Received October 25, 2020; revised January 7, 2021; accepted January 31, 2021. Date of publication 8 February 2021;
date of current version 25 February 2021. The review of this article was arranged by Editor M. K. Radhakrishnan.

Digital Object Identifier 10.1109/JEDS.2021.3057662

Alleviation of Negative-Bias Temperature Instability in Si p-FinFETs With ALD W Gate-Filling Metal by Annealing Process Optimization

LONGDA ZHOU [1,2], QIANQIAN LIU [1], HONG YANG [1,2], ZHIGANG JI [3] (Member, IEEE), HAO XU [1], GUILEI WANG [1,2], EDDY SIMOEN [4] (Senior Member, IEEE), HAOJIE JIANG [1], YING LUO [1], ZHENZHEN KONG [1], GUOBIN BAI [1], JUN LUO [1,2], HUAXIANG YIN [1,2] (Senior Member, IEEE), CHAO ZHAO [1,2], AND WENWU WANG [1,2]

1 Key Laboratory of Microelectronics Devices & Integrated Technology, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China
2 The School of Electronic, Electrical, and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
3 National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Shanghai Jiaotong University, Shanghai 200240, China
4 Unit Process Modules Department, IMEC, 3001 Leuven, Belgium

CORRESPONDING AUTHOR: H. YANG and G. WANG (e-mail: yanghong@ime.ac.cn; wangguilei@ime.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 92064002, and in part by the National Key Research and Development Program of China under Grant 2016YFA0301701.

ABSTRACT In this article, we present an experimental study on the impact of post-metallization annealing conditions on the negative-bias temperature instability (NBTI) of Si p-channel fin field-effect transistors (p-FinFETs) with atomic layer deposition tungsten (ALD W) as the gate-filling metal. The effects of annealing conditions on the tensile stress of the W film, impurity element concentration in the gate stack, fresh interface quality, threshold voltage shift (ΔV_T), pre-existing traps (ΔN_{HT}), generated traps, and their relative contributions were studied. The time exponents of ΔV_T, the impacts of stress bias and temperature on NBTI degradation, and the recovery kinetics of the generated traps were analyzed. For devices with a B_2H_6-based W-filling metal, a 34% reduction in the fresh interface states, reduced ΔV_T, and a 29% improvement in the operation overdrive voltage could be achieved by optimizing the annealing conditions. The NBTI is alleviated mainly because of the reduction in the generated traps, while the energy distribution of ΔN_{HT} is insensitive to the annealing conditions. Furthermore, the relative contribution of the generated bulk insulator traps to the total number of generated traps could be reduced by optimizing the annealing conditions.

INDEX TERMS Reliability, negative-bias temperature instability (NBTI), Si p-FinFETs, post-metallization annealing, ALD W gate-filling metal, trap generation.

I. INTRODUCTION

In the replacement metal gate (RMG) fin field-effect transistor (FinFET) technology, atomic layer deposition (ALD) tungsten (W) has been widely implemented as a gate-filling metal owing to its excellent step coverage and conformity to the deposited thin film [1]–[3]. In particular, the WF_6-based ALD process with reducing agents B_2H_6 or SiH_4 [4]–[8] has been systematically investigated in terms of the gap filling capability, fluorine concentration in the W film, impact on the channel stress, and electrical characteristics of the capacitors and transistors [6], [8]. Despite the promising initial performance, more attention should be paid on reliability issues. The influence of ALD W gate-filling metal on the negative-bias temperature instability (NBTI) of RMG Si FinFETs originates from two aspects: tensile strain and residual fluoride in the W film [9]. The compressive strain along the gate length direction of the channel induced by the gate-filling metal with an inherent tension [2] causes a reduction

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

TABLE 1. Process parameters for gate stack.

Process	Precursors	Growth Temperature
ALD HfO_2	TEMAH+H_2O	300 °C
ALD TiN	$TiCl_4$+NH_3	400 °C
ALD TaN	$TaCl_5$+NH_3	400 °C
CVD TiN	TDMAT	400 °C
ALD W	WF_6+B_2H_6/SiH_4	300 °C

in the interface trap generation owing to the lower forward reaction rate constant [10], [11]. The diffusion of F from the W film into the gate stack passivates the Si dangling bonds and forms strong Si–F bonds at the interfaces, which in turn suppresses trap generation under NBTI stress [9]. Further improvement in device reliability can be expected by independently modulating the tensile strain in the W film and the F concentration in the gate stack by process optimization. Considering that both F diffusion and bonding are thermally activated processes [12], optimization of post-metallization annealing could serve as an efficient and convenient way to increase the F concentration at the interfaces. Therefore, it is necessary to investigate the impact of annealing conditions on the tensile strain in the W film and NBTI of Si FinFETs with ALD W gate-filling metal.

In this work, the tensile stress of the ALD W films and the F concentration at the interfaces under different annealing conditions were measured and compared. Annealing-optimized p-FinFETs were fabricated under the optimized annealing condition. The NBTI degradation, key parameters, and lifetimes of baseline and annealing-optimized devices are compared, and the underlying physical mechanism is discussed.

II. DEVICE AND EXPERIMENTAL

RMG Si p-FinFETs were fabricated using a full-gate-last process. Fig. 1a shows the major steps of the gate stack process: (1) Dummy poly-Si/SiO_2 gate stack removal; (2) Growth of the interfacial layer (IL) of SiO_2 via chemical oxidation of the ozone; (3) ALD deposition of HfO_2 as a high-k layer; (4) Deposition of a multilayer work function metal stack, i.e., ALD TiN/ALD TaN/CVD TiN; (5) ALD deposition of a 1000-Å-thick B_2H_6-based W film as the gate-filling metal.

