Abstract

Purpose – The purpose of this paper was to present the results of the three types of FG transistors that were investigated. The reliability issues of oxide thickness due to programming, fabrication defects and process variation may cause leakage currents and thus charge retention failure in the floating gate (FG).

Approach – The tunnelling and electron injection methods were applied to program FG devices of different length (180 and 350 nm) and coupling capacitor sizes. The drain current and threshold voltage changes were determined for both gate and drain voltage sweep. The devices were fabricated using IBM 130 nm process technology.

Findings – Current leakages are increasing with device scaling and reduce the charge retention time. During programming, charge traps may occur in the oxide and prevent further programming. Thus, the dominant factors are the reliability of oxide thickness to avoid charge traps and prevent current/charge leakages in the FG devices. The capacitive coupling (between the tunnelling and electron injection capacitors) may contribute to other reliability issues if not properly considered.

Originality/value – Although the results have raised further research questions, as revealed by certain reliability issues, the results have shown that the use of FGs with nanoscale technology is promising and may be suitable for memory and switching applications.

Keywords Electron traps, nanoscale, threshold voltage shift, CMOS switches, floating gates

Paper type Research paper

1. Introduction

The circuit density and interconnections are increased in nanoscale applications (Kurokawa et al., 2009) and thus demand reliable packaging, thermal compensation and controllable fabrication process parameters (Zhang et al., 2008; Sylvester et al., 2008). The controllability of the floating gate (FG) transistor threshold voltage shift could provide solutions to overcome some of these challenges and can be further explored in tuning circuits, switching circuits and also in millimetre-wave application, among other applications (Mabuza et al., 2009). Different FG structures have been proposed in literature (Raguet et al., 2008; Bartolomeo et al., 2009; Lin and Sun, 2007; Bouchakour et al., 2001).

The power supply is limited by the thermal voltage (Foty, 2008). Thus, the supply saturates at about 1 V and its scaling factor is less than the actual technology downsizing, which then violates constant-field scaling and causes velocity saturation (Sylvester et al., 2008). Thus, the field in the oxide increases with technology downsizing and gives rise to transistor short channel effects, device parameter variations and excessive sub-threshold, junction, oxide charge trapping and gate oxide leakage currents (Chen, 2007). The intrinsic parameters are also not scaling as desired.
The reliability of FGs depends on the oxide thickness (the primary limiting factor), which may lead to charge leakages (Chung et al., 2001; Ning et al., 2012) and thus threshold voltage instability (Rao and Irrera, 2010). High dielectric material is proposed to overcome some of the reliability issues (Rao and Irrera, 2010; Sebastiani et al., 2007). Charge trapping and de-trapping are possible within the oxide.

The parasitic capacitors and resistors are not scaling as desired with the technology. Thus, the device dimensions will scale more than the parasitics and the overall oxide thickness will also scale down more. The scaling of voltage is also not as desired, owing to thermal voltage, thus the oxide voltage will scale less than the oxide thickness (Foty, 2008; Ning et al., 2012). Based on this, the electric field keeps increasing with technology and a constant field is not maintained, which causes reliability issues due to oxide field overstress.

The input capacitance affects the programming time of the devices and has less effect on the achievable output current. The capacitance effect on the output current is significant at low bias voltages. A smaller input capacitance reduces the programming time for both injecting and tunnelling of the FG devices. The technology downsizing results in smaller devices and thus, smaller output currents. The leakage currents reduce the charge retention time of the FG devices.

The initial charge or voltage of the FG is unknown and dependent on the implementation structure. In addition, programming of the FG is dependent on tunnelling (Chiou et al., 2001) and channel hot electron injection (Chen and Teng, 1992), which is modelled with fit parameters that can be extracted experimentally for accurate modelling.

This paper investigated FG devices with different sizes and programming capacitors. Tunnelling and electron injection programming methods were applied to control the charge in the FG and thus vary the threshold voltage shift. Possible charge losses, charge trapping and charge retention were evaluated. The paper is organized as follows: Section 2 presents the material and methods used to enable measurements, Section 3 discusses the biasing and programming conditions for the FG transistors, Section 4 presents the measured results and Section 5 is the discussion and conclusions.

