Development of an Ultrafast On-the-Fly I_{DLIN} Technique to Study NBTI in Plasma and Thermal Oxynitride p-MOSFETs

Vrajesh D. Maheta, *Member, IEEE*, E. Naresh Kumar, Shweta Purawat, Christopher Olsen, Khaled Ahmed, *Senior Member, IEEE*, and Souvik Mahapatra, *Senior Member, IEEE*

***Abstract*—An ultrafast on-the-fly technique is developed to study linear drain current (I_{DLIN}) degradation in plasma and thermal oxynitride p-MOSFETs during negative-bias temperature instability (NBTI) stress. The technique enhances the measurement resolution ("time-zero" delay) down to 1 μs and helps to identify several key differences in NBTI behavior between plasma and thermal films. The impact of the time-zero delay on time, temperature, and bias dependence of NBTI is studied, and its influence on extrapolated safe-operating overdrive condition is analyzed. It is shown that plasma-nitrided films, in spite of having higher N density, are less susceptible to NBTI than their thermal counterparts.**

***Index Terms*—Field acceleration, negative-bias temperature instability (NBTI), plasma oxynitride, p-MOSFET, safe-operating voltage, temperature activation, thermal oxynitride, time exponents.**

I. Introduction

THE negative-bias temperature instability (NBTI) of device parameters (linear drain current I_{DLIN}, threshold voltage V_T, transconductance G_M, etc.) is a serious reliability issue for Si oxynitride (SiON) p-MOSFETs [1]–[14]. Similar to other degradation mechanisms, the NBTI lifetime is determined by recording device parameter degradation during accelerated stress test and subsequently using suitable bias and time-dependent models [15]–[24] to extrapolate measured data from stress (high voltage, short time) to operating (low voltage, long time) conditions. However, unlike other degradation mechanisms, NBTI degradation recovers substantially once the stress is removed [25], [26]. Therefore, the conventional stress–measure–stress (SMS) technique (interrupting stress at logarithmic time intervals and performing transfer I–V sweep to extract device parameters) results in incorrect degradation magnitude and rate of temporal buildup and, hence, cannot be employed for reliable NBTI characterization.

In recent years, the aforementioned issue was addressed by employing ultrafast (UF) SMS [27], [28] and on-the-fly (OTF) I_{DLIN} [7], [8], [14], [25], [29] techniques. In UF SMS, the stress is interrupted only for a short duration ($\sim$1 μs), and the gate bias (V_G) is ramped from stress (V_{GSTRESS}) to 0 V and back to V_{GSTRESS}. During this V_G ramp, I–V is measured by using a current–voltage converter (IVC) and digital-storage oscilloscope (DSO) at the source. This technique provides a direct estimation of V_T for each stress interval and, hence, $\Delta V_T (= V_T(t) - V_{T0}$, with V_{T0} being prestress V_T) but requires smoothing of as-measured I–V data (both before and during stress) that get significantly impacted by noise inherent to the IVC–DSO measurement system. Moreover, it is not possible to determine the degradation for stress time that is much shorter than 1 s. In conventional OTF techniques [7], [8], [25], V_{GSTRESS} is held for the entire stress duration, and I_{DLIN} is continuously monitored using a standard source measure unit (SMU) such as Agilent 4156 C. Device degradation is expressed as $\Delta V = -\Delta I_{\mathrm{DLIN}}/I_{\mathrm{DLIN0}} * (V_G - V_{T0})$, which is related but not equal to ΔV_T, as the mobility degradation is not taken into account [30]. This implementation is less noisy than the UF SMS technique due to the use of standard SMU, which, however, puts a lower limit ($\sim$1 ms) on the delay in time (t_0) between the application of stress and the measurement of the first I–V point (I_{DLIN0}) at $V_G = V_{\mathrm{GSTRESS}}$, although degradation can be measured from the stress time of 1 ms and higher. Another OTF technique [29] superimposes a small ac signal on V_{GSTRESS} and measures both I_{DLIN} and G_M, and calculates the degradation as $\Delta V_T = \int \Delta I_{\mathrm{DLIN}}/G_M$. Although I_{DLIN0} is not required to be measured with a minimum possible delay after the application of stress, such an ac OTF technique gives erroneous ΔV_T [31] due to the use of a simplified G_M expression and the error due to G_M measurement. Moreover, using this technique degradation cannot be obtained for a stress time below 1 s.

In this paper (an extended version of our IEDM07 paper [14]), a UF OTF technique is developed, which is capable of measuring I_{DLIN0}, with a minimum t_0 of 1 μs, and is also capable of providing device degradation from a stress time of 1 μs and higher. The time, temperature (T), and oxide-field (E_{OX}) dependence of NBTI is studied on SiON p-MOSFETs

Manuscript received December 13, 2007; revised June 11, 2008. Current version published September 24, 2008. The review of this paper was arranged by Editor J. Suehle.

V. D. Maheta and S. Mahapatra are with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India (e-mail: vrajesh@ee.iitb.ac.in; souvik@ee.iitb.ac.in).

E. N. Kumar was with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India. He is now with the Taiwan Semiconductor Manufacturing Company, Hsinchu 300, Taiwan, R.O.C.

S. Purawat was with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India. She is now with Intel Technology India Pvt. Ltd., Bangalore 560 017, India.

C. Olsen and K. Ahmed are with the Front End Products Group, Applied Materials, Inc., Santa Clara, CA 94085 USA.

Digital Object Identifier 10.1109/TED.2008.2003224

0018-9383/$25.00 © 2008 IEEE

having different gate insulator processes, and the UF OTF results are compared with those obtained by a conventional OTF method. The impact of t_0 on extracted safe-operating V_G for a given lifetime criterion is evaluated. It is shown that the UF OTF technique captures certain features of the NBT degradation, particularly at an early stress phase normally missed by the conventional OTF method (and fast SMS). It results in a slightly different time dependence than that of the conventional OTF method, and the obtained safe-operating V_G gets impacted by time-zero delay for $t_0 \geq 10$ μs. It is also shown that the N density at the Si/SiON interface controls the magnitude, time evolution, T activation, E_{OX} acceleration, and the extrapolated safe-operating condition of the NBTI. Therefore, the SiON process that results in lower interfacial N density is less susceptible to NBTI in spite of having a higher total N dose, which opens up interesting leakage-versus-reliability-tradeoff opportunities [11].

II. Device and Experimental Details

Experiments were performed on fully processed p-MOSFETs $(W/L = 15/0.16$ μm), with plasma-nitrided oxide (PNO) and rapid thermal-nitrided oxide (RTNO) as gate insulators. The equivalent oxide thickness (EOT) and atomic $N\%$ density of the gate insulators are as follows: PNO1 (2.35 nm, 17%), PNO2 (2.14 nm, 29%), PNO3 (1.99 nm, 36%), and RTNO (2.2 nm, 6%). The PNO films were subjected to proper two-step postnitridation anneal (PNA), as described in [32], which reduces the N density at and near the Si/SiON interface. The $N\%$ values were calculated using XPS, and CV measurements followed by quantum–mechanical corrections were done to estimate the EOT of these devices.

