Delft University of Technology

Challenges and Solutions in Emerging Memory Testing

Vatajelu, E.I.; Prinetto, Paolo; Taouil, Mottaqiallah; Hamdioui, Said

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Challenges and Solutions in Emerging Memory Testing

Elena Ioana Vatajelu, Member, IEEE, Paolo Prinetto, Senior Member, IEEE, Mottaqiallah Taouil, Member, IEEE, and Said Hamdioui, Senior Member, IEEE

Abstract— The research and prototyping of new memory technologies are getting a lot of attention in order to enable new (computer) architectures and provide new opportunities for today's and future applications. Delivering high quality and reliability products was and will remain a crucial step in the introduction of new technologies. Therefore, appropriate fault modelling, test development and design for testability (DfT) is needed. This paper overviews and discusses the challenges and the emerging solutions in testing three classes of memories: 3D stacked memories, Resistive memories and Spin-Transfer-Torque Magnetic memories. Defects mechanisms, fault models, and emerging test solutions will be discussed.

Index Terms— Emerging Memories, 3D-SIC, Resistive RAM, Magnetic RAM, Defects, Fault Models, Test.

——————————— ◆ ———————————

1 INTRODUCTION

SEMICONDUCTOR memories have been integral and critical components of all computing systems, including sensors, desktops, servers, etc. [1]. Their function has been evolving over time; e.g., SRAM for primary memory, DRAM for secondary or main memory, and Flash for mass-storage. Recent application (such as data-intensive applications) and architecture (such as multi-core, GP-GPUs) trends do not only exacerbate the old requirements, but have additional requirements for the memories and the memory systems (e.g., higher bandwidth, higher density, lower power, sustainable scaling, lower cost, lower latency) [2]. Conventional memories such as SRAM, DRAM, and Flash are unlikely to satisfy all the requirements. Cell scaling, which has sustainably supported cost reduction for decades, is facing major challenges and is expected to end soon [3]. Hence, a lot of effort is put on searching and developing new memory alternatives; examples are 3D stacked memories [4, 5, 6], resistive memories [7, 9], etc. Not all emerging memories are brand new ideas; many of them are based on concepts and physical mechanisms that have been known for years, but have not received enough attention for several reasons, such as immature technology, no market demand, limited resources that are used on more rewarding technologies, etc. For example, the memristor as a resistive device was invented in 1971 by Leon Chua and just started getting attention in 2008 after HP prototyped the device [7].

As already mentioned, the limitations of conventional memories (scalability, cost, reduced noise margins, bandwidth) and the requirements of today's applications are forcing the community to urgently find new alternative memory technologies. There is a wide variety of new emerging memories [3], both volatile and non-volatile. Among the volatile memory technologies that have as target to replace or compete with CMOS SRAM and DRAM, there are the corbon-nanotube based SRAMs (CN-SRAM), zero-capacitor RAM (Z-RAM), tyristor RAM (T-RAM). In addition, there are several proposal of non-volatile emerging memory technology, such as Magnetic Random Access Memory (MRAM), Ferroelectric Random Access Memory (FeRAM), Phase Change Memory (PCM), Resistive Random Access Memory (RRAM), three-dimensional, crosspoint and quantum dot memories, memories based on deoxyribonucleic acid, organic and polymer materials. Some of these emerging memories are already in early stage production, such as FeRAM, MRAM and PCM, while others are still too immature to be considered for mass fabrication, such as DNA and quantum dot [8]. Only two of them seem to enter the mainstream market on the short-term, if not already, probably starting with niche applications. These are: (a) resistive memories (RRAM) [9,10] which have the potential of providing order of magnitude lower latency and exponentially greater endurance than NAND flash, and high potential to replace DRAMs; (b) Spin-Transfer-Torque Magnetic RAMs (STT-MRAM) [11,12] which are viable alternatives for the replacement of DRAM and low level cache SRAM due to the programming speed, endurance and non-volatility. Together with the introduction of new technologies, a new integration paradigm is introduced to improve the quality of today's memories, i.e., the 3D integration. In a 3D stacked memory [4] the power consumption and memory bottleneck in computer systems are reduced by acheiving a wider bandwidth and using short vertical interconnects referred to as Through-Silicon Vias (TSVs). In 2015, Micron has introduced the 3D XPoint memory, which is a transistorless crosspoint architecture which incorporates the benefits of emerging

————————————————

- *E. I. Vatajelu is with Univ. Grenoble Alps, TIMA Laboratory, Grenoble, France, E-mail: ioana.vatajelu@univ-grenoble-alpes.fr.*
- *P. Prinetto is with CINI Cybersecurity National Lab & Politecnico di Torino, Italy, E-mail: paolo.prinetto@polito.it.*
- *M. Taouil and S. Hamdioui are with Delft University of Technology, Delft, the Netherlands, E-mail: {m.taouil, s.hamdioui} @tudelft.nl*

memory technology and 3D integration [128].

In this work we focus on the potential of the 3D stacked memories, RRAM, and STT-MRAM to work as primary and secondary memory, and on the fault modelling and test challenges they have to face.

The remainder of this paper is organized as follows. Sections 2, 3 and 4 describe the working principle, state-of-the-art, defect mechanisms, fault models and test solutions for 3D-Stacked ICs (3D-SICs), RRAMs and MRAMs, respectively. Finally, Section 5 concludes this paper.

2   3D-STACKED-ICs

2.1 Working PrInciple and Classification

Working principle: In *3D Die Stacking*, different manufactured tiers can be stacked and bonded to other tiers using a direct communication link between vertically adjacent tiers. A 3D-SIC consists of two or more dies stacked in the vertical direction. The interconnection between the dies can be implemented physically by micro-bumps and/or TSVs, or via contactless communication based on capacitive [13,14] or inductive coupling [15,16]. Among the interconnection schemes, TSVs are the most promising, especially for the power network due to unstable power delivery [17].

Figure 1 depicts a two-layer 3D-SIC with a face-to-back stacking configuration. Compared to off-chip wire-bonds, TSVs enable extremely short connections as they go straight through the substrate of the dies. Between the stacked dies, micro-bumps are used to connect the TSVs from Die 2 to Die 1. TSV-based 3D-SICs can be used to empower More-Moore and More-than-Moore systems and have considerable advantages over planar ICs and SiPs, such as high-speed, low power consumption, small form factor, and heterogeneous integration [18-21].

2.5D-Stacked ICs (2.5D-SICs) are a special class of 3D Die Stacking are in which two or more active dies are stacked side by side Face-to-Face (F2F) on a large passive silicon interposer. The interposer is only used to connect the active dies by means of TSVs and wires. 2.5D- SICs are in general easier to manufacture, but their advantages are also typically less than those of 3D-SICs (e.g., interconnect power dissipation, bandwidth, off-chip I/O density) [22].
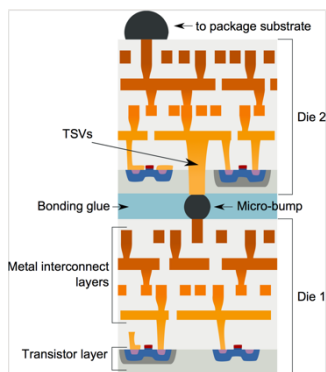

Fig. 1. TSV-based 3D die stacking

Classification: Partitioning memories across multiple device layers of a 3D-SIC can take place at different granularities, resulting in three different architectures. A top to bottom perspective is presented in the following [23].

1) **Stacked banks** - The coarsest granularity partitioning of memory takes place at the bank level, by stacking banks on top of each other. Each bank consists of a complete memory system (i.e., memory cell array, address decoder, write drivers, etc.). An overall reduction in wire length is obtained (about 50% for certain configurations), resulting into significant reduction in both power and delay [24, 25]. A 3D manufactured DRAM based on the stacking of banks manufactured by Samsung is described in [4].

2) **Cell arrays stacked on logic** - This approach, in contrast to the previous one, separates the peripheral logic (row decoders, sense amplifier, column select logic, etc), from the cell arrays. The peripheral logic is placed on the bottom layer, while the cell array is split across one or multiple layers. This is considered to be the *true 3D memory* [24]. Research in this area has been performed for both SRAMs [24, 26] and DRAMs [19, 27]. By using this separation method, the peripheral logic can be independently optimized for speed, while the cell arrays can be arranged to meet different criteria (density, footprint, thermal, etc). Instances of 3D-DRAMs based on *cell arrays stacked on logic* have been manufactured by NEC Electronics, Elpida Memory [5], Tezzaron [6], Samsung [28] and SK Hynix [29]. The array layer can be further classified in:
   a) **divided-columns:** in which bitlines are split and mapped onto different layers;
   b) **divided-rows:** in which wordlines are split and mapped onto different layers.
   Both organizations reduce latency and power due to reduced wordline/bitline lengths.

3) **Intra-cell (bit) partitioning** - Here, memory cells are split among one or more layers. At this fine granularity level, the relative small size of the cell and the size of the TSV make the splitting across layers a difficult task [26].

2.2 Oportunities and Challenges

Opportunities: A key condition to shift from the design and prototype phase to large-scale production is a manageable cost figure [42]. 3D-SICs are able to reduce cost by splitting up large dies over multiple smaller layers. A benefit of this approach is that the compound yield of the 3D-SIC with smaller die sizes may exceed the yield of the single large die [43]. Another way to reduce cost in 3D-SICs is by integrating multiple stand-alone chips. For example, the bandwidth is significantly improved by stacking DRAM on logic. In addition, vertically stacking reduces the footprint, the volume, and the weight of the memory device, which in turn increases the package density. In spite of these major advantages, cost is still a limiting factor for wide acceptance of 3D-SICs, as it depends on the yield lerning curve driven by the cumulative produced 3D-SICs.

