NBTI Lifetime Prediction and Kinetics at Operation Bias Based on Ultrafast Pulse Measurement

Zhigang Ji, L. Lin, Jian Fu Zhang, Ben Kaczer, and Guido Groeseneken

***Abstract*—Predicting negative bias temperature instability (NBTI) lifetime can be dangerous since it is difficult to assess its safety margin. The common technique uses gate bias V_g acceleration to reduce the test time, and the data were typically obtained from quasi-dc measurements. Recently, it has been shown that substantial recovery occurs during the quasi-dc measurement, and the suppression of recovery requires using ultrafast pulse measurement, where time was reduced to the order of microseconds. In a real circuit, different transistors have different levels of recovery, and the worst case scenario is when recovery is suppressed. At present, there is little information on how this worst case NBTI lifetime can be predicted and whether the traditional V_g acceleration technique can still be used. This work will show that the prediction based on the V_g acceleration results in a substantial error, and its cause will be analyzed. To predict the worst case lifetime, a model for NBTI kinetics under operation gate bias is developed. This kinetics includes contributions from both as-grown and generated defects, and it no longer follows a simple power law. Based on the new kinetics, a single-test prediction method is proposed, and its safety margin is estimated to be 50%.**

***Index Terms*—Bias temperature instability, defects, degradation, device lifetime prediction, gate dielectrics, MOSFETs, NBTI, positive charging, reliability.**

I. Introduction

Negative bias temperature instability (NBTI) is limiting the lifetime of pMOSFETs with SiON gate dielectric and has recently received much attention [1]–[12]. The NBTI lifetime is typically defined as the time for the threshold voltage shift ΔV_{th} to reach a preset level [6]–[11]. Under an operation gate bias V_{gop}, the required lifetime is ten years, and the degradation under V_{gop} can be too low to reliably be measured within a practical stress time. To predict the lifetime under V_{gop}, multiple accelerated tests are carried out with stress biases V_{gst} higher than V_{gop}. The accelerated lifetime is typically fitted with $|V_{\mathrm{gst}}|^{-\alpha}$ [6], [7] or $\exp(-|V_{\mathrm{gst}}|)$ [8], [9], and then extrapolated to V_{gop}, so that the lifetime under V_{gop} can be estimated. Recent works [2]–[5], however, have raised several questions on the validation of the preceding prediction technique, as described here.

Manuscript received June 24, 2009; revised September 16, 2009. Current version published December 23, 2009. The review of this paper was arranged by Editor J. Suehle.

Z. Ji, L. Lin, and J. F. Zhang are with the School of Engineering, Liverpool John Moores University, L3 3AF Liverpool, U.K. (e-mail: J.F.Zhang@ljmu.ac.uk).

B. Kaczer is with the Inter-university MicroElectronics Centre (IMEC), 3001 Leuven, Belgium.

G. Groeseneken is with the IMEC, 3001 Leuven, Belgium, and also with Katholieke Universiteit Leuven, 3001 Leuven, Belgium.

Digital Object Identifier 10.1109/TED.2009.2037171

First, the aforementioned prediction technique was developed based on quasi-dc measurements with a measurement time for transfer characteristics I_d–V_g up to seconds [7]–[9], [12]. Recent works [2]–[5] have shown that substantial recovery occurs during the measurement, and to suppress the recovery, the measurement time must be reduced to the order of tens of microseconds by using the ultrafast pulse (UFP) technique [2], [4], [5], [10]. In a real circuit, different transistors will experience different levels of recovery, and the worst case scenario will be no recovery. It is not known whether the V_g acceleration technique previously described is still applicable in this case.

Second, the threshold voltage is typically evaluated by extrapolating the transfer characteristic I_d–V_g, and the sensing V_g used here is close to V_{th}. In a practical circuit, the operation bias is higher than V_{th}, and an implicit assumption of early works is that ΔV_{th} is insensitive to the sensing V_g. Recent work [4], [10], however, shows that ΔV_{th} increases with the sensing V_g, and its value at V_{gop} can double the level extracted by extrapolating I_d–V_g. There is no information on how the lifetime can be predicted by including this dependence on the sensing V_g.

Third, not only recovery but also degradation can occur during the measurement [4], [5], [10]. In an NBTI test, the first measured point is used as the reference value [3], [13], and if the reference is degraded, every subsequent data point will be affected. For some "on-the-fly" technique recently proposed [3], [13], the sensing V_g is the stress bias, and it is reported that substantial degradation can occur for the first point when measured by a typical quasi-dc parameter analyzer [4], [10]. The lifetime cannot reliably be predicted based on a degraded reference.

In this paper, we will show that, when the recovery is suppressed, ΔV_{th} does not follow a simple power law against stress time, and the kinetics at different V_{gst}'s are not in parallel, making the conventional V_g acceleration technique inapplicable here. To predict the worst case lifetime, a physics-based NBTI kinetics at V_{gop} is proposed, which includes contributions from both as-grown and generated defects. An effort will be made to estimate the safety margin of this new lifetime prediction method.

II. Devices and Experiments

A. Devices

To test the applicability of the proposed NBTI kinetics and lifetime prediction method, samples from six different

0018-9383/$26.00 © 2009 IEEE

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

TABLE I
WAFERS AND THE FITTED PARAMETERS AT 125 °C

Device number	Gate Dielectrics	A (mV/sec)	n	c (mV)	t* (μs)
A	1.85nm 12s Plasma SiON	0.23	0.36	11.91	30
B	1.4nm Plasma SiON	7.70	0.13	18.39	4390
C	2.7nm Thermal SiON	26.27	0.07	22.07	80
D	2.0nm 45s Plasma SiON	4.10	0.12	7.49	740
E	2.0nm 20s Plasma SiON	3.91	0.12	1.89	110
F	TiN, ALCVD 2.0 nm / 1 nm HfSiON / SiON	12.21	0.14	40.86	30

processes were used, as shown in Table I. The SiON samples have p+ poly-si gate; four of them were plasma nitrided for different times, and one was thermally nitrided. One HfSiON/SiON stack was also prepared by atomic layer chemical vapor deposition (ALCVD) with 80% Hf and a TiN gate.

B. Experiments and Measurements

The typical NBTI test procedure is to stress pMOSFETs under a constant $V_{\rm gst}$ and temperature for a prespecified time, and the degradation was then monitored by interrupting the stress and measuring I_d–V_g [2], [12], [13]. This "stress-then-sense" procedure is also used for this paper. Since the objective is to study the NBTI lifetime prediction and kinetics with recovery suppressed, the UFP I_d–V_g is used. Recent works [2], [4], [10] have shown that recovery during measurement can effectively be suppressed to within the measurement resolution when the pulse edge time was reduced to the order of tens of microseconds. Both the pulse rise and fall times are 5 μs in this work. The majority of tests were carried out at 125 °C, with a typical test time ranging from microseconds to 75 h.

We now address the degradation in the reference value. In early works, the reference value for an NBTI test was typically obtained by using a quasi-dc parameter analyzer, and the time for measuring one point t_r was 10–150 ms [4], [10]. To show that this quasi-dc reference is degraded, Fig. 1 compares the quasi-dc I_d–V_g with the UFP I_d–V_g when the edge time is 5 μs. When the sensing V_g is low and around $V_{\rm th} \approx -0.3 \sim -0.5$ V, I_d degrades little, and the quasi-dc value can be used as the reference. At $|V_{\rm gop}| = 1.2$ V, however, the degradation of the quasi-dc reference becomes observable. Moreover, some on-the-fly technique [3], [13] uses stress bias $|V_{\rm gst}| > |V_{\rm gop}|$ as the sensing V_g, and Fig. 1 shows that the quasi-dc reference further degrades for higher $|V_g|$.

