NEUROTEC I: Neuro-inspired Artificial Intelligence Technologies for the Electronics of the Future

Melvin Galicia[1], Stephan Menzel[8], Farhad Merchant[1], Maximilian Müller[2], Hsin-Yu Chen[8], Qing-Tai Zhao[9], Felix Cüppers[10], Abdur R. Jalil[10], Qi Shu[10], Peter Schüffelgen[10], Gregor Mussler[9], Carsten Funck[3], Christian Lanius[5], Stefan Wiefels[8], Moritz von Witzleben[3], Christopher Bengel[3], Nils Kopperberg[3], Tobias Ziegler[3], R. Walied Ahmad[8], Alexander Krüger[10], Leticia Pöhls[5], Regina Dittmann[8], Susanne Hoffmann-Eifert[10], Vikas Rana[10], Detlev Grützmacher[9,10], Matthias Wuttig[2,10], Dirk Wouters[3], Andrei Vescan[4], Tobias Gemmeke[5], Joachim Knoch[6], Max Lemme[7], Rainer Leupers[1] and Rainer Waser[3,8,10]

[1]Institute for Communication Technologies and Embedded Systems,[2]Institute of Physics, Physics of Novel Materials, [3]Institute of Materials in Electrical Engineering II, [4]Compound Semiconductor Technology, [5]Institute of Integrated Digital Systems and Circuit Design, [6]Institute of Semiconductor Electronics, [7]Chair of Electronic Devices, RWTH Aachen University, Aachen, Germany, [8]Peter-Grünberg Institute-7, [9]Peter-Grünberg Institute-9, [10]Peter-Grünberg Institute-10, Forschungszentrum Jülich GmbH and JARA-Fit, Germany

***Abstract*—The field of neuromorphic computing is approaching an era of rapid adoption driven by the urgent need of a substitute for the von Neumann computing architecture. NEUROTEC I: "Neuro-inspired Artificial Intelligence Technologies for the Electronics of the Future" project is an initiative sponsored by the German Federal Ministry of Education and Research (BMBF for its initials in German), that aims to effectively advance the foundations for the utilization and exploitation of neuromorphic computing. NEUROTEC I stands at its successful "final stage" driven by the collaboration from more than 8 institutes from the Jülich Research Center and the RWTH Aachen University, as well as collaboration from several high-tech industry partners. The NEUROTEC I project considers the field interplay among materials, circuits, design and simulation tools. This paper provides an overview of the project's overall structure and discusses the scientific achievements of its individual activities.**

I. Introduction

Pattern recognition is a task that requires little time and energy when performed by the human brain, however this same task requires several orders of magnitude more time and energy when performed by a top-notch computer that uses the von Neumann architecture [1]. This is just one of many tasks where the brain outperforms current computers. Neuromorphic computing aims to replicate the calculation process that takes place in the brain by developing artificial synapses or "computing-in-memory" (CIM) concepts for artificial neural networks. Hence neuromorphic computers have the potential to become the computer generation of the future. To accelerate the achievement of such an ambitious goal, NEUROTEC I brings together an interdisciplinary and complementary stack of research disciplines as illustrated in Fig. 1. Here, various memristive materials and circuit configurations are investigated. Several integration methodologies for synaptic networks are developed and simulation models and tools have been created in order to implement, test and validate characterization methods for memristive circuits. The project focuses on a wide-ranging investigation of different memristive materials at different technology levels. Thus, it aims to develop novel components for new computer architecture which have the potential for fast and energy-efficient computations that can be extensively exploited

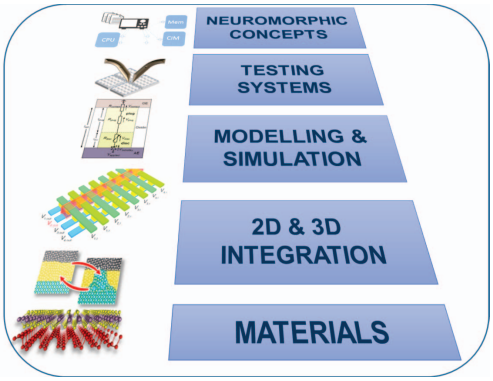


Fig. 1. Scientific and technological research fields in NEUROTEC I.

in applications of artificial intelligence (AI). In the following sections, the main scientific and technological research activities and their outcomes are described in more detail.

