Gate current scaling rules characterization, an efficient tool for gate oxide optimisation in 0.12 µm CMOS technologies

Willem-Jan Toren
Philips semiconductors
willem-jan.toren@st.com

Franck Arnaud
STMicroelectronics
franck.arnaud@st.com

Abstract

In this paper, the gate current scaling rules are characterized. This analysis justifies a method to determine the gate oxide thickness variation over the wafer with an accuracy of about 0.2 Å. A simple but accurate model is applied, which also allows determining the total off-state power consumption of digital circuits.

1. Introduction

The today advanced CMOS technologies (0.12 µm and smaller) suffer, besides the classic transistor off-current, also from gate leakage current, which is also showed in [1,2]. This last current is due to a direct tunnel current through the ultra thin oxides (around 18 Å) [3,4]. For digital applications it is mainly this current that will determine the off-state power consumption. For analogue applications this effect can deteriorate the circuits performance (e.g. transistor matching because of a large variation in oxide thickness). In this paper the gate current is characterized with respect to both applications by determining the power consumption as a function of transistor dimensions and determining the dispersion of the oxide thickness. The devices are made in a conventional 0.12 µm CMOS technology, with gate oxide thickness in the range of 17 to 21 Å.

The gate current is measured in two modes. First the gate current, Ig_on, is measured with all terminals grounded, and the gate potential at VDD. The second mode measures the gate current, Ig_off, where all terminals are grounded, and the drain potential is at VDD. This is schematically shown in figure 1. Both modes are present in e.g. inverter stages.

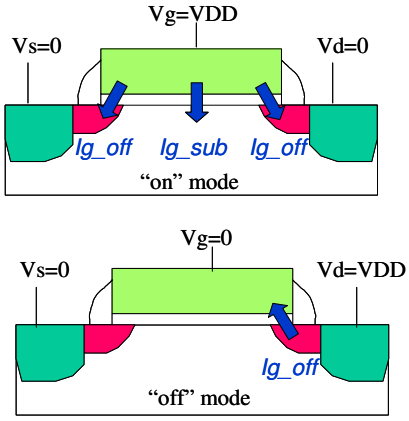

Figure 1. Schematical drawing of the gate current in two transistor modes

This paper will first characterize both gate currents. Then a model for the scaling rules is defined, which will be verified by the measurements. Finally the within wafer oxide thickness variations are calculated.

2. Gate current dependence

First for both modes the gate current relation with bias and oxide thickness is verified.

Gate current versus bias

Figure 2 shows the Ig_on and Ig_off for nMOSTs and pMOSTs as a function of the bias.

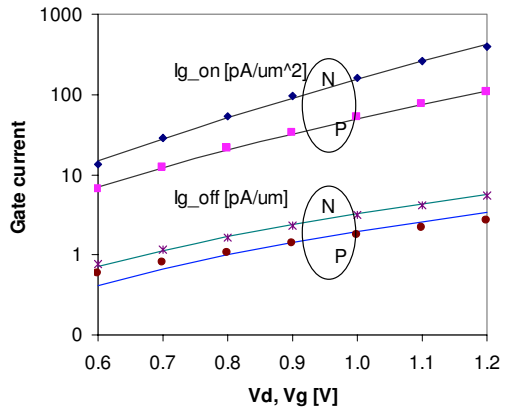

Figure 2. Ig_on and Ig_off as a function of the bias, measured on a 100/100 N and P MOS transistor

The lines represent the fit with the direct tunnel model [3,4]. It fits both modes, which is expected. The data plotted in this figure and also in the next figure are the average value of the measurement of about 30 dies. In paragraphs 3 it will be shown that the dispersion is very large.

Gate current versus Tox

Figure 3 shows the gate current in both modes, as a function of the oxide thickness. It is clear that the demand for a better control over the transistor channel by thinner gate oxide is paid by a very large increase in gate current.

