Multi-level storage in phase-change memory devices

A. Sebastian*, M. Le Gallo, W. W. Koelmans, N. Papandreou, H. Pozidis and E. Eleftheriou
IBM Research - Zurich, 8803 Rüschlikon, Switzerland
*ase@zurich.ibm.com

ABSTRACT

Phase-change memory (PCM) is a promising technology for both storage class memory and emerging non-von Neumann computing systems. For both applications, a key enabling technology is the ability to store multiple resistance levels in a single device. Multi-level storage is achieved by modulating the size of the crystalline/amorphous phase configuration. A key challenge, in this respect, is the device variability, which can be addressed by iterative programming schemes. When retrieving the stored information, the two additional challenges are resistance drift and low-frequency noise. Resistance drift is attributed to a spontaneous structural relaxation of the unstable amorphous states to a more stable "ideal glass" state and is well captured by a collective relaxation model. This model, in conjunction with the electrical transport models, provides a complete description of the time/temperature dependence of electrical transport in PCM devices. To counter resistance drift, several strategies have been devised, such as drift-resilient read-out mechanisms as well as coding and detection schemes. These techniques have helped to demonstrate storing up to 8 levels of information in a single PCM device. Yet another fascinating new approach is that of drift-resilient device architectures. Experimental results on prototype devices show remarkable promise in terms of eliminating drift as well as low-frequency noise.

Key words: Phase change memory, multi-level storage, resistance drift

1. INTRODUCTION

Computing is becoming progressively more data-centric than compute-centric [1]. Ultra-fast, high-density, nonvolatile memory that could occupy the space between conventional memory (SRAM,DRAM) and storage (Flash, HDD) also known as storage class memory, could play a key role in future computing systems [2]. It is also becoming increasingly clear that the conventional von Neumann architecture is highly inefficient when dealing with such data-centric computing because of the need to shuttle data back and forth between the physically separated memory and processing units. Several non-von Neumann computing architectures are being researched, such as the brain-inspired neuromorphic computing and memcomputing. Non-volatile memory devices could also play a key role in these computing paradigms to emulate neuronal or synaptronic devices or as elements of computational memory in a memcomputing processor [3], [4].

Phase-change memory (PCM) is arguably the most advanced non-volatile memory candidate that could serve many of these applications. And a key distinguishing aspect is the ability to store multiple levels of resistance information in a single device. In the context of storage class memory, this provides a clear path to reduce the cost per bit, a key obstacle to a wide-spread commercialization of PCM. For applications in non-von Neumann computing, multi-level resistance storage is central to the realization of several neuromorphic and memcomputing algorithms [5], [6].

In this article, we provide an overview of the state of the art in multi-level storage in PCM. First we present ways to achieve multiple resistance levels in a reliable manner. Subsequently we address the key challenge of resistance variations and ways to counter it via a combination of read-out mechanisms and signal processing and coding schemes. It can be shown that with these techniques, it is possible to store up to 8 levels of information in PCM devices. Finally we present the concept of projected PCM that could have a significant impact on multi-level PCM.

2. PROGRAMMING

In a PCM device, a tiny volume of phase-change material is sandwiched between two electrodes. When a RESET pulse is applied to the device, an amorphous region is created via the melt-quench process. When SET pulses are applied, the size of the amorphous region decreases because of crystallization. In these nanoscale devices, crystal growth appears to be the dominant mechanism with which crystallization occurs [8]. Multiple resistance levels are achieved by varying the size of the amorphous region by the application of

either RESET or SET pulses. One could vary the amplitude or the width of the pulse or also the width of its trailing edge. A typical programming curve that shows the variation of the programmed resistance as a function of the programming current (obtained by modulating the gate voltage) is shown in Figure 1(a). The left part of the programming curve is unidirectional and involves mostly amorphous-to-crystalline phase transition. The right part of the programming curve is mostly bidirectional, with the melt-quench process dominating the phase transition.