Fig. 1(a) shows the schematic of the measurement setup. A pulse generator is connected to the gate, which gives the stress pulse when triggered. The source is connected to a custom IVC set for a gain of $10^3 - 10^4$ and DSO. The drain is connected to a switch matrix which, in turn, is connected to a dc power supply and SMU. For the first 35 ms of stress, the switch matrix keeps the drain connected to the dc supply and then switches over to SMU. This helps to prevent RC-related issues (time lag between rise in gate voltage and drain current) that affect I_{DLIN} transients at a short time. A drain bias of 0.1 V is set for the entire stress duration. The SMU is triggered first, which, in turn, triggers the DSO, and the DSO then triggers the pulse generator. This sequence makes the DSO (and SMU, although it does not capture anything meaningful until 40 ms) to begin capturing data before the gate is pulsed. Note that it is important to choose a proper rise time of the gate pulse to avoid ringing in V_G and, hence, overshooting in the measured I_{DLIN} as V_G reaches V_{GSTRESS}, which can result in the overestimation of I_{DLIN0} and the error in the calculated ΔV. A rise time of 5 μs is used in the present experiments. The DSO captures the I_{DLIN} transient up to 180 ms, and the SMU makes the measurement from 40 ms until the end of the stress time. The overlap period of 40–180 ms is used to check (and calibrate when required) I_{DLIN} readings obtained from the DSO and SMU. Note that some amount of noise always corrupts the I_{DLIN} measurement (more so for the early duration measured using the DSO) in spite of employing best practices for shielding and grounding. Therefore, a 20-point moving averaging is performed to smooth the as-measured I_{DLIN} data. Fig. 1(b) shows the applied V_G pulse and resultant I_{DLIN} transient captured using the DSO at the initiation of stress (for the PNO1 device). Note that I_{DLIN} is recorded every 1 μs as V_G rises to V_{GSTRESS}, and therefore, peak $I_{\mathrm{DLIN}} (= I_{\mathrm{DLIN0}})$ can be obtained within 1 μs($\equiv t_0$ delay) of V_G, becoming equal to V_{GSTRESS}.

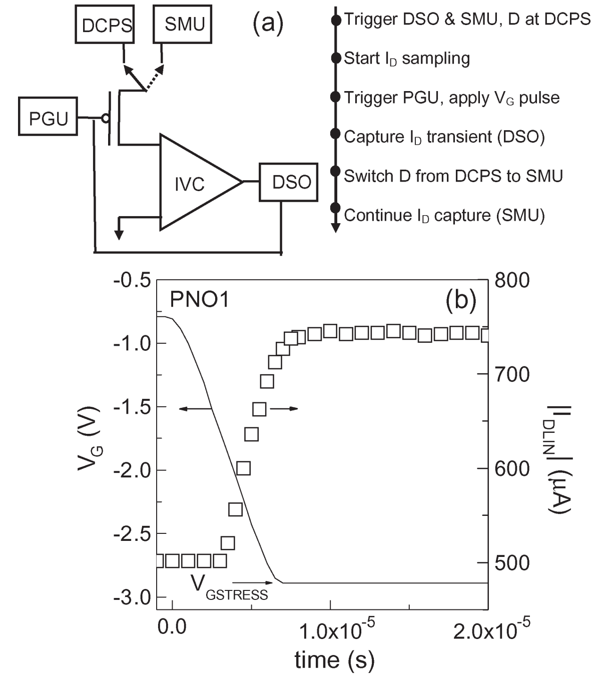

Fig. 1. (a) Schematic of UF OTF I_{DLIN} setup and sequence for NBTI stress measurement. (b) Gate voltage (V_G) and resultant I_{DLIN} transition captured using DSO during initiation of stress phase for PNO1 device. Peak I_{DLIN} $(= I_{\mathrm{DLIN0}})$ is obtained within 1 μs from $V_G = V_{\mathrm{GSTRESS}}$.

III. Results and Discussion

A. Stress Time and Temperature Dependence

Fig. 2(a) shows the captured I_{DLIN} transients for the entire stress duration (1 μs to 1000 s) for the PNO1 and RTNO devices. The oxide field (E_{OX}) and temperature (T) during stress were kept identical for both devices, and the captured I_{DLIN} is normalized to I_{DLIN0} (at $t_0 = 1$ μs) for easy comparison (as the measured I_{DLIN} was very different for these devices). Fig. 2(b) shows the time evolution of ΔV for PNO1 and RTNO devices calculated using the I_{DLIN} transients of Fig. 2(a), with I_{DLIN0} obtained at t_0 delays of 1 μs (UF OTF) and 1 ms (conventional OTF). Under identical stress E_{OX} and T, RTNO shows a much larger I_{DLIN} and, hence, ΔV degradation compared with PNO1. Due to a very high initial I_{DLIN} degradation, the RTNO device shows a very large ΔV at an early stress phase. As an example, for t_0 delay of 1 μs and stress time of 1 ms, the measured ΔV is 25 mV for RTNO, which is $\sim$10$\times$ larger than that obtained for PNO1. It is important to note that the conventional OTF technique $(t_0 = 1$ ms) fails to capture this large initial I_{DLIN} degradation for RTNO and, hence, results in a lower ΔV magnitude (due to lower value of captured

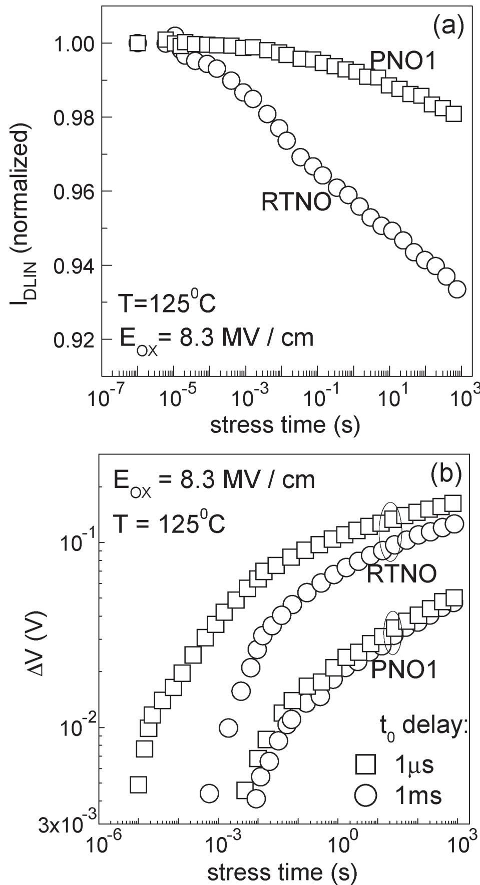

Fig. 2. (a) Captured I_{DLIN} transients over nine decades of time for PNO1 and RTNO devices stressed under identical oxide field (E_{OX}) and temperature (T). (b) Time evolution of ΔV degradation extracted from I_{DLIN} transients for t_0 delay of 1 μs and 1 ms for PNO1 and RTNO devices.

