Experimental Characterization of Random Telegraph Noise and Hot Carrier Aging of Nano-scale MOSFETs

AZRIF BIN MANUT

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

November 2020

Acknowledgements

Firstly, I owe my deepest gratitude to my principle supervisor, Professor Jian F. Zhang. Without your guidance, encouragement and unconditional support throughout my entire research project, this thesis would not have been made possible.

I also would never have made it this far but for the unwavering supervision, encouragement and support given by Dr. Zhigang Ji. I am forever indebted to him. Words could not express my gratitude to him.

I must also thank my dear (ex-)colleagues in LJMU; Dr. Gao Rui, Dr. Meng Duan, Dr. Jigang Ma, Dr. Zheng Chai, Dr. Brahim Benbakhti and Ms. Mehzabeen Mehedi. Thank you all.

Last but not least, I send my deepest thanks to my loving wife, Khairul Nisha Mohd Kharuddin and also to my parents; Kelthom Ibrahim and Manut Haji Yusoff. Thank you for being there always all throughout my journey. And, to my children, Anis Sofia, Fahim and Imran thank you for the support and love.

Abstract

One of the emerging challenges in the scaling of MOSFETs is the reliability of ultra-thin gate dielectrics. Various sources can cause device aging, such as hot carrier aging (HCA), negative bias temperature instability (NBTI), positive bias temperature instability (PBTI), and time dependent device breakdown (TDDB). Among them, hot carrier aging (HCA) has attracted much attention recently, because it is limiting the device lifetime. As the channel length of MOSFETs becomes smaller, the lateral electrical field increases and charge carriers become sufficiently energetic ("hot") to cause damage to the device when they travel through the space charge region near the drain.

Unlike aging that causes device parameters, such as threshold voltage, to drift in one direction, nano-scale devices also suffer from Random Telegraph Noise (RTN), where the current can fluctuate under fixed biases. RTN is caused by capturing/emitting charge carriers from/to the conduction channel. As the device sizes are reduced to the nano-meters, a single trap can cause substantial fluctuation in the current and threshold voltage.

Although early works on HCA and RTN have improved the understanding, many issues remain unresolved and the aim of this project is to address these issues. The project is broadly divided into three parts: (i) an investigation on the HCA kinetics and how to predict HCA-induced device lifetime, (ii) a study of the interaction between HCA and RTN, and (iii) developing a new technique for directly measuring the RTN-induced jitter in the threshold voltage.

To predict the device lifetime, a reliable aging kinetics is indispensable. Although early works show that HCA follows a power law, there are uncertainties in the extraction of the time exponent, making the prediction doubtful. A systematic experimental investigation was carried out in Chapter 4 and both the stress conditions and measurement parameters were carefully selected. It was found that the forward saturation current, commonly used in early work for monitoring HCA, leads to an overestimation of time exponents, because part of the damaged region is screened off by the space charges near the drain. Another source of errors comes from the inclusion of as-grown defects in the aging kinetics, which is not caused by aging. This leads to an underestimation of the time exponent. After correcting these errors, a reliable HCA kinetics is established and its predictive capability is demonstrated.

There is confusion on how HCA and RTN interact and this is researched into in Chapter 5. The results show that for a device of average RTN, HCA only has a modest impact on RTN. RTN can either increase or decrease after HCA, depending on whether the local current under the RTN traps is rising or reducing. For a device of abnormally high RTN, RTN reduces substantially after HCA and the mechanism for this reduction is explored.

The RTN-induced threshold voltage jitter, ΔV_{th}, is difficult to measure, as it is typically small and highly dynamic. Early works estimate this ΔV_{th} from the change in drain current and the accuracy of this estimation is not known. Chapter 6 focuses on developing a new 'Trigger-When-Charged' technique for directly measuring the RTN-induced ΔV_{th}. It will be shown that early works overestimate ΔV_{th} by a factor of two and the origin of this overestimation is investigated.

This thesis consists of seven chapters. Chapter 1 introduces the project and its objectives. A literature review is given in Chapter 2. Chapter 3 covers the test facilities, measurement techniques, and devices used in this project. The main experimental results and analysis are given in Chapters 4-6, as described above. Finally, Chapter 7 concludes the project and discusses future works.

List Of Abbreviations

Abbreviation	Signification
CP	Charge Pumping
C-V	Capacitance-Voltage
CVS	Constant Voltage Stress
EOT	Equivalent Oxide Thickness
HCA	Hot Carrier Ageing
HfO_2	Hafnium Oxide
IL	Interfacial Layer
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
NBTI	Negative Bias Temperature instability
PBTI	Positive Bias Temperature instability
RTN	Random Telegraph Noise
SILC	Stress Induced Leakage Current
SiON	Silicon Oxide Nitride
SRAM	Static Random Access Memory
TDDB	Time Dependent Dielectric Breakdown
TDV	Time-Dependent Variability
TWC	Trigger-When-Charged
WDF	Within-Device-Fluctuations
VSST	Voltage Step Stress Technique

List Of Symbols

Symbol	Description	Unit
C_{ox}	Oxide capacitance	F
E_f	Fermi level	eV
g_m, G_m	Transconductance	S
I_d/I_{ds}	Drain current	A
L	Mask channel length	µm
L_D	Debye length	cm
N_A	Substrate doping density	cm^{-3}
N_{it}	Interface states density	cm^{-2}
n_i	concentration in Si substrate	cm^{-3}
q	One electron charge	C
R	Feedback resistance	Ω
T	Temperature	^oC
V_d/V_{dd}	Drain voltage	V
V_{fb}	Flat band voltage	V
V_g	Voltage applied on the gate	V
V_{th} / V_t	Threshold voltage	V
ΔV_{th}	Threshold voltage shift	V
W	Mask channel width	µm

List of Figures

Figure 1.1 Moore's Law shows Doubling of Transistors vs Year [1]. 1

Figure 2.1 Locations and names of charges associated with thermal oxidized silicon [40]. 9

Figure 2.2 Schematic description of the R-D model used to describe the BTI phenomenon. 13

Figure 2.3 Illustration of two-staged model, which is the HDL model for a switching oxide trap coupled to the creation of a dangling bond at the interface. Figure from [63]. 15

Figure 2.4 Effects of hot carrier injection. 17

Figure 2.5 Cross-section of nMOSFET during hot carrier injection process. The figure is from [72]. 19

Figure 2.6 Random telegraph signals in a small MOSFET measured at the indicated gate voltages. Active device area is 0.4 μm^2, V_{ds}=4 mV, T=293 K [86]. 21

Figure 2.7 Time trace of RTN for typical 2-level RTN despite slight fluctuation at the base current. 22

Figure 2.8 Schematic representation of two-level RTN 23

Figure 2.9 Single-level RTN (a) nMOSFFET (b) pMOSFET [91]. 24

Figure 2.10 Time trace of complex RTN. (a) Four states of RTN originated from two traps can be observed, (b) multiple states of RTN from multiple traps can be observed [91]. 25

Figure 3.1 (a) Block diagram of the conventional measurement system. (b) picture of the Cascade probe station cable to connect the testing device and circuit and (c) connection of pulse generator and oscilloscope. 29

Figure 3.2 Some of the equipment for the measurement consists of Pulse Generator, Oscilloscope and Voltage source. 31

Figure 3.3 Working screen of Keithley 4200 Advanced Parameter Analyser. 32

Figure 3.4 Carbolite CTF 12/100/900 Ceramic Tube Furnace 33

Figure 3.5 The threshold voltage V_t is extracted by the g_m-max extrapolation method. 36

Figure 3.6 the threshold voltage V_t is extracted by the constant current method. 36

Figure 3.7 Equivalent circuit of a MOS structure 38

Figure 3.8 Configuration of spilt C-V measurement technique to obtain (a) $C_{gate\text{-}channel}$ against the Gate voltage and (b) $C_{gate\text{-}bulk}$ against the Gate voltage. 40

Figure 3.9 Measurement of spilt CV profiles (a) separate C-V measurement were measured and (b) the final profile acquired by the combination of the C_{gc} and C_{gb} measurement. 40

Figure 3.10(a) Traditional NBTI test sequence (b) The 2nd order OTF measurement sequence 45

Figure 3.11 The measurements of n^{th} and n-1^{th} I_d, combined with the transconductance $g_{m(n)}$, can provide the threshold voltage shift, ΔV_t of n^{th} and n-1^{th} measurement points..... 46

Figure 3.12 A typical result measured with pulse technique. a) A screenshot of V_g and V_{out} acquired by the oscilloscope. V_g was supplied by pulse generator. b) The extracted TC curve with 5 µs pulse edge time under V_d=25 mV; both up and down edges can be used to obtain I_d-V_g curve [13]. 47

Figure 3.13 Schematic set-up for the pulse I_d-V_g technique proposed by [118]..... 48

Figure 3.14 Schematic of our modified pulse I_d-V_g. 50

Figure 3.15 The connection of op-amp circuit without connecting to a device for calibration. (a) the circuit when the 10 kΩ resistor is grounded. (b) the connection when the 10 kΩ resistor is connected to the pulse generator. (c) and (d) are the measured I_d obtained from configuration (a) and (b) respectively [119]..... 52

Figure 3.16 Result of pulse measurement calibration with DUT (90 nm x 70 nm) connected by repeating the same measurement many times. (a)waveform of pulse V_g used and (b) shows the measurement variation at a constant voltage..... 53

Figure 3.17 Schematic diagram of measurement set-up for DMP technique. 54

Figure 3.18 Trigger mechanism on Channel 2 and Channel 3 of oscilloscope. 55

Figure 3.19 Steps involved in measurement..... 55

Figure 3.20 Flowchart of the process 56

Figure 3.21 Fresh IV plots from a device under test with V_g from +0.45 V to +1.05 V with 25 mV between the each V_g. 57

Figure 3.22 Raw WDF signal captured as each level of DMP 58

Figure 3.23 Triggering level after raw WDF (a) shows correct trigger level setting while (b) show trigger levels were far from the Max value. 59

Figure 3.24 Trigger point selected at each V_g level. First point is maximum value from WDF/RTN signal and subsequent points is trigger level set. (a) Sampling rate = 10 MSa/s and (b) 100 MSa/s. 10 seconds interval is because the time allocated for the trigger settings comply with pre-set condition. 60

Figure 3.25 I_{d0} measurement at true V_g pre-set at pulse generator (a) shows gate voltage signal and 'red box' is where value of true V_g is measured and (b) shows drain current signal and 'red box' is where value of I_{d0} is measured. Inset is V_g signal (in (a)) and I_{d0} signal (in (b)). 62

Figure 3.26 Comparison of the fresh IV with IV at discharge. 63

Figure 4.1 A comparison of Hot Carrier Aging (HCA) with BTIs reported by early works. (a) and (b) are re-plots of data from refs. [63] and [125], respectively..... 67

Figure 4.2 (a) Downscaling L increases HCA. The stress was at 125 ºC for 1000 sec. (b) comparison of HCA and BTIs for L=27 nm used in this work. Stresses were under the same $|V_g|$, with $V_d=V_g$ for HCA and $V_d=0$ for PBTI and NBTI. (c) A comparison of their recovery under $V_g=V_d=0$..... 68

Figure 4.3 HCA under $V_g=V_d$ is more than HCA under $V_g=V_d/2$ for L=27 nm. 69

Figure 4.4 The access nMOS in a SRAM during read-0 has a HCA duty factor of ~50%. When transistor flips from OFF to ON, or vice versa, HCA happens. 69

Figure 4.5 Variation of time exponent, '*n*' (the line slope), with HCA time. A re-plot of data from ref. [63]. 70

Figure 4.6 Verification of predicting HCA under use-V_{dd}. The model was extracted from accelerated VSS test data given in (a) with $V_g=V_d$ rising from 1.3 to 1.7 V (For details, see Figure 4.14). The symbols in (b)-(e) were measured from 4 devices and not used for fitting. The lines in (b)-(e) are the predicted HCA. The lines in (f) and (g) were obtained from Figure 4.10 by converting ΔV_{th} to $\Delta I_d/I_d(V_g=V_d=0.9$ V). 71

Figure 4.7 (a) CSR wafer used in our investigation. The shiny parts are the devices. Devices with gate width of 900 nm and 90 nm are selected. (b) Location of device that has been tested. 73

Figure 4.8 Recovery of large device after HCA, NBTI and PBTI 74

Figure 4.9 (a) I-V curve for 50 DUTs. (b) Variation of V_{th0} for 50 DUTs 75

Figure 4.10 A physical model for interface-traps generation [8]. 76

Figure 4.11 Good agreement of I_{sub}/I_d for large and small device 77

Figure 4.12 I_{sub}/I_d does not represent HCA-stress well for nm-devices, as it has a device-to-device variation (DDV) at stress=0 (a) and its DDV does not correlate with that of HCA-induced $\Delta I_d/I_d$. 78

Figure 4.13 HCA is too low to establish kinetics reliably under use-V_{dd}=0.9 V and acceleration (e.g. 1.3 V) is needed. 79

Figure 4.14 (a) The time exponents under $V_g=V_d$ is smaller than under $V_g=V_d/2$. (b) The time exponent is insensitive to stress biases under $V_g=V_d$ 80

Figure 4.15 The AC and DC HCAs agree well when using equivalent stress time, i.e. time×Duty Factor (DF). The AC stress conditions are given in the format of 'AC_DF_Frequency' and $V_g=V_d=2$ V.81

Figure 4.16 Voltage-Step-Stress (VSS) technique for HCA. (a) One device was stressed for a time T and the stress $V_g=V_d$ was then stepped up. ΔV_{th} is plotted against linear (b) and log (c) 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)) through Eqs. 4.3-4.5. The dashed line has n=0.29 and m=9. 83

Figure 4.17 Although test data (**o** and □) show $(\Delta I_d/I_d\_F)<(\Delta I_d/I_d\_R)$, higher '*n*' for $\Delta I_d/I_d\_F$ leads to incorrect $(\Delta I_d/I_d\_F)>(\Delta I_d/I_d\_R)$ when extrapolating. $\Delta I_d/I_d\_F$ does not sense the defects above space charges. The 'Δ' in inset is calculated from ($At^{0.29}$-Constant), which fits well with $Bt^{0.34}$ (**black line**). Subtracting a constant from a real power law can give an 'apparent' higher '*n*'. I_d was measured under $V_g=V_d=0.9$ V..... 85

Figure 4.18 The forward and reverse ΔV_{th} measured under $V_d=0.1$ V agrees well..... 86

Figure 4.19 Relation of HCA-induced ΔV_{th} under $V_d=0.1$ V with $\Delta I_d/I_d$ under $V_g=V_d=0.9$ V. The open symbols are the mean of 50 small devices (90 × 27nm) and the filled symbols are the mean of 50 large devices (900 nm×27 nm)..... 86

Figure 4.20 Evaluation of lifetime vs. V_{dd} based on the model extracted from VSS technique as in Figure 4.16. 87

Figure 4.21 HCA of two W=90 nm devices shows large DDV. WDF, UE, and LE is 'within-a-device-fluctuation', the upper- and the lower- envelope. 88

Figure 4.22 (a) HCA kinetics for the mean of 50 of W=90 nm devices. UE, DC, and LE have different '*n*' (inset). Incorrect inclusion of an as-grown component, '**C**', gives an apparent lower '*n*' at short time. (b) The definition of UE, DC, and LE. DC is the average over 10ms. 89

Figure 4.23 (a) For L×W=27 nm × 90 nm, LE increases with HCA, but WDF does not. (b) The WDF_mean of 50 devices and its sigma do not increase with stresses. 90

Figure 4.24 (a) LE_F correlates with LE_R. (b) WDF_F does not correlate with WDF_R..... 91

Figure 4.25 Normal probability plot for $\Delta I_d/Id_0$ for different stress times..... 92

Figure 4.26 Different $V_g=V_d$ of 1,000s stress time..... 93

Figure 4.27 Statistics of LE DDV after different stress time (a & b) and voltage (c & d). The lines are fitted with the defect-centric distribution WDF (Eqs. 4.6 & 4.7). 95

Figure 4.28 (a) The mean of 50 90×27nm agrees well with one 900nm ×27 nm for VSS stresses. (b) Sigma versus mean. The fitted exponent is 0.55, agreeing well with Eq. 4.7. 96

Figure 4.29 Impact of DDV on use-V_{dd}. When $\Delta I_d/I_d$ reaches 10% at i×σ, the mean $\Delta I_d/I_d$, μ, of defect-centric distributions reduces for higher i (a). This in turn requires a lower use-V_{dd} (b). For the reverse: '■'--- HCA only and '●' --- HCA and RTN/WDF. 97

Figure 5.1 A comparison of RTN/WDF in a fresh device with that after stressing for 10 s and 1000 s (a) at constant V_g = 0.85 V and (b) at constant V_g = 0.9 V 100

Figure 5.2 Amplitude of RTN/WDF at two V_g level; at V_g = 1.0 V (symbol ■) and at V_g = 0.9 V (symbol ●). (45 nm process) 101

Figure 5.3 Typical fresh devices with (a) and without (b) clear RTN. 103

Figure 5.4 The device-to-device variation of RTN/WDF amplitude. The devices were fabricated by (a) 45 nm and (b) 22 nm processes. The dashed line is the average RTN/WDF for 50 devices (45nm) and 24 devices (22nm). In (a), the 'A1' and 'A2' mark two devices of RTN/WDF close to the average and 'O1' and 'O2' mark two outliers. In (b), 'O3' and 'O4' mark two outliers. The results for these six devices are given as representatives and their fresh RTN/WDF are marked out by the red '●'. 104

Figure 5.5 Typical impact of HCA on the devices of RTN/WDF close to average, marked as 'A1' and 'A2' in Figure 5.4(a). RTN/WDF can either increase by 24% (a)&(b) or decrease (c)&(d) after HCA. 107

Figure 5.6 I-V measurement for device marked 'A2' before and after HCA stress. 108

Figure 5.7 Typical impact of HCA on an outlier, marked as 'O1' in Figure 5.4(a). (a) is fresh and RTN/WDF reduces by 66% after HCA (b). (c) shows that RTN/WDF amplitude returns to its fresh level after an anneal at 400 °C. 108

Figure 5.8 The tests were similar to those in Figure 5.7, but the device 'O4' was fabricated by a 22 nm processes and stressed under V_g=V_d=2 V for 1 ks. 109

Figure 5.9 A schematic illustration of HCA-induced de-sensitization of a critical trap. (a) shows that a critical trap is at the location where the current density peaks, causing abnormal high RTN/WDF before HCA. (b) shows how a change of current distribution after HCA can reduce the current

density under this trap, de-sensitizing it. This diagram is used to highlight the possible change of current distribution before and after HCA. It does not mean that current path is always strongly localized. 110

Figure 5.10 The impact of HCA on a device of outlier RTN/WDF, marked as 'O2' in Figure 5.4(a). (a) & (b) shows that RTN/WDF reduces by 65% after HCA. (c) shows that RTN/WDF remains low after an anneal at 400 ºC, supporting defect loss. 112

Figure 5.11 I-V measurement for device marked as 'O2' for (a) Fresh measurement compared to after HCA (V_g=V_d=2.2 V) and (b) comparison between fresh measurement and after annealing process. 113

Figure 5.12 The tests were similar to that in Figure 5.10 but the device 'O3' was fabricated by a 22 nm processes. 113

Figure 6.1 (a-b) A comparison of ΔV_{th} evaluated by $\Delta I_d/g_m$ (□) with ΔV_{th} directly measured at V_g~V_{th} (*dashed line*). The $\Delta V_{th}=\Delta I_d/g_m$ with V_g~V_{dd} (region A) is ×2 of the real ΔV_{th}. (c) I_{do} measurement captured the impact of RTN on device 90 nm x 27 nm 118

Figure 6.2 Example of RTN/WDF signals. (a) not analysable RTN signals and (b) shows complex WDF. 119

Figure 6.3 A schematic illustration of the impact of ΔI_d and ΔV_{th} on timing: (a) circuits and (b) waveform. V_{out} switches when $V_g \approx V_{th}$, which is delayed by a lower charging current, I_d-ΔI_d, supplied through the transistor 1 and a higher $V_{th}=V_{th0}+\Delta V_{th}$ of the transistor 2. 120

Figure 6.4 Comparison between system noise and true signal measured at V_g=-0.60 V. 122

Figure 6.5 (a) As-grown hole traps reduces for lower $|V_g|$ and (b) WDF probes traps near E_f. 124

Figure 6.6 The 'Trigger-When-Charged (TWC)' technique. (a) Test procedure: After a stabilization period, the RTN-induced ΔI_d is monitored under $V_g=V_{dd}$ and the upper envelope (UE) is determined. The trigger-level for subsequent pulse (3 μs) I_d-V_g (p-IV) is then set just below UE to measure ΔV_{th}. 50 p-IVs were measured in (b) and their average is given in (c). The TWC p-IV captures the RTN-induced ΔV_{th}, while the Traditional p-IV at pre-set time often misses the charge and is inapplicable. 127

Figure 6.7(a) Test configuration for "Trigger-When-Charged (TWC)" measurement technique. A high-speed operational amplifier based circuit is used to convert I_d to V_{out} that is connected to both channels 2 and 3. The "Trigger out" of the oscilloscope is connected to the "External trigger in" of

the pulse generator. (b) The V_{out} fluctuation is captured by both channels 2 and 3, as they are physically connected. (c) A screen-shot of the TWC p-IV measurement waveform. Channel 2 keeps its fine scale for accurate triggering, while channel 3 is switched to a coarse scale to capture the whole "TWC" p-IV. (d) A screen-shot of the traditional p-IV measurement at a preset time, where the trapped charge is missed. 132

Figure 6.8(a) Early works estimated RTN-induced ΔV_{th} from $\Delta I_d/g_m$ at V_{dd}=0.9 V (Point 'B'), rather than directly measuring it at $V_g=V_{th}$ (Point 'A'). The two insets are enlarged p-IV at the two points. The black p-IV is reference and the blue p-IV is the TWC p-IV. (b) The poor correlation between $\Delta I_d/g_m$ at V_{dd} and ΔV_{th} at $V_g=V_{th}$. Each point was taken from a different device. The dotted lines mark the mean values..... 133

Figure 6.9 Examples of the device specific dependence of the apparent ΔV_{th} on the sensing V_g, $V_g$sense. (a)-(e) were obtained from five different devices. The apparent ΔV_{th} at a $V_g$sense was obtained from the shift of TWC p-IV from the reference at $V_g$sense. The ΔV_{th} is normalized against its value at $V_g$sense=V_{th}. As the lowest $|V_g$sense| is close to V_{th}, the data starts from ~1 in all devices. 136

Figure 6.10 A schematic illustration of different impacts of traps at different locations on a device at threshold condition. The current can follow a percolation path under $V_g=V_{th}$. The trap in green corresponds to the device in Figure 6.9a: it is away from the critical current path, so that it only has a small effect on the device at V_{th}. The trap in red corresponds to the device in Figure 6.9b: it is on top of the current critical path and has a large effect on the device at V_{th}. 136

Figure 6.11 (a) The impact of sensing V_g on the apparent ΔV_{th} of a large device of 3×1 μm. The deep level traps were filled under V_g=-0.9 V for 100 sec before the measurement. ΔV_{th} is normalized against its value at $V_g=V_{th}$. (b) The stochastic variation of 135x27 nm devices. Each line is from one device. The symbols are the average. (c) Dependence of σ and σ /μ on the sensing V_g..... 139

Figure 6.12(a) The cumulative distribution of ΔV_{th}. The symbols are test data and the lines are fitted with the Defect-Centric model that assumes the number of traps per device following the Poisson's distribution and the ΔV_{th} induced by a trap following exponential distribution. (b) Standard deviation versus mean. Lines show that the data follow the prediction of Defect-Centric model well with a power exponent of 0.5. The different pairs of (μ,σ) are obtained by varying the time window of "I_d monitor" from 10 μs to 100 sec in Figure 6.6a. 141

List of Tables

Table 1.1 The impact of the scaling down geometry of MOSFET on $|\Delta V_{th}|$ [38] 4

Table 3.1 List of the devices used in this thesis 34

Table 3.2 Summary of the settings and type of data captured during the test. 61

Table 5.1 Details of the sample of devices used 105

Table of Contents

Acknowledgements i
Abstract ii
List Of Abbreviations i
List Of Symbols i
List of Figures vii
List of Tables xiv
Table of Contents xv
1 Introduction 1
1.1 Motivation 1
1.2 Research Concerns and Rationale 3
1.3 Objectives of this project 5
1.4 Structure of the thesis 5
2 Literature Review 7
2.0 Introduction 7
2.1 Bias Temperature Instability 8
2.2 Modelling of Bias Temperature Instability 12
2.2.1 Hole Trapping 12
2.2.2 Reaction-Diffusion Model 12
2.2.3 Two-stage Model 14
2.2.4 A-G Model 16
2.3 Hot Carrier Model and Characterization 16
2.3.1 Lucky-Electron Model 18
2.3.2 Energy Driven Model 20
2.4 Random Telegraph Noise 20
2.4.1 Single Defect responsible for RTN 22
2.5 Variability 25
2.6 Challenges 27
3 Experimental Facilities and Measurement Techniques 28
3.1 Introduction 28
3.2 Facilities 31
3.3 Devices 33
3.4 Conventional characterization and stress techniques 34
3.4.1 Conventional I_d-V_g technique 34
3.4.2 Conventional Capacitance-Voltage (C-V) technique 37
3.5 On-The-Fly (OTF) techniques 42
3.6 Pulse I_d-V_g techniques 46
3.6.1 Experimental Set-up 48
3.6.2 Calibration of Pulse Measurement System 50
3.7 Measurement Set-up for Nano-Devices 53
Set-up for Discharge-based Multi-Pulsed Technique (DMP) 53
3.8 Statistical Methods 63
3.8.1 Standard Deviation 63
3.9 Conclusions 64
4 Hot Carrier Aging and its Variation under Use-bias on Large and Small Devices 66

4.1 Introduction 66
4.2 Devices and Experiments 71
4.3 Prediction of HCA Lifetime on Large Device 76
4.3.1 Select the parameter representing stresses 76
4.3.2 Select HCA acceleration 78
4.3.3 DC versus AC 80
4.3.4 Application of VSST technique 81
4.3.5 Selecting parameter for extracting '*n*' 84
4.3.6 Prediction 87
4.4 Prediction of HCA for small devices 88
4.4.1 Characterizing HCA in nm-width devices 88
4.4.2 Statistic HCA 92
4.4.3 Impact on use-V_{dd} 96
4.5 Conclusions 97
5 Impact of Hot Carrier Aging on the Random Telegraph Noise and Within-Device-Fluctuation 98
5.1 Introduction 98
5.2 Device and Experiments 99
5.3 Results and Discussion 106
5.4 Conclusion 115
6 Development of a Technique for Directly Measuring RTN and BTI-Induced V_{th} Fluctuation under Use-V_{dd} 116
6.1 Introduction 116
6.2 Devices and measurement technique 121
6.2.1 Devices 121
6.2.2 Selection of test conditions 123
6.2.3 Test procedures 126
6.3 Results and discussions 132
6.4 . Conclusions 142
7 Conclusion and Future Works 144
7.1 Conclusions 144
7.1.1 Conclusions on the kinetics and prediction of HCA 144
7.1.2 Conclusion on impact of hot carrier aging (HCA) on random telegraph noise (RTN) 145
7.1.3 Conclusion on Trigger-When-Charged technique 146
7.2 Future Works 147
7.2.1 HCA and BTI Coupling 147
7.2.2 Defect losses 148
References 149
List of Publications 169
Journals 169
Conferences 169

1 Introduction

1.1 Motivation

Integrated circuits (ICs) were first invented by J. Kilby at Texas Instruments in 1958. According to the Moore's Law, the transistor's number per chip doubles every two years as shown in Figure 1.1 [1]. This law has been the driving engine for innovation in semiconductor industries for more than five decades. At present, the rise in density of transistors is accomplished by constant-field scaling of complementary MOSFETs (CMOS) [2]. The constant-field scaling keeps the electric-field constant thus making it possible to reduce device dimensions and power-supply voltage. This has led to numerous advantages such as increased density of MOSFETs on ICs, better performance, and less power consumption.

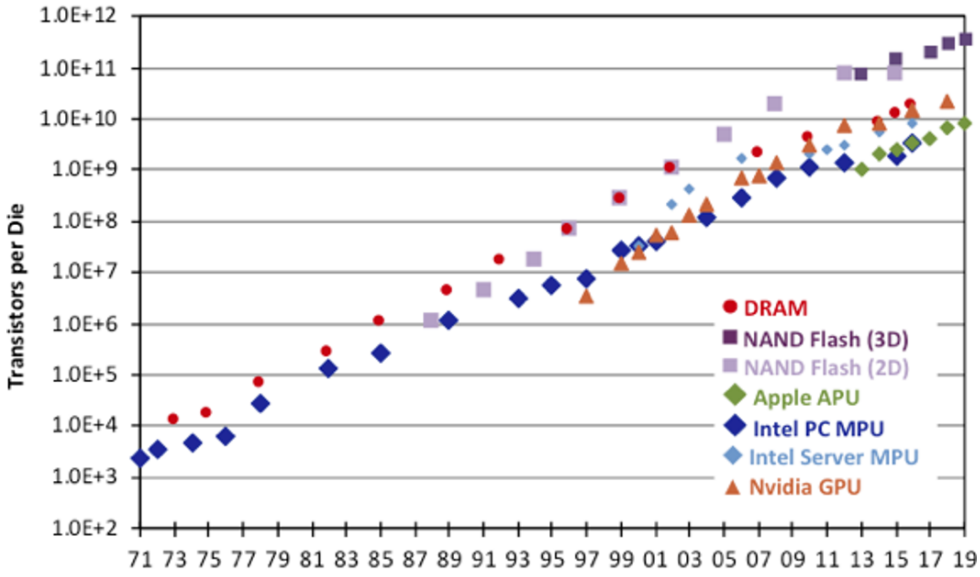


Figure 1.1 Moore's Law shows Doubling of Transistors vs Year [1].

On one hand, scaling of MOSFETs has played a major role in the expansion of the semiconductor industry for forty years [3],[2]. On the other hand, they also contribute to many reliability issues [4], [5]. In 1970s, reliability issues were concentrated on the contamination such as mobile ions and induced instability which were overcome with the introduction of cleanroom technology. As the down-scaling continued in the 1980s, the operation voltage of devices was reduced and maintained at 5 V. This resulted in a higher electrical field in the device. The lifetime of the nMOSFETs and pMOSFETs was limited by hot carriers and this became the main reliability issue [6].

The phenomenon called hot carrier injection is the injection of energetic carriers into the gate oxide [7]. Hot carrier injection contributes to the damage of gate oxide permanently through carrier trapping and interface trap generation, resulting in a shift in device parameters such as threshold voltage, sub-threshold slope, and transconductance [8], [9]. This is called hot carrier induced aging (HCA) degradation in MOSFETs [10].

The number of hot carriers created by impact ionization near the drain junction of pMOSFET's compared with the nMOSFET's of the same channel length, is 2-4 orders smaller because of the lower mobility of holes [11].

In year 2009 and 2011, International Technology Roadmap for Semiconductors (ITRS) identified Random Telegraph Noise (RTN) as a concern for static random access memory (SRAM) scaling. This is because the acceptable noise margins are becoming narrower due to increasing V_{th} variability [12]. In 2018, IEEE International Roadmap for Devices

and Systems: More Moore again highlighted RTN as near-term (2017-2024) device reliability challenges [13].

RTN is caused by single trap at gate dielectric capturing and emitting a carrier [14]. Its impact increases with the down-scaling of MOSFETs. As flash memory and SRAM typically use the minimum sized devices to achieve high density, they are especially vulnerable to RTN.

1.2 Research Concerns and Rationale

One of the emerging challenges in the scaling of MOSFETs is the reliability of ultra-thin gate dielectrics. Various sources can cause instabilities to the device, such as hot carrier injection, negative bias temperature instability (NBTI), positive bias temperature instability (PBTI), and time dependent device breakdown (TDDB) etc. A lot of studies have concentrated on NBTI [15]–[20] because it was limiting the lifetime of pMOSFETs [21]–[23]. HCA has been studied for more than 20 years [10], [24]–[26]. For modern CMOS technologies, HCA limits the device lifetime.

Another major challenge for sub-nanometre MOSFETs is statistical variability. It originates from the discreteness of charge and granularity of matter [27], [28]. Variability can be divided into device-to-device variability (DDV) and within-device-variability (WDF).

Sources of DDV in transistors can be categorized into time-zero variability (TZV) [27], [29]–[31], and time-dependent variability (TDV) [32]–[36]. TZV is directly related to the micro-fabrication process including random discrete dopants, line edge roughness, polysilicon granularity and oxide thickness fluctuations [37]. On other hand, TDV results from aging-induced defects. It is a product of a process of filling and generating new defects. TDV shows stochastic behaviour and varies with devices [38]. Researchers around the world have reported that both sources affect the reliability of MOSFETs.

With the device progressively scaling down, the number of defects within the dielectric decrease, but the impact from each single defect increases. This results in a Within-device-fluctuation (WDF) phenomenon.

Table 1.1 The impact of the scaling down geometry of MOSFET on $|\Delta V_{th}|$ [38]

FET-name and dimension (in µm)	wide		narrow		minimal	
	W	L	W	L	W	L
	10	0.1	0.2	0.12	0.11	0.1
number of carriers in channel at Vg=VT-200mV	15000		370		170	
number #Nit at a density DNit=1E11/cm^2	1000		24		11	
ΔDNit causing a ΔVT=50mV (in cm^{-2})	4.9×10^{11}		4.9×10^{11}		4.9×10^{11}	
makes a number Δ#Nlt	4900		120		50	
ΔVT caused by a single trapped carrier (at interface)	0.01mV		0.43mV		1.0 mV	

1.3 Objectives of this project

The objectives of this project are:

i. To study HCA and develop models for predicting the device lifetime.

ii. To investigate the relationship between RTN and HCA .

iii. To develop a direct measurement technique for RTN-induced ΔV_{th}.

1.4 Structure of the thesis

This thesis is organized as follows:

In Chapter 1, a brief background of MOSFETs reliability has been given. Challenges for current and future MOSFETs are also mentioned, together with the project's objectives.

In Chapter 2, a literature review on the model and characterization techniques of bias temperature instability (BTI) and hot carrier injection (HCI) are presented. Defects responsible for random telegraph noise (RTN) and challenges are also reviewed.

Chapter 3 comprises the experiment facilities and measurement techniques that have been used in this project.

Chapter 4 studies hot carrier aging and variation under use-bias.

Chapter 5 investigates the impact of hot carrier aging on RTN.

In Chapter 6, the Trigger-When-Charged method is proposed to directly measure RTN induced jitter on V_{th} of nano-devices.

In Chapter 7, the project is summarised and future work is discussed.

2 Literature Review

2.0 Introduction

Integrated circuit (IC) technology has had a great impact on our daily life for the past 60 years. The physical dimensions of MOSFETs have been successfully down-scaled to increase operation speed and density. This follows the Moore's Law. As the result, the channel length now reaches the nanometre range while the effective oxide thickness (EOT) is below 1 nm. This makes reliability an increasing concern. Product failures due to electronics circuit malfunctions or them not operating as intended can contribute huge losses to a company. When the problem is not detected during functional testing, it can cause gradual performance reduction and even total failure of the products. This happened to Intel Corporation's Cougar Point chipset in 2011. The chipset was utilized in Intel's latest Sandy Bridge's processor when they detected a potential reliability problem in the chipset's serial-ATA channels in about 5% of integrated circuits [39].

Instabilities can be caused by many sources, such as positive bias temperature instability (PBTI), negative bias temperature instability (NBTI), and hot carrier injection (HCI) etc. Among them, NBTI and HCI are becoming the harshest reliability issues because they are aggravated by the use of the nitrogen to prevent the boron penetration through the gate dielectric.

For nano-size MOSFETs, another major challenge is the fluctuations induced by random charge-discharge of traps in gate dielectric. Recently, fluctuations in nano-metre size has become a major concern for circuit design. Fluctuations are commonly observed as a jitter or random telegraph noise (RTN).

With the migration of semiconductor manufacturing technology node, the complexity of the circuit increases. Reliability of the circuits needs to be checked at design optimization stage before the fabrication starts. As an example, the operating voltage, V_{dd} of a circuit must be properly selected. Increasing the value of V_{dd} provides better switching speed and improves MOSFET drive current. But unfortunately it will also increase the degradation. Thus, a lifetime prediction model is important to evaluate a trade-off between reliability and performance.

2.1 Bias Temperature Instability

Bias temperature instability (BTI) can be divided into Negative Bias Temperature Instability (NBTI) and Positive Bias Temperature Instability (PBTI). NBTI can lead to shorter lifetime of pMOSFETs and has attracted much attention. PBTI, on other hands, has a significant impact on nMOSFETs.

BTI is originated from the creation or filling of traps in the gate dielectric. These extra charges in the gate dielectric will increase $|V_{th}|$ by imposing an inverted electric field in terms of gate voltage. $|I_g|$ also increases due to created traps that assist carrier hopping.

In 1980, B. E. Deal [40] suggested traps and charges can be classified into four categories. They are (1), Mobile ionic charges Q_m, (e.g. sodium ions) which are mobile within the oxide when under BTI stress condition; (2) Oxide trapped charges Q_{OT}, which are distributed inside the oxide layer and can be created by X-ray radiation or hot-electron injection; (3) Fixed oxide charges Q_f, located at or near the interface are immobile under an applied electric field; (4) Interface states N_{IT} and trapped charges Q_{IT} are located at the Si-SiO_2 interface with energy states within the silicon forbidden bandgap and able to exchange charges with silicon in a short time.

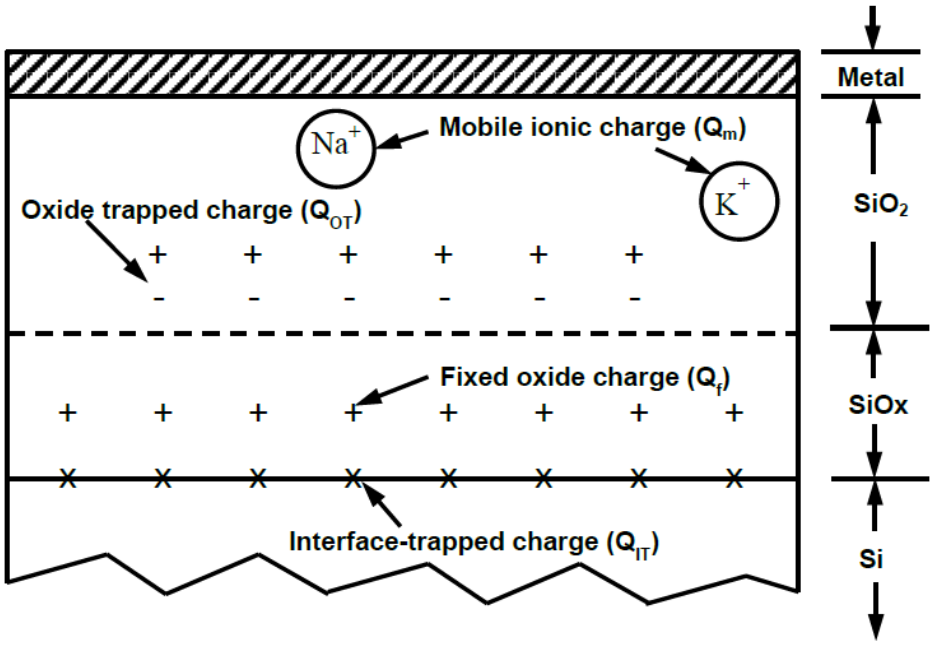


Figure 2.1 Locations and names of charges associated with thermal oxidized silicon [40].

Expected device lifetime is typically 10 years, which is too long to reach in a laboratory if using normal procedure. Thus, the accelerated BTI test is applied to shorten the time. The standard criterion of BTI lifetime is that $|\Delta V_{th}|$ should not exceed a certain level (typically 100 mV) after the device reached its lifetime. An accelerated BTI test can be done by applying a much higher V_g (severer stress) than operating condition and within an acceptable testing time (usually < 10^6 seconds). Then, using the time-evolution model, the device lifetime under operating voltage can be projected. BTI kinetics typically follow a power law against both stress time and overdrive voltage $|V_g\text{-}V_{th}|$. Joint Electron Device Engineering Council (JEDEC) [41] suggested that BTI kinetics can be described in Equation (2.1):

$$\Delta V_{th} = A \cdot (|Vg - V_{th}|)^m \cdot t_{str}^n \tag{2.1}$$

where t_{str} is BTI stress time, A, m, and n are fitting parameters from the accelerated test. However, when fast characterization was introduced, researchers realized that slow characterization failed to capture a small portion of BTI due to recovery. This makes the Equation (2.1) no longer relevant to predict BTI lifetime because the time exponent, n, is sensitive to the measurement condition.

The degradation of BTI can recover once the stress voltage is removed. This fast recovery during the measurement resulted in an underestimation of the total BTI degradation. The typical time using slow measurement ranges from milli to tens of seconds to measure one IV curve. Due to this fast recovery, the measurement result is highly dependent on BTI characterization speed. Fast characterization was introduced to overcome this problem. Researchers can now measure BTI kinetics within several microseconds by using ultrafast measurement.

BTI is not a newly discovered problem. It is one of the earliest instabilities reported for metal-oxide-semiconductor field effect transistors (MOSFETs). BTI is caused by a temperature accelerated degradation under bias. It manifests itself as a decrease of drain current, I_{ds}, and transconductance, g_m, and an increase of off current, I_{off}, and the magnitude of threshold voltage, V_{th}. Four possibilities that can happen during circuit operation are negative bias on pMOSFET, positive bias on pMOSFET, negative bias on nMOSFET, and positive bias on nMOSFET. Although there are four possibilities, most of the researchers focus on negative bias on pMOSFET because it is the most severe degradation condition.

BTI was first reported by Miura and Matukura in the year 1966 [42]. Detailed characterization of BTI was first showed by Bell Telephone Laboratories' researchers by using metal gate devices on 100 nm oxides and then stressed at 10^6 V/cm at 300 ºC [43]. They concluded that interface trap density, D_{it}, peaked in the lower half of the bandgap, increased with gate voltage and with time. Although equal amounts of interface state and positive charge generation occur for both n- and p-type silicon substrates, the net effect on threshold voltage, ΔV_t, is greater for p-FETs, because in this case the positive oxide charge and positive interface charge are additive [44]. This is the reason why NBTI is of more concern on p-MOSFET than n-MOSFET.

Various BTI models have been proposed, including Holes Trapping, Reaction-Diffusion (R-D) Model, 2-Stage Model, and the As-grown-Generation (AG) model.

2.2 Modelling of Bias Temperature Instability

2.2.1 Hole Trapping

Hole trapping is one of the factors responsible for BTI [20]. Hole traps can be generated through the interaction of oxides with either free holes or hydrogen. Several models have been proposed for hole traps, including bond-strain gradient, oxygen vacancy, and self-trapping of hydrogen on bridging oxygen [45], [46].

The trapped hole can be neutralized once the gate bias has been removed, leading to a rapid recovery of BTI. Suppression of this recovery is important in measurement of NBTI. As such, techniques such as On-The-Fly (OTF) and Ultra-Fast Single Pulse Measurement (UFSP) were introduced to suppress recovery [18], [47]–[50]. The OTF technique measures the degradation at stress gate bias, while the UFSP probes the degradation at threshold voltage level [51].

2.2.2 Reaction-Diffusion Model

Jeppson and Svensson first introduced the R-D model in the year 1977 [4] and it is the most popular model. This model explains the BTI effect in terms of electrochemical reaction at the SiO_2/Si interface which converts the precursors into interface states and releases a hydrogenous species, as shown in equation (2.2) [4]:

$$\equiv Si\text{-}H + hole \rightarrow \equiv Si\bullet + H \qquad (2.2)$$

In their model, the assumption was that the silicon interface contains a large number of defects. These defects are electrically inactive and can be activated through chemical reaction. It can be described that N_{it} generation (recovery) occurs in two distinct phases: (i) Fast generation (recovery) ascribed to spontaneous de-passivation (re-passivation) of Si-H (Si-) bonds by holes (hydrogen) at the interface; (ii) slow generation (recovery) rate-limited by hydrogen diffusion away from (towards) the interface[52].

M. A. Alam (2003) demonstrated that the numerical and analytical solution of the R-D model can provide an adequate framework to explain BTI [15]. This has brought the R-D model into people's attention again and has been a popular model for BTI in the past decade. The model can be described thus: the holes in the inversion layer interact with the Si-H bonds at Si-SiO_2 interface at elevated temperature. This create interface traps, Si+ and hydrogen species, as shown in Figure 2.2.

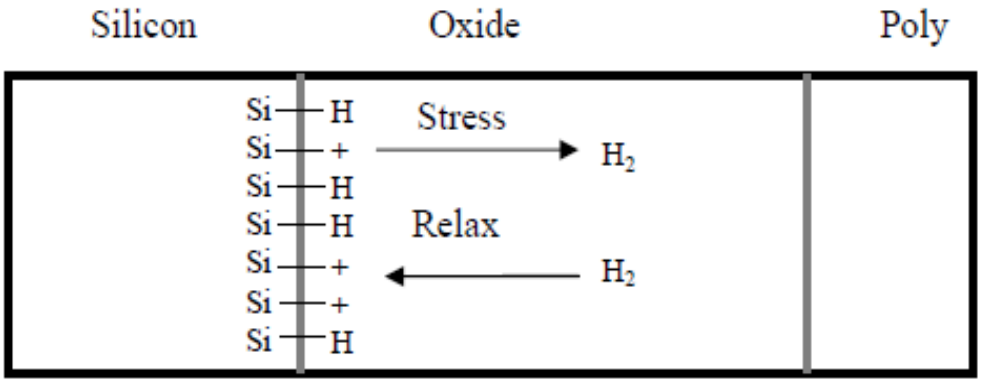


Figure 2.2 Schematic description of the R-D model used to describe the BTI phenomenon.

2.2.3 Two-stage Model

The two-stage model is based on the established properties of *E′* centres, which is based on the combination of interface trap and hole trapping/de-trapping effect [53]. It was developed based on the Harry-Diamond-Laboratories (HDL) model by T. Grasser et al. [53], [54]. Creation of *E′* centres from their oxygen vacancy precursors is suggested to occur via a multiphonon-field-assisted hole trapping mechanism. These models arose from the perceived failure of the pure R-D model (without hole trapping) in interpreting the ultra-fast recovery [55]. This model describes that the recoverable component is due to hole trapping in oxygen vacancies near the interface and the permanent component is due to the creation of interface traps (P_b-centres) [56]. It is suggested that this model is capable of explaining the hole trapping and de-trapping mechanism in NBTI degradation, particularly on the transformation of hole traps into a more permanent form [57].

Earlier works assumed that the degradation caused by NBTI is permanent. However, it was observed that NBTI degradation recovers considerably within 1 sec. Recovery or relaxation happens when the negative bias is removed from the gate [58]. Chen *et al* first referred to relaxation as dynamic NBTI in 2002. They concluded that the channel inversion layer disappeared when the gate bias polarity is reversed to positive or zero. This will interrupt the breaking of the Si–H bond due to a lack of holes, and at the same time, H moves back to the SiO_2/Si interface under the influence of positive gate voltage and passivates the Si dangling bond [59].

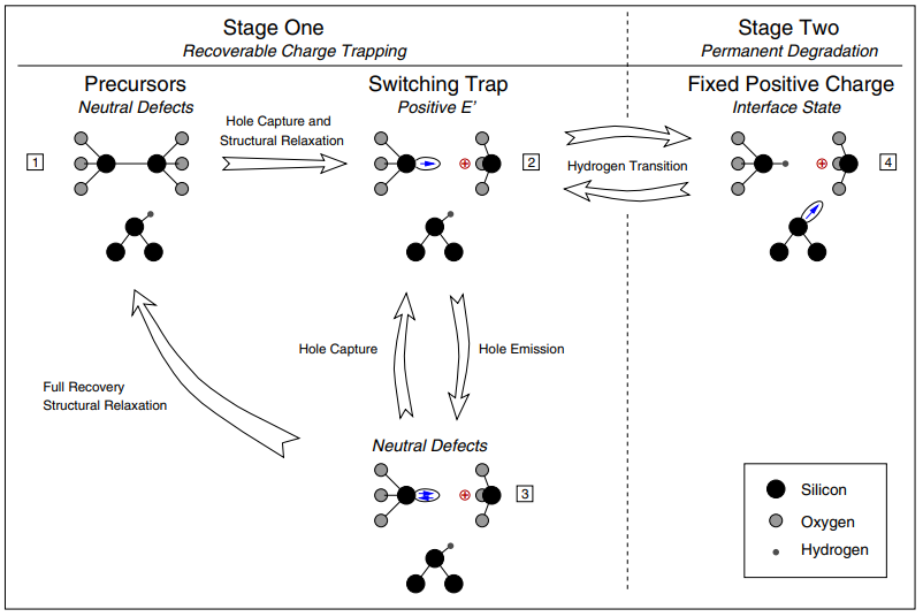


Figure 2.3 Illustration of two-staged model, which is the HDL model for a switching oxide trap coupled to the creation of a dangling bond at the interface. Figure from [63].