Utilizing the third dimension might be the only way to significantly reduce memory latency and power consumption for future generations of multi-core microprocessors [20]. Stacking provides additional benefits such as reduced power consumption (up to 50% and 25% for standby and active power respectively for four stacked memory dies) [4], reduced noise levels due to the shorter global interconnects and the need of smaller I/O drivers [44]. In general, any efficient partitioning of IP cores reduces long global wires and hence the delay and power dissipation [17, 45]. However, special care must be taken to maintain a stable clock and power distribution [46-48].

Another benefit of the 3D stacked memories is vertical redundancy. Traditionally, yield improvement for 2D memories is based on the use of spare rows and/or columns [49-51]. 3D stacked memories provide additional repair features in the vertical dimension as spares can be accessed on neighbor dies. Preliminary research results show the significant benefits of using this vertical direction [21, 23, 52]. However, TSVs need to scale by at least one order of magnitude to make such schemes viable [53]. Other research publications analyzed the impact of TSV redundancy schemes [54-56]; however, they typically come with a high area overhead.

Challenges: 3D-SIC manufacturing requires additional processing steps as compared to conventional ICs; these include the forming of TSVs, thinning wafers, and stacking and bonding wafers or dies. Each of these additional steps may introduce new defects to the system. Example of defects will be discussed in more detail in 2.3. Testing for these defects is one of the biggest challenges of 3D-SICs, and it will be discussed in detail in 2.5.

2.3 Defects

A 3D-SIC consists of multiple stacked interconnected dies. Both the dies and the interconnects are susceptible to defects during wafer manufacturing, stacking, and packaging. In essence, we distinguish between three defect sources:

1) Wafer manufacturing
2) Stacking defects
3) Assembly and packaging defects

In this paper, we focus only on defects related to stacking, since they are the only defects that are specific to 3D-SICs. These stacking defects can be categorized into defects related to the interconnect (TSVs and micro-bumps), and defects related to the die. Example defects of both categories are summarized below. Interconnect defects include:

- Pinhole defects along TSV walls create shorts or low resistance paths between TSVs and the substrate. This causes degradation of the signal quality in terms of strength and speed [59, 63-65].
- An incomplete fill of TSVs (voids) may originate from insufficient wetting during plating. Voids cause partial opens and increase resistance [59, 63-65].
- Coefficient of thermal expansion (CTE) mismatch between TSV metal (e.g., copper) and substrate may lead to TSV cracks and sidewall delamination. Both lead to increased path resistance [64-68].

- Pinch-off of TSVs during plating could lead to increased TSV resistance or partial opens [63].
- Missing contacts between TSVs and transistors or metal layers cause opens [63, 69].
- A misalignment of TSVs and μ-bumps increase the resistance and cause (partial) opens [63-65].
- Crosstalk between different TSVs [65, 70].
- Damage in underlying BEOL [71].
- Weak bonding due to buckled thinned Si chip [71].
- Variation in TSV heights may cause tin to be squeezed out from μ-bump causing shorts between μ-bumps [57, 71].
- Electromigration causes voids and cracks in the joints, resulting in higher resistive μ-bumps, or opens [72].
- Cracks in μ-bumps may be formed due to a CTE mismatch between copper, silicon, and silicon-oxide [63].

In the literature, no new types of defects are reported for the dies. However, stacking might impact the parametric yield. For example, the mechanical stress induced by TSVs might impact (negatively or positively) the transistor speed [73]. In addition, thinning of dies leads to a shift in the transistor current-voltage (I-V) characteristic, impacting both speed and power [74].

2.4 Fault Models

The faults that may occur due to defects in traditional planar dies are well known and classified [75]. Of interest to this work are the defects that occur in the 3D interconnects and their corresponding interconnect faults [76]. The interconnect fault classification is depicted in Fig. 2, where the faults are grouped into static and dynamic faults, including stuck-at faults (SAF), bridge faults, path delay faults (PDF), stuck-open faults (SOF), and crosstalk faults. The authors concluded that dynamic faults embody most defects and therefore it is essential to test for them.

2.5 Test and Design-for-Test

Testing is one of the biggest challenges of 3D-SICs due to its number of potential test phases. Figure 3 shows the conventional 2D test flow for planar wafers [57, 58]; it consists of two test phases: a wafer test before packaging and a final test after packaging. The 2.5D/3D-SICs, however, require additional test phases. In general, four test phases can be distinguished for a 3D-SICs consisting of n dies, as depicted in Fig. 2(b): (1) n *pre-bond* wafer tests, (2) n-2 *mid-bond* tests, (3) one *post-bond* test before packaging and (4) one *final* test. This results into 2·n test phases [59].
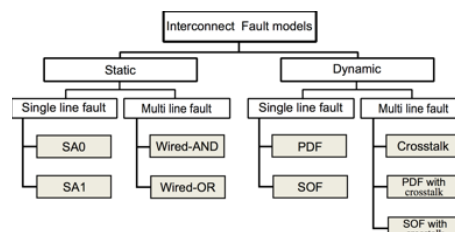

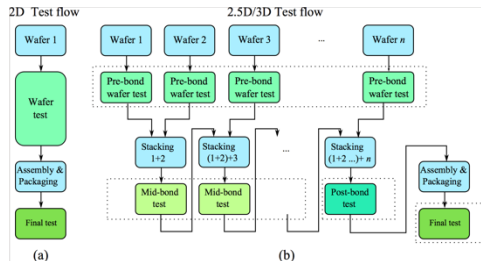Fig. 2. Fault Model Clasification for Interconnects [76]

Fig. 3. 2D versus 2.5D/3D D2W test flows.

The test challenges can be sub-divided into two main categories: (1) test access and (2) test flow optimization.

1) **Test Access:** The test access can be divided further into two subcategories: external and internal access (or Design-for-Test, i.e., DfT architecture). Assuming that only bottom dies have external access, the pre-bond testing of non-bottom wafers comes with extra challenges as they are not designed with external I/O pins. One option to access these wafers is to use dedicated test pads which add area overhead and undesired load capacitance in the final stack. Performing direct probing on the micro-bumps [60] would make the probe pads superfluous, however, the manufacturing of a fine pitch probe card is challenging. In addition, probing dies that are already thinned may lead to serious IC damage [61].

2) **Test Flow Optimization:** Test flow optimization can also be divided into two subcategories: test content and test order. Each test covers a set of faults (which are higher abstract representations of defects). Generating test patterns may follow a similar flow as used in traditional 2D. However, defects can be introduced into the stack during different phases. In addition, the newly introduced components, i.e., the TSV based interconnects, should also be tested. One of the main challenges is to test the interconnects at-speed due to their low latency. Once the test content is defined, one should determine the order in which interconnects and dies are tested and during which phases. Early testing might prevent further assembly costs, such as the stacking of good dies on defective partial stacks, but may also impact the cost negatively by having a test overkill. The number of test phases further complicates finding the optimal test flow [62].

As 3D-SICs are quickly gaining more ground, the need for a standardized test becomes more important. Several solutions have been proposed [77-81], but with many limitations, such as the ability to test the dies or interconnects separately. However, one of the most promising standards under development is the IEEE P1838 [82], which focuses on the dies as key components in the stack. The stack-level architecture routes both data and control signals up and down through the stack (TestTurns and TestElevators) to reach each particular die in the stack. The architecture supports both intra-die test (INTEST) and inter-die test (EXTEST) during all test phases as depicted in Fig. 4. In the pre-bond phase, dedicated pads can be used to test dies. In the mid-bond and post-bond

phases, both dies in a partial and complete stack respectively can be tested (INTEST) with this test architecture. EXTESTs can be performed for the interconnects during mid-bond and post-bond and are be based on the IEEE 1149.1 (JTAG) [83] and IEEE 1500 [84] standards. The final test (post-packaging) consists of the same tests options. In [85], the authors showed how to extend such an architecture to test JEDEC Wide-I/O memory. In [86] and [87] this architecture was used to perform at speed interconnect testing in the presence of "shore logic", i.e., logic outside the die wrapper boundary register.

With respect to testing TSVs and interconnects, several schemes have been proposed. For the sake of brevity, we focus on post-bond testing. In the past, (DRAM) memory vendors were typically not in favor of integrating JTAG on their devices [88]. Besides JTAG, other solutions have been proposed to test the interconnects. In [89] and [90] authors present hardwired BISTs with at-speed testing capability for crosstal faults. These methods have several drawbacks, such as lacking flexibility to alter test patterns, require DfT area overhead of approximately 8%, and can handle only uni-directional lines. In [91], a Memory Based Interconnect Test (MBIT) methodology is presented, used to test and diagnose defective interconnects by using the CPU to write and read from the memory. This approach does not require any DfT as it entirely reuses existing components in the stack. In addition, it supports at-speed testing and detects static and dynamic faults. An additional benefit is that it is very flexible in altering test patterns simply by modifying software instructions and also has an extremely short test execution time.
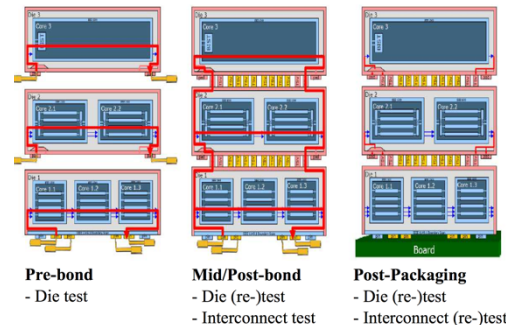

Fig. 4. IEEE P1838 [82].

