Key Issues and Solutions for Characterizing Hot Carrier Aging of Nanometer Scale nMOSFETs

Meng Duan, Jian Fu Zhang, Zhigang Ji, Wei Dong Zhang, Ben Kaczer, and Asen Asenov, *Fellow, IEEE*

***Abstract*— Silicon bandgap limits the reduction of operation voltage when downscaling device sizes. This increases the electrical field within-a-device and hot carrier aging (HCA) is becoming an important reliability issue again for some CMOS technologies. For nanodevices, there are a number of challenges for characterizing their HCA: the random charge–discharge of traps in gate dielectric causes "within-a-device-fluctuation (WDF)," making the parameter shift uncertain after a given HCA. This can introduce errors when extracting HCA time exponents and it will be shown that the lower envelope of the WDF must be used. Nanodevices also have substantial device-to-device variation (DDV) and multiple tests are needed for evaluating their standard deviation (σ) and mean value (μ). Repeating the time-consuming HCA tests is costly and a voltage-step-stress method is applied to reduce the number of tests by 80%. For a given number of devices under tests (DUTs), there is a little information on the accuracy of the extracted σ and μ. We will develop a method to provide this information, based on the defect-centric model. For 40 DUTs with an average of ten traps per device, the extracted μ and σ has an accuracy of ±14% and ±24%, respectively, with a 95% confidence.**

***Index Terms*— Aging, Bias Temperature Instability, defects, device-to-device variations (DDV), fluctuation, hot carriers, instabilities, random telegraph noise, reliability, time-dependent variations.**

I. Introduction

In 1980s, hot carrier aging (HCA) was the most important reliability issue as downscaling device size without reducing operation voltage (V_{dd}) increases electrical field within the device [1]–[3]. HCA was alleviated since 1990s, because of the reduced V_{dd}. As V_{dd} approaches the limit imposed by the silicon bandgap, downscaling the channel length leads to a rapid rise of HCA [4]–[6]. It has been reported that, for some CMOS technologies, HCA can even be more severe than bias temperature instabilities [4]–[6] and HCA has been revisited by many researchers recently [4]–[17]. There are important differences between the HCA of nanodevices and the classical HCA in 1980s. For example, the worst HCA used to occur under $V_g = V_d/2 = V_{dd}/2$, but HCA of nanodevices under $V_g = V_d = V_{dd}$ is substantially higher than that under $V_g = V_{dd}/2$ [5], [9], [11]. It is proposed that the HCA of nanodevices is driven by carrier energy and carrier–carrier interaction [11], [14], [16], [17] and multivibration excitation [11] plays important roles. The objective of this paper is to investigate some key issues in characterizing the HCA of nano-nMOSFETs.

The HCA kinetics follows a power law against both stress time and biases [1], [18]

$$\text{HCA} = CV^m t^n \tag{1}$$

where C is a constant. To predict the long-term HCA under operation V_{dd}, it is important to extract the exponent m and n accurately, but there are a number of challenges for this extraction from nanodevices. In our recent iedm work [5], we addressed two key issues.

1) In the presence of "within-a-device-fluctuation (WDF)" [19], [20] for nanodevices, one must use the lower envelope (LE) of the WDF when extracting the time exponent.
2) The device-to-device variation (DDV) requires repeating the tests many times to obtain the statistical properties [21], [22]. Aging tests are time consuming and its repetition is costly. The voltage-step-stress technique can reduce the number of tests by 80%.

In this paper, in addition to describe the above two issues in more detail, we extend the iedm work [5] by addressing another two key issues.

1) The accuracy of the evaluated standard deviation (σ) and the mean value (μ) of DDV increases with the number of devices under tests (DUTs). In practice, however, the number of DUTs is limited by test time. When a limited number of DUTs is used for evaluating σ and μ, there is little information on their accuracy. For the first time, we will develop a method for estimating their accuracy against the number of DUTs.
2) During HCA, positive bias temperature instability (PBTI) occurs near the source [12]. The effect of PBTI on the kinetics of HCA will be assessed.

II. Devices and Experiments

The MOSFETs used were fabricated by a 28-nm planar CMOS technology. The channel length and width are 27 × 90 nm with HK/metal gate. A wide channel length of 900 nm was also used, reducing the DDV to $< \pm 8\%$. The gate dielectric stack consists of a Hafnium oxide and a SiON interfacial layer with a 1.2-nm equivalent oxide thickness.

Manuscript received February 9, 2017; revised March 24, 2017; accepted March 25, 2017. Date of publication April 26, 2017; date of current version May 19, 2017. This work was supported by the Engineering and Physical Science Research Council of U.K. under Grant EP/L010607/1. The review of this paper was arranged by Editor C. M. Compagnoni. *(Corresponding author: Jian Fu Zhang.)*

M. Duan, J. F. Zhang, Z. Ji, and W. D. Zhang are with the Department of Electronics and Electrical Engineering, Liverpool John Moores University, Liverpool, L3 3AF, U.K. (email: J.F.Zhang@ljmu.ac.uk).