II. New materials and structures for memristors

A. Optimization of nanoscale switching VCM systems

Nanoscale memristive devices based on filamentary valence change mechanism (VCM) are widely researched candidates for emerging memories and neuromorphic applications. Co-integration with complementary metal oxide semiconductor (CMOS) chips is one of the important requirements on the redox-based resistive random access memories (ReRAM). We developed and characterized nanoscale memristors constituted from asymmetric metal electrodes that sandwich a thin layer of transition metal oxide grown by atomic layer deposition. The main structure consists of 3 nm HfO_2 and Pt and Ti electrodes. Previous works suggested inclusion of an intentionally grown off-stoichiometric TiO_{2-x} interface layer in the $Ti/TiO_{2-x}/HfO_2/Pt$ stack, Fig. 2(b) [2]. Studying the switching response to fast voltage pulses of the fabricated $(100\ nm)^2$ crossbar devices, shown in Fig. 2(a), revealed that gradual and abrupt binary switching coexist. The binary mode is characterized by significant switching stochasticity associated with the inherent randomness of ionic motion, filament geometry as well

978-3-9819263-6-1/DATE22/©2022 EDAA

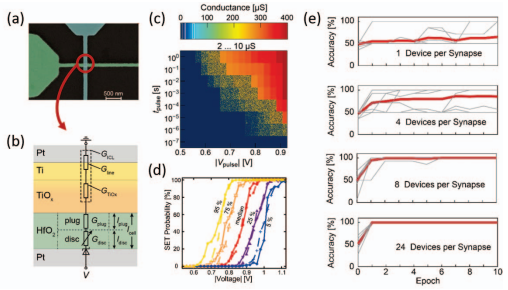


Fig. 2. (a) Top view of a $(100\ \text{nm})^2$ crossbar. (b) Stack order and eqivalent circuit. (c) SET kinetics conductance map of the binary mode. (d) Experimental and simulated SET probability statistics for a pulse duration of 1 µs. (e) Spiking Neural Network results.

as vacancy concentration [3], [4]. By combining the statistical data of a SET probability measurement ensemble with the JART VCM v1b, Fig. 2(d) [5], we were able to study the concept of parallel variable, binary devices as fundamental building block for synapses in neuromorphic architectures. By implementing this block into a Spiking Neural Network simulation, it was shown that the network's performance increases when the synapses comprise a higher number of devices, due to a larger window of programming voltage as well as a larger number of accessible intermediate resistance states [6].

B. Optimization of interface switching PCM systems

Phase-change materials (PCMs) have attracted considerable interest for neuromorphic computing. Yet, at present there are open issues regarding the speed limits of the switching process. For this reason, we have investigated clear design rules for application relevant properties such as the minimum time to crystallize. Establishing such rules would be facilitated by a fundamental understanding of the relationship between the achievable switching speed and the stoichiometry as well as the chemical bonding. For this purpose, a quantum-chemical map of bonding is utilized in our studies [7]. This map reveals that two characteristic quantities, the number of electrons transferred (ET) and the number of electrons shared (ES) between adjacent atoms are good quantum-chemical bonding descriptors, enabling a separation of ionic, covalent, metallic and metavalent bonding [8]–[10]. These two quantities are also good property predictors, as shown in Fig. 3, where trends for the minimum crystallization time as a function of material stoichiometry, described by ES and ET. Going from GeSe to GeTe and subsequently towards SnTe, a reduction of the minimum time for crystallization by 6 orders of magnitude has been achieved. At the same time, Fig. 3(b) shows the stability

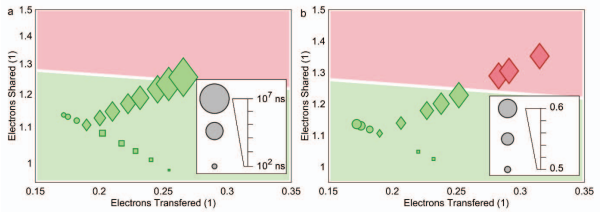


Fig. 3. Change of switching kinetics with chemical bond indicators. (a) Minimum crystallization time and (b) reduced onset temperature of the glass transition as a function of two quantum-chemical bonding indicators.

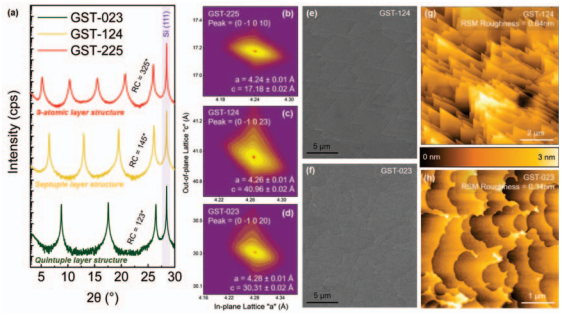


Fig. 4. Structural characterization of GST alloys via (a) XRD $2\theta/\theta$ scans (b-d) reciprocal space maps of GST-225, 124, 023 phases respectively. (e-f) Scanning electron micrographs and (g-h) atomic force microscopy images confirm the entirely coalescent epilayer with ultra-smooth surfaces of various GST alloys.

of the amorphous phase decreases too, as characterized by the reduced onset temperature of the glass transition [11].

