Development of characterization techniques for negative bias temperature instabilities

Lin Lin

Liverpool John Moores University

2012

THE FOLLOWING HAVE NOT BEEN COPIED ON INSTRUCTION FROM THE UNIVERSITY

Acknowledgements

The PhD degree is the final reward of a long educational journey. The journey with difficulties, puzzle, disappointment and surprises.

This thesis is results of this long journey and supported by many people. I cannot explain in words what I feel today, when finally I have the opportunity to express my gratitude for all of them.

First and foremost, I owe my deepest gratitude to Professor John Zhang. From the very first day, you have proven to be a very kind and helping person. Having you as my supervisor is the best thing that happened to me during my PhD. Thank you for your continuous unconditional support, encouragement, guidance and advice in all possible way.

I also owe a lot of gratitude to my second supervisor Dr. Weidong Zhang. He has been so kindly to support me and solve any questions I had.

I am also indebted to Dr. Zhigang Ji. This thesis would not have been possible without his continuous support, encouragement. He always tried his best to help me with my puzzles.

Abstract

The requirements for ever faster circuits and higher packing density have driven the continuous downscaling of the transistor sizes in the last 50 years or so. This leads to higher electrical field and operation temperature and, in turn, accelerates the degradation. One of the most serious reliability issues for the current CMOS technology is the negative bias temperature instability (NBTI). This project will focus on investigating the NBTI and the positive charges responsible for it.

Modern MOSFETs use gate dielectrics in the nanometer range and the degradation will recover rapidly. To suppress the recovery, high speed characterization technique is needed. In this project the measurement speed has been improved from 5μs to 200 ns for Id-Vg measurements and 800ns for C-V measurement.

As a Hf-dielectric/SiON stack is replacing SiON as the gate dielectric, the task is to identify which layer of the stack dominates positive charging (PC). A main achievement in this project is the finding that PCs are dominated by the interfacial layer (IL) and they do not pile up at the HfSiON/(IL) interface.

Evaluating the conventional threshold voltage shift measured by extrapolating transfer characteristics, $\Delta Vth(ex)$, underestimates the NBTI-induced degradation of drain current, ΔId. In this project we proposed the effective threshold voltage shift, $\Delta Veff$, in order to evaluate the devices degradation correctly.

The next task was to develop a lifetime prediction method, based on ΔV_{eff}. To predict the worst-case lifetime which is recovery free, a model for NBTI kinetics under operation gate bias was developed. This kinetics includes contributions from both as-grown and generated defects and it no longer follows a simple power law. Based on the new kinetics, a single test prediction method was proposed and its safety margin is estimated to be 50%.

A fast single pulse charge pumping (SPCP) technique was developed in this project, reducing the measurement time to microseconds. By exploring the differences in the transient currents corresponding to the two edges of the gate pulse, the net charges pumped into devices can be obtained and their saturation level is used to evaluate interface states. For the first time, SPCP allows the recovery of interface states to be monitored with a time resolution in microseconds. The results show that the recovery of stress-induced interface states is substantial within $100\mu s$, which would be missed if conventional charge pumping were used.

Contents

List of abbreviations

Abbreviation	Signification
ALCVD	Atomic Layer Chemical Vapor Deposition
ALD	Atomic Layer Deposition
CP	Charge Pumping
DPN	Decoupled Plasma Nitridation
EOT	Equivalent Oxide Thickness
FG	Forming gas
HCl	Hot carrier injection
HfO2	Hafnium dioxide
HfSiON	Hafnium Silicate Oxide Nitride
IL	Interfacial layer
MOCVD	Metal Organic Chemical Vapor Deposition
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
NBTI	Negative Bias Temperature instability
PDA	Post Deposition Anneal
PVD	Physical Vapor Deposition
SILC	Stress Induced Leakage Current
SiON	Silicon Oxide Nitride
TaN	Tantalum Nitride
TDDB	Time Dependent Dielectric Breakdown
TiN	Titanium Nitride

List of symbols

Symbol	Description	Unit
μ_{eff}	Effective mobility	$cm^{-2}/V\text{-}s$
ε_{SiO2}	Dielectric constant of SiO_2	
ε_{Si}	Dielectric constant of Si	
ε_0	Electric permittivity of vacuum	F/cm
ε_{IL}	Dielectric constant of the interfacial layer	
C_{ox}	Oxide capacitance	F
$C_{s,lf}$	Low-frequency substrate capacitance	F
E_f	Fermi level	eV
E_{eff}	Effective surface field in the Si substrate	MV/cm
f	Frequency	Hz
g_d	Drain conductance	S
g_m, G_m	Transconductance	S
I_d	Drain current	A
I_g	Gate current	A
J_g	Gate current density	A/cm^2
L	Mask channel length	μm
L_D	Debye length	cm
N_A	Substrate doping density	cm^{-3}
N_{it}	Interface trap density	cm^{-2}
n_i	Intrinsic carrier concentration in Si substrate	cm^{-3}
N^0	Total density of Si-H bonds	
$N_H(0)$	Free H_0 at the interface	
q	One electron charge	C
Q_{inv}	Inversion charge density	C/cm^2

Q_b	Depletion charge density	C/cm^2
R	Feedback resistance	Ω
R_{sd}	Series resistance at source and drain	Ω
R_{ch}	Channel resistance	Ω
T	Temperature	$^{\circ}C$
X_{ox}	Equivalent oxide thickness	nm
X_c	Centroid of the charge	
U_F	Normalized Fermi potential	
U_S	Normalized surface potential	
V_d	Drain voltage	V
V_{fb}	Flat band voltage	V
V_g	Voltage applied on the gate	V
V_{ox}	Voltage drop across the oxide	V
V_t	Threshold voltage	V
ΔV_t	Threshold voltage shift	V
W	Mask channel width	μm

1 | A Review of the Degradation and Defects of MOSFETs

1.1   Introduction

We have witnessed the microelectronics-led revolution and its substantial impact on our everyday life in the last 50 years or so. The most successful microelectronic technology is the CMOS and it has captured over 90% of the market share now. This success heavily relies on the excellent insulating properties of gate silicon dioxides/oxynitrides and their near perfect interface with silicon. The gate silicon dioxides/oxynitrides are amorphous insulators with a large bandgap of about 9eV and high energy barriers for free electrons (3.2eV) and holes (4.8eV) from silicon [1]. They can be grown on Si with low defect density.

To increase the packing density and the operation speed, the transistor size has been downscaled ever since the invention of integrated circuits in 1958. In 1965, Gordon Moore, one of the founders for Intel, predicted that the number of transistors used per chip would double every 18~24 months [2]. The semiconductor industry has followed his prediction since then and this is known as the Moore's Law [3,4] now.

Before 1990s, the operation voltage of CMOS technologies was maintained at 5V, when the sizes of MOSFETs were downscaled. This leads to an increase of electrical field within the devices, as shown in Fig.1.1 [5]. A higher field increases the leakage current and decreases the device lifetime. To control the leakage current and achieve

the required lifetime, the operation voltage has been reduced gradually since 1990. This allowed the electrical field being broadly kept as a constant in 1990s, but the electrical field is increasing again recently, as illustrated by Fig.1.1. Since the bandgap of silicon is 1.12eV and it does not change with downscaling, the operation voltage cannot be a lot lower than 0.8V. For a modern CMOS technology, the operation voltage is already close to 1V, so that there is little room for further reduction of operation voltage. The electrical field is expected to increase further with downscaling in the future. The basic equation to calculate electrical field is shown in below

$$E_{ox} = \frac{V_g - V_{poly} - \emptyset_s - \emptyset_{ms}}{t_{ox}}$$

Where $\emptyset_s$ is the substrate surface potential and $\emptyset_{ms}$ is the work function.

Fig.1.1 Evolution of oxide and silicon electric fields showing 3 different scaling scenario periods [5].

A higher electrical field always accelerates the degradation of MOSFETs and shortens their lifetime. After device fabrication, the degradation can occur during the

device operation in a number of ways, including hot carrier stresses [6-15], gate dielectric breakdown [16-27], Fowler-Nordheim injections (FNI) [28-37], positive bias temperature instabilities [38-47], and negative bias temperature instabilities [48-62].

Degradation can also happen during device fabrication. For a modern CMOS technology, plasma processing is essential and plasma process-induced damages (PID) must be controlled [63-68]. The PID is most severe during the plasma etching of metals, where the metal wires connected to a floating gate act as antenna and collect charges. This builds up a high field over the gate dielectrics and can result in latent damage [63-68].

Apart from the high field-induced degradations mentioned above, irradiation is another source of device degradation. Here degradation can happen both during [69-74] and post [75-78] irradiation. During irradiation, the energetic photons can create electron-hole pairs in the gate oxides. The electrons are typically swept out the oxide, but holes will be captured and form positive charges [69-74]. After irradiation, interface states can increase continuously under a low positive gate bias [75-78].

In the rest of this chapter, a brief review will be given to the main degradation processes: section 1.2 on hot carriers, section 1.3 on time-dependent dielectric breakdown (TDDB), section 1.4 on the post-stress degradation, section 1.5 on latent defect creation, section 1.6 on positive bias temperature instabilities (PBTI), and finally section 1.7 on negative bias temperature instabilities (NBTI). Through this

review, the rationale for the selection of research topics in this project will be explained.

1.2   Hot carrier stresses

Hot carriers are the energetic charge carriers that can cause damages to MOSFETs. When MOSFETs operate in their saturation mode, Fig.1.2 shows that the conduction channel will be pinched off and there is a space charge region near the drain. This results in a high lateral field that accelerates charge carriers. Some of the 'lucky' charge carriers can gain a high energy before collision. When they eventually collide with other atoms, they can cause damage. Since the high field only exists near the drain, the hot carriers and their damage also are localized near the drain. Although hot carriers induce damage in both nMOSFETs and pMOSFETs, the damage is typically higher in nMOSFETs, since electron mobility is around three times of hole mobility in silicon, allowing electrons gaining a higher speed and energy for a given electrical field. As a result, the attention will be focused on hot electron induced damage in nMOSFETs hereafter.
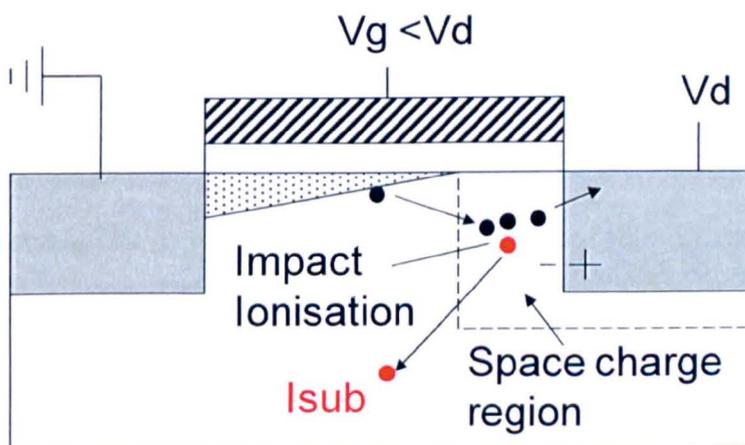
Fig.1.2 Generation of hot carriers in the space charge region when an nMOSFET operates in the saturation region.

Hot carriers can cause damage in two ways: creating interface states and forming space charges in the dielectrics. At the interface between SiO_2 and silicon, the number of oxygen atoms is inadequate to bond every silicon atom, so that some silicon atoms have dangling bonds [79-81]. It is well accepted that these dangling bonds, so-called Pb centers, introduce states into the bandgap of silicon [79-81]. To passivate these interface states, devices are annealed in an ambient containing hydrogen at a temperature around 400°C, which is a standard industry process. As illustrated in Fig.1.3, this leads to the formation of Si-H bonds:

$$Si_3 \equiv Si\bullet + H_2 \rightarrow Si_3 \equiv Si\text{-}H + H.$$
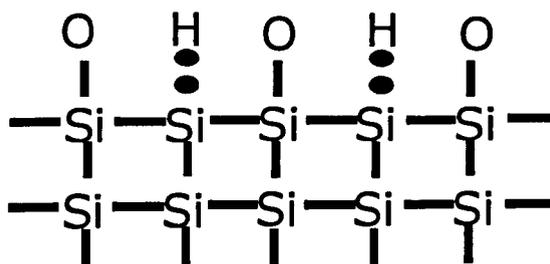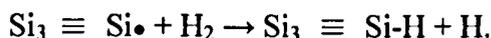
Fig.1.3 Passivation of silicon dangling bonds at the SiO_2/Si interface by annealing in a hydrogen ambient.

Through bombarding the interface, hot electrons can rupture the Si-H bonds and create interface states.

Some hot electrons can gain enough energy to be injected into the oxides. They can then be captured by electron traps in the oxides, and form space charges, as shown in Fig.1.4 On the relative importance of interface state creation versus electron trapping in oxides, it is generally agreed that the hot electron induced damage is dominated by the interface state creation [6,7]. This is partially because the most severe hot electron stress occurs under the bias condition of Vg=Vd/2 [9], so that the gate of

nMOSFETs is negatively biased against drain, impeding electron injection into the gate oxide. For the high quality thin oxides used in modern CMOS technologies, there are few pre-existing electron traps [82-84] and steady electron trapping is difficult due to detrapping by tunneling.
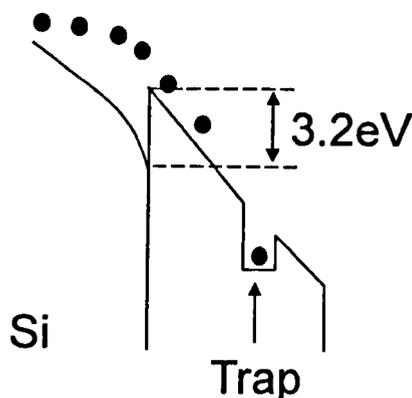
Fig.1.4 Formation of space charges in the gate oxide by capturing the injected hot electrons.

To explain why the hot electron stress is most severe under Vg=Vd/2, two factors must be considered: the electrical field in the space charge region near the drain and the number of electrons passing through this region. On one hand, for a given Vd, an increase of Vg allows a larger number of electrons through the space charge region near the drain, resulting in more hot electrons initially. On the other hand, a higher Vg will reduce the difference between Vg and Vd and in turn the size of space charge region near the drain. A smaller space charge region gives weaker electrical field and less hot electrons. When Vg>Vd/2 approximately, the weakening field has a larger effect on the hot electrons than the increased number of electrons passing through it, so that overall hot electrons reduce. As a result, hot electrons are negligible at the steady state and they mainly occur during the switching of gate bias.

The hot electron induced damage has attracted a lot of attention since 1980s. When the channel length becomes smaller, the leakage between source and drain increases, since the pn junction at the source can be affected by the bias at the drain through the drain-induced barrier lowering (DIBL). To control the leakage current, the doping density of silicon has been increased, since a higher density of space charges will be able to better screen the source from the drain. Unfortunately, a higher density of space charges will produce a higher electrical field and in turn more hot electrons. The hot electron induced damage becomes so severe that it is generally accepted that it limits the lifetime of nMOSFETs. The lifetime of nMOSFETs used to be shorter than that of pMOSFETs for CMOS technologies earlier than the 0.18μm generation.

At present, hot carrier induced degradation is still important and it is an industrial standard test for qualifying every new CMOS process. It is generally believed, however, that pMOSFETs have shorter lifetime than nMOSFETs due to the enhanced negative bias temperature instabilities (NBTI) for the current CMOS technologies, as to be described in section 1.7. As a result, a lot of attention has been switched from hot carrier induced degradation to the NBTI recently. This project will also focus on the NBTI, rather than hot carrier induced degradation.

1.3 Time dependent dielectric breakdown (TDDB)

For the early generation of CMOS technologies, gate oxide breakdown was typically caused by extrinsic defects, such as pinholes and contaminating species. There is little carrier injection into the gate oxides by Fowler-Nordheim tunneling during the operation and defect creation in the bulk of oxides is insignificant. As the oxide

thickness drops below ~3 nm approximately, electrons can now pass through the gate oxide by direct tunneling during the normal device operation. Strictly speaking, the gate oxides are no longer an insulator in this case and the intrinsic breakdown of gate oxides becomes a severe problem for industry.

When electrons are transported through gate oxides under a high electrical field, defects will be created. Although the number of electrons passing through the oxides is typically much higher than any other species, it is generally agreed that electrons will not create defects directly [16,28,85-90]. The electrons will release a damaging species that interacts with the oxides and generating defects. Agreement has not been reached on the identity of the damaging species. Some researchers [85-88] believe that they are hydrogenous species, while others [16,28,89,90] propose that they are holes. Recent experimental evidences show that both hydrogenous species and holes can cause defect creation [91].
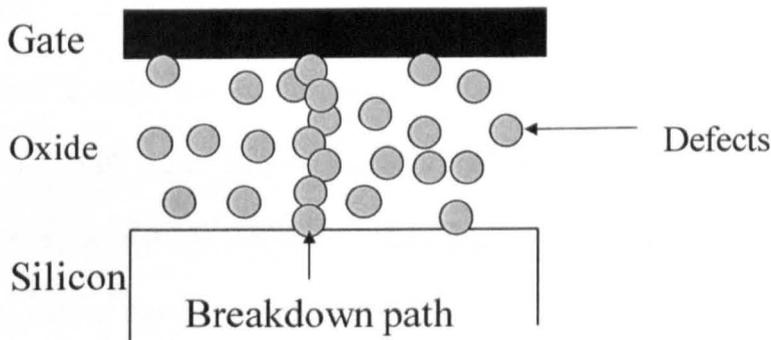
Fig.1.5 The generated defects in the gate oxide overlap with each other and form a conduction path through the oxide, triggering the breakdown.

The most successful model for oxide breakdown is the percolation model [16-19,22]. This model assumes that each created defect has a finite size and the defect distribution is completely random in the oxides. As the generated defects accumulate,

they overlap with each other and Fig.1.5 shows that a conduction path is eventually formed when the defect bridges the two electrodes, which triggers the breakdown. This model has been used to successfully predict the dependence of oxide breakdown on the gate area and the oxide thickness [16-19].

Different types of defects can be generated during the breakdown tests: interface states, electron traps, and hole traps, but they do not contribute equally to the breakdown [83]. The interface states are located at the oxide/silicon interface and do not agree with the random spatial distribution required by the percolation model. Hole trap generation is not thermally activated [83], disagreeing with the thermal acceleration of breakdown process [18,19]. Among the created electron traps, some of them can only be filled at low oxide field (e.g. 2MV/cm) and were referred to as "low-field traps", whilst others can capture electrons at an oxide field over 10 MV/cm (the so called "high-field traps") [82,83]. The creation of low-field electron traps saturates as stress increases, so that they will not trigger breakdown.

Among high-field traps, two capture cross sections are clearly identified, one is in the order of $10^{-13} \sim 10^{-14}$ cm^2 and the other in the order of $10^{-15} \sim 10^{-16}$ cm^2 [84]. The generation of traps with the smaller capture cross section again saturates as stress increases and cannot cause the breakdown. The creation of traps with a capture cross section in the order of $10^{-13} \sim 10^{-14}$ cm^2, however, does not saturate. Its distribution through the oxide is uniform on a macroscopic scale, in agreement with a statistically random distribution on a microscopic scale [83]. Its physical size of the order of nanometer also agrees well with the defect size independently determined by the breakdown tests [16,22]. This strongly supports that the high-field electron traps

with a capture cross section in the order of $10^{-13} \sim 10^{-14}$ cm^2 is responsible for the oxide breakdown.

It should be noted that the thickness of the gate oxides is around 1nm for the state-of-the-art MOSFETs, which is comparable with the electron trap size. As a result, generating one electron trap will be sufficient to give a stress-induced leakage current (SILC) through the gate oxide and the overlap of two traps is sufficient to trigger breakdown for modern MOSFETs.

Although a gate voltage ramp was used in the early breakdown tests, the standard breakdown test is carried out under either a constant voltage or a constant current nowadays. The time and electron fluency at the breakdown are recorded and the Weibull distribution is used to present the results. To achieve a reliable statistical distribution, the number of test samples has to be in the order of hundreds [27]. The test procedure for the time dependent dielectric breakdown (TDDB) is well established in industry now and this project will not address it further.

1.4 Post-stress degradation

The post-stress degradation was investigated after a device was subjected to either irradiation [75-78] or electrical stresses [92-98]. It was found that under a moderate positive gate bias (e.g. + 1MV/cm), interface states build-up continues after terminating the stress. Three theories have been proposed: hydrogen transportation [75-78], trapped hole conversion [92] and hydrogen emission [98].

The hydrogen transportation model was proposed mainly to explain the degradation post-irradiation [75-78]. Here, mobile hydrogenous species were formed in the bulk of dielectrics during the irradiation, as illustrated by Fig.1.6 After the irradiation, these mobile hydrogenous species gradually move to the SiO_2/Si interface under a positive gate bias. Once arrived at the interface, they create interface states through a reaction such as,

$$Si_3 \equiv SiH + H^+ + e \rightarrow Si_3 \equiv Si\bullet + H_2.$$
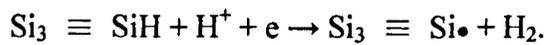
This reaction is considered to be rapid and the generation rate is believed to be controlled by the hydrogen transportation. Although this model has been used successfully to explain the delayed build-up of interface states following irradiation, it is found that the hydrogen transportation through a relatively thin oxides (<15nm) is too fast to be responsible for the slow increase of interface states in the modern MOSFETs post electrical stresses [94-98].
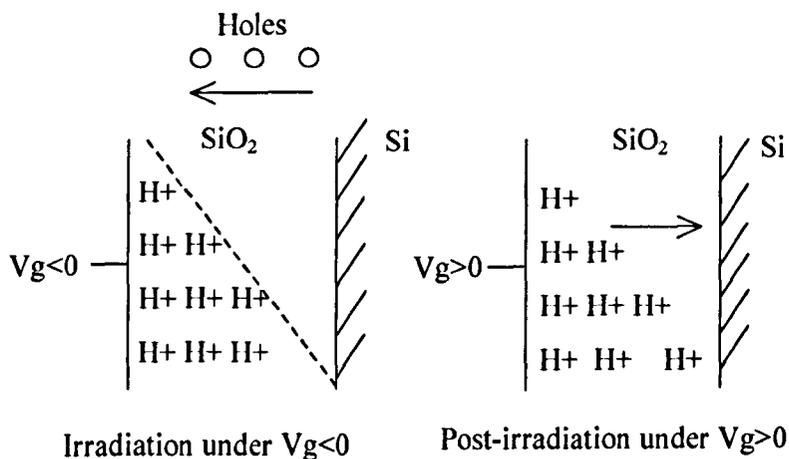
Fig.1.6 A schematic illustration of the hydrogen transportation model for interface state generation post-irradiation.

The trapped hole conversion model is schematically shown in Fig.1.7 It is assumed that there are strained bonds near to the interface, which can be hole traps [92]. The bond is broken by capturing a hole. After the trapped hole is neutralized, the strained bond will not be restored. Instead, a weak or broken bond is produced, which acts as an interface state. This model predicts that the generation of interface states should be proportional to the number of holes detrapped. The experimental evidences, however, do not agree with this prediction [94-97].
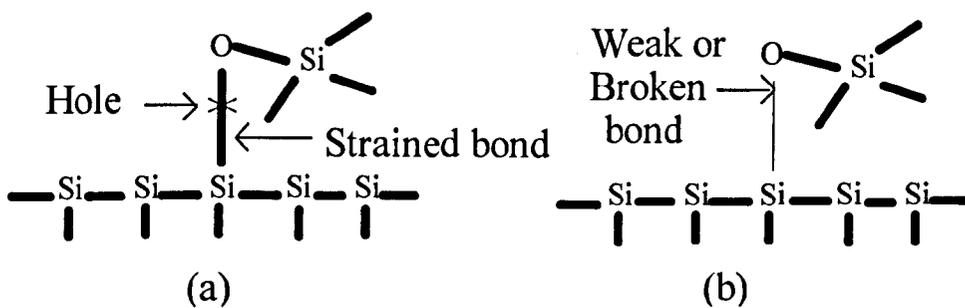
Fig.1.7 A schematic illustration of the trapped hole conversion model for interface state generation.

The hydrogen emission model assumes that, following the neutralization of trapped holes, the emission of a neutral hydrogenous species is the rate-limiting process. To facilitate the discussion, the defect responsible for the post-stress degradation can be schematically represented by D—H. The hole trapping and detrapping can be schematically shown as:

$$D—H + hole \rightarrow D—H^+$$

and

$$D—H^+ + electron \rightarrow D---H.$$

After the detrapping, the original D—H bond can be weakened and is represented by the dashed line. This leads to the emission of neutral hydrogenous species, since the hydrogenous species is neutral, the defect should remain neutral.

$$D\text{---}H \rightarrow D + H. \tag{1}$$

Like the hydrogen transportation model, the hydrogen then moves to the SiO_2/Si interface and create interface states. Unlike the hydrogen transportation model, however, the generation rate is limited by the hydrogen emission process shown in (1), rather than its transportation. This model has been used successfully to explain the slow build-up of interface states following substrate hot hole stresses [98].

1.5   Latent defect generation

The latent defect generation was reported after plasma processing. Fig.1.8 shows that the plasma etch induced damage can be annealed in forming gas at $400^{\circ}C$. The annealed sample, however, is not the same as the sample that did not experience plasma etching. When the plasma-etched and then annealed sample was stressed, the generated defect is substantially higher than that in a wet-etched one for the same electrical stress. This means that the defects created by the plasma process were made 'latent' by the forming gas anneal, but they were not eliminated.

Fig.1.8 Generation of latent defects by plasma etching. (a) is the interface states measured after plasma etching. (b) is the interface states after annealing the plasma etched sample. (c) is the interface states after electrical stressing the plasma-etched and then annealed samples. The dashed line is the interface states after the same electrical stress on a wet-etched sample [63].

It has been identified that the trapped holes and hydrogen exposure are the two essential conditions for the formation of latent defects. For example, Fig.1.9 shows that, if the trapped holes are neutralized before the hydrogen exposure by FN injection, the formation of latent defects is effectively suppressed. This observation leads to the proposal of the two-stage H_2-cracking model [68]. In the first stage, H_2 is cracked at a trapped hole site into reactive hydrogenous species, such as H^+ and H^o. These reactive hydrogenous species then react with the device to form the latent defects. Based on this model, the guide-line for suppressing the latent defects is that the positive charges in the oxides induced by plasma processing must be neutralized before the forming gas anneal [68].

Fig.1.9 (a) Generation of latent interface states. After the substrate hole injection (SHI), one device was exposed to forming gas (FG) directly and another device was subjected to FN injection first and then the same FG anneal. When both devices were stressed by SHI again, the interface states created in the device with FNI is much lower than that without FNI (the dashed line). (b) shows the behavior of trapped holes. [68].

1.6   Positive bias temperature instabilities (PBTI)

Two different types of PBTIs were reported. The first type is for pMOSFETs [38-41]. It is found that substantial amount of positive charges can be formed when a positive gate bias was applied to pMOSFETs at elevated temperature. The source of this PBTI has been identified as the defects involving both boron and water [38]. After exposing pMOSFETs to wet nitrogen, PBTI will increase substantially [38]. Moreover, Fig.1.10 illustrates that the water diffuses into the device through the gate edges [38]. This means that this type of PBTI for pMOSFETs can be effectively suppressed through proper capsulation.
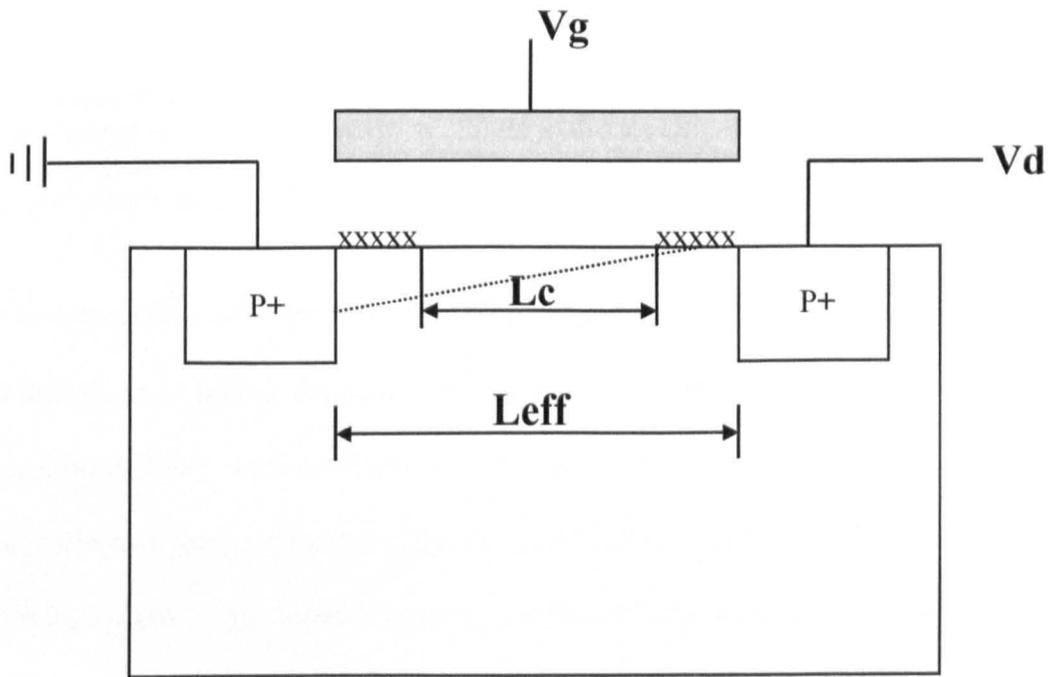
(a)

Fig.1.10 (a) A schematic diagram showing the nonuniform degradation of MOSFET's under PBTS. The degraded region is marked out by 'x'. (b) shows that the PBTI effect increases as channel length reduces, because the degraded region is a larger percentage of the whole channel for devices of smaller channel length. [38]

The second type of PBTI is for nMOSFETs with high-k dielectric stacks. During the initial stage of high-k dielectric investigation, it is widely reported that substantial electron trapping occurs under a positive gate bias [42-47]. For a given positive gate bias, electron trapping increases the threshold voltage of nMOSFETs and reduces the driving current. The electron trapping is rapid and can reach a substantial level even during a typical transfer characteristics measurement with a time in the order of seconds. When this trapping-reduced drain current is used to estimate the electron mobility, an underestimation occurs. To extract the real electron mobility, pulse measurement with a time in the order of tens of microseconds must be used [42-45]. The pulse measurement is also essential for probing the real density of electron traps and assessing their capture cross sections [43,44]. Early work [99] used quasi-DC

measurement and only captured a fraction of electron traps, due to the detrapping during the measurement. This leads to a gross underestimation of the capture cross section [99]. It is found that the detrapping can be frozen, only when the measurement time is reduced to the order of tens of microseconds [43,44].

Unlike SiO_2 that has little pre-existing electron traps, the electron traps in HfO_2 is as-grown, rather than generated by electrical stresses [100,101]. Despite of the difference between HfO_2 and SiO_2, it is found that the two capture cross section of as-grown electron traps in HfO_2 is similar to those for the generated electron traps in SiO_2: one in the order of 10^{-14} cm^2 and the other in the order of 10^{-16} cm^2 [44].

On the special location, it is reported that the as-grown electron traps are mainly in the high-k layer and there is little traps in the interfacial SiO_2 or $SiON$ [100,101]. Importantly, Fig.1.11 shows that electron trapping becomes negligible when the Hf-dielectric layer is thinner than 2 nm approximately, which is the relevant thickness for its real application in CMOS technologies. As a result, the electron trapping induced PBTI for industrial nMOSFETs is not as severe as that observed earlier on HfO_2 of 3 nm or thicker [100,101]. The PBTI of nMOSFETs will not be addressed further in this project.

Fig.1.11 The location of as-grown electron traps in Hf-dielectric/SiO_2 gate stack is the shaded region. The electron-trapping induced PBTI is insignificant for Hf-dielectric layer thinner than 2 nm. [101]

1.7 Negative bias temperature instabilities (NBTI)

Negative bias temperature instability (NBTI) takes place in pMOSFETs under negative gate voltage at elevated temperature. NBTI has a number of adverse effects on devices, such as a reduction of drain current and transconductance g_m, and an increase of the magnitude of threshold voltage. It is the earliest instability reported for MOS devices. In 1967, Deal et al. [48] reported that both the interface trap density N_{it}, and oxide charge density N_{ox}, increased upon negative bias stress. The rate of increase of N_{it} and N_{ox} was similar.

The next milestone is the proposal of the reaction-diffusion model (R-D) by Jeppson and Svensson in 1977 [50]. In this model, the degradation will start with an electrochemical reaction at the SiO_2/Si interface, which converts the precursors (e.g. Si-H) into interface states and release a hydrogenous species. The hydrogenous

species then diffuse away from the interface and this diffusion process limits the generation of interface states. The R-D model successfully explains the power law dependence of the interface state generation against time.

Fig.1.12 The transition of lifetime limitation mechanisms as a function of gate oxide thickness. When the thickness is below 3.5 nm, degradation due to NBTI becomes to limit the device lifetime [102].

NBTI received little attentions in the 1980s and most part of 1990s, since the efforts were focused on the hot carrier induced degradation in this period, as mentioned in section 1.2. As the downscaling continues, the boron penetration from the p+ poly gate through the gate oxides becomes a serious issue. To suppress boron penetration, an increasing amount of nitrogen has been used in the oxynitrides. Nitridation increases NBTI [102-104]. This, together with an increase of operation temperature and oxide field, have raised NBTI to such a level that pMOSFETs has shorter lifetime than the hot-carrier induced lifetime of nMOSFETs, as illustrated in Fig. 1.12 [102]. As a result, the NBTI research has experienced a renaissance in the new millennium and it is the most important reliability issue for CMOS technologies now.

This project will focus on studying NBTI and developing the required characterization techniques.

References

1. S. M. Sze, "Physics of semiconductor devices," J. Willy & Sons, 1981.

2. G. E. Moore, "Cramming more components onto integrated circuits," Electronics, **38**, 114 (1965).

3. R. R. Schaller, "Moore's law: past, present, and future," IEEE Spectrum, **34**, 52 (1997).

4. R. D. Issac, "The future of CMOS technology," IBM J. Res. Dev., **44**, 369 (2000).

5. G. Groeseneken, R. Degraeve, B. Kaczer, K. Martens, "Trends and perspectives for electrical characterization and reliability assessment in advanced CMOS technologies," in Proc. European Solid-State Device Research Conference (ESSDERC), 64, 2010.

6. C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-electron-induced MOSFET degradation --- Model, Monitor, and Improvement," IEEE J. Solid-State Cir., **SC-20**, 295 (1985).

7. G. Groeseneken, R. Bellens, G. Van den bosch and H .E. Maes, "Hot-carrier degradation in submicrometre MOSFETs: from uniform injection towards the real operating conditions," Semicond. Sci. Tech., **10**, 1208 (1995).

8. J. F. Zhang and W. Eccleston, "Effects of high field injection on the hot carrier induced degradation of submicrometer pMOSFET's," IEEE Trans. Elec. Dev., **42**, 1269 (1995).

9. C. T. Wang, "Hot carrier design consideration for MOS devices and circuits," Van Nostrand Reinhold, New York, 1992.

10. P. E. Cottrell, R. R. Troutman and T. H. Ning "Hot-electron emission in n-channel IGFETS," IEEE Trans. Elec. Dev., **26**, 520 (1979).

11. W. L. Chen and T. P. Ma, "Channel hot-carrier induced oxide charge trapping in NMOSFET's," in Tech. Digest IEEE 37th Int. Electron Devices Meeting, Washington, p.731, USA (1991).

12. D. J. DiMaria, "Defect generation in field-effect transistors under channel-hot-electron stress," J. Appl. Phys., **87**, 8707 (2000).

13. M. Brox and W. Weber, "Dynamic degradation in MOSFET's – Part I: The physical effects," IEEE Trans. Elec. Dev., **38**, 1852 (1991).

14. W. L. Chen, A. Balasinski and T. P. Ma, "Lateral profiling of oxide charge and interface traps near MOSFET junctions," IEEE Trans. Elec. Dev., **40**, 187 (1993).

15. V. H. Chan and J. E. Chung, "The impact of nMOSFET hot-carrier degradation on CMOS analog subcircuit performance", IEEE J. Solid-St. Circ., **30**, 644 (1995).

16. R. Degraeve, G. Groeseneken, R. Bellens, J. L. Ogier, M. Depas, P. J. Roussel and H. E. Maes, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," IEEE Trans. Elec. Dev., **45**, 904 (1998).

17. R. Degraeve, B. Kaczer and G. Groeseneken, "Ultra-thin oxide reliability: searching for the thickness scaling limit," Microelectronics Reliability, **40**, 697 (2000).

18. B. Kaczer, R. Degraeve, N. Pangon, T. Nigam and G. Groeseneken, "Investigation of temperature acceleration of thin oxide time-to-breakdown," Microelectronic Eng., **48**, 47 (1999).

19. B. Kaczer, R. Degraeve, N. Pangon and G. Groeseneken, "The influence of elevated temperature on degradation and lifetime prediction of thin silicon-dioxide films," IEEE Trans. Elec. Dev., **47**, 1514 (2000).

20. D. J. DiMaria, E. Cartier and D. Arnold, "Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon," J. Appl. Phys., **73**, 3367 (1993).

21. D. J. DiMaria, "Temperature-dependence of trap creation in silicon dioxide," J. Appl. Phys., **68**, 5234 (1990).

22. J. H. Stathis, "Percolation models for gate oxide breakdown," *J. Appl. Phys.*, vol. **86**, no. 10, pp. 5757–5766, Nov. 1999.

23. D. J. Dumin, J. R. Maddux, R. S. Scott and R. Subramoniam, "A model relating wearout to breakdown in thin oxides," IEEE Trans. Elec. Dev., **41**, 1570 (1994).

24. B. Schlund, C. Messick, J. S. Suehle and P. Chaparala, "A new physics-based model for time-dependent-dielectric-breakdown," in Proc IEEE 34th Int. Reliability Phys. Symp., Dallas, p.84, USA (1996).

25. J. W. McPherson and H. C. Mogul, "Disturbed bonding states in SiO2 thin-films and their impact on time-dependent dielectric breakdown," in Proc IEEE 36th Int. Reliability Phys. Symp., Reno, p.47, USA (1998).

26. M. Kimura, "Field and temperature acceleration model for time-dependent dielectric breakdown," IEEE Trans Elec. Dev., **46**, 220 (1999).

27. E. Y. Wu, J. H. Stathis, and L. K. Han, "Ultra-thin oxide reliability for ULSI applications," Semicond. Sci. Technol., **15**, 425, (2000).

28. J. F. Zhang, S. Taylor, and W. Eccleston, "Electron trap generation in thermally grown SiO_2 under Fowler-Nordheim stress," J. Appl. Phys., **71**, 725 (1992).

29. J. F. Zhang, "Traps," Wiley Encyclopedia of Electrical and Electronics Eng., **22**, 540 (1999).

30. K. F. Schuegraf and C. M. Hu, "Metal-Oxide-Semiconductor field-effect-transistor substrate current during Fowler-Nordheim tunneling stress and silicon dioxide reliability," J. Appl. Phys., **76**, 3695 (1994).

31. H. Satake and A. Toriumi, "Substrate hole current generation and oxide breakdown in Si MOSFETs under Fowler-Nordheim electron tunneling injection," in Tech. Digest IEEE 39th Int. Electron Devices Meeting, Washington, p.337, USA (1993).

32. E. Cartier, J. C. Tsang, M. V. Fischetti and D. A. Buchanan, "Light emission during direct and Fowler-Nordheim tunneling in ultra thin MOS tunnel junctions," Microelectronic Eng., **36**, 103 (1997).

33. M. Rasras, D. de Wolf, G. Groeseneken, B. Kaczer, R. Degraeve and H. E. Maes, "Photo-carrier generation as the origin of Fowler-Nordheim-induced substrate hole current in thin oxides," in Tech. Digest IEEE 45th Int. Electron Devices Meeting, Washington, p.465, USA (1999).

34. R. S. Scott, N. A. Dumin, T. W. Hughes, D. J. Dumin and B. T. Moore, "Properties of high voltage stress generated traps in thin silicon oxides," in Proc IEEE 33rd Int. Reliability Phys. Symp., Las Vegas, p.131, USA (1995).

35. E. M. Vogel, J. S. Suehle, M. D. Edelstein, B. Wang, Y. Chen and J. B. Bernstein, "Reliability of ultrathin silicon dioxide under combined substrate hot-electron and constant voltage tunneling stress," IEEE Trans. Elec. Dev., **47**, 1183 (2000).

36. H. Uchida and T. Ajioka, "Electron trap center generation due to hole trapping in SiO2 under Fowler-Nordheim tunneling stress," Appl. Phys. Lett., **51**, 433 (1987).

37. J. F. Zhang, S. Taylor and W. Eccleston, "A quantitative investigation of electron detrapping in SiO_2 under Fowler-Nordheim stress," J. Appl. Phys., **71**, 5989 (1992).

38. J. F. Zhang and W. Eccleston, "Positive bias temperature instability in MOSFET's," IEEE Trans. Elec. Dev., **45**, 116 (1998).

39. H. Ushizaka and Y. Sato, "The process dependence on positive bias temperature aging instability of p+(B) polysilicon-gate MOS devices," IEEE Trans. Electron Devices, vol. 40, p. 932 (1993).

40. F. Faggin, D. D. Forsythe, and T. Klein, "Room temperature instabilities observed on silicon gate devices," in Proc. IEEE IRPS, p. 35. (1970).

41. H. Nakayama, Y. Osada, and M. Shindo, "Room temperature instabilities of p-channel silicon-gate MOS transistors," J. Electrochem. Soc., vol. 125, p. 1302, (1978).

42. J. H. Sim, S.C. Song, P. D. Kirsch, C. D. Young, R. Choi, D. L. Kwong, B. H. Lee, and G. Bersuker, "Effects of ALD HfO_2 thickness on charge trapping and mobility," Microelectronic Eng., **80**, 218 (2005).

43. C. Z. Zhao, M. B. Zahid, J. F. Zhang, G. Groeseneken, R. Degraeve, and S. De Gendt, "Properties and dynamic behavior of electron traps in HfO_2/SiO_2 stacks," Microelectronic Eng., vol.80, pp.366-369 (2005).

44. C. Z. Zhao, J. F. Zhang, M. B. Zahid, B. Govoreanu, G. Groeseneken, and S. De Gendt, "Determination of capture cross sections for as-grown electron traps in $HfO_2/HfSiO$ stacks," J. Appl. Phys., **100**, Art. No.093716, 2006.

45. T. P. Ma, H. M. Bu, X. W. Wang, L. Y. Song, W. He, M. Wang, H.-H. Tseng, and P. J. Tobin, "Special reliability features for Hf-based high-κ gate dielectrics," *IEEE Trans. Device Mater. Rel.*, vol. 5, no. 1, pp. 36–44, Mar. 2005.

46. R. Degraeve, A. Kerber, P. Roussel, E. Cartier, T. Kauerauf, L. Pantisano, and G. Groeseneken, "Effect of bulk trap density on HfO_2 reliability and yield," in *IEDM Tech. Dig.*, 2003, pp. 935–938.

47. Z. Ji, J. F. Zhang, W. Zhang, G. Groeseneken L. Pantisano, S. De Gendt, M. M. Heyns, "An assessment of the mobility degradation induced by remote charge scattering," Applied Phys. Lett., vol. 95, No. 26, Art. No. 263502, 2009.

48. B. E. Deal, M. Skalar, A. S. Grove and E. H. Snow, "Characteristics of the surface-state charge (Qss) of thermally oxidized silicon," J. Electrochem. Soc., **114**, 266 (1967).

49. A. Goetzberger, A. D. Lopez and R. J. Strain, "On the formation of surface states during stress aging of thermal Si-SiO interfaces," J. Electrochem. Soc., **120**, 90 (1973).

50. 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., **48**, 2004 (1977).

51. G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng and D. L. Kwong, "Dynamic NBTI of p-MOS transistors and its impact on MSFET scaling," IEEE Elec. Dev. Lett., **23**, 734 (2002).

52. C. E. Blat, E. H. Nicollian, E. H. Poindexter, "Mechanism of negative-bias-temperature instability," J. Appl. Phys., **69**, 1712 (1991).

53. S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang, W. Y. Teo and L. Chen, "Linear relationship between H+-trapping reaction energy and defect generation: insight into nitrogen-enhanced negative bias temperature instability," Appl. Phys. Lett., **83**, 530 (2003).

54. X. J. Zhou, L. Tsetseris, S. N. Rashkeev, D. M. Fleetwood, R. D. Schrimpf, S. T. Pantelides, J. A. Felix, E. P. Gusev and C. D'Emic, "Negative bias-temperature instabilities in metal-oxide-silicon devices with SiO2 and SiOxNy/HfO2 gate dielectrics," Appl. Phys. Lett., **84**, 4394 (2004).

55. D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: road to cross in deep submicron silicon semiconductor manufacturing," J. Appl. Phys., **94**, 1 (2003).

56. S. Ogawa, M. Shimaya and N. Shiono, "Interface-trap generation at ultrathin SiO2 (4-6nm)-Si interface during negative-bias temperature aging," J. Appl. Phys., **77**, 1137 (1995).

57. S. Ogawa and N. Shiono, "Generalized diffusion-reaction model for the low-field charge-buildup instability at the Si-SiO$_2$ interface," Phys. Rev. B, **51**, 4218 (1995).

58. P. K. Hurley, M. A. Negara, T. Van Hemert, K. Cherkaoui, "The influence of oxide charge on carrier mobility in HfO2/TiN gate silicon MOSFETs," ECS Transactions, 19, 379 (2009).

59. M. H. Chang and J. F. Zhang, "On positive charges formed under negative bias temperature stresses," J. Appl. Phys., 101, Art. no. 024516, 2007.

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

61. J. F. Zhang, M. H. Chang, and G. Groeseneken,"Effects of measurement temperature on NBTI," IEEE Electron Dev. Lett., Vol.28, No.4, pp.298-300, 2007.

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 Dev., vol. 56, No.5, pp. 1086-1093, 2009.

63. H. C. Shin and C. M. Hu, "Thin gate oxide damage due to plasma processing," Semicond. Sci. Tech., **11**, 463 (1996).

64. T. Brozek and C. R. Viswanathan, "Increased hole trapping in gate oxides as latent damage from plasma charging," Semicond. Sci. Tech., **12**, 1551 (1997).

65. T. Brozek, L. H. Peng and C. R. Viswanathan, "Role of test stress levels in detection of process-induced latent charging damage in MOS transistors," in Proc 1st Int. Symp. on Plasma Process-Induced Damage, Santa Clara, p.81, USA (1996).

66. J. C. King and C. M. Hu, "Effect of low and high-temperature anneal on process induced damage of gate oxide," IEEE Elec. Dev. Lett., **15**, 475 (1994).

67. M. Creusen, H. C. Lee, S. Vanhaelemeersch and G. Groeseneken, "The effect of plasma damage and different annealing ambients on the generation of latent interface states," in Proc 3rd Int. Symp. on Plasma Process-Induced Damage, Honolulu, p.217, USA (1998).

68. J. F. Zhang, H. K. Sii, R. Degraeve, and G. Groeseneken, "Mechanism for the generation of interface state precursors," J. Appl. Phys., 87(6), pp.2967-2977 (2000)

69. M. Knoll, D. Braunig and W. R. Fahrner, "Generation of oxide charge and interface states by ionizing-radiation and by tunnel injection experiments," IEEE Trans. on Nucl. Sci., **29**, 1471 (1982).

70. P. M. Lenahan and P. V. Dressendorfer, "Effects of bias on radiation-induced paramagnetic defects at the silicon-silicon dioxide interface," Appl. Phys. Lett., **41**, 542 (1982).