I_{DLIN0}), as shown. However, the difference in ΔV obtained by conventional OTF and UF OTF techniques is much smaller for PNO1, because the short-time I_{DLIN} degradation is small.

Fig. 3 shows the time evolution of ΔV for PNO1 and RTNO devices, obtained at different T (identical stress E_{OX}) for a t_0 delay of 1 μs. The PNO1 device shows a clear T activation for the entire duration of stress. This is contrary to the RTNO which shows negligible T dependence up to a stress time of about 1 ms and a weaker T activation for a longer stress time. It is interesting to note that the large difference in ΔV transients (magnitude and shape) between the RTNO and PNO1 can mostly be attributed to this early, T-independent degradation observed for the former device. Moreover, note that the conventional OTF technique fails to capture this T independence of ΔV transients at an early stress time for RTNO (not shown). Figs. 2 and 3 clearly show that the time evolution of I_{DLIN} and resultant ΔV degradation is different for the SiON films fabricated using different processing conditions, and is likely due to the difference in underlying physical mechanisms, as discussed elsewhere [12], [14], [33]. Although not the primary focus of this paper, the clear T dependence of ΔV seen for the PNO1 device, particularly at an early stress time, suggests interface-trap-generation (ΔN_{IT})-driven NBTI for these films [7], [8], [10]–[14]. On the other hand, the large T-independent degradation observed for RTNO devices indicates that hole

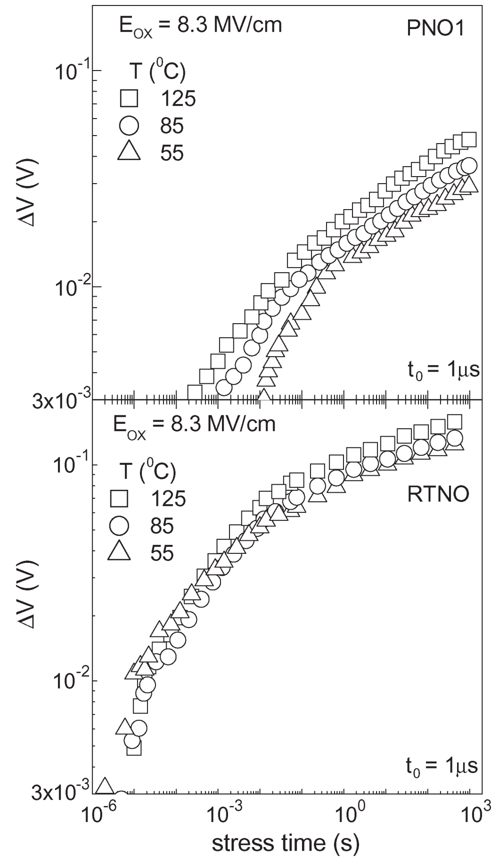

Fig. 3. Time evolution of ΔV degradation for (top) PNO1 and (bottom) RTNO devices obtained at different stress T (identical stress E_{OX}) for a t_0 delay of 1 μs.

trapping (ΔN_h) plays an equally important (or perhaps dominant) role for these films [12], as also independently suggested by others [4], [5], [29]. Finally, we wish to point out that PNO films that were not subjected to optimal PNA show a very large T-independent degradation at the early stress time, similar to those reported in [28], which is most likely due to significant hole trapping such as in the RTNO devices. A detailed study of the impact of different SiON processing conditions on NBTI using the UF OTF technique has been presented elsewhere [33].

Note that the measured NBTI data are usually extrapolated to the end of life by using power-law time dependence [1]–[3], [6]–[14]. Furthermore, the value of the obtained power-law time exponent (n), as well as its (in)dependence on stress T and E_{OX}, has been used to determine the physical mechanism of NBTI [8], [20], [23], [28]. Therefore, it is important to extract the correct value of n not only for the reliable determination of device lifetime but also for understanding the underlying physical process and for developing robust NBTI models. It is shown in Figs. 2 and 3 that the time evolution of ΔV does not follow a single power law for the entire (approximately microseconds to 1 ks) duration of stress. However, ΔV asymptotically converges to a power-law dependence for a relatively longer stress time. Fig. 4(a) shows n as a function of the t_0 delay for the PNO1 device, extracted from the log–log ΔV versus time plot over different time intervals of stress at particular E_{OX} and T. The error bar indicates the spread in n due to noise in the I_{DLIN0} measurement. Note that for a particular t_0 delay, n reduces

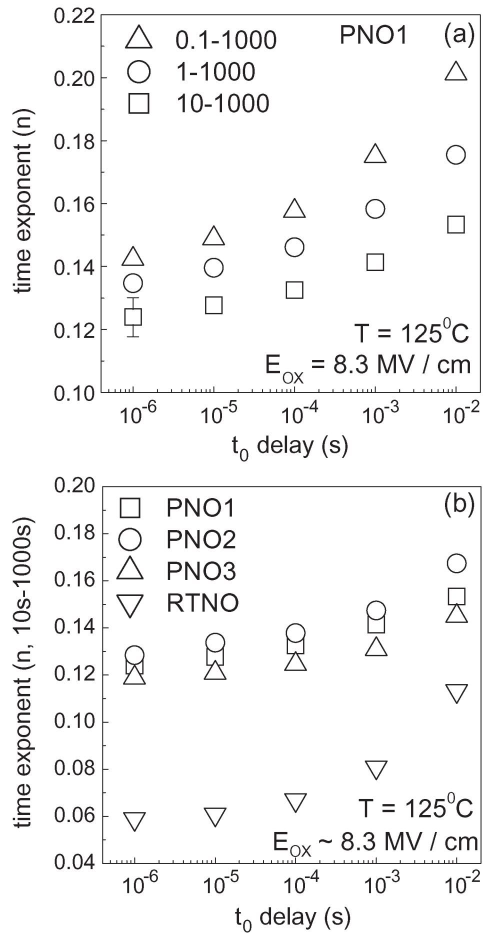

Fig. 4. (a) Extracted power-law time exponent (n) over the last two, three, and four decades of stress time as a function of t_0 delay for PNO1 device. (b) Extracted n over the last two decades of stress time as a function of t_0 delay for all RTNO and PNO devices under identical stress T and E_{OX}. Maximum error in n due to noise-induced scatter in I_{DLIN0} is ± 0.005.

