Characterization of negative bias temperature instability and lifetime prediction for pMOSFETs

Zhigang Ji

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

June 2010

Abstract

In order to achieve high speed and packing density, the size of the transistor has shrunk aggressively. The gate dielectric, as the most critical component in the transistor, is undergoing rapid and substantial changes with the adoption of ultra-thin plasma nitrided oxide and more recently high-k dielectrics. As the nitrogen concentration in silicon oxynitrides (SiON) increases, the negative bias temperature instability (NBTI) rises and becomes a limiting factor for device lifetime.

The NBTI can recover significantly during typical measurement time when using conventional instruments. To suppress this recovery, several fast techniques have been developed, including ultra-fast pulse I_d-V_g technique and the On-The-Fly technique. These techniques, however, give different threshold voltage degradation (ΔV_t) after the same stress. The interpretation of this difference is still controversial. The objective of chapter 3 is to bridge the gap between the ΔV_t extracted from these techniques. Degradation and recovery during measurement, measurement and truncation errors, and evaluation of transconductance are examined. After taking these factors into account, the gap in ΔV_t still cannot be filled. The effect of the sensing V_g on ΔV_t is considered and it is found that $|\Delta V_t|$ increases with sensing $|V_g|$. The popular assumption of

ΔV_t being independent of sensing V_g is invalid, thereafter. After taking both the effect of sensing V_g and recovery into account, the gap in ΔV_t is successfully bridged. The difference between the effect of sensing V_g and recovery is explored, and the results show that they are two different phenomena.

The recovery suppression and the sensing V_g effect challenge the applicability of the traditional lifetime prediction technique. In a large circuit with roughly 10^6 MOSFETs, there will always be some of them under the worst case condition, namely constant stress without recovery. The failure of one of these MOSFETs can lead to the malfunction of the whole circuit. At present, there is little information on how this worst case NBTI lifetime can be predicted and whether the traditional V_g acceleration technique can be applied. In chapter 4, the worst case lifetime prediction is investigated. It is found that the prediction based on the V_g acceleration results in substantial errors. To predict the worst-case lifetime, a model for NBTI kinetics under operation gate bias is 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 is proposed and its safety margin is estimated to be 50%.

Mobility reduction is another important issue when oxide thickness becomes thinner. It is reported that when the gate SiON becomes thinner than 2 nm or the interfacial layer in high-k stack is thinner than 2.5 nm, carrier mobility reduces. Agreement has not yet been reached on the level of reduction, or on

the underlying mechanism. Remote charge scattering (RCS) has been proposed to be responsible for this mobility reduction. However, one weakness of earlier work is that different samples were used when experimentally studying the RCS and this introduces uncertainties. For example, a reduction in oxide thickness does not only bring the gate closer to the substrate, but also modulates other factors such as surface roughness. In chapter 5, the importance of RCS is assessed by varying charge in the same device through either processing or electron trapping, to remove the uncertainties from using different devices. It is found that by increasing charge density at 0.56 ~ 1 nm from the substrate interface to the order of 10^{20} cm^{-3}, both electron and hole mobility change little.

Acknowledgements

I would like to thank my supervisor, Prof. J. F. Zhang for all the help and continuous guidance, advice, encouragement and fruitful discussions throughout the project. Without his investment of time and efforts, the project would never reach the fruition.

I also deeply appreciate Dr. Wei for his advice and support throughout the project. Many thanks are given to other members of our Microelectronic research group, Mr. X. F. Zheng, Dr. M. H. Chang, Mr. L. Lin and Mr. C. Robinson for their discussion.

Finally, I would like to express my deep gratitude to my wife, Mrs. J. Huang, for her love and support and my parents and all my family for their encouragement during all these years.

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
HCI	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	°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

Table of contents

1 Introduction

1.1 Challenges for scaling CMOS devices

Moore's law [1-5] has been the guiding principle for the integrated circuit (IC) industry since 1965. In order to follow the Moore's Law (≈16% compound annual growth rate), device scaling is the key attribute for planar MOSFET circuitry based on the work of Dennard et al [6-8].

From 1970s to mid 1990s, the scaling of MOSFETs largely followed the constant voltage scaling rule. As the hot-carrier induced reliability issue became difficult to handle, the supply voltage was reduced as the MOSFET scales, and the constant field scaling is now used. Since mid 90s, the semiconductor industry association (SIA) started to publish technology roadmaps, which include an outlook of future scaling of MOSFET technology. It later became an international effort as the International Technology Roadmap for Semiconductors (ITRS) [9]. The scaling trend projection in ITRS is determined from transistor performance targets and power dissipation constraints, together with a sophisticated compact model of MOSFET transistors. In all scaling

schemes, the gate oxide thickness scales down with the transistor feature size at a steady rate.

Decreasing device feature sizes leads to short channel effects and reduced gate control. To suppress the short channel effects, an equivalent gate oxide thickness (EOT) of less than 1 nm is required for a 70 nm channel length high performance devices, as proposed in 2006. At this thickness of SiO_2, devices are hampered by increased gate leakage, questionable reliability, large poly-silicon depletion effects, and boron penetration into the channel.

Poly-silicon depletion can decrease the capacitance of the device and becomes significant as the EOT is decreased. Increasing the doping does minimize influencing the depletion capacitance, but this leads to greater boron penetration and threshold voltage shifts. Furthermore, the limit of poly-silicon doping is ~10^{21} cm^{-3} which may not be enough to recover the capacitance. It is suggested that replacing the gate electrode with metals, which have at least an order of magnitude higher carrier concentration and do not suffer from depletion, will tackle these issues. Many investigations are ongoing concerning alternative metal gate electrodes with work on titanium, tantalum, ruthenium, tungsten as well as their nitrides, silicates, and alloys [10-14]. Metal gate electrodes have been successfully incorporated into the CMOS process.

Gate leakage current reduces the on-off current ratio of the device, creating heat and possibly breaking down the device. While increasing the thickness of the SiO_2 will decrease this leakage, the capacitance and consequently the speed and drive capability of the device will decrease. The efforts are made

mainly in two areas. First, in order to decrease the leakage current while maintaining an appropriate gate capacitance, SiO_2 is replaced by a material possessing a higher dielectric constant denoted as high-k dielectric. This is because the dielectric constant, k, is a measure of an insulator's ability to concentrate an electric field. If one gate oxide has twice the dielectric constant of another, a given voltage will draw twice as much charge into the transistor channel or the same amount of charge will accumulate, if the higher k dielectric is made twice as thick. There are many difficulties, however, with high-k dielectric, including process integration, lower carrier mobility, threshold voltage instability, and poor yield [15]. Although these problems delayed the introduction of high-k dielectrics, high-k dielectrics were successfully used in production for CMOS technology in 2007.

The second area is focused on developing new device structures, such as Ultra-thin Body MOSFET, Double Gate MOSFETs (including FinFET), and various Multi-gate MOSFETs [16]. The short-channel effects are suppressed and relatively thick dielectrics can be used in these new transistors. For both SiON and high-k layers, Negative Bias Temperature Instability (NBTI) is a severe problem.

1.2 New materials in advanced gate dielectrics

As one of the solutions to suppress gate leakage current, an alternative gate dielectric has been sought. The requirements of this new dielectric include: an increase in the dielectric constant (k) [17], a large band gap with high barrier heights to both electrons and holes [18], and compatibility with conventional

planar CMOS processing. There are many potential candidates for high-k gate dielectrics with k values ranging from 7 ~ 80. These candidates include: Si_3N_4, Al_2O_3, Y_2O_3, La_2O_3, Ta_2O_5, TiO_2, ZrO_2, and HfO_2 [19]. However, TiO_2 and Ta_2O_5 have low thermal stability which makes them incompatible with conventional CMOS processing [19]. Al_2O_3 alone, which has a k value of 9, may not be sufficient to achieve sub 1 nm equivalent oxide thickness (EOT) values. However, when incorporated into HfO_2 as an Hf-aluminate, the film enhances scalability toward sub-nm EOT applications [20]. Although scaling is possible, other device parameters such as transconductance are degraded when comparing Hf-aluminates to conventional HfO_2 [21]. La_2O_3 and Y_2O_3 have higher k values of 30 and 15, respectively, but lose their k values in conventional processing due to silicate formation [22]. La_2O_3 could still be a possible candidate with continued process optimization. Hf-based dielectrics have eventually become the winner as the first high-k dielectric used in CMOS technology [20, 23-32].

1.3 Negative Bias Temperature Instability (NBTI)

Negative Bias Temperature Instability (NBTI) is not a new problem. It is actually one of the earliest instabilities indentified for metal-oxide-semiconductor field effect transistors (MOSFETs). NBTI occurs in pMOSFETs stressed with negative gate voltage at elevated temperatures. It manifests itself as absolute drain current I_{dsat}, and transconductance g_m, decrease and absolute "off" current I_{off}, and threshold voltage V_t increase. Typical stress temperatures are in the range of 100 ~ 250 °C. The oxide electric fields used to be below 6 MV/cm, but much higher field has been used in more recent work. Such field and

temperature are typically encountered during burning-in, but are also approached in high performance ICs during routine operation. Either negative gate voltages or elevated temperatures can produce NBTI, but a stronger and faster effect is produced by their combined action. It occurs primarily in pMOSFETs with negative gate voltage bias and appears to be smaller for positive gate voltage and for either positive or negative gate voltage in nMOSFETs. In MOS circuits, it occurs most commonly during the high output state of inverter operation. It also leads to timing shifts and potential circuit failure due to increased spreads in signal arrival in logic circuits. Asymmetric degradation in timing paths can lead to non-functionality of sensitive logic circuits and hence lead to product failure.

NBTI was first studied by Deal in 1967 [33] and later was studied on FAMOS and EEPROMs [34]. Deal named it Instability Number VI [35]. The original technological motivation for NBTI studies was the shift of threshold voltage of p-channel MNOS [35,36] or FAMOS non-volatile memories [34]. In 1973, Goetzberger et al at Bell lab were one of the first groups to show detailed characterization of NBTI [36]. They used metal gate devices on 100 nm oxide, stressed at -10^6 V/cm at 300 °C and found an interface trap density D_{it} peak in the lower half of the band gap. The higher the starting D_{it}, the higher the final stress-induced D_{it}. For positive gate voltage, they noted a very small D_{it} increase.

In 1977, Jeppson and Svensson [37] investigated the kinetics of NBTI. They observed that the defect generation follows a power law with a power factor in the range of 0.2 -0.3. The physical process responsible for NBTI includes an

electrochemical reaction at the interface, followed by a diffusion of hydrogenous species. The interfacial reaction leads to a rupture of hydrogen bonds and the release of hydrogenous species. It is believed that the reaction is relatively fast and the generation rate is limited by the subsequent diffusion process.

Though NBTI is not important on buried channel (BC) pMOSFETs [38] because of the naturally reduced oxide field for the same gate voltage compared to surface-channel devices, the introduction of CMOS technology and the dual poly process finally replaced the buried channel PMOS devices. Although the circuit performance is improved, the NBTI performance is actually worse [39].

NBTI is not the only type of electrical stress induced instability. Other instabilities include hot carrier induced instability, Fowler-Nordheim or direct tunnelling induced damage and the time-dependent-dielectric-breakdown (TDDB). Compared with these instabilities, NBTI received less attention until about 7 or 8 years ago when oxide dielectrics became thinner. NBTI is becoming increasingly important in recent years, mainly for two reasons. First, for each new generation of CMOS process, both operation temperature and electrical field increase. Second, to suppress the boron penetration and increase the dielectric constant, the nitrogen density in the gate oxide is increasing rapidly. Nitridation enhances NBTI and positive charge formation [40-42]. For a gate oxide thinner than 3.5 nm, the NBTI replaces hot carrier induced degradation as the limiting mechanism for device lifetime.

Schroder and Babcock gave a comprehensive review on the subject in early 2003 [43], which summarized the understanding on NBTI at the time, its origin,

its effect on transistors, and most importantly its dependence on process and device parameters. Research on NBTI prior to 2003 largely followed the paradigm set in the 1970s.

A typical NBTI study involves stressing a MOS device at an elevated temperature under negative gate voltage. In the earlier work [37], the gate voltage was chosen so that the electric field in the oxide is less than about 6 MV/cm to avoid degradation due to Fowler-Nordheim tunneling. After a preset stress time, the sample is cooled down to room temperature and its device characteristics are measured. Typically the parametric shift in threshold voltage, flat-band voltage and linear drain current are recorded. Generation of interface states is commonly measured by one of the many available techniques, while the accompanying generation of fixed charge is calculated indirectly. This stress-measure cycle is repeated and extends to, typically, 10^3–10^5 s.

1.3.1 NBTI models

Various NBTI models have been proposed, of which the Reaction–Diffusion (R–D) model is the most prevalent [37,44-52]. It was first proposed by Jeppson and Svensson [37]. In their model, it is assumed that the silicon interface contains a large number of defects. Those defects are electrically inactive and can be activated through chemical reaction like:

$$Si_3 \equiv SiH + O_3 \equiv SiOSi \equiv O_3 \rightleftharpoons Si_3 \equiv Si\cdot + O_3 \equiv Si^+ + O_3 \equiv SiOH + e^-$$

Where $Si_3 \equiv SiH$ is the surface defect, $Si_3 \equiv Si\cdot$ is the surface trap, $O_3 \equiv Si^+$ is the oxide charge. When the defect is activated, the H of the SiH bond is

released by some dissociation mechanism and reacts with the SiO_2 lattice to form an OH group bonded to an oxide atom, leaving a trivalent Si atom in the oxide to form a fixed charge and one trivalent Si atom at the Si surface constituting an interface state. The rule $N_{it} \sim t^{1/4}$ was observed and mathematically proven by assuming the process is diffusion rather than reaction-rate limited.

The model described above is called the Electro-chemical model. Since then, many variants have been proposed by assuming different dissociation mechanisms to meet the observations from the experiments. Two popular ones are given below.

High-electric field dissociation

A hydrogen terminated interface trap precursor can be represented by,

$$Si_3 \equiv SiH$$

High electric field can dissociate the silicon-hydrogen bond, leading to:

$$Si_3 \equiv SiH \rightarrow Si_3 \equiv Si \cdot + H^0$$

where H^0 is a neutral interstitial hydrogen atom or atomic hydrogen. Atomic hydrogen is highly reactive and considered to be a fast diffuser in oxide [53]. The availability of SiH bonds for dissociation under high electric field is the rate-limiting process for this reaction. Recent first-principles calculations show that the positively charged hydrogen or proton (H^+) is the only stable charge state of hydrogen at the interface. H^+ reacts directly with SiH to form an interface trap, according to the reaction [54]:

$$Si_3 \equiv SiH + H^+ \rightarrow [Si_3 \equiv Si]^+ + H_2$$

The above reaction uses the fact that the SiH complex (or passivated dangling bond chemical species) is polarized such that a more positive charge resides near the Si atom and a more negative charge resides near the hydrogen atom. Mobile H^+ migrates towards the negatively charged dipole region of the SiH molecule, then reacts with the H^- to form H_2, leaving behind a positively charged Si dangling bond. In this model, H_2 can later dissociate to act as a catalyst to disrupt additional SiH bonds. This process, in theory, can continue so long as hydrogen is available and SiH bonds are available to react. The reaction between hydrogenous species (H^+ or H_2) and SiH bonds is the rate-limiting process for this reaction model.

Hole induced dissociation

This model explains NBTI induced trap formation by the interaction of SiH with "hot holes" or holes near or at the Si/SiO_2 interface [43,49,55,56]. Dissociation involving holes is given by

$$Si_3 \equiv SiH + h^+ \rightarrow Si_3 \equiv Si \cdot + H^+$$

During NBTI stress, holes are attracted to the SiO_2/Si interface. This model is consistent with the results that a positive substrate bias accelerates NBTI degradation, reduces the device lifetime [42]. Hole-induced SiH bond dissociation is the rate-limiting process for this reaction model.

1.3.2 NBTI recovery

Earlier work treated the degradation as permanent. Since 2003, however, the NBTI community started to realize that the recovery effect is not negligible [57-61]. It was observed that NBTI degradation has appreciable recovery within a time less than 1 s. If an AC stress voltage is applied, the sample is stressed for half of the period in each cycle. During the other half-cycle, existing NBTI degradation can be healed. As a result, NBTI degradation is much less severe for samples under AC stressing than those under constant voltage stressing. In modern circuit design with roughly 10^6 MOSFETs, there will always be some devices under the worst case condition, namely stress is constantly applied most of the time. Failure of one MOSFET can potentially lead to the failure of the whole circuit. Therefore, it is crucial to model the device and predict its lifetime at its worst case condition [62].

The traditional stress-measure procedure is problematic because of the recovery effect. Cooling the wafer from the stress temperature to the measurement temperature (normally room temperature) takes minutes, and during this time, the NBTI degradation is substantially recovered. Even if the cooling procedure is eliminated and the measurement taken immediately after the electrical stress, the commonly used parametric analyzer still takes significant time, typically between 0.5 s to 10 s. It should be noted that the time stated in the specification from the manufacturer does not include the hold and delay time presets for the instrument. If these are included, the measurement time can even double.

Rangan et al were the first to propose a solution for investigating the recovery transient [63]. It involves converting the change in drain current ΔI_d, at stress bias into the change in threshold voltage, ΔV_t. Since the drain current is measured at stress bias and only one drain current point is measured, the procedure can be considered as recovery free. One problem of this solution is that when the change in drain current is converted into that in threshold voltage, the mobility degradation is ignored. In order to remove this uncertainty, an improved On-The-Fly (OTF) technique (the 2^{nd} order OTF technique) was proposed [64] to take mobility variation into consideration. However, since these researchers used slow measurement instruments, the non-negligible NBTI degradation during measurement itself seriously affects the result. To reduce the measurement time, customized instrumentation was needed. Kerber et al were the first to design their own customized circuit to implement fast measurements [65]. Since then, a number of variants have been proposed [66-70]. The basic principle for all these fast measurements is to combine a pulse generator with a digital oscilloscope. The pulse generator is used to supply the gate voltage, and the corresponding current will be converted to voltage and captured by the oscilloscope.

The suppression of recovery in the measurement invokes an alternative theory that accounts for the NBTI phenomenon. The NBTI is modeled by charge trapping/de-trapping. It is well known that injected carriers can be trapped in an insulator as they pass through it. Charge trapping in dielectrics under Fowler-Nordheim stress was extensively studied in the 1980s. Trapped charge in the gate dielectric of MOSFET will shift the threshold voltage V_t and the flat-band voltage V_{fb}. Positive charge trapping will increase the $|V_t|$ of pMOSFETs.

Ushio et al proposed that hole-trapping near the Si/SiON interface caused the enhanced NBTI in SiON gate dielectric [71,72]. After recovery was suppressed, more groups began to report hole trapping. Huard and Denais used their OTF technique to demonstrate that NBTI degradation can be entirely explained by hole trapping [64]. The group at LJMU further separated these traps into three different types [73].

Tan et al tried to reconcile the hole trapping with the R-D model and suggested that hydrogen released from broken Si-H bonds can be trapped at nitrogen sites at the interface to form fixed charge [40,74]. The role of hole trapping in SiON gate dielectric has been one of the most controversial topics in the study of NBTI.

Recent work shows that the ΔV_t obtained from customized instrumentation can be five times higher than that from conventional techniques. However, the ΔV_t obtained from the On-The-Fly technique can be one order of magnitude higher than that from conventional techniques and the interpretation of this large difference is quite controversial. The reconciliation of all the existing characterization methods is the subject in chapter 3.

The NBTI recovery model also brings question into the applicability of existing lifetime prediction techniques, which were developed without taking recovery time into account. This issue will be thoroughly investigated in chapter 4.

1.4 Carrier mobility degradation in state-of-the-art transistors

One of the most important parameters of a MOSFET is the channel mobility: a measure of the ease with which the inversion carriers travel along the channel. The mobility is sensitive to the quality of the channel, the dielectric, and even the gate electrode. Properties such as oxide trapped charge, lattice vibrations scatter the moving carriers and decrease their mobility, reducing the effectiveness of the device. Among the scattering mechanisms, coulomb scattering, phonon scattering, and surface roughness are the most important ones in CMOS devices.

Phonon Scattering

Phonon scattering of the inversion carriers originates from the acoustic and optical mode vibrations of the Si lattice. As the temperature increases, lattice vibrations also increase and thus the phonon scattering rate increases. At room temperature, phonon scattering plays an important role on the carrier mobility in inversion layers. It has been shown experimentally by Takagi et al that mobility limited by phonon scattering can be determined from [75],

$$\mu_{ph} = AE_{eff}^{-0.3}T^{-1.75}$$

where A is a constant with the value of 2×10^5 and 6.1×10^4 for electrons and holes on (100) Si, respectively, E_{eff} is the effective vertical field in silicon at the interface, and T is the temperature in Kelvin.

Coulomb Scattering

Scattering from charged centers such as fixed charge, interface states, and ionized impurities in the depletion layer of the substrate is designated as Coulomb scattering. As the vertical field increases and more carriers occupy the inversion layer, electrostatic screening occurs and Coulomb scattering therefore reduces [75]. This means that Coulomb scattering limited mobility is dominant at lower fields where the surface carrier density is small. An increase in doping concentration and fixed charge density enhances coulomb scattering. The coulomb scattering weakens as temperature increases, since the velocity of the inversion carriers increases and they interact less effectively with stationary impurities.

Surface Roughness

Since the silicon-insulator interface is not perfect, the surface potential can fluctuate on the molecular level and the small deviations of the interface from the ideal plane causes perturbations in the inversion carrier mobility. The surface roughness of the interface decreases the mobility of carriers in the inversion layer. As the vertical field increases, the inversion carriers are pulled closer to the interface and surface roughness limited mobility becomes increasingly smaller. Therefore, the mobility limited by surface roughness has an inverse relationship with effective field with a power factor,γ, close to 2 [75, 76].

Multiple Scattering Mechanisms

When multiple scattering mechanisms are present, the overall mobility is generally described by Matthiessen's rule:

$$\frac{1}{\mu} = \sum_i \frac{1}{\mu_i} = \frac{1}{\mu_{ph}} + \frac{1}{\mu_{coul}} + \frac{1}{\mu_{sr}} + \cdots$$

Since there is a reciprocal relationship, the effective mobility is dominated by the process with the lowest mobility. Coulomb scattering limited mobility is dominant at low fields since there is no screening of the inversion layer and surface roughness scattering is dominant at higher fields since the carriers are pulled closely to the imperfect interface.

The above mechanisms are very well established for thick EOT SiO_2 devices with poly-silicon electrodes. However, these mechanisms may not be sufficient to explain the poor mobility shown by the devices with high-k dielectrics and metal gate electrodes [77].

When the gate SiO_2 thickness, T_{ox}, is over 5 nm, carrier mobility is insensitive to T_{ox} [78]. Once it is below 2–3 nm, however, mobility reduces [79-84] and the reported reduction near threshold voltage varies substantially, from a factor over 2 [79] to an insignificant level [81]. As gate voltage increases from threshold, the relative mobility reduction can either decrease [79,80] or increase [81,82]. In order to explain this phenomenon, several scattering sources have been proposed. Agreement has not yet been reached on the origin for such phenomenon and the proposed mechanisms include remote phonon scattering [85], remote surface roughness [81,82] , long-range Coulomb interaction

between carriers in the gate and in the inversion layer [84] and remote charge scattering from impurities in the depleted poly-Si gate [79,80,83]. The explanation is controversial and Remote Charge Scattering from impurities in depleted poly-Si gate has attracted much attention. A brief introduction of these remote scattering mechanisms is given below.

Remote phonon scattering

Remote phonon scattering was proposed by Fischetti et al [85]. He predicted that the remote phonon scattering originating from the gate dielectric cannot be neglected in MOSFET with high-k materials. Since the bonds in most high-k dielectrics are the metal-oxygen type, they are highly polarisable and result in a large static permittivity, desirable for advanced CMOS scaling. However, inherent in these soft bonds are low energy lattice oscillations which are optical in nature and derive from their ionic characteristics. These low energy phonons trigger frequent emission and absorption processes by thermal electrons in the inversion layer, resulting in a remote phonon scattering mechanism which, due to the dipole field of the insulator, decays away from the bulk of the insulator into the inversion layer. Conversely, SiO_2 yields hard Si-O bonds and a reduced ionic polarization with high energy optical phonons. It is difficult for thermal electrons to emit excitations at these energies especially at room temperature where the number of absorbed thermally excited phonons is low.

Remote surface roughness scattering

Remote surface roughness scattering was first proposed by Li and Ma in 1987

[86]. They introduced this new scattering mechanism to explain the experimental observation that the mobility of carriers in the MOSFET channel degrades as the gate oxide gets thinner. When the oxide is thick, the channel charges are far from the metal/oxide interface and the scattering of carriers by that interface can be neglected. As the oxide becomes thinner, the imperfection at the metal/oxide interface (not perfectly smooth) can give rise to a potential variation seen by the carriers inside the channel which serves to degrade the mobility.

Long-range Coulomb interaction

In this theory, mobility degradation is explained by long-range coulomb interaction between electrons in the inverted channel of the substrate and electrons in the gate. The idea of coulomb interaction originates from the work in investigating the mutual drag between a two-dimensional electron gas (2DEG) and a 3DEG in compound-semiconductor system [87-90]. When the oxide dielectric thickness is thinner, the electrons in the channel of the Si substrate can interact strongly with the electrons in the gate. Since the electrons in the gate are almost at rest, while the channel electrons drift from the source to the drain under the action of the applied bias, a 'gate drag' can be expected and therefore electrons in the channel will lose momentum to those in the gate, leading to the mobility degradation.

Remote charge scattering from impurities in depleted poly-Si gate

The remote charge scattering (RCS) from impurities was extensively studied in

III-V hetero-structure semiconductors [91-94]. In 1999, it was introduced to metal-oxide-semiconductor devices on Si by Krishnan et al [95]. They developed a model for RCS by computing the impurity potential which is obtained by solving the Poisson equation, taking into account the size quantization effect and 2D screening of the carriers. In their calculation, they used semi-analytical expressions for the momentum relaxation time (MRT) that Stern and Howard originally derived from coulomb centres in the substrate silicon with very thick oxide dielectric [91]. However, free carriers can also produce screening effect and if this is taken into account, the impact from RCS can be negligible [83].

1.5 Organization of the thesis

This thesis is organized as follows:

Chapter 2 comprises a review of the characterization techniques used in investigating Negative Bias Temperature Instability (NBTI), including conventional and fast techniques. The DC transfer characteristic, capacitance (C-V), On-The-Fly (OTF) and ultra fast pulse I_d-V_g measurement (UFP) are reviewed. The focus is on the implementation and validation of the ultra fast pulse I_d-V_g technique. This technique is used later in the thesis.

Chapter 3 contains a comprehensive study on the characterization method of NBTI. The impact of various factors and uncertainties on ΔV_t evaluation by the OTF-V_{gst} technique is analyzed first, including degradation during the measurement that affects the reference I_d and initial transconductance. The

popular assumption that $|\Delta V_t|$ is independent of sensing V_g is not used, rather a new evaluation method is proposed and justified. Based on the new evaluation method, the difference in ΔV_t evaluated by different techniques is explained. Chapter 4 employs threshold voltage shift at device operating level to develop a model for NBTI kinetics which includes contributions from both as-grown and generated defects. The inapplicability of existing lifetime prediction techniques will be demonstrated. Based on the new kinetics, a single-test novel lifetime prediction method is proposed, and the safety margin is estimated.

Chapter 5 comprise a discussion of the impact of RCS on mobility. The impact is studied in the same device by varying charge density either by annealing or trapping. Therefore, the uncertainty from varying distance into the oxide by using different devices is removed. The work shows that the contribution of RCS to mobility degradation is insignificant.

Chapter 6 summarizes the work presented in this thesis. Finally, the direction for future work is suggested.

References

1. R. R. Schaller, "Moore's law: past, present and future," *IEEE Spectrum,* vol. 34, pp. 52-56, 1997.
2. G. E. Moore, "Microprocessors and Integrated Electronic Technology," *Proceedings of the IEEE*, vol. 64, pp. 837-841, 1976.
3. G. E. Moore, "The role of Fairchild in silicon technology in the early days of silicon valley," *Proceedings of the IEEE*, vol. 86, pp. 53-62, 1998.
4. G. E. Moore, "Cramming more components onto the integrated circuits" *Proceedings of the IEEE,* vol. 86, pp. 82-85, 1998
5. R. Goodall, D. Fendel, A. Allan, P. Landler, and H. R. Huff, "Long-term productivity mechanisms of the Semiconductor Industry," in *Electrochemical Society Conference (ECS)*, pp. 203-210, 2002.
6. R. H. Dennard, "U.S. Patent No. 3,387,286." USA, June 4, 1968.
7. R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, "Design of ion-implanted MOSFET's with very small physical dimensions, " *Proceedings of the IEEE*, vol. 87, pp. 668-678, 1999.
8. R. H. Dennard, "Evolution of the MOSFET Dynamic Ram - a Personal View," *IEEE Electron Dev. Lett.,* vol. 31, pp. 1549-1555, 1984.
9. International Technology Roadmap for Semiconductors, http://public.itrs.net.
10. J. Lee, H. Zhong, Y. S. Suh, G. Heuss, J. Gurganus, B. Chen and V. Misra, "Tunable work function dual metal gate technology for bulk and non-bulk CMOS," in *International Electron Devices Meeting (IEDM) Digest*, pp. 359-362, 2002.
11. V. Misra, H. Zhong, and H. Lazar, "Electrical properties of Ru-based alloy gate electrodes for dual metal gate Si-CMOS," *IEEE Electron Dev. Lett.,* vol. 23, pp. 354-356, 2002.
12. Y. S. Suh, G. Heuss, H. Zhong, S. N. Hong, and V. Misra, "Electrical characteristics of TaSixNy gate electrodes for dual gate Si-CMOS devices," in *Symposia on VLSI Technology and Circuits*, pp. 47-48, 2001.
13. H. Zhong, "*Ph.D. dissertation - Ru-based Gate Electrodes for Advanced Dual-Metal Gate CMOS Devices*," North Carolina State University, Raleigh, 2001.
14. R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz, "High-k/Metal-gate stack and its MOSFET characteristics," *IEEE Electron Dev. Lett.,* vol. 25, pp. 408-410, 2004.
15. G. Groeseneken, L. Pantisano, L. A. Ragnarsson, R. Degraeve, M. Houssa, T. Kauerauf, P. Roussel, S. De Gendt, M. Heyns, "The achievement and challenges for the electrical performance of MOSFETs with high-k gate dielectrics," in *International Symposia on Physical and Failure Analysis of IC*, pp. 147-155, 2004.

16. 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.*, vol. 114, pp. 266-274, 1967.
17. H. R. Huff, A. Hou, C. Lim, Y. Kim, J. Barnett, G. Bersuker, G. A. Brown, C. D. Young, P. M. Zeitzoff, J. Gutt, P. Lysaght, M. I. Gardner, and R. W. Murto, "High-k gate stacks for planar, scaled CMOS integrated circuits," *Microelectronic Engineering*, vol. 69, pp. 152-167, 2003.
18. J. Robertson, "Band offsets of high dielectric constant gate oxides on silicon," *Journal of Non-Crystalline Solids*, vol. 303, pp. 94-100, 2002.
19. G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High-k gate dielectrics: Current status and materials properties considerations," *J. Appl. Phys.*, vol. 89, pp. 5243-5275, 2001.
20. T. H. Hou, J. Gutt, C. Lim, S. Marcus, C. Pomarede, E. Shero, H. D. Waard, C. Werkhoven, M. Gardner, R. W. Murto, and H. R. Huff, "Improved Scalability of High-k Gate Dielectrics by Using Hf-Aluminates," in *1st International Symposium on High Dielectric Constant Materials*, pp. 133-137, 2002.
21. J. H. Lee, Y. S. Kim, H. S. Jung, J. H. Lee, N. I. Lee, H. K. Kang, J. H. Ku, H. S. Kang, Y. K. Kim, K. H. Cho, and K. P. Suh, "Poly-Si gate CMOSFETs with HfO2-Al2O3 laminate gate dielectric for low power applications," in *Symposia on VLSI Technology and Circuits*, pp. 84-85, 2002.
22. A. Chin, Y. H. Wu, S. B. Chen, C. C. Liao, and W. J. Chen, "High Quality La2O3 and Al2O3 Gate Dielectrics with Equivalent Oxide Thickness 5-10 Å," in *Symposia on VLSI Technology and Circuits*, pp. 16-17, 2000.
23. J. Barnett, D. Riley, T. C. Messina, P. Lysaght, and R. Carpio, "Wet etch enhancement of HfO2 films by implant processing," *Diffusion and Defect Data Part B Solid State Phenomena*, vol. 92, pp. 11-14, 2003.
24. H. Dedong, K. Jinfeng, L. Changhai, and H. Ruqi, "Reliability characteristics of high-K gate dielectrics HfO2 in metal-oxide semiconductor capacitors," *Microelectronic Engineering*, vol. 66, pp. 643-647, 2003.
25. X. Garros, C. Leroux, D. Blin, J. F. Damlencourt, A. M. Papon, and G. Reimbold, "Investigation of HfO2 dielectric stacks deposited by ALD with a mercury probe, in *European Solid State Device Research Conference*, pp. 411-414, 2002.
26. M. Gutowski, J. E. Jaffe, L. Chun Li, M. Stoker, R. I. Hegde, R. S. Rai, and P. J. Tobin, "Thermodynamic stability of high-k dielectric metal oxides ZrO2 and HfO2 in contact with Si and SiO2," in *Symposia on VLSI Technology and Circuits*, Vol. 716, pp. B3.2.1-B3.2.5, 2002.
27. D. D. Han, J. F. Kang, C. H. Lin, and R. Q. Han, "Annealing characteristics of ultra-thin high-K HfO2 gate dielectrics," *Chinese Physics*, vol. 12, pp. 325-327, 2003.
28. H. Hang, Z. Chunxiang, Y. F. Lu, Y. H. Wu, T. Liew, M. F. Li, B. J. Cho, W. K. Choi, and N. Yakovlev, "Physical and electrical characterization of HfO2 metal insulatormetal capacitors for Si analog circuit applications," *J. Appl. Phys.*, vol. 94, pp. 551-557, 2003.

29. Y. Kim, C. Lim, C. D. Young, K. Matthews, J. Barnett, B. Foran, A. Agarwal, G. A. Brown, G. Bersuker, P. Zeitzoff, M. Gardner, R. W. Murto, L. Larson, C. Metzner, S. Kher, and H. R. Huff, "Conventional Poly-Si Gate MOS-transistors With a Novel, Ultra-Thin Hf-oxide Layer," in *Symposia on VLSI Technology and Circuits*, pp. 167-168, 2003.
30. S. J. Lee, C. H. Choi, A. Kamath, R. Clark, and D. L. Kwong, "Characterization and reliability of dual high-k gate dielectric stack (poly-Si-HfO2-SiO2) prepared by in situ RTCVD process for system-on-chip applications," *IEEE Electron Dev. Lett.*, vol. 24, pp. 105-7, 2003.
31. P. F. Lee, J. Y. Dai, K. H. Wong, H. L. W. Chan, and C. L. Choy, "Growth and characterization of Hf-aluminate high-k gate dielectric ultrathin films with equivalent oxide thickness less than 10 Å," *J. Appl. Phys.*, vol. 93, pp. 3665-3667, 2003.
32. C. Lim, Y. Kim, A. Hou, J. Gutt, S. Marcus, C. Pomarede, E. Shero, H. Waard, C. Werkhoven, L. Chen, J. Tamim, N. Chaudhary, G. Bersuker, J. Barnett, C. D. Young, P. Zeitzoff, G. A. Brown, M. Gardner, R. W. Murto, and H. R. Huff, "Effect of Deposition Sequence and Plasma Treatment on ALCVD(TM) HfO2 n- MOSFET Properties," in *International Symposium on High Dielectric Constant Materials*, 2002.
33. B. E. Deal, M. Sklar, A. S. Grove and E. H. Snow, "Characteristics of the surface-state charge (Qss) of thermally oxidized silicon," *J. Electrochem. Soc.*, vol. 114, pp. 266-274, 1967.
34. D. F. Bentchkowsky, "A fully decoded 2048-bit electrically programmable FAMOS read-only memory," *IEEE Solid-State Circuit,* vol. 6 pp. 301-306, 1971.
35. B. E. Deal, "The current understanding of charges in the thermally oxidized silicon structure," *J. Electrochem. Soc.*, vol. 121, pp. 198-205, 1974.
36. R. J. Strain, A. Goetzberger and A. D. Lopez, "On the formation of surface states during stress aging of thermal Si–SiO2 interfaces," *J. Electrochem. Soc.*, vol. 120, pp. 90-96, 1973.
37. K. O. Jeppson, and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices," *J. Appl. Phys.*, vol. 48, pp. 2004-2014, 1977.
38. W. Schemmert, L. Gabler and B. hoeffinger, "Conductance of ion-implanted buried-channel MOS transistors," *IEEE Trans. Electron Dev.*, vol. 23 pp. 1313-1319, 1976.
39. Y. Nishida, H. Sayama, K. Oda, M. Katayama, Y. Inoue, H. Morimoto, and M. Inuishi, "SoC CMOS technology for NBTI/HCI immune I/O and analog circuits implementing surface and buried channel structures," in *International Electron Devices Meeting (IEDM) Digest*, pp. 869-873 , 2001.
40. 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. 1881-1884, 2003.

41. 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 *International Electron Devices Meeting (IEDM) Digest*, pp. 73-74, 1999.
42. 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 μm gate CMOS generation," in *Symposia on VLSI Technology and Circuits*, pp. 92-93, 2000.
43. D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, pp. 1-18, 2003.
44. S. Ogawa and N. Shiono, "Generalized diffusion-reaction model for the low-field charge-buildup instability at the Si-SiO2 interface," *Phys. Rev. B*, vol. 51, pp. 4218 , 1995.
45. M. A. Lam , B. Weir, P. Silverman, "The prospect of using thin oxides for silicon nanotransistors," in *International Electron Devices Meeting (IEDM) Digest*, pp. 30-34, 2001.
46. M. A. Alam, "A critical examination of the mechanism of dynamic NBTI for PMOSFETs," in *International Electron Devices Meeting (IEDM) Digest*, pp. 346-350, 2003.
47. H. Kufluoglu and M. A. Alam, "A geometrical unification of the theories of NBTI and HCI time-exponents and its implications for ultra scaled planar and surround gate MOSFETs," in *International Electron Devices Meeting (IEDM) Digest*, pp. 113-116, 2004
48. A. T. Krishnan, V. Reddy, S. Chakravathi, J. Rodriquez, S. John and S. Krishnan, "NBTI impact on transistor and circuit: models, mechanisms and scaling effects [MOSFETs]," in *International Electron Devices Meeting (IEDM) Digest*, pp. 15.4.1-15.4.4, 2003
49. C. E. Blat, E. H. Nicollian, and E. H. Poindexter, "Mechanism of negative-bias-temperature instability," *J. Appl. Phys.*, vol. 69, pp. 1712-1720, 1991.
50. J. Kakalios, R. A. Street, and W. B. Jackson, "The stretched-exponential relaxation arising from dispersive diffusion of hydrogen in amorphous silicon," *Phys. Rev. Lett.* vol. 59, pp. 1037-1040, 1987.
51. S. Zafar, B. H. Lee, J. Stathis, A. Callegari, T. Ning, "A model for negative bias temperature instability (NBTI) in oxide and high κ pFETs," in *Symposia on VLSI Technology and Circuits*, pp. 208-209, 2004.
52. K. Hess, A. Haggag, W. McMahon, B. Fischer, K. Cheng, J. Lee, and J. Lyding, "Simulation of Si-SiO2 defect generation in CMOS chips: from atomistic structure to chip failure rates," in *International Electron Devices Meeting (IEDM) Digest*, pp. 93-96, 2000.
53. D. L. Griscom, "Diffusion of radiolytic molecular-hydrogen as a mechanism for the post-irradiation buildup of interface states in SiO2-on-Si structures," *J. Appl. Phys.*, vol. 58, pp. 2524-2533, 1985.

54. S. N. Rashkeev, D. M. Fleetwood, R. D. Schrimpf and S. T. Pantelides, "Proton-induced defect generation at the Si-SiO2 interface," *IEEE Trans. Nucl. Sci.*, vol. 48, pp. 2086-2092, 2001.
55. R. B. Flair and R. C. Sun, "Threshold-voltage instability in MOSFETs due to channel hot-hole emission," *IEEE Trans. Electron Dev.*, vol. 28, pp. 83-93 ,1981.
56. F. B. McLean, "A framework for understanding radiation-induced interface states in SiO MOS structures," *IEEE Trans. Nucl. Sci.*, vol. 27, pp. 1651-1657,1980.
57. G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng and D. L. Kwong, "Dynamic NBTI of p MOS transistors and its impact on MOSFET scaling," *IEEE Electron Dev. Lett.*, vol. 23, pp. 734–736, 2002.
58. G. Chen, K. Y. Chuah, M. F. Li, D. S. H. Chan, C. H. Ang, J. Z. Zheng, Y. Jin and D. L. Kwong, "Dynamic nbti of pmos transistors and its impact on device lifetime," in *international reliability physics symposium (IRPS)*, pp. 196–202, 2003.
59. B. Zhu, J. S. Suehle, Y. Chen and J. B. Bernstein, "The negative bias temperature instability of deep sub-micron p-MOSFETs under pulsed bias stress," in *Integrated Reliability Workshop Final Report (IIRW)*, pp. 125–129, 2002.
60. M. Ershov, R. Lindley, S. Saxena, A. Shibkov, S. Minehane, J. Babcock, S. Winters, H. Karbasi, T. Yamashita, P. Clifton and M. Redford, "Transient effects and characterization methodology of negative bias temperature instability in Pmos transistors," in *international reliability physics symposium (IRPS)*, pp. 606–607, 2003.
61. 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.
62. H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin and C.Schlunder, "Analysis of NBTI degradation and recovery behaviour based on ultra fast Vt measurements," in *international reliability physics symposium (IRPS)*, pp. 448–453, 2006.
63. S. Rangan, N. Mielke and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability PMOSFETs," in *International Electron Devices Meeting (IEDM) Digest*, pp. 14.3.1–14.3.4, 2003.
64. M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. R. Tauriac and N. Revil, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's," in *International Electron Devices Meeting (IEDM) Digest*, pp. 109-112, 2004.
65. A. Kerber, E. Cartier, L. A. Ragnarsson, M. Rosmeulen, L. Pantisano, R. Degraeve, T. Kauerauf, G. Groeseneken, H. E. Maes and U. Schwalke. "Characterization of the VT-instability in SiO2/HfO2 gate dielectrics," in *International Electron Devices Meeting (IEDM) Digest*, pp. 41-45, 2003.

66. C. Shen, M. F. Li, X. P. Wang, Y. C. Yeo, D. L. Kwong, "A fast measurement technique of MOSFET Id-Vg characteristics," *IEEE Electron Dev. Lett.*, vol. 27, pp. 55-57, 2006.
67. C. Shen, M. F. Li, C. E. Foo, T. Yang, D. M. Huang, T. C. Chua, F. Nouri, G. S. Samudra and Y. C. Yeo. "Characterization and physical origin of fast *Vth* transient in BTI of pMOSFETs with SiON dielectric," in *International Electron Devices Meeting (IEDM) Digest*, pp. 333-336, 2006.
68. H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin and C. Schlunder, "Analysis of NBTI degradation and recovery behaviour based on ultra fast Vt measurements," in *international reliability physics symposium (IRPS)*, pp. 448–453, 2006.
69. D. V. Singh, P. Solomon, E. P. Gusev, G. Singeo and Z. Ren, "Ultra-fast measurement of the inversion charge in MOSFETs and impact on measured mobility in high-k MOSFETs," in *International Electron Devices Meeting (IEDM) Digest*, pp. 863-866,2003.
70. E. N. Kumar, V. D. Maheta, S. Purawat, A. E. Islam, 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) Idlin Technique," in *International Electron Devices Meeting (IEDM) Digest*, pp. 809-812, 2007.
71. K. K. Abdelghafar, K. Watanabe, J. Ushio and E. Murakami, "Effect of nitrogen at SiO2/Si interface on reliability issues —negative-bias temperature instability and Fowler-Nordheim-stress degradation," *Appl. Phys. Lett.*, vol. 81, pp. 4362-4364, 2002.
72. J. Ushio, T. Maruizumi and K. K. Abdelghafar," Interface structures generated by negative-bias temperature instability in Si/SiO2 and Si/SiOxNy interfaces," *Appl. Phys. Lett.*, vol. 81, pp. 1818-1820, 2002.
73. M. H. Chang and J. F. Zhang, "On positive charges formed under negative bias temperature stresses," *J. Appl. Phys.*, vol. 101, pp. 1312-1316, 2007.
74. S. S. Tan, T. P. Chen, C. H. Ang and L. Chan, "Relationship between interfacial nitrogen concentration and activation energies of fixed-charge trapping and interface state generation under bias-temperature stress condition," *Appl. Phys. Lett.*, vol. 82, pp. 269-271, 2003.
75. S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration," *IEEE Trans. Electron Dev.*, vol. 41, pp. 2357-2362, 1994.
76. Y. C. Cheng and E. A. Sullivan, "On the role of scattering by surface roughness in silicon inversion layers," *Surface Science*, vol. 34, pp. 717-731, 1973.
77. E. P. Gusev, "Ultrathin high-K gate stacks for advanced CMOS devices," in *International Electron Devices Meeting (IEDM) Digest,* pp. 20.1.1 - 20.1.4, 2001.
78. M. S. Liang, J. Y. Choi, P. K. Ko, and C. Hu, "Inversion-layer capacitance and mobility of very thin gate-oxide MOSFET's," *IEEE Trans. Electron Dev.*, vol. 33, pp. 409-413, 1986.

79. M. S. Krishnan, Y. C. Yeo, Q. Lu, T. J. King, J. Bokor, and C. Hu, "Remote charge scattering in MOSFETs with ultra-thin gate dielectrics," in *International Electron Devices Meeting (IEDM) Digest,* pp. 571-574, 1998.
80. M. Casse, L. Thevenod, B. Guillaumot, L. Tosti, F. Martin, J. Mitard, O. Weber, F. Andrieu, T. Ernst, G. Reimbold, T. Billon, M. Mouis, and F. Boulanger, "Carrier transport in HfO2/metal gate MOSFETs: physical insight into critical parameters" *IEEE Trans. Electron Dev.,* vol. 53, pp. 759-768, 2006.
81. N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman, "Estimation of the effects of remote charge scattering on electron mobility of n-MOSFET's with ultrathin gate oxides," *IEEE Trans. Electron Dev.,* vol. 47, pp. 440-447, 2000.
82. Q. Xiang, G. Yeap, D. Bang, M. Song, K. Ahmed, E. Ibok, and M. R. Lin, "Performance and reliability of sub-100nm MOSFETs with ultra thin direct tunneling gate oxides," in *Symposia on VLSI Technology and Circuits*, pp. 160-161, 1998.
83. D. Esseni and A. Abramo, "Modeling of electron mobility degradation by remote Coulomb scattering in ultrathin oxide MOSFETs," *IEEE Trans. Electron Dev.,* vol. 50, pp. 1665-1674, 2003.
84. M. V. Fischetti, "Long-range Coulomb interactions in small Si devices. Part II. Effective electron mobility in thin-oxide structures," *J. Appl. Phys.,* vol. 89, pp. 1232-1250, 2001.
85. M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering," *J. Appl. Phys.,* vol. 90, pp. 4587-4608, 2001.
86. J. Li and T. P. Ma, "Scattering of silicon inversion layer electrons by metal/oxide interface roughness," *J. Appl. Phys.,* vol. 62, pp. 4212-4215, 1987.
87. M. B. Pogrebinskii, F. T. Poluprovodn, "Mutual drag of carriers in a semiconductor-insulator- semiconductor system," *J. Appl. Phys.,* vol. 11, pp. 327, 1977.
88. P. J. Price, "Hot electron effects in heterolayers," *Physica B & C,* vol. 117, pp. 750-752, 1983.
89. I. I. Boiko, N. Yu, Z. Sirenko and T. Fiz, "Scattering of Two-Dimensional Electron Gas on the Semibounded Three-Dimensional Electron Gas," in *J. Sov. Phys. Tech.,* vol. 58, pp. 967, 1988.
90. B. Leikhtman and P. M. Solomon, "Mutual drag of two- and three-dimensional electron gases in heterostuctures," *Phys. Rev. B*, vol. 41, pp. 9921, 1990.
91. F. Stern and W. E. Howard, "Properties of semiconductor surface inversion layers in the electric quantum limit," *Phys. Rev. B*, vol. 163, pp. 816, 1967.
92. K. Hess, "Impurity and phonon scattering in layered structures," *Appl.Phys. Lett.*, vol. 35, pp. 484, 1979.

93. Y. Takeda, H. Kamei, and A. Sasaki, "Mobility calculation of two-dimensional electron gas in GaAs/AlGaAs heterostructure at 4.2 K," *Electron. Lett.*, vol. 18, pp. 309, 1982.
94. A. A. Grinberg and M. S. Shur, "Effect of image charges on impurity scattering of two-dimensional electron gas in AlGaAs/GaAs," *J. Appl. Phys.*, vol. 58, p. 382, 1985.
95. M. S. Krishnan, Y. C. Yeo, Q. Lu, T. -J. King, J. Boker and C. Hu, "Remote charge scattering in MOSFETs with ultra-thin gate dielectrics," in *International Electron Devices Meeting (IEDM) Digest,* pp. 571-574, 1998.

