The Incremental Frequency Charge Pumping Method: Extending the CMOS Ultra-Thin Gate Oxide Measurement Down to 1nm

***Steve S. Chung**, NCTU Chair Professor*

Department of Electronic Engineering, National Chiao Tung University, Taiwan
TEL: 886-3-5731830 Fax: 886-3-5734608 email: schung@cc.nctu.edu.tw

Abstract- Interface characterization is fundamental to the understanding of device reliability as well as the gate oxide process monitoring, in particular for the development of an efficient tool for analyzing the hot carrier reliability of state-of-the-art CMOS devices. This paper will give an overview of more advanced charge pumping technique for extending the reliability characterization of ultra-thin gate oxide CMOS in the direct tunneling regime which conventional CV measurement can not meet. This talk will address the basics of charge pumping (CP) method, an advanced technique called IFCP(Incremental Frequency CP), and the applications to the device interface characterization as well as CMOS reliabilities. Its potential use for the device reliability study, such as hot-carrier, NBTI, and oxide process quality monitoring for the advanced CMOS technology will be presented. More recent developments for sub-100nm strained-silicon CMOS device applications will also be demonstrated.

1. Introduction

As indicated from the SIA roadmap, scaling of sub-100nm device in 2005 used a t_{ox} in the range 10-15 Å. As gate oxide reduces to the direct tunneling regime, the leakage current through the gate oxide will induce measurement error. Therefore, reliability measurement limitation becomes a critical issue for the gate oxide as thin as 20Å-30 Å and below. The characterization of interface traps (N_{it}) and oxide trapped charges (Q_{ot}) at the oxide interface can provide sufficient information for designing reliable CMOS devices.

On the other hand, it is also well known that two pronounced effects occur as a result of the gate oxide scaling below 20A-30 Å, i.e., direct tunneling gate leakage and the quantum mechanical effect [1-2]. This makes the device characterization more difficult. Although conventional CV method [3] has been used for almost over 4 decades, it requires a large area capacitor for the measurement and has faced difficulties when gate leakage current exists. In recent years, several other gate oxide reliability analysis techniques, such as CP (Charge- Pumping) [4-5], GD (Gated-Diode) [6] etc. have been elaborated for such a purpose. Until recently, the CP method on a thin gate oxide measurement has aroused much interest [7-8]. Basically, the above methods can all be employed to determine the interface traps in a CMOS device. However, for ultra-thin gate oxide below 20 Å, the aforementioned methods are limited by the tunneling current during the measurement.

In this talk, first, an overview of more advanced charge pumping(CP) technique for the interface characterization of CMOS reliabilities will be introduced. Its potential use for the device reliability study, and process monitoring for the state-of-the-art CMOS technology will be presented. More recent developments for its applications to the HC reliability and process characterization will then be demonstrated.

2. Recent Development of CP Technique

Basically, the charge pumping technique is mainly used for the profiling of N_{it} and/or Q_{ot} along the channel direction (e.g., spatial distribution)[4,9] such that the device reliability relating to the generation of these interface or oxide traps (e.g., the spatial distribution of N_{it} and Q_{ot}) can be determined. The schematic of a CP method is illustrated in Fig. 1(a), where the source and drain are grounded and by applying a gate pulse at the gate, the device channel will switch between accumulation and inversion, such that a recombination current can be observed from the substrate, called charge pumping current. If the device is operated under high field stress, the N_{it}/Q_{ot} will detect more recombination centers and a larger CP current.

As a result of the gate oxide scaling, an anomalous leakage current will induce measurement error for ultra-thin gate oxide applications. Only until recently, the CP method on a thin gate oxide measurement has been reported [7]. However, this method is in-accurate when $t_{ox} \leq$ 12A. As a consequence, a so-called IFCP(Incremental Frequency CP) technique [8] developed by our group, seems to be one of the most efficient CP approach for measuring ultra-thin gate oxide reliability.

