Extended Abstracts of the 2016 International Conference on Solid State Devices and Materials, Tsukuba, 2016, pp119-120

B-7-04

Investigation of variability in Vertical Resistive RAM (VRRAM): Physical Model

Giuseppe Piccolboni[1], Gabriel Molas[1], Daniele Garbin[1], Thilo Werner[1], Elisa Vianello[1], Barbara De Salvo[1], Gérard Ghibaudo[2], and Luca Perniola[1]

(1) CEA, LETI, MINATEC Campus, GRENOBLE, France, giuseppe.piccolboni@cea.fr, (2) IMEP LAHC CNRS Grenoble, France

Abstract—In this paper, we propose a physical model based on a spring effect and trap assisted tunnelling (TAT) conduction to explain the presence of a short term (≈40 cycles) correlation among resistances during ReRAM cycling. The memory cells used to attain our experimental results are HfO_2-based Vertical Resistive RAM (VRRAM) which show up to 10^7 cycling endurance. The paper finally shows the impact of the variability reduction thanks to this correlation on specific neuromorphic applications where a limited number of cycles is required, reducing the integration constraints.

Introduction

Thanks to their promising performances (scalability, switching speed and data retention [1]) Metal Oxide based Resistive RAMs (OxRAM) were proposed as a possible future replacement for flash technology. One of the areas where researchers are still investigating is the switching mechanism undergoing in the metal oxide and in particular the variability linked to the latter [2]. Variability increases for high resistances [3] and hence affects low power operations. In this study, following [4], a physical model based on TAT conduction fits experimental data showing the existence of a short range (≈40) resistances correlation during cycling. To efficiently reproduce this effect, an analogy with a spring effect is used, correlating the filament shape to the 1st and previous operating cycle of the memory. We thus propose a new physical interpretation of the phenomena governing switching variability that could help better understanding the origin of the cycle-to-cycle resistance variations. Finally we quantify the impact of this variability reduction for Convolutional Neural Network applications, requiring a reduced number of cycles.

Technological details

The measurements of this work were conducted on a 1T-1R HfO_2-based Vertical OxRAM (VRRAM) obtained with a MESA structure (fig.1). TiN which acts as Bottom Electrode (BE) is 40 nm thick. The metal oxide, namely HfO_2 is deposited by ALD its thickness being 5 nm. 10nm Ti Top Electrode (TE) is deposited on top of the resistive layer to gather oxygen from it and ensure a bipolar behavior of the cell.

VRRAM performances

1. Functionality

These cells can operate up to 10^7 cycles with a memory window of at least one decade as depicted in Fig.2. 3 extractions of 10 cycles from the endurance characteristic are reported in Fig.3. For each extraction the resistances are tightly distributed around a given average while the average resistance from different extraction varies significantly, especially for HRS. This observation if fundamental for understanding the experimental results presented in the next section.

2. Resistance variability

Previous works [3,1] showed that resistance distributions obtained by cycling OxRAM cells are characterized by a cell-to-cell and cycle-to-cycle variability which are well described by a lognormal distribution. This was confirmed also by data collected from our VRRAM samples on a high number of cycles (>100). On the other hand, considering a shorter cycling range, we observe a correlation among the resistances as witnessed also in [5]. This means that in this range the resistance at a given cycle is influenced by the resistances of the previous ones. The correlation is quantified with the correlation coefficient (CC) as thoroughly reported in [4]. Fig.4 shows the evolution of the CC for the HRS resistances obtained for I_{SET} = 100 μA and Fig.5 shows a 1-D cut of it. These data show how resistances are correlated up to λ≈40 cycles. Fig.6 shows the dependence of the correlation length (λ) as a function of I_{SET}, a higher programming current leading to higher correlation length.