2 Conventional characterization review and development of pulse I_d-V_g technique

2.1 Introduction

In the 1980s, highly integrated and automated electrical measurement instruments became commonly available, including Hewlett Packard 4140B pA meter/DC voltage source [1], followed by Hewlett Packard 4145 [2], 4155/4156 semiconductor parameter analyzers [3], Agilent B1500 semiconductor device analyzer [4], and similar products from other manufacturers. The semiconductor parameter analyzers are the de facto standard for measuring the I-V characteristics of transistors.

Normally the measurements performed by these analyzers are called quasi-static or DC measurements. DC measurements can accurately measure small currents down to 10^{-15} A which is critical for MOSFET characterization. In order to reach such high accuracy, normally the measurement is slow. In practice, to achieve sub-pico-ampere accuracy, the measurement time at each bias point (integration time) should be at least 20 ms, or one power-line-cycle, in order to minimize interference from the power supply. In addition, small bias steps in measurement and a delay time before measurement at each step are recommended, to ensure that all transients caused by the bias step relax and

the true steady-state characteristics are measured. In the field of reliability research, though this is widely accepted when investigating on thick oxide, it is no longer valid when the oxide thickness becomes very small.

The degradation of MOSFET characteristics under electrical stress can recover after the stress is removed. This recovery produces strong transients when measuring MOSFET device parameters, notably in the threshold voltage V_t, in the time scale of micro-seconds to tens of seconds. The typical measurement time with semiconductor parameter analyzers ranges from milli to tens of seconds, which overlaps with that of the transient in device parameters. During this overlap time, transistor parameters (e.g. threshold voltage) have already changed. It is therefore necessary to remove this overlap by applying fast measurements for transistor characteristics.

In this chapter, the conventional techniques including measurement of transfer characteristic (I-V), capacitance (C-V) and the fast techniques including On-The-Fly (OTF) and ultra fast pulse I_d-V_g (UFP) are reviewed first. The focus is on the implementation and justification of the ultra fast pulse I_d-V_g technique and this technique will be used later in the thesis.

2.2 Review of conventional characterization techniques

2.2.1 V_t by conventional I_d-V_g technique

Conventionally, V_t degradation is evaluated using the measure-stress-measure (MSM) method with a DC parametric analyzer. This is also called the "slow"

technique or DC technique since the total time for one I_d-V_g measurement is typically in the order of 1~10 seconds, and the threshold voltage is normally extracted by using extrapolation [5] or the constant current method [6]. Figure 2.1 (a) shows how the threshold voltage V_t is extracted by using an extrapolation method, from the typical I_d-V_g curves measured before and after stress. The transconductance is firstly 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 transconductance. Figure 2.1 (b) shows a typical V_t shift with stress time under NBTI stress.

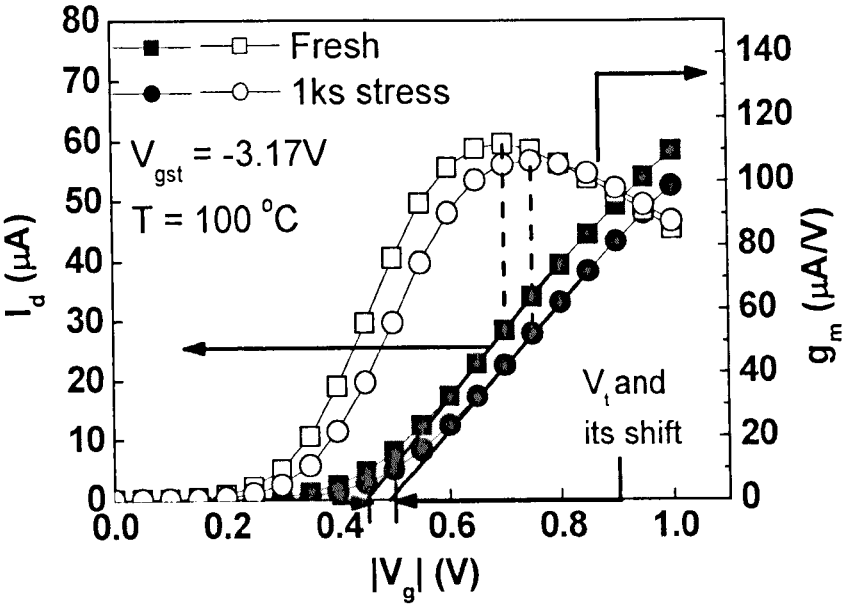
2.2.2 Conventional C-V technique

The capacitance-voltage behaviour of a MOS device can be described using the equivalent circuit shown in figure 2.2 [7] where C_{OX} is the oxide capacitance, C_S the substrate capacitance, C_{it} the interface state capacitance, R_S the series resistance and $1/R_P$ the parallel conductance.

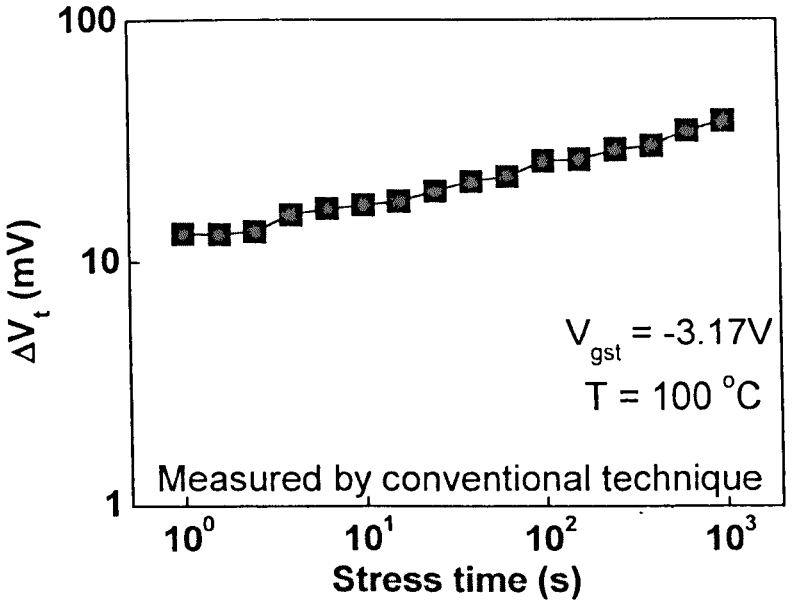
The capacitance of a MOS capacitor is defined as

$$C = \frac{dQ_g}{dV_g} \tag{2.1}$$

based on charge neutrality, $Q_g = -(Q_s + Q_{it})$ with Q_s the substrate charge and Q_{it} the trapped interface charge. This assumes no charge trapping in the dielectric. The gate voltage is partially dropped across the dielectric and the semiconductor substrate. This gives $V_g = V_{fb} + V_{ox} + \phi_s$, where V_{fb} is the flat-band voltage, V_{ox} the voltage drop across the oxide and ϕ_s the Si surface potential.



(a)



(b)

Fig 2.1 (a) Typical results obtained by using conventional technique. I_d-V_g curves were measured after stressing the device under NBTI for a period of time. After stress, I_d-V_g curve is shifted towards higher $|V_g|$. Threshold voltage is extracted by using the maximum g_m extrapolation method. The transconductance 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 from the maximum transconductance. (b) Typical NBTI degradation with stress time.

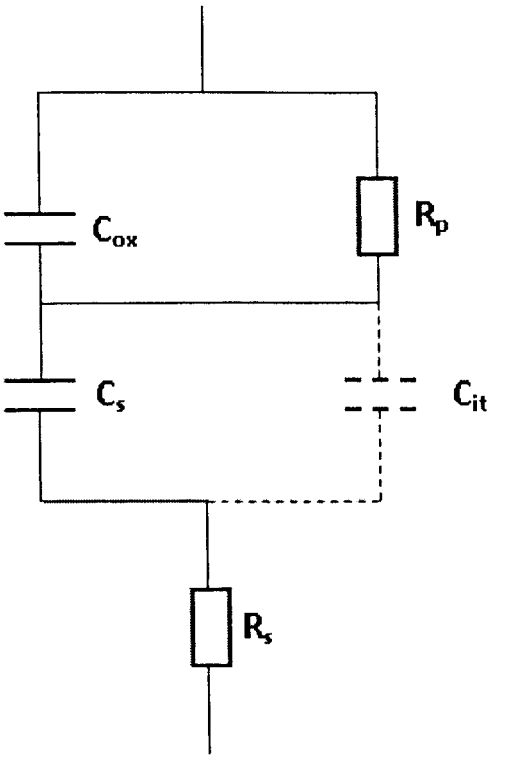


Fig 2.2 Equivalent circuit of an MOS structure.

Therefore equation (2.1) can be written as

$$C = -\frac{dQ_s + dQ_{it}}{dV_{ox} + d\phi_s} \tag{2.2}$$

Depending on the Si surface potential, the contribution of majority, minority or depletion charge associated with the substrate varies.

From the equivalent circuit, the total gate capacitance can also be written as

$$C = \frac{1}{1/C_{ox} + 1/(C_s + C_{it})} \tag{2.3}$$

The low-frequency substrate capacitance is given by [7]

$$C_{s,lf} = \hat{U}_s \frac{\varepsilon_{si}\varepsilon_0}{2L_D} \frac{e^{U_F}(1-e^{-U_S}) + e^{-U_F}(e^{U_S}-1)}{F(U_s,U_F)} \tag{2.4}$$

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

$$F(U_S,U_F) = \sqrt{e^{U_F}(e^{-U_S}+U_S-1) + e^{-U_F}(e^{U_S}-U_S-1)} \tag{2.5}$$

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

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

$$\hat{U}S = \frac{|U_S|}{U_S} \tag{2.6}$$

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

$$L_D = \sqrt{\frac{\varepsilon_{Si}\varepsilon_0 kT}{2q^2 N_A}} \tag{2.7}$$

High Frequency C-V measurement

In a High Frequency Capacitance-Voltage (HF C-V) measurement, a small AC signal superimposed on a DC bias is applied to the MOS device and the response is analyzed with respect to gain and phase. A schematic drawing of the HF C-V measurement setup is shown in figure 2.3 (a).

The gain and phase analysis is based on the equivalent circuits given in figure 2.3 (b) and figure 2.3 (c) where in addition to the measured capacitance, either a parallel conductance term or a series resistance is considered. The impedance Z_{serial} of the serial circuit is given by

$$Z_{serial} = R_{serial} + \frac{1}{j\varpi C_{serial}} \tag{2.8}$$

where R_{serial} is the series resistance, $\varpi = 2\pi f$ is the angular frequency and C_{serial} the serial capacitance. For the parallel circuit the admittance $Y_{parallel}$ can be written as

$$Y_{parallel} = G_{parallel} + j\varpi C_{parallel} \tag{2.9}$$

where $G_{parallel}$ is the parallel conductance and $C_{parallel}$ the parallel capacitance.

To discuss the limitations of the equivalent circuits the serial circuit is transformed into a parallel circuit and vice versa. The inverse of the serial impedance is given by

$$\begin{aligned} Y_{serial} &= \frac{1}{Z_{serial}} = \frac{1}{R_{serial} + 1/j\varpi C_{serial}} \\ &= \frac{1/R_{serial}}{1+Q_{serial}^2} + \frac{j\varpi C_{serial}}{1+1/Q_{serial}^2} \equiv Y_{parallel} \end{aligned} \tag{2.10}$$

with $Q_{serial} = 1/\varpi C_{serial} R_{serial}$ being the quality factor of the serial circuit. For the inverse of the parallel admittance, write

$$\begin{aligned} Z_{parallel} &= \frac{1}{Y_{parallel}} = \frac{1}{G_{parallel} + j\varpi C_{parallel}} \\ &= \frac{1/G_{parallel}}{1+Q_{parallel}^2} + \frac{1/j\varpi C_{parallel}}{1+1/Q_{parallel}^2} \equiv Z_{serial} \end{aligned} \tag{2.11}$$

with $Q_{parallel} = \varpi C_{parallel} / G_{parallel}$ being the quality factor of the parallel circuit. For the case where $G_{parallel} < \varpi C_{parallel}$ and $R_{serial} < 1/\varpi C_{serial}$, both

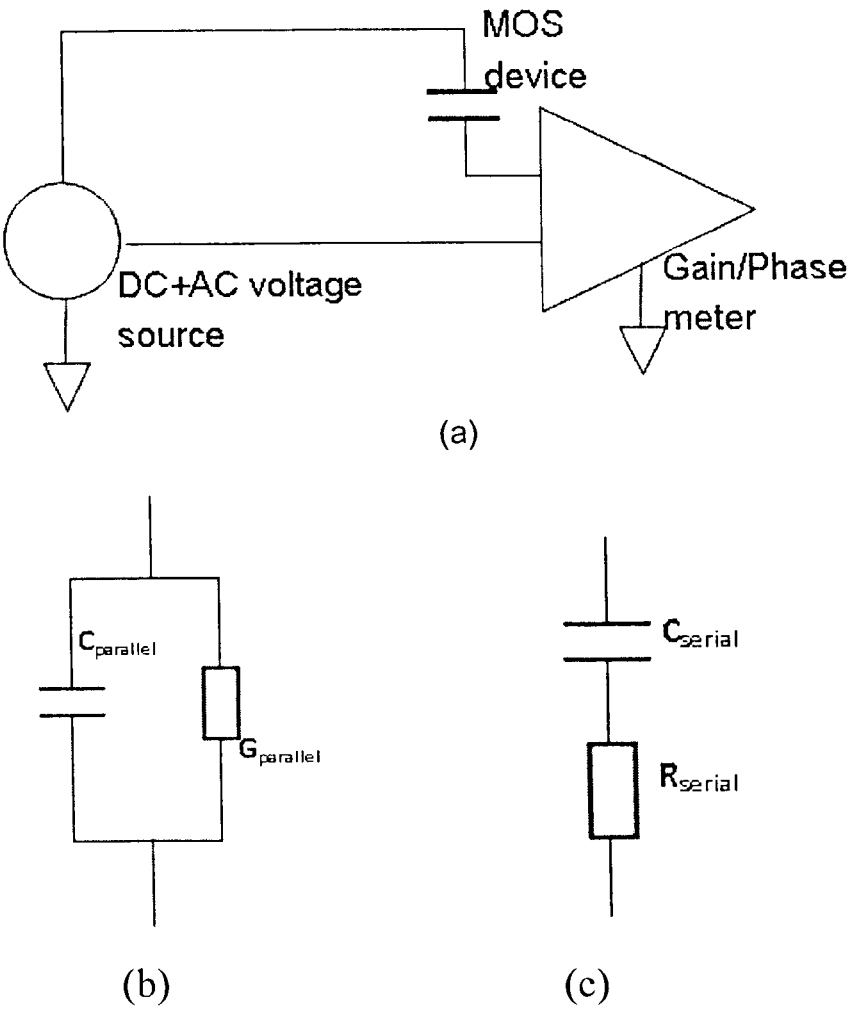


Fig 2.3 (a) Schematic configuration of high frequency C-V measurement. The DC and AC bias is applied to one terminal of the capacitor whereas the gain and phase of the AC signal is measured on the second terminal. (b) Equivalent circuit of a HF C-V taken in the parallel mode, (c) Equivalent circuit of a HF C-V taken in a serial mode

measurement modes yield equivalent results and the extracted capacitance becomes $C_{parallel} = C_{serial}$. Therefore, the measurement result is independent of the selected measurement mode. However, considering the simplified equivalent circuit of the MOS devices which normally contains the parallel conductance and the serial resistance, as shown in figure 2.4.

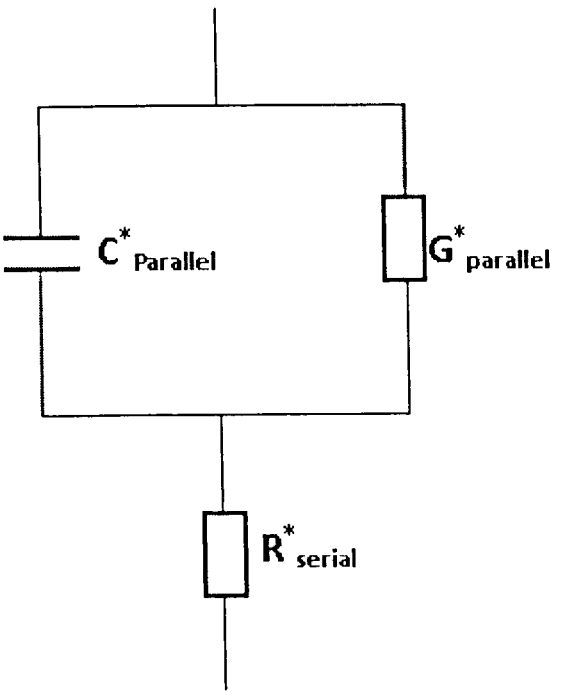


Fig 2.4 Simplified equivalent circuit of a MOS device with parallel conductance and serial resistance.

The impedance Z* and admittance Y* of the circuit are given by

$$\begin{aligned} Z^* &= R^*_{serial} + \frac{1}{G^*_{parallel} + j\varpi C^*_{parallel}} \\ &= R^*_{serial} + \frac{1/G^*_{parallel}}{1+Q^{*2}_{parallel}} + \frac{1/j\varpi C^*_{parallel}}{1+1/Q^{*2}_{parallel}} \end{aligned} \tag{2.12}$$

$$Y^* = \frac{1}{Z^*} = \frac{R^*_{serial} + \frac{1/G^*_{parallel}}{1+Q^{*2}_{parallel}} + \frac{j\varpi C^*_{parallel}}{\varpi^2 C^{*2}_{parallel}(1+1/Q^{*2}_{parallel})}}{(R^*_{serial} + \frac{1/G^*_{parallel}}{1+Q^*_{parallel}})^2 + (\frac{1/\varpi C^*_{parallel}}{1+1/Q^{*2}_{parallel}})^2}$$

where $C_{parallel}$ is the equivalent capacitance measured in the parallel mode [8].

The limiting factor at higher frequencies is the series resistance whereas at low frequencies the parallel conductance has to be taken into consideration.

Quasi static C-V measurement

An alternative technique to measure the capacitance-voltage characteristic of MOS devices is the quasi static C-V measurement. This technique requires measurement of either the displacement current or displacement charge, which directly reflect the capacitance of the test structure.

One of the commonly used quasi static C-V measurement techniques involves application of a linear voltage ramp to one terminal of the device and measurement of the current as a function of time at the second terminal, as shown in figure 2.5 (a). The capacitance is then given by the relation

$$C = \frac{J}{dV/dt} \tag{2.13}$$

The voltage ramp dV/dt can be varied over a range from ~10 V/s down to <0.01 V/s.

The practical limitation of the voltage ramp is given by the sensitivity of the ammeter and the leakage current of the measurement setup. The basic gate leakage requirements can be estimated as follows

$$J_g < C_{ox} \cdot dV / dt \tag{2.14}$$

For an EOT of ~ 2 nm (1.5 μF/cm^2) and a voltage ramp of 1 V/s a leakage current of $J_g < 1.5$ μA/cm^2 would be required.

Equilibrium controlled quasi static C-V

The equilibrium controlled quasi static C-V measurement [9] is an alternative technique. In contrast to the linear voltage ramp technique, discrete voltage steps are applied and the displacement charge is measured using the electrometer, as illustrated in figure 2.5 (b). The capacitance is calculated from the measured displacement charge and the voltage step following the relationship in equation (2.15).

$$C(V)=\frac{\Delta Q}{\Delta t}=\frac{Q(V+\frac{\Delta V}{2})-Q(V-\frac{\Delta V}{2})}{\Delta V} \qquad (2.15)$$

The strength of the quasi static C-V technique is the ability to measure interface states and it has been used extensively in the past decades. However, as the device scaling progresses, the gate leakage current has become so large that the quasi static C-V measurement cannot be used for a SiO_2 thinner than 3 nm approximately.

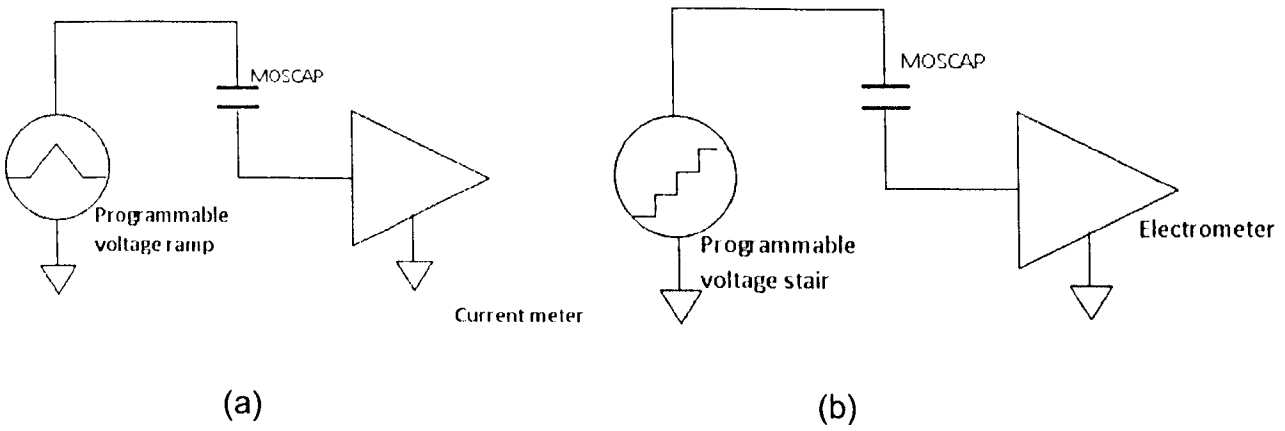


Fig 2.5 (a) Schematic configuration of quasi-static measurement setup using a linear voltage ramp; (b) schematic configuration of equilibrium controlled quasi static measurement setup where discrete voltage steps are applied and the displacement charge is measured using an electrometer.

2.3 On-The-Fly (OTF) technique

The OTF technique was motivated by the desire to measure the NBTI induced threshold voltage shift without recovery. The recovery during the conventional measurement has long been thought negligible until recently when it was found that most of the V_t degradation induced by NBTI recovers during the measurement when using conventional techniques [10-14]. In addition, the studies of charge trapping in high-k dielectrics in recent years have also shown significant recovery even within 1 ms delay [15–16]. Therefore, an accurate V_t measurement technique is required to capture all the degradation.

Some attempts on reducing the measurement time with the semiconductor parameter analyzers like the Agilent 41xx series were made initially. The minimum measurement time for a single bias point can be as fast as 80 μs as specified by the manufacturer [3]. This is much faster than the configuration in previous module. However, it was soon realized that significant overhead time must be added to the quoted minimum, as shown in figure 2.6 (a). The real measurement time can be much longer than the value preset in the instrument.

One of the earliest measurement solutions proposed to mitigate the recovery during measurement is the on-the-fly (OTF) technique, and a few variants have been proposed [17-21]. The common feature is that the stress voltage is always applied to the gate, and the degradation of drain current is measured at stress voltage. Since the stress voltage is not removed from the gate, the on-the-fly method is claimed to be free from the fast recovery of NBTI. It soon became popular as it can be performed on commercial semiconductor parameter

analyzers, without additional equipment requirement.

The first OTF technique (the 1st order OTF technique) was proposed by Rangan et al [17]. An initial I_d-V_g curve is measured with V_g swept to the stress voltage, and both the drain current I_{d0} at $V_g = V_{gst}$ and the threshold voltage V_{t0} are recorded. During the following electrical stress, the drain current is continuously sampled at $V_g = V_{gst}$ and the change of drain current $\Delta I_d = I_d - I_{d0}$ can be calculated. The threshold voltage shift is then calculated from

$$\Delta V_t = \frac{\Delta I_d}{I_{d0}} \cdot (V_g - V_{t0}) \tag{2.16}$$

Apparently, equation (2.16) ignored the mobility variation with V_g, and this assumption introduces a large uncertainty into the value of the extracted threshold voltage shift. Therefore this technique is not widely used.

In order to remove this uncertainty, an improved OTF technique (the 2nd order OTF technique) was proposed by Denais et al [18]. The mobility degradation is taken into consideration by evaluating the transconductance, g_m during each measurement. As shown in figure 2.6 (b), three points are measured at V_{gst} and V_{gst}±DV. Therefore, g_m can be obtained by using equation (2.17).

$$g_m = \frac{I_d(V_g + DV) - I_d(V_g - DV)}{2DV} \tag{2.17}$$

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

$$\Delta I_d(n) = I_d(n) - I_d(n-1) \tag{2.18}$$

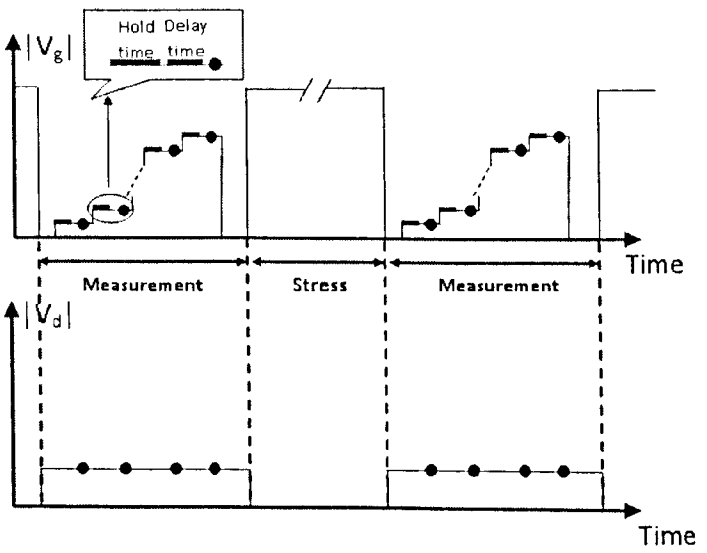
The shift of threshold voltage between these two points can be evaluated by

$$\Delta V_t(n) = -\frac{\Delta I_d}{g_m(n)} \tag{2.19}$$

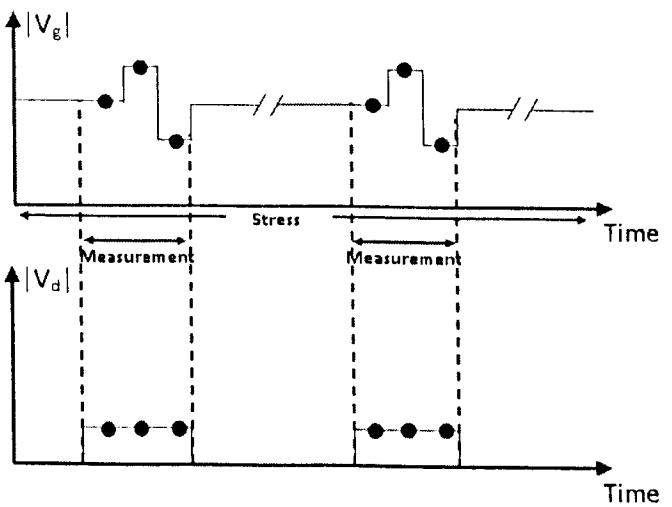
The accumulative shift of threshold voltage is

$$\Delta V_t = -\sum_{n=1}^{M} \frac{I_d(n) - I_d(n-1)}{g_m(n)} \tag{2.20}$$

Where M+1 is the number of I_d measurements and $g_m(n)$ is the mean value of the transconductance between the n^{th} and $(n-1)^{th}$ measurements, as shown in figure 2.7. Hence, periodical, three-point I_d measurements are enough to monitor ΔI_d, g_m, and ΔV_t during stress.



(a)



(b)

Fig 2.6 (a) Traditional NBTI test sequence, (b) The 2nd order On-The-Fly measurement sequence.

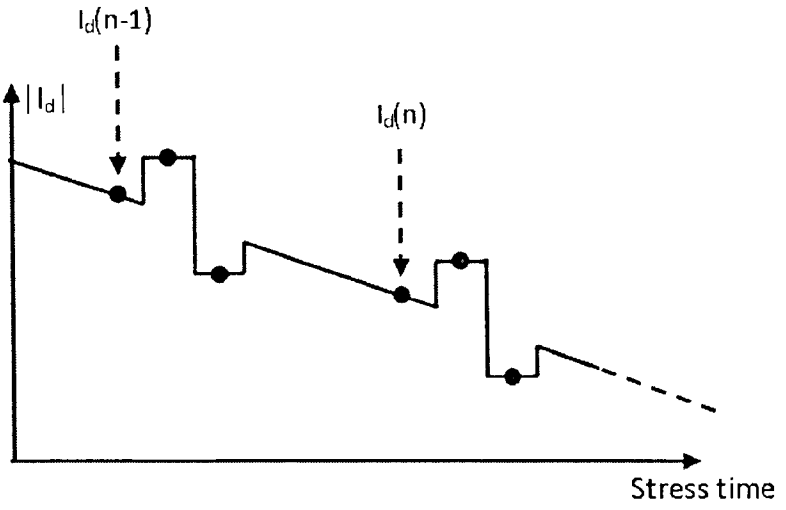


Fig 2.7 The n^{th} and $n\text{-}1^{th}$ Id measurements, together with the transconductance $g_m(n)$, can give the threshold voltage shift, ΔV_t, between n^{th} and $n\text{-}1^{th}$ measurement points.

In equation (2.20), g_m is assumed to be constant between the two measurement points and the average $\overline{g_m}$ is used, $\overline{g_m(n)} = (g_m(n) + g_m(n-1))/2$. To estimate the error caused by this assumption, figure 2.8 shows a comparison of the ΔV_t calculated by using $g_m(n)$, $g_m(n\text{-}1)$ and $\overline{g_m(n)}$. It is clear that the error is insignificant.

Since the 'On-The-Fly' measurement is based on I_d degradation at stress voltage, NBTI still occurs during the measurement. When I_d was measured for the first time, it typically took 0.15 s and some degradation occurs during this period. This degradation in the reference I_d leads to an underestimation of ΔV_t. It also affects the g_m evaluation. The error of the 2^{nd} order OTF technique will be discussed further in chapter 3.

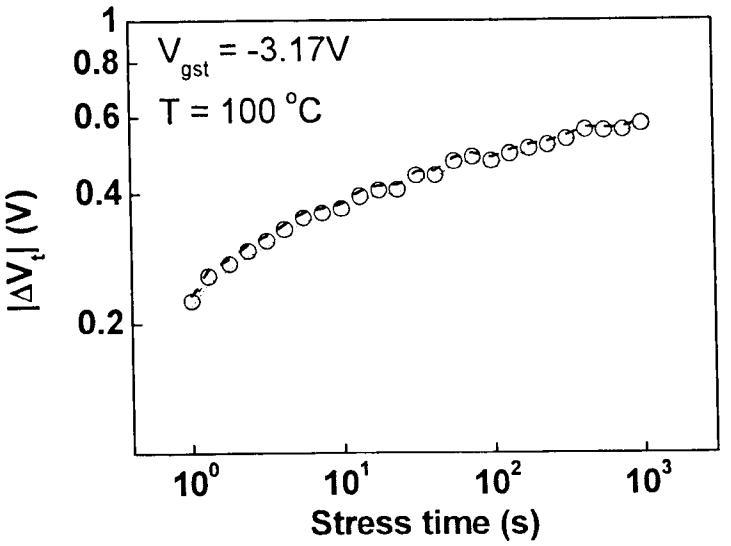


Fig 2.8 The effect of g_m on ΔV_t. Symbol 'O' is calculated using the average g_m value between two measurement points, while the dashed line used the previous g_m value, and the dot line uses the current gm value. It is clear that ΔV_t is weakly dependent on the g_m value used.

2.4 Pulse I_d-V_g Technique

2.4.1 Experimental setup

Another direction for suppressing recovery is to implement fast measurement by using specific circuits. Kerber et al [15] were the first to develop a pulse I_d–V_g technique to investigate the large charge trapping in high-k dielectric. A I_d–V_g curve can be measured within 10–100 μs using their technique. The schematic measurement setup for the pulsed I_d-V_g measurement is given in figure 2.9. In this setup, the MOSFET is used in an inverter circuit with a resistive load (R). A small DC bias is applied to the resistor which acts together with the channel resistor as a voltage divider. The I_d-V_g characteristic is obtained by applying a trapezoidal (triangular) pulse to the gate and recording the drain voltage using a

digital oscilloscope. From the measured voltage traces the I_d-V_g characteristic can be constructed using the following relationship [15]:

$$I_d = \frac{100mV}{V_d}\left(\frac{100mV - V_d}{R}\right) \tag{2.21}$$

where V_d is the measured drain voltage and R the resistive load of the inverter.

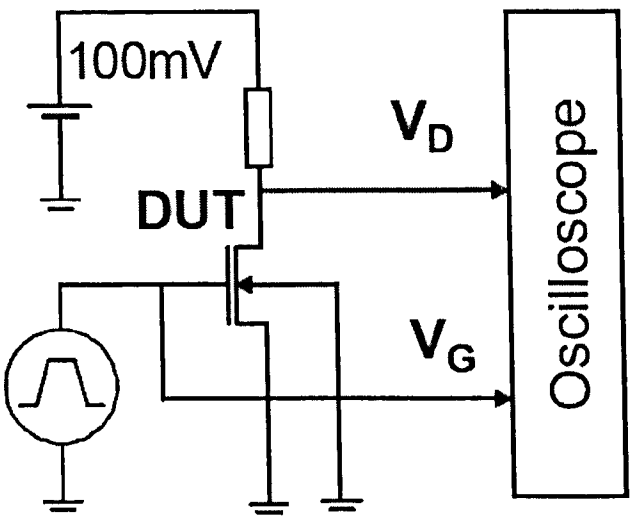
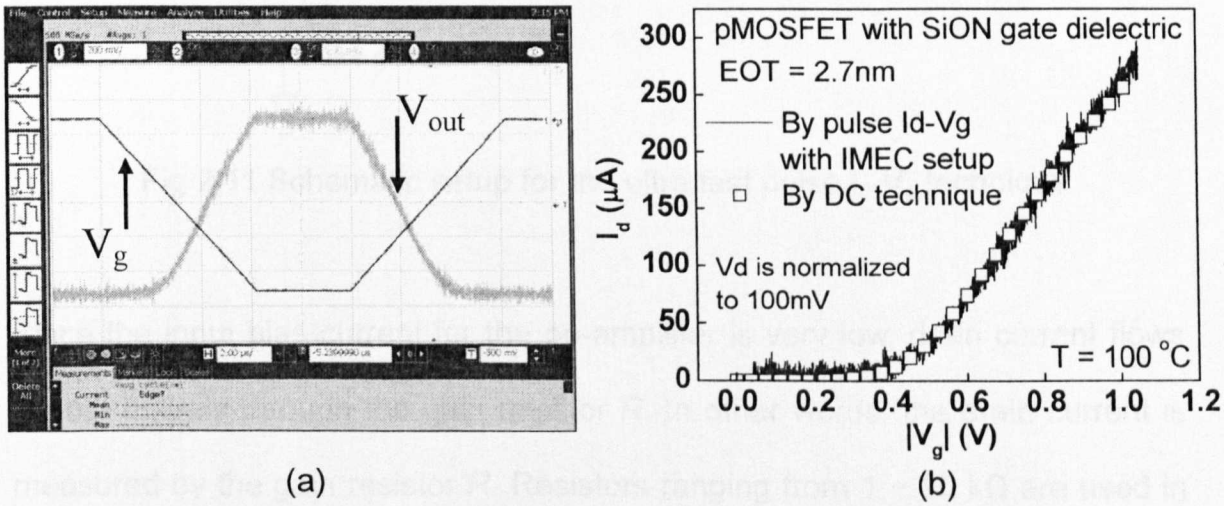


Fig 2.9 Schematic setup for the pulse I_d-V_g measurement proposed by A.Kerber et al. The MOSFET was used in an inverter circuit with a resistive load (R). From the voltage waveform ($V_g(t)$, $V_d(t)$), the I_d-V_g characteristics were extracted.

Due to the use of a voltage divider, the drain voltage changes during the measurement. In order to eliminate this effect, the extracted drain current is normalized to a constant drain voltage, which is given by the term $100mV/V_d$ in equation (2.21).

This method suffers from the following limitations. Firstly, in order to reduce noise and obtain accurate data from the oscilloscope, the impedance along the signal path should be matched and therefore the resistive load should be around 50Ω. This limits the gain of the circuit significantly. Figure 2.10 (a) shows a screen shot of one measurement using this technique. Figure 2.10 (b) shows

the extracted I_d-V_g curve. It can be seen, how considerable noise is observed, and this technique is affective with large drive currents. Moreover, due to the use of a voltage divider, the drain voltage changes during the measurement. The normalization used in the technique is only correct when the MOSFET is operated in the linear regime, and this gives uncertainties on the drain current measured at higher V_g voltage level where the device series resistance cannot be ignored.



(a) (b)

Fig 2.10 (a) A screenshot of the signals from the oscilloscope connected as shown in figure 2.9. (b) The extracted I_d-V_g curve by using equation (2.21), V_d is normalized to 100mV. The device used here is 2.7nm SiON with a channel length = 0.15 μm and channel width = 10 μm.

In order to increase the gain of the circuit while maintaining impedance matching along the signal path, an improved approach was proposed, where an op-amplifier is used [22]. The schematic setup is shown in figure 2.11. The drain of the MOSFET is connected to negative input of the op-amplifier (op-amp). Since the voltages at the two input terminals are approximately equal when negative feedback is present through *R*, the drain voltage of the MOSFET is fixed at V_d supplied by the voltage source.

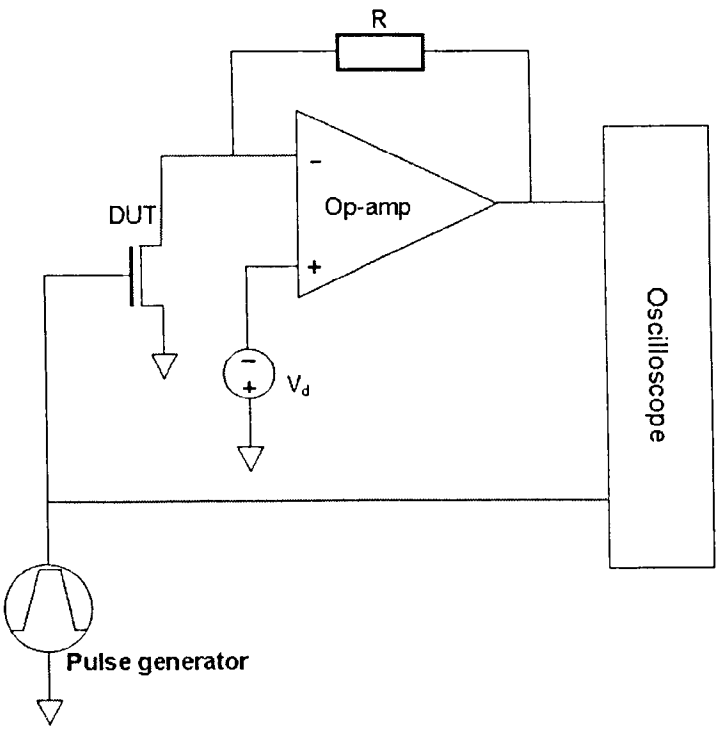
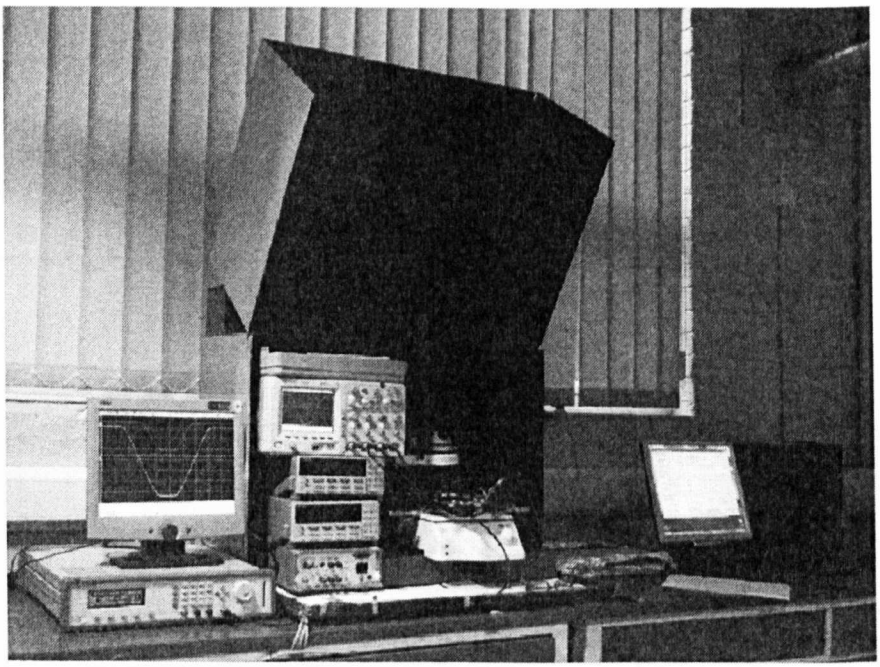


Fig 2.11 Schematic setup for the ultra fast pulse I_d-V_g technique

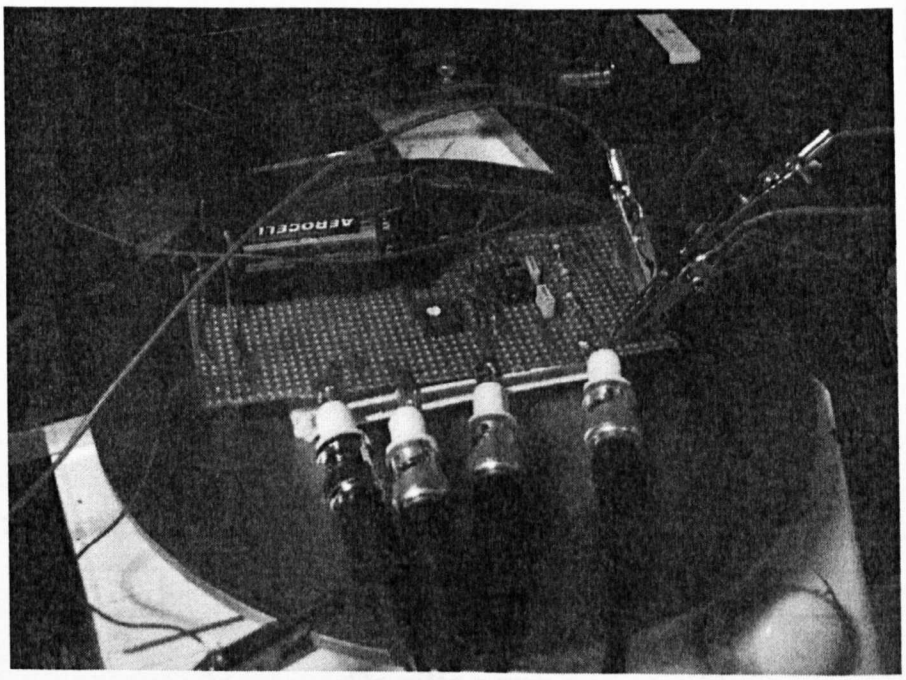
Since the input bias current for the op-amplifier is very low, drain current flows almost entirely through the gain resistor *R*. In other words, the drain current is measured by the gain resistor *R*. Resistors ranging from 1 – 10 kΩ are used in this study for different gain. The output voltage from the op-amplifier is related to the MOSFET drain current by,

$$V_{out} = I_d \cdot R + V_d \tag{2.22}$$

This technique was implemented, as shown in figure 2.12. A high-speed op-amplifier f with 60 MHz bandwidth (AD844) is used to achieve fast measurement [23]. As to be discussed in more detail later, accurate and fast measurement primarily relies on the minimization of the length of signal paths. The corresponding circuit schematic diagram is shown in figure 2.13 and the impedance controlled cables are labeled explicitly, while thin lines represent wires and probe tips. The components enclosed by the dashed box (except the DUT) is mounted inside the probe station, as shown in figure 2.12.



(a)



(b)

Fig 2.12 Photographs of the ultra fast pulse measurement setup

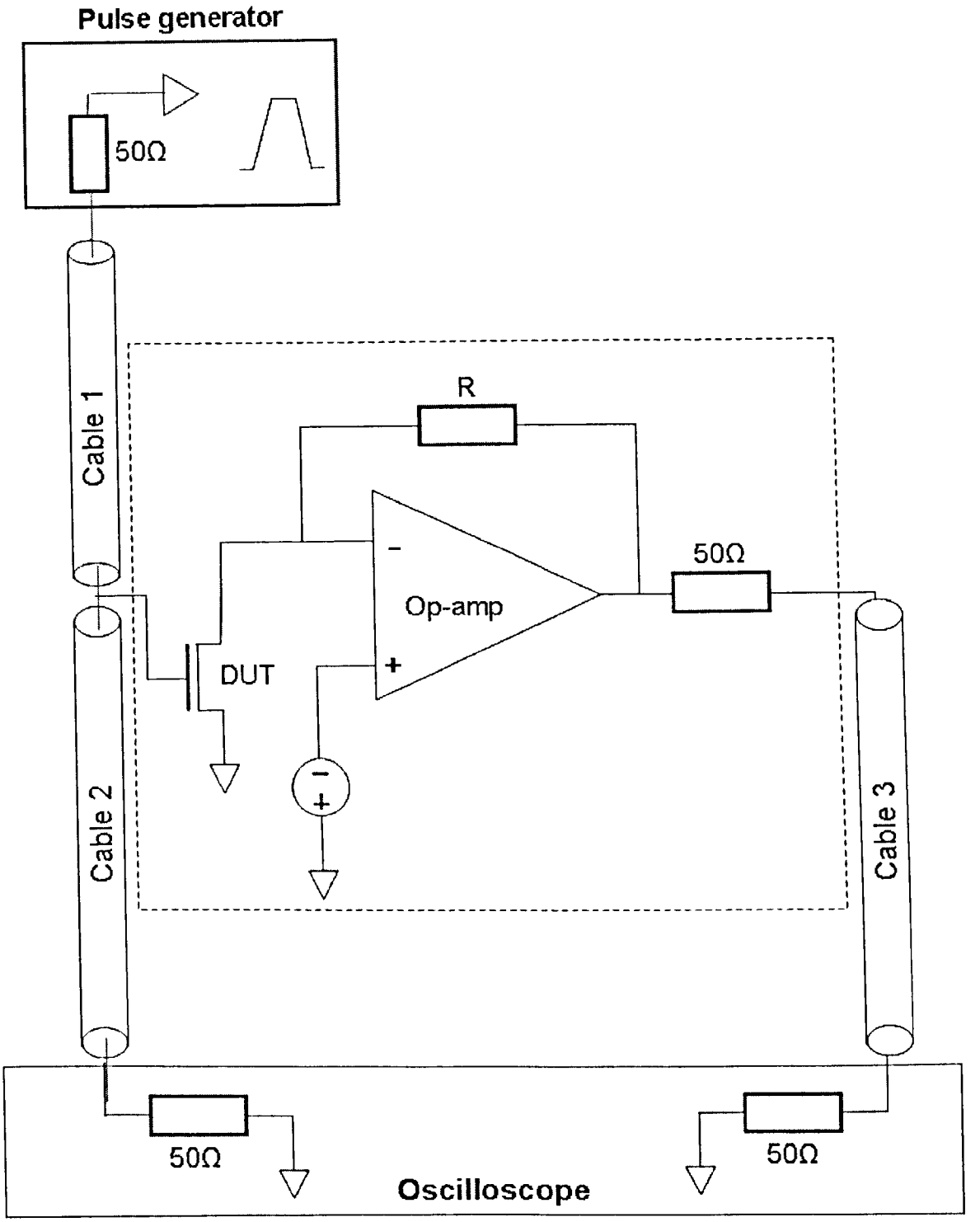


Fig 2.13 Schematic setup for ultra fast pulse measurement by using op-amplifier (AD844) with matched impedance and controlled cable delay.

The signal path of any non-impedance controlled section (*e.g.* from the drain of the transistor to the input of op-amplifier, or from the gate to the junction between cable 1 and cable 2) is less than 10 cm, in order to minimize parasitic

effects. All the transmission lines are 50 Ω co-axial cables, and cables 2 and 3 have the same length to minimize the difference in cable delay. The output impedance of the pulse generator and the input impedance of the oscilloscope are adjusted to 50Ω as well. Since the MOSFET gate has very high impedance, the short wire connecting the gate does not break the impedance matching between cable 1 and 2. A 50 Ω Resistor is used at the output of the op-amplifier to match the cable impedance, so the voltage recorded by the oscilloscope, through cable 3, is ½ of V_{out} from the op-amplifier.