However, given the device-to-device variations across an array, it is difficult to achieve identical states across a large array by applying single programming pulses. Even though we can achieve a certain phase configuration in a single PCM device by applying an appropriate programming pulse, it is impossible to achieve the same state in a large collection of devices by application of the same programming pulse. In Figure 1(b), we show a resistance distribution corresponding to 4 programmed levels obtained via single-pulse programming. Because of the variability, the distributions overlap. Phase-change devices exhibit both inter-device and intra-device variability. Inter-device variability arises predominantly from structural variability during the fabrication of phase change devices. The physical attributes of the device, such as the thickness of the phase-change layer, the electrode dimensions etc., can vary across the entire wafer. This variability is likely to increase further when the devices scale into future technology nodes. In contrast, the most likely reason for intra-device variability is that the amorphous regions created via the melt-quench process are likely to have a slightly different atomic configuration with a different distribution of crystalline nuclei each time [9].

To counter this variability, iterative programming is needed where a sequence of write-and-verify steps is used in a feedback loop to minimize the error between the programmed and a specified target resistance level [7]. With iterative programming, one can achieve tight distributions of resistance levels as shown in Figure 1(c). The iterative programming operation needs to be efficient to be of practical significance. To achieve low energy consumption, several schemes have been proposed that minimize the application of high-energy melting pulses. To reduce latency, a scheme has been proposed that implement the measurement of the current error and the subsequent corrections to the amplitude of the write pulse in the analog domain [7], [10]. A 512 Mb PCM chip was fabricated in the 90 nm CMOS process that implements iterative programming schemes [11], [12]. MLC programming at 2 bits/device has been accomplished using this chip; in particular, convergence of the programming algorithm was achieved in fewer than 13 iterations for 99.9% of the devices, resulting in a write access time of 9.8 µs.

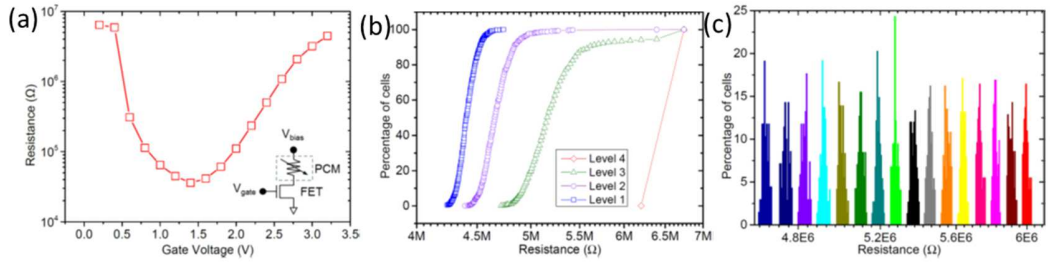

Fig. 1. (a) A typical programming curve indicating the programmed resistance vs. the programming current. In this particular case, the programming current is varied by varying the voltage applied to the gate of the FET access device under constant bias voltage. (b) Resistance distribution corresponding to 4 programmed levels using single-pulse programming. Variations cause the distributions to overlap because the same applied voltage pulse leads to different temperatures in different devices. (c) Tight, well-controlled distributions can be achieved using iterative programming as described in Papandreou et al. [7].

3. READ-BACK

3.1 Resistance Variations

A key challenge while retrieving the stored information is the resistance variations with time and temperature. These resistance variations are caused mostly by the phase-change material in the amorphous phase. Because electrical transport in amorphous phase-change materials is thermally activated (the activation energy is typically in the range of 0.2 to 0.4 eV), the programmed resistance levels exhibit a significant temperature dependence. But what is more detrimental is the temporal variations in properties, such as the activation energy for carrier emission. This naturally translates into a temporal evolution of the resistance typically referred to as the resistance drift. For example, at constant ambient temperature, the resistance typically

exhibits a temporal dependence characterized by $R(t) = R(t_0)(t/t_0)^{\upsilon_R}$, where $R(t_0)$ is the resistance measured at time t_0. The drift coefficient υ_R, which has a typical value of 0.1, exhibits significant inter-device and intra-device variability. This drift variability is arguably the most significant challenge for multi-level storage in PCM [14]. Another key challenge is that of noise. When a bias voltage, V_R, is applied to read back the resistance levels, significant fluctuations on top of the mean current, I_R, are observed. This relatively low frequency noise exhibits a characteristic constant normalized current spectral density S_I/I_R^2, with a $1/f$ frequency dependence [15].