The detailed process parameters for the gate stack are summarized in Table 1. During the post-metallization annealing process, all the samples were annealed at 400 °C for 30 min using 90% N_2 + 10% H_2 forming gas (FG). A part of the samples underwent additional forming gas annealing (FGA) at 450 °C for 60 min, while the rest were taken as baseline (BL) samples.

To investigate the impacts of annealing conditions on the stress of the W film and impurity element concentration in the gate stack, we additionally prepared some test samples without patterning, by growing on blanket wafers; these

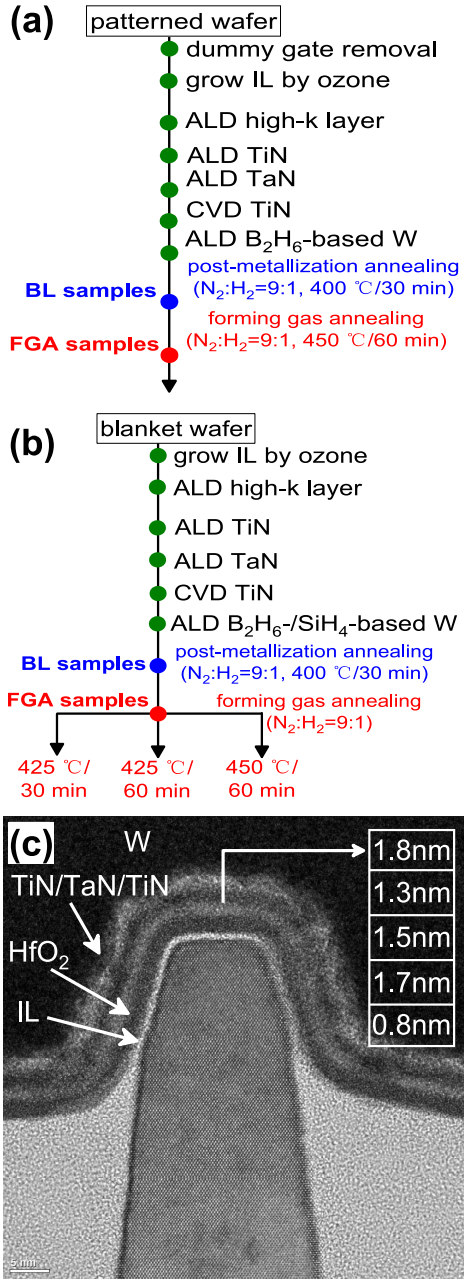

FIGURE 1. (a) Gate process flow of baseline (BL) and forming gas annealing (FGA)-optimized p-FinFETs fabricated on patterned wafers. (b) Gate process flow of test samples with different ALD W processes and annealing conditions fabricated on blanket wafers. (c) TEM image of the gate structure of the baseline p-FinFET with B_2H_6-based W film as the gate-filling metal. The thicknesses of the gate dielectric layers on the top wall are specified in the box.

test samples contained the same gate stack structures as the FinFETs and were prepared using the same gate process, the details of which are shown in Fig. 1b. Both B_2H_6-based W and SiH_4-based W were used. For each ALD W process, three different FGA conditions were applied for comparison. The TEM image in Fig. 1c shows the gate structure of the baseline p-FinFET with B_2H_6-based W; the thicknesses of the gate stacked layers are shown in Fig. 1c.

The test samples prepared on the blanket wafers were used for secondary ion mass spectroscopy (SIMS) measurements and film stress measurements. The tensile stress in the ALD W films was evaluated using a Toho FLX–2320–S film stress measurement system.

Electrical characterization was performed on patterned wafers using a Keysight B1500A semiconductor analyzer. The tested devices with a channel length of 500 nm and 4800 fins were used for the charge pumping (CP) measurement and direct-current current voltage (DCIV) measurements, whereas devices with a channel length of 500 nm and 32 fins were used for measuring the time kinetics of the threshold voltage shift (ΔV_T) and for multi-level discharging-based multi-pulse energy profiling (Multi-DMP) measurement [13], [14].

The fresh interface quality was evaluated using the CP method [15], [16], and the trap generation under constant-voltage stress was measured using the DCIV method [17]. The threshold voltage was extracted using the constant-current method, and the target I_{DLIN} was 100 nA $\times$ W/L. The ΔV_T time kinetics were characterized under a constant field stress [18] to eliminate the field reduction effect on parameter extraction [19], with a measurement delay of $\sim$10 μs. Multi-DMP characterization was used to extract the energy distribution of the pre-existing traps (ΔN_{HT}). A measurement delay of 10 μs was used for the pulsed IV sweep in the Multi-DMP test.

III. RESULTS AND DISCUSSION

A. IMPACTS OF ANNEALING CONDITION ON METAL STRAIN AND INTERFACE QUALITY

Fig. 2 shows the measured tensile stresses in the B_2H_6- and SiH_4-based W films under different annealing conditions. In terms of the baseline process, compared with SiH_4-based ALD W, which has a polycrystalline phase, B_2H_6-based ALD W exhibits an amorphous phase and thus shows a lower tensile stress [8]. For the SiH_4-based W film, the additional annealing process results in stress relaxation [20], and the tensile stress decreases with increasing annealing temperature and time (e.g., 46% loss of tensile stress under the 450 °C/60 min annealing condition). In contrast, the tensile stress in the B_2H_6-based W film is independent of the annealing conditions. The above results indicate that, compared with SiH_4-based W, optimization of the annealing process is more suitable for B_2H_6-based W because it can help adjust the impurity element concentration in the gate stack without compromising the tensile stress of the W film.

