Characterization of the diode leakage current in advanced 0.12 μm CMOS technology
(Does stress play a role?)

Willem-Jan Toren[1], Emmanuel Perrin[2], Meindert Lunenborg[1]
[1]Philips semiconductors, 38921 Crolles France email: willem-jan.toren@philips.com
[2]STMicroelectronics, 38921 Crolles France

Abstract
Anomalous N+/Pwell diode leakage current in advanced 0.12 um CMOS technology is investigated. This current is tunnel enhanced, and geometry dependent. Doping level variations can explain this current behaviour, as well as other width scalable deviations like the mobility. Stress in STI is the most probable cause of these doping variations.

Introduction
The dimensions of the transistors of the today state of the art CMOS technology (0.1 um node and below) are reduced to such dimensions that new parasitic effects like reduced saturation current and strongly increased leakage current become visible. Smeys et. al. [1] is contributing the increase of leakage current to bandgap narrowing caused by stress. Steegen et. al. [2] observes a relation between stress and leakage current without explaining the physical effect. Scott et. al. [3] shows the relation between NMOST drive current and stress is caused by mobility reduction caused by lattice deformation. The presence of stress causes by STI processing is evidenced by a large number of authors. However, the translation between stress and device characteristics is not always easy. This paper will show an increase of doping concentration for devices with higher stress by analyzing the leakage current. First, it will be shown that the linear geometry scaling rules are no longer valid, because smaller devices generate more leakage current. The temperature analyses indicates a trap assisted tunnel leakage current, which is confirmed by fitting the current with standard tunnel enhanced generation equations, with the doping concentration as fit parameter.

Set-up
The characterization is done on a set of 9 devices, all with the same gate length, different width, and different S/D length (See figure 1, table 1).

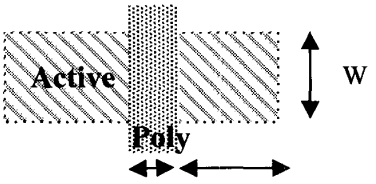

Figure 1: Layout of the transistors, for values of L and W see table 1

L \ W	0.2	2.68	5.56
0.4	1	4	7
3.0	2	5	8
5.12	3	6	9

Table 1. Dimensions of the 9 devices [um], numbers are the device number and serve as a reference for the results

In order to gain resolution, multiple devices were connected in parallel. The measured currents were directly normalized to one device in all cases. From these devices the diode leakage current is measured as a function of the reverse bias, by connecting only the drain and bulk contact. The results were verified and in agreement with the method proposed by [4].
The technology used is a 0.12 um CMOS for system on chip applications.

Model
Many papers reported already that the leakage current in today CMOS technologies is strongly enhanced by the electric field [5,6]. The physical idea behind the model is that the traps,

ISBN 4-89114-019-4/020-8

which normally generate the leakage current by a SRH mechanism, are assisted by a tunnel component. This means that the carriers do not need to have the total energy to cross the bandgap, as expected by the SRH theory. Therefore the leakage current increases.
The electric field of this tunnel-enhanced generation depends on the bias over the junction, and the depletion width. The latter depends on the doping concentration, which makes the relation leakage current vs. doping concentration.
The traps for the generation are interface states situated in a depleted region. In MOS transistors there are roughly two different regions, one under the gate, and two at the STI interface in the bulk (see figure 2).

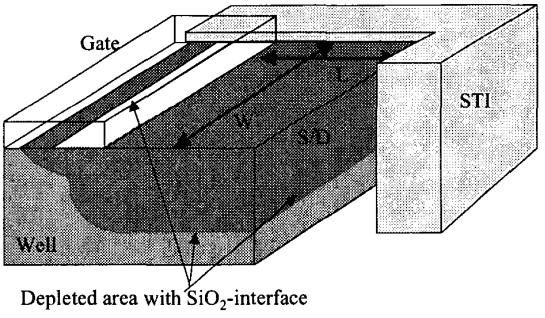

***Figure 2:** Cross-section of a transistor, to indicate the depleted areas adjacent to silicon-oxide. The first one is under the gate (which is drawn transparently), and the second one is besides the STI oxide.*

The surface leakage component of diodes (which is generated in the depleted S/D bulk region) is very low, which means that the bulk trap concentration is very low.

To extract the two different components of gate leakage current and STI leakage current a simple linear scaling model is used (1)

$$I_{tot} = I^{o}_{sti} * P_{sti} + I^{o}_{gate} * W \qquad (1)$$

I_{tot} = the total measured diode leakage current
I^{o}_{sti} = Generation current on STI edge per STI perimeter P_{sti} (2 * L + W)
I^{o}_{gate} = Generation current under the gate per transistor width W
In former technologies, I^{o}_{sti} and I^{o}_{gate} were geometry independent, which changes (as we will see) in the presented technology.

