State-of-the-art and future of silicon on insulator technologies, materials, and devices

Sorin Cristoloveanu *

Laboratoire de Physique des Composants à Semiconducteurs (UA–CNRS & INPG), ENSERG, B.P. 257, 38016 Grenoble Cedex 1, France

Abstract

The context of SOI technologies is briefly presented in terms of wafer fabrication, configuration/performance of typical SOI devices, and operation mechanisms in partially and fully depleted MOSFETs. The future of SOI is tentatively explored, by discussing the further scalability of SOI transistors as well as the innovating architectures proposed for the ultimate generations of SOI transistors. © 2000 Elsevier Science Ltd. All rights reserved.

1. Introduction

Silicon on insulator (SOI) technology, originally developed for the niche of radiation-hard circuits, has experienced three decades of continuous improvement in material quality, device physics, and processing. Recently, SOI has joined the microelectronics roadmap: SOI circuits are indeed attractive because of their enhanced performance (higher speed, lower power-voltage) and scalability.

The aim of this article is to provide a synthetic view of the present status and future developments in SOI technologies and devices. The fabrication methods for SOI materials and the family of SOI devices will be reviewed in Sections 2 and 3, respectively. Section 4 will be dedicated to the main mechanisms involved in the operation of fully and partially depleted SOI MOSFETs. In Section 5, it will be demonstrated that, based on scalability and flexibility arguments, SOI is capable of further extending the limits and performance of bulk silicon technology.

Revolutionary SOI solutions include: ground-plane, double-gate, and ultra-thin transistors, dynamic-threshold MOSFETs, 3-D structures, and microsensors.

2. Synthesis of SOI wafers

In the last 20 years, a variety of SOI structures have been conceived with the aim of separating, using a buried oxide (BOX), the active device volume from the Si substrate.

Silicon-on-sapphire (SOS, Fig. 1a1) is fabricated by epitaxial growth of a Si film on Al2O3. The electrical properties may suffer from lateral stress, in-depth inhomogeneity of the film, and defective transition layer [1,2]. SOS films have been improved by solid-phase epitaxial regrowth (implantation-induced amorphisation and annealing). Good quality 100 nm thick films, on 6° SOS wafers [3], are now available. The 'infinite' thickness of the insulator renders SOS promising for the integration of RF (66 MHz, low noise [3]) and radiation-hard circuits.

The epitaxial lateral overgrowth (ELO) method consists of growing a single-crystal Si film on a seeded and patterned oxide (Fig. 1a2). The ELO process requires a post-epitaxy thinning of the Si film and is limited by the lateral extension of defect-free, single-crystal regions. The ELO technique serves for the integration of 3-D stacked circuits.

In the last decade, the dominant SOI technology was SIMOX which is synthesized by internal oxidation during the deep implantation of oxygen ions into a Si wafer. Annealing at a high temperature (1320°C, for 6 h) restores the crystalline quality of the film. 8° SIMOX wafers have good thickness uniformity, low defect density (except threading dislocations: 10⁵–10⁶ cm⁻²), sharp
Si–SiO₂ interface, robust BOX, and high carrier mobility [4]. SIMOX comes in several flavors (Fig. 1b): (i) thin and thick Si films fabricated by adjusting the implant energy, (ii) low-dose ($4 \times 10^{17} \text{O}^+ / \text{cm}^2$) SIMOX with a 0.1 μm thick BOX (Fig. 1b1), (iii) standard oxygen dose ($1.8 \times 10^{18} \text{O}^+ / \text{cm}^2$) SIMOX where the thicknesses of the Si film and BOX are 0.2 and 0.4 μm, respectively (Fig. 1b2), (iv) double SIMOX (Fig. 1b3), where the Si layer sandwiched between the two oxides serves for interconnects, wave guiding, additional gates, or electric shielding, (v) interrupted oxides (Fig. 1b4) which can be viewed as SOI regions integrated into a bulk Si wafer.

Wafer bonding (WB) and etch-back is another mature SOI technology. An oxidized Si wafer is mated to a second Si wafer (Fig. 1a3). After bonding, the structure is thinned down to reach the target thickness of the silicon film. Etch-stop layers can be achieved by doping steps (P+/P −, P/N) or porous silicon (Eltran) [5].