B. Kaczer is with Imec, B-3001 Leuven, Belgium.

A. Asenov is with the Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow, G12 8QQ, U.K.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2017.2691008

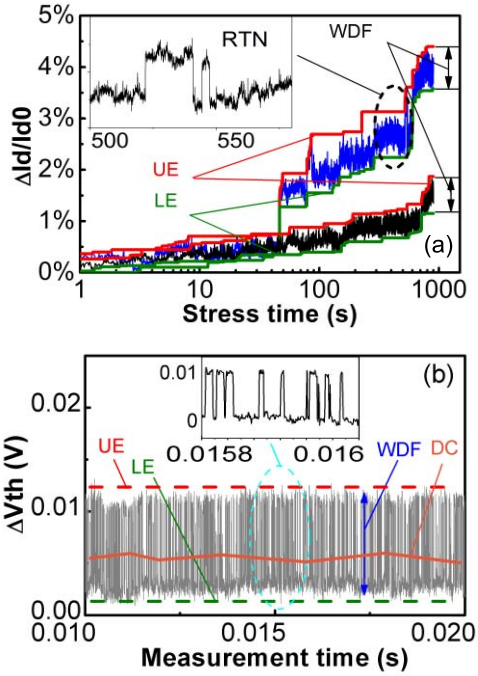


Fig. 1. (a) HCA ($V_g = V_d = 1.3$ V) of two W = 90-nm devices shows large DDV. WDF, UE, and LE are "within-a-device-fluctuation," the upper envelope, and the lower envelope. (b) Simplest form of WDF: a two-level RTN. The "dc" marked out the average value within 10 ms, as used in a typical Source-and-Measure-Unit.

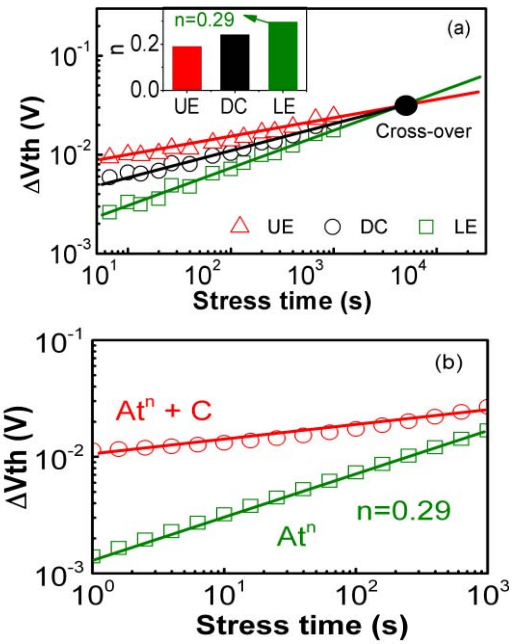


Fig. 2. (a) HCA kinetics for the mean of 40 W = 90-nm devices. UE, dc, and LE have different "*n*" (inset). (b) Incorrect inclusion of an as-grown component "*C*" gives an apparent lower "*n*."

The HCA was carried out under $V_g = V_d$ at 125 °C, rather than room temperature, as it was reported that HCA increases with temperature for modern CMOS nodes [12], [23]. The threshold voltage (Vth) was monitored from the V_g shift under a fixed drain current of 100 nA × W/L [5], [24]. The nanometer devices have an as-fabricated DDV at time zero. Prasad *et al.* [22] reported no correlation between the time-zero DDV and time-dependent DDV, while Kerber and Nigam [25] observed a weak correlation. In this paper, the effect of this time-zero DDV on the follow-on time-dependent DDV is taken into account by using the time-zero I_d-V_g of each device as its own reference. For the planar CMOS process used in this paper, HCA is more severe for nMOSFETs than for pMOSFETs when stressed under $|V_g| = |V_d|$ [11], so that this paper will focus on the HCA of nMOSFETs.

Fig. 1(a) shows a typical aging process for a 27 × 90 nm device, where the data were recorded by an oscilloscope at a sampling rate of 10^6 points/sec, giving a time resolution of 1 μ s, which is fast enough to capture the charge–discharge of traps in gate dielectric [26].

III. Results and Discussion

A. Extraction of Time Exponents n

Some typical HCA results are plotted against stress time for two nanodevices in Fig. 1(a). In addition to a substantial DDV, there is a considerable WDF [19], [20]. This fluctuation is not caused by the soft breakdown of the gate dielectric, since the gate current is two orders of magnitude less than the fluctuation in the drain current. Moreover, in its simplest form, the WDF only has two levels, a signature of random telegraph noise rather than breakdown, and one example is given in Fig. 1(b). This supports that the fluctuation in Fig. 1(a) originates from the random charge/discharge of traps in gate dielectrics.

The large DDV of WDF in Fig. 1(a) can have two sources: a large variation of trap number per device and a large variation of the impact of one trap on devices. To explain the latter, one should note that the current flow in the device is not uniform [21]. The trap will have a larger impact on a device when the local current beneath it is high [27]. It has been shown that the impact of a trap on the device follows an exponential distribution [21].

The WDF introduces uncertainty to the HCA after a given stress: the parameter shifts can be anywhere between the upper envelope (UE) and LE of the WDF [19], [20]. LE is caused by the defects that do not discharge. Fig. 1(a) shows that LE increases with HCA stress levels, so that these defects were charged by the HC stress. Once they are charged, they remain charged during the measurement. In contrast, "UE" is the upper envelope of the fluctuation. It contains two components: LE and the fluctuation. It represents the "total" degradation level. When a commercial "dc" source-and-measure unit is used, it effectively takes the average within, for example, 10 ms, as the "dc" line marked out in Fig. 1(b).

