Low Leakage Reliability Characterization Methodology for Advanced CMOS with Gate Oxide in the 1nm Range

Steve S. Chung[1], H. J. Feng[1], Y. S. Hsieh[2], Alex Liu[2], W. M. Lin[2], D. F. Chen[2], J. H. Ho[2], K. T. Huang[2], C. K. Yang[2], Osbert Cheng[2], Y. C. Sheng[2], D. Y. Wu[2], W. T. Shiau[2], S. C. Chien[2], Kuan Liao[2], and S. W. Sun[2]

[1]*Department of Electronic Engineering, National Chiao Tung University, Taiwan*
[2]*United Microelectronics Corporation (UMC), Central R&D Division,, Taiwan*

Abstract- A low leakage characterization technique for the lateral profiling of interface and oxide traps in a 12Å-16Å range gate oxide CMOS devices has been demonstrated. The approach being taken includes an incremental frequency Charge-Pumping(IFCP) measurement and a neutralization procedure such that interface and oxide traps can be separated. The most critical steps are the elimination of leakage current during measurement and a neutralization procedure, which enables accurate determination of interface and oxide traps. This method has been demonstrated successfully for an advanced sub-100nm CMOS devices. As an important merit for its application, evaluations of HC reliability and NBTI effect have also been demonstrated. Evaluations of gate oxide qualities with plasma nitridation in both n- and p-MOSFET reliabilities have been properly described based on the current analysis technique.

1. Introduction

There are three major techniques, i.e., charge pumping[1-2], gated-diode[3-4], and DCIV[5], for the characterization of interface/oxide traps(N_{it}/Q_{ot}) in CMOS devices. However, for sub-100nm device with t_{ox} in the range 10-20A, as of today, none of the above has been provided for the *quantitative* calculation of N_{it}/Q_{ot}. A recent one reported by the author using gated-diode for a *thicker oxide* is given in [6] where N_{it} and Q_{ot} can be separated. The gated-diode[3-4] and DCIV[5] can measure the generation-recombination current due to existing N_{it} and/or Q_{ot} in the form of drain or substrate current, however, for *very thin* gate oxide, a *quantitative analysis* is difficult and not available so far. In [7], we have demonstrated successfully an IFCP(Incremental Frequency Charge Pumping) method for the interface characterization, in which a method has been developed to eliminate the leakage current error during the measurement. However, further effort is needed for a full characterization of device reliability, including hot carrier (HC) effect and Negative Bias Temperature Instability(NBTI) studies in particular for pMOSFET devices with $t_{ox} \leq 20A$.

In this paper, a low leakage reliability evaluation technique has been provided for the characterization of HC and NBTI effects, with focus on the profiling of interface/oxide traps. Results have been demonstrated for more advanced technology with effective channel length (60nm) and ultra-thin (12-16A) gate oxide CMOS devices.

2. Device Preparation

The devices in this study were fabricated by the state-of-the-art IC manufacturing. 12-16A gate oxides CMOS devices were formed by advanced plasma nitridations and with several splits of nitride density (PN1-PN3). More detailed descriptions of these samples can also be found from [7].

3. Results and Discussion

A. Principle of the Low Leakage CP Method

Figure 1 shows the schematic diagram of a low leakage IFCP measurement developed by us previously in [7]. With both S/D grounded and by applying a gate pulse with a fixed high level and a varying low voltage (V_{gl}), the channel will operate between accumulation and inversion. This gives rise to the charge pumping current $I_{CP}(=I_B)$ measured from the bulk. Fig. 1(b) shows the normal (t_{ox}>30Å) and abnormal (t_{ox}<30Å) CP curves, from which it can be observed that when t_{ox}<30Å, leakage occurs as a result of the tunneling between the gate and bulk. This will affect the calculation accuracy of the interface traps since interface trap is proportional to the value of measured I_{CP}. Obviously, as revealed in Fig. 2, the leakage current increases largely with a reducing gate oxide thickness (e.g., V_{gh} < 0V for the ultra-thin oxide sample EOT= 12A).

