Bias Temperature Instability Modelling and Lifetime Prediction on Nano-scale MOSFETs

Rui Gao

A thesis submitted in partial fulfilment of the

requirements of Liverpool John Moores University for

the degree of Doctor of Philosophy

July 2018

Acknowledgements

First I would like to express my greatest gratitude to my first supervisor Dr. Zhigang Ji for his continuous support, guidance and encouragement throughout my entire PhD research project.

I would also thank my second supervisor Prof. Jianfu Zhang and third supervisor Prof. Weidong Zhang for the fruitful discussions which made the project much more efficient. Thanks to Dr. Benbakhti Brahim for reviewing my transfer report.

I want to thank all the members working together (used to work) in LJMU microelectronics reliability and characterization group. They are: Dr. Meng Duan, Dr. Jigang Ma, Dr. Zheng Chai, Mr. Xiong Zhang, Mr. Azrif Manut, Dr. Pengpeng Ren, Mr. Steven Duffy, Mr. James Brown, Mr. Dale Hodgkinson and Ms. Mehzabeen Mehedi. Dr. Meng Duan's pioneering work on variability and Dr. Jigang Ma's outstanding work on *Energy Alternating Defects* laid a solid foundation for the modelling work in this project. Dr. Jigang Ma also helped me a lot to improve my electric characterization skills in my first year.

Thanks to Mrs. Helen Pottle, Mrs. Caroline O'Rourke for helping me out when my scholarship was delayed.

Last, but not least, I would like to express my deepest love to my family.

Abstract

Bias Temperature Instability (BTI) is one of the most important reliability concerns for Metal Oxide Semiconductor Field Effect Transistors (MOSFET), the basic unit in integrated circuits. As the development MOSFET manufacturing technology, circuit designers need to consider device reliability during design optimization. An accurate BTI lifetime prediction methodology becomes a prerequisite.

Typical BTI lifetime standard is ten years, accelerated BTI tests under high stress voltages are mandatory. BTI modelling is needed to project BTI lifetime from high voltages (accelerated condition) to operating voltage. The existing two mainstream BTI models: 1). The Reaction-Diffusion (R-D) framework and 2). The Two-Stage model cannot provide accurate lifetime prediction. Quite a few fitting parameters and unjustifiable empirical equations are needed in the R-D framework to predict the lifetime, questioning its predicting capability. The Two-stage model cannot project device lifetime from high voltages to operating voltage.

Moreover, the scaling down of MOSFET feature size brings new challenges to nano-scale device lifetime prediction: 1). Nano-scale devices' current is fluctuating due to the impact of a single charge is increasing as MOSFET scaling down, repetitive tests need to be done to achieve meaningful averaged results; 2). Nano-scale devices have significant Device-to-Device variability, making the lifetime a distribution instead of a single value.

In this work a comprehensive As-grown Generation (A-G) framework based on the A-G model and defect centric theory is proposed and successfully predicts the Time Dependent Variability and lifetime on nano-scale devices. The predicting capability is validated by the good agreement between the test data and predicted values. It is speculated that the good predicting capability is due to the correct understanding of different types of defects.

In the A-G framework, Time Dependent Variability is experimentally separated into Within-Device Fluctuation and the averaged degradation. Within-Device Fluctuation can be directly measured and the averaged degradation can be modelled using the A-G model. The averaged degradation in the A-G model contains: Generated Defects, As-grown Traps and Energy Alternating Defects. These defects have different kinetics against stress time thus need separate modelling. Various patterns such as Stress-Discharge-Recharge, multi-Discharging-based Multiple Pulses are designed to experimentally separate these defects based on their different charging/discharging properties. Fast-Voltage Step Stress technique is developed to reduce the testing time by 90% for the A-G framework parameter extraction, making the framework practical for potential use in industry.

List of Abbreviations

Abbreviations	Signification
A(H/E)T	As-grown (Hole/Electron) Traps
A-G model	As-grown Generation model
AN(PC/ET)	Anti-Neutralization (Positive Charges/Electron Traps)
CET	Capture Emission Time
CP	Charge Pumping
CPC/CET	Cyclic Positive Charges/Cyclic Electron Traps
CV	Capacitance-Voltage
EAD	Energy Alternating Defects
EoS/EoR	End of Stress/End of Recovery
EoT	Effective Oxide Thickness
Freq	Frequency
IV	Drain current (I_d)~Gate voltage (V_g)
Meas	Measurement
MOS(FET)	Metal Oxide Semiconductor (Field Effect Transistor)
MSM	Measure-Stress-Measure
N/PBTI	Positive/Negative Bias Temperature Instability
Pts	Points
R-D model	Reaction-Diffusion model
RTN	Random Telegraph Noise
SDR	Stress-Discharge-Recharge
SMU	Source Measure Unit
SR	Sampling Rate
T0V/TDV	Time-zero Variability/Time Dependent Variability
TDDS	Time Dependent Defect Spectroscopy
UE/LE	Upper Envelope/Lower Envelope
VRS	Voltage Ramp Stress
(F-)VSS	(Fast-) Voltage Step Stress
WDF	Within Device Fluctuation

List of Symbols

Symbol	Description	Unit
C_{ox}	oxide capacitance	F
E_a'	activation energy	eV
E_c	bottom of the conduction band	eV
E_f/E_F	Fermi level	eV
E_v	top of the valence band	eV
g_m	transconductance of an I_d-V_g curve	S
I_d	drain current	A
I_{doff}	"off" state drain current	A
I_{dcc}	constant I_d value for V_{th} extraction	A
I_g	gate leakage current	A
N_c	effective density of states in conduction band	cm^{-3}
N_{IT}/N_{it}	interface states density	cm^{-3}
N_v	effective density of states in valence band	cm^{-3}
T	Temperature	
T_w	time window	s
V_{dd}	drain supply voltage /Circuit operating voltage	V
V_g	gate voltage	V
V_{gch}	charging gate voltage	V
V_{gdisch}	discharging gate voltage	V
V_{grech}	recharging gate voltage	V
$V_{gst}/V_{gstress}$	stress gate voltage	V
V_{th}	threshold voltage	V
W/L	MOSFET channel width/length	m
ΔV_{th}	threshold voltage shift	V
σ_i	capture cross section	cm^{-2}
φ_s	surface potential	eV

List of Figures

List of Tables

Table of Contents

1 Introduction of Bias Temperature Instability (BTI) Lifetime Prediction

1.1 A review of BTI lifetime prediction

Nowadays electronic products are widely used in daily life from all aspects. Reliability is one of the most important parameters for electronic products. Product failures due to circuit malfunction can result in excessive warranty costs and severe brand damage to any integrated circuit company. In 2011, for example, Intel was about to launch their Sandy Bridge processor when a potential reliability problem was detected. The problem, which was not spotted during extensive functional testing, was a gradual performance reduction and even total failure of the serial-ATA channels in about 5% of the manufactured integrated circuits [1]. Unfortunately, this is not rare. A similar event also happened to Cisco, costing $655 million to fix [2].

As semiconductor manufacturing technology node migrates, the complexity of circuits nowadays requires the reliability to be evaluated while design optimization before the fabrication. One example is the circuit operating voltage V_{dd} selection. Higher V_{dd} can enhance MOSFET drive current and increase the switching speed to achieve a higher performance, but the degradation will also be higher. To make a trade-off between performance and reliability, a lifetime prediction model becomes a prerequisite.

1.1.1 Introduction of BTI

Bias Temperature Instability (BTI) is a key reliability issue in MOSFET. It describes a phenomenon that degrades the performance of a device when a bias is applied on the gate of the MOSFET and turns on that channel. BTI on one hand, will cause device threshold voltage $|V_{th}|$ to increase, resulting in a reduction of device driving current $|I_d|$ (Fig. 1.1a) and circuit operating frequency; on the other hand, BTI will enhance the absolute "off" current $|I_{doff}|$ (Fig. 1.1b) and Gate Leakage $|I_g|$ (Fig. 1.1c), resulting in the increase of circuit power consumption. BTI is normally characterized by threshold voltage shift ($|\Delta V_{th}|$) against stress time. Elevated temperature will accelerate BTI. 125°C is the temperature people normally used to do BTI tests, which is also used hereafter if not specified.

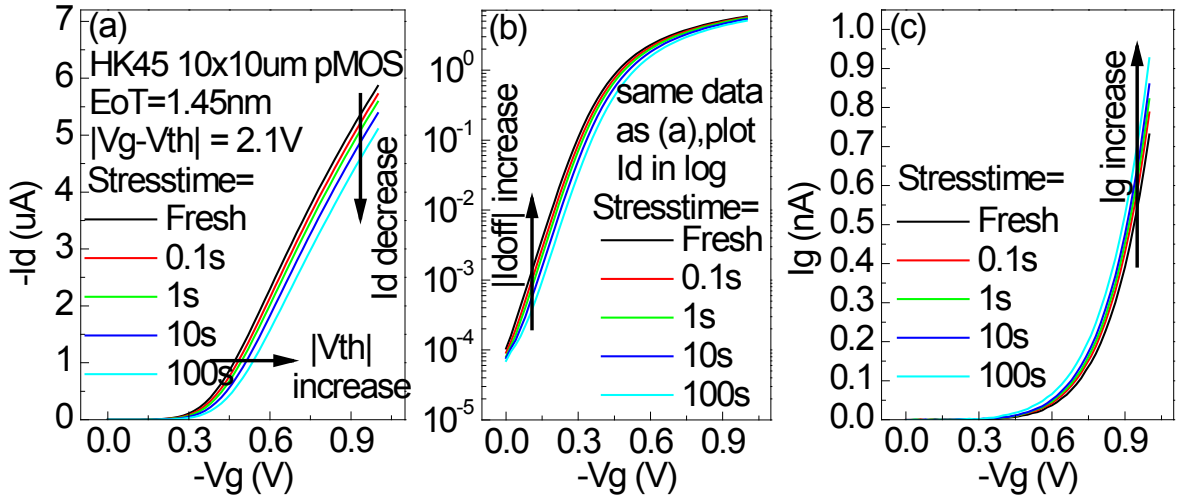

Fig. 1.1 MOSFET I_d-V_g (IV) curves show both $|V_{th}|$ (a) and $|I_{d\_off}|$ increases under NBTI stress. (c) I_g-V_g measurements show gate leakage (I_g) also increases.

Most integrated circuits suffer from BTI degradation. A typical example is a Complementary Metal Oxide Semiconductor (CMOS) inverter, which is the basic logic cell in integrated circuits. Fig. 1.2 shows pMOS/nMOS suffers from Negative/Positive Bias Temperature Instability when the CMOS inverter is operating.

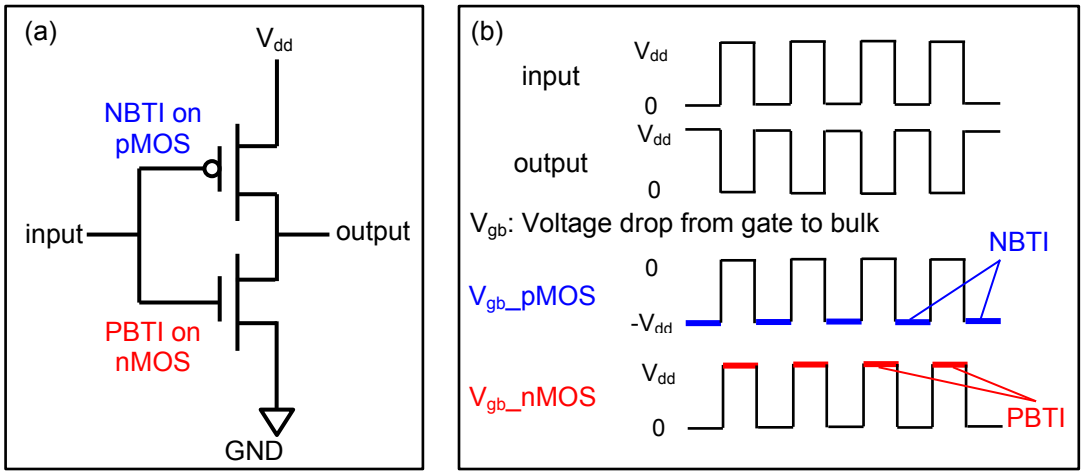

Fig. 1.2 (a) Illustration of the operation condition of a CMOS inverter in circuits. (b) With input 0/1 (GND/V_{dd}), output is 1/0 (V_{dd}/GND) and the pMOS/nMOS device (top/bottom) is under uniform negative/positive gate bias towards its bulk in the marked phases.

The complexity of circuit fabrication increases geometrically with MOSFET feature size scaling down and circuit integration goes up. Circuit designers need to consider the circuit lifetime as soon as they start to design the circuits before fabrication. One example is use-bias V_{dd} selection. Higher V_{dd} means higher circuit performance, but meanwhile the degradation will also be higher and lifetime is sacrificed. To make a proper trade-off between performance and reliability, lifetime prediction is needed. As one of the most important reliability mechanism, BTI lifetime prediction becomes a necessity for industry.

1.1.2 Origin of BTI

BTI is caused by the creation or filling of traps in gate dielectric. The extra charges in gate dielectric will impose an inverted electric field in terms of gate voltage, resulting $|V_{th}|$ to increase. Created traps will also assist carrier hopping thus $|I_g|$ also increases.

Traps and charges in gate dielectric are usually classified into four categories (Fig. 1.3), as suggested by B.E. Deal [3] in 1980. They are: 1). Interface states N_{IT} and trapped charges Q_{IT}, which are located at the Si-SiO_2 interface with energy states within the silicon forbidden bandgap and can exchange charges with silicon in a short time; Interface traps can possibly be produced by excess silicon (trivalent silicon), broken Si-H bonds, excess oxygen and impurities; 2). Fixed oxide charges Q_f, which are located at or near the

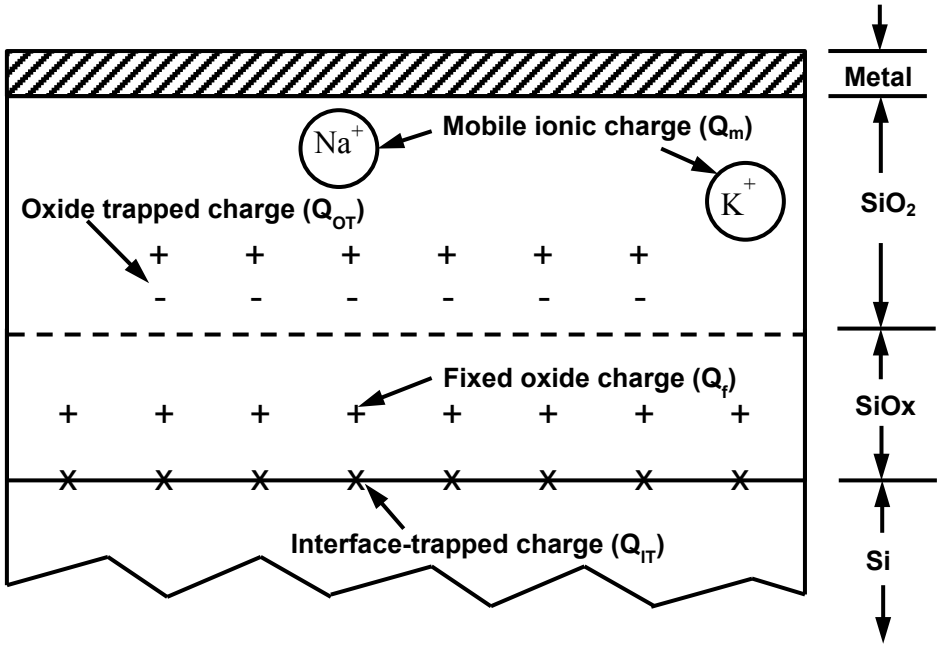

Fig. 1.3 Names and locations of charges associated with thermal oxidized silicon [3].

interface and are immobile under an applied electric field; 3). Oxide trapped charges Q_{OT}, which can be created, for example, by X-ray radiation or hot-electron injection; these traps are distributed inside the oxide layer; 4). Mobile ionic charges Q_m, such as sodium ions, which are mobile within the oxide under BTI stress condition.

When BTI was first observed in 1967 [4], it was mainly caused by ionic contamination [5, 6] and electrically active surface states passivation at Si-SiO_2 interface [7, 8] from the immature fabrication technology. With the development of semiconductor manufacturing [9-11], ionic contamination and pre-existing interface states can be well controlled, now it's mainly the creation of interface states and oxide traps that are responsible for BTI degradation.

With the scaling down of MOSFET feature size, the status of BTI compared to other reliability mechanisms has been changing at different stages. In the 1970s, Hot Carrier Aging (HCA) was the most important as circuits are driven by a high V_{dd}; in 1980s with V_{dd} scaling down with feature size HCA was much alleviated, while the thin gate oxide makes Time Dependent Dielectric Breakdown (TDDB) start to be a major problem; in 2000s High-κ material is widely used to suppress TDDB. High-κ material can have the same amount of capacitance with a much larger thickness, TDDB is no longer a major problem then. But High- κ material will enhance BTI degradation [12, 13], and BTI starts to become one of the most important sources of reliability concern [14, 15].

1.1.3 Accelerate characterization of BTI

The standard criteria of BTI lifetime is that $|\Delta V_{th}|$ will not exceed a certain level (typically 100mV) after device reaches its lifetime. Typical device lifetime is 10 years, which is too long to reach in a lab. Accelerated BTI test by applying a much severer stress (higher V_g) than operating condition is mandatory. The device lifetime under operating voltage is then projected from the accelerated tests within acceptable testing time ($<10^6$ seconds) using a time-evolution model. Predicting capability of the model can be tested by comparing the predicted value with test data under operating condition, as shown in Fig. 1.4.

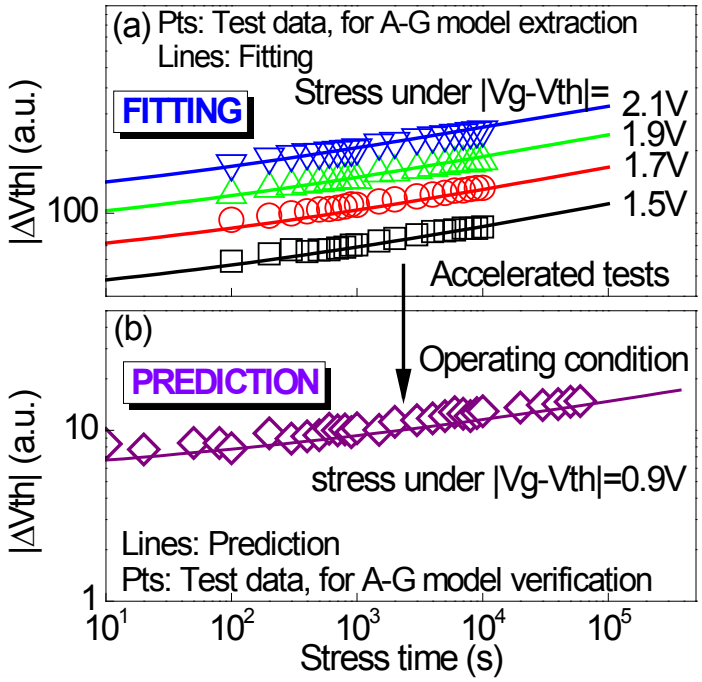

Fig. 1.4 An example of BTI lifetime projection from accelerated tests to operating condition. (a) Test data under high $|V_g-V_{th}|$ (accelerated stress) condition is used to extract model parameters and predict device lifetime. (b) The accuracy of the prediction is usually verified by the comparison of test data under use-bias and model prediction. Note the test data in (b) should not be used to extract model parameters.

Typical BTI kinetics follow a power law against both stress time and overdrive voltage $|V_g\text{-}V_{th}|$, and can be described in Equation (1.1), suggested by the Joint Electron Device Engineering Council (JEDEC) [16].

$$\Delta V_{th} = A \cdot (|V_g - V_{th}|)^m \cdot t_{str}^n \quad (1.1)$$

Where t_{str} is BTI stress time, A, m, and n are fitting parameter from accelerated tests.

Equation (1.1) works well for slow DC BTI characterization (detailed in section 2.3.1), as shown in Fig. 1.5.

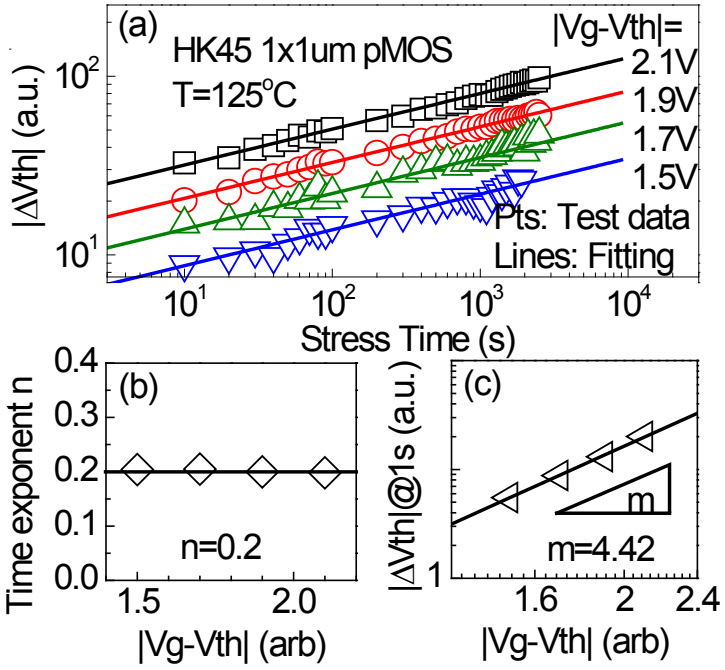

Fig. 1.5 BTI kinetics measured by the slow DC characterization can be well described by Equation (1.1). NBTI is adopted as an example here. $|\Delta V_{th}|$ kinetics under different $|V_g\text{-}V_{th}|$ stress (a) can be well fitted by power law with the same time exponent n (b), $|V_g\text{-}V_{th}|$ power exponent m can then be extracted from fitted $|\Delta V_{th}|$@1s against $|V_g\text{-}V_{th}|$ (c).

However, after the fast characterization method was introduced, people realize slow characterization can only capture a small portion of BTI due to fast recovery. The time

exponent n, is sensitive to the measurement condition, resulting in that Equation (1.1) is no longer applicable to predict BTI lifetime.

1.1.4 BTI Recovery

BTI degradation contains a lot of fast recoverable traps, which will recover once the stress voltage is removed. Due to the fast recovery, the measurement result is highly dependent on BTI characterization speed. In early BTI works, due to the limitation of measurement facilities, slow DC measurements were adopted to measure BTI degradation, which usually took hundreds milliseconds to several seconds to measure one IV curve and extract the $|\Delta V_{th}|$. During the measurement most of the fast recoverable traps have already escaped, resulting in an underestimation of the total BTI degradation, as shown in Fig. 1.6.

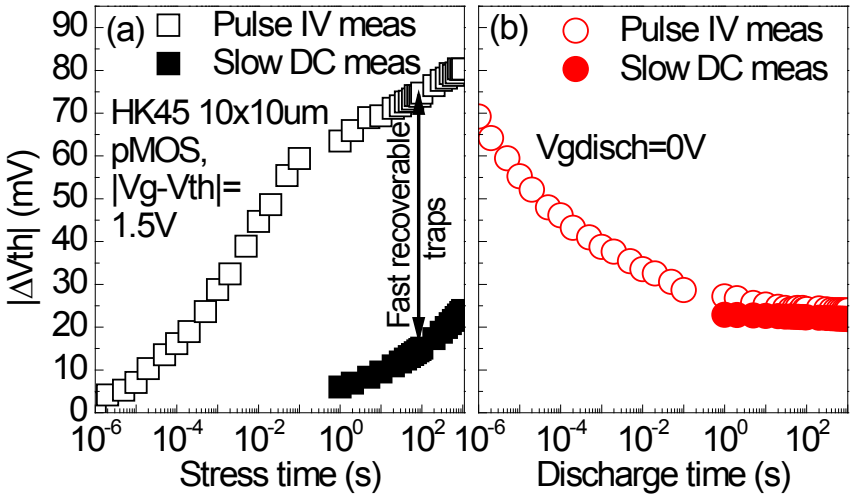

Fig. 1.6 NBTI stress (a) and recovery (b) kinetics measured by fast pulse IV measurement and slow DC measurement. For each data point, the hollow one is measured in 3 microseconds while the solid one is measured in 100 milliseconds. Slow DC measurement only captures the slow traps induced $|\Delta V_{th}|$ thus underestimates the total $|\Delta V_{th}|$. Refer to section 2.3 for details.

With the development of measurement facilities, nowadays people can measure BTI kinetics within several microseconds. Fig. 1.6 shows the comparison between the fast pulse measurement results and slow DC measurement results under the same $|V_g-V_{th}|$ stress. More than 75% of the total $|\Delta V_{th}|$ has been underestimated in the stress phase (Fig. 1.6a) with slow DC measurements. Unless specified, all the BTI degradation is measured by fast pulse measurement hereafter.

1.2 A review of Bias Temperature Instability modelling

BTI modelling is indispensable to project the lifetime from accelerated stress condition to operating condition. Various models [17-21] have been proposed, among them the most important two models are the Reaction-Diffusion model and the Two-Stage model.

1.2.1 Reaction-Diffusion (R-D) model

The origin of the Reaction-Diffusion (R-D) model can be traced back to 1977. K. O. Jeppson et al [11] reported interface states in the oxide were an important source of the BTI degradation. A hydrogen-diffusion controlled mechanism was proposed to explain the creation and annihilation of interface states, corresponding to the degradation and recovery observed in the experiments, as shown in Fig. 1.7

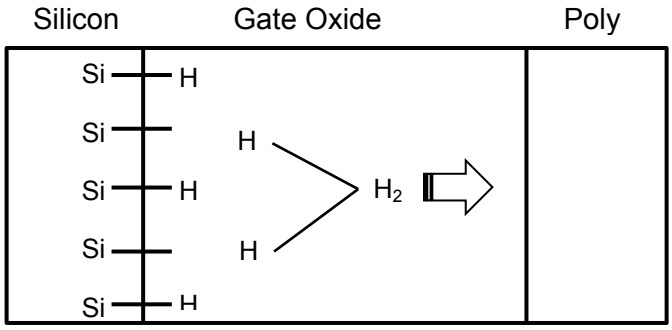

Fig. 1.7 Schematic illustration of the R-D model to interpret interface trap generation.

Along with this thinking, S. Mahapatra [20] and M. Alam [22] proposed the R-D model to interpret the broad experimental dataset [11, 23, 24] from previous people's work and to provide a physics-based analytical model to predict BTI lifetime.

The R-D model assumes that when a bias is applied on the gate, a field-dependent reaction will be initialized at the Si/SiO_2 interface, where the passivated Si-H bonds are broken and interface traps are generated (Fig. 1.7) [25]. High temperature will weaken the existing Si-H covalent bond hence will also increase BTI. Meanwhile the generated H will also react with the Si- bond and neutralize it, so it is a competing process.

At stress phase,

$$\mathrm{Si-H} \leftrightarrow \mathrm{Si}\cdot + \mathrm{H} \tag{1.2}$$

The generated hydrogen will accumulate near the interface and inevitably diffuse away from the interface, leaving behind a positively charged interface state (N_{it}). The R-D model is described by a group of differential equations (1.3)-(1.6). The meaning of these parameters is given in Table 1.1.

$$dN_{it}/dt = k_F(N_0 - N_{it}) - k_R N_H N_{it} \quad (x = 0) \quad (1.3)$$

$$dN_{it}/dt = D_H(dN_H/dx) + (\delta/2)dN_H/dt \quad (x < \delta) \quad (1.4)$$

$$D_H(d^2N_H/dx^2) = dN_H/dt \quad (\delta < x < T_{PHY}) \quad (1.5)$$

$$D_H(dN_H/dx) = K_P N_H \quad (x > T_{PHY}) \quad (1.6)$$

Table 1.1 Explanation of parameters used in the R-D Model

x	Distance towards the gate, 0 means at the interface
N_{it}	Numbers of interface states at any given x
N_0	Initial number of unbroken Si-H bonds
N_H	Hydrogen concentration at any given x
k_F	Oxide field dependent forward dissociation rate
k_R	Recombination rate
D_H	Hydrogen diffusion coefficient
δ	Interface thickness
T_{PHY}	Oxide thickness
k_P	Surface recombination velocity at oxide/poly interface

By solving the Equations (1.3)-(1.6) the solutions suggest there are five different regimes of time-dependent interface trap generation at the stress phases, as shown in Fig. 1.8.

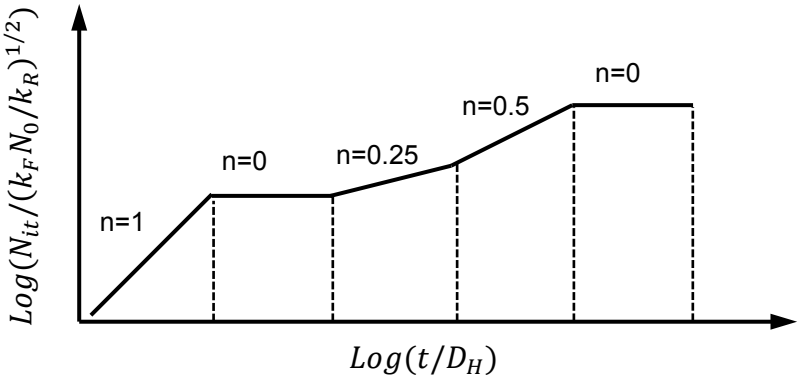

Fig. 1.8 Five phases obtained from the general solution of the R-D equations during NBTI stress

At the very beginning, both N_{it} and N_H are negligible compared to N_0, so it is reaction dominated, thus by Equation (1.3),

$$N_{it} = k_F N_0 t \tag{1.7}$$

As time passes, more N_{it} and N_H are generated, the reaction rate $k_F(N_0-N_{it})$ decreases while the recombination rate $k_R N_H N_{it}$ increases, a dynamic equilibrium is reached and Equation (1.3) becomes

$$dN_{it}/dt = k_F(N_0 - N_{it}) - k_R N_H N_{it} = 0 \tag{1.8}$$

On the other hand, the generated N_H does not have time to diffuse from the interface yet, which suggests

$$N_H = N_{it} \tag{1.9}$$

Substitute Equation (1.9) into Equation (1.8), the solution of N_{it} becomes

$$N_{it} = \left(\frac{k_F N_0}{k_R}\right)^{0.5} \tag{1.10}$$

In state-of-the-art modern MOSFETs, all these phases will take place in sub-microseconds, thus are not observed in most measurements [25].

As stress time lapses, hydrogen diffusion starts to take over the control of the interface trap generation, under the approximation that the very slow trap generation is negligible compared to the diffusion [25], the solution to Equations (1.3)-(1.6) will be

$$N_{it} \sim (k_F N_0/k_R)^{1/2}(D_H t)^{1/4} \tag{1.11}$$

For most advanced devices, before the end of this regime the diffusion length $x = (D_H * t)^{1/4}$ is already in the poly region, so this n = 1/4 regime only needs to be considered for very thick oxides.

Once the hydrogen arrives at the oxide/poly-interface, part of the hydrogen will reflect back and the other part keeps diffusing in the poly layer, which has a different diffusion coefficient from the oxide. Under the approximation that the trap generation rate is slow compared to the generation and annealing fluxes in Equation (1.5) (i.e., $dN_{it}/dt \sim 0$), the solution to Equations (1.3)-(1.6) will be [26]:

$$N_{it} = A \cdot (k_F N_0 / k_R)^{1/2} (D_H t)^{1/2} \qquad (1.12)$$

The final case n=0 means all the Si–H bonds have been broken, $N_{it} \sim N_0$ = Constant, which is hard to observe during measurements, since before device reaches this regime other failure mechanisms like TDDB have caused the oxide breakdown.

Equations (1.3)-(1.6) can also be used to solve the recovery phase with assigning $k_F=0$. A typical simulation result of a stress-recovery NBTI is given in Fig. 1.9a.

The R-D model soon became widely accepted since it explains the two key features of NBTI: the power law dependence against stress time and the recovery phenomenon. However, different values of time exponent n were reported by different groups, which differ from any of the value in Fig. 1.8. [15, 27-29]. Great efforts have been made to reconcile the gap between the measured data and the theory. For example, it is found that by changing the type of hydrogen species, the time exponent can be adjusted:, 1). either the neutral H or the H_2 only is involved in diffusion and thus leads to the 0.17 and 0.25 respectively; 2). H^+ can be affected by both diffusion and drift or H_2 or maybe all of them are responsible for the degradation [30-32]. By combining multiple types of species, thus modification is capable of explaining any time exponent observed in the experiment.

The R-D model is assumed to be absolutely correct since the bulk of literature appeared to be unanimously supportive of this model [22, 34, 35]. A break of the R-D model correctness was brought by the fast recovery investigation using fast pulse measurement, which directly contradicts the prediction from any variants of R-D theory [36, 37]. A comparison of the test result of BTI recovery and the R-D model predicted value is shown in Fig. 1.9. The BTI recovery is too fast to be explained by the R-D theory.

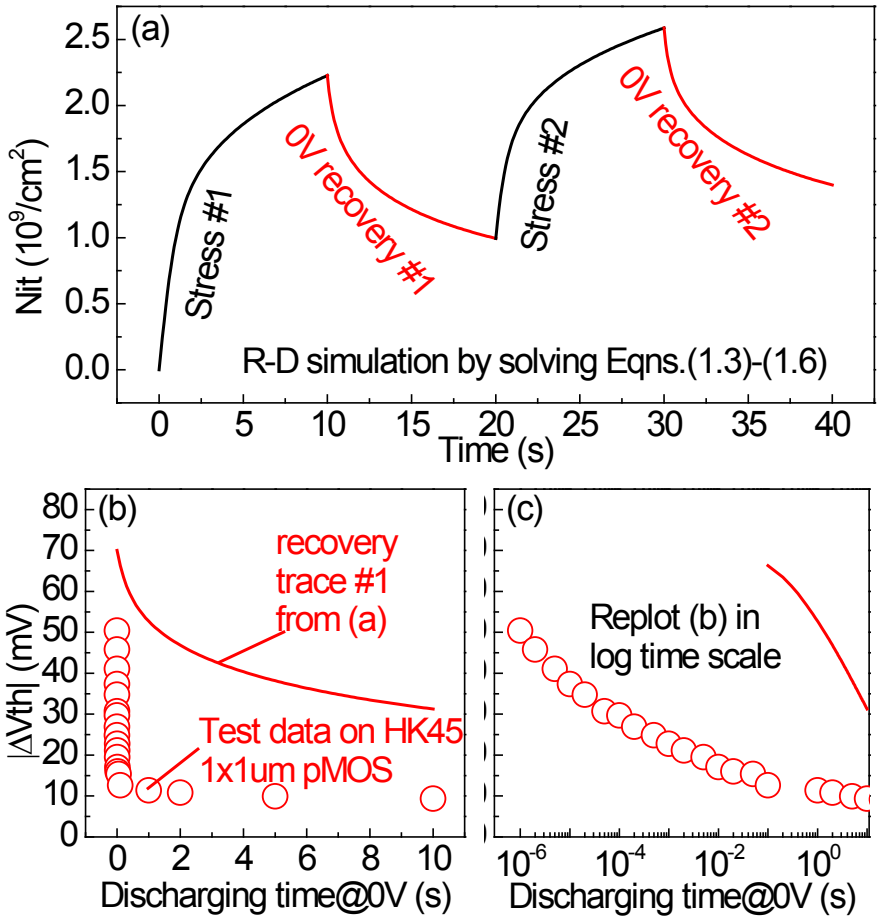

Fig. 1.9. (a) The R-D model based simulation of a BTI stress-recovery-stress procedure. (b) Test data on HK45 1x1um pMOS. (c) Plot the data from (b) in log scale. It's unlikely that almost half of the hydrogen diffused in 10 seconds will diffuse back in less than 1 millisecond.

To fill the gap between measurements and theory, the R-D model added hole traps (ΔN_{HT}) and oxide traps (ΔN_{OT}) to reconcile the fast BTI recovery. These three components are uncorrelated with each other. The name of the new model also changed to "the R-D framework" [21, 38].

The R-D framework extracts parameters by fitting the experiment data with [21]:

$$\Delta V_{th} = -\frac{q}{C_{ox}} \cdot (\Delta N_{it} + \Delta N_{HT} + \Delta N_{OT}) \quad (1.13)$$

$$\Delta N_{HT} = B \cdot \left|V_{gov}\right|^{\Gamma\_HT} \cdot e^{-\frac{E_A\_HT}{kT}} \cdot \left(1 - e^{-\left(\frac{t_{str}}{\tau_s}\right)^{\beta_{HTS}}}\right) \quad (1.14)$$

$\Gamma\_HT = \Gamma\_IT, E_A\_HT = 0.03eV, \tau_s = \tau_{0s} \cdot e^{-\left(\frac{E_A\_\tau_s}{kT}\right)}$

B, τ_{0s}, and β_{HTS} are device dependent parameters

$$\Delta N_{OT} = C \cdot \left(1 - e^{-\left(\frac{t_{str}}{n}\right)^{\beta_{OT}}}\right) \quad (1.15)$$

$\eta = 5e^{12}, \beta_{OT} = 0.36eV, \Gamma\_OT = 9, n = \eta \cdot \left|V_{gov}\right|^{\Gamma\_OT} \cdot e^{-\frac{E_A\_OT}{kT\beta_{OT}}}$

C is device dependent parameters

As there are so many device dependent parameters in the fitting, mathematically there will always be a good fitting, as shown in Fig. 1.10a&b. To validate the predicting capability of the R-D model, the comparison between the test data which is not used for the parameter extraction and the predicted value should be used. Fig. 1.10 shows the R-D framework cannot predict the lower stress voltage NBTI, questioning its predicting capability under operating voltage at device lifetime.

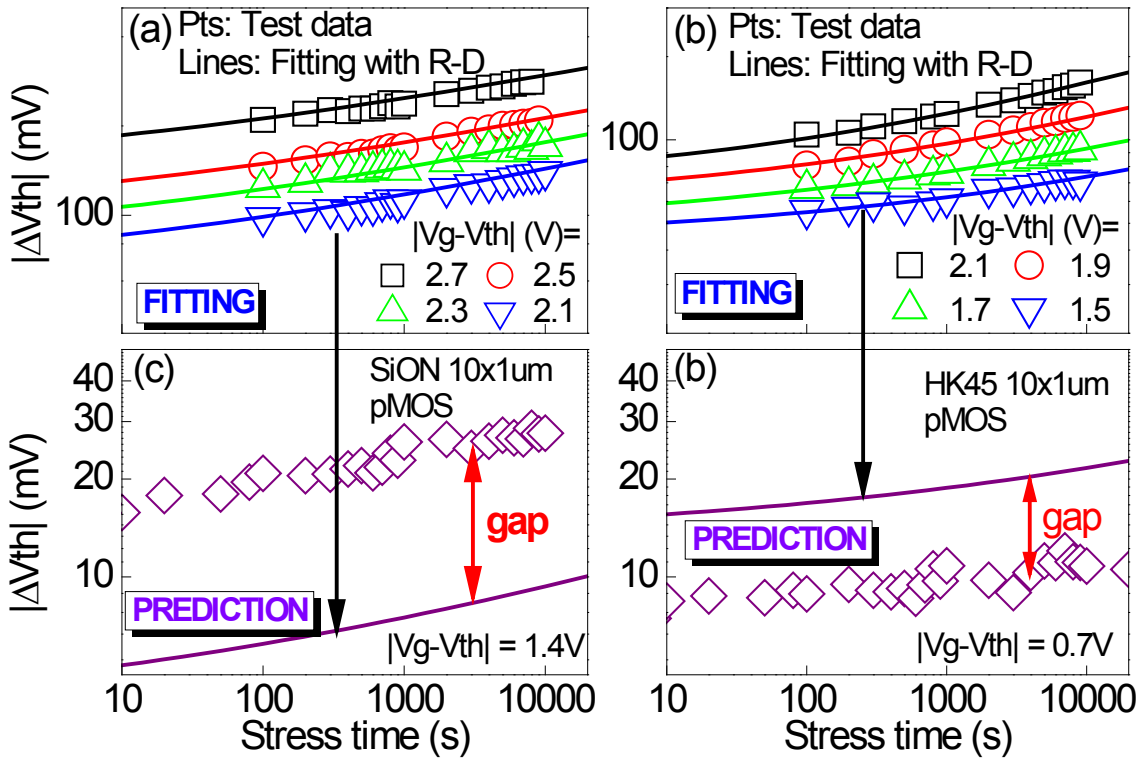

Fig. 1.10 The R-D framework prediction on 2 different processes of pMOSFETs. Parameters are extracted by fitting the experiment data and R-D framework prediction (upper), however they cannot predict the NBTI behavior under a much lower voltage (lower).

1.2.2 Two-Stage Model

The two-stage model was proposed by T. Grasser [32, 36] from an energy-level perspective to explain NBTI recovery. Most of the publications [36, 37, 39, 40] showed the clear existence of two components during NBTI stress: a fast recoverable component "R" and another "permanent" or slowly recovering component "P" [41, 42]. Experiments showed the degradation of the drain current obtained at different temperatures and voltages can be made to overlap using simple scaling, which suggests "R" and "P" in fact

are not in parallel but act in series and are therefore two facets of a single degradation mechanism proceeding in steps [19].

The two-stage model considers the Si-H systems as having two closely coupled energetic minima, the equilibrium and a slightly higher second minimum separated by an energy barrier [43, 44], which may be excited by thermal excitation quantum-mechanical tunnelling[44, 45]. The introduction of the electric field will lower the second minimum below the equilibrium minimum, making it the preferred energetic state. A time-dependent transition between the two wells is observed, depending on the transition mechanism and the statistical distribution of these barrier energies. Once the electric field is removed, the equilibrium well will be occupied again, which commonly involves long recovery transients, as shown in Fig. 1.11.

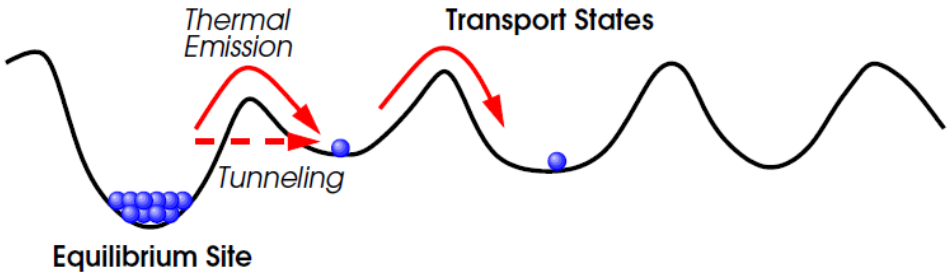

Fig. 1.11 The Si-H dissociation from energy-levels perspective. Hydrogen is released from a Si-H bond over a barrier into transport states (tetrahedral interstitial sites, bond center sites, etc.). Emission may occur via thermal emission or, at very low temperatures, via tunneling [46].

Fig. 1.12 shows the quantitative impact of the electric field on energy level in two wells based two-stage model, and the final degradation $|\Delta V_{th}|$ can be deducted from a series of quantum mechanics calculation [46]:

$$\Delta V_{th} = -\frac{q \cdot N}{C_{OX}} \cdot \left(\frac{1}{1+\exp\left(\frac{V_{3,m} - \left(V_{1,m} + 2\Delta\right)}{\sigma_3}\right)} - \frac{1}{1+\exp\left(\frac{V_{3,m} - V_{1,m}}{\sigma_3}\right)}\right) \quad (1.16)$$

Experimental data is then used to fit the parameters in Equation(1.16).

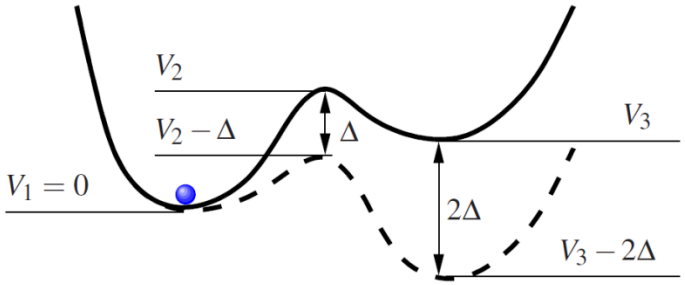

Fig. 1.12. Impact of the electric field on energy level in double well model. In equilibrium, the first well (V_1) is the energetically preferred configuration, application of the electric field tilts the wells favoring the second well (V_3) [46].

The Two-stage model explained data under lower temperatures well. But the high temperature calculation deviates from the experiment data and the permanent/slow component cannot be explained. By increasing the number of the wells, which means there are multi energy-level for hydrogen transition, both the stress and recovery traces can be well explained under different temperatures. But again, although physically possible, more wells will need more fitting parameters, which makes the model more complex and less convincing.

As a physics-based model from in-depth study of NBTI recovery, the two-stage model itself cannot be directly adopted to do a lifetime prediction. Based on the defect spectroscopy extracted from Time Dependent Defect Spectroscopy (TDDS) technique, CET mapping method is proposed to evaluate the device lifetime prediction. CET

mapping predicts $|\Delta V_{th}|$ follow the sum of two error-functions, which results in increasingly lower n with increasing stress time until complete saturation. However, currently there is no evidence which shows BTI will reach saturation before it reaches the lifetime. Some publications report the decreasing time exponent as stress time increases, but it's already been proved that that is due to the contamination of recoverable components or the deduction of the electric field under a constant gate voltage stress.

1.2.3 As-grown Generation (A-G) Model

The R-D framework, although taking both components into consideration by adding ΔN_{HT} and ΔN_{OT}, cannot project the measurement data beyond the voltages which are used for parameter extraction (Fig. 1.10). The Two-stage model cannot project device lifetime from high voltages to operating voltage.

The A-G model is proposed under such circumstances. The basic underlying idea of the A-G model is that there are different types of defects involved under BTI stress and these need to be understood and modelled separately. The idea itself is not brand new as most people now accept that BTI consists of a fast recoverable component "R" and a "permanent" component "P", they follow different time evolution kinetics thus need separate modelling. The main novelty of the A-G model is defects separation based on their charging/discharging behaviors. By applying the specified test pattern, different types of defects can be experimentally separated and modelled respectively, instead of using the ambiguous Recoverable Component "R" and Permanent Component "P".

For the simplest case: DC NBTI stress, the A-G model can be described by a simple equation

$$\Delta V_{th} = A + Gt_{str}^{n} \quad (1.17)$$

Where A is the saturation level of As-grown Traps and Gt_{str}^{n} is the defect generation.

Equation (1.17) is succinct but it can only deal with DC stress on big devices, as the understanding of different types of defects improves, the A-G model can be further extended to more complicated stress conditions, as detailed in chapter 5 & 6.

1.3 Challenges in predicting BTI lifetime on nano-scale devices

As semiconductor manufacturing technology migrates to nanometre scale, variability in nano-CMOS transistors start to become a challenge for further scaling down [47]. The challenge in predicting nano-scale devices BTI lifetime is two-fold: Device to Device Variability and Within Device Fluctuation, as detailed in the following two sections.

1.3.1 Device-to-Device Variability

As the feature size migrates to sub-65 nm, there are only a handful of defects in one device. Originating from various sources during the fabrication, including random discrete dopants, line edge roughness, polysilicon granularity and oxide thickness fluctuations [48], the fresh threshold voltages $|V_{th0}|$ at time zero illustrates a Gaussian distribution instead of a constant value. As shown in Fig. 1.13. Fig. 1.13 is a normal probability plot generated

by Matlab. The purpose of a normal probability plot is to graphically assess whether the data follows a normal distribution. If the data are normal the plot will be linear. Other distribution types will introduce curvature in the plot. Quite a few normal probability plots will be used in chapter 6. This $|V_{th0}|$ variability is independent of stress thus called Time-zero Variability.

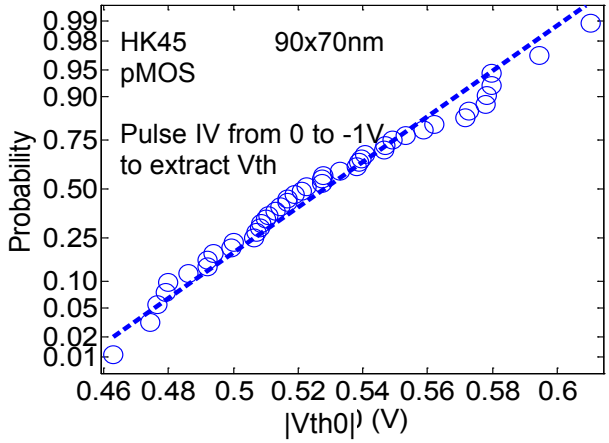

Fig. 1.13 Fresh threshold voltages $|V_{th0}|$ measured on 50 HK45 90x70nm pMOSFETs follow a normal distribution.

Time Dependent Variability, as a process of filling and generating new defects, also shows stochastic behaviour and varies with devices [49]. As a consequence, BTI kinetics and device lifetime becomes a stochastic distribution instead of a single value with MOSFET size scaling down, as shown in Fig. 1.14. **For the convenience of discussion hereafter, "big device" is used to represent MOSFET whose size is larger than (including) W*L=1um*1um, "nano-scale device" is used to represent MOSFET smaller than W*L=0.1um*0.1um (in this work 90nm*70nm MOSFET is used). Variability on "big device" is negligible but significant on "nano-scale devices".**

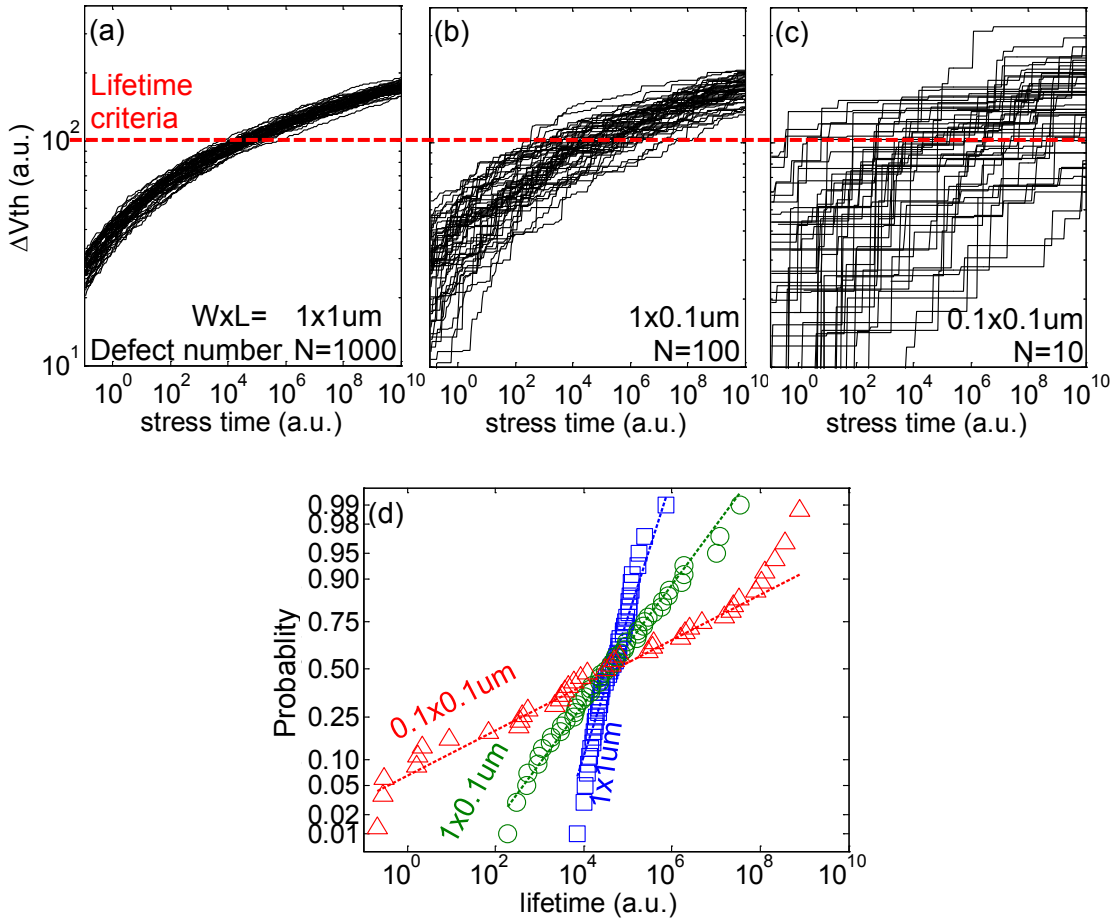

Fig. 1.14 Simulated BTI kinetics against stress time on different size of pMOSFETs (a-c) based on the defect centric theory in [50]. 50 devices are simulated for each size, a criteria of $|\Delta V_{th}|$=100 is adopted to extract the lifetime of all 50 devices and their distributions are plotted in (d). It is clearly shown that as device size scaling down, Device-to-Device Variability of BTI kinetics becomes larger, resulting in a wider spreading in lifetime for each device. This is because large device contains many defects and their random properties average out.

Industry normally uses "n·σ" standard to define the lifetime of nano-scale devices, which means after stressing under use bias for device lifetime, the Cumulative Distribution Function (CDF) at "n·σ" of a standard normal distribution will pass the lifetime criteria, for example "3σ" lifetime criteria means at circuit lifetime, at least 99.87% of devices in this circuit should pass. Typical "n·σ" corresponded CDF values are shown in Table 1.2.

Table 1.2 Cumulative Distribution Function (CDF) in a standard normal distribution

standard	σ	2σ	3σ	4σ	5σ
CDF	84.13%	97.72%	99.87%	99.9968%	99.999971335%

1.3.2 Within Device Fluctuation

With the scaling down of device geometry, the number of defects within the dielectric becomes smaller. Based on the simple charge sheet approximation, the impact from each single defect on $|\Delta V_{th}|$ is expected to increase, as shown in Table 1.3.

Table 1.3 Impact of a single trap as the scaling down of MOSFET size [49]

Device size (in um)	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 $V_g=V_{th}$-0.2	15,000		370		170	
Number of #Nit at a density DN_{it}=1e11/cm^2	1,000		24		11	
ΔD_{it} causing a $\|\Delta V_{th}\|$=50mV (in cm^{-2})	4.9e11		4.9e11		4.9e11	
Makes a number $\Delta\#N_{it}$	4,900		120		50	
$\|\Delta V_{th}\|$ caused by a single trapped carrier (at interface)	0.01mV		0.43mV		1.0mV	

In practice, both the defect numbers and their impact on $|\Delta V_{th}|$ are found to be stochastic, contributing to the large current fluctuation. The variation of the impact of one single trap is normally explained with the percolation path theory, which was firstly proposed by H.S. Wong et al [51] in 1993, they used atomic simulation to show channel current fluctuations can be caused by the percolation paths in a non-uniform profile. Many people use percolation paths theory to explain their experimental results [52-54].

As the impact of a single defect is increasing and the defect number in a single device is decreasing with device scaling down, I_d fluctuation caused by individual defect on nano-scale devices cannot be averaged out, resulting in a Within Device Fluctuation phenomenon. Repeated tests need to be done on nano-scale devices to achieve meaningful results, as detailed in chapter 6.

1.4 Defect centric theory

Defect centric theory was firstly proposed by B. Kaczer [55] in 2010 from the statistical analysis of vast numbers of defects. Their experimental data shows that the impact of a single defect on a nano-scale device follows an exponential distribution [52], and the average impact η, is inversely proportional to the device area ([52], Fig. 1.15). The number of traps on different nano-scale devices follows a Poisson distribution.

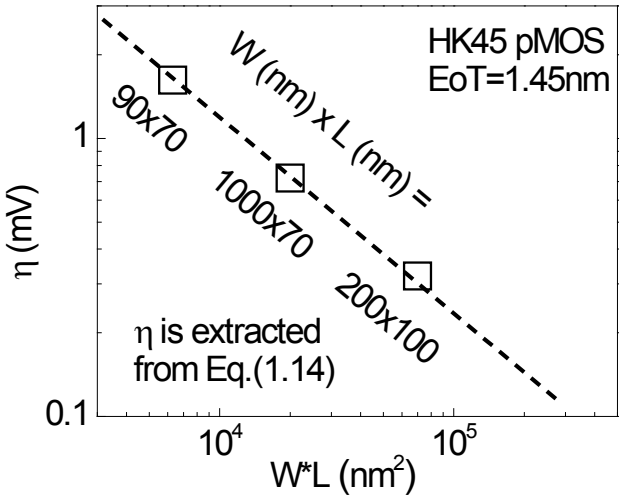

Fig. 1.15 Average impact of a single defect (η) values extracted on 3 different sizes of HK45 pMOSFETs show an inverse proportional relation against device area W*L.