Results

Temperature behaviour
To investigate the nature of the diode leakage current, the activation energy of this current is plotted as a function of reverse bias for the smallest and the largest device (figure 3).

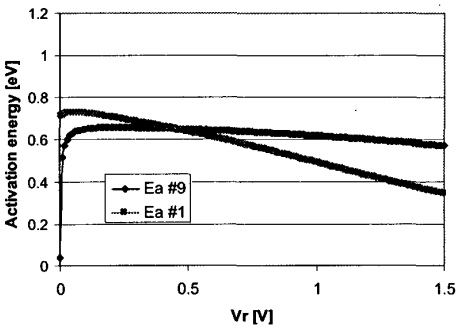

***Figure 3:** Activation energy of the diode leakage current as a function of the reverse bias for device 1 and 9*

For device 9 the Ea is about 0.6 eV, and constant with bias. This is also expected for standard Schotckey-Read-Hall (SRH) generation. However, for device 1 the Ea drops from about 0.6 eV to 0.3 eV. This is due to the tunnel enhanced leakage generation as is reported in [5,6]. At higher temperatures the carriers have enough energy to cross the bandgap, which makes the tunnel enhancement less effective, and causes so a lower activation energy. This proves that there is a different behaviour between small and large devices, not only in amplitude, but also on temperature behaviour.

Reverse bias dependence
The diode leakage current of the 9 NMOSTs is plotted in figure 5 (at the end of this paper) as a function of the reverse bias. The first thing that can be noticed is that the leakage current of a small device (e.g. device 1) is larger than for a large device (e.g. device 9). For this reason the linear scaling model cannot be applied anymore. For larger devices, this still can be applied, as is show in figure 5. The first fit (dashed black lines) is obtained by a least square fit using the 4 largest devices, 5,6,8,9. The fit and

measurements are in good agreement. The result is used to calculate the leakage current of the smaller devices using (1). It is clear that the smaller devices generate much more current, than was expected from linear scaling. If the linear fit was applied for all devices, an even worse fit would be obtained.
To model the leakage current more accurate, the tunnel enhanced generation model of [6] is used. The electric field is the reverse bias divided by the depletion width of the junction, which is calculated using standard equations.
If the well doping concentration is increased for the smaller devices, which means that the depletion width over the junction decreases, and so the electric field and the tunnel enhancement increases, the second fit is obtained (purple lines in figure 5), which fits very good for all the devices. The needed doping levels are plotted in table 2.

L \ W	0.28	2.68	5.56
0.4	2.3 / 0.9	1.66 / 0.42	1.66 / 0.35
3.0	1.78 / 0.37	1.35 / 0.29	1.35 / 0.29
5.12	1.66 / 0.37	1.35 / 0.29	1.35 / 0.29

Table 2. Doping levels (Ngate / Nsti *10^{18} cm^{-3}) of the 9 devices used to fit the date in figure 5.

This means that the leakage current can be fitted by two doping concentrations adjustments. The larger doping concentration of smaller devices compared with larger devices can be explained by stress in the silicon caused by STI processing. Capacitance measurements have shown a segregation of boron concentration at the STI border. A compressive stress will make this segregation less effective, because the diffusion will be slower, which leads to higher concentrations at the STI border. Also the diffusion of the boron under the gate will be impacted by the stress.

Extracted parameters
Figure 4 shows the leakage current per um of the STI region and under the gate for device 1 and 9. For low reverse bias values, the diode leakage, generated under the gate is much lower that the diode leakage close to the STI. Because the increase with reverse bias is not very high, we can simplify the model by just SRH generation. That means that the number of interface states in the depleted region

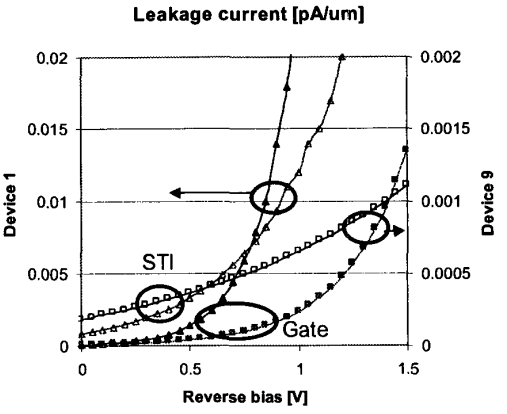

Figure 4: The extracted parameters Isti and Igate for two different devices.

under the gate is much less than at the STI interface. Causes for this are aggressive etch of STI trench, less cleaning possible, etc. However, if the reverse bias is increased, the gate current becomes larger than the STI current. Looking to the shape of the curves (a very large increase with the reverse bias) concludes that the current is tunnel enhanced, in particular for device 1. Because of the faster increase of the gate current, means a higher electric field, what is expressed in table 2 with a higher doping level.

Discussion

Doping level difference
The difference in well doping concentration between the devices is about a factor 3. This is well possible if a higher compressive stress squeezes the lattice, the diffusion of the doping atoms will be slowed down, which means that the doping level will remain higher. This higher doping level will also reduce the mobility as is reported in [1].