Fig. 3 shows the steps of this IFCP method. We need to remove the leakage current during the CP measurement. Here, 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. Table 1 lists the equations of this CP technique, in which a correct value of CP current is calculated by removing the leakage component by using two CP curves under different frequencies. Again in Fig. 4, it shows the linearity of the measured I_{CP} with measurement frequency in that from the measured CP curves under various frequencies, the I_{CP} for any frequency can be derived from a known frequency values. For example, I_{CP} for 1MHz is twice the value of that for 500kHz, and is 5 times the value measured at 200kHz.

B. Lateral Profiling of Interface/Oxide Traps Along the Channel Direction

As is well known that the lateral profiling of the generated interface or oxide traps, needs to have two key elements, one is the charge pumping current and the other one is the so called local threshold voltage distribution, such that the distribution of traps can be delineated along the channel direction. In the past, several successful implementations of the method can be found from those reported in [1-2]. Based on the basic measurement of a device characteristics as shown in Fig. 5, it shows the device drain currents of a fresh (before-stress, curve 1) and a stressed one(after-stress, curve 2). For the presence of the interface and oxide traps, as can be seen from the measurement of GIDL current in Fig.6, we see that there are oxide traps generated in the gate oxide. Before calculating the profiling of either interface or oxide trap, we need to separate the contribution of charge pumping current contributed from both traps. As a consequence, to separate the N_{it} from Q_{ot}, the steps are as follows.

1. First, 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. 5.
2. To identify whether Q_{ot} is generated, we monitor the GIDL current as given in Fig. 6, 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 three-step, in which curve(2) is moved to curve(3) and then aligned with the fresh one, curve(1). Specifically, first, we inject negative charges into the gate

0-7803-8684-1/04/$20.00 ©2004 IEEE

oxide from the gate edge (V_G= 0.5V, V_D= -2V, time= 5sec). Second, we inject negative charges by FN tunneling under the gate channel (V_G= 2.5V, time=5sec). Finally, we use the bias condition (V_G= 0.5, V_D=V_B= -2V, time= 30sec) to eliminate the hole traps completely. In Fig. 6, we see the GIDL shifts back to the fresh state.

3. At the same time, I_{CP}'s are measured as shown in Fig. 7, the difference between curves (1) and (3) gives the ΔI_{CP}. In accordance with Eq. (2) in Table II, we calculate the value of $d\Delta I_{CP}/dV_{gl}$ and dV_{gl}/dx to obtain the lateral distribution of $N_{it}(x)$. Here, $dV_{gl}(x)/dx$ or $dV_T(x)/dx$ can be obtained in a similar way as those described in [2]. On the other hand, the difference between curves (2) and (3) give the $Q_{ot}(x)$ by using Eq. (4) in Table II. The calculation of N_{it} and Q_{ot} is thus completed.

Results of the N_{it} and N_{ot}(= Q_{ot}/q) distributions are shown in Fig. 8, in which the drain junction is located at $\Delta L/2$= 0.02μm(Fig. 10). The method to determine the physical lateral junction of source or drain with the body can be found in [7] (also in Fig. 9) and the results for the total N_{it}'s are shown in Fig. 10. These results indicate that the device with higher nitrogenation contents exhibits a larger slope and hence a larger amount of N_{it}. Besides, from the calculated value for several different test samples in Fig. 11, it reveals that the one with additional RPN treatment has a larger amount of interface traps.

C. Applications to Evaluating Device HC Reliability

To evaluate the HC reliability, in Fig. 12, 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= 1.4V, 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. 12 and 13 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.

As an application of the profiling of interface/oxide traps, it can be used to identify the quality of gate oxide using different dosages of nitrogen. Fig. 14 shows the comparison of the oxide quality between the PN1 and the PN3 with the same gate oxide thickness (EOT= 16Å). We observed the PN1 sample is worse than PN3 sample after V_G= V_D= 2V stress. As shown in this figure, the $N_{it}(x)$ of PN1 sample is larger than that of PN3 sample whether in the gate channel or in the overlap region. Here, we see that a higher concentration of plasma nitridation is better to resist the channel hot electron injection since PN3 sample reliability has been improved with a high plasma density and N_2 content close to the poly-SiO_2 interface. In other words, higher concentration of plasma nitridation has better HC stress reliability for nMOSFET.