71. D. M. Fleetwood, M. J. Johnson, T. L. Meisenheimer, P. S. Winokur, W. L. Warren and S. C. Witczak, "1/f noise, hydrogen transport, and latent interface-trap buildup in irradiated MOS devices," IEEE Trans. Nucl. Sci., **44**, 1810 (1997).

72. R. K. Freitag, D. B. Brown and C. M. Dozier, "Evidence for two types of radiation-induced trapped positive charge," IEEE Trans. Nucl. Sci., **41**, 1828 (1994).

73. D. M. Fleetwood, "Effects of hydrogen transport and reactions on microelectronics radiation response and reliability," Microelectronics Reliability, **42**, 523 (2002).

74. F. B. McLean, "A framework for understanding radiation-induced interface states in SiO MOS structures," IEEE Trans. Nucl. Sci., **27**, 1651 (1980).

75. D. L.Griscom, "Diffusion of radiolytic molecular-hyderogen as a mechanism for the post-irradiation buildup of interface states in SiO2-on-si structures," J. Appl. Phys., **58**, 2524 (1985).

76. N. S. Saks, C. M. Dozier and D. B. Brown, "Time-dependence of interface trap formation in MOSFETs following pulsed irradiation," IEEE Trans. Nucl. Sci., **35**, 1168 (1988).

77. D. B. Brown and N. S. Saks, "Time-dependence of radiation-induced interface trap formation in metal-oxide-semiconductor devices as a function of oxide thickness and applied field," J. Appl. Phys., **70**, 3734 (1991).

78. R. E. Stahlbush, A. H. Edwards, D. L. Griscom, and B. J. Mrstik, "Post-irradiation cracking of H2 and formation of interface states in irradiated metal-oxide-semiconductor field-effect-transistors," J. Appl. Phys., **73**, 658-667 (1993).

79. A. Stesmans, "Dissociation kinetics of hydrogen-passivated Pb defects at the (111)Si/SiO$_2$ interface," Phys. Rev. B, **61**, 8393 (2000).

80. G. J. Gerardi, E. H. Poindexter, P. J. Caplan, and N. M. Johnson, "Interface traps and Pb centers in oxided (100) silicon-wafers," Appl. Phys. Lett., **49**, 348 (1986).

81. W. D. Zhang, J. F. Zhang, M. J. Uren, G. Groeseneken, R. Degraeve, M. Lalor, and D. Burton, "On the interface states generated under different stress conditions," Appl. Phys. Lett., vol.79, no.19, 3092-3094 (2001).

82. W. D. Zhang, J. F. Zhang, M. Lalor, D. Burton, G. Groeseneken, and R. Degraeve, "Two types of neutral electron traps generated in the gate silicon dioxide," IEEE Trans. Electron Dev., vol.49, no.11, 1868-1875 (2002).

83. W. D. Zhang, J. F. Zhang, C. Z. Zhao, M. H. Chang, G. Groeseneken and R. Degraeve, "Electrical signature of the defect associated with gate oxide breakdown," IEEE Electron Dev. Lett., vol.27, no.5, pp.393-395 (2006).

84. M. H. Chang, J. F. Zhang, and W. D. Zhang, "Assessment of capture cross sections and effective density of electron traps generated in silicon dioxides," IEEE Trans. Electron Dev., vol.53, no.6, pp.1347-1354 (2006).

85. E. Cartier and J. H. Stathis, "Atomic hydrogen-induced degradation of the Si/SiO₂ structure," Microelectronic Eng., **28**, 3 (1995).

86. D. A. Buchanan, A. D. Marwick, D. J. DiMaria and L. Dori, "Hot-electron-induced hydrogen redistribution and defect generation in metal-oxide-semiconductor capacitors," J. Appl. Phys., **76**, 3595 (1994).

87. E. Cartier, D. A. Buchanan and J. G. Dunn, "Atomic hydrogen-induced interface degradation of reoxidized-nitrided silicon dioxide on silicon," Appl. Phys. Lett., **64**, 901 (1994).

88. D. J. DiMaria, "Defect generation under substrate-hot-electron injection into ultrathin silicon dioxide layers," J. Appl. Phys., **86**, 2100 (1999).

89. S. Ogawa, N. Shiono and M. Shimaya, "Neutral electron trap generation in SiO₂ by hot holes," Appl. Phys. Lett., **56**, 1329 (1990).

90. J. D. Bude, B. E. Weir and P. J. Silverman, "Explanation of stress-induced damage in thin oxides," in Tech. Digest IEEE 44th Int. Electron Devices Meeting, San Francisco, p.179, USA (1998).

91. M. H. Chang and J. F. Zhang, "On the role of hydrogen in hole-induced electron trap creation," Semiconductor Sci. and Technology, vol.19, pp.1333-1338 (2004).

92. S. K. Lai, "Interface trap generation in silicon dioxide when electrons are captured by trapped holes," J. Appl. Phys., **54**, 2540 (1983).

93. R. Bellens, E. De Schrijver, G. Van den bosch, G. Groeseneken, P. Heremans and H. E. Maes, "On the hot-carrier-induced poststress interface-trap generation in n-channel MOS-transistors," IEEE Trans. Elec. Dev., **41**, 413 (1994).

94. I.S.Al-kofahi, J.F.Zhang and G.Groeseneken, "Continuing degradation of the SiO2/Si interface after hot hole stress," J. Appl. Phys., **81**, 2686-2692 (1997).

95. J. F. Zhang, I. S. Al-kofahi, and G. Groeseneken, "Behaviour of hot hole stressed SiO_2/Si interface at elevated temperature," J. Appl. Phys., 83(2), 843-850 (1998).

96. C. Z. Zhao, J. F. Zhang, G. Groeseneken, R. Degraeve, J. N. Ellis, and C. D. Beech, "Interface state generation post-hole injection," J. Appl. Phys., vol.90, no.1, pp.328-336 (2001).

97. J. F. Zhang, H. K. Sii, G. Groeseneken, and R. Degraeve, "Degradation of oxides and oxynitrides under hot hole stress," IEEE Trans. Electron Devices, 47(2), pp.378-386 (2000).

98. J. F. Zhang, C. Z. Zhao, G. Groeseneken, and R. Degraeve, "Analysis of the kinetics for interface state generation following hole injection," J. Appl. Phys., vol.93, no.10, pp.6107-6116 (2003).

99. S. Zafar, A. Callegari, E. Gusev, and M. V. Fischetti, "Charge trapping related threshold voltage instabilities in high permittivity gate dielectric stacks," J. Appl. Phys., **93**, 9298 (2003).

100. J. F. Zhang, C. Z. Zhao, M. B. Zahid, G. Groeseneken, R. Degraeve, and S. De Gendt, "An assessment of the location of as-grown electron traps in $HfO_2/HiSiO$ stacks," IEEE Electron Dev. Lett., vol.27, no.10, pp.817-820 (2006).

101. J. F. Zhang, "Defects and Instabilities in Hf-dielectric/SiON stacks," Proc. Of The 16th Biennial Conference on Insulating Films on Semiconductors (INFOS), pp.1883-1887, Cambridge University, UK, 2009.

102. N. Kimizuka, T. Yamamoto, T. Mogami, K. Yamaguchi, K. Imai and T. Horiuchi, "The impact of bias temperature instability for direct-tunneling ultra-thin gate oxide on MOSFET scaling," in Tech. Digest Symp. on VLSI Technology, Kyoto, p73, Japan (1999).

103. N. Kimizuka, K. Yamaguchi, K. Imai, T. Iizuka, C. T. Liu, R. C. Keller and T. Horiuchi, in "NBTI enhancement by nitrogen incorporation into ultrathin gate oxide for 0.10 m gate CMOS generation," in Tech. Digest Symp. on VLSI Technology, Honolulu, p.92, USA (2000).

104. S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang and L. Chen, "Nitrogen-enhanced negative bias temperature instability: an insight by experiment and first-principle calculations," Appl. Phys. Lett., **82**, 1881 (2003).

2 | Experimental Facilities and Techniques

2.1   Introduction

The standard system for wafer level tests typically involves a computer, a semiconductor parameter analyser, a pulse generator, and a probe station with micro-positioners, as illustrated in Fig.2.1. At the start of this project, such a system was already available in the laboratory and was used for the quasi-DC tests in Chapter 3.

This test system is fully controlled by the computer through an IEEE 488 port. The control program was specifically designed for this system and was written in turbo C language. The probe station allows connecting the test devices at wafer level with the electronic equipment. The pulse generator is used to generate pulses required by the charge pumping measurements. The parametric measurement mainframe (Agilent E5270A) has 4 medium power source and measurement units (SMU), which can be used to supply/measure the gate, drain, well and substrate voltage/current.

It typically takes 20~150ms for measuring one point by the industrial standard parameter analyzers, such as Agilent E5270A, 4145, 4155/4156 [1] and Keithley 4200. To obtain a transfer characteristic, tens of points are needed and the total measurement time will be in the order of seconds. This kind of measurement is often referred to as quasi-static measurement or quasi-DC measurement. Two assumptions

in the quasi-DC measurement are that the measurement itself will not cause any degradation for the device under test (DUT) and there will be no recovery of degradation during the measurement. For MOSFETs with relatively thick oxides (e.g. >5nm), this is widely accepted. For thinner oxides, however, both recovery and degradation can occur during the quasi-DC measurement and it becomes essential to increase the measurement speed by using the pulse measurements [2-9].
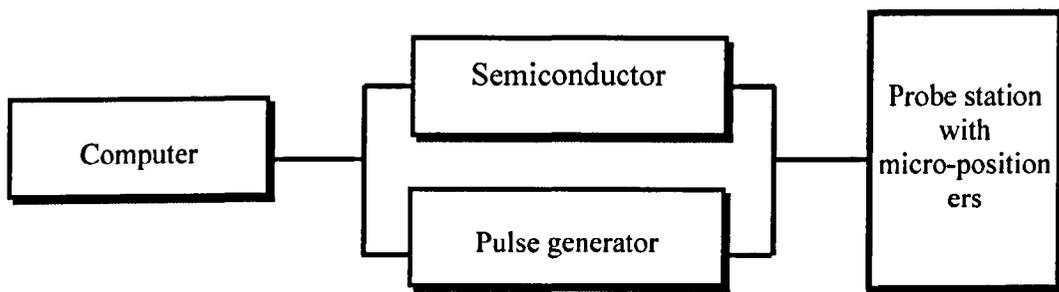
Figure 2.1. A schematic diagram of the testing system.

In this chapter, section 2.2 will describe the stress-measure-stress (SMS) methodology widely used for characterizing the reliability of MOSFETs. Section 2.3 will cover the threshold voltage extraction from the quasi-DC transfer characteristic. Section 2.4 will describe the pulse technique existed in the laboratory and shows that the highest measurement speed achievable is 5µs. A major effort has been made in this project to improve the measurement speed and section 2.5 will show how the noise is minimized. Section 2.6 will demonstrate that the re-designed and improved system can achieve a measurement speed of 200 ns for Id-Vg measurement and 800ns for the capacitance-voltage measurement. Finally, section 2.7 addresses how to correct the system delay and gate leakage and how to separate the displacement current from the drift current.

2.2   Stress-measure-stress methodology

The common reliability test will start with characterising the properties of fresh devices, such as the threshold voltage, interface states, and charges in the gate dielectric [2,3,5-8]. These values will be used as the reference for measuring the shift of parameters during the stress due to degradation. To produce a measurable amount of degradation during a practical test time, the biases applied are typically considerably higher than that used in the real operation. The role of temperature is more complex. In some types of tests, such as negative bias temperature stress [2,3,10-12] and time dependent dielectric breakdown (TDDB) [13-15], it is often increased to accelerate the degradation. In hot carrier stress, however, room temperature is typically used. This is because an increase of temperature can lower the carrier mobility and the energy of hot carriers, resulting in lower degradation, even though the degradation process itself is thermally activated [16].

To monitor the extent of degradation, the stress is often interrupted at a preset time, so that the typical measurements, such as transfer characteristics and charge pumping, can be performed. The bias applied during the measurement is generally lower than the stress bias and its polarity can also be different from that of the stress bias. The implicit assumption is that this interruption will not disturb the degradation, but this is not true for thin dielectrics [2-9]. The common practice is to stress and measure the degradation at the same temperature, to minimize the interruption. Changing temperature of a thermal chuck is a slower process than the electrical measurements.

After recording the degradation, the stress bias is reapplied and the degradation continues. The degradation generally will not happen in a constant speed and the defect generation rate will slow down for longer stress. A power law kinetics is typically observed for the bias temperature stresses [17,18] and hot carrier stresses [16], with a power factor in the range of 0.1-0.5 for the former [17,18] and around 0.2-0.4 for the latter [16]. As a result, a logarithmic time scale is often preferred to a linear time scale for setting the measurement points for probing the degradation [2,3,16-18].

2.3 Threshold voltage extraction from quasi-DC Id-Vg

Threshold voltage, Vth, is one of the most important parameters for MOSFETs and its shift is often monitored during the electrical stresses. Traditionally, the threshold voltage is measured from the quasi-DC transfer characteristic by using extrapolation [2,3] and this will be described in this section. Recently, it is widely reported that the quasi-DC measurement is too slow and degradation can recover substantially during the quasi-DC measurement [2-9]. To minimize the measurement time, several new techniques have been developed recently, including on-the-fly (OTF) method [10,19], pulse technique [2,3,20], and the pulse Vg characterisation [3,20]. The pulse techniques will be given in section 2.4.

By definition, a MOSFET will be off when Vg<Vth and switched on when Vg>Vth and the drain current Id is zero at Vth. In reality, a MOSFET will not be turned on suddenly and the Id will not be zero in the subthreshold region. This means that the evaluation of Vth will involve a certain degree of approximation.

The classical method for determining Vth is to extrapolating the quasi-DC Id-Vg measured at a sufficiently low drain bias (e.g. Vd<0.1V) [2,3,20]. One example is given in Fig.2.2. After measuring Id-Vg, transconductance, gm=dId/dVg, is evaluated from it. Fig.2.2 shows that gm has a peak at a certain Vg(gm=max). The Id-Vg is then linearly extrapolated from Vg(gm=max) to Id=0 to find the Vth, as illustrated in Fig.2.2. After negative bias stresses, the Id-Vg was measured again to evaluate the degraded Vth and its shift, ΔVth.
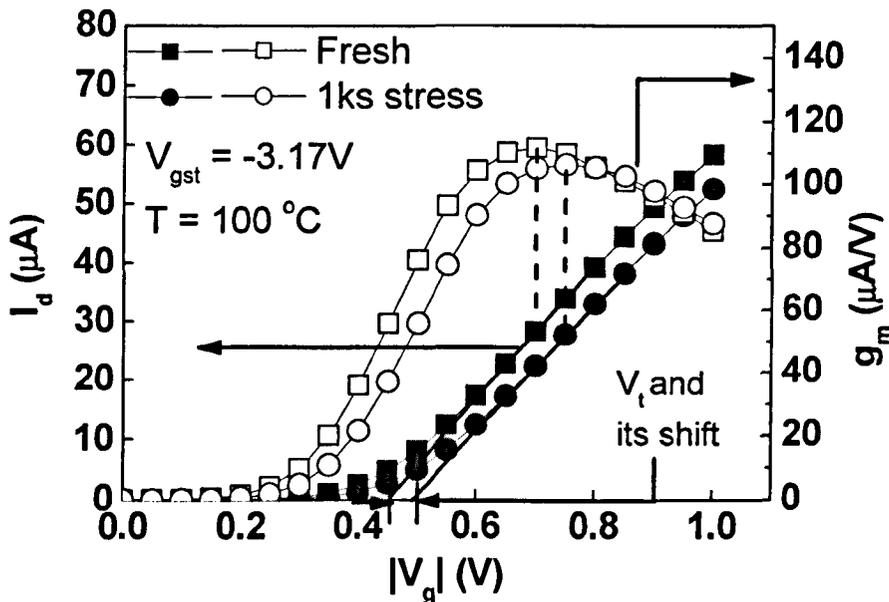
Fig.2.2 Typical experimental results obtained by using conventional technique. I_d-V_g curves are measured before and after stressing the device for a certain period of time. After the stress, the I_d-V_g curve is shifted towards higher |V_g|. Threshold voltage is extracted by using maximum gm extrapolation method. The trans-conductance is first calculated by differentiating the I_d-V_g curve and threshold voltage is defined as the gate-voltage axis intercept of the linear extrapolation on the I_d-V_g curve at maximum trans-conductance. [20]

2.4   The existing pulse technique and its shortcoming

The principle of pulse technique for evaluating Vth is essentially the same as that for the quasi-DC Id-Vg technique. As illustrated by Fig.2.3, the stress is interrupted and

Id-Vg is recorded again and the extrapolation is made from the maximum transconductance point. The difference is the measurement speed. A pulse generator is used here to replace the quasi-DC parameter analyzer as the voltage source for the gate bias.
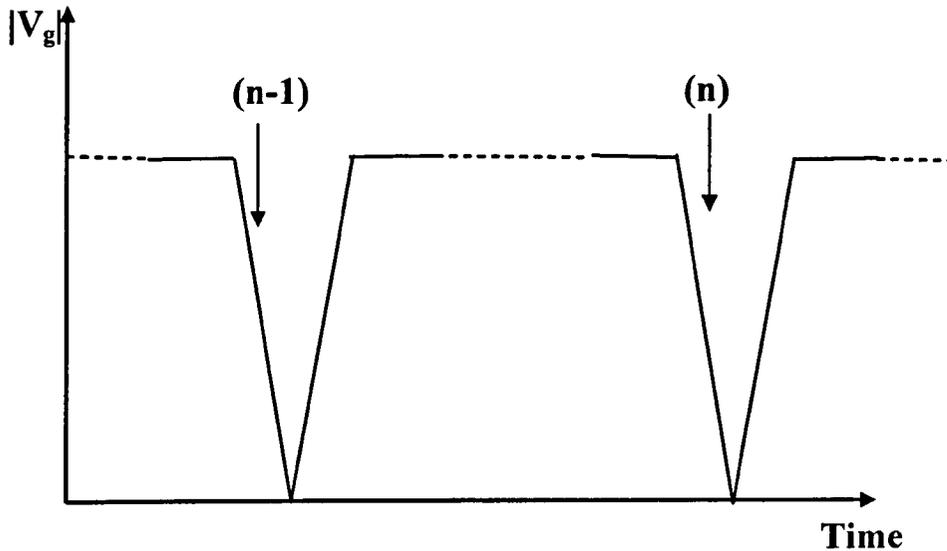
Fig.2.3 Gate voltage waveform used by the UFP-ex and UFP-V_g techniques. '(n-1)' and 'n' indicate two neighboring measurement points. [3]

The circuit used to implement the pulse technique is given in Fig.2.4. The drain to source bias, V_d, was pinned at a low level (e.g. -25 mV), supplied by a battery through a voltage divider and an operational amplifier. The gate bias was provided by an Agilent 81101A pulse generator and the typical width/rise/ fall time is 10/5/5 µs, respectively. Both V_g and the V_{out} in Fig.2.4 were recorded. The V_{out} is converted to I_d by,

$$I_d=(V_{out}-V_d)/R,$$

where the feedback resistor, R, typically has a value of 1 kΩ.
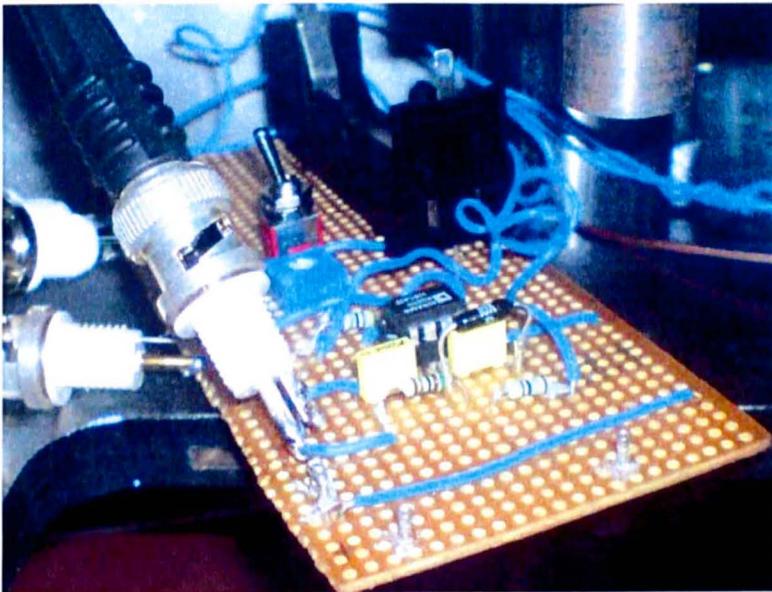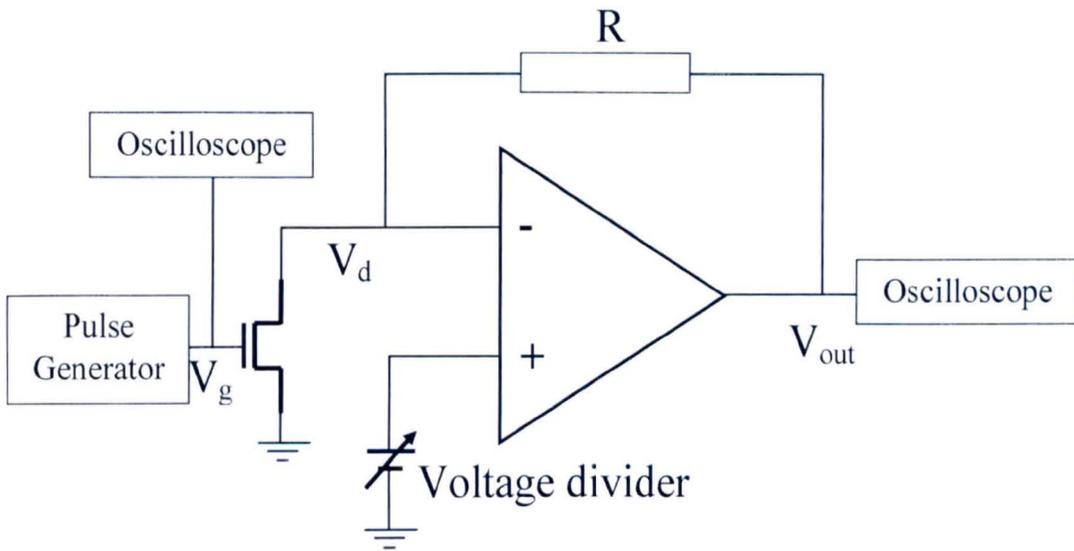
Fig.2.4 The circuit and its photograph used by the existing pulse technique. The drain current is converted from V_{out} through $I_d=(V_{out}-V_d)/R$ and V_d is set through a voltage divider and an operational amplifier. [3]

Some typical Id-Vg characteristics measured by using the existing pulse technique are given in Fig.2.5. It can be seen that the results are good for a measurement time (i.e. the pulse edge time) of 5µs or longer, but the Id was seriously distorted for

shorter time. Fig.2.6 show that the typical split capacitance-voltage (C-V) characteristics measured by the existing pulse technique. It uses the relationship I=C*(dv/dt). Distortion is again observed when the measurement time is below 5μs for both the gate-to-channel capacitance, Cgc (Fig.2.6(a)), and the gate-to-body capacitance, Cgb (Fig.2.6(b)) . A lot of efforts have been made in this project to improve the measurement speed and the details are given in the next section.
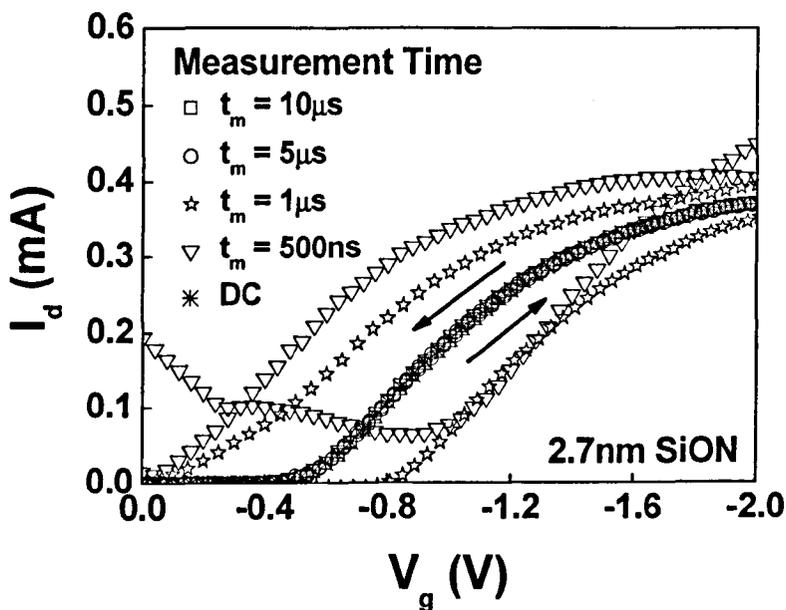
Fig.2.5 Typical Id-Vg measured by the existing pulse technique. The measurement speed is limited at 5μs.
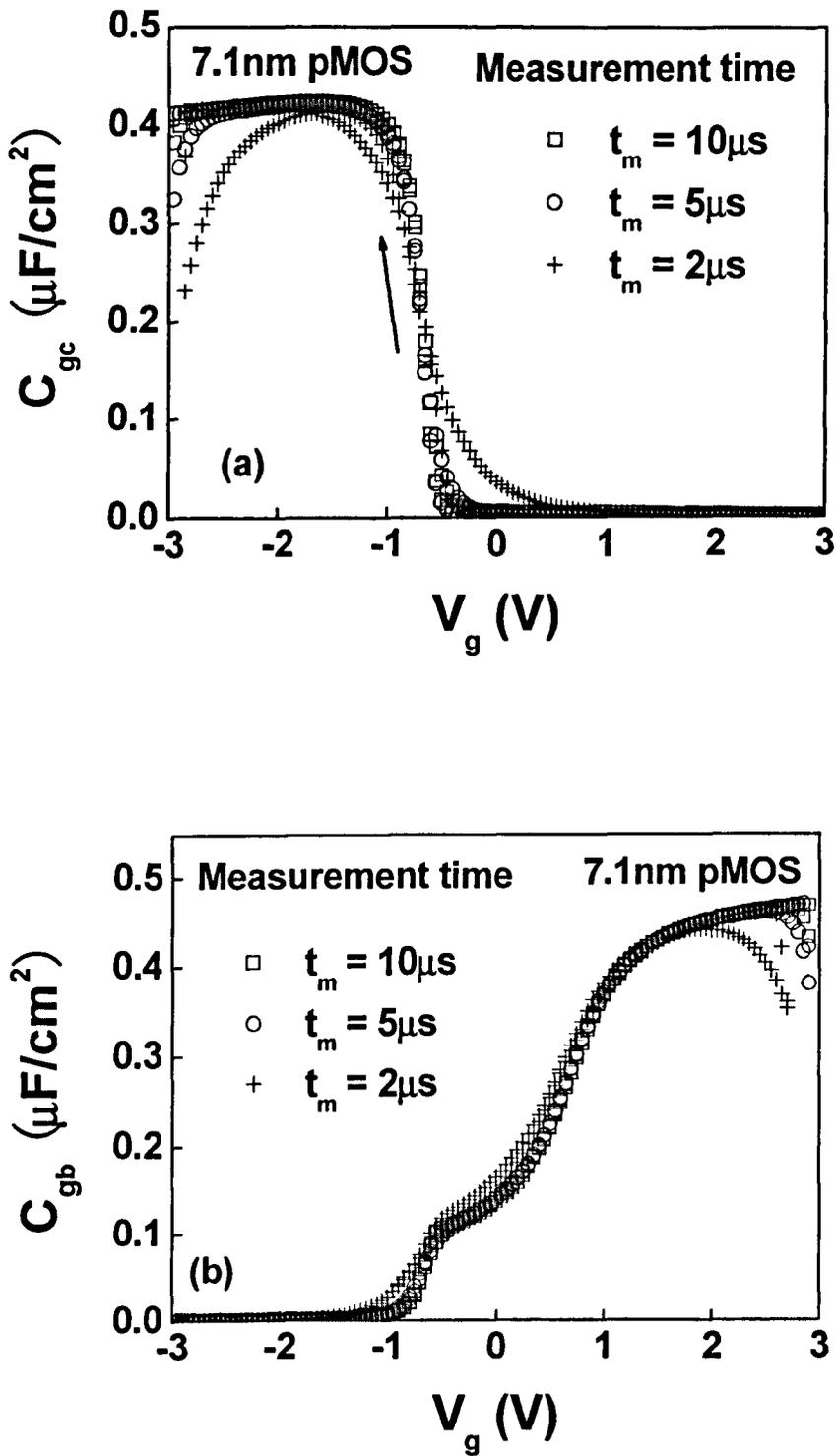
Fig.2.6 Typical split C-V measured by the existing pulse technique. (a) The gate-to-channel capacitance and (b) The gate-to-body capacitance. Distortion can be observed when measurement time is less than 5μs.

Fig.2.6 shows there is a sudden change in C_{gb} under the same voltage level when C_{gc} also present a sudden change. The reason is under this voltage level, conduction channel is created, charges in the conduction channel response the gate bias instead of the charges in the substrate. This change is rapid which will cause the sudden change in C_{gb}.

2.5   Design and implementation of a high speed pulse technique

2.5.1   High performance amplifier

For the existing pulse technique, the operational amplifier used is the AD844, which has a relatively low bandwidth of 80MHz. This limits the measurement speed. It is of great importance to reduce the measurement time further. For pulse I-V measurement, it is reported that in order to measure the intrinsic Id-Vg, the measurement time has to be in the order of 100ns. In addition, the accuracy of the pulse C-V technique used in Chapter 5 relies on the ramp rate, dV/dt, since the displacement current is proportional to the ramp rate.

To improve the measurement speed, one wide bandwidth and high gain op-amp (TI OPA657, 1.6GHz) is selected to replace the AD844. It has high precision Analog-to-Digital converter with very low input noise voltage and only requires a quiescent current of 14mA. This makes it suitable for the present application, since the low level signals can be significantly amplified in a single gain stage with exceptional bandwidth.

The designed circuit diagram with the new op-amp is shown in Fig.2.7. Like the existing technique, it was first implemented by using a veroboard, as shown in Fig. 2.8.

The test results given in Fig.2.9 show that the noise level of the circuit in Fig.2.8 is too large to be acceptable. This is addressed in the next section.
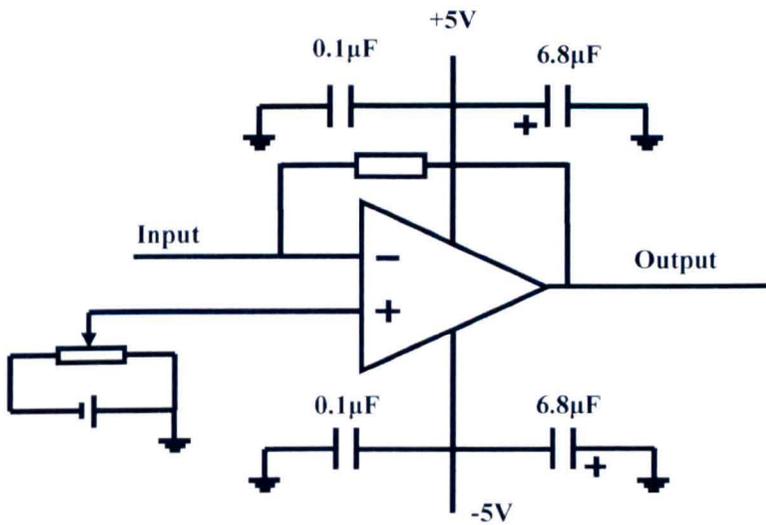
Fig.2.7 The new OPA657 amplifier circuit designed for the pulse technique.
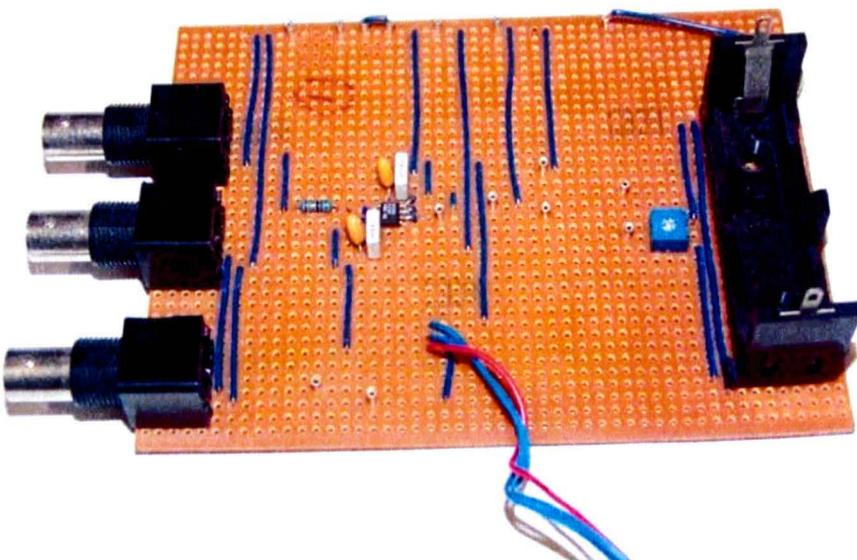
Fig.2.8 The photograph of the first OPA657 pulse Id-Vg circuit.
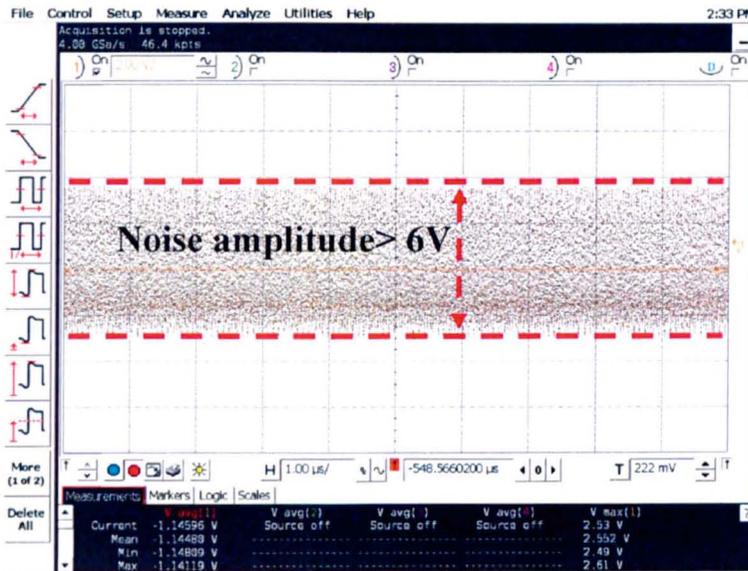
Fig.2.9 Output signal from the OPA657 pulse I-V circuit given in Fig.2.8

2.5.2 Noise reduction

The high noise level for the circuit given in Fig.2.9 suggests that the high performance op-amp requires a dedicated assembly and layout. The printed circuit board (PCB) was used and attention is paid to minimize the possible noise sources. After carefully studying RF circuit requirements, several noise sources were identified and addressed, including socketing, grounds, external components, power supply, and layout.

2.5.2.1 Socketing

As shown in Fig.2.8, the standard PCB socket was soldered onto veroboard to allow connection of input signal cable. This type of socketing was not designed to shield noise and interference for a high-speed circuit and must be replaced.

The pin configuration of the new OPA657 amplifier is SO-8, designed for surface

mount. In order to solder the amplifier onto veroboard, extra wires were used to extend the length of the 8 pins. Such socketing however, is not recommended for a high-speed amplifier like the OPA657. The increased lead length and pin-to-pin capacitance introduced by the socket can create a troublesome parasitic network which makes it almost impossible to achieve a smooth and stable frequency response. Best results were obtained by soldering the OPA657 onto the board directly. [21]

2.5.2.2 Ground plane

A ground plane in PCB assembly is a layer of copper that appears to most signals as an infinite ground potential. This helps reducing noises and ensures that different signals have the same reference potential. It also makes the circuit design easier, allowing the designer to reach ground without having to run multiple tracks. The component needing ground is routed directly to the ground plane on another layer. [21]

The improved circuit was assembled on a double layer PCB board, shown in Fig. 2.10. The parasitic capacitance to ground is minimized for all I/O pins. Parasitic capacitance on the output and inverting input pins can cause instability. On the noninverting input, it can react with the source impedance and unintentionally limit frequency band. To reduce the unwanted capacitance, a window around the signal I/O pins should be opened in the ground and power planes. Apart from these windows, ground and power planes should be unbroken elsewhere on the board.
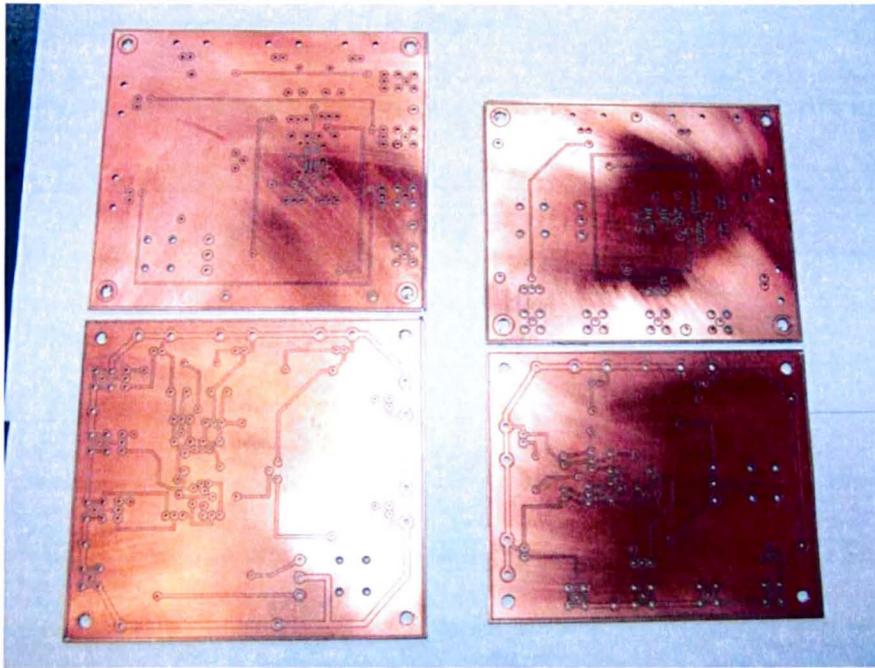
Fig. 2.10 The double layer PCB board for the improved circuit.

2.5.2.3   Selection and placement of external components

Careful selection and placement of external components will preserve the high frequency performance of the OPA657. Resistors should be of a very low reactance type. Surface-mount resistors work best and allow a tighter overall layout. Metal film and carbon composition axially leaded resistors provide good high-frequency performance. Their lead lengths are made as short as possible. Since the output pin and inverting input pin are the most sensitive ones to parasitic capacitance, the feedback resistor are positioned as close as possible to the output pin. Other network components, such as noninverting input terminal resistors, should also be placed close to the package. Even with a low parasitic capacitance shunting the external resistors, excessively high resistor values can create significant time constants that degrade performance. Good axial metal film or surface-mount resistor have approximately 0.2 pF in shunt with the resistor. [21]

2.5.2.4 Power supply conditioning and layout

To eliminate the noise from the input pin, the circuit was tested first under no input signal condition, which means there is no MOS device has been probed during the circuit test. This allows examining other noise sources, such as power supply and components layout.

For the power supply section, based on the RF circuit noise requirement, the power supply connections should always be decoupled with capacitors. In Fig.2.7, the small decoupling capacitors on the supply pins are effective at high frequency and the large decoupling capacitors are effective at low frequency. The distance from the power-supply pins to the high-frequency $0.1\mu F$ decoupling capacitor is minimized to within 0.25". The ground and power plane should not be in close proximity to the signal I/O pins.[21] The voltage regulators are also essential to stabilize the voltage supply.

The second circuit was designed after considering the possible noise sources mentioned above. The schematic circuit diagram is given in Fig.2.11 and a photograph of the PCB circuit is given in Fig.2.12. Fig.2.13 shows that the noise level was reduced to 600 mV. This is still too high to be acceptable and further improvement is essential.
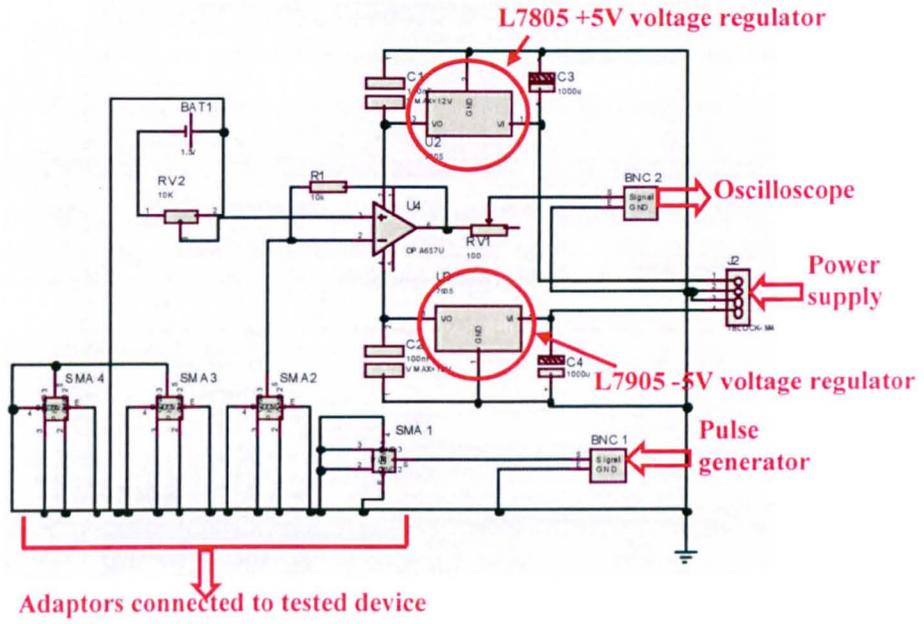
Fig.2.11 The schematic improved circuit
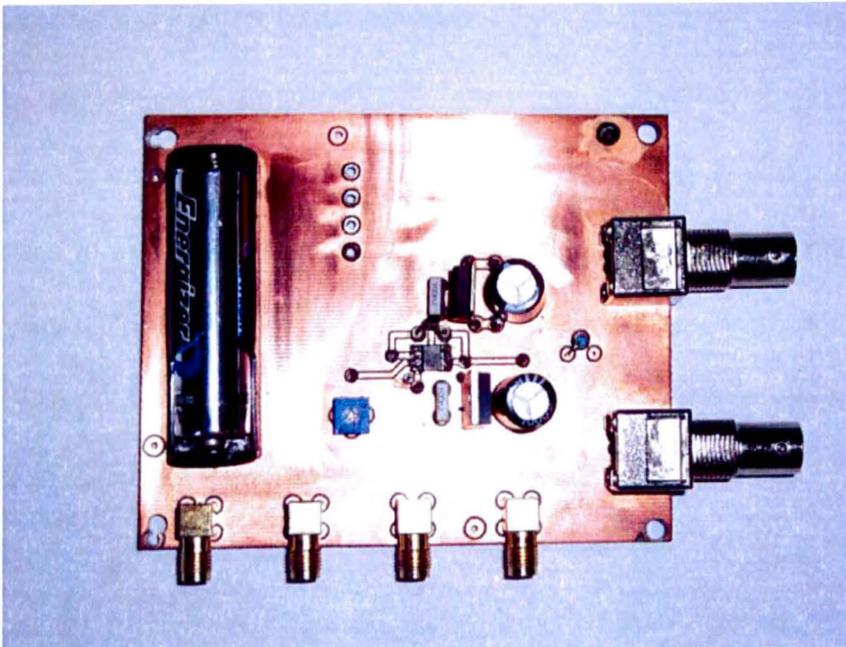
Fig.2.12 A photograph of the improved circuit.

Fig.2.13. After the improvement, system noises were reduced to 600mV.

2.5.2.5 Cabling and further improvements

During the experiment, SSMC-to-SMA cables are needed to connect the 657 amplifier circuit and the micropositioners of the probe station, as illustrated in Fig. 2.14. In order to avoid impedance mismatch, all system components must possess 50Ω impedance. The length of SSMC-to-SMA cables is minimized. Moreover, BNC cables between circuits and oscilloscope must have the same length in order to synchronise the multiple output channels.

Variable resistor replacement: The variable resistor in Fig.2.7 introduces noise into the system due to its turning mechanism. It is replaced by a basic voltage divider consisting of two fixed value resistors.

Location of decoupling capacitors: The amplifier requires +/-5V power supply, each needs two decoupling capacitors and one voltage regulator. The small decoupling capacitors were placed as close as possible to the output pin. Test results show that the circuit had a better performance when the voltage regulator and the large decoupling capacitor were placed close to each other.
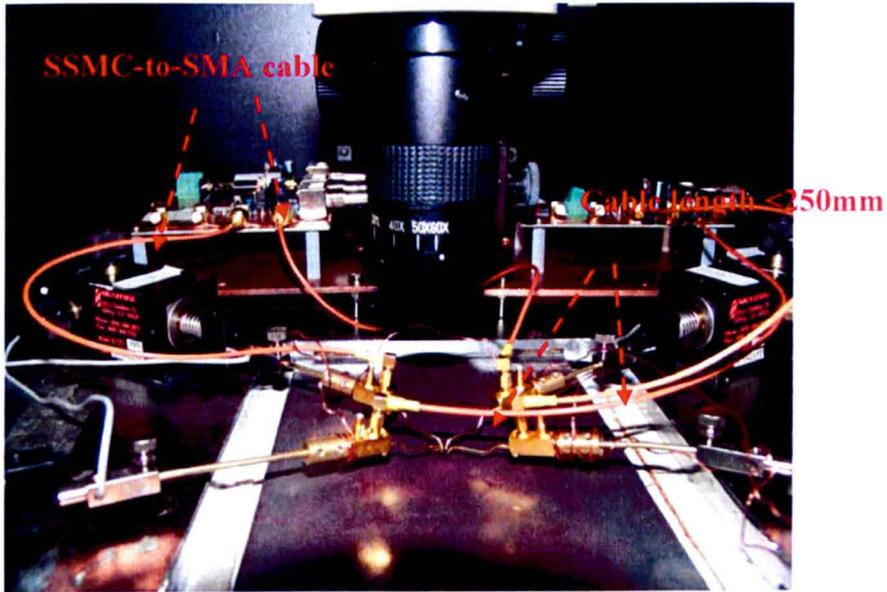
Fig.2.14   SSMC-to-SMA BNC cable are using to connect testing device and circuit.

The circuit was re-designed again and is shown in Fig.2.15.
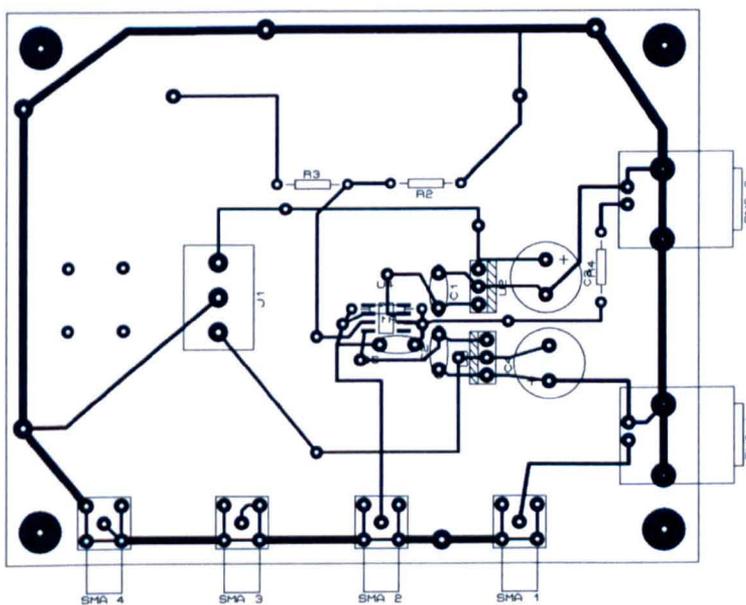
Fig.2.15 The new circuit after 2nd re-design to include the further improvement.