Cyclability is the main factor in this model (Figure 2.3). When the bond between the two silicon atoms captured an electron (emitted the hole), the bond does not fully created. This bond can easily lose an electron again. This act of switching traps by E' centre has been suggested by Lelis *et al* [60]. As illustrated in Figure 2.3, at state 1, a neutral precursor exists. Upon hole capture, the Si-Si bond breaks and this results in a positively charged E' centre in state 2. The positively charge E' center can attract the H. This resulted the hydrogen passivate a silicon dangling bond at the interface can move to the E' center. Thus can be locked in the positive charge as in state 4. In state 3, hole emission (electron capture) neutralizes the E' centre. In this state, two options exist; a hole can be captured again causing a transition to state 2, or the structure can relax back to its equilibrium configuration as in state 1.

2.2.4 A-G Model

In the A-G Model, NBTI originates from positive charge (PC) formation within the gate dielectric and interface states [51]. These positive-charges (PC) formation within the gate dielectric and the generation of interface states contribute to NBTI [49]. It is proposed by J. F. Zhang *et al* [18], [51], [61] that three types of PCs contribute to NBTI: As-grown Holes Traps (AHT), Cyclic Positive Charges (CPC) and Anti-Neutralization Positive Charges (ANPC). AHT is pre-existing defects in the device and has energy levels below the top edge of the silicon valence band, making it the easiest to neutralize, but hard to charge. CPC is energetically located within the band gap and can be repeatedly charged and discharged by alternating gate-bias polarity. ANPC has energy levels above the bottom edge of the silicon conduction band, so that they are easy to charge but hard to neutralize.

2.3 Hot Carrier Model and Characterization

For the last 20 years, BTI has been the most studied CMOS degradation mechanism. Recently, due to aggressive CMOS scaling, Hot Carrier aging (HCA) comes back to challenge device and circuit reliability [62]–[64]. Theories based on long channel devices are no longer enough to explain the device response due HCA. This is because the operation voltage can drastically change the energy of hot carriers. Different physical processes such as carrier injection into oxide, impact ionization, charge trapping and interface trap creation can be present and contribute differently to the degradation.

The injection of hot carriers from the channel to the gate oxide is the main cause for hot carrier degradation. The two factors that determine the carrier injection at any given location along the channel are, the electric field at the point of the injection and the concentration of the carriers in the channel. As the gate bias increases, the concentration of the carriers in the channel also increases. This results in electron and hole injection current increases. While in the sub-threshold region, the injection of the carriers into the gate oxide is negligible due to the concentration of carriers in the channel being low [65].

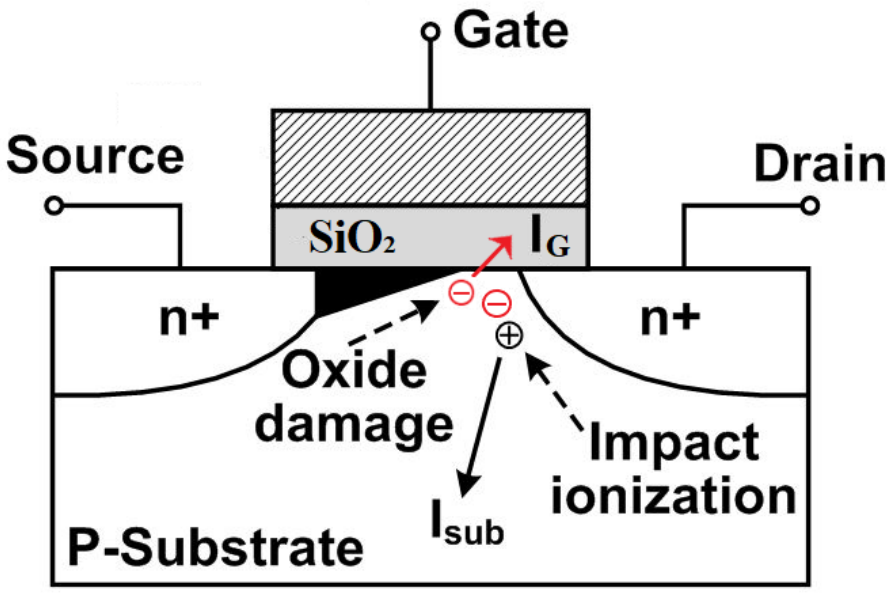


Figure 2.4 Effects of hot carrier injection.

In the 1950s, theoretical work on hot carrier injection and its effects on the devices had already started but the modelling was not started until 1980 [63]. The reason is mainly because there was an ongoing debate regarding the different mechanisms involved in hot carrier injection. For examples, what charges get injected and whether charge injection is needed to cause degradation of a device [64], [65].

It has been recognized that hot-carrier induced aging (HCA) degradation can pose a major threat to the scaling of device [62]. HCA affects electrical characteristics of MOSFETs even under nominal operating conditions, resulting in enhanced circuit failure [42-43]. HCA models have developed from phenomenological or empirical approaches [8], [68] to more sophisticated models to explain the rich physics behind this effect [62], [69]–[71], as detailed in the following.

2.3.1 Lucky-Electron Model

Early attempts linked the device life time to the electric field. Hot carrier can be demonstrated as in Figure 2.5. The large voltage drop across the pinch-off region results in a high lateral electric field close to the drain region. The carriers traversing this high field region accumulate considerably higher energies than the equilibrium thermal energy in the semiconductor lattice. These carriers with high energy are called ‘hot-carriers’. If the carriers gain enough energy, it can be injected into the gate oxide causing interfacial damage and will introduce instabilities in a MOSFET’s electrical characteristics. The damage rate is dependent on the lateral electric field. This phenomenon is explained by C. M. Hu *et al* [10] using the ‘Lucky Electron Model’.

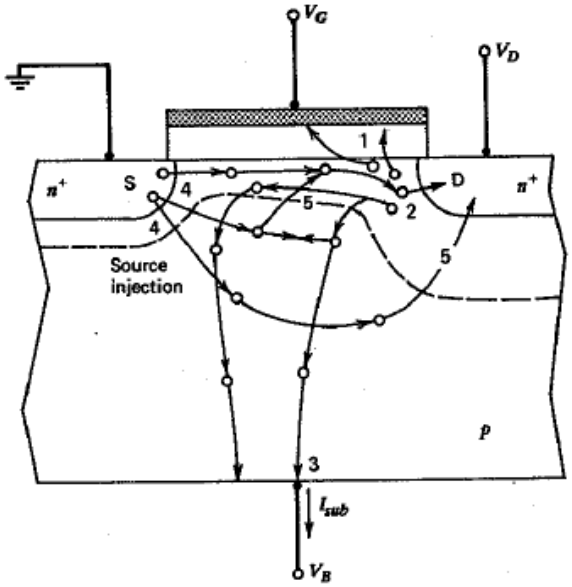


Figure 2.5 Cross-section of nMOSFET during hot carrier injection process. The figure is from [72].

The Lucky Electron Model (LEM) has been used for decades as the guiding principle of most industry standard model and projection methodologies associated with hot carriers. In this model, the fundamental concept can be explained by interaction of free electrons accelerated by an electric field until they collide with an atom. The collision will ionize the atom resulting in two free electrons. This is where the term ‘impact ionization’ is coming from.

2.3.2 Energy Driven Model

Another model for HCI is called the Energy Driven Model [51-52]. It was proposed in 2005 as a replacement for the Lucky Electron Model in a short channel regime. The driving force in this model is energy rather than peak lateral electric field as it is in LEM.

Understanding the damage mechanism as well its related reliability problems is the main concern. HCA is caused by injection of energetic carriers into the gate dielectric near the drain side. This is the area where defects are generated. Charge pumping techniques have been used to characterize the location of trap generation for both border traps (N_b) and interface traps (N_{it}) to understand the damage mechanism [75] [76]. Chun-Chang Lu *et al* [77] showed a modified charge pumping technique to better understand HCA reliability problems. It was shown that HCA stress causes interface trap generation through the whole channel and substantial border trap generation at the edge of the gate region.

2.4 Random Telegraph Noise

Since the days of vacuum electronics, Random Telegraph Noise (RTN) was already observed. It was often referred to as burst of popcorn noise [78][79]. RTN in MOSFET was reported by Ralls et al [80] in 1985 and followed by Uren et al [81] in 1986. Since then, RTN has been found in many different semiconductor devices including MOSFETs, flash memories, RRAM and LEDs [82], [83].

In 2009, for the first time, the International Technology Roadmap for Semiconductors (ITRS) identified Random Telegraph Noise (RTN) as a concern for static random access memory (SRAM) scaling due to increasing V_{th} variability including RTN [12]. This makes the acceptable noise margin narrower. This issue is mentioned again by ITRS in 2011 [84] and by IEEE International Roadmap for Devices and Systems: More Moore in 2018 [13].

RTN in its primary form can be described as random switching of a current between a high and a low state in a semiconductor device [85]. This dual state can be schematically represented in Figure 2.6. We can notice that for randomly distributed up and down times, the amplitude is normally fixed and can be easily calculated from time trace. RTN is normally caused by the emission and capture of charge carrier to and from the conduction channel. RTN has become the harshest reliability issue in the aggressively scaled CMOS technology. It is because the device size is so small that the trapping/de-trapping of one individual defect has a greater impact on the device performance [79].

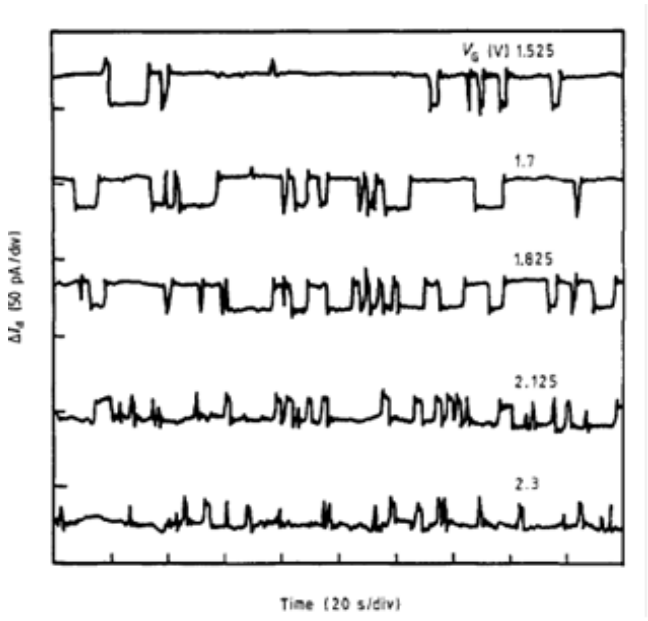


Figure 2.6 Random telegraph signals in a small MOSFET measured at the indicated gate voltages. Active device area is 0.4 μm^2, V_{ds}=4 mV, T=293 K [86].

RTN can be studied either in the frequency domain by applying a Fourier Transform or in the time domain by investigating the time trace directly [80], [87], [88]. In frequency domain, the signal exist within a given frequency band concerning a range of frequencies. While in time domain, it shows the changes in of a signal over a period of time. Fourier transform can converts a time function into an integral sine-waves of multiple frequencies. Figure 2.7 shows the time trace of an RTN.

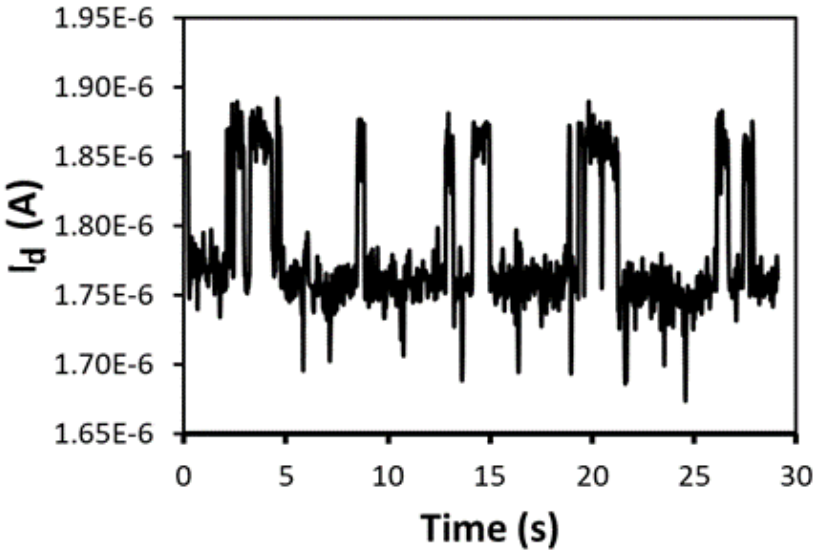


Figure 2.7 Time trace of RTN for typical 2-level RTN despite slight fluctuation at the base current.

2.4.1 Single Defect responsible for RTN

As the scaling down of feature size, a single defect is sufficient to shift threshold voltage and driving current in nano-scaled MOSFETs. It was reported by IBM in 2011 that a single RTN can make the device reach lifetime criteria [89]. The impact of single trap on I_d become larger as MOSFET's scaled down.

We could observe the random step of threshold voltage, V_{th} or/and drain current, I_d during charging and discharging of this single defect. RTN in its most basic form consists of the random switching of the current between a high and low state as depicted in

Figure 2.8. For the two-level RTN, it can be characterized by up time (t_u), down time (t_d) and amplitude of ΔI_D. For an adequate amount of time, the probability distribution functions of the switching time follows an exponential distribution.

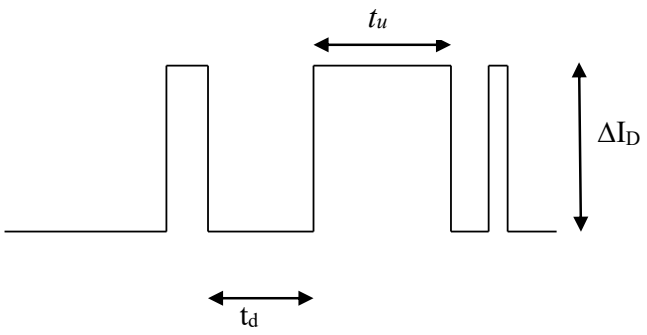


Figure 2.8 Schematic representation of two-level RTN

It was reported by Liu et al [90] that hot carrier stress has no impact on RTN. This is true for normal RTN. For abnormal/large RTN, we have observed that hot carrier stress has an impact on the magnitude of amplitude and also in some cases completely suppresses RTN. This interesting phenomenon will be investigated thoroughly in Chapter 5.

We can express ΔV_{th} due to RTN as

$$\Delta V_{th} = \frac{q}{L_{eff} W_{eff} C_{ox}} \tag{2.3}$$

where q is the elementary charge, L_{eff} is the effective channel length, W_{eff} is the effective channel width and C_{ox} is the gate capacitance per unit area [87].

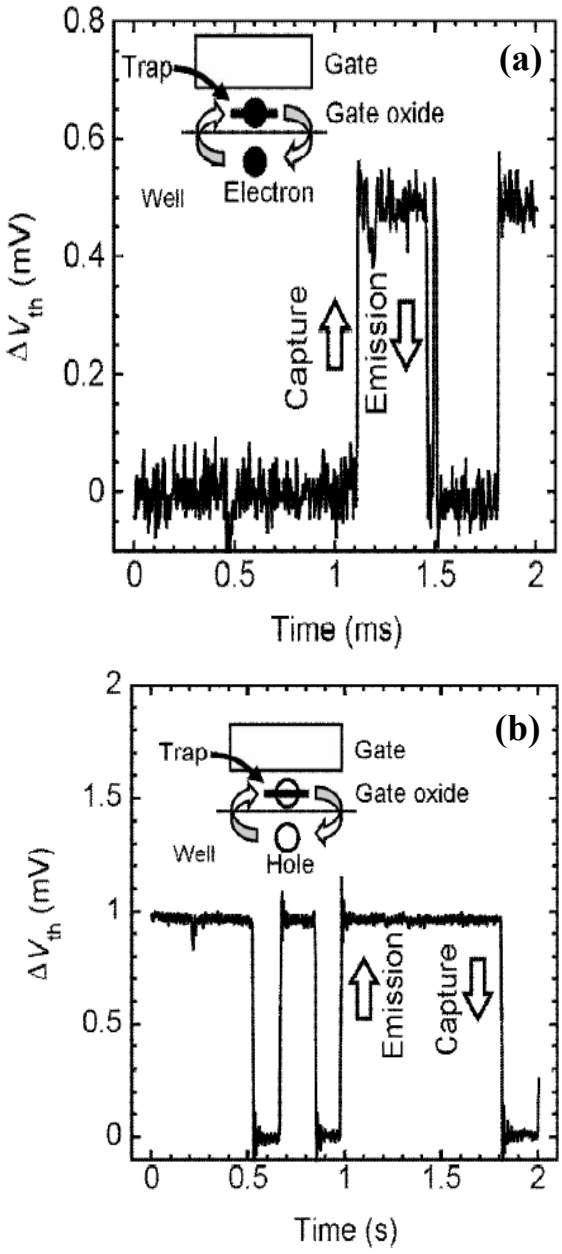


Figure 2.9 Single-level RTN (a) nMOSFFET (b) pMOSFET [91].

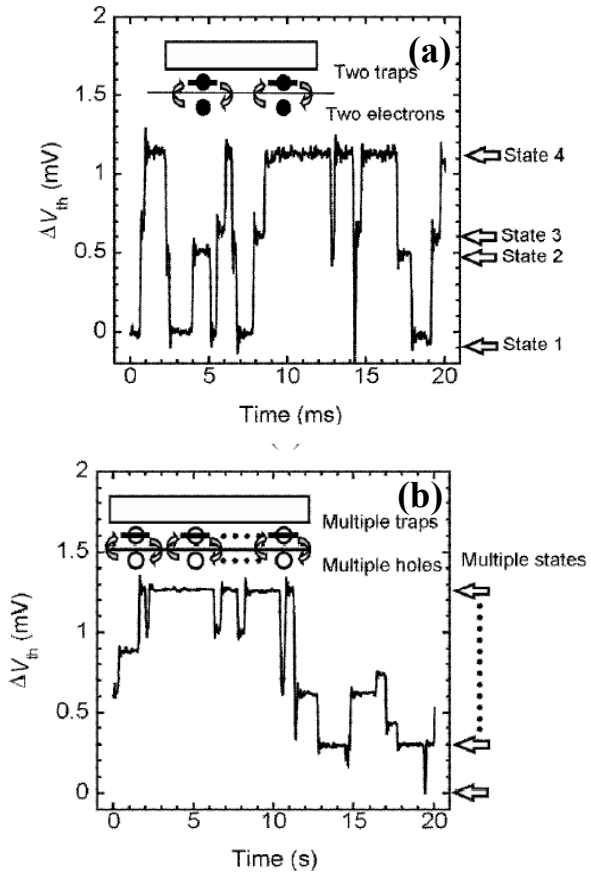


Figure 2.10 Time trace of complex RTN. (a) Four states of RTN originated from two traps can be observed, (b) multiple states of RTN from multiple traps can be observed [91].

2.5 Variability

It is well-known that variability in characteristics of devices is a major concern to down-scaling the supply voltage of current and future complementary metal-oxide-semiconductor (CMOS) technologies [41-42]. In sub-100 nm technologies, much attention has been given to solving and minimizing variability issues. As mentioned earlier, variability can be specified into two categories: Time-zero variability (TZV) and Time-dependent variability (TDV).

TZV is related to fabrication process variations, such as random discrete dopants (RDD), oxide thickness variation (OTV), and line edge roughness (LER) [94][95].

TDV is linked directly to MOSFET reliability [96]. The aging-induced TDV arises from wear-out mechanisms such as Hot Carriers, Negative Bias Temperature Instability (NBTI), and Time-Dependent Dielectric Breakdown (TDDB) [7], [37], [55–58]. It has been reported that TDV originates from aging induced discrete changes in oxide or interface and is independent of AFV [19]. Investigation on BTI variability has been widely reported [21–23] and it can be accurately described by the number of Poisson-distributed defects in the oxide layer and the distribution of each individual defect [19], [104]. The importance of TDV attracted the interest of the researcher because one single defect can shift the threshold voltage over 30 mV, enough to cause failure on a device [19], [105].

In the last two decades, Bias Temperature Instability (BTI) and Time-Dependent Dielectric Variation (TDDB) have attracted the attention of more researchers [6], [57], [60–62] than Hot Carrier Aging (HCA). This is because it was reported by N. Kimizuka [5] that HCA effects are smaller. This is not valid any more for the recent nano-scaled CMOS nodes [93], [63]. ITRS also mentioned in 2009, 2011 and IEEE in 2018 the importance of addressing the RTN issue. These show that HCA and RTN are important issues to be studied and researched into.

2.6 Challenges

The rapid growth of the internet and telecommunications has been attributed to the aggressive down-scaling of the device dimensions [108], [109]. Current developments in the fabrication process has made it possible to integrate more than 1 billion transistors on one single processor die [110]. This high integration density has to be accompanied by stringent efforts to increase the reliability of each transistor. The failure of a single transistor can lead to complete failure of the whole system.

Several studies on HCA variability have been done and it is proposed that the HCA variability could be explained by the BTI variability model [25-26]. Despite recent extensive research on HCA variability, a deeper understanding of the nature of HCA variability such as the connections with the fabrication process and technology, the physical origin, the possible scaling trend [113] on deeply scaled MOSFETs is still missing.

RTN did not capture the attention of researchers until the 2000s. Previously, researchers were interested in RTN as a fundamental characteristic of trap at gate dielectric, and placed low priority on RTN as a reliability problem. In 2006, it was reported for the first time, RTN became a reliability issue in high capacity flash memory [14].

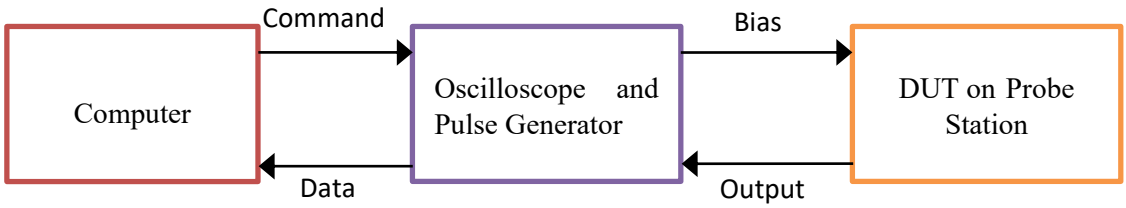
3 Experimental Facilities and Measurement Techniques

3.1 Introduction

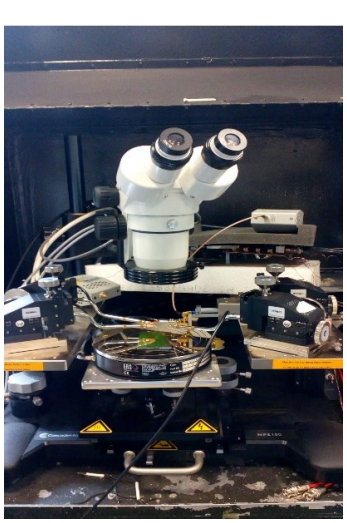
Recent developments in equipment and measurement techniques provide enormous opportunity in achieving a breakthrough in research. Advanced test facilities together with the right measurement skills and methods will generally produce outstanding outcomes.

In order to achieve good results, one must know the fine details of the equipment set-up. Figure 3.1 illustrates a standard wafer level device characterization facility. It consists of a Cascade probe station housed in a black box to minimize interference from the outside environment during the measurement. The Cascade probe station comprises of a thermal-controlled stage and four micro-positioners. The device-under-test (DUT) is placed on the stage of the probe station. Connections to the device's four terminals are made by using four micro-positioners where each positioner's needle contacts one of the four DUT terminals. The connections from the micro-positioner to the pulse generator and to the semiconductor analyser are controlled by a personal computer.

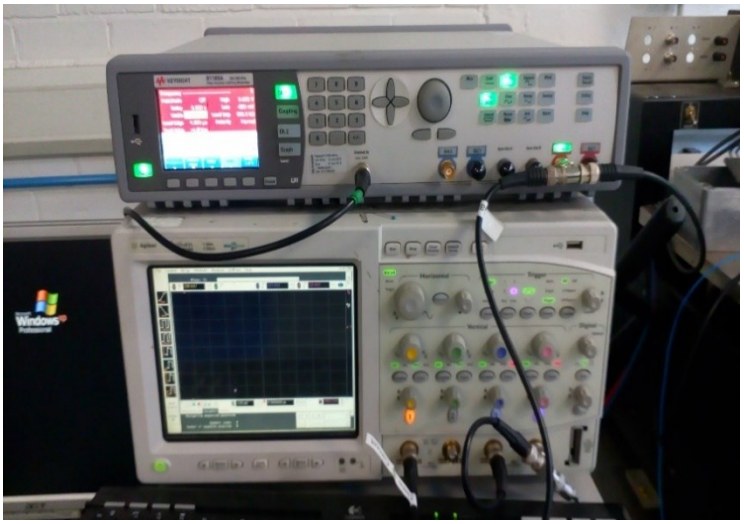
Figure 3.1(b) presents a photo of the micro-positioners located in the probe station. As portrayed in the photograph, the SSMC-to-SMA cables are needed to connect the amplifier circuit to the micro-positioners. In order to avoid impedance mismatch, all system components must possess a 50 Ω impedance and the length of the SSMC-to-SMA cables is minimized. The BNC cables between the circuit and the oscilloscope are required to be of the same lengths in order to synchronise the multiple output channels.



(a)



(b)



(c)

Figure 3.1 (a) Block diagram of the conventional measurement system. (b) picture of the Cascade probe station cable to connect the testing device and circuit and (c) connection of pulse generator and oscilloscope.

Pulse measurement system used in this work comprises of Arbitrary Waveform Generator Agilent HP81150A, an external circuit which includes amplifier to measure drain current, a Visual Basic programming code, a Cascade probe station and Oscilloscope. The

programming code is to control the pulse generator and to automate measurements. This system is capable of delivering a minimum measurement time of 100 ns with a noise margin of 2 mV.

For the quasi-DC measurement, an industrial standard parameter analyser Agilent E5270A and the Keithley 42000-SCS were used. It generally takes about 20~150 ms for measuring one point and tens of points are needed in order to obtain a transfer characteristic. Total measurement time normally will be in the order of seconds. Most of the experiments conducted in this work characterizes thin (<3 nm) gate oxides, it becomes essential to increase the measurement speed by using the pulse measurements because both recovery and degradation can occur during the quasi-DC measurement. This phenomenon will be discussed later in this chapter. Although DC measurement is considered as slow measurement, it will still be carried out to compare the work conducted in this thesis to the standard slow measurement typically applied in the industry.

One of the important parameters during the measurement is drain voltage, V_d. If V_d is too low, the signal is too weak when compared with the noise. If V_d is too high, the non-uniformity of the channel can be substantial. It is found that a suitable V_d is 100 mV with a feedback resistance of 10 kΩ.

In this chapter, the test systems for the conventional measurement techniques, including the transfer characteristics (I-V), threshold voltage and capacitance-voltage (C-V) will be

reviewed. Fast measurements, including the On-The-Fly (OTF), fast pulse I_d-V_g and single point measurements are further described, due to their significance to this project. The devices used in the experiments will also be described.

3.2 Facilities

The test set-up for measurement includes a desktop computer, a Cascade probe station with micro-positioners, Agilent 81150A pulse generator, Keithley 4200 advanced parameter analyser, I-V converter with OPA657 amplifier made in-house and a 4-channel Agilent MSO8104A oscilloscope. Figure 3.2 presents the some of the equipment used for the measurement.

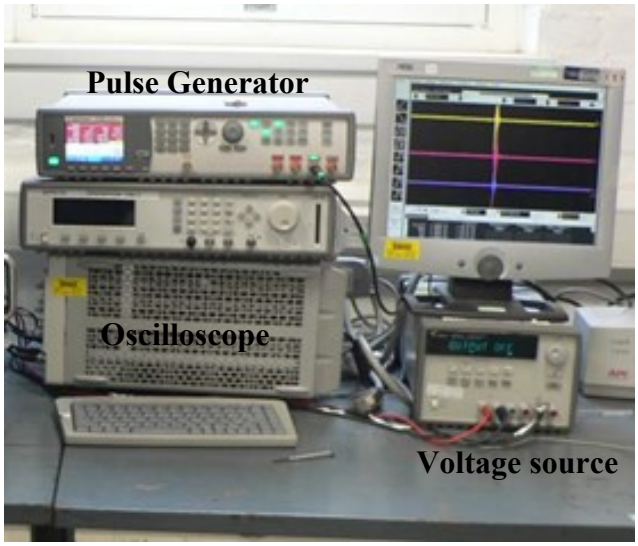


Figure 3.2 Some of the equipment for the measurement consists of Pulse Generator, Oscilloscope and Voltage source.

During the measurement, the quasi-DC system is fully controlled by the desktop computer. Turbo C language is used to write the control program for the system. The pulse system comprises of a desktop computer, an in-house circuit with amplifier (OPA657) to convert current (I_{ds}) to voltage and for amplification. The system is also connected to an Agilent MSO8104A oscilloscope for monitoring applied gate voltage and for circuit output voltage acquisition. The Agilent 81150A pulse generator supplies voltage to the gate of MOSFET. A Keithley 4200 advanced parameter analyser is used to integrate DC-IV, C-V and ultra-fast I-V functions as shown in Figure 3.3.

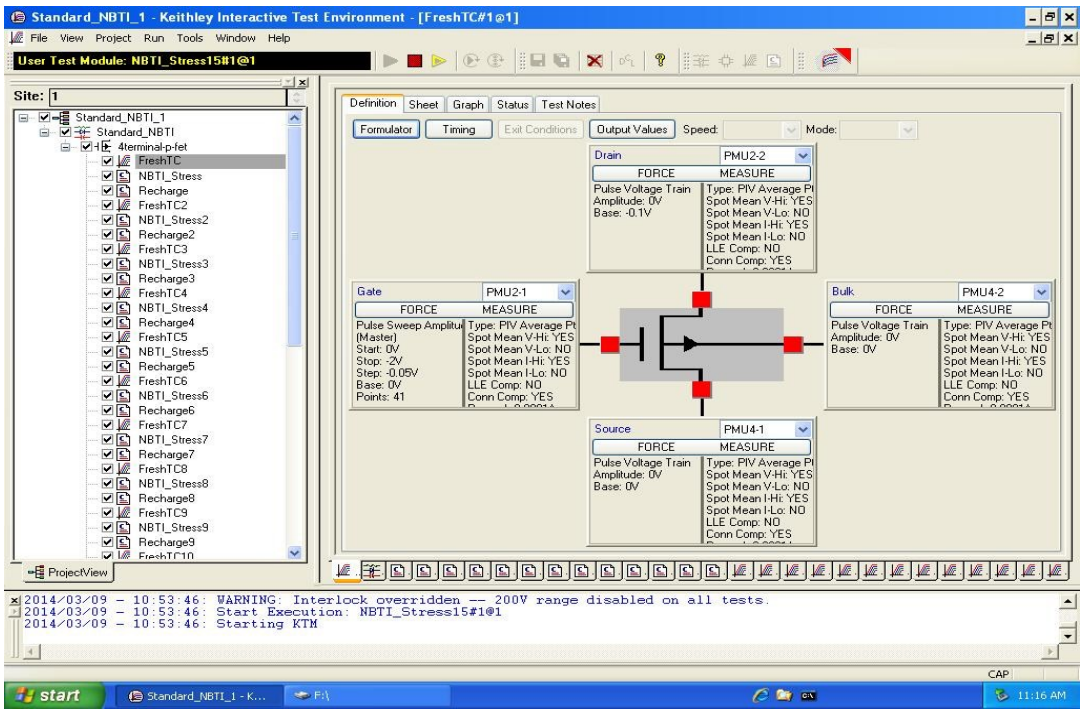


Figure 3.3 Working screen of Keithley 4200 Advanced Parameter Analyser

The annealing process is done using Carbolite CTF 12/100/900, it is a ceramic tube furnace with resistance wire as presented in Figure 3.4. It has the uniform zone length of 375 mm and can withstand the maximum temperature of 1,200 °C. During the annealing process, the temperature ranges from 150 °C to 400 °C. High purity nitrogen or forming

gas (combination of 10% H_2 and 90% N_2) is used during the process. The flow rate is 2 litres per minute and the upstream pressure is 1.2 ~ 1.4 bars.

Figure 3.4 Carbolite CTF 12/100/900 Ceramic Tube Furnace

3.3 Devices

The devices used in Chapter 5 were fabricated at Interuniversity Microelectronics Research Centre (IMEC), Belgium. These samples were processed based on 45 nm technology and 22 nm technology. Samples used in Chapter 4 and Chapter 6 are manufactured at Taiwan Semiconductor Manufacturing Company (TSMC). It was fabricated using TSMC's N28HPL high-k metal gate technology. A summary of the devices used in this thesis is shown in Table 3.1.

Table 3.1 List of the devices used in this thesis

Process	Gate material	Dielectric material	EOT	Size (W x L, um)
28 nm	HKMG	HfO_2/SiON	1.2 nm	2.7x0.9, 0.09x0.027
45 nm	HKMG	HfO_2/SiON	1.45 nm	0.09x0.07
22 nm	HKMG	HfO_2/SiON	1.0 nm	0.09x0.07

The gate dielectric stack used in this project consists of HfO_2 and interfacial SiON with an equivalent oxide thickness (EOT) between 1 nm and 1.45 nm.

3.4 Conventional characterization and stress techniques

In this section, the conventional characterization and stress techniques will be presented. It includes system and equipment set-up, samples used in the experiments, techniques for characterizing degradation, and techniques for stressing devices.

3.4.1 Conventional I_d-V_g technique

The conventional technique for characterizing I_d-V_g is also called slow measurement because the total time taken for one I_d-V_g measurement is in the order of 1-10 seconds. It is measured by applying DC voltages using a DC parametric analyser. The threshold voltage, abbreviated as V_{th}, of a MOSFET is usually defined as the gate voltage where a

strong inversion layer forms at the interface of the substrate. The formation of this strong inversion layer allows the flow of electrons/holes through the drain-source. There are several methods to extract the threshold voltage. Threshold voltage can be extracted from the I_d-V_g curve by using the constant current method [114] or the maximum conduction, g_m-max, method [115]. The measurement is generally carried out using low drain voltages, so that the device operates in the linear region. These two methods are demonstrated in Figure 3.5 and Figure 3.6. The g_m-max method requires the transconductance which is calculated by differentiating the I_d-V_g curve. V_{th} is extracted from the gate voltage axis intercept of the linear extrapolation of the I_d-V_g curve at maximum transconductance.

In the previous chapter, it is explained that negative bias temperature instability (NBTI) is a key reliability issue in MOSFETs. The conventional measurement method for negative bias temperature instability (NBTI) is carried out using the measure-stress-measure (MSM) methodology. It starts by characterising the properties of a fresh device, such as measuring the threshold voltage. The fresh value is used as the reference for measuring the shift of parameters during the stress due to degradation. The stress biases applied are typically considerably higher than that used in the real operation to produce a measurable degradation within a practical test time. During the stress, measurement is interrupted at preset times to measure the I_d-Vg transfer characteristics. For NBTI, stressed I_d-V_g transfer characteristics are expected to be shifted in the negative direction, which signifies the formation of positive charges. It is worth mentioning that the device under test can be stressed either at the room temperature or a higher temperature, such as 125^{o}C, which is a typical temperature used in industry for device qualification.

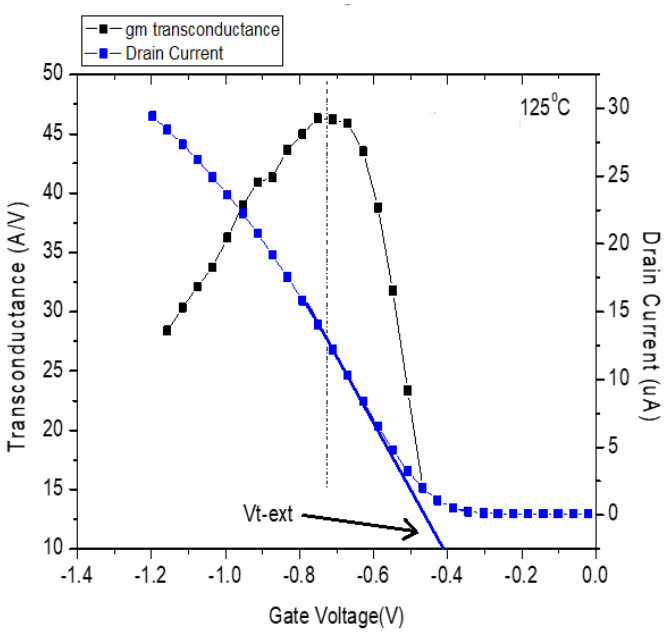


Figure 3.5 The threshold voltage V_t is extracted by the g_m-max extrapolation method.

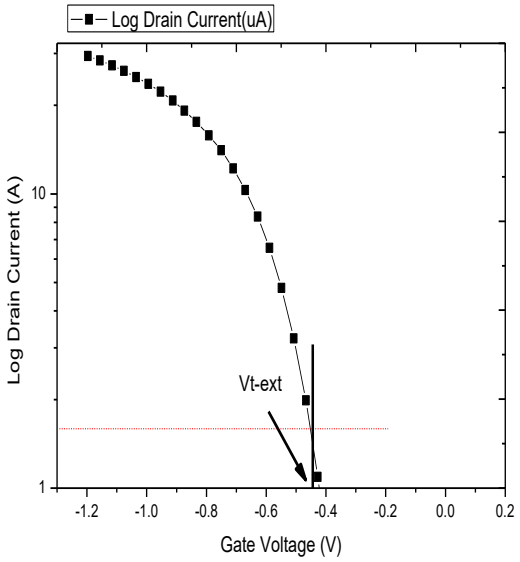


Figure 3.6 the threshold voltage V_t is extracted by the constant current method.

3.4.2 Conventional Capacitance-Voltage (C-V) technique

One of the standard measurements to characterize gate-oxide quality is capacitance-voltage (C-V) measurement. By using C-V measurement, various MOS devices parameters such as the oxide thickness, flatband voltage, threshold voltage, bulk and interface charges information can be extracted.

Figure 3.7 shows the equivalent circuit of an MOS device. Capacitance of an MOS capacitor can be described by the change in the charge (Q_g) of a device when varying voltage (V_g):

$$C = \frac{dQ_g}{dV_g} \tag{3.1}$$

The concept of charge neutrality is upheld whereby Q_g = - (Q_s + Q_{it}) by assuming that there is no charge trapping in the dielectric. Q_s is the substrate charge and Q_{it} is the trapped interface charge. Gate voltage has a partial drop across the dielectric and the semiconductor substrate: $V_g = V_{fb} + V_{ox} + \phi_S$, where V_{fb} is the flat-band voltage, V_{ox} is the voltage drop across the oxide and the ϕ_s is the Si surface potential. Using this assumption, the equation (3.1) is re-written as

$$C = -\frac{dQ_s + dQ_{it}}{dV_{ox} + d\phi_s} \tag{3.2}$$

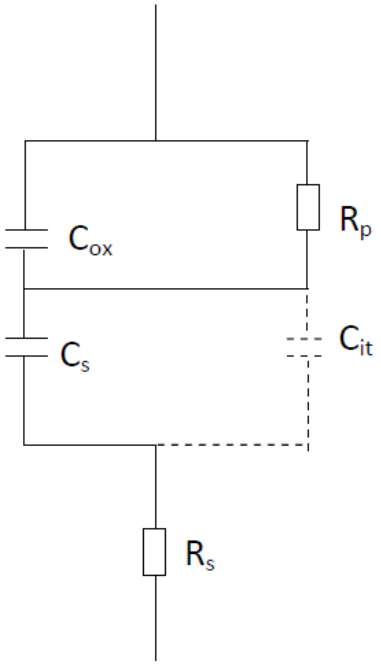


Figure 3.7 Equivalent circuit of a MOS structure

Depending on the Si surface potential, the involvement of the majority, minority and the depletion charge associated with the substrate varies. From the equivalent circuit, the total gate capacitance can also be written as:

$$C = \frac{1}{1/C_{ox} + 1/(C_s + C_{it})} \tag{3.3}$$

The low-frequency substrate capacitance is given by [116]:

$$C_{s,lf} = \hat{U}_S \frac{\varepsilon_{Si}\varepsilon_o}{2L_D} \frac{e^{U_F}(1-e^{-U_S}) + e^{-U_F}(e^{U_S}-1)}{F(U_s, U_F)} \tag{3.4}$$

where the dimensionless surface electric field $F(U_S, U_F)$ is defined by:

$$F(U_S,U_F)=\sqrt{e^{U_F}(e^{-U_S}+U_S-1)+e^{-U_F}(e^{U_S}-U_S-1)} \tag{3.5}$$

U_S and U_F are normalized potentials, defined as $U_S = {}^{q\varphi_s}/_{kT}$ and $U_F = {}^{q\varphi_F}/_{kT}$. The Fermi potential is calculated by $\varphi_F = ({}^{kT}/_{q})ln({}^{N_A}/_{n_i}$ where N_A is the acceptor concentration and n_i the intrinsic carrier concentration in the Si substrate.

The symbol $\hat{U}_S$ stands for the sign of the surface potential and is given by

$$\hat{U}_S = \frac{|U_S|}{U_S} \tag{3.6}$$

where $\hat{U}_S = 1$ for $U_S > 0$ and $\hat{U}_S = -1$ for $U_S < 0$. The extrinsic Debye length L_D is:

$$L_D = \sqrt{\frac{\varepsilon_{Si}\varepsilon_o kT}{2q^2 N_A}} \tag{3.7}$$

Figure 3.8 illustrates the typical setup of the C-V measurement in this work. Figure 3.8(a) and (b) show where the gate-channel capacitance, C_{gc} and gate-bulk capacitance, C_{gb} is separately obtained through the split C-V technique [117]. Figure 3.9(a) and (b) respectively present the C_{gc} and the C_{gb}, and the combination of these two measurements. We can measure the total gate capacitance by combining the C_{gc} and the C_{gb}. Parasitic capacitance will lead to an offset of the measurement and this has been considered by nulling back to zero. In our case, parasitic capacitance has been accounted for in the measurements.

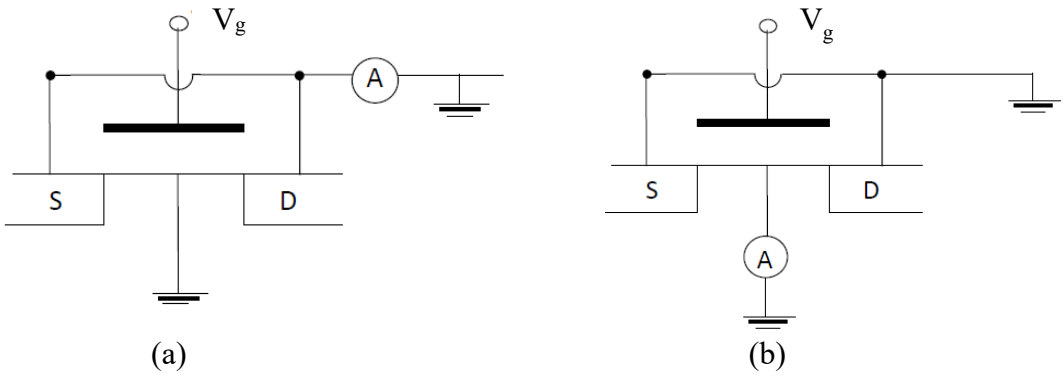


Figure 3.8 Configuration of spilt C-V measurement technique to obtain (a) $C_{gate\text{-}channel}$ against the Gate voltage and (b) $C_{gate\text{-}bulk}$ against the Gate voltage.

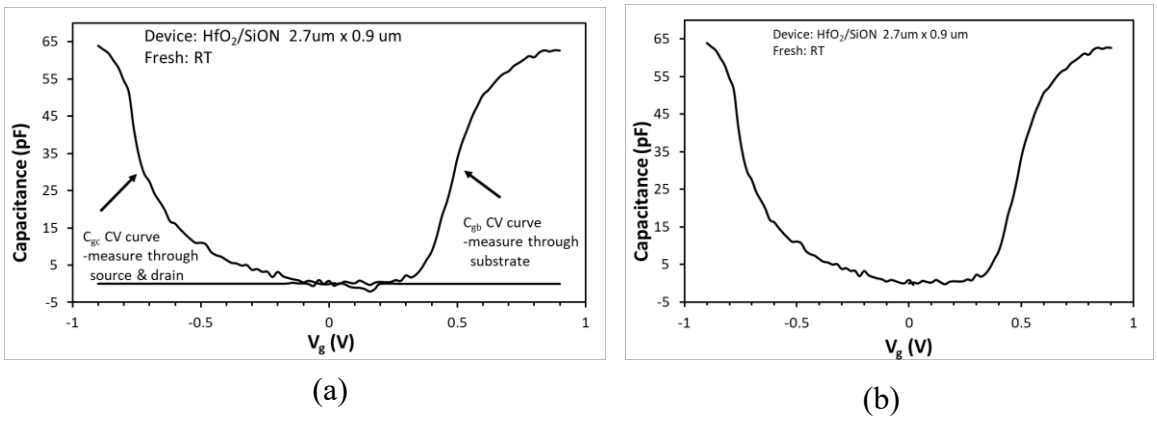


Figure 3.9 Measurement of spilt CV profiles (a) separate C-V measurement were measured and (b) the final profile acquired by the combination of the C_{gc} and C_{gb} measurement.

Another technique to measure the total gate capacitance measurement is by using a single C-V measurement, conducted in this work. This can be done by superimposing a small oscillating AC voltage on a DC voltage, which is applied to the gate. The resulting AC current through the source, drain or substrate is measured from which the capacitance, which is the change in charge in response to the AC voltage, is calculated.

Basically, the combined C-V can be categorized into three regions: accumulation, depletion, and inversion, as described below for a p-type substrate.

Accumulation Region

The accumulation region of the C-V curve can be observed when a negative voltage is applied to the gate of a p-type substrate MOS capacitor. The negative polarity will attract the holes, which are the majority carriers, towards the gate. These holes will accumulate at the oxide/substrate interface due to the oxide being a good insulator. The C-V measurement measures the oxide capacitance in the strong accumulation region at which the voltage is negative enough and the C-V curve is essentially flat. Hence, the oxide thickness can be extracted from the oxide capacitance.

Depletion Region

The holes are repelled from the substrate oxide interface as the gate voltage moves toward the positive. Subsequently, a carrier- depleted area forms beneath the oxide. As the gate voltage becomes more positive, the depletion zone becomes deeper. The depletion capacitance thus becomes smaller and the total measured capacitance becomes smaller consequently.

Inversion Region

As the gate voltage increases, the mid-band energy level eventually falls below the Fermi-level at the interface. In this case, the interfacial region is inverted from p-type into n-type. The positive gate bias attracts the electrons, which are minority carriers, towards the gate. These minority carriers will pile up at the oxide/substrate interface and form an inversion layer due to the oxide being a good insulator. The consequence is that the positive charges on the gate are separated from the electrons in the substrate by the oxide and the total capacitance returns to the oxide capacitance. The electrons in the inversion layer screen the positive charges on the gate from the substrate, so that the depletion depth will not increase further with V_g.

3.5 On-The-Fly (OTF) techniques

The development of the OTF technique was primarily motivated to measure NBTI induced voltage shift without recovery. Numerous studies have found that the V_{th} degradation induced by NBTI recovers during the interruption of stress for the conventional method [6-10].

The common characteristic of the OTF technique is the stress voltage always applied to the gate, and the degradation of the drain current is measured at stress voltage. In order to maintain the stress is always applied to the gate, measurement of ΔV_{th} is done at the stress voltage.

The OTF technique monitors both I_d and the transconductance, g_m, at preset intervals under a low drain bias. To evaluate g_m, the stress V_g is perturbed by a small amount of ± DV and the corresponding current variation is recorded. The g_m at a time "*n*" is estimated from the equation 3.9.

Rangan et al [11] were the first to propose the OTF technique in the year 2003. The technique initially measured I_d-V_g with the V_g ramped to the stress voltage, both the drain current I_{d0} at $V_g = V_{gst}$ and the threshold voltage V_{t0} are recorded. As the electrical stress is continuously applied, the drain current is sampled non-stop at V_g=V_{gst}. The threshold voltage shift is then calculated from the following equation

$$\Delta V_t = \frac{\Delta I_d}{I_{d0}}\left(V_g - V_{t0}\right) \tag{3.8}$$

where the change in the drain current is $\Delta I_d = I_d - I_{d0}$. Due to the uncertainty in the extracted threshold voltage shift, introduced by ignoring the mobility variation with V_g in equation 3.8, this technique is not widely accepted.