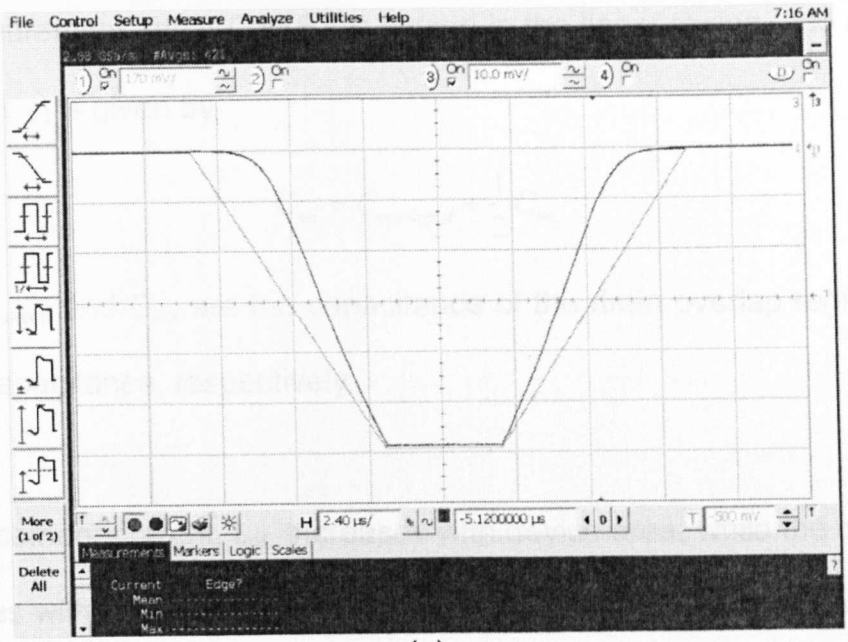
When a voltage pulse is applied to the gate, the transistor will be turned on and the drain current will induce a corresponding voltage pulse in the output voltage. Both pulses are recorded by the oscilloscope, as shown in figure 2.14 (a), and conversion from V_{out} to I_d is calculated by using equation (2.22). An plot of $V_g(t)$ and $I_d(t)$ yields the familiar I_d-V_g curve. The device used in this measurement is the same as that used in figure 2.10 (a) and clearly the accuracy is improved.

2.4.2 Source of errors

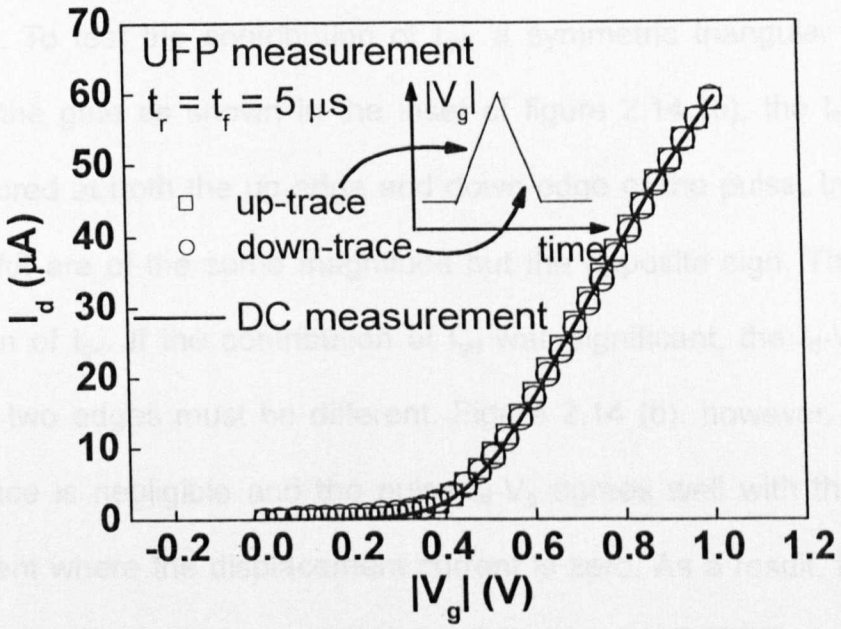
Displacement current, I_{gd}

In principle, when a pulse is applied to the gate, the current measured from the drain has two components: drift current, I_d, and displacement current, I_{gd}. The I_{gd} results from the presence of gate-to-drain capacitor, C_{gd}. By taking I_{gd} into account, equation (2.22) becomes,

$$V_{out} = (I_d - I_{gd}) \cdot R + V_d \qquad (2.23)$$



(a)



(b)

Fig 2.14 (a) Screenshot of the signals from the oscilloscope connected as shown in figure 2.13. (b) The extracted I_d-V_g curve by using equation (2.22), with V_d = -25mV during the measurement. The device used here is the same as the one used in figure 2.10. Measurement accuracy has improved significantly.

The I_{gd} is proportional to the pulse rate,

$$I_{gd} = C_{gd} \cdot \frac{dV_{gd}}{dt} \tag{2.24}$$

In the measurement, the MOSFET is biased in the linear region, and as a rough estimation, C_{gd} is given by,

$$C_{gd} = C_{overlap,d} + \frac{1}{2}C_{inv} \tag{2.25}$$

where $C_{overlap,d}$ and C_{inv} are the capacitance of the drain overlap region and the inversion capacitance, respectively.

The displacement current, I_{gd}, increases with device area, while the drift current, I_d, increases with W/L. To minimize I_{gd} and ensure $I_{gd}<<I_d$, MOSFETs with W>>L should be used. In this work, the typical L and W is 0.15 μm and 10 μm, respectively. To test the contribution of I_{gd}, a symmetric triangular pulse was applied on the gate as shown in the inset of figure 2.14 (b), the I_d-V_g curves were measured at both the up-edge and down-edge of the pulse. In these two edges, dV_g/dt are of the same magnitude but the opposite sign. This changes the direction of I_{gd}. If the contribution of I_{gd} was significant, the I_d-V_g obtained from these two edges must be different. Figure 2.14 (b), however, shows that the difference is negligible and the pulse I_d-V_g agrees well with that from DC measurement where the displacement current is zero. As a result, the effect of the displacement current is negligible for the size of MOSFETs and the pulse speed used here. The typical edge time used is limited to 5 μs, as discussed below.

Parasitic capacitors

As shown in figure 2.14 (a), V_{out} must be synchronized to V_g in order to produce the correct I_d-V_g curve. A delay of δt between V_{gs} and V_{out} will generate

approximately a horizontal shift of I_d-V_g curve,

$$\delta V_{gs} = \frac{dV_{gs}}{dt} \cdot \delta t \tag{2.26}$$

When measurement time (rise/fall time) is reduced, dV_g/dt increases, and the distortion of the I_d-V_g curve worsens. One source of this delay difference arises from the unmatched signal path length, for example, between the two long cables 2 and 3 in figure 2.13. One meter of cable length difference causes 50 ns delay time skew. If in the fast measurement, V_g ramps from 0 to 3 V in 5μs, 50 ns introduces 30 mV shift in I_d-V_g curve. This can be comparable with the degradation induced by NBTI stress. Using cables of equal length is essential therefore.

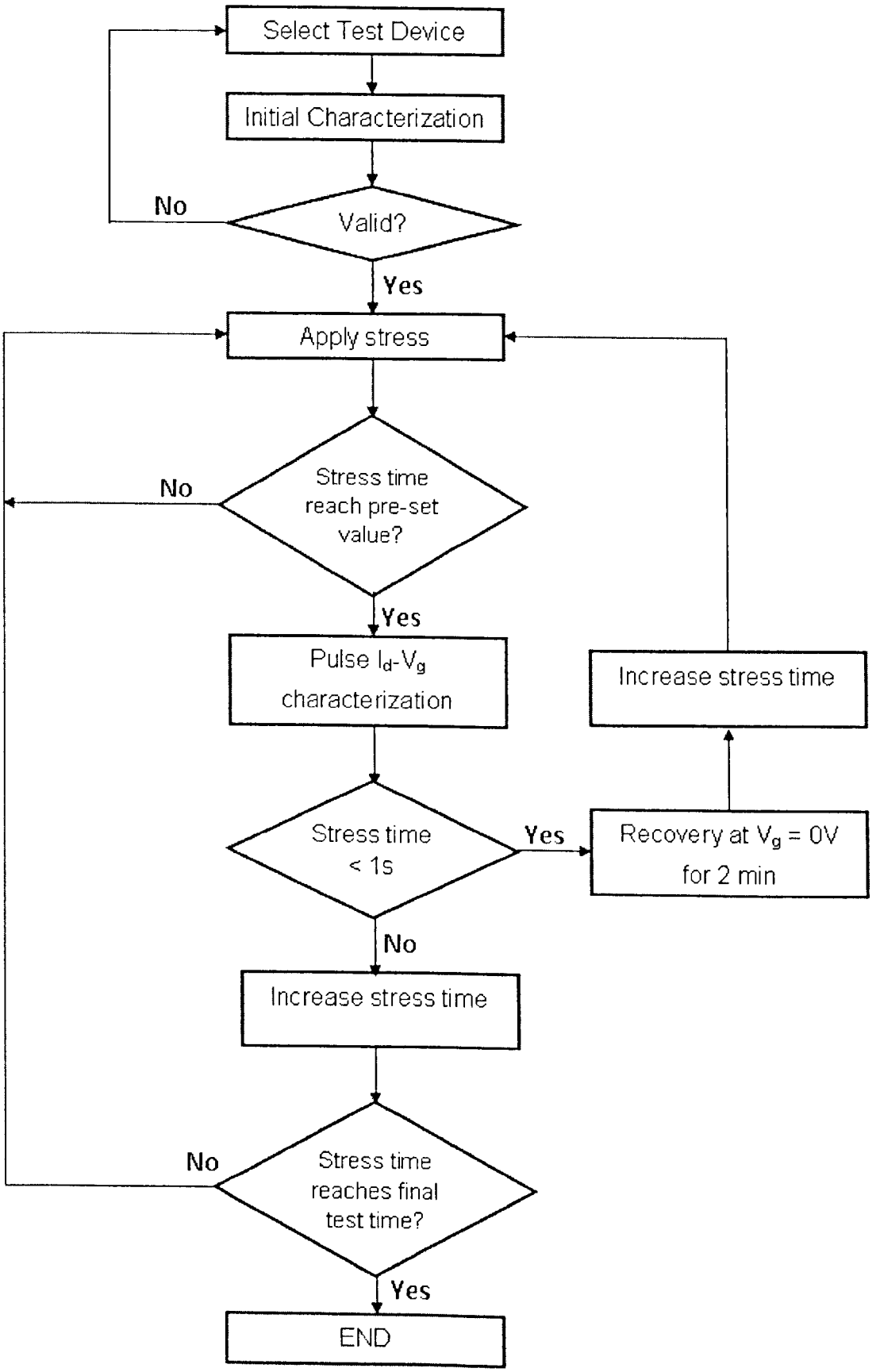
Another potential source of the delay arises from the parasitic capacitance in parallel with the feedback resistor *R*. For *R* = 10 kΩ used, a capacitance as little as 1 pF would cause a delay of 10 ns in the op-amplifier output waveform. For distortion of the I_d-V_g curve to be less than 10 mV, the maximum ramp rate for V_g need to be 5 V over 5 μs. In practice, small stray capacitance is unavoidable and its control is essential.

An error can also originate from the parasitic capacitance seen at the inverting input terminal of the op-amplifier, which consists of the intrinsic input capacitance of the op-amplifier, the capacitance between wires in the circuit and the wires connecting the probe, and finally the probe tips. It was found that, extremely short wires in the current setup are absolutely necessary.

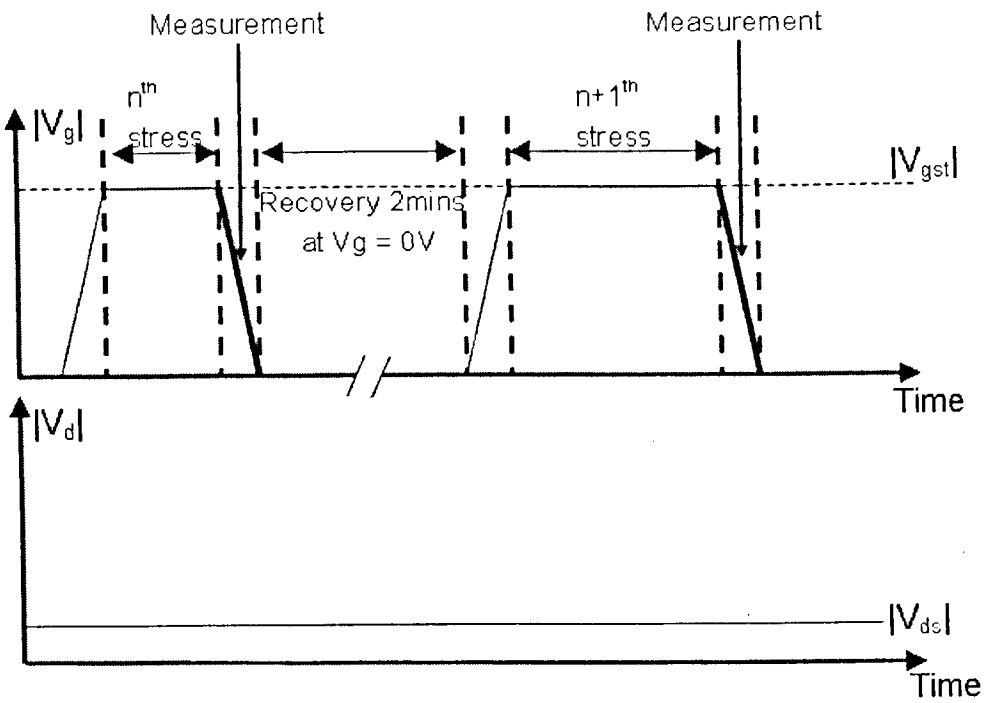
2.5 Application to the investigation of NBTI

This pulse I_d-V_g technique was applied to investigate Negative Bias Temperature Instability on devices with thin oxides. A pMOSFET with a 2.7nm SiON gate dielectric is used for demonstration here.

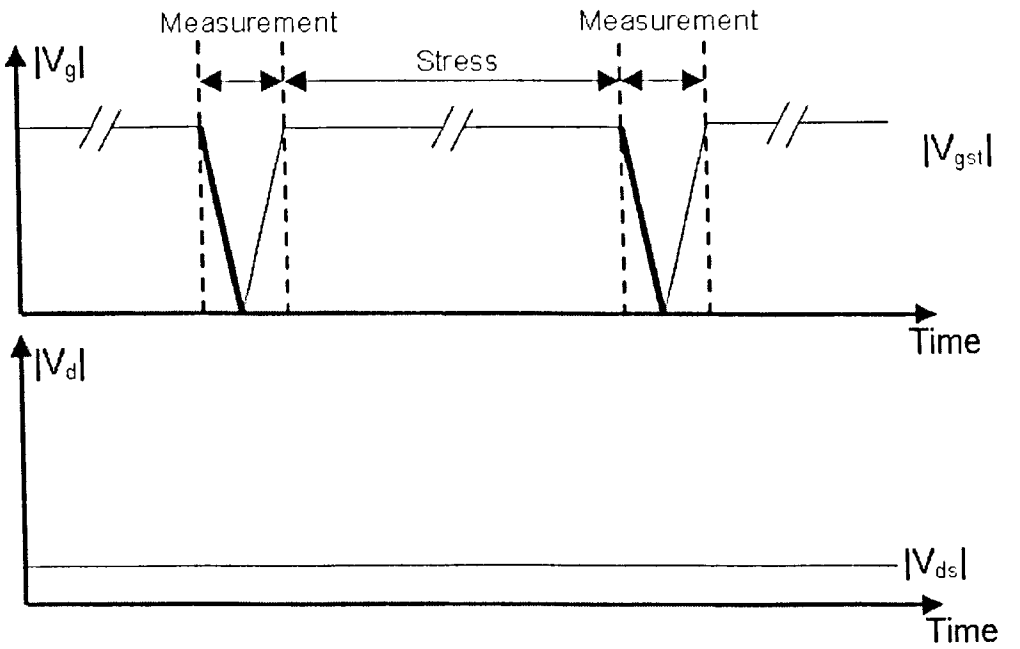
The test procedure is shown in figure 2.15 (a). Since the time for data transmission and storage for the computer is much longer than 1 s, only one I_d-V_g curve can be obtained within a 1s stress cycle. Therefore, it is necessary to repeat the stress and measurement within 1 s stress. The waveform within 1 s is schematically illustrated in figure 2.15 (b) and the waveform after 1 s is shown in figure 2.15 (c). After stress for a certain time, an I_d-V_g measurement is performed at the falling edge and then 0V is applied on the gate for 2 minutes to allow recovery. It is assumed that the device will recover to its fresh condition after 2 minutes. In order to check this assumption, an I_d-V_g curve is measured each time after 2 minutes recovery and the extracted threshold voltage is shown in figure 2.16. It can be seen that, V_t can be accurately obtained by extracting I_d-V_g measured directly after the stress and the V_t is also almost constant after 2 minutes recovery and agrees with its fresh value. This validates the assumption. One uncertainty from the procedure is that a small voltage (100mV) is always applied on the drain even during the stress, as shown in figure 2.15 (b) and (c). Therefore, the applied stress is not 100% uniform. In order to check whether this small V_d affects the degradation, different V_d was used in the measurement. As we can see in figure 2.17, V_d up to 100mV will not affect the results, so that V_d = 100mV is normally used in this work.


Select Test Device
Initial Characterization
Valid?
No
Yes
Apply stress
Stress time reach pre-set value?
No
Yes
Pulse I_d-V_g characterization
Increase stress time
Stress time < 1s
Yes
Recovery at V_g = 0V for 2 min
No
Increase stress time
Stress time reaches final test time?
No
Yes
END

(a)



(b)



(c)

Fig 2.15 (a) Typical NBTI test procedure used in this project. The device is characterized by applying a pulse on the gate. The edge time of the pulse is set at 5 μs. (b) The waveform used for measuring NBTI degradation within 1s stress time. Since the time for data transmission and storage is much longer than the stress time (<1 s), stress is repeated after recovery. The stress time is the peak time of the pulse, and the falling edge of the pulse is used to measure I_d-V_g curve within 5 μs. The edge for the measurement is represented by the thick black line in the diagram. Between two

neighbouring stress tests, 0V is applied on the gate for 2 minutes to allow the device recoverying to its fresh state. (c) The waveform used for stress time over 1 s. Data transmission and storage can be performed during the stress, and the recovery stage used in (b) is not needed here. DC voltage is applied on the gate to stress the device and one pulse is triggered when the stress time reached the pre-set level for monitoring the degradation. The edge time is 5 μs as well.

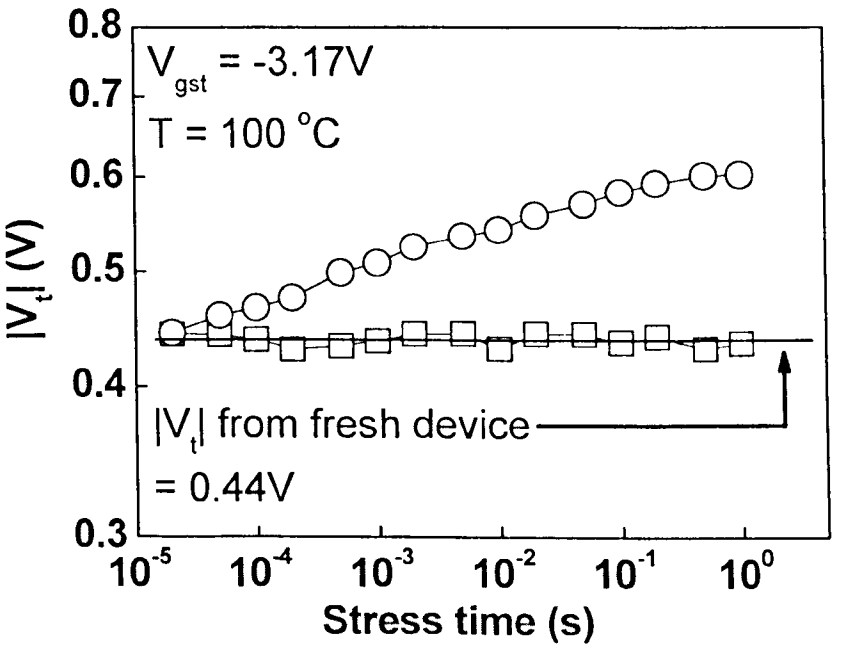


Fig 2.16 When stress time is within 1 s. The device is stressed, measured and recovered repeatedly. After each stress, the device is left with V_g = 0V for 2 minutes for recovery. After the recovery, I_d-V_g is measured before applying the next stress. This allows threshold voltage of the device after 2 minutes recovery to be compared with the fresh value. The square symbols show that the threshold voltage after recovery and the solid line is the threshold voltage extracted from fresh device.

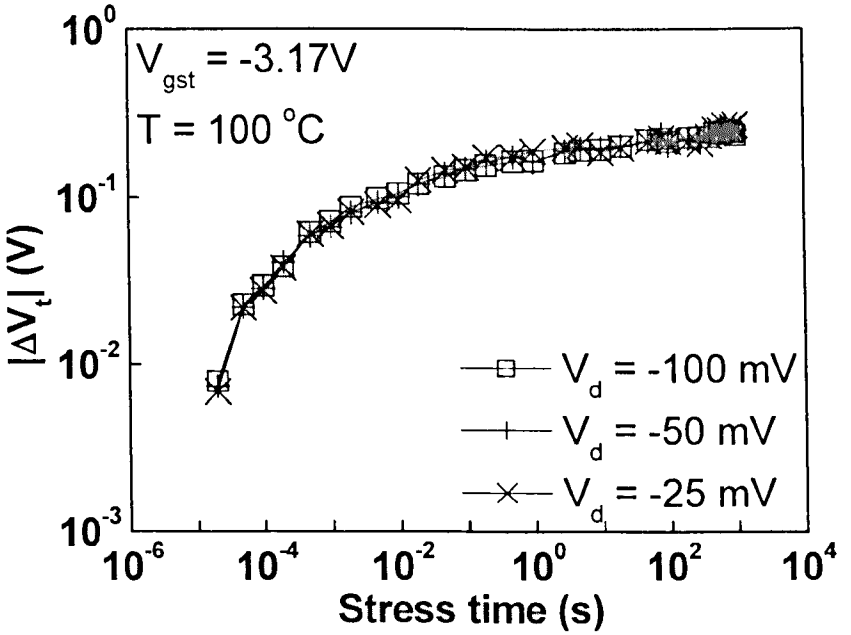


Fig 2.17 Effect of drain voltage on the NBTI degradation. Drain voltage up to 100mV will not affect the experimental result.

Finally, this technique is used to demonstrate the huge recovery phenomenon after removing stress. In figure 2.18, the device is stressed for 1000s and I_d-V_g is measured at the falling edge with different speed. The extracted V_t shift is plotted against falling edge time, i.e. I_d-V_g measurement time. It is clearly observed that as measurement time increases, the measured V_t shift drops. And the V_t measured by conventional technique is only 1/6 of the value before recovery. Also from the figure, it can be seen the measurement time of 5 μs is fast enough to freeze the recovery. As a result, 5μs measurement time is used hereafter, unless otherwise specified.

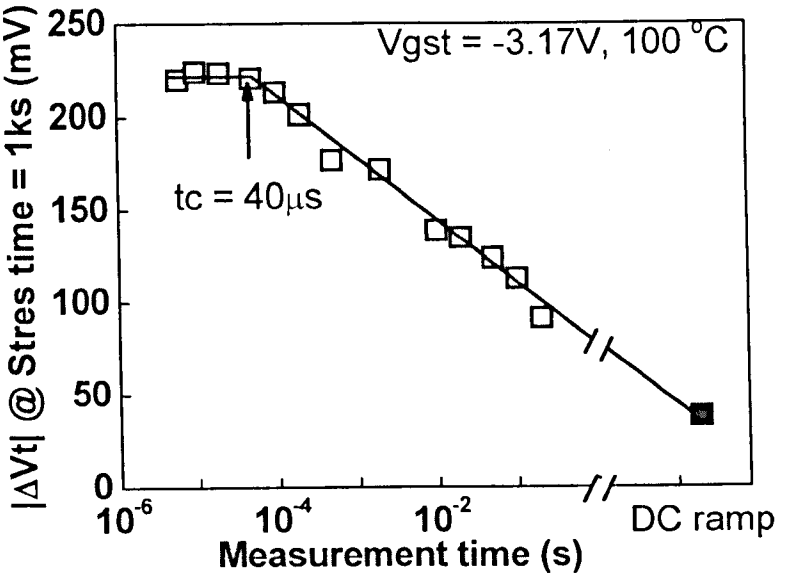


Fig 2.18 pMOSFETs with SiON gate dielectrics were stressed for 1000 s with V_g = -3.17V, after which I_d-V_g was measured with different falling edge time. The threshold voltage shift before and after stress (ΔV_t) is plotted against the measurement time. The threshold voltage shift obtained from DC measurement is also shown for comparison. The recovery during a quasi-DC measurement can reduce $|\Delta V_t|$ by a factor of 6. When the measurement time is less than t_c = 40 μs, the recovery during measurement is negligible.

2.6 Summary

In this chapter, the conventional techniques are firstly reviewed, including the measurement of transfer characteristics (I-V), capacitance (C-V), followed by the fast techniques including On-The-Fly (OTF) and ultra fast pulse I_d-V_g (UFP). The development of ultra fast pulse technique was then described. The experimental setup is described, and the sources of errors were analyzed. The details for implementing this ultra fast pulse technique have not been documented in the literature. Finally, the applicability of this technique to NBTI investigation was demonstrated together with typical test procedures used in this work.

References

1. *Operation and service Manual: Model 4140B pA Meter/DC Voltage Source,* Hewlett Packard Company, 1980.
2. *Operation Manual: Model 4145B Semiconductor Parameter Analyzer,* Hewlett Packard Company, 1989.
3. *Agilent 4155C/4156C Manual: including User's Guide, Programmer's Guide and Command Reference,* Agilent Technologies, 2003.
4. *Agilent B1500 Device Analyzer Series: Programming guide,* Agilent Technologies, Ed.7 2008.
5. A. O. Conde, F. J. G. Sanchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectronics Reliability*, vol. 42, pp. 583-596, 2002
6. D. S. Ang and S. Wang, "The recovery of the NBTI-stressed ultrathin gate pMOSFET: The role of deep-level hole traps," *IEEE Electron Dev. Lett.*, vol. 27, pp. 914-916, 2006.
7. *Semiconductor Material and Device Characterization*, D. K. Schroder, John Wiley and Sons Inc., 2nd edition, 1998.
8. W. K. Henson, K. Z. Ahmed, E. M. Vogel, J. R. Hauser, J. J. Wortman, R. D. Venables, M. Xu, and D. Venables. "Estimating Oxide Thickness of Tunnel Oxides Down to 1.4 nm Using Conventional Capacitance-Voltage Measurements on MOS Capacitors," *IEEE Electron Dev. Lett.,* vol. 20, pp. 179-181, 1999.
9. M. Kerber and U. Schwalke, "The equilibrium Controlled Static C-V Measurement," *Solid-State Electronics*, vol. 34, pp. 1141-1148, 1991.
10. G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng and D. L. Kwong, "Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling," *IEEE Electron Dev. Lett.,* vol. 23, pp. 734-736, 2002.
11. B. Zhu, J. S. Suehle, Y. Chen and J. B. Bernstein, "The negative Bias Temperature Instability of deep sub-micron pMOSFETs under pulsed bias stress," in *Integrated Reliability Workshop Final Report (IIRW),* pp. 125-129,2002.
12. M. Ershov, R. Lindley, S. Saxena, A. Shibkov, S. Minehane, J. Babcock, S. Winters, H. Karbasi, T. Yamashita, P. Clifton and M. Redford, "Transient effects and characterization methodology of negative bias temperature instability in pMOS transistors," in *international reliability physics symposium (IRPS)*, pp. 606-607, 2003.
13. 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.
14. H. Usui, M. Kanno and T. Morikawa, "Time and voltage dependence of degradation and recovery under pulsed negative bias temperature stress," in *international reliability physics symposium (IRPS)*, pp. 610-611, 2003.
15. A. Kerber, E. Cartier, L. A. Ragnarsson, M. Rosmeulen, L. Pantisano, R. Degraeve, T. Kauerauf, G. Groeseneken, H. E. Maes and U. Schwalke, "Characterization of the VT-instability in SiO2/HfO2 gate dielectrics," in *international reliability physics symposium (IRPS)*, pp. 41-45, 2003.

16. C. Shen, M. F. Li, X. P. Wang, H. Y. Yu, Y. P. Feng, A. T. L. Lim, Y. C. Yeo, D. S. H. Chan and D. L. Kwong, "Negative-U traps in HfO2 gate dielectrics and frequency dependence of dynamic BTI in MOSFETs," in *International Electron Devices Meeting (IEDM) Digest*, pp. 733-736, 2004.
17. S. Rangan, N. Mielke and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability PMOSFETs," in *International Electron Devices Meeting (IEDM) Digest*, pp. 14.3.1–14.3.4, 2003.
18. M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. R. Tauriac and N. Revil, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's," in *International Electron Devices Meeting (IEDM) Digest*, pp. 109-112, 2004.
19. M. Denais, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, N. Revil and A. Bravaix, "New methodologies of NBTI characterization eliminating recovery Effects," in *Solid-State Device Research conference (ESSDERC)*, pp. 265–268, 2004.
20. A. T. Krishnan, C. Chancellor, S. Chakravarthi, P. E. Nicollian, V. Reddy and A. Varghese, "Material dependence of hydrogen diffusion: Implications for NBTI degradation," in *International Electron Devices Meeting (IEDM) Digest*, pp. 345-358, 2005.
21. D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri and M. Alam, "On the dispersive versus Arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: Measurements, theory and implications," in *International Electron Devices Meeting (IEDM) Digest*, pp. 684-687, 2005.
22. C. Shen, M. F. Li, C. E. Foo, T. Yang, D. M. Huang, T. C. Chua, F. Nouri, G. S. Samudra and Y. C. Yeo, "Characterization and physical origin of fast *Vth* transient in BTI of pMOSFETs with SiON dielectric," in *International Electron Devices Meeting (IEDM) Digest*, pp. 333-336, 2006.

3 Real threshold voltage instability of pMOSFETs under practical operational conditions

3.1 Introduction

Gate dielectric breakdown [1,2] and negative bias temperature instability (NBTI) [3] –[10] are two major reliability issues for current CMOS technologies. To suppress boron penetration into the substrate, high nitridation concentration has been used in gate SiON, which increases NBTI [3]. Now NBTI is limiting the lifetime of pMOSFETs and it has received a lot of attention [3]-[10]. NBTI-induced threshold voltage shift, ΔV_t, was traditionally evaluated from the shift of the quasi-DC transfer characteristic I_d-V_g and its measurement can take several seconds [4].

Recently, it has been shown that significant recovery of ΔV_t can occur during this measurement delay [5,6], which, in turn, leads to an underestimation of degradation. To suppress this recovery during measurement, the "On-The-Fly (OTF)" and "ultra-fast pulse (UFP)" techniques have been developed [5-7]. The OTF technique evaluates ΔV_t at the stress gate bias V_{gst} so that the stress is

maintained during the measurement [5,6]. Since ΔV_t is effectively sensed at V_{gst}, this technique will be referred to as "OTF-V_{gst}" hereafter. The UFP technique replaces the quasi-DC I_d-V_g by a pulse I_d-V_g and the recovery can be suppressed to within measurement resolution [7,8]. Like conventional measurement, ΔV_t is also evaluated by extrapolating the pulse I_d-V_g, and it will be referred to as "UFP-ex" hereafter.

Earlier work shows that the ΔV_t obtained from the UFP-ex technique can be five times higher than that from conventional techniques and it is widely accepted that the difference is because of the recovery [7]. The ΔV_t from the OTF-V_{gst} technique can be one order of magnitude higher than that from conventional techniques and the interpretation of this large difference is still controversial [5,6]. Furthermore, ΔV_t obtained from these two fast techniques is also different and this difference is not fully understood at present.

In this chapter, the impact of various factors and uncertainties on ΔV_t evaluation by the OTF-V_{gst} technique will be analyzed first, including the degradation during measurement that affects the reference I_d and initial transconductance. The popular assumption that $|\Delta V_t|$ is independent of sensing V_g will be removed and a new evaluation method will be proposed and justified. Based on the new evaluation method, the gap in ΔV_t evaluated by different techniques will be explained. Finally, the possible physical mechanism is discussed.

3.2 Samples

The MOSFETs used in this study were manufactured at Interuniversity Microelectronics Research Centre (IMEC), Belgium. The pMOSFET used for NBTI tests has a surface channel and a p+ poly-Si gate. The channel length is 0.15 μm and the channel width is 10 μm. The gate dielectric is silicon oxynitride (SiON) with an equivalent thickness of 2.7 nm, oxidized at 850 °C and then nitrided in NO at 1050 °C for 10 s.In addition to the 2.7nm thermally nitrided sample, three wafers nitrided by DPN for different time were also used, as detailed in table 3.1 and a top view of the MOSFET layout is given in figure 3.1.

Table 3.1 Nitridation conditions of gate dielectrics and their thickness

Wafer No.	Nitridation time (s)	Gate Dielectric Thickness (nm)
1	12	1.85
2	20	2.0
3	45	2.0

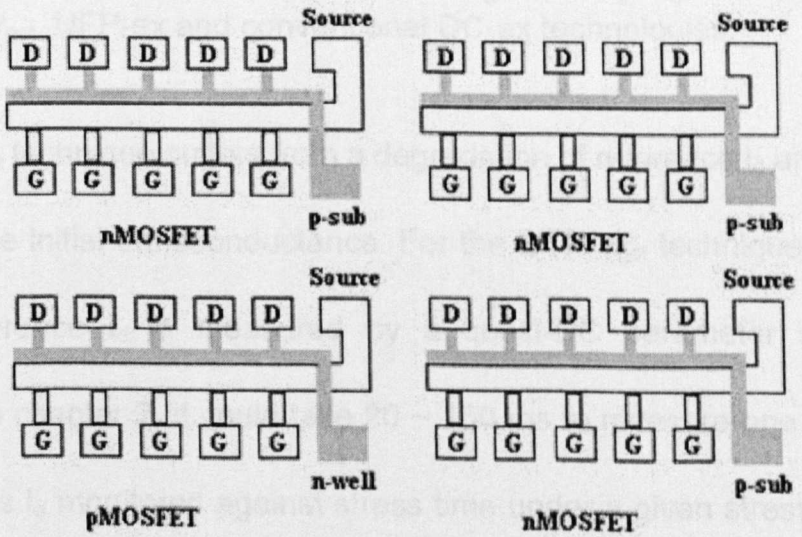


Fig 3.1 Layout of MOSFETs used in the experiments

3.3 Shortcomings of the OTF-V_{gst} technique

The comparison of threshold voltage shift, $|\Delta V_t|$, measured by OTF-V_{gst}, UFP-ex and conventional DC-ex techniques is shown in figure 3.2. Although OTF-V_{gst} and UFP-ex are both fast techniques, the obtained $|\Delta V_t|$ are different.

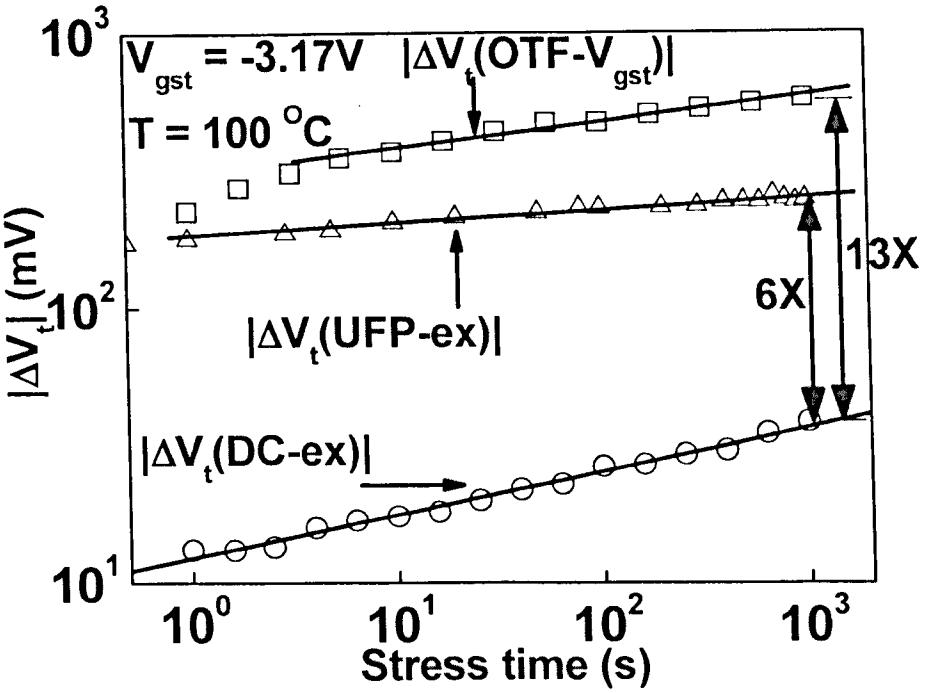
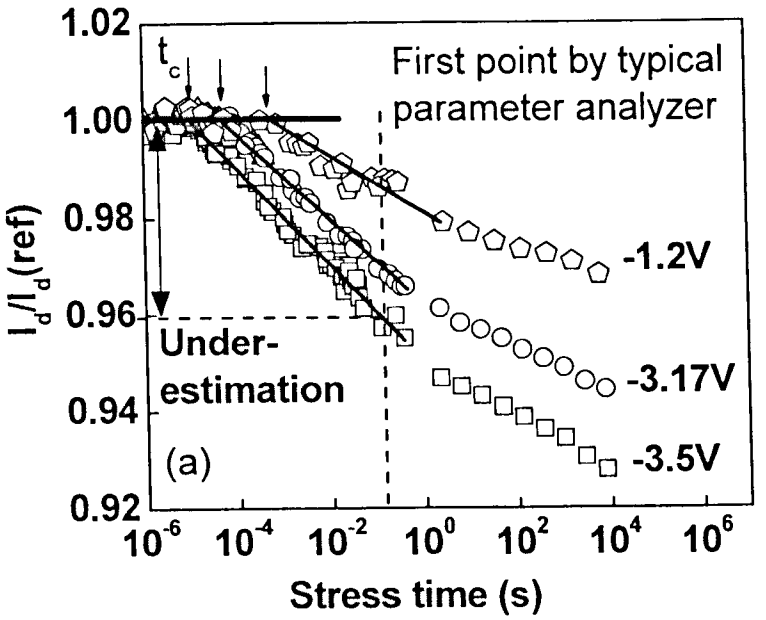


Fig 3.2 Comparison of threshold voltage shift, $|\Delta V_t|$, measured by OTF-V_{gst}, UFP-ex and conventional DC-ex technologies.

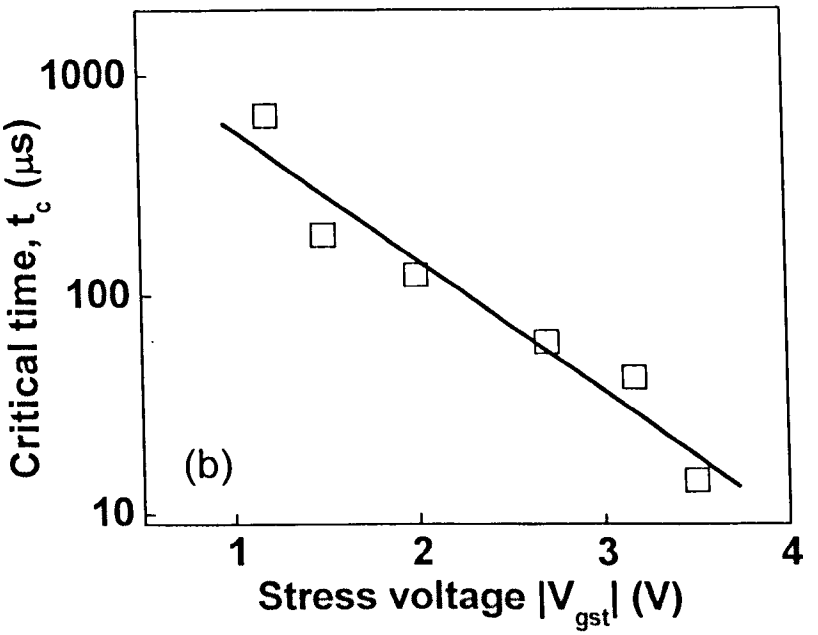
The OTF-V_{gst} technique suffers from a degradation of reference I_d and an error in evaluating the initial transconductance. For the OTF-V_{gst} technique proposed in [5], the reference I_d is measured by a quasi-DC parameter analyzer. As mentioned in chapter 2, it could take 20 ~ 150 ms to measure one point. Figure 3.3 (a) shows I_d monitored against stress time under a given stress bias. In this test, data for time within and over 1 s were measured by an oscilloscope and a

standard parameter analyzer, respectively. By using a parameter analyzer, the 1st point is measured with a rate of 150 ms per point and is marked out in figure 3.3 (a). Apparently, I_d is substantially degraded. To suppress I_d degradation during measurement to a level within a measurement resolution of 0.3%, the measurement time must be less than a critical value, t_c, which is defined as the time within which I_d is flat and degradation cannot be observed, as shown in figure 3.3 (a). When the measurement time is longer than t_c, I_d degradation becomes observable. At V_{gst}= -3.17 V, t_c is approximately 30 μs. Figure 3.3 (b) shows that t_c increases to 600 μs when V_{gst}= -1.2 V. A reduction in $|V_{gst}|$ results in a slower degradation and consequently, an increase of t_c. If the reference I_d is already degraded, ΔV_t will be under-estimated.

Unlike UFP-ex, OTF-V_{gst} requires the evaluation of g_m. Figure 2.6 (b) shows that two measurement points at $V_g=V_{gst}\pm DV$ are needed to evaluate the first g_m. The impact of I_d degradation during the measurement on g_m should also be assessed. When the measurement time for one point is 150 ms, figure 3.4 shows that the initial g_m depends on the perturbation sequence of gate bias. When $|V_g|$ was firstly decreased to $|V_{gst}-DV|$ and then increased to $|V_{gst}+DV|$, g_m was underestimated since the continuing degradation reduced the increase of I_d at $|V_{gst}+DV|$. The opposite occurs if $|V_g|$ was increased first.



(a)



(b)

Fig 3.3 Dependence of drain current on stress time and the definition of a critical time, t_c (a). (b) shows that t_c reduces for higher $|V_{gst}|$. The solid lines in (a) and (b) are guides-to-the-eye.

To minimize the I_d degradation, measurement time for the initial g_m should be limited to within t_c and figure 3.4 also gives the g_m evaluated in this way. As the stress time increases over 1 s, figure 3.4 shows that the g_m measured by two

perturbation sequences merged. This indicates that the measurement induced degradation of g_m becomes insignificant at time longer than 1 s even for a standard parameter analyzer.

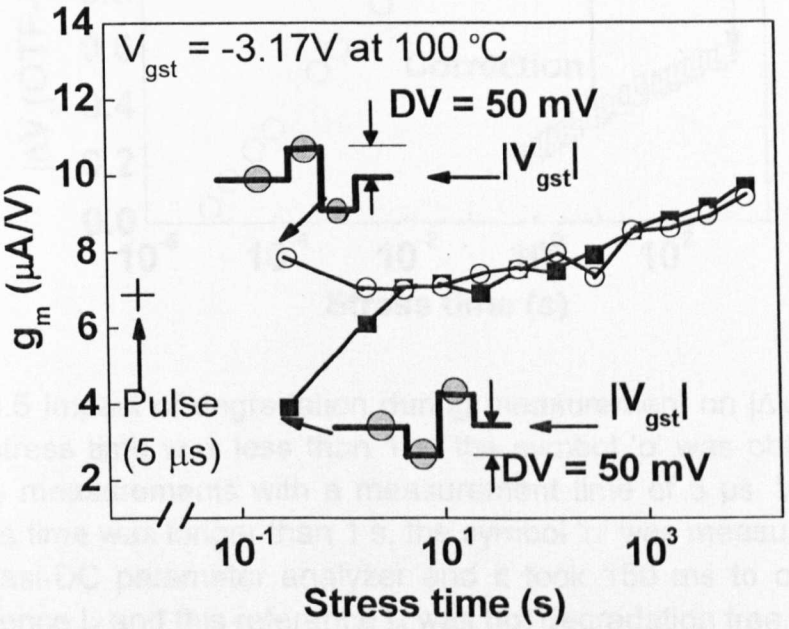


Fig 3.4 The classical transconductance, g_m, during NBTI test. The symbol '+' was obtained from a pulse (5 μs) I_d~V_g and should be used as the initial g_m. The symbol 'o' and '■' were measured by a quasi-DC parameter analyzer and the time for obtaining one I_d point is 150 ms.

By using a degradation-free reference I_d given in figure 3.3 (a) and the corrected initial g_m given in figure 3.4, $|\Delta V_t|$ is re-calculated. Figure 3.5 compares the $|\Delta V_t|$ evaluated with a measurement time of 5 μs and 150 ms. When the degraded I_d with the measurement time of 150 ms was used as the reference, $|\Delta V_t|$ was underestimated by a factor of 2.5 after 1000 s stress.

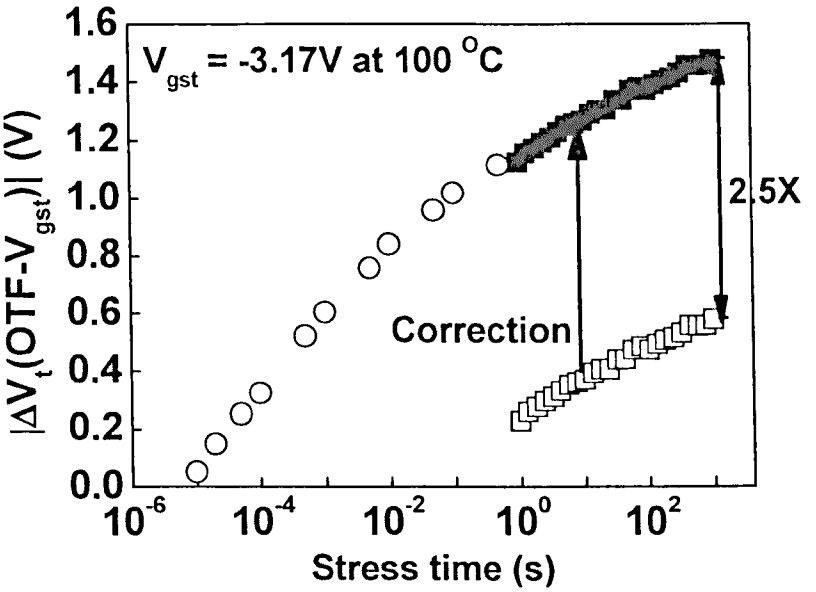


Fig 3.5 Impact of degradation during measurement on $|\Delta V_t|$. When the stress time was less than 1 s, the symbol 'o' was obtained by pulse measurements with a measurement time of 5 μs. When the stress time was longer than 1 s, the symbol '□' was measured using a quasi-DC parameter analyzer and it took 150 ms to obtain the reference I_d and this reference I_d was not degradation free.

As a result, suppression of degradation during measurement further enhances $|\Delta V_t|$ and widens its gap from the $|\Delta V_t|$ evaluated by pulse I_d-V_g technique. The $|\Delta V_t|$ now rose to above 1 V and is over one order of magnitude higher than the $|\Delta V_t|$ typically reported in earlier work [9-10].

To check whether this large $|\Delta V_t|$ originates from the poor quality of the samples, $|\Delta V_t|$ is compared with that reported in literature. Threshold voltage shift is transformed to charge density here to make direct comparison for samples of different thickness. First, figures 3.6 (a) (b) (c) show that the degradation of the SiON sample is similar to that of SiON samples reported in the literature if

measured by the traditional quasi-DC extrapolation technique [9,10,14]. Secondly, figure 3.6 (d) shows that the degradation of SiON samples is also similar to those reported in the literature if it is measured by the pulse technique and extracted by the extrapolation method [7]. Finally, the OTF I_d degradation was compared. Figure 3.7 shows that OTF $\Delta I_d/I_{d0}$ is again similar to that reported in the literature [15]. Based on the good agreement with earlier work, it is concluded that the quality of these samples is similar to the typical SiON used in earlier work.

The next question is whether the high $|\Delta V_t|$ is caused by the use of exceptionally severe stress conditions. The stress voltage V_{gst} used in figure 3.5 is -3.17V. Figure 3.2 shows that $|\Delta V_t|$ is in the order of tens of mV under these stress conditions, when measured by the conventional DC-ex technique. This level of $|\Delta V_t|$ is typically observed in earlier work [9-10] and consequently, this stress is not exceptionally harsh.