The performance of the circuit after the second re-design is shown in Fig.2.16. When compared with Fig.2.13, the noise in the output signal is reduced significantly to about 30mV. There is, however, an oscillation in the signal and efforts were made to remove it.
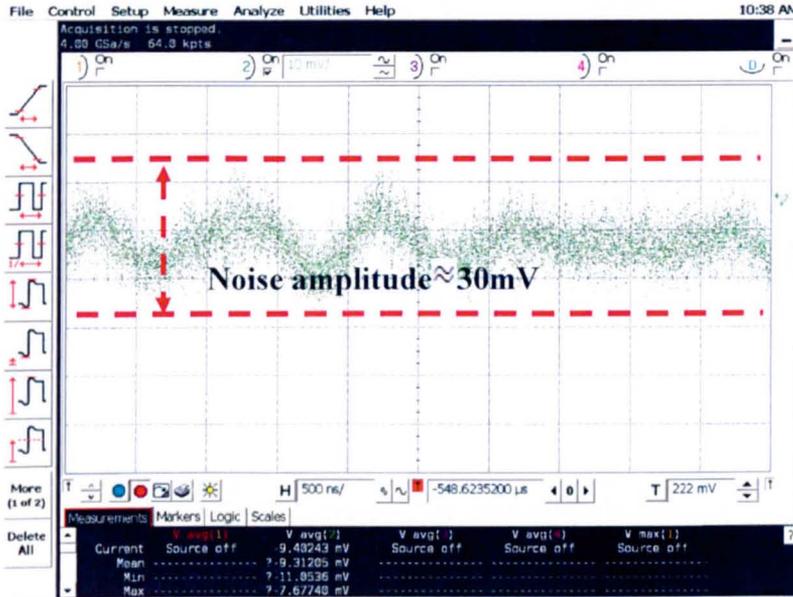
Fig.2.16 The noise of output signal for the circuit in Fig.2.15.

2.5.2.6 Shielding and the final circuit

In order to connect to other equipments within the testing system, right angle metal SMA and BNC adaptors were assembled on the PCB board. Even with the out-layer of the adaptor leads soldered on the PCB board and connected to the ground, the metal surface still act like antenna and act as a source of noise. To achieve better shielding performance, these adaptors were wrapped with copper wire and connected into circuit GND.

After minimizing all possible noise sources and implementing the improvements, Fig. 2.17 shows that the system noise was reduced to within 5mV, which is acceptable for pulse measurements. The final circuit is given in Fig.2.18.
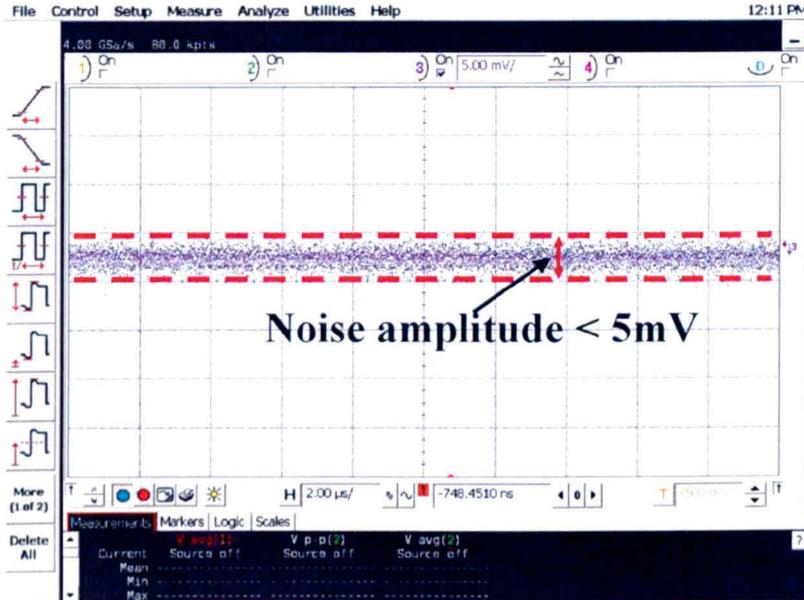
Fig.2.17 The noise of output signal for the circuit in Fig.2.18.
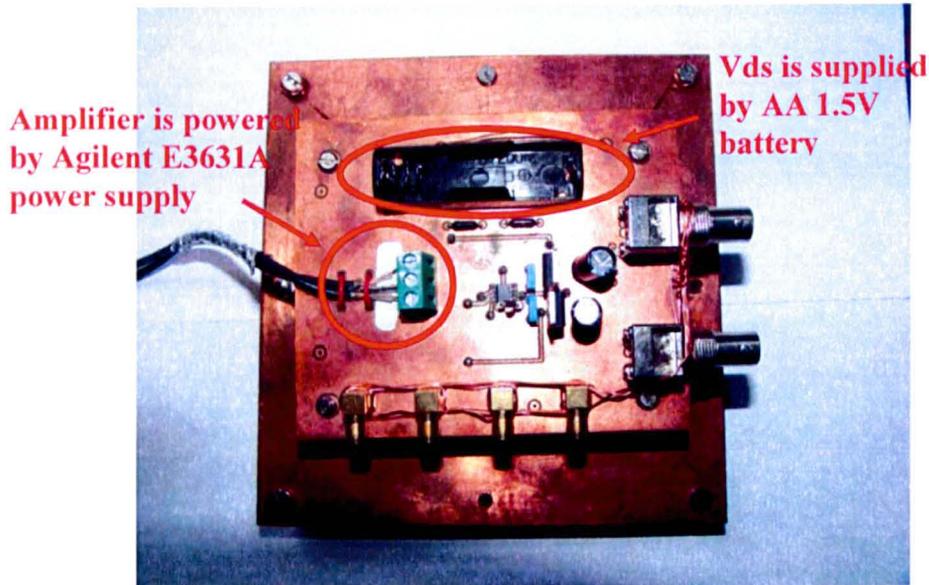
Fig.2.18 The final circuit.

2.6   Improvements in measurement speed

By using the circuit given in Fig.2.18, Fig.2.19 shows that the measurement speed

has been improved from the previous 5µs to 200 ns for Id-Vg and 800 ns for C-V.
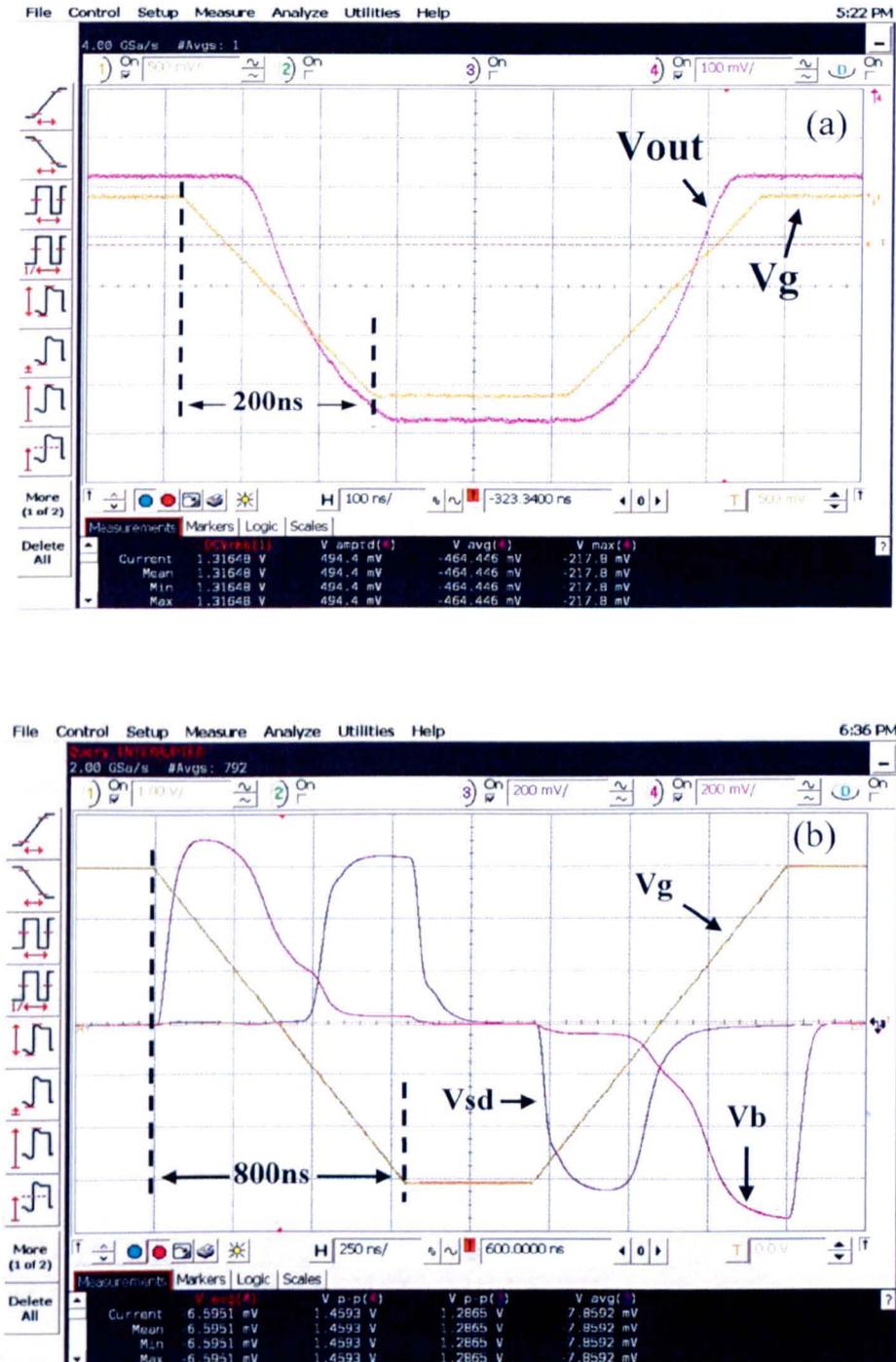
Fig.2.19 The screen shot of pulse measurement of Id-Vg (a) and C-V (b).

2.7 Corrections and calibration

2.7.1 System delay

Fig.2.20 gives the pulse C-V test system configuration.
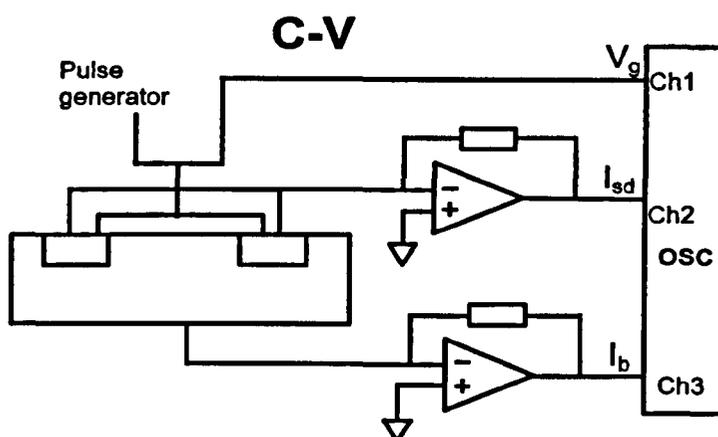
Fig.2.20 Pulse C-V test system configuration

As shown in the figure above, a pulse is applied to gate and captured by the oscilloscope directly. At the same time, currents Isd and Ib flowing out of the device will be converted into voltage by this home-made circuit and captured by the oscilloscope. The gate-to-channel Cgc-V measurements with a pulse edge time, tm, of $10\mu s$ and $1\mu s$ are given in Figs.2.21(a)&(b) respectively. Though good agreement has been observed between switching on and off edges when tm is $10\mu s$, a parallel shift appears when tm reduces to $1\mu s$. This is caused by the delay of the op-amp circuit. The gate voltage is directly fed into the oscilloscope. The signals corresponding to Isd and Ib, however, are converted through the op-amp and were delayed. The effect of system delay on the results is illustrated by Fig.2.22. It shifts the output signal towards negative Vg by ΔV, when Vg was swept towards negative, which turned on a pMOSFET. Similarly, it shifts the output signal towards positive

Vg, when Vg was swept towards positive by ΔV, which turned off a pMOSFET. Therefore, the difference between off-to-on and on-to-off edge will be twice of ΔV. The same explanation can also be applied to the gate-to-body Cb-V measurement given in Figs.2.23(a)&(b).
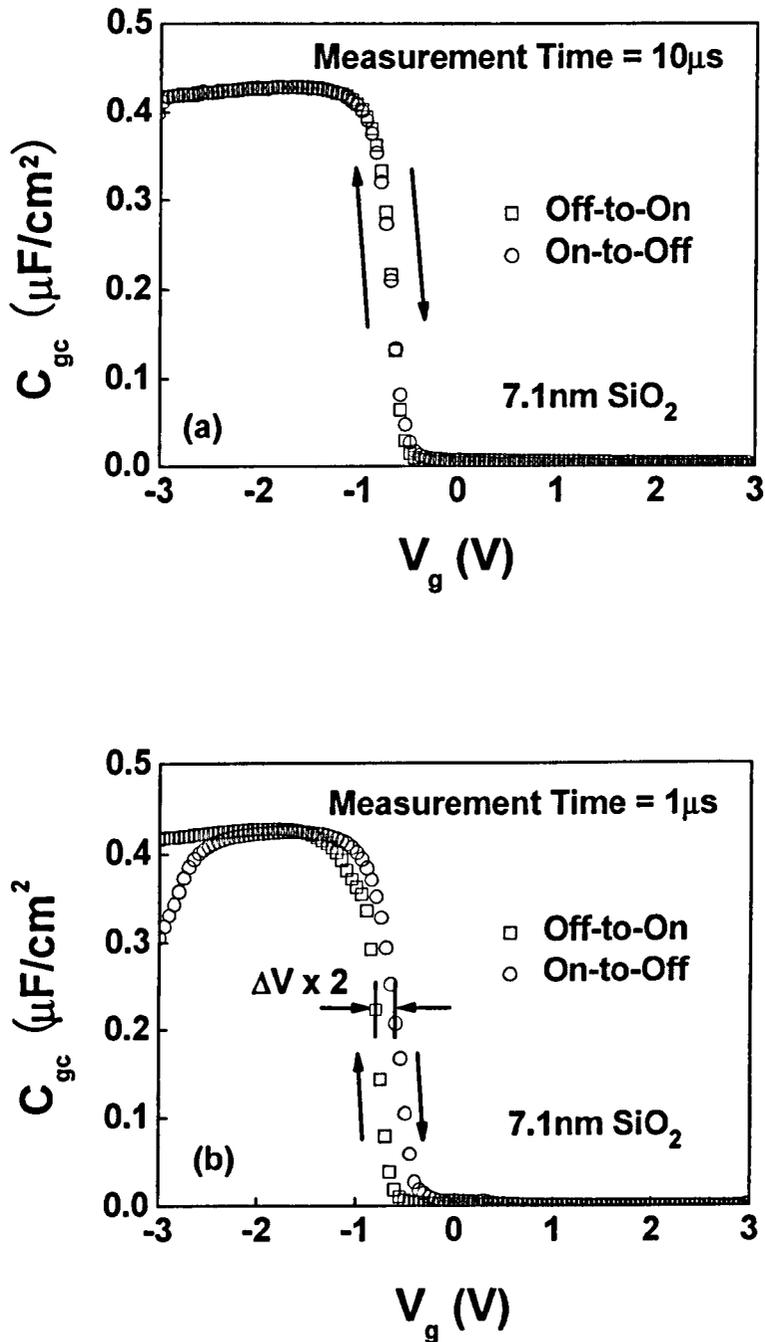
Fig.2.21 Cgc-V results at different measurement speed. (a) When measurement speed is 10μs, the system delay is negligible. (b) When measurement time is 1μs, a clear shift occurs.
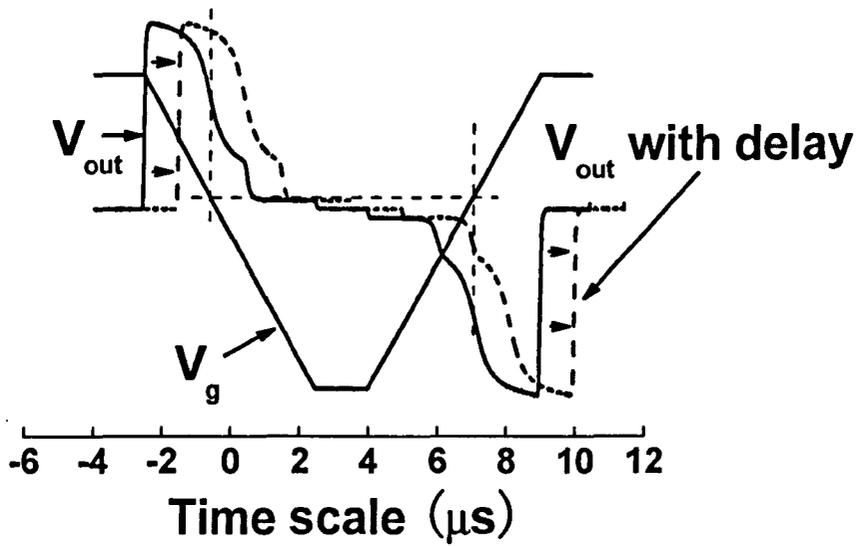
Fig.2.22 An illustration of the system delay on the results. The system delay induces a shift in the time scale for the output signal relative to Vg.
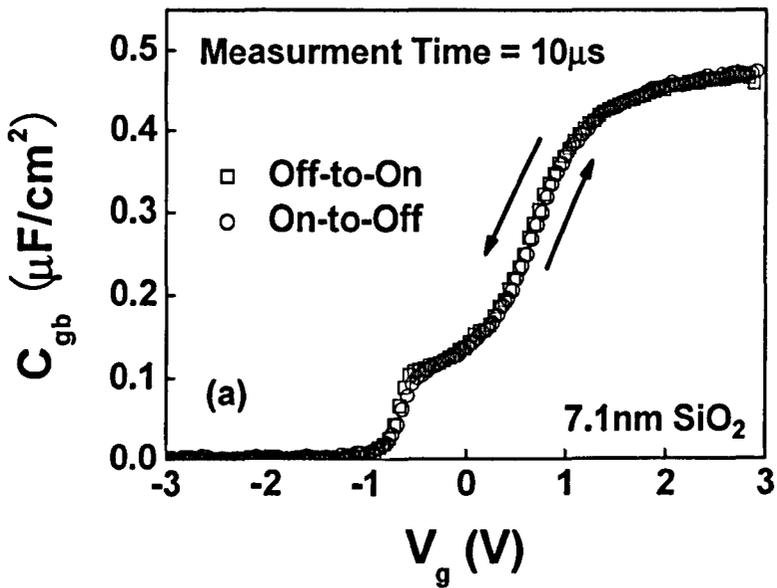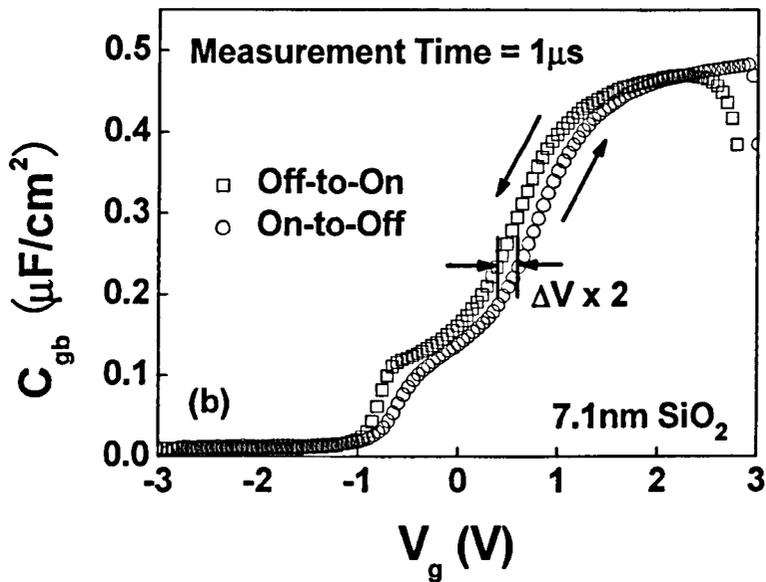
Fig.2.23 Cgb-V results at different measurement speed. (a) When measurement speed is 10μs, the system delay is negligible. (b) When measurement time is 1μs, a clear shift occurs.

This delay time, ΔT, is determined by the op-amp and it is a constant and independent of measurement conditions and devices. It can be obtained from $\Delta V/[dV/dt]$, where, ΔV is the delay induced shift, as illustrated in Fig.2.23(b), and dV/dt is its ramp rate. For a constant delay time, ΔV is proportional to the ramp rate, as shown by Fig.2.24. The system delay can be determined from the slope of the line in Fig.2.24 and it equals to 34.9 ns for the present circuit.
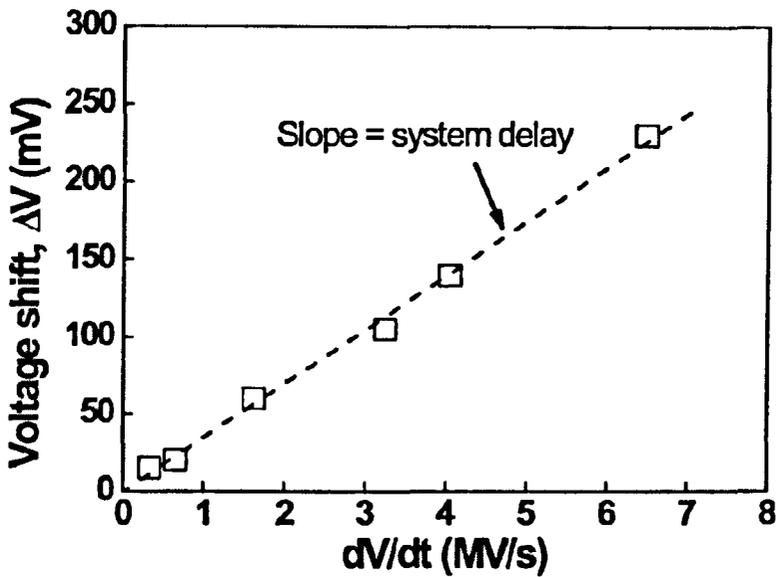
Fig.2.24 The relationship between ramp rate and the voltage shift.

After taking this 34.9ns system delay time into account, the C-V results measured at different speed (i.e. dV/dt) is corrected. Some examples are given in Fig.2.25(a) for Cgb-V and in Fig.2.25(b) for Cgc-V. It can be seen that the C-V from the two pulse edges merges and the agreement for different edge times, tm, is good for tm as low as 800ns.
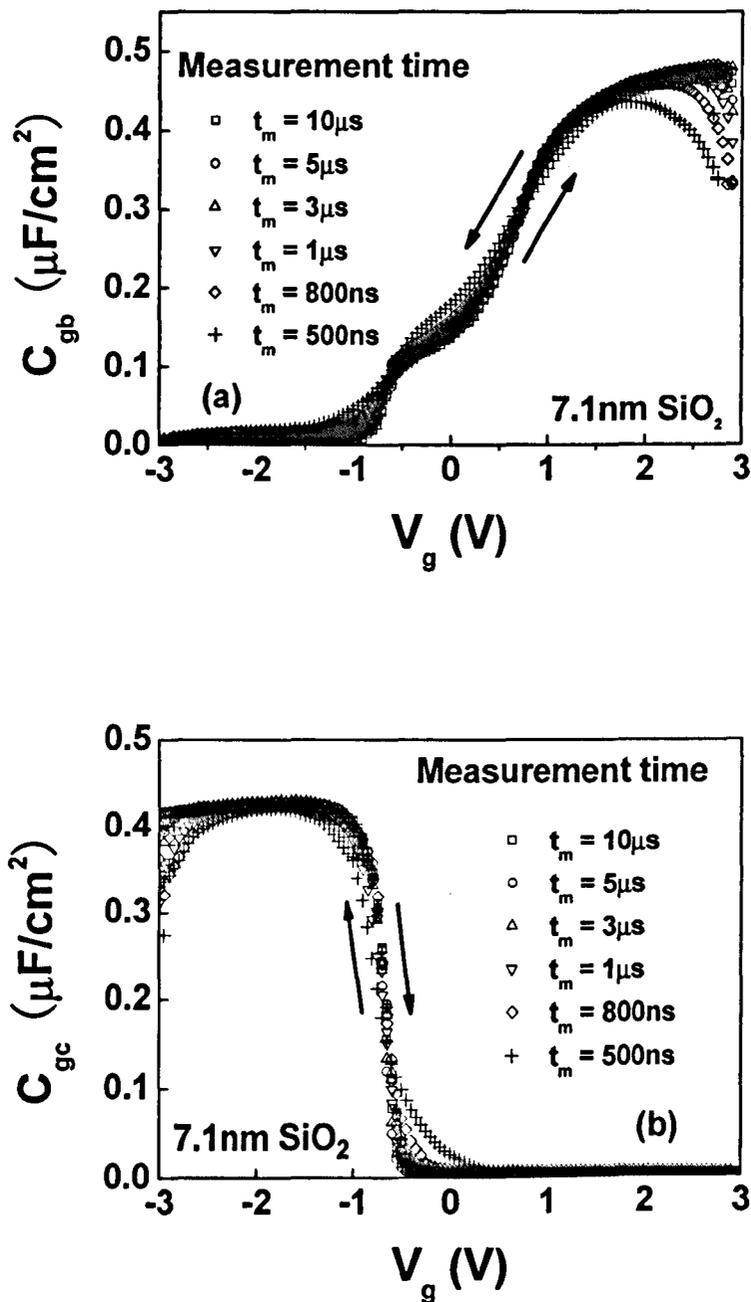
Fig.2.25 After correcting the system delay, off-to-on and on-to-off edge agrees well for a measurement time down to 800ns both Cgb (a) and Cgc (b).

Fig.2.25 shows both C_{gb} and C_{gc} data agrees well when measurement time is slower than 800ns, and distorted when measurement time is less than 800ns. System parasitic components are responsible for the distortion, since they cannot response to the rapid changing gate voltage.

2.7.2   Leakage current correction

For modern MOSFETs, the thickness of their gate dielectrics has shrunk to the nano-meter range. As a consequence, leakage current increases exponentially. When applying the pulse technique to a 1.85nm dielectrics, the measured Cgc and Cgb are shown in Fig.2.27. At high |Vg| region, curves for the off-to-on sweep start to deviate from those for the on-to-off sweep. This deviation is largest for the longest edge time and can be reduced only partially even at the shortest edge time. To correct for the gate leakage, the average value for the two edges can be used and the details will be presented in Chapter 5. Fig.2.28 shows that results after such a correction.
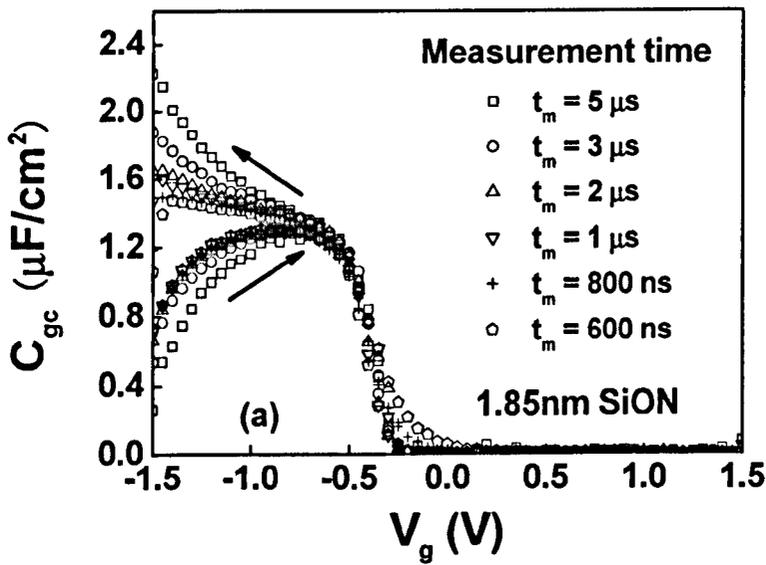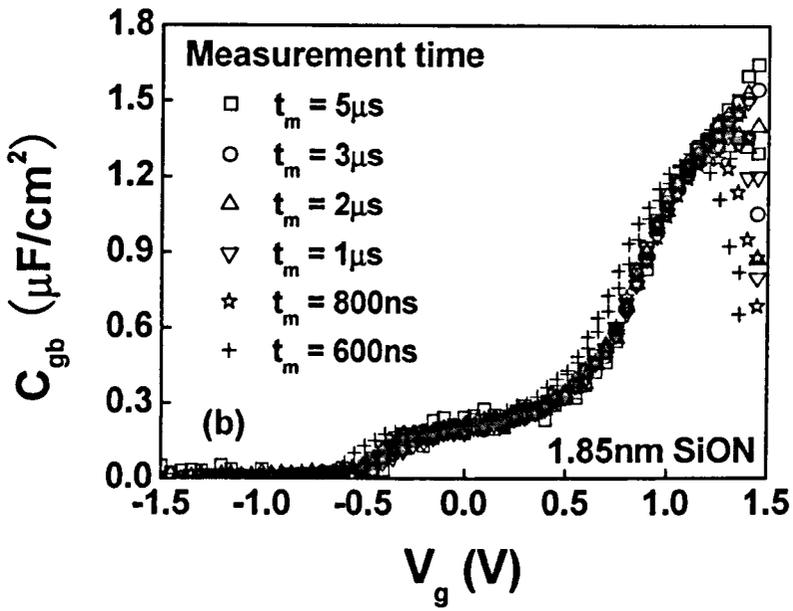
Fig.2.27 Impact of gate leakage current on the measured Cgc (a) & Cgb (b). The gate leakage causes a separation for the two C-Vs corresponding to the two pulse edges.
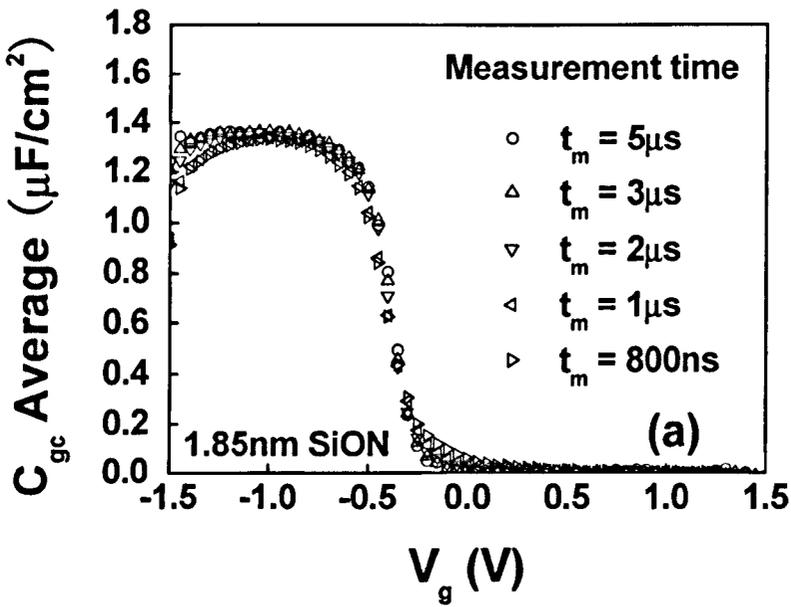
Fig.2.28 Results after correcting the gate leakage current for the data given in Fig.2.27.

2.7.3 Displacement current correction

In section 2.7.1, the effect of system delay on C-V has been addressed and corrected. For the I-V measurement, the same system delay has to be corrected. Even after the system delay is corrected, Fig.2.29 shows that the pulse I-V still shows clear difference between on-to-off and off-to-on edges. This is because at high measurement speed, the displacement currents can contribute to the measured drain current considerably. As it is shown in Fig.2.30, on one hand, when the device is switched from off to on, the direction of the displacement current Igd is opposite to the drift channel current. On the other hand, when the device is switched from on to off, the directions of these two current become the same. This leads to the hysterisis in Fig.2.29. The shorter the edge time is, the larger the displacement current and the hysterisis become.
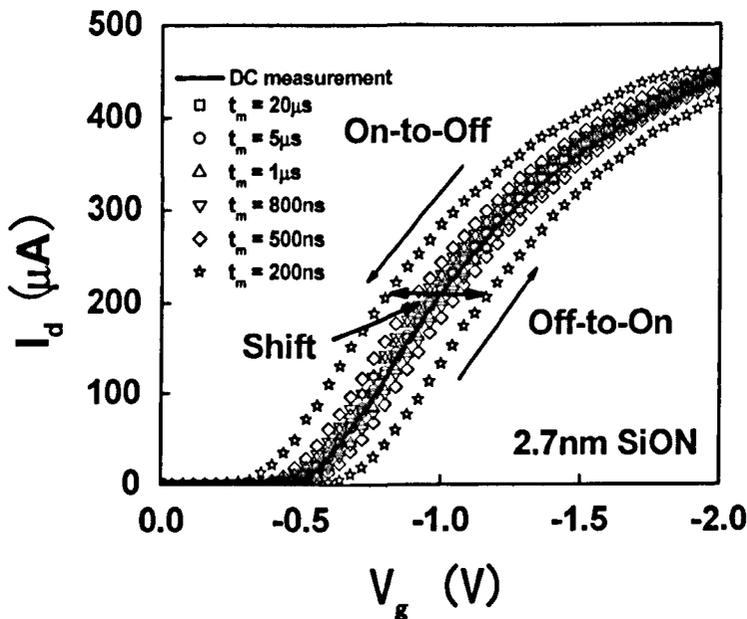
Fig.2.29 After the system delay correction, the pulse I-Vs still have hysterisis between the on-to-off and off-to-on edges.
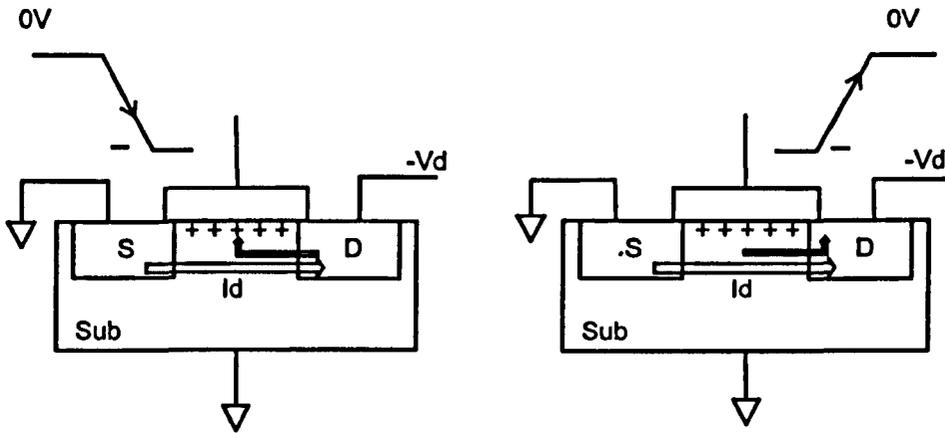
Fig.2.30 A comparison of the directions of drift current, Id, with the displacement currents for the two edges of the gate pulse.

Because the displacement current direction is opposite for the two pulse edges, it can be corrected by using the average current as the drift current, namely,

$$I_d = \left[I_{(on-to-off)} + I_{(off-to-on)} \right] / 2$$

After the correction, Fig.2.31 shows that Id become insensitive to the measurement time.
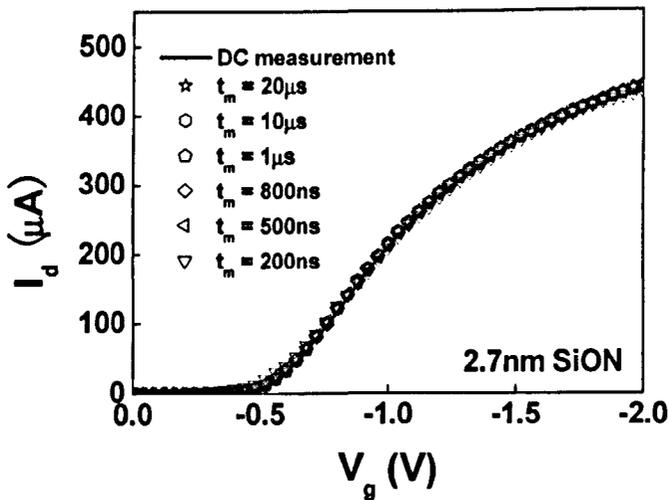
Fig.2.31 After the displacement current is corrected, pulse I-Vs are insensitive to the measurement time.

References

1. Operation Manual: Model 4145B Semiconductor Parameter Analyzer. Hewlett Packard Company, 1989.

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

3. 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 Dev., vol. 56, No.5, pp. 1086-1093, 2009.

4. J. F. Zhang, C. Z. Zhao, M. B. Zahid, G. Groeseneken, R. Degraeve, and S. De Gendt, "An assessment of the location of as-grown electron traps in $HfO_2/HiSiO$ stacks," IEEE Electron Dev. Lett., vol.27, no.10, pp.817-820 (2006).

5. J. F. Zhang, "Defects and Instabilities in Hf-dielectric/SiON stacks," Proc. Of The 16th Biennial Conference on Insulating Films on Semiconductors (INFOS), pp.1883-1887, Cambridge University, UK, 2009.

6. C. Z. Zhao, M. B. Zahid, J. F. Zhang, G. Groeseneken, R. Degraeve, and S. De Gendt, "Properties and dynamic behavior of electron traps in HfO_2/SiO_2 stacks," Microelectronic Eng., vol.80, pp.366-369 (2005).

7. C. Z. Zhao, J. F. Zhang, M. B. Zahid, B. Govoreanu, G. Groeseneken, and S. De Gendt, "Determination of capture cross sections for as-grown electron traps in $HfO_2/HfSiO$ stacks," J. Appl. Phys., **100**, Art. No.093716, 2006.

8. T. P. Ma, H. M. Bu, X. W. Wang, L. Y. Song, W. He, M. Wang, H.-H. Tseng, and P. J. Tobin, "Special reliability features for Hf-based high-κ gate dielectrics," *IEEE Trans. Device Mater. Rel.*, vol. 5, no. 1, pp. 36–44, Mar. 2005.

9. Z. Ji, J. F. Zhang, W. Zhang, G. Groeseneken L. Pantisano, S. De Gendt, M. M. Heyns, "An assessment of the mobility degradation induced by remote charge scattering," Applied Phys. Lett., vol. 95, No. 26, Art. No. 263502, 2009.

10. M. H. Chang and J. F. Zhang, "On positive charges formed under negative bias temperature stresses," J. Appl. Phys., 101, Art. no. 024516, 2007.

11. J. F. Zhang, M. H. Chang, and G. Groeseneken, "Effects of measurement temperature on NBTI," IEEE Electron Dev. Lett., Vol.28, No.4, pp.298-300, 2007.

12. P. K. Hurley, M. A. Negara, T. Van Hemert, K. Cherkaoui, "The influence of oxide charge on carrier mobility in HfO2/TiN gate silicon MOSFETs," ECS Transactions, 19, 379 (2009).

13. R. Degraeve, G. Groeseneken, R. Bellens, J. L. Ogier, M. Depas, P. J. Roussel and H. E. Maes, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," IEEE Trans. Elec. Dev., **45**, 904 (1998).

14. B. Kaczer, R. Degraeve, N. Pangon, T. Nigam and G. Groeseneken, "Investigation of temperature acceleration of thin oxide time-to-breakdown," Microelectronic Eng., **48**, 47 (1999).

15. B. Kaczer, R. Degraeve, N. Pangon and G. Groeseneken, "The influence of elevated temperature on degradation and lifetime prediction of thin silicon-dioxide films," IEEE Trans. Elec. Dev., **47**, 1514 (2000).

16. C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-electron-induced MOSFET degradation --- Model, Monitor, and Improvement," IEEE J. Solid-State Cir., **SC-20**, 295 (1985).

17. 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., **48**, 2004 (1977).

18. J. F. Zhang, C. Z. Zhao, M. H. Chang, M. B. Zahid, A. R. Peaker, S. Hall, G. Groeseneken, L. Pantisano, S. De Gendt, and M.Heyns, "Impact of different defects on the kinetics of negative bias temperature instability of hafnium stacks," *Applied Physics Letters*, **92**, Art.no. 013502 (2008).

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

20. Z. Ji, "Characterization of negative bias temperature instability and lifetime prediction for pMOSFETs," Ph.D thesis, Liverpool John Moores University (2010).

21. TI OPA657 amplifier data sheet

3 | Dominant layer for NBTI positive charges in Hf-based gate stack

3.1 Introduction

As the SiON thickness approaches 1 nm, the gate leakage current increases rapidly and the power consumption is becoming intolerable. One of the most important developments for CMOS technologies is replacing SiON by high-k dielectric/SiON stacks, which is often described as an open heart operation for MOSFETs. In 2007, Intel announced that they used Hf-based dielectric/SiON stacks with metal gate for their 45 nm CMOS technology [1]. This allows the industry to continue following the Moore's law longer than otherwise possible.

The development of high-k dielectric/SiON stack has encountered many difficulties, such as process integration, low carrier mobility, and high instabilities. As mentioned in section 1.6, the electron-trapping induced PBTI of nMOSFETs was an obstacle in the early stage of developing high-k dielectric layer [2-7]. The problem was successfully overcome by using Hf-dielectric layer thinner than 2 nm, where electron trapping becomes negligible [8,9].

As the PBTI of nMOSFETs is suppressed, attentions have been switched to the positive charging induced NBTI of pMOSFETs. Unfortunately, it was reported that NBTI can be significant even when the Hf-dielectric is thinner than 2nm [10-13]. Fig.3.1 compares the impact of Hf-dielectric layer thickness on PBTI of nMOSFETs

with that on NBTI of pMOSFETs. As the Hf-dielectric layer becomes thinner, the

reduction of PBTI-induced threshold voltage shift, ΔVth, is dramatic and becomes

negligible at 1.5nm, the reduction of PBTI-induced ΔVth is modest and it is still

important even for a 1nm Hf-dielectric layer [10]. The question is why NBTI has

such a different dependence on the thickness from PBTI and finding the answer is

the objective of this chapter.
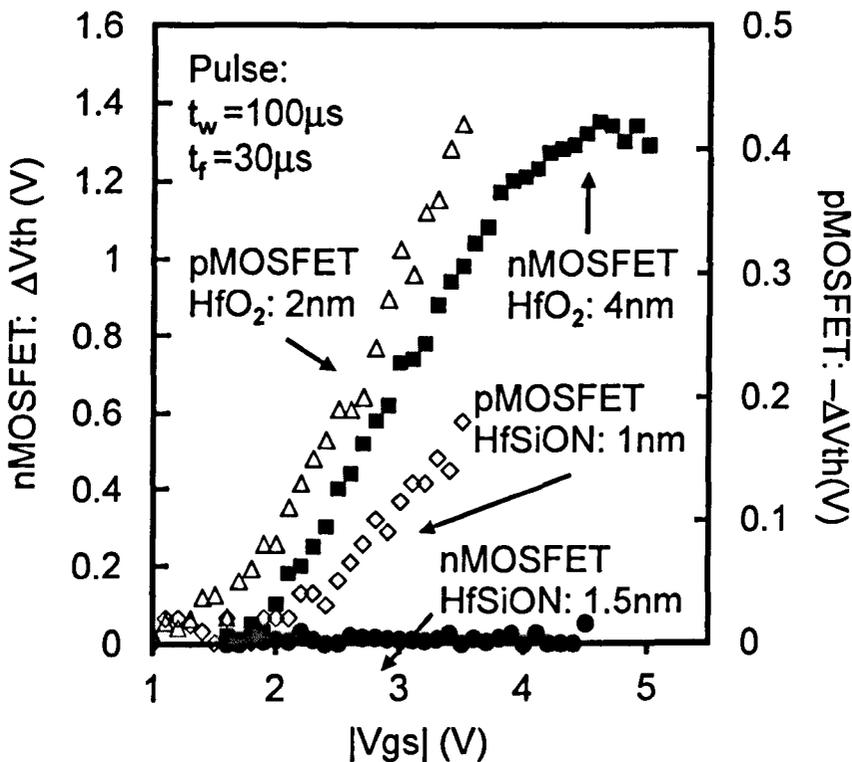
Fig.3.1 A comparison of the impact of Hf-dielectric layer thickness on PBTI of nMOSFETs and NBTI of pMOSFETs. The filled symbols represent nMOSFETs. The instability is negligible in the nMOSFETs with a sub-2nm Hf-silicate. The open symbols represent pMOSFETs and considerable instability occurs in a sub-2nm Hf-silicate. [10]

To answer the above question, one needs to know the difference in the spatial

distribution of NBTI positive charges from that of PBTI electron trapping. The

determination of charge spatial distribution is a challenging task. Traditionally,

photo-IV technique has been used to determine the spatial distribution of electron

traps for thick dielectrics [14], but this is a quasi-DC measurement and is not suitable

for probing the spatial distribution of electron trapping in nm dielectrics [2-9]. In this work, we will assess which dielectric layer in the stack dominates the NBTI positive charging by systematically varying the thickness of one layer and keeping the other layer thickness fixed. To prepare the background for this work, a brief review will be first carried out on the recent progress in understanding the positive charging in SiO_2. The reported similarity and difference in NBTI between Hf-dielectric/SiON stack and SiO_2 also will be summarized.

3.2 Recent progresses in understanding positive charges

3.2.1 Types of positive charges

Positive charging in SiO_2 has puzzled the international community for decades. Some positive charges behave so strangely that they are called by various names, such as anomalous positive charges [15], slow states [16] and border traps [17]. It has been reported recently that the anomalous behavior originates from the simultaneous presence of different types of positive charges in SiO_2: cyclic positive charge (CPC), anti-neutralization positive charge (ANPC) and as-grown hole trapping (AIIT) [18-21]. This framework for the classification of positive charges is illustrated by Fig.3.2.