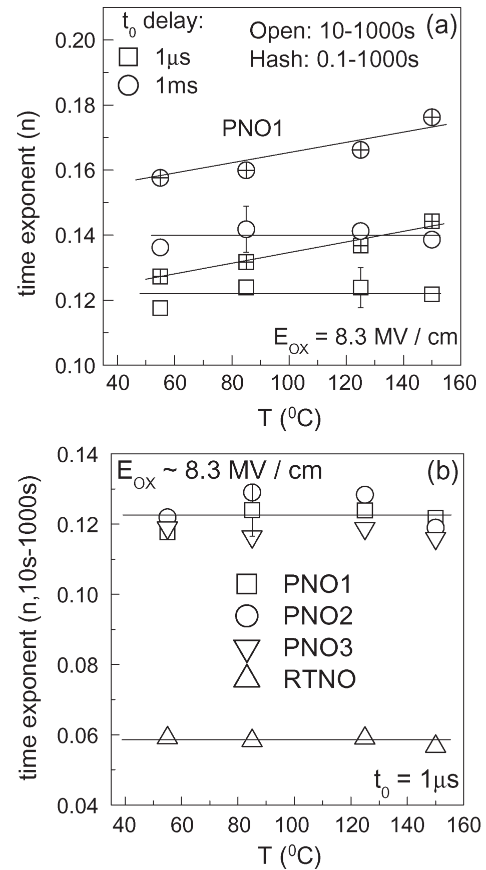

Fig. 5. (a) Extracted n over the last two and four decades of stress time as a function of stress T for a t_0 delay of 1 μs and 1 ms for PNO1 devices. (b) Extracted n over the last two decades of stress time as a function of stress T for a t_0 delay of 1 μs for all PNO and RTNO devices. Stress E_{OX} held constant for all devices. Lines are a guide to the eye.

when extracted over smaller decades of time intervals from the final stress point, i.e., n for the last two decades of stress time is less than that for the last three decades, and so on. On the other hand, for a given stress time interval, n reduces with the reduction in the t_0 delay, which is now well known [25], [26]. Note that a long-time n (extracted for $t \geq 10$ s) tends to saturate as t_0 is reduced below $\sim$10 μs and the variation (with t_0) becomes smaller than the calculated error bar. Fig. 4(b) shows the t_0 dependence of n extracted for the last two decades of stress (under identical T and E_{OX}) for all RTNO and PNO devices used in this paper. The robustness of the extracted n ($t_0 = 1$ μs, t stress ≥ 10 s) is evident from nearly identical values for all PNO devices. Note that for all values of t_0, RTNO shows a much lower n than that of the PNO devices due to differences in the underlying physical mechanism for these devices, as mentioned before and as explained in [12], [14], and [33]. However, the reduction in n with the reduction in t_0 once again saturates as t_0 becomes lower than $\sim$10 μs, which is observed for all devices. Hence, it is apparent that while the conventional OTF method results in a slightly higher n than the present UF OTF method, a faster OTF implementation (with $t_0 < 1$ μs) is unlikely to produce a significantly lower n than that obtained with the present technique.

Fig. 5(a) shows the stress T dependence (stress E_{OX} is kept constant) of n for the PNO1 device, extracted over the last two (10 s–1 ks) and last four (0.1 s–1 ks) decades of stress time for t_0 delay of 1 μs and 1 ms. Once again, the error bar denotes the variation in n due to noise in I_{DLIN0}. Note that the n extracted over the last two decades of stress, although higher for higher t_0 delay, which is as expected [25], [26], is independent of stress T [8]. However, n shows a linear T dependence when extracted over the last four decades of stress, and while n for a particular T is higher for a higher t_0 delay, the slope of n versus T is identical for $t_0 = 1$ μs and 1 ms. Note that the mechanism behind the T dependence of n for a shorter stress duration and T independence of n for a longer stress duration has been discussed before [8]. Therefore, when n is extracted over a larger time span covering short and long durations, the short-time part influences the overall T dependence, as has been observed. Fig. 5(b) shows the stress T dependence of n (under identical stress E_{OX}) for all RTNO and PNO devices used in this paper, extracted for a longer ($t \geq 10$ s) stress time for t_0 delay of 1 μs. A lower n is observed for RTNO when compared with PNO, as mentioned before. Identical values of n (within error) for all PNO devices and the T independence of n for all devices indicate that the ΔV time evolution at a longer stress time indeed follows a power law. We wish to point out that the T independence of n (when properly extracted)

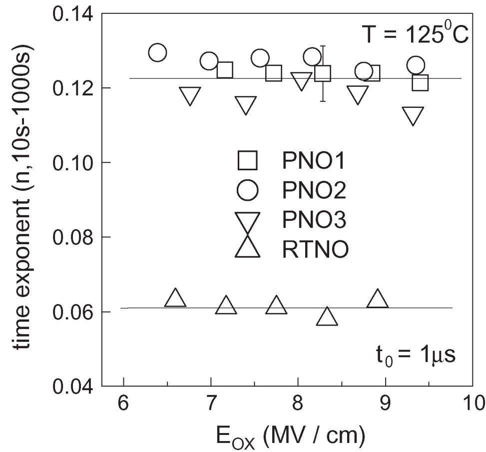

Fig. 6. Extracted n over the last two decades of stress time as a function of stress E_{OX} for all PNO and RTNO devices, obtained for a t_0 delay of 1 μs under identical stress T. Lines are a guide to the eye.

shown in this paper is in contrast to some earlier reports of n being linearly dependent on T (observed when obtained using delay I_{DLIN} measurements) [23]. However, note that such T dependence of n [23] has been shown to be an artifact of the measurement delay [8]. Figs. 4 and 5 show that the value of n, a parameter of practical as well as theoretical significance and a matter of debate in recent literature [7]–[14], [23]–[26], [28], is significantly affected by measurement condition, extraction criterion, as well as SiON process [33] and needs careful attention. It is important to use minimum time-zero delay (less than 10 μs) and longer stress time (time range of 10 s and higher for fitting) for the reliable determination of the power-law time exponent.

B. Stress Time and Field Dependence

Fig. 6 shows the stress E_{OX} dependence of n for all RTNO and PNO devices used in this paper, extracted over a longer ($t \geq 10$ s) stress time for a t_0 delay of 1 μs (under identical stress T). Note that n is independent of stress E_{OX} and a similar E_{OX} independence is also obtained for conventional OTF measurements with $t_0 = 1$ ms (not shown in this paper). Note that the invariance of n (extracted over a long stress time) as stress T and/or E_{OX} are varied, which is universally observed in SiON films fabricated using a wide range of conditions, provides the credibility of using the power-law extrapolation of measured ΔV in time to the end of life. The invariance of n with the stress E_{OX} also suggests the absence of an additional degradation mechanism (such as bulk-trap generation) in the range of E_{OX} used in these measurements [21], [34]. This is important as accelerating only the defects responsible for NBTI during stress ensures the estimation of the correct field acceleration factor, which is a prerequisite for the reliable extrapolation of measured degradation at stress to operating condition [34]. Once again, the identical values of n (within error) for all PNO devices and the E_{OX} independence of n for all devices indicate that the ΔV time evolution at a longer stress time indeed follows a power law. Finally, we wish to point out that the measured long-time exponent reduces by less than 0.01 for additional two decades of stress beyond 10^3 s (verified over a few stress conditions, not shown in this paper), possibly due to the reduction in stress E_{OX}, as explained in [35].

