Avalanche injection of hot holes in the gate oxide of LDMOS transistors

S. Manzini *, A. Gallerano

TPA Groups, STMicroelectronics, Via Tolomeo, 1, 20010 Cornaredo, Milano, Italy

Abstract

A new degradation mode of n-channel lateral double-diffused MOS (LDMOS) transistors has been investigated. The degradation, resulting in a large increase of the drain saturation current at the onset of strong inversion, is attributed to avalanche-generated hot holes injected and trapped in the gate oxide above the n-type drift region of LDMOS transistors operating at a high drain voltage and a low gate voltage near threshold. Worst-case static gate-bias condition, drain voltage dependence, maximum operating drain voltage, and the effect of varying some geometrical parameters of the device are studied. A method, based on gate-to-drain capacitance measurements, to characterize the spatial extension of the damaged region and the amount of trapped holes, is presented. © 2000 Published by Elsevier Science Ltd. All rights reserved.

1. Introduction

Hot-carrier effects in MOSFETs have received much interest for at least a quarter of the century [1] and are well known to impose severe limitations to the long-term reliability of scaled devices. For pure logic circuits fabricated in standard CMOS technology, the scaling of the devices is generally accompanied by the scaling of the supply voltage which partially alleviates hot-carrier effects. The scenario for power integrated circuits is somehow different. Circuits for smart power applications require a combination of high-voltage, double-diffused MOS (DMOS) power transistors, low-voltage, dense CMOS signal blocks and, possibly, precision bipolar transistors for analog stages. On the one hand, the supply voltage and therefore the voltage rating of high-voltage DMOS transistors, being a constraint of the specific application, is not scalable, while on the other hand, progress in microlithography allows a drastic reduction of the drain-source on-state resistance of the power DMOS transistors integrable in a given silicon area. These two opposite issues are expected to drive DMOS transistors merged in VLSI mixed processes [2] towards hot-carrier limitations.

Hot-carrier effects have been evidenced recently in n-channel lateral/vertical DMOS transistors (LDMOS/VDMOS) operating at large overdrive voltage [3-6] and have been attributed to hot-electron injection and trapping in the gate oxide near the source end of the channel. In this article, a new type of degradation induced by hot carriers, affecting high-voltage LDMOS transistors operating near threshold voltage is investigated. This degradation mode limits the maximum operating drain voltage the devices can withstand in the on-state while maintaining the drift of the drain saturation current within prefixed values for the useful life of the circuit.

In n-channel LDMOS and conventional n-channel MOS transistors, the kind of degradation induced by hot carriers is distinctly different. In fact, the two structures are somehow complementary. While the body/drain junction is “shallow” in scaled n-channel MOS transistors, the body/drain junction is virtually infinite in n-channel LDMOS transistors. This causes the current path and the carrier multiplication spot to be much less confined near the Si/SiO₂ interface which, in turn, reduces the probability of hot carrier injection into the gate oxide. This implies, for example, that a large multiplication current, carried by secondary holes generated by impact ionization in silicon, does not necessarily cause a large degradation of the Si/SiO₂ interface or that the degradation of the device is not correlated with the peak body current [7]. Another substantial
difference is that the doping concentration of the drain diffusion is generally lower than the doping concentration of the body diffusion and the area of the drain diffusion under the gate oxide is comparable or larger than the area of the body diffusion. From an experimental point of view, these two circumstances allow one to study the damage generated by hot carrier at the Si/SiO₂ interface when hot holes are avalanche-injected into the oxide.

2. Devices and experiments

The devices which are characterized, schematically shown in Fig. 1, are LDMOS transistors with LOCOS-terminated field plate, integrable in a 0.6 μm Bipolar-CMOS–DMOS mixed process [2]. The thickness of the gate oxide is 180 Å. The p-body of the intrinsic n-channel enhancement-mode transistor is formed by a large tilt angle implantation technique [2] to avoid high-temperature, long-time diffusion thermal steps incompatible with the simultaneous realization of short-channel CMOS transistors. The intrinsic effective channel length, Lch, as measured by capacitance–voltage methods [8] and transmission electron microscopy, is 0.54 μm. The peak doping concentration of the p-body diffusion near the source end of the channel is in the low 10¹⁷ cm⁻³ range. The extrapolated threshold voltage in the linear region (drain voltage, V_D = 100 mV) is about 1 V. Devices with gate oxide length, L_G, between the end of the poly-Si gate and the end of the active area from 1.0 to 1.8 μm have been evaluated. The source width, W, ranges from 16 to 3000 μm. The devices with W = 3000 μm have a conventional multi-finger (interdigitated source–drain) layout. However, no dependence of the degradation rate on W and layout is observed. The doping concentration of the drain (n-well) region is about 4 x 10¹⁶ cm⁻³. The drain-source breakdown voltage (gate grounded) ranges between 24 and 26 V, slightly increasing with L_G and independent of W. Most of the devices have the n⁺-source diffusion internally shorted to the p-body diffusion in order to increase the snap-back voltage. Some devices with separate contacts to n⁺-source and p-body diffusion are used to measure the p-body avalanche current.