2. Material and methods

The process used for design implementation is an IBM 8HP process, a 130 nm SiGe BiCMOS technology node. The process was chosen over 0.18 µm and other less scaled devices because of the higher unity gain frequency, thus making this process technology suitable for mm-wave applications. All simulation models and parameters were included in the 8HP process design kit supplied by the foundry. The Cadence “Analogue Design Environment” was used for schematic simulations and layout design. The FG devices were programmed using electron injection and tunnelling methods. The final multi-project wafer (MPW) that was fabricated is shown in Figure 1.

Figure 1 Fabricated chip of the MPW has an area of 4 × 4 mm². The charge pump and FG highlighted devices were fabricated

The other circuits shown in Figure 1 were developed for different research projects in the MPW. There are three FG devices. The fabricated integrated circuit (IC) was placed on a printed circuit board (PCB) (as shown in Figure 2), which enabled integration with an external power supply and the measurement equipment.
Figure 2 PCB designed for experimental measurements. Physical dimensions: 7 cm × 7 cm with a height of 2 cm

The PCBs designed for all the FG devices have the same structure as in Figure 2. The FG terminal includes the control gate voltage, tunnelling voltage, the output signal (drain/source) terminal and the supply voltage. The diced wafer is placed on a gold-plated Kovar carrier plate by using conductive epoxy. The Kovar plate is used to provide a ground plane for the chip. Aluminum substrate, which is used for wirebond pad connections, is soldered onto the Kovar plate. The Kovar carrier plate is screwed onto a larger nickel-plated aluminum casing, which has feedthrough connections to enable coupling with measurement equipment. The bondwires used are 18 µm in diameter. The equipment used during measurements is shown below in Figure 3.

Figure 3 The parameter analyser used for DC sweep measurements and its test fixture

The test equipment presented in Figure 3 was connected to a computer to capture the measured results. Before and after programming the FG devices, the parameter analyser (Agilent 4155B Semiconductor Parameter Analyser) and its test fixture (Agilent 16442A Test Fixture) were used to measure the gate and drain voltage sweep of the FG devices. The FG devices summarised in Table I were fabricated and measured.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>FG 1</th>
<th>FG 2</th>
<th>FG 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$ (µm)</td>
<td>0.18</td>
<td>0.18</td>
<td>0.35</td>
</tr>
<tr>
<td>$W$ (µm)</td>
<td>8</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>$C_{TN}$ (µm²)</td>
<td>0.8 x 0.8</td>
<td>0.8 x 0.8</td>
<td>1 x 1</td>
</tr>
<tr>
<td>$C_{CG}$ (µm²)</td>
<td>8 x 8</td>
<td>8 x 8</td>
<td>10 x 10</td>
</tr>
</tbody>
</table>

Table I Summary of parameters for the fabricated FG devices
All the devices in Table I were fabricated using IBM process technology. One of the FG devices in Table I has a length of 0.35 µm and the other two devices have a length of 0.18 µm. The capacitors, $C_{CG}$ and $C_{TUN}$, are formed by shorted transistor that both have an oxide thickness of 3.2 nm. The ratio between $C_{CG}$ and the $C_{TUN}$ area is 10. The layout designs are shown in Figure 4.

**Figure 4** For (a) and (b), tunnelling capacitor $W = 0.8 \mu m$ and $L = 0.8 \mu m$; control capacitor $W = 8 \mu m$ and $L = 8 \mu m$; pFET transistor $L = 0.18 \mu m$ and $W = 8 \mu m$ (a) or 1.5 µm (b). For (c), tunnelling capacitor $W = 1 \mu m$ and $L = 1 \mu m$; control capacitor $W = 10 \mu m$ and $L = 10 \mu m$; pFET transistor $L = 0.35 \mu m$ and $W = 2 \mu m$. The simplified circuit diagram is presented in (d).

The fabrication process uses five levels of global layers; two levels of 1 copper (M1 and M2), one level of 2 copper (MQ) and two levels of analogue metal (LY and AM). The common wiring level vias are V1, V2 and VL.

3. Biasing results and programming conditions

The FG drain voltage sweep was done for different control gate voltages and the results were independent of the control gate voltage, since the control gate capacitor blocks the DC path. The maximum drain currents were 1.8 mA, 1.1 mA and 0.52 mA for FG 1, FG 2 and FG 3, respectively. The gate voltage sweep measurements are presented in Figure 5 for different FG drain-to-source voltages, as indicated by the legends. As the drain-to-source voltage increases, the drain current also increases. The threshold voltage influence is dominantly visible.
for low drain-to-source voltages. It can be noted that for voltages below the threshold (ignoring the device behaviour around the threshold) the logarithm of the drain current is nearly constant at its minimum level and starts increasing linearly above the threshold.

