Impact of high-k dielectric defects on the mobility and reliability of InGaAs channel MOSFETS

E. Cartier, A. Majumdar, K. A. Jenkins, J.-B. Yau, M. M. Frank, T. Ando, J. Rozen, J. Bruley, P. Kerber, M. Hopstaken, E. Levrau, W-E Wang, K.-T. Lee, C. Liang, C.-W. Cheng, J. Bruley, X. Sun, R. T. Mo, C.-C. Yeh, E. Leobandung and V. Narayanan

IBM, T. J. Watson Research Center
Yorktown Heights, NY 10598, USA
**Corresponding author: ecartier@us.ibm.com*

Abstract

Significant progress has been made in the fabrication of short channel InGaAs-based nMOSFETs. While scaling and performance are key elements in establishing technology feasibility, the reliability of such new technology elements needs to be established in a timely manner. In this presentation, we summarize our experimental efforts to quantify and understand the role of the high-k gate stack and its defects for the performance *and* the stability of InGaAs nMOSFETs and we discuss methods to improve performance and reliability of the transistors.

1. Introduction

InGaAs-based nFETs have been under intense study for future technology nodes due to their higher electron mobility and higher electron injection velocity than Si nFETs [1]. Recently, it was shown that short-channel FinFETs on InGaAs on-silicon wafers using the aspect ratio trapping (ART) technique exhibit excellent short channel control down to 20 nm gate [2].

In state-of-the-art HfO_2-based MGHK nFETs, the presence of a high quality interfacial SiO_2-layer remains essential to minimize charge trapping and to provide the necessary threshold-voltage stability. The oxides of InGaAs are of low electrically quality and need to be replaced by deposited layer. We will discuss the impact of this approach on the stability of InGaAs nFETs with a Al_2O_3/HfO_2 dual-layer. The charge trapping can be substantially reduced by high pressure deuterium annealing leading to large performance improvements. But the stability of the gate stack remains significantly inferior to that of MGHK nFETs with comparable thickness scaling.

2. Experimental

We will discuss device fabrication [3] with a focus on gate stack formation and annealing [4] (Fig. 1-3). Various experimental techniques with fast sensing of charge trapping are than applied to investigate the stability of transistors (Fig. 4-6). [5, 6] Fast sensing is a key element in the qualification of these transistors.

3. Results and Discussion

The impact of high-pressure deuterium annealing is summarized in Fig. 2. These anneals are found to lead to substantial mobility enhancement in InGaAs nFETs with an Al_2O_3/HfO_2 dual-layer gate stacks, yielding motilities comparable to those measured at 35 K where charge trapping is strongly suppressed (Fig. 3). The instability is largely due to fast electron trapping (Fig. 4) and is significantly improved with annealing (Fig. 6). In addition to a reduction of electron trapping and enhanced V_t-stability, a true increase in mobility is observed consistent with a significant reduction of the interface state density.

The reduction of the interface-state density and of the electron trapping in the gate stack allows detailed studies of the interface stability. Using conductance measurements on MOS capacitors, we demonstrate, that significant amounts of interface-states are generated at intended use conditions of a low power technology for example in a 5 nm node. This leads to new reliability concerns with InGaAs nFETs. In a 'hybrid' technology with a Ge/SiGe pFET and an InGaAs nFET, it should be anticipated that a possible reliability impasse will come from the nFET, as the pFet has been demonstrated to be very robust. [7]

4. Conclusion

The performance and reliability of InGaAs-based nFETs with Al_2O_3/HfO_2TiN dual-dielectrics gate stacks can be greatly improved, yielding scaled nFETs with good performance. Further improvement of the gate stack stability will however be needed to meet reliability criteria used in state-of-the-art CMOS technologies with MGHK gate stacks on Si.

References

[1] J. A. del Alamo, Nature 479, pp. 317–323, (2011).
[2] X. Sun, *et al.,* VLSI Tech. Dig. T40 (2017).
[3] A. Majumdar, *et al.,* IEEE Trans. Electron Dev. 61, pp. 3399-3304 (2014).
[4] M. M. Frank, presented at SISC (2017), to be published
[5] E. Cartier, *et al*., Solid-State Dev. Res. Conf. (ESSDERC), pp. 292-295, 47th European (2017).
[6] E. Cartier, *et al*., Int. Rel. Symp. (IRPS) (2018).
[7] [6] P. Hashemi et al., VLSI Tech. Digest., p 120 (2017)
[8] K. A. Jenkins, to be published

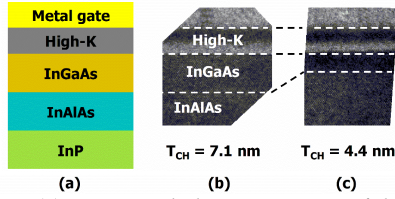

Fig. 1. (a) Gate stack layer structure of long-channel $In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}As$ FETs on InP substrate. (b), (c) High-resolution TEM images of fabricated devices with channel thickness TCH = 7.1 nm and 4.4 nm, respectively.

