3D monolithic integration: Technological challenges and electrical results


⇑ Corresponding author.
E-mail address: maud.vinet@cea.fr (M. Vinet).

ARTICLE INFO

Article history:
Available online 19 November 2010

Keywords:
Transistor
CMOS
3D
Germanium

ABSTRACT

After a short reminder of the principle of monolithic 3D integration, this paper firstly reviews the main technological challenges associated to this integration and proposes solutions to assess them. Wafer bonding is used to have perfect crystalline quality of the top layer at the wafer scale. Thermally stabilized silicide is developed to use standard salicidation scheme in the bottom layer. Finally a fully depleted SOI low temperature process is demonstrated for top layer processing (overall temperature kept below 650°C). In a second part the electrical results obtained within this integration scheme are summarized: mixed Ge over Si inverter is demonstrated and electrostatic coupling between top and bottom layer is used to shift the threshold voltage of the top layer. Finally circuit opportunities such as stabilized SRAM or gain in density are investigated.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

3D integration is regularly mentioned for its potential in decreasing costs, variability [1] and delay in interconnections limiting nowadays IC’s performance [2]. 3D monolithic integration is the only option enabling a full use of the third dimension at the cell scale thanks to its high alignment precision (σMONO ~10 nm [3] compared to σTSV ~0.5 μm [4]). In addition at the CMOS cell level 3D monolithic integration offers the unique additional benefit to allow for an independent optimization of n-FET and p-FET allowed by stacking entire p-FET onto n-FET layers suppressing thus lots of technological challenges. Especially it arouses interest first to alleviate gate stack challenges in term of EOT and work function adjustment and then to control strain in the devices channel. Indeed firstly to address the control of the threshold voltage, many techniques have been proposed [5] all relying on additional lithography and etching steps. 3D integration can easily overcome all these integration challenges in separating n and p layers. Secondly, in regular planar integration, different ways to induce strain are used in order to improve electron and hole mobilities. One is to use (1 0 0) and (1 1 0) oriented substrates for n and pFETs, respectively, which leads to tricky and expensive processes [6] another is to induce strain in the channel through the design of stressor liner but as n and pFETs requires opposite strain two different materials are required for each transistor type and the benefit of such liner decreases as the pitch is scaled down. Finally, transport can also be optimized thanks to the use of high mobility channel material such as germanium for pMOSFETs [7]. Since the introduction of high k dielectric in production germanium has indeed proven to be a promising material [8–10] as it allows to benefit from high hole mobility. Within this context 3D monolithic integration appears as an opportunity both for Si over Si integration and for hybrid Ge based CMOS integration, both transport and gate stack can be optimized within this integration scheme.

After a short reminder of the principle of monolithic 3D integration, in this paper we firstly review the main technological challenges associated to this integration and we propose solutions to assess them. In a second part we will summarize the electrical results obtained within this integration scheme.

2. Monolithic 3D process flow

The basic integration principle is described below for Ge CMOS integration [11] and as well for Si only integration. Firstly, optimized nMOS transistors are fabricated on the lower level, where any substrate (bulk, SOI – shown here – or sSOI (1 0 0)) can be used. For salicidation, a platinum incorporation followed by tungsten and fluorine implantation in nickel is introduced to stabilize the silicide (Fig. 1a). Low temperature molecular bonding is then used for upper active layer realization (Fig. 1b). pMOS are the processed on the upper layer (Ge or (1 1 0) SOI or standard SOI) using HfO2 gate dielectric and SPE (Solid Phase Epitaxial) dopant activation (Fig. 1c) in order to limit the overall thermal budget to...
Finally a single lithography is used to realize the 3D contacts (Fig. 1d). Two different gate stacks can be integrated in the two stacked layers without any additional lithography.

3. Technological challenges for 3D integration and their assessments

3.1. High quality upper layer

Several studies have already shown successful monolithic 3D integrations [12–15]. These approaches however fail to address sub-45 nm node due to imperfect top substrate quality (thickness control, and defects). The use of Wafer Bonding (WB) provides crucial advantages compared to previously employed techniques for top active layer realization, namely: 1/top substrate perfect quality, 2/no need for seed-windows which limits density, 3/independent optimization of n and p-MOSFETs when processed on distinct layers via the choice of strain options, channel and surface orientations and channel material independently for each layer.