Bias-stress experiments are performed at a constant gate voltage, V_G, and a constant drain voltage, V_D. The drain current, I_D, is monitored continuously during the bias-stress and the degradation of I_D, Δ, is defined as the relative (percent) increase of I_D referred to its initial value as measured immediately after V_G and V_D are applied. The time-to-fail (time-to-drift), τ, corresponding to an arbitrary degradation of I_D of p percent, is implicitly defined by Δ(τ) = p. This τ is a function of V_D and V_G applied during the bias-stress. Also, the threshold voltage in the linear region, V_G(τ), the drain-source on-state resistance, R_s = V_D/I_D at V_G = 5 V, and the gate-to-drain capacitance, C GD, are monitored at exponentially spaced time-steps. C GD vs. V_G characteristics are measured at 1 MHz with the d.c. bias and the a.c. small signal applied to the gate, the drain connected to the virtual ground of the lock-in phase amplifier, and the common n⁺-source/p-body terminal shorted to ground.

Measurements are performed at the wafer level, in the dark, and at the controlled temperature of 27 ± 0.1°C.

3. Degradation of the drain saturation current

A preliminary set of experiments is performed in order to establish the worst-case (static) gate-bias condition. The time-to-fail at p = 10% is measured as a function of V_G for a few values of V_D. Typical results are reported in Fig. 2. The gate voltage, hereafter referred to as V G(min), corresponding to the minimum of τ occurs at V_G ≈ 0.95 V, i.e., at V_G ≈ V T,ch. V G(min) – the worst-case gate-bias condition – is independent of V_D and L_G. All the subsequent measurements are performed at V_G = V G(min), i.e., with the device in strong saturation (I_D ≡ I D,sat) at low I D,sat at the onset of strong inversion. The worst-case gate-bias condition for the present degradation mode is not correlated with the p-body current whose maximum value occurs at higher V_G.

The degradation of I D,sat is shown in Fig. 3 as a function of the stress time, t, with V_D as a parameter. Δ is positive and increases logarithmically with t, at least for large values of Δ, or, alternatively, it can be represented with the more familiar power law, Δ ∝ t^α, often observed in n-channel MOS transistors [9], but with α ≪ 1. Note that Δ can attain considerably high values, of the order of 50% (measured up to about 70 h of cumulative stress.

Fig. 1. Schematic (not to scale) vertical section of a lateral DMOS transistor. L_G is the thin gate oxide between poly-Si gate edge and drain/LOCOS edge. p-type substrate and n⁺ buried layer(s) are not shown.
time at high $V_D$). No systematic shift of $V_{T,eh}$ is observed while a minor (a few percent) decrease of $R_{on}$ is observed. As shown in Fig. 4, $\tau$ as a function of $V_D$ can be fitted over several orders of magnitude to the empirical law:

$$\tau = \tau_0 \exp \left( \frac{V_0}{V_D} \right)$$

with $V_0$ basically independent of $L_G$. The data shown in Fig. 4 are average values – in the sense of log-normal distributions with standard deviations of about 0.1 decade – of measured, not extrapolated $\tau$’s. From Fig. 4, the maximum operating drain voltage, $V_{D,max}$, in the worst-case gate-bias condition can be extrapolated to the target device life time. $V_{D,max}$ is proportional to $L_G$ in the range of $L_G$, which has been investigated.

