Latent Damage Generation in Thin Oxides of Metal-Oxide-Semiconductor Devices under High-Field Impulse Stress and Damage Characterization Using Low-Frequency Noise Measurement

2001 Jpn. J. Appl. Phys. 40 6770

(http://iopscience.iop.org/1347-4065/40/12R/6770)

Jpn. J. Appl. Phys. Vol. 40 (2001) pp. 6770–6777
Part 1, No. 12, December 2001

Latent Damage Generation in Thin Oxides of Metal-Oxide-Semiconductor Devices under High-Field Impulse Stress and Damage Characterization Using Low-Frequency Noise Measurement

Wai Kin CHIM and Peng Soon LIM

Center for Integrated Circuit Failure Analysis and Reliability, Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576

(Received June 29, 2001; accepted for publication August 14, 2001)

Latent damage generation in thin oxides of metal-oxide-semiconductor (MOS) devices, caused by high-field impulse or electrostatic discharge (ESD) stress applied to the gate electrode, can lead to increased trap generation during subsequent hot-carrier stressing. However, the charge-to-breakdown (Q_{bd}) of such impulse prestressed devices is not significantly affected by the latent damage, and hence it is difficult to characterize such damage using Q_{bd} measurements. Monitoring of the latent damage can be carried out by detecting the change in an appropriate electrical parameter of the device or by extracting the generated interface states and oxide trap charges. However, such electrical measurements have their own limitations. It was found that low-frequency noise measurement is a more sensitive method than the above-mentioned electrical measurements for characterizing the ESD stress-induced latent damage in thin oxides.

KEYWORDS: silicon dioxide, electrostatic discharge, hot carrier, interface state, oxide trap charge, flicker noise, latent damage

1. Introduction

Decreasing feature sizes in submicrometer metal-oxide-semiconductor (MOS) devices have resulted in a deterioration of the oxide-breakdown protection margin under electrostatic discharge (ESD) or high-field impulse stress. This means that the difference between the junction breakdown trigger voltage and oxide breakdown voltage for a device is becoming smaller as device dimensions shrink. For example, for a sub-quarter micron complementary-MOS (CMOS) process with an oxide thickness of 4 nm, the oxide-breakdown protection margin against ESD stress is estimated to decrease to zero.[1)] The use of epi-CMOS technology with decreasing epitaxial layer thickness for CMOS latchup prevention can also raise the trigger voltages of the silicon controlled rectifiers used in ESD protection circuits to levels higher than the oxide breakdown voltage. All these would imply an increasing susceptibility of the device to high-field impulse stress with the resultant effect of catastrophic oxide breakdown being reached at lower ESD dose levels. In addition, there will also be situations when the oxide does not break down catastrophically under low-level ESD transients but may suffer sufficient latent damage, impacting the long-term reliability of the device.[2)] The detection of such latent damage is usually performed either by monitoring the change in an appropriate electrical parameter (*e.g.*, the threshold voltage, linear drain current or transconductance) of the device.[2)] Besides detecting the ESD latent damage, it is also useful to obtain a better understanding of the types of damage generated under high-field impulse stress. Although the types and distribution of traps generated under constant-current or constant-voltage stress have been studied extensively, information on damage caused by short-duration, high-field impulse stress is not widely available. Such knowledge will be useful in minimizing the latent damage generation under high-field impulse stress through optimization of the fabrication processes and/or materials.

In this article, we will first show that the conventional charge-to-breakdown (Q_{bd}) measurement is not sufficiently sensitive in detecting latent damage in oxides subjected to high-field impulse stress applied between the gate and substrate terminals. This is despite the fact that latent damage in the form of trapped holes and neutral electron traps is generated by such high-field impulses applied between the gate and substrate electrodes, as determined from capacitance–voltage (C–V) and transient current (I–t) measurements on the capacitor test structures. Measurements of the transistor subthreshold characteristics also show oxide trap charge and interface state generation in devices subjected to high-field impulse stress. It will be shown that there is a correlation between the maximum transconductance ($G_{m,max}$) of the transistor device and the magnitude of the low-frequency noise. We will also show that low-frequency (flicker or $1/f$) noise measurement can provide a more sensitive alternative for detecting oxide latent damage compared to $G_{m,max}$ measurement.

It is to be noted that the high-field impulse stressing in this study was implemented with high-voltage impulses applied between the gate and substrate terminals of the MOS field-effect-transistor (MOSFET). While this type of stress does occur in some input circuits, the dominant type of ESD stress in most input/output circuits is that when the impulse stress is applied to the drain terminal of the MOSFET device. This is because most ESD protection circuits have additional circuit elements protecting the input, while the output driver circuit must bear the brunt of an ESD event. Impulse stressing applied to the drain terminal of the MOSFET device can also generate latent damage in the oxide, especially under very fast and short duration transients. However such stressing and latent damage generation have been studied previously.[2)]

2. Experimental Details

The MOS transistors used for the subthreshold and low-frequency noise measurements are commercial grade n^+-polysilicon gate/n-channel (n^+/nMOSTs) and p^+-polysilicon gate/p-channel (p^+/pMOSTs) surface-channel MOS transistors. The n^+/nMOSTs were fabricated using a 0.35 μm technology with an electrical oxide thickness (T_{ox}) of 7 nm, as determined from C–V measurements. The n^+/nMOSTs have

different mask gate lengths (L) of 0.25, 0.35 and 0.5 μm and a gate width (W) of 50 μm. The gate oxide was grown in a partial wet ambient at 850°C. The source and drain of the n^+/nMOSTs were formed by arsenic implantation at 50 keV to a dose of $4 \times 10^{15}\,cm^{-2}$. The low-doped drain (LDD) implant was performed at 25 keV with a phosphorous dose of $4.5 \times 10^{13}\,cm^{-2}$. The p^+/pMOSTs have different mask gate lengths (L) of 0.24, 0.34, 0.5 and 1.0 μm, a similar gate width (W) of 50 μm and an electrical oxide thickness (T_{ox}) of 4.3 nm. The LDD implant for the p^+/pMOSTs was carried out at 10 keV with boron difluoride (BF_2), to a dose of $2 \times 10^{14}\,cm^{-2}$. The source and drain were formed by BF_2 implantation at 30 keV to a dose of $4 \times 10^{15}\,cm^{-2}$.

