Development of a technique for characterizing bias temperature instability-induced device-to-device variation at SRAM-relevant conditions

M. Duan, J. F. Zhang, Z. Ji, W. Zhang, B. Kaczer, T. Schram, R. Ritzenthaler, G. Groeseneken, and A. Asenov

***Abstract*— SRAM is vulnerable to device-to-device variation (DDV), since it uses minimum-sized devices and requires device-matching. In addition to the as-fabricated DDV at time-zero, aging induces a time-dependent DDV (TDDV). Bias temperature instability (BTI) is a dominant aging process. A number of techniques have been developed to characterize the BTI, including the conventional pulse IV, random telegraph noises (RTN), time dependent defect spectroscopy (TDDS), and TDDV accounting for the within-device fluctuation (TVF). These techniques, however, cannot be directly applied to SRAM, because their test conditions do not comply with typical SRAM operation. The central objective of this work is to develop a technique suitable for characterizing both the NBTI and PBTI in SRAM. The key issues addressed include the SRAM relevant sensing Vg, measurement delay, capturing the upper-envelope of degradation, sampling rate, and measurement time window. The differences between NBTI and PBTI are highlighted. The impact of NBTI and PBTI on the cell-level performance is assessed by simulation, based on experimental results obtained from individual devices. The simulation results show that, for a given static noise margin, test conditions have a significant effect on the minimum operation bias.**

I. INTRODUCTION

SRAM can occupy over 50% of the space for system-on-a-chip products, forcing it to use the minimum-sized devices. The device-to-device variability (DDV) increases as their size reduces [1-15]. SRAM is especially vulnerable to DDV, since it has a high packing density and requires device-match for its two cross-coupled inverters in Fig. 1. Apart from the as-fabricated DDV at time-zero [1-6], aging introduces a time-dependent DDV (TDDV) [7-15]. Aging originates from a number of sources: negative bias temperature instability (NBTI) [16-19], hot carriers [20], and oxide breakdown [21]. For high-k/SiON gate dielectric stack, positive bias temperature instability (PBTI) can also be important [22-24]. Both NBTI for pMOSFETs and PBTI for nMOSFETs are investigated in this work.

BTI can induce TDDV in two ways. On one hand, different devices in a circuit can suffer from different BTIs. For example, the pMOSFET 'PR' in Fig. 1 suffers from NBTI stress, whilst 'PL' does not, so that TDDV between PR and PL increases with time. On the other hand, even if two devices were stressed under the same conditions, the stochastic nature of charging-discharging the as-grown defects [8,11] and generating new defects [14,15] will result in TDDV.

A number of techniques have been developed to characterize the BTI and the TDDV, including the conventional pulse IV [25-26], random telegraph noises (RTN) [11-13, 27-29], time dependent defect spectroscopy (TDDS) [8], and TDDV accounting for the within-a-device fluctuation (TVF) [14,15]. Although these techniques have provided valuable information on the defects, they cannot be directly applied to SRAM, because their test conditions do not comply with the SRAM operation, as analyzed in section III. The objective of this work is to develop a technique suitable for characterizing both the NBTI and PBTI in SRAM. The key issues addressed include the sensing Vg, measurement delay, capturing the upper-envelope of degradation, sampling rate, and measurement time window. The impact of BTI-induced TDDV on the static noise margin (SNM) and the minimum operation voltage of SRAM will be simulated and their sensitivity to test conditions will be highlighted.

II. DEVICES AND EXPERIMENTS

Both pMOSFETs and nMOSFETs have a channel length of 50 nm and a width of 90 nm. The gate dielectric stack is HfO_2 with an Al_2O_3 cap layer and the equivalent oxide thickness is 1.45 nm. The gate is TiN.

The experimental setup is given in Fig. 2(a) and Id was measured at |Vd|=0.1 V through a fast operational amplifier [30]. To find the response time of the setup, a step Vg was applied to the input and Fig. 2(b) shows that the output Id can response in 50 ns. Before aging, a reference Id-Vg was taken in 3 μs and is shown in Fig. 3. The degradation during this short measurement time is negligible [14,26].

Manuscript received MMM DDD, 2014. This work was supported by the Engineering and Physical Science Research Council of UK under the grant no. EP/I012966/1 and EP/L010607/1. The review of this paper was arranged by Editor Elyse Rosenbaum.

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

B. Kaczer, T. Schram, R. Ritzenthaler, and G. Groeseneken are with IMEC, Leuven B3001, Belgium. G. Groeseneken is also with KU Leuven, Leuven, Belgium.

A. Asenov is with Department of Electronics and Electrical Engineering, University of Glasgow.

The test follows a 'stress-then-sense' procedure [25,26] and the technique developed in this work requires the device being stressed under 'use' voltage. The stress was carried out at Vg=+1.4 V for PBTI of nMOSFETs and Vg=-1.4 V for NBTI of pMOSFETs. During stress, Id was continuously monitored on-the-fly against time. To assess the aging on the SRAM trip voltage, |Vg| was ramped down from 1.4 V to the trip point of the inverter, |Vtr|=0.7 V, in 3 µs to minimize the recovery [26] and Id was measured. Fig. 3(a) shows that the stress lowered Id by ΔIds at |Vtr|=0.7 V. The shift of trip voltage, ΔVtr, was taken against the reference IV. Fig. 3(b) plots the ΔVtr against the ΔId/Id measured at Vg=-1.4 V for 21 devices. They have a linear relation for all tested devices, which is used to convert ΔId/Id to ΔVtr.

All tests and measurements were carried out at 125 °C. Two channels of the oscilloscope were used for Id to obtain two different resolutions for the stress and measurement phases, respectively [15].

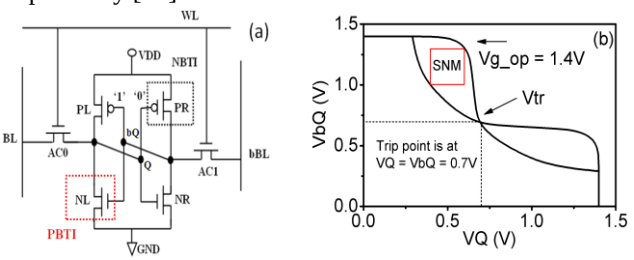


Fig. 1 (a) A standard symmetric 6-transistor SRAM bitcell with PR, but not PL under NBTI stress. NL, but not NR, under PBTI stress when node Q is '0' and bQ is '1'. If the bitcell does not flip, the stress is DC. (b) The butterfly characteristics during read. Vtr is the trip point.