An improved OTF technique (2nd order) is then suggested by Denais *et al* [12] as shown in Figure 3.10(b). In this method, mobility degradation is taken into consideration by evaluating the transconductance, g_m. Three points are measured at V_{gst} and V_{gst}± DV as shown if Figure 2.6(b). In order to estimate the transconductance, $g_m(n)$, V_g was perturbed by a small ±DV, where D signifies perturbation.

$$g_m(n) = \frac{I_d(V_g + DV) - I_d(V_g - DV)}{2DV} \tag{3.9}$$

The degradation of drain current between two measurement points '*n*' and '*n*-1' is,

$$\Delta I_d(n) = I_d(n) - I_d(n-1) \tag{3.10}$$

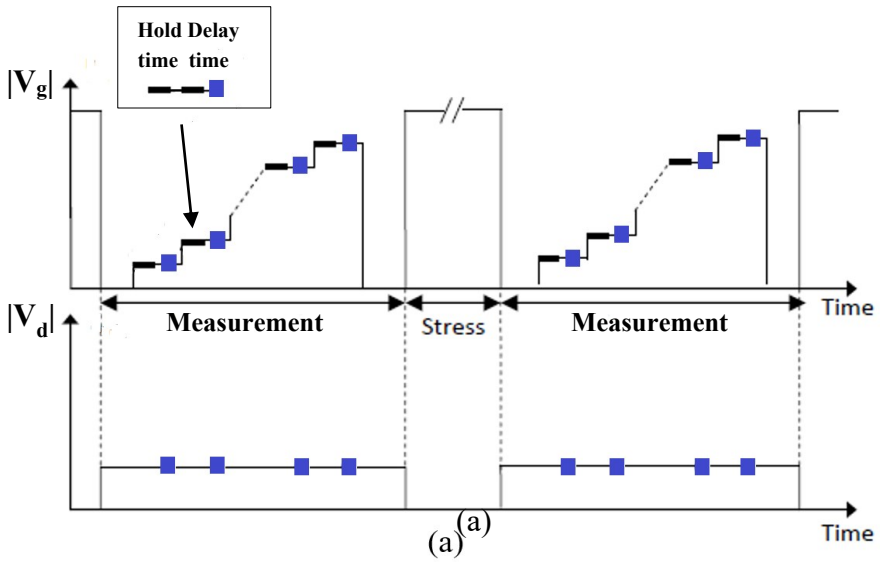
The shift of threshold voltage between these two points can be evaluated by,

$$\Delta V_t(n) = -\frac{\Delta I_d(n)}{g_m(n)} \tag{3.11}$$

The accumulative shift of threshold voltage is,

$$\Delta V_t = -\sum_{n=1}^{M} \frac{I_d(n) - I_d(n-1)}{g_m(n)} \tag{3.12}$$

where M is the number of I_d measurements and $g_m(n)$ is the mean value of the trans-conductance between the n^{th} and n-1^{th} I_d measurements, as shown in Figure 3.11. Hence, periodical three-point I_d measurements are enough to monitor ΔI_d, g_m, ΔV_t during stress.



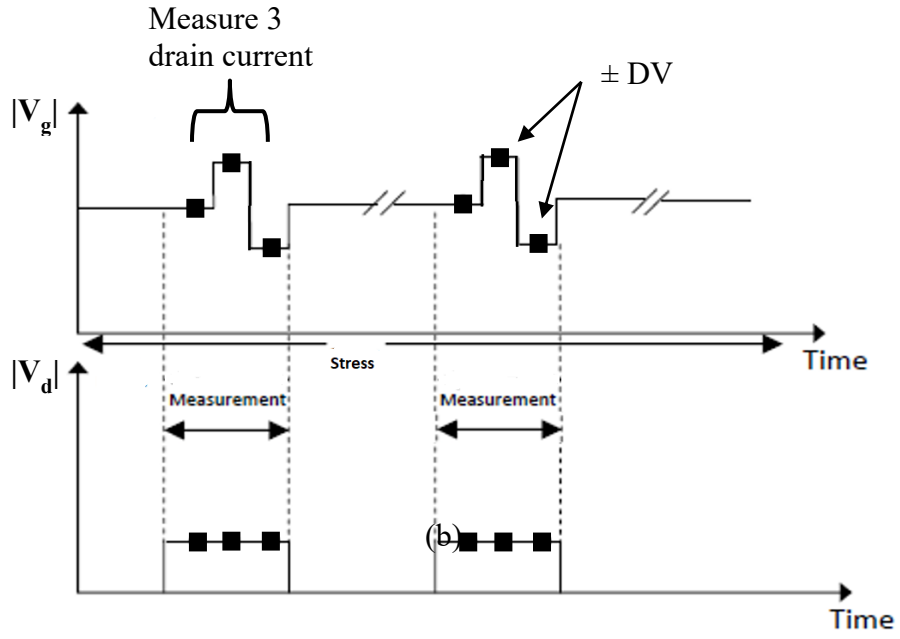


Figure 3.10(a) Traditional NBTI test sequence (b) The 2nd order OTF measurement sequence

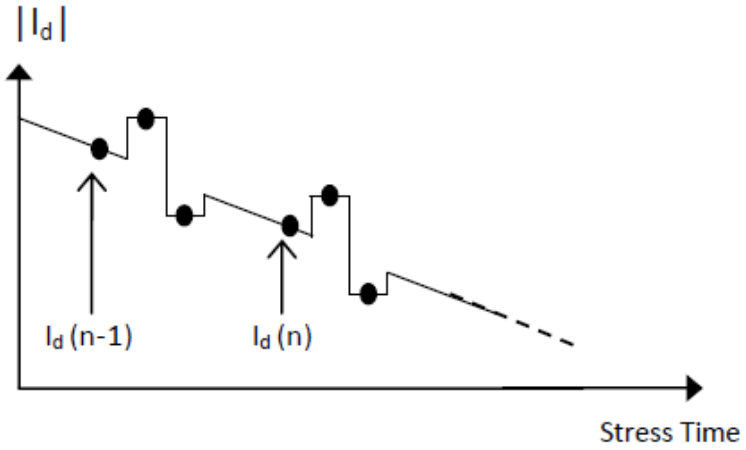


Figure 3.11 The measurements of n^{th} and n-1^{th} I_d, combined with the transconductance $g_m(n)$, can provide the threshold voltage shift, ΔV_t of n^{th} and n-1^{th} measurement points.

3.6 Pulse I_d-V_g techniques

The Pulse I_d-V_g technique becomes popular, since NBTI recovery can be substantial when using a conventional measurement technique. Pulse I_d-V_g works by applying a pulse signal generated by the pulse generator to the gate of the transistor as presented in Figure 3.12. A digital oscilloscope is used to record drain current during the pulse edges. The transfer characteristic I_d~V_g can be determined from the gate voltage and the corresponding drain current. The benefit by using this technique is that the threshold voltage (V_t) can be determined after the application of a pulse by capturing the I_d~V_g during the falling edge of the stress pulse. This can minimize the trapping/de-trapping during measurement.

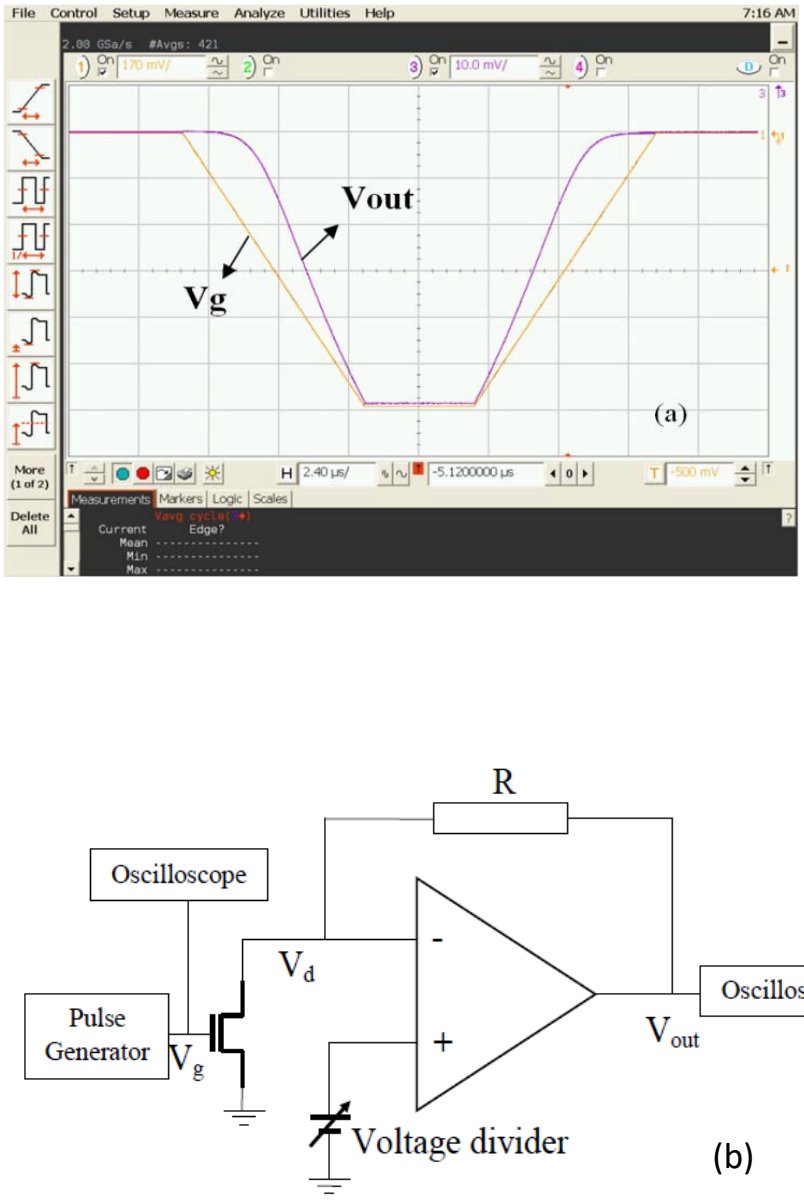


Figure 3.12 A typical result measured with pulse technique. a) A screenshot of V_g and V_{out} acquired by the oscilloscope. V_g was supplied by pulse generator. b) The extracted TC curve with 5 μs pulse edge time under V_d=25 mV; both up and down edges can be used to obtain I_d-V_g curve [13].

3.6.1 Experimental Set-up

Pulse I_d-V_g technique was first developed by Kerber *et al* [14] to investigate the large charge trapping occurring in high-k dielectric. In this technique, the MOSFET is connected to an inverter circuit with the load resistor, R_L. The resistor and the MOSFET's channel form a voltage divider. The schematic measurement set-up of this pulse measurement is illustrated in Figure 3.13. By applying a trapezoidal (triangular) pulse to the gate, the drain voltage, V_D, is recorded using a digital oscilloscope.

From the measured V_D, the I_d~V_g characteristic can be determined using

$$I_D = \frac{100\ mV}{V_D}\left(\frac{100\ mV - V_D}{R_L}\right) \tag{3.13}$$

where R_L is the resistive load of the inverter circuit.

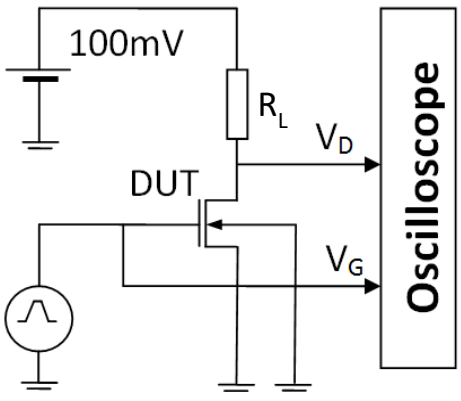


Figure 3.13 Schematic set-up for the pulse I_d-V_g technique proposed by [118].

The voltage divider used in this circuit potentially caused changes of drain voltage during the measurement. However, this effect can be suppressed by normalizing the extracted drain current to a constant drain voltage, which is given by the term 100 mV/V_D in equation (3.13).

In order to suppress the noise and increase the accuracy of the measurement, impedance along the signal path of this circuit needs to be matched and hence the resistive load should be around 50 Ω. But this limits the gain of the circuit significantly.

A better approach to increase the gain of the circuit is to use op-amp (operational amplifier). This at the same time maintains impedance matching along the signal path. The new circuit is presented in the Figure 3.14. In this schematic circuit, the MOSFET's drain is connected to the negative input of the op-amp. Since the voltages at the two input terminals are approximately equal when negative feedback is present through R, the drain voltage of the MOSFET is fixed at V_d supplied by the voltage source.

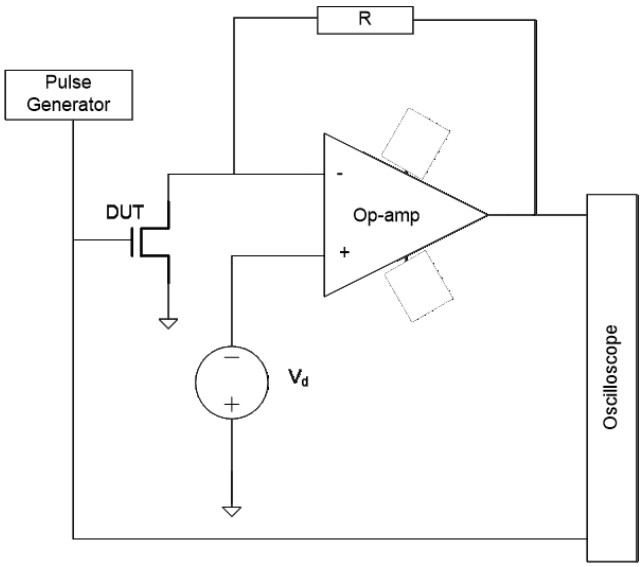


Figure 3.14 Schematic of our modified pulse I_d-V_g.

In principle, the input bias current of the op-amp is very low, the drain current flows almost entirely through the gain resistor, R. Resistors ranging from 1 to 10 kΩ are used for different gain. The output voltage of the op-amp, in terms of the MOSFET drain current is given by the following equation:

$$V_{out} = I_d R + V_d \tag{3.14}$$

3.6.2 Calibration of Pulse Measurement System

In order to ensure that our measurement system has an acceptable margin of noise level and performance, the details the various checks conducted on the pulse measurement system are described in this section.

Op-amp circuit calibration without connection to device under test

Before start the measurement, a calibration exercise was first carried out. This was done by connecting only the op-amp circuit without the device under test. The 10 kΩ resistor is grounded and the system noise is checked, without connecting the pulse generator. The system noise should be less than 0.1%. Figure 3.15(a) shows the connection prepared. Figure 3.15(c) shows that the accuracy of 0.09% of the I_d measurement using the similar system, so that the noise is negligible [119]. The pulse was connected to the input and the current at the edge of the pulse was measured as presented in Figure 3.15(b),. Different values of pulse edge time were applied and the comparison of the measured current with different pulse edge times is shown in Figure 3.15(d). The good agreement cemented that the respond of op-amp circuit in the range of a few micro-seconds.

Op-amp circuit calibration with connection to device under test

The op-amp circuit then connected to the device under test (transistor 90nm x 70 nm). the V_g waveform is applied to the circuit for calibration as shown in Figure 3.16(a). the I_d-V_g was measured from the pulse edges of 5 µs. The V_g was applied for 50 times and each I_{d0} is calculated at constant coltage as shown in Figure 3.16(b). Accuracy of 0.8 µs can be achieved when the same measurement was repeated many times, less than 0.2%.

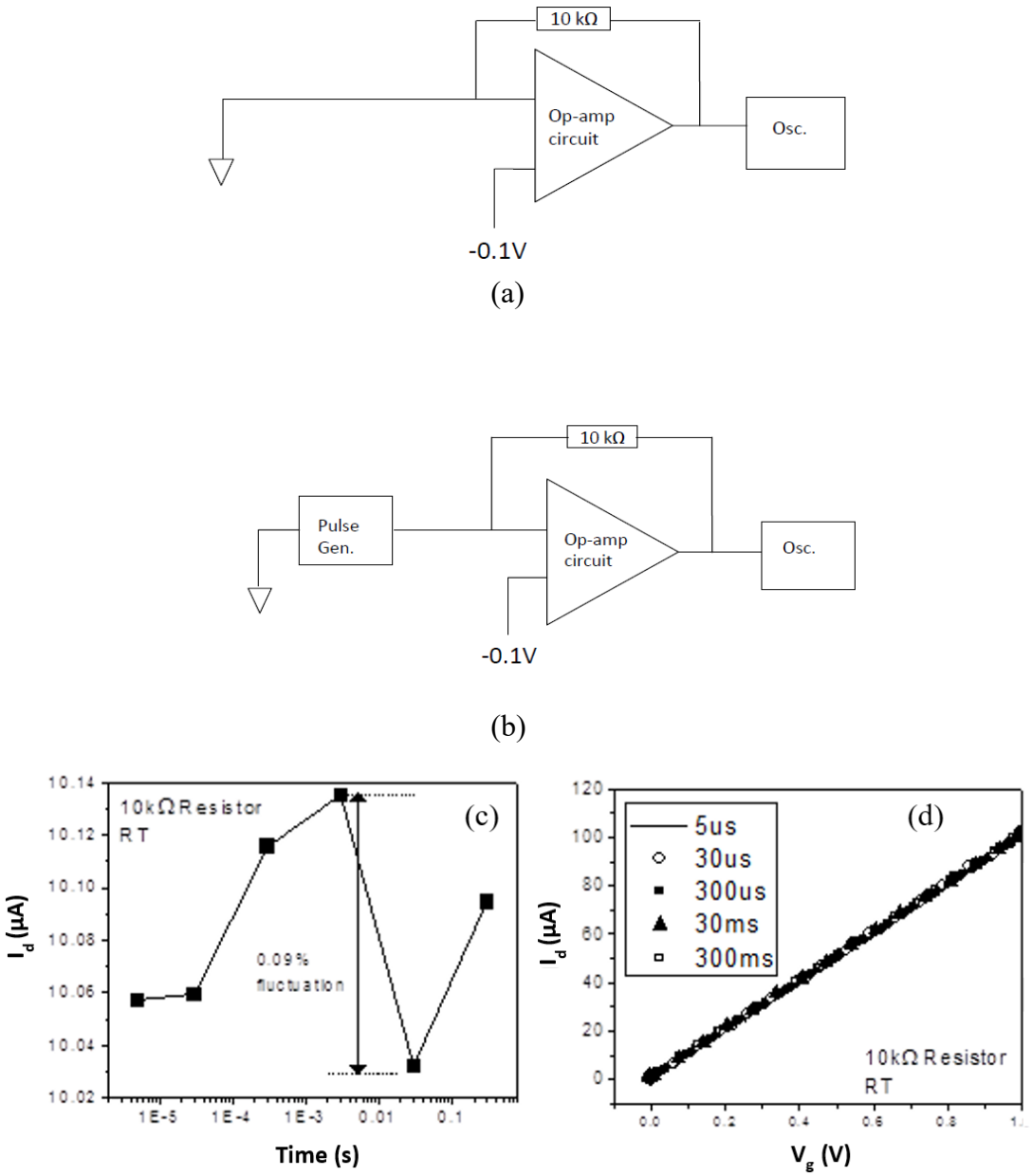


Figure 3.15 The connection of op-amp circuit without connecting to a device for calibration. (a) the circuit when the 10 kΩ resistor is grounded. (b) the connection when the 10 kΩ resistor is connected to the pulse generator. (c) and (d) are the measured I_d obtained from configuration (a) and (b) respectively [119].

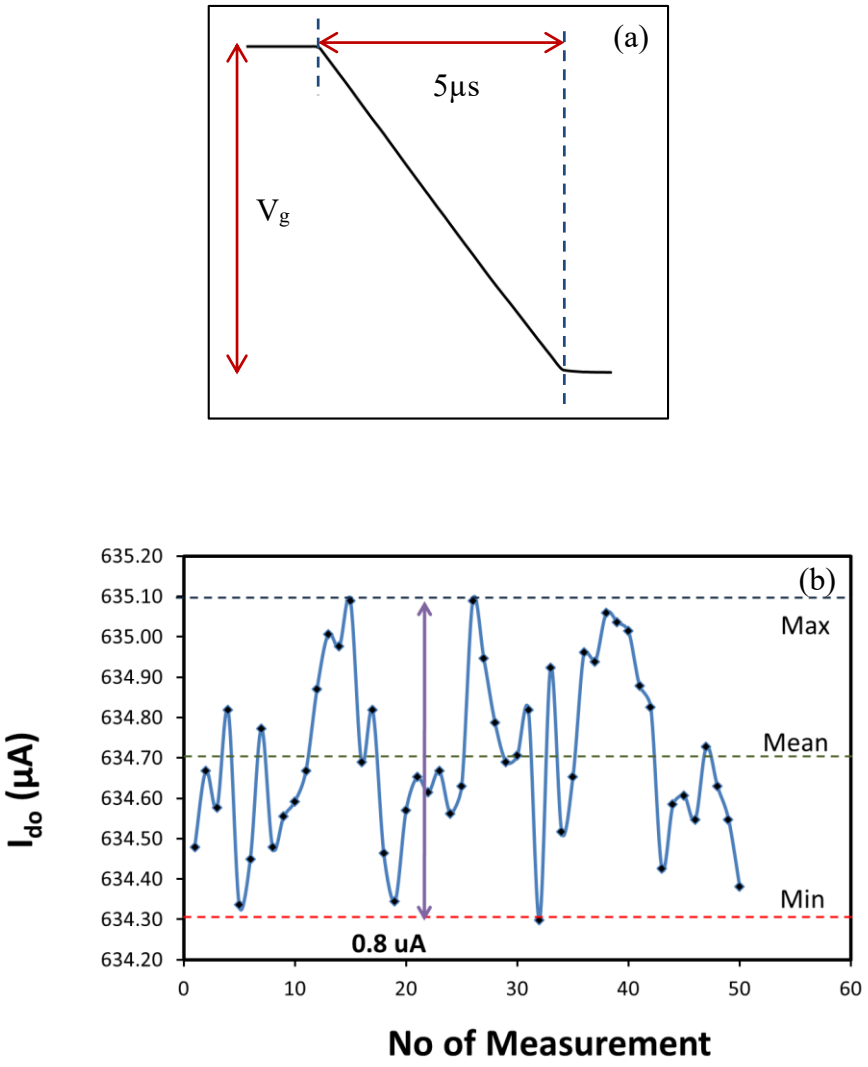


Figure 3.16 Result of pulse measurement calibration with DUT (90 nm x 70 nm) connected by repeating the same measurement many times. (a)waveform of pulse V_g used and (b) shows the measurement variation at a constant voltage.

3.7 Measurement Set-up for Nano-Devices

Set-up for Discharge-based Multi-Pulsed Technique (DMP)

In Chapter 6, discharge-based multi-pulsed technique (DMP) is applied on the nano-devices for probing traps. In this measurement set-up, a desktop computer is used to

control the pulse signal generator and oscilloscope. A Cascade probe station with four micro-positioners is used to probe the nano-device. An in-house developed circuit is used as an amplifier. On the oscilloscope, Channel 3 and Channel 4 are used to monitor and measure output current, I_{ds} of the devices, as shown in Figure 3.17.

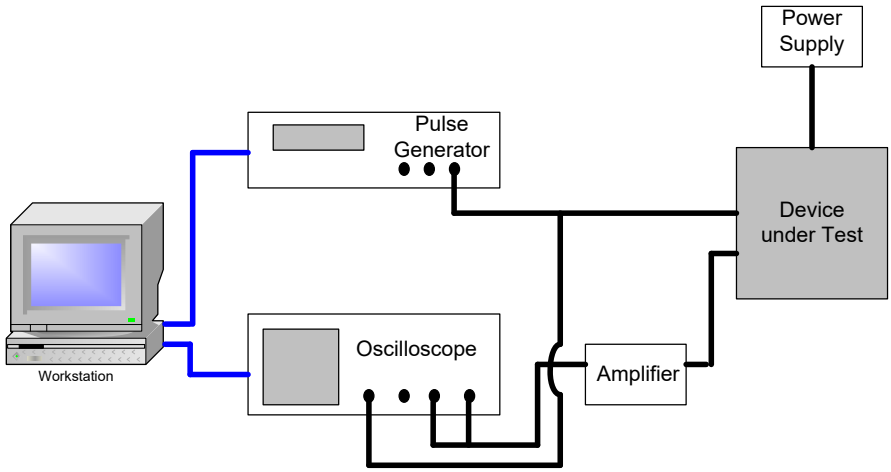


Figure 3.17 Schematic diagram of measurement set-up for DMP technique.

Measurement for a nano-device is very challenging due to its small signal and noise effects. The oscilloscope must be properly set up to be sensitive enough to measure the raw signal. Channel 2 and Channel 3 of the oscilloscope are used to monitor Vsd of the device. Basically, in this measurement set-up, Channel 3 is used to trigger the pulse generator when the signal reaches the Max value of the measured within-device-fluctuation/random telegraph noise (WDF/RTN). While Channel 2 is used to capture I-V measurement when it is triggered. Figure 3.18 presents the waveform of Channel 3 and Channel 4.

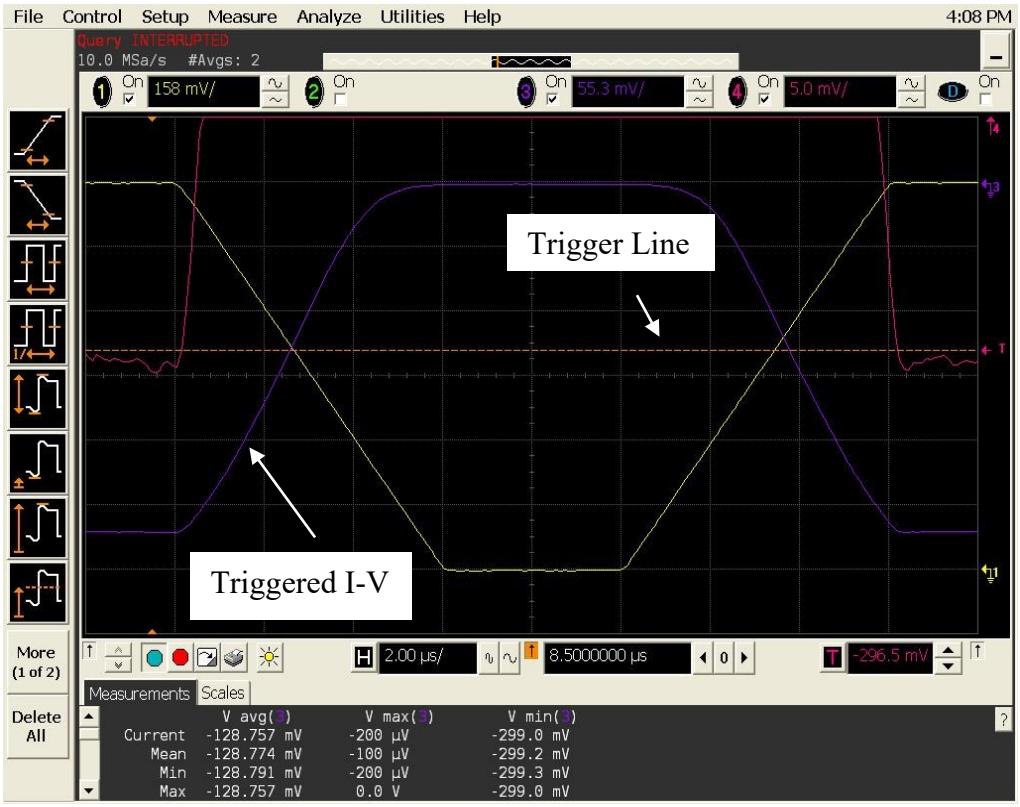


Figure 3.18 Trigger mechanism on Channel 2 and Channel 3 of oscilloscope.

Figure 3.19 shows the steps involved during the measurement. The test starts from measuring spot-IV, from gate voltage of -0.45 V to -1.05 V with 25 mV interval. At each of the V_g intervals, +1 V is applied for 40 s to discharge all the traps if there are traps that have been charged.

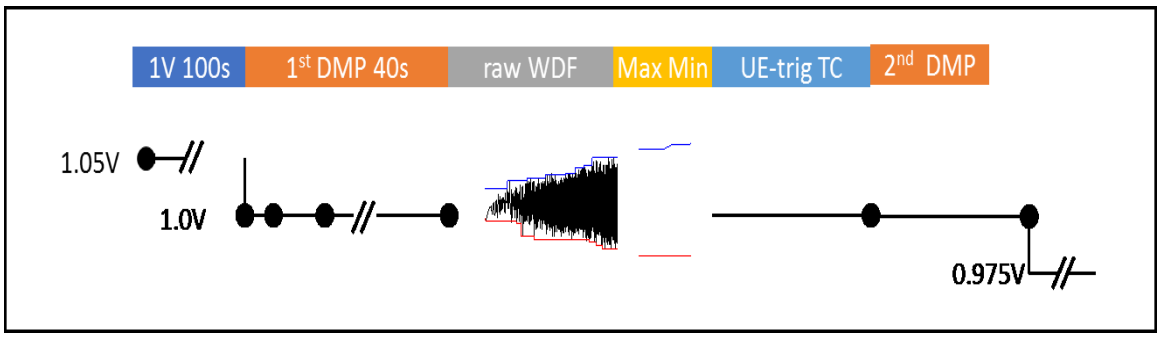


Figure 3.19 Steps involved in measurement.

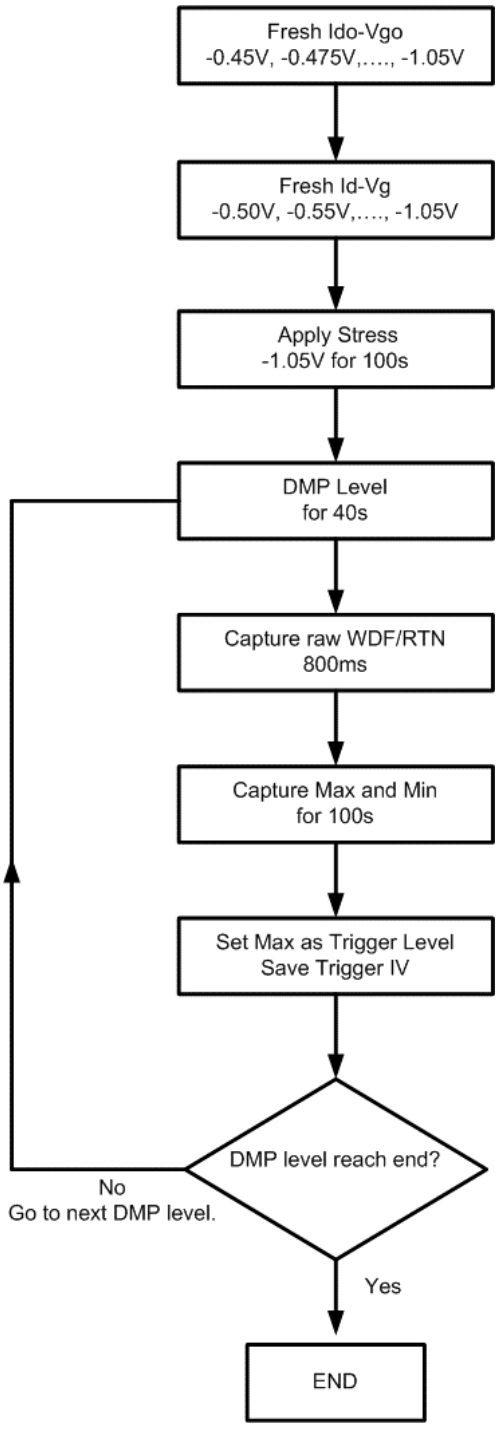


Figure 3.20 Flowchart of the process

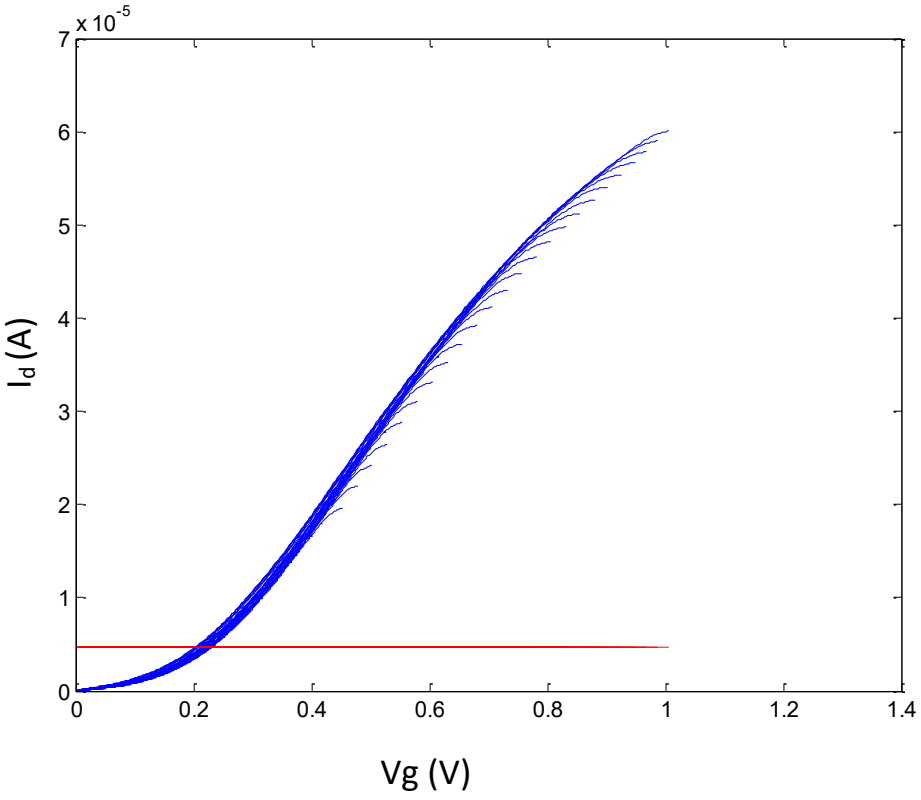


Figure 3.21 Fresh IV plots from a device under test with V_g from +0.45 V to +1.05 V with 25 mV between the each V_g.

After that, the devices under test are subjected to NBTI stress at 1.05 V for 100 s. Then, measurement is started with the first level of V_g = 1.0 V. Gate voltage is applied for 40 s and then the raw WDF signal is captured for 800 ms using oscilloscope. Subsequently, the Max (maximum value of V_d) and Min (minimum value of V_d) measured by the oscilloscope are captured for 100 s. Figure 3.22 presents the raw WDF signal and the max and min values measured during the time window (Tw) of 100 s time period. The reasons for only capturing Max and Min is to control the size of each data file and to reduce the time for saving each item of the raw signal data on the hard-disk.

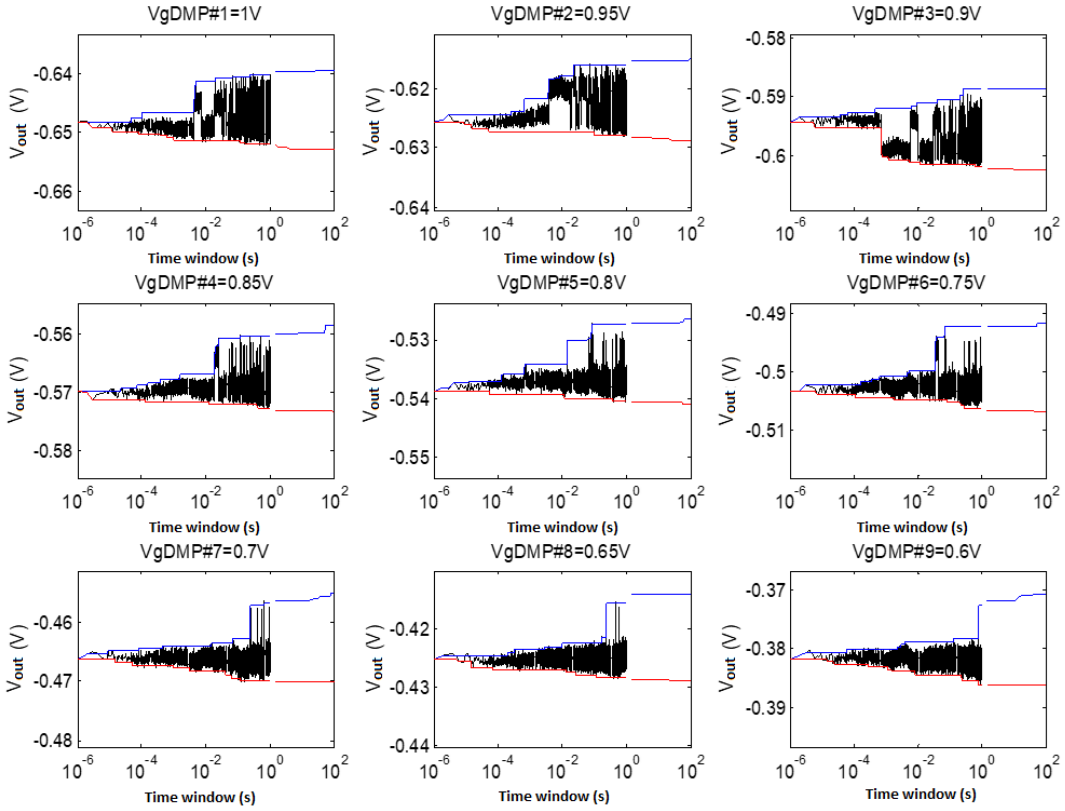


Figure 3.22 Raw WDF signal captured as each level of DMP

When the measurement of max and min comes to the end at 100 s, subsequently the Max value will be used as our trigger level as shown in Figure 3.23 . During this stage, IV will be triggered and captured when the WDF signal reaches the set trigger value. The average IV from 100 repeated measurements will be saved in the computer and the test procedure will move to the next V_g value of 0.95 V. These procedures will be repeated until V_g reaches 0.55 V or the lowest V_g value for the oscilloscope able to trigger Max set value.

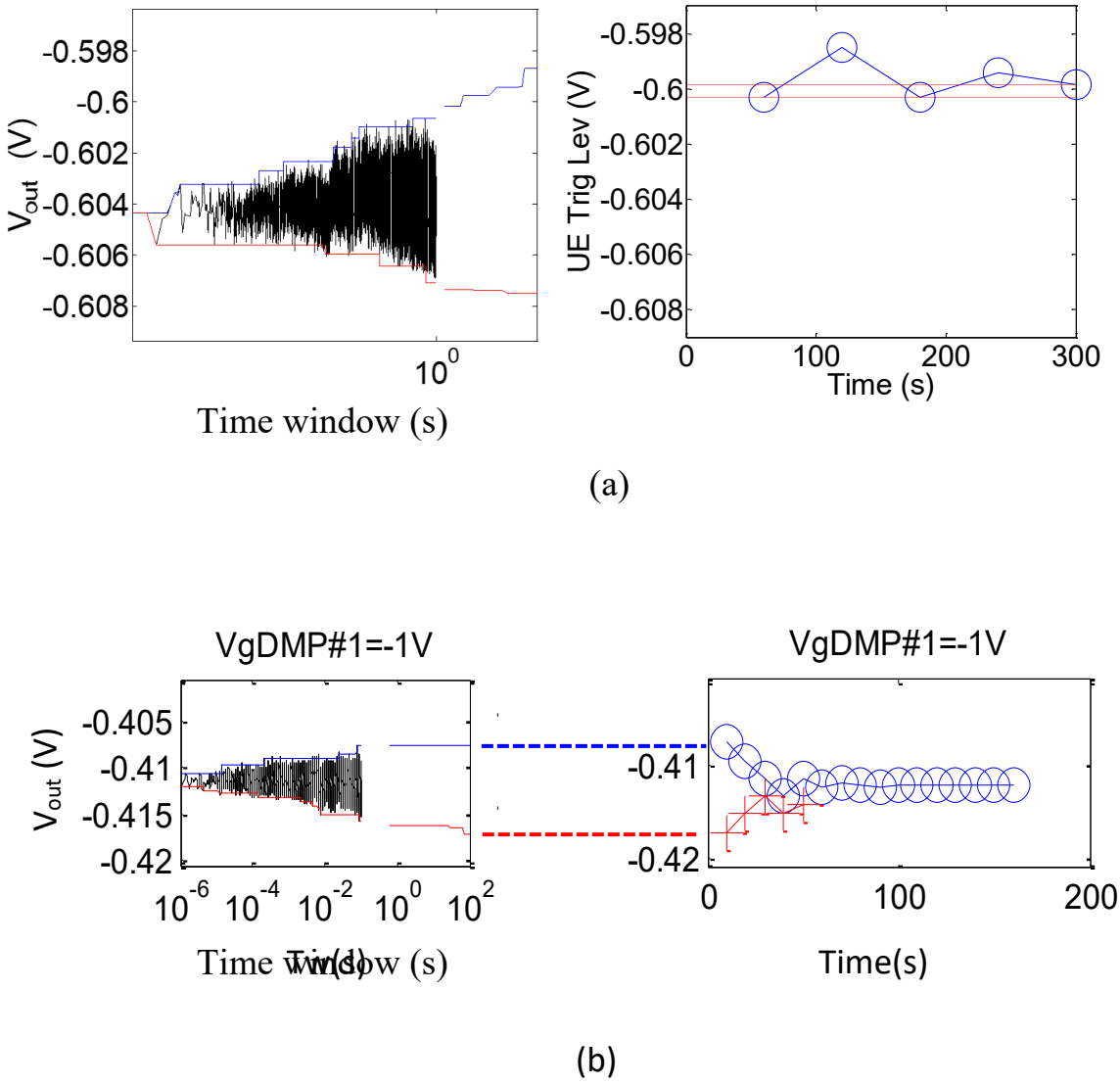


Figure 3.23 Triggering level after raw WDF (a) shows correct trigger level setting while (b) show trigger levels were far from the Max value.

To check that our trigger setting is triggered approximately at the Max level detected at the WDF-signal, we measure each of the trigger-levels set by the program. Figure 3.24 presents the triggering method used and the accuracy comparison between these two triggering methods, using sampling rates of 10 MSa/s (Figure 3.24(a)) and 100 MSa/s Figure 3.24(b)). The results show that 100 MSa/s gives better accuracy and is faster to meet trigger conditions compared to 10 MSa/s sampling rate. Table 3.2 summarizes

settings of the oscilloscope and type of data measured during each step of the measurement.

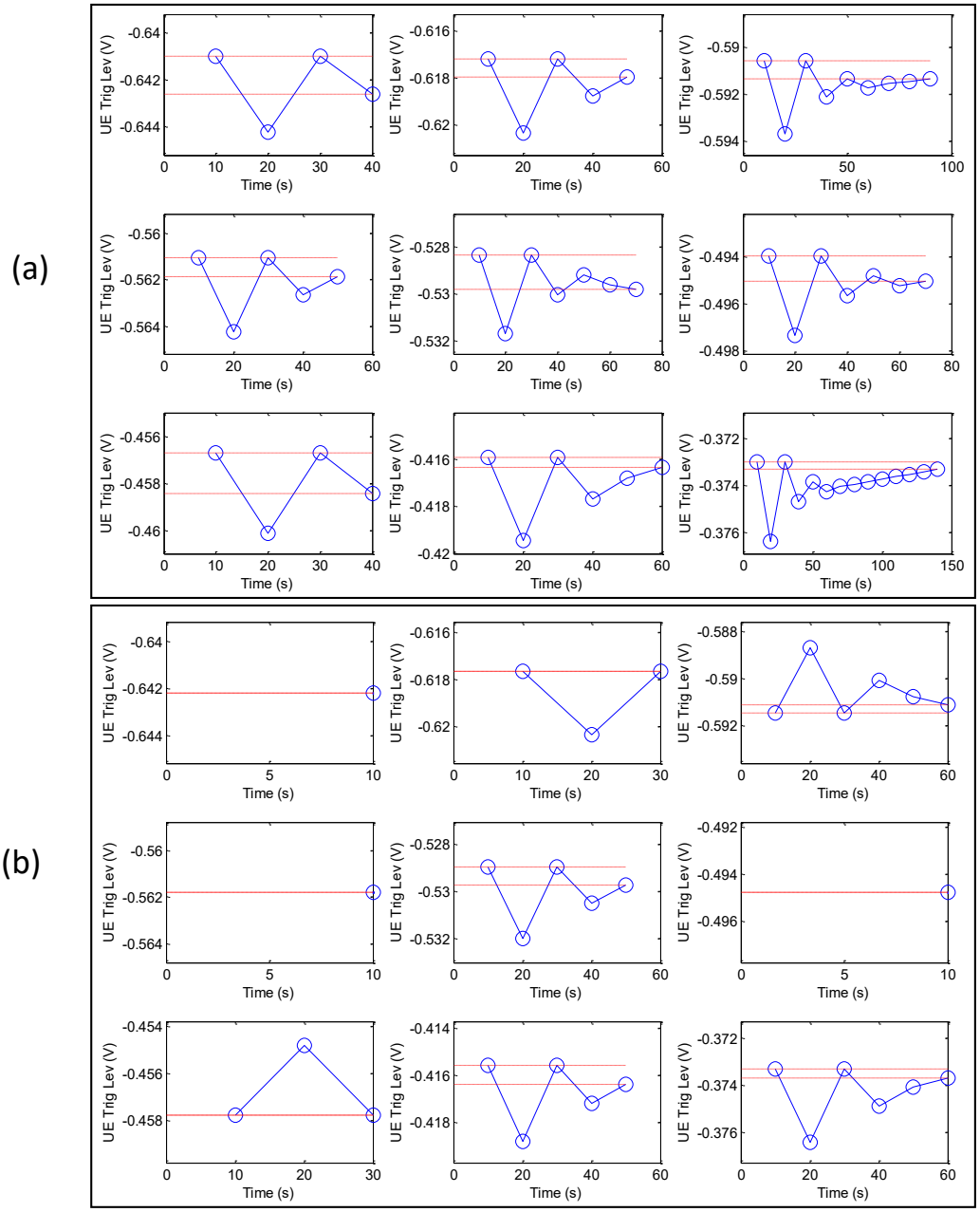


Figure 3.24 Trigger point selected at each V_g level. First point is maximum value from WDF/RTN signal and subsequent points is trigger level set. (a) Sampling rate = 10 MSa/s and (b) 100 MSa/s. 10 seconds interval is because the time allocated for the trigger settings comply with pre-set condition.

Table 3.2 Summary of the settings and type of data captured during the test.

Step	***Sampling Rate***	***Channel***	***Data Captured***
Fresh-IV	100 MSa/s	Ch1 & Ch. 3	I-V
Charging	100 MSa/s	Ch1 & Ch. 3	I-V
DMP	100 MSa/s	Ch1 & Ch. 3	I-V
WDF/RTN	1 MSa/s	Ch.1, Ch. 3 & Ch.4	Raw Signal
UE-Trig	100 MSa/s	Ch.1, Ch. 3 & Ch.4	I-V

In this measurement method, fresh I_d value and V_g value at each of the steps are measured at the flat region during the Fresh-IV measurement as illustrated in Figure 3.25. From the figure, each I_{d0} and V_g at each V_g-step is translated in one fresh I-V that will be used as our I_{d0}. In this way, we recorded the I_{d0} at true V_g.

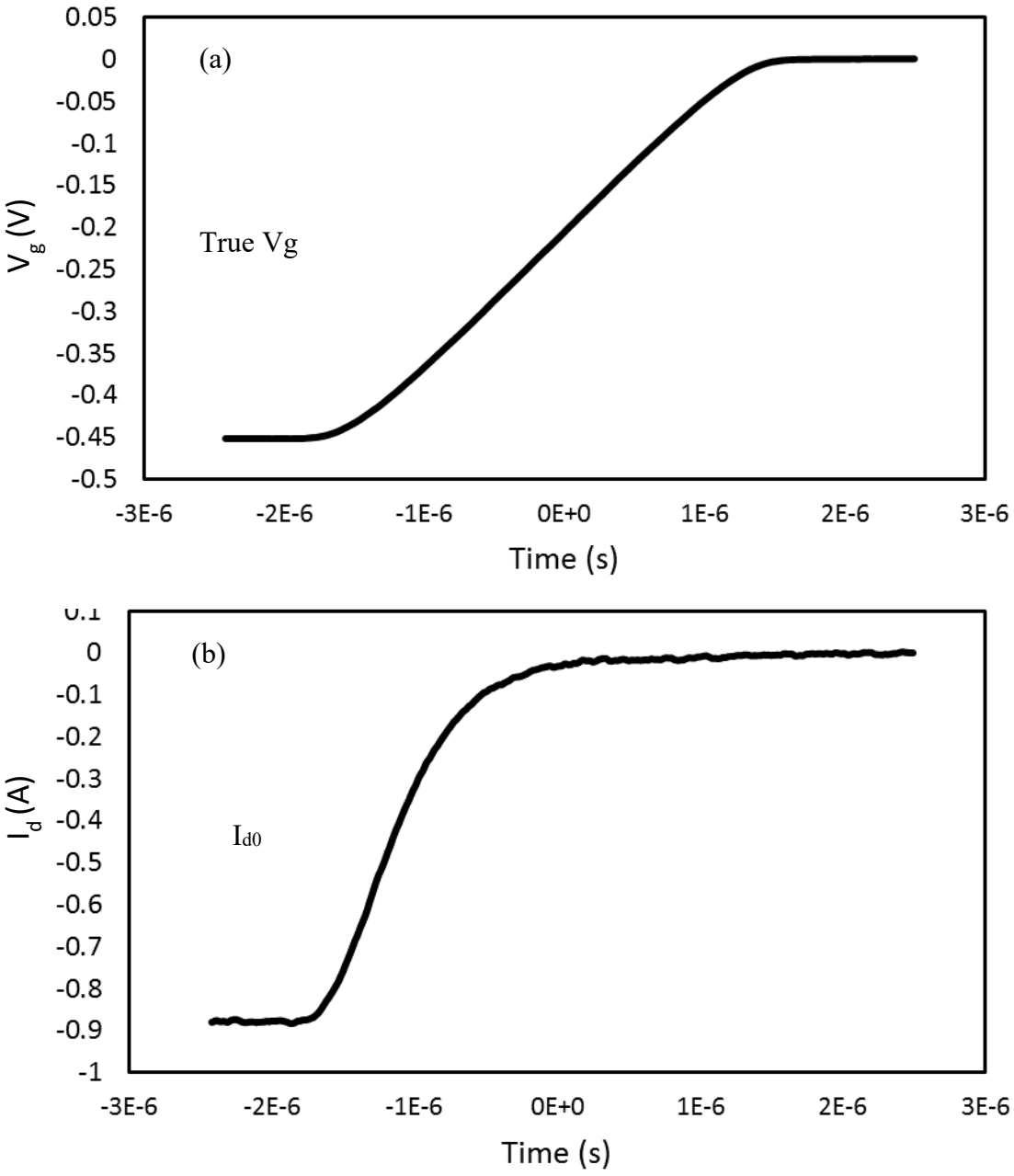


Figure 3.25 I_{d0} measurement at true V_g pre-set at pulse generator (a) shows gate voltage signal and 'red box' is where value of true V_g is measured and (b) shows drain current signal and 'red box' is where value of I_{d0} is measured. Inset is V_g signal (in (a)) and I_{d0} signal (in (b)).