The influence of the annealing conditions on the distribution of the impurity elements in the gate stack was studied via SIMS measurements, as shown in Fig. 3. During the ALD process, TiN and $TiCl_4$ reacted with NH_3 to generate TiN and HCl, while WF_6 reacted with B_2H_6/SiH_4 to generate W and other F-containing byproducts during the ALD W process. Therefore, both F and Cl could be introduced into the gate stack. Post-metallization annealing facilitated the residual F in the W film and residual Cl in the TiN film

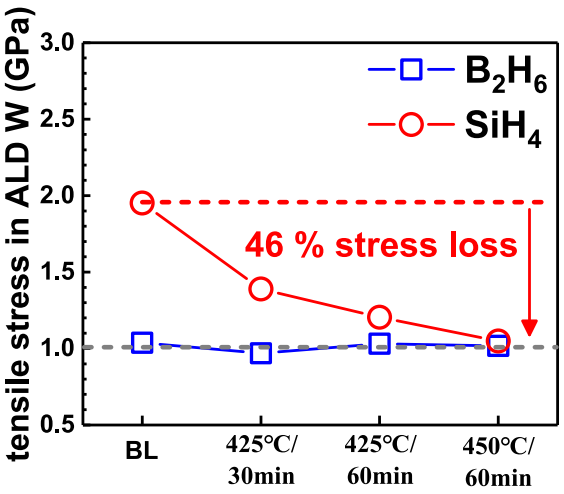

FIGURE 2. Measured stress of W films under different annealing conditions for B_2H_6- and SiH_4-based ALD W.

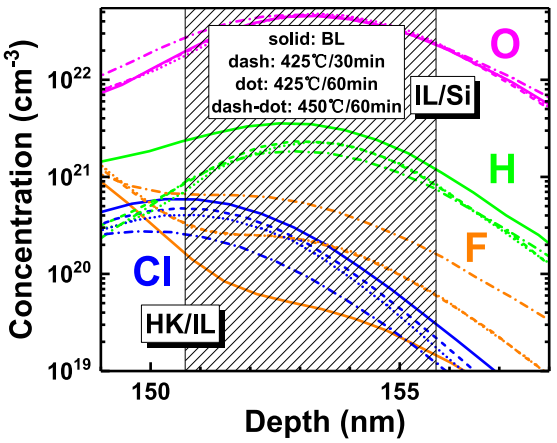

FIGURE 3. SIMS profiles of O, H, Cl and F in gate stacks for B_2H_6-based ALD W. Samples were fabricated using baseline process and different extra annealing conditions.

to diffuse into the gate stacks and provide the necessary activation energy to passivate the Si dangling bonds at the interfaces. Consequently, the annealing condition is vital for adjusting the H, Cl, and F concentrations at the interfaces and improving the interface quality. Fig. 4 shows the measured concentrations of H, Cl, and F at the interfaces. The additional annealing process reduces the H concentrations at the interfaces, and this reduction is independent of the annealing temperature and time. The Cl concentrations at the interfaces decreased after the additional annealing process and continued to decrease with increasing annealing temperature and time. On the contrary, the F concentrations at the interfaces show a significant increase after the additional annealing process with annealing temperature and time. Si–F bonds are stronger and more difficult to break under stress than Si–H and Si–Cl bonds [21]–[23]; thus, we chose an additional FGA process at 450 °C for 60 min as our optimized annealing condition for fabricating the annealing-optimized p-FinFETs, as shown in Fig. 1a.

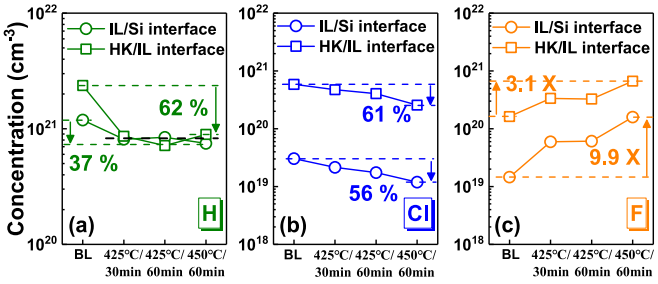

FIGURE 4. Measured concentrations of (a) H, (b) Cl, and (c) F at IL/Si and HK/IL interfaces from the SIMS results shown in Fig. 3 and plotted as a function of the annealing conditions.

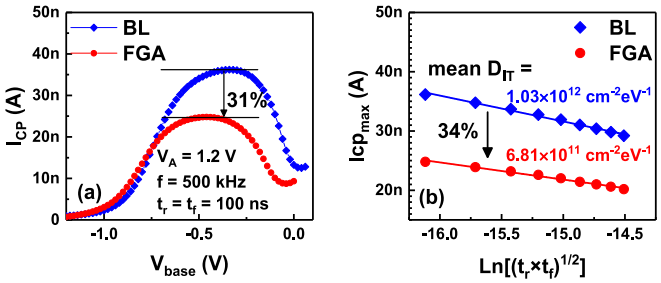

FIGURE 5. (a) Charging pumping current comparison between p-FinFETs fabricated using baseline process and optimized-annealing process. (b) Icp_{max} as a function of $ln[(t_r \times t_f)^{1/2}]$ for baseline device and annealing-optimized device; the extracted mean D_{IT} is shown.