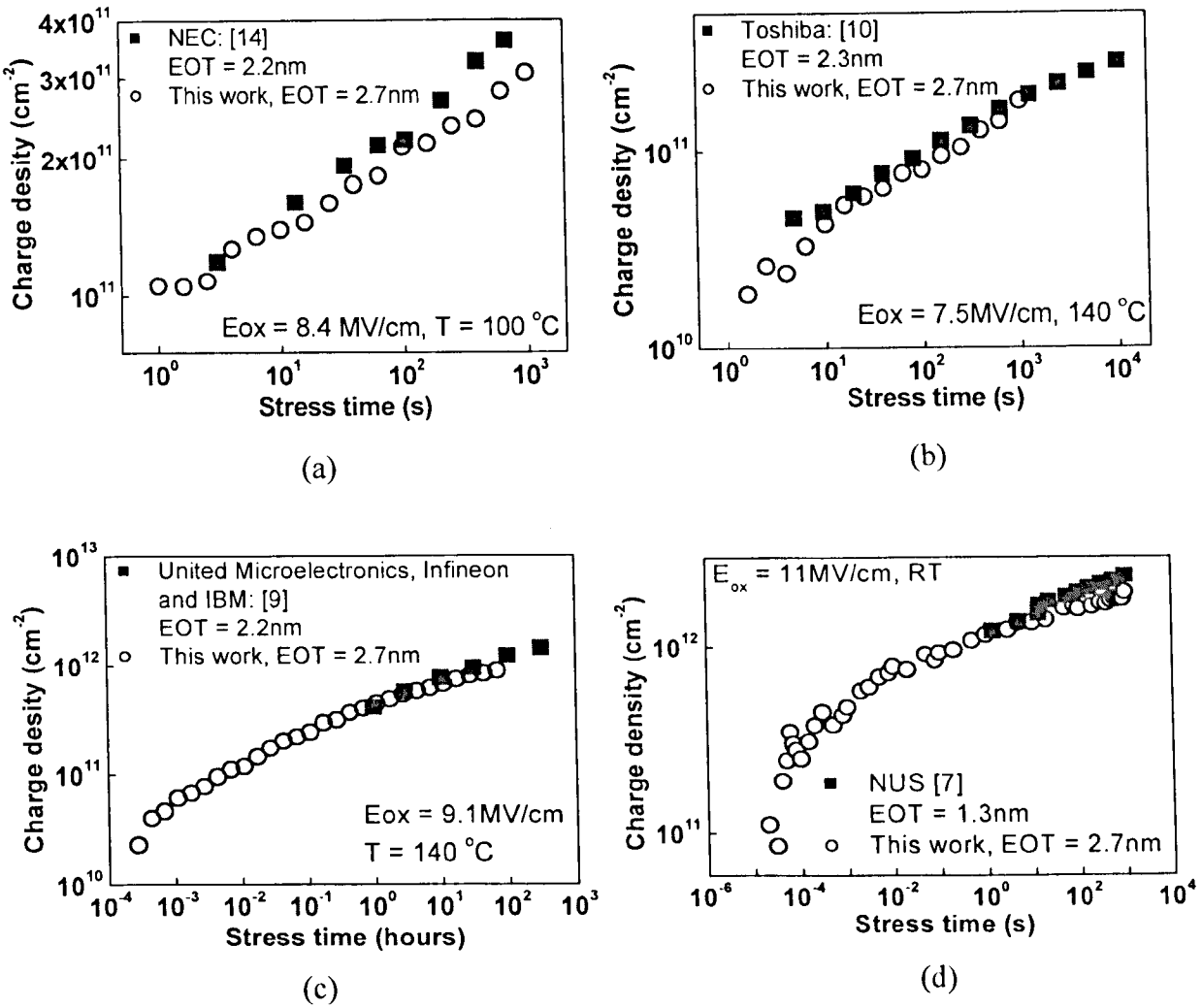


Fig 3.6 Comparison of these results with NEC [14] (a), Toshiba [10] (b), United Microelectronics, Infineon and IBM joint work [9] (c), and the (d). (a)-(c) was measured by conventional technique, but (d) was measured by pulse I_d-V_g technique. In all cases, threshold voltage is evaluated by extrapolation. The threshold voltage shift is transformed into charge density for comparing samples of different thickness.

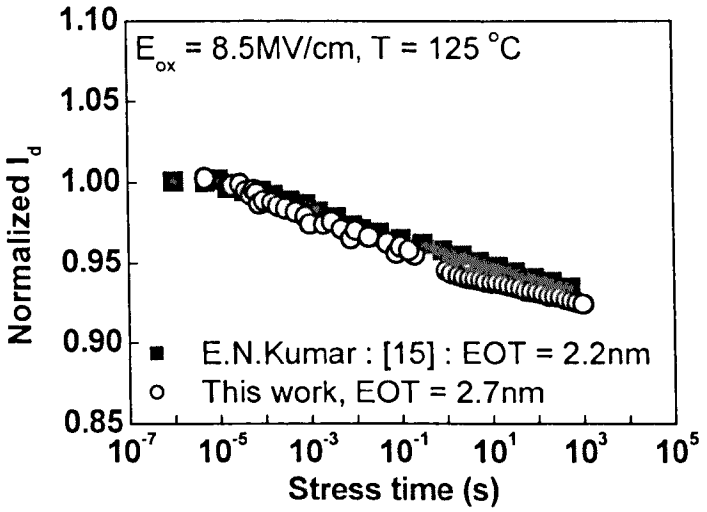


Fig 3.7 Drain current degradation comparison of the result with that of Kumar et al [15]

After clarifying that the high $|\Delta V_t|$ in figure 3.5 is not caused by the sample quality and exceptionally severe stress conditions, the next question is to understand why the $|\Delta V_t|$ given by OTF-V_{gst} is larger than that given by the UFP-ex under the same stress condition. The differences between these two techniques should be noted, namely: (i) the sensing V_g is around V_{gst} for OTF-V_{gst}, but close to V_t for UFP-ex, and (ii) OTF-V_{gst} requires the evaluation of g_m, but UFP-ex does not.

3.4 A novel evaluation method for $|\Delta V_t|$ based on the pulse I_d-V_g technique

A popular assumption is that $|\Delta V_t|$ is independent of sensing V_g, although this was not experimentally verified. To test this assumption, ΔV_t is evaluated for different sensing V_g. Using the pulse I_d-V_g technique, the whole I_d-V_g curve can be obtained within 5μs which is short enough to suppress recovery within

measurement accuracy as has been demonstrated in chapter 2. Based on the I_d-V_g curves, the degradation of I_d at each sensing $|V_g|$, $\Delta I_{d,n}(V_g)$, is evaluated from the difference in the I_d~V_g recorded at two neighbouring measurement points, as illustrated in figure 3.8 (a), and the result is given in figure 3.8 (b). The threshold voltage shift between these two points (n-1) and (n) at a given V_g, $|\Delta V_{t,n}|$, can be evaluated by using the same analysis as those for OTF-V_{gst}, namely, the equations (2.1) and (2.2). By combining the measurement of UFP-ex with the analysis of OTF-V_{gst} in this way, the range of sensing V_g from V_{gst} used by OTF-V_{gst} is extended to around V_t used by UFP -ex. The result for $|\Delta V_{t,n}|$ is plotted against sensing V_g in figure 3.8 (c) and it clearly increases with sensing $|V_g|$. The assumption that $|\Delta V_t|$ is independent of sensing V_g used in OTF-V_{gst} technique is not justified therefore. The detailed procedure and evaluation method are given next.

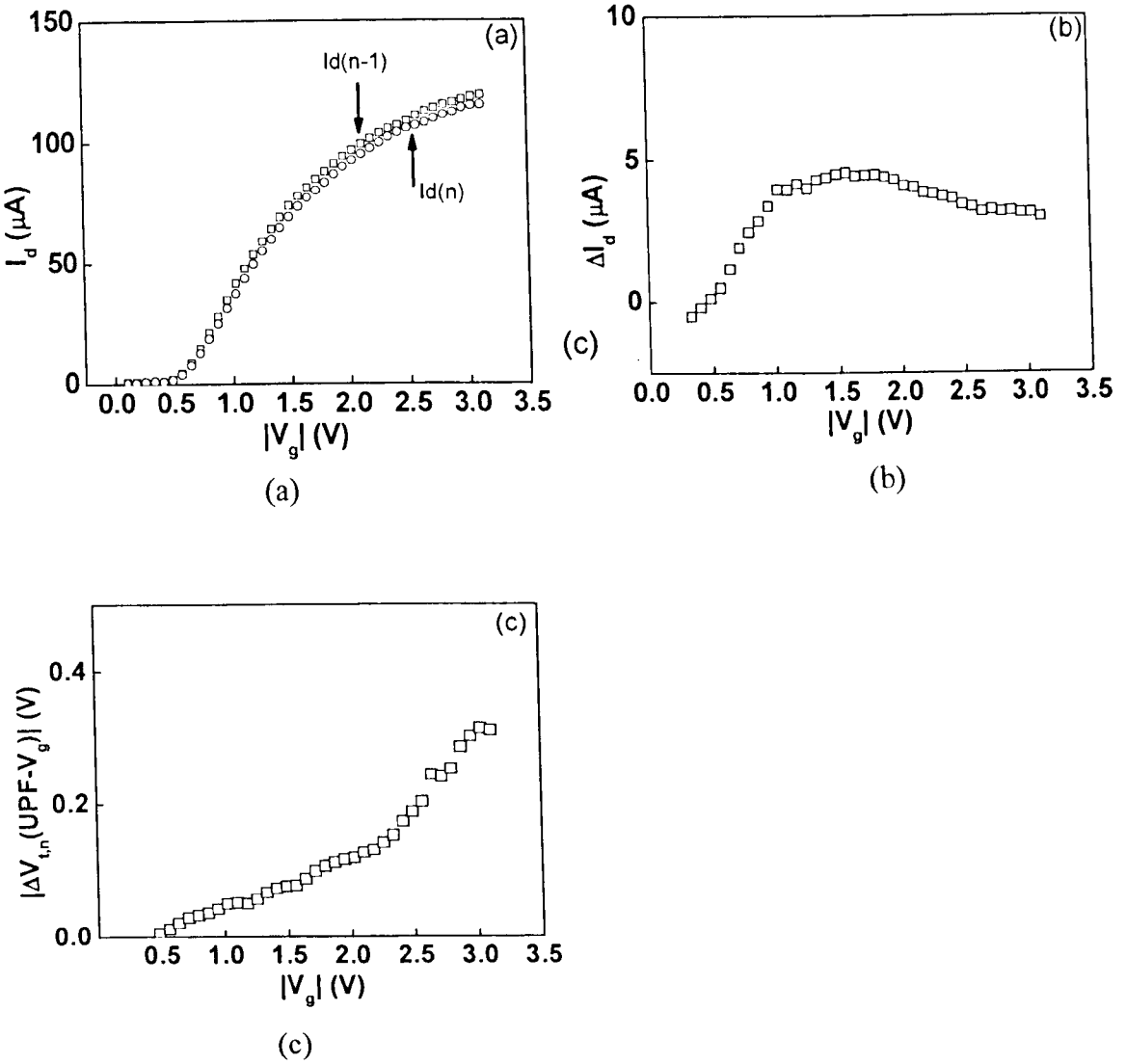


Fig 3.8 The ΔV_t evaluation by combining the measurement of UFP-ex technique with the analysis of OTF-V_{gst}. Two pulse I_d~V_g after a stress of 0.5 ms and 3 s is given in (a). The drain current difference in these two I_d~V_g, $\Delta I_{d,n}$, is caused by NBTI between the two measurement points and is plotted against sensing V_g in (b). (c) shows the threshold voltage shift between these two measurement points, $\Delta V_{t,n}$. To make the difference in the two I_d~V_g in (a) observable, a large time step, 0.5 ms ~ 3 s, is purposely used here for illustration.

3.4.1 Procedure and evaluation method

The evaluation of g_m is now examined and it is shown how to make the correction. In this measurement, V_d was set at -25 mV and $|V_d| << |V_g - V_t|$. Under these conditions, the relation between the drain current and gate bias can be simplified to

$$I_d(t) = \mu C_{ox} \cdot \frac{W}{L} \cdot V_d \cdot [V_g - V_t(t)] \tag{3.1}$$

where μ is the effective carrier mobility and C_{ox} the effective oxide capacitance per unit area. Since the I_d is clearly driven by (V_g-V_t), rather than V_g, a general definition for transconductance should be

$$G_m = \frac{dI_d}{d(V_g - V_t)} \tag{3.2}$$

The physical meaning of G_m is the variation of drain current per unit overdrive voltage, (V_g-V_t), a measure of the ability to control I_d with the gate taking into account the offset effect of threshold voltage on V_g into account. If V_t were a constant, the G_m in equation (3.2) would reduce to the classical g_m as shown in equation (2.17). When ΔV_t changes with sensing V_g, the transconductance should be evaluated by using equation (3.2). The threshold voltage shift between two neighbouring measurement, n-1 and n, $\Delta V_{t,n}$ is

$$\Delta V_{t,n} = \frac{I_{d,n} - I_{d,n-1}}{(G_{m,n} + G_{m,n-1})/2} \tag{3.3}$$

Where

$$G_{m,n-1} = \frac{dI_d}{d[V_g - V_{t,n-1}]}$$

$$G_{m,n} = \frac{dI_d}{d[V_g - V_{t,n-1} - \Delta V_{t,n}]}$$

and the accumulative degradation is

$$\Delta V_t = \sum_{n=1}^{N} \Delta V_{t,n} \tag{3.4}$$

In order to solve equations (3.3) and (3.4), iteration is needed. The flow chart of iteration is shown in figure 3.9. The iteration sequence begins with an assumed threshold voltage degradation between n-1 and n, $\Delta V_{t,n}(V_g)$. It is V_g dependent and firstly, it is calculated by using $\Delta I_d/[(g_{m,n-1}+g_{m,n})/2]$ as normally used in the OTF-V_{gst} technique. This permits a calculation of the $G_{m,n}$ and then the new $\Delta V_{t,n}(V_g)$ by using equation (3.3).The newly calculated value of $\Delta V_{t,n}(V_g)$ is compared with the assumed one. The iteration continues until the difference between two iterations becomes acceptable. In this case, the difference is set at 1mV.

3.4.2 Dependence of $|\Delta V_t|$ on sensing V_g

After using the correct G_m, figure 3.10 shows the dependence of the accumulative $|\Delta V_t|$ on sensing V_g. The correction in transconductance by using equation (3.2) reduces $|\Delta V_t|$ moderately at low sensing $|V_g|$, but strongly at high sensing $|V_g|$. This drop is because an increase of $|\Delta V_t|$ with $|V_g|$ results in a smaller denominator in equation (3.2), making G_m larger than the g_m. This in turn leads to a reduction of $|\Delta V_t|$. Even after correcting transconductance, figure 3.10 shows that $|\Delta V_t|$ can still rise substantially for higher sensing $|V_g|$. The

dependence of $|\Delta V_t|$ on sensing $|V_g|$ will be named the **Vg effect**. Since the evaluation gives $|\Delta V_t|$ at different sensing V_g, the technique will be referred to as 'UFP-V_g' hereafter.

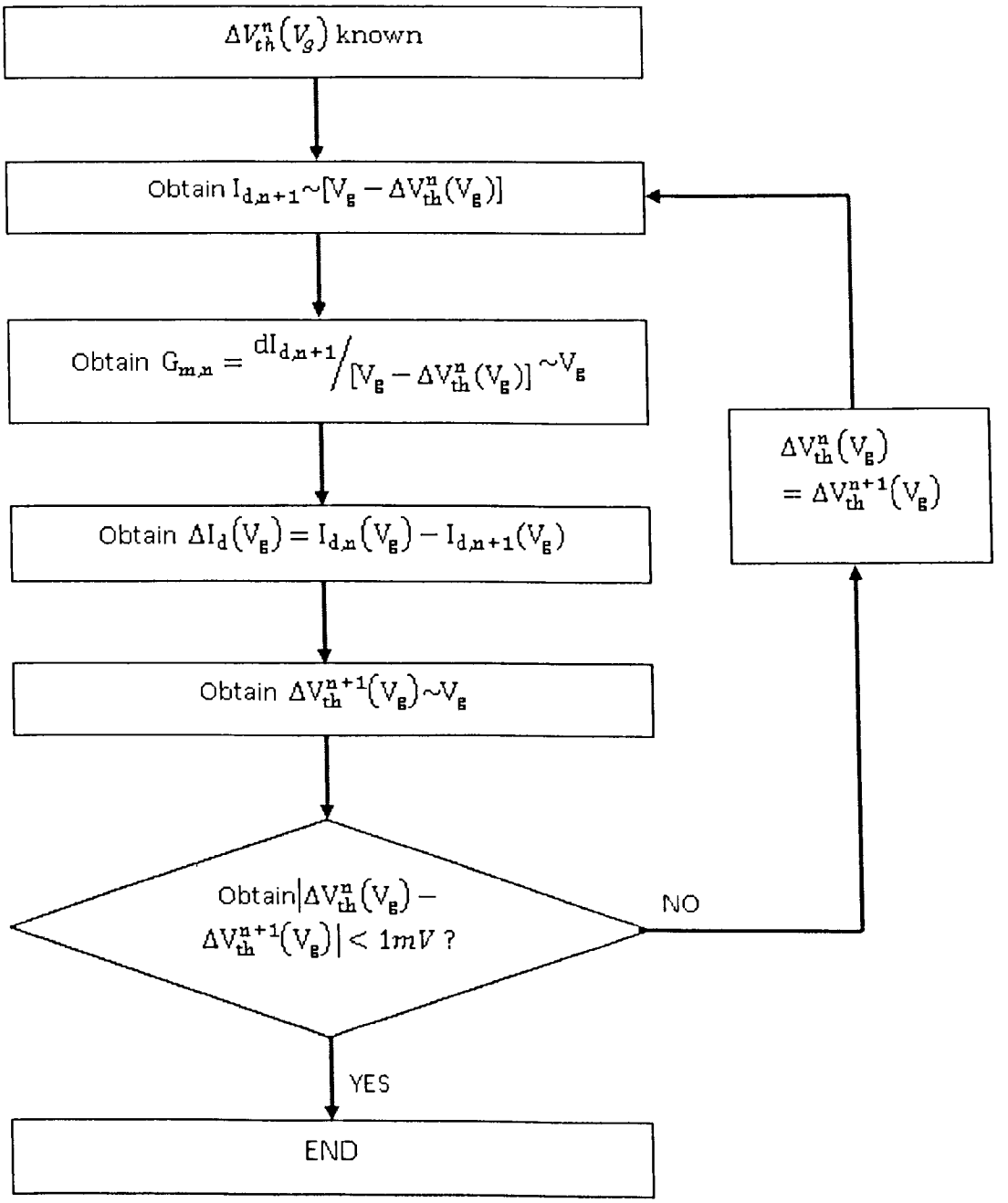


Fig 3.9 Flow chart of the $|\Delta V_t|$ evaluation by iteration.

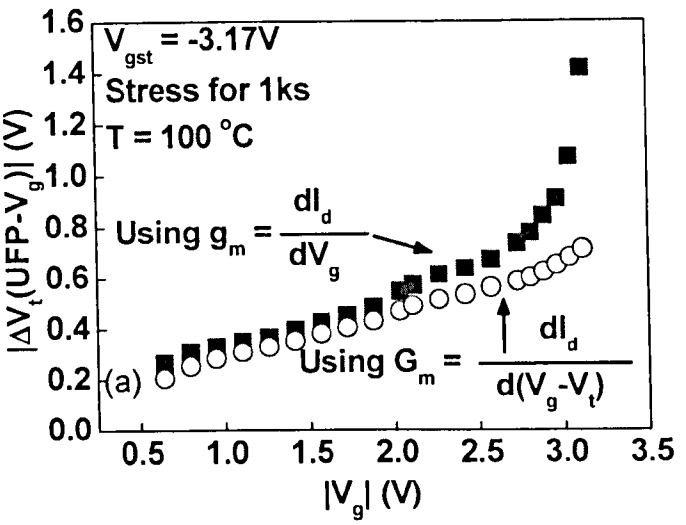


Fig 3.10 Impact of transconductance evaluation on the accumulative $|\Delta V_t|$. The sample is stressed at -3.17V for 1000 s. When the classical $g_m=dI_d/dV_g$ was used, $|\Delta V_t|$ was substantially overestimated at high sensing $|V_g|$.

In order to check whether this Vg effect only happens under high stress conditions, relatively low stress voltage V_{gst} = -2.0V is used in the next experiment. Figure 3.11 shows that the feature of the V_g dependence on $|\Delta V_t|$ is not changed.

The result shown above was obtained on the thermally nitrided SiON dielectric. To ensure the Vg effect is not process-specific, the same technique is applied on 1.85nm – 2.0nm SiON gate devices which have been nitrided by DPN. Figure 3.12 shows that the feature of the Vg effect always occurs, indicating it is insensitive to processing. The degradation is higher after longer nitridation as expected.

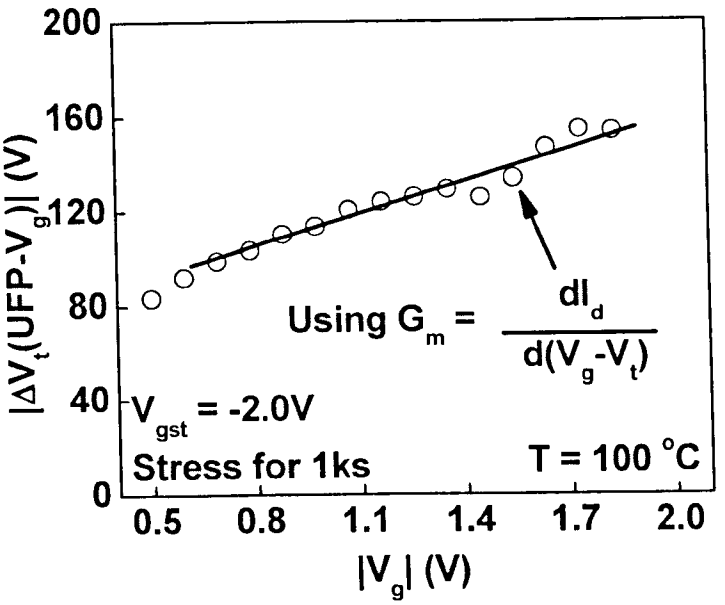


Fig 3.11 Dependence of $|\Delta V_t|$ on sensing V_g after stressing at a relatively low V_{gst} = -2.0V for 1000 s. The V_g effect is still substantial.

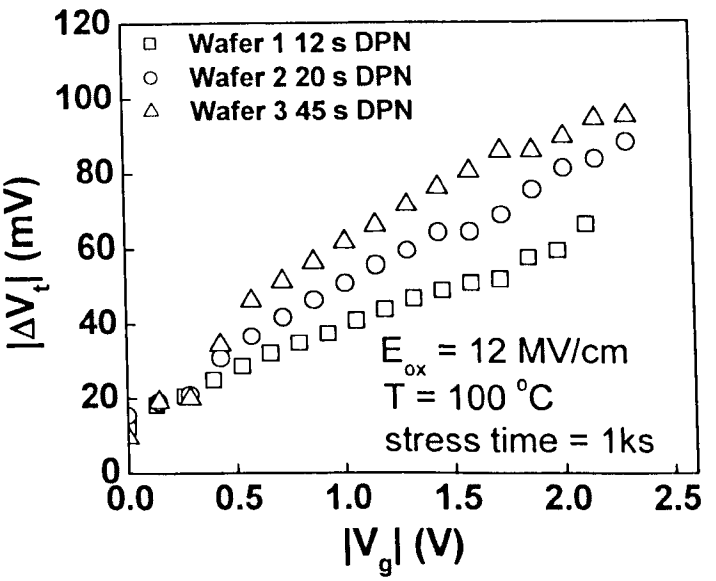


Fig 3.12 Insensitivity of the Vg effect to fabrication processes. The samples used here are nitrided by DPN with different nitridation time, leading to a different nitridation concentration.

3.5 Justification of the Vg effect on $|\Delta V_t|$

In order to justify the Vg effect on $|\Delta V_t|$, possible sources of error should be carefully examined, including:

1) The accuracy of the measurement.
2) Truncation errors from the evaluation procedure
3) Gate leakage current contribution to the drain current.

The measurement accuracy is addressed first. Recently, Reisinger et al [16] have pointed out that the measurement accuracy should be as high as $10^{-4}*I_d$. They show that error propagation outside the measurement range of V_g can cause substantial errors in the $|\Delta V_t|$ evaluation, as shown in figure 3.13.

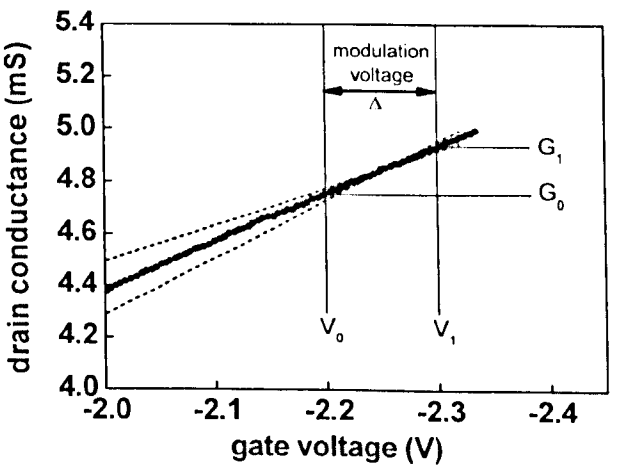


Fig 3.13 Figure from Reisinger et al work [16]. The dashed straight lines illustrate how even a small change in conductance at stress level leads to a large variation at the lower V_g where the threshold voltage is evaluated.

Reisinger et al [16] reported that if the I_d-V_g measured near to the stress V_g is extrapolated towards lower $|V_g|$, both statistic and systematic errors will be

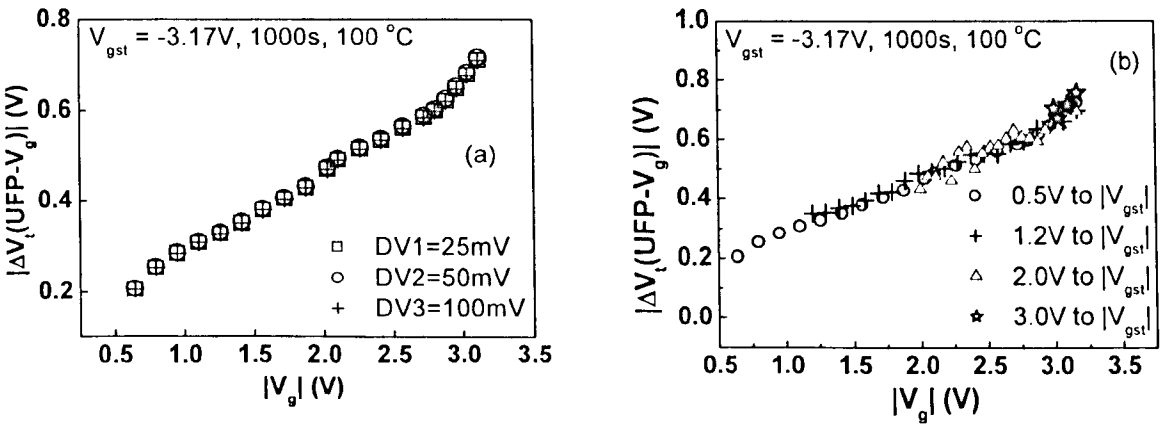
enlarged. This means that the I_d measured at a high $|V_g|$ close to the stress level should not be used to determine the $|\Delta V_t|$ at the conventional low $|V_g|$. In this evaluation, however, the I_d-V_g obtained between 'Vo' and 'V1' were used, as shown in figure 3.13, to determine the $|\Delta V_t|$ at the sense voltage. Furthermore, between 'Vo' and 'V1' no extrapolation of I_d-V_g was used. To obtain the $|\Delta V_t|$ at a lower $|V_g|$, the I_d-V_g covering that $|V_g|$ is used. As a result, errors do not propagate in this evaluation process.

Within DV, I_d is assumed to be a linear function of V_g and this assumption introduces truncation errors. The larger DV, the larger will be the truncation error.

To control this error, DV must be sufficiently small so that a further reduction hardly improves the results. Figure 3.14 (a) shows that there is little difference in the result between DV=100 mV and DV=25 mV and so DV=25 mV is used in this work. The effect of the I_d~V_g range on the $|\Delta V_t|$ is also studied and figure 3.14 (b) shows that $|\Delta V_t|$ at a given sensing V_g is insensitive to how low $|V_g|$ was ramped down, so long that the I_d-V_g covers this sensing V_g.

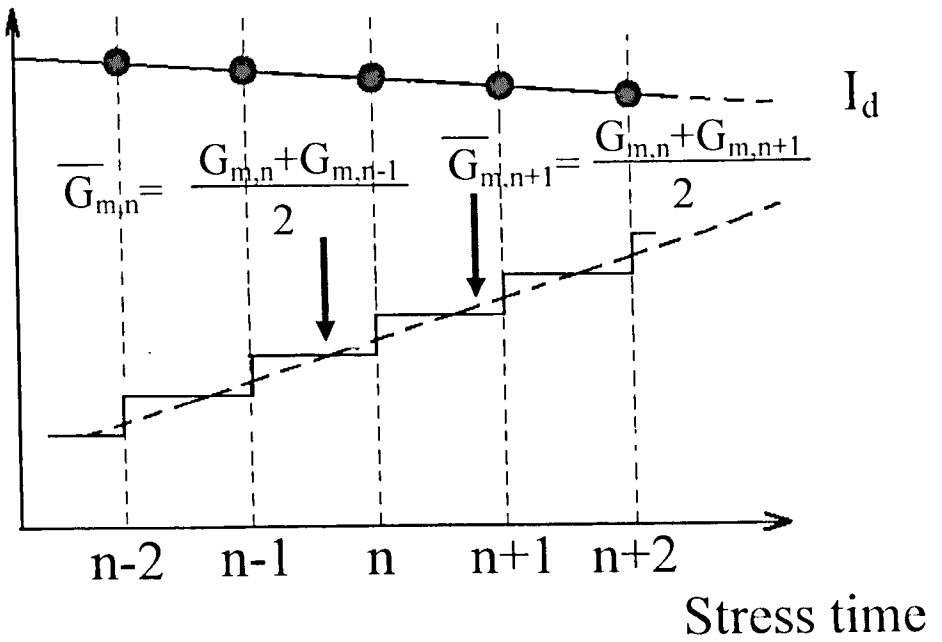
During NBTI tests, figure 3.15 (a) illustrates that G_m was evaluated only at pre-specified points, rather than continuously. This means that the continuing variation of G_m is approximated by a step function and the G_m between two neighbouring points is treated as a constant. The step in time must be sufficiently small to control the truncation error in time. Figure 3.15 (b) shows that the use of

8 points per decade is adequate. A constant number of points per decade time is used, instead of a constant Δt, since degradation rate slows down and larger Δt can be used for longer stress time.

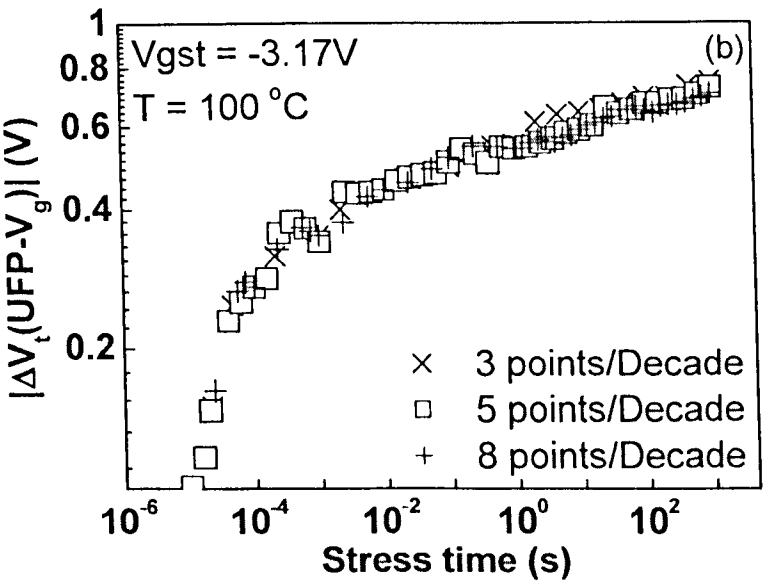


(a) (b)

Fig 3.14 An assessment of the effect of evaluation errors on ΔV_t. (a) compares ΔV_t obtained by using three different DV for estimating G_m through equation (3.2). The negligible difference in ΔV_t confirms that truncation errors in DV are insignificant. (b) compares ΔV_t obtained by ramping down $|V_g|$ from $|V_{gst}|$=3.17 V to 3, 2, 1.2 and 0.5 V, respectively. It shows that the range of I_d~V_g has little effect on ΔV_t at a given sensing V_g, so long that this sensing V_g is covered by the I_d~V_g.



(a)



(b)

Fig 3.15 An assessment of the effect of time truncation error on ΔV_t. (a) shows that the continuous variation of G_m with stress time is approximated by a step function. (b) shows that an increase of number of points per decade of stress time has little effect on ΔV_t, indicating that the time truncation error is negligible.

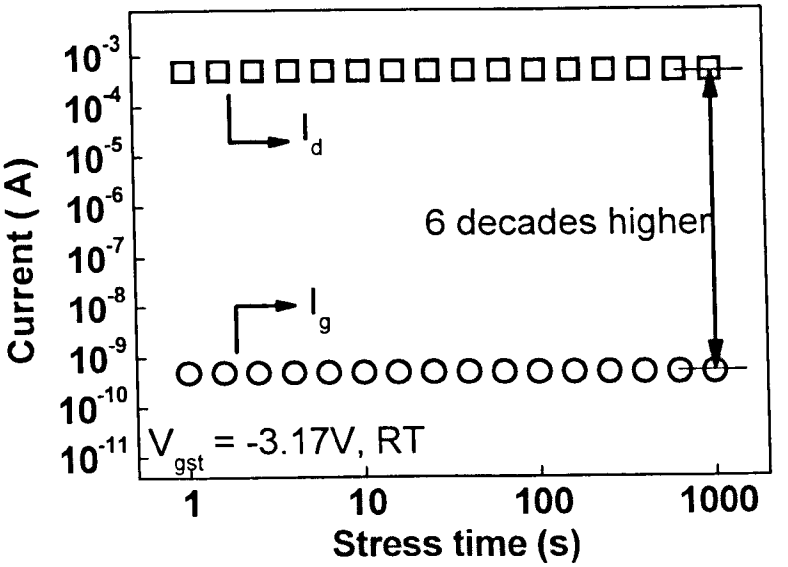
One requirement for the new OTF-V_g technique is that gate leakage current should be negligible when compared with the drain current during the whole stress time and also over the whole range of sensing V_g. For this sample, $I_g << I_d$ is always valid as shown in figure 3.16 (a) and (b). Therefore, it will not introduce errors to V_t extraction.

After the correction, $|\Delta V_t|$ can be as large as 0.7V, which corresponds to a charging level of 6×10^{12} cm^{-2}, approximately, for a 2.7 nm SiON. Although such a high level of charging is rarely reported by earlier work on NBTI, it was routinely observed under other types of stress, such as substrate hole injection [17], as shown in figure 3.17. As a result, it is physically realistic and achievable.

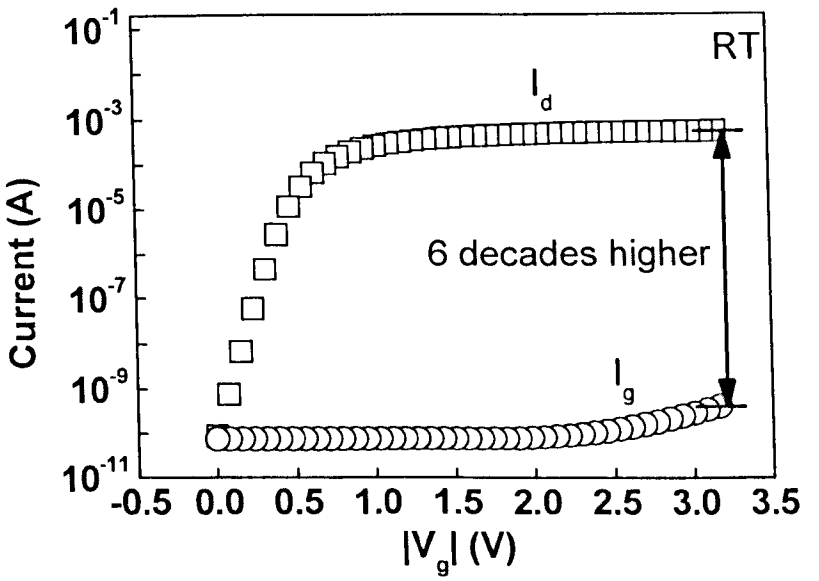
Though $|\Delta V_t|$ can be as large as 0.7V, a MOSFET can still be switched on. Figure 3.18 shows that, $|V_t|=|V_{t0}+\Delta V_t|$ increases with $|V_g|$ at a slower rate than $|V_g|$ itself and consequently, $|V_g-V_t|$ still increases with $|V_g|$. For example, even if $|\Delta V_t|=0.7V$ gave $|V_t|$ = 1.14V at $|V_g|=3.17V$, there is still an over-drive voltage of $|V_g-V_t|$= 2.03V, so that the MOSFET will be switched on.

3.6 Effect of mobility variation in the evaluation

The OTF-V_g technique can give $|\Delta V_t|$ over a wide range of sensing $|V_g|$. When $|\Delta V_t|$ is sensed near the threshold condition, hole concentration in the p-channel is negligible and the impact of mobility variation on $|\Delta V_t|$ is normally ignored.



(a)



(b)

Fig 3.16 (a) I_d and I_g current variation with stress time. $I_g << I_d$ is valid over the whole degradation stress period. (b) I_d and Ig current variation with sensing V_g. $I_g << I_d$ is valid over the whole range of V_g. 2.7nm SiON sample is used here.

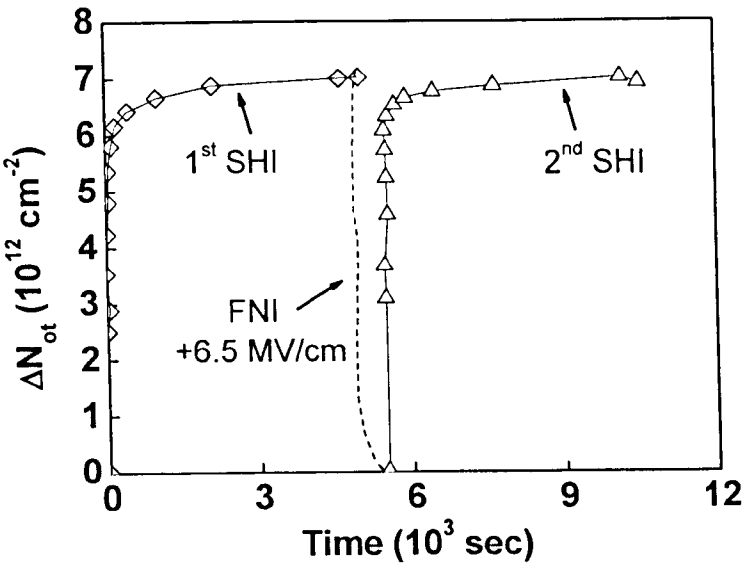


Fig 3.17 Positive charging during substrate hole injection (SHI). The charging density can be routinely over $6x10^{12}$ cm^{-2}. The figure is taken from [17].

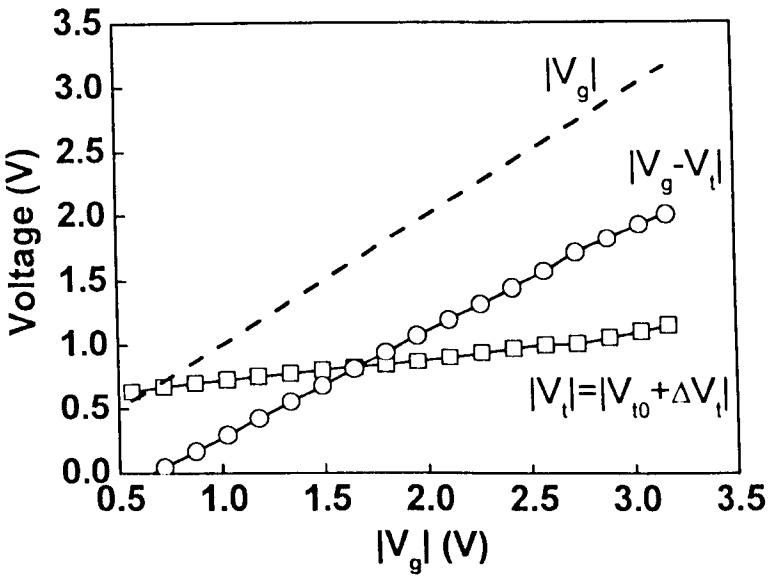


Fig 3.18 The device is stressed at -3.17V for 1ks under 100 °C condition. $|\Delta V_t|$ can be as large as 0.7V. Although an increase of $|\Delta V_t|$ results in an increase of $|V_t|$ for higher $|V_g|$, $|V_g-V_t|$ still increases with $|V_g|$, so that the MOSFET can be switched on and I_d increases with $|V_g|$.

When $|\Delta V_t|$ is sensed in strong inversion at higher $|V_g|$, there is a strong channel and the impact of mobility variation on $|\Delta V_t|$ must be evaluated. Degradation of carrier mobility during stress results in a reduction of I_d and will give an apparent $|\Delta V_t|$, even though the threshold voltage was not shifted. In this sense, $|\Delta V_t|$ measured by the OTF-V_g technique could be over-estimated. It is not explicit whether the effect of mobility variation has been taken into account in the new technique. This uncertainty is addressed in the following.

With numerical approximation, one can have

$$\Delta I_d \approx A \cdot (V_g - V_t) \cdot \Delta\mu + A \cdot \mu \cdot \Delta(V_g - V_t) \tag{3.5}$$

As expected, both mobility variation, $\Delta\mu$, and a shift of $(V_g - V_t)$ can change I_d.

To facilitate the discussion, a parameter is defined,

$$\alpha = \frac{(V_g - V_t) \cdot \Delta\mu}{\mu \cdot \Delta(V_g - V_t)}$$

If mobility variation were neglected, $\alpha=0$ would be required. It should be emphasized that the OTF-V_g technique does not assume $\alpha=0$. If $\alpha=0$ were assumed, $G_m = A^*\mu$, which is not valid under the present test condition.

From the evaluation, although G_m was treated as a constant between two neighbouring measurement points, it has not been assumed as constant during the whole stress period. Figure 3.15 (a) illustrates schematically the approximation used. Since the analytical expression of G_m against stress time is not known, the continuous function of G_m vs time (the dashed line) is replaced by

the step function. In the time step between 'n-1' and 'n', the G_m is considered as a constant, that is the average of its value at 'n-1' and 'n'. For the next time step between 'n' and 'n+1', it is considered as constant again, but this constant is changed and is now the average of its values at 'n' and 'n+1'. As a result, although they are approximated as constants within one time step, they are changing from one step to the next.

It is well known that mobility is reduced for higher $|V_g|$. This, however, will not cause an increase of $|\Delta V_t|$ for higher $|V_g|$. In figure 3.19 (a), an I_d-V_g is shifted by 0.1 V to simulate the case that the charging does not change with V_g. Figure 3.19 (b) shows the g_m before and after the shift and the average value. It confirms that g_m indeed reduces for higher $|V_g|$. Figure 3.19 (c), shows that ΔI_d also reduces for higher $|V_g|$, as a consequence of smaller g_m. Figure 3.19 (d) clearly shows that the $|\Delta V_t| = |\Delta I_d/g_m(\text{average})|$ is independent of V_g and equals the expected value of 0.1 V. Therefore, the increase of $|\Delta V_t|$ with $|V_g|$ observed in these tests is not caused by a reduction of g_m for higher $|V_g|$.

To evaluate the variation of mobility during NBTI, from equations (3.1) and (3.5), write

$$\frac{\Delta\mu(n)}{\mu(n)} = \frac{\Delta I_d(n)}{I_d(n)} + \frac{\Delta V_t(n)}{V_g - V_t(n)} \tag{3.6}$$

The accumulative mobility variation is

$$\frac{\Delta\mu(accumulative)}{\mu_0} \approx -\sum_{n=1}^{N} \frac{\Delta\mu(n)}{\mu_0} \tag{3.7}$$

Under the typical test conditions, the relative change of drain current, threshold voltage and the effective mobility are given in figure 3.20 (a) and (b), respectively. The $\Delta V_t/(V_g-V_t)$ can be over 30%, corresponding to a shift of V_t larger than 0.7V. This, however, only leads to about 10% percent degradation of drain current, because the effective mobility at the given stress voltage actually increased.

During stress, it is reported that the carrier mobility will degrade, contradicting the mobility increase observed in figure 3.20 (b). To explain it, it should be pointed out that the effective mobility used in figure 3.20 (b) is different from the low field mobility, μ_{LF}. The relation between these two can be expressed as [18]

$$\mu = \frac{\mu_{LF}}{1+\theta\cdot\left(V_g - V_t - 0.5\cdot V_d\right)} \tag{3.8}$$

The mobility degradation reported in earlier work [19,20] is for μ_{LF}. Under the test condition employed here, there is little doubt that μ_{LF} is also degraded. To show this degradation, the I_d-V_g before and after a typical NBTI stress is plotted in figure 3. 21 (a) and the corresponding transconductance is given in figure 3. 21 (b). The maximum g_m is proportional to the low field mobility, μ_{LF}, and it is clear that stress reduces the maximum g_m.

Equation (3.8) shows that the effective mobility increases for lower $|V_g-V_t|$. The stress was carried out under a constant $|V_g|$. As time increases, the magnitude of threshold voltage increases, leading to a reduction of $|V_g-V_t|$. A smaller $|V_g-V_t|$

can over-compensate the effect of μ_{LF} degradation, resulting in an increase of μ. An increase of μ is also supported by the increase of g_m at stress gate bias in figure 3.4.

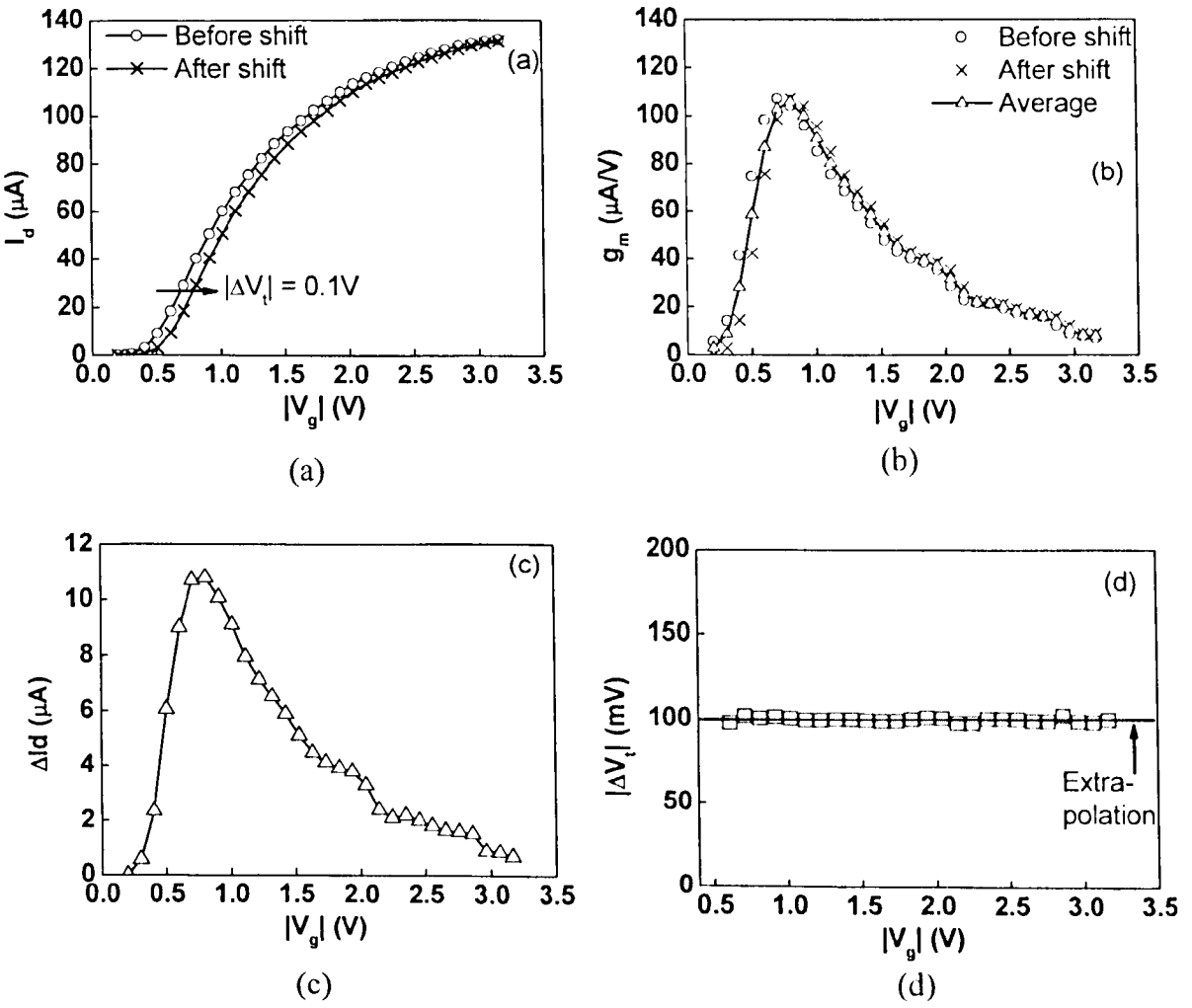
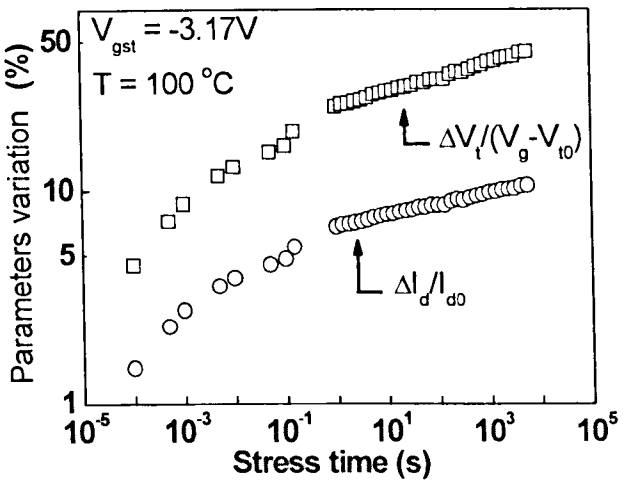
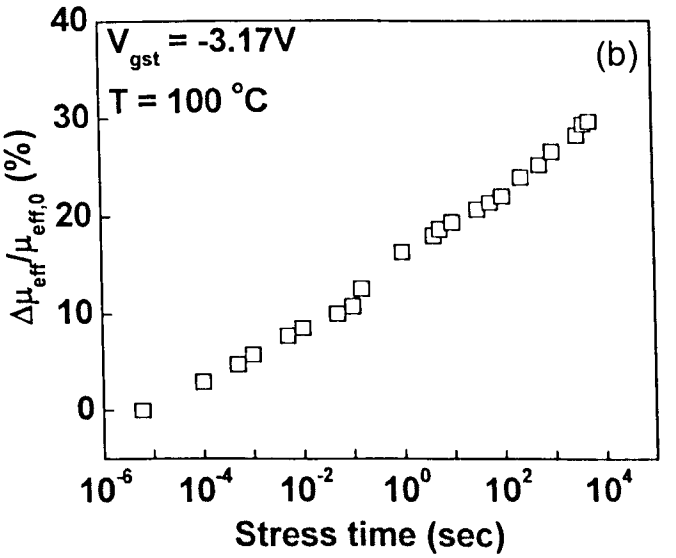


Fig 3.19 A demonstration that a reduction of g_m for higher $|V_g|$ does not result in an increase of $|\Delta V_t|$ for higher $|V_g|$. In (a), the I_d-V_g was shifted in parallel by 0.1 V. In (b), g_m was evaluated from the I_d-V_g both before and after the shift. The average value of these two is also given. In (c), the difference in I_d for the two curves in (a) was given. In (d), $|\Delta V_t|$ was evaluated by using $|\Delta V_t| = |\Delta I_d / g_m(\text{average})|$.