One of the ways to improve accuracy of the measurement is by making sure that the drain current, I_d, during the UE-trigger is compared with the reference I_{d0} at the same V_g. In case the I_{d0} is not available at the V_g=UE-trigger, we use interpolation to obtain it. To

ensure the accuracy of the interpolated I_{d0}, the reference I_{d0}-V_g must have a sufficient number of points. For a pulse edge time of 3 µs, a 10 M/sec sampling rate gives 30 points. Figure 3.26 shows that this is inadequate. When a 100 M/sec sampling rate was used, there are 300 points, giving an acceptable accuracy.

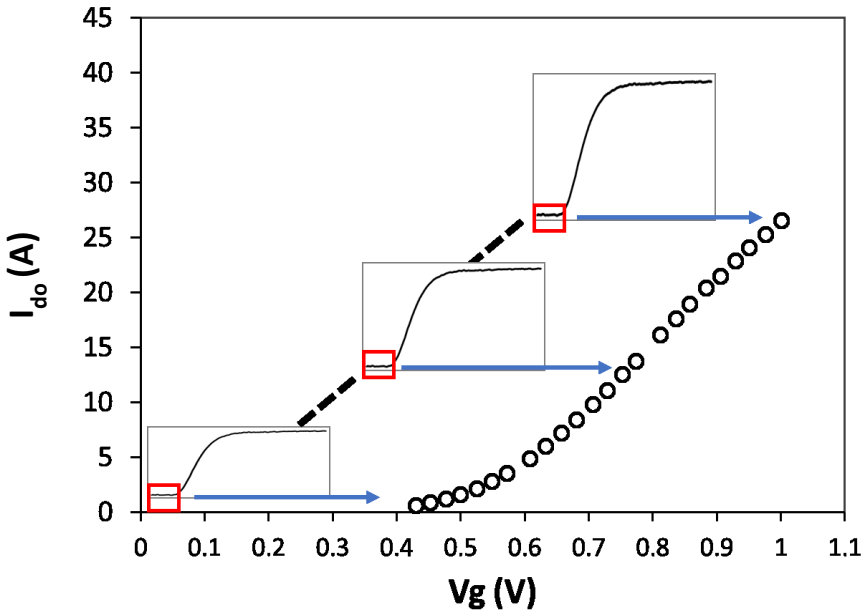


Figure 3.26 Comparison of the fresh IV with IV at discharge.

3.8 Statistical Methods

3.8.1 Standard Deviation

To study device-to-device variations of nano-scaled devices, statistical properties of a device's parameters such as drain current degradation need to be measured. The important

statistical properties for a parameter '*X*' include its 'mean (μ)' and 'standard deviation (σ)'. 'μ' is defined as the arithmetic average of the population (equation 3.9) and 'σ' is defined as the root-mean-square (RMS) deviation of the values from the mean (equation 3.10). A low standard deviation indicates that the data points tend to be close to the mean, whilst a high standard deviation indicates that the data points are spread out over a large range of values.

$$\mu(X) = \frac{\sum_{i=1}^{N} X_i}{N} \tag{3.15}$$

$$\sigma(X) = \sqrt{\frac{\sum_{i=1}^{N} (X_1 - \mu(X))^2}{N}} \tag{3.16}$$

In this application, *X* can be an electric parameter such as V_{th}, I_{dS}, $\Delta I_{dS}/I_{dS0}$, ΔV_{th}, etc. *N* is the number of tested samples. In chapter 6, the mean and sigma value against NBTI stress time will be investigated.

3.9 Conclusions

In this chapter, the principles of various techniques for characterising the gate dielectrics in MOS devices are described and corresponding working principles are briefly reviewed. The different methods for extracting the threshold voltage have been demonstrated. The C-V, the OTF technique, and the pulse I-V measurement system have been presented and

the mechanism behind the measurements has been discussed. In addition, some techniques specifically for deeply scaled devices is also discussed, such as RTN. The accuracy of the pulse I-V measurement is calibrated.

4 Hot Carrier Aging and its Variation under Use-bias on Large and Small Devices

4.1 Introduction

Variability in the characteristics of a device is the major obstacle to the down-scaling of the supply voltage of current and future generations of CMOS technologies [63], [71], [90], [92], [120]–[123]. This issue may exacerbate reliability problems. HCA, NBTI and RTN introduce additional sources on a device's variability.

HCA is a well-known detrimental phenomenon where it impacts the performance of MOS transistors, the fundamental building block of modern microelectronics. According to the Lucky Electron Model proposed by Hu [8], an electron gains enough energy without suffering an energy stripping collision in the channel is emitted into oxide by overcoming the local energy barrier at the Si-SiO2 interface.

Recent results in Figure 4.1 show that HCA can be a severe factor for current and future CMOS nodes [62], [63], [124][120]. Figure 4.1(a) shows HCA gives lower device lifetime compared to BTIs and Figure 4.1(b) shows higher degradation for a device under HCA.

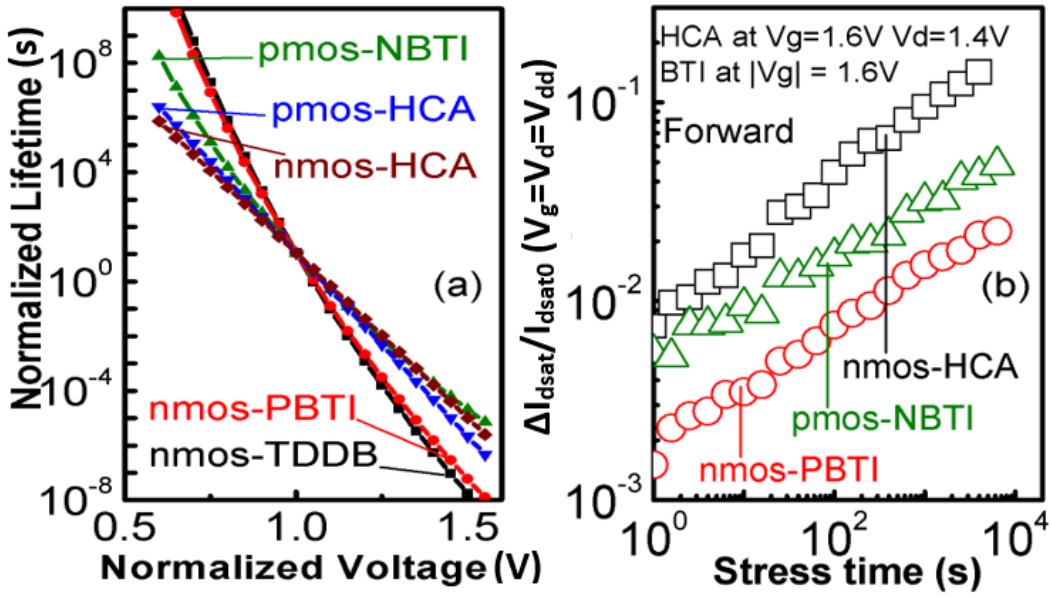


Figure 4.1 A comparison of Hot Carrier Aging (HCA) with BTIs reported by early works. (a) and (b) are re-plots of data from refs. [63] and [125], respectively.

HCA can cause damage due to:

(i) Channel length down-scaling enhances HCA as shown in Figure 4.2(a). With the reduced gate length, effect of HCA is increased. For some sub-30 nm processes, HCA degradation can be higher than BTI as in Figure 4.1(b) and Figure 4.2(b).

(ii) HCA can have larger time exponents (Figure 4.1(b) and Figure 4.2(b)) [120], [124], [126]–[128] and its importance increases with aging.

(iii) NBTI recovery [129] is higher than HCA as shown in Figure 4.2(c). This makes HCA more important for AC operation, where NBTI recovers.

(iv) Conventionally, the worst HCA occurs during switch near $V_g \sim V_d/2$ and duty factor (DF) is typically low (1~2%) [129][130]. For modern CMOS, however, more damage occurs under $V_g=V_d$ (Figure 4.3) [120], [126], [128] and DF can be high. For example, during 'read 0' in an SRAM cell, one access nMOSFET.

can suffer HCA for ~50% of time (Figure 4.4).

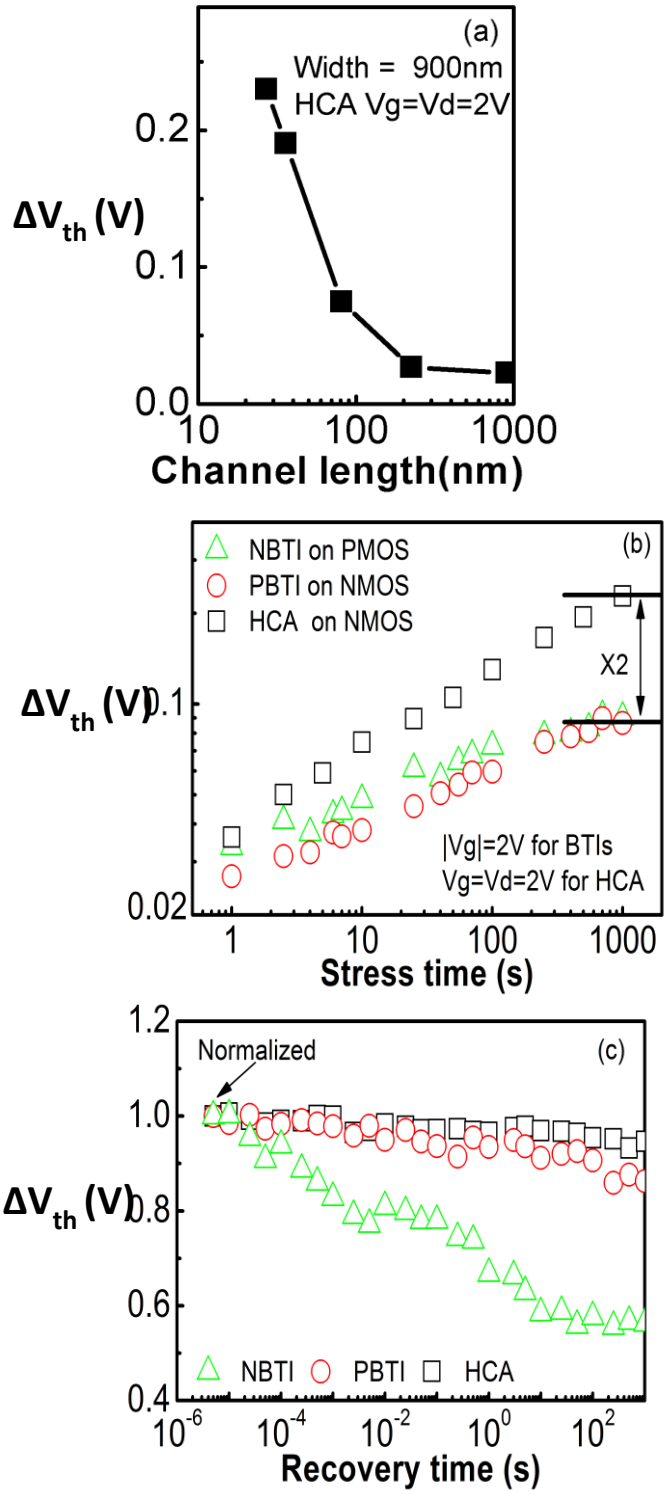


Figure 4.2 (a) Downscaling L increases HCA. The stress was at 125 °C for 1000 sec. (b) comparison of HCA and BTIs for L=27 nm used in this work. Stresses were under the same $|V_g|$, with $V_d=V_g$ for HCA and $V_d=0$ for PBTI and NBTI. (c) A comparison of their recovery under $V_g=V_d=0$.

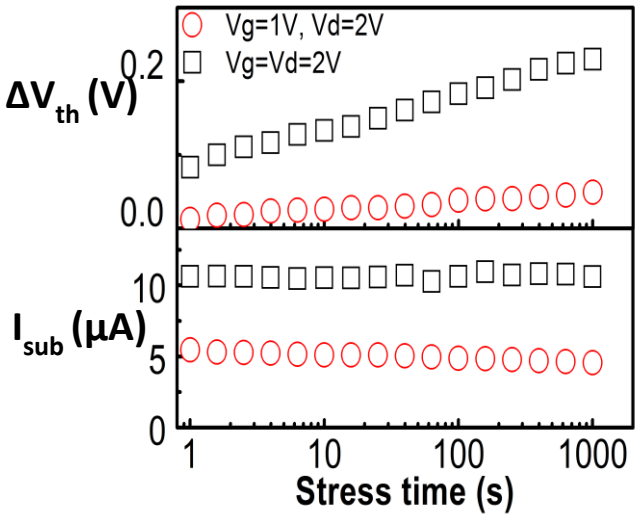


Figure 4.3 HCA under V_g=V_d is more than HCA under V_g=V_d/2 for L=27 nm.

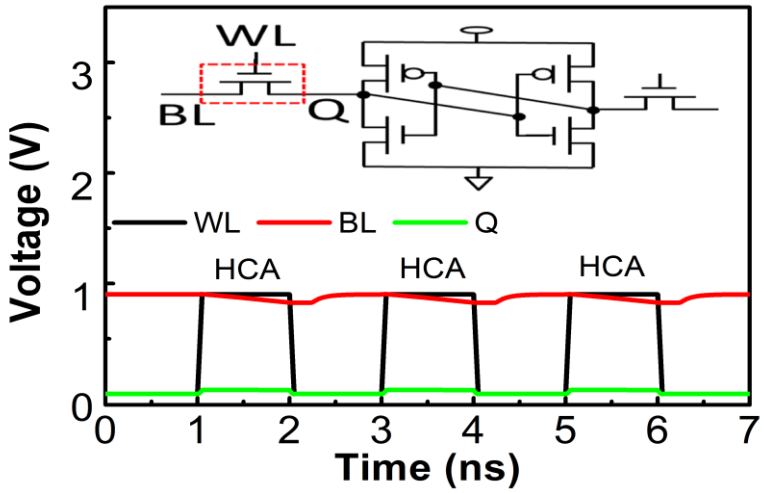


Figure 4.4 The access nMOS in a SRAM during read-0 has a HCA duty factor of ~50%. When transistor flips from OFF to ON, or vice versa, HCA happens.

Static random access memory (SRAM) is the majority of transistor in modern microprocessor. The most dominating wearout mechanism that resulted in increase of threshold voltage of a transistor is BTI and HCA. HCA would affect transistor when it

flips from being OFF to ON or vice versa. While NBTI affects pMOS transistor when gate is applied LOW and PBTI affects nMOS when the gate is applied HIGH [131]. This make HCA is more important to study in SRAM.

The renewed HCA-threat has motivated its re-visit [63], [64], [124]–[128][132]. It is that reported aging mechanisms and time exponent, '*n*' (Eq. 4.3), are different under different stress biases [64], [125], [128]. '*n*' can also vary with time (e.g Figure 4.5) [63], [64], [124], [127], challenging the lifetime prediction based on extrapolation that requires a constant '*n*' [130], [133], [134]. The recent works have focused on bias-accelerated HCA [63], [64], [124]–[128][132] and there is little data on HCA under use-bias. For test engineers, two pressing questions are: can lifetime under use-bias still be predicted by the established JEDEC method based on extrapolation and how to evaluate '*n*' correctly for HCA?

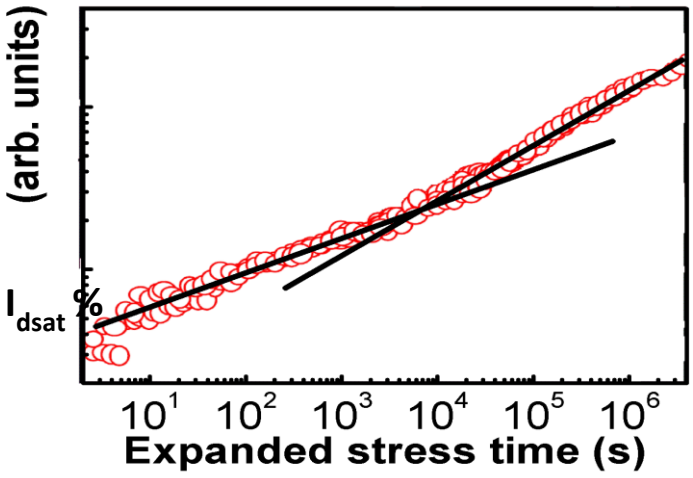


Figure 4.5 Variation of time exponent, '*n*' (the line slope), with HCA time. A re-plot of data from ref. [63].

A key advance of this chapter is answering them and finding the pitfalls for extracting '*n*'. For the first time, the capability of predicting HCA under use-bias is experimentally verified (Figure 4.6).

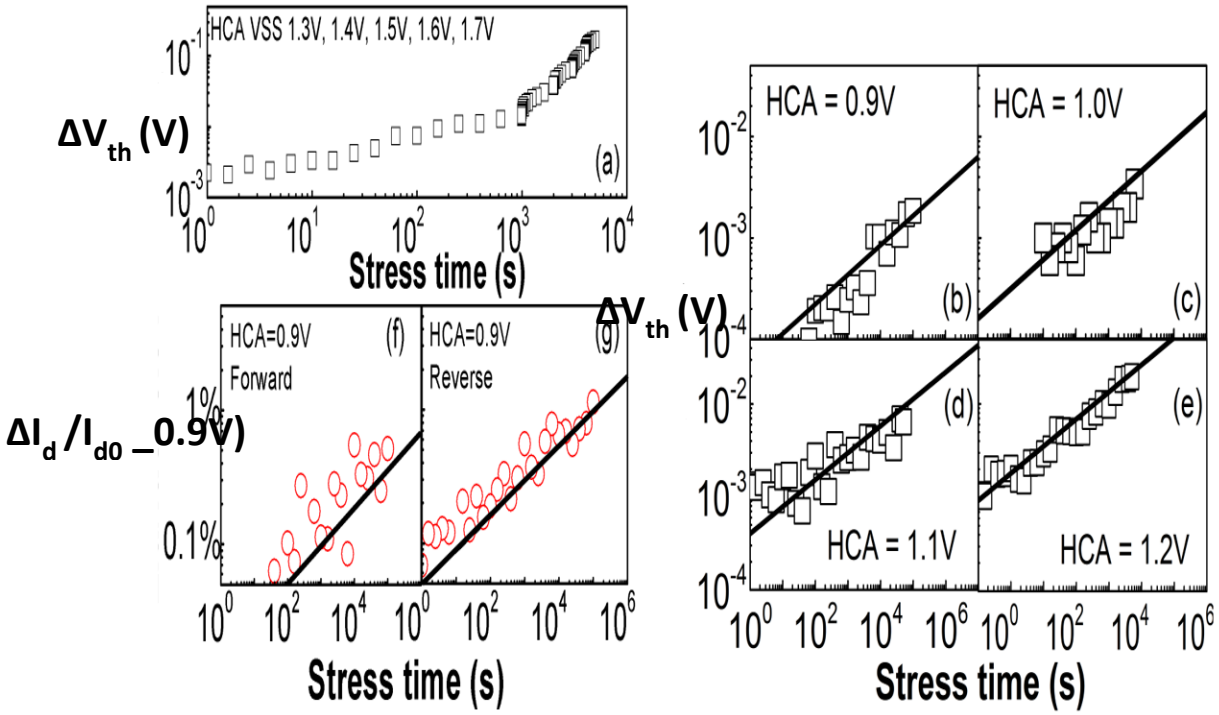


Figure 4.6 Verification of predicting HCA under use-V_{dd}. The model was extracted from accelerated VSS test data given in (a) with V_g=V_d rising from 1.3 to 1.7 V (For details, see Figure 4.14). The symbols in (b)-(e) were measured from 4 devices and not used for fitting. The lines in (b)-(e) are the predicted HCA. The lines in (f) and (g) were obtained from Figure 4.10 by converting ΔV_{th} to $\Delta I_d/I_d$(V_g=V_d=0.9 V).

4.2 Devices and Experiments

Devices used in this investigation are supplied by CSR plc. It was fabricated using 28 nm high-k metal gate technology at Taiwan Semiconductor Manufacturing Company (TSMC). The 12-inch wafer consists of nMOS and pMOS devices with various gate

length and width. For this investigation, the devices used have gate size of 900 nm (W) x 27 nm (L) and 90 nm (W) x 27 nm (L). A device with a gate width of 900 nm is considered a large device while a device with a gate width of 97 nm is considered a small device. Figures 4.7(a) and (b) show the CSR wafer and its use-V_{dd} is 0.9 V. V_d=V_g is chosen for the hot carrier stress, as I_{sub}/I_d has a device-to-device variation (DDV) at stress-0 for nm-devices as shown in Figure 4.12(a). DDV also does not correlate with that of HCA-induced $\Delta I_d/I_d$ as depicted in Figure 4.12(b). In this chapter, all tests were done at a temperature of 125^{o}C unless stated otherwise. The structure and layout of devices were not shown due to confidentiality.

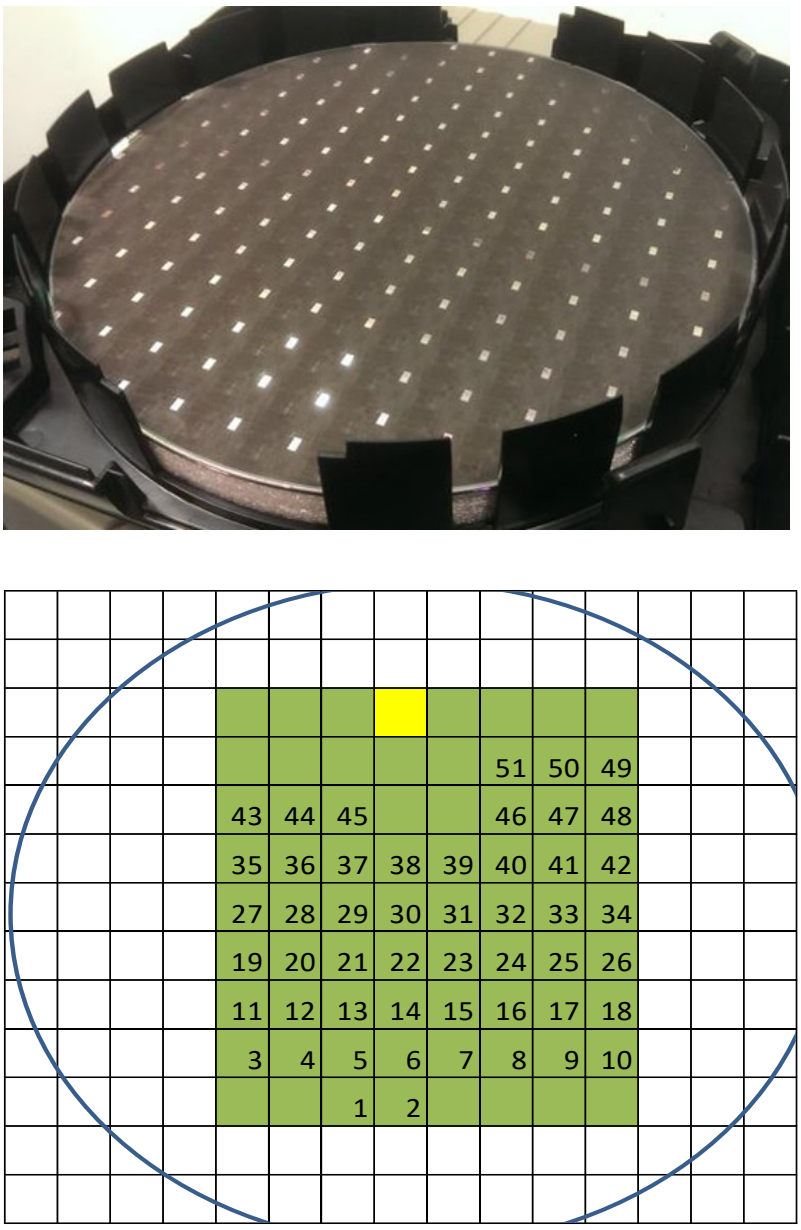


Figure 4.7 (a) CSR wafer used in our investigation. The shiny parts are the devices. Devices with gate width of 900 nm and 90 nm are selected. (b) Location of device that has been tested.

In order to confirm slow measurement (SMU) can be used to measure HCA degradation, a device of similar size (900 nm x 27 nm) is stressed with HCA (V_g=V_d=2 V) on nMOS, PBTI (V_g = 2 V) on nMOS and NBTI (V_g=-2 V) on pMOS for 1000 s. Then, the recovery

of the device is measured. The result in Figure 4.8 showed that after HCA stress, there is no significant recovery. For PBTI and NBTI, there is recovery as we expected. Based on the results, slow measurement using SMU can be used to measure HCA on a large device.

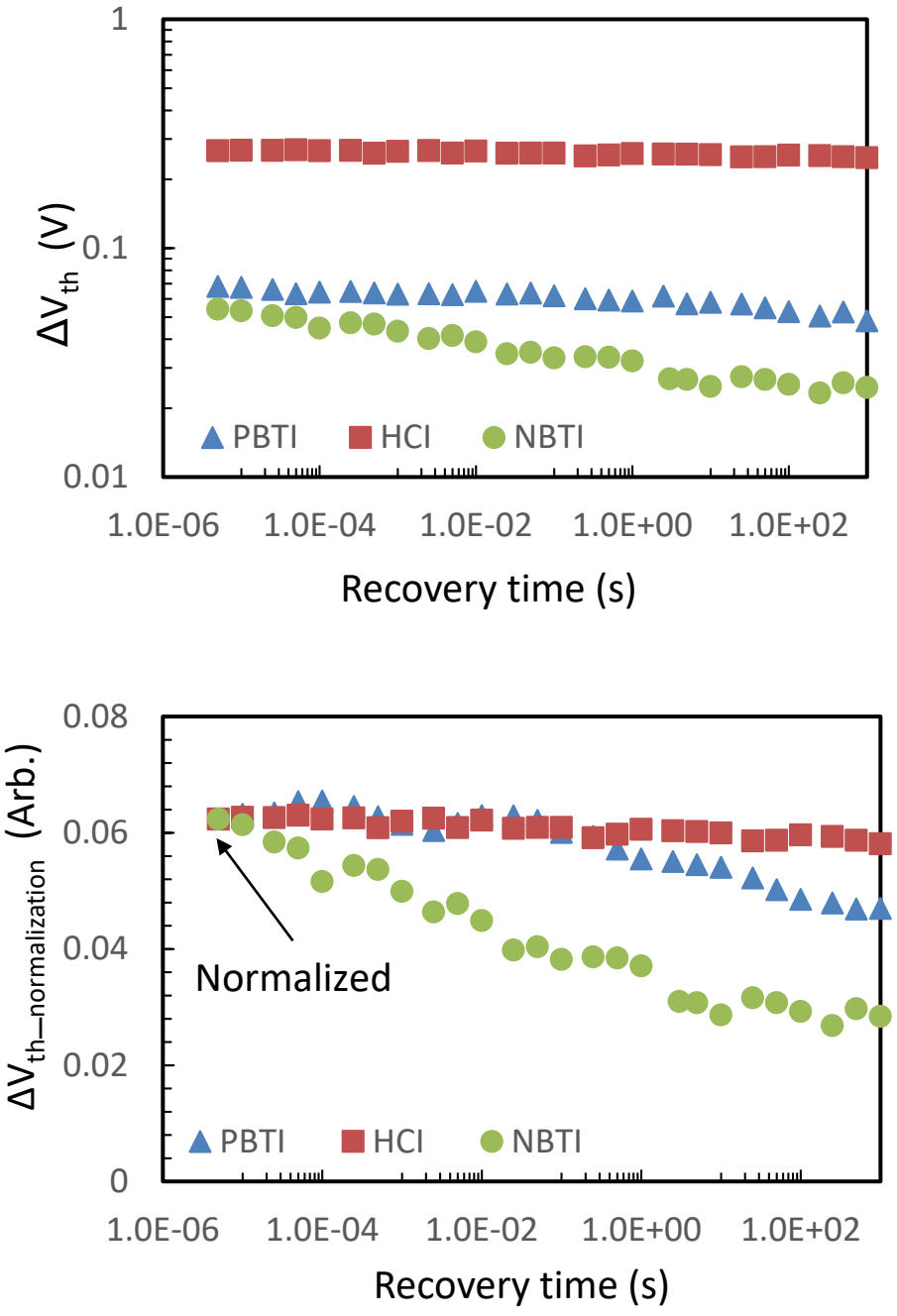


Figure 4.8 Recovery of large device after HCA, NBTI and PBTI

Figure 4.9 shows the variations of 50 of 27 nm (length) x 90 nm (width) devices before stress. It can be seen that the as-fabricated threshold voltage can fluctuate by ~ 100 mV between 0.35 and 0.45 V.

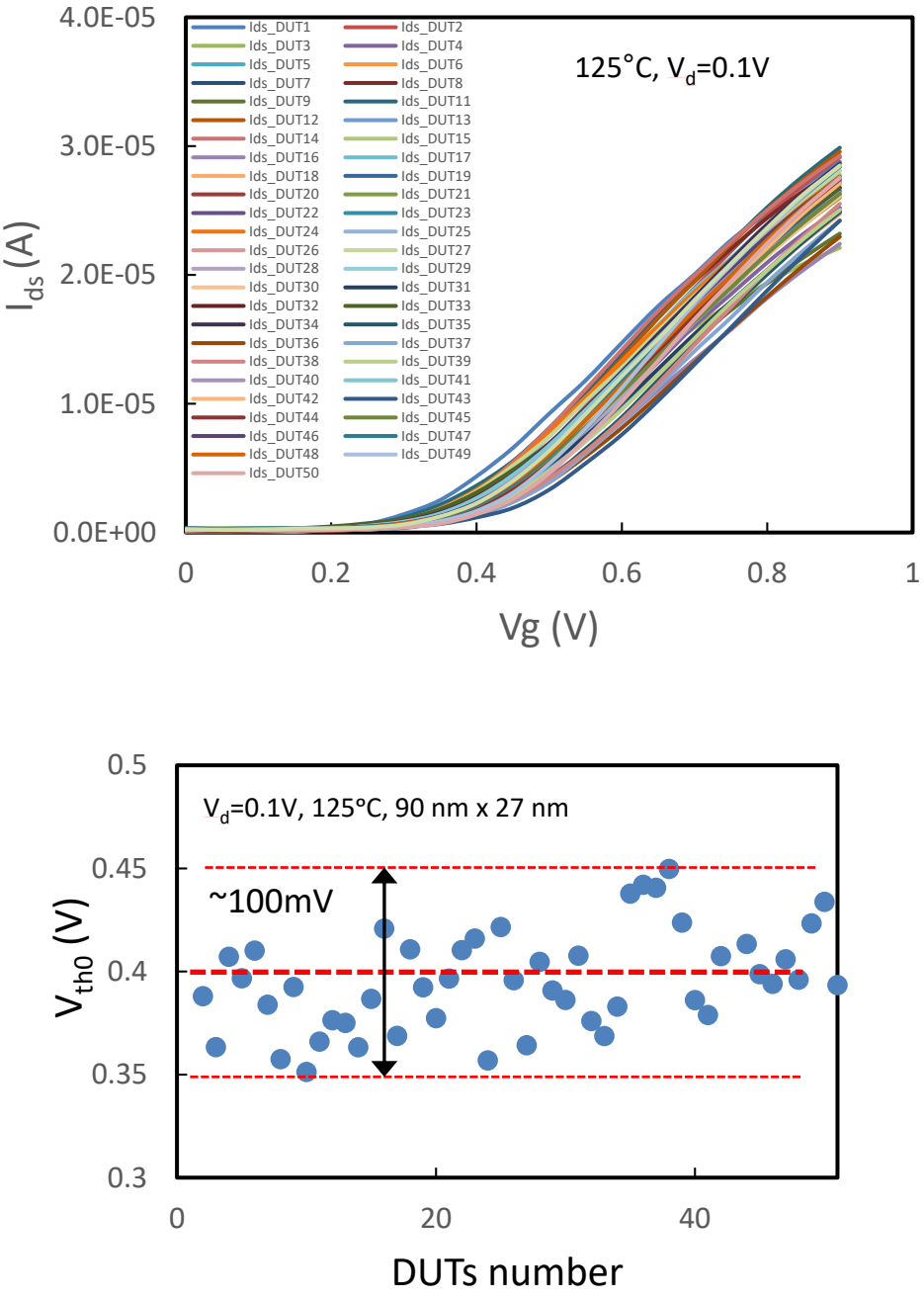


Figure 4.9 (a) I-V curve for 50 DUTs. (b) Variation of V_{th0} for 50 DUTs

4.3 Prediction of HCA Lifetime on Large Device

Wide devices with size of 900nm (width) x 27 nm (length) are used in this test.

4.3.1 Select the parameter representing stresses

Lucky Electron Model in the form of equations (4.1) and (4.2) by C. M. Hu proposed that HCA stress can be represented by the ratio of substrate to drain current I_{sub}/I_d [8]. Hot carriers break the Si-H bond at the interface, as illustrated in Figure 4.10.

$$\frac{I_{sub}}{I_d} = C_1 \cdot \exp(-\varphi_i / q\lambda E_m) \tag{4.1}$$

$$\Delta N_D = C_3 \cdot t^n \cdot (\frac{I_d}{W})^n \cdot (\frac{I_{sub}}{I_d})^{n \cdot \varphi_{it}/\varphi_i} \tag{4.2}$$

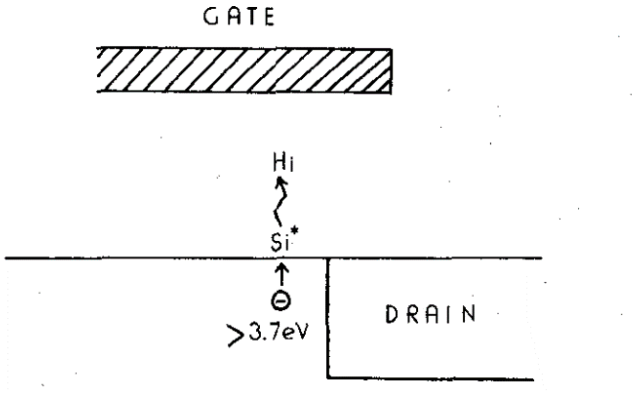


Figure 4.10 A physical model for interface-traps generation [8].

Figure 4.11(a) shows that, for a wide device of 900 nm x 27 nm, there is a good correlation between HCA and I_{sub}/I_d, justifying its use for representing the stress. For smaller devices, Figure 4.11(b) shows that their average I_{sub}/I_d agrees well with the wide device and their average HCA also correlates well with the mean I_{sub}/I_d. This may indicate that I_{sub}/I_d can

be used as a parameter for representing HCA stress of small devices, but its problem will be shown in the following.

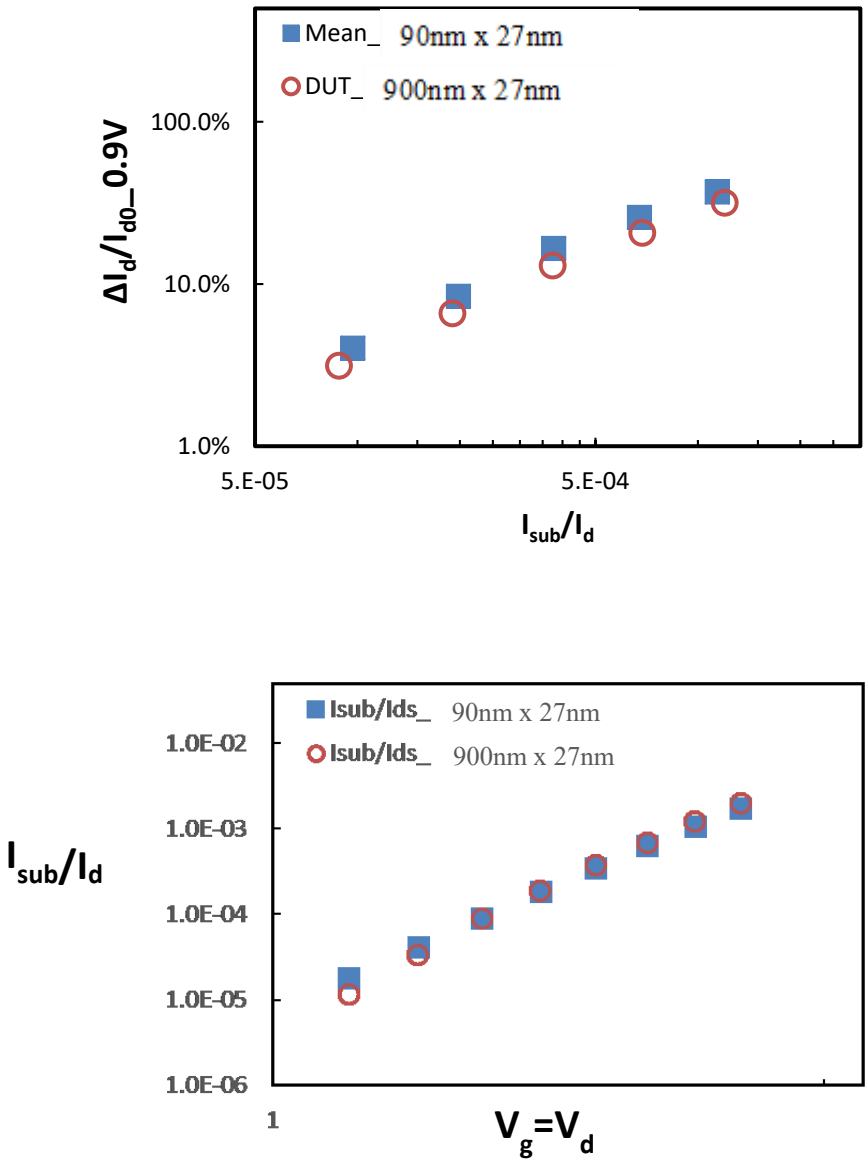


Figure 4.11 Good agreement of I_{sub}/I_d for large and small device

Figure 4.12 clearly shows there is a device-to-device variation in the as-fabricated I_{sub}/I_d for small devices, but there is no correlation between it and the HCA. Thus, I_{sub}/I_d cannot be directly used to characterize small device HCI degradation.

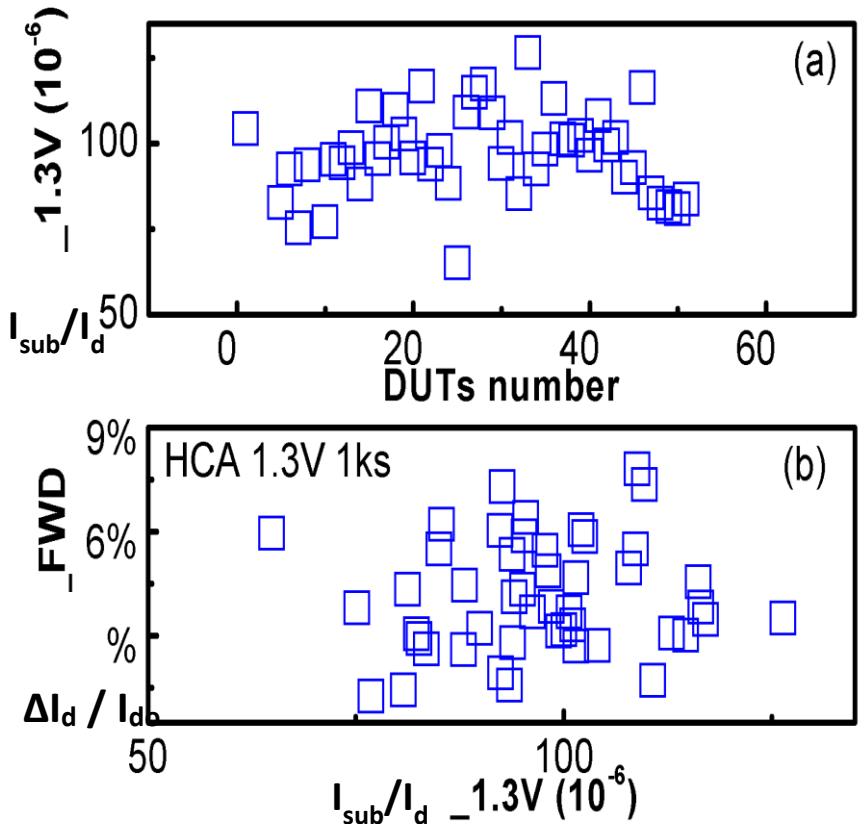


Figure 4.12 I_{sub}/I_d does not represent HCA-stress well for nm-devices, as it has a device-to-device variation (DDV) at stress=0 (a) and its DDV does not correlate with that of HCA-induced $\Delta I_d/I_d$.

4.3.2 Select HCA acceleration

Figure 4.13 shows that HCA under the used V_{dd}=0.9 V is too low for establishing a reliable aging kinetics within a practical test time, so that acceleration is needed. HCA can be accelerated in two ways: by raising V_d under a constant V_g and by raising both V_g and V_d with V_g=V_d. The problem is that the time exponents obtained by these two methods is different, as shown in Figure 4.14 and which acceleration method should be used.

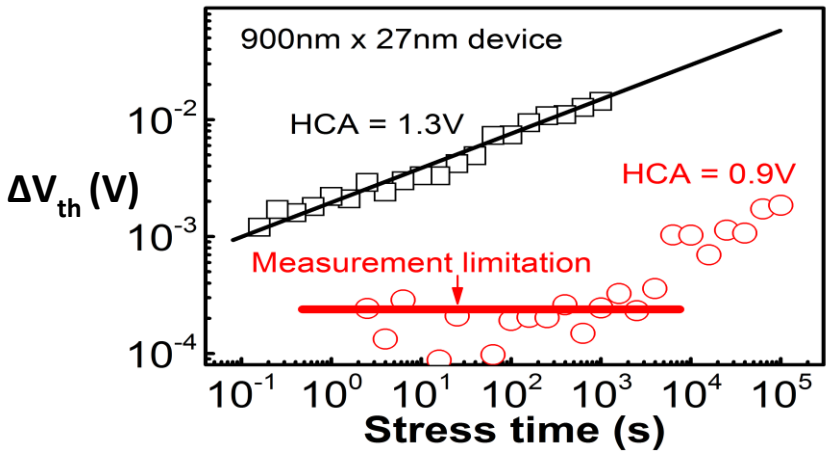


Figure 4.13 HCA is too low to establish kinetics reliably under use-V_{dd}=0.9 V and acceleration (e.g. 1.3 V) is needed.

SRAM often is used for qualifying new processes [132], where the access nMOSFETs suffer the worst HCA under $V_g \approx V_d$ (Figure 4.4). As a result, HCA under use-$V_g=V_d$ must be predicted, so that the acceleration should be carried out under $V_g=V_d$. Figure 4.14(b) confirms '*n*' is bias-independent under $V_g=V_d$, meeting the requirement for prediction from accelerated tests [18].

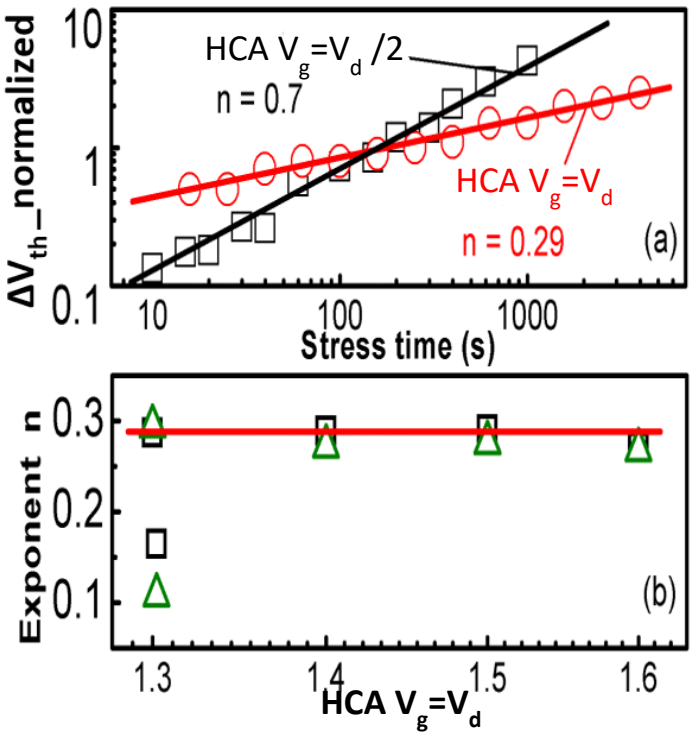


Figure 4.14 (a) The time exponents under $V_g=V_d$ is smaller than under $V_g=V_d/2$. (b) The time exponent is insensitive to stress biases under $V_g=V_d$

4.3.3 DC versus AC

Unlike NBTI (AC) < NBTI (DC), the AC and DC HCAs agree well, regardless of frequency and duty-factor (DF) for the same equivalent stress time, i.e. DF × time as shown in Figure 4.15, confirming the frequency-independence [3]. As a result, DC will be selected for convenience.

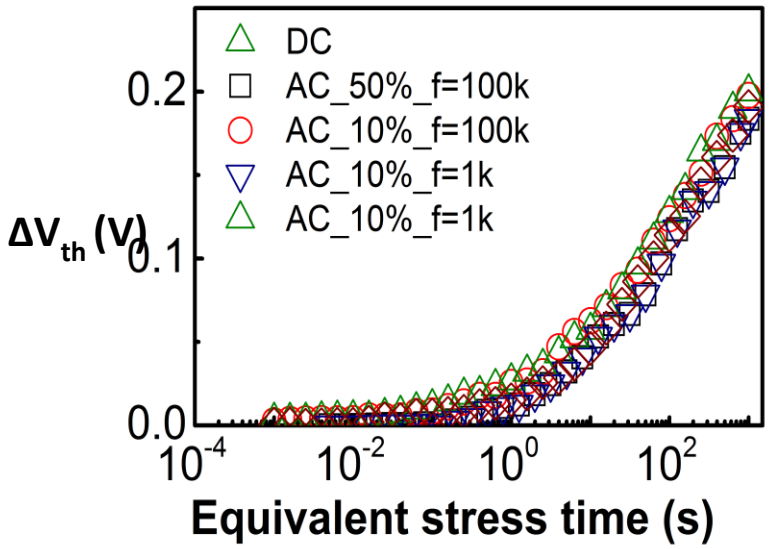


Figure 4.15 The AC and DC HCAs agree well when using equivalent stress time, i.e. time×Duty Factor (DF). The AC stress conditions are given in the format of 'AC_DF_Frequency' and V_g=V_d=2 V.

4.3.4 Application of VSST technique

BTI is a major source of degradation. Rapid recovery of BTI happens when stress voltage is removed. In industrial practice for qualification screening, attention is focused on degradation after recovery. BTI degradation can be expressed by a power law against time and voltage in equation (4.3) [135]:

$$\Delta V_t = A \ V_g^m \ t^n \tag{4.3}$$

where A is a pre-factor, n is the time exponent and m is the voltage exponent. Lifetime under nominal operation can be predicted if these parameters are known. Conventionally, they are extracted from accelerated tests under constant voltage stress (CVS).

Characterization of HCA degradation is also done by the constant voltage stress (CVS) method [JEDEC 2011]. However, the disadvantages of the CVS method are that it requires multiple identical devices and will take days to complete a single set of tests.