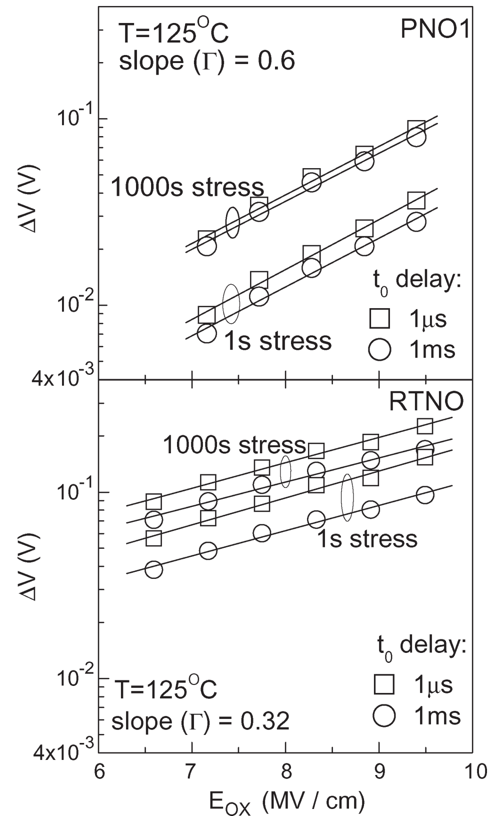

Fig. 7. E_{OX} dependence of ΔV degradation for a t_0 delay of 1 μs and 1 ms for (top) PNO1 and (bottom) RTNO devices measured at 1 and 1000 s stress time under identical stress T. Maximum error in field acceleration slope (Γ) is $\pm$0.02 cm/MV.

Fig. 7 shows ΔV versus stress E_{OX} (at constant T) for the PNO1 and RTNO devices, extracted at t stress of 1 and 1000 s for t_0 of 1 μs and 1 ms. Irrespective of t_0 and t stress, ΔV is higher for RTNO compared with PNO1 for the stress E_{OX} range studied. The impact of t_0 induced reduction in ΔV, while, more prominent at a shorter t stress, is higher for the RTNO compared with the PNO1 device. As shown in Section III-A, this is due to the large I_{DLIN} degradation observed at the short-time stress for the RTNO device. For both PNO1 and RTNO (and verified for other PNO devices, not shown in this paper), the slope (Γ) of the ΔV versus E_{OX} curve is independent of t_0 and the stress time. This t_0 invariance of the E_{OX} dependence is in contrast with that observed for the time dependence of NBTI. Fig. 8 shows the E_{OX} dependence of ΔV (normalized to EOT) for all PNO and RTNO devices used in this paper. ΔV is extracted at the stress time of 100 s for the measurement with $t_0 = 1$ μs. For the PNO devices, ΔV increases, whereas Γ reduces with an increase in the N dose. However, the RTNO devices show the largest ΔV and smallest Γ among all devices, in spite of having a much lower N density compared with the rest of the PNO. Note that for the PNO (+PNA) process, a peak N density is observed at the SiON/poly-Si interface, and the N density near the Si/SiON interface is very low, contrary to RTNO films [32], [37], [38]. Fig. 8 clearly shows that the control of the N density at the

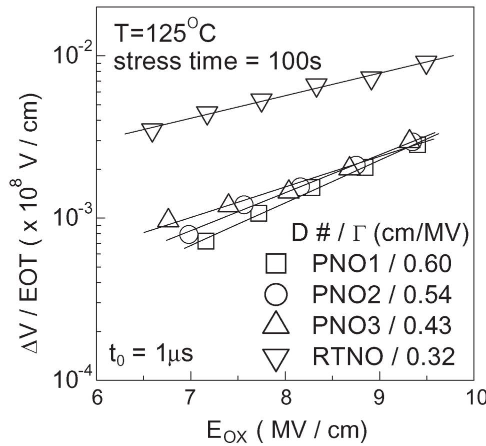

Fig. 8. E_{OX} dependence of ΔV degradation (normalized to EOT) measured at 100-s stress time and constant T with a t_0 delay of 1 μs for all PNO and RTNO devices.

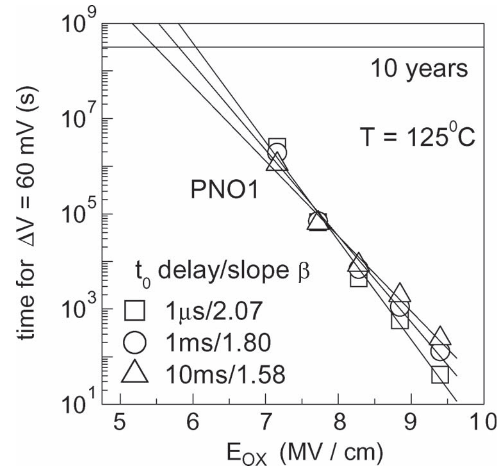

Fig. 9. Time to reach ΔV degradation of 60 mV as a function of E_{OX} for PNO1 device, obtained for different t_0 delays under identical stress T. The field acceleration slope (β) is shown.

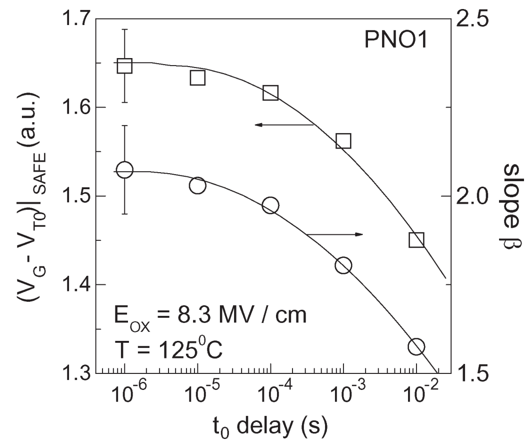

Fig. 10. Impact of t_0 delay on β and extracted safe overdrive voltage $(V_G - V_{T0})|_{\mathrm{SAFE}}$ for ten years lifetime (to reach $\Delta V = 60$ mV) for PNO1 device under identical stress E_{OX} and T. Lines are a guide to the eye. Maximum errors in β and in $(V_G - V_{T0})|_{\mathrm{SAFE}}$ are ±0.12 and ±0.04, respectively.

Si/SiON interface is very crucial to control NBTI, which is investigated in detail in [33]. Finally, it is important to note that a higher Γ (for PNO with a high N dose or RTNO) implies a higher degradation and, therefore, a lower lifetime when stress data are extrapolated to the measurement condition, as discussed in the following.