3 RRAM

3.1 Working Principle and Classification

Working principle: The Resistive Random Access Memory (RRAM) is a non-volatile RAM. Its data storage element is a two terminal device that switches between resistive states, i.e., the high resistance state (HRS) and the low resistance state (LRS), when triggered by an electrical input. The resistive storage is a three-layer device, consisting of a dielectric sandwiched between two metal electrodes (as in Fig. 5a). There are many materials which can be used for the electrodes and dielectric, but the underling operation principle remains the same. RRAM relies on the formation (corresponding to low resistance) and

the rupture (corresponding to high resistance) of conductive paths in the dielectric layer. Once the conduction path is formed, it may be RESET (the path broken, transition from LRS to HRS) or SET (the path re-formed, transition from HRS to LRS), as shown in Fig. 5a. Usually, the samples right after fabrication, i.e., the pristine samples, have a very high electrical resistance (approximately 1GΩ) and a large voltage is required for the first SET operation, also known as the forming process; this drastically reduces the device resistance (to about 10KΩ) triggering the switching behaviour in the subsequent cycles.

A common architecture for high density resistive devices (mainly used for data storage) is the crossbar array (Fig. 6a). It consists of two sets of perpendicular nanowires with resistive devices located at intersection points. One set of parallel wires is used as bit-lines, while the other as word lines. Crossbar architectures suffer from sneak paths (i.e, unintended electrical paths within the circuit) which may affect read and write operations.
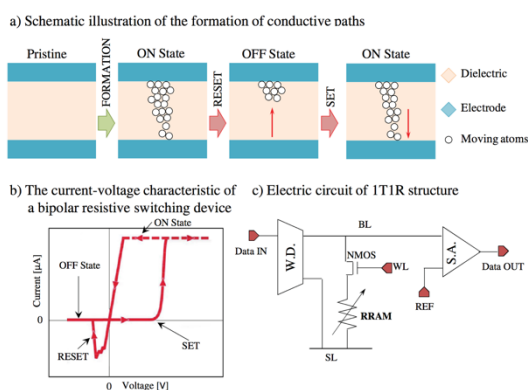

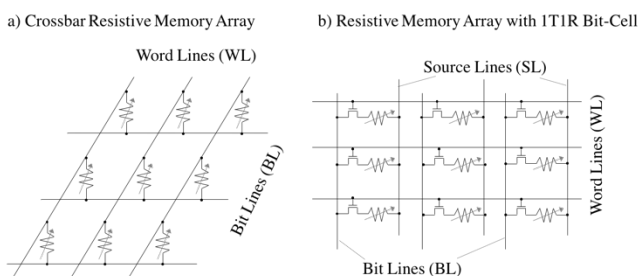Fig. 5. Operation Principle of generic Resistive RAM

Fig. 6. Resistive Memory Array Architectures

A typical architecture for embedded RRAM is based on the *1T1R memory cell* (Fig. 6b) which consists of a resistive storage device (1R) and an access device, typically an NMOS transistor (1T), as shown in Fig. 5c (here W.D. is the Write Driver, S.A. is the Sense Amplifier, REF is the reference current during read, BL is the Bit Line, SL is the Source Line and WL is the Word Line). To read the data from a RRAM cell, a small bias voltage is applied to detect whether the cell is in low or high resistive state. The decision is taken by a sense amplifier, which compares the current passing through the device against a reference current. The write operation is performed by the write driver and uses one of two possible RRAM switching modes: unipolar switching, which depends solely on the amplitude of the applied voltage and not its polarity, i.e.,

SET and RESET are controlled by the same polarity; and bipolar switching (illustrated in Fig. 5b) in which SET and RESET are controlled by reverse polarities.

Classification: There exist a large variety of resistive memory technologies. They can be classified by the dominant physical switching mechanism in: Phase Change Memories, Electrostatic/ Electronic Effects Memories, and Redox Memories. Various resistive switching mechanisms have been proposed to efficiently perform the SET and RESET operations. They include the formation and rupture of conductive paths, charge trapping, electrode- limited conduction [30, 31]. The low-resistance path can be either localized (filamentary) or homogeneous.

One of the most versatile resistive memories is the Redox RAM [3, 32], where the RESET and SET processes, breakdown and regrow of the conductive filaments, involve oxidation and reduction (i.e., redox reaction). These are Metal-Insulator-Metal (MIM) structures, in which the switching mechanism is electrochemical and it can occure in the insulator-layer, or at the insulator-layer/metal contact interfaces. The MIM structures can be classified by their underlying switching mechanism as follows [33]:

1) **The Valence Change Mechanism (VCM):** relies on the fact that the dielectric layer can act as an electrolyte. The migration of oxygen vacancies within the applied electric field exhibits a bipolar operation. The mobile species, contributing to conductive path formation, are the oxygen anions (positively charged oxygen vacancies). The band diagram features an electrostatic barrier which defines the electric current. The SET operation is performed by applying a negative bias voltage, which causes a local redox reaction, therefore an increase in conductivity. The RESET is performed by reversing the bias polarity and allowing the recombination of oxygen. The most common examples of VCM RRAMs use TaOx, HfOx and TiOx [34, 35] devices.

2) **The Electrochemical Mechanism (ECM):** relies on an electrochemically active electrode metal such as Ag or Cu. The mobile metal cations drift in the ion conducting layer and discharge at the counter-electrode, leading to a growth of conductive metallic filaments in the isolation layer - i.e., the SET mechanism. The RESET mechanism is performed by reversing the polarity of the applied voltage, resulting in the electrochemical dissolution of the conductive filaments [36].

3) **The Thermochemical Mechanism (TCM):** relies on a filament modification due to Joule heating. Conductive filaments, composed of the electrode metal transported into the insulator, are formed during the forming process prior to memory cyclic switching. The SET operation is achieved by Joule heating; it triggers local redox reactions that facilitate the formation of oxygen deficient ions and metallic filaments. The RESET operation is a thermally activated process resulting in a local decrease of the metallic species. TCMs are unipolar switching devices. NiO has emerged as the reference material for resistive switching based on the TCM [37].

As a reasonably representative example, in the subsequent sections, the focus will be on HfOx-based VCM RRAMs, as they seem the most promising. Note that the focus of the paper is on device test where the quality of the conductive path formation is relevant, regardless of the physical mechanism.

3.2 Opportunities and Challenges

Opportunities: Emerging memory technologies are on the way to revolutionary change the classical memory/storage architectures. There are several emergent memory technologies that attempt to address the technical challenges and constraints faced by today's memories. The new memories should meet the high demands of tomorrow applications, like high performance and high density, good endurance, small devices sizes, good integration, low power profile, resistance to radiation, and ability to scale below 20 nm [92, 93]. One of the most promising emerging memory is the Resistive RAM.

Resistive RAM is considered as one of the strong candidates to replace Flash memory due to their potential advantages such as the high storage density and 3D packaging (allowing layers of memory devices to be integrated in one chip), fast switching, low energy consumption per switching cycle, and compatibility with the current silicon fabrication process.

The simple device structure (metal-insulator-metal) of a RRAM device, its compatibility with CMOS process, the scaling opportunities below 8nm, its large on/off ratio, and fast operating speed make the RRAM devices ideal candidates to eventually be used as embedded memories.

Challenges: Amongst the greatest challenges faced by today's RRAM devices is their relatively low endurance ($10^5 - 10^{10}$ cycles [118]) and poor uniformity. The low endurance limits their efficiency as embedded memories, while the poor uniformity causes extreme variability and limited reproducibility.

Another challenge is the large number of new materials (and combinations of materials) which can be used for the resistive stack formation. This makes standardization of the fabrication process hard. The introduction of new materials in RRAM fabrication does not give enough time to collect and generate the required data to guarantee a sufficient yield; the process integration task is often not supported by consolidated knowledge. These issues, which are common to all emerging technologies, introduce aggressive challenges on defect and fault modelling and possible test solutions.

3.3 Defects

The resistive memory fabrication is performed using two processing steps: the standard process and a non-standard process for the resistive stack deposition at the back-end-of-line [94]. This assures that there is no interference with the logic process, however *front-end contamination* may arise. The resistive stack is deposited at higher metal layers, usually between layers M4 and M5 or M3 and M4, as shown in Fig. 7a [95, 96].

The standard CMOS fabrication process might introduce defects which often are caused by impurity deposi-

tions. These defects behave as resistive defects at the electrical level. At cell level, they could lead to resistive open defects in the metal lines connecting the NMOS transistor source, gate and drain (Df1, Df2 and Df3 in Fig. 7b). In essence, these resistive defects model the lumped effect of broken or irregular shaped metal lines, narrow, cracked or non-existent vias, and dust particles deposited between the layers impeding proper electric conductivity.