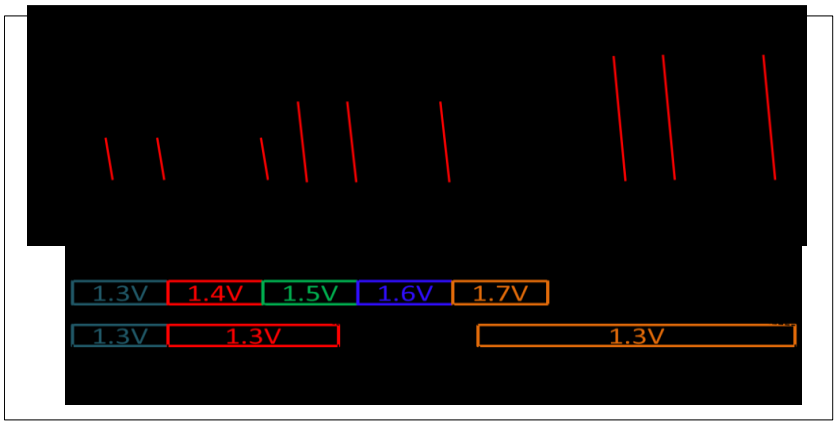
Voltage Step Stress Technique (VSST) was introduced by Z. Ji *et al* [133]. This method was developed to predict the lifetime of a device under nominal operation. It is initially developed for NBTI. The technique improves the CVS method by significantly shortening the time taken to complete a test and reducing numbers of devices needed to only one device.

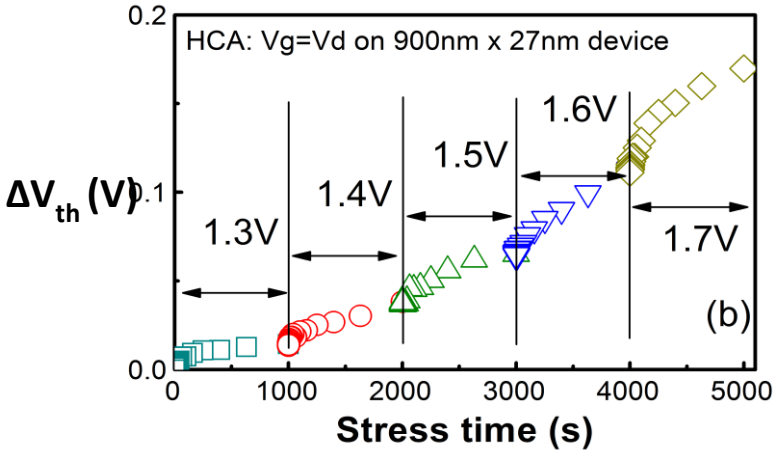
Stressing the device at higher stress gate voltage, V_{gst} for the same time, ΔT, is more effective to generate defects compared with only stress at V_1 in Figure 4.16. For stress time of ΔT under high V_{gst}, degradation can be equivalent to that under a V_1 for an effective longer stress time, ΔT_{eff} [133], [136]. ΔT_{eff} can be evaluated by equations (4.4) and (4.5),

$$\Delta T_{eff} = \Delta T \cdot \left(\frac{V2}{v1}\right)^{m/n} \tag{4.4}$$

$$A \cdot \Delta T^{n} \cdot {V_2}^{m} = A \cdot {\Delta T_{eff}}^{n} \cdot (V_1)^{m} \quad (V_2 > V_1) \tag{4.5}$$

For an L × W=27 nm × 900 nm (large device), stress under each V_g=V_d lasted for T=1ks and biases then stepped up as in Figure 4.16, lifting HCA up from the power law as shown in Figure 4.16(c).





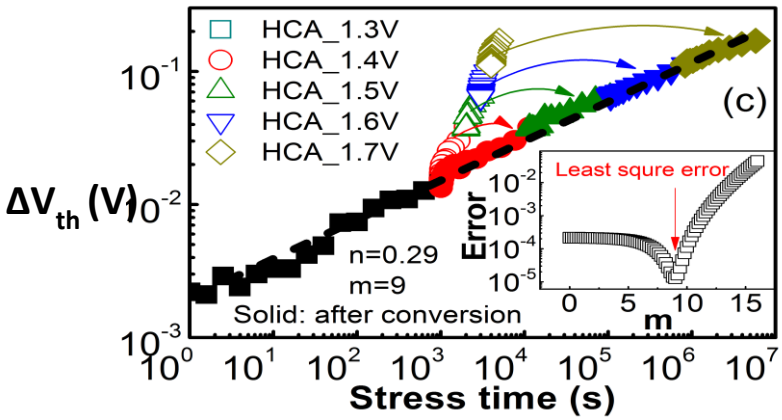


Figure 4.16 Voltage-Step-Stress (VSS) technique for HCA. (a) One device was stressed for a time T and the stress V_g=V_d was then stepped up. ΔV_{th} is plotted against linear (b) and log (c) stress time. The stress time under high bias is converted to an equivalent longer time at low bias by fitting the voltage exponent 'm' (inset of (c)) through Eqs. 4.3-4.5. The dashed line has n=0.29 and m=9.

Based on equations (4.3)-(4.4), HCA under a high V_g=V_d was converted into a longer equivalent stress time under a low V_g=V_d as shown in Fig. 4.16(c) and ΔV_{th} follows a power law well beyond the 10% HCA lifetime criterion and allowing reliable extraction of '*n*' and '*m*'.

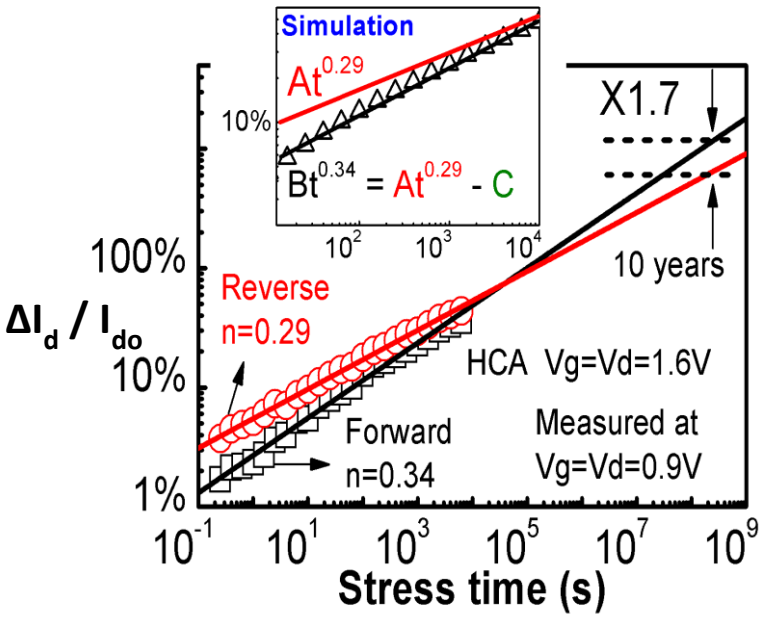
A typical result of a device measured with VSST is shown in Figure 4.16(c). The device is stressed with V_1 at 1.3 V (V_g=V_d=1.3 V) for 1,000 s and continuously stressed with V_2, V_3…V_N of 1.4 V, 1.5 V, 1.6 V and 1.7 V. The time period for each voltage is 1,000 s. During the HCA stresses, IV measurements are taken periodically.

The value of voltage exponent, *m*, is fitted by converting the VSS data in Figure 4.16(c) to a power law with the least square error, as shown by the inset of Figure 4.16(c).

4.3.5 Selecting parameter for extracting '*n*'

HCA was widely monitored by the shift in forward saturation current $\Delta I_d/I_d$_F, although reverse $\Delta I_d/I_d$_R and ΔV_{th}($V_d \leq 0.1$ V) also were used [50], [63], [64], [71], [124]–[128]. The issue is '*n*' for $\Delta I_d/I_d$_F is larger than '*n*' for $\Delta I_d/I_d$_R, leading to their incorrect cross-over and errors in prediction at 10 years (Figure 4.17), highlighting the importance of '*n*'-accuracy. Under V_g=V_d, $\Delta I_d/I_d$_F does not sense the HCA-defects above the space charge region (Figure 4.17), resulting in an apparent larger '*n*', as simulated by subtracting a constant from real power law (inset of Figure 4.17). The '*n*' extracted from forward $\Delta I_d/I_d$

is erroneous, therefore. To capture all defects, $\Delta V_{th}(V_d=0.1V)$ should be used for extracting '*n*', as $\Delta V_{th}\_F=\Delta V_{th}\_R$ as shown in Figure 4.18. Once ΔV_{th} is predicted, $\Delta I_d/I_d\_F$ and $\Delta I_d/I_d\_R$ can be evaluated from their measured relation with ΔV_{th} as referred to in Figure 4.19.



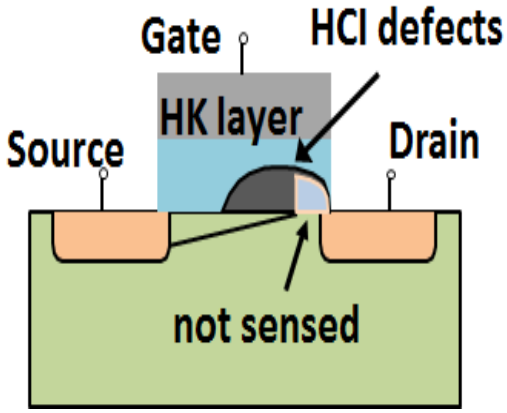


Figure 4.17 Although test data (**o** and □) show $(\Delta I_d/I_d\_F)<(\Delta I_d/I_d\_R)$, higher '*n*' for $\Delta I_d/I_d\_F$ leads to incorrect $(\Delta I_d/I_d\_F)>(\Delta I_d/I_d\_R)$ when extrapolating. $\Delta I_d/I_d\_F$ does not sense the defects above space charges. The 'Δ' in inset is calculated from ($At^{0.29}$-Constant), which fits well with $Bt^{0.34}$ (**black line**). Subtracting a constant from a real power law can give an 'apparent' higher '*n*'. I_d was measured under $V_g=V_d=0.9$ V.

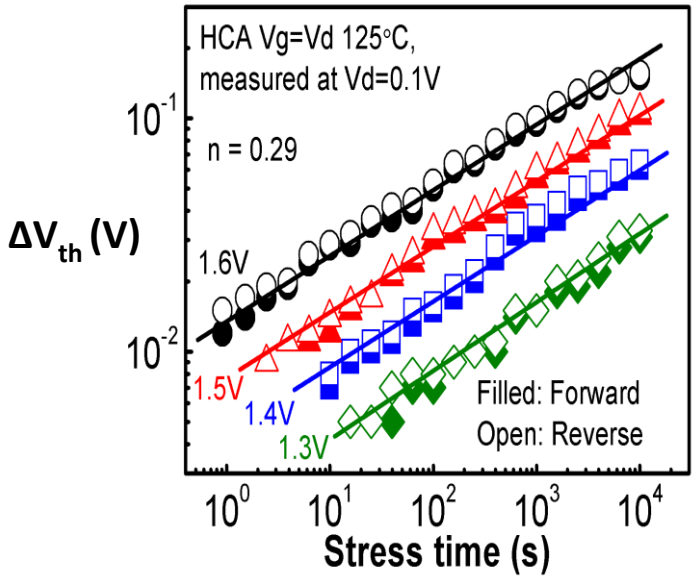


Figure 4.18 The forward and reverse ΔV_{th} measured under V_d=0.1 V agrees well.

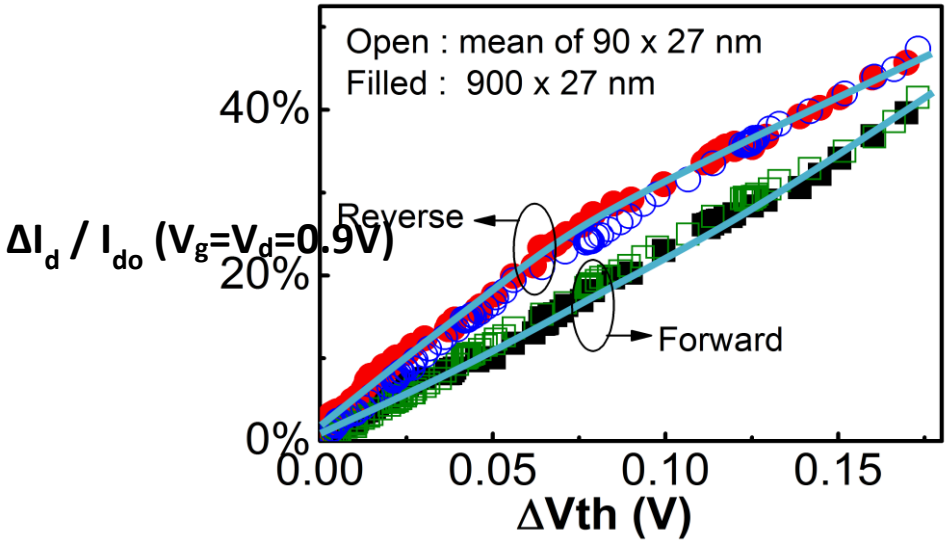


Figure 4.19 Relation of HCA-induced ΔV_{th} under V_d=0.1 V with $\Delta I_d/I_d$ under V_g=V_d=0.9 V. The open symbols are the mean of 50 small devices (90 × 27nm) and the filled symbols are the mean of 50 large devices (900 nm×27 nm).

4.3.6 Prediction

The data obtained from the experiment in Figure 4.16 is used to verify our prediction based on the VSS technique. It agrees well with the test data in Figure 4.6(b)-(g). Both figures show the highest ΔV_{th} is ~2-orders above ΔV_{th} under 0.9 V (Figure 4.6(b) and this verify its prediction capability. Test data from Figure 4.6(b)-(g) were not used for fitting and the model was extracted only from Figure 4.6(a). The extracted model from Eq. 4.3 can be used for evaluating HCA under any bias and time. It can also be used for predicting lifetime and operation V_{dd}. The device lifetime is the time for ΔV_{th} reaching 20-50 mV [137], [138], under operation voltage the required lifetime is 10 years [18]. Figure 4.20 shows evaluation of lifetime for forward and reverse mode. Based on the result plotted for lifetime in Figure 4.20, it can be concluded to achieve 10 years lifetime, the operation voltages are 0.9V for the reverse mode and 1.0V for the forward mode.

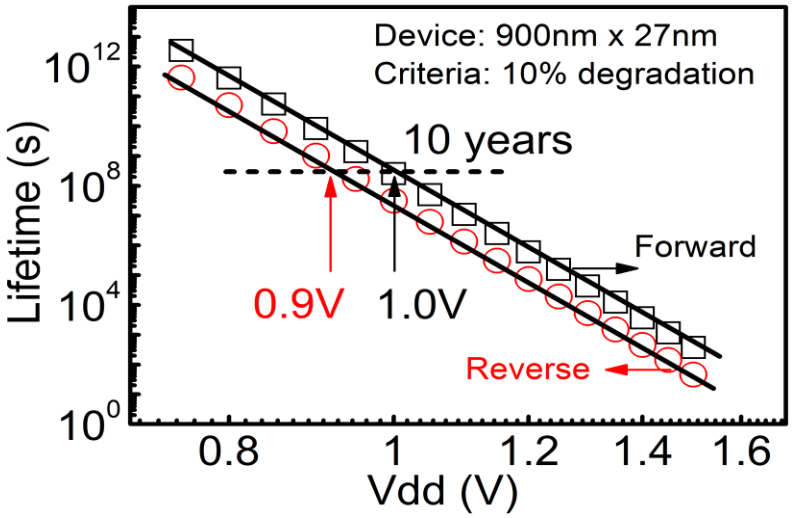


Figure 4.20 Evaluation of lifetime vs. V_{dd} based on the model extracted from VSS technique as in Figure 4.16.

4.4 Prediction of HCA for small devices

4.4.1 Characterizing HCA in nm-width devices

A Scaling device provides a handful of advantages but it also introduces a few reliability problems such as random telegraph noise (RTN). In scaled devices also we can observe large device-to-device variation as shown in Figure 4.9. Unlike L×W=27 × 900 nm devices, 27 × 90 nm devices suffer from RTN-like within-a-device fluctuation (WDF) and large DDV as shown in Figure 4.21.

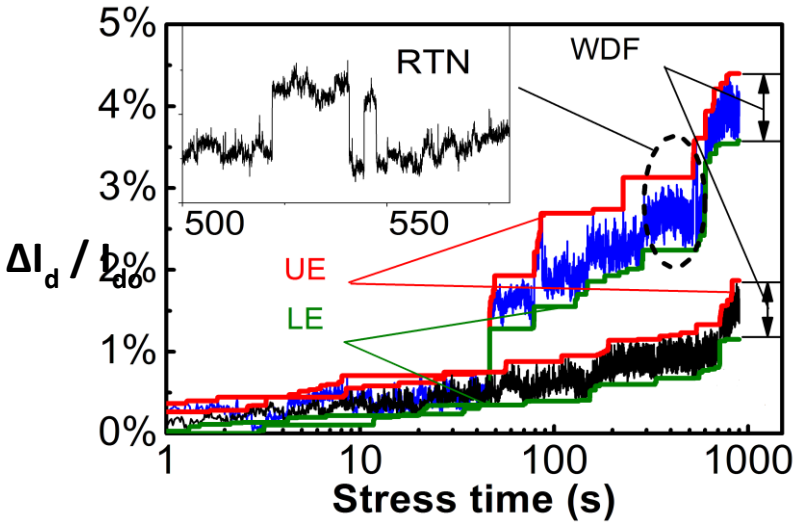


Figure 4.21 HCA of two W=90 nm devices shows large DDV. WDF, UE, and LE is 'within-a-device-fluctuation', the upper- and the lower- envelope.

To extract HCA kinetics, one has to use the smooth mean of 50 devices, but '*n*' depends on how data is taken (Figure 4.22(a)). After a stress, ΔV_{th} fluctuates and one can use its up-envelope (UE), lower-envelope (LE) [134], or average over a period of time, e.g. ~10ms (Figure 4.22(b)), as a typical quasi-DC Source-Measure-Unit (SMU) does. The

'*n*' from UE and DC (inset of Figure 4.22(a)) is smaller than the '*n*=0.29' from W=900 nm (Figure 4.16(c), but the '*n*' from LE agrees well with it. The smaller '*n*' for UE incorrectly takes it below LE when extrapolating (see 'cross-over', Figure 4.22(a)).

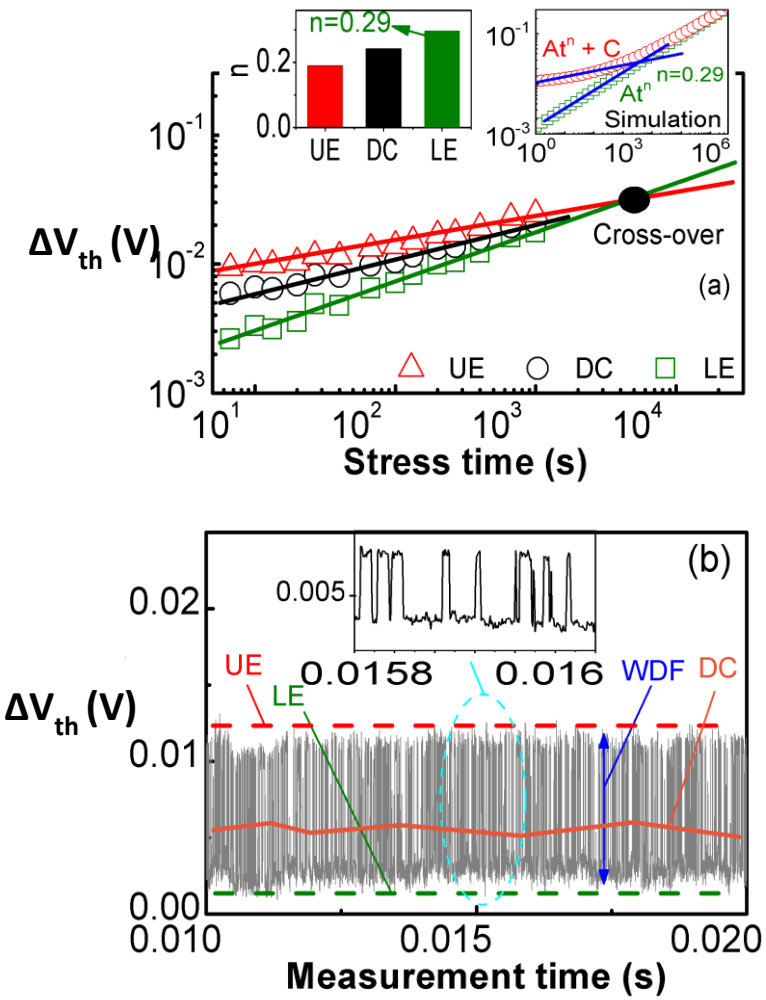


Figure 4.22 (a) HCA kinetics for the mean of 50 of W=90 nm devices. UE, DC, and LE have different '*n*' (inset). Incorrect inclusion of an as-grown component, '**C**', gives an apparent lower '*n*' at short time. (b) The definition of UE, DC, and LE. DC is the average over 10ms.

To explain the whether UE or LE should be used in extracting '*n*', Figure 4.23(a) shows that LE increases with HCA time, but the WDF=(UE-LE) does not increase with aging. This explain that LE is caused by HCA, while WDF is not. WDF must originate from as-grown defects and should be excluded from aging kinetics, so that LE must be used for extracting '*n*'. Figure 4.23(b) supports earlier statement by showing that the WDF mean for 40 devices is constant over HCA time.

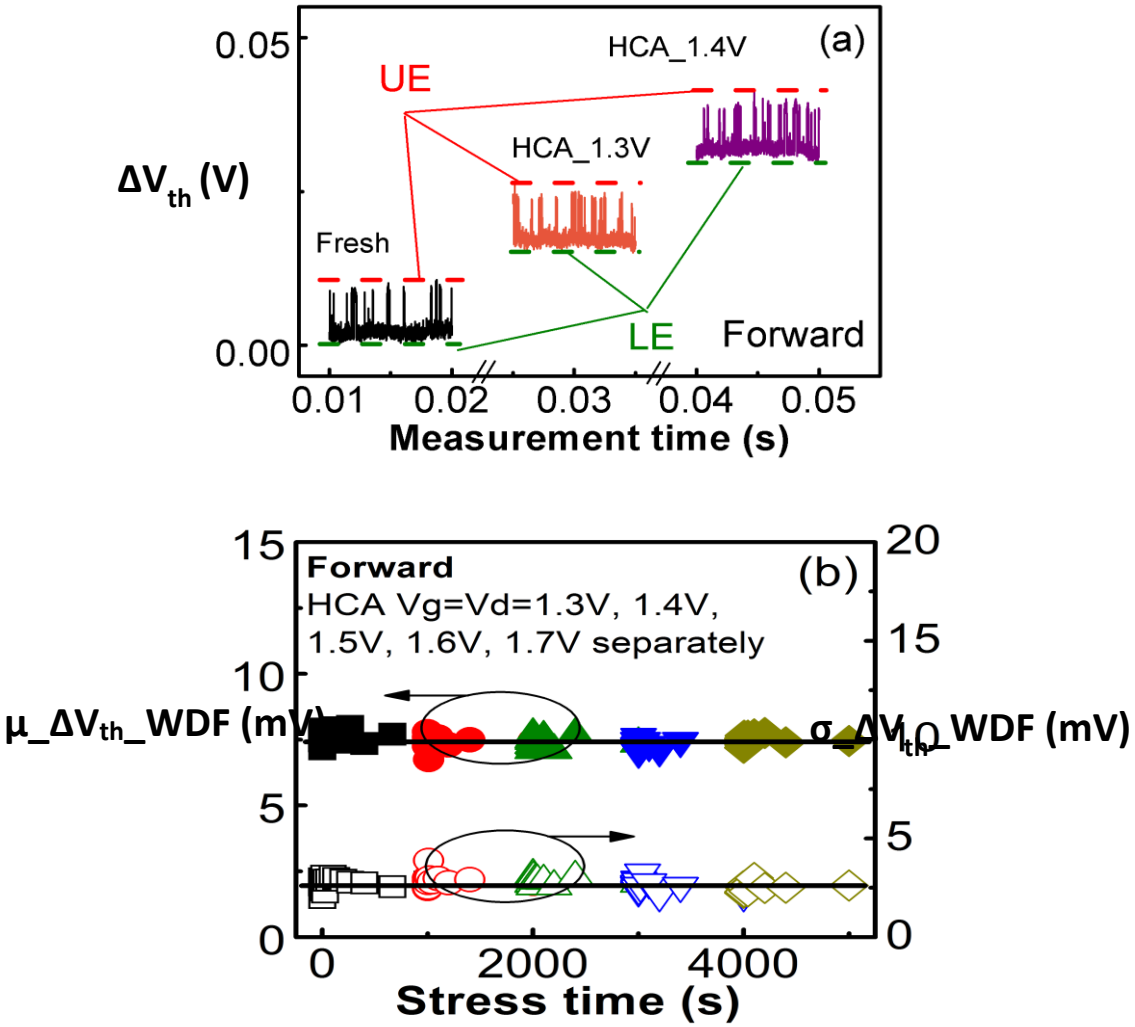


Figure 4.23 (a) For L×W=27 nm × 90 nm, LE increases with HCA, but WDF does not. (b) The WDF_mean of 50 devices and its sigma do not increase with stresses.

To further support that LE and WDF are originates from different sources, Figure 4.24(a) shows LE_F and LE_R correlates, but WDF_F and WDF_R does not as shown in Figure 4.24(b), supporting their different origins.

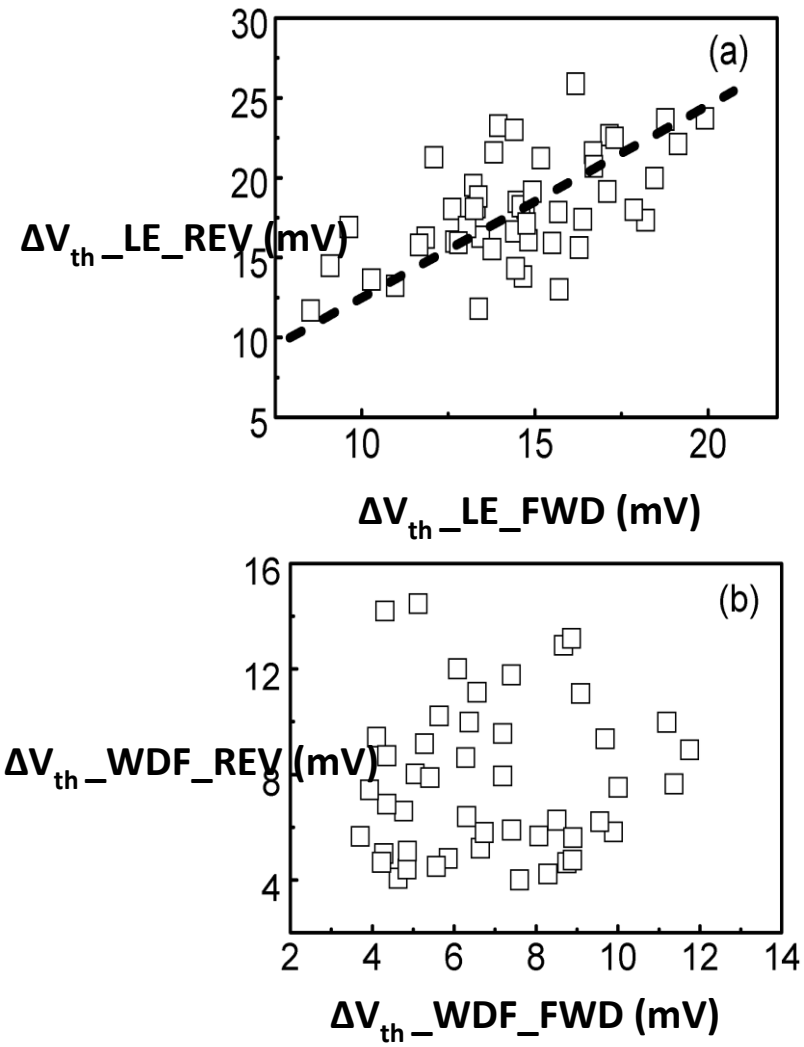


Figure 4.24 (a) LE_F correlates with LE_R. (b) WDF_F does not correlate with WDF_R.

Since HCA-recovery is insignificant as shown in Figure 4.2(c), one may think it can be measured by a quasi-DC SMU [120][122]. This, however, gives an erroneous lower '*n*' (Figure 4.22(a)) by including some as-grown WDF. Adding a constant to a power law

leads to an apparent lower '*n*' at short time and a variation of '*n*' with time (inset, Figure 4.22(a)) [130].

4.4.2 Statistic HCA

50 small devices were tested and Figure 4.25 shows variation of $\Delta I_d/I_{d0}$ at different stress times.

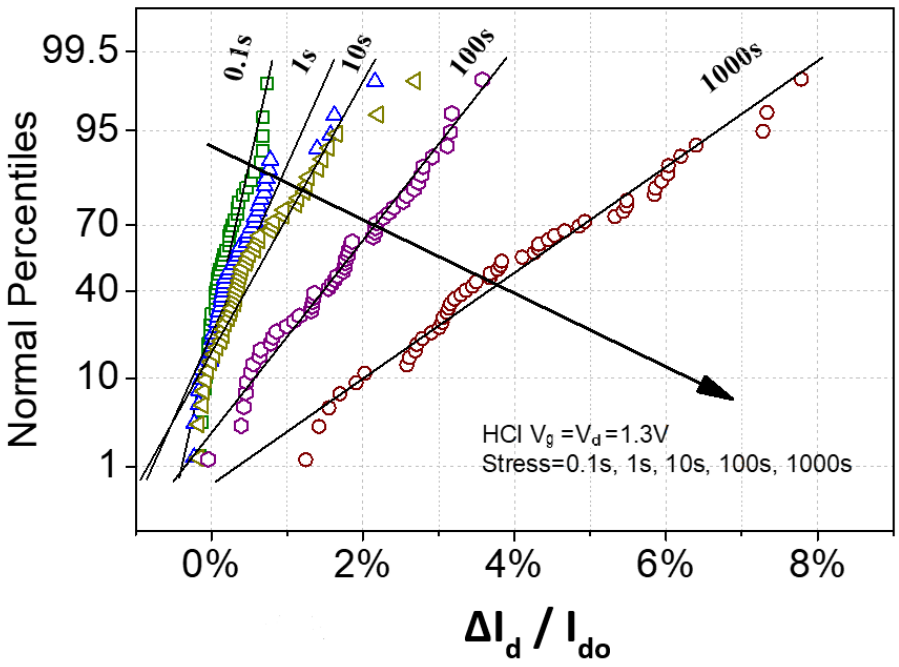


Figure 4.25 Normal probability plot for $\Delta I_d/Id_0$ for different stress times

HCA under different HCI stress voltage also showed large variations, as shown in Figure 4.26. Based on these results, it can be concluded that HCA induces a time-dependent device-to-device variation.

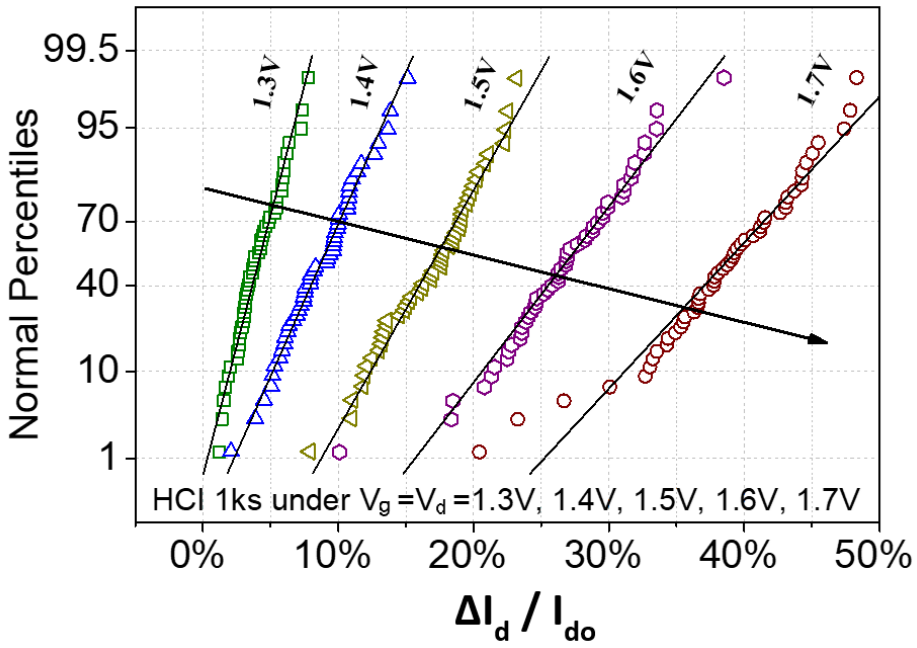


Figure 4.26 Different V_g=V_d of 1,000s stress time.

Defect-Centric distribution, developed to explain variability in BTI, has been reported to be able to describe HCA degradation in nMOSFETs [139][140][141]. The Defect-Centric distribution is based on two assumptions: the threshold voltage, V_{th} produced by a single charged trap has an exponential distribution (with mean value η), while the total number of charged traps per device is Poisson-distributed [142]. The basic equations describing the Defect-Centric distribution are:

$$F_N(\Delta Vth, \eta) = a_0 + \sum_{k=0}^{\infty} \frac{e^{-N} N^k}{k!} F_k(\Delta Vth, \eta) \tag{4.6}$$

$$N = \frac{2\mu^2}{\sigma^2} \qquad \eta = \frac{\sigma^2}{2\mu} \tag{4.7}$$

where F_N is the cumulative Defect-Centric distribution, η is the mean value of ΔV_{th} produced by a single charge (exponential distribution), *N* is the dimensionless mean value of the total number of charged traps per device (Poisson distribution), $\Delta Vth, \eta$ is the expected value of F_N and F_k $(\Delta Vth, \eta)$ is cumulative distribution function [19], [139].

The DDV of LE at different times (Figure 4.27(a) and Figure 4.27(b)) and voltage (Figure 4.27(c) and Figure 4.27(d)) follows the defect-centric distribution (Eqs.4.6 and 4.7) well [139]. LE_mean of 50 W=90 nm agrees well with ΔV_{th} of one W=900 nm (Figure 4.28(a)) and can be predicted by the same method (Figure 4.6 and Figure 4.16). After knowing LE_mean, the standard deviation, σ, can be evaluated from its power law relation with the mean as shown in Figure 4.28(b).

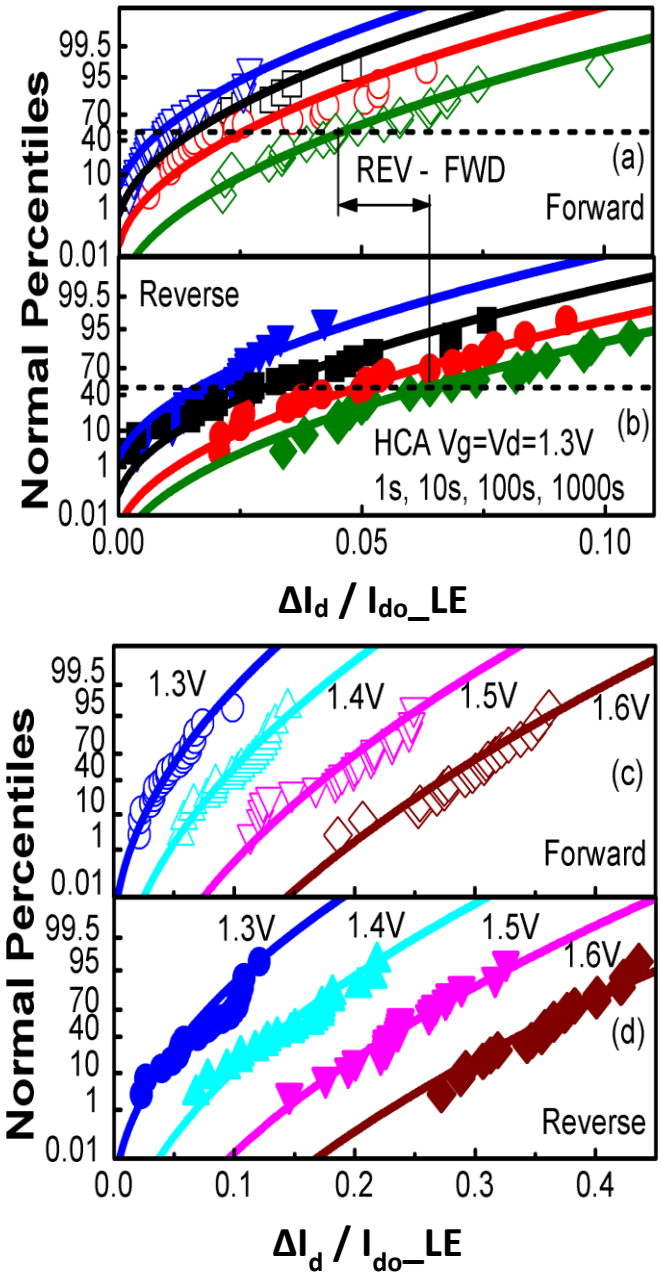


Figure 4.27 Statistics of LE DDV after different stress time (a & b) and voltage (c & d). The lines are fitted with the defect-centric distribution WDF (Eqs. 4.6 & 4.7).

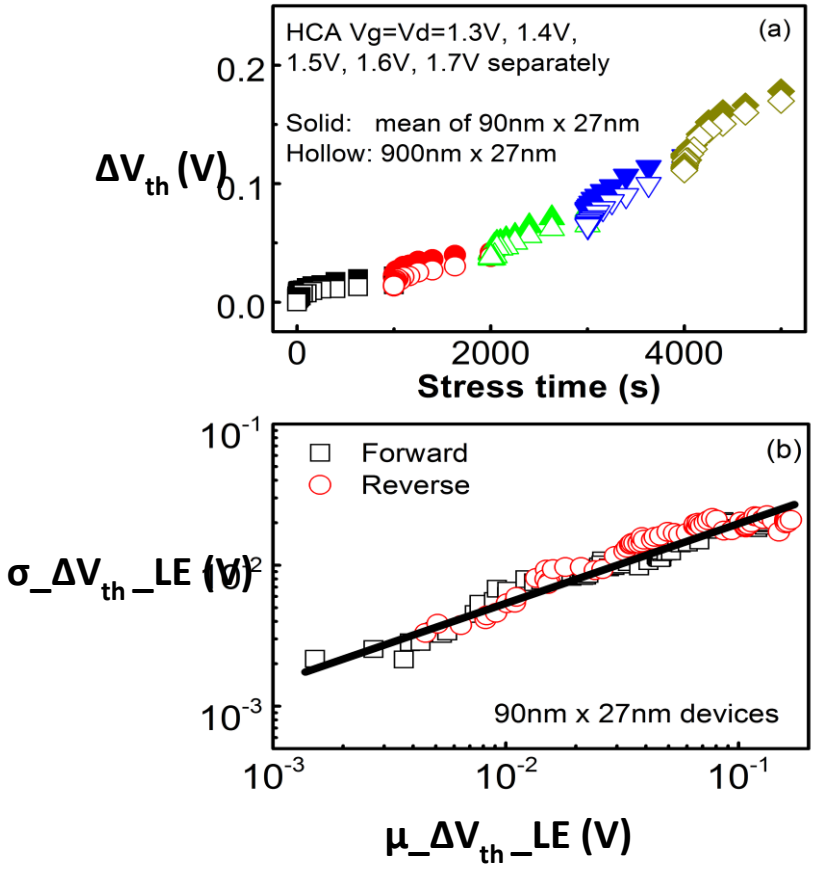


Figure 4.28 (a) The mean of 50 90×27nm agrees well with one 900nm ×27 nm for VSS stresses. (b) Sigma versus mean. The fitted exponent is 0.55, agreeing well with Eq. 4.7.

4.4.3 Impact on use-V_{dd}

To have a yield corresponding to i×σ, $\Delta I_d/I_d$=10% is required at i×σ, resulting in smaller mean value (Figure 4.29(a)) and in turn lower use-V_{dd} (Figure 4.29(b)) for higher i. For a yield of 3×σ (99.7%), HCA-only and HCA+WDF reduces V_{dd} from its zero-spread value by 75 mV and 100 mV, respectively.

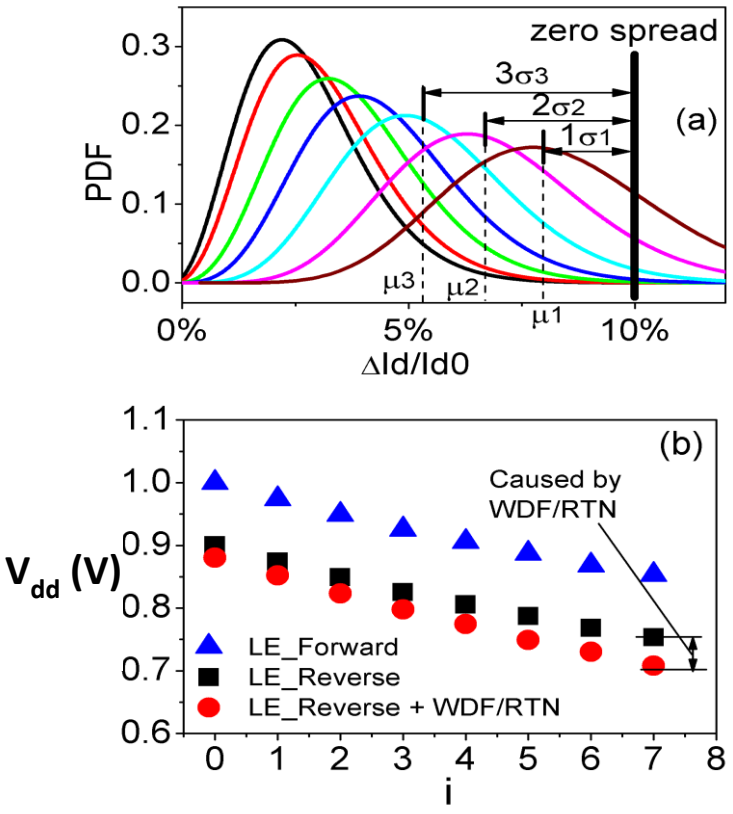


Figure 4.29 Impact of DDV on use-V_{dd}. When $\Delta I_d/I_d$ reaches 10% at i×σ, the mean $\Delta I_d/I_d$, μ, of defect-centric distributions reduces for higher i (a). This in turn requires a lower use-V_{dd} (b). For the reverse: '■'--- HCA only and '●' --- HCA and RTN/WDF.

4.5 Conclusions

As CMOS scales down, HCA scales up. For the first time, this chapter experimentally verifies that the HCA under use-V_{dd} can be predicted by the power law extracted from the VSST-method, provided that correct acceleration and '*n*'-evaluation are made. It points out the forward saturation $\Delta I_d/I_d$ and HCA measured by SMU gives erroneous '*n*' for nm-width devices. The model requires only 3 fitting parameters (Eq.4.3), making it readily implementable.

5 Impact of Hot Carrier Aging on the Random Telegraph Noise and Within-Device-Fluctuation

5.1 Introduction

One of the most popular topics on small devices is Random Telegraph Noise (RTN). RTN behaviour was first studied by Kandiah and co-workers in 1978 [143]. As a device's area is scaled down, the impact of a single charge in the gate dielectric scales up [19], [87], [134], [144]–[146]. Only a handful of defects exist in small area devices and the emission and capture of a single hole and electron will cause a discrete measurable step-like change in source-drain current. For current and future CMOS nodes, charging and discharging a single trap induces a random telegraph noise in I_d under a given V_g. When there are more than a few (e.g. 4) traps, it becomes difficult to separate them and the complex RTN signals appear in the form of within-a-device-fluctuation (WDF) [134], [147]. This may also exacerbate by white and $1/f$-noises present in the output signal [148].

RTN/WDF is usually observed at time-zero, i.e. in fresh devices[134][106]. It is becoming a major challenge for low power circuits. The low (V_g-V_{th}) used in low power circuits has less headroom to tolerate a given V_{th} shift, ΔV_{th}, since $\Delta V_{th}/(V_g\text{-}V_{th})$ is higher and the impact of ΔV_{th} on the driving current is relatively stronger. For instance, it has

been reported that a single charge can cause a V_{th} shift of ~30 mV [19], while a shift of only several mV can cause errors in circuits like successive approximation analogue-to-digital converters.

In addition to RTN/WDF, aging also occurs through either bias temperature instabilities [19][149]–[151] or hot carrier stresses [63], [90]. Unlike RTN/WDF, aging causes a gradual shift of device parameters in one direction. The interaction between RTN/WDF and aging is not fully understood and is of importance to optimize circuit performance. It has been reported that aging can either increase RTN/WDF or has little contribution to it [106].

This chapter is dedicated to investigating the relation between the amplitude of RTN/WDF and hot carrier aging (HCA) for nMOSFETs. The outputs show that the impact of HCA on devices of average RTN/WDF is typically modest, but can be substantial on devices of abnormally high RTN/WDF. The mechanism will be explored.

5.2 Device and Experiments

nMOSFETs were used in this experiment. Devices were fabricated by a 45 nm HK/MG process, having a typical channel length/width of 50/90 nm and an equivalent oxide thickness (EoT) of 1.45 nm. To ensure that the findings are not process specific, tests were also carried out on nMOSFETs fabricated by a 22 nm process. These devices were

fabricated with a typical channel length of 90 nm and channel width of 70 nm and an EoT of 1 nm.

The implementation of the RTN/WDF measurement is relatively easy. A V_g is applied generally above the threshold voltage and V_d is kept at a constant value of 10-200mV [87]. In this chapter, we run the test by capturing I_d at V_g=0.85~1.0 V and V_d=0.1 V. This V_g are selected to be near to the operational gate voltage, V_{gop}, of the device. The time window for the measurement is set to 60 s. This is long enough to capture the typical RTN/WDF in devices. After measuring RTN on fresh a device, Hot Carrier Aging (HCA) was applied to the device for 1,000s under V_g =V_d = 2.2 V. After that, we measured RTN in I_d at V_g = 0.85~1.0V and V_d = 0.1V again, to compare it with that in the fresh device.

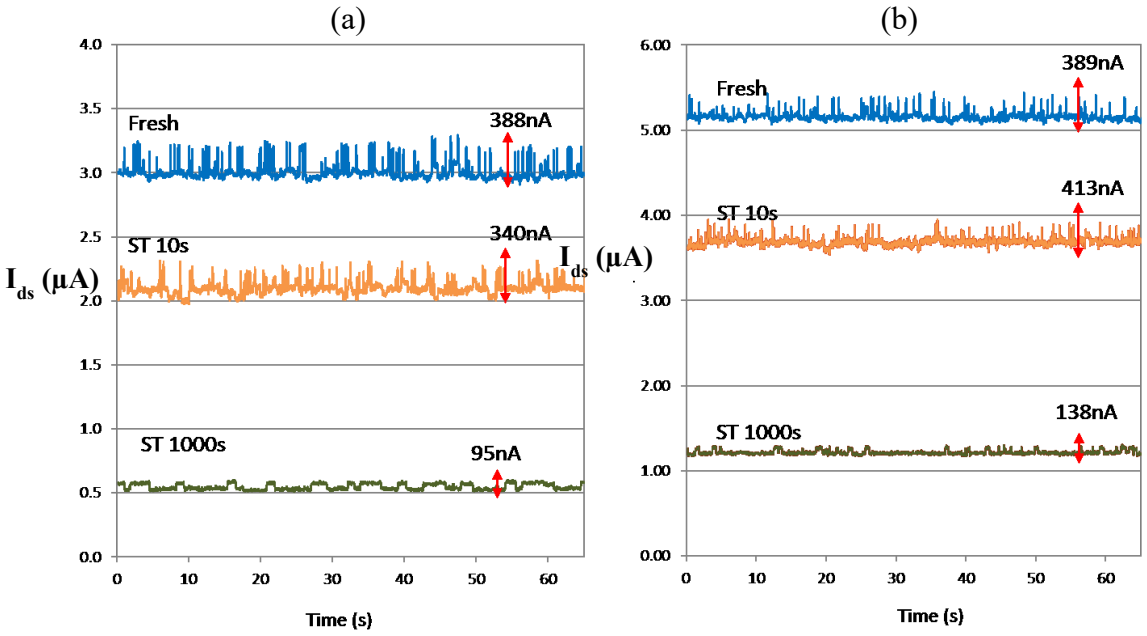


Figure 5.1 A comparison of RTN/WDF in a fresh device with that after stressing for 10 s and 1000 s (a) at constant V_g = 0.85 V and (b) at constant V_g = 0.9 V

The RTN/WDF signals before and after stress are compared in Figure 5.1. HCA stresses on the device can change the amplitude of RTN/WDF and this will be investigated in this chapter.

For a 45 nm process, measurements of RTN/WDF were carried out on 50 devices to demonstrate the device-to-device variation in Figure 5.2. Based on the measurement, V_g = 1.0 V is selected because more devices have abnormally large RTN/WDF at this V_g.