In Fig.3.2(a), positive charges were first formed during a substrate hole injection (SHI) with an oxide field of $E_{ox}=-5MV/cm$. A Fowler-Nordheim electron injection was then used to neutralize them. This is followed by applying $Vg<0$ and $Vg>0$

(b) Vg>0: Neutralization (c) Vg<0: Charging
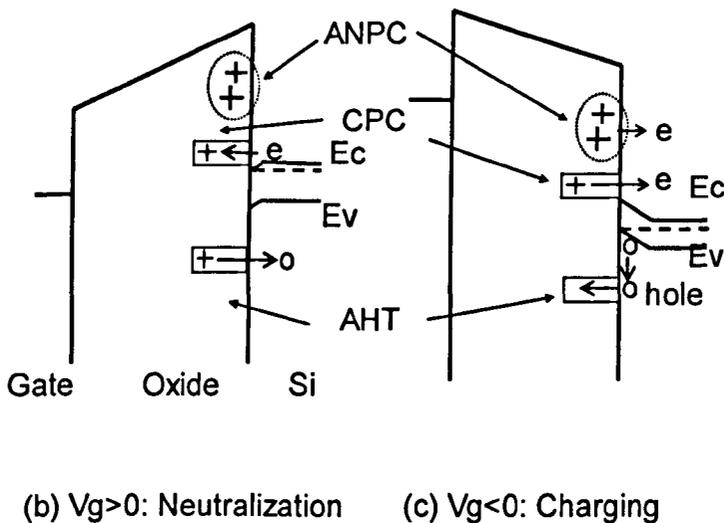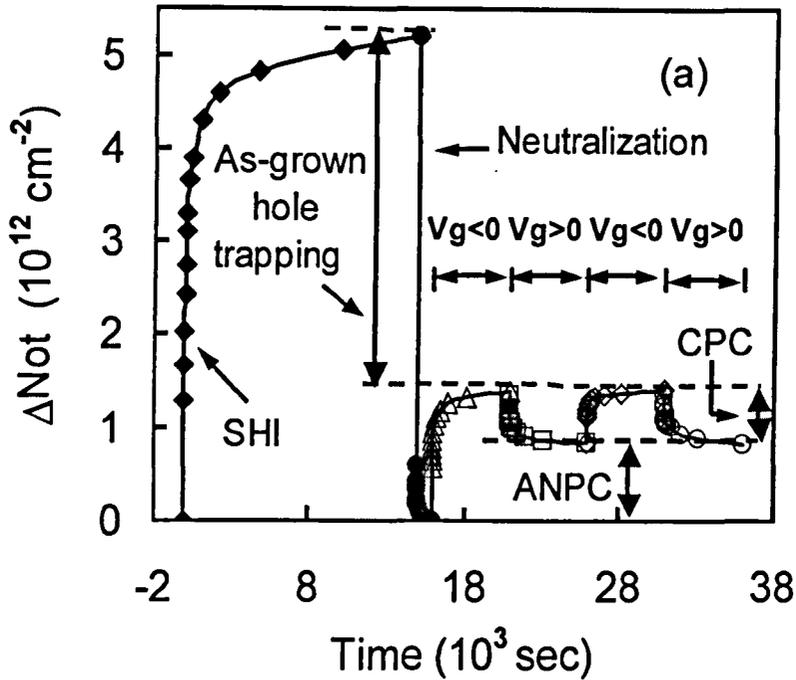
Fig.3.2 The presence of three different types of positive charges in a 5.5nm SiO₂. (a) shows the test sequence. Initially, positive charging built up during a substrate hole injection (SHI) under an oxide field of Eox=-5MV/cm, a n-well and p-substrate bias of 6V and 7V, respectively. The neutralization was under Eox=+8MV/cm. This was followed by applying Vg<0 (Eox=-5MV/cm) and Vg>0 (Eox=+5MV/cm) alternately, with all other terminals grounded. The neutralization (b) and charging (c) of cyclic positive charge (CPC) only involve electron tunneling at the same energy level. The anti-neutralization positive charge (ANPC) has an energy level above the conduction band edge of Si, making its neutralization difficult. For the same Eox=-5MV/cm, charging level without switching on SHI is well short of that with SHI, since as-grown hole traps (AHT) cannot be filled by holes near the top edge of Si valence band (c). [18]

Alternatively with an oxide field of Eox=±5MV/cm. Under Vg<0, some of the neutralized positive charges can be recharged without resuming hole injection. When the gate bias polarity is switched, part of the positive charges can be repeatedly neutralized and recharged, so that they are referred to as cyclic positive charge (CPC). The neutralization of some positive charge under Vg>0 is more difficult than their charging under Vg<0 and they are referred to as anti-neutralization positive charge (ANPC).

Figs.3.2 (b) & 3.2 (c) show that the energy level of ANPC is higher than that of CPC and is above the bottom edge of Si conduction band. This explains the difficulty for neutralizing ANPC. Under the same Eox=-5MV/cm, the recharging under Vg<0 without substrate hole injection (SHI) clearly did not reach the level when SHI was switched on. This indicates that some defects are difficult to recharge without accelerating holes in the substrate, as illustrated by Fig.3.2 (c), and they are as-grown hole traps (AHT) [18-21].

3.2.2   Similarity in NBTI defects between Hf-stack and SiON

Fig.3.3 shows the behavior of positive charges formed in an Hf-silicate/SiON stack after negative bias temperature stress. A comparison with the positive charges formed in SiO_2 in Fig.3.2(a) indicates that the same three types of positive charges exist despite of the material difference [13]. This observation is reinforced by Fig. 3.4 that shows that, on one hand, CPC is insensitive to the measurement temperature, since electron tunneling at the same energy level is not accelerated by thermal energy. On the other hand, the ANPC increases when the measurement temperature

reduces. This is because lowering temperature will reduce the number of free electrons at energy levels above Ec, so that more ANPC survives the neutralization. This dependence on temperature is the same as what was observed for positive charges in SiO_2 [18-21].
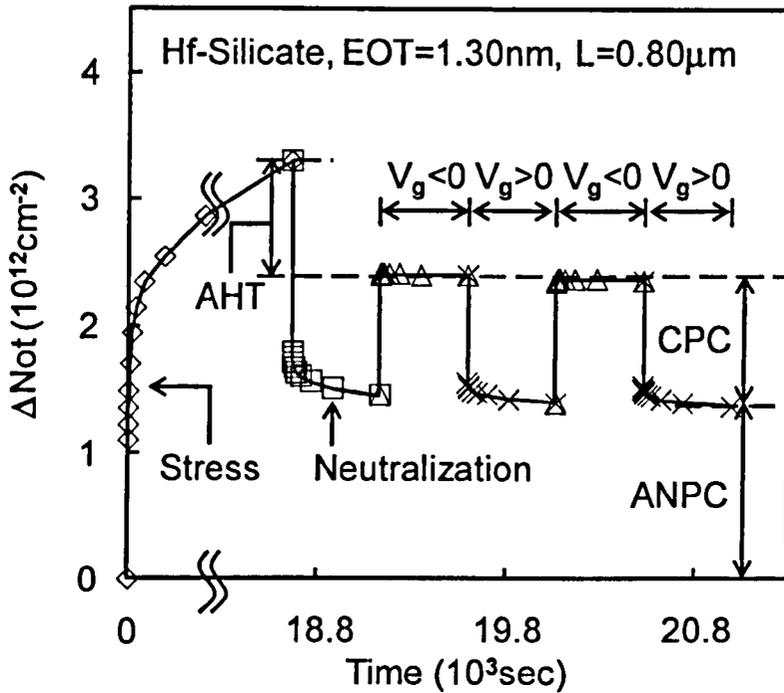
Fig.3.3 Types of positive charges in Hf-silicate (70% Hf). The Hf-silicate was first stressed for a relatively long time (18,662sec) under Vg= –2.8 V. The stress period is not drawn to scale here. To determine the types of positive charges, the sensing phase started with the neutralization carried out under Eox(EOT)=+6.5MV/cm, followed by applying Vg<0 and Vg>0 corresponding to Eox(EOT)=±5MV/cm, respectively. A comparison with Fig.3.2(a) indicates that the same three types of positive charges were formed in SiO_2 and Hf-dielectrics. [13]

To further support that different types of positive charges exist in Hf-dielectric/SiON stack, Fig.3.5 compares the dependence of ANPC and CPC on stress time and temperature. While ANPC increases for longer stress time and higher temperature, CPC is insensitive to the time and temperature. Such differences in ANPC and CPC support that they originate from different defects.
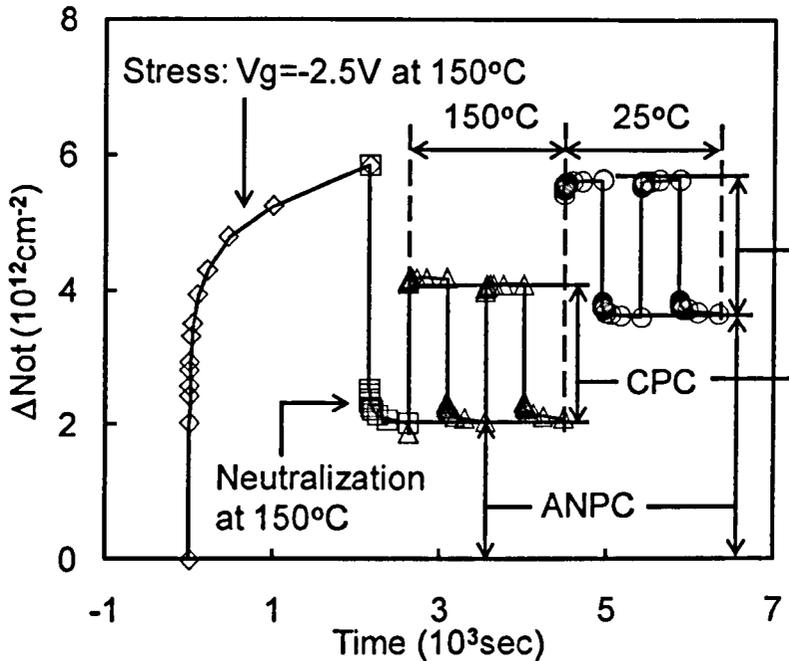
Fig.3.4 Dependence of different types of positive charges on measurement temperature. The initial test sequence is the same as that in Fig.3.3. After measurement at the stress temperature of 150°C, the device was cooled down to 25°C and both CPC and ANPC were assessed again. It is clear that ANPC increases for lower measurement temperature, but CPC does not. [13]
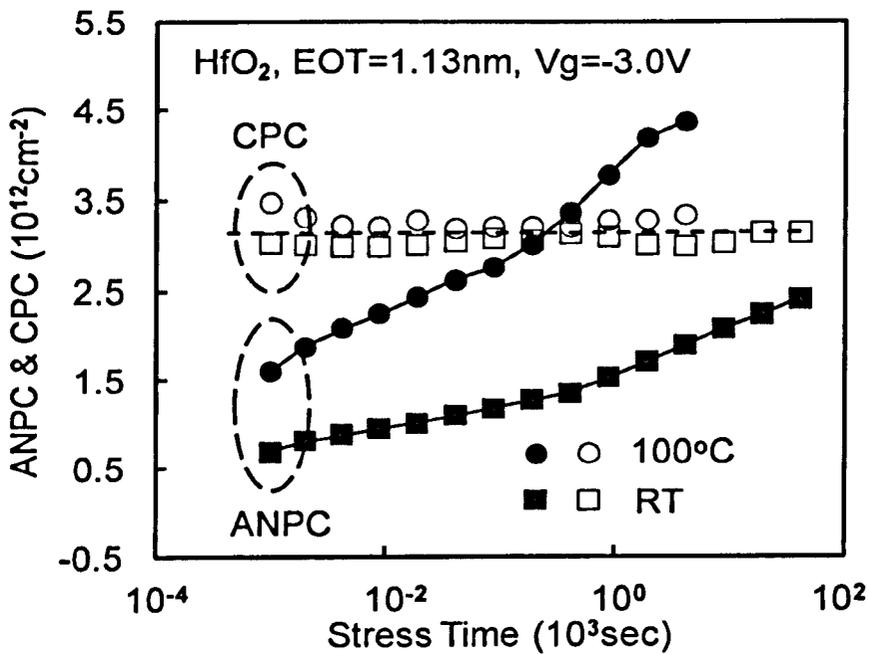
Fig.3.5 Dependence of ANPC and CPC on stress time and temperature. ANPC increases with both stress temperature and time, but CPC does not. [13]

3.2.3 Differences in NBTI properties between Hf-stack and SiON

After reporting that the same framework can be used for positive charging in SiO_2

and Hf-dielectric/SiON stack, the reported difference in the NBTI for these two is

given in Fig.3.6. It can be seen that the power factor of NBTI kinetics for Hf-stack is

substantially lower than that for SiO_2. A close examination of the ΔVth for the

Hf-stack at short stress time shows that there is a flat region in Fig.3.7. This flat

region originates from the as-grown CPC in the Hf-stack, while the as-grown CPC is

negligible in SiO_2, as shown in Fig.3.8. Fig.3.7 shows that how to remove the

contribution of this flat region to the NBTI kinetics. After such a correction, Fig.3.6

shows that the power factor of Hf-stack agrees well with that of SiO_2.
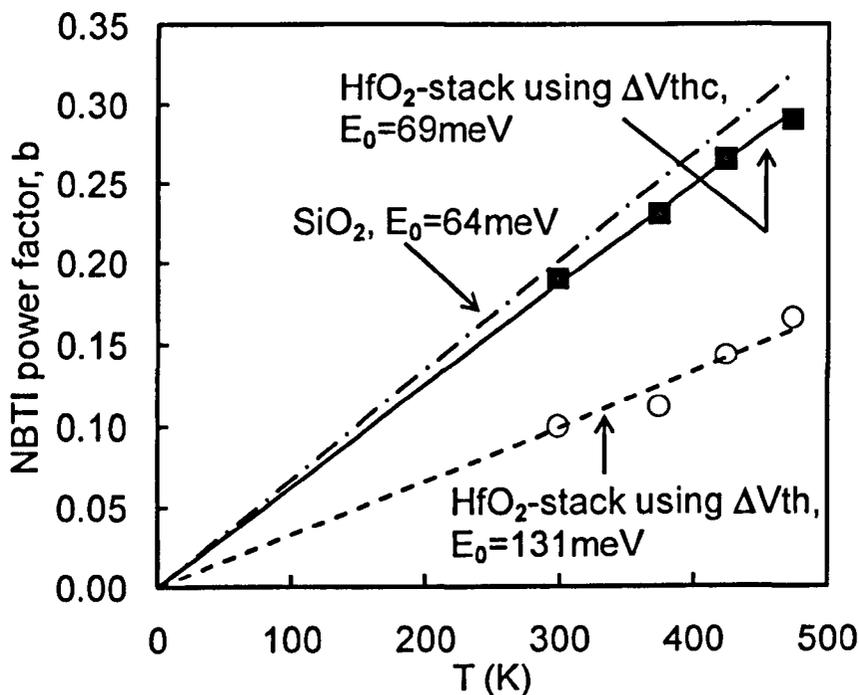
Fig.3.6 A comparison of the power factor of the NBTI kinetics for Hf-stacks with that for SiO_2. The power factor extracted from the total ΔVth (the symbol 'o') is substantially less than that of SiO_2. The symbol '■' is the power factor for $\Delta Vthc$ shown in Fig.3.7. The result for SiO_2 was taken from ref. 22. [12]
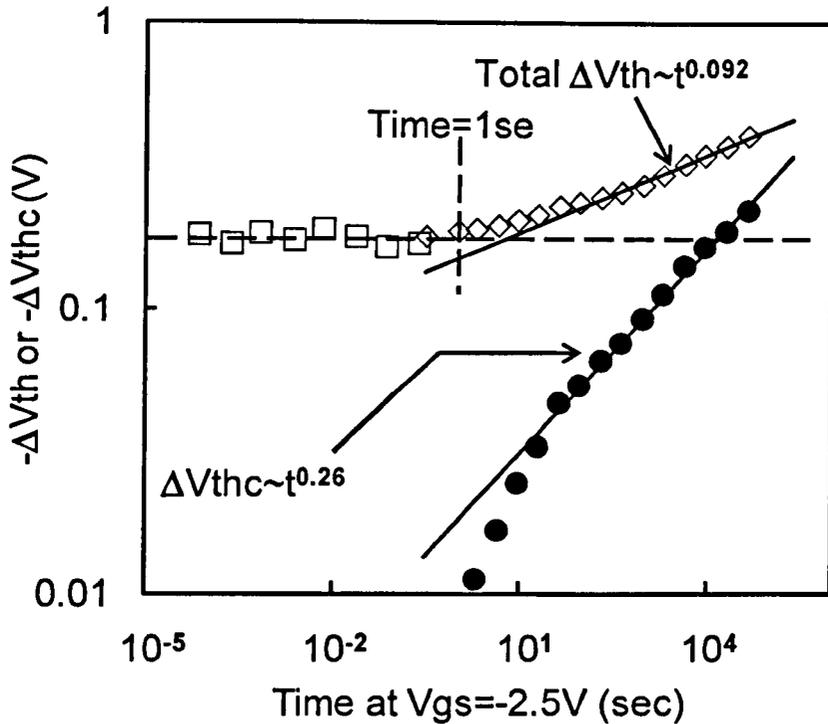
Fig.3.7 The 'flat-then-rise' feature of NBTI kinetics for Hf-stacks. The total ΔVth is 'flat' when stress time is less than 1 sec and only starts 'rise' at longer time. The symbols '�口' and '◊' were measured by using pulsed and quasi-DC Id~Vg, respectively. The ΔVthc (symbol '●') represents the rising part of ΔVth. The power factor increases from 0.092 to 0.26 when the total ΔVth was replaced by ΔVthc. The stress was at a gate bias of -2.5 V and room temperature. [12]
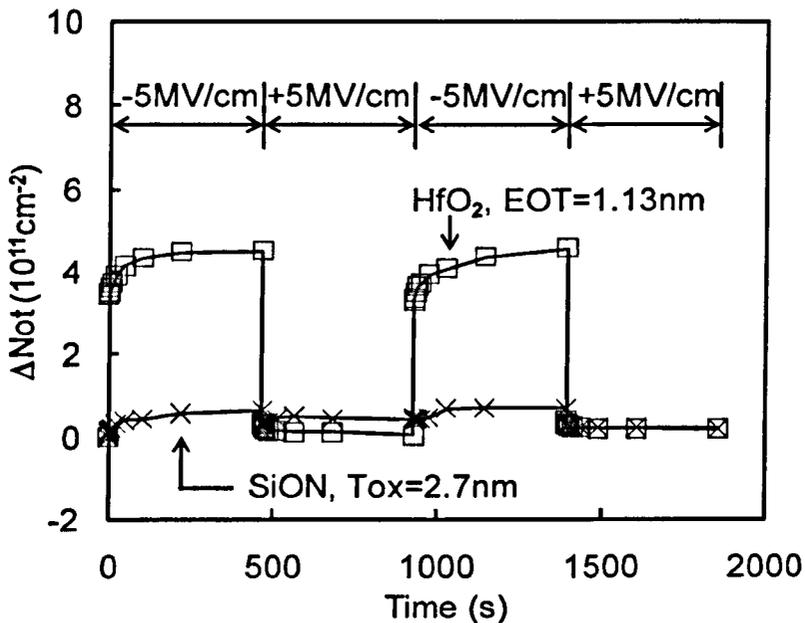
Fig.3.8. A comparison of cyclic positive charging (CPC) in fresh SiON and HfO₂. Although the CPC is negligible for a fresh SiON, it is considerable in a fresh HfO₂. [12]

To explain why there can be a large amount of as-grown CPC in a Hf-stack, but not in SiO_2, it has been pointed out that the samples experienced different amount of hydrogen exposure during their fabrication: the Hf-stack was prepared by ALCVD and SiO_2 was thermally grown. Fig.3.9 shows that when an SiO_2 MOS structure was subjected to an extensive hydrogenation, there could also be a large amount of as-grown CPC.
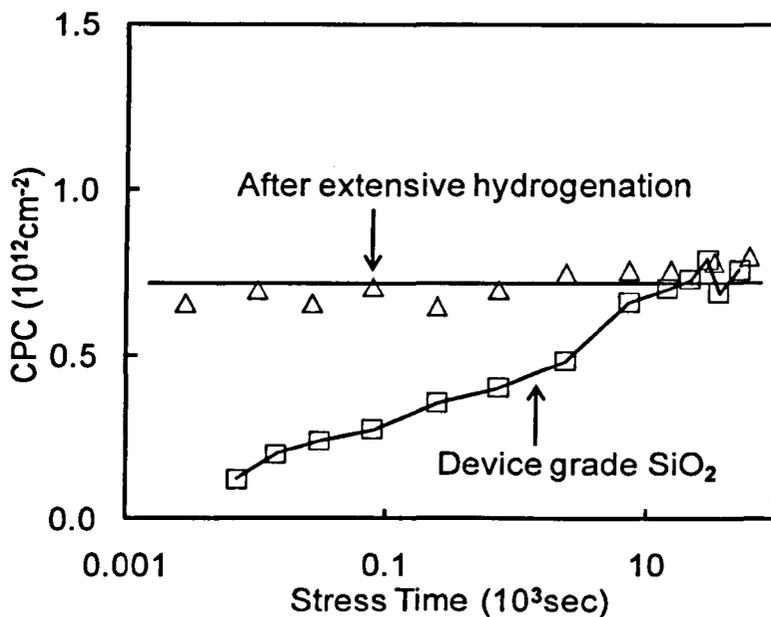
Fig.3.9 Dependence of CPC in SiO_2 on stress level. For a device grade SiO_2, CPC built up gradually with stress. After an extensive hydrogenation, CPC in SiO_2 becomes insensitive to stress and behaves like the Hf-stack [9].

Through the above review, it becomes clear that the HfO_2 based dielectric has not introduced any new type of positive charges and its NBTI kinetics is similar to that of SiO_2. This raises the possibility that the positive charging in the Hf-stack is actually dominated by the interfacial SiON layer, although this layer is generally sub 1 nm and thinner than the Hf-dielectric layer. In the rest of this chapter, we will present unambiguous experimental evidence to show that this is indeed the case.

3.3 Test Samples

To assess which layer in the Hf-dielectric/SiON stack dominates the positive charge, one would like to vary the thickness of one layer and keep the rest of process conditions identical. Two groups of samples were used in this chapter: one with variable Hf-dielectric thickness and the other with variable interfacial SiON layer thickness.

Ideally, one hopes to change the Hf-dielectric layer thickness on the same wafer to minimize the potential wafer-to-wafer variations, but such samples were not available. In this work, Hf-silicates (80% Hf) were prepared by atomic layer deposition to a thickness of 1, 2 and 3nm on three IMEC-cleaned wafers. The interfacial layer (IL) is fixed at 1nm for all three samples and the gate is TiN. Both the channel length and width is 1 μm. The maximum thickness of Hf-silicates is limited to 3nm, since further increase can cause crystallization of the film.

The thickness of the interfacial layer is varied by using a slant-etched wafer, as illustrated in Fig.3.10. The HfSiON layer is fixed at 2 nm and the IL thickness changes between 1.7 and 3.6 nm across the wafer.

HfSiON

SiON
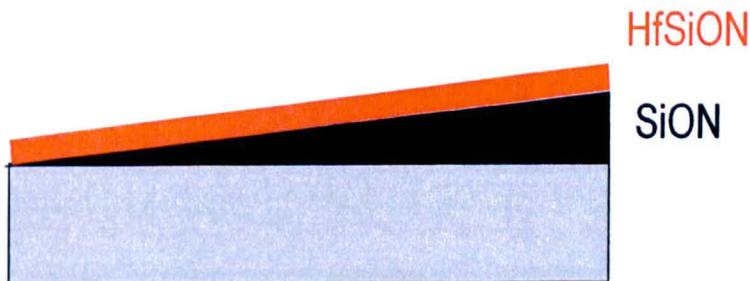
Fig.3.10 A schematic illustration of the slant-etched wafer with variable interfacial layer thickness.

3.4 The dielectric charges versus interface states

To determine the dominant layer for positive charges in gate dielectric, ΔNot, three samples with different thickness of HfSiON layers were used. The problem is that positive charges increase with stress level and a comparison of different samples is meaningful only if they were subjected to the same stress. Under the same effective oxide field, Fig.3.11 shows that the difference in their density of generated interface states, ΔDit, is negligible, confirming that they were subjected to the same stress level.
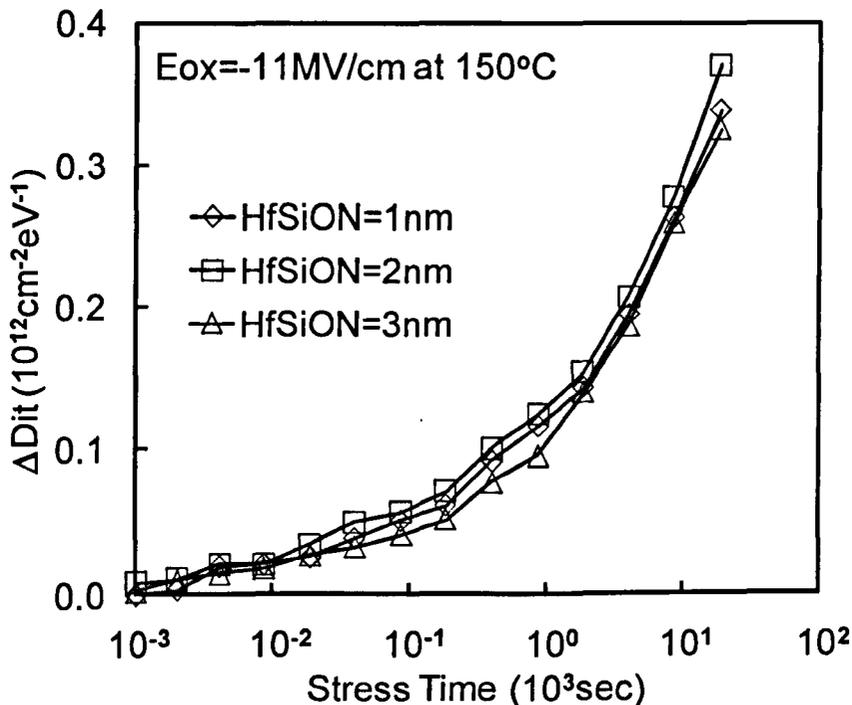
Fig.3.11 The density of interface states generated, ΔDit, in three samples of different HfSiON layers. The electrical field, Eox=-11MV/cm, is the field strength over the equivalent oxide thickness. The negligible difference in ΔDit for the three samples confirms that they were subjected to the same stress level for a given time.

Before assessing the spatial location of ΔNot in Hf-based stacks, we examine its relative importance to the NBTI-induced threshold voltage shift. As described in section 2.3, the classical measurement of threshold voltage is to linearly extrapolate the transfer characteristics from the Vg where the transconductance is at its maximum

to Id=0. The shift of threshold voltage measured in this way, ΔVth, contains contribution from both the generated interface states and the positive charges in the gate dielectric formed during the negative bias temperature stress (NBTS), so that

ΔVth = ΔVth(ΔDit) + ΔVth(ΔNot).

Early works [23,24] reported that the generated interface states are acceptor-type above the midgap of Si and donor-type below the midgap. As a result, ΔVth(ΔNot) can be determined from the shift of gate voltage at the midgap of Si, where the net charge of interface states is negligible.
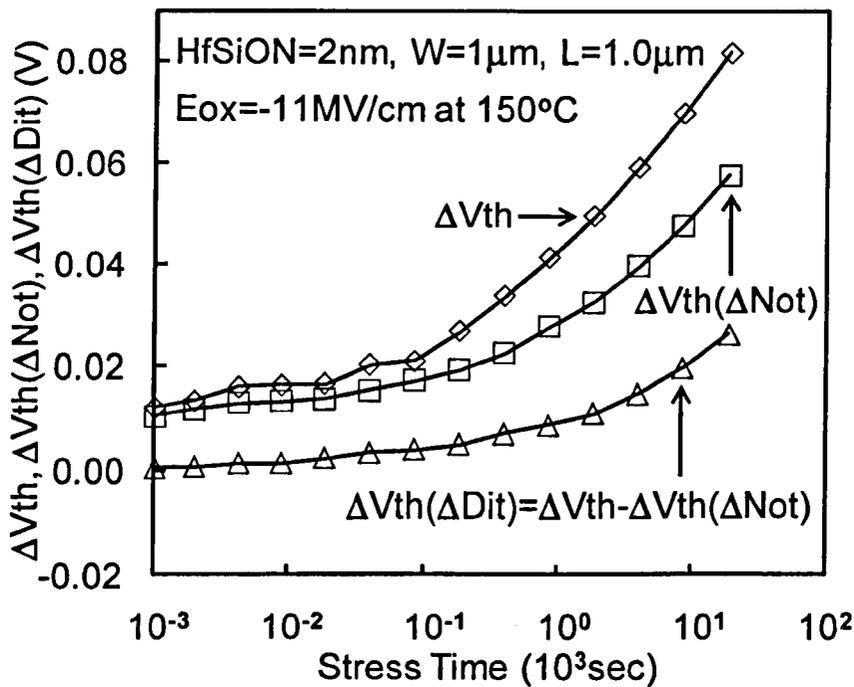
Fig.3.12 The threshold voltage shift, ΔVth, and its two components. ΔVth(ΔNot) and ΔVth(ΔDit) is the contribution from the positive charges in dielectric and the generated interface states, respectively. ΔVth(ΔNot) is clearly larger than ΔVth(ΔDit) for the Hf-based dielectric stack.

Fig.3.12 compares the contributions of ΔNot and ΔDit to ΔVth. Initially, ΔDit is negligible and ΔVth is dominated by ΔNot, since there are pre-existing CPC in the stack [12,13], as described in section 3.2.3. As stress time increases, the contribution of ΔDit rises and reaches about one third of ΔVth for the last point in Fig.3.12.

3.5 Volume distribution of dielectric positive charges

3.5.1 Uniform volume distribution

As the first attempt for assessing the spatial distribution of positive charges, we study the case that they are uniformly distributed throughout the whole stack with a volume density of ρ. $\Delta Vth(\Delta Not)$ is related to ρ by,

$$\Delta Vth(\Delta Not) = -\frac{q \cdot \rho \cdot X_{HF}^2}{2 \cdot \varepsilon_0 \cdot k_{HF}} - \frac{q \cdot \rho}{2 \cdot \varepsilon_0 \cdot k_{IL}} \cdot \left[X_{IL}^2 + 2 \cdot \frac{k_{IL}}{k_{HF}} \cdot X_{IL} \cdot X_{HF} \right], \tag{3.1}$$

where X is the dielectric layer thickness and k the dielectric constant (k_{HF}=16.3 for the Hf-silicates here). The subscripts 'HF' and 'IL' represent HfSiON and interfacial layer, respectively.
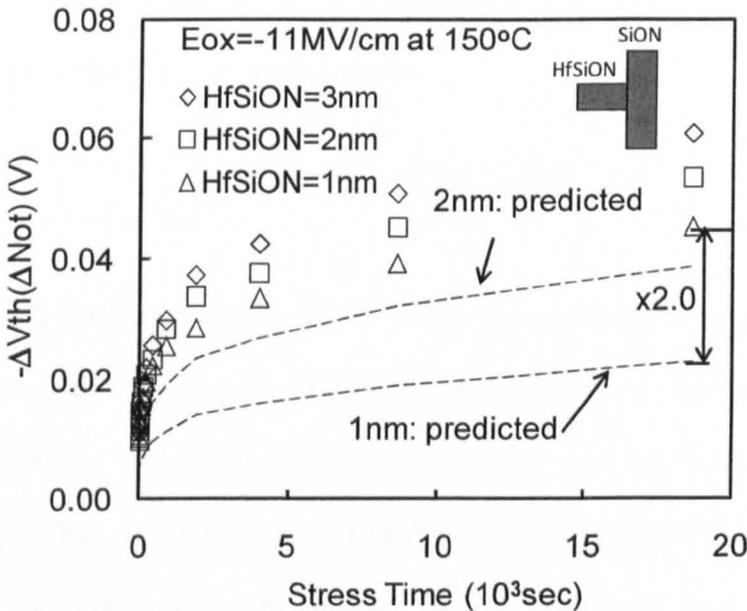
Fig.3.13 Uniform distribution of positive charges throughout the HfSiON/SiON stack. The dashed line is the fitted curve with equation (3.1) and the poor agreement with test data indicates that positive charges are not uniformly distributed in the stack. The shaded area of the inset shows the location of positive charges.

Fig.3.13 plots the ΔVth(ΔNot) measured at different stress time for X_{HF}=1, 2 and

3nm. At a given time point, ΔVth(ΔNot) at X_{HF}=3nm and the equation (3.1) were

used first to calculate the volume density ρ. This ρ is then used to predict the

ΔVth(ΔNot) at X_{HF}=1 and 2nm based on equation (3.1). The predicted values are

compared with the measured value for samples with X_{HF}=1 and 2nm. It can be seen

that the predicted value does not agree with the measured ones. For X_{HF}=1nm, the

predicted value is only about half of the measured value. As a result, the assumption

of a uniform distribution throughout the stack does not agree with the test data.

3.5.2 All charges in HfSiON layer

The next assumption to examine is that all positive charges are located within the

HfSiON layer with a volume density of ρ_{HF} and there are no positive charges in the

interfacial layer. ΔVth(ΔNot) is related to ρ_{HF} by,

$$\Delta Vth(\Delta Not) = -\frac{q \cdot \rho_{HF} \cdot X_{HF}^2}{2 \cdot \varepsilon_0 \cdot k_{HF}}. \tag{3.2}$$

Again at a given time point, ΔVth(ΔNot) at X_{HF}=3nm and the equation (3.2) were

used to calculate the volume density ρ_{HF}. This ρ_{HF} is then used to predict the

ΔVth(ΔNot) at X_{HF}=1 and 2nm based on equation (3.2). The predicted ΔVth(ΔNot)

is represented by the two dotted lines in Fig.3.14 and they depart further from the

test data. The predicted ΔVth(ΔNot) at X_{HF}=1nm is only 15% of the test data, so

that we can rule out that positive charges are dominated by the bulk of HfSiON. This

is in sharp contrast with the location of electron traps, which are mainly in the

Hf-dielectric layer. It also rules out the possibility that electron traps and positive

charges have a common origin.
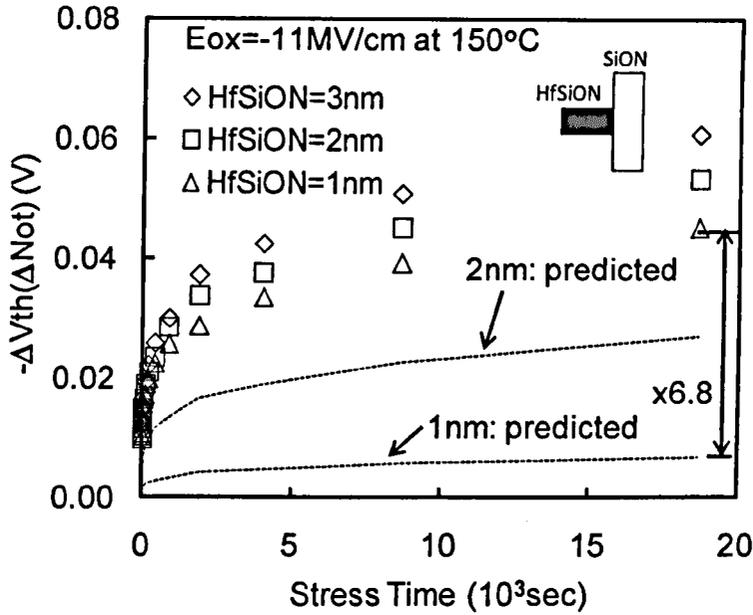
Fig.3.14 All positive charges are located within the HfSiON and there are no charges in the interfacial layer. The dotted line is the fitted curve with equation (3.2) and the poor agreement with test data indicates that positive charges are not dominated by the HfSiON bulk. The shaded area of the inset shows the location of positive charges.

3.5.3   All charges in the interfacial SiON layer

Since positive charges are not dominated by the HfSiON layer, it is natural to examine whether they are dominated by the interfacial SiON layer. If one assumes that the interfacial layer has a volume density of ρ_{IL} and ρ_{HF} =0, $\Delta Vth(\Delta Not)$ is related to ρ_{HF} by,

$$\Delta Vth(\Delta Not) = -\frac{q \cdot \rho_{IL}}{2 \cdot \varepsilon_0 \cdot k_{IL}} \cdot \left[X_{IL}^2 + 2 \cdot \frac{k_{IL}}{k_{HF}} \cdot X_{IL} \cdot X_{HF} \right], \qquad (3.3)$$

The predicted $\Delta Vth(\Delta Not)$ based on equation (3.3) is represented by the two solid lines in Fig.3.15 and they are closest to the test data among the three assumptions made so far. This supports that the IL dominates positive charges.
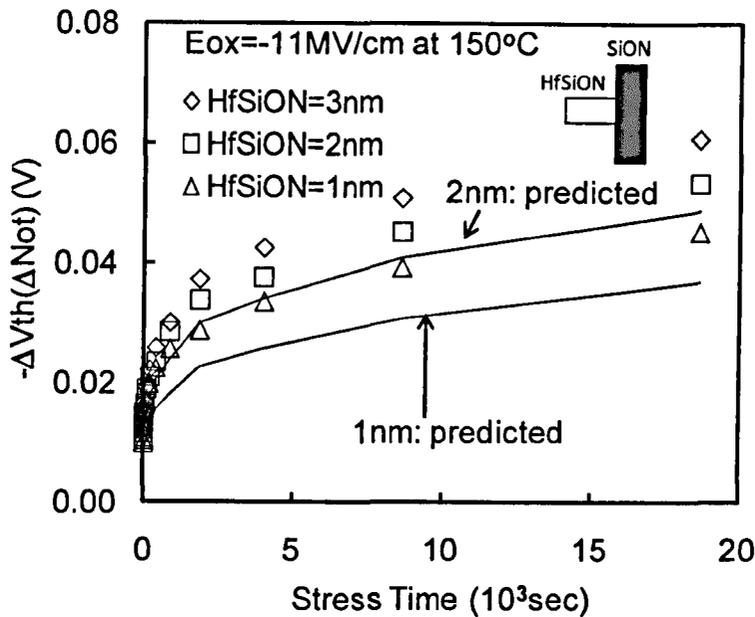
Fig.3.15 All positive charges are located within the interfacial layer and there are no charges in the HfSiON. The solid line is the fitted curve with equation (3.3). The shaded area of the inset shows the location of positive charges.

3.6 Sheet distribution of dielectric positive charges

The difference between the prediction and test date in Fig.3.15 is still considerable, indicating that the positive charges are not uniformly distributed in the interfacial layer. To explore the location of positive charge further and to improve the agreement between prediction and test data, the sheet distribution will be explored in this section.

3.6.1 Positive charges at the HfSiON/SiON interface

In comparison with conventional MOSFETs with SiON as gate dielectric, one new feature for an Hf-based stack is the presence of an HfSiON/IL interface and the

question is whether positive charges pile up at this interface. If one assumes that all

charges are at the HfSiON/IL interface with an area density of Q, it will give,

$$\Delta Vth(\Delta Not) = -\frac{q \cdot X_{HF} \cdot Q}{\varepsilon_0 \cdot k_{HF}}.$$ (3.4)

At a given time point, $\Delta Vth(\Delta Not)$ at $X_{HF}=3nm$ and the equation (3.4) were used

first to calculate the area density Q. This Q is then used to predict the $\Delta Vth(\Delta Not)$ at

$X_{HF}=1$ and 2nm based on equation (3.4). The predicted value is shown as the two

dashed lines in Fig.3.16 and it is less than half of the test data for $X_{HF}=1nm$.
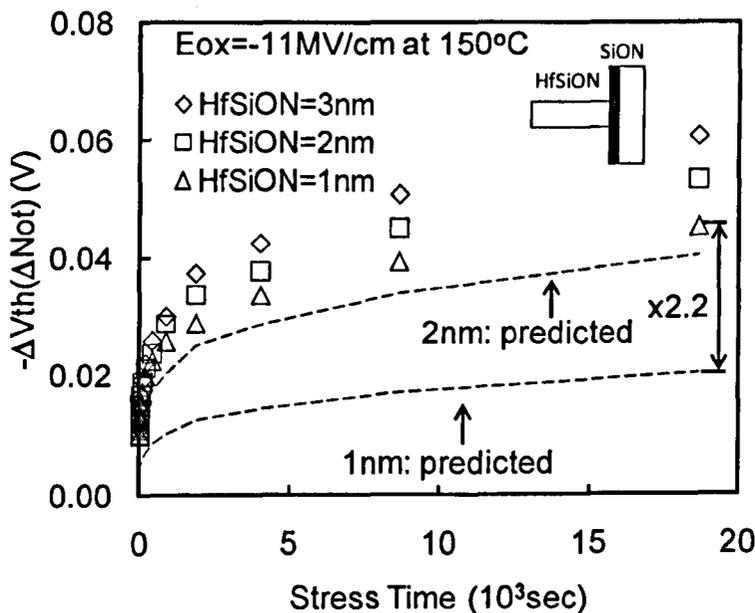
Fig.3.16 A comparison of test data (symbols) with the prediction (dashed lines) based on sheet distributions. The prediction is based on the equation (3.4) under the assumptions that the positive charges are at the HfSiON/IL interface. The thick gray line in the inset shows the location of positive charges.

To further rule out the possibility that positive charges pile up at the HfSiON/IL

interface, the dependence of $\Delta Vth(\Delta Not)$ on the IL thickness can be examined, by

using the slant-etched samples described in section 3.3. Eq. (3.4) requires

$\Delta Vth(\Delta Not)$ to be independent of IL thickness. This is against our test results in Fig.

3.17 that shows ΔVth(ΔNot) increasing for thicker IL. It can be concluded that positive charges are not concentrated at the HfSiON/IL interface.
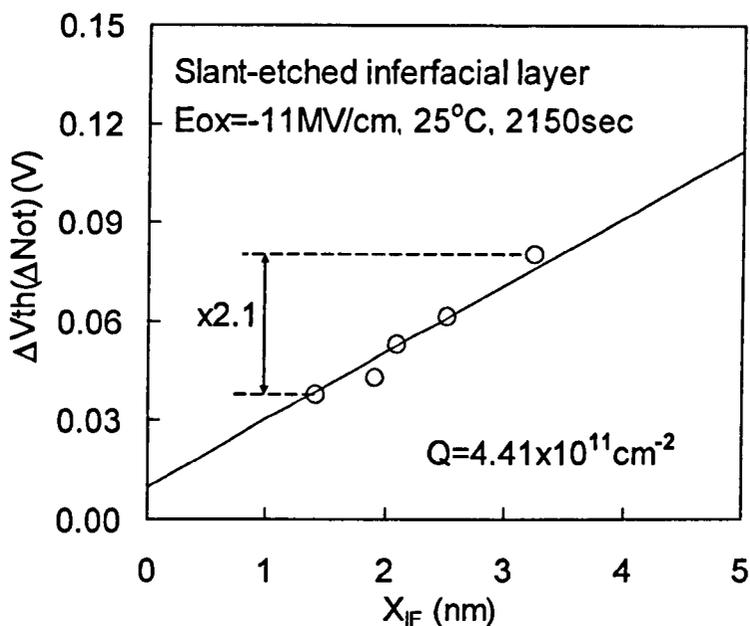
Fig.3.17 Effects of interfacial layer thickness, X_{IF}, on ΔVth(ΔNot). The HfSiON layer is fixed at 2nm. The solid line is fitted with eq. (3.5) by assuming that charges pile up at the substrate interface.

3.6.2 Positive charges at the SiON/substrate interface

After ruling out the HfSiON/IL interface, it is natural to study if positive charges pile up at the IL/Si interface. By assuming all charges at the substrate interface, we have,

$$\Delta Vth(\Delta Not) = -qQ\left(\frac{X_{IL}}{\varepsilon_0 k_{IL}} + \frac{X_{HF}}{\varepsilon_0 k_{HF}}\right).$$ (3.5)

The predicted value is shown as the two solid lines in Fig.3.18 and good agreement with test data is achieved. This supports that positive charges pile up at the IL/substrate interface. The smaller measured ΔVth(ΔNot) at X_{HF}=1nm in Fig.3.18 results from a larger gate capacitance, when compared with that at X_{HF}=3nm.

The assumption that positive charges pile up at the IL/substrate interface leads to the prediction that ΔVth(ΔNot) increases linearly with both the Hf-dielectric and interfacial layer thickness in equation (3.5). Fig.3.18 shows that this prediction agrees with the test data when varying the Hf-dielectric thickness. To further support this assumption, Fig.3.17 shows that this prediction also agrees well with the test data when the IL thickness is changed. It can be concluded that positive charges pile up at the IL/substrate interface. Consequently, a reduction of HfSiON thickness has little effect on positive charging.
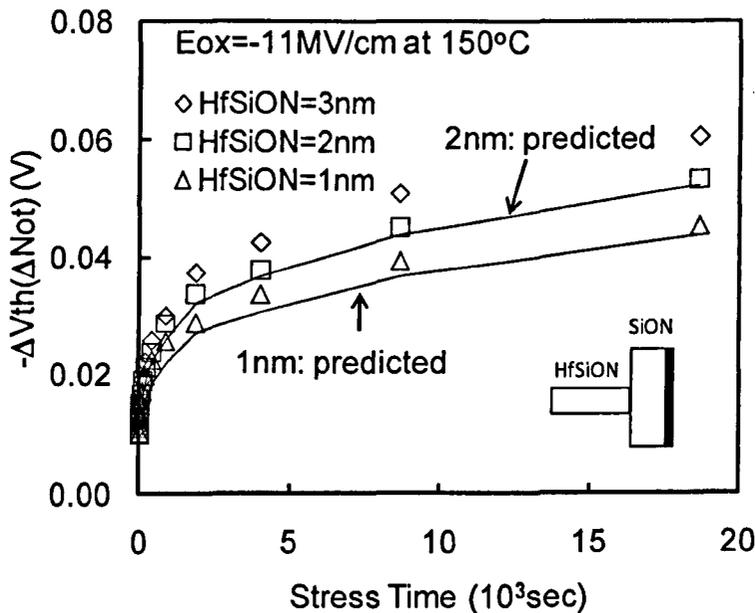
Fig.3.18 A comparison of test data (symbols) with the prediction (lines) based on sheet distributions. The prediction is based on the equation (3.5) under the assumptions that the positive charges are at the IL/Silicon interface. The thick gray line in the inset shows the location of positive charges.

To explain this pile-up, it should be noted that the oxygen vacancy, $\equiv$Si—Si$\equiv$, has been proposed as the main source for positive charges [25]. It is reasonable to assume that there are more $\equiv$Si—Si$\equiv$, when moving from dielectric towards silicon substrate. The data, however, does not give direct information on the origin or microscopic structure of positive charges.

3.7   Sensitivity to stress and measurement conditions

The test results used up to here were obtained under an electrical field over the equivalent oxide thickness of Eox= -11MV/cm and at 150 °C. The measurement was carried out using quasi-DC transfer characteristics that took 6sec. The question is whether the dominant layer is sensitive to the stress bias and measurement speed. To test the impact of stress field, temperature, and measurement time, the devices were stressed under Eox=-10MV/cm at room temperature. The pulse transfer characteristics were used, allowing the measurement time reducing from 6 sec to 5µs [26-28]. Fig.3.19 compares the difference caused by the measurement time.
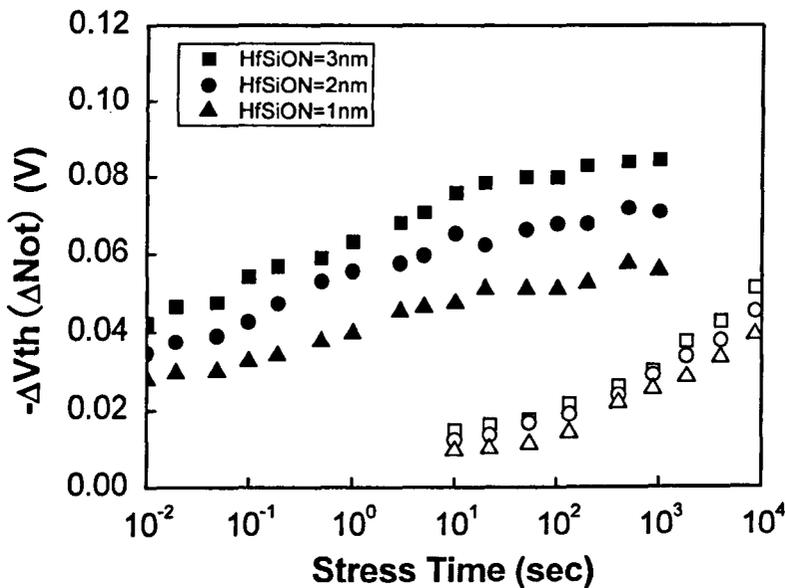
Fig.3.19 A comparison of test data with quasi-DC measurement and UFP measurement. The solid symbols present the data measured by UFP method. The empty symbols present the data measured by quasi-DC method. The difference will be explained in chapter 4 and 5.

Fig.3.20 compares the test data with the predictions based on volume distribution of positive charges. The worst agreement is for the assumption that the HfSiON layer

dominates the positive charges and the agreement is still poor under the assumption of a uniform distribution through the stack. The best agreement is obtained for the assumption that the interfacial layer dominates the positive charges.

