Demonstration of an InGaAs gate stack with sufficient PBTI reliability by thermal budget optimization, nitridation, high-k material choice, and interface dipole

J. Franco, A. Vais[1], S. Sioncke, V. Putcha[1], B. Kaczer, B.-S. Shie, X. Shi, R. Mahlouji, L. Nyns, D. Zhou, N. Waldron, J.W. Maes[2], Q. Xie[2], M. Givens[3], F. Tang[3], X. Jiang[3], H. Arimura, T. Schram, L.-Å Ragnarsson, A. Sibaja Hernandez, G. Hellings, N. Horiguchi, M. Heyns[1], G. Groeseneken[1], D. Linten, N. Collaert, A. Thean imec, Kapeldreef 75, 3001 Leuven – Belgium;
[1]also at KU Leuven; [2]ASM Belgium; [3]ASM America, Phoenix, AZ, USA e-mail: Jacopo.Franco@imec.be

Introduction: Remarkable performance has been recently demonstrated with high-mobility IIIV channels in experimental devices fabricated on 2" InP substrates [1,2], and in advanced VLSI-compatible finFET [3] and Gate-All-Around (GAA) architectures [4] on 300mm Si wafers. While acceptable interface quality is typically achieved by exploiting the so-called surface self-cleaning effect of TriMethylAluminum-based ALD deposition of Al_2O_3 on InGaAs [5], charge trapping in oxide border traps remains excessive, causing unacceptable time-dependent variability [6] and jeopardizing the device reliability due to large Positive Bias Temperature Instability (PBTI) [7,8].

In this paper, we compare charge trapping data of InGaAs channel devices across several device architectures, spanning from planar MOS capacitors and FETs [9], to finFETs [6], GAA's [4], Vertical NanoWires (VNW) [10], and even to a more advanced device concept as Tunnel-FETs [11]. We show that the same amount of charge trapping is observed irrespective of the device type and channel material. We argue that *the reliability performance is not intrinsically limited by the use of a IIIV channel per se, but by the high-k layer quality which is limited by the process thermal budget imposed by the channel material properties.*

We discuss IIIV-compatible low temperature anneals, and high-k nitridation for improved device reliability. Furthermore, we characterize oxide defect distributions in Al_2O_3 and HfO_2 deposited on InGaAs. We find that in HfO_2 a minimum defect density exists ~0.2eV below InGaAs conduction band (E_C). *We demonstrate a sufficiently reliable InGaAs/ HfO_2 gate stack by introducing an interface dipole* induced by a novel ALD interfacial layer (IL) developed on a ASM Pulsar® 3000 reactor [4], which aligns the HfO_2 minimum defect density level to the channel E_C.

Experimental: Charge trapping is benchmarked by flatband voltage (V_{fb}) hysteresis measurements. For a fair comparison of gate stacks with different Capacitance Equivalent Thickness (CET), the measured V_{fb} shifts are converted into equivalent charge sheet density ($\Delta N_{eff}=\Delta V_{fb}*C_{ox}/q$) and plotted against the applied equivalent oxide field [$E_{ox}\approx(V_G-V_{fb0})$/CET]. To estimate the maximum operating overdrive voltage (V_{ov}), CV-BTI measurements [12] have been performed at room temperature, with a sense delay of 20ms. PBTI reliability target is set at 10 year operation, with a $\Delta V_{fb}\leq$30mV. Assuming a typical time power-law exponent of ~0.13, and a target EOT of 1nm (CET≈1.6nm including quantum correction for InGaAs), this failure criterion projects to $\Delta N_{eff}\leq 3\times10^{10}cm^{-2}$ (hysteresis<2mV, at 1s charging time) at operating condition (V_{DD}=0.5-0.8V, $E_{ox}=(2/3)*V_{DD}$/CET≈2-3.5 MV/cm, cf. Fig. 1).

