Time-dependent variation: A new defect-based prediction methodology

M. Duan, J. F. Zhang, Z. Ji, W. Zhang,
B. Kaczer*, T. Schram*, R. Ritzenthaler*, A. Thean*, G. Groeseneken*, and A. Asenov**
School of Engineering, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK (j.f.zhang@ljmu.ac.uk)
*IMEC, Leuven B3001, Belgium, **Dept. Electronics and Electrical Engineering, University of Glasgow

Introduction: Variation of nm-devices is a threat especially to the VLSI circuits requiring device-matching, such as SRAM. The discreteness of aging-induced charges causes a Time-Dependent Variation (TDV), which has received many attentions recently [1-4], but a widely-accepted technique for predicting the long-term TDV is missing. TDV is widely characterized by random telegraph noises (RTN) [5-9]. This work will show typical RTN substantially underestimating the TDV. Early works focused on short stress time (e.g. ≤1,000 sec) under the implicit assumption that all defects were as-grown [5-10]. We will show that the generation of new defects is significant and their properties and kinetics are distinctively different from those of as-grown traps. Based on the 'As-gown-Generation (AG)' model, for the first time, a methodology for predicting the long term TDV of threshold voltage shift, ΔVth, under a given operation bias, Vg_op, is proposed and *its prediction-capability is verified*.

Devices and Experiments: pMOSFETs have HfO_2 with an Al_2O_3 cap-layer and an EOT of 1.45 nm. The channel L×W is 50×90 nm. A Vg_op=-1.4 V was used for demonstration purpose. ΔVth was measured in 3 μs to minimize recovery during measurement [2]. Test temperature is 125 °C and the sampling rate is 10 M/sec [11].

TDV has two components: a within-a-device fluctuation (WDF) and a device-to-device variation (DDV). WDF originates from the stochastic charging/discharging for a given number of defects in the same device, while DDV is caused by a device-to-device variation in defects number (**Fig. 1**). Early works either address only WDF [5-9] or mix WDF with DDV [12]. We will consider them separately to enable reliable predictions.

Capturing the worst WDF: A circuit should tolerate the worst charging level when it fluctuates. For a measurement time window, tw, the defects with a charging/discharging time, t*≤tw, dominate the WDF. Early works [10, 12] carried out multiple tests with the same tw to increase chances for capturing defects (**Fig. 2**). In principle, the same defect can also be captured by a single long test with tw>>t*. **Fig. 2** compares the WDF from 100 repeated short tests of tw=20 ms with a single long test of tw=100×20 ms, confirming their statistical equivalence. A single long tw is preferred for its simplicity. To minimize missing defects, we propose monitoring ΔVth continuously for as long as practical and then extrapolating it (**Fig. 3**).

Defects and properties of WDF: WDF mainly originates from the traps located close to Ef at the interface, where trapping probability changes rapidly (**Fig. 4a**). Ef is below Ev at the interface for typical Vg_op and the hole traps below Ev are as-grown (**Fig. 4b**) [13]. As a result, *WDF originates from as-grown hole traps (AHT)*. The energy density of AHT increases when energy level is lowered (**Fig. 4c**), leading to the increase of WDF for higher |Vg| (**Fig. 4d**). This means that the RTN recorded under |Vg|<|Vg_op| by some early works [5-9] underestimates WDF. **Fig. 5a** shows that the WDF appears increasing with stress level, but this is an artifact caused by the increasing tw (**Fig. 3**). The *WDF does not increase with stress level for a fixed tw* (**Fig. 5b**), confirming their 'as-grown' nature.

Inclusion of generated defects: In addition to WDF, there is a component that does not discharge under a given Vg_op and increases with stress, corresponding to the lower envelope ('**LE**') in **Fig. 5a**. The 'LE' originates from traps sufficiently above Ef so that they do not discharge (**Fig. 4a**). Although AHTs below Ev do not increase with stress time (**Fig. 4b**), new traps clearly are generated above Ev (**Fig. 4b**). Since they do not discharge under Vg_op (**Fig. 4a**), they do not give RTN signals, but contribute to ΔVth through LE. The real ΔVth is WDF+LE (**Fig. 5a**) and *its up-envelope '**UE**' is substantially higher than WDF/RTN* recorded with a tw≤1000 sec (**Fig. 5b**), the typical tw used by early RTN works [5-9].

Device-to-device variation (DDV): **Fig. 6a** gives the WDF for 56 devices and each point was measured as the lower dashed line in **Fig. 5b**. WDF has a substantial DDV: spreading by nearly ×3. It follows a Gaussian distribution (**Fig. 6b**) and the σ increases with tw (**Fig. 6c**). For a given tw, however, σ_WDF is independent of stress time (**Fig. 6d**), since the defects are as-grown.