To show that the I_d at $V_{\rm gop} = -1.2$ V degrades little when measured with a pulse edge of 5 μs, Fig. 2(a) gives the I_d, which is repeatedly measured 30 times. Degradation was not observed, and a measurement resolution of $\pm 0.23\%$ was achieved.

In this paper, we intend to stress the device at $|V_{\rm gop}| = 1.2$ V, which is lower than the typical stress bias used in early NBTI tests [2]–[5], [10]. For low degradation, noises must be minimized. The typical commercial parameter analyzers minimize noises by repeating I_d measurement for a given bias many times and then use their average value. For a stress time of less than 1 s, we also repeated the pulse measurement many times (i.e., 30) and used their average value, as shown in Fig. 2(b). For stress longer than 1 s, we used a quasi-dc parameter analyzer that automatically uses average values. As will be shown in Sections III-C and III-D, the measurement accuracy achieved in this way is good enough to establish a kinetic model that allows lifetime to be predicted with a safety margin of 50%.

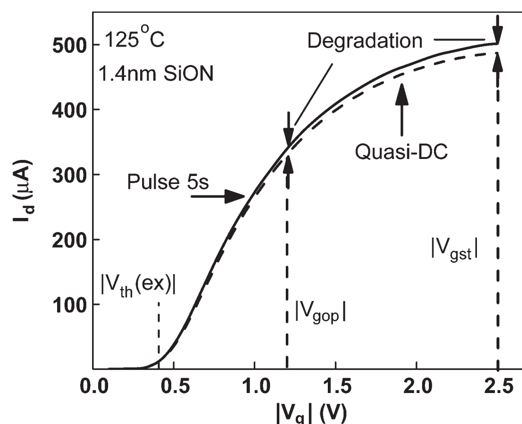

Fig. 1. Comparison of I_d–V_g measured by UFP and quasi-dc techniques on fresh devices. The edge time of the pulse is 5 μs, and it took 150 ms for recording one point for the quasi-dc technique. Although the degradation is negligible near the threshold voltage $V_{\rm th}$(ex), it increases and becomes observable at $|V_{\rm gop}| = -1.2$ V for the quasi-dc measurement. Therefore, the quasi-dc I_d at $|V_{\rm gop}| = -1.2$ V should not be used as the reference value.

C. $\Delta V_{\rm th}$ Extraction at the Operation Gate Bias

Traditionally, NBTI-induced $\Delta V_{\rm th}$ was evaluated by extrapolating I_d–V_g, as shown in the inset of Fig. 3(a), and the assumption is that $\Delta V_{\rm th}$ is insensitive to the sensing V_g. Our recent work [4], [10], however, shows that $|\Delta V_{\rm th}|$ substantially increases for higher sensing $|V_g|$. Since measuring $\Delta V_{\rm th}$ at different sensing V_g's is rarely reported in the past, a description is given here.

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

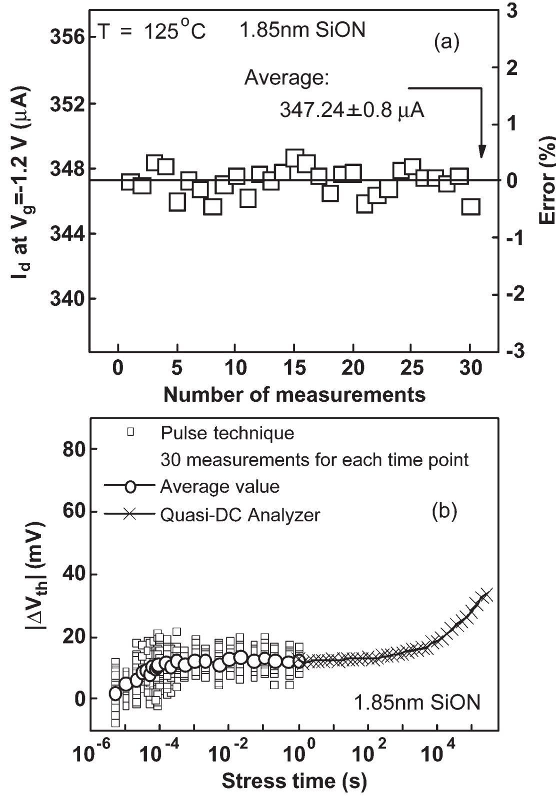

Fig. 2. Measurement resolution and minimization of noises. (a) I_d at $V_g = -1.2$ V was measured 30 times from the pulsed I_d–V_g with an edge time of 5 μs. A measurement resolution of $\pm 0.23\%$ was achieved, and I_d degradation during the measurement cannot be observed within this resolution. (b) The pulse measurement (symbol '□') was repeated 30 times for a stress time of less than 1 s, and the average value (symbol '○') was used to minimize the noise. For stress longer than 1 s, we used a quasi-dc parameter analyzer that automatically used the average value (symbol '×').

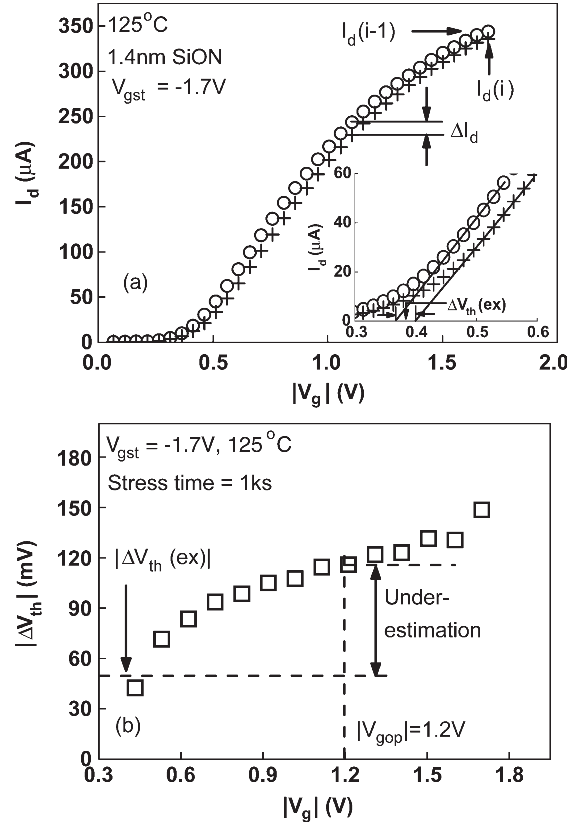

Fig. 3. Evaluation of the threshold voltage shift at different sensing V_g's. (a) I_d–V_g at two stress time points, i.e., $(i-1)$ for 1 s and i for 1000 s, which were measured by the UFP technique with an edge time of 5 μs. ΔI_d represents the NBTI-induced current degradation between these two time points at a given sensing V_g. The inset shows the evaluation of the threshold voltage shift by extrapolating I_d–V_g. (b) $|\Delta V_{\mathrm{th}}|$ increases with the sensing $|V_g|$, so that $|\Delta V_{\mathrm{th}}|$ at $|V_{\mathrm{gop}}| = 1.2$ V is substantially higher than the $|\Delta V_{\mathrm{th}}(\mathrm{ex})|$ obtained by extrapolation.

Fig. 3(a) shows that the drain current degradation ΔI_d at a given sensing V_g can be measured, and the ΔV_{th} can be evaluated by

$$\Delta V_{\mathrm{th}} \approx \sum_{i=1}^{N} \Delta V_{\mathrm{th}}(i) = -\sum_{i=1}^{N} \frac{\Delta I_d(i)}{\left[G_m(i) + G_m(i-1)\right]/2} \quad (1)$$

where i and $(i-1)$ represents two measurement points, and $\Delta V_{\mathrm{th}}(i)$ is the incremental degradation between these two points. G_m is the transconductance and is evaluated from I_d–V_g by

$$G_m = \frac{dI_d}{d(V_g - V_{\mathrm{th}})}. \quad (2)$$

$d(V_g - V_{\mathrm{th}})$ is used in the denominator of the preceding formula since I_d is driven by $(V_g - V_{\mathrm{th}})$ and V_{th} varies with V_g here [4], [10]. A typical result is given in Fig. 3(b), and $|\Delta V_{\mathrm{th}}|$ at the operation $V_{\mathrm{gop}} = -1.2$ V is much higher than $|\Delta V_{\mathrm{th}}|$ obtained by extrapolating I_d–V_g. Since the objective of this work is to investigate the worst case NBTI lifetime and kinetics, we use ΔV_{th} at V_{gop} hereinafter, unless otherwise specified.