Fig. 1(a) shows the schematic diagram of the popular CP setup and the leakage component (I_G gate-to-bulk leakage) occurred in ultra-thin gate oxide. Fig. 1(b) shows the normal (t_{ox}>30Å) and abnormal (t_{ox}<30Å) CP curves. It is noted that when t_{ox}<30Å, leakage occurs as a result of the tunneling between the gate and bulk. Fig. 2 shows the measured data for a set of ultra-thin gate oxide devices, in which a large current incurred for a very thin (12Å) device. In a newly developed IFCP (Incremental Frequency Charge Pumping) method, we use the approach by measuring CP currents at two different frequencies. For example, first, we measure the

I_{CP} at two different frequencies as in Fig. 3, where f_1=2MHz and f_2=1MHz, from which we take the difference of I_{CP} for these two frequencies. This I_{CP} becomes the I_{CP} value at a new frequency $f_1 - f_2$ = 1Mz. Again in Fig. 4, it was shown the measured CP curves under various frequencies, in which the measured curves can be multiplied by any factor. For example, both $I_{CP,1M\text{-}500KHz}$ and 5/2*$I_{CP,1M\text{-}800KHz}$ can be regarded as the I_{CP} at 500KHz.

3. Process Monitors by the Charge Pumping Method

A. Determination of the Effective Channel Length

The following equation is used for the extraction of interface traps and effective channel length L_{eff}, (Fig. 5) i.e.,

$$I_{CP,max} = f\, q W\, (L_{eff})\, N_{it} = f\, q W\, (L\text{-}\Delta L_0)\, N_{it}. \qquad (1)$$

Since $I_{cp,\,max}$ is proportional to the generated interface traps, N_{it}, as source and drain touch with each other, it is assumed that there is no N_{it} and $I_{cp,\,max}$ is zero. Therefore, we can extract ΔL_0 from Fig. 5, where the calculated results of effective channel length is shown. Note that: (1) For various gate lengths, the total lateral diffusion length (physical length) ΔL_0 in nMOSFET (pMOSFET) is equal to 0.08um (0.06um), in this case. (2) This ΔL_0 represents the case when I_{CP}= 0 or the lateral length of S/D junction.

B. Gate Oxide Quality Analysis by IFCP Method

To verify the efficacy of the above IFCP method, it has been applied to evaluate interface traps in RTNO and PRN treated gate oxides. The devices used in this study were fabricated by the state-of-the-art IC manufacturing. 12-16A gate oxides were formed by rapid thermal nitric oxide (RTNO). From the measured CP data (e.g., Fig. 3), we may calculate interface traps for various RTNO and PRN treated gate oxide as shown in Fig. 6. The steeper the slope is, the poorer is the oxide quality. Results show that: (1) RPN treated gate oxide cause larger amount of interface traps than RTNO gate oxide, since the bottom layer of base oxide (RTNO) has re-oxidation during thermal process, as shown in Fig. 7. (2) Oxide quality in pMOSFET's is not as good as that in nMOSFET's, since it generates more interface traps.

4. Applications to the HC and NBTI reliability

In general, the drain current (I_D) degradation and/or threshold voltage (V_t) shift is used as an indicator of the device reliability. The IFCP method, can be used to characterize the hot carrier degradation. Its correlation with the drain current degradation will be presented as follows.

A. Applications to nMOSFET Hot Carrier Studies

The spatial distributions of N_{it} and Q_{ot}, can be described by the following separation method:

1. For a fresh device, the drain current (at V_G= 2V) is measured, curve (1). Again, the device is stressed and its current, curve(2), is measured, Fig. 8.
2. To identify whether Q_{ot} is generated, we monitor the GIDL current as given in Fig. 9, from which we see a shift between curve(1) and curve(2) corresponding to a threshold voltage shift, V_T, caused by the Q_{ot}. A neutralization is performed in two-step, in which curve is moved to curve (3) and then be aligned with the fresh one, curve(1).
3. Simultaneously, I_{CP}'s are measured as in Fig. 10, the difference between curve(1) and (3) gives the value of N_{it}, while the difference between (2) and (3) gives the value of Q_{ot}.
4. A local $V_T(x)$ is calculated, following [4] or [5], and from Eq. (3) in Table II, calculation of N_{it} and Q_{ot} is completed.

Results are shown in Fig. 11, drain junction is located at $\Delta L/2$= 0.02μm. The comparison between $I_{B,max}$ and V_G=V_D stresses are given in Fig. 12, and is consistent with the drain current degradation in Fig. 13, where V_G=V_D stress has larger I_D degradation.

