Heterogeneous integration of ReRAM crossbars in 180 nm CMOS BEoL process

J. Sandrini [a,*], M. Thammasack [a], T. Demirci [a], P.-E. Gaillardon [a], D. Sacchetto [b], G. De Micheli [a], Y. Leblebici [a]

[a] Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[b] CSEM, Neuchâtel, Switzerland

ARTICLE INFO

ABSTRACT

This work reports on a heterogeneous integration of resistive memories into the Back-End-of-the-Line of 180 nm standard CMOS foundry chips. A TaO_x-based ReRAM technology with materials and processes fully CMOS compatible has been developed and characterized. A low-cost integration method is applied to the developed TaO_x-based memories to achieve chip level ReRAM–CMOS integration. The integrated memory devices show working voltages compatible with CMOS circuits operations. Measured SET and RESET voltages of the ReRAM integrated cells are -1 V and $+1.3$ V, respectively, demonstrating suitability for low-voltage applications.

1. Introduction

Resistive Random Access Memories (ReRAMs) are intensively investigated as a possible replacement for Flash memories. When compared to mainstream NAND Flash technology, the main advantages of ReRAMs are low working voltage, fast operating speed [1], good endurance [2] and, thanks to their simple structure, high scalability [3]. Furthermore, ReRAMs can be fabricated with materials that are fully compatible with CMOS technology and with a suitable thermal budget that allows *Back-End-of-the-Line* (BEoL) integration. Combined ReRAM–CMOS processes enable the creation of a memory layer directly superposed to the CMOS front-end, reaching a memory density, over n number of ReRAM layers, of $4F^2/n$ (where F is the feature size). Besides standalone memories [4,5], ReRAMs are a fundamental building block for many promising applications such as neuromorphic circuits [6] and future generation FPGAs [7].

In this paper, we present a heterogeneous integration of ReRAMs with standard CMOS technology by post-processing the BEoL of fully-finished CMOS chips. In Section 2, the fabrication and characterization of CMOS-compatible ReRAM cells are described. In Section 3, the post-processing technique is discussed and, in Section 4, the electrical results of the integrated cells are presented. The conclusion is given in Section 5.

2. ReRAM technology development

In the framework of integrating a ReRAM technology compatible with CMOS processes, we developed a TaO_x-based memory technology targeting low voltage operations. Low temperature sputtered TiN has been used as electrode material for its compatibility with the BEoL, as it is a common diffusion barrier layer for the metal lines. TaO_x has been selected as memory layer for its retention and endurance characteristics [8]. Fig. 1(a) shows a schematic drawing of the fabricated 20 nm-thick TaO_x-based ReRAM devices. The advantage of a technology development based on standalone memory cells, instead of using CMOS-integrated ReRAMs, is an higher fabrication throughput and a simpler test setup.

First, the bottom electrode is formed by a 200 nm TiN sputtering on a passivated 4″ silicon wafer followed by a Cl_2/BCl_3-based *Reactive Ion Etching* (RIE). Then, a 100 nm layer of SiO_2 is deposited by Low-Temperature LPCVD (LTO) at 400 °C for electrical passivation. To delimit the ReRAM memory stack, VIAs are patterned by optical lithography. The VIAs, which have diameters of 1, 2, 3, 5 and 10 μm, are etched into the SiO_2 passivation by BHF. During the same process step the LTO is opened to give access to the bottom electrode pads. Then, 10 nm Ta, 20 nm TaO_x and 100 nm TiN layers are deposited by sputtering. The top electrode is defined by optical lithography followed by RIE. The micrograph in Fig. 1(b) shows a 10 μm cross-point ReRAM cell.

* Corresponding author. Tel./fax: +41 21 6936949.
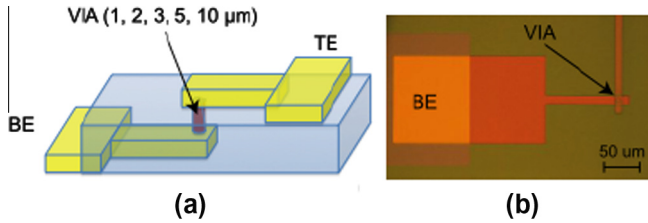 E-mail address: jury.sandrini@epfl.ch (J. Sandrini).

Fig. 1. (a) Schematic drawing of the TiN/Ta/TaO$_x$/TiN ReRAM devices. (b) Micrograph of the fabricated ReRAM cells.