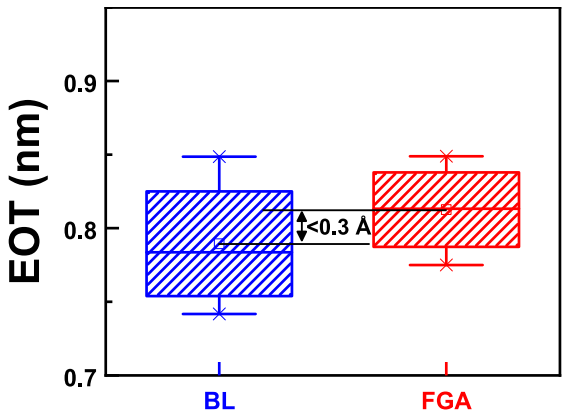

FIGURE 6. Extracted EOT values for baseline (BL) devices and forming gas annealing (FGA)-optimized devices.

Charge pumping measurements were conducted to evaluate the fresh interface states in the baseline and annealing-optimized devices, as depicted in Fig. 5. Notably, a 34% reduction in the mean D_{IT} can be achieved by optimizing the annealing conditions. Based on the SIMS results, this improvement in the interface quality can be attributed to an increase in the F concentration at the interfaces.

B. NBTI DEGRADATION

Fig. 6 plots the box charts of the equivalent oxide thickness (EOT) values. The mean EOT difference for the baseline and annealing optimized devices is within 0.3 Å. Such a small EOT difference makes the reliability comparison between different annealing conditions straightforward, and we will compare NBTI degradation under the same V_{OV}.

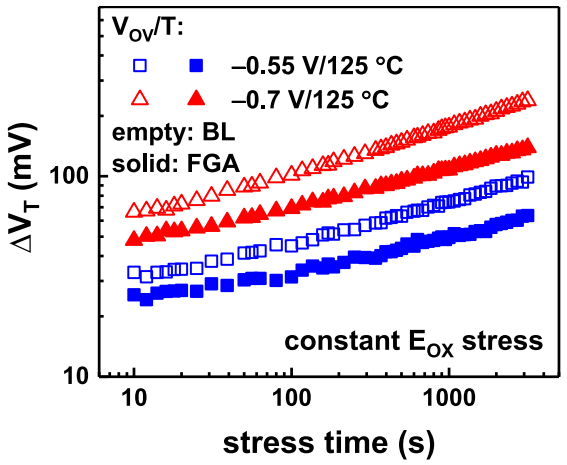

FIGURE 7. ΔV_T time kinetics at different V_{OV} and T under constant field stress for baseline devices and annealing-optimized devices. The extracted E_A and power-law FAE are shown in Fig. 9.

Fig. 7 plots the measured ΔV_T time kinetics under a constant field stress; the stress bias was increased with ΔV_T during stress to ensure a fixed V_{OV}. Compared with the baseline devices, the annealing-optimized devices exhibit lower NBTI degradation. To further study the physical mechanism behind this annealing condition dependence of NBTI, the key NBTI parameters are extracted and analyzed.

Fig. 8 shows the extracted long-term time exponents (n) of ΔV_T under different stress conditions. Notably, n of the baseline device is much higher than that of the annealing-optimized device. Based on the BAT framework [11], [24], the total degradation ΔV_T consists of three uncorrelated subcomponents: hole trapping in the pre-existing traps (ΔV_{HT}), generated interface traps (ΔV_{IT}), and generated bulk insulator traps (ΔV_{OT}). The ΔV_{HT} subcomponent represents the inherent process-induced gate insulator traps and typically saturates within 10 s, exhibiting n $\sim$ 0 at long stress times [11], [24]. The ΔV_{IT} subcomponent has a power law dependence on the time, and the long-term n is $\sim$1/6, which is confirmed by the calculated results of the double-interface reaction-diffusion (R-D) model as well as the DCIV experimental results [25]. The reported n of the ΔV_{OT} subcomponent is in the range of $\sim$1/3–1/4 [11]. Thus, the observed higher value of n for the baseline device mainly results from the higher relative contribution of the generated traps whose time exponents are much higher than that of ΔV_{HT}. Moreover, the extracted n shows an increasing trend (from 0.12 to 0.27) with V_{OV} and T, indicating an increasing contribution from ΔV_{OT}.

Figs. 9a and 9b respectively show the stress bias and temperature dependences of ΔV_T, including the extracted Arrhenius activation energy (E_A) and power-law field acceleration exponent (FAE). The baseline devices show higher values of FAE and E_A than the annealing-optimized devices. The generated traps are typically reported to exhibit stronger

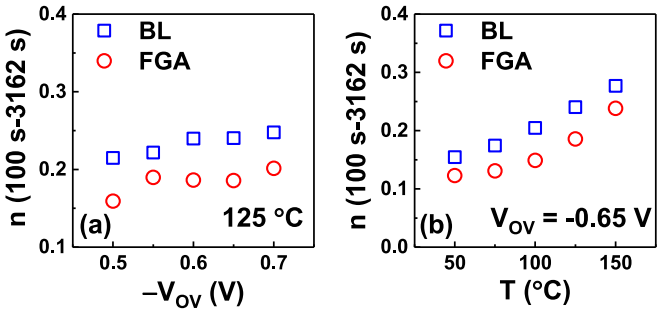

FIGURE 8. Extracted long-term power-law time exponents of ΔV_T as a function of (a) V_{OV} at 125 °C and (b) T at $V_{OV} = -0.65$ V.