To study the dependency of the HC reliability on the gate oxide thickness, we found that a thicker sample has worse reliability after the stress. This result is consistent with lateral profiling for two different gate oxide thickness samples as in Fig. 15. The thicker sample (16Å) has larger area of N_{it} than thinner sample (14Å) so that it has worse hot carrier reliability. We think the N_{it} at p-substrate of a fresh device is donor type interface trap, so above the quasi-Fermi level E_{fp}, the polarity of N_{it} is positive. Since a thinner sample (14Å) has smaller N_{it} than thicker sample (16Å) ones for a fresh device, the N_{it} of thinner sample (14Å) provides less attractive force to channel hot electron injection and hence induced less N_{it} after the stress. Based on this reasoning, we believe that the N_{it} of a fresh device will affect the amount of hot electron injection for nMOSFET device.

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 shown in Fig. 16. The experimental steps to do the neutralization for pMOSFET are shown in Fig. 17 and the profiling results for N_{it} and N_{ot} are given in Fig. 18. Similar results for the drain current degradation shown in Fig. 19 reveals that the largest degradation occurs at V_G= V_D stress condition. 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. 20 and 21 for different N_2 contents and different gate oxide thickness, respectively. Here, it was noted that the PN3 sample is worse than PN1 sample after V_G= V_D= -2V stress. This might be due to an acceptor type interface trap at the n-substrate and if the interface traps are under the quasi-Fermi level E_{fn}, the polarity of N_{it} is negative which can attract hot hole injection. Since PN3 has larger amount of N_2, it creates larger attractive force for hot hole injection during stress and then leads to more generation of N_{it}.

D. 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. 23 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. 17 should be made. From the results of both Figs. 23 and 24, we can see that N_{it} is largely enhanced on the drain side, which is attributed to a combination of NBTI and HC effect [3]. As a final verification of NBTI, both samples were subject to asymmetrical stress and with profiling results given in Fig. 25, where PN3 with high density plasma nitridation exhibits a largest N_{it} as a result of larger N_2 content and its interaction with H_2.

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. 23), (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. 24), and (3) larger amount of N_{it} has been generated in a heavily nitrided pMOSFET as shown in Fig. 25.

In summary, the low leakage CP measurement (IFCP), equipped with a neutralization step, has been demonstrated successfully for the interface characterization of ultra-thin gate oxide devices. It allows fast determination of both N_{it} and Q_{ot} generated under both HC and NBTI stresses. In particular, the quantitative profiling of N_{it}/Q_{ot} is *the first one reported to date* for t_{ox}<20 Å devices. Evaluation of gate oxide qualities with plasma nitridation in both n- and p-MOSFET reliabilities have been adequately described based on the current analysis technique. Results have shown that the present profiling technique has been a useful approach toward an understanding of the device HC reliability and NBTI effects for CMOS devices with different techniques of nitridation and with ultra-thin gate oxide thickness.

Acknowledgments This work was supported in part by the National Science Council, Taiwan, under contract NSC91-2215-E009-040.

References

[1] S. S. Chung et al., ***Proc. SSDM***, pp. 841-843, 1993.
[2] C. Chen et al., ***IEEE T-ED.***, Vol. 45, No. 2, pp. 512-519, 1998.
[3] S. S. Chung et al., in ***Tech. Dig. IEDM***, pp. 513-516, 2002.
[4] S. -J. Chen et al., ***Proceedings of IRPS,*** pp. 203-207, 2003.
[5] C. T. Sah, ***Proceedings of 6th ICSICT***, pp. 1-15, 2001.
[6] S. S. Chung et al., ***Symposium on VLSI Tech.***, pp. 19-20, 1997.
[7] S. S. Chung et al., ***Symposium on VLSI Tech.***, pp. 19-20, 2002.