There is wide consensus that resistance drift is caused by the structural relaxation of the amorphous phase. When the molten phase-change material is quenched rapidly, the atomic configurations are frozen into a highly stressed glass state that relaxes to lower-energy states [16]. Recent first-principles calculations provide significant insights into the microscopic picture of structural relaxation and the nature of the "ideal glass" [17], [18]. They show that the increase in resistance over time is correlated with the elimination of defects in the amorphous phase, accompanied by a slow evolution of the bond network towards structures that have a first-neighbor topology similar to that of the crystalline phase, but without the required long-range order.

One approach to model the dynamics of structural relaxation is via a two-state model for the relaxation of defects [19]. This is based on the popular relaxation model proposed by Gibbs [20]. The essential idea is that there are structural defects that can be removed by relaxation. To remove these defects, different activation energies are required. As the relaxation proceeds, defects with lower activation energy will be removed first, followed by those with higher activation energy. The distribution of activation energies for the relaxation of defects serves as the parameter that tracks the state of relaxation of the material at any instance in time. But there are some drawbacks to this approach. To match experimental observations, it is required to have a relatively flat distribution of activation energies over the whole energy range in all materials exhibiting drift behavior. Moreover, according to the Gibbs picture, the defects that have undergone relaxation once no longer participate in subsequent relaxation processes.

An alternative microscopic picture is that the atomic configurations collectively relax towards a more favorable equilibrium state (a possible ideal glass state). The relaxation proceeds in a sequence of transitions between neighboring un-relaxed amorphous states as shown schematically in Figure 2(a) [13]. An order parameter capturing the distance of the unrelaxed states from the ideal glass state serves as tracking parameter for the state of relaxation. Such a model can describe the time/temperature dependence of low-field resistance variations remarkably well, as shown in Figure 2(b). Recently we showed that the collective relaxation model can also be used to describe the time/temperature dependence of the high-field transport regime [21].

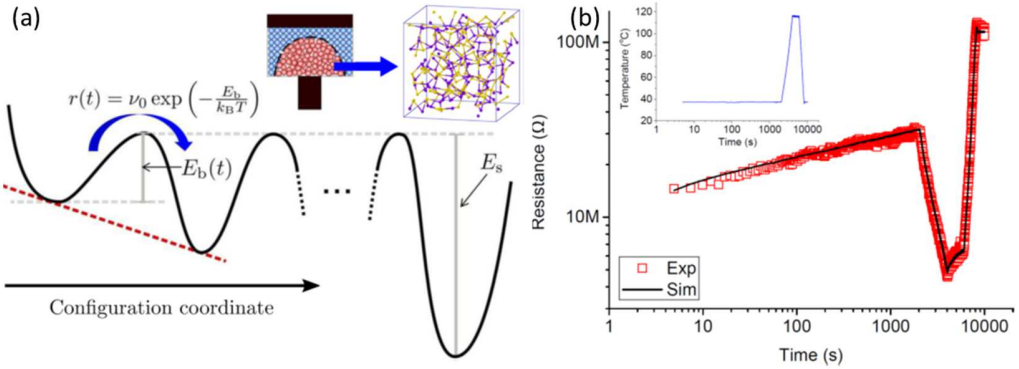

Fig. 2. (a) Schematic illustration of the structural relaxation model. The amorphous state created via the melt-quench process is in an unstable state and proceeds through a sequence of transitions between neighboring unrelaxed amorphous states towards the energetically more favorable "ideal glass" state. (b) Resistance variation associated with a high resistance state as a function of time and temperature when a temperature profile as shown in the inset is applied to the PCM device. The collective relaxation model captures the resistance variation remarkably well [13].