B. Applications to pMOSFET Hot Carrier Studies

In a similar manner, experiments have been performed for pMOSFET, in which a pulse with high state V_{gh} keeping fixed and with a varying pulse low voltage V_{gl} applied at the gate, the charge pumping current can be measured as illustrated. To evaluate the HC reliability, in Fig. 13, we measure the I_D degradation of nMOSFET device after $I_{G,max}$ (V_G= -0.5V, V_D= -2V), $I_{B,max}$ (V_G= -1.2V, V_D= -2V), and V_G= V_D (V_G= V_D= -2V) stress conditions, respectively. The maximum I_D degradation is found at V_G= V_D stress condition. Figs. 13 and 14 show the consistency that V_G=V_D stress exhibits a largest I_D degradation and larger N_{it} distribution comparing to those of $I_{B,max}$ stress. Using the same approach as that in nMOSFET to separate N_{it} and Q_{ot}, we have the lateral profiling of interface traps $N_{it}(x)$ and oxide traps $Q_{ot}(x)$, as shown in Figs. 14 for pMOSFET's under V_G= V_D and $I_{B,max}$ conditions.

C. Applications to pMOSFET NBTI Studies

To explore further merit of the present approach, NBTI induced degradations were studied. There are two different modes of NBTI. For symmetrical NBTI, a negative bias is applied at the gate while S/D and substrate are grounded. Fig. 15 shows the profiling results of N_{it} at 100°C. For asymmetrical NBTI, V_G=V_D stress is applied at 100°C. The measured I_D and GIDL currents and the procedures similar to Fig. 9 should be made. From the results of both Figs. 15 and 16, we can see that N_{it} is largely enhanced on the drain side,

which is attributed to a combination of NBTI and HC effect.

As a conclusion of the observed NBTI effects: (1) the symmetrical NBTI effect will generate a double hump(N_{it}) at the S/D junction region(Fig. 15), (2) N_{it} is greatly enhanced for device under asymmetrical NBTI stress, where both HC and pure NBTI effect can be seen at the drain side(Fig. 16).

5. Applications to Advanced Strained-Si Devices

In more recent years, strained-Si devices have evolved as a potential candidate for high speed and low power logic CMOS technologies beyond 65nm. Depending on device types and structures, these devices exhibit mobility enhancement with a factor of 50% to 100% over that of bulk devices. Our observation indicates that there is a trade-off between device performance and reliability, i.e., a larger enhancement of mobility may adversely degrade the device reliability. In terms of the channel engineering using strained-Si/SiGe, the out-diffusion of Ge as well as the Si/SiGe induced trap generation is unclear. The IFCP can be an effective tool to analyze the enhanced degradation effect generated by the Si/SiGe interface [11].

Fig. 17 shows how to identify the two-level CP curves. Two interface traps(N_{it}) are located at SiO_2/Si and Si/SiGe interfaces respectively, Fig. 1(a). In Fig.1(b), curve (1) is the contribution from SiO_2/Si interface, while curve(2) is contributed from Si/SiGe interface. Combining (1) and (2), we have curve (3) which is the so-called *two-level CP curves*. Based on the IFCP method that we developed in [8], Fig. 17 shows the CP curves for a strained, curve (2), and the control sample(bulk device, curve(1)). Obviously, for a strained-Si device, a two-level CP curve was observed. An abrupt increase of the current at high gate bias, V_{gh}>0.1V is the contribution from the Si/SiGe interface. By employing the technique in [8] and Fig. 3, S/D lateral diffusion length ΔL_0, and the profile of generated N_{it} along the channel can be delineated (Figs. 18 and 19). Similarly, results have been shown for p-MOSFET's with different Ge composition. To identify the observed Si/SiGe interface effect, we see that in n-MOSFET: (1) the Ge effect is more pronounced at the drain junction region and with Si/SiGe interface generated N_{it} near the channel center, and (2) it has a larger generated N_{it} since n-channel exhibits a much larger strain.

6. Summary and Conclusion

To summarize, with the scaling of device dimension and the gate oxide, low leakage interface characterization technique is essential for the physical understanding of the mechanism and for evaluating reliabilities of sub-100nm CMOS devices and beyond. The IFCP method has played a major breakthrough for ultra-thin gate oxide as this as 1nm. This method is superior to the conventional CV method for N_{it} characterization in that the latter needs a large area capacitor samples, while IFCP can deal with a device with small geometry. This IFCP serves many purposes for device channel determination, hot carrier reliability, NBTI reliability, as well as oxide quality process monitors. More recent results on using these methods have been reviewed in this paper.

