Investigation of Bias-Temperature Instability in work-function-tuned high-*k*/metal-gate stacks

B. Kaczer,[a] A. Veloso, and Ph. J. Roussel
IMEC, B-3001 Leuven, Belgium

T. Grasser
Christian Doppler Laboratory for TCAD at the Institute for Microelectronics, TU Wien, A-1040 Vienna, Austria

G. Groeseneken
IMEC, B-3001 Leuven, Belgium and ESAT Department, KU Leuven, B-3001 Heverlee, Belgium

(Received 8 June 2008; accepted 17 November 2008; published 9 February 2009)

The impact of V_{th}-adjusting layers on high-*k*/metal-gate *n*- and *p*FinFET V_{th} stability is investigated. Additional insight is gained by monitoring ΔV_{th} recovery transients over several decades in time. V_{th}-adjusting capping layers deposited directly on top of the gate dielectric degrade the V_{th} stability. The V_{th} stability improves as the capping layers are buried in the gate metal and are thus isolated from the gate dielectric. A combination of a capping layer thickness and depth in the metal gate is found that practically eliminates *n*FET V_{th} instability. © *2009 American Vacuum Society.*
[DOI: 10.1116/1.3054352]

I. INTRODUCTION

The introduction of high-*k*/metal-gate stacks into complementary metal-oxide semiconductor (CMOS) technology requires separately "tuning" the gate work function for *n*- and *p*FET devices. Adjustment of the field-effect transistor (FET) threshold voltage V_{th} can be achieved, e.g., by insertion of "capping" layers into the gate stack, as discussed elsewhere.[1] In this work we investigate the impact of these layers on the bias-temperature instability (BTI) of V_{th} in *n*- and *p*FinFETs with HfSiO dielectrics and TiN gates. An extension of a standard BTI evaluation technique[2] allows us to monitor V_{th} *transients* and thus to better understand and compare the influence of the layers.

We find that capping layers deposited directly on top of the gate dielectric degrade the V_{th} stability. Conversely, *p*FET capping layers separated sufficiently from the gate dielectric, i.e., "buried" in the gate metal, have V_{th} stability identical to reference stacks. Similarly, the *n*FET V_{th} stability improves as the capping layers are isolated from the gate dielectric. Finally, we report that a capping layer suitably spaced from the gate dielectric *practically eliminates* *n*FET V_{th} instability.

II. SAMPLES

Silicon-on-insulator (SOI) *n*- and *p*FinFETs with $L_{FIN}=1\ \mu$m, $W_{FIN}=10\times 1\ \mu$m, and $H_{FIN}=65$ nm were used. AlO and DyO capping layers were, respectively, introduced into the *p*FET and *n*FET stacks, resulting in V_{th} and capacitance equivalent thickness (CET) as given in Fig. 1.[1] Conservative dielectrics thicknesses were chosen to avoid reliability issues not related to negative and positive BTI (NBTI and PBTI). The completed stacks were exposed to 1050 °C during FET junction formation.

[a]Electronic mail: kaczer@imec.be

III. MEASUREMENT TECHNIQUE

The BTI is typically assessed using the measure-stress-measure (MSM) technique,[2] during which the threshold voltage *shift* ΔV_{th} with respect to the initial V_{th} is measured after each stress phase. Such evaluation tacitly assumes a *time-independent* ΔV_{th}, while in reality strong transients are observed.[2] The V_{th} relaxation *trends* thus carry additional information about the BTI processes in the studied device, which is otherwise lost in the single-V_{th} MSM measurements.

Here we therefore use the recently introduced *extended* MSM (eMSM) technique (Ref. 2) to monitor the relaxation transients after each stress phase [Fig. 2(a)]. Since the BTI relaxation typically occurs over many time scales, the transients are recorded and plotted on the logarithmic time scale from ~1 ms to ~10 s, i.e., over four decades in time. An example of such plot is shown in Fig. 2(b), with the time t_{relax} always measured from the end of the preceding stress phase. The last relaxation transient is typically monitored up to 3–10 ks, i.e., over $6\frac{1}{2}$–7 decades in time.