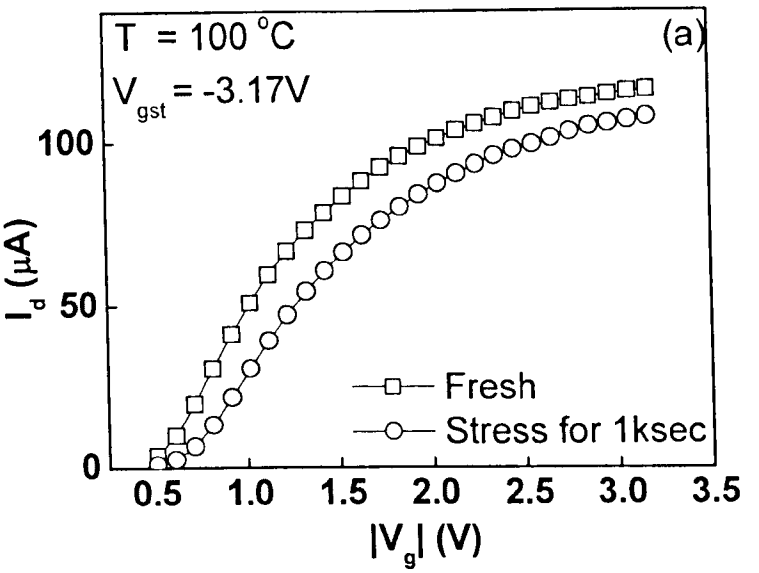


(a)

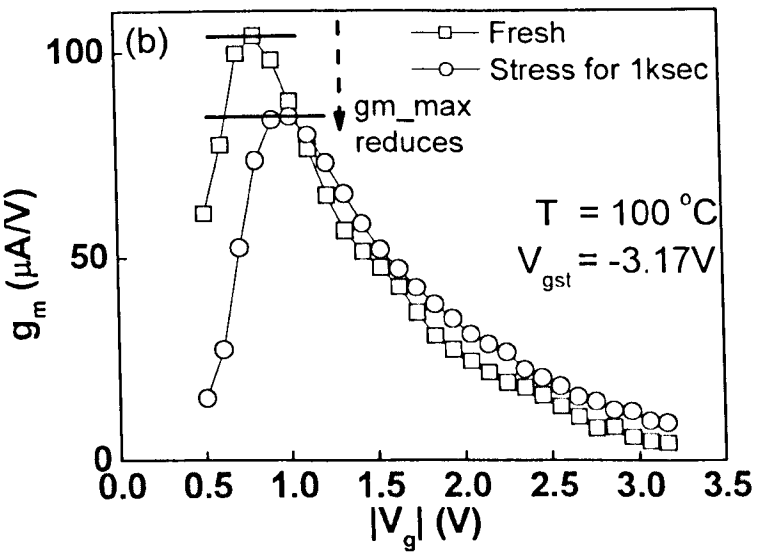


(b)

Fig 3.20 (a) Variation of $\Delta I_d/I_{do}$ and $\Delta V_t/(V_g-V_{to})$ during stress. In contrast with the relatively small $\Delta I_d/I_{do}$ (<10%), $\Delta V_t/(V_g-V_{to})$ can be over 30%. (b) Variation of the effective mobility during stress. Against popular expectation of mobility degradation, the effective mobility at the stress gate bias actually increases with stress time.



(a)



(b)

Fig 3.21(a) Pulse I_d-V_g measurement was performed before and after stress at -3.17V at 100 °C for 1000 s. I_d-V_g curve measured after stress shifts negatively, indicating the generation of positive charges. (b) from the I_d-V_g curves, the gm before and after stress are obtained. The maximum transconductance, gm_max, reduced after NBTI stress, as illustrated by the dashed arrow. After the stress, the g_m-V_g is shifted toward higher $|V_g|$. The gm at a given $|V_g|$=3.17V actually increased.

3.7 Bridging $|\Delta V_t|$ evaluated by different techniques

An attempt is now made to bridge the gap in $|\Delta V_t|$ evaluated by the different techniques. Figure 3.22 compares $|\Delta V_t|$ evaluated by various techniques. First of all, after replacing g_m by G_m, the $|\Delta V_t|$ by OTF-V_{gst} agrees well with the $|\Delta V_t|$ obtained from UFP-V_g when the sensing V_g is at V_{gst}, although quasi-DC modulation was used by OTF-V_{gst}. This indicates that the recovery is insignificant even during the quasi-DC measurement, if the modulation from V_{gst} is small (DV= 25 mV). Secondly, the $|\Delta V_t|$ obtained by UFP-ex also agrees with the $|\Delta V_t|$ from the UFP-V_g when the sensing $|V_g|$ is reduced to close to $|V_t|$. As a result, the difference in $|\Delta V_t|$ evaluated by OTF-V_{gst} and UFP-ex is caused by the difference in sensing V_g. Finally, the on-the-fly and quasi-DC $|\Delta V_t|$ are compared. The symbol '□' in figure 3.16 was obtained by ramping down $|V_g|$ from $|V_{gst}|$ in a quasi-DC speed and the total measurement time was 7.65 s. The difference between $|\Delta V_t|$ by the UFP-V_g and the quasi-DC results is small when sensing V_g is near to V_{gst}, since recovery is limited here. As $|V_g|$ decreases and measurement time increases, the recovery becomes larger, resulting in lower quasi-DC $|\Delta V_t|$ when compared with the $|\Delta V_t|$ by UFP-V_g. When the sensing V_g is close to V_t, the quasi-DC $|\Delta V_t|$ evaluated through G_m agrees with that by extrapolating the quasi-DC I_d~V_g, $|\Delta V_t(\text{DC-ex})|$, further endorsing the new evaluation technique.

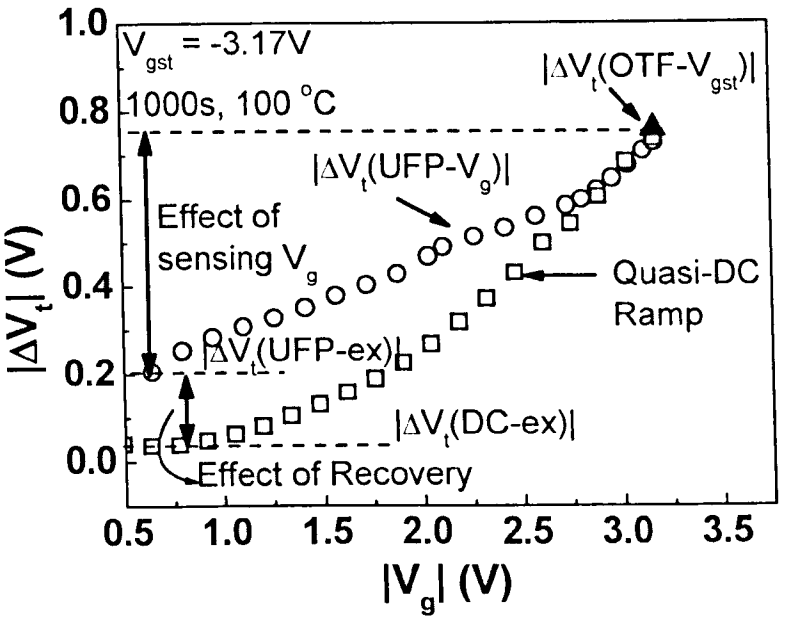


Fig 3.22 Bridging the gap in ΔV_t evaluated by different techniques. The symbol '▲', $|\Delta V_t$ (OTF-$V_{gst})|$ was obtained by the OTF-V_{gst} technique with V_{gst} modulated by ±25 mV at quasi-DC speed. The symbol 'o' and '□' was obtained by using the new UFP-V_g method, where $|V_g|$ was ramped down from $|V_{gst}|$ to 0.5 V in 5 μs (pulse) and 7.65 s (quasi-DC), respectively. The ΔV_t evaluated by extrapolating the pulse and quasi-DC I_d~V_g is represented by the two dashed horizontal lines, labelled as $|\Delta V_t$ (UFP-ex)| and $|\Delta V_t$ (DC-ex)|. By taking both the recovery and sensing V_g effects into account, the new method successfully bridges the gap in $|\Delta V_t$ (OTF-$V_{gst})|$, $|\Delta V_t$ (UFP-ex)| and $|\Delta V_t$ (DC-ex)|.

In summary, the difference between $|\Delta V_t$(DC-ex)| and $|\Delta V_t$(UFP-ex)| is caused by the recovery reported in earlier work. The difference between $|\Delta V_t$(OTF-$V_{gst})|$ and $|\Delta V_t$(UFP-ex)| is caused by the sensing V_g effect. Both the recovery and the V_g effect contribute to the difference between $|\Delta V_t$(DC-ex)| and $|\Delta V_t$(OTF-$V_{gst})|$. In this way, the large difference in $|\Delta V_t|$ evaluated by various techniques is bridged by the new method.

3.8 Possible physical process

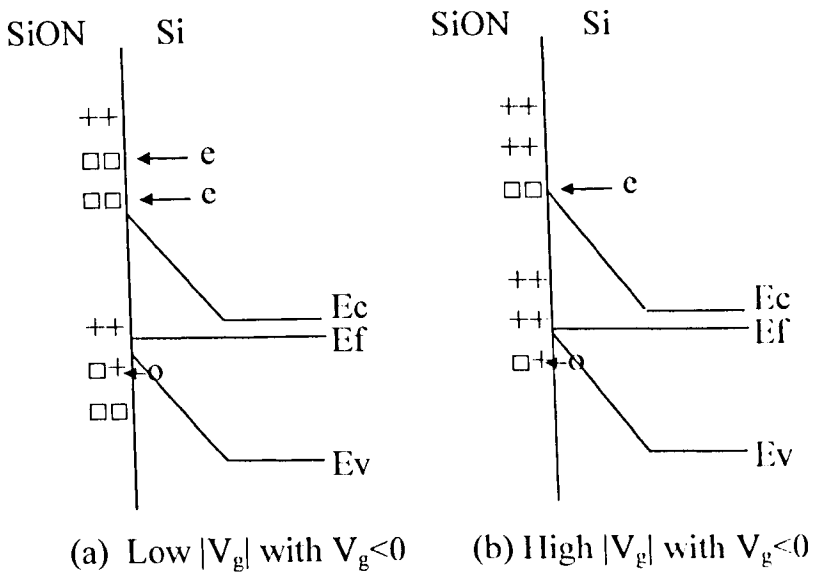
Threshold voltage of pMOSFETs can be defined as [21]

$$V_t = \phi_{ms} + 2\phi_F - \frac{\sqrt{4\varepsilon_{Si} q N_A |\phi_F|}}{C_{ox}} - \frac{X_c Q}{X_{ox} C_{ox}} \qquad (3.9)$$

In some text books [22] a thermal potential term is added to equation (3.5), but it makes no difference to the discussion here. In physical terms, V_t is the minimum V_g needed to switch on a MOSFET and equation (3.1) shows that I_d is proportional to (V_g-V_t). For a degradation-free pMOSFET, Q is fixed and V_t is a constant. During stress, however, Q increases with time, resulting in a shift in V_t. During recovery, Q and $|\Delta V_t|$ decreases.

The phenomenon that $|\Delta V_t|$ increases with $|V_g|$ can be explained if Q is higher at higher $|V_g|$, as schematically illustrated in figure 3.23. The defect responsible for the sensing V_g effect is not known at present. It has been reported that some positive charges can have an energy level above the bottom edge of the silicon conduction band and their neutralization was sensitive to temperature [23-25]. One possibility is that, under low $|V_g|$, free electrons in silicon driven by thermal energy can still bombard the interface and neutralize a portion of them, as illustrated by figure 3.23 (a). A higher negative V_g raises the defect energy level and impedes free electrons from reaching the interface and more positive charges can survive the neutralization shown in figure 3.23 (b). Another possibility is that there are defects close to the top edge of the silicon valence

band. An increase of $|V_g|$ also raises the energy level of these defects with respect to the Fermi level, resulting in more positive charge. Whether they are interface states is not known, but even if they are, they cannot be probed by the conventional technique like charge pumping, since they are close to the band edge [26].



(a) Low $|V_g|$ with $V_g<0$ (b) High $|V_g|$ with $V_g<0$

Fig 3.23 Schematic illustrations of the defects responsible for the sensing V_g effect. The symbol '+' and '□' represents positively charged and neutralized defects, respectively. 'e' and 'o' represents free electrons and holes. (a) shows that when $|V_g|$ is relatively low, some defects above Ec can be neutralized by free electrons driven to the interface by thermal energy. Some defects below Ef can also be neutral. (b) shows that when V_g is more negative, positive charges can increase.

3.9 Difference between the sensing Vg effect and the recovery effect

When $|V_g|$ was ramped down from the stress level, figure 3.22 shows that both the recovery and the effect of sensing $|V_g|$ can lead to a reduction of $|\Delta V_t|$.

Although a loss of charges must have occurred in both cases, there are good reasons for believing that the V_g effect is not the same as the recovery reported in the literature, as discussed below:

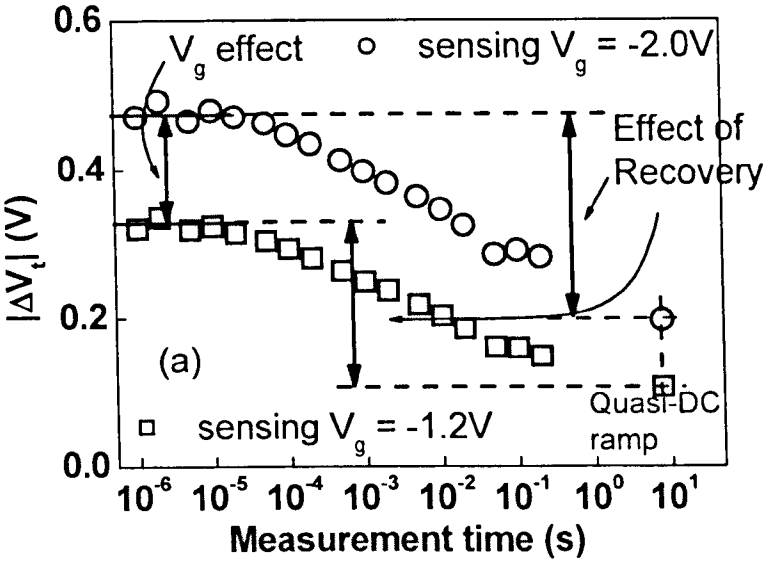
Recovery occurs under the same sensing V_g: Since $|\Delta V_t|$ is conventionally measured with V_g close to V_t, the sensing V_g changes little and it is widely observed that recovery is substantial. To test this further, figure 3.24 (a) shows that recovery is also substantial under other fixed sensing V_g. As a result, the recovery occurs when the sensing V_g effect is eliminated.

Relative recovery is insensitive to the sensing V_g: After normalizing the $|\Delta V_t|$ in figure 3.24 (a), figure 3.24 (b) shows that the relative recovery under different sensing V_g agrees well. As a result, although the sensing $|V_g|$ affects $|\Delta V_t|$, it has little effect on the relative recovery.

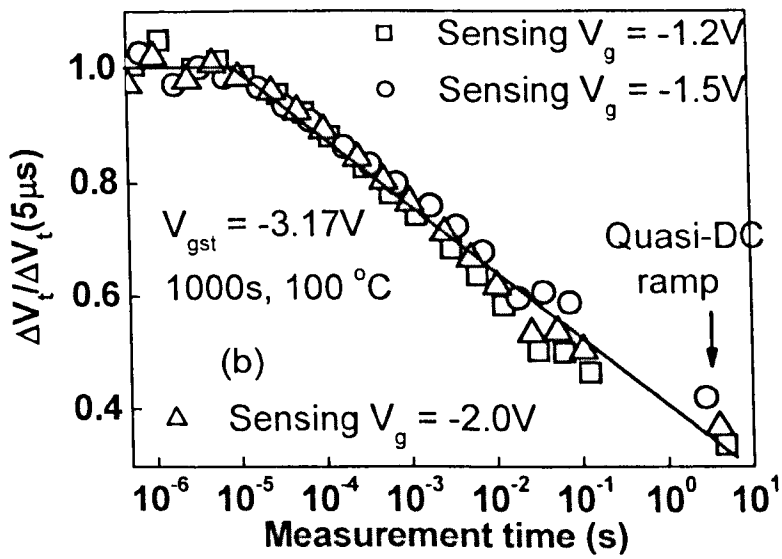
V_g effect is substantial when recovery is minimized: By using pulse I_d-V_g technique with 5 μs measurement time, the recovery can be minimized to within the measurement resolution. Substantial V_g effect, however, was observed at the same short time in figure 3.24 (a). This supports that they are different phenomena.

When recovery is essentially over, V_g effect remains: Although it is difficult, if not impossible, for the recovery rate to drop to zero, it reduces exponentially as time

increases, as shown in figure 3.24 (b). To study what happens to the V_g effect after recovery is essentially over, one device was allowed to recover for 1000 s at 100 °C, as shown in figure 3.25 (a). The pulsed I_d~V_g was measured both before and after the recovery to evaluate the V_g effect and figure 3.25 (b) shows that the relative V_g effect remains substantial after the recovery.

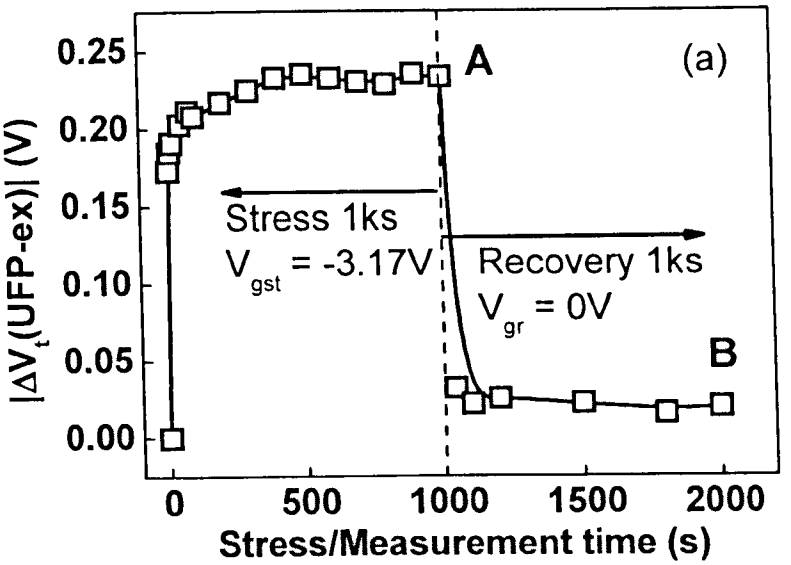


(a)

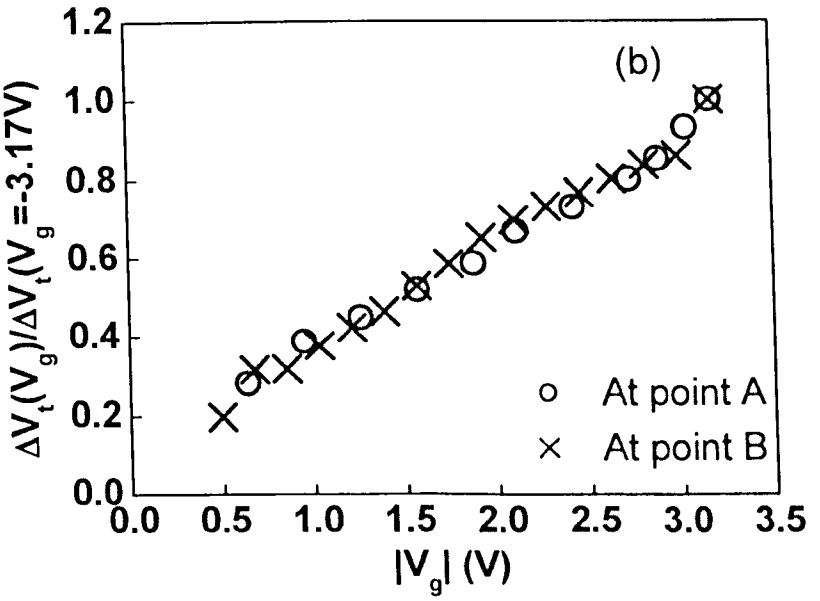


(b)

Fig 3.24 Difference between the conventional recovery and the sensing V_g effect. (a) shows that the V_g effect exists when the recovery is negligible at short measurement time. (b) shows that the normalized recovery is independent of the sensing V_g.



(a)



(b)

Fig 3.25 A comparison of the sensing V_g effect before and after recovery. In (a), the point 'A' is before recovery and further recovery beyond the point 'B' is insignificant. The solid lines are guides-to-the-eye. The pulse (5 μs) I_d~V_g was taken at both 'A' and 'B' and the normalized V_g effect is compared in (b). The V_g effect is substantial both before and after the recovery.

3.10 Advantage of $|\Delta V_t|$ over $\Delta I_{dsat}/I_{dsat0}$ as a measure for NBTI

Apart from ΔV_t, $\Delta I_{dsat}/I_{dsat0}$ is also used to monitor NBTI. Compared with $\Delta I_{dsat}/I_{dsat0}$, ΔV_t has two advantages.

Firstly, ΔV_t allows the estimation of the effective charge density, ΔN, from $\Delta N=\Delta V_t \times C_{ox}/q$, which cannot be directly obtained from $\Delta I_{dsat}/I_{dsat0}$. Secondly and more importantly, ΔV_t is independent of the source and drain series resistance, R_{sd}, but $\Delta I_{dsat}/I_{dsat0}$ depends on it. To provide experimental evidence, a resistor, R_{ext}, was externally connected to the drain. Figure 3.26 confirms that $\Delta Id_{sat}/Id_{sat0}$ measured at $V_g=V_d=-1.2$ V drops for higher R_{ext}, while ΔV_t remains a constant.

In a real circuit, R_{sd} depends on the size and layout and it can change substantially for different devices [27]. As a result, the $\Delta I_{dsat}/I_{dsat0}$ obtained from the test in a laboratory may not be directly used to predict the circuit performance, where the device size and layout change.

Moreover, V_t is a key parameter for the BSIM-based circuit simulators [28,29]. As a result, ΔV_t is indispensable.

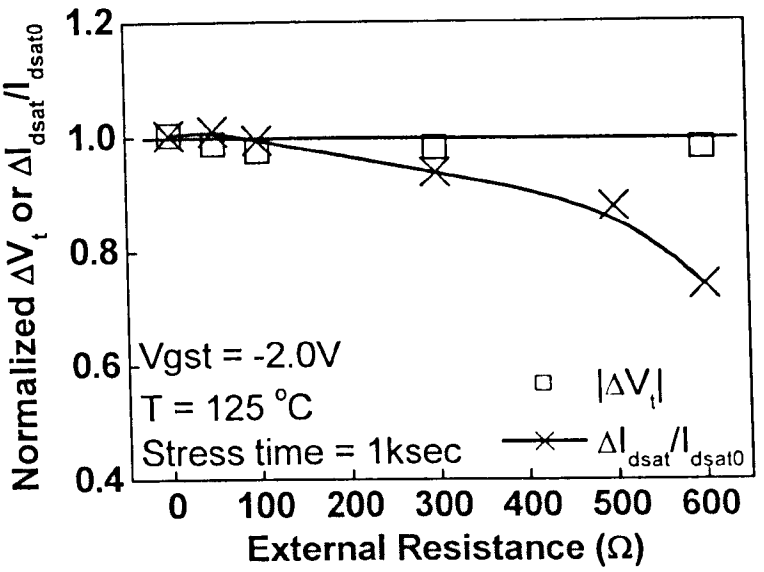


Fig 3.26 Dependence of $\Delta I_{dsat}/I_{dsat0}$ and $|\Delta V_t|$ on the series resistance. The 'external resistance', R_{ext}, is a resistor externally connected to the drain of the devices under test. Both $\Delta Id_{sat}/I_{dsat0}$ and $|\Delta V_t|$ were normalized to their value at R_{ext}=0.

3.11 Summary

In this chapter, the NBTI-induced threshold voltage shift evaluated by different techniques is compared, including on-the-fly (OTF-V_{gst}) sensed at stress V_{gst}, ultra-fast pulse (UFP-ex) technique with extrapolation, and the conventional quasi-DC measurement with extrapolation. Attention was paid to the impact of various factors reported recently on ΔV_t. This includes the degradation during measurement that affects the reference I_d and initial transconductance. The sample quality and errors during both measurement and evaluation were addressed. After clarifying these issues, it is found that recovery alone cannot explain the difference in ΔV_t extracted from different techniques and it is essential to consider the effect of sensing V_g on ΔV_t. The evaluation of ΔV_t at

different sensing V_g was achieved by calculating the ΔI_d and transconductance at the same V_g and the classical $g_m=dI_d/dV_g$ should be replaced by $G_m=dI_d/d(V_g-V_t)$. The results clearly show that $|\Delta V_t|$ increases with sensing $|V_g|$ because of more positive charge at higher $|V_g|$ and the popular assumption of ΔV_t being independent of sensing V_g is shown to be invalid. After taking both recovery and the sensing V_g effect into account, the large gap in ΔV_t obtained from different techniques in earlier work has been successfully bridged. The results show that recovery reported by earlier work and the sensing V_g effect observed here are two different phenomena.

Test engineers should take into account the increase in $|\Delta V_t|$ with sensing $|V_g|$ when assessing the impact of NBTI on circuit performance. In a worst case scenario, $|\Delta V_t|$ should be sensed at the highest possible oxide field during operation without recovery by using the UFP-V_g technique and care must be exercised to ensure the reference I_d degradation at this high sensing $|V_g|$ is insignificant. In reality, different devices in a circuit can have different recovery dynamics and the oxide field can also change with time and along the channel, resulting in smaller $|\Delta V_t|$.

References

1. J. S. Suehle, "Ultrathin gate oxide reliability: physical models, statistics, and characterization," *IEEE Trans. Electron Dev.*, vol. 49, pp. 958-971, 2002.
2. W. D. Zhang, J. F. Zhang, C. Z. Zhao, M. H. Chang, G. Groseneken, and R. Degraeve, "Electrical signature of the defect associated with gate oxide breakdown," *IEEE Electron Dev. Lett.*, vol. 27, pp. 393-395, 2006.
3. 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.
4. J. F. Zhang, M. H. Chang and G. Groeseneken, "Effects of measurement temperature on NBTI," *IEEE Electron Dev. Lett.*, vol. 28, pp. 298-300, 2007.
5. M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. R. Tauriac and N. Revil, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's," in *International Electron Devices Meeting (IEDM) Digest*, pp.109-112, 2004.
6. 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.
7. 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 *Symposia on VLSI Technology and Circuits*, pp. 92-93, 2005.
8. J. F. Zhang, Z. Ji, M. H. Chang, B. Kaczer, and G. Groeseneken, "Real Vth instability of pMOSFETs under practical operation conditions," in *International Electron Devices Meeting (IEDM) Digest*, pp.817-820, 2007.
9. C. H. Liu, M. T. Lee, C.-Y. Lin, J. Chen, K. Schruefer, J. Brighten, N. Rovedo, T. B. Hook, M. V. Khare, S.-F. Huang, C. Wann, T. C. Chen, and T. H. Ning., "Mechanism and process dependence of negative bias temperature instability (NBTI) for pMOSFETs with ultrathin gate dielectrics," in *International Electron Devices Meeting (IEDM) Digest*, pp. 861–864, 2001.
10. Y. Mitani, M. Nagamine, H. Satake and A. Toriumi, "NBTI mechanism in ultra-thin gate dielectric: Nitrogen-originated mechanism in SiON," in *International Electron Devices Meeting (IEDM) Digest*, pp. 509–513, 2002.
11. S. Rangan, N. Mielke, and E. C. C. Yeh, "The universal recovery behaviour of negative bias temperature instability," in *International Electron Devices Meeting (IEDM) Digest*, pp. 341–344, 2003.
12. C. Shen, M. -F. Li, X. P. Wang, Y. –C. Yeo, and D. -L. Kwong, "One fast measurement technique of MOSFET Id-Vg characteristics," *IEEE Electron Dev. Lett.*, vol. 27, pp. 55-57, 2006.
13. C. Shen, M. -F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra, and Y.-C. Yeo, "Characterization and physical origin of fast Vth transient in

NBTI of pMOSFETs with SiON dielectric," in *International Electron Devices Meeting (IEDM) Digest*, pp. 333-336, 2006.

14. N. Kimizuka, K. Yamaguchi, K. Imai, T. Iizuka, C. T. Liu, R. C. Keller and T. Horiuchi, "NBTI enhancement by nitrogen incorporation into ultrathin gate oxide for 0.10-μm gate CMOS generation," in *Symposia on VLSI Technology and Circuits*, pp. 92–93, 2000.
15. E. N. Kumar, V. D. Maheta, S. Purawat, A. E. Islam, 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 *International Electron Devices Meeting (IEDM) Digest*, pp. 809–812, 2007.
16. H. Reisinger, U. Brunner, W. Heinrigs, W. Gustin and C. Schlunder, "A comparison of fast methods for measuring NBTI degradation," *IEEE Trans. Device Mater. Rel.*, vol. 7, pp. 531-539, 2007.
17. 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.
18. *Operation and modelling of the MOS transistor,* Y. Tsividis, WCB /McGraw-Hill, 1999.
19. D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol.94, pp.1-17, 2003.
20. V.Huard, M.Denais and C.Parthasarathy, "NBTI degradation: From physical mechanisms to modeling," *Mircoelectronics Reliability*, vol. 46, pp. 2101, 2006.
21. *Fundamentals of Modern VLSI devices*, T. Yuan and T. H. Ning, Cambridge University Press, 1998.
22. *MOSFET modeling for VLSI simulation: Theory and Practice,* N. Arora, World Scientific Press, 2006.
23. 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.
24. 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.
25. 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.
26. 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.
27. A. Dixit, A. Kottantharayil, N. Collaert, M. Goodwin, M. Jurczak and K. D. Meyer, "Analysis of the parasitic S/D resistance in multiple-gate FETs," *IEEE Trans. Electron Dev.*, vol. 52, pp. 1132-1140, 2005.

28. *BSIM3v3 manual*, Y. H. Cheng, M. Chan, K. Hui, M. C. Jeng, Z. H. Liu, J. H. Huang, K. Chen, J. Chen, R. Tu, K. K. Ping and C. Hu, University of California, Berkeley.
29. N. K. Jha, P. S. Reddy, D. K. Sharma, and V. R. Rao, "NBTI degradation and its impact for analog circuit reliability," *IEEE Trans. Electron Dev.*, vol. 52, pp. 1132-1140, 2005.

4 NBTI Lifetime prediction and kinetics at operation bias by pulse I_d-V_g technique

4.1 Introduction

Negative bias temperature instability (NBTI) is nowadays the most critical device degradation mechanism and becomes a limiting factor in the scaling of modern CMOS technologies [1-12]. This leads to a strong demand for accurate lifetime prediction. The NBTI lifetime is typically defined as the time for the threshold voltage shift, ΔV_t, to reach a preset level [6-11]. Under an operational 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 $|V_{gst}|^{-\alpha}$ [6,7] or $\exp(-|V_{gst}|)$ [8,9] and then extrapolated to V_{gop}, so that the lifetime under V_{gop} can be estimated. Recent work [2-5], however, has raised several questions on the validation of the above prediction technique, as described below.

First of all, the above prediction technique was developed based on quasi-DC measurements with a measurement time for a transfer characteristics, I_d-V_g, up to seconds [7-9, 12]. Recent work [2-5] has shown that substantial recovery occurs during the measurement and to suppress the recovery, measurement time must be reduced to the order of tens of microseconds by using the ultra-fast pulse (UFP) technique [2,4,5,10]. In a real circuit, different transistors will experience different levels of recovery and the worst-case scenario will be no recovery. It is not known whether the V_g acceleration technique described above is still applicable in this case.

Secondly, the threshold voltage is typically evaluated by extrapolating the transfer characteristic, I_d-V_g, and the sensing V_g used here is close to V_t. In a practical circuit, the operation bias is higher than V_t and an implicit assumption of earlier work is that ΔV_t is insensitive to the sensing V_g. In chapter 3, however, it was shown that ΔV_t increases with sensing V_g and its value at V_{gop} can double the level extracted by extrapolating I_d-V_g. There is no information on how the lifetime can be predicted by including this dependence on sensing V_g.

In this chapter, lifetime prediction will be based on results measured by the pulse I_d-V_g technique, and the degradation will be evaluated at the operational bias level. Therefore recovery will be suppressed during the measurement and the worst-case lifetime will be estimated. For the first time, current acceleration methods, which are normally used for predicting NBTI lifetime, will be evaluated

based on all measured data. Substantial errors from these acceleration methods will be demonstrated and causes will be explored. To predict the worst-case lifetime, a model for NBTI kinetics under operational gate bias is developed which includes contributions from both as-grown and generated defects. Based on the new kinetics, a single test prediction method is proposed. An effort will be made to estimate the safety margin of this new lifetime prediction method.

4.2 Samples

The MOSFETs used in this study were manufactured at Interuniversity Microelectronics Research Centre (IMEC), Belgium. In order to test the applicability of the proposed NBTI kinetics and lifetime prediction method, samples from six different processes were used, 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 nitrided for different time and one was thermally nitrided. One HfSiON/SiON stack was also prepared by ALCVD with 80% Hf and a TiN gate.

Table 4.1 Samples used in the experiments

Wafter No.	Gate material	Gate dielectrics	EOT (nm)
A	p^+ poly-Si	12 s 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	45 s plasma nitrided SiON	2.0
E	p^+ poly-Si	20 s plasma nitrided SiON	2.0
F	TiN	2 nm ALCVD HfSiON + 1 nm SiON	1.53

4.3 Measurement problems and solutions

The typical lifetime prediction is based on so-called measure-stress-measure (MSM) procedures. The pMOSFETs are stressed under constant V_{gst} at 125 °C for a pre-specified time and the degradation then monitored by interrupting the stress and measuring I_d-V_g by using a standard semiconductor parametric analyzer [2,12,13]. The degradation is normally evaluated by using the extrapolation method. Therefore, the predicted lifetime can be over-estimated by the recovery during the measurement and the ΔV_t used for prediction not evaluated at the real operational bias as discussed in chapter 3.

In order to study NBTI lifetime prediction with recovery suppressed, the UFP I_d-V_g technique will be used. In chapter 2, it was shown that the recovery during measurement can be effectively suppressed to within the measurement resolution when the pulse edge time was reduced to the order of tens of microseconds. In this chapter, 5 μs is used.

In order to predict lifetime at operational conditions, the sensing V_g should be at operational level. The sensing V_g at operational level will be referred to as V_{gop}, hereafter and ΔV_t(DC-ex) will be replaced by ΔV_t(UFP-V_g).

In this chapter, devices are stressed at $|V_{gop}|$=1.2 V, which is lower than the typical stress bias used in earlier NBTI tests [2-5,10]. For low level degradation,

noise must be minimized. The typical commercial parameter analyzers minimize noise by repeating the I_d measurement for a given bias many times to produce an average value as illustrated in figure 4.1 (a) [14]. For a stress time less than 1s, the pulse measurement was repeated 30 times and an average value was used, as shown in figure 4.1 (b). For stress longer than 1s, a quasi-DC parameter analyzer that automatically uses average values was used.

In order to avoid device to device variation, the repeated measurement must be performed on the same sample. Within 1 s of stress time, the device was stressed for a pre-set time, and then measured within 5μs. After that, it was left with 0 V applied on the gate for 2 minutes to recover to its fresh state. In order to justify that recovery is completed, figure 4.1 (a) gives the I_d measured repeatedly for 30 times on a device stressed for 1 s and recovery for 2 minutes. The recovery is complete and degradation was not observed. A measurement resolution of ±0.23% was achieved. As to be shown later in this chapter, the measurement accuracy achieved in this way is good enough to establish a kinetic model that allows lifetime to be predicted with a safety margin of 50%.

Figure 4.1 (b) also shows that $|\Delta V_t|$ becomes higher than 10 mV in less than 1 ms under V_{gop} = -1.2 V. This rapid increase of $|\Delta V_t|$ does not mean that the quality of the samples used here is poor. This difference is caused partially by suppressing recovery with the pulse measurement and partially by using $|V_{gop}|$=1.2 V as the sensing bias, rather than the conventional extrapolation to

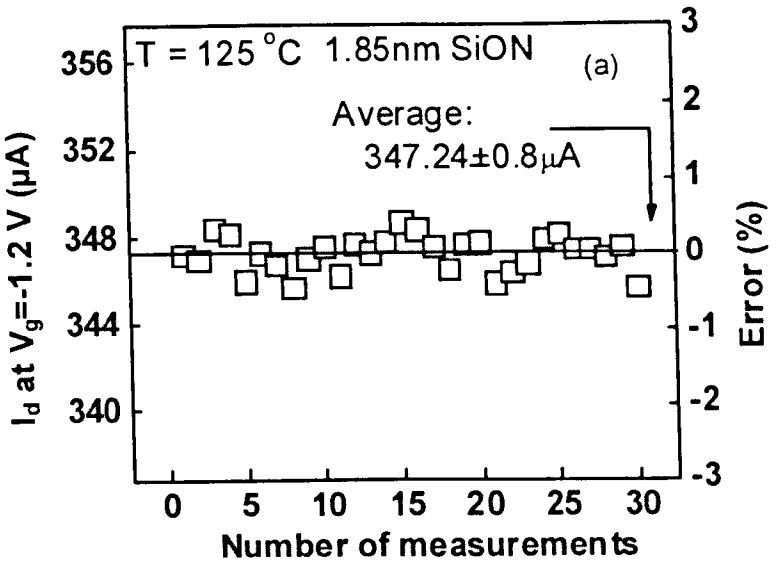
$|V_t|$~0.4 V, as shown in chapter 3. Therefore, the degradation can be easily underestimated if recovery is not suppressed or improper sensing V_g is used and this can have a very significant impact on lifetime prediction.

4.4 Theories for the NBTI dynamics

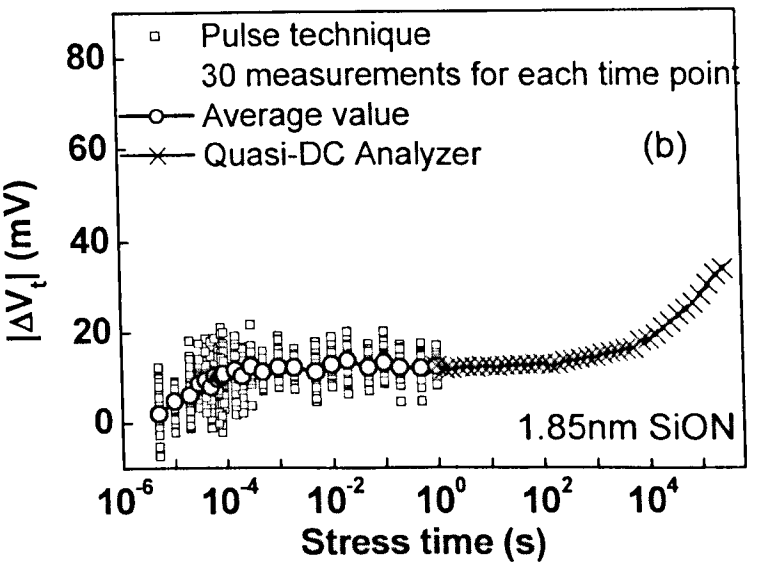
In this section, two theories for the NBTI degradation, namely, the reaction-diffusion model (R-D model) and the charge-trapping model are outlined.

4.4.1 Reaction-Diffusion model

Since reaction-diffusion model was first proposed in 1977 [15], it has been the most successful and most widely accepted model for NBTI. However, it remains controversial as to the exact form of diffusing hydrogen species, the interfacial chemical reaction and the appropriate boundary conditions [13].



(a)



(b)

Fig 4.1 Measurement resolution and minimization of noises. In (a), I_d at V_g=-1.2 V was measured 30 times on a device stressed for 1 s and recovery for 2minutes from the pulsed I_d-V_g with an edge time of 5 μs. In (b), the pulse measurement (symbol '□') was repeated 30 times for stress time less than 1 s and the average value (symbol 'o') was used to minimize the noise. For stress longer than 1 s, we used a quasi-DC parameter analyzer that automatically used the average value (symbol '×').

According to the reaction-diffusion theory, interface trap generation is a result of

the electrochemical reaction at the Si/SiO_2 interface. The exact reaction is unclear, but it is widely accepted that:

- A Si-H bond is broken, hydrogen species are released, and leaving behind a silicon dangling bond (interface trap);
- Holes at the interface catalyze or participate in the reaction;
- The forward reaction is accelerated by stress voltage and temperature.

A hypothetical equation of this interface reaction under electrical stress is given in (1), where the interaction of a hole, h^+, with the Si-H bond results in a donor-like interface trap and a free H that can diffuse away from the interface or passivate a Si dangling bond. The trap density (N_{it}) increases with the net reaction given in (2).

$$\equiv SiH + h^+ \underset{passivation}{\overset{stress}{\rightleftharpoons}} \equiv Si^+ + H^0 \tag{4.1}$$

$$\frac{dN_{it}}{dt} = k_F (N_0 - N_{it}) - k_R \cdot N_H^{(0)} \cdot N_{it} \tag{4.2}$$

At the beginning of the stress, both N_{it} and $N_H^{(0)}$ are negligible ($N_0 >> N_{it}$), and so is the second term in the right side of (2). Therefore, the increase in N_{it} is generation-limited, as shown in region (i) ($N_{it} \sim t^1$) of figure 4.2. The reaction later reaches a dynamic equilibrium state in region (ii) ($N_{it} \sim t^0$) when the annealing term of (2) becomes comparable to the generation term. After sufficient build-up of hydrogen near the semiconductor/oxide interface, the diffusion of H into the oxide begins to dominate, so the net reaction rate becomes limited by the hydrogen diffusion in region (iii) ($N_{it} \sim t^n$), and gives the characteristic time

evolution of the NBTI degradation. Eventually, in region (iv), all the Si-H bonds are broken ($N_{it} = N_0$), so that the trap density saturates [16].

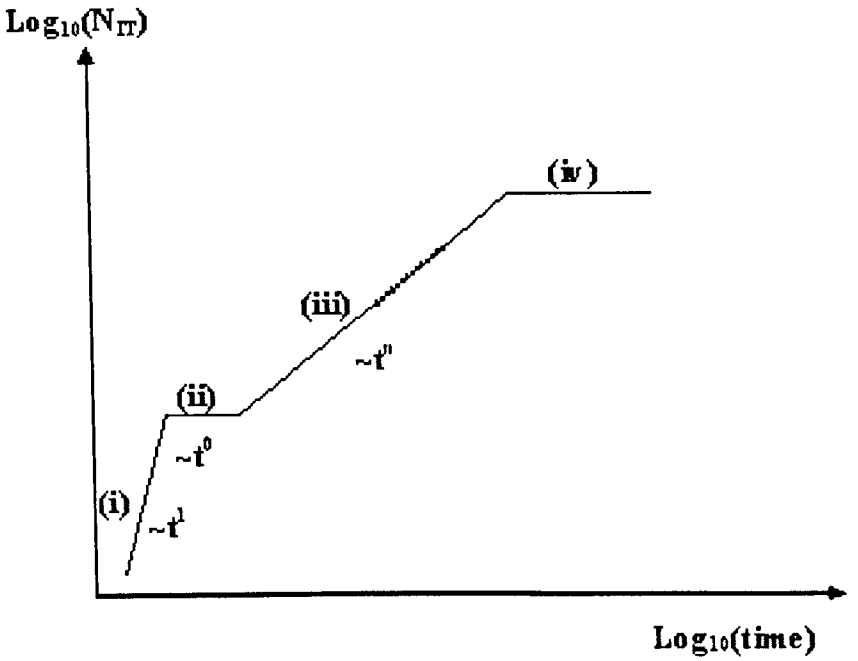


Fig 4.2 Schematic time evolution of the classic R-D model. Regions: (i) generation-limited, (ii) dynamic equilibrium, (iii) diffusion-limited, and (iv) saturation. Region (iii) gives the power-law behaviour for NBTI.

The reaction-limited process and the following dynamic equilibrium is considered to be very fast and has never been monitored by measurement. Therefore, the measured results are normally explained by using the third stage which is diffusion-limited. An asymptotic solution was derived by Jeppson to show that N_{it} grows with time as $N_{it} \sim t^{0.25}$ in this diffusion-limited situation [15]. Alam, by assuming a triangular hydrogen concentration profile, arrived at the same result [17]. As shown in figure 4.3, the hydrogen concentration away from the interface ($x = 0$) is approximated with a simple linear function, with characteristic diffusion length $\sqrt{Dt}$. The hydrogen concentration at the interface is $N_H^{(0)}$. Since each broken Si-H bond releases one atomic hydrogen, the number of interface trap

equals the number of hydrogen atoms in diffusion.

$$\begin{aligned} N_{it}(t) &= \int_0^{\infty} N_H(x,t)dx \\ &= \int_0^{\sqrt{Dt}} N_H(0,t)(1-\frac{x}{\sqrt{Dt}}) \\ &= \frac{1}{2} N_H(0,t)\sqrt{Dt} \end{aligned} \tag{4.3}$$

Since the reaction rate is much faster than the diffusion rate, the system is assumed to be close to equilibrium even in the presence of slow diffusion. Therefore, the $dN_{it}/dt \approx 0$ and equation (4.2) roughly equals

$$(\frac{k_F N_0}{k_R}) \approx N_H(0) \cdot N_{it} \tag{4.4}$$

Combining equations (4.3) and (4.4), gives

$$N_{it} = \sqrt{\frac{k_F}{2k_R} N_0 (Dt)^{0.25}} \tag{4.5}$$

A power-law dependence with an exponent of 0.25 is identified.

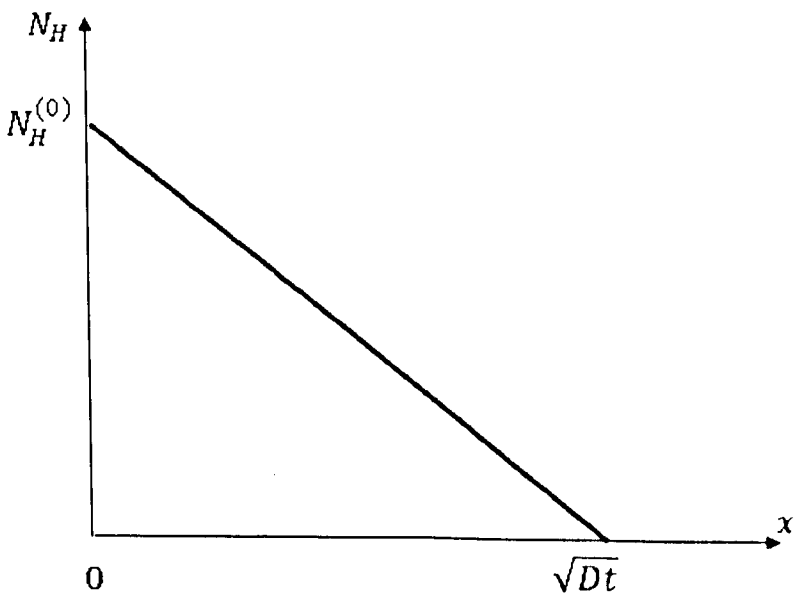


Fig 4.3 Approximate concentration of H^0 in the diffusion process.

Apart from H^0, other diffusion species are proposed. Chakravarthi et al

enumerated a possible signature of neutral H_2 diffusion with n = 1/6 [18]. Ogawa and Shiono demonstrated that H^+ diffuses with n = 1/2 [19]. The mathematical derivation is similar to that presented above and is not repeated here.

4.4.2 Charge Trapping/De-trapping model

As the gate dielectric is scaled down into the direct tunneling regime, a significant content of nitrogen is always added to the SiO_2 dielectric. Since the silicon-nitrogen bond is weaker than the silicon-oxygen bond, it is well known that nitrogen incorporation leads to higher NBTI. Carriers can be trapped in the insulator as they transport through it [20-23]. Trapped charge in the gate dielectric of a MOSFET will shift the threshold voltage V_t. Positive charge trapping (hole trapping) would increase the $|V_t|$ of a pMOSFET.