3.2 Drift-resilient Metrics

To counter the resistance variations, one approach is to design alternate metrics or read-out schemes that are more representative of the phase configuration, which is drift invariant. Owing to the strong field dependence of electrical transport in amorphous phase-change materials [23], we can get a better measure of the phase configuration if we explore the high-field regime of every programmed state [24]. But one cannot apply

arbitrarily high read voltages to all states because of threshold switching [25], [26] and subsequent read disturb. In the absence of a priori knowledge of the programmed state, the only way to explore the high-field regime of every programmed state is by applying a varying read voltage and then detecting the voltage or time at which a certain current threshold (I_{th}) is reached. This voltage or time value (typically referred to as the M-metric) is used as the measure of the programmed state. I_{th} may be fixed or varied as a function of the instantaneous voltage value [22], [27]. Figure 3 shows that one can mitigate the effect of drift significantly using the M-metric. Recently, a circuit realization of an M-metric (with I_{th} varying with the voltage value) was proposed with remarkably low read latency [28].

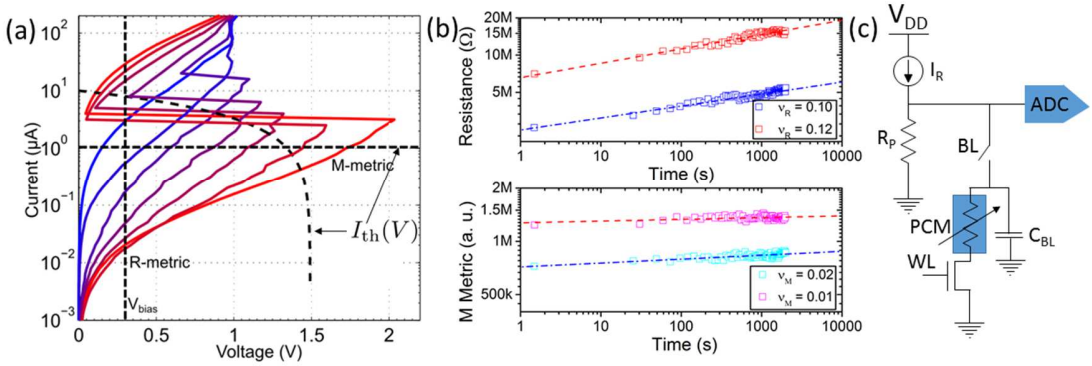

Fig. 3. (a) Measured *I-V* characteristics of a typical PCM device programmed at various states. The dashed lines indicate the ways in which the programmed states can be measured. The R-metric is measured by biasing the device at a low bias voltage. The M-metric is measured by detecting the voltage that corresponds to a threshold current I_{th} that could also vary with the bias voltage. (b) (top) The temporal evolution of resistance corresponding to two resistance states. (bottom) The corresponding temporal evolution of alternate metrics such as the M metric exhibit substantially reduced drift behavior [22]. (c) Schematic illustration of a low latency circuitry for M metric read-out where I_{th} varies with the bias voltage.

3.3 Signal Processing and Coding

A complementary approach to cope with the drift and variability in MLC PCM is through coding and signal processing. Drift and drift variability across devices in a PCM array cause stored level distributions to shift and broaden over time. Non-volatility, i.e., the ability of memory devices to retain the information stored over time, is seriously challenged by drift. Typical industry approaches of detecting stored data by fixed thresholds cannot provide adequate data retention in the presence of such signal distortion. One possible solution is to refresh the stored information at regular time intervals, but this is not practical for nonvolatile memories because of their large storage capacity and associated penalty in power, device wear and latency. A more viable approach is to adapt the detection thresholds according to the changing characteristics of the memory devices, as is typically done today in Flash memory. This is accomplished by using reference devices, which however, sacrifices user capacity and silicon area.