4. Degradation of the gate-to-drain capacitance

Fig. 5 shows two $C_{GD}$ vs. $V_G$ characteristics before and after a static bias-stress at $V_G = V_{G,min}$ and $V_D = 20$ V. A generalized increase of $C_{GD}$ is observed up to, relatively, large negative $V_G$. Fig. 6 shows the modification of the $dC_{GD}/dV_G$ vs. $V_G$ characteristics obtained by direct numerical differentiation of the data in Fig. 5. The
flat-band voltage, \( V_{FB} \), and threshold voltage, \( V_T \), in the n-well region are, operatively, defined as the gate voltage corresponding to the two peaks of the \( dC_{GD}/dV_G \) vs. \( V_G \) curve. The error associated with this operative definition of \( V_{FB} \) and \( V_T \) can be calculated from the classical MOS theory [10]. Each peak of the virgin device, referred to as \( V_{FB}^{(0)} \) and \( V_T^{(0)} \), respectively, evolves separately into two distinct peaks, labeled 1 and 2 in Fig. 6. The corresponding flat-band voltage and threshold voltage after stress will be denoted by \( V_{FB}^{(1)} \), \( V_{FB}^{(2)} \), \( V_T^{(1)} \), and \( V_T^{(2)} \) slightly increase indicating a minor generation of negative charge (trapped electrons) in the gate oxide and/or at the Si/SiO₂ interface while \( V_{FB}^{(2)} \) and \( V_T^{(2)} \) show a large shift towards more negative values indicating a substantial generation of positive charge (trapped holes). A full set of \( dC_{GD}/dV_G \) vs. \( V_G \) characteristics as changing in time is shown in Fig. 7. The time evolution of the four peaks is shown in Fig. 8. The difference \( V_T^{(2)} - V_{FB}^{(2)} \) (\( \approx -1.4 \) V) is nearly constant in time which also implies that the surface state density, integrated over the Si band gap between flat-band and threshold is constant in time in the n-well region, where injection and trapping of hot holes occurs. The (equivalent) positive charge per unit area at the Si/SiO₂ interface, associated with the trapped holes, is given by:

\[
\sigma^+ = \frac{\varepsilon_0}{t_{ox}} [V_{FB}(0) - V_{FB}^{(2)}] = \frac{\varepsilon_0}{t_{ox}} [V_T(0) - V_T^{(2)}].
\]

The spatial extension of the hole trapping region can be determined from the \( C_{GD} \) characteristics, making some simplifying assumptions. So far, in fact, only a partial information – relative to the position and to the shift of the inflexion points along the \( V_G \)-axis – has been used. We neglect electron trapping and we attribute the increase of \( C_{GD} \) after stress in Fig. 5 to hole trapping only. The total gate-to-drain capacitance of the virgin device is approximated as the parallel combination of three terms:

\[
C_{GD} = C_1L_1W + C_2L_2W + C_3,
\]

where \( L_1 + L_2 = L_G - L_{ch} - x_{JL} \) is the length of the n-well region under the gate oxide and \( x_{JL} \) is the lateral n⁺-source/p-body junction depth. We estimate \( x_{JL} = 0.1 \mu \text{m} \).
and therefore $L_1 + L_2 = 0.36 \ \mu m$ if $L_G = 1.0 \ \mu m$. $C_3$ is any parasitic capacitance, including poly-Si/bird’s beak/n-well, poly-Si/field oxide/n-well, fringing fields, and gate-to-drain metallization capacitances. $C_1$ and $C_2$ are specific (per unit area) capacitances. $C_1$, $C_2$, and $C_3$, in general, depend on $V_G$. The total gate-to-drain capacitance of the stressed device is approximated by:

$$C_{GD}^{(s)} = C_1 L_1 W + C_2^{(s)} L_2 W + C_3.$$  

$L_2$ is therefore the spatial extension of the hole trapping region in a box approximation. The integral:

$$Q^+ = \int_{-\infty}^{+\infty} (C_{GD}^{(s)} - C_{GD}) \, dV_G$$

$$= L_2 W \int_{-\infty}^{+\infty} (C_2^{(s)} - C_2) \, dV_G$$

(2)

is the total equivalent charge at the Si/SiO$_2$ interface generated by the bias-stress. In Ref. [6], the above statement, expressed by Eq. (2), was demonstrated empirically via numerical simulations. From Eqs. (1) and (2), $L_2$, in terms of measured quantities, is given by

$$L_2 = \frac{1}{W} \frac{Q^+}{\sigma^+}$$

$$= \frac{1}{W} \frac{t_{ox}}{C_{ox}} \frac{1}{V_{FB}(0) - V_{FB}^{(2)}} \int_{-\infty}^{+\infty} (C_{GD}^{(s)} - C_{GD}) \, dV_G,$$

(3)

and plotted in Fig. 9. Neglecting electron trapping causes $Q^+$ and then $L_2$ to be underestimated, which, in turn, causes the weak sub-logarithmic behavior of both $Q^+$ and $L_2$ at long stress time.