The MOS capacitors used for the charge-to-breakdown (Q_{bd}) studies were n^+-polysilicon gate/p-substrate (n^+/p-Si) capacitors from the same 0.25 μm technology process and the same wafer as the p^+/pMOSTs. The capacitor test structures have a gate area of $2.33 \times 10^{-4}\,cm^2$ and $T_{ox} = 4.3$ nm. The capacitors were subjected to varying degrees of impulse prestress treatment with the positive-voltage impulses applied either to the gate with substrate grounded (substrate electron injection condition) or to the substrate with gate grounded (gate electron injection condition) before the Q_{bd} test. The results of the Q_{bd} tests were then compared to that of virgin capacitors without the impulse prestress treatment.

The high-field impulse prestress treatment was performed using a single-pulsing transmission line technique.[3] Several high-voltage pulses of 4 to 7 ns rise time and 200 ns duration were applied to the gate (or substrate) electrodes of the MOS device (i.e., MOS transistor or capacitor) to simulate a low-level ESD event. Varying number (and amplitude) of high-field impulses were applied to generate varying degrees of degradation. The number and amplitude of impulses were chosen such that a substantial amount of latent damage, indicated by significant shifts in the threshold voltage (V_{th}) and maximum transconductance ($G_{m,max}$), was created without resulting in oxide breakdown. For the n^+/nMOSTs, 6 to 25 high voltage (9 to 15 V across the oxide) impulses were applied to the gate electrode with the substrate, source and drain terminals connected to ground. In the subsequent study on p^+/pMOSTs, the high-field impulses were applied to the body or substrate (i.e., n-well) with the gate connected to ground and source/drain terminals floated to prevent oxide breakdown from occurring at the gate-to-source/drain overlapping regions. 15 to 150 impulses, at a higher voltage of 19.8 V across the oxide, were applied to the substrate of the p^+/pMOSTs. The ability of the p^+/pMOSTs to withstand a higher impulse stress voltage without oxide breakdown as compared to the n^+/nMOSTs could be attributed to the floating of the source/drain terminals of the former. The polarity of impulses applied to both n^+/nMOSTs and p^+/pMOSTs were chosen such that the channel region was driven into deep-depletion/inversion in both cases.

A schematic of the single-pulsing transmission line setup is shown in Fig. 1. During operation, the transmission line is charged to the supply voltage V_S through the 10-MΩ charging resistor. When the relay switch S_1 is closed, a discharge to the device-under-test (DUT) occurs, producing a high current event and a square voltage pulse of amplitude $V_S/2$. A calibrated current probe and a storage oscilloscope were used to monitor the current and voltage waveforms across the DUT during the high-field ESD stress.

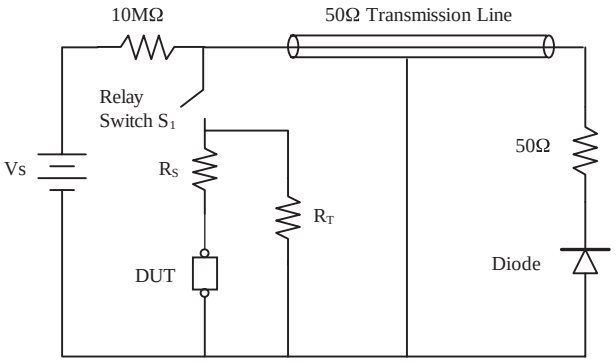


Fig. 1. Single-pulsing transmission line setup used for the high-field impulse or ESD stressing of the devices.

The hot-carrier stressing of the n^+/nMOSTs and p^+/pMOSTs was performed under the maximum degradation conditions. For the n^+/nMOSTs, hot-carrier stressing was carried out under maximum substrate current ($I_{sub,max}$) condition at a drain voltage (V_d) of 4 V and a gate voltage (V_g) of 1.55 V, with source and body (substrate or p-well) grounded. For the p^+/pMOSTs, hot-carrier stressing was carried out under maximum gate current ($I_{g,max}$) condition at $V_d = -7$ V and $V_g = -1.15$ V, with source and body (substrate or n-well) grounded. The gate current measured during $I_{sub,max}$ and $I_{g,max}$ stress were < 50 fA and ~ 400 nA, respectively.

The number of interface traps (δN_{it}) and oxide trap charges (δN_{ot}) generated by the hot-carrier and high-field impulse stressing were obtained from changes in the subthreshold current characteristics of the virgin and stressed n^+/nMOSTs devices using the technique proposed by McWhorter and Winokur,[4] commonly known as McWhorter's method. For the p^+/pMOSTs, δN_{ot} and N_{it} were extracted using the dc current–voltage (DCIV) technique.[5] This is because the extracted trap densities, obtained using McWhorter's method, are not accurate in the presence of nonuniformly distributed oxide trap charges. Such lateral non-uniform charge distributions in the gate oxide are typically generated under impulse stressing applied between the gate and substrate electrodes with the device biased into deep depletion.[6] To overcome such limitation, the DCIV technique, configured in the top-emitter configuration, was used to estimate δN_{ot} and δN_{it} from the peak base (n-well) recombination current density ($I_{base,peak}$) versus gate voltage (V_g) plot. The shift of $I_{base,peak}$ in the positive (negative) gate-voltage direction (i.e., δV_{peak}) and the change in the $I_{base,peak}$ amplitude (i.e., $\delta I_{base,peak}$) are proportional to the number of negative (positive) oxide trap charges and interface states generated, respectively. δN_{ot} following electrical stress can be calculated using ($C_{ox}\delta N_{ot}/q$), where C_{ox} is the oxide capacitance per unit area and q is the charge of an electron. It should be noted that the DCIV measurement requires the presence of a vertical bipolar transistor structure. Since the available samples (i.e., n^+/nMOSTs and p^+/pMOSTs) were all fabricated on a p-type substrate with a dual-well structure, the DCIV measurement can only be carried out on the p^+/pMOSTs, and not on the n^+/nMOSTs. For wafers with a p-type substrate, the vertical bipolar structure is a pnp transistor structure formed from the p-type substrate, n-well and p^+-source/drain diffusion and this is present only in the p^+/pMOSTs, and not in the n^+/nMOSTs.