An alternate approach for adapting the level thresholds introduces modulation coding (data shaping) schemes in which user data are grouped into sets of symbols and encoded in short codewords [29], [30]. Instead of representing symbols by absolute voltage levels, information is partially stored in the relative order of levels within a codeword. As this relative ordering is typically maintained over time, such codes offer robustness to drift. By additionally requiring codes to be approximately balanced, i.e., to contain roughly equal numbers of each of the stored levels, changing level thresholds are estimated better. A schematic illustration of the level estimation and detection path is shown in Fig. 4.

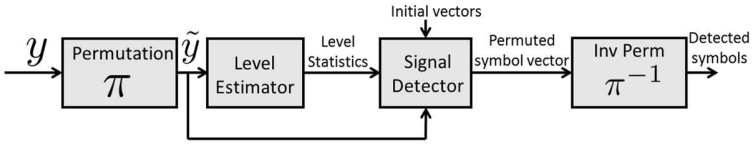

Fig. 4. Schematic illustration of level estimation and detection path. A vector of signal levels, y is first permuted ($\tilde{y}$) before being stored on the PCM devices. Upon readout, estimates of the mean level values are obtained and used to detect the permuted vector. Symbols are finally recovered by reverse permutation.

4. DEMONSTRATION OF STORING THREE BITS PER DEVICE

The combination of coding, adaptive threshold detection and judicious placement of target programmed levels leads to remarkable tolerance to drift and variability, even for prolonged periods of time after programming. It has recently been shown that a PCM array of 64k devices can be cycled 1 million times, then programmed at 8 levels/device (3 bits/device), and subsequently reliably detected 10 days later amidst ambient temperature variations between 25°C and 75°C [31]. In Fig. 5(a), the temperature profile and the corresponding M-metric measurements form the 8 programmed levels are shown. As shown in Fig. 5(b), the bit-error rate remained below 3×10^{-4} throughout the 10-day retention period. The experiments were conducted on a prototype PCM chip with a 4-bank interleaving architecture. The PCM devices are based on doped $Ge_2Sb_2Te_5$ and are integrated with the peripheral read/write circuitry at 90nm CMOS baseline technology.

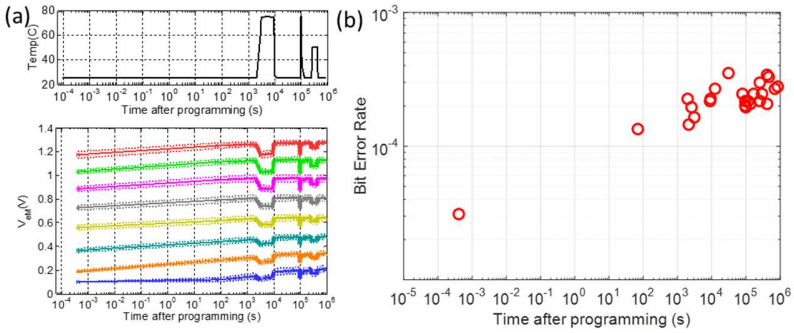

Fig. 5. (a) Time-temperature profile and the corresponding M-metric measurements from 8 programmed levels. (b) Bit-error rate for 3 bits/device storage.

5. PROJECTED PHASE-CHANGE MEMORY

An emerging avenue of research is that of designing memory devices specifically for multi-level storage. Although phase-change materials have excellent phase-transition properties, that is, they can undergo phase transitions on the nanosecond timescale and down to nanoscale dimensions, their highly disordered nature and high defect density make them susceptible to highly undesirable electrical effects, such as noise and drift. This leads to the difficult task of having to optimize the phase-change and the electrical properties in one and the same material.

