Low Thermal Budget Dual-Dipole Gate Stacks Engineered for Sufficient BTI Reliability in Novel Integration Schemes

J. Franco*, Z. Wu[1], G. Rzepa[2], A. Vandooren, H. Arimura, D. Claes, N. Horiguchi, N. Collaert, D. Linten, T. Grasser[2], B. Kaczer

imec, Leuven – Belgium, [1]also at KU Leuven – Belgium , [2]TU Wien – Austria *Jacopo.Franco@imec.be

(Invited paper)

Abstract

Low thermal budget gate stacks will be required for novel integration schemes, such as 3D sequential stacking of CMOS tiers. We study the impact of a reduced thermal budget on BTI reliability, and we demonstrate interface dipole engineering to suppress the carrier-defect interaction and achieve sufficient nMOS PBTI and pMOS NBTI reliability without the customary post-deposition anneals.

Introduction

3D sequential integration is a recently envisioned approach to increase CMOS functionality per die area by stacking transistors on top of each other (**Fig. 1**), or to co-integrate heterogeneous technologies on multiple tiers of the same wafer within a single fabrication flow [1-3]. Thermal budget management represents the most crucial challenge of this approach: thermal steps for the fabrication of the top tier should be limited to preserve the functionality of the bottom tier devices and interconnects (BEOL). Gate stack reliability is extremely sensitive to process temperature, particularly for high-k/metal gate (HKMG) technologies. In a gate-first integration flow, the high-k layer is exposed to the source/drain (S/D) activation anneal (~1000℃), which reduces the dielectric defect densities and minimizes Bias Temperature Instability (BTI) [4]. In more contemporary Replacement Gate (RMG) flows, a dedicated high-temperature 'reliability' anneal (~900℃) is typically performed after the deposition of the final gate stack to ensure sufficient stability [5,6]. Such high temperature steps are not suitable for top tier fabrication as they would degrade the BEOL of the bottom tier.

In this paper, we discuss first the impact of a limited thermal budget on the BTI reliability of the HKMG stack. By using our recently introduced physics-based BTI modeling framework Comphy ("Compact Physical", **Fig. 2** [7]), we compare the oxide defect properties in an as-deposited low thermal budget SiO_2/HfO_2/TiN gate stack with the same gate stack exposed to a typical 'reliability' anneal, and with a commercial high-k-first 28nm HKMG technology, revealing larger oxide defect density at energy levels around the Si channel band edges. We then use the pathfinder predictive capabilities of Comphy to assess the impact of dipoles insertions at the interface between SiO_2 and HfO_2 (by depositing a thin dipole-former interlayer with ALD) to 'shift up' or 'down' the energy levels of the high-k defects with respect to the Si conduction or valence band, for improved nMOS and pMOS reliability. This approach is then experimentally demonstrated on hardware.

The interface dipole-based low thermal budget gate stacks discussed here open up the reliable use of high-k gate stacks for 3D sequential integration, but also for other novel integration concepts as, e.g., the embedding of thin film transistors in the BEOL, or to conveniently re-arrange the integration flow of standard Si CMOS which is currently dictated by the highest temperature steps, e.g., the S/D contacts are fabricated only after the gate stack 'reliability' anneal instead of directly after the S/D epitaxy.

Impact of low thermal budget on HKMG reliability

Top tier devices which did not receive any high temperature anneal after the HKMG deposition show poor PBTI and NBTI reliability (**Fig. 3**): extremely large threshold voltage shifts (ΔV_{th}: ~20× and ~10× larger than the acceptable target for nMOS and pMOS, respectively) and detrimentally weak BTI voltage accelerations are observed, in contrast to the same gate stack exposed to a ~900℃-1s 'reliability' anneal, and to a commercial 28nm HKMG technology.

To understand the origin of the poor top tier reliability, we used Comphy [7] to model the ΔV_{th} measured in nMOS and pMOS devices subjected to a complex stress/recovery waveform [8], comprising various gate voltages applied at different temperatures (**Fig. 4**). An excellent match of the modeled BTI kinetics to the experimental data is achieved by properly calibrating shallow and deep defect band properties for both the SiO_2 and HfO_2 dielectric; these defect properties are compared in **Table I** to the ones extracted in [7] from BTI measurements of a commercial 28 nm technology. The reduced thermal budget results in a ~2× larger defect density in HfO_2, and especially in a significantly reduced mean trap energy level and mean thermal barrier for charge capture for both the SiO_2 and HfO_2 traps, possibly due to a less stiff, more disordered amorphous oxide.