Fig.3.21 compares the test data with the prediction based on sheet distribution. The difference between the test data and the prediction is substantial, when assuming that the positive charges are located at the HfSiON/SiON interface. The agreement is good under the assumption that the positive charges pile up at the interfacial layer/substrate interface. As a result, the conclusion of the interfacial layer dominating positive charging will not change with stress and measurement conditions.
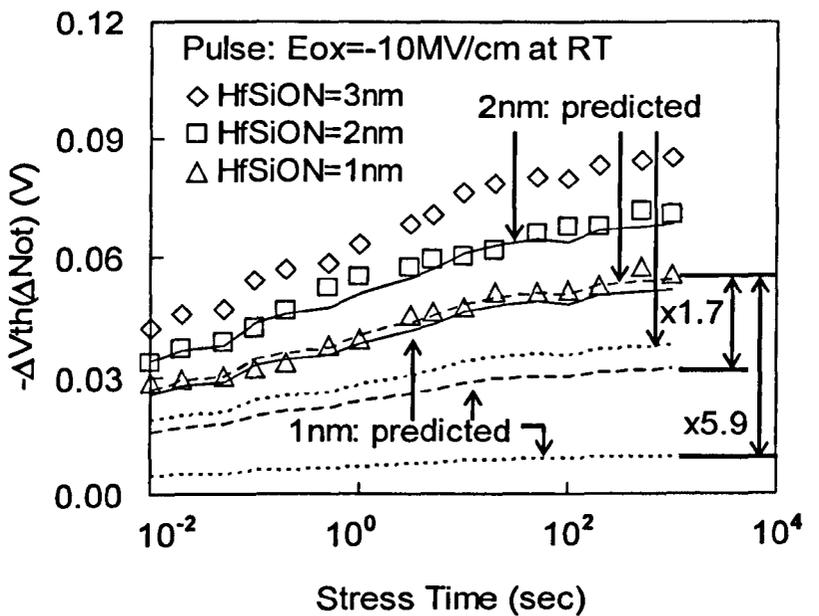
Fig.3.20 A comparison of test data (symbols) with the prediction (lines) based on volume distributions. The prediction is under three assumptions: (i) $\rho_{HF}=\rho_{IF}$, dashed lines; (ii) $\rho_{IF}=0$, dotted lines; and (iii) $\rho_{HF}=0$, solid lines. The pulse measurement time is 5μsec.
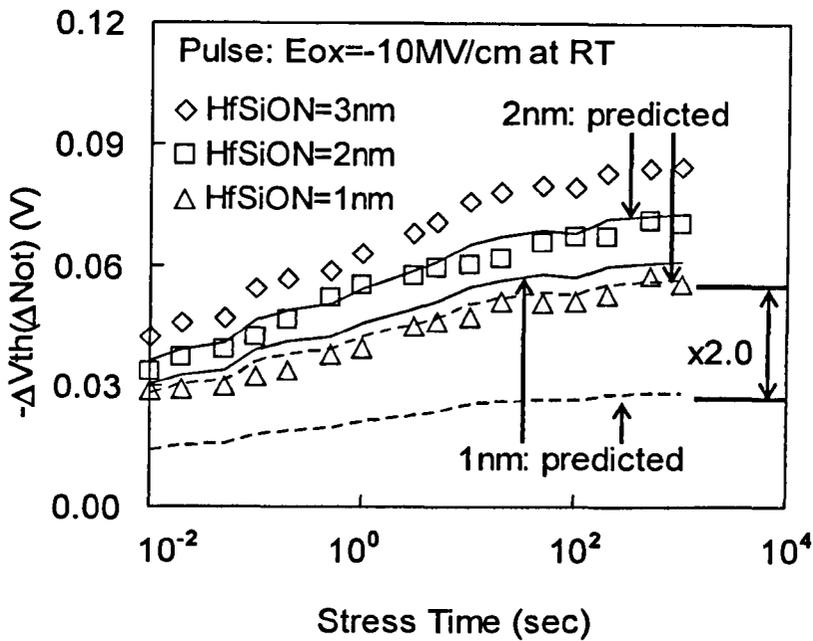
Fig.3.21 A comparison of test data (symbols) with the prediction (lines) based on sheet distributions. The prediction is under two assumptions for positive charges: (i) at HfSiON/IL interface, dashed lines and (ii) at the IL/substrate interface, solid lines. The pulse measurement time is 5μs.

3.8 Conclusion

In this chapter, the recent progresses on understanding the positive charges in dielectric are reviewed. Three different types of positive charges were identified: as-grown hole traps, cyclic positive charges, and anti-neutralization positive charges. Despite of the difference observed in the NBTI kinetics for SiON and the Hf-dielecric/SiON stacks, it is found that the same framework applies to both of them. This indicates that positive charging could be dominated by the interfacial SiON layer in the stack.

Efforts were focused on assessing the dominant layer for the stress-induced positive charge in Hf-based stacks by varying the thickness of HfSiON and interfacial layers. It is concluded that positive charges in the stack are indeed dominated by the

interfacial layer. It is ruled out that positive charges pile up at the HfSiON/IL interface. The results support that positive charges are located close to the IL/substrate interface. Consequently, unlike electron trapping, a reduction of HfSiON thickness will not reduce positive charges.

Since the NBTI positive charging for the Hf-based dielectric stack is dominated by the interfacial SiON layer, the rest of this project will be carried out mainly on SiON samples. The process-to-process variations of SiON samples are smaller and there are more SiON samples available to this project. It is expected that the conclusions obtained from the SiON samples are also applicable to the Hf-based dielectric stacks.

References

1. K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neirynck, S. Pei, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Schifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki, "A 45 nm logic technology with high-k+metal gate transistors, strained silicon, 9 Cu interconnect layers, 193 nm dry patterning, and 100% Pb-free packaging,", in *IEDM Tech. Dig.*, 247 (2007).

2. J. H. Sim, S.C. Song, P. D. Kirsch, C. D. Young, R. Choi, D. L. Kwong, B. H. Lee, and G. Bersuker, "Effects of ALD HfO_2 thickness on charge trapping and mobility," *Microelectronic Eng.,* **80**, 218 (2005).

3. C. Z. Zhao, M. B. Zahid, J. F. Zhang, G. Groeseneken, R. Degraeve, and S. De Gendt, "Properties and dynamic behavior of electron traps in HfO_2/SiO_2 stacks," *Microelectronic Eng.,* **80**, pp.366-369 (2005).

4. C. Z. Zhao, J. F. Zhang, M. B. Zahid, B. Govoreanu, G. Groeseneken, and S. De Gendt, "Determination of capture cross sections for as-grown electron traps in HfO_2/HfSiO stacks," *J. Appl. Phys.,* **100**, Art. No.093716 (2006).

5. T. P. Ma, H. M. Bu, X. W. Wang, L. Y. Song, W. He, M. Wang, H.-H. Tseng, and P. J. Tobin, "Special reliability features for Hf-based high-κ gate dielectrics," *IEEE Trans. Device Mater. Rel.,* vol. 5, no. 1, pp. 36–44 (2005).

6. R. Degraeve, A. Kerber, P. Roussel, E. Cartier, T. Kauerauf, L. Pantisano, and G. Groeseneken, "Effect of bulk trap density on HfO_2 reliability and yield," in *IEDM Tech. Dig.*, pp. 935–938 (2003).

7. Z. Ji, J. F. Zhang, W. Zhang, G. Groeseneken L. Pantisano, S. De Gendt, M. M. Heyns, "An assessment of the mobility degradation induced by remote charge scattering," *Applied Phys. Lett.*, **95**, Art. No. 263502 (2009).

8. J. F. Zhang, C. Z. Zhao, M. B. Zahid, G. Groeseneken, R. Degraeve, and S. De Gendt, "An assessment of the location of as-grown electron traps in HfO_2/HiSiO stacks," *IEEE Electron Dev. Lett.*, **27**, no.10, pp.817-820 (2006).

9. J. F. Zhang, "Defects and Instabilities in Hf-dielectric/SiON stacks," *Proc. Of The 16th Biennial Conference on Insulating Films on Semiconductors (INFOS)*, pp.1883-1887, Cambridge University, UK (2009).

10. C. Z. Zhao, M. B. Zahid, J. F. Zhang, G. Groeseneken, R. Degraeve, and S. De Gendt, "Threshold voltage instability of p-channel metal-oxide-semiconductor field effect transistors with Hafnium based dielectrics," *Applied Physics Letters*, **90**, Art.no.143502 (2007).

11. C. Z. Zhao, J. F. Zhang, M. B. Zahid, G. Groeseneken, R. Degraeve, and S. De Gendt,   "Impact of gate materials on positive charge formation in HfO2/SiO2 stacks," *Applied Physics Letters*, **89**, Art.No. 023507 (2006).

12. J. F. Zhang, C. Z. Zhao, M. H. Chang, M. B. Zahid, A. R. Peaker, S. Hall, G. Groeseneken, L. Pantisano, S. De Gendt, and M.Heyns, "Impact of different defects on the kinetics of negative bias temperature instability of hafnium stacks," *Applied Physics Letters*, **92**, Art.no. 013502 (2008).

13. C. Z. Zhao, J. F. Zhang, M. H. Chang, A. R. Peaker, S. Hall, G. Groeseneken, L. Pantisano, S. De Gendt, and M. Heyns, "Stress-induced positive charge in

Hf-based gate dielectrics: impact on device performance and a framework for the defect," *IEEE Trans. Electron Dev.,* **55,** pp.1647-1656 (2008).

14. D. Felnhofer, E. P. Gusev, P. Jamison, D. A. Buchanan, "Charge trapping and detrapping in HfO_2 high-k MOS capacitors using internal photoemission," *Microelectronic Eng.,* **80,** 58 (2005).

15. D. R. Young, E. A. Irene, D. J. DiMaria, R. F. De Keersmaecker, and H. Z. Massoud "Electron trapping in SiO_2 at 295 and 77°K," *J. Appl. Phys.,* **50,** pp.6366-6372 (1979).

16. S. K. Lai and D. R. Young, "Effects of avalanche injection of electrons into silicon dioxide — generation of fast and slow interface states," *J. Appl. Phys.,* **52,** pp.6231-6240, (1981).

17. D. M. Fleetwood, "Effects of hydrogen transport and reactions on microelectronics radiation response and reliability," *Microelectronics Reliability,* **42,** pp.523-541 (2002).

18. 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 Dev.,* **51,** pp. 1267-1273 (2004).

19. C. Z. Zhao, J. F. Zhang, G. Groeseneken, and R. Degraeve, "Hole-traps in silicon dioxides - Part II: Generation mechanism," *IEEE Trans. Electron Dev.,* **51,** pp. 1274-1280 (2004).

20. C. Z. Zhao and J. F. Zhang, "Effects of hydrogen on positive charges in gate oxides," *J. Appl. Phys.,* **97,** Art. No. 073703 (2005).

21. M. H. Chang and J. F. Zhang, "On positive charges formed under negative bias temperature stresses," *J. Appl. Phys.,* **101,** Art. no. 024516 (2007).

22. B. Kaczer, V. Arkhipov, M. Jurczak, and G. Groeseneken, "Negative bias temperature instability (NBTI) in SiO_2 and SiON gate dielectrics understood through disorder-controlled kinetics," *Microelectronic Eng.*, **80**, 122 (2005).

23. G. A. Scoggan and T. P. Ma, "Effects of electron-beam radiation on MOS structures as influences by the silicon dopant," *J. Appl. Phys.*, **48**, pp.294-300 (1977).

24. J. F. Zhang, S. Taylor, and W. Eccleston, "Electron trap generation in thermally grown SiO_2 under Fowler-Nordheim stress," *J. Appl. Phys.*, **71**, pp. 725-734 (1992).

25. H. S. Witham and P. M. Lenahan, "Nature of the E' deep hole trap in metal-oxide-semiconductor oxides," *Appl. Phys. Lett.*, **51**, pp. 1007-1009 (1987).

26. T. Yang, M. F. Li, C, Shen, C. H. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu, and D. L. Kwong, "Fast and slow dynamic NBTI components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application," in *VLSI Symp. Tech. Dig.*, pp. 92-93 (2005).

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

28. 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 Dev.*, **56**, pp. 1086-1093 (2009).

4 | Effective threshold voltage shift and its use for lifetime prediction

4.1   Introduction

As described in section 2.3, one of the most common threshold voltage measurement technique is the linear extrapolation method with the drain current measured as a function of gate voltage at a low drain voltage of typically 50~100mV to ensure operation in the linear region [1-4]. One problem is that the threshold voltage shift measured by this extrapolation method, ΔVth(ex), cannot be used to evaluate the degradation of drain current under operation condition. Fig.4.1 shows the difference between the measured drain current degradation and that calculated from ΔVth(ex) (see Fig.4.1). The degradation of drain current is considerably underestimated.

The potential causes for this underestimation will be analyzed in section 4.2. One of the objectives of this chapter is to find a simple method for evaluating ΔVth, that can be used to evaluate the drain current degradation under operation conditions. Section 4.3 will show that this can be achieved by using the effective threshold voltage shift, ΔVeff.

After evaluating ΔVeff, attention will be focused on how it can be used to predict the device lifetime. Predicting NBTI lifetime can be dangerous since it is difficult to assess its safety margin. The common technique uses gate bias Vg acceleration to reduce the test time. Although this technique has been widely used when ΔVth is obtained from quasi-dc measurements [5-8], its applicability to the UFP ΔVth has

not been investigated. It has been widely reported that substantial recovery occurs during the quasi-dc measurement, so that the quasi-DC ΔVth is considerably smaller than the UFP ΔVth [1-4].
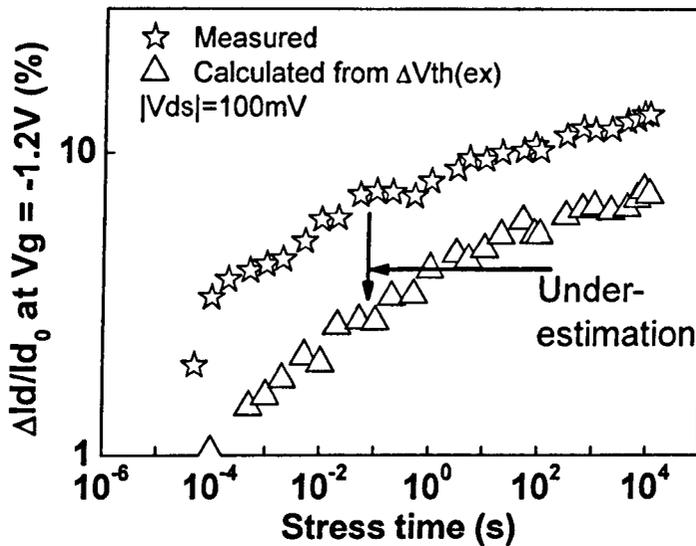
Fig.4.1 A comparison of the measured drain current degradation under Vg=-1.2V with that estimated from ΔVth(ex). The ΔVth(ex) clearly underestimated the degradation of drain current.

To convert from ΔVth(ex) to $\Delta I_d / I_d$ current degradation, the following equation is used

$$\frac{\Delta I_d}{I_{d0}} = \frac{\Delta Vth_{(ex)}}{V_g \times \Delta Vth_{0(ex)}}$$

In section 4.4, it will be shown that the traditional Vg acceleration technique cannot be used to predict device lifetime, once the recovery is suppressed by the UFP technique. A new kinetic model for the UFP ΔVeff will be developed in section 4.5 and, based on that, a new single test lifetime prediction technique will be proposed in section 4.6. Section 4.7 will show the applicability of this single test lifetime

prediction technique to samples fabricated by different processes and estimate the error margin. Finally, conclusions will be drawn in section 4.8.

The MOSFETs used in this chapter were manufactured at interuniversity Microelectronics Research Centre (IMEC), Belgium. In order to test the applicability of the proposed NBTI dinetics and lifetime prediction method, samples from six different processes were sued, as shown in Table 4.1. The samples with silicon oxynitride (SiON) as the gate oxide have p+ poly-Si gate and four of them were plasma nitride for different time and one was thermally nitride. One HfSiON/SiON stack was also prepared by ALCVD with 80% Hf and a TiN gate.

Wafter No.	Gate material	Gate dielectrics	EOT (nm)
A	p$^+$ poly-Si	12s plasma SiON	1.85
B	p$^+$ poly-Si	Plasma SiON	1.4
C	p$^+$ poly-Si	Thermal nitried SiON	2.7
D	p$^+$ poly-Si	45sec plasma nitrided SiON	2.0
E	p$^+$ poly-Si	20sec plasma nitrided SiON	2.0
F	TiN	2nm ALCVD HfSiON and 1nm SiON	1.53

Table 4.1 Samples used in this chapter

4.2   Potential causes for the underestimation of ΔId/Id by ΔVth(ex)

There are two possible causes for the underestimation of drain current degradation shown in Fig.4.1 by ΔVth(ex): mobility degradation [9,10] and the impact of sensing Vg effect on ΔVth [1-3].

Since ΔVth(ex) is evaluated near to Id=0 and the number of charge carriers in the channel is low, it is generally believed that ΔVth(ex) mainly originates from charges both in the gate dielectrics and at the oxide/substrate interface [1-4] and the contribution from mobility degradation is insignificant. Under operation condition, however, the number of charge carriers in the conduction channel are substantial and the impact of mobility degradation on ΔId/Id may be considerable [9,10]. Since ΔVth(ex) hardly takes the mobility degradation into account, it may underestimate the current degradation. The evaluation of mobility degradation, $\Delta\mu$, however, can be problematic, controversial and undesirable for test engineers, as detailed below.

To evaluate mobility, the split capacitance-voltage (CV) characteristic was traditionally used [10-12]. One problem is that CV measurement does not separate the inversion charge carriers captured by interface states from mobile charge carriers, since it is difficult to measure interface states near the band edges [13]. Another problem is that CV measurements sometimes take seconds [14-16] during which substantial NBTI recovery occurs [1-4,17,18]. A fast method is to measure the slope of Id/gm$^{0.5}$ versus Vg-Vth [9], where gm is transconductance. However, Id/gm$^{0.5}$ versus Vg-Vth is not always linear [19]. This method is of limited use, since it only gives the low field mobility.

Another method for estimating mobility is using the transfer characteristics, Id-Vg, measured at low drain bias, Vd [11]. It, however, requires knowing the threshold voltage, Vth. It has been shown that mobility evaluated in this way can be substantially underestimated if |Vg-Vth| is overestimated [20]. It is proposed that Coulombic scattering from the NBTI-induced charges is responsible for mobility

degradation, but this is controversial [21,22]. Moreover, NBTI-induced device lifetime was typically defined as the stress time for ΔV_{th} to reach a pre-specified value [5-8]. Introducing additional parameters like $\Delta\mu$ will make such definition inapplicable and one must model both ΔV_{th} and $\Delta\mu$ now. This is undesirable.

The second potential cause for the underestimation of $\Delta I_d/I_d$ by $\Delta V_{th}(ex)$ is the impact of sensing V_g on threshold voltage shift. Recently, it has been pointed out that the threshold voltage shift evaluated by the extrapolation method, $\Delta V_{th}(ex)$, underestimates the real threshold voltage shift under the operation condition, because ΔV_{th} increases with the sensing V_g [1-3] (see Fig.4.5). The sensing $|V_g|$ used by the extrapolation method is around V_{th}, which is typically around 0.3V (see Fig. 4.3 (a)) and lower than the operation voltage of 1.2V.

In this chapter, attempts will not be made to separate the potential contributions of each of the two causes to the underestimation. Instead, efforts will be made to find an effective parameter that can take into account the full degradation whatever are causes.

4.3   Effective threshold voltage shift

4.3.1   The concept of effective parameters

When the quantity, distribution, or origin of a parameter is not known, the concept of "effective" or "equivalent" parameter is widely used. The "effective" parameter should have two properties: it has the same impact on the electrical performance of devices as the real parameter and it can be readily evaluated.

One example is the "equivalent oxide thickness (EOT)" for characterizing high-k/SiON stack [16,20]. Although the real dielectric constant can have complex distribution across the stack, one can assume it being the same as SiO_2 and readily measure the EOT from the capacitance at strong inversion, so that the difficulty in determining its spatial distribution is avoided.

Another example is the popular "effective density" for the trapped charges in gate dielectric [23]. Generally, we do not know the spatial distribution of trapped charges and its determination is difficult. In most cases, however, what is important is their impact on device performance. By assuming that they are at the SiON/substrate interface, we can determine their effective density from midgap voltage shift [24].

In the present case, the individual contribution from the ΔVth(ex) and the $\Delta\mu$ to ΔId/Id is not known. The challenge is how to find a parameter that includes the effects of both ΔVth(ex) and $\Delta\mu$ on ΔId, without actually evaluating $\Delta\mu$. In the next section, such a parameter will be selected.

4.3.2   Selection of the effective parameter

Two parameters have been used to characterize NBTI: ΔVth(ex) [1-3] and ΔId/Id [10]. The problem with ΔId/Id is its dependence on device size, W/L, and source/drain resistance, Rsd. Under a low Vd, one can use,

$$Id = (CoxW / L) \times \mu \times (Vg - Vth) \times (Vd - IdRsd) \qquad (1)$$

This leads to,

$$\frac{\Delta Id}{Id} = \frac{1}{1+(CoxW\ /\ L)\mu(Vg-Vth)Rsd}[\frac{\Delta\mu}{\mu}-\frac{\Delta Vth}{(Vg-Vth)}]\qquad(2)$$

Eq. (2) predicts a reduction of $\Delta Id/Id$ for higher $RsdW/L$. To test it, Rsd is purposely increased by externally connecting a resistor to the drain. Fig.4.2(a) confirms the reduction of $\Delta Id/Id$ for larger Rsd and, consequently, $\Delta Id/Id$ obtained from one device cannot be used to predict $\Delta Id/Id$ for a device of different $RsdW/L$. Since $RsdW/L$ can vary substantially [25], $\Delta Id/Id$ is not a preferred parameter.

Eq. (2) shows that both $\Delta\mu$ and ΔVth contribute to ΔId. As described in the introduction, evaluating $\Delta\mu$ is problematic and controversial. It is therefore desirable to avoid it by attributing its potential effect on ΔId to an additional threshold voltage shift. Central to this chapter is to propose using an "Effective threshold voltage shift", $\Delta Veff$, which has the same effect on ΔId as the combined effect from $\Delta\mu$ and $\Delta Vth(ex)$. Figs.4.2 (b) & (c) shows that such a $\Delta Veff$ is independent of Rsd, so that it has a clear advantage over $\Delta Id/Id$. The extraction of $\Delta Veff$ from standard NBTI tests will be addressed next.
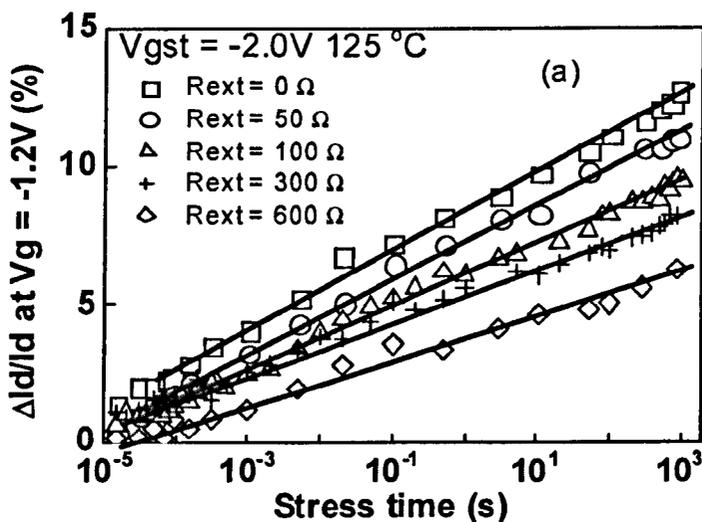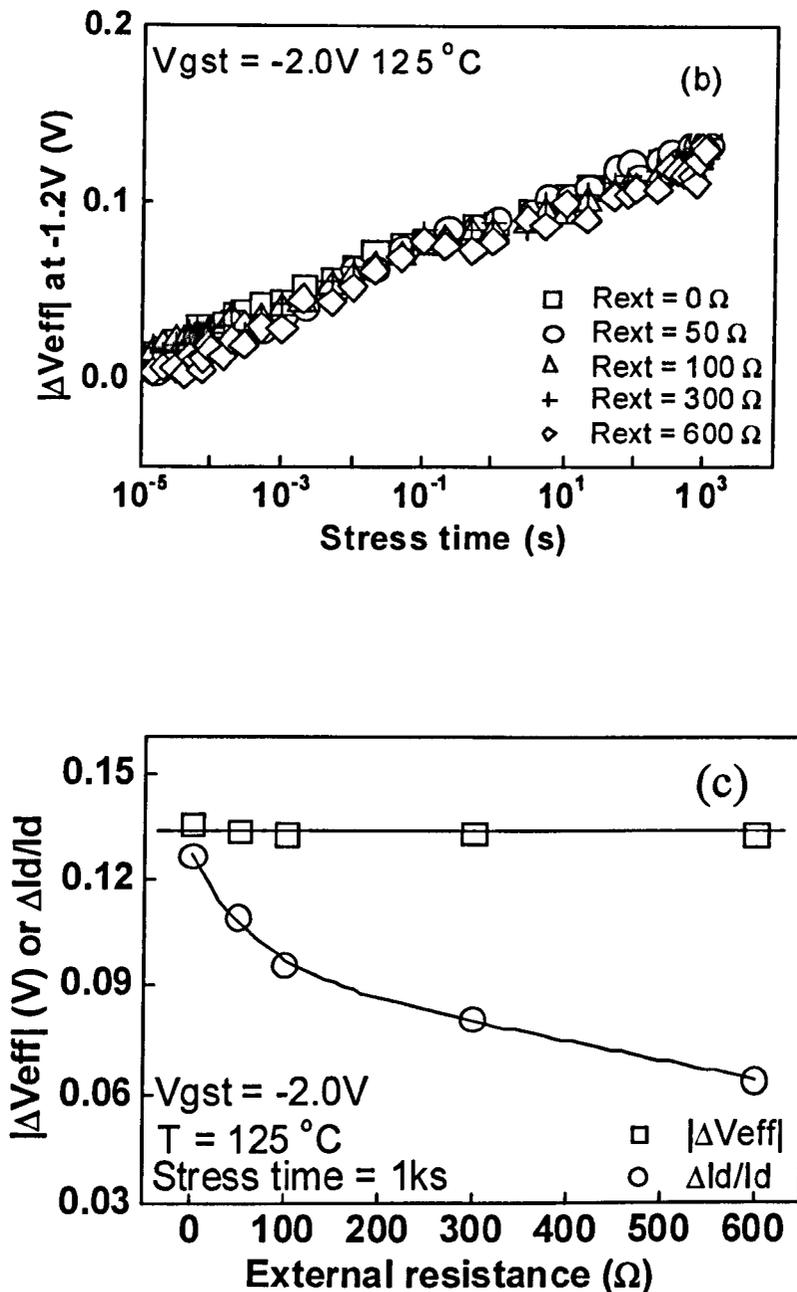
Fig.4.2 Effect of source/drain series resistance, Rsd, on ΔId/Id (a) and ΔVeff (b). Rsd was increased by externally connecting a resistor, Rext. The ΔVeff and ΔId/Id were evaluated from the same set Id-Vg at Vg= -1.2 V. ΔId/Id is sensitive to Rsd, but ΔVeff is not (c).

4.3.3   Evaluation of the effective threshold voltage shift

If we use subscript '0' and '1' to represent parameters before and after stress, at the same current, namely $Id_0 = Id_1$, we have,

$$\mu_0 \times (Vg_0 - Vth_0) = \mu_1 \times (Vg_1 - Vth_1).$$

To attribute the effect of a varying μ on Id to a change in Vth, we assume that μ is a constant for a given (Vg-Vth), so that $\mu_0 = \mu_1$. This leads to,

$$\Delta Veff = Vth_1 - Vth_0 = Vg_1 - Vg_0. \qquad (3)$$

$\Delta Veff$ can be evaluated from the gate bias shift under a constant Id, therefore.
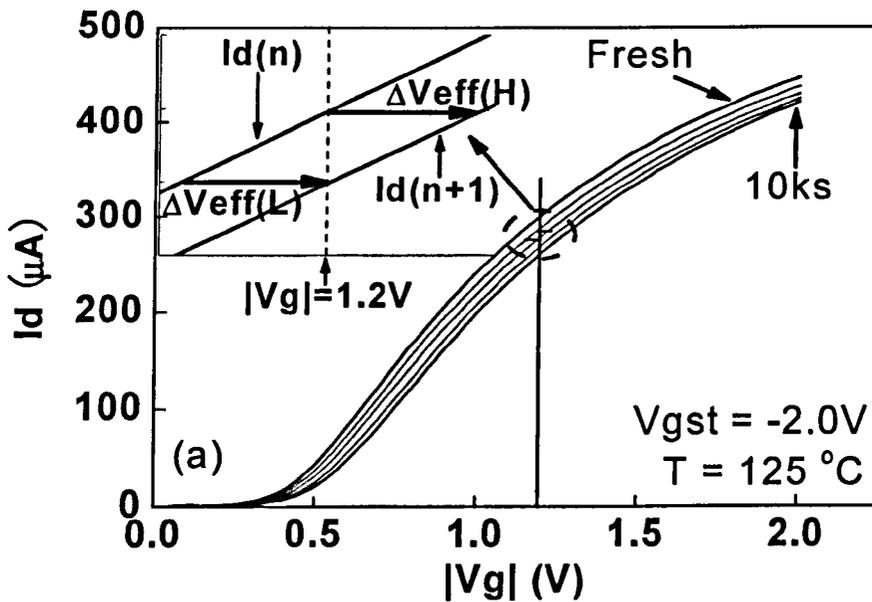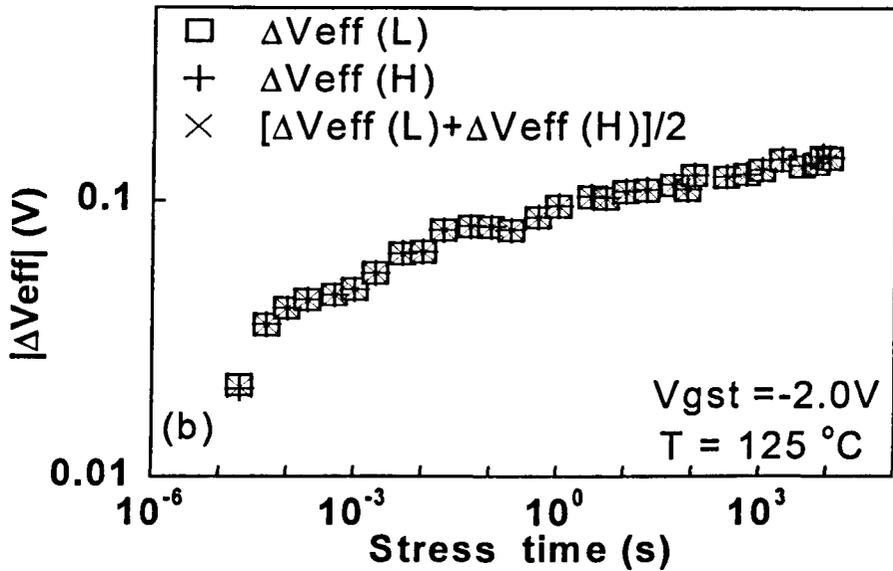
Fig.4.3 (a) Evaluation of ΔVeff between the measurement point 'n' and 'n+1' under a sensing $|Vg|$=1.2 V. ΔVeff(H) and ΔVeff(L) represents ΔVeff measured at Id(n) and Id(n+1), respectively. (b) The accumulative ΔVeff against stress time.

A typical result is shown in Fig.4.3(a) for evaluating ΔVeff between two measurement points 'n' and 'n+1'. Under an operation gate bias, say Vg= -1.2V, one can use either Id(n) or Id(n+1) as the constant current level for measuring ΔVeff, as represented by ΔVeff(H) and ΔVeff(L), respectively in Fig.4.3(a). Fig.4.3(b) shows the accumulative ΔVeff against stress time and the difference in ΔVeff(H) and ΔVeff(L) is insignificant. On the time step size, it is found that one point per decade of stress time is adequate, as shown in Fig.4.4. Eq. (3) also confirms that ΔVeff does not depend on RsdW/L, in contrast to ΔId/Id given by eq. (2).
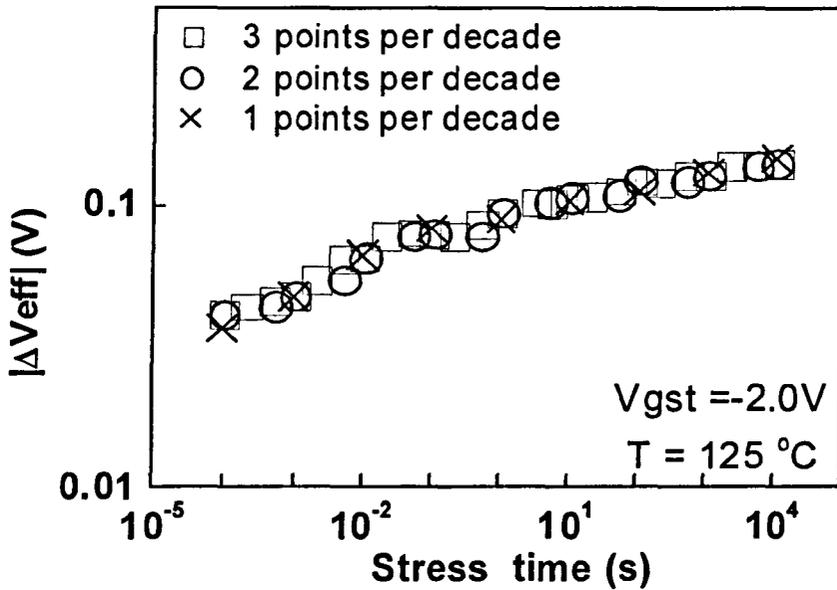
Fig.4.4 Effects of number of points per decade of stress time on ΔVeff evaluated at Vg=-1.2V.

Fig.4.5 shows the dependence of ΔVeff on the sensing Vg. $|\Delta$Veff$|$ clearly increases with $|$Vg$|$ and $|\Delta$Veff$|$ at Vg=-1.2V is substantially higher than the conventional $|\Delta$Vth(ex)$|$.
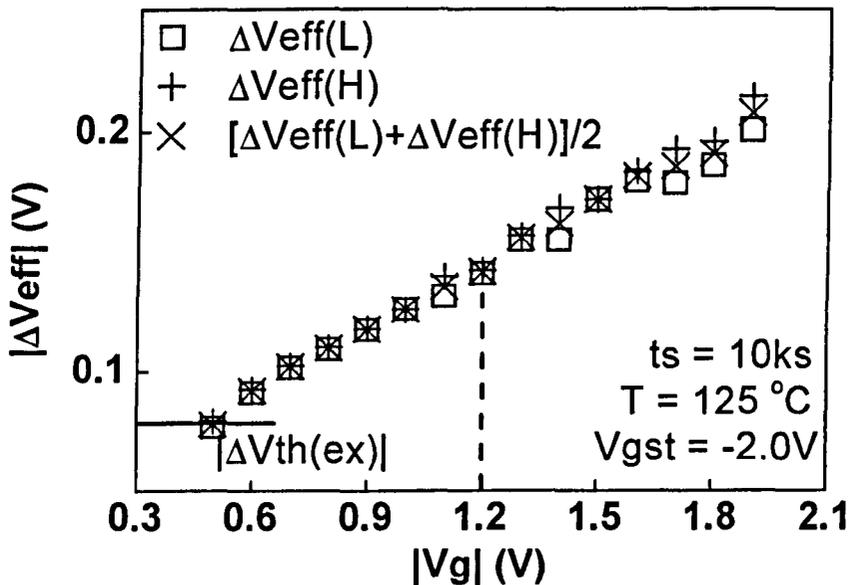
Fig.4.5 Increase of $|\Delta$Veff$|$ with sensing $|$Vg$|$. Under an operation bias, say $|$Vg$|$=1.2V, $|\Delta$Veff$|$ is substantially higher than the conventional $|\Delta$Vth(ex)$|$.

4.3.4   Justification of the effective threshold voltage shift

The objective for introducing ΔVeff is to characterize Id degradation without evaluating $\Delta\mu$. We now show that ΔVeff can be such a parameter. The Id-Vg measured on a fresh device is used to determine the un-degraded mobility, μ_0, first. Both μ_0 and ΔVeff is then used to determine ΔId/Id. Fig.4.6 confirms that the ΔId/Id calculated from ΔVeff agrees well with the measured one.

Fig.4.6 also shows that if ΔVeff is replaced by the on-the-fly (OTF) technique [1], ΔId/Id at Vg=-1.2V is clearly overestimated. The OTF uses the stress Vgst as the sensing Vg and Vgst is typically higher than the operation |Vg|=1.2V. As a result, |ΔVth(OTF)|>|ΔVeff| leads to the overestimation of ΔId/Id.
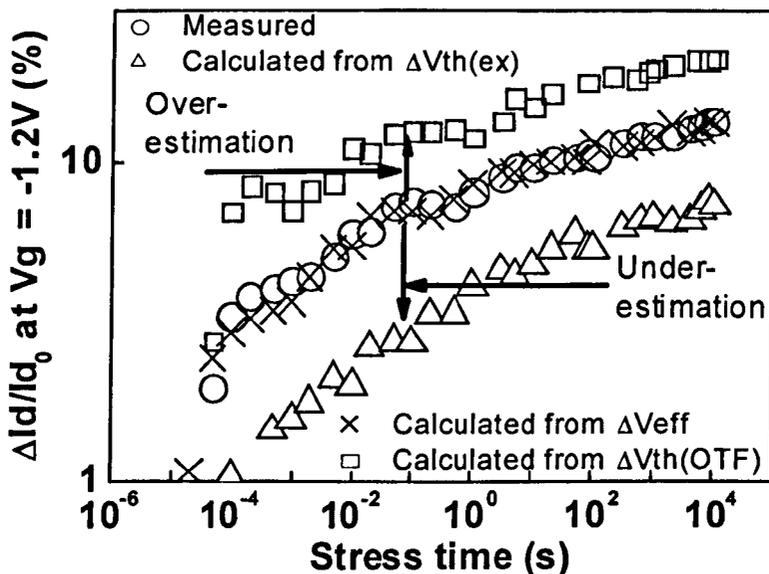
Fig.4.6 A comparison of measured Id degradation at Vg=-1.2V with that predicted from ΔVeff, ΔVth(ex) and ΔVth(OTF). The stress was at Vgst= -2V.

4.4   An analysis of the traditional lifetime prediction technique

One purpose for monitoring and evaluating the degradation is to predict device lifetime under real operation conditions. After evaluating effective threshold voltage shift, efforts will be made to use it for predicting the device lifetime. Specifically, the worst-case lifetime will be predicted, where the recovery of degradation will be suppressed by using the ultra-fast pulse measurement described in Chapter 2. The traditional Vg acceleration technique will be briefly reviewed first and its applicability to the worst-case lifetime prediction will then be investigated.

4.4.1   The traditional Vg acceleration technique for lifetime prediction

The NBTI lifetime is typically defined as the time for the threshold voltage shift, ΔV_{th}, to reach a preset level [5-8]. Under an operation gate bias, V_{gop}, the required lifetime is 10 years and the degradation under V_{gop} can be too low to be measured reliably within a practical stress time. To predict the lifetime under a V_{gop}, multiple accelerated tests are carried out with stress biases, V_{gst}, higher than V_{gop}. The accelerated lifetime is typically fitted with $|V_{gst}|^{-\alpha}$ [5,6] or $\exp(-|V_{gst}|)$ [7,8] and then extrapolated to V_{gop}, so that the lifetime under V_{gop} can be estimated.

To demonstrate the traditional lifetime prediction technique, the $|V_{gst}|^{-\alpha}$ acceleration will be used as an example in the following. The $|V_{gst}|^{-\alpha}$ acceleration requires:

$$|\Delta V_{th}| = B\,|V_{gst}|^m\,t^n \tag{4}$$

where 'B' is a constant for a given temperature. At a device lifetime of t=τ, ΔV_{th}

reaches the specified $\Delta V_{th}(\tau)$ and we have,

$$\log(\tau) = \frac{1}{n}\log[\frac{|\Delta V_{th}(\tau)|}{B}] - \frac{m}{n}\log(|V_{gst}|) \tag{5}$$

The equation (5) requires the $\log|\Delta V_{th}|\sim\log(t)$ being shifted in parallel for different

V_{gst} and the power factor against time, 'n', being insensitive to V_{gst}, so that $\log(\tau)$ is a

straight line against $\log(|V_{gst}|)$. For the conventional ΔV_{th} measured by extrapolating

the quasi-DC I_d-V_g, Figs.4.7(a) & (b) show that these requirements can be met, so

that the prediction agrees well with measurement in Fig.4.7 (c).
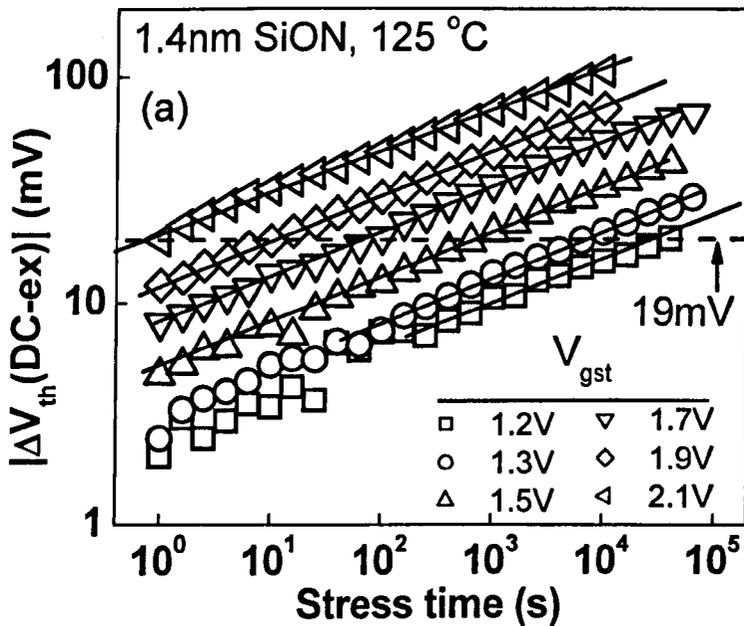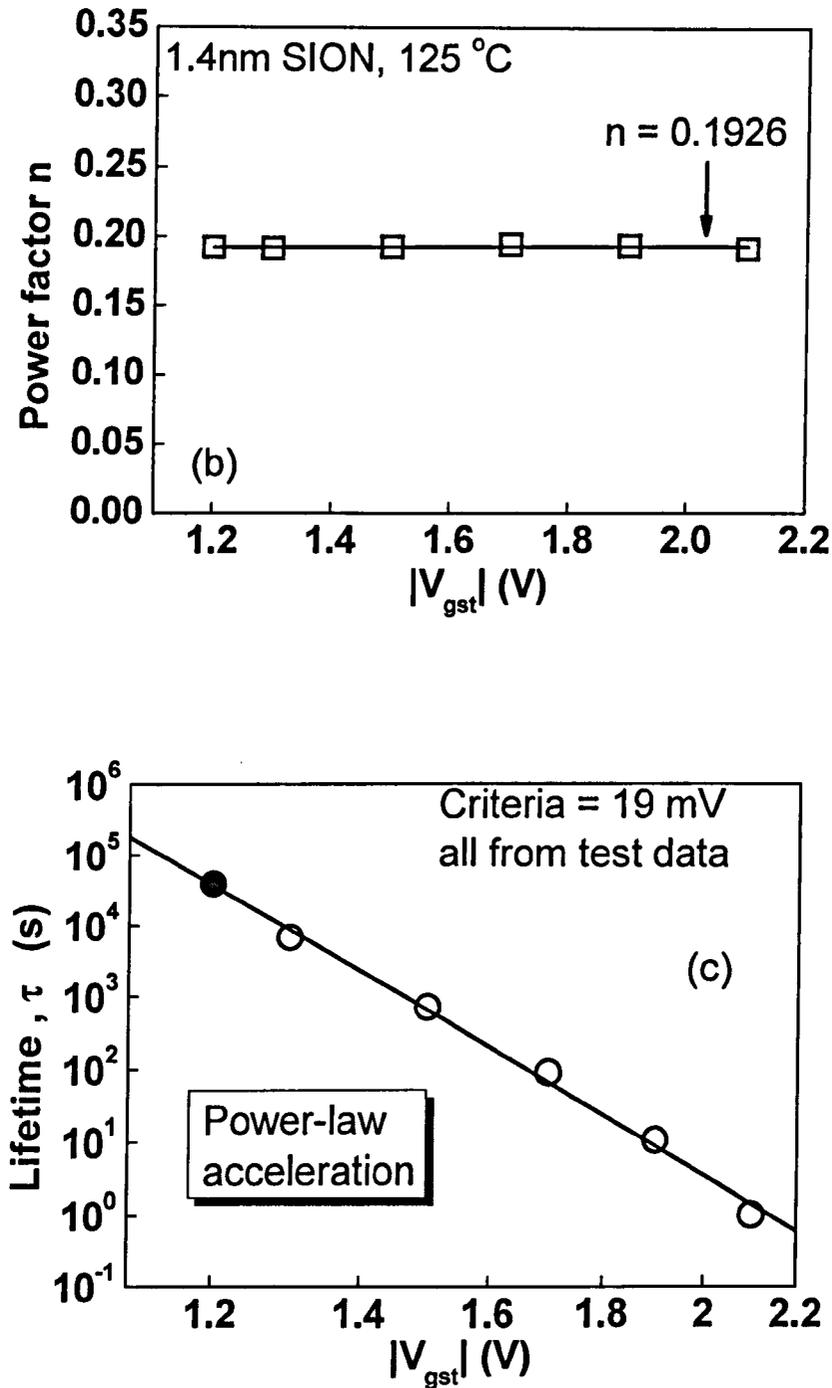
Fig.4.7 Lifetime prediction by V_g acceleration technique for the conventional ΔV_{th}(DC-ex) measured by extrapolating the quasi-DC I_d-V_g. (a) shows that ΔV_{th}(DC-ex) follows the power law. The horizontal dashed line represents $|\Delta V_{th}|$= 19 mV that is used to define the lifetime. (b) shows that the power factor is insensitive to the stress bias. The solid line represents the average power factor of n=0.1926. In this case, (c) shows that the prediction (solid line) agrees with the measurement (symbol '●').

4.4.2   The inapplicability of traditional Vg acceleration technique for the worst-case lifetime prediction

Fig.4.8(a) shows the typical NBTI kinetics when ΔVeff is evaluated from the UFP measurement and the recovery is suppressed. The traditional Vg acceleration method will be used for this set of data to investigate its applicability. When the Vg acceleration technique is used for predicting device lifetime, its safety margin is generally not known. To study its applicability to the set of data in Fig.4.8(a), however, one must assess the prediction safety margin. This requires directly measuring ΔVeff at $V_{gst}=V_{gop}$. As V_{gop} does not reduce proportionally with the SiON thickness, the oxide field during device operation has increased to such a level that Fig.4.8(a) shows that the ΔVeff at $V_{gop}=-1.2V$ can now be reliably measured. This allows comparing the measured stress time for ΔVeff to reach a given level under $V_{gst}=V_{gop}$ with that predicted by using V_g acceleration, so that the safety margin of prediction can be estimated.