Theoretical Model

Correlation among subsequent cycles can be explained by the fact that a physical parameter, linked to the cell resistance, is influenced by its values in the previous cycles. Previous works [3, 1] suggested that the resistances are log normally distributed because the oxide region where the filament was dissolved (L_{GAP}) varies normally and the tunnelling current flowing in the insulating gap has an exponential dependence with the distance. Our experiments showed that in the small cycling range (≈40) resistances have a constant average meaning that their values gravitate around the initial one. This suggests that also for L_{GAP} there must be a given relation during cycling. In order to verify this assumption, we first have to find a law governing the L_{GAP} evolution and then to verify that the corresponding resistances show a correlation among each other. Knowing that the resistance in the studied range has a constant mean value (no resistance drift) we suppose that L_{GAP} varies around a given value *and* is correlated from cycle to cycle. In other words we make an analogy with the motion of an object around the equilibrium point of a spring: the object is free to oscillate but always stays around the spring's equilibrium point (Fig.7). In order to have such an evolution L_{GAP} must somehow be related to its initial value (for the average) and its previous cycles (for the correlation). So, given an initial L_{GAP} (called $L_{GAP(0)}$) corresponding to an initial resistance ($R_{(0)}$), we say that L_{GAP} at the n^{th} cycle depends both on $L_{GAP(0)}$ and $L_{GAP\ n-1}$. In order to maintain a variability when sorting L_{GAP} values, we work with Gaussian probability density functions, calculating the probability of L_{GAP} at the n^{th} cycle as:

$$P_{L_{GAP\,n}} = N\left(\mu_{L_{GAP(0)}}, \sigma_{SPRING}\right)\ x\ N(\mu_{L_{GAP\,n-1}}, \sigma_{CYCLING}) \quad \text{Eq.1}$$

where σ_{SPRING} is the standard deviation of the Gaussian centered around the initial L_{GAP} value and $\mu_{L_{GAP\,(0)}}$ is its average. We indicated the standard deviation of the distribution centered around the initial value as "spring" because it has the effect of keeping L_{GAP} around the initial value during cycling. $\sigma_{CYCLING}$ represents the standard deviation of the Gaussian centered around the previous cycle (n-1th) L_{GAP} and $\mu_{L_{GAP\,n-1}}$ is its average. L_{GAP} distributions during cycling is represented in Fig. 9. Once L_{GAP} probability at a given cycle (n) is calculated, the corresponding resistance is obtained using a TAT model presented in [7]. This model calculates through a 3D network of resistances that simulates the oxygen vacancies (V_O) present in the HfO_2, the TAT current flowing in the device for different gap lengths (L_{GAP}). Fig.10 and 11 depict the V_O spatial distribution and density in the case of a large L_{GAP} corresponding to the High Resistive State (HRS). As shown in Fig.12 this model is able to correctly fit our experimental data. Fig.8 resumes the simulation steps that relate the gap length to the correlation existing among the cell resistance during cycling.

Neuromorphic Applications

ReRAM have been pointed out as one of the probable candidates for neuromorphic applications [8]. In this work we address a particular application, namely Convolutional Neural Networks (CNN, Fig.13) for handwritten digits recognition where we show the impact of the reduce variability on the circuit performance. Since this application needs few cycles [7], the correlation, that reduces cycle-to-cycle variability [6], could play an important role in the neural network performances. In our approach, the synapse is emulated by a vertical pillar composed by VRRAM operating in parallel [6]. The recognition rate of the circuit depends on the number of VRRAM cells per synapse. In Fig. 14 we compare this recognition rate for the overall resistance distribution (blue) and the distributions taking into account only a limited number of cycles (red). We see how the second one needs less devices per synapse to have the same performances. In particular, to reach 98% recognition rate (RR), 8 stacked levels are required when the variability is lower thanks to correlation, while 10 stacked levels are needed when we consider the full resistance distribution. A detailed study on the correlation existing among cycle can hence help properly dimension neural circuits where the variability is a key factor in the performances.

Conclusions

In this work we addressed the resistance correlation existing in the short cycling range (≈40) of OxRAM. A spring effect analogy combined with a Trap Assisted Tunneling-based physical model is used to fit our experimental results. We quantified the advantage of this correlation effect on the circuit performances of Vertical RRAM based Convolutional Neural Network requiring a limited number of cycles.

References

[1] M. Ueki et al., 2015 VLSI Tech. Dig., pp. T108 - T109.
[2] F.M. Puglisi et al., 2013 ICICDT Tech. Dig., pp. 85-88.
[3] D. Garbin et al., 2015 EUROSOI-ULIS Tech. Dig., pp. 125-128.
[4] G. Piccolboni et al., IEEE Elec. Dev. Lett., 2016, in press.
[5] A. Calderoni et al., 2014 Int. Mem. Workshop, pp. 1-4.
[6] G. Piccolboni et al., 2015 IEDM Tech. Dig., pp. 17.2.1-17.2.4.
[7] D. Garbin et al., 2014 IEDM Tech. Dig., pp. 661-664.
[8] D. Kuzum et al., Nanotechnology, p. 382001, 2013.