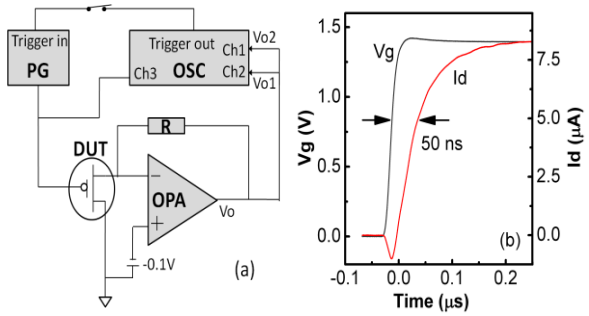


Fig. 2 (a) The test circuit and (b) The characteristic response time of the system. Vg was stepped with an edge time of 10 ns and Id=(Vo-Vd)/R rises to 63% of its peak in 50 ns.

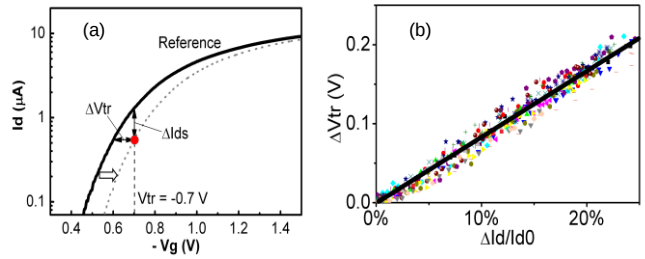


Fig. 3 (a) ΔVtr evaluation. The reference IV was taken from a fresh device. After stress, Id at the trip point, Vg=Vtr, was measured and ΔVtr was taken from the Vg shift from the reference IV at this Id. (b) The linear relation between ΔId/Id measured under Vg=-1.4 V and ΔVtr. The data were obtained from **21** devices.

III. SHORTCOMINGS OF EXISTING TECHNIQUES

BTI tests were generally carried out on individual devices [7-16] and care must be excised when using these test data for assessing the impact on SRAM. In principle, the aging of a device in a circuit is the same as that for an individual device, provided the same voltage is applied during both stress and the measurement. To apply the test data obtained from individual devices to SRAM, it is crucial to align the measurement condition with the SRAM operation. In this section, the typical operation conditions of SRAM will be briefly reviewed first and the misalignment between the existing measurement techniques and SRAM operation will then be pointed out.

A. *Typical operation conditions of SRAM*

SRAM has three basic operation modes: read, write, and hold. BTI mainly occurs in hold-mode and Fig. 1(a) shows that, when Q='0' and bQ= '1', NL and PR suffers PBTI and NBTI, respectively, but NR and PL does not. This weakens NL and PR and maximizes the mismatch of the two inverters. In another word, the effects of PBTI and NBTI are 'adding', rather than cancelling, for SRAM. If a bitcell's content does not change, NL and PR will be under the worst BTI stress: DC stress without recovery.

The SNM for a standard 6-transistor SRAM is smaller during read than hold, because the pre-charged BL in Fig. 1(a) will partially pull-up Q through voltage dividing between AC0 and NL. The BTI weakens NL and PR by reducing their overdrive voltage |Vg-Vth| and in turn, the SNM. In contrast, this reduction of |Vg-Vth| does not reduce the write noise margin, since a weakened NL and PR will make the bitcell easier to flip. As a result, SRAM is most vulnerable to BTI during read. The typical read time is in the order of tens of nano-seconds, during which Q and bQ can approach Vtr in Fig. 1(b) and flip the bitcell erroneously. The SRAM-relevant BTI should use the operation bias for hold, Vg_op, as the stress voltage and the degradation should be sensed at Vg=Vtr.

B. *Shortcomings of existing techniques*

The conventional pulse IV (p-IV): The p-IV allows measuring threshold voltage shift at a pre-set stress time in ~µs that minimizes recovery during measurement [25,26]. It worked well for large devices where Id has little fluctuation, as shown in Fig. 4(a). For nm-size devices, however, Id fluctuates substantially (Fig. 4b) due to the discrete nature of charging-discharging. Fig. 4(c) shows that for the same stress, the degradation varies substantially, depending on the charging level immediately before triggering the measurement. The conventional p-IV makes one measurement and gives only one ΔVth at a pre-set time [25,26]. It does not take this within-a-device-fluctuation (WDF) into account and does not give the range of threshold voltage shift for a given stress time. The DC measurement of conventional parameter analyzer

typically uses the average value within a measurement time window of, say 10 ms, as represented by the point 'C' in Fig. 4(b). It does not capture the fluctuation well.

The random telegraph noises (RTN): Unlike the conventional p-IV, the RTN captures the WDF by monitoring Id against time [11-13,27-29]. The problem with the standard RTN technique is that it requires both charging and discharging a defect to produce a 'telegraph-noise', as shown in Fig. 5(a). However, there are substantial amount of defects that do not discharge under Vg_op= -1.4 V, as marked by the 'LE' for the lower envelope of ΔId/Id in Fig. 5(b) and they are not probed by RTN.

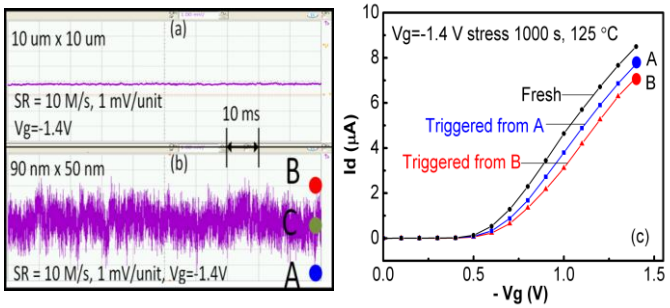


Fig. 4 After a 1000 sec stress, Id under Vg=-1.4 V fluctuates little for a 10×10 μm device (a), but substantially for a 90×50 nm device (b). The time window in (a) and (b) is 0.1 sec. Note the output signal Vo in Fig. 2(a) is negative, so that Id at 'B' is smaller than Id at 'A'. The point 'C' represents a typical DC measurement at a speed of 10 ms per point. When triggering Vg-switch from the point 'A' and 'B', the recorded degradation is significantly different (c).

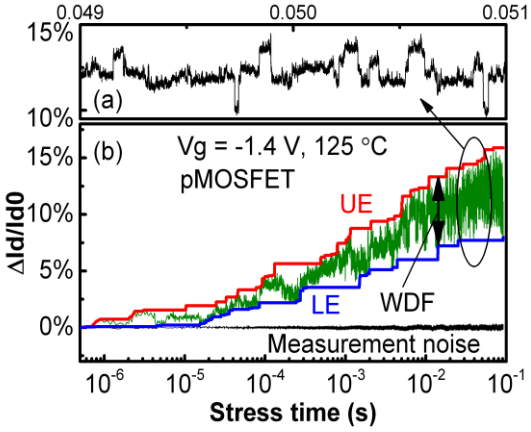


Fig. 5 (a) For a given defect number, the RTN signal by enlarging the circled region of (b). Defects, however, increase with time. The 'UE' and 'LE' represent the upper and lower envelop of raw data and their difference is caused by the within-a-device-fluctuation (WDF) under a given Vg. The system measurement noise is negligible.

