Magneto-modulation of gate leakage current in 65 nm nMOS transistors: Experimental, modeling, and simulation results


1. Introduction

Modern integrated circuits contain millions of transistors that are continuously switching currents at different amplitudes and at different frequencies as well, which lead to the generation of on-chip magnetic fields in the range of tens of mili-Teslas (mT) [1]. Such a mixture of static and dynamic switching currents results also in an on-chip mixed static and dynamic magnetic field. The tri-dimensional nature of the B field lines, and the way the metal layer interconnections and circuits are distributed on the silicon chip, causes some of the B field lines to cross through the channel of the MOS transistors, which can result in a modification of the electrical performance. Therefore, it is the aim of this work to study the influence of magnetic B fields on the electrical performance of 65 nm nMOS transistors. For this purpose we perform a series of experiments where a controlled static and dynamic magnetic field is externally applied to the transistor. A description of the MOS transistor technology used for the experimental part as well as the experimental setup, together with the obtained experimental results are given in Section 2. In Section 3 we introduce and describe the proposed model that explains the experimental results for both static and dynamic B fields. The numerical simulations that validate the static B field model, which are based on Minimos-NT [2], are also described in this section. The dynamic B field simulations, which are based on a static–dynamic mixed macro-model, are also described in detail at the end of this section. The implications and conclusions of this work are reviewed in Section 4, where we infer a potentially larger impact of the on-chip magnetic field as the MOS transistor cell approaches a low-dimensional semiconductor system, which is the case for MOS technologies with minimum dimensions below 65 nm and gate oxides thinner than 2 nm.

2. Experimental results

An nMOS transistor fabricated in 65 nm CMOS technology with a Nitrogen-doped Silicon Oxide (Si₃N₄) gate oxide thickness Tₒx of 1.9 nm, a (W/L) ratio of (2 μm/65 nm), and a S/D-B junction of 15 nm was used in the experiments. The nMOS transistor was inserted in between the poles of a GMW 5403AC electromagnet that produced the static and dynamic B field, while the electrical test was performed with an Agilent B1500A Semiconductor Device Analyzer. The magnetometer has a large inductance value that impeded the generation of fast time-varying B fields. Therefore...
the experiments were only conducted at a maximum frequency of 50 Hz.

We define the x-axis along the channel length L, the y-axis along the transistor width W, and the z-axis normal to the x−y plane. A positive normal magnetic field \( B_z \) comes from above the gate into the bulk, while a positive parallel magnetic field \( B_y \) is defined as entering from \( y = 0 \) into the width \( W \) of the transistor. As a first approach an squared magnetic \( B_y \) pulse of an amplitude varying from \(-100 \text{ mT}\) to \(+100 \text{ mT}\), switched at a frequency of 5 Hz and with a raising/falling feature (\( dB_y/dt \)) of 10 mT/ms, was applied to the transistor biased at \( V_{ds} = 0.1 \text{ V} \) and \( V_{gs} = 1.0 \text{ V} \). From the experimental results shown in Fig. 1 we readily see the gate current \( I_g \) shows a positive peak \( \Delta I_g = 4.0 \text{ nA} \) for the negative transient of \( B_y \) and a negative peak \( \Delta I_g = 3.4 \text{ nA} \) for the positive transient. At the steady or static state there is also a modulation of \( I_g \). The first experimental finding is that the transient (dynamic) \( B_y \) field leads to a larger \( \Delta I_g \) current modulation than the \( B_y \) static case. Therefore we proceed with an additional experiment by applying a \( B_y \) field with three different (\( dB_y/dt \)) rates of 5.0, 6.6, and 10 mT/ms. The dynamic magneto-modulation of the gate current \( \Delta I_g \) increases with the increase of the (\( dB_y/dt \)) rate as shown in Fig. 2. At \( B_y \) field intensities close to \( \pm 100 \text{ mT} \) the \( I_g \) current show instabilities that we believe are because of “lattice disorders” at the Si–SiO\(_2\) interface [3] that lead to longitudinal optical (LO) magneto-phonon oscillations [4]. A first interpretation of the experimental results is that the surface electron concentration is being spatially and time modulated. Therefore, we should expect a channel current \( I_{ds} \) magnetic modulation as well, which is confirmed by the experimental results shown in Fig. 3. At the transient of the \( B_y \) field the \( I_{ds} \) current shows also a peak excursion, while at the steady-state there is a positive increase \( I_{ds} = 227.72 \mu\text{A} \) for a \( +100 \text{ mT} \) and a negative decrease \( I_{ds} = 227.6 \mu\text{A} \) for \(-100 \text{ mT} \). In order to understand the experimental results we proceed to the analysis of the data, and the development of an analytical model supported by electro-magnetic numerical simulations.