After the CMOS is fabricated, the resistive layers are deposited. Chemical and physical conditions during the bottom electrode deposition can affect the composition and the microstructure of the deposited thin film, and imprint residual stresses [99], which in turn affect the quality of the forming process (the amplitude of the signal required for the forming process), and consequently, the value of LSR. In extremis, this effect can prevent the forming process entirely, which results in an open circuit-like behaviour. The subsequent polishing process could leave the metal surface rough, leading to large resistance variations. The deposition of the resistive switching material is vulnerable to various problems related to precursors and cleaners; therefore, it can cause defects such as thik or thin localized spots [97, 98]. A deficient capping layer deposition can lead to large variations in the characteristics of the forming process and in the efficiency of the switching process. Moreover, the top electrode deposition might induce parameter variations and defects. The last step, the pillar eching, targets the achievement of steep edges of the resistive stack and to control the device critical dimensions. Improper etching causes wide resistance variations and resistive defects (shunt and contact). These defects can be modelled by Df4 and Df5 in Fig. 7b.
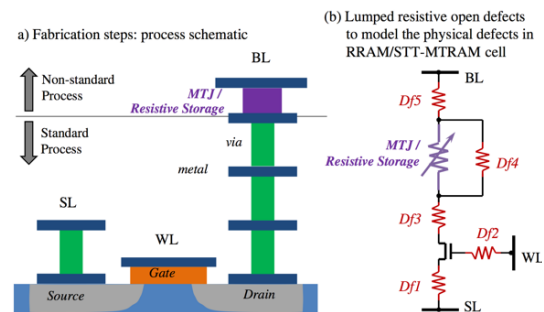
Fig. 7. RRAM/STT-MRAM bit-cell

There are several works in the literature dealing with resistive defects, such as resistive opens, resistive shorts and bridges [101, 102, 103, 104, 105, 106], as well as defects leading to large parameter variability [107, 108, 109, 110]. The next subsection describes the fault models which can be extrapolated based on the described possible defects.

3.4 Fault Models

Defect based analysis requires (i) to simulate the defective device under a significant set of input stimuli (i.e., sequences of read and write operations) able to sensitize faulty behaviour, (ii) to observe the behaviour of the memory output in response to the input stimuli, and (iii) to classify the observed faulty behaviours in a set of high level functional fault models.

Since the resistive devices are analog in nature, the fault modelling can be performed in a similar manner as for most analog devices. The faults can be classified in two main categories: (i) *catastrophic (hard) faults*: the component is opened or shorted, (ii) *parametric (soft) faults*: the defects shift the resistance value outside the tolerated boundaries.

Memory functional fault models have been deeply studied in the literature, and this survey focuses on those faults observed as a consequence of RRAM specific defects:

– **Transition Fault (TF):** the cell fails to undergo a down-transition (up-transition) when write 0, i.e., the RESET operation (write 1, i.e., the SET operation) is performed. These faults caused by the resistive defects (Df2, Df3 and Df4 in Fig. 7b) are mainly hard faults [104].

– **Stuck-at-Fault (SAF):** the cell is always in LRS (Stuck-at-0) or in HRS (Stuck-at-1). The cell will act as a Stuck-at-0 (hard fault) when these faults are caused by the presence of large resistive open defects on the word line (Df2), a permanent open switch (the state of the access transistor is stuck-at OFF), the presence of resistive defect Df4 large enough to shunt the resistive device (Fig. 7b). SAFs can also be soft faults, when they are caused by an incorrect forming process. If the cell is over-formed, it behaves as a Stuck-at-LRS (Stuck-at-0) meaning that the LRS value is lower than its nominal value, and the limited strength of the write driver may fail to complete the RESET operation. When the forming operation fails completely, the resistive switching is not activated, and the cell behaves as Stuck-at-HRS (Stuck-at-1) [102, 104].

– **Write Disturbance Fault (WDF):** these faults are coupling-like faults and are caused by a defective transistor. If the access transistor is stuck-at-ON, a writing operation on a cell (aggressor) sharing the bit or source line with the victim cell can result in an unintentional write to the victim cell. If this fault occurs in 1 cycle, the fault is called static (WDF), if it requires several consecutive cycles it is called dynamic (dWDF) [104].

– **Undefined Write Fault (UWF):** the cell is set to an undefined state by a write 1 (0) operation; the stored data corresponds to an arbitrary logic value. These faults occur in weak cells, i.e., when LRS (HRS) is smaller (larger) than the nominal value. In this situation, the device remains in an *intermediary* state, causing a random logic value to be read from the defective RRAM cells [102] (effect supported also by cycle-to-cycle variation of the RRAM resitive levels). This fault can be observed in the fresh cell, caused by extreme process variation, or in the aged cell, due to resistance value shifting over time.

– **Slow Write Fault (SWF):** the cell fails to undergo a write 0 (write 1) operation in the allotted time. These faults can be hard faults when caused by the presence of small resistive defects in the memory cell at locations Df2, Df3, and Df4 in Fig. 7b, or soft faults when caused by a weak access transistor, improper capping layer deposition or improper stack etching, which affects the efficiency of the state transition (SET or RESET), or resistance drift due to aging [109].

– **Incorrect Read Fault (IRF):** the cell returns an incorrect logic value when a read operation is performed, while the data stored by the cell is correct and not affected by the read operation. These faults are mainly hard faults, caused by the presence of resistive defects in the memory cell such as Df3, Df4, and Df5 in Fig. 7b [103].

– **Read Disturb Fault (RDF):** the cell returns a correct logic value when a read operation is performed, while the data stored by the cell is flipped by the read operation. These faults are mainly soft faults and occur in weak cells, i.e., when LRS (HRS) is larger (smaller) than the nominal value. The small bias current during the read operation is sufficient to complete a RESET (SET) operation. In the case of bipolar switching devices, only one state is affected by this fault, while in the case of unipolar switching devices both states are prone to RDF [110].

– **Unknown Read Fault (URF):** the read operation returns an arbitrary logic value, irrespective of the applied bias. These are soft faults caused by a combination of parametric variations which result in similar values of LRS and HRS, close to the nominal averaged LRS and HRS value, i.e., the reference value for read operation [107].

Some of these faults have been studied in relation with the Phase Change Memory device but they extend to any RRAM device type. Aside from these RRAM specific faults, all other faults (static or dynamic, single or multiple cell, coupling, etc.) introduced and studied for traditional memories are likely to occur. However, they are out of the scope of this paper, since they have been extensively studied in the past, for SRAM and DRAM memories. They have similar effects on the RRAMs and the same detection methods can be implemented.

3.5 Test and Design-for-Test

The traditional memory faults, i.e., TF, SAF, IRF, RDF, can be detected by March tests which consist of sequences of March elements. A March element consists of a sequence of operations (read and write) applied to each cell in the memory, before proceeding to the next cell [119]. Several works are centred on the analysis and detection of resistive defects, by exploiting a traditional memory fault analysis. Test strategies based on expansion of traditional march algorithms are proposed to identify faulty cells. Amongst these works, [105, 107, 108, 109] propose fast march test algorithms which employ sneak-path sensing to detect faults in the memory array. The testing schemes use sneak paths inherent in crossbar memories, to test multiple memory elements at the same time, thereby reducing testing time. However, these test algorithms are not suited for the 1T1R-RRAM structure.

Most RRAM faults can be detected by a traditional March C test [110]. However, the SAF (caused by over-

forming) and RDF (caused by non-fixed high and low RRAM resistance due to variation) need additional testing steps for detection as they can be dynamic faults (more than one operation is required in order to sensitize these faults). An extended (modified) March C algorithm has been proposed that contains two extra consecutive read operations to test for these dynamic faults [110].

Another extended March test, i.e., the March-1T1R is presented in [104] which also targets potential faults caused by the access transistor. This algorithm can detect the WDF and dWDF, as well as the other discussed faults.

Special attention has been given to the detection of UWFs, which cannot be detected by a March sequence alone. It requires stressing the cells in such a way that faulty cells flip their state from undefined to wrong, while healthy cells remain unchanged. Two DfT schemes have been proposed for RRAM test [103, 107] based on (i) the duration of the access pulse, referred to as *Sort Write Time Scheme*, (ii) the amplitude of the write voltage bias, referred to as *Low Write Voltage Scheme*, respectively. In these scenarios, when the DfT mode is enabled, a weak write operation is performed by setting a shorter duration of the access pulse or a lower amplitude of the write pulse, respectively. In order to detect the defective cells, a standard write operation is performed, followed by the proposed weak write operation. The faulty cells are detected by performing read operations and identify the cells which have undergone the state flip.

4 MRAM

4.1 Working Principle and Classification

Working principle: The Magnetic Random Access Memory (MRAM) is a non-volatile RAM. Its data storage element is a three layer Magnetic Tunnelling Junction (MTJ) device [38] which consists of one oxide barrier layer sandwiched between two ferromagnetic layers (FLs). One of the two magnetic layers, referred to as fixed layer, has a fixed magnetic orientation set at fabrication time, whereas the other, called free layer, has a freely rotating magnetic orientation that can be dynamically changed by forcing sufficient tunnelling currents across the device (Fig. 8a). The conductance of such a tunnelling junction can vary depending on whether the magnetizations of the FLs have parallel (high conductance) or anti-parallel (low conductance) orientations. This effect is called Tunnelling Magneto-Resistance Effect (TMR) and is characterized by the TMR ratio, the ratio between the conductances of the two relative orientations [39]. The voltage-resistance behaviour of an MTJ device exhibits a hysteresis characteristic [38] (see Fig. 8b).