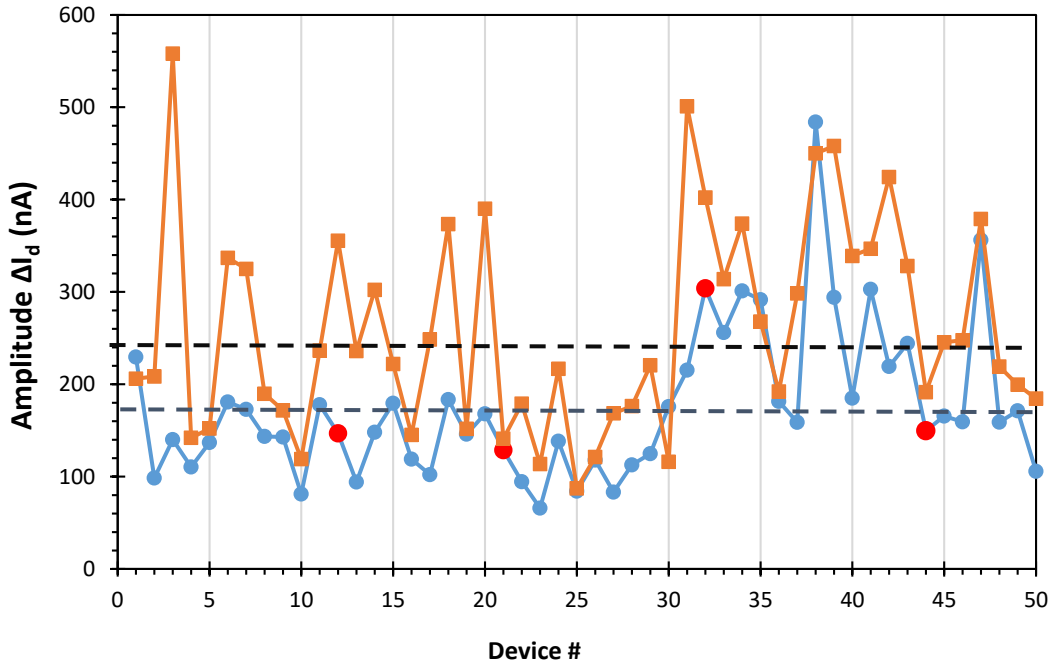


Figure 5.2 Amplitude of RTN/WDF at two V_g level; at V_g = 1.0 V (symbol ■) and at V_g = 0.9 V (symbol ●). (45 nm process)

Tests start by measuring RTN/WDF of I_d under V_g=1.0 V and V_d=0.1 V, and typical results are given in Figures 5.3(a) and 5.3(b). It can be seen that RTN has two clear states

due to trapping and de-trapping of a carrier at a single trap in the gate dielectric. We have tested 50 devices from 45nm process and 24 devices from 22nm process to measure amplitude of RTN/WDF in each of the devices. Figure 5.4 shows the device-to-device variation of RTN/WDF amplitude for 50 devices (45nm) and 24 devices (22nm). We can see that amplitude of RTN/WDF of the devices from 45nm process varies from as high as 558nA to as low as 88nA. For 22nm technology, the amplitude of RTN/WDF changes from a minimum of 96 nA to a maximum of 478 nA.

The ΔI_d and Amplitude in the figures are calculated as below:

$$\Delta I_d = I_d \text{ (measured)} - I_d \text{ (minimum)} \quad (5.1)$$

$$\text{Amplitude} = I_d \text{ (maximum)} - I_d \text{ (minimum)} \quad (5.2)$$

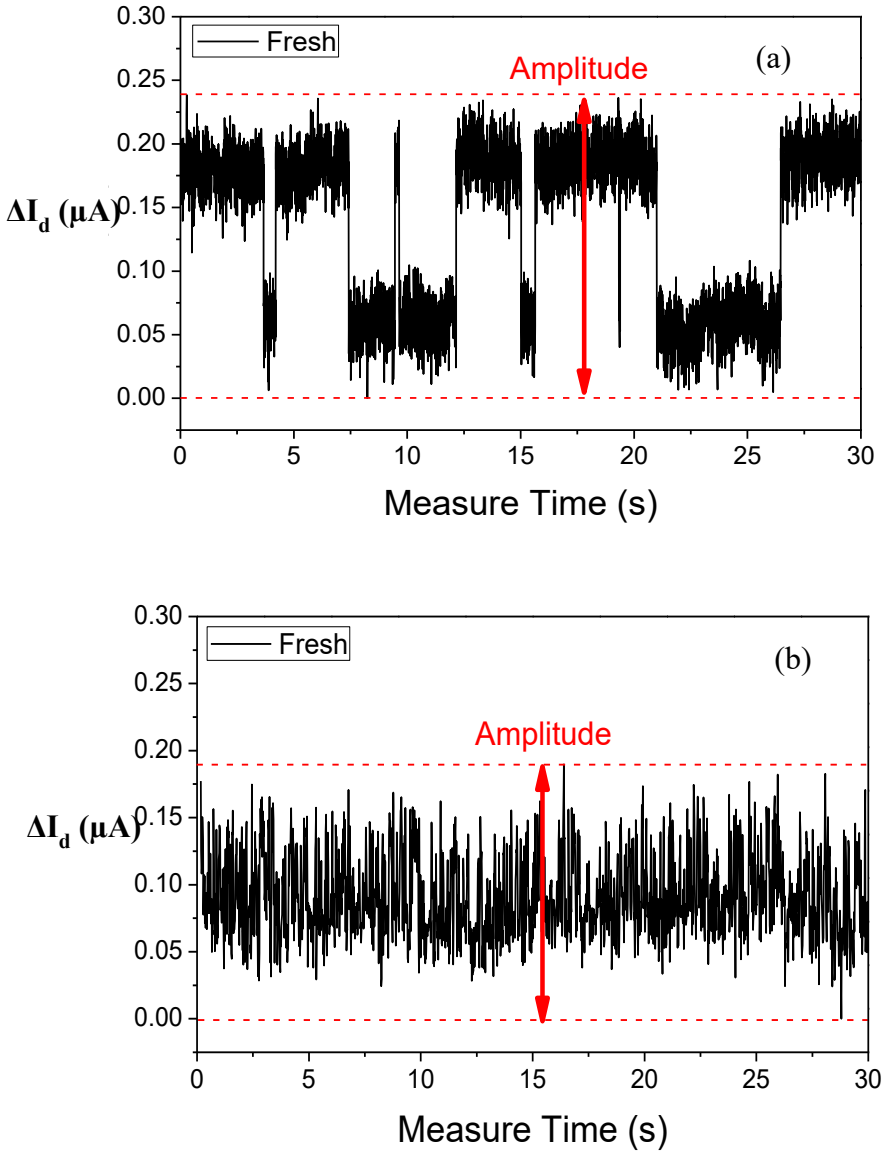


Figure 5.3 Typical fresh devices with (a) and without (b) clear RTN.

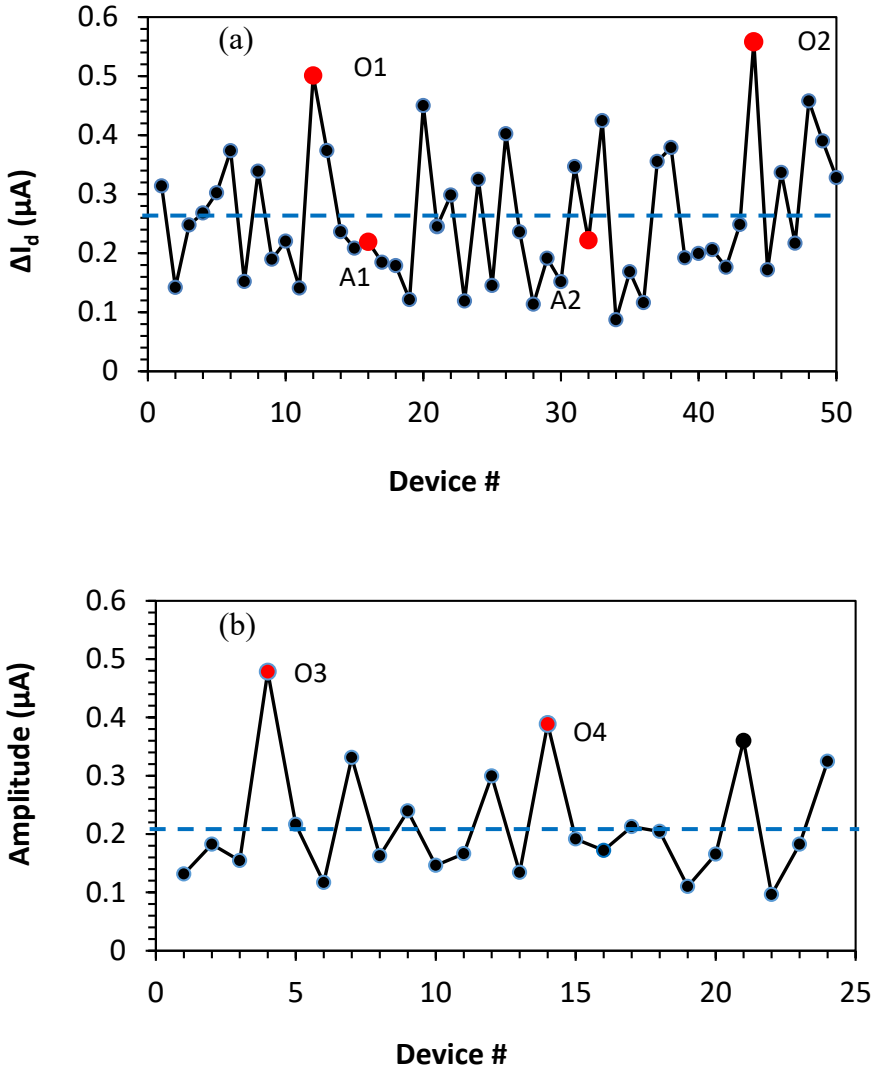


Figure 5.4 The device-to-device variation of RTN/WDF amplitude. The devices were fabricated by (a) 45 nm and (b) 22 nm processes. The dashed line is the average RTN/WDF for 50 devices (45nm) and 24 devices (22nm). In (a), the 'A1' and 'A2' mark two devices of RTN/WDF close to the average and 'O1' and 'O2' mark two outliers. In (b), 'O3' and 'O4' mark two outliers. The results for these six devices are given as representatives and their fresh RTN/WDF are marked out by the red '●'.

Two types of devices were selected for both 45 nm process and 22 nm process: one with average and one with abnormally high RTN/WDF. Table 5.1 shows details of the samples selected.

Table 5.1 Details of the sample of devices used

DEVICE	PROCESS	AMPLITUDE (RTN/WDF)	SIZE (RTN/WDF)
O1	45 nm	501 nA	Abnormal
O2	45 nm	558 nA	Abnormal
A1	45 nm	219 nA	Average
A2	45 nm	222 nA	Average
O3	22 nm	478 nA	Abnormal
O4	22 nm	388 nA	Abnormal

After HCA was applied under V_g=V_d=2.2 V for 1 ks, RTN/WDF was measured at the same I_d as that for a fresh device. This is to ensure that the Si surface potential is kept approximately the same. After the HCA, some of the devices then were selected to undergo the annealing process in forming gas (10% H_2) for 45 min at 400 ºC.

Hydrogen or forming gas annealing is often used at the end of the CMOS process to passivate the defects [24]. In our case, annealing is used to return the HCA stressed device to fresh condition. It is assumed that the defect returned to its original precursor status after annealing [23].

5.3 Results and Discussion

For a selected device of average RTN/WDF, Figures 5.5(a)-(d) show that HCA can either increase or decrease RTN/WDF modestly and the typical variation range is ±25%, which is smaller than the 6× device-to-device variation in Figure 5.4(a). This agrees with the early works [90], [106] and the verdict that RTN/WDF is dominated by as-grown defects [19][152]. One may argue that fresh devices shown in Figure 5.5(a) and Fig 5.5(c) do not represent RTN-like behaviour. This is because there are multiple (typically >3) traps and they will be referred to as "within a device fluctuation" (WDF). It is difficult to determine numbers of RTN-levels and the capture-emission times for each trap from such a complex signal. This Chapter focuses on the total RTN/WDF total amplitude, as defined in Fig. 5.4(b). Figure 5.6 shows the I_d versus V_g for the device 'A2' before and after HCA. HCA causes a clear reduction of I_d through aging.

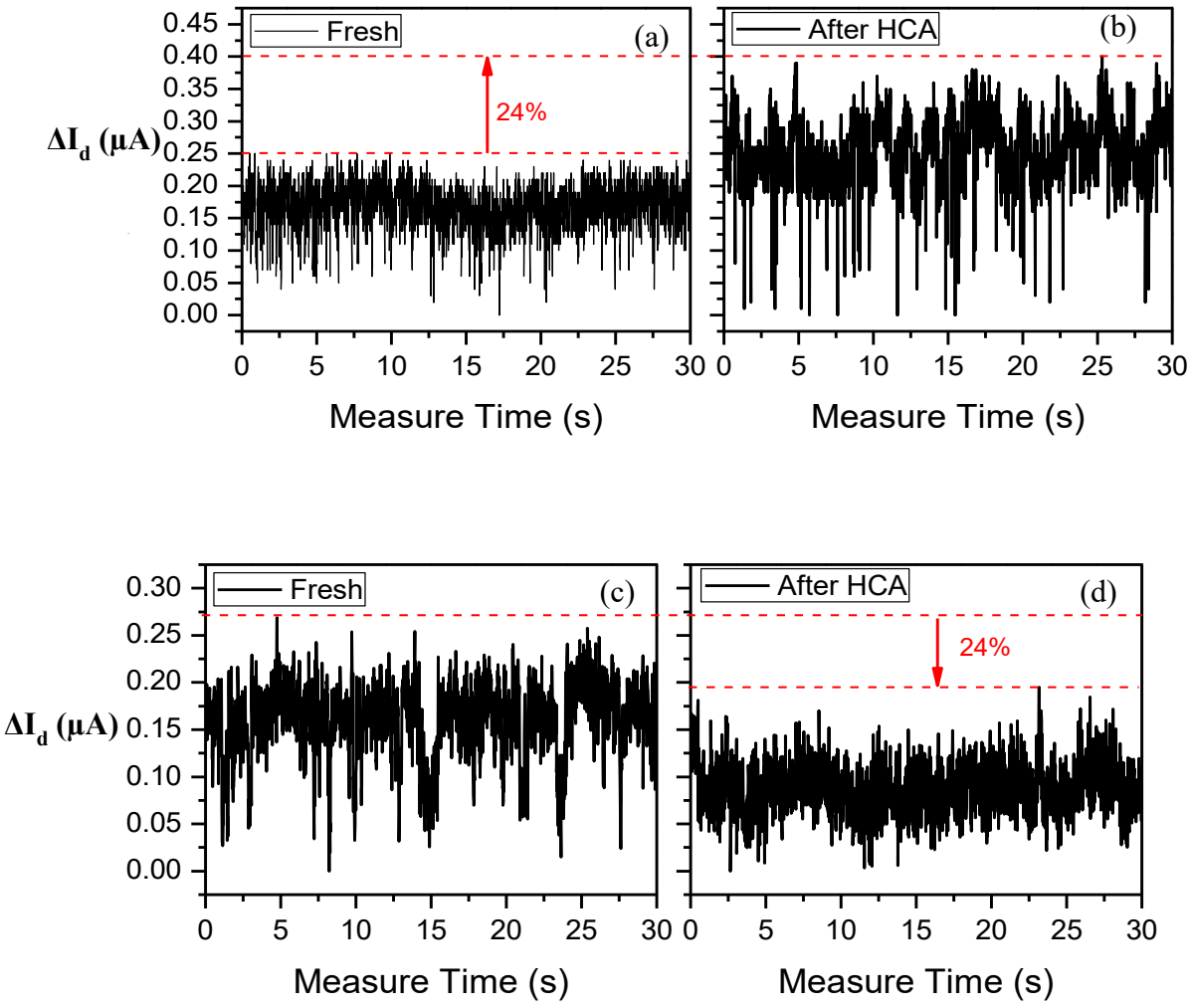


Figure 5.5 Typical impact of HCA on the devices of RTN/WDF close to average, marked as 'A1' and 'A2' in Figure 5.4(a). RTN/WDF can either increase by 24% (a)&(b) or decrease (c)&(d) after HCA.

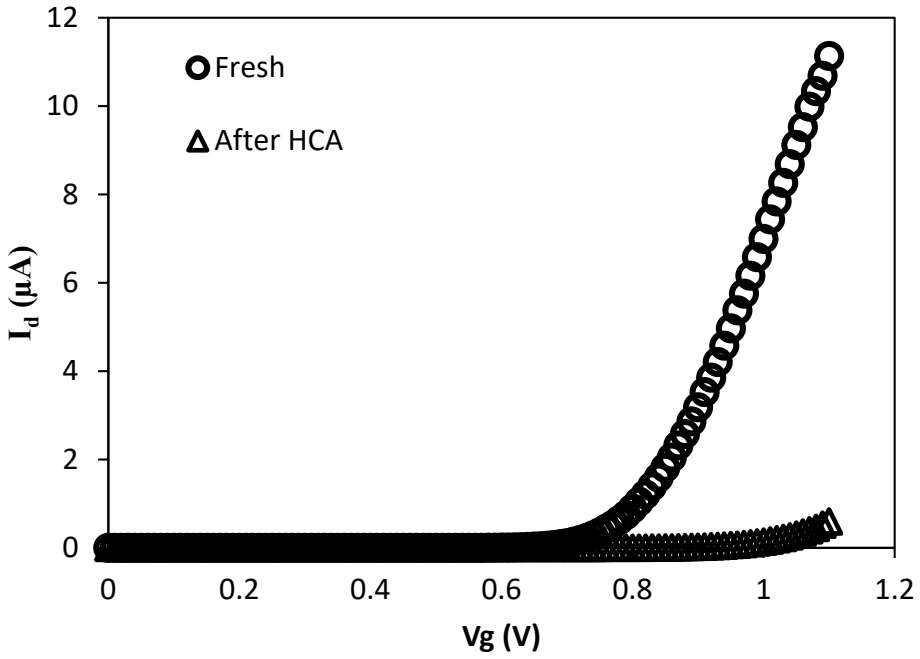


Figure 5.6 I-V measurement for device marked 'A2' before and after HCA stress.

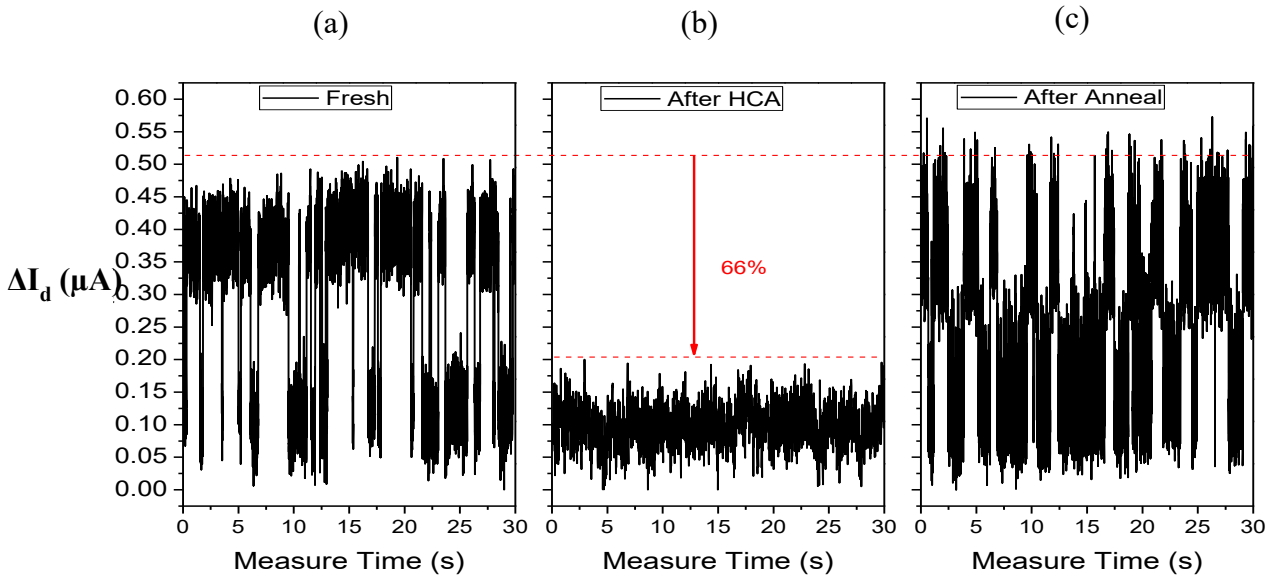


Figure 5.7 Typical impact of HCA on an outlier, marked as 'O1' in Figure 5.4(a). (a) is fresh and RTN/WDF reduces by 66% after HCA (b). (c) shows that RTN/WDF amplitude returns to its fresh level after an anneal at 400 °C.

The HCA-defects can affect RTN/WDF in two possible ways. On one hand, they may directly contribute to RTN/WDF by their charging/discharging. On the other hand, even if their charges do not alternate, they still can affect RTN/WDF by changing the current distribution within a device [153]. For the same as-grown defects, their effects on RTN/WDF will be different when the current density beneath it changes [19], [153]. If the HCA defects contribute directly to RTN/WDF by alternating their charging/discharging, an increase of defects by HCA should lead to a higher RTN/WDF. This is, however, against the reduction in Figure 5.5(c) and (d). As a result, it appears that HCA affects RTN/WDF through changing the current distribution, which will be further explored. This also agrees with early reports [154] that HCA defects recover little, so that they will not contribute to RTN/WDF by not discharging.

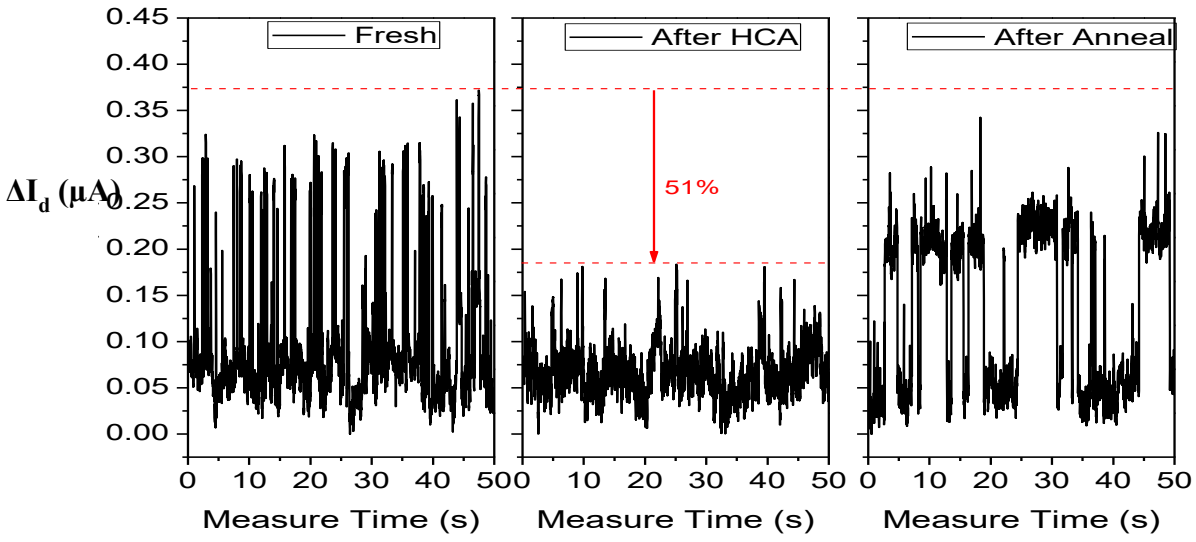


Figure 5.8 The tests were similar to those in Figure 5.7, but the device 'O4' was fabricated by a 22 nm processes and stressed under V_g=V_d=2 V for 1 ks.

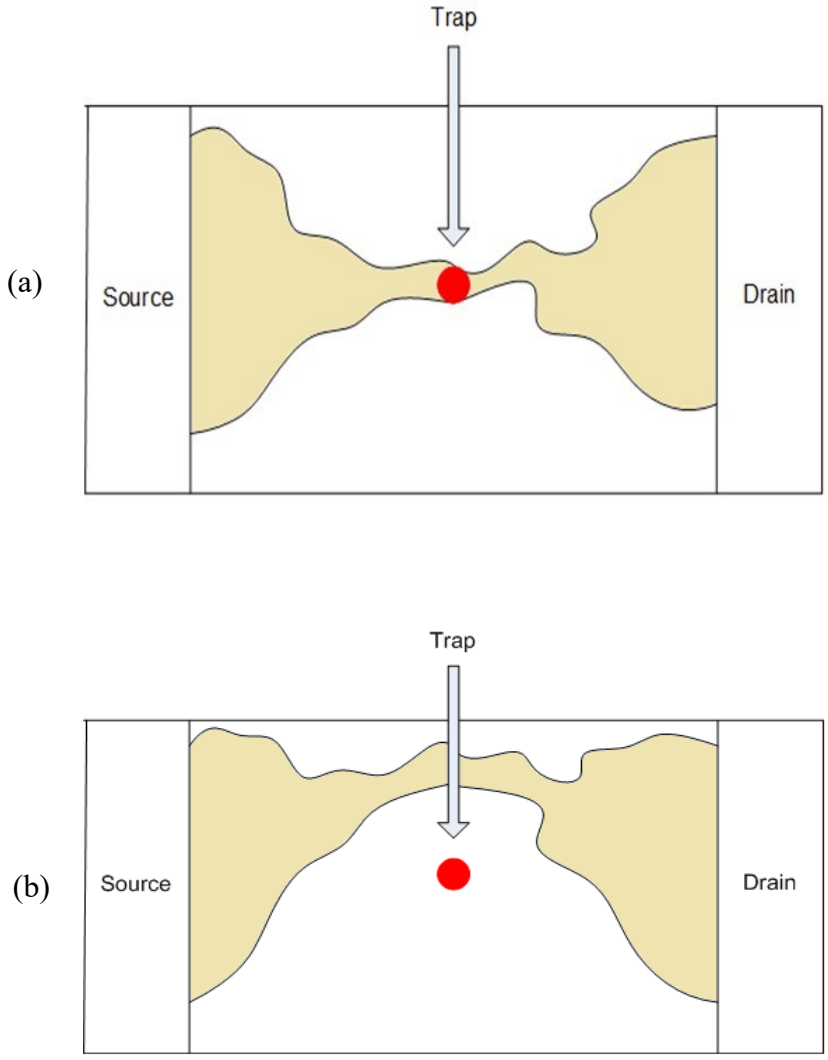


Figure 5.9 A schematic illustration of HCA-induced de-sensitization of a critical trap. (a) shows that a critical trap is at the location where the current density peaks, causing abnormal high RTN/WDF before HCA. (b) shows how a change of current distribution after HCA can reduce the current density under this trap, de-sensitizing it. This diagram is used to highlight the possible change of current distribution before and after HCA. It does not mean that current path is always strongly localized.

The impact of HCA on RTN/WDF through changing the current distribution will be further explored. For a device of abnormally high RTN/WDF, Figure 5.7(a) and (b) shows, for the first time, that RTN/WDF actually can be substantially (66%) reduced post

HCA. After the reduction, the RTN/WDF is around the average level shown in Figure 5.4(a).

To confirm that this reduction is not process-specific, Figure 5.8(a) and (b) show a substantial reduction again for a device of abnormally high RTN/WDF, fabricated by a 22 nm process. There are two possible explanations for the HCA-induced RTN/WDF reduction: a loss of defects [155][137] or a change of current distribution [153].

It has been reported that RTN/WDF in pMOSFETs can be unstable during measurements and some defects can disappear and then reappear [144]. The reduction in Figure 5.7(b), however, is a different phenomenon, since the RTN/WDF in nMOSFETs observed here is stable and does not disappear both before and after HCA during measurements.

The impact of a charged defect on the current will depend on its relative position against the current flow. On one hand, if there is a strong current flow directly beneath a charged defect, the impact of this defect on the current will be large. On the other hand, if there is little current flowing below a charged defect, its impact on the current will be weak.

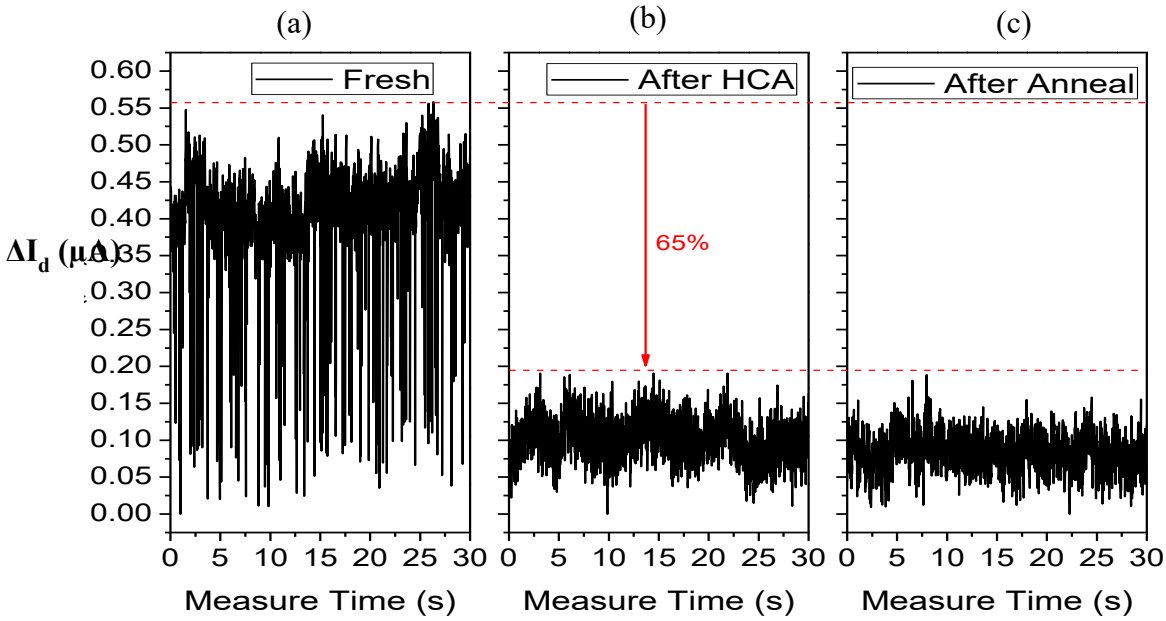
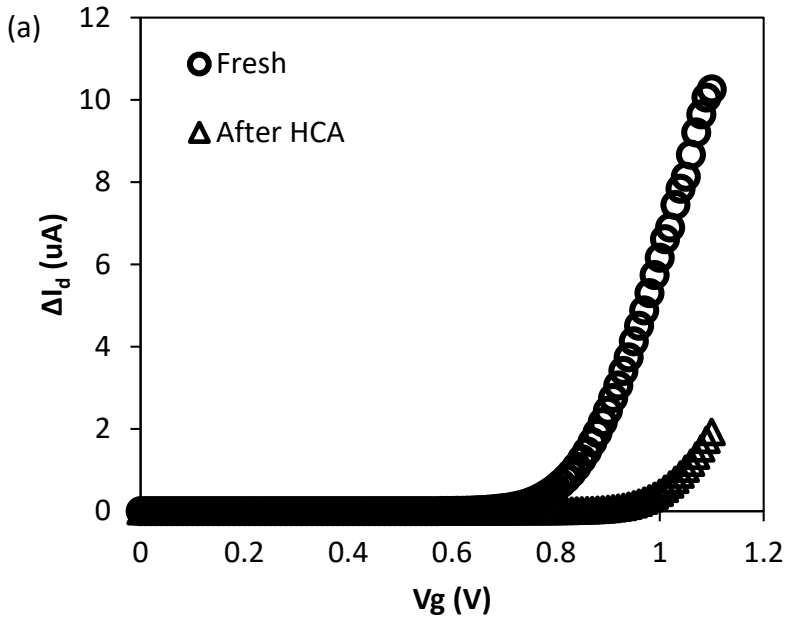


Figure 5.10 The impact of HCA on a device of outlier RTN/WDF, marked as 'O2' in Figure 5.4(a). (a) & (b) shows that RTN/WDF reduces by 65% after HCA. (c) shows that RTN/WDF remains low after an anneal at 400 °C, supporting defect loss.



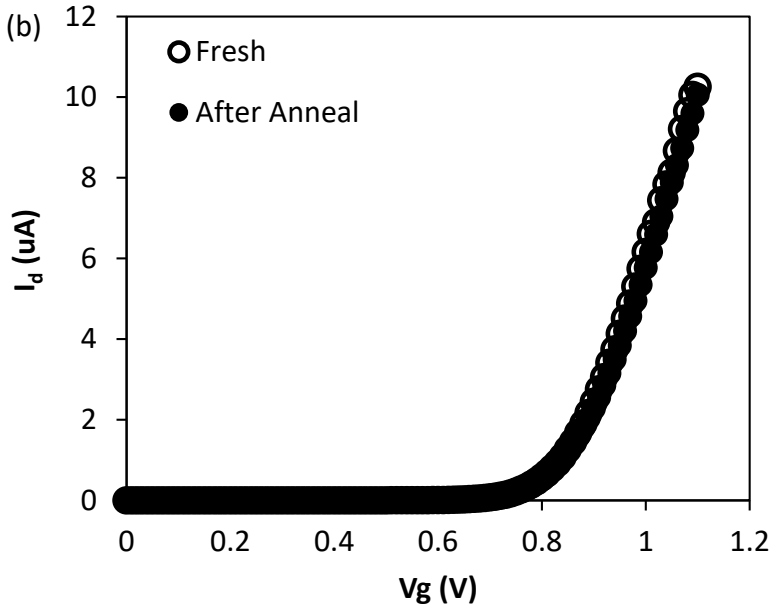


Figure 5.11 I-V measurement for device marked as 'O2' for (a) Fresh measurement compared to after HCA (V_g=V_d=2.2 V) and (b) comparison between fresh measurement and after annealing process.

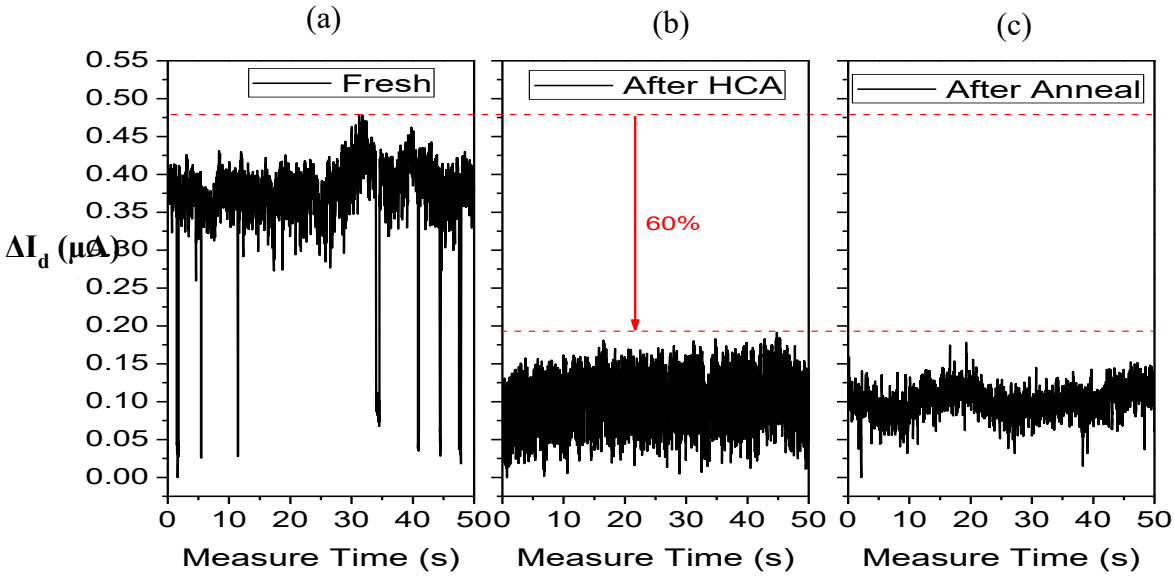


Figure 5.12 The tests were similar to that in Figure 5.10 but the device 'O3' was fabricated by a 22 nm processes.

The abnormally high RTN/WDF can originate from the presence of a critical trap: the current density peaks just beneath it, so that its charging/discharging has an abnormally large impact on I_d, as illustrated in Figure 5.9(a) [153].Large number of traps can contribute to multi-level of RTN states [156]. Figure 5.9(b) shows that HCA can modify the current distribution, reducing the density beneath this trap and de-sensitizing I_d to it. For a device of average RTN/WDF, the current distribution can be less localized and there is no critical trap where the current density peaks. Since the impact of each trap in these devices is close to average, current density beneath it changes typically modestly. On one hand, if the density increases, RTN/WDF will rise. On the other hand, if the density reduces, RTN/WDF will decrease. As a result, the HCA has a relatively modest impact on the RTN/WDF in both directions, as shown in Fig. 5.5.

To further investigate the origin of the HCA-induced reduction, the devices were annealed at 400^{o}C for 45 min in H_2 as the ambient gas, which removed the HCA-generated defects and restored I_d to its fresh level [137], [155], as confirmed in Fig. 5.11. When measured again post-anneal, Figure 5.7(c) and 5.8(c) shows that the amplitude of RTN/WDF nearly returns to its pre-stress abnormal high level for most devices (~75%). The anneal restores the original current distribution and, in turn, the abnormal RTN/WDF. This supports an HCA-induced change of current distribution as the origin of abnormal RTN/WDF reduction.

It is noted that the RTN/WDF, after the stress then anneal, behaves differently from that in fresh devices, although their magnitudes before and after the anneal are similar in Figure 5.7 and 5.8. It is not known what causes these differences at present. One may speculate that the as-grown traps in the fresh devices were also affected by the stress and

the subsequent anneal, leading to the changes in their RTN/WDF behaviour. Since most of the devices tested in this work have a complex WDF, rather than a clear RTN, it is difficult to extract the capture and emission time reliably.

5.4 Conclusion

In this Chapter, the impact of HCA on the magnitude of RTN/WDF of nMOSFETs is investigated. For the devices of average RTN/WDF, HCA typically can either increase or reduce it modestly (±25%). For the devices of abnormally high RTN/WDF, however, the HCA generally can reduce it substantially. After an anneal at 400 ºC, RTN/WDF returns to the abnormal level for most devices, supporting an HCA-induced change of current distribution as the origin of the reduction. There are cases, however, where RTN/WDF does not return to its pre-stress level after the annealing process, suggesting defect losses.

6 Development of a Technique for Directly Measuring RTN and BTI-Induced V_{th} Fluctuation under Use-V_{dd}

6.1 Introduction

As CMOS nodes scale down, the fluctuations induced by random charge-discharge of traps in the gate dielectric scale up. Smaller devices have larger statistical spreading because of fewer traps per device and the larger impact of a single charge [19], [157] on them. The increased number of devices per chip also leads to a larger statistical spread [19], [157] and the higher data transmission rate requires tighter control of jitters [158]. Fluctuations have become a major concern for circuit design and have attracted much attention recently [19], [32], [48], [89], [146], [147], [157]–[165]. It has been reported that current fluctuation in some fresh devices can be over the typical device lifetime criterion of 10% [89], reducing the yield.

Fluctuations are commonly observed as the random telegraph noises (RTN) in the drain current, ΔI_d, under a given gate bias, V_g, and early works [32], [48], [154], [160]–[162], [165]–[167] have focused on them. ΔI_d_RTN allows probing individual traps and an analysis of their mean capture and emission time dependence on V_g gives the trap energy and spatial locations [89], [160], [165], [167]. This has improved the understanding

substantially. There are, however, few direct measurements of the RTN-induced jitter in threshold voltage, ΔV_{th}. This is because the measurement is difficult; the charge-discharge of traps for RTN is highly dynamic and the average ΔV_{th} is typically low. As a result, the RTN-induced ΔV_{th} often was either not given [89], [161] or estimated from dividing ΔI_d by trans-conductance, i.e. $\Delta V_{th} \approx \Delta I_d / g_m(V_{dd})$ [32], [146], [154], [160], [165]. The accuracy of the ΔV_{th} evaluated in this way was not given in these works [32], [146], [154], [160], [165].

Figure 6.1(a) and Figure 6.1(b) shows that the ΔV_{th} evaluated in this way (marked as 'A') is substantially different from that directly measured with sensing-V_g near to V_{th} (Marked as 'B'). The data shown is measured from 28 nm device.

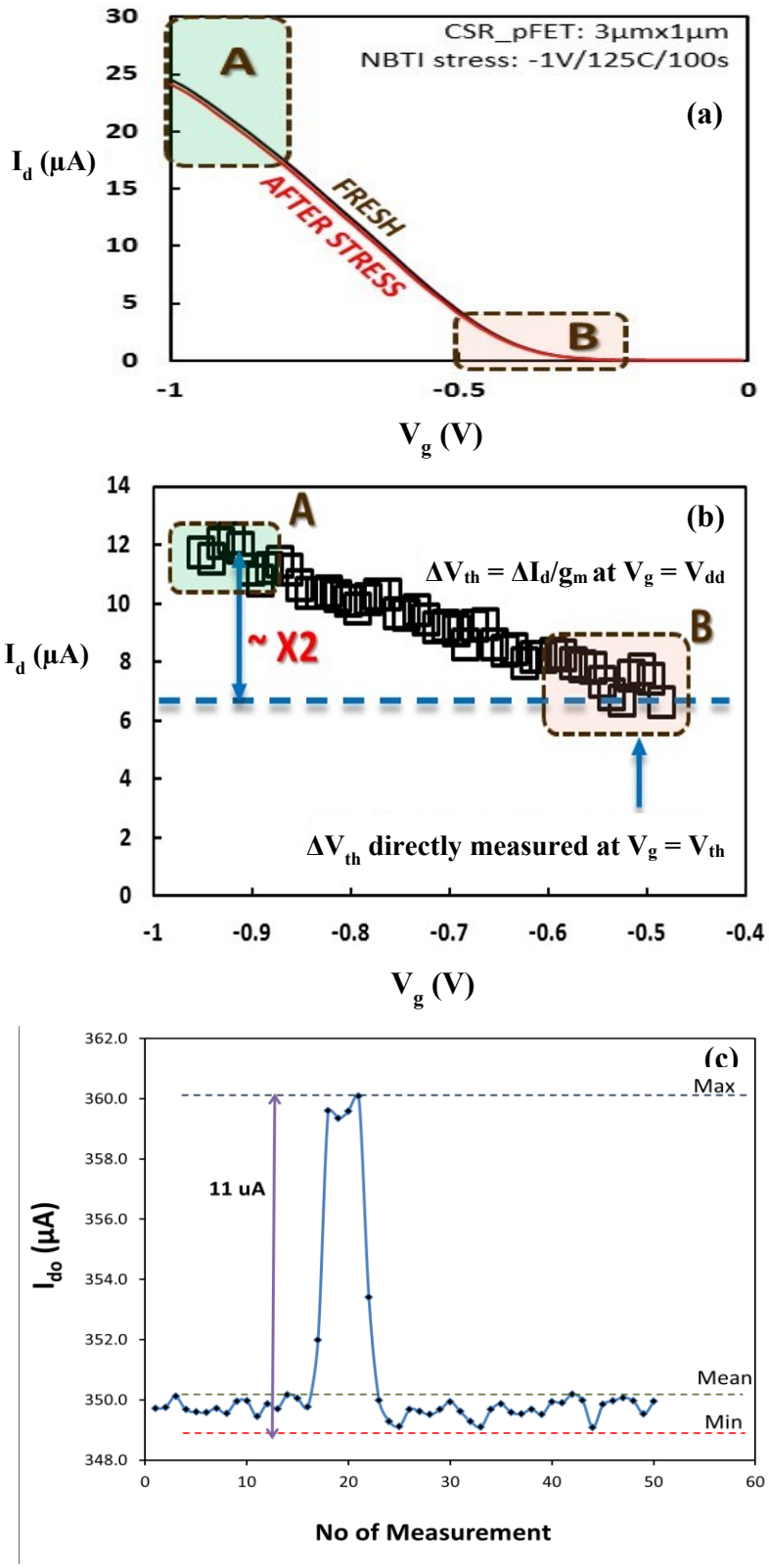


Figure 6.1 (a-b) A comparison of ΔV_{th} evaluated by $\Delta I_d/g_m$ (□) with ΔV_{th} directly measured at $V_g \sim V_{th}$ (*dashed line*). The $\Delta V_{th}=\Delta I_d/g_m$ with $V_g \sim V_{dd}$ (region A) is ×2 of the real ΔV_{th}. (c) I_{do} measurement captured the impact of RTN on device 90 nm x 27 nm

There are devices that do not give analysable RTN signals in terms of extracting mean capture/emission time [161] and ΔI_d can appear as a complex within-a-device-fluctuation (WDF) [48], as shown in Figure 6.2. The charge-discharge of traps for RTN is highly dynamic and the average ΔV_{th} is typically low. These devices were simply deselected in some early works [160], [162], making the real device-to-device variations (DDV) of fluctuation unobtainable.

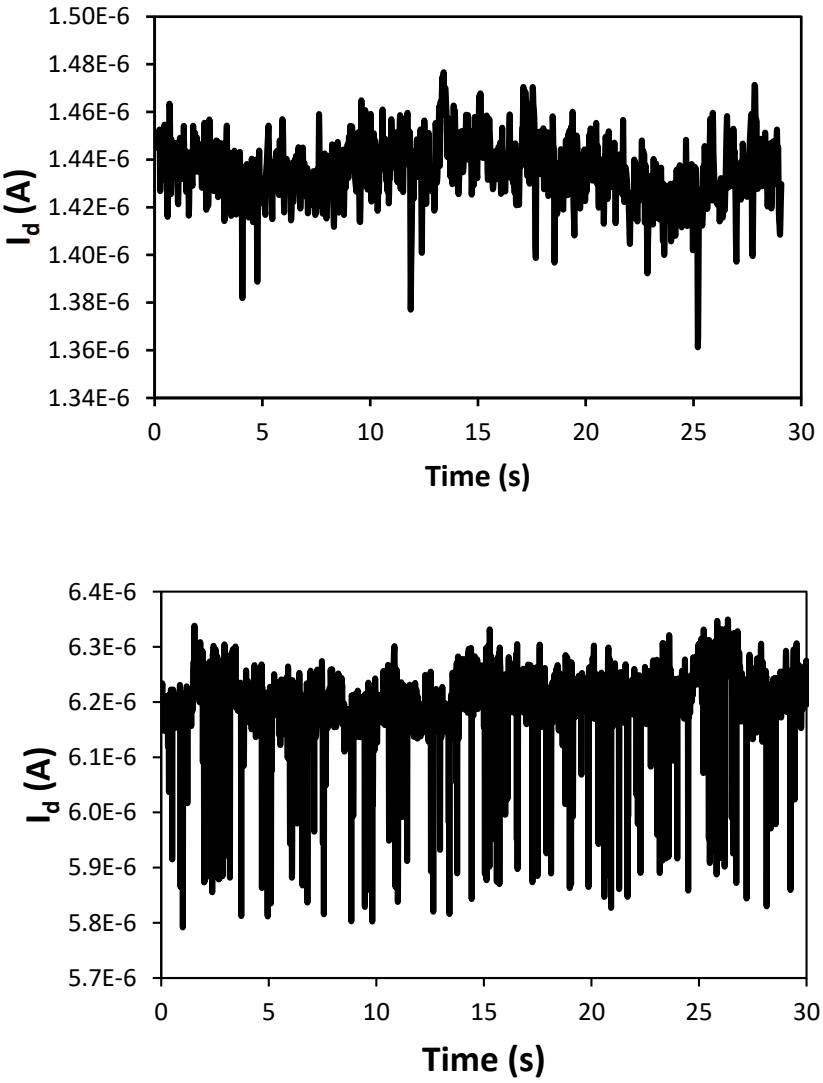


Figure 6.2 Example of RTN/WDF signals. (a) not analysable RTN signals and (b) shows complex WDF.

To model the impact of RTN on the margin of SRAM [147] and the timing error [163], one needs both ΔI_d and ΔV_{th}. For example, RTN in the pass transistor 1 in Figure 6.3(a) can reduce the driving current by ΔI_d and slow down the V_g rise of transistor 2 in reaching its threshold voltage, V_{th0}, by $\Delta t(\Delta I_d)$. RTN in the transistor 2 can increase its V_{th} by ΔVth and results in a further delay, $\Delta t(\Delta V_{th})$. Therefore, there is a need to obtain both accurate ΔI_d and ΔV_{th}.

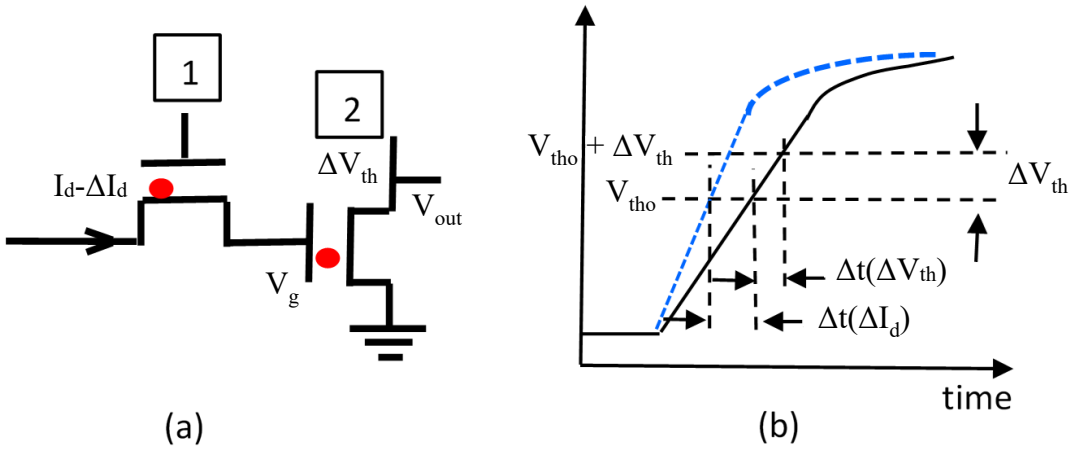


Figure 6.3 A schematic illustration of the impact of ΔI_d and ΔV_{th} on timing: (a) circuits and (b) waveform. V_{out} switches when $V_g \approx V_{th}$, which is delayed by a lower charging current, I_d-ΔI_d, supplied through the transistor 1 and a higher $V_{th}=V_{th0}+\Delta V_{th}$ of the transistor 2.