3. Analysis, modeling and simulations

Under the experimental conditions used in this work, the dynamic modulation of \( \Delta I_g \) follows a behavior modeled by Eq. (1).

\[
\Delta I_g = \Delta I_{g0} + \frac{1}{\zeta + \kappa} \left( \frac{\partial B_y}{\partial t} \right) B_y 
\]

(1)

where \( \Delta I_{g0} \) has a value that, for this particular conditions, ranges between 0.18 and 0.27 nA, but it depends on the bias conditions as governed by tunneling [5]. \( \zeta = -0.00156 \text{nA/mT} \), and \( \kappa = 0.00465 \text{nA s/T}^2 \). The \( I_g \) current is also linearly dependent on the static \( B_y \) field, with a modulation rate of 60 pA/mT as shown in Fig. 2.

Depending on the direction, the steady-state magnetic field, applied perpendicular to the surface and to channel current, pushes electrons either to the right or the left side along the width axis, which induces a two-dimensional space modulation of the surface electron concentration on the transistor channel. When a time-varying \( B \) field is also applied perpendicular to the surface and to the channel current, an electromotive force on the surface channel is induced, which either reduces or increases the channel current. The space static magneto-modulation of the channel electron concentration \( \Delta n \) is proportional to the magnitude of the time-varying \( B \) field, while the induced electromotive voltage \( V_{em} \) is proportional to the time-varying rate (\( dB/dt \)). Therefore the induced \( V_{em} \) voltage changes the effective channel voltage \( V_{dc} \) as.

\[
V_{dc} = V_{ds} \pm (W \times L) \cdot \frac{\partial}{\partial t} B
\]

(2)
Both, the space and time-magneto-modulation of the electron channel concentration induce a two-dimensional and time space potential modulation, or in other words a sort of "sea-wave of electrons" causes an image force oxide barrier modulation that leads to the gate current modulation and channel current interference.

Numerical simulations of the static magneto-redistribution of the surface electron inversion channel along the width are shown in Fig. 4. In this case a $B_z$ field of +100 mT is applied to the transistor biased at $V_{gs} = 0.1$ V and $V_{ds} = 1.0$ V. This leads electrons to accumulate at the left side and depopulate the right side. The electrons are space-modulated with a Gaussian distribution, with its peak $\Delta n$s located at 35 nm from the width side, and with an average broadness $\Delta V_{acc}$ of 70 nm. Ideally one should expect the inversion layer charge to end at the gate edge, but because gate edge effects, such as fringing fields and thickening of the gate oxide, the maximum of the magneto-modulated surface channel occurs under the gate at 35 nm from the gate edge. The peak value $\Delta n$s of this electron modulation is a linear function of $B_z$ that increases at a rate of $1.8 \times 10^{10}$ cm$^{-2}$/mT. This static magneto-modulation of the surface inversion charge also leads to a modulation of the oxide potential barrier $\Phi_0$ at a rate of 11 $\mu$eV/mT.

When $B_z$ becomes dynamic, with a switching rate ($d\Phi_0/dt$), the $\Delta n$ charge not only changes in space but in time as well. Such a space- and time-magneto-modulation induces a gate current $\Delta I_G = (\Delta Q n)/(d_t/dt)$. The induced $\Delta I_G$ current increases ($\Delta I_{G}^+$) or decreases ($\Delta I_{G}^-$) with respect to its value at $B_z = 0$ because of the induced electromotive voltage $V_{em}$ and the capacitive coupling with the gate electrode.