In Fig.4.8(a), the last measured $|\Delta$Veff$|$ reached 60 mV under $V_{gst}=V_{gop}$ and if one uses $|\Delta$Veff$|$=60 mV to define lifetime, the stress time for this last point will be the lifetime under $V_{gop}=-1.2V$. This measured lifetime is compared with the prediction based on V_g acceleration in Figs.4.8(b) and 4.8(c). Two popular V_g acceleration methods have been used in literature: $|V_{gst}|^{-\alpha}$ [5,6] and $\exp(-|V_{gst}|)$ [7,8]. Figs.4.8(b) and 4.8(c) show that there is a substantial difference between the predicted and measured time and V_g acceleration cannot be used to predict lifetime in our case, therefore.
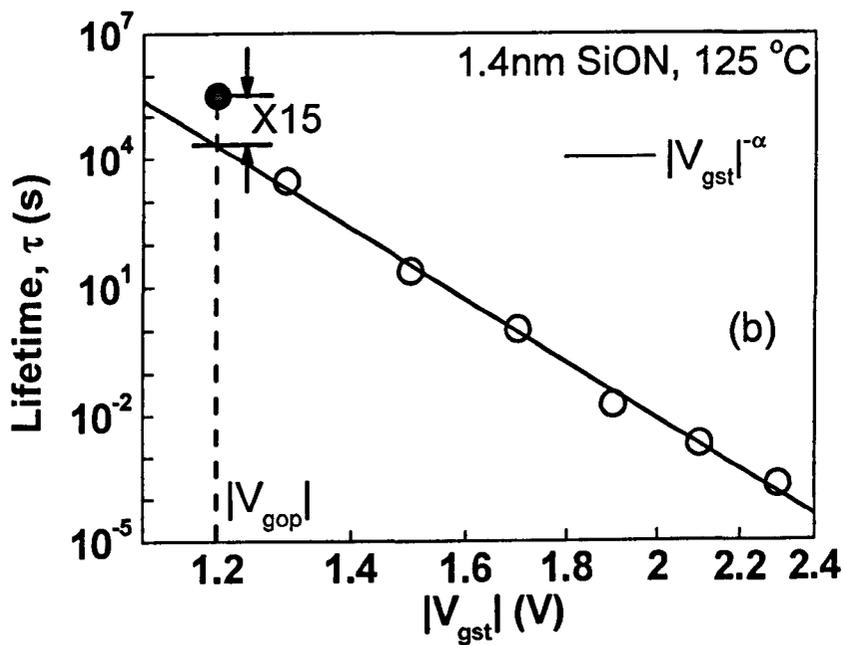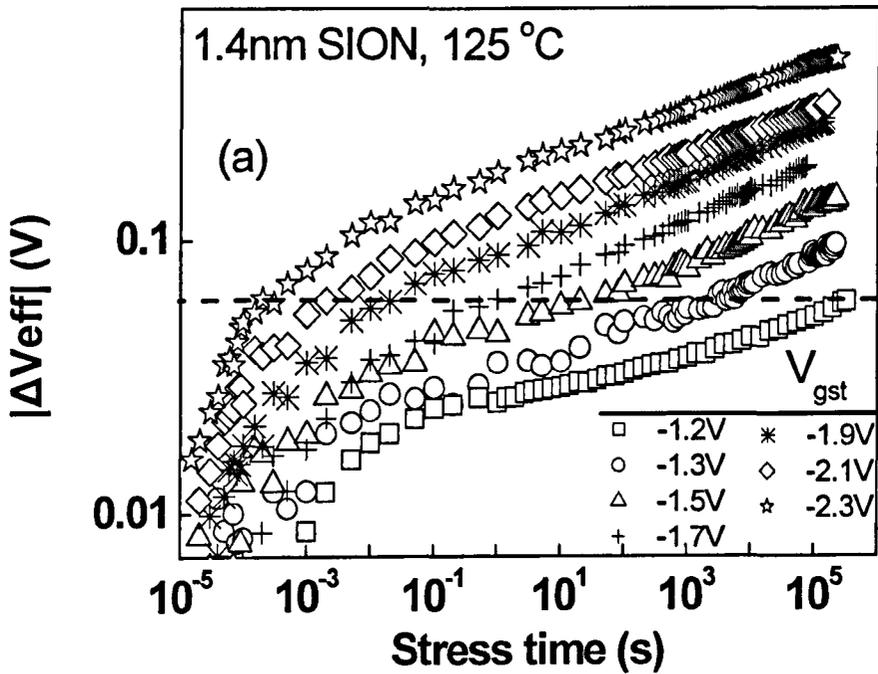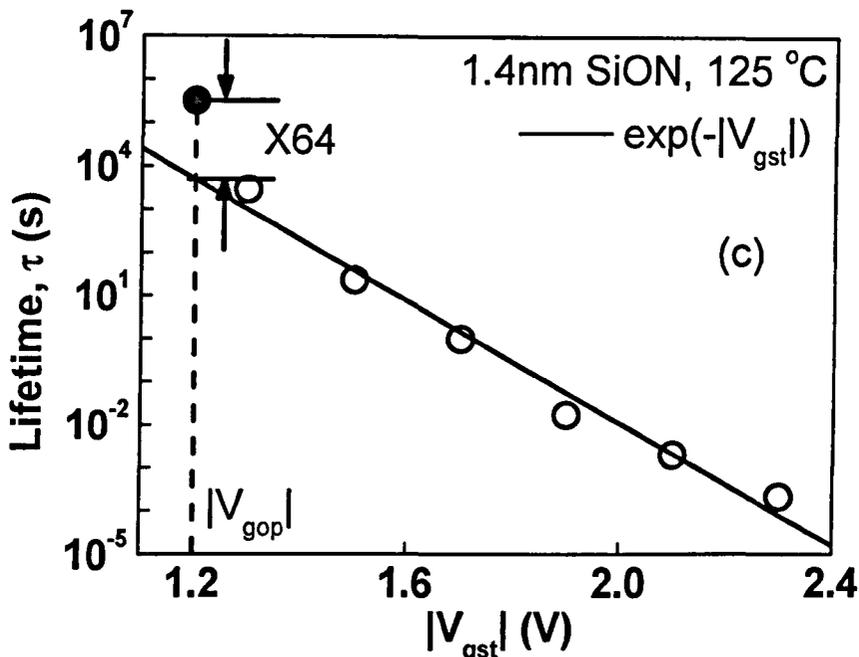
Fig.4.8 Inapplicability of V_g acceleration technique for predicting NBTI lifetime at $V_{gop}=-1.2V$ measured by UFP technique. In (a), the dynamic behaviour of $|\Delta Veff|$ under different stress bias, V_{gst}, is compared and the kinetics does not follow a simple power law. The horizontal dashed line represents $|\Delta Veff|= 60$ mV that is used to define the lifetime. In (b) and (c), the measured lifetime is compared with the predicted one based on $|V_{gst}|^{-\alpha}$ and $exp(-|V_{gst}|)$, respectively. The symbol '●' represents the measured lifetime under $|V_{gst}|=1.2V$. The solid line was obtained by fitting the data at higher $|V_{gst}|$. The error in prediction is substantial.

To explore the reason for the inapplicability of the traditional Vg acceleration prediction method to the ΔVeff in Fig.4.8(a), it should be noted that ΔVeff in Fig.4.8(a) no longer follows a simple power law and log|ΔVeff|~log(t) at different V_{gst} is generally not a parallel shift, when the recovery is suppressed and $V_{gop}=-1.2V$ is used as the sensing V_g. A clear example for the non-parallel shift is given in Fig.4.9. The V_{gst} effect can no longer be separated into a 'pre-factor' like that in equation (4) and this explains the inapplicability of V_g acceleration technique to the case where recovery is suppressed.
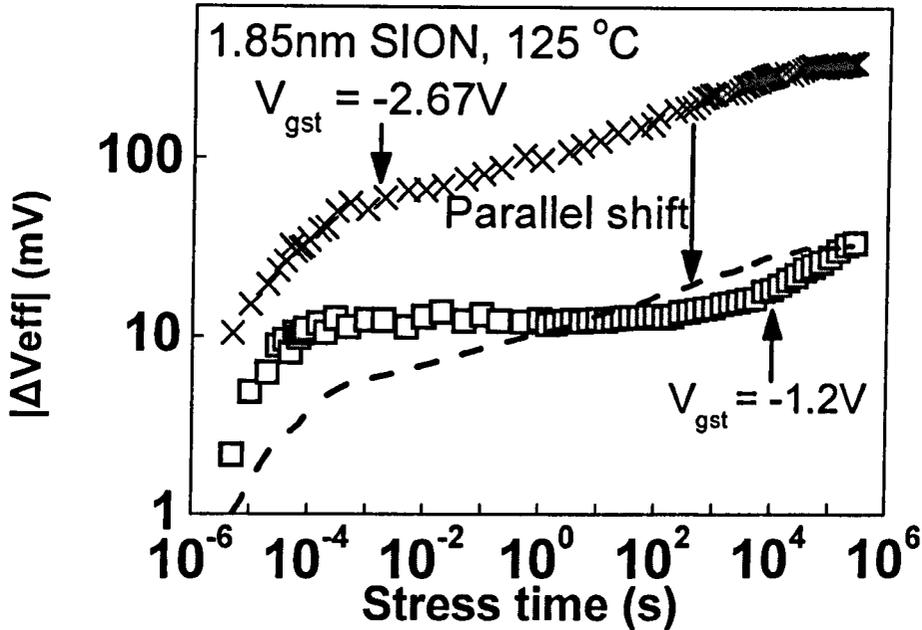
Fig.4.9 The kinetics at different V_{gst} is not shifted in parallel for the UFP ΔVeff sensed at $|V_g|=1.2V$. The dashed curve was obtained by shifting the symbol '×' downward in parallel.

4.5   A new kinetic model for ΔVeff

Since the existing lifetime time prediction technique cannot be applied to the condition where the recovery is suppressed, there is a need for developing new lifetime prediction method for the worst-case scenario. Fig.4.9 shows that the degradation under the operation gate bias $V_{gop}=-1.2V$ is large enough to be measured, once the recovery is suppressed and the sensing $Vg=-1.2V$ is used. In principle, the lifetime of a device can now be predicted from a single test under $Vg=-1.2V$ by extrapolating the stress to the allowed ΔVeff(t=τ) for the lifetime definition, as schematically illustrated by Fig.4.10.
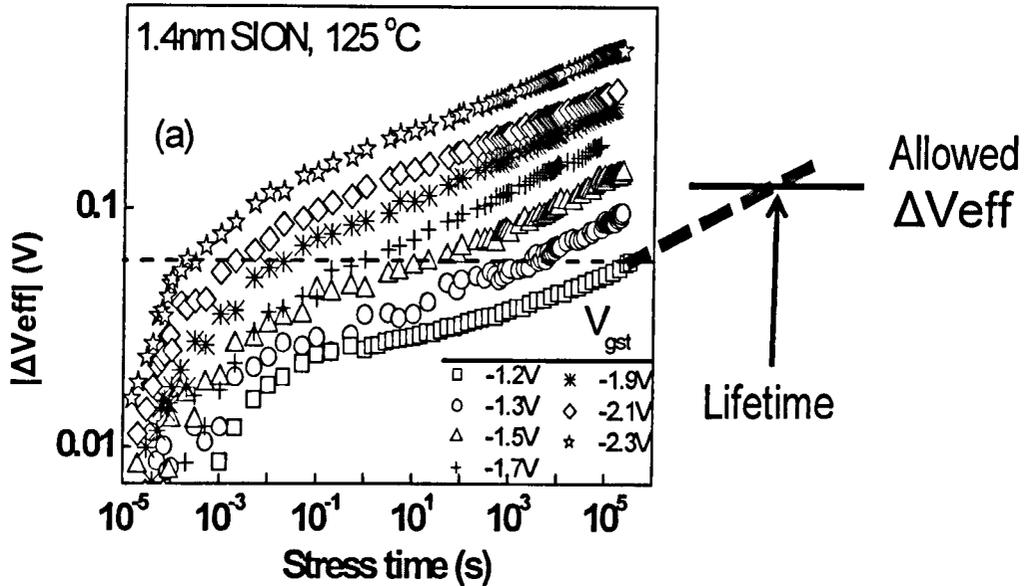
Fig.4.10 A schematic illustration for the single test lifetime prediction technique.

The success of this technique will depend on the availability of a NBTI kinetic model for the extrapolation. When quasi-DC measurement was used, it is well known that the NBTI kinetics is a power law (see Fig.4.7). The data in Fig.4.10, however, clearly do not follow a straight line. The kinetics no longer follows a simple power law against stress time when recovery is suppressed and efforts should be made to develop a model that can describe its dynamic behavior. For the ΔV_{eff} under $V_{gst}=V_{gop}=-1.2V$, Fig.4.11 shows that an outstanding feature of the kinetics is the presence of a 'shoulder'. This indicates that there is an initial period when as-grown defects dominate and the saturation of their charging results in the 'shoulder'. At longer stress time, generation of new defects becomes increasingly important and is responsible for the rise above the 'shoulder'.
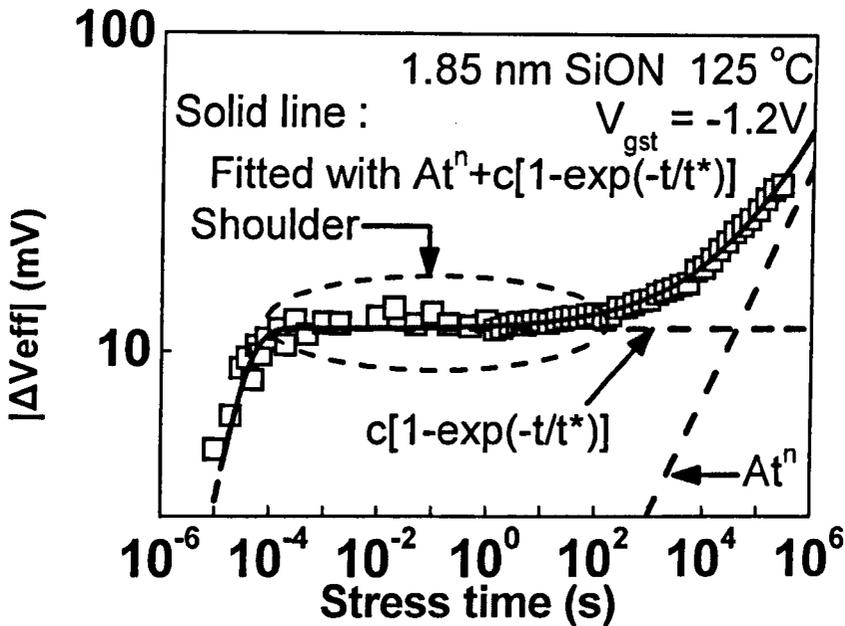
Fig.4.11 The kinetic feature of the UFP ΔVeff sensed at $|V_g|$=1.2V: a "shoulder". By combining the 1st order model for as-grown hole traps with the power law for defect generation, ΔVeff can be fitted over 10 orders of magnitude in time, as shown by the solid line. The dashed lines show that ΔVeff is dominated by as-grown hole traps initially, but the generated defects become important at longer stress time.
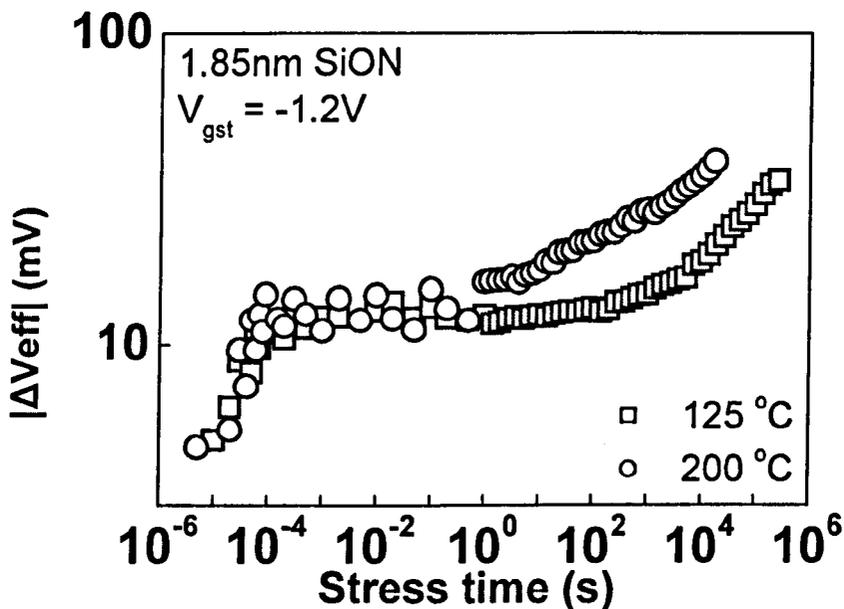
Fig.4.12 Effect of temperature on NBTI kinetics. The height of the shoulder is insensitive to temperature, but the generation of defects above the shoulder is thermally accelerated.

To support the above suggestion, two tests were carried out. In the first test, we checked the effect of temperature on the shoulder height. The saturation level of as-grown defects should be insensitive to temperature [16,27] and if it dominates the shoulder, the shoulder height should be insensitive to temperature. This is confirmed by Fig.4.12. Fig.4.12 also shows that the rise above the shoulder is thermally activated, supporting that defect generation is thermally accelerated.

In the second test, we compare the charging and discharging rate of the defect responsible for the shoulder. As described in Chapter 3, early works [28-30] identified three different types of positive charges in the dielectric: anti-neutralization positive charges (ANPC), cyclic positive charges (CPC), and as-grown hole trapping (AHT). For self-completeness, this framework for positive charges is illustrated by Fig.4.13 again. ANPC has an energy level above the bottom edge of silicon conduction band, Ec, making its discharging more difficult than charging. CPC has an energy level close to Ec and its charging rate is similar to the discharging rate. In contrast, AHT is below the top edge of silicon valence band and that there are far more valence electrons for discharging than hot holes required for charging. As a result, AHT has the signature that discharging is much faster than charging. Fig.4.14 shows that, when the stress time corresponds to the shoulder, the charging and discharging properties of the defect agree with the signature of AHT, supporting that AHT dominates the shoulder.

Vg < 0 : Charging   **Vg > 0 : Discharging**
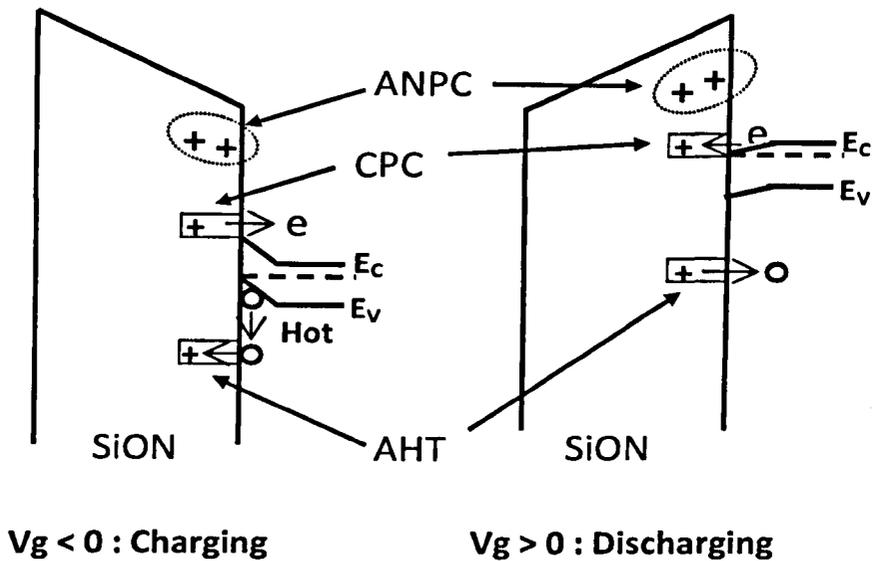
Fig.4.13 Energy band diagram of different types of positive charges. The anti-neutralization positive charges (ANPC) have energy level above Ec, making them difficult to neutralize. The cyclic positive charges (CPC) have energy level near to Ec, resulting in similar charging and discharging rate. The as-grown hole traps have energy level below the top edge of silicon valence band, Ev. Their charging requires hot holes, leading to charging slower than discharging.
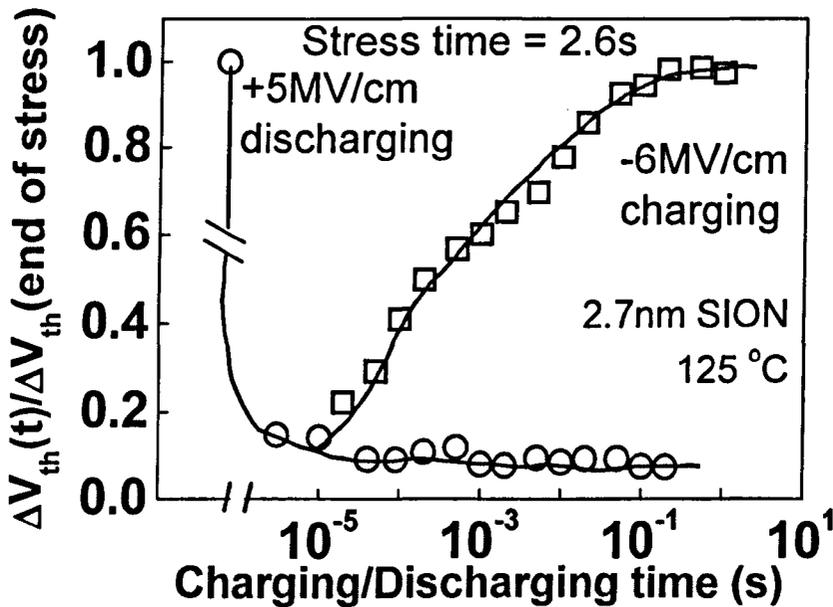
Fig.4.14 A comparison of the charging and discharging rate for the as-grown defects. The stress time is 2.6sec that corresponds to the region where as-grown defects dominate. The discharging under $V_g>0$ is much faster than charging under $V_g<0$: a unique signature of as-grown hole traps (AHT). The solid lines are guides-for-the-eye. It should be noted that the rapid discharging within 5μs observed here was achieved by applying a positive gate bias. For normal NBTI test, however, positive gate bias was not applied and the AHT discharging at V_g=-1.2 V within 5μs was negligible.

On the kinetics, the charging of AHT generally follows the first order reaction model [31,32], while the generation of new defects follows a power law [5,33,34]. By combining these two, we have:

$$\Delta Veff = At^n + c(1 - e^{-t/t^*})$$ (6)

For a given stress temperature and bias, 'A', 'n', 'c', and t* are constants and were obtained by fitting test data with the least square errors and their values are given in Table 4.2. There is only one 'n' for the whole stress period and this 'n' is not the slope of the data in Fig.4.11, namely $n \neq d(\log|\Delta V_{th}|)/d[\log(t)]$. Fig.4.11 shows that equation (6) can model the 'shoulder' and this simple physics-based model can fit the ΔV_{th} over ten orders of stress time. The two dashed lines represent the contribution from AHT and generated defects, respectively. AHT clearly dominates initially, but generated defects become more important for longer stress time. On the nature of degraded defects, the early works [1-4,28-32] show that both interface states and new hole traps are created by stresses. The new hole traps are further separated into anti-neutralization positive charges and cyclic positive charges, each with unique signatures. A detailed discussion, however, is out of the scope of this work.

Table 4.2. The wafers and the fitted parameters at 125 °C

Device number	Gate Dielectrics	A (mV/sec)	n	c (mV)	t* (μs)
A	1.85nm 12s Plasma SiON	0.23	0.36	11.91	30
B	1.4nm Plasma SiON	7.70	0.13	18.39	4390
C	2.7nm Thermal SiON	26.27	0.07	22.07	80
D	2.0nm 45s Plasma SiON	4.10	0.12	7.49	740
E	2.0nm 20s Plasma SiON	3.91	0.12	1.89	110
F	TiN, ALCVD 2.0 nm / 1 nm HfSiON / SiON	12.21	0.14	40.86	30

4.6   A single test lifetime prediction technique

The principle of the single test lifetime prediction technique is already illustrated by

Fig.4.10. The required extrapolation range should be estimated. The affordable test

time is typically in the order of days and the data were used to predict lifetime in years, so that a kinetic model should have the ability to predict at least two decades ahead. To test the prediction ability of a model, the test data in the last two decades are not used for fitting the model and the ΔV_{eff} at the last test point is considered as $\Delta V_{eff}(\tau)$ and the time for the last point is treated as the measured lifetime τ_m, although τ_m is of course not the real device lifetime.

The departure of kinetics from a simple power law was noted in the past and suggestions were made on how lifetime prediction method should be modified to take this departure into account [34,35]. One proposed method is only using the data with a stress time over 10 s to fit the power law against time [34], but there is no information on the prediction accuracy. Fig.4.15 shows that this method can overestimate τ by a factor of 75000.

Another proposed method is to fit $\Delta V_{eff}(t)-\Delta V_{eff}(1\ sec)$ with a power law [35], but Fig.4.16 shows that it underestimates τ by a factor of 10. By applying the present model of equation (6) to the same set of data, Fig.4.17 shows that good agreement is achieved between the measurement and the prediction with $\tau_p/\tau_m=1.03$.
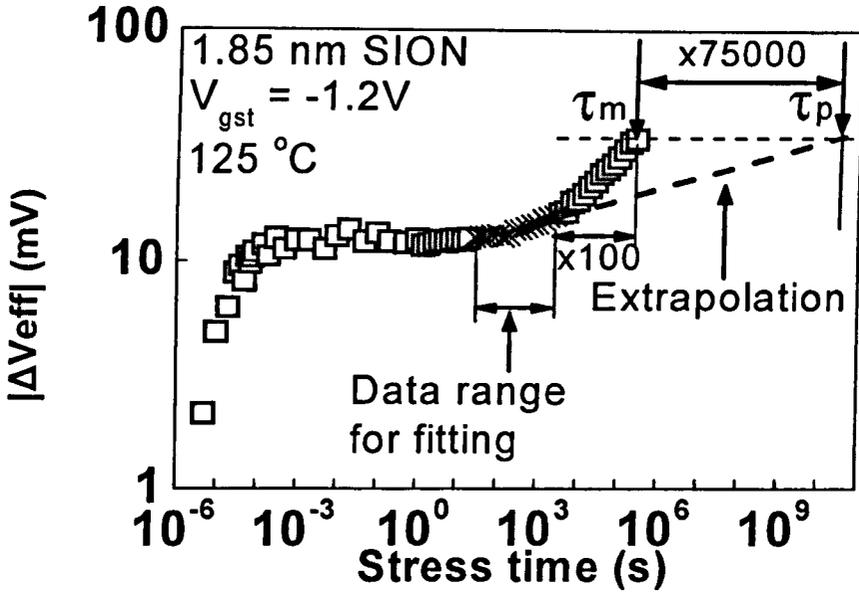
Fig.4.15 Lifetime prediction based on the method proposed in ref. 34. All symbols are test data but only symbol "×" was used for fitting with a power law in the range of 26.8 s < t < 2680 s. The thick dashed line is extrapolated from the fitted line for prediction. τ_m is the time for the last test point and τ_p is the predicted time. τ is over-estimated by a factor of 75000.
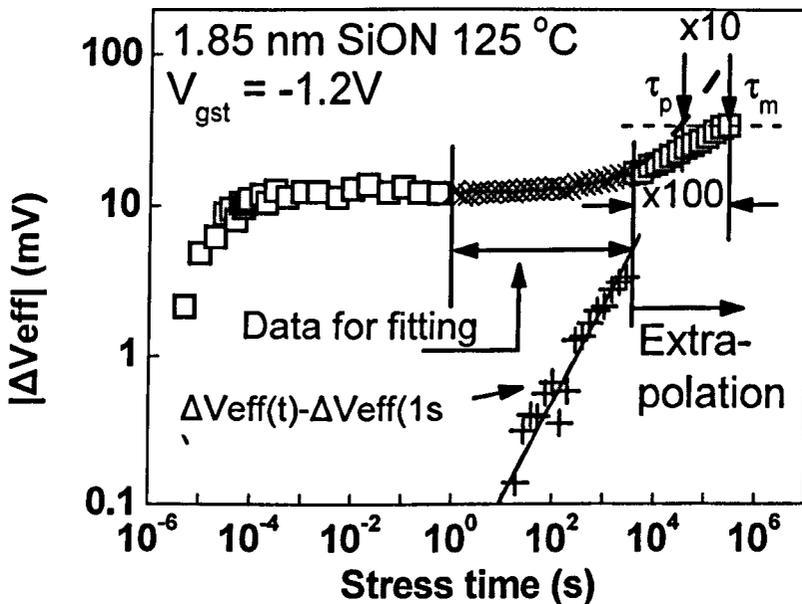
Fig.4.16 Lifetime prediction based on the method proposed in ref. 35. All symbols are test data but only symbol "×" was used for fitting. ΔVth(t) - ΔVth(1 s) (the symbol "+") was fitted with a power law and ΔVth(1 s) was then added back. The dashed line is extrapolated for prediction. τ_m is the time for the last test point and τ_p is the predicted time. This method under-estimates τ by a factor of 10.
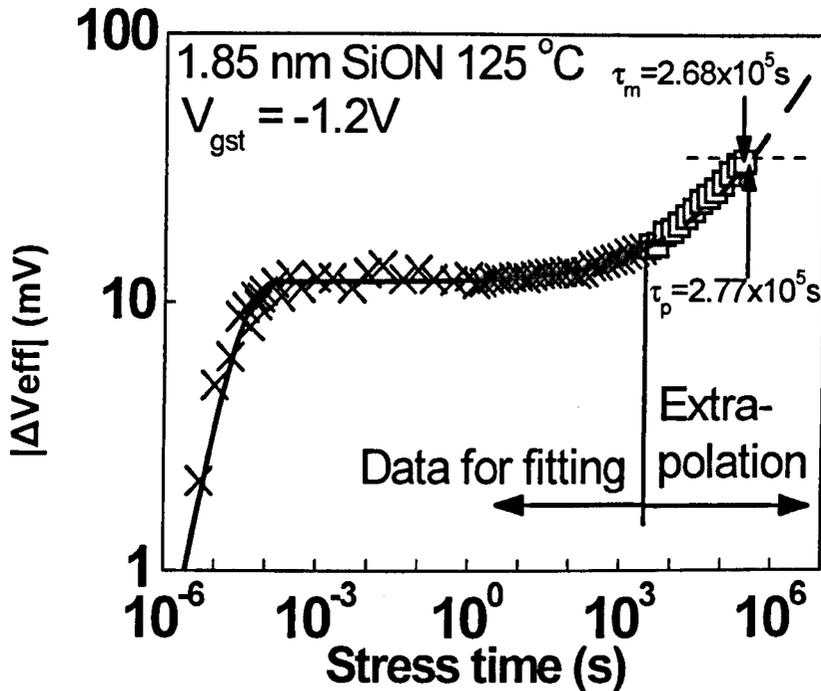
Fig.4.17 Lifetime prediction based on the present model: the equation (6). The test data is the same as those in Figs.4.15 and 4.16 and only symbol "×" was used for fitting. Our model reduces the prediction error substantially and τ_p/τ_m=1.03, here. The test sample is a plasma nitrided 1.85nm SiON.

4.7 Applicability of the single test prediction method for samples fabricated by different processes

Although a good prediction is achieved in Fig.4.17, it is inadequate to demonstrate that a prediction method works for one process. For a prediction technique to be useful, it must be applicable to samples fabricated by a wide range of processes. We now test its applicability for four other SiON layers with different nitrogen concentrations and nitrided either by plasma or thermally. Moreover, an ALCVD HfSiON/SiON stack is also tested and the thickness of these samples is given in Table 4.1. Figs.4.18(a)-(e) shows that ΔVeff follows equation (6) in all cases, although the 'shoulder' in some samples is less apparent. The prediction achieved a

safety margin of 50% or less in all processes tested, making us confident that the single test technique is generally applicable.

Table 4.1 shows that the fitted power factors for different processes have a range of 0.07~0.36, which agrees with the range reported by early works [36-42]. Early works reported that the variation of power factor could come from two sources: different hydrogenous species and different nitrogen densities and distributions. On the hydrogenous species, it was reported that the power factor for H^+ [40], atomic hydrogen [41], and H_2 [42] is 0.5, 0.25 and 0.16, respectively. On the nitridation effect, it was reported that an increase of nitrogen reduces the power factor [36,37]. Moreover, for the same area density of nitrogen, the power factor of a thermally nitrided SiON is typically lower than that of a plasma nitrided SiON [36]. The power factor in Table 4.1 agrees with these trends. The sample A has the lowest nitrogen density and the highest power factor. The sample C was thermally nitrided and has the lowest power factor.

(a) 1.4nm SiON 125 °C, $V_{gst} = -1.2V$. $\tau_m = 3.17 \times 10^5 s$, $\tau_p = 3.08 \times 10^5 s$. Data for fitting and Extrapolation regions of $|\Delta Veff|$ (mV) versus Stress time (s).
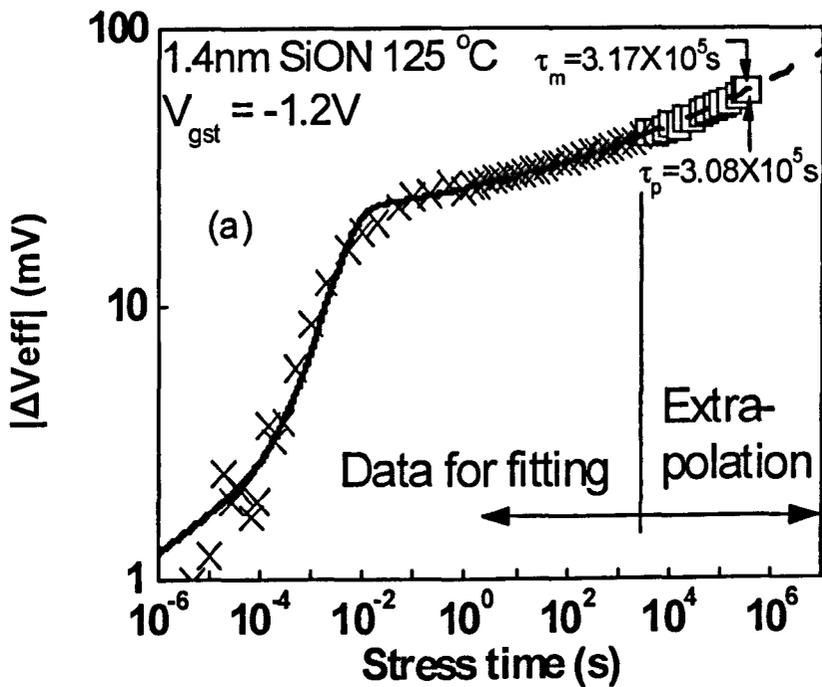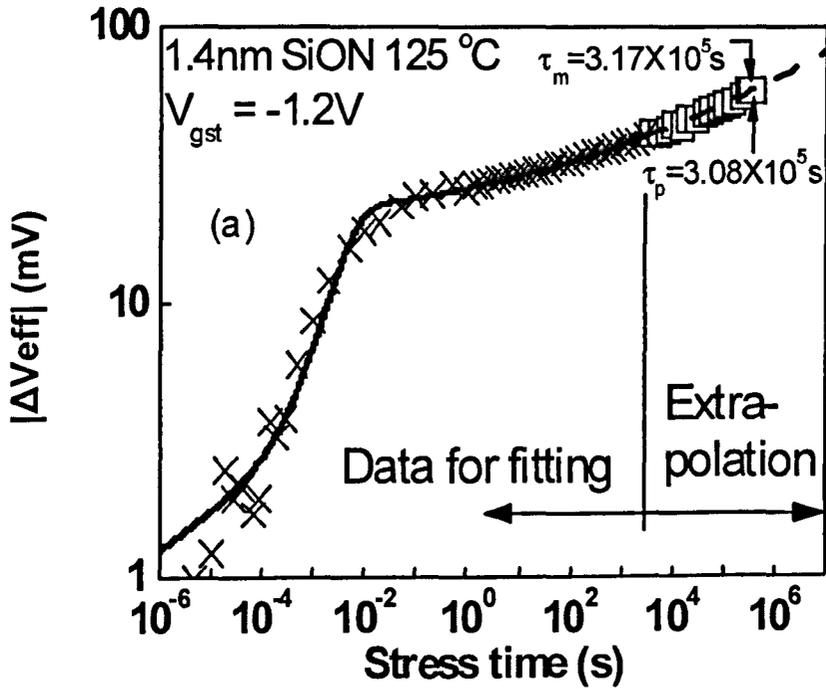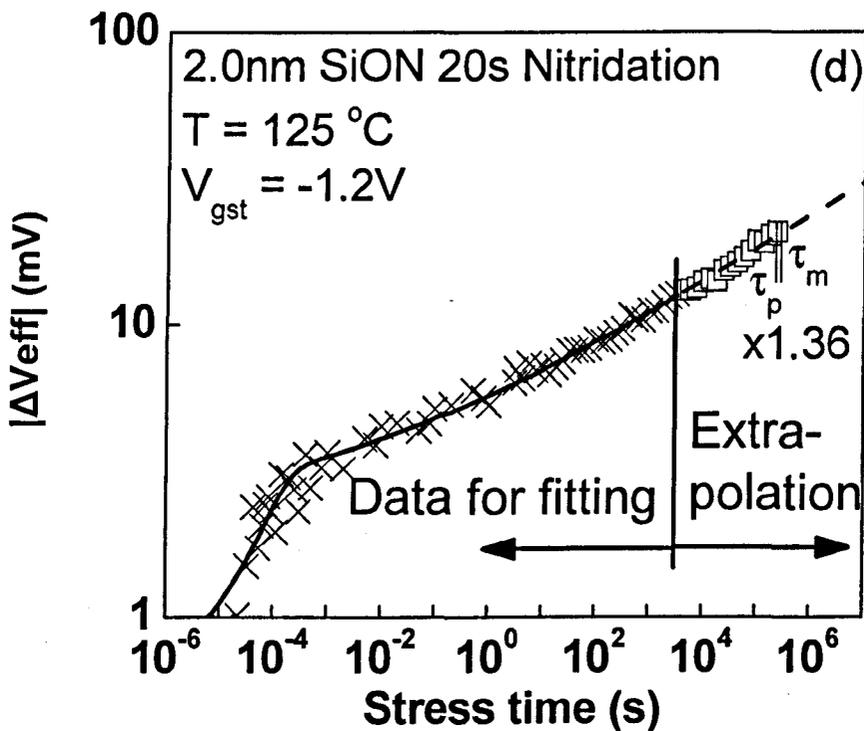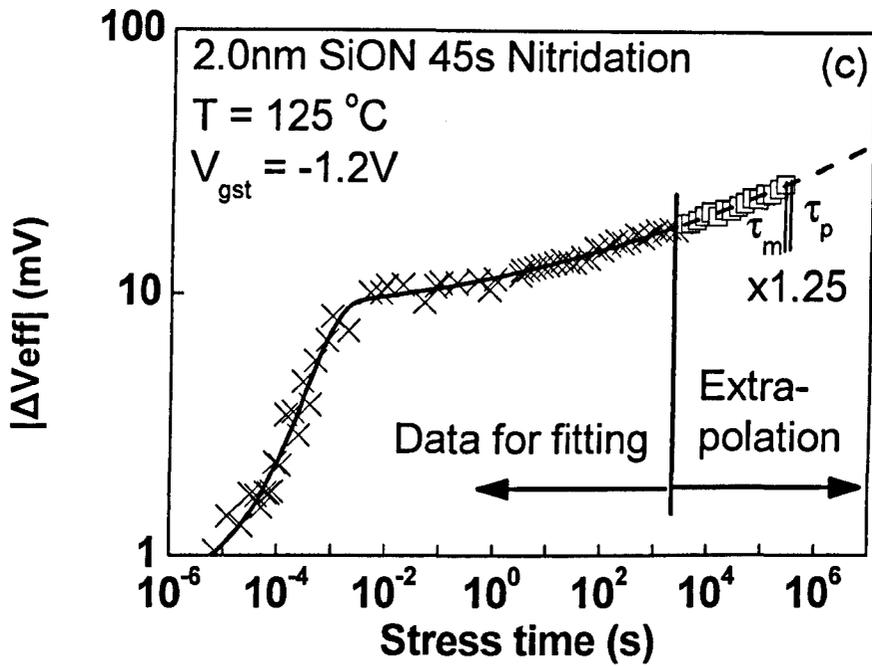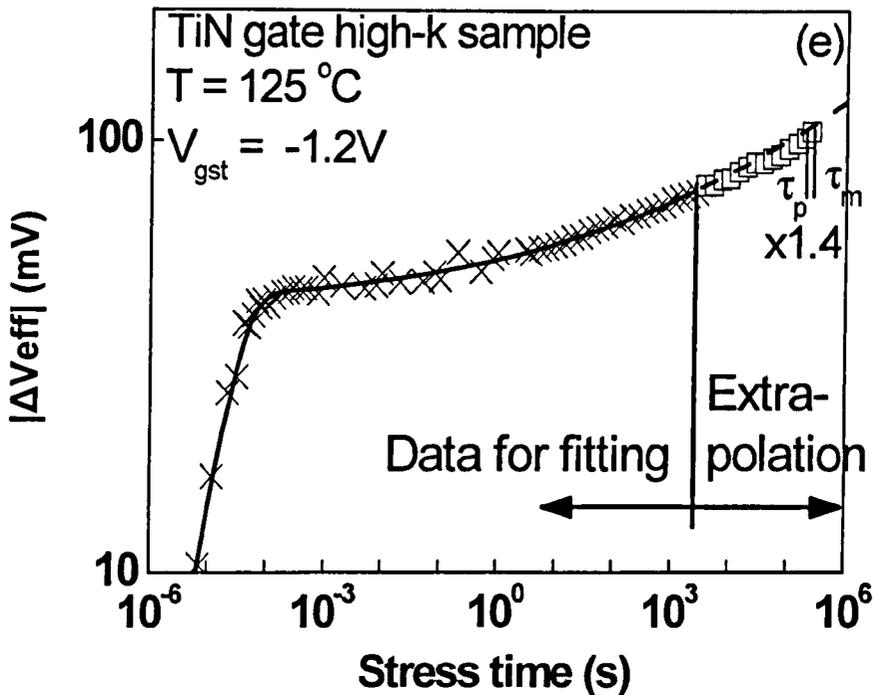
Fig.4.18 Applicability of the single test lifetime prediction technique for different fabrication processes: (a) 1.4nm plasma SiON, (b) 2.7nm thermal SiON, (c) 2.0nm 45sec plasma SiON, (d) 2.0nm 20sec plasma SiON, and (e) 2.0nm/1.0nm HfSiON/SiON stack prepared by ALCVD with TiN gate. The safety margin for the prediction is within 50% in all cases.

4.8   Conclusions

Evaluating NBTI-induced mobility degradation is problematic, controversial, and undesirable for test engineers. Central to this chapter is to propose using an effective threshold voltage shift, ΔVeff, as the single parameter for characterizing NBTI, which fully takes account of any potential contribution from mobility degradation. By assuming that mobility does not change under a constant (Vg-Vth), the evaluation of mobility variation is avoided. A method for extracting ΔVeff has been proposed, which only requires standard NBTI tests and can be easily implemented.

The results show that Id degradation under operation gate bias predicted from ΔVeff is in good agreement with the measured data, but ΔId is substantially underestimated or overestimated if the ΔVth(ex) or the OTF ΔVth is used, respectively. ΔVeff is preferred over ΔId/Id for characterizing NBTI, since it does not depend on device size and source/drain series resistance.

The ΔVeff is then used for the NBTI lifetime prediction in the worst case scenario where the recovery is suppressed and ΔVeff is sensed at the operation gate bias. In this case, the conventional V_g acceleration prediction is inapplicable, because the NBTI kinetics no longer follow a simple power law and an increase of stress bias does not lead to a parallel shift of log|ΔVeff|.

To predict the lifetime at the operation gate bias based on the UFP measurement, NBTI kinetics and defects are examined. An outstanding feature of the kinetics is the presence of a 'shoulder', which is insensitive to temperature and must be dominated by the charging of as-grown defects. The charging and discharging properties of the defect agree well with the signature of as-grown hole traps. By combining the first order model for the as-grown hole traps and the power law for generating new defects, ΔVeff can be modeled over ten orders of stress time. This kinetic model is then used to predict the NBTI lifetime, based on a single test at the operation temperature and bias. For the six different processes tested, the safety margin of the single test prediction technique is within 50%, which is substantially better than the methods proposed in early works.

References

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

2. 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 Dev., vol. 56, No.5, pp. 1086-1093, 2009.

3. Z. Ji, "Characterization of negative bias temperature instability and lifetime prediction for pMOSFETs," Ph.D thesis, Liverpool John Moores University (2010).

4. M. H. Chang, "Electron trapping in gate dielectrics and NBTI of MOSFETs," Ph.D thesis, Liverpool John Moores University (2006).

5. A. Aoulaiche, B. Kaczer, D. Jaeger, M. Houssa, K. Martens, R. Degraeve, P. Roussel, J. Mitard, S. De Gendt, H. E. Maes, G. Groeseneken, M. Meuris and M. M. Heyns, "Negative bias temperature instability on Si-passivated Ge-interface", in Proc. Int. Rel. Phys. Symp., 2008, pp.262-264.

6. C. Schlunder, W. Heinrigs, W. Gustin, and H. Reisinger, "On the impact of the NBTI recovery phenomenon on lifetime prediction of modern p-MOSFETs," in IRW final report, 2006, pp.1-4.

7. M. Ershov, S. Saxena, H. Karbasi, S. Winters, S. Minehane, J. Babcock, R. Lindley, P. Clifton, M. Redford, and A. Shibkov, "Dynamic recovery of negative bias temperature instability in p-type metal-oxide-semiconductor field-effect transistors," Appl. Phys. Lett., vol.83, pp.1647-1649, 2003.

8. H. Puchner and L. Hinh, "NBTI reliability analysis for a 90nm CMOS technology," in ESSDERC, 2004, pp.257-260.

9. Y. Z. Hu, D. S. Ang, and G. A. Du, in Proc. 46th Int. Reliability Physics Symp., p.743 (2008).

10. A.T. Krishnan, V. Reddy, S. Chakravathi, J. Rodriquez, S. John, S. Krishnan, "NBTI impact on transistor and circuit: models, mechanisms and scaling effects [MOSFETs]," in *IEDM Tech. Dig.*, pp.15.4.1-15.4.4, (2003).

11. C. G. Sodini, T. W. Ekstedt, and J. L. Moll, Soli-State Electronics, **25**, p.833 (1982).

12. Z. Ji, J. F. Zhang, W. Zhang, G. Groeseneken L. Pantisano, S. De Gendt, M. M. Heyns, "An assessment of the mobility degradation induced by remote charge scattering," Applied Phys. Lett., vol. 95, No. 26, Art. No. 263502, 2009.

13. W. D. Zhang, J. F. Zhang, M. J. Uren, G. Groeseneken, R. Degraeve, M. Lalor, and D. Burton, "On the interface states generated under different stress conditions," Appl. Phys. Lett., vol.79, no.19, 3092-3094 (2001).

14. E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology* (New York: Wiley, 1982).

15. M. J. Uren and K. M. Brunson, "An improved technique for the evaluation of surface Fermi energy in metal-oxide-semiconductor capacitors," *Semicond. Sci. Technol.*,   vol.9, pp.1504-1510, 1994.

16. C. Z. Zhao, J. F. Zhang, M. H. Chang, A. R. Peaker, S. Hall, G. Groeseneken, L. Pantisano, S. De Gendt, and M. Heyns, "Stress-induced positive charge in Hf-based gate dielectrics: impact on device performance and a framework for the defect," *IEEE Trans. Electron Dev.*, vol. 55, pp.1647-1656, 2008.

17. T. Yang, M. F. Li, C, Shen, C. H. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu, and D. L. Kwong, "Fast and slow dynamic NBTI

components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application," in *VLSI Symp. Tech. Dig.*, 2005, pp. 92-93.

18. M. H. Chang and J. F. Zhang, "On positive charges formed under negative bias temperature stresses," J. Appl. Phys., 101, Art. no. 024516, 2007.

19. A. Ortiz-Conde, F. J. Garcia Sanchez, J. J. Liou, A.   Cerdeira, M. Estrada, and Y. Yue, Microelectronic Rel., p.583 (2002).