Given this uncertainty, the challenge is how to extract the time exponent (n) reliably for nanodevices. One effective method for suppressing the fluctuations in Fig. 1 is to use the mean value of multiple devices. Fig. 2(a) shows that smooth data are obtained for the UE, LE, and dc, when the mean of 40 devices was used. The n extracted from these three, however, are different, with UE giving the lowest and LE having the highest n. This leads to the crossover of LE from UE, when

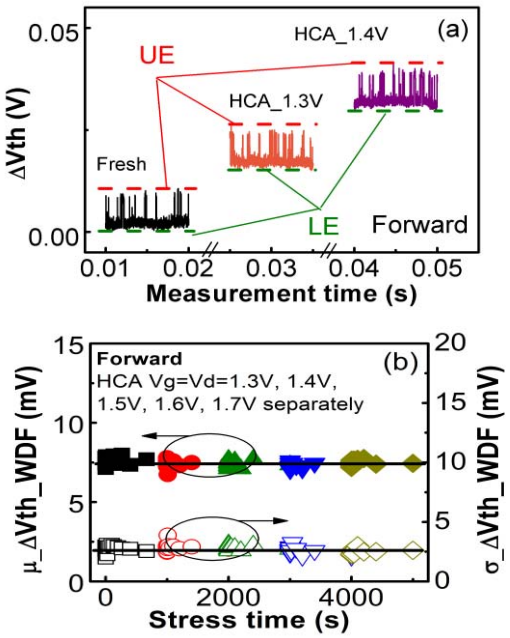


Fig. 3. (a) For $L \times W = 27 \times 90$ nm, LE increases with HCA, but WDF = UE – LE does not. (b) μ_WDF of 40 devices and its sigma does not increase with stress time.

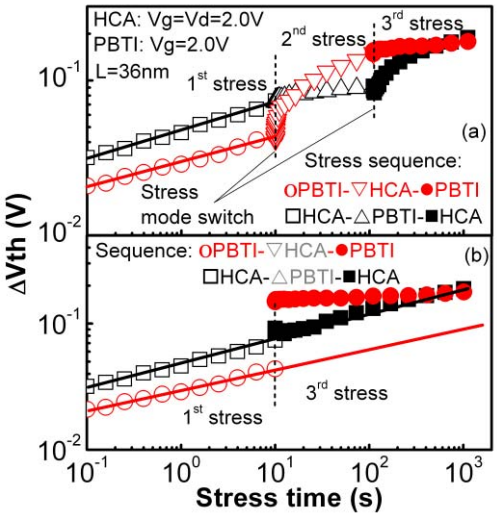


Fig. 4. (a) Black set of symbols follow the test sequence of HCA (first stress), PBTI (second stress), and HCA (third stress). The red set of symbols follow the test sequence of PBTI (first stress), HCA (second stress), and PBTI (third stress). (b) second stress periods were removed for both test sequences. The HCA kinetics is hardly affected by the preceding PBTI, but PBTI kinetics is substantially affected by the preceding HCA.

extrapolating ahead, which is not physically meaningful, as the LE should never be higher than UE [19], [20].

To investigate whether UE or LE should be used for extracting n, we examine their dependence on HCA. Fig. 3(a) clearly shows that LE increases progressively with HCA time, but the WDF = UE − LE remains the same. As a result, LE is caused by HCA, while WDF is not. WDF originates from the "as-grown" defects in fresh devices [27], [28]. To further support this, Fig. 3(b) shows that the mean of WDF for 40 devices is a constant against HCA time.

Since WDF is not caused by HCA, it should not be included when extracting the HCA time exponent [27]–[29]. In another word, n should be extracted from LE, rather than UE. UE gives a lower apparent n, because it contains as-grown traps. This can be demonstrated in Fig. 2(b): adding a constant to a power law leads to an apparent lower n.

B. Contribution of PBTI

When stressed under $V_g = V_d$, the electrical field over the gate dielectric is not uniform. At the source end, the device suffers from PBTI [12], while HCA dominates at the drain end for short channel. For long-channel nMOSFETs (e.g., 1.5 μm), it is well known that HCA reduced for higher temperature [30]. Both HCA and PBTI, however, rise with temperature for modern CMOS nodes [12], [23]. PBTI is process-dependent [31], and we now assess the relative contribution of PBTI to the aging for our devices under 125 °C. Two test sequences are used in Fig. 4:

1) HCA (first stress)-PBTI (second stress)-HCA (third stress);
2) PBTI (first stress)-HCA (second stress)-PBTI (third stress).

The same V_g were used for all stresses. A comparison of the two "first stress" in Fig. 4(a) shows that the HCA is clearly stronger than PBTI. The PBTI [second stress, the symbol "Δ" in Fig. 4(a)] after the HCA (first stress) only produces modest further aging. In Fig. 4(b), we remove the second stress period, so that the HCA (third stress, "■") is joined together with the HCA (first stress, "□"). It can be seen that two HCA essentially follows the same kinetics, so that the impact of the PBTI (second stress) on the HCA kinetics is modest. As a result, the HCA kinetics reported in this paper is dominated by the HCA process.

On the other hand, when the HCA (second stress, "▽") was applied after the PBTI (first stress), Fig. 4(a) shows that ΔVth rises substantially above the level extrapolated from the power law line of the PBTI (first stress), confirming the dominance of HCA. When HCA (second stress) was removed and the PBTI (third stress) is joined with the PBTI (first stress), Fig. 4(b) shows that the two PBTIs do not follow the same kinetics.

It should be pointed out that, although HCA dominates the aging kinetics under our test conditions ($V_g = V_d$ and channel length less than 36 nm), the relative strength of PBTI against HCA will increase for longer channel and higher V_g/V_d.

