Understanding Workfunction Tuning in HKMG by Lanthanum Diffusion Combining Simulations and Measurements

A. Spessot, C. Caillat, P. Fazan
Micron Technology Belgium BVBA
Kapeldreef 75, 3001 Leuven, Belgium
aspessot@micron.com

R. Ritzenthaler, T. Schram
imec, Kapeldreef 75, 3001 Leuven, Belgium

***Abstract*— In this paper, we propose a method to extract effective diffusion coefficients for Lanthanum in HfO_2 for an HKMG technology. TCAD diffusion simulations is combined to the analysis of theoretically expected Work Function shift due to Lanthanum at the HfO_2/SiO_2 interface and experimentally extracted Work Function value under various thermal budgets, obtaining a good agreement between simulations and experimental data.**

***Keywords*— HKMG, Diffusion, Lanthanum, Work Function Gate Stack**

I. INTRODUCTION

With the widespread introduction of High-K / Metal Gate (HK/MG) stacks in advanced CMOS flows [1], arises an increasing need for a refined characterization of physical mechanisms at work with HK Work Function (WF) shifters such as La, Y or Al [2]. In this type of approach, a WF shift is achieved by diffusing such atoms into the thin HK in order to create dipoles close to the channel (at the SiO2 interlayer interface), shifting WF and hence V_T. Thus, guaranteeing a good V_T control implies a good control of this type of process, justifying the growing interest for detailed mechanisms descriptions [3-6].

In this context, we have experimentally studied the diffusion of La into a thin (2nm) HfO_2 layer standing on a SiO_2 interfacial layer, by applying series of thermal budgets and measuring their impact on the electrical WF extracted on large MOS capacitors. Since such thin layers prevent from performing reliable quantitative physical analysis such as SIMS, extracting diffusion coefficients is challenging.

II. EXPERIMENTAL DATA

Fig 1 shows a sketch of the experimental devices used for the WF extraction. The gate stack consists of SiO_2 (1 nm)\ HfO_2 (2 nm)\La cap (1 nm)\Metal Gate (5 nm). Five splits (2 wafers per split) were processed by adding on top of the standard process flow the additional thermal budgets as shown in Table 1. The applied additional thermal budgets were longer than 60 minutes and in the range of 600°C to 750°C, mimicking real conditions required by a DRAM process flow [7,8]. The initial La concentration in the capping layer was

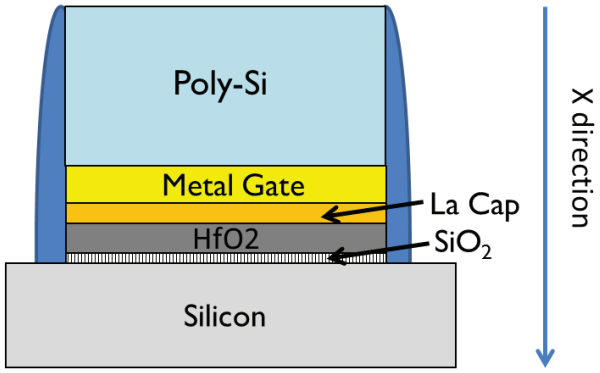

Fig 1: Sketch of the experimental devices used for the WF extraction. The gate stack consists of SiO_2 (1 nm) \ HfO_2 (2 nm) \ La cap (1 nm) \ Metal Gate (5 nm).

estimated to be 8.8E21 at/cm^3 by XRD measurements on various blanket wafers with different deposited La thickness. The WF is extracted measuring the C-V curves on large capacitors (50 µm x 50 µm) and then using the CVC method [9]. Fig 2 shows the variation in the Work Function measured with respect to the case of no La cap.

	Step	Temp (ºC)	Duration [x min]			Split #
			< 5	5<x<60	>60	
Common	Spacers Formation	750	x			
Common	Junction Anneal	1050	x			
Common	Forming anneal + Sintering Anneal	450		x		
Common	Thermal Budget 1	600			x	1
Optional	Thermal Budget 2	700			x	2
Optional	Thermal Budget 3	700			x	3
Optional	Thermal Budget 4	750			x	4
Optional	Thermal Budget 5 (1+2+3+4)	600-750			x	5

Table 1: List of the main thermal budgets applied in the process flow. On top of standard steps common to the various wafers (highlighted in orange), specific additional steps were applied (in yellow). The split column shows the label of the splits