η can be extracted from the standard deviation (σ) against the average (μ) of $|\Delta V_{th}|$ relationship in equation (1.18):

$$\sigma^2(|\Delta V_{th}|) = 2 \cdot \eta \cdot \mu\_\Delta V_{th} \quad (1.18)$$

Equation (1.18) is derived from the assumption that the number of the defects (ΔN) charged/generated under BTI stress follows a Poisson distribution. Variation of a Poisson distribution equals to its average value:

$$\text{Variation}(\Delta N) = \sigma^2(\Delta N) = \mu\_\Delta N \quad (1.19)$$

On nano-scale devices, the variation in ΔN is only part of the $|\Delta V_{th}|$ variation, the random spatial distribution of the charges also plays an important role. This was shown by Asenov [56] through simulation of random dopant distributions.

To reconcile with the results in [56], an empirical factor of 2 [57] needs to be added in the denominator to convert $|\Delta V_{th}|$ to ΔN:

$$\Delta N = \frac{|\mu\_\Delta V_{th}|}{2\eta} \quad (1.20)$$

Substitute Equation (1.20) into Equation (1.19),

$$\frac{\sigma^2(|\Delta V_{th}|)}{4\eta^2} = \frac{|\mu\_\Delta V_{th}|}{2\eta} \quad (1.21)$$

Equation (1.21) is equivalent to Equation (1.18) and is verified by experiment data from different groups [58, 59].

Based on equation (1.18) η can be extracted by

$$\eta = \frac{\sigma^2_{\Delta Vth}}{2 \cdot \mu_{\Delta Vth}} \quad (1.22)$$

Equation (1.22) is used to extract the η of different types of defects in chapter 6.

As a summary of statistical analysis of traps during recovery phase, defect centric theory alone cannot predict device lifetime, but it's a very useful tool. With defect centric theory, people can use monte-carlo simulation to construct the Time Dependent Variability of each single device from the mean value and Device-to-Device Variability, as detailed in chapter 6.

1.5 Organization of this thesis

A brief review of BTI and its modelling for lifetime prediction is discussed in chapter 1.

In chapter 2 the standard characterization method and some specialized techniques used in this work will be introduced.

In chapter 3 a novel Stress-Discharge-Recharge technique is developed to investigate the defect generation under BTI stress. The newly developed Stress-Discharge-Recharge method captures the entire Generated Defects by adding a recharge phase and makes the time exponent independent of measurement condition.

Pre_Existing defects are investigated in chapter 4. Pre_Existing defects related tests need to be done on heavily stressed devices on which Generated Defects' distortion is suppressed. Experiment data shows the existence of two types of Pre_Existing defects: As-grown Traps and Energy Alternating Defects. By applying so-called multi-

Discharging-based Multiple Pulses technique, these two types of defects can be separated and then modelled respectively.

In Chapter 5 the As-grown Generation (A-G) model is proposed base on the defect generation in chapter 3 and Pre_Existing defects in chapter 4. Good predicting capability has been verified under both NBTI and PBTI stress on big devices. In addition, a new fast wafer-level test scheme for parameter extraction is developed, reducing test time to one hour per device and significantly improving the efficiency for variability tests of nanoscale devices.

In chapter 6 a comprehensive A-G framework is proposed based on the A-G model in chapter 5 and defect centric theory to do the lifetime prediction on nano-scale devices. The framework has excellent predictive capability, as validated by comparison with experimental data. After that the framework is implemented into a commercial simulator and its applicability for circuit level simulation is demonstrated.

Finally the conclusion and future work are given in chapter 7.

2 Characterization methodology

2.1 Introduction

BTI was discovered almost 50 years ago [4] and people now still haven't reached an agreement on its underlying mechanism and modelling. One major difficulty causing this is the BTI characterization. In early works which use slow DC measurements, $|\Delta V_{th}|$ is underestimated to a large extent, due to 1). In Measure-Stress-Measure (MSM) scheme, BTI will recover instantly once stress is removed, before the measurement is taken [60]. 2). In On-The-Fly scheme, fresh reference is contaminated by the fast charging once stress is applied, resulting in a smaller $|\Delta V_{th}|$. In the early stage of switching from slow DC measurement to fast BTI characterization people were puzzled by the inconsistency of results from different characterization methods.

This chapter is organized into eight sections: sample information is firstly given in section 2.2. Typical electric characterization methods are introduced in section 2.3 and BTI-related characterization schemes are introduced in section 2.4. After that an energy profiling technique, Discharging-based Multiple Pulses (DMP) technique is introduced in section 2.5, which is widely used in the As-grown Generation (A-G) model parameter extraction. In section 2.6 two mainstream characterization methods of variability on nano-scale device: Random Telegraph Noise (RTN) technique and Time Dependent Defect Spectroscopy (TDDS) technique are briefly reviewed. Repetitive tests need to be done on

multiple nano-scale transistors for variability statistical analysis, which is time-consuming. The fast wafer-level characterization is usually used. Two popular fast BTI parameter extraction techniques: Voltage Ramp Stress and Voltage Step Stress are introduced in section2.7. Finally a summary is drawn in section 2.8.

2.2 Sample preparation in this work

The sample information used in this work is summarized in Table 3.1. Both nMOSFET/pMOSFET fabricated by five different processes from two suppliers: 1). Interuniversity Microelectronics Center (IMEC) and 2). CSR plc (formerly Cambridge Silicon Radio, acquired by Qualcomm in 2015). The CSR sample is manufactured in Taiwan Semiconductor Manufacturing Company (TSMC) with its most profitable 28nm technology nodes.

Table 2.1 Test sample information

Sample name	**HK45**	**HK22**	**SiON**	**FF**	**CSR**
Gate Material	Metal	Metal	Poly-Si	Metal	Metal
Dielectric	HK stack	HK stack	SiON	HK stack	HK stack
EoT (nm)	1.45	1	2.5	1	1.2
Technology Nodes (nm)	45	22	45	22	28
Structure	Planar	Planar	Planar	FinFET	Planar
Supplier	Interuniversity Microelectronics Center (IMEC)				CSR plc (acquired by Qualcomm in 2015)

Samples listed in Table 3.1 cover a wide range of fabrication conditions: from immature processes under development (HK45, HK22, SiON and FF from IMEC) to mature

commercial processes (CSR). The versatility of a proposed model can be guaranteed if it is applicable to all the processes. Device structure and the corresponding energy band diagram during NBTI stress condition are given in Fig. 2.1.

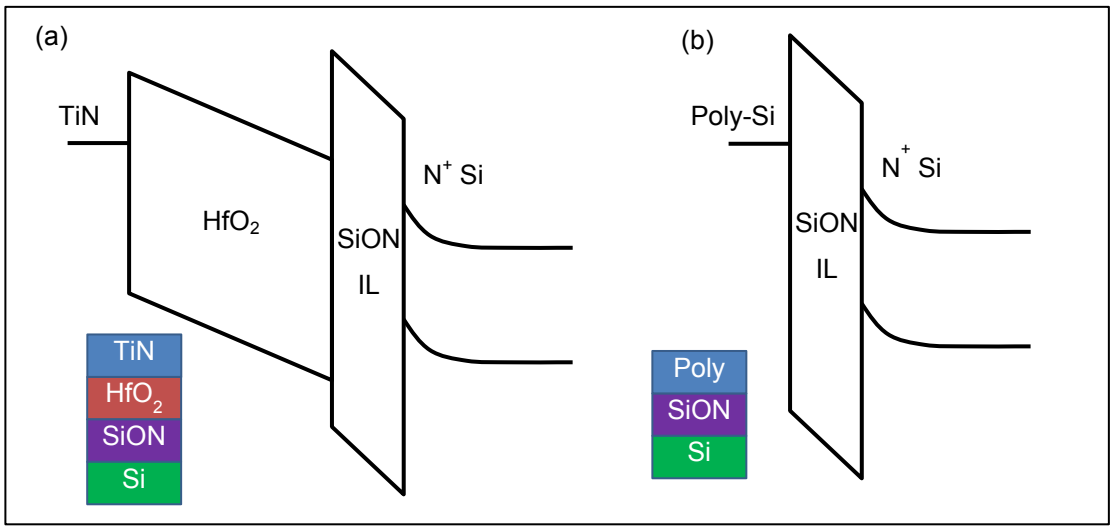

Fig. 2.1 Device structure and the corresponding energy band diagram during NBTI stress condition of High-k Metal Gate (a) and Poly-Si Gate SiON (b).

2.3 Typical electric characterization method

2.3.1 I_d-V_g (IV) measurement

As BTI is normally characterized by "threshold voltage shift" $|\Delta V_{th}|$, I_d-V_g measurement and V_{th} extraction is the basic element in BTI characterization.

IV sweep measurement is the most conventional characterization method with commercial Source Measure Units (SMUs) [61]. The test waveform and typical results are shown in Fig. 2.2. An SMU supplied gate voltage (V_g) sweep is applied to the MOSFET gate, while

the drain is biased under a constant drain voltage (V_d). Both voltage and current values are measured during each V_g level. The typical measurement speed for one I_d measurement is from several to hundreds of milliseconds, making the full IV Sweep testing time from one to ten seconds, thus this method is also called as quasi-static or slow DC measurements.

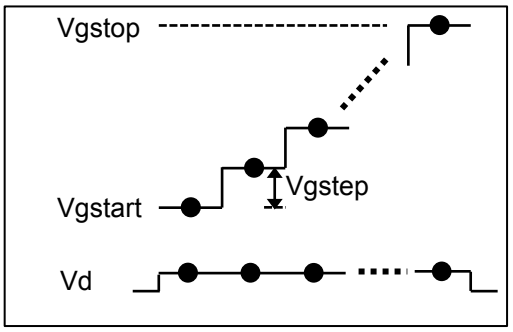

Fig. 2.2 Test waveform of IV sweep measurement. An SMU supplied V_g sweep is applied on the gate while another SMU supplied constant drain voltage V_d is applied on the drain. Both voltage and current can be captured simultaneously by the SMUs. Measurement time for each point is from several to hundreds of milliseconds.

IV sweep provides the most accurate measurement results (sub-pA) due to the high resolution of SMU. Gate leakage and sub-threshold current can be accurately measured. But because of the slow measurement speed, IV sweep cannot capture the "real" $|\Delta V_{th}|$ or "real" fresh Id reference (I_{d0}), resulting in an underestimation of BTI degradation in both Measure-Stress-Measure (MSM) measurement scheme and On-The-Fly measurement scheme, as detailed below:

People normally use 0V as V_{gstart}. In MSM scheme if IV sweep is used to measure $|\Delta V_{th}|$, at the starting V_g=0V measurement, stress is interrupted and most of the fast-recoverable traps have already discharged after tens of milliseconds measurement, resulting in an underestimation of total BTI degradation [60, 62]. In On-The-Fly scheme a "real" fresh

I_{d0}, free from any trap, under stress V_g is needed as reference. However, if V_{gstop} is set equal to or higher than the stress voltage, during that I_{d0} measurement, a lot of fast traps have already been charged and the measured I_{d0} is already degraded.

To overcome the shortages of the IV sweep measurement fast IV measurement becomes a necessity. Spot IV measurement is one of the widely used fast IV characterization methods. The test waveform is shown in Fig. 2.3. The measurement is taken at the flat region of each pulse and the typical measurement time is about microseconds or sub-microseconds, depending on the accuracy requirement. Unlike IV sweep in which V_g is monotonic and incremental, V_g in Spot IV will drop to 0 after each measurement to avoid unintentional traps to be charged during the measurement.

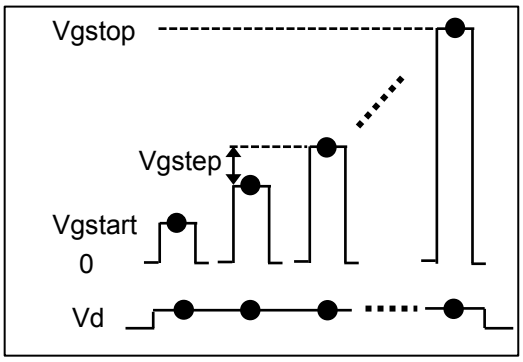

Fig. 2.3 Test waveform of Spot IV measurement. After each measurement V_g firstly drops to zero and then rise to the next level to avoid potential stress during the measurement. Pulse units are adopted in spot-IV measurements so the measurement time for each point can be as low as (sub-) microseconds.

With the development of commercial semiconductor analyzers, high speed measurement enables the full monitoring of I_d current during a pulse on the gate of MOSFET. This is the pulse IV measurement, as shown in Fig. 2.4.

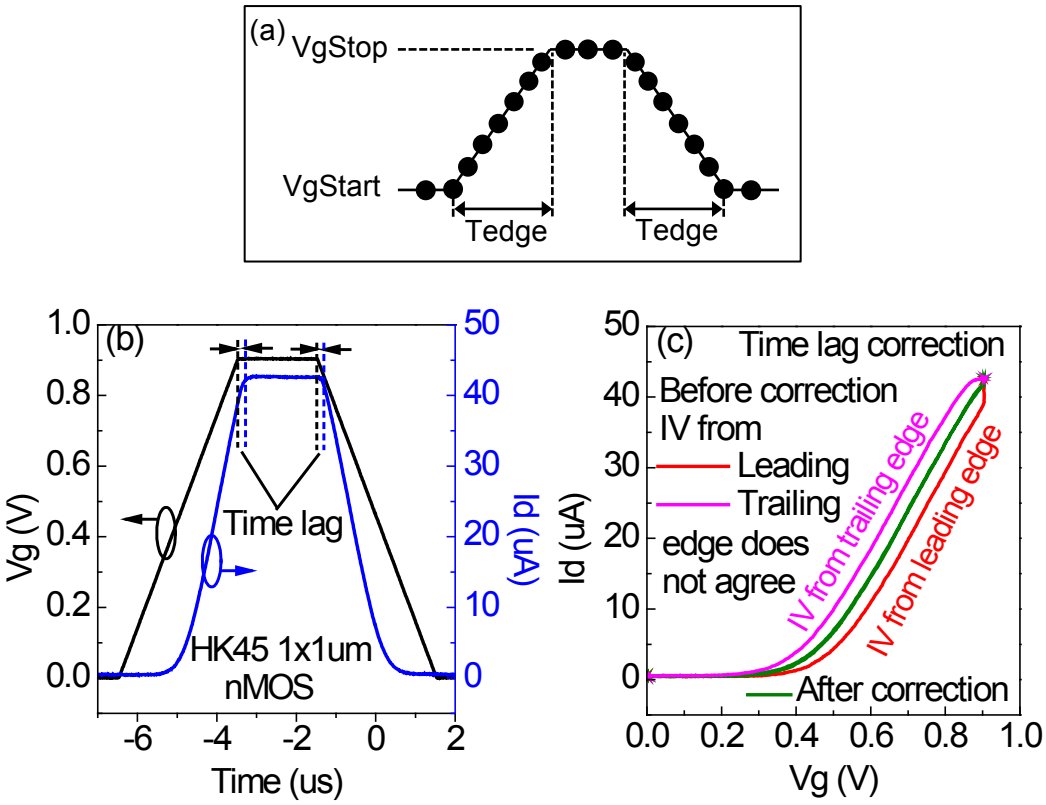

Fig. 2.4 (a) Typical waveform of a Pulse IV measurement. (b) Typical results of pulse IV. Dash lines are drawn at the switching point of both V_g (black) and I_d (blue), clearly I_d measurement is lagging compared to V_g measurement. This time lag results in a gap between the two IV curves measured from the leading and trailing edge of the V_g pulse, as show in (c). This gap can be corrected by shift Id data forward against time to offset the time lag, as the green line shows in (c).

Pulse IV can be used to measure both fresh and stress IV curves as the entire measurement only takes several microseconds thus most of the degradation is "frozen".

Note in a pulse IV I_d measurement is lagging to the V_g measurement (Fig. 2.4b). This results in a gap between the pulse IV from the leading edge and from the trailing edge. The gap can be corrected by shifting the I_d data to offset the lagging time in terms of V_g, as shown in Fig. 2.4c.

After measuring the I_d-V_g curve V_{th} needs to be extracted to get $|\Delta V_{th}|$. Various methods to extract V_{th} from an IV have been proposed [63], here only two mainstream methods are discussed: max-g_m method and constant current method.

Max-g_m method is one of the most popular V_{th} extraction methods. The following is the procedure: Firstly transconductance, g_m, of the IV curve is calculated by first derivative (slope), secondly V_g at max-g_m ("o") is determined, after that, a tangent line through V_g at max-g_m point on the IV, whose slope is the max-g_m value, will intercept V_g axis at threshold voltage V_{th}, as shown by the "□" symbol in Fig. 2.5.

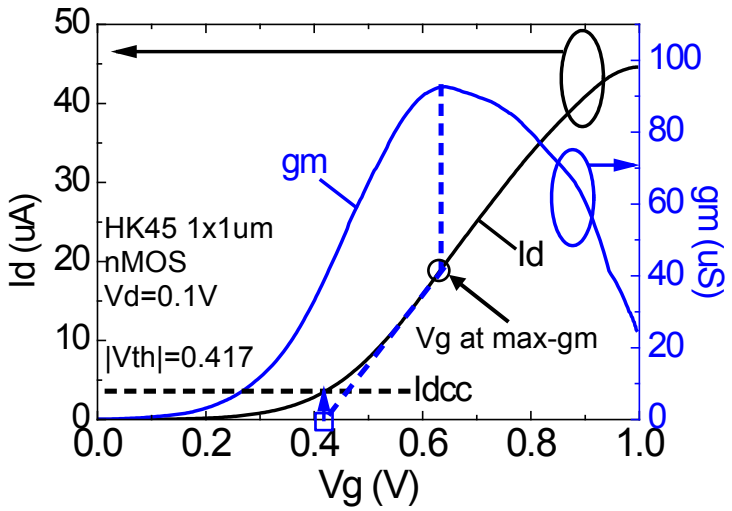

Fig. 2.5 Max-g_m (blue dash lines) and constant current (black dash line) method extracted V_{th}. Here I_{dcc} is properly selected to intercept at the same V_{th} as max-g_m method on a fresh device.

Constant current is another widely used method to extract V_{th} due to its simplicity, versatility and better accuracy compared with max-g_m. The procedure is also shown in Fig. 2.5: firstly draw a line I_d = constant drain current (I_{dcc}), whose typical value is 100nA·W/L, where W and L are the Channel width and Channel Length, respectively. The V_g value of the cross point of this line and IV curve is V_{th}.

Both max-g_m and constant current method have their advantages and disadvantages. Max-g_m extracted V_{th} is a definite value, while in the constant current method the extracted V_{th} is dependent on the I_{dcc} selection, which is an arbitrary value. But as a sacrifice of first derivative, max-g_m extracted $|\Delta V_{th}|$ is less accurate compared to the constant current method. Moreover, for some IV curves whose V_g stops at $|V_{th}|$+0.1 or $|V_{th}|$+0.2V, the max-g_m method cannot be applied as V_g point at max g_m is beyond the measurement, as shown in Fig. 2.5, max-g_m is located at $|V_{th}|$+0.23V.

To combine the advantages of these two methods, in this work the following procedure is used to extract V_{th} if not specified: Firstly, measure a fresh IV curve from zero to operating voltage; secondly use the max-g_m method to extract the fresh V_{th0}; thirdly, calculating the I_d value on the IV curve corresponding to $|V_{th0}|$, and use this I_d value as I_{dcc} for CC method for all the rest $|V_{th}|$ extraction. In this way, the constant current method will extract the same V_{th0} on fresh device as max-g_m, which is of physical meaning, and retain all the benefits of constant current method.

2.3.2 Split Capacitance Voltage (CV) measurement

Split CV measurement is widely used to evaluate the defect properties in MOSFETs [64-66]. "Split" means Gate-Channel capacitance (C_{gc}) and Gate-Bulk capacitance (C_{gb}) need to be measured separately, as shown in Fig. 2.6a&b. It was firstly proposed by J. Koomen [67] in 1973 to extract the charge trapped in interface states in weak and intermediate inversion bulk doping density.

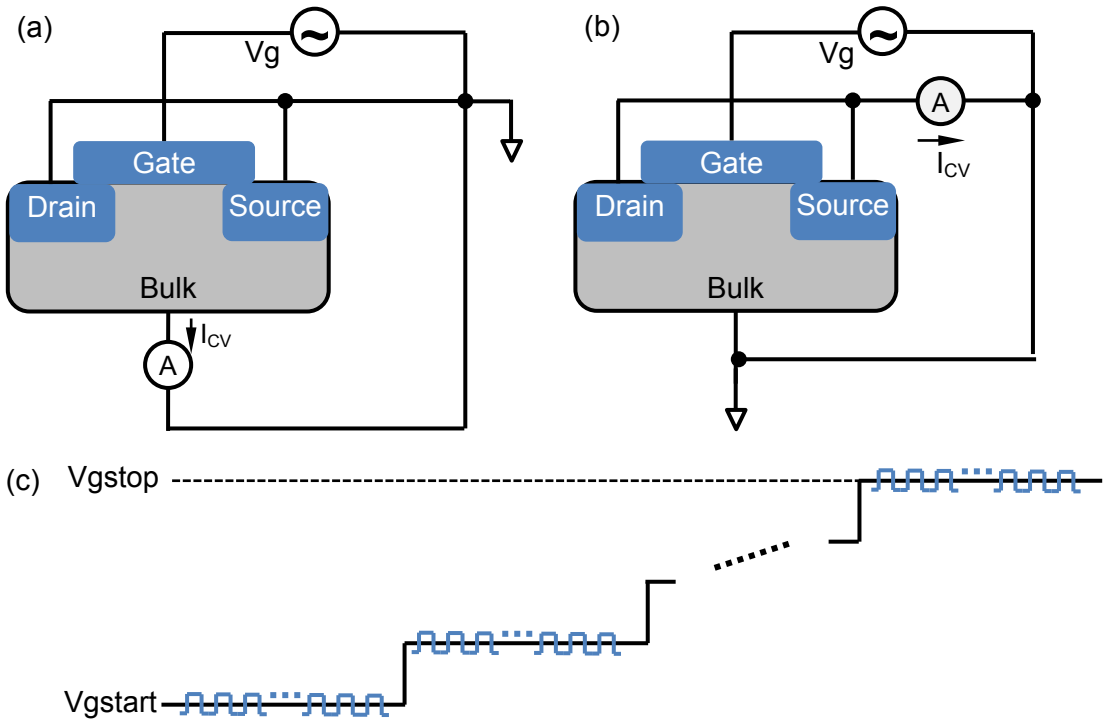

Fig. 2.6 Typical configuration and waveform for MOSFET split CV measurements. (a) Drain and Source are both grounded to measure C_{gb}, (b) Bulk is grounded, Drain and Source are connected together to measure C_{gc}. (c) Illustration of the V_g waveform applied on the gate for both C_{gb} and C_{gc}. A V_g-sweep signal is applied on the gate to form the MOS capacitor, then a high frequency (10kHz~10MHz) and small amplitude (~30mV) AC signals are added on top of each V_g level, the AC current through the capacitor is then measured and capacitance is calculated.

Lots of useful information can be extracted from the split CV such as flat band voltage, bulk doping, surface potential (φ_s) etc. A physics based CVC simulator [68] including polysilicon depletion and surface quantum mechanical effects is used in this work to extract all these device parameters, as shown in Fig. 2.7b. The CVC simulator needs only two inputs from the user: measured full CV data and MOSFET geometry (Width and Length), a theoretical CV which best fits the measured data is then generated with all the fitting parameters related to CV, as listed in Table 2.2.

V_{th} can also be extracted from split CV [63], but due to the split CV is even slower than IV sweep, and C_{gb} & C_{gc} need different measurement configuration, it is not used to measure BTI $|\Delta V_{th}|$ in this work. It is the φ_s~V_g relationship, one of the outputs from CVC simulator, mainly used in this work. φ_s~V_g relationship can be used to convert the V_{gdisch} to E_f-E_V, as detailed in section 2.5. Note if split CV test is carried out on a heavily stressed device, Generated Defects and stress induced gate leakage will affect the measurement results. Pre_Existing defects' charging won't affect the CV, as shown in Fig. 2.7d.

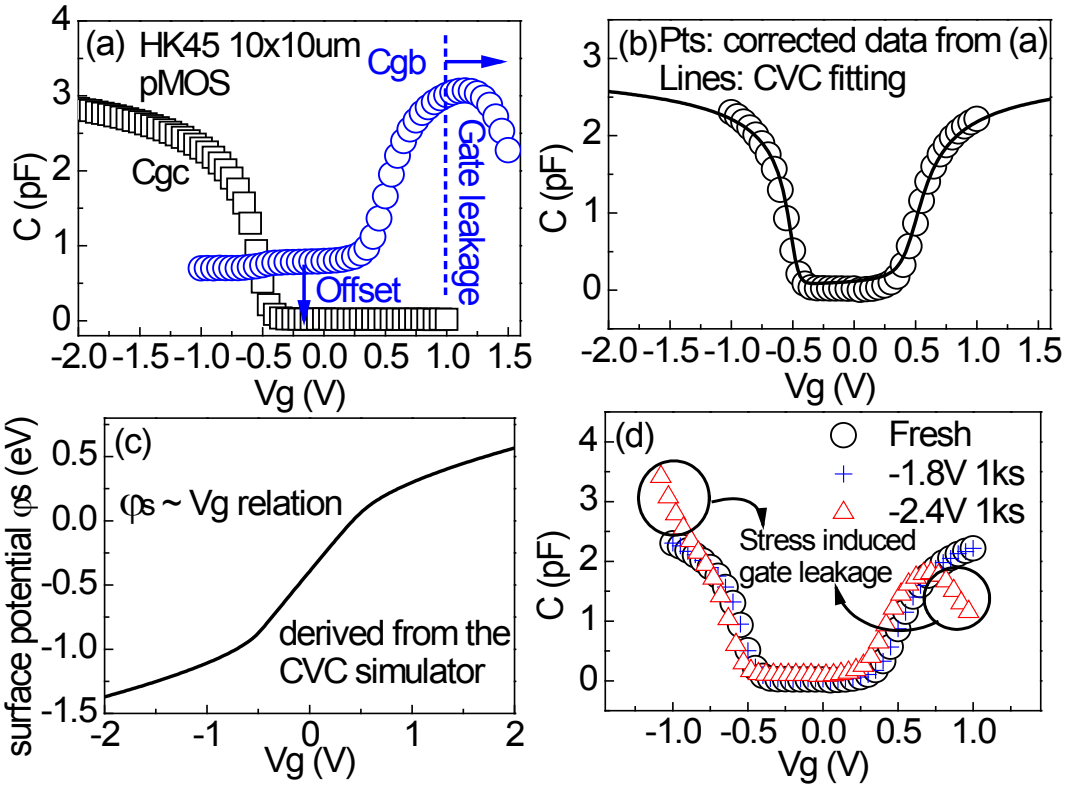

Fig. 2.7 (a) Split CV results on a HK45 WxL=10x10 micrometers pMOS. Note gate leakage starts to become comparable to CV current when V_g exceeds 1V, thus only $|V_g|$<1V data is used for CVC fitting in (b). (b) CVC simulator provided in [68] is used to fit the experiment data. (c) Surface potential against V_g relationship is generated by the CVC simulator with fitting parameters from (b). (d) A comparison of split CV results on fresh and stress device.

Table 2.2 MOSFET parameters extracted from fitting the measured CV with the CVC simulator

parameters	value
Flat band voltage (V)	0.578
Surface doping density (cm^{-3})	2.0e17
Effective oxide thickness (nm)	1.45
Oxide capacitance (pF)	2.11
Bulk doping (cm^{-3})	2.0e17
Interface charge density (cm^{-3})	5.85e12
Bulk potential (V)	-0.426
Threshold voltage(V)	-0.42

Note $|V_{th}|$=0.42V is extracted from the split CV results. This agrees very well with max-g_m extracted $|V_{th}|$=0.417V from the IV measurement, supporting the correctness of the fitting from the CVC simulator.

2.3.3 Charge pumping measurement

Charge pumping is one of the most conventional methods to measure interface states density. A typical charge pumping configuration is shown in Fig. 2.8a, gate is applied a high frequency (~MHz) AC square pulses while charge pumping current is measured at the substrate. J.S. Brugler et al first [69] used such a configuration to investigate MOSFET switching behavior in 1969. They observed a net DC "pumping current" on the bulk when applying periodic pulses to the gate. The current was found to be proportional to the gate area and the frequency of the applied gate pulses. J.S. Brugler showed that the current originates from recombination of minority and majority carriers at traps at the Si-SiO_2

interface, but he did not provide a quantitative formula to depict the pumping current and interface states, which limited the use of this method.

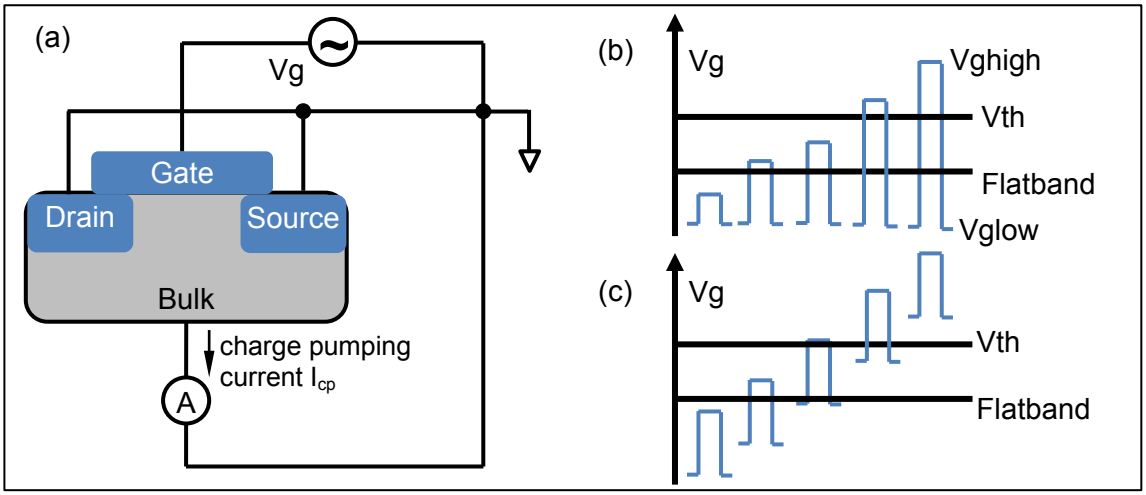

Fig. 2.8 (a) Typical configuration of the charge pumping measurements on nMOSFET. (b) Fixed base charge pumping V_g waveform. (c) Fixed Amplitude charge pumping V_g waveform.

A breakthrough was made by G. Groeseneken et al [70] in 1984 for a thorough investigation and correct explanation of the charge pumping measurement. After that it has become a routine test to measure interface states ever since. A brief introduction and explanation of the charge pumping test and how to calculate the interface states are given hereafter. For more details please refer to [70].

Depending on the V_g pattern, charge pumping tests can be clarified as Fixed Base charge pumping (Fig. 2.8b) and Fixed Amplitude charge pumping (Fig. 2.8c). The former one is used in the first charge pumping paper and the latter one is firstly proposed by A.B.M Elliot [71] in 1976. For both V_g patterns there must be a V_g phase in which V_{ghigh} is higher than V_{th} while V_{glow} is lower than flatband voltage, I_{CP} will reach its max value under this condition, as explained below:

When the gate is biased under V_{ghigh}, the nMOSFET is pulsed into inversion, the surface becomes deeply depleted and electrons will flow from the N^+ source and drain regions into the channel to build the inversion layer, where some of the electrons will be captured by the interface states.

When the gate is biased under V_{glow} (V_{gbase} in Fig. 2.8b&c), nMOSFET is pulsed from inversion into accumulation from inversion, the mobile electrons in the channel drift back to the source and drain, but the electrons trapped in the interface states will recombine with the holes from the P^+ substrate and give rise to a net DC current through the substrate. This is the charge pumping current I_{CP}. The charge interface states, Q_{it}, which will recombine is then given by

$$Q_{it} = A_G \cdot q \cdot \int D_{it}(E)dE \tag{2.1}$$

Equation (2.1) can also be expressed as

$$Q_{it} = A_G \cdot q \cdot \overline{D_{it}} \cdot \Delta\varphi_s \tag{2.2}$$

Where

A_G is the channel area of the nMOS

$D_{it}(E)$ is the interface states density at energy level E

$\overline{D_{it}}$ is the mean interface states density averaged over the energy level swept by E_f

$\Delta\varphi_s$ is the total sweep of surface potential

The total amount of interface states density across the full energy level N_{it} is

$$N_{it} = \overline{D_{it}} \cdot \Delta\varphi_s \tag{2.3}$$

Substitute Equation (2.3) into Equation (2.2),

$$I_{CP} = f \cdot q \cdot A_G \cdot N_{it} \quad (2.4)$$

$$N_{it} = I_{CP}/(q \cdot f \cdot A_G) \quad (2.5)$$

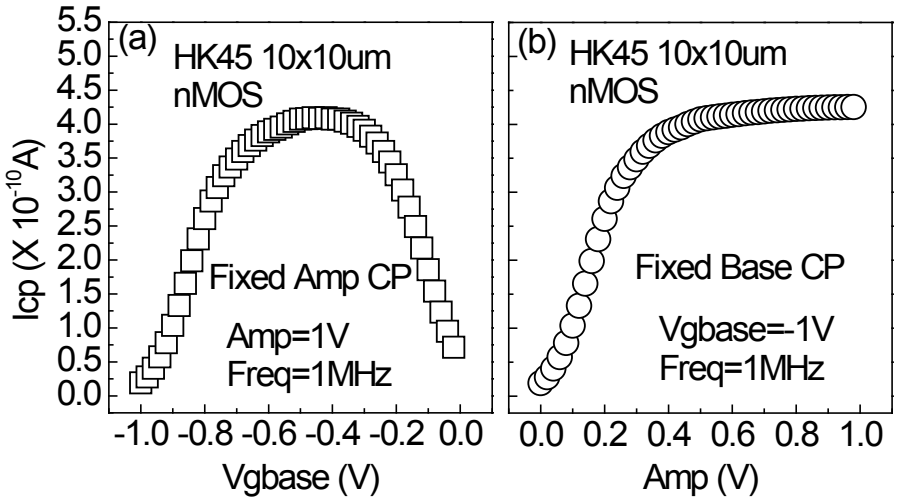

Fig. 2.9 Typical results of (a) Fixed Amplitude charge pumping and (b) Fixed Base charge pumping measurements on a HK45 10x10um nMOSFET.

2.4 Typical BTI Characterization scheme

2.4.1 Measure-Stress-Measure (MSM) characterization method

Most of the conventional BTI characterization methods follow a straightforward Measure-Stress-Measure (MSM) scheme, as shown by the flow chart in Fig. 2.10. Before applying any stress, fresh characteristics such as IV, split CV, charge pumping etc. is firstly measured as fresh reference [10]. Note the fresh measurements must not bring any degradation, otherwise the lifted reference will lead to an underestimation of total BTI degradation. After that, the device is then stressed for a preset period of time whose

interval can be either linear or logarithmic time based. For the tests aiming at the end-of-life prediction, due to the fact that many decades in time have to be covered with these characterization procedures, a logarithmic time base is more frequently employed. When the specified time is reached, the stress condition will be interrupted, and the same measurement on fresh is repeated. This MSM sequence will continue until the total stress time reaches the preset value. If necessary, a post-stress measurement can be carried out to gather more information about the BTI degradation, like discharging kinetics, recovery after annealing etc.

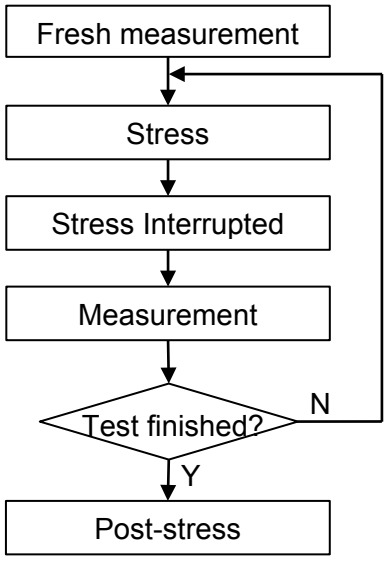

Fig. 2.10 Flow chart of a Measure-Stress-Measure (MSM) test scheme.

Fig. 2.11 shows a typical charge pumping MSM test to study the interface states generation under NBTI stress. $|V_g-V_{th}|$=2.0V NBTI Stress is interrupted by a fixed amplitude charge pumping measurement at the pre-defined log incremental stress time to measure the interface states generation kinetics. After all the charge pumping tests are

finished, the device is floating for one day, the post-stress measurement: another charge pumping measurement is carried out on the stressed device, which shows interface states will not be discharged.

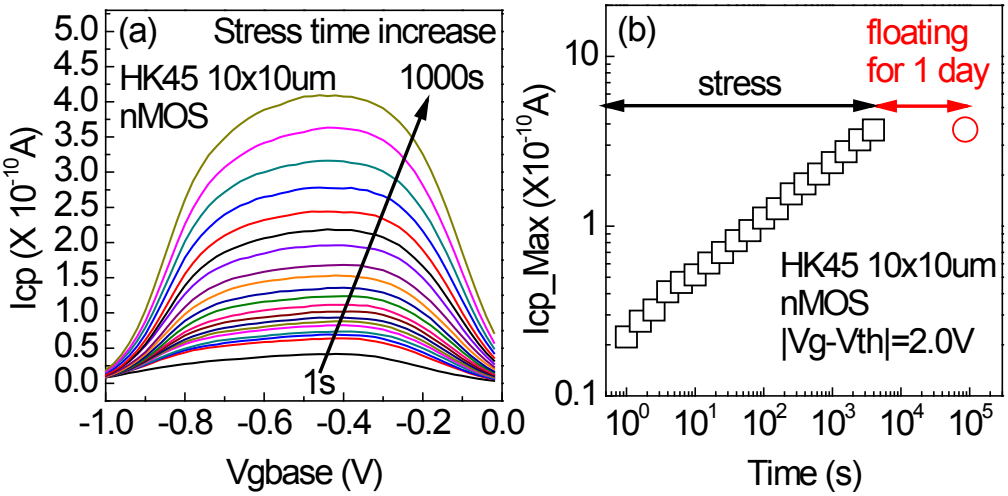

Fig. 2.11 (a) Typical test data of fixed amplitude charge pumping MSM on a HK45 10x10um nMOSFET under $|V_g-V_{th}|$=2.0V PBTI stress. (b) Plot the max value of each line from (a) against stress time, indicating the interface states generation during NBTI stress. After stress the device is floating for one day, and then re-do the charge pumping. This is so-called "Post stress measurement". Results show interface states generated under NBTI will not be discharged.

Charge pumping MSM characterizes the interface states generation during BTI, but interface states are just parts of BTI degradation. The total BTI degradation needs to be characterized by $|\Delta V_{th}|$. Depending on the $|\Delta V_{th}|$ extraction method, BTI MSM can be classified as IV sweep MSM, Spot I_d sense MSM, Extend I_d sense MSM, ultrafast sense MSM, as detailed below:

IV sweep MSM was the most conventional characterization method. It can be used to extract almost all the information such as V_{th}, transconductance g_m, sub-threshold swing (SS), etc. A typical result of IV sweep MSM is shown in Fig. 2.12.

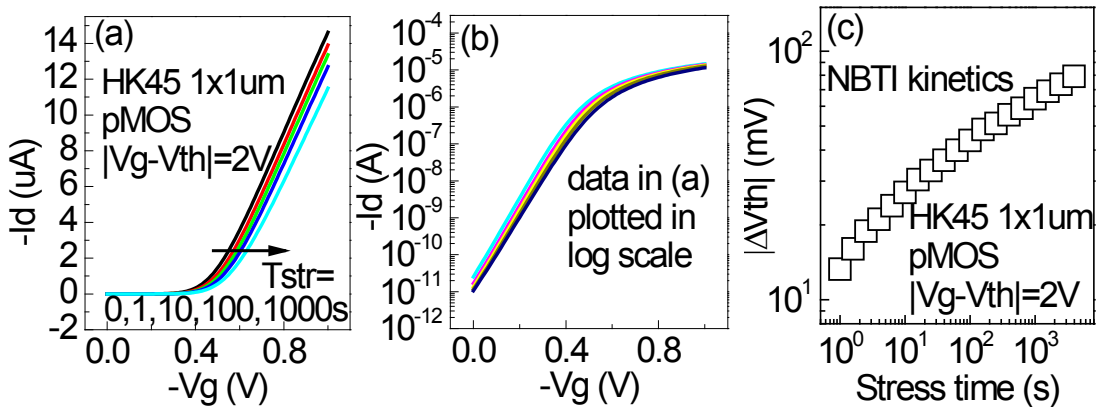

Fig. 2.12 Typical IV sweep MSM measurement results under $|V_g-V_{th}|$=2.0V NBTI stress on a HK45 1x1um pMOS. (a) The linear plot of all the IVs during stress, constant I_{dcc}=1uA is used to extract all the $|V_{th}|$ values. (b) A replot of data in (a) but in log I_d scale to show SS region is well captured by IV sweep measurement. (c) Plot $|\Delta V_{th}|$ from (a) against stress time.

The pitfall of IV sweep MSM has been mentioned in section 2.3.1, the slow measurement speed cannot capture the fast recoverable traps, resulting a big underestimation of the "real" $|\Delta V_{th}|$ [60, 62], epically for NBTI on pMOSFETs, which usually contains a lot of fast recoverable traps (refer to Fig. 1.6).

After realizing the underestimation due to the fast recovery during the slow IV sweep MSM, people started to develop new techniques to minimize the recovery and characterize total BTI. A straightforward thinking is reducing the measurement time. Spot-Id sense MSM is proposed to achieve this purpose [72]. A schematic depiction of the test waveform of Spot I_d sense MSM is illustrated in Fig. 2.13a. A Spot IV measurement is firstly carried out on a fresh device to measure the reference, then the stress is applied on the gate. When the preset stress time is reached, V_g is lowered down to the sensing V_g (V_{gsense}) to measure only one I_d value instead of a full IV, $|\Delta V_{th}|$ is then extracted by the comparison of I_d and fresh I_{d0} (I_d at V_{gsense} on fresh IV), as shown in Fig. 2.13b.

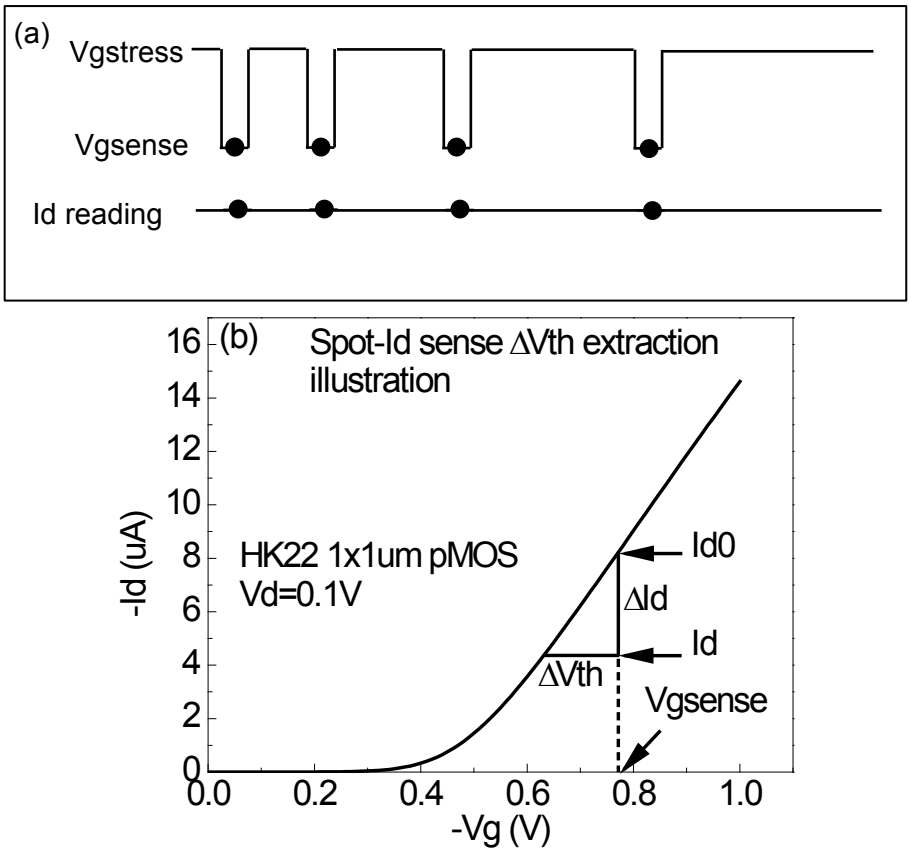

Fig. 2.13. (a) Test waveform of Spot I_d sense MSM method. (b) Illustration of $|\Delta V_{th}|$ extraction of Spot I_d sense MSM measurement.

The Spot I_d sense MSM method reduced the measurement time to microseconds or sub microseconds by using the pulse measurement and only measure one I_d, hence the entire degradation can be captured. But the $|\Delta V_{th}|$ extraction method contains uncertainty. Note in Fig. 2.13b $|\Delta V_{th}|$ is sensing under a constant sensing V_g (V_{gsense}) instead of a constant I_d. As $|\Delta V_{th}|$ increases, a change of charged traps in the gate dielectric will change the substrate surface potential, so that the $|\Delta V_{th}|$ are actually measured at different surface potential. As can be seen in Fig. 2.14, if the degradation is small (<30mV), $|\Delta V_{th}|$ extracted from constant V_{gsense} and constant I_d is comparable. However, when $|\Delta V_{th}|$

further increases, degradation from constant V_{gsense} method becomes lower than constant I_d method. This is because high $|\Delta V_{th}|$ lowers E_f if V_g is kept constant.

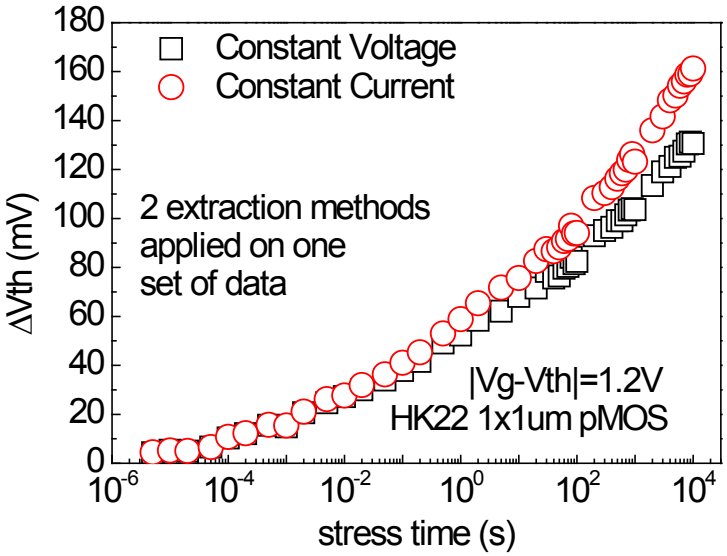

Fig. 2.14. The impact of different sensing techniques on the degradation. The increase of $|\Delta V_{th}|$ will reduce the surface potential if the constant sensing V_g is used and thus leads to the lower degradation when compared with sensing at the constant current level which is approximately at the same surface potential.

A solution is proposed to overcome this problem as depicted in Fig. 2.15. A Pulse IV MSM test is firstly carried out prior to the Spot Id sense MSM. From the Pulse IV measurements both $|\Delta V_{th}|$ from constant current and $\Delta I_d/I_{d0}$ at a constant sensing V_g can be achieved. $|\Delta V_{th}|$~$\Delta I_d/I_{d0}$ is plotted in Fig. 2.15. It is shown that the correlation between them is irrespective of stress conditions and can be fitted by a cubic equation well. As long as this unique correlation is set up, it can be used to convert the $\Delta I_d/I_{d0}$ measured under a constant sensing V_{gsense} to the corresponding $|\Delta V_{th}|$ measured at the same surface potential (constant current).

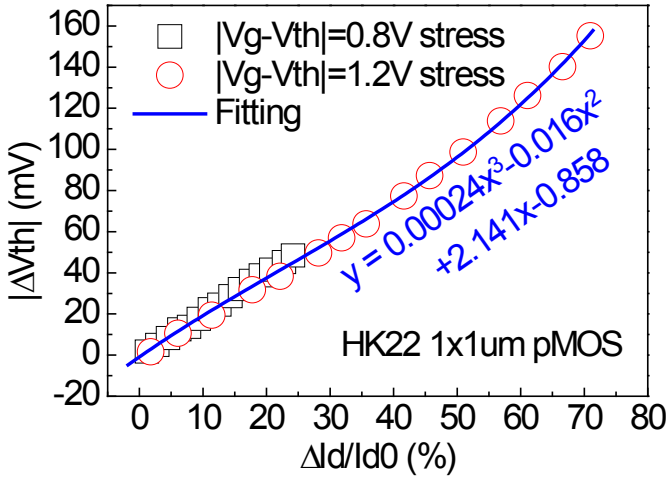

Fig. 2.15. Experimental extracted unique relationship between $\Delta I_d/I_{d0}$ and $|\Delta V_{th}|$. $\Delta I_d/I_{d0}$ is taken from the constant V_{gsense} and the $|\Delta V_{th}|$ is from constant current method. This relationship is independent of stress condition and can be used to convert $\Delta I_d/I_{d0}$ into $|\Delta V_{th}|$.

As a key feature, BTI recovery will help people to understand BTI mechanism thus is worthwhile to investigate. Extended Id sense MSM method is proposed to study the BTI recovery kinetics [73]. A schematic illustration of this technique is given in Fig. 2.16.

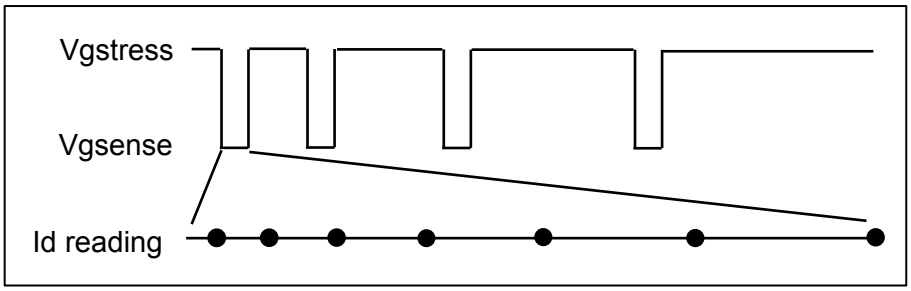

Fig. 2.16 Test waveform illustration of the extended Id sense measurement technique. The waveform is very similar to spot-I_d MSM, except extended Id sense technique takes multiple measurements instead of one at the V_{gsense} level.

Extended I_d sense MSM method is very similar to the Spot I_d sense MSM. The only difference is Extended I_d sense MSM measures multiple I_d values under V_{gsense} at a preset discharging time sequence to monitor BTI recovery kinetics.

With the development of semiconductor analyzers, such as high-speed oscilloscope, microseconds or sub-microseconds measurements become available. This is the so-called "ultrafast sense" measurement.

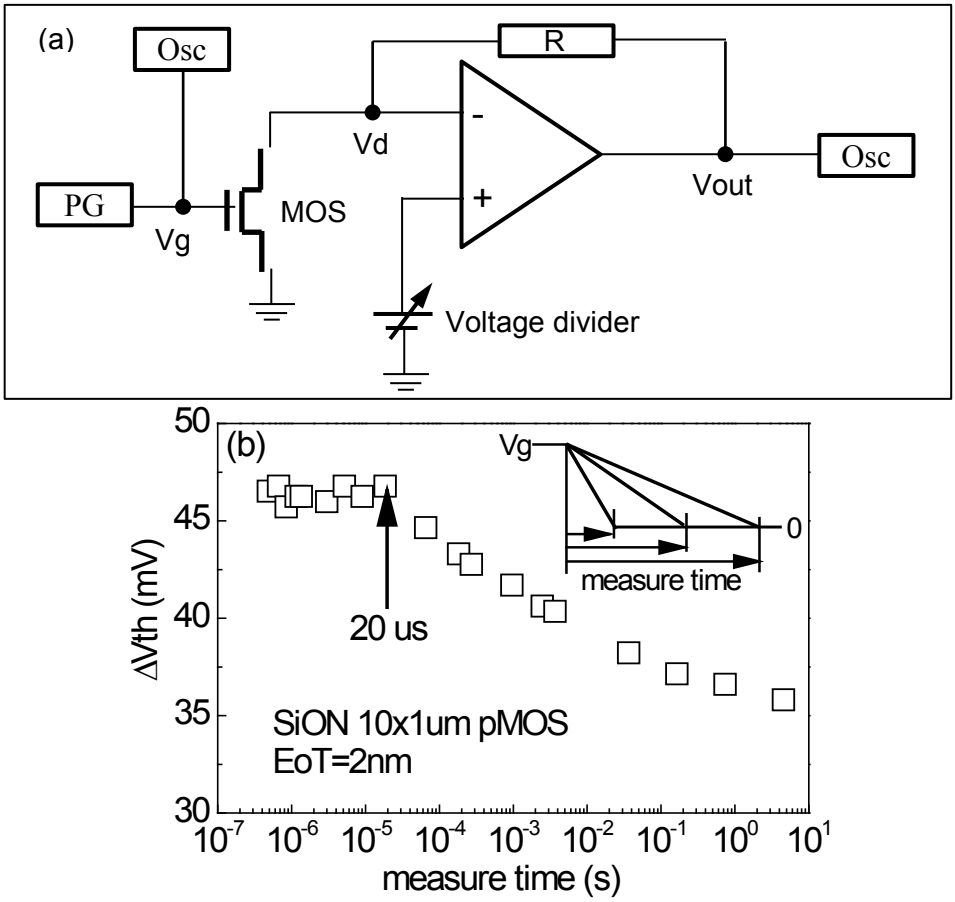

Fig. 2.17. (a) Test configuration of ultrafast sense measurement technique. (b) Impact of measurement speed on $|\Delta V_{th}|$, inset shows the pulse IV waveform with different measure time. It is shown that on SiON pMOS 20 microseconds is capable to capture the entire NBTI degradation

The first ultrafast sense technique was proposed in [74], the fast speed is achieved by a high speed current-voltage Amplifier which can convert the I_d current driven by a ~microseconds pulse on the gate, to voltage, thus can be monitored by a high-speed oscilloscope. Fig. 2.17a shows the "ultrafast sense" measurement configuration used in this work.

As a sacrifice of the fast measurement speed, the accuracy of the "ultra-fast sense" MSM will decrease with higher speed. Fig. 2.17b shows the impact of measurement speed on $|\Delta V_{th}|$, it is found that on SiON process, as long as the "ultrafast" pulse IV is taken within 20 microseconds, the entire $|\Delta V_{th}|$ can be captured. Considering the process difference and the accuracy of the facilities used in this work, 3 microseconds "ultra-fast" pulse IV is adopted to measured $|\Delta V_{th}|$ hereafter if not specified.

2.4.2 On-the-Fly (OTF) characterization method

Apart from reducing the testing time, another way to suppress the recovery is to avoid interrupting the stress while measuring V_{th}. On-The-Fly characterization method was proposed [75, 76] following such thinking. A typical On-The-Fly test pattern is depicted in Fig. 2.18. The stress voltage V_g is always applied on the gate, when it reaches the specified stress time, ΔI_d current will be monitored right at V_g to extract $|\Delta V_{th}|$ by dividing the local transconductance g_m, which is calculated from the I_d measurements at another two voltages close to V_g, V_g-ΔV, V_g+ΔV, as illustrated in Fig. 2.18.

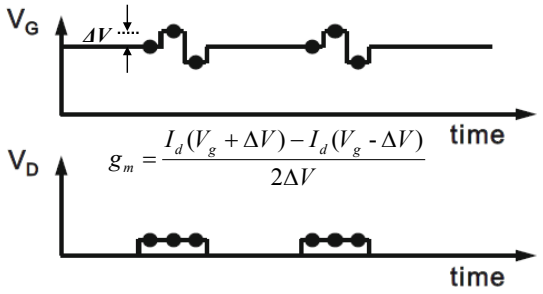

Fig. 2.18 On-The-Fly characterization method illustration [12]. $|\Delta V_{th}|$ is calculated by $\Delta I_d/g_m$. To measure the real-time g_m another two V_g levels V_g+ΔV and V_g-ΔV is applied on the gate. ΔV is very small so the impact on the stress condition is negligible.

One important issue for this On-The-Fly technique is the selection of the reference I_{d0}. Since all the $|\Delta V_{th}|$ is converted from $\Delta I_d/g_m = (I_{d0}-I_d)/g_m$, this means that a wrong I_{d0} will affect all the $|\Delta V_{th}|$ results [60, 62]. Slow DC measured I_{d0} cannot be used because it has already been contaminated by the charging of fast recoverable BTI traps during tens or hundreds of milliseconds I_{d0} measurement under stress V_g. Fast measurement like pulse IV or spot IV has to be used to get the "real" fresh reference I_{d0}.