All magnetic nanostructures abide by the thermally activated magnetization reversal. According to Néel-Brown theory, at finite temperature, there is a finite probability for the magnetization to flip and reverse its direction. The thermally activated magnetization reversal of an MTJ device is given by the ratio between the height of the energy barrier between the two magnetization states of the free layer and the energy scaling factor $k_B T$ (k_B the Boltzmann constant and T the operation temperature).

This is the underlying phenomenon on which the write operation of such devices is based, and the main cause of reliability concern, since it can cause spontaneous state reversal.

A typical MRAM cell is the 1T1MTJ memory cell which consists of an MTJ device (1MTJ) and an access device, typically an NMOS transistor (1T), as shown in Fig. 8c (here W.D. is the Write Driver, S.A. is the Sense Amplifier, REF is the reference current during read, BL is the Bit Line, SL is the Source Line and WL is the Word Line). To perform a read operation, a small voltage is applied to the BL, while the SL is grounded; subsequently, a current proportional to its electrical resistance passes through the MTJ device. The decision is taken by a sense amplifier, which compares the current flowing through the device against a reference current. The read operation is performed in the same manner for all types of MRAMs, while the write operation differs, dependent on the
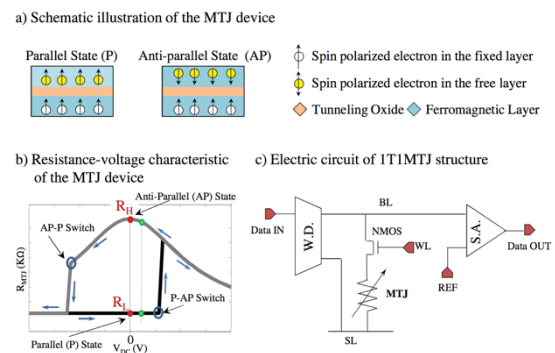
MRAM class. The MRAM classes are described next.

Fig. 8 Operation Principle of generic Magnetoresistive RAM.

Classification: The types of magnetoresistive memories are:

1) **Conventional MRAM:** relies on the fact that an external magnetic field influences the magnetization direction in a ferromagnetic layer. These memory devices feature an external field line. In the simplest design, each cell lies between a pair of perpendicular write lines which create a magnetic field when a current flows through them, which sets the magnetization direction of the free ferromagnetic layer. This approach requires a large current to generate the field, making it inapplicable for low-power applications. Moreover, device scaling is limited as the induced field may cause false writes to neighbour cells.

2) **Toggle MRAM:** the MRAM bit state can be programmed via a toggling mode. It relies on the unique behaviour of a synthetic anti-ferromagnet free layer formed from two ferromagnetic layers (with different net anisotropy) separated by a non-magnetic coupling spacer layer [40]. To achieve the spin flip desired during write operation, there exists a critical field at which magnetizations of the two anti-parallel layers will rotate to be orthogonal to the applied field.

3) **Spin-Transfer Torque MRAM:** relies on the ability of a spin-polarized current to flip the magnetization

direction of a ferromagnetic layer. The spin-torque effect is a result of conservation of angular momentum in layered magnetic devices [41]. This implementation does not require a field line; it is solely controlled by current flowing through the MTJ device. By passing through a ferromagnetic layer (the fixed layer), the current becomes spin polarized and maintains this polarization as it passes through the nonmagnetic oxide layer and the second ferromagnetic layer (the free layer). This leads to the change of the polarization orientation of the free layer.

4) **Thermal assisted switching MRAM:** In these devices, the magnetic tunnel junction is briefly heated up during the write process to facilitate magnetization reversal [122]. The written state remains stable at a colder temperature the rest of the time.

As a reasonably representative example, in the subsequent sections the focus will be on the STT-MRAM devices, since they have been reported as the most promising candidates to replace today's RAMs [3].

4.2 Opportunities and Challenges

Opportunities: Spin-Transfer Torque technology shows good potential as a replacement for both SRAM and DRAM memories. It shows good integration capabilities and potentially reduced fabrication cost, high operation speed, potentially to the level of L1 cache, high endurance and non-volatility [38, 120, 121].

Challenges: The asymmetric write operations and the susceptibility to spontaneous magnetic reversal are the greatest challenges faced by today's STT-MRAM devices. The write asymmetry limits their power efficiency and operation speed, since the memory is as fast as its slowest operation [120]. The spontaneous magnetic reversal limits the data retention time, causes read destructive faults, and imprints a probabilistic behaviour to the write operation [121].

Similar to other emerging technologies, an important challenge related to this memory is the large number of steps needed for the fabrication process and the relative lack of experience. This restricts the product dependability due to large fabrication-induced process variability and fabrication defects. These issues, in conjunction with the intrinsic stochasticity of the magnetic pillar, imposes great challenges on fault modelling and possible test solutions.

4.3 Defects

Similar to RRAM, MRAM devices are CMOS compatible. The CMOS front-end-of-line process can introduce defects such as resistive opens in the metal lines connecting the NMOS transistor source, gate and drain (Df1-Df3 in Fig. 7b).

After the CMOS logic is fabricated, the wafer is prepared for magnetic stack deposition by Chemical Mechanical Polishing (CMP) process. If this stage is not successfully completed, it causes issues such as low breakdown voltage or orange peel (Néel) coupling, leading to offset fields which affect the hysteresis curve. Both issues cause significant variations in the electrical characteristics

of the fabricated device. On the other hand, over- polishing during the CMP process can cause dishing and/or voids on the metal strap, or leave behind residual slurry particles [124]. The effect of these imperfections during the fabrication process causes a defective behaviour which can be modelled as a resistive open. Since this defect resides between the MTJ element and the CMOS logic, it contributes to the overall value of the resistive open defect Df3.

After the polishing process, the magnetic stack is deposited and annealed. The main issues that can arise during this fabrication step are: material contamination, rough surface layers, and reduced integrity of the oxide barrier. These issues can lead to a wide variation in the cell resistance and switching current [123, 124].

The magnetic stack deposition is followed by an etching process to obtain the desired MTJ pillar. This is the fabrication step which is the most difficult to control; therefore, the step is more prone to parameter variations and defects. The target of such a process is to obtain steep MTJ pillar edges, to prevent side-walls re-depositions, prevent magnetic layer corrosion, and control the device critical dimensions [123, 124]. Improper etching causes large variations in resistance and TMR ratio distributions over the fabricated devices, and hinders the switching process. Improper etching is also a source of resistive defects (shunt and contact) mainly due to side-walls re-depositions. These defects can be modelled by Df4 and Df5 in Fig. 7b. The MTJ etching process is today the main cause of weak and defective STT-MRAM cells.

There are several works in literature dealing with resistive defects. Examples are resistive opens, resistive shorts and bridges [111, 112, 115], as well as defects leading to large parameter variability [111, 113, 114, 115, 116, 117]. Most of these works focus on the defect, fault modelling, and test of Toggle-MRAM and TAS-MRAM memories. Only a few works are dedicated to the newer and more efficient STT-MRAM memory, which is the object of this study. The next subsection describes the fault models which are abstracted from the described possible defects in STT- MRAM devices.

4.4 Fault Models

Possible faults in MRAM device include TF, SAF, SWF, RDF, IRDF and URF, which are already defined for RRAM; therefore, their definitions will be omitted in this subsection. In case the defects originate from the same sources as for RRAM, the faults are omitted entirely (i.e., TF, SAF).

– **Undefined Write Fault (UWF):** these faults are mainly soft faults and occur as a consequence of the stochastic nature of the write operation, as the magnetization reversal is a probabilistic phenomenon. For instance, a device fabricated with a large free layer volume will require a large current to have a high probability of successful write operation [121]. This means that when such a cell is written with nominal bias, the probability of magnetization reversal will be low. This results in the cell settling in an arbitrary state. There is a fundamental difference between

UWFs in MRAM and RRAM. The undefined state in RRAM means that the cell resistance is settled to an intermediate state between LRS and HRS, while in MRAM the undefined state means that the cell settles in either low or high resistive state with a certain probability.

- **Slow Write Fault (SWF):** can be both hard and soft faults. The hard faults are caused by the presence of small resistive defects in the memory cell at locations Df2, Df3 and Df4 in Fig. 7b). The soft faults are caused by a weak access transistor, magnetic layer corrosion due to improper etching, by the Néel coupling (offset of hysteresis curve) or by large variations in the device critical dimension, which affect the efficiency of the magnetization reversal process.

- **Incorrect Read Fault (IRF):** these faults are mainly hard faults caused by the presence of resistive defects in the memory cell. They can also be soft faults as a result of fabrication-induced variability (such as deviations in the critical dimensions of the tunnelling layer), leading to significant variations in the resistance ratio, i.e., TMR [113].

- **Read Disturbance Fault (RDF):** is a soft fault caused by large variations in the device critical dimension or magnetic layer corrosion due to improper etching. This fault occurs due to the fact that the read and write paths are shared. Even if the read current is much lower than the critical write current, it can still induce a magnetic disturbance in the MTJ device. This may lead to magnetization reversal (a probabilistic fault). For low variability cells the occurrence probability of this fault is low, however, the probability of magnetization reversal increases as the number of consecutive read operations increases [113]. Therefore, the occurrence probability of a dynamic RDF is larger than the occurrence probability of a static RDF.

