VDE SPEC

Neuromorphic computing based on new devices – a layer model for developing AI hardware

VDE SPEC 90033 V1.0 (en)

VDE

Preface

Release date of this VDE SPEC: 01.11.2024

No draft has previously been published for this VDE SPEC.

This VDE SPEC was developed according to the public VDE SPEC specification. It is being developed in VDE SPEC consortia and does not require the participation of all potential interest groups.

This VDE SPEC is not part of the VDE set of regulations or the German set of standards. In particular, this VDE SPEC is not a technical rule within the meaning of Section 49 EnWG.

The authors of this VDE SPEC are:

- Beyer, Sven, GlobalFoundries
- Bolzani Pöhls, Leticia, RWTH Aachen
- Dittmann, Regina, FZ Jülich
- Dudek, Damian, VDE ITG
- Gemmeke, Tobias, RWTH Aachen
- Gude, Michael, CologneChip
- Joseph, Jan Moritz, Roofline AI
- Kohlstedt, Hermann, University of Kiel
- Leupers, Rainer, RWTH Aachen
- Mikolajick, Thomas, TU Dresden
- Nielen, Lutz, aixACCT Systems
- Paintz, Christian, Melexis
- Thiem, Steffen, X-FAB Semiconductor Foundries
- Waser, Rainer, FZ Jülich
- Wehn, Norbert, RPTU Kaiserslautern
- Wenger, Christian, IHP
- Wiefels, Stefan, FZ Jülich
- Wirth, Matthias, VDE WIN
- Ziegler, Martin, Kiel University

There are currently no rulings applying to this topic in any German standards.

Despite great efforts to ensure the correctness, reliability and precision of technical and non-technical descriptions, the VDE SPEC project group can neither explicitly nor implicitly guarantee the correctness of the document. This document is used in the knowledge that the VDE SPEC project group cannot be made liable for damage or loss of any kind. The application of the present VDE SPEC does not release the user from responsibility for their own actions and is therefore at their own risk.

In the course of the manufacture and/or introduction of products into the European internal market, the manufacturer shall carry out a risk analysis in order to first determine which risks the product may entail. After performing the risk analysis, he evaluates these risks and, if necessary, takes suitable measures to effectively eliminate or minimize the risks (risk assessment). The present VDE SPEC does not release the user from this responsibility.

Attention is drawn to the possibility that some elements of this document may affect patent rights. VDE, DKE, and DIN are not responsible for identifying any or all of the related patent rights.

Executive Summary

The development of our digital world is accompanied by an ever greater demand for computing capacity and for energy, with corresponding consequences for our climate. It is forecast that by the year 2030, around one fifth of global electricity production will be needed for information technology (IT) (Dhar, 2020). One of the fastest growing fields of information technology is currently artificial intelligence (AI). However, this has had a highly inefficient energy balance hitherto (Jones, 2018), due to the conventional CMOS technology used as computing hardware. But it is not only hardware that needs redesigning but also the algorithms used on the hardware, which traditionally follow the „von Neumann concept".

These limitations make it more important to develop new energy-efficient technologies that protect resources. In turn, this demands new concepts for developing new materials, technologies and microelectronics to process, save and transmit information.

The *neuromorphic computing (NMC)* approach focuses on the structure and functionality of biological nervous systems in an attempt to achieve realistic simulation of neural networks. Biological neurons (nerve cells) are capable of both processing and saving information. In doing so, the human brain consumes about 20 watts in power, compared to several megawatts in server farms with GPU-based computing nodes. Even if the comparison depends greatly on the specific field of application, it can still be used for certain tasks, clearly illustrating the drastic difference in energy consumption between technical systems and nature.

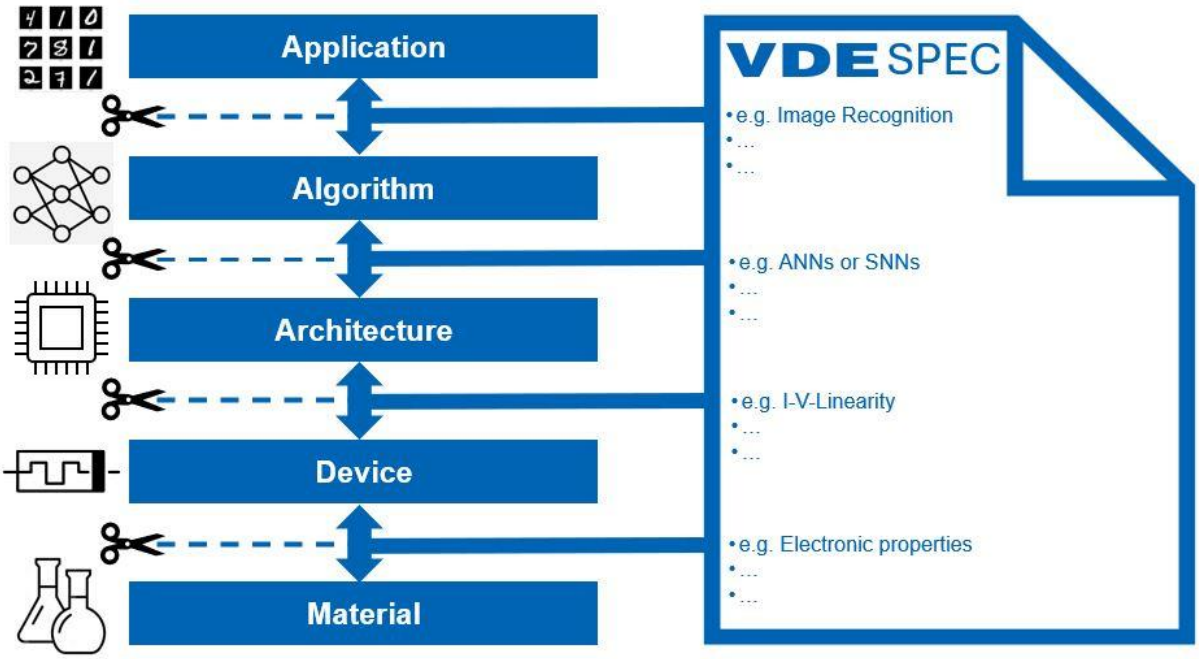

Figure 1 – Layer model with the respective interfaces

The NMC layer model approach entails describing sensor and information processing with many different technical developments, then allocating these to the corresponding layers. Particular attention is paid to the interfaces between the layers, as the handover involved at these points is rarely given clear definition, or is missing entirely. This serves to advance further development of NMC technology as a whole and provides concepts for defined applications.

The layer model defines five superimposed layers with their respective interfaces. Depending on the function, devices in the same layer with clearly defined interfaces are interchangeable.

While academic development pursues research in various directions, only technologically resilient, verified concepts can be turned into systems that are then available as added value for the ongoing development process in products and services. This VDE SPEC on NMC is therefore an initial stipulation of framework parameters to facilitate this transfer from research for research and application through to product development, following a structured path. In its role as independent technology association that focuses on electrical, electronic and information technologies, the VDE is publishing this coordinated VDE SPEC on NMC and using it as the basis to build a testing and validation platform.

The validation platform will be built provisionally on a controller-based Device-under-test (DUT) with a bus system that has adapters for integrated NMC switchgear units primarily on memristive devices (Fig. 2). To record the energy efficiency, it is necessary to use the chip under typical conditions, i.e. test patterns are generated and validated on the chip. Initially, X-bar arrays are assumed to work with different pulse shapes – not only with DC signals.