C. Extraction of Voltage Exponent m

The voltage exponent (m) is conventionally extracted by repeating the HCA under several (e.g., 4–6) different stress biases with one new device used for each bias, as shown in Fig. 5. This is acceptable for large devices where the DDV is negligible and only one test is needed for each bias. For nanometer devices, however, multiple tests have to be carried out to take their considerable DDV into account [21], [22]. The test time becomes costly, and there is a need to reduce it.

A voltage-step-stress (VSS) technique has been proposed for negative bias temperature instability [32] that allows m being extracted from just one device and reduces the number of tests by ∼80%. We investigate the applicability of VSS to

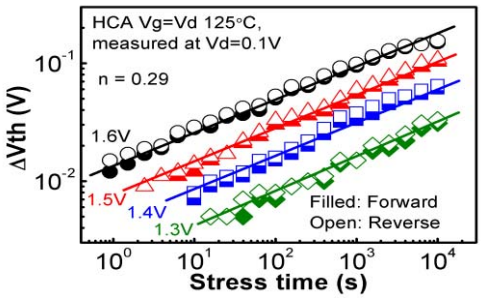


Fig. 5. Typical tests for extracting the voltage exponent repeat the tests under several different biases. The forward and reverse ΔVth measured under $V_d = 0.1$ V agree well here. The drain for stress and measurement is the same for the forward measurement, while the drain was swapped with the source after stress for the reverse measurement.

HCA here, first on a large device (27 × 900 nm) and then on nanodevices.

The principle of VSS is given in Fig. 6(a). The stress bias was applied for a given time and then raised in steps, so that different stress biases were applied to the same device. For the same stress time, HCA is higher under higher biases and typical results are given in Fig. 6(b). A HCA under a higher bias is equivalent to a HCA under a lower bias for a longer time, as illustrated in Fig. 6(a) and (c) and this equivalent time can be evaluated by [32]

$$T_N = T(V_N/V)^{m/n}. \tag{2}$$

With n extracted from the first stress step, m can be extracted by converting the data in Fig. 6(b) to a power law in Fig. 6(c). m is fit here by minimizing the least-square error between the test data and the power law, as shown by the inset of Fig. 6(c).

We now apply the VSS technique to nanodevices. Although there is a large DDV, Fig. 7 shows that their mean value agrees well with that of a large device. As a result, the time and voltage exponents for the mean value of nanodevices can be extracted in the same way as that used for a large device: the n can again be extracted by fitting the first stress step and m is extracted by fitting the data at other voltage steps with the power law, as illustrated in Fig. 6(c).

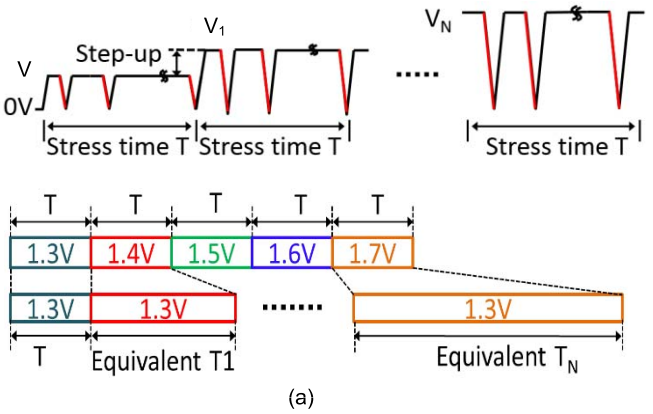
D. Verification of the Extracted Model

As the original mission for developing a model is to use it to make prediction, a model should be validated by verifying its prediction capability.

The test data under high stress biases (1.3–1.7 V) in Fig. 6 were used to extract the HCA model. We now verify its capability of predicting the HCA under low operation biases (0.9–1.2 V). Fig. 8 shows that the model prediction (lines) agrees well with the mean test data. It should be emphasized that the test data in Fig. 8 themselves were not used for fitting the model parameters.

Based on the extracted model with $n = 0.29$ and $m = 9$, Fig. 8 projects a mean ΔVth_LE = 18 mV under an operation voltage of 0.9 V for 10 years. Adding the WDF, we have ΔVth_UE = 25.5 mV. Once the mean (μ) is predicted, the next task is to determine the standard deviation (σ).

It has been proposed that the time-dependent DDV follows a defect-centric model [21], [22], [33]. This model predicts



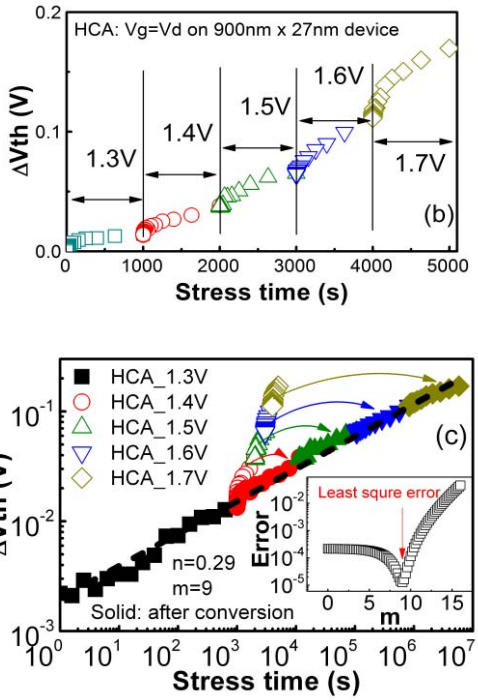


Fig. 6. VSS technique for HCA. (a) One device was stressed for a time T, and the stress $V_g = V_d$ was then stepped up. ΔVth is plotted against (b) linear and (c) log stress time. The stress time under high bias is converted to an equivalent longer time at low bias by fitting the voltage exponent "m" [inset of (c)], based on (2). The dashed line has $n = 0.29$ and $m = 9$.

that the relation between μ and σ is [33]

$$\sigma = \sqrt{2\eta\mu} \tag{3}$$

where η is the average impact of a trap on the device. Fig. 9(a) shows that HCA-induced DDV follows this relationship well. For a given predicted μ, the corresponding σ can be determined from (3) fit in Fig. 9(a). For example, for $\mu = 25.5$ mV, the corresponding σ is 13.0 mV. Fig. 9(b) shows that the statistical distribution of HCA-induced DDV agrees well with the defect-centric model.

