 'read', or keep 'idle' or 'global'[9] values. All opamps are also availed with the flexibility to tune the feedback components in the case, if needed for stability reasons. The switches between the opamps and crossbar-terminals are digitally controlled by FPGA to make sure that the desired biases are applied for the corresponding operation. The switches are connected such that, a 3-bit control word, $\{FPGA_A, FPGA_B, FPGA_C\}$ is used to decide the needed operation performed on the OxRAM in the targeted 1T1R device in the crossbar, whose possible control

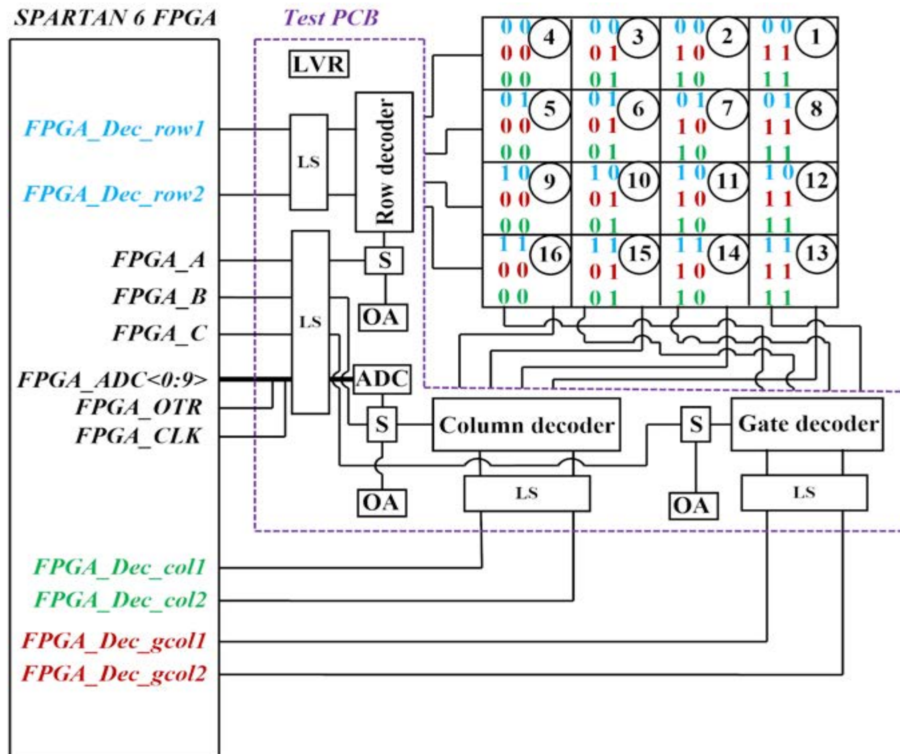


Figure 2-12: Functional block diagram of the test-circuit for testing 4 × 4 1T1R crossbar.

[9]'Idle' or 'Global' operation of OxRAM is biasing the crossbar terminals with equal voltage amplitudes and keeping the gate-biases to 0V.

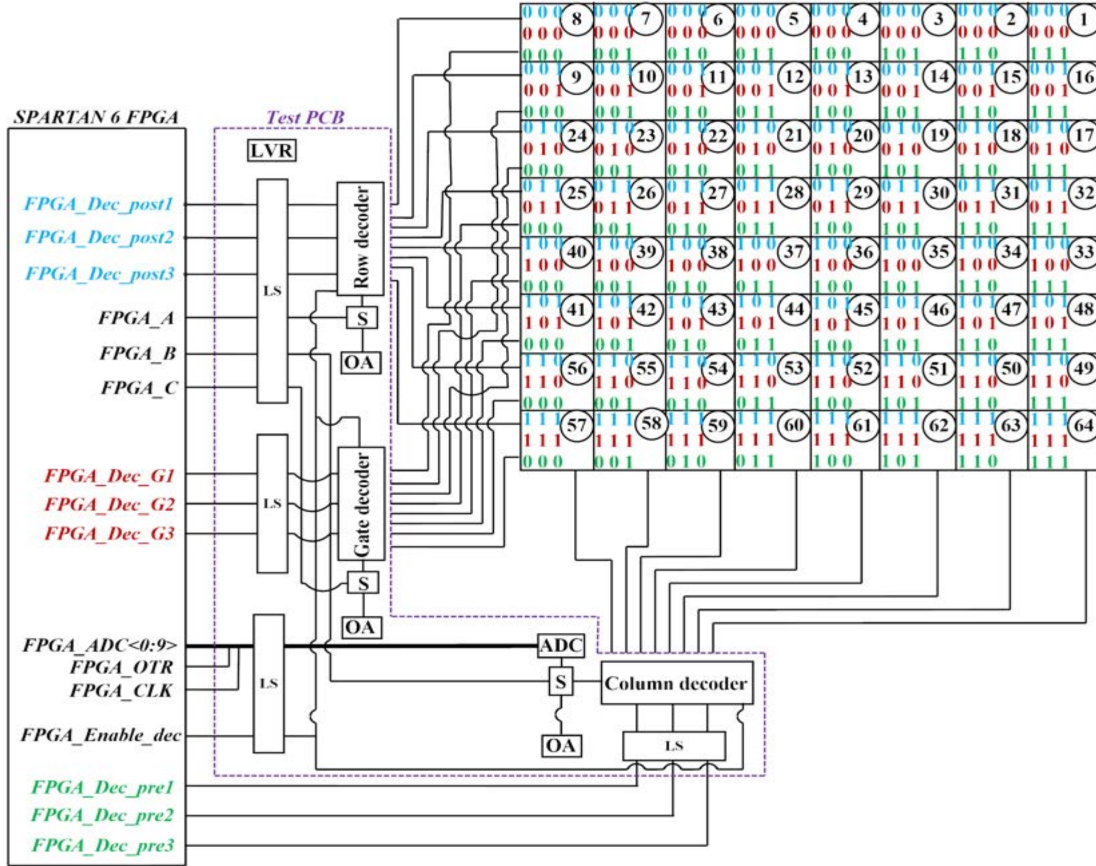


Figure 2-13: Functional block diagram of the test-circuit for testing 8 × 8 1T1R crossbar.

bits and its corresponding OxRAM operations are shown in Table 2.1. Here, 'ON' indicates 4.8 V (vdd) and 'OFF' indicates 0 V.

2.3.3 Assembly and mounting of test-PCBs and customized boards

Test-PCBs are made according to the test-circuits described in Fig. A-1, Fig. A-5, Fig. A-6 and Fig. A-7. Two four layered-PCBs are designed using OrCAD® CIS[10] and Allegro®[11], whose design details are described in appendices A.1 and A.2. One

[10]OrCAD® Capture is one of the most widely used schematic design solutions for the creation and documentation of electrical circuits. Coupled with the optional OrCAD CIS product for component data management, the designer can use components in the schematic that can be used later in Allegro® PCB designer suite to design PCBs.

[11]Allegro® PCB Designer of Cadence® is a scalable, proven PCB design environment that addresses technological and methodological challenges thereby, making the design cycles shorter and

Table 2.1: Control-bit for performing different operations for OxRAM in the targeted 1T1R device in the crossbar.

OxRAM operation	FPGA_A	FPGA_B	FPGA_C
'Form'	ON	ON	ON
'Erase'	OFF	ON	ON
'Write'	OFF	OFF	ON
'Read'	OFF	OFF	OFF
'Idle' or 'Global'	ON	OFF	ON

of them is designed exclusively for testing 4×4 1T1R crossbar and the other is for testing different circuits in the outer-ring. 8×8 1T1R crossbar is a part of the circuits in the outer-ring. The components used for the PCB are chosen carefully by going through their data-sheets. Parameters such as, operating voltage range, bandwidth, switching characteristics, etc. are considered for choosing the components, and for those components which have PSpice[12] model— are simulated to visualize their characteristics. Once the PCBs are designed the resulting gerber[13] files are sent to a PCB manufacturing company and the components in the BOM list are ordered. The PCB components are soldered to the PCB and the PCB is mounted using corner screws. Some guidelines for a better PCB assembly and mounting practice are narrated in appendix A.3. Along with the PCBs, few auxiliary boards are made to avail of some buttons (to facilitate programming of driver) and to ease probing of terminals of crossbars.

2.3.4 Description and working of the experimental set-up

Experimental set-up for characterizing both 4×4 and 8×8 1T1R crossbars is similar except few changes. The main components in the experimental set-ups are the pack-

predictable.

[12]PSpice models of more than 33k components can be simulated in Cadence® before making PCB layouts to observe their working characteristics at circuits-level.

[13]Gerber format is an open 2D binary vector image file format. It is the standard file used by PCB industry software to describe the printed circuit board images: copper layers, solder mask, legend, etc. The gerber files can be viewed in any third-party gerber viewer such as ViewMate, etc. for verification. Gerber files are the raw files needed by the PCB manufacturing company to make PCBs.

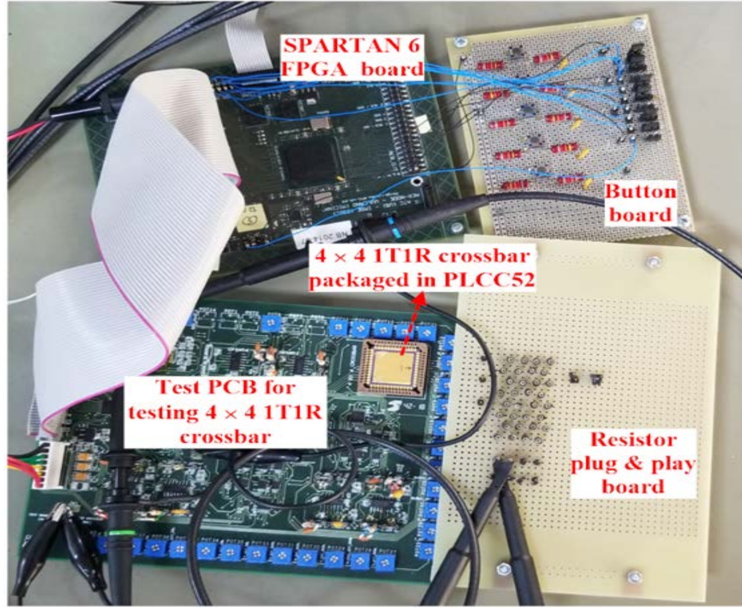


Figure 2-14: Experimental set-up for characterizing 4 × 4 1T1R crossbar.

aged chip, test-boards, a resistor plug-&-play board, button board, an FPGA driver board, and a mixed-signal oscilloscope. Fig. 2-14 and Fig. 2-14 show the experimental set-up for testing 4 × 4 and 8 × 8 1T1R crossbars respectively. Equipment like— regulated power supplies, banana cables, connecting wires, etc. are also used. The test-PCB is sourced with regulated power supplies of 10 V, 4.8 V, and 3.3 V. A 20 × 2 header bus connects the test-PCB with the FPGA board. Some digital output pins of FPGA are connected manually to the pins of button-board to establish connections for the push-button to use them. The resistor plug-&-play board consists of a crossbar arrangement of header-pins where the resistors can be plugged if needed. It is connected to the crossbar terminals using wires that run underneath both boards. The plug-&-play board facilitates testing with resistors before moving to test the chip and eases probing of crossbar terminals during testing. A separate section that dedicates to the choice of the driver is explained in Section 3.4.4 of Chapter 3. Fig. 3-30 of Section 3.4.4 shows the comparison of ON/OFF pulses of different widths (5 ns, 10 ns, 15 ns, 20 ns, 25 ns, and 30 ns) programmed using Spartan®-6 FPGA. It is observed that a full digital strength of 3.3 V can be reached at about 10 ns fast pulse using Spartan®-6 FPGA. Spartan®-6 FPGA board can deliver fast pulses by carefully

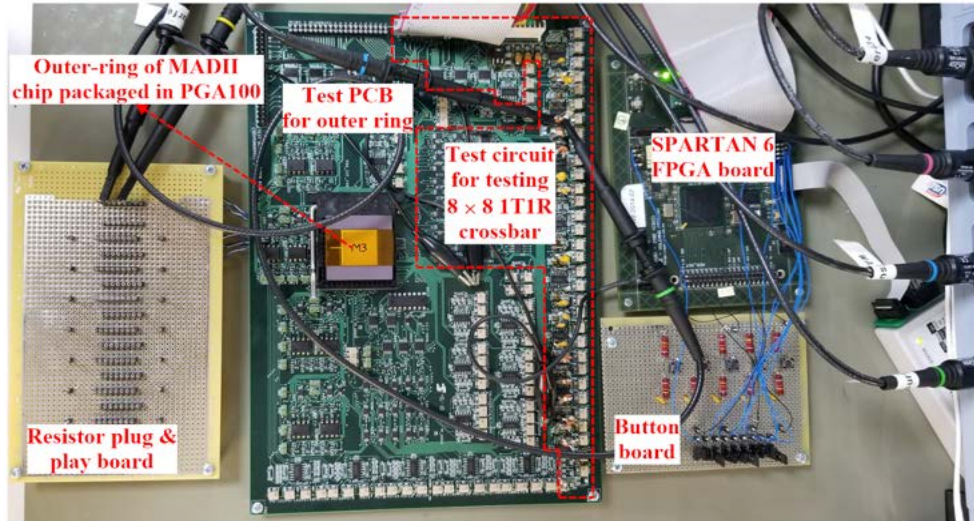


Figure 2-15: Experimental set-up for characterizing 8 × 8 1T1R crossbar.

configuring clocking wizard tool in core generator. Configurations such as slew rate (QUIETIO mode), drive strength (2 mA), etc. are set to have less peak over-shoot and ringing when pulses are applied. Here, in our experiments Spartan®-6 FPGA driver is chosen to program the test-PCBs. The filter in the oscilloscope probes is also tuned to have smoothly settled voltage pulses with no or minimal overshoot or ringing. A button-board is made and added to the experimental set-up, which has header-pins with jumpers and exclusive buttons that are programmed to do OxRAM operations like 'form', 'erase', 'write', 'read' or keep 'idle' or 'global' in a particular sequence. The header-pins with jumpers are used to provide input bits to the decoders on the test-PCB to choose the active and defaults rows, columns, and gates thereby, targeting a particular 1T1R device in the crossbar.

An FSM is programmed in FPGA using VHDL language to define functions of each button on the button-board. Different states are made in FSM as shown in Fig. 2-16 in such a way that push-buttons on the button board are programmed to establish different OxRAM operations in a sequence on the targeted 1T1R device. One of the push-buttons is programmed to do 'Form_Global_Read' task, while another button is programmed to do 'Erase_Global_Read' task. Other buttons are dedicatedly programmed to do 'Write_Global_Read' and a separate 'Read' operation. The push-buttons inherently have bouncing effects [107] for a short time when immedi-

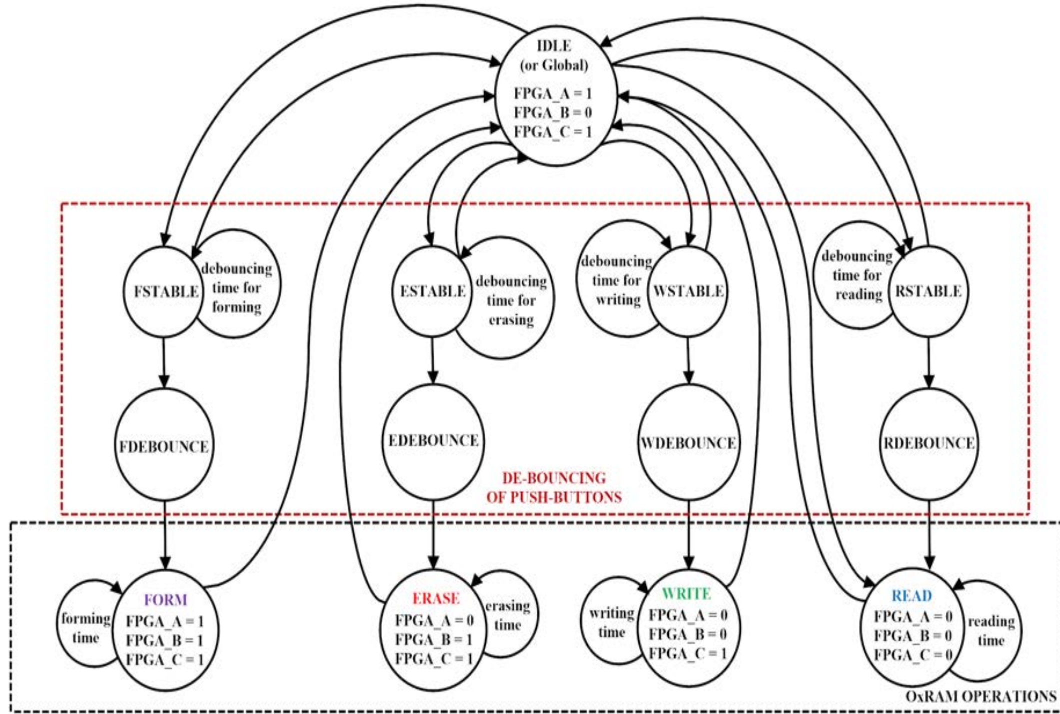


Figure 2-16: FSM used to characterize 1T1R crossbars.

ately pressed and released and these can easily cause short unwanted transients to the crossbar terminals, which can be harmful to OxRAMs and hence they must be eliminated. To overcome this, a software-based debouncing technique is used, which is waiting for some time until bouncing finishes and settles to start the next state.

The characterization of OxRAMs is done in three steps once the FPGA's flash is loaded with the generated programming file via JTAG[14] using the iMPACT™ tool of ISE® design suite[15]. The first step is to use jumpers on button-board to give inputs to the decoders. This is to target the 1T1R device in the crossbar. The second step is to press and release the push-button for the needed operation performed on the

[14]Joint Test Action Group (JTAG) is the IEEE standard 1149.1 that allows users to test all the different interconnects in the FPGA by connecting various integrated circuits, without having to physically probe the connections. This is an advantage when programming board, as this can all be done by software. JTAG makes a boundary scan cell that latches each pin on the device to test the various inputs and outputs. This data is then compared with the expected results from the circuit to find and faults in the interconnects. The biggest advantage of JTAG is that it allows for quicker test times, which is critical when trying to implement designs quickly.

[15]ISE® design suite of Xilinx® is a software tool for synthesis and analysis of HDL designs by enabling developers to synthesize their designs, perform timing analysis, analyze RTL diagrams, simulate the design for different stimuli and generate programming file which is configured to the target device using the iMPACT™ tool.

Table 2.2: Bias conditions used for characterizing a single 1T1R device in the crossbars.

OxRAM operation	V_{TS}	V_{GS}	Pulse-width
‘Form’	4.8 V	1.1 V	10 µs
‘Erase’	- 3 V	4.8 V	400 ns
‘Write’	2.4 V	1.5 V	400 ns
‘Read’	0.3 V	3.8 V	(in the range of µs to ms)
‘Idle’ or ‘Global’	0 V	0 V	(in the range of µs to ms)

targeted OxRAM. The third step is to determine the resistance state of the OxRAM from the ‘Read’ operation performed in the previous step.

Fig. 2-17 shows the active and default biases of the rows, columns, and gate-terminals when the sequential operation— ‘Form_Global_Read’ is performed on a targeted OxRAM in a 4 × 4 crossbar. Fig. 2-18 shows the active and default biases of the rows, columns, and gate-terminals when the sequential operation— ‘Erase_Global_Read’ is performed on a targeted OxRAM in a 4 × 4 crossbar. Fig. 2-19 shows the active and default biases of the rows, columns, and gate-terminals when the sequential operation— ‘Write_Global_Read’ is performed on a targeted OxRAM in a 4 × 4 crossbar. Fig. 2-20 shows the active and default biases of the rows, columns, and gate-terminals when the sequential operation— ‘Read’ is performed on a targeted OxRAM in a 4 × 4 crossbar.

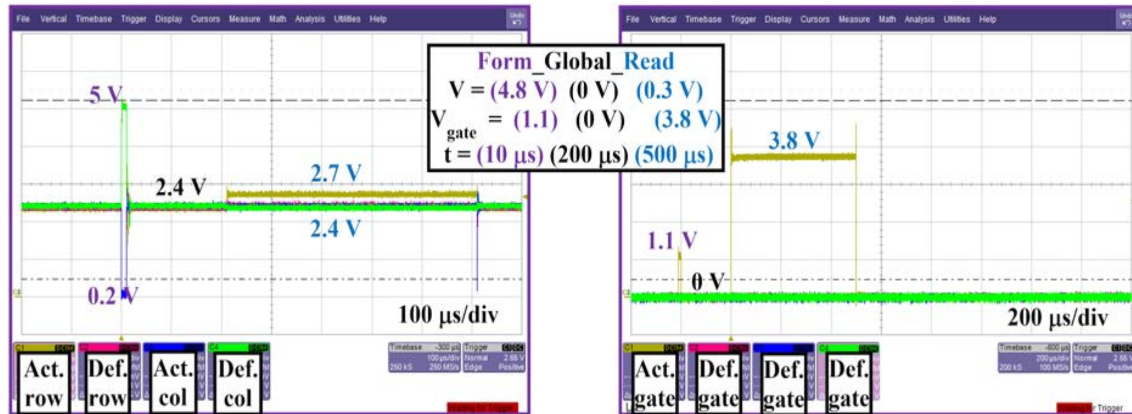


Figure 2-17: Default and active biases (in the form of pulses) applied across rows, columns and gates of the 4 × 4 crossbar for the sequential OxRAM operations- ‘Form_Global_Read’.

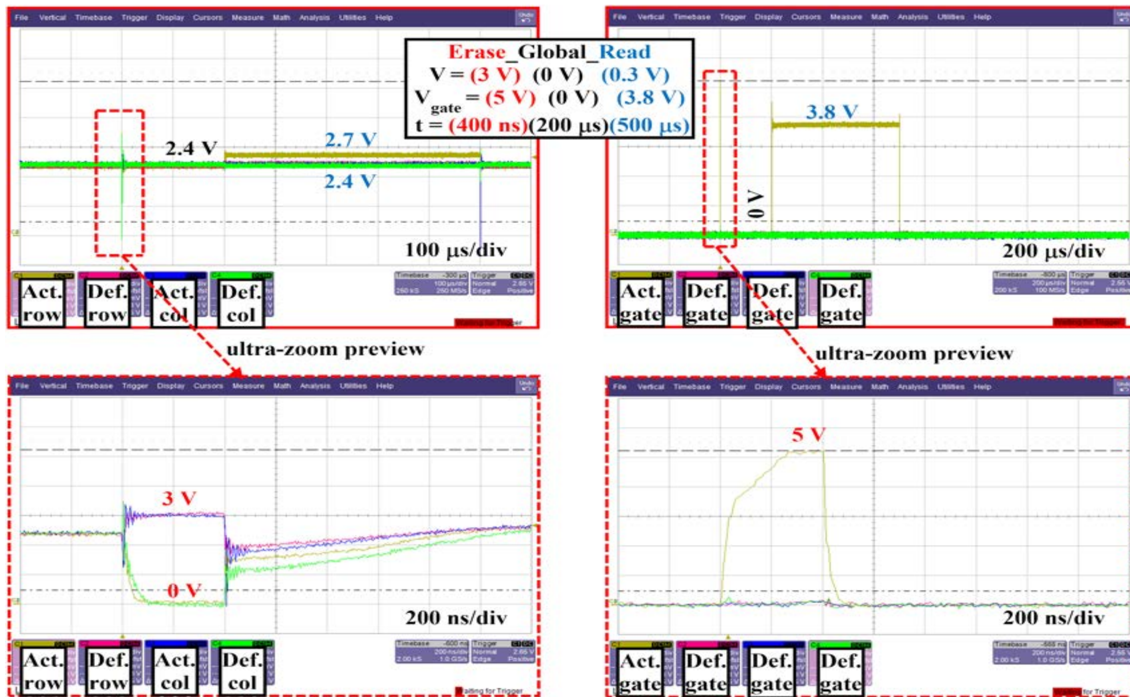


Figure 2-18: Default and active biases (in the form of pulses) applied across rows, columns and gates of the 4 × 4 crossbar for the sequential OxRAM operations- 'Erase_Global_Read'.

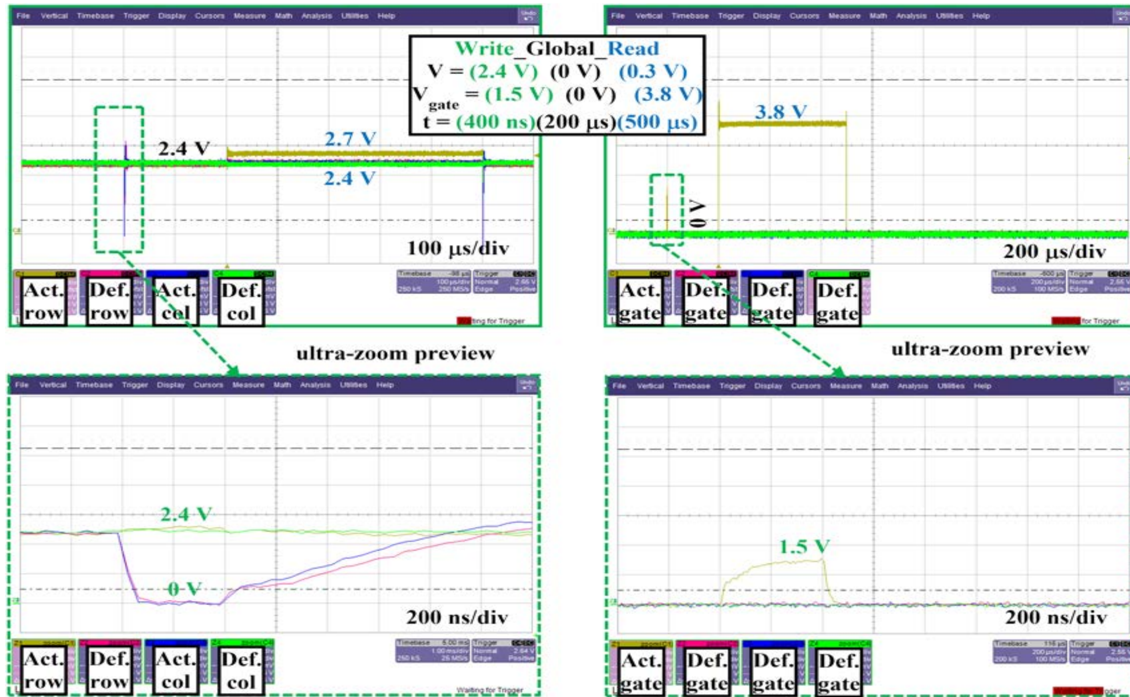


Figure 2-19: Default and active biases (in the form of pulses) applied across rows, columns and gates of the 4 × 4 crossbar for the sequential OxRAM operations- 'Write_Global_Read'.

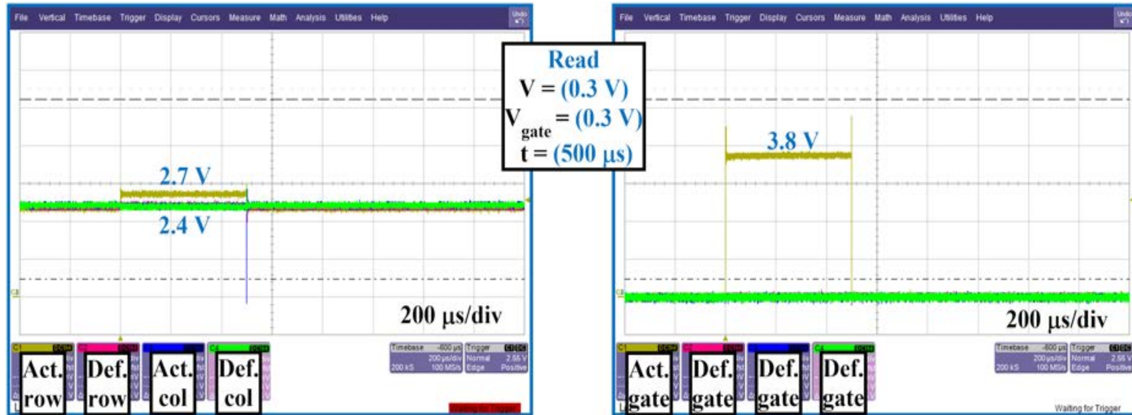


Figure 2-20: Default and active biases (in the form of pulses) applied across rows, columns and gates of the 4 × 4 crossbar for the sequential OxRAM operations- 'Read'.

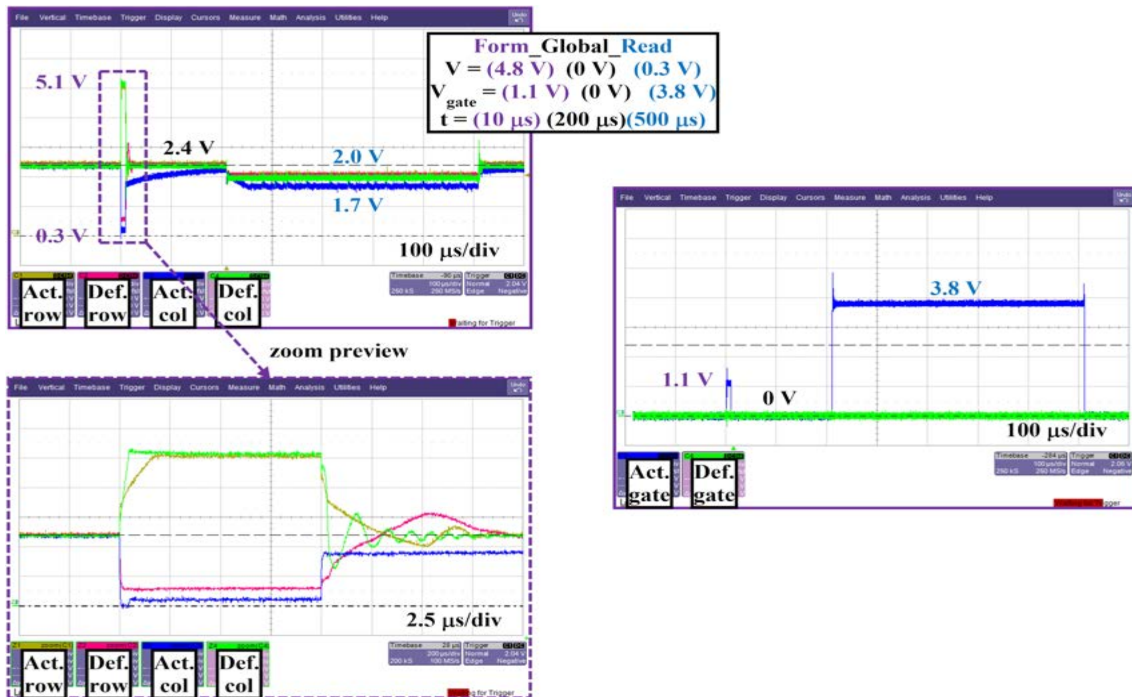


Figure 2-21: Default and active biases (in the form of pulses) applied across rows, columns and gates of the 8 × 8 crossbar for the sequential OxRAM operations- 'Form_Global_Read'.

Similarly, Fig. 2-21 shows the active and default biases of the rows, columns, and gate-terminals when the sequential operation— 'Form_Global_Read' is performed on a targeted OxRAM in a 8 × 8 crossbar. Fig. 2-22 shows the active and default biases of the rows, columns, and gate-terminals when the sequential operation— 'Erase_Global_Read' is performed on a targeted OxRAM in a 8 × 8 crossbar.

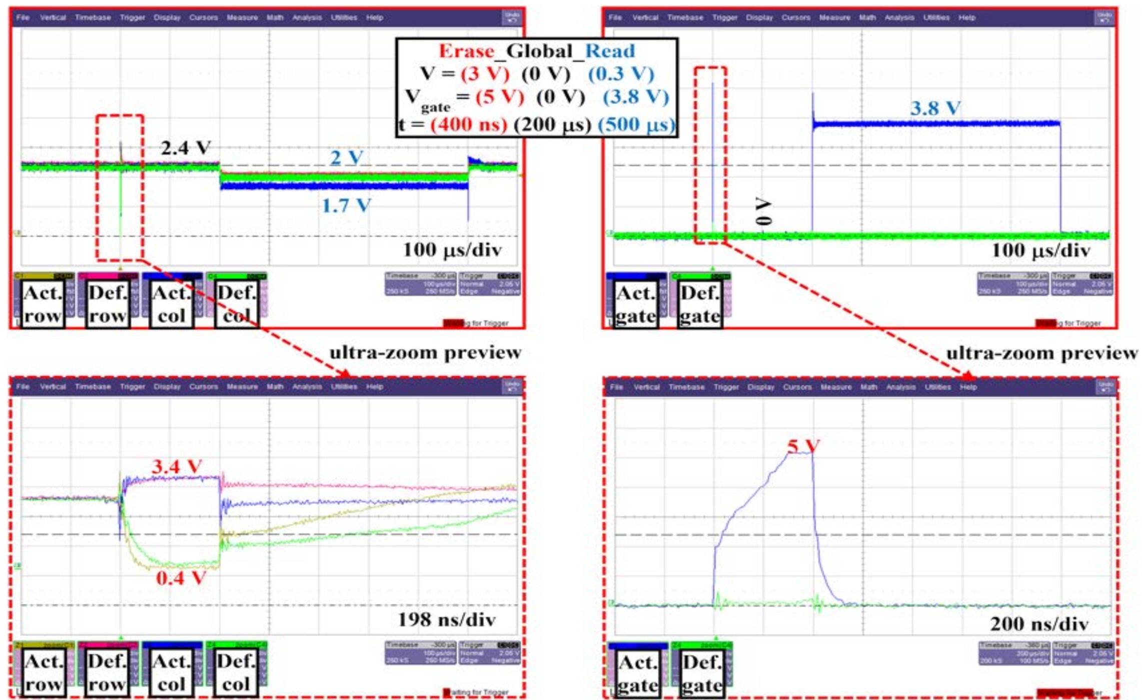


Figure 2-22: Default and active biases (in the form of pulses) applied across rows, columns and gates of the 8 × 8 crossbar for the sequential OxRAM operations- 'Erase_Global_Read'.

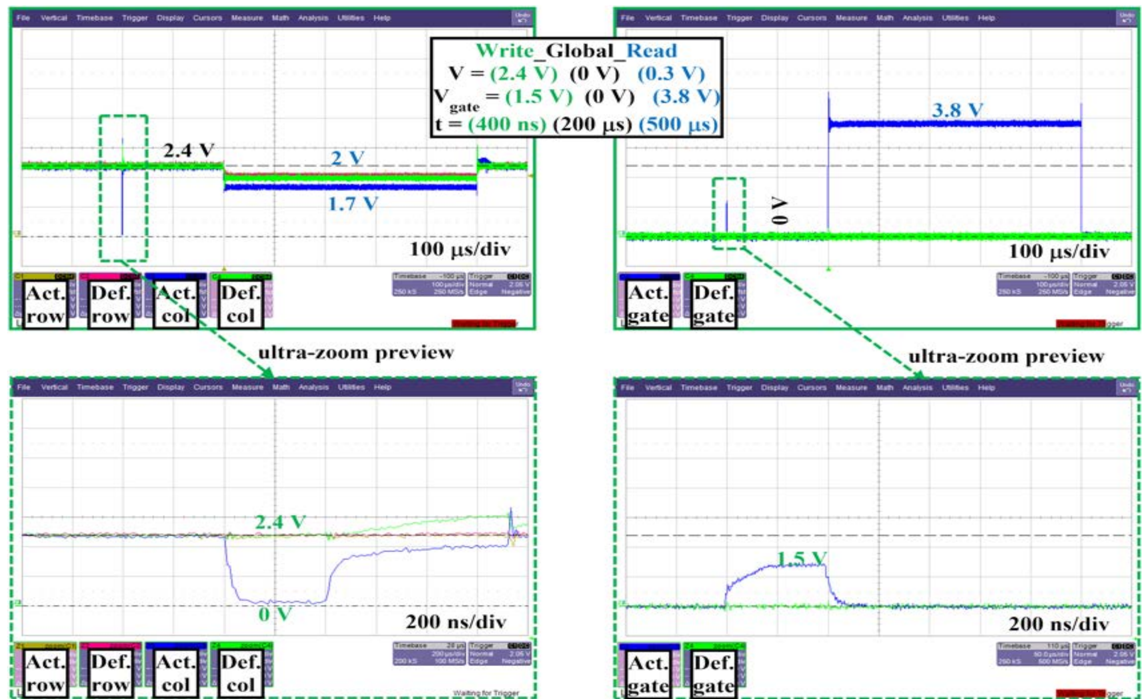


Figure 2-23: Default and active biases (in the form of pulses) applied across rows, columns and gates of the 8 × 8 crossbar for the sequential OxRAM operations- 'Write_Global_Read'.

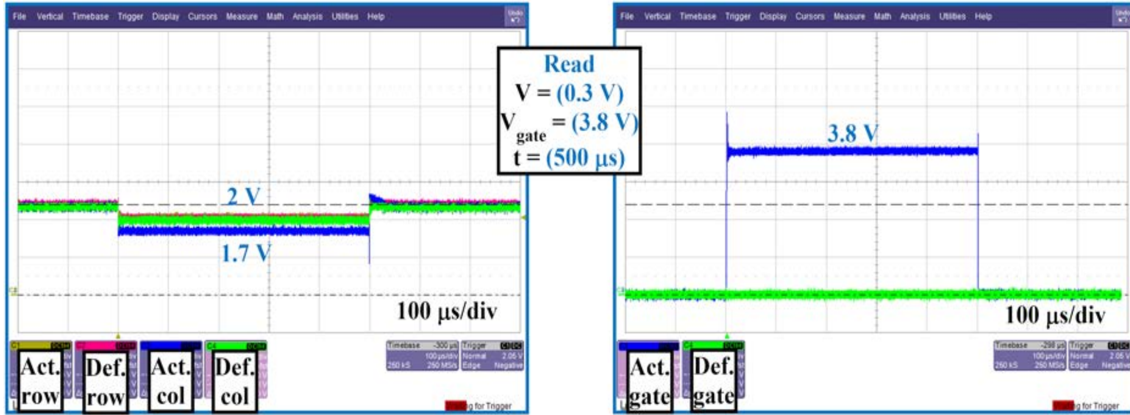


Figure 2-24: Default and active biases (in the form of pulses) applied across rows, columns and gates of the 8 × 8 crossbar for the sequential OxRAM operations- 'Read'.

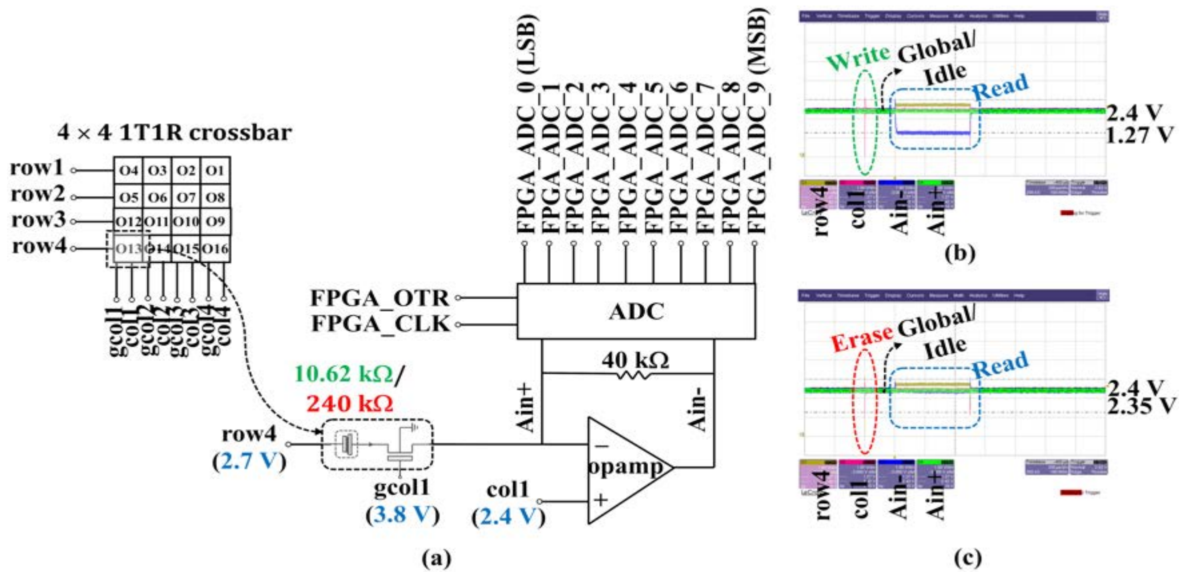


Figure 2-25: Read operation in a single 1T1R device in the 4 × 4 crossbar: (a) Read scheme and biases showing a targeted 1T1R device in a 4 × 4 crossbar, (b) Observation of terminals of the 'Read' scheme in oscilloscope after a 'Write_Global_Read' operation performed on the targeted 1T1R device, (c) Observation of terminals of the read scheme in oscilloscope after an 'Erase_Global_Read' operation performed on the targeted 1T1R device.

Fig. 2-23 shows the active and default biases of the rows, columns, and gate-terminals when the sequential operation— 'Write_Global_Read' is performed on a targeted OxRAM in a 8 × 8 crossbar. Fig. 2-24 shows the active and default biases of the rows, columns, and gate-terminals when the sequential operation— 'Read' is performed on a targeted OxRAM in a 8 × 8 crossbar. In all the above figures that

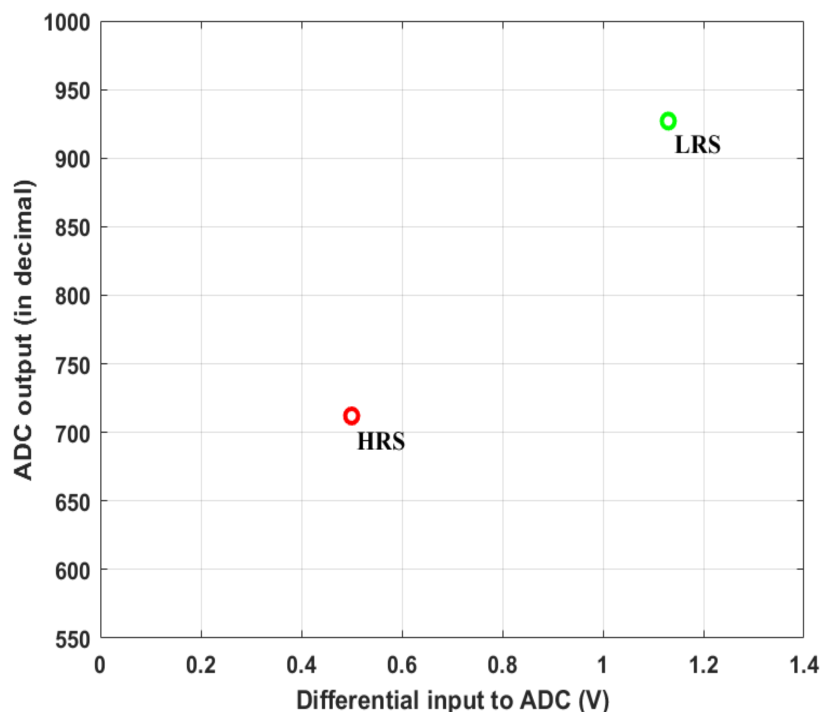


Figure 2-26: Output of ADC Vs differential input of ADC for the whole resistance range thereby, highlighting both LRS and HRS of OxRAM.

shows the biases (in the form of pulses) needed for different OxRAM operations, the amplitudes and pulse-widths needed are duly labeled on them. The bias conditions for various OxRAM operations are shown in Table 2.2, whose values vary slightly from the one recommended by Leti.

The read procedure can be done in two ways. One is by directly observing the differential inputs of the ADC and another way is to check the output of ADC. Fig. 2-25 shows both the read approaches. Here, Fig. 2-25 (a) shows the 'Read' scheme implemented where the targeted 1T1R device is operated with a 'Read' voltage of 0.3 V. The 'Read' task is performed via an opamp and an ADC in order to get the resistance of the OxRAM.

When a 'Write_Global_Read' operation is performed with a 'Read' voltage of 0.3 V, the differential input to ADC becomes 2.4 V - 1.27 V = 1.13 V. This results in,

$$LRS = (\frac{40\,\mathrm{k\Omega}}{2.4\,\mathrm{V} - 1.27\,\mathrm{V}}) \times 0.3\,\mathrm{V} = 10.62\,\mathrm{k\Omega} \tag{2.1}$$

where, 40 kΩ is the read resistor placed across the read opamp. This resistor across the terminals- Ain+ and Ain- is shown in Fig. 2-25 (a). Similarly, when a 'Erase_Global_Read' operation is performed with a 'Read' voltage of 0.3 V, the differential input to ADC becomes 2.4 V - 2.35 V = 0.5 V. This results in,

$$HRS = (\frac{40\,\mathrm{k\Omega}}{2.4\,\mathrm{V} - 2.35\,\mathrm{V}}) \times 0.3\,\mathrm{V} = 240\,\mathrm{k\Omega} \tag{2.2}$$

Fig. 2-25 (b) and Fig. 2-25 (c) show various biases of the read circuitry when sequential tasks— 'Write_Global_Read' and 'Erase_Global_Read' are performed on a targeted 1T1R device. One can see how— bias of the terminal 'Ain-' changes in both LRS (resistance after the sequential operation- 'Write_Global_Read') and HRS (resistance after the sequential operation- 'Erase_Global_Read') during 'Read' part of the waveform. The second way to determine the resistance state of the targeted OxRAM is to observe the output of ADC, which can be stored in a register in the FPGA for further processing, when doing some learning. The 10-bit ADC is clocked with a 1.25 MHz clock signal and it uses only first 9 bits to quantify the differential input or the resistance of the targeted OxRAM, as the signal of MSB (which is the 10-bit) depends on the sign of the difference of the input signals [108]. The MSB signal– 'FPGA_ADC_9' goes high (3.3 V) when the difference of the ADC's inputs becomes less than zero and goes low (0 V) when the difference of the ADC's inputs becomes equal to or greater than zero. Signal 'FPGA_ADC_OTR' goes low (0 V) when the difference of the ADC's inputs is in between the range of voltage– -1.030 and +1.36 V for the whole input span of about 2 V and it becomes high (3.3 V) when it is out of this range. Fig. 2-26 shows the output of ADC in decimal for various differential input values of ADC where both LRS and HRS are highlighted.

2.3.5 Experimental results of characterization of 1T1R based memristors

The OxRAMs in the crossbar are formed initially so that, they are in LRS. Then they are switched between HRS and LRS. All OxRAM operations like 'form', 'erase', 'write' and 'read' are carried out as explained in Section 3.4.4. Two tests have been carried out. The first test is a single switch between LRS and HRS for all 64 1T1R devices of 8 × 8 crossbar. The results of this soft-test are shown in Fig. 2-27. The second test is carried out for a 1T1R device of 4 × 4 crossbar by switching the device between LRS and HRS for 10^3 cycles. The switching characteristics of this hard-test are shown in Fig. 2-28. Fig. 2-29 and Fig. 2-30 shows the cycle-to-cycle variability results for LRS and HRS for 10^3 cycles. It is observed from the variability results that standard deviation of HRS, σ_{HRS} is high and as cycle number increases, the device degrades with HRS approaching LRS. The variability results of LRS appear fine with a σ_{HRS} of 4.4 kΩ and μ_{LRS} of 22 kΩ.

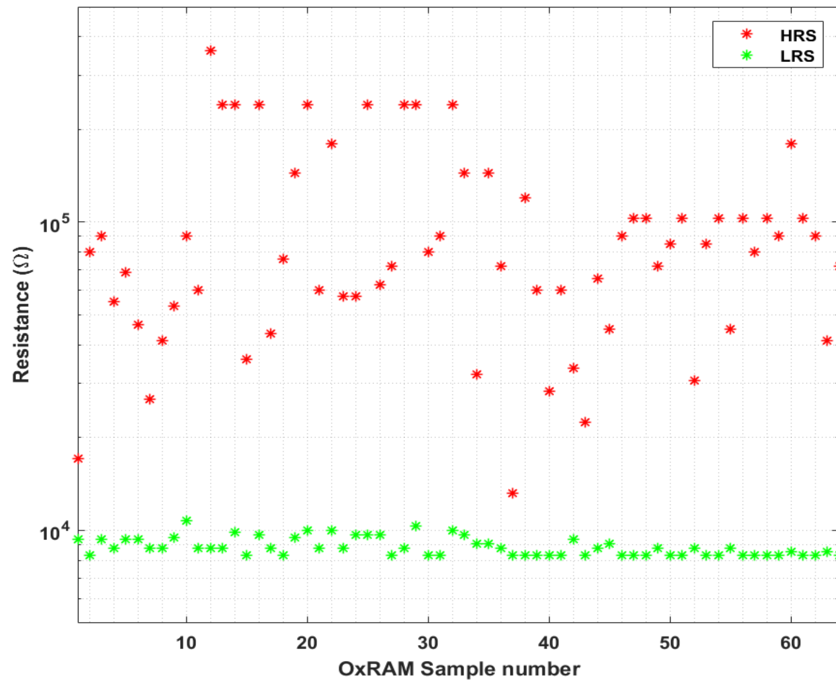


Figure 2-27: HRS and LRS values of OxRAMs of all 64 1T1R devices in 8 × 8 crossbar.

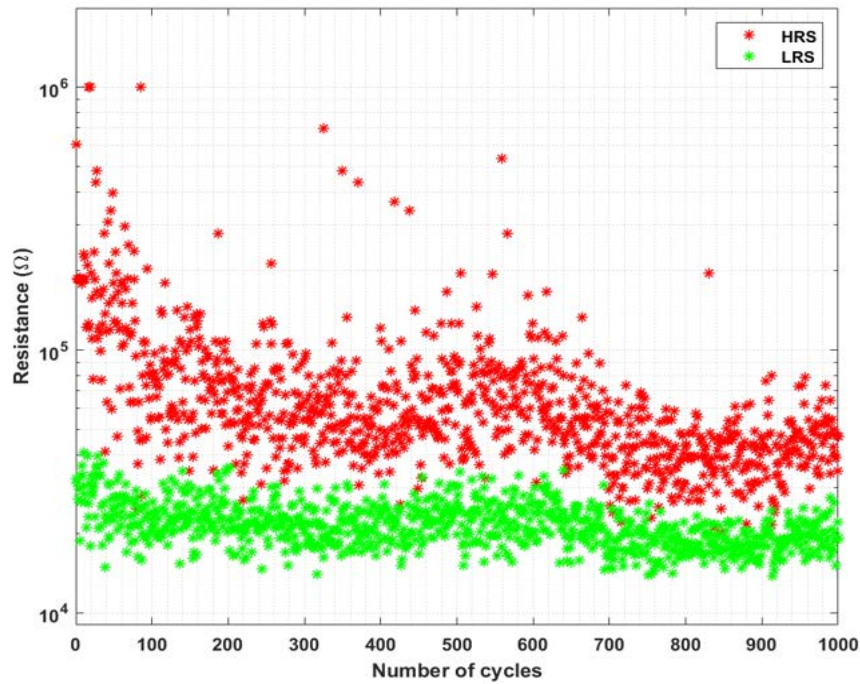


Figure 2-28: Switching of resistance of OxRAM of a 1T1R device between HRS and LRS for 10^3 cycles.

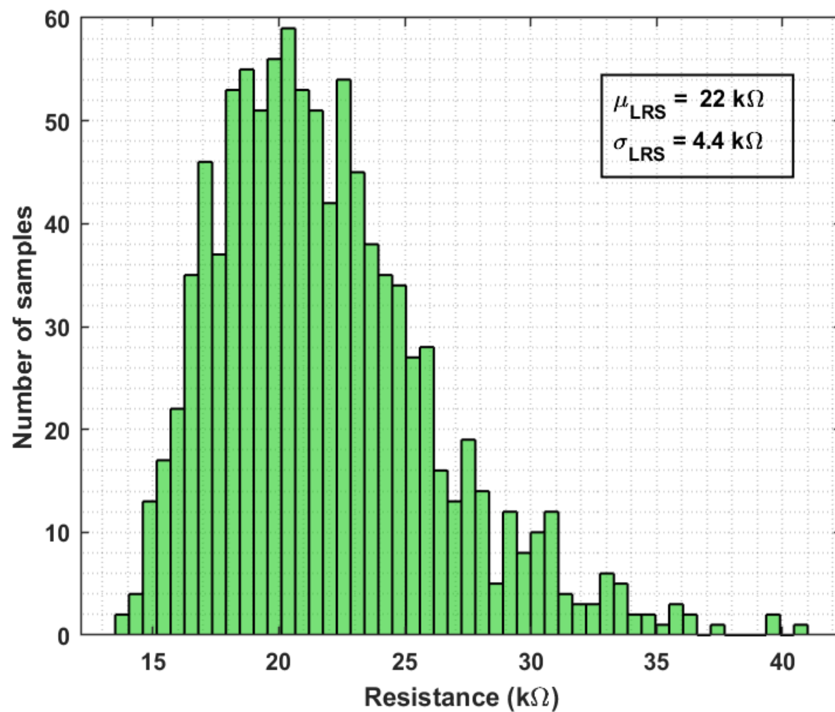


Figure 2-29: Cycle-to-cycle variability results of LRS for 10^3 switching cycles.