Fig. 2(b) shows that $|\Delta V_{\mathrm{th}}|$ becomes higher than 10 mV in less than 1 ms under $V_g = -1.2$ V. This rapid increase in $|\Delta V_{\mathrm{th}}|$ does not mean that the quality of the samples used here is poor [10]. It is caused partially by suppressing recovery with the pulse measurement and partially by using $|V_g| = 1.2$ V as the sensing bias, rather than the conventional extrapolation to $|V_{\mathrm{th}}| \sim 0.4$ V, as shown in our early works [4], [10].

III. Results and Discussion

A. Inapplicability of the V_g Acceleration Prediction Technique

Fig. 4(a)–(c) shows the common procedure for predicting lifetime: Multiple V_g accelerated tests were carried out, and the lifetime τ at V_{gop} was estimated by an extrapolation against V_{gst}. V_g acceleration is used to allow obtaining a reliable NBTI kinetics within a practical length of stress time. The safety margin of the prediction, however, is generally not known. To assess the prediction safety margin, it is essential to be able to directly measure ΔV_{th} at $V_{\mathrm{gst}} = V_{\mathrm{gop}}$. As V_{gop} does not proportionally reduce with the SiON thickness, the oxide field during device operation has increased to such a level that Fig. 4(a) shows that the ΔV_{th} at $V_{\mathrm{gop}} = -1.2$ V can now reliably be measured. This allows comparing the measured stress time for ΔV_{th} to reach a given level under $V_{\mathrm{gst}} = V_{\mathrm{gop}}$ with that predicted by using V_g acceleration, so that the safety margin of prediction can be estimated.

In Fig. 4(a), the last measured $|\Delta V_{\mathrm{th}}|$ reached 60 mV under $V_{\mathrm{gst}} = V_{\mathrm{gop}}$, and if we use $|\Delta V_{\mathrm{th}}| = 60$ mV to define lifetime,

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

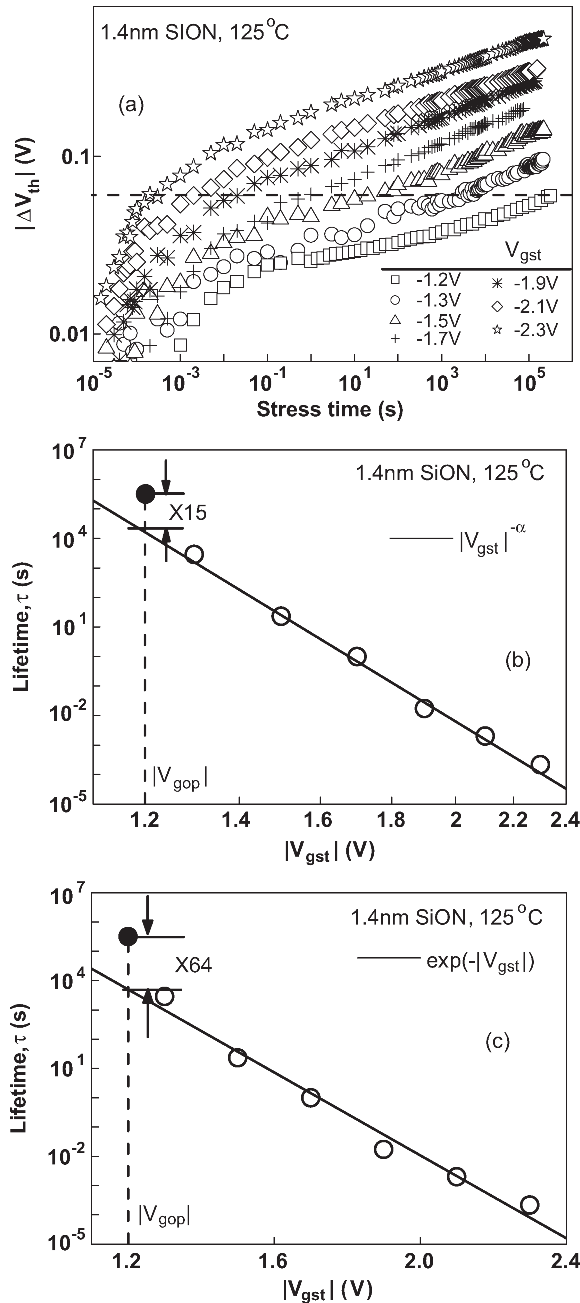

Fig. 4. Inapplicability of the V_g acceleration technique for predicting NBTI lifetime at $V_{\mathrm{gop}} = -1.2$ V measured by the UFP technique. (a) Dynamic behavior of $|\Delta V_{\mathrm{th}}|$ under different stress biases V_{gst} is compared, and the kinetics does not follow a simple power law. The horizontal dashed line represents $|\Delta V_{\mathrm{th}}| = 60$ mV that is used to define the lifetime. (b) and (c) Measured lifetime is compared with the predicted one based on $|V_{\mathrm{gst}}|^{-\alpha}$ and $\exp(-|V_{\mathrm{gst}}|)$, respectively. The symbol '•' represents the measured lifetime under $|V_{\mathrm{gst}}| = 1.2$ V. The solid line was obtained by fitting the data at higher $|V_{\mathrm{gst}}|$. The error in prediction is substantial.

the stress time for this last point will be the lifetime under $V_{\mathrm{gop}} = -1.2$ V. This measured lifetime is compared to the prediction based on the V_g acceleration in Fig. 4(b) and (c). Two popular V_g acceleration methods have been used in the literature: $|V_{\mathrm{gst}}|^{-\alpha}$ [6], [7] and $\exp(-|V_{\mathrm{gst}}|)$ [8], [9]. Fig. 4(b) and (c) shows that there is a substantial difference between the predicted and measured time, and therefore, V_g acceleration cannot be used to predict lifetime in our case.

To analyze the reason for the inapplicability of the V_g acceleration technique, we will use the $|V_{\mathrm{gst}}|^{-\alpha}$ acceleration as an example. The $|V_{\mathrm{gst}}|^{-\alpha}$ acceleration requires

$$|\Delta V_{\mathrm{th}}| = B|V_{\mathrm{gst}}|^m t^n \tag{3}$$

where B is a constant for a given temperature. At a device lifetime of $t = \tau$, ΔV_{th} reaches the specified $\Delta V_{\mathrm{th}}(\tau)$, and we have

$$\log(\tau) = \frac{1}{n}\log\left[\frac{|\Delta V_{\mathrm{th}}(\tau)|}{B}\right] - \frac{m}{n}\log\left(|V_{\mathrm{gst}}|\right). \tag{4}$$

Equations (3) and (4) require the $\log|\Delta V_{\mathrm{th}}| \sim \log(t)$ being shifted in parallel for different V_{gst}'s and the power factor against time n, being insensitive to V_{gst}, so that $\log(\tau)$ is a straight line against $\log(|V_{\mathrm{gst}}|)$. For the conventional ΔV_{th} measured by extrapolating the quasi-dc I_d–V_g, Fig. 5(a) and (b) shows that these requirements can be met, so that the prediction agrees well with measurement in Fig. 5(c). Once the recovery is suppressed and $V_{\mathrm{gop}} = -1.2$ V is used as the sensing V_g, however, ΔV_{th} in Fig. 4(a) no longer follows a simple power law, and $\log|\Delta V_{\mathrm{th}}| \sim \log(t)$ at different V_{gst}'s is generally not a parallel shift. A clear example for the nonparallel shift is given in Fig. 6. The V_{gst} effect can no longer be separated into a "prefactor" like that in (3), and this explains the inapplicability of the V_g acceleration technique to the case where recovery is suppressed.