The objective of this chapter is to develop a Trigger-When-Charged (TWC) technique for directly measuring the RTN-induced ΔV_{th}. By ensuring that the measurement is taken when a trap is charged, the accuracy is substantially improved. It is found that the $\Delta I_d/g_m(V_{dd})$ correlates poorly with the directly measured ΔV_{th} and the former doubles the latter on average. The discrepancy originates partly from the device-to-device variation

of relative local current density beneath a trap at $V_g=V_{th}$ [105], [168], [169][170] and partly from the charge-induced mobility degradation. The TWC developed in this work is applicable to devices with or without analysable RTN signals and it will be used to evaluate the device-to-device variation.

6.2 Devices and measurement technique

6.2.1 Devices

The pMOSFETs used in this work were fabricated by a 28 nm commercial CMOS process with a use V_{dd} of 0.9 V. They have a metal gate and a high-k dielectric stack with an equivalent thickness of 1.2 nm. The channel length and width are 27 nm and 135 nm, respectively. For comparison purposes, big devices of 3 µm (W) x 1 µm (L) were also used, which has insignificant device-to-device variation. All tests were performed at 125 ^{o}C.

Noise from the measurement system is one of common problems encountered during the experiment. In order to assess the effect of system noise on the RTN/WDF measurement, a comparison of system noise and with typical RTN/WDF is shown in Figure 6.4. From the figure, system noise is < 12% of the RTN/WDF at V_g=0.6 V

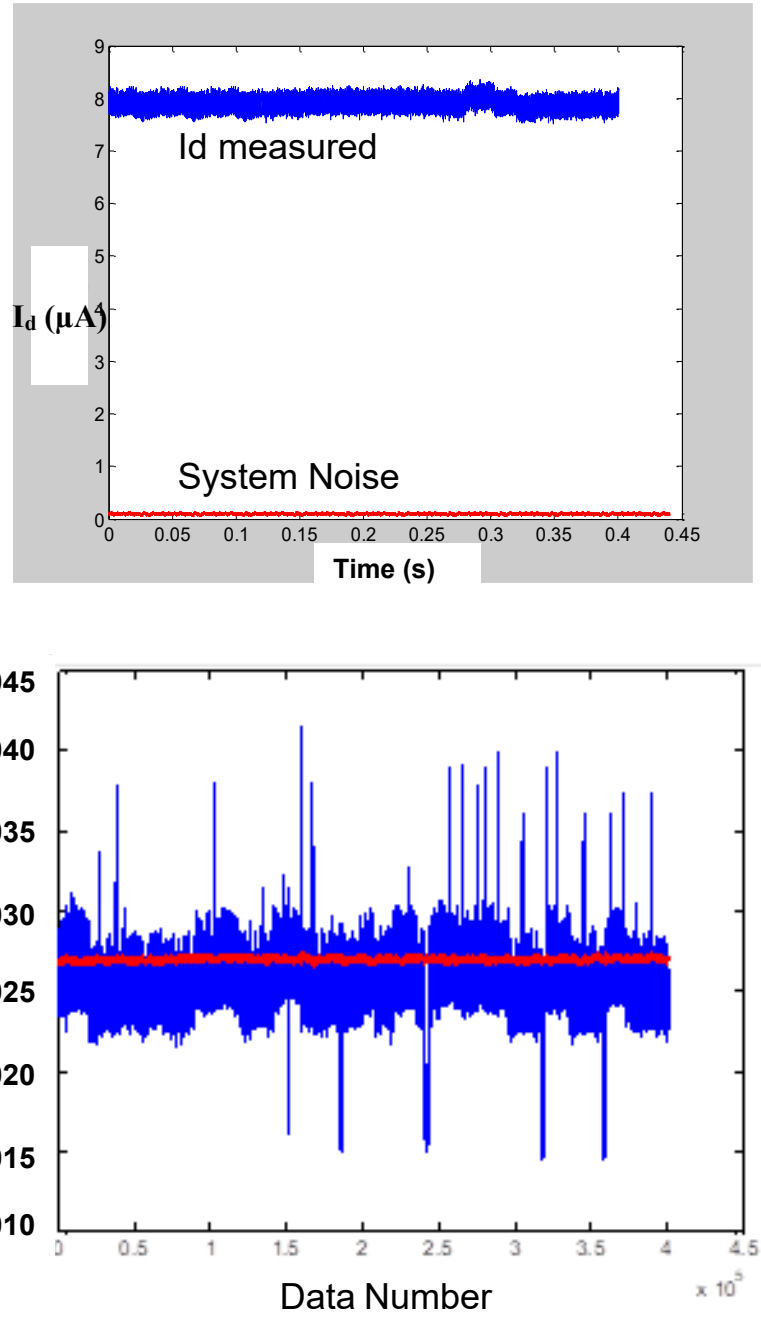


Figure 6.4 Comparison between system noise and true signal measured at V_g=-0.60 V.

I-V measurements are used to monitor the threshold voltage shift. In order to freeze the discharging during IV measurements, the pulse method is used with measurement time less than 5 us [22]. In principle, discharging can occur under a given $V_{discharge}$ and during the measurement itself. Hence, it is required to freeze the discharge during the

measurement under a given $V_{discharge}$. If the DC measurement is used under a given $V_{discharge}$, discharging induced by the measurement itself reduces the measurement accuracy.

6.2.2 Selection of test conditions

The V_g range for measuring ΔI_d on-the-fly:

For negative bias temperature instability (NBTI) tests, V_g-acceleration is generally used [171] to give a measurable ΔV_{th} within an acceptable time. Although Figure 6.5(a) shows that there are more 'as-grown' traps [130] for higher $|V_g|$, this V_g-acceleration is, however, not suitable for characterizing WDF in nano-scaled devices. Fig. 6.5(b) shows that WDF is dominated by traps close to E_f, where occupancy changes rapidly. An increase of $|V_g|$ lowers E_f and the WDF would be dominated by a different group of traps from those under use V_{gop}. Early works report that different traps have different impacts on a device [19], so that ΔI_d should be measured on-the-fly under the use V_{gop}, rather than a raised $|V_g|$.

This chapter focuses on the V_g range of practical use: 0.6 V to 1.0 V. For a large device, Figure 6.1(a) shows that the ΔV_{th} is insignificant in this range, making its measurement a challenge.

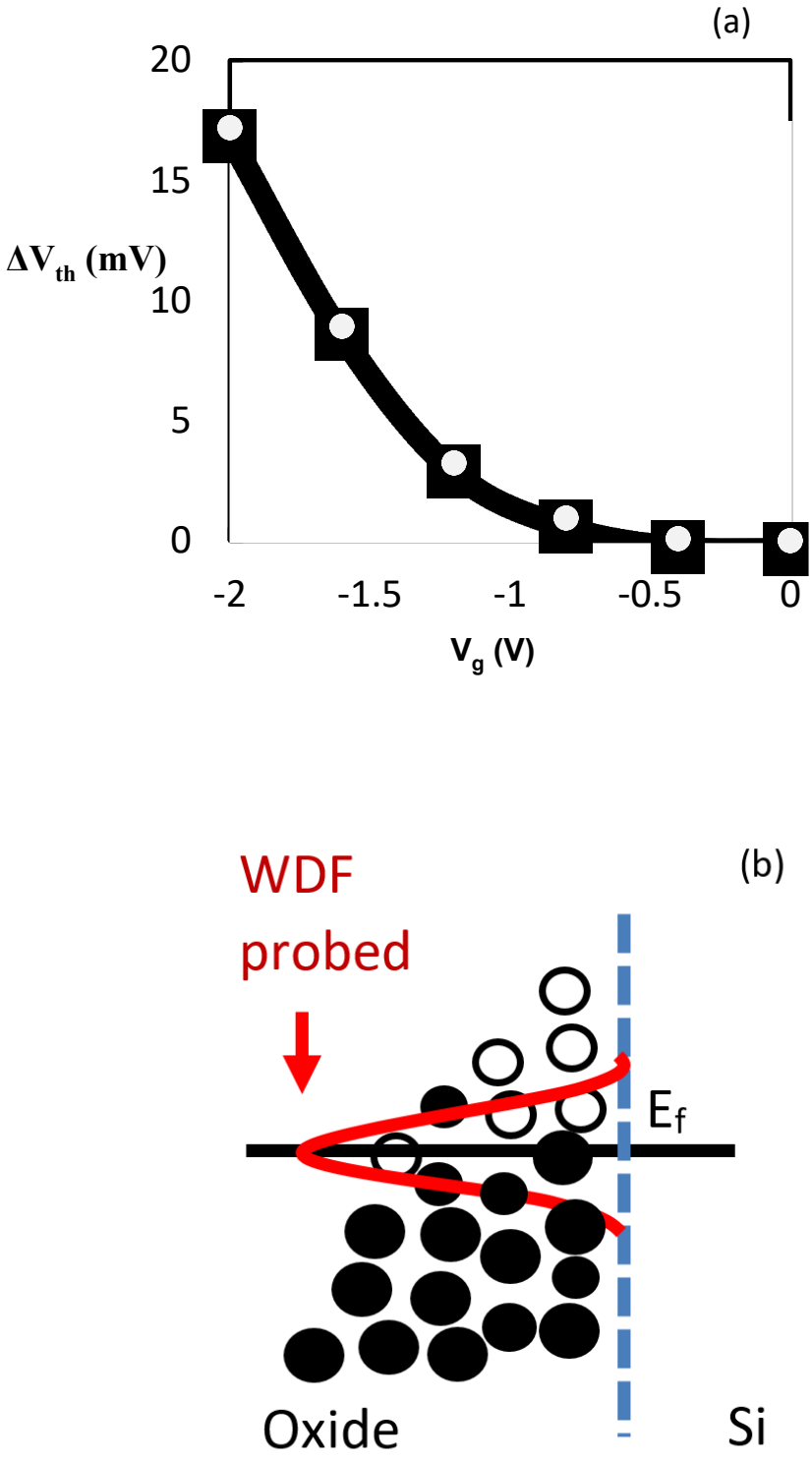


Figure 6.5 (a) As-grown hole traps reduces for lower $|V_g|$ and (b) WDF probes traps near E_f.

The measurement delay for ΔV_{th} and the sensing V_g:

In addition to ΔI_d, ΔV_{th} is needed. As mentioned earlier, a fluctuation in V_{th} causes timing errors in switching-on of a device (Figure 6.3). For an SRAM cell, although a pMOSFET can be biased under V_{gop} in 'hold' mode, V_g can be close to V_{th} during read/write. It is well known that ΔV_{th} reduces with longer measurement delay due to de-trapping [172][173] and the question is what is the 'right delay'. On one hand, for modern digital circuits, the transition time from V_{gop} to V_{th} can be in the order of tens of pico-seconds. On the other hand, the discharge or emission time typically is reported being micro-seconds or longer when V_g is above or close to V_{th} [134], [173]. The discharge during the transition between V_{gop} and V_{th} should be frozen, therefore.

There are two ways of freezing the discharge during V_{th} measurement. One is measuring ΔI_d 'on-the-fly' without reducing V_g to V_{th} and then evaluating ΔV_{th} from $\Delta I_d/g_m$ [15], [112], [128], [133], [134 . Figure 6.6, however, shows that $\Delta V_{th}=\Delta I_d/g_m$ is different from the real ΔV_{th}, when the sensing V_g is reduced to ~V_{th}. As a result, the correct sensing V_g should be ~V_{th}, rather than V_{gop}.

The other is to apply a V_g pulse sufficiently fast (e.g. <5 μs) to freeze the discharge during the transition from V_{gop} to V_{th} [134], [173] . In this work, a 3 μs pulsed I_d-V_g was used and ΔV_{th} will be measured from the V_g shift at V_{th}.

6.2.3 Test procedures

Figure 6.6 gives the V_g waveform and measurement steps. After recording the reference I_d-V_g on a fresh device, the test starts by filling the traps under V_{gop}=-1 V for 40 seconds, as early work [130] shows that charging traps below E_f completes within tens of seconds.

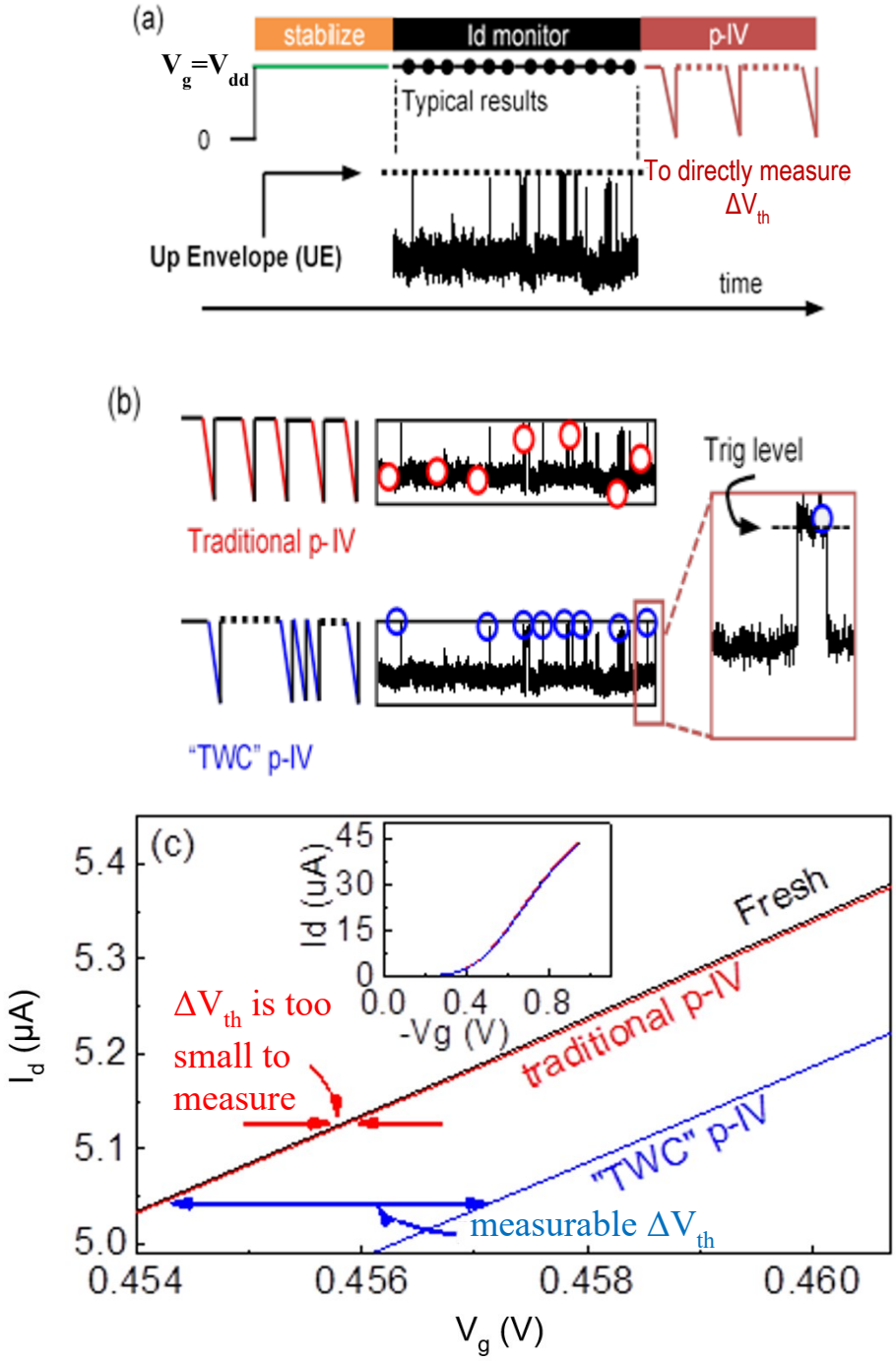


Figure 6.6 The 'Trigger-When-Charged (TWC)' technique. (a) Test procedure: After a stabilization period, the RTN-induced ΔI_d is monitored under $V_g=V_{dd}$ and the upper envelope (UE) is determined. The trigger-level for subsequent pulse (3 μs) I_d-V_g (p-IV) is then set just below UE to measure ΔV_{th}. 50 p-IVs were measured in (b) and their average is given in (c). The TWC p-IV captures the RTN-induced ΔV_{th}, while the Traditional p-IV at pre-set time often misses the charge and is inapplicable.

Difficulties with standard measure-stress-measure methods:

For aging-induced ΔV_{th} under stresses such as negative bias temperature instability (NBTI) [174][175] and hot carriers [176], [177], the degradation is commonly measured at preset time. This is acceptable, as the V_g-acceleration used in the stress generally leads to a large-enough ΔV_{th} that is measurable and deterministic at a preset time. There are, however, two difficulties in applying this method to deeply scaled devices under use-V_{dd}, where ΔV_{th} mainly exhibits as Random Telegraph Noise (RTN). First, there are only a few active traps and the average ΔV_{th} is typically low. Second, charge-discharge of these traps is highly dynamic: they are often neutral at the preset time for measurement, as shown by the red circle symbols in Figure 6.6, and would be missed by the measurement.

One way to avoid these difficulties is selecting devices that only have one trap, which induces a high enough ΔV_{th} (e.g. 30 mV) and its emission time is long enough for completing the measurement [105], [168]. This has improved our understanding of the interaction between a trap and the current. Such devices, however, are rare and the required device selection precludes obtaining real device-to-device variations. A

measurement technique is needed for measuring ΔV_{th} at $V_g=V_{th}$ without such device selection, therefore.

Test procedure of Trigger-When-Charged technique:

Figure 6.6(a) gives the V_g waveform. After recording the reference I_d-V_g on a fresh device, the test starts by a 'stabilization' period of 40 sec under $V_g=V_{dd}$=-0.9 V. If there are any traps at deep energy level in a device, they will be filled during this period [178]. ΔI_d under V_g=-0.9 V is then monitored for a period, e.g. 100 sec, as marked by 'I_d monitor' in Fig. 6.6(a). A sampling rate of 1 M/sec was used [134]. The trapping-induced upper envelope (UE) of ΔI_d is obtained.

To measure the trapping-induced ΔV_{th}, one must ensure that the measurement was taken when the traps are charged. This is achieved by setting the trigger level of the oscilloscope and the pulse generator for V_g just below the UE, as shown in Figure 6.6(b). Once triggered, the pulse I_d-V_g (p-IV) is recorded in 3 μs to minimize the discharge [134], [178].

Although a sampling rate of 1 M/sec can be used to monitor ΔI_d under a fixed V_g=-0.9 V, it only gives 3 points in 3 μs and is too slow for the p-IV. To have a sufficient number of points for p-IV, a higher rate of 100 M/sec is used. The p-IV was repeatedly measured for 50 times and their average is used to reduce the system noise to ~1 mV.

ΔV_{th} is evaluated from the difference between the Trigger-When-Charged (TWC) p-IV and the reference p-IV. The reference p-IV was obtained also from the average of 50 p-IV with the same sweep rate, performed on fresh devices before filling the deep traps by applying the waveform in Figure 6.6(a). When measuring these 50 p-IV, it is possible that a trap can be filled during the measurement. These outlier p-IVs were excluded from the reference p-IV. This ensures capturing the ΔV_{th} induced by both RTN and deep traps, if they are present. In case one is interested in capturing RTN-induced ΔV_{th} only, the reference p-IV should be taken after filling the deep traps.

Figure 6.6(c) demonstrates that a single trap induced ΔV_{th} of ~2 mV is successfully captured by the TWC technique, which often would be missed by the traditional p-IV recorded at a preset time, as illustrated by the red circles in Figure 6.6(b). The measured $\Delta V_{th}/\Delta I_d$ ratio is used to convert ΔI_d to ΔV_{th}.

Measurement set-up:

As the main objective of this chapter is to develop a technique for measuring the RTN-induced ΔV_{th} under use V_{dd}, the detailed measurement set-up is given in Figure 6.7. I_d under V_d=0.1 V was converted to a voltage, V_{out}, by a home-made operational amplifier circuit. During the 'I_d monitor' phase in Figure 6.6(a), V_{out} was monitored by both channels 2 and 3 of an oscilloscope and one example is given in Figure 6.7(b).

In the following 'p-IV' phase of Figure 6.6(a), when V_{out} is above the 'trigger level' in Figure 6.7(c), the oscilloscope triggers and simultaneously sends out a signal to trigger the pulse generator for V_g. Both the pulse applied to the gate and the corresponding V_{out} are captured, as shown in Figure 6.7c. Two channels are needed here: channel 2 is at a fine scale to ensure capturing the small V_{out} fluctuation with good accuracy and channel 3 is switched to a coarse scale to capture the whole p-IV. As a comparison, Figure 6.7(d) shows an example triggered at a preset time that missed the trapped charge.

The UE in Figure 6.6(a) can be caused by either a single trap or multiple traps. In the latter case, the UE results from the combined charges of multiple traps. This removes the need for selecting devices of a single trap and makes the method applicable to all devices.

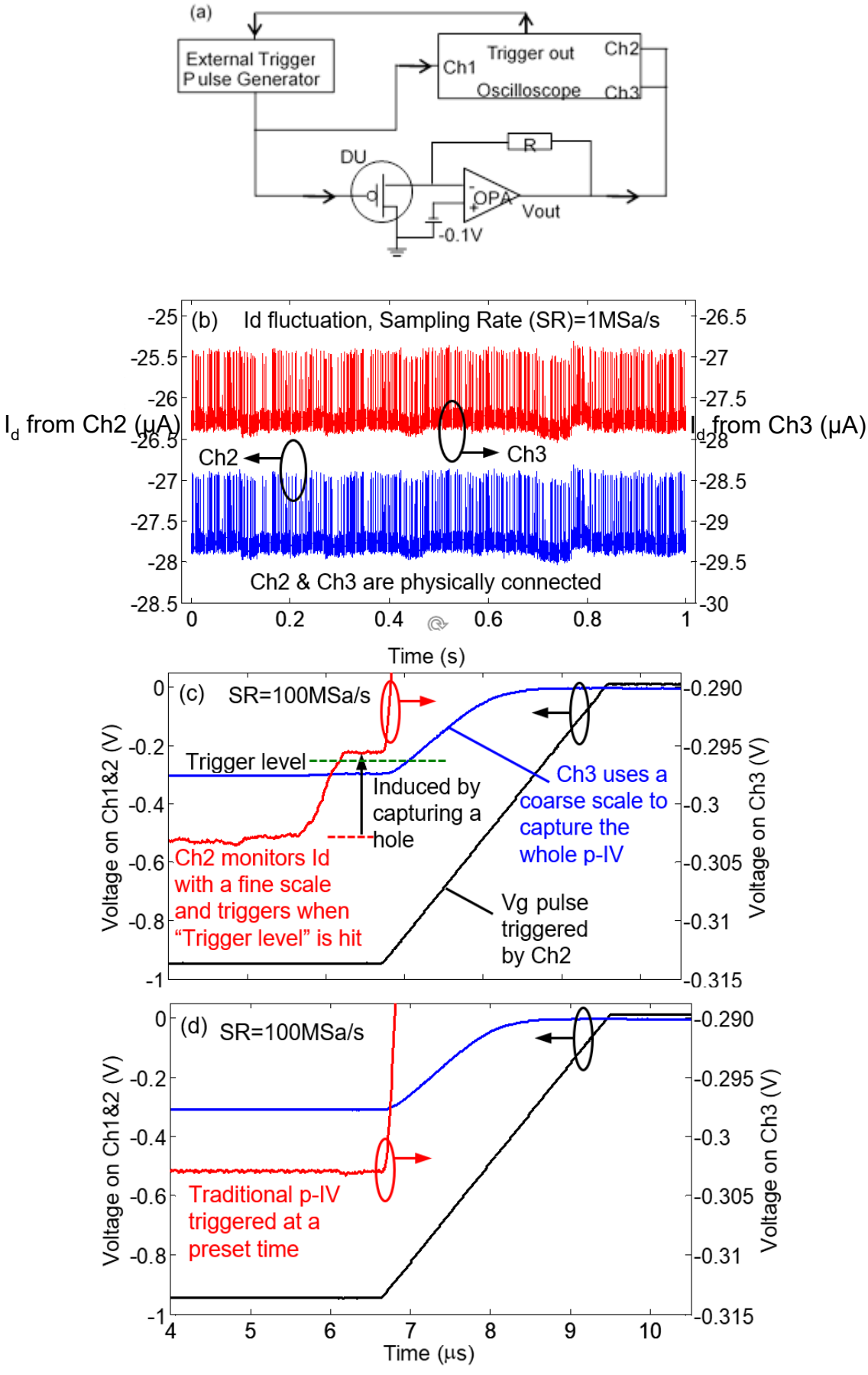

(a)
External Trigger Pulse Generator
Ch1
Trigger out
Ch2
Oscilloscope
Ch3
R
DU
OPA
Vout
-0.1V
(b) Id fluctuation, Sampling Rate (SR)=1MSa/s
Id from Ch2 (μA)
Id from Ch3 (μA)
Ch2
Ch3
Ch2 & Ch3 are physically connected
Time (s)
(c) SR=100MSa/s
Trigger level
Induced by capturing a hole
Ch2 monitors Id with a fine scale and triggers when "Trigger level" is hit
Ch3 uses a coarse scale to capture the whole p-IV
Vg pulse triggered by Ch2
Voltage on Ch1&2 (V)
Voltage on Ch3 (V)
(d) SR=100MSa/s
Traditional p-IV triggered at a preset time
Time (μs)

Figure 6.7(a) Test configuration for "Trigger-When-Charged (TWC)" measurement technique. A high-speed operational amplifier based circuit is used to convert I_d to V_{out} that is connected to both channels 2 and 3. The "Trigger out" of the oscilloscope is connected to the "External trigger in" of the pulse generator. (b) The V_{out} fluctuation is captured by both channels 2 and 3, as they are physically connected. (c) A screen-shot of the TWC p-IV measurement waveform. Channel 2 keeps its fine scale for accurate triggering, while channel 3 is switched to a coarse scale to capture the whole "TWC" p-IV. (d) A screen-shot of the traditional p-IV measurement at a preset time, where the trapped charge is missed.

6.3 Results and discussions

A. Comparison between $\Delta I_d/g_m(V_{dd})$ and $\Delta V_{th}(V_{th})$

As mentioned in the introduction, early works [32], [146], [154], [160], [165] often estimated ΔV_{th} by $\Delta I_d/g_m(V_{dd})$, where both ΔI_d and g_m were obtained under $V_g=V_{dd}$. This is effectively measuring the shift of IV at $V_g=V_{dd}$, as marked by the point 'B' in Figure 6.8(a) and the corresponding inset. The real ΔV_{th}, however, should be evaluated from $V_g=V_{th}$ at the point 'A' in Figure 6.8(a). In this work, V_{th} is extracted by extrapolating from the maximum g_m point and V_{th}=-0.45 V in Figure 6.8(a). The shift in V_{th}, ΔV_{th}, at a given sensing V_g is evaluated from $\Delta I_d/g_m(V_g sense)$. We now compare the ΔV_{th} evaluated at $V_g sense=V_{th}$ ('A' in Figure 6.8(a)) with that at $V_g sense=V_{dd}$ ('B' in Figure 6.8(a)).

Figure 6.8(b) plots $\Delta V_{th}(V_{th})$ against $\Delta V_{th}(V_{dd})=\Delta I_d/g_m(V_{dd})$ measured on 63 devices. Both of them have a large DDV, but the correlation between them is poor. For similar $\Delta I_d/g_m(V_{dd})$, ΔV_{th} can spread from its minimum to its maximum approximately. As a result, errors are large if $\Delta I_d/g_m(V_{dd})$ is used as ΔV_{th}, so that it is essential to measure ΔV_{th} directly at $V_g=V_{th}$. Although both of them have a maximum close to the typical device

lifetime definition of 30~50 mV, the average $\Delta I_d/g_m(V_{dd})$ doubles that of ΔV_{th}, as shown by the two dashed lines in Figure 6.8(b). This is because many devices have $\Delta V_{th}(V_{th})$ close to zero, but $\Delta I_d/g_m(V_{dd})$ are above 10 mV. The origin of the differences between these two will be analysed next.

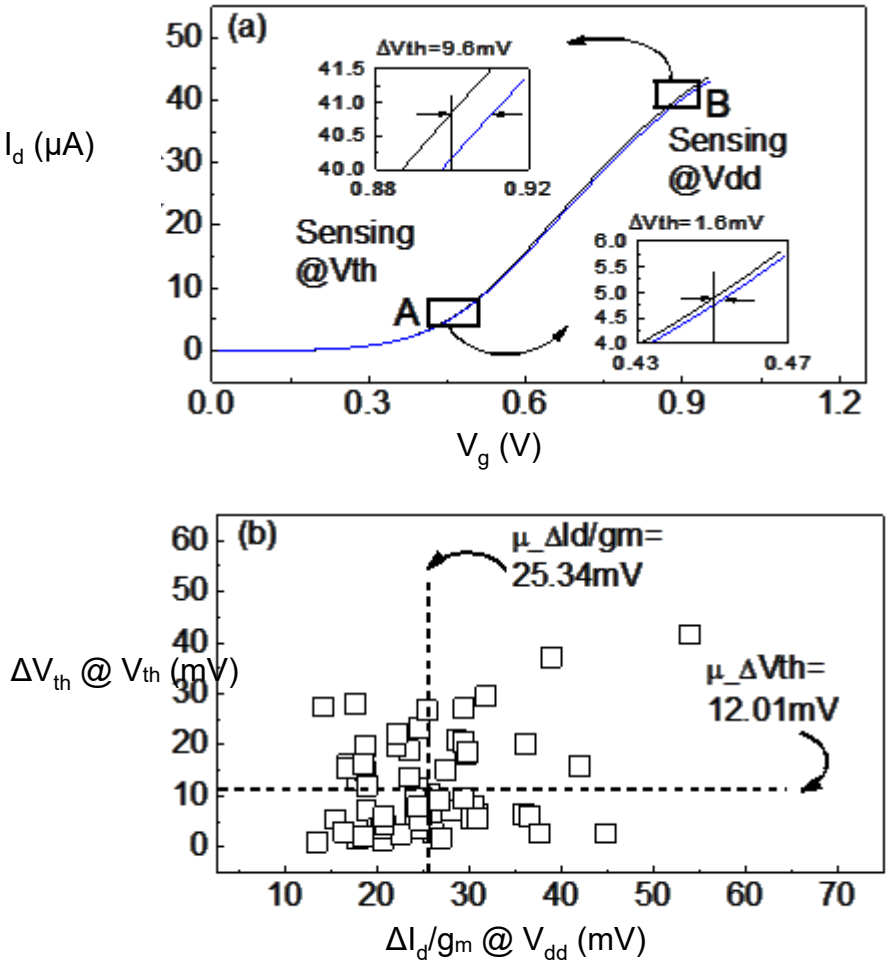


Figure 6.8(a) Early works estimated RTN-induced ΔV_{th} from $\Delta I_d/g_m$ at V_{dd}=0.9 V (Point 'B'), rather than directly measuring it at V_g=V_{th} (Point 'A'). The two insets are enlarged p-IV at the two points. The black p-IV is reference and the blue p-IV is the TWC p-IV. (b) The poor correlation between $\Delta I_d/g_m$ at V_{dd} and ΔV_{th} at V_g=V_{th}. Each point was taken from a different device. The dotted lines mark the mean values.

B.Effects of sensing V_g on ΔV_{th}

In Figure 6.8(a), the sensing V_g for ΔV_{th} is -0.9 V for the point B and V_{th}=-0.45 V for the point A. Since the whole I_d~V_g was measured, one can also extract the "apparent ΔV_{th}" at other sensing V_g by using $\Delta I_d/g_m(V_g sense)$. The "apparent ΔV_{th}" here refers to the ΔV_{th} evaluated in this way under $V_g sense \neq V_{th}$. Typical examples obtained from different devices are given in Figure 6.9(a-e).

The dependence of the apparent ΔV_{th} on the sensing V_g has strong device-to-device variations (DDV), agreeing with that observed for single traps [105], [168]. On one hand, Figure 6.9(a) corresponds to Figure 6.8(a), where ΔV_{th} increases monotonically with $|V_g|$ and ΔV_{th} at $|V_g|$=0.9 V is 6 times of the real $\Delta V_{th}(V_{th})$. On the other hand, ΔV_{th} can also reduce by almost half over the same voltage range, as shown in Figure 6.9(b). There are also cases where (i) ΔV_{th} is almost a constant (Figure 6.9(c)); (ii) ΔV_{th} increases initially and then reduces (Figure 6.9(d)); and (iii) ΔV_{th} decreases initially and then increases (Figure 6.9(e)).

It is known that channel current can have a narrow percolation path near V_{th} and the impact of a charged trap on a deeply scaled device depends on the relative local current density beneath the trap [105], [168]–[170]. This can explain the device specific dependence observed in Figure 6.9. As schematically illustrated in Figure 6.10, for the device in Figure 6.9(a), the trap is located far away from the current percolation path at V_{th}, so that it has little impact and $\Delta V_{th}(V_{th})$ is low. The many close-to-zero $\Delta V_{th}(V_{th})$

points in Figure 6.8(b) indicates that this is often the case. As V_g increases, the current becomes more evenly spread and its relative density under this trap rises, leading to the increase of ΔV_{th} with V_g. As there is current flowing beneath each trap at V_{dd}, there is no close-to-zero apparent ΔV_{th} in Figure 6.8b, when evaluated by $\Delta I_d/g_m(V_{dd})$.

For the device in Figure 6.9(b), however, the trapped charge is on top of the current percolation path at V_{th}, resulting in a large ΔV_{th} at V_{th}. As V_g increases, the current path is widened, so that the impact of the same charge on the device reduces and the ΔV_{th} decreases with $|V_g|$ in Figure 6.9(b). Similarly, the relative current density under the trap in Figure 6.9(c) changes little with V_g and ΔV_{th} is insensitive to V_g. The dependence of relative current density under a trap on V_g may not be monotonic, which can explain the behaviour in Figure 6.9(d&e). For instance, in Figure 6.9(d), it may increase initially and then decrease. When there are multiple traps, some can behave like Figure 6.9(a) and some like Figure 6.9(b). A combination of them can give the complex dependence in Figure 6.9(d&e).

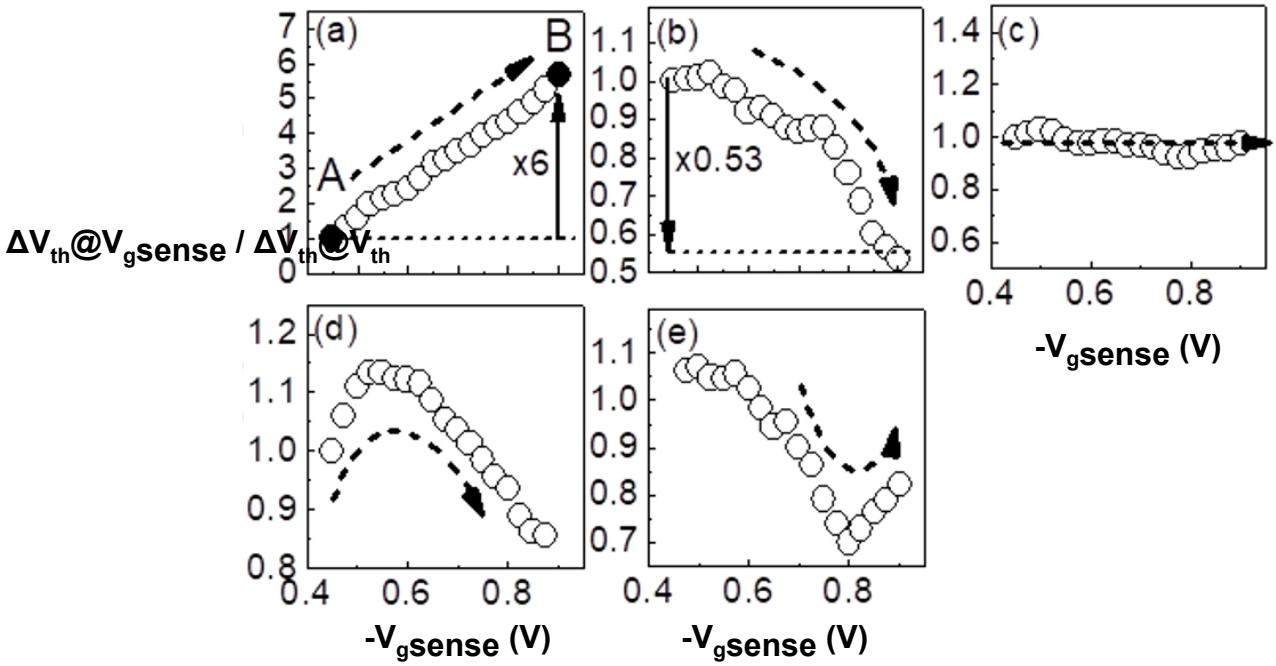


Figure 6.9 Examples of the device specific dependence of the apparent ΔV_{th} on the sensing V_g, $V_g$sense. (a)-(e) were obtained from five different devices. The apparent ΔV_{th} at a $V_g$sense was obtained from the shift of TWC p-IV from the reference at $V_g$sense. The ΔV_{th} is normalized against its value at $V_g$sense=V_{th}. As the lowest |$V_g$sense| is close to V_{th}, the data starts from ~1 in all devices.

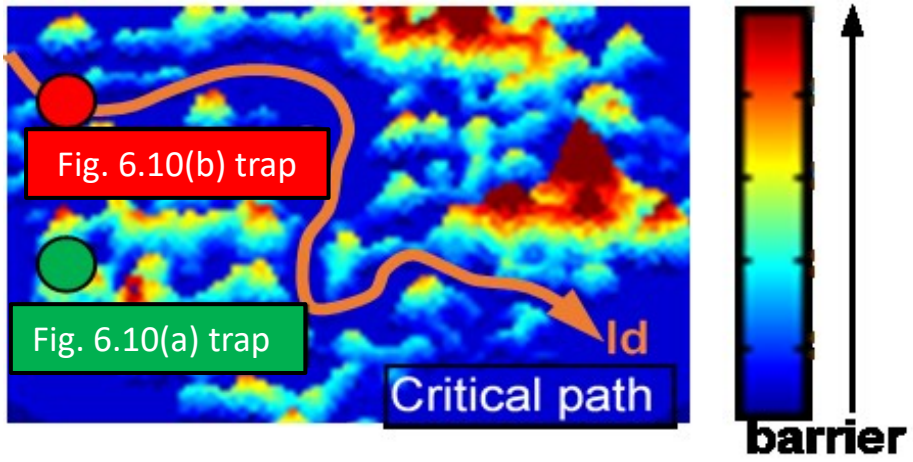


Figure 6.10 A schematic illustration of different impacts of traps at different locations on a device at threshold condition. The current can follow a percolation path under V_g=V_{th}. The trap in green corresponds to the device in Figure 6.9a: it is away from the critical current path, so that it only has a small effect on the device at V_{th}. The trap in red corresponds to the

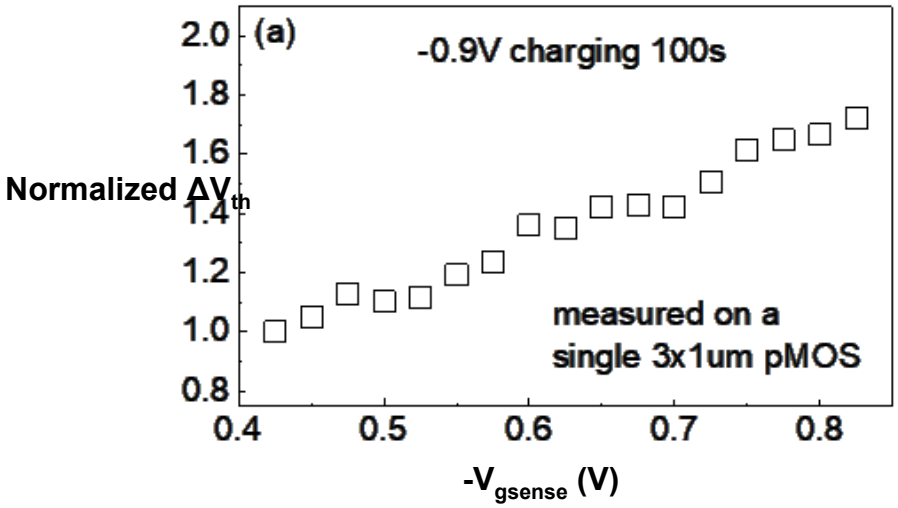
device in Figure 6.9b: it is on top of the current critical path and has a large effect on the device at V_{th}.

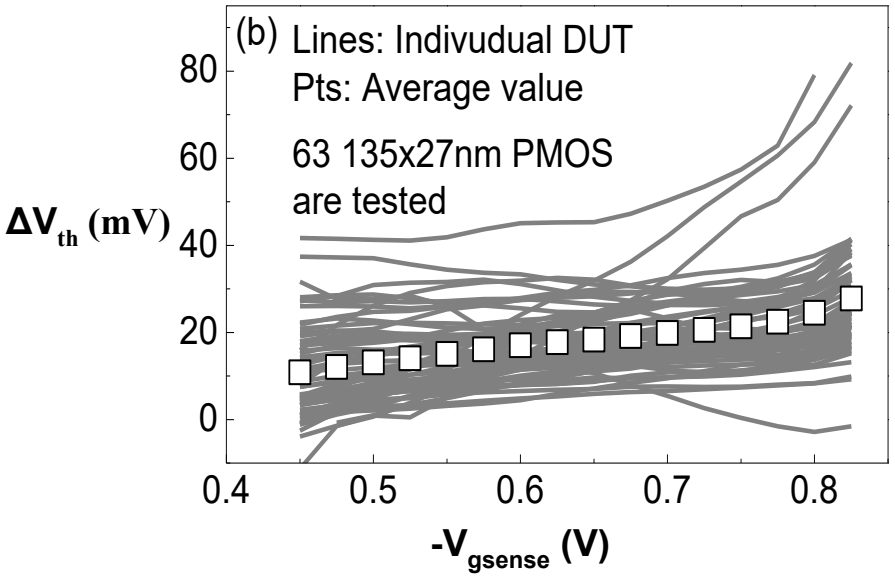
Although the deeply scaled device-specific dependence of ΔV_{th} on sensing V_g can be explained by the interaction between the trap and the relative local current density beneath it, there is also a device independent ΔV_{th} dependence on the sensing V_g. For a large 3×1 μm device where DDV is insignificant, Figure 6.11(a) shows that ΔV_{th} also increases with $|V_g sense|$. On one hand, a more evenly distributed I_d at higher $|V_g sense|$ allows more traps making an effective impact. On the other hand, the charge induced columbic scattering causes mobility degradation [179], [180], which leads to ΔI_d(mobility). When the apparent ΔV_{th} is evaluated from ΔI_d(measured)/g_m, the ΔI_d(mobility) is treated as if it was caused by ΔV_{th}. In other words, the apparent ΔV_{th}= ΔI_d(measured)/g_m includes the contribution from mobility degradation to ΔI_d. As the effect of mobility degradation increases with $|V_g sense|$, it contributes to the increase in the apparent ΔV_{th} for higher $|V_g sense|$.

C. Statistics

As there is hardly any information on the statistical properties of the directly measured RTN-induced ΔV_{th}, especially in terms of its dependence on $V_g sense$, we report the DDV of this dependence here. Each line in Figure 6.11(b) represents one device and the first impression is that the apparent ΔV_{th} broadly increases for higher $|V_g sense|$. Although the ΔV_{th} for some devices can reduce for higher $|V_g sense|$ as shown in Figure 6.9(b), it is rare for a trap to be above a localized percolation path. As a result, the average (symbols in

Figure 6.11(b) increases monotonically for higher |$V_g$sense|, which is partly driven by the mobility degradation.





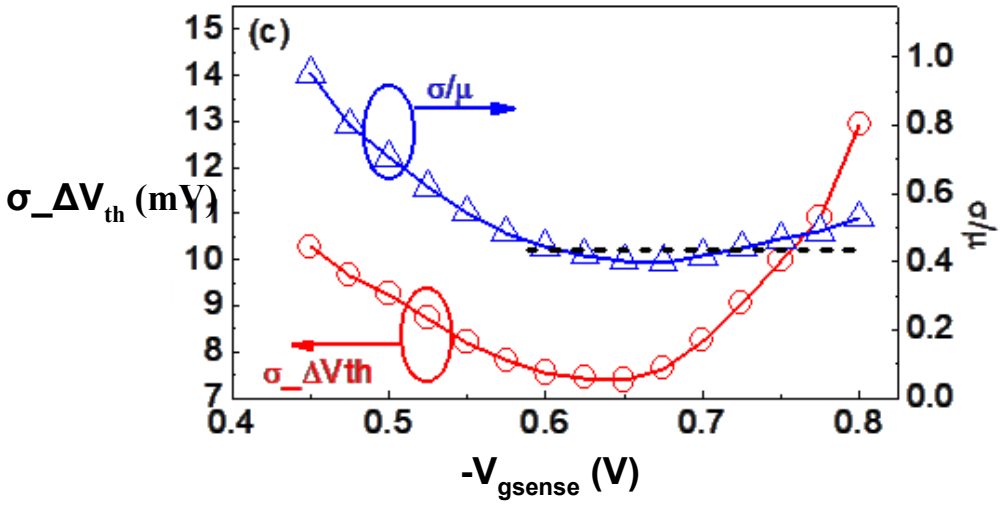


Figure 6.11 (a) The impact of sensing V_g on the apparent ΔV_{th} of a large device of 3×1 μm. The deep level traps were filled under V_g=-0.9 V for 100 sec before the measurement. ΔV_{th} is normalized against its value at $V_g=V_{th}$. (b) The stochastic variation of 135x27 nm devices. Each line is from one device. The symbols are the average. (c) Dependence of σ and σ /μ on the sensing V_g.

The standard deviation, σ, is plotted against $V_g$sense in Figure 6.11(c). It can be divided into two regions: as |$V_g$sense| increases, σ decreases first and then increases. The minimum point is around 0.65 V. To explore this further, the relative variation, σ/μ, is also plotted in Figure 6.11(c). When |$V_g$sense|>0.65 V, σ/μ only rises modestly, so that the higher σ is mainly caused by the higher μ, as shown by the symbols in Figure 6.11(b). Below 0.65 V, however, σ increases and μ decreases for lower |$V_g$sense|, resulting in a rising σ/μ. When |$V_g$sense| lowers towards |V_{th}|, the current path becomes increasingly localized, leading to higher statistical variations, even though the trapped charges remain the same.

The cumulative distribution probability of ΔV_{th} is given in Figure 6.12(a) and σ is plotted against μ in Figure 6.12(b) for $V_g$sense=V_{th}. The RTN of nMOSFETs is smaller than that

of pMOSFETs. σ follows μ by a power law with an exponent of ~0.5, agreeing with the prediction of the Defect-Centric model [19], [105], [139]. According to this model, the average ΔV_{th} induced by a trap, η, is,

$$\eta = \frac{\sigma^2}{2\mu}. \tag{6.1}$$

Using the fitted line in Figure 6.12(b), η ~ 3.2 mV is obtained. This η is ~2×q/C_{ox} approximately, where q is one electron charge and C_{ox} the gate oxide capacitance. This agrees well with the value reported for the recoverable component of NBTI of pFinFETs [180], although the test samples used here are planar pMOSFETs from a different supplier. The average number of traps, N, per device is,

$$N = \frac{\mu}{\eta}. \tag{6.2}$$

A μ~12 mV in Fig. 6.8(b) gives N~4.