All measurements are done at temperature $T=125$ °C. The measurement voltage V_{meas} [Fig. 2(a)] is chosen close to the initial threshold voltage $V_{th,initial}$ (Fig. 1).[3] The stress voltages V_{stress} are chosen with respect to $V_{th,initial}$, and the observations are therefore reported at the "overdrive" voltage $V_{s\text{-}th}=V_{stress}-V_{th,initial}$. The integrity of the gate dielectric after the entire MSM sequence is checked by assessing the gate leakage at V_{meas}.[2]

IV. PFET RESULTS (NBTI)

In all the *p*FET high-*k* stacks studied, ΔV_{th} transients show a behavior qualitatively identical to what we previously observed on conventional SiO_2 and SiON stacks, with the $\log(t_{relax})$-like relaxation [Fig. 2(b)], indicating an underlying dispersive process.[4,5] Following the same analysis as in

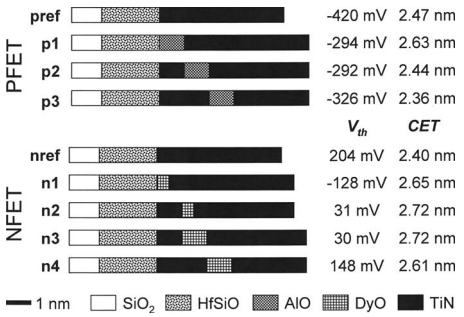

FIG. 1. Investigated *p*- and *n*FinFET gate stacks. All film thicknesses are to the scale indicated in the figure. Capping layers at different positions in gate stacks adjust V_{th} and slightly influence CET.

Ref. 5, assuming *universality of relaxation*,[6] we quantify the ΔV_{th} transients [Fig. 2(b)] by separating them into the relaxing ("recoverable" or "*R*") and the permanent ("unrecoverable" or "*P*") degradation components, i.e., $\Delta V_{th}=R+P$. With the above assumption, we describe the recoverable component *R* by the universal relaxation function in the following form:

$$R(t_{\text{stress}},t_{\text{relax}})=\frac{R(t_{\text{stress}},t_{\text{relax}}=0)}{1+B\xi^{\beta}}, \quad (1)$$

where $\xi=t_{\text{relax}}/t_{\text{stress}}$ is the *universal* (dimensionless) relaxation time, t_{stress} is the *total* (*cumulative*) stress time, and *B* is a scaling parameter. The functional form of Eq. (1) has properties similar to a stretched exponential, which is often invoked to describe the relaxation of dispersive systems, and β has the attributes of a dispersion parameter.[4]

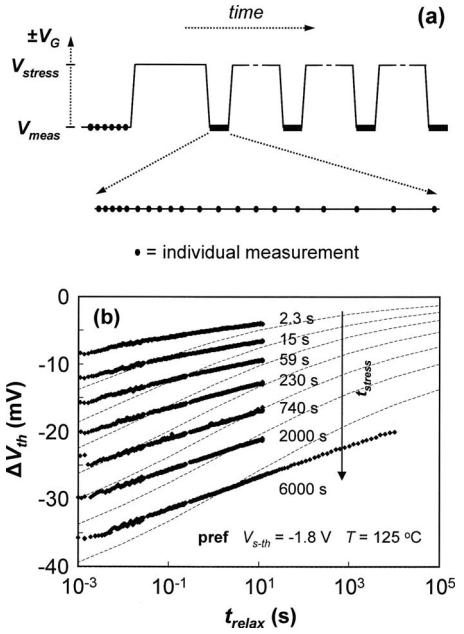

FIG. 2. (a) Extended MSM stress/relaxation voltage sequence applied at FET gate. The FET source current is monitored at V_{meas} after each stress phase and converted to ΔV_{th} *relaxation transient* spanning several decades. (b) $\Delta V_{th}<0$ V transients at increasing cumulative stress times measured on stack "pref" (symbols). ΔV_{th} transients calculated by fitting relaxation data obtained on 1.4 nm SiON *p*FET stressed at similar oxide electric field (Ref. 5) are shown for comparison (dashed lines).