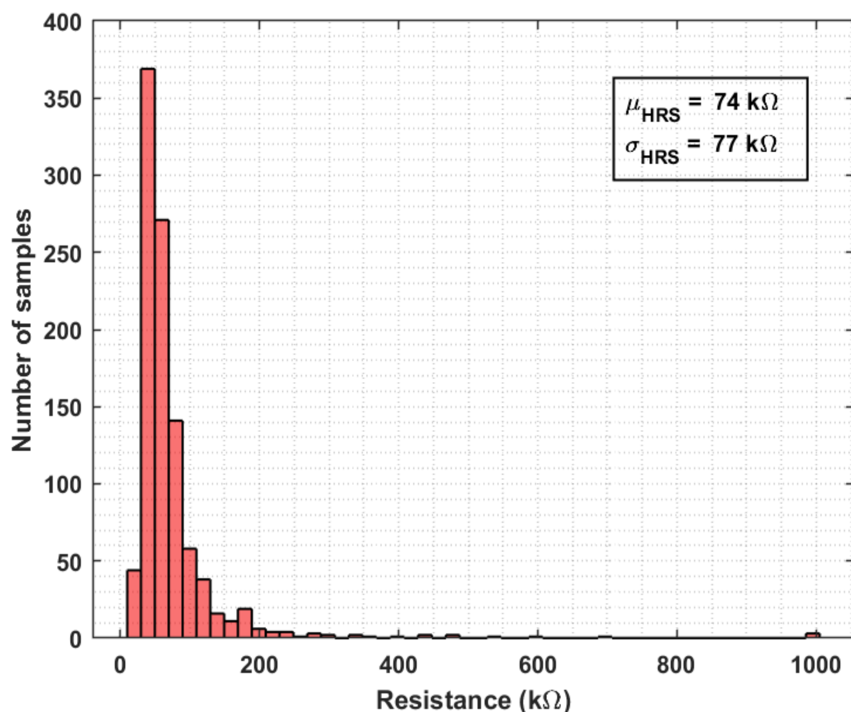


Figure 2-30: Cycle-to-cycle variability results of HRS for 10^3 switching cycles.

2.4 Characterization of MIM-based memristors using SPA

Wafer-level I-V characteristic-measurements are performed on MIM based memristors. The MIM based memristors are built on silicon oxide of thickness 200 nm. Above it, a 200 nm thick layer of Tungsten (W) is built, which acts as the bottom electrode. A dielectric layer is deposited above it, whose thickness and composition depend on the ALD process used. There's a TiN layer of thickness 200 nm built on top of the dielectric, which acts as the top electrode. Structurally, MIM based memristors are three-layered (as shown in Fig. 2-31), while OxRAMs are 4 layered. When it comes to the fabrication process, 1T1R based memristors are taped out in a complex monolithic integration by growing OxRAMs above the CMOS part while MIM based memristors are grown in layers using Atomic Layer Deposition (ALD) process.

ALD process- typically uses two chemicals to create an alternate, saturated, chem-

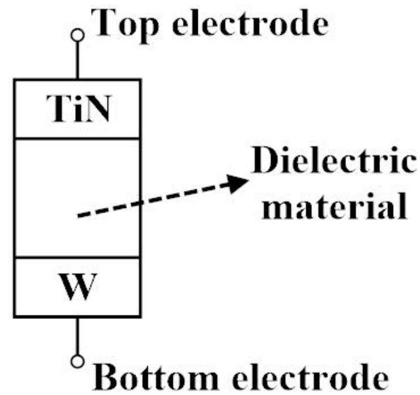


Figure 2-31: Structure of a MIM based memristor.

ical reactions on the surface, resulting in a unique self-limiting growth with many excellent features like conformality, uniformity, repeatability and accurate thickness control [109]. These chemicals (precursors) do not exist in the gas phase at the same time as typical CVD processes. Rather, precursors are pulsed sequentially in an inert carrier gas through a heated batch of substrates, with a purge time between the pulses to prevent vapor phase reactions. ALD is, therefore, suitable for the synthesis of thin solid layers of inorganic materials even as thin as one molecular monolayer. Savannah 200 is the equipment used for ALD and it is shown in Fig. 2-32.

Figure 2-32: Savannah 200 from Cambridge NanoTech Inc. [109].

2.4.1 Experimental set-up

In our experiments, we plan to do wafer-level I-V measurements for nine wafers, whose dielectric compositions are obtained from different ALD processes, which are shown in Table 2.3. The I-V characteristics are obtained on 15 × 15 µm^2 sized MIM based memristors on all nine wafers. Fig. 2-33 shows the wafer map with a preview of a chip and a zoom preview of 5 × 5 µm^2 and 15 × 15 µm^2 sized MIM based memristors.

Table 2.3: Different dielectrics used in ALD process

Wafer no.	**ALD process**	**Thickness (nm)**
W_1	HfO_2 standard process	10.6±0.18
W_2	HfO_2 process with short H_2O pulse	10.5±0.2
W_3	HfO_2 process with large N_2 flow rates	10.2±0.14
W_4	[HfO_2 (6 cycles) + Al_2O_3 (1 cycle)]× 14 + HfO_2 (4 cycles)	11.6±0.08
W_5	Al_2O_3 process at 225°C	99±0.04
W_6	[Al_2O_3 (6 cycles) + HfO_2 (1 cycle)]× 14 + Al_2O_3 (4 cycles)	12.7±0.14
W_7	[Al_2O_3 (16 cycles) + HfO_2 (75 cycles)]× 14 + Al_2O_3 (16 cycles)	13±0.1
W_8	[HfO_2 75 cycles) + Al_2O_3 (16 cycles)	11.9±0.08
W_9	HfO_2 process at 150°C	10.5±0.2

The characterization experiments are performed mainly using the MPITS2000 probe station[16] [110] and HP4155B SPA. Placing the wafer in MPITS2000 probe station is done in three steps. At first, the station is initiated and the wafer is placed on the chuck either by front or loading from side thereby, turning on the vacuum. The second step is to train the contact needles and make two-point alignment of the wafer. The third step is to establish contact on the designed structures using needles for taking measurements. All these are done using the MPI Sentio®[17] software suite. The tri-axial cables connect both the probes of MPITS2000 and the SMUs of

[16]MPITS2000-SE from MPI is the first ever 200 mm automated probe system. The probe station is known for its ultra-low noise, very accurate and highly reliable DC/CV, RF and high power measurements.

[17]MPI Sentio® is the multi-touch prober control software suite used to probe wafer in MPITS2000 probe station. The GUI of Sentio® is used to make precise alignment of wafers and establish contacts for measurement.

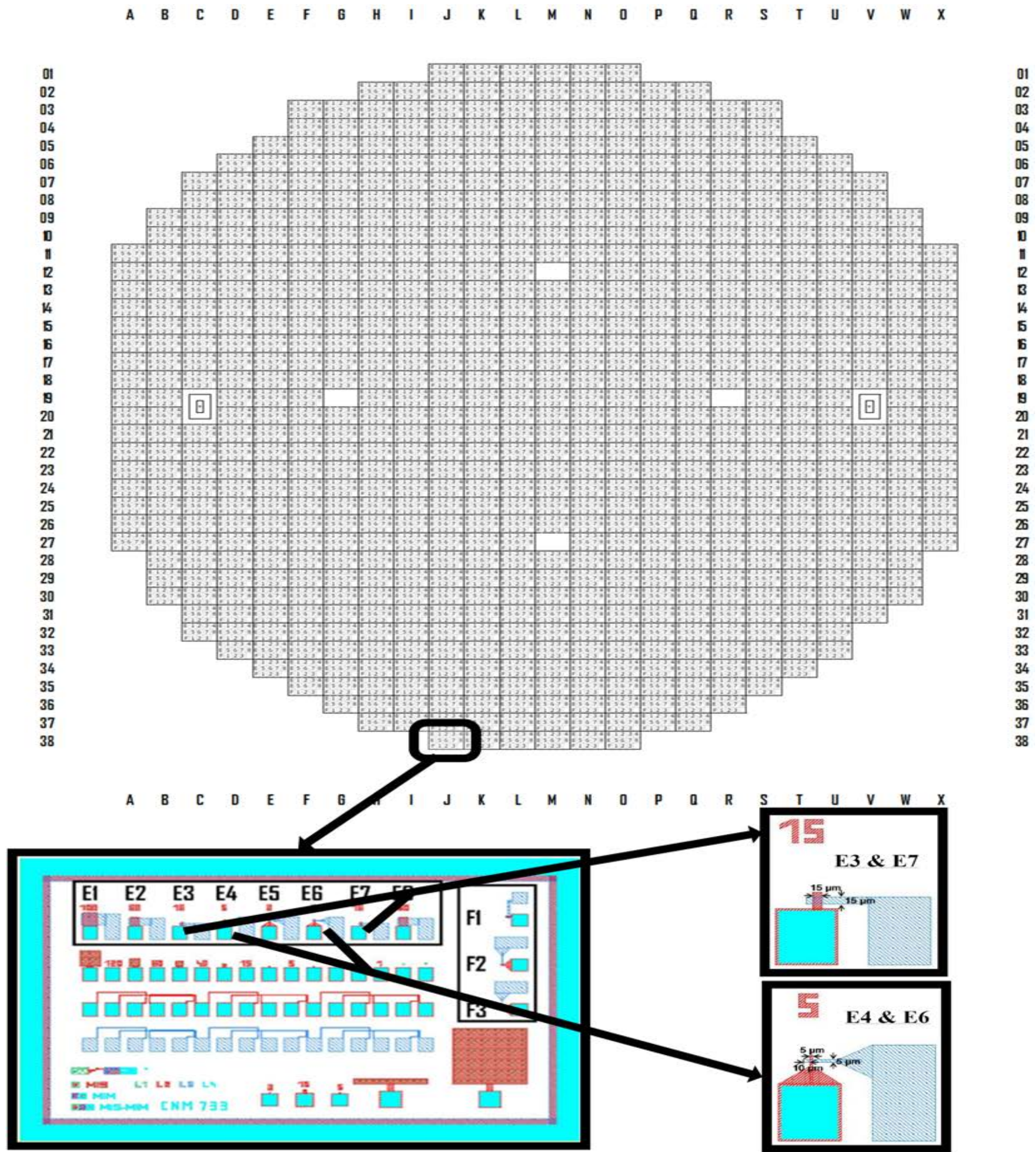


Figure 2-33: Wafer map showing a preview of a single chip with zoom previews of 5 × 5 µm^2 and 15 × 15 µm^2 sized MIM based memristors.

HP4155B SPA. The experimental set-up is shown in Fig. 2-34. One can perform experiments directly using the front panel of SPA or using MATLAB scripts running in a computer by connecting it to SPA by GPIB. Here, the experiments are carried out using MATLAB scripts running in a computer by connecting it to SPA by GPIB.

Once the needles are placed on the pads of the MIM structured memristors, DC voltage sweeps are done to characterize the memristors. A full cycle of switching

between LRS and HRS is attempted on devices in wafers- W_1, W_2, W_3 and W_4, whose I-V characteristics are plotted. Initially, when a positive voltage sweep is done from 0 to V_{max}, at a certain voltage (called as V_{BD}) the dielectric filament breakdown and conduction occurs. This breakdown voltage, V_{BD} depends on the dielectric composition and thickness used in the ALD process. Memristors that are thicker or with Al_2O_3 dielectric need higher V_{BD} for the dielectric to breakdown.

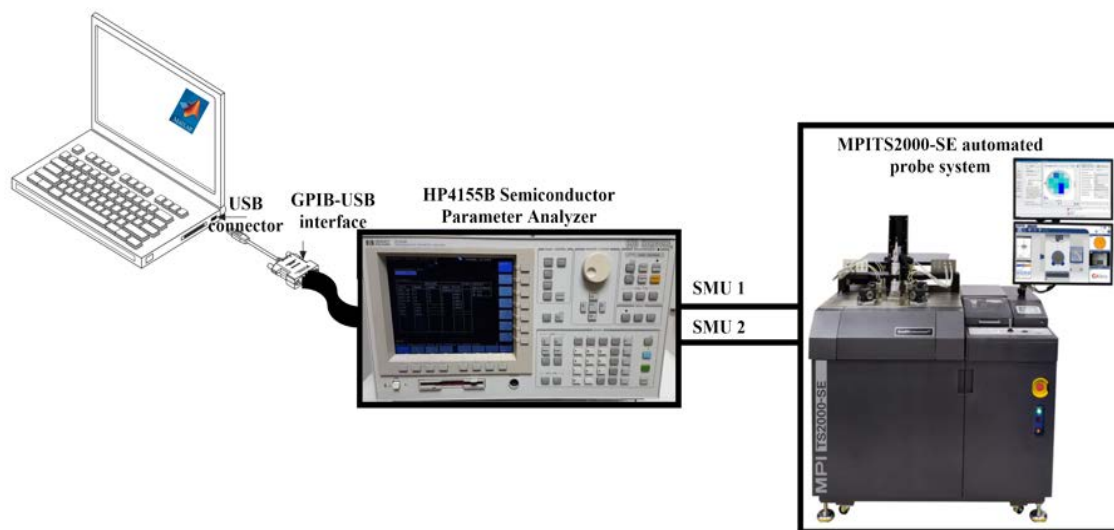


Figure 2-34: Experimental set-up for characterization of MIM based memristors.

When a negative voltage is swept from 0 to V_{min}, at a certain voltage (called as V_{reset}) the current drops. This corresponds to the change of resistance from LRS to HRS and this operation is called 'Reset' or 'Erase'. When a positive voltage sweep is done from 0 to V_{max}, at a certain voltage (called as V_{set}) the current suffers an increase than at lower voltages. This corresponds to the change of resistance from HRS to LRS and this operation is called 'Set' or 'Write'. When we apply the positive and negative voltage sweeps alternatively in a sequence, the first positive voltage sweep will results in V_{BD} (by causing a breakdown of filament) and then the negative voltage results in V_{reset}, followed by V_{set}, V_{reset}, V_{set}, V_{reset} and so on, depending on the number of cycles of the voltage sweeps. The resistance state of the device is read at a low voltage of V_{read} = -0.1 V. A forming compliance current of 1 mA is kept and a compliance current of 100 mA is kept for both positive and negative voltage sweeps. A fixed number of steps of 180 is used in both voltage sweeps.

2.4.2 Characterization results of MIM-based memristors

The I-V characteristic measurements performed on MIM based memristors on wafers- W_1, W_2, W_3, and W_4 are plotted. These figures are shown in Fig. 2-35, Fig. 2-36, Fig. 2-37 and Fig. 2-38. The sub-figures in them show separately the I-V characteristics during the breakdown of filament, I-V characteristics when switching between HRS and LRS and currents- I_{LRS} and I_{HRS} for a read voltage of V_{read} = -0.1 V during 19 full cycles of switching between HRS and LRS. Experimental results taken from first 4 wafers reveal that— wafers that have Al_2O_3 needed a lower V_{min} for resetting the device when compared to other wafers. The comparison of V_{max}, V_{min}, V_{set}, and V_{reset} are shown in Table 2.4.Unlike 1T1R devices, MIM based memristors need low voltage levels for switching. Due to this, MIM based memristors can be future potential candidates for its integration with deep nanometer CMOS technology.

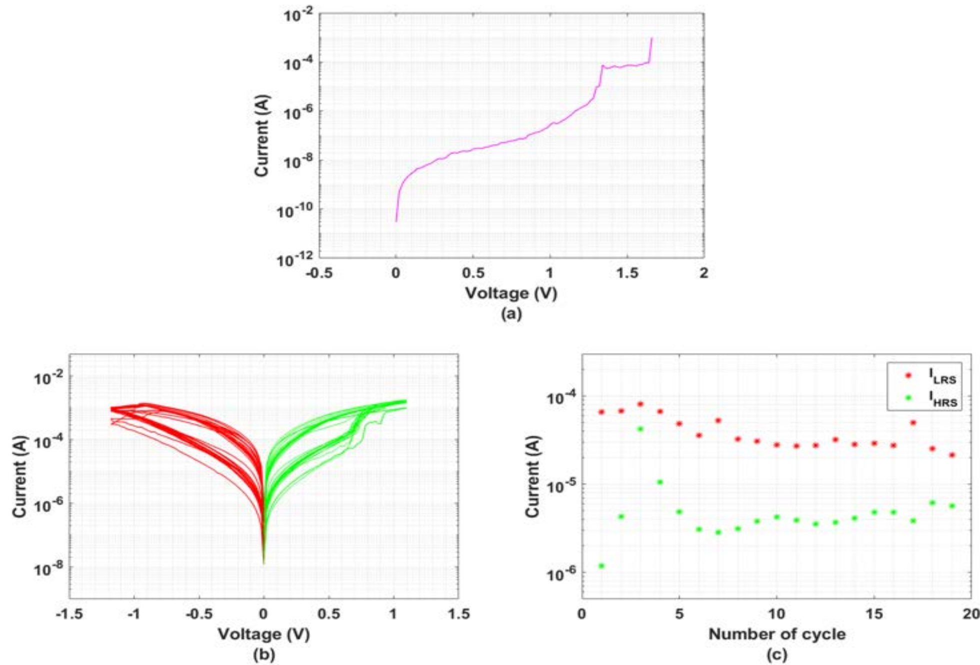


Figure 2-35: Experimental results of a 15 × 15 µm^2 sized MIM based memristor in wafer, W_1: (a) I-V characteristics during a breakdown attempt, (b) I-V characteristics during 19 full-cycles of switching between HRS and LRS, (c) Currents- I_{LRS} and I_{HRS} for a read voltage of V_{read} = -0.1 V during 19 full-cycles of switching between HRS and LRS.

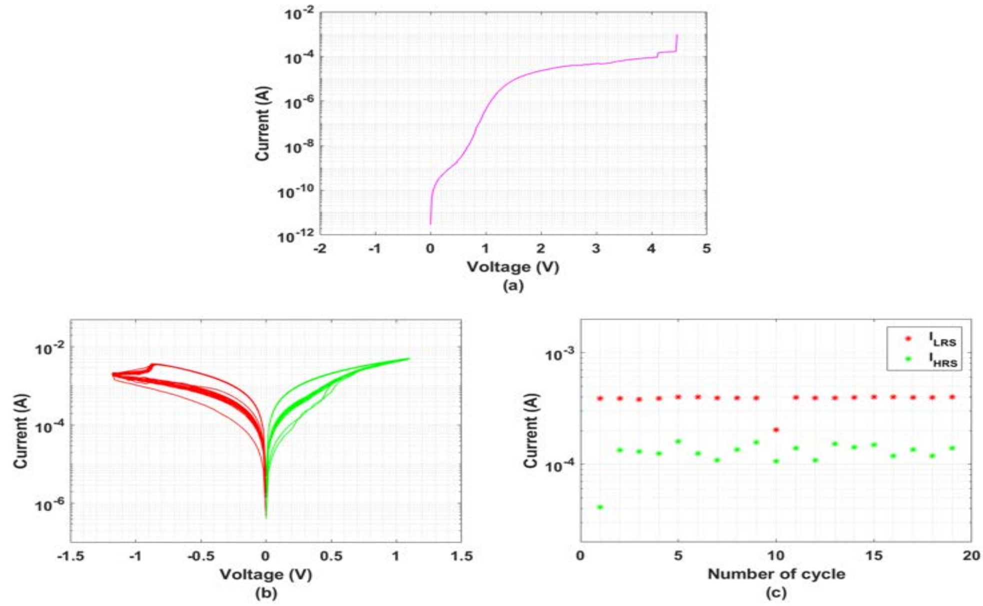


Figure 2-36: Experimental results of a 15 × 15 µm^2 sized MIM based memristor in wafer, W$_2$: (a) I-V characteristics during a breakdown attempt, (b) I-V characteristics during 19 full-cycles of switching between HRS and LRS, (c) Currents- I$_{LRS}$ and I$_{HRS}$ for a read voltage of V$_{read}$ = -0.1 V during 19 full-cycles of switching between HRS and LRS.

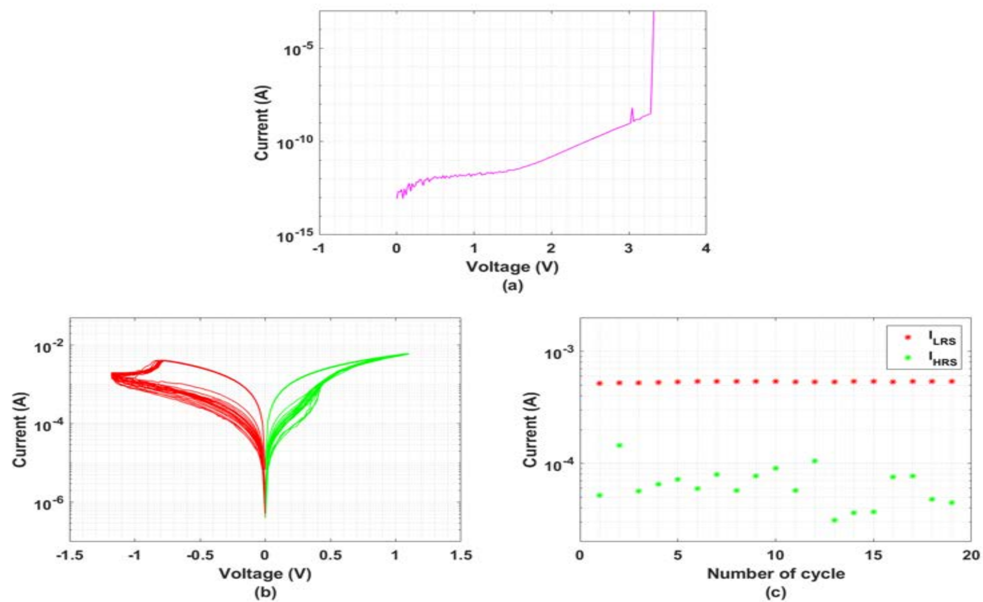


Figure 2-37: Experimental results of a 15 × 15 µm^2 sized MIM based memristor in wafer, W$_3$: (a) I-V characteristics during a breakdown attempt, (b) I-V characteristics during 19 full-cycles of switching between HRS and LRS, (c) Currents- I$_{LRS}$ and I$_{HRS}$ for a read voltage of V$_{read}$ = -0.1 V during 19 full-cycles of switching between HRS and LRS.

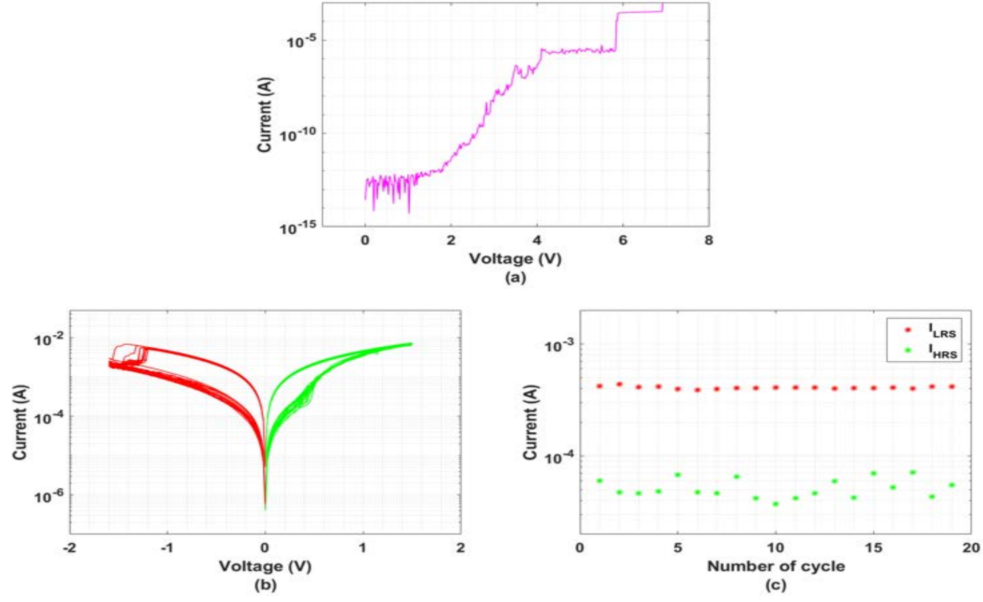


Figure 2-38: Experimental results of a 15 × 15 µm^2 sized MIM based memristor in wafer, W_4: (a) I-V characteristics during a breakdown attempt, (b) I-V characteristics during 19 full-cycles of switching between HRS and LRS, (c) Currents- I_{LRS} and I_{HRS} for a read voltage of V_{read} = -0.1 V during 19 full-cycles of switching between HRS and LRS.

Table 2.4: Comparison of V_{max}, V_{min}, V_{set} and V_{reset} for wafers- W_1, W_2, W_3 and W_4.

Wafer no.	V_{max} (V)	V_{min} (V)	V_{set} (V)	V_{reset} (V)
W_1	1.2	-1.2	0.7	-0.9
W_2	1.2	-1.2	0.5	-0.88
W_3	1.2	-1.2	0.4	-0.8
W_4	1.5	-1.6	0.45	-1.2

Chapter 3

Bulk-based three-stage DC offset Calibration Scheme for Memristive Crossbar

3.1 Need for offset calibration

Typical neuromorphic circuits based on the use of OxRAM devices comprise an m×n 1T1R array, also referred to as 'crossbar', where 1T1R synapses are used as programmable interconnecting elements. When each wordline in the crossbar is simultaneously driven by inference spike pulses, the full memristor array can become active. This makes a majority of them drive a few mA in their LRS, which makes power dissipation severe. This can easily limit the maximum crossbar size and the driving capability of the peripheral circuit [111].

An approach is to reduce the array power consumption by limiting the read voltage, V_{read} applied to the devices. Fig. 3-1 shows the simulated OxRAM currents for read voltage pulses less than 1 V for nominal LRS (13.65 kΩ) and HRS (836.4 kΩ) values. In order to investigate how small one can make the inference read voltage pulses (10 mV, 1 mV or even less), the crossbar lines need to be set to an identical voltage level with an inter-line error lower than the pulse amplitudes. Keeping the

read pulse amplitude little becomes non-trivial as the offset voltages of the system affect the results. To overcome this problem, the opamps used, either in the neuron circuits or the buffers need to be finely calibrated to reduce their input DC offset voltage, which ultimately sets the resting voltage level of the crossbar lines higher than electrical noise.

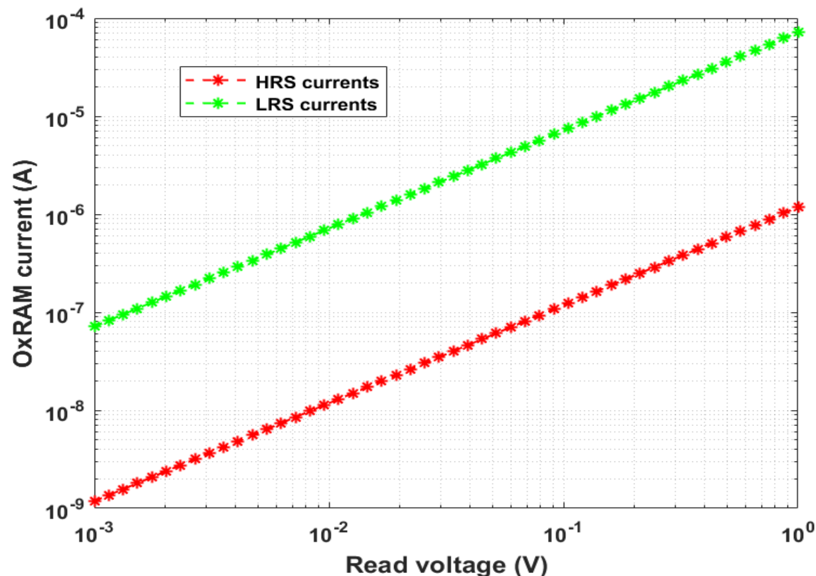


Figure 3-1: OxRAM currents for low read voltage pulses.

3.2 Three-stage bulk-based DC offset calibration approach

As conventional calibration schemes can compensate for offset ranges in the order of a few mV [112], we propose a finer calibration technique. Our proposed calibration scheme is based on compensating the DC offset by varying the bulk voltage of one of the transistors that form the input differential pair of the amplifiers. To this end, a cascade of resistor ladders is used, which allows us to increase the accuracy of the reference voltage, yielding to a calibration step lower than 0.1 mV. The calibration scheme is implemented on a 4×4 1T1R crossbar, whose conceptual diagram is shown in fig. 3-2.

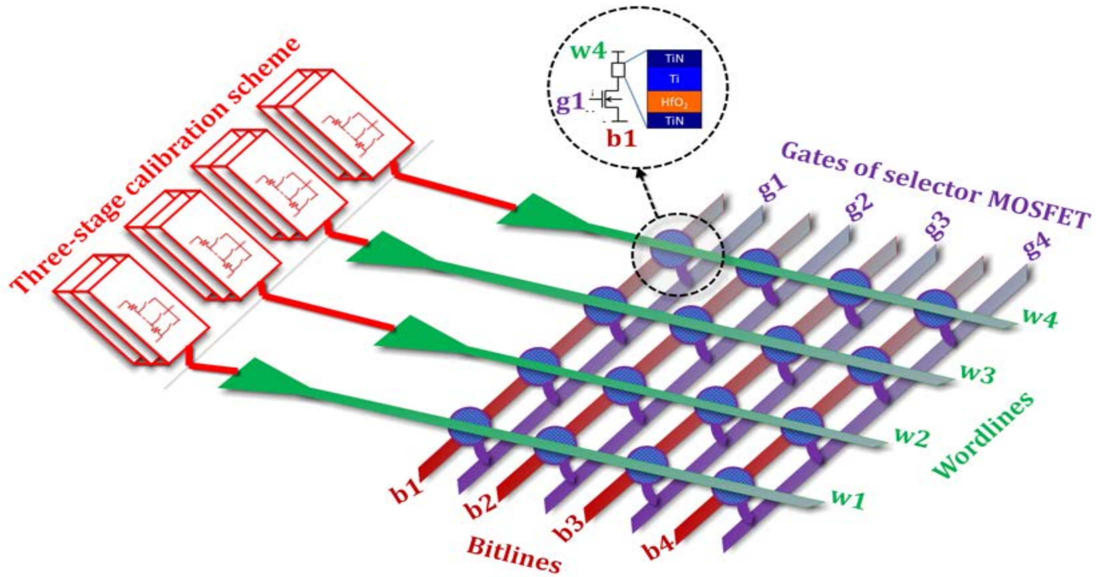


Figure 3-2: Conceptual diagram of the proposed three-stage calibration scheme in the 4 × 4 1T1R crossbar.

3.3 Design of 1T1R crossbar with three-stage DC offset calibration scheme

Fig. 3-3 shows the scheme of 1T1R synapses used as programmable interconnects in a 4×4 crossbar with the circuitry of DC offset voltage calibration in each wordline. Each wordline has its pre-synaptic driver and each pre-synaptic driver has its opamp (used in buffer configuration), an I-pot (a current bias circuit for opamp) and a three-stage DC offset calibration circuit. There is an edge-triggered D-flip flop based shift register to load control-bits for the current-bias circuit and the calibration circuit. There are 16 1T1R structures arranged in a 4×4 matrix, whose wordlines- $w_{1,2,3,4}$ are connected to the output of their corresponding pre-synaptic driver. The gates, $g_{1,2,3,4}$ are pulled out bitlinewise and a post-synaptic neuron follows the source of the NMOS in each bitline of the 1T1R array. The main sub-circuits used to implement the bulk-based calibration scheme in a 4×4 1T1R crossbar are the two-stage PMOS-based differential opamp, pulse-shaping digital block, three-stage calibration scheme, I-pots, shift-register, and post-synaptic drivers. The design details of all these sub-circuits are explained in the forthcoming sub-sections of this chapter, which follows additional

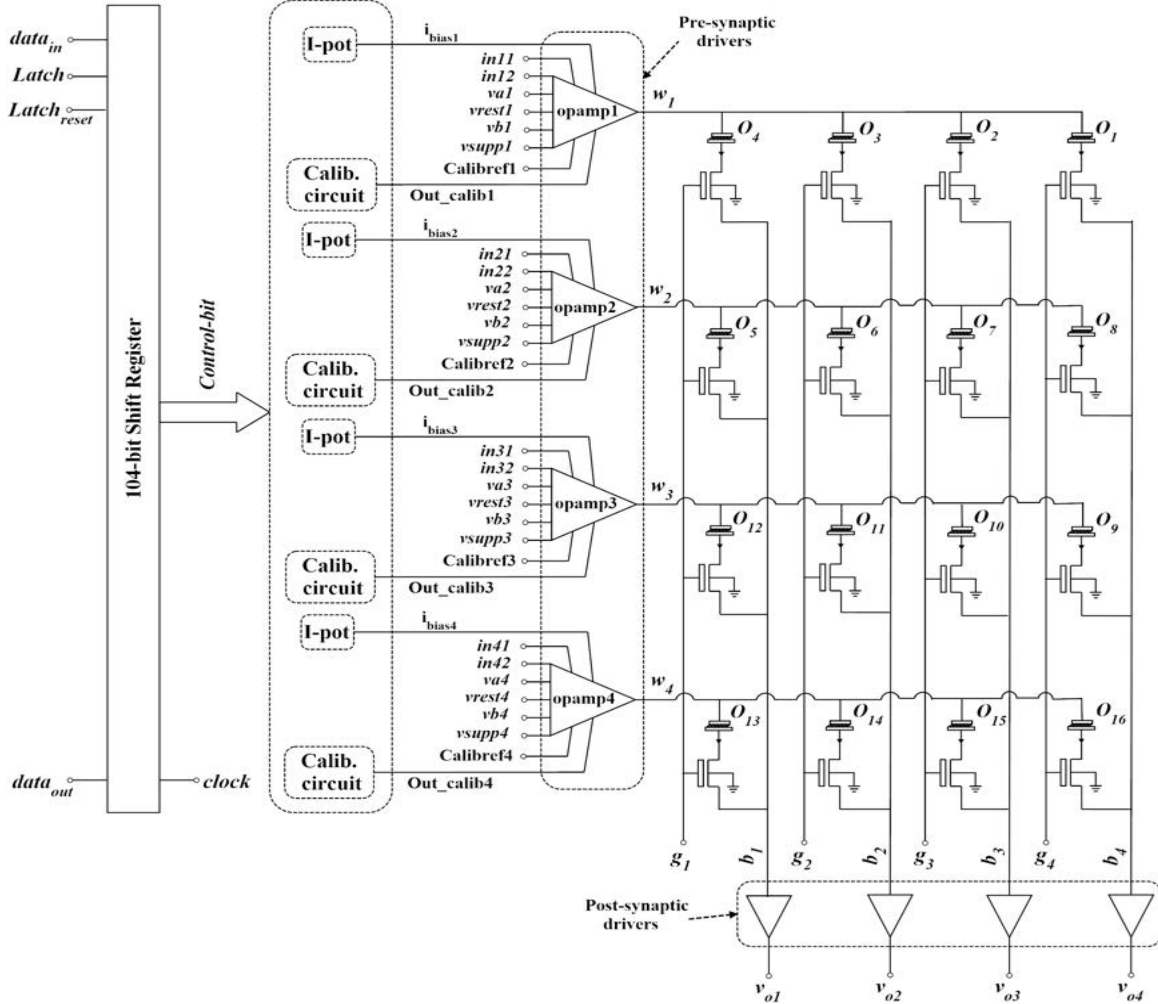


Figure 3-3: Scheme of a 4×4 1T1R crossbar with DC offset voltage calibration in each wordline.

circuits' detail, their transistor-level electrical simulation results (done using spectre or eldoD simulator[1]), description and working of the experimental set-up, and finally, the verification of experimental results of the calibration scheme.

All circuits, except the post-synaptic drivers, are designed using MAD200 PDK, where the OxRAM devices are monolithically integrated above the CMOS part of the chip. Layout-extracted simulations, taking into account technology-process corners, PVT variations, noise effects, and temperature variation feature that worst-case offset voltage in the order of 3 mV can be compensated down to 200 µV. This compensated

[1]Spectre is a SPICE-class circuit simulator developed at Cadence Design Systems®. EldoD is a pure SPICE simulator developed by Mentor Graphics®. Spectre emerged as a fast, more accurate, and reliable simulator when compared to the SPICE simulator [113].

DC offset voltage is further limited by fabrication defects in OxRAM such as nano battery effect, which is due to non-homogeneous ion distribution in the electrolyte [114, 115].

3.3.1 Design of two-stage PMOS-based differential pair opamp

We intend to design an opamp that meets the design criteria and specifications of DC offset calibration scheme- implemented in a crossbar. The main design challenge is to put the differential pair MOSFETs of the opamp in separate wells, as we need to access bulk-terminals of the MOSFETs. The specification we need to meet is to implement a fine calibration in the order of 100 µV steps or even less. Of course, this calibration output will be limited further by mismatch, electrical noise, parasitics in layout (if critical), experimental set-up, and other factors.

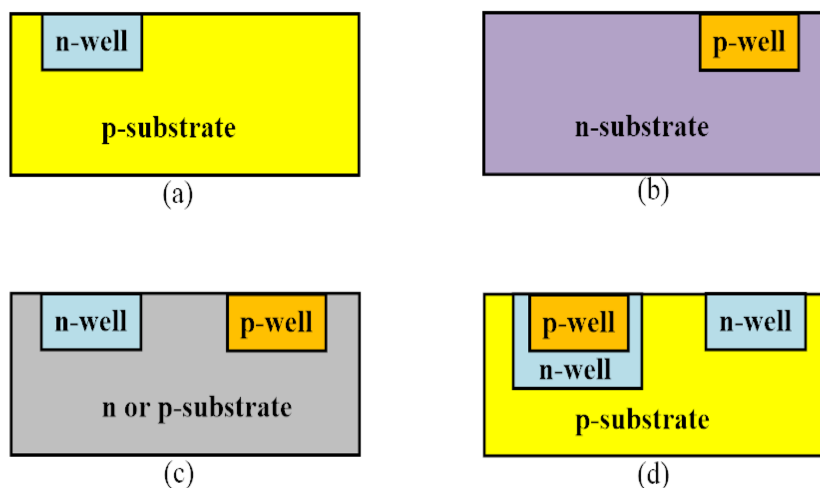


Figure 3-4: Possible wells used in bulk CMOS process: (a) n-well process, (b) p-well process, (c) Twin-well process, (d) Triple-well process.

MAD200 PDK has front-end design features such as- shallow trench isolation, triple well[2] (NISO[3]), twin-tub (or twin-well)[4], single poly CMOS process using a

[2]Triple-well is using different n-wells on p-substrate, which is lightly doped. It is mainly used to allow bodies of the MOSFETs to be at different potentials. The added n-well (that houses p-well) form a diode, which electrically isolates the p-well from the substrate, as shown in fig. 3-4(d).

[3]NISO stands for 'N Isolation'. It represents burying N-layer to isolate the p-well and underneath the NMOS devices to enable forward bias and back bias.

[4]Twin-well is using two wells on the same substrate (either p-substrate or n-substrate), which

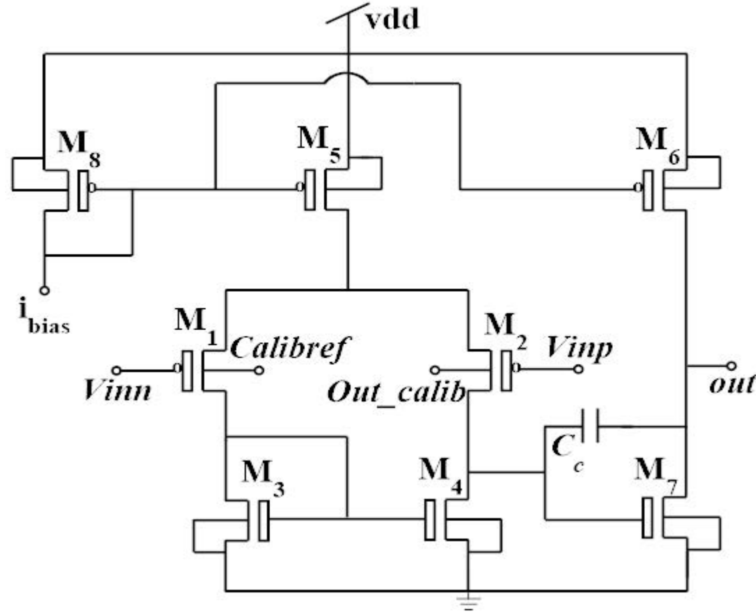


Figure 3-5: Schematic view of the two-stage differential opamp.

type <100>P_EPI(4 µm; 11.5 Ω.cm)P+substrate(10 mΩ.cm) and back-end design featuring damascene copper process[5] for all four metal levels. Fig. 3-4 shows the possible wells used in the bulk CMOS process [116]. Here, in our work, we used n-well CMOS process. In the n-well CMOS process, if we use NMOS differential pair for the opamp we need to have all the bulks connected to the substrate and hence we opted to keep PMOS for the differential pair. Here, we can make separate wells and bias them at different voltages.

The design approach of the opamp is based on iterating between- designing, analyzing by simulation, modifying the design; and obtaining new specifications. The opamp specifications such as gain, phase margin, bandwidth, ICMR values, power dissipation, slew rate, etc. are initially set to target values by considering its use as a buffer across the wordline to have two states of resistances as load condition when

is lightly doped to reduce excessive doping effects. Twin-well is used when bodies of NMOS are all biased with the same potential. Twin-well is shown in fig. 3-4(c)

[5]Damascence copper process - is a novel method of copper metalization to overcome problems like fast diffusion of Cu into Si and SiO2, poor oxidation/corrosion resistance, poor adhesion to SiO2 and difficulty in the conventional dry-etching technique. Unlike the conventional method, it is done by CMP and by using special barrier layers like Ta, TaN, TiN, and TiW to prevent intermixing of materials above and below the barrier.

used in a crossbar architecture. We also considered the operating frequency range of the pulses we should apply, bandwidth, and a worst-case of 5 pF load capacitor when setting the initial values of the specifications. We decided to keep mismatch low, wherever possible in the design. After determining the specifications of the opamp through various analyses the design parameters- like the sizing of MOSFETs and capacitance of the compensation capacitor are changed accordingly. Fig. 3-5 shows the schematic view of the two-stage differential opamp, whose differential pair MOSFETs' body-inputs are put in separate wells for biasing separately. Opamp comprises MOSFETs- $M_{1,2,....8}$ and miller compensation capacitor, C_c. The opamp design is carried out by biasing i_{bias} with 40 µA and keeping the bulks ($Calibref$ and Out_calib) at vdd. Fine-tuning of opamp specifications are carried out iteratively by keeping the following design relationships [117, 118] and optimizations-

1. For keeping Phase Margin (PM) of minimum 60° (targeted initial value), the condition for compensation capacitor's capacitance,

$$C_c \geq 0.22 \times C_L \tag{3.1}$$

   is met. Here, a worst-case capacitance of 5 pF is considered as C_L. When starting the design, compensation capacitor's capacitance is considered by taking $C_c \geq (2\ or\ 3) \times 0.22 C_L$ by considering the parasitic effect and to keep flexibility at later design stages for tuning parameters such as gain bandwidth product, slew rate, etc.

2. Slew rate (SR) is tuned by altering the bias current, i_{bias} and the sizes of the MOSFETs- M_8 and M_5. Slew rate is related to the drain current of MOSFET, M_5 by the relation,

$$SR = \frac{I_{D5}}{C_c} \tag{3.2}$$

   Here, I_{D5} is the drain current of MOSFET- M_5 and C_c is the compensation capacitor's capacitance.

3. DC operating point parameters (like β_{eff}, mobility of ions, threshold voltages, etc.) are obtained by simulating the diode-connected (or when MOSFETs are put in saturation region) PMOS and NMOS transistors of size- W = 10 µm and L = 2 µm. Though the minimum channel length of the MOSFET by PDK's design rule is 0.5 µm, we kept the minimum length of all MOSFETs in our design as 2 µm for mismatch reasons, as short channels, can degrade matching the behavior of MOSFETs by extra mismatch effects like RSCE[6].

4. Gain Bandwidth Product (GBW) is tuned by changing sizes of differential pair MOSFETs- $M_{1,2}$ by the relation,

$$GBW = \frac{g_{m1}}{C_c} \tag{3.3}$$

   Here, g_{m1} is the transconductance of MOSFET- M_1 and C_c is the compensation capacitor's capacitance.

5. ICMR+ value can be tuned by changing sizes of current mirror MOSFETs- $M_{3,4}$ and ICMR- value can be tuned by resizing MOSFET- M_5. The relationship of the ICMR values with MOSFET-sizes of $M_{3,4,5}$ is given by,

$$\left(\frac{W}{L}\right)_{3,4} = \frac{2I_{D3}}{\mu_n C_{ox}[V_{DD} - ICMR + +V_{t3max} \mid V_{t1min} \mid]^2} \tag{3.4}$$

$$\left(\frac{W}{L}\right)_{5} = \frac{2I_{D5}}{\mu_p C_{ox}[ICMR - -\sqrt{\frac{2I_{D1}}{\mu_p C_{ox}}} - \mid V_{t1min} \mid]^2} \tag{3.5}$$

   Here, I_{D3} and I_{D5} are the drain current of MOSFETs- M_3 and M_5 respectively. μ_n and μ_p are the mobilities of electrons and holes respectively. I_{D1} is the drain current of MOSFET- M_1. V_{t3max} is the threshold voltage of MOSFET-

[6]RSCE (Reverse Short Channel Effect) is an increase of threshold voltage with decreasing channel length. At short-channel length the halo doping of the source overlaps that of the drain, increasing the substrate doping concentration in the channel area, and thus increasing the threshold voltage. This increased threshold voltage requires a larger gate voltage for channel inversion.

M$_3$ for maximum common-mode voltage and V_{t1min} is the threshold voltage of MOSFET- M$_1$ for maximum common-mode voltage.

6. By current or transconductance relation,

$$\frac{\left(\frac{W}{L}\right)_7}{\left(\frac{W}{L}\right)_4} = \frac{g_{m7}}{g_{m4}} \tag{3.6}$$

we can further tune the size of MOSFETs- M$_4$ and M$_7$ or even phase margin (if needed), as for 60° of phase margin, $g_{m7} \geq 10 \times g_{m1}$. Here, g_{m1} and g_{m7} are transconductances of MOSFETs- M$_1$ and M$_7$ respectively. g_{m1} is determined earlier when tuning GBW using equation 3.3.

7. The design condition,

$$\frac{\left(\frac{W}{L}\right)_7}{\left(\frac{W}{L}\right)_4} = 2 \times \frac{\left(\frac{W}{L}\right)_6}{\left(\frac{W}{L}\right)_5} \tag{3.7}$$

to keep the opamp's schematic balanced and to keep minimal systematic DC offset of the opamp is met. Care is also taken that the opamp's gain is not underestimated during AC analysis. This is done by carefully tuning the DC component of VCM during AC analysis for getting the maximum gain, which is obtained previously in the DC sweep analysis.

8. MOM capacitor is used as a Miller compensation capacitor [119], C_c because of its characteristics like high density, good matching, and low parasitics.

9. The following relationships are further considered when tuning different parameters and specifications of the two-stage opamp-

$$C_c \propto PM \propto \frac{1}{GBW} \tag{3.8}$$

$$L_6 \propto GBW \propto \frac{1}{PM} \propto \frac{1}{Gain} \tag{3.9}$$

$$W_7 \propto PM \propto \frac{1}{GBW} \propto \frac{1}{Gain} \tag{3.10}$$

$$Flicker\, noise \propto \frac{1}{WL} \tag{3.11}$$

$$Mismatch^7 \propto \frac{1}{WL} \tag{3.12}$$

10. We also used 'multiplier' or 'number of fingers' option when making layouts and split the width of the MOSFET to make parallelly-connected MOSFET structures that are arranged in a stack, such that all sources and drains of each element are connected in parallel by suitable metal connections, where some of the drain and source connections serve two different elements thereby, reducing the silicon area and its associated reduction of the parasitic[8] capacitance of the two junctions- source-substrate and drain-substrate.

11. Layouts of the current-mirror MOSFET pairs- $M_{5,8}$ and $M_{3,4}$ are made using techniques like inter-digitized scheme and common-centroid technique to consume less silicon area and keep parasitic capacitance of the substrate junctions low [120].

Various analyses such as DC sweep analysis- to determine DC systematic offset, maximum DC open-loop gain, AC analysis- to determine AC gain, phase margin, gain bandwidth product, slew-rate determination experiment, Monte Carlo DC offset variation for mismatch and noise analyses- to determine summarized and input-referred noise— are all done for different load configurations. On top of this, a post-layout simulation was done to observe the effect of parasitics in the layout. Table 3.1 shows the values of the MOSFET sizes of the designed opamp. Fig.3-6(a) shows the layout

[7]Mismatch reduces on using less area. Mismatch also reduces when QUAD (square-shaped) layout structures are used. One way to do this is to connect MOSFETs in parallel by dividing its width.

[8]Parasitic effects are the spurs that appear because of interference between interconnections' lines and they come from the substrate or can be the consequence of opening switches. Parasitic effects can be reduced by making symmetrical layouts.

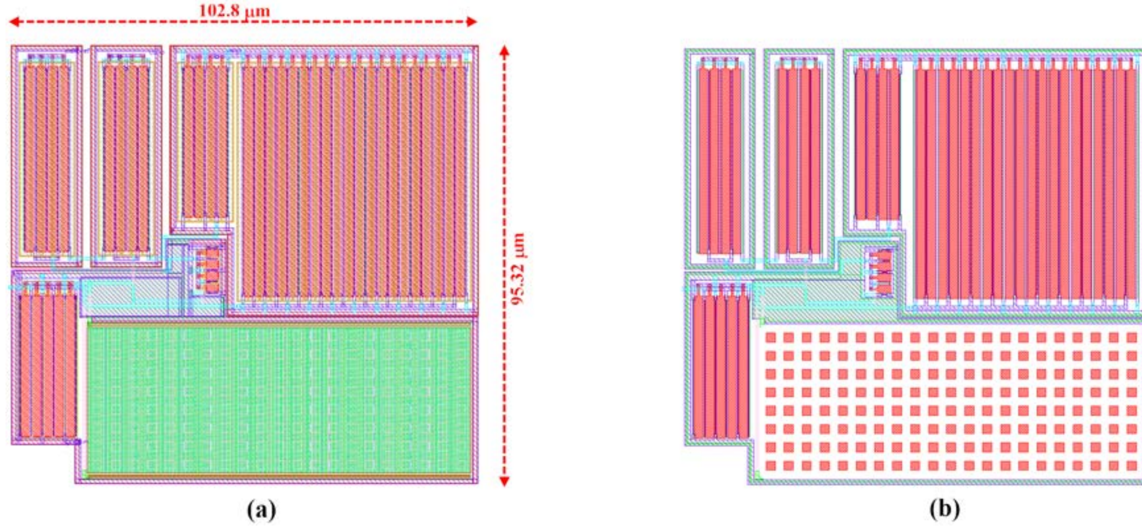


Figure 3-6: (a) Layout view of the two-stage differential opamp, (b) Parasitics extracted layout view of the two-stage differential opamp.

view of the two-stage opamp and fig.3-6(b) shows the layout view of the opamp after parasitic extraction.

Fig.3-7(a) shows the technology-process corner[9] and fig.3-7(b) shows Monte Carlo[10] variation (300 runs) of DC transfer curve of the opamp. Fig.3-7(c) shows the Monte Carlo (300 runs) distribution of the DC offset voltage of opamp and comparison of nominal and layout-extracted simulated DC transfer curve of the opamp is plotted in fig.3-7(d). A shift of about 100 µV between the schematic and layout-extracted simulated DC transfer curve of the opamp is due to the real parasitic components in the extracted layout.

The opamp has a gain of 101 dB, a phase margin of 60° and a slew rate of 15.07 V/µs when connected with a capacitive load of 5 pF. It has a gain of 69.3 dB, a phase margin of 66.2° and a slew rate of 13 V/µs when connected to a R_LC_L load of R_L= 2 kΩ; C_L= 5 pF. It is observed to have a gain of 98.5 dB, a phase margin of 60° and a slew rate of 15.04 V/µs when connected to a R_LC_L load of R_L= 225 kΩ; C_L= 5 pF. The load resistors are chosen based on the HRS and LRS values of the 1T1R device.

[9]Technology-process corners include Front End Of Line (FEOL) corners. The best-case and worst case corners are classified based on different design parameters like mobility, vth variation, the resistance of the actives, body coefficient, oxide thickness, and PVT variations.

[10]Monte Carlo simulations include both process (wafer-to-wafer variations), On-Chip Variations (OCV) like device mismatch.

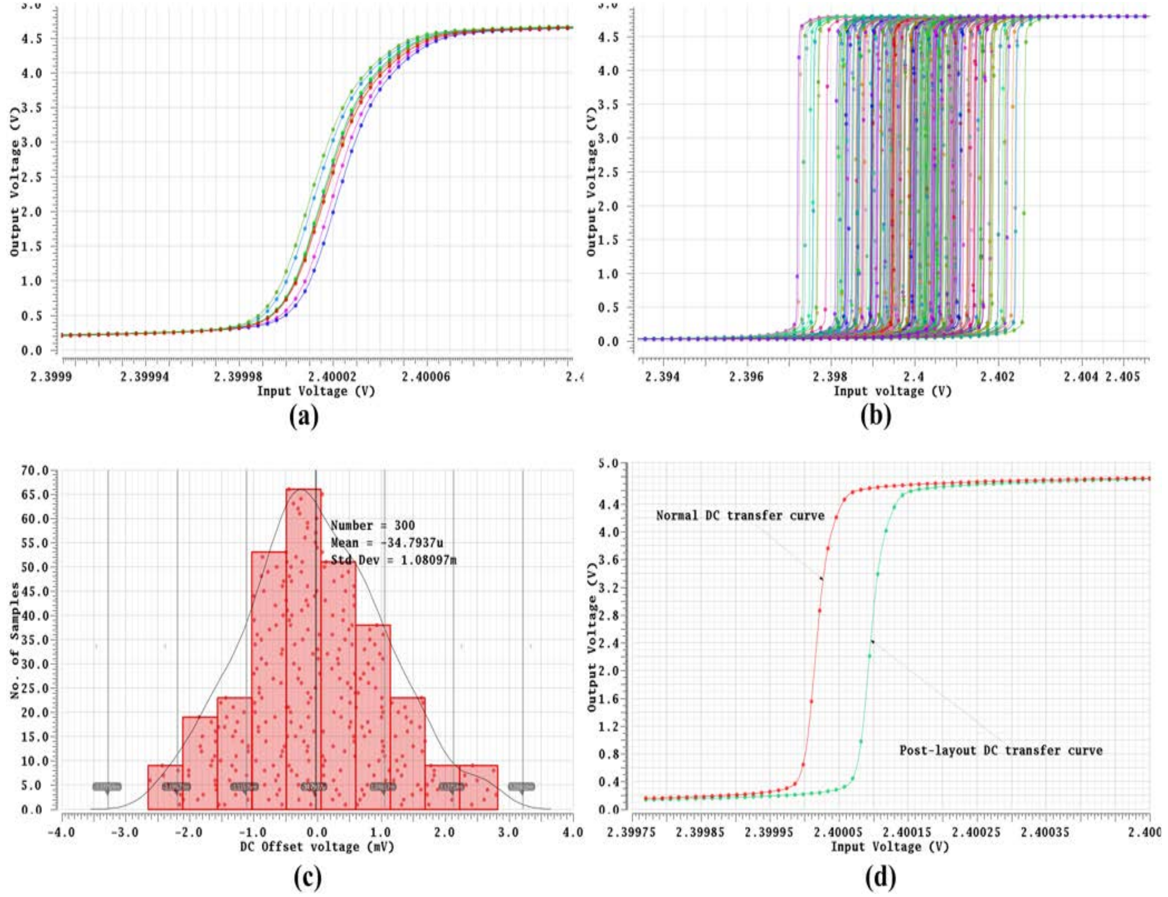


Figure 3-7: (a) Technology-process corner variation of DC transfer curve of opamp, (b) Monte Carlo variation of DC transfer curve of the opamp, (c) Monte Carlo distribution of DC offset voltage of opamp, (d) Comparison of nominal and layout-extracted simulated DC transfer curves of opamp.