E. Assessing the Accuracy of Statistical Properties μ and σ

As mentioned earlier, HCA tests are time consuming and only a limited number of DUTs can be used in practice for

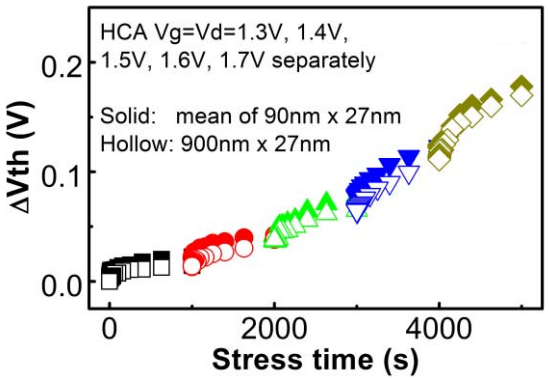


Fig. 7. Mean of 40 90 × 27 nm devices agrees well with one 900 × 27 nm for VSS stresses.

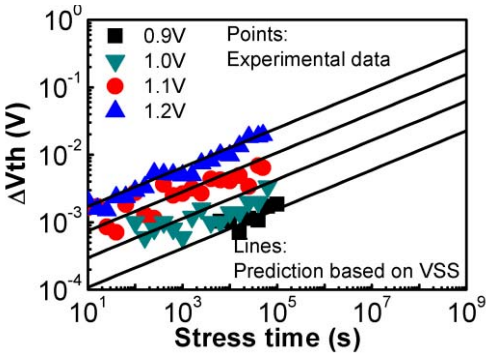


Fig. 8. Comparison of the model prediction with the test data for ΔVth. The model parameters in (1) were extracted from VSS accelerated tests (see Fig. 6). The test data at lower voltages in this figure were not used for extracting the model parameters.

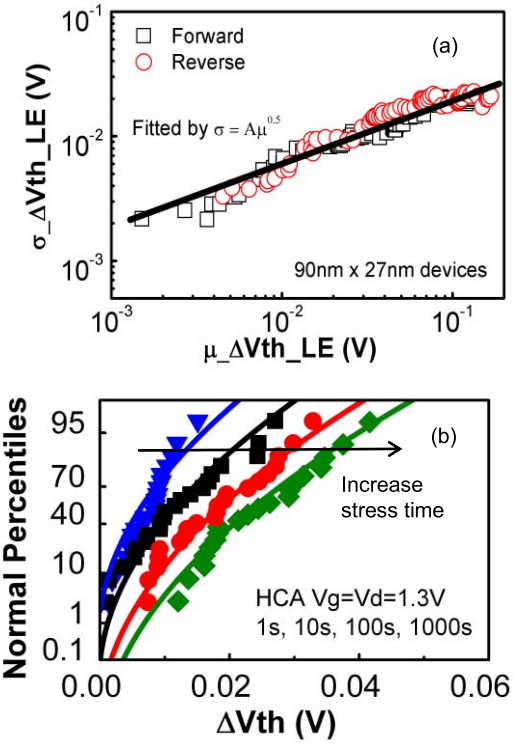


Fig. 9. Statistics of HCA-induced DDV. The lines are fit with the defect-centric distribution. (a) Sigma versus mean. (b) Distribution after different stress time.

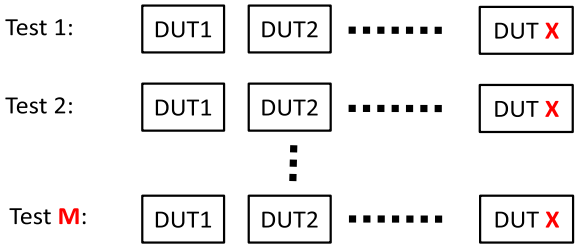


Fig. 10. Illustration of statistical tests: in a hypothetic Test 1, engineer 1 used X DUTs for extracting the μ and σ of HCA. In test 2, engineer 2 also used X DUTs, but will obtain different μ and σ, because a different set of devices were used.

extracting the statistical properties: mean (μ) and standard deviation (σ). For a given number of DUTs, the question is how accurate the extracted μ and σ is. This information is missing from early works and we will develop a new method to address it next. The results in this section are from simulation.

The defect-centric model has been verified based on the test results of 92 000 DUTs from 4000 lots [22], and the HCA reported here also follows it well. We can use this model to assess the accuracy of μ and σ extracted from a given number of DUTs by generating HCA in each hypothetic device, as detailed below.

To determine the statistical distribution of the defect-centric model, two parameters are needed: the average number of traps per device Nt and the average ΔVth induced by one trap η. The η can be estimated from (3) and Fig. 9(a) and is ~ 3.4 mV. With a typical lifetime criteria of 25–50 mV, Nt will be in the range of 7–15.