Material characterization of the deposited TaO$_x$ layer has been carried out with *Energy Dispersive X-ray* (EDX) analysis using a FEI Tecnai Osiris TEM. Full wafer 20 nm TaO$_x$ and TiN layer were sputtered under the same conditions than for the ReRAM cell. Fig. 2(a) and (b) shows the EDX result and the depth profiles of the layers, respectively. Quantification of Ta and O gives 65.1% of O and 34.9% of Ta or TaO$_{2-0.14}$.

Electrical characterizations were carried out with an Agilent B1500 semiconductor device analyzer. Fig. 3 shows the *I–V* curve of a 10 μm cross-point ReRAM cell. A DC voltage is applied between the top and the bottom electrode, sweeping between −2 V and 2 V. The ReRAM forming operation occurs during the first sweep from 0 V to +2 V. Then, a sweep from 0 V to −2 V allows the RESET of the memory, while a DC sweep in the opposite voltage polarity allows the SET operation. The ReRAM cell shows low-voltage forming behavior ($V_{forming}$ = 2 V), therefore a high voltage forming step was not needed. No current compliance was applied to the cell. The observed resistances are 8 kΩ for the *Low Resistance State* (LRS) and 160 kΩ for the maximum *High Resistance State* (HRS), with a maximum HRS/LRS ratio of 20. The reset voltage is 1.8 V and the SET voltage is −1.5 V.

The measured *I–V* curve shows a linear current behavior in the LRS, exhibiting an ohmic conduction mechanism, and an exponential *I–V* relation in the HRS. The HRS conduction mechanism is
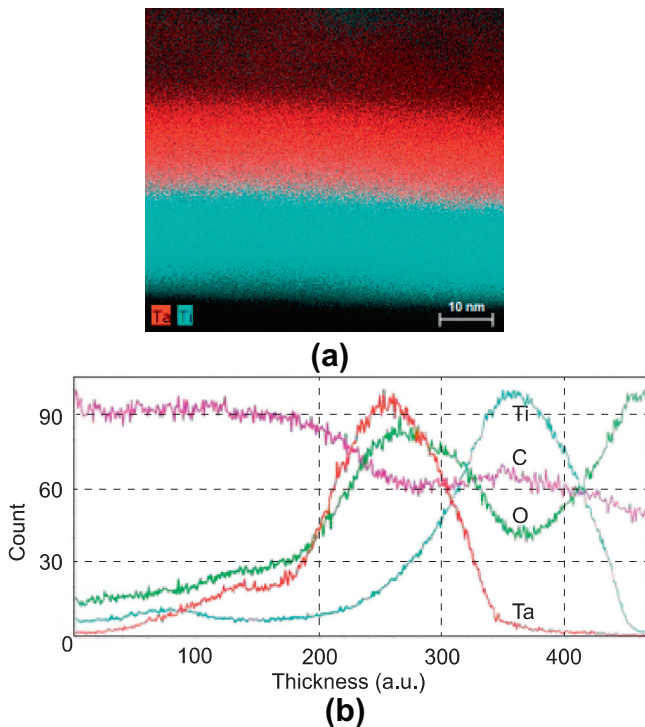

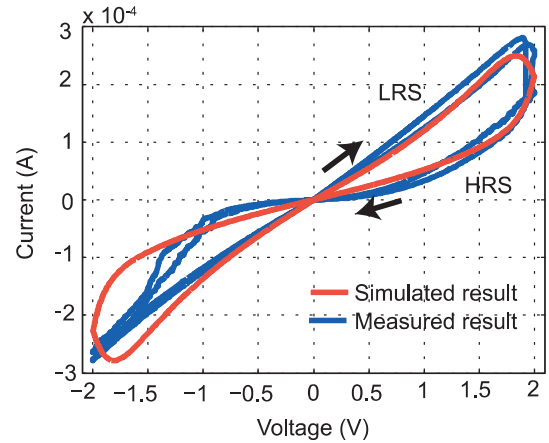**Fig. 2.** EDX analysis showing (a) Ti and Ta atoms and (b) depth profile.