B. NBTI Kinetics at the Operation Gate Bias Based on the UFP Measurement

Since ΔV_{th} does not follow a simple power law against stress time when recovery is suppressed, efforts should be made to develop a model that can describe its dynamic behavior. For the ΔV_{th} under $V_{\mathrm{gst}} = V_{\mathrm{gop}} = -1.2$ V, Fig. 7 shows that an outstanding feature of the kinetics is the presence of a "shoulder." This indicates that there is an initial period when as-grown defects dominate, and the saturation of their charging results in the "shoulder." At longer stress time, generation of new defects becomes increasingly important and is responsible for the rise above the "shoulder."

To support the aforementioned suggestion, two tests were carried out. In the first test, we checked the effect of temperature on the shoulder height. The saturation level of as-grown defects should be insensitive to temperature [14], [15], and if it dominates the shoulder, the shoulder height should be insensitive to temperature. This is confirmed by Fig. 8. Fig. 8 also shows that the rise above the shoulder is thermally activated, supporting that defect generation is thermally accelerated.

In the second test, we compare the charging and discharging rates of the defect responsible for the shoulder. Early works [15]–[18] identified three different types of positive charges in the dielectric, i.e., antineutralization positive charges (ANPCs), cyclic positive charges (CPCs), and as-grown hole traps (AHTs), as illustrated by Fig. 9. ANPC has an energy level above the bottom edge of silicon conduction band Ec, making its discharging more difficult than charging. CPC has an energy level close to Ec, and its charging rate is similar to the

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

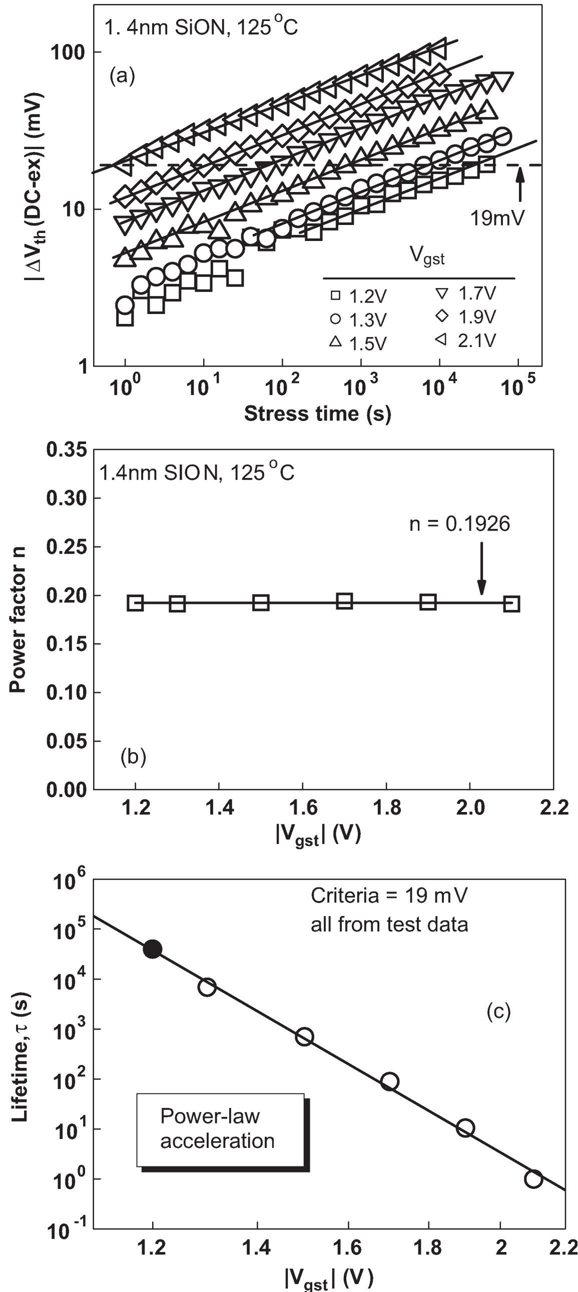

Fig. 5. Lifetime prediction by the V_g acceleration technique for the conventional $\Delta V_{\rm th}$ (dc-ex) measured by extrapolating the quasi-dc I_d–V_g. (a) $\Delta V_{\rm th}$ (dc-ex) follows the power law. The horizontal dashed line represents $|\Delta V_{\rm th}| =$ 19 mV that is used to define the lifetime. (b) Power factor is insensitive to the stress bias. The solid line represents the average power factor of $n = 0.1926$. (c) Prediction (solid line) agrees with the measurement (symbol '•').

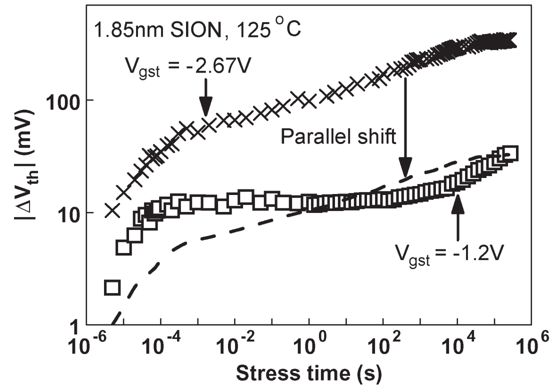

Fig. 6. Kinetics at different $V_{\rm gst}$'s is not shifted in parallel for the UFP $\Delta V_{\rm th}$ sensed at $|V_g| = 1.2$ V. The dashed curve was obtained by shifting the symbol '×' downward in parallel.

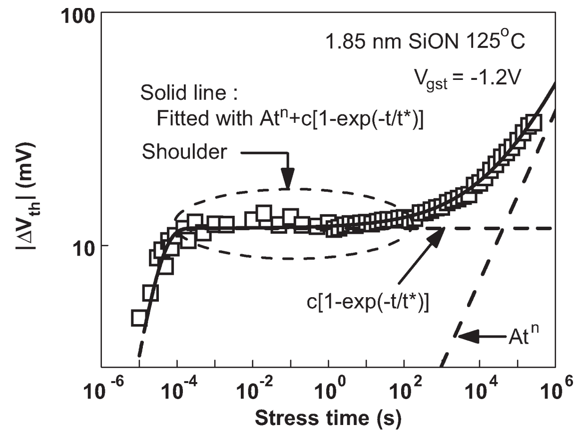

Fig. 7. Kinetic feature of the UFP $\Delta V_{\rm th}$ sensed at $|V_g| = 1.2$ V: a "shoulder." By combining the first-order model for AHTs with the power law for defect generation, $\Delta V_{\rm th}$ can be fitted over ten orders of magnitude in time, as shown by the solid line. The dashed lines show that $\Delta V_{\rm th}$ is initially dominated by AHTs, but the generated defects become important at longer stress time.

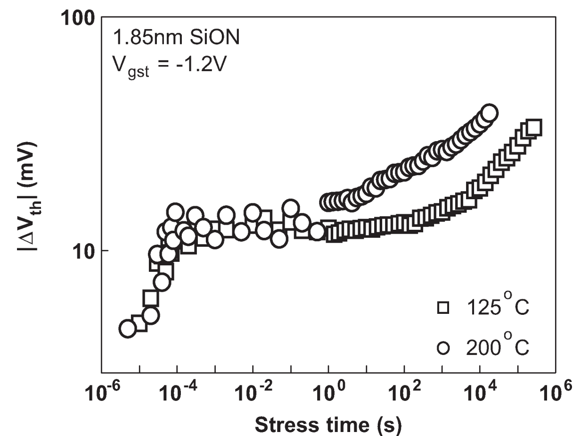

Fig. 8. Effect of temperature on NBTI kinetics. The height of the shoulder is insensitive to temperature, but the generation of defects above the shoulder is thermally accelerated.

discharging rate. In contrast, AHT is below the top edge of the silicon valence band, and there are far more valence electrons for discharging than hot holes required for charging. As a result, AHT has the signature that discharging is much faster than charging. Fig. 10 shows that, when the stress time corresponds to the shoulder, the charging and discharging properties of the defect agree with the signature of AHT, supporting that AHT dominates the shoulder.