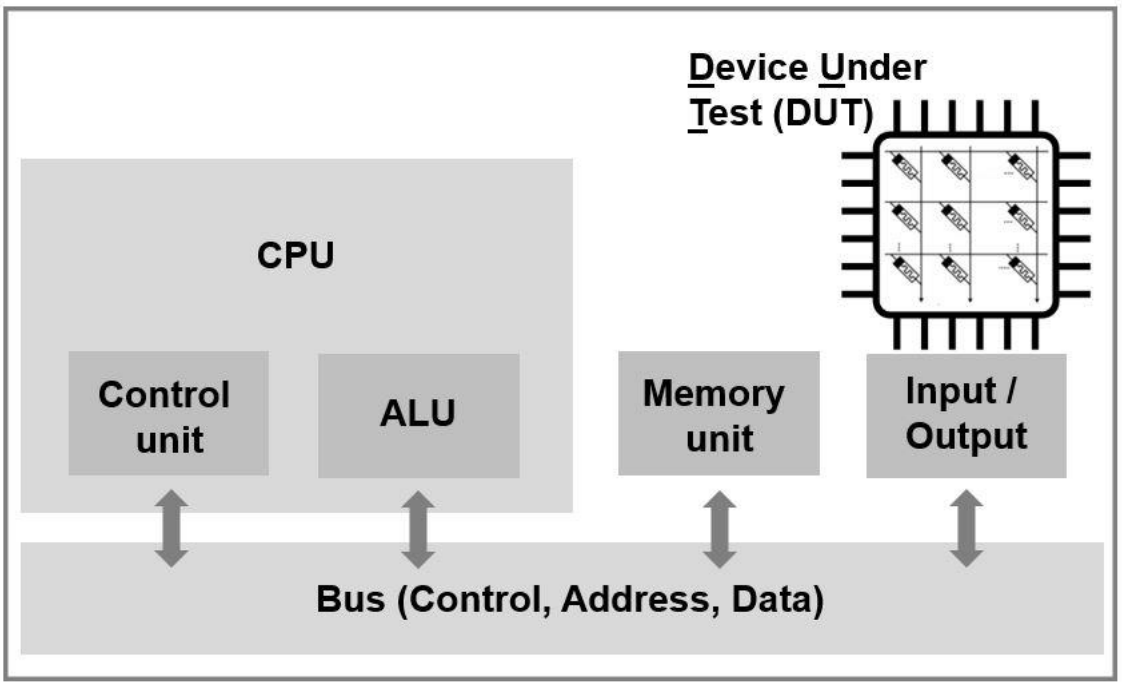

Figure 2 – Circuit board with Device-under-test (DUT) and bus system

This unit acts as exchangeable module that can be connected to further periphery (Fig. 3) for conducting signal measurements and validating switching statuses, but also for checking the energy efficiency and benchmarking algorithms and stipulated use cases compared to conventional information processing systems.

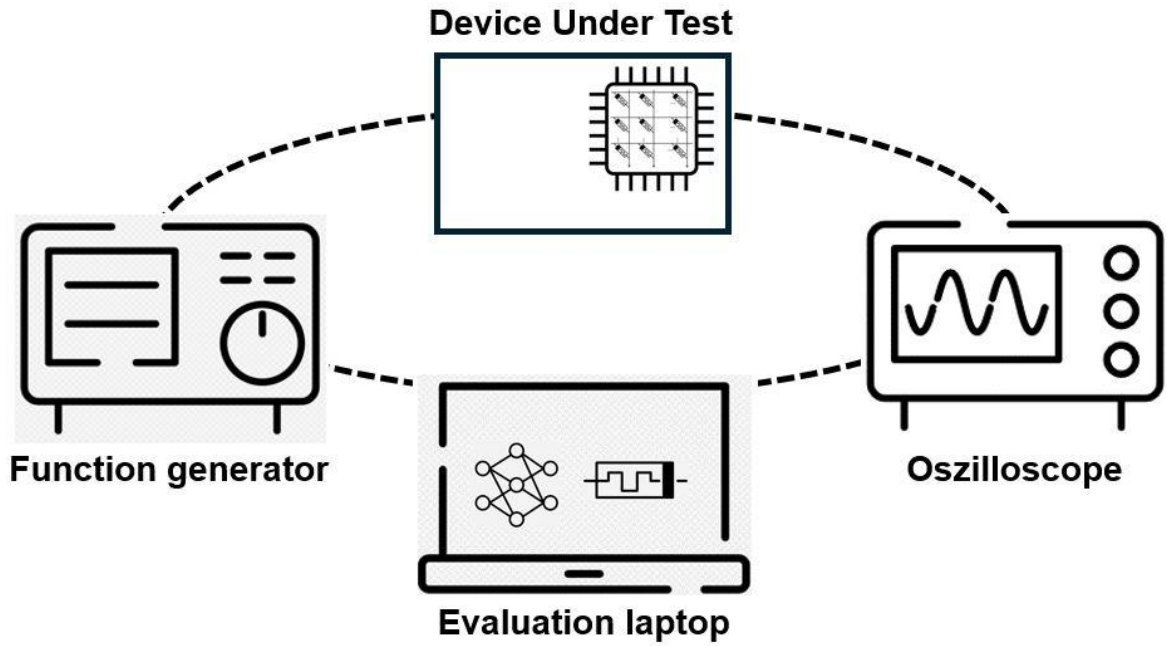

Figure 3 – Device-under-test (DUT) with peripherals

External analysis is used in advance for material testing and device parametrization. Although „on wafer” and „on chip” measurements are an essential part of further planning, they are being put back for the time being in order to prepare the technological development initially with the approach described here, and to support both the academic sector and industry with know-how transfer.

Table of contents

List of figures

List of tables

1 Scope

Potential users of neuromorphic devices and systems should be able to take a neutral approach to checking and assessing the advantages and performance of various concepts. This presumes a previously coordinated set of metrics, provided in this case as a collaborative effort by the expert committee initiated by the VDE. The content of this document is an essential part of this comparability for more efficient transfer to application and commercialization of this technology.

This VDE SPEC defines requirements and criteria for assessing devices and systems in the context of memristive switching devices and systems based on neuromorphic concepts.

Users of such pioneering technologies benefit on the different technical levels with a transparent comparison for their developments, thus pursuing the objective of making it possible to assess innovation cycles with greater efficiency and transparency. The specifications defined here serve in the next stage for building a test center and certifying assemblies. This provides effective support for developing innovative applications with benchmarking on the basis of neutral tests.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

DIN and DKE maintain terminological databases for use in standardization at the following addresses:

- DIN-TERMinologieportal: available at https://www.din.de/go/din-term
- DKE-IEV: available at https://www.dke.de/DKE-IEV

3.1
Algorithm

In the context of artificial neural networks (ANN), algorithms are abstracted models of connected artificial neurons that simulate biological living systems. They are used to offer practical solutions for complex tasks from various application areas.

3.2
Function unit

In the design hierarchy being considered here, this is a physical monolithic block (e.g. a crossbar including periphery and connection to a bus system).

3.3
Neuromorph

Ancient Greek: νεῦρον – *neuron*, nerve, and μορφή – *morphé*, shape, form

3.4
Neuromorphic computing

The concept of neuromorphic computing is a technical simulation of biological systems for information processing. It is expected to achieve an increase in computing capacity and also improve energy efficiency.

3.5
System

In the context of a design hierarchy for hardware/software co-design, a system is understood as the interaction between system-on-chip, subsystems, function units and devices.

Other definitions and abbreviations can be found in the annex with the various tables (see below).