On-The-Fly method can eliminate the recovery completely compared to MSM as the stress has never been removed. By using the proper fresh reference, $|\Delta V_{th}|$ sensing at stress V_g can be reliably extracted. However, On-The-Fly method extracted $|\Delta V_{th}|$ is sensing at the stress V_g instead of real V_{th0}, for NBTI on silicon devices, due to mobility degradation, $|\Delta V_{th}|$ sensing at stress V_g is usually much larger than sensing at V_{th0}, as shown in Fig. 2.19.

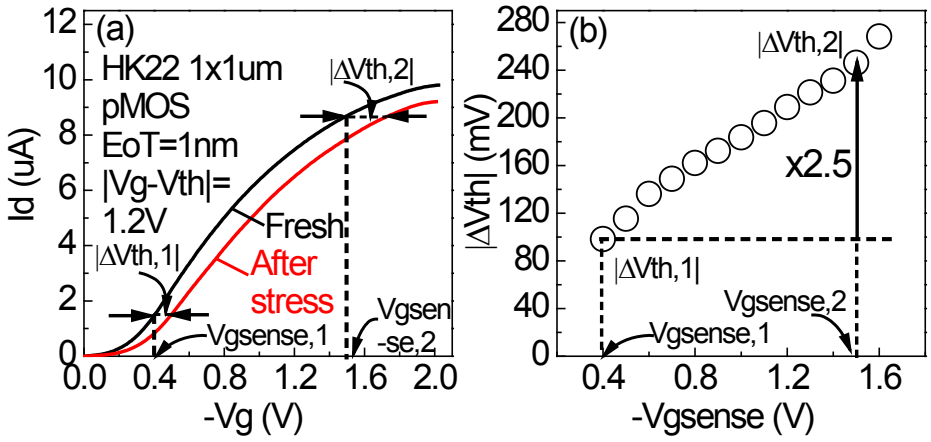

Fig. 2.19 (a) Illustration of $|\Delta V_{th}|$ extracted at different sensing V_g. Pulse IV MSM is carried out on HK22 1x1um pMOSFET under $|V_g-V_{th}|$=1.2V 1ks NBTI stress. (b) $|\Delta V_{th}|$ against V_{gsense} plot. $|\Delta V_{th}|$ extracted from V_{gsense}=1.5V is 2.5 times higher compared to V_{gsense}=-0.4V.

The phenomenon in Fig. 2.19 is called "V_g-Effect" as detailed in [60]. The solution is to use $|\Delta V_{th}|$~$\Delta I_d/I_{d0}$ correlation measured from pulse IV MSM to convert $\Delta I_d/I_{d0}$ into $|\Delta V_{th}|$, as shown in Fig. 2.15.

Another simplified On-The-Fly pattern is as shown in Fig. 2.20. Constant stress V_g is applied on the gate and constant V_d is applied on the drain. I_d is continuously monitored against stress time. This On-The-Fly measurement can capture I_d fluctuation continuously thus is widely used on nano-scale devices. The conventional Within Device Fluctuation measurement in chapter 6 is using this pattern. Fresh g_m at stress V_g from fresh device is divided by ΔI_d to get $|\Delta V_{th}|$, which also suffers from "V_g-Effect". In this work, $\Delta I_d/I_{d0}$~$|\Delta V_{th}|$ relationship is used to convert ΔI_d to $|\Delta V_{th}|$, as detailed in section 6.3.1.

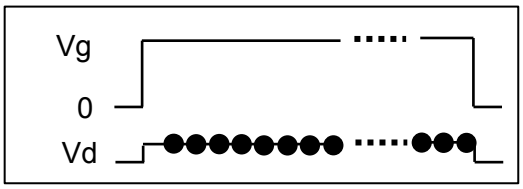

Fig. 2.20 Test waveform of the simplified On-The-Fly measurement scheme. I_d is continuously monitored under a constant V_g stress.

2.5 Discharging-based Multiple Pulse (DMP) technique

2.5.1 Conventional DMP technique

Defects in gate dielectric play an important role during BTI stress, the characterization of these defects' energy distribution is very useful for both process qualification in industry

and better understanding for modelling. A Discharge-based multi-pulse technique (DMP) technique was proposed to do this by X. Zheng et al in 2010 [77]. The typical waveform is shown in Fig. 2.21. Note nMOSFET is demonstrated in Fig.2.21, for pMOSFET all the V_g should be in the opposite polarity.

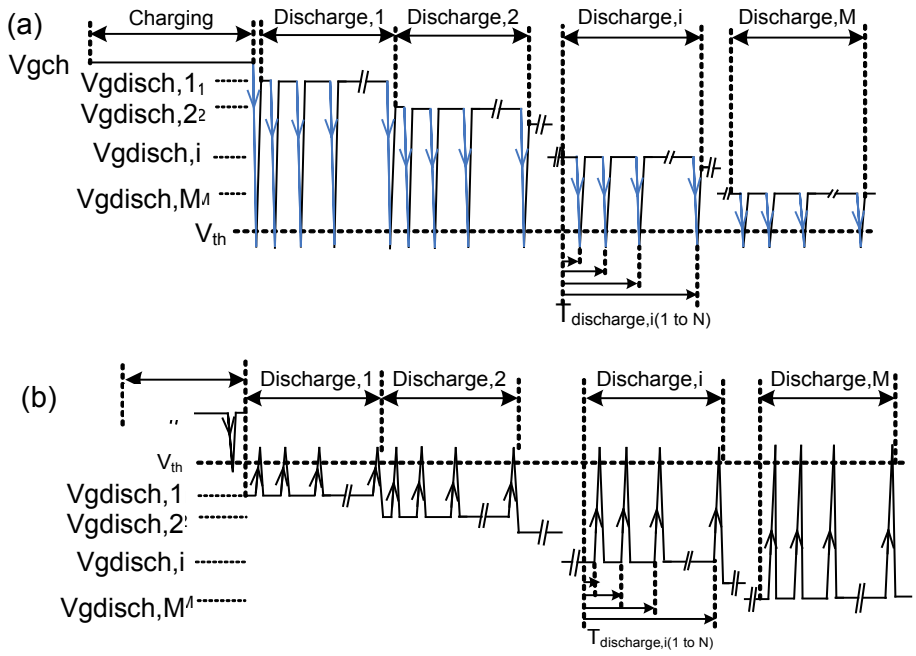

Fig. 2.21 Test waveform of the conventional DMP technique when (a) $|V_{gdisch}|>|V_{th}|$ (b) $|V_{gdisch}|<|V_{th}|$. A charging voltage $|V_{gch}|$ is firstly applied to fill all the chargeable traps, $|V_g|$ is then lowered down step by step to discharge the traps at different energy location. Ultrafast pulse IV is adopted to capture $|\Delta V_{th}|$.

Although the DMP technique is proposed for electron trap energy distribution probing for Flash memory application, it is also applicable to both n-type and p-type MOSFETs [78, 79]. The principles of DMP on a pMOSFET are shown in Fig. 2.22:

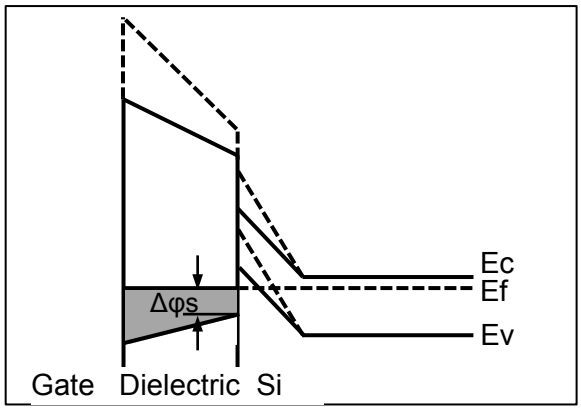

Fig. 2.22 The energy diagram to show the principles of the DMP technique on a pMOSFET. When V_{gdisch} was stepped toward positive direction each time, a shaded area with an energy depth of $\Delta\varphi_s$ at the interface falls below E_f and the positive charges within it start discharging.

A negative V_{gch} is firstly applied on a pMOSFET to fill the hole traps and form the positive charges. Then $|V_g|$ will be lowered to $|V_{gdisch,1}|$, some of the charged positive charges will fall below Fermi level, Ef, as marked by the grey shade in Fig. 2.22. As a first order approximation [78-81], below E_f, positive charges are assumed to be neutralized throughout the oxide given a sufficient discharge time. When the V_{gdisch} moves further towards positive for each step, the energy level of positive charges is lowered against the substrate, bringing a new dash area below E_f for discharging. Through monitoring the $|\Delta V_{th}|$ during each V_{gdisch} step, the profile of the positive charges can be evaluated.

2.5.2 Spot Id DMP technique

Due to the complexity of the pattern, conventional DMP pattern cannot be directly transferred to a mainstream semiconductor analyzer like Keithley 4200A-SCS, thus

making it unsuitable for industry use. To conquer this difficulty a simplified Spot I_d DMP technique is proposed as shown in Fig. 2.23.

A typical result of the Spot I_d DMP technique is given in Fig. 2.24. A charging V_g V_{gch}=-1.6V 10,000 seconds stress is apllied on a HK22 1x1um pMOSFET to filled all the chargeable defects. V_g is then stepping down and the filled postive charges in the gate dielectric start to discharge. Most of the positive charges will discharge within 20 seconds, the discharing trace after 20 seconds is a flat line compared to the obvious drop of $|\Delta V_{th}|$ at the beginning.

The profile of the positive charges induced $|\Delta V_{th}|$ can be extracted by plotting the $|\Delta V_{th}|$ after 20 seconds discharging against the corresponding V_{gdisch}, as shown in Fig. 2.24b. $|\Delta V_{th}|$ is then converted into the effective charge density [82, 83] with Equation (2.6) and (2.7):

$$\Delta N_{OX} = |\Delta V_{th}| \cdot C_{OX}/q \tag{2.6}$$

$$C_{OX} = \varepsilon_0 \cdot \varepsilon_{SiO_2} \cdot W \cdot L/EoT \tag{2.7}$$

Where ε_0 is the vacuum permittivity, ε_{SiO2} is the SiO_2 dielectric constant, W/L is the channel Width/Length and EoT is the effective oxide thickness.

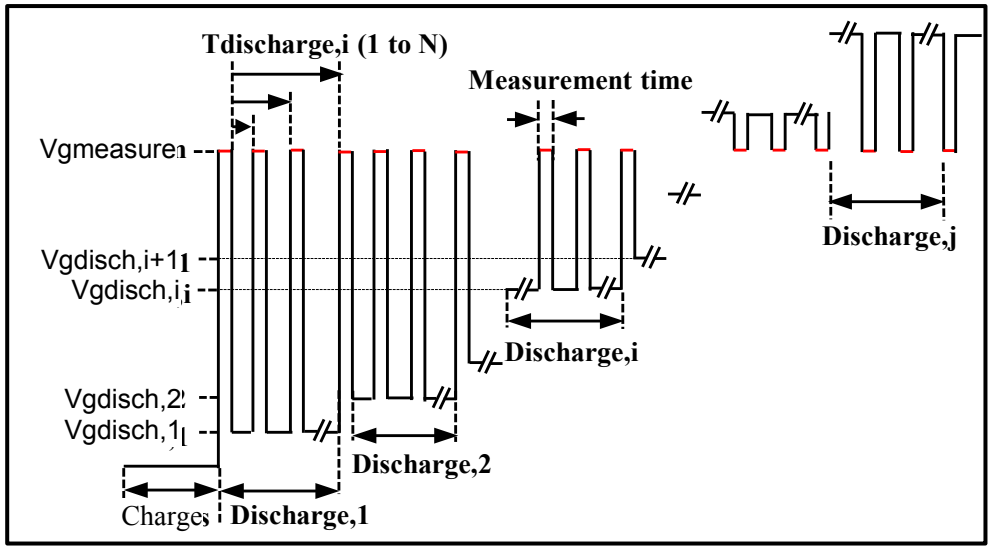

Fig. 2.23. Test waveform of the Spot Id DMP technique.

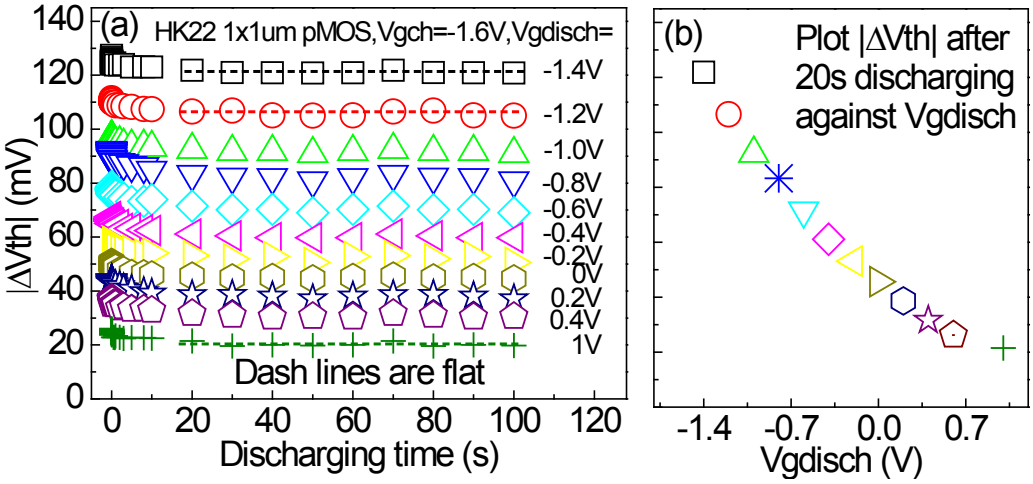

Fig. 2.24. (a) Typical results for discharging kinetics under different V_{gdisch}, The discharge time is the time under a given V_{gdisch} as marked in Fig. 2.23. The device is stressed at -1.6 V for 10 ks before the discharging. (b) Profile of the positive charges induced $|\Delta V_{th}|$.

The number of positive charges $|\Delta N_{ox}|$ is plotted against V_{gdisch} in Fig. 2.25a. $|\Delta N_{ox}|$ extracted from different discharging time overlaps, indicating the impact of discharging time is negligible. The impact of interface states (N_{it}) is also evaluated as shown in Fig. 2.25b, which shows N_{it} is also negligible.

To obtain the energy distribution, V_{gdisch} must be converted to the energy level E_f relative to E_v, i.e. E_f-E_v. As shown in Fig. 2.25(c), E_f-E_v = $E_g/2$+ φ_B- φ_S. A theoretical (E_f-E_v) ~ V_g curve is first calculated using the CVC simulator [68]. By shifting the theoretical curve towards the left until E_f-E_v=$E_g/2$-φ_B (i.e. the strong inversion condition) occurs at the measured V_{th}, the E_f-E_v versus V_{gdisch} relation is obtained and then used to convert V_{gdisch} into E_f-E_v. Fig. 2.26a plots ΔN_{OX} against E_f-E_v. By differentiating ΔN_{OX} against E_f-E_v, the energy density, ΔD_{OX}, is also obtained as shown in Fig. 2.26b.

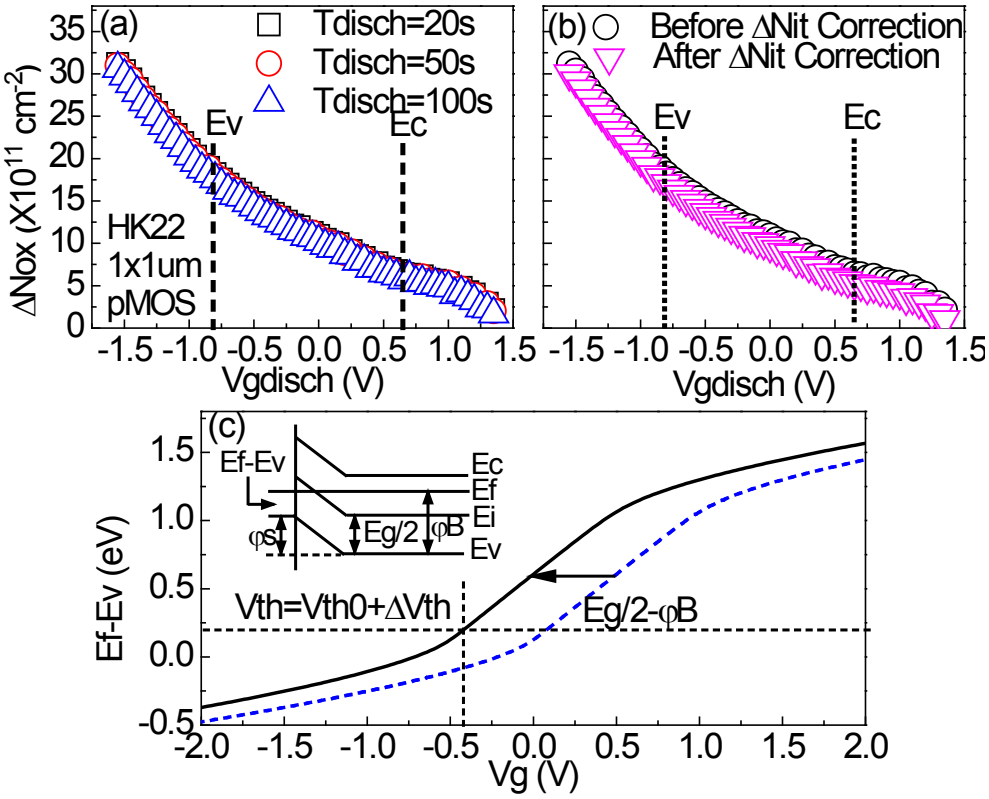

Fig. 2.25 (a) A comparison of the positive charge density ΔNox extracted at different discharging time. (b) A comparison of ΔN_{OX} before and after Nit correction. (c) Illustration of the Ef-Ev versus V_{gdisch} relation extraction. The dashed curve is the theoretical (E_f-E_v) ~ V_g, calculated from the CVC simulator**.** By aligning E_f-E_v=$E_g/2$- φ_B to the threshold voltage of stressed device, the solid line is obtained with which E_f-E_v can be found for a given V_g=V_{gdisch}. The inset of (c) shows the relationship between E_f-E_v and surface potential, φs and φ_B.

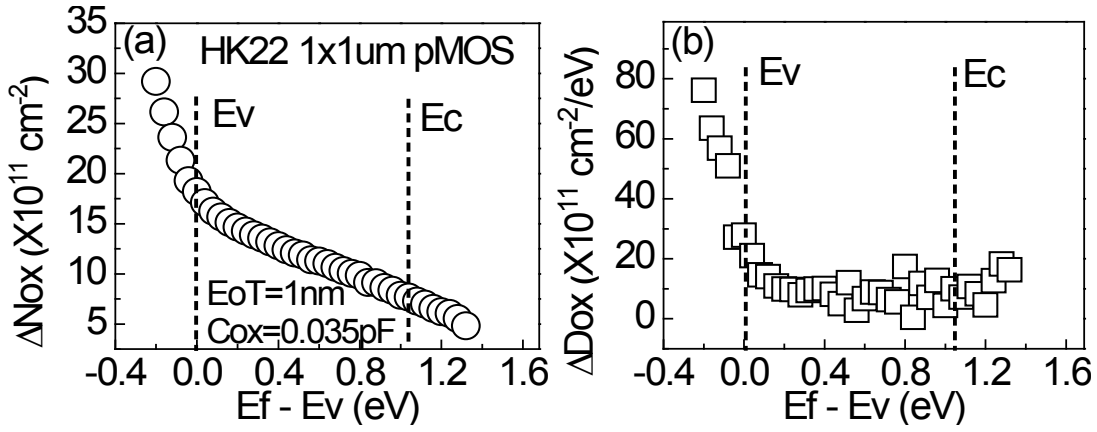

Fig. 2.26 (a) Energy profile ΔN_{OX} of the positive charges against their energy level. (b). Energy density ΔD_{OX} calculated by differentiating the data in (a).

2.6 Variability characterization method

2.6.1 Random Telegraph Noise (RTN) technique

I_d on nano-scale devices is fluctuating under a constant gate voltage due to the charging and discharging of a handful of defects in the gate dielectric. Sometimes when V_g is properly biased (normally around V_{th}), one trap will dominate the current fluctuation and an RTN signal can be observed, as shown in Fig. 2.27.

RTN on deeply scaled MOSs was observed and well modelled in 1989[84, 85]. As the scaling down of feature size, single trap impact on I_d becomes larger and larger as MOSFET's size scaling down [55]. IBM reported a single RTN can already reach device lifetime criteria in 2011 [86]. Meanwhile as the voltage scales down, BTI is much alleviated resulting in RTN playing a more and more important role in modern circuits.

RTN has three basic parameters: amplitude, capture time and emission time. RTN amplitude reflects the impact of the RTN trap, the averaged capture time (τ_c) and emission time (τ_e) can be used to extract the trap's energy information according to Equation (2.8).

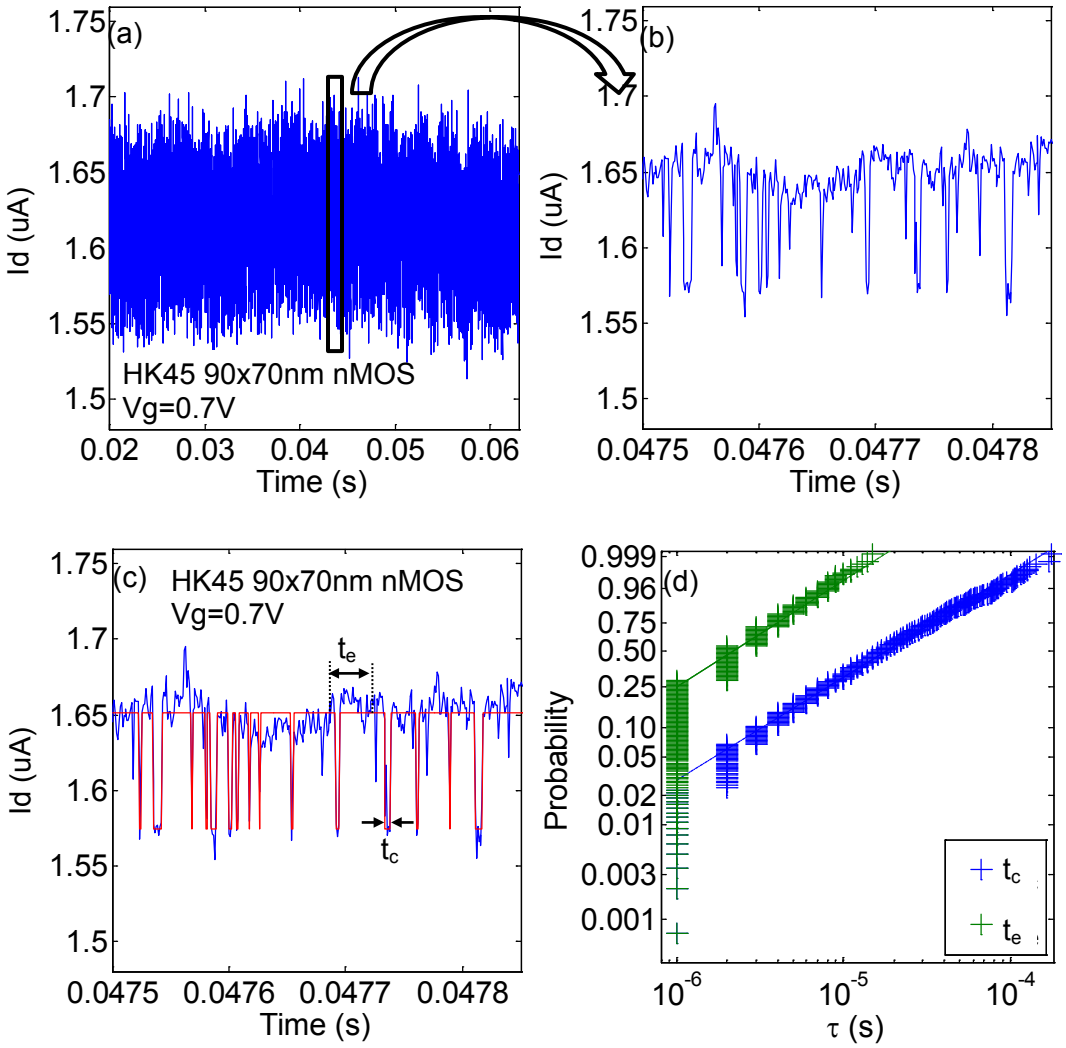

Fig. 2.27 (a) RTN measured on a HK45 90x70nm nMOS. V_g is biased at $V_g=V_{th}+0.2=0.7V$, drain is biased at constant V_d= +100mV, sampling rate=1MSa/s to capture the fast RTN signal. (b) Enlarge the data marked in black rectangle from (a) to show a clear RTN is observed. (c) RTN information: Amplitude, t_c and t_e extraction. (d) t_c and t_e follow a good exponential distribution.

$$\frac{\tau_c}{\tau_e} = \exp(\frac{E_T - E_F}{kT}) \quad (2.8)$$

Where E_T is the RTN responsible trap energy, E_F is the fermi level.

Once enough RTN traps are captured and statistically analyzed, people will be able to evaluate the behavior of these traps under actual operating conditions.

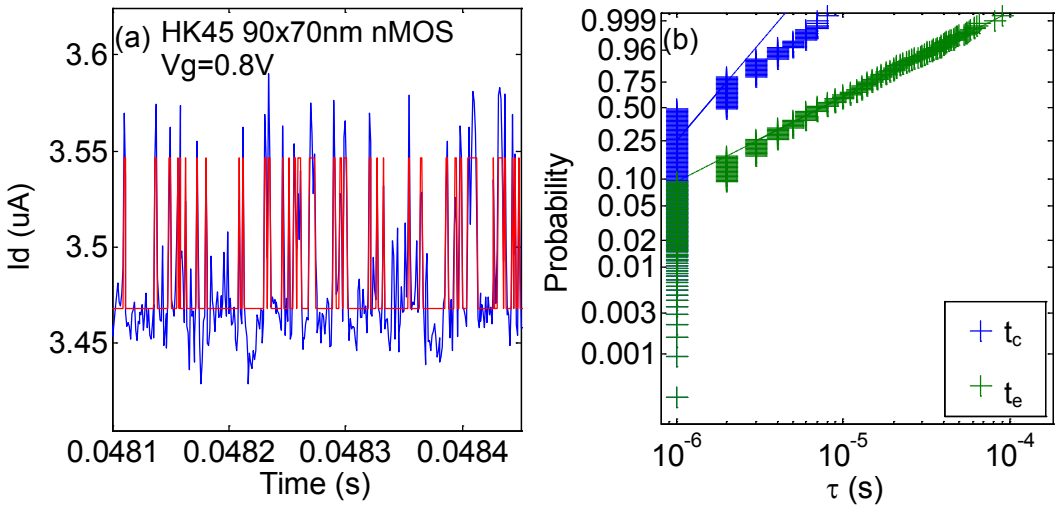

Fig. 2.28 (a) RTN measurement and analysis on the same device under V_g=0.8V, clearly t_c decrease and t_e increases under a higher V_g compared to Fig. 2.27c. (b) τ_c and τ_e under 0.8V also follow a good exponential distribution.

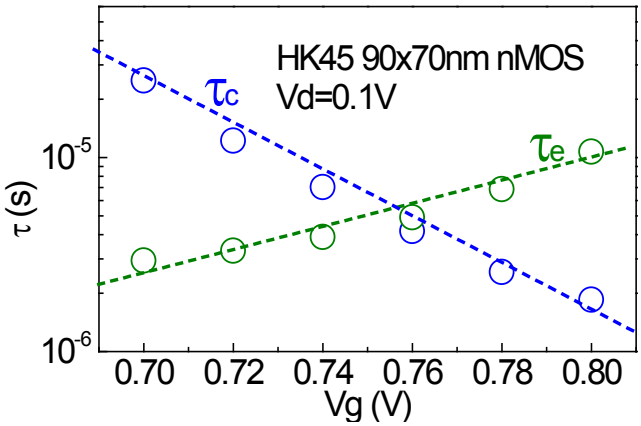

Fig. 2.29 RTN τ_c and τ_e dependence on V_g. Trap energy information can be extracted from this.

2.6.2 Time Dependent Defect Spectroscopy (TDDS) technique

T. Grasser et al [87] proposed the Time Dependent Defect Spectroscopy (TDDS) technique after in-depth study of recovery behavior on nano-scale devices.

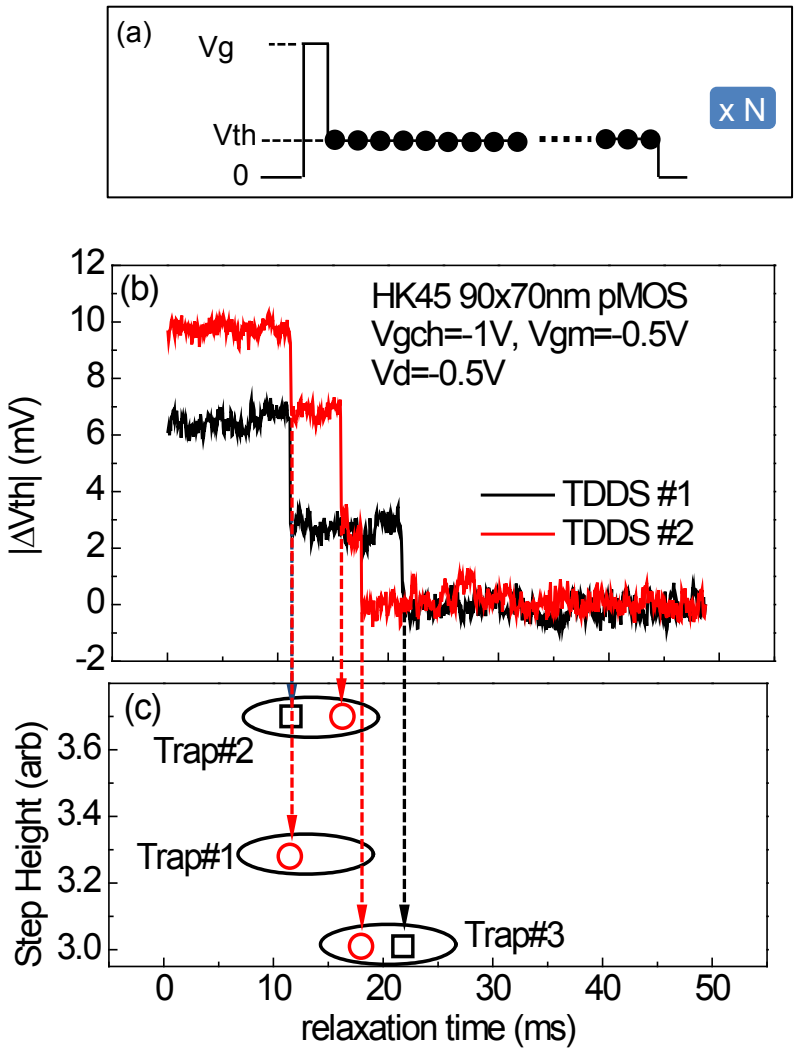

Fig. 2.30 (a) Test waveform of the Time Dependent Defect Spectroscopy (TDDS) technique. (b) Typical TDDS measurement results, each step represents a single defect emission, the amplitude and emission time of this defect is then extracted. (c) By repeating the TDDS measurements a defect spectroscopy in terms of emission time and amplitude can be extracted.

TDDS test pattern illustration is shown in Fig. 2.30a. A stress V_g is firstly applied to charge up multiple traps, charge time is very short to avoid defect generation, $|V_g|$ then lowers down close to $|V_{th}|$ and monitor the I_d recovery continuously. For some nano-scale devices the discharging traces can have clear steps, the emission times τ_e of these steps can then be captured. Finally all the traps will be discharged and then the same procedure is repeated to gain another group of τ_e, finally a defect spectroscopy can be extracted and Time Dependent Variability can be evaluated.

Compared with RTN technique, TDDS can extract multiple traps at one time, and it can also capture some traps which have long capture time thus cannot be captured by RTN. The sacrifice is that the capture time information cannot be extracted in TDDS.

2.7 Accelerated BTI parameter extraction technique

Traditional BTI degradation follows a power law, which can be described as:

$$\Delta Vth = A \cdot (|V_g - V_{th}|)^m \cdot t_{str}^n \quad (2.9)$$

According to JEDEC. Time exponent n can be extracted through Constant Voltage Stress and multiple stress voltages are used to extract A and m values, the testing time usually takes a few days. In industry, the fast BTI parameter extraction method is always preferred in lots of circumstances, like fast wafer screening, process optimization etc. Moreover, as nowadays most circuits are fabricated with nano-scale devices. Nano-scale devices need multiple devices average to get meaningful results, fast BTI parameter extraction methods

become a necessity. In this section two popular accelerated BTI parameter extraction methods are introduced: Voltage Ramp Stress and Voltage Step Stress technique.

2.7.1 Voltage Ramp Stress technique

Voltage Ramp Stress test waveform is shown in Fig. 2.31. In terms of BTI characterization method it belongs to Spot Id MSM measurement as previous discussion. Instead of a constant stress voltage and a log-incremental stress time, the Voltage Ramp Stress technique employs a linear-incremental stress voltage and a linear-incremental stress time. Since the m value in Equation (2.7) is usually much larger than n value, voltage acceleration is a much more effective way to accelerate BTI compared to stress time, thus the testing time is reduced. V_{gsense} is selected around V_{th} and measurement delay is one millisecond, $|\Delta V_{th}|$ is extracted by interpolation to the reference IV on fresh.

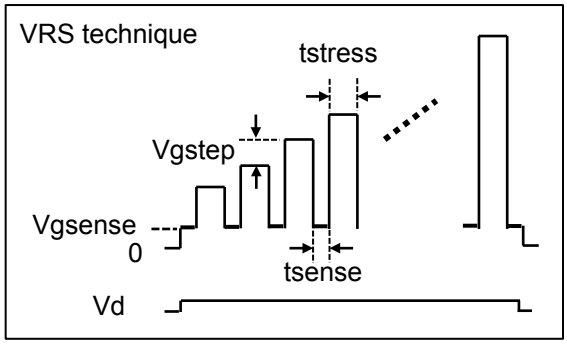

Fig. 2.31 Test waveform of the Voltage Ramp Stress (VRS) technique.

Ramp Rate (RR) in Fig. 2.31 is defined by:

$$RR = \Delta Vg/\Delta t \tag{2.10}$$

A, m and n extraction from the Voltage Ramp Stress pattern are shown as below:

$$\Delta V_{th}(T_{CVS}, V_{CVS}) = A \cdot V_{CVS}^{m} \cdot T_{CVS}^{n} \tag{2.11}$$

From RR definition:

$$V_i = RR \cdot \Delta t \cdot i \tag{2.12}$$

The i^{th} Voltage Ramp Stress phase can be equal to device stressed under V_{CVS} for effective stress time

$$\Delta t_i(V_{CVS}) = \frac{\Delta V}{RR} \cdot \left(\frac{V_i}{V_{CVS}}\right)^{m/n} \tag{2.13}$$

The total equivalent stress time under Constant Voltage Stress for the whole Voltage Ramp Stress procedure is:

$$t_{CVS}(V_{CVS}) = \int_1^N \Delta t_i(V_{CVS}) \tag{2.14}$$

Substituted Equation (2.11) into Equation (2.8),

$$\Delta V_{th}(RR, V_i) = \frac{A}{(m/n+1)^n} \cdot \frac{V_i^{m+n}}{RR^n} \tag{2.15}$$

Fig. 2.32d-f show how m and n is extracted in the Voltage Ramp Stress technique. $|\Delta V_{th}|$ against stress voltage is plotted in Fig. 2.32d, they also follow a power law before $|V_g|$ goes beyond too high (1.9V) which may generate some unintentional traps. Fitted power law will intercept V=1V at ΔV_{th} at 1V, which is

$$\Delta V_{th}(RR, 1) = \frac{A}{(m/n+1)^n} \cdot \frac{1}{RR^n} \qquad n = -\text{slope} \tag{2.16}$$

$$\frac{A}{(m/n+1)^n} = \Delta V_{th} \text{ at } 1V|_{RR=1} \tag{2.17}$$

Power exponent in (d) is m+n, as long as n is extracted, m can be determined, A then can be easily calculated by substituting m and n value into Equation (2.17).

The Voltage Ramp Stress technique is trying to capture the total degradation (1 millisecond measure delay) while the parameters are deduced from the power law. For some immature process which has a lot of As-grown traps the power law is no longer valid, resulting in the wrong parameter extraction, as detailed in chapter 5.

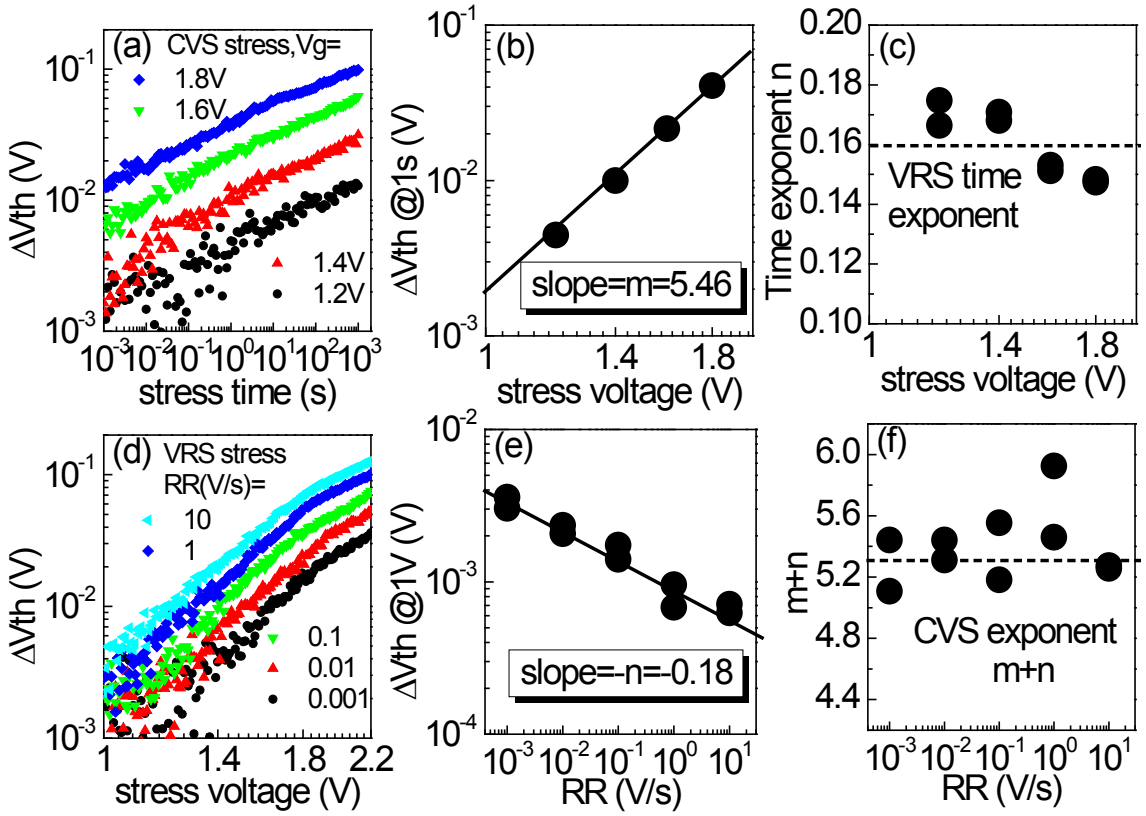

Fig. 2.32 A replot of the Voltage Ramp Stress technique results and analysis in [88] to show parameter extraction from the Voltage Ramp Stress technique. PBTI on nMOSFETs is adopted here. a-c show the test data and A,m,n extraction using conventional Constant Voltage Stress technique, d-f show the test data and parameter extraction of Voltage Ramp Stress technique. The agreement of parameters extracted from these two different techniques indicates the correctness of the Voltage Ramp Stress technique.

2.7.2 Voltage Step Stress technique

Another fast BTI parameter extraction method is Voltage Step Stress technique, as illustrated in Fig. 2.33. Prior to stress, a reference IV curve is recorded. The device is stressed at $|V_g| = |V_1|$ for a pre-specified time, Δt and then $|V_g|$ is then 'stepped up' to next $|V_g|$ for another Δt. IV is measured at pre-defined intervals during each step. This procedure will continue until reaching the pre-set maximum N.

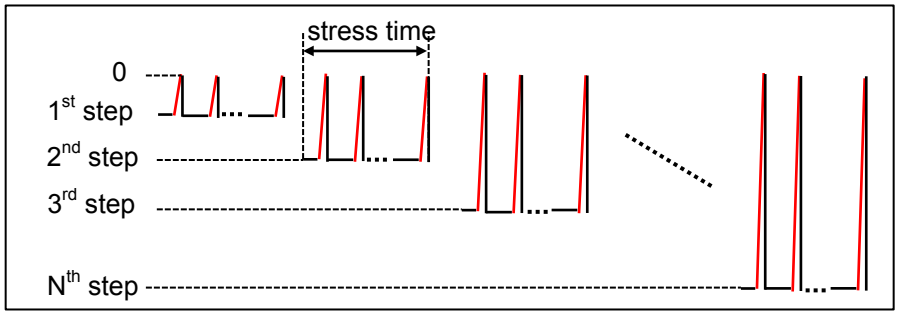

Fig. 2.33 Voltage Step Stress test waveform illustration.

Compared to the Voltage Ramp Stress technique which attempts to capture the whole $|\Delta V_{th}|$, the Voltage Step Stress method employs slow DC measurement to only capture defects which will not be discharged under 0V, thus the power law is valid for both mature and immature processes. Due to high accuracy brought by the slow DC measurement employed here, time exponent can be reliably extracted from the first $|V_1|$ stress phase, as shown in Fig. 2.34, meanwhile $A(|V_1-V_{th}|)^m$ is also achieved in the first stress phase, once m value is extracted, A can be easily calculated.

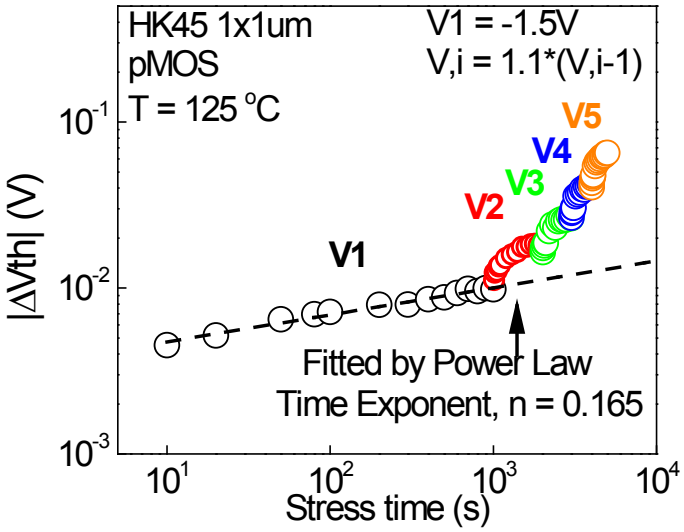

Fig. 2.34 Typical results of applying the Voltage Step Stress technique on a HK45 1x1um pMOSFET. Firstly BTI kinetics under $V_g=V_1$ is used to extract A & n, the rest data under high V_g is then used to extract m.

Stress time in the Voltage Step Stress pattern can be converted to equivalent stress time under Constant Voltage Stress at V_1, which is

$$\Delta t_{eff} = \sum_1^N \left(\frac{V_i - Vth}{V_1 - Vth}\right)^{m/n} \cdot \Delta t \qquad (2.18)$$

$|\Delta V_{th}|\sim\Delta t_{eff}$ after Equation (2.18) conversion should then follow the same trace as the first stress phase as if the devices is stressed under Constant Voltage Stress=V_1 for a very long time.

The analysis of A & m extraction from the Voltage Step Stress technique is shown in Fig. 2.35.

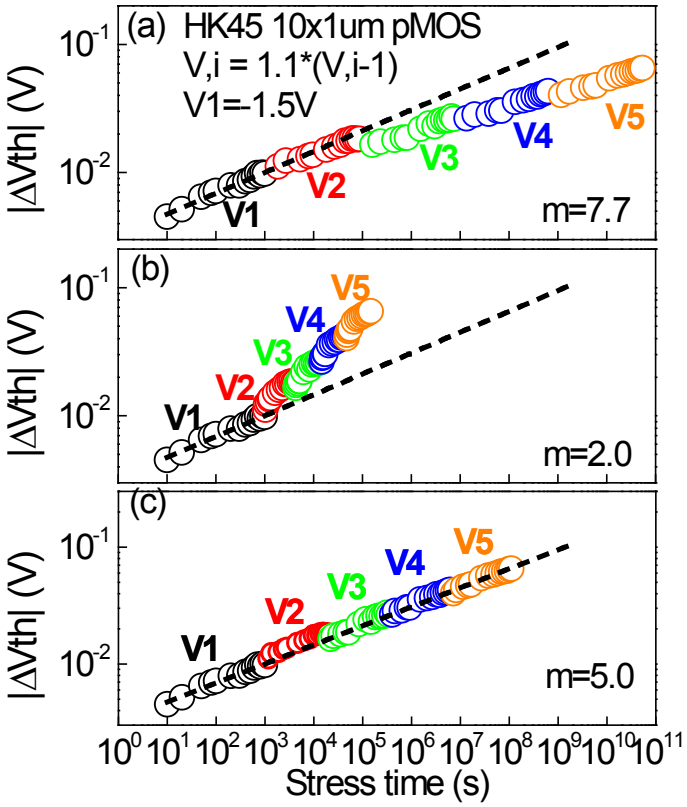

Fig. 2.35 Stress time transformation with different voltage exponent, m. Only when the correct m is used, they agree well with the power law predetermined in the first step (dashed line).

A guessed m value (any positive integer bigger than 1 is fine) is substituted into Equation (2.18) to calculate the effective stress time Δt_{eff}. Then check if $|\Delta V_{th}| \sim \Delta t_{eff}$ follows the same power law as the first stress phase. Only when the correct m is found, can the power law be restored. Fig. 2.36a-c shows three cases with m = 2, 5, 7.7 respectively. If m is too far from its real value, the transformed curve deviates from the power law extended from step #1. The correct value of m can be obtained when the best agreement is reached.

To check the convergence of the fitting and whether m extraction is unique, the least square error between measured $|\Delta V_{th}|$ and $A(|V_1 - V_{th}|)^m \Delta t_{eff}^n$ is given in Fig. 2.36,

obviously the fitting is convergent and m is unique. As soon as m is extracted, together with $A(|V_1 - V_{th}|)^m$ value from the first stress phase, A is also extracted.

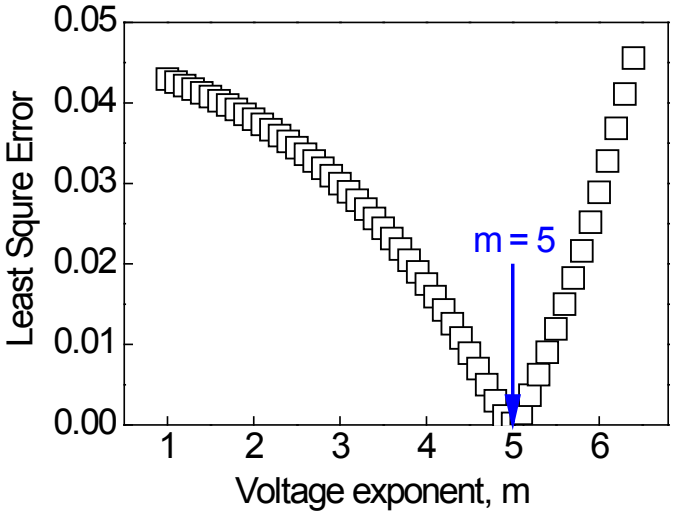

Fig. 2.36 Least Square Error between the transformed curve (the points in Fig. 2.35) and the dashed line extrapolated from the first step (dash line in Fig. 2.35).

2.8 Summary

In this chapter, mainstream methodologies for the reliability and variability of BTI characterization are discussed. Due to the fast recovery and also fast charging, slow DC measurement will result in a big underestimation of BTI degradation. Advanced fast measurement and OTF measurement are able to capture the majority of BTI degradation but the degradation needs to be converted to $|\Delta V_{th}|$ sensing at a constant surface potential, namely a constant current.

Energy profile is an effective way to understand BTI behavior. Discharging-based Multiple Pulses (DMP) technique can be used to achieve this purpose, but the waveform is

complicated and not easy to transfer to a mainstream analyzer like the Keithley SCS4200. The conventional DMP technique has been modified to be accommodated into the Keithley 4200. $\Delta Id/Id_0 \sim |\Delta V_{th}|$ conversion is used to extract $|\Delta V_{th}|$ to ensure the degradation is sensing at a constant surface potential.

Time Dependent Variation brings an extra challenge to BTI characterization. On one hand the current is fluctuating and people normally use Random Telegraph Noise (RTN) or Time Dependent Defect Spectroscopy (TDDS) to characterize it. On the other hand, Device-to-Device Variability tests make the fast BTI parameter extraction method a necessity.

Two popular fast methods, Voltage Ramp Stress and Voltage Step Stress are briefly introduced. Both methods use the power law to accelerate the parameter extraction.

3 Defect Generation under BTI stress

3.1 Introduction & Motivation

Defect generation under BTI can be traced back as early as 1977[11], K. O. Jeppson et al suggested a hydrogen-diffusion controlled interface state creation is responsible for such degradation.

With the development of measurement facilities, recovery within microseconds can be measured. Publications about NBTI recovery mushroomed in the past two decades. Now it is widely accepted that NBTI consists of a recoverable component "R" and a so-called "permanent" component "P" which is difficult to recover. However, due to the logarithmic discharging behavior of NBTI, there is no clear boundary of "R" and "P", making it very difficult to study "R" and "P" respectively and understand their underlying mechanisms.

In this chapter, defect generation is investigated by experimentally separating it according to its test behavior. The definition for Generated Defects is: Defects which do not exist on a fresh device. Test results show Generated Defects have two components depending on the measurement condition, one can repetitively charge/discharge, another one is very difficult to discharge even under a positive bias. The absolute amount of these 2 types of defects is related to the measurement condition, but if they were added together and form

the entire Generated Defects, the total amount is a constant value independent of measurement condition.

3.2 Generated Defects under NBTI stress

3.2.1 Energy profile of Generated Defects: Cyclic Positive Charges (CPC) and Anti-Neutralization Positive Charges (ANPC)

Based on the Discharging-based Multiple Pulses (DMP) technique introduced in chapter 2, defect energy profile is extracted on a fresh and then stressed device as shown in Fig. 3.1. Generated Defects, by its definition, can be extracted by subtracting the fresh from the stressed energy profile (Fig. 3.1b). -1.2V is selected as DMP charging voltage to avoid any potential generation on fresh device, which is confirmed in Fig. 3.1a that $|\Delta V_{th}|$ return to zero at the end of DMP, the last discharging V_g, $V_{gdisch,end}$=+1.6V. The device is then stressed by a heavy stress $V_{gstress}$=-2.5V for 1,000 seconds, after that the same DMP procedure is repeated on the stressed device.

From the energy profile point of view, Fig. 3.1b clearly shows Generated Defects are mainly located both within the Si bandgap and beyond Si conduction band. As observed in Fig. 3.1a, charging-discharging can be cycled by alternating V_g polarity for the defects within the bandgap, so that they are referred to as Cyclic Positive Charges (CPC) [89]. On the other hand, the defects above Si Ec are more difficult to neutralize and they are called Anti-Neutralization Positive Charges (ANPC).

Note in Fig. 3.1 the defect profile is in $|\Delta V_{th}|$ instead of the density. The annotation in Fig. 3.1b "CPC" and "ANPC" actually represents "CPC induced $|\Delta V_{th}|$" and "ANPC induced $|\Delta V_{th}|$". Considering the lifetime prediction only requires the $|\Delta V_{th}|$ value, and also for the convenience of discussion, "CPC" and "ANPC" hereafter mean their induced $|\Delta V_{th}|$ instead of the defects themselves hereafter. This applies for all the defects including: Generated Defects, As-grown Traps, Energy Alternating Defects, Pre_Existing defects.

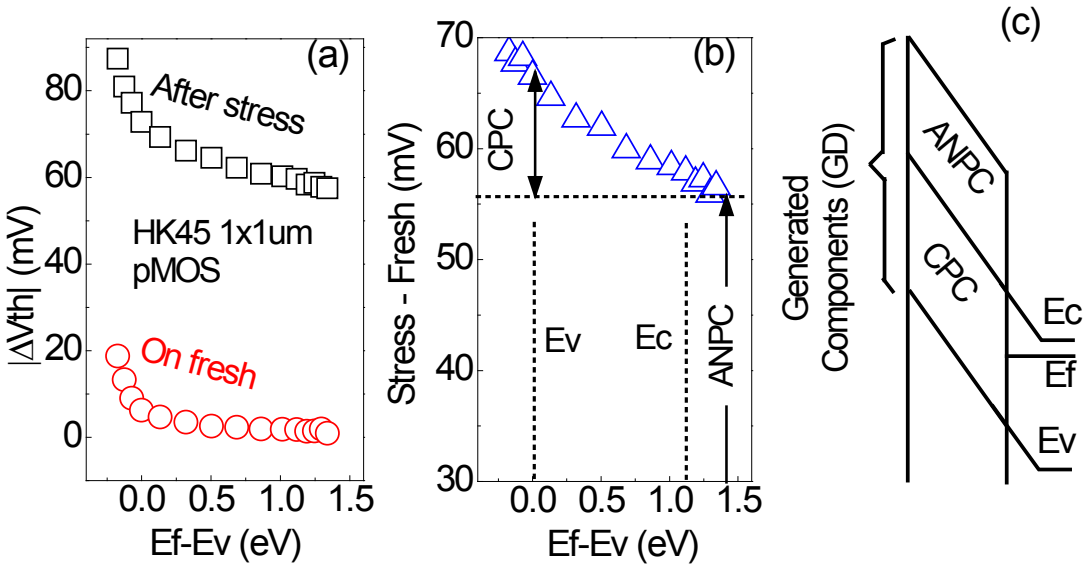

Fig. 3.1 (a) A comparison of the energy profiles before and after stress, $|\Delta V_{th}|$ impact on energy level E_f-E_v has already been taken into consideration. (b) By subtracting the fresh profile from the stressed one, the profile of Generated Defects was extracted. (c) Illustration of the energy range for the Cyclic Positive Charges (CPC) and the Anti-Neutralization Positive Charges (ANPC). Note the illustration is just a "rule-of-thumb" here for convenience, Ec and Ev is not the precise boundary to separate CPC and ANPC from energy location point of view.

3.2.2 Characterization method of CPC and ANPC

As the definition in 3.3.1, unlike ANPC that can be measured after a certain delay time under a certain discharge V_g due to its "Anti-Neutralization" nature, CPC cannot be directly measured due to: 1). CPC shared the energy profile with part of the Pre_Existing defects (as detailed in chapter 4); 2). CPC's repetitive charging discharging behavior is the same as Pre_Existing defects. Thus again, subtraction is used to extract the CPC components.

Actually due to the non-saturation of ANPC discharging (Fig. 3.2) kinetics, CPC and ANPC here are "pragmatic" definitions. Any defect which cannot be discharged under +1.6V 10s will be ascribed to ANPC.

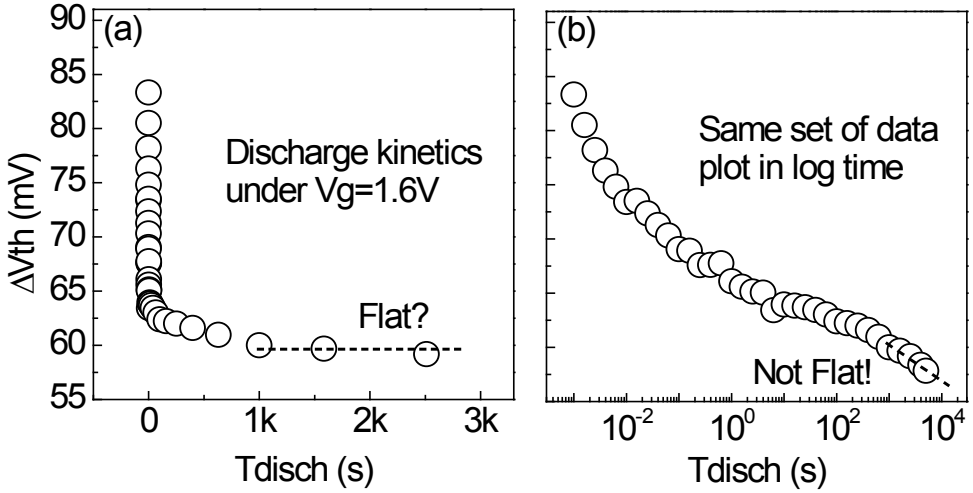

Fig. 3.2 Discharging kinetics under positive gate voltage in (a) linear time scale and (b) log time scale. Clearly the discharging won't become flat although it looks so in linear time plot in (a).