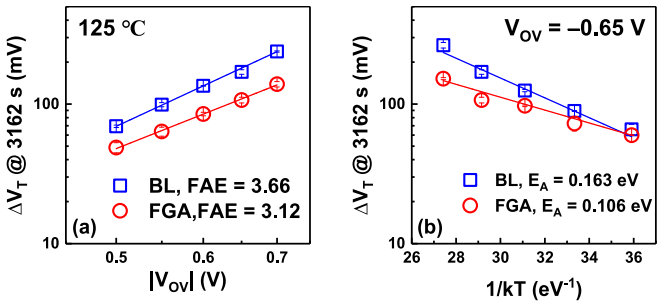

FIGURE 9. Fixed-time ΔV_T plotted as a function of (a) V_{OV} at 125 °C and (b) T at $V_{OV} = -0.65$ V for baseline devices and annealing-optimized devices.

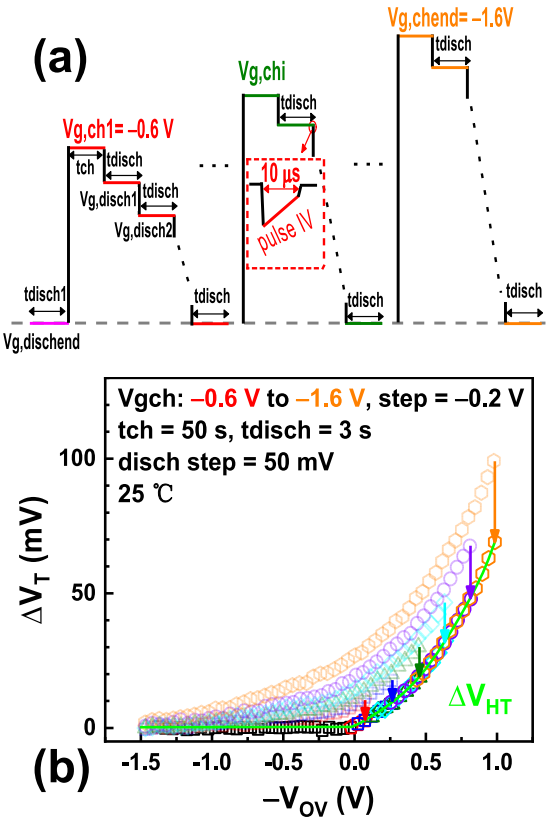

FIGURE 10. (a) V_G waveform for multi-DMP test used in this study. (b) Extraction procedure of $\Delta V_{HT} - -V_{OV}$ curve: the lowest black curve is measured on a fresh device when $V_{g,ch} =\sim V_{T0}$. The other ΔV_T–V_{OV} curves are measured on a heavily stressed device. Using the black curve as the base, data for next higher $V_{g,ch}$ are shifted down to align these two curves at the end of the lower one. This shifting procedure is carried out until the highest $V_{g,ch} = -1.6$ V is reached. The obtained ΔV_{HT}–V_{OV} curve is indicated by a thick green line.

stress bias and temperature dependences than the pre-existing traps [10], [11], [25]. Therefore, the higher values of FAE and E_A for the baseline devices could be attributed to the higher relative contribution of the generated traps, consistent with the findings of the time exponent discussed in the previous part.

C. TRAP ANALYSIS

The Multi-DMP technique was used to extract ΔV_{HT} and its corresponding trap density, ΔN_{HT}. Fig. 10a shows the V_G waveform of Multi-DMP test. First, a heavily stressed device was used and discharged under a high positive voltage ($V_{\mathrm{g,dischend}}$) to de-trap the holes in the pre-existing and generated traps; when the discharging process was completed, V_{T} was measured and used as the reference V_{T} (V_{T0}) of the Multi-DMP test. The device was then charged at a low voltage ($V_{\mathrm{g,ch1}}$) and subsequently discharged at different discharging voltages ($V_{\mathrm{g,disch}}$). V_{T} was measured at the end of each discharging process to extract ΔV_T, and the overdrive voltage V_{OV} was calculated as $V_{OV} = V_{\mathrm{g,disch}} - V_T$. Finally, this charging–discharging process was repeated at multiple higher $V_{\mathrm{g,ch}}$ values; the extracted $\Delta V_T - V_{OV}$ curves at different $V_{\mathrm{g,ch}}$ values are shown in Fig. 10b. The procedure for separating the $\Delta V_{\mathrm{HT}} - V_{OV}$ curve from the $\Delta V_T - V_{OV}$ curves is also illustrated in Fig. 10b. ΔV_{HT} is converted to ΔN_{HT} using the equation of the charge-sheet model, $\Delta N_{\mathrm{HT}} = (C_{OX} * \Delta V_{\mathrm{HT}})/q$, where C_{OX} is the oxide capacitance, and q is the electronic charge. The $V_{OV} - E_f - E_v$ relationship can be obtained by fitting the measured C–V results to the C–V simulations, where E_f corresponds to the detected trap energy level E_T.

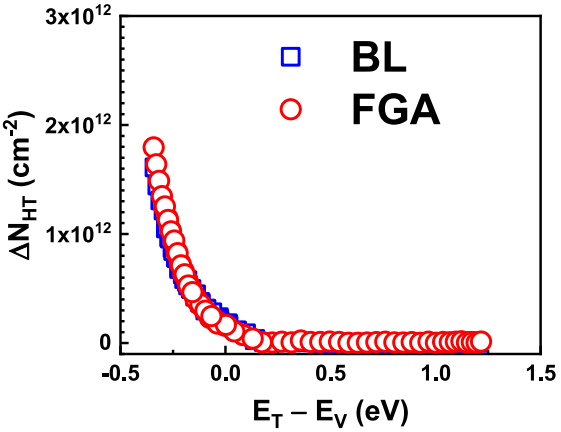

FIGURE 11. Comparison of the energy distribution of ΔN_{HT} between baseline devices and annealing-optimized devices. E_T is the energy level of the traps.