- ***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 ? f*** **2(a)**

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

- ***Incremental Frequency CP Methodology***

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

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

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

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

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

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

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

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

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

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

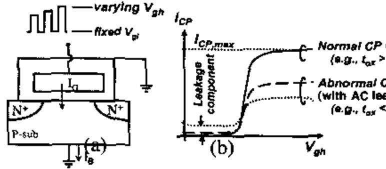

Fig. 1 (a) The schematic of charge pumping (CP) for nMOSFET measurement. (b) Leakage currents occurs when t_{ox}< 20A. [7]

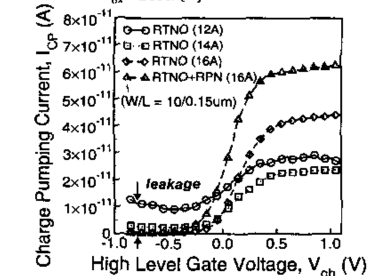

Fig. 2 Measured CP currents for ultra thin (12~16A) gate oxide. Note that 12A gate oxide has large leakage currents for V_{gh} < 0V. [7]

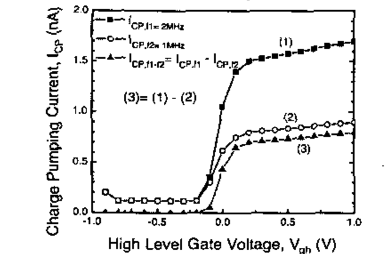

Fig. 3 Incremental frequency charge pumping (IFCP) methodology- Using two-leakage CP curves, $I_{CP,f1}$ and $I_{CP,f2}$, to obtain a correct CP curve, $I_{CP,?f=f1-f2}$.

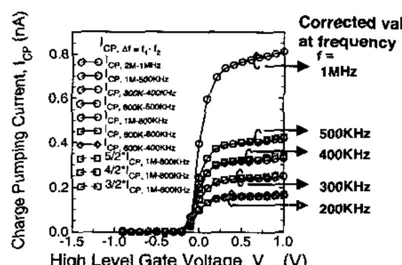

Fig. 4 Corrected CP curves for the IFCP technique. Note that these CP curves have linear relationship between the ? f = f_1 -f_2, and can be multiplied by an algebraic factor.

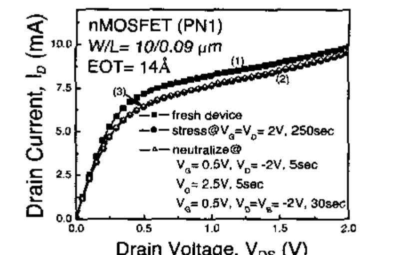

Fig. 5 Measured I_D currents for fresh(1), stress(2), and after neutralization (3) under V_G=V_D stress.

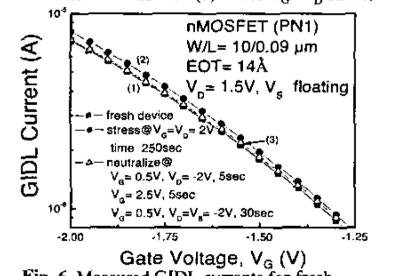

Fig. 6 Measured GIDL currents for fresh, stressed(2), and after neutralization(3). Note that hole trap is eliminated in the neutralization step.

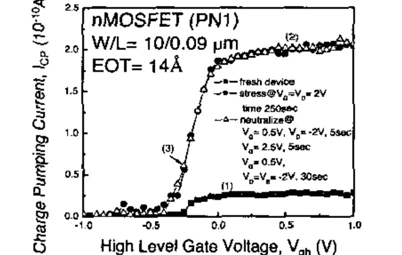

Fig. 7 Measured I_{CP}- curve(1), fresh, curve(2) stress, and curve(3) after neutralization.

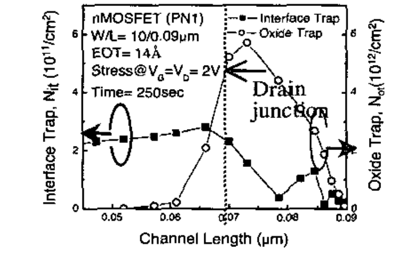

Fig. 8 Calculated Lateral distribution of N_{it} and N_{ot}(=Q_{ot}/q) along the channel length under V_G= V_D stress.

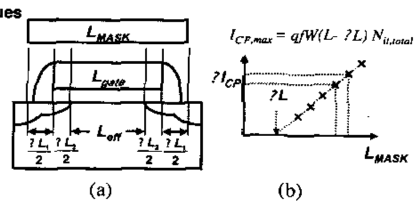

Fig. 9 (a) Definition of L and ? L. (b) The method to extract the lateral junction of S/D junction.