Fig. 3. *I–V* curve of the fabricated 10 μm TiN/Ta/TaO$_x$/TiN-based ReRAM and the simulated curve from the model.

classically associated to tunneling phenomena, i.e., trap-assisted tunneling, Poole–Frenkel emission, Fowler–Nordheim tunneling and direct tunneling [9]. The measurement results are compared with the simulations obtained from the ReRAM Verilog-A compact model released by Stanford University [10]. This model interprets the various conduction mechanisms previously mentioned defining the current as an exponential function of the voltage and the gap according to the following equations:

$$I(g, V) = I_0 e^{-\frac{g}{g_0}} \sinh^{\frac{V}{V_0}}$$

$$\frac{dg}{dt} = v_0 e^{-\frac{E_a}{kT}} \sinh^{\frac{qa\gamma V}{t_{ox}kT}}$$

where g is the gap between the filament end and the electrode, v_0 is velocity containing the attempt to escape frequency, E_a is the activation energy, a is the hopping site distance, γ an enhancement factor (related to the polarizability of the material and to the non-uniform potential distribution) and t_{ox} is the switching material thickness.

The simulation result, obtained adapting the model's fitting parameters (t_{ox} = 20.55 nm, g_0 = 350 pm, V_0 = 1.2 V, v_0 = 19 nm/ns, I_0 = 258 μA, γ_0, the first value for γ = 15.85, g_{0max}, the maximum value for g_0 = 1.7 nm), fits well the measurements, corroborating the hypothesis that the HRS conduction mechanism is based on tunneling-based phenomena.

3. Post-processing technique

An integration method for passive memory arrays on top of CMOS chip front-end is discussed in this section. Integration is done by post-processing a fully processed CMOS die. Chip-level post-processing is attractive because it is cheaper compared to standard wafer-based solutions [11]. From a circuit perspective, the CMOS die embeds the read/write circuitry used to control the memory array.

We post-process the chips fabricated in 180 nm CMOS technology. A micrograph of the post-processed test chip is shown in Fig. 4. The chip includes an 8 × 8 crossbar array, single memory cross-points and various test structures. A TaO$_x$ thin film is deposited between the CMOS chip Metal 5 (M5) and an intermediate metal layer between M5 and M6, which is used in the target CMOS technology to fabricate *Metal–Insulator–Metal* (MIM) capacitors. Fig. 5 shows a scanning electron micrograph of the MIM capacitor, which is electrically connected through VIAs to M6. The M5, M6 and the MIM layers are made of Al, with an outer
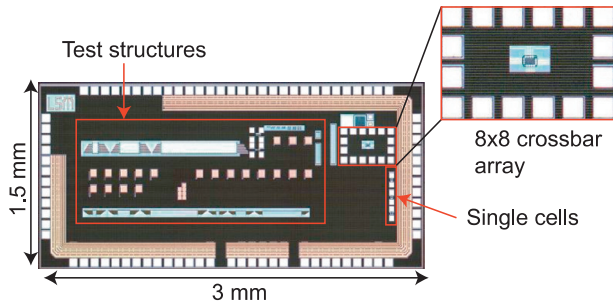
Fig. 4. Micrograph of the post-processed chip. Inset shows the 8×8 array of TaO_x devices together with probe pads.

Fig. 6. Fabrication process steps for the ReRAM–CMOS BEoL integration: (a) carrier wafer creation by Si DRIE. (b) CMOS chip assembly. (c) Parylene coating. (d) Parylene RIE. (e) Chip passivation opening. (f) TaO_x sputtering.

TiN barrier layer. There is indeed no difference between the cell materials of the integrated ReRAM cells and the single cross-points described in Section 2.

In Fig. 6, the main fabrication steps to fabricate ReRAMs cells are shown. A carrier wafer is fabricated from a standard 4″ Si wafer after a lithography process to delimit the area where the chip will be located. The notch to allocate the CMOS chip is created by a Si *Deep Reactive Ion Etching* (DRIE) corresponding to the chip high using a standard Bosh process with SF_6 and C_4F_8. It is critical to avoid any high mismatch between the Si wafer surface and the chip to improve the lithography resolution during the next steps. The chip is then placed in the created notch, which is designed to compensate the chip dicing tolerances. Fig. 7 shows a micrograph of the CMOS chip embedded into the carrier wafer. Within the chip it is possible to notice the array structures where the TaO_x-based memories will be created. The cavity containing the chip has one aperture to facilitate the chip release. After embedding the chip in the carrier wafer, a parylene deposition is performed to protect the chip and to ensure its mechanical stability during the subsequent processing steps. Afterwards, the parylene layer is selectively etched by O_2 RIE to open the areas where the TaO_x must be deposited. Subsequently, the chip passivation is opened down to Metal 5 (M5) by BHF wet etching and TaO_x is deposited by sputtering with the same deposition recipe used for the devices reported in Section 2. The parylene layer is then removed from the chip by lift-off. Fig. 8 depicts the Word line (in M6) and the Bit line (in M5) of an 8×8 crossbar memory array. The cross section A–A′ on the bottom right shows a single cross-point included between M5 (bottom electrode) and the MIM capacitor layer (top electrode). The deposited TaO_x creates a resistive switching layer between the boundaries of the top electrode (the MIM capacitor layer) and M5, as highlighted in the dashed circle.