Depending on the applications, various methods as aforementioned are demonstrated for the monitoring of ultra-thin gate oxide quality as well as for the hot carrier reliability studies. However, efforts still need to be paid for developing more advanced techniques, e.g., gated-diode measurement [6], since the tunneling leakage increases exponentially with the applied gate or drain voltages. In this case, the difficulty of those developed methods will be limited by these tunneling leakage as the scaling of gate oxide continues.

[1] C. -H. Choi, J. S. Goo, T. -Y. Oh, Z. Yu, and R. Dutton, "C-V and Gate Tunneling Current Characterization of Ultra-Thin Gate Oxide MOS(tox– 1.3-1.8nm)," in ***Symp. on VLSI Tech.***, pp. 63-64, 1999.
[2] K. Yang, T. J. King, and C. Hu, "Quantum Effect in Oxide Thickness Determination from Capacitance Measurement," in ***Symp. on VLSI Tech.***, pp. 77-78, 1999.
[3] L. M. Terman, ***Solid State Electronics***, vol. 5(5), pp. 285-299, 1962.
[4] C. Chen and T. P. Ma, "Direct Lateral Profiling of Hot Carrier Induced Oxide Charge and Interface Traps in Thin Gate MOSFET's," ***IEEE Trans. Electron Devices***, Vol. 45, p. 512, 1998.
[5] S. S. Chung et al., "Characterization of Hot Electron Induced Interface States in LATID MOS Devices Using an Improved Charge Pumping Method," ***Ext. Abs. of SSDM***, Chiba, Japan, p. 841, 1993.
[6] P. Speckbacher, J. Berger, A. Asenov, F. Koch, and W. Weber, "The Gated-Diode Configuration in MOSFET's- A Sensitive Tool for Characterizing Hot-Carrier Degradation," ***IEEE Trans. Electron Devices***, Vol. 42, p. 1287, 1995.
[7] P. Masson, J. -L. Autran, and J. Brini, "On the Tunneling Component of Charge Pumping Current in Ultrathin Gate Oxide MOSFET's," ***IEEE Electron Device. Lett.***, vol. 20, pp. 92-94, 1999.
[8] S. S. Chung et al., "A Novel and Direct Determination of the Interface Traps in Sub-100nm CMOS Devices with Direct Tunneling Regime (12-16A) Gate Oxide," in ***Symp. on VLSI Tech.***, p. 74, 2002.
[9] S. M. Cheng, C. M. Yih, J. C. Yeh, S. N. Kuo, and S. S. Chung, "A Unified Approach to Profiling the Lateral Distributions of Both Oxide Charge and Interface States in n-MOSFET's Under Various Bias Stress Conditions," ***IEEE Trans. Electron Devices***, vol. ED-44, p. 1908, 1997.
[10] S. S. Chung et al., "Low Leakage Reliability Characterization Methodology for Advanced CMOS with Gate Oxide in the 1nm," in ***IEDM Tech. Digest***, p. 477, 2004.
[11] S. S. Chung et al., "A New Observation of the Germanium Outdiffusion Effect on the Hot Carrier and NBTI Reliabilities in Sub-100nm Technology Strained-Si/SiGe CMOS Devices," in ***Symp. On VLSI Tech.***, p. 86, 2005.

Acknowledgments This work was supported by the National Science Council, Taiwan, under contract number NSC93-2215-E009-026. The authors would like to thank UMC Central R&D division for continuous support during the course of this study.

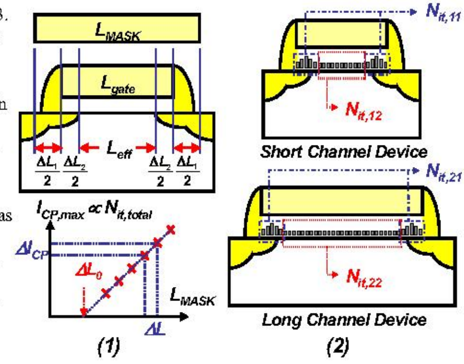


Fig. 5 Illustration of ΔL_0 extraction from CP data.
(1) Parameter definition and extraction method.
(2) Interface traps distribution in short and long channel length devices.