**Figure 5** The gate voltage sweep before programming (logarithmic scale of the drain current); the measured threshold voltage is (a) 2.3 V for FG 1, (b) 3.3 V for FG 2 and (c) 3.1 V for FG 3.

Since the gate voltage bias has no influence on the drain voltage sweep measurements, a bias of zero volts was used for the drain voltage sweep measurements after programming the devices. The results show that the devices are dependent on the drain-source bias condition. The difference in the threshold deep in Figure 5 (a-c) is due to the initial charge in the floating gate, which is unknown after fabrication. Another contributing factor is the device dimensions. The area based on the length and width is higher for FG 1, followed by FG 3 and then FG 2. Thus, a smaller area resulted in higher threshold voltage and a higher area resulted in lower threshold voltage. It can be noted that critical behaviour is experienced at zero bias voltage where there is a narrow deep in the drain current. Thus, the results presented in the next section are compared to this critical behaviour. The DC programming conditions are presented in Table II.

**TABLE II** The programming conditions for tunnelling and electron injection

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Tunnelling</th>
<th>Injection</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{CG}$ (V)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>$V_{TUN}$ (V)</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>$V_{S}$ (V)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>$V_{D}$ (V)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pulse width (s)</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

The tunnelling and injection voltage was 5 V, but this voltage was also increased to 5.5 V in certain conditions. The programming results are presented in the next section. The initial charge after fabrication is unknown and different for all devices, thus there is uncertainty about the results achieved by the measurements before programming the devices used as reference.
4. Results

For the results provided here, a comparison is made between the measurements taken with the device in its initial state from fabrication (INIT), after tunnelling the device from the initial state (TUN5₁), after electron injection following the first tunnelling (INJ55), 10 hours after injection for possible charge losses (10H) and then after tunnelling following injection (TUN55₂). The results of the gate voltage sweep measurements for FG 1, FG 2 and FG 3 are shown in Figure 6.

**Figure 6** Gate voltage sweep comparison for: (a) FG 1, (b) FG 2, and (c) FG 3

After the first tunnelling (TUN5₁), the threshold voltage shifted from about 2.3 V to about 0.7 V for FG 1, as shown in Figure 6 (a). There was no significant change in threshold voltage after injection (INJ55), but there was an increase in the drain current. The device was measured again 10 hours after injection (10H); for gate voltages below 3 V, similar results were achieved and there was a slight change above 3 V. Another tunnelling was done after injection (TUN55₂); although there was change in current, the variation was within 0.025 mA. This value was much smaller compared to the difference between the first tunnelling (TUN5₁) and the injection (INJ55), which was about 0.19 mA. For the second tunnelling after injection (TUN55₂), the device did not return to the level of the first tunnelling. Subsequent programming with either injection or tunnelling did not cause threshold voltage shift. The cause of this behaviour, as noted in the results, can be attributed to a number of factors. The fact that the threshold was 0.7 V after first tunnelling could be caused by voltage overstress, which may have cause over-erasure of the device. In cases of over-erasure, the device may have acquired properties preventing it from responding to further programming. Thus, since there was almost insignificant change in the threshold voltage after the first tunnelling of FG 1, voltage overstress contributed to the response of the device, as seen in the measurement in Figure 6 (a). The minor difference between the injection results (INJ55) and the measurements taken 10 hours after injection (10H) shows that there were leakages, although at this stage it is not fully clear which leakage components made a greater contribution, i.e. sub-threshold leakage, edge-direct tunnelling current or gate-induced drain-leakage.