To achieve high quality molecular bonding several challenges are raised. Firstly, the base wafer topography due to the bottom transistor layer (~100 nm) must be suppressed to enable full film transfer. Secondly, note that the Inter-Layer Dielectric (ILD) thickness has to be as minimized down to around 100 nm to allow dense 3D contacts. Indeed this additional depth, specific to 3D technology must be minimized to enable the contact scalability as its etching and filling become critical. Low temperature bonding of SOI substrate followed by etch back of handle wafer is used to transfer the stacked thin film. A low temperature annealing (200 °C) was performed to strengthen the bonding interface before mechanical exfoliation. No bonding defect has been observed at the wafer scale either in acoustic or infrared characterization as shown in Fig. 2 Middle and Right. Fig. 2 Left shows a SEM cross-section at the transistor scale. The Inter-Layer Dielectric is around 100 nm between the top of the bottom gate and the bottom of the top silicon channel. The bonding interface is located roughly in the middle of the ILD. Its thickness can be further decreased (as shown later on) by using a thin thermal oxide on the handle wafer instead a deposited one.

Note that wafer bonding, contrary to other techniques for upper thin film realization based on recrystallization, offers the possibility to naturally co-integrate different surface and channel orientations such as $h_{100}$ for bottom active layer and $h_{110}$ for top layer without any additional process challenges. Furthermore this mature process step leads to a high quality crystalline top film with low thermal budget.

3.2. Upper layer low temperature process

Low temperature processing of upper layer is necessary to protect the underlying layer performance: no additional dopant diffusion, no salicide degradation, limited work function shift... The main costly step in term of thermal budget is dopant activation. Ge completely fulfils the low temperature dopant activation requirements. Thus its integration as an upper layer is quite straightforward. On the other hand for Silicon integration a promising way to achieve excellent activation levels is the Solid Phase Epitaxy (SPE) technique. SPE is based on the low temperature recrystallization of amorphous Si that results in an above equilibrium activation. This technique can be used at low temperatures (typically below 600 °C) that suppress dopant diffusion and facilitate ultra-shallow junction formation. Fig. 3 shows the sheet

![Fig. 1. 3D monolithic integration process flow.](image-url)
resistance $R_s$ junction depth $X_j$ trade-off that can be obtained in bulk silicon for P-dopants, for conventional (Rapid Thermal Processing, RTP) and advanced (SPE, Laser, Flash) annealing techniques [16,17]. Sheet resistance measurements obtained on thin Si films (<30 nm) with SPE-activated boron are also plotted for comparison. Ge pre-amorphization was performed prior to boron implantation. For these implant conditions, excellent results are obtained that are in line with the most advanced annealing techniques on bulk and clearly outperform conventional RTP results.

Additionally to dopant activation, other steps such as gate dielectric, spacers, and passivation layers realization are classically processed at high temperature. It is necessary to keep these thermal budgets as low as possible. For gate dielectric realization, thermally grown SiO$_2$ (at ∼1000 °C) is logically replaced by high $K$ dielectric. In our case HfO$_2$ deposition using Atomic Layer Deposition (ALD) at 350 °C followed by a thermal anneal at 515 °C for 5 min was used. Low temperature deposited oxides are used for spacers and passivation layers process. This allows us to be in line with the target: the thermal budget is kept below 600 °C.

### 3.3. Silicides

Silicides are greatly sensitive to thermal budget. For 600 °C thermal budget, the classical NiSi agglomerates, leading to a strongly increased sheet resistance. To stabilize NiSi, an original treatment based on platinum incorporation associated with fluorine and tungsten implantation is proposed. Fig. 4 clearly shows the benefits of this NiSi treatment which ensure its stability up to (650 °C, 40 min) whereas NiSi is agglomerated in less than 1 min at this temperature.

The electrical results were confirmed by Scanning Electron Microscopy (SEM) observations of the silicide layer. After 40 min anneal at 650 °C, the stabilized silicide layer still presented a
continuous aspect unlike the classical NiSi, which is strongly agglomerated.

Development of a stabilized silicide up to 650 °C is an essential breakthrough, as it determines the ability of realizing top and bottom MOSFET with performance compatible with ITRS requirements for sub-45 nm node:

- For bottom MOSFET, the access salidication is mandatory for reaching ITRS values in terms of series resistance.
- For top MOSFET, a 650 °C thermal budget allows epitaxy for raised source-drain which is compulsory for advanced nodes fully depleted SOI structures.

Then, this stabilized silicide enables to obtain equivalent performance for top and bottom MOSFETs for the sub-32 nm nodes.

3.4. 3D contacts

Contacting top and bottom layer is a problem specific to 3D monolithic integration scheme. To save one lithography step, a single lithography can be used for realization of contact reaching both top and bottom MOSFETs electrodes. In that configuration, a highly selective etch is needed to open contact stretching down to bottom layer without passing trough the upper layer. It was demonstrated that the bottom layer can be reached without passing trough the silicide of the top active layer, thus giving a low contact resistivity [3].