To reduce the discharge time and observe it within the measurement time window, tw, the RTN is typically monitored at a |Vg| lower than the operation |Vg_op| [27,28]. As analyzed in the section III.A, for SRAM, Vg_op can be applied indefinitely during the 'hold' of a bitcell, but Vg may approach Vtr during read for a typical time of only ~10 ns. As a result, for SRAM, the charge fluctuation should be monitored at Vg_op, rather than at |Vg|<|Vg_op| [27,28]. Fig. 6 presents the transient ΔVtr when |Vg| was stepped down from |Vg_op| to |Vtr|. The ΔVtr is flat in ~μs range [11,26], although discharge happened at longer time. It should be pointed out that the flat ~μs region in Fig. 6 is not caused by system limitation, since its response time is 50 ns (Fig. 2b). As a result, there is no RTN signal in the time domain relevant to SRAM operation at Vtr, so that RTN cannot be used to measure ΔVtr in a time-domain relevant to reading a SRAM cell. Moreover, RTN is difficult to analyze when there are more than 4 traps [29].

Time dependent defect spectroscopy (TDDS): TDDS [8] probes individual defects by monitoring their discharge after Vg switching from Vg_op to a low level close to Vth, as illustrated in Fig. 6. When triggered from the same point 'B' repeatedly, the same SRAM-relevant charge level in ~μs was observed. The subsequent discharge, however, introduces a considerable variation due to the stochastic nature of discharge. As a result, there is no TDDS signal at Vtr for SRAM operation condition and there is no unique relation between the ΔVtr at ~ μs and the amount of discharge measured within a limited time window. To further explore this point, Fig. 7 compares two cases triggered from 'A' and 'B'. As expected, ΔVtr('B')>ΔVtr('A') in ~μs, but the two curves actually cross over later, confirming that there is no unique relation between ΔVtr at ~ μs and ΔVtr at longer time.

Another difficulty with the TDDS is that discharge does not always complete within a practical time window, especially after a relatively long stress (e.g. >1000 sec), because of the 'permanent component' [16,18] originating from the generated anti-neutralization positive charges [31-33]. Moreover, TDDS does not directly give the within-a-device-fluctuation (WDF) at Vg_op, which plays an important role in the BTI-induced TDDV, as to be shown in the sections IV.B and IV.D.

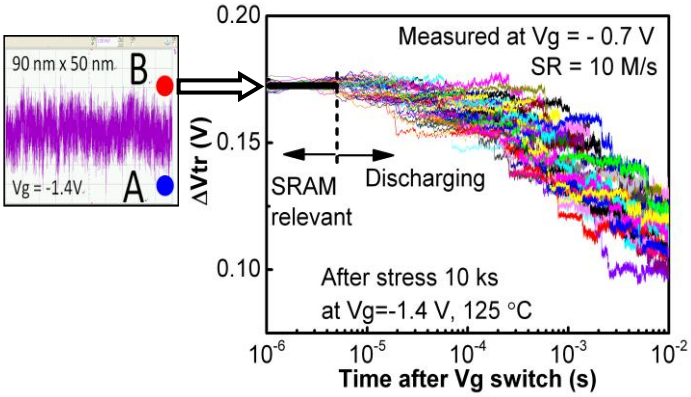


Fig. 6 Vg-switch from -1.4 V to -0.7 V was triggered from the point 'B' and ΔVtr is monitored for 10^{-2} sec under Vtr=-0.7 V. Vg=-1.4 V was then re-applied until the Id reached the point 'B' and the Vg-switch triggered again. This sequence was repeated 50 times. The discharge is absent for <5 μs. The stochastic discharge at longer time induces considerable variation, which is irrelevant to SRAM during read operation.

The TDDV accounting for the WDF (TVF): The TVF [14,15] is based on the measurement in Fig. 5(b), where the upper envelop, 'UE', was divided into two components: a WDF and the defect that does not discharge under Vg_op, i.e. the lower envelope (LE). To capture the UE and LE, ΔId/Id was monitored continuously under Vg_op=-1.4 V. By comparing the WDF, LE, and UE of different devices, TVF allows separating device-to-device variation (DDV) from WDF.

The TVF proposed in our early work [14,15] monitors the degradation by ΔId/Id under Vg_op, but SRAM is most vulnerable to BTI during read when Vg approaching Vtr. The SRAM-aging should be characterized by ΔVtr, therefore. In addition the sampling rate used in [14,15] was only 100 point per second, which substantially underestimated the WDF, as to be shown in the section IV.A. The applicability of TVF to PBTI of nMOSFETs is not tested, either.

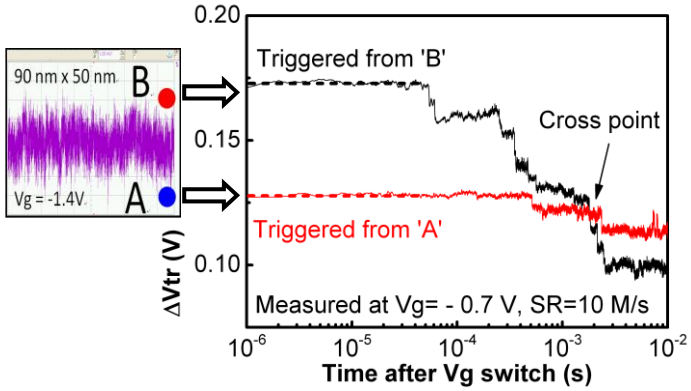


Fig. 7 After triggering from the points 'A' and 'B' in Fig. 4b respectively, ΔVtr('A')<ΔVtr('B') in ~ μs, but ΔVtr('A')>ΔVtr('B') at 10 ms, indicating the ΔVtr measured at typical DC speed is not a reliable representation of the SRAM-relevant ΔVtr in ~ μs.

IV. CHARACTERIZE BTIs-INDUCED TDDV FOR SRAM

The TVF technique [14,15] will be revised and extended to nMOSFETs to probe the BTIs-induced TDDV for SRAM by addressing a number of key issues, including the sensing Vg, measurement delay, capturing the UE of degradation, sampling rate, and measurement time window. We will first study how to measure the SRAM-relevant BTI for one device reliably and then investigate the device-to-device variation.

A. *Measuring a single device*

AC or DC stress: Digital circuits are typically under AC stress with a reduced degradation when compared with DC stress. A SRAM bitcell, however, can hold its content indefinitely and the DC stress is used here, therefore.

Sensing Vg and measurement delay: As analyzed in the section III.A, the SRAM is most vulnerable to BTI during read, when Vg can approach Vtr in ~ ns. A degradation of Vtr can cause a flip, so that BTI should be assessed by measuring ΔVtr at a sensing Vg~Vtr [26], as shown in Fig. 3. Ideally, the measurement delay should be only ~ ns, but this cannot be achieved for wafer level measurements. Fig. 7 shows that a measurement delay of ~ μs is adequate to minimize the recovery during measurement.