4 Abbreviations

Abbreviation	description
ADC	Analog-to-digital converter
AFM	Atomic force microscopy
AI	Artificial intelligence
ANN	Artificial neural networks
CIM	Computing-in-memory
CMOS	Complementary metal-oxide semiconductor
DAC	Digital-to-analog converter
DDR	Double data rate
DUT	Device-under-test
HW	Hardware
ITG	Information Technology Society (of the VDE; https://www.vde.com/de/itg)
MAC	Multiply-Accumulate (Operation)
ML	Machine learning
MTJ	Magnetic tunnel junction
MVM	Matrix-Vector multiplication
NMC	Neuromorphic Computing
NoC	Network-on-Chip
SEN	Scanning Electron Microscopy
SoC	System-on-Chip
SPEC	Spezification (here: VDE SPEC)
STM	Scanning Tunneling Microscopy
SW	Software
VDE	German Association for Electrical, Electronic & Information. (www.vde.de)
XRD	X-Ray Diffraction

5 Contents

The content of this VDE SPEC has the following structure.

Section I.I. explains the significance of NMC respectively neuromorphic concepts for current applications in information technology, with particular reference to artificial intelligence.

Section I.II. addresses the significance of a hardware/software co-design for efficient development of complex architectures and systems.

Sections 5.1. to 5.5. take a comprehensive system perspective covering aspects **from application via algorithms, architectures and devices to materials**. One essential aspect consists in describing the respective interfaces between two layers and the requirements made of the technical assemblies or systems in these layers.

■ I.I. Neuromorphic computing and AI hardware

NMC addresses energy-efficient, performant information processing in biological systems, with potential for an alternative to conventional CMOS-based von-Neumann architecture.

So-called memristive devices offer a practical way of achieving the objectives of NMC. These are primarily electronic devices with variable, resistor-based memory function, with the implied effect of both saving and also processing information by flexible adaptation of other signal inputs. This comes closer to information processing in biological systems, such as the neuron (Ziegler, 2020).

First and foremost, hardware for the NMC should not replace conventional CMOS technology for digital circuits but supplement it in the corresponding application fields. There is therefore a special focus on using materials for memristive devices that can be integrated in the existing CMOS technology. Potential materials being considered in this context are transition metal oxides and, most recently, transition metal dichalcogenides, as well as memristors based on phase change materials and ferroelectric materials (Waser, 2019).

■ I.II. Hardware/software co-design

Technical testing of systems for NMC needs adequate methods and equipment together with an unequivocal description of the test procedure (Leupers, 2024). It is necessary to bridge the gap between the requirements for system applications, circuit architecture and electronic devices and the materials that are available, which requires stipulation of the system layers with their interfaces. A validation platform with the Device-under-test acts as a practical utensil, connecting the integrated circuit or so-called neuromorphic hardware with inputs and outputs. Such a flexible test environment facilitates software development for migrating artificial intelligence applications to the new memristive device-based circuits. This is one variant for the evaluation and validation of neuromorphic system architecture together with the corresponding software.

Thanks to the compact structure of this measurement and characterization platform, it can be used without complex add-ons or special requirements. The platform flexibility deserves a special mention: it offers inputs and outputs with a specially developed connection matrix for variable assignment. The Platform is thus independent of specific chip packages and pin assignments. It permits fully automatic generation of all necessary input pulses. The specially implemented transimpedance amplifier allows for high-precision current measurement. To correctly assess the performance and energy efficiency of the neuromorphic circuit, peripheral elements such as ADCs or transimpedance amplifiers must be measured and taken into account separately.

The platform is controlled by a microcontroller containing a python interface for simple implementation of the necessary measurement routines.

The focus is on developing corresponding software for the NMC, facilitated by the specially implemented computing-in-memory interfaces. Developed mapping and scheduling algorithms can thus be tested and verified early on in terms of scalable semiconductor technology.

A multi-stage process is necessary for applying AI models and methods to systems for the NMC.

- **Stage 1** entails loading an AI product as software. In specific terms, this means that the Tensorflow or PyTorch model is imported. The model is then adapted with three functions. First, the model is analyzed and adapted e.g. by suitable quantization to the properties of the integrated circuit. Second, a first estimation is given of the application's performance (e.g. the expected latency) for later validation. Third, the model is available in a formal data format so the compiler can continue processing it in the next steps.

- **In stage 2**, the imported and possibly quantized AI software (typically the model of a specific neural network) is processed by the compiler. The model is viewed using different mathematical abstractions and optimized for NMC characteristics. A specific example in this case consists of viewing the model as a network of neurons that are saved in memristors; the memory location defines how the integrated circuit has to move the data between the neurons.

- **In stage 3**, the compiler produces a general abstraction of the AI model for the computing-in-memory concept of the NMC integrated circuits. The abstraction is selected so that all integrated circuits can be addressed. Possible examples include memory operations (LOAD/STORE) and matrix vector multiplications (MVM). Synchronization between several calculation kernels is also necessary.

- **Stage 4** specifies the general computing-in-memory abstraction for a specific integrated circuit, translating the abstract commands (e.g. LOAD) into specific commands (e.g. a bit code). Separating stages 3 and 4 makes it possible to produce a “retargetable” compiler for different integrated circuits.

- **Stage 5** executes the generated code for an AI model. This is possible either on the integrated circuit of the user to verify their AI model, or (at an earlier point in the design cycle) in a hybrid simulation platform including the test platform. The latter allows for system verification prior to scalable technology implementation, or for precise evaluation of various integrated circuit architectures for serial NMC development.

5.1 Application layer

The application scope for NMC is varied and diverse. It is a technology that is seen to have great potential for improving the currently available conventional technologies for modeling in artificial intelligence (AI) and machine learning (ML) when it comes to performance, complexity and also reduced energy consumption, with a great added value in terms of information processing.

Implementation of these requirements depends above all on a transparent comparison of the various technological approaches, including metrics and structured classification of the technological demand in individual steps and groups. The application areas play an essential role in stipulating these metrics as the determining factors in demand for memory, logic and data processing capacity. Today, different research and development approaches are available but are not easily brought together. With this VDE SPEC we hope to smooth the way in this direction, showing a few important examples of possible applications.

- **Edge computing and Internet-of-Things (IoT)**

NMC is ideal for edge computing applications and on IoT devices with low power consumption thanks to local operation and the power supply. Real-time processing plays a major role, with information processing taking place directly on the SoC. NMC-based integrated circuits would allow for sensor data processing, pattern recognition and efficient decision-making in a defined context directly ‘on the edge’.

- **Sensors**

NMC can be used for more efficient building of sensor data and sensory processing systems with learning algorithms integrated directly in the same circuit with sensor data capture. An additional microcontroller or CPU is then superfluous. Furthermore, the algorithms are based on biological mechanisms for perceiving sensory data, with real-time processing making efficient use of primary energy. This is also technologically less elaborate because complex von-Neumann computer architecture is not necessary. Multimodal data capture can thus be scaled over several channels and efficient processing for recognizing patterns and extracting relevant information.

- **NMC platforms**

As NMC technology matures further, specialized platforms will emerge that provide the hardware, software and development tools for a wide range of applications in sensor data processing of large data volumes, also in real time. NMC supplements, extends or even replaces von Neumann architectures, resulting in greater flexibility in designing information processing systems, as with integrated circuits

for edge computing. These NMC platforms can be used for scaling and verifying the existing technologies, while also opening up experimental space in academic and industrial research.