Ushio et al first proposed that hole-trapping near the Si/SiON interface caused enhanced NBTI in SiON gate dielectric [24]. After recovery is suppressed in recent work, more groups began to believe in hole trapping. For example, Huard and Denais used their OTF technique to demonstrate that NBTI degradation can be entirely explained by hole trapping [3].

Tan et al tried to reconcile the hole trapping with the prevailing R-D model and suggested that hydrogen released from broken Si-H bonds can be trapped at nitrogen sites near the interface to create fixed charge [25-30]. The role of hole

trapping in NBTI has been one of the most active research topics recently.

4.5 Traditional acceleration method for lifetime prediction

Accelerated stress testing is widely accepted in the semiconductor industry. It allows the projection of long lifetime (10 years or more) in a relative short test time. After the accelerated stress, NBTI lifetime estimated at stress condition has to be extrapolated to nominal gate voltage [31]. The extrapolation procedure is based on fitting experimental data points $\tau(Vg)$ with an analytical model, and then using it to estimate the lifetime for operational bias. Unfortunately, there is no analytical model or theory for the dependence of NBTI lifetime on stress gate voltage or electric field in the gate oxide [1,32] . As a result, the empirical or phenomenological models are used.

Two of the most frequently used models for extrapolation along the voltage axis are the exponential [8,9] and power laws [6,7]. The lifetime predicted by them can be considerably different. In the following, these two models will be outlined first and their applicability is then addressed.

4.5.1 Exponential law

Extrapolation by using an exponential law was derived from the study of hot carrier injection experiments as early as 1983 [33]. The model depends on two

observations from experimental results. One is the relationship between degradation and the stress time,

$$\Delta V_t = A \cdot t^n \tag{4.6}$$

Another is the relationship between the magnitude of degradation and V_d, the voltage applied on the drain,

$$A \propto \exp(-\alpha / V_d) \tag{4.7}$$

Using (6) and (7), the lifetime τ, of MOS devices under a certain criterion can be expressed as,

$$\Delta V_t(\tau) = A \cdot \tau^n \tag{4.8}$$

$$\tau = [\frac{1}{A} \Delta V_t(\tau)]^{-1/n} \propto A^{-1/n} \tag{4.9}$$

Since α and n is constant for different V_d, lifetime τ and $1/V_d$ show an exponential relationship.

When the stress mode changes from HCI to NBTI, the same exponential relationship is used. This relationship is used because it can fit the measured data [8, 9, 34-37].

4.5.2 Power law

For the thinner oxides, another empirical model was proposed [38], namely.

$$\Delta V_t = B \cdot |V_{gs}|^m t^n \tag{4.10}$$

The lifetime τ of MOS devices under a certain criterion can be expressed as

$$\log(\tau) = \frac{1}{n} \cdot [\log(\frac{\Delta V_t(\tau)}{B}) - m \cdot \log(|V_{gst}|)] \quad (4.11)$$

It indicates a power law relationship between the lifetime and V_{gst} [39]. This power law relationship became popular after 2004 [6,7,38,40].

4.5.3 Applicability of the acceleration technique

The applicability of the two acceleration techniques will be examined first for the data measured from quasi-DC I_d-V_g by extrapolation, ΔV_t(DC-ex) and then for the data extracted from ultra-fast pulse I_d-V_g at V_{gop}, ΔV_t(UFP-V_{gop}). Figure 4.4 a-c show the common procedure for prediction of lifetime based on ΔV_t(DC-ex): multiple V_g accelerated tests were carried out and the lifetime, τ, at V_{gop} was estimated by an extrapolation against V_{gst}. V_g acceleration is used to allow reliable NBTI kinetics to be measured within a practical length of stress time. The safety margin of the prediction, however, is generally not known. To assess the prediction of a safety margin, it is essential to be able to directly measure ΔV_t at V_{gst}=V_{gop}. This allows comparison of the measured stress time for ΔV_t 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.

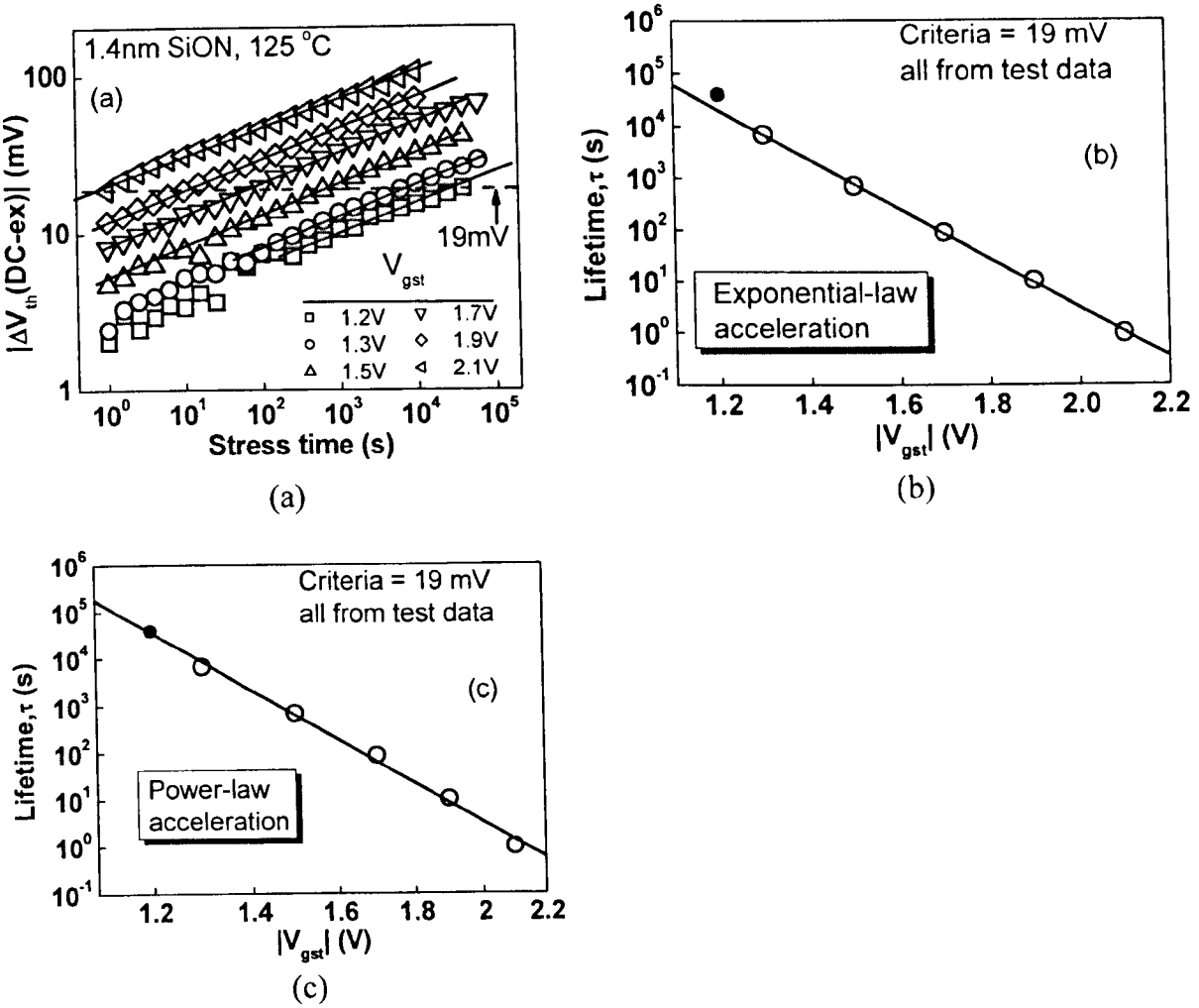


Fig 4.4 Lifetime prediction by V_g acceleration technique for the conventional ΔV_t(DC-ex) measured by extrapolating the quasi-DC I_d-V_g. (a) shows that ΔV_t(DC-ex) follows the power law. The horizontal dashed line represents $|\Delta V_t|$= 19 mV that is used to define the lifetime. (b) shows that the prediction (solid line) by using exponential extrapolation does not agree with the measurement. (c) shows that the prediction (solid line) by power law extrapolation agrees better with the measurement (symbol '•') when compared with the prediction by exponential law extrapolation.

In figure 4.4 (a), the last measured $|\Delta V_t|$ reaches 19 mV under V_{gst}=V_{gop} and using $|\Delta V_t|$=19 mV to define lifetime, the stress time for this last point will be the lifetime under V_{gop}=-1.2 V. This measured lifetime is compared with the prediction based on V_g acceleration in figures 4.4 (b) and (c). Power law $|V_{gst}|^{-\alpha}$ [6,7] and exponential law $\exp(-|V_{gst}|)$ [8,9] are both used. A comparison of figures

4.4 (b) and (c) shows that lifetime projected by the power law agrees better with the measured value. This supports the earlier work [6,7], which reported that the power law is better than the exponential law for predicting lifetime of thin oxides.

In order to assess the uncertainties from recovery and different sensing voltage, measurement by pulse I_d-V_g technique replaced the conventional quasi-DC technique to suppress recovery. Threshold voltage shifts at operational level was obtained by using UPF-V_g technique, as discussed in chapter 2.

As V_{gop} does not reduce proportionally with the SiON layer thickness, the oxide field during device operation increases to such a level that figure 4.5 (a) shows that the ΔV_t at V_{gop}=-1.2 V can now be reliably measured.

In figure 4.5 (a), $|\Delta V_t|$ = 60 mV is used to define lifetime since it is the last value reached under $V_{gst} = V_{gop}$ = -1.2V. The same V_g acceleration methods are used and the measured lifetime is compared with the predicted one. Figure 4.5 (b) and (c) show that there is a substantial difference between the predicted and measured lifetime and neither of the V_g acceleration methods can be used to predict lifetime in this case, therefore.

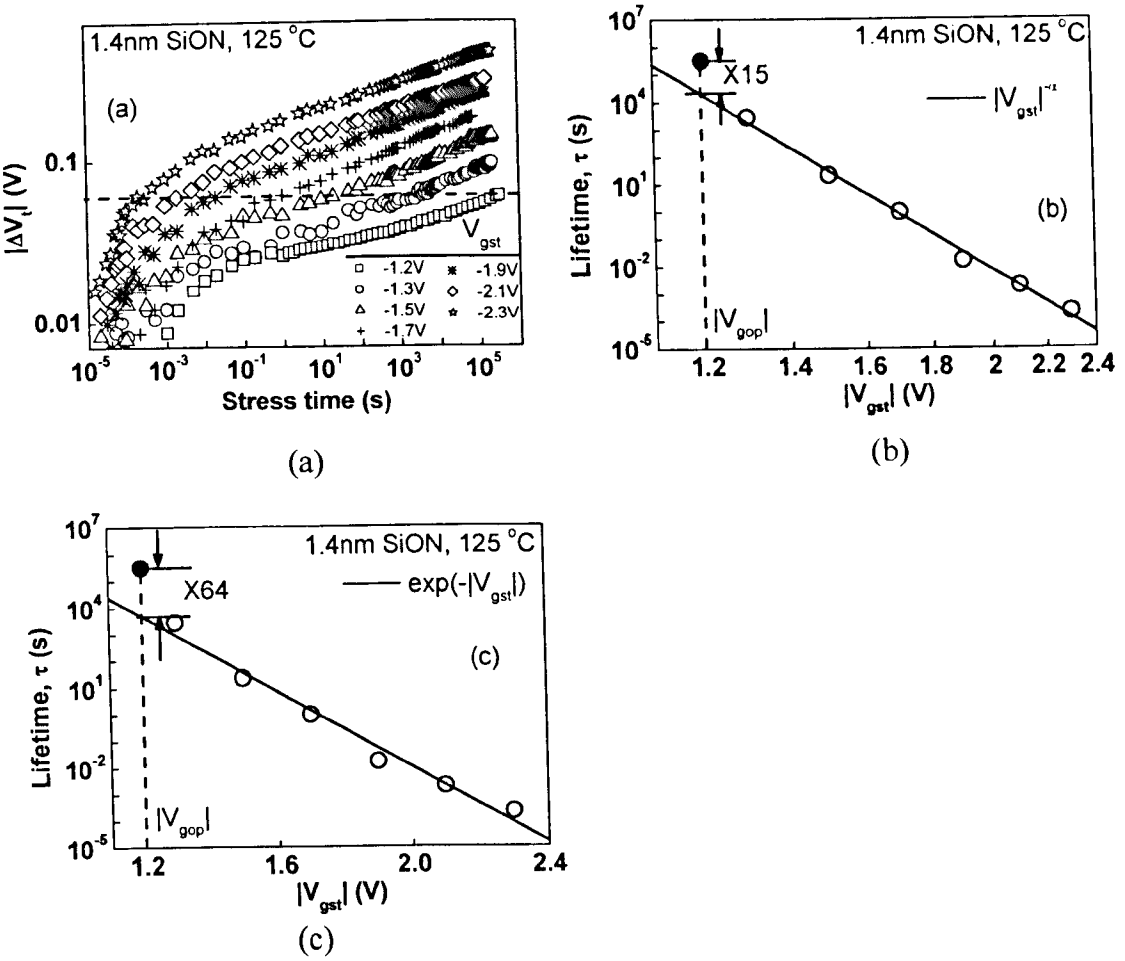


Fig 4.5 Inapplicability of V_g acceleration technique for predicting NBTI lifetime at V_{gop}=-1.2 V measured by UFP technique. In (a), the dynamic behaviour of $|\Delta V_{th}|$ under different stress bias, V_{gst}, is compared and the kinetics does not follow a simple power law. The horizontal dashed line represents $|\Delta V_{th}|$= 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 $|V_{gst}|$=1.2 V. The solid line was obtained by fitting the data at higher $|V_{gst}|$.

To analyze the reason for the inapplicability of the V_g acceleration technique, the $|V_{gst}|^{-\alpha}$ acceleration rule is used as an example. At a device lifetime of $t = \tau$, ΔV_t reaches the specified $\Delta V_t(\tau)$ and the equations (4.10) and (4.11) require the $\log|\Delta V_t| \sim \log(t)$ to be 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_t measured by extrapolating the quasi-DC

I_d-V_g which is shown in figure 4.4 (a), these requirements can be met, as can be seen from figure 4.6. Once the recovery is suppressed and V_{gop}=-1.2 V is used as the sensing V_g, however, ΔV_t in figure 4.5 (a) no longer follows a simple power law and $\log|\Delta V_t| \sim \log(t)$ at different V_{gst} is generally not a parallel shift. A clear example for the non-parallel shift is given in figure 4.7. The V_{gst} effect can no longer be separated into a 'pre-factor' like that in equation (4.10) and this explains the inapplicability of V_g acceleration technique to the case where recovery is suppressed.

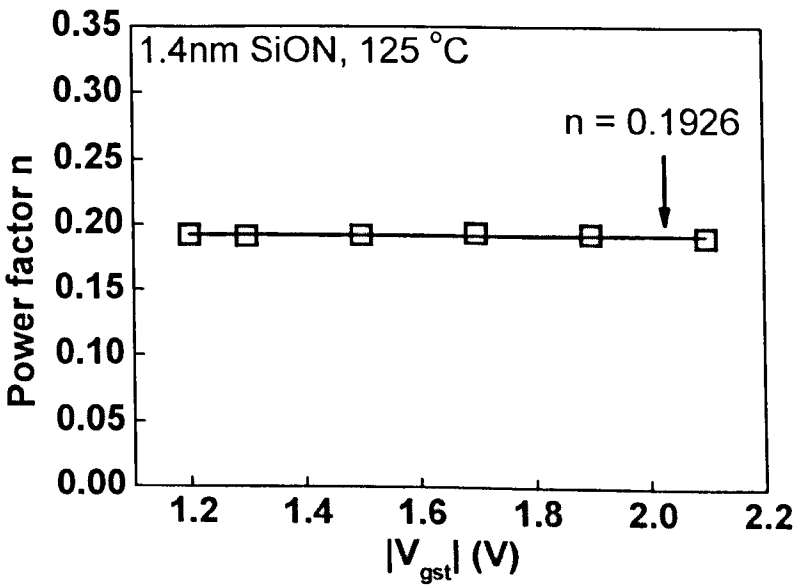


Fig 4.6 The power factor n from DC-ex measurement. It is insensitive to the stress bias. The solid line represents the average power factor of n=0.1926.

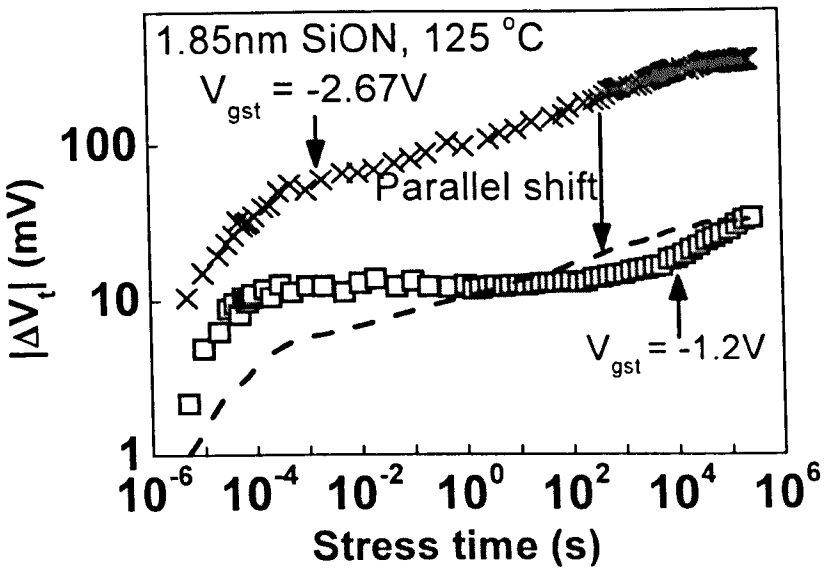


Fig 4.7 The kinetics at different V_{gst} is not shifted in parallel for the UFP ΔV_t sensed at $|V_g|$=1.2 V. The dashed curve was obtained by shifting the symbol '×' downward in parallel.

4.6 New model for NBTI kinetics at operational gate bias

Since ΔV_t does not follow a simple power law against stress time when recovery is suppressed, efforts should be made to develop a model that can describe this dynamic behaviour. For the ΔV_t under $V_{gst}=V_{gop}$=-1.2 V, figure 4.8 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'.

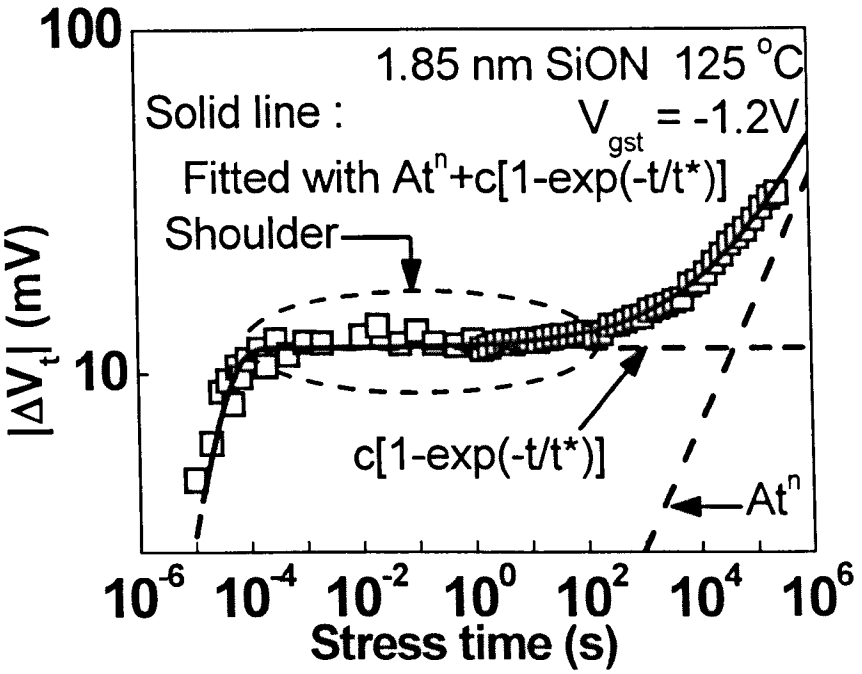


Fig 4.8 The kinetic feature of the UFP ΔV_t sensed at $|V_g|$=1.2 V: a "shoulder". By combining the 1st order model for as-grown hole traps with the power law for defect generation, ΔV_t can be fitted over 10 orders of magnitude in time, as shown by the solid line. The dashed lines show that ΔV_t is dominated by as-grown hole traps initially, but the generated defects become important at longer stress time.

To support the above suggestion, two tests were carried out. In the first test, the effect of temperature on the shoulder height was checked. The saturation level of as-grown defects should be insensitive to temperature [41,42] and if it dominates the shoulder, the shoulder height should be insensitive to temperature. This is confirmed by figure 4.9. Figure 4.9 also shows that the rise above the shoulder is thermally activated, supporting the proposition that defect generation is thermally accelerated.

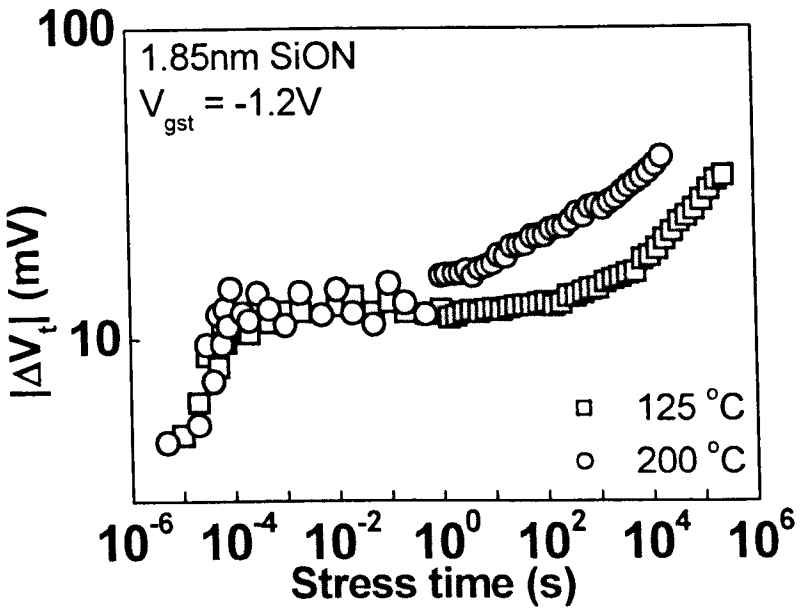
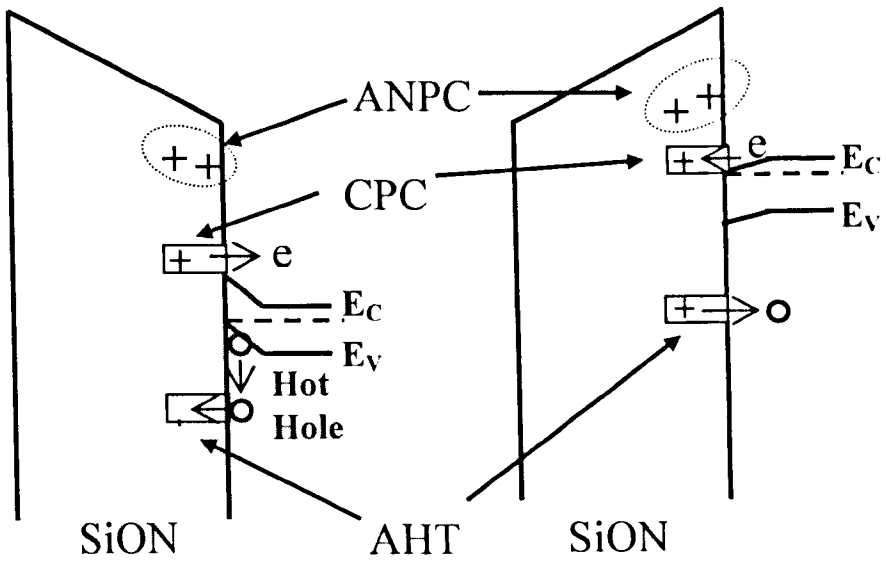


Fig 4.9 Effect of temperature on NBTI kinetics. The height of the shoulder is insensitive to temperature, but the generation of defects above the shoulder is thermally accelerated.

In the second test, the charging and discharging rate of the defect responsible for the shoulder is compared. Earlier work [42-45] 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), as illustrated by figure 4.10. 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 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. Figure 4.11 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. 10 Energy band diagram of different types of positive charges. The anti-neutralization positive charges (ANPC) have energy level above the bottom edge of silicon conduction band, 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.

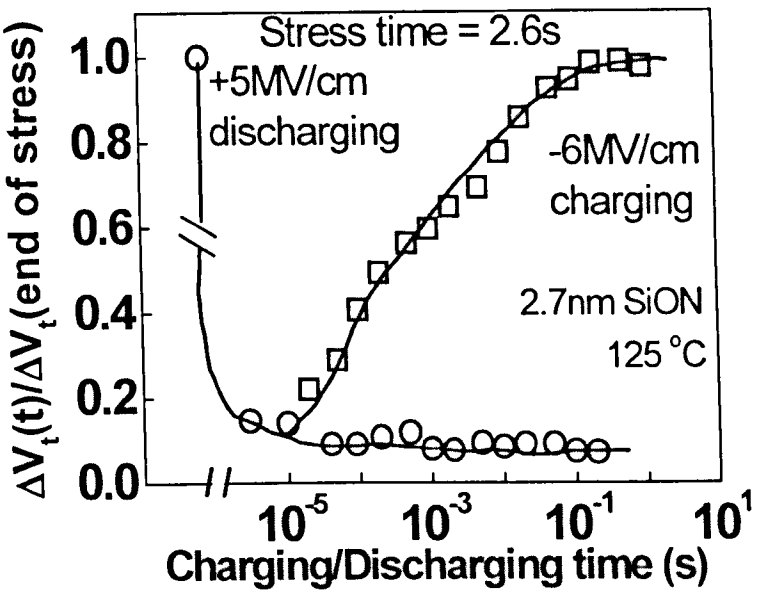


Fig 4.11 A comparison of the charging and discharging rate for the as-grown defects. The stress time is 2.6 s that corresponds to the region where as-grown defects dominate. The discharging under $V_g>0$ is much faster than charging under $V_g<0$: a unique signature of as-grown hole traps (AHT). The solid lines are guides-for-the-eye. It should be noted that the rapid discharging within 5 μs 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 5 μs was negligible.

It is now discussed how hot holes can be generated under a modest bias of V_{gop}=-1.2 V. Hot holes can arise from three sources. Firstly, the surface quantization effect gives energy subbands [46,47]. Although the carrier density reduces as the energy increases, there are holes in the higher sub bands. Secondly, the process "a" in figure 4.12 shows that an electron tunneling from the gate can recombine with a hole in the substrate and the released energy can create hot holes [48]. Third, it has been proposed that photons can be emitted by electron-hole recombination in the gate [49], as illustrated by the process "b" in figure 4.12. Holes can become hot by absorbing the photons.

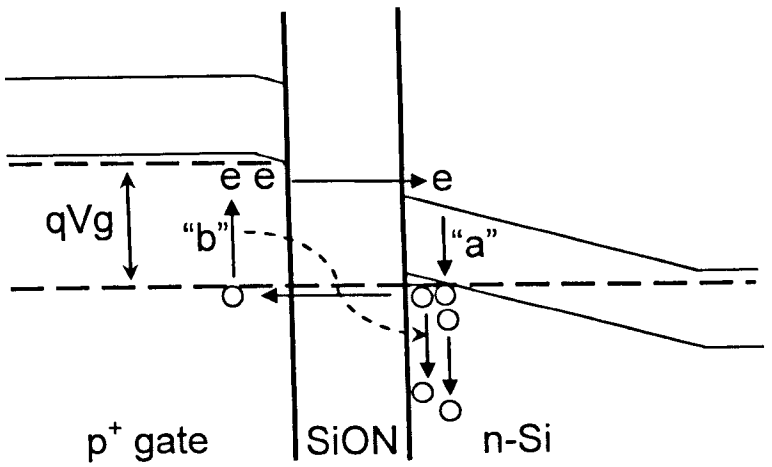


Fig 4.12 A schematic diagram showing the physical processes for generating hot holes under V_g=-1.2 V. The process "a" shows that the energy released by the electron-hole recombination in the substrate can create hot holes. The process "b" shows that the photons emitted from the gate can be absorbed to generate hot holes.

On the kinetics, the charging of AHT generally follows the first order reaction model [50,51], while the generation of new defects follows a power law [5,6,27]. By combining these two, write:

$$\Delta V_t = At^n + c(1 - e^{-t/t^*}) \tag{4.12}$$

For a given stress temperature and bias, 'A', 'n', 'c', and t* are constants and were obtained by fitting test data with a least square technique. There is only one 'n' for the whole stress period and this 'n' is not the slope of the data in figure 4.8, namely $n \neq d(\log|\Delta V_t|)/d[\log(t)]$. Figure 4.8 shows that equation (4.12) can model the 'shoulder' and this simple physics-based model can fit the ΔV_t 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 generated defects, earlier work [12,13,42-45,50,51] shows that both interface states and new hole traps are created by stressing. The new hole traps are further separated into anti-neutralization positive charges and cyclic positive charges, each with a unique signature.

4.7 Lifetime prediction based on a single test

The V_g acceleration technique was developed for predicting the lifetime caused by either hot carrier [52] or the time dependent dielectric breakdown (TDDB) [53]. For TDDB, intrinsic dielectric breakdown generally does not occur within an affordable test time under the operation bias and multiple V_g accelerated tests are essential. For NBTI with a sensing $|V_g|$=1.2 V based on UFP measurement, it has been shown that the prediction based on V_g acceleration is inapplicable and an obvious question is how to assess lifetime in this case. One possibility is to use thermal acceleration. Figure 4.9, however, shows that $\log|\Delta V_t|$ at different temperatures is not shifted in parallel, so that it does not offer a reliable prediction method. Since ΔV_t can be measured at V_{gop} (figure 4.8), in principle, lifetime can be estimated by extrapolating the ΔV_t against stress time based on a single test at V_{gop}. This requires a reliable kinetic model that can not only fit the existing test data, but also predict the future trend.

4.7.1 Verification of prediction method

The affordable test time is typically in the order of days and the data can be 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_t at the last test point is considered as $\Delta V_t(\tau)$. The time for the last point is treated as the measured lifetime τ_m, although τ_m is not the actual device lifetime. A 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 [5,54]. One proposed method is to only use data with a stress time over 10 s to fit the power law against time [5], but there is no information on the prediction accuracy. Figure 4.13 shows that this method can overestimate τ by a factor of 75000. Another proposed method is to fit $\Delta V_t(t)$-ΔV_t(1 s) with a power law [54], but figure 4.14 shows that it underestimates τ by a factor of 10. By applying the model of equation (4.12) to the same set of data, figure 4.15 shows that good agreement is achieved between the measurement and the prediction with $\tau_p/\tau_m = 1.03$.

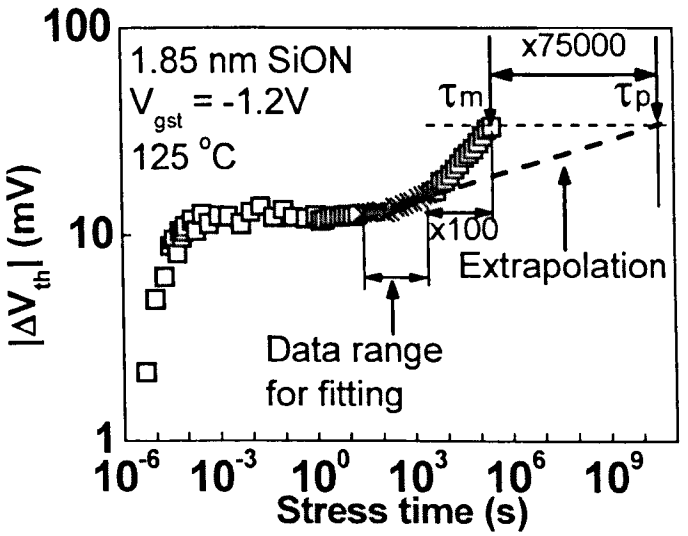


Fig 4.13 Lifetime prediction based on the method proposed in ref. 5. All symbols are test data but only symbol "×" was used for fitting with a power law in the range of 26.8 s < t < 2680 s. The thick dashed line is extrapolated from the fitted line for prediction. τ_m is the time for the last test point and τ_p is the predicted time.

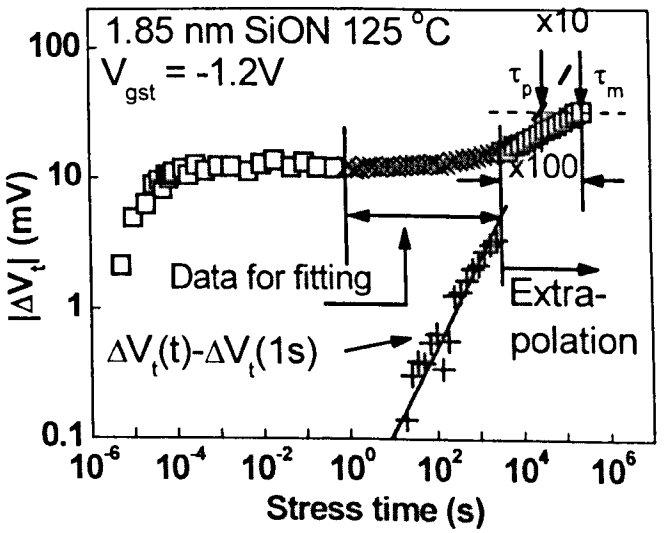


Fig 4.14 Lifetime prediction based on the method proposed in [53]. All symbols are test data but only symbol "×" was used for fitting. ΔVt(t) - ΔVt(1 s) (the symbol "+") was fitted with a power law and ΔVt(1 s) was then added back. The dashed line is extrapolated for prediction. τ_m is the time for the last test point and τ_p is the predicted time.

4.7.2 Applicability to samples from different processes

Although a good prediction is achieved in figure 4.15, it is inadequate to demonstrate that a prediction method works for only one process. For a prediction technique to be useful, it must be applicable to samples fabricated by a wide range of processes. The applicability is now tested for four other SiON layers with different nitrogen concentrations and nitrided either by plasma or thermal methods. Moreover, an ALCVD HfSiON/SiON stack is also tested. Figures 4.16 (a)-(e) shows that ΔV_t follows equation (4.12) in all cases, although the 'shoulder' in some samples is less apparent. Importantly, the prediction achieved a safety margin of 50% or less in all processes tested, giving confidence that the single test technique is generally applicable.

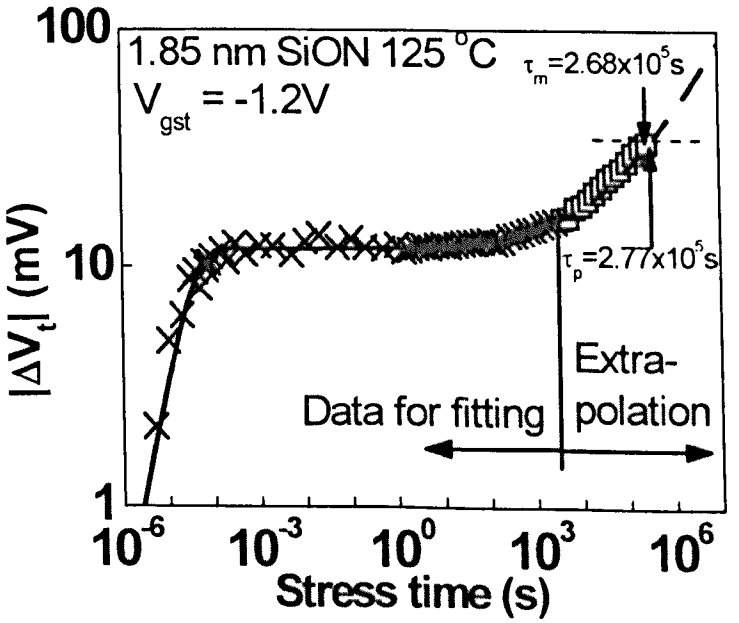
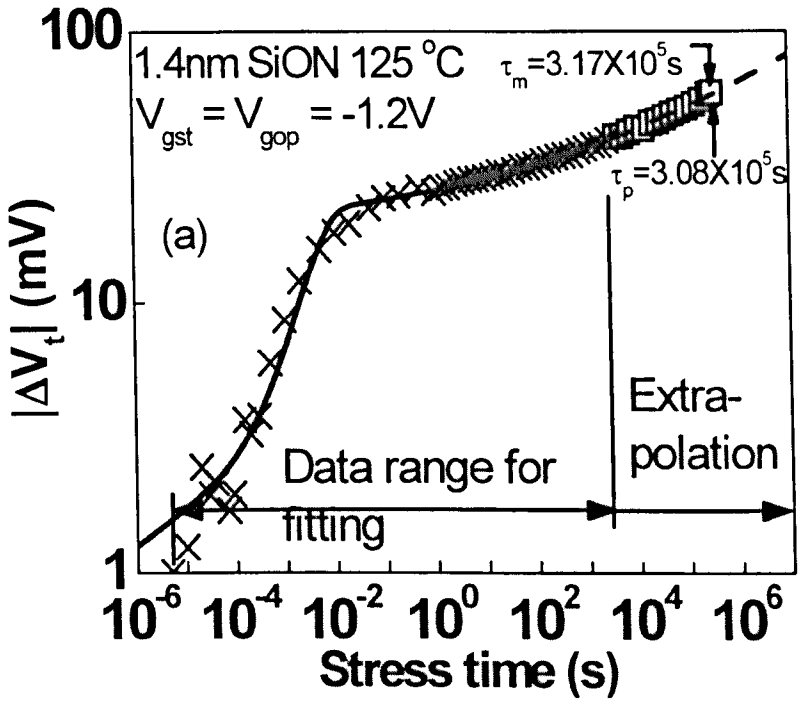
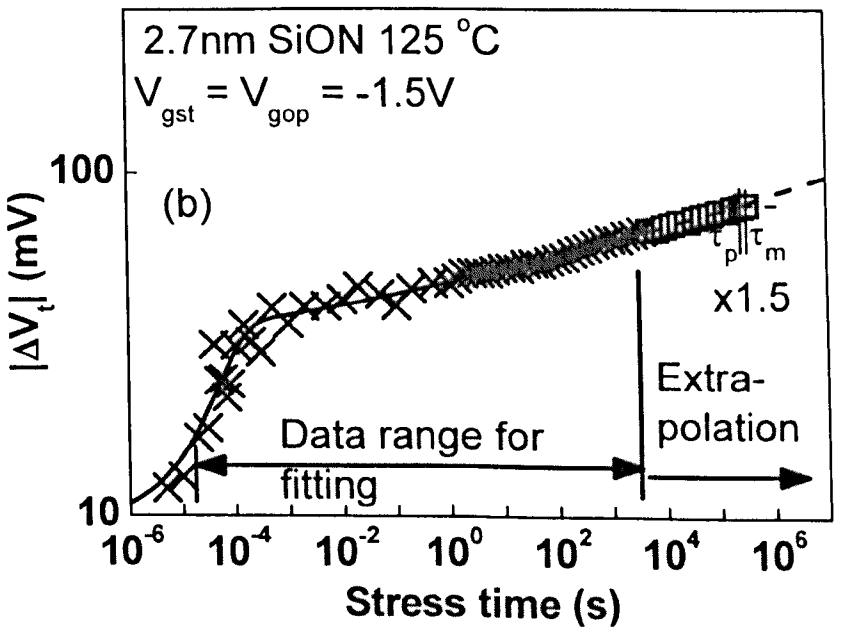


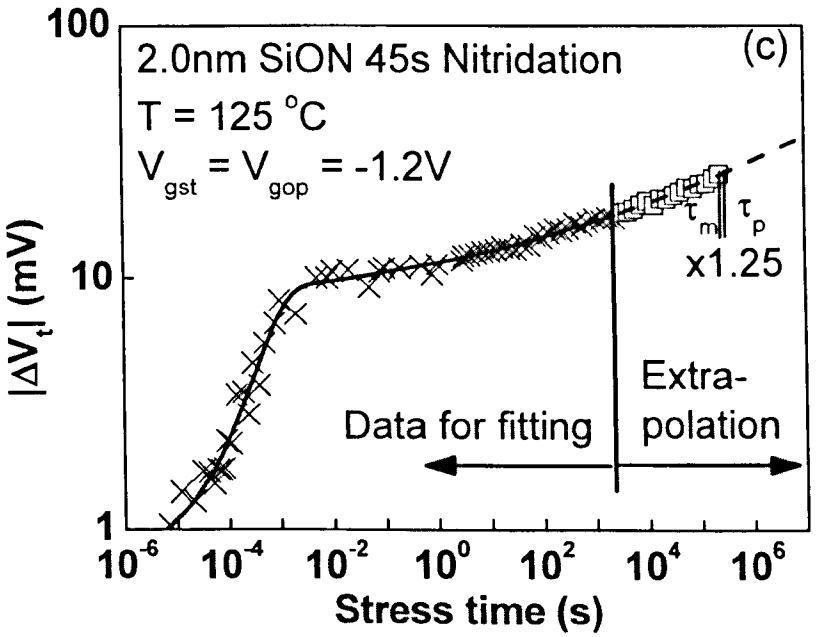
Fig 4.15 Lifetime prediction based on our model: the equation (4.12). The test data is the same as those in Figs. 13 and 14 and only symbol "×" was used for fitting.


1.4nm SiON 125 °C
τ_m=3.17X10^{5}s
$V_{gst} = V_{gop}$ = -1.2V
τ_p=3.08X10^{5}s
(a)
Data range for fitting
Extra-polation
$|\Delta V_t|$ (mV)
Stress time (s)

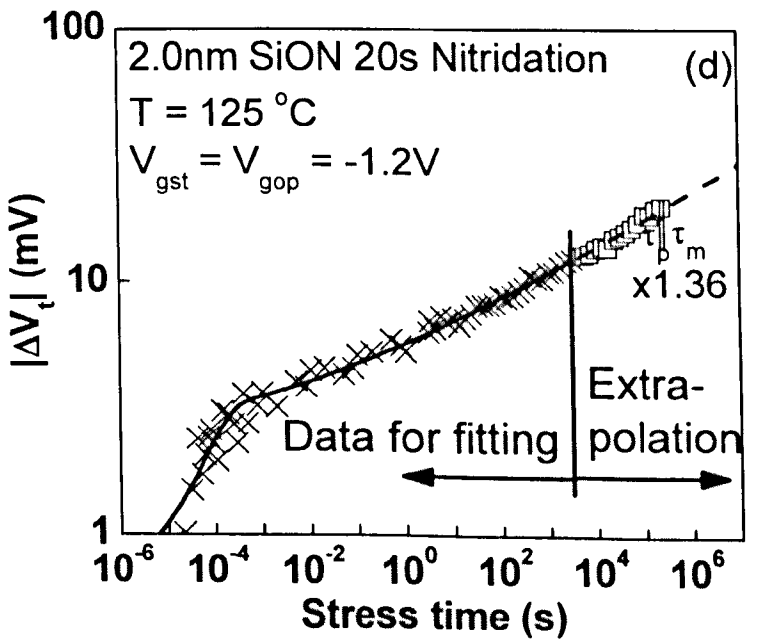
(a)


2.7nm SiON 125 °C
$V_{gst} = V_{gop}$ = -1.5V
(b)
τ_p τ_m
x1.5
Data range for fitting
Extra-polation
$|\Delta V_t|$ (mV)
Stress time (s)

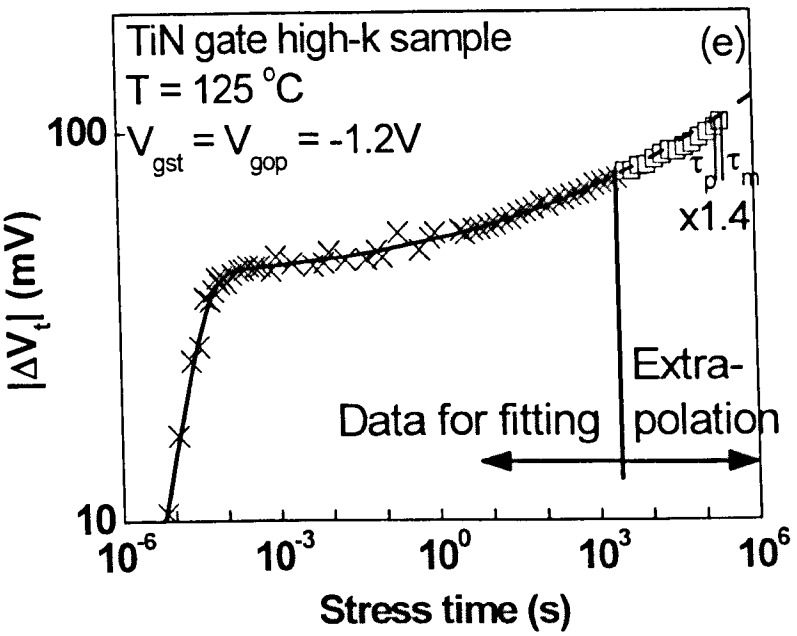
(b)


100
10
1
2.0nm SiON 45s Nitridation
(c)
T = 125 °C
V_gst = V_gop = -1.2V
τ_m
τ_p
x1.25
Extra-
polation
Data for fitting
|ΔV_t| (mV)
10^-6 10^-4 10^-2 10^0 10^2 10^4 10^6
Stress time (s)

(c)


100
10
1
2.0nm SiON 20s Nitridation
(d)
T = 125 °C
V_gst = V_gop = -1.2V
τ_p
τ_m
x1.36
Extra-
polation
Data for fitting
|ΔV_t| (mV)
10^-6 10^-4 10^-2 10^0 10^2 10^4 10^6
Stress time (s)

(d)



(e)

Fig 4.16 Applicability of the single test lifetime prediction technique for different fabrication processes: (a) 1.4 nm plasma SiON, (b) 2.7 nm thermal SiON, (c) 2.0 nm 45 s plasma SiON, (d) 2.0 nm 20 s plasma SiON, and (e) 2.0 nm/1.0 nm HfSiON/SiON stack prepared by ALCVD with TiN gate. The safety margin for the prediction is within 50% in all cases.

4.7.3 An analysis of extracted power factors

Table 2 shows that the fitted power factors for different processes have a range of 0.07~0.36, which agrees with the range reported by earlier work [55-58].

Earlier work reported that the variation of power factor could come from two sources: different hydrogenous species and different nitrogen densities and distributions.

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

Device number	Gate Dielectrics	A (mV/s)	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

On the hydrogenous species, the work of Ogawa et al [19] showed that the power factor for atomic hydrogen should be 0.25. More recently, however, it has been observed that the power factor can be either considerably higher or lower than 0.25. On one hand, it was reported that the power factor could rise to 0.5, if the hydrogenous species is H^+, where the transport was field-assisted. On the other hand, it was reported that the power factor should be 0.16, if the hydrogenous species is H_2 [59].

On the nitridation effect, it was reported that an increase of nitrogen reduces the

power factor through increased dispersion for the transportation of hydrogenous species [60]. 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. This was attributed to the pile-up of nitrogen towards the SiON/substrate interface in a thermally nitrided sample [55].

The power factor reported in this work generally agrees with the trend observed by the earlier work. The sample A with the highest power factor of 0.36 has the lowest nitrogen density and figure 4.17 shows that the pure SiO_2 has high power factors [56-58]. The sample C was thermally nitrided and has the lowest power factor of 0.07 [59-62], again agreeing with the trend in figure 4.17.

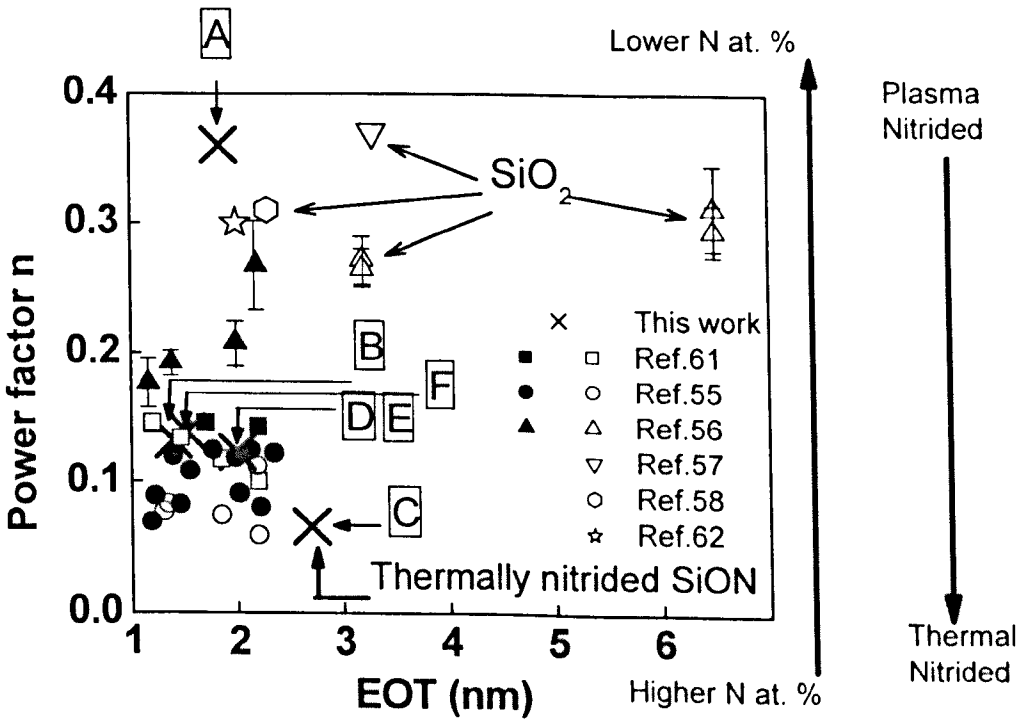


Fig 4.17 A comparison of the power factor obtained in this work with that reported by earlier work.