Results and Discussion: IIIV/high-k gate stacks show excessive charge trapping already at low operating voltages, with a very weak dependence on the applied charging E_{ox}, as compared to Si gate stacks (Fig. 1). We have ascribed this poor behavior to a wide distribution of oxide defect levels around the channel Fermi level [8]. A typical $\Delta N_{eff}\sim3\times10^{11}cm^{-2}$ is measured at operating conditions, i.e., 10× larger than an acceptable reliability target. *Similar* ΔN_{eff} *is observed consistently in various device architectures, from planar MOS capacitors and FETs, to advanced FinFETs, GAAs, VNWs and TFETs* (Fig. 2). In the case of simple test vehicles with no backend-of-line processing (MOS capacitors, lab MOSFETs), an even larger ΔN_{eff} is observed if a Post-Metal Anneal (PMA, 450C, 5', FGA) is not performed.

In order to understand whether this excessive charge trapping is intrinsically related to the use of an InGaAs channel (e.g., due to surface native oxides, or to diffusion of IIIV elements into the high-k layer), we compare with n-type Si capacitors, with same high-k ALD and metal gate (TiN). No high temperature anneal was performed after high-k deposition, to mimic IIIV process thermal budget limitations. Large ΔN_{eff} values and weak E_{ox}-dependence are observed in such gate stacks (Fig. 3a). In particular, *Al_2O_3 and Al_2O_3/HfO_2 stacks show comparable ΔN_{eff} on Si and on InGaAs*. HfO_2 shows a reasonably low ΔN_{eff} on Si when a ~1nm SiO_2 IL is used [13,14], while large ΔN_{eff} values, comparable to IIIV gate stacks, are observed when the SiO_2 IL is thinned down (Fig. 3b), unveiling insufficient high-k quality. A significant reliability improvement is, however, observed when a 1035C PMA is performed to mimic a Gate-First integration flow (Fig. 4), confirming the crucial role of thermal budget. IIIV-compatible, low temperature (T<600C) anneals can also yield some reliability improvement (Fig. 5): at low temperatures, no clear difference is observed when performing the anneal right after high-k deposition (PDA) or after metal gate deposition (PMA). In contrast, high temperature PDA (850C) leads to degraded reliability. By repeating twice the annealing step (PDA+PMA), a larger improvement is observed, suggesting longer anneal durations are favorable. Ultra-fast laser anneals show higher optimal temperature, suggesting high-k quality depends on the total anneal thermal budget (i.e., duration *and* temperature), and are therefore less suitable for the IIIV process window. *By optimizing the low-T anneal condition, an improvement of the max. operating V_{ov} is demonstrated* (Fig. 6). *Further improvement is obtained by incorporating Nitrogen in the high-k layer* [15] with a Decoupled Plasma Nitridation process (Fig. 7). High-k improvement is also crucial to fulfill device variability specs: the anneal optimization also reduces the V_{fb} and its spread across wafer, possibly thanks to a reduced density of fixed-charge in deep oxide traps (Fig. 8).

We have recently proposed a technique based on measuring hysteresis for increasing/decreasing charging/discharging voltages allowing to characterize shallow and deep oxide defect level distributions [16]. Here we apply this technique to characterize Al_2O_3 and HfO_2 deposited on InGaAs (Fig. 9). For both dielectrics, two defect bands are observed at shallow and deep energy levels. While Al_2O_3 shows a lower peak defect density, its shallow and deep defect bands partially overlap, yielding a wide distribution of defect levels around the InGaAs E_C. Apparent non-Arrhenius temperature dependence of charge trapping is observed in Al_2O_3, irrespective of the channel material, due to the deep defect band contribution at low temperatures (Fig. 10). In contrast, a minimum defect density is observed in HfO_2, ~0.2eV below InGaAs E_C (Fig. 9e), while the deep defect band does not contribute to hysteresis at operating voltages (Fig. 10c). By inducing an interface dipole (Fig. 9f-g), the minimum defect density level can be shifted towards the device operating energy range. This is achieved by depositing a recently developed IL (ASM® [4]) on InGaAs prior to HfO_2 deposition. A significant ΔN_{eff} reduction is achieved ($\Delta N_{eff}\sim3\times10^{10}cm^{-2}$ at E_{ox}~3MV/cm, Fig. 11a), with a stronger E_{ox} dependence beneficial for low V_{DD} operation. A max. operating V_{ov} of ~0.45V is demonstrated for this InGaAs MOS stack (Fig. 11b), *in line with IIIV operating specifications*.