-1.2V/+1.6V is applied alternatively on a fresh device and then stressed device for CPC extraction. The device is already heavily stressed under -2.5V 1ks. Due to the power law

intrinsic of NBTI further defects generation is suppressed. -1.2V charging on this stressed device is a purely filling procedure. Since the Pre_Existing defects' charging kinetics is independent of stress (as detailed in chapter 4), CPC charging kinetics can be easily extracted by:

$$\mathrm{CPC} = |\Delta V_{\mathrm{th}}| - \mathrm{Pre\_Ex} - \mathrm{ANPC} \tag{3.1}$$

Where Pre_Existing defects' charging kinetics is measured on fresh device and ANPC is calculated by subtracting the fresh $|V_{th0}|$ from local V_{th} measured after stress, as shown by the gap in Fig. 3.3b.

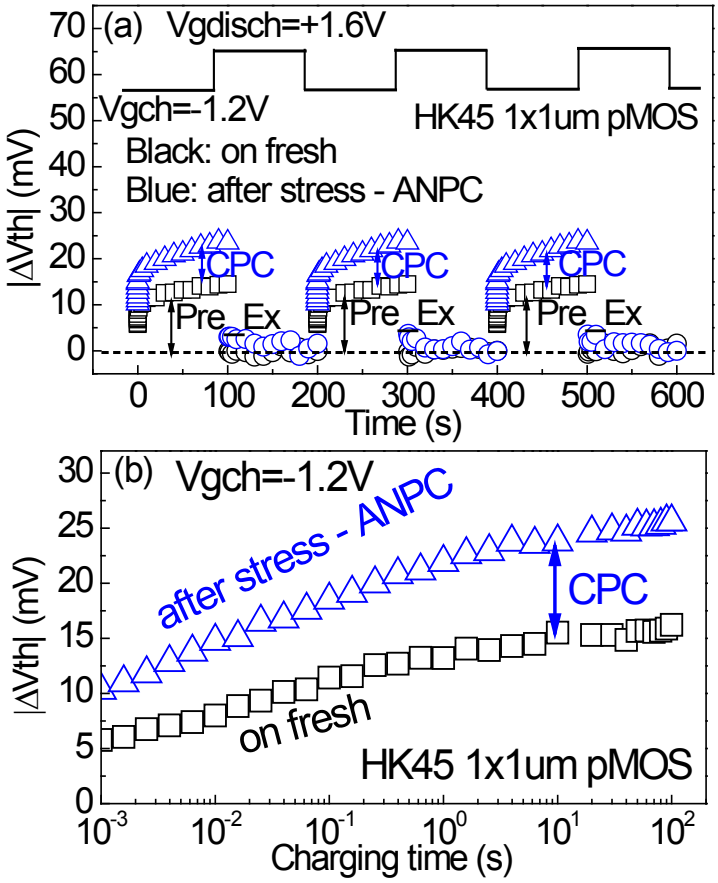

Fig. 3.3 (a) Test waveform and typical results for CPC extraction. Note ANPC is already subtracted as a reference for the charging kinetics on stressed device. (b) The averaged kinetics of CPC extracted from (a) in log time scale.

Clearly under -1.2V, CPC charging will saturate within a short period (10s in Fig. 3.4). If a higher charging voltage -1.4V is applied, which is still much lower than the stress voltage -2.5V, the total amount of CPC will saturate at the same level as -1.2V, but at an earlier saturation time. This can be understood by the fact that the filling efficiency increases due to the higher current.

It is interesting to point out that the CPC can saturate after such short times under low voltage because they have been generated already. This means CPC have two different processes: first, they have to be generated/activated from their precursors; second, after the generation, CPC have a filling process shown in Fig. 3.4.

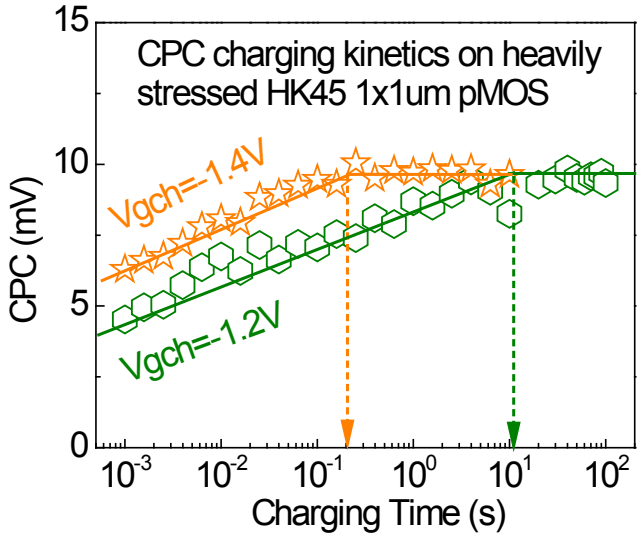

Fig. 3.4 Voltage dependence of CPC charging kinetics on stressed device.

The repetitive charging and discharging behavior and non-existence on fresh device clearly prove CPC are Generated Defects; however, ANPC cannot be proved in this way as it is difficult to be neutralized at the end of the recovery period, due to the non-saturation of ANPC discharging kinetics (Fig. 3.2).

To support that ANPC is also part of Generated Defects, the activation energy of CPC and ANPC are compared. Fig. 3.5 shows the same stress voltage applied for 1ks under 4 different temperatures. The activation energy, Ea', for both CPC and ANPC, can be easily extracted from the slope in Fig. 3.5a. It turns out Ea' for both CPC and ANPC are around 0.14eV, much larger than the Ea' of ~0.04eV reported for Pre_Existing defects [90], supporting the hypothesis that defect generation is a different process from filling As-grown traps. And also the extracted Ea' of 0.14eV agrees well with the value reported in early generation-related works [91, 92]. What is worth noting is that in physics, the activation energy varies with different individual defect, therefore, the extracted effective activation energy represents the average effect on temperature activation.

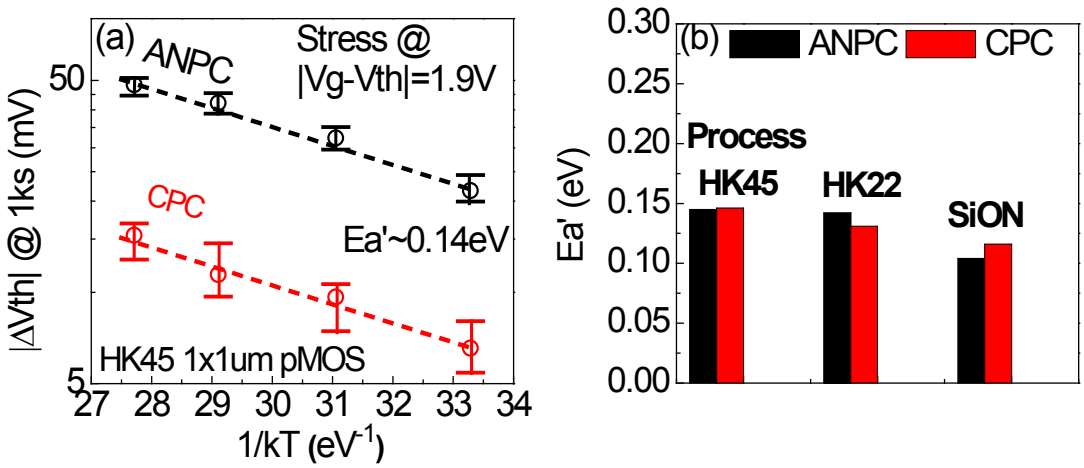

Fig. 3.5 (a) The effective activation energy extraction for CPC and ANPC on HK45 1x1um pMOSFET after stress under DC $|V_g-V_{th}|$=1.9V for 1ks. Wherein, ANPC here is measured after +1.6V discharge 30s to remove all the non-GD component. After the ANPC measurement, V_{grech}=-1.2V is applied to get CPC amount. 5 devices are used for each temperature and the error bar is given. The corresponding activation energy, Ea', can be obtained by Ea'/n = 0.7eV, where n is the time exponent (n=0.2 as shown in Fig. 3.14). (b) The extracted Ea' from another two processes: HK22 & SiON show similar results as HK45.

3.2.3 Speculation of the origin of CPC and ANPC

Although the electronic measurements above do not give direct evidence for the physical origin of CPC and ANPC, a speculation based on the current results and past work [89, 93] can be made here. Fig. 3.5 shows that CPC and ANPC have the same thermal activation. Based on the common thermal activation of CPC and ANPC, on one hand, one may speculate that they have the same chemical structure and a variation of properties, such as bond length/angle, which in amorphous oxides results in a distribution of energy levels. On the other hand, one may also speculate that they are different products of a common controlling aging process. Hydrogenous species are likely involved in the generation of both CPC and ANPC. For example, an increase of hydrogen exposure of the sample will significantly enhance the efficiency of CPC generation [89]. And in some samples under development, substantial CPC Pre-exist on fresh devices [93], possibly because of high exposure to hydrogenous species during fabrication.

3.3 ANPC and CPC generation kinetics

3.3.1 Novel Stress-Discharge-Recharge technique

In previous sections the CPC and ANPC components can be extracted respectively on a stressed device, but the generation kinetics is still unknown. A novel Stress-Discharge-Recharge test procedure is designed to investigate the kinetics of the entire Generated Defects.

Stress-Discharge-Recharge test procedure is shown in Fig. 3.6. After stressing under $V_{gstress}$, an opposite polarity of discharging V_g (V_{gdisch}) was applied to accelerate the discharging of Pre_Existing defects, as used in early works [94, 95]. The low use-bias under real operation V_{gch} ($|V_{gch}| << |V_{gstress}|$) was then applied to refill all the traps that contribute under use condition. The pulse I_d-V_g curves are recorded during each step with a measurement time of three microseconds. The threshold voltage degradation is monitored by sensing at a constant Id of 500nA·W/L around threshold voltage.

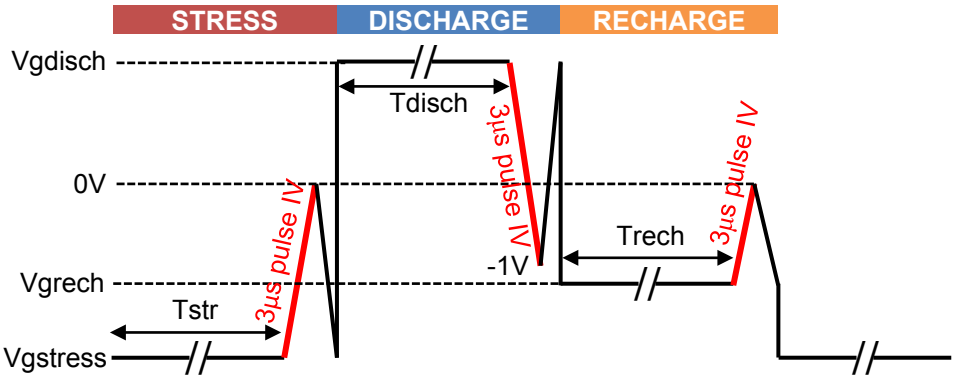

Fig. 3.6 The test follows a Stress-Discharge-Recharge sequence. At the end of each step, $|\Delta V_{th}|$ was monitored from a corresponding IV, which was taken from the 3µs pulse edge with V_d = -0.1V applied. $|\Delta V_{th}|$ was extracted at a constant current of 500nA·W/L.

To reliably extract the CPC and ANPC generation kinetics, the test parameters for the Stress-Discharge-Recharge waveform should be carefully chosen. First of all, V_{gdisch} should be high enough to remove all the Pre_Existing defects, meanwhile it cannot be too high to introduce any electron traps on a fresh device. Secondly recharge voltage (V_{grech}) should be able to fill all the CPC components generated by the stress phase within a relatively short (no longer than ten seconds) time, otherwise the testing time will be too long as this recharge time (T_{rech}) is needed in every stress phase.

Before applying the stress voltage to generate defects, Pre_Existing defects amount under V_{gch} needs to be measured, which later will be subtracted as a reference to get CPC on stressed device. This can be done be applying the Stress-Discharge-Recharge pattern with $V_{gstress}=V_{grech}$ for once. Although the Stress-Discharge-Recharge waveform is more complicated compared with how people normally measure the "permanent" component by introducing a delay at zero volt, the advantage of this pattern is it can give us a reliable time exponent of Generated Defects regardless of measurement condition, stress voltage, temperature, and frequency on different process, as detailed in section 3.5.

Fig. 3.7 is a typical Stress-Discharge-Recharge pattern result with. Clearly both ANPC and CPC follow the power law, with the same time exponent 0.2, which further supports their generation intrinsic. Note ANPC is seven times higher than CPC on HK45 process, the ratio is similar for the other three processes used in this chapter. This ratio will be used in chapter 4 when measuring the kinetics of Pre_Existing defects.

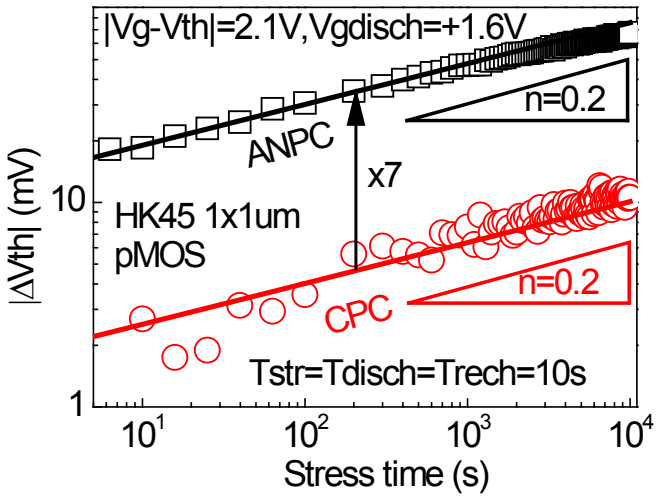

Fig. 3.7 Both ANPC and CPC follow power law with the same power exponent, supporting they belong to the same type of defect.

3.3.2 Contribution between interface states and Oxide traps

Note Anti-Neutralization Positive Charges (ANPC) contain the interface states, as interface states will not discharge. Interestingly, when the interface states generation kinetics measured by the charging pumping technique is subtracted from the kinetics of Generated Defects, a good 1:1 relationship is achieved between the interface states and the remaining generated oxide traps (Fig. 3.8). This well agrees with the reports in early works [4, 11, 96]. Noted in V. Huard's work, gate current is used to characterize the NBTI degradation, early works show Pre_Existing defects charging/discharging will have negligible impact on gate current, it is speculated that Generated Defects which caused gate current increase. This is another evidence supporting that, Generated Defects, as a whole, is one type of defects, both CPC and ANPC are just two different facets in terms of a specific measurement condition.

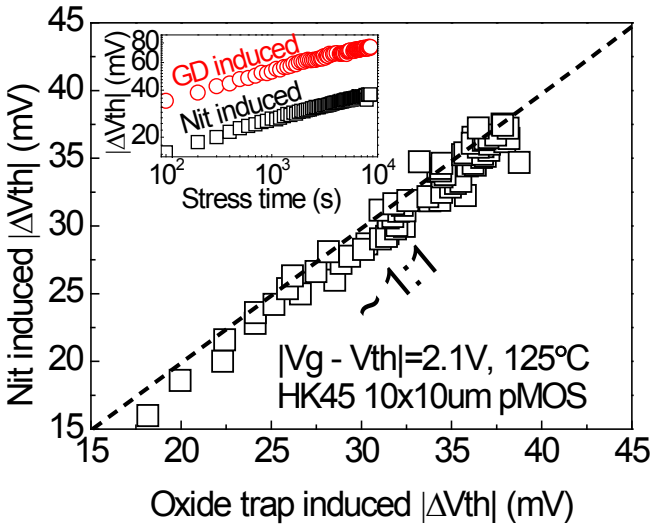

Fig. 3.8 Interface states generation versus oxide trap generation. Generated Defects are measured using the Stress-Discharge-Recharge technique on a HK45 10x10um pMOSFET, while interface states are measures by charge pumping on another HK45 10x10um pMOSFET.

3.4 Reliable NBTI Generated Defects' time exponent extraction

3.4.1 Existing values of NBTI Generated Defects' time exponent

Power-law based extrapolation by JEDEC is the standard method for lifetime prediction of commercial circuits. With the introduction of the fast measurement into BTI characterization, since the whole NBTI is captured, the power law relationship is no longer valid for most of the state-of-the-art technique devices and circuits. To restore the power law, a common practice is to introduce a delay between the stress and measurement [97, 98], during which NBTI partially recovers [41, 62, 75, 98-101]. However, even if the test data can be fitted well with a power law after such a delay, the time exponent, n, depends on measurement condition (discharging voltage, measurement speed etc.) [98, 102], as shown in Fig. 3.9.

Due to the intrinsic nature of power law, a small change in time exponent will lead to unacceptable errors when extrapolated to 10 years. Fig. 3.10 shows for time exponent n changes from 0.16 to 0.2, the predicted lifetime can have a significant uncertainty as big as sixteen times. Thus there is a strong need to find the reliable and accurate time exponent of n for NBTI lifetime prediction.

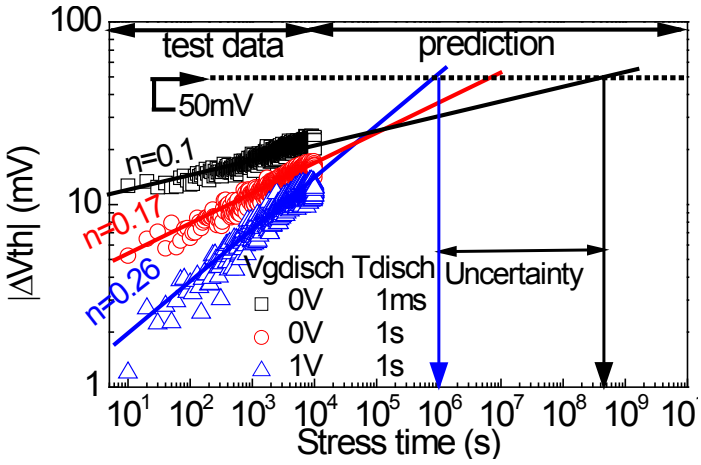

Fig. 3.9 NBTI measured by conventional method under different discharge V_g (V_{gdisch}) and time (T_{disch}). Different time exponents lead to substantial uncertainty in lifetime extraction when extrapolating to 50mV. The stress was under $|V_g-V_{th}|$= 1.2V.

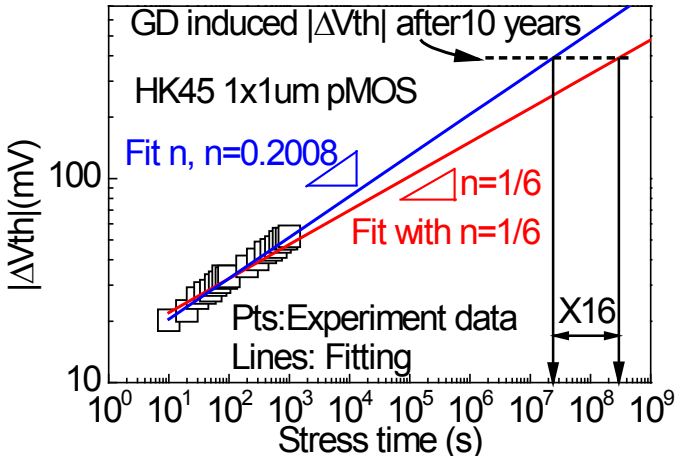

Fig. 3.10 Fit the test data with/without assuming the n=1/6, respectively, and then extrapolated to 10 years. Lifetime will be overestimated by 16 times if n is wrongly extracted as 1/6.

3.4.2 Independent time exponent of Generated Defects

The root of the confusion about the NBTI time exponent lies in ambiguous separation of different kinds of defects. Most people now agree that NBTI consists of a Recoverable (R) and a so-called Permanent (P) component, but due to the non-saturation of discharging behavior, P is very difficult to measure. In terms of the A-G model point of view, P is the

same as ANPC. Fig. 3.9 shows if different V_{gdisch} and T_{disch} are applied, ANPC will be different, resulting in the big change on the time exponent.

As discussed in previous sections, ANPC is just part of Generated Defects, there are also CPC generated on stressed device which can repetitively charge/discharge. The same activation energy (Fig. 3.5) and time exponent (Fig. 3.7) strongly support the hypothesis that they belong to the same type of defects.

Note in previous sections the definition of Cylic Positive Charges (CPC) and Anti-Neutralization Positive Charges (ANPC) is pragmatic. To get a reliable time exponent, the impact of discharge condition on CPC and ANPC needs to be clarified.

In Fig. 3.11 different T_{disch} (a&b) and V_{gdisch} (d&e) are applied on a heavily stressed device to evaluate their impact on ANPC (a&c) & CPC (b&e). ANPC are measured right after the given discharge condition, after that -1.2V 10s is applied to refill and extract the CPC. Note ANPC measured right before this -1.2V 10s are used as the reference in Equation (4.1).

Interestingly, once both CPC and ANPC are included, Generated Defects become independent of the measurement conditions, i.e. the discharge time and voltage. This lays a solid foundation to extract the accurate time exponent of Generated Defects for reliable life time prediction when extrapolated to 10 years.

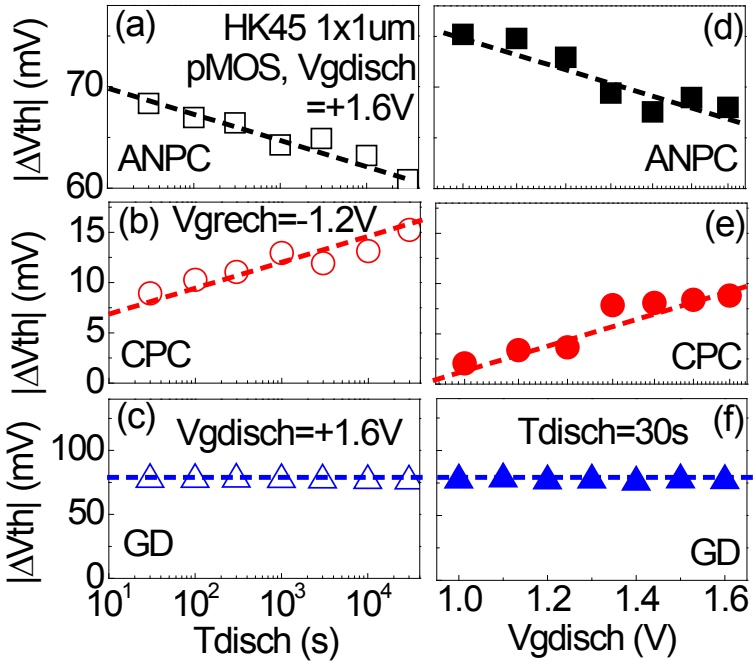

Fig. 3.11 Dependence on discharge conditions for (a&d) ANPC, (b&e) CPC and (c&f) GD=ANPC+CPC. V_{gdisch}= +1.6V for (a-c), T_{disch}=30s for (d-f).

3.4.3 Constant V_g and Constant E_{ox} stress

Although BTI is considered an electric-field driven phenomenon [103], the tests are usually performed under constant V_g throughout the entire test. The underlying assumption is that the total degradation, ΔV_{th}, is much smaller than the applied voltage and thus the electric field on the dielectric will not be disturbed. However, to investigate the time exponent usually long time stress needs to be carried out on the device, the impact of this effect will accumulate and make a big difference. Fig. 3.13 compared Generated Defects under the constant V_g and constant E_{ox} condition, wherein, constant E_{ox} is approximately achieved by consistently increasing stress V_g, by the amount of ΔV_{th},

which is measured in the last step (Fig. 3.12). Unless specified, Generated Defects components discussed in this chapter are using constant E_{ox} stress.

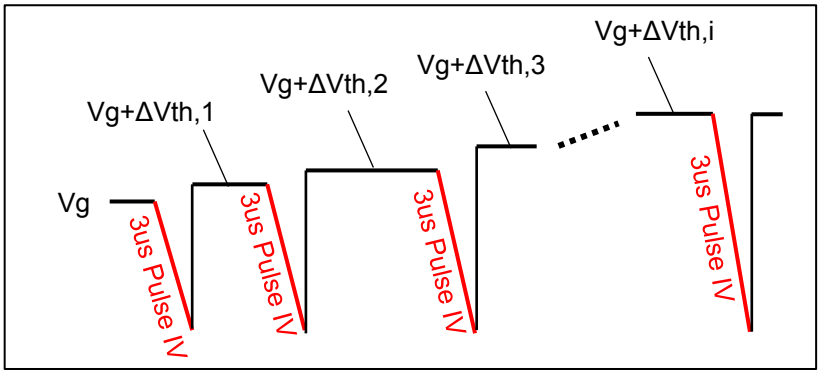

Fig. 3.12 Illustration of the on-site constant E_{ox} correction. Once a pulse IV is measured, $|\Delta V_{th}|$ can be extracted instantly and then added to the initial $V_{gstress}$ to calculate the next stress voltage, note the previous stress voltage is still applied during the pulse IV data saving and analysis.

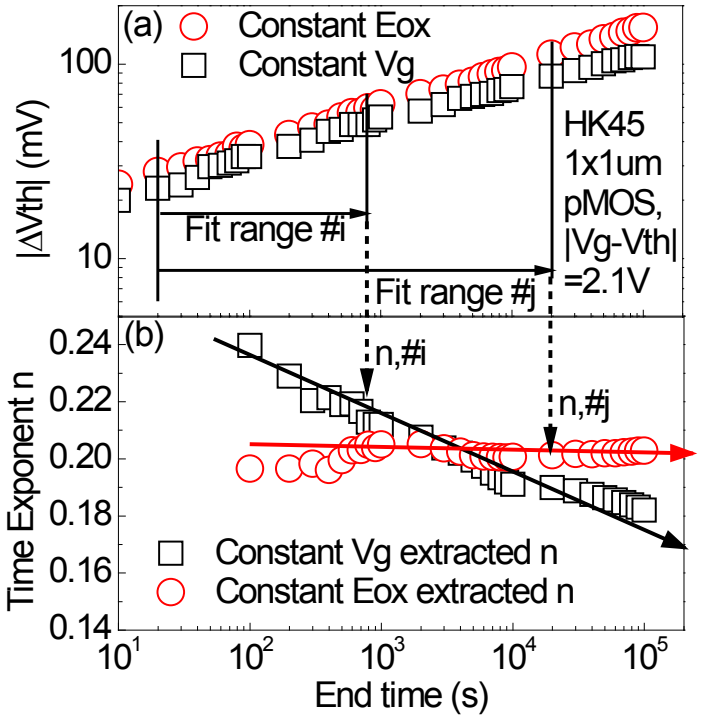

Fig. 3.13 (a) Kinetics of Generated Defects measured under constant V_g and E_{ox} NBTI stress. (b) The extracted time exponent n using data between 10s and variable end-time. n varies little after 10^3s for constant Eox stress, in which the overdrive $|V_{gstress}-V_{th}|$ is kept constant during the test. If the test is carried out under constant V_g stress, a gradual reduction can be observed.

Similar to early works [37, 104, 105] reported, the constant V_g results also show a reduction of n as stress time increases caused Eox deduction, while the constant-Eox results follow a good power law in the whole time domain. It should be noted that this independent n from long stress time is achieved under 125°C and Generated Defects are within 200mV, if the temperature is high and the stress is heavy, device might be running out of defect-precursors, meanwhile the simultaneous annealing effect at the high temperature [96, 104-106] may also contribute to the n reduction.

3.4.4 Voltage dependence under different temperature

With the reliable time exponent by capturing the whole GD components under constant-Eox stress, GD voltage dependence can be easily extracted under 125°C (Fig. 3.14a) and room temperature (Fig. 3.14b). The independent time exponent of GD n=0.2 regardless of temperature further supports their generation property.

The NBTI kinetics given by JEDEC is:

$$GD = g \cdot \exp(Ea'/k_B T)(|V_g - V_{th})^m t^n \quad (3.2)$$

E_a' is the activation energy of Generated Defects and can be extracted from Fig. 3.5, and g_0, m, and n can be easily extracted from Fig. 3.14. Equation (3.2) was then used to predict Generated Defects under low stress conditions and Fig. 3.15 shows that the prediction agrees well with the experiment data. It should be pointed out the measured data in Fig. 3.15 themselves were not used for extracting the parameters in Equation (3.2).

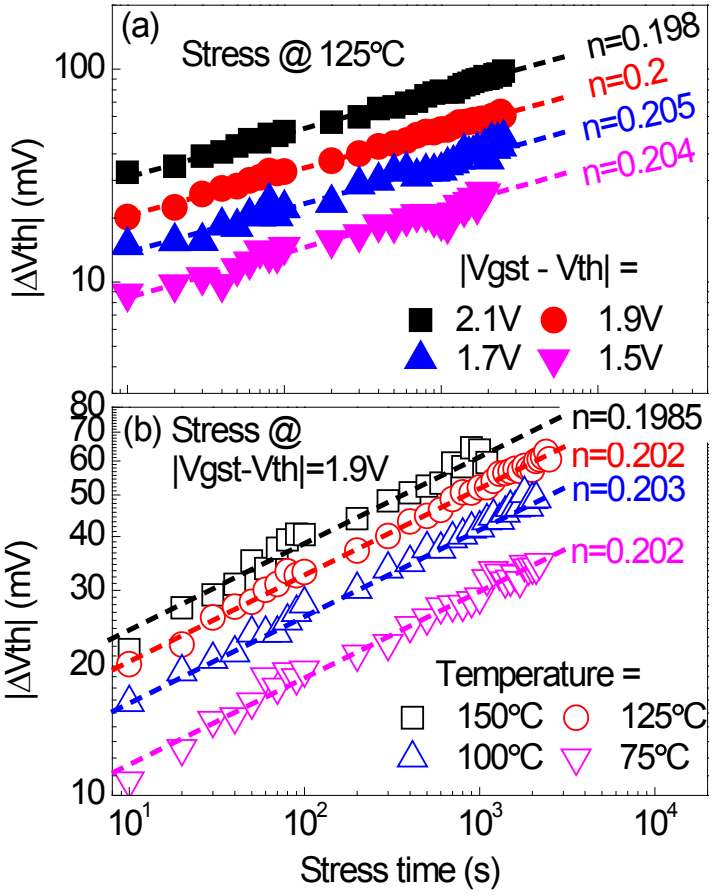

Fig. 3.14 Kinetics of Generated Defects under different (a) stress voltages and (b) temperatures. The time exponents equal to 0.2 for all circumstances.

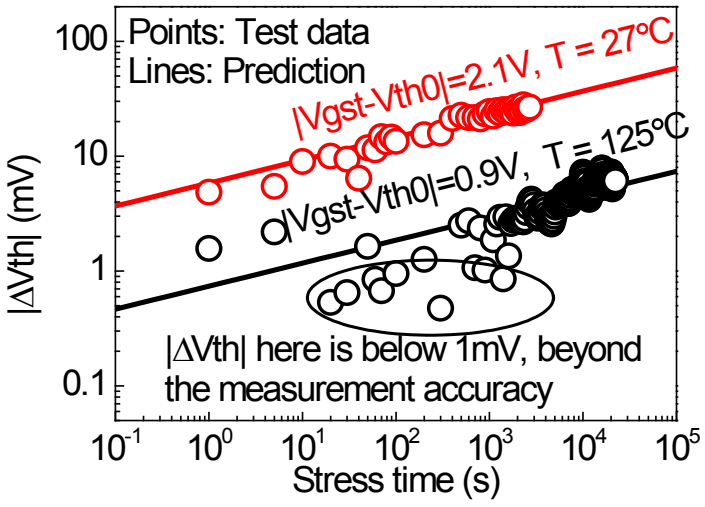

Fig. 3.15 Verification of Generated Defects calculated with Equation (3.2). Note the test data is not used for parameter extraction.

3.4.5 Generated Defects under AC NBTI stress

In most of the commercial circuits, CMOS operates under AC conditions. To predict the device lifetime under AC stress, Generated Defects under AC stress conditions need to be modelled.

To compare with DC, AC Stress-Discharge-Recharge waveform is developed by replacing the stress phase with Freq=10kHz, Duty Factor=0.5 AC stress, as shown in Fig. 3.16.

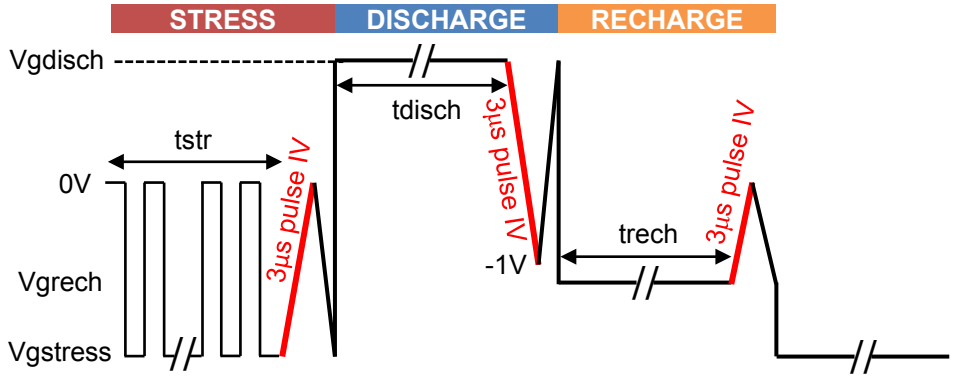

Fig. 3.16 Test waveform of the AC Stress-Discharge-Recharge technique.

Fig. 3.17 compared the DC and AC Stress-Discharge-Recharge measured Generated Defects results under the same $|V_{gstress}-V_{th}|$, which clearly shows Generated Defects are mainly driven by the Equivalent stress time [33] (i.e. stress time * duty factor). Later in the next section, Process Dependence, it is again shown that the activation energy Ea' (Fig. 3.19) for CPC and ANPC, voltage exponent m (Fig. 3.20) is also independent of DC or AC stress.

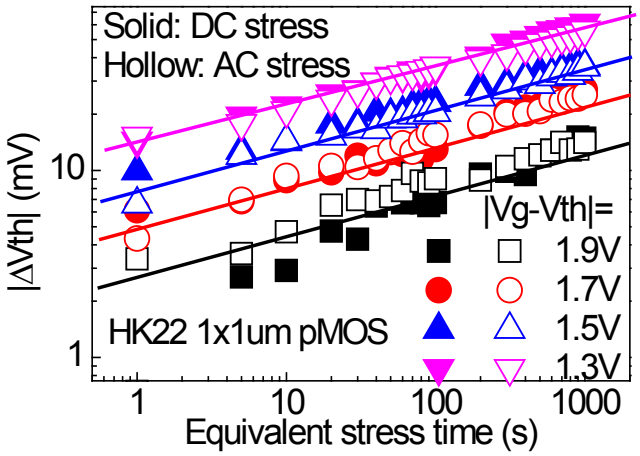

Fig. 3.17 Generated Defects measured under DC and AC Stress-Discharge-Recharge method agrees very well. Equivalent stress time for AC is calculated by stress time*duty factor.

3.4.6 Process dependence

NBTI on different processes can have significant discrepancies, for example, the time exponent dependence on different processes, as shown by the time exponent n summary from early works in Fig. 3.18. However, in terms of Generated Defects, n shows a much smaller variation of process dependence, as shown by the measured n results on four different processes in Fig. 3.18.

Moreover, on HK45, HK22 and SiON process, Activation energy Ea' (Fig. 3.19) and $|V_g-V_{th}|$ exponent m (Fig. 3.20) extracted under DC & AC stress have the same values, further support that Generated Defects are driven by an equivalent stress time.

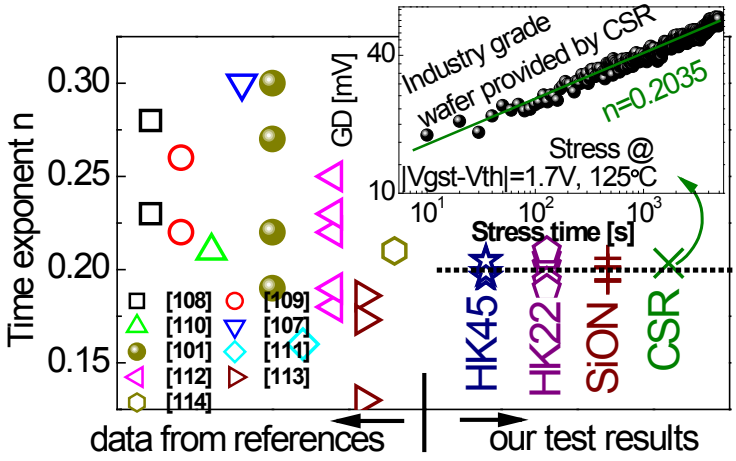

Fig. 3.18 A comparison of NBTI time exponent reported by early works [107-114] with Generated Defects' time exponent values extracted using the SDR method. Multiple points for each process represent the values from different stress voltage and temperature. The inset shows results on the CSR sample.

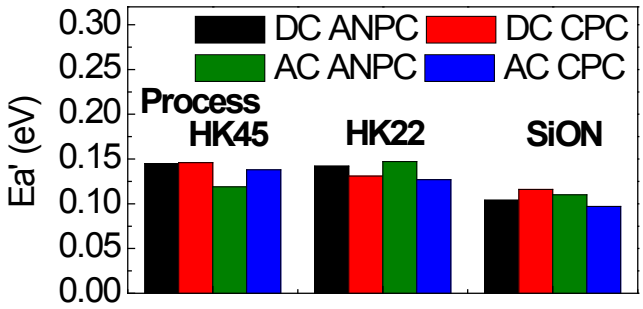

Fig. 3.19 A summary of the activation energy Ea' of ANPC and CPC under DC/AC NBTI stress condition on 3 different processes.

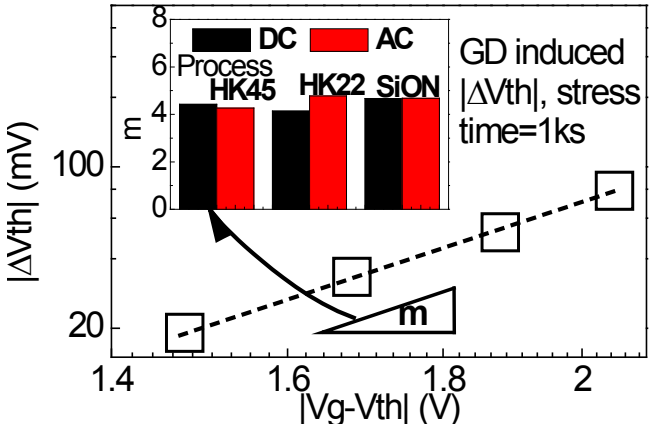

Fig. 3.20 $|V_g-V_{th}|$ exponent m extraction. The data points ('□') is taken from Fig. 3.14a. The inset shows the extracted m value for DC and AC stress conditions across 3 different processes.

3.5 Generated Defects under PBTI stress

Existence of defect generation under PBTI is actually much less debatable compared to NBTI. Similar to NBTI, experiment data shows PBTI on nMOS also have two components in Generated Defects: Cyclic Electron Traps (CET) and Anti-Neutralization Electron Traps (ANET). Again ANET time exponent is measurement condition related (Fig. 3.21a), by adding the CET part, the entire Generated Defects under PBTI illustrates an independent time exponent n=0.32 (Fig. 3.21b). It's worth noting a similar time exponent n=0.3 is also observed by IBM [115], which is using gate current to characterize the PBTI degradation, and normally people believe Stress Induced Gate Current (SILC) is caused by defects generation. Due to the intrinsic of power law, the much larger PBTI time exponent will make PBTI increase much more quickly compared to NBTI as time approaches the device lifetime, as detailed in the next section.

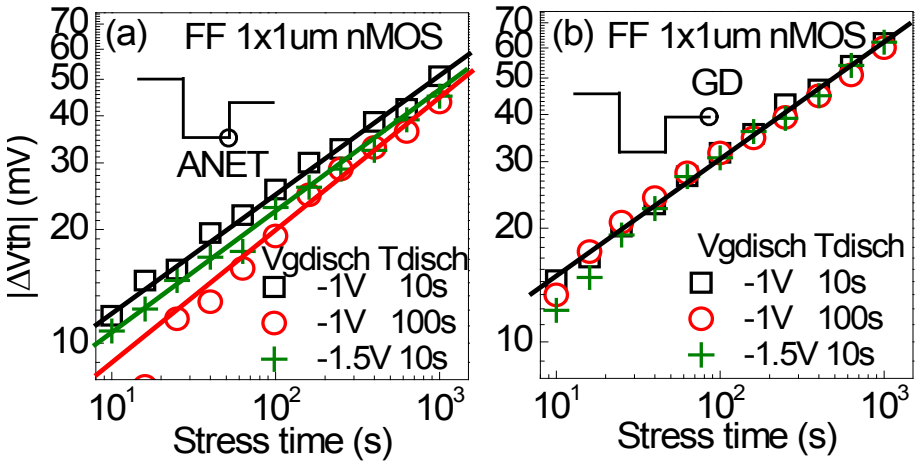

Fig. 3.21 Typical results of Stress-Discharge-Recharge technique on nMOSFETs. The degradation is monitored by measuring pulse IV in 3us. It is clearly shown that Anti-Neutralization Electron Traps (ANET) vary with different $|V_{gdisch}|$ and T_{disch} (a) while the entire Generated Defects are independent of measurement condition (b).

3.6 Evaluation of contribution from NBTI/PBTI Generated Defects at device lifetime

PBTI attracted much less attention in the past decades compared to NBTI due to the much smaller degradation within a feasible experiment time domain. The uncertainty of NBTI time exponent obscures the contribution of N/PBTI at device lifetime. By applying the newly developed Stress-Discharge-Recharge method, both N/PBTI defects generation time exponent can be reliably extracted independent of measurement condition, thus laying a solid foundation to evaluate the impact of N/PBTI defect generation at device lifetime.

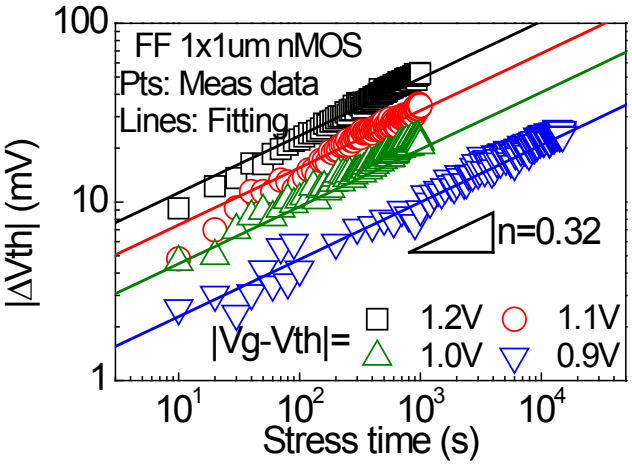

Fig. 3.22 Generated Defects under different $|V_g-V_{th}|$ PBTI stress also follow a power law but with a larger much time exponent n=0.32. Tests are following the Stress-Discharge-Recharge procedure, V_{gdisch}=-1V, V_{grech}=0.5V.

Fig. 3.23 shows the comparison between the calculated Generated Defects under the same oxide electric field E_{ox}=5MV/cm N/PBTI stress. From Generated Defects point of view,

within a feasible test time (<10^7s), PBTI is far smaller compared to NBTI, but due to the much larger time exponent, PBTI eventually play a more important role than NBTI.

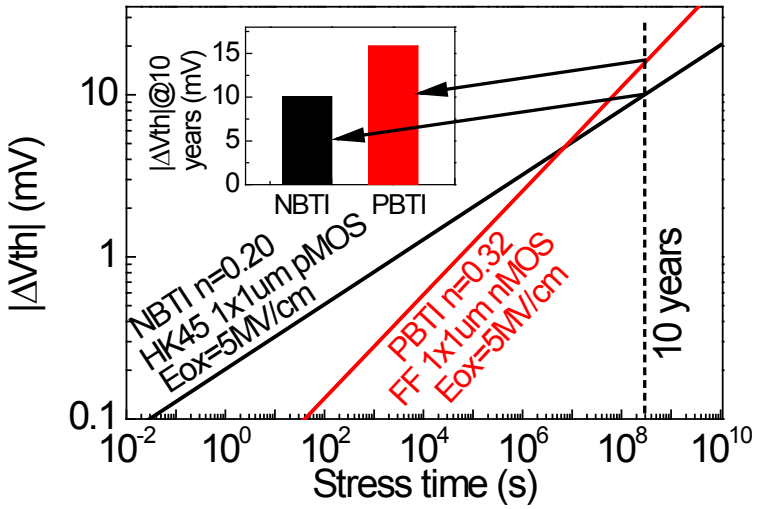

Fig. 3.23 Comparison of the kinetics of Generated Defects under oxide electric field E_{ox}=5MV/cm, T=125°C N/PBTI stress. The inset compares Generated Defects under N/PBTI after 10 years.

3.7 Summary

In this chapter, defect generation under both NBTI and PBTI are investigated in detail from different aspects.

NBTI is taken as the example. Firstly it is clarified that Generated Defects contain two categories of defects from the energy profile point of view: Cyclic Positive Charges (CPC) and Anti-Neutralization-Positive-Charges (ANPC). CPC are clearly generated while ANPC's anti-neutralization property makes it difficult to investigate. The activation energy of CPC and ANPC extracted from different temperatures equals, and is much

higher than Pre_Existing defects reported by early work, strongly supporting that both CPC and ANPC are generated.

A new Stress-Discharge-Recharge method was proposed study the kinetics of Generated Defects. By adding CPC and ANPC together, it is shown that the entire Generated Defects are independent of measurement conditions. In terms of the time exponent of Generated Defects, very similar n~0.2 is achieved across the four different processes, significantly bridging the spread of n reported in the early NBTI works. Based on the reliably extracted time exponent, voltage, temperature and process dependence is studied and they can be well described by a power law, NBTI under low stress 125^{o}C and 35^{o}C can be accurately predicted, validating the accuracy of the extracted time exponent.

PBTI Generated Defects on nMOS is also measured with the Stress-Discharge-Recharge method. Time exponent n=0.32 is achieved regardless of discharge conditions. The much larger n in PBTI raises an alert for its reliability at device lifetime, highly likely its impact is underestimated, PBTI Generated Defects might be even larger at device lifetime compared with NBTI.

4 Pre_Existing defects under BTI stress

4.1 Introduction & Motivation

Most people now agree that NBTI contains two components: A "Permanent" component "P", whose underlying mechanism is still under debate, and another recoverable component "R", which Pre-Existed as the device is fabricated. In chapter 3 Generated Defects have been discussed in great details. From the definition in chapter 3 obviously the "P" part people normally measured is Anti-Neutralization Positive Charges (ANPC), while another component in Generated Defects, Cyclic Positive Charge (CPC)'s charging/discharging behavior is similar to "R" after being generated. To differ from other papers [105], and also as an echo to Generated Defects, "Pre_Existing defects" is used instead of "R" to represent defects which Pre-existed on a fresh device. Obviously, by definition,

$$\mathrm{BTI} = \mathrm{GD\ (Generated\ Defects)} + \mathrm{Pre\_Ex(Pre\_Existing\ defects)} \qquad (4.1)$$

Equation (4.1) is used to build the A-G model in chapter 5.

Although NBTI recoverable component has already been studied in great detail [36, 74, 116, 117], it's still of great importance to investigate Pre_Existing defects' properties. Firstly Pre_Existing defects are needed to build up the complete A-G model to predict the lifetime. Secondly the underlying mechanism of Pre_Existing defects is still under debate. For example IMEC used to consider there is only one type of Pre_Existing defects but

now they have to take two or even more types of Pre_Existing defects into consideration to fit the experiment data [118]. Ambiguous understanding of Pre_Existing defects makes it difficult to understand nano-scale devices behavior and predict its lifetime. As circuits operating voltage lowers down Pre_Existing defects become more and more important in nano-scale devices' reliability, currently there is still no feasible model which can predict the lifetime on nano-scale devices within a tolerable testing time, one of the reasons lies in the lack of understanding on Pre_Existing defects, none of the existing models which work on big devices can be extended to nano-scaled devices by just doing multiple average. In chapter 6 the A-G model is extended to nano-scale devices and good predicting capability is achieved, which is due to the correct understanding of different types of defects.

4.2 Two components in Pre_Existing defects

Experiment data indicates Pre_Existing defects have two components: As-grown Traps, whose energy level will not change after charging/discharging; Energy Alternating Defects, whose energy will alternate while charging/discharging as the name suggests.

4.2.1 Energy profile evidence for two components in Pre_Existing defects

Fig. 4.1 shows the defect profile extracted from charge up and DMP. The DMP procedure has already been introduced in chapter 2. The charge up pattern, is the inverse direction of

DMP procedure, namely DMP $|V_g|$ is stepping down while charge up $|V_g|$ is stepping up. Charging time is ten seconds for each voltage during charge up, which equals to the discharging time during DMP. Generated Defects are negligible as $|\Delta V_{th}|$ lowers down to zero under positive V_{gdisch}=1.6V, indicating all the defects in Fig. 4.1 belong to Pre_Existing defects.

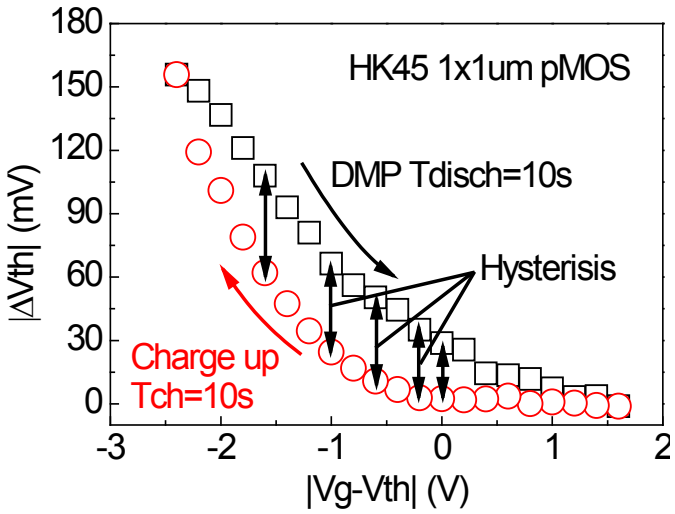

Fig. 4.1 Pre_Existing defects' profile extraction from charge up and DMP pattern. A big hysteresis is observed.

As-grown Traps have been investigated in great details in past works [78, 119], they can charge/discharge rapidly under a constant $|V_g|$ and cause current fluctuation on nano-scale devices [120], indicating its energy level is not changing during charging/discharging. Normally within one second As-grown Traps will be fully charged and reach saturation when the charging voltage ($|V_{gch}|$) is applied, or fully discharge once $|V_{gch}|$ is removed. As-grown Traps alone cannot explain the big hysteresis in Fig. 4.1. Clearly there are a certain number of traps whose energy is alternating. For example, when the discharging voltage ($|V_{gdisch}|$) equals to zero in the DMP trace ('□') there are still ~30 mV $|\Delta V_{th}|$ cannot be

discharged after ten seconds, indicating their energy located below $E_f@V_g$=0V, if this is their original energy position they should be able to charge up under $|V_{gch}|$=0V, but none of defects are charged under this condition in the charge up trace ('o'), the only explanation is these defects' energy is pulled down after they are charged, these defects are named Energy Alternating Defects.

4.2.2 Principle for multi-DMP (m-DMP) technique

To avoid the distortion from Generated Defects, Pre_Existing defects' kinetics is normally measured on stressed device. Fig. 4.2 shows the typical charging kinetics of Pre_Existing defects. A fresh device is firstly stressed under a heavy stress $|V_g-V_{th}|$=1.9V for 10s to prohibit Generated Defects' distortion during later Pre_Existing defects' charging, then a constant $|V_g-V_{th}|$=1.3V is applied to measure Pre_Existing defects' charging kinetics. To confirm there is no Generated Defects' distortion on the same device a second $|V_g-V_{th}|$=1.9V stress is then re-applied for 10ks, charging kinetics under $|V_g-V_{th}|$=1.3V is then re-measured, the perfect agreement indicates Generated Defects are excluded in Fig. 4.2.

Pre_Existing defects' charging in Fig. 4.2 keeps rising as time evolves in log scale, a turning point is observed around one hundred milliseconds charging time, Early works [78, 121, 122] show charging kinetics in Fig. 4.2 is a sum of a fast saturated As-grown Traps' charging and a power law Energy Alternating Defects' charging.

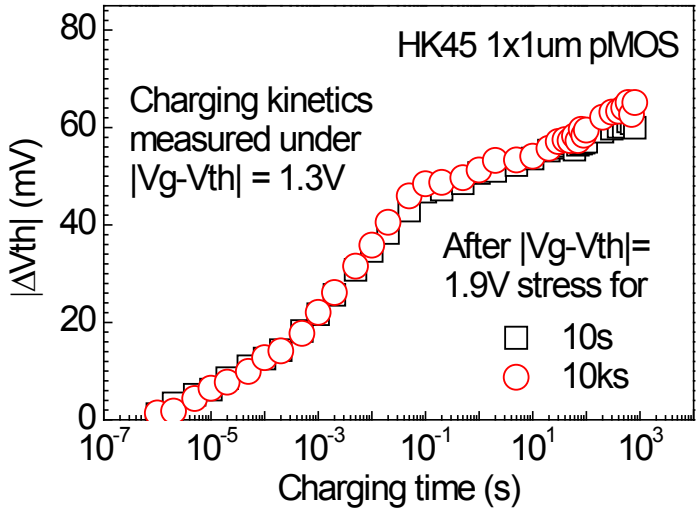

Fig. 4.2 Charging kinetics under the same $|V_g-V_{th}|$=1.3V after stressing under $|V_g-V_{th}|$=1.9V for 10s and 10ks. Stressed Vth is used as reference so both kinetics starts from 0. The perfect agreement indicates the charging here is purely the filling of Pre_Existing defects, which is independent of stress time.

Z. Ji et al in 2015 [122] proposed a multi-DMP (m-DMP) technique to extract the profile of As-grown Trap. The technique is based on the different behavior of trap energy during charging and discharging, as shown in Fig. 4.3. After 100 seconds charging under $|V_{gch}|$, all the chargeable As-grown Traps and Energy Alternating Defects under $|V_{gch}|$ are charged. After being charged, As-grown Traps remain at the same energy level while Energy Alternating Defects' energy alternates to a lower energy level. Thus in the DMP measurement, As-grown Traps will be discharged once $|V_{gdisch}|$ lowers down, while Energy Alternating Defects cannot be discharged under the same $|V_{gdisch}|$ because its energy level has alternated. Moreover, alternated Energy Alternating Defects will distort As-grown Traps' profile which locates at the same energy level as the alternated Energy Alternating Defects.

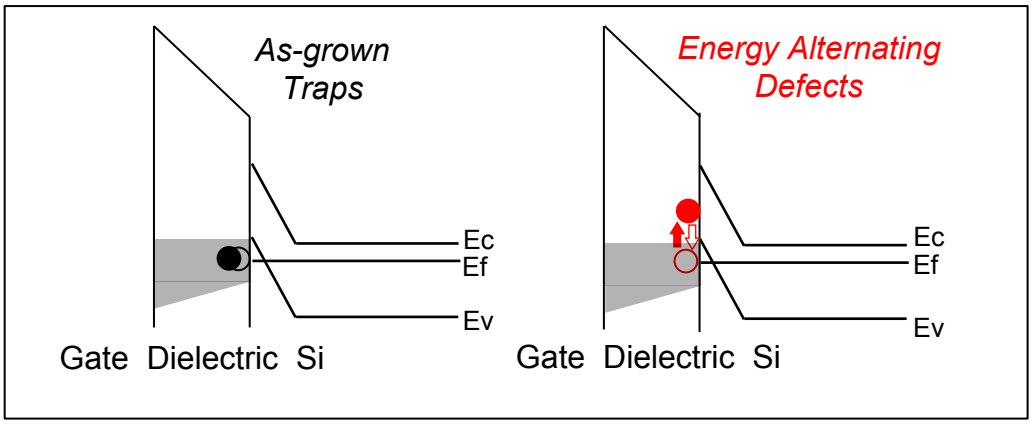

Fig. 4.3 Illustration of two components in Pre_Existing defects: As-grown Traps and Energy Alternating Defects. Charged As-grown Traps will be discharged once $|V_g|$ lowers down, while charged Energy Alternating Defects cannot be discharged as its energy is alternating. For details about the DMP principle refer to section 2.5.