C. Extraction of Device Lifetime

As, measured data are extrapolated in time to a particular failure criterion to estimate the device lifetime, it is important to implement a UF OTF method (with $t_0 \leq 10$ μs) and choose a proper stress time interval $(t \text{ stress} \geq 10 \text{ s})$ for reliable extraction of time exponents and resultant time to failure for a given stress condition. As mentioned in Section I, ΔV is related but not equal to ΔV_T, because the impact of mobility degradation on I_{DLIN} degradation is not considered. However, as shown in [30] and [36], mobility correction introduces a multiplicative correction factor to ΔV, which obviously does not affect the power-law time exponent extracted from the log–log ΔV versus time plot. However, note that as mentioned in [36], such postmeasurement mobility correction to convert ΔV to ΔV_T works well only for devices that have a lower N dose. As the main aim of this part is to investigate the impact of the t_0 delay on the extracted safe-operating condition for devices having different (low and high) N contents, a failure criterion of $\Delta V = 60$ mV is used. The main conclusion regarding the impact of the t_0 delay does not change when a similar criterion involving ΔV_T is used (verified for the PNO1 device and not shown in this paper).

The time to reach a particular failure criterion (in this paper, it is set as $\Delta V = 60$ mV) for a given stress E_{OX} and T is determined from the log–log ΔV versus time plot. For a higher stress E_{OX}, the failure criterion is satisfied before the end of stress, whereas for lower stress E_{OX}, the time to fail is obtained by making the necessary extrapolation using n obtained over the last two decades $(t \geq 10 \text{ s})$ of stress time. Fig. 9 shows the obtained time to failure for different stress E_{OX} (fixed stress T) for the PNO1 device, obtained for different t_0 delays of the ΔV measurement. Note that the time to fail obtained at a higher stress E_{OX} increases with the increase in the t_0 delay due to the reduction in ΔV. However, at a lower stress E_{OX}, the increase in n due to the increase in the t_0 delay results in a lower time to failure (this is normally true for most devices and certainly for all devices used in this paper—not explicitly shown—unless a dramatic reduction is seen in long-time ΔV magnitude as t_0 is increased). Therefore, the slope (β) of the time to fail versus E_{OX} plot reduces at a higher t_0, as shown (not to be confused with the t_0 invariance of Γ). Hence, the extrapolated (a simple exponential field dependence is used in this paper) E_{OX} for meeting a particular lifetime criterion (in this paper, it is set as ten years for dc stress) reduces for the higher t_0 delay and results in the reduction of safe overdrive $(V_G - V_{T0})|_{\mathrm{SAFE}}$ (which is calculated back from extrapolated E_{OX}). Fig. 10 shows the impact of the t_0 delay on β and calculated $(V_G - V_{T0})|_{\mathrm{SAFE}}$ for the PNO1 device for $T = 125$ °C stress. It can be clearly seen that for $t_0 \geq 10$ μs, a nonnegligible reduction is observed in β and extrapolated $(V_G - V_{T0})|_{\mathrm{SAFE}}$. Similar results were obtained for other PNO and RTNO devices used in this paper. It is important to note that the extracted $(V_G - V_{T0})|_{\mathrm{SAFE}}$ using conventional OTF

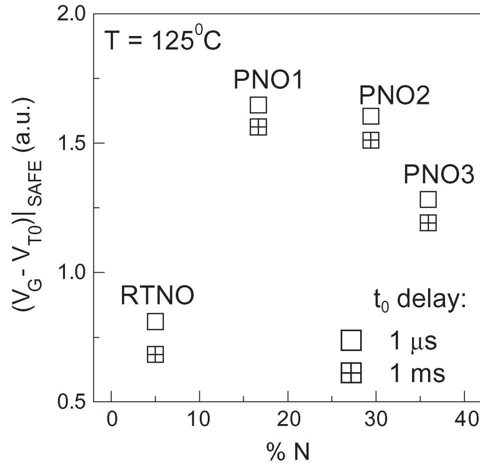

Fig. 11. $(V_G - V_{T0})|_{\mathrm{SAFE}}$ at $T = 125$ °C as a function of atomic $N\%$ extracted for a t_0 delay of 1 μs and 1 ms for all PNO and RTNO devices.

($t_0 = 1$ ms) is always underestimated, and a process qualified with such a method is likely to pass qualification when also using UF OTF.

Fig. 11 shows the calculated $(V_G - V_{T0})|_{\mathrm{SAFE}}$ as a function of atomic $N\%$, extracted for PNO and RTNO devices for the t_0 delay of 1 μs and 1 ms at $T = 125$ °C stress. Note that the extracted $(V_G - V_{T0})|_{\mathrm{SAFE}}$ is higher for the lower t_0 delay for all devices used in this paper, and the difference between the UF OTF and conventional OTF is slightly more for RTNO compared with that for the PNO devices. Finally, note that in spite of a lower $N\%$, the calculated $(V_G - V_{T0})|_{\mathrm{SAFE}}$ is much lower for RTNO compared with that for the PNO devices, whereas for the PNO devices, $(V_G - V_{T0})|_{\mathrm{SAFE}}$ reduces with an increase in $N\%$. This is fully consistent with higher ΔV and lower Γ observed (see Fig. 8), as the N density is increased at the Si/SiON interface.

IV. Conclusion

To summarize, a novel UF OTF technique is developed, which is capable of measuring I_{DLIN} degradation during NBTI from 1 μs of the application of stress bias (time-zero delay), which is much faster than that (1 ms) available for a conventional OTF technique. The technique has been employed to study the time, temperature, and field dependence of degradation during NBTI stress, on PNO and RTNO p-MOSFETs having different N density, and distribution in the gate insulator. The time-zero delay (and, therefore, whether a UF OTF or conventional OTF technique is used) impacts the magnitude, time, and T dependence but does not impact the field dependence of degradation. In spite of a lower $N\%$, RTNO devices show very large, T-independent degradation at early (t stress up to $\sim$1 ms) stress time and, as a result, a much larger overall degradation compared with that of PNO devices. It is shown that the Si/SiON interfacial N density (more for RTNO compared with PNO) controls the NBTI degradation; an increase in the interfacial N density increases the magnitude but reduces the long-time power-law exponent, T activation, and field acceleration. Finally, the field dependence of time to failure and extrapolated safe-operating bias is shown to be influenced by the time-zero delay for $t_0 > 10$ μs, and PNO devices show a much larger safe-operating bias than RTNO for a given failure criterion in spite of a higher overall N dose but lower N density at the Si/SiON interface.

Acknowledgment

The authors would like to thank M. A. Alam of Purdue University, West Lafayette, IN, for the useful discussion.