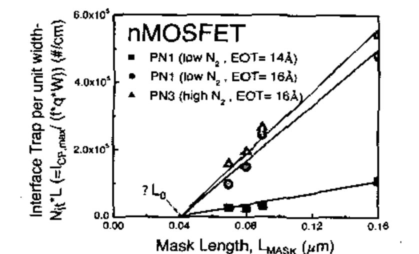

Fig. 10 Calculated N_{it} from $I_{CP,MAX}$ in Fig. 1. It also shows the extraction of offset length ? L_0 .

Gate Oxide	Type	N_{it} (#/cm^2)
RTNO 16A	nMOSFET	~ 3.86x10^{10}
	pMOSFET	~ 7.37x10^{10}
RTNO+RPN 16A	nMOSFET	~ 5.86x10^{10}
	pMOSFET	~ 9.49x10^{10}

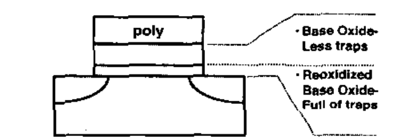

Fig. 11 **(a)** Amount of interface traps in various types of gate oxide devices (Fig. 2).
(b) Plasma nitrided gate oxide. The highly nitrided has larger amount of interface traps.

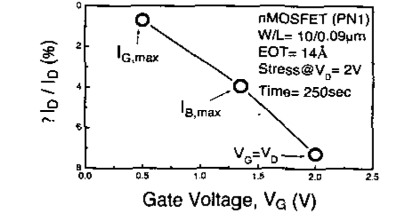

Fig. 12 Measured device I_D degradation at $I_{G,max}$, $I_{B,max}$, and V_G= V_D stress conditions.

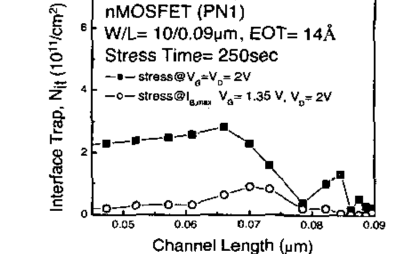

Fig. 13 Comparison of N_{it} distribution between $I_{B,max}$ and V_G= V_D stress conditions. Note that V_G= V_D has much larger values of N_{it}.

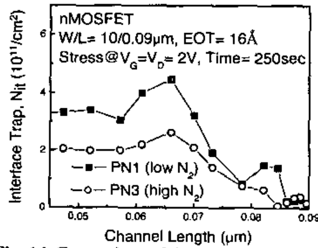

Fig. 14 Comparison of the generated N_{it} for two different plasma nitrided samples, where PN3 reliability has been improved with a high plasma density and N_2 content close to poly-Si.

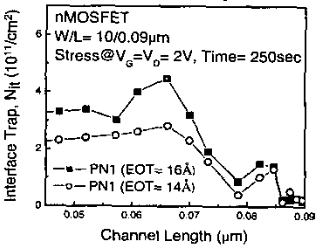

Fig. 15 Comparison of the generated N_{it} for two different gate oxide thicknesses, where thinner oxide shows better reliability.

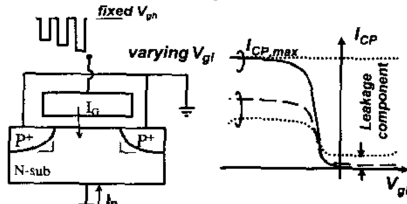

Fig. 16 (left) The schematic of charge pumping (CP) for pMOSFET measurement. (right) Measured I_{CP} as a function of low varying gate voltages

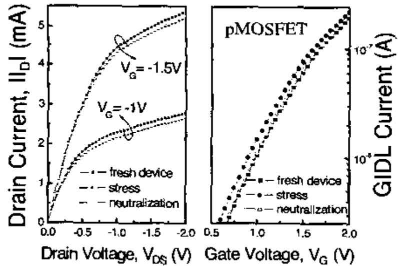

Fig. 17 (Left) Drain currents for fresh, stressed, and after neutralization. N_{it} is dominant from these curves. (Right) The neutralization procedure is similar to Fig. 5

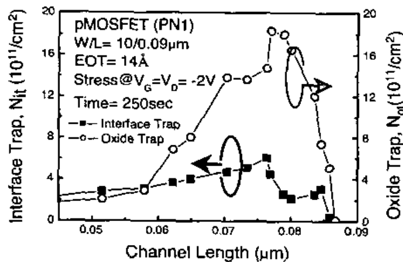

Fig. 18 Calculated Lateral distribution of N_{it} and N_{ot}(=Q_{ot}/q) along the channel length under V_G= V_D stress for pMOSFET device.