The LE also has a considerable DDV (**Figs. 7a&b**) and the trap creation is stochastic. At 1000 sec, the spread is around ×3 and its distribution is given in **Fig. 7b**. σ_LE increases with stress time (**Fig. 7c**), unlike the constant σ_WDF (**Fig. 6d**). This supports that LE and WDF are dominated by different types of defects.

Modelling: The different time dependence of LE (**Fig. 8a**) and WDF (**Fig. 8b**) should be modeled by different kinetics. Their average of 56 devices, μ, is a smooth function of time, making reliable modelling possible. μ will be modeled first, followed by σ.

The μ_ LE does not follow a power law in **Fig. 8a**. Since LE does not fluctuate with time, it also should be captured by large devices and follow the same model. The ΔVth of large devices (e.g. 10×1 μm) follows the 'As-gown-Generation (AG)' model (**Fig. 9a**). After experimentally separating the generated defects (GD) from AHT [14], ΔVth(GD) follows a power law well for both large and small devices (**Fig. 9b**), laying the foundation for prediction. The 'AG' model works equally well for the μ_ LE (**Fig. 8a**). The μ_WDF in **Fig. 8b** has a linear relation with log(tw) over 9 decades.

Once μ is known, σ can be obtained through its power law relation with μ for both LE (**Fig. 8c**) and WDF (**Fig. 8d**). An exponent of 0.38 for WDF agrees well with the value reported by early work [15], but the exponent for LE is only 0.20.

Predictions: A model is of value only if it can predict. The required lifetime often is 10 years, while the practical test is typically limited to days, so that a model should have the capability to predict two decades ahead. To test this prediction capability, the test data in the last two decades were not used for fitting. The predicted μ_UE and σ_UE agree well with the measured value (**Figs. 10a&b**).

A step-by-step guide for predicting the long-term TDV: (i) Monitor ΔVth under the Vg_op continuously (**Fig. 5a**) for multiple devices; (ii) Obtain the DDV μ_LE and μ_WDF (**Figs. 8a&b**); (iii) Apply 'AG' model to μ_LE and Fit μ_WDF (**Figs. 8b&b**); (iv) Evaluate σ from μ (**Figs. 8c&d**). (v) Calculate μ_UE=μ_LE+μ_WDF and $\sigma\_UE=(\sigma\_LE^2+\sigma\_WDF^2)^{0.5}$ (**Figs. 10a&b**).

Prediction of the long term yield: **Fig. 11a** verifies that ΔVth distribution can be predicted reliably two decades ahead and **Fig. 11b** gives the lifetime-induced yield. The distribution is narrower for larger ΔVth(lifetime), since the variation, σ/μ, reduces for larger μ (inset of **Fig.10b**). Other circuit-specific parameters, such as SNM for SRAM, can be converted from ΔVth (e.g. **Fig. 12**).

Conclusions: For the first time, different impacts of as-grown and generated defects on nm-sized devices are demonstrated. As-grown hole traps are responsible for WDF, which increases with Vg_op and tw. The generated defects are substantial, but do not contribute to WDF and consequently are not detected by RTN. The non-discharging component follows the same model as that for large devices: the 'AG' model. Based on this defect framework, a new methodology is proposed for test engineers to predict the long term TDV and yield and its prediction-capability is verified.

Acknowledgement: This work is supported by the EPSRC of UK under Grant Nos. EP/I012966/1 and EP/L010607/1.

[1] E. R. Hsieh et al, IEDM2013, p770. [2] M. Duan et al., TED2013, p2505. [3] M. Toledano-Luque, et al., VLSI 2011, p152. [4] A. Asenov, et al., DATE 2011, p1. [5] J. Zou, et al., VLSI 2013, p186. [6] N. Tega, et al., VLSI 2009, p.50. [7] T. Nagumo, et al., IEDM 2009, p.759. [8] K. Takeuchi, et al., VLSI 2009, p.54. [9] H. Miki, *et al.*,VLSI 2011, p.148.[10] T. Grasser et al., IRPS2010, p16. [11] M. Duan et al., IEDM 2013. [12] C. Liu, *et al.*, IEDM 2011, p.571. [13] S.F. W. M. Hatta, et al., TED2013, 1745. [14] Z. Ji, et al., IEDM 2013. [15]. S. Pae et al., TDMR2008, p519.

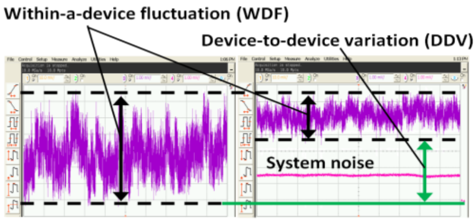

Fig. 1 The charging-discharging induced within-a-device fluctuation (WDF) under Vg=-1.4 V and the device-to-device variation (DDV) of WDF. The system noise also is shown.