Low-frequency ($f = 10$ Hz) flicker noise measurements were carried on the n^+/nMOSTs and p^+/pMOSTs by obtaining the drain current noise (S_{Id}). The device was biased in the linear region at $|V_d| = 0.1$ V and at a constant effective gate voltage $|V_g - V_{th}| = 0.6$ V, where V_{th} is the threshold voltage of the device at the particular stress time point, for the low-frequency noise measurements.[7] The input-referred noise power, S_{V_g} (equal to S_{Id}/G_m^2), was then obtained, with the transconductance (G_m) measured at $|V_g - V_{th}| = 0.6$ V. The noise measurements were normalised to frequency, with S_{V_g} expressed in units of V^2/Hz.

3. Results and Discussion

Figure 2 shows the Weibull probability plot of the charge-to-breakdown (Q_{bd}) distribution of the n^+/p-Si MOS capacitors ($T_{ox} = 4.3$ nm). The first group of capacitors received no impulse prestress treatment. The second group are capacitors prestressed with twenty-three high-field (~ 61 V, ~ 65 kA/cm^2) impulses applied to the gate (substrate injection), while the third group are capacitors prestressed with twenty high-field (~ 48 V, ~ 75 kA/cm^2) impulses applied to the substrate (gate injection). For gate-injection impulse prestress, 20 pulses were chosen because most devices can only withstand a maximum of 20 to 30 impulses at the stated stressing conditions before oxide breakdown. For the substrate-injection impulse prestress, 23 pulses were chosen so that an equivalent amount of fluence (~ 0.3 C/cm^2) was injected as in the case of gate injection, despite the fact that most devices could withstand more than 30 pulses without oxide breakdown. The Q_{bd} measurements were performed at a constant current density of 1.72 mA/cm^2 (gate current $I_g = -400$nA) with the capacitors biased into accumulation. In view of the thin oxide used, the onset of soft breakdown was taken as the criterion for oxide breakdown in the Q_{bd} measurements. The results show that there is hardly any difference in the Q_{bd} values of the impulse prestressed devices (both substrate injection and gate injection) as compared to the devices without the impulse prestress treatment. This is despite the fact that latent damage, in the form of trapped holes are present in the devices subjected to high-field impulse stressing, as reported previously.[8] One possible explanation for this could be the small amount of cumulative charge injected during the impulse prestress treatment (~ 0.3 C/cm^2) as compared to a median Q_{bd} of about 20 C/cm^2. Similar Q_{bd} measurements have been performed previously on n^+-polysilicon gate/n-substrate (n^+/n-Si) MOS capacitors with a thicker oxide ($T_{ox} = 7$–8 nm),[9] and the conclusion was similar to that for Fig. 2.

It was suggested previously that for ultrathin oxides (< 5 nm thickness), constant-voltage stress (CVS) should be used for the Q_{bd} measurement instead of constant-current stress (CCS).[10] This is because for very thin oxides, the injected electrons can tunnel ballistically through the oxide without interacting with the silicon dioxide lattice. This means that the electron energy at the anode is determined by the voltage difference between the cathode and anode, which corresponds to the applied gate voltage. As such, it was argued that the Q_{bd} measurements should be performed using CVS for ultrathin oxides. Figure 3 shows the Weibull probability plot of the charge-to-breakdown (Q_{bd}) distribution of similar n^+/p-Si MOS capacitors as in Fig. 2 but measured under CVS with an applied gate voltage $V_g = -6$ V (accumulation mode). Again, the results show that there is no significant difference in the Q_{bd} values of the impulse prestressed devices (both substrate injection and gate injection) as compared to the devices without the impulse prestress treatment. However, the impulse prestressed devices do show slightly higher Q_{bd} values, due possibly to preconditioning or strengthening of the oxide by the prestress as has been reported previously. Martin *et al*. suggested two possible explanations for the increase in Q_{bd} in a prestressed device.[11] First, the elimination of possible electrical breakdown sites that trigger eventual thermal breakdown could lead to a better oxide reliability. Secondly, the trapped electrons during the pre-stress treatment could decrease the rate of charge injection and build-up during the CVS Q_{bd} test. Notwithstanding the correct expla-

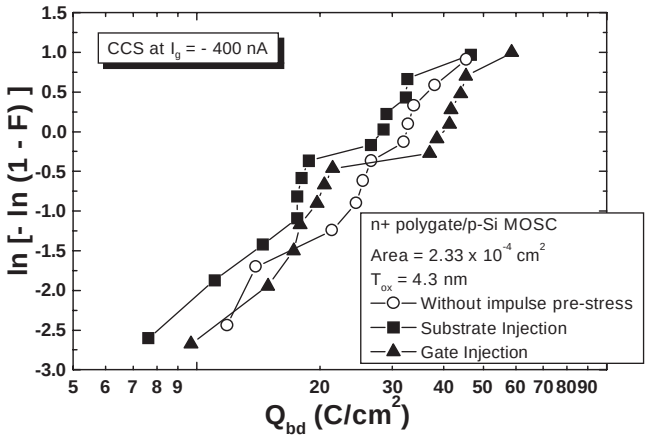


Fig. 2. Charge-to-breakdown (Q_{bd}) distribution under constant-current stress (CCS) of n^+/p-Si MOS capacitors (MOSCs) without the impulse prestress treatment, capacitors prestressed with 23 high-field (~ 61 V, ~ 65 kA/cm^2) impulses applied to the gate (substrate injection) and capacitors prestressed with 20 high-field (~ 48 V, ~ 75 kA/cm^2) impulses applied to the substrate (gate injection). The Q_{bd} measurements were performed at a constant current density of 1.72 mA/cm^2. The n^+/p-Si MOSCs have an oxide thickness of 4.3 nm and a gate area of 2.33×10^{-4} cm^2.