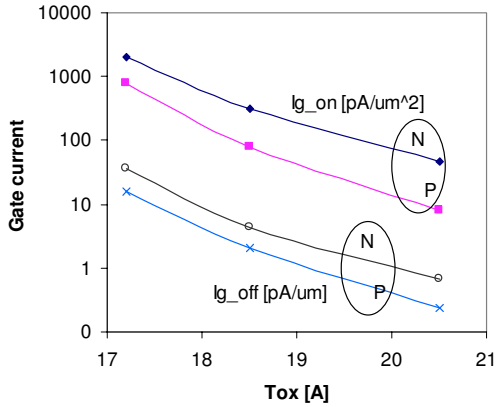

Figure 3. Ig_on and Ig_off as a function of oxide thickness, measured on a 100/100 N and P MOS transistor

3. Scaling rules of the gate current

This paragraph first shows the model for the scaling rules of the transistor gate current. Then results are shown from which the parameters have been extracted.

Model

To predict the total power consumption of digital circuits it is important to know the gate current as a function of its geometry. This is done with the help of figure 1, from which a simple, but adequate, model is derived (eq.1)

$$Ig^0 = Const \cdot f(tox, techno)\ [A/cm^2]$$
$$Ig\_on = Ig^0 \cdot W \cdot (L - 2\Delta L + 2\alpha \cdot \Delta L) \qquad (1)$$
$$Ig\_off = Ig^0 \cdot W \cdot \alpha \cdot \Delta L$$
$$Ig\_sub = Ig^0 \cdot W \cdot (L - 2\Delta L)$$

Ig^0 is the gate current density, which depends amongst others on the gate oxide thickness and process parameters (like LDD dose as is showed in this paper).

ΔL is the gate-LDD overlap, which depends on the LDD dose. ΔL is the

effective length, over which the maximum gate current flows.

α is the difference of gate current density between the substrate and the LDD region. The value of α will deviate from 1 because of different doping levels in the channel and LDD region. The measurements will show that Ig_off is equal for positive and negative (Vg-Vd). This model is applied on the measurements in the next paragraph.

Characterization

Figure 4 shows the Ig_on as a function of Ig_off for 10/0.16 NMOS transistors for three different splits in oxide thickness and for about 30 dies per split.

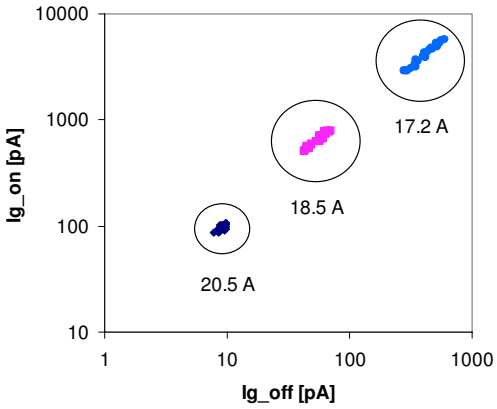

Figure 4. Ig_on versus Ig_off for three splits in oxide thickness, measured on 10/0.16 NMOSTs

The first observation is that there is a strong relation between Ig_on and Ig_off over a wide range of currents. This means that if the substrate current is large, because of a thinner oxide, the overlap current is also large. The large dispersion of the gate current is due to gate oxide thickness variations over the wafer. The next figure shows a zoom in the 18.5 group, for three different variants in LDD dose.

A remarkable increase is observed on the Ig_off for the different splits. This is explained in the model by a larger ΔL due to a larger LDD dose. Ig_on increases with different LDD doses because it also contains

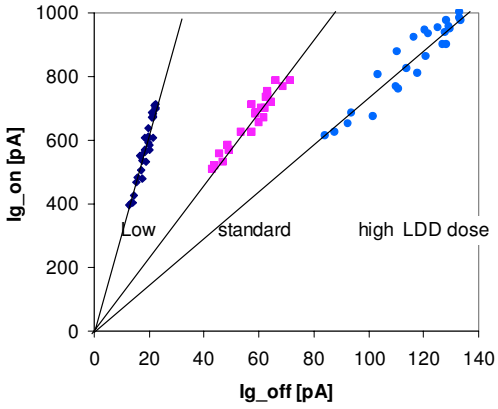

Figure 5. Ig_on versus Ig_off for three splits in LDD dose, measured on 10/0.16 NMOSTs

twice Ig_off (eq. 1). If for every die Ig_sub (Ig_on-2*Ig_off) is plotted versus Ig_off, it is seen that Ig_sub depends almost not on the LDD dose (see fig 6)