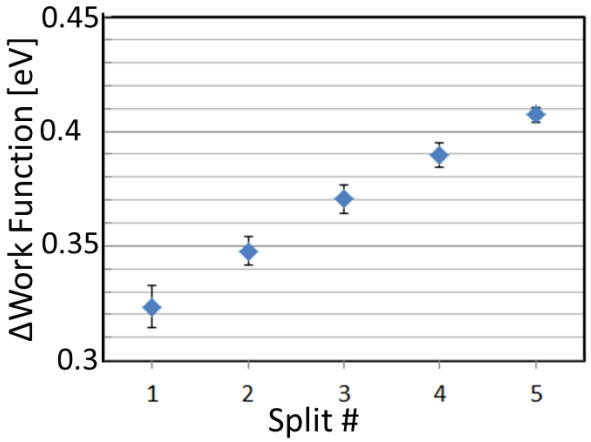

Fig 2: Experimental shift of the Work Function [eV] measured with respect to the case of no La cap. The split description corresponds to Table 1. Each point shows the median of the distribution of the dies measured on two wafers per split.

III. Methodological and Theoretical Background

Theoretical calculations using ab-initio simulations can predict the band offset induced by substitutional La atoms in a HfO_2 lattice (a La atom replacing a Hf atom– see figure 3) at the HK/SiO_2 interface [3].

The origin of this band offset lies in the creation of dipoles due to a relative difference of electronegativity of La vs. Hf [4]. Thus, it is predicted in [3] that 50% substitutional La induces a WF shift of -0.5eV (i.e. towards more N-type). Taking this figure and considering the local concentration of Hf atom in a monoclinic HfO_2 mesh at the interface, we will assume in the following an effect of La concentration to WF shift of α_{La}=1.466E22 at.cm^{-3}/eV. We will also assume that the range of WF shift experimentally observed is close enough to the 50% substitutional La case of [1] for the effect to be considered as proportional to the dose (Table 2).

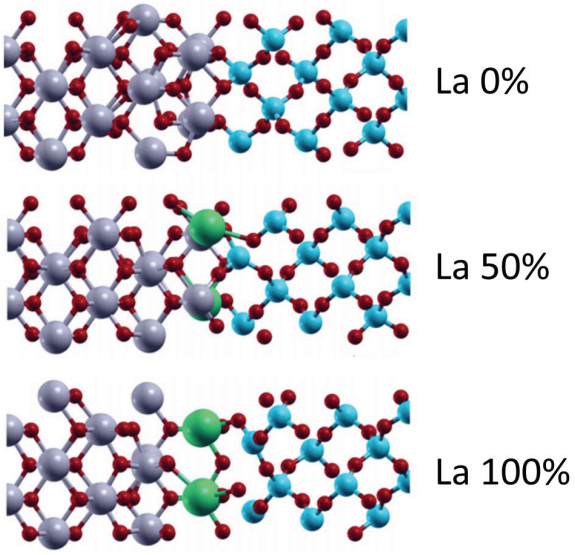

Fig 3:Relaxed Structure of HfO2/SiO2 for 0%, 50% and 100% La substitutional cases. We used in this work the 50% configuration as a starting point for WF shift. The illustrations are taken from [3]

Split #	ΔWF (eV)	Corresponding La concentration(*) (at/cm3)
1	0.321	4.71E+21
2	0.346	5.07E+21
3	0.365	5.36E+21
4	0.388	5.69E+21
5	0.407	5.97E+21

(*) Expected effect of 100% of substitutional La: 1.466E22 at.cm-3/eV

Table 2: Expected concentration at the HfO_2/SiO_2 interface for the measured WF shift.

Thus, we can calculate the expected concentration variation at the HfO_2/SiO_2 interface for the observed WF shift range (Table 2).

IV. Simulation Methodology

Starting from an existing deck calibrated for a standard process flow under Sentaurus[10], we derived a simpler 1D model to simulate the La diffusion in the gate stack. The stack was modified as follows: a thick high-K layer of HfO_2 (20nm) on top of SiO_2 and Silicon was used for the gate stack; a deposited layer with uniform concentration of La covered by a Metal Gate layer was used for the La cap. To determine the La diffusion coefficients in HfO_2, TCAD simulations corresponding to the various thermal budgets have been performed, as detailed in Table 1. The La diffusion was simulated using the constant diffusion model and assuming La diffusion as governed by the Fick equation (Eq. 1):

$$\frac{\partial La}{\partial t} = D\frac{\partial^2 La}{\partial x^2} \qquad \textit{(Eq. 1)}$$

where D is the intrinsic diffusivity of the La, calculated assuming the Arrhenius law (*Eq. 2*).

$$D = D_0 e^{-(E_A/k_B T)} \qquad \textit{(Eq. 2)}$$

It was observed that the diffusivity equation is not able to fully explain the diffusion saturation in the La diffusivity when high temperatures are applied (up to 1080°C) [6]; however, due to the lower temperature used here for the additional thermal budgets, we expect that the saturation does not play a significant role.