We now discuss how hot holes can be generated under a modest bias of $V_{\rm gop} = -1.2$ V. Hot holes can come from three sources: First, the surface quantization effect gives energy subbands [19], [20]. Although the hole density reduces as the energy increases, there are holes in the higher subbands. Second, process "a" in Fig. 11 shows that an electron tunneling from the gate can recombine with a hole in the substrate, and the released energy can create hot holes [21]. Third, it has

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

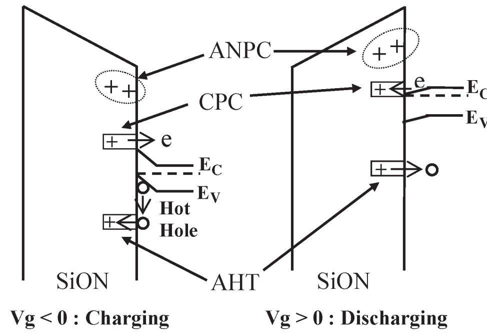

Fig. 9. Energy band diagram of different types of positive charges. The ANPCs have an energy level above the bottom edge of the silicon conduction band Ec, making them difficult to neutralize. The CPCs have an energy level near Ec, resulting in similar charging and discharging rates. The AHTs have an energy level below the top edge of the silicon valence band Ev. Their charging requires hot holes, leading to charging that is slower than discharging.

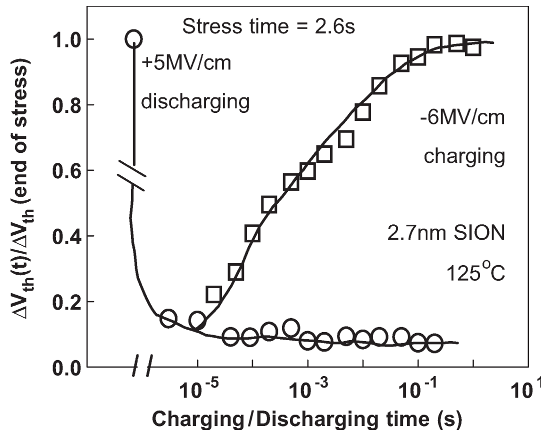

Fig. 10. Comparison of the charging and discharging rates for the as-grown defects. The stress time is 2.6 s, which corresponds to the region where as-grown defects dominate. The discharging under $V_g > 0$ is much faster than the charging under $V_g < 0$, which is a unique signature of AHTs. The solid lines are guides for the eye. It should be noted that the rapid discharging within 5 μs observed here was achieved by applying a positive gate bias. For a normal NBTI test, however, positive gate bias was not applied, and the AHT discharging at $V_g = -1.2$ V within 5 μs was negligible [4], [10].

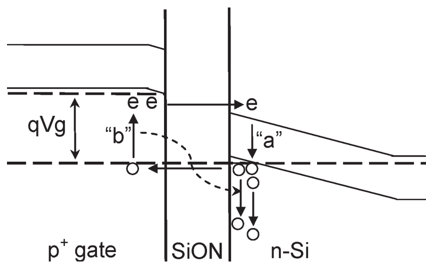

Fig. 11. Schematic showing the physical processes for generating hot holes under $V_g = -1.2$ V. Process "a" shows that the energy released by the electron–hole recombination in the substrate can create hot holes. Process "b" shows that the photons emitted from the gate can be absorbed to generate hot holes.

been proposed that photons can be emitted by electron–hole recombination in the gate [22], as illustrated by process "b" in Fig. 11. Holes can become hot by absorbing the photons.

On the kinetics, the charging of AHT generally follows the first-order reaction model [23], [24], whereas the generation of

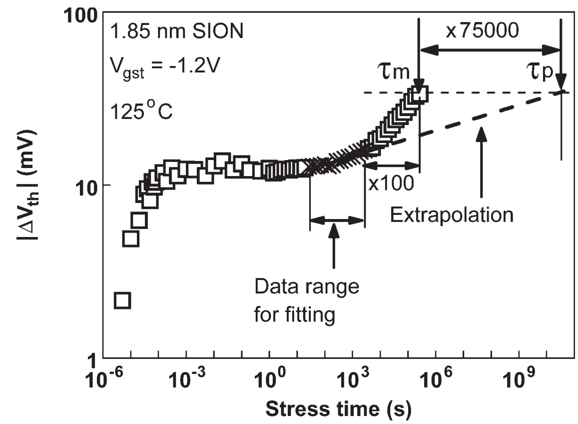

Fig. 12. Lifetime prediction based on the method proposed in [5]. All symbols are test data, but only symbol "×" was used for fitting with a power law in the range of 26.8 s $< t <$ 2680 s. The thick dashed line is extrapolated from the fitted line for prediction. τ_m is the time for the last test point, and τ_p is the predicted time. τ is overestimated by a factor of 75 000.

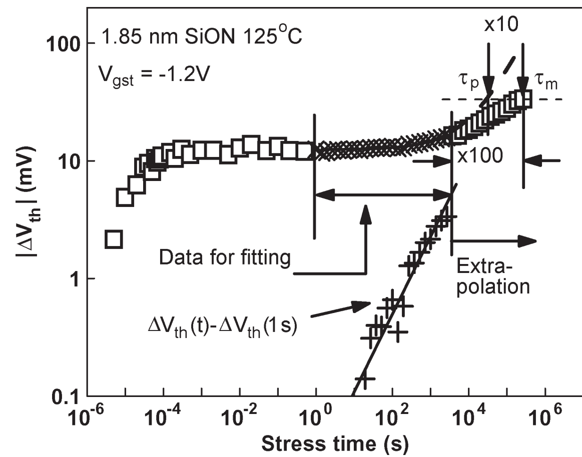

Fig. 13. Lifetime prediction based on the method proposed in [27]. All symbols are test data, but only symbol "×" was used for fitting. $\Delta V_{\rm th}(t) - \Delta V_{\rm th}$ (1 s) (symbol "+") was fitted with a power law, and $\Delta V_{\rm th}$(1 s) was then added back. The dashed line is extrapolated for prediction. τ_m is the time for the last test point, and τ_p is the predicted time. This method underestimates τ by a factor of 10.

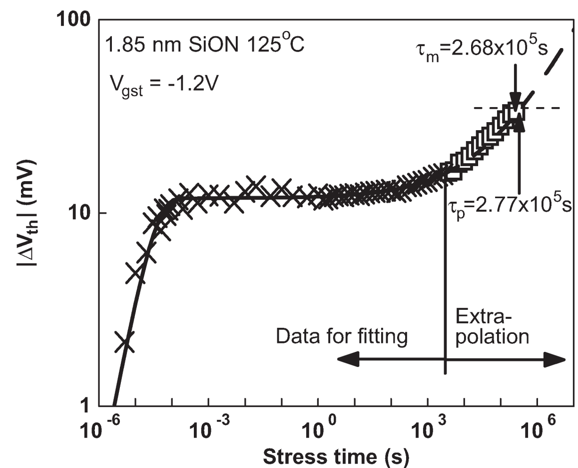

Fig. 14. Lifetime prediction based on our model: (5). The test data are the same as those in Figs. 12 and 13, and only symbol "×" was used for fitting. Our model substantially reduces the prediction error, and $\tau_p/\tau_m = 1.03$. The test sample is a plasma-nitrided 1.85-nm SiON.