Table 3.1: Design parameters of the two-stage PMOS-based differential-pair opamp.

Parameter	**Value**
Supply voltage	$vdd = 4.8$ V
Bias current for opamp	$i_{bias} = 40$ µA
PMOS-size in differential pair	$(\frac{W}{L})_{1,2} = (\frac{40\ \mu m}{2\ \mu m})$
PMOS-size in input current mirror	$(\frac{W}{L})_{5,8} = (\frac{32.5\ \mu m}{2\ \mu m})$
NMOS-size in input current mirror	$(\frac{W}{L})_{3,4} = (\frac{2.5\ \mu m}{2\ \mu m})$
Second stage PMOS-size	$(\frac{W}{L})_6 = (\frac{49.75\ \mu m}{2\ \mu m})$
Second stage NMOS-size	$(\frac{W}{L})_7 = (\frac{30.6\ \mu m}{2\ \mu m})$
Capacitance of compensation capacitor	$C_c = 2.521$ pF

A factor of about 200 is considered between the two states and hence the opamp is analyzed at both 2 kΩ and 225 kΩ. Additionally, the opamp is also analyzed for 7 kΩ load conditions. A worst-case capacitance load of 5 pF is considered along the wordlines of the crossbar. A summary of the design specifications of the opamp for four different load configurations is tabulated in Table B.1 in Appendix B. The input-referred noise of the opamp is 72.12 µV/$\sqrt{Hz}$. The opamp also has a low mean statistical DC offset mismatch of 35 µV.

3.3.2 Design of pulse-shaping digital blocks across wordlines of memristive crossbar

Each pre-synaptic driver is made up of a digital pulse shaping block and an opamp. Fig.3-8 shows the pre-synaptic driver circuit, whose output is connected to wordline, w_1. Fig.3-9 shows the layout view of the pre-synaptic driver circuit. There are similar pre-synaptic drivers in the other three wordlines of the crossbar. The digital controls of the pulse shaping digital block– in11, in21, and vsupp1 are used to control the analog biases– the upper magnitude of the voltage test pulse, va1, the lower magnitude, vb1, and the resting voltage, vrest1 from reaching the wordline, w_1 of

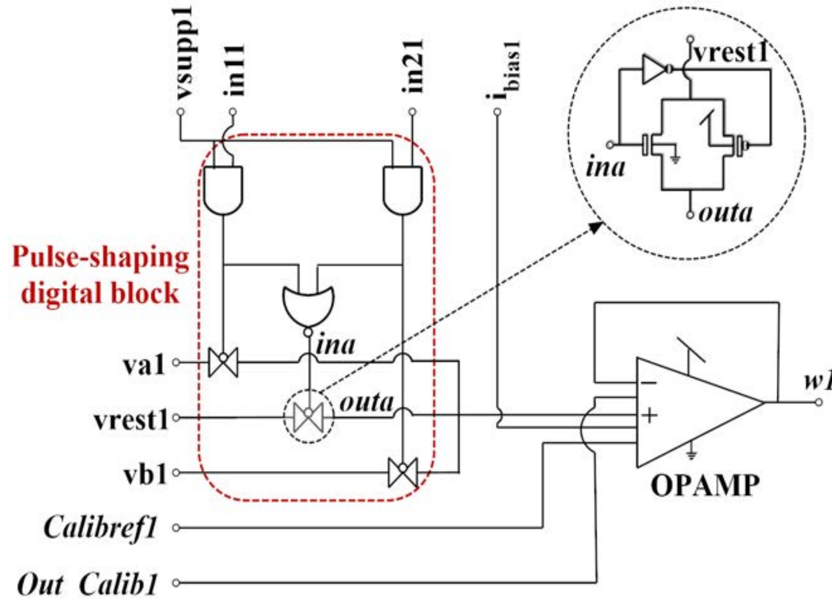


Figure 3-8: Schematic view of the pre-synaptic driver across wordline, $w1$.

Table 3.2: Output of pre-synaptic driver of wordline, $w1$ for different combination of digital inputs.

vsupp1	**in11**	**in21**	**w1**
ON	ON	OFF	va1
ON	OFF	ON	vb1
ON	OFF	OFF	vrest1
ON	ON	ON	intermediate value

crossbar through switches and an opamp, which is connected in a buffer configuration. Table 3.2 shows how the various combination of the digital inputs is used to set the output of the pre-synaptic driver. Here, 'ON' indicates 4.8 V (vdd), and 'OFF' indicates 0 V. The 'intermediate' value is due to the unconfigured 4th digital input combination, which depends on all three analog biases.

The opamp's DC offset voltage is calibrated before setting a low-amplitude infer-

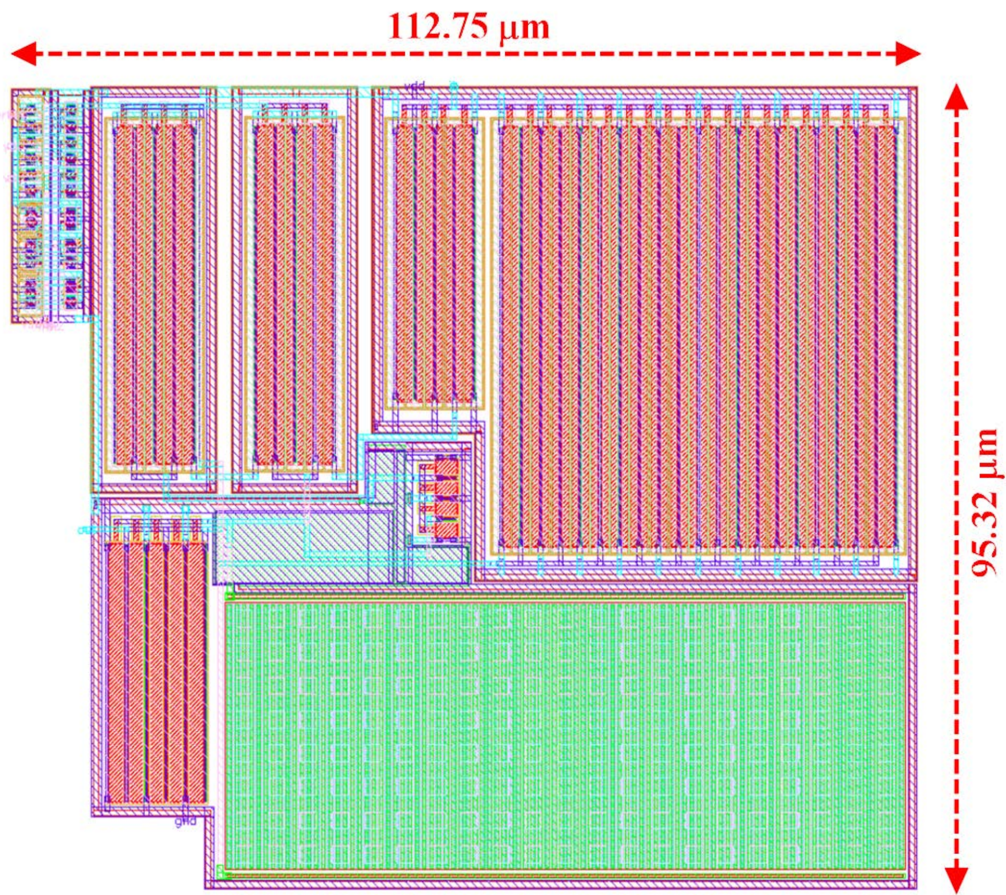


Figure 3-9: Layout view of a pre-synaptic driver.

ence pulse in the read-out wordline path and the opamp is biased with the output current, i_{bias1} from I-pot [121]. Each pre-synaptic driver has its calibration circuit and I-pot. The pre-synaptic drivers are simulated and checked for both load conditions i.e. LRS and HRS during its design. Both the bulk terminals of the differential pair MOSFETs of the opamp– '*Calibref1*' and '*Out_ Calib1*' are pulled, which are used by the calibration circuit for offset calibration.

3.3.3 Design of body-input three-stage offset calibration scheme

During inference read operation mode, the current pulses associated with the OxRAM states in the crossbars are integrated with the neurons or read by external buffers for testing purposes. Since low-power consumption and scalability are two major concerns, it becomes non-trivial to investigate how small the pulses can be made if the offset voltage in the crossbar's wordlines is decisively high to affect the results. Hence, it is important to calibrate and compensate for such an offset voltage. A possible offset calibration technique is to adjust the bulk voltage of the PMOS-based differential pair in the opamp by a digitally-controlled word, which is stored in a register at start up,

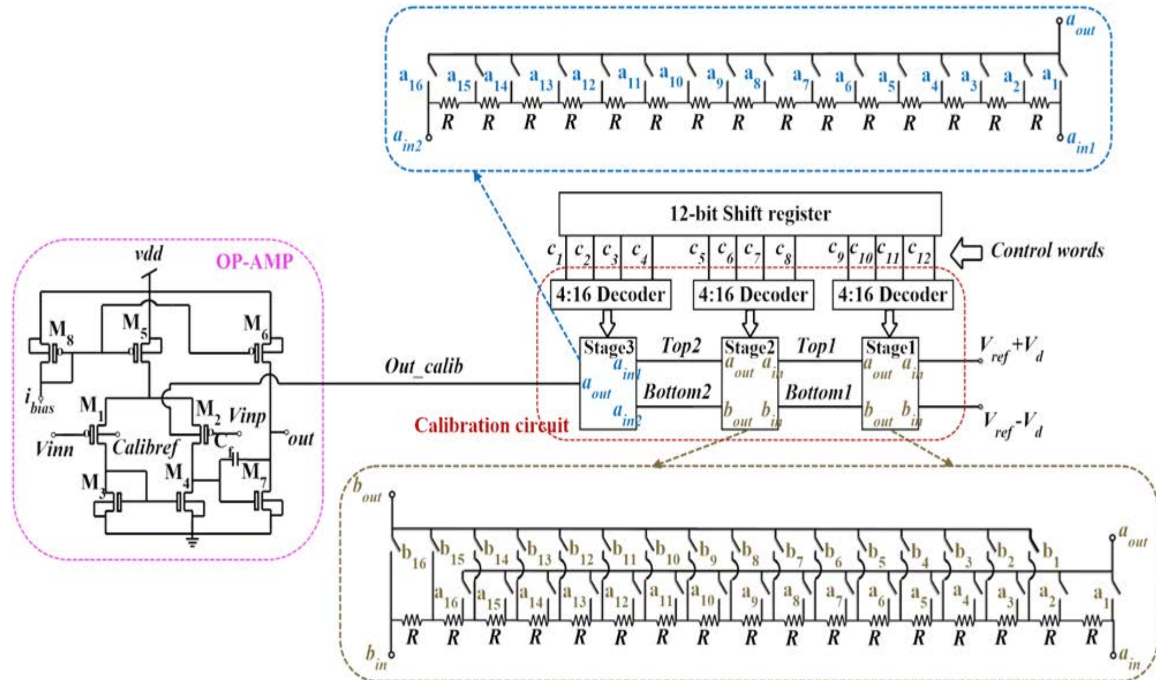


Figure 3-10: Schematic view of the three-stage calibration scheme.

or permanently with memristor-based non-volatile registers. This strategy, however, allows a coarse calibration only. For example, assuming an offset spread of 50 mV among all uncalibrated opamps, a 10-resistor ladder with a calibration differential voltage of $V_d = 25$ mV, would lead to a calibration step of 5 mV.

Overcoming this, a three-stage DC offset calibration scheme is proposed, as shown in Fig.3-10, which can eventually result in a much finer calibration scheme. It consists of a cascade of three resistor ladders of high ohmic unsalicided N+ polysilicon resistors–17 resistors in the first two stages and 15 resistors in the third one. PMOS switches are used to select the resistor ladder output, as the reference voltage, V_{ref} is set near vdd. This is because we want to calibrate via PMOS differential pair MOSFET. The first stack of resistors is connected to the references voltages V_{ref} +

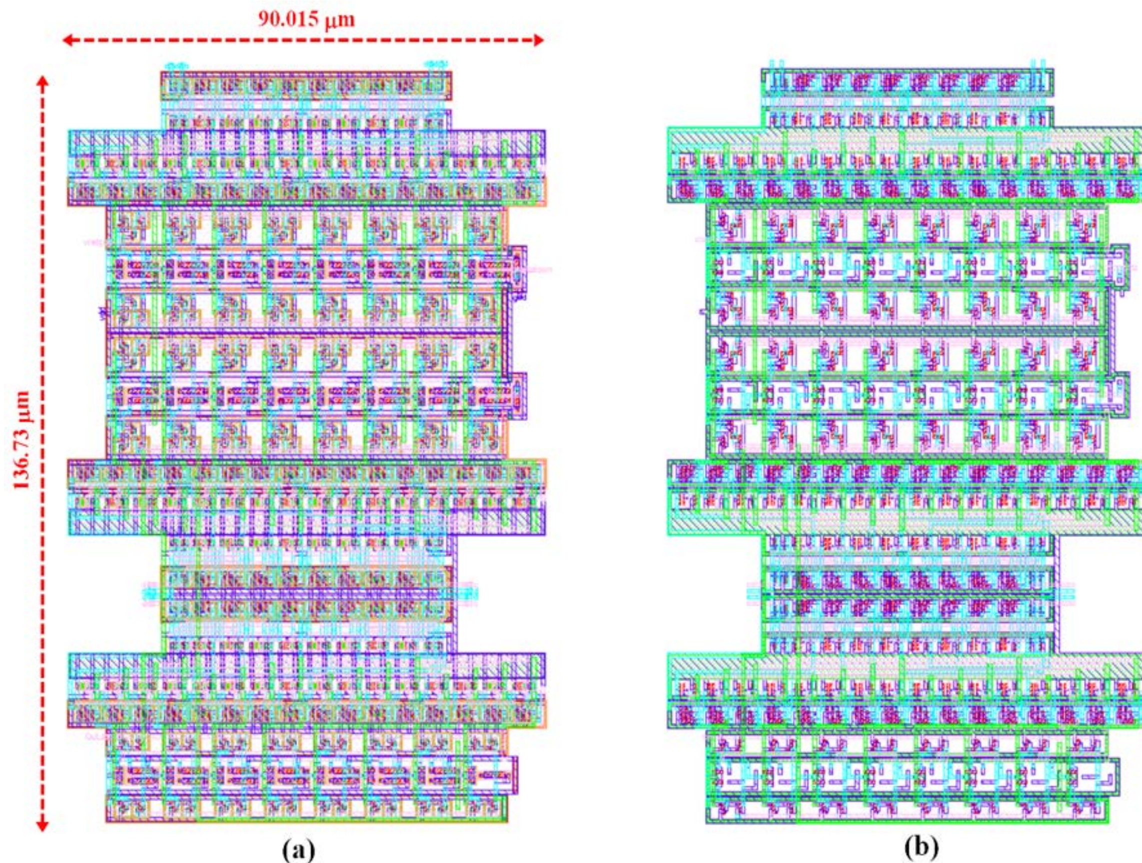
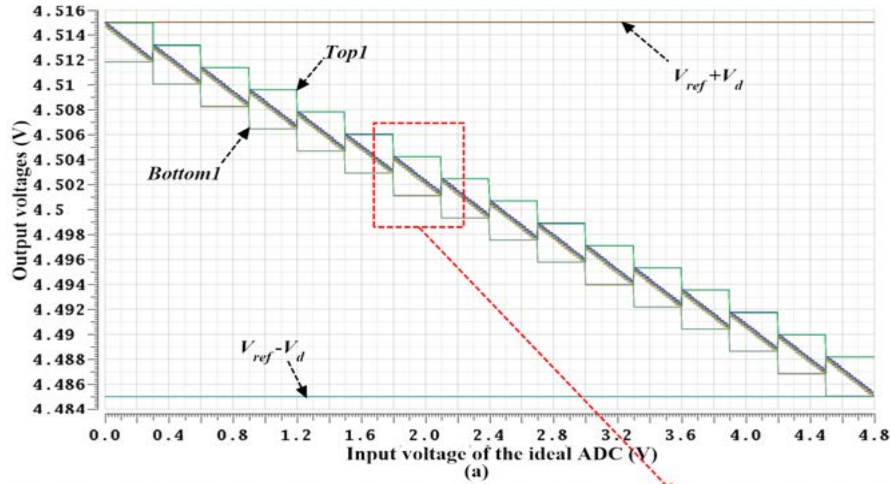


Figure 3-11: (a) Layout view of the three-stage calibration scheme with the decoders, (b) Parasitic extracted layout view of the three-stage calibration scheme with the decoders.

Table 3.3: MOSFET sizing and biasing parameters of the three-stage calibration scheme.

Parameter	**Value**
Supply voltage	$vdd = 4.8\ V$
Calibration reference voltage	$V_{ref} = 4.5\ V$
Calibration differential voltage	$V_d = 15\ mV$
PMOS-size used in calibration scheme	$(\frac{W}{L}) = (\frac{1\ \mu m}{0.5\ \mu m})$
Resistance of resistor in cali. resistor-bank	$R = 30\ k\Omega$
Size of resistor in cali. resistor-bank	$(\frac{W}{L})_R = (\frac{0.64\ \mu m}{3.218\ \mu m})$
Inv. terminal PMOS diff. pair bulk voltage	$Calibref = 4.5\ V$



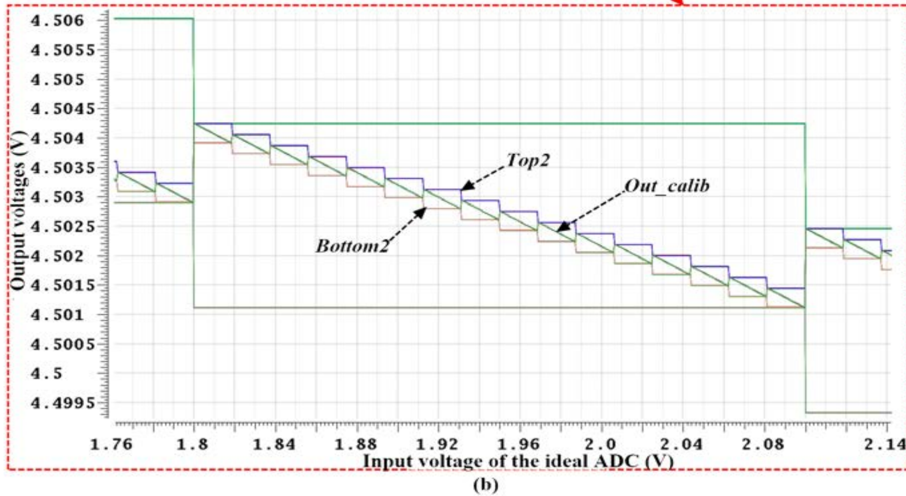


Figure 3-12: Technology-process corner variation results of the three-stage calibration scheme: (a) Technology-process corner variation results of signals - V_{ref} - V_d, V_{ref} + V_d, $Top1$ and $Bottom1$ (b) A zoom-preview of technology-process corner variation results showing signals - $Top2$, $Bottom2$ and Out_calib.

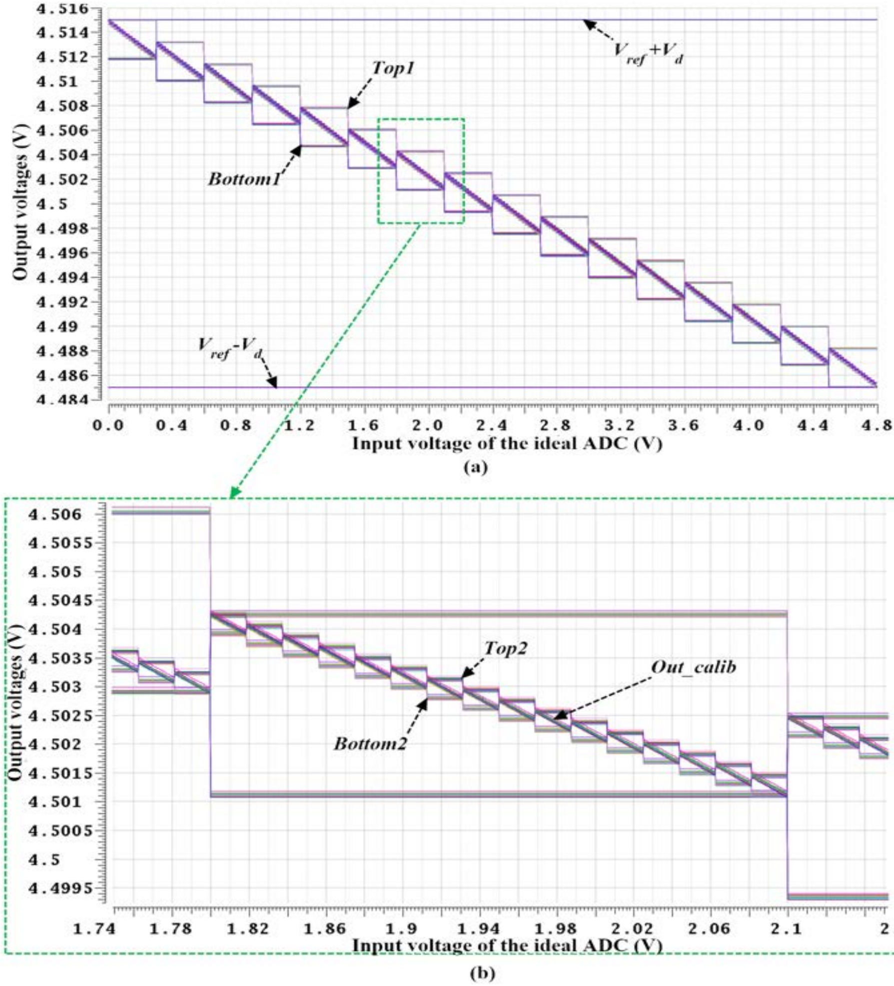
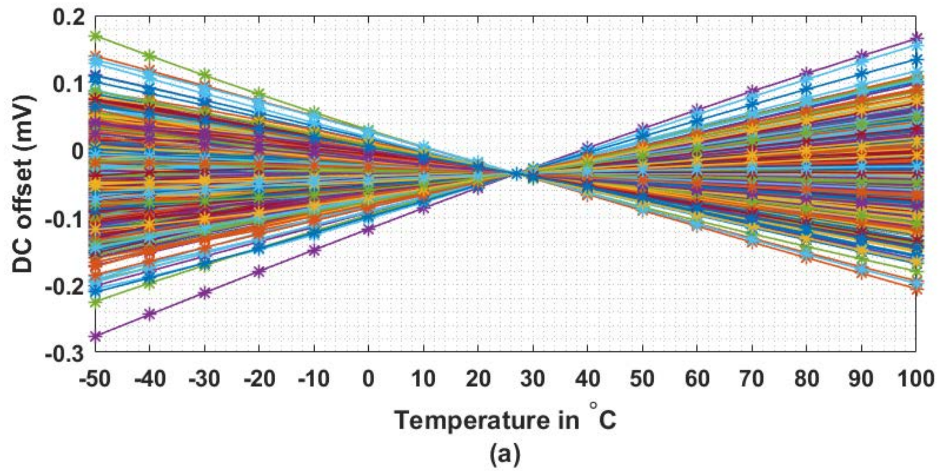


Figure 3-13: Monte Carlo variation results of the three-stage calibration scheme: (a) Monte Carlo variation results of signals - V_{ref} - V_d, V_{ref} + V_d, $Top1$ and $Bottom1$ (b) A zoom-preview of monte Carlo variation results showing signals - $Top2$, $Bottom2$ and Out_calib.

V_d and V_{ref} - V_d, which are set to choose a coarse (stage 1) voltage range, which in turn is used to pick finer (stage 2 and stage 3) ranges in the next proceeding stacks. Here, V_d is the differential voltage and V_{ref} is the reference voltage.

In the first stage, it is possible to obtain a $(\frac{2\times V_d}{16})$ coarse voltage step, whereas in the second stage a $(\frac{2\times V_d}{16\times16})$ fine voltage step is obtained and in the third stage it is possible to get a $(\frac{2\times V_d}{16\times16\times16})$ finer voltage step. MOSFET-sizing of switches and biasing parameters of the calibration scheme is shown in Table 3.2. The PMOS switches are controlled by three 4:16 decoders whose control bits are loaded from a 12-bit shift register. The decoders are designed with less number of gates. The analog buffers



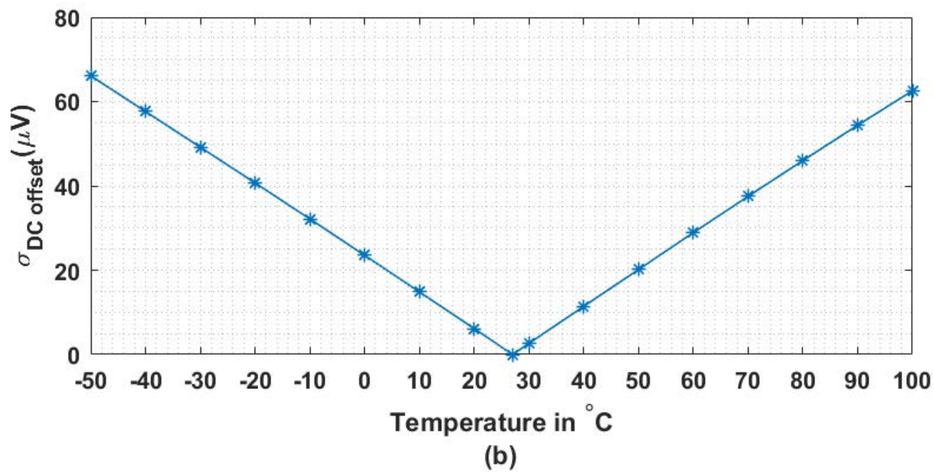


Figure 3-14: (a) Monte Carlo variation of DC offset voltage due to temperature with calibration at 27°C, (b) Sigma of DC offset voltage due to temperature variation with calibration at 27°C.

between the different stages in the calibration circuit are not used since they can introduce additional offset voltages. For calibration, one has to properly select the switches using decoders and set the bias for coarse, fine and finer offset calibration at one of the bulk-terminals (in Fig. 3-10 it is 'Out_calib') of the differential pair MOSFETs, while a reference voltage is set at the other bulk-terminal, '$Calibref$'. Fig.3-11(a) shows the layout view of the three-stage calibration scheme with its decoders and Fig.3-11(b) shows the layout view of the calibration scheme with parasitics content, with which a post-layout simulation is also performed to observe the effect

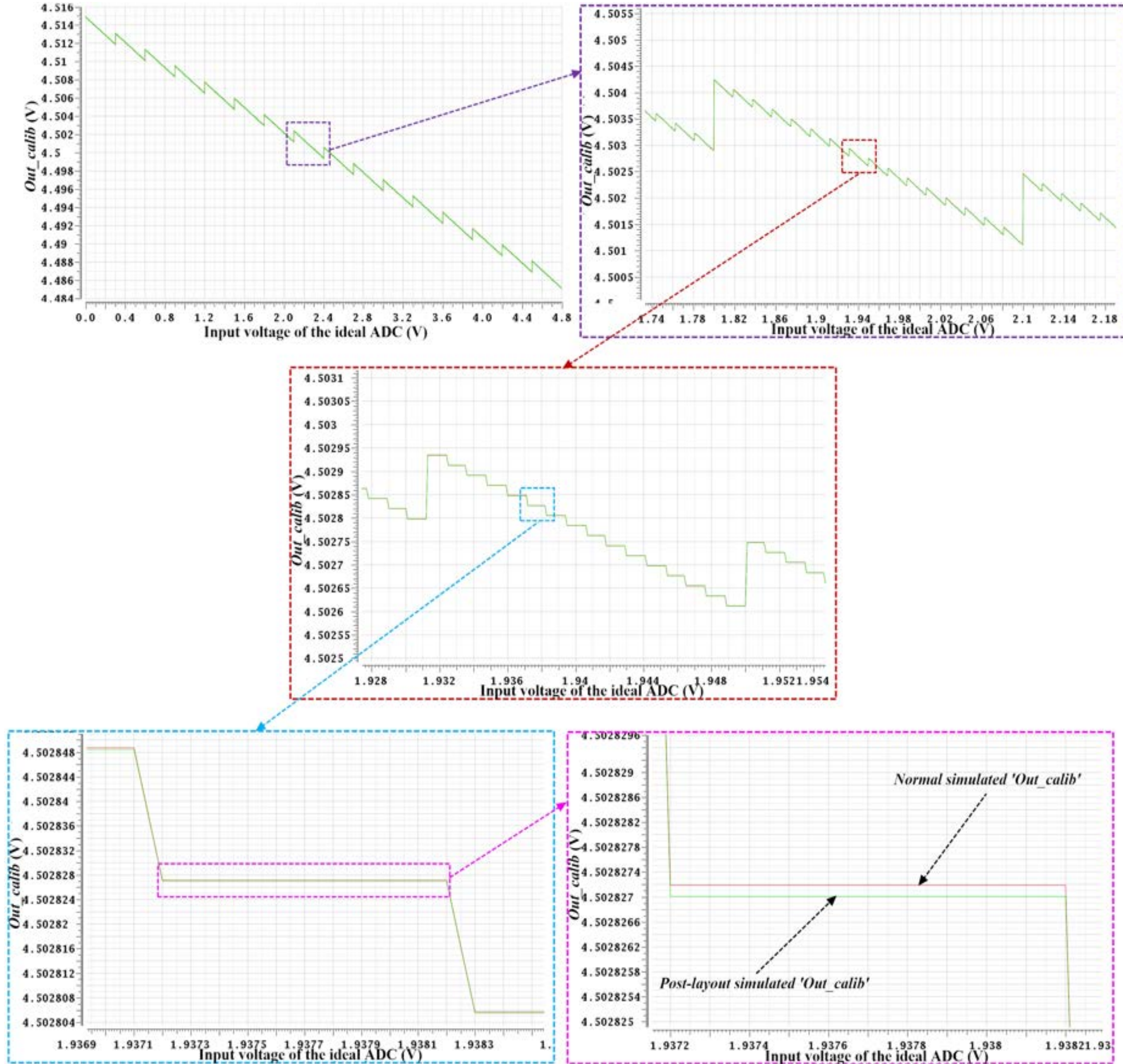


Figure 3-15: Comparison of nominal and layout-extracted simulated output, *Out_calib* of the calibration scheme.

of parasitics in the layout, which can be critical during the finer (stage 3) calibration.

Simulations considering variations due to mismatch, technology process corners, temperature, and parasitics present in the layout are done during design. Fig. 3-12 shows the technology-process corner variation results of the calibration scheme of wordline, w_1, and Fig. 3-13 shows the Monte Carlo variation (12 runs) results of the calibration scheme of wordline, w_1 when its full control words are swept using an ideal ADC. The slight upward jumps in the output voltage of the calibration scheme are due to the interleaved approach of the calibration scheme. Fig. 3-14(a)

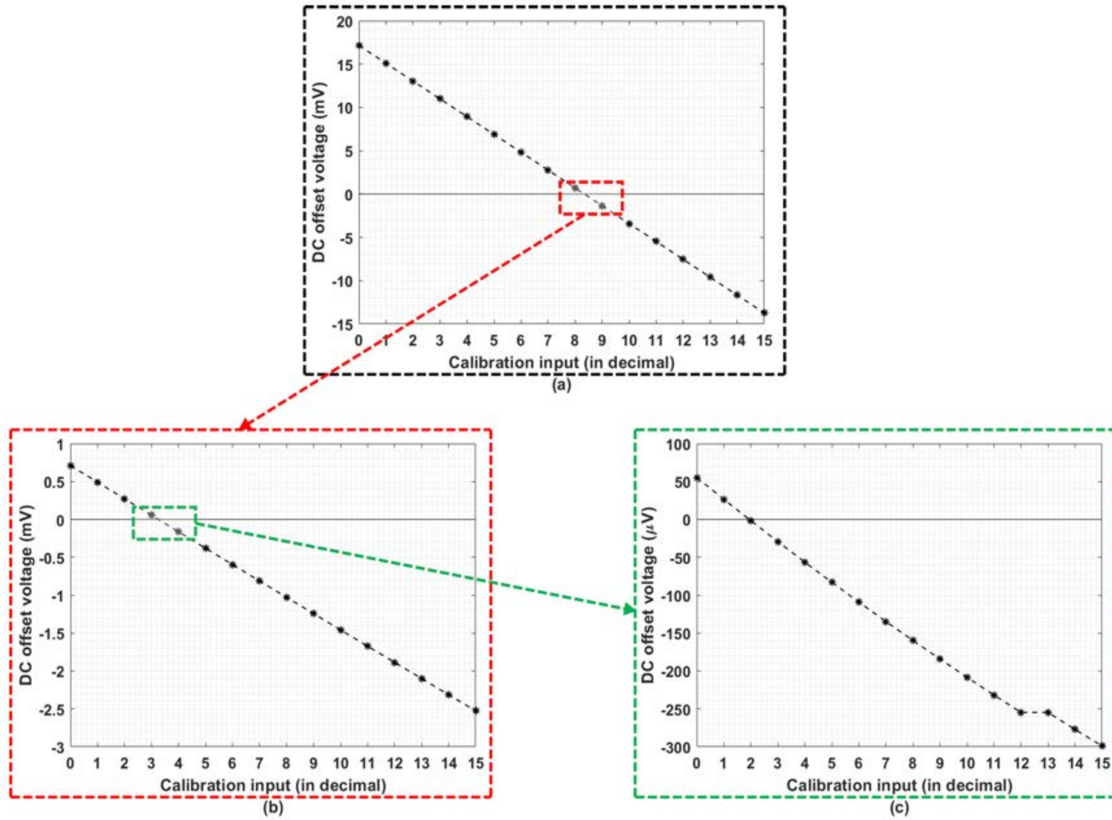


Figure 3-16: (a) Simulation results during coarse (stage 1) calibration of DC offset voltage across wordline, $w1$ targeting the zero-crossing region, (b) Simulation results during fine (stage 2) calibration of DC offset voltage across wordline, $w1$ targeting the zero-crossing region, (c)Simulation results during finer (stage 3) calibration of DC offset voltage across wordline, $w1$ targeting the zero-crossing region.

shows the Monte Carlo (500 runs) variation of DC offset voltage due to different temperatures[11] when DC offset is calibrated and compensated at 27 °C. Fig. 3-14(b) shows the sigma variation of the DC offset voltage due to different temperature when DC offset is calibrated and compensated at 27 °C. Fig. 3-15(a) shows the comparison of the layout-extracted simulation results with the nominal simulation results of the calibration scheme of wordline, w_1 when its full control words are swept using an ideal ADC. A very small voltage difference of about 2 μV is observed by this comparison, which is due to the presence of real parasitic capacitors and resistors in the extracted layout. Fig. 3-16 shows the simulation results during coarse (stage 1), fine (stage 2),

[11] A separate statistical temperature variation result is obtained along with device mismatch. Here 27 °C is the nominal simulation temperature.

and finer (stage 3) calibration of DC offset voltage across wordline, $w1$ of the crossbar targeting the zero-crossing regions.

3.3.4 Design of 4 × 4 1T1R crossbar

A 4×4 1T1R crossbar is designed using MAD200 PDK, whose schematic and layout views are shown in Fig. 3-17. The crossbar's wordlines are connected to the output of the buffer configured opamps that facilitate the calibration of DC offset across the wordlines. The basic operations of the 1T1R devices in a crossbar are simulated and analyzed before designing the crossbar, whose details are explained previously in Section 2.3.1 of Chapter 2. The working principle is to target a synapse using decoders and perform basic OxRAM operations. Initially, each OxRAM device is at very high resistance state of 100 GΩ, called PRS, and hence it has to be electro-formed to make the conductive filament in the oxide layer for the first time [122]. After forming, the OxRAM typically reaches LRS, and hence a RESET (or erase) operation has to be

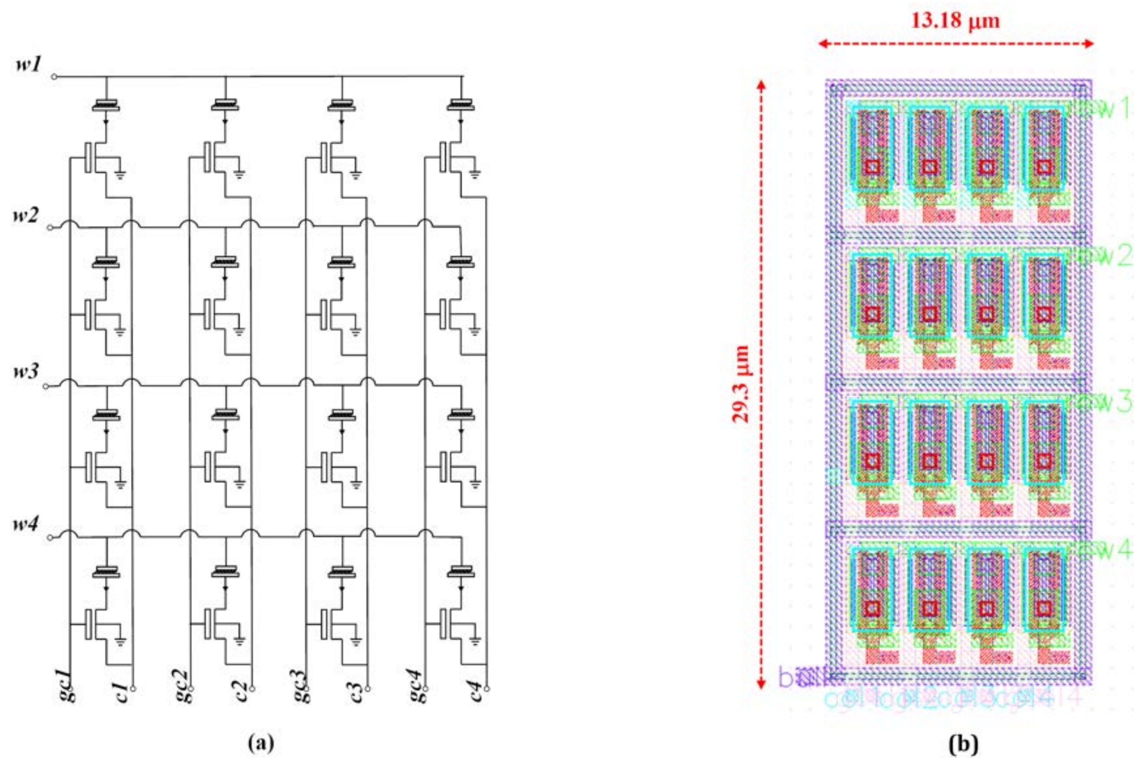


Figure 3-17: (a) Schematic view of the 4 × 4 1T1R crossbar, (b) Layout view of the 4 × 4 1T1R crossbar.

carried out to push it to HRS. Now, a SET (or write) operation has to be carried out to push the OxRAM back to LRS. Like this, the OxRAM device can be operated in binary mode, i.e. switching between LRS and HRS. The NMOS transistor connected in series with the memristive device acts as a current limiter for initial high forming current.

3.3.5 Design of I-pots

I-pots are digitally programmable current sources which, from a reference current, can provide any desired current with high precision, down to pA [121]. I-pot circuits are used as current bias source circuits for the opamp. The schematic view of the I-pot is shown in Fig. 3-18. It has a decade current splitter, a fine current splitter, and a current sign selector and tester. The current splitter circuits are MOS ladder structures [123]. The decade current splitter has 6 current splitting possibilities, where the reference current, Iref is split by 10 in each stage of a MOS ladder structure. The fine current splitter has 2^8 combinations, where the output of decade current splitter is split by 2 in each combination using the MOS ladder structure. The control bits

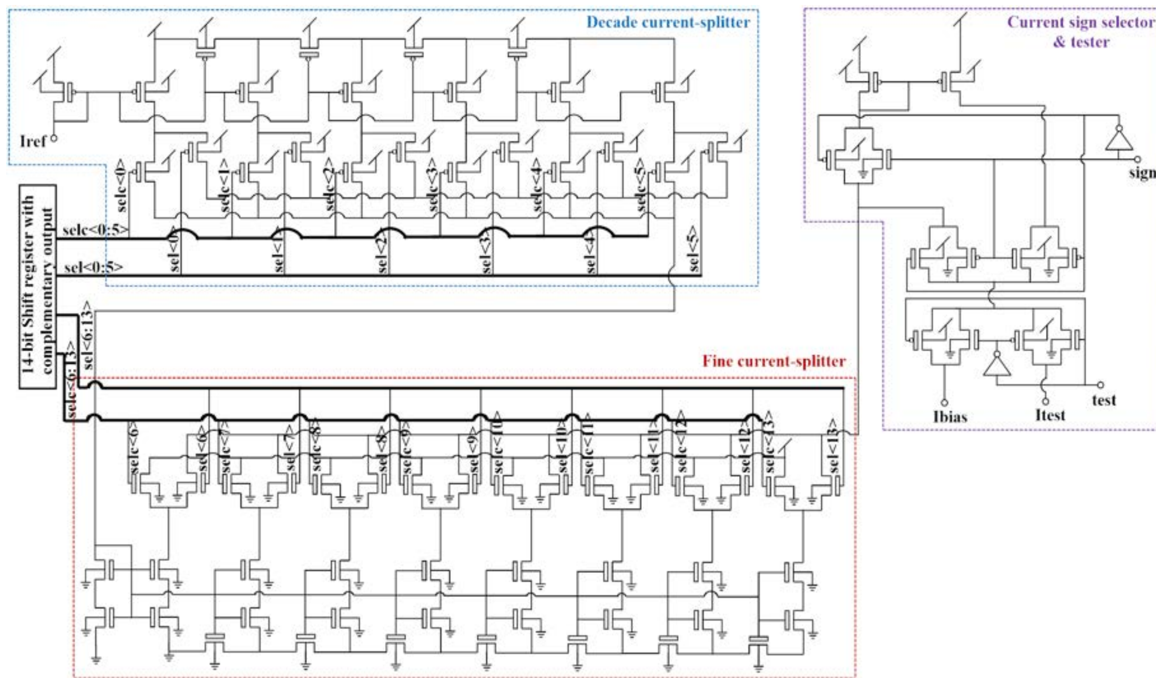


Figure 3-18: Schematic view of the I-pot.

Table 3.4: Values of the control signals for setting output current of I-pot to 40 µA.

Control signals	Value
sel<0>	ON
sel<1>	OFF
sel<2>	OFF
sel<3>	OFF
sel<4>	OFF
sel<5>	OFF
sel<6>	ON
sel<7>	OFF
sel<8>	ON
sel<9>	ON
sel<10>	OFF
sel<11>	ON
sel<12>	OFF
sel<13>	OFF

for these current splitters are loaded from a 14-bit shift register. The signal- 'sign' is used to switch the sign of the current-after splitting, and the signal- 'test' is used to test the current before it is used to bias the opamp.

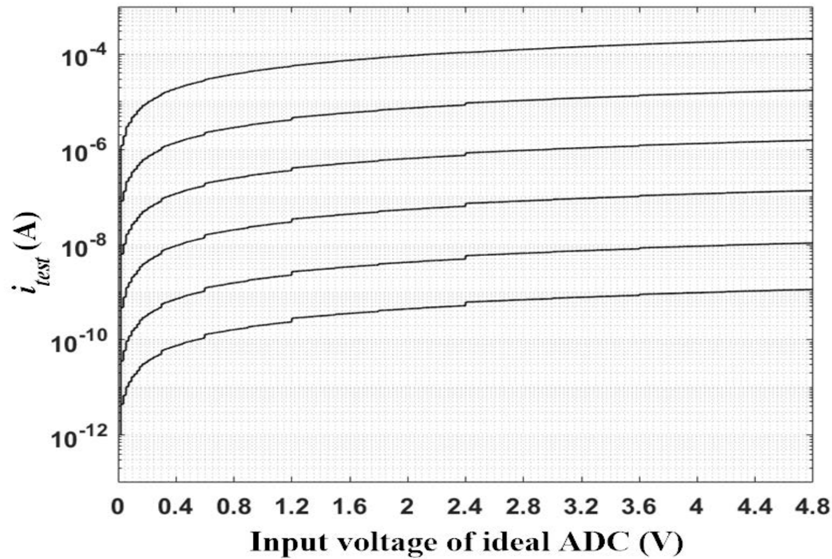


Figure 3-19: Simulated output currents of I-pot for 6 possible control bits of decade current splitter for the reference current, i_{ref} = 100 µA.

$Ibias_{1,2,3,4}$ are the bias currents of the opamps in all the four wordlines of 4 × 4 1T1R crossbar. Fig. 3-19 shows the output current of an I-pot for the 6 possible control bits of the decade current splitter for the reference current, Iref = 100 µA when the input voltage of an ideal ADC– connected to the digital controls of the decade and fine current splitters is swept for the full range of control word. Table 3.4 shows the values of the control signals for setting the output current of I-pot to 40 µA. Here, 'ON' represents 4.8 V (vdd) and 'OFF' represents 0 V.

3.3.6 Design of D-flip flop based shift-register

The calibration schemes and I-pots for all four pre-synaptic drivers of the 4×4 1T1R crossbar needs 104-bit control lines to control them digitally. Controlling all 104-bit digital lines can easily increase the number of pads, which eventually can increase the number of driver boards or can add complexities in making the test-bench. To overcome this, a shift register is used to load the control bits for the I-pot circuits and the calibration circuit. In each wordline of the 4×4 crossbar, the I-pot needs a 14-bit data control word and the calibration scheme needs a 12-bit data control word. So, a 104-bit latched D-flip-flop-based shift register is used to load all the control

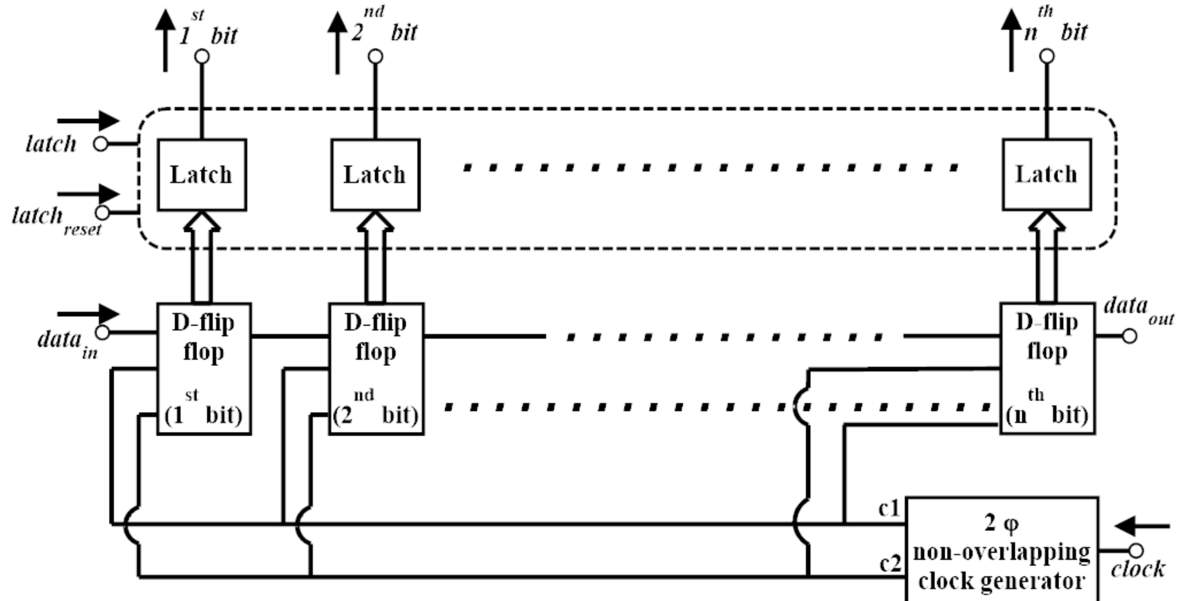


Figure 3-20: Scheme of n-bit D-flip flop based shift register.

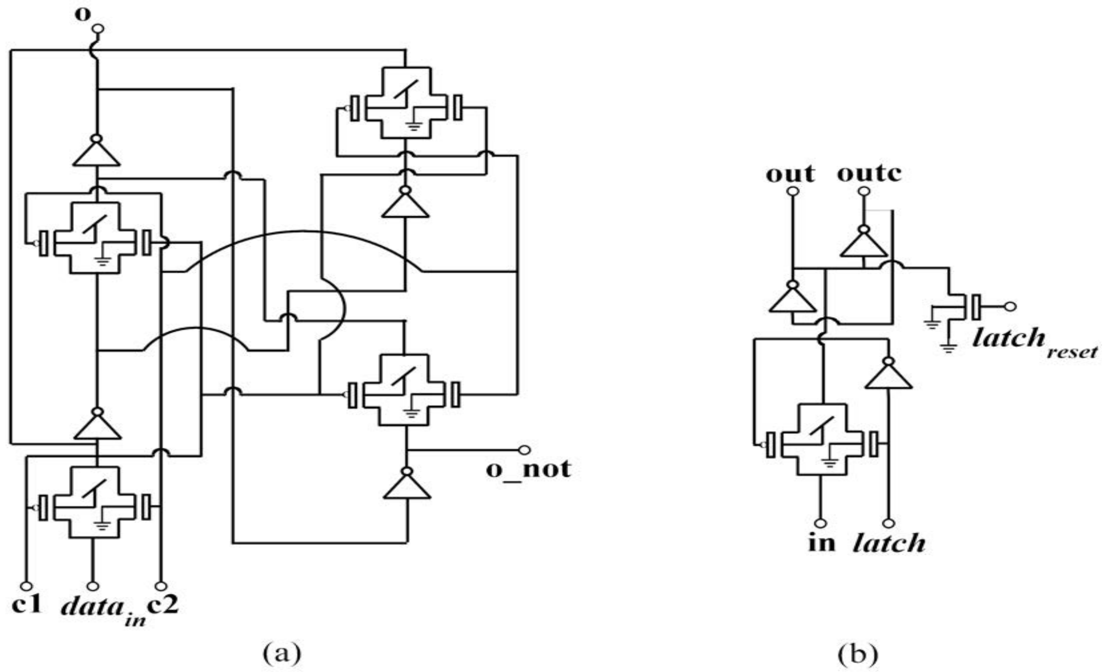


Figure 3-21: (a) Schematic view of edge-triggered D-flip flop, (b) Schematic view of the latch.

bits for the I-pots and calibration circuits for all the four wordlines in the crossbar. Fig. 3-20 shows the scheme of D-flip flop based shift register and Fig. 3-21 shows the schematic view of edge-triggered D-flip flop and latch. A 2-phase non-overlapping clock generator is used to prevent overlapping of clock signals.

Fig. 3-22 shows the transistor-level electrical simulation (done in spectre) results

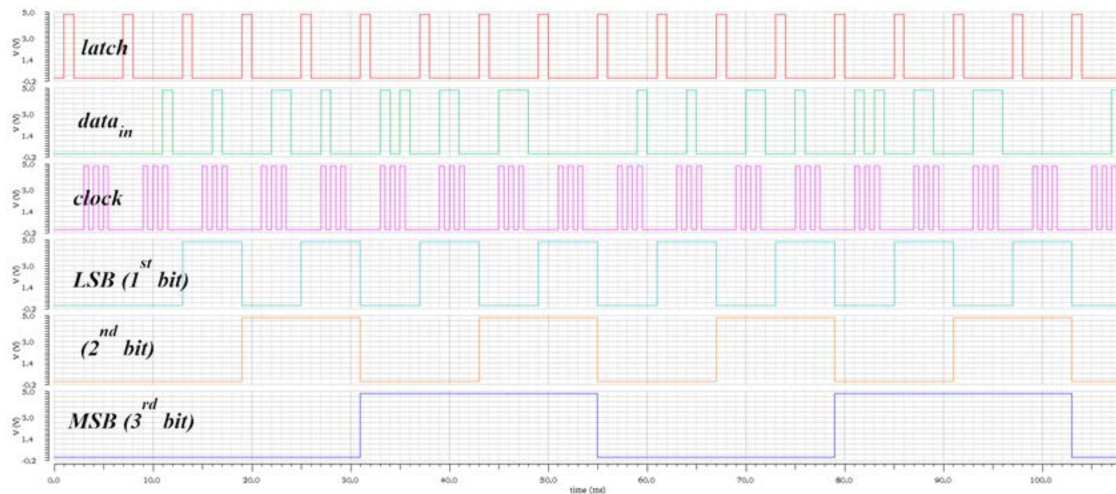


Figure 3-22: Simulation results of a 3-bit shift register.

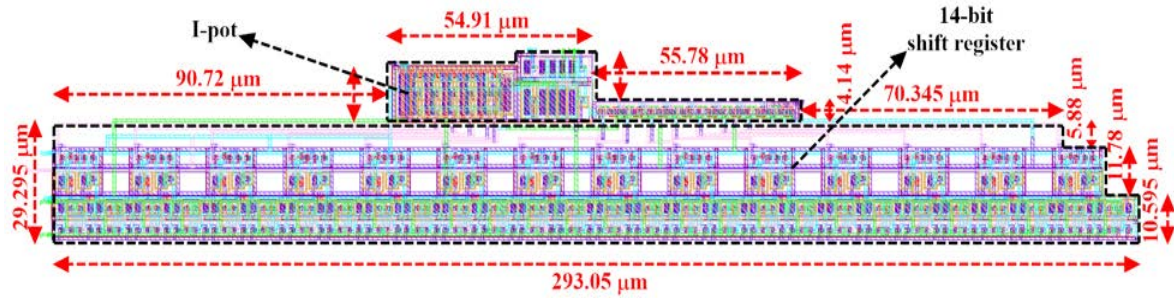


Figure 3-23: Layout view of I-pot with 14-bit shift-register.

of a 3-bit D-flip flop based shift register when loading all possible control-bits in a repeating sequence. After $data_{in}$ is loaded, *clock* is stopped and the *latch* signal is turned ON. This will hold the $data_{in}$ value, which will be passed as digital inputs of the calibration scheme and I-pots. Since complementary control-bits are needed for some digital inputs of I-pots, complementary outputs are also taken from the shift-registers.