The recent, revolutionary UNIBOND process uses the deep implantation of hydrogen (dotted line in Fig. 1a3) to generate microcavities [6]. After bonding and annealing, the wafers separate naturally at a depth defined by the location of hydrogen microcavities. This mechanism, referred to as Smart-Cut, is completed by touch-polishing. The Smart-Cut approach has several outstanding advantages: (i) the etch-back step is avoided, (ii) the second wafer (Fig. 1a4) being recyclable, UNIBOND is a single-wafer process, (iii) only conventional equipment is needed for mass production, (iv) relatively inexpensive 12" wafers are manufacturable, and (v) unlimited combinations of BOX and film thicknesses can be achieved in order to match most device configurations (ultra-thin CMOS or thick-film power transistors and sensors). The defect density in the film is very low, the electrical properties are excellent, and the BOX quality is comparable with that of the original thermal oxide. The Smart-Cut process is adaptable to a variety of materials: SiC or III–V compounds on insulator, silicon on diamond, etc.

Other SOI technologies are full isolation by oxidized porous silicon (FIPOS) and zone melting recrystallization (ZMR) [1].

3. SOI devices

SOI circuits consist of single-device islands dielectrically isolated from one another and from the underlying substrate (Fig. 4a). The lateral isolation offers more compact design and simplified technology than in bulk silicon: there is no need of wells or interdevice trenches. In addition, the vertical isolation renders the latch-up mechanism impossible.

The source/drain regions extend down to the buried oxide, hence the junction surface, leakage current and junction capacitance are minimized. Obvious implications are improved speed, lower power dissipation, wider temperature range, and attenuated short-channel effects.

As far as the reliability is concerned, SOI MOSFETs are extremely robust to transient radiation effects. It is the permanent radiation-induced damage in the BOX which may be an issue. Another problem is the integrity of the gate oxide which is governed by the defect density in the film. Assistance of specialists in dielectrics is still needed for understanding the reliability aspects and the microstructure of the buried oxide.

It is in the highly competitive domain of low power/voltage circuits that SOI is most attractive. SOI offers the possibility to achieve a quasi-ideal subthreshold slope (60 mV/decade at room temperature), hence a threshold voltage below 0.3 V. Low leakage currents limit the static power dissipation, as compared to bulk Si, whereas the dynamic power dissipation is minimized by the combined effects of low parasitic capacitances and reduced voltage supply.

It has been repeatedly demonstrated that SOI circuits of generation (n) and bulk-Si circuits from the next generation (n+1) perform comparably. The speed record is for a CMOS technology with excellent short-channel behavior down to $L = 45$ nm, and 8 ps inverter delay [7,8]. More complex mainstreaming SOI circuits...
have also been fabricated: 0.5 V–200 MHz microprocessor [9], 4 Mbit SRAM [10], 16 Mbit and 1 Gbit DRAM [11], etc. [1,12]. Several companies (IBM, Motorola, Sharp) have announced the imminent commercial deployment of ‘SOI-enhanced’ PC processors and mobile communication devices.

The family of SOI devices also includes bipolar transistors (with a lateral configuration), high-voltage DMOS, smart power devices, and 3-D circuits [13]. Most innovative devices make use of the possibility to (i) combine bulk Si and SOI on a single chip (Fig. 2a), (ii) adjust the thickness of the Si overlay and buried oxide, and (iii) implement additional gates in the buried oxide (Fig. 2b). The interrupted BOX (Fig. 1b) allows the controlling of the vertical power devices, located in the bulk region of the wafer, by a low-power CMOS on SOI (Fig. 2a). Double SIMOX (Fig. 1b 3) has also been used to achieve a double-shielded, intelligent, high-voltage circuit [14].

SOI is an ideal material for microsensors (pressure, acceleration, gas flow, temperature, radiation, magnetic field, etc.) because the Si/BOX interface gives a perfect etch-stop mark, making it possible to fabricate very thin membranes, as shown in Fig. 2c [1,15].

The Gate all–around (GAA) transistor is fabricated by etching a cavity into the BOX and wrapping the oxidized transistor body by a poly-Si gate (Fig. 2d) [12].

4. Special mechanisms in SOI MOSFETs

4.1. Fully-depleted SOI transistors

In SOI MOSFETs, inversion channels can be activated at both the front Si–SiO₂ interface and the back Si–BOX interface (back-gate biasing \( V_{G2} \)) (Fig. 4a). Full depletion means that the depletion region covers the whole transistor body and does not extend with gate bias. A strong coupling develops between the gate bias and the inversion charge, leading to enhanced drain current. The front- and back-surface potentials become coupled too and the electrical characteristics of one channel vary remarkably with the bias applied to the opposite gate [16]. This interface coupling causes the front-gate measurements to depend on the back gate bias and quality of the BOX oxide and interface. Totally new \( I_D(V_G) \) relations apply to fully depleted SOI MOSFETs.