Conclusions: We have shown that the poor PBTI reliability of IIIV/high-k gate stacks is universally related to process thermal budget limitations. Low temperature anneal optimization and high-k nitridation reduce oxide defect density. In contrast to a wide distribution of defect levels in Al_2O_3, HfO_2 on InGaAs shows a minimum defect density ~0.2eV below the channel E_C. By introducing an interface dipole, a significant reliability boost was demonstrated. While low thermal budget high-k quality and IIIV interface thermal stability constitute challenges, our results show that a reliable IIIV/high-k gate stack can be fabricated.

References: [1] S. Lee, VLSI14; [2] A. Vais, submitted VLSI16; [3] N. Waldron, VLSI14; [4] N. Waldron, IEDM15; [5] M.L. Huang, APL(87); [6] J. Franco, IEDM14; [7] S. Deora, TDMR13(4); [8] J. Franco, IRPS14; [9] A. Alian, IEDM13; [10] T. Ivanov, submitted VLSI16; [11] A. Alian, IEDM15; [12] E. Bury, IRPS13; [13] E. Cartier, IEDM11; [14] H. Arimura, IRPS14; [15] S.A. Krishnan, IRPS06; [16] A. Vais, APL(107).

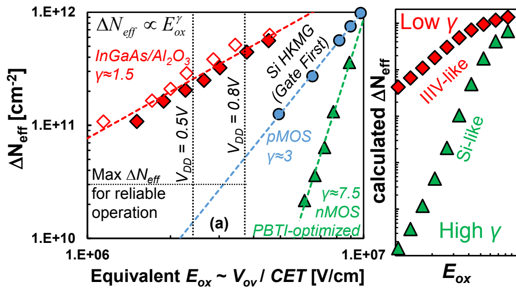


Fig. 1. **(a)** BTI shifts (t_{st}=1s) converted into effective trapped charge sheet density (ΔN_{eff}=$\Delta V_{fb}C_{ox}/q$) and plotted vs. the equivalent oxide field (E_{ox}≈V_{ov}/CET). IIIV/high-k gate stacks show larger ΔN_{eff} and weaker E_{ox}-dependence (exponent γ~1.5) as compared to Si devices (Si pMOS and nMOS Gate-First data shown for comparison) ascribed to **(b)** a wide energy distribution of oxide defects close to InGaAs E_C.

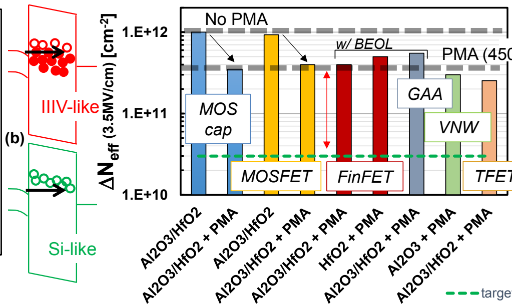


Fig. 2. IIIV/high-k gate stack hysteresis benchmark across a broad range of test vehicles (planar MOS capacitors and FETs, FinFETs, Gate-All-Around devices, Vertical NanoWires, and Tunnel-FETs). Irrespective of the device architecture, the same ΔN_{eff}~3x10^{11}cm^{-2} is observed. Simple vehicles without backend processing show larger ΔN_{eff} if a thermal step is not performed after metal gate deposition (No PMA vs. PMA 450C).

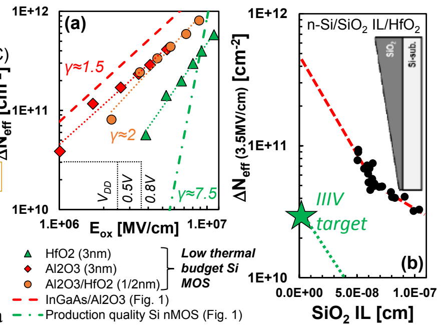
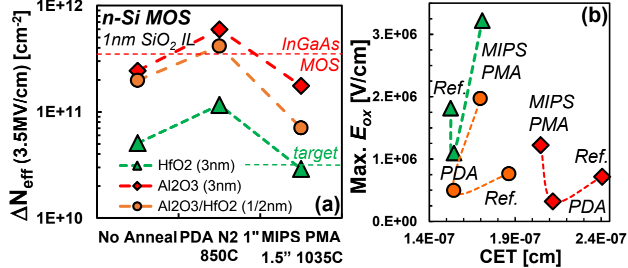
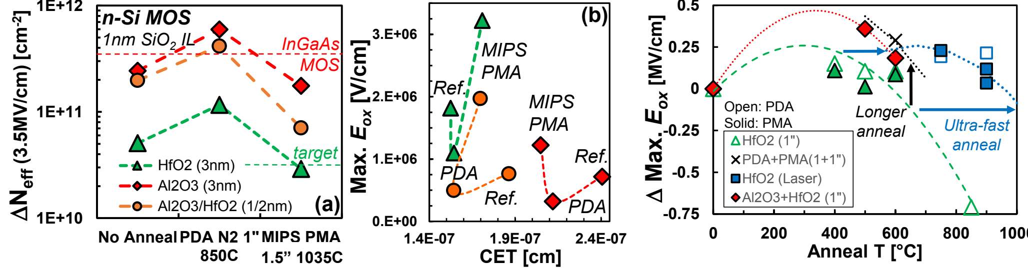