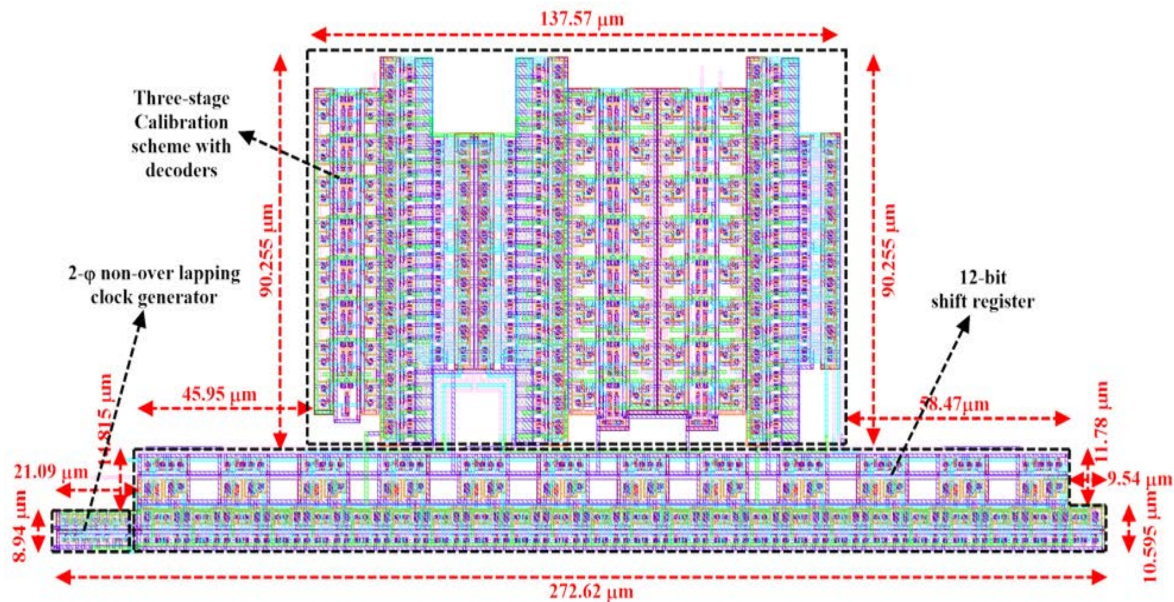


Figure 3-24: Layout view of calibration scheme with its decoders and 12-bit shift-register.

Fig. 3-23 shows the layout view of I-pot and the 14-bit shift-register, which is used to load its control-bits. Fig. 3-24 shows the layout view of the three-stage calibration scheme with its decoders and the 12-bit shift-register, which is used to

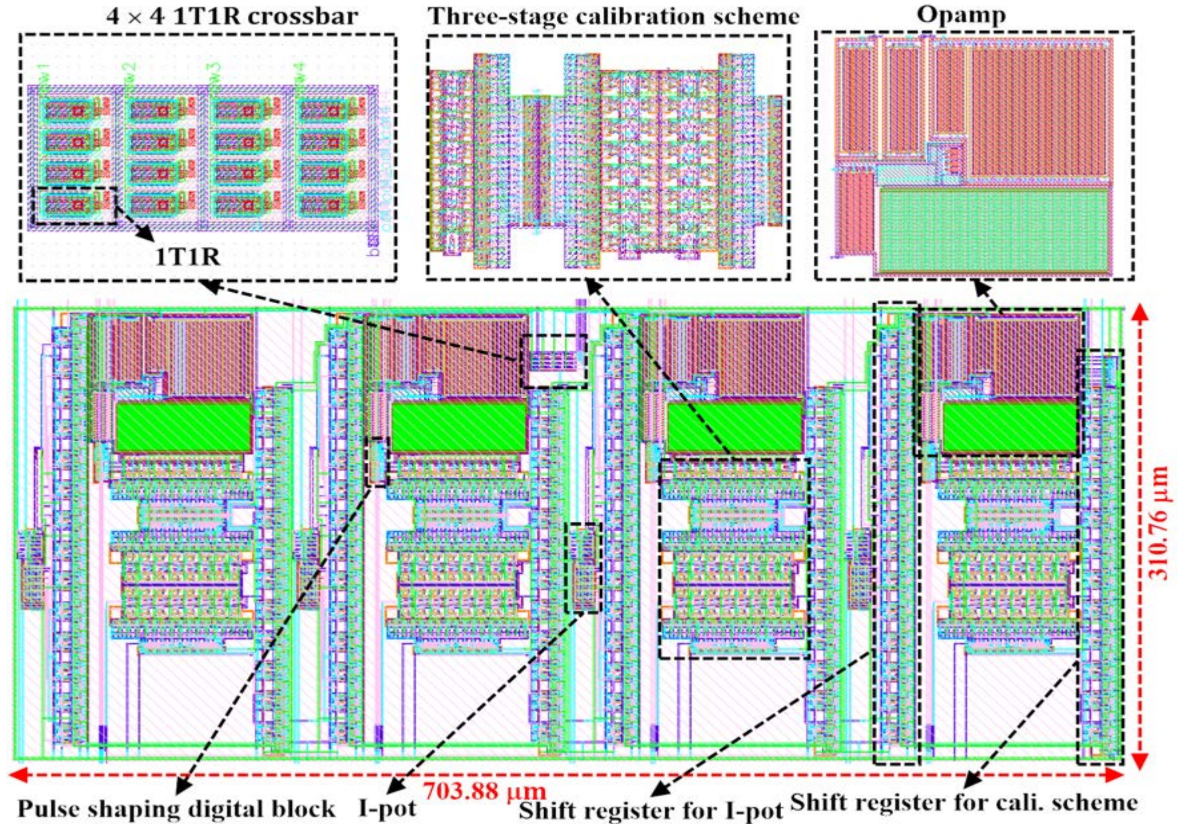


Figure 3-25: Layout view of the three-stage calibration scheme implemented along the wordlines of a 4 × 4 1T1R crossbar with highlighted different sub-circuits in the outer-ring.

load its control-bits. The 2-phase non-overlapping clock generator is shown only in the 12-bit shift-register of calibration scheme, as its output clock signals- c1 and c2 are the clock sources for all flip-flops throughout the shift-register. Moreover, the shift-registers are connected in a fashion such that the 104-bit control word comprises the sequence- 14-bit control lines for I-pot$_1$ of wordline- w_1, 12-bit control lines for calibration circuit$_1$ of wordline- w_1, 14-bit control lines for I-pot$_2$ of wordline- w_2, 12-bit control lines for calibration circuit$_2$ of wordline- w_2, 14-bit control lines for I-pot$_3$ of wordline- w_3, 12-bit control lines for calibration circuit$_3$ of wordline- w_3, 14-bit control lines for I-pot$_4$ of wordline- w_4 and 12-bit control lines for calibration circuit$_4$ of wordline- w_4. Fig. 3-25 shows the layout view of the three-stage calibration scheme implemented along the wordlines- $w_{1,2,3,4}$ of a 4 × 4 1T1R crossbar, where circuits such as I-pot, crossbar, opamp, three-stage calibration scheme, pulse shaping

digital block and shift register are highlighted. The three-stage calibration scheme implemented on a 4 × 4 1T1R crossbar is one of the three circuits, which comprises the outer-ring of the MAD200 chip. Fig. 3-26 shows the layout view of the MAD200 chip, where calibration scheme along with its pads are highlighted in the outer-ring that is duly labeled. Separate *vdd* and *gnd* signal lines are used for analog and digital circuits.

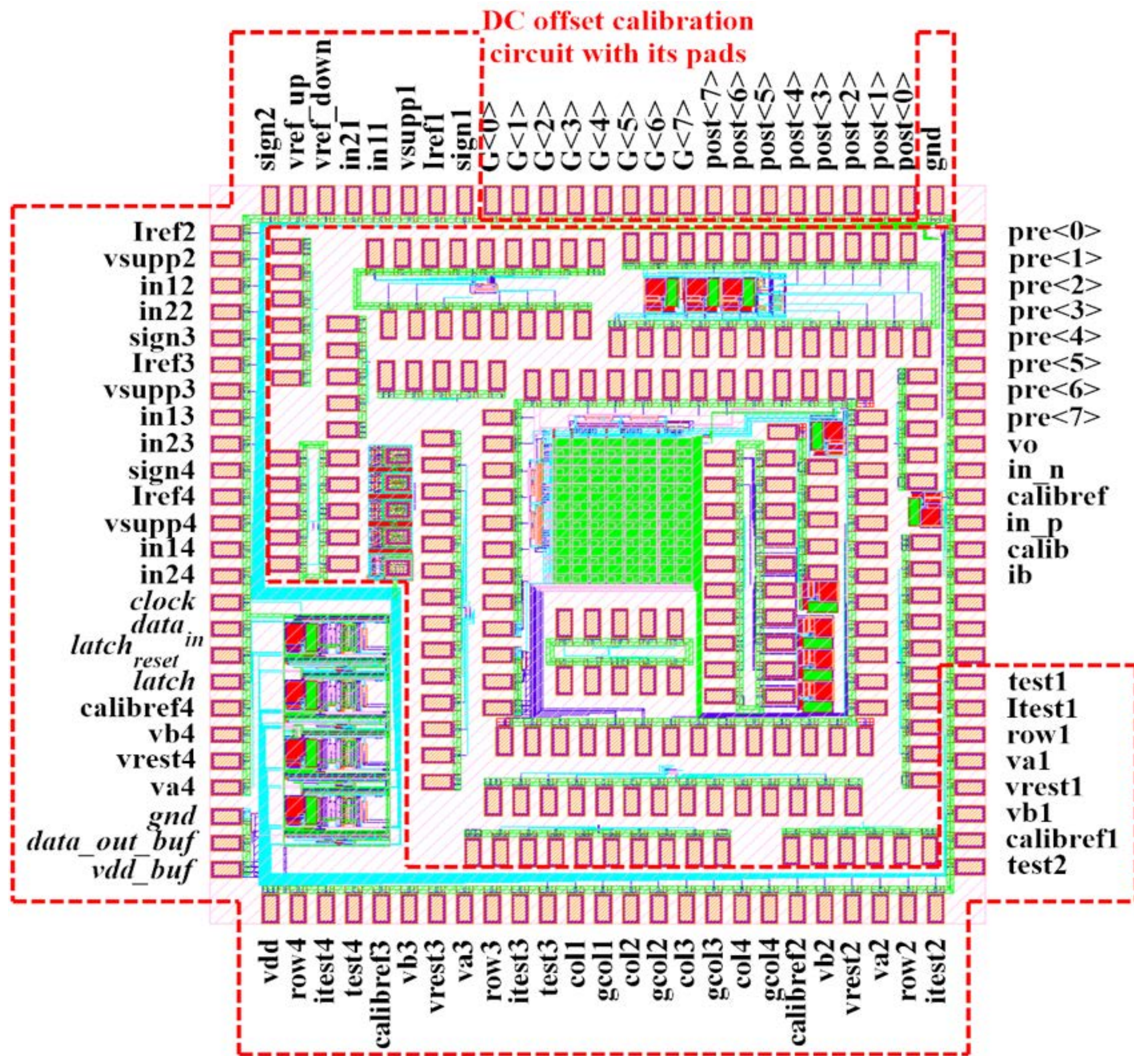


Figure 3-26: Layout view of the MAD200 chip highlighting the bulk-based calibration scheme implemented along the wordlines of a 4 × 4 1T1R crossbar with its pads duly labeled.

3.4 Preparing an experimental set-up for calibration scheme

Once the design of the three-stage calibration scheme implemented on a 4×4 1T1R crossbar is made, the taped-out chip is packaged, which is followed by planning circuits for the test-board to test the packaged-chip, designing the test-PCB, assembling components on PCB, mounting PCB, making auxiliary boards, setting up the experimental test-bench and programming driver to test the chip.

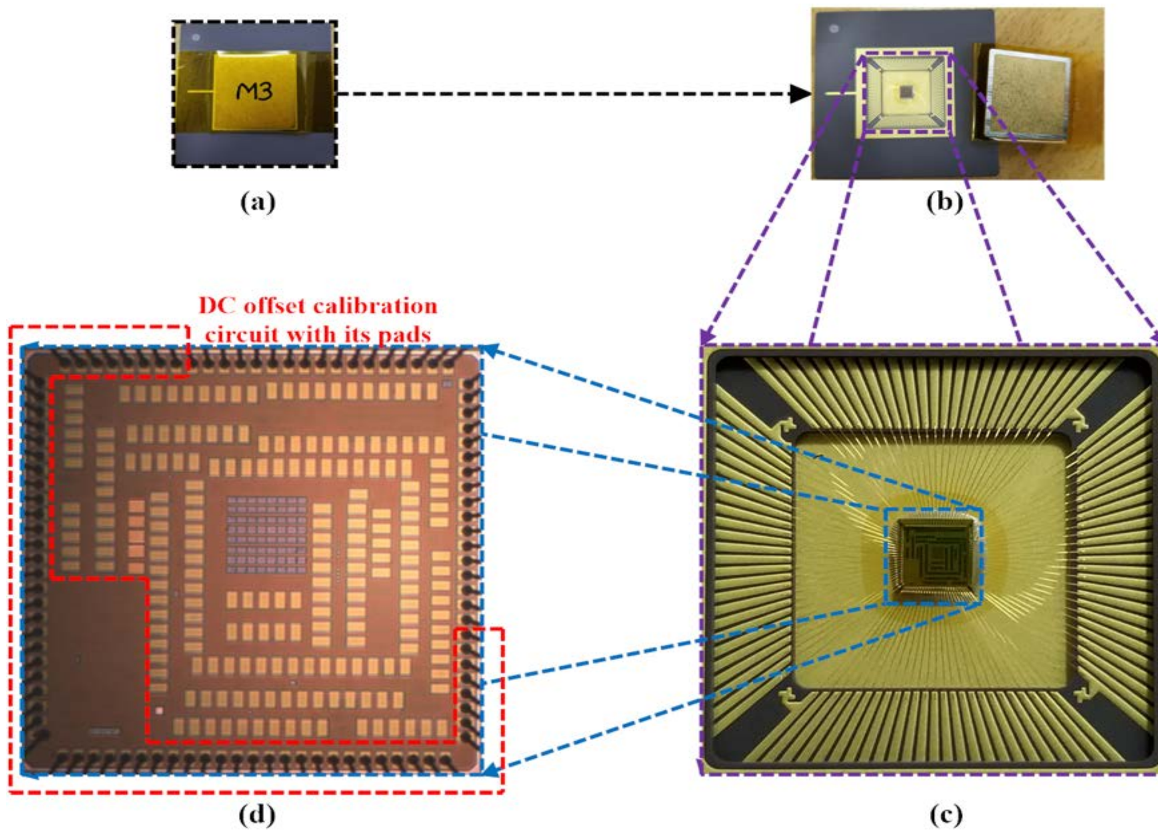


Figure 3-27: (a) Top view of the chip Packaged in PGA100 package, (b) Top view of the packaged chip after gently removing the top wrapper stuck above the package, (c) A zoom preview of the top view of the packaged chip, (d) An ultra-zoom preview of the top view of the chip captured using a microscope with the highlighted offset calibration circuits with its pads.

3.4.1 Packaging of chip

The outer-ring of the MAD200 chip is packaged in the PGA100 package and the top-side of the chip is covered by a protective wrap. Different previews of the top view of the packed chip are shown in Fig. 3-27 (a, b, and c). Fig. 3-27 (d) shows the microscopic top view of the layout of the chip. The packed chip is planned to be rested on a 14 × 14 PGA ZIF socket, which is mounted on a custom-designed PCB, whose details are discussed in Section 3.4.2 and Section 3.4.3. Care is taken that the pin number or address of the packaged chip correctly matches the planned pin number or address of the PGA ZIF socket. The pin addresses for different views of the packaged chip and the ZIF socket are listed in Table A.1 of Appendix A. Different views of the packaged chip and the ZIF socket are shown in Fig. A-11 of appendix A with their pin numbers or addresses duly labeled.

3.4.2 Design of circuits for the test-PCB

Test-circuits are designed to facilitate testing of the designed three-stage DC offset calibration scheme implemented in 4 × 4 1T1R crossbar. Fig. 3-28 shows the functional block diagram of the test-circuits made for testing the three-stage calibration scheme. The main functional blocks in the test-circuit are the opamps, switches, level-shifters, decoders, LVRs, and SPARTAN®-6 FPGA board. The nomenclatures used for the functional block are— 'S' for switches, 'OA' for Operational Amplifiers' and 'LVR' for Linear Voltage Regulator. A detailed schematic of the test-circuits for testing calibration scheme implemented on a 4 × 4 1T1R crossbars is shown in Fig. A-6 of appendix A.

Opamps are used to bias reference voltages signals of calibration scheme such as vref_up (= V_{ref} + V_d), vref_down (= V_{ref} - V_d), and *Calibref{1, 2, 3, 4}*. Opamps are also used to set optimal values of active and default biases for different OxRAM operations such as 'form', 'erase', 'write', 'read', and to keep 'idle' or 'global'[12] values.

[12]'Idle' or 'Global' operation of OxRAM is biasing the crossbar terminals with equal voltage amplitudes and keeping the gate-biases to 0V, so that the targeted OxRAM is not disturbed from its current state.

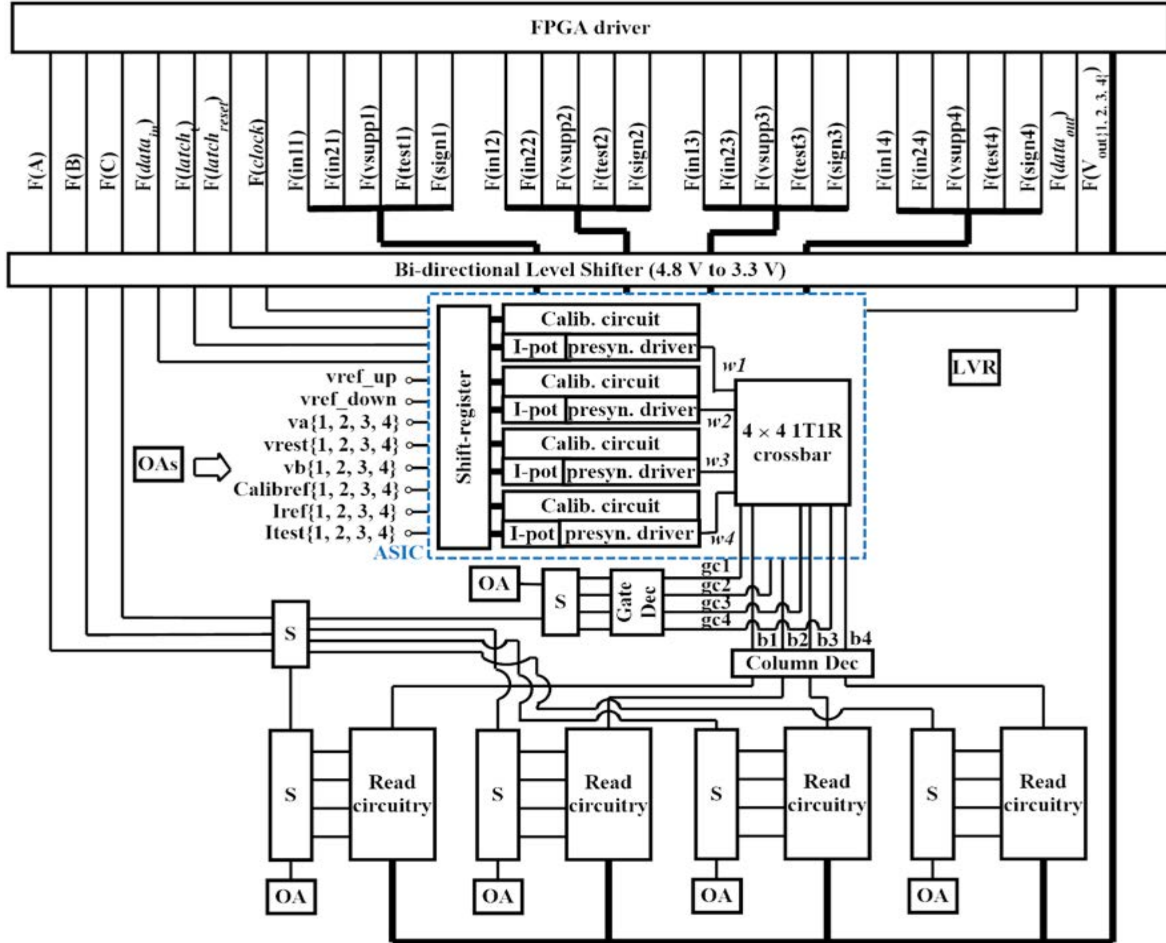


Figure 3-28: Functional block diagram of the test-circuit for testing the three-stage calibration scheme.

Table 3.5: Control-bit for performing different operations for OxRAM in the targeted 1T1R device in the crossbar, whose wordlines are calibrated for DC offset voltage.

OxRAM operation	**F(A)**	**F(B)**	**F(C)**
'Form'	ON	ON	ON
'Erase'	OFF	ON	ON
'Write'	OFF	OFF	ON
'Read'	OFF	OFF	OFF
'Idle' or 'Global'	ON	OFF	ON

For wordlines, the biases for different OxRAM operations are directly applied across the terminals- va{1, 2, 3, 4}, vrest{1, 2, 3, 4} and vb{1, 2, 3, 4}, which are digitally controlled using signals- F(in11), F(in21), F(in11), F(in21), F(in13), F(in23), F(in14), F(in24) and F(supp{1, 2, 3, 4}). It is to be noted that all four terminal of va{1, 2,

3, 4} are shorted on PCB such that va{1} = va{2} = va{3} = va{4}. The same is done for signals– vrest{1, 2, 3, 4}, vb{1, 2, 3, 4} and *Calibref{1, 2, 3, 4}*. For gates and bitlines, the biases for different OxRAM operations are applied via switches and by using decoders. Digital circuits (such as gates and switches) for scrutinizing the needed opamp-biases across the wordlines- are inside ASIC whereas, digital circuits for scrutinizing the needed opamp-biases across gates and bitlines are included in the test-PCB. Switches (such as SPST and SPDT) and decoders are connected between the opamps and the gate or bitline terminals, which are used to choose default and active gates and bitlines. They are also used to apply the desired operation on the targeted 1T1R device. For this purpose, the dedicated 3-bit control line- F(A, B, C) is used, whose possible control-bits and corresponding OxRAM operations are

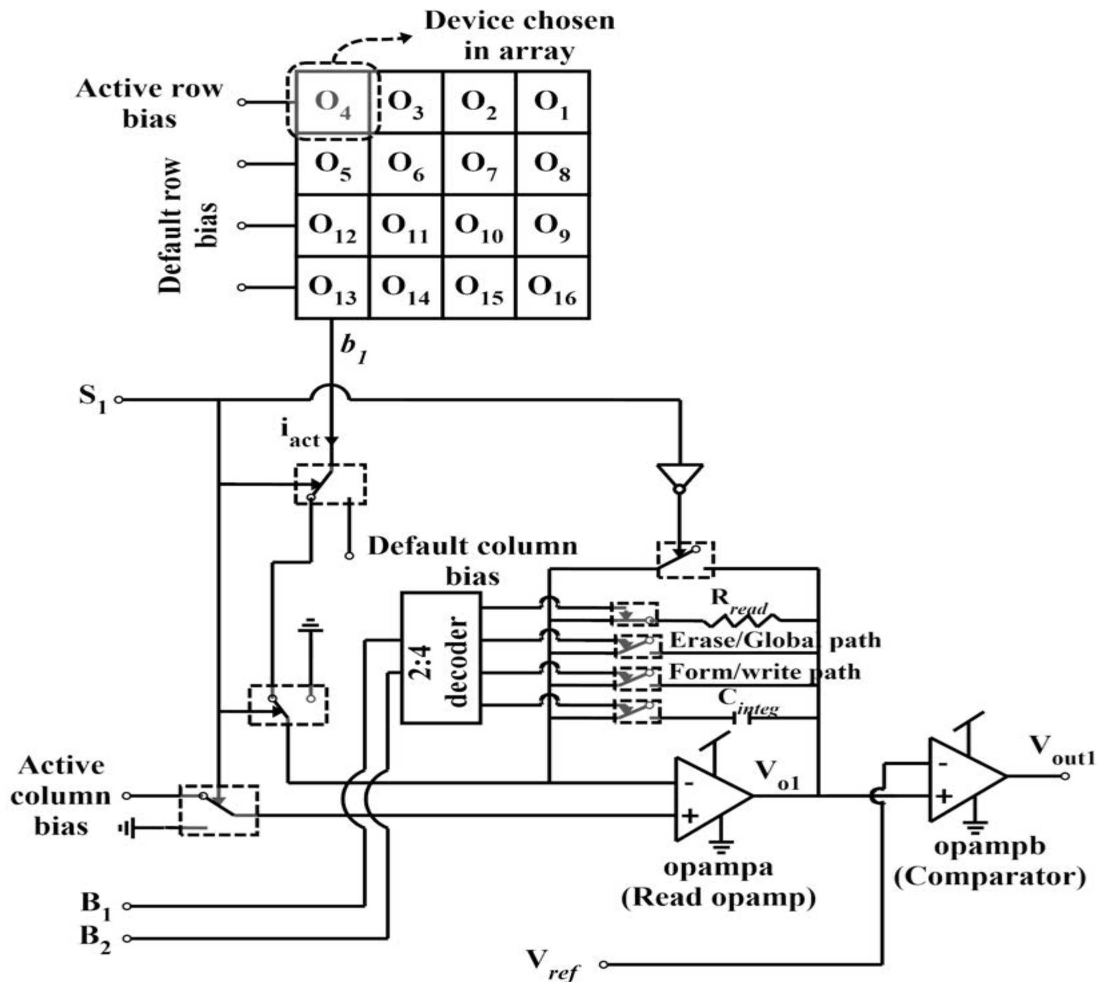


Figure 3-29: Scheme of the read circuitry implemented for the calibration scheme.

shown in Table 3.5. Here, 'ON' indicates 4.8 V (vdd) and 'OFF' indicates 0 V. Spartan®-6 FPGA board is used to program and digitally control the PCB (that has the test-circuits) and the ASIC part. For wordlines, the three-bit control is established by programming the digital lines of the pulse-shaping digital block through FPGA. Different supply voltages such as 10 V, 3.3 V and 4.8 V for the test-circuits are ensured by the LVR and level-shifters are used for bi-directional voltage level conversion between 4.8 V and 3.3 V, which is needed when using FPGA driver for controlling the overall testing of the chip. A read circuitry is made at each bitline, whose scheme is shown in Fig. 3-29. It mainly comprises switches, decoders, and opamps. Opampa (or Read opamp) and control lines- $B_{1,2}$ are used to properly set the value of the feedback component during OxRAM operations- such as 'form', 'erase', 'write', 'read'. The objective is to keep the needed bias of the inverting terminal of opampa or bitline, b_1 for various OxRAM operations, as shown in Table 2.2 of Section 2.3.4 in Chapter 2. During a 'read' operation, opampa can also be used as an integrator, whose output is compared with V_{ref} by a comparator (using opampb). Here R_{read} = 26.71 kΩ and C_{integ} = 1 µF. Here S_1 is chosen by a bitline decoder (not shown in Fig. 3-29). When bitline, b_1 is chosen as active bitline, signal S_1 is kept 'ON'. This connects the bitline, b_1 to 'active column bias'. Alternatively, when bitline, b_1 is chosen as default bitline, signal S_1 is kept 'OFF'. This connects the bitline, b_1 to 'default column bias'. Both 'active column bias' and 'default column bias' for different OxRAM operations are established through switches and decoders in such a way that anyone of the bitlines will be kept as active thereby, leaving the rest to default at an instant. The control lines- F(A, B, C) and $B_{1,2}$ together facilitate choosing the desired OxRAM operation and keeping appropriate biases across the bitlines.

3.4.3 Assembly and mounting of test-PCB and auxiliary boards

A test-PCB is made according to the test-circuits described in the detailed Fig. A-6 (of Appendix A). A four layered-PCB is designed using OrCAD® CIS[13] and Allegro®[14], whose design details are described in appendix A.2. A three-stage bulk-based calibration scheme is part of the circuits in the outer-ring. The components used for the PCB are chosen carefully by going through their data-sheets. Parameters such as, operating voltage range, bandwidth, switching characteristics, etc. are considered for choosing the components and for those components which have PSpice[15] model— are simulated to visualize their characteristics. The main purpose of the PCB is to assure desired analog biases at specific terminals of the chip, which are controlled by switches and digital circuits.

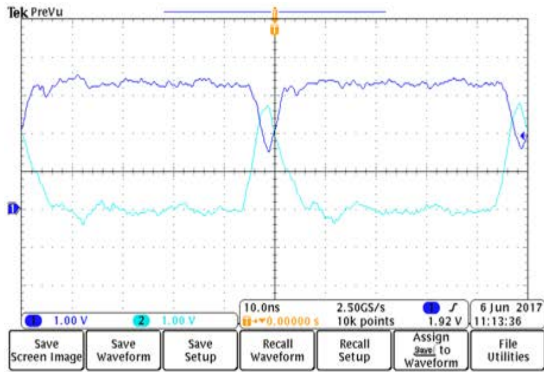
Once the PCBs are designed the resulting gerber files are sent to a PCB manufacturing company and the components in the BOM list are ordered. The PCB components are soldered to the PCB and the PCB is mounted using corner screws. The assembled and mounted PCB is shown in Fig. A-10 (of appendix A). Some guidelines for a better PCB assembly and mounting practice are narrated in appendix A.3. Along with the PCBs, few auxiliary boards are made to avail of some buttons (to facilitate programming of driver) and to ease probing of terminals of the crossbar. A button-board is made and added to the experimental set-up, which has header-pins with jumpers, exclusive buttons, and level-shifting ICs. The header-pins with jumpers are used to target a synapse in a crossbar and to provide input bits to the calibration scheme. The buttons are programmed to do OxRAM operations like 'form', 'erase', 'write', 'read' or keep 'idle' or 'global' in a particular sequence. The level-shifters on button-board are used to convert the output of the comparator from 4.8 V to 3.3 V

[13]OrCAD® Capture is one of the most widely used schematic design solutions for the creation and documentation of electrical circuits. Coupled with the optional OrCAD CIS product for component data management, the designer can use components in the schematic that can be used later in Allegro® PCB designer suite to design PCBs.

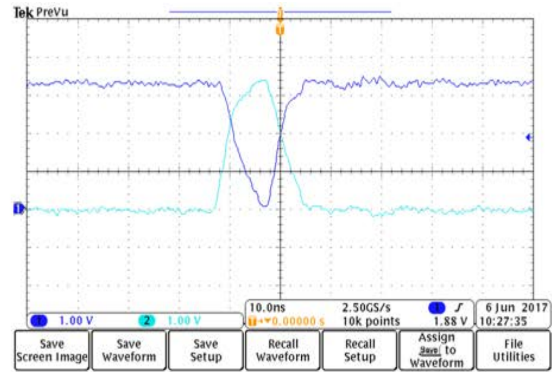
[14]Allegro® PCB Designer of Cadence® is a scalable, proven PCB design environment that addresses technological and methodological challenges thereby, making the design cycles shorter and predictable.

[15]PSpice models of more than 33k components can be simulated in Cadence® before making PCB layouts to observe their working characteristics at circuits-level.

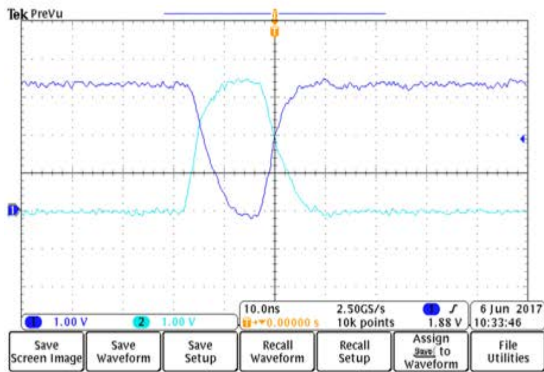
for being used in FSM during programming.



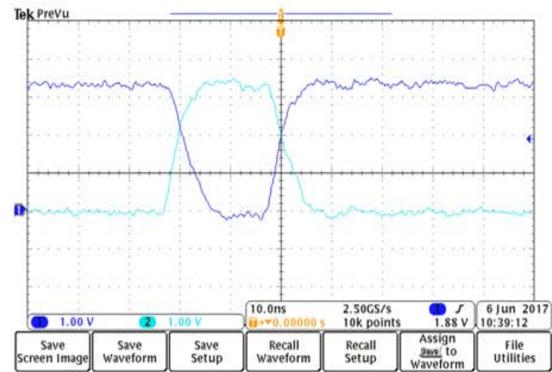
(a)



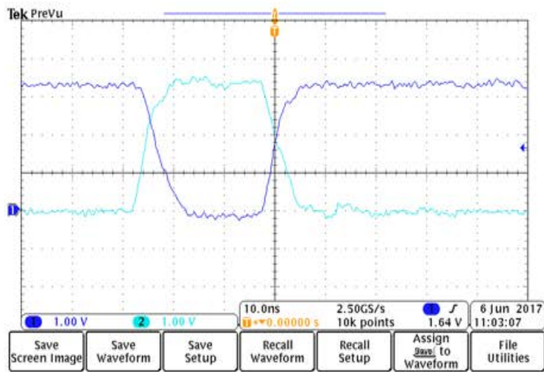
(b)



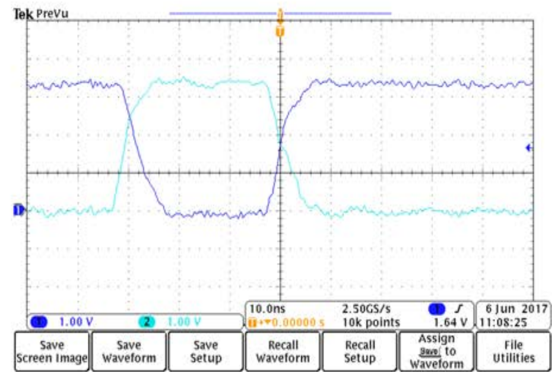
(c)



(d)



(e)



(f)

Figure 3-30: Comparison of ON/OFF pulses of different widths: (a) 5 ns ON/OFF pulse, (b) 10 ns ON/OFF pulse, (c) 15 ns ON/OFF pulse, (d) 20 ns ON/OFF pulse, (e) 25 ns ON/OFF pulse, (f) 30 ns ON/OFF pulse.

3.4.4 FPGA SPARTAN®-6 driver board

We need a driver to program the test-PCB to test the chip which rests on the ZIF socket of the PCB. Drivers such as microcontroller- STMF4DISCOVERY kit [124], FPGA SPARTAN®-3 [125], and FPGA SPARTAN®-6 [126] are considered as possible programming platforms for controlling the test-board digitally.

It is observed that a full digital strength of 3.3 V can be reached at about 10 ns fast pulse using Spartan®-6 FPGA. Fig. 3-30 shows a comparison of ON/OFF pulses of different widths (5 ns, 10 ns, 15 ns, 20 ns, 25 ns, and 30 ns) programmed using Spartan®-6 FPGA. Here, in Fig. 3-30 (a) the purple-colored trace does not working correctly for the 'OFF' pulse, as it did not reach the magnitude equivalent to the 'ON' pulse of the cyan-colored trace. In our experiments Spartan®-6 FPGA driver (also called as 'Node Board'[16] [127]) is chosen to program the test-PCB, due to its configurable speed (programming fast pulses), availability of more number of programmable pins and features like QUIETIO mode (slew rate option), which keeps low peak over-shoot and less ringing. The filter in the oscilloscope probes is also tuned to have smoothly settled voltage pulses with no or minimal overshoot or ringing.

3.5 Description and working of the experimental set-up of the calibration scheme

Fig. 3-31 shows the experimental setup of the DC offset calibration scheme. It mainly comprises the test-PCB that includes the chip under test, a SPARTAN®-6 driver board, a button-board, a resistor plug-and-play board, a mixed-signal oscilloscope, and its digital pod. The test-PCB is controlled through the SPARTAN®-6 driver board. The button-board has dedicated buttons to perform OxRAM operations in a sequence- like 'Form_Global_Read', 'Erase_Global_Read', 'Write_Global_Read', 'Read' and to calibrate DC offset during 'Read'. These tasks or operations are highlighted in Fig. 3-31 for the corresponding buttons on the button-board. The button-

[16] The node board contains a XC6SLX150T Spartan®-6 FPGA and four SATA connectors that renders 76 programmable pins.

board and SPARTAN®-6 driver board are used together for two main purposes: (i) Target a synaptic 1T1R device in the crossbar and perform operations like 'form', 'erase', 'write' and 'read' through a 3-bit control signal, F(A, B, C) and (ii) Set the 12-bit control word for calibration scheme and perform calibration of DC offset during a 'read' operation. The resistor plug-and-play board is used to facilitate easy probing of the crossbar terminals and to test with resistor-banks before testing the chip. The SPARTAN®-6 driver board is powered by a 5 V supply line with the short ground to keep the influence of noise low. Calibration can be done once in the beginning irrespective of the OxRAM operation. As we intend to investigate how small 'Read' pulses can be used and how we can calibrate DC offset during such tiny 'read' pulses, we are doing DC offset calibration during a 'Read' operation.

An FSM is programmed in FPGA using VHDL language to define functions of each button on the button-board. Different states are made in FSM as shown in Fig. 3-32 in such a way that push-buttons on the button-board are programmed to establish different OxRAM operations in a sequence on the targeted 1T1R device. One of the push-buttons is programmed to do 'Form_Global_Read' task, while another button is programmed to do 'Erase_Global_Read' task. Other buttons are exclusively

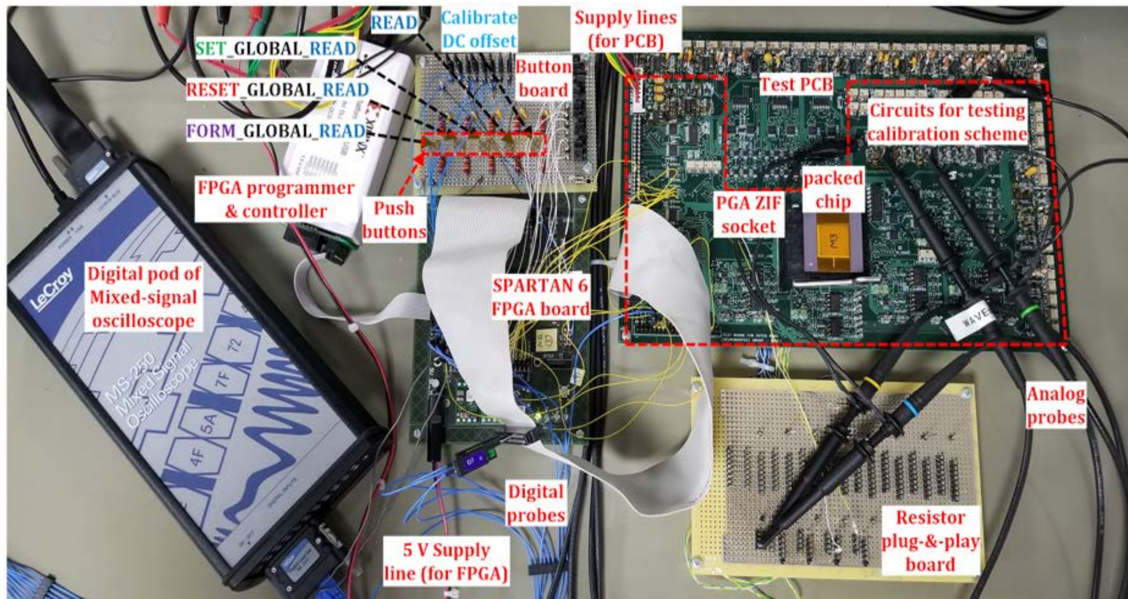


Figure 3-31: Experimental set-up of the DC offset calibration scheme.

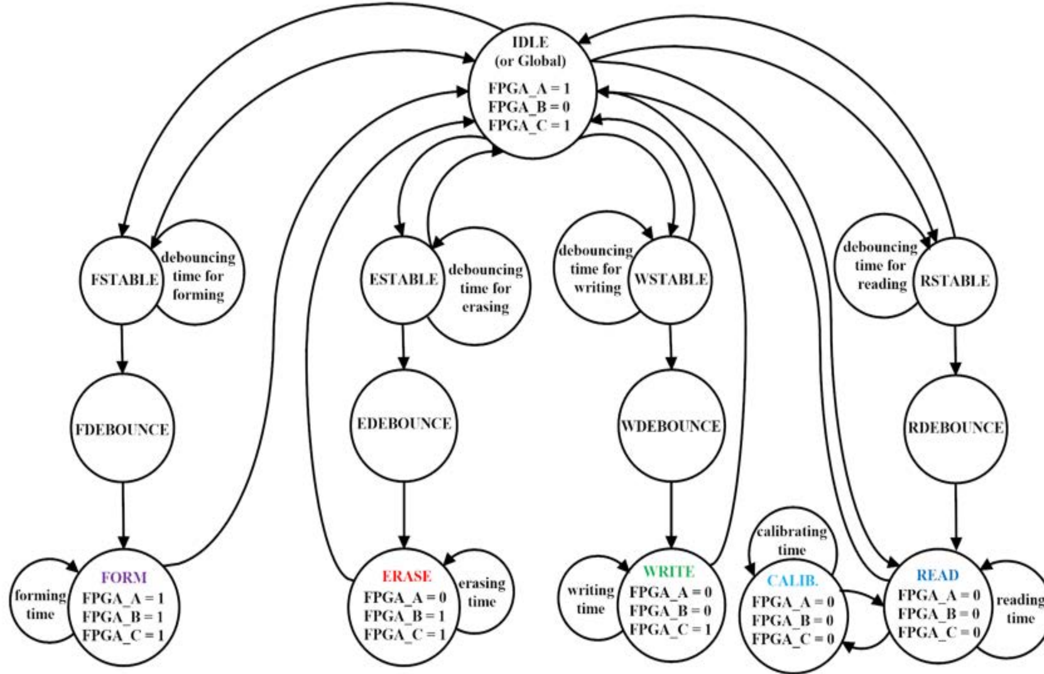


Figure 3-32: FSM used to calibrate DC offset voltage and perform OxRAM operations.

programmed to do 'Write_Global_Read', a separate 'Read' operation and to perform offset calibration during 'Read' operation. Before calibrating offset, the control-bits for calibration scheme are set as desired by using jumpers on the auxiliary boards. The push-buttons inherently have bouncing effects [107] for a short time when immediately pressed and released and these can easily cause short unwanted transients to the crossbar terminals, which can be harmful to OxRAMs and hence they must be eliminated. To overcome this, a software-based debouncing technique is used, which is waiting for some time until bouncing finishes and settles to start the next state.

3.6 System-level simulation results for pattern recognition

Before starting with the experimental application on an offset calibrated 4×4 crossbar for pattern recognition, it is important to verify the results by simulation. Two system-level simulations are carried out on the crossbar for pattern recognition. The

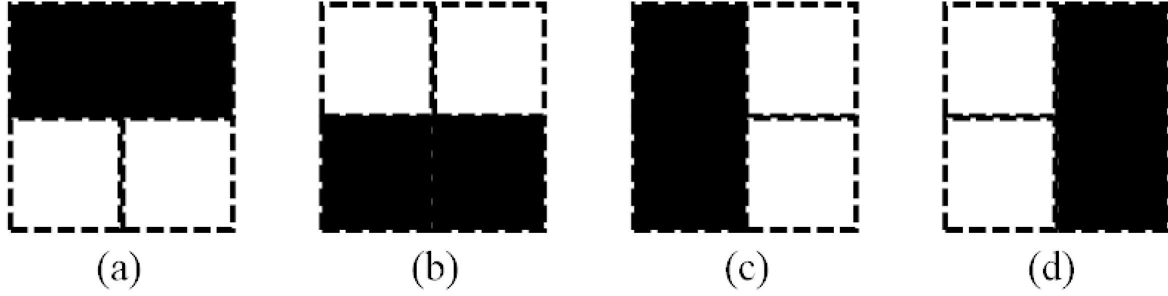


Figure 3-33: Patterns used for recognition-task using 4×4 crossbar: (a) Pattern-1, (b) Pattern-2, (c) Pattern-3, (d) Pattern-4.

first method is by using Supervised Single-Shot Programming (SSSP) for template matching. Here, there isn't any learning of features, rather a weight-update is done to all synapses for specific post-synaptic spikes to fire earlier. The second method is by using STDP learning, where the weights are changed based on the STDP learning rule. STDP is a learning mechanism by which, change in synaptic weight is a function of the time difference between the pre-synaptic spike and post-synaptic spike [78, 79]. The patterns (Pattern-1, Pattern-2, Pattern-3, and Pattern-4) considered are shown in Fig. 3-33 and a model-based design using Simulink of MATLAB® is used to verify the results.

Fig. 3-34 shows the behavioral system-level simulation for pattern recognition implemented on a 4×4 crossbar. The main blocks are the pattern generator, crossbar, integrators & comparator block, scopes or displays, and STDP processor. The pattern generator is used to feed the pixels of the patterns in the form of 'read' pulses- tpre{a, b, c, d}. The read pulses or the pre-synaptic pulses are fed into the system (across wordlines, $w_{1,2,3,4}$ of the crossbar) in the sequence as they appear in Fig. 3-33. The read pulses have an amplitude of 0.3 V and a pulse width of 50 ms. The total time of each pixel is 200 ms. Since there are 4 pixels in each pattern, 800 ms is used to feed all the pixels of the patterns. The crossbar comprises wordlines, $w_{1,2,3,4}$ and bitlines, $b_{1,2,3,4}$ interconnected by synapses, W{1, 2,...16}.

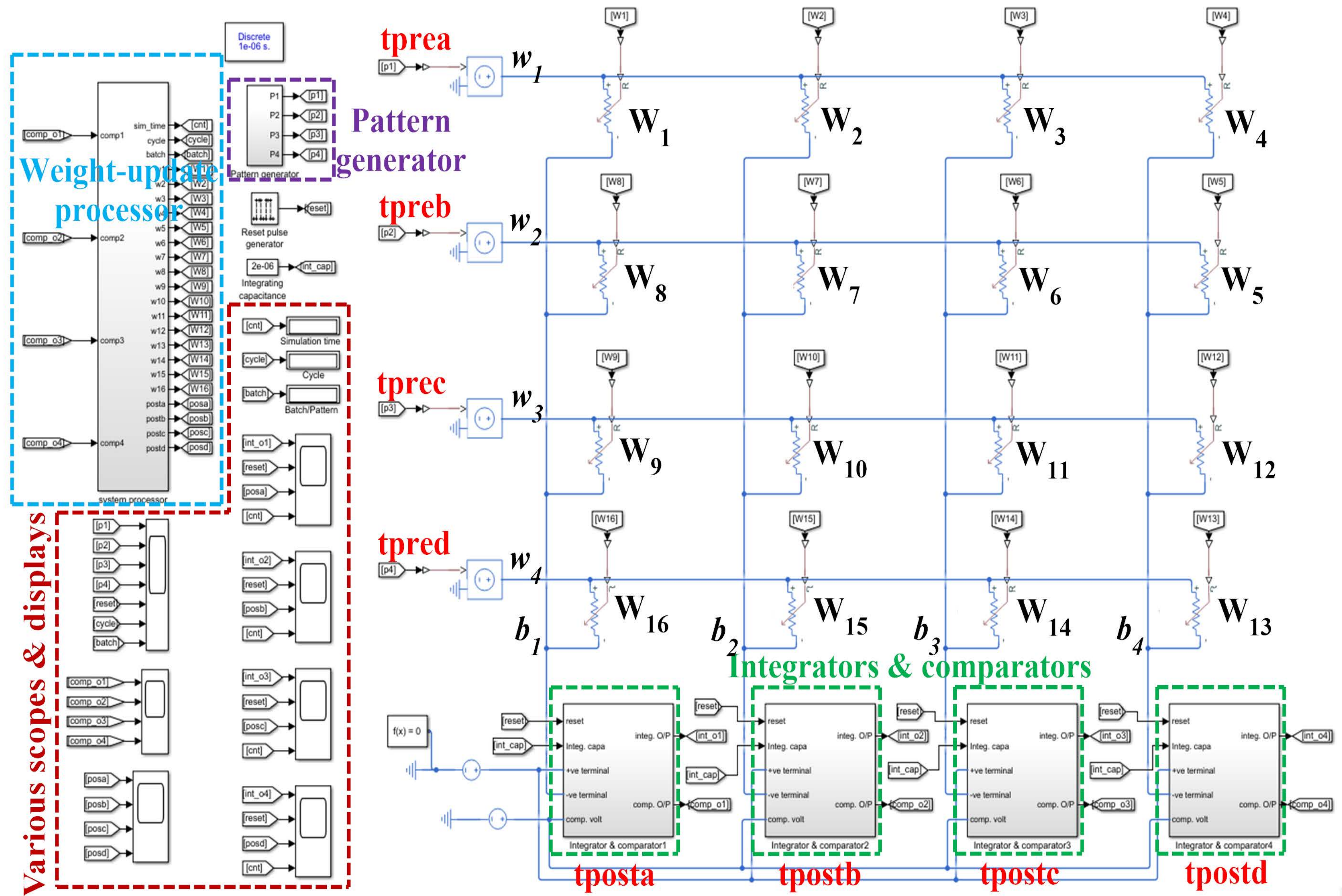


Figure 3-34: Model based simulation environment implemented in Simulink environment.

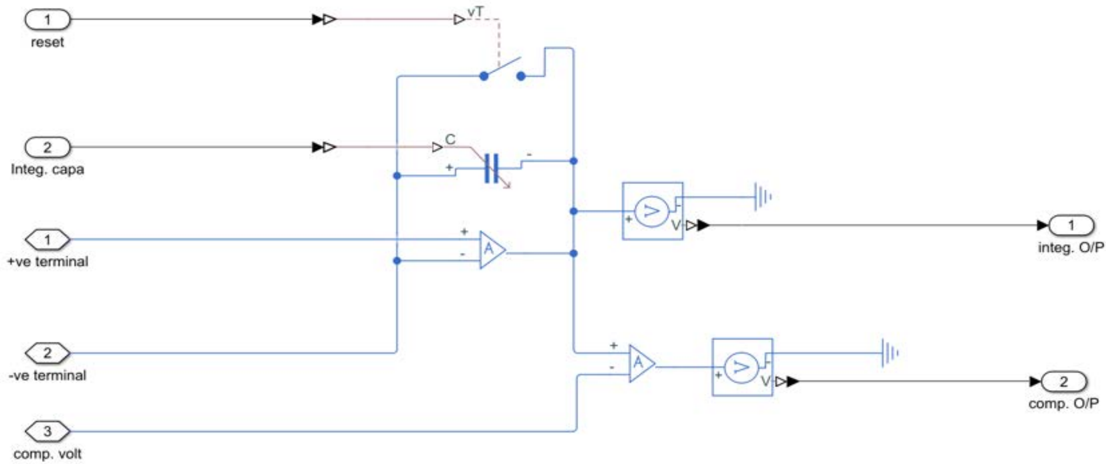


Figure 3-35: Scheme of integrator and comparator implemented in Simulink environment.

The synapses have binary weights and they can be either in LRS or HRS. Fig. 3-35 shows the integrator & comparator block, which mainly comprises an integrating opamp, integrating capacitor (of capacitance 2 µF) across a reset switch, an opamp used as a comparator, voltage sensors, and converters to convert physical signal to a simulink signal and vice versa. The accumulated currents across bitlines are integrated, whose output voltage is compared with a reference voltage to generate the post-synaptic pulses- tpost{a, b, c, d}. A reset signal is applied at the end of each pixel to reset the integrator so that it starts integrating current for the next pixel.