- **Retention Fault (RtF):** the cell can lose its state over time. This fault is due to thermal noise; this is a soft failure, resulting from large variation in the MTJ's thermal stability factor. The cell's thermal stability is strongly dependent on the volume of the free layer and on the uniaxial anisotropy, which can be strongly affected by the fabrication process, especially by the CMP and etching processes [124].

Other functional faults affecting the magnetic type memory behaviour have been featured in literature, and the most prominent seams to be the *Write Disturbance Fault (WDF)*, i.e., the state of a cell is flipped when a write operation is performed on an adjacent cell. However, in order for such fault to happen, the write paths of the aggressor and receptor cells must not be separated. This situation occurs for classical MRAM, toggle MRAM and even TAS-MRAM devices, but it does *not* occur in the case of the STT- MRAM cells, therefore this fault, and other similar faults remain out of the scope of this paper.

The MRAM memory is likely to suffer from other faults (static or dynamic, single or multiple cell, coupling) studied for traditional memories. The description and characterization of these faults remains out of the scope of

this paper, since they occur mostly at CMOS level and they have been extensively studied in the past (in relation to SRAM and DRAM memories). They have similar effects on MRAMs, consequently the same detection methods can be implemented.

4.5 Test and Design-for-Test

Much like in the case of RRAMs, the traditional faults occurring in an MRAM memory (i.e., TF, SAFs, IRF) can be detected by March tests. Several works are centred on the analysis and detection of resistive defects by exploiting a traditional memory fault analysis. Most of the work dedicated to test and design for test of MRAMs are specific to conventional, toggle or thermally assisted MRAMs [112, 114]. The proposed test techniques are mainly targeting the WDF which can occur with high probability in these memory devices. However, this is not the case for STT-MRAMs, which have the read destructive fault as one of the most common occurring faults.

In STT-MRAMs, however, the RDF is the specific fault most likely to occur. Consequently, research efforts have been dedicated to develop test algorithms and DfT solutions targeting these faults. For instance, two similar DfT techniques have been proposed in [116, 117]. They are based on tracing the MTJ current during the read operation. More specifically, they trace the ratio of the read current with respect to the reference current. If the STT-MRAM cell operates correctly, the read current is either always larger, or always smaller, than the reference current. If an RDF occurs, the ratio between currents flips at some point, after the read operation is completed. The DfTs are based on tracking the ratio between the read current and the reference current throughout the duration of the read operation, even after the outputs of the sense amplifier are stable. A read operation is activated by sufficient difference between active and reference currents (a differential sense amplifier is used), while the RDF detection is activated by a flip in the current ratio (current mirrors are used).

5 Conclusion

This paper discussed the test challenges and emerging solutions of 3D stacked memories, Resistive memories and Spin-Transfer- Torque Magnetic memories. From a test perspective, 3D stacked memories face the least challenges and are closest to enter the market. RRAM and STT-RAM, however, have besides the traditional faults also unique non-deterministic faults, as RRAM and STT-RAM devices suffer heavily from parametric variations. Currently, only a few works deal with RRAM and MRAM testing and they mainly propose structural fault model-based testing. The fault coverage of these tests are not correlated to the design specification and therefore many unique non-deterministic faults could remain undetected. With the new data storage paradigm and the ever-increasing performance demands, a viable companion to structural testing would be specification-based testing. In the latter, testing would not rely on fault models as it is entirely based on the design specification. However, this

testing approach still requires extensive research.

ACKNOWLEDGMENT

REFERENCES

[1] A.K. Sharma, *Advanced Semiconductor Memories: Architectures, Designs, and Applications*. Wiley-IEEE Press, 2009.

[2] M. Pavlovic *et al.*, "On the memory system requirements of future scientific applications: Four case-studies," in *IISWC*, Nov 2011, pp. 159–170.

[3] "The International Technology Roadmapfor Semiconductors 2015 Edition," ITRS, 2015. [Online]. Available: http://www.itrs.net

[4] U. Kang *et al.*, "8 Gb 3-D DDR3 DRAM Using Through-Silicon-Via Technology," *JSSC*, pp. 111–119, Jan. 2010.

[5] M. Kawano *et al.*, "A 3D Packaging Technology for 4 Gbit Stacked DRAM with 3 Gbps Data Transfer," in *IEDM*, Dec 2006, pp. 1–4.

[6] T. Zhang *et al.*, "A customized design of DRAM controller for on-chip 3D DRAM stacking," in *CICC*, Sept 2010, pp. 1–4.

[7] D.B. Strukov *et al.*, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, 2008.

[8] J. S. Meena, S. Min Sze, U. Chand, T.-Y. Tseng, "Overview of emerging nonvolatile memory technologies," Nanoscale Research Letters, vol. 9, pp. 526-559, 2014.

[9] I. Newsroom. 3d xpoint memory. [Online]. Available: https://newsroom.intel.com/press-kits/introducing-intel-optane-technology-bringing-3d-xpoint-memory-to-storage-and-memory-products/, 2015.

[10] Arstechnica. [Online]. Available: http://arstechnica.com/gadgets/2015/10/hp-and-sandisk-join-forces-to-finally-bring-memristor-like-tech-to-market/, 2015.

[11] EETimes. Avalanche samples spin trans- fer torque magnetic ram. [Online]. Available: http://www.eetimes.com/document.asp?doc id=1327122, 2015

[12] ExtremeTech. Toshibas new mram cache could reduce cpu power consump- tion by 60http://www.extremetech.com/extreme/184183 -toshibas-new-mram-cache-could-reduce-cpupowerconsumption-by-60, 2014

[13] S. Mick *et al.*, "Buried Bump and AC Coupled Interconnection Technology," *IEEE Trans. on Advanced Packaging*, pp. 121–125, Feb. 2004.

[14] K. Kanda *et al.*, "1.27Gb/s/pin 3mW/pin Wireless Superconnect (WSC) Interface Scheme," in *ISSCC*, vol. 46, no. 1, Feb. 2003, pp. 186–487.

[15] N. Miura *et al.*, "A 1TB/s 3W Inductive-Coupling Transceiver Chip," in *ASPDAC*, Jan. 2007, pp. 92–93.

[16] J. Xu *et al.*, "AC Coupled Interconnect for Dense 3-D ICs," *IEEE Trans. on Nuclear Science*, vol. 51, no. 5, pp. 2156–2160, Oct. 2004.

[17] W. Davis *et al.*, "Demystifying 3D ICs: The Pros and Cons of Going Vertical," *IEEE Design Test of Computers*, pp. 498–510, Nov. 2005.

[18] T. Jiang *et al.*, "3D Integration-Present and Future," in *10th Electronics Packaging Technology Conference*, Dec. 2008, pp. 373 –378.

[19] R. Anigundi *et al.*, "Architecture design exploration of three-dimesional (3D) integrated DRAM," in *ISQED*, March 2009.

[20] P. Garrou *et al.*, *Handbook of 3D Integration: Volumes 1 and 2 – Technology and Applications of 3D Integrated Circuits*. Weinheim, Germany: John Wiley & Sons, 2008.

[21] R. Patti, "Three-Dimensional Integrated Circuits and the Future of System-on-Chip Designs," *Proc. of the IEEE*, pp. 1214–1224, June 2006.

[22] J. Knickerbocker *et al.*, "2.5D and 3D Technology Challenges and Test Vehicle Demonstrations," in *ECTC*, May 2012, pp. 1068–1076.

[23] M. Taouil *et al.*, "Layer Redundancy Based Yield Improvement for 3D Wafer-to-Wafer Stacked Memories," in *ETS*, May 2011, pp. 45–50.

[24] K. Puttaswamy *et al.*, "3D-Integrated SRAM Components for High-Performance Microprocessors," *TC*, pp. 1369–1381, Oct 2009.

[25] P. Reed *et al.*, "Design aspects of a microprocessor data cache using 3D die interconnect technology," in *ICICDT*, May 2005, pp. 15–18.

[26] Y.F. Tsai *et al.*, "Design Space Exploration for 3-D Cache," *TVLSI*, 2008.

[27] G.H. Loh, "3D-Stacked Memory Architectures for Multi-core Processors," in *ISCA*, June 2008, pp. 453–464.

[28] H. Mujtaba, Samsung Begins Mass Production of HBM2 DRAM - 4 GB HBM2 for HPC, 8 GB HBM2 Production Commences This Year. [Online]. Available: http://wccftech.com/samsung-hbm2-dram/, 2016.

[29] H. Mujtaba. SK Hynix to Commence Mass Production of 4 GB HBM2 DRAM In Q3 2016 - Aiming at NVIDIA Pascal and AMD Plaris GPUs. [Online]. Available: http://wccftech.com/sk- hynix-hbm2-mass-production-q3-2016, 2016.

[30] J.K. Lee *et al.*, "Accurate analysis of conduction and resistive-switching mechanisms in double-layered resistive-switching memory devices," *App. Phys. Lett.*, 2012.

[31] L. Zhu *et al.*, "An overview of materials issues in resistive random access memory," *Journal of Materiomics*, vol. 1, no. 4, pp. 285 – 295, 2015.

[32] S. Yu, "Overview of resistive switching memory (rram) switching mechanism and device modeling," in *ISCAS*, 2014, pp. 2017–2020.

[33] E.I. Vatajelu *et al.*, "Nonvolatile memories: Present and future challenges," in *IDT*, Dec 2014, pp. 61–66.

[34] E.W. Lim *et al.*, "Conduction mechanism of valence change resistive switching memory: A survey," *Electronics*, vol. 4, no. 3, p. 586, 2015.