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

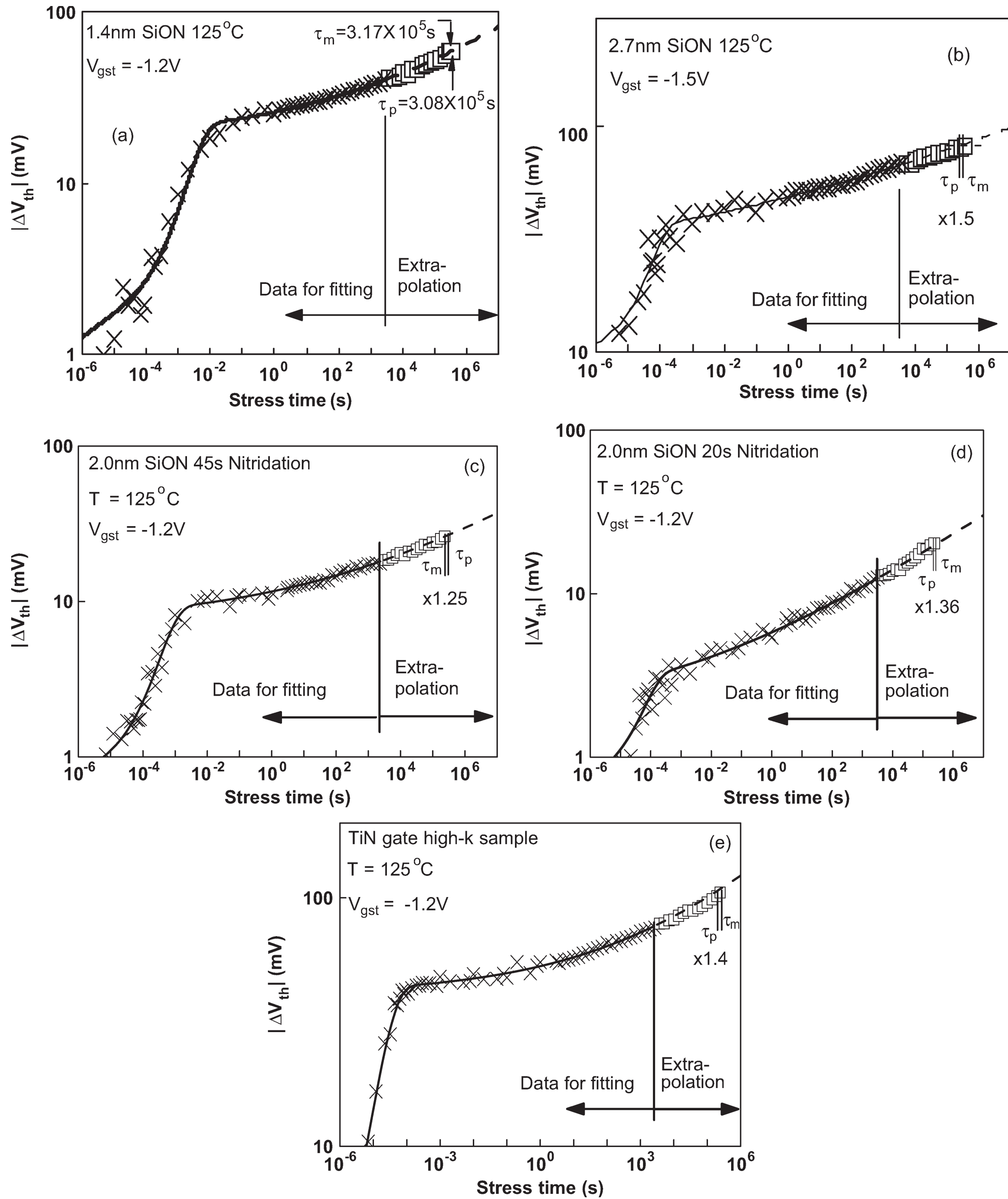

Fig. 15. Applicability of the single-test lifetime prediction technique for different fabrication processes. (a) 1.4-nm plasma SiON. (b) 2.7-nm thermal SiON. (c) 2.0-nm 45-s plasma SiON. (d) 2.0-nm 20-s plasma SiON. (e) 2.0-nm/1.0-nm HfSiON/SiON stack prepared by ALCVD with TiN gate. The safety margin for the prediction is within 50% in all cases.

new defects follows a power law [5], [6], [25]. By combining these two, we have

$$\Delta V_{\rm th} = At^n + c(1 - e^{-t/t^*}). \qquad (5)$$

For a given stress temperature and bias, A, n, c, and t^* are constants and were obtained by fitting test data with the least-square errors, and their values are given in Table I. There is only one n for the whole stress period, and this n is not the slope of the data in Fig. 7, i.e., $n \neq d(\log|\Delta V_{\rm th}|)/d[\log(t)]$. Fig. 7 shows that (5) can model the "shoulder," and this simple physics-based model can fit the $\Delta V_{\rm th}$ over ten orders of stress time. The two dashed lines represent the contribution from AHT and generated defects, respectively. AHT clearly dominates initially, but generated defects become more important for longer stress time. On the nature of degraded defects, our early works [12], [13], [15]–[18], [23], [24] show that both interface states and new hole traps are created by stresses. The new hole traps are further separated into ANPCs and CPCs, each with

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

unique signatures. A detailed discussion, however, is out of the scope of this work.

C. Lifetime Prediction Based on a Single Test

The V_g acceleration technique was developed for predicting the lifetime caused by time-dependent dielectric breakdown (TDDB) [26]. For TDDB, intrinsic dielectric breakdown generally does not occur within an affordable test time under the operation bias, and multiple V_g accelerated tests are essential. For NBTI with a sensing $|V_g| = 1.2$ V based on UFP measurement, we have shown that the prediction based on V_g acceleration is inapplicable, and an obvious question is how to assess lifetime in this case. One possibility is to use thermal acceleration. Fig. 8, however, shows that $\log|\Delta V_{\rm th}|$ at different temperatures is not shifted in parallel, so that it does not offer a reliable prediction method. Since $\Delta V_{\rm th}$ can be measured at $V_{\rm gop}$ (Fig. 7), in principle, lifetime can be estimated by extrapolating the $\Delta V_{\rm th}$ against stress time based on a single test at $V_{\rm gop}$. This requires a reliable kinetic model that can not only fit the existing test data but also predict the future.

The affordable test time is typically on the order of days, and the data were used to predict lifetime in years, so that a kinetic model should have the ability to predict at least two decades ahead. To test the prediction ability of a model, the test data in the last two decades are not used for fitting the model, the $\Delta V_{\rm th}$ at the last test point is considered as $\Delta V_{\rm th}(\tau)$, and the time for the last point is treated as the measured lifetime τ_m, although τ_m is not the real device lifetime. The departure of kinetics from a simple power law was noted in the past, and suggestions were made on how the lifetime prediction method should be modified to take this departure into account [5], [27]. One proposed method is only using the data with a stress time of more than 10 s to fit the power law against time [5], but there is no information on the prediction accuracy. Fig. 12 shows that this method can overestimate τ by a factor of 75 000. Another proposed method is to fit $\Delta V_{\rm th}(t) - \Delta V_{\rm th}$ (1 s) with a power law [27], but Fig. 13 shows that it underestimates τ by a factor of 10. By applying our model of (5) to the same set of data, Fig. 14 shows that good agreement is achieved between the measurement and the prediction with $\tau_p/\tau_m = 1.03$.