PCMs exhibit at least two distinctive structural and electronic states i.e. the crystalline (low resistive) and the amorphous (high resistive) that provide a platform for the basic logic states "1" and "0". Among various PCMs including Sb_xTi_{1-x} and Ge_xSb_{1-x} alloys, the $Ge_xSb_yTe_z$ (GST-xyz) stoichiometric alloys are the most promising candidates for the realization of neuromorphic and non-volatile memory applications. However, the realization of a GST-based memory cell requires both, high structural quality epitaxial growth of GST crystals and their incorporation into CMOS compatible fabrication technology. Using the technique of Metal-Organic Vapor Phase Epitaxy, we achieved strain-free and high crystal quality epilayers of several GST alloys on Si(111) substrates. XRD scans, depicted in Fig. 4(a), indicate the high crystal quality of of GST epilayers. To evaluate the lattice parameters, the reciprocal space maps are acquired at the corresponding asymmetric peaks shown in Fig. 4(b-d) and no strain related defects are observed. In the second stage, the growth parameters are systematically modified to improve the surface quality. Eventually, the entirely coalescent epilayers are achieved and can be observed via micrographs in Fig. 4(e-f) and topographical images in Fig. 4(g-h). In future, produced epilayers will be subjected to the pre-patterned substrates to achieve selective epitaxy in the nanostructures to incorporate GST alloys into various quantum devices. [12]–[15].

C. Stack layers from 2D materials

To bring all the fascinating properties of two-dimensional (2D) transition metal dichalcogenides into relevant applications, developing scalable synthesis methods is of great importance. In this realm, metal-organic chemical vapor deposition (MOCVD) has been found to be a promising candidate. Using state of the art MOCVD reactors, and characterization techniques, we have paved the way for the realization of (opto)electronic devices based on 2D materials by providing high-quality 2D layers and heterostructures. For this purpose, optimized processes to obtain coalesced 2D-WS_2 and 2D-MoS_2 films with high homogeneity and minimized carbon contamination have been developed [16], [17]. The layers have been used for the demonstration of flexible photodetectors [18]. The AIXTRON close coupled showerhead reactor was used to successfully perform experiments on the upscaling

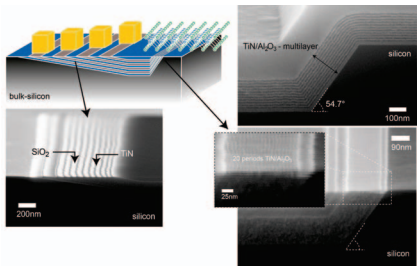


Fig. 5. A buried multi-electrode (BMEs) platform.

of the deposition process to 100 mm diameter substrates. Moreover, extensive effort has been expended to develop 2D heterostructures in a seamless one-step growth approach [19].

D. Buried multi-electrode devices for 2D materials

Local properties of materials are usually tested by restricting the investigated volume by nanoscale patterning and/or using appropriate contact/gate-electrodes. However, etching a nanostructure likely leads to issues related to surface defects and electrodes are either too large or too inflexible to facilitate a reliable local characterization. To study the memristive properties of 2D materials and transistion metal oxides locally, we have developed a buried multi-electrode (BMEs) platform that allows the fabrication of approximately 20 individually addressable (gate-)electrodes with a lateral length down to 5 nm and with 5 nm interdistance between adjacent electrodes, see Fig. 5. BMEs are manufactured by etching a mold into a silicon substrate with a steep and on a shallow etch flank. While the steep flank is created with wet chemical, ansisotropic silicon etching, the shallow flank is realized with controlled resist erosion in a SF_6/O_2 plasma. The BME substrates are currently used to study the local properties of two-dimensional materials. Furthermore, the local properties of memristive transition metal oxides are studied by depositing the memristive material and a top electrode and then using the BMEs as nanoscale contact electrodes.

III. Integration for 2D and 3D Synaptic Networks

A. Integration of TaO_x/Ta devices with FDSOI MOSFET

To eliminate the sneak path current in the ReRAM crossbar architecture, an 1T-1R integration platform on fully depleted silicon-on-insulator (SOI) substrate is developed. This architecture enables neuromorphic computing applications based on TaO_x/Ta memristive crossbar array. Fig. 6 shows the process flow, illustrative cross-section of the 1T-1R device. For the MOSFET device fabrication, a gate-first process with self-aligned NiSi S/D is developed on the 30 nm-thick SOI substrate. Subsequently, the TaO_x (10 nm)/Ta(15 nm) ReRAM is integrated in the backend of the SOI MOSFET as shown in Fig. 6. The gate length of the SOI MOSFET was 2 µm, whereas the ReRAM cell area was 2x2 μm^2. Fig. 6 shows the electroforming process and bipolar switching of TaO_x/Ta devices under different gate voltages (V_{gs}= 1.0 V – 2.0 V) in the 1T-1R configuration.