978-1-5386-6508-4/19/$31.00 ©2019 IEEE

Dual interface dipole-based CMOS gate stacks for sufficient BTI reliability at low thermal budget

A. Comphy pathfinding study: nMOS and pMOS

Properly engineered interface dipoles can 'shift up' the shallow high-k defect band w.r.t. the Si conduction band (for improved PBTI), or 'shift down' the deep defect band w.r.t. the Si valence band (for possibly improved NBTI), and therefore reduce the density of accessible defects towards a level comparable to the high temperature HKMG stack. Comphy simulations show that a dipole at the SiO_2/HfO_2 interface is very effective to improve PBTI: a ΔV_{th}<50mV is projected for 10-year operation at V_{ov}=0.7V at 125℃ if the high-k shallow defect energy is increased by 0.4eV (**Fig. 5**). This improvement is virtually independent of the SiO_2 thickness, as PBTI is mostly controlled by high-k electron traps (SiO_2 electron traps are negligible, cf. Table I). Interestingly, Comphy simulations show that a similar approach can be effective also for improving NBTI, despite that in pMOS this mechanism is commonly ascribed to near-interface traps: in a low thermal budget HKMG stack with a thin SiO_2 IL, the deep defects in HfO_2 have a sufficiently high density to surpass the SiO_2 hole traps (**Fig. 6a**, dashed vs. dotted lines), and therefore a proper interface dipole shift can reduce the overall charge trapping to a tolerable level. However, if a standard thick SiO_2 IL would be used (**Fig. 6b**, 1nm vs. 0.6nm), the total density of hole traps in the larger SiO_2 volume would remain dominant, defeating the effectiveness of an interface dipole at the SiO_2/HfO_2 interface for improving the overall NBTI.

B. Experimental demonstration

A ~0.3nm thin $LaSiO_x$ layer ALD-deposited between SiO_2 and HfO_2 can induce a ~0.4eV negative V_{fb} shift (**Fig. 7a**) [9,10], and therefore tune the nMOS V_{th} towards low values without requiring complex low work function metal stacks. The PBTI ΔV_{th} measured on these $LaSiO_x$-inserted nMOSFETs is significantly reduced (~8×) w.r.t. the reference low thermal budget "top tier" gate stack, in excellent quantitative agreement with the Comphy prediction (**Fig. 7b**).

To experimentally verify the effectiveness of interface dipoles for NBTI reliability, Al_2O_3 can be used to form a negative dipole on SiO_2 [11]. In order to achieve a sufficient interface dipole density to induce a ~0.4eV shift, a ~1nm thick Al_2O_3 layer is necessary, (**Fig. 8a**, quantitatively in-line with [11]). Additional trapping in such a thick layer might defeat the reliability improvement strategy. However, our previous BTI study of bulk Al_2O_3 oxide defects suggests a negligible trap density at energies close to the Si valence band, making Al_2O_3 suitable for pMOS use (while its wide defect distribution around the Si conduction band makes it unsuitable for nMOSFETs [12]). Furthermore, a sizeable dipole shift is observed by inserting Al_2O_3 layers as thin as 0.2nm, with marginal EOT penalty.

By depositing Al_2O_3 on SiO_2 before HfO_2, a NBTI ΔV_{th} reduction up to ~10× is demonstrated (**Fig. 8**). In particular, a stronger NBTI voltage acceleration is observed when inserting Al_2O_3, quantitatively in line with the Comphy prediction (solid lines), confirming an effective decoupling of the defect energy level w.r.t. the channel Fermi level [12]. A ~0.2nm thin Al_2O_3, inducing ~0.24eV of dipole shift, is already sufficient to bring NBTI ΔV_{th} within specs.

The proposed dual dipole-based stacks are shown in Fig. 9 to enhance the PBTI and NBTI reliability of the as-deposited high-k gate stack sufficiently to match the maximum operating V_{ov} of a standard high-k gate stack exposed to the high temperature reliability anneal. For the pMOS gate stack, the reliability improvement is observed only at scaled SiO_2 IL (**Fig. 9b**), in line with the Comphy prediction (cf. Fig. 6). For both the pMOS and nMOS gate stacks, the marginal EOT penalty induced by the insertion of the dipole forming layers can be compensated by slightly scaling the SiO_2 IL, still maintaining sufficient reliability thanks to the achieved defect level decoupling.