The gate voltage sweep measurements for FG 2 are shown in Figure 6 (b). The initial device threshold (INIT) is higher than the other thresholds of either tunneled or injected states. After the first tunnelling (TUN5₁), the threshold voltage was reduced and the current increased. For injection (INJ55), the current was reduced and the threshold voltage was also reduced, but opposite results were expected in terms of the threshold voltage. The measurements done 10 hours after injection (10H) present other undesirable effects (as shown in Figure 6 (b)). The results (10H) were supposed to be similar to the injection results (INJ55), with a slight variation, as in the case of FG 1 (as shown in Figure 6 (a)). The results (10H) were similar to the first tunnelling case (TUN5₁),
which could then be related to possible current leakages. FG 2 has the smallest transistor area and is thus likely to be affected by process variation and current leakages, compared to the other devices presented here. Because of the smaller area, the amount of charge that can be contained in the FG is small and that is why small variations or leakages would have a higher influence. It is also interesting to note that tunnelling measurements after injection (\(TUN55\)) have reduced the threshold voltage, as expected when tunnelling a device. The inconsistencies with regard to the measurements, especially for the injection (\(INJ55\)) and second tunnelling, (\(TUN552\)) could be related to a few different issues. Firstly, as the initial charge in the floating gate is unknown, it is possible that the first tunnelling (\(TUN55\)) was not sufficient (the programming time could have been shorter) for the whole charge transfer at that stage (refer to Figure 6 (b)). Thus, while doing injection programming, the results seemed to contradict the expected threshold voltage shift direction. The actual cause could be attributed to the shorter programming time of the first tunnelling, which was insufficient to overcome the initial charge effects. This can be further supported by looking at the second tunnelling done after the injection. Based on that, it is clear that in this case the tunnelling had reduced the threshold voltage as expected compared to the prior injection programming. Secondly, the results taken 10 hours (\(10H\)) after injection (\(INJ55\)) were determined to test the retention of charge. It was noted that the device had failed the retention test and had actually returned to the state it was in before the injection programming, which was after the first tunnelling (\(TUN55\)). This device, FG 2, had the smallest device dimensions and thus it was expected that it would suffer more from the short-channel effects and thus, current leakages. Although this might have been the case, unlike FG 1, the FG 2 device showed some variations in the threshold voltage and it was not trapped into one state. Hence, in the case of the FG 2 device, there was no voltage overstress.

The measurement results for the gate voltage sweep of FG 3 are shown in Figure 6 (c). The initial state measurements (\(INIT\)) have a threshold voltage lower than the first tunnelling results (\(TUN5\)). This is rather different from FG 1 and FG 2 presented above. The first tunnelling increased the threshold voltage, while injection (\(INJ55\)) reduced it to within 0.1 V compared to the initial state threshold, as shown in Figure 6 (c). The tunnelling and injection yielded opposite results with regard to threshold voltage shift changes. Further tunnelling (\(TUN552\)) had almost no visible effects, as there was only the slightest change compared to the injection results. The results in Figure 6 (c) show that the threshold voltage was varied between 2.9 and 3.6 V, with the initial state threshold voltage of 3.1 V. FG 3 has a longer gate length and the total threshold voltage variation is 0.7 V (between minimum and maximum values); this value is lower compared to FG 1 and FG 2. The contributing factor is the fact that the FG 3 threshold voltage shifted to a higher value after the first tunnelling, which is opposite to the results achieved in FG 1 and FG 2. Since reference is based on the initial measurement, there is uncertainty because the initial charge is unknown for these devices. The device dimensions and also the coupling capacitor dimensions are different. Owing to different transistor area, the initial charge for each device is expected to be different and thus, different initial threshold voltages are expected. The measurements taken immediately after injection (\(INJ55\)) and the measurements taken 10 hours after injection (\(10H\)) have a direct correlation and no variation. Voltage overstress and charge trapping are the potential causes of the behaviour of the FG 3 device. Even in this case, the uncertainty of the initial charge has an important effect on the interpretation of results. For example, the FG 3 initial charge threshold is smaller than the first tunnelling threshold (\(TUN55\)), but since all subsequent results are around the same threshold voltage for both the second tunnelling and injection programming, this was probably caused by charge trapping. Hence the device did not respond to further programming. The minor deviation between the threshold voltages can be associated with charge leakages. At this stage, it is not clear which leakage component(s) is more dominant.
The drain voltage sweep measurements for FG 1 are shown in Figure 7 (a). The first tunnelling reduced the threshold voltage and thus resulted in an increase in the drain current \((TUN5)\). Although the gate voltage sweep measurements showed no change in the threshold voltage, as shown in Figure 6 (a), the drain voltage sweep reveals that following injection \((INJ55\) and \(10H\)), the drain current was reduced, which translates to an increase in the threshold voltage. This corresponds to what was expected: injection reduces the output current. Tunnelling after injection was supposed to reduce the threshold voltage and increase the current, but the second tunnelling did not have an effect. Subsequent results were recorded for injection and tunnelling with the programming voltage increased from 5 V to 5.5 V and then to 6 V, but the responses did not change and the device was trapped in one state with no further programming effects.