Fig. 11 shows the extracted energy profile of ΔN_{HT}. We previously [14] reported that the ΔN_{HT} energy distribution

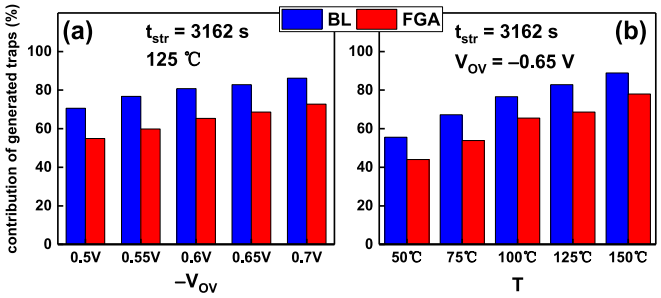

FIGURE 12. Ratio of generated traps to ΔV_T at (a) fixed T and different V_{OV} and (b) fixed V_{OV} and different T.

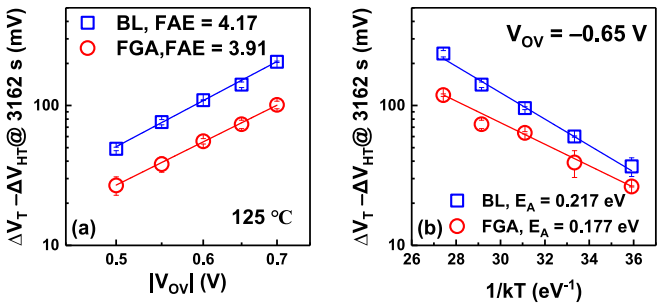

FIGURE 13. Fixed-time generated traps plotted as a function of (a) V_{OV} at 125 °C and (b) T at V_{OV} = –0.65 V for baseline devices and annealing-optimized devices.

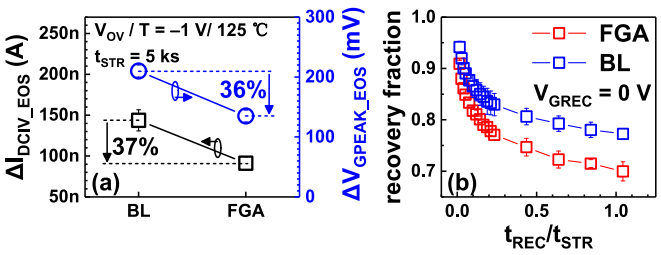

FIGURE 14. (a) Measured ΔI_{DCIV_EOS} and ΔV_{GPEAK_EOS}, where ΔI_{DCIV_EOS} is the generation of the DCIV peak current at the end of the stress, and ΔV_{GPEAK_EOS} is the shift in the DCIV peak voltage at the end of the stress. (b) Measured recovery fraction for the baseline devices and annealing-optimized devices; the recovery fraction is calculated by $\Delta I_{DCIV\_REC}/\Delta I_{DCIV\_EOS}$, where ΔI_{DCIV_REC} is the measured degradation with a recovery time of t_{REC}.

is independent of the stress conditions, indicating its as-grown nature. In this work, it is noteworthy that the ΔN_{HT} energy distribution was found to be insensitive to the annealing conditions. The generated traps can be calculated by subtracting ΔV_{HT} from ΔV_T. Fig. 12 shows the relative contribution of the generated traps at different V_{OV} and T, where the baseline devices exhibit a higher relative contribution of the generated traps than the annealing-optimized devices, as expected.

Fig. 13 compares the V_{OV} and T dependences of the generated traps. Based on the SIMS and charge pumping results discussed above, compared with the baseline devices, ΔV_{IT} of the annealing-optimized devices is expected to exhibit higher values of FAE and E_A because of the large number of strong Si–F bonds formed at the interfaces. However, our

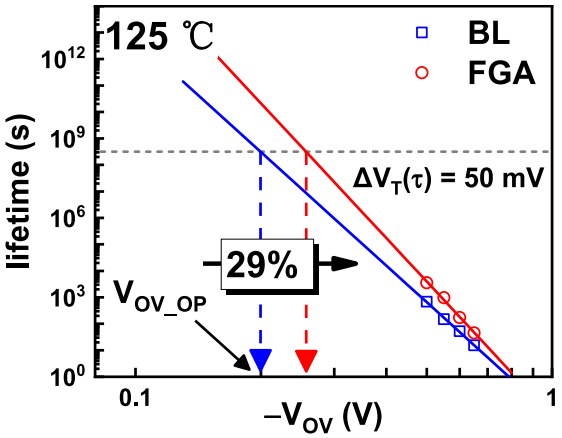

FIGURE 15. Comparison of extrapolated lifetimes between baseline devices and annealing-optimized devices.