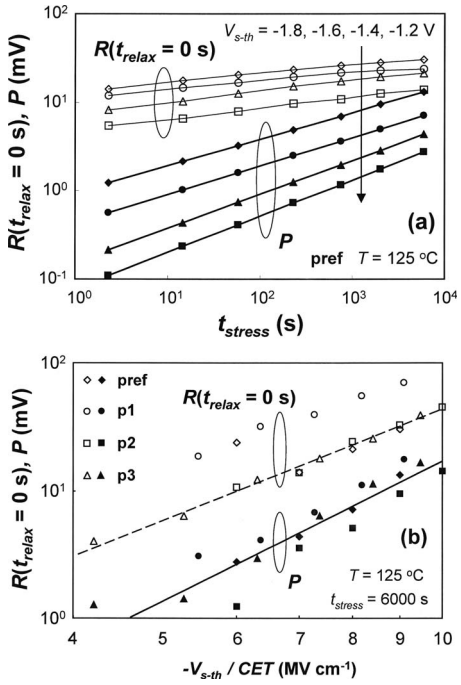

FIG. 3. (a) Permanent *P* and recoverable *R* components extracted from Fig. 2(b) (and other V_{stress} values on stack "pref") vs t_{stress}. Unlike in Ref. 5, a simple power-law dependence on t_{stress} is assumed for the permanent component during extraction. (b) Extracted *P* and *R* components at $t_{\text{stress}}=6000$ s for all studied *p*FET stacks. Stacks with buried capping layers ("p2" and "p3") behave identically to the reference stack "pref" (spread in the *P* component data is within the precision of the extraction technique).

The recoverable component *R* is reported at $t_{\text{relax}}=0$ s, i.e., the moment *immediately* after the removal of stress gate voltage. The permanent component $P(t_{\text{stress}})$ reported here then represents the remaining ΔV_{th} after the full relaxation of the recoverable component, corresponding to $t_{\text{relax}}=\infty$ s in Eq. (1) (the possibility of long-term recovery of the *P* component is discussed in Ref. 7). The result of this extrapolation, done after each stress phase (i.e., as a function of the total t_{stress}) and for each stress voltage condition, is shown in Fig. 3(a).

As NBTI is typically assumed to be accelerated by the electric field at the substrate/dielectric interface,[8] both components are compared at $V_{\text{s-th}}$/CET. In order to reduce measurement and extraction inaccuracies, the comparison is done at the maximum degradation at $t_{\text{stress}}=6000$ s. The comparison of all *p*FET stacks in Fig. 3(b) shows that the permanent degradation is identical for all *p*FET stacks, irrespective of the capping layer presence or its position. In stacks "*p*2" and "*p*3," where the capping layer is buried in the gate electrode, the recoverable component behaves identically to the reference.

Interestingly, the recoverable component increases considerably in stack "*p*1," which coincides with significant mobility degradation observed in the *p*FETs with this stack.[1] Both effects are most probably due to new states created by outdiffusion of Al, likely as far as the SiO_2 interfacial layer.

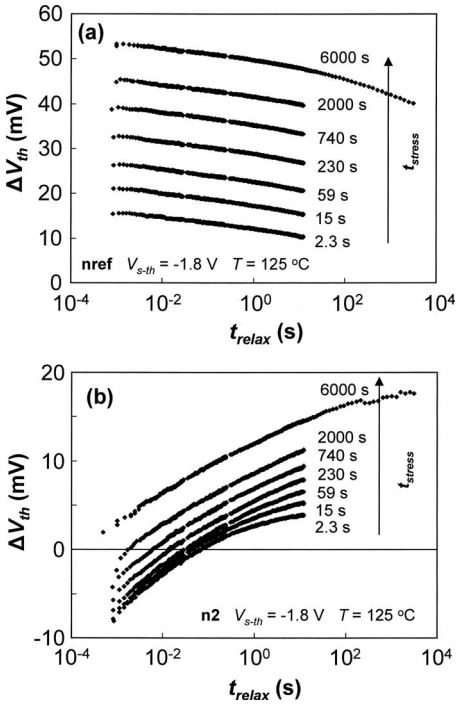

FIG. 4. (a) Relaxation of $\Delta V_{th} > 0$ V due to negative charge in reference *n*FET stack and (b) opposite trend due to insertion of DyO is superimposed on the overall positive shift in ΔV_{th} with increasing stress, same as in (a).