- **Processing audiovisual data**

One of the prime applications for NMC is character and image recognition with the previous complexity levels of circuit technology. However, other applications are also conceivable, for instance in medical technology. Information processing of audio signals for hearing aids and hearing implants such as the cochlea implant are a good area of application with clear performance parameters, clearly defined information processing and local energy consumption. Although other application areas are emerging, including sensor data processing for autonomous driving, the corresponding development horizons are currently still too broad for inclusion in this VDE SPEC for NMC. Depending on maturity and development, such applications are sure to play an important role as NMC comes into play, but this SPEC focuses on clearly foreseeable parameter space with a manageable degree of complexity.

Altogether, NMC has the potential of being a technology that drives innovation and progress in a broad range of areas, offering new possibilities for improving the efficiency, autonomy and intelligence of systems and devices. Given the ongoing nature of R&D efforts, it can be presumed that NMC will be deployed increasingly in large-scale applications that benefit from its unique capabilities.

Interfaces in the layer model

The following section explains the layers and corresponding interfaces of the NMC layer model.

The layer model defines and categories the function units, and specifies the definitions, thus allowing for a structured overview and facilitating exchange between the individual disciplines pursuing research and development work in the individual layers. This permits troublefree communication and makes it possible to forward parameters needed for optimization or provision and for implementation of new parameters in the overall system.

Many of the exchanged parameters are integrated in a complex overall system, that appears to be simple only at first glance. In fact, many details have to be fulfilled in terms of functionality, reliability, security and efficiency. The tasks to be solved in the individual layers extend from electronic material behaviour via signal routing in the devices and circuits through to information processing in hard- and software on the higher abstraction levels. Furthermore, the NMC layer model has a regulated sequence for processing the selected application areas. VDE test procedures are then derived accordingly as a follow-up.

Interface 1 <> 2: Application layer <> Algorithmic layer

It is clear from the above discussion that the applications play an essential role in the structure of the layer model. The selected, stipulated applications make it possible to derive the requirements for a higher abstraction level on top of the underlying levels of lesser abstraction. The purpose of this structure consists in defining the workflow and handover points to make them clear for all stakeholders building this system, establishing a shared understanding with known parameters in the respective layers and interfaces. This is referred to below as specification framework.

Specification framework
Specification of the task being performed in order to derive the corresponding algorithm The measurement results are assessed using selected algorithms implemented on the respective NMC hardware. This entails defining a suitable dataset for assessing energy efficiency and throughput (MNIST, possibly CIFAR-10). The focus is on quantitative assessment of the new system components. Problem complexity is of secondary importance as long as it does not impact on the functional properties of the system components as such and in interaction with other system components. Consideration should also be given to stipulating the usage profile in order to safeguard a uniform workflow of the algorithms to the applications. This refers for example to daily updating of the parameters for a stipulated period of time. Other scenarios are also conceivable, depending on the requirements profile of an application. The NMC hardware being used together with product validation and testing (PVT) must include exceptions, so-called „worst-case corners” according to the requirement profile

Assessment on system level

Quantitative capture of the measured data is essential for validation of the results with fixed indicators. In each case, the measurement data are pigeon-holed and classified using the layer model and interfaces shown here.

Algorithm accuracy is defined according to the specific NMC hardware being used. If not deterministic, units are given individual mean values as for N runs.

Latency is significant in terms of assessing NMC system efficiency. It is important to measure the time intervals for the process of the first input/output data (IO) from hardware to IO result; this must be recorded as a parameter.

Energy demand is a central optimization parameter. Stipulating measurement of this variable is therefore very important for measuring the efficiency of the individual system and also in comparison with other approaches. A conventional von-Neumann computer is used as reference system and stipulated for the run of the algorithm. Energy consumption could be captured as a quotient run of first data to IO with determined presumption of the energy costs for IO communication, such as energy / IO bit.

Surface demand on the chip is taken as a comparison parameter. Consideration is given to units such as on-chip memory, control unit, computing units (ALU), IO bus amplifiers or also integrated units such as the memristive crossbar array. The integration advantage is also validated during test operation to allow for conclusions about design and technology process through to the material layer. Multiple use of operation steps in the algorithm is also possible and considered during benchmarking.

Structured documentation of the results is the basis for validating and testing the NMC systems; measurement is according to models, circuit simulations, database for a virtual tape-out and physical parameters of the NMC hardware, and also according to component breakdown.

Tolerances are also an integral aspect of the test parameters, including definition of the following, among others: considered losses in chip surface yield, tolerable worst-case results including the process capability indices, such as 3-sigma as standard deviation to the mean for non-deterministic execution.

Proactive documenting and assessment of additional measures for executing the test algorithms is particularly necessary for memristive and other new devices, including

- initializing the algorithm workflow and energy consumption for using the NMC hardware
- loading parameters, training runs for the crossbar arrays and other alternatives measured according to latency, run efficiency and energy demand
- data retention behavior and refreshing data in the memory cells and units based on the test parameters multiple writing, reading, read disturb, soft errors and, error correction methods
- load tests for error tolerance and ageing with multiple unit triggering also with higher signal pulse voltages
- stipulating the necessary test methodology edge steepness, time intervals such as “time-0” and “in-the-field”, depending on the NMC hardware being tested

Algorithm for validating NMC systems

Basically, the choice of problem solution algorithm is arbitrary as long as it fulfils the specified minimum application requirements so that structured solution of the task is possible. The algorithm must undergo reproduction for traceability reasons. This is done as follows:

- algorithm for solving the task (e.g. network model including parameters) formulated as pseudocode
- if the algorithm matches the NMC hardware, the training data must also be available in the corresponding format
- training algorithm documented according to training/validation/test datasets and hyper-parameters
- stipulated precision data should encompass at least n>10 independent (emulated) inputs and n>10 trainings each

Addendum: security aspects and open points in the algorithm

To assess security-relevant aspects, firstly the viewed attack vectors must be defined, followed by assessing weaknesses in the algorithm, system architecture and architectural components. That is not part of this VDE SPEC but is taken into consideration in an addendum with a focus on security aspects.

General, internationally used definitions relating to NMC systems and machine learning referring to pattern recognition:

Table 1 – Characteristics and Characteristic Expressions at the interface between Applications and Algorithms

Characteristic	Characteristic Expressions					
Applications	Classification	Prediction	Clustering	Association	…	
Neural Networks	ANN (SLP, MLP)	BNN	CNN	RNN (LSTM, …)	SNN	…
Datasets	MNIST	CIFAR-10	ImageNet	…		
Inference / Training	Inference	Training	Inference + Training	…		

Definitions and Abbreviations (alphabetical):

- *CIFAR-10* – the CIFAR-10 dataset (Canadian Institute For Advanced Research) is a collection of images that are commonly used to train machine learning and computer vision algorithms
- *Datasets* – integral part in the field of machine learning, consisting of texts, numerical data, audio, images, videos etc. for solving and analysing AI challenges
- *Inference* – applying the formerly trained capability to new data
- *MNIST* – the MNIST dataset (Modified National Institute of Standards and Technology database) is a large dataset of hand-written digits that is commonly used for training various image processing systems
- *Neural Networks (NN)* – different architectures of Artificial-NN (ANN) such as Single-Layer-Perceptron (SLP), Multi-Layer-Perceptron (MLP), Binarized-NN (BNN), Convolutional-NN (CNN), Recurrent-NN (RNN), Long-Short-Term-Memory-NN (LSTM), Spiking-NN (SNN), and others
- *Training* – learning a new capability from existing data

5.2 Algorithmic layer

NMC requires comprehensive consideration of a large design space. Established datasets are often used to this end for quantitative comparisons in terms of precision and target values. (Gemmeke, 2024).