Capturing the UE of ΔVtr: When a bitcell's content does not change, WDF occurs under Vg_op. A bitcell can be read many times and there can be millions of bitcells in a SRAM. It is inevitable that some bitcells will be read when the charging reaches its upper-envelop (UE), i.e. the point 'B' in Figs. 4(b)&(c). It is important to capture the UE of ΔVtr during test, therefore. This requires monitoring Id under Vg_op, rather than under a |Vg|<|Vg_op|.

It has been reported that defects can have a wide range of charging and discharging time [8,11-13]. To capture the fast trap, the sampling rate, SR, must be sufficiently high. To capture the slow trap, the measurement time window must be sufficiently wide.

Sampling rate: To investigate the dependence of WDF on SR, it is desirable to fix the number of defects during the measurement. This can be achieved by first stressing a device heavily, so that further defect generation is negligible during the subsequent measurement. After a stress time of 1,000 sec or longer, tests show that further increase in defect number in the subsequent 40 sec is less than 0.27 mV, which is within the test resolution. Fig. 8(a) gives the WDF measured within a time window of tw=1 sec at different SR after 50 ksec stress. The WDF increased substantially with SR, but Fig. 8(b) shows that a saturation is reached around SR=1 M/s. This indicates that the defects responsible for WDF have a charging/discharging time larger than ~ μs, in agreement with the lack of fluctuation in the ~ μs region in Fig. 7. We will use SR=10 M/s for the on-the-fly measurement of ΔId/Id hereafter and emphasize that it is enough to capture the fast traps for NBTI.

Measurement window: Fig. 9(a) indicates that the WDF is approximately constant as the measurement time window, tw, increases. This is, however, an artifact. Once the tw is plotted in logarithmic scale, Fig. 9(b) shows that WDF increases with tw. As a result, tw should be made as long as possible. Since Id is monitored on-the-fly, the longest possible tw is tw=stress time and this is achieved by recording Id continuously during test. The time window here is for measuring ΔId/Id under Vg_op=-1.4 V and it increases with stress time.

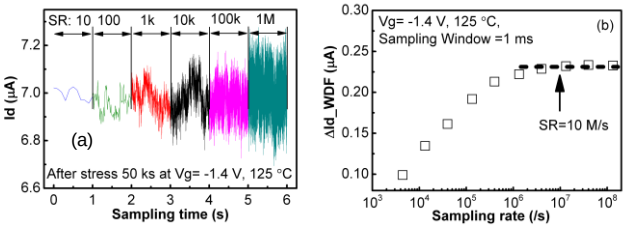


Fig. 8 Dependence of within-a-device-fluctuation (WDF) on the sampling rate for pMOSFETs. Id fluctuation increases with sampling rate (SR) when SR<1 M/s (a), but saturates after SR>1 M/s (b). The device was stressed for 50 ksec to ensure that further degradation during the measurement itself is negligible.

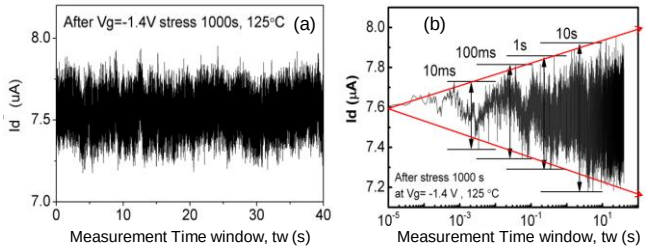


Fig. 9 Dependence of within-a-device-fluctuation on the measurement time window for pMOSFETs. Although the Id fluctuation appears insensitive to time in a linear scale in (a), it clearly increases with time when plotted in a logarithmic scale in (b) for a pMOSFET.

After examining the WDF caused by the NBTI of pMOSFETs, we turn our attention to the PBTI of nMOSFETs. As illustrated in Fig. 1, when the pMOSFET (PR) is suffering NBTI, the nMOSFET (NL) is suffering PBTI at the same time. Fig. 10(a) shows that the PBTI-induced WDF increases by a factor over 3 with sampling rate, and then saturates for SR≥400 k/s. Similar to the NBTI-induced WDF, the PBTI-induced WDF also increases with measurement time window in Fig. 10(b). The revised TVF proposed for NBTI can be used to probe the PBTI of nMOSFETs.

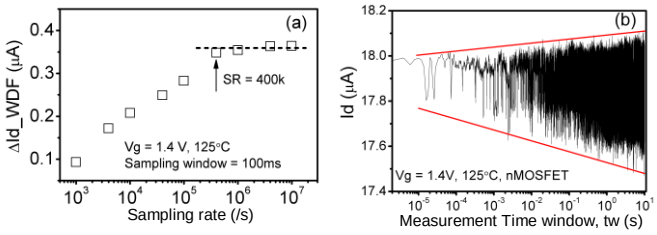


Fig. 10 Within-a-device-fluctuation of PBTI in an nMOSFET. The nMOSFET was pre-stressed for 1 ks under Vg=+1.4 V to ensure negligible new defect creation during the measurement itself. (a) Dependence on sampling rate. Id fluctuation increase with sampling rate (SR), but saturates after SR>400 k/s. (b) Dependence on the measurement time window.

In short, the TVF monitors ΔId/Id on-the-fly and then convert it to ΔVtr at the trip point based on the pulse-IV. It is different from the conventional on-the-fly method that measures ΔVth at the stress Vg, rather than at the trip point. It is also different from the conventional pulse-IV that did not monitor the ΔId/Id on-the-fly and did not capture the upper-envelope of ΔId/Id.

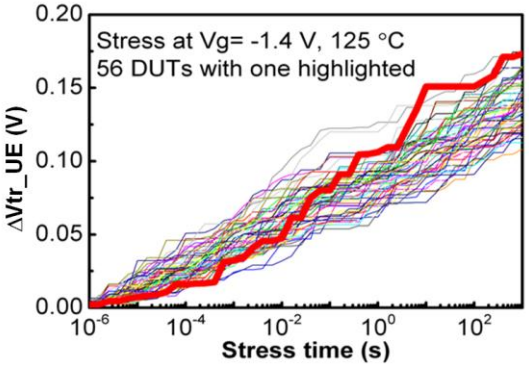


Fig. 11 The ΔVtr_UE against stress time, recorded for 56 90×50 nm pMOSFETs. For each device, ΔId/Id0 was continuously monitored under Vg=-1.4 V and its UE was extracted, as shown in Fig. 5(b). This ΔId/Id0_UE was then converted to ΔVtr_UE by using their relation in Fig. 3(b). The thick line highlights the device with the highest degradation at 1000 sec, although it is close to the bottom at short time. The step-like change is caused by the discreteness of charges.