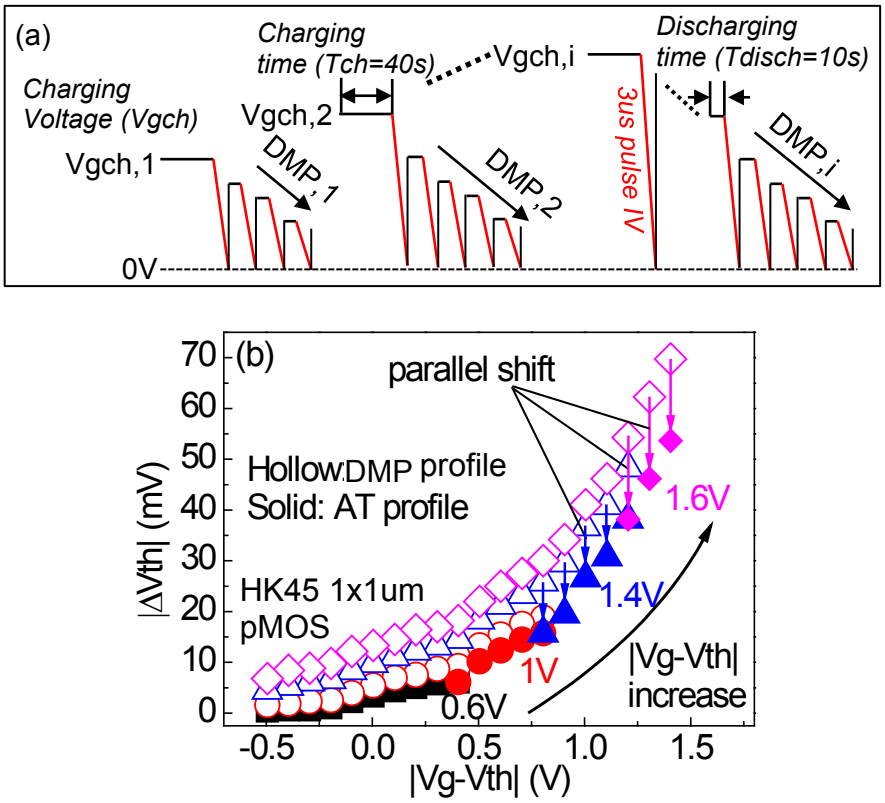

Fig. 4.4 (a) Illustration of the m-DMP test waveform. (b) Illustration of As-grown Traps' profile extraction using the m-DMP technique.

m-DMP, as the name suggests, consists of multiple DMP tests. Test waveform of m-DMP technique is shown in Fig. 4.4a. The charging voltage, V_{gch}, in the first DMP, is set close to V_{th}. Due to $V_{gch,1}$ is too low, no traps will be filled thus the DMP profile is zero. Use the first DMP profile as the start of As-grown Traps' profile. V_{gch} then will increase by a ΔV to $V_{gch,2}$. In the second DMP profile, the initial dischargeable defects, from $V_{gch,2}$ to $V_{gch,1}$, is As-grown Traps. This initial $|\Delta V_{th}|$ drop in the second DMP profile, then will append to As-grown Traps' profile. Repeat the same analysis, append the initial $|\Delta V_{th}|$ drop from i^{th} DMP profile to As-grown Traps' profile. Finally the entire As-grown Traps' profile can be extracted as the solid points in Fig. 4.4(b). Extracted As-grown Traps' profile on HK45 1x1um pMOSFET can be fitted by Equation (4.2):

$$AT = p_1 \cdot \exp[p_2 \cdot (|V_g - V_{th}|)] \tag{4.2}$$

4.2.3 As-grown Traps and Energy Alternating Defects separation

After As-grown Traps' profile is extracted from the m-DMP technique, the saturated level in Pre_Existing defects' charging kinetics can be calculated by interpolating $|V_g\text{-}V_{th}|$ to As-grown Traps' profile. As-grown Traps saturates within one second, by subtracting the saturated As-grown Traps from over one second Pre_Existing defects' charging, Energy Alternating Defects' kinetics is extracted and it follows a power law. Energy Alternating Defects under different $|V_g\text{-}V_{th}|$ shows they can be well fitted by

$$\Delta V_{th\_EAD} = A_2 \cdot (|V_g - V_{th}|)^{m_2} \cdot t^{n_2} \tag{4.3}$$

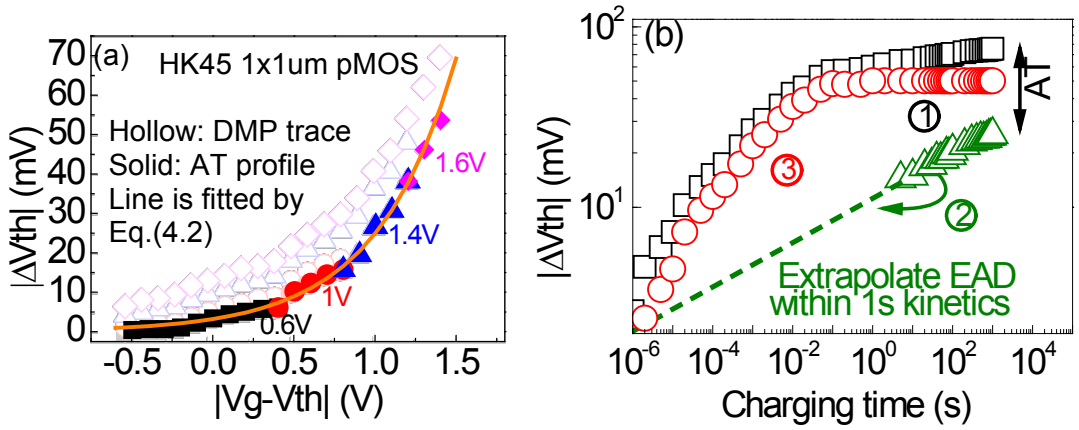

Fig. 4.5 Charging kinetics extraction of two types of Pre_Existing defects. (a) Extract the profile of As-grown Traps (AT) using m-DMP technique. (b) The kinetics of Energy Alternating Defects (EAD) ('Δ') is obtained by subtracting the saturated AT in (a) from the total Pre_Existing trap over 1s charging kinetics ('□'). AT kinetics ('o') is then obtained by subtracting Energy Alternating Defects' power law kinetics (dashed line) from total ('□').

Although Equation (4.3) is exactly the same as Generated Defects' kinetics, both time exponent and $|V_g-V_{th}|$ exponent is different from Generated Defects thus Energy Alternating Defects must be separated for modelling.

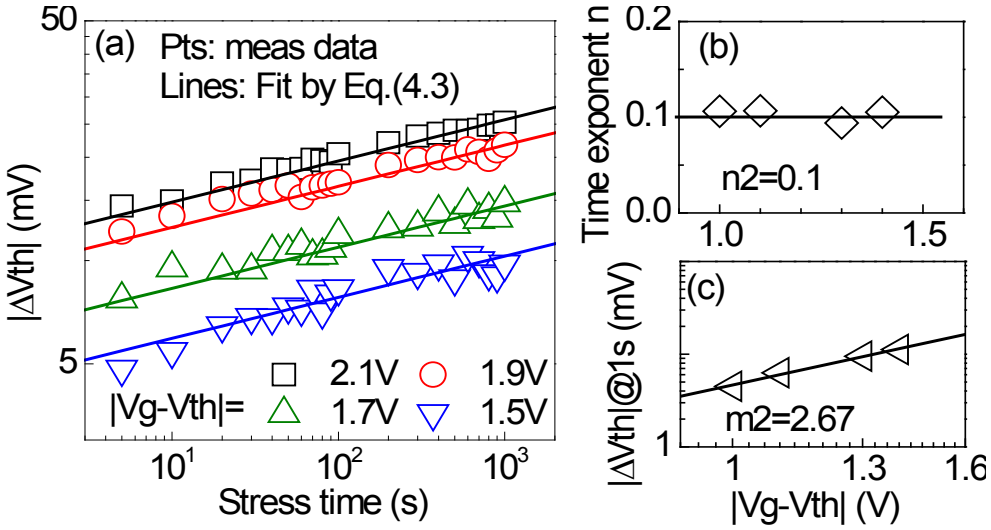

Fig. 4.6 Energy Alternating Defects' charging follows a power law (a) with a $|V_g-V_{th}|$ independent time exponent (b). EAD also follows power law against $|V_g-V_{th}|$ (c).

After all the Energy Alternating Defects' parameters in Equation (4.3) are extracted, within one second Energy Alternating Defects' charging can be extrapolated and As-grown Traps' charging kinetics is obtained by subtracting Energy Alternating Defects' kinetics from Pre_Existing defects' charging, as shown in Fig. 4.7. As-grown Traps' charging under different $|V_g\text{-}V_{th}|$ can be well normalized and described with Equation (4.4):

$$f(t_{ch}) = 1 - e^{-\left(\frac{t_{ch}}{\tau}\right)^{\gamma}} \tag{4.4}$$

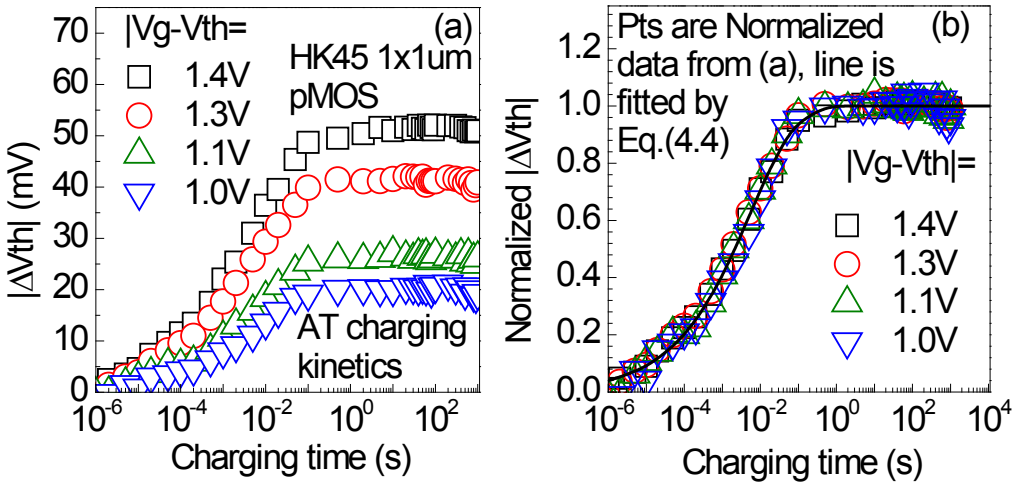

Fig. 4.7 As-grown Traps' charging under different $|V_g\text{-}V_{th}|$ (a). (b) By normalizing against the saturated level at each $|V_g\text{-}V_{th}|$, the kinetics overlaps each other and follows stretched exponent kinetics as described by Equation (4.4).

4.3 Temperature dependence of As-grown Traps and Energy Alternating Defects

To further verify As-grown Traps and Energy Alternating Defects are two types of defects, temperature dependence of As-grown Traps and Energy Alternating Defects are

investigated. Fig. 4.8 shows the As-grown Traps' profiling extracted with m-DMP technique under different temperatures. As NBTI recovery will slow down as temperature decreases, Tdisch=30s is used to ensure all the As-grown Traps will be fully discharged under room temperature. The perfect agreement shows As-grown Traps' profile is independent of temperature, which agrees with the observation on III-V [122] and Germanium [123].

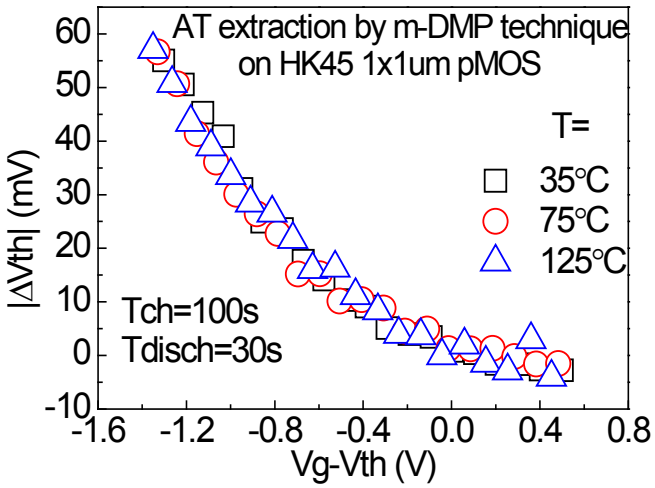

Fig. 4.8 As-grown Traps' profile extracted using m-DMP technique under different temperature.

The charging kinetics of As-grown Traps and Energy Alternating Defects under different temperatures is also extracted as illustrated in Fig. 4.9.

Fig. 4.9 shows although As-grown Traps' profile is independent of the temperature, the charging kinetics will be accelerated under high temperatures. More Energy Alternating Defects will be charged under higher temperature. This again supports that As-grown Traps and Energy Alternating Defects are two different types of defects.

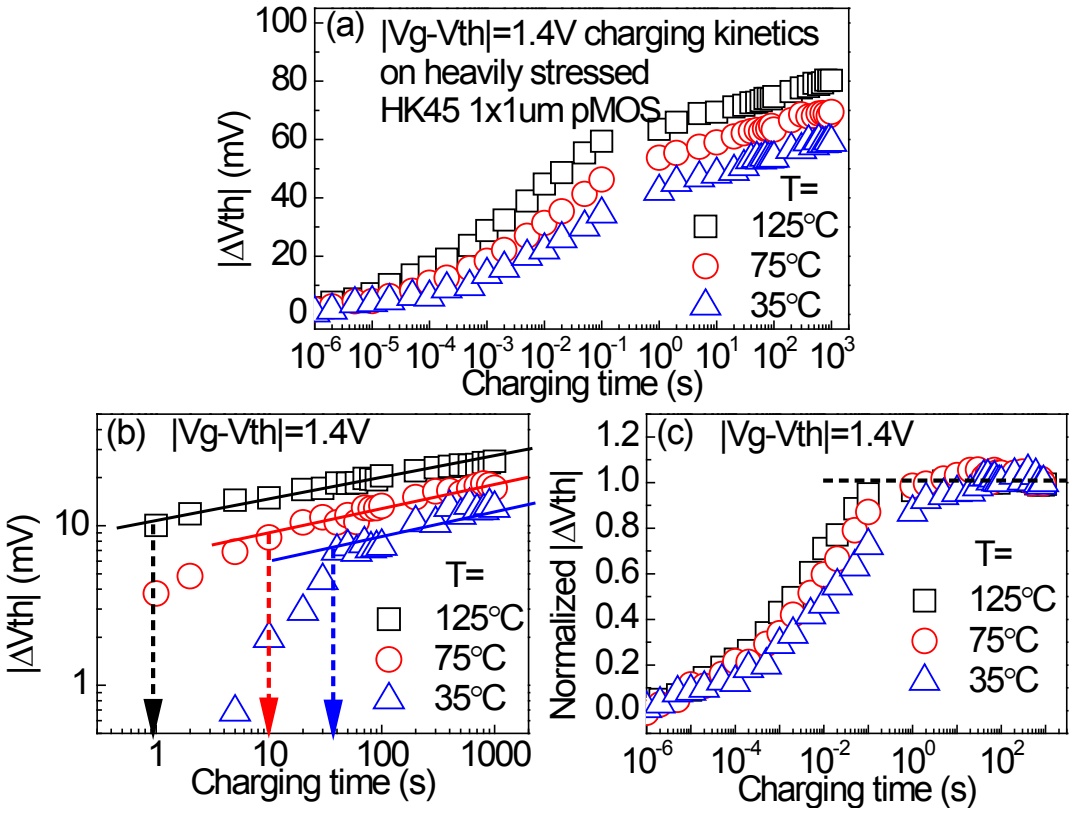

Fig. 4.9 (a) Pre_Existing defects' charging kinetics at the same $|V_g-V_{th}|$ under different temperatures. (b) Energy Alternating Defect (EAD) charging kinetics under different temperature, calculated by subtracting the same amount of saturated As-grown Traps (AT) from Pre_Existing defects' charging kinetics in (a). Note under 75°C and 35 °C, AT charging is much slower compared to 125°C thus 1s is not long enough for AT charging to reach saturation. EAD power law parameters should be extracted from a much higher starting stress time under low temperatures, as shown in the dashed arrows. (c) Normalized AT charging kinetics after subtracting the extrapolated EAD kinetics fitted with the solid lines in (b).

4.4 Summary

In this chapter, Pre_Existing defects is investigated and modelled, to build up a complete BTI model together with Generated Defects as will be discussed in chapter 5.

Experiment results, both defect profile and charging kinetics, support that there are two types of Pre_Existing defects: As-grown Traps and Energy Alternating Defects. As-grown Traps' energy will not alternate during charging/discharging while Energy Alternating Defects' energy will alternate to a lower level after charging and restore after discharging. multi-Discharging-based Multiple Pulses (m-DMP) technique can be applied to extract As-grown Traps' profile and separate As-grown Traps and Energy Alternating Defects' charging kinetics. m-DMP and charging kinetics results show As-grown Traps' profile is independent of temperature but more Energy Alternating Defects will be charged under higher temperature. As-grown Traps' charging will quickly reach saturation, the speed is affected by temperature, charging is faster under higher temperature.

5 As-grown Generation (A-G) model under BTI stress

5.1 Introduction & Motivation

As semiconductor manufacturing technology scaling down, reliability concerns such as BTI have already become a challenge and need to be considered during circuit design.

Lifetime is an important standard for circuit designers. However, in existing mainstream BTI models, the R-D framework has an unsatisfactory predicting capability, the two-stage model fails to do the lifetime projection to long time under use-bias. Moreover, none of the existing models can predict the lifetime for nano-scaled devices, which is really what industry needs.

The A-G model has solved these difficulties via modelling BTI based on understanding different types of defects. In this chapter, it is firstly shown how the framework works on a large device, in the next chapter a demonstration will be given how this framework can be further extended to nano-scale devices.

This chapter is organized into six sections: in section 5.2 a brief review of the A-G model is given; In section 5.3 the application of the A-G model under NBTI stress is demonstrated. A fast characterization method is also given to reduce testing time, DC/AC NBTI under use-bias validates the predicting capability of the model; In section 5.4 the A-

G model is shown to be also applicable under PBTI stress; In section 5.5 a comparison between the A-G model with the Reaction-Diffusion model and Two-Stage model is made; section 5.6 summarizes this chapter.

5.2 A review of the A-G model

Although the A-G model was firstly proposed in 2013 [78], the underlying idea, defect separation, can be traced back to 2004 [124, 125]. Jian F. Zhang et al [124] report new hole trap generation under electrical stress apart from As-grown Hole Traps (AHT), that clearly shows generated hole traps consist of two components: Cyclic Positive Charges (CPC) and Anti-Neutralization Positive Charges (ANPC). Generated mechanism of CPC & ANPC is discussed in [125], but very little modelling work was involved. Moreover, all the measurements here are slow DC measurements. After the fast pulse measurement was introduced, the recoverable component attracts most people's attention.

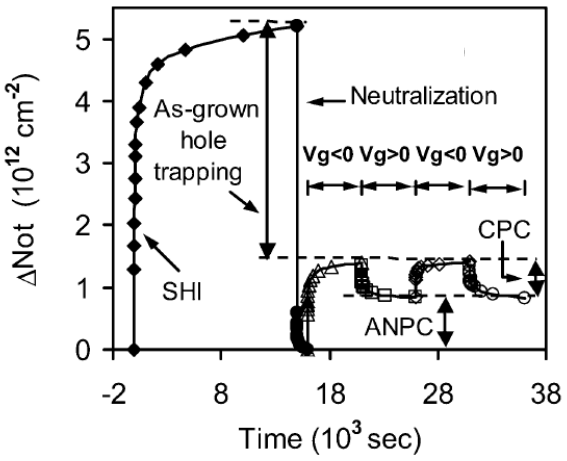

Fig. 5.1 First paper clearly separates hole trap generation into 2 parts: Cyclic Positive Charges (CPC) & Anti-Neutralization Positive Charges (ANPC) [124].

After X. Zheng [77] et al proposed the DMP technique to investigate the energy information of defects on memory devices, the technique was then applied to pMOSFETs to study the energy profile of AHT, CPC & ANPC [79]. The energy profile of SiON samples with different nitridation shows AHT is insensitive to stress time and temperature, located below the valence band edge and is substantially higher in thermal SiON. Generated defects are above the valence-band edge, with CPC located within the bandgap and ANPC located above the conduction band edge. This laid a solid foundation for the separation of As-grown Hole Traps & Generated Defects. The same year, the A-G model was proposed by Z. Ji et al in IEDM, 2013 [78].

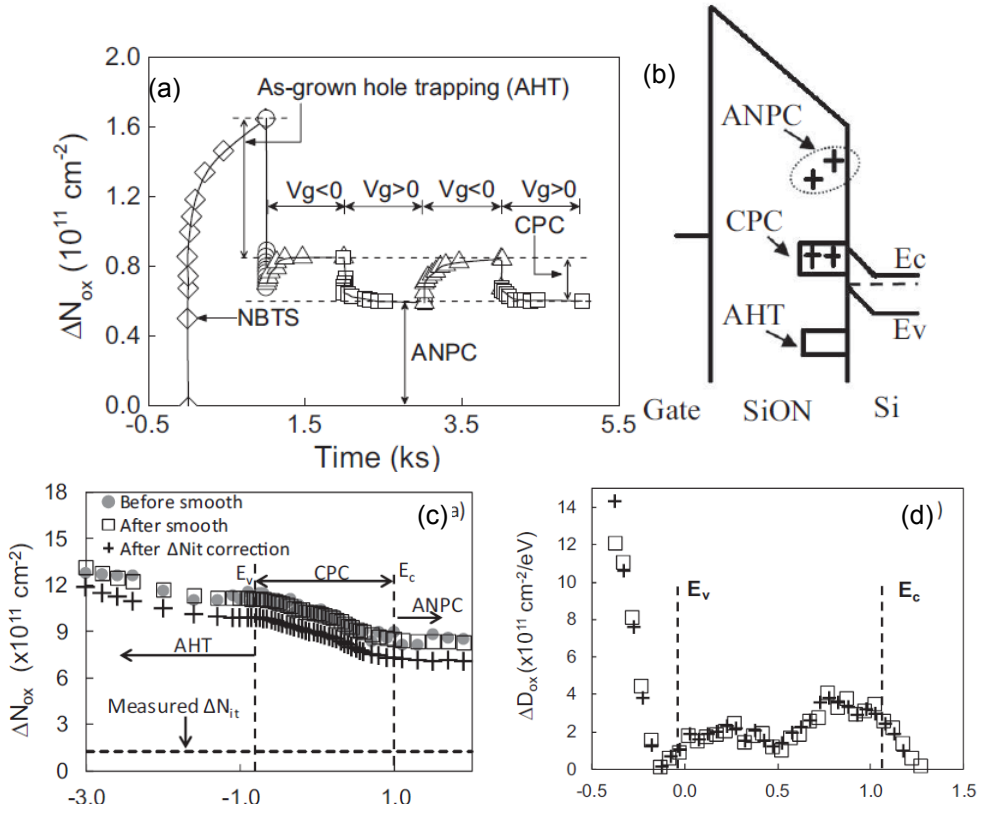

Fig. 5.2 Illustration of energy location of As-grown Hole Traps (AHT), Cyclic Positive Charges (CPC) & Anti-Neutralization Positive Charges (ANPC) (a&b), based on the defect profile results (c&d) measured with Discharging-based Multiple Pulses (DMP) technique.

As the name suggests, the A-G model separates defects into two parts: As-Grown Hole Traps and Generated Defects. As-Grown Hole Traps are independent of stress and will saturate within a short time, while Generated Defects follow a power law. Excellent predicting capability was confirmed on two different processes under DC NBTI stress.

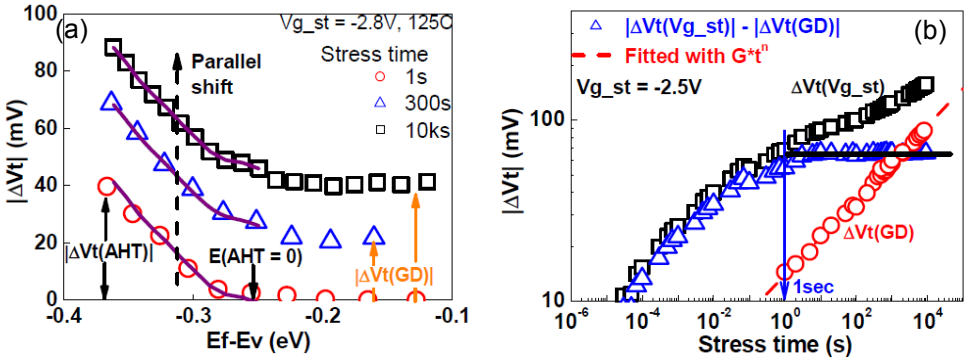

Fig. 5.3 Illustration of the A-G model proposed by Z. Ji et al [78]. (a) Independent As-grown Hole Traps' profile of stress time. (b). $|\Delta V_{th}|$ consists of a fast saturated As-grown Hole Traps component plus Generated Defects , which follow a power law.

After DC verification, the A-G model was then extended to AC stress condition [33]. As discharging kinetics is needed for AC modelling, CPC and ANPC must be separately modelled due to their different discharging properties. A pragmatic V_g=0 is used in [33] as the criteria to separate CPC & ANPC, as circuits normally operate under an AC bias between operating voltage and zero. Although excellent predicting capability was proven under different frequency and duty factor, the complicated test pattern makes the testing time for parameter extraction very long: more than two days for a given process. This is not acceptable for nano-scale devices as repeated tests needs to done. Defect separation technique in [33] cannot be applied on single nano-scaled devices, as it is impossible to

find a constant gate voltage as the boundary of As-grown Traps & Generated Defects for a bunch of nano-scale devices which have stochastic fresh $|V_{th}|$.

After Si devices, the A-G model was again applied on other material like Ge [121, 123] and III-V [122]. It was found that unlike SiON devices, Ge pMOSFETs have a significant amount of Energy Alternating Defects, which has not been considered in the A-G model proposed in [33]. Further investigation showed Si High-K devices also have this Energy Alternating Defects component. After taking this Energy Alternating Defects component into consideration, the up-to-date A-G model is illustrated in Fig. 5.4. In the following sections of this chapter, an A-G model with good predicting capability for various kinds of stress condition, NBTI & PBTI, DC & AC, Si & Ge, is proposed. In chapter 6 it is further shown the A-G model is capable of predicting the lifetime of nano-scale devices.

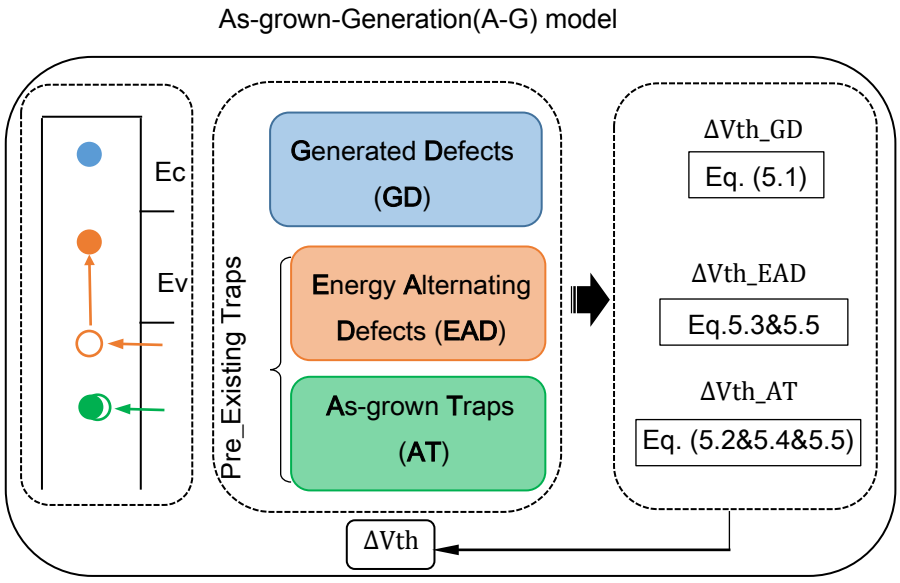

Fig. 5.4 Illustration of the A-G model based on the understanding and separation of different type of defects.

5.3 Application of the A-G model to NBTI

5.3.1 Generate Defects modelling

Generated Defects, although widely observed [126], are usually ignored [87] because of the lack of a proper characterization method. Conventionally, Generated Defects are simply determined after discharging [105], which does not saturate, making the extraction sensitive to the discharging time. This is also applied in early A-G model: Generated Defects was extracted by discharging under a selected V_g level, which is determined from As-grown Traps' profile results by parallel shift, and T_{disch}=T_{stress} is used to discharge all the As-grown Traps [33]. However, there is no justification of this T_{disch} and it's difficult to understand why As-grown Traps & Generated Defects can have a clear boundary which separates them into blocks. In the updated A-G model in this chapter, Generated Defects are measured after discharging under a positive bias and then recharging under use bias, as detailed in chapter 3. By doing so the entire Generated Defects are extracted and the time exponent is independent of discharging time and voltage. Generated Defects' kinetics under different overdrive voltage, $|V_g\text{-}V_{th}|$, is found to follow classical power law (Fig. 5.5) thus can be modelled with Equation (5.1) for long term prediction.

$$GD = g_1 \cdot \left(|V_g - V_{th}|\right)^{m_1} t^{n_1} \tag{5.1}$$

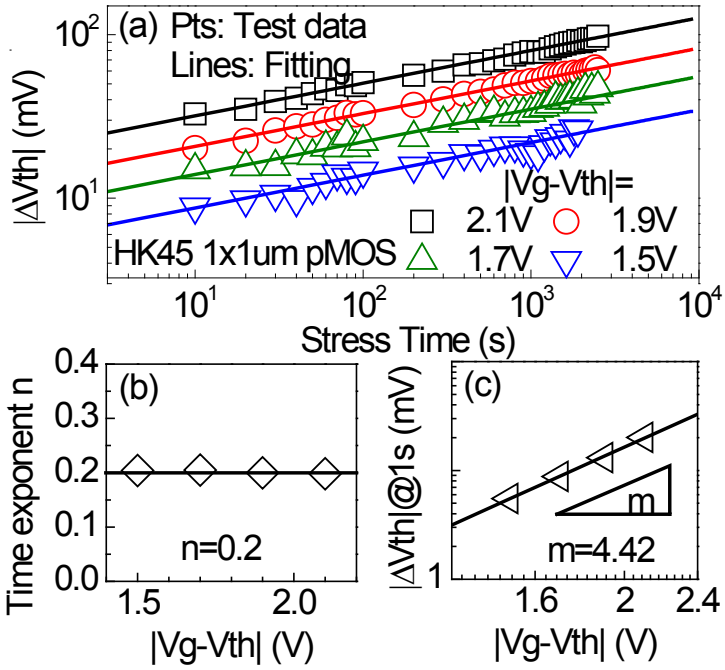

Fig. 5.5 (a) Kinetics of Generated Defects under different $|V_g-V_{th}|$ stress condition. (b) Generated Defects' time exponent n is independent from $|V_g-V_{th}|$. (c). $|V_g-V_{th}|$ exponent extraction.

5.3.2 Pre_Existing defects modelling

As detailed in chapter 4, by exploring the different impacts of charging on the energy distribution of Energy Alternating Defects and As-grown Traps, their charging kinetics can be successfully separated. As-grown Traps saturate rapidly (Fig. 5.6a). The saturation level increases exponentially with $|V_g-V_{th}|$ in Equation (5.2), and the normalized kinetics is independent of $|V_g-V_{th}|$ (Fig. 5.6b) and can be described by Equation (5.3). Energy Alternating Defects' charging follows power law in both time and voltage (Fig. 5.7a-c) and can be modelled with Equation (5.4). Note although both Generated Defects and Energy Alternating Defects follow a power law relationship, they have different time exponents of 0.2 (Fig. 5.6b) and 0.1 (Fig. 5.7b) and thus need separate modelling.

$$AT = p_1 \cdot \exp[p_2 \cdot (|V_g - V_{th}|)] \quad (5.2)$$

$$f(t_{ch}) = 1 - e^{-\left(\frac{t_{ch}}{\tau}\right)^{\gamma}} \quad (5.3)$$

$$EAD = g_2 \cdot (|V_g - V_{th}|)^{m_2} t^{n_2} \quad (5.4)$$

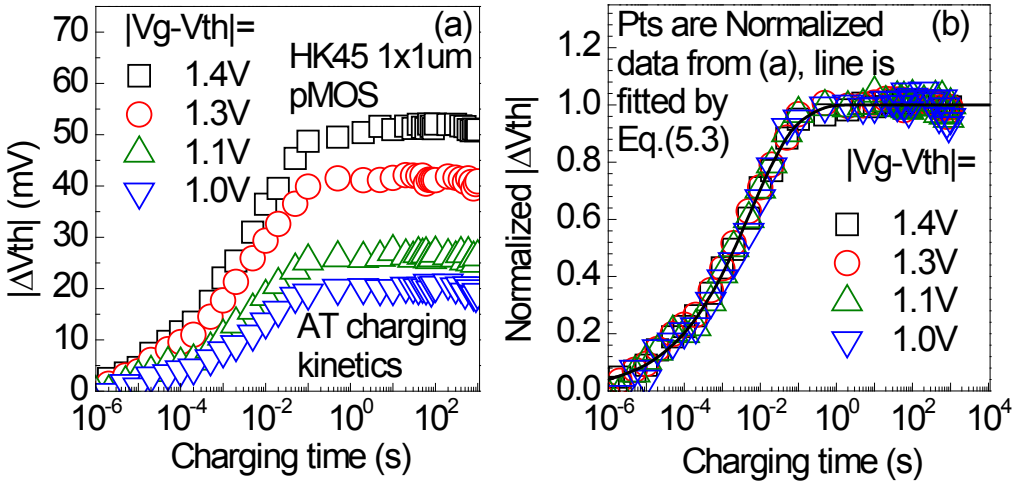

Fig. 5.6 As-grown Traps' charging under different $|V_g\text{-}V_{th}|$ (a). (b) By normalizing against the saturated AT under different $|V_g\text{-}V_{th}|$, the kinetics overlaps each other and follows stretched exponent kinetics in Equation (5.3).

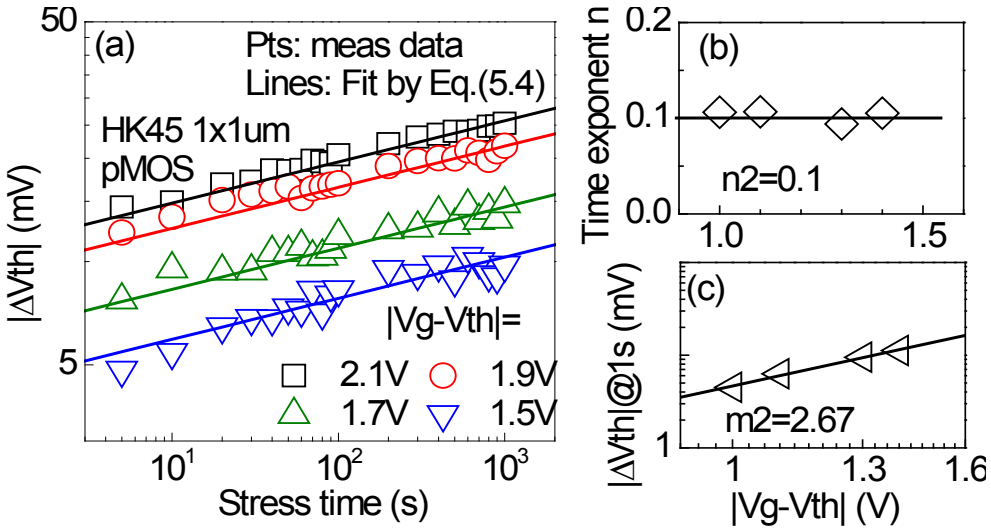

Fig. 5.7 (a) Energy Alternating Defects charging kinetics under different $|V_g\text{-}V_{th}|$ on a heavily stressed pMOSFET. (b) EAD charging time exponent is independent of $|V_g\text{-}V_{th}|$ and its value is much smaller compared to Generated Defects. Voltage exponent is given in (c).

5.3.3 Fast-Voltage Step Stress technique for the A-G model parameter extraction

To accelerate NBTI stress, stress voltage is raised above its normal operation level. Accelerated test under Constant Voltage Stress scheme is normally applied with typical total testing time of a few days [78].

This is tolerable for big devices, where Device-to-Device Variability is negligible. However, nano-scale devices require repetitive tests to retrieve statistical properties of Device-to-Device Variability and the task becomes laborious.

The Voltage Ramp Stress methodology was proposed in 2009 [88, 127] to accelerate g_1 and m_1 extraction and soon became widely used by industry, whose test pattern has already been introduced in chapter 2. It's an efficient screening tool for comparing different processes, but the accuracy is less satisfactory. Fig. 5.8 shows there is a large gap between $|\Delta V_{th}|$ measured by Constant Voltage Stress tests and prediction by the calculation with parameters extracted from the Voltage Ramp Stress tests.

Voltage Step Stress is another fast characterization method proposed by Z. Ji et.al [128] in 2014. Compared to the Voltage Ramp Stress pattern which attempts to capture the whole $|\Delta V_{th}|$, the Voltage Step Stress method uses DC slow measurements to ensure the power law and then extracts g_1,m_1,n_1 on a single device. g_1 & n_1 is extracted from the first stress phase, the following phases then can be used to extract m_1.

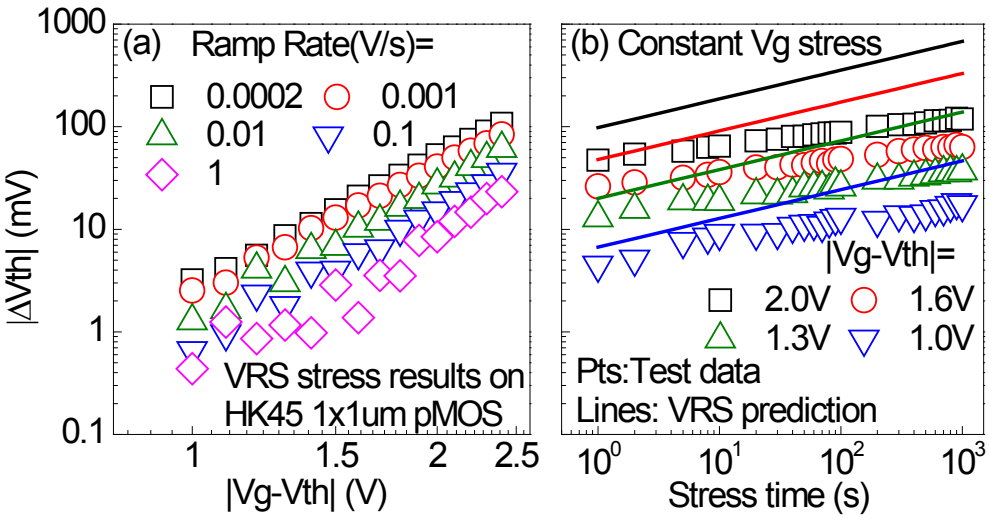

Fig. 5.8 (a) Typical results of applying Voltage Ramp Stress technique on HK45 pMOSFET. (b) Comparison of the test data measured under conventional Constant Voltage Stress with the Voltage Ramp Stress technique predicted value.

Although the testing time has been reduced to within two hours by applying the Voltage Step Stress technique, there is still room to further accelerate this technique. Note g_1 & n_1 are extracted from the first stress phase, the remaining 80% of the testing time, is trying to extract the m_1 value. With a known time exponent n, it is unnecessary to stress 1,000 seconds in the first phase to get n_1 then. Actually the first stress phase is usually a light or moderate stress to leave room for accelerated $|\Delta V_{th}|$ in latter step stress, thus the total $|\Delta V_{th}|$ amount in the first stress phase will not be very significant, questioning the accuracy of g_1 & n_1 value. In chapter 3 a Stress-Discharge-Recharge technique has been proposed to get the reliable and accurate time exponent of Generated Defects regardless of discharging time, bias and temperature. Based on this apparent time exponent a further accelerated Voltage Step Stress technique Fast Voltage Step Stress is developed.

The test waveform used by Fast Voltage Step Stress is given in Fig. 5.9: $|V_g-V_{th}|$ starts from a low value (i.e. -0.35V) and then gradually increases with a small step (typical value is ~20mV). Each step is applied for a pre-defined time, Δt (10s here). At the end of each step, Generated Defects are monitored by the procedure in Fig. 5.9a.

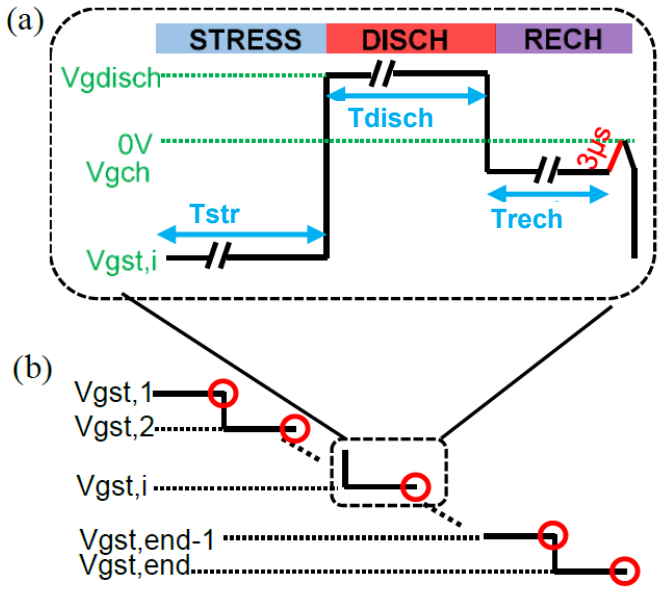

Fig. 5.9 Illustration of the Fast-Voltage Step Stress waveform. (a) The test procedure under each voltage step. A stress-discharge-recharge sequence is used. At the end of the recharge step, $|\Delta V_{th}|$ was monitored from a corresponding IV, which was taken from the 3µs pulse edge with V_d=-0.1V applied. Tdisch and Tch have negligible impact on Generated Defects' extraction [129]. In this work, $T_{disch}=T_{ch}$=10s is used. (b) The V_g waveform for the Fast-Voltage Step Stress technique.

A typical result is shown in Fig. 5.10. Under low $|V_g-V_{th}|$, there is a flat plateau. This is because generation is negligible in 10 seconds when $|V_g-V_{th}|$ is low. The $|\Delta V_{th}|$ in this plateau actually comes from the As-grown Traps corresponding to the charging V_{gch}. As-grown Traps will soon reach saturation and remain the same, so long as the same V_{gch} is applied. This explains the plateau.

As $|V_g\text{-}V_{th}|$ increases, Generated Defects induced $|\Delta V_{th}|$ becomes higher, resulting in the rise in Fig. 5.10. By subtracting AHT in the plateau from the total ('◊'), Generated Defects induced $|\Delta V_{th}|$ under different $|V_g\text{-}V_{th}|$ can be obtained ('□'). According to equation (5.1) [18], to generate the same amount of Generated Defects induced $|\Delta V_{th}|$, a stress under a $|V_g\text{-}V_{th}|$ for a time Δt is equivalent to a stress under another overdrive voltage V_{geff} for the effective stress time of Δt_{eff}:

$$\Delta t_{eff} = \left(V_{gov}/V_{geff}\right)^{m/n} \cdot \Delta t \tag{5.5}$$

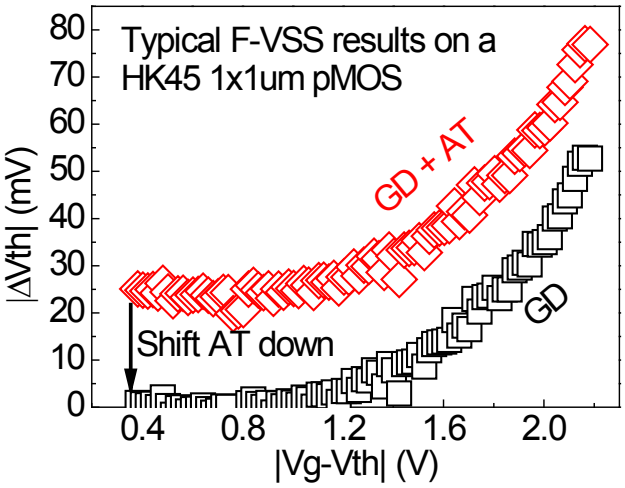

Fig. 5.10 Typical results of Fast-Voltage Step Stress technique with $|V_g\text{-}V_{th}|$ gradually increasing from 0.4V to 2.2V under 125 °C. Each volage lasts 10s and the voltage step of 20mV is used. The total degradation includes both As-grown Traps and Generated Defects. AT can be determined under low stress condition where Generated Defects are neglible. By subtracting AT from the total, Generated Defects induced $|\Delta V_{th}|$ can be extracted.

With the time exponent, n, extracted from the Constant Voltage Stress measurement (0.2 for this process) [17], Δt for each step in the Fast-Voltage Step Stress test can be converted to Δt_{eff} under any given m value using Equation (5.5). To minimize the measurement inaccuracy, Only Generated Defects induced $|\Delta V_{th}|$ over 5mV is used for the analysis. Fig. 5.11a shows that $|\Delta V_{th}|$~Δt ('□') is transformed to $|\Delta V_{th}|$~Δteff kinetics under

three different m. With the larger m value, the device takes longer to reach the same degradation, leading to smaller apparent time exponent, n'. Only when m is correctly determined, the n' from the transformed kinetics equals the predetermined n. The corresponding parameter g_0 can then be determined simultaneously, as shown in Fig. 5.11b.

When extracting the parameters in the power law, the Voltage Ramp Stress technique uses the total degradation, i.e. the sum of As-grown Traps and Generated Defects. As-grown Traps saturate quickly and will distort the power law of Generated Defects. This is the reason for the inaccuracy of the Voltage Ramp Stress method. The Fast-Voltage Step Stress method can reliably remove As-grown Traps and only fit Generated Defects induced $|\Delta V_{th}|$ with the power law. This delivers the accuracy.

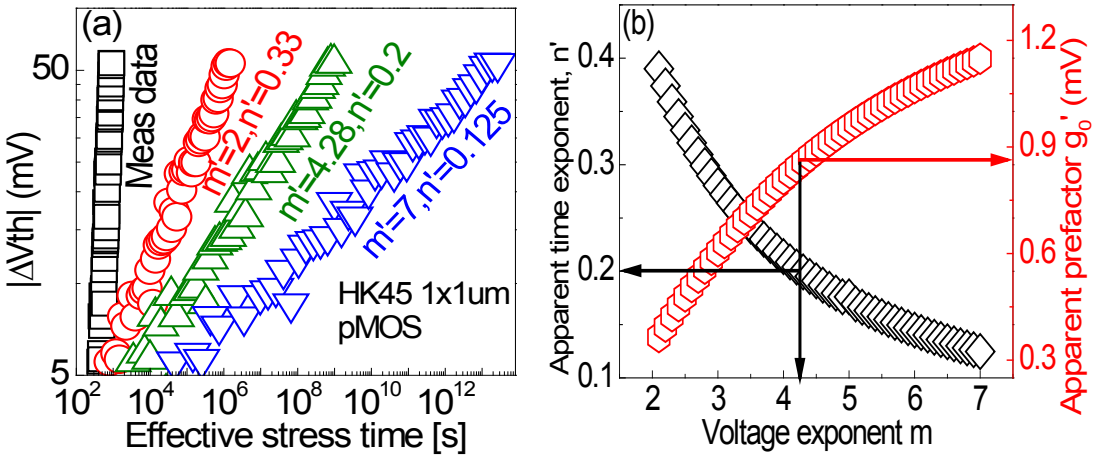

Fig. 5.11 (a) The measured degradation is plotted against time ('□'). By assuming different m', the real stress time under each stepping voltage can be transformed to effective stress time under a certain constant effective $|V_g-V_{th}|$ (1V used). (b) The extracted n' and g_0' under given m'. The correct m and g0 corresponds to n=0.2.

To validate the Fast-Voltage Step Stress technique, the extracted parameters are used to predict Generated Defects under various constant $|V_g-V_{th}|$ stress on two processes, as shown in Fig. 5.12, Generated Defects were measured under constant $|V_g-V_{th}|$ and these data were not used to fit the g_1 & m_1. Good agreement has been achieved even when $|V_g-V_{th}|$ is as low as 0.9 V, validating the proposed Fast-Voltage Step Stress method. Note none of the data in Fig. 5.12 is used for parameter extraction.

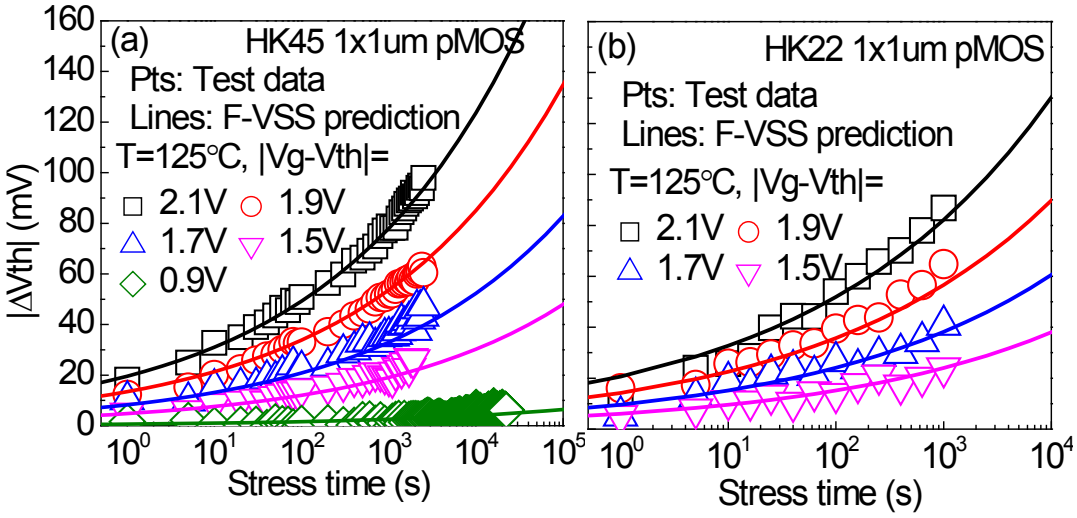

Fig. 5.12 Validation of the extacted parameters using Fast-Voltage Step Stress technique. Constant $|V_g-V_{th}|$ Stress measured data under different $|V_g-V_{th}|$ is compared with F-VSS prediction. Good agreement is achieved on both HK45 and HK22 processes.

5.3.4 Experimental Validation under both DC & AC

For DC NBTI stress, the total $|\Delta V_{th}|$ should be the sum of the $|\Delta V_{th}|$ caused by Generated Defect, Energy Alternating Defects & saturated As-grown Traps. Note constant E_{ox} stress is used to extract the framework parameters, while real circuits work under a constant operating V_g (V_{gop}). A flow chart as shown in Fig. 5.13 is used to calculate $|\Delta V_{th}|$.

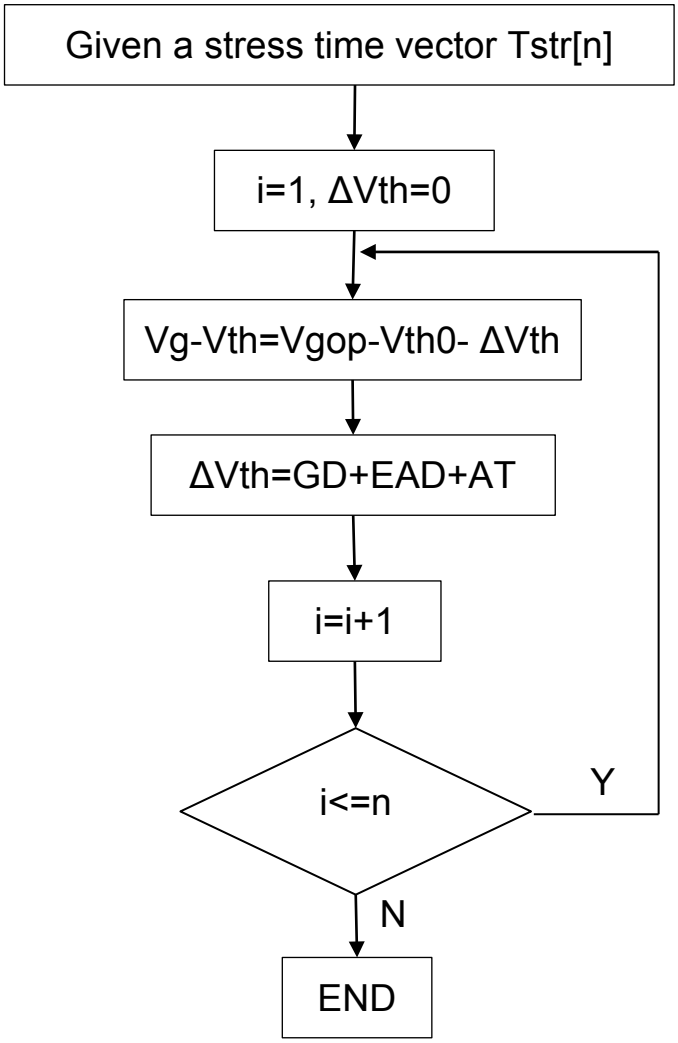

Fig. 5.13 Procedure to calculate $|\Delta V_{th}|$ under constant voltage stress with parameters extracted under constant $|V_g-V_{th}|$.

Fig. 5.14 shows the good agreement between experiment data and the A-G model prediction under DC NBTI, as shown in Fig. 5.14. It should be noted that the time exponents of Generated Defects & Energy Alternating Defects extracted under constant Eox high gate voltages, voltage exponent is extracted using Fast-Voltage Step Stress and the experiment data in Fig. 5.14 has not been used in parameter extraction.

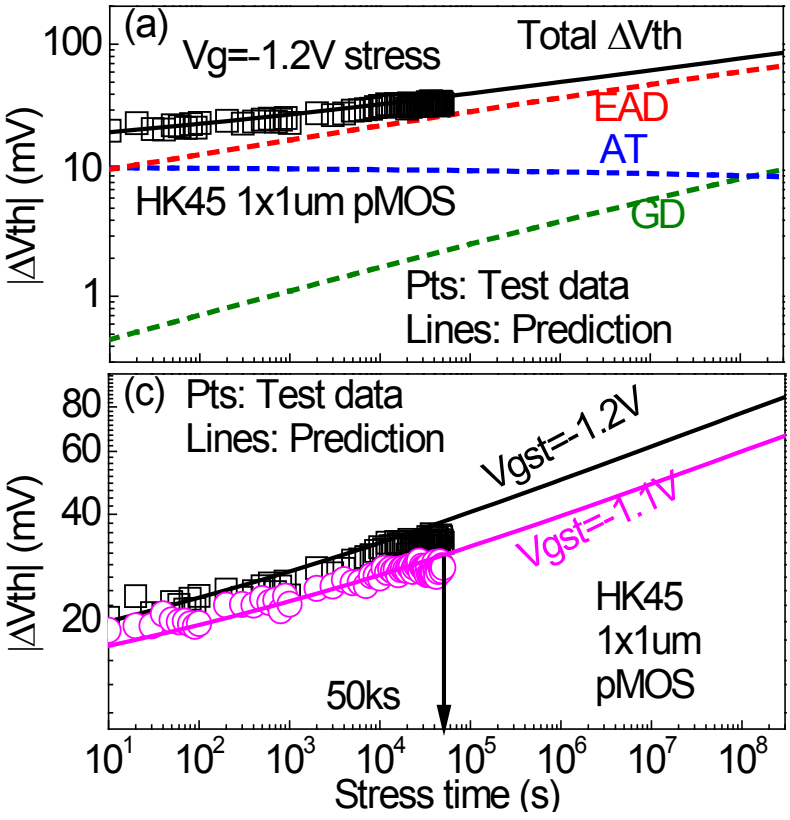

Fig. 5.14 (a) V_g=-1.2V DC constant voltage stress experiment data can be well predicted by the A-G model. The solid line total $|\Delta V_{th}|$ is the sum of 3 dash lines corresponding to different components. Note AT is slightly decreasing against stress time, this is due to the $|V_g-V_{th}|$ deduction as $|\Delta V_{th}|$ increase. Another V_g=-1.1V is also given in (b).

To clarify the different between the "prediction" and "fitting", the Reaction-Diffusion (R-D) framework "prediction" from [130] is replotted in Fig. 5.15 to compare with the A-G model's prediction. Fig. 5.15(b) seems to show the agreement between the measured total |ΔVth| and the "predicted" values from the R-D framework, but in fact this is the "parameter extraction" instead of "prediction". Because the measured total |ΔVth| data itself is used to extract the R-D framework parameters. The R-D Parameters are extracted by empirically adjusting the parameters to best fit the measured total |ΔVth|, a good "fitting" will be achieved anyway. In terms of the A-G model prediction, all the

parameters are extracted from different patterns and different accelerated stress conditions separately, the measured total $|\Delta V_{th}|$ under much lower stress condition is then compared with the predicted values.