3.6.1 Using Supervised Single-Shot Programming (SSSP)

The Supervised Single-Shot Programming (SSSP) for template matching is based on both - (i) updating weights by determining the time of occurrence of pre-synaptic and post-synaptic pulses and (ii) doing opposite weight update for the rest of the synapses that don't contribute to the targeted post-synaptic spike. The times of occurrence of tpost{a,b,c,d} and tpre{a,b,c,d} pulses are stored in memory for each pixel throughout the simulation. Hence, we will have 4 values of times of occurrence for each

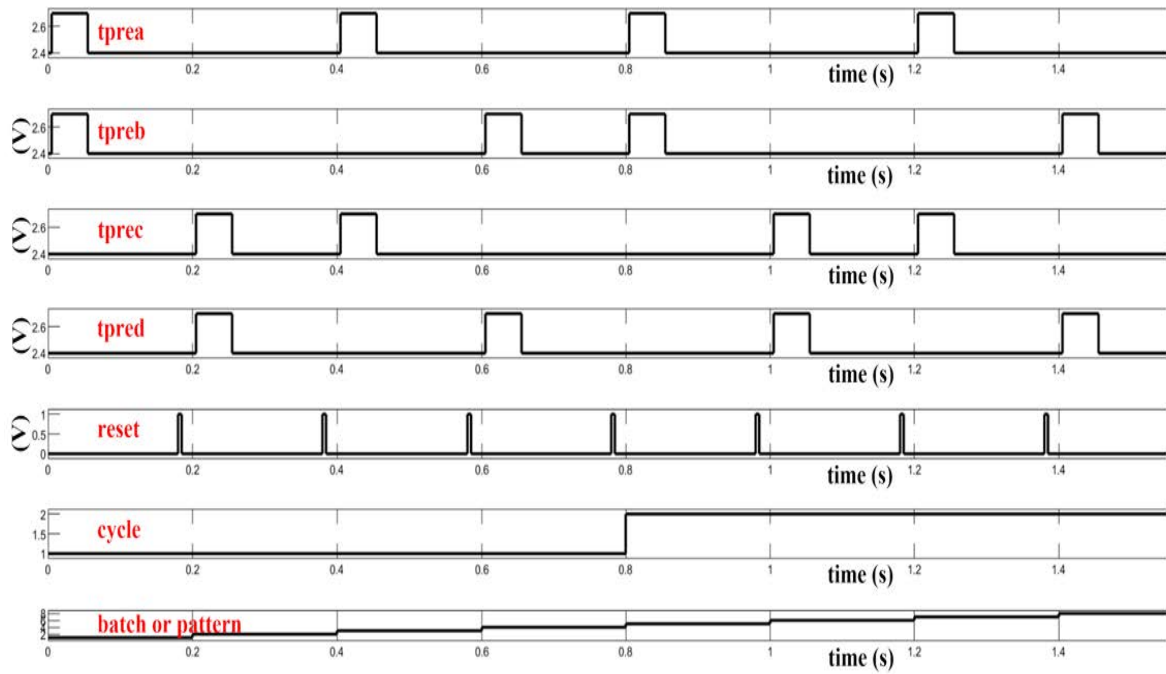


Figure 3-36: Pre-synaptic pulses, reset, cycle and batch signals.

post-synaptic pulse when the whole pattern is fed into the system. The internally stored time of occurrence of tpost{a,b,c,d} is compared with the time of occurrence of

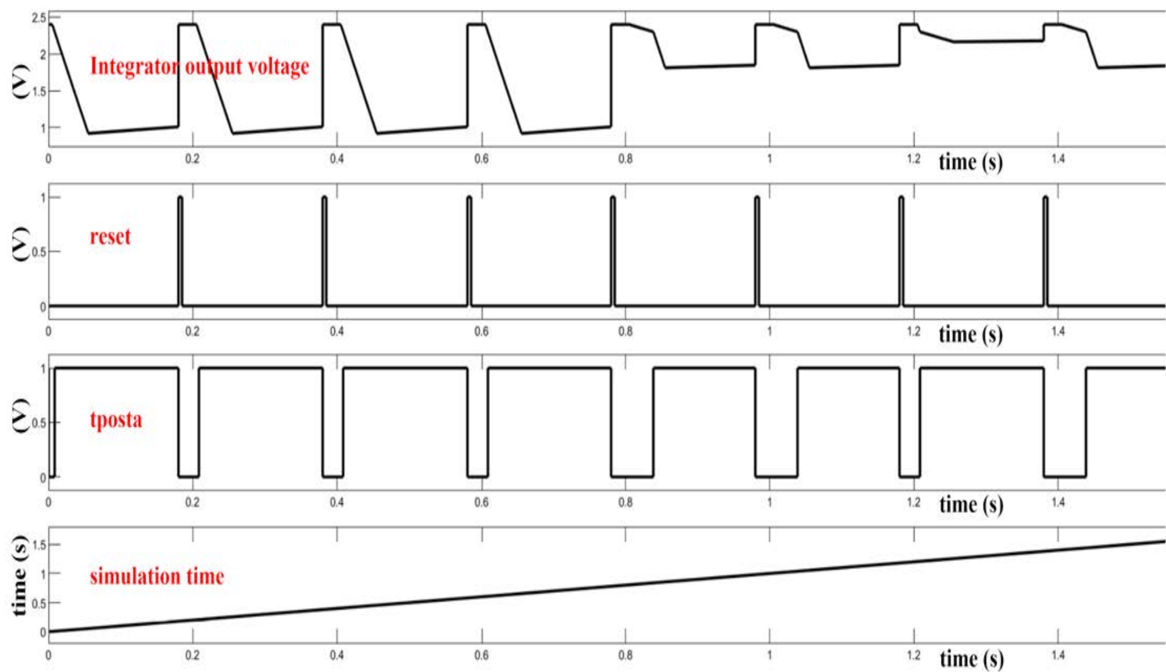


Figure 3-37: Integrator output voltage, reset, tposta and simulation time signals.

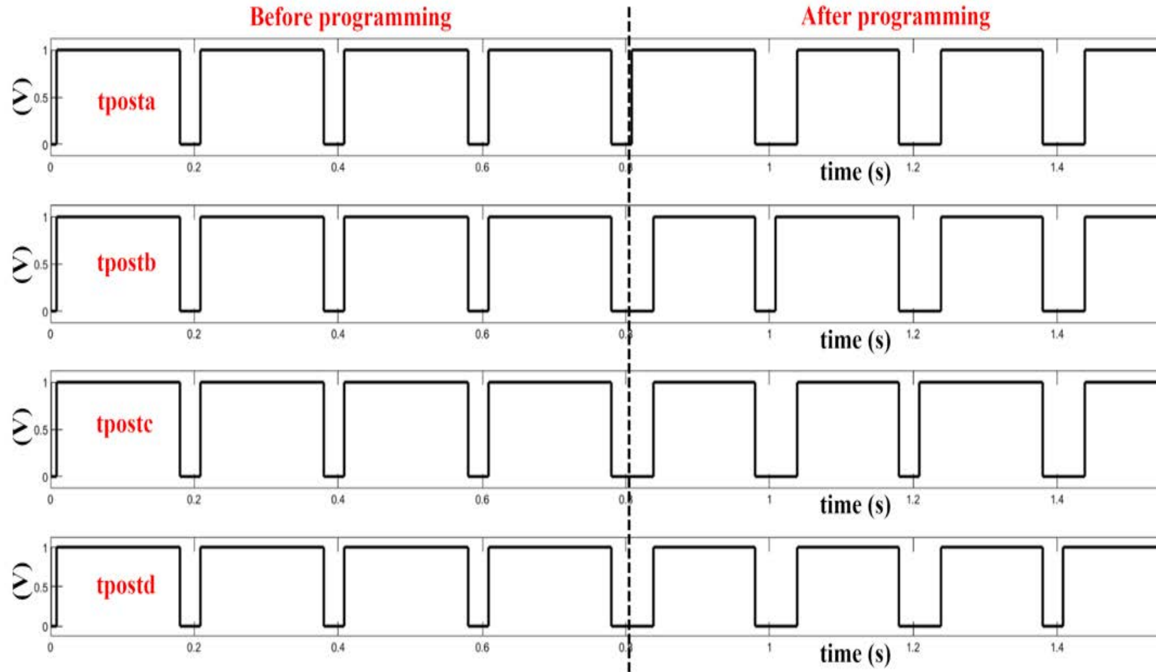


Figure 3-38: Post-synaptic pulses- before and after programming.

its corresponding tpre{a,b,c,d} and the appropriate weight W{1,2,. . . 16} is updated in the weight-update processor based on certain priorities. When two post-synaptic spikes generate at the same time, the priority is given to those with 'minimum index' and those which didn't get programmed earlier. We target the post-synaptic pulses- 'tposta' to spike earlier for 'pattern-1', 'tpostb' to spikes earlier for 'pattern-2', 'tpostc' to spike earlier for 'pattern-3' and 'tpostd' to spike earlier for 'pattern-4'. This is done by doing weight update to the synapses that are responsible to make a particular post-synaptic pulse to spike earlier for a given pattern. For the same pattern, the opposite weight update is done for the remaining synapses. For example- the synaptic weights- W{1,8,9,16} mainly contribute to tposta (which is the targeted bitline for 'pattern-1'). We want to program tposta spike earlier for pattern-1 and so we apply weight update on these synapses for pattern-1. Here, if ΔT = tposta - tprea is positive, a 'write' operation is performed on the synapse- W{1}, and when ΔT is negative an 'erase' operation is performed on W{1}. A similar weight update is applied on W{2} by computing (tposta - tpreb) and comparing its value with 'zero'. For the same pattern-1, the rest of the weights- W{9,16}, which don't contribute to

causing post-synaptic spike will face opposite weight update. For the same pattern if we move to other bitlines of the crossbar, the synapses will have an opposite weight update relation when compared to the targeted bitline. This will cumulatively make tposta to spike earlier. Similarly, the rest of the post-synaptic pulses are trained to spike earlier for their corresponding patterns.

The synapses are initially set to random weights i.e. either LRS or HRS. Two cases have been further considered in the simulation. The first case is an ideal case where LRS = 10 kΩ and HRS = 100 kΩ are considered. The second case is the non-ideal case that has mismatch variability taken into account which is obtained from the experimental results of endurance tests, similar to the results obtained in Fig. 2-28 and Fig. 2-27 in chapter 2. The second case is considered to test the tolerance of the system during the variability of resistances. It is observed that in both cases, we can program a particular post-synaptic pulse to spike earlier for a given pattern. Fig. 3-36 shows the pre-synaptic pulses for the input pattern (as shown in Fig. 3-33), cycles, and batch. It also shows the reset pulse of the integrator, which is applied after each pixel input of the pattern. Fig. 3-37 shows integrator output voltage, reset, and post-synaptic pulse of bitline, b_1. Fig. 3-38 shows the post-synaptic pulses before and after programming. Here, tposta spike faster by getting trained for Pattern-1 in one cycle. Likewise, tpost{b, c, d} also spikes faster by getting trained for Pattern-{2, 3, 4} respectively.

3.6.2 Using STDP learning rule

Pattern recognition using the STDP learning rule is based on- updating the weights of the synapses by STDP rule i.e. by determining the time of occurrence of pre-synaptic pulse and post-synaptic pulse. When the post-synaptic pulse spike after the pre-synaptic pulse, the weight of the corresponding synapse is strengthened by decreasing the resistance. Alternately, when the pre-synaptic pulse spike after the post-synaptic pulse or when there is no pre-synaptic pulse, the weight of the corresponding synapse is weakened by increasing the resistance. This weight update is done in step for the synapses in the crossbar such that the weights evolve from 'random values' to reach

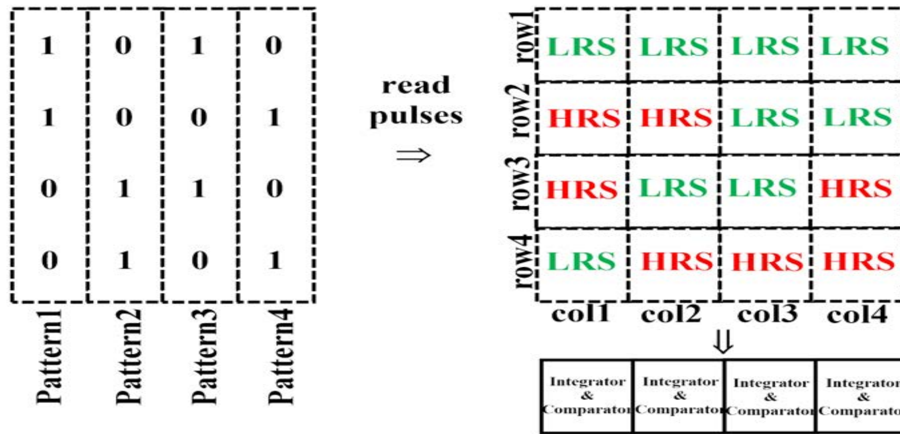


Figure 3-39: Conceptual diagram showing patterns applied on a crossbar that has random weights.

'learned values'. The main difference between pattern recognition using the earlier approach and the STDP learning rule is- in SSSP all synaptic weights are changed during a weight update whereas, in STDP, weight update is done in steps only to the synapses that contribute post-synaptic pulse to spike earlier. This causes the weights to learn gradually. The former approach is more or less like supervised one-time programming while the latter is unsupervised learning.

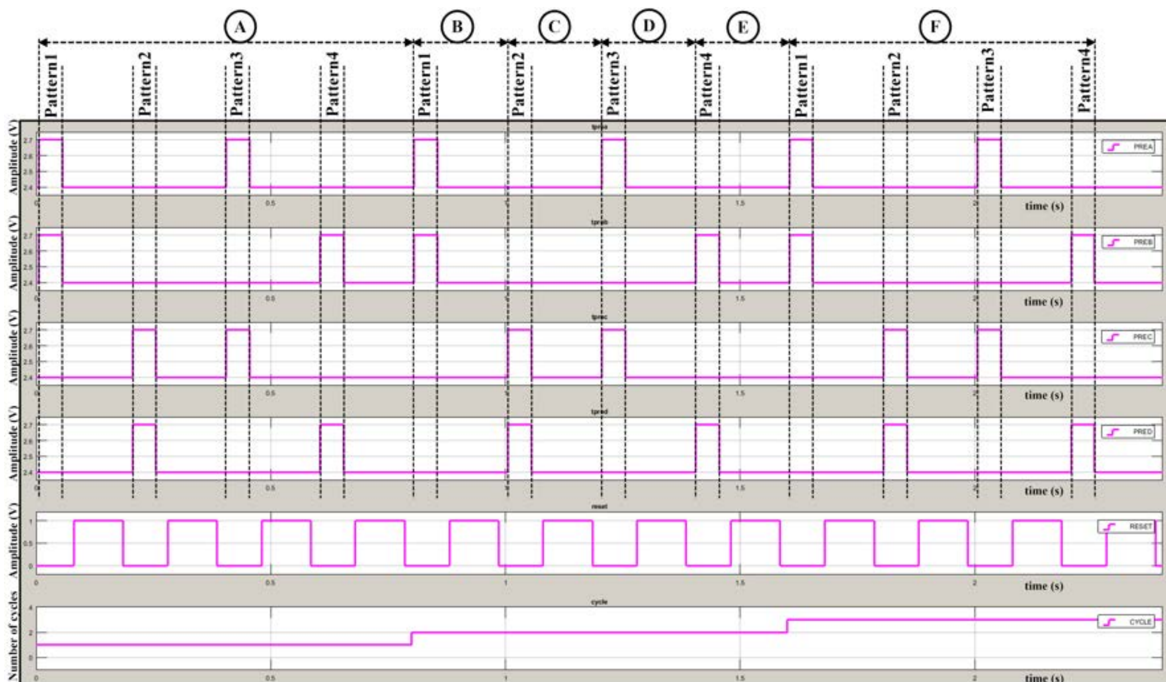


Figure 3-40: Simulated Pre-synaptic pulses, reset signal and number of cycles showing different regions- A, B, C, D, E and F using STDP learning rule.

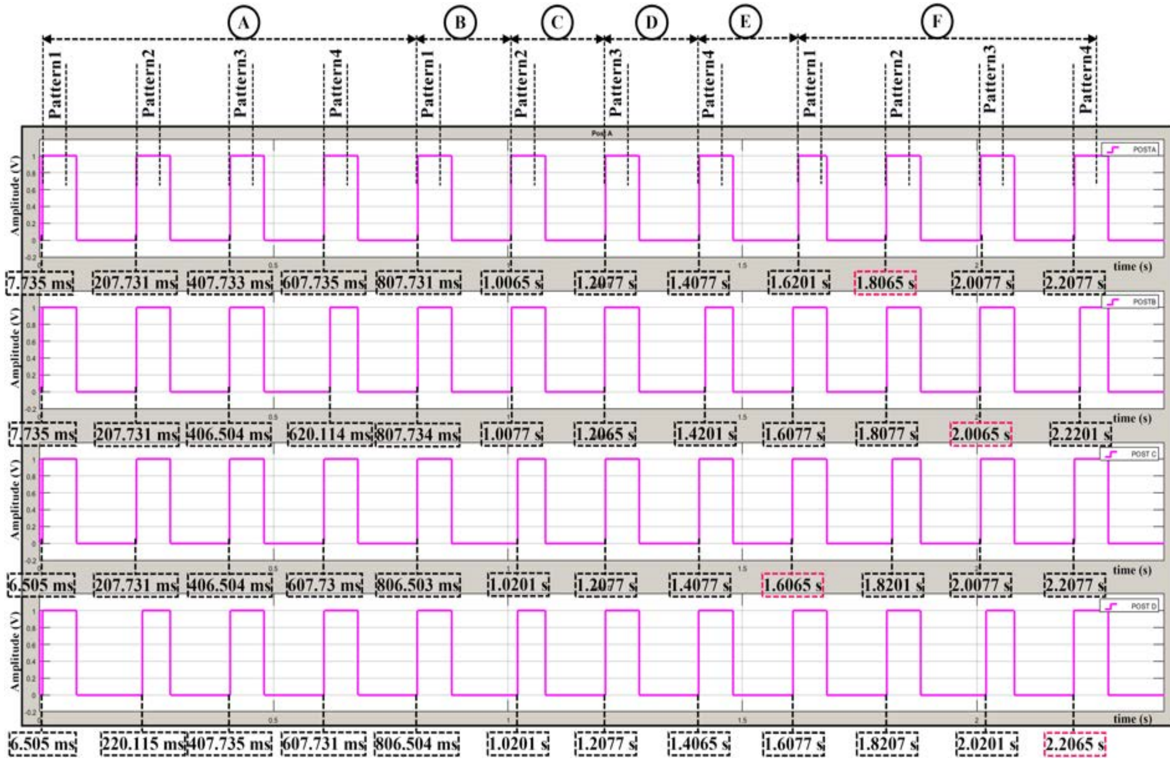


Figure 3-41: Simulated Post-synaptic pulses showing different regions- A, B, C, D, E and F using STDP learning rule.

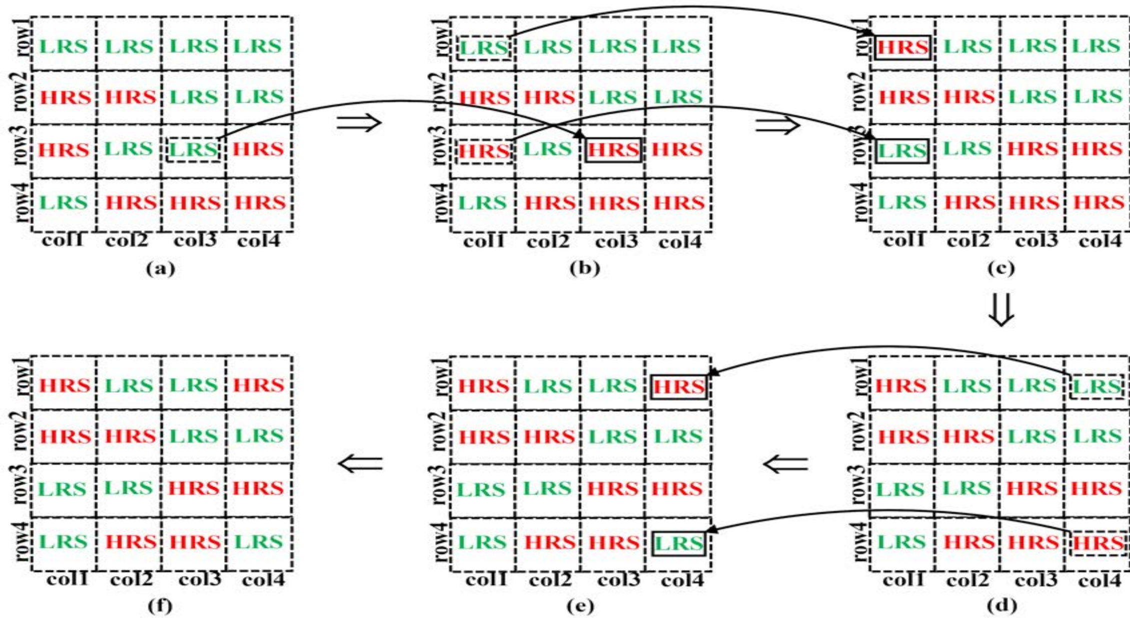


Figure 3-42: Crossbar showing binary weights and its evolution- (a) Initial random weight, (b) 1st weight update, (c) 2nd weight update, (d) 3rd weight update, (e) 4th weight update, (f) Final weights.

The times of occurrence of tpost{a,b,c,d} and tpre{a,b,c,d} pulses are stored in memory for each pixel throughout the simulation. Hence, we will have 4 values of times of occurrence for each post-synaptic pulse when the whole pattern is fed into the system. The internally stored time of occurrence of tpost{a,b,c,d} is compared with the time of occurrence of its corresponding tpre{a,b,c,d} and the appropriate weight W{1,2,...16} is updated in the weight-update processor based on the STDP learning rule and a condition. The condition is- when two or more post-synaptic pulses spike at the same time, priority for weight-update is given to the one with a 'minimum index' number and the one that hasn't faced a weight update earlier. After learning, the final results of pattern recognition using the STDP learning rule differ for different initial weights, whose values are considered random. Fig. 3-39 shows the conceptual diagram where the patterns are applied as read pulses across wordlines of a 4×4 crossbar, whose weights are initially set to random values. Fig. 3-40 shows the waveforms of the pre-synaptic pulses, reset signal, and the number of cycles showing different regions such as 'A', 'B', 'C', 'D', 'E' and 'F'. Region 'A' shows a direct inference of all 4 patterns. Region 'B' shows the inference of 'pattern-1' after 1st weight update whereas, Region 'C' shows the inference of 'pattern-2' after 2nd weight update. Region 'D' shows the inference of 'pattern-3' after 3rd weight update and Region 'E' shows the inference of 'pattern-4' after 4th weight update. Region 'F' shows the inference of all 4 patterns after learning. Fig. 3-41 shows the waveforms of the post-synaptic pulses, where the start-time of the pulses and the regions- 'A', 'B', 'C', 'D', 'E' and 'F' are duly labeled. Fig. 3-42 shows how the weights evolve from random weights to learned values. From 3-41 and 3-42 we can see that after learning 'col3' spikes early for 'pattern-1', 'col1' spikes early for 'pattern-2', 'col2' spikes early for 'pattern-3' and 'col4' spikes early for 'pattern-4'. Fig. 3-43 shows the evolution of weights during simulation using STDP learning with five different initial random weights.

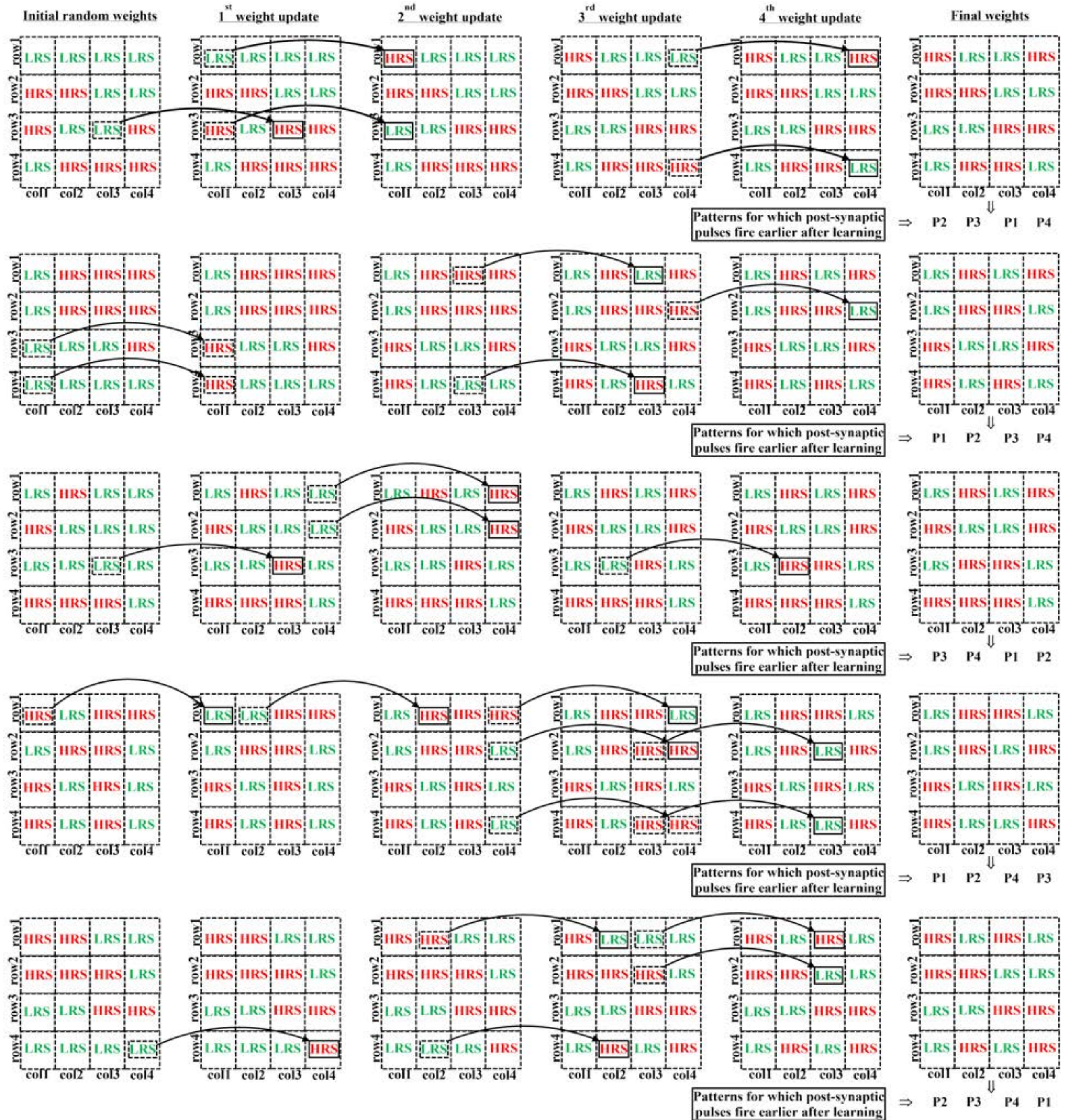


Figure 3-43: Simulation results showing STDP weight updates for five different random initial weights.

3.7 Experimental results of memristive processor facilitated with bulk-based calibration scheme across wordlines

Results of three-stage bulk-based calibration scheme implemented on a 4×4 1T1R memristive crossbar comprise different tests' results conducted on the experimental setup, shown in Fig. 3-31. It includes results from preliminary tests, calibration scheme results, results of characterization of OxRAMs, pattern recognition results using SSSP on the calibrated crossbar, and pattern recognition results using STDP learning rule on the calibrated crossbar. During preliminary tests of validating the working of I-pots and shift-register, care is taken that– biases of the power supply opamps do not disturb the OxRAMs in the crossbar. This is done by setting the voltage difference between the top and bottom terminals of 1T1R devices and its gate-biases to 0 V.

3.7.1 Preliminary test results

The preliminary test-procedure is to validate the functionality of test-PCB and the working of on-chip I-pots and shift-register. The main objective of testing the PCB is to ensure the smooth functioning of all PCB components and to verify the amplitude and pulse-width of the pulses that are applied on the crossbar terminals (by keeping a substitute resistor-bank) consider the different OxRAM operations. The test-PCB is tested by using carbon resistors plugged on resistor plug-&-play board and by applying pulses based on sequential OxRAM operations such as- 'FORM_GLOBAL_READ' or 'ERASE_GLOBAL_READ' or 'WRITE_GLOBAL_READ' or 'READ'. The power supply opamps are biased with the optimal values for keeping active and default values during various OxRAM operations. Waveforms of the applied pulses are observed by probing the crossbar terminals (-wordlines, gates, and bitlines), which are checked for the desired ones, as shown in Fig. 2-17, Fig. 2-18, Fig. 2-19, and Fig. 2-20 of chapter 2. In this way, the test-PCB is tested for the needed functionality.

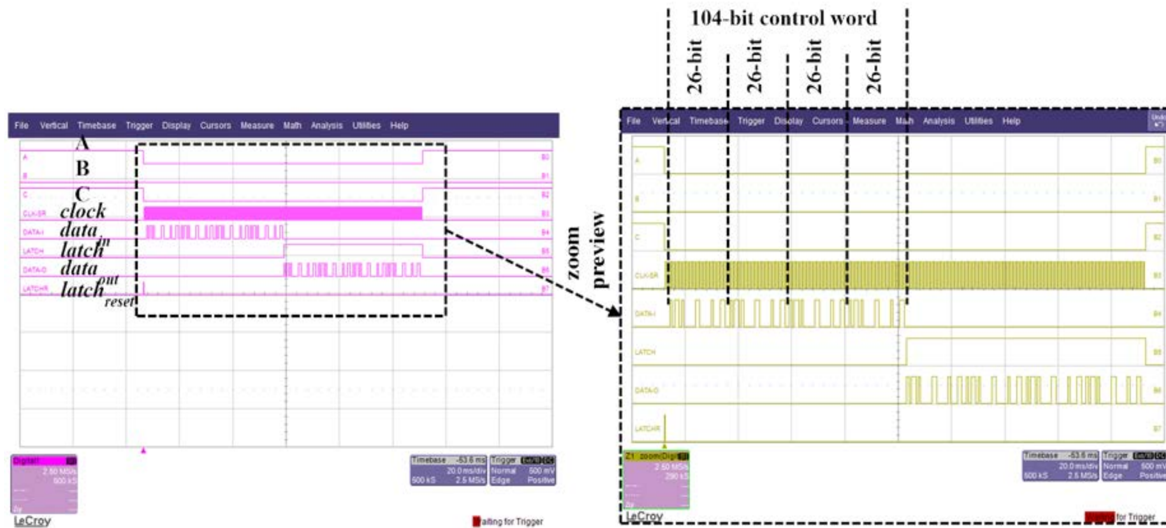


Figure 3-44: Preview of output screen when testing shift-register.

For testing I-pot, the input reference currents for the I-pots, Iref{1, 2, 3, 4} are set to 100 µA using potentiometers, and the digital signals test{1, 2, 3, 4} are kept 'ON'. The control-bits for I-pot (as shown in Table 3.4) are programmed together with a *clock* signal of 2 kHz frequency using SPARTAN®-6 driver board and are loaded to the shift-register. After $data_{in}$ is loaded, the *clock* is stopped and the *latch* signal is turned 'ON' to observe the output currents, Itest{1, 2, 3, 4} across the load resistors. A current of about 40 µA (as required for the nominal bias of two-stage differential opamp) is observed across the load resistor, which validates the working of the I-pots.

For testing shift-register, the 104-bit control word including the control-bits for calibration is programmed using the SPARTAN®-6 driver board and is loaded to the shift-register. A *clock* signal of 2 kHz frequency is used and the calibration control-bits are set such that the bulks- Out_Calib{1, 2, 3, 4} and Calibref{1, 2, 3, 4} are both set at 4.8 V, as calibration is not included in the test. Once the data sequence is loaded, without turning 'OFF' the *clock* signal, $data_{out}$ or *data_out_buf* signal is observed. Fig. 3-44 shows the observed results of the shift-register, where the data sequence of control-word of $data_{out}$ of the last D flip-flop is similar to the input data sequence of $data_{in}$. This validates the working of shift-register.

3.7.2 Three-stage bulk-based calibration scheme results of full input control-word

After testing shift-register and I-pot, calibration bits are carefully set using the button-board such that a coarse (stage 1) calibration is first carried out. Before calibration, the OxRAMs are carefully formed in a three-step process. The first step is to load shift-register with 104-bit $data_{in}$, *clock* (of 2 kHz frequency) signal, *latch* and other control signals (like digital inputs of pre-synaptic drivers, 3-bit control signal F(A, B, C), etc.) by programming through SPARTAN®-6 driver board in such a way that– when the 'FORM_GLOBAL_READ' button is pressed, the resistance of the formed OxRAM is obtained by probing the terminals of opampa (read opamp as shown in Fig. 3-29). The second step is to use the button-board to configure both the

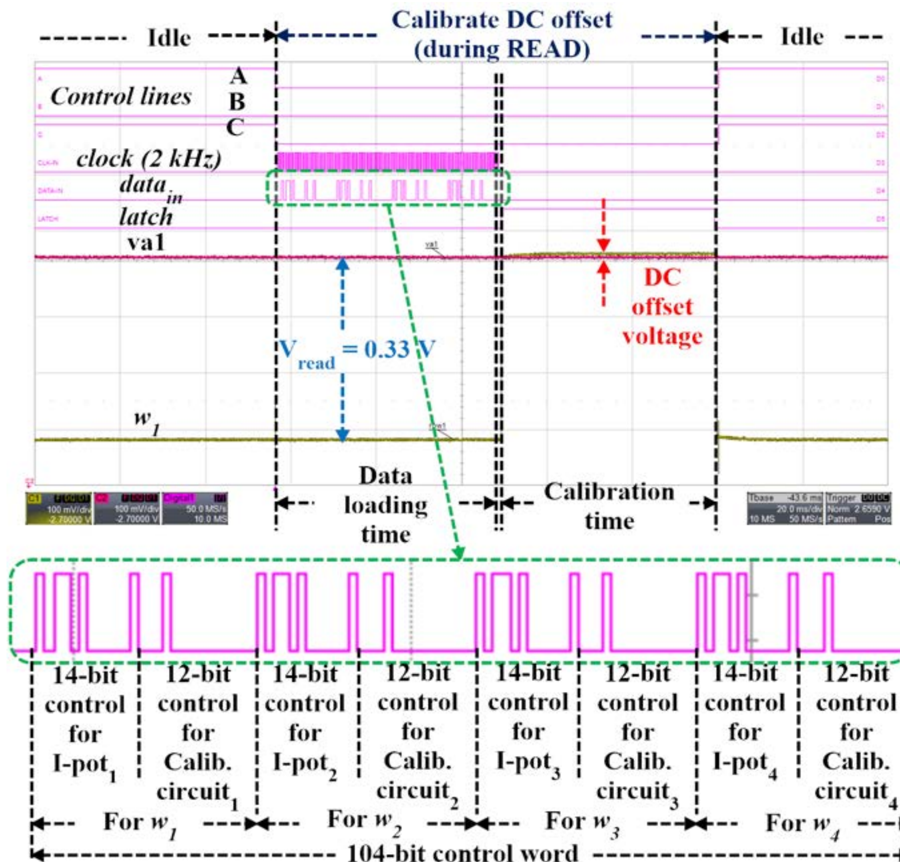


Figure 3-45: Preview of output screen when calibrating DC offset across wordline, w_1.

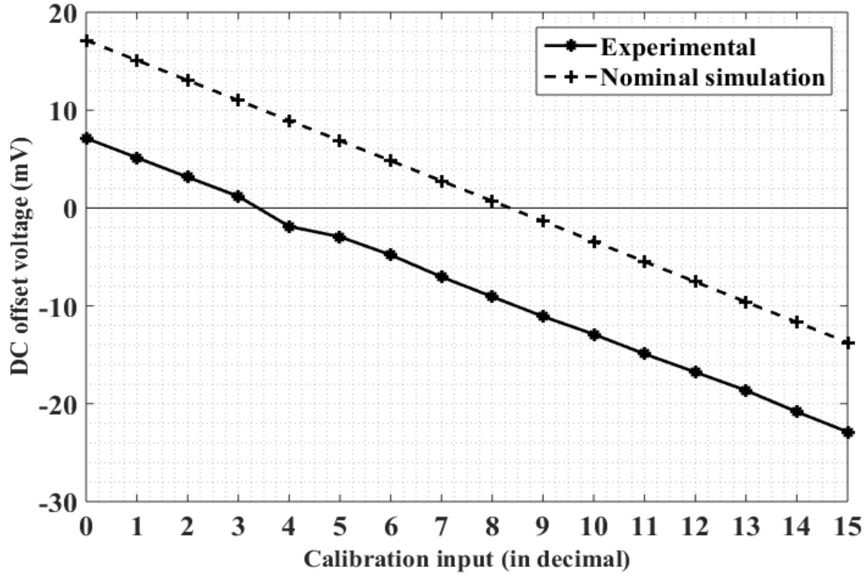


Figure 3-46: Comparison of experimental and simulation results during stage 1 calibration of DC offset voltage across wordline, w_1.

input control-bits for calibration and the input bits for targeting the 1T1R synapse in the crossbar. The final step is to obtain post-forming resistance after pressing the 'FORM_GLOBAL_READ' button.

Once all OxRAMs are formed, calibration of DC offset is done in a 'Read' operation by a three-step process. The first step is to load shift-register with 104-bit $data_{in}$, *clock* (of 2 kHz frequency) signal, *latch* and other control signals by programming through SPARTAN®-6 driver board in such a way that- when the 'Calibrate DC offset' button is pressed on the button-board, the *clock* runs as long as the 104-bit data sequence is loaded, which is followed by stopping the *clock* and turning 'ON' the *latch*. The second step is to set the calibration input control-bits on the button-board. The third step is to observed DC offset voltage by pressing the 'Calibrate DC offset' button.

Fig. 3-45 shows a preview of the output screen when wordline w_1 is calibrated during a 'Read' operation of V_{read}= 0.33 V. Here $va1$ is the input voltage of opamp1 (as shown in Fig. 3-3) and opamp1's output leads to wordline, w_1 of the crossbar. The 'Data loading time' comprises the time for loading $data_{in}$ and *clock*. Once the data-sequence is loaded, the *latch* is turned 'ON' and the DC offset voltage is calibrated.

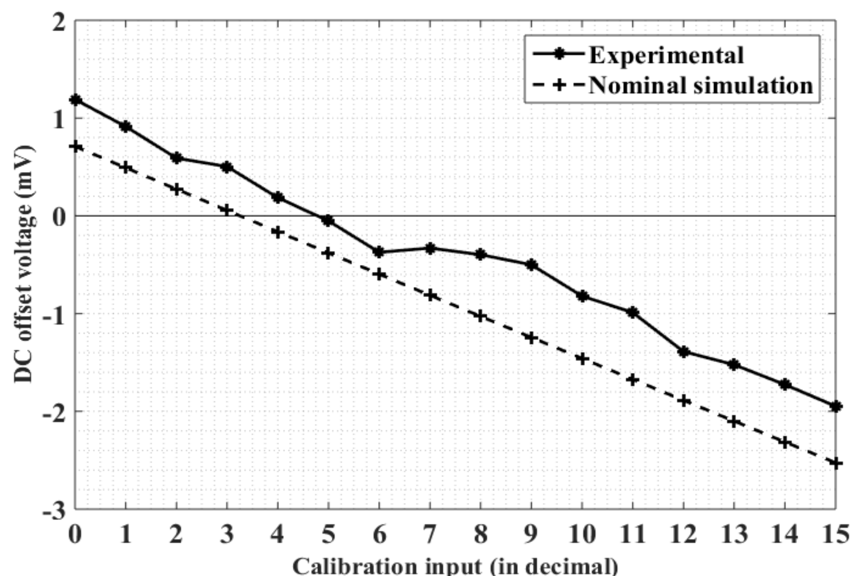


Figure 3-47: Comparison of experimental and simulation results during stage 2 calibration of DC offset voltage across wordline, w_1 targeting the zero-crossing region.

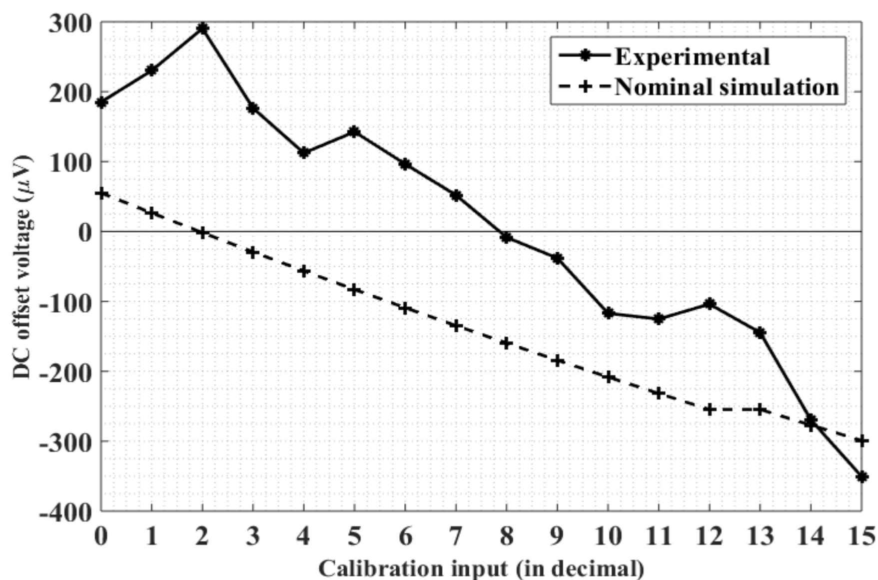


Figure 3-48: Comparison of experimental and simulation results during stage 3 calibration of DC offset voltage across wordline, w_1 targeting the zero-crossing region.

Here, DC offset voltage is the difference between the input and output of the opamp connected in buffer configuration across the wordline, w_1. The results are taken by averaging 100 million samples to filter out noise, whose standard-deviation is about 200 µV. The power dissipation during inference 'READ' operation is about

0.8 µW for a 4×4 crossbar when using a 50 mV read pulse whose DC offset voltage is finely calibrated. Fig. 3-46 shows the comparison of experimental and simulation results when wordline, w_1 of the crossbar is calibrated for DC offset voltage during coarse (stage 1) calibration for V_{Read}= 0.33 V. The zero-crossing region in Fig. 3-46 is targeted and DC offset voltages are calibrated during fine (stage 2) and finer (stage 3) calibration, whose results are shown in Fig. 3-47 and Fig. 3-48. Experimental results of the three-stage calibration scheme match simulation results.

3.7.3 Characterization results of OxRAMs in 4×4 1T1R crossbar with on-chip DC offset calibration across wordlines

In this section, several characterization results of the OxRAMs of 4×4 1T1R crossbar where calibration of DC offset is implemented are discussed. Unlike the characterization of individual crossbars (as discussed in Section 2.3.5), characterizing OxRAMs in the crossbar that is equipped with a calibration scheme is not straight-forward.

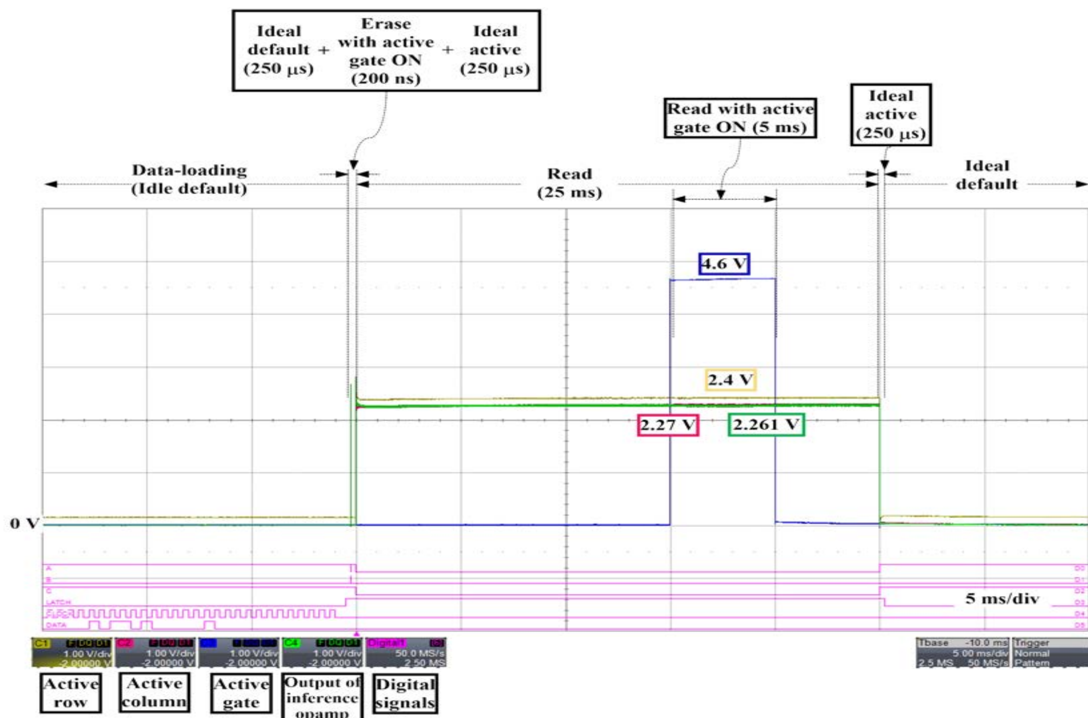


Figure 3-49: Active row, column and gate biases applied in the form of pulses- showing a read operation after an erase operation.

It is due to the presence of pre-synaptic drivers across the wordlines, whose control-bit for calibration and i-pots are loaded from the shift-register. Therefore, there is always a 'data-loading' part, where the *clock* is 'ON' and $data_{in}$ is loaded. Following this, the *clock* is stopped, the *latch* is turned 'ON'and the sequential OxRAM operation such as- 'FORM_GLOBAL_READ' or 'ERASE_GLOBAL_READ' or 'WRITE_GLOBAL_READ' or 'READ'- is carried out. Utmost care is taken to the OxRAMs such that none of them are disturbed during the 'data-loading' part. Also for the 'ERASE', 'WRITE', and 'READ' part of the pulses the active gate is switched 'ON' after a short delay to avoid transients, which could be harmful to the OxRAMs. Like other tests, here also the active and default wordlines, bitlines, and gates are chosen using jumper pins on the button-board to pick a 1T1R device.

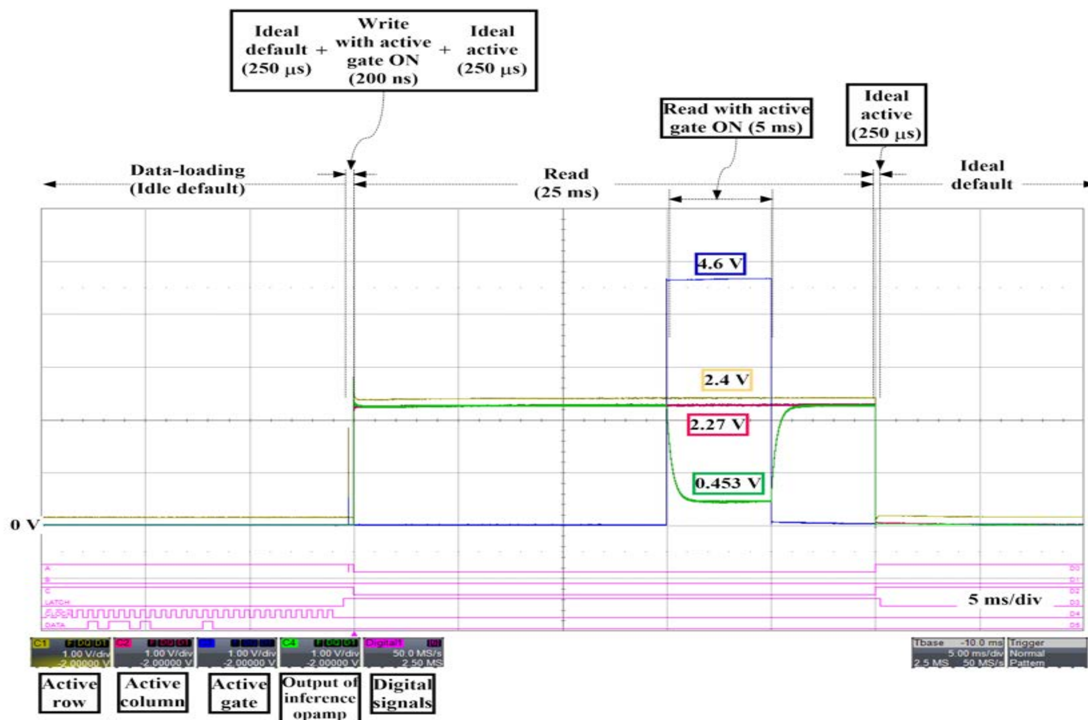


Figure 3-50: Active row, column and gate biases applied in the form of pulses- showing a read operation after a write operation.

Once a device is targeted the desired sequential operation is carried out on it. All OxRAMs are carefully formed and are switched between LRS and HRS. Fig. 3-49 shows the biases of the row (or wordline), column (or bitline), gate bias, output of inference opamp, and digital signals when 'ERASE_GLOBAL_READ' operation

is carried out. Fig. 3-50 shows the biases of the row, column, gate bias, output of inference opamp, and digital signals when 'WRITE_GLOBAL_READ' operation is carried out. Here, during 'Ideal default' and 'Active default' both top and bottom terminals of the targeted 1T1R device are biased at 0 V and the gate is turned 'OFF'.

Initially, the 'data-loading' part is programmed for 52.5 ms. This follows 'Ideal default', 'Erase', 'Ideal active', 'read', and 'Ideal active'. During 'read', first, the row and column biases are switches and after 15 ms time, the gate for is turned 'ON'. The values of the control signals- A, B, and C for different OxRAM operations are digitally set based on Table 3.5. The LRS and HRS resistance values of the OxRAM are calculated from Fig. 3-49 and Fig. 3-50, where we have the values of the 'Active row', 'Active column', 'Active gate', and 'Output voltage of the inference opamp'. A feedback resistor of resistance value 26.71 kΩ is used at the inference opamp. The 'Read' voltage is 2.4 - 2.27 = 0.13 V. The output of the inference opamp goes to 2.261 V during a 'Read' after 'Erase' when the active gate is turned 'ON'. Similarly, the

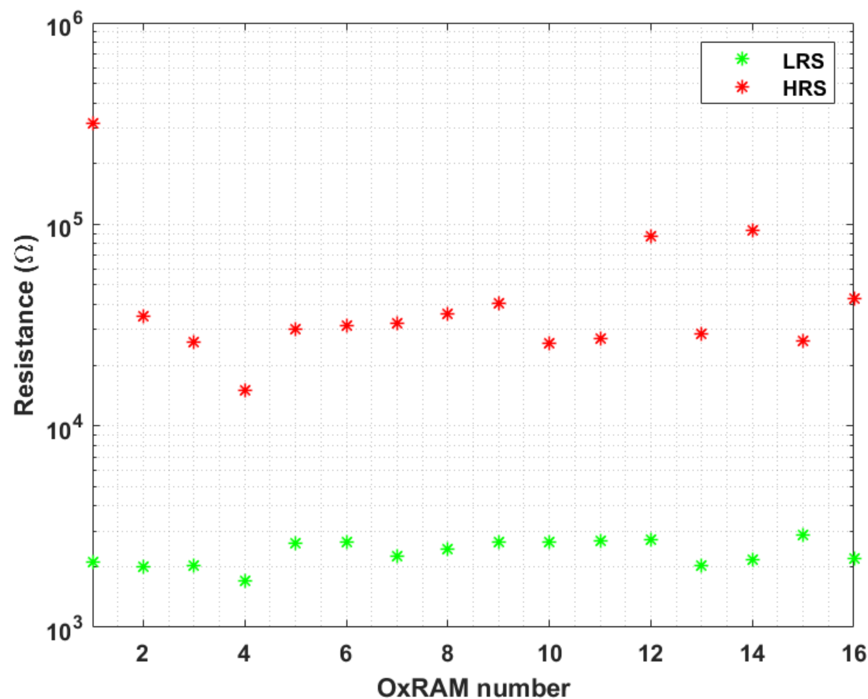


Figure 3-51: LRS and HRS values of OxRAMs of all 16 1T1R devices in the 4×4 crossbar where calibration of DC offset voltage is carried out.

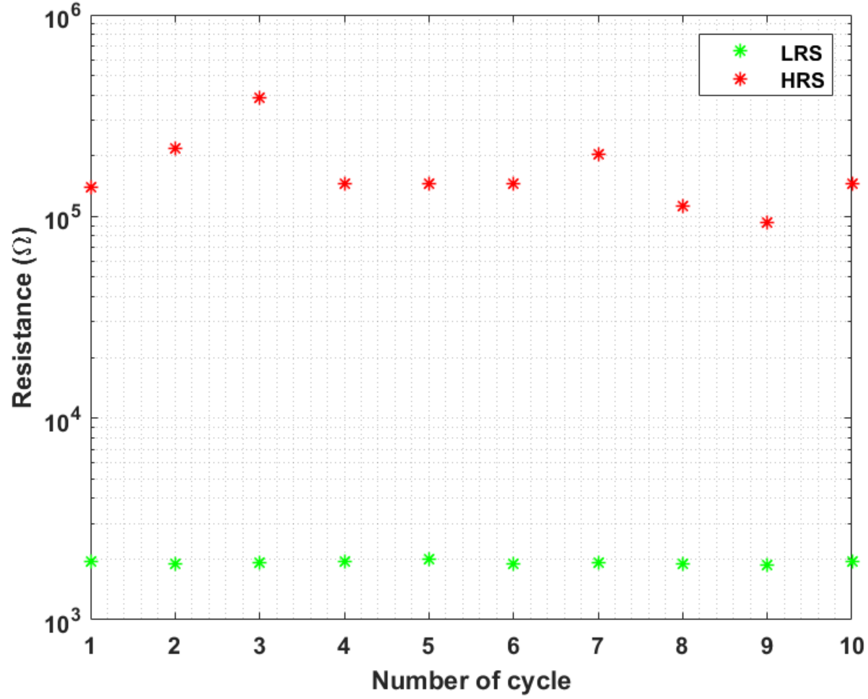


Figure 3-52: LRS and HRS values for 10 switching cycles of an OxRAM of the 4×4 1T1R crossbar where calibration of DC offset voltage is carried out.

output of the inference opamp goes to 0.453 V during a 'Read' after 'Write' when the active gate is turned 'ON'. The LRS and HRS values are calculated as follows-

$$\mathrm{LRS} = \left(\frac{26.71\,\mathrm{k\Omega}}{2.27\,\mathrm{V} - 0.453\,\mathrm{V}}\right) \times 0.13\,\mathrm{V} = 1.91\,\mathrm{k\Omega} \tag{3.13}$$

$$\mathrm{HRS} = \left(\frac{26.71\,\mathrm{k\Omega}}{2.27\,\mathrm{V} - 2.261\,\mathrm{V}}\right) \times 0.13\,\mathrm{V} = 385.81\,\mathrm{k\Omega} \tag{3.14}$$

Fig. 3-51 shows the LRS and HRS values of OxRAMs of all 1T1R devices in the 4×4 crossbar where calibration of DC offset voltage is implemented. Fig. 3-52 shows the LRS and HRS values for 10 switching cycles of an OxRAM of the 4×4 1T1R crossbar where calibration of DC offset voltage is implemented. Fig. 3-53 shows the LRS and HRS values for 400 switching cycles of an OxRAM of the 4×4 1T1R crossbar where calibration of DC offset voltage is implemented. This plot shows high variability in the distribution of HRS.