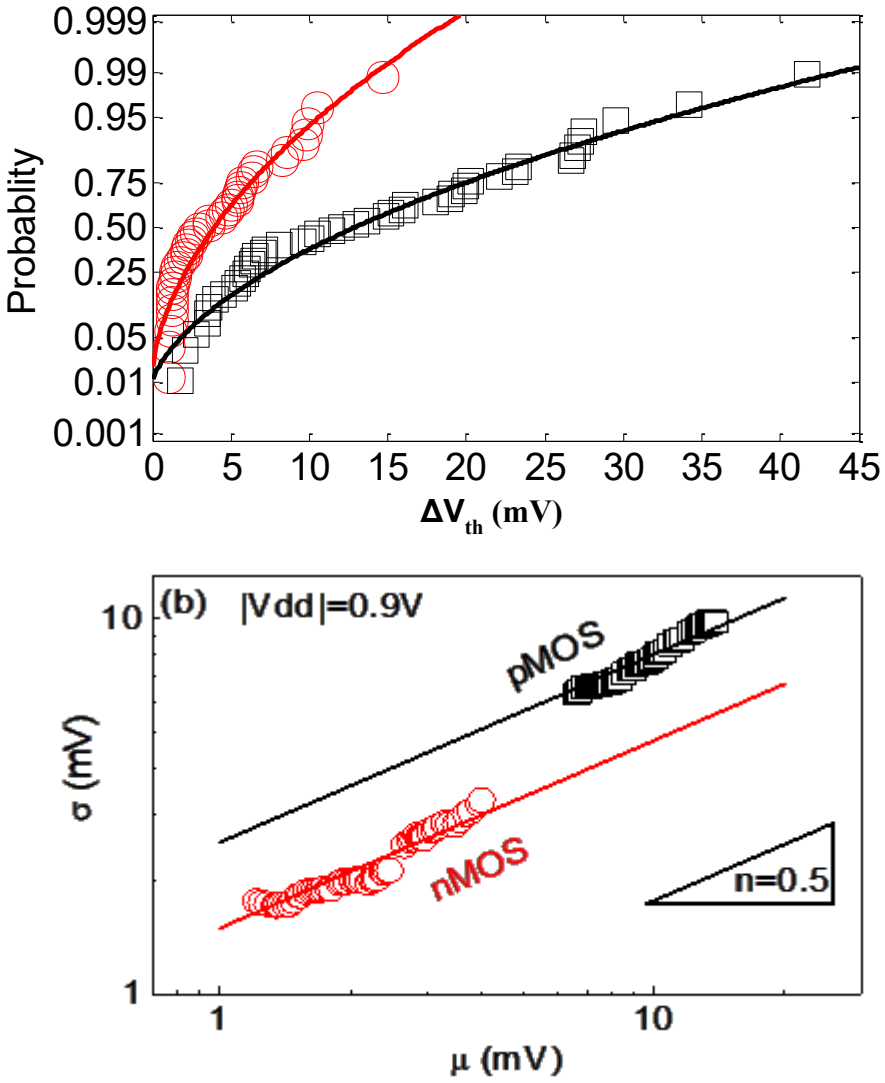


Figure 6.12(a) The cumulative distribution of ΔV_{th}. The symbols are test data and the lines are fitted with the Defect-Centric model that assumes the number of traps per device following the Poisson's distribution and the ΔV_{th} induced by a trap following exponential distribution. (b) Standard deviation versus mean. Lines show that the data follow the prediction of Defect-Centric model well with a power exponent of 0.5. The different pairs of (μ,σ) are obtained by varying the time window of "I_d monitor" from 10 μs to 100 sec in Figure 6.6a.

For nMOSFETs, the corresponding values are $\mu \sim 6.5$ mV, $\eta \sim 1.1$ mV, and $N \sim 6$. When compared with pMOSFETs, the lower RTN in nMOSFETs is caused by smaller η. Although there are more traps in nMOSFETs, they are in the *high-k* layer and further away from the conduction channel and induce a smaller ΔV_{th} [181].

6.4. Conclusions

The conventional method of 'Measure-Stress-Measure' at preset times is inapplicable for the RTN-induced ΔV_{th}, since the trap can be neutral when pulse IVs are taken. Early works estimate the RTN-induced ΔV_{th} by $\Delta I_d/g_m$ at $V_g=V_{dd}$ and its accuracy is not known. In this chapter, a 'Trigger-When-Charged (TWC)' method is proposed for directly measuring the real ΔV_{th} at $V_g=V_{th}$. By setting the trigger level close to the upper envelope of trapping-induced ΔI_d, it ensures that the pulse IV is taken when traps are charged.

Results show that there is no unique relationship between $\Delta I_d/g_m$ at $V_g=V_{dd}$ and the directly measured ΔV_{th} and their correlation is poor. The device-specific dependence of the apparent ΔV_{th} on the sensing V_g originates from the device-to-device variation (DDV) of relative local current density under a trap at V_{th}. Moreover, on average, $\Delta I_d/g_m(V_{dd})$ doubles $\Delta V_{th}(V_{th})$ and the charge-induced mobility degradation through columbic scattering plays a role.

The TWC is applicable to devices with or without analysable RTN signals. For the first time, it is used for assessing the statistical properties of RTN-induced ΔV_{th}, especially in terms of its dependence on $V_g$sense. For the same trapped charges, it is found that σ has a minimum around $|V_g\text{sense}|=0.65$ V. The increase in σ when $|V_g|$ lowers toward $|V_{th}|$ is explained by an increased localization of current path. The DDV follows the Defect-Centric model. For the 135 × 27 nm pMOSFETs used in this work, the average ΔV_{th}

induced per trap is ~3.2 mV and not far compared to ~3.4 mV for device of 90 x 27 nm as reported in [177] .

7 Conclusion and Future Works

7.1 Conclusions

This project studies the hot carrier aging (HCA), random telegraph noise (RTN), and their interactions. As down-scaling of device sizes continues, HCA become increasingly important. The HCA kinetics is established through detailed measurements and its capability in predicting the device lifetime is demonstrated. The attentions were then turned to the interaction between HCA and RTN. Finally, a new Trigger-When-Charged (TWC) technique is developed to enable direct measurement of RTN-induced jitter in the threshold voltage.

7.1.1 Conclusions on the kinetics and prediction of HCA

Hot carrier aging can be a critical factor for current and future CMOS. HCA can cause higher degradation to the device due to channel length down-scaling and larger time exponents. HCA also has low recovery compared to NBTI as shown in Figure 4.2(c). This new threat makes it worthwhile to re-visit HCA. Lifetime prediction requires constant time exponent '*n*'. However, it is reported that the aging mechanism and time exponent '*n*' are changed under different bias. Two pressing questions are: Can the JEDEC method predict lifetime under use-bias? and how can we correctly evaluate '*n*' for HCA?

The results show that there are two sources of errors for extracting '*n*'. One of them is from using the forward saturation current to monitor the HCA. HCA above the space charge near the drain is screened off and this leads to an increase of '*n*'. To correct this error, '*n*' should be extracted from the linear threshold voltage shift. The other source of error is the inclusion of as-grown defects that were not caused by HCA. This error can be removed by using the lower envelope of the fluctuation.

In this study, the Voltage Step Stress Technique (VSST) method is used to stress the device at V_g=V_d from 1.3 V to 1.7 V. During each stress, IV measurements are taken periodically. Based on the results obtained for large devices, it shows good agreement between prediction and experiment data. For small devices, HCA induces a time-dependent device-to-device variation (TDDDV). This study verifies that the HCA under use-V_{dd} can be predicted by the power law extracted from VSST method.

7.1.2 Conclusion on impact of hot carrier aging (HCA) on random telegraph noise (RTN)

Random telegraph noise (RTN) is becoming a major challenge for low power circuits. HCA, as has been reported, either can increase the RTN or has little effect on it. RTN is caused by a single trap at the gate dielectric capturing and emitting a carrier. When there are more than a few traps, it become complex RTN signals in the form of within-device-fluctuation (WDF).

For scaling devices, aging also occurs in the form of bias temperature instabilities (BTI) or hot carrier stresses. It has been reported that aging can either increase the RTN/WDF or makes little contribution to it. The interaction between RTN/WDF and hot carrier aging (HCA) is not fully understood.

Based on the study in this project, it is found that HCA can either increase or reduce the average RTN modestly. The HCA influences RTN mainly through changing the density of current flow under a trap. When the local current increases, RTN rises. For devices with abnormally large RTN, RTN reduces substantially after HCA. The abnormally large RTN is caused by traps above the percolation path of current. HCA changes the current path, so that the percolation path of the current is moved away from these traps, resulting in the reduction of RTN.

7.1.3 Conclusion on Trigger-When-Charged technique

Although RTN has become a major concern for the circuit design and has attracted much attention, there are no direct measurements of the RTN-induced jitter in V_{th}. This is because, the charge-discharge of traps is highly dynamic and the ΔV_{th} is low. There are also devices that show complex RTN. These devices were simply deselected in some early works.

The Trigger-When-Charged (TWC) method is introduced for directly measuring the real ΔV_{th} at $V_g=V_{th}$. This was done by setting the trigger level close to the upper envelope of trapping-induced ΔI_d. Thus, it ensures that the pulse IV is taken when traps are charged.

The experimental results show that there is no unique relationship between $\Delta I_d/g_m$ at $V_g=V_{dd}$ and the directly measured ΔV_{th} and their correlation is poor. The device-specific dependence of the apparent ΔV_{th} on the sensing V_g originates from the device-to-device variation (DDV) of relative local current density under a trap at V_{th}.

7.2 Future Works

Hot carrier aging (HCA) and random telegraph noise (RTN) are becoming important to understand as the device scaling. Although progress has been made in this project, there are a lot of interesting topics awaiting further investigations, including but not limited to the following:

7.2.1 HCA and BTI Coupling

Bias temperature instability is another important degradation mechanism in nano-scale devices. Devices such as SRAM also experience BTI while the circuit is in operation. HCA is caused by hot carrier injection on the drain side while BTI is caused by the filling

and generation of defects across the gate dielectrics. It is worthwhile investigating the coupling of these two degradation mechanisms.

Under both BTI and HCA, it is reported that the aging kinetics follows a power law. In real circuits, devices can suffer both BTI and HCA and the there is little information on the kinetics combining these two mechanisms and on how to predict the device lifetime. Further works should investigate the similarity and differences in defects created by these two mechanisms. If they originate from the same precursors, the consumption of these precursors by one mechanism will make them unavailable to the other mechanism, so that one will affect the other.

7.2.2 Defect losses

Defect losses were observed in this project. After stress and then anneal, there are cases where RTN/WDF did not return to its pre-stress high level and an example is given in Figure 5.10. This agrees with the defect loss reported in early works [137], [155]. If the critical trap is lost after the stress and annealing process, the RTN/WDF cannot return to its fresh abnormally high level, even though the original current distribution is restored. Figure 5.12 confirms that similar loss also occurs for a device fabricated by the 22 nm process. There is little information available on defect losses at present and the underlying mechanism is not clear. Further work is needed in this area.

References

[1] S. Davis, “Transistor Count Trends Continue to Track with Moore’s Law,” *Semiconductor Digest*, 2020. [Online]. Available: https://www.semiconductor-digest.com/2020/03/10/transistor-count-trends-continue-to-track-with-moores-law/. [Accessed: 13-Oct-2020].

[2] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, “Design of Ion-Implanted MOSFET’s With Very Small Physical Dimensions,” *IEEE J. Solid-State Circuits*, 1974.

[3] G. E. Moore, “Cramming more components onto integrated circuits,” *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.

[4] K. O. Jeppson and C. M. Svensson, “Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices,” *J. Appl. Phys.*, vol. 48, no. 5, pp. 2004–2014, 1977.

[5] Y. Kiinizuka *et al.*, “The impact of bias temperature instability for direct-tunneling ultra-thin gate oxide on MOSFET scaling,” *Syrnposlurn VLSl Technol.*, pp. 73–74, 1999.

[6] 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, 1995.

[7] A. Kumar, T. H. Ning, M. V. Fischetti, and E. Gusev, “Hot-carrier charge trapping and reliability in high-k dielectrics,” in *2002 Symposium on VLSI Technology. Digest of Technical Papers (Cat. No.01CH37303)*, 2002, pp. 152–153.

[8] Chenming Hu, Simon C. Tam, Fu-Chieh Hsu, Ping-Keung Ko, Tung-Yi Chan, and K. W. Terrill, “Hot-Electron-Induced MOSFET Degradation - Model, Monitor, and Improvement,” *IEEE J. Solid-State Circuits*, vol. 20, no. 1, pp. 295–305, 1985.

[9] V. H. Chan and J. E. Chung, “The Impact of NMOSFET Hot-Carrier Degradation on

CMOS Analog Subcircuit Performance," *IEEE J. Solid-State Circuits*, 1995.

[10] C. Hu, "LUCKY-ELECTRON MODEL OF CHANNEL HOT ELECTRON EMISSION Chenming," vol. 2, pp. 22–25, 1979.

[11] T. C. Ong, P. K. Ko, and C. Hu, "Hot-Carrier Current Modeling and Device Degradation in Surface-Channel p-MOSFET's," *IEEE Trans. Electron Devices*, 1990.

[12] "ITRS 2009 web site:http://www.itrs2.net/itrs-reports.html." .

[13] IEEE, "THE INTERNATIONAL ROADMAP FOR DEVICES INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS 2017 EDITION MORE MOORE," 2018.

[14] H. Kurata *et al.*, "The impact of random telegraph signals on the scaling of multilevel flash memories," in *IEEE Symposium on VLSI Circuits, Digest of Technical Papers*, 2006.

[15] M. A. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *Technical Digest - International Electron Devices Meeting*, 2003, pp. 14.4.1-14.4.4.

[16] J. H. Stathis and S. Zafar, "The negative bias temperature instability in MOS devices: A review," *Microelectron. Reliab.*, vol. 46, no. 2–4, pp. 270–286, Feb. 2006.

[17] S. Mahapatra, S. De, K. Joshi, S. Mukhopadhyay, R. K. Pandey, and K. V. R. M. Murali, "Understanding Process Impact of Hole Traps and NBTI in HKMG p-MOSFETs Using Measurements and Atomistic Simulations," *IEEE Electron Device Lett.*, vol. 34, no. 8, pp. 963–965, Aug. 2013.

[18] Z. J. Z. Ji, L. Lin, J. F. Z. J. F. Zhang, B. Kaczer, and G. Groeseneken, "NBTI Lifetime Prediction and Kinetics at Operation Bias Based on Ultrafast Pulse Measurement," *IEEE Trans. Electron Devices*, vol. 57, no. 1, pp. 228–237, 2010.

[19] B. Kaczer *et al.*, "Origin of NBTI variability in deeply scaled pFETs," *2010 IEEE Int. Reliab. Phys. Symp.*, pp. 26–32, 2010.

[20] M. Denais *et al.*, “Interface Trap Generation and Hole Trapping Under NBTI and PBTI in Advanced CMOS Technology With a 2-nm Gate Oxide,” *IEEE Trans. Device Mater. Reliab.*, vol. 4, no. 4, pp. 715–722, Dec. 2004.

[21] B. Cheng, A. R. Brown, and A. Asenov, “Impact of NBTI / PBTI on SRAM Stability Degradation,” *IEEE Electron Device Lett.*, vol. 32, no. 6, pp. 740–742, Jun. 2011.

[22] Z. J. Z. Ji, J. F. Z. J. F. Zhang, M. H. C. M. H. Chang, B. Kaczer, and G. Groeseneken, “An Analysis of the NBTI-Induced Threshold Voltage Shift Evaluated by Different Techniques,” *IEEE Trans. Electron Devices*, vol. 56, no. 5, pp. 1086–1093, 2009.

[23] J. Franco, B. Kaczer, M. Cho, G. Eneman, G. Groeseneken, and T. Grasser, “Improvements of NBTI reliability in SiGe p-FETs,” in *IEEE International Reliability Physics Symposium Proceedings*, 2010, pp. 1082–1085.

[24] T. H. Ning, C. M. Osburn, and H. N. Yu, “Emission probability of hot electrons from silicon into silicon dioxide,” *J. Appl. Phys.*, 1977.

[25] K. N. Quader, E. R. Minami, W. J. Wei-Jen, H. Huang, P. K. Ko, and C. Hu, “Hot-Carrier-Reliability Design Guidelines for CMOS Logic Circuits,” *IEEE J. Solid-State Circuits*, 1994.

[26] E. Takeda, “A cross section of VLSI reliability-hot carriers, dielectrics and metallization,” *Semiconductor Science and Technology*. 1994.

[27] A. Asenov, S. Kaya, and A. R. Brown, “Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness,” *IEEE Trans. Electron Devices*, vol. 50, pp. 1254–1260, 2003.

[28] H. Fukutome *et al.*, “Suppression of poly-gate-induced fluctuations in carrier profiles of sub-50nm MOSFETs,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2006.

[29] S. Borkar, “Designing Reliable Systems from Unreliable Components: The Challenges of

Transistor Variability and Degradation," *IEEE Micro*, vol. 25, no. 6, pp. 10–16, Nov. 2005.

[30] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 fjim MOSFET's: A 3-D 'atomistic' simulation study," *IEEE Trans. Electron Devices*, vol. 45, pp. 2505–2513, 1998.

[31] S. Pae *et al.*, "BTI reliability of 45nm high-k/metal-gate process technology," *2008 IEEE Int. Reliab. Phys. Symp.*, pp. 352–357, 2008.

[32] K. Takeuchi, T. Nagumo, S. Yokogawa, K. Imai, and Y. Hayashi, "Single-Charge-Based Modeling of Transistor Characteristics Fluctuations Based on Statistical Measurement of RTN Amplitude," pp. 2008–2009, 2009.

[33] a Asenov and a R. Brown, "Statistical aspects of NBTI/PBTI and impact on SRAM yield," *2011 Des. Autom. Test Eur.*, pp. 1–6, Mar. 2011.

[34] F. Adamu-Lema *et al.*, "Accuracy and issues of the spectroscopic analysis of RTN traps in nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 60, no. 2, pp. 833–839, 2013.

[35] P. Jain, A. Paul, X. Wang, and C. H. Kim, "A 32nm SRAM reliability macro for recovery free evaluation of NBTI and PBTI," in *2012 International Electron Devices Meeting*, 2012, pp. 9.7.1-9.7.4.

[36] C. Ho, S. Member, M. K. Hassan, S. Y. Kim, and K. Roy, "Analysis of Stability Degradation of SRAMs Using a Physics-Based PBTI Model," *IEEE Electron Device Lett.*, vol. 35, no. 9, pp. 951–953, Sep. 2014.

[37] A. Asenov, "Simulation of statistical variability in nano MOSFETs," in *Digest of Technical Papers - Symposium on VLSI Technology*, 2007.

[38] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, "The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-

stress," in *IEEE International Reliability Physics Symposium Proceedings*, 2010, pp. 7–15.

[39] "Intel Identifies Chipset Design Error, Implementing Solution | Intel Newsroom." [Online]. Available: https://newsroom.intel.com/news-releases/intel-identifies-chipset-design-error-implementing-solution/#gs.x4oetn. [Accessed: 17-Feb-2020].

[40] B. E. Deal, "Standardized Terminology for Oxide Charges Associated with Thermally Oxidized Silicon," *IEEE Trans. Electron Devices*, 1980.

[41] JEDEC, "Failure Mechanisms and Models for Semiconductor Devices," *Pap. No. JEP*, vol. 122, no. MARCH, p. 2010, 2010.

[42] Y. Miura and Y. Matukura, "Investigation of Silicon-Silicon Dioxide Interface Using MOS Structure," *Japanese J. Appl. Physics, Part 1 Regul. Pap. Short Notes Rev. Pap.*, 1966.

[43] A. Goetzberger and H. E. Nigh, "Surface Charge After Annealing of AlSi0,Si Structures Under Bias," *Proc. IEEE*, vol. 5, p. 1966, 1966.

[44] A. K. Sinha, H. J. Levinstein, and T. E. Smith, "THERMAL STRESSES AND CRACKING RESISTANCE OF DIELECTRIC FILMS (SiN, Si//3N//4, AND SiO//2) ON Si SUBSTRATES.," *Journal of Applied Physics*, vol. 49, no. 4. pp. 2423–2426, 1978.

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

[46] J. Lee, A. S. Oates, and S. Member, "Characterization of NBTI-Induced Interface State and Hole Trapping in SiON Gate Dielectrics of p-MOSFETs," *IEEE Trans. Device Mater. Reliab.*, vol. 10, no. 2, pp. 174–181, 2010.

[47] a. E. E. Islam, S. Mahapatra, S. Deora, V. D. Maheta, and M. a. A. Alam, "On the differences between ultra-fast NBTI measurements and Reaction-Diffusion theory," *2009*

IEEE Int. Electron Devices Meet., no. c, pp. 1–4, Dec. 2009.

[48] 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.

[49] 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.

[50] JEDEC Standard, "A Procedure for Measuring P-Channel MOSFET Negative Bias," *JEDEC Eng. Stand. Publ.*, no. November, 2004.

[51] J. F. Zhang, "Oxide Defects," in *Bias Temperature Instability for Devices and Circuits*, Springer New York, 2014, pp. 253–285.

[52] Z. Q. Teo, D. S. Ang, and K. S. See, "Can the reaction-diffusion model explain generation and recovery of interface states contributing to NBTI?," *2009 IEEE Int. Electron Devices Meet.*, pp. 1–4, Dec. 2009.

[53] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A two-stage model for negative bias temperature instability," *2009 IEEE Int. Reliab. Phys. Symp.*, pp. 33–44, Apr. 2009.

[54] T. Grasser *et al.*, "Recent advances in understanding the bias temperature instability," in *2010 International Electron Devices Meeting*, 2010, pp. 4.4.1-4.4.4.

[55] S. Gupta, B. Jose, K. Joshi, A. Jain, M. A. Alam, and S. Mahapatra, "A Comprehensive and Critical Re-assement of 2-Stage Energy Level NBTI Model," *2012 IEEE Int. Phys. Symp.*, pp. 1–6, 2012.

[56] a. Narendiran and B. Bindu, "Simulation studies of Negative Bias Temperature Instability in FinFETs using two-stage model," *2012 Int. Conf. Devices, Circuits Syst.*, pp. 555–557, Mar. 2012.

[57] R. Russin, M. Muhamad, S. Shahabuddin, N. Soin, and M. F. Bukhori, “A Study using Two Stage NBTI Model for 32 nm high-k PMOSFET,” *2013 IEEE Int. Conf. ofElectron Devices Solid-State Circuits*, pp. 2–3, 2013.

[58] C. R. Parthasarathy, M. Denais, V. Huard, G. Ribes, E. Vincent, and A. Bravaix, “New Insights Into Recovery Characteristics During PMOS NBTI and CHC Degradation,” *IEEE Trans. Device Mater. Reliab.*, vol. 7, no. 1, pp. 130–137, Mar. 2007.

[59] G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng, and D. L. Kwong, “Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling,” *IEEE Electron Device Lett.*, vol. 23, no. 12, pp. 734–736, Dec. 2002.

[60] A. J. Lelis and T. R. Oldham, “Time Dependence of Switching Oxide Traps,” *IEEE Trans. Nucl. Sci.*, 1994.

[61] J. F. J. F. J. F. Zhang *et al.*, “Hole traps in silicon dioxides. Part I. Properties,” *IEEE Trans. Electron Devices*, vol. 51, no. 8, pp. 1267–1273, 2004.

[62] Y. M. Randriamihaja *et al.*, “New Hot Carrier degradation modeling reconsidering the role of EES in ultra short N-channel MOSFETs,” in *2013 IEEE International Reliability Physics Symposium (IRPS)*, 2013, p. XT.1.1-XT.1.5.

[63] J. H. Stathis *et al.*, “Reliability challenges for the 10nm node and beyond,” in *2014 IEEE International Electron Devices Meeting*, 2014, pp. 20.6.1-20.6.4.

[64] M. Cho *et al.*, “Channel Hot Carrier Degradation Mechanism in Long/Short Channel <formula formulatype="inline"><tex Notation="TeX">n</tex></formula>-FinFETs,” *IEEE Trans. Electron Devices*, vol. 60, no. 12, pp. 4002–4007, Dec. 2013.

[65] I. C. Chen, J. Y. Choi, T. Y. Chan, and C. Hu, “The Effect of Channel Hot-Carrier Stressing on Gate-Oxide Integrity in MOSFET’s,” *IEEE Trans. Electron Devices*, 1988.

[66] 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 40nm NMOS node at

high temperature," in *IEEE International Reliability Physics Symposium Proceedings*, 2009, pp. 531–548.

[67] V. Huard, C. R. Parthasarathy, A. Bravaix, C. Guerin, and E. Pion, "CMOS device design-in reliability approach in advanced nodes," in *IEEE International Reliability Physics Symposium Proceedings*, 2009, pp. 624–633.

[68] J.-S. Goo, Y.-G. Kim, H. L'Yee, H.-Y. Kwon, and H. Shin, "An analytical model for hot-carrier-induced degradation of deep-submicron n-channel LDD MOSFETs," *Solid. State. Electron.*, vol. 38, no. 6, pp. 1191–1196, Jun. 1995.

[69] 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 *2012 International Electron Devices Meeting*, 2012, pp. 30.5.1-30.5.4.

[70] S. Tyaginov, M. Bina, J. Franco, Y. Wimmer, and D. Osintsev, "A Predictive Physical Model for Hot-Carrier Degradation in Ultra-Scaled MOSFETs," pp. 6–9.

[71] M. Bina *et al.*, "Predictive Hot-Carrier Modeling of n-Channel MOSFETs," *IEEE Trans. Electron Devices*, vol. 61, no. 9, pp. 3103–3110, Sep. 2014.

[72] R. S. Muller, T. I. Kamins, and M. Chan, "Device Electronics for Integrated devices," *Wiley*. p. 560, 2003.

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

[74] W. Arfaoui *et al.*, "Energy-driven hot-carrier model in advanced nodes," *IEEE Int. Reliab. Phys. Symp. Proc.*, vol. 33, no. 0, pp. 1–5, 2014.

[75] X. M. Li and M. J. Deen, "Determination of interface state density in MOSFETs using the spatial profiling charge pumping technique," *Solid. State. Electron.*, vol. 35, no. 8, pp. 1059–1063, Aug. 1992.

[76] C.-Y. Lu, K.-S. Chang-Liao, P.-H. Tsai, and T.-K. Wang, "Depth Profiling of Border

Traps in MOSFET With High-<tex>$kappa$</tex>Gate Dielectric by Charge-Pumping Technique," *IEEE Electron Device Lett.*, vol. 27, no. 10, pp. 859–862, Oct. 2006.

[77] C. Lu *et al.*, "Profiling of Channel-Hot-Carrier Stress-Induced Trap Distributions Along Channel and Gate Dielectric in High-k Gated MOSFETs by a Modified Charge Pumping Technique," *IEEE Trans. Electron Devices*, vol. 61, no. 4, pp. 936–942, Apr. 2014.

[78] W. H. Card and P. K. Chaudhari, "Characteristics of burst noise," *Proc. IEEE*, vol. 53, no. 6, pp. 652–653, 1965.

[79] E. Simoen, B. Kaczer, M. Toledano-Luque, and C. Claeys, "Random Telegraph Noise: From a Device Physicist's Dream to a Designer's Nightmare," in *ECS Transactions*, 2011, vol. 39, no. 1, pp. 3–15.

[80] K. S. Ralls *et al.*, "Discrete resistance switching in submicrometer silicon inversion layers: Individual interface traps and low-frequency (1f?) noise," *Phys. Rev. Lett.*, vol. 52, no. 3, pp. 228–231, Jan. 1984.

[81] M. J. Uren, D. J. Day, and M. J. Kirton, "1/f and random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 47, no. 11, p. 1195, Jan. 1985.

[82] T. Kang *et al.*, "Random telegraph noise in GaN-based light-emitting diodes," *Electron. Lett.*, 2011.

[83] C. Monzio Compagnoni, M. Ghidotti, A. L. Lacaita, A. S. Spinelli, and A. Visconti, "Random telegraph noise effect on the programmed threshold-voltage distribution of flash memories," *IEEE Electron Device Lett.*, 2009.

[84] "ITRS 2011 web site: http://www.itrs2.net/2011-itrs.html." .

[85] K. Kandiah, "Random telegraph signal currents and low-frequency noise in junction field effect transistors," *IEEE Trans. Electron Devices*, vol. 41, pp. 2006–2015, 1994.

[86] E. Simoen and C. Claeys, "Random telegraph signal: A local probe for single point defect

studies in solid-state devices," in *Materials Science and Engineering B: Solid-State Materials for Advanced Technology*, 2002, vol. 91–92, pp. 136–143.

[87] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/ *f*) noise," *Adv. Phys.*, vol. 38, no. 4, pp. 367–468, 1989.

[88] M. J. Kirton, M. J. Uren, S. Collins, M. Schulz, A. Karmann, and K. Scheffer, "Individual defects at the Si:SiO2 interface," *Semicond. Sci. Technol.*, 1989.

[89] H. Miki *et al.*, "Understanding short-term BTI behavior through comprehensive observation of gate-voltage dependence of RTN in highly scaled high-κ / metal-gate pFETs," in *Digest of Technical Papers - Symposium on VLSI Technology*, 2011.

[90] 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," pp. 836–839, 2014.

[91] N. Tega, "Study on Impact of Random Telegraph Noise on Scaled MOSFETs," no. February, 2014.

[92] C. Liu *et al.*, "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," pp. 571–574, 2011.

[93] D. Angot *et al.*, "BTI variability fundamental understandings and impact on digital logic by the use of extensive dataset," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 15.4.1-15.4.4.

[94] N. Seoane, G. Indalecio, E. Comesana, M. Aldegunde, A. J. Garcia-Loureiro, and K. Kalna, "Random Dopant, Line-Edge Roughness, and Gate Workfunction Variability in a Nano InGaAs FinFET," *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 466–472, Feb. 2014.

[95] S. K. Saha, "Modeling Process Variability in Scaled CMOS Technology," *IEEE Des. Test Comput.*, vol. 27, no. 2, pp. 8–16, Mar. 2010.

[96] J. Martin-Martinez, R. Rodriguez, M. Nafria, and X. Aymerich, "Time-Dependent Variability Related to BTI Effects in MOSFETs: Impact on CMOS Differential Amplifiers," *IEEE Trans. Device Mater. Reliab.*, vol. 9, no. 2, pp. 305–310, Jun. 2009.

[97] R. Gautam, M. Saxena, R. S. Gupta, and M. Gupta, "Hot-Carrier Reliability of Gate-All-Around MOSFET for RF/Microwave Applications," *IEEE Trans. Device Mater. Reliab.*, vol. 13, no. 1, pp. 245–251, Mar. 2013.

[98] S. Poli *et al.*, "Temperature Dependence of the Threshold Voltage Shift Induced by Carrier Injection in Integrated STI-Based LDMOS Transistors," *IEEE Electron Device Lett.*, vol. 32, no. 6, pp. 791–793, Jun. 2011.

[99] G. Sasse, "Device degradation models for circuit reliability simulation," in *2013 IEEE International Integrated Reliability Workshop Final Report*, 2013, pp. 42–42.

[100] A. Kerber, D. Lipp, M. Trentzsch, B. P. Linder, and E. Cartier, "Impact of TDDB in MG/HK devices on circuit functionality in advanced CMOS technologies," in *2011 International Electron Devices Meeting*, 2011, pp. 18.1.1-18.1.4.

[101] Y. Mitani and A. Toriumi, "Experimental study on origin of V<inf>TH</inf> variability under NBT stress," *2011 Int. Reliab. Phys. Symp.*, p. PL.2.1-PL.2.6, Apr. 2011.

[102] G. Rosa, W. Ng, S. Rauch, R. Wong, and J. Sudijono, "Impact of NBTI Induced Statistical Variation to SRAM Cell Stability," *2006 IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 274–282, 2006.

[103] J. Martin-martinez, R. Rodriguez, M. Nafria, X. Aymerich, R. Rodríguez, and S. Member, "Time-Dependent Variability Related to BTI Effects in MOSFETs: Impact on CMOS Differential Amplifiers," *IEEE Trans. Device Mater. Reliab.*, vol. 9, no. 2, pp. 305–310, Jun. 2009.

[104] S. E. Rauch, “The statistics of NBTI-induced V/sub T/ and β mismatch shifts in pMOSFETs,” *IEEE Trans. Device Mater. Reliab.*, vol. 2, no. 4, pp. 89–93, Dec. 2002.

[105] J. Franco *et al.*, “Impact of single charged gate oxide defects on the performance and scaling of nanoscaled FETs,” in *2012 IEEE International Reliability Physics Symposium (IRPS)*, 2012, pp. 5A.4.1-5A.4.6.

[106] M. Duan *et al.*, “Time-dependent variation: A new defect-based prediction methodology,” in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, 2014, pp. 1–2.

[107] T. Aichinger, M. Nelhiebel, and T. Grasser, “On the temperature dependence of NBTI recovery,” *Microelectron. Reliab.*, vol. 48, no. 8–9, pp. 1178–1184, Aug. 2008.

[108] T. Ghani, “Innovations to extend CMOS nano-transistors to the limit,” *Proceeding Thirteen. Int. Symp. Low power Electron. Des. - ISLPED ’08*, pp. 301–302, 2008.

[109] J. F. Zhang, H. K. Sii, G. Groeseneken, S. Member, and R. Degraeve, “Degradation of Oxides and Oxynitrides Under Hot Hole Stress,” vol. 47, no. 2, pp. 378–386, 2000.

[110] B. Stackhouse *et al.*, “A 65 nm 2-Billion Transistor Quad-Core Itanium Processor,” *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 18–31, Jan. 2009.

[111] Y. Wang, S. Cotofana, and L. Fang, “A unified aging model of NBTI and HCI degradation towards lifetime reliability management for nanoscale MOSFET circuits,” *Proc. 2011 IEEE/ACM Int. Symp. Nanoscale Archit. NANOARCH 2011*, pp. 175–180, 2011.

[112] J. B. Velamala, V. Reddy, R. Zheng, S. Krishnan, and Y. Cao, “On the bias dependence of time exponent in NBTI and CHC effects,” in *2010 IEEE International Reliability Physics Symposium*, 2010, pp. 650–654.

[113] L. Ma *et al.*, “Physical understanding of hot carrier injection variability in deeply scaled nMOSFETs,” vol. 15.

[114] D. S. Ang and S. Wang, “Recovery of the NBTI-stressed ultrathin gate p-MOSFET: The

role of deep-level hole traps," *IEEE Electron Device Lett.*, 2006.

[115] a. Ortiz-Conde, F. J. García Sánchez, J. J. Liou, a. Cerdeira, M. Estrada, and Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectron. Reliab.*, vol. 42, pp. 583–596, 2002.

[116] D. K. Schroder, *Semiconductor Material and Device Characterization: Third Edition*. 2005.

[117] C. Measurements, "Application Note Series Gate Dielectric Capacitance-Voltage Characterization Using the Model 4200 Semiconductor Characterization System Understanding MOS-Capacitor."

[118] A. Kerber *et al.*, "Characterization of the VT-instability in SIO2/HFO2 gate dielectrics," in *IEEE International Reliability Physics Symposium Proceedings*, 2003.

[119] S. F. W. M. Hatta, "Probing Technique for Energy Distribution of Positive Charges in Gate Dielectrics and Its Application To Lifetime Prediction a Thesis Submitted in Partial Fulfilment of the Requirements of Liverpool John Moores University for the Degree of Doctor of Philo," Liverpool John Moores University, 2013.

[120] M. Cho *et al.*, "Channel Hot Carrier Degradation Mechanism in Long/Short Channel n-FinFETs," vol. 60, no. 12, pp. 4002–4007, 2013.

[121] P. Magnone *et al.*, "Impact of Hot Carriers on nMOSFET Variability in 45- and 65-nm CMOS Technologies," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2347–2353, Aug. 2011.

[122] N. Hui-Hsin Hsu, E. Chen, L. S. Huang, O. Cheng, and I. C. Chen, "Intrinsic hot-carrier degradation of nMOSFETs by decoupling PBTI component in 28nm high-k/metal gate stacks," in *2012 IEEE International Reliability Physics Symposium (IRPS)*, 2012, p. XT.13.1-XT.13.4.

[123] E. Amat *et al.*, "Competing Degradation Mechanisms in Short-Channel Transistors Under

Channel Hot-Carrier Stress at Elevated Temperatures," *IEEE Trans. Device Mater. Reliab.*, vol. 9, no. 3, pp. 454–458, Sep. 2009.

[124] F. Cacho, P. Mora, W. Arfaoui, X. Federspiel, and V. Huard, "HCI/BTI COUPLED MODEL: THE PATH FOR ACCURATE AND PREDICTIVE RELIABILITY SIMULATIONS," *2014 IEEE Int. Reliab. Phys. Symp.*, vol. 2009, pp. 4–8, Jun. 2014.

[125] A. Bravaix *et al.*, "Impact of the gate-stack change from 40nm node SiON to 28nm High-K Metal Gate on the Hot-Carrier and Bias Temperature damage," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 1–9, 2013.

[126] G. T. Sasse, F. G. Kuper, and J. Schmitz, "MOSFET degradation under RF stress," *IEEE Trans. Electron Devices*, 2008.

[127] 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*, 2011.

[128] A. Bravaix, V. Huard, D. Goguenheim, and E. Vincent, "Hot-carrier to cold-carrier device lifetime modeling with temperature for low power 40nm Si-bulk NMOS and PMOS FETs," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2011, pp. 27.5.1-27.5.4.

[129] S. Ramey and J. Hicks, "SILC and gate oxide breakdown characterization of 22nm tri-gate technology," in *IEEE International Reliability Physics Symposium Proceedings*, 2014.

[130] Z. Ji *et al.*, "Negative bias temperature instability lifetime prediction: Problems and solutions," *2013 IEEE Int. Electron Devices Meet.*, no. 1, pp. 15.6.1-15.6.4, Dec. 2013.

[131] T. Liu, C.-C. Chen, and L. Milor, "Reliability and Aging Analysis on SRAMs Within Microprocessor Systems," in *Dependability Engineering*, InTech, 2018.

[132] U. Bhattacharya *et al.*, "45nm SRAM Technology Development and Technology Lead

Vehicle.," *Intel Technol. J.*, 2008.

[133] Z. Ji *et al.*, "A single device based voltage step stress (VSS) technique for fast reliability screening," *2014 IEEE Int. Reliab. Phys. Symp.*, no. 1, p. GD.2.1-GD.2.4, Jun. 2014.

[134] M. Duan *et al.*, "Key issues and techniques for characterizing time-dependent device-to-device variation of SRAM," *2013 IEEE Int. Electron Devices Meet.*, no. 1, pp. 31.3.1-31.3.4, Dec. 2013.

[135] F. Mechanisms and S. Devices, "JEDEC JEP122E - Failure Mechanisms and Models for Semiconductor Devices," 2009.

[136] D. Varghese, P. Moens, and M. A. Alam, "on-State Hot Carrier Degradation in Drain-Extended NMOS Transistors," *IEEE Trans. Electron Devices*, vol. 57, no. 10, pp. 2704–2710, Oct. 2010.

[137] M. Duan *et al.*, "New insights into defect loss, slowdown, and device lifetime enhancement," *IEEE Trans. Electron Devices*, 2013.

[138] M. Aoulaiche *et al.*, "Negative bias temperature instability on Si-passivated Ge-interface," in *IEEE International Reliability Physics Symposium Proceedings*, 2008, pp. 358–362.

[139] 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. PP, no. 99, pp. 1–1, 2014.

[140] L. M. Procel *et al.*, "A Defect-Centric perspective on channel hot carrier variability in nMOSFETs," *Microelectron. Eng.*, 2015.

[141] B. Kaczer *et al.*, "Origins and implications of increased channel hot carrier variability in nFinFETs," in *IEEE International Reliability Physics Symposium Proceedings*, 2015.

[142] L. M. Procel, F. Crupi, L. Trojman, J. Franco, and B. Kaczer, "A Defect-Centric Analysis of the Temperature Dependence of the Channel Hot Carrier Degradation in nMOSFETs," *IEEE Trans. Device Mater. Reliab.*, 2016.

[143] C. Z. Zhao, J. F. Zhang, G. Groeseneken, and R. Degraeve, "Hole-traps in silicon dioxides - Part II: Generation mechanism," *IEEE Trans. Electron Devices*, 2004.

[144] T. Grasser *et al.*, "Hydrogen-related volatile defects as the possible cause for the recoverable component of NBTI," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 15.5.1-15.5.4.

[145] P. Ren *et al.*, "New observations on complex RTN in scaled high-κ/metal-gate MOSFETs — The role of defect coupling under DC/AC condition," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 31.4.1-31.4.4.

[146] K. Ota, M. Saitoh, C. Tanaka, D. Matsushita, and T. Numata, "Systematic study of RTN in nanowire transistor and enhanced RTN by hot carrier injection and negative bias temperature instability," *Dig. Tech. Pap. - Symp. VLSI Technol.*, vol. 10, no. c, pp. 2013–2014, 2014.

[147] 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.

[148] S. Singh, E. Mannila, D. Golubev, J. Peltonen, and J. Pekola, "Determining the parameters of a random telegraph signal by digital low pass filtering."

[149] J. F. Zhang, "Defects and instabilities in Hf-dielectric/SiON stacks (Invited Paper)," *Microelectron. Eng.*, vol. 86, no. 7–9, pp. 1883–1887, Jul. 2009.

[150] M. H. Chang and J. F. Zhang, "On positive charge formed under negative bias temperature stress," *J. Appl. Phys.*, vol. 101, no. 2, pp. 1–7, 2007.

[151] Y. Z. Y. Zhao, M. X. M. Xu, and C. T. C. Tan, "A new observation of hot-carrier induced interface traps spatial distribution in 0.135 μm n-MOSFET by gate-diode method," *Proceedings. 7th Int. Conf. Solid-State Integr. Circuits Technol. 2004.*, vol. 2, 2004.

[152] T. Grasser, H. Reisinger, P. J. Wagner, and B. Kaczer, "Time-dependent defect spectroscopy for characterization of border traps in metal-oxide-semiconductor transistors," *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2010.

[153] L. Gerrer *et al.*, "Modelling RTN and BTI in nanoscale MOSFETs from device to circuit: A review," *Microelectron. Reliab.*, 2014.

[154] C. Liu, K. T. Lee, H. Lee, Y. Kim, S. Pae, and J. Park, "New Observations on the Random Telegraph Noise induced Vth Variation in Nano-scale MOSFETs Impacts of Worst Bias Condition , BTI Aging Effects and Random Switching Property," *IEEE Int. Reliab. Phys. Symp. Proc.*, pp. 4–8, 2014.

[155] M. Duan *et al.*, "Defect Loss: A New Concept for Reliability of MOSFETs," *IEEE Electron Device Lett.*, vol. 33, no. 4, pp. 480–482, Apr. 2012.

[156] R. Wang, S. Guo, Z. Zhang, J. Zou, D. Mao, and R. Huang, "Complex random telegraph noise (rtn): What do we understand?," in *Proceedings of the International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA*, 2018, vol. 2018-July.

[157] C. Prasad *et al.*, "Bias temperature instability variation on SiON/Poly, HK/MG and trigate architectures," in *IEEE International Reliability Physics Symposium Proceedings*, 2014.

[158] M. P. Li, "Jitter challenges and reduction techniques at 10 Gb/s and beyond," *IEEE Trans. Adv. Packag.*, 2009.

[159] P. Ren *et al.*, "New Observations on Complex RTN in Scaled High- κ / Metal-gate MOSFETs – the Role of Defect Coupling under DC / AC Condition," pp. 778–781, 2013.

[160] 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," *2010 Int. Electron Devices Meet.*, pp. 628–631, Dec. 2010.

[161] C. Y. Chen *et al.*, "Correlation of Id- and Ig-random telegraph noise to positive bias temperature instability in scaled high-κ/metal gate n-type MOSFETs," in *IEEE*

International Reliability Physics Symposium Proceedings, 2011.

[162] a. Teramoto, T. Fujisawa, K. Abe, S. Sugawa, and T. Ohmi, "Statistical evaluation for trap energy level of RTS characteristics," *Dig. Tech. Pap. - Symp. VLSI Technol.*, vol. 38, no. c, pp. 99–100, 2010.

[163] J. W. Lu *et al.*, "Device-level PBTI-induced timing jitter increase in circuit-speed random logic operation," in *Digest of Technical Papers - Symposium on VLSI Technology*, 2014.

[164] J. Zou *et al.*, "New Insights into AC RTN in Scaled High- κ / Metal-gate MOSFETs under Digital Circuit Operations," *Symp. VLSI Technol. (2012*, pp. 139–140, 2012.

[165] H. Miki *et al.*, "Voltage and temperature dependence of random telegraph noise in highly scaled HKMG ETSOI nFETs and its impact on logic delay uncertainty," in *Digest of Technical Papers - Symposium on VLSI Technology*, 2012.

[166] M. Yamaoka *et al.*, "Evaluation methodology for random telegraph noise effects in SRAM arrays," in *2011 International Electron Devices Meeting*, 2011, pp. 32.2.1-32.2.4.

[167] K. Ota, M. Saitoh, C. Tanaka, D. Matsushita, and T. Numata, "Systematic Study of RTN in Nanowire Transistor and Enhanced RTN by Hot Carrier Injection and Negative Bias Temperature Instability," *Dig. Tech. Pap. - Symp. VLSI Technol.*, vol. 273, no. 1991, pp. 2013–2014, 2014.

[168] B. Kaczer *et al.*, "The relevance of deeply-scaled FET threshold voltage shifts for operation lifetimes," in *2012 IEEE International Reliability Physics Symposium (IRPS)*, 2012, pp. 5A.2.1-5A.2.6.

[169] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS amplitudes in decananometer MOSFETs: 3-D simulation study," *IEEE Trans. Electron Devices*, 2003.

[170] M. F. Bukhori, S. Roy, and A. Asenov, "Simulation of statistical aspects of charge trapping and related degradation in bulk mosfets in the presence of random discrete dopants," *IEEE Trans. Electron Devices*, 2010.

[171] J. F. Zhang, M. H. Chang, and G. Groeseneken, "Effects of measurement temperature on NBTI," *IEEE Electron Device Lett.*, vol. 28, no. 4, pp. 2007–2009, 2007.

[172] W. J. Liu *et al.*, "On-the-fly interface trap measurement and its impact on the understanding of NBTI mechanism for p-MOSFETs with SiON gate dielectric," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2007.

[173] Z. Ji, J. F. Zhang, M. H. Chang, B. Kaczer, and G. Groeseneken, "An analysis of the NBTI-induced threshold voltage shift evaluated by different techniques," *IEEE Trans. Electron Devices*, 2009.

[174] J. F. Zhang *et al.*, "Dominant layer for stress-induced positive charges in Hf-Based gate stacks," *IEEE Electron Device Lett.*, 2008.

[175] R. Gao *et al.*, "Predictive As-grown-Generation (A-G) model for BTI-induced device/circuit level variations in nanoscale technology nodes," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2017.

[176] M. Duan *et al.*, "Insight into Electron Traps and Their Energy Distribution under Positive Bias Temperature Stress and Hot Carrier Aging," *IEEE Trans. Electron Devices*, vol. 63, no. 9, pp. 3642–3648, 2016.

[177] M. Duan, J. F. Zhang, Z. Ji, W. D. Zhang, B. Kaczer, and A. Asenov, "Key Issues and Solutions for Characterizing Hot Carrier Aging of Nanometer Scale nMOSFETs," *IEEE Trans. Electron Devices*, vol. 64, no. 6, pp. 2478–2484, 2017.

[178] J. F. Zhang, Z. Ji, and W. Zhang, "As-grown-generation (AG) model of NBTI: A shift from fitting test data to prediction," *Microelectron. Reliab.*, 2018.

[179] Z. Çelik-Butler, S. P. Devireddy, H. H. Tseng, P. Tobin, and A. Zlotnicka, "A low-frequency noise model for advanced gate-stack MOSFETs," *Microelectron. Reliab.*, 2009.

[180] W. Zhu, J. P. Han, and T. P. Ma, "Mobility measurement and degradation mechanisms of MOSFETs made with ultrathin high-k dielectrics," *IEEE Trans. Electron Devices*, 2004.

[181] M. Toledano-Luque *et al.*, “From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation,” *2011 Symp. VLSI Technol. - Dig. Tech. Pap.*, pp. 152–153, 2011.

List of Publications

Journals

1. **A. B. Manut**, R. Gao, J. F. Zhang, Z. Ji , M. Mehedi, W. D. Zhang, D. Vigar, A. Asenov, and B. Kaczer, "Trigger-When-Charged: A Technique for Directly Measuring RTN and BTI-Induced Threshold Voltage Fluctuation Under Use-V_{dd}" in *IEEE Transactions on Electron Devices*, vol. 66, no. 3, pp. 1482-1488, March 2019.

2. **A. B. Manut**, J. F. Zhang, M. Duan, Z. Ji, W. Zhang, B. Kaczer, T. Schram, N. Horiguchi, and G. Groeseneken, "Impact of Hot Carrier Aging on Random Telegraph Noise and Within a Device Fluctuation," in *IEEE Journal of the Electron Devices Society*, vol. 4, no. 1, pp. 15-21, Jan. 2016.

Conferences

1. M. Duan, J. F. Zhang, **A. Manut**, Z. Ji ; W. Zhang, A. Asenov, L. Gerrer, D. Reid, H. Razaidi D. Vigar, V. Chandra, R. Aitken, B. Kaczer, and G. Groeseneken, "Hot carrier aging and its variation under use-bias: Kinetics, prediction, impact on V_{dd} and SRAM," *2015 IEEE International Electron Devices Meeting (IEDM)*, Washington. 2015.

2. **A.B. Manut**, J. F. Zhang, Z. Ji, W. Zhang, "Interaction Between Random Telegraph Noise And Hot Carrier Ageing," *IEEE China Semiconductor Technology International Conference (CSTIC), ShangHai*, March 18-19, 2019.

3. J. F. Zhang, **A. Manut**, R. Gao, M. Mehedi, Z. Ji, W. Zhang, and J. Marsland, "An assessment of RTN-induced threshold voltage jitter," *Proceedings of 2017 13th IEEE International Conference on ASIC (ASICON)*, Oct. 2019.