To promote research into the development of new devices and function units with improved quantitative comparability, examples of trained neural nets are available in the public domain, for example https://lnkd.in/e-F9wce6 or https://lnkd.in/ec-9Cicu. These work with binary values (+1 and -1) and supply results that approximate floating point calculations for simple datasets such as MNIST or CIFAR-10. The repository named above contains parameters, python code and details for (re-) training these networks.

In a binary neural network, floating point weights are usually replaced by binary weights. This optimizes memory capacity and computing work and is therefore ideal for devices with locally restricted resources - so-called systems-on-chip (SoC). While binary weights result in acceleration, binary neural networks do not yet achieve the same precision as their equivalents with floating point weights in 32-bit networks.

Basically, optimizing the number representation is and will remain a top research issue, also when it comes to complex information data networks, with great relevance for social, economic and industrial applications.

Interface 2 <> 3: Algorithmic layer <> Architecture layer

Specification framework
The correlation between algorithm and NMC hardware (HW) is based on a definition of the HW architecture. This HW architecture consists of various components (see chapter 5.3), to which instructions of the algorithm description are applied in pseudocode. Besides function units transforming data to the corresponding computing units, the architecture includes data storage and data exchange components.

General, internationally used definitions relating to NMC systems and machine learning referring to von-Neumann computing units or function blocks of a microprocessor.

Table 2 – Characteristics and Characteristic Expressions at the interface between Algorithms and Architectures

Characteristic	Characteristic Expressions			
ADCs	on-Chip	off-Chip	…	
DACs	on-Chip	off-Chip	not implemented	…
Sensing Electronics	on-Chip	off-Chip	charge integration	…
Activation Functions	sigmoid	hyperbolic tangent	rectified linear unit	…

Definitions and Abbreviations (alphabetical):

- *Activation Functions* – so-called neuron activation functions – most common are: Sigmoid, Hyperbolic Tangent (tanh) and Rectified Linear Unit (ReLU). Activation functions can be realized either in SW or in HW
- *ADC* – system that converts an analog signal into a digital signal, e.g. converting an analog input voltage or current to a digital number representing the magnitude of the voltage or current
- *DAC* – system that converts a digital signal into an analog signal – performing the reverse function of an ADC
- *Sensing Electronics* – e.g. a transimpedance amplifier (TIA) for converting current to voltage or a charge-based accumulation circuit

5.3 Architecture layer

The HW architecture as a whole is to be seen as a selection of subsystems and their connectivities, and must offer a certain functionality.

It is appropriate to have an undercut between function units and subsystems. A function unit for example can be a monolithic physical block such as a RAM, cache or ROM memory, and also a crossbar array that needs additional subsystems for it to be used in an SoC. Such additional subsystems can be the control unit, for example (here: controller for programming the crossbar array) or the interface logic to the network (here: network-on-chip, NoC).

The HW architecture is an essential element in chip design and its optimization plays a crucial role for efficient execution of an algorithm. At the same time, optimization scope is feasible by varying and rating the function units, such as memories, processors and NoCs. Such scope can be measured with the above mentioned parameters (see interface 1<>2:).

A common approach for artificial neural networks is to outsource data such as weights, input data and interim results to an area outside the viewed architecture, in other words, to an external, peripheral memory. On the other hand, it makes sense to assess how efficiency is measured with the listed parameters on the system level so that the relevant scope for optimizing the chip architecture is not unnecessarily restricted by stipulating constants and parameters. This ensures that data exchange is considered from input to output after the task has been fulfilled.

Measuring algorithm efficiency alone is not deemed expedient. As already explained in chapter 5.1, the reason for this is the focus on the overall system and solving the task for a quantitative check on the obtained result and the related energy consumption. Simply checking on the level of an individual computing operation on the system (here e.g. MAC operation) is not expedient. On the one hand, various papers have shown that the energy demand of the individual computing operation can be neglected for various use cases and architectures. On the other hand, there are numerous NMC approaches where the MAC operation can be carried out more efficiently. Direct comparison of the results from individual computing operations in the algorithm layer is therefore not possible.

Defining a reference implementation while stating the characterization in a conventional logic family of CMOS technology provides the basis for further development and testing. This NMC system is the starting point as a reference system that allows for optimization and proprietary solutions in all layers of the model. Control and assessment is then automated in the layer, with the same assumptions for all proposed solutions. This NMC HW system is the reference system for submitting a simple SoC with processor, memory and accelerator to testing and benchmarking in comparison with conventional

systems. As already mentioned in chapter 5.2, peripheral function blocks such as DDR and ADC etc. may be added to the overall system after choosing the stipulated application routine.

To take account of the properties i.e. parameter spaces of the NMC function block, the assessment methods must be adapted and parameterized accordingly. Examples include switching frequency, endurance and stability of the written values under load, subsumed under “reliability” of the used memristive or other devices. It is important to distinguish whether this is deterministic or non-deterministic.

This context already indicates that optimization of an individual memristive (or other) device does not play a decisive role for measurement, testing and benchmarking. This would only allow for characterization and modeling of a crossbar array as function unit. By contrast, system-level assessment considers modeling in its stipulated parameters, including measuring the obtained solution with corresponding comparison to the reference system. Greater effort is thus required for adapting and modifying the HW architecture, as this entails adapting an algorithm to the new architecture.

Interface 3 <> 4: Architecture layer <> Device layer

Specification framework
The definition of this interface between the architecture and device layers plays an essential role in this model. As far as devices are concerned, pure research has taken a great many approaches, each with differing degrees of technical maturity so that they are only rarely implemented in circuits and subsequently in HW architectures. Furthermore, the parameter space for designing integrated circuits is not firmly defined so that it is rarely feasible in practical terms for random devices to be implemented in memristive technology, for example. Stipulating defined parameters at this interface therefore plays a decisive role for circuit design and for integrating the periphery (SoC, subsystems, function units). This then facilitates transfer from research into the system design, establishing content-related communication between two disciplines. For each function unit, successful integration at this interface is measured by latency for the computing operation in the documented HW system, energy consumption and the space required for the circuit on the chip or in the package. The reference system is the conventional von-Neumann computer architecture in CMOS technology, on the basis of signal paths of binary Boolean algebra. Deviation of coding in the architectures being checked demands reconciliation and an addition to the prevailing stipulation. In view of the fact that the requirements made of the functional properties of the devices are described by parameters and depend on the characteristics of an HW function unit, it is not possible to provide a complete list of relevant parameters. Checking and adapting to the system level is decisive and practical in this case. Methods for error correction, redundancy or calibration are used in the system level where they are validated and measured. Particularly in the algorithm layer, the joint design (HW/SW co-design for the NMC) results in additional degrees of freedom that enhance the parameter space with new functionalities of memristive devices or other possible technologies. Assessing the possible gains offered by these new device technologies in terms of energy efficiency, error correction, latency and necessary space urgently needs test parameters to be defined, which is an integral part of this VDE SPEC. Whether this refers to local optimization in the respective layer or even cross-layer „global” optimization of the overall system depends on the approach and relation to the layer model. „Global” optimization and benchmarking in the layer model presented here requires modeling with the *application*<> *algorithmic* <> *architecture* <> *device* <> *material layers*. Modeling in the architecture and device layers is guided by parameters assigned to the respective layer or interface while satisfying the overriding solution of the task (deterministic and suitable number representation) from the stipulated application. While considering the parameters of the devices, it is important when checking the NMC HW architecture to have models that depict not only I/U non-linearity and device behavior that is dynamic in time, but also variability and latency. Corresponding models are stipulated and supplemented by defining the device parameters. These parameters are summarized in table 3. The respective devices are measured by means of voltage pulses with a specified signal sequence to ascertain the parameters. A uniform measurement protocol is defined for these pulse measurements to ensure comparability between various device groups. These physical parameters obtained by measurements are the basis for the described models used in the design of the NMC HW architecture. Tolerance space is given the same consideration as uncertainties and dynamic behaviour through transient analysis. These models contain both physical and abstract devices, and circuit properties for the circuit simulations in the architecture layer (chapter 5.3) for rating the system. The level of abstraction from the measured parameters is insignificant in the devices layer and the interface described here.