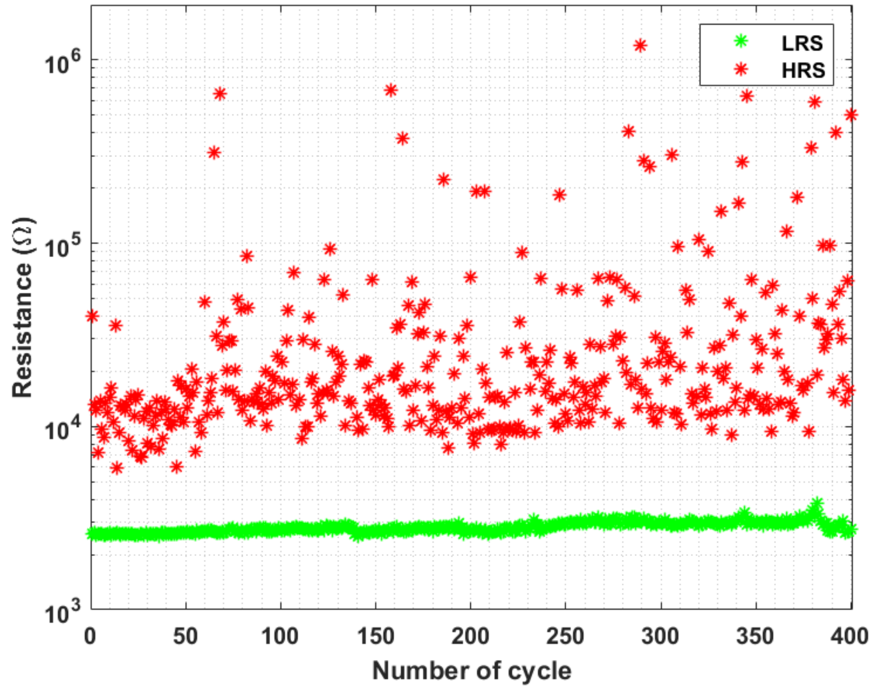


Figure 3-53: Switching of resistance of OxRAM of a 1T1R device between HRS and LRS for 400 cycles.

3.7.4 Template-matching results implemented on a calibrated 4×4 1T1R crossbar

This section discusses various template-matching results done using a 4×4 1T1R crossbar where calibration scheme is implemented across wordlines. The patterns for which template-matching is done is shown in Fig. 3-33. Fig. 3-54 shows the conceptual block diagram showing how patterns are fed as 'read' pulses across the rows of the calibrated crossbar, whose synapses are switched to learned weights. Initially, the OxRAMs in 4×4 1T1R crossbar are formed one-by-one and are set to learned weights as shown in the crossbar of Fig. 3-54. Following this, an inference is carried out across bitlines by carefully reading the OxRAMs using tiny read pulses. For a specific pattern when the contributing weights are strengthened (or when resistance is low), the pulse-width of the inverted version of the output of the comparator becomes minimum. Hence, achieving minimum pulse-width during inference using a pattern is an indication of achieving learned weights for that pattern.

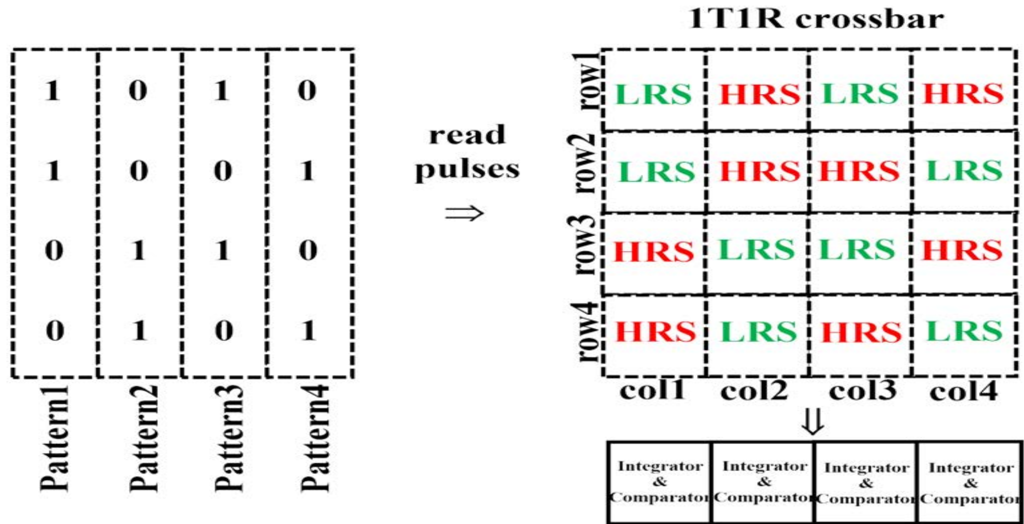


Figure 3-54: Conceptual block diagram showing patterns fed as read pulses across the wordlines (or rows) of the calibrated crossbar, whose synaptic weights are switched to learned values.

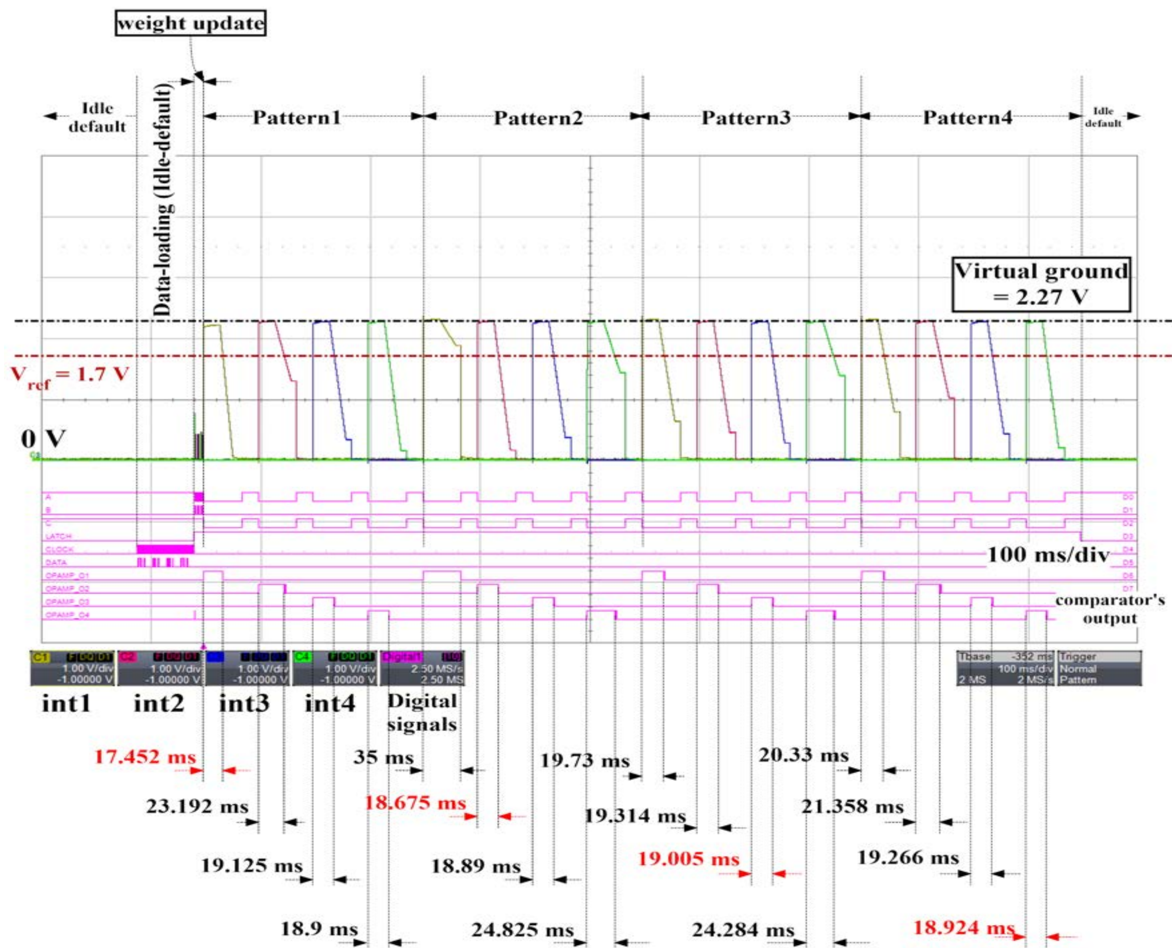


Figure 3-55: Output of integrators and digital signals for template matching for read voltage of 0.13 V.

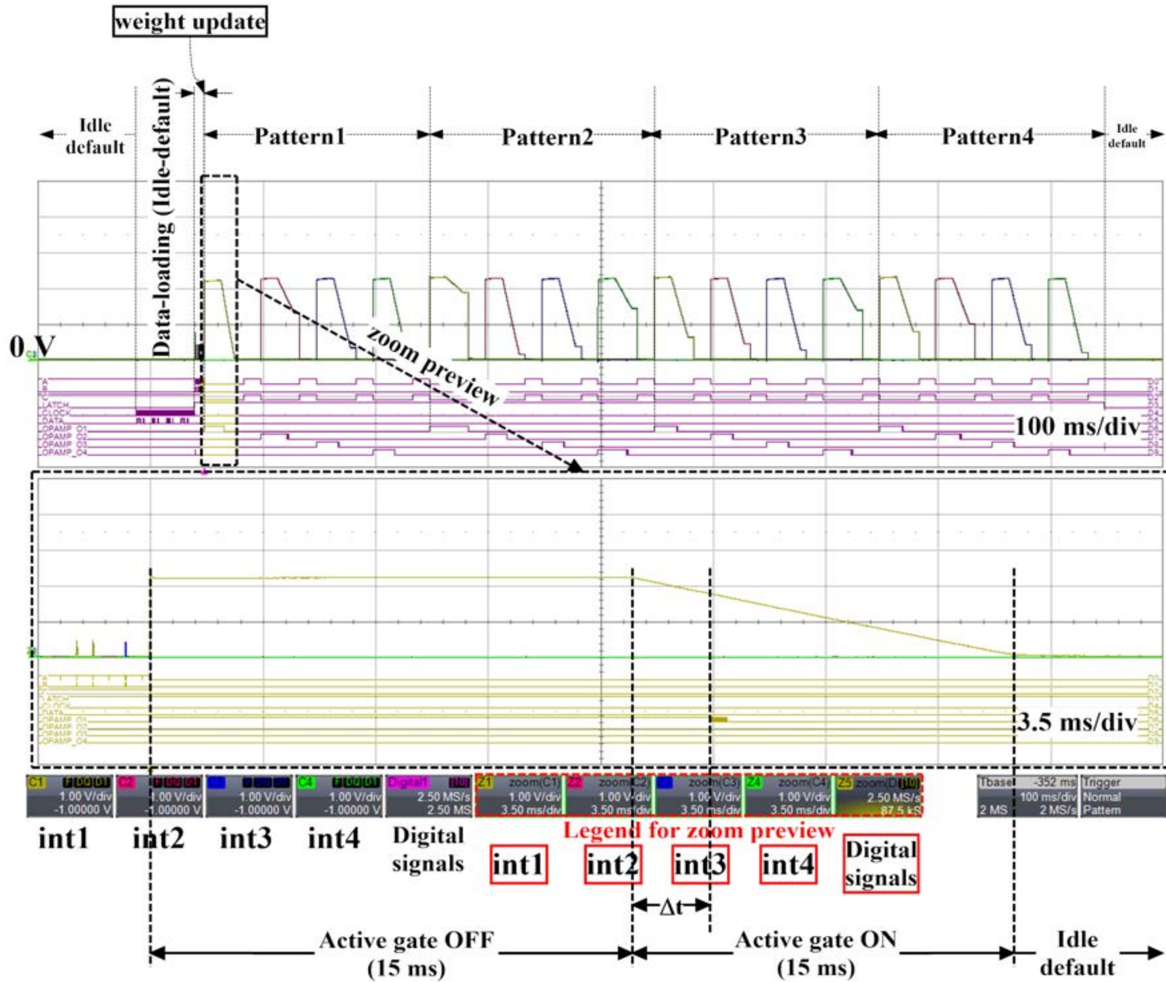


Figure 3-56: Output of integrators and digital signals for template-matching using read voltage of 0.13 V- with a zoom preview showing the integrated voltage.

In the beginning, the inference is done with a read pulse of amplitude 0.13 V. There are integrating opamps across the bitlines of the crossbar whose feedback capacitors have a capacitance value of 1 µF. Fig. 3-55 shows the output of integrators and digital signals for the template-matching experiment performed on the calibrated crossbar by using read pulses of amplitude 0.13 V. It comprises initially a 'data-loading' time, which is followed by 'weight-update' and 'inference'. The digital signals- 'OPAMP_O1', 'OPAMP_O2', 'OPAMP_O3', and 'OPAMP_O4' are the output of the comparator across 4 bitlines. The other digital signals include- the control signals- A, B, and C, the *latch*, *clock*, and $data_{in}$ (or DATA). The weight-update is done to keep the synapses with learned weights, as shown in the crossbar of Fig. 3-54. The inference

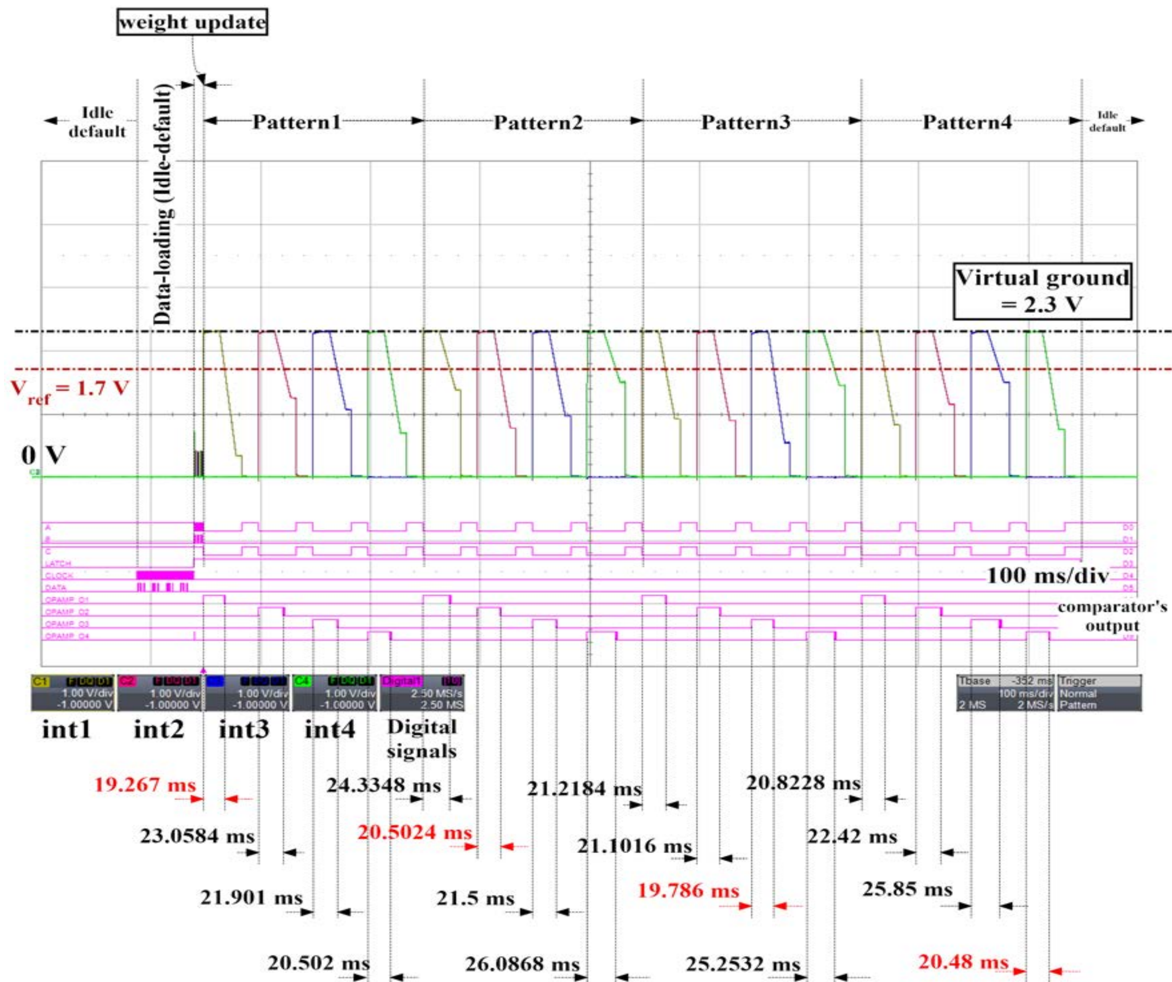


Figure 3-57: Output of integrators and digital signals for template-matching using read voltage of 100 mV.

is done for each pattern and during each pattern, each bitline is separately read, as the test-bench facilitates picking only one active bitline at an instant and keeping the rest as default. The task 'inference' here denotes integrating the accumulated current and comparing the integrated output voltage with a reference voltage (here, it is 1.7 V). The integrating time set during each inference is 15 ms. The results clearly show that- when learned weights are set for the synapses, they result in the desired pulse-widths after inference. In other words, inference across bitline1 during pattern1 results in the least pulse-width, the inference across bitline2 during pattern2 results in the least pulse-width, and so on. The 'Pulsewidth' here is a direct measurement of Δt, which is the difference between the time of occurrence of pre-synaptic pulse

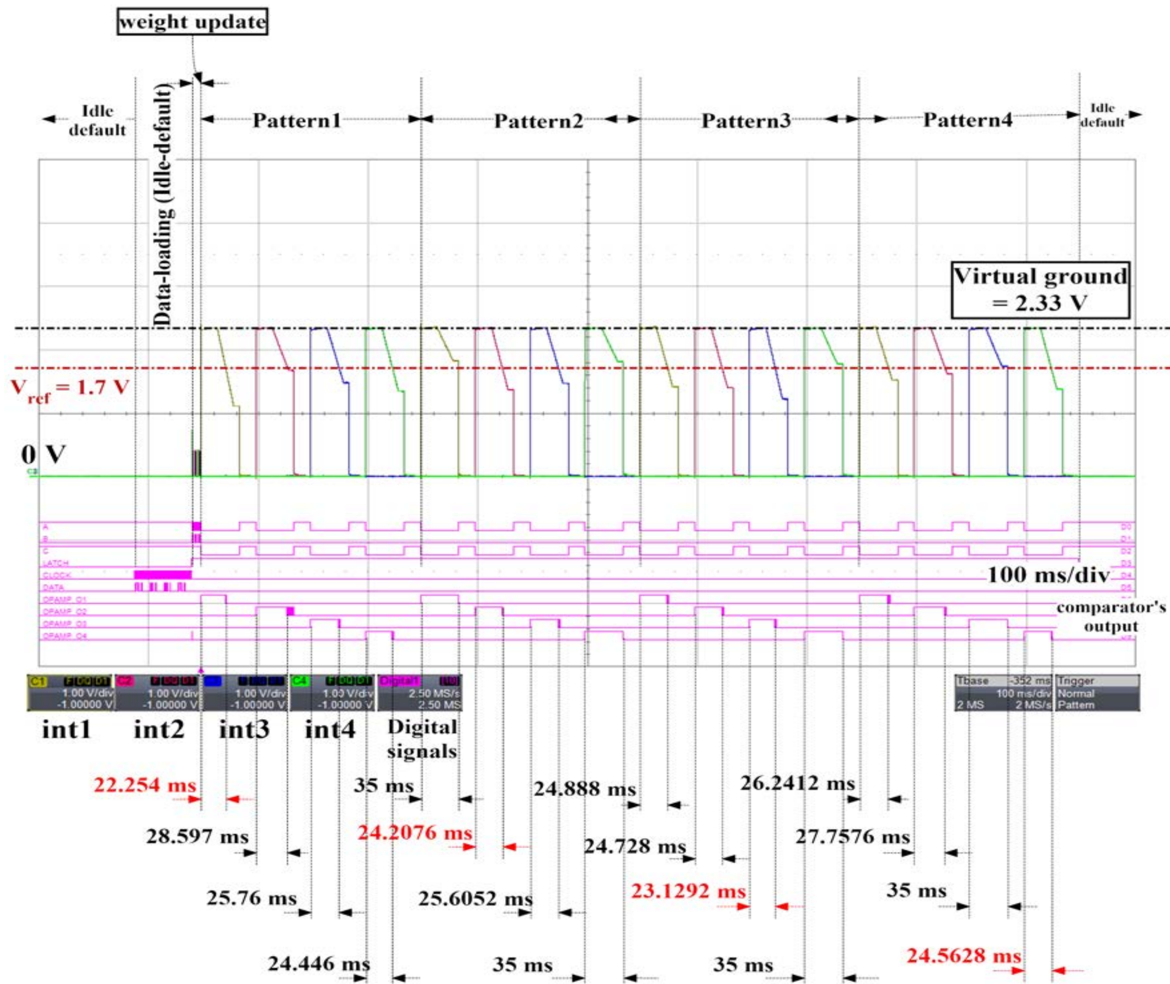


Figure 3-58: Output of integrators and digital signals for template-matching using read voltage of 70 mV.

and post-synaptic pulse. During inference every labeled pulse-width results in its Δt by subtracting 15 ms, as the active gate is switched 'ON' after 15 ms. Fig. 3-56 shows a zoom preview of the output of integrators and digital signals for template matching using read voltage of 0.13 V. Here Δt is the difference between the time when the active gate is switched 'ON' during 'read' and the time when the output of its corresponding comparator turns 'OFF'.

To test the robustness of the system for tiny read pulses, whose DC offset can be finely calibrated the amplitude is reduced in steps. For this, decreasing the higher-level voltage (2.4 V) is not recommended, as this bias enters the crossbar's wordlines via the pre-synaptic driver and we need this bias for 'write' and 'erase' operations.

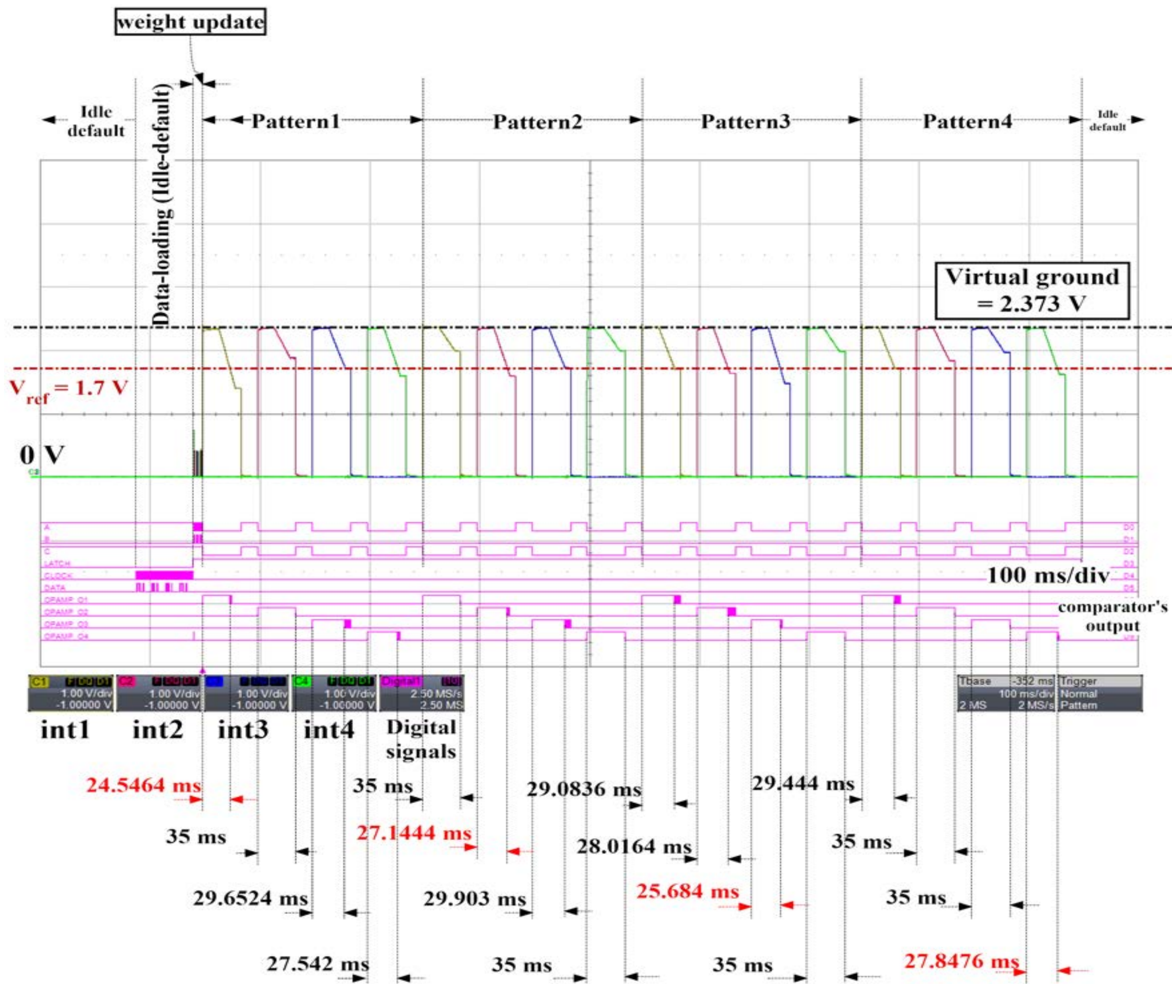


Figure 3-59: Output of integrators and digital signals for template-matching using read voltage of 30 mV.

Instead, the virtual ground (2.27 V) is increased in steps such that virtual ground is set to 2.3 V, 2.33 V, and 2.373 V, whose read voltages are 100 mV, 70 mV, and 30 mV respectively. On top of this, there is low-frequency noise spread in the range between 15 and 20 mV in the space between the top and bottom signals of the read pulse. Fig. 3-60 shows the low-frequency noise when a 2.4 V signal is directly probed to view on the oscilloscope. Here, 3-bit 'ERes' noise filter option of the oscilloscope is used, which is its maximum noise filtering capability. Here, we can see the noise spread of 17 mV on top of the signal. Hence, the low-frequency noise spread has to be subtracted from the above-measured read amplitudes to get precise read amplitudes. These tiny read amplitudes after excluding the noise floors by which the systems work

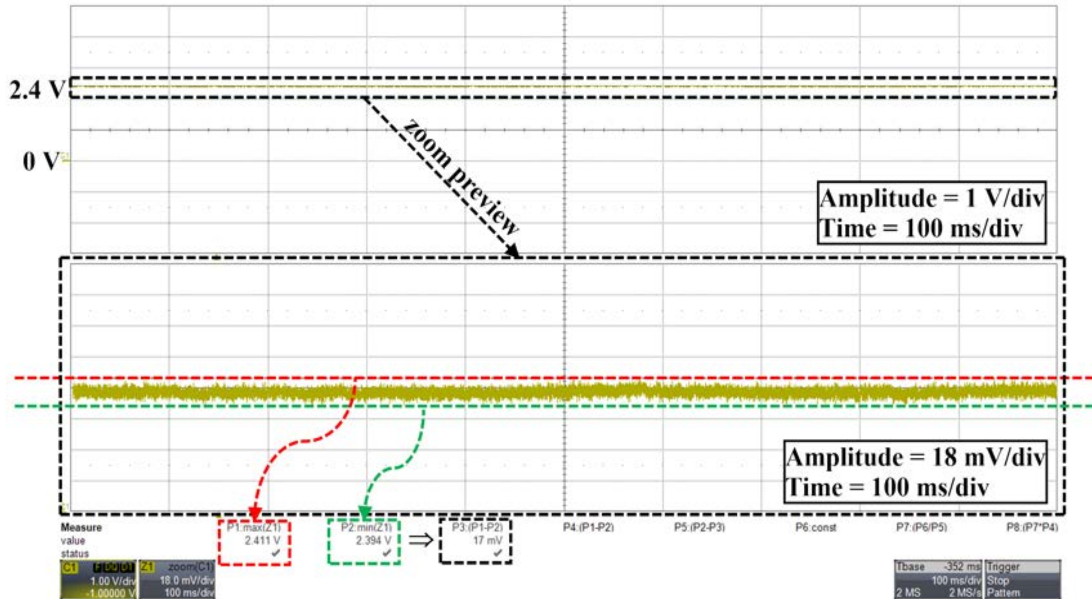


Figure 3-60: 2.4 V signal from power supply directly observed on oscilloscope with 3 bit 'ERes' noise filter option.

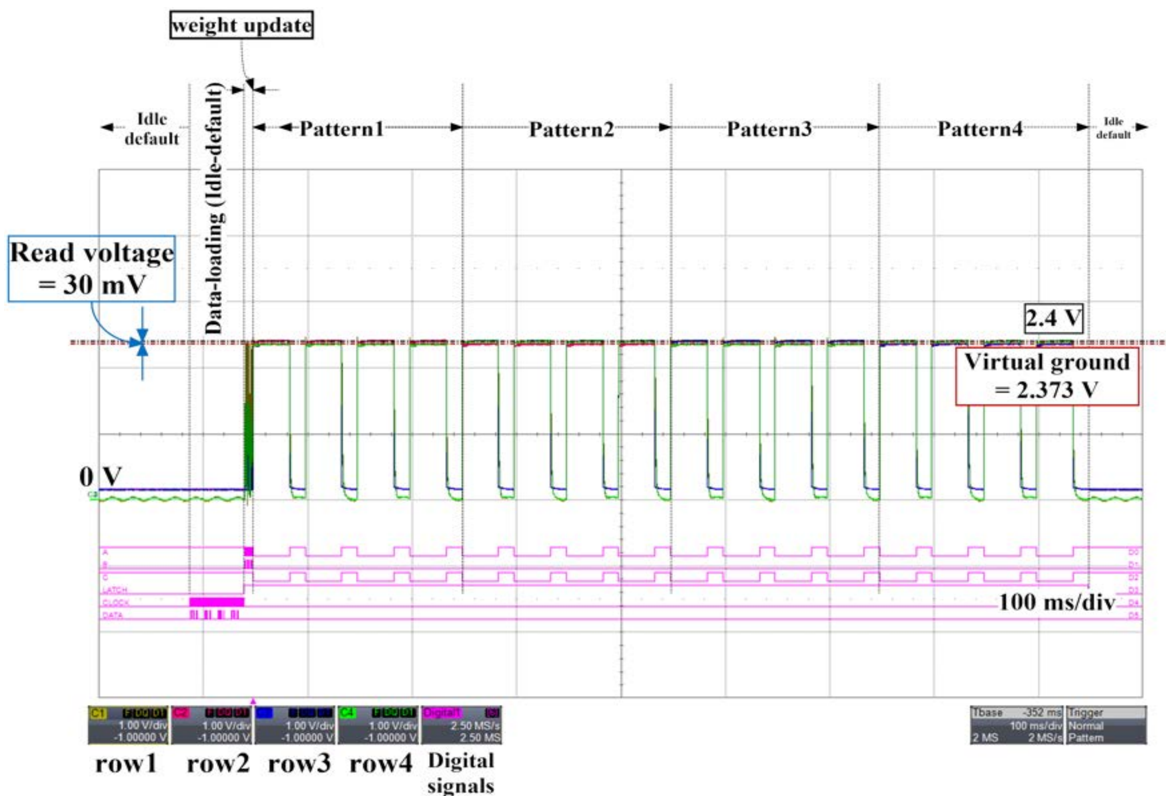


Figure 3-61: Wordlines (or Rows) and digital signals for template-matching using read voltage of 30 mV.

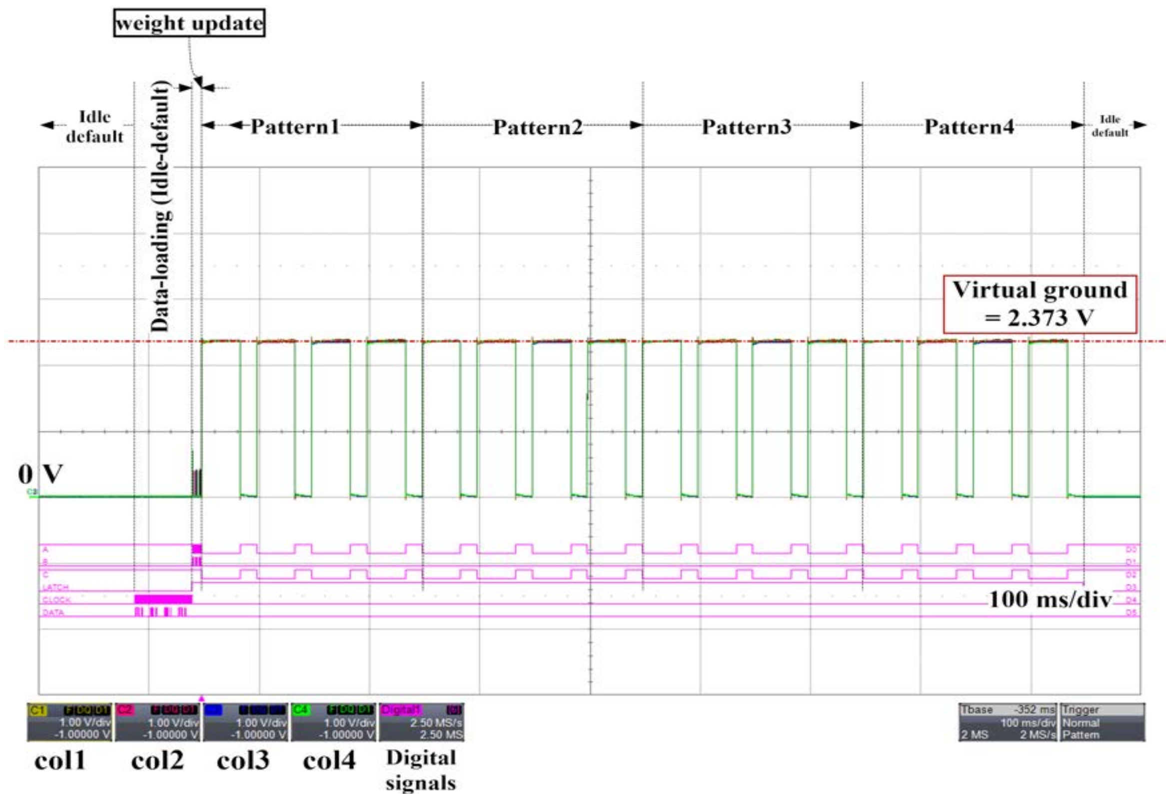


Figure 3-62: Bitlines (or columns) and digital signals for template-matching using read voltage of 30 mV.

as desired for template matching are 84 mV, 54 mV, and 14 mV respectively.

Fig. 3-57 shows the output of integrators and digital signals for the template-matching experiment performed on the calibrated crossbar by using 'read' pulses of amplitude 100 mV. Fig. 3-58 shows the output of integrators and digital signals for the template-matching experiment performed on the calibrated crossbar by using 'read' pulses of amplitude 70 mV. And Fig. 3-59 shows the output of integrators and digital signals for the template-matching experiment performed on the calibrated crossbar by using 'read' pulses of amplitude 30 mV. Fig. 3-59 also shows how the system is capable to result in the desired output during inference even when read with tiny read pulses that have amplitude as low as 30 mV. Fig. 3-61 shows the wordlines (or rows) and digital signals for the template-matching experiment performed on the calibrated crossbar by using 'read pulses' of amplitude 30 mV. The read voltage of 30 mV is carefully set such that none of the signal's data-point after noise filtration touches

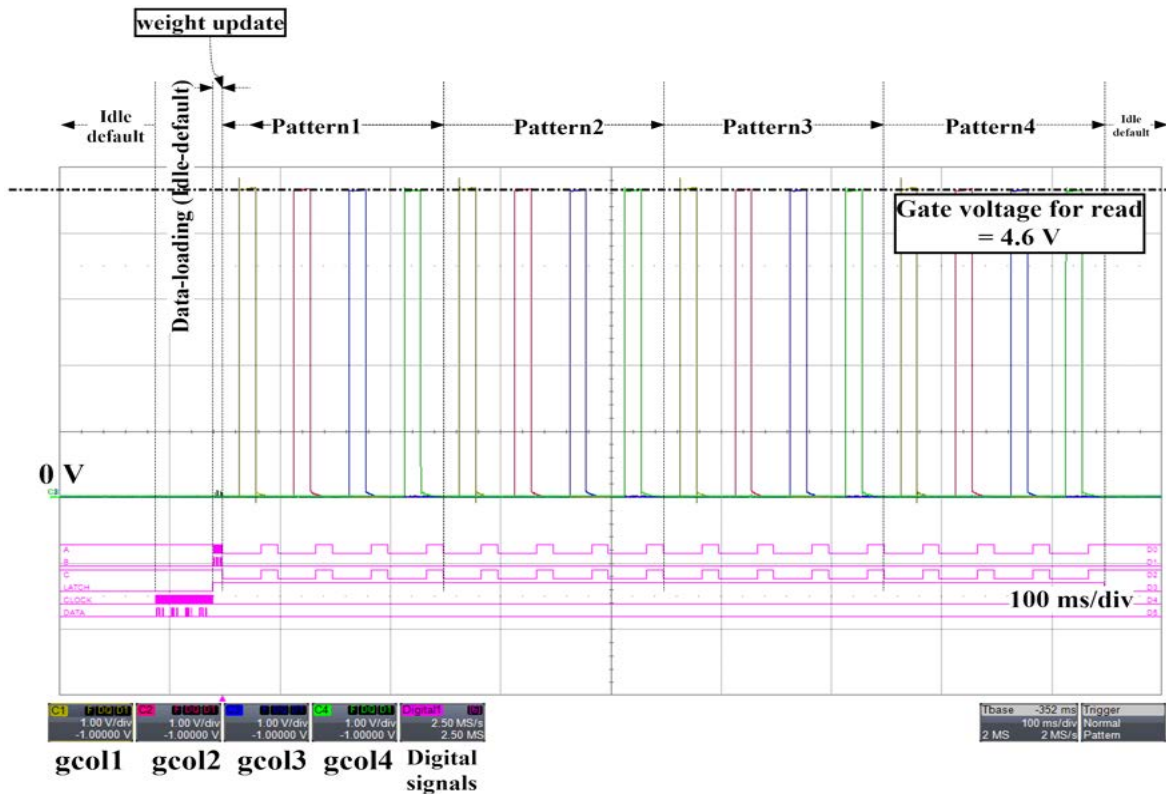


Figure 3-63: Gate biases and digital signals for template-matching using read voltage of 30 mV.

or exceeds the high level of the read pulse (2.4 V). Fig. 3-62 shows the bitlines (or columns) and digital signals for the template-matching experiment performed on the calibrated crossbar by using 'read' pulses of amplitude 30 mV. Fig. 3-63 shows the gate signals (or gcol{1,2,3,4}) and digital signals for the template-matching experiment performed on the calibrated crossbar by using 'read' pulses of amplitude 30 mV. A bias of 4.6 V is applied to the gate of the 1T1R devices to read the current of the OxRAMs and it is applied for 15 ms after switching the wordlines and bitlines to avoid transients from harming the OxRAMs.

3.7.5 Pattern recognition results using SSSP on calibrated crossbar

In this section, the experimental results of Supervised Single Shot Programming (SSSP) for pattern1 on calibrated crossbar are discussed. With reference to the model-

based system-level simulation carried out for pattern recognition in Section 3.6.1, each pattern is targeted for a particular post-synaptic pulse to spike earlier. The synapses that contribute weight for a particular pattern to spike earlier are programmed to face a weight update, while the rest of the synapses that don't contribute face an opposite weight update. Similarly, other post-synaptic pulses are programmed to spike earlier for other specific patterns by their corresponding weight updates. Here, we target 'col1' to spike earlier for 'pattern-1', 'col2' to spike earlier for 'pattern-2', 'col3' to spike earlier for 'pattern-3', and 'col4' to spike earlier for 'pattern-4'. The flowchart

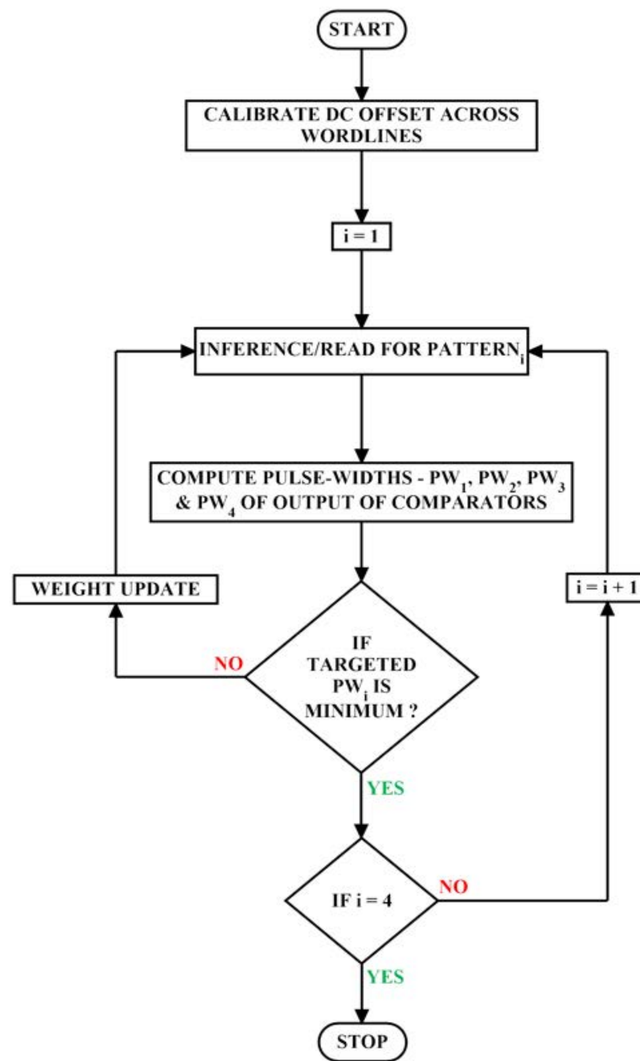


Figure 3-64: Flowchart for implementation of SSSP on calibrated crossbar by targeting- 'col1' to spike earlier for 'pattern-1', 'col2' to spike earlier for 'pattern-2', 'col3' to spike earlier for 'pattern-3' and 'col4' to spike earlier for 'pattern-4'.

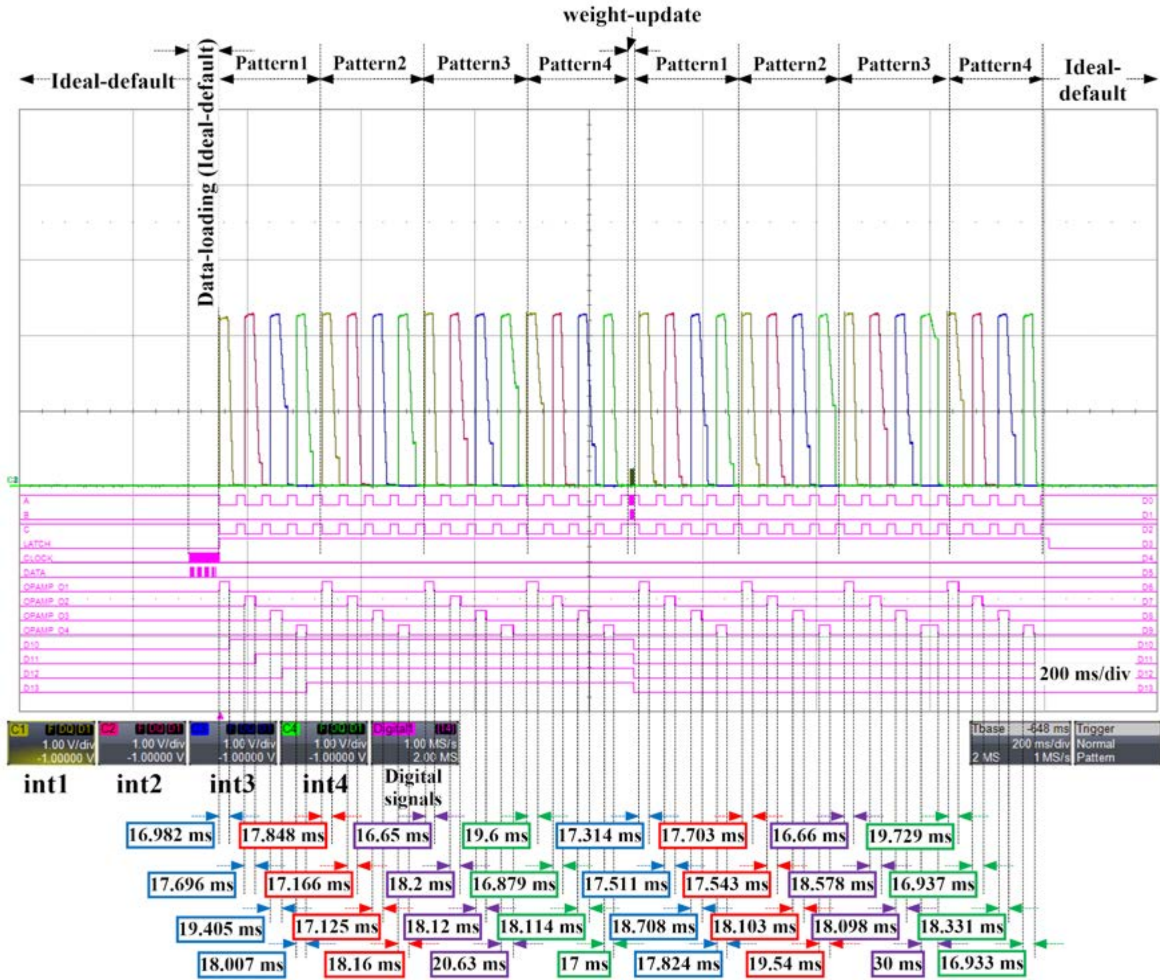


Figure 3-65: Inference results by programming with the single shot variant of STDP targeting pattern1.

for implementing the SSSP on calibrated crossbar is shown in fig. 3-64.

The FPGA driver board is programmed such that the user can initially calibrate DC offset and then choose the pattern for programming. Dedicated push-buttons are used to do calibration and weight update. When a particular pattern is chosen and the weight update is carried out, the inference outputs of both before and after weight update is visualized to see if the targeted post-synaptic pulse is programmed to spike earlier for the specific pattern or not. If not the push-button is pressed again to visualize the results. Fig. 3-65 shows the inference results by programming a single shot variant of STDP targeting pattern1. Here the push-button is pressed for the 3rd time and observed that- bitline1 (or column1) spiked earlier for pattern1, which

results in a shorter pulse-width of the comparator's output. It also appears that the bitline1 spiked earlier for pattern1 when the push-button is pressed for the 1st or 2nd time. Similarly, other patterns can be chosen for targeted post-synaptic pulses to spike earlier by programming the devices.

3.7.6 Pattern recognition results using STDP learning rule on calibrated crossbar

In this section, experimental results of pattern recognition using 'STDP learning rule' implemented on calibrated crossbar are discussed. With reference to the model-based system-level simulation carried out for pattern recognition in Section 3.6.2, weight updates are done based on the time of occurrence of pre-synaptic and post-synaptic pulses. When the post-synaptic pulse spike after the pre-synaptic pulse, the weight of the corresponding synapse is strengthened by decreasing the resistance. Alternately, when the pre-synaptic pulse spike after the post-synaptic pulse or when there is no pre-synaptic pulse, the weight of the corresponding synapse is weakened by increasing the resistance. A condition (to give priority to minimum index and to those which did not face weight update earlier) is used when two or more post-synaptic pulses spike at the same time. Initially, the weights are kept random or even unknown. Fig. 3-66 shows the flowchart for implementing the STDP learning rule on the calibrated crossbar for pattern recognition.

The FPGA driver board is programmed such that the user can initially calibrate DC offset and then go for the pattern recognition task. The objective to keep tiny read inference pulses, calibrate DC offset for it, and use the same tiny read pulses for pattern recognition using the STDP learning rule- is all done. Dedicated push-buttons are used to do tasks such as calibration and weight update with inferences. For a specific pattern when the contributing weights are strengthened (or when resistance is low), the pulse-width of the inverted version of the output of the comparator becomes minimum. Hence, achieving minimum pulse-width during inference using a pattern is an indication of achieving learned weights for that pattern. The STDP weight

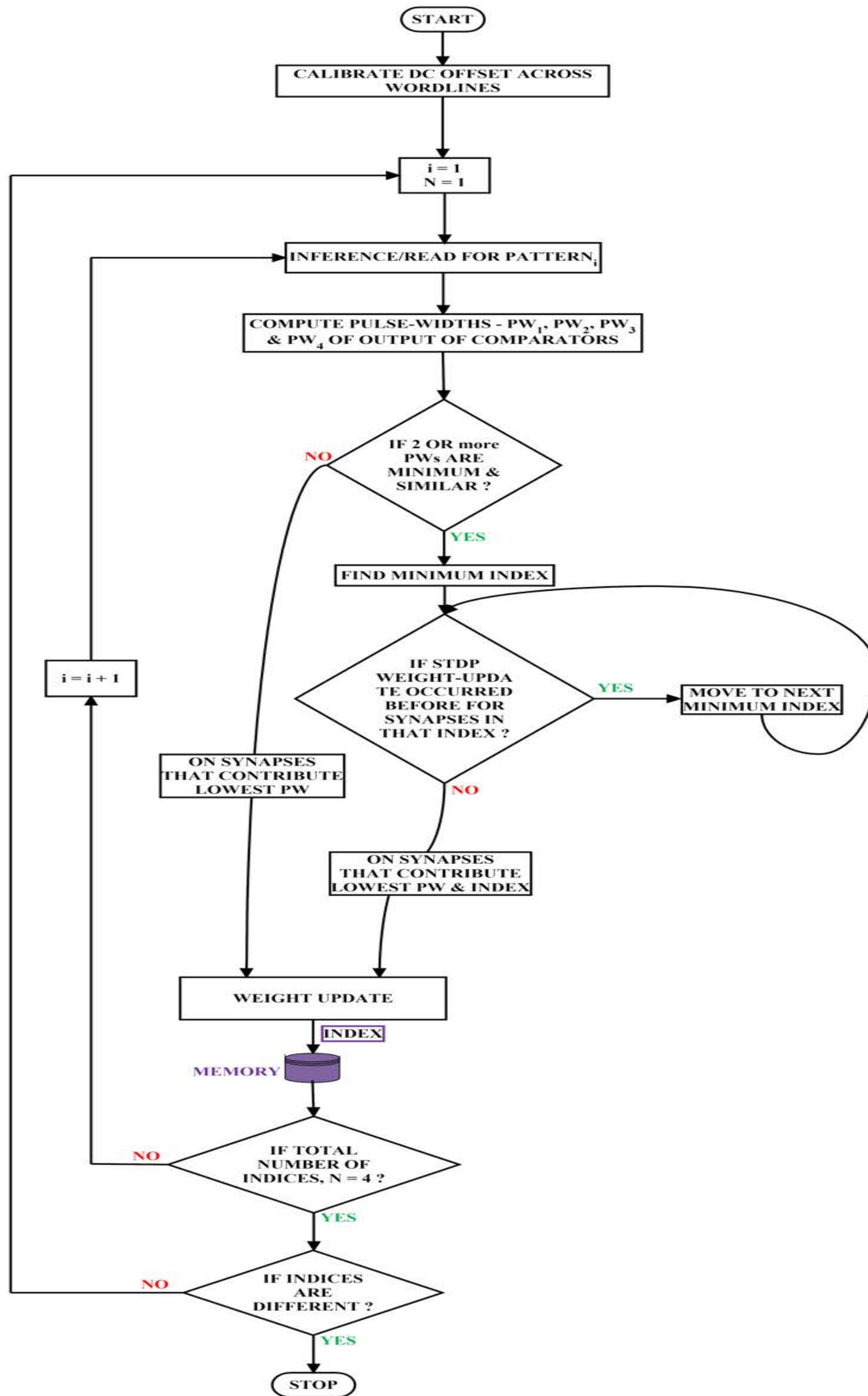


Figure 3-66: Flowchart for implementation of STDP learning rule on calibrated crossbar for pattern recognition.