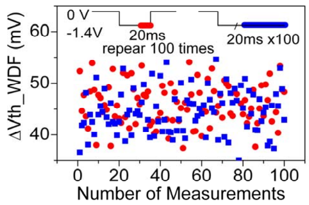

Fig.2 The WDF (see Fig. 1) measured within a time window of 20 ms. The insets give the Vg waveforms. '●' were obtained by repeating the short measurements 100 times. '■' were obtained from a continuous test with 100 sequential 20 ms windows.

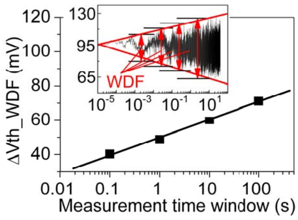

Fig. 3 Increase of WDF with the time window, tw. This device was stressed first for 1000 sec under Vg=-1.4 V and 125 °C to ensure that there is little changes in the number of defects during the subsequent measurement. A larger tw allows capturing slower traps, leading to the increase of WDF.

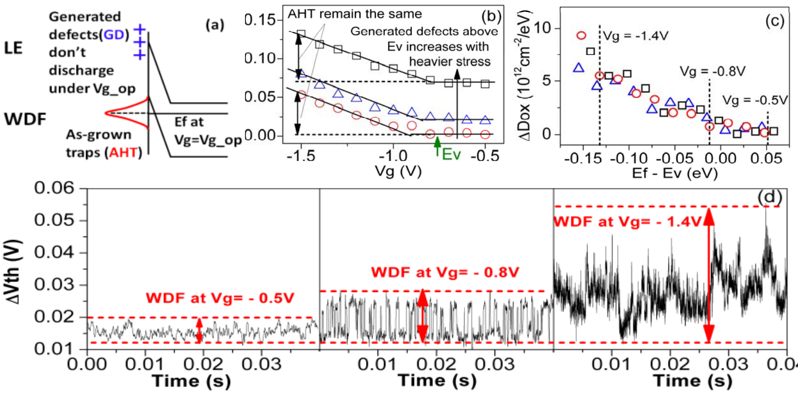

Fig. 4 (a) As-grown hole traps (AHT) near the interface Ef cause WDF and the traps above Ef will not discharge under Vg_op and cause the 'LE' in Fig. 5a. (b) Two groups of defects: The AHT are located below Si Ev and does not increase with stress; The generated defects (GD) are above Ev and increases with stress. (c) The energy distribution of hole traps: As Ef moves lower from Ev, hole trap density, ΔDox, rises, resulting in a larger WDF for more negative Vg in (d).

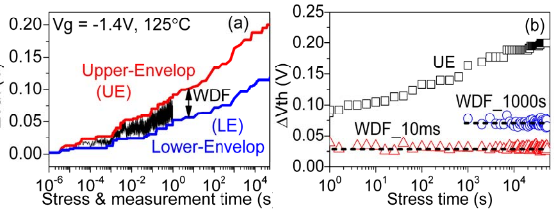

Fig. 5 (a) The ΔVth measured continuously under Vg_op=-1.4 V. Apart from the WDF, there are defects that do not discharge under a given Vg, giving rise to the lower, 'LE'. The upper-envelope, 'UE', equals to WDF+LE. The raw data also are presented for an initial 1 sec. (b) The WDF measured at a **fixed tw** of 10 ms and 1000 sec, taken at different stress time, does not increase with stress time. Even with tw=1000 sec, WDF substantially underestimates UE.

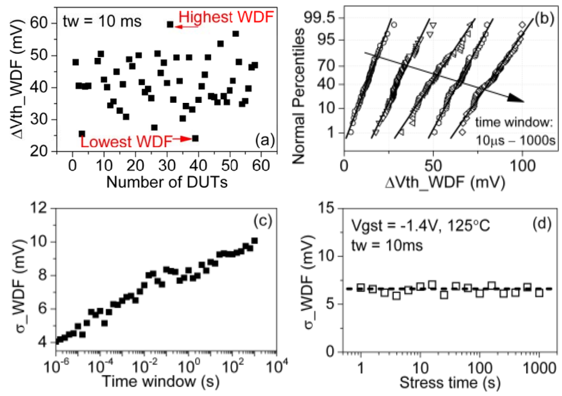

Fig. 6 (a) Device-to-device variation (DDV) of WDF. Each point represents one device and was obtained from the lower dashed line in Fig. 5(b). (b) The DDV distribution of WDF. The lines are fitted with the Gaussian distribution. The standard deviation of DDV increases with tw (c), but not with stress time (d).