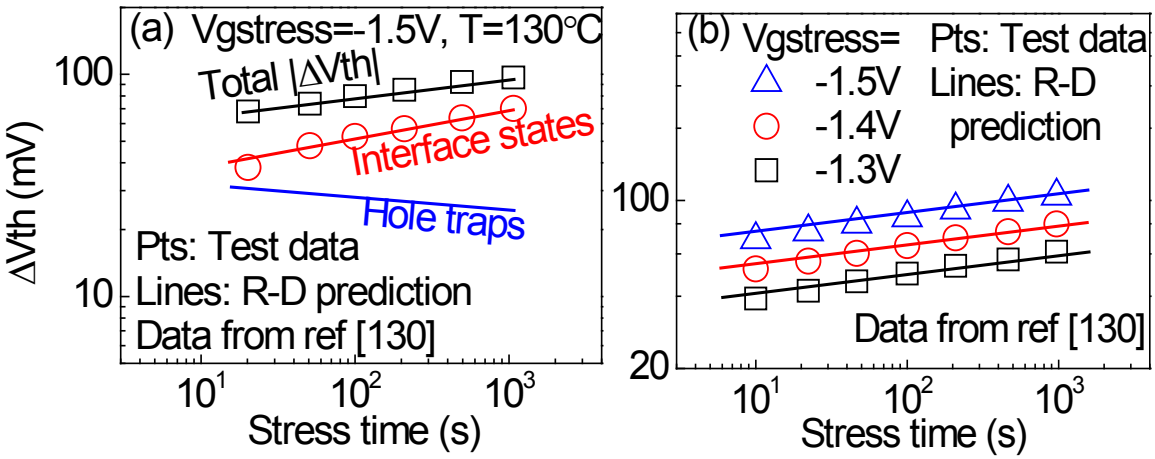

Fig. 5.15 A replot of the Reaction-Diffusion (R-D) framework predicting capability verification data from [130]. (a) Total |ΔVth| under -1.5V NBTI stress contains two components: interface states and Hole Traps induced $|\Delta V_{th}|$, the former one is fitted with time exponent n=1/6 and the latter one is fitted with the empirical equations, as detailed in section 1.2.1 Reaction-Diffusion framework. (b) Measured total $|\Delta V_{th}|$ can be fitted the Reaction-Diffusion framework.

Most of circuit work under AC stress. Unlike DC the stress is always applied thus Energy Alternating Defects follow a power law and As-grown Traps saturate, V_g is alternating between $V_{gstress}$ and 0V according to the frequency and duty factor under AC stress. For Generated Defects, it is shown in chapter 3 that they are equivalent stress time (stress time*duty factor) driven process and follow the same kinetics under both DC and AC. For Pre_Existing defects, their charging kinetics under $V_{gstress}$ and discharging kinetics under 0V of are needed to predict AC BTI kinetics [33].

Discharging kinetics of Pre_Existing defects under 0V can be directly measured on heavily stressed devices, on which further generation is negligible. Since there is Cyclic Positive Charges' charging discharging on heavily stressed device, so before Pre_Existing defects' kinetics measurement, $|V_g-V_{th}|$=0.35V is applied 100s to charge up the Cyclic Positive Charges and then its discharging kinetics under 0V can be directly measured. Results shown in Fig. 5.16 have already subtracted Cyclic Positive Charges' discharging kinetics thus Cyclic Positive Charges' impact have already been excluded.

$$\Delta Vth_{tdisch} = \Delta Vth_{tdisch=0} \cdot \left(1 + B \cdot t_{disch}^{\beta}\right)^{-1} \tag{5.6}$$

The discharging of the total Pre_Existing defects after different charging times or voltages can be well scaled by a universal recovery curve in Equation (5.6).

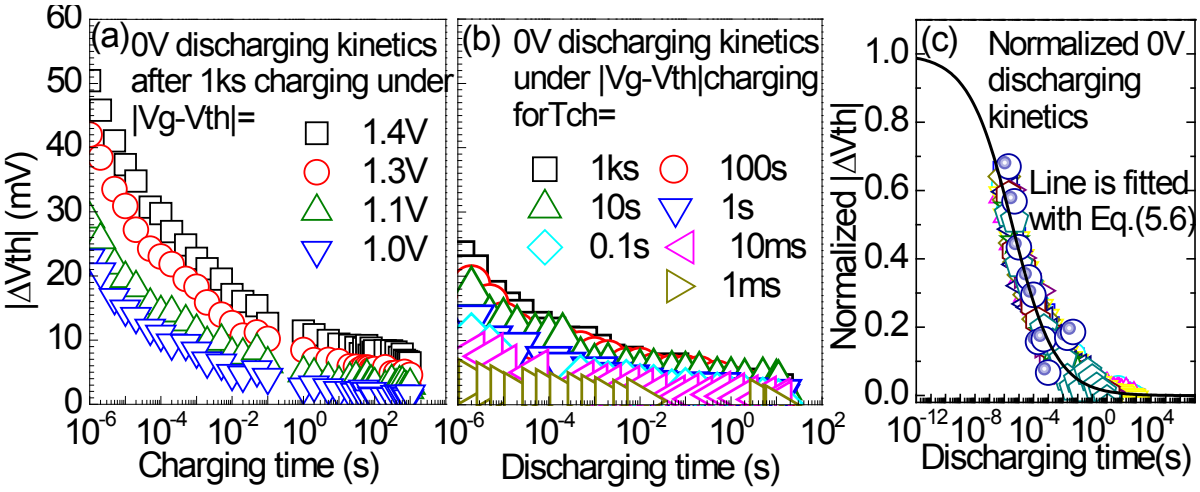

Fig. 5.16 Discharging kinetics under 0V of the Pre_Existing defects after different charging voltage (a) and charging time (b). (c) They can be normalized and modelled by a universal recovery trace in Equation (5.6).

With the discharging kinetics parameters all the parameters for the A-G model have been extracted. By solving these coupled equations, DC/AC NBTI degradation under any operating voltage, Frequency and Duty Factor can be predicted.

Fig. 5.17 shows the good agreement between experiment data and prediction. Stress time =1ks, Stress voltage is specially selected to make the total degradation $|\Delta V_{th}|$ under 1kHz ~30mV so $|\Delta V_{th}|$ can be still observable under high frequency and small duty factor. On this process $|V_g|$=1.3V is used. Extracted parameters are given in Table 5.1.

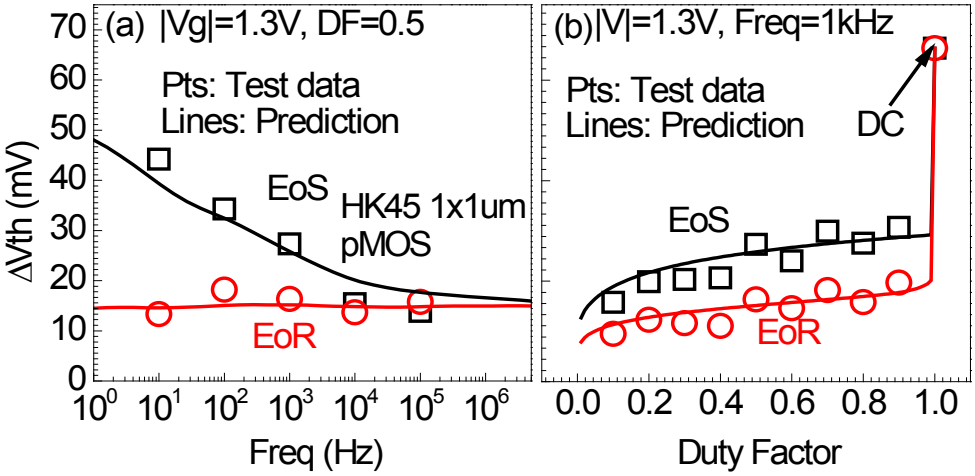

Fig. 5.17 Verification of the A-G model with parameters extracted from F-VSS: the prediction (lines) agrees well with test data (symbols) for AC/DC NBTI (a&b). The test data here were not used for fitting the parameters. The stress time is 1ks.

Table 5.1 The A-G model parameters extracted on HK45 1x1um pMOS under NBTI stress.

Defects	GD		EAD		AT				Discharge	
Parameters	g_1(mV)	0.69	g_2(mV)	5.18	p_1(mV)	1.65	τ (s)	5.8e-3	B	19.88
	m_1	4.42	m_2	2.30	p2	2.52	γ	0.36	β	0.24
	n_1	0.20	n_2	0.10						

5.3.5 Evaluate each component contribution under use-bias at device lifetime

With parameters in Table 5.1, $|\Delta V_{th}|$ under DC/AC any given voltage can be calculated. An evaluation of each component contribution at 10 years is given in Fig. 5.17 (DC) and Fig. 5.18 (AC).

Fig. 5.18 shows under DC NBTI stress condition, As-grown Traps still play an important role as all the chargeable As-grown Traps are filled under DC stress condition. Energy Alternating Defects are the dominating defects at device lifetime due to its much larger pre-factor (g_2) compared with Generated Defects. The lower the operating voltage drops, the higher Energy Alternating Defects will contribute. This is due to the fact that Generated Defects have a much larger voltage exponent compared to Energy Alternating Defects thus will drop much faster as $|V_g|$ lowers down.

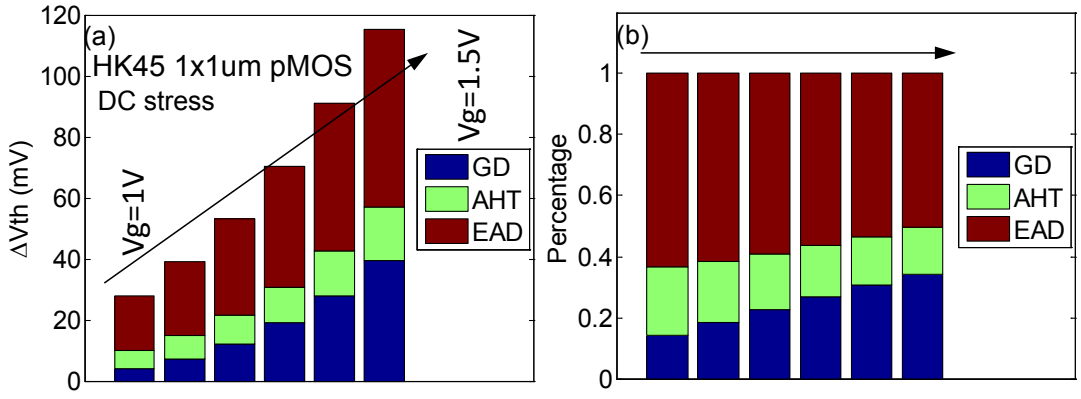

Fig. 5.18 (a) The A-G model predicted $|\Delta V_{th}|$ of different types of defects after 10 years DC stress under different operating V_g. (b) The contribution of each type of defects

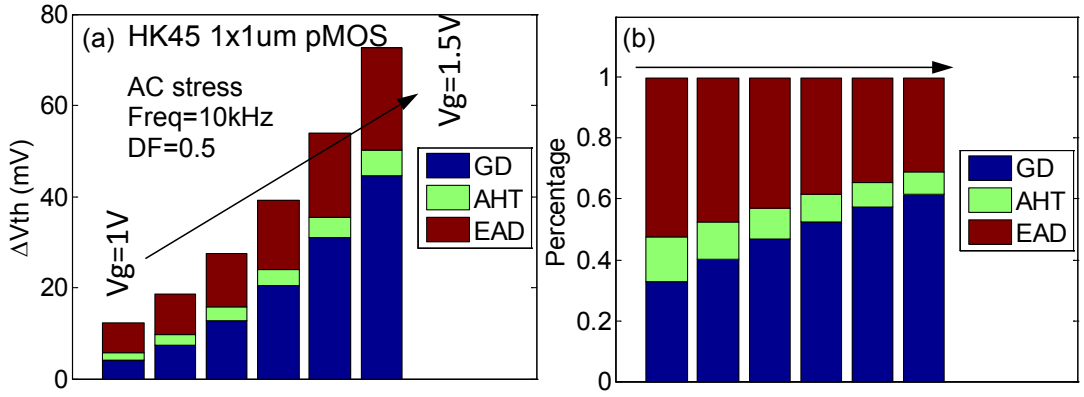

Fig. 5.19 The A-G model predicted $|\Delta V_{th}|$ of different types of defects after 10 years AC stress under different operating voltage V_g. (b) The contribution of each type of defects

Fig. 5.19 shows under AC NBTI stress, As-grown Traps are less important compared to DC as only a small portion can be filled under high frequency AC stress. Energy Alternating Defects are still important but no longer dominating. Generated Defects' become an important component and cannot be neglected.

5.4 Application of the A-G model to PBTI

PBTI has attracted much less attention in the past decades due to the fact that PBTI $|\Delta V_{th}|$ is much smaller within the feasible testing time ($<10^7$ s) compared to NBTI. Although some early works [107, 130] reported a big time exponent n for PBTI, the time exponent n values are dependent on the measurement condition thus had not attracted people's attention. By Applying the Stress-Discharge-Recharge technique, Generated Defects are extracted under different $|V_g-V_{th}|$ PBTI stress as shown in Fig. 5.20. Note that the extracted Generated Defects show the time exponent of 0.32 which is much larger than the

value reported by most groups [131, 132]. However it is very close to the value reported by IBM when electron trapping has been removed [133] or suppressed [134].

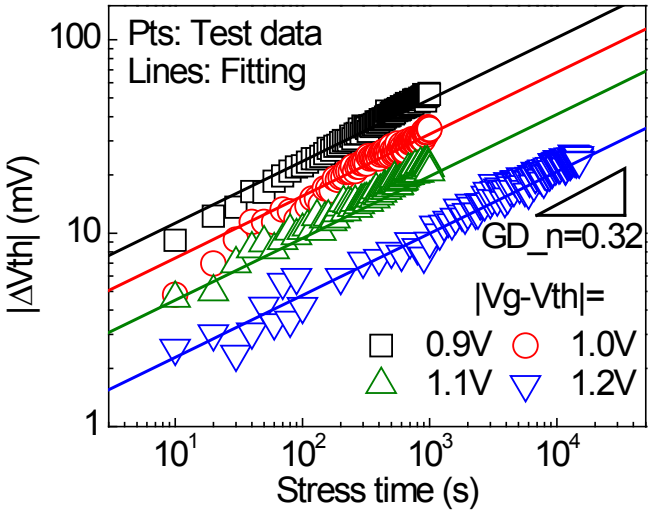

Fig. 5.20 Generated Defects measured by Stress-Discharge-Recharge technique under different $|V_g-V_{th}|$ PBTI stress. Lines are fitted with power law to extract parameters of the A-G model. Note the time exponent here is much larger (0.32) compared to NBTI (0.2 as shown in Fig. 5.5a).

Following the same test procedure and analysis as NBTI, Pre_Existing defects' parameters in the A-G model under PBTI stress can be extracted as shown in Fig. 5.21. m-DMP technique is firstly applied to get As-grown Traps' profile (Fig. 5.21a). Saturated As-grown Traps induce $|\Delta V_{th}|$ calculated from the profile is then subtracted from the over one second Pre_Existing defects' charging kinetics to get Energy Alternating Defects charging kinetics (Fig. 5.21b). Energy Alternating Defects follow a power law against both charging time and charging $|V_g-V_{th}|$ and can be well fitted with Equation (5.4). Within one second Energy Alternating Defects' charging is then calculated by Equation (5.4) and is subtracted from Pre_Existing defects' charging kinetics to get As-grown Traps' charging kinetics under different $|V_g-V_{th}|$, as shown in Fig. 5.21d. By adding the kinetics of Generated Defects and Pre_Existing defects, $|\Delta V_{th}|$ kinetics can be calculated under any

$|V_g|$ DC PBTI stress. A comparison between the predicted value and test data under different $|V_g|$ DC PBTI stress is shown in Fig. 5.22, Good agreements are achieved.

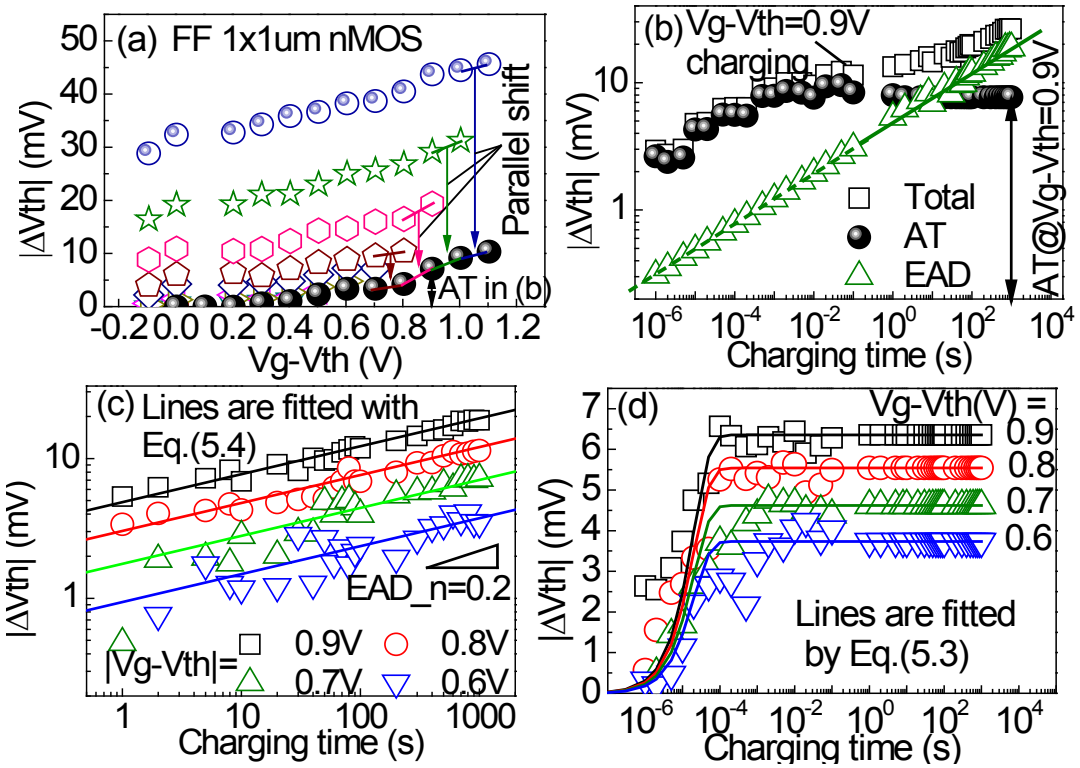

Fig. 5.21 Pre_Existing defects' parameter extraction under PBTI. (a) As-grown Traps (AT) profile is extracted with the m-DMP technique. (b) Energy Alternating Defects (EAD) charging kinetics extraction. (c) EAD under different $|V_g-V_{th}|$ m, note EAD time exponent is much smaller (0.2) compared to Generated Defects (0.32, Fig. 5.19). (d) AT charging kinetics under different $|V_g-V_{th}|$.

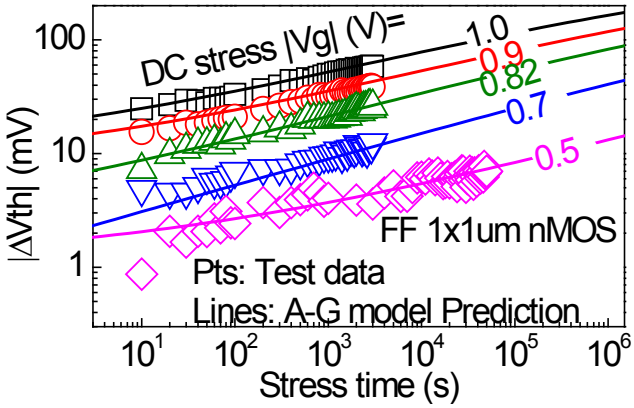

Fig. 5.22 DC PBTI stress under various V_g can be well predicted by the A-G model. The measured data (symbols) were not used for the A-G model parameter extraction.

To model the AC PBTI stress, discharging kinetics of Pre_Ex under 0V are measured after charging under different $|V_g-V_{th}|$, as shown in Fig. 5.23. They can be well fitted with the same Equation (5.6) as NBTI. The test data under $|V_g|$=1.3V different frequency and duty factor is well predicted by the A-G model, as shown in Fig. 5.24.

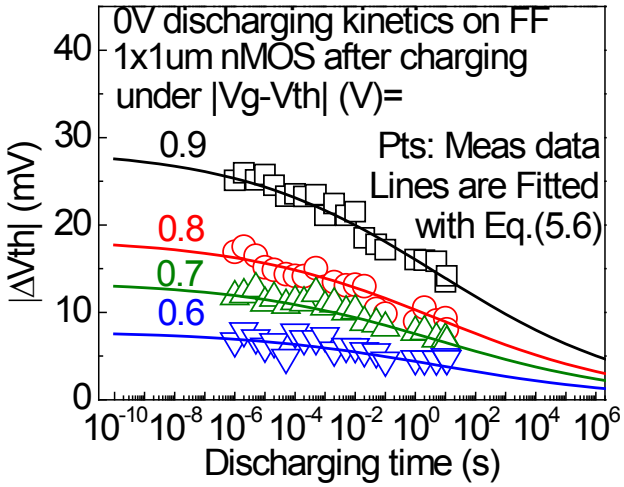

Fig. 5.23 Pre_Existing defects discharging kinetics under 0V after charging under different $|V_g-V_{th}|$ PBTI stress.

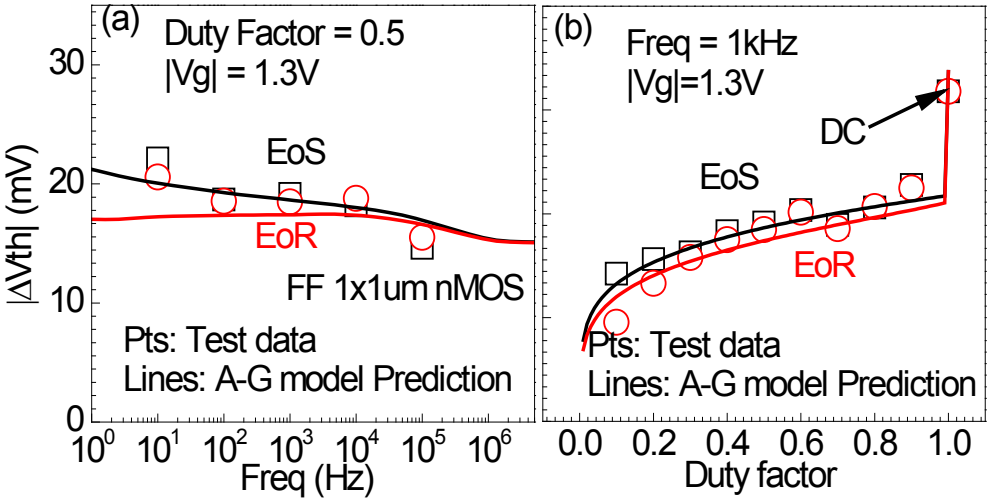

Fig. 5.24 $|\Delta V_{th}|$ measured under $|V_g-V_{th}|$=0.9V AC PBTI stress for various frequency (a) and duty factor (b) can be well predicted by the A-G model.

Table 5.2. The A-G model parameters under PBTI stress

Defects	GD		EAD		AT				Discharge	
Parameters	g_1(mV)	2.75	g_2(mV)	7.28	p_1(mV)	0.29	τ(s)	2.0e-4	B	0.79
	m_1	5.23	m_2	3.28	p_2	3.55	γ	1.0	β	0.13
	n_1	0.32	n_2	0.2						

With the NBTI parameters in Table 5.1 and PBTI parameters in Table 5.2, NBTI and PBTI degradation under use-bias is compared, as shown in Fig. 5.25. Contrary to People normally expect PBTI is less important compared with NBTI [58, 135, 136], PBTI could be as important or even more important at device lifetime due to its larger time exponent for both Generated Defects (PBTI:0.32 > NBTI:0.2) and Energy Alternating Defects (PBTI:0.2 > NBTI:0.1).

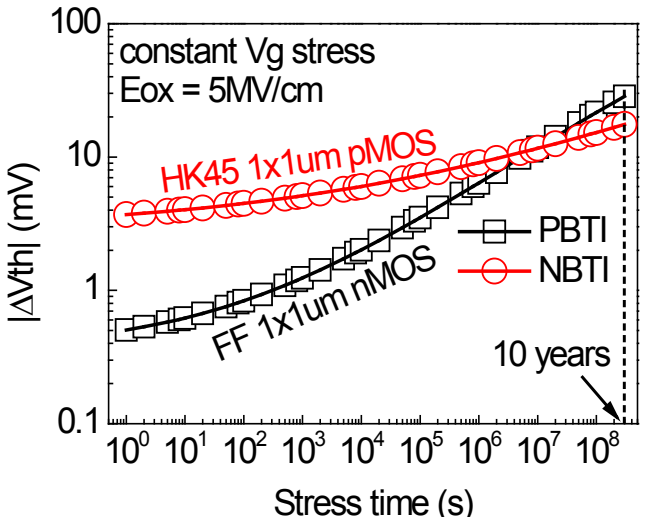

Fig. 5.25 NBTI & PBTI induced total $|\Delta V_{th}|$ under operating condition E_{ox}=5MV/cm predicted by the A-G model. Although PBTI('□') is much smaller compared to NBTI('o') within typical testing time (<10^6s), it increases much faster due to the much larger time exponent of Generated Defects & EAD, and eventually surpasses NBTI at device lifetime 10 years.

5.5 Comparison of the A-G model with other models

As a newly proposed model, it is worthwhile to have a comparison between the A-G model and two other popular models: the Reaction-Diffusion (R-D) model and the Two-Stage model.

The advantages of the A-G model are: 1) it can predict device lifetime accurately with parameters extracted within a short period of testing time; 2). Later in chapter 6 it will be shown this framework can also be well adopted on nano-scaled devices to do the lifetime prediction, which currently cannot be achieved by other models; 3). All the equations in the A-G model are succinct. No complex differential equations need to be solved.

The R-D model is the traditional model. As detailed in chapter 1, the newest R-D framework also considered NBTI with three components, N_{it}, ΔN_{HT} and $\Delta N_{OT.}$ Unlike in the A-G model different components are separated and directly measured, in the R-D framework, ΔN_{it} is measured by Direct-current current-voltage (DCIV) technique [137] and then used to determine the parameters in R-D theory differential equations. After that $\Delta N_{HT,}$ ΔN_{OT} are calculated by empirical equations to best fit the total $|\Delta V_{th}|$. This is questionable to validate the R-D framework predicting capability as the data itself has been used for model parameter extraction, as discussed in [33].

The advantages of the R-D framework lie in its simplicity of test pattern and can be easily integrated with EDA tools. Although lots of studies are against this framework [37, 94], it's still widely used in industry as there is no alternative solution.

T. Grasser proposed the Two-stage model in 2009 [19] to interpret the universal recovery and temperature dependence of NBTI. Although it well explains NBTI recovery, this model is difficult to predict the device lifetime. Later T. Grasser developed Time Dependent Defect Spectroscopy (TDDS) technique to extract the properties of single trap on nano-scaled devices, based on TDDS. Capture Emission Time (CET) mapping method was then proposed to describe the dynamic behavior under BTI condition, however, cannot be extrapolated outside the measurement window for long term lifetime prediction [138].

5.6 Summary

In this chapter, a comprehensive As-grown Generation (A-G) model under BTI stress is proposed. By understanding different types of defects, the A-G model shows good predictive capability for both NBTI & PBTI. To accelerate the model parameter extraction, Fast-Voltage Step Stress method is developed to reduce the testing time and significantly improve the efficiency of model parameter extraction, which lays a solid foundation to extend the A-G model to nano-scale devices, as detailed in the next chapter.

6 Defect-induced Time Dependent Variability modelling

6.1 Introduction & Motivation

Device variability is emerging as a fundamental challenge to IC design in scaled CMOS technology as the technology node migrates to sub- 65nm [139-141]. Due to the profound impact on nearly all aspects of circuit performance, it must be considered for design optimization [142]. Compared with variability reduction in process improvement, the device/circuit co-design approach has recently been considered as an alternative and effective solution, in which the physical-based model with accurate predictive capability becomes a prerequisite [143-145].

Time-zero Variability has been well understood and modelled [146]. Accurate modelling for ageing-induced Time Dependent Variability, however, is still a challenge [57, 118, 138, 147]. Time Dependent Variability increases gradually with time and thus the assessment of the end-of-life variability is a necessity. In current Time Dependent Variability models, Capture-Emission-Time (CET) mapping model well describes the dynamic behavior under BTI condition, however, cannot be extrapolated outside the measurement window for long term lifetime prediction [138]; Defect-centric theory for Time Dependent Variability modelling has been proposed [57, 118], only one type of trap is considered in its classic unimodal model [55], although the model is recently generalized by taking two or more types into consideration [118]. Three key issues,

however, remain unanswered: 1). how many types of defects should be taken into account? 2). how can the parameters for each type be extracted in an efficient way? 3). how reliable is the predicted Time Dependent Variability lifetime? These issues are resolved by a comprehensive As-grown Generation (A-G) framework based on the A-G model and defect centric theory proposed in this chapter.

This chapter will be organized as follows: Firstly in section 6.2 a brief introduction of 2 existing mainstream techniques of Time Dependent Variability characterization and modeling are given, secondly in 6.3 the Within Device Fluctuation (WDF) phenomenon on nano-scale devices is discussed. WDF is one of the main difficulties for nano-scale device characterization. An "Envelope" analysis [148] is used to capture the worst case of BTI degradation to predict device lifetime. Next in section 6.4 the averaged degradation of BTI is investigated. WDF is excluded by taking the averaged IV. It is found that, the mean value of the averaged degradation across multiple nano-scaled, behaves just the same as a big device thus can be well modeled by the A-G model; Device-to-Device Variability of both WDF and averaged degradation is modelled in section 6.5. In section 6.6 the procedure of Time Dependent Variability modelling based on the As-grown Generation (A-G) model together and defect-centric theory is given. The comprehensive modelling is named as "the A-G framework". The A-G framework's predicting capability is proven by the agreement of test data and prediction. After validation, the A-G framework is then implemented into a commercial simulator and its applicability for circuit level simulation is demonstrated in section 6.7. Finally a summary is wrapped up in section 6.8.

6.2 Deficiency of existing techniques for Time Dependent Variability characterization

RTN and Time Dependent Defect Spectroscopy (TDDS) are two mainstream techniques to characterize Time dependent Variability. The test procedure and analysis of these two techniques have already been briefly introduced in section 2.5. Both techniques are really powerful and improve people's understanding on Time Dependent Variability significantly. However, in terms of evaluating the lifetime of Time Dependent Variability on nano-scale devices, both techniques have their deficiencies, as detailed below:

1). Both RTN and TDDS technique need to select devices. RTN tests must be carried out on devices which have clear RTN, TDDS tests also requires the devices to have clear discharging steps during the recovery trace. However, only a small portion of nano-scale

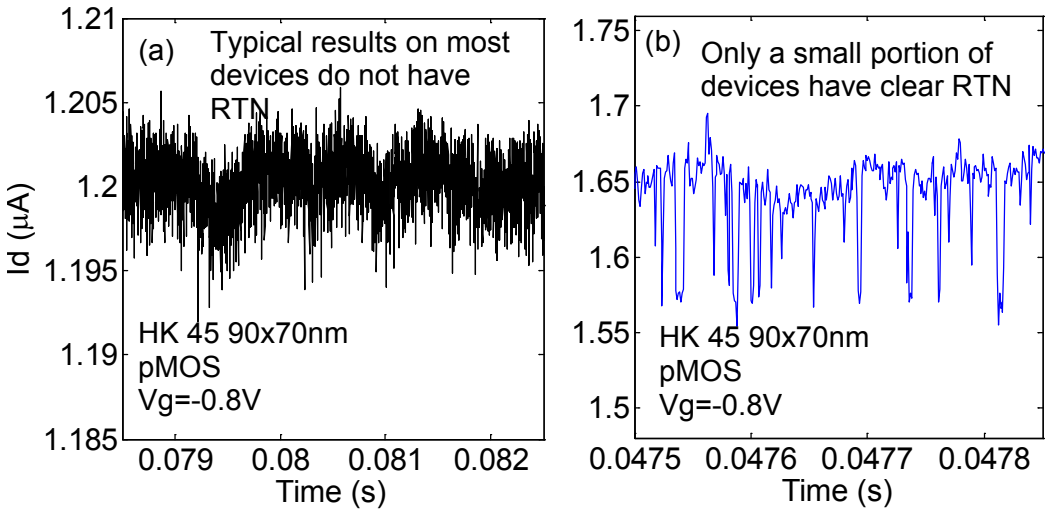

Fig. 6.1 Most nano-scale devices show a stochastic current fluctuation (a) instead of a clear RTN (b). Sampling rate = 1MSa/s is used here.

devices can have clear RTN signals (Fig. 6.1) or I_d drop during TDDS. This introduces uncertainty to the Device-to-Device Variability assessment;

2). The Time Window (Tw), used in a typically RTN test is short (e.g. <=1 second [149-151]) and cannot capture slow traps. Although RTN can be observed within a short Tw, increasing Tw gives a rising fluctuation which must be properly considered.

3). Although many methodologies for modelling RTN have been proposed [1-8], their predictive capability is difficult to validate.

4). Both RTN and TDDS technique are quite time consuming, and nano-scale devices characterization usually needs a lot of repeating tests, thus making these two techniques inefficient for industry use.

Considering the above deficiencies, a test-proven model which can evaluate the lifetime of nano-scale devices within a practical testing time is still missing. A comprehensive A-G framework is proposed in this chapter to fulfill this task.

6.3 Within Device Fluctuation (WDF) technique on nano-scale devices

6.3.1 A review Within Device Fluctuation (WDF) technique

As discussed in section 6.2, only a small portion of devices have clear RTN or Time Dependent Defect Spectroscopy (TDDS) discharging steps, questioning the Device-to-

Device Variability of these two mainstream techniques. To characterize I_d fluctuation on all the nano-scaled devices, M. Duan et al [148, 152] proposed a measurement technique called "Within Device Fluctuation (WDF)", to capture the worst case of nano-scale devices BTI degradation.

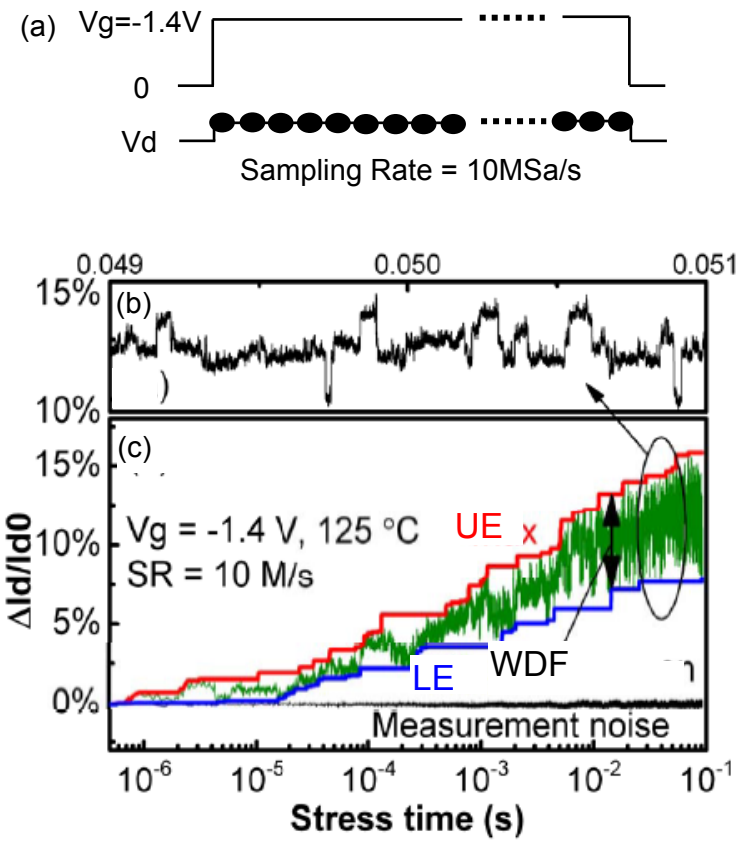

Fig. 6.2 (a) Test waveform of WDF measurement technique. Typical results are shown in (c), the green lines are measured raw data of Id fluctuation, the red and blue lines are extracted UE and LE based on Equation (6.1) & (6.2). (b) is plotted with part of WDF in linear time scale, indicating WDF is formed by a convolution of multiple RTN signals [148, 152].

The test waveform of WDF measurement in [152] is shown in Fig. 6.2a. V_g=-1.4V On-The-Fly (OTF) measurement is carried out on a nano-scale fresh pMOS with Sampling Rate of 10MSa/s, I_d fluctuation is continuously monitored against stress time. To analyze and model I_d fluctuation, also to reduce the data storage, "Envelope" instead of raw Id data

is used to record the worst/best BTI degradation as Upper Envelope (UE) and Lower Envelope (LE), as shown in Fig. 6.2c. UE and LE in [152] are defined as

$$UE(t') = \text{Maximum of } \Delta I_d/I_{d0} \text{ as time increases from0 to t'} \quad (6.1)$$

$$LE(t') = \text{Minimum of } \Delta I_d/I_{d0}, \text{ as time decreases from t(end of stress) to t'} \quad (6.2)$$

WDF is then defined by:

$$WDF = UE - LE \quad (6.3)$$

Fig. 6.2 shows the worst case of BTI degradation can be delivered by the Upper Envelope (UE). Some preliminary UE modelling work [120] has already been done based on the A-G model. It is interpreted that the Lower Envelope (LE) is caused by Generated Defects due to it's very difficult to discharge [129]. WDF is caused by As-grown Traps, as 1). As-grown Traps can have a rapid charging/discharging under a constant V_g; 2). Both WDF and As-grown Traps show they are independent of stress [148, 152]. The averaged LE results across multiple devices can be modelled with a power law. The averaged WDF results are found to follow a logarithmic relationship against time window. The worst case, UE, is then the sum of LE and WDF. However, the modelling in [120] is inadequate to model Time Dependent Variability under use-bias, as 1). There is no V_g dependence for Time Dependent Variability, V_g=-1.4V is not the operating voltage; 2). The verification testing data itself in [120] is used to extract model parameter, questioning its predicting capability; 3). Device-to-Device Variability results for both LE and WDF, have much smaller power exponent than 0.5, thus defect-centric theory cannot be applied to simulate Time Dependent Variability distribution.

To conquer these difficulties a new WDF measurement technique is developed in this chapter as shown in Fig. 6.3a. V_g is firstly applied on the device for 100 seconds to measure the device charging kinetics, each pulse IV (blue line in Fig. 6.3a) is the average of 30 IV measurements. WDF is averaged out by 30 time average thus the measurement results are called the averaged degradation. After 100 seconds charging the Id fluctuation is continuously monitored (red line in Fig. 6.3). Typical results of the new WDF measurement technique are given in Fig. 6.3b. After 100s charging, the device is already stabilized, most of the chargeable traps have already been filled thus further charging is suppressed, thus the new WDF measurement technique separate the charging kinetics and I_d fluctuation. It will be shown later (in section 6.5) that new WDF Device-to-Device Variability can be fitted with a power law whose time exponent is 0.5, and defect centric theory can be applied. To differentiate from the conventional WDF technique, "Uppder Envelope" is called "WDF^+" and "Lower Envelope (LE)" WDF^-, new WDF is defined by

$$\text{new WDF} = WDF^+ - WDF^- \tag{6.4}$$

Unless specified, the terminology "WDF" stands for the "new WDF" in Equation (6.4) hereafter in this chapter.

In the new WDF measurement technique, the worst BTI degradation is delivered by the sum of averaged degradation and WDF^+.

$$\text{worst BTI} = \text{averaged degradation} + WDF^+ \tag{6.5}$$

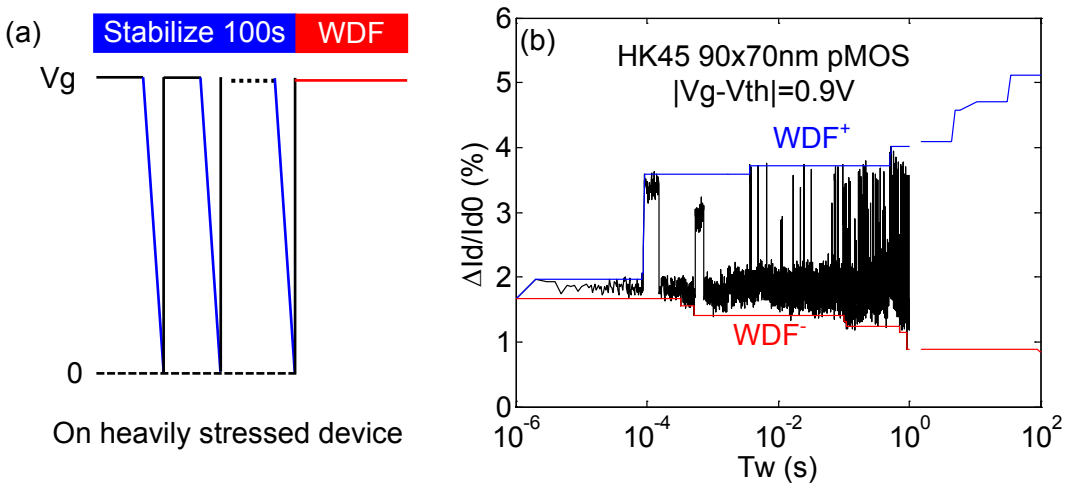

Fig. 6.3 (a) Test waveform of the new WDF measurement technique. (b) Typical results of the new WDF technique.

WDF^+ in $\Delta I_d/I_{d0}$ needs to be converted into $|\Delta V_{th}|$. People normally use $\Delta I_d/g_m$ to do this, however this could lead to a big (as high as 2.5 in [60]) overestimation of the real $|\Delta V_{th}|$ [60]. The correlation between $\Delta I_d/I_{d0}$ and $|\Delta V_{th}|$ measured during the charging kinetics is used to do this conversion, as shown in Fig. 6.4.

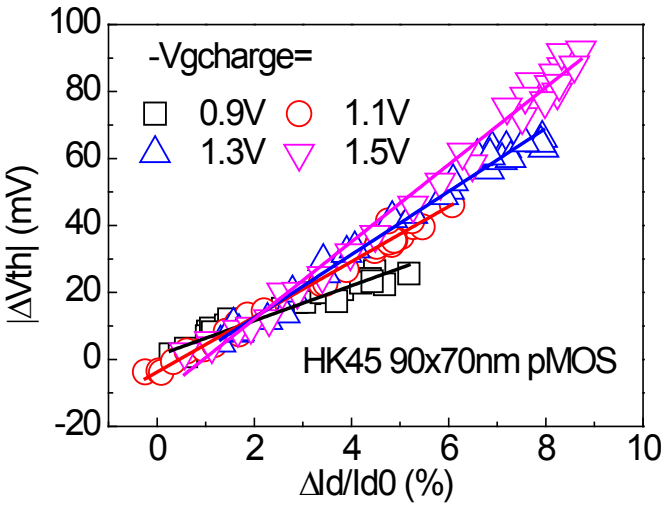

Fig. 6.4 $|\Delta V_{th}|$~ $\Delta I_d/I_{d0}$ relationship extracted from the charging kinetics on a HK45 90x70nm pMOS under different $|V_g-V_{th}|$. Note the conversion is device specified.

6.3.2 Sampling rate dependence

WDF is dependent with measurement sampling rate, since current fluctuation is averaged out under low sampling rate, as shown in Fig. 6.5. When sampling rate < 1MSa/s, WDF is proportional to log(sampling rate). Note Fig. 6.5a is measured on a stressed device thus further BTI is suppressed, as Id remains constant. However, it is found as sampling rate reaches 1MSa/s, WDF saturates vs high sampling rate, indicating 1MSa/s sampling rate can capture the entire WDF. In the following sections, SR=1MSa/s if not specified.

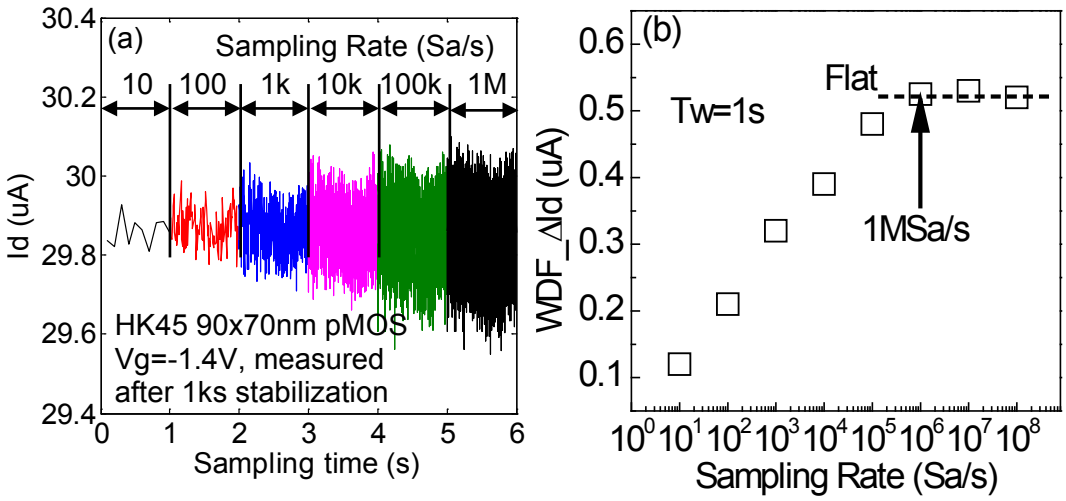

Fig. 6.5 (a) I_d fluctuation increases with sampling rate due to measurement results are averaged out under low sampling rate. Stressed device is used here to suppress further degradation during the measurement. (b). WDF in ΔI_d at 1 second time window (Tw) extracted from (a) is plotted against sampling rate, which indicates 1MSa/s is capable to capture the entire WDF.

6.3.3 Stress dependence

M. Duan et al already shows the conventional WDF is independent of stress time [148, 152]. This conclusion is still valid for the new WDF, as shown in Fig. 6.6. WDF under

V_g=-0.8V is firstly measured on a fresh HK45 90x70nm pMOS, the device is then stressed under $|V_g-V_{th}|$=1.7V for ten seconds (Stress #1), after that the device is floating for five minutes to stabilize and another WDF measurement under V_g=-0.8V is carried out on the stressed device. The same procedure is repeated with a second stress under $|V_g-V_{th}|$=1.7V for one thousand seconds (Stress #2).

Note WDF increases with the measurement Time Window (Tw). A larger Tw allows capturing slower traps, leading to the increase of WDF.

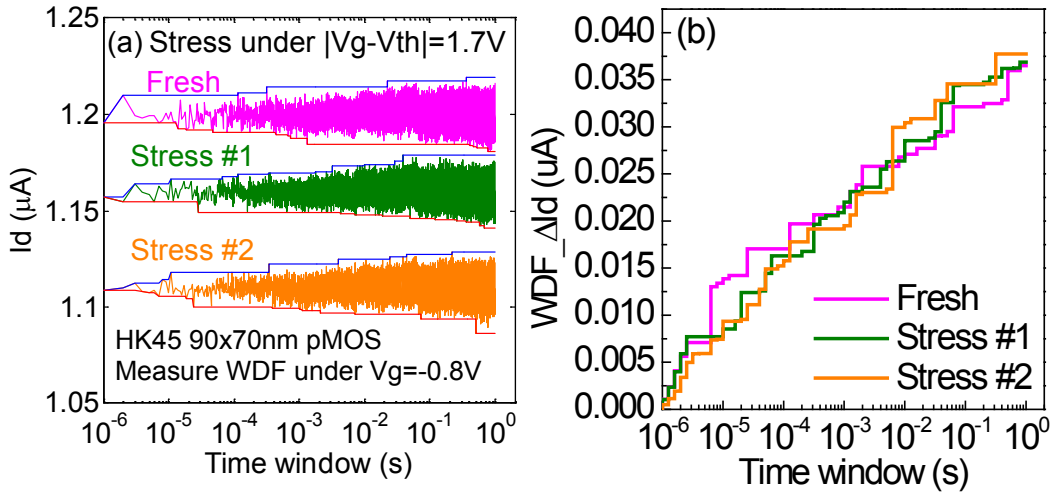

Fig. 6.6 WDF measurement results on fresh and stress device show they are independent of stress. (a) Id fluctuation measured on fresh and stressed device and WDF extraction. (b) Extracted WDF from Id fluctuation in (a) is independent of stress condition.

6.3.4 Frequency & Duty factor dependence

Most circuits work under AC condition. To implement the A-G framework in a circuit simulator, WDF^+ under AC stress condition must be modelled. The test waveform of

WDF measurement as shown in Fig. 6.3a is only applicable to DC stress, as BTI kinetics cannot remains at a constant value under AC stress.

An AC WDF measurement technique is developed as shown in Fig. 6.7b. To better describe the principles of the AC WDF measurement technique, illustration of the DC WDF is also given in Fig. 6.7a as a comparison.

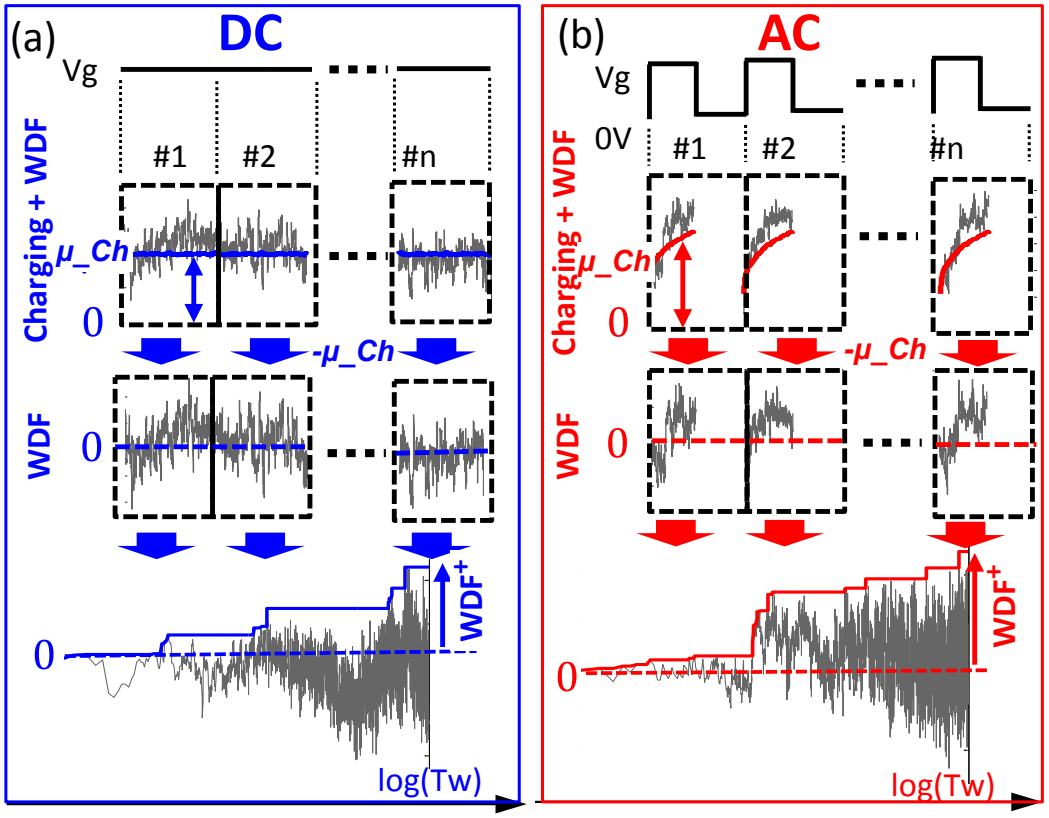

Fig. 6.7 Illustration of WDF extraction under DC (a) and AC (b) NBTI stress. WDF is obtained by subtracting average from total. DC WDF is continuously recorded. AC WDF can be formed by joining each "ON" phase.

AC WDF can only be captured in the ON phase since I_d=0 in the OFF phase. Both DC and AC WDF measurement need to stabilize the traps before monitoring I_d fluctuation. For DC case, after stabilization all the chargeable traps are filled thus further charging is negligible, WDF can be directly measured as illustrated in Fig. 6.3a. However, under AC stress condition, V_g is alternating all the time, traps are repetitively charge/discharge during the ON/OFF phase. To separate the charging and WDF during the ON phase, measured $|\Delta V_{th}|$ raw data (converted from ΔI_d in the ON phase) from different AC cycles is averaged out to achieve an Averaged Charging kinetics (μ_Ch in Fig. 6.7). WDF is averaged out by doing this average. After μ_Ch is extracted, AC WDF in each cycle is then calculated by subtracting μ_Ch from the measured $|\Delta V_{th}|$ raw data. WDF^+ in all the cycles during one AC WDF measurement is then concatenated and plotted against equivalent time window (Time window * Duty actor). AC WDF^+ is extracted.

WDF^+ is different with multiple measurements due to the stochastic (dis-)charging of the defects. Fig. 6.8a shows 100 AC WDF^+ measurements (grey lines), the average value of these 100 results (μ_WDF^+) follows a good logarithmic relation against time window. Fig. 6.8b shows μ_WDF^+ is independent of Frequency and duty factor, which indicates WDF^+ under AC stress can be directly obtained from DC measurements after taking duty factor into account. It is speculated that this is due to the probability for charging a trap only depends on the accumulative time under a given $|V_g-V_{th}|$.

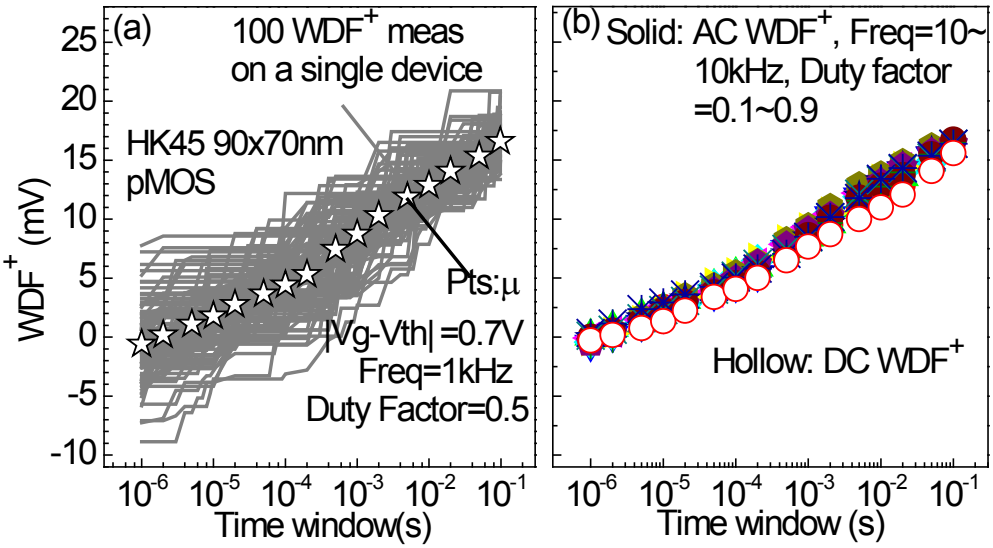

Fig. 6.8 (a) Multiple WDF^+ measurements and their averaged value under AC NBTI. (b) Averaged WDF^+ overlaps for DC and AC of different frequency and duty factor.

6.3.5 Within Device Fluctuation (WDF) modelling

DC WDF^+ measurements under different $|V_g\text{-}V_{th}|$ are carried out on 50 stressed HK45 90x70nm pMOSFETs, as shown by the grey lines in Fig. 6.9a. The mean value of WDF^+, μ_WDF^+, as shown by the red circles, can be empirically fitted with Equation (6.6).

$$\Delta V_{th}\_WDF^+ = p_3 \cdot \exp\left(p_4 \cdot |V_g - V_{th}|\right) \cdot \log_{10}(t_w/t_0) \qquad (6.6)$$

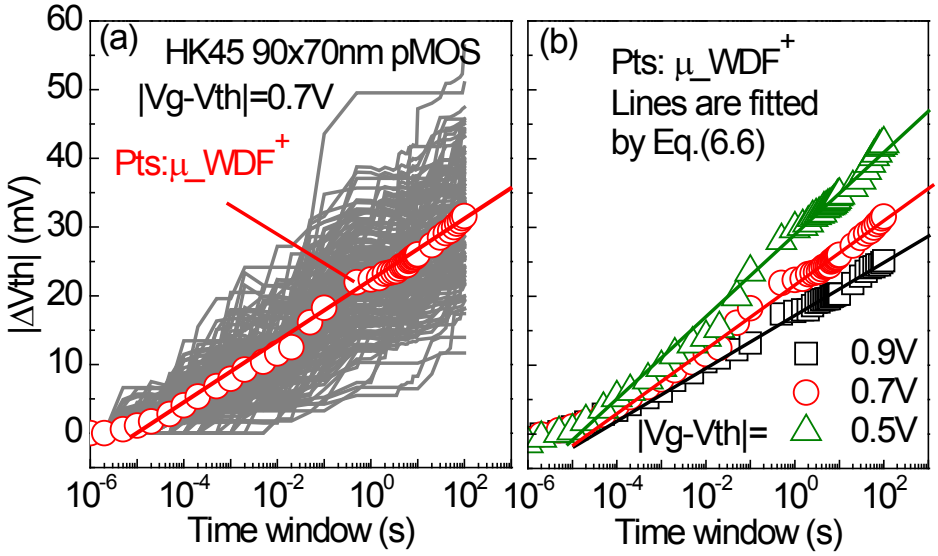

Fig. 6.9 (a) DC WDF^+ measurement results from multiple HK45 90x70nm pMOSFETs and its average value under $|V_g-V_{th}|$=0.7V. (b) Mean value of WDF^+ (μ_WDF^+) under different $|V_g-V_{th}|$ is fitted with Equation (6.6).