As a result, this work confirms the trend, but throws little new light on the power factor. What is new for this work is that it proposes a method for predicting the worst case device lifetime without using the conventional V_g acceleration and it shows the applicability of this method for processes over a wide range of power factors.

4.8 Summary

This chapter investigates the NBTI lifetime prediction in the worst case scenario where the recovery is suppressed and ΔV_t is sensed at the operational 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 $\log|\Delta V_t|$.

To predict the lifetime at the operational 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, ΔV_t can be modelled 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 earlier work.

References

1. D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, pp. 1-17, 2003.
2. 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 *Symposia on VLSI Technology and Circuits*, pp. 92-93, 2005.
3. M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. R. Tauriac and N. Revil, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's," in *International Electron Devices Meeting (IEDM) Digest*, pp. 109-112, 2004.
4. J. F. Zhang, Z. Ji, M. H. Chang, B. Kaczer, and G. Groeseneken, "Real Vth instability of pMOSFETs under practical operation conditions," in *International Electron Devices Meeting (IEDM) Digest*, pp. 817-820, 2007.
5. 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 *International Electron Devices Meeting (IEDM) Digest*, pp. 809-812, 2007.
6. A. Aoulaiche, B. Kaczer, D. Jacger, M. Houssa, K. Martens, R. Degraeve, P. Roussel, J. Mitard, S. D. Gendt, H. E. Maes, G. Groeseneken, M. Meuris and M. M. Heyns, "Negative bias temperature instability on Si-passivated Ge-interface", in *international reliability physics symposium (IRPS)*, pp. 262-264, 2008.
7. 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 *Integrated Reliability Workshop Final Report (IIRW)*, pp.1-4, 2006.
8. 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.
9. H. Puchner and L. Hinh, "NBTI reliability analysis for a 90nm CMOS technology," in *European Solid-State Device Research Conference (ESSDERC)*, pp. 257-260, 2004.

10. 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.
11. M. Rafik, X. Garros, G. Ribes, G. Ghibaudo, C. Hobbs, A. Zauner, M. Muller, V. Huard, and C. Ouvrard, "Impact of TiN metal gate on NBTI assessed by interface states and fast transient effect characterization," in *International Electron Devices Meeting (IEDM) Digest*, pp. 825-828, 2007.
12. J. F. Zhang, M. H. Chang, and G. Groeseneken, "Effects of measurement temperature on NBTI," *IEEE Electron Dev. Lett.*, vol. 28, pp. 298-300, 2007.
13. M. H. Chang and J. F. Zhang, "On positive charges formed under negative bias temperature stresses", *J. Appl. Phys.*, vol. 101, pp. 1201-1206, 2007.
14. *Agilent 4155C/4156C Manuals: including User's Guide, Programmer's Guide and Command Reference*. Agilent Technologies, 2003.
15. K. O. Jeppson and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices," *J. Appl. Phys.*, vol. 48, pp. 2004-2014, 1977.
16. H. Kufluoglu and M. A. Alam, "The theory of interface-trap-induced NBTI degradation for reduced cross section MOSFETs," *IEEE Trans. Electron Dev.*, vol. 53, pp. 1120-1130, 2006.
17. M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *International Electron Devices Meeting (IEDM) Digest*, pp. 14.4.1–14.4.4, 2003.
18. 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 *international reliability physics symposium (IRPS)*, pp. 273-282, 2004.
19. 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.
20. C. Shen, M. -F. Li, X. P. Wang, Y. -C. Yeo and D.-L. Kwong, "One fast measurement technique of MOSFET Id-Vg characteristics," *IEEE Electron Dev. Lett.*, vol. 27, pp. 55-57, 2006.
21. C. Shen, M. -F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra and Y.-C. Yeo, "Characterization and physical origin of fast Vth transient in NBTI of pMOSFETs with SiON dielectric," in *International Electron Devices Meeting (IEDM) Digest,* pp. 333-336, 2006.
22. Y. N. Cohen, J. Shappir and D. F. Bentchkowsky. "Dynamic model of trapping-detrapping in SiO2," *J. Appl. Phys.*, vol. 58, pp. 2252–2261, 1985.

23. C. H. Liu, M. T. Lee, C. Y. Lin, J. Chen, K. Schruefer, J. Brighten, N. Rovedo, T. B. Hook, M. V. Khare, S. –F. Huang, C. Wann, T. -C, Chen and T. H. Ning. "Mechanism and process dependence of negative bias temperature instability (NBTI) for pMOSFETs with ultrathin gate dielectrics," in *International Electron Devices Meeting (IEDM) Digest,* pp. 39.2.1–39.2.4, 2001.
24. K. K. Abdelghafar, K. Watanabe, J. Ushio and E. Murakami, "Effect of nitrogen at SiO2/Si interface on reliability issues negative-bias-temperature instability and Fowler-Nordheim-stress degradation," *Appl. Phys. Lett.,* vol. 81, pp. 4362-4364, 2002.
25. L. Duong, E. Li, V. Gopinath, S. Prasad, J. Lin, D. Pachura and V. Hornback, "Effect of nitrogen incorporation on PMOS negative bias temperature instability in ultrathin oxi-nitrides," in *Integrated Reliability Workshop Final Report (IIRW),* pp.128–130, 2003.
26. S. S. Tan, T. P. Chen, C. H. Ang and L. Chan, "Relationship between interfacial nitrogen concentration and activation energies of fixed-charge trapping and interface state generation under bias-temperature stress condition," *Appl. Phys. Lett.,* vol. 82, pp. 269-271, 2003.
27. S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang and L. Chan, "Nitrogen-enhanced negative bias temperature instability: An insight by experiment and first-principle calculations," *Appl. Phys. Lett.,* vol. 82, pp. 1881-1883, 2003.
28. S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang, W. Y. Teo and L. Chan, "Linear relationship between H+-trapping reaction energy and defect generation: Insight into nitrogen-enhanced negative bias temperature instability," *Appl. Phys. Lett.,* vol. 83, pp. 530-532, 2003.
29. S. S. Tan, T. P. Chen, C. H. Ang and L. Chan, "Atomic modelling of nitrogen neighboring effect on negative bias temperature instability of pmosfets," *IEEE Electon Device Letters*, vol. 25, pp. 504–506, 2004.
30. J. Ushio, T. Maruizumi and K. K. Abdelghafar, "The interface structures generated by negative-bias temperature instability in Si/SiO2 and Si/SiOxNy interfaces," *Appl. Phys. Lett.,* vol. 81, pp. 1818-1820, 2002.
31. *JEDEC/FSA Foundry Process Qualification Guidelines (JP-001),* JEDEC Solid State Technology Association, 2002.
32. M. Ershov, S. Saxena, S. Minehane, P. Clifton, M. Redford, R. Lindley, H. Karbasi, S. Graves and S. Winters, "Degradation dynamics, recovery, and characterization of negative bias temperature instability," *Microelectronics Rel.,* vol. 45 pp. 99-105, 2005.

33. E. Takeda and N. Suzuki, "An empirical model for device degradation due to hot-carrier injection," *IEEE Electron Dev. Lett.*, vol. 4, pp. 111-113, 1983.
34. A. Narr and A. Lill, "Lifetime prediction for PMOS and NMOS devices biased on a degradation model for gate-bias-stress," *Microelectronics Rel.*, vol. 37, pp. 1433–1436, 1997.
35. 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 *Symposia on VLSI Technology and Circuits,* pp. 73-74, 1999.
36. T. Yamamoto, K. Uwasawa and T. Mogami, "Bias temperature instability in scaled p+ polysilicon gate p-MOSFET's," *IEEE Trans. Electron Dev.*, vol. 46, pp. 921-926, 1999.
37. G. Chen, K. -Y. Chuah, M. -F. Li, D. S. H. Chan, C. H. Ang, J. Z. Zheng, Y. Jin, and D. -L. Kwong, "Dynamic NBTI of PMOS transitors and its impact on device lifetime," in *international reliability physics symposium (IRPS)*, pp. 196-202, 2003.
38. H. Aono, E. Murakami, K. Okuyama, A. Nishida, M. Minami, Y. Ooji, and K. Kubota, "Modeling of NBTI saturation effect and its impact on electric field dependence of the lifetime," *Microelectronics Rel.*, vol. 45 pp. 1109-1114, 2005.
39. X. Li B. Huang, J. Qin, X. Zhang, M. Talmor, Z. Gur, and J. B. Bernstein, "Deep submicron CMOS integrated circuit reliability simulation with SPICE," in *International Society for Quality Electronic Design (ISQED)*, pp.1-8, 2005.
40. W. Heinrigs, H. Reisinger, W. Gustin and C. Schlunder, "Consideration of recovery effects during NBTI measurements for accurate lifetime predictions of state-of-the-art pMOSFETs," in *international reliability physics symposium (IRPS)*, pp.288-292, 2007.
41. G. V. D. 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, pp. 2073-2080, 1994.
42. C. Z. Zhao, J. F. Zhang, M. H. Chang, A. R. Peaker, S. Hall, G. Groeseneken, L. Pantisano, S. D. 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.
43. 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.

44. 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.
45. C. Z. Zhao and J. F. Zhang, "Effects of hydrogen on positive charges in gate oxides," *J. Appl. Phys.*, vol. 97, pp. 1231-1234, 2005.
46. Y. T. Hou, M. F. Li, Y. Jin, and W. H. Lai, "Direct tunneling hole currents through ultrathin gate oxides in metal-oxide-semiconductor devices," *J. Appl. Phys.*, vol. 91, pp. 258-264, 2002.
47. C. Z. Zhao, J. F. Zhang, M. B. Zahid, B. Govoreanu, G. Groeseneken, and S. D. Gendt, "Determination of capture cross sections for as-grown electron traps in HfO_2/HfSiO stacks," *J. Appl. Phys.*, vol. 100, pp. 3201-3204, 2006.
48. J. D. Bude, B. E. Weir, and P. J. Silverman, "Explanation of stress-induced damage in thin oxides," in *International Electron Devices Meeting (IEDM) Digest*, pp. 102-107, 1998.
49. M. Xu, C. Tan, and M. F. Li, "Extended Arrhenius law of time-to-breakdown of ultrathin gate oxides," *Appl. Phys. Lett.*, vol. 82, pp. 2482-2484, 2003.
50. 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.
51. 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.
52. A. Acovic, G. L.Rosa and Y. -C. Sun, "A review of hot-carrier degradation mechanisms in MOSFETs," *Mircoelectron. Reliab.*, vol. 36, pp. 845-869, 1996.
53. E. Y. Wu, J. Aitken, E. Nowak, A. Vayshenker, P. Varekamp, G. Hueckel, J. McKenna, D. Harmon, L. -K. Han, C. Montrose, and R. Dufresne, and R.-P. Vollertsen, "Voltage-dependent Voltage-acceleration of oxide breakdown for ultra-thin oxides," in *International Electron Devices Meeting (IEDM) Digest*, pp. 541-544, 2000.
54. 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.
55. 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.

56. 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 *International Electron Devices Meeting (IEDM) Digest*, pp. 691-694, 2005.
57. 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 *Integrated Reliability Workshop Final Report (IIRW)*, pp. 98-101, 2000.
58. Y. Mitani, M. Nagamine, H. Satake, and A. Toriumi, "NBTI mechanism in ultra-thin gate dielectric - nitrogen-originated mechanism in SiON," in *International Electron Devices Meeting (IEDM) Digest,* pp. 509-512, 2002.
59. M. A. Alam, H. Kufluoglu, D. Varghese and S. Mahapatra, "A comprehensive model for PMOS NBTI degradation: Recent progress," *Microelectronics Rel.*, vol. 47, pp. 853-862, 2007.
60. J. B. Yang, T. P. Chen, S. S. Tan, C. M. Ng and L. Chan, "Influence of hydrogen dispersive diffusion in nitrided gate oxide on negative bias temperature instability," *Appl. Phys. Lett.*, vol. 93, pp. 1093-1097, 2008.
61. S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, "On the physical mechanism of NBTI in Silicon oxynitride p-MOSFETs: can differences in Insulator Processing conditions resolve the interface trap generation versus hole trapping controversy?" in *International Electron Devices Meeting (IEDM) Digest,* pp. 1-9, 2007.
62. Y. Mitani, "Influence of Nitrogen in ultra-thin SiON on Negative Bias Temperature Instability under AC stress," in *International Electron Devices Meeting (IEDM) Digest,* pp. 117-120, 2004.

5 Effects of remote charge scattering on mobility degradation

5.1 Introduction

One driving force for downscaling metal-oxide-semiconductor field effect transistors (MOSFETs) is to improve circuit operation speed. When the gate SiO_2 thickness, T_{ox}, is over 5 nm, carrier mobility is insensitive to T_{ox} [1]. Once it is below 2–3 nm, however, mobility reduces for thinner T_{ox} [2-7]. The reported reduction near threshold voltage varies substantially, from a factor over 2 [2] to an insignificant level [4]. As gate voltage increases from threshold, the relative mobility reduction can either decrease [2,3] or increase [4,5]. Agreement has not yet been reached on the origin for such phenomena and the proposed mechanisms include remote charge scattering (RCS) from impurities in the depleted poly-Si gate [2,3,6], increased surface roughness [4,5], and long-range Coulomb interaction between carriers in the gate and in the inversion layer [7]. Even among the various RCS theories, depending on the assumptions used in the model, the calculated RCS mobility can vary over two orders of magnitude

[2,6]. For the high-*k*/SiO_2 stack, it has been reported that a reduction in the thickness of interfacial SiO_2 below 2.5 nm can progressively reduce carrier mobility [3], indicating the presence of remote scattering. Soft optical phonons [8,9], remote surface scattering [10], and RCS from charges either in the bulk of high-*k* layer or at the high-*k*/SiO_2 interface [3,6] can all contribute to a lower mobility.

One weakness of earlier work is that different samples were used when experimentally studying the RCS, which leads to uncertainties. For example, a reduction in T_{ox} can not only bring the gate closer to the substrate, but could also modulate other factors such as surface roughness [4,5]. In this work, these uncertainties will be removed to allow a study of the impact of RCS on mobility by varying charges in the same device through either processing or electron trapping.

5.2 Samples selection

The typical doping level for poly-Si gates is 10^{19}–10^{20} cm^{-3} and it is proposed that the ionized impurity reduces effective mobility when $T_{ox} \leq 2nm$ [3,6]. As a result, the criteria for selecting test samples are as follows: the dielectric charging can vary in the order of 10^{20} cm^{-3} and must be within 2 nm from the substrate. Four different samples were selected and the details are given in table 5.1. All samples have a substrate doping of 5×10^{17} cm^{-3}. The channel length

is 0.25–0.8 μm and the channel width is 10 μm. The samples with an IMEC Clean process were activated by a 1000 °C spike anneal.

Table 5.1 The samples

Sample	A	B	C	D
Device	nMOSFET	pMOSFET	nMOSFET	nMOSFET
T_{HfO2}	2 nm	2nm	2nm	1.5nm HfSiON
HfO_2: Material	HfO_2	HfO_2	HfO_2	ALD
Process	ALD	ALD	PVD	
T_{SiO2}	0.56 nm	0.73 nm	1 nm	0.77nm
SiO_2: Material	SiO_2	SiON	SiO_2	SiO_2
Process	SPER 650 °C	IMEC Clean	IMEC Clean	IMEC Clean
EOT	0.96 nm	1.13nm	1.62nm	1.16nm
Gate material	TaN/TiN	TaN/TiN	TaN/TiN	Poly-Si

5.3 Conventional mobility characterization technique

In this chapter, the principle of the conventional mobility measurement will be introduced, and then the problems of this technique for the devices with high-k stacks will be raised.

5.3.1 Principle of conventional mobility measurement

In order to study the impact of RCS on mobility, accurate measurements of high-k mobility must be performed. For a MOSFET working in the linear region, the drain current I_d can be written as [11]

$$I_d = \frac{W}{L} \cdot \mu_{eff} \cdot Q_{inv}(V_g) \cdot V_d \qquad (5.1)$$

μ_{eff} is the effective mobility of the carriers in the inversion layer and Q_{inv} is the inversion charge density, which is a function of the gate bias. Rearranging equation (5.1), μ_{eff} can be determined from

$$\mu_{eff} = \frac{L}{W} \cdot \frac{I_d(V_g)}{V_d \cdot Q_{inv}(V_g)} = \frac{L}{W} \cdot \frac{g_d}{Q_{inv}} \tag{5.2}$$

where the drain conductance g_d is defined as

$$g_d = \frac{I_d}{V_d} \tag{5.3}$$

The most important parameters in equation (5.2) are Q_{inv} and I_d. To find inversion charge density Q_{inv}, two approaches are commonly used. Q_{inv} can be approximated by

$$Q_{inv}(V_g) = C_{ox} \cdot (V_g - V_t) \tag{5.4}$$

where C_{ox} is the oxide capacitance per unit area and V_t is the threshold voltage of the device. Since the threshold is not uniquely defined and different definitions can lead to different V_t [12], equation (5.4) can give inaccurate estimation of the inversion charge, especially near V_t.

A better approximation is to use the method called split CV [13]

$$Q_{inv}(V_g) = \int_{-\infty}^{V_g} C_{gc}\, dV_g \tag{5.5}$$

where C_{gc} is the gate to channel capacitance. C_{gc} is measured by connecting the capacitance meter between the gate and the source-drain while grounding the substrate as illustrated by figure 5.1 (a). At gate voltages below threshold, the

capacitance is equal to the total overlap capacitance of the gate to the source and drain, $2C_{ov}$. As the channel starts to invert and inversion charge appears, the capacitance increases.

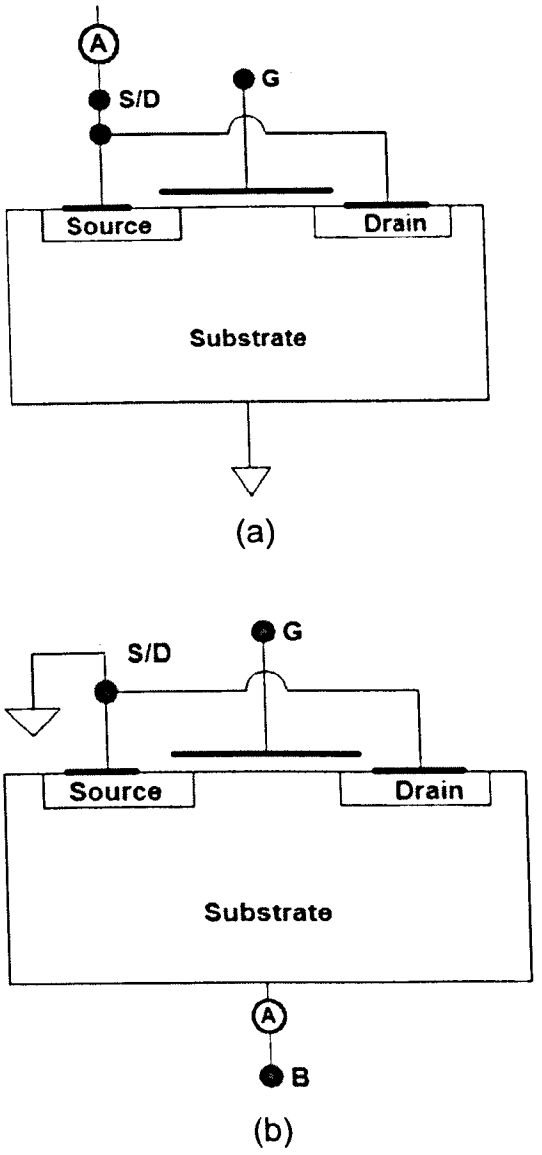


Fig 5.1 Configuration for split C-V technique. Conventionally, split C-V technique is performed by using LCR meters. (a) In order to measure gate-to-channel capacitance, C_{gc}, the terminal marked HI is connected to the gate and LO to the source and drain, the substrate of the device is connected to GND. (b) In order to measure gate-to-substrate capacitance, C_{gb}, the terminal HI is connected to the gate and LO to the substrate.

As the voltage is increased further, the capacitance saturates to some inversion

value which is a combination of the overlap capacitance and the channel capacitance, $2C_{ov} + C_{ch}$. Subtracting $2C_{ov}$ and integrating this curve gives Q_{inv} as a function of V_g. The gate to substrate capacitance, C_{gb}, can be measured in a similar way. The connection is shown in figure 5.1 (b). The bulk charge density, Q_b, can be formed by integration in a similar way to equation (5.5),

$$Q_b(V_g) = \int_{V_{fb}}^{V_g} C_{gb} dV_g \tag{5.6}$$

By using this split C-V technique, Q_{inv} and Q_b can be measured separately. Q_{inv} is used to calculate mobility and both Q_{inv} and Q_b are used to calculated effective surface field.

The drain conductance, g_d, can be determined from eqn. (3) using a single low drain bias, typically 50–100 mV [14], and measuring the DC drain current.

Carrier mobility given in Eqn. (2) is usually plotted versus the effective surface field in the Si substrate which is given by

$$E_{eff} = \frac{1}{\varepsilon_{Si}\varepsilon_0}(Q_b + \eta \cdot Q_{inv}) \tag{5.7}$$

The factor η is given as 1/2 for electron mobility and 1/3 for hole mobility [15]. The depletion charge can either be determined experimentally, or be estimated by [11]

$$Q_b = \sqrt{4\varepsilon_{Si}\varepsilon_0\phi_f qN_A} \tag{5.8}$$

Where

$$\phi_F = \frac{kT}{q}\ln(\frac{N_A}{n_i}) \quad (5.9)$$

5.3.2 Challenge for conventional measurements on high-k devices

Generally, the electron trapping in thin dielectric stack is highly unstable [16, 17]. The slow gate voltage ramp used in split C-V measurement can result in significant charge trapping during the measurement, giving rise to a stretch-out of the C-V curve, thus results in an over-estimation of Q_{inv} [18]. Additionally, a threshold voltage shift caused by the change of oxide trapping density during the I_d-V_g sweep can result in an underestimation of I_d current. These factors lead to large errors in mobility evaluation [19].

In order to suppress trapping during the measurement, two fast techniques are proposed. Pulse I_d-V_g technique [20] is used to measure drain current and pulse C-V technique is developed to measure Q_{inv}. The details of the pulse I_d-V_g technique is given in chapter 2. In the next section, the development of pulse C-V will be described.

5.4 Novel pulse C-V based mobility characterization technique

5.4.1 Shortcomings of existing setups for pulse C-V techniques

The principle of the pulse C-V technique is similar to that of the quasi-static C-V measurement. By using a linear voltage ramp, the corresponding displacement current is measured and then converted to capacitance. One advantage of pulse C-V is that, by raising the ramp rate, the displacement current can be several orders of magnitude higher than that in the conventional quasi static C-V measurement. Two setups were proposed for the pulse C-V measurement. A commercial I/V converter (e.g. Keithley 428) is used to convert the displacement current to a voltage which is recorded with a digital oscilloscope [21,22]. The schematic setup is shown in figure 5.2 (a). This setup gives good measurement accuracy, but the ramp rate is limited to around 10 kV/s [22]. At this rate, only large area devices can be used for the test. Although large area capacitors are normally provided by the manufacturer, it is better to obtain MOSFETs that are large enough for this setup. Nowadays the largest MOSFET available on the test mask is typically 10x10 μm^2. Figure 5.3 shows that if the device with an EOT = 1.53nm and area = 100 μm^2 is used, the displacement current at 10 kV/s is too small. Only when the ramp rate is increased to 80 kV/s, the displacement current can be measured accurately.

Another setup was proposed by Singh et al [18]. They connected the MOSFET

with a capacitor, C_R, to measure inversion charges directly, and it is reported that the pulse edge transition time can reach around 70ns. The schematic setup is shown in figure 5.2 (b). The shortcoming for this setup is that it is very sensitive to parasitic effects from the cable connection and consequently it needs a special probe card with ground plane to minimize inductive effects and reflections [18]. Because of this inconvenience, this technique is seldom used. Since both setups have their shortcomings, a new setup was developed for the pulse C-V technique.

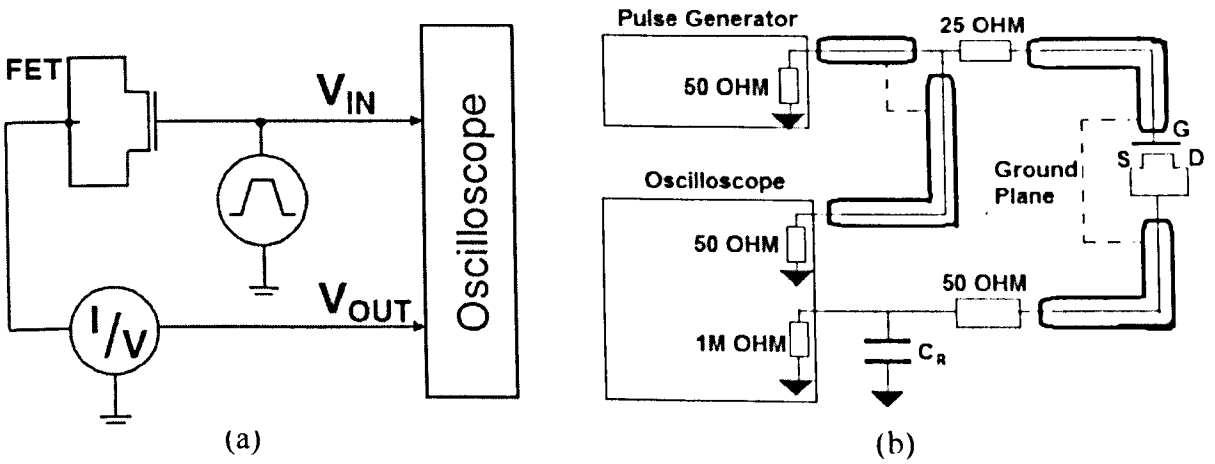


Fig 5.2 (a) Schematic setup used to obtain pulse C-V characteristics. The voltage pulse applied to the gate (V_{in}) is recorded using a digital oscilloscope together with the output voltage (V_{out}) of the current-voltage amplifier which convert displacement current to the voltage by I/V converter Keithley 428. (b) Schematic illustration of the setup used for pulse Q_{inv} measurement, Q_{inv} is measured directly by using an external capacitor connected in parallel with MOSCAP. Parasitic effect must be minimized and attentions should be paid to impedance match and ground plane connection.

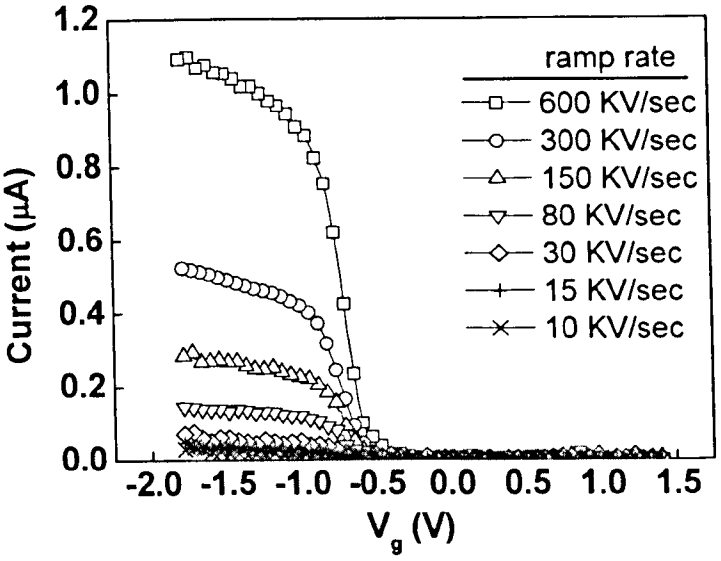


Fig 5.3 Displacement current measured from the MOSFET with an EOT of 1.53 nm and a typical area of 100 μm^2. The magnitude of the displacement current drops for lower ramp rate. When ramp rate is lower than 30 kV/s, the displacement current is below 10^{-8} A and it is not easy to measure accurately.

5.4.2 The new pulse C-V setup

The schematic setup for the new pulse C-V technique is the same as that used by Deleruyell et al [21]. In order to increase the measurement speed, a new circuit was designed to replace the commercial I/V converter. The circuit configuration is shown in figure 5.4 (a). The principle is similar to that for pulse I_d-V_g but the drain is virtually grounded here. The source/drain and substrate are connected to separate circuit so that C_{gc} and C_{gb} can be measured at the same time. By following the same cautions mentioned in chapter 2, the ramp rate can be increased up to 600 kV/s. In order to improve accuracy, 200 times averaging is used. An accuracy of 10^{-8} A is reached, which allows measuring capacitances

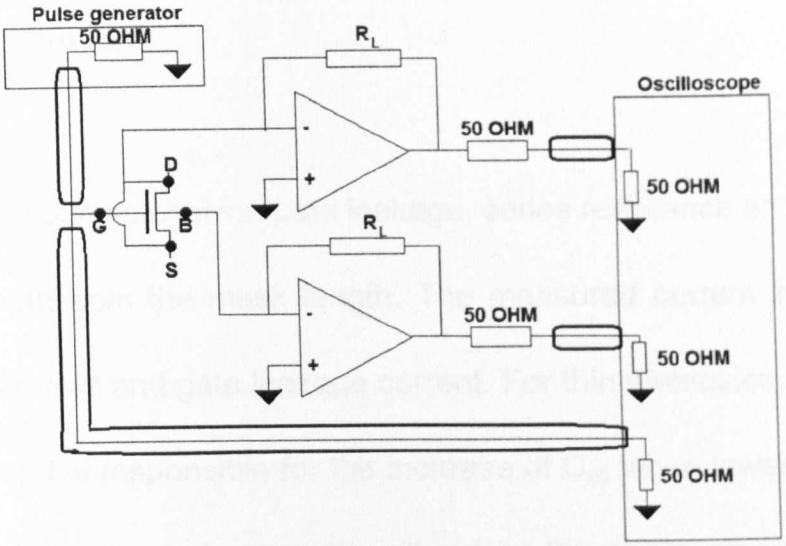
of the order of pico-farads.

When a pulse is applied to the gate, the displacement current is converted to a voltage, V_{out}, and recorded by a digital oscilloscope. The C-V characteristic can then be extracted using the following relationship:

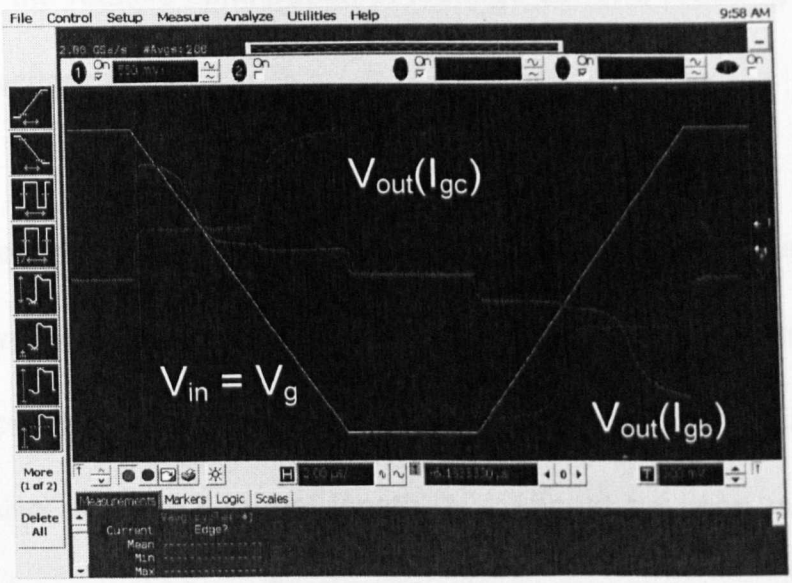
$$C = \frac{V_{out} - V_{offset}}{R \cdot dV / dt} \tag{5.10}$$

where dV/dt is the voltage ramp rate, V_{offset} is the offset voltage of the current-voltage amplifier and R is the feedback resistor between input and output which control the gain of the circuit. The input (V_{in}) and output (V_{out}) voltages correspond to the terminals of the circuits as illustrated by figure 5.4 (a). In figure 5.4 (b), a typical screen-shot is given. The gate voltage $V_g = V_{in}$, and output voltages $V_{out}(I_{gc})$ and $V_{out}(I_{gb})$ are captured by the oscilloscope. A comparison between the pulse C-V technique and the standard HF C-V is given in figure 5.5 (a).

Figure 5.5 (a) shows that the C_{gc}~V_g measured by pulse C-V technique can be different at high $|V_g|$ for different ramp rate. This is caused by gate leakage current and its correction will be addressed next.



(a)



(b)

Fig 5.4 (a) The configuration of new developed pulse C-V setup. In order to measure C_{gb} and C_{gc} at the same time, two amplifiers are used. One amplifier is connected to the source and drain and another one to the substrate. Both amplifiers are connected the ground, effectively grounding the source, drain and substrate. When a pulse is applied on the gate, the displacement current from the source/drain and substrate are converted into voltage traces using two independent circuits and recorded with a digital oscilloscope. (b) A typical screen shot for V_{in}, $V_{out}(I_{gc})$, and $V_{out}(I_{gb})$.

5.4.3 Sources of error

There are three sources of error: gate leakage, series resistance and a reduction of channel length from the mask length. The measured current includes both displacement current and gate leakage current. For thin dielectrics, leakage can be significant and is responsible for the increase of C_{gc} when lowering |dV/dt| in figure 5.5 (a). A decrease of ramp rate will reduce the displacement current and this increases the relative contribution of leakage current, resulting in the larger apparent capacitance.

According to equation (5.2), in order to obtain the true mobility, drain current and effective channel length should also be obtained precisely. Therefore, series resistance from the source and drain should be found for correcting the drain current, I_d. Also effective channel length, L_{eff}, should be used rather than the mask length, L_{mask}, in the specification.

5.4.3.1 Gate leakage current

The displacement current is proportional to dV/dt, but the leakage current, I_g, is independent of it. This allows I_g to be corrected by comparing two currents I_1, I_2, measured at different dV/dt [23].

$$C \cdot dV / dt\big|_{f_1} + I_g = I_1 \tag{5.11}$$

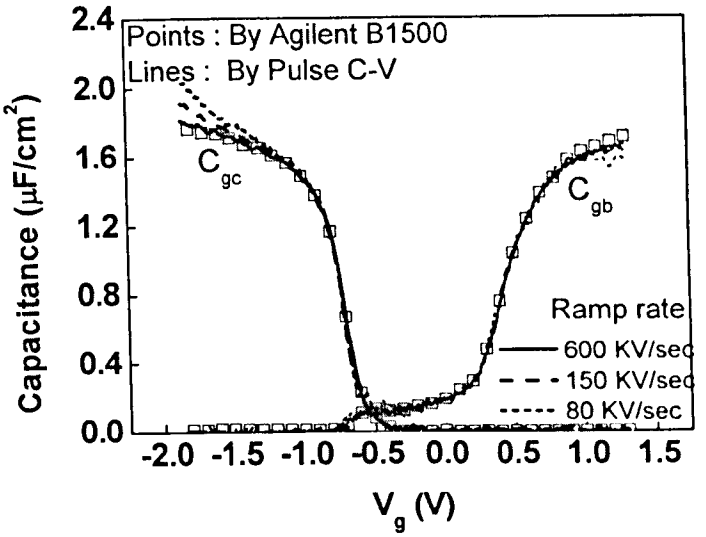
$$C \cdot dV / dt\big|_{f_2} + I_g = I_2 \tag{5.12}$$

Subtracting (12) from (11), we have,

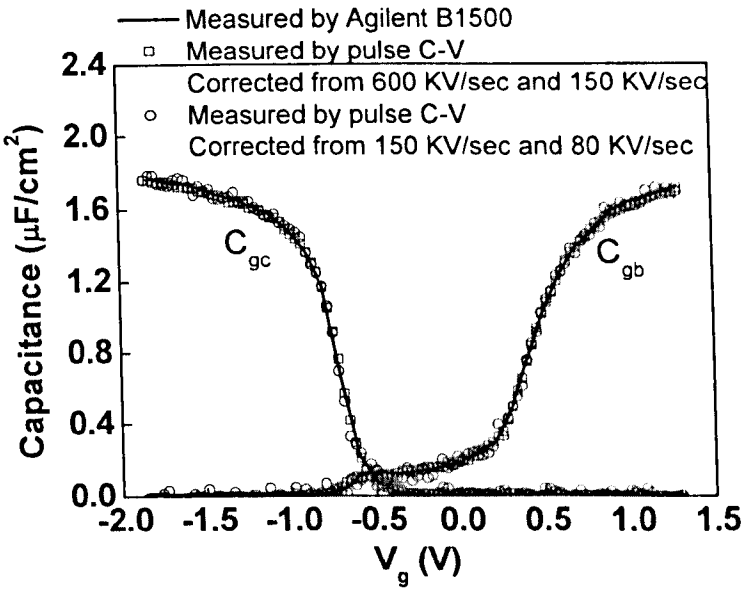
$$C = \frac{I_1 - I_2}{dV / dt\big|_{f_1} - dV / dt\big|_{f_2}} \tag{5.13}$$

Figure 5.5 (b) compares the C-V corrected by using dV/dt = 600 and 150 kV/s with that corrected by using dV/dt = 80 and 150 kV/cm. Unlike the disagreement shown in figure 5.5 (a), after correction, the C-V curves become independent of ramp rate and they also agree well with the standard HF C-V. Inversion charge Q_{inv} and the bulk charge Q_b are integrated by using equation (5.5) and (5.6) and they are shown in figure 5.6. The Q_b calculated from equation (5.8) is also compared.

Figure 5.7 shows the results obtained for the thinner sample B (EOT = 1.13nm).The difference before and after correction can be clearly seen so that a correction is essential. Figure 5.7 (b) shows the inversion charge integrated from C_{gc} before and after correction. Large leakage current gives higher apparent Q_{inv} and may under-estimate the mobility.



(a)



(b)

Fig 5.5 (a) A comparison of the C-V characteristics measured using our pulse C-V technique and the conventional High frequency technique (f = 5 MHz). Three ramp rates were used for the pulse C-V. The pulse C-V varies with ramp rate and does not agree with the high frequency C-V, due to gate leakage current. (b) The gate leakage current is corrected by using C-V at two frequencies. After the correction, all pulse C-V curves agree well with the high frequency C-V.

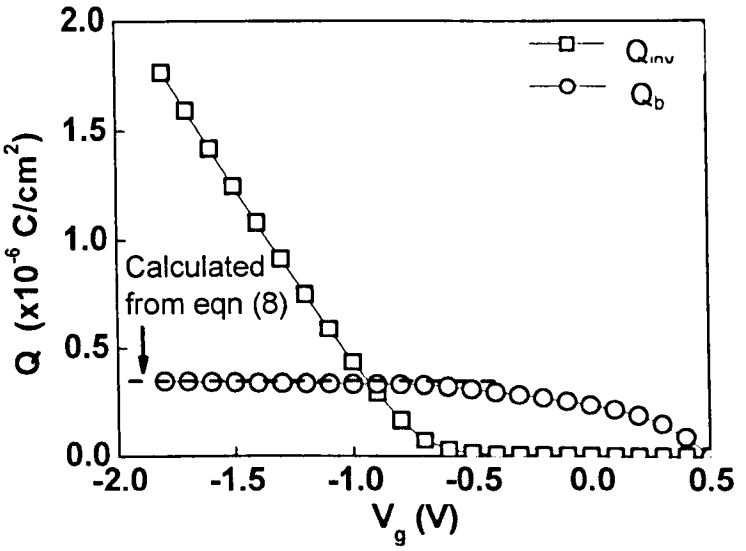


Fig 5.6 Inversion charge Q_{inv} and bulk charge density Q_b obtained from equation (5.5) and equation (5.6). Q_b calculated from equation (5.8) is represented by the dashed line. The substrate doping in our sample is $5x10^{17}$ cm^{-3}.

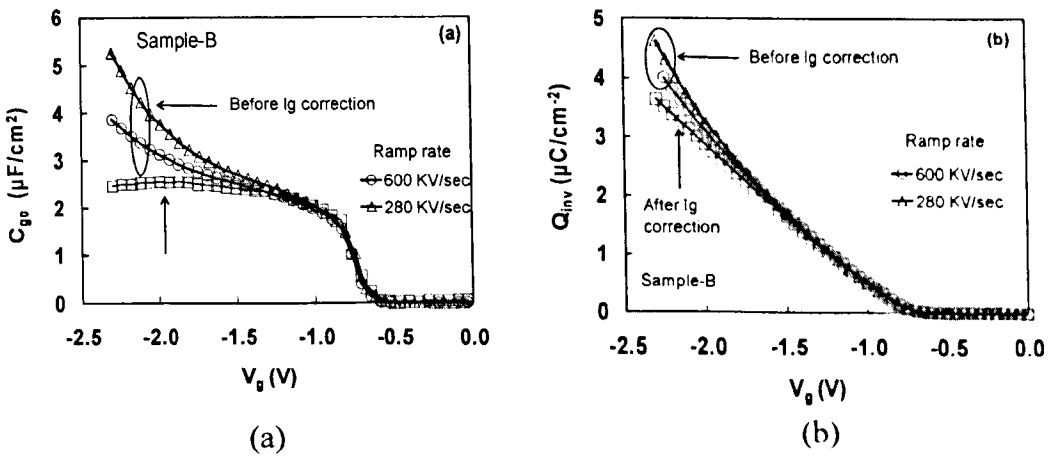


Fig 5.7 (a) C_{gc} is measured by the pulse C-V technique with 600 kV/s and 280 kV/s ramp rate respectively. The thickness of the device used in the test is 1.13 nm and therefore the leakage current is much more severe than the case shown in fig 5.5. C_{gc} before and after correction agrees reasonably well within -1.2V and then begin separating. (b) shows the inversion charge integrated by using C_{gc} before and after correction.

5.4.3.2 Series resistance and effective channel length

In order to extract series resistance and effective channel length, the Shift-and-Ratio method [24,25] is used and a very brief description is given below.

The Shift and Ratio method starts with equation (5.14)

$$R_{tot} \equiv \frac{V_d}{I_d} = R_{sd} + R_{ch} = R_{sd} + L_{eff} \cdot f(V_g - V_t) \tag{5.14}$$

where R_{tot} is the total resistance between source and drain, including series resistance R_{sd} and channel resistance R_{ch}.

By differentiating equation (5.14) with respect to V_g, one can obtain

$$S(V_g) \equiv \frac{dR_{tot}}{dV_g} = \frac{dR_{sd}}{dV_g} + L_{eff} \frac{df(V_g - V_t)}{dV_g} \tag{5.15}$$

Since the parasitic resistance R_{sd} is either independent or a weak function of V_g, the first term in equation (5.15) can be neglected to give

$$S(V_g) \equiv \frac{dR_{tot}}{dV_g} = L_{eff} \frac{df(V_g - V_t)}{dV_g} \tag{5.16}$$

S&R extraction is usually carried out with two devices: one long-channel and one short-channel:

$$S^0(V_g) = L_{eff}^0 \frac{df(V_g - V_t^0)}{dV_g} \tag{5.17}$$

and

$$S^1(V_g) = L^1_{eff}\frac{df(V_g - V_t^1)}{dV_g} \tag{5.18}$$

where the superscript 0 represents the long channel device and 1 represents the short device.

If $V_t^0 = V_t^1$, S^0 and S^1 would be similar functions of V_g and L_{eff} would be extracted from the ratio $S^1/S^0 = L^1_{eff}/L^0_{eff} \simeq L^1_{eff}/L^0_{mask}$. In general, however, $V_t^0 \neq V_t^1$ due to the short-channel effect [25], and the S-functions must be shifted with respect to each other before the ratio is taken.

$$r(\delta, V_g) \equiv \frac{S^0(V_g)}{S^1(V_g - \delta)} \tag{5.19}$$

The purpose is to find the δ-value for which the ratio r is a constant, independent of V_g. The δ is the threshold voltage difference between the two devices.

The procedure is automated by computing the average r and the mean square deviation of r from its average value, i.e.

$$\langle r \rangle = \frac{\int_{\Delta V_g} r(\delta, V_g) dV_g}{\int_{\Delta V_g} dV_g} \tag{5.20}$$

$$\langle \sigma^2 \rangle = \langle r^2 \rangle - \langle r \rangle^2 \tag{5.21}$$

Once the correct shift is found, r is evaluated from equation (5.19) and one example is given in figure 5.8 (a). ΔL and R_{sd} can be then obtained by using

equation (5.22) and (5.23).

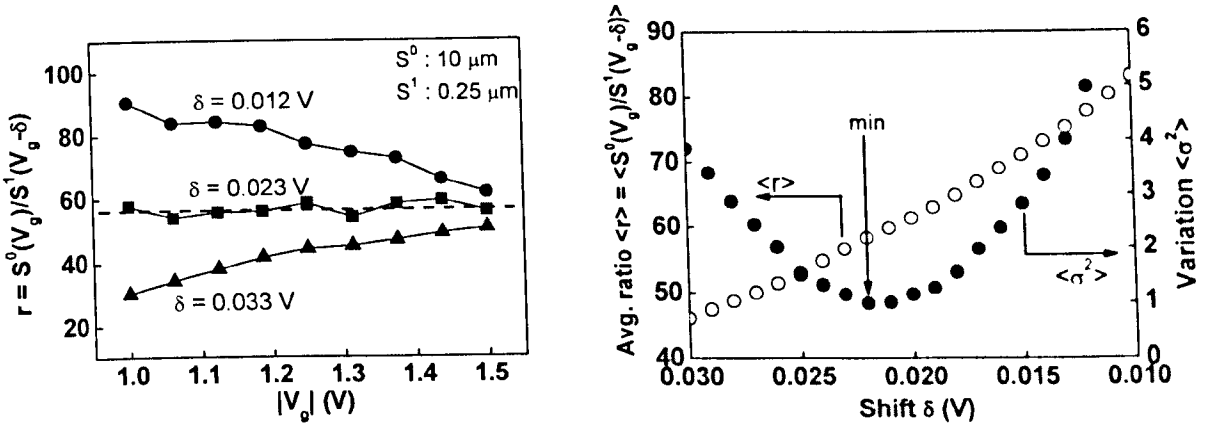
$$r|_{\delta} = \frac{L^0_{eff}}{L^1_{eff}} = \frac{L^0_{mask} - \Delta L}{L^1_{mask} - \Delta L} \tag{5.22}$$

And

$$Rsd = \frac{r|_{\delta} \cdot R^1_{tot} - R^0_{tot}}{r|_{\delta} - 1} \tag{5.23}$$

The correct δ results in the minimum $< \sigma^2 >$, as illustrated by figure 5.8 (b).

For the data in figure 5.8, the extracted R_{sd} = 76 Ω and the ΔL = 0.08 μm.



(a) (b)

Fig 5.8 (a) The dependence of the ratio of S-function for three different amounts of shift, δ. Only when the correct shift is found, r becomes independent on V_g. (b) Demonstration of extracting series resistance (R_{sd}) and effective channel length by using Shift and Ratio (S&R) method. The longer and shorter channel is 10 μm and 0.25 μm respectively. The channel width is 10 μm. Average raito <r> and variance $<\sigma^2>$ depends on the gate voltage shift, δ.

By using the same S&R procedure, the series resistance and channel length offset are extracted for four samples we used in this work and the results are given in table 5.2.

Table 5.2 Extracted R_{sd} and ΔL_{eff} by using S&R method

Sample	R_{sd} (Ω)	ΔL_{eff} (μm)
A	64	0.07
B	76	0.08
C	57	0.056
D	75	0.013

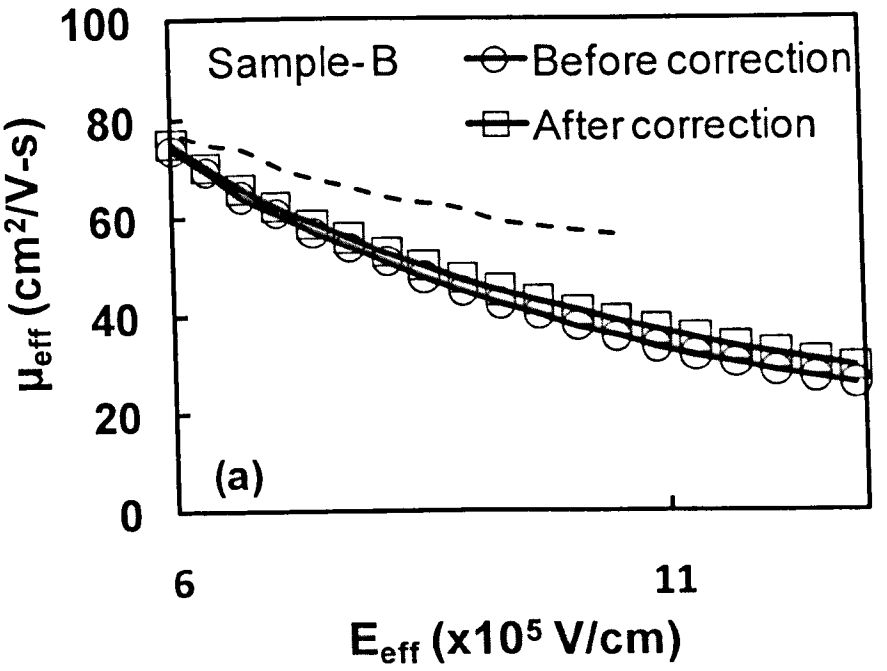
A comparison of the mobility before and after the correction is shown in figure 5.9 (a) and (b). It will be seen that without the correction, the mobility can be under-estimated.