[35] A. Wedig *et al.*, "Nanoscale cation motion in taox, hfox and tiox memristive systems," *Nature Nanotech.*, vol. 11, pp. 67–74, 1 2016.

[36] R. Waser, "Electrochemical and thermochemical memories," in *IEEE IEDM*, 2008.

[37] D. Ielmini *et al.*, "Thermochemical resistive switching: materials, mechanisms, and scaling projections," *Phase Transitions*, vol. 84, no. 7, 2011.

[38] M. Hosomi *et al.*, "A novel nonvolatile memory with spin torque transfer magnetization switching: spin-ram," in *IEEE IEDM*, 2005.

[39] M. Julliere, "Tunneling between ferromagnetic films," *Physics Letters A*, vol. 54, no. 3, pp. 225 – 226, 1975.

[40] B.N. Engel *et al.*, "A 4-mb toggle mram based on a novel bit and switching method," *IEEE T. on Magnetics*, vol. 41, no. 1, pp. 132–136, Jan 2005.

[41] J. Katine *et al.*, "Device implications of spin-transfer torques," *J. Magnetism and Magnetic Mat.*, vol. 320, no. 7, pp. 1217 – 1226, 2008.

[42] M. Taouil, "Yield and cost analysis for 3d stacked ics," Ph.D. dissertation, Delft University of Technology, 2014.

[43] G. Smith *et al.*, "Yield Considerations in the Choice of 3D Technology," in *ISSM*, Oct. 2007, pp. 1–3.

[44] Cadence, "3D ICs with TSVs Design Challenges and Requirements," 2011. [Online]. Available: https://www.cadence.com/rl/resources/white papers/3dic wp.pdf

[45] J. Davis *et al.*, "Interconnect Limits on Gigascale Integration (GSI) in the 21st Century," *Proc. of the IEEE*, pp. 305–324, March 2001.

[46] H. He *et al.*, "Analysis of tsv geometric parameter impact on switching noise in 3d power distribution network," in *25th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC 2014)*, May 2014, pp. 67–72.

[47] T.Y. Kim *et al.*, "Bounded skew clock routing for 3d stacked ic designs: Enabling trade-offs between power and clock skew," in *International Green Computing Conference*, Aug 2010, pp. 525–532.

[48] A. Todri-Sanial *et al.*, "Worst-Case Power Supply Noise and Tempera- ture Distribution Analysis for 3D PDNs with Multiple Clock Domains," in *IEEE 11th International New Circuits and Systems Conference*, June 2013, pp. 1–4.

[49] W.K. Huang *et al.*, "New Approaches for the Repairs of Memories with Redundancy by Row/Column Deletion for Yield Enhancement," *TCAD*, pp. 323–328, March 1990.

[50] I. Kim *et al.*, "Built in Self Repair for Embedded High Density SRAM," in *ITC*, Oct. 1998, pp. 1112–1119.

[51] R. Adams, *High Performance Memory Testing: Design Principles, Fault Modeling and Self-Test*. Springer, 2003.

[52] L. Jiang *et al.*, "Yield Enhancement for 3D-Stacked Memory by Redundancy Sharing Across Dies," in *ICCAD*, Nov. 2010, pp. 230–234.

[53] M. Lefter *et al.*, "Is tsv-based 3d integration suitable for inter-die memory repair?" in *DATE*, March 2013, pp. 1251–1254.

[54] A.C. Hsieh *et al.*, "TSV Redundancy: Architecture and Design Issues in 3-D IC," *TVLSI*, pp. 711–722, 2012.

[55] J. Jung *et al.*, "Cost-Effective TSV Redundancy Configuration," in *VLSI-SoC*, Oct. 2012, pp. 263–266.

[56] P.C. Chew *et al.*, "Through Silicon Via (TSV) Redundancy - a High Reliability, Networking Product Perspective," in *EMAP*, Dec. 2012.

[57] E.J. Marinissen, "Testing TSV-Based Three-Dimensional Stacked ICs," in *DATE*, March 2010, pp. 1689–1694.

[58] M. Taouil *et al.*, "Quality versus Cost Analysis for 3D Stacked ICs," in *IEEE 32nd VLSI Test Symposium*, April 2014, pp. 1–6.

[59] E.J. Marinissen and Y. Zorian, "Testing 3D Chips Containing Through- Silicon Vias," in *International Test Conference*, Nov. 2009, pp. 1–11.

[60] K. Smith *et al.*, "Evaluation of TSV and Micro-Bump Probing for Wide I/O Testing," in *ITC*, Sept. 2011, pp. 1–10.

[61] H.H.S. Lee *et al.*, "Test Challenges for 3D Integrated Circuits," *IEEE Design & Test of Computers*, vol. 26, no. 5, pp. 26–35, Sept. 2009.

[62] M. Agrawal *et al.*, "Test-Cost Modeling and Optimal Test-Flow Selection of 3-D-Stacked ICs," *TCAD*, pp. 1523–1536, Sept 2015.

[63] K. Chakrabarty *et al.*, "TSV Defects and TSV-Induced Circuit Failures: The Third Dimension in Test and Design-for-Test," in *IRPS*, April 2012.

[64] S. Kannan *et al.*, "Fault Modeling and Multi-Tone Dither Scheme for Testing 3D TSV Defects," *JETTA*, pp. 39–51, Feb. 2012.

[65] F. Ye *et al.*, "TSV Open Defects in 3D Integrated Circuits: Characterization, Test, and Optimal Spare Allocation," in *DAC*, June 2012.

[66] C.W. Kuo *et al.*, "Thermal Stress Analysis and Failure Mechanisms for Through Silicon Via Array," in *ITherm*, May 2012, pp. 202–206.

[67] Q.C. X. Liu *et al.*, "Failure Mechanisms and Optimum Design for Electroplated Copper Through-Silicon Vias (TSV)," in *ECTC*, May 2009, pp. 624–629.

[68] M. Jung *et al.*, "Full-Chip Through-Silicon-Via Interfacial Crack Analysis and Optimization for 3D IC," in *ICCAD*, Nov. 2011, pp. 563–570.

[69] A. Papanikolaou *et al.*, *Three Dimensional System Integration*. Springer US, 2011.

[70] A. Engin *et al.*, "Modeling of Crosstalk in Through Silicon Vias," *TEMC*, vol. 55, no. 1, pp. 149–158, Feb. 2013.

[71] E. Beyne *et al.* (2013, July) Failure Analysis for 3D TSV Systems. [Online]. Available: http://www.sematech.org/meetings/archives/3d/10124/pres/Beyne.pdf

[72] D. Jung *et al.*, "Disconnection failure model and analysis of TSV-based 3D-ICs" in *EDAPS*, Dec. 2012, pp. 164–167.

[73] S. Deutsch *et al.*, "TSV Stress-Aware ATPG for 3D Stacked ICs," in *ATS*, Nov. 2012, pp. 31–36.

[74] A. Ikeda *et al.*, "Design and Measurements of Test Element Group Wafer Thinned to 10 μm for 3D System in Package," in *ICMTS*, March 2004.

[75] N.K. Jha *et al.*, *Testing of Digital Systems*. Cambridge University Press, 2002. New York, NY, USA:

[76] M. Taouil *et al.*, "Interconnect Test for 3D Stacked Memory-on-Logic," in *DATE*, March 2014, pp. 1–6.

[77] D.L. Lewis *et al.*, "A Scanisland Based Design Enabling Prebond Testability in Die-Stacked Microprocessors," in *ITC*, Oct 2007, pp. 1–8.

[78] X. Wu *et al.*, "Scan chain design for three-dimensional integrated circuits (3D ICs)," in *ICCD*, Oct 2007, pp. 208–214.

[79] X. Wu *et al.*, "Test-access mechanism optimization for core-based three-dimensional SOCs," in *ICCD*, Oct 2008, pp. 212–218.

[80] L. Jiang *et al.*, "Test architecture design and optimization for three-dimensional socs," in *DATE*, April 2009, pp. 220–225.

[81] L. Jiangs *et al.*, "Layout-driven test-architecture design and optimization for 3D SoCs under pre-bond test-pin-count constraint," in *ICAD*, Nov 2009.

[82] IEEE 3D-Test Working Group (3DT-WG). (2014). [Online]. Available: http://grouper.ieee.org/groups/3Dtest/

[83] "Ieee standard for test access port and boundary-scan architecture - redline," *IEEE Std 1149.1-2013 - Redline*, pp. 1–899, May 2013.

[84] "Ieee standard testability method for embedded core-based integrated circuits," *IEEE Std 1500-2005*, pp. 1–117, June 2012.

[85] S. Deutsch *et al.*, "DfT architecture and ATPG for Interconnect tests of JEDEC Wide-I/O memory-on-logic die stacks," in *ITC*, Nov 2012.

[86] JEDEC. (2011) JEDEC, Wide I/O Single Data Rate (Wide I/O SDR) JESD229. [Online]. Available: http://www.jedec.org/standards-documents/results/jesd229

[87] K. Shibin *et al.*, "At-speed testing of inter-die connections of 3d-sics in the presence of shore logic," in *ATS*, Nov 2015, pp. 79–84.

[88] H. Ehrenberg *et al.*, "IEEE Std 1581 - A standardized test access methodology for memory devices," in *ITC*, Sept 2011, pp. 1–9.