B. *NBTI-induced TDDV*

After studying how to characterize the SRAM-relevant BTI for a single device by the improved TVF, the same test like Fig. 5 was repeated for 56 different pMOSFETs to study the NBTI-induced time-dependent device-to-device variation (TDDV). The 'ΔId/Id0_UE' measured in Fig. 5 is converted to ΔVtr_UE by using Fig. 3(b) for each device and the device-to-device variations are given in Fig. 11. It increases in steps and the gap between two steps varies due to the discreteness and stochastics of charges. The thick line represents the devices of the largest UE at a stress time of 1000 sec. Their positions relative to other devices change with time. For example, the device of largest UE at 1000 sec had one of the lowest UE at short time. As a result, it is essential to measure the NBTI in all devices at all time.

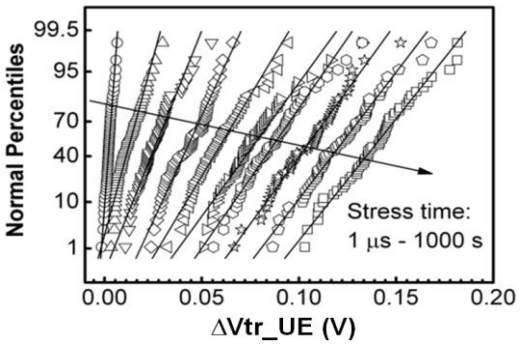


Fig. 12 The statistical distribution of ΔVtr_UE for NBTI of pMOSFETs. An increase of stress time raised not only the average, but also the variation. The solid lines were fitted with the Gaussian distribution.

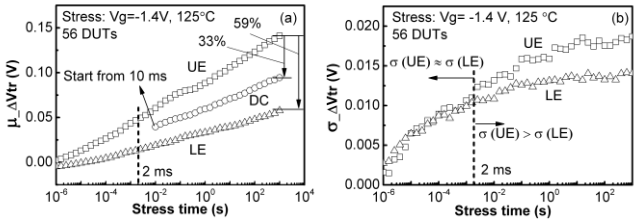


Fig. 13 NBTI of pMOSFETs: The kinetics of the average, μ (a), and standard deviation, σ (b). μ_UE> 2×μ_LE, due to the contribution of WDF under a given Vg. The 'DC' represents the typical results measured by a quasi-DC parameter analyzer. σ_UE≈σ_LE within 2 ms, but σ_UE>σ_LE afterwards.

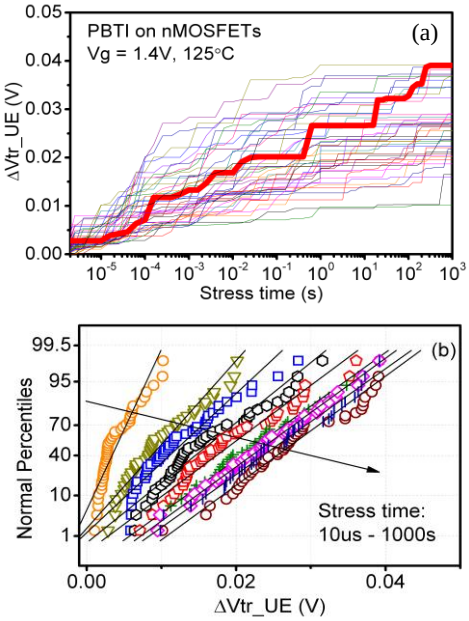


Fig. 14 PBTI of 45 nMOSFETs. Stress conditions were Vg=+1.4 V under 125 °C. The thick line in (a) highlights the device with the highest degradation at 1000 s, although it was close to the bottom at short stress time. (b) Statistical distribution of ΔVth_UE, solid lines were fitted with the Gaussian distribution.

The distributions of UE are given in Fig. 12 and they can be fitted reasonably with the Gaussian distribution. The kinetics of the average (μ) and standard deviation (σ) is given in Figs. 13(a)&(b), respectively. As time increases, both μ and σ increase. μ_UE more than doubles μ_LE and the typical DC measurement also substantially underestimates μ_UE. Although μ_LE<μ_UE/2 in Fig. 13(a), σ _LE ≈ σ _UE initially (i.e. < 2 ms) in Fig. 13(b), indicating the WDF contributes little to DDV initially. For longer stress, however, σ _UE is clearly above σ _LE and the WDF does vary for different devices.

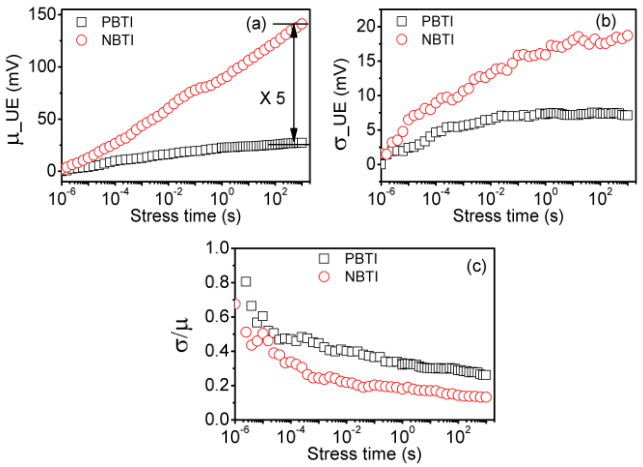


Fig. 15 A comparison of the kinetics: PBTI versus NBTI under |Vg|=1.4 V at 125 °C. (a) The average, μ; (b) the standard deviation, σ; and (c) σ/μ. NBTI is substantially higher than PBTI, but PBTI has higher σ/μ, indicating a higher relative variation.

C. *PBTI-induced TDDV*

Similar results were obtained for PBTI in nMOSFETs, as shown in Figs. 14(a)&(b). Figs. 15(a)-(c) compare the UE of NBTI and PBTI when stressed under |Vg|=1.4 V. After 1000 sec stress, the average of NBTI is 5 times of that of PBTI. The difference in their standard deviation, however, is smaller. This leads to a higher σ/μ for PBTI (Fig. 15(c)), indicating the device-to-device variation is relatively larger for PBTI of nMOSFETs.

D. *Impact on SRAM*

Ideally, to assess the impact of BTI on SRAM performance, one would like to insert the aged devices into a SRAM cell and measure it directly [34]. Such test structure, however, is not available to this work, so that we simulate the impact. The design and optimization of SRAM must meet multiple constraints, such as static and dynamic margins and power consumptions. In this work, we focus only on one key parameter, the static read noise margin (SNM).