B. Integration concepts for multi-dimensional networks

Electrochemical metallization (ECM) materials can mimic synaptic functionality by creating conductive filament with adjustable strength in-between two metallic electrodes (consisting e.g. of Pt and Cu) as illustrated in Fig. 7(a). 2D memristive crossbar arrays allow only a limited number of independent interconnects to be realized and applying appropriate voltages to the bar-shaped electrodes may result in parasitic filament growth or it may modify the synaptic weight of existing interconnects. Thus, we currently explore shrinking the crossed bars to electrodes with a T-cross-section (see Fig. 7(b)) positioned alternating on a regular square/cubic grid. This allows independent interconnect formation of each electrode with its nearest neighbors, strongly increasing the connectivity of such a synaptic network. A central ingredient for the realization of such a synaptic network is the fabrication of appropriate lateral ECM cells with a true nanoscale interelectrode distance in lateral direction. To this end, we developed a combined spacer and damascene process which is schematically shown in Fig. 7(c): after the fabrication of a first Pt electrode, Cu-doped SiO_2 as ECM material is deposited conformally with a suitable sputter process. Subsequently, Cu is deposited followed by chemical-mechanical polishing. Since the lateral interdistance of the electrodes is determined by a deposition and not a lithography process, sub-10 nm separations can in principle be realized. First lateral ECM cells show memristive switching behavior featuring a sub-30 nm electrode interdistance, see the electron micrograph in Fig. 7(c).

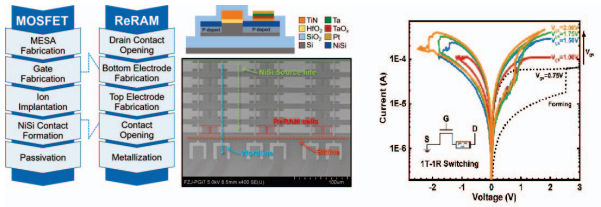


Fig. 6. Schematic diagram of the fabrication steps for the FD-SOI MOSFET and TaO_x/Ta ReRAM integration. SEM image of the 16x16 1T-1R crossbar array. *I-V* characteristics (electroforming and switching) of the TaO_x/Ta ReRAM in the 1T-1R configuration, where MOSFET serves as current-limiter.

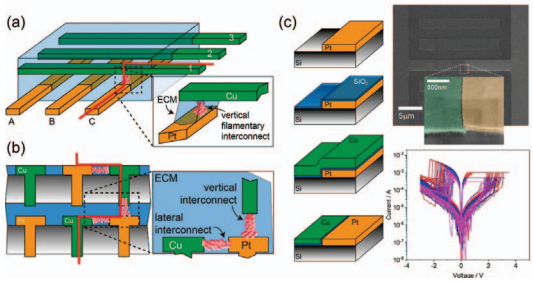


Fig. 7. (a) Memristive crossbar with vertical ECM cells (b) T-junction array for higher connectivtiy of synaptic networks (c) Fabrication scheme for lateral ECM cells and *I-V* characteristic of a Pt/SiO_2/Cu cell.

IV. Modeling, Design and Simulation Tools

A. Physics-based modeling

To enable proper circuit design using memristors, reliable compact models for this emerging type of devices need to be developed. This comprises the modeling of the deterministic switching behavior as well as modelling of reliability aspects. Within the project we developed compact models including variability for filamentary switching VCM devices

[5]. The model reproduces the switching dynamics of the Pt/HfO_2/TiO_x/Ti, Pt/TaO_x/Ta and Pt/ZrO_x/Ta devices developed in this project. It captures the D2D and C2C variability [4], [20] including the switching probability at certain voltages according to Fig. 2. Moreover, read instability and shaping failures can be well reproduced [21]. To understand the trade-off between retention behavior (Fig. 9b) and read instability, a kinetic Monte Carlo Model was developed [22]. High jump rate of the oxygen vacancies within the domain accounts for the read-instability, while the lower jump rate from domain to domain determine the retention. Using density functional non-equilibrium Green's function calculations, we could reveal two major types of conduction in VCM cells [23]. In both cases, the electron transport is dominated (at least at lower voltages) by the tunneling to one of the metal/oxide interfaces, but in one case the electrons tunnel into the conduction band of the oxide, and in the other case they tunnel into defect states induced by nearby oxygen vacancies.