• *Frequency Dependent CP Curves*

$I_{CP(f1,\ with\ leakage)} = I_{CP(f1,\ correct)} + I_{CP(AC\ leakage\ at\ f1)}$ 1(a)

$I_{CP(f2,\ with\ leakage)} = I_{CP(f2,\ correct)} + I_{CP(AC\ leakage\ at\ f2)}$ 1(b)

• *Frequency Dependent CP Components*

$I_{CP(f,\ correct)}$ = strong function of $f \propto f$ 2(a)

$I_{CP(AC\ leakage\ at\ f)}$ = weak function of $f \cong$ constant 2(b)

• *Incremental Frequency CP Methodology*

When $(f_1\text{-}f_2) << f_1$ and $f_1 \rightarrow$ high f,

$I_{CP(AC\ leakage\ at\ f1)} \cong I_{CP(AC\ leakage\ at\ f2)}$ 3(a)

$I_{CP(f1\text{-}f2,\ correct)} \cong I_{CP(f1,\ correct)} - I_{CP(f2,\ correct)}$

$\cong I_{CP(f1,\ with\ leakage)} - I_{CP(f2,\ with\ leakage)}$ 3(b)

Table I Equations for the frequency dependent CP curves, CP components, and the new IFCP method

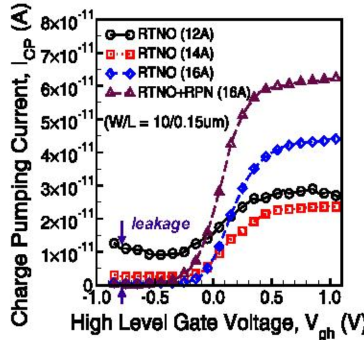


Fig. 2 Measured CP currents for ultra thin (12~16A) gate oxide. Note that 12A gate oxide has large leakage currents for V_{gh} < 0V. [8]

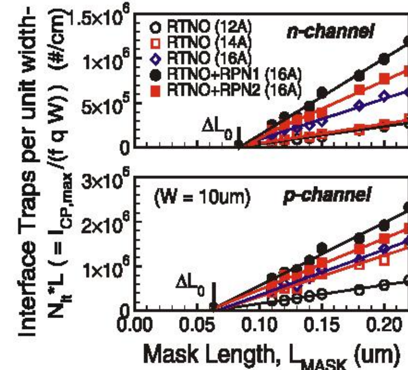


Fig. 6 Calculated N_{it} from $I_{CP,MAX}$ in Fig. 7. It also shows the extraction of offset length ΔL_0 ($\approx \Delta L_1$ ($\approx$ 0.04um) + ΔL_2).

$$\Delta I_{CP} = qfW\int_0^{\infty} N_{it}(x)dx \quad (1)$$

$$N_{it} = \frac{1}{qfW}\frac{d\Delta I_{CP}}{dV_{gl}}\frac{dV_{gl}}{dx} \quad (2)$$

where

$$\frac{dV_{gl}}{dx} = \frac{dV_T(x)}{dx} \quad (3)$$

and

$$N_{ot}(x) = Q_{ot}(x)/q = \frac{C_{ox}\Delta V_G}{q} \quad (4)$$

Table II Equations used to calculate the distributions of N_{it} and N_{ot}.

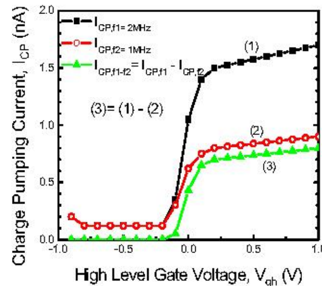


Fig. 3 Incremental frequency charge pumping (IFCP) methodology- Using two-leakage CP curves, $I_{CP,f1}$ and $I_{CP,f2}$, to obtain a correct CP curve, $I_{CP,\Delta f\ \ f1-f2}$.