The drain voltage sweep measurements for FG 2 are shown in Figure 7 (b). The tunnelling measurements \((TUN5)\) show an increase in the drain current from 1.1 mA to about 2.7 mA from the initial fabrication state \((INIT)\). Injection resulted in a decrease in the drain current from 2.7 to 2 mA, reflecting an increase in the threshold voltage. There was a slight increase (about 0.1 mA) in the drain current with the second tunnelling after injection.

The drain voltage sweep measurements for FG 3 are shown in Figure 7 (c). The tunnelling measurements \((TUN5)\) show an increase in the drain current from 0.51 mA to about 1.48 mA from the initial fabrication state \((INIT)\). Injection \((INJ55\) and \(10H\)) resulted in a decrease of the drain current from 1.48 to 1 mA, reflecting an increase in the threshold voltage. There is a slight increase (about 0.1 mA) in the drain current with the second tunnelling after injection, almost similar to FG 2.

This section presented the practical measurement results for the FG devices. The injection and tunnelling characteristics were measured with three different devices, which differed in width or area of input capacitive coupling. The results showed that as the device area was reduced, the achievable output current was also reduced. Thus, when larger currents are required, a larger device area must be used. The effect is that more die area is taken up by these large devices, thus an area penalty is paid for high-power applications.

5. Discussion and conclusion

The maximum measured drain current for three devices was 3 mA, 2.7 mA and 1.4 mA for FG 1, 2 and 3, respectively. FG 1 has an area of 1.44 µm², FG 2 has an area of 0.27 µm² and FG 3 has an area of 0.7 µm². The
area size determines the amount of charge or current in the FG. A longer device length minimises process variation and prevents depletion region overlap between the drain and the source regions (Ning et al., 2012).

The ratio between the ON state switching current (after tunnelling) and OFF state switching current (after injection) was 6, 1.35 and 1.5 for FG 1, 2 and 3, respectively. The ratio is higher for devices which have a larger transistor area; this relates to the amount of charge that can be contained in the FG transistor: a larger area requires more programming time and higher programming voltages. The ratios correlated with the expected results in terms of FG transistor area: a high area results in a higher ratio. The ratio of less than 2 achieved for FG 2 and 3 is less suitable for switching or tuning applications: a smaller ratio between the ON and OFF states could result in higher losses and demand much smaller bit-line swing voltage for sense amplifiers. The measured drain currents should not be interpreted as the highest possible output current that can be achieved, but the highest values based on relevant measurements for similar programming conditions.

The reliability issues are dominated by the oxide thickness of the transistor, which relates to the programming methods. Uncertainties are introduced by oxide voltage overstress, charge trapping in the oxide and leakage currents through the substrate. Reliability issues were experienced for FG 1 and 2. For FG 1, tunnelling was performed and measurements taken, but when the device was measured after 24 hours, similar measurements were not obtained. After tunnelling the device with the same voltage, but for a longer period, the charge was maintained and similar results were achieved after 24 hours. For FG 2, similar behaviour was experienced after injection and comparison of the results with 10 hours’ difference between measurements: the effect was rather visible for the gate voltage sweep results.

It can also be noted from the gate voltage sweep experimental results that the threshold voltage shift showed no significant variation for subsequent programming; the significant change only happened when the devices were programmed from the initial state. Even though this was the case, the drain voltage sweep measurements showed the actual effects expected for drain current changes after programming the devices.

The conclusion based on this is that even though for the p-type transistor there is no change in the threshold voltage, the drain current does change: this effect of no threshold voltage shift has also been identified in previous work (Takeda et al., 2002). The behaviour of having no significant change in the threshold voltage, but a noticeable change in the drain current, is interesting to observe and poses a challenge for further investigative work in future. The drain current changes correlate with the expected behaviour of the devices: tunnelling increases current and injection reduces current. The experimental results further raise uncertainties caused by the oxide thickness and the unknown initial charge in the FG after fabrication; the initial charge was different for all devices.

References