B. Neuromorphic accelerator system level simulation

We developed a high level virtual platform "NeuroVP" integrating CIM accelerators [24]. Using NeuroVP to evaluate performance and power consumption at the electronic system level (ESL), it has been corroborated that the execution of neural network applications, mainly vector-matrix multiplications (VMM) operations, using neuromorphic accelerators yields higher power efficiency speedup relative to a von Neumann computing system as shown in Fig. 8. The power consumption of CMOS devices is a combination of leakage and dynamic power. Leakage power can be assumed as constant as long as temperature and supply voltage changes are not taken into consideration. Dynamic power, on the other hand, depends on short circuit and switching power caused by the total activity performed by the device. The dynamic power depends on control signals which initiate actions or instructions performed by the device. Hence, to estimate the dynamic power consumption at ESL, it is sufficient to trace the control signals.

V. Development of Test Systems

A. Sub-nanosecond electrical characterization

To test whether memristors, developed in section II, can also operate at GHz frequencies, their switching kinetics were studied in the sub-nanosecond range. For this purpose, an automated radio frequency setup was built, with which the change in resistance can be measured on a 10 ps-timescale [25]. It could be shown that VCM devices can switch within 50 ps from the HRS to LRS, which is, to our knowledge, the

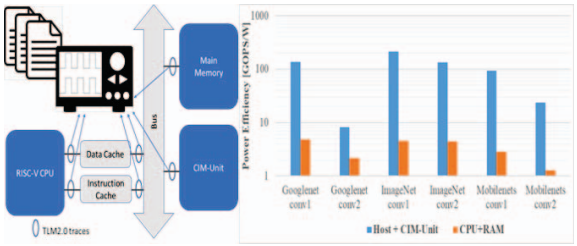
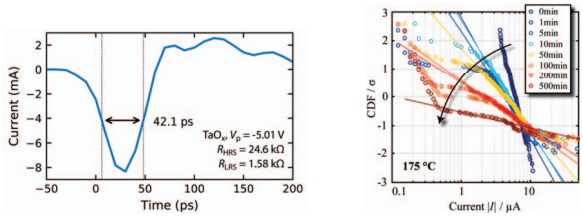


Fig. 8. NeuroVP high level architecture and performance results.



(a) (b)

Fig. 9. (a) Current response of a Pt/TaO_x/Ta-based VCM device, reprinted from [25] with permission of AIP Publishing, (b) Retention of HRS read current distribution, reprinted from [22] with permission of ACS.

fasted reported SET time of a VCM device (see Fig. 9a). It could also be shown that the SET kinetics in the sub-10 ns range are only limited by the electrical charging of the devices [26]. Neither the migration of oxygen vacancies, nor the heating time of the conductive filament cause a detectable delay in the SET kinetics at timescales down to 50 ps. Recently, we investigated the RESET kinetics and could show that they are limited intrinsically at approx. 480 ps by the coexistence of a unipolar switching mode [27]. Faster RESET operations require higher voltages, which, however, trigger only the unipolar SET, occurring at the same voltage polarity as the RESET. The occurrence of the unipolar SET is attributed to an oxygen exchange at the active electrode of the VCM device. This exchange could be suppressed by placing an oxygen blocking layer at the interface between the oxide and the active Pt electrode, with which it became possible to also achieve 50 ps fast RESET times.

B. Array characterization

We fabricated VCM ReRAM devices in a 32 x 1 line-array structure, with a (30 nm Pt/5 nm ZrO_2/20 nm Ta/30 nm Pt) stack [21]. The devices were contacted via a wedge probe card with 32 probes and programmed using a custom array tester based on the µController module platform by aixACCT Systems. Using this setup, the most relevant reliability aspects, being their variability [21], retention [22] and endurance [28] were investigated. We demonstrated that VCM ReRAM exhibits a characteristic variability from read to read comprising random jumps of the programmed resistance state, originating from the random walk of conduction supporting oxygen vacancies [21]. If multiple devices are considered, the instability of all individual devices accumulates to intrinsic statistics (c.f. Fig. 9b, 0 min) which are remarkably stable over time scales of seconds and minutes [21]. Over longer periods or at elevated temperatures, we could show that the intrinsically stable resistance distributions begin to degrade as depicted by Fig. 9b. Thus, we demonstrated that the retention of VCM ReRAM can be described as changes in their intrinsic statistics. In particular, a broadening of the distribution (i.e. increasing standard deviation) could be identified as limiting factor of the long term retention [22]. Furthermore, we investigated the endurance of the line-array structure. Here, we developed an algorithm to optimize the biasing conditions of each device towards maximum endurance. Using this algorithm, we compared the optimized endurance of VCM devices with different ohmic electrodes (Zr, Hf, Ti, Ta). We could demonstrate that

metals with a higher work function reduce the amount of excessively generated oxygen vacancies and thus prolong the endurance [28].