Gate Oxide	Type	N_{it} (#/cm²)
RTNO 16A	nMOSFET	~ 3.86×10¹⁰
	pMOSFET	~ 7.37×10¹⁰
RTNO 16A + RPN	nMOSFET	~ 5.86×10¹⁰
	pMOSFET	~ 9.49×10¹⁰

(a)

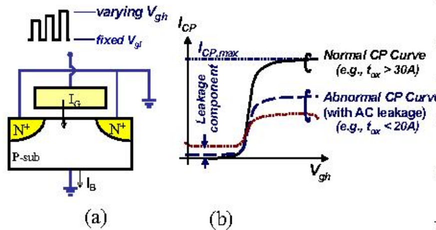


Fig. 1 (a) The schematic of charge pumping (CP) for nMOSFET measurement. (b) Leakage currents occurs when t_{ox}< 20A. [8]

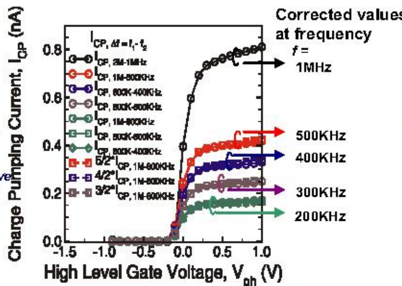
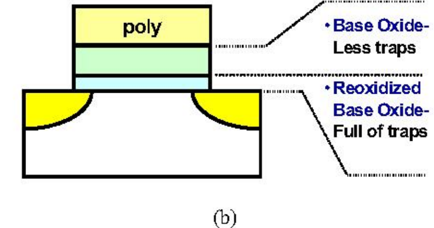


Fig. 4 Corrected CP curves for the IFCP technique. Note that these CP curves have linear relationship between the $\Delta f - f_1 - f_2$, and can be multiplied by an algebraic factor.



(b)

Fig. 7 (a) Amount of interface traps in various types of gate oxide devices.
(b) RPN treated RTNO gate oxide. The bottom layer has been reoxidized which causes more oxide traps.

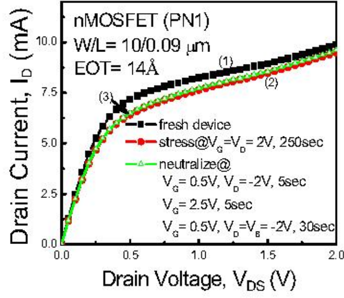


Fig. 8 Measured I_D currents for fresh(1), stress(2), and after neutralization (3) under V_G–V_D stress [10].

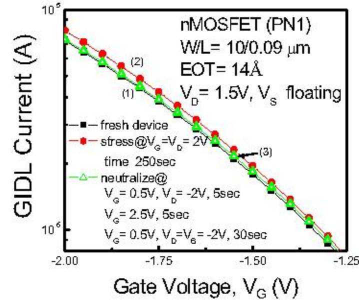


Fig. 9 Measured GIDL currents for fresh, stressed(2), and after neutralization(3). Note that hole trap is eliminated in the neutralization step.

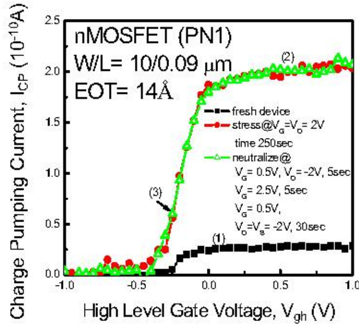


Fig. 10 Measured I_{CP}- curve(1), fresh, curve(2) stress, and curve(3) after neutralization.

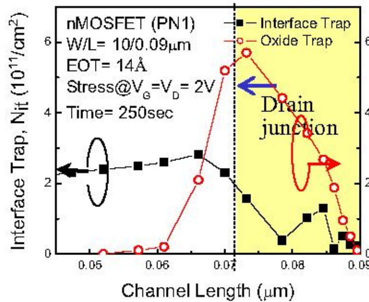


Fig. 11 Calculated lateral distribution of N_{it} and $N_{ot}(-Q_{ot}/q)$ along the channel length under V_G– V_D stress.

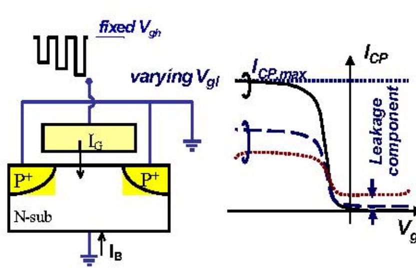


Fig. 12 (left) The schematic of charge pumping (CP) for pMOSFET measurement.
(right) Measured I_{CP} as a function of low varying gate voltages.