[89] V. Pasca *et al.*, "Configurable Thru-Silicon-Via interconnect Built-In Self-Test and diagnosis," in *LATW*, March 2011, pp. 1–6.

[90] Y.J. Huang *et al.*, "Post-bond test techniques for TSVs with crosstalk faults in 3D ICs," in *VLSI-DAT*, April 2012, pp. 1–4.

[91] M. Taouil *et al.*, "Post-bond interconnect test and diagnosis for 3-d memory stacked on logic," *TCAD*, vol. 34, no. 11, pp. 1860–1872, Nov 2015.

[92] M. Indaco *et al.*, "On the impact of process variability and aging on the reliability of emerging memories (embedded tutorial)," in *ETS*, May 2014, pp. 1–10.

[93] S. Ghosh, "Embedded memory design for future technologies: Challenges and solutions," in *VLSI-Design*, Jan 2014, pp. 14–15.

[94] D. Ielmini *et al.*, *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*. John Wiley and Sons, 2015.

[95] N. Banno *et al.*, "A fast and low-voltage cu complementary-atom-switch 1mb array with high-temperature retention," in *VLSI-Technology*, June 2014, pp. 1–2.

[96] G. Jurczak, "Advances and trends of rram technology," in *SEMICON*, 2015.

[97] G. Niu, et al., "Geometric conductive filament confinement by nanotips for resistive switching of HfO2-RRAM devices with high performance," Scientific Reports 6, Article number: 25757, 2016.

[98] D. Carta, et al., "Spatially resolved TiOx phases in switched RRAM devices using soft X-ray spectromicroscopy," Scientific Reports 6, Article number: 21525, 2016

[99] J. Singh *et al.*, "An overview: electron beam-physical vapor deposition technology - Present and future applications," *App. Research Lab., Penn. State Uni.*, 1999.

[100] L.G. Wang *et al.*, "Excellent resistive switching properties of atomic layer-deposited al2o3/hfo2/al2o3 trilayer structures for non-volatile memory applications," *Nanoscale Research Letters*, 2015.

[101] O. Ginez *et al.*, "Design and test challenges in resistive switching RAM (ReRAM): An electrical model for defect injections," in *IEEE ETS*, May 2009, pp. 61–66.

[102] N.Z. Haron *et al.*, "On defect oriented testing for hybrid cmos/memristor memory," in *IEEE ATS*, Nov 2011.

[103] N.Z. Haron *et al.*, "DfT schemes for resistive open defects in RRAMS," in *DATE*, March 2012, pp. 799–804.

[104] Y.X. Chen *et al.*, "Fault modeling and testing of 1T1R memristor memories," in *IEEE VTS*, April 2015.

[105] S. Kannan *et al.*, "Modeling, detection, and diagnosis of faults in multilevel memristor memories," *TCAD*, pp. 822–834, May 2015.

[106] S. Hamdioui *et al.*, "Testing open defects in memristor-based memories," *IEEE Transactions on Computers*, pp. 247–259, Jan 2015.

[107] S. Kannan *et al.*, "Sneak-path testing of crossbar-based nonvolatile random access memories," *TNano*, pp. 413–426, May 2013.

[108] S. Kannan *et al.*, "Detection, diagnosis, and repair of faults in memristor-based memories," in *IEEE VTS*, April 2014, pp. 1–6.

[109] S.N. Mozaffari *et al.*, "Fast march tests for defects in resistive memory," in *IEEE/ACM NANOARCH*, July 2015, pp. 88–93.

[110] C.Y. Chen *et al.*, "RRAM defect modeling and failure analysis based on march test and a novel squeeze-search scheme," *TC*, pp. 180–190, Jan 2015.

[111] C.L. Su *et al.*, "Mram defect analysis and fault modeling," in *ITC*, Oct 2004, pp. 124–133.

[112] J. Azevedo *et al.*, "A complete resistive-open defect analysis for thermally assisted switching MRAMs," *IEEE T. on VLSI*, vol. 22, no. 11, pp. 2326–2335, Nov 2014.

[113] A. Chintaluri *et al.*, "Analysis of defects and variations in embedded spin transfer torque (STT) MRAM arrays," *JETCAS*, pp. 1–11, 2016.

[114] C.L. Su *et al.*, "Write disturbance modeling and testing for mram," *IEEE T. on VLSI*, vol. 16, no. 3, pp. 277–288, March 2008.

[115] R. Robertazzi *et al.*, "Analytical mram test," in *ITC*, Oct 2014.

[116] R. Bishnoi *et al.*, "Read disturb fault detection in STT-MRAM," in *ITC*, Oct 2014.

[117] Y. Ran *et al.*, "Read disturbance issue for nanoscale stt-mram," in *NVMSA*, 2015.

[118] "Resistive Switching: From fundamentals of nanoionic redox process to memristive device applications," edited by D. Ielmini and R Waser, Wiley-VCH Verlag GmbH&Co, ISBN: 978-3-527-68094-8, 2016

[119] A. J. Van De Goor, "Using march tests to test SRAMs," in *IEEE Design & Test of Computers*, vol. 10, no. 1, pp. 8-14, March 1993.

[120] K. W. Kwon, S. H. Choday, Y. Kim and K. Roy, "AWARE (Asymmetric Write Architecture with REdundant Blocks): A High Write Speed STT-MRAM Cache Architecture," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 22, no. 4, pp. 712-720, April 2014.

[121] X. Fong, Y. Kim, R. Venkatesan, S. H. Choday, A. Raghunathan and K. Roy, "Spin-Transfer Torque Memories: Devices, Circuits, and Systems," in Proceedings of the IEEE, vol. 104, no. 7, pp. 1449-1488, July 2016.

[122] S. Chaudhuri, W. Zhao, J. O. Klein, C. Chappert and P. Mazoyer, "Design of TAS-MRAM prototype for NV embedded memory applications," 2010 IEEE International Memory Workshop, Seoul, 2010, pp. 1-4.

[123] W. Zhao, et al., "Failure Analysis in Magnetic Tunnel Junction Nanopillar with Interfacial Perpendicular Magnetic Anisotropy," *Materials*, 9(1), 41, 2016.

[124] B. Dieny, R. Goldfarb, K.-J. Lee "Introduction to Magnetic Random-Access Memory," The Institute of Electrical and Electronics Engineers (IEEE), published by John Wiley & Sons, ISBN: 978-1-119-00974-0, 2017.

[125] Micron 3D Xpoint technology. [Online]. Avaliable at: https://www.micron.com/about/emerging-technologies/3d-xpoint-technology

Elena Ioana Vatajelu received her PhD degree in electronics engineering from Universitat Politecnica de Catalunya (UPC), Barcelona, Spain in September 2011. She is currently a researcher at TIMA Laboratory in Grenoble, France. Her current research activity is focused on emerging memory technologies with special emphasis on spin-based devices. She is mainly focusing on the characterization of fabrication-induced process variability, fault modeling and defect characterization; design-for-reliability, design-for-test and design-for-security. She is strongly involved in the international community as a member of the Program and Organizing Committees of major conferences.

Paolo Prinetto is full professor of Computer Engineering at the Dipartimento di Automatica e Informatica of the Politecnico di Torino, Torino (Italy), and Adjoint Professor of the University of Illinois at Chicago, IL (USA). He is the President of CINI: Consorzio Interuniversitario Nazionale per l'Informatica and Vice-Chair of the IFIP Technical Committee TC 10 - Computer Systems Technology. From 2010 to 2014 he was Member of the Scientific Committee of the French "Centre National de la Recherche Scientifique" (C.N.R.S.). Research activities mainly focused on Digital Systems Design & Test, System Dependability, Hardware Security and Trust, Emerging Memories, FPGA-based Reconfigurable System Design, Assistive Technologies and ICT for people with disabilities.

Mottaqiallah Taouil received the M.Sc. and Ph.D. degrees (both with Hons.) in computer engineering from the Delft University of Technology, Delft, The Netherlands. He is currently a Post-Doctoral Researcher with the Dependable Nano-Computing Group, Delft University of Technology. His current research interests include reconfigurable computing, embedded systems, very large scale integration design and test, built-in-self-test, and 3-D stacked integrated circuits, architectures, design for testability, yield analysis, and memory test structures.

Said Hamdioui is currently a Chair Professor on Dependable and Emerging Computer Technologies at the Computer Engineering Laboratory of the Delft University of Technology (TUDelft), the Netherlands. Prior to joining TUDelft, Hamdioui worked for Intel Corporation (Califorina, USA), Philips Semiconductors R&D (Crolles, France) and for Philips/ NXP Semiconductors (Nijmegen, The Netherlands). His research focuses on two domains: Dependable CMOS nano-computing (including Reliability, Testability, Hardware Security) and emerging technologies and computing paradigms (including 3D stacked ICs, memristors for logic and storage, in-memory-computing). He owns one patent and has published one book and co-authored over 170 conference and journal papers. He delivered dozens of keynote speeches, distinguished lectures, and invited presentations and tutorial at major international forums/conferences/schools and at leading semiconductor companies. Hamdioui is a Senior member of the IEEE, Associate Editor of IEEE Transactions on VLSI Systems (TVLSI), and he serves on the editorial board of IEEE Design & Test, and of the Journal of Electronic Testing: Theory and Applications (JETTA). He is also member of AENEAS/ENIAC Scientific Committee Council (AENEAS =Association for European NanoElectronics Activities).