Conclusions

We have demonstrated the use of ALD dipole-forming interlayers ($LaSiO_x$ for nMOS and Al_2O_3 for pMOS) to engineer the energy alignment of the high-k defect levels w.r.t. the Si channel band edges, and achieve sufficient BTI reliability in low thermal budget gate stacks, relevant for novel integration schemes such as 3D sequential stacking. This strategy was identified by using the pathfinding predictive capabilities of the imec/T.U. Vienna BTI simulation framework Comphy, and were demonstrated experimentally on Si hardware.

References

[1] P. Batude *et al.*, in *Proc*. VLSI Tech. 2015;
[2] A. Vandooren *et al.*, in *Proc*. VLSI Tech. 2018;
[3] A. Mallik *et al.*, in *Proc*. IEDM 2017;
[4] H. Arimura *et al.*, in *Proc*. IRPS 2014;
[5] G. Rzepa *et al.*, in *Proc*. IRPS 2017;
[6] B. Linder *et al.*, in *Proc*. IRPS 2016;
[7] G. Rzepa *et al.*, in *Micr. Rel.* 85, pp. 49-65 , 2018;
[8] Z. Wu *et al.*, submitted for IRPS 2019;
[9] J. Franco *et al.*, in *Proc*. IRPS 2017;
[10] H. Arimura *et al.*, in *Proc*. IEDM 2016;
[11] Y. Kamimuta *et al.*, in *Proc*. IEDM 2007;
[12] J. Franco *et al.*, in *Proc*. IEDM 2017.

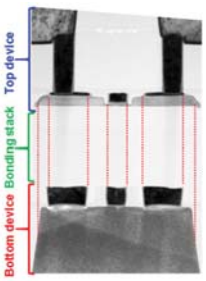

Fig. 1: TEM of a 3D structure showing stacked top and bottom tier devices with nanometric alignment [2].

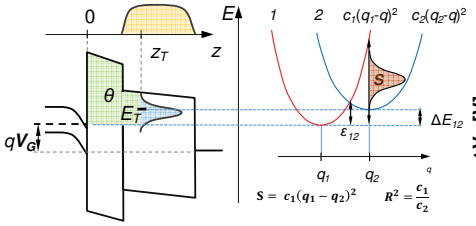

Fig. 2: Schematic of the physics-based BTI model implemented in the imec/T.U. Vienna simulation framework Comphy [7]: each defect band in the dielectric layers is represented as a Normal distribution ($\langle E_t\rangle \pm \sigma_{Et}$) of defect levels and a density N_t, assumed uniform in space. Each defect is represented as a two-well system with distributed parameters ($\langle S\rangle \pm \sigma_S$, R) describing the activation energy of the charge capture and emission processes.

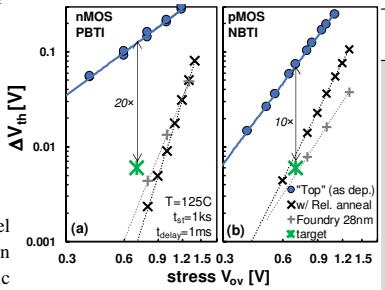

Fig. 3: Low thermal budget HKMG stack ("top") shows poor **(a)** PBTI and, **(b)** NBTI compared to the same stack with a high temperature 'reliability' anneal, or compared to a 28nm commercial technology [5,7].