5.5 Effect of remote charges on mobility degradation

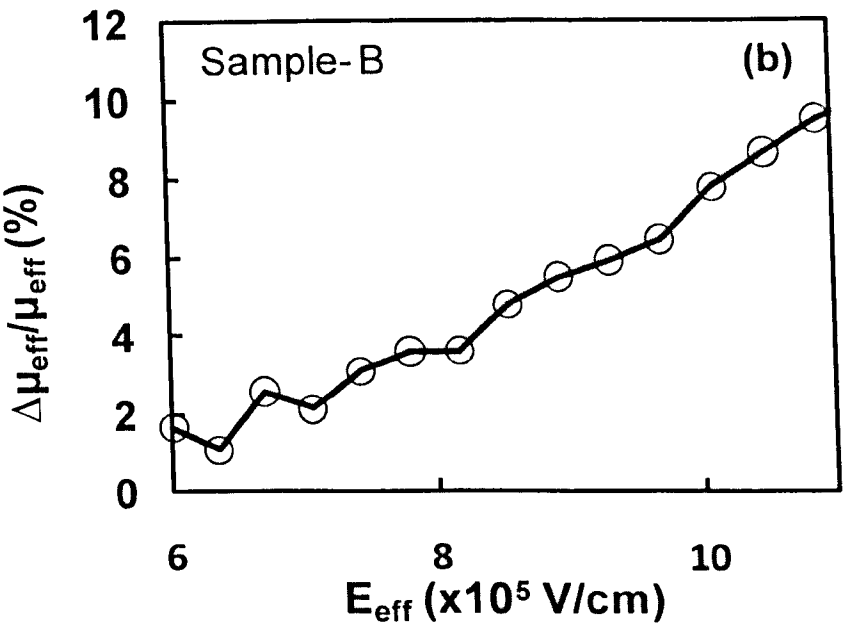
The impact of process-induced positive charge (PIPC) on the effective mobility can now be studied followed by an assessment of the effect of electron trapping.

5.5.1 Impact of PIPC on electron mobility

It has been reported that annealing in forming gas (10% H_2) at 500–550 °C can generate positive charge in the HfO_2 /SiO_2 stack [26,27]. I_d-V_g curves before and after 60 minutes forming gas annealing for sample A are shown in figure 5.10 (a) and the corresponding mobility is shown in figure 5.10 (b).

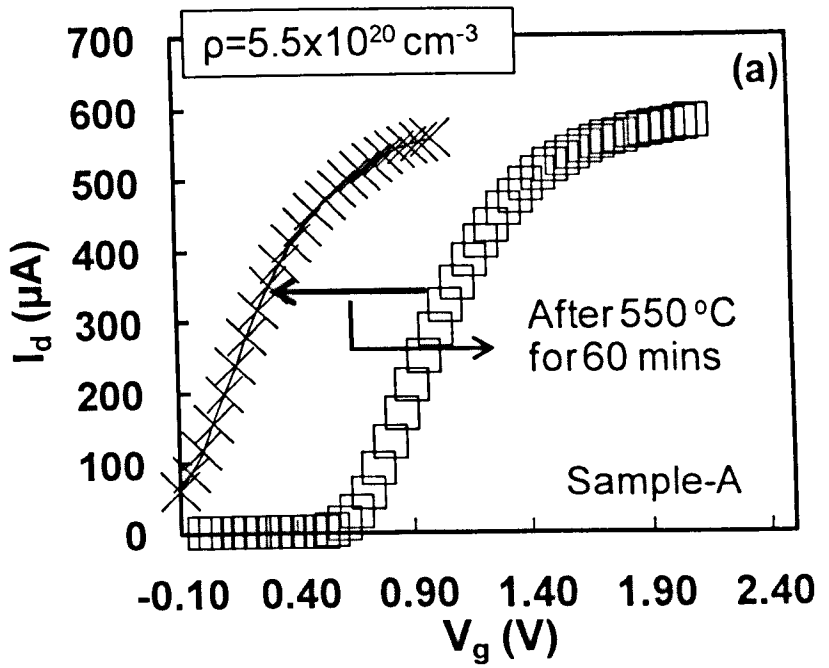


(a)

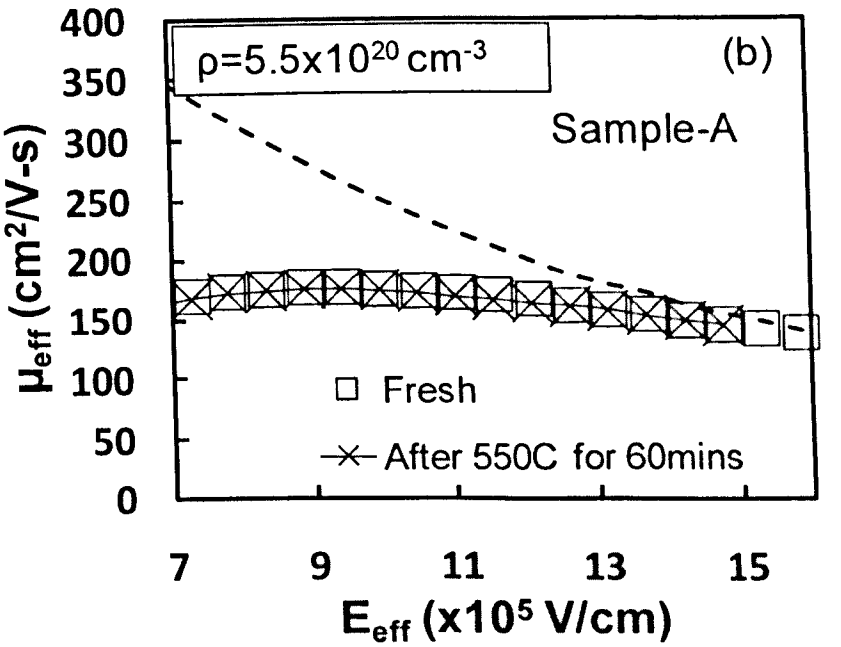


(b)

Fig 5.9 (a) A comparison of mobility before and after the correction for series resistance and channel length offset. The mobility increases after the correction. (b) shows the percentage of the effective mobility due to correction. Measurement without the proper correction can under-estimate mobility by as large as 10%.



(a)



(b)

Fig 5.10 The transfer characteristics of sample A before and after a 60 minutes annealing are shown in (a). A positive charging of $5.5x10^{20}$ cm^{-3} is generated in HfO_2. The effective electron mobility before and after the 60 minutes annealing is shown in (b). The dash line is the universal curve.

Since the detailed spatial distribution of PIPC is not available, its density and

separation from the substrate interface must be estimated.

On one hand, as a "rule-of-the-thumb" estimation of the farthest possible distance from the substrate, it is assumed that all PIPC are uniformly distributed in the HfO_2 layer and there is no PIPC in the interfacial SiO_2. The volume density, ρ, can be determined from [28]

$$\rho = -\frac{2\varepsilon_0 k_{Hf} \Delta V_t}{q T_{Hf}^2} \tag{5.24}$$

where k_{Hf} and T_{Hf} are the dielectric constant and thickness of high-k layer, respectively. By using equation (5.24), a ΔV_t=0.8 V from figure 5.10 (a) gives an area density of 1.1×10^{14} cm^{-2} and a volume density of $\rho = 5.5 \times 10^{20}$ cm^{-3}, which is in the same order as the typical doping density for a modern poly-Si gate [4,6] and is among the highest charging level observed for MOS devices.

On the other hand, as an estimation of the nearest possible distance, we assume that all PIPC are uniformly distributed in the interfacial layer. The volume density, ρ, can be determined from [28]

$$\rho = -\frac{2\varepsilon_0 k_{IL} \Delta V_t}{q(T_{IL}^2 + 2\frac{k_{IL}}{k_{Hf}} \cdot T_{IL} \cdot T_{Hf})} \tag{5.25}$$

where k_{IL} and T_{IL} is the dielectric constant and thickness of interfacial layer, respectively. By using equation (5.25), a ΔV_t=0.8 V from figure 5.10 (a) gives a volume density of $\rho = 5.2 \times 10^{20}$ cm^{-3}, so that the volume density is insensitive

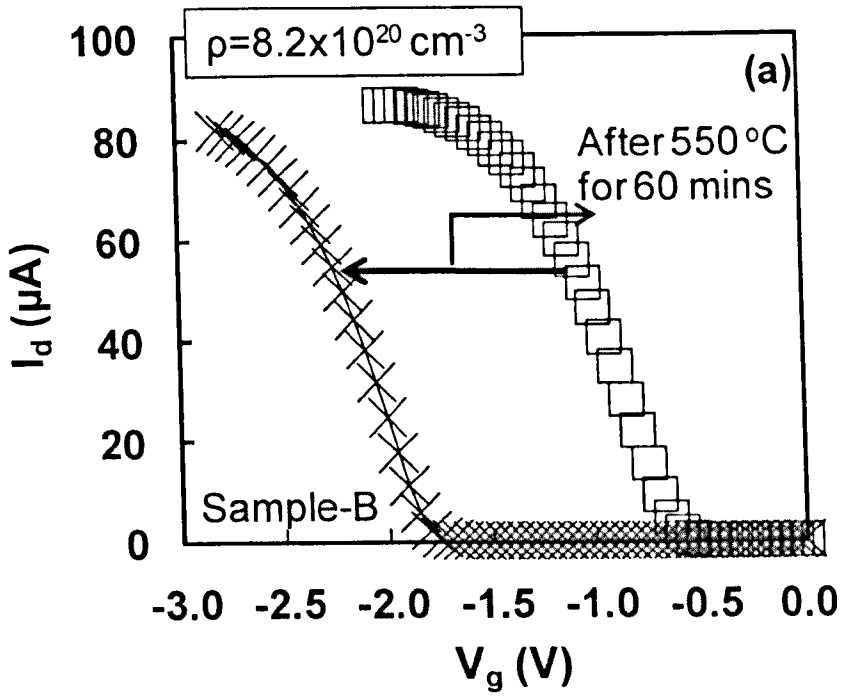
to the assumption. We will use the ρ calculated by using equation (5.24) hereafter. The effective electron mobility before and after PIPC generation is evaluated and figure 5.10 (b) clearly shows that it changes little through the whole range of effective vertical field. We conclude that the RCS induced by a PIPC of $\rho = 5.5 \times 10^{20}$ cm^{-3} at 0.56 nm from the substrate has an insignificant effect on electron mobility.

5.5.2 Impact of PIPC on hole mobility

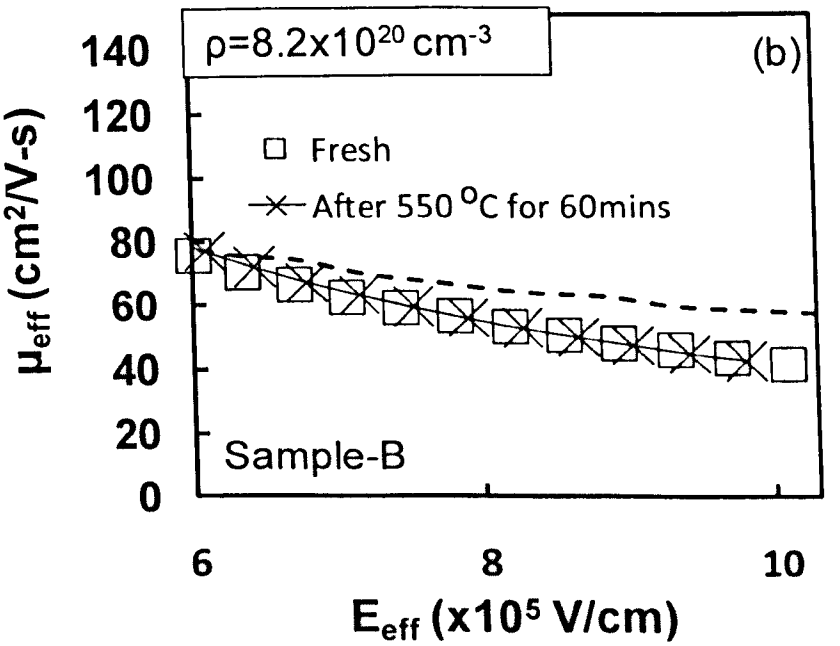
To test effects of PIPC on hole mobility, the sample B is used. After 60 minutes exposure to forming gas, the threshold voltage shift caused by PIPC can reach 1.2 V, as shown in figure 5.11 (a) and the corresponding volume density can reach $\rho = 8.2 \times 10^{20}$ cm^{-3} at 0.73 nm away from the substrate interface. Figure 5.11 (b) shows that hole mobility changes little after the PIPC formation.

5.5.3 Impact of electron trapping on electron mobility

The PIPC originates from hydrogenous species [26,27] and it may be assumed that its interaction with channel electrons is weaker than the interaction between channel electrons and the ionized impurities in the gate, although there is no experimental evidence for this assumption. To test the sensitivity of RCS to the type of charge, the impact of electron trapping in the dielectric on electron mobility is investigated next.



(a)



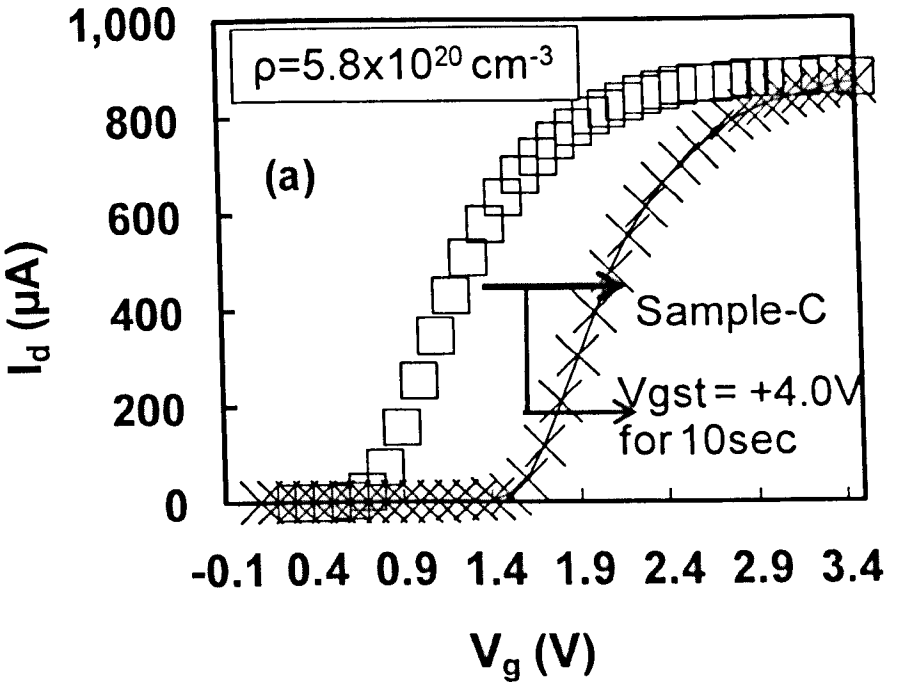
(b)

Fig 5.11 The transfer characteristics before and after 60 minutes annealing of sample B are shown in (a). The effective hole mobility before and after the 60 minutes annealing is shown in (b). The dash line is the universal curve.

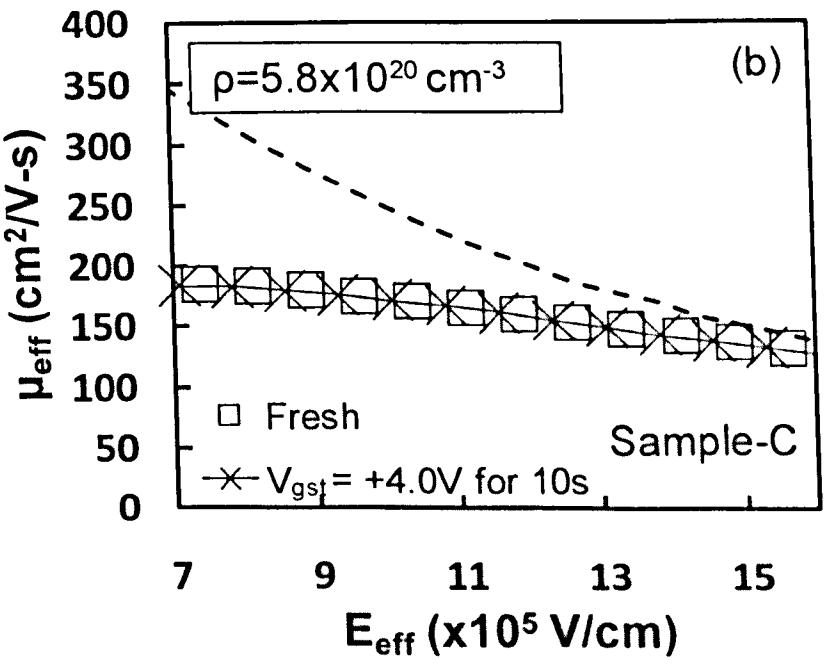
Although pre-existing electron traps in SiON are negligible [29], their volume

density can reach the order of 10^{20} cm^{-3} in a 4 nm HfO_2 prepared by atomic layer deposition (ALD) [17]. The problem is that the electron trapping decreases sharply as the HfO_2 becomes thinner and becomes insignificant for a 2 nm ALD HfO_2 [3,17]. Sufficient electron trapping within 2 nm from the substrate interface cannot be obtained by using ALD HfO_2, therefore. The RCS influence decays exponentially with distance [4,6] and it is essential to have charges within 2 nm, where RCS was reported to be effective [2-4]. This means that we cannot use ALD HfO_2.

To overcome the above difficulty, a 2 nm PVD HfO_2 (sample C) is prepared. Figure 5.12 (a) shows that electron trapping leads to a substantial positive shift of I_d-V_g. Since there is little pre-existing electron traps in the interfacial SiO_2 [17,29], a uniform distribution of electron traps in the HfO_2 is assumed. A ΔV_t=0.84 V leads to a volume density of $\rho = 5.8 \times 10^{20}$ cm^{-3}. Figure 5.12 (b) clearly shows that electron mobility changes little after an electron trapping of $\rho = 5.8 \times 10^{20}$ cm^{-3} at 1 nm from the substrate interface. As a result, changing the type of charges has not changed the conclusion.



(a)



(b)

Fig 5.12 The sample C was stressed at +4.0V for 10 s and an electron trapping of 5.8x10^{20} cm^{-3} are formed. The transfer characteristics before and after electron trapping are shown in (a). The effective electron mobility before and after electron trapping is shown in (b). The dash line is the universal curve.

5.5.4 Impact of gate material on mobility

The samples used up to here have a TaN/TiN gate and the poly-Si gated sample is now tested. After forming gas annealing for 120 minutes, it can be seen from figure 5.13 (a), threshold voltage shift reached 0.34V and volume density of PIPC up to $\rho = 2.7 \times 10^{20}$ cm^{-3} were generated at 0.77 nm from the substrate interface. Little change in the effective mobility can be observed in figure 5.13 (b). Our conclusion holds for devices with either metal or poly gate.

5.5.5 Discussion

The slight degradation observed in these samples may be attributed to the large amount of charge pre-existing in the samples. If this is the case, the mobility in these samples should have already been heavily degraded. When more charge was added by forming gas annealing or trapping, further degradation could be much smaller. In order to test this possibility, the mobility of these samples is compared with those reported from earlier work for similar dielectric stacks and gate materials:

- The sample B: figure 5.14 (a) shows that the hole mobility is similar to that reported by the group at IMEP/CEA-LETI [30].
- The sample D: figure 5.14 (b) shows that the electron mobility is similar to that reported by the group at Intel [9].

- The samples A and C: figure 5.14 (c) shows that the electron mobility is similar to that reported by the group at Sematech [31] and figure 5.14 (d) shows that the electron mobility is similar to that reported by the group at IMEP/CEA-LETI [30].

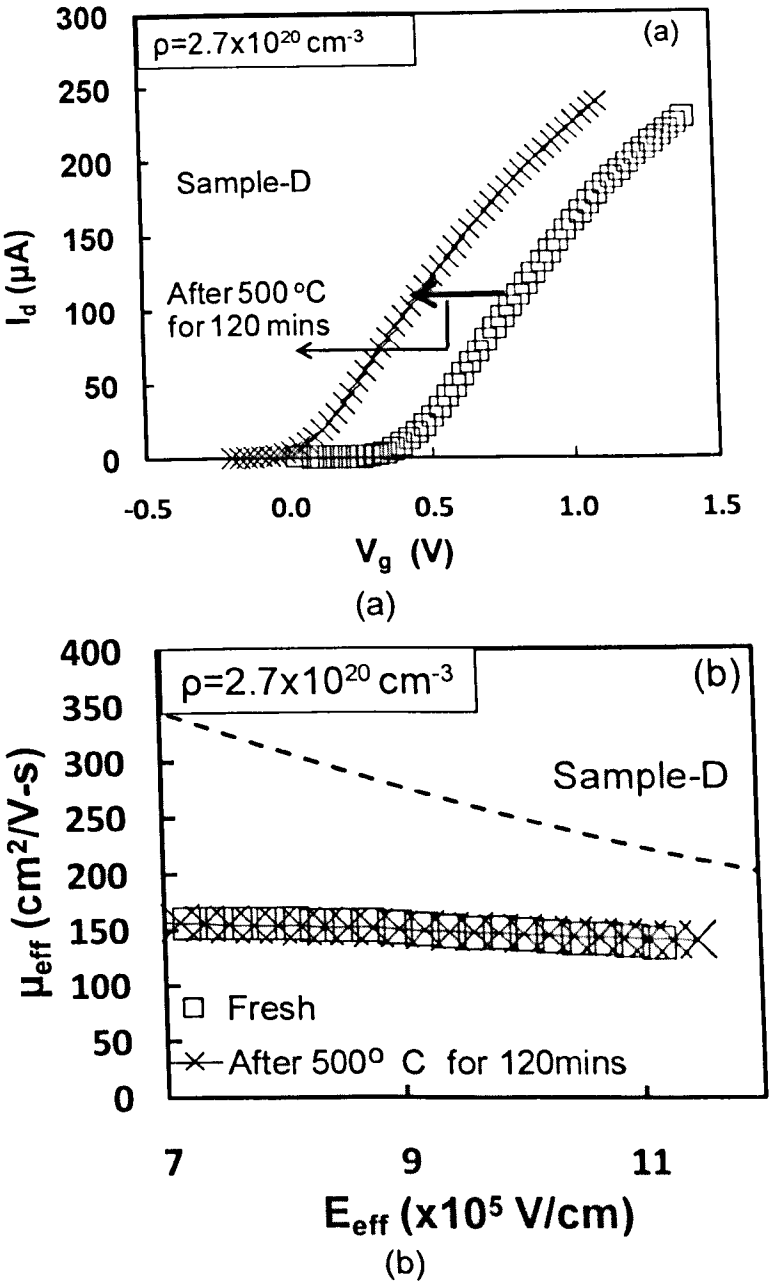
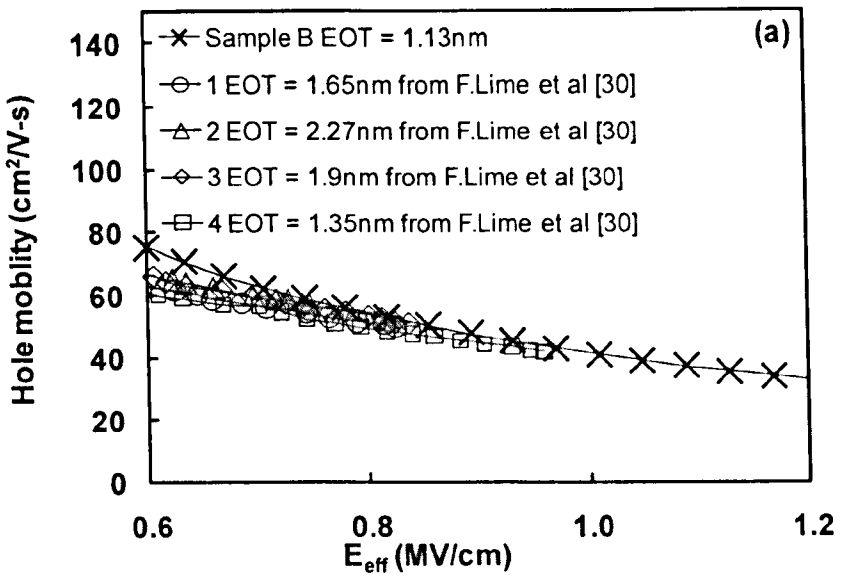
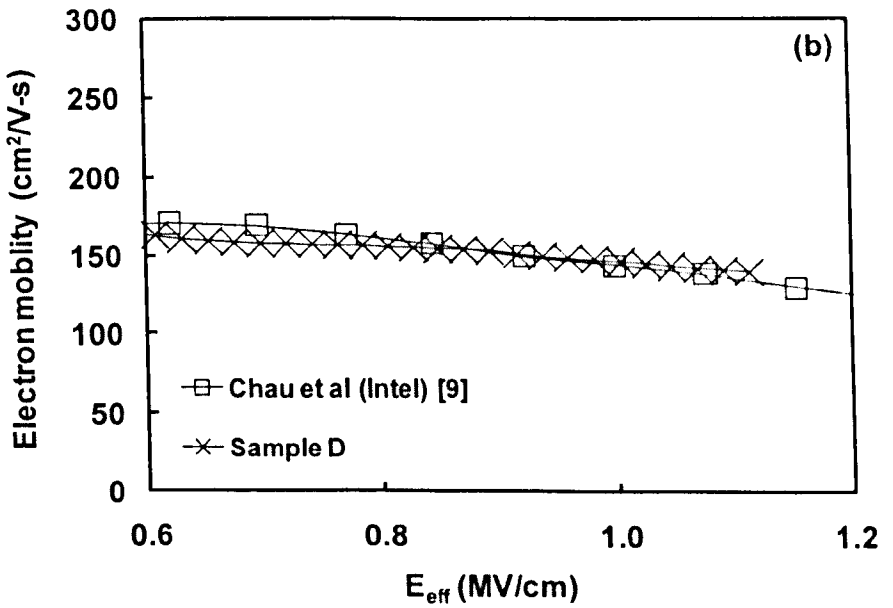


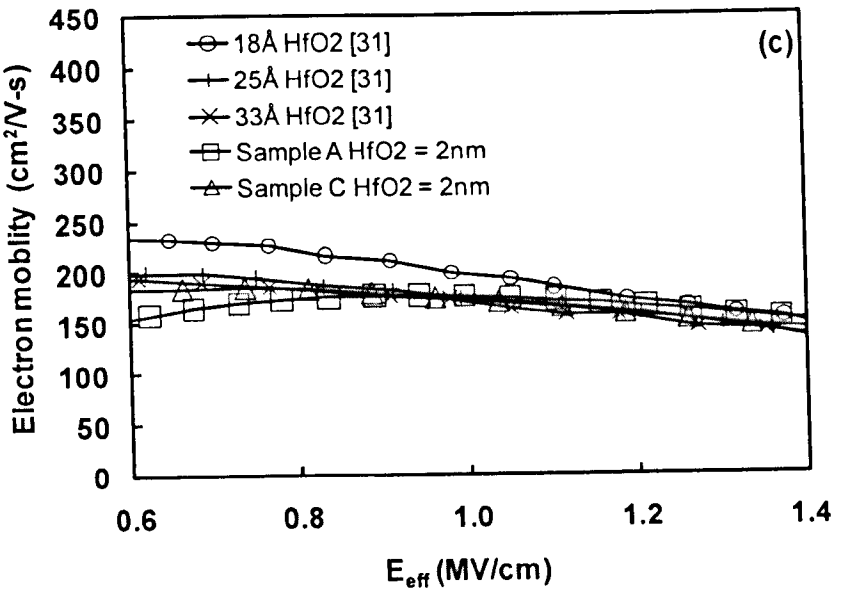
Fig 5.13 The transfer characteristics before and after 120 minutes annealing of the sample D are shown in (a). A PIPC of $2.7x10^{20}$ cm^{-3} was generated. The effective electron mobility before and after the 60 minutes annealing is shown in (b). The dash line is the universal curve.


(a)
Sample B EOT = 1.13nm
1 EOT = 1.65nm from F.Lime et al [30]
2 EOT = 2.27nm from F.Lime et al [30]
3 EOT = 1.9nm from F.Lime et al [30]
4 EOT = 1.35nm from F.Lime et al [30]
Hole mobility (cm²/V-s)
140
120
100
80
60
40
20
0
0.6
0.8
1.0
1.2
E_eff (MV/cm)

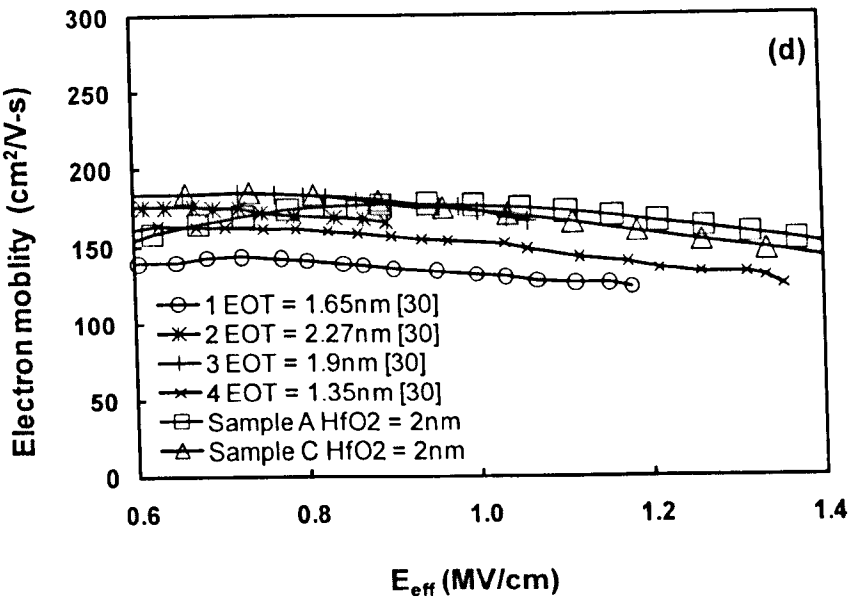
(a)


(b)
Electron mobility (cm²/V-s)
300
250
200
150
100
50
0
Chau et al (Intel) [9]
Sample D
0.6
0.8
1.0
1.2
E_eff (MV/cm)

(b)



(c)



(d)

Fig 5.14 (a) A comparison of the hole mobility in our sample B with that reported by the group at IMEP and CEA-LETI [30]. (b) A comparison of electron mobility in our sample D of a poly-Si gate with that reported by the group at Intel [9]. (c) A comparison of the electron mobility in our sample A and C of a TiN gate with that reported by the group at SEMATECH [31]. (d) A comparison of the electron mobility in our sample A and C with that reported by the group at IMEP and CEA-LETI [30].

As a result, any pre-existing charge in these samples is not excessive and is similar to those in earlier work [9,30,31]. To quantify these pre-existing charges,

we compare the measured threshold voltage with its theoretical value next.

Figure 5.15 compares the measured fresh threshold voltage with its theoretical value. Two values agree quite well, indicating insignificant pre-existing charges in the oxides.

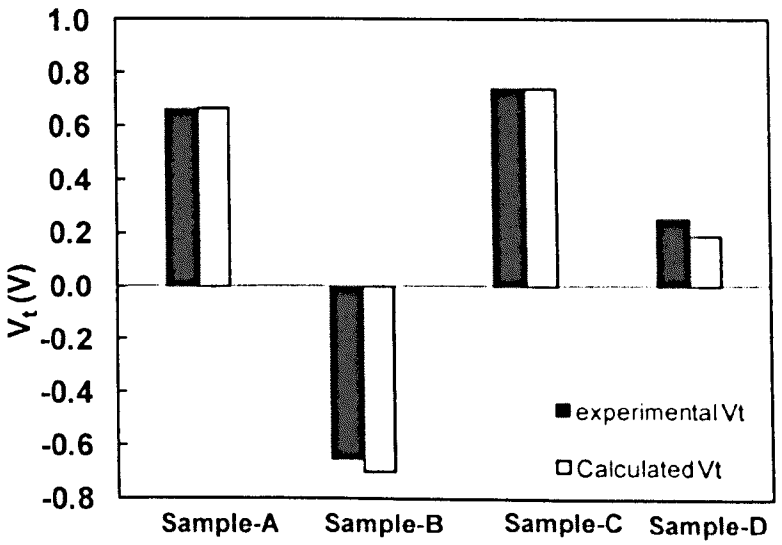


Fig 5.15 A comparison of the measured threshold voltage for fresh devices with their theoretical values. The theoretical value is calculated by using equation (5.26). The experimental value is extracted by using constant current method at the level of $1x10^{-6}$*W/L.

To explore why RCS is not as important as previously thought, it is noted that the strength of RCS is sensitive to the assumptions used in the model [33,34]. For example, the inclusion of gate free carrier screening substantially reduces the effect of RCS on channel mobility [33,34] and it can become negligible when the free carrier density in the gate is sufficiently high [33].

It should be clarified that this work does not rule out that remote charge can degrade mobility in principle. The degradation was not observed for a charge

density up to the order of 10^{20} cm^{-3} at a location of 0.56~1 nm from the substrate interface. If the remote charge is closer to the substrate or have a higher density, it may degrade the mobility.

5.6 Summary

In this chapter, the focus has been on investigating the impact of remote charge scattering on mobility degradation. In order to fix the charging during the measurement, a new measurement setup was developed for the pulse C-V technique with the capability to measure C-V curve up to 600 kV/s ramp rate. For a 10x10 μm^2 MOSFET, this produces a displacement current large enough for an accurate measurement. To evaluate mobility, corrections must be made for gate leakage current, series resistance and the channel length offset.

Unlike earlier work where different samples were used for investigating the impact of remote charge scattering on mobility, the same sample is used in this work. The charge is introduced either by forming gas annealing or trapping, and the volume density can reach the order of 10^{20} cm^{-3} at 0.56nm -1nm from the substrate interface. These experiments clearly showed that these remote charges have little effect on the channel mobility, although it is not ruled out that RCS can degrade mobility in principle.

References

1. M. S. Liang, J. Y. Choi, P. K. Ko, and C. Hu, "Inversion-layer capacitance and mobility of very thin gate-oxide MOSFET's," *IEEE Trans. Electron Dev.*, vol. 33, pp. 409-413, 1986.
2. M. S. Krishnan, Y. C. Yeo, Q. Lu, T. J. King, J. Bokor, and C. Hu, "Remote charge scattering in MOSFETs with ultra-thin gate dielectrics," in *International Electron Devices Meeting (IEDM) Digest*, pp. 571-574, 1998.
3. M. Casse, L. Thevenod, B. Guillaumot, L. Tosti, F. Martin, J. Mitard, O. Weber, F. Andrieu, T. Ernst, G. Reimbold, T. Billon, M. Mouis, and F. Boulanger, "Carrier transport in HfO2/metal gate MOSFETs: physical insight into critical parameters" *IEEE Trans. Electron Dev.*, vol. 53, pp. 759-768, 2006.
4. N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman, "Estimation of the effects of remote charge scattering on electron mobility of n-MOSFET's with ultrathin gate oxides," *IEEE Trans. Electron Dev.*, vol. 47, pp. 440-447, 2000.
5. Q. Xiang, G. Yeap, D. Bang, M. Song, K. Ahmed, E. Ibok, and M. R. Lin, "Performance and reliability of sub-100nm MOSFETs with ultra thin direct tunneling gate oxides," in *Symposia on VLSI Technology and Circuits*, pp. 160-161, 1998.
6. D. Esseni and A. Abramo, "Modeling of electron mobility degradation by remote Coulomb scattering in ultrathin oxide MOSFETs," *IEEE Trans. Electron Dev.*, vol. 50, pp. 1665-1674, 2003.
7. M. V. Fischetti, "Long-range Coulomb interactions in small Si devices. Part II. Effective electron mobility in thin-oxide structures," *J. Appl. Phys.*, vol. 89, pp. 1232-1250, 2001.
8. M. V. Fischetti, D. A. Neumayer, E. A. Cartier, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering," *J. Appl. Phys.*, vol. 90, pp. 4587-4608, 2001.
9. R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz, "High-k/Metal-gate stack and its MOSFET characteristics," *IEEE Electron Dev. Lett.*, vol. 25, pp. 408-410, 2004.
10. S. I. Saito, K. Torii, Y. Shimmamoto, S. Tsujikawa, H. Hamamura, O. Tonomura, T. Mine, D. Hisamoto, T. Onai, J. Yugami, M. Hiratani, and S. Kimura, "Effects of remote-surface scattering on carrier mobility in field-effect-transistors with ultrathin gate dielectrics," *Appl. Phys. Lett.*, vol. 84, pp. 1395-1397, 2004.
11. *Physics of Semiconductor devices*, S. M. Sze, 3rd edition, John Wiley & Sons Inc, 2007.
12. A. O. Conde, F.J. G. Sanchez, J. J. Liou, A. Cerdeira, M. Estrada, Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectronics reliability*, vol. 42, pp. 583-596, 2002.

13. C. G. Sodini, T. W. Ekstedt and J. L. Moll, "Charge accumulation and mobility in thin dielectric MOS transistors," *Solid State electron.* Vol. 25 pp. 833-841, 1982.
14. *Semiconductor material and device characterization*, D. K. Schroder, 2nd Edition, John Wiley & Sons inc, 1998.
15. F. Babarada, M. D. Profirescu and A. Rusu, "MOSFET mobility degradation modeling," in *international Semiconductor Conference (CAS)*, pp. 304-307, 2003.
16. C. Z. Zhao, J. F. Zhang, M. B. Zahid, B. Govoreanu, G. Groeseneken, and S. D. Gendt, "Determination of capture cross sections for as-grown electron traps in HfO_2/HfSiO stacks," *J. Appl. Phys.*, vol. 100, pp.349-359, 2006.
17. J. F. Zhang, C. Z. Zhao, M. B. Zahid, G. Groeseneken, R. Degraeve, and S. D. Gendt, "An assessment of the location of as-grown electron traps in HfO_2/HiSiO stacks," *IEEE Electron Dev. Lett.*, vol. 27, pp. 817-820, 2006.
18. D. V. Singh, P. Solomon, E. P. Gusev, G. Singco and Z. Ren, "Ultra-fast measurements of the inversion charge in MOSFETs and impact on measured mobility in high-k MOSFETs," in *International Electron Devices Meeting (IEDM) Digest,* pp. 863-866, 2004.
19. W. Zhu, J. P. Han, and T. P. Ma, "Mobility measurement and degradation mechanisms of MOSFETs made with ultrathin high-k dielectrics," *IEEE Trans. Electron Dev.*, vol. 51, pp. 98-104, 2004.
20. C. Shen, M. -F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra, and Y. -C. Yeo, "Characterization and physical origin of fast Vth transient in NBTI of pMOSFETs with SiON dielectric," in *International Electron Devices Meeting (IEDM) Digest*, pp. 333-336, 2006.
21. D. Deleruyelle, J. Cluzel, B. D. Salvo, D. Fraboulet, D. Mariolle, N. Buffet and S. Deleonibus, "Electrical characterization of fast transient phenomena in a Si-rich based non-volatile random access memory," *Solid State Electron.*, pp. 1641-1644, 2003
22. Z. Xu, B. Kaczer, J. Johnson, D. Wouters and G. Groeseneken, "The charge trapping in the metal-ferroelectric-insulator-semiconductor structure with the SrBi2Ta2o9 /Al2O3 /SiO2 stack," *J. Appl. Phys.*, vol. 96, pp.1614-1619, 2004.
23. J. Schmitz, M. H. H. Weusthof and A. J. Hof, "Leakage current correction in quasi-static C-V measurements," *in IEEE Conference on Microelectronic Test Structure (ICMTS),* pp. 22-25, 2004.
24. T. Yuan, D. S. Zicherman, D. R. Lombardi, J. R. Phillip, C. H. Hsu, I. H. Hussein, W. R. Matthew, D. Bijan and G. S. Ghavam, "A new shift and ratio method for MOSFET channel-length extraction," *IEEE Electron Dev. Lett.*, vol. 13, pp. 267-269, 1992.
25. *Fundamentals of modern VLSI devices,* T. Yuan and T. H. Ning, pp. 206-211, Cambridge University Press 1998.
26. C. Z. Zhao, J. F. Zhang, M. H. Chang, A. R. Peaker, S. Hall, G. Groeseneken, L. Pantisano, S. D. Gendt, and M. Heyns, "Process-induced positive charges in Hf-based gate stacks," *J. Appl. Phys.*, vol. 103, pp.1241-1245, 2008.

27. M. H. Chang, C. Z. Zhao, Z. Ji, J. F. Zhang, G. Groeseneken, L. Pantisano, S. De Gendt, and M. M. Heyns, "On the activation and passivation of precursors for process-induced positive charges in Hf-dielectric stacks," *J. Appl. Phys.*, vol. 105, pp. 1679-1684, 2009.
28. J. F. Zhang, M. H. Chang, J. Zhigang, L. Lin, I. Ferain, G. Groeseneken, L. Pantisano, S. D. Gendt and M. M. Heyns, "Dominant layer for stress-induced positive charges in Hf-based gate stacks," *IEEE Electron Dev. Lett.*, vol. 29, pp. 1-4, 2008.
29. M. H. Chang, J. F. Zhang and W. D. Zhang, "An assessment of capture cross sections and effective density of electron traps generated in silicon dioxides," *IEEE Trans. Electron Dev.*, vol. 53, pp. 1347-1354, 2006.
30. F. Lime, K. Oshima, M. Casse, G. Ghibaudo, S. Cristoloveanu, B. Guillaumot and H. Iwai, "Carrier mobility in advanced CMOS devices with metal gate and HfO2 gate dielectric," *Solid-State Electron*, vol. 47, pp.1617-1621, 2003.
31. 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 HfO2 thickness on charge trapping and mobility," *Solid-State Electron*, vol.80, pp.218-221, 2005.
32. H. Y. Y. Ren, Y.-C.Yee, J. F. Kang, X. P. Wang, H. H. H. Ma, M. F. Li, D. S. H. Chan, D. -L, Kwong, "Fermi Pinning-induced thermal instability of metal-gate work function", *IEEE Electron Dev. Lett.*, vol. 25, pp. 337-339, 2004
33. F. Gamiz and M. V. Fischetti, "One type of the remote Coulomb scattering in the metal–oxide–semiconductor field effect transistors: Screening by electrons in the gate," *Appl. Phys. Lett.*, vol. 83, pp. 4848-4850, 2003.
34. T. Ishihara, J. Koga, K. Matsuzawa and S. Takagi, " Modeling of screening effect on remote Coulomb scattering due to gate impurities by nonuniform free carriers in poly-Si gate," *J. Appl. Phys.*, vol. 102, pp. 5329-5334, 2007.

6 Conclusions and Future work

6.1 Conclusions

This project is divided into three parts: the reconciliation of characterization techniques for NBTI, the worst case NBTI lifetime prediction, and the impact of remote charge scattering on mobility degradation. Conclusions for each part are given below.

6.1.1 Conclusions on the real threshold voltage instability under operational bias

The focus was on bridging the gap in threshold voltage shift evaluated by different techniques after NBTI stress. Firstly, the NBTI-induced threshold voltage shift evaluated by different techniques was compared, including the on-the-fly technique sensed at stress V_{gst} (OTF-V_{gst}), ultra-fast pulse technique with extrapolation (UFP-ex), and the conventional quasi-DC measurement with extrapolation (DC-ex). The impact of various factors reported recently on ΔV_t was examined one by one. These include the degradation during measurement that affects the reference I_d and initial transconductance, the recovery during measurement, and the evaluation of transconductance.

After clarifying these issues, it is found that the gap in ΔV_t extracted from different techniques cannot be bridged and it is essential to consider the effect of sensing V_g on ΔV_t. The evaluation of ΔV_t at different sensing V_g was achieved by calculating the ΔI_d and transconductance at a given V_g and the classical transconductance $g_m=dI_d/dV_g$ should be replaced by $G_m=dI_d/d(V_g-V_t)$. The results clearly show that $|\Delta V_t|$ increases with sensing $|V_g|$ because of more positive charge trapping at higher $|V_g|$ and the assumption of ΔV_t being independent of sensing V_g was shown to be invalid. After taking both recovery and the sensing V_g effect into account, the large gap in ΔV_t obtained from different techniques in earlier work has been bridged successfully. The results show that recovery reported by earlier work and the sensing V_g effect observed here are two different phenomena.

6.1.2 Conclusions on NBTI lifetime prediction and kinetics at operational bias

The objective was to predict the worst case NBTI lifetime under operational bias. It is found that when the recovery is suppressed by using ultra-fast pulse measurement and ΔV_t is sensed at the operational gate bias, the conventional V_g acceleration prediction is not applicable. This is because the NBTI kinetics no longer follows a simple power law and an increase of stress bias does not lead to a parallel shift of $\log|\Delta V_t|$.

To predict the worst case lifetime at the operational gate bias, NBTI kinetics and

defects were 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_t can be modelled 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 earlier work.

6.1.3 Conclusions on the impact of remote charge scattering on mobility degradation

In this part, the impact of remote charge scattering on mobility degradation is investigated. In order to freeze the charging during the measurement, a new setup for the pulse C-V technique was developed, To evaluate mobility, corrections were made for gate leakage current, series resistance and the effective channel length.

Unlike earlier work where different samples were used for investigating the impact of remote charge scattering on mobility, the same sample was used in our work. The charge is introduced either by forming gas annealing or trapping, and the volume density can reach the order of 10^{20} cm^{-3} at 0.56 – 1 nm from the

substrate interface. The experiments clearly showed that these remote charges have little effect on either electron or hole mobility.

6.2 Future work

Despite the progress made in this project, there are many problems remaining to be solved, including but not limited to, the following:

Further improvement of the ultra-fast I_d-V_g technique

This project relied heavily on the ultra-fast I_d-V_g technique implemented in chapter 2, which, to the best knowledge of this author, has reached the highest accuracy compared with the implementation by other groups. However, the speed is limited to 5 μs at present. This is because the op-amplifier used only has 60 MHz bandwidth, which limited the measurement speed. The difficulty for using higher bandwidth op-amp is these high bandwidth op-amplifiers are usually sensitive to parasitic components in the circuit, and therefore, they need to be mounted by using the surface mount technology on double layer PCB.

The defects responsible for the sensing Vg effect

This work clearly shows that threshold voltage shift after NBTI stress is dependent on sensing V_g. However, the understanding of the defects responsible for the sensing V_g effect is poor. For example, it is not clear whether

they are hole traps close to the interface or interface states. The generation mechanism of these defects is not known either. These issues should be addressed in future work.

NBTI lifetime prediction and kinetics at operational bias

In this work, the proposed kinetics can predict NBTI lifetime at the operational gate bias in the worst case scenario where the recovery is suppressed. However, this lifetime prediction is based on a single device. In the real world, one functional circuit can include millions of these transistors and only the lifetime of the functional circuit is meaningful to the manufacturer and the customer. Therefore, not only the worst case lifetime is needed, the impact of recovery should be included in the kinetics model as well. Furthermore, it could be useful if the model can be verified by implementing it into a circuit simulator.

Impact of remote charge scattering on mobility degradation

This work clearly shows that the remote charges have little effect on the carrier mobility for the device studied. However, the reason for the observed mobility degradation when the gate oxide becomes thinner is still not clear and needs further work.

List of publications

Journal:

1. **Z. Ji**, L. Lin, J. F. Zhang, B. Kaczer, and G. Groeseneken, "NBTI lifetime prediction and kinetics at operation bias based on ultrafast measurement," *IEEE Trans. Electron Dev.*, vol. 57, pp. 228-237, 2010.
2. **Z. Ji**, J. F. Zhang, W. D. Zhang, G. Groeseneken, L. Pantisano, S. D. Gendt, and M. M. Heyns, "An assessment of the mobility degradation induced by remote charge scattering," *Appl. Phys. Lett.*, vol. 95, Art.no. 263502, 2009.
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, pp. 1086-1093, 2009.
4. M. H. Chang, C. Z. Zhao, **Z. Ji**, J. F. Zhang, G. Groeseneken, L. Pantisano, S. D. Gendt, and M. M. Heyns, "An assessment of the mobility degradation induced by remote charge scattering," *J. Appl. Phys.*, vol. 105, Art.no. 054505, 2009.
5. J. F. Zhang, M. H. Chang, **Z. Ji**, L. Lin, I. Ferain, G. Groeseneken, L. Pantisano, S. D. Gendt and M. M. Heyns, "Dominant layer for stress-induced positive charges in Hf-Based gate stacks," *IEEE Electron Dev. Lett.*, vol. 29, pp. 1-4, 2008.

Conference:

1. **Z. Ji**, L. Lin, and J. F. Zhang, "Impact of sensing gate bias on NBTI of Hf-based dielectric stacks," in *215th Electrochemical Conference Society Meeting (ECS)*, San Francisco, CA, 2009.
2. J. F. Zhang, M. H. Chang, **Z. Ji**, and W. D. Zhang, "Recent progress in understanding the instability and defects in gate dielectrics," in *9th International Conference on Solid-State and Integrated-Circuit Technology (ICSICT)*, 2008.
3. J. F. Zhang, **Z. Ji**, M. H. Chang, B. Kaczer, and G. Groeseneken, "Real Vth instability of pMOSFETs under practical operation conditions," in *International Electron Devices Meeting (IEDM) Digest*, Washington DC, pp. 817-820, 2007.