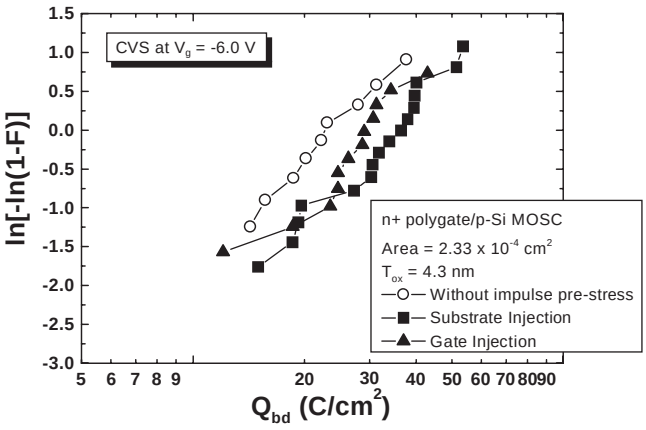
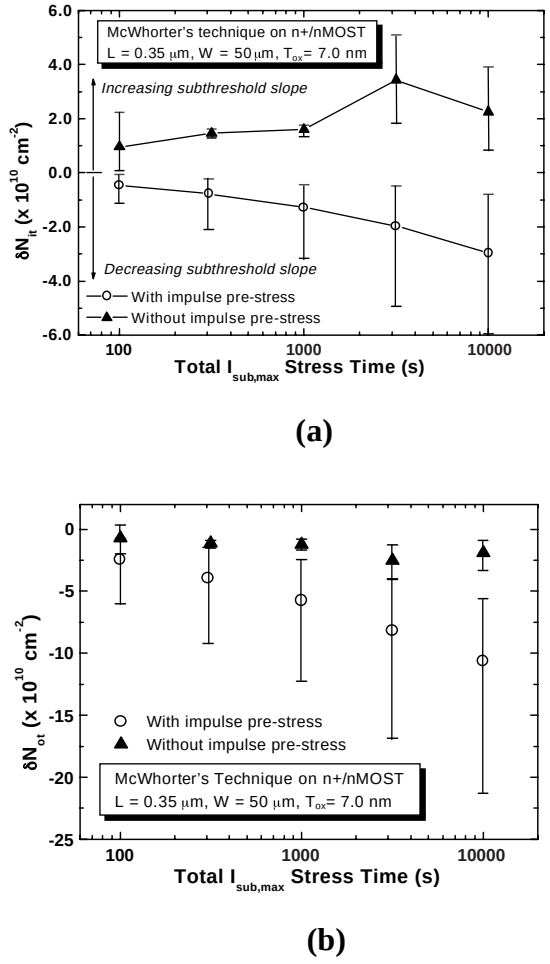


Fig. 3. Charge-to-breakdown (Q_{bd}) distribution under of n^+/p-Si MOS capacitors (MOSCs) without the impulse prestress treatment, capacitors prestressed with 23 high-field (~ 61 V, ~ 65 kA/cm^2) impulses applied to the gate (substrate injection) and capacitors prestressed with 20 high-field (~ 48 V, ~ 75 kA/cm^2) impulses applied to the substrate (gate injection). The Q_{bd} measurements were performed at a constant gate voltage, $V_g = -6$ V. The n^+/p-Si MOSCs have an oxide thickness of 4.3 nm and a gate area of 2.33×10^{-4} cm^2.

nation, the above results show that Q_{bd} measurements are not sufficiently sensitive to detect the latent damages generated by the impulse stress. It should also be noted that impulse stress generally generates a large amount of positive oxide trap charge because of the regenerative (feedback) injection mechanism, as reported previously.[8] For CVS or CCS, the amount of positive charge trapping in the oxide is much less and occurs during the early duration of stress. For longer stress duration, electron or negative charge trapping tends to dominate.[3]

The high-field impulse prestress was also carried out on n^+/nMOSTs (gate length $L = 0.35\,\mu$m, gate width $W = 50\,\mu$m and oxide thickness $T_{ox} = 7$ nm). Some of the impulse prestressed n^+/nMOSTs were subjected to a subsequent $I_{sub,max}$ hot-carrier stress step to investigate the effect of the impulse prestress on the subsequent hot-carrier reliability. Figure 4 shows the change in interface trap density (δN_{it}) and oxide trap charge density (δN_{ot}) generated during the $I_{sub,max}$ hot-carrier stressing for the n^+/nMOSTs. The δN_{it} and δN_{ot}, for both sets of devices, were taken to be zero at time $t = 0$ of the $I_{sub,max}$ stress. The symbol and line bar at each stress time represents the average and range of the δN_{it} and δN_{ot} values obtained from four devices. In Fig. 4(a), a positive (negative) δN_{it} value means that the subthreshold slope of the device becomes larger (smaller) after stress. The n^+/nMOSTs with and without the impulse prestress treatment were observed to exhibit opposite signs in δN_{it}, which implies a change in subthreshold slope in the opposite manner, during the hot-carrier stress. For the n^+/nMOSTs with the impulse prestress treatment, the increasingly negative δN_{it} with increase in hot-carrier stress time seems to be counterintuitive. This discrepancy could possibly be explained by the annihilation of nonuniformly distributed positive charges generated by the impulse prestress dominating the generation of interface traps during hot-carrier stress. The former gives rise to a smaller subthreshold slope while the latter results in a larger subthreshold slope. In Fig. 4(b), it is seen that the impulse prestress treatment results in greater electron trapping (indicated by negative values for δN_{ot}) during the subsequent hot-carrier stress.



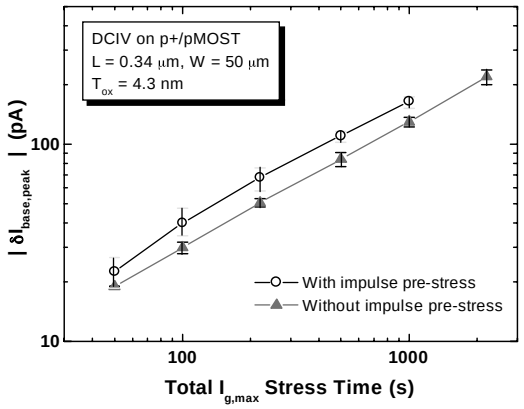
(a)

(b)

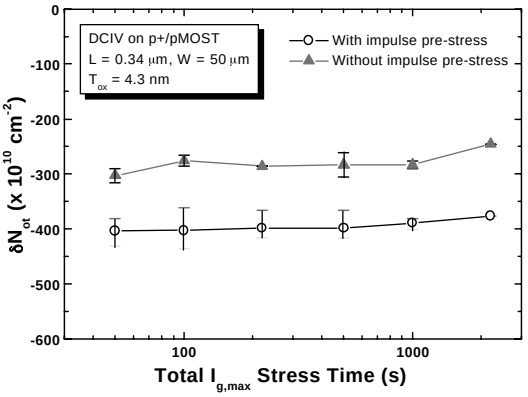
Fig. 4. Change in (a) interface trap density (δN_{it}) and (b) oxide trap charge density (δN_{ot}) during maximum substrate current ($I_{sub,max}$) hot-carrier stress for n^+/nMOSTs (gate length $L = 0.35\,\mu$m, gate width $W = 50\,\mu$m and oxide thickness $T_{ox} = 7$ nm) with (open circles) and without (solid triangles) the impulse prestress treatment. The symbol and line bar at each stress time represent the average and range of the δN_{it} and δN_{ot} values obtained from four devices. In (a), a positive (negative) δN_{it} value means that the subthreshold slope of the device becomes larger (smaller) after stress. In (b), a negative δN_{ot} value implies electron trapping. The $I_{sub,max}$ hot-carrier stressing of the n^+/nMOSTs was performed at a drain voltage (V_d) of 4 V and a gate voltage (V_g) of 1.55 V. The impulse prestress treatment was performed by applying 6 to 25 high-voltage (9 to 15 V across the oxide) pulses to the gate electrode (inversion condition during impulse prestress) of the n^+/nMOSTs, with the substrate, source and drain grounded.