References

[1] Y. Mitani, M. Nagamine, H. Satake, and A. Toriumi, "NBTI mechanism in ultra-thin gate dielectric–nitrogen-originated mechanism in SiON," in *IEDM Tech. Dig.*, 2002, pp. 509–512.
[2] S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang, W. Y. Teo, and L. Chan, "Neighboring effect in nitrogen-enhanced negative bias temperature instability," in *Proc. Solid State Devices Mater.*, 2003, pp. 70–71.
[3] S. Tsujikawa, T. Mine, K. Watanabe, Y. Shimamoto, R. Tsuchiya, K. Ohnishi, T. Onai, J. Yugami, and S. Kimura, "Negative bias temperature instability of pMOSFETs with ultra-thin SiON gate dielectrics," in *Proc. Int. Rel. Phys. Symp.*, 2003, pp. 183–188.
[4] V. Huard, F. Monsieur, G. Ribes, and S. Bruyere, "Evidence for hydrogen-related defects during NBTI stress in p-MOSFETs," in *Proc. Int. Rel. Phys. Symp.*, 2003, pp. 178–182.
[5] V. Huard and M. Denais, "Hole trapping effect on methodology for DC and AC negative bias temperature instability measurements in PMOS transistors," in *Proc. Int. Rel. Phys. Symp.*, 2004, pp. 40–45.
[6] Y. Mitani, "Influence of nitrogen in ultra-thin SiON on negative bias temperature instability under AC stress," in *IEDM Tech. Dig.*, 2004, pp. 117–120.
[7] A. T. Krishnan, C. Chancellor, S. Chakravarthi, P. E. Nicollian, V. Reddy, A. Varghese, R. B. Khamankar, and S. Krishnan, "Material dependence of hydrogen diffusion: Implications for NBTI degradation," in *IEDM Tech. Dig.*, 2005, pp. 688–691.
[8] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. Alam, "On the dispersive versus arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: Measurements, theory, and implications," in *IEDM Tech. Dig.*, 2005, pp. 684–687.
[9] K. Sakuma, D. Matsushita, K. Muraoka, and Y. Mitani, "Investigation of nitrogen-originated NBTI mechanism in SiON with high-nitrogen concentration," in *Proc. Int. Rel. Phys. Symp.*, 2006, pp. 454–460.
[10] G. Gupta, S. Mahapatra, L. Leela Madhav, D. Varghese, K. Ahmed, and F. Nouri, "Interface-trap driven NBTI for ultrathin (EOT $\sim$ 12Å) plasma and thermal nitrided oxynitrides," in *Proc. Int. Rel. Phys. Symp.*, 2006, pp. 731–732.
[11] A. E. Islam, G. Gupta, S. Mahapatra, A. Krishnan, K. Ahmed, F. Nouri, A. Oates, and M. A. Alam, "Gate leakage vs. NBTI in plasma nitrided oxides: Characterization, physical principles, and optimization," in *IEDM Tech. Dig.*, 2006, pp. 329–332.
[12] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, "On the physical mechanism of NBTI in silicon oxynitride p-MOSFETs: Can differences in insulator processing conditions resolve the interface trap generation versus hole trapping controversy?" in *Proc. Int. Rel. Phys. Symp.*, 2007, pp. 1–9.
[13] D. Varghese, G. Gupta, L. Madhav, D. Saha, K. Ahmed, F. Nouri, and S. Mahapatra, "Physical mechanism and gate insulator material dependence of generation and recovery of negative bias temperature instability in p-MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 7, pp. 1672–1680, Jul. 2007.
[14] E. N. Kumar, V. D. Maheta, S. Purawat, A. E. Islam, C. Olsen, K. Ahmed, M. Alam, and S. Mahapatra, "Material dependence of NBTI physical mechanism in silicon oxynitride (SiON) p-MOSFETs: A comprehensive study by ultra-fast on-the-fly (UF-OTF) I_{DLIN} technique," in *IEDM Tech. Dig.*, 2007, pp. 809–812.
[15] K. O. Jeppson and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MOS devices," *J. Appl. Phys.*, vol. 48, no. 5, pp. 2004–2014, May 1977.
[16] S. Ogawa and N. Shiono, "Generalized diffusion-reaction model for the low-field charge-buildup instability at $Si-SiO_2$ interface," *Phys. Rev. B, Condens. Matter*, vol. 51, no. 7, pp. 4218–4230, Feb. 1995.
[17] G. V. Gadiyak, "Numerical simulation of hydrogen redistribution in thin SiO_2 films under electron injection in high fields," *Appl. Surf. Sci.*, vol. 113/114, pp. 627–630, 1997.

[18] A. Haggag, W. McMahon, K. Hess, K. Cheng, J. Lee, and J. Lyding, "High-performance chip reliability from short-time-tests-statistical models for optical interconnect and HCI/TDDB/NBTI deep-submicron transistor failures," in *Proc. Int. Rel. Phys. Symp.*, 2001, pp. 271–279.

[19] M. Alam, "A critical examination of the mechanics of dynamic NBTI for pMOSFETs," in *IEDM Tech. Dig.*, 2003, pp. 345–348.

[20] S. Chakravarthi, A. T. Krishnan, V. Reddy, C. F. Machala, and S. Krishnan, "A comprehensive framework for predictive modeling of negative bias temperature instability," in *Proc. Int. Rel. Phys. Symp.*, 2004, pp. 273–282.

[21] M. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectron. Reliab.*, vol. 45, no. 1, pp. 71–81, Jan. 2005.

[22] S. Zafar, "Statistical mechanics based model for negative bias temperature instability induced degradation," *J. Appl. Phys.*, vol. 97, no. 10, pp. 103 709-1–103 709-9, May 2005.

[23] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification," in *Proc. Int. Rel. Phys. Symp.*, 2005, pp. 381–387.

[24] T. Grasser, W. Gos, V. Sverdlov, and B. Kaczer, "The universality of NBTI relaxation and its implications for modeling and characterization," in *Proc. Int. Rel. Phys. Symp.*, 2007, pp. 268–280.

[25] S. Rangan, N. Mielke, and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability," in *IEDM Tech. Dig.*, 2003, pp. 341–344.

[26] M. Ershov, S. Saxena, H. Karbasi, S. Winters, S. Minehane, J. Babcock, R. Lindley, P. Clifton, M. Redford, and A. Shibkov, "Dynamic recovery of negative bias temperature instability in p-type metal–oxide–semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 83, no. 8, pp. 1647–1649, Aug. 2003.

[27] T. L. Yang, M. F. Li, C. Shen, C. H. Ang, Z. Chunxiang, Y. C. Yeo, G. Samudra, S. C. Rustagi, and M. B. Yu, "Fast and slow dynamic NBTI components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application," in *VLSI Symp. Tech. Dig.*, 2005, pp. 92–93.