Measuring approaches for characterizing resistive memory elements (Nielen, 2023)

Special analytics is required for measured characterization of resistive memory elements, here also memristive devices. A distinction is made between characterizing individual cells and characterizing memory arrays/matrices.

Depending on configuration, individual cells need either two (here a resistive resistor, 1R) or up to four combination possibilities (here a combination of a transistor and a resistive resistor, 1T1R).

In the devices layer, characterizing the composition of such individual cells (crossbar arrays) as relevant for NMC applications demands more contacts or connection possibilities (up to now, the upper limit is 64 –128 – state of the art). Both individual cells and memory matrices must be contacted, ideally this should be automated. Contact needles are positioned using a microscope for measurement. Temperature-dependent characteristics are measured with a thermochuck. Automation by so-called waver probes („on-wafer measurement") warrants a secure, reliable process.

The high degree of flexibility and the range of configuration possibilities allows the tester to carry out any number of test scenarios for high-resolution characterization of memristive devices. This is done with specially rated matrix test systems for collecting the named physical parameters. One essential aspect consists in adapting the amplifier circuits to the demands of fast current limiting and wideband coverage of voltage ranges up to 10 V in this case, as explained in the following specifications.

Characterization of individual cells

High bandwidths are to be provided for the characterization of resistive individual cells, particularly for ultrafast pulse measurements in this case of up to 250 MHz bandwidth with fast current limiting here of up to 30 ns reaction time, and current measurements here with a bandwidth of up to 100 MHz.

The measurement routines include the following aspects:

- electroforming, derived from galvanization to describe the diffusion processes in solid state
- unipolar and bipolar switching (I/U characteristics and pulse forms)
- retention, here: parameter for long-term stability of saved data
- endurance, here: parameter for describing the number of deleting/programming cycles without data degradation
- life-time acc testing (in combination with opt. thermochuck), degradation test under load and stress tests
- switching kinetics, here also device dynamic and transient analysis
- quantized conductivity (histogram), here: parameter for conductivity and resistance behavior

Characterization of memory matrices

The characterization of resistive memory matrices, here with up to 32 x 32 channels for digital, analog and NMC applications, requires bandwidths of up to 100 MS/s incl. fast current limiting with a reaction time of up to <100 ns and fast current measurement up to 100 MHz.

The measurement routines include the following aspects:

- electroforming
- unipolar and bipolar switching (I/U characteristics and pulse forms)
- write and read methods for analog voltage weighting, programming of the crossbar array
- signal sequences resolved in time, as here „spike current integration" for crossbar arrays
- computing-in-memory (ACN, seq. Logic)
- spike timing dependent plasticity (STDP, completely variable pre- and post-synaptic signal generation)
- short and long-time plasticity

Table 3 – Characteristics and Characteristic Expressions at the interface between Architectures and Devices

Characteristic	Characteristic Expressions (Beyer, 2024)						
Crossbar Dimensions*	8x8	32x32	64x64	128x128	256x256	512x512	…
Cell Elements	1R (passive CB)	1D1R (active CB)	1T1R (active CB)	…			
CMOS Process Nodes	14 nm	55 nm	65 nm	90 nm	130 nm	150 nm	180 nm	…
Feature Size	0,5 µm x 0,5µm	1 µm x 1µm	…				
Forming Voltage	suitable for the CMOS node, e.g. 130nm/VForm <1.5 V						
Write Voltage	<1 V	<3 V	<5 V	<10 V	>10 V	…	
Write Time - Switching Time	<10 ns	<100 ns	>100 ns	…			
Read Time	<10 ns	>10ns	…				
Write Energy - VoltagexTime	<1 fJ/bit	<10 fJ/bit	<100 fJ/bit	<1 pJ/bit	<10 pJ/bit	>10 pJ/bit	…
Retention Time	<10 years (at 125°C)	>10 years (at 125°C)	…				
Number of States	1-10	10-100	100-1000	…			
I-V-Linearity	none	low	medium	high	…		
Endurance Cycles	10^4	…	10^{15}	…			
Variability	Device-to-Device (D2D) Variability	Cycle-to-Cycle (C2C) Variability	…				
on/off-Ratio	1	10	100	1000	10^4	…	

Definitions and Abbreviations (alphabetical):

- *1R* – single resistive switch in passive crossbar arrays, feature size $4F^2$, easy to fabricate, excellently applicable for 3D integration
- *1T1R* – combination of a transistor (1T) and a resistive switch (1R) in resistive random access memory (ReRAM), feature size 6 − $8F^2$, 3D integration difficult
- *CMOS Process Node* – refers to a specific semiconductor manufacturing process and its design rules
- **Crossbar Dimensions* – crossbar array consisting of n word lines (WL), m bit lines (BL) and n x m memory cells
- *Device Variability* – deviation of the main chosen parameters and figures of merit of the device(s)
- *Endurance Cycles* – ability of a memristive device to sustain a certain number of operational cycles before its memristive states become unstable and difficult to maintain
- *Feature Size* – is determined by the width of the smallest lines that can be patterned in a semiconductor fabrication proces
- *I-V-Linearity* – relationship between the current through an electronic device and the voltage across its terminals (⇒ I-V-characteristic of the device)
- *Number of States* – distinguishable conductance levels on each memristive device
- *on/off-Ratio* – ratio of the on-state and off-state current without any applied gate voltage. High on/off ratio means a low leakage current (= improved device performance)
- *Retention Time* – measures the duration for which memristive states can persist without significant degradation or relaxation
- *Write Time / Switching Time* – time taken by a switch to go from an ON state to an OFF state or vice versa

5.4 Device layer

Simulating neurobiological mechanisms for information processing and storage within memristive devices requires an understanding of quantum mechanical and physical material characteristics (Ziegler, 2024). Scaling some of these materials allows for geometrical dimensions in the magnitude of the matter wavelength of the electronics in memristive devices. This needs the latest technological production methods, interfacial investigations into the atomic scale and appropriate metrology. The following two related sections describe the research, development and application activities in this context.

Section 1 – Parametrization of memristive devices

Established processes and new methods of material analysis and material characterization provide a basic understanding of the chemical and physical mechanisms of memristive devices. In this context it should be noted that while it is possible to model the functionality of most memristive devices on the macroscopic scale, only a limited fundamental understanding is available on the atomic scale, due particularly to the fact that these are complex, coupled chemical and physical processes.

Close dovetailing is required of electronic measurements on the individual device characterized by U/I and C/V measurements, together with automated „on-wafer" measurements, temperature and time measurements, together with analytical processes for characterization and structural investigation of active layers with conventional materials science methods such as AFM, STM, REM, XRD (see chapter 4 for abbreviations). This is necessary to understand the material characteristics so that these can flow into the corresponding technology processes in order to optimize the memristive devices for the subsequent layers of the NMC layer model described here. These parameter measurements are categorized and serve to extend substantiated statistics for optimizing the devices. This approach provides the basis for ongoing research over and beyond the other layers of the NMC model and offers the possibility for transfer into the technological industrial setting.