It should be noted that the results in Fig. 4 were obtained using McWhorter's method,[4] which estimates net trap densities throughout the entire oxide. In situations where traps of opposite polarity are generated, the oppositely-charged traps tend to cancel each other. McWhorter's method also applies well for situations of uniform damage but is less accurate for cases of nonuniform damage generation, as could occur during the impulse prestress and hot-carrier stress. In view of the limitations of McWhorter's method, the trap measurement was repeated using an alternative technique that does not have such limitations. This alternative trap measurement method is the DCIV technique.[5] The DCIV measurements, however, require a vertical bipolar transistor structure. As the transistor devices were fabricated on wafers with a p-type substrate and a dual-well structure, the DCIV measurements can only be carried out on p-channel MOS transistors (i.e., p^+/pMOSTs fabricated on the same wafer as the n^+/p-Si MOS capacitors used in the Q_{bd} measurements), as explained previously. The p^+/pMOSTs ($L = 0.35\,\mu$m, $W = 50\,\mu$m and $T_{ox} = 4.3$ nm), with and without the impulse prestress treatment, were subjected to a hot-carrier stress step under $I_{g,max}$ bias (maximum hot-carrier degradation condition of p-channel MOS transistors). Figures 5(a) and 5(b) show the change in interface trap density (proportional to $|(I_{base,peak}|)$ and oxide trap charge density (δN_{ot}), respectively, during hot-carrier stress, extracted using DCIV measurements. It is seen that the p^+/pMOSTs with the impulse prestress treatment show higher interface trap generation and electron trapping after the first 50 s of hot-carrier stress. The electron trapping did not change much for a longer duration of hot-carrier stress after the first 50 s, while the interface trap generation continues to increase for longer hot-carrier stress duration.

The effects of the generated interface traps and oxide trap charges on the fractional change in the maximum transconductance ($|\delta G_{m,max}/G_{m,max}|$) were investigated. Figure 6 shows the relationship between the interface trap density (i.e., δN_{it} or $|\delta J_{base,peak}|$) and oxide trap charge density (δN_{ot}) to $|\delta G_{m,max}/G_{m,max}|$. The maximum transconductance was

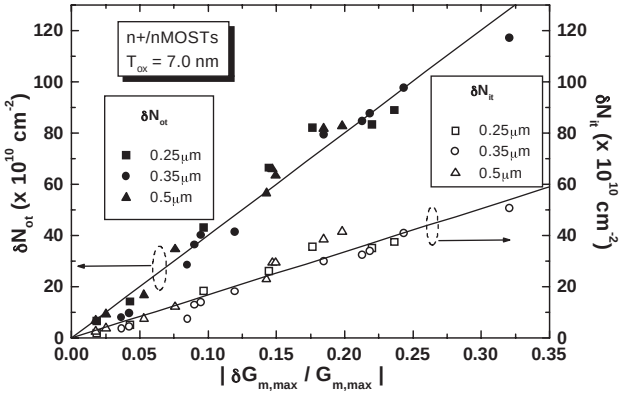


(a)

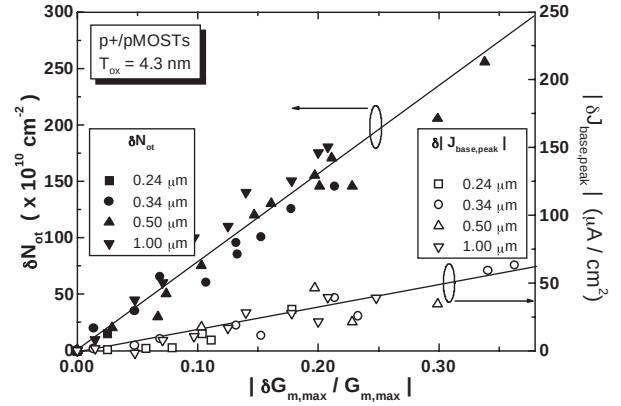


(b)

Fig. 5. Change in (a) the magnitude of the peak base current (i.e., $|\delta I_{\mathrm{base,peak}}|$) and (b) oxide trap charge density (δN_{ot}) during maximum gate current ($I_{\mathrm{g,max}}$) hot-carrier stress for p^+/pMOSTs (gate length $L = 0.34\,\mu$m, gate width $W = 50\,\mu$m and oxide thickness $T_{\mathrm{ox}} = 4.3$ nm) with (open circles) and without (triangles) the impulse prestress treatment. The change in the interface trap density is proportional to $|\delta I_{\mathrm{base,peak}}|$. The symbol and line bar at each stress time represent the average and range of the δN_{it} and δN_{ot} values obtained from four devices. In (b), a negative δN_{ot} value implies electron trapping. The $I_{\mathrm{g,max}}$ hot-carrier stress of the p^+/pMOSTs was performed at a drain voltage (V_{d}) of -7 V and a gate voltage (V_{g}) of -1.15 V. The impulse prestress treatment was performed by applying 15 to 150 high-voltage (19.8 V across the oxide) pulses to the substrate electrode (gate grounded) of the p^+/pMOSTs, with the source and drain floated.