5. Discussion and conclusions

Fig. 9 shows that the hole trapping region extends over a significant portion ($\approx 0.15 \ \mu m$) of the n-well region as $L_2 \approx L_1$ after a bias-stress of about 70 h. However gate-to-drain capacitance measurements alone are not sufficient to univocally identify the portion of the n-well region – either near the p-body/n-well junction or near the LOCOS/gate oxide edge – where hole trapping takes place. Some insight about the location of hot hole injection can be achieved by a numerical simulation of the schematic device shown in Fig. 1 biased in the worst-case condition. The physical models and the solution of the transport equation in hot carrier regime are themselves an intricate problem. However, we used device simulation (in the so-called hydrodynamic approximation [11]) only as a qualitative tool to identify the most likely point of hot-hole injection. The component of the hole current density, normal to the Si/SiO$_2$ interface (also in the bird’s beak region) and evaluated at the Si/SiO$_2$ interface is found to be the largest at the gate oxide/field oxide encroachment (Fig. 10).

Avalanche injected holes are known to be readily trapped in SiO$_2$ with an initial trapping efficiency in excess of 99% [12], due to their large capture cross-section in SiO$_2$. The local decrease of the flat-band voltage, in the high injection region, acts as a negative feedback mechanism and limits the hot-hole current at the Si/SiO$_2$ interface. This is known [13] to cause the development of a front of (positive) charge propagating from the point of initial peak injection ($x = 0$) towards the source with a characteristic logarithmic time dependence. The local positive charge causes the n-well region to be less depleted with the device in the saturation

![Fig. 9. Time evolution of the spatial extension of the hole trapping region.](image)

Fig. 10. Component of the hole current density normal to the Si/SiO$_2$ interface, evaluated at the interface, as a function of the x-coordinate along the interface. $x = 0$ at the gate oxide/field oxide edge. $J^+$ is positive in the direction from silicon to silicon dioxide.
mode and more accumulated with the device in the linear mode. This justifies the observed increase of drain current after stress in both modes of operation, whereas, the p-body region is not affected by charge trapping and the intrinsic threshold voltage of the device, \( V_{T,\text{th}} \), does not change. Fig. 9 shows that the growth of the extension of the trapping region is apparently unbounded and, in principle, can invade the p-body region causing a reduction of the saturation threshold voltage of the device – a short channel effect – and eventually an increase of drain leakage current in the off-state (gate grounded) due to hot-hole-induced punchthrough. In this respect, the present failure mechanism is similar but complementary to the failure mechanism reported for conventional p-channel MOS transistors [14]. This degradation mode is therefore of special interest for LDMOS transistors in analog stages (e.g., high-voltage LDMOS in a current mirror configuration) even if it can affect the long-term reliability of LDMOS-based power switches causing hot-hole-induced punchthrough in devices with poor margin against punchthrough itself. However, the devices which have been evaluated show neither any increase of the drain leakage current in the off-state nor any significant degradation of the sub-threshold characteristic.

Also, the degradation (relative increase) of the drain-to-gate capacitance, sometimes referred to as \( C_{\text{RSS}} \), during a static bias-stress at \( V_D = 20 \text{ V} \) and \( V_G = V_{G,\text{min}} \) has been evaluated. \( C_{\text{RSS}} \) is measured in the same configuration as \( C_{GD} \), but with the gate and drain terminals interchanged. The two measurements, indeed, gave the same result in terms of \( L_2 \). The degradation of \( C_{\text{RSS}} \) may cause, in principle, a long-term deterioration of the performance of LDMOS transistors operating at a high frequency.

The present degradation mode is a true hot-carrier effect taking place with the device in the on-state at \( V_G \approx V_{T,\text{th}} \approx V_{G,\text{min}} \) at room temperature. However, the same degradation mode may occur at a high temperature with the device in the off-state in a classical, long-term, high-temperature reverse-bias stress test (HTRB, gate grounded, drain at high voltage). In an HTRB stress test, primary electrons, thermally generated in the n-well depletion region, are accelerated in a configuration of electric potentials and fields similar to the room-temperature, \( V_G \approx V_{G,\text{min}} \) case. A small increase of the gate-to-drain capacitance after a 1000 h HTRB stress test has been observed [15] in field-oxide terminated LDMOS transistors similar to those characterized in this work.

References