Table I

		Foundry 28 nm	Low Thermal Budget
Shallow SiO_2	N_t [/cm³]	1.5×10^{19}	1.17×10^{19}
	$\langle E_t\rangle \pm \sigma_{Et}$ [eV]	1.13 ± 0.15	$\mathbf{0.55} \pm 0.11$
	$\langle S\rangle \pm \sigma_S$ [eV]	3.82 ± 1.36	$\mathbf{2.66} \pm 1.07$
	R [1]	0.407	0.852
Shallow HfO_2	N_t [/cm³]	5.52×10^{20}	2.56×10^{20}
	$\langle E_t\rangle \pm \sigma_{Et}$ [eV]	1.2 ± 0.156	$\mathbf{0.5} \pm 0.105$
	$\langle S\rangle \pm \sigma_S$ [eV]	3.19 ± 0.77	$\mathbf{2.34} \pm 0.7$
	R [1]	0.587	0.556
Deep SiO_2	N_t [/cm³]	1.42×10^{20}	$\mathbf{2.05 \times 10^{20}}$
	$\langle E_t\rangle \pm \sigma_{Et}$ [eV]	-1.26 ± 0.25	$\mathbf{-0.92} \pm 0.23$
	$\langle S\rangle \pm \sigma_S$ [eV]	5.63 ± 2.67	5.49 ± 1.88
	R [1]	1.8	1.15
Deep HfO_2	N_t [/cm³]	2.95×10^{20}	$\mathbf{5.17 \times 10^{20}}$
	$\langle E_t\rangle \pm \sigma_{Et}$ [eV]	-0.17 ± 0.14	$\mathbf{-0.6} \pm 0.1$
	$\langle S\rangle \pm \sigma_S$ [eV]	6.5 ± 1.99	6.08 ± 1.53
	R [1]	0.514	1.58

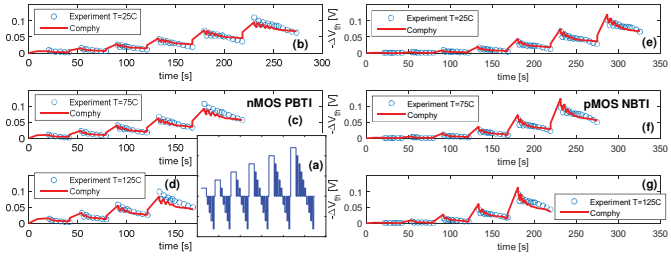

Fig. 4 (left): Defect models (see parameters in Table I) were calibrated in Comphy [7] to excellently reproduce the nMOS PBTI (b,c,d) and (e,f,g) pMOS NBTI kinetics in a variety of conditions included in **(a)** complex waveform [8] comprising increasing stress voltages and decreasing discharge voltages, performed at **(b,e)** 25ºC, **(c,f)** 75ºC, **(d,g)** 125ºC.

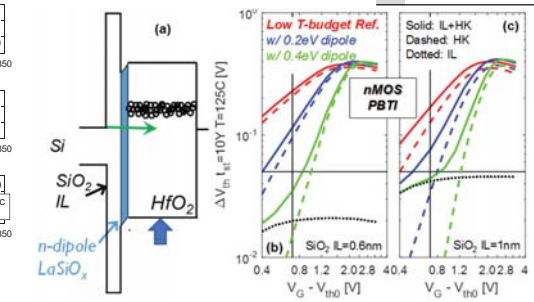

Fig. 5: **(a)** Band diagram of the proposed nMOS gate stack: a thin LaSiOx layer can be deposited on SiO_2 to form a dipole and thus shift the HfO_2 bandgap up to 0.4eV w.r.t. Si. **(b, c)** Comphy simulations of PBTI ΔV_{th} (10Y,125ºC) for increasing V_{ov} (defect model for low thermal budget, cf. Table I). The impact of a dipole at the SiO_2/HfO_2 interface is emulated by shifting up the energy level of the high-k traps by 0.2 and 0.4eV. The dashed and dotted lines show the respective contributions of HfO_2 and SiO_2 traps. Two scenarios are considered **(b)** 0.6nm thin SiO_2 IL, **(c)** conventional 1nm SiO_2 IL. In both cases, a 0.4eV dipole is predicted to maintain PBTI ΔV_{th}(10Y,125ºC) within a 50mV target at V_{ov}=0.7V.

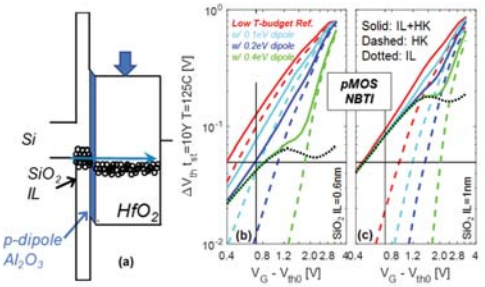

Fig. 6: Same as Fig. 6, for pMOS: **(a)** band diagram of the proposed gate stack: Al_2O_3 can be deposited on SiO_2 to form an opposite dipole. The dipole impact is emulated by shifting the energy level of the HfO_2 deep traps down by 0.1, 0.2 and 0.4eV, for **(b)** a 0.6nm SiO_2 IL, and **(c)** a 1nm SiO_2 IL. A dipole shift is effective in improving NBTI only in the former case, while in the latter the reliability is limited by hole traps in the 1nm SiO_2 IL. (Note: the kink in the contribution of SiO_2 traps is due to gate interaction on trap occupancy at high V_G).