Fig. 3. **(a)** ΔN_{eff} vs. applied E_{ox} measured on low thermal budget (i.e., no anneal after high-k deposition) n-type Si capacitors. **(b)** A slant etch was used to modulate the IL thickness across wafer (inset). Note the weak E_{ox}-dependence and large ΔN_{eff}, similar to IIIV gate stacks.





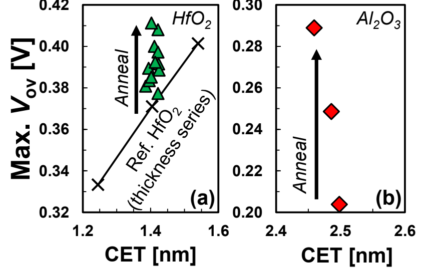


Fig. 4. **(a)** ΔN_{eff} measured on th different thermal budgets: without anneal, with 850C PDA, or with Gate-First-like 1035C PMA (Metal-Inserted-PolySi scheme). Different high-k layers are compared: 3nm HfO$_2$, 3nm Al$_2$O$_3$, or 1/2nm Al$_2$O$_3$/HfO$_2$. The use of Al$_2$O$_3$ induces a reliability penalty even on Si channels. **(b)** Maximum operating E_{ox}: only the Gate-First-like thermal budget yields sufficient reliability.

Fig. 5. Improvement of the max. operating E_{ox} vs. anneal temperature (open: PDA, solid: PMA), on low thermal budget Si high-k gate stacks. IIIV-compatible anneals (<600C) can improve gate stack reliability. Ultra-fast Laser anneals shift optimum T to higher values.

Fig. 6. Low temperature anneals improve maximum operating overdrive voltage, with no CET penalty on **(a)** HfO$_2$, and **(b)** Al$_2$O$_3$ gate stacks. Note the lower max. V_{ov} (i.e., poorer reliability) with Al$_2$O$_3$, despite the larger CET.

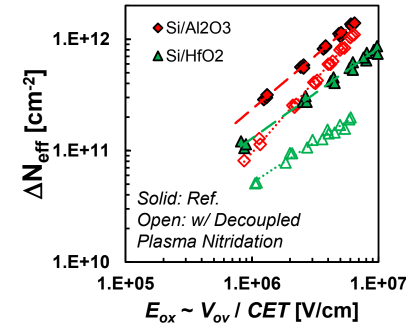
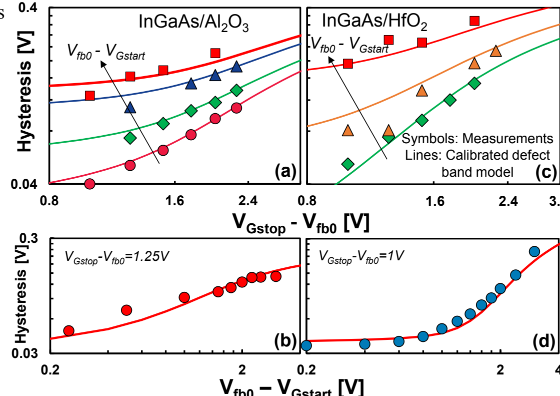


Fig. 7. ΔN_{eff} measured on low thermal budget Si MOS with Al$_2$O$_3$ or HfO$_2$, without and with Decoupled Plasma Nitridation. Incorporation of Nitrogen in the high-k yields some oxide trap passivation in both high-k layers.