We recently proposed the concept of projected PCM devices, where the device comprises a carefully designed segment consisting of a non-insulating material (projection segment) that is parallel to the phase-change segment as shown in Fig. 6(a) [32]. The resistance of this projection segment is judiciously chosen such that it has only a marginal influence on the write operation, but a significant influence on the read operation. This is indeed possible because of the highly nonlinear nature of the electrical transport in amorphous phase-change materials. At high fields, the amorphous material undergoes threshold switching, leading to a low-resistance ON state [26]. Thus, if during the high-field write process the resistance of the projection component is significantly higher than the ON-state resistance of the amorphous region, most of the current will still flow through the phase-change segment. During the low-field read process, however, the current bypasses the highly resistive amorphous region and flows through that part of the projection segment that is parallel to it. Hence, the resistance of the device is dominated by the resistance of that part of the projection segment, and thus is a good measure of the amorphous/crystalline phase configuration. The phase-configurational information is, in a sense, projected onto the projection component.

To demonstrate the concept of projection, several projected PCM devices with a lateral device geometry were fabricated. Both the temporal resistance drift and noise were practically eliminated in these devices. The drift coefficient υ_R was orders of magnitude lower than the value of 0.073 in identical phase-change devices without projection (see Fig. 6(b)). As shown in Fig. 6(c), the $1/f$ noise is also reduced significantly via projection, approaching noise figures close to the thermal noise floor.

The concept of projection could have significant consequences for multi-level storage. For example, to reduce the power needed to program a device, the phase-change component can be made as small as or as resistive

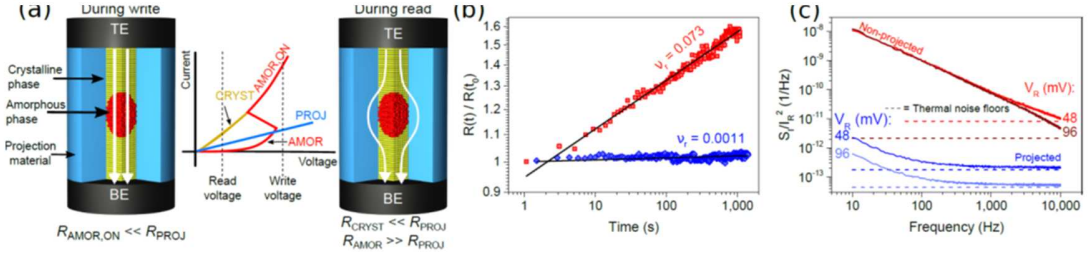

Fig. 6. (a) Compared to a conventional PCM device, the projected PCM device has an additional segment in parallel to the phase-change segment. The desired *I-V* characteristics corresponding to those two segments are shown schematically. The amorphous phase is denoted as AMOR with the corresponding resistance, R_{AMOR}, and the crystalline as CRYST with resistance R_{CRYST}. The projection segment is denoted as PROJ with resistance R_{PROJ}. In write mode, the write voltage exceeds the threshold voltage, and the amorphous section goes into the ON-state with a resistance $R_{AMOR,ON}$ that is lower than R_{PROJ}. This ensures that most of the current flows through the phase-change segment. In read mode, because R_{PROJ} is chosen to be much lower than R_{AMOR} at low fields, the current preferentially flows through the section of the projection segment that is parallel to the amorphous section. Elsewhere, the current will preferentially flow through the crystalline section. (b) Comparison between the resistance drift in a regular and a projected PCM device. (c) Comparison between the normalized spectral density of the current noise in a regular and a projected PCM device measured for two different read voltages. The dashed lines show the thermal noise floors for each bias voltage and resistance. [32].

as possible, as long as the phase change is facilitated. Moreover, the amorphous phase-change material can have undesirable electrical properties as long as the projection material has excellent electrical characteristics. Finally, the projection component can have very innovative geometries and resistance variations to facilitate both fabrication and multi-level storage.