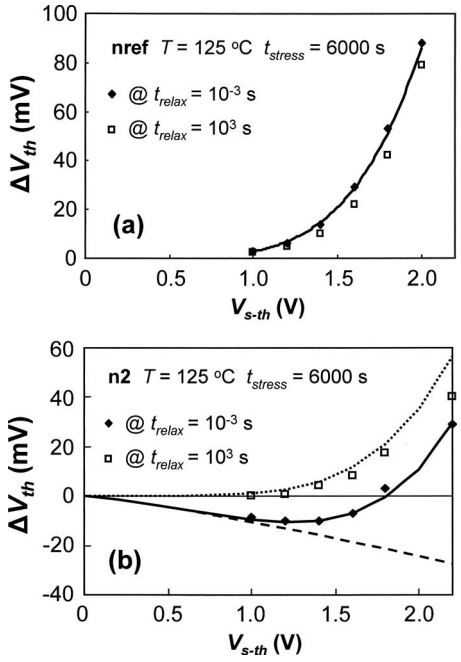

FIG. 5. (a) In the *n*FET reference stack, ΔV_{th} at $t_{relax} = 10^{-3}$ s vs $V_{s\text{-}th}$ can be fitted with $\sim V_{s\text{-}th}^{5}$. (b) ΔV_{th} at 10^{-3} s in stack "*n*2" can be fitted with two competing trends (respectively, proportional to $V_{s\text{-}th}^{5}$ and $-V_{s\text{-}th}^{1.5}$). Stacks *n*2 and "*n*3" behave similarly.

In stacks *p*2 and *p*3 the TiN barrier prevents intermixing of Al during the high-temperature junction formation, which preserves the NBTI behavior seen in the reference stack.

We have already noted that although the studied *p*FET devices contain a high-permittivity gate dielectric layer, their NBTI relaxation behavior is comparable to "conventional" SiO_2 and SiON *p*FETs. This similarity is not only qualitative but also quantitative. The relaxation *dispersion* is very similar in both types of stacks—the dispersion parameter β is typically between 0.12 and 0.16 in the high-*k* stacks discussed here, while ~0.16 was found for SiON *p*FETs.[5] Also the *magnitude* of both components is comparable to the conventional *p*FET, as is documented in Fig. 2(b). This is an indication of a similar NBTI mechanism in both the conventional and the high-*k* *p*FETs, controlled by the properties of the oxide close to the substrate/dielectrics interface.[9]

V. NFET RESULTS (PBTI)

In *n*FETs with the reference gate stack "nref" following positive BTI stress, ΔV_{th} transients of the opposite polarity (i.e., corresponding to negative charge) are observed in Fig. 4(a).[10] Note that PBTI stress again results in very long, $\log(t_{relax})$-like ΔV_{th} relaxation transients, just like those typically observed in the NBTI/*p*FET case [Fig. 2(b)].

In the stacks with the DyO capping layer, the ΔV_{th} transients reveal an *additional* mechanism corresponding to positive charging [Fig. 4(b)].[11] Note that the existence of the two "competing" components would be completely overlooked if only a traditional single-point MSM technique were used.[4] The presence of the two transient components in stacks "*n*1"-"*n*4," however, precludes the use of the analysis based on relaxation universality (cf. Sec. IV). We therefore evaluate ΔV_{th} at the shortest and longest relaxation times as a function of $V_{s\text{-}th}$ (Fig. 5). The reference stack positive ΔV_{th} appears to follow the power-law dependence on $V_{s\text{-}th}$ [Fig. 5(a)]. In contrast to that, ΔV_{th} at 10^{-3} s in stacks "*n*2" and "*n*3" changes from negative to positive as $V_{s\text{-}th}$ is increased. This trend can be reproduced with two opposite components, both depending on $V_{s\text{-}th}$ [Fig. 5(b)]. The fit representing the positive component (dotted line) tracks ΔV_{th} at 10^{3} s, suggesting the negative transient abating after 10^{3} s.

While depositing the DyO layer directly on top of the dielectric (stack *n*1) leads to large instabilities and *n*FET mobility degradation[1] likely due to intermixing of the layers, separating the DyO layer by the gate metal substantially reduces outdiffusion of Dy and, in fact, improves the V_{th} sta-

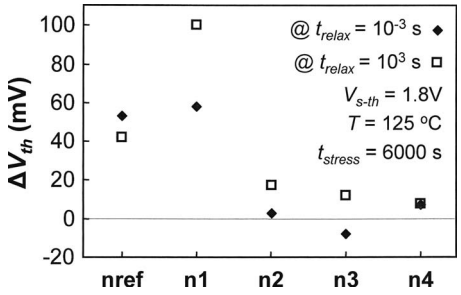

FIG. 6. Depositing the DyO layer directly on top of the dielectric (stack "*n*1") aggravates the V_{th} instability over the reference stack "nref," while burying the DyO layer in the gate metal (stacks "*n*2"—"*n*4") substantially improves it.