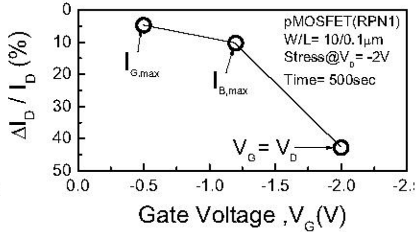


Fig. 13 Measured device I_D degradation at $I_{G,max}$, $I_{B,max}$, and V_G– V_D stress conditions.

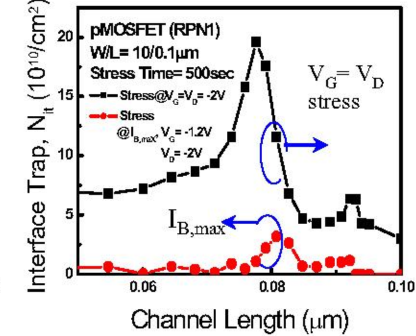


Fig. 14 Comparison of N_{it} distribution between $I_{B,max}$ and V_G– V_D stress conditions. Note that V_G– V_D has much larger values of N_{it}.

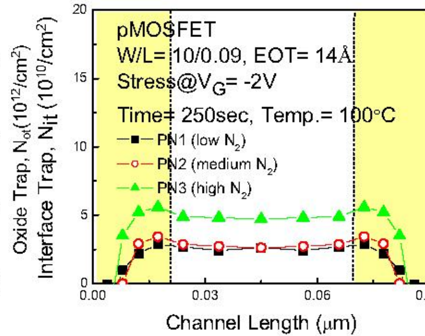


Fig. 15 Symmetrical NBTI stress where D and S are grounded and V_G– -2V, stressed at T– 100°C. N_{it} distribution with double-hump can be seen at the S/D side.

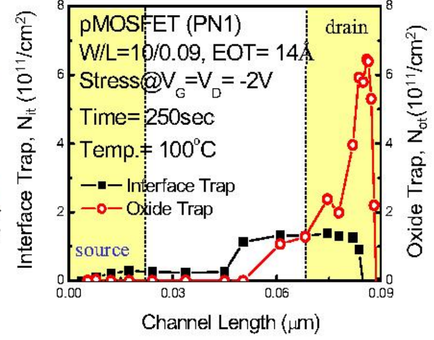


Fig. 16 Asymmetrical NBTI stress where source is grounded while V_G– V_D– -2V is applied at the drain side. Note that N_{it} is dominant of the device degradation since N_{it} has larger values inside the channel region.

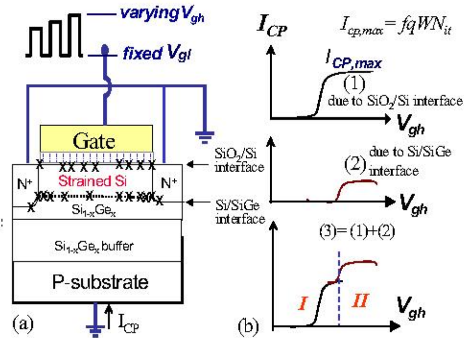


Fig. 17 (a) The CP measurement and the generated interface traps. (b) A two-level CP curve, in which region I is the generated N_{it} at the SiO_2/Si interface, region II is contributed from the Si/SiGe interface [11].

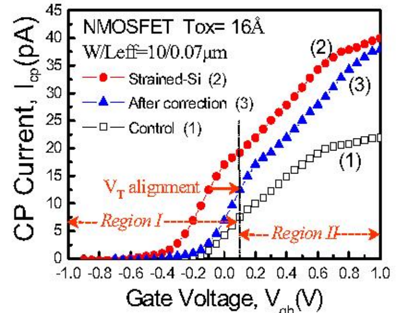


Fig. 18 Measured I_{CP} currents for control(bulk)(1), strained-Si (2), and after correction (3).

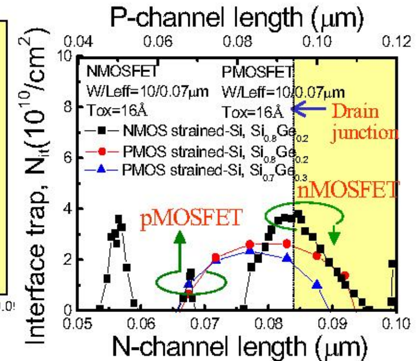


Fig. 19 Calculated Si/SiGe interface generated N_{it} distributions for both n- and p-MOSFET's along the device channel.