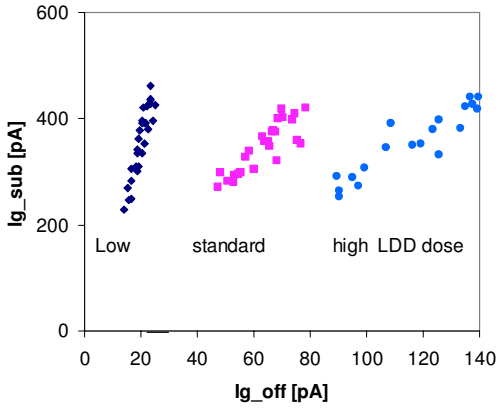

Figure 6. Ig_sub versus Ig_on for different splits in LDD dose, measured on 10/0.12 NMOSTs

The amplitude of Ig_sub still shows a large dispersion, but the mean value is constant for the three splits in LDD dose.

This means that the surface contribution is separated from the transistor-width contribution, such that the scaling rules can be determined. The fact that Ig_sub does not depend on the LDD dose variations also

means that the same Ig_off is measured with positive or negative Vg-Vd.

Parameter determination

The parameters of eq. 1 can easily be extracted, if the Ig_sub of two transistors with different lengths are measured, with exactly the same oxide thickness. Because this is not known yet, the slopes of the Ig_sub vs. Ig_off curves is taken for two transistor lengths. Slope(0.16)/Slope(0.12) gives (0.16-2.ΔL)/(0.12-2.ΔL). Inline SEM measurements give the right poly length. From this equation, the ΔL can be calculated, without knowing the exact oxide thickness of every device. Now α and Ig^0 are easily determined too. The results are plotted in table 1.

LDD dose	ΔL [nm]	α	Ig^0 mean [pA/μm^2]	
			10/0.16	100/100
Low	1.6	3.8	305	301
Std	6.5	2.9	305	312
high	8.2	3.8	374	380

Table 1. Overview of the obtained parameters for NMOSTs

α is almost equal for all splits, as is expected. The ΔL increases for higher doses of LDD, but gives a very small value. The Ig^0 is plotted for a 100/100 transistor to estimate the error, which is about 1%. For PMOSTs the same calculation can be done.

The 100/100 Ig^0 current seems to increase with the LDD dose, however this is most probably due to a wafer-to-wafer average oxide thickness variation of about 0.3 Å.

4. Oxide thickness variations

Because of the known parameters in table 1 and the relation of Ig^0 with the oxide thickness of figure 4, the oxide thickness variation between the transistors (from figure 4) can be determined. The distribution for the three different splits in oxide thickness is plotted in figure 7.

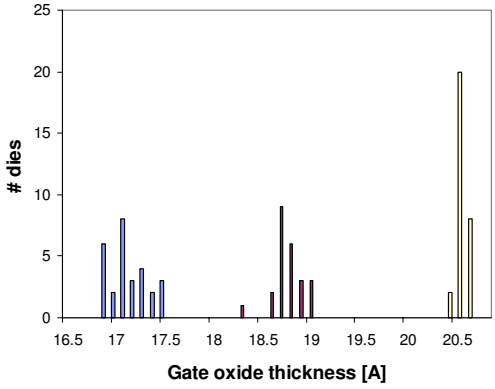

Figure 7 Histogram of the oxide thickness variations of three splits in oxide thickness, measured on 10/0.16 NMOSTs.

It is clear that if the gate oxide thickness decreases, its dispersion increases. With this method, it is possible to determine the thickness variation with an accuracy of about 0.2 Å. It is observed that for the 17Å split the dispersion is about 10%.

5. Conclusions

Scaling rules of the gate current are determined, and given in a simple but adequate model, which allows calculating the total circuit power consumption.

It has been shown, that this method can determine the gate oxide thickness variation with an accuracy of about 0.2Å, which makes it very effective for the oxide thickness optimization. The method is applied on short channel transistors, where large capacitance (which are normal used) will be less accurate, and are sensible to defects.

A large number of devices have to be measured, because of the large dispersion of the oxide thickness over the wafer.

Reference:

[1] W. Kirklen Henson et.al., IEEE Trans. el. Dev. VOL 47. NO. 7, july 2000

[2] G. Timp et.al., IEDM 98-1041

[3] Hisayo Sasaki Momose et.al., IEEE Trans. el. Dev. VOL 43. NO. 8, aug. 1996

[4] Nian Yang et.al., IEEE Trans. el. Dev. VOL 46. NO. 7, july 1999