Due to the finiteness of the La source, it is expected that after the common thermal budget of the Table 1 the initial La concentration in the cap is reduced due to the diffusion of La in both direction, downward toward the high-K and upwards toward the Metal Layer. This concept is shown in Fig. 4: at the beginning La is confined in the cap (Fig 4a), while the additional thermal budgets cause La diffusion (Fig 4b). As a result, both high-k and Metal Gate will contain La, while the cap itself is depleted (Fig 4c). To properly account for the competitive diffusion of La in the Metal Gate and the consequent reduction of the La concentration in the capping layer, we have assumed a finite La source during the simulation of the standard thermal budgets of Table 1. Then,

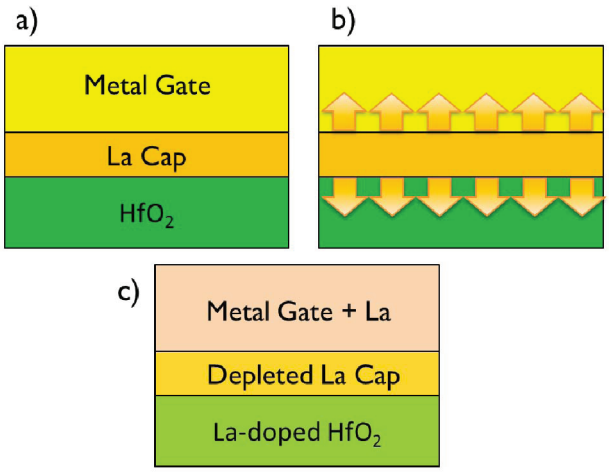

Fig 4: Sketch of the simulated La diffusion. a) Initial configuration before thermal anneals, with HfO_2/La cap/Metal Gate b) During the thermal treatments of Table 1, the La diffuses from the cap downward and upward. c) After the La diffusion, the La cap is depleted, and both high-K and Metal Gate contain La.

we used the depletion of the initial La concentration in the Metal Gate/La cap after the thermal diffusion as a fitting parameter (D_E), keeping the concentration of La at the high-K/La cap interface constant for the additional anneals.

Fig 5 shows a sketch of the simulated model we have used. The purple curve shows the initial La concentration as deposited in the capping layer. Two different La concentration profiles have been simulated for the split 1 and 5 of Table 2, calculated assuming the same pair of diffusion coefficients and the same depletion level. It is worth noting that the higher the thermal budget, the higher is the simulated La concentration in the HfO_2, in particular at 2 nm from the high-K/La cap interface (identified by line (a)), where we should expect the real interface with SiO_2.

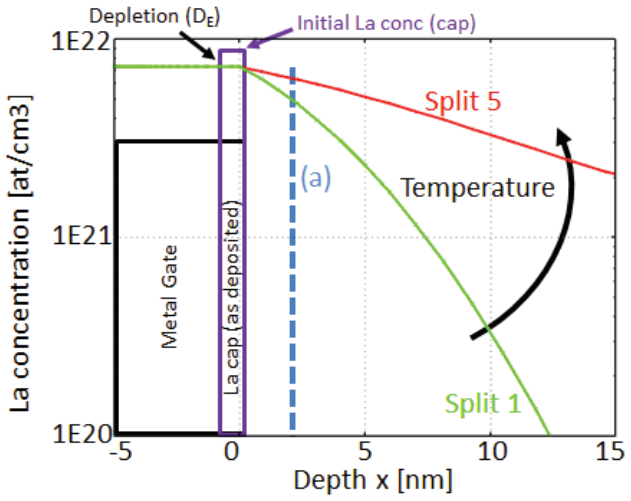

Fig 5: Example of the simulated model. The purple curve shows the initial La concentration as deposited in the capping layer. Two different La concentration profiles were simulated for the split 1 and 5 of Table 2, calculated assuming the same coefficients (D_E=16%, D_0=9.14E-12cm^2/s, E_A=1.05eV). Higher thermal budget gives higher simulated La concentration in the high-K, in particular at 2 nm from the high-K/La cap interface (identified by line (a)).

V. Results and Discussion

Fig 6 shows the expected La concentration determined at 2 nm from the high-K/La cap interface, corresponding to the line (a) of Fig 5, for the various splits. The experimental values are evaluated using the expected effect α_{La} and the estimated concentration range at HfO_2/SiO_2 interface (Table 2) as boundary conditions. The simulated curves are calculated according to the method presented in Section IV and using D_0, E_A and D_E as variable parameters. Fitting the experimental points, we were able to determine the best pair of diffusion coefficients (D_0, E_A) and the depletion level D_E of La. The resulting simulations are shown by the empty circles of Fig 6. To illustrate the possible simulated trends, two other sets of values are included in the graphs, corresponding to higher (red triangles) and lower La diffusivity (blue circles), respectively.