Section 2 – Modeling and simulating memristive devices

A targeted design of device materials depending on the application layer requires a physical description in accordance with section 1 above.

Models then range from equivalent substitute circuits for the corresponding devices via data-based models through to detailed analytical physical descriptions based on drift-diffusion equations of the material transitions into the respective memristive devices.

Consideration at this point is given to the individual cells (1R but also 1T1R), so that their parameters can be transferred to higher layers. Together with these parameters, the definitions stipulated here should ensure adequate exchange with the subsequent layers for optimization regarding the requirements from the application layer.

Interface 4 <> 5: Device layer <> Material layer

Specification framework
This specification defines uniform parameters for exchange between the device and material layers. These allow for uniform exchange within the overall layer model for optimization but also for comparability of research results and subsequent research. The requirements made of materials and devices result from the overlying layers, provided that the envisioned application is fulfilled. In the interface described here between material systems and NMC devices, technical maturity is crucial for further integration and use of materials available in the CMOS fabrication process. Research and development of devices in the industrial setting should therefore satisfy the following parameters in the framework of the layer model of this VDE SPEC: ■ forming voltage – voltage pulse for forming the memory status ■ write voltage – voltage pulse for writing the memory status ■ write time - switching time – time interval for the write cycle and generally for switching the memristive individual cell ■ read time – time interval for the read cycle and reading the memristive individual cell ■ write energy - voltage x time – energy input for writing the memristive individual cell ■ retention time – parameter for long-term stability of a saved data ■ number of states – stipulation of the memory states of the memristive individual cell ■ current limitation value – stipulation of the current limitation of the memristive individual cell To this end, material parameters have to be categorized according to the following table. Validating the material selection according to maturity, reproducibility and scaling plays a crucial role in rating and integrating the memristive devices. Adapting the parameters provided here depends on the state of research, because it is possible that the definition of this VDE SPEC no longer covers new device concepts, with the need for corresponding adaptation in an updated version.

Table 4 – Characteristics and Characteristic Expressions at the interface between Devices and Materials/Mechanisms

Characteristic	Characteristic Expressions							
Crystallinity	crystalline	poly-crystalline	amorph	…				
Morphology	Shape	Size	Structure	…				
Electronic Properties	Resistivity (10^{-6}-10^{-7} Ωm)	Band Gap (1–3 eV)	Resistance Change (R_{on}/R_{off} >10)					
	Mobilities Electrons (700-8500 cm²/Vs) and Ions	…						
Mechanical Properties	Strength	Stiffness	Elasticity	Plasticity	Ductility	Toughness		
	Brittleness / Malleability	Resilience	Hardness	…				
Photonic Properties	Reflection	Adsorption	Transmission	Refraction	…			
Magnetic Properties	Type of magnetic Material	Saturation Magnetization	Magnetic Anisotropy	…				
Thermal Properties	Thermal Conductivity (0.5–5 W/Kcm)	…						
Technolog. Properties	Process Temperature	CMOS Compatibility	…					
Switching Mechanism/ Devices	ECM	PCM	VCM	ReRAM	FeRAM	FeFET	STT-MRAM	…

Definitions and Abbreviations (alphabetical):

- *ECM* – Electrochemical Metallization; *PCM* – Phase Change Memory; *VCM* – Valence Change Memory; *ReRAM* - Resistive Random-Access Memory; *FeRAM* – Ferroelectric Random-Access Memory; *FeFET* – Ferroelectric Field-Effect Transistor; *STT-MRAM* – Spin-Transfer Torque Magnetic Random-Access Memory

5.5 Materials layer

Selected material compounds of ferroelectrics, metal oxides, chalkogenides, 2D van-der-Waals materials or organic materials for memristive devices that change their conductivity according to the electrical pre-voltage are highly promising candidates for energy-efficient information processing. They are used in research as artificial synapses and neurons in neuromorphic circuits. This is a new approach compared to conventional binary non-volatile memory cells, where memristive devices are already used (Ziegler, 2024). (Wiefels, 2024)

Two technical factors play a crucial role in the further development of memory devices:

- **scaling** the materials from the sub-nanometer scale to macroscopic dimensions of a 300 mm wafer and
- **compatibility** in existing contamination-free CMOS technologies.

The materials for memristive devices must be tunable and reproducible in a defined layer thickness and roughness and satisfy the demands for switching dynamics - transient analysis in this case - to ensured physical switching and reading behavior on all abstraction levels. The multitude of different physical phenomena of the memristive switching dynamics are grouped into electronic effects, ionic effects and structural or magnetic respectively ferroelectric effects.

Some of the materials for memristive devices have several advantageous characteristics, such as fast access time in the magnitude of a few ten picoseconds. These are several orders of magnitude higher than in conventional non-volatile memory cells such as flash memories. Retention can also be modified by choosing suitable materials.

However, these advantages come with a few challenges. One such challenge is the broad distribution of memory states within the individual cell, which is a disadvantage when it comes to evaluating logic operations. The central challenge in the context of in-memory logic operations is the limited precision resulting from signal noise and conductivity drift. Temperature-related conductivity fluctuations can also be a problem. A further challenge is stoichiometric stability during the write pulse, with the possibility of ion migration effects.

The integration of memristive devices in established CMOS technology processes is an essential focus of this section, while not being restricted solely to material selection. The parameters defined here have to meet differing requirements, depending on whether the crossbar array described in chapter 5.3 is on the chip, on the IC, in the package (structure and connection technology) or connected to the IO bus as a peripheral function unit.

Integration and Scaling

As already mentioned in the section on the “Devices” (see chapter 5.4), two key objectives for the optimization of necessary materials for the integration in electronic devices and functional units like x-bar-arrays for the NMC are of high interest the **scaling** and the **increase in energy efficiency**. For STT-MRAM, scaling to 11 nm cells and the realization of 2 Mbit embedded MRAM in 14 nm FinFET CMOS is already demonstrated. However, due to the lower resistance ratio, the readout of magnetic tunnel junctions (MTJs) is more difficult to control technically. Nevertheless, a 64 x 64 MTJ array was recently fabricated in 28 nm CMOS technology. To increase the resistance ratio of MTJs in the future, advances are needed on the material side.

One challenge is to scale ferroelectric devices based on HfO_2, as the thickness of the material to enable 3D capacitance with a 10 nm node and to achieve uniform polarization on the nanoscale of a material is currently not reproducible. The different material phases are the reason. Therefore, the current path is towards ultra-thin layers with the pure ferroelectric orthorhombic phase and without dead layers at the interfaces in order to approach the range below 20 nm for ferroelectric devices on HfO_2.

Phase change devices (PCDs) can be manufactured in the sub-10 nm range. The limiting factor for CMOS-integrated PCDs is the high RESET current required to implement larger access transistors. Commercially available ReRAM cells with conventional geometries have been co-integrated in 28 nm CMOS technology. By using a sidewall technique and nano-thin Pt electrodes, small arrays of 1 nm × 3 nm HfO_2 cells and 3 × 3 arrays of Pt/HfO_2 /TiO_x /Pt cells with a structure size of 2 nm and a half pitch of 6 nm, respectively, were fabricated.

In terms of ultimate scaling, the loss of oxygen to the environment could limit retention times for ReRAM devices scaled below 10 nm. However, filaments as small as 1-2 nm can be stable if stabilized by structural defects such as grain boundaries or dislocations. Therefore, finding a material solution for limiting oxygen vacancies to the nanoscale could ensure the required retention for devices in the size range of a few nm.