For this case the induced drain current variation $\Delta I_D$ is not only dependent on ($d\Phi_0/dt$) but on the magnitude of $B_z$ as well. In the linear regime the effective drain current $I_{D\text{eff}}$ is roughly approximated by the following equation:

$$I_{D\text{eff}} = \left[1 + \left(\frac{g}{m}\cdot\tau\cdot B_z\right)^2\right]^{-1} \cdot \left(\frac{W}{L}\right) \mu C_0 (V_{gs} - V_T) V_{ds}$$

where $\tau$ is the scattering time. Notice from Fig. 3 that for the steady-state value of $+B_z = 100$ mT $I_D$ increases to $I_0^+$, while for $-B_z = 100$ mT $I_D$ reduces to $I_0^-$. These variations are in the range of hundreds of nano-amperes and have a parabolic behavior with respect to $B_z$ as shown in Fig. 5. This parabolic behavior is the result of a classical Hall magneto-resistance [6]. The curvature of the parabola (see Eq. (2)) depends on the scattering time $\tau$, which is a complex function of both $V_{gs}$ and $V_{ds}$. For the particular case of $V_{gs} = 1.0$ V and $V_{ds} = 0.1$ V $\tau$ is equal to 1.029 ps, which is in the range (0.8–1.2 ps) reported for a two-dimensional electron gas on (1 1 1) silicon [7].

When the magnetic field is also applied perpendicular to the channel current flow, but parallel to the surface ($B_y$), the inversion channel charge experiences a vertical deflection (along the $z$-axis). For positive $B_y$ electrons are pushed towards the Si–SiO$_2$ interface, while for $-B_y$ electrons are pulled down the substrate.

The experimental results of the modulated drain current $\Delta I_D (I_d$ under applied magnetic field minus $I_d$ with no applied magnetic field) are also shown in Fig. 5. From measurements and simulations at different $V_{gs}$ and $V_{ds}$ conditions we found out that in general the scattering time $\tau$ gets larger than 1 ps as $V_{ds}$ gets closer to 0, which is an indication of a reduction of the scattering process at very low $V_{ds}$. For large values of $V_{gs}$, $\tau$ reduces down to 0.2 ps at $V_{ds} = 0.1$ V. For the magnetic field $B_y$ range from 0 to 50 mT $I_d$ increases and gets a maximum value at 25 mT. This increase is due to the increase in electron concentration at and near the surface. The centroid [8] of the quantized inversion layer, which in this simulations is placed 0.2 nm below the surface, have a variation of about 15% at $B_y = 25$ mT, which indicates the differential vertical increase of the surface electron concentration $n_z$ is responsible for the increase of $I_d$. The larger vertical modulation of the surface electron concentration is responsible for the larger $\Delta I_d$ vertical modulation when compared to horizontal $\Delta I_d$ modulation. For $B_y > 25$ mT the surface scattering process increases, which results in a reduction of the carrier mobility $\mu$ that compensates the increase, and thus results in a reduction of $I_d$ at larger $B_y$. The induced gate current modulation $\Delta I_G$ has the same behavior as the case of a perpendicular magnetic field $B_z$, but its magnitude is about three times larger. According to Minimos-NT simulations this increase of $\Delta I_G$ is due to the larger vertical surface modulation of the electron concentration (see Fig. 6). The modified version of Minimos-NT [9] allows the steady-state calculation of space-varying electron concentration in the inversion channel of the MOS transistor. For the applied $B_y$ field, the electron inversion channel is vertically modulated as shown in Fig. 6. A positive $B_y$ pushes electrons towards the surface, while a negative $-B_y$ pull electrons down the bulk.

The vertical modulation depends on the position along the channel as seen in Fig. 7. From the source-bulk metallurgical.
junction ($x = \Delta L$) up to approximately $x = 85$ nm. $\Delta n$ stays in the range of around $1 \times 10^{14}$ cm$^{-2}$. However, in and near the high longitudinal electric field region ($x = 85–100$ nm) the $\Delta n$ concentration goes up to levels of $1 \times 10^{16}$ cm$^{-2}$. The larger $\Delta n$ modulation at the drain side implies a higher gate current injection at the drain side. This space-modulated electron inversion charge $q \Delta n$ is then used to develop a non-quasi-static electro-magnetic (nqsem) macro-model. This "nqsem" model is implemented in Spice by considering the inversion channel layer as a current source $i_{em}$ given by the following model:

$$i_{em} = q \frac{\partial \Delta n}{\partial t}$$  \hspace{1cm} (4)