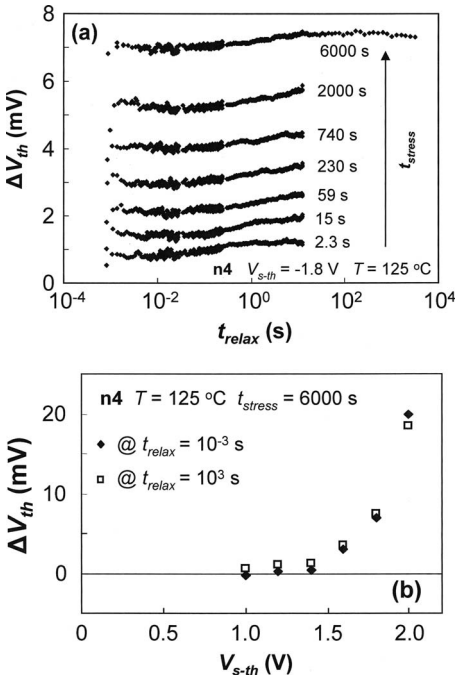

FIG. 7. (a) In stack "*n*4," the two components in ΔV_{th} transients appear to be compensating, with minute overall ΔV_{th} shift. (b) At $V_{s\text{-}th}$=1 V, the overall ΔV_{th} after 6000 s of stress is ~1 mV.

bility with respect to the reference stack, as shown in Fig. 6. Most notably, in stack *n*4 with the DyO layer separated by 2 nm of TiN, after 6000 s of T=125 °C and $V_{s\text{-}th}$=1.8 V stress, ΔV_{th} is mere 7 mV and stable, as is also documented by Fig. 7(a). This *n*FET high-*k*/metal gate stack shows extremely small instability, of the order of 1 mV or less at $V_{s\text{-}th}$=1 V [Fig. 7(b)]. A similarly negligible instability can be expected at operating conditions ($V_{s\text{-}th} \approx 0.6-0.8$ V and 3×10^8 s). We speculate that this extremely stable V_{th} could be due to compensation of defect states in the dielectric by Dy.

VI. CONCLUSIONS

The impact of V_{th}-adjusting layers on high-*k*/metal-gate *n*- and *p*FinFETs V_{th} stability is investigated. Additional insight is gained by application of the recently developed extended MSM technique to monitor ΔV_{th} transients. Capping layers deposited directly on top of the gate dielectric degrade the V_{th} stability, most probably due to new states created by intermixing of the capping layer with the dielectric. V_{th} stability improves as the capping layers are buried in the gate metal, thus reducing their outdiffusion into the gate dielectric. A combination of a capping layer thickness and depth in the metal gate is found that hugely improves *n*FET V_{th} stability.

[1]A. Veloso *et al.*, Proceedings of the 2008 International Symposium on VLSI Technology (IEEE, Piscataway, NJ, 2008), p. 14.
[2]B. Kaczer, T. Grasser, Ph. J. Roussel, J. Martin-Martinez, R. O'Connor, B. J. O'Sullivan, and G. Groeseneken, Proceedings of the 2008 IEEE International Reliability Physics Symposium, 2008 (IEEE, Piscataway, NJ, 2008), p. 20.
[3]T. Grasser, P.-J. Wagner, Ph. Hehenberger, W. Gös, and B. Kaczer, Proceedings of the 2007 International Integrated Reliability Workshop, 2007 (IEEE, Piscataway, NJ, 2007), p. 6.
[4]B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, Proceedings of the 2005 IEEE International Reliability Physics Symposium, 2005 (IEEE, Piscataway, NJ, 2005), p. 381.
[5]T. Grasser, B. Kaczer, P. Hehenberger, W. Gös, R. O'Connor, H. Reisinger, W. Gustin, and C. Schlünder, Proceedings of the 2007 International Electron Devices Meeting, 2007 (IEEE, Piscataway, NJ, 2007), p. 801.
[6]M. Denais *et al.*, Proceedings of the 2006 IEEE International Reliability Physics Symposium, 2006 (IEEE, Piscataway, NJ, 2006), p. 735.
[7]T. Grasser, B. Kaczer, and W. Gös, Proceedings of the 2008 IEEE International Reliability Symposium, 2008 (IEEE, Piscataway, NJ, 2008), p. 28.
[8]A. E. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, and M. A. Alam, IEEE Trans. Electron Devices **54**, 2143 (2007).
[9]S. Zafar (private communication).
[10]D. Heh, C. D. Young, and G. Bersuker, IEEE Electron Device Lett. **29**, 180 (2008).
[11]S. Z. Chang *et al.*, Proceedings of the 2008 International Symposium on VLSI Technology (IEEE, Piscataway, NJ, 2008), p. 62.