Reduction of the bandgap
[2] explains the increase of leakage current by a stress induced bandgap narrowing model. Indeed, if the bandgap is reduced, the reverse diode leakage current will increase. However, the bias dependence as is shown in this paper is not explained by this effect. Figure 4 shows that the STI current at low bias values is higher. At higher bias values, this is inverted. Trap assisted

ISBN 4-89114-019-4/020-8

leakage current explains this, where as bandgap narrowing cannot.
[1] compares the drive current of devices with more and less stress. High stress devices give a lower drive current which is explained by a lower mobility value due to stress. The results presented in this paper can also explain the lower mobility value by a higher doping concentration.

Conclusions

In this paper the N+/P-well diode leakage current is investigated. The found leakage behaviour can be explained by a tunnel enhanced generation process at the interface between Si/SiO_2 . The linear scaling model of the leakage current cannot be applied anymore, because smaller devices generate more current than larger ones. This is only because the tunnel enhancement in more effective for smaller devices at higher bias levels, due to a higher electric field. We propose a hypothesis that this is due to a higher doping level because of a less effective segregation of boron in the SiO_2 layer. Stress due to STI processing is a probable cause of this effect, because it squeezes the lattice, which slows down the diffusion. This also explains the lower mobility, which causes a lower saturation current.

Reference

[1] P. Smeys, et.al. IEEE trans. El. Dev. Vol. 46, no. 6, june 1999
[2] A. Steegen et. al. VLSI 2000
[3] G. Scott, et.al. IEDM 99-827
[4] Dachs, ESSDERC 2001, page 175
[5] Tse-en Chang et. al. IEEE trans. El. Dev., VOL 42, NO 4, April 1995, page 738
[6] F. Hurkx, IEEE Trans. El. Dev. Vol. 40, no. 12, pp. 2273, December 1993

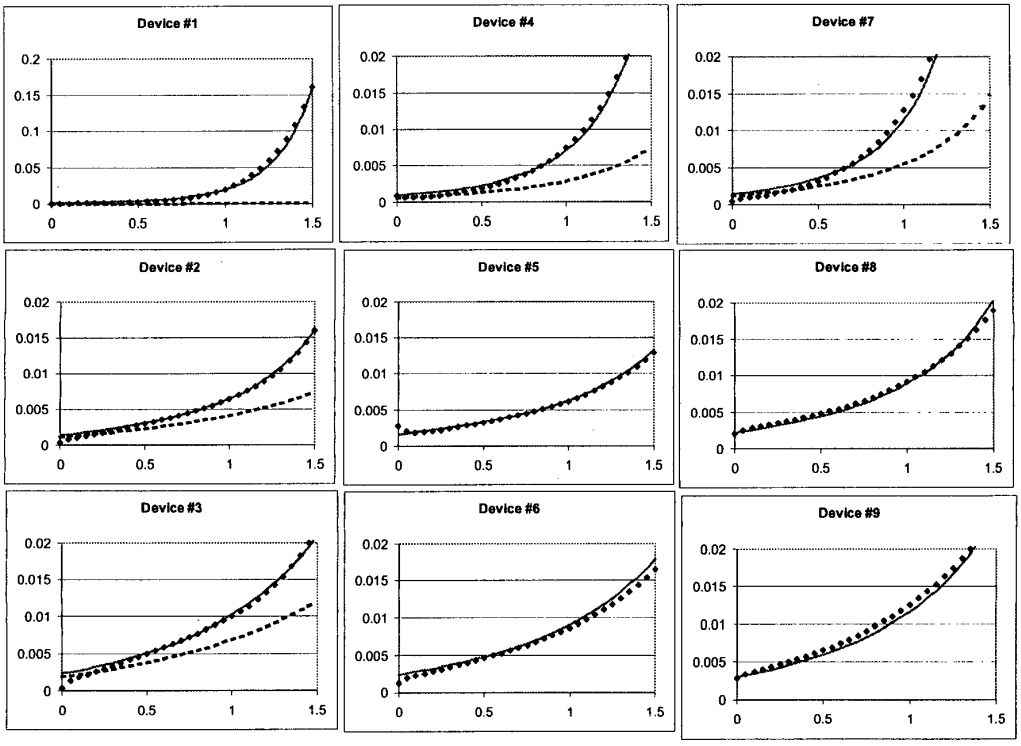

Figure 5: Diode leakage current NMOSTs (in pA) for the 9 different devices (table 1) as a function of the reverse bias. The blue marker lines are the measurements; the black dotted lines are the fit, using only the four largest devices, and calculate the contribution of the smaller devices. The calculation for device 1 gives values close to the 0 pA axes, and Is almost not visible. The purple lines are obtained from the same fit, but now the generation is adapted by changing the doping levels.