It is worth noting that to obtain a La concentration at the high-K/SiO_2 interface compatible with the measured WF and with the initial La cap concentration of 8.8E21 at/cm^3, we had to assume a depletion of La source of about 16%. Another consideration is related to the experimental range of La concentration: it is confined between 4.7E21 and 6E21 at/cm^3, implying that the expected concentration of La at 2 nm is between ~50% and ~70% of the initial La at the interface between high-K and La cap before the thermal diffusion. As a final note, we should mention that the observed depletion of the active La dose diffusing in the high-K could also account for La segregation taking place at the interface between the capping layer and the high-K.

VI. Conclusions

We propose a method to match the simulation of La diffusion in the high-K to experimental WF measurements. A good agreement between the experimental value of the La concentration at the high-K/SiO_2 interface and the corresponding simulations was obtained for different thermal budgets, using the La diffusion coefficients and the depletion level of La concentration in the Cap/Metal Gate stack as fitting parameters.

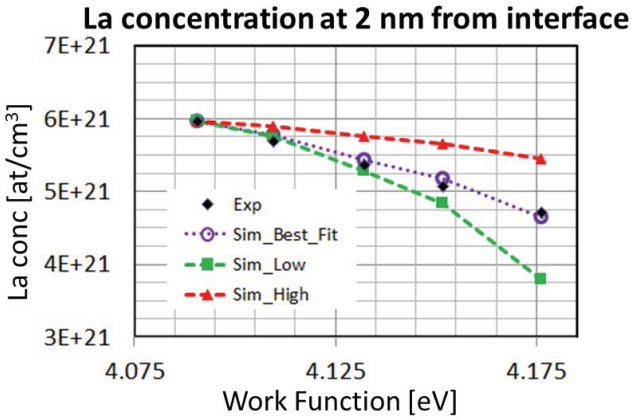

Fig 6: Expected La concentration at 2 nm from the high-K/La cap interface, calculated according to the experimental Work Function shift of Table 2. The corresponding simulations obtained by the best fit (D_E=16%, E_A=1.05eV, D_0=9.14E-12) are shown (open purple circles). Red triangles and Green points show two other simulations for the case of lower (E_A=1.2eV, D_0=5E-12) and higher diffusivity of La (E_A=0.8eV, D_0=5E-14), respectively

ACKNOWLEDGMENT

The authors would like to acknowledge Sergiu Clima and An De Keersgieter of imec and Augusto Benvenuti from Micron for the valuable discussions and the imec's core partners program.

REFERENCES

[1] Mistry, K. *et al,* "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging"*, IEDM Tech Dig.* p 247 (2007).

[2] M. M. Frank, "High-K/Metal Gate Innovations Enabling Continued CMOS scaling", *Proc ESSDERC*, p. 25 (2011).

[3] J. A. Caraveo-Frescas, H. Wang, U. Schwingenschlögl, and H. N. Alshareef, "Experimental and theoretical investigation of the effect of SiO2 content in gate dielectrics on work function shift induced by nanoscale capping layers", Appl. Phys. Lett. Vol 101, p 112902 (2012).

[4] L. Lin and J. Robertson "Atomic mechanism of dipole formed at high-K: SiO2 interface", JAP Vol 109, Pag 094502 (2011).

[5] E. Martinez *et al.*, "Lanthanum diffusion in the TiN/LaOx/HfSiO/SiO2/Si stack", Microelectronic Engineering Vol 88 pp 1349–1352 (2011).

[6] Z. Essa *et al.*, "Evaluation and modeling of lanthanum diffusion in TiN/La2O3/HfSiON/SiO2/Si high-k stacks", Appl. Phys. Lett. Vol 101, p 182901 (2012).

[7] C. Ortolland *et al.*, "Ion-implantation-based low-cost HK/MG process for CMOS low-power application", *Proc Symp VLSI*, pp. 185-186 (2010).

[8] C.N. Ni *et al*, "Gate-First TiAlN P-Gate Electrode for Cost Effective High-k Metal Gate Implementation", *Proc Symp VLSI-TSA* (2012).

[9] J. R. Hauser and K. Ahmed, "Characterization of ultra-thin oxides using electrical C-V and I-V measurements" *AIP Conf. Proc.*, pp. 235-239 (1998).

[10] Synopsys, Zurich, Switzerland, "Sentaurus Process User Guide" Version G-2012.06 ed., 2012.