D. Applicability of the Single-Test Prediction Method for Different Processes

Although a good prediction is achieved in Fig. 14, it is inadequate to demonstrate that a prediction method works for one process. For a prediction technique to be useful, it must be applicable to samples fabricated by a wide range of processes. We now test its applicability for four other SiON layers with different nitrogen concentrations and nitrided either by plasma or thermally. Moreover, an ALCVD HfSiON/SiON stack is also tested, and the thickness of these samples is given in Table I. Fig. 15(a)–(e) shows that $\Delta V_{\rm th}$ follows (5) in all cases, although the "shoulder" in some samples is less apparent. The prediction achieved a safety margin of 50% or less in all the processes tested, making us confident that the single-test technique is generally applicable.

Table I shows that the fitted power factors for different processes have a range of 0.07–0.36, which agrees with the range reported by early works [28]–[34]. Early works reported that the variation of power factor could come from two sources, i.e., different hydrogenous species and different nitrogen densities and distributions. On the hydrogenous species, it was reported that the power factor for H^+ [32], atomic hydrogen [33], and H_2 [34] is 0.5, 0.25, and 0.16, respectively. On the nitridation effect, it was reported that an increase in nitrogen reduces the power factor [28], [29]. Moreover, for the same area density of nitrogen, the power factor of a thermally nitrided SiON is typically lower than that of a plasma-nitrided SiON [28]. The power factor in Table I agrees with these trends. Sample A has the lowest nitrogen density and the highest power factor. Sample C was thermally nitrided and has the lowest power factor.

IV. Conclusion

This work has investigated the NBTI lifetime prediction in the worst case scenario where the recovery is suppressed and $\Delta V_{\rm th}$ is sensed at the operation gate bias. In this case, the conventional V_g acceleration prediction is inapplicable, because the NBTI kinetics no longer follow a simple power law, and an increase in stress bias does not lead to a parallel shift of $\log|\Delta V_{\rm th}|$.

To predict the lifetime at the operation gate bias based on the UFP measurement, NBTI kinetics and defects are examined. An outstanding feature of the kinetics is the presence of a "shoulder," which is insensitive to temperature and must be dominated by the charging of as-grown defects. The charging and discharging properties of the defect agree well with the signature of AHTs. By combining the first-order model for the AHTs and the power law for generating new defects, $\Delta V_{\rm th}$ can be modeled over ten orders of stress time. This kinetic model is then used to predict the NBTI lifetime, based on a single test at the operation temperature and bias. For the six different processes tested, the safety margin of the single-test prediction technique is within 50%, which is substantially better than the methods proposed in early works.

References

[1] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, no. 1, pp. 1–17, Jul. 2003.

[2] T. Yang, M. F. Li, C. Shen, C. H. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu, and D. L. Kwong, "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.

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

[4] J. F. Zhang, Z. Ji, M. H. Chang, B. Kaczer, and G. Groeseneken, "Real $V_{\rm th}$ instability of pMOSFETs under practical operation conditions," in *IEDM Tech. Dig.*, 2007, pp. 817–820.

[5] E. N. Kumar, V. D. Maheta, S. Purawat, A. E. Eslam, C. Olsen, K. Ahmed, M. A. 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) $I_{\rm DLIN}$ technique," in *IEDM Tech. Dig.*, 2007, pp. 809–812.

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

[6] A. Aoulaiche, B. Kaczer, D. Jacger, M. Houssa, K. Martens, R. Degraeve, P. Roussel, J. Mitard, S. De Gendt, H. E. Maes, G. Groeseneken, M. Meuris, and M. M. Heyns, "Negative bias temperature instability on Si-passivated Ge-interface," in *Proc. Int. Rel. Phys. Symp.*, 2008, pp. 262–264.

[7] C. Schlunder, W. Heinrigs, W. Gustin, and H. Reisinger, "On the impact of the NBTI recovery phenomenon on lifetime prediction of modern p-MOSFETs," in *Proc. IRW Final Report*, 2006, pp. 1–4.

[8] 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.

[9] H. Puchner and L. Hinh, "NBTI reliability analysis for a 90 nm CMOS technology," in *Proc. ESSDERC*, 2004, pp. 257–260.

[10] Z. Ji, J. F. Zhang, M. H. Chang, B. Kaczer, and G. Groeseneken, "An analysis of the NBTI-induced threshold voltage shift evaluated by different techniques," *IEEE Trans. Electron Devices*, vol. 56, no. 5, pp. 1086–1093, May 2009.

[11] M. Rafik, X. Garros, G. Ribes, G. Ghibaudo, C. Hobbs, A. Zauner, M. Muller, V. Huard, and C. Ouvrard, "Impact of TiN metal gate on NBTI assessed by interface states and fast transient effect characterization," in *IEDM Tech. Dig.*, 2007, pp. 825–828.

[12] J. F. Zhang, M. H. Chang, and G. Groeseneken, "Effects of measurement temperature on NBTI," *IEEE Electron Device Lett.*, vol. 28, no. 4, pp. 298–300, Apr. 2007.

[13] M. H. Chang and J. F. Zhang, "On positive charges formed under negative bias temperature stresses," *J. Appl. Phys.*, vol. 101, no. 2, p. 024 516, Jan. 2007.

[14] G. Van den bosch, G. Groeseneken, H. E. Maes, R. B. Klein, and N. S. Saks, "Oxide and interface degradation resulting from substrate hot-hole injection in metal–oxide–semiconductor field-effect transistors at 296 and 77 K," *J. Appl. Phys.*, vol. 75, no. 4, pp. 2073–2080, Feb. 1994.

[15] C. Z. Zhao, J. F. Zhang, M. H. Chang, A. R. Peaker, S. Hall, G. Groeseneken, L. Pantisano, S. De Gendt, and M. Heyns, "Stress-induced positive charge in Hf-based gate dielectrics: Impact on device performance and a framework for the defect," *IEEE Trans. Electron Devices*, vol. 55, no. 7, pp. 1647–1656, Jul. 2008.

[16] J. F. Zhang, C. Z. Zhao, A. H. Chen, G. Groeseneken, and R. Degraeve, "Hole traps in silicon dioxides—Part I: Properties," *IEEE Trans. Electron Devices*, vol. 51, no. 8, pp. 1267–1273, Aug. 2004.

[17] C. Z. Zhao, J. F. Zhang, G. Groeseneken, and R. Degraeve, "Hole traps in silicon dioxides—Part II: Generation mechanism," *IEEE Trans. Electron Devices*, vol. 51, no. 8, pp. 1274–1280, Aug. 2004.

[18] C. Z. Zhao and J. F. Zhang, "Effects of hydrogen on positive charges in gate oxides," *J. Appl. Phys.*, vol. 97, no. 7, p. 073 703, Apr. 2005.

[19] Y. T. Hou, M. F. Li, Y. Jin, and W. H. Lai, "Direct tunneling hole currents through ultrathin gate oxides in metal–oxide–semiconductor devices," *J. Appl. Phys.*, vol. 91, no. 1, pp. 258–264, Jan. 2002.

[20] C. Z. Zhao, J. F. Zhang, M. B. Zahid, B. Govoreanu, G. Groeseneken, and S. De Gendt, "Determination of capture cross sections for as-grown electron traps in $HfO_2/HfSiO$ stacks," *J. Appl. Phys.*, vol. 100, no. 9, p. 093 716, Nov. 2006.

[21] J. D. Bude, B. E. Weir, and P. J. Silverman, "Explanation of stress-induced damage in thin oxides," in *IEDM Tech. Dig.*, 1998, pp. 179–182.

[22] M. Xu, C. Tan, and M. F. Li, "Extended Arrhenius law of time-to-breakdown of ultrathin gate oxides," *Appl. Phys. Lett.*, vol. 82, no. 15, pp. 2482–2484, Apr. 2003.

[23] J. F. Zhang, H. K. Sii, G. Groeseneken, and R. Degraeve, "Hole trapping and trap generation in the gate silicon dioxide," *IEEE Trans. Electron Devices*, vol. 48, no. 6, pp. 1127–1135, Jun. 2001.