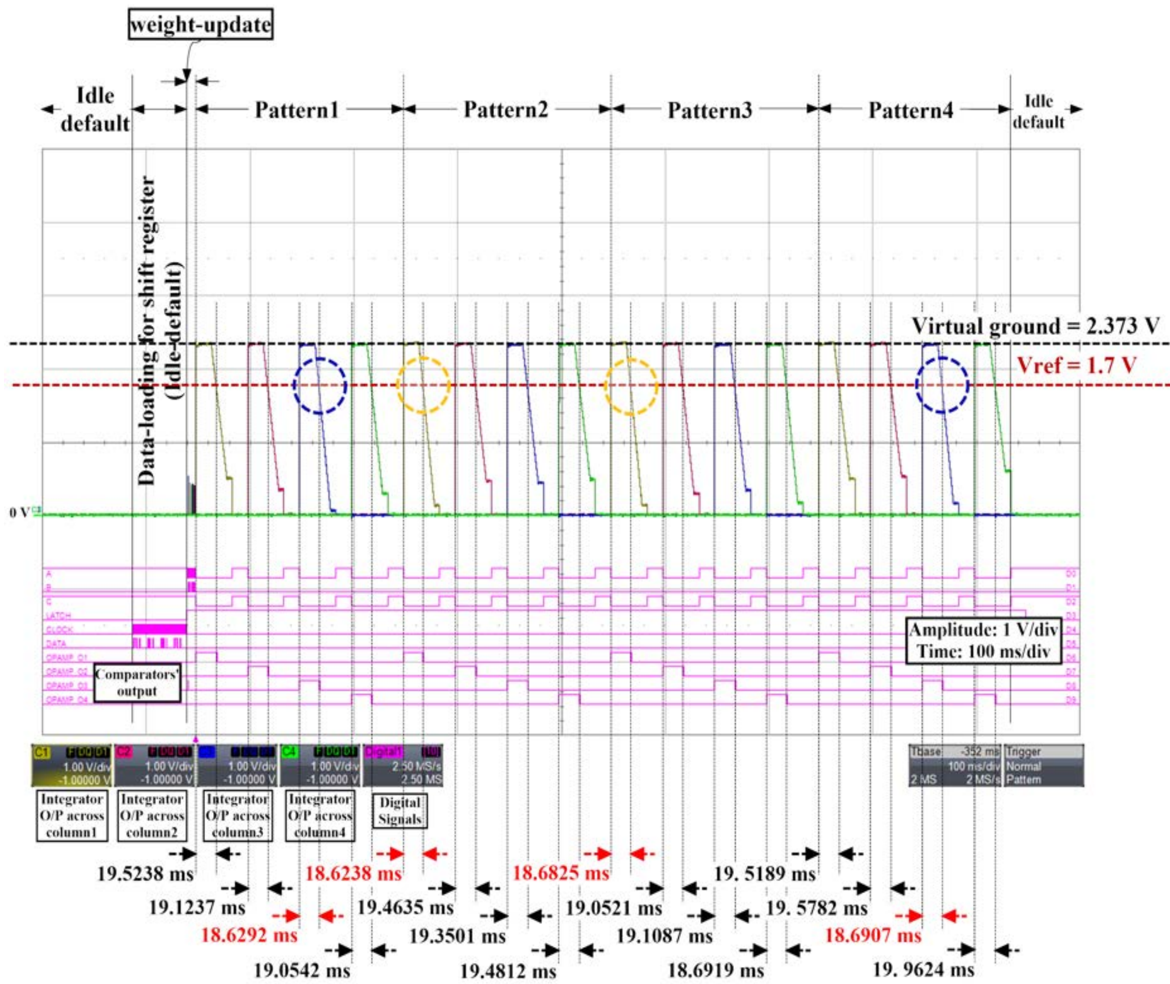


Figure 3-67: Output of integrators and comparators with other digital signals during 1st weight-update.

update is done on the synapses that contribute minimum pulse-width, whereas the rest of the weights are kept untouched. When two or more pulse-widths are similar and minimum, priority is given to the one that has minimum index number and those contributing synapses which did not face weight update earlier. The index number is the number of the bitline or column. The trained patterns and its index number are both stored in a register, which is used later to learn future patterns in order to check if the synapses faced weight update earlier or not when two or more pulse-widths become minimum and similar. The integrators have feed-back capacitors whose capacitance is 0.64 µF. Fig. 3-67 shows the output of the integrators and comparators with other digital signals during 1st weight-update, where 'col3' spiked earlier for both 'pattern-

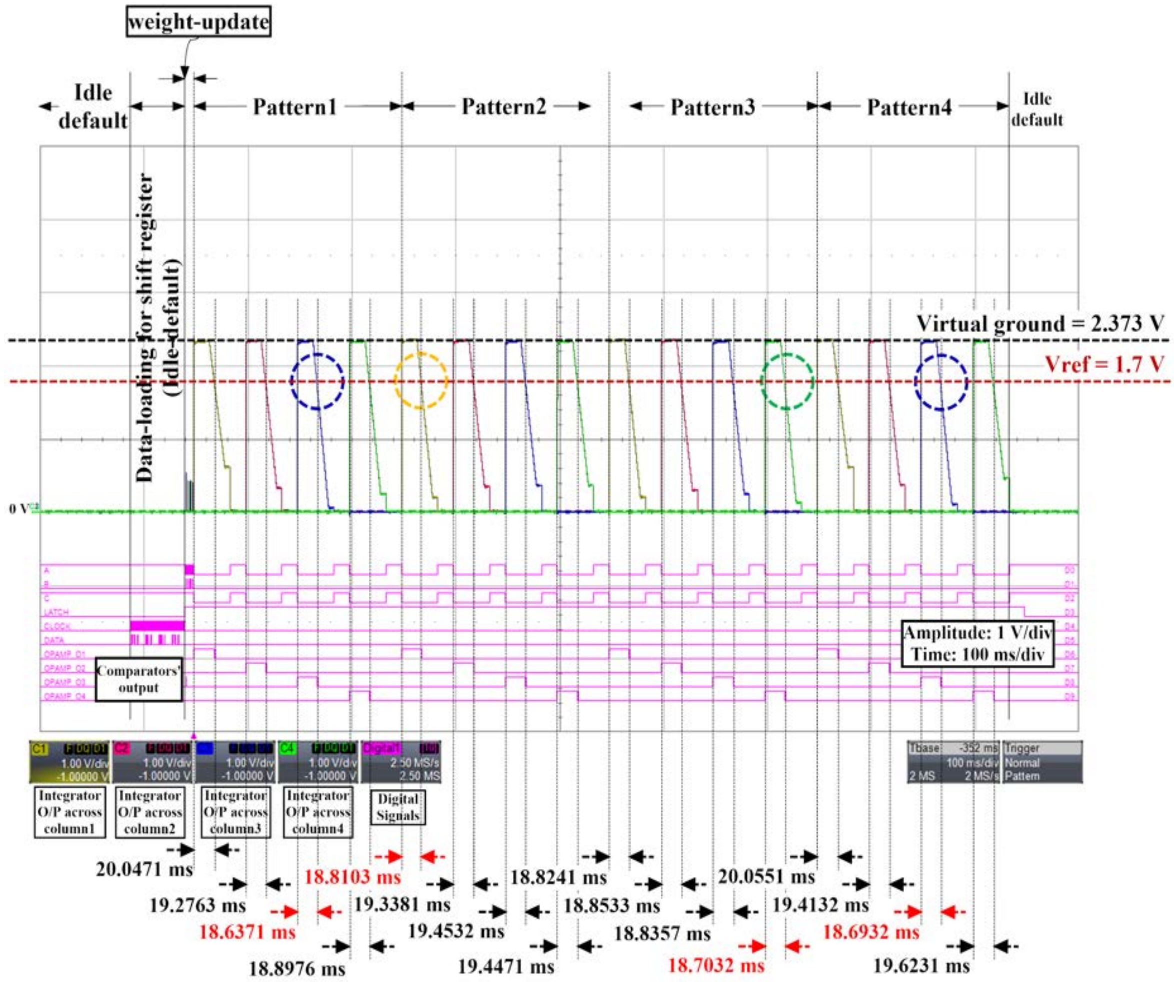


Figure 3-68: Output of integrators and comparators with other digital signals during 2nd weight-update.

1' and 'patterns-4'. Also, 'col1' spiked earlier for both 'pattern-2' and 'patterns-3'. Fig. 3-68 shows the output of the integrators and comparators with other digital signals during 2nd weight-update, where 'col3' spiked earlier for both 'pattern-1' and 'patterns-4'. Here, there is some learning and due to this 'col1' spiked earlier for 'pattern-2' and 'col4' spiked earlier for 'pattern-3'. Fig. 3-69 shows the output of the integrators and comparators with other digital signals during 3rd weight-update, where the synapses had almost learned weights. Here, 'col1' spiked earlier for 'pattern-2', 'col2' spiked earlier for 'pattern-4', 'col3' spiked earlier for 'pattern-1', and 'col4' spiked earlier for 'pattern-3'. Until we have different index numbers the weight-update is carried out.

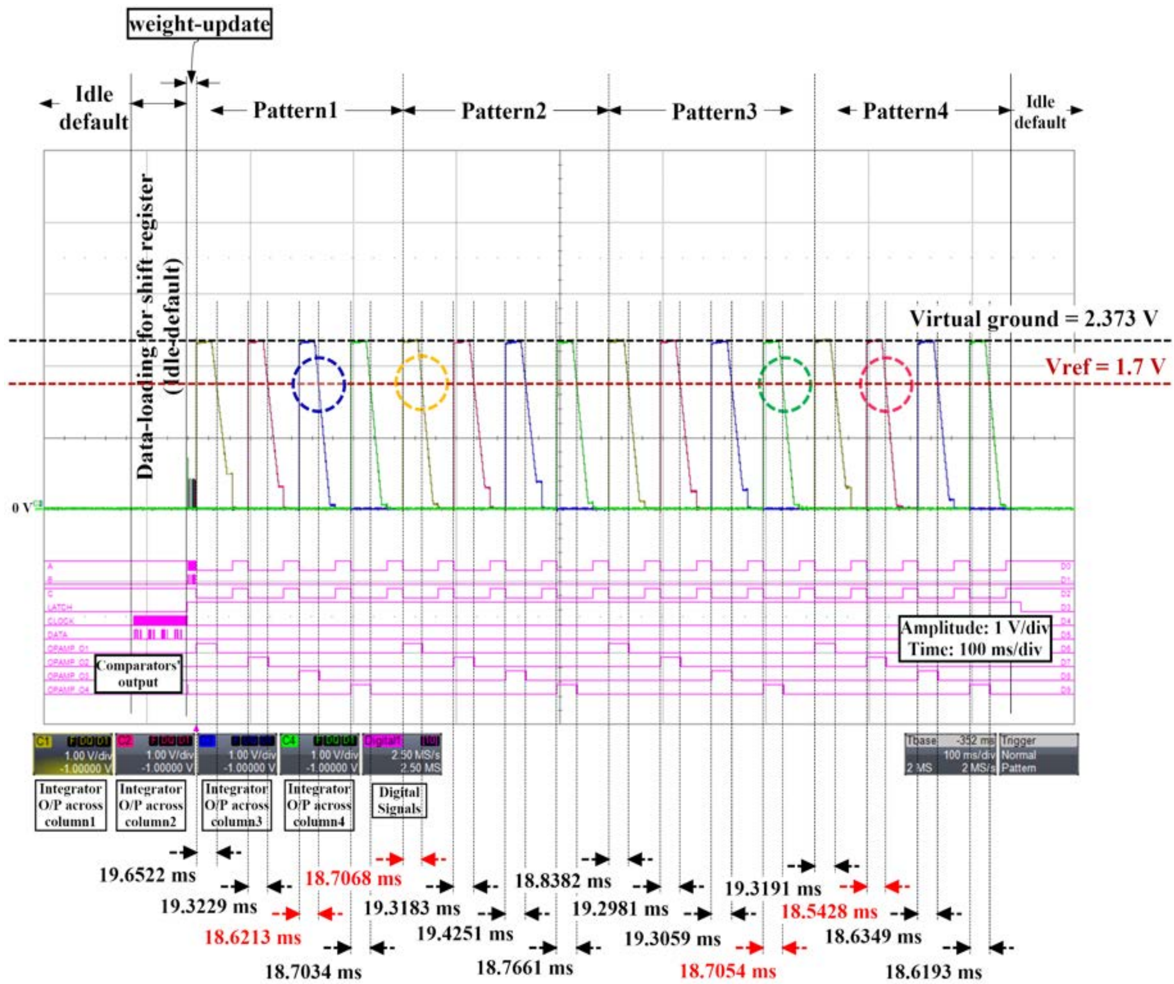


Figure 3-69: Output of integrators and comparators with other digital signals during 3rd weight-update.

Chapter 4

MCN Attenuator for Efficient Memristive Crossbar Read-Out[1]

4.1 Need for a current attenuator

In a typical memristor-based fully connected feed-forward neural network that mainly comprises the pre-synaptic neurons, the memristive crossbar and the post-synaptic neurons [129]. As the LRS current across a crossbar line is high due to the typical low resistance during an inference read operation (after a 'write' operation), an extremely large integrating capacitor would be needed (larger than nFs) at the post-synaptic neuron for reasonable integration speed, making IC integration impossible. Hence, a current attenuator is needed to scale down the read current.

4.2 Design of Modified Current Normalizer (MCN attenuator)

Several current attenuating strategies exist such as Gilbert's current normalizer circuit [130], MOS-ladder [131] and the WTA [132] based current attenuator [133]. The proposed current normalizer considers Gilbert's current normalizer circuit as a refer-

[1]This chapter has been published as a paper [128]

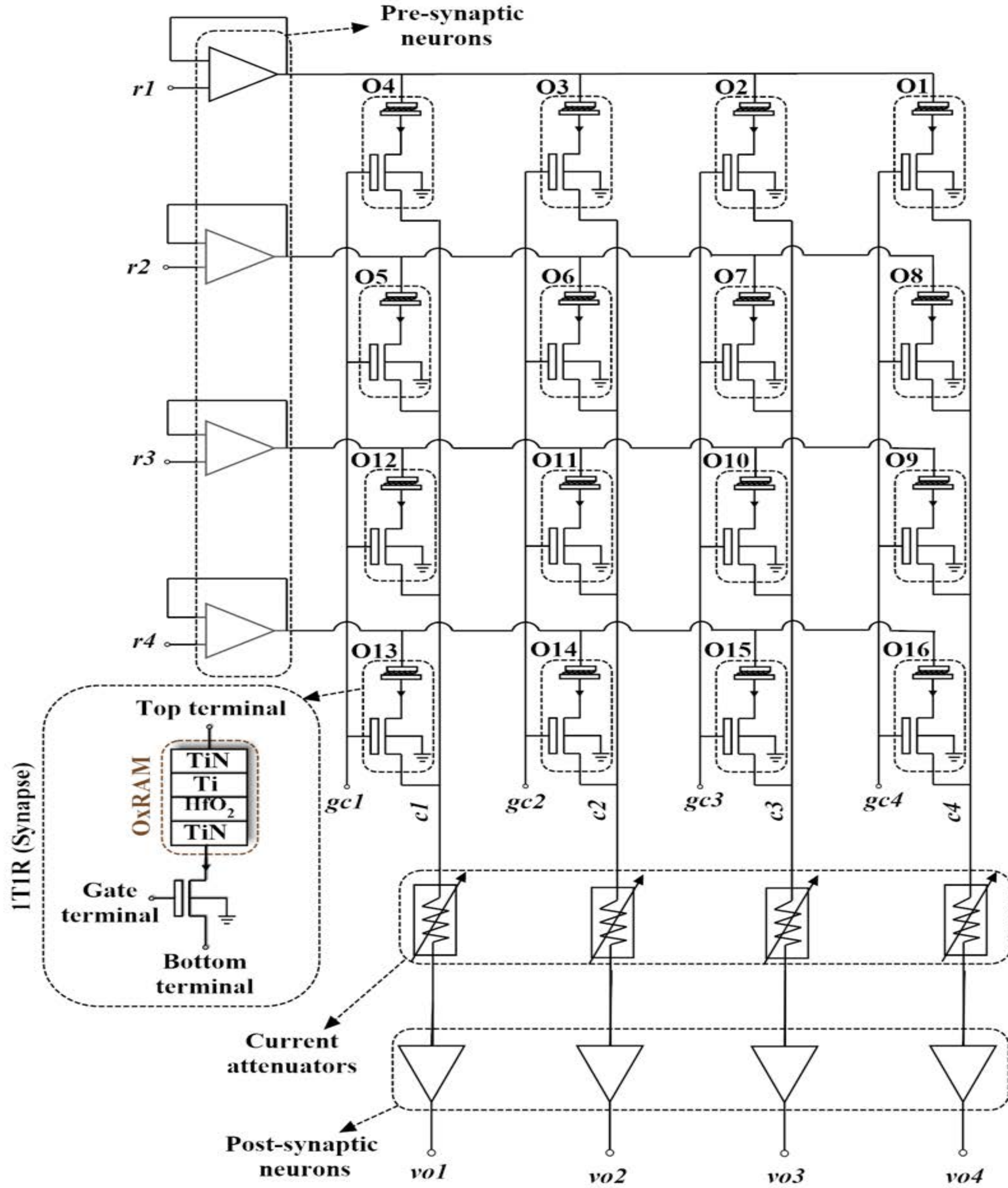


Figure 4-1: Scheme of a 4 × 4 1T1R crossbar with pre-synaptic neurons, current attenuators and post-synaptic neurons.

ence, on which modifications are done to ease the attenuation of an inference current. The idea is based on creating a splitting of the inference current by a factor of about two by using a MOS biased in the ohmic region. This Chapter proposes a Modified Current Normalizer (MCN) circuit for attenuating a crossbar read-out line current. Simulation results taking into account the effect of PVT variations are shown to validate the proposed circuit technique.

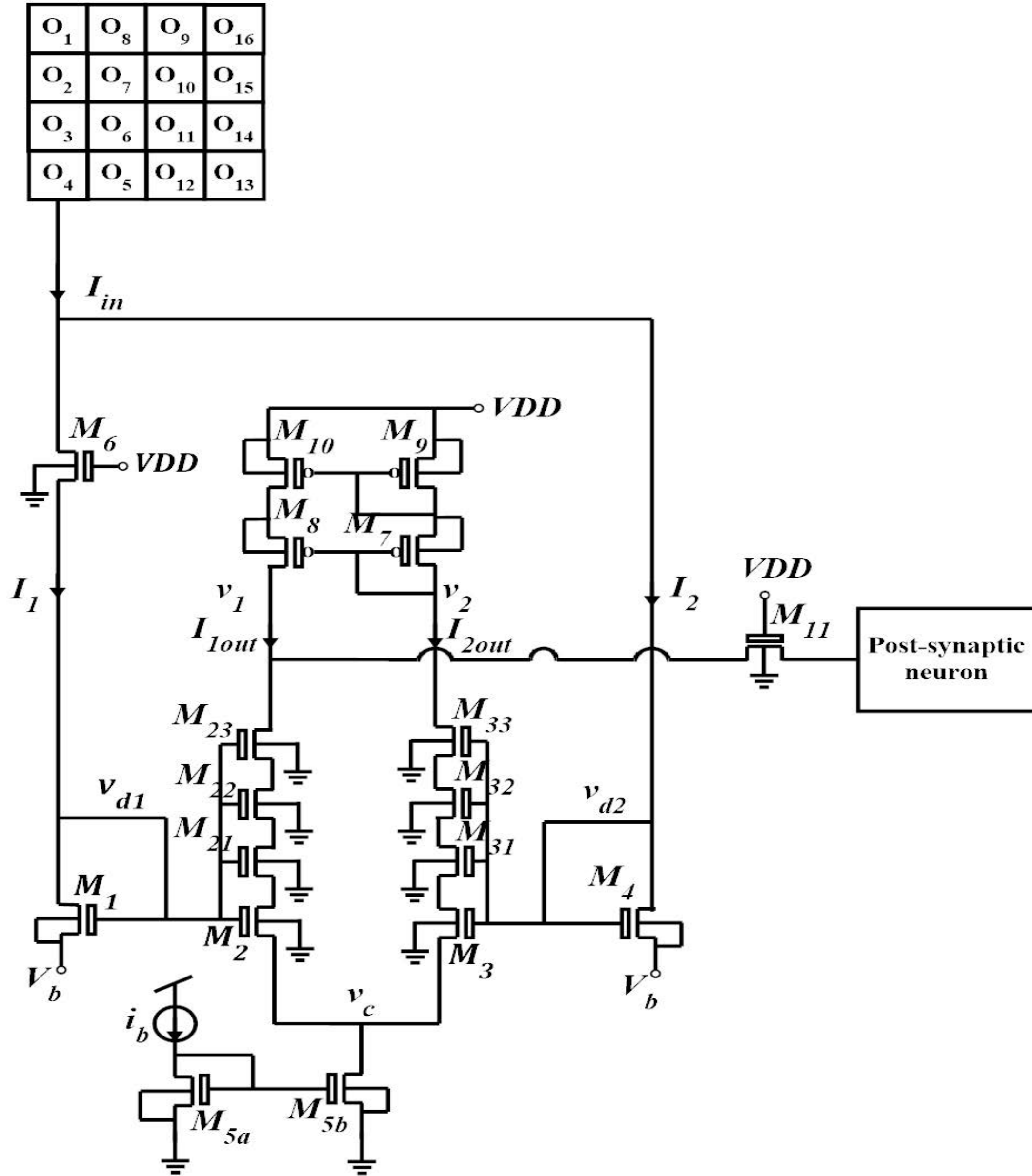


Figure 4-2: Details of MCN circuit schematics.

Fig. 4-1 shows a fully connected 4 × 4 feed-forward neural network with pre-synaptic neurons, current attenuators, and post-synaptic neurons. Circuit elements labeled as $O_{1,2,3,....,16}$ comprise the 1T1R based memristor-selector synaptic devices. Each row has a pre-synaptic neuron, which is made up of a pulse-shaping digital block and an opamp that is finely calibrated [134]. Each column is connected to a current attenuator circuit, followed by a post-synaptic neuron. The post-synaptic neuron comprises a CMOS integrate-and-fire neuron, which integrates the inference current

Table 4.1: MOSFET-sizes and biases of proposed MCN circuit.

Parameter	Value
Supply voltage	$VDD = 4.8$ V
Size of $M_{1,4}$	$(\frac{W}{L})_{1,4} = (\frac{10\,\mu m}{1\,\mu m})$
Size of $M_{2,3}$	$(\frac{W}{L})_{2,3} = (\frac{1\,\mu m}{1\,\mu m})$
Size of $M_{21,22,23}$	$(\frac{W}{L})_{21,22,23} = (\frac{1\,\mu m}{1\,\mu m})$
Size of $M_{31,32,33}$	$(\frac{W}{L})_{31,32,33} = (\frac{1\,\mu m}{1\,\mu m})$
Size of $M_{5,7,8}$	$(\frac{W}{L})_{5,7,8} = (\frac{1\,\mu m}{1\,\mu m})$
Size of M_6	$(\frac{W}{L})_6 = (\frac{8\,\mu m}{1\,\mu m})$
Size of $M_{9,10,11}$	$(\frac{W}{L})_{9,10,11} = (\frac{1\,\mu m}{1\,\mu m})$
V_b	*0.5* V
i_b	*25* nA

and generates an output spike when a threshold is reached. The OxRAM devices are initially formed to bring them from the Pristine Resistance State (PRS). Later, they can be switched in binary mode, i.e. either SET or RESET [43]. For reading the memristance of the OxRAMs, small read voltages of 0.2 or 0.3 V are applied across the rows, so that the aggregated inference read currents across columns are scaled-down and integrated.

We propose an MCN based attenuator, which mainly includes having a MOS-resistor and a current-mirror at the load on Gilbert's two-input current normalizer circuit [130]. Fig. 4-2 shows the proposed MCN circuit connected between a column of the crossbar and a post-synaptic neuron. It works by inserting an ohmic biased MOS (M_6) which creates a current splitting of the crossbar column current I_{in} by a factor of about two. This MOS-resistor creates a small resistance R_{M_6} which is given by

$$R_{M_6} = \frac{1}{\mu_n \cdot C_{ox} \cdot (V_{GS} - V_T)} \cdot \frac{L_{M_6}}{W_{M_6}} \tag{4.1}$$

where μ_n is the charge-carrier effective mobility, C_{ox} is the gate oxide capacitance, V_{GS} is the gate-to-source voltage, V_T is the threshold voltage, L_{M_6}, and W_{M_6} are the length and width of the MOS-resistor. From Fig. 4-2, and knowing transistors M_1 and M_4 will be biased in strong inversion because of the large input current I_{in}

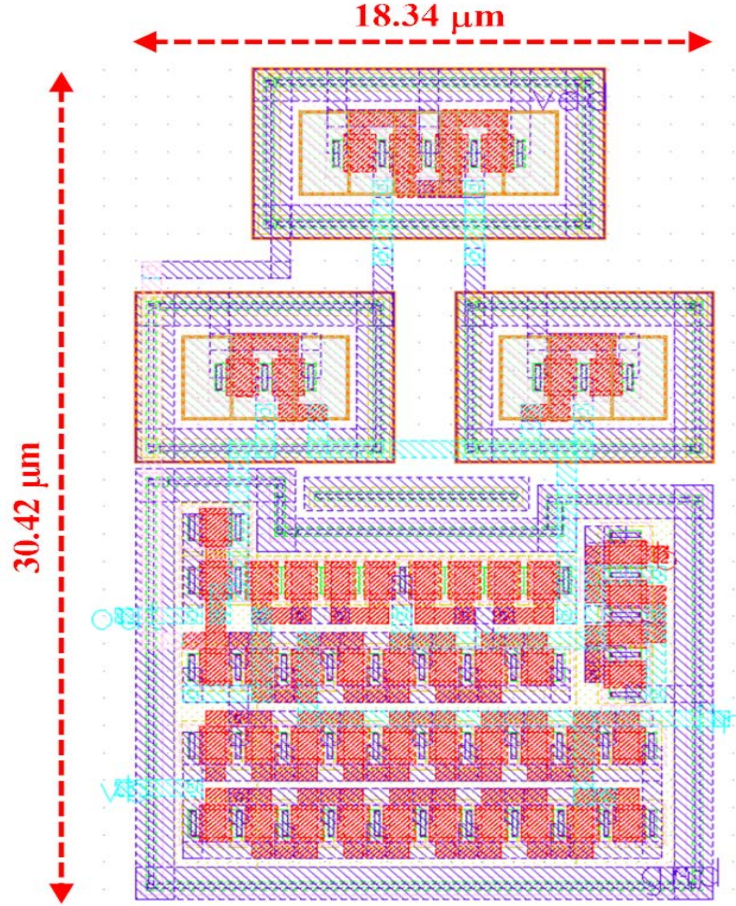


Figure 4-3: Layout view of the MCN circuit.

$$\Delta V_d = v_{d2} - v_{d1} = I_1 R_{M_6} \tag{4.2}$$

$$I_i = \frac{\mu_n \cdot C_{ox} \cdot W_{1,4}}{2L_{1,4}}(v_{di} - V_b - V_T)^2 \tag{4.3}$$

which results in

$$\frac{I_1}{I_2} = (1 - \frac{\Delta V_d}{v_{d2} - V_b - V_T})^2 = (1 - \frac{\sqrt{\frac{\mu_n \cdot C_{ox} \cdot W_{1,4}}{2L_{1,4}}} I_1 R_{M_6}}{\sqrt{I_2}})^2 \tag{4.4}$$

I_1 and I_2 are the input currents shown in Fig.3. MOSFETs M_1 and M_4 have same size. If we can assume that $R_{M_6} \ll \sqrt{I_2}/(\sqrt{\frac{\mu_n \cdot C_{ox} \cdot W_{1,4}}{2L_{1,4}}} I_1)$, then

$$I_1 \simeq I_2 \simeq \frac{I_{in}}{2} \qquad with \qquad R_{M_6} \ll \frac{1}{\sqrt{\frac{\mu_n \cdot C_{ox} \cdot W_{1,4}}{2L_{1,4}}}} \sqrt{\frac{2}{I_{in}}} \tag{4.5}$$

From eq. (4.2) we obtain

$$\Delta V_d \simeq \frac{I_{in} R_{M_6}}{2} \tag{4.6}$$

The differential pair M_2, M_3 in Fig. 4-2 will be biased in weak inversion because i_b is intentionally made very small. Consequently,

$$I_{outi} = \frac{2n \cdot \mu_n \cdot C_{ox} \cdot U_T^2 \cdot W_{2,3}}{L_{2,3}} \cdot e^{\frac{v_{di} - v_c}{nU_T}} \tag{4.7}$$

$$i_b = I_{out1} + I_{out2} \tag{4.8}$$

where n is the subthreshold slope factor and U_T is the thermal voltage. From here

$$\frac{I_{out2}}{I_{out1}} = e^{\frac{\Delta V_d}{nU_T}} = e^x \tag{4.9}$$

where from eq. (4.6) $x = (I_{in} R_{M_6})/(2nU_T)$. Straight forward calculations yield

$$I_{out2} - I_{out1} = i_b \frac{e^x - 1}{e^x + 1} \tag{4.10}$$

If we can assume that $x \ll 1$ (which is equivalent to $R_{M_6} \ll 2nU_T/I_{in}$), then

$$I_{out} = I_{out2} - I_{out1} \simeq i_b \frac{x}{2} = \frac{i_b R_{M_6}}{4nU_T} I_{in} \tag{4.11}$$

Therefore, the MCN circuit output current is, under some assumptions for R_{M_6}, proportional to I_{in}, and the proportionality factor can be controlled by bias current i_b. This will allow us to scale down the input current by four orders of magnitude. Transistor M_{11} is an optional switch, which is used to isolate the post-synaptic neuron for test purposes. Differential MOSFET groups $M_{2,21,22,23}$ and $M_{3,31,32,33}$ are split along the length dimension to keep them in square shape and exploit inter-digitated layout to minimize mismatch due to gradients. The MCN circuit is implemented

using a 130nm CMOS technology. Fig. 4-3 shows the layout view of the designed MCN circuit. Table 4.1 shows the sizes of MOSFETs and biases used in the proposed MCN circuit, shown in Fig. 4-2. An optimal design procedure is carried out on the MCN circuit to keep the mismatch as low as possible.

4.3 Simulation Results

Since we did not have the chance to submit the design of the attenuator for fabrication, we are presenting only simulation results. Various simulation results of the proposed MCN approach are obtained in comparison with other approaches (MOS-ladder and WTA type circuit) considering mismatch-and-process variations, temperature, and input-referred noise. Fig. 4-4 (a) shows how the output current of the MCN circuit depends on the input current I_{in}.

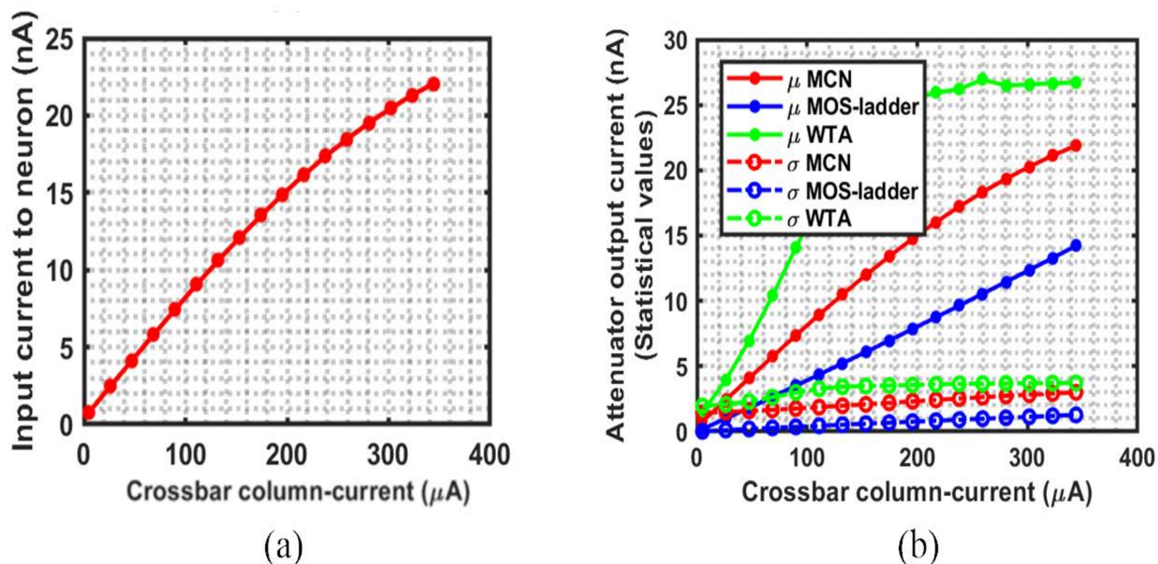


Figure 4-4: (a) Minimum to maximum crossbar column inference current Vs Input current to neuron, (b) Comparison of average and standard deviation of output current for different attenuators: MCN circuit, MOS-ladder circuit, and WTA circuit considering process and mismatch variations with 100 Monte Carlo runs.

For a read voltage of 0.3 V, when the 1T1R memristive device used here is in typical LRS of 13.9 kΩ the inference read current is 21.5 µA. And when the device is in typical HRS of 1MΩ the inference read current results in 333 nA. The minimum

column current is when all the OxRAMs in an active column are in HRS and the maximum column current is when all the OxRAMs in an active column are in LRS. Fig. 4-4 (a) shows that the MCN output current (input current to neuron) is linear when the inference current stays below about 300 µA.

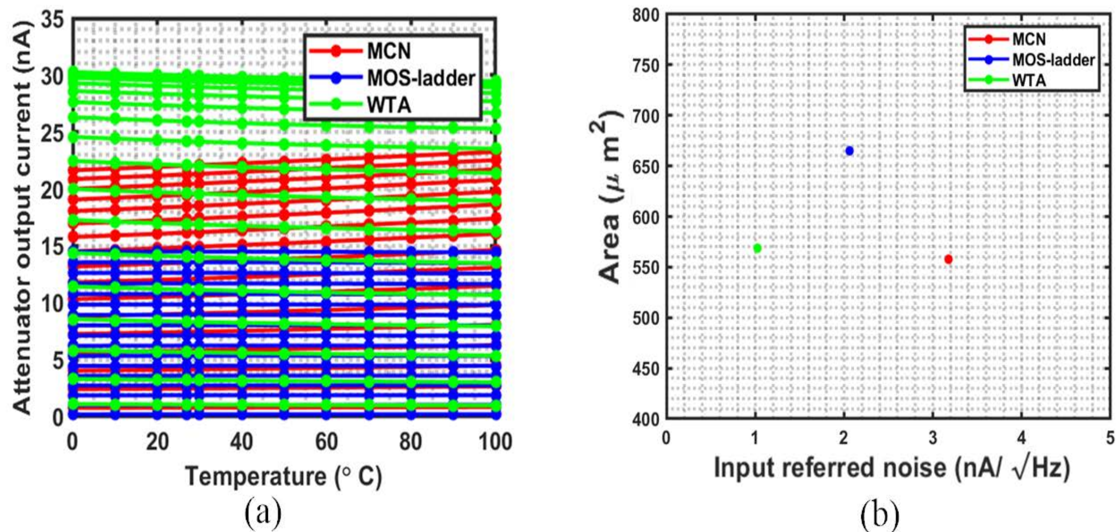


Figure 4-5: (a) Temperature variations of the output current of the MCN circuit, MOS-ladder and WTA circuit for different values of output currents, (b) Area and input-referred noise of the MOS circuit, MOS-ladder, and WTA circuit.

Fig. 4-4 (b) shows the statistical output current results (both mean-µand sigma-σ) for 100 Monte Carlo [2] runs of the MCN circuit, a MOS-ladder, and a WTA type circuit considering both process and mismatch variations. Fig. 4-5 (a) shows how different column inference currents (minimum to maximum) vary due to temperature. We can see that temperature variations have a fairly small impact on the output currents (less than 10%). The input-referred noise for the frequency range 0.01 Hz to 100 MHz is 3.178 $n\text{A}/\sqrt{Hz}$ for the MCN circuit, 2.06 $n\text{A}/\sqrt{Hz}$ for the MOS-ladder and 1.02 $n\text{A}/\sqrt{Hz}$ for the WTA type circuit. A comparison of both area and input-referred noise for different attenuators is shown in Fig. 4-5 (b).

[2]Monte Carlo simulations includes both process (wafer-to-wafer variations), On Chip Variations (OCV) like device mismatch.

Chapter 5

Conclusion and Future work

In the first part of this thesis, we discussed- three different test-benches for characterizing memristors. The first method uses a commercial memristor- ‘NeuroBit’ and using the ‘ArC One memristor characterization platform’ to characterize it. The second test-bench comprises a full-custom designed experimental set-up, whose design involved implementing crossbar in MAD200 PDK that underwent hybrid monolithic integration of OxRAM-based memristors above the CMOS. Customized PCBs are designed to test both 4×4 and 8×8 1T1R crossbars. The third test-bench is a wafer-level characterization made on MIM-based memristors that were grown using the ALD process using different dielectrics.

The second part of the thesis finds a way to increase scalability and explains the problem of DC offset, which becomes the bottleneck for applying tiny read pulses across memristive crossbars. The thesis also comes with a solution to implement bulk-based three-stage DC offset calibration technique on a 4×4 1T1R crossbar. This chip is designed using MAD200 PDK, which is followed by a custom-made test-PCB to test the chip. The experimental results of the calibration scheme were verified with the simulation one. System-level experimental results were obtained on the calibrated crossbar- for template matching, for recognizing patterns using SSSP, and for recognizing patterns using STDP learning rule.

The third part of the thesis involves identifying the layout-area constrain in a typical memristor-based single-layer neural network. The thesis also comes with a new

current conveyor to attenuate current by a factor of 10^4, whose results are compared with existing ones in terms of mismatch, noise, and area.

The future work relies on challenging scalability (by increasing crossbar size) for such offset calibration techniques and exploring other on-chip calibration schemes for implementation on the memristive crossbar, which can be eventually used for low-power learning or programming.

References

[1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics Magazine*, p. 4, 1965.

[2] C. Hu, "New sub-20nm transistors — Why and how," pp. 460–463, June 2011.

[3] W.Arden, M.Brillouët, P.Cogez, M.Graef, B.Huizing, and R.Mahnkoph, "More-than-Moore," *International Technology Roadmap for Semiconductors*, 2008.

[4] J. von Neumann, "The Computer and the Brain," *New Haven/London: Yale University Press*, 1958.

[5] C. Mead, "Analog VLSI and Neural Systems," *Addison Wesley*, 1989.

[6] C. A. Mead and M. A. Mahowald, "A silicon model of early visual processing," *Neural Networks*, vol. 1, no. 1, pp. 91–97, 1988.

[7] K. Boahen, "Neurogrid: Emulating a Million Neurons in the Cortex," *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, p. 6702, August 2006.

[8] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, "Neurogrid: Emulating a Million Neurons in the Cortex," *Proceedings of the IEEE*, vol. 102, pp. 699–716, April 2014.

[9] J. Schemmel, D. Brüderle, A. Grübl, M. Hock, K. Meier, and S. Millner, "A wafer-scale neuromorphic hardware system for large-scale neural model-

ing," *2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1947–1950, May 2010.

[10] J. Schemmel, L. Kriener, P. Müller, and K. Meier, "An accelerated analog neuromorphic hardware system emulating NMDA- and calcium-based non-linear dendrites," *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 2217–2226, May 2017.

[11] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, pp. 668–673, August 2014.

[12] S. B. Furber, D. R. Lester, L. A. Plana, J. D. Garside, E. Painkras, S. Temple, and A. D. Brown, "Overview of the SpiNNaker System Architecture," *IEEE Transactions on Computers*, vol. 62, pp. 2454–2467, Dec 2013.

[13] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y.-H. Weng, A. Wild, Y. Yang, and H. Wang, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," *IEEE Micro*, vol. 38, pp. 82–99, January 2018.

[14] D. Ma, J. Shen, Z. Gu, M. Zhang, X. Zhu, X. Xu, Q. Xu, Y. Shen, and G. Pan, "Darwin: A neuromorphic hardware co-processor based on spiking neural networks," *Journal of Systems Architecture*, vol. 77, pp. 43–51, January 2017.

[15] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, and G. Indiveri, "A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses," *Front. Neurosci.*, vol. 9, April 2015.

[16] C. Frenkel, M. Lefebvre, J.-D. Legat, and D. Bol, "A 0.086 — mm^2 12.7-pJ/SOP 64k-Synapse 256-Neuron Online-Learning Digital Spiking Neuromorphic Processor in 28-nm CMOS," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, pp. 145–158, Feb 2019.

[17] S. Moradi, N. Qiao, F. Stefanini, and G. Indiveri, "A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs)," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, pp. 106–122, February 2018.

[18] S. Furber, "Large-scale neuromorphic computing systems," *Journal of Neural Engineering*, vol. 13, pp. 1–14, Feb 2016.

[19] L. Camuñas-Mesa, B. L. Barranco, and T. S. Gotarredona, "Neuromorphic Spiking Neural Networks and Their Memristor-CMOS Hardware Implementations," *Materials*, vol. 12, August 2019.

[20] "Neuroshield: Neuromem neural network as a shield or a usb extension." `https://general-vision.com/hardware/neuroshield/`.

[21] L.O.Chua, "Memristor- the missing circuit element," *IEEE Transactions on Circuit Theory*, vol. 18, pp. 507–509, Sept. 1971.

[22] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. William, "The missing memristor found," *Nature*, vol. 453 (7191), pp. 80–83, 2008.

[23] S. H. Jo, T. Chang, T. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," *Nano Lett.*, vol. 10, pp. 1297–1301, 2010.

[24] Z. Biolek, D. Biolek, and V. Biolková, "Spice model of memristor with nonlinear dopant drift," *Radio Eng.*, vol. 18 (2), pp. 210–214, 2009.

[25] J. A. Pérez-Carrasco, T. S.-G. C. Zamarreño-Ramos, and B. Linares-Barranco, "On neuromorphic STDP memristive systems," *in proceedings of the IEEE In-*

ternational Symposium on Circuits and System (ISCAS 2010), pp. 1659–1662, 2010.

[26] S.Shin, K.Kim, and M.S.-Kang, "Compact models for memristors based on charge-flux constitutive relationships," *IEEE Trans. CAD Int. Circuits Syst.*, vol. 29, pp. 590–598, 2010.

[27] R. Waser and A. Aono, "Nanoionics-based resistive switching memories," *Nat. Mater*, vol. 6, pp. 833–840, 2007.

[28] J. J. Yang, F. Miao, M. D. Pickett, D. A. A. Ohlberg, D. R. Stewart, C. N. Lau, and R. S. Williams, "The mechanism of electroforming of metal oxide memristive switches," *Nanotechnology*, vol. 20, no. 21, 2009.

[29] S. A. Wolf, J. Lu, M. R. Stan, E. Chen, and D. M. Treger, "The promise of nanomagnetics and spintronics for future logic and universal memory," *Proc. IEEE 98*, pp. 2155–2168, 2010.

[30] A. Sheikholeslami and P. G. Gulak, "A survey of circuit innovations in ferro-electric random-access memories," *Proc. IEEE 88*, pp. 667–689, 2000.

[31] W. Zhao, S. Chaudhuri, C. Accoto, J. O. Klein, D. Ravelosona, C. Chappert, and P. Mazoyer, "High density Spin-Transfer Torque (STT)-MRAM based on cross-point architecture," *in 2012 4th IEEE International Memory Workshop (IMW)*, May 2012.

[32] R. W. (ed.), "Nanoelectronics and Information Technology," *3rd edn. (Wiley-VCH), Berlin*, 2012.

[33] J. J. Yang, M. D. Pickett, X. Li, D. A. A. Ohlberg, D. R. Stewart, and R. S. Williams, "Memristive switching mechanism for metal/oxide/metal nanodevices," *Nat. Nanotechnology*, vol. 3, 429, 2008.

[34] S. Yu, X. Guan, and H. P. Wong, "Conduction mechanism of TiN/HfOx/Pt resistive switching memory: a trap-assisted-tunneling model," *Appl. Phys. Lett.*, vol. 99, July 2011.

[35] I. Valov, R. Waser, J. R. Jamerson, and M. N. Kozicki, “Electrochemical metallization memories- fundamentals, applications, prospects,” *Nanotechnology*, vol. 22, May 2011.

[36] “NeuRAM3- WP3 update,” *Review meeting- Internal document*, August 2016.

[37] R. Waser, R. Dittmann, G. Staikov, and K. Szot, “Redox-based resistive switching memories-nanoionic mechanism, prospects, and challenges,” *Adv. Mater.*, vol. 21 (25-26), 2009.

[38] B. J. Choi, D. S. Jeong, S. Kim, C. Rohde, S. Choi, J. H. Oh, H. J. Kim, C. S. Hwang, K. Szot, R. Waser, B. Reichenberg, and S. Tiedke, “Resistive switching mechanism of TiO_2 thin films grown by atomic-layer deposition,” *J.Appl.Phys.*, vol. 98, no. 3, 2005.

[39] Y. M. Kim and J. S. Lee, “Reproductive resistance switching characteristics of Hafnium oxide-based nonvolatile memory devices,” *J.Appl.Phys.*, vol. 104, no. 11, 2008.

[40] Y. Wu, S. Yu, B. Lee, and P. Wong, “Low-power $TiN/Al_2O_3/Pt$ resistive switching device with sub-20μA switching current and gradual resistance modulation,” *J.Appl.Phys.*, vol. 110, no. 9, 2011.

[41] L. Chen, Q. Q. Sun, J. J. Gu, Y. Xu, S. J. Ding, and D. W. Zhang, “Bipolar resistive switching characteristics of atomic layer deposited Nb_2O_5 this films for nonvolatile memory application,” *J.Appl.Phys.*, vol. 11, no. 3, pp. 849–852, 2011.

[42] K. Szot, W. Speier, G. Bihlmayer, and R. Waser, “Switching the electrical resistance of individual dislocations in single-crystalline $SrTiO_3$,” *Nat.Mater.*, vol. 5, pp. 312–320, 2006.

[43] D. Garbin, O. Bicher, E. Vianello, Q. Rafhay, C. Gamrat, L. Perniola, G. Ghibaudo, and B. DeSalvo, “Variability-tolerant Convolutional Neural

Network for Pattern Recognition applications based on OxRAM synapses," *2014 IEEE International Electron Devices Meeting*, pp. 28.4.1–28.4.4, December 2014.

[44] D. Garbin, E. Vianello, O. Bicher, Q. Rafhay, C. Gamrat, G. Ghibaudo, B. DeSalvo, and L. Perniola, "HfO_2-based OxRAM Devices as Synapses for Convolutional Neural Networks," *IEEE Transactions on Electron Devices*, vol. 62, pp. 2494–2501, August 2015.

[45] S. D. Ha and S. Ramanathan, "Adaptive oxide electronics: a review," *J.Appl.Phys.*, vol. 110, August 2011.

[46] J. N. Reynolds, "Crossbar switch," *U.S. Patent 1, 131, 734*, March 1915.

[47] S. C. Goldstein and M. Budiu, "Nanofabrics: spatial computing using molecular electronics," *Proceedings 28th Annual International Symposium on Computer Architecture*, pp. 178–189, August 2002.

[48] A. DeHon, "Array-based architecture for molecular electronics," *Presented at the 1st Workshop Non-Silicon Computation (NSC-1), Boston, MA.*, August 2002.

[49] P. J. Kuekes, J. R. Heath, and R. S. Williams, "Molecular wire crossbar memory," *U.S.Patent 6 128 214*, Oct. 2000.

[50] A. R. Pease, J. O. Jeppesen, J. F. Stoddart, Y. Luo, C. P. Collier, and J. R. Heath, "Switching devices based on interlocked molecules," *Acc. Chem. Res.*, vol. 34, pp. 433–444, April 2001.

[51] S. Folling, O. Turel, and K. Likharev, "Single-electron latching switches as nanoscale synapses," *IJCNN'01. International Joint Conference on Neural Networks. Proceedings*, pp. 216–221, July 2001.

[52] O. Turel and K. Likharev, "Cross-nets possible neuromorphic networks based on nanoscale components," *Int. J. Circuit Theory Appl.*, vol. 31, pp. 37–53, Januray 2003.

[53] Y. Cassuto, S. Kvatinsky, and E. Yaakobi, "Sneak-path constraints in memristor crossbar arrays," *2013 IEEE International Symposium on Information Theory*, pp. 156–160, July 2013.

[54] E. Linn, R. Rosezin, C. Kugeler, and R. Waser, "Complementary resistive switches for passive nanocrossbar memories," *Nat. Mater.*, vol. 9, pp. 403–406, April 2010.

[55] H. Manem, G. S. Rose, X. He, and W. Wang, "Design considerations for variation tolerant multilevel CMOS/Nano memristor memory," *GLSVLSI '10: Proceedings of the 20th symposium on Great lakes symposium on VLSI*, pp. 287–292, May 2010.

[56] M. Zidan, A. Eltawil, F. Kurdahi, H. Fahmy, and K. Salama, "Memristor multiport readout: A closed-form solution for sneak paths," *Nanotechnology, IEEE Transactions on*, vol. 13, pp. 274–282, March 2014.

[57] M. E. Fouda, A. M. Eltawil, and F. J. Kurdahi, "On one step row readout technique of selector-less resistive arrays," *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 72–75, August 2017.

[58] M. Shevgoor, N. Muralimanohar, R. Balasubramonian, and Y. Jeon, "Improving memristor memory with sneak current sharing," *2015 33rd IEEE International Conference on Computer Design (ICCD)*, pp. 549–556, October 2015.

[59] P. O. Vontobel, W. Robinett, P. J. Kuekes, D. R. Stewart, J. Straznicky, and R. S. Williams, "Writing to and reading from a nano-scale crossbar memory based on memristors," *Nanotechnology*, vol. 20, pp. 425 204–425 223, September 2009.

[60] M. Zackriya, H. M. Kittur, and A. Chin, "A novel read scheme for large size one-resistor resistive random access memory array," *Scientific reports*, vol. 7, February 2017.

[61] A. Ciprut and G. F. Eby, "Hybrid Write Bias Scheme for Non-Volatile Resistive Crossbar Arrays," *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018.

[62] A. Ciprut and G. F. Eby, "Energy-efficient write scheme for nonvolatile resistive crossbar arrays with selectors," *IEEE Transactions on Very Large Scale Integration (*VLSI*) Systems*, vol. 26, no. 4, pp. 711–719, 2018.

[63] F. Bedeschi, R. Fackenthal, C. Resta, E. M. Donze, M. Jagasivamani, C. E. Buda, F. Pellizzer, W. D. Chow, A. Cabrini, G. M. A. Calvi, R. Faravelli, A. Fantini, G. Torelli, D. Mills, R. Gastaldi, and G. Casagrande, "A bipolar-selected phase change memory featuring multi-level cell storage," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 1, pp. 217–227, 2009.

[64] D. B. Strukov and K. K. Likharev, "CMOL FPGA: a reconfigurable architecture for hybrid digital circuits with two-terminal nanodevices," *Nanotechnology*, vol. 16, pp. 888–900, April 2005.

[65] G. S. Snider and R. S. Williams, "Nano/CMOS architectures using a field-programmable nanowire interconnect," *Nanotechnology*, vol. 18, pp. 1–10, January 2007.

[66] W. Robinett, G. Snider, D. Stewart, J. Straznicky, and R. Williams, "Demultiplexers for Nanoelectronics Constructed From Nonlinear Tunneling Resistors," *IEEE Transactions on Nanotechnology*, vol. 6, pp. 280–290, May 2007.

[67] W. Gerstner, R. Ritz, and J. L. van Hemmen, "Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns," *Biological Cybernetics*, vol. 69, pp. 503–515, 1993.

[68] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann, "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs," *Science*, vol. 275, pp. 213–215, January 1997.

[69] G. Q. Bi and M. M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *J. Neurosci.*, vol. 18, pp. 10464–10472, December 1998.

[70] G. Q. Bi and M. M. Poo, "Synaptic Modification by Correlated Activity: Hebb's Postulate Revisited," *Annual Review of Neuroscience*, vol. 24, pp. 139–166, March 2001.

[71] L. I. Zhang, H. W. Tao, C. E. Holt, W. A. Harris, and M. ming Poo, "A critical window for cooperation and competition among developing retinotectal synapses," *Nature*, vol. 395, pp. 37–44, September 1998.

[72] D. E. Feldman, "Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex," *Neuron*, vol. 27, pp. 45–56, July 2000.

[73] Y. Mu and M. M. Poo, "Spike timing-dependent LTP/LTD mediates visual experience-dependent plasticity in a developing retinotectal system," *Neuron*, vol. 50, pp. 115–125, April 2006.

[74] S. Cassenaer and G. Laurent, "Hebbian STDP in mushroom bodies facilitates the synchronous flow of olfactory information in locusts," *Nature*, vol. 448, pp. 709–713, June 2007.

[75] V. Jacob, D. J. Brasier, I. Erchova, D. Feldman, and D. E. Shulz, "Spike Timing-Dependent Synaptic Depression in the In Vivo Barrel Cortex of the Rat," *J. Neurosci.*, vol. 27 (6), pp. 1271–1284, February 2007.

[76] J. E. Rubin, R. C. Gerkin, G. Q. Bi, and C. C. Chow, "Calcium time course as a signal for spike-timing-dependent plasticity," *J. Neurophysiol.*, vol. 93, pp. 2600–2613, May 2005.

[77] D. O. Hebb, "The Organization of Behavior: A Neuropsychological Study," *New York: Wiley*, 1949.

[78] A. Delorme, L. Perrinet, and S. J. Thorpe, “Networks of integrate-and-fire neurons using Rank Order Coding B: Spike timing dependent plasticity and emergence of orientation selectivity,” *Neurocomputing*, vol. 38-40, pp. 539–545, June 2001.

[79] R. Guyonneau, R. Vanrullen, and S. J. Thorpe, “Networks of integrate-and-fire neurons using Rank Order Coding B: Spike timing dependent plasticity and emergence of orientation selectivity,” *J. Physiol Paris*, vol. 984-986, pp. 487–497, July 2004.

[80] T. Masquelier and S. J. Thorpe, “Unsupervised learning of Visual Features through Spike Timing Dependent Plasticity,” *PLOS Computational Biology*, vol. 3, February 2007.

[81] T. Masquelier and S. J. Thorpe, “Learning to recognize objects using waves of spikes and Spike Timing-Dependent Plasticity,” *The 2010 International Joint Conference on Neural Networks (IJCNN)*, vol. 3, July 2010.

[82] J. M. Young, W. J. Waleszczyk, C. Wang, M. B. Calford, B. Dreher, and K. Obermayer, “Cortical reorganization consistent with spike timing–but not correlation-dependent plasticity,” *Nature Neuroscience*, vol. 10, May 2007.

[83] L. A. Finelli, S. Haney, M. Bazhenov, M. Stopfer, and T. J. Sejnowski, “Synaptic Learning Rules and Sparse Coding in a Model Sensory System,” *PLOS Computational Biology*, vol. 4, April 2008.

[84] T. Masquelier, R. Guyonneau, and S. J. Thorpe, “Spike Timing Dependent Plasticity Finds the Start of Repeating Patterns in Continuous Spike Trains,” *PLOS ONE*, vol. 3, January 2008.

[85] T. Masquelier, R. Guyonneau, and S. J. Thorpe, “Competitive STDP-based spike pattern learning,” *Neural Comput.*, vol. 21, pp. 1259–1276, May 2009.

[86] U. Weidenbacher and H. Neumann, "Unsupervised Learning of Head Pose through Spike-Timing Dependent Plasticity," *PIT 2008:Perception in Multimodal Dialogue Systems*, vol. 5078, pp. 123–131, 2008.

[87] Y. Yang and R. Huang, "Probing memristive switching in nanoionic devices," *Nature Electronics*, vol. 1, pp. 274–287, 2018.

[88] Y. Yang, X. Zhang, L. Qin, Q. Zeng, X. Qiu, and R. Huang, "Probing nanoscale oxygen ion motion in memristive systems," *Nature Communications*, vol. 8, pp. 1–10, May 2017.

[89] B. D. Hoskins, G. C. Adam, E. Strelcov, N. Zhitenev, A. Kolmakov, D. B. Strukov, and J. J. McClelland, "Stateful characterization of resistive switching TiO_2 with electron beam induced currents," *Nature Communications*, vol. 8, pp. 1–11, December 2017.