The worst case BTIs for SRAM: To illustrate the potential impact of BTI on SRAM, we use simulation based on a 45 nm technology. The SNM was simulated with both access transistors, AC0 and AC1 in Fig. 1(a) switched on. The bias at Q, VQ, is swept from zero to 1.4 V and the VbQ is recorded to obtain the voltage transfer characteristics (VTC) for the inverter on the right hand side (PR-NR). We then sweep VbQ and record VQ to obtain the VTC for the inverter on the left hand side (PL-NL). These two VTCs form the 'butterfly' in Fig. 1(b).

We will first consider the NBTI and PBTI separately and then combine them. When reading a bitcell, the worst case is that both NBTI and PBTI reach their UE, which maximizes the mismatch between the two inverters. As an example, Fig. 16 shows that this occurred three times for two devices. Given that there are often multi-millions bitcells in a SRAM, one should consider the combined impact of UE_NBTI and UE_PBTI.

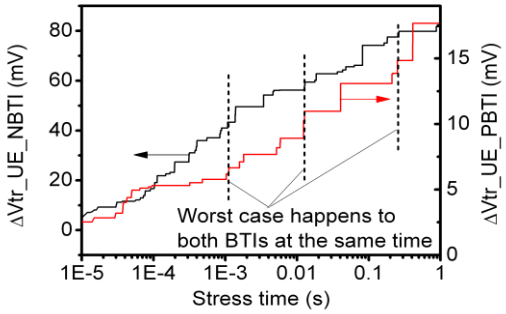


Fig. 16 The UE of NBTI and PBTI. The worst TDDV occurs when NBTI and PBTI reach their UE simultaneously, as marked by the vertical dotted lines.

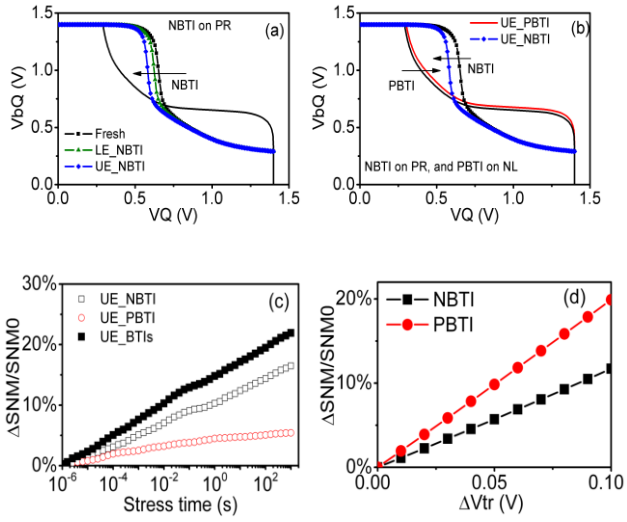


Fig. 17 Simulation of the BTI impact on SRAM for a 45 nm technology. (a) The NBTI impact on SRAM. The degradation is much larger for the UE_NBTI than the LE_NBTI. (b) Degradation by both NBTI and PBTI. (c) SNM degradation caused by UE_NBTI, UE_PBTI, and their combination. (d) For the same |ΔVtr|, PBTI results in a larger ΔSNM/SNM.

BTIs impact on SNM: One of the most important parameters for SRAM read stability is the Static Noise Margin (SNM) [35]. SNM is typically measured from the size of the square that can be fitted into the two voltage transfer characteristics (VTC) for the butterfly (Fig. 1(b)). To simulate the impact of NBTI, we assume that PR is aged by ΔVtr(NBTI) and simulate the VTC for the PR-NR inverter. Fig. 17(a) shows the VTC(PR-NR) is shifted towards left, reducing SNM. The impact of UE_NBTI is significantly higher than that of LE_NBTI, so that it is important to capture the UE.

Similarly, to simulate the impact of PBTI, we assume that NL is aged by ΔVtr(PBTI) and simulate the VTC for the PL-NL inverter and Fig. 17(b) shows the VTC(PL-NL) is shifted

towards right. Since the two VTCs were shifted in the opposite direction by NBTI and PBTI, respectively, their effect on the SNM reduction is adding.

Fig. 17(c) shows that reduction of SNM caused by UE_NBTI and UE_PBTI separately and by combining them. Although NBTI has a larger impact, the contribution of PBTI is around one-fourth of the combined and must be taken into account. Early works [23,36] show that PBTI of high-k/SiON stack is sensitive to processing conditions and the smaller PBTI in Fig. 17(c) could be achieved through process optimization. For the same |ΔVtr|, however, Fig. 17(d) shows that PBTI has a larger effect on SNM degradation. This is because during read, VQ in Figs. 1(a)&(b) is not at zero, due to the voltage dividing between AC0 and NL [37].

Fig. 18 gives the distribution of ΔSNM/SNM. A combination of NBTI and PBTI increases not only its average, but also its variation.

BTIs impact on the minimum V_{DD}: To reduce power consumption, lowering V_{DD} is desirable. A lower V_{DD}, however, reduces SNM, as shown in Fig. 19. For a given SNM, the required V_{DD} can be substantially increased by the BTI-induced TDDVs. For example, for a 45nm CMOS technology, to keep a SNM=180 mV, V_{DD} is about 0.9 V before BTI. It increases to 1.2 V when considering the UE_NBTI and rises further to 1.38 V after combining UE_NBTI and UE_PBTI.

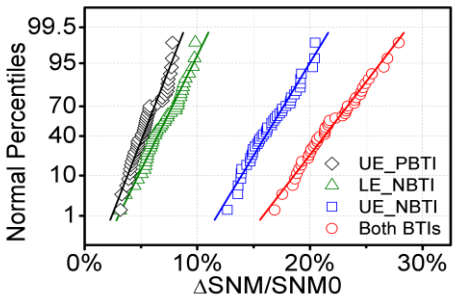


Fig. 18 Statistical distributions of ΔSNM/SNM caused by NBTI, PBTI, and NBTI+PBTI. The devices were stressed under |Vg|=1.4 V for 1000 sec at 125 °C. The SNM was simulated based on the experimental UE_NBTI and UE_PBTI.

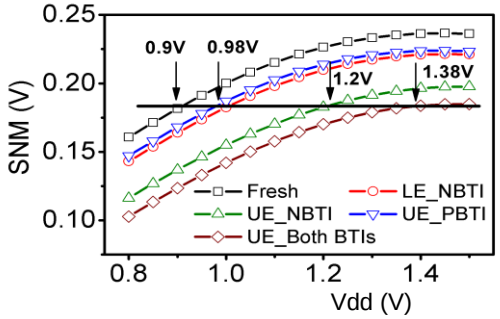


Fig. 19 SNM versus V_{DD} with NBTI, PBTI, and NBTI+PBTI. For a required SNM, BTIs increase the minimum V_{DD} substantially. The SNM was simulated based on the experimental UE_NBTI and UE_PBTI.