We acknowledge the support of the PCM teams at IBM Research - Zurich and at the IBM T. J. Watson Research Center.

REFERENCES

1. J. Kelly III, S. Hamm, *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Columbia Univ. Press, 2013.
2. H.-S. P. Wong and S. Salahuddin, Nature Nanotechnology **10**(3) (2015) 191.
3. S. Saïghi *et al.*, Frontiers in Neuroscience **9** (2015) 51.
4. M. Di Ventra and Y. V. Pershin, Nature Physics **9(**4) (2013) 200.
5. T. Tuma *et al.*, Nature Nanotechnology (2016). DOI: 10.1038/nnano.2016.70.
6. P. Hosseini *et al.*, IEEE Electron Device Letters **36(**9) (2015) 975.
7. N. Papandreou *et al.*, Proc. IEEE Int'l Symp. Circuits and Systems (ISCAS 2011) 329.
8. A. Sebastian *et al.*, Nature Communications **5** (2014) 4314.
9. M. Le Gallo *et al.*, Proc. European Solid-State Device Research Conf. (ESSDERC 2016), in press.
10. H. Pozidis *et al.*, Proc. European Symp. Phase Change and Ovonic Science (E\PCOS 2011) 150.
11. G. F. Close *et al.*, Proc. 2011 IEEE Symp. on VLSI Circuits (VLSIC 2011) 202.
12. G. Close *et al.*, IEEE Trans. Circuits and Systems I **60(**6) (2013) 1521.
13. A. Sebastian *et al.*, Proc. IEEE Intl. Reliability Physics Symposium (IRPS 2015) MY.5.1.
14. H. Pozidis *et al.*, Proc. IEEE Intl. Memory Workshop (IMW 2012) 1.
15. P. Fantini, A. Pirovano, D. Ventrice, A. Redaelli, Applied Physical Letters **88** (2006). 263506.
16. D. Ielmini, S. Lavizzari, D. Sharma, A. Lacaita, Proc. IEEE Intl. Electron Dev. Meeting (IEDM 2007) 939.
17. J. Y. Raty *et al.*, Nature Communications **6** (2015) 7467.
18. F. Zipoli, D. Krebs, A. Curioni, Physical Review B **93** (2016) 115201.
19. S. Lavizzari, D. Ielmini, D. Sharma, A. L. Lacaita, IEEE Trans. Electron Dev. **56** (2009) 1078.
20. M. Gibbs, J. Evetts, J. Leake, J. Materials Science **18**(1) (1983) 278.
21. M. Le Gallo *et al.*, Proc. IEEE Intl. Reliability Physics Symposium (IRPS 2016), in press.
22. A. Sebastian *et al.*, Journal of Applied Physics**110**, (2011) 084505.
23. M. Le Gallo, M. Kaes, A. Sebastian, D. Krebs, New Journal of Physics **17**(9) (2015) 093035.
24. N. Papandreou *et al.*, Solid-State Electronics **54**(9) (2010) 991.
25. A. Pirovano *et al.*, IEEE Trans. Electron Devices **51**(3) (2004) 452.
26. M. Le Gallo, A. Athmanathan, D. Krebs, A. Sebastian, Journal of Applied Physics **119 (**2016) 025704.
27. N. Papandreou *et al.*, Proc. IEEE Intl. Electron Devices Meeting (IEDM 2011) 3.5.1.
28. J. Cheon *et al.*, Proc. IEEE Custom Integrated Circuits Conf. (CICC 2015) 1.
29. N. Papandreou *et al.*, Proc. IEEE Intl. Memory Workshop (IMW 2011) 147
30. H. Pozidis *et al.*, Proc. IEEE Intl. Memory Workshop (IMW 2013) 100.
31. M. Stanisavljevic *et al.*, Proc. IEEE 8th Intl. Memory Workshop (IMW 2016) 1.
32. W. W. Koelmans *et al.*, Nature Communications **6** (2015) 8181.