[90] Y. Yang and W. D. Lu, "Progress in the Characterizations and Understanding of Conducting Filaments in Resistive Switching Devices," *IEEE Transactions on Nanotechnology*, vol. 15, pp. 465–472, May 2016.

[91] G.-S. Park, Y. B. Kim, S. Y. Park, X. S. Li, S. Heo, M.-J. Lee, M. Chang, J. H. Kwon, M. Kim, U.-I. Chung, R. Dittmann, R. Waser, and K. Kim, "In situ observation of filamentary conducting channels in an asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structure," *Nature Communications*, vol. 4, pp. 1–9, September 2013.

[92] "ArC One Memristor Characterisation Platform." `https://arc-instruments.co.uk/products/arc-one/`.

[93] R. Berdan, A. Serb, A. Khiat, A. Regoutz, C. Papavassiliou, and T. Prodromakis, "A μ-controller-based system for interfacing selectionless RRAM crossbar arrays," *IEEE Transactions on Electron Devices*, vol. 62, pp. 2190–2196, July 2015.

[94] A. Serb, A. Khiat, and T. Prodromakis, "An RRAM Biasing Parameter Optimizer," *IEEE Transactions on Electron Devices*, vol. 62, pp. 3685–3691, November 2015.

[95] Y. Hirose and H. Hirose, "Polarity-dependent memory switching and behavior of Ag dendrite in Ag-photodoped amorphous As_2S_3 films," *J. Appl. Phys.*, vol. 47, pp. 2767–2772, August 2008.

[96] F. Wang, W. Dunn, M. Jain, C. D. Leo, and N. Vickers, "The effects of active layer thickness on programmable metallization cell based on Ag–Ge–S," *Solid-State Electronics*, vol. 61, pp. 33–37, July 2011.

[97] K. A. Campbell and J. T. Moore, "Silver-selenide/chalcogenide glass stack for resistance variable memory," *U.S. Patent 7, 151, 273*, December 2006.

[98] K. A. Campbell and J. T. Moore, "Resistance variable memory device and method of fabrication," *U.S. Patent 7, 348, 209*, March 2008.

[99] K. A. Campbell, "Method of forming a PCRAM device incorporating a resistance-variable chalcogenide element," *U.S. Patent 7, 354, 793*, April 2008.

[100] A. S. Oblea, A. Timilsina, D. Moore, and K. A. Campbell, "Silver chalcogenide based memristor devices," *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–3, July 2010.

[101] "Neuro-Bit memristor- user manual." `https://wiki.telavivmakers.org/images/0/0f/The_Neuro-Bit_Memristor_user_manual.pdf`.

[102] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for boltzmann machines," *Cognitive Science*, vol. 9, no. 1, pp. 147–169, 1985.

[103] D. H. Goldberg, G. Cauwenberghs, and A. G. Andreou, "Probabilistic synaptic weighting in a reconfigurable network of VLSI integrate-and-fire neurons," *Neural Netw.*, vol. 14, no. 6-7, pp. 781–793, 2001.

[104] F. Corradi, C. Eliasmith, and G. Indiveri, "Mapping arbitrary mathematical functions and dynamical systems to neuromorphic VLSI circuits for spike-based neural computation," *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 269–272, 2014.

[105] P. Wijesinghe, A. Ankit, A. Sengupta, and K. Roy, "An all-memristor deep spiking neural computing system: A step toward realizing the low-power stochastic brain," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 5, pp. 345–358, 2018.

[106] M. Al-Shedivat, R. Naous, G. Cauwenberghs, and K. N. Salama, "Memristors Empower Spiking Neurons With Stochasticity," *IEEE Journal on emerging and selected topics in circuits and systems*, vol. 5, no. 2, pp. 242–253, 2015.

[107] "Switch Bounce and How to Deal with It." https://www.allaboutcircuits.com/technical-articles/switch-bounce-how-to-deal-with-it/.

[108] "Analog to digital converter- ad9200." https://www.analog.com/media/en/technical-documentation/data-sheets/AD9200.pdf.

[109] "Savannah 100, 200 & 300 Atomic Layer Deposition System-Maintenance Manual." https://static1.squarespace.com/static/57b26cc76b8f5b7524bf9ed2/t/57fd1cf237c5817117c38cfe/1476205812733/Savannah_ALD_Maintenance_Manual_Rev_1.1.pdf.

[110] "MPI TS2000-SE." https://www.mpi-corporation.com/ast/engineering-probe-systems/mpi-automated-systems/ts2000-se-probe-system/.

[111] E. Chicca, "Neuromorphic electronic circuits for building autonomous cognitive systems," *Proceedings of the IEEE*, vol. 102, pp. 1367–1388, September 2014.

[112] G. Nagy, D. Arbet, and V. Stopjakova, "Digital methods of offset compensation in 90 nm cmos operational amplifiers," *Design and Diagnostics of Elec-*

tronic Circuits Systems (DDECS), 2013 IEEE 16th International Symposium on, pp. 124–127, July 2013.

[113] K. S. Kundert, "The Designer's Guide to SPICE & SPECTRE," *Kluwer Academic Publishers*, 1995.

[114] S. Tappertzhofen, E. Linn, U. Böttger, R. Waser, and I. Valov, "Nanobattery effect in rrams- implications on device stability and endurance," *IEEE Electron Device Letters.*, vol. 35, no. 2, pp. 208–210, 2014.

[115] I. Valov, E. Linn, S. Tappertzhofen, S. Schmelzer, J. van den Hurk, F. Lentz, and R. Waser, "Nanobatteries in redox-based resistive switches require extension of memristor theory," *Nature Commun.*, vol. 4, pp. 1771–1–1771–9, March 2013.

[116] R. J. Baker, "CMOS- Circuit Design, Layout, and Simulation, 4th Edition," *Wiley*-IEEE P*ress*, July 2019.

[117] P. E. Allen and D. R. Holberg, "CMOS Analog Circuit Design- Second Edition," OXFORD U*niversity* P*ress*, 2002.

[118] A. Sedra and K. Smith, "Microelectronic Circuits," N*ew* Y*ork:* H*olt,* R*inehart, and* W*inston*, 1982.

[119] R. G. Eschauzier, L. P. T. Kerklaan, and J. H. Huising, "A 100-MHz 100-dB Operational Amplifier with MULTIPATH Nested Miller Compensation Structure.," *IEEE J. of Solid-State Circuits*, vol. 27, pp. 1709–1717, December 1992.

[120] F. Maloberti, "Analog Design for CMOS VLSI Systems," *Kluwer Academic Publishers*, 2001.

[121] R. Serrano-Gotarredona, L. Camunas-Mesa, T. Serrano-Gotarredona, J. A. Lenero-Bardallo, and B. Linares-Barranco, "The Stochastic I-Pot: A Circuit Block for Programming Bias Currents," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, pp. 760–764, Sept. 2007.

[122] H.-S. P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen, and M.-J. Tsai, "Metal—Oxide RRAM.," *Proceedings of the IEEE*, vol. 100, pp. 1951–1970, June 2012.

[123] B. Linares-Barranco, T. Serrano-Gotarredona, and R. Serrano-Gotarredona, "Compact low-power calibration mini-DACs for neural arrays with programmable weights.," *IEEE Transactions on Neural Networks*, vol. 14, pp. 1207–1216, Sept. 2003.

[124] "STM32F4DISCOVERY kit." `https://www.st.com/en/evaluation-tools/stm32f4discovery.html`.

[125] "Spartan-3 FPGA Starter Kit Board User Guide." `https://www.xilinx.com/support/documentation/boards_and_kits/ug130.pdf`.

[126] "Spartan-6." `https://www.xilinx.com/support/documentation-navigation/silicon-devices/fpga/spartan-6.html`.

[127] "Node Board." `http://www2.imse-cnm.csic.es/neuromorphs/index.php/AER-Vision-Processors`.

[128] C. Mohan, J. M. de la Rosa, E. Vianello, L. Periniolla, C. Reita, B. Linares-Barranco, and T. Serrano-Gotarredona, "A Current Attenuator for Efficient Crossbars Read-Out," *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019.

[129] C. Zamarreño-Ramoz, L. A. Camuñas-Mesa, J. A. Pérez-Carraso, T. Masquelier, T. S. Gotarredona, and B. Linares-Barranco, "On Spike-Timing-Dependent-Plasticity, Memristive Devices, and Building a Self-Learning Visual Cortex," *Front. Neurosci.*, 17 March 2011.

[130] B. Gilbert, "Current-mode circuits from a Translinear Viewpoint: A Tutorial," *In: Analogue IC design: the current-mode approach*, pp. 11–91, Ed. by C. Tomazou, F. J. Lidgey and D. G. Haigh. Stevenage, Herts., UK: Peregrinus, 2011, Chap. 2.

[131] K. Bult and J. Geelen, "An inherently linear and compact MOST-only current division technique," *IEEE J. Solid-State Circuits*, vol. 27, pp. 1730–1735, Dec. 1992.

[132] J. Lazzaro, S. Ryckebush, M. A. Mahowald, and C. A. Mead, "Winner take-all networks of O(N) complexity," *in Advances in Neural Information Processing Systems 1*, Morgan Kaufmann Publishers, San Francisco, CA, 1989.

[133] M. . Nair, L. K. Muller, and G. Indiveri, "An inherently linear and compact MOST-only current division technique," *Nano Futures*, vol. 1, Nov. 2017.

[134] C. Mohan, L. Camuñas-Mesa, E. Vianello, L. Periniolla, C. Reita, J. M. de la Rosa, T. Serrano-Gotarredona, and B. Linares-Barranco, "Calibration of offset via bulk for low-power HfO_2 based 1T1R memristive crossbar read-out system," *Microelectronic Engineering, Elsevier*, October 2018.

Appendix A

PCB design details and guidelines

Test-PCBs for testing the chip for characterization of 1T1R OxRAM based memristors and calibration scheme— are designed in Allegro PCB editor, whose design details are furnished below.

A.1 Test-PCB for testing 4×4 1T1R crossbar

Fig. A-1 shows the designed test-circuit for testing the 4×4 1T1R crossbar. Fig. A-2 shows the 4-layered PCB designed using Allegro PCB editor. A 3-D view of the designed PCB is shown in fig. A-3. Fig. A-4 shows the PCB dully assembled with PCB components like opamps, switches, decoders, ADC, etc. The chip is packaged in PLCC52 package and is to be placed in the socket (at the right-top corner) of the PCB during testing.

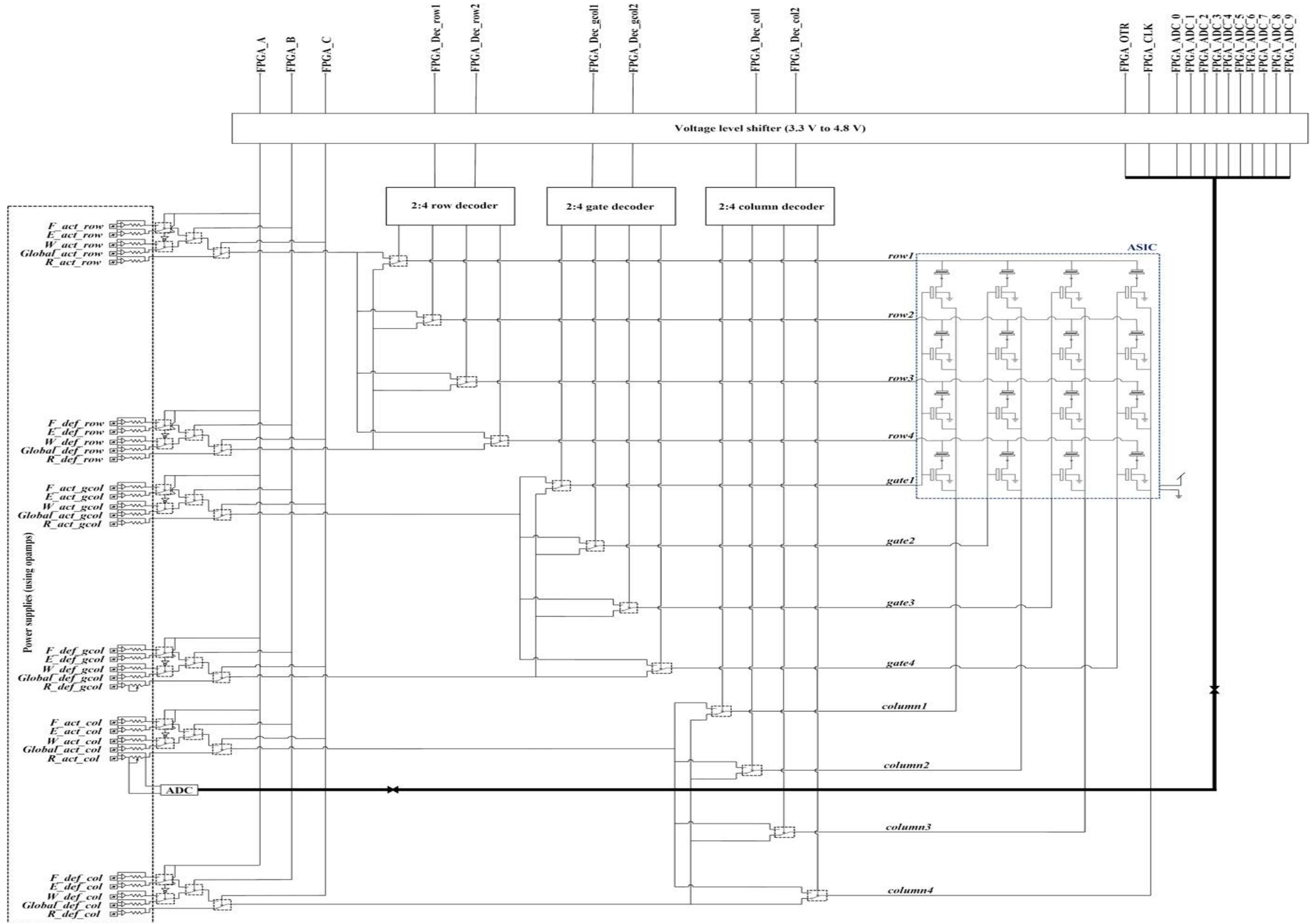


Figure A-1: Schematic view of the test-circuit for testing 4×4 1T1R crossbar along with its ASIC.

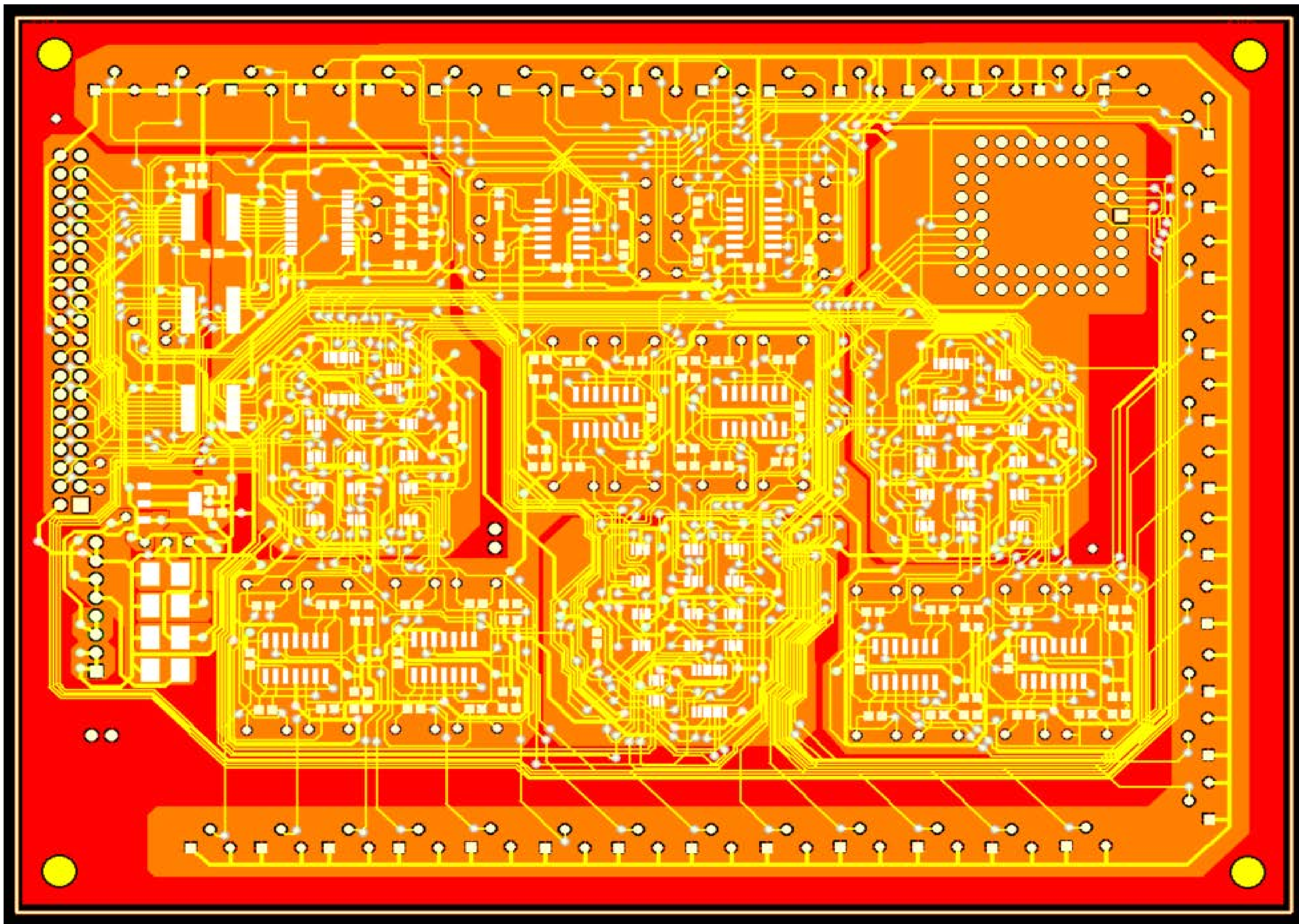


Figure A-2: Layout view of the PCB used for testing 4×4 1T1R crossbar.

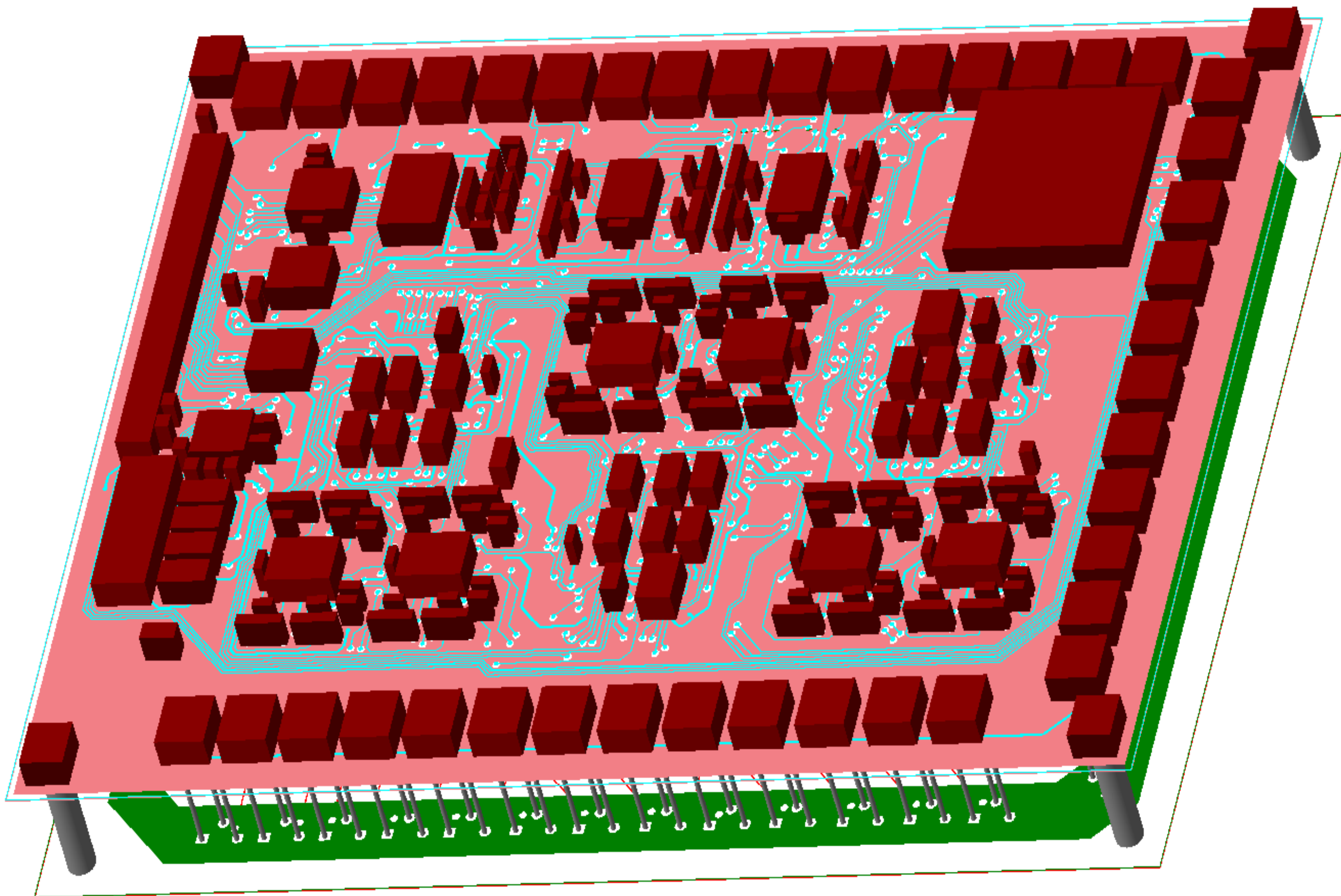


Figure A-3: 3-D view of the designed PCB for testing 4×4 1T1R crossbar.

Figure A-4: Assembled and mounted PCB for testing 4×4 1T1R crossbar.

A.2 Test-PCB for testing different circuits in the Outer-ring

Outer-ring of the MAD200 chip comprises an individual 8×8 1T1R crossbar, a three-stage DC-offset calibration scheme implemented on a 4×4 1T1R crossbar and a two-stage opamp. Fig. A-5 shows the designed test-circuit for testing the 8×8 1T1R crossbar. Fig. A-6 shows the designed test-circuit for testing the three-stage calibration scheme implemented in 4×4 1T1R crossbar and fig. A-7 shows the designed test-circuit for testing opamp. The layout view of the PCB used for testing different circuits in the outer-ring is shown in fig. A-8 and fig. A-9 shows the 3-D view of the designed PCB for testing different circuits in the outer-ring. Fig. A-10 shows the PCB duly assembled with PCB components like opamps, switches, decoders, ADC, etc. The chip is packaged in the PGA100 package and is to be placed in the PGA socket of the PCB during testing. Fig. A-11 (a) shows the top view of the front-side of the packaged chip dully labeled with pin numbers. Fig. A-11 (b) shows the top view of the rear-side (mirrored) of the packaged chip dully labeled with pin numbers or addresses. Fig. A-11 (c) shows the top view of the front-side of the PGA ZIF 14 × 14 socket dully labeled with pin numbers or addresses. The pin location that the packaged chip uses to sit on the ZIF socket is highlighted in A-11 (c). Table A.1 shows the addresses or locations of the signals of the Outer-ring on packaged chip (for both front and rear side views) and for PGA ZIF socket.

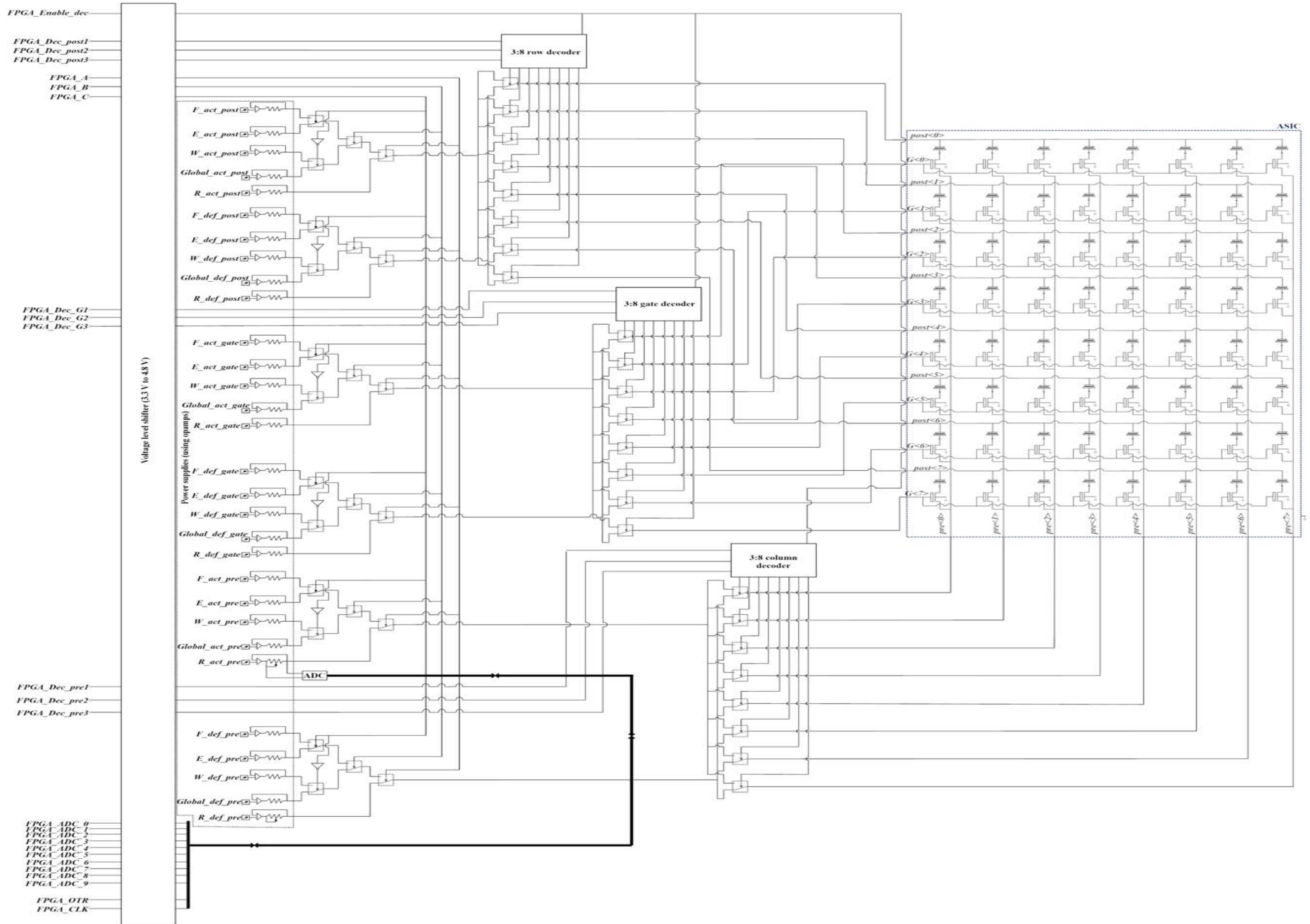


Figure A-5: Schematic view of the test-circuit for testing 8×8 1T1R crossbar along with its ASIC.

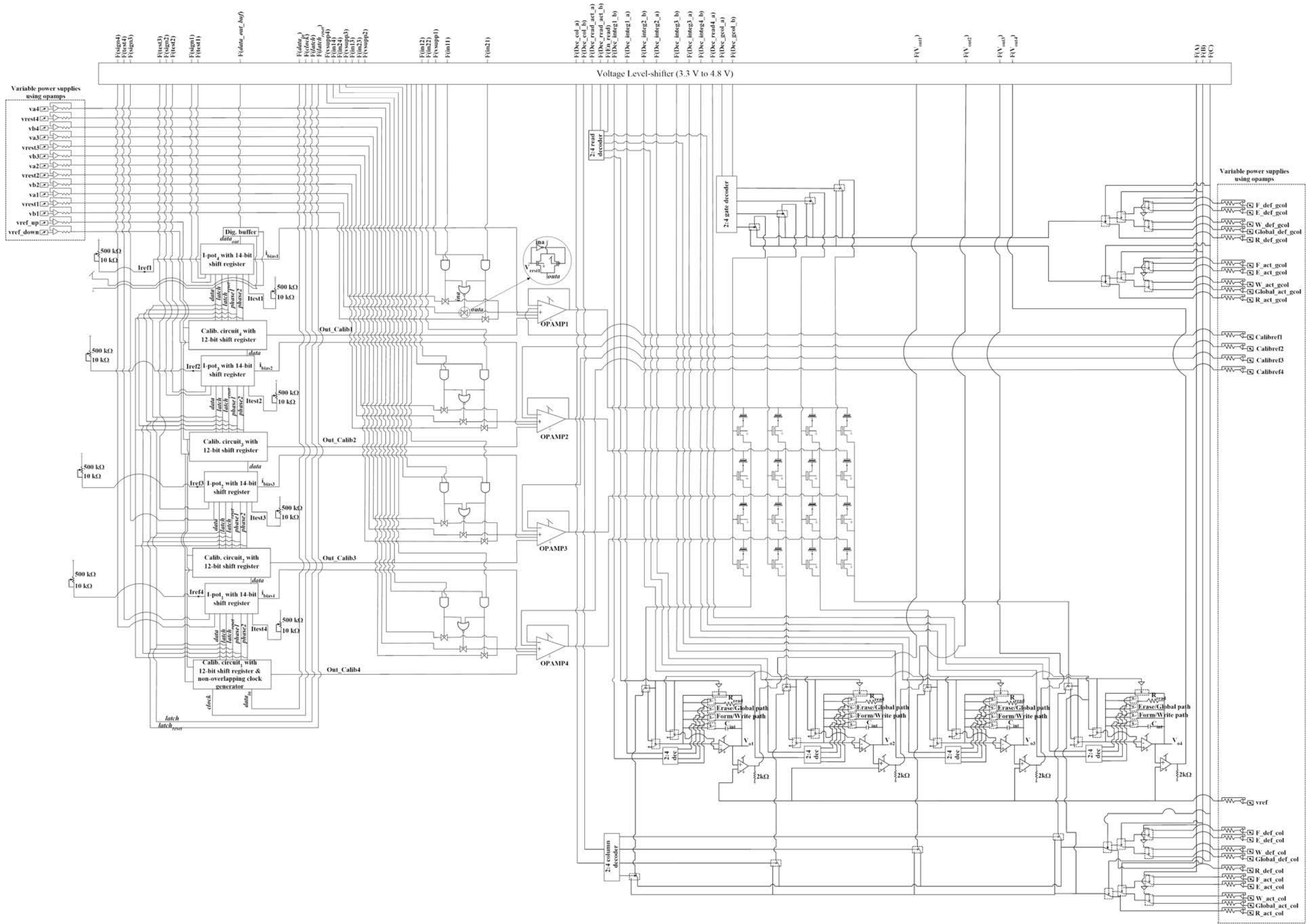


Figure A-6: Schematic view of the test-circuit for testing calibration scheme implemented in 4×4 1T1R crossbar.

Figure A-7: Schematic view of the test-circuit for testing opamp.

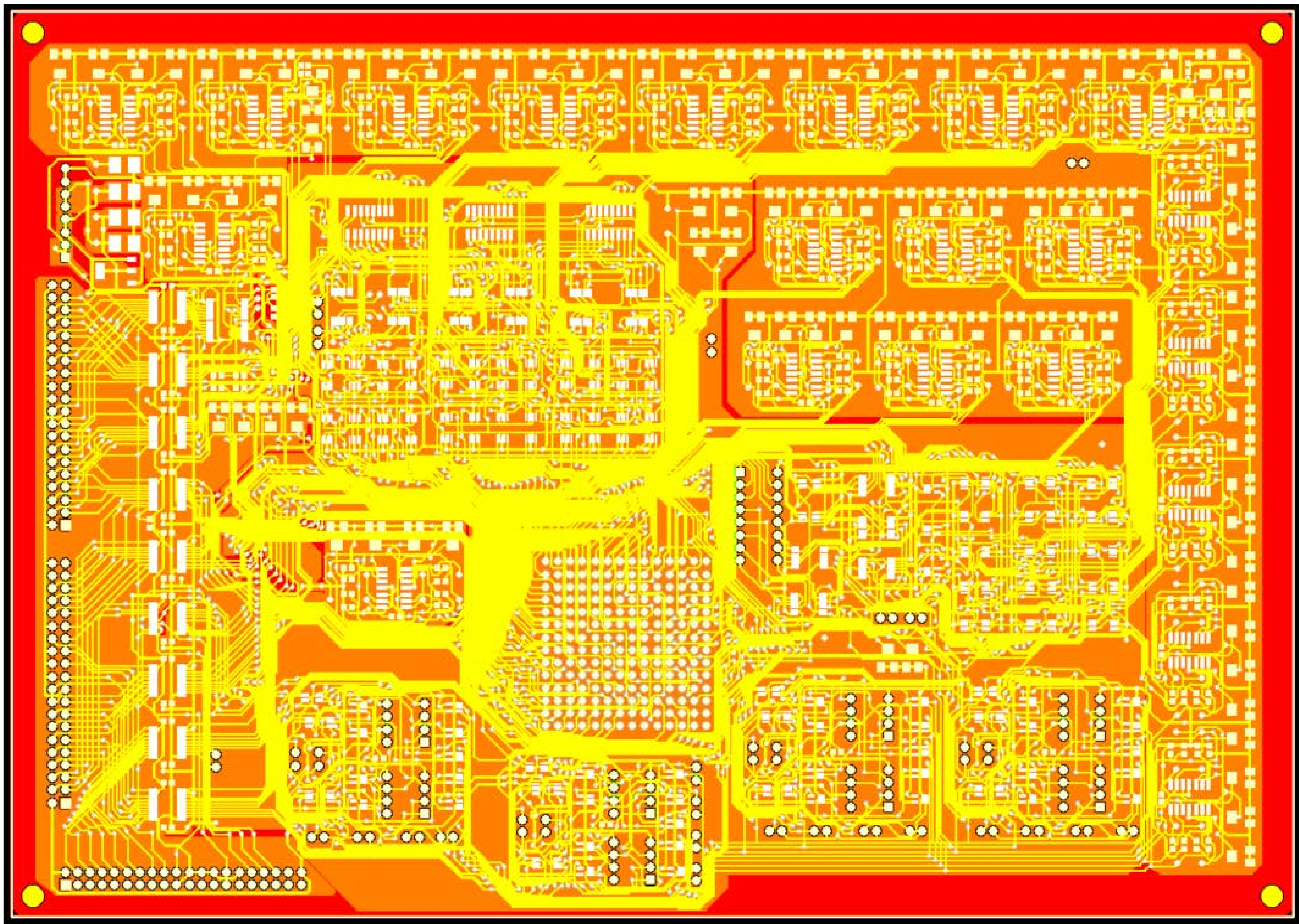


Figure A-8: Layout view of the PCB used for testing different circuits in the outer-ring.

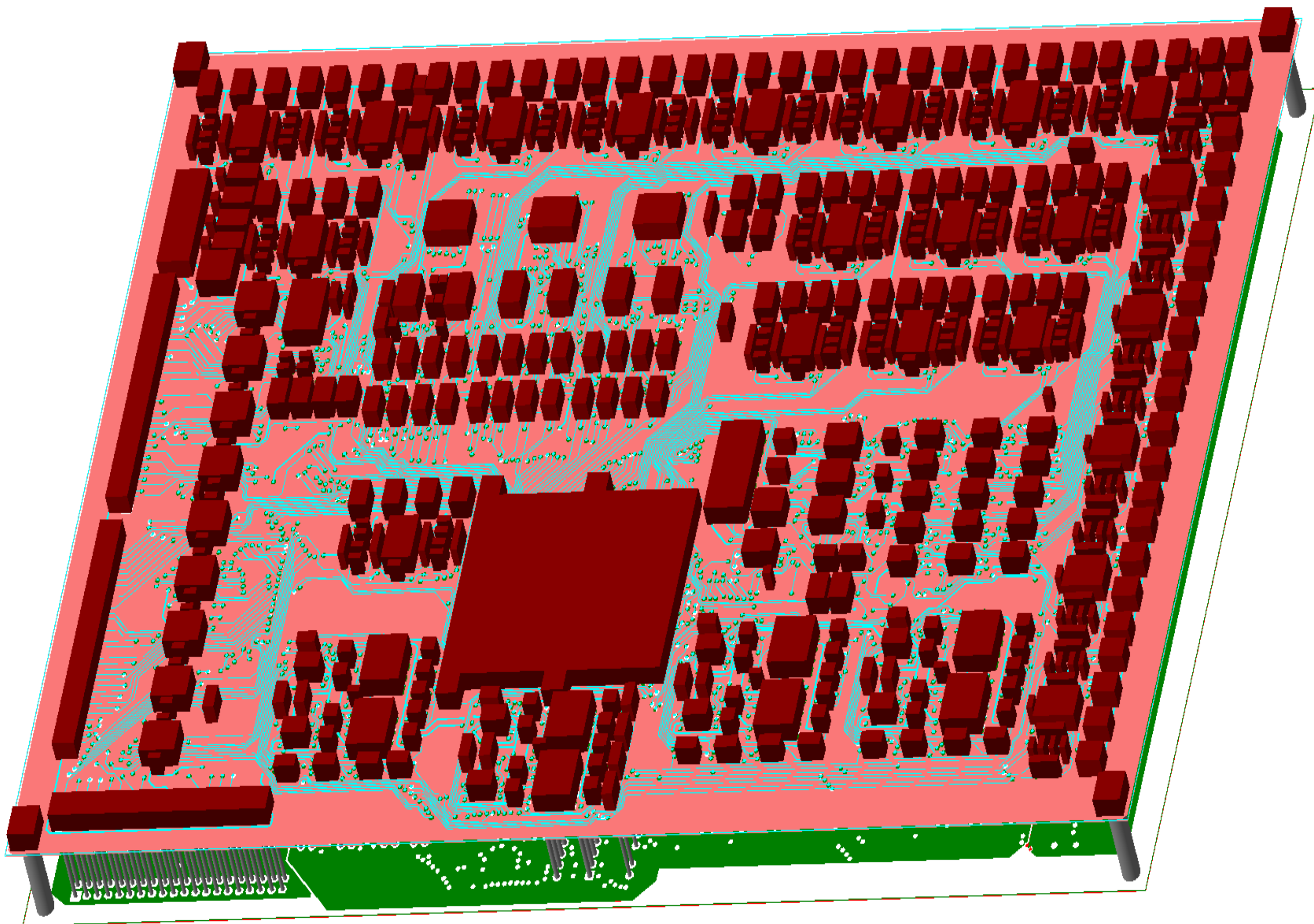


Figure A-9: 3-D view of the designed PCB for testing different circuits in the outer-ring.

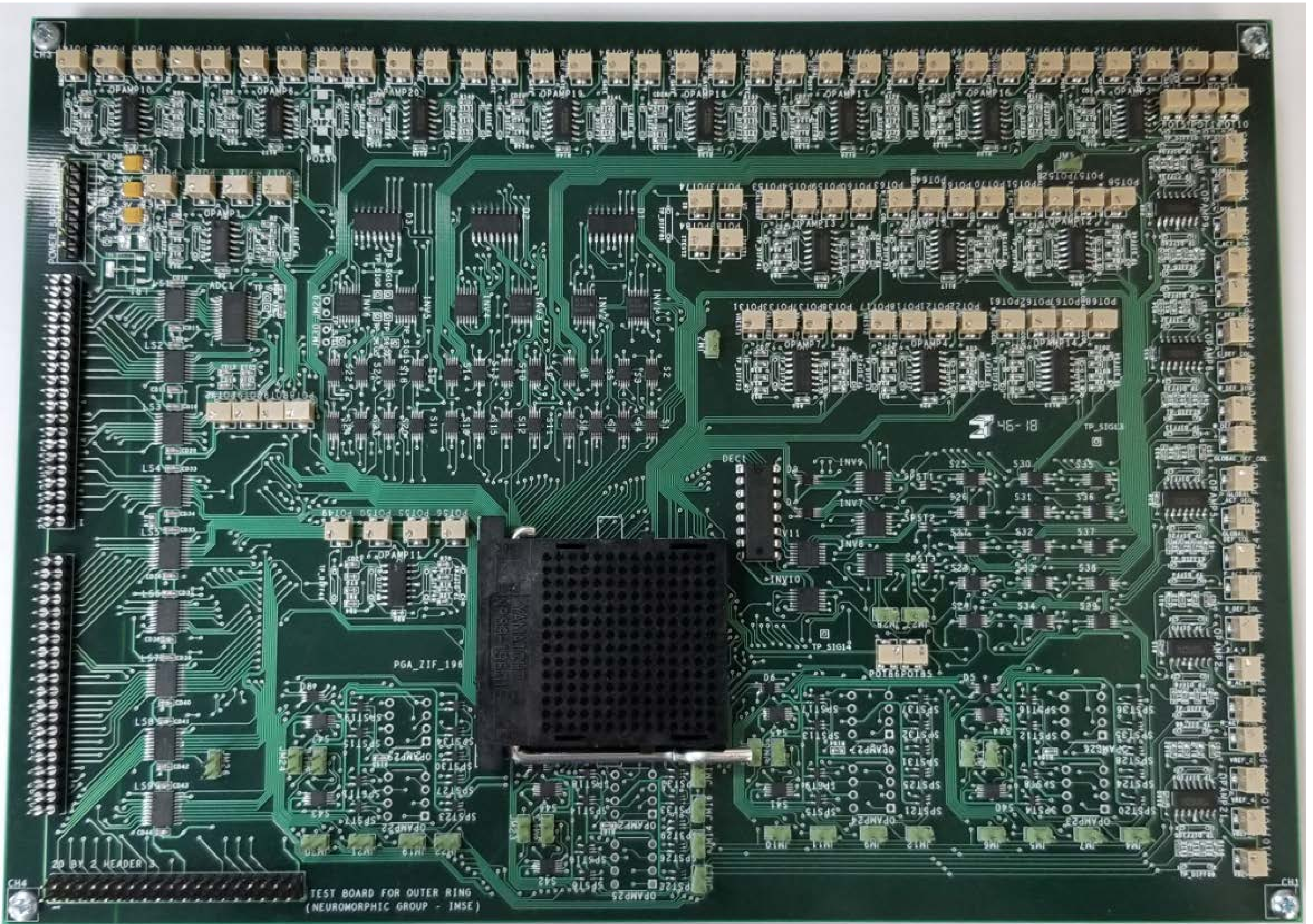


Figure A-10: Assembled and mounted PCB for testing different circuits in the outer-ring.

Table A.1: Pin addresses for packaged chip and PGA socket for different signals.

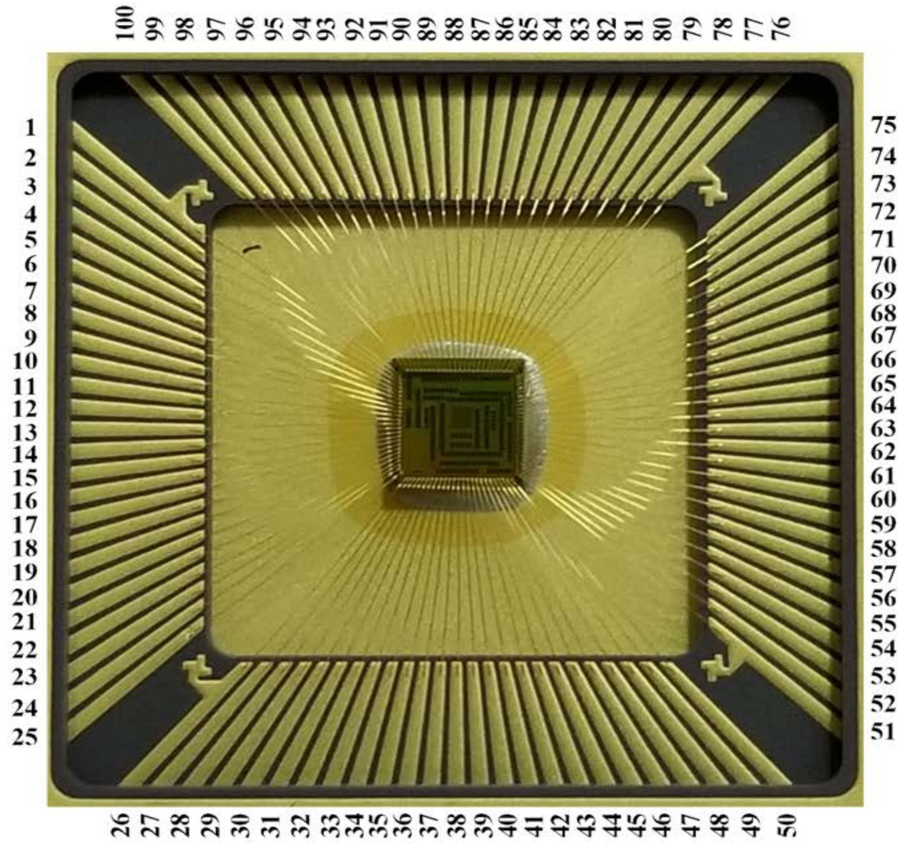
Sl.No.	**Signal name**	**Address or location**		
		On top view of packaged chip		**On PGA ZIF socket**
		Front-side	**Rear-side**	
1	Iref2	1	B2	C13
2	vsupp2	2	B1	B13
3	in12	3	C2	C12
4	in22	4	C1	B12
5	sign3	5	D2	C11
6	Iref3	6	D1	B11
7	vsupp3	7	E2	C10
8	in13	8	E1	B10
9	in23	9	F3	D9
10	sign4	10	F2	C9
11	Iref4	11	F1	B9
12	vsupp4	12	G2	C8
13	in14	13	G3	D8
14	in24	14	G1	B8
15	$clock$	15	H1	B7
16	$data_{in}$	16	H2	C7
17	$latch_{reset}$	17	H3	D7
18	$latch$	18	I1	B6
19	Calibref4	19	I2	C6
20	vb4	20	J1	B5
21	vrest4	21	J2	C5
22	va4	22	K1	B4
23	gnd	23	L1	B3
24	$data_out_buf$	24	K2	C4
25	vdd_buf	25	M1	B2

Continuation of Table A.1				
Sl.No.	Signal name	Address or location		
		On top view of packaged chip		On PGA ZIF socket
		Front-side	Rear-side	
26	vdd	26	L2	C3
27	row4	27	M2	C2
28	itest4	28	L3	D3
29	test4	29	M3	D2
30	Calibref3	30	L4	E3
31	vb3	31	M4	E2
32	vrest3	32	L5	F3
33	va3	33	M5	F2
34	row3	34	K6	G4
35	itest3	35	L6	G3
36	test3	36	M6	G2
37	col1	37	L7	H3
38	gcol1	38	K7	H4
39	col2	39	M7	H2
40	gcol2	40	M8	I2
41	col3	41	L8	I3
42	gcol3	42	K8	I4
43	col4	43	M9	J2
44	gcol4	44	L9	J3
45	Calibre2	45	M10	K2
46	vb2	46	L10	K3
47	vrest2	47	M11	L2
48	va2	48	M12	M2
49	row2	49	L11	L3
50	itest2	50	M13	N2

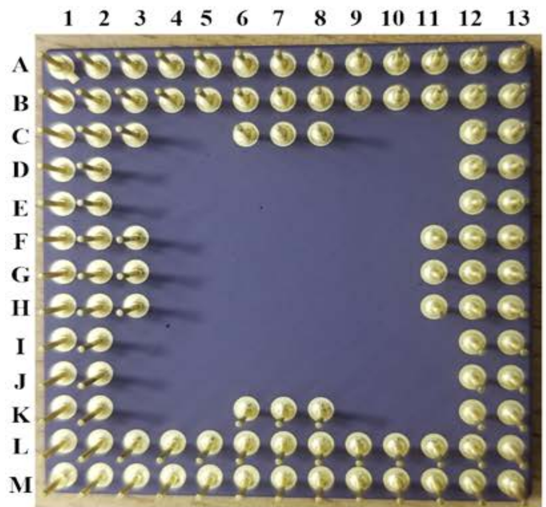
Continuation of Table A.1

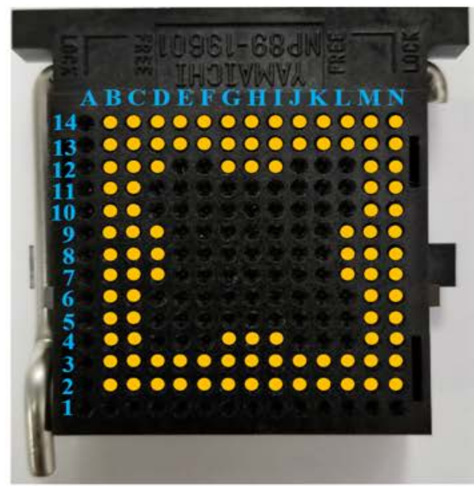
Sl.No.	Signal name	Address or location		
		On top view of packaged chip		On PGA ZIF socket
		Front-side	Rear-side	
51	test2	51	L12	M3
52	Calibref1	52	L13	N3
53	vb1	53	K12	M4
54	vrest1	54	K13	N4
55	va1	55	J12	M5
56	row1	56	J13	N5
57	Itest1	57	I12	M6
58	test1	58	I13	N6
59	—	59	H11	L7
60	—	60	H12	M7
61	—	61	H13	N7
62	ib	62	G12	M8
63	calib	63	G11	L8
64	in_p	64	G13	N8
65	calibref	65	F13	N9
66	in_n	66	F12	M9
67	vo	67	F11	L9
68	pre<7>	68	E13	N10
69	pre<6>	69	E12	M10
70	pre<5>	70	D13	N11
71	pre<4>	71	D12	M11
72	pre<3>	72	C13	N12
73	pre<2>	73	B13	N13
74	pre<1>	74	C12	M12
75	pre<0>	75	A13	N14

Continuation of Table A.1				
Sl.No.	Signal name	Address or location		
		On top view of packaged chip		On PGA ZIF socket
		Front-side	Rear-side	
76	gnd	76	B12	M13
77	post<0>	77	A12	M14
78	post<1>	78	B11	L13
79	post<2>	79	A11	L14
80	post<3>	80	B10	K13
81	post<4>	81	A10	K14
82	post<5>	82	B9	J13
83	post<6>	83	A9	J14
84	post<7>	84	C8	I12
85	G<7>	85	B8	I13
86	G<6>	86	A8	I14
87	G<5>	87	B7	H13
88	G<4>	88	C7	H12
89	G<3>	89	A7	H14
90	G<2>	90	A6	G14
91	G<1>	91	B6	G13
92	G<0>	92	C6	G12
93	sign1	93	A5	F14
94	Iref1	94	B5	F13
95	vsupp1	95	A4	E14
96	in11	96	B4	E13
97	in21	97	A3	D14
98	vref_down	98	A2	C14
99	vref_up	99	B3	D13
100	sign2	100	A1	B14



(a)





(b) (c)

Figure A-11: (a) Top view of the front-side of the packaged chip dully labelled with pin numbers, (b) Top view of the rear-side (mirrored) of the packaged chip dully labelled with pin numbers or addresses, (c) Top view of the front-side of the PGA ZIF 14 $\times$ 14 socket dully labelled with pin numbers or addresses and marked with the location where the packaged chip sits on it.

A.3 Guidelines for a better PCB assembling and mounting practice

Soldering PCB includes initial practice sessions followed by few guidelines, safety precautions, and thumb-rules. Some of them are discussed below-

A better practice is to start soldering the SMD components first and then move to solder the through-hole components on the PCB. Also, within SMD components, it is better to start soldering the components whose pitch[1] distance is less.

Using optimal soldering temperature, the optimal amount of soldering flux, optimal amount of soldering tin, and optimal inclination of soldering iron tip during soldering for different pitch distances— are always the mandatory guidelines to be followed and this depends on practicing soldering of components for different packages. The same also applies for desoldering them. Soldering iron with a narrow tip is used for soldering components with small pitch distance and the one with a large tip is used for soldering components that have large pitch distance.

Cleaning soldering tip using a wet sponge and tinning the soldering iron tip improves conductivity. It also makes soldering easier and quicker.

It is advisable to keep the solder smoke absorber always ON throughout the soldering session, as prolonged exposure to soldering smoke, is injurious.

It is also a must to wear an anti-static wrist strap to prevent damage to active components during soldering, which is highly sensitive to static charges accumulated.

During PCB design, placing a via on the pad for passive components is not recommended in automated PCB assemblers, as the flux can flow-out through the via. But doing so can be beneficial in manual-soldering, as it facilitates a back-up plan to establish a connection through via when the pad is destroyed or removed accidentally.

After the soldering, the PCB is drenched in an ultrasonic bath for removal of flux, which is then dried using a high-pressure air blower. The air blower is thoroughly used all over the PCB on both sides to make sure that no water droplets are present on PCB to prevent short-circuits during testing.

[1] Pitch is the center-to-center spacing between conductors, such as pads and pins on a PCB.

Appendix B

MADII circuit design details

B.1 Specifications of opamp

Table B.1 shows the specifications of the two-stage opamp for four different load conditions, considering the resistive states of the OxRAM.

Table B.1: Design specifications of the two-stage opamp.

Parameter	**With C load (C = 5 pF)**	**With RC load (C = 5 pF)**		
		With R = 2 kΩ	**With R = 7 kΩ**	**With R = 225 kΩ**
Gain (A_v) in dB	100.96	69.33	79.13	98.47
GBW in MHz	15.12	12.8	14.38	15.1
Phase Margin (PM) in °	59.93	66.23	62	60
ICMR+ in V	3			
ICMR- in V	0.7			
Slew Rate (SR) in V/µs	15.07	13	14.3	15.04
Input bias current in µA	40.2			
2nd stage drain current in µA	634.48	623.8	624.37	631.25
Power dissipation (P_{diss}) in mW	3.29	3.2	3.2	3.22
DC systematic offset in µV	89.6			
DC offset voltage variation (mismatch- 300 runs)	µ= -34.8 µV	-	µ= -315.25 µV	µ= -41.951 µV
	σ = 1.081 mV	-	σ = 1.0806 mV	σ = 1.081 mV
Total input referred noise in V^2	8.34e-11			