C. Manufacturing Testing

From testing perspective, a complete understanding of possible manufacturing failure mechanisms, their associated defects, and the resulting faulty behaviour of memristors is mandatory. We conducted activities to develop more suitable manufacturing test strategies able to guarantee the memristor's quality. Firstly, a review regarding the manufacturing process related to VCM-based memristors, and a comparative analysis of the CMOS and memristor manufacturing processes was presented in [29]. We identified possible manufacturing failure mechanisms that may affect these novel devices, completing the list of the already known mechanisms. Analysis about possible faulty behaviours was performed, allowing the definition of more accurate fault model [29]. In parallel, a study regarding process variation's impact on the behaviour of ReRAMs has been developed [30]. Electrical simulations considering a ReRAM (3x3 word cell array) were performed to properly identify the tolerated variability as well as the interval where extreme variability causes faults. The obtained results show that different parameters can degrade the functionality of the ReRAM cell and demonstrate the existence of a clear relation between the performed operating sequence and the tolerated percentage of variability. Finally, a new Design-for-Testability (DFT) strategy was specified, implemented, and validated in [31]. The strategy combines the introduction of an on-chip sensor able to measure the voltage of a 1T1R ReRAM cell with the execution of a predefined operating sequence. The obtained results show that the proposed DFT strategy can detect conventional and unique faults.

VI. Neuromorphic circuit concepts

A. Memristive content addressable memory

Here, our aim is the development of neuro-compute-units, based on memristors and focusing on 2 types: 1) memory-arrays based on memristors as possible enablers for efficient CIM, and 2) units that can realize associative operations like ternary content-addressable memory TCAM. We studied operations in analog and digital domains. We developed a new associative memory concept based on complementary resistive switching (CRS) cells, see Fig. 10, wherein both the binary data and its complement are stored. This CAM enables the direct analog computation of the Hamming Distance (HD), i.e. the difference in number of bits between the input and stored data vector. This HD can be used in binary neural networks for very efficient calculation of the VMMs required during inference operation. This concept was used to simulate the inference step on a 1-layer fully connected neural network trained on a binarized version of the MNIST data set, achieving a prediction accuracy of approximately 86% [32].

B. Analog CIM array based on memristor

Here we investigated the computational principles such as VMM in 1T1R arrays and the effect of different of array

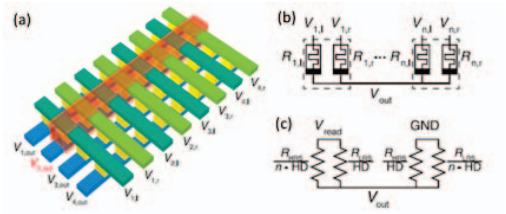


Fig. 10. (a) Lateral CRS-based passive crossbar array. Highlighted part maps to the circuit diagram of (b). (b) Circuit schematic of multiple lateral CRS cells in parallel with a common center electrode. (c) Equivalent circuit.

Fig. 11. Simulated energy consumption in 8x8 1T1R array for VMM operation.

architectures (memory, vertical, and pseudo-crossbar 1T1R) and peripheral circuits for readout and multilevel programming of ReRAM devices. As an example, Fig. 11 shows the calculated energy consumption.

We also looked into the feasibility study of 1S1R array for VMM operation. "Raw", i.e. selector-less ReRAM memory arrays suffer from read errors due to sneak path currents, and increase the power consumption during both read and write due to leakage currents in partially selected devices. Therefore, a selector device is mandatory. A standard CMOS transistor (T) device solves the above problems, however it is a 3-terminal device that dominates the 1T1R cell area, while it has to be integrated in the Si substrate. Alternative 2-terminal selector (S) devices are proposed, with threshold switching devices being the most promising types. The 1S1R cell results in the densest integration of ReRAM arrays and enables 3D stacking.

However, while appropriate for memory array application, the non-linear selector conduction may affect the use of the 1S1R array for in-memory computations. Therefore, we investigated the feasibility of VMM in 1S1R arrays using simulation. The possibility of VMM for different array sizes as well as wire resistances was investigated: for larger array sizes (128x128), the effect of series resistance becomes more important and resistance <2 mOhm is required for proper VMM operations.