results show that the baseline devices have higher values of FAE and E_A of the generated traps than the annealing-optimized devices. Considering that ΔV_{OT} has stronger stress bias and temperature dependences than ΔV_{IT} [10], one possible explanation is that the ΔV_{OT} subcomponent is suppressed and its relative contribution to the generated traps decreases by annealing optimization. To provide experimental evidence for this hypothesis, DCIV characterization was performed to measure the generated traps during and after the NBTI stress. The devices were stressed under 5 ks of DC stress and subsequently recovered at 0 V for 5 ks to record the recovery kinetics. Fig. 14a shows the measured ΔI_{DCIV} and ΔV_{GPEAK} at the end of the stress; notably, the annealing-optimized devices have lower ΔI_{DCIV_EOS} and ΔV_{GPEAK_EOS} than the baseline devices, indicating that the trap generation is suppressed by annealing optimization. Moreover, Fig. 14b shows that the annealing-optimized devices have a lower recovery fraction than the baseline devices, which can be attributed to the different recovery features of the generated interface traps (ΔN_{IT}) and bulk insulator traps (ΔN_{OT}). The passivation of ΔN_{OT} is much slower than that of ΔN_{IT}, and ΔN_{OT} is typically modeled as a quasi-permanent subcomponent [25], [26], whereas the recovery of ΔN_{IT} is modeled by the R-D model [11]. The ratio of the quasi-permanent ΔN_{OT} to the recoverable ΔN_{IT} decreases with annealing optimization, resulting in a lower recovery fraction for the annealing-optimized devices.

D. LIFETIME COMPARISON

Fig. 15 shows the extrapolated NBTI lifetimes at 125 °C for the baseline devices and annealing-optimized devices. The lifetime criterion was set to 50 mV, and the corresponding operation overdrive voltage V_{OV_OP} was determined. Notably, V_{OV_OP} is improved by 29% by optimizing the annealing conditions, mainly attributed to the reduction in the number of generated traps.

IV. CONCLUSION

In this study, we experimentally investigated the impact of post-metallization annealing conditions on the NBTI of Si p-FinFETs with ALD W-filling metal. In terms of the tensile stress of the W film, unlike the SiH_4-based W, the stress in B_2H_6-based W was found to be independent of the annealing condition, indicating that optimization of the annealing process is more applicable for B_2H_6-based devices than for SiH_4-based devices. For devices with B_2H_6-based W, a 34% reduction in the fresh interface states can be achieved by optimizing the annealing conditions, attributed to the increase in F concentration at the interfaces. NBTI degradation decreased, and V_{OVOP} could be improved by 29% by choosing an appropriate annealing condition, resulting from the reduction in the number of generated traps. The energy profile of ΔN_{HT} was insensitive to the annealing conditions, leading to a lower relative contribution of the generated traps to NBTI degradation in annealing-optimized devices compared to baseline devices. Moreover, the relative contribution from the bulk trap generation to the total generated traps could be reduced by optimizing the annealing conditions. These results indicate that optimization of the annealing process is a convenient and efficient method to alleviate the NBTI of p-FinFETs with a B_2H_6-based W gate-filling metal.

REFERENCES

[1] H. Mertens *et al.*, "Si-cap-free SiGe p-channel FinFETs and gate-all-around transistors in a replacement metal gate process: Interface trap density reduction and performance improvement by high-pressure deuterium anneal," in *Proc. Symp. VLSI Technol. (VLSI-T)*, Kyoto, Japan, Jun. 2015, pp. T142–T143, doi: 10.1109/VLSIT.2015.7223654.

[2] A. Nainani *et al.*, "Is strain engineering scalable in FinFET era?: Teaching the old dog some new tricks," in *IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, Dec. 2012, pp. 1–4, doi: 10.1109/IEDM.2012.6479065.

[3] P.-Y. Yang, "Effect of gate-line-end-induced stress and its impact on device's characteristics in FinFETs," *IEEE Electron Device Lett.*, vol. 37, no. 7, pp. 910–912, Jul. 2016, doi: 10.1109/LED.2016.2565901.

[4] J. W. Klaus, S. J. Ferro, and S. M. George, "Atomic layer deposition of tungsten using sequential surface chemistry with a sacrificial stripping reaction," *Thin Solid Films*, vol. 360, nos. 1–2, pp. 145–153, Feb. 2000, doi: 10.1016/s0040-6090(99)01076-7.

[5] S.-H. Kim, N. Kwak, J. Kim, and H. Sohn, "A comparative study of the atomic-layer-deposited tungsten thin films as nucleation layers for W-plug deposition," *J. Electrochem. Soc.*, vol. 153, no. 10, pp. G887–G893, Jul. 2006, doi: 10.1149/1.2222966.

[6] G. L. Wang *et al.*, "Application of atomic layer deposition tungsten (ALD W) as gate filling metal for 22 nm and beyond nodes CMOS technology," *ECS J. Solid State Sci. Technol.*, vol. 3, no. 4, pp. P82–P85, Feb. 2014, doi: 10.1149/2.015404jss.

[7] Q. Xu *et al.*, "Application of ALD W films as gate filling metal in 22 nm HKMG-last integration: Evaluation and improvement of the adhesion in CMP process," *Microelectron. Eng.*, vol. 137, pp. 43–46, Apr. 2015, doi: 10.1016/j.mee.2015.01.007.

[8] G. L. Wang *et al.*, "pMOSFETs featuring ALD W filling metal using SiH4 and B_2H_6 precursors in 22 nm Node CMOS technology," *Nanoscale Res. Lett.*, vol. 12, pp. 1–6, Apr. 2017, doi: 10.1186/s11671-017-2080-2.

[9] L. Zhou *et al.*, "Comparative study on NBTI kinetics in Si p-FinFETs with B_2H_6-based and SiH4-based atomic layer deposition tungsten (ALD W) filling metal," *Microelectron. Rel.*, vol. 107, Apr. 2020, Art. no. 113627, doi: 10.1016/j.microrel.2020.113627.