Acknowledgments

This work has been partially supported by the LabEx Minos ANR-10-LABX-55-01. It is also financed, within the PANACHE project, by European ENIAC Joint Undertaking and the French Government.

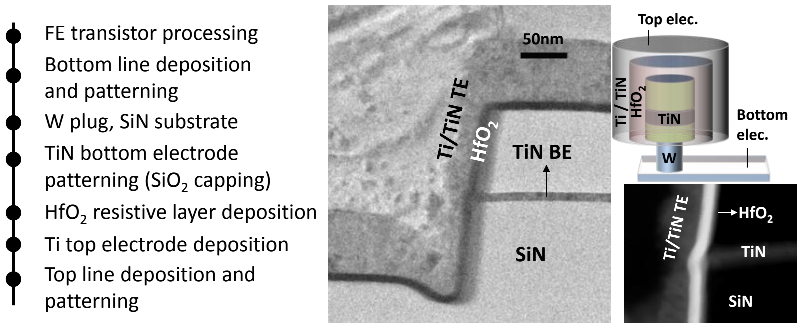

Fig 1. Description of the integration flow, and TEM cross sections (high resolution and dark field) of the TiN/HfO_2/Ti/TiN VRRAM.

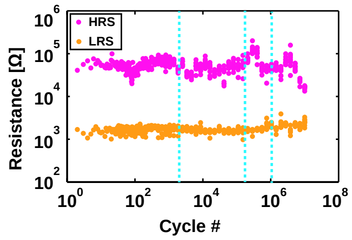

Fig.2 10^7 cycles endurance characteristic of HfO_2-based VRRAM.

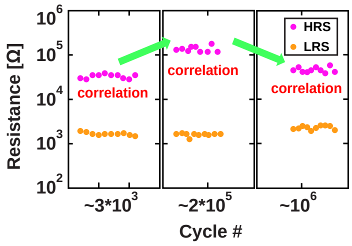

Fig.3 3 Extractions from Fig.2 to illustrate the tight resistances distribution.

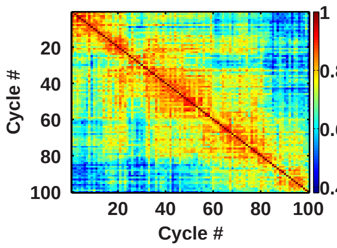

Fig.4 Correlation coefficient matrix for HRS resistances (50 cells, 100 cycles).

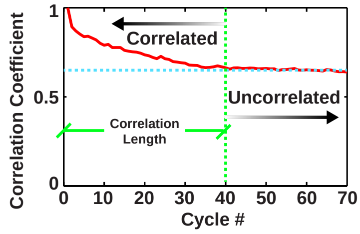

Fig.5 1-D cut of Fig.4. Correlation length λ is defined.

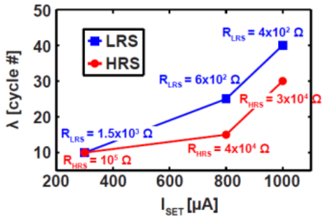

Fig.6 Evolution of the correlation length (λ) as a function of I_{SET}.

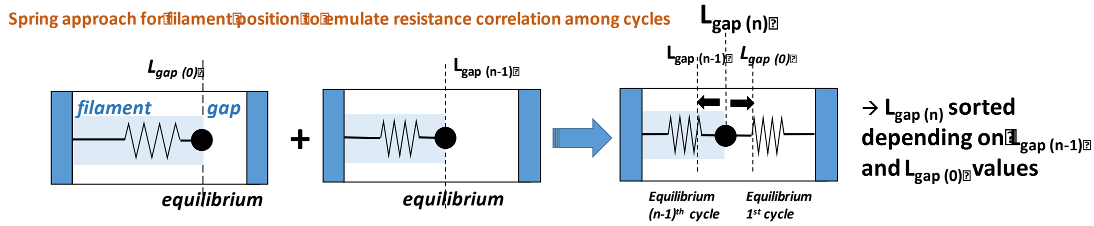

Fig.7 Spring Approach: the L_{GAP} at the n^{th} cycle ($L_{GAP\,(n)}$) depends both on the initial value ($L_{GAP\,(0)}$) and on the previous one ($L_{GAP\,(n-1)}$). The system behaves hence similarly to an object in an harmonic motion.