3.5. Alignment

Density of 3D structure is mainly linked to alignment performances. The huge advantage of monolithic integration compared to other 3D integration schemes mainly relies on the fact that the alignment performance between layers only depends on lithographic alignment capability.

The overlays of bottom active level with bottom gate level and top active layer are displayed in Table 1. We clearly observe that the alignment performance is not degraded for the upper level location. This enables connections at the transistor scale and represents the major advantage of 3D monolithic over parallel integration.

4. Electrical results

We have firstly demonstrated that low temperature process yields thinner EOT than its Rapid Thermal Annealing counterpart. Capacitance measurements were performed on N and PFETs of top and bottom layers. With a 5 nm HfO2 gate dielectric for both types of devices we observed a 4 Å EOT reductions: 1.5 nm (resp. 1.6 nm) for top NFET (resp. PFET) compared to 1.9 nm (resp. 2 nm) for bottom NFET (resp. PFET). The EOT decrease is due to the reduction of interfacial oxide growth as shown by the cross sectional TEM images that were realized through both gate stacks. Gate current was plotted as a function of EOT and this figure of merit clearly underlines the advantage of low temperature process: 4 Å EOT decreases it exhibits lower Jc than 1050 °C process [20].

We have also demonstrated functional Ge/Si and Si/Si devices. GeOI top p-MOSFET presents at the same time high hole mobility (>1.5 compared to HfO2/TiN gate stack on SOI [18]) with EOT = 1.9 nm controlled Ion/Ioff ratio, well-suited Vth (Fig. 5). 3D SRAMs [19,20] and inverters have been fabricated (Fig. 5) using this technology. Equilibrated inverter transfer voltage characteristic has been obtained with a Ge p-MOSFET width equal to the Si n-MOSFET one, demonstrating the interest of Ge over Si co-integration for high density CMOS integration.

We performed a complete optimized 3D all Si integration where (1 1 0) wafer was used for top PFETs. Ion(Vg) output characteristics show well behaved devices for top and bottom layers [20]. Thus a functional equilibrated inverter was demonstrated with (1 1 0) PFET on top of (1 0 0) NFET.

The unique 3D electrostatic coupling between layers was quantified: depending on the targeted Inter-Layer Dielectric (ILD) the top transistors can be considered either as FDSOI with thick BOX (thin ILD, Fig. 6 left) or as Double Gate like devices (ultra-thin ILD, Fig. 6 right). In that latter case a 130 mV VTH shift is observed when the top NFET lies above a bottom transistor as compared to the isolated case (q_m of the bottom gate is the one of N+ doped polysilicon i.e. 5.1 eV). This VTH is due to the electrostatic coupling.

Table 1
Alignment performance with stepper 248 nm for bottom gate on bottom active and top active on bottom active. σ is the standard deviation of the overlay measurement.

<table>
<thead>
<tr>
<th></th>
<th>Bottom gate/bottom active (nm)</th>
<th>Top active/bottom active (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Y</td>
<td>10</td>
<td>7</td>
</tr>
</tbody>
</table>

Fig. 5. Ion-Vg characteristics of stacked p-MOSFETs (either on GeOI and SOI) on salicided bottom n-SOI MOSFET. Transfer Voltage characteristic of an inverter: top SOI (1 1 0) pFET and bottom SOI (1 0 0) nFET.
between the stacked layers. It demonstrates that the threshold voltage of top FETs can be dynamically tuned by biasing the bottom FET. This characteristic presents strong interest for the stabilization of SRAM cells [21,22].

5. Circuit advantages

Regarding memories, embedded SRAMs play a key role in VLSI circuits. Today, they represent more than 60% of the total chip area and between 20% and 50% of chip power [23]. As the technologies scale down into the deca-nanometer range, transistor leakage and intrinsic variability are becoming a major concern. Designing a reliable, compact and low power SRAM cell, using conventional planar transistors on BULK Si wafers, is becoming increasingly difficult [24]. To cope with this issue, thin film SOI technologies (FDSOI, DGMOS, Gate-All-Around) and innovative design architectures are extensively investigated. Thin film SOI devices allow an improvement of both sub-threshold slope and short channel effects thanks to a better electrostatic channel control, reducing performance gain, required for logic and memory aggressive scaling.

6. Conclusion

Major technological challenges of this integration have been reviewed and successfully settled. Functional inverters and SRAM featuring either Si over Si or Ge over Si have been fabricated. And it has been shown to be promising in terms of density (average density gain ~40% compared to planar 45 nm integration) and performance gain, required for logic and memory aggressive scaling.

Acknowledgement

This work was supported by the French National Research Agency (ANR) through Carnot Institute funding.

References