(a)

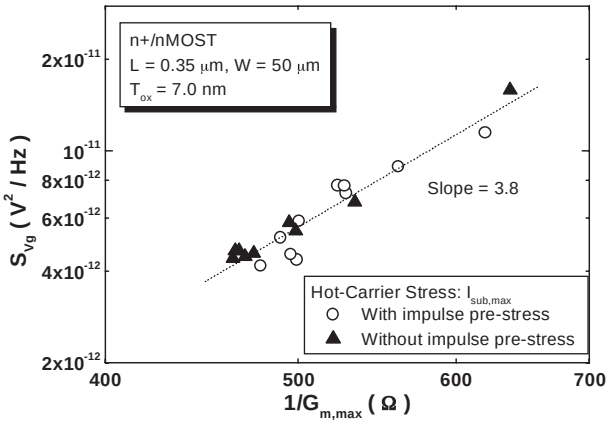


(b)

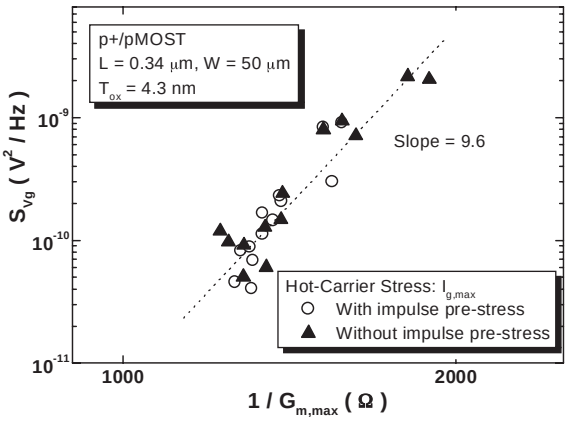
Fig. 6. Relationship of the interface trap density (δN_{it} or $|\delta J_{\mathrm{base,peak}}|$) and oxide trap charge density (δN_{ot}) to the absolute fractional change in the maximum transconductance ($|\delta G_{\mathrm{m,max}}/G_{\mathrm{m,max}}|$). Impulse stresses of varying degrees were applied to (a) n^+/nMOSTs (gate length $L = 0.25$, 0.35 and 0.5 μm; gate width $W = 50\,\mu$m and oxide thickness $T_{\mathrm{ox}} = 7$ nm) and (b) p^+/pMOSTs ($L = 0.24$, 0.34, 0.5 and 1 μm; $W = 50\,\mu$m and $T_{\mathrm{ox}} = 4.3$ nm). The maximum transconductance was measured at a drain voltage magnitude $|V_{\mathrm{d}}|$ of 0.1 V.

measured at a drain voltage magnitude ($|V_{\mathrm{d}}|$) of 0.1 V. Impulse stresses of varying degrees were applied to n^+/nMOSTs ($T_{\mathrm{ox}} = 7$ nm) and p^+/pMOSTs ($T_{\mathrm{ox}} = 4.3$ nm) of varying gate lengths, the results of which are shown in Figs. 6(a) and 6(b), respectively. The impulse stress was chosen over the hot-carrier stress because it produces a wider range of variation in $G_{\mathrm{m,max}}$. It is seen that there is a linear relationship between δN_{it} (or $|\delta J_{\mathrm{base,peak}}|$) and $|\delta G_{\mathrm{m,max}}/G_{\mathrm{m,max}}|$ and also between δN_{ot} and $|\delta G_{\mathrm{m,max}}/G_{\mathrm{m,max}}|$. The positive values of δN_{ot} denote hole trapping, which occurred during the high-field impulse stress. This is consistent with results obtained using C–V and I–t measurements on MOS capacitor structures subjected to high-field impulses applied to the gate electrode.[3] It is also known that neutral electron traps are also generated during the high-field impulse stress.[3] However, these neutral traps have to be filled by an electron injection phase before they have an effect on the electrical characteristics of the device. The multitude of trap formation and filling do complicate the detection of latent damage generated by the impulse stress as it is difficult to set a criterion for latent damage based on the densities of the measured traps. An alternative is to monitor the $G_{\mathrm{m,max}}$ degradation with the impulse stress, as $\delta G_{\mathrm{m,max}}$ encompasses the combined effect of the generated interface states and bulk traps (i.e., hole traps and filled neutral electron traps). However, the change in $G_{\mathrm{m,max}}$ may not be sufficiently sensitive for monitoring small levels of latent damage induced by the high-field impulse stress, especially in thin-oxide devices. The sensitivity limitation may be overcome by using low-frequency noise measurements for latent damage detection.

Figures 7(a) and 7(b) show the relationship between the input-referred noise power ($S_{V_{\mathrm{g}}}$ measured at frequency $f = 10$ Hz) and the reciprocal of the maximum transconductance ($1/G_{\mathrm{m,max}}$) on a log–log plot for n^+/nMOSTs



(a)



(b)

Fig. 7. The relationship between the input-referred noise power (S_{V_g} measured at frequency $f = 10$ Hz) and the reciprocal of the maximum transconductance ($1/G_{m,max}$) on a log–log plot for (a) n^+/nMOSTs (gate length $L = 0.35\,\mu$m; gate width $W = 50\,\mu$m and oxide thickness $T_{ox} = 7$ nm) and (b) p^+/pMOSTs ($L = 0.34\,\mu$m; $W = 50\,\mu$m and $T_{ox} = 4.3$ nm) after varying degrees of hot-carrier stress [i.e., $I_{sub,max}$ stress in (a) and $I_{g,max}$ stress in (b)]. Devices with and without the impulse prestress treatment are indicated by open circles and solid triangles, respectively. The dotted line indicates the best fit to the measured data points. A slope greater than unity was observed in both (a) and (b) indicating the greater sensitivity of low-frequency noise measurements as compared to $G_{m,max}$ measurements.

($L = 0.35\,\mu$m, $W = 50\,\mu$m and $T_{ox} = 7$ nm) and p^+/pMOSTs ($L = 0.34\,\mu$m, $W = 50\,\mu$m and $T_{ox} = 4.3$ nm), respectively, after applying varying degrees of hot-carrier stress (i.e., $I_{sub,max}$ stress for n^+/nMOSTs and $I_{g,max}$ stress for p^+/pMOSTs). It is observed that S_{V_g} increases as $G_{m,max}$ decreases (or $1/G_{m,max}$ increases). The dotted line indicates the best fit to the measurement results. A slope greater than unity was observed in both Figs. 7(a) and 7(b). The significance of this greater than unity slope will be discussed below.