Once η and Nt is known, the number of traps in a hypothetic device nt can be randomly generated by using Poisson distribution and the threshold voltage shift induced by a trap, ΔVth,i, can be obtained by using the exponential distribution, according to the defect-centric model [21], [33]. The total ΔVth of this device is the sum of each-trap induced shift

$$\Delta V\text{th} = \sum_{i=1}^{\text{nt}} \Delta V\text{th}, i. \tag{4}$$

We will use Nt = 7.5 and η = 3.4 mV to demonstrate the method. The ΔVth for a hypothetic device, DUT1, is calculated according to (4). If we assume X devices have been used for evaluating μ and σ in a test, i.e., the test 1 in Fig. 10, we can statistically calculate the ΔVth for these X devices. These X ΔVth can then be used to calculate one μ and one σ, as represented by a data point in Fig. 11(a) and (b), respectively.

If another test engineer repeated the same test, i.e., the test 2 in Fig. 10, a different group of X devices would be used, producing a different μ and σ and give another data point in Fig. 11. Fig. 11 shows the statistical spread for 1000 tests, i.e., M = 1000 in Fig. 10, when X DUTs were used for each test. The X was varied between 10 and 1000. As expected, the spread becomes increasingly larger when a smaller number of DUTs were used for the test.

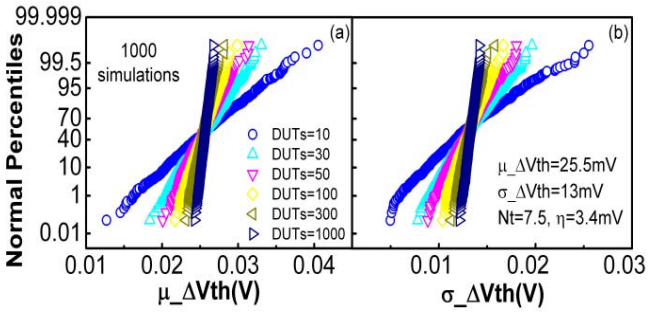


Fig. 11. (a) μ and (b) σ extracted for different DUTs (X in Fig. 10). For a given X, the tests were repeated 1000 times ($M = 1000$ in Fig. 10).

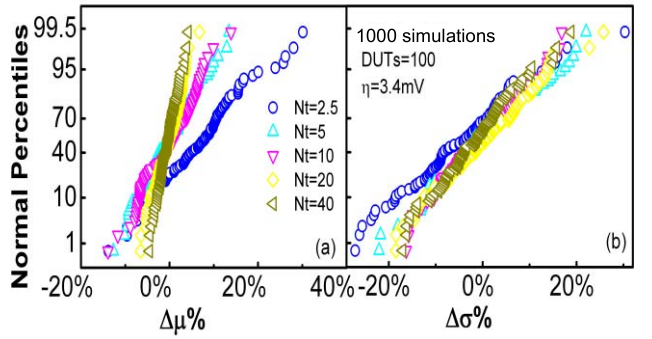


Fig. 12. Impact of the average number of traps (Nt) per DUT on the (a) μ and (b) σ extracted for DUTs = 100 when the tests were repeated 1000 times ($M = 1000$ in Fig. 10).

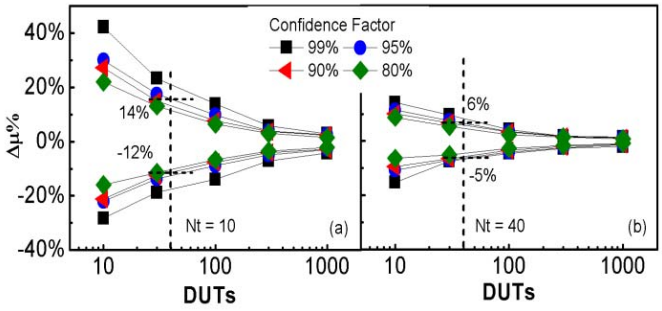


Fig. 13. Dependence of the accuracy of mean value (μ) on the number of DUTs used in a test for (a) Nt = 10 and (b) Nt = 40. The accuracy with a 95% confidence is marked out for 40 devices.

The number of DUTs is not the only parameter controlling the accuracy of μ and σ. Fig. 12 shows that for a given $X = 100$, the spread reduces for higher Nt, because a higher number of traps per device averages out device variations to a certain extent.

This work used 40 DUTs and we now assess the accuracy of the evaluated μ and σ. Fig. 13(a) shows that the evaluated μ has an accuracy within $\pm 14\%$ for Nt = 10, with a 95% confidence. To assess the impact of Nt, Fig. 13(b) shows that the accuracy reaches $\pm 6\%$ when Nt = 40. If 1000 DUTs were used, the accuracy will improve to $\pm 2.6\%$ for Nt = 10 and $\pm 1.3\%$ for Nt = 40.

The corresponding σ is given in Fig. 14. When DUTs = 40, the evaluated σ has an accuracy within $\pm 24\%$ for Nt = 10, not as accurate as μ. An increase of Nt to 40 only makes a modest improvement to $\pm 22\%$. With 1000 DUTs, an accuracy of $\pm 5\%$ can be achieved for Nt = 10.