[24] J. F. Zhang, H. K. Sii, A. H. Chen, C. Z. Zhao, M. J. Uren, G. Groeseneken, and R. Degraeve, "Hole trap generation in gate dielectric during substrate hole injection," *Semicond. Sci. Technol.*, vol. 19, no. 1, pp. L1–L3, Jan. 2004.

[25] S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang, and L. Chen, "Nitrogen-enhanced negative bias temperature instability: An insight by experiment and first-principle calculations," *Appl. Phys. Lett.*, vol. 82, no. 12, pp. 1881–1883, Mar. 2003.

[26] E. Y. Wu, J. Aitken, E. Nowak, A. Vayshenker, P. Varekamp, G. Hueckel, J. McKenna, D. Harmon, L.-K. Han, C. Montrose, R. Dufresne, and R.-P. Vollertsen, "Voltage-dependent Voltage-acceleration of oxide breakdown for ultra-thin oxides," in *IEDM Tech. Dig.*, 2000, pp. 541–544.

[27] A. Neugroschel, G. Bersuker, R. Choi, and B. H. Lee, "Effect of the interfacial SiO_2 layer in high-k HfO_2 gate stacks on NBTI," *IEEE Trans. Device Mater. Rel.*, vol. 8, no. 1, pp. 47–61, Mar. 2008.

[28] V. D. Maheta, C. Olsen, K. Ahmed, and S. Mahapatra, "The impact of nitrogen engineering in silicon oxynitride gate dielectric on NBTI of pMOSFETs: A study by ultrafast On-The-Fly Idlin technique," *IEEE Trans. Electron Devices*, vol. 55, no. 7, pp. 1630–1638, Jul. 2008.

[29] 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. 691–694.

[30] Y. F. Chen, M. H. Lin, C. H. Chou, W. C. Chang, S. C. Huang, Y. J. Chang, K. Y. Fu, M. T. Lee, C. H. Liu, and S. K. Fan, "Negative bias temperature instability (NBTI) in deep sub-micron p-gate pMOSFETs," in *Proc. IIRW*, 2000, pp. 98–101.

[31] 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.

[32] 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. IRPS*, 2004, pp. 273–282.

[33] S. Ogawa and N. Shiono, "Generalized diffusion-reaction model for the low-field charge-buildup instability at the $Si\text{-}SiO_2$ interface," *Phys. Rev. B, Condens. Matter*, vol. 51, no. 7, pp. 4218–4230, Feb. 1995.

[34] M. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, "A comprehensive model for PMOS NBTI degradation: Recent progress," *Microelectron. Reliab.*, vol. 47, no. 6, pp. 853–862, Jun. 2007.

Zhigang Ji received the B.Eng. degree in electric engineering from Tsinghua University, Beijing, China, in 2003 and the M.Eng. degree in microelectronics from Peking University, Beijing, in 2006. He is currently working toward the Ph.D. degree in the Reliability Laboratory, School of Engineering, Liverpool John Moores University, Liverpool, U.K.

He is focusing on semiconductor device reliability issues, particularly negative bias temperature instability and the new measurement techniques for characterizing nano-CMOS devices.

L. Lin received the B.Sc. degree in electronic engineering from the University of Central Lancashire, Preston, U.K., in 2001 and the M.Sc. degree in microelectronics system design from Liverpool John Moores University, Liverpool, U.K., in 2003. He is currently working toward the Ph.D. degree in the Reliability Laboratory, School of Engineering, Liverpool John Moores University.

He is focusing on semiconductor device reliability issues, particularly negative bias temperature instability and new measurement techniques for characterizing nano-CMOS devices.

Jian Fu Zhang received the B.Eng. degree in electrical engineering from Xi'an Jiao Tong University, Xi'an, China, in 1982 and the Ph.D. degree in electrical engineering from the University of Liverpool, Liverpool, U.K., in 1987.

From 1986 to 1992, he was a Senior Research Assistant with the University of Liverpool, where he worked on the dielectric recovery of plasma in accelerating gas flow, plasma processing of semiconductors, and reliability of MOS devices. In 1992, he joined the School of Engineering, Liverpool John Moores University, Liverpool, U.K., as a Senior Lecturer. He became a Reader in Microelectronics in 1996 and a Professor in 2001. He has authored or coauthored more than 100 journal and conference proceeding papers, including ten invited papers at several international conferences. His current research interests include the performance, degradation, and defect characterization of MOSFETs and nonvolatile memories.

Dr. Zhang was a member of the technical program committee for the IEEE Semiconductor Interface Specialists Conference (SISC).

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.

Ben Kaczer received the M.S. degree in physical electronics from Charles University, Prague, Czech Republic, in 1992 and the M.S. and Ph.D. degrees in physics from The Ohio State University, Columbus, in 1996 and 1998, respectively.

In 1998, he joined the Reliability Group of the Inter-university MicroElectronics Centre (IMEC), Leuven, Belgium, where his activities have included research on the degradation phenomena and reliability assessment of SiO_2, SiON, high-k, and ferroelectric films; planar and multiple-gate FETs and circuits; and characterization of Ge/III-V and MIM devices. He has authored or coauthored more than 200 journal and conference proceeding papers, and presented eight invited presentations and three International Reliability Physics Symposium (IRPS) tutorials.

Dr. Kaczer has served or is serving in various functions at the IEEE International Electron Device Meeting, IRPS, International Integrated Reliability Workshop, IEEE Semiconductor Interface Specialists Conference, and Insulating Films on Semiconductor conference. He received the OSU Presidential Fellowship and support from Texas Instruments, Inc., Dallas, for his Ph.D. research on the ballistic-electron emission microscopy of SiO_2 and SiC films. He was the recipient of the Best and the Outstanding Paper Awards at IRPS and the Best Paper Award at the International Symposium on the Physical and Failure Analysis of Integrated Circuits.

Guido Groeseneken received the M.Sc. degree in electrical and mechanical engineering and the Ph.D. degree in applied sciences from Katholieke Universiteit Leuven (KU Leuven), Leuven, Belgium, in 1980 and 1986, respectively.

In 1987, he joined the R&D Laboratory, Interuniversity Microelectronics Center (IMEC), Leuven, where he is responsible for research in reliability physics for deep-submicrometer CMOS technologies. From October 2005 to April 2007, he was also responsible for the IMEC Post-CMOS Nanotechnology Program within IMEC's core partner research program. Since 2001, he has been a Professor with KU Leuven, where he is also the Program Director of the Master in Nanoscience and Nanotechnology and is also coordinating a European Erasmus Mundus Master Program in nanoscience and nanotechnology. He has authored or coauthored more than 500 publications in international scientific journals and international conference proceedings, and six book chapters. He is the holder of ten patents in his fields of expertise. He has made contributions to the fields of nonvolatile semiconductor memory devices and technology, reliability physics of VLSI technology, hot carrier effects in MOSFET's, TDDB of oxides, negative-bias-temperature instability effects, ESD protection and testing, plasma-processing-induced damage, electrical characterization of semiconductors, and characterization and reliability of high-k dielectrics. His current research interests include nanotechnology for post-CMOS applications, such as carbon nanotubes for interconnect applications and tunnel FETs for alternative nanowire devices.

Dr. Groeseneken became an IMEC Fellow in 2007. He has served as technical program committee member of several international scientific conferences, among which are the IEEE International Electron Device Meeting (IEDM), the European Solid-State Device Research Conference (ESSDERC), the International Reliability Physics Symposium, the IEEE Semiconductor Interface Specialists Conference (SISC), and the EOS/ESD Symposium. From 2000 to 2002, he was the European Arrangements Chair of IEDM. In 2005, he was the General Chair of the Insulating Films on Semiconductor (INFOS) conference, which was organized in Leuven, Belgium.

Authorized licensed use limited to: LIVERPOOL JOHN MOORES UNIVERSITY. Downloaded on January 28, 2010 at 10:36 from IEEE Xplore. Restrictions apply.