20. J. H. Sim, S.C. Song, P. D. Kirsch, C. D. Young, R. Choi, D. L. Kwong, B. H. Lee, and G. Bersuker, "Effects of ALD HfO_2 thickness on charge trapping and mobility," Microelectronic Eng., **80**, 218 (2005).

21. P. K. Hurley, M. A. Negara, T. Van Hemert, K. Cherkaoui, "The influence of oxide charge on carrier mobility in HfO2/TiN gate silicon MOSFETs," ECS Transactions, 19, 379 (2009).

22. M. Khare, X. W. Wang and T. P. Ma, IEEE Electron Dev. Lett., 20, p.57 (1999).

23. I.S.Al-kofahi, J.F.Zhang and G.Groeseneken, "Continuing degradation of the SiO2/Si interface after hot hole stress," J. Appl. Phys., 81, 2686-2692 (1997).

24. J. F. Zhang, M. H. Chang, Z. Ji, L. Lin, I. Ferain, G. Groeseneken, L. Pantisano, S. De Gendt, and M. M. Heyns, "Dominant layer for stress-induced positive charges in Hf-based gate stacks," IEEE Electron Device Letters, p.1360 (2008).

25. A. Dixit, A. Kottantharayil, N. Collaert, M. Goodwin, M. Jurczak, and K. De Meyer, IEEE Trans. Electron Dev., p.1132 (2005).

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

27. G. Van den bosch, G. Groeseneken, H. E. Maes, R. B. Klein, and N. S. Saks, "Oxide and interface degradation resulting from substrate hot-hole injection in

metal-oxide-semiconductor field-effect transistors at 296 and 77 K," *J. Appl. Phys.*, vol.75, p.2073-2080, 1994.

28. 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 Dev.*, vol.51, pp.1267-1273, 2004.

29. C. Z. Zhao, J. F. Zhang, G. Groeseneken and R. Degraeve, "Hole traps in silicon dioxides --- Part II: Generation mechanism," *IEEE Trans. Electron Dev.*, vol.51, pp.1274-1280, 2004.

30. C. Z. Zhao and J. F. Zhang, "Effects of hydrogen on positive charges in gate oxides," *J. Appl. Phys.*, vol.97, Art. No. 073703, 2005.

31. J. F. Zhang, H. K. Sii, G. Groeseneken, and R. Degraeve, "Hole trapping and trap generation in the gate silicon dioxide," *IEEE Trans. Electron Dev.*, vol.48, pp.1127-1135, 2001.

32. J. F. Zhang, H. K. Sii, A. H. Chen, C. Z. Zhao, M. J. Uren, G. Groeseneken and R. Degraeve, "Hole trap generation in gate dielectric during substrate hole injection," *Semiconductor Sci. and Technology*,   vol.19, pp.L1-L3, 2004.

33. S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang, and L. Chen, "Nitrogen-enhanced negative bias temperature instability: An insight by experiment and first-principle calculations," *Appl. Phys. Lett.*, vol.82, pp.1881-1883, 2003.

34. E. N. Kumar, V. D. Maheta, S. Purawat, A. E. Eslam, C. Olsen, K. Ahmed, M. A. Alam, and S. Mahapatra "Material dependence of NBTI physical mechanism in silicon oxynitride (SiON) p-MOSFETs: A comprehensive study by ultra-fast on-the-fly (UF-OTF) I_{DLIN} technique," in *IEDM Tech. Dig.*, 2007, pp.809-812.

35. A. Neugroschel, G. Bersuker, R. Choi, and B. H. Lee, "Effect of the interfacial SiO$_2$ layer in high-k HfO$_2$ gate stacks on NBTI," *IEEE Trans. Dev. Materials Rel.*, vol.8, pp.47-61, 2008.

36. V. D. Maheta, C. Olsen, K. Ahmed, and S. Mahapatra, "the impact of nitrogen engineering in silicon oxynitride gate dielectric on NBTI of pMOSFETs a study by ultra OTF Idlin technique," *IEEE Trans. Electron Dev.*, vol.55, pp.1630-1638,2008.

37. A. T. Krishnan, C. Chancellor, S. Chakravarthi, P. E. Nicollian, V. Reddy, A. Varghese, R. B. Khamankar, and S. Krishnan, "Material Dependence of Hydrogen Diffusion Implications for NBTI Degradation," in *IEDM Tech. Dig.*, 2005, pp.691-694.

38. Y. F. Chen, M. H. Lin, C. H. Chou, W. C. Chang, S. C. Huang, Y. J. Chang, K. Y. Fu, M. T. Lee, C. H. Liu, and S. K. Fan, "Negative bias temperature instability (NBTI) in deep sub-micron p-gate pMOSFETs," in *IIRW*, 2000 pp. 98-101.

39. Y. Mitani, M. Nagamine, H. Satake, and A. Toriumi, "NBTI mechanism in ultra-thin gate dielectric - nitrogen-originated mechanism in SiON," in *IEDM Tech. Dig.*, 2002, pp.509-512.

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

41. S. Ogawa and N. Shiono, "Generalized diffusion-reaction model for the low-field charge-buildup instability at the Si-SiO$_2$ interface," *Phys. Rev. B*, vol.51, pp.4218-4230, 1995.

42. M. Alam, H. Kufluoglu, D. Varghese, and, S. Mahapatra, "A comprehensive model for PMOS NBTI degradation: Recent progress," *Microelectron. Reliab.,* vol.47, pp.853-862, 2007.

5 | A single pulse charge pumping technique for fast measurement of interface states

5.1 Introduction

The near perfect SiO_2/Si interface plays a major role in the success of silicon-based CMOS technologies. This high quality of interface, however, is achieved only after an anneal in a hydrogen environment. Without such an anneal, the typical interface states are in the order of 10^{12} cm^{-2} [1-3]. It is widely believed that the interface states originate from silicon atoms of a dangling bond [1], although different types of interface states have been reported [3,4]. Annealing in hydrogen at a temperature around 400°C forms Si-H bonds and passivates the interface states. When compared with Si-O bonds, Si-H bonds at the interface are relatively weak and can be ruptured by a number of physical processes, such as irradiation [5-7], hot carriers [8-10], positive [11,12] and negative [13-18] biased temperature stresses. Moreover, the generation of interface states can continue even post irradiation [5-7] and electrical stresses [19-21]. It has been proposed that the generation of interface states controls the hot carrier lifetime of nMOSFETs [8], substantially degrades the lifetime of pMOSFETs due to the negative bias temperature instability (NBTI) [13-18], and reduce the transfer efficiency of charge coupled devices (CCDs). Characterizing interface states is an important task, therefore.

As mentioned in section 2.4, several techniques have been developed to measure interface states, such as high-frequency capacitance-voltage (HFCV), namely the Terman's method [22], quasi-static capacitance-voltage (QSCV) [23,24], conductance [23,24], subthreshold swing [25-27] and charge pumping [4,28-30]. All of these techniques, however, suffer from a common drawback: slow measurement speed and the typical measurement time is in the order of seconds [22-30]. It has been shown recently that the degradation measured at such speed is only a fraction of the real degradation, because of the rapid recovery of degradation after removing stresses [15-17,31,32]. Ultra-fast measurement techniques have been developed and the results show that measurement time has to be reduced to the order of microseconds to minimize the recovery [15-17,31,32]. However, these fast techniques are based on monitoring the transient transfer characteristics and the measured drain current is affected by both created interface states and charges trapped in the gate dielectrics [26,32]. The interface state density cannot be extracted from these fast measurements. The central task of this chapter is to develop a fast characterization technique that allows direct evaluation of interface states with a time in the order of microseconds.

This chapter will be organized in the followings. Section 5.2 will analyze the potential of each existing technique for fast measurements and justify the selection of charge pumping method. The experimental setup for overcoming the shortcomings of the conventional charge pumping technique will then be described in section 5.3. Section 5.4 presents the principle of single pulse charge pumping (SPCP) technique and the formula needed for extracting interface states. Section 5.5 will calibrate the interface states measured by SPCP against those from the well-accepted charge

pumping technique. Section 5.6 examines the effect of pulse edge time on the measurement and section 5.7 presents the results for nMOSFETs. The issues and applicability of SPCP to thin dielectric will be investigated in section 5.8. Section 5.9 studies the recovery of interface states with a time resolution of microseconds. Finally, conclusions will be drawn in section 5.10.

5.2   Selection of techniques

The most sensitive and complete technique for measuring interface states is the conductance method, which can detect interface states as low as the order of 10^9 cm^{-2}. It also gives the capture cross section of interface states [23]. Unfortunately, this is the most time consuming technique. Under each gate bias, a conductance against frequency sweep is typically carried out, that can take tens of seconds. In addition, to find the interface potential under a given gate bias, a low frequency CV has to be measured. As a result, it is impossible to use this technique for fast measurements.

The Terman's method is one of the first techniques used for measuring interface states, developed as early as in 1962 [22]. Although it measures capacitance at high frequency, this does not mean that it can be done at a high speed. During the measurement, the gate bias consists of a quasi-DC ramp and a small high frequency probing signal. The quasi-DC ramp rate must be sufficiently slow that the thermal equilibrium with interface states is maintained. The typical ramp rate used is between 5 and 50 mV/sec [24] and it can take over 20 sec to sweep one volt. The other CV techniques such as quasi-static CV and high-low frequency CV [22-24] all

require the same quasi-DC ramp for the gate bias, so that they cannot be used for fast measurements.

The subthreshold swing (SS) technique [25-27] only requires measuring the transfer characteristics in the low gate bias region, making it suitable for samples with thin gate dielectrics where gate leakage is problematic for other techniques. Unlike charge pumping, the SS does not need a contact to the substrate, so that it can be applied to silicon on insulator devices, where a connection to the substrate is often not available. Unfortunately, this technique is also based on the assumption that the interface states are in thermal equilibrium with the gate bias sweep, so that fast pulse cannot be applied here.

The charge pumping can be used for small MOSFETs and is a popular technique [4,28-30]. It is a technique that actually requires the interface states not being in thermal equilibrium with the measurement signal, offering the potential for fast measurements. Although the base level of gate bias pulse is often swept, such sweep is not essential. In principle, the interface states can be determined by measuring a single charge pumping current, so long that the gate bias pulse covers the range from flatband voltage to threshold voltage. Conventionally, multiple gate pulses are applied during the measurement and the measured charge pumping current is an average DC current. For a typical parameter analyzer, measuring one DC point will take 10-150 ms [17]. Although this is a significant improvement when compared with the seconds needed by other techniques, it is still too slow since recovery was observed in the order of microseconds [15-17,31]. One potential solution to the problem is to replace the DC current measurement by recording transient currents. It

will be shown that this is indeed achievable and a single pulse charge pumping (SPCP) technique will be developed in this Chapter.

5.3   Devices and Experimental

5.3.1   Devices

The equivalent gate oxide thickness used in current industry is less than 2nm and the gate leakage current can be substantial during the charge pumping measurement. To simplify the experimental conditions, a relatively thick (7.1nm) SiO_2 layer will be used first to develop the new SPCP technique, so that the interference from the gate leakage is eliminated. The pMOSFETs have a p+ poly-si gate and nMOSFETs have an n+ poly-si gate. The n well was doped to a level of $2 \times 10^{17} cm^{-3}$ approximately. The channel length and width is 10x200μm. Each device has contacts to the well and substrate, allowing the substrate hot carrier injection being used to accelerate the generation of interface states [33].

The applicability of the SPCP to thin dielectrics will then be demonstrated on two samples. One of them is a plasma nitrided 2nm SiON with a p+ poly-si gate. The other is an HfSiON (1nm)/SiON (1nm) stack with a TiN gate and an equivalent oxide thickness of 1.22nm. Both samples have a channel length and width of 10×10μm. The detailed stress and measurement conditions used are given in the figure captions.

5.3.2   Experimental setup

Fig.5.1 shows the experimental setup. The source and drain of a MOSFET is tied together and a pulse is applied to the gate. Three transient voltages were recorded by an oscilloscope: gate bias V_g and the outputs of two operational amplifiers, $V(I_{sd})$ and $V(I_b)$. The transient current in the channel, I_{sd}, and in the substrate, I_b, is converted from $V(I_{sd})$ and $V(I_b)$ by,

$$I_{sd}=V(I_{sd})/R, \text{ and } I_b=V(I_b)/R, \tag{1}$$

where R is the feedback resistance and is typically set at 10 KΩ.

A screen shot of the oscilloscope is shown in Fig.5.2(a) for a stressed device. It confirms that there are no voltage overshoots at V_g corners from measurement errors. Spikes, however, appear in both $V(I_{sd})$ and $V(I_b)$. To rule out that these spikes originate from parasitic effects such as geometric components [34], Fig.5.2(b) shows that they are absent in the screen shot for a fresh device. As a result, they must come from the stress generated defects and we will show how they can be used to extract interface states next. The gate bias pulse always has an equal rise and fall time, which is typically 6μs.
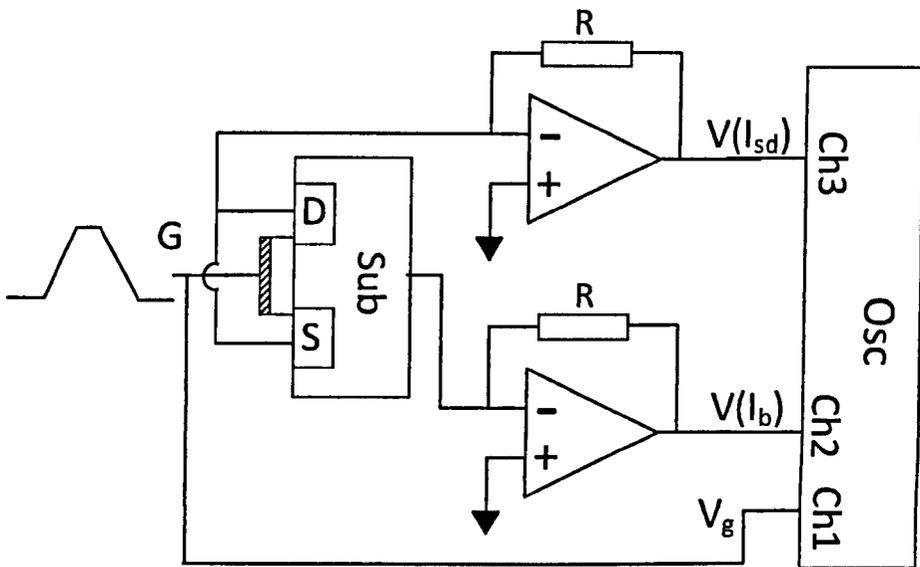
Fig.5.1 The experimental setup for the single pulse charge pumping (SPCP) measurements.

5.4   The principle of single pulse charge pumping technique

The gate bias switches from +1 to -3 V during the first edge of the pulse in Fig.5.2(a), which turns on the pMOSFET. During the pMOSFET switch on, displacement current of C_{gc} charges formed the current I_{sd}. The corresponding transient channel current, I_{sd}, is obtained by converting $V(I_{sd})$ in Fig.5.2(a) through equation (1) and is plotted against V_g as the solid curve in Fig. 5.3(a). The pMOSFET is switched off during the second edge of gate pulse and the I_{sd} also is given in Fig.5.3(a) as the dashed curve. To facilitate the comparison of I_{sd} for the two pulse edges, we define I_{sd} as positive for both edges, although I_{sd} actually flows in the opposite directions since V_g is swept in the opposite directions in Fig.5.2(a).

Fig.5.3(a) shows that I_{sd} has a profound peak when switched from off-to-on, but this peak is absent during the on-to-off transition. To understand this difference, we analyze the charging and discharging of interface states during the edges. The charge

neutrality level of SiO$_2$/Si interface is close to the middle of silicon bandgap, E$_i$, with acceptor-like and donor-like states above and below E$_i$, respectively [35,36]. At the off-mode, V$_g$>0 and Fig.5.3(b) shows that acceptor-like states are below the Fermi-level, E$_f$, and negatively charged, while the donor-like states are neutral. For the off-to-on switching, the acceptor-like states sufficiently below the edge of conduction band, E$_c$, cannot emit their electrons in time before the device is inverted. These negative states will recombine with holes flowing into the channel from source and drain, resulting in a recombination channel current. In addition, the donor-like states above the Fermi level will become positively charged by capturing holes, leading to a charging current. The recombination and charging currents together give rise to the spike in off-to-on I$_{sd}$ in Fig.5.3(a).
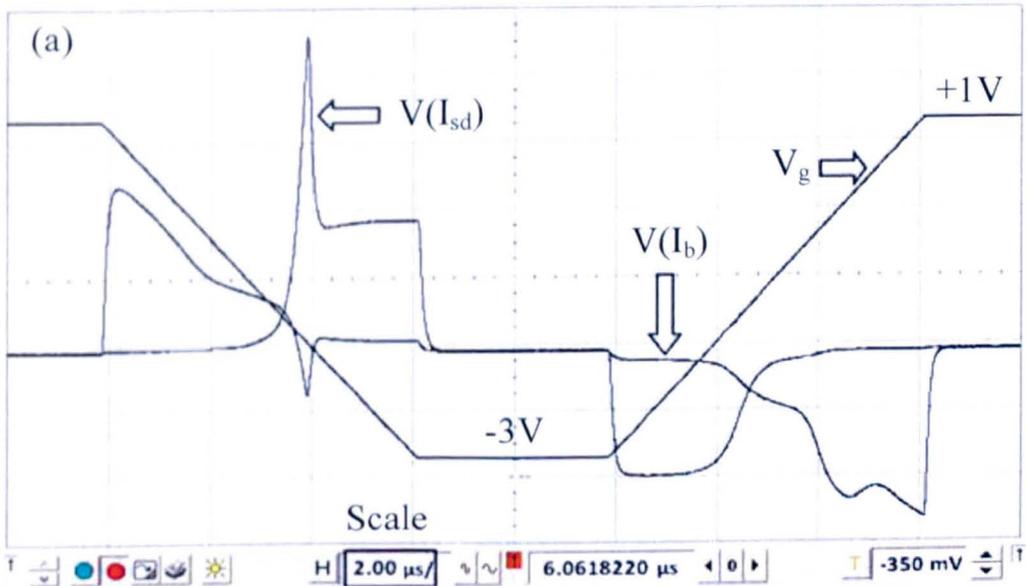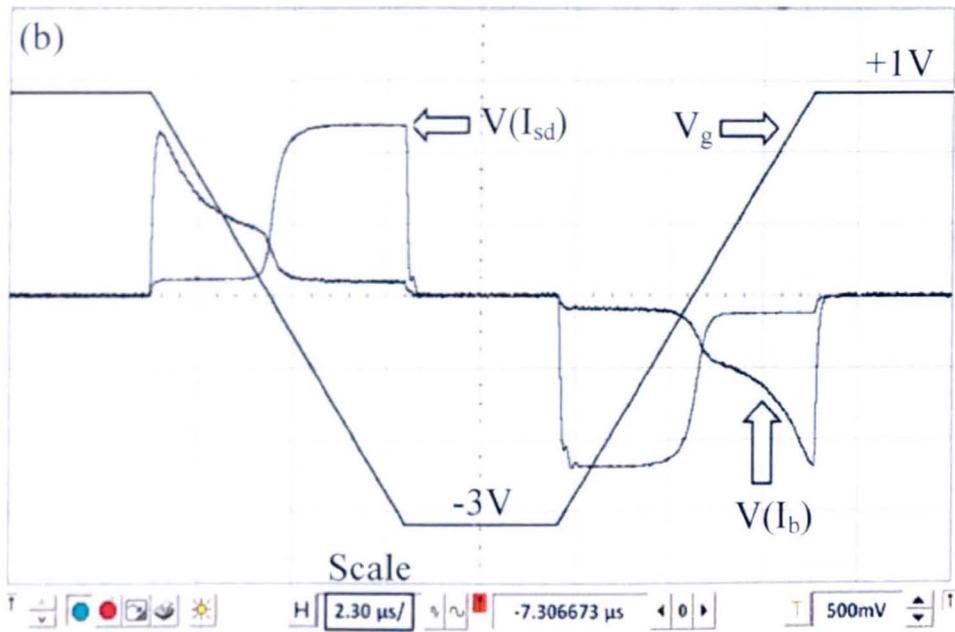
Fig.5.2 The screen shot of the oscilloscope showing typical waveforms for a stressed device (a) and a fresh device (b). The device in (a) was stressed under an oxide field of -5MV/cm by the substrate hot hole injection (SHI) with a n-well bias of +5V and p-substrate bias of +5.7V. The gate dielectric is a 7.1 nm SiO_2 layer.
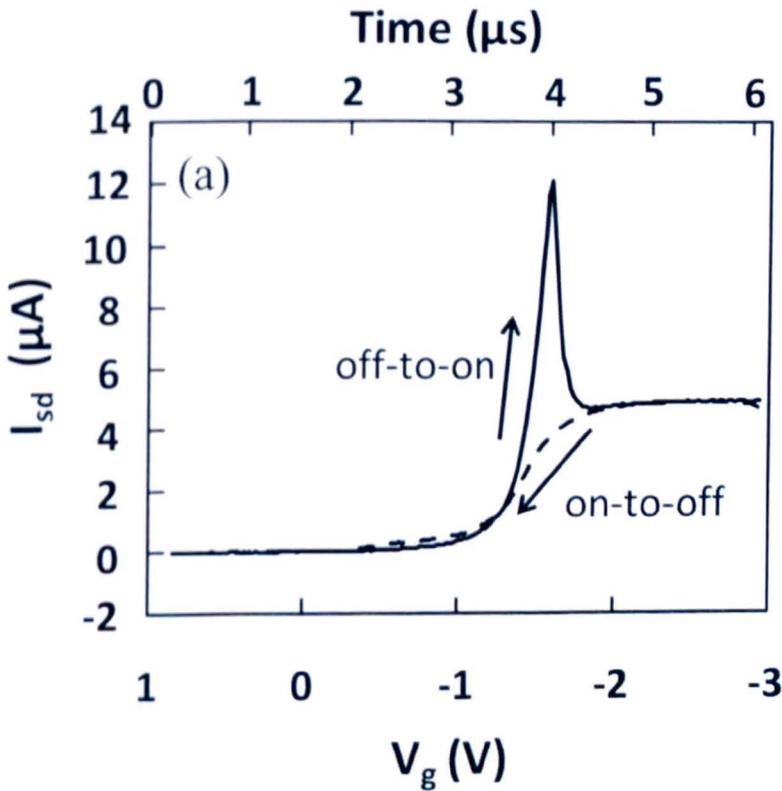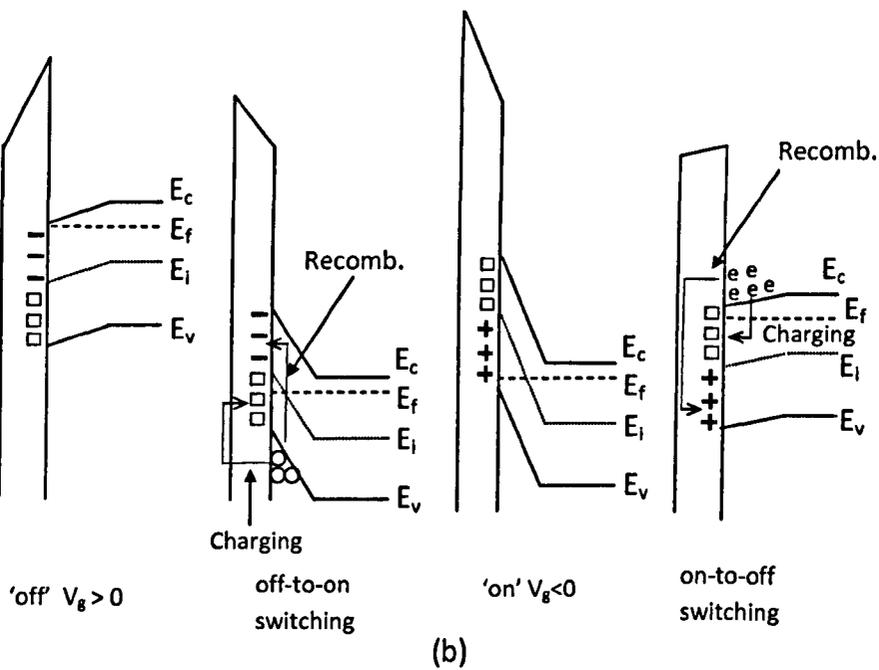
Fig.5.3 (a) The channel current corresponding to the two edges of the gate pulse in Fig.5.2(a). (b) Energy band diagrams under different operation conditions. '-', '□' and '+' represents negative, neutral and positive states and 'o' and 'e' are holes and electrons.

During the on-to-off switching, however, the recombination is between the positive donor-like states and electrons from the substrate and the charging of acceptor-like states also is by substrate electrons, as shown in Fig.5.3(b). Here the recombination and charging does not contribute to channel current and I_{sd} is mainly a displacement current, explaining the lack of peak during the on-to-off transition.

Based on the above explanation, the recombination and charging of states during the on-to-off switching requires an additional electron flow from the substrate and this should give rise to a peak in the substrate current, I_b. The dashed curve in Fig.5.4 shows that this is indeed the case. When compared with the peak in Fig.5.3(a), the 'on-to-off' peak in Fig.5.4 is broader. To explain this difference, we note from the inset of Fig.5.4 that the rise of capacitance from its minimum towards accumulation is not as sharp as that towards strong inversion. This is to say that the rise towards

accumulation occurs over a larger range of V_g. The rise of capacitance towards accumulation originates from electrons running towards the oxide/substrate interface from the substrate, during which electrons neutralize donor-like states and charge acceptor-like states and give rise to the 'on-to-off' peak.   When it occurs over a larger range of V_g, it will take longer for a given dV_g/dt. For a fixed amount of interface states related charges, a longer time results in a lower current and a less profound peak.

It is intriguing that there is also a downward peak in the 'off-to-on' I_b in Fig.5.4. Fig. 5.2(a) shows that the location of this peak is the same as the I_{sd} peak, so that it must be related to the interface states. During the 'off-to-on' transition and before the recombination occurring, there are negative acceptor-like states that are not in thermal equilibrium with gate pulse. These negative states cause a transient deep depletion. As they are neutralized by recombination, the deep depletion region collapses, resulting in an electron flow marked by 'f' in the inset of Fig.5.4. Since 'f' is in the opposite direction from the off-to-on displacement current in substrate, it leads to the downward peak in Fig.5.4.
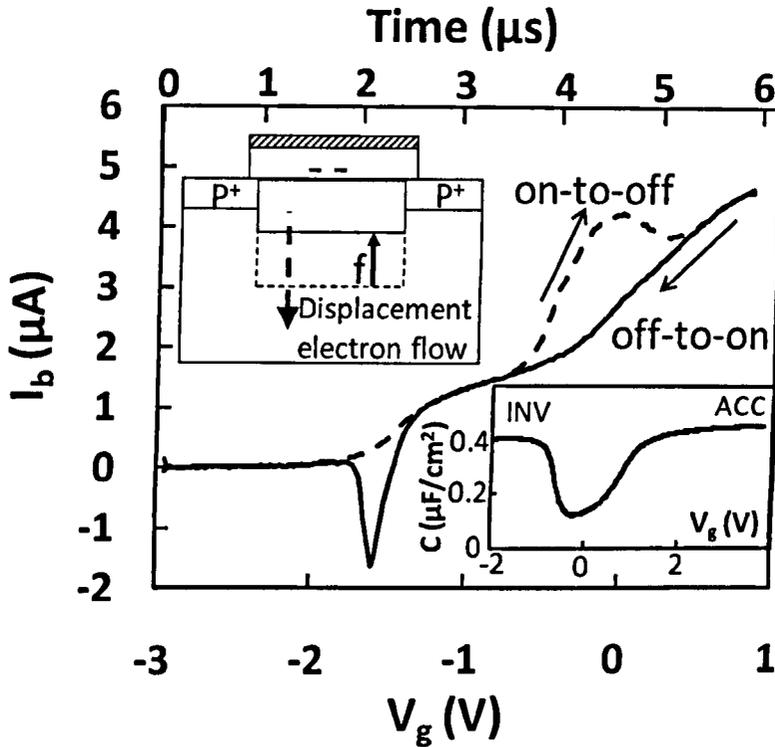
Fig.5.4 The substrate current corresponding to the two edges of the gate pulse in Fig.5.2(a). The top-left inset shows the deep depletion caused by negative states during 'off-to-on' switching. As the negative states are neutralized, the deep depletion collapses with an electron flow marked by 'f' in the opposite direction of the displacement current. The bottom-right inset shows that the transition from the minimum capacitance to accumulation takes longer than that to inversion.

The above analysis reveals that the differences in the transient currents for the two edges originate from interface states. We now explore how they can be used to extract interface states. To evaluate the amount of charges flowing in or out the device during switching, the transient current should be integrated against the edge time in Figs.5.3(a) and 5.4:

$$\Delta Q_b = \int_0^t [I_b(on-to-off)-I_b(off-to-on)]dt \qquad (2)$$

and

$$\Delta Q_{sd} = \int_0^t [I_{sd}(off-to-on)-I_{sd}(on-to-off)]dt \qquad (3)$$

where ΔQ_b and ΔQ_{sd} is the net charges flowing into the device from substrate and source/drain, respectively. Despite of the different waveform of I_{sd} and I_b, Fig.5.5 shows that ΔQ_{sd} and ΔQ_b saturates at the same level, since they originates from the same interface states. The number of states per unit area, N_{it}, can be evaluated from this saturation level,

$$N_{it} = \Delta Q_{sat}/(q * L * W) \tag{4}$$

where L and W are channel length and width, respectively, and q is one electron charge.
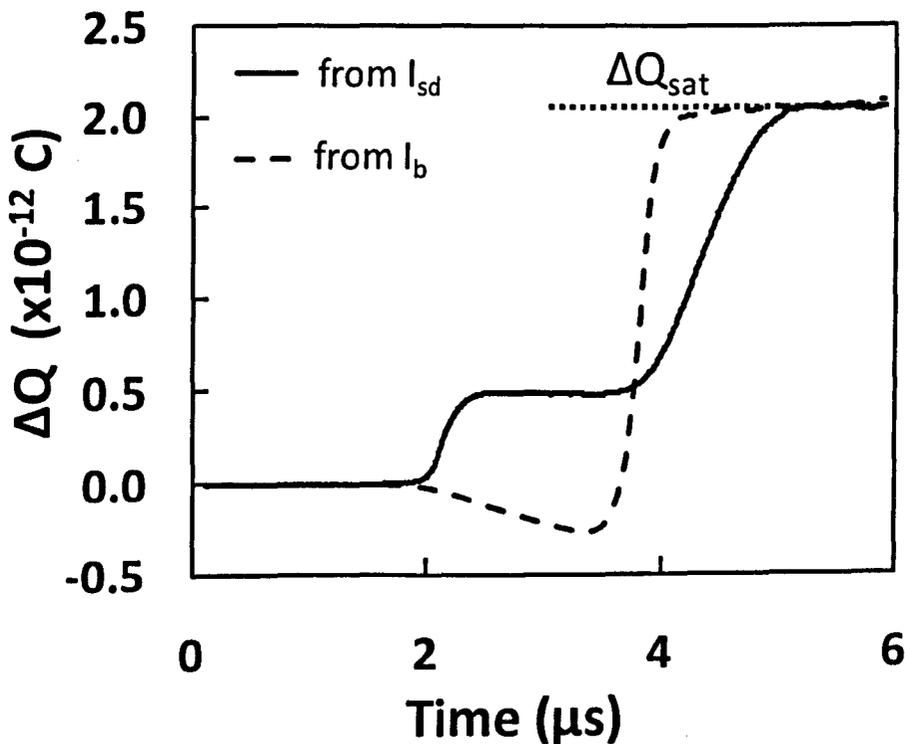
Fig.5.5 The net charge pumped into the devices from the substrate and the source/drain, calculated from the difference in the two curves in Figs. 5.3(a) and 5.4. ΔQ_{sat} is the saturation level.

5.5   Calibration of single pulse charge pumping technique

To calibrate the SPCP method, the N_{it} extracted through eq. (4) will be compared with that from the conventional CP (CCP) technique for a 7.1nm SiO_2 sample. To create interface states, a high oxide field is often applied, which could break down the oxide before a large number of interface states were created. In order to generate large number of interface states at moderate electrical field, substrate hole injection (SHI) technique can be used [33]. Here, the n-well/p-substrate junction was forward biased during the stress, as illustrated in Fig.5.6. The source and drain were grounded and the holes flowing into the n-well were driven towards the interface by the positive bias applied on the n-well. They bombard the interface and accelerate the generation of interface states.
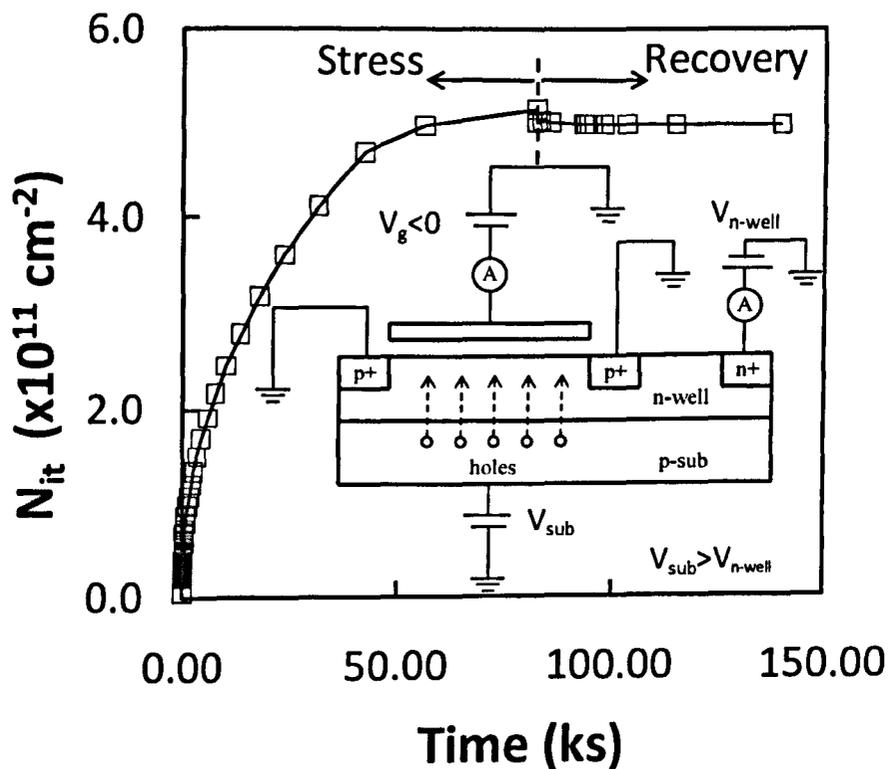
Fig.5.6 The interface states generated during the substrate hole injection (SHI) and their recovery under $V_g=0$ after the SHI. The inset schematic diagram shows the biases of SHI. The conventional charge pumping was used here for the measurement.

To calibrate SPCP against CCP, the interface states must remain the same during both CCP and SPCP measurements. To ensure this, the device was allowed to recover for 10mins following stress and before measurements. Fig.5.6 shows that interface states become stable 5mins after stress, so that they will not change for the subsequent CCP and SPCP measurements. Fig.5.7 shows that N_{it} extracted from SPCP agrees well with that from CCP under the same pulse edge time, justifying SPCP as a technique for measuring interface states. Although SPCP uses transient currents that have a lower accuracy than the DC current used by CCP, Fig.5.7 shows that interface states in the order of 10^{10} cm^{-2} can be reliably measured by SPCP.
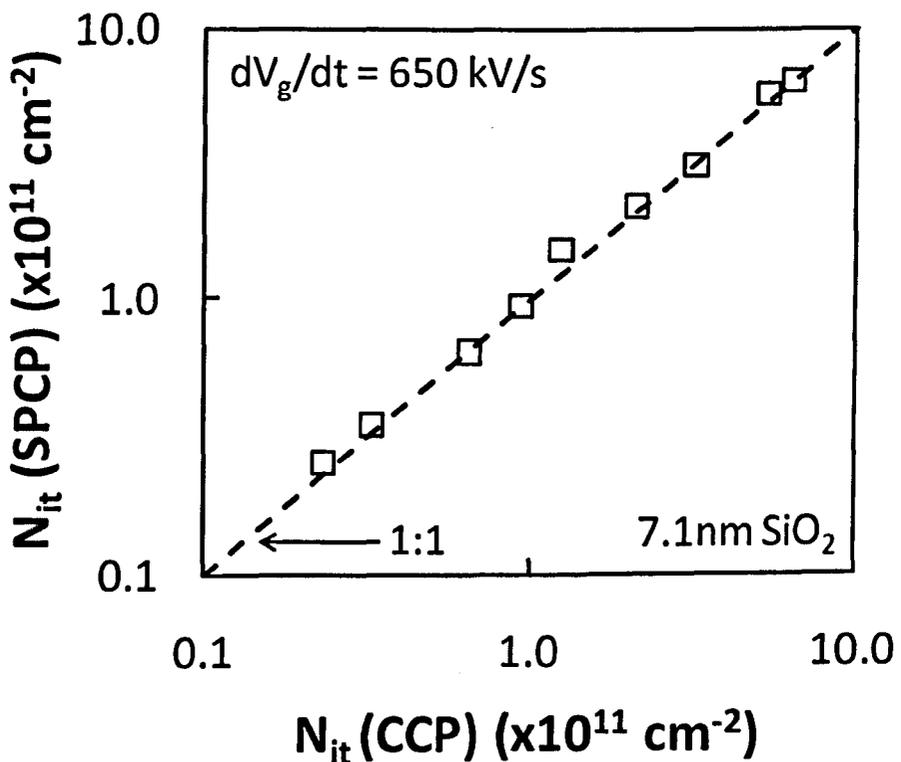
Fig.5.7 Calibration of the single pulse CP technique against the conventional CP method. After stress, the device was allowed to recover with gate floating for 10 min before measurements. The interface states were first measured by conventional CCP and then by SPCP. The two measurements agree well when interface states remain steady during the measurements.

5.6   Single pulse charge pumping with different pulse edge time

The charge pumping occurs mainly during the edge time of the gate pulse and it is interesting to study the effect of pulse edge times. Figs.5.8(a)-(c) show the impact of pulse edge time on the measurement. A longer edge time leaves less interface states available for charge pumping [4], but good agreement between the SPCP and CCP is achieved for all edge times used here.
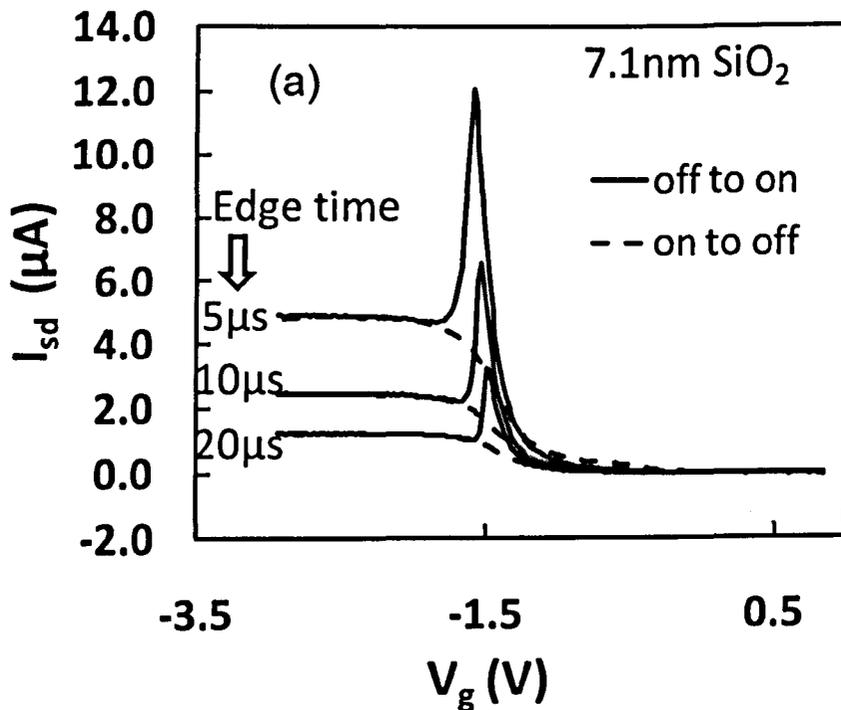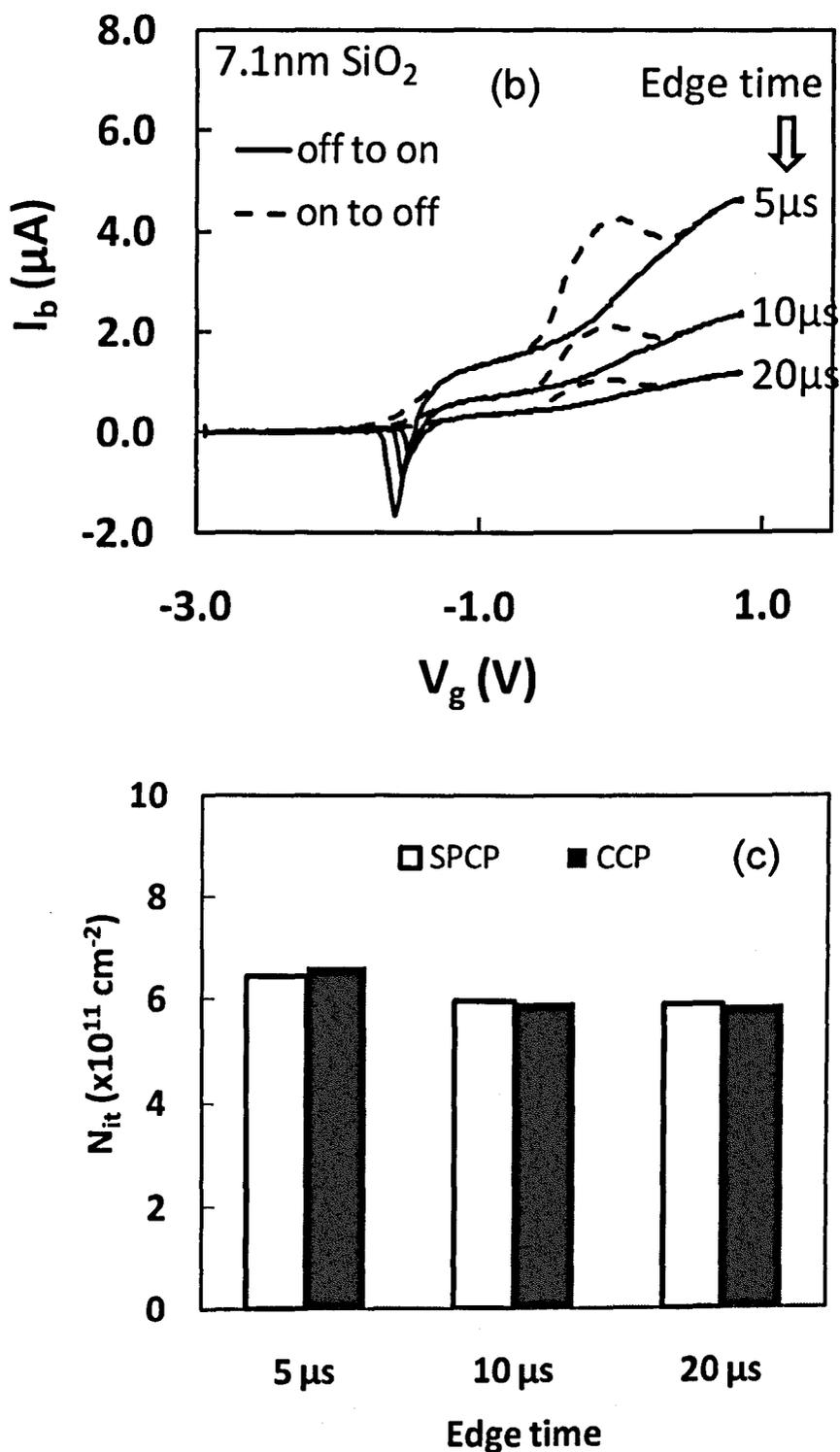
Fig.5.8 The impact of pulse edge times on the channel (a) and substrate (b) currents. (c) shows that the interface states extracted by SPCP agree well with that by CCP for all edge times used here.