The proposed integration method has the advantage of avoiding high-resolution lithography steps during the chip post-processing, reducing the total lithography step number and the process cost.
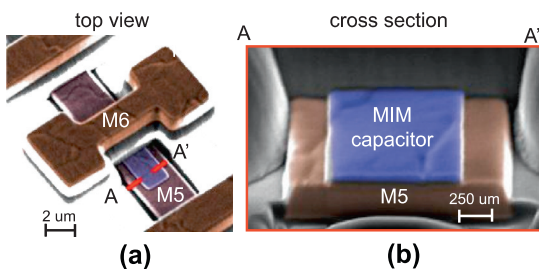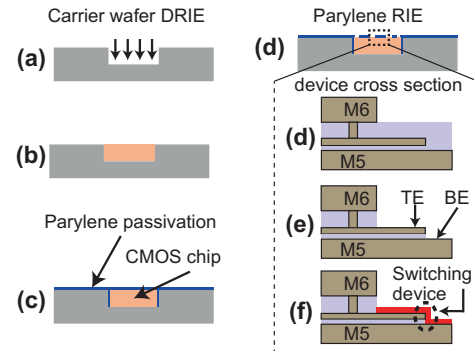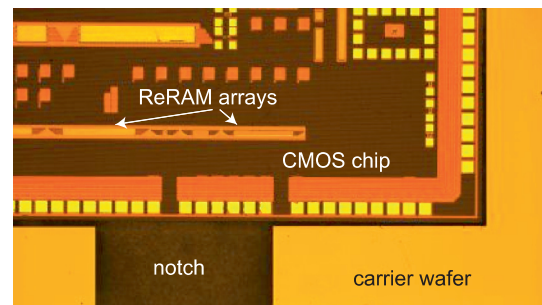

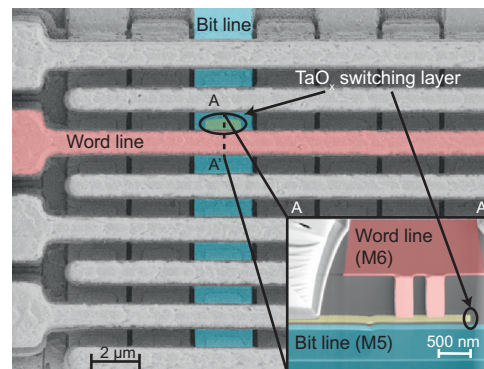**Fig. 7.** Micrograph of the carrier wafer with the embedded CMOS chip.

Fig. 8. Scanning electron micrograph of an 8×8 array with Word and Bit lines in M6 and M5 metal lines. Inset shows the cross-point where the TaO_x active region is located after the post-processing (at the edge between the MIM layer and the M5).

4. Electrical results

Electrical tests have been carried out as shown in Fig. 9. The measurement has been realized with the same test setup previously described in Section 2. A first read operation shows that the cells were already in the LRS (no current compliance has been set during this operation). The pristine resistance value is most probably caused by the square shape of the cell, which acts as a field enhancement structure, and because the TaO_x film, that is not embedded into a passivation layer, may have been reduced during the lift-off process. A DC voltage is applied between the top and the bottom electrode, sweeping between −1 V and 1.5 V. For the set operation a current compliance is applied with an external parameter analyzer. Low-voltage operation has been observed with a SET voltage of −1 V and a RESET voltage of +1.3 V. The
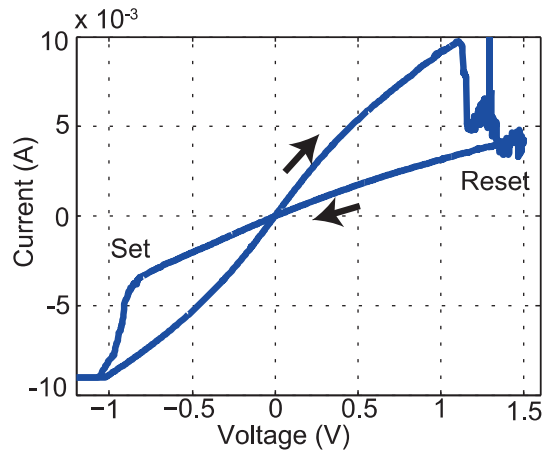

Fig. 5. (a) Scanning electron micrograph of the MIM capacitor. (b) FIB-SEM cross-section.