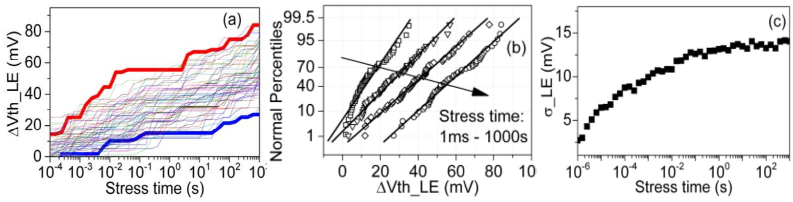

Fig. 7 (a) The DDV of LE. Each curve represents one device (see Fig. 5a). Two thick lines represent the highest and lowest LE at 1000 sec. (b) Distribution of LE. The lines are fitted with the Gaussian distribution. (c) The standard deviation of LE increases with stress time, in contrast with the σ_WDF in Fig. 6(d).

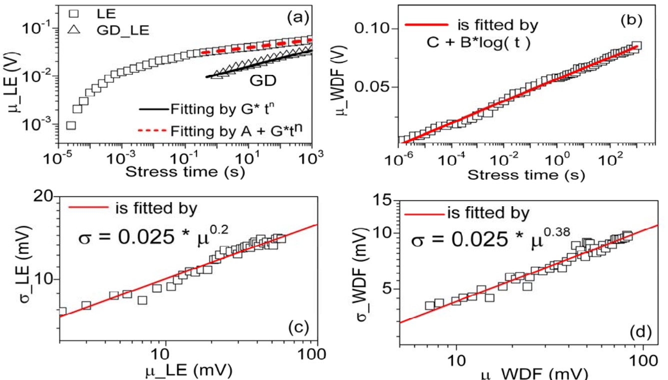

Fig. 8 (a) The average of LE, μ_LE, follows the As-grown-Generation (AG) model in Fig. 9. The generated defects (GD) follow a power law and was separated from the as-grown defects experimentally. (b) The average of WDF, μ_WDF, follows a linear relation with log(tw). The relation between σ and μ is shown for LE (c) and WDF(d). The lines are fitted.

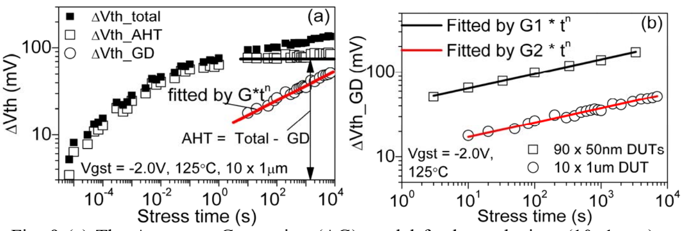

Fig. 9 (a) The As-grown-Generation (AG) model for large devices (10×1 μm): ΔVth=A+Gt^n. The filling of as-grown hole trap saturates and 'A'=constant for > ~1 sec. (b) The generated defects (GD) follow a power law for both large and small devices. They were separated from 'A' experimentally. The details of the separation procedure are given in ref. 14. The lines are fitted.

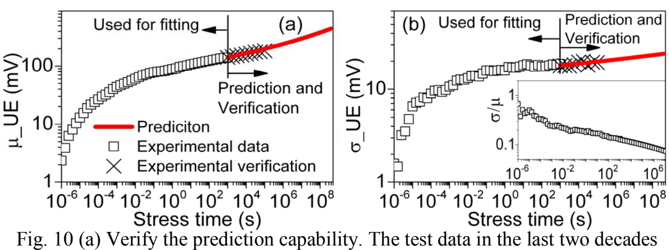

Fig. 10 (a) Verify the prediction capability. The test data in the last two decades ('×') were not used for fitting and agree well with the prediction (red lines) for both μ (a) and σ (b). The inset of (b) shows σ/μ reducing with stress time.

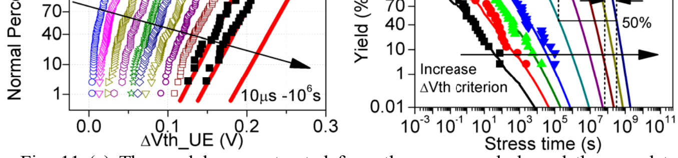

Fig. 11 (a) The model was extracted from the open symbols and then used to predict the distribution two decades ahead. The prediction (solid red lines) agrees well with the test data ('■'). (b) The distribution of lifetime-induced yield for different ΔVth(lifetime) criteria.

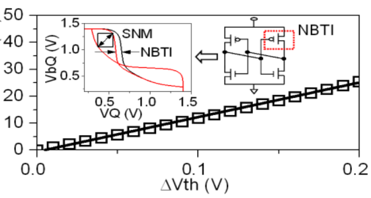

Fig. 12 Conversion of ΔVth to ΔSNM/SNM for a 6T SRAM, based on simulation for a 45 nm technology. One pMOSFET was subjected to NBTI, while the other was not.