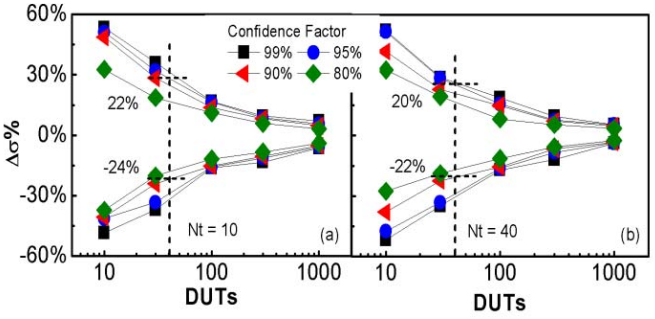


Fig. 14. Dependence of the accuracy of standard deviation (σ) on the number of DUTs used in a test for (a) Nt = 10 and (b) Nt = 40. The accuracy with a 95% confidence is marked out for 40 devices.

IV. Conclusion

This paper investigates the key issues and provides solutions for characterizing the HCA of nanodevices. It is shown that the WDF is not caused by the HCA, so that they must be excluded, when extracting the HCA time exponent. This can be achieved by using the LE of WDF. The commercial source-and-measure unit measures a data point by taking the average within a period. This includes a part of WDF, resulting in an under-estimation of time exponent. The voltage exponent can be extracted by using the VSS technique, reducing the number of tests by ~80%. HCA follows the defect-centric model well. Based on this model, the accuracy of the mean and standard deviation of DDV can be estimated for a given number of DUTs. For 40 DUTs with an average ten traps per device, the accuracy for μ and σ is $\pm 14\%$ and $\pm 24\%$, respectively, with a 95% confidence.

Acknowledgment

The authors would like to thank D. Vigar from Cambridge Silicon Radio for supplying test samples.

References

[1] C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-electron-induced MOSFET degradation—Model, monitor, and improvement," *IEEE Trans. Electron Devices*, vol. ED-32, no. 2, pp. 375–385, Feb. 1985.

[2] P. Heremans, R. Bellens, G. Groseneneken, and H. E. Maes, "Consistent model for the hot-carrier degradation in n-channel and p-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 35, no. 2, pp. 2194–2209, Feb. 1988.

[3] J. F. Zhang and W. Eccleston, "Effects of high field injection on the hot carrier induced degradation of submicrometer pMOSFET's," *IEEE Trans. Electron Devices*, vol. 42, no. 7, pp. 1269–1276, Jul. 1995.

[4] A. Bravaix *et al.*, "Impact of the gate-stack change from 40 nm node SiON to 28 nm high-*k* metal gate on the hot-carrier and bias temperature damage," in *Proc. IRPS*, 2013, pp. 2D.6.1–2D.6.9.

[5] M. Duan *et al.*, "Hot carrier aging and its variation under use-bias: Kinetics, prediction, impact on Vdd and SRAM," in *Proc. IEDM*, 2015, pp. 547–550.

[6] B. Kaczer *et al.*, "Origins and implications of increased channel hotcarrier variability in nFinFETs," in *Proc. IRPS*, 2015, pp. 3B.5.1–3B.5.6.

[7] N. H. Hsu *et al.*, "Intrinsic hot-carrier degradation of nMOSFETs by decoupling PBTI component in 28 nm high-*k*/metal gate stacks," in *Proc. IRPS*, 2012, pp. XT.13.1–XT.13.4.

[8] J. H. Stathis *et al.*, "Reliability challenges for the 10 nm node and beyond," in *Proc. IEDM*, 2014, pp. 522–525.

[9] G. T. Sasse, F. G. Kuper, and J. Schmitz, "MOSFET degradation under RF stress," *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 3167–3174, Nov. 2008.

[10] A. J. Scholten, D. Stephens, G. D. J. Smit, G. T. Sasse, and J. Bisschop, "The relation between degradation under DC and RF stress conditions," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2721–2728, 2011.

[11] A. Bravaix, V. Huard, D. Goguenheim, and E. Vincent, "Hot-carrier to cold-carrier device lifetime modeling with temperature for low power 40 nm Si-bulk NMOS and PMOS FETs," in *Proc. IEDM*, 2011, pp. 622–625.

[12] K. T. Lee *et al.*, "PBTI-associated high-temperature hot carrier degradation of nMOSFETs with metal-gate/high-*k* dielectrics," *IEEE Trans. Electron Devices*, vol. 29, no. 4, pp. 389–391, Apr. 2008.

[13] C. Liu, K. T. Lee, S. Pae, and J. Park, "New observations on hot carrier induced dynamic variation in nano-scaled SiON/poly, HK/MG and FinFET devices based on on-the-fly HCI technique: The role of single trap induced degradation," in *Proc. IEDM*, 2014, pp. 836–839.

[14] S. E. Rauch and G. La Rosa, "The energy-driven paradigm of NMOSFET hot-carrier effects," *IEEE Trans. Device Mater. Rel.*, vol. 5, no. 4, pp. 701–705, Dec. 2005.

[15] P. Magnone, F. Crupi, N. Wils, H. P. Tuinhout, and C. Fiegna, "Characterization and modeling of hot carrier-induced variability in subthreshold region," *IEEE Trans. Electron Devices*, vol. 59, no. 8, pp. 2093–2099, Aug. 2012.

[16] M. Bina, K. Rupp, S. Tyaginov, O. Triebl, and T. Grasser, "Modeling of hot carrier degradation using a spherical harmonics expansion of the bipolar Boltzmann transport equation," in *Proc. IEDM*, 2012, pp. 713–716.

[17] Y. M. Randriamihaja, A. Zaka, V. Huard, M. Rafik, D. Rideau, and D. Roy, "Hot carrier degradation: From defect creation modeling to their impact on NMOS parameters," in *Proc. IRPS*, 2012, pp. XT.15.1–XT.15.4.