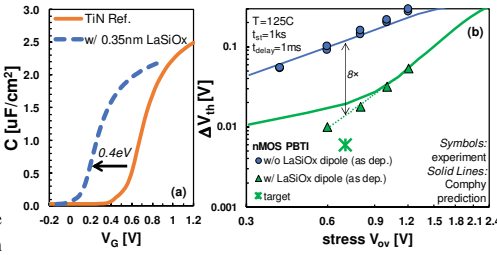

Fig. 7: **(a)** C-V curves of *n*-substrate MOS capacitors w/o and w/ $LaSiO_x$ interlayer, documenting the effective work function shift induced by the dipole. **(b)** Measured PBTI ΔV_{th} (t_{st}=1ks, 125ºC) for increasing V_{ov} on the 'as-dep.' low thermal budget gate stack, w/o and w/ the $LaSiO_x$ layer. A ~8× ΔV_{th} reduction is observed, quantitatively in line with the Comphy prediction based on dipole shift (solid lines).

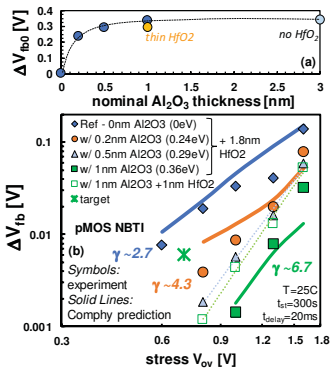

Fig. 8: **(a)** Measured V_{fb} tuning for increasing thicknesses of Al_2O_3 between SiO_2 and HfO_2. **(b)** NBTI shifts for increasing V_{ov} (t_{st}=300s, T=25ºC) measured on the low thermal budget gate stack w/o and w/ Al_2O_3 interlayers of increasing thicknesses. The insertion of a ~0.2nm Al_2O_3 layer brings the reliability within specs. Further improvement is obtained for larger dipole shifts, quantitatively in line with the Comphy prediction (solid lines).

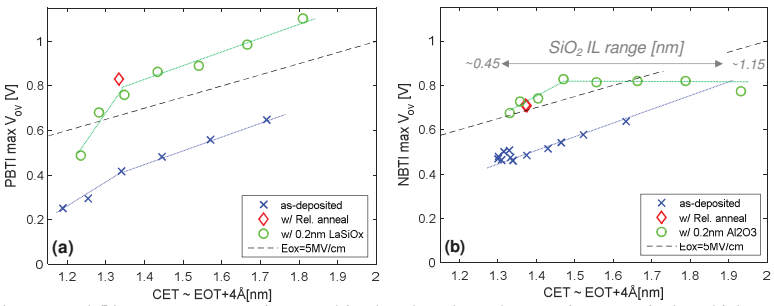

Fig. 9: **(a)** PBTI and **(b)** NBTI max operating overdrive benchmark vs. the capacitance equivalent thickness of each gate stack. Three HKMG stacks are compared: as-deposited (low thermal budget), w/ a typical reliability anneal (~900C-1s; high thermal budget), or w/ a dipole former interlayer deposited between the SiO_2 IL and the HfO_2 (low thermal budget). Different SiO_2 IL thicknesses, ranging between ~0.45nm and ~1.15nm were fabricated to test the dipole effect on nMOS and pMOS. In the former case, the dipole shift induced by a 0.2nm $LaSiO_x$ interlayer brings the PBTI reliability on par with the high thermal budget ref. (max E_{ox}>5MV/cm) irrespective of the SiO_2 thickness. In the pMOS case, the dipole shift induced by a 0.2nm Al_2O_3 bridges the reliability gap w.r.t. the high thermal budget process if combined with a thin SiO_2 IL; with a SiO_2 IL thicker than 1nm the dipole-induced benefit on NBTI fades out, as correctly predicted by Comphy (cf. Fig. 6c).