Basically, the time of pulse edge time corresponding to the range of energy gap. As shown in fig.5.9, the faster pulse edge time will allow us to probe larger range of energy gap.
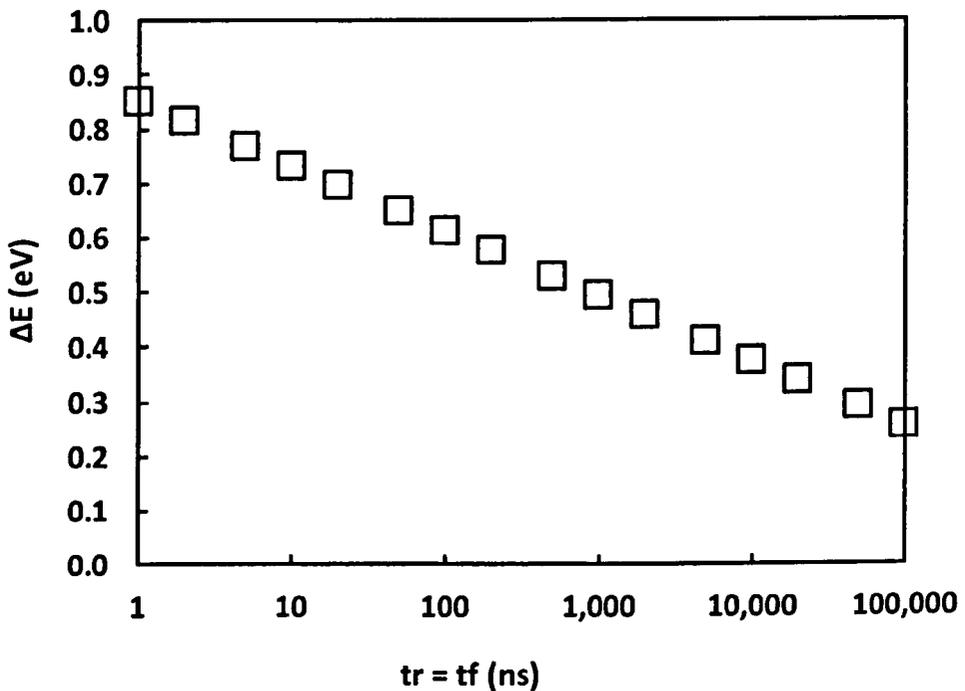
Fig.5.9 Different pulse edge time can probe different range of energy gap

5.7   Applicability of SPCP to nMOSFETs

Up to now, the test was carried out on pMOSFETs. In this section, tests are carried out on nMOSFETs. Fig.5.10 presents the waveforms of $V(I_{sd})$ and $V(I_b)$ for a stressed nMOSFET. Similar to the results of pMOSFETs, peaks can be observed in Fig.5.10, especially for $V(I_{sd})$. These peaks originate from the generated interface states, since they are absent for fresh sample. Figs.5.11(a) and 5.11(b) compare the I_{sd} and I_b for the two pulse edges. The I_{sd} and I_b for the two edges are clearly different here due to charge pumping of the created interface states.
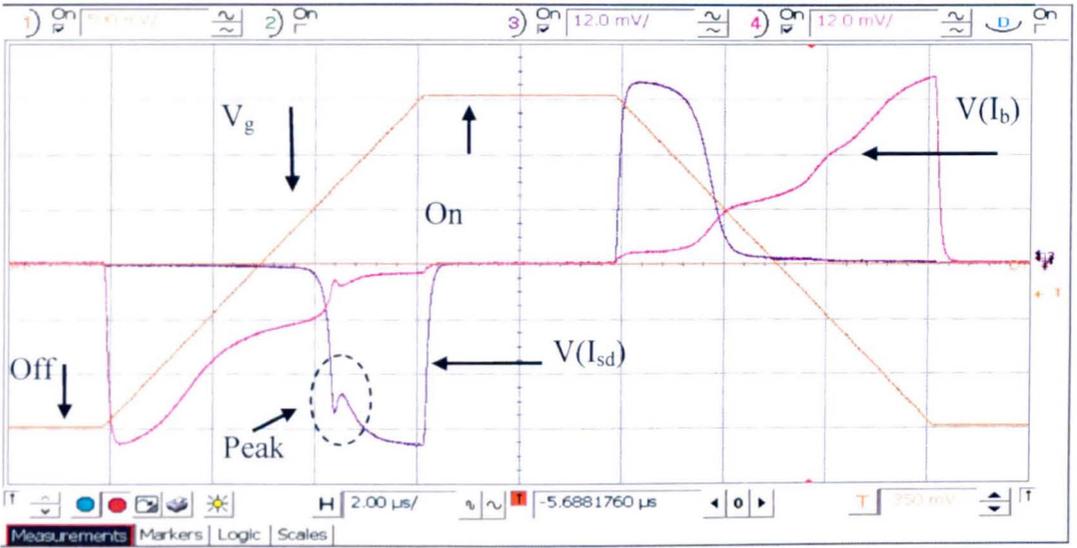
Fig.5.10 A screen shot of the signals when a gate pulse, V_g, is applied to a stressed nMOSFET. Apart from a change of polarity, note the differences in $V(I_{sd})$ and $V(I_b)$ for the two pulse edges.
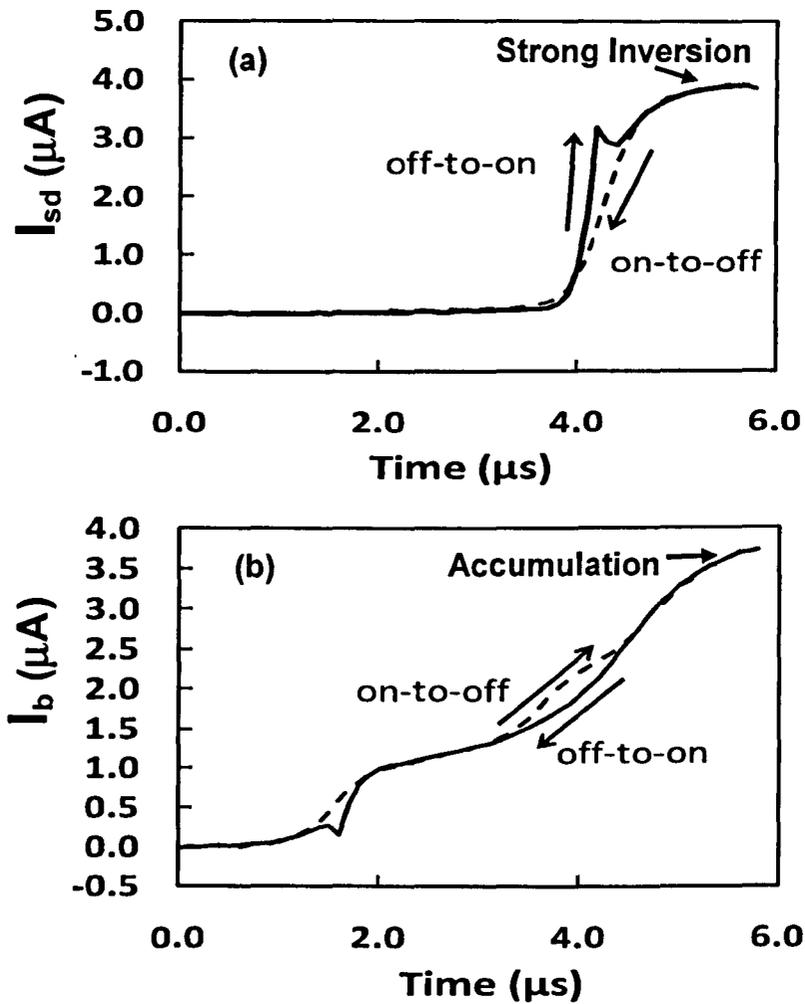
Fig.5.11 The transient channel current (a), I_{sd}, and substrate current (b), I_b, converted from the $V(I_{sd})$ and $V(I_b)$ in Fig.5.9. After stress, clear differences are observed in both I_{sd} and I_b between the two edges, caused by the stress induced interface states.

By using equations (2) and (3), the ΔQ_{sd} and ΔQ_b are calculated and given in Fig.5.11. As expected, both ΔQ_{sd} and ΔQ_b saturates, since the number of interface states are limited. Importantly, ΔQ_{sd} and ΔQ_b saturate again at the same level, despite the different waveform of I_{sd} and I_b in Figs.5.11(a) and 5.11(b). The interface states per unit area, N_{it}, can be evaluated from this saturation level by using the equation (4).
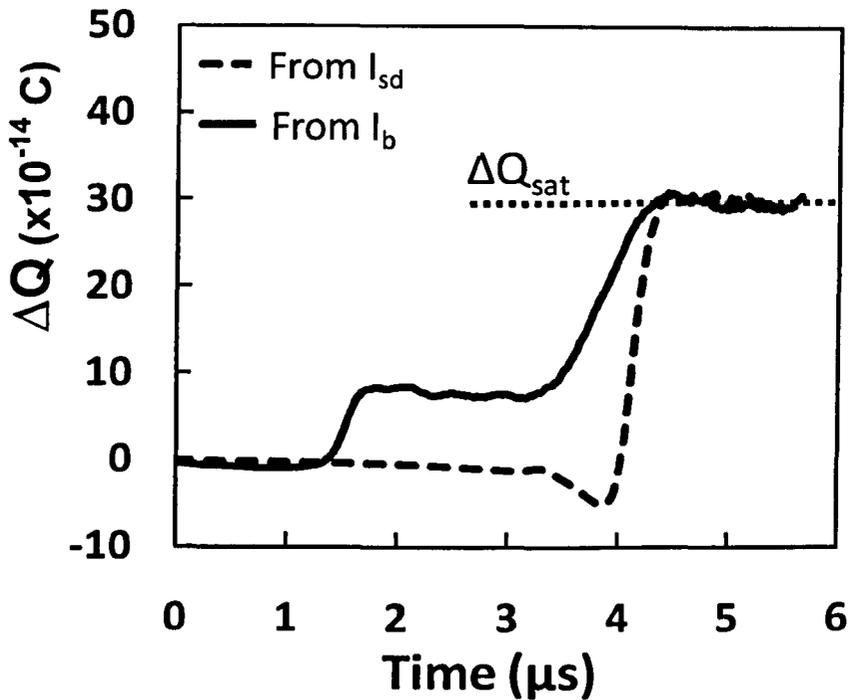
Fig.5.11 The net charges pumped into the device from the substrate (the solid line) and the source and drain (the dashed line). Although they have different shapes, both of them saturate at the same level.

To calibrate the interface states evaluated by the SPCP technique, the results are again compared with those evaluated by the conventional charge pumping (CCP) technique. The pulse edge times used for the CCP are the same as those for SPCP, but the pulse was repeatedly applied for the CCP and the DC charge pumping current was measured. Care also was excised to ensure that the same number of interface states were present during both measurements. Fig.5.13 shows that the N_{it} extracted from SPCP agrees well with that obtained from CCP. When the stress increases, N_{it} obtained by both techniques increases in step and their ratio remains at one. Moreover, although the accuracy of transient measurements is generally not as good as that of DC measurements, Fig.5.13 shows that the SPCP can be used to measure the degradation of N_{it} in the order of 10^{10} cm^{-2}, which is adequate for typical stress tests [10]. This gives us the confidence that the SPCP can be used to monitor the

transient behavior of interface states with a time resolution in the order of microseconds.
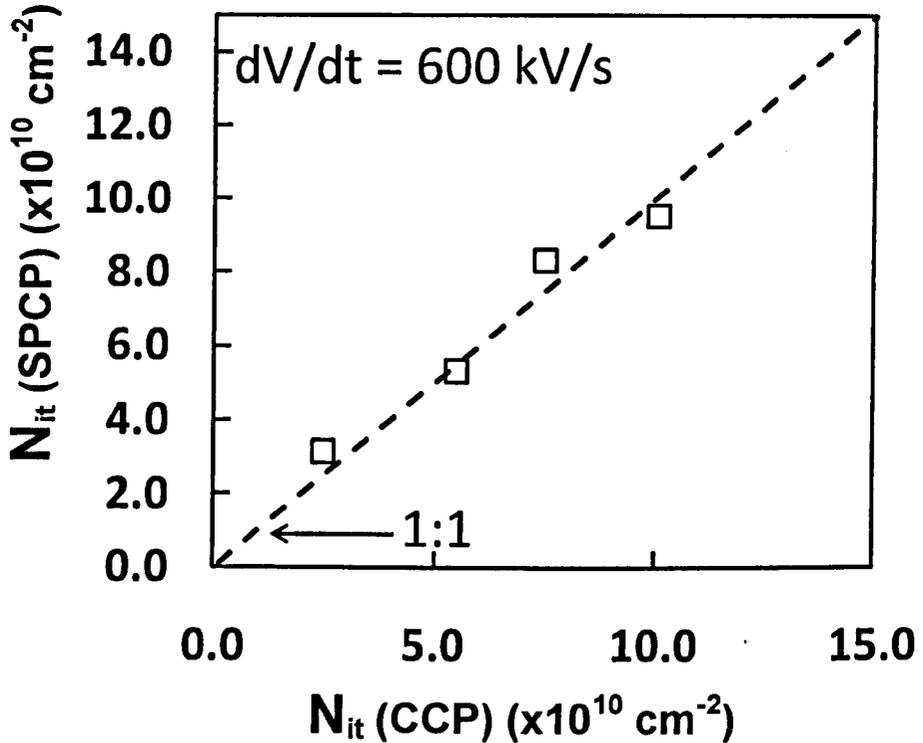
Fig.5.13 Calibration of the single pulse CP technique (SPCP) against the conventional CP (CCP) method. The interface states were first measured by CCP and then by SPCP. The interface states measured by SPCP agree well with those by CCP.

5.8 Applicability of SPCP to thin dielectrics

The main challenge of thin gate dielectrics on interface state measurements is that the high gate leakage can bury the net charge pumping current. One example is given in Fig.5.14 for a 2nm SiON layer. Unlike the 7.1nm sample in Fig.5.3(a) where I_{sd} for two edges merges at high negative V_g, Fig.5.14(a) shows the I_{sd} now diverges due to gate leakage current through the channel, I_{gsd}. During the off-to-on switching, the inset of Fig.5.14(b) shows that the displacement current and I_{gsd} is in the same

direction, resulting in the higher I_{sd} for more negative V_g. For the on-to-off switching, however, the displacement current changes direction, but I_{gsd} does not, so that they are in the opposite directions now. The inset of Fig.5.14(a) shows that the two I_{sd} curves do not merge as V_g is swept towards the negative direction, so that the net charge pumped from the source and drain in Fig.5.14(b) does not have a saturation level, making the evaluation of interface states through equation (4) problematic.

Fig.5.15 gives the corresponding substrate current. Although the two I_b also diverge for negative V_g, they still merge for the positive V_g, allowing the net pumped charge to saturate, as shown by the inset of Fig.5.15. This makes the SPCP applicable to thin dielectrics. To explore why the gate leakage current I_g has less impact on I_b, we compare the DC leakage I_b and I_{sd} in Fig.5.16, which is caused by I_g. At the V_g where the transient I_{sd} peaks in Fig.5.14(a), the leakage is well over two orders of magnitude higher than that where the transient I_b peaks in Fig.5.15. This is because the work function difference is against electron tunneling to the gate from silicon valence band under 1 V>V_g>0V where I_b peaks, but it assists the channel hole tunneling to the gate under -1 V<V_g< 0V where I_{sd} peaks, as illustrated by the inset of Fig.5.16.
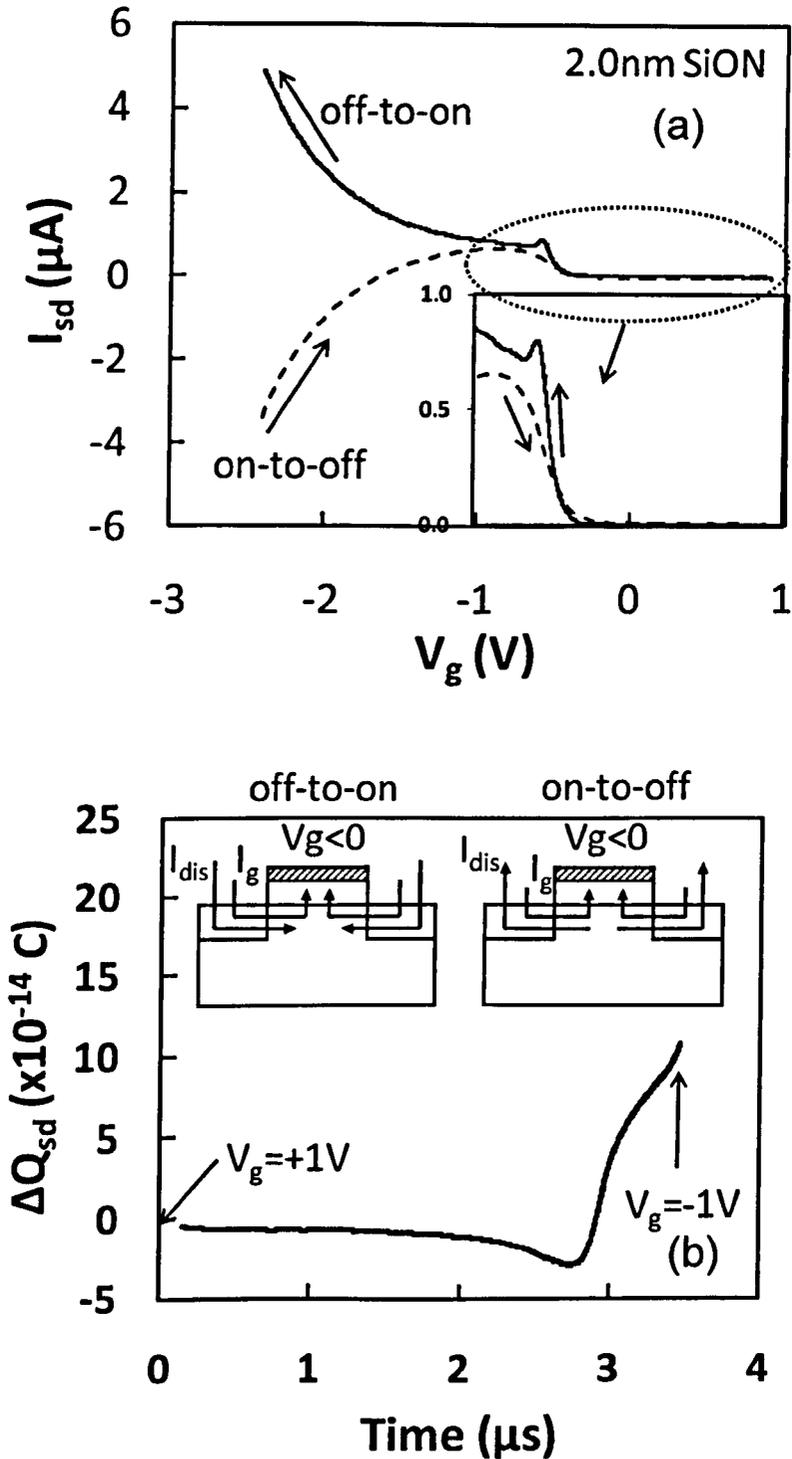
Fig.5.14 (a) The channel current for a 2nm SiON. The divergence at high negative V_g is caused by gate leakage. (b) shows that the net charges pumped into the device from the source/drain do not give a clear saturation level. The inset of (b) shows that the gate leakage current from the channel is in the same direction as the channel displacement current during the off-to-on transition, while they are in the opposite directions during the on-to-off switching.
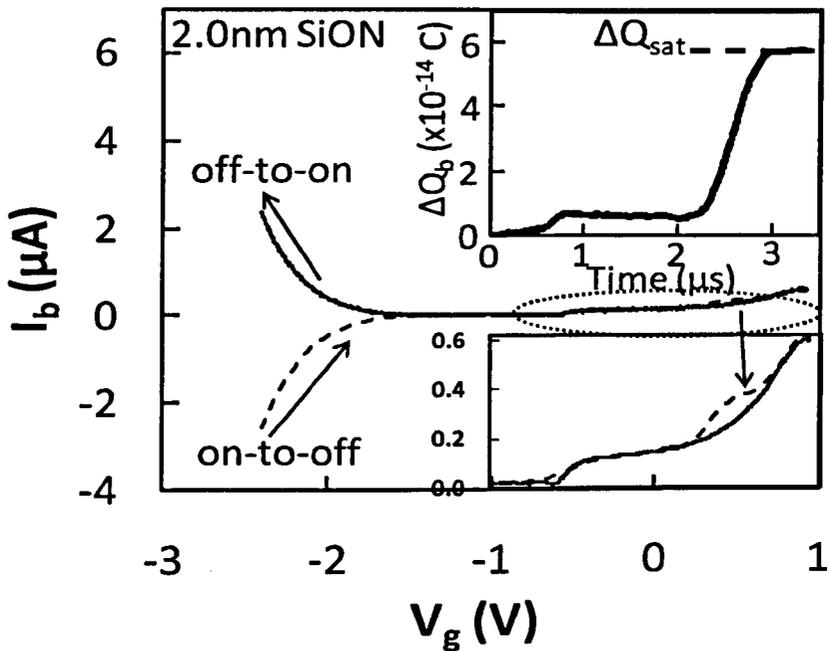
Fig.5.15 The substrate current for a 2nm SiON corresponding with Fig.5.13(a). The divergence at high negative V_g is caused by gate leakage. The inset shows that the two I_b merges after the peak at positive V_g and the net charges pumped into the device saturate.
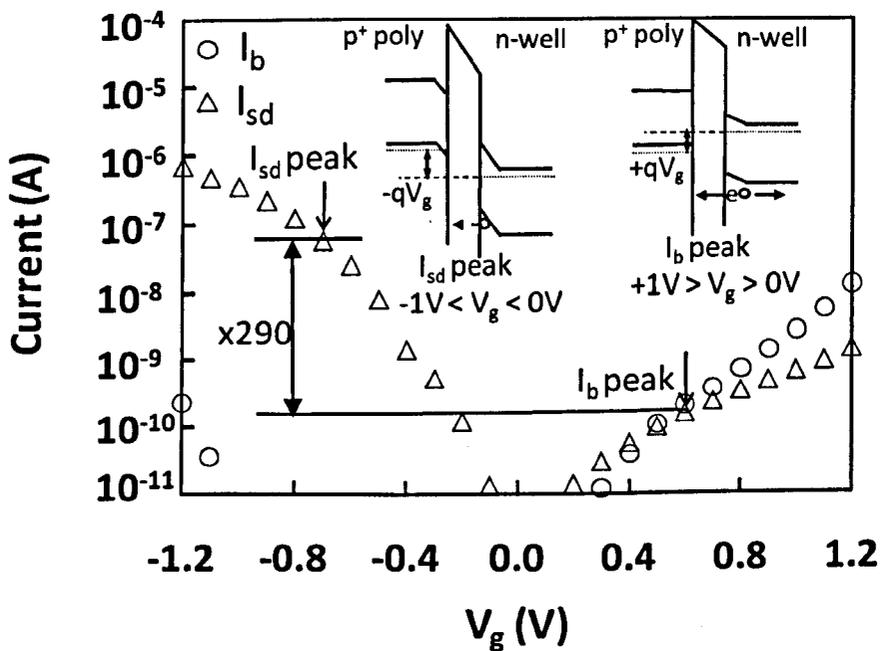
Fig.5.16 A comparison of channel and substrate DC leakage currents. The 'I_{sd} peak' and 'I_b peak' represent the V_g position where the I_{sd} and I_b reach their peaks in Figs. 5.13a & 5.14, respectively. The inset shows that the work function difference impedes electron tunnelling at I_b peak, but assist hole tunnelling at I_{sd} peak, resulting in lower leakage at the I_b peak.

The result for an HfSiON (1 nm)/SiON (1 nm) stack is given in Fig.5.17. Here the two I_b can again be used to determine the net charge pumped and in turn the interface states. Fig.5.17 also shows that, after merging, the two I_b diverge again at longer time that corresponds to higher V_g and I_g. This leads to the up-swing of ΔQ_b after saturation in Fig.5.17. The conventional CP technique measures the total charges pumped and this up-swing will be included in evaluating N_{it}. For SPCP, however, Fig.5.17 allows the high leakage region being excluded in measuring N_{it}, so that the SPCP is less vulnerable to the gate leakage.
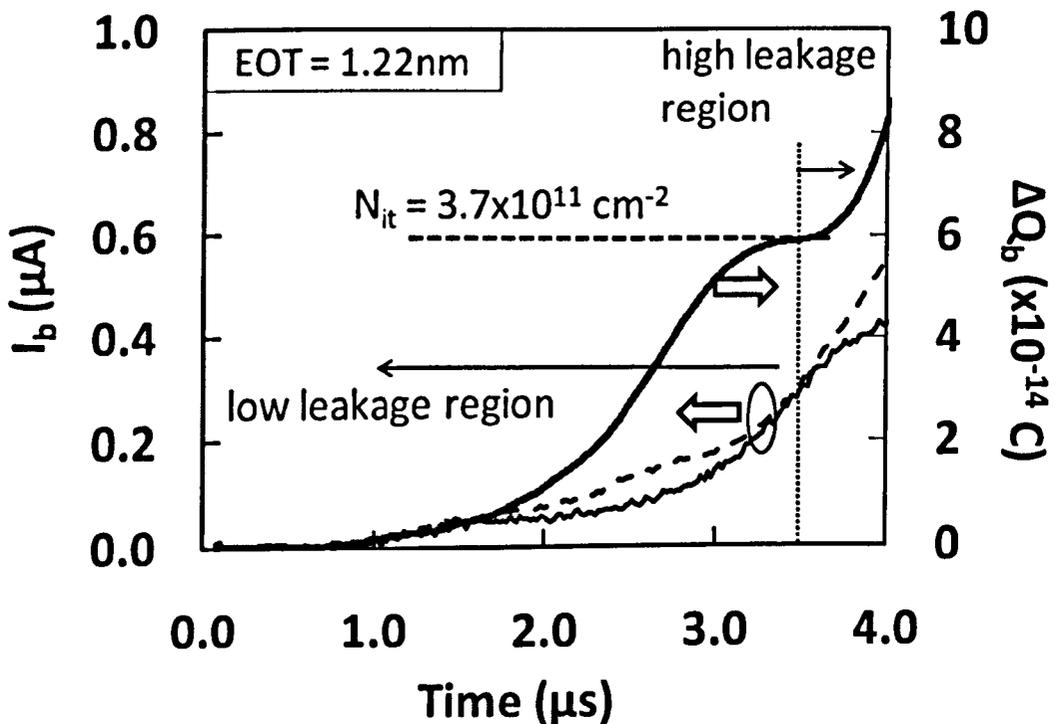
Fig.5.17 A demonstration of the applicability of SPCP for an HfSiON (1 nm) /SiON (1 nm) stack. N_{it} can be evaluated from ΔQ_b before the leakage becomes important.

Another advantage of SPCP over CCP is its better tolerance to the interference from defects in the dielectrics. For thin dielectrics, these defects in the dielectrics can communicate with the substrate mainly during the plateau of gate pulse [30], which contributes to the net pumped charges. In SPCP, however, the plateau region was not

used for measurements, so that the contribution from defects in the dielectrics is minimized.

5.9   Recovery of stress-induced interface states

Fig.5.6 shows that, when monitored by CCP, interface state recovery is apparently insignificant, in agreement with early work [26]. If this were true, it would mean that there were no needs for developing fast techniques for measuring interface states. The CCP measurement takes seconds and one possibility is that the recovery of stress-induced interface states is too fast to be captured by CCP. In the following, it will be shown that this is indeed the case.

The inset of Fig.5.18(a) shows the test procedure to capture the fast recovery. Unlike the early work [15] that requires sweeping V_g from stress level, V_g is limited to within $\pm 1V$ in the SPCP measurement here. The defects within the dielectric layer will be charged at the high negative stress level [31] and interfere with the measurement of interface states [30]. Through limiting V_g to within $\pm 1V$, this interference is suppressed for the SPCP measurement. After stress, the device is allowed to recover for a time of t_r before the SPCP measurement. Fig.5.18(a) gives the ΔQ_b with t_r as a parameter and Fig.5.18(b) plots the evaluated N_{it} against the total recovery time. To the best of our knowledge, for the first time, the recovery of interface states has been recorded with microseconds resolution. Within 8 μs, the recovery is negligible, justifying that the SPCP speed in the order of microseconds used in this work will capture the stress-induced states in full.
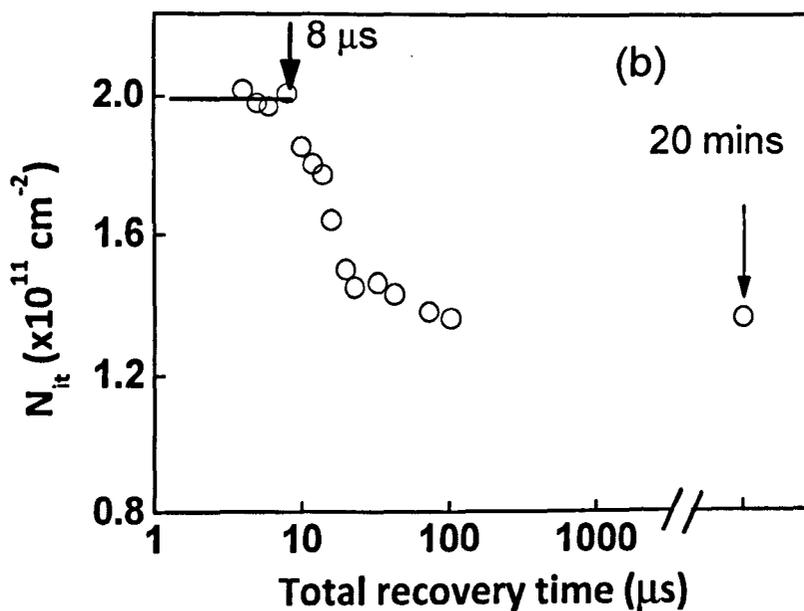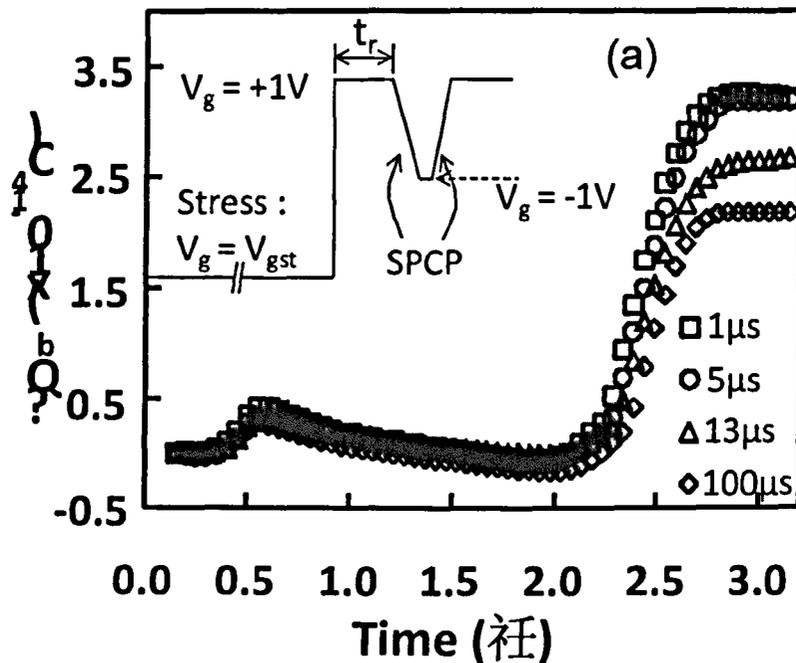
Fig.5.18 Recovery of stress-induced interface states at microseconds scale. The test procedure is given in the inset of (a).   After stress, recovery is allowed for a time of t_r. The N_{it} is then monitored by SPCP. (a) gives the net charges pumped into the device after different t_r. (b) shows the interface states evaluated against the total recovery time that is the sum of t_r and the time for ΔQ_b reaching saturation. The test sample is a 2nm SiON.

Fig.5.18(b) shows that the recovery of generated interface states becomes observable after a few microseconds. This agrees well with the early results of the ultra-fast measurements, where recovery of degradation in drain current and threshold voltage also starts in the order of microseconds after terminating stresses [31,37,38]. Fig.5.18(b) also shows that the recovery becomes insignificant after 100μs, in contrast with the continued recovery of drain current observed by the early work [31,38]. This difference can be explained by noting that the degradation in drain current originates from both the generated interface states and the positive charges trapped in the dielectric. It is well known that detrapping time increases exponentially with distance from the dielectric/substrate interface [10], leading to the continued recovery of drain current.

The negligible recovery of interface states between 100μs and 20min explains why recovery is insignificant when measured by CCP in Fig.5.6. Since CCP can take seconds, the recovery is essentially over well before the completion of CCP measurements. As a result, the rapid recovery within 100μs must be missed by CCP. The mechanism for this rapid recovery is not clear at present, but the SPCP technique opens the way for its exploration.

5.10   Conclusions

A fast single pulse charge pumping (SPCP) technique has been developed for characterizing interface states in this work to improve the measurement speed. It is based on the non-thermal equilibrium of interface states with the gate pulse, which gives rise to peaks in the channel and substrate currents and is responsible for the

differences in the currents corresponding to the two edges of gate pulse. These differences are used to evaluate the net charges pumped into the device and it has been shown that interface states can be evaluated from the saturation level of the net charges. Although in principle either channel or substrate current can be used, the latter is preferred for thin dielectrics, since it suffers less from gate leakage. The SPCP has successfully reduced the measurement time to microseconds from the seconds needed by the conventional CP, CV, conductance and subthreshold swing techniques. It is also more tolerant to the interferences from gate leakage and defects within the dielectrics, since the contribution from the plateaus of the pulse to the measurement is excluded. By using SPCP, the recovery of stressed-induced interface states was examined with a time resolution in microseconds for the first time. It is found that the recovery is substantially within 100µs, which would be missed if the conventional techniques were used. Further recovery beyond 100µs, however, is insignificant.

References

1. A. Stesmans and V. V. Afanas'ev, "Electrical activity of interfacial paramagnetic defects in thermal (100) Si/SiO$_2$," *Phys. Rev. B.*, **57**, 10030 (1998).

2. D. J. DiMaria, E. Cartier and D. Arnold, "Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon," *J. Appl. Phys.*, **73**, 3367 (1993).

3. J. F. Zhang, H. K. Sii, R. Degraeve, and G. Groeseneken, "Mechanism for the generation of interface state precursors," J. Appl. Phys., 87(6), pp.2967-2977 (2000)

4. W. D. Zhang, J. F. Zhang, M. J. Uren, G. Groeseneken, R. Degraeve, M. Lalor, and D. Burton, "On the interface states generated under different stress conditions," *Appl. Phys. Lett.*, vol.79, pp.3092-3094, 2001.

5. N. S. Saks, C. M. Dozier and D. B. Brown, "Time-dependence of interface trap formation in MOSFETs following pulsed irradiation," IEEE Trans. Nucl. Sci., **35**, 1168 (1988).

6. R. E. Stahlbush, A. H. Edwards, D. L. Griscom, and B. J. Mrstik, "Post-irradiation cracking of H2 and formation of interface states in irradiated metal-oxide-semiconductor field-effect-transistors," J. Appl. Phys., **73**, 658-667 (1993).

7. D. B. Brown and N. S. Saks, "Time-dependence of radiation-induced interface trap formation in metal-oxide-semiconductor devices as a function of oxide thickness and applied field," J. Appl. Phys., **70**, 3734 (1991).

8. C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-electron-induced MOSFET degradation --- Model, Monitor, and Improvement," IEEE J. Solid-State Cir., **SC-20**, 295 (1985).

9. J. F. Zhang and W. Eccleston, "Effects of high field injection on the hot carrier induced degradation of submicrometer pMOSFET's," IEEE Trans. Elec. Dev., **42**, 1269 (1995).

10. I. S. Al-kofahi, J. F. Zhang, and G. Groeseneken, "Continuing degradation of SiO2/Si interface after hot hole stress," J. Appl. Phys., vol. 81, pp.2686-2692, 1997.

11. J. F. Zhang and W. Eccleston, "Positive bias temperature instability in MOSFET's," IEEE Trans. Elec. Dev., **45**, 116 (1998).

12. H. Ushizaka and Y. Sato, "The process dependence on positive bias temperature aging instability of p+(B) polysilicon-gate MOS devices," IEEE Trans. Electron Devices, vol. 40, p. 932 (1993).

13. B. E. Deal, M. Skalar, A. S. Grove and E. H. Snow, "Characteristics of the surface-state charge (Qss) of thermally oxidized silicon," J. Electrochem. Soc., **114**, 266 (1967).

14. 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., **48**, 2004 (1977).

15. M. F. Li, D. Huang, W. J. Liu, Z. Y. Liu, and X. Y. Huang, "New insights of BTI degradation in MOSFETs with SiON gate dielectrics," *ECS Trans.*, vol. 19, pp.301-318, 2009.

16. J. F. Zhang, M. H. Chang, Z. Ji, L. Lin, I. Ferain, G. Groeseneken, L. Pantisano, S. De Gendt, and M. M. Heyns, "Dominant layer for stress-induced positive

charges in Hf-based gate stacks," *IEEE Electron Device Letters,* vol. 29, pp.1360-1363, 2008.

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

18. S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang and L. Chen, "Nitrogen-enhanced negative bias temperature instability: an insight by experiment and first-principle calculations," Appl. Phys. Lett., **82**, 1881 (2003).

19. C. Z. Zhao, J. F. Zhang, G. Groeseneken, R. Degraeve, J. N. Ellis, and C. D. Beech, "Interface state generation post-hole injection," J. Appl. Phys., vol.90, no.1, pp.328-336 (2001).

20. J. F. Zhang, C. Z. Zhao, G. Groeseneken, and R. Degraeve, "Analysis of the kinetics for interface state generation following hole injection," J. Appl. Phys., vol.93, no.10, pp.6107-6116 (2003).

21. J. F. Zhang, H. K. Sii, G. Groeseneken, and R. Degraeve, "Degradation of oxides and oxynitrides under hot hole stress," *IEEE Trans. Electron Devices*, vol. 47, pp.378-386, 2000.

22. L. M. Terman, "An investigation of surface states at a silicon/silicon oxide interface employing metal-oxide-silicon diodes," *Solid-State Electron,* vol.5, pp.285-299, 1962.

23. E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology* (New York: Wiley, 1982).

24. M. J. Uren and K. M. Brunson, "An improved technique for the evaluation of surface Fermi energy in metal-oxide-semiconductor capacitors," *Semicond. Sci. Technol.,* vol.9, pp.1504-1510, 1994.

25. C. Tan, M. Xu, and Y. Wang, "Application of the difference subthreshold swing analysis to study generation of interface traps in MOS structures due to Fowler-Nordheim aging," *IEEE Electron Dev. Lett.*, vol. 15, pp. 257-259, 1994.

26. M. H. Chang and J. F. Zhang, "On positive charges formed under negative bias temperature stresses", *J. Appl. Phys.*, vol.101, Art.no.024516, 2007.

27. J. R. Brews, "Subthreshold behavior of uniformly and nonuniformly doped long-channel MOSFET", IEEE Trans. Elec. Dev., **26**, 1282 (1979).

28. G. Groeseneken, H. E. Maes, N. Beltran, and R. F. De Keersmaecker, *IEEE Trans. Electron Devices*, vol. ED-31, pp.42-53, 1984.

29. P. Heremans and J. Witters, "Analysis of the charge pumping technique and its application for the evaluation of MOSFET degradation", IEEE Trans. Elec. Dev., **36**, 1318 (1989).

30. M. B. Zahid , R. Degraeve, J. F. Zhang, G. Groeseneken, "Defects Generation in SiO$_2$/HfO$_2$ Studied With Variable t$_{charge}$-t$_{discharge}$ Charge Pumping (VT^{2}CP)," in *Proc. of Int. Reliability Phys. Symp. (IRPS)*, pp.55-60, 2007.

31. 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 Dev.*, vol. 56, pp. 1086-1093, 2009.

32. Z. Ji, L. Lin, J. F. Zhang, B. Kaczer, and G. Groeseneken, "NBTI lifetime prediction and kinetics at operation bias based on ultrafast pulse measurement," *IEEE Trans. Electron Dev.*, vol. 57, pp. 228-237, 2010.

33. J. F. Zhang, C. Z. Zhao, G. Groeseneken, and R. Degraeve, "Analysis of the kinetics for interface state generation following hole injection," *J. Appl. Phys.*, vol.93, pp.6107-6116, 2003.

34. G. Van den bosch, G. Groeseneken, and H. E. Maes, "On the geometric component of charge-pumping current in MOSFET's," *IEEE Electron Device Letters,* vol. 14, pp.107-109, 1993.

35. G. A. Scoggan and T. P. Ma, "Effects of electron-beam radiation on MOS structures as influenced by the silicon dopant," *J. Appl. Phys.,* vol.48, pp.294-300, 1977.

36. J. F. Zhang, S. Taylor, and W. Eccleston, "Electron Trap Generation in Thermally Grown Silicon Dioxide Under Fowler-Nordheim Stress," *J. Appl. Phys.,* vol.71, pp.725-734, 1992.

37. S. Deora, V. D. Maheta, A. E. Islam, M. A. Alam and S. Mahapatra, "A common framework of NBTI generation and recovery in plasma-Nitrided SiON p-MOSFETs," *IEEE Electron Dev. Lett.,* vol. 30, pp.978-980, 2009.

38. T. Yang, M. F. Li, C, Shen, C. H. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, S. C. Rustagi, M. B. Yu, and D. L. Kwong, "Fast and slow dynamic NBTI components in p-MOSFET with SiON dielectric and their impact on device life-time and circuit application," in *VLSI Symp. Tech. Dig.,* 2005, pp. 92-93.

6 | Conclusions and Future Work

6.1 Conclusions

The work in this project has been focused on the negative bias temperature instabilities (NBTI) and the positive charges responsible for them. Chapter 1 reviewed the important reliability issues and defects responsible for them and explained the rationale for the selection of research topics. Chapter 2 described the test facilities and efforts made to improve the measurement speed from 5μs to 200ns for Id-Vg measurements and 800ns for the C-V measurements. The main research works are divided into three parts and covered in the next three chapters: Chapter 3 on the dominant layer for positive charges in Hf-dielectric/SiON stack; Chapter 4 on the effective threshold voltage shift and its use for a single test lifetime prediction method; and Chapter 5 on the development of a fast single pulse charge pumping method. Conclusions for each part are given below:

6.1.1 The dominant layer for the positive charges in Hf-based dielectric stacks

It has been reported that the same three types of positive charges exist in both SiON and Hf-based stacks: cyclic positive charges, anti-neutralization positive charges, and as-grown hole traps. Since the presence of Hf-dielectric layer has not introduced

any new type of positive charges, it is possible that positive charges in the Hf-stack are dominated by the interfacial SiON layer. What is missing is clear experimental evidence to confirm this possibility. This is chosen as the first topic, because, if the positive charging in the Hf-based stack is also dominated by the interfacial SiON layer, the rest of the project can then be concentrated on SiON samples that are relatively plentiful, in anticipation that the conclusion is also applicable to the Hf-based stack.

By varying the thickness of HfSiON and keeping a fixed interfacial layers, the obtained test result clearly do not agree with either the assumption that positive charges are distributed through the stack or the assumption that positive charges are dominated by the bulk of Hf-dielectric layer. This means that the positive charges in the stack can only be dominated by the interfacial layer. If the positive charges were located at the Hf-dielectric/IL interface, the threshold voltage shift should be independent of the IL thickness. This is, however, against the observation that ΔV_{th} increases with the IL thickness. The results support that positive charges are located close to the IL/substrate interface. Consequently, unlike electron trapping, a reduction of HfSiON thickness will not reduce positive charges and NBTI remains as an important reliability issue for future CMOS technologies.

6.1.2 The effective threshold voltage shift and its application in lifetime prediction

Evaluating NBTI-induced mobility degradation is problematic, controversial, and undesirable for test engineers. Central to this work is to propose using an effective threshold voltage shift, ΔV_{eff}, as the single parameter for characterizing NBTI,

which fully takes account of any potential contribution from mobility degradation. By assuming that mobility does not change under a constant (Vg-Vth), the evaluation of mobility variation is avoided. A method for extracting ΔVeff has been proposed, which only requires standard NBTI tests and can be easily implemented. The results show that Id degradation under operation gate bias predicted from ΔVeff is in good agreement with the measured data, but Id degradation is substantially underestimated or overestimated if the extrapolated ΔVth or the 'on-the-fly' ΔVth is used, respectively. ΔVeff is preferred over ΔId/Id for characterizing NBTI, since it does not depend on device size and source/drain series resistance.

Based on ΔVeff, the NBTI lifetime prediction is investigated for the worst case scenario where the recovery is suppressed. In this case, the conventional Vg acceleration prediction is inapplicable, because the NBTI kinetics no longer follow a simple power law and an increase of stress bias does not lead to a parallel shift of $\log|\Delta V_{th}|$. To predict the lifetime, NBTI kinetics and defects are examined. An outstanding feature of the kinetics is the presence of a 'shoulder', which is insensitive to temperature and must be dominated by the charging of as-grown defects. The charging and discharging properties of the defect agree well with the signature of as-grown hole traps. By combining the first order model for the as-grown hole traps and the power law for generating new defects, ΔV_{th} can be modeled over ten orders of stress time. This kinetic model is then used to predict the NBTI lifetime, based on a single test at the operation temperature and bias. For the six different processes tested, the safety margin of the single test prediction technique is within 50%, which is substantially better than the methods proposed in early works.

6.1.3 A single pulse charge pumping technique for fast measurements of interface states

The existing techniques for characterizing interface states typically take several seconds and the degradation can recover substantially during the measurement. In this work, the potential of the established techniques for fast measurements is analyzed. The conductance, high-frequency CV, quasi-static CV, and subthreshold swing all requires a gate voltage sweep that must be sufficiently slow to maintain the thermal equilibrium of interface states with the gate bias. This makes them fundamentally unsuitable for fast measurements. In contrast, charge pumping does not need this thermal equilibrium and can be used for fast measurement in principle. The speed of conventional charge pumping method, however, is limited by measuring a DC charge pumping current.

In this work, a fast single pulse charge pumping (SPCP) technique has been developed for characterizing interface states to improve the measurement speed. It is based on the non-thermal equilibrium of interface states with the gate pulse, which gives rise to peaks in the channel and substrate currents and is responsible for the differences in the currents corresponding to the two edges of gate pulse. These differences are used to evaluate the net charges pumped into the device and it has been shown that interface states can be evaluated from the saturation level of the net charges. Although in principle either channel or substrate current can be used, the latter is preferred for thin dielectrics, since it suffers less from gate leakage. The SPCP has successfully reduced the measurement time to microseconds from the seconds. It is also more tolerant to the interferences from gate leakage and defects

within the dielectrics, since the contribution from the plateaus of the pulse to the measurement is excluded. By using SPCP, the recovery of stressed-induced interface states was examined with a time resolution in microseconds for the first time. It is found that the recovery is substantially within 100µs, which would be missed if the conventional techniques were used. Further recovery beyond 100µs, however, is insignificant.

6.2 Future works

Despite of the effort made in this project, many questions remain to be answered. These include, but are not limited to, the followings:

NBTI lifetime prediction with recovery:

In this project, a prediction method has been proposed only for the worst case scenario: zero recovery. In reality, different MOSFETs in a circuit will experience different levels of recoveries. For the quasi-DC measurement used in industry, a duty factor is introduced to take into account of the recovery. Since the NBTI kinetics will be different when recovery exists, the challenge is how to combine the model proposed in this project with the classical method to take recovery into account when predicting device lifetime in the future.

Defects responsible for the fast recovery of interface states:

Although the work shows that there is a fast recovery phase for the interface states, little is known about the defect responsible for it. Traditionally, interface states are

generally believed to originate from Pb centers. If there is only one type of interface states, one would expect that the recovery will continue as time increases. However, test results show that the recovery stopped abruptly after 100 µs. This raises the possibility that the recoverable interface states are different from the non-recoverable ones. Further work is needed to explore this issue.

NBTI for nanometer MOSFETs:

The device sizes, such as 10 µm, used in this work are relatively large to increase the signal strength. For the devices in tens of nanometers, however, there can be only a few of defects created during the stress. There will be large device-to-device variations and challenge the classical definition for device lifetime. Further work is clearly needed in this area.

NBTI for multi-gate MOSFETs and nano-wire MOSFETs:

To increase the gate control on the substrate and reduce the drain-induced barrier lowering leakage current, multi-gate MOSFETs and nano-wire MOSFETs were investigated. There is little information available on their NBTI properties and one can expect that the device-to-device variations can be significant again. It is a question whether the lifetime prediction technique developed earlier is applicable to these devices.

NBTI for other dielectric/semiconductor structures:

As the downscaling of silicon based MOSFETs approaches its end, a lot of attentions have been paid to the further improvement of transistor speed without downscaling its physical sizes. Both Germanium and III-V semiconductor MISFETs have been

intensively investigated. The dielectric stack used in these transistors may not involve SiON and there is little data available on their instability. It is not clear whether NBTI will still be the most important reliability issue and what will be the physical process controlling the lifetime of devices. Will these MISFET of high mobility semiconductors more reliable than Si MOSFETs? Will they be ever stable enough to be used to build microprocessors?

List of publications

1. J. F. Zhang, M. H. Chang, Z. Ji, L. Lin, I. Ferain, G. Groeseneken, L. Pantisano, S. De Gendt, and M. M. Heyns

   "Dominant layer for stress-induced positive charges in Hf-based gate stacks"

   IEEE Electron Device Letters, vol. 29, pp.1360-1363, 2008.

2. Z. Ji, L. Lin and J. Zhang

   "Impact of sensing gate bias on NBTI of Hf-based dielectric stacks"

   ECS transactions on Silicon Nitride, Silicon Dioxide, and Emerging Dielectrics 10, Vol. 19, No.2, pp.351-362, 2009.

3. J. F. Zhang, Z. Ji, L. Lin, and W. Zhang (Invited)

   "Effective threshold voltage shift: a measure for NBTI removing uncertainty in mobility Degradation"

   Proc. of IEEE 10th International Conference on Solid-State and Integrated-Circuit Technology, pp.1600-1603, 2010.

4. Z. Ji, L. Lin, J. F. Zhang, B. Kaczer, and G. Groeseneken

   "NBTI lifetime prediction and kinetics at operation bias based on ultrafast pulse measurement"

   IEEE Trans. Electron Dev., vol. 57, No.1, pp. 228-237, 2010.

5. L. Lin, Z. Ji, J. F. Zhang, W. D. Zhang, B. Kaczer, S. De Gendt, and G. Groeseneken

   "A single pulse charge pumping technique for fast measurements of interface states"

IEEE Trans. Electron Dev., vol. 58, No.5, pp. 1490-1498, 2011.

6. L. Lin,   Z. Ji, J. F. Zhang and W. D. Zhang

    "Development of a fast technique for characterizing interface states"

    ECS transactions on Silicon Nitride, Silicon Dioxide, and Emerging Dielectrics

    11, Vol. 35, No.4, pp.81-93, 2011.

7. J. F. Zhang, Z. Ji, W. Zhang, L. Lin,   B. Kaczer, S. De Gendt, and G.

    Groeseneken (Invited)

    "Advanced characterization techniques for MOS devices"

    China Semiconductor Technology International Conference (CSTIC), 2012.