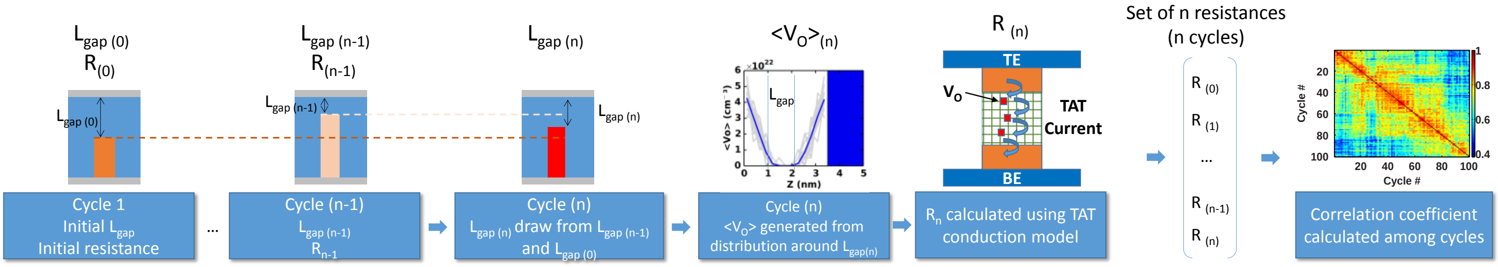

Fig.8 Simulation Framework: the filament gap at the n^{th} cycle is calculated from the initial one, that grants the constant resistance average in the cycle span considered, and the $(n-1)^{th}$ one, that introduce the correlation effect that will be measured in the resistance values caluclated via a trap assisted tunnelling model.

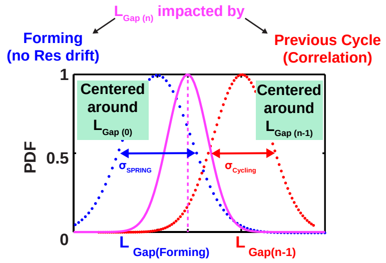

Fig.9 L_{GAP} probability density used to sort the n^{th} cycle ($L_{GAP(n)}$), using distributions created around the initial state gap ($L_{GAP(0)}$) and the gap at the previous cycle ($L_{GAP(n-1)}$).

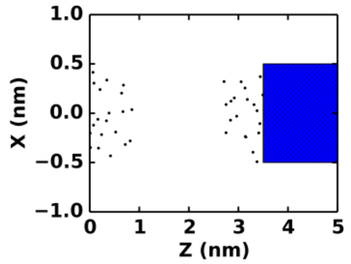

Fig.10 V_O spatial distribution. Parameters for TAT current calculation: energy barrier (E_B) = 2.5 eV, e^- effective mass (m_{eff}) = 0.1 m_0, interatomic distance (a) = 0.25 nm

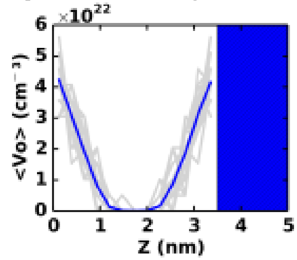

Fig.11 V_O density in the gap region of the metal oxide. The width of the spatial region containing low V_O density is related to the overall TAT flowing in the metal oxide.

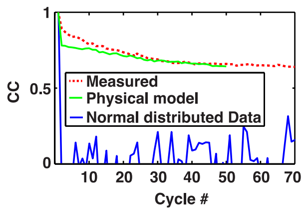

Fig.12 Correlation coefficient of measured (red), simulated (green) and normaly distributed (blue) resistances. Only the physical model fits the experimental data.

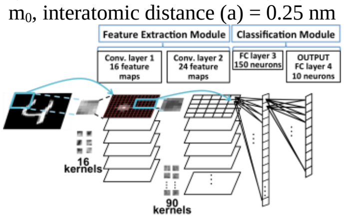

Fig.13 Convolutional Neural Networks architecture for handwritten digits recognition.

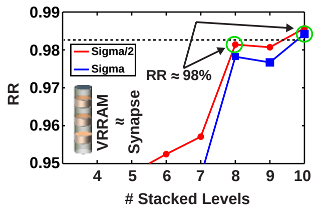

Fig.14 Recognition rate as a function of the number of stacked VRRAM in a synapse. Correlation (σ/2 for R distribution) allows (RR) of ≈ 98% with less stacked levels.