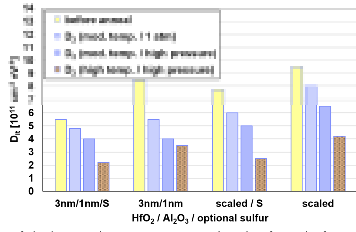

Fig. 2. D_{it} of bilayer/InGaAs stacks before/after anneal in D_2 at various pressures. Substantial reduction in D_{it} (from 100 kHz conductance) can be achieved. [4]

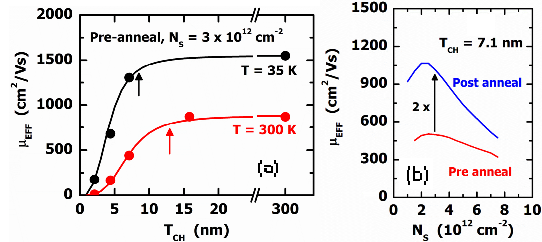

Fig. 3. (a) $In_{0.53}Ga_{0.47}As$ channels at carrier density $N_S = 3 \times 10^{12}$ cm^{-2} and temperature T = 35 and 300 K. The arrows indicate where μ_{EFF} drops by 10% of its thick-channel value. (b) Pre- and post-anneal room-temperature mobility μ_{EFF} (obtained from DC ID-VG) vs. carrier density N_S of $In_{0.53}Ga_{0.47}As$ channels with thickness T_{CH} = 7.1 nm. At $N_S = 3 \times 10^{12}$ cm^{-2}, optimized D_2-annealing lead to 2 x higher mobility. [5]

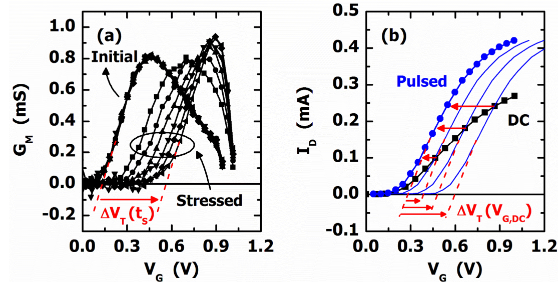

Fig. 4. Pulsed stress data for $In_{0.53}Ga_{0.47}As$ channel with thickness T_{CH} = 7.1 nm. (a) Pulsed linear trans-conductance G_M vs. V_G for stress times t_S = 1 µs to 100 ms (10 × steps) showing only threshold voltage shift ΔV_T with stress at V_G = 1 V. (b) Schematic illustrating the origin of the large stretch out in the measured I_D-V_G curve during a DC sweep under steady-state charge trapping. The V_T of the pulsed I_D-V_G curves (1 µs rise time) is shifted to match the DC sweep. [5]

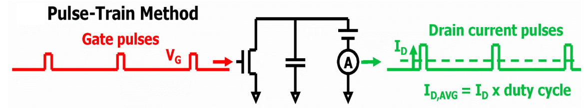

Fig. 5. Illustration of a pulse-train method to measure drain currents in FETs at high speed in large area FETs. The duty cycle is kept constant at 1 %, allowing for long charge relaxation times after each on-pulse. A low-pass filter is used to integrate the drain current over many pulses to boost the measurement sensitivity. Pulse times as short as 1 ns are obtained by optimizing the experimental setup. [8]

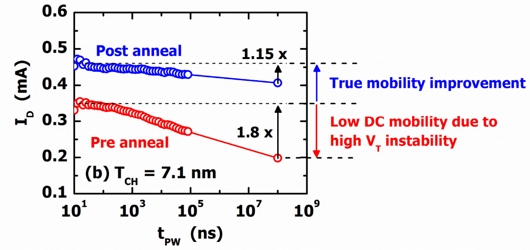

Fig. 6. Averaged drain current, $I_{D,AVG}$ vs. pulse width, t_{PW}, for InGaAs channels with thickness T_{CH} = 7.1 nm before and after advanced anneal at room temperature. Both, electron trapping (slope) and N_{it} are reduced (vertical shift at short times). [5]

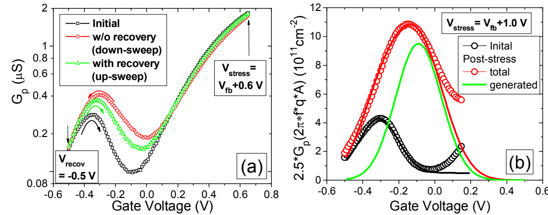

Fig. 7. (a) Evolution of the parallel conductance, G_P@100kHz as measured during a stress sequence.
(b) Decomposition of post-stress G_P-peak into an 'initial' and post-stress interface-state component. New states appear to be shifted in gate voltage with respect to the pre-existing interface-states. [6]

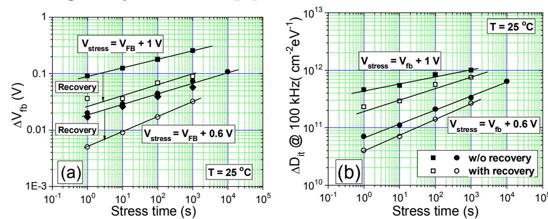

Fig. 8: (a) Evolution of the post-stress and post-recovery flatband voltage during PBTI stress as a function of stress time at V_g = 0.6 V and 1.0 V, respectively, for high pressure deuterium passivated $Al_2O_3/HfO_2/TiN$ gate stacks on n-type InGaAs at 25 °C. (b) Evolution of the post-stress and post-recovery interface state density during PBTI stress as a function of stress time at V_g = 0.6 V and 1.0 V, respectively, for high pressure deuterium passivated $Al_2O_3/HfO_2/TiN$ gate stacks on n-type InGaAs at 25 °C. [6]