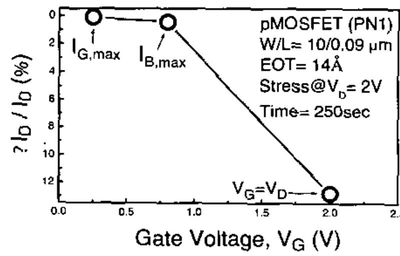

Fig. 19 (pMOSFET) Measured device I_D degradation at $I_{G,max}$, $I_{B,max}$, and V_G= V_D stress conditions.

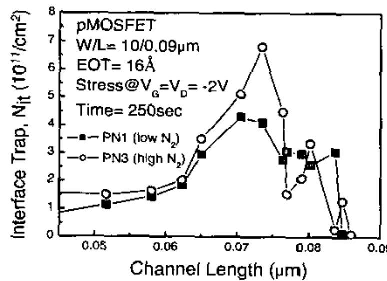

Fig. 20 (pMOSFET) Comparison of the generated N_{it} for two different plasma nitrided samples, where PN1 device reliability has been improved with a lower plasma density and a lower N_2 content.

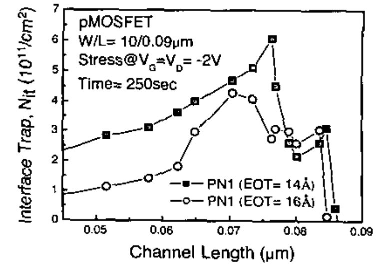

Fig. 21 (pMOSFET) Comparison of the generated N_{it} for two different gate oxide thicknesses, where thinner oxide shows better reliability.

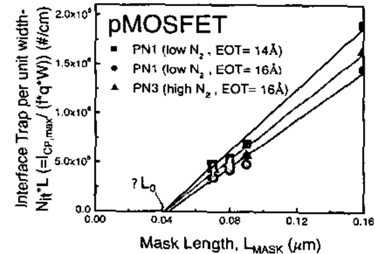

Fig. 22 Calculated N_{it} from $I_{CP,max}$ in Fig. 16. It also shows the extraction of offset length ? L_0.

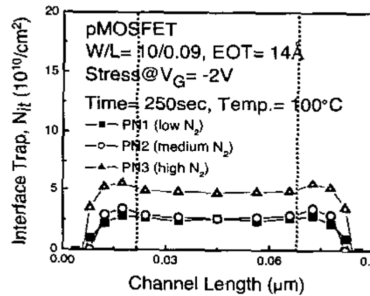

Fig. 23 Symmetrical NBTI stress where D and S are grounded and V_G= -2V, stressed at T= 100°C. N_{it} distribution with double-hump can be seen at the S/D side.

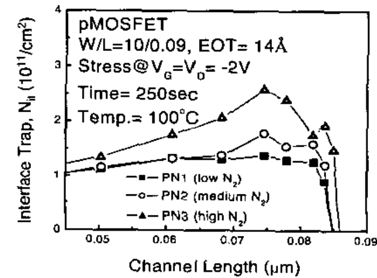

Fig. 25 Asymmetrical NBTI stress where source is grounded while V_G= V_D= -2V is applied at the drain side. Note that PN1 has a lower plasma nitridation density and a better reliability.

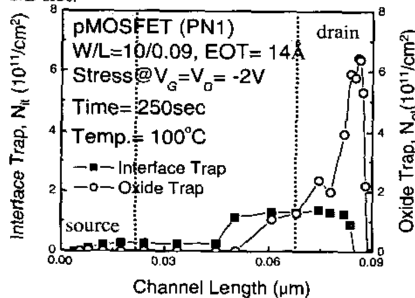

Fig. 24 Asymmetrical NBTI stress where source is grounded while V_G= V_D= -2V is applied at the drain side. Note that N_{it} is dominant of the device degradation since N_{it} has larger values inside the channel region.