Latency and Time Dynamics

In addition to increasing the access time, increasing parallelization is an important optimization factor, although it is not primarily about ever higher clock frequencies. Nevertheless, it makes sense to determine the ultimate speed limits of NVM concepts in order to estimate the maximum learning rates and investigate the effects of short spiking stimulations. In addition, novel computing concepts, as discussed in layer *algorithms* (chapter 5.2), could benefit from higher access times. For MRAM, reliable switching at 250 ps was demonstrated by using double spin-torque MTJs consisting of two reference layers, a tunnel barrier and a non-magnetic spacer.

Memory arrays based on FeRAM can successfully switch with 14 ns at 2.5 V. Ferroelectric field-effect transistors (FeFETs) have been shown to switch with pulses <50 ns in 1 Mbit memory arrays. PCM devices can be switched with pulses <10 ns. In general, their speed is limited by the crystallization time of the material. Using Ge_x-Sn_y-Te samples as an example, it was shown that this time can be set in a wide range from 25 ns to 10 ms by adjusting the material composition. This offers great potential for adapting the operating time of an NMC system to the respective application. SET and RESET circuits with 50 ps and 400 ps were demonstrated for VCM ReRAM. Both have so far been limited by extrinsic effects and component errors rather than intrinsic physical rate limiting steps.

Endurance

All memristive memory devices have a limited lifetime because the storage is based on the physical displacement of atoms, just like ReRAM, PCM and ferroelectric effects in these devices. For silicon-based FeFETs, the lifetime is usually in the order of 10^5, which is mainly due to a dielectric breakdown in the SiO_2 at the Si-HfO_2 interface. As for the lifetime of VCM ReRAM, convincing statistics have shown that $>10^6$ cycles are realistic.

In some reports, maximum cycle numbers of more than 10^{10} cycles have been reached. Depending on the material system, different mechanisms of failure for endurance are discussed. The microstructure of the switching material could deteriorate or be irreversibly penetrated by metal atoms. In VCM ReRAM, excessive generation of oxygen vacancies was discussed as an endurance limiting factor. Novel

material solutions that limit the ions to the intended radius of action could be a way to increase the lifetime of ReRAM devices. For PCM, it has been proposed to implement multi-PCM synapses. Arbitration across multiple memory elements could circumvent both the endurance and variability issues.

A typical limiting factor with regard to the reliable operation of ferroelectric memories is the “wake-up effect”, which causes an increase in polarization after a few cycles, and fatigue, which leads to a decrease in polarization at high numbers of cycles. Both are caused by the displacement of defects such as oxygen gaps and are needed to be addressed in future.

Retention

After training, the state of the non-volatile memory device must be stable at an operating temperature of 85 °C for about 10 years. However, these requirements are highly dependent on the application environment of the NMC system. From a thermodynamic point of view, the states in ferroelectric or ferromagnetic memories could both be stable. In contrast, ReRAM and PCM devices store information in the form of the allocation of atoms, where both LRS and HRS are metastable states and the storage is determined by material parameters such as the diffusion coefficient of the respective species. Here, degradation is not a digital flipping of states, but a gradual process. In PCM, the drift of the resistive state is caused by the structural relaxation of the melt-quenched amorphous phase.

Apart from a drifting of the state, a broadening of the programmed state distribution (e.g. resistance) is typically observed for ReRAM. Furthermore, since analog or multi-level programming is highly relevant for NMC, it should be considered that intermediate resistance states might have a reduced retention compared to the edge cases of high and low resistive states as demonstrated for PCM devices.

Read Disturb

During inference, frequent reading of the memory elements is required, which should not change the learned state. In a bipolar ReRAM memory, read disturbance in the HRS/LRS occurs mainly when reading with SET/RESET polarity, as the read disturbance can be seen as extrapolation of SET/RESET kinetics to lower voltages. Nevertheless, the HRS state in bipolar filamentary VCMs has been shown to be stable by extrapolation over years at read voltages up to 350 mV.

Variability and Adjustment of the memory states

The variability is particularly pronounced in systems based on the stochastic movement and redistribution of atoms, such as ReRAM and PCM. Here the variability from device to device (D2D), from cycle to cycle (C2C) and even from one read operation to the next (R2R) must be distinguished. The D2D variability can be kept comparatively low by optimizing the manufacturing processes. In contrast, the C2C variability of filamentary resistive ReAM and PCM can be significant due to the random nature of filament and crystal growth. With intelligent programming algorithms, however, the C2C variability can be very well reduced to a minimum.

However, R2R fluctuations persist in the form of readout noise in filamentary VCMs. They are usually attributed to the activation and deactivation of traps or the random redistribution of defects and depend strongly on the material. In PCM, the R2R fluctuations are caused by 1/f noise and temperature-induced resistance fluctuations. One approach to solve these problems as well as the drift is to use the so-called projected phase change memory with a non-isolating projection segment in parallel to the PCM segment.

Although variability is a challenge for the defined applications, it is possible to design NMC systems to take advantage of it.

For most computing concepts described in the section layer Algorithms (see chapter 5.2), operation with binary memory devices is severely limited, and the ability to set multiple states opens up a new complexity space. For devices with thermodynamically stable states such as ferroelectric or magnetic memories, the intermediate states depend on the presence of domains. As a result, the performance strongly depends on the specific domain structure, and the scaling may be limited by the size of the domains. Nevertheless, multi-level switching has been demonstrated for both FTJ and FeRAM.

Although there is no direct link between the *application* layer (see chapter 5.1) in the model presented, the parameters in the interfaces are nevertheless decisive for the functionality and fulfilment of the requirements by the specified and defined application. The selection and use therefore has an implicit effect on the design of the overall NMC system. This is the overriding objective of this VDE SPEC.

6 Literature and sources

Beyer, S. W. (2024). Wenger, C., Ziegler, M. *Information from GlobalFoundries, IHP (Frankfurt/O.) and Kiel University.*

Dhar, P. (2020). *The carbon impact of artificial intelligence. Nat. Mach. Intell., 2(8), 423-425.*

Gemmeke, T. (2024). *Information from the department for integrated digital systems and circuit design (IDS) at RWTH Aachen.*

Jones, N. (2018). *How to stop data centres from gobbling up the world's electrici-ty. Nature, 561(7722), 163-167.*

Leupers, R. (2024). *Information from the Institute for Communication Technologies and Embedded Systems (ICE) at RWTH Aachen.*

Nielen, L. (2023). *aixACCT Systems, data sheet RS/FE Memory Analyzer - fully integrated, modular, semi-automatic characterization system for resistive and ferroelectric memories.*

Waser, R. (2019), Dittmann, R., Menzel, S., & Noll, T. *Introduction to new memory paradigms: memristive phenomena and neuromorphic applications. Faraday discussions, 213, 11-27.*

Wiefels, S. D. (2024), Dittmann, R. *Materials challenges and perspectives, in: Roadmap to Neu-romorphic Computing with Emerging Technologies, Page 40, arXiv:2407.02353.*

Ziegler, M. (2020). *Novel hardware and concepts for unconventional computing. Scientific reports, 10(1), 1-3.*

Ziegler, M. (2024). *Information from the Chair of Energy Materials and Devices Department of Materials Science, Kiel University.*

7 Committees

DKE/K 631: Semiconductor devices

DKE/K 682: Structure and connection technology for electronic assemblies

GMM Technical Group 1.1.4: Testing equipment and methods

VDE Verband der Elektrotechnik
Elektronik Informationstechnik e.V.

Merianstraße 28
63069 Offenbach am Main
Germany
Phone +49 69 6308-0
service@vde.com
www.vde.com

VDE