[28] C. Shen, M. F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra, and Y. C. Yeo, "Characterization and physical origin of fast V_{th} transient in NBTI of pMOSFETs with SiON dielectric," in *IEDM Tech. Dig.*, 2006, pp. 1–4.

[29] M. Denais, C. Parthasarathy, G. Ribes, Y. Rey-Tauriac, N. Revil, A. Bravaix, V. Huard, and F. Perrier, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFETs," in *IEDM Tech. Dig.*, 2004, pp. 109–112.

[30] A. E. Islam, E. N. Kumar, H. Das, S. Purawat, V. Maheta, H. Aono, E. Murakami, S. Mahapatra, and M. A. Alam, "Theory and practice of on-the-fly and ultra-fast V_T measurements for NBTI degradation: Challenges and opportunities," in *IEDM Tech. Dig.*, 2007, pp. 805–808.

[31] H. Aono, E. Murakami, K. Shiga, F. Fujita, S. Yamamoto, M. Ogasawara, Y. Yamaguchi, K. Yanagisawa, and K. Kubota, "A study of SRAM NBTI by OTF measurement," in *Proc. Int. Rel. Phys. Symp.*, 2008, pp. 67–71.

[32] C. Olsen, "Two-step post nitridation annealing for lower EOT plasma nitrided gate dielectrics," U.S. Patent 0 175 961A1, Sep. 9, 2004.

[33] V. D. Maheta, C. Olsen, K. Ahmed, and S. Mahapatra, "The impact of nitride engineering on silicon oxynitride gate dielectric in negative-bias temperature instability of p-MOSFETs: A study by using ultrafast on-the-fly $\mathrm{I_{DLIN}}$ technique," *IEEE Trans. Electron Devices*, vol. 55, no. 7, pp. 1630–1638, Jul. 2008.

[34] S. Mahapatra, P. Bharath Kumar, and M. A. Alam, "Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-MOSFETs," *IEEE Trans. Electron Devices*, vol. 51, no. 9, pp. 1371–1379, Sep. 2004.

[35] A. E. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, and M. A. Alam, "Recent issues in negative-bias temperature instability: Initial degradation, field dependence of interface trap generation, hole trapping effects and relaxation," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2143–2154, Sep. 2007.

[36] A. E. Islam, H. Das, S. Mahapatra, and M. A. Alam, "Mobility degradation due to interface traps in plasma oxynitride PMOS devices," in *Proc. Int. Rel. Phys. Symp.*, 2008, pp. 87–96.

[37] J. R. Shallenberger, D. A. Cole, and S. W. Novak, "Characterization of silicon oxynitride thin films by X-ray photoelectron spectroscopy," *J. Vac. Sci. Technol. A, Vac. Surf. Films*, vol. 17, no. 4, pp. 1086–1090, Jul. 1999.

[38] S. Rauf, S. Lim, and P. L. G. Ventzek, "Model for nitridation of nanoscale SiO_2 thin films in pulsed inductively coupled N_2 plasma," *J. Appl. Phys.*, vol. 98, no. 2, 024 305, Jul. 2005.

Vrajesh D. Maheta (M'06) received the B.E. degree in electronics from Sardar Patel University, Gujarat, India, in 1993 and the M.E. degree in microelectronics from Birla Institute of Technology and Science, Pilani, India, in 2002. He is currently working toward the Ph.D. degree in electrical engineering in the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India.

His research interests are in the field of semiconductor device physics and simulation, modeling, and characterization of CMOS silicon devices. He is working on gate insulator reliability issues, particularly bias temperature instability in advanced CMOS silicon devices.

E. Naresh Kumar received the M.Sc. degree in physics from Sri Sathya Sai Institute of Higher Learning, Prasanthinilayam, India, in 2005 and the M.Tech. degree in electrical engineering from the Indian Institute of Technology Bombay (IIT Bombay), Mumbai, India, in 2007.

He was with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India. Since 2007, he has been with the Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan, R.O.C. His current research interests are in the areas of physics and simulation of novel semiconductor devices. He has also worked on negative-bias temperature instability in SiON and high-k dielectrics.

Shweta Purawat received the B.Eng. degree in electronics and instrumentation from Shri G.S. Institute Of Technology and Science, Indore, India, in 2005 and the M.Tech. degree in electrical engineering from the Indian Institute of Technology Bombay (IIT Bombay), Mumbai, India, in 2007.

She was with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India. She is currently with Intel Technology India Pvt. Ltd., Bangalore, India, where she is currently working on the circuit design on 32-nm technology node. She has also worked on negative-bias temperature instability in SiON and high-k dielectrics.

Christopher Olsen received the B.S. degree in materials engineering from the University of California (UC), Los Angeles, in 1991 and the M.S and Ph.D. degrees in materials science and engineering from the UC, Berkeley, in 1994 and 1998, respectively, with research focus on semiconductor epitaxy, characterization, and far infrared detector fabrication.

Since 2000, he has been a Senior Member of the Technical Staff with the Front End Products Group, Applied Materials, Inc., Santa Clara, CA. Prior to joining Applied Materials, Inc., he was a Senior Process Engineer at VLSI Technology, San Jose, CA, and focused on 0.15-μm CMOS gate dielectric development and shallow-trench isolation (STI) corner-rounding techniques. He holds patents on a methodology for $N\%$ determination in oxynitrides and STI corner-rounding techniques.

Khaled Ahmed (M'97–SM'04) received the B.S. and M.S. degrees in electrical engineering from Ain Shams University, Cairo, Egypt, in 1991 and 1994, respectively, and the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, in 1998.

In 1997, he was an Intern with AMD, Sunnyvale, CA, where he worked on advanced gate dielectrics for high-performance CMOS applications, and in 1998, he was a Senior Engineer, where he worked on MOS device characterization and reliability for logic and nonvolatile memory applications. In 1999, he was with Conexant Systems, Newport Beach, CA, where he worked on advanced gate dielectrics for CMOS applications. Since 2001, he has been with the Front End Products Group, Applied Materials, Inc., Santa Clara, where he is currently a Senior Member of the Technical Staff and has been working on the electrical characterization and reliability of ultrathin oxynitrides and high-k gate dielectrics for CMOS applications. He has authored or coauthored over 30 publications.

Dr. Ahmed has served as a technical program committee member for the IRPS and the IEEE Semiconductor Interface Specialists Conference.

Souvik Mahapatra (M'02–SM'07) received the Ph.D. degree in electrical engineering from the Indian Institute of Technology Bombay (IIT Bombay), Mumbai, India, in 1999.

He was with Bell Laboratories, Murray Hill, NJ, in 2000–2001. Since 2002, he has been with the Department of Electrical Engineering, IIT Bombay, where he is currently an Associate Professor. He has published more than 70 papers in peer-reviewed international journals and conferences; delivered invited talks at major international conferences, including the IEDM; given tutorials at IRPS; and served as a Reviewer for several international journals and conferences. His research interests are CMOS devices and Flash memory reliability.