Fig. 9. Measured I–V curve of an integrated TaO_x-based ReRAM device. The low SET and RESET voltages are compatible with low-voltage applications.

voltage levels are compatible with the selected 180 nm CMOS technology. The 50 nm-thick integrated ReRAM has a LRS of 80 Ω and a HRS of 320 Ω, making it a potential candidate for applications such as non-volatile FPGAs, where multiplexers can be replaced by ReRAM devices [7].

The lower operating voltages of the integrated cell, compared to the standalone ones, are caused by the field enhancement effect of the MIM capacitor shape. The relatively low obtained resistance levels are originated by the pristine LRS value. Because the LRS has a quite small value, the cells, that have high current leakage, are difficult to reset properly. Moreover, in the integrated cell I–V curve reported in Fig. 9 it is possible to notice some current fluctuations during the reset process. Although the origin of these phenomena is not clear yet, recent interpretations [12] suggest that it is related to the generation of oxygen vacancies under the effect of a high electrical field and high currents. These events occur predominantly in the early reset phase, when the distance between the cathode and the filament is still modest. The nature of these phenomena appears to be stochastic: they are present just for some reset operations. The influence of the MIM capacitor (3.6 fF for the measured devices), which is electrically in parallel to the ReRAM cell, is negligible for the obtained measurements. The RC time constant is indeed less than 1 ps, which is much

smaller than the performed IV sweep step time and smaller than any ReRAM switching time reported in the literature. Furthermore the set operation, which is the operation that suffers more from the capacitor effect due to the transition from HRS to LRS, does not present any sharp resistance change. Consequently the capacitor can be discharged gradually avoiding over-forming effects.

5. Conclusion

$TiN/TaO_x/TiN$-based memories have been fabricated and characterized, targeting CMOS compatibility and low-voltage operations. We demonstrated a low cost technique for heterogeneous integration of ReRAM devices in the BEoL of fully processed 180 nm CMOS chip. We applied the presented integration scheme to fabricate standalone memories and achieved ReRAM–CMOS post-processing integration. The measured integrated ReRAMs show characteristics with a SET voltage and RESET voltage of -1 V and $+1.3$ V respectively.

Acknowledgments

References

[1] A.C. Torrezan, J.P. Strachan, G. Medeiros-Ribeiro, S. Williams, Nanotechnology 22 (48) (2011).
[2] M.-J. Lee, C.B. Lee, D. Lee, S.R. Lee, et al., Nat. Mater. 10 (2011) 625–630.
[3] B. Govoreanu, G.S. Kar, Y.-Y. Chen, V. Paraschiv, et al., IEDM (2011) 729–732.
[4] T.-Y. Liu, T.H. Yan, R. Scheuerlein, Y. Chen, et al., ISSCC se. 12 (2013) 210–211.
[5] R. Fackental, M. Kitagawa, W. Otsuke, K. Prall, et al., ISSCC se. 19 (2014) 338–339.
[6] S. Park, A. Sheri, J. Kim, J. Noh, et al., IEDM (2013) 625–628.
[7] P.-E. Gaillardon, D. Sacchetto, G.B. Beneventi, M.H. Benjamaa, et al., IEEE Trans. Nanotechnol. 12 (1) (2013).
[8] Y.-B. Kim, Se R. Lee, D. Lee, C.B. Lee et al., in: Symp. of VLSI Tech. Dig. of Tech. Papers, 2012, pp. 52–53.
[9] H.-S.P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, et al., Proc. IEEE 100 (6) (2012) 1951–1970.
[10] X. Guan, S. Yu, H.-S. Wong, IEEE Electron Device Lett. 33 (10) (2012) 1405–1407.
[11] Z. Wei, Y. Kanzawa, K. Arita, Y. Katoh, et al., IEDM (2008) 1–4.
[12] S. Yu, X. Guan, H.-S.P. Wong, IEEE Trans. Electron Device 49 (4) (2012) 1183–1188.