C. Digital CIM

Starting from a classic digital design flow, we assessed opportunities to integrate neuromorphic computing principles in the resulting modules. As overarching principle, we embrace the concept of near- or in-memory computing. Hence, a first step was to establish a flow capable to automatically contruct macros of high regularity considering the placement of standard cells and the involved routing. As development vehicle a standard cell based memory [33] is used, that can seemingly be integrated in surrounding digital logic realizing a near-memory

computing macro. This flow was applied to the design of a TCAM [34], which is 1.8x faster than [35] and has a 2.4x lower energy per operation than [36]. Today, the design methodology serves as a strong foundation for regular structures supporting various other algorithms. Other neuromorphic techniques have been added on top explore a wider design space. Besides others, we considered the natural attention mechanism to put more effort or energy on difficult cases resulting in cascaded classifiers [37], and assessed the potential of limiting data to binary values as observed in the signal potential of spikes on the axons resulting in binary neural networks [38]. Optimizations of the binary neural network improve the energy efficiency by up 3.5x compared to previous work [39], while the usage of cascaded classifiers can reduce the number of operations by up to 5.7x. Leaving the classic deterministic and full-precision computing in the digital domain behind, we assessed approximate computing techniques as well as computing in the time domain. The most promising concepts have been fabricated in a 22 nm FDSOI technology. Overall, this enabled us to quantify the underlying trade-off between throughput, accuracy and efficient in neuromorphic hardware realizations.

VII. Conclusion

The researchers from the NEUROTEC I project have published several findings and made available many contributions to the neuromorphic research community. After its official conclusion in late 2021, we plan for another five years with the consequent "NEUROTEC II" project which is planned as a concretization of the investigated technologies in NEUROTEC I, including the realization of chip demonstrators.

Acknowledgement

This work was supported by BMBF in Germany within NEUROTEC, project numbers 16ES1134 and 16ES1133K.

References

[1] F. Staudigl *et al.*, "A Survey of Neuromorphic Computing-in-Memory: Architectures, Simulators and Security," *IEEE Design & Test*, 2021.

[2] A. Hardtdegen *et al.*, "Improved Switching Stability and the Effect of an Internal Series Resistor in HfO2/TiOx Bilayer ReRAM Cells," *IEEE Trans. on Electron Devices*, vol. 65, no. 8, pp. 3229–3236, 2018.

[3] F. Cüppers *et al.*, "Exploiting the switching dynamics of HfO2/TiOx-based ReRAM devices for reliable analogue memristive behaviour," *APL materials*, vol. 7, p. 091105, 2019.

[4] C. Bengel *et al.*, "Variability-Aware Modeling of Filamentary Oxide-Based Bipolar Resistive Switching Cells Using SPICE Level Compact Models," *IEEE Trans. on Circuits and Systems I: Regular Papers*, vol. 67, no. 12, pp. 4618–4630, 2020.

[5] "JART, Juelich Aachen Resistive Switching Tools," https://www.emrl.de/JART.html, accessed: 2021-11-01.

[6] C. Bengel *et al.*, "Utilizing the Switching Stochasticity of HfO2/TiOx-Based ReRAM Devices and the Concept of Multiple Device Synapses for the Classification of Overlapping and Noisy Patterns," *Frontiers in Neuroscience*, vol. 15, p. 621, 2021.

[7] J.-Y. Raty *et al.*, "A Quantum-Mechanical Map for Bonding and Properties in Solids," *Advanced Materials*, vol. 31, 11 2018.

[8] B. Kooi *et al.*, "Chalcogenides by Design: Functionality through Metavalent Bonding and Confinement," *Advanced materials*, vol. 32, 2020.

[9] L. Guarneri *et al.*, "Metavalent bonding in crystalline solids: how does it collapse?" 2020.

[10] M. Wuttig *et al.*, "Incipient Metals: Functional Materials with a Unique Bonding Mechanism," *Advanced Materials*, vol. 30, p. 1803777, 2018.

[11] C. Persch *et al.*, "The potential of chemical bonding to design crystallization and vitrification kinetics," *Nature communications*, vol. 12, no. 1, p. 4978, August 2021.

[12] P. Schüffelgen *et al.*, "Selective area growth and stencil lithography for in situ fabricated quantum devices," *Nature Nanotechnology*, vol. 14, p. 1, 09 2019.

[13] D. Rosenbach *et al.*, "Quantum Transport in Topological Surface States of Selectively Grown Bi2Te3 Nanoribbons," *Advanced Electronic Materials*, vol. 6, 07 2020.

[14] J. Kölzer *et al.*, "Phase-coherent loops in selectively-grown topological insulator nanoribbons," *Nanotechnology*, vol. 31, 04 2020.

[15] J. Kölzer *et al.*, "In-plane magnetic field-driven symmetry breaking in topological insulator-based three-terminal junctions," 12 2020.

[16] A. Grundmann *et al.*, "H2S-free Metal-Organic Vapor Phase Epitaxy of Coalesced 2D WS2 Layers on Sapphire," *MRS Advances*, vol. 4, no. 10, p. 593–599, 2019.

[17] M. Marx *et al.*, "Metalorganic Vapor-Phase Epitaxy Growth Parameters for Two-Dimensional MoS2," *J. of Electronic Materials*, vol. 47, 11 2017.