V. CONCLUSIONS

The existing techniques are not suitable for probing the BTI-induced TDDV for SRAM and a technique suitable for this task has been developed by improving the Time-dependent Variation accounting within-device Fluctuation (TVF). The key issues addressed include the sensing Vg, measurement delay, capturing the upper-envelop of degradation, sampling rate, and measurement time window. The TVF allows capturing both the fast and slow traps, minimizing the missing of a defect during measurement. The results show that the within-a-device fluctuation is significant and it should be captured at the operation bias, rather than at a Vg close to threshold level. Both NBTI for pMOSFETs and PBTI for nMOSFETs were investigated and compared. The NBTI is substantially higher than the PBTI, but PBTI also makes considerable contribution to TDDV for the high-k/SiON stack. For the same average ΔVtr, PBTI has a large variation and also a large effect on the SNM.

REFERENCES

[1] A. Asenov, B. Cheng, D. Dideban, U. Kovac, N. Moezi, C. Millar, G. Roy, A. R. Brown, S. Roy, and Ieee, "Modeling and Simulation of Transistor and Circuit Variability and Reliability," *IEEE Custom Integrated Circuits Conference*, pp. 1-8, 2010.

[2] K. J. Kuhn, "Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale CMOS," in *IEDM Tech. Dig.*, 2007, pp. 471-474.

[3] P. A. Stolk, and D. B. M. Klaassen, "The effect of statistical dopant fluctuations on MOS device performance," in *IEDM Tech. Dig.*, 1996, pp. 627-630.

[4] T. Mizuno, J. Okamura, and A. Toriumi, "Experimental-Study of Threshold Voltage Fluctuation Due to Statistical Variation of Channel Dopant Number in Mosfets," *IEEE Trans. Elec. Dev.*, vol. 41, no. 11, pp. 2216-2221, Nov, 1994.

[5] J. T. Horstmann, U. Hilleringmann, and K. F. Goser, "Matching analysis of deposition defined 50-nm MOSFET's," *IEEE Trans. Elec. Dev.*, vol. 45, no. 1, pp. 299-306, Jan, 1997.

[6] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterization and Modeling of Mismatch in Mos-Transistors for Precision Analog Design," *IEEE J. Solid-State Circuits*, vol. 21, no. 6, pp. 1057-1066, Dec, 1986.

[7] Y. Mitani, "Experimental Study on Origin of V_{TH} Variability under NBT Stress," in *Proc. Int. Rel. Phys. Symp.*, 2011, pp. 857-862.

[8] T. Grasser, H. Reisinger, P.-J.Wagner, and B.Kaczer, "Time-dependent defect spectroscopy for characterization of border traps inmetal-oxide-semiconductor transistors," *Phys. Rev. B*, vol. 82, art.no. 245318, 2010.

[9] S. Pae, J. Maiz, C. Prasad, and B. Woolery, "Effect of BTI Degradation on Transistor Variability in Advanced Semiconductor Technologies," *IEEE Trans. Dev. Mat. Rel*, vol. 8, no. 3, pp. 519-525, Sep. 2008.

[10] S. E. Rauch, "Review and reexamination of reliability effects related to NBTI-induced statistical variations," *IEEE Trans. Dev. Mat. Rel*, vol. 7, no. 4, pp. 524-530, Dec, 2007.

[11] C. Z. Liu, J. B. Zou, R. S. Wang, R. Huang, X. Q. Xu, J. H. Liu, H. M. Wu, and Y. Y. Wang, "Towards the Systematic Study of Aging Induced Dynamic Variability in nano-MOSFETs: Adding the Missing Cycle-to-Cycle Variation Effects into Device-to-Device Variation," in *IEDM Tech. Dig.*, 2011, pp.571-574.

[12] J. B. Zou, R.S. Wang, N.B. Gong, R. Huang, X. Q. Xu, J. J. Q, C. Z. Liu, J. P. Wang, J. H. Liu, J. G. Wu, S. F. Yu, P. P. Ren, H. M. Wu, S. W. Lee, Y. Y. Wang, "New Insights into AC RTN in Scaled High-κ / Metal-gate MOSFETs under Digital Circuit Operations," in *VLSI Tech. Symp.*, 2012, p139-140.

[13] P. P. Ren, P. Hao, C. Z. Liu, R. S. Wang, X. B. Jiang, Y. X. Qiu, R. Huang, S. F. Guo, M. L. Luo, J. B. Zou, M. Li, J. P. Wang, J. G. Wu, J. H. Liu, W. H. Bu, W. S. Wong, S. Yu, H. M. Wu, S. W. Lee, Y. Y. Wang, "New Observations on Complex RTN in Scaled High-κ/Metal-gate MOSFETs