Figure 8 shows the gate-length (L) dependence of the power-law relationship between S_{V_g} (measured at $f = 10$ Hz) and $1/G_{m,max}$ obtained from p^+/pMOSTs ($T_{ox} = 4.3$ nm and $W = 50\,\mu$m) of different gate lengths ($L = 0.24$, 0.34, 0.5 and $1\,\mu$m). The low-frequency noise measurements were performed on the p^+/pMOSTs before and after different degrees of impulse stress. Again, it is observed that S_{V_g} increases

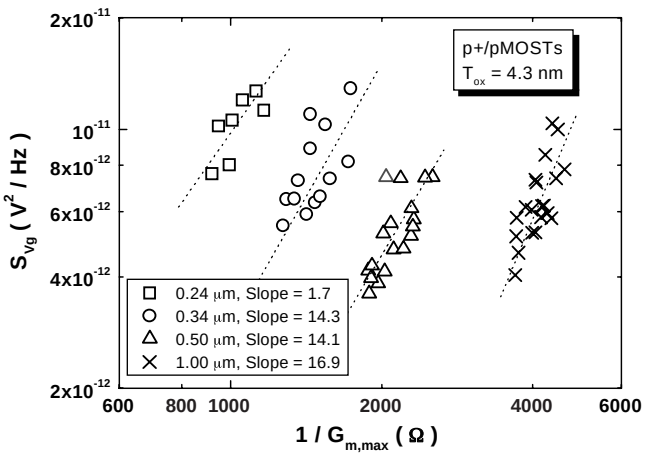


Fig. 8. Gate-length (L) dependence of the power-law relationship between the input-referred noise power (S_{V_g} measured at frequency $f = 10$ Hz) and the reciprocal of the maximum transconductance ($1/G_{m,max}$) obtained from p^+/pMOSTs ($L = 0.24$, 0.34, 0.5 and $1\,\mu$m; $W = 50\,\mu$m and $T_{ox} = 4.3$ nm). The low-frequency noise measurements were performed on the p^+/pMOSTs before and after different degrees of impulse stress. A straight line was fitted to each gate length and a slope greater than unity was obtained for all cases.

as $G_{m,max}$ decreases (or $1/G_{m,max}$ increases). A straight line was fitted to each gate length and a slope greater than unity was obtained for all cases. The slopes are however different for different gate lengths, indicating that the relative sensitivity of low-frequency noise measurements over that of $G_{m,max}$ measurements is not a constant but varies with gate length. The exact reason for this is not clear at the moment, although it is possibly related to some short-channel effect.

It has been reported that interface traps and border oxide traps generate flicker noise[12,13] in accordance with either McWhorter's carrier number fluctuation theory[14] or Hooge's mobility fluctuation mechanism.[15] The unified flicker noise model proposed by Hung and co-workers incorporates both the number fluctuation and surface mobility fluctuation mechanisms as[16]

$$S_{V_g}(f) = \frac{kTq^2}{\gamma f W L C_{ox}^2}(1 + \alpha\mu N)^2 N_t(E_{fn}), \qquad (1)$$

where k is the Boltzmann constant, T is the temperature, q is the electronic charge, γ is the attenuation coefficient of the electron wave function in the oxide, f is the measurement frequency, W is the gate width, L is the gate length, C_{ox} is the oxide capacitance per unit area, α is the Coulomb scattering coefficient, μ is the carrier mobility, N_t is the oxide trap density and E_{fn} is the quasi-fermi level of electrons. Since N is proportional to $C_{ox}(V_g - V_{th})$, the input referred noise power has a term proportional to C_{ox}^2 (the number fluctuation model) and another proportional to $C_{ox}^{-1}(V_g - V_{th})$. The measurement of S_{V_g} was performed at a constant effective gate voltage ($|V_g - V_{th}|$) so as to minimize the bias dependency of S_{V_g} as the threshold voltage V_{th} of the device changes with electrical stressing. It is seen from Figs. 7 and 8 that S_{V_g} has a power-law relation to $1/G_{m(max)}$ which can be expressed as

$$S_{V_g} = A\left[\frac{1}{G_{m(max)}}\right]^B, \qquad (2)$$

where A and B are constants for a particular gate length. The values of A and B for each gate length can be obtained from

fitting a straight line to the measurement results. It was found that the value of B, or the slope of the best-fit straight line on a log–log plot, ranges from 1.7 to 16.9 based on the results in Figs. 7 and 8. Equation (2) can be differentiated once to obtain the following equation:

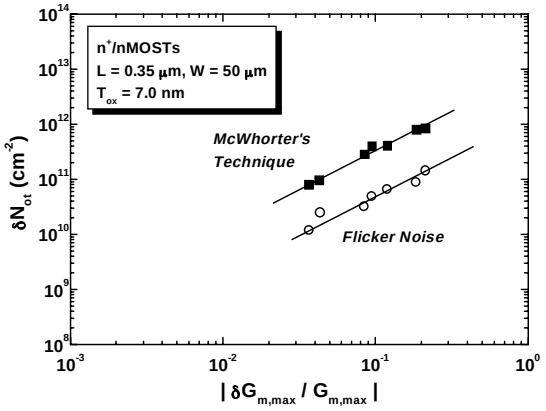
$$\frac{\delta(S_{V_g})}{S_{V_g}} = -B\frac{\delta(G_{m(max)})}{G_{m(max)}}. \quad (3)$$

It is seen from eq. (3) that the percentage change in the input-refered noise power, $\delta(S_{V_g})/S_{V_g}$, is larger than the percentage change in the maximum transconductance, $\delta(G_{m(max)})/G_{m(max)}$, by the factor B. Since B is larger than unity, it implies that the characterization of the latent damage arising from high-field impulse stress can be performed more sensitively using low-frequency noise measurements as compared to monitoring the degradation in the maximum transconductance.