6.4 Application of the A-G model on the averaged degradation

With the new measurement pattern and analysis of WDF^+ as shown in Fig. 6.7, the worst case of BTI degradation is WDF^+ plus the averaged degradation, under both DC and AC stress condition. Averaged degradation can be measured by averaging IV curves to exclude WDF. For simplicity, in the following contents, different types of defects like Generated Defects, As-grown Traps and Energy Alternating Defects on nano-scale devices represent their averaged degradation, the terminology "averaged degradation" are omitted.

6.4.1 Generated Defects modelling on nano-scale devices

Normally on a mature process, multiple nano-scale devices BTI after averaging will have the same amount of $|\Delta V_{th}|$ as a big device [153]. However, the sample tested in this chapter is a developing process and nano-scale devices show much higher $|\Delta V_{th}|$ than big devices. But the time exponent of Generated Defects remains the same (Fig. 6.10), this also agrees with what IMEC reported that the time exponent, n, of Generated Defects is independent of device geometry [154]. With a known time exponent, Fast-Voltage Step Stress technique can be applied to reduce the testing time.

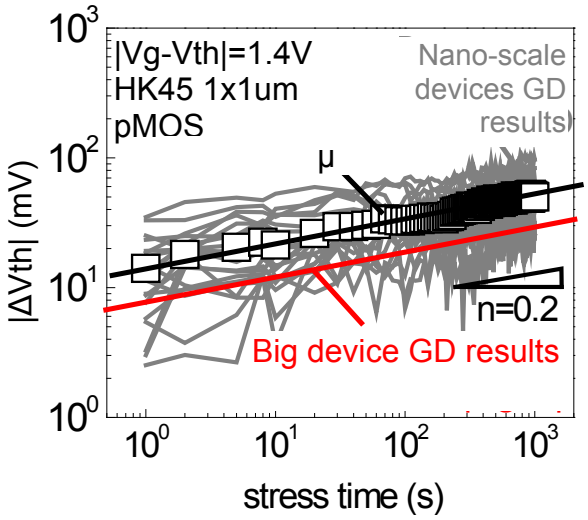

Fig. 6.10 A comparison of Generated Defects (GD) results on nano-scale devices and on a big device. Stress-Discharge-Recharge technique is adopted. It is shown that the averaged Generated Defects on nano-scale devices and has the same time exponent of 0.2 as on a big device.

Fig. 6.11 shows typical results of applying Fast-Voltage Step Stress pattern (as detailed in section 5.3.3) on a single pMOS device. 100 IV curves are captured at the end of recharging phase with 100 milliseconds interval to save the data on oscilloscope. V_{th} fluctuation of these 100 IV curves reflects the WDF under the recharge V_g (V_{grech}).

Probability plot of these 100 $|\Delta V_{th}|$ after different $|V_g-V_{th}|$ is given in Fig. 6.11a, the parallel shift again confirms the independency of WDF on NBTI stress. The flat plateau

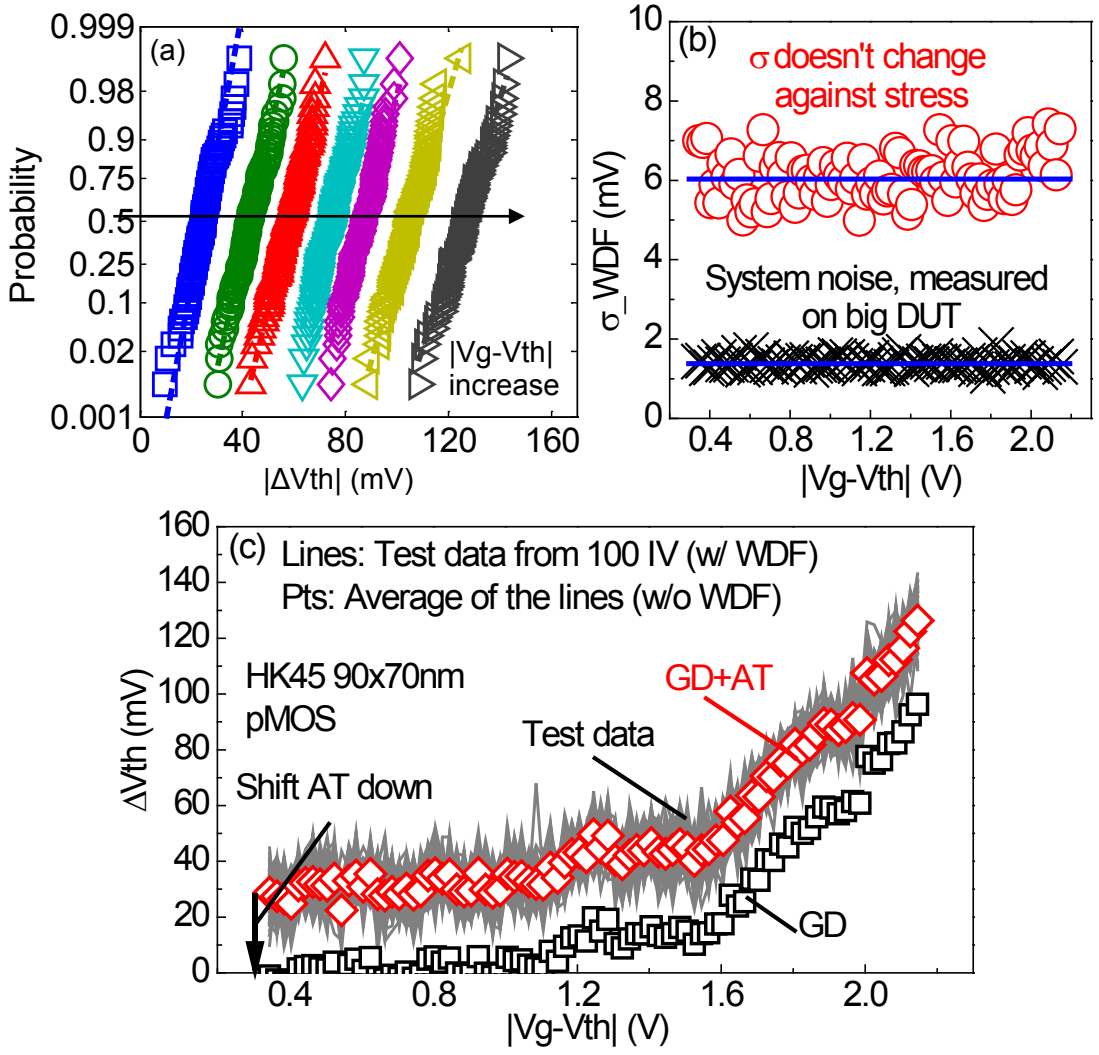

Fig. 6.11 Typical results of Generated Defects (GD) measured on a **SINGLE** nano-scale device. (a) Statistical Within-Device-Fluctuation distrbiution of $|\Delta V_{th}|$ from 100 IV curves measured after each $|V_g-V_{th}|$ step. (b) The standard deviation for the data in (a) ('o') and the data from large device ('x'). σ_WDF from nano-scaled device is much larger than the one from large device which is dominated by system noise. In addition, σ_WDF changes little with stress $|V_g-V_{th}|$, confirming that it is dominated by As-grown Traps (AT). (c) For each $|V_g-V_{th}|$, the 100 $|\Delta V_{th}|$ were ploted as lines and their mean value as '◊'. Generated Defects ('□') were obtained by subtracting As-grown Traps in the flat plateau.

due to As-grown Traps under V_{grech} is then subtracted to get Generated Defects induced $|\Delta V_{th}|$.

The same test procedure in Fig. 6.11 is then repeated to measure Generated Defects on multiple nano-scale devices, as shown in Fig. 6.12. Their average value is then used to extract the parameters of Generated Defects.

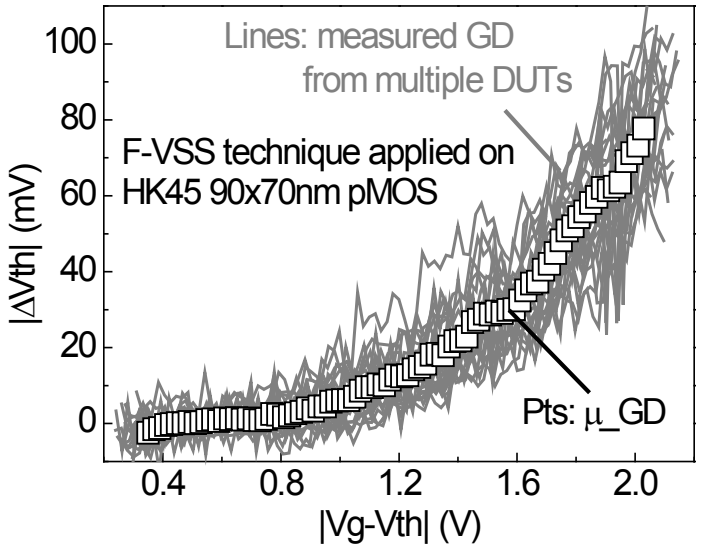

Fig. 6.12 Generated Defects measured on multiple HK45 90x70nm pMOSFETs. Each grey line represents Generated Defects kinetics in Fig. 6.11. The points show the average value of all the grey lines.

With apparent time exponent n=0.2 from the big device, the same analysis procedure as described in section 5.3.1 can be applied on μ_ GD for g_1 and m_1 extraction, as shown in Fig. 6.13a&b. μ_ GD is then modelled by

$$\mu\_GD = g_1 \cdot \left(V_g - V_{th}\right)^{m_1} t_{str}^{n_1} \tag{6.7}$$

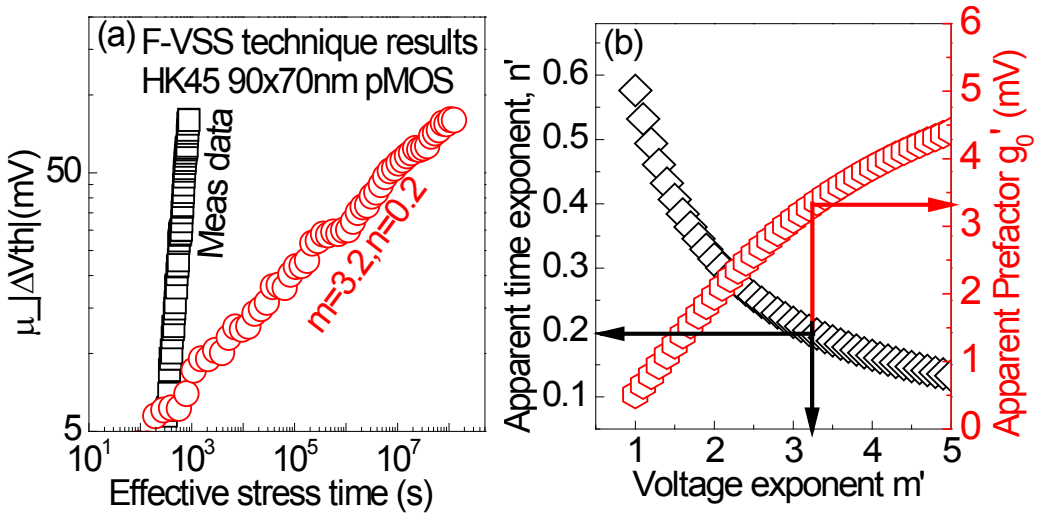

Fig. 6.13 (a) Generated Defects are ploted against real stress time ('□') and effective stress time ('**o**'), defined by Eqn (2) which is a function of voltage exponent m. m is selected to give a time exponent of 0.2. (b) The apperaent time exponent, n', prefactor, g0' under any given voltage expoenet, m'. The real m and g0' value can be determined when the apparent n' equals to 0.2, which is the real time exponent extracted from big device.

6.4.2 Pre_Existing defects modelling on nano-scale devices

After applying the Fast-Voltage Step Stress technique to extract Generated Defects parameters, the Pre_Existing defects charging and discharging kinetics can be measured on these stressed devices whose further generation is suppressed. Fig. 6.14a&d show μ_Pre_Ex traps charging/discharging results after averaging on multiple HK45 90x70nm pMOSFETs, Fig. 6.14b/e show the $|V_g-V_{th}|$ dependence of μ_Pre_Ex.

Due to the stochastic behavior of nano-scale devices m-DMP technique cannot be applied on nano-scale devices, μ_AT profile in Equation (6.7) thus cannot be directly measured. But Energy Alternating Defects' time exponent n_2 is already known from big device. With

a known time exponent n_2, Pre_Existing defects' charging kinetics can be directly fitted with Equation (6.11). The discharging kinetics of Pre_Existing Defects can be described with Equation (6.12). The parameters of Pre_Existing defects are summarized in Table 6.1 in section 6.6.

$$\mu\_AT = p_1 \cdot \exp[p_2 \cdot (V_g - V_{th})] \tag{6.8}$$

$$\mu\_AT_{tch} = AT \cdot (1 - e^{-\left(\frac{t_{ch}}{\tau}\right)^{\gamma}}) \tag{6.9}$$

$$\mu\_EAD = g_2 \cdot (V_g - V_{th})^{m_2} t^{n_2} \tag{6.10}$$

$$\mu\_Pre\_Ex_{tch} = \mu\_AD\_AT_{tch} + \mu\_AD\_EAD_{tch} \tag{6.11}$$

$$\mu\_Pre\_Ex_{tdisch} = Pre\_Ex_{tdisch=0} \cdot \left(1 + B \cdot t_{disch}^{\beta}\right)^{-1} \tag{6.12}$$

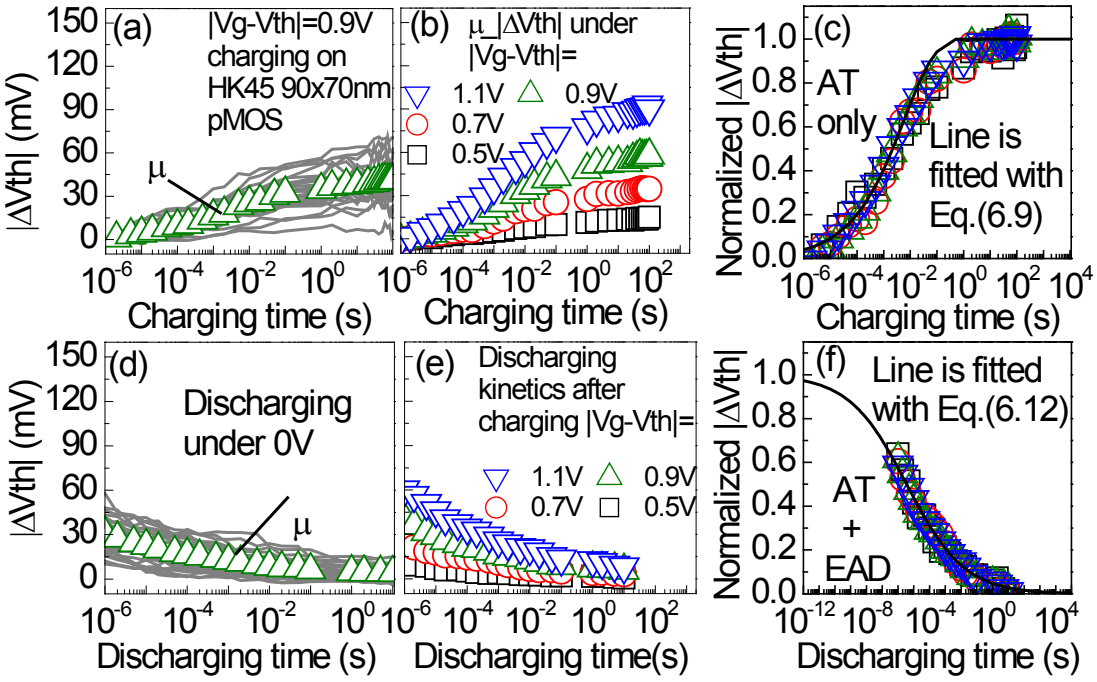

Fig. 6.14 Measured Device-to-Device Variability from charging (a) and discharging (d) kinetics and their averaged value under different $|V_g\text{-}V_{th}|$ (b&e). The normalized kinetics for charging (c) and discharging (f) can be well fitted with Equation (6.9) & (6.12), similar to the large devices.

6.5 Device-to-Device Variability modelling

Nano-scaled devices show significant Device-to-Device Variability. People normally used the standard deviation (σ) versus average (μ) plot to characterize Device-to-Device Variability [59, 120, 153]. Both theoretical simulation [57] and experiment data [58, 59] shows σ~μ follows a power law and the power factor is 0.5. The single trap impact, η, can be extracted using Equation (6.13).

$$\eta = \frac{\sigma^2_{\Delta Vth}}{2 \cdot \mu_{\Delta Vth}} \tag{6.13}$$

6.5.1 Device-to-Device Variability of WDF^+

Device-to-Device Variability of WDF^+ can be plotted as the σ value of WDF^+ at different time window versus the corresponding μ value, as shown in Fig. 6.15b. Fig. 6.15a is a replot of Fig. 6.9a which is the source data of the σ and μ value in Fig. 6.15b. Power exponent of 0.5 is achieved for σ~μ, indicating WDF^+ can be described by the defect centric theory and the average impact of a single WDF^+ η_{WDF+} can be extracted using Equation (6.13). WDF^+ Device-to-Device Variability under different $|V_g-V_{th}|$ can be fitted with the same line, indicating η_{WDF+} is independent of $|V_g-V_{th}|$.

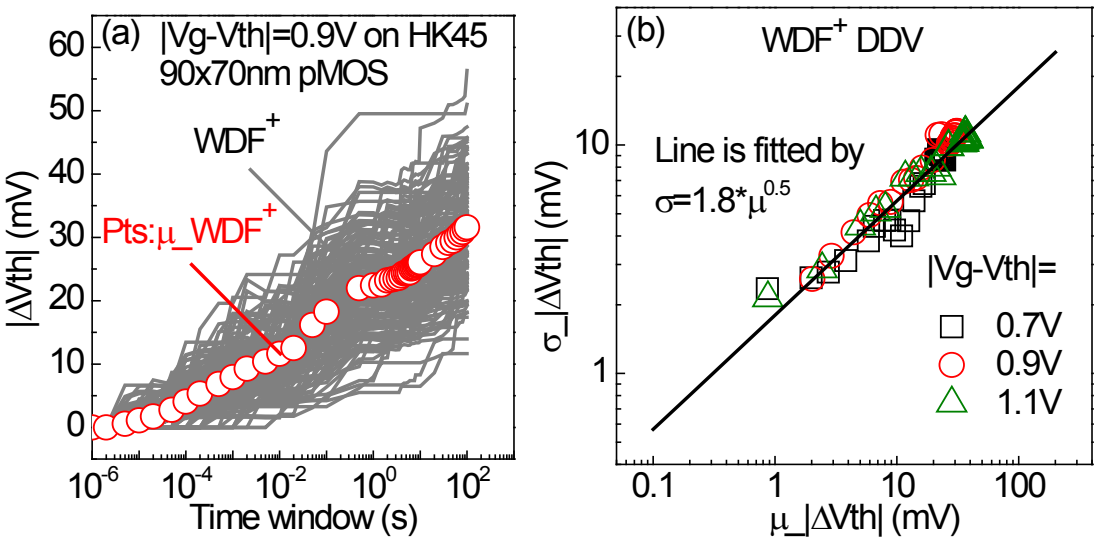

Fig. 6.15 (a) A replot of Fig. 6.9a. WDF^+ measurement results (grey lines) against time window (T_w) under $|V_g-V_{th}|$=0.9V on multiple devices. (b) Plot the standard deviation (σ) of WDF^+ from (a) against the average WDF^+ to extract the average impact of a single defect (η) of WDF^+. Another two measurement condition $|V_g-V_{th}|$=0.7V and 1.1V results are also plotted together with $|V_g-V_{th}|$=0.9V. The same fitting equation indicates η is independent of $|V_g-V_{th}|$.

6.5.2 Device-to-Device Variability of the averaged degradation

Following the same analysis as Fig. 6.15, Defects Device-to-Device Variability of Generated Defects can be extracted from the Fast-Voltage Step Stress, as shown in Fig. 6.16. Pre_Existing defects' Device-to-Device Variability can be extracted from the charging and discharging kinetics on stressed devices, as shown in Fig. 6.17. They can both be well fitted with a power law whose power exponent is 0.5, indicating they also follow the defect centric theory.

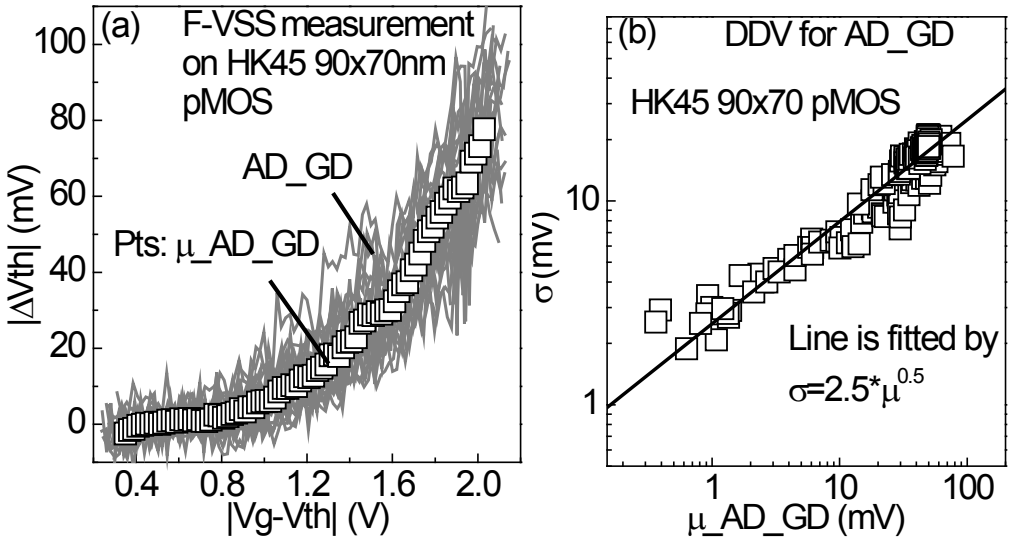

Fig. 6.16 (a) A replot of Fig. 6.12. Generated Defects results of Fast-Voltage Step Stress technique. (b) Plot σ~μ from (a) to extract η of Generated Defects.

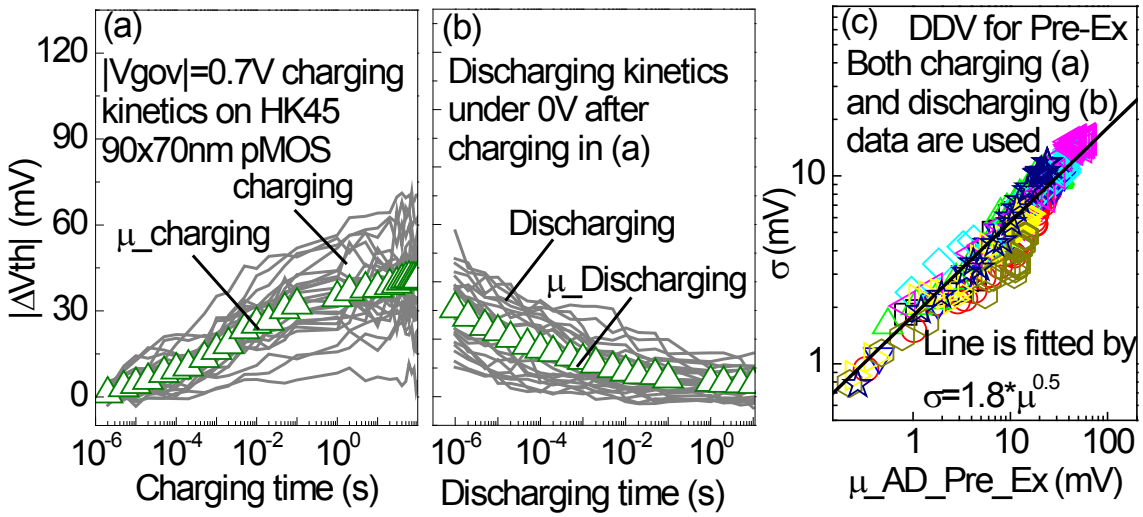

Fig. 6.17 A replot of Fig. 6.14a Pre_Existing defects charging kinetics (a) and Fig. 6.14d discharging kinetics under 0V (b) σ~μ plot from both charging and discharging under different $|V_g-V_{th}|$ is given in (c), which follow a power law well with an exponent of 0.5.

η of Generated Defects, Pre_Existing defects and WDF^+ are summarized in Fig. 6.18 under both NBTI and PBTI stress. Fig. 6.18 shows Generated Defects has a much larger η

compared to the Pre_Existing defects. It is speculated that this is because Generated Defects are more likely located above the current percolation path due to channel-carrier assisting defect generation, as illustrated in Fig. 6.19. η reported by Time Dependent Defect Spectroscopy technique on the same process [155] is similar to that of the Pre_Existing defects, because TDDS only captures the dischargeable traces which are dominated by Pre_Existing defects.

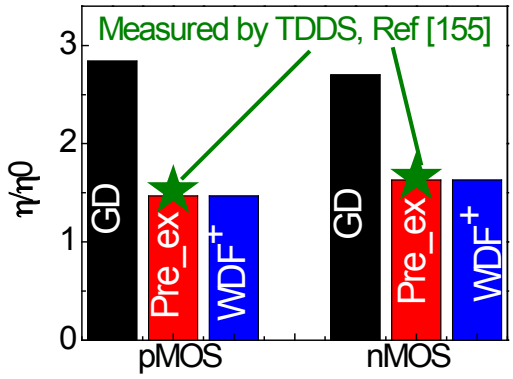

Fig. 6.18 Trap impact, η/η_0, for Generated Defects, Pre_Existing defects and WDF^+ for pMOS and nMOS. η_0 is the impact of a single charge in the charge sheet approximation calculated by $\eta_0=q/C_{OX}$, where q is the elementary charge and Cox the gate oxide capacitance in inversion. η/η_0 extracted using Time Dependent Defect Spectroscopy (TDDS) [29] is also plotted for comparison.

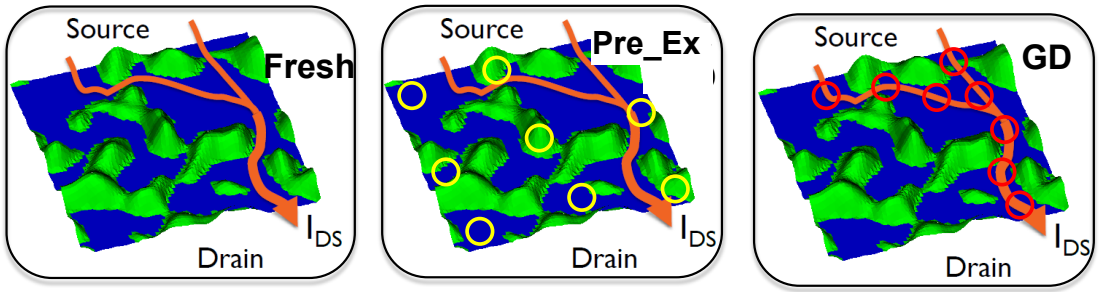

Fig. 6.19 Illustration of a potential explanation for the larger average impact of a single defect (η) of Generated Defects compared with Pre_Existing defects. Generated Defects are more likely located above the current percolation path due to channel-carrier assisting defect generation, resulting in a higher η compared with the random distributed Pre_Existing defects.

6.6 Time Dependent Variability modelling based on the A-G model and Defect-Centric theory and experimental validation

Defect-centric theory is proposed by B. Kaczer [55] in 2010 and has been widely adopted to evaluate the Time Dependent Variability [52, 54, 156]. The average impact of a single defect η is a fundamental parameter determining the variability of nano-scale devices. η is inversely proportional to device area and can be extracted from Device-to-Device Variability whose power exponent is 0.5. Impact of a single defect within a specific follows an exponential distribution and the trap number on multiple devices follows a Poisson distribution.

Although defect centric theory is based on the observation of TDDS measured traps which does not include Generated Defects, Fig. 6.16b shows the Device-to-Device Variability of Generated Defects can also be fitted with power exponent of 0.5 hence defect centric theory is still applicable.

The mean value of the averaged degradation on multiple nano-scale devices can be modelled by the A-G model, as detailed in chapter 5 on a big device. μ_WDF^{+} has already been modelled in section 6.3. Together with defect-centric theory, Time Dependent Variability on each device can be well modelled.

Generated Defects are firstly used as an example to show the procedure of constructing Time Dependent Variability on multiple pMOSFETs by statistical simulation. The same procedure is then repeated on Pre_Existing defects and also WDF^{+}. By adding Generated Defects, Pre_Existing defects and WDF^{+} together, the worst case of BTI degradation is achieved and device lifetime can be extracted.

The procedure for the statistical simulation of Generated Defects is illustrated in Fig. 6.20.

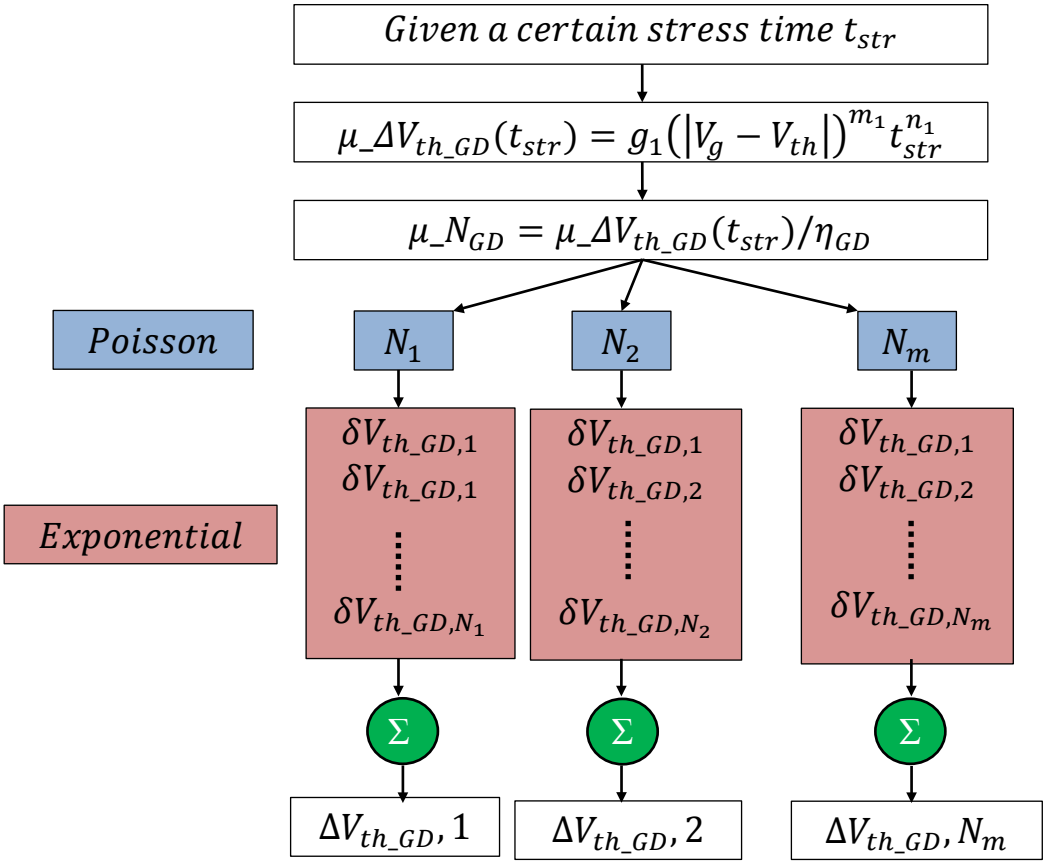

Fig. 6.20 Procedure to construct Generated Defects' distribution with the defect centric theory. This procedure is irrelavant to defect preoperty thus is also applicable to As-grown Trap, Energy Alternating Defects and WDF^{+}.

With the experimentally extracted parameters in Table 6.1, $\mu\_\Delta V_{th\_GD}$ and its corresponding defect number μ_N_{GD} at a given $|V_g\text{-}V_{th}|$ and stress time can be calculated by Equation (6.13) and (6.14), respectively.

$$\mu\_\Delta V_{th\_GD} = g_1 \cdot \left(|V_g - V_{th}|\right)^{m_1} t_{str}^{n_1} \tag{6.14}$$

$$\mu\_N_{GD} = \mu\_\Delta V_{th}\_GD/\eta_{GD} \tag{6.15}$$

In Fig. 6.20 M nano-scale devices are simulated. For each device, the number of Generated Defects is generated by randomly choose an element from a Poisson distribution with an average value of μ_N_{GD}. For each defect on this device, the impact of this defect is generated by randomly choose an element from an exponential distribution with an average value of η_{GD}. The total Generated Defects for a single device is obtained by summing the impact of all the Generated Defects.

A typical simulation result is given in Fig. 6.21a. Each '+' represents Generated Defects on a single nano-scale device under $|V_g\text{-}V_{th}|$=1.5V DC NBTI stress at the specified stress time. 3,000 devices are simulated as there are 3,000 '+' points for each stress time. The μ value of these 3,000 '+' points is plotted as '**o**', which represents $\mu\_\Delta V_{th\_GD}$ kinetics. To verify the accuracy of the statistical simulation, $\mu\_\Delta V_{th\_GD}$ and $\sigma\_\Delta V_{th\_GD}$ from the simulation is compared with the independently measured values in Fig. 6.21b&c. For the measurement, 20 nano-scale devices were stressed under a constant $|V_g\text{-}V_{th}|$=1.5 V using Stress-Discharge-Recharge technique. Good agreement between the simulation and test data for both $\mu\_\Delta V_{th\_GD}$ (Fig. 6.21b) and $\sigma\_\Delta V_{th\_GD}$ (Fig. 6.21c) justifies the simulation procedure and confirms the A-G model's predicting capability on Generated Defects.

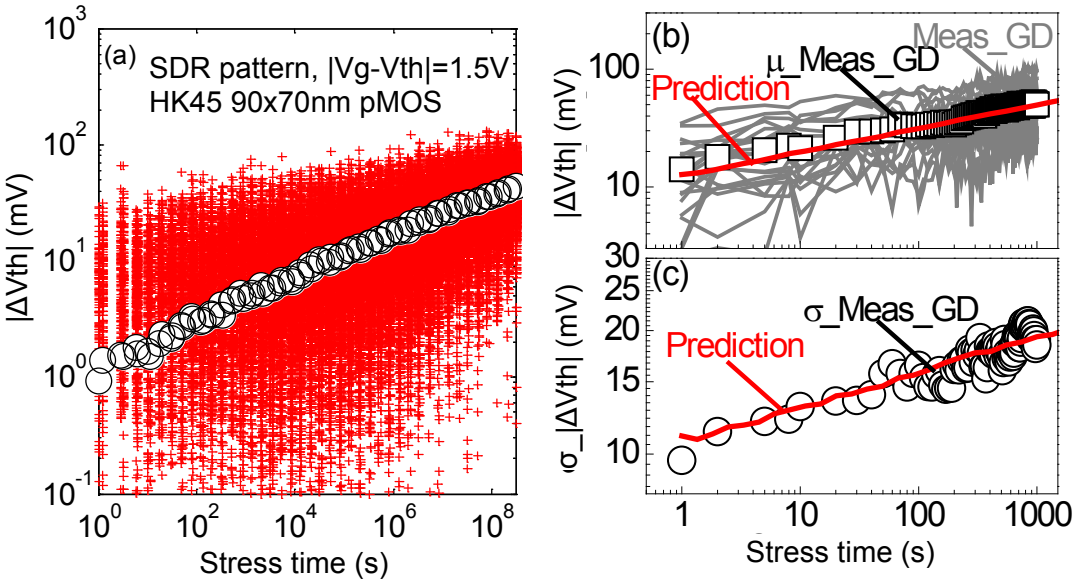

Fig. 6.21 (a) Simulated kinetics of Generated Defects (GD) under $|V_g-V_{th}|$=1.5V NBTI. Each '+' represents one simulated device. The 'o' at each stress time is the average of 3000 devices. Good agreement is achieved between the test data and simulaton of (b) $\mu\_\Delta V_{th\_GD}$ and (c) $\sigma\_\Delta V_{th\_GD}$.

Note the test data in Fig. 6.21b&c were not used for the A-G model parameter extraction. Probability plot of Fig. 6.21 comparison is also given in Fig. 6.22. Note the simulation can capture the "tail" in the distribution which is a big concern for circuit reliability.

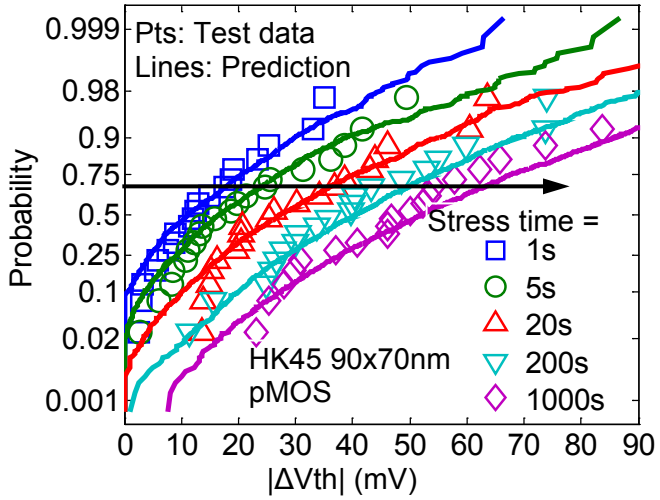

Fig. 6.22 Probability plot of data in Fig. 6.21a. The distribution of the test data (symbols) can be well predicted by the simulation results (lines).

Following the same procedure in Fig. 6.20, As-grown Traps, Energy Alternating Defects and WDF^{+} distribution can be simulated using the A-G model and defect centric theory. The worst case of Time Dependent Variability on nano-scale devices can be simulated following the procedure as shown in Fig. 6.23. This Time Dependent Variability modelling methodology based on the A-G model and defect centric is named as the "A-G framework", as illustrated in Fig. 6.24. Extracted parameters on HK45 process is given in Table 6.1.

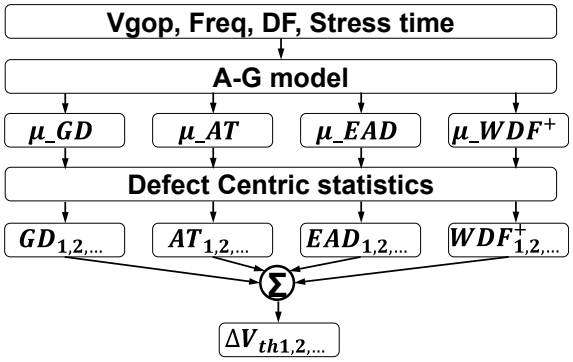

Fig. 6.23 Procedure to simulate Time Dependent Variability on nano-scale devices based on the A-G framework. "Defect Centric statistics" represent the entire procedure in Fig. 6.20.

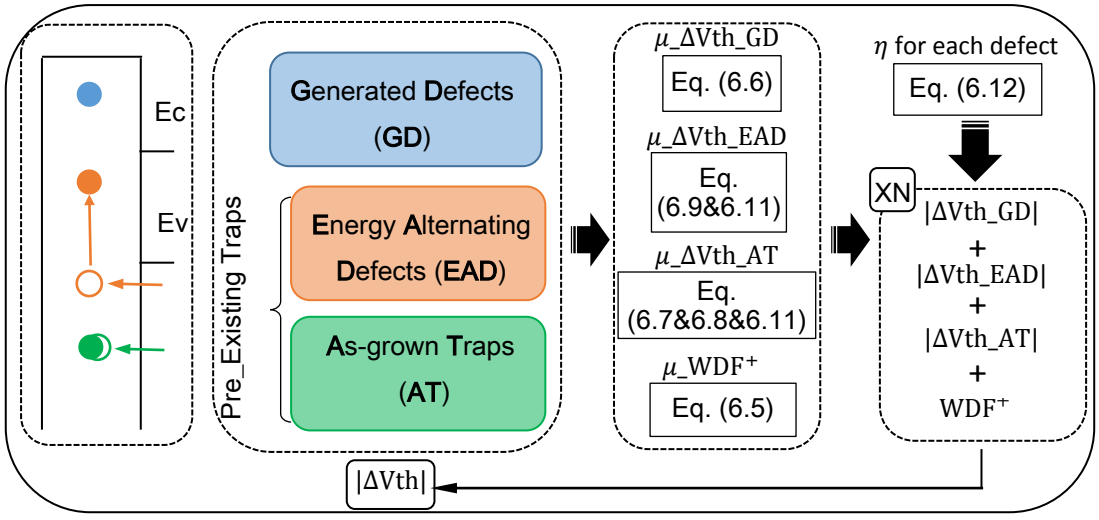

Fig. 6.24 Illustration of the comprehensive A-G framework based on the A-G model and defect centric theory.

Table 6.1 Extracted parameters of the A-G framework under NBTI/PBTI stress on HK45 90x70nm pMOSFET/nMOSFETs. All the $|\Delta V_{th}|$ related parameters (g_1, g_2, p_1, p_3, η_{GD}, η_{Pre_Ex}, η_{WDF+}) are in millivolts and timing related parameters (τ, t_0) are in seconds.

Defects	GD		EAD		AT				Discharge		WDF^+	
pMOSFET parameters	g_1	3.31	g_2	10.0	p1	3.46	τ	5.80e-3	B	19.88	p_3	0.76
	m_1	3.24	m_2	1.80	p2	2.89	γ	0.36	β	0.24	p_4	1.28
	n_1	0.2	n_2	0.1							t_0	1e-5
	η_{GD}	3.13	η_{Pre_Ex}	1.62							η_{WDF+}	1.62
nMOSFET parameters	g_1	0.38	g_2	0.27	p1	2.71	τ	2.00e-4	B	0.79	p_3	0.40
	m_1	4.03	m_2	5.65	p2	0.35	γ	1.0	β	0.13	p_4	1.27
	n_1	0.32	n_2	0.2							t_0	1e-5
	η_{GD}	4.05	η_{Pre_Ex}	2.45							η_{WDF+}	2.45

The predicting capability of the A-G framework needs to be verified by comparison between the prediction and test data. $|V_g|$=1.3V 10 kHz Duty factor=0.5 AC NBTI/PBTI stress is applied on multiple HK45 90x70nm pMOSFETs/nMOSFETs to measure the averaged $|\Delta V_{th}|$. Monto-Carlo simulation is performed to predict the degradation on 1,000 nano-scale devices. Good agreement has been achieved for both averaged degradation (Fig. 6.25a&c) and its distribution (Fig. 6.25b&d), confirming the predicting capability of the A-G framework. Note 1). All the parameters in Table 6.1 were extracted from the measurements under accelerated DC constant $|V_g-V_{th}|$ stresses only, and the AC test data in Fig. 6.25 is measured under constant $|V_g|$ stress; 2). The test data is not used for model parameter extraction.

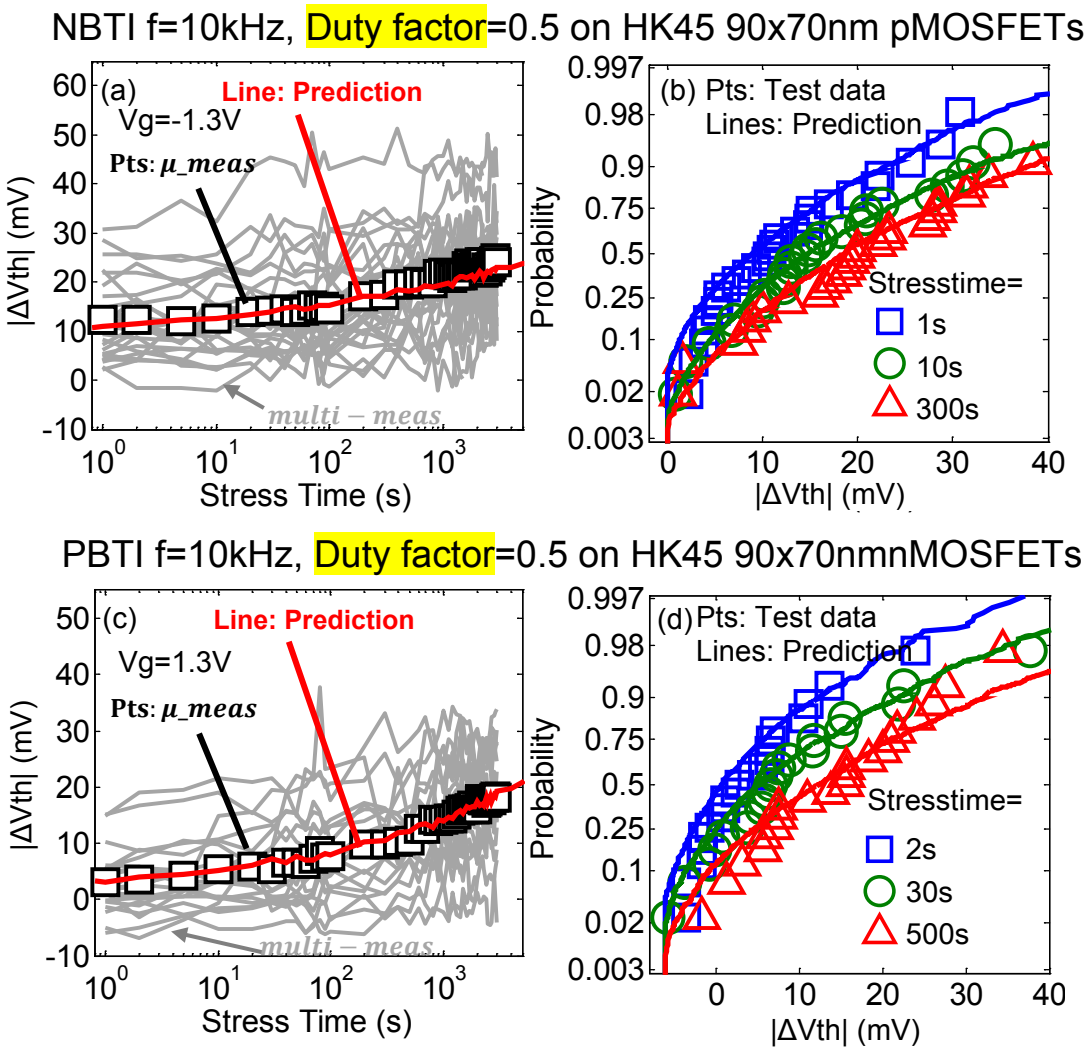

Fig. 6.25 Demonstration of the A-G frameowrk's predicting capability on HK45 90x70nm devices under AC NBTI (a&b) and PBTI (c&d) stress. a&c compare the averaged value ('□')from multiple device measurements (grey lines) and the prediction from 1000 Monto-Carlo simulation with the A-G framework considering both Time-zero Variability and Time Dependent Variability. b&d compares the distributions at different stress time.

6.7 Device & Circuit lifetime Prediction under use bias

With the extracted parameters in Table 6.1, $|\Delta V_{th}|$ distribution of each component under use-bias ($|V_g-V_{th}|$=0.5V) AC stress for both NBTI and PBTI at device lifetime can be

predicted based on the A-G framework, as shown in Fig. 6.26. Note each line is calculated by adding a new component to the neighboring left line, so it is the gap between two curves represents the contribution of a specified component.

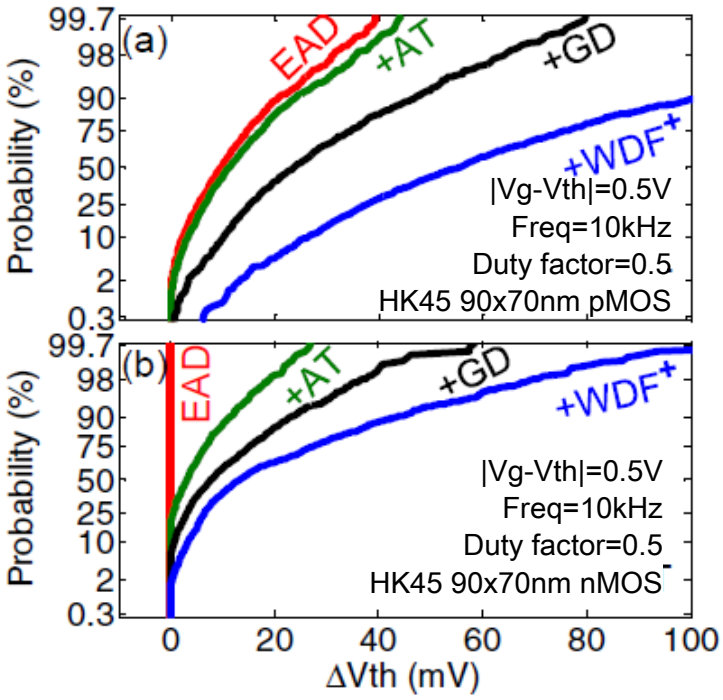

Fig. 6.26 10-year degradation under AC NBTI (a) and PBTI (b) with contribution from different type of defects. $|V_g-V_{th}|$=0.5V, Freq=10kHz, Duty factor=0.5. 1000 Monto-Carlo simulations were performed.

Fig. 6.26 shows WDF^+ (blue-black) brings a significant $|\Delta V_{th}|$ on top of the BTI averaged degradation. This agrees with people's observation on RTN [86, 157] that Id fluctuation is becoming more and more important as MOSFET feature size scaling down and $|V_{dd}|$ lowering down. Generated Defects (black-green), which is usually ignored, also plays a very important role after 10 years. This is due to their larger time exponent compared to Energy Alternating Defects. Energy Alternating Defects and As-grown Traps are supposed to be the majority defects studied in Time Dependent Defect Spectroscopy

(TDDS) technique. They are also very important in the total $|\Delta V_{th}|$ after 10 years but not the dominating components.

The proposed A-G framework is then implemented into the commercial circuit simulator HSPICE (Simulation Program with Integrated Circuit Emphasis) for variability-aware analysis. A 31-stage Ring Oscillator (RO) is used for demonstration. Time-zero Variability introduces fresh threshold voltage (V_{th0}) variation. By directly measuring Time-zero Variability distribution on multiple fresh devices [158], both Time-zero Variability and Time Dependent Variability are taken into consideration. The frequency distribution at fresh and 10 years can be assessed, as shown in Fig. 6.27. It is also found that PBTI can introduce extra degradation and should be included in the simulation.

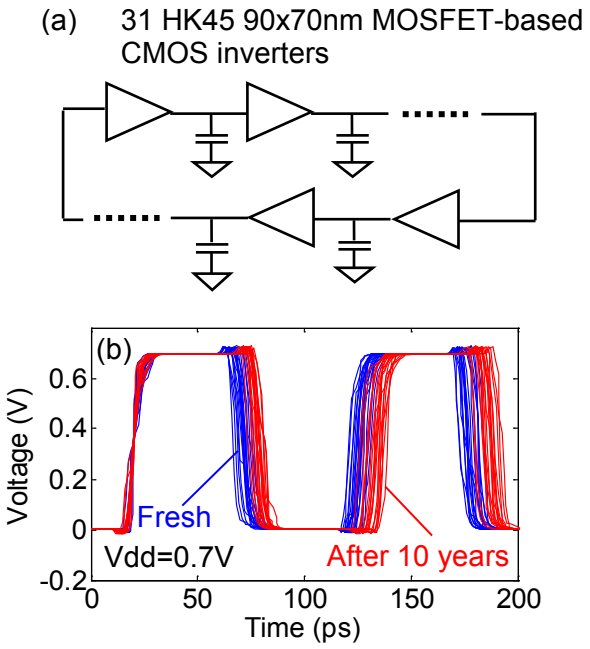

Fig. 6.27 (a) A typical 31-stage Ring Oscillator (RO) circuit used for the HSPICE circuit simulation. (b) Snapshot of the waveform from one node of the RO before and after 10 years under the operating voltage V_{dd}=0.7V. 1000 run Monto-Carlo simulation is used.

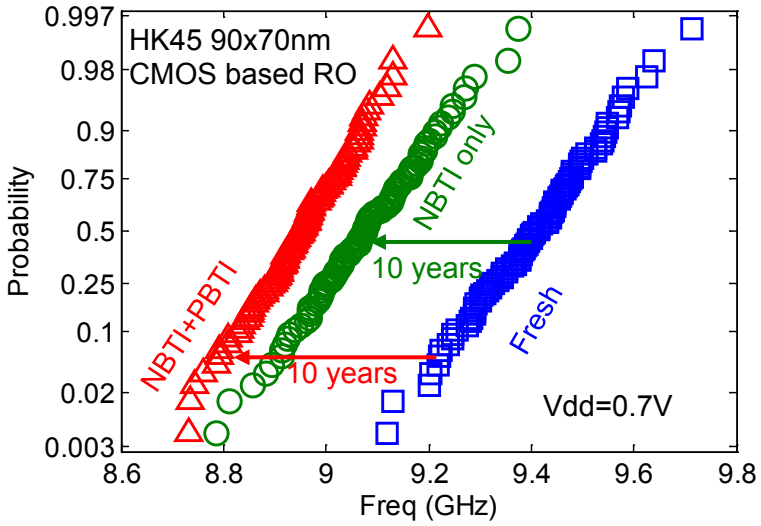

Fig. 6.28 Probablity plot of the RO frequency before and after 10 years. The simulation with taking only NBTI into consideration is also compared with the simulation with both NBTI and PBTI. The difference reveals that PBTI should not be neglected.

Increasing the circuit operating voltage V_{dd} to gain higher performance is widely used in circuit design. As shown in Fig. 6.29a, higher V_{dd} can enhance RO frequency effectively and meanwhile reduce Time-zero Variability impact, leading to the improvement of the circuit yield. However, higher V_{dd} will introduce larger degradation thus reduce circuit lifetime. Circuit designers need to make a tradeoff depending on the specified circuit requirements. Lifetime prediction under different V_{dd} is indispensable.

The proposed A-G framework can fulfill this demand on nano-scale devices based circuits. Fig. 6.29 shows the RO frequency degradation distribution after 10 years under different V_{dd} can be predicted. With this $\Delta f/f_0 \sim V_{dd}$ relationship, lifetime under any failure criteria can be extracted. Fig. 6.30 shows the max V_{dd} allowance predicted by the A-G framework with different failure criteria (from 6% to 10%).

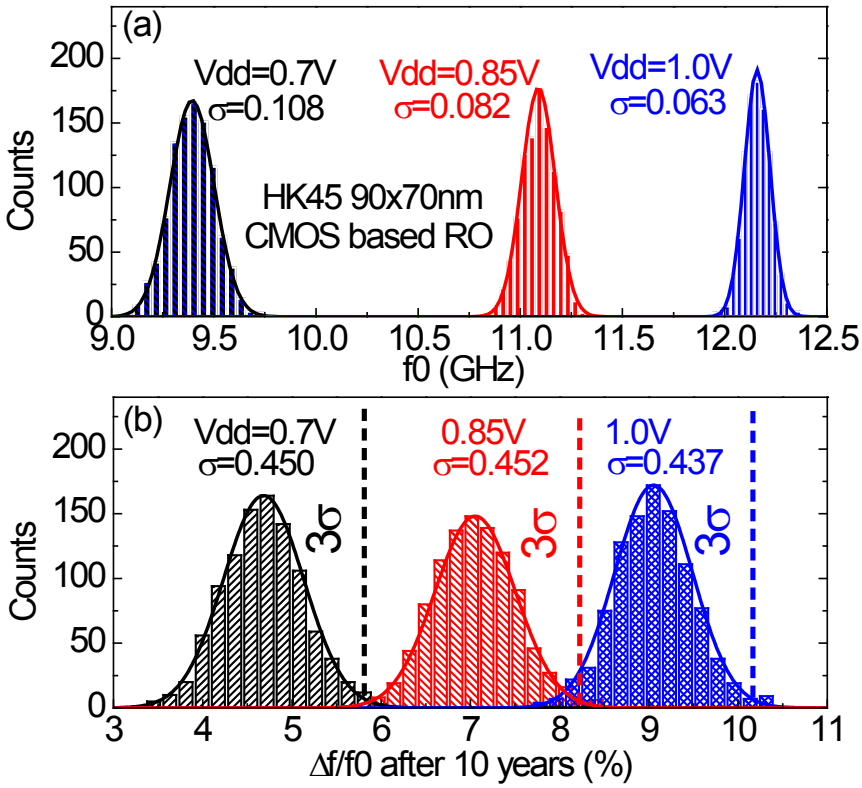

Fig. 6.29 Simulation results of 1000 ROs based on HK45 90x70nm CMOS. (a) Distribution of the fresh RO frequency (f_0) under different V_{dd}. By increasing V_{dd}, RO performance increases (higher f_0) and the impact of Time-zero Variability becomes smaller (decreasing σ). (b) The RO frequency degradation ($\Delta f/f_0$) after 10 years is increasing with higher V_{dd}.