– the Role of Defect Coupling under DC/AC Condition," in *IEDM Tech. Dig.*, 2013, pp.778-781.
[14] M. Duan, J. F. Zhang, Z. Ji, W. Zhang, B. Kaczer, T. Schram, R. Ritzenthaler, G. Groeseneken, and A. Asenov, "New analysis method for time-dependent device-to-device variation accounting for within-device fluctuation," *IEEE Trans. Elec. Dev.*, vol. 60, no. 8, pp. 2505-2511, 2013.
[15] M. Duan, J. F. Zhang, Z. Ji, J. G. Ma, W. Zhang, B. Kaczer, T. Schram, R. Ritzenthaler, G. Groeseneken, and A. Asenov, "Key issues and techniques for characterizing Time-dependent Device-to-Device Variation of SRAM," in *IEDM Tech. Dig.*, 2013, pp.774-777.
[16] T. Grasser, T. Aichinger, G. Pobegen, H. Reisinger, P. J. Wagner, J. Franco, M. Nelhiebel, and B. Kaczer, "The 'Permanent' Component of NBTI: Composition and Annealing," in *Int. Rel. Phys. Symp.*, 2011, pp.605-613.
[17] M. H. Chang, and J. F. Zhang, "On positive charge formed under negative bias temperature stress," *J. Appl. Phys.*, vol. 101, no. 2, art.no. 024516, 2007.
[18] M. Duan, J. F. Zhang, Z. Ji, W. Zhang, B. Kaczer, S. De Gendt, and G. Groeseneken, "New insights into defect loss, slowdown, and device lifetime enhancement," *IEEE Trans. Elec. Dev.*, vol. 60, no. 1, pp. 413-419, Jan., 2013.
[19] J. F. Zhang, M. H. Chang, Z. Ji, L. Lin, I. Ferain, G. Groeseneken, L. Pantisano, S. De Gendt, and M. M. Heyns, "Dominant Layer for Stress-Induced Positive Charges in Hf-Based Gate Stacks," *IEEE Elec. Dev. Lett.*, vol. 29, no. 12, pp. 1360-1363, Dec, 2008.
[20] J. F. Zhang, and W. Eccleston, "Effects of High Field Injection on the Hot Carrier Induced Degradation of Submicrometer pMOSFETs," *IEEE Trans. Elec. Dev.*, vol. 42, no. 7, pp. 1269-1276, Jul, 1995.
[21] W. D. Zhang, J. F. Zhang, C. Z. Zhao, M. H. Chang, G. Groeseneken, and R. Degraeve, "Electrical signature of the defect associated with gate oxide breakdown," *IEEE Elec. Dev. Lett.*, vol. 27, no. 5, pp. 393-395, May, 2006.
[22] J. F. Zhang, "Defects and instabilities in Hf-dielectric/SiON stacks (Invited Paper)," *Microelectron. Eng.*, vol. 86, no. 7-9, pp. 1883-1887, Jul-Sep, 2009.
[23] 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, art. no.093716, 2006.
[24] C. Z. Zhao, M. B. Zahid, J. F. Zhang, G. Groeseneken, R. Degraeve, and S. De Gendt, "Properties and dynamic behavior of electron traps in HfO_2/SiO_2 stacks," *Microelectron. Eng.*, vol.80, pp.366-369, 2005.
[25] T. Yang, M. F. Li, C. Shen, C. H. Ang, C. X. 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 Tech. Symp.*, 2005, pp. 92-93.
[26] Z. G. 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. Elec. Dev*, vol. 56, no. 5, pp. 1086-1093, May, 2009.
[27] T. Nagumo, K. Takeuchi, T. Hase, and Y. Hayashi, "Statistical Characterization of Trap Position, Energy, Amplitude and Time Constants by RTN Measurement of Multiple Individual Traps," in *IEDM Tech. Dig.*, 2010, pp. 628-631.
[28] N. Tega, H. Miki, F. Pagette, D. J. Frank, A. Ray, M. J. Rooks, W. Haensch, and K. Torii, "Increasing Threshold Voltage Variation due to Random Telegraph Noise in FETs as Gate Lengths Scale to 20 nm," in *VLSI Symp. Tech. Dig.*, 2009, pp. 50-51.
[29] H. Kurata, S. Kamohara, K. Tokami, Y. Ikeda, R. Yamada, N. Tega, H. Miki, T. Osabe, A. Kotabe, K. Otsuga, "Anomalously Large Threshold Voltage Fluctuation by Complex Random Telegraph Signal in Floating Gate Flash Memory," in *IEDM Tech. Dig.*, 2006, pp. 218-221.
[30] Z. Ji, J. F. Zhang, and W. Zhang, "A new mobility extraction technique based on simultaneous ultra-fast Id-Vg and Ccg-Vg measurements in MOSFETs," *IEEE Trans. Elec. Dev.*, vol. 59, no. 7, pp. 1906-1914, 2012.
[31] C. Z. Zhao and J. F. Zhang, "Effects of hydrogen on positive charges in gate oxides," *J. Appl. Phys.*, vol. 97, no. 7, art.no. 073703, Apr. 2005.
[32] 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. Elec. Dev.*, vol. 55, pp.1647-1656, 2008.
[33] J. F. Zhang, M. H. Chang, and G. Groeseneken, "Effects of measurement temperature on NBTI," *IEEE Elec. Dev. Lett.*, vol.28, no.4, pp.298-300, 2007.
[34] A. T. Krishnan, V. Reddy, D. Aldrich, J. Raval, K. Christensen, J. Rosal, C. O'Brien, R. Khamankar,A. Marshall, W-K. Loh, R. McKee, S. Krishnan, "SRAM Cell Static Noise Margin and VMIN Sensitivity to Transistor Degradation" in *IEDM Tech. Dig.*, 2006, pp. 77-80.
[35] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells", *IEEE J. Solid-State Cir.*, vol. SC-22, no. 5, pp.748-754, 1987.
[36] M. B. Zahid, R. Degraeve, J. F. Zhang, G. Groeseneken, "Impact of Process Conditions on Interface and High-k Trap Density Investigated by Variable Tcharge-Tdischarge Charge Pumping (VT^2CP)," *Microelectronic Eng.*, vol.84, pp.1951-1955, 2007.
[37] A. Bansal, R. Rao, J. J. Kim, S. Zafar, J. H. Stathis, "Impact of NBTI and PBTI in SRAM Bit-cells: Relative Sensitivities and Guidelines for Application-Speci fic Target Stability/Performance," in *Int. Rel. Phys. Symp.*, 2009, pp.745-749.

Meng Duan received the B.Eng. and M.Eng. degrees from Xidian University, Xi'an, China, in 2000 and 2003, respectively, and Ph.D degree from Liverpool John Moores University in 2013. He is current with Liverpool John Moores University as a Post-Doctoral Researcher in the field of aging-induced time-dependent device variability.

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.

He has been a Professor of Microelectronics with Liverpool John Moores University, Liverpool, since 2001. His current research interests include the aging, variability, characterization, and modelling of nanometer-size devices.

Zhigang Ji received the B.Eng degree from Tsinghua University in 2003, M.Sc degree from Peking University in 2006, and Ph.D. degree from Liverpool John Moores University (LJMU), Liverpool, U.K., in 2010. He has been a Lecturer at LJMU since 2011. His research interests include modelling and defect characterization of MOS devices

Wei Dong Zhang received the Ph.D. degree from Liverpool John Moores University (LJMU), Liverpool, U.K., in 2003. He has been a Reader in Microelectronics at LJMU since 2010. His current research interests include reliability characterization and quality assessment of CMOS and memory devices.

Ben Kaczer received the M.S. degree in physical electronics from Charles University, Prague, Czech Republic, and the M.S. and Ph.D. degrees in physics from the Ohio State University, Columbus, OH, USA, in 1992, 1996, and 1998, respectively. He joined the Reliability Group of IMEC, Leuven, Belgium, in 1998, where he is a Principal Scientist. Dr. Kaczer is currently serving on the IEEE TRANSACTIONS ON ELECTRON DEVICES Editorial Board.

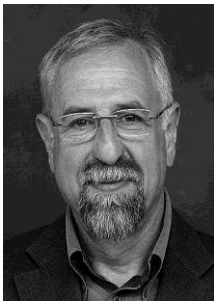
T. Schram: Not available.

Romain Ritzenthaler (M'09) received the Ph.D degree from Institut National Polytechnique de Grenoble, Grenoble, France, in 2006.

He is a Device Engineer with Interuniversity Micro-Electronics Center, Leuven, Belgium, since 2011.

Guido Groeseneken (F'05) received the Ph.D. degree in applied sciences from KU Leuven, Leuven, Belgium, in 1986.

He has been with Interuniversity Micro-Electronics Center, Leuven, Belgium, since 1987.



Asen Asenov (M'96–SM'05–F'11) received the Ph.D. degree in solid-state physics from the Bulgarian Academy of Sciences in 1989.

He is currently with Gold Standard Simulations, Ltd., Glasgow, U.K., and with Glasgow University, Glasgow.