[18] D. Schneider *et al.*, "Highly responsive flexible photodetectors based on movpe grown uniform few-layer mos 2," *ACS Photonics*, 05 2020.

[19] A. Grundmann *et al.*, "MOVPE of Large-Scale MoS 2 /WS 2 , WS 2 /MoS 2 , WS 2 /Graphene and MoS 2 /Graphene 2D-2D Heterostructures for Optoelectronic Applications," *MRS Advances*, vol. 5, pp. 1–9, 02 2020.

[20] C. Bengel *et al.*, "Implementation of multinary lukasiewicz logic using memristive devices." 2021 IEEE Intl. Symposium on Circuits and Systems (ISCAS), Daegu, Korea, 22-28 May, 2021, 2021.

[21] S. Wiefels *et al.*, "Hrs instability in oxide based bipolar resistive switching cells," *IEEE Trans. Electron Devices*, vol. 67, pp. 4208–4215, 2020.

[22] N. Kopperberg *et al.*, "A consistent model for short-term instability and long-term retention in filamentary oxide-based memristive devices," *ACS Appl. Mater. Interfaces*, 2021.

[23] C. Funck *et al.*, "Comprehensive model of electron conduction in oxide-based memristive devices," *ACS Appl. Electron.*, vol. 3, pp. 3674–3692, 2021.

[24] M. Galicia *et al.*, "NeuroVP: A System-Level Virtual Platform for Integration of Neuromorphic Accelerators," in *IEEE Intl. System-on-Chip Conf.,in press*, 2021.

[25] M. von Witzleben *et al.*, "Study of the SET switching event of VCM-based memories on a picosecond timescale," *Journal of applied physics*, vol. 127, no. 20, pp. 204–501, 2020.

[26] M. von Witzleben *et al.*, "Determining the electrical charging speed limit of ReRAM devices," *IEEE J. Electron Devices SoC*, pp. 667–678, 2021.

[27] M. von Witzleben *et al.*, "Intrinsic RESET speed limit of valence change memories," ACS Appl. Mater. Interfaces, in press, doi: 10.1021/acsaelm.1c00981, 2021.

[28] S. Wiefels *et al.*, "Impact of the Ohmic Electrode on the Endurance of Oxide-Based Resistive Switching Memory," *IEEE Trans. on Electron Devices*, vol. 68, no. 3, pp. 1024–1030, 2021.

[29] L. M. B. Poehls *et al.*, "Review of manufacturing process defects and their effects on memristive devices," *Journal of Electronic Testing*, vol. 37, no. 4, pp. 427–437, Aug 2021.

[30] E. Brum *et al.*, "Evalutating the Impact of Process Variation on RRAMs," IEEE 22nd Latin-American Test Symposium (LATS), 2021.

[31] S. Copeti *et al.*, "Validating a DFT Strategy's Detection Capability regarding Emerging Faults in RRAMs," 29th IFIP/IEEE Intl. Conf. on Very Large Scale Integration (VLSI-SoC), 2021.

[32] T. Ziegler *et al.*, "In-Memory Binary Vector–Matrix Multiplication Based on Complementary Resistive Switches," *Advanced Intelligent Systems*, vol. 2, no. 10, p. 2000134, 2020.

[33] X. Fan *et al.*, "Synthesizable Memory Arrays Based on Logic Gates for Subthreshold Operation in IoT," *IEEE Trans. on Circuits and Systems I: Regular Papers*, vol. 66, no. 3, pp. 941–954, 2019.

[34] X. Fan *et al.*, "Compiling All-Digital Embedded Content Addressable Memories on Chip for Edge Application," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1–1, 2021.

[35] C.-X. Xue *et al.*, "A 28-nm 320-Kb TCAM Macro Using Split-Controlled Single-Load 14T Cell and Triple-Margin Voltage Sense Amplifier," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 10, pp. 2743–2753, 2019.

[36] I. Arsovski *et al.*, "A 32 nm 0.58-fJ/Bit/Search 1-GHz Ternary Content Addressable Memory Compiler Using Silicon-Aware Early-Predict Late-Correct Sensing With Embedded Deep-Trench Capacitor Noise Mitigation," *IEEE J. of Solid-State Circuits*, vol. 48, no. 4, pp. 932–939, 2013.

[37] C. Latotzke *et al.*, "Cascaded Classifier for Pareto-Optimal Accuracy-Cost Trade-Off Using off-the-Shelf ANNs," 2021.

[38] T. Stadtmann *et al.*, "From Quantitative Analysis to Synthesis of Efficient Binary Neural Networks," in *ICMLA*, 2020, pp. 93–100.

[39] D. Wan *et al.*, "TBN: Convolutional Neural Network with Ternary Inputs and Binary Weights," in *ECCV*, 2018.