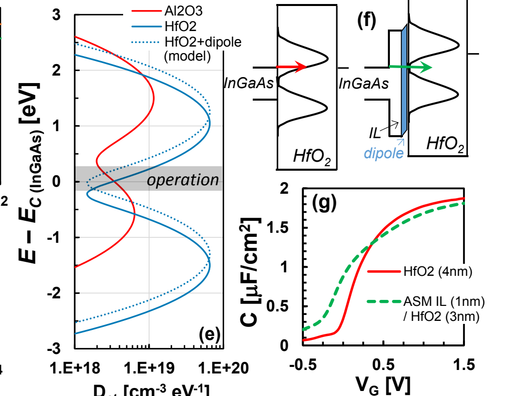


Fig. 9. Hysteresis measured for varying maximum V_G (i.e., charging voltage) and minimum V_G (i.e., discharging voltage), in **(a-b)** InGaAs/Al$_2$O$_3$ and **(c-d)** InGaAs/HfO$_2$ gate stacks. **(e)** A defect band model, including two normal distributions of defect levels above and below InGaAs E_C, calibrated to match the experimental data [cf. lines in (a-d)]. While the peak defect densities are lower in Al$_2$O$_3$, a partial overlap of the defect bands induces an almost uniform distribution of defect levels around InGaAs E_C. On the contrary, HfO$_2$ on InGaAs shows a minimum defect density ~0.2eV below the channel E_C, which **(f)** can be exploited for improving gate stack reliability by introducing an interface dipole. **(g)** 100kHz C-V curves of InGaAs/HfO$_2$ and InGaAs/ASM-IL/HfO$_2$ MOS; note the ~0.2V V_{fb} tuning.

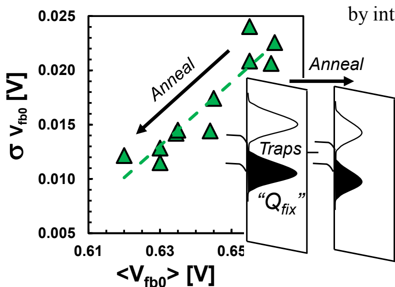


Fig. 8. Standard deviation of V_{fb} across Si MOS capacitors vs. the mean V_{fb} of each wafer. Anneals reduce both V_{fb} and its variability, suggesting (inset) a reduction of both shallow electron traps (i.e., improved reliability) and of deep traps which are normally filled at operating condition (i.e., behave as fixed oxide charge).

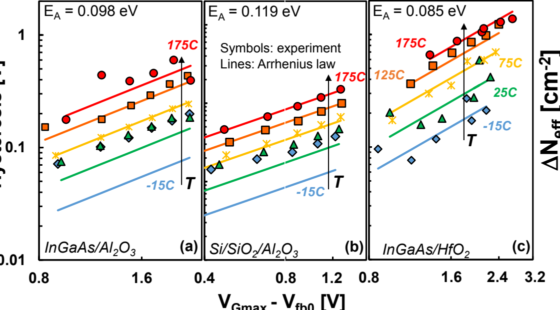


Fig. 10. Temperature dependence of hysteresis in **(a)** InGaAs/Al$_2$O$_3$, **(b)** Si/SiO$_2$/Al$_2$O$_3$, **(c)** InGaAs//HfO$_2$ gate stacks. A single Arrhenius activation does not describe correctly the temperature acceleration of charge trapping in Al$_2$O$_3$ gate stacks, irrespective of the channel material, due to the contribution of both shallow and deep defect bands. In contrast, HfO$_2$-based IIIV gate stack shows a typical Arrhenius dependence of hysteresis, confirming only the shallow defect band interacts with channel carriers.

Fig. 11. **(a)** ΔN_{eff} measured in InGaAs/HfO$_2$ gate stacks, without and with ASM-IL (cf. record GAA drive demonstrated in [4] with the same layer). Note the stronger E_{ox}-dependence of hysteresis and the reduced ΔN_{eff} at operating condition for the latter. **(b)** Extrapolated PBTI time-to-failure of the optimized gate stack: a max. operating overdrive of ~0.45V for 10 year reliability is demonstrated, sufficient for targeted low V_{DD} operation.