[18] *Failure Mechanisms and Models for Semiconductor Devices*, JEDEC, Alexandria, VA, USA, 2011.

[19] M. Duan *et al.*, "New analysis method for time-dependent device-to-device variation accounting for within-device fluctuation," *IEEE Trans. Electron Devices*, vol. 60, no. 8, pp. 2505–2511, Aug. 2013.

[20] M. Duan *et al.*, "Development of a technique for characterizing bias temperature instability-induced device-to-device variation at SRAM-relevant conditions," *IEEE Trans. Electron Devices*, vol. 61, no. 9, pp. 3081–3089, Sep. 2014.

[21] B. Kaczer *et al.*, "Origin of NBTI variability in deeply scaled pFETs," in *Proc. IRPS*, 2010, pp. 26–32.

[22] C. Prasad *et al.*, "Bias temperature instability variation on SiON/poly, HK/MG and trigate architectures," in *Proc. IRPS*, 2014, pp. 6A.5.1–6A.5.7.

[23] A. Bravaix, C. Guerin, V. Huard, D. Roy, J. M. Roux, and E. Vincent, "Hot-carrier acceleration factors for low power management in DC-AC stressed 40 nm NMOS node at high temperature," in *Proc. IRPS*, 2009, pp. 531–548.

[24] S. W. M. Hatta *et al.*, "Energy distribution of positive charges in gate dielectric: Probing technique and impacts of different defects," *IEEE Trans. Electron Devices*, vol. 60, no. 5, pp. 1745–1753, May 2013.

[25] A. Kerber and T. Nigam, "Correlation of BTI induced device parameter degradation and variation in scaled metal gate/high-*k* CMOS technologies," in *Proc. IRPS*, 2014, pp. 6A.6.1–6A.6.6.

[26] M. Duan *et al.*, "Key issues and techniques for characterizing time-dependent device-to-device variation of SRAM," in *Proc. IEDM*, 2013, pp. 774–777.

[27] M. Duan *et al.*, "Time-dependent variation: A new defect-based prediction methodology," in *Proc. VLSI Technol. Symp.*, 2014, pp. 74–75.

[28] R. Gao *et al.*, "Predictive As-grown-generation (A-G) model for BTI-induced device/circuit level variations in nanoscale technology nodes," in *Proc. IEDM*, 2016, pp. 778–781.

[29] Z. Ji *et al.*, "Negative bias temperature instability lifetime prediction: Problems and solutions," in *Proc. IEDM*, 2013, pp. 413–416.

[30] F.-C. Hsu and K. Y. Chiu, "Temperature dependence of hot-electron-induced degradation in MOSFET's," *IEEE Electron Device Lett.*, vol. 5, no. 5, pp. 148–150, May 1984.

[31] 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, 2006, Art. no. 093716

[32] Z. Ji *et al.*, "A single device based voltage step stress (VSS) technique for fast reliability screening," in *Proc. IRPS*, 2014, pp. GD.2.1–GD.2.4.

[33] L. M. Procel, F. Crupi, J. Franco, L. Trojman, and B. Kaczer, "Defect-centric distribution of channel hot carrier degradation in nano-MOSFETs," *IEEE Electron Device Lett.*, vol. 35, no. 12, pp. 1167–1169, Dec. 2014.

Meng Duan received the B.Eng. and M.Eng. degrees from Xidian University, Xi'an, China, in 2000 and 2003, respectively, and the Ph.D. degree from Liverpool John Moores University, Liverpool, U.K., in 2013.

He is currently a Post-Doctoral Researcher with the University of Glasgow, Glasgow, U.K., where he is involved in the field of device modeling. His current research interests include nanometer MOSFETs reliability, variability, and device modeling on 1T-DRAM.

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

Since 2001, he has been a Professor of Microelectronics with Liverpool John Moores University, Liverpool. His current research interests include the qualification, characterization, and modeling of nanometer-size devices.

Dr. Zhang has served or is serving as a member of the technical program committees for several international conferences, including IEDM.

Zhigang Ji received the Ph.D. degree from Liverpool John Moores University (LJMU), Liverpool, U.K., in 2010.

He is currently a Senior Lecturer with LJMU. His current research interests include characterization and modeling on future logic and memory devices.

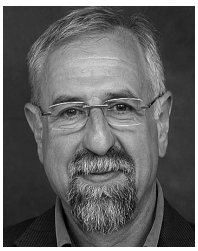
Wei Dong Zhang received the Ph.D. degree from Liverpool John Moores University (LJMU), Liverpool, U.K., in 2003.

He is currently a Professor of Nanoelectronics with LJMU. His research interests include the characterization and quality assessment of nanoscale resistive switching and flash memory devices, CMOS devices, and GaN HEMT devices.

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, OH, USA, in 1996 and 1998, respectively.

In 1998, he joined the Reliability Group, IMEC, Leuven, Belgium, where he is currently a Principal Scientist.

Dr. Kaczer has been on the IEEE Transactions Electron Devices Editorial Board since 2011.



Asen Asenov (M'96–SM'05–F'11) was the Founder and the CEO at Gold Standard Simulations Ltd, Glashow, U.K., acquired by Synopsys in 2016. He is currently the James Watt Professor of Electrical Engineering, and the Leader of the Glasgow Device Modeling Group, The University of Glasgow, Glasgow, U.K., and also splits his time between Glashow University and Synopsys. He has over 800 publications and over 200 invited talks.

Mr. Asenov is a fellow with the Royal Society of Edinburgh.