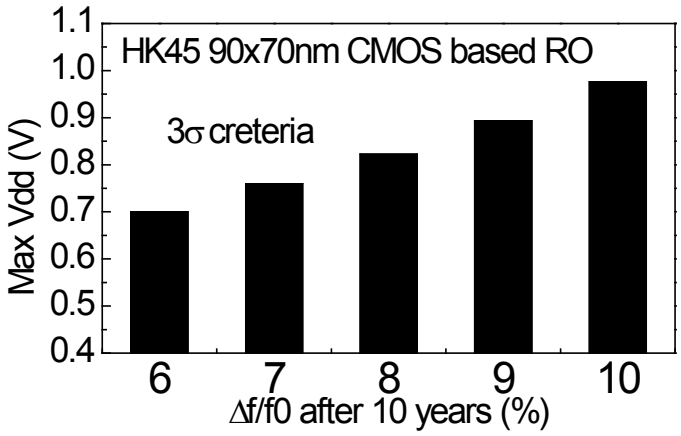

Fig. 6.30 Max V_{dd} predicted by the A-G framework based RO simulation with different failure criteria ($\Delta f/f_0$ from 6% to 10%).

6.8 Summary

A comprehensive As-grown-Generation (A-G) framework for assessing Time Dependent Variation under AC BTI is proposed in this chapter. Compared with the traditional characterization techniques like Random Telegraph Noise (RTN) and Time Dependent Defect Spectroscopy (TDDS), which need to select devices, WDF^{+} measurements on all the devices is used in the A-G framework to capture the worst case of BTI degradation [120]. WDF can be excluded by measuring the averaged degradation, the degradation for scaled devices can be considered as the convolution of the averaged degradation and WDF^{+}.

Averaged degradation on multiple devices can be modelled by the A-G model as detailed in chapter 4. The Fast-Voltage Step Stress technique is used to reduce the test time and significantly improve the efficiency for model parameter extraction. By implementing the model into a commercial circuit simulator, the Time Dependent Variability analysis on circuit level can be made and the model should be helpful for the variability-aware circuit design in the nano-era.

7 Conclusions and future work

7.1 Conclusions

This project is targeted at the modelling of reliability and variability on modern nano-scale devices. Currently there is no practical model for nano-scale devices' lifetime prediction. Based on the understanding of different types of defects, a comprehensive A-G framework is proposed and successfully achieved a test-proven lifetime prediction. Moreover, Fast-Voltage Step Stress acceleration technique is developed to reduce the testing time by 90% for the A-G framework parameter extraction, making the model practical for potential use in industry.

7.1.1 Conclusions on reliable time exponent extraction of long term BTI

As BTI degradation follows a power law, a small change (0.17→0.2) in the time exponent will result in a big error (16 times underestimation) for lifetime prediction. The time exponent reported by different groups is spread over a wide range, and the value changes under different measurement conditions, questioning the accuracy of the predicted lifetime.

The reason lies on the non-saturation of BTI discharging, which makes it very difficult to separate the "Permanent" component "P" and "Recoverable" component "R".

Instead of using ambiguous "P" and "R" to separate defects, the A-G framework classifies the defects into As-grown Traps, Cyclic Positive Charges (CPC) and Anti-Neutralization Positive Charges (ANPC), according to their behaviour during charging and discharging. Both CPC and ANPC are Generated Defects, their sum, Generated Defects, follow a power law and the time exponent is independent of measurement conditions. That is because, although ANPC decreases as discharging time increases, the extra dischargeable defects can be easily recharged under use-bias and give rise to a higher CPC. Based on this observation, "Stress-Discharge-Recharge" pattern is proposed to capture the entire Generated Defects; NBTI time exponents extracted from the Stress-Discharge-Recharge method on four different processes turns out to be ~0.2.

The Stress-Discharge-Recharge technique is also applicable for PBTI, it is found PBTI has a much bigger time exponent (~0.32) compared to NBTI, which means at device lifetime, PBTI degradation is comparable or even larger than NBTI, although it's much smaller than NBTI within a short time.

7.1.2 Conclusions on the Fast-Voltage Step Stress acceleration technique

Accelerated characterization methods are normally preferred or even indispensable for nano-scale devices as they need multiple devices average to get meaningful results. The Voltage Ramp Stress technique is one of the accelerated methods widely used by industry, however, the predicting capability is unsatisfactory due to this method using the total BTI

degradation and power law for parameter extraction. BTI degradation does not follow a power law, especially on immature processes which contain a lot of As-grown Traps.

Voltage Step Stress improves the predicting capability by measuring "P" with slow DC measurements, due to the fact that "P" follows power law well, no matter whether on mature or immature processes. The testing time is around 2 hours for BTI power law parameter extraction and only one device is needed.

Based on the conclusion in section 7.1.1, the Voltage Step Stress technique can be further accelerated and the Fast-Voltage Step Stress technique is proposed. The further acceleration is realized by the known time exponent. Meanwhile the Stress-Discharge-Recharge method is applied to measure Generated Defects instead of slow DC measurement, which delivers the accuracy of BTI time exponent n.

7.1.3 Conclusions on Pre_Existing defects

Pre_Existing defects kinetics is investigated on heavily stressed devices on which further generation is supressed. It is found that Pre_Ex consists of 2 components: As-grown Traps, whose energy will not change during charging and discharging, and Energy Alternating Defects, whose energy will be pulled down after charging, and back to the original level after discharging.

Based on the different behaviour of As-grown Traps and Energy Alternating Defects, multi-Discharging-based Multiple Pulses (m-DMP) technique is developed to separate these two components, results show As-grown Traps will saturate within one second while Energy Alternating Defects follow a power law, but the time exponent is different from Generated Defects.

7.1.4 Conclusions on the test-proved A-G framework lifetime prediction on nano-scale devices

After Generated Defects and Pre_Existing defects are modelled, BTI under an arbitrary voltage, frequency, duty factor and stress time can be calculated according to the A-G framework. Good agreements between measurement data under low gate voltage on big devices and prediction, deliver the good predicting capability of this framework.

Nano-scaled devices have two differences from big devices and make the lifetime prediction difficult: Within Device Fluctuation and Device-to-Device Variability. To predict the Time Dependent Variability of nano-scale devices, WDF is firstly excluded out of BTI by taking the averaged IV measurement. The results, averaged degradation, are collected across multiple devices. Device-to-Device Variability of the averaged degradation is then calculated as the standard deviation against the mean value on multiple devices. The mean value of the averaged degradation turns out to behave like a big device thus can be modelled by the A-G model. WDF, which is independent of stress, frequency and duty factor, is then directly measured on stressed nano-scale devices and modelled.

Together with defect centric theory, the A-G framework is able to simulate the behaviour of each nano-scaled device under any operating condition. Test data under a constant operating voltage on 30 devices is predicted well by the framework-based simulation, validating the predicting capability of the A-G framework. Lifetime is predicted under different criteria (2σ, 3σ etc.), a 31-stage ring oscillator constructed with HK45 90nmx70nm devices is simulated and 10-year V_{dd} is predicted, which is one of the most important parameters for the circuit designers.

7.2 Future work

The A-G model predicts BTI lifetime based on the understanding of different types of defects, the underlying idea is defect separation according to their experimental behaviour, which is also applicable to other degradation mechanisms like Hot Carrier Aging (HCA). By employing the A-G model to other circumstances, a lot of interesting topics are worthwhile to investigate, including but not limited to, the following:

7.2.1 BTI and Hot Carrier Aging (HCA) coupling

Hot Carrier Aging (HCA) is another very important degradation mechanism in modern circuits. A lot of devices are under HCA stress condition instead of BTI while circuits operate, for example SRAM [159].

Fig. 7.1 shows the voltages applied on each terminal while an MOS is under BTI and HCA stress, BTI is mainly caused by the filling and generation of defects uniformly across the gate dielectric while HCA is caused by the hot carrier injection on the drain side. Clearly BTI degradation also exists under the HCA stress condition, and it's a non-uniform BTI. The coupling of these two degradation mechanisms is worthwhile to investigate. Quite a few promising results have already been carried out by M. Duan et. al [160-162] on PBTI and HCA coupling on nMOSFETs.

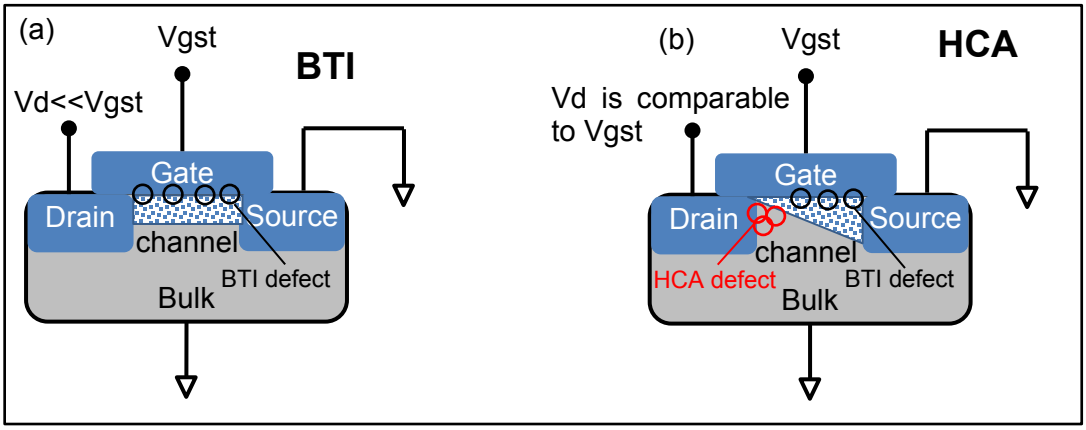

Fig. 7.1 Illustration of the MOSFET voltages under (a) BTI stress and (b) Hot Carrier Aging (HCA) stress.

7.2.2 Within Device Fluctuation (WDF) simulation

Fig. 7.2 shows WDF^{+} is the dominating component that limits nano-scale device lifetime, and the predicted lifetime is much shorter than people's expectation. The reason lies in the definition of WDF^{+} is too rigid, as long as one point hit a higher WDF^{+}, WDF^{+} will then remain at this value and never be back, and the device will fail once WDF^{+} exceeds the lifetime criteria. This is not true for real operating circuits, they can still work with a tolerable jitter.

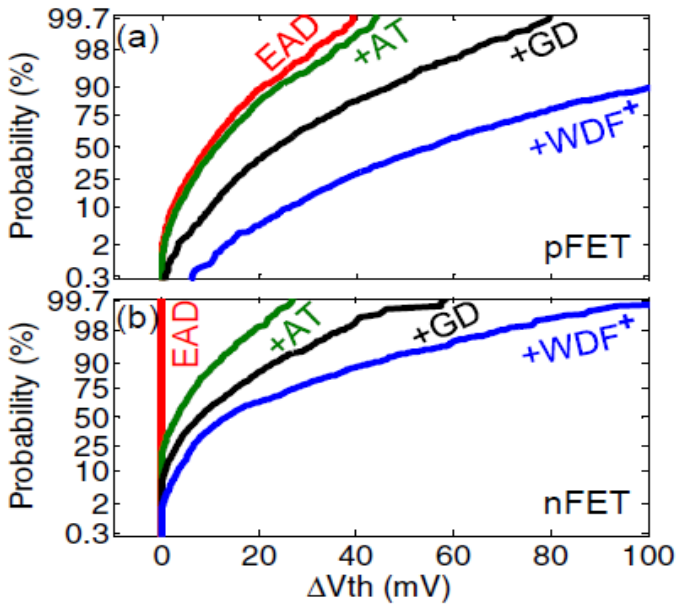

Fig. 7.2 A replot of Fig. 6.26. 10-year degradation under AC NBTI (a) and PBTI (b) with contribution from different types of defects. $|V_g-V_{th}|$=0.5V, Freq=10kHz, Duty factor=0.5. 1000 Monto-Carlo simulations were performed.

One solution is to simulate WDF^+ measurement results with a convolution of RTN signals, which has mature circuit simulation tools already. The A-G framework suggests WDF is mainly caused by As-grown Traps, whose energy profile and charging kinetics can be extracted as described in chapter 4. Based on As-grown Traps' profile and defect centric theory, Monte-Carlo simulation can be used to construct multiple devices with As-grown Traps and mimic the $|\Delta V_{th}|$ kinetics by adding every single trap induced RTN signals.

$$\tau_c = 1/(n_s \cdot v \cdot \sigma_i \cdot e^{\Delta E_B/kT}) \quad (7.1)$$

$$\tau_e = \tau_c \cdot e^{(E_f - E_t)/kT} \quad (7.2)$$

Where σ_i is the average capture cross section, n_s is the carrier density in the inversion layer, v is the average velocity of the carriers, ΔE_B is the thermal activation barrier to capture a carrier, E_f is the fermi level and E_t is the trap energy level [84].

Single trap induced RTN's timing information is generated by Equation (7.1) & Equation (7.2), σ_i and ΔE_B are the only parameters to be determined, they can be extracted by fitting the RTN with As-grown Traps' charging kinetics. Some promising preliminary results have already been carried out as shown in Fig. 7.3.

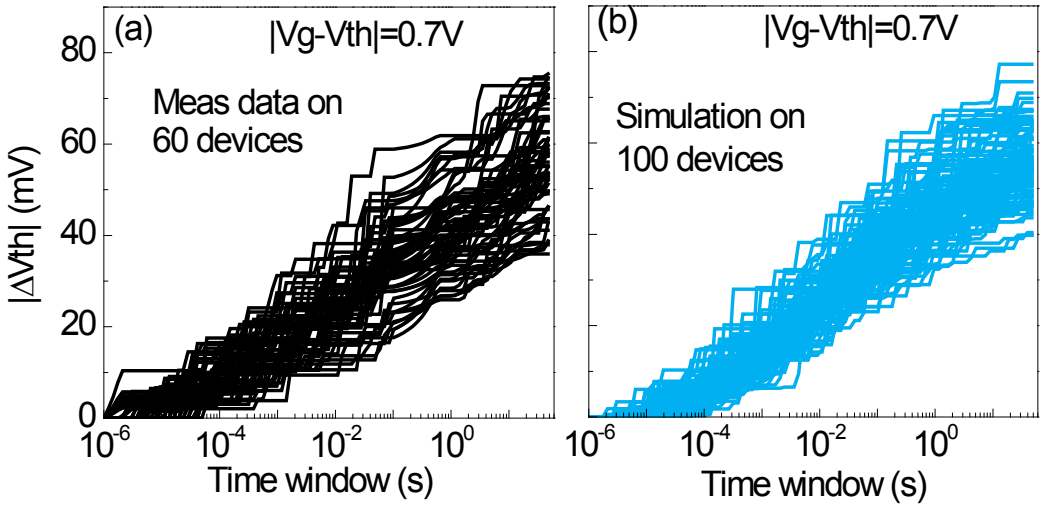

Fig. 7.3 Measured WDF data on 60 devices (a) can be well reproduced by 100 Monte-Carlo WDF simulation (b).

7.2.3 Circuit simulation

Once the work in section 7.2.2 and section 7.2.3 can be done, the A-G framework can be employed to simulate the degradation of various types of circuits like Ring Oscillator (RO), Static Random-Access Memory (SRAM) etc, under real operating conditions instead of pure BTI.

8 References

[1] "https://newsroom.intel.com/news-releases/intel-identifies-chipset-design-error implementing-solution"

[2] "https://www.cisco.com/c/en/us/about/supplier-sustainability/memory.html."

[3] B. E. Deal, "Standardized terminology for oxide charges associated with thermally oxidized silicon," *IEEE Trans. Electron Devices,* vol. 27, no. 3, pp. 606-608, 1980.

[4] B. E. Deal, M. Sklar, A. Grove, and E. Snow, "Characteristics of the Surface-State Charge (Qss) of Thermally Oxidized Silicon," *J. Electrochem. Soc.,* vol. 114, no. 3, pp. 266-274, 1967.

[5] E. Snow, A. Grove, B. Deal, and C. Sah, "Ion transport phenomena in insulating films," *J. Appl. Phys.,* vol. 36, no. 5, pp. 1664-1673, 1965.

[6] M. Kuhn and D. Silversmith, "Ionic contamination and transport of mobile ions in MOS structures," *J. Electrochem. Soc.,* vol. 118, no. 6, pp. 966-970, 1971.

[7] P. Balk, "Effects of hydrogen annealing on silicon surfaces," in *Electrochemical Society Spring Meeting*, 1965, pp. 237-240.

[8] A. Grove, B. Deal, E. Snow, and C. Sah, "Investigation of thermally oxidised silicon surfaces using metal-oxide-semiconductor structures," *Solid-State Electron.,* vol. 8, no. 2, pp. 145-163, 1965.

[9] Y. Miura and Y. Matukura, "Investigation of silicon-silicon dioxide interface using MOS structure," *Jpn. J. Appl. Phys.,* vol. 5, no. 2, p. 180, 1966.

[10] A. Kerber and E. Cartier, "Bias temperature instability characterization methods," in *Bias Temperature Instability for Devices and Circuits*, ed: Springer, 2014, pp. 3-31.

[11] 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.

[12] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, and R. Chau, "A 45nm logic technology with high-k+ metal gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging," in *IEDM Tech. Dig.*, 2007, pp. 247-250.

[13] M. Chudzik, B. Doris, R. Mo, J. Sleight, E. Cartier, C. Dewan, D. Park, H. Bu, W. Natzle, and W. Yan, "High-performance high-κ/metal gates for 45nm CMOS and beyond with gate-first processing," in *VLSI Symp. Tech. Dig.*, 2007, pp. 194-195.

[14] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, and S. Krishnan, "NBTI impact on transistor and circuit: models, mechanisms and scaling effects [MOSFETs]," in *IEDM Tech. Dig.*, 2003, pp. 14.5. 1-14.5. 4.

[15] N. K. Jha, P. S. Reddy, D. K. Sharma, and V. R. Rao, "NBTI degradation and its impact for analog circuit reliability," *IEEE Trans. Electron Devices,* vol. 52, no. 12, pp. 2609-2615, 2005.

[16] J. S. S. T. Association, "Failure mechanisms and models for semiconductor devices," *JEDEC Publication JEP122-B,* 2003.

[17] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification," in *Proc. IRPS*, 2005, pp. 381-387.

[18] H. Reisinger, T. Grasser, K. Ermisch, H. Nielen, W. Gustin, and C. Schlünder, "Understanding and modeling AC BTI," in *Proc. IRPS*, 2011, pp. 6A. 1.1-6A. 1.8.

[19] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A two-stage model for negative bias temperature instability," in *Proc. IRPS*, 2009, pp. 33-44.

[20] S. Mahapatra and M. Alam, "A predictive reliability model for PMOS bias temperature degradation," in *IEDM Tech. Dig.*, 2002, pp. 505-508.

[21] S. Mahapatra, N. Goel, S. Desai, S. Gupta, B. Jose, S. Mukhopadhyay, K. Joshi, A. Jain, A. Islam, and M. Alam, "A comparative study of different physics-based NBTI models," *IEEE Trans. Electron Devices,* vol. 60, no. 3, pp. 901-916, 2013.

[22] M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *IEDM Tech. Dig.*, 2003, pp. 14.4. 1-14.4. 4.

[23] C. Blat, E. Nicollian, and E. Poindexter, "Mechanism of negative-bias-temperature instability," *J. Appl. Phys.,* vol. 69, no. 3, pp. 1712-1720, 1991.

[24] G. Gadiyak, "Numerical simulation of hydrogen redistribution in thin SiO2 films under electron injection in high fields," *Applied surface science,* vol. 113, pp. 627-630, 1997.

[25] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectron. Rel.,* vol. 45, no. 1, pp. 71-81, 2005.

[26] S. Mahapatra, P. B. Kumar, and M. Alam, "A new observation of enhanced bias temperature instability in thin gate oxide p-MOSFETs," in *IEDM Tech. Dig.*, 2003, pp. 14.2. 1-14.2. 4.

[27] M. Houssa, M. Aoulaiche, S. De Gendt, G. Groeseneken, M. Heyns, and A. Stesmans, "Reaction-dispersive proton transport model for negative bias temperature instabilities," *Appl. Phys. Lett.,* vol. 86, no. 9, p. 093506, 2005.

[28] V. Huard, M. Denais, F. Perrier, N. Revil, C. Parthasarathy, A. Bravaix, and E. Vincent, "A thorough investigation of MOSFETs NBTI degradation," *Microelectron. Rel.,* vol. 45, no. 1, pp. 83-98, 2005.

[29] B. Kaczer, V. Arkhipov, M. Jurczak, and G. Groeseneken, "Negative bias temperature instability (NBTI) in SiO2 and SiON gate dielectrics understood through disorder-controlled kinetics," *Microelectron. Eng.,* vol. 80, pp. 122-125, 2005.

[30] H. Kufluoglu and M. A. Alam, "A Generalized Reaction–Diffusion Model With Explicit H-H_2 Dynamics for Negative-Bias Temperature-Instability (NBTI) Degradation," *IEEE Trans. Electron Devices,* vol. 54, no. 5, pp. 1101-1107, 2007.

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

[32] A. E. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, and M. A. Alam, "Recent issues in negative-bias temperature instability: Initial degradation, field dependence of interface trap generation, hole trapping effects, and relaxation," *IEEE Trans. Electron Devices,* vol. 54, no. 9, pp. 2143-2154, 2007.

[33] Z. Ji, J. Zhang, L. Lin, M. Duan, W. Zhang, X. Zhang, R. Gao, B. Kaczer, J. Franco, and T. Schram, "A test-proven As-grown-Generation (AG) model for predicting NBTI under use-bias," in *VLSI Symp. Tech. Dig.*, 2015, pp. T36-T37.

[34] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.,* vol. 94, no. 1, pp. 1-18, 2003.

[35] S. Chakravarthi, A. Krishnan, V. Reddy, C. Machala, and S. Krishnan, "A comprehensive framework for predictive modeling of negative bias temperature instability," in *Proc. IRPS*, 2004, pp. 273-282.

[36] C. Shen, M.-F. Li, C. Foo, T. Yang, D. Huang, A. Yap, G. Samudra, and Y. Yeo, "Characterization and physical origin of fast Vth transient in NBTI of pMOSFETs with SiON dielectric," in *IEDM Tech. Dig.*, 2006, pp. 1-4.

[37] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanisms to modelling," *Microelectron. Rel.,* vol. 46, no. 1, pp. 1-23, 2006.

[38] N. Goel, S. Mukhopadhyay, N. Nanaware, S. De, R. Pandey, K. Murali, and S. Mahapatra, "A comprehensive DC/AC model for ultra-fast NBTI in deep EOT scaled HKMG p-MOSFETs," in *Proc. IRPS*, 2014, pp. 6A. 4.1-6A. 4.12.

[39] D. Ang, S. Wang, and C. Ling, "Evidence of two distinct degradation mechanisms from temperature dependence of negative bias stressing of the ultrathin gate p-MOSFET," *IEEE Electron Device Lett.,* vol. 26, no. 12, pp. 906-908, 2005.

[40] A. Haggag, G. Anderson, S. Parihar, D. Burnett, G. Abeln, J. Higman, and M. Moosa, "Understanding SRAM high-temperature-operating-life NBTI: Statistics and permanent vs recoverable damage," in *Proc. IRPS*, 2007, pp. 452-456.

[41] T. Grasser, B. Kaczer, P. Hehenberger, W. Gos, R. O'connor, H. Reisinger, W. Gustin, and C. Schlunder, "Simultaneous extraction of recoverable and permanent components contributing to bias-temperature instability," in *IEDM Tech. Dig.*, 2007, pp. 801-804.

[42] T. Grasser, W. Gos, V. Sverdlov, and B. Kaczer, "The universality of NBTI relaxation and its implications for modeling and characterization," in *Proc. IRPS*, 2007, pp. 268-280.

[43] W. Kauzmann, "Dielectric relaxation as a chemical rate process," *Reviews of Modern Physics,* vol. 14, no. 1, p. 12, 1942.

[44] J. R. Jameson, W. Harrison, P. Griffin, J. Plummer, and Y. Nishi, "A semiclassical model of dielectric relaxation in glasses," *J. Appl. Phys.,* vol. 100, no. 12, p. 124104, 2006.

[45] W. Phillips, "Tunneling states in amorphous solids," *Journal of Low Temperature Physics,* vol. 7, no. 3, pp. 351-360, 1972.

[46] T. Grasser, B. Kaczer, and W. Goes, "An energy-level perspective of bias temperature instability," in *Proc. IRPS*, 2008, pp. 28-38.

[47] G. Declerck, "A look into the future of nanoelectronics," in *VLSI Symp. Tech. Dig.*, 2005, pp. 6-10.

[48] A. Asenov, "Simulation of statistical variability in nano MOSFETs," in *VLSI Symp. Tech. Dig.*, 2007, pp. 86-87.

[49] 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 *Proc. IRPS*, 2010, pp. 7-15.

[50] B. Kaczer, S. Mahato, V. V. de Almeida Camargo, M. Toledano-Luque, P. J. Roussel, T. Grasser, F. Catthoor, P. Dobrovolny, P. Zuber, and G. Wirth, "Atomistic approach to variability of bias-temperature instability in circuit simulations," in *Proc. IRPS*, 2011, pp. XT. 3.1-XT. 3.5.

[51] H.-S. Wong and Y. Taur, "Three-dimensional" atomistic" simulation of discrete random dopant distribution effects in sub-0.1/spl mu/m MOSFET's," in *IEDM Tech. Dig.*, 1993, pp. 705-708.

[52] J. Franco, B. Kaczer, M. Toledano-Luque, P. J. Roussel, J. Mitard, L. Å. Ragnarsson, L. Witters, T. Chiarella, M. Togo, N. Horiguchi, G. Groeseneken, M. F. Bukhori, T. Grasser, and A. Asenov, "Impact of single charged gate oxide defects on the performance and scaling of nanoscaled FETs," in *Proc. IRPS*, 2012, pp. 5A.4.1-5A.4.6.

[53] M. Agostinelli, S. Pae, W. Yang, C. Prasad, D. Kencke, S. Ramey, E. Snyder, S. Kashyap, and M. Jones, "Random charge effects for PMOS NBTI in ultra-small gate area devices," in *Proc. IRPS*, 2005, pp. 529-532.

[54] J. Franco, B. Kaczer, N. Waldron, P. J. Roussel, A. Alian, M. A. Pourghaderi, Z. Ji, T. Grasser, T. Kauerauf, and S. Sioncke, "RTN and PBTI-induced time-dependent variability of replacement metal-gate high-k InGaAs FinFETs," in *IEDM Tech. Dig.*, 2014, pp. 20.2. 1-20.2. 4.

[55] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs," in *Proc. IRPS*, 2010, pp. 26-32.

[56] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3-D “atomistic” simulation study," *IEEE Trans. Electron Devices*, vol. 45, no. 12, pp. 2505-2513, 1998.

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

[58] S. Rauch, "The statistics of NBTI-induced VT and β mismatch shifts in pMOSFETs," *IEEE Trans. Device Mater. Rel.*, vol. 2, no. 4, pp. 89–93, Dec. 2002.

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

[60] J. Zhang, Z. Ji, M. Chang, B. Kaczer, and G. Groeseneken, "Real Vth instability of pMOSFETs under practical operation conditions," in *IEDM Tech. Dig.*, 2007, pp. 817-820.

[61] https://www.keysight.com/en/pc-1862522/source-measure-units?cc=US&lc=eng

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

[63] A. Ortiz-Conde, F. G. Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectron. Rel.,* vol. 42, no. 4, pp. 583-596, 2002.

[64] P. Habas, S. Mileusnic, and T. Zivanov, "Characterization of power VDMOSFETs by split CV measurements," in *Proc. Microelectronics*, 2000, pp. 339-342.

[65] S. Mileusnic, P. Haba, and M. Zivanov, "New characterization technique for oxide degradation in power VDMOSFET based on split CV measurements," in *Proc. Semiconductor*, 1999, pp. 165-168.

[66] S. Mileusnic, M. Zivanov, and P. Habas, "MOS transistors characterization by split CV method," in *Proc. Semiconductor*, 2001, pp. 503-506.

[67] J. Koomen, "Investigation of the MOST channel conductance in weak inversion," *Solid-State Electron.,* vol. 16, no. 7, pp. 801-810, 1973.

[68] J. Hauser and K. Ahmed, "Characterization of ultra-thin oxides using electrical CV and IV measurements," in *AIP Conference Proceedings*, 1998, pp. 235-239.

[69] J. S. Brugler and P. G. Jespers, "Charge pumping in MOS devices," *IEEE Trans. Electron Devices,* vol. 16, no. 3, pp. 297-302, 1969.

[70] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. De Keersmaecker, "A reliable approach to charge-pumping measurements in MOS transistors," *IEEE Transactions on Electron Devices,* vol. 31, no. 1, pp. 42-53, 1984.

[71] A. B. Elliot, "The use of charge pumping currents to measure surface state densities in MOS transistors," *Solid-State Electron.,* vol. 19, no. 3, pp. 241-247, 1976.

[72] C. D. Young, Y. Zhao, M. Pendley, B. H. Lee, K. Matthews, J. H. Sim, R. Choi, G. A. Brown, R. W. Murto, and G. Bersuker, "Ultra-short pulse current–voltage characterization of the intrinsic characteristics of high-κ devices," *Jpn. J. Appl. Phys.,* vol. 44, no. 4S, p. 2437, 2005.

[73] B. Kaczer, T. Grasser, P. J. Roussel, J. Martin-Martinez, R. O'Connor, B. O'sullivan, and G. Groeseneken, "Ubiquitous relaxation in BTI stressing—New evaluation and insights," in *Proc. IRPS*, 2008, pp. 20-27.

[74] H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin, and C. Schlunder, "Analysis of NBTI degradation-and recovery-behavior based on ultra fast VT-measurements," in *Proc. IRPS*, 2006, pp. 448-453.

[75] S. Rangan, N. Mielke, and E. Yeh, "Universal recovery behavior of negative bias temperature instability [PMOSFETs]," in *IEDM Tech. Dig.*, 2003, pp. 14.3. 1-14.3. 4.

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

[77] X. Zheng, W. Zhang, B. Govoreanu, J. Zhang, and J. Van Houdt, "A discharge-based multi-pulse technique (DMP) for probing electron trap energy distribution in high-k materials for Flash memory application," in *IEDM Tech. Dig.*, 2009, pp. 1-4.

[78] Z. Ji, S. F. W. M. Hatta, J. F. Zhang, J. G. Ma, W. Zhang, N. Soin, B. Kaczer, S. D. Gendt, and G. Groeseneken, "Negative bias temperature instability lifetime prediction: Problems and solutions," in *IEDM Tech. Dig.*, 2013, pp. 15.6.1-15.6.4.

[79] S. Hatta, Z. Ji, J. F. Zhang, M. Duan, W. D. Zhang, N. Soin, B. Kaczer, S. De Gendt, and G. Groeseneken, "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, 2013.

[80] T. Aichinger, M. Nelhiebel, S. Einspieler, and T. Grasser, "Observing two stage recovery of gate oxide damage created under negative bias temperature stress," *J. Appl. Phys.,* vol. 107, no. 2, p. 024508, 2010.

[81] X. Ji, Y. Liao, F. Yan, Y. Shi, G. Zhang, and Q. Guo, "The energy distribution of NBTI-induced hole traps in the Si band gap in PNO pMOSFETs," in *Proc. IRPS*, 2012, pp. XT. 12.1-XT. 12.5.

[82] M. Chang and J. Zhang, "On positive charge formed under negative bias temperature stress," *J. Appl. Phys.,* vol. 101, no. 2, p. 024516, 2007.

[83] C. Zhao and J. Zhang, "Effects of hydrogen on positive charges in gate oxides," *J. Appl. Phys.,* vol. 97, no. 7, p. 073703, 2005.

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

[85] K. Hung, P. Ko, C. Hu, and Y. C. Cheng, "Random telegraph noise of deep-submicrometer MOSFETs," *IEEE Electron Device Lett.,* vol. 11, no. 2, pp. 90-92, 1990.

[86] H. Miki, M. Yamaoka, N. Tega, Z. Ren, M. Kobayashi, C. D'Emic, Y. Zhu, D. Frank, M. Guillorn, and D.-G. Park, "Understanding short-term BTI behavior through comprehensive observation of gate-voltage dependence of RTN in highly scaled high-κ/metal-gate pFETs," in *VLSI Symp. Tech. Dig.*, 2011, pp. 148-149.

[87] T. Grasser, H. Reisinger, P. J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," in *Proc. IRPS*, 2010, pp. 16-25.

[88] A. Kerber, S. A. Krishnan, and E. A. Cartier, "Voltage Ramp Stress for Bias Temperature Instability Testing of Metal-Gate/High-Stacks," *IEEE Electron Device Lett.,* vol. 30, no. 12, pp. 1347-1349, 2009.

[89] J. Zhang, "Defects and instabilities in Hf-dielectric/SiON stacks," *Microelectron. Eng.,* vol. 86, no. 7, pp. 1883-1887, 2009.

[90] B. Djezzar, H. Tahi, A. Benabdelmoumene, A. Chenouf, M. Goudjil, and Y. Kribes, "On the permanent component profiling of the negative bias temperature instability in p-MOSFET devices," *Solid-State Electron.,* vol. 106, pp. 54-62, 2015.

[91] S. Pae, M. Agostinelli, M. Brazier, R. Chau, G. Dewey, T. Ghani, M. Hattendorf, J. Hicks, J. Kavalieros, and K. Kuhn, "BTI reliability of 45 nm high-K+ metal-gate process technology," in *Proc. IRPS*, 2008, pp. 352-357.

[92] J. Campbell, P. Lenahan, A. Krishnan, and S. Krishnan, "NBTI: An atomic-scale defect perspective," in *Proc. IRPS*, 2006, pp. 442-447.

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

[94] V. Huard, F. Monsieur, G. Ribes, and S. Bruyere, "Evidence for hydrogen-related defects during NBTl stress in p-MOSFETs," in *Proc. IRPS*, 2003, pp. 178-182.

[95] D. Heh, P. D. Kirsch, C. D. Young, and G. Bersuker, "A new dielectric degradation phenomenon in nMOS high-k devices under positive bias stress," in *Proc. IRPS*, 2008, pp. 347-351.

[96] V. Huard, "Two independent components modeling for negative bias temperature instability," in *Proc. IRPS*, 2010, pp. 33-42.

[97] J. Hicks, D. Bergstrom, M. Hattendorf, J. Jopling, J. Maiz, S. Pae, C. Prasad, and J. Wiedemer, "45nm Transistor Reliability," *Intel Technology Journal,* vol. 12, no. 2, pp. 131-144, 2008.

[98] S. Mukhopadhyay, J. Franco, A. Chasin, P. J. Roussel, B. Kaczer, G. Groeseneken, N. Horiguchi, D. Linten, and A. Thean, "Fundamental study of the apparent voltage-dependence of NBTI kinetics by constant electric field stress in Si and SiGe devices," in *Proc. IRPS*, 2016, pp. 5A-3-1-5A-3-7.

[99] V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammase, D. Barge, and C. Ouvrard, "New characterization and modeling approach for NBTI degradation from transistor to product level," in *IEDM Tech. Dig.*, 2007, pp. 797-800.

[100] D. Breed, "A new model for the negative voltage instability in MOS devices," *Appl. Phys. Lett.,* vol. 26, no. 3, pp. 116-118, 1975.

[101] A. Krishnan, C. Chancellor, S. Chakravarthi, P. Nicollian, V. Reddy, A. Varghese, R. Khamankar, and S. Krishnan, "Material dependence of hydrogen diffusion: Implications for NBTI degradation," in *IEDM Tech. Dig.*, 2005, pp. 4 pp.-691.

[102] S. Mahapatra, V. Huard, A. Kerber, V. Reddy, S. Kalpat, and A. Haggag, "Universality of NBTI-From devices to circuits and products," in *Proc. IRPS*, 2014, pp. 3B. 1.1-3B. 1.8.

[103] M. Cho, J. D. Lee, M. Aoulaiche, B. Kaczer, P. Roussel, T. Kauerauf, R. Degraeve, J. Franco, L, x00C, Ragnarsson, and G. Groeseneken, "Insight Into N/PBTI Mechanisms in Sub-1-nm-EOT Devices," *IEEE Trans. on Electron Devices,* vol. 59, no. 8, pp. 2042-2048, 2012.

[104] G. Pobegen and T. Grasser, "On the distribution of NBTI time constants on a long, temperature-accelerated time scale," *IEEE Trans. Electron Devices,* vol. 60, no. 7, pp. 2148-2155, 2013.

[105] T. Grasser, M. Waltl, Y. Wimmer, W. Goes, R. Kosik, G. Rzepa, H. Reisinger, G. Pobegen, A. El-Sayed, A. Shluger, and B. Kaczer, "Gate-Sided Hydrogen Release as the Origin of "Permanent" NBTI Degradation: From Single Defects to Lifetimes," in *IEDM Tech. Dig.*, 2015, pp. 20.1.1 - 20.1.4.

[106] M. Duan, J. Zhang, Z. Ji, W. Zhang, B. Kaczer, S. De Gendt, and G. Groeseneken, "Defect loss: A new concept for reliability of MOSFETs," *IEEE Electron Device Lett.,* vol. 33, no. 4, pp. 480-482, 2012.

[107] D. Ang, S. Lai, G. Du, Z. Teo, T. Ho, and Y. Hu, "Effect of hole-trap distribution on the power-law time exponent of NBTI," *IEEE Electron Device Lett.,* vol. 30, no. 7, pp. 751-753, 2009.

[108] L. Jin, M. Xu, and C. Tan, "An Investigation on the Permanent Component of NBTI Degradation in a 90nm CMOS Technology," in *Proc. ICSICT*, 2006, pp. 1147-1149.

[109] A. Kerber, K. Maitra, A. Majumdar, M. Hargrove, R. J. Carter, and E. A. Cartier, "Characterization of Fast Relaxation During BTI Stress in Conventional and Advanced CMOS Devices With Gate Stacks," *IEEE Trans. Electron Devices,* vol. 55, no. 11, pp. 3175-3183, 2008.

[110] Z. Liu, D. Huang, W. Liu, C. Liao, L. Zhang, Z. Gan, W. Wong, and M.-F. Li, "Comprehensive studies of BTI effects in CMOSFETs with SiON by new measurement techniques," in *Proc. IRPS*, 2008, pp. 733-734.

[111] J.-J. Kim, R. M. Rao, J. Schaub, A. Ghosh, A. Bansal, K. Zhao, B. P. Linder, and J. Stathis, "PBTI/NBTI monitoring ring oscillator circuits with on-chip Vt characterization and high frequency AC stress capability," in *VLSI Symp. Cir. Dig.*, 2011, pp. 224-225.

[112] M. Jin, C. Tian, G. La Rosa, S. Uppal, W. Mcmahon, H. Kothari, Y. Liu, E. Cartier, W. Lai, and A. Dasgupta, "Impact of hydrogen in capping layers on BTI degradation and recovery in high-κ replacement metal gate transistors," in *Proc. IRPS*, 2013, pp. PI. 3.1-PI. 3.5.

[113] H. Kukner, P. Weckx, J. Franco, M. Toledano-Luque, M. Cho, B. Kaczer, P. Raghavan, D. Jang, K. Miyaguchi, and M. G. Bardon, "Scaling of BTI reliability in presence of time-zero variability," in *Proc. IRPS*, 2014, pp. CA. 5.1-CA. 5.7.

[114] C. Prasad, M. Agostinelli, J. Hicks, S. Ramey, C. Auth, K. Mistry, S. Natarajan, P. Packan, I. Post, and S. Bodapati, "Bias temperature instability variation on SiON/Poly, HK/MG and trigate architectures," in *Proc. IRPS*, 2014, pp. 6A.5.1 - 6A.5.7.

[115] D. P. Ioannou, S. Mitt, and G. LaRosa, "Positive bias temperature instability effects in advanced high-k metal gate NMOSFETs," in *Proc. IEEE IRW Final Rep.*, 2008, pp. 55-57.

[116] D. Ang and S. Wang, "Recovery of the NBTI-stressed ultrathin gate p-MOSFET: The role of deep-level hole traps," *IEEE Electron Device Lett.,* vol. 27, no. 11, pp. 914-916, 2006.

[117] T. Grasser and B. Kaczer, "Negative bias temperature instability: Recoverable versus permanent degradation," in *Proc. ESSDERC*, 2007, pp. 127-130.

[118] P. Weckx, B. Kaczer, C. Chen, J. Franco, E. Bury, K. Chanda, J. Watt, P. J. Roussel, F. Catthoor, and G. Groeseneken, "Characterization of time-dependent variability using 32k transistor arrays in an advanced HK/MG technology," in *Proc. IRPS*, 2015, pp. 3B. 1.1-3B. 1.6.

[119] Z. Ji, J. F. Zhang, L. Lin, M. Duan, W. Zhang, X. Zhang, R. Gao, B. Kaczer, J. Franco, T. Schram, N. Horiguchi, S. D. Gendt, and G. Groeseneken, "A test-proven As-grown-Generation (A-G) model for predicting NBTI under use-bias," in *VLSI Symp. Tech. Dig.*, 2015, pp. T36-T37.

[120] M. Duan, J. Zhang, Z. Ji, W. Zhang, B. Kaczer, T. Schram, R. Ritzenthaler, A. Thean, G. Groeseneken, and A. Asenov, "Time-dependent variation: A new defect-based prediction methodology," in *VLSI Symp. Tech. Dig.*, 2014, pp. 1-2.

[121] P. Ren, R. Gao, Z. Ji, H. Arimura, J. F. Zhang, R. Wang, M. Duan, W. Zhang, J. Franco, S. Sioncke, D. Cott, J. Mitard, L. Witters, H. Mertens, B. Kaczer, A. Mocuta, N. Collaert, D. Linten, R. Huang, A. V.-Y. Thean, and G. Groeseneken, "Understanding charge traps for optimizing Si-passivated Ge nMOSFETs," *VLSI Symp. Tech. Dig.,* 2016.

[122] Z. Ji, X. Zhang, J. Franco, R. Gao, M. Duan, J. F. Zhang, W. D. Zhang, B. Kaczer, A. Alian, and D. Linten, "An investigation on border traps in III–V MOSFETs with an In 0.53 Ga 0.47 As channel," *IEEE Trans. Electron Devices,* vol. 62, no. 11, pp. 3633-3639, 2015.

[123] J. Ma, W. Zhang, J. Zhang, B. Benbakhti, Z. Ji, J. Mitard, J. Franco, B. Kaczer, and G. Groeseneken, "NBTI of Ge pMOSFETs: understanding defects and enabling lifetime prediction," in *IEDM Tech. Dig.*, 2014, pp. 34.2. 1-34.2. 4.

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

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

[126] A. Neugroschel, G. Bersuker, and R. Choi, "Applications of DCIV method to NBTI characterization," *Microelectron. Rel.,* vol. 47, no. 9, pp. 1366-1372, 2007.

[127] A. Kerber and E. Cartier, "Application of VRS Methodology for the Statistical Assessment of BTI in MG/HK CMOS Devices," *IEEE Electron Device Lett.,* vol. 34, no. 8, pp. 960-962, 2013.

[128] Z. Ji, J. F. Zhang, W. Zhang, X. Zhang, B. Kaczer, S. D. Gendt, G. Groeseneken, P. Ren, R. Wang, and R. Huang, "A single device based voltage step stress (VSS) technique for fast reliability screening," in *Proc. IRPS*, 2014, pp. GD.2.1-GD.2.4.

[129] R. Gao, A. B. Manut, Z. Ji, J. Ma, M. Duan, J. F. Zhang, J. Franco, S. W. M. Hatta, W. D. Zhang, and B. Kaczer, "Reliable Time Exponents for Long Term Prediction of Negative Bias Temperature Instability by Extrapolation," *IEEE Trans. Electron Devices,* vol. 64, no. 4, pp. 1467-1473, 2017.

[130] S. Mukhopadhyay, K. Joshi, V. Chaudhary, N. Goel, S. De, R. K. Pandey, et al., "Trap Generation in IL and HK layers during BTI / TDDB stress in scaled HKMG N and P MOSFETs," in *Proc. IRPS*, 2014, pp. GD.3.1-GD.3.11.

[131] G. Krause, R. Geilenkeuser, M. Trentzsch, F. Graetsch, and L. Herrmann, "Application of fast wafer-level reliability PBTI tests for screening of High-k/Metal Gate process splits," in *Proc. IEEE IRW Final Rep.*, 2009, pp. 66-69.

[132] J. Yang, M. Masuduzzaman, K. Joshi, S. Mukhopadhyay, J. Kang, S. Mahapatra, and M. A. Alam, "Intrinsic correlation between PBTI and TDDB degradations in nMOS HK/MG dielectrics," in *Proc. IRPS*, 2012, pp. 5D. 4.1-5D. 4.7.

[133] D. P. Ioannou, S. Mittl, and G. L. Rosa, "Positive Bias Temperature Instability Effects in nMOSFETs With HfO_2/TiN Gate Stacks," *IEEE Trans. Device Mater. Rel.,* vol. 9, no. 2, pp. 128-134, 2009.

[134] C. Schlünder, H. Reisinger, S. Aresu, and W. Gustin, "On the PBTI degradation of pMOSFETs and its impact on IC lifetime," in *Proc. IEEE IRW Final Rep.*, 2011, pp. 7-11.

[135] V. P.-H. Hu, M.-L. Fan, C.-Y. Hsieh, P. Su, and C.-T. Chuang, "FinFET SRAM cell optimization considering temporal variability due to NBTI/PBTI, surface orientation and various gate dielectrics," *IEEE Trans. Electron Devices,* vol. 58, no. 3, pp. 805-811, 2011.

[136] T. T.-H. Kim, P.-F. Lu, K. A. Jenkins, and C. H. Kim, "A ring-oscillator-based reliability monitor for isolated measurement of NBTI and PBTI in high-k/metal gate technology," *IEEE Trans. Very Large Scale Integr. Syst.,* vol. 23, no. 7, pp. 1360-1364, 2015.

[137] A. Neugroschel, C.-T. Sah, K. M. Han, M. S. Carroll, T. Nishida, J. T. Kavalieros, and Y. Lu, "Direct-current measurements of oxide and interface traps on oxidized silicon," *IEEE Trans. Electron Devices,* vol. 42, no. 9, pp. 1657-1662, 1995.

[138] K. Zhao, J. Stathis, B. Linder, E. Cartier, and A. Kerber, "PBTI under dynamic stress: From a single defect point of view," in *Proc. IRPS*, 2011, pp. 4A. 3.1-4A. 3.9.

[139] Y. Mitani and A. Toriumi, "Experimental study on origin of V TH variability under NBT stress," in *Proc. IRPS*, 2011, pp. PL. 2.1-PL. 2.6.

[140] R. Pengpeng, R. Wang, Z. Ji, H. Peng, J. Xiaobo, G. Shaofeng, L. Mulong, M. Duan, J. F. Zhang, W. Jianping, L. Jinhua, B. Weihai, W. Jingang, W. Waisum, Y. Shaofeng, W.

Hanming, L. Shiuh-Wuu, X. Nuo, and H. Ru, "New insights into the design for end-of-life variability of NBTI in scaled high-k/metal-gate Technology for the nano-reliability era," in *IEDM Tech. Dig.*, 2014, pp. 34.1.1-34.1.4.

[141] V. Huard, F. Cacho, X. Federspiel, W. Arfaoui, M. Saliva, and D. Angot, "Technology scaling and reliability: Challenges and opportunities," in *IEDM Tech. Dig.*, 2015, pp. 20.5. 1-20.5. 6.

[142] A. Bansal, R. Rao, J.-J. Kim, S. Zafar, J. H. Stathis, and C.-T. Chuang, "Impacts of NBTI and PBTI on SRAM static/dynamic noise margins and cell failure probability," *Microelectron. Rel.,* vol. 49, no. 6, pp. 642-649, 2009.

[143] G. Sasse, "Device degradation models for circuit reliability simulation," in *Proc. IEEE IRW Final Rep.*, 2013, pp. 42-42.

[144] C.-Y. Chen, Q. Ran, H.-J. Cho, A. Kerber, Y. Liu, M.-R. Lin, and R. W. Dutton, "Correlation of I d-and I g-random telegraph noise to positive bias temperature instability in scaled high-κ/metal gate n-type MOSFETs," in *Proc. IRPS*, 2011, pp. 3A. 2.1-3A. 2.6.

[145] A. S. Oates, "Reliability challenges for the continued scaling of IC technologies," in *Proc. CICC*, 2012, pp. 1-4.

[146] Y. Wang, B. Cheng, X. Wang, E. Towie, C. Riddet, A. Brown, S. Amoroso, L. Wang, D. Reid, and X. Liu, "Variability-aware TCAD based design-technology co-optimization platform for 7nm node nanowire and beyond," in *VLSI Symp. Tech. Dig.*, 2016, pp. 1-2.

[147] S. Ramey, Y. Lu, I. Meric, S. Mudanai, S. Novak, C. Prasad, and J. Hicks, "Aging model challenges in deeply scaled tri-gate technologies," in *Proc. IEEE IRW Final Rep.*, 2015, pp. 56-62.

[148] M. Duan, J. F. Zhang, Z. Ji, W. D. Zhang, B. Kaczer, S. De Gendt, and G. Groeseneken, "New insights into defect loss, slowdown, and device lifetime enhancement," *IEEE Trans. Electron Devices,* vol. 60, no. 1, pp. 413-419, 2013.

[149] H. Miki, M. Yamaoka, D. Frank, K. Cheng, D.-G. Park, E. Leobandung, and K. Torii, "Voltage and temperature dependence of random telegraph noise in highly scaled HKMG ETSOI nFETs and its impact on logic delay uncertainty," in *VLSI Symp. Tech. Dig.*, 2012, pp. 137-138.

[150] 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," in *VLSI Symp. Tech. Dig.*, 2014, pp. 1-2.

[151] 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," in *Proc. IRPS*, 2014, pp. XT. 17.1-XT. 17.5.

[152] M. Duan, J. F. Zhang, Z. Ji, W. D. Zhang, B. Kaczer, T. Schram, R. Ritzenthaler, G. Groeseneken, and A. Asenov, "New Analysis Method for Time-Dependent Device-To-Device Variation Accounting for Within-Device Fluctuation," *IEEE Trans. Electron Devices,* vol. 60, no. 8, pp. 2505-2511, 2013.

[153] M. Duan, J. Zhang, A. Manut, Z. Ji, W. Zhang, A. Asenov, L. Gerrer, D. Reid, H. Razaidi, and D. Vigar, "Hot carrier aging and its variation under use-bias: Kinetics, prediction, impact on Vdd and SRAM," in *IEDM Tech. Dig.*, 2015, pp. 20.4. 1-20.4. 4.

[154] M. Toledano-Luque, B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, and G. Groeseneken, "Toward a streamlined projection of small device bias temperature instability lifetime distributions," *J. Vac. Sci. Technol. B,* vol. 31, no. 1, p. 01A114, 2013.

[155] M. Toledano-Luque, B. Kaczer, J. Franco, P. J. Roussel, T. Grasser, T. Y. Hoffmann, and G. Groeseneken, "From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation," in *VLSI Symp. Tech. Dig.*, 2011, pp. 152-153.

[156] M. Toledano-Luque, B. Kaczer, J. Franco, P. J. Roussel, M. Bina, T. Grasser, M. Cho, P. Weckx, and G. Groeseneken, "Degradation of time dependent variability due to interface state generation," in *VLSI Symp. Tech. Dig.*, 2013, pp. T190-T191.

[157] H. Miki, N. Tega, M. Yamaoka, D. Frank, A. Bansal, M. Kobayashi, K. Cheng, C. D'Emic, Z. Ren, and S. Wu, "Statistical measurement of random telegraph noise and its impact in scaled-down high-κ/metal-gate MOSFETs," in *IEDM Tech. Dig.*, 2012, pp. 19.1. 1-19.1. 4.

[158] A. Zhang, C. Huang, T. Guo, A. Chen, S. Guo, R. Wang, R. Huang, and J. Xie, "Reliability variability simulation methodology for IC design: An EDA perspective," in *IEDM Tech. Dig.*, 2015, pp. 11.5. 1-11.5. 4.

[159] M. Duan, J. Zhang, Z. Ji, J. Ma, W. Zhang, B. Kaczer, T. Schram, R. Ritzenthaler, G. Groeseneken, and A. Asenov, "Key issues and techniques for characterizing time-dependent device-to-device variation of SRAM," in *IEDM Tech. Dig.*, 2013, pp. 31.3. 1-31.3. 4.

[160] M. Duan, J. F. Zhang, Z. Ji, W. D. Zhang, D. Vigar, A. Asenov, L. Gerrer, V. Chandra, R. Aitken, and B. Kaczer, "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.

[161] 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.

[162] M. Duan, J. Zhang, J. Zhang, W. Zhang, Z. Ji, B. Benbakhti, X. Zheng, Y. Hao, D. Vigar, and F. Adamu-Lema, "Interaction between hot carrier aging and PBTI degradation in nMOSFETs: Characterization, modelling and lifetime prediction," in *Proc. IRPS*, 2017, pp. XT-5.1-XT-5.7.

List of Publications

Journals

1. **R. Gao**, A. B. Manut, Z. Ji, J. Ma, M. Duan, J. F. Zhang, J. Franco, S. W. M. Hatta, W. D. Zhang, and B. Kaczer, "bility by Extrapolation," *IEEE Trans. Electron Devices,* vol. 64, no. 4, pp. 1467-1473, 2017.
2. **R. Gao**, Z. Ji, J. F. Zhang, W. D. Zhang, S. F. W. M. Hatta, J. Niblock, P. Bachmayr, L. Stauffer, K. Wright, and S. Greer, "A Discharge-Based Pulse Technique for Probing the Energy Distribution of Positive Charges in Gate Dielectric," *IEEE Trans. Semicond. Manuf.,* vol. 28, no. 3, pp. 221-226, 2015.
3. **R. Gao**, Z. Ji, A. B. Manut, J. F. Zhang, J. Franco, S. W. M. Hatta, W. D. Zhang, B. Kaczer, D. Linten, and G. Groeseneken, "NBTI-Generated Defects in Nanoscaled Devices: Fast Characterization Methodology and Modeling," *IEEE Trans. Electron Devices*, vol. 64, no. 10, pp. 4011-4017, 2017.
4. Z. Ji, X. Zhang, J. Franco, **R. Gao**, M. Duan, J. F. Zhang, W. D. Zhang, B. Kaczer, A. Alian, and D. Linten, "An investigation on border traps in III–V MOSFETs with an In 0.53 Ga 0.47 As channel," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3633-3639, 2015.

Conference

1. **R. Gao**, Z. Ji, S. Hatta, J. Zhang, J. Franco, B. Kaczer, W. Zhang, M. Duan, S. De Gendt, and D. Linten, "Predictive As-grown-Generation (AG) model for BTI-induced device/circuit level variations in nanoscale technology nodes," in *IEDM Tech. Dig.*, 2016, pp. 31.4. 1-31.4. 4.
2. P. Ren, **R. Gao**, Z. Ji, H. Arimura, J. F. Zhang, R. Wang, M. Duan, W. Zhang, J. Franco, S. Sioncke, D. Cott, J. Mitard, L. Witters, H. Mertens, B. Kaczer, A. Mocuta, N. Collaert, D. Linten, R. Huang, A. V.-Y. Thean, and G. Groeseneken, "Understanding charge traps for optimizing Si-passivated Ge nMOSFETs," in *VLSI Symp. Tech. Dig.*, 2016, pp. 1-2.
3. Z. Ji, J. F. Zhang, L. Lin, M. Duan, W. Zhang, X. Zhang, **R. Gao**, B. Kaczer, J. Franco, T. Schram, N. Horiguchi, S. D. Gendt, and G. Groeseneken, "A test-proven As-grown-Generation (A-G) model for predicting NBTI under use-bias," in *VLSI Symp. Tech. Dig.*, 2015, pp. T36-T37.
4. Z. J. Brown, **R. Gao**, Z. Ji, J. Chen, J. Wu, J.F. Zhang, B. Zhou, Q. Shi, J. Crowford, and W.D. Zhang “A low-power and high-speed True Random Number Generator using generated RTN,” in *VLSI Symp. Tech. Dig.*, 2018 (to be published).