For sufficiently long and wide MOS transistors, and for $V_{gs}$ values above the threshold voltage $V_{th}$ and very small values of $V_{ds}$, one should expect the transversal and longitudinal electric fields $E_t$ and $E_l$ to be homogeneously distributed on the Si–SiO$_2$ interface, and thus the inversion channel charge to be also homogeneously distributed on the channel surface. Under this condition, the magneto-modulation of the inversion charge $\Delta n$ should produce homogeneous stripes of depopulated or accumulated electrons either at the left or right side of the channel. However, for short devices, these stripes do not have a homogeneous distribution along the channel as can be seen from Fig. 8, where the simulations of a (W/L) = (1 μm/65 nm) nMOS transistor are shown. Electrons are magneto-deflected to the right creating an accumulated stripe of electrons along the right side of the transistor, while at the left side there is a stripe of depopulated electrons. The largest magneto-modulation occurs close to the drain ($x = 60$ nm) and source sides ($x = 10$ nm), while the lowest one occurs at the middle of the channel ($x = 30$ and 40 nm). From the numerical simulation results the $\Delta n$ charge can be modeled as

$$q \Delta n = q(a + b I_{ds}) e^{E_l/E_c}$$  \hspace{1cm} (5)

where $a$, $b$, and $E_c$ are factors that depend on the geometry and bias conditions. But, in general we found out that the shorter the transistor the smaller the value of $E_c$, the critical longitudinal electric field at which the electric field dominates over the injection of electrons driven by the channel current $I_{ds}$.

As the $\Delta n$ charge is $x$-dependent, then one should take the integral from the source to the drain to calculate the total magneto-modulated charge $q \Delta n$.

$$i_{em} = \int_{x=0}^{x=L} \frac{q}{\partial t} \left( (a + b I_{ds}) e^{-E_l/E_c} \right) dx$$  \hspace{1cm} (6)
The inversion channel, where the magneto-current source $I_{leak}$ is generated by the influence of a time-varying magnetic field, is surrounded by a network of gate $C_{gem}$ and bulk $C_{bem}$ capacitors. The value of the $C_{gem}$ is determined by the gate oxide thickness and the transistor area ($W \times L$), while the value of the bulk capacitor $C_{bem}$ is given by the inversion channel layer-bulk depletion region. Both capacitors are bias-dependent. Therefore, their values used for the Spice macro-model need to be updated for the different bias conditions. In our particular case we implemented a network of 10 distributed $I_{leak}$ sources along the channel, 10 $C_{gem}$ capacitors, 10 $C_{bem}$ capacitors for the bottom plate, and five $C_{bem}$ capacitors for each lateral side of the transistor. The result of this distributed "nqsem" macro-model is shown compared to experimental results in Fig. 9. The measured pulsed bulk current probes the magnetic induction of bulk currents through the $C_{bem}$ capacitors, and validates the Spice macro-model.

4. Conclusions

For the first time a variation or magneto-modulation of the gate leakage current (in a 65 nm nMOS transistor) induced by a pulsed magnetic field is reported. This experimental observation is explained in terms of the magnetically induced space and time modulation of the surface electron inversion charge. The space and time magnetic modulation of the surface inversion electron charge also results in an electro-magnetic interference with the channel current. The interference of the steady-state magnetic field is attributed to the magnetic modulation of the channel conductance, while the transient interference is attributed to the induced electromotive voltage along the channel axis. The theoretical explanation is validated with composed Minimos-NT and Spice simulations. The minimum magnetic field that the transistor was able to detect was about 500 $\mu$T, a magnitude that is in the range of on-chip generated magnetic fields [10,11]. This is an indication that on-chip magnetic fields are self-inducing both channel current noise and gate leakage current. This preliminary report is the foundation for a future development of a full electro-magnetic model for nano-meter MOS technologies. It is also remarkable the high-speed reaction of both ($dl/dB/dt$) and ($dl/dB/dt$), which indicates the 65 nm nMOS transistor is capable to sense very high-speed on-chip magnetic fluctuations.

Acknowledgments

E. Gutiérrez thanks Intel and CONACyT (through Grant 100028) for their partial fundings, and IBM for providing the test samples.

References