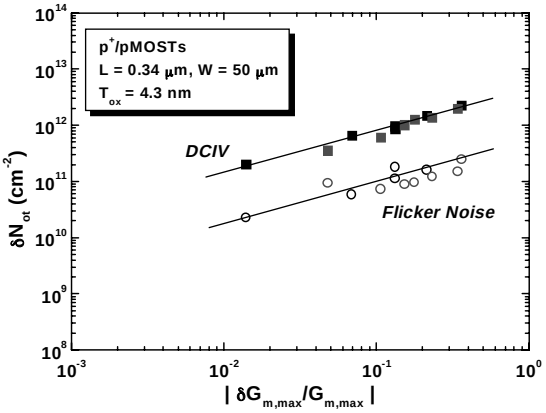
The amount of impulse-stress generated traps (δN_t) was also obtained from low-frequency noise measurements. δN_t was calculated using eq. (1) by noting the change in the noise magnitude (δS_{V_g}) after the impulse stress. Assuming the carrier number fluctuation mechanism is the dominant mode of flicker noise generation, which is reasonably true with a low gate voltage bias for a surface-channel device, eq. (1) reduces to

$$\delta S_{V_g}(f) = \frac{kTq^2}{\gamma f WLC_{ox}^2}\delta N_t(E_{fn}). \quad (4)$$

Using the flicker noise measured at $f = 10$ Hz for the impulse-stressed n^+/nMOSTs ($L = 0.35\,\mu$m, $W = 50\,\mu$m and $T_{ox} = 7$ nm) and p^+/pMOSTs ($L = 0.34\,\mu$m, $W = 50\,\mu$m and $T_{ox} = 4.3$ nm), and substituting γ as $10^8\,\text{cm}^{-1}$, the generated trap density, δN_t ($\text{cm}^{-3}\,\text{eV}^{-1}$), was calculated. Assuming that the traps contributing to low-frequency noise arise from a volume of the oxide of up to 2 nm from the interface, the calculated trap density is converted from per unit volume to per unit area (δN_{ot}) and plotted with respect to $|\delta G_{m(max)}/G_{m(max)}|$ in Fig. 9. It is seen that the trap densities estimated from flicker noise data are comparable in magnitude with those calculated using McWhorter's method and DCIV measurements. It is difficult to say how much the difference in the extracted oxide trap charge density using the different techniques is due to nonuniform damage generation or to the different measurement methods used. However, the second reason is more likely as flicker noise measurements result consistently in smaller oxide trap charge densities which is further explained below. Also, the results in Fig. 9 were obtained using varying degrees of impulse stress, for which nonuniform damage generation may not be as serious as that under hot-carrier stress. It is to be noted that the slopes of the plots in either Figs. 9(a) or 9(b) are quite similar suggesting a reasonable correlation between the flicker noise measurement and either measurements from McWhorter's method or the DCIV technique. The different slopes between Figs. 9(a) and 9(b) could possibly be a result of a difference in oxide thickness, assuming that the quality of the oxide-silicon interface of the different device structures [i.e., n^+/nMOSTs in Fig. 9(a) and p^+/pMOSTs in Fig. 9(b)] are similar. Although the assumption of similar interface quality is difficult to ensure and to verify, the smaller slope for the thinner oxide device in Fig. 9(b) could be an indication of the reduced charge trapping in the thinner oxide. It is worthwhile to note that both McWhorter's method and DCIV measurements estimate trap densities throughout the entire oxide. The low-frequency noise calculation, on the other hand, only accounts for the active traps very near the interface (i.e., "border traps") and also assumes a uniform oxide trap distribution. It is therefore not surprising that the oxide trap densities calculated from low-frequency noise data are much smaller than those calculated using McWhorter's method or DCIV measurements.



(a)



(b)

Fig. 9. Comparison of oxide trap charge density (δN_{ot}) calculated from low-frequency (frequency $f = 10$ Hz) noise measurements to that obtained using (a) McWhorter's method and (b) DCIV measurements. Calculations of δN_{ot} were performed on (a) n^+/nMOSTs (gate length $L = 0.35\,\mu$m; gate width $W = 50\,\mu$m and oxide thickness $T_{ox} = 7$ nm) and (b) p^+/pMOSTs ($L = 0.34\,\mu$m; $W = 50\,\mu$m and $T_{ox} = 4.3$ nm) subjected to varying degrees of impulse stress.

4. Conclusion

This study has shown that latent damage in thin oxides, caused by high-field impulse stress applied to the gate electrode, has a significant impact on the subsequent trap generation within the device under hot-carrier stressing but not on the charge-to-breakdown Weibull distribution. Characterization of such latent damage could be carried out more sensitively using low-frequency noise measurements than with the conventional method of monitoring the change in the maximum transconductance of the transistor device.

Acknowledgements

This project is supported under the National University of Singapore Research Project RP3992677 (R-263-000-094-112). The authors are grateful to Chartered Semiconductor Manufacturing Ltd. for providing the test wafers.

1) A. Amerasekera and C. Duvvury: *ESD in Silicon Integrated Circuits* (Wiley, Chichester, 1995) p. 181.
2) S. Aur, A. Chatterjee and T. Polgreen: IEEE Trans. Electron Devices **35** (1988) 2189.
3) W. K. Chim and G. L. Teh: Jpn. J. Appl. Phys. **37** (1998) 1671.
4) P. J. McWhorter and P. S. Winokur: Appl. Phys. Lett. **48** (1986) 133.
5) A. Neugroschel, C. T. Sah, K. M. Han, M. S. Carroll, T. Nishida, J. T. Kavalieros and Y. Lu: IEEE Trans. Electron Devices **42** (1995) 1657.
6) W. K. Chim and P. S. Lim: IEEE Trans. Electron Devices **47** (2000) 473.
7) W. K. Chim, B. P. Yeo, P. S. Lim and D. S. H. Chan: IEEE Electron Device Lett. **19** (1998) 363.
8) P. S. Lim and W. K. Chim: Appl. Phys. Lett. **77** (2000) 2719.
9) W. K. Chim, P. S. Lim, B. P. Yeo, D. S. H. Chan and G. L. Teh: Ext. Absr. Int. Conf. Solid State Devices and Materials, 1999, p. 334.
10) T. Nigam, R. Degraeve, G. Groeseneken, M. M. Heyns and H. E. Maes: IEEE Proc. Int. Reliability Phys. Symp., 1998, p. 62.
11) A. Martin, P. O'Sullivan and A. Mathewson: Microelectron. Reliab. **37** (1997) 1045.
12) M. Aoki and M. Kato: IEEE Electron Device Lett. **17** (1996) 118.
13) M. H. Tsai and T. P. Ma: IEEE Electron Device Lett. **14** (1993) 256.
14) S. Christensson, I. Lundstrom and C. Svensson: Solid-State Electron. **11** (1968) 797.
15) F. N. Hooge: Physica B **83** (1976) 14.
16) K. K. Hung, P. K. Ko, C. Hu and Y. C. Cheng: IEEE Trans. Electron Devices **37** (1990) 654.