[10] N. Parihar, R. Tiwari, and S. Mahapatra, "Modeling channel length scaling impact on NBTI in RMG Si p-FinFETs," in *Proc. Int. Conf. Simulat. Semicond. Process. Devices (SISPAD)*, Austin, TX, USA, Sep. 2018, pp. 176–180, doi: 10.1109/SISPAD.2018.8551740.

[11] S. Mahapatra and N. Parihar, "Modeling of NBTI using BAT framework: DC–AC stress-recovery kinetics, material, and process dependence," *IEEE Trans. Device Mater. Rel.*, vol. 20, no. 1, pp. 4–23, Mar. 2020, doi: 10.1109/TDMR.2020.2967696.

[12] A. Shickova *et al.*, "Novel, effective and cost-efficient method of introducing fluorine into metal/Hf-based gate stack in MuGFET and planar SOI devices with significant BTI improvement," in *Proc. Symp. VLSI Technol. (VLSI-T)*, Kyoto, Japan, Jun. 2007, pp. 112–113, doi: 10.1109/vlsit.2007.4339748.

[13] Z. Ji *et al.*, "An investigation on border traps in III–V MOSFETs with an $In_{0.53}Ga_{0.47}As$ channel," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3633–3639, Nov. 2015, doi: 10.1109/TED.2015.2475604.

[14] L. Zhou *et al.*, "Comparative study on the energy profile of NBTI-related defects in Si and ferroelectric p-FinFETs," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Dallas, TX, USA, Apr. 2020, pp. 1–6, doi: 10.1109/IRPS45951.2020.9129562.

[15] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. D. Keersmaecker, "A reliable approach to charge-pumping measurements in MOS transistors," *IEEE Trans. Electron Devices*, vol. 31, no. 1, pp. 42–53, Jan. 1984, doi: 10.1109/T-ED.1984.21472.

[16] J.-P. Han *et al.*, "Energy distribution of interface traps in high-k gated MOSFETs," in *Symp. VLSI-Technol. Dig. Tech. (VLSI-T)*, Kyoto, Japan, Jun. 2003, pp. 161–162, doi: 10.1109/VLSIT.2003.1221135.

[17] A. Neugroschel, G. Bersuker, and R. Choi, "Applications of DCIV method to NBTI characterization," *Microelectron. Rel.*, vol. 47, nos. 9–11, pp. 1366–1372, Sep.–Nov. 2007, doi: 10.1016/j.microrel.2007.07.037.

[18] R. Gao, Z. Ji, J. F. Zhang, J. Marsland, and W. D. Zhang, "As-grown-generation model for positive bias temperature instability," *IEEE Trans. Electron Devices*, vol. 65, no. 9, pp. 3662–3668, Sep. 2018, doi: 10.1109/TED.2018.2857000.

[19] L. Zhou *et al.*, "An investigation of field reduction effect on NBTI parameter characterization and lifetime prediction using a constant field stress method," *IEEE Trans. Device Mater. Rel.*, vol. 20, no. 1, pp. 92–96, Mar. 2020, doi: 10.1109/TDMR.2019.2960471.

[20] D. Choi *et al.*, "Phase, grain structure, stress, and resistivity of sputter-deposited tungsten films," *J. Vac. Sci. Technol. A*, vol. 29, no. 5, p. 8, Sep. 2011, doi: 10.1116/1.3622619.

[21] K. I. Seo, R. Sreenivasan, P. C. McIntyre, and K. C. Saraswat, "Improvement in high-$k(HfO_2/SiO_2)$reliability by incorporation of fluorine," *IEEE Electron Device Lett.*, vol. 27, no. 10, pp. 821–823, Oct. 2006, doi: 10.1109/led.2006.882564.

[22] M. Chang, M. Jo, H. Park, H. Hwang, B. H. Lee, and R. Choi, "Effect of F_2 postmetallization annealing on the electrical and reliability characteristics of HfSiO gate dielectric," *IEEE Electron Device Lett.*, vol. 28, no. 1, pp. 21–23, Jan. 2007, doi: 10.1109/LED.2006.887941.

[23] W.-C. Wu *et al.*, "Fluorinated HfO2 gate dielectrics engineering for CMOS by pre- and post-CF4 plasma passivation," in *IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, Dec. 2008, pp. 1–4, doi: 10.1109/IEDM.2008.4796706.

[24] N. Parihar, N. Goel, S. Mukhopadhyay, and S. Mahapatra, "BTI analysis tool—Modeling of NBTI DC, AC stress and recovery time kinetics, nitrogen impact, and EOL estimation," *IEEE Trans. Electron Devices*, vol. 65, no. 2, pp. 392–403, Feb. 2018, doi: 10.1109/TED.2017.2780083.

[25] N. Parihar, R. G. Southwick, M. Wang, S. Member, J. H. Stathis, and S. Mahapatra, "Modeling of NBTI kinetics in RMG Si and SiGe FinFETs, part-I: DC stress and recovery," *IEEE Trans. Electron Devices*, vol. 65, no. 5, pp. 1699–1706, May 2018, doi: 10.1109/TED.2018.2819023.

[26] N. Parihar, U. Sharma, R. G. Southwick, M. Wang, J. H. Stathis, and S. Mahapatra, "Ultrafast measurements and physical modeling of NBTI stress and recovery in RMG FinFETs under diverse DC–AC experimental conditions," *IEEE Trans. Electron Devices*, vol. 65, no. 1, pp. 23–30, Jan. 2018, doi: 10.1109/TED.2017.2773122.