The threshold voltage decreases linearly with \( V_{G2} \), the subthreshold slope is a maximum for depletion at the back interface, and the transconductance reflects the possible activation of the back channel.

4.2. Partially depleted SOI transistors

In partially depleted SOI MOSFETs, the depletion charge does not extend from an interface to the other, and a neutral region subsists. If body contacts are not supplied, the so-called floating-body effects arise, leading to detrimental consequences. The kink effect (Fig. 3a) is due to majority carriers, generated by impact ionization, which collect in the body and increase the body potential (lower threshold voltage). In weak inversion and for high drain bias, a similar positive feedback is responsible for negative resistance regions, hysteresis in log \( I_D(V_G) \) curves, and eventually transistor latch (Fig. 3b).

The floating body may also induce transient effects. A drain current overshoot is observed when the gate is turned on (Fig. 3c). Majority carriers are expelled from the depletion region and collect in the neutral body. The drain current decreases gradually with time during electron–hole recombination. A reciprocal undershoot
occurs when the gate is switched from strong to weak inversion: the current now increases with time (Fig. 3c) as the majority carrier generation allows the depletion depth to shrink gradually.

5. Ultimate SOI MOSFETs

5.1. Short-channel effects

In both fully and partially depleted MOSFETs with submicron length, the lateral bipolar transistor is easily activated leading to positive (extra current) or negative (premature breakdown, vertical arrow in Fig. 3d) consequences. The breakdown voltage is especially lowered for n-channels, shorter devices, thinner films, and higher temperatures [17].

The self-heating effect, induced by the poor thermal conductivity of the buried oxide, is responsible for the mobility degradation, threshold voltage shift, and negative differential conductance (Fig. 3d).

The threshold voltage is reduced by charge sharing effect (CSE) and drain-induced barrier lowering (DIBL). An extra DIBL effect is due to the field penetration into the buried oxide and Si substrate. Fringing fields are responsible for an increase in the potential at the interface, film–BOX, as if the back gate (substrate) was driven from depletion to weak inversion. This drain-induced virtual substrate biasing (DIVSB) is critical in sub-0.1 μm FD MOSFETs, where the front-channel threshold voltage is lowered and the subthreshold swing is degraded [18].

The hot-carrier degradation mechanisms are more complex in SOI MOSFETs than in bulk Si for several reasons: presence of two oxides, two channels and related coupling mechanisms, different field distributions, and additional BOX damage which can affect the properties of integrated circuits [19]. In n-channels, the defects are created at the interface where the electrons flow; exceptionally, injection into the opposite interface may arise when the transistor is biased in the breakdown region. Although the device lifetime is relatively similar in bulk Si and SOI, the influence of stressing bias is different: SOI MOSFETs degrade less than bulk Si MOSFETs for $V_G \approx V_D/2$ (i.e. for maximum substrate current) and more for $V_G \approx V_T$ (i.e. enhanced hole injection). The device aging is accelerated by accumulating the back interface [19]. In p-channels, the electrons generated by front-channel impact ionisation may be trapped into the buried oxide. An apparent degradation of the front interface occurs via coupling [19].

5.2. Scalability of SOI MOSFETs

The key parameters in SOI are the doping level, film thickness, and BOX thickness [20]. Increasing the doping reduces the threshold voltage roll-off $\Delta V_T(L)$ resulting from both CSE and DIBL. There is no influence of the film thickness as long as the device is partially depleted. The worst case occurs for a doping–thickness combination that corresponds to the transition between partial and full depletion.

In fully depleted (FD) MOSFETs, the thinning of the film is always beneficial. The possibility to use ultra-thin films, with lower doping (i.e. higher mobility), is actually the main asset of SOI, as long as the $V_T$ roll-off is of primary concern. However, the control of the nominal threshold voltage and DIBL becomes more difficult in low-doped films [21]. To extend the design window, mid-gap metal gates and thin buried oxides (reduced self-heating and DIVSB, but increased parasitic capacitance) are being considered.

5.3. Innovating architectures

5.3.1. Dynamic-threshold MOSFET

The DT-MOSFET is a partially depleted transistor, where the gate and the body are interconnected. As the gate voltage increases in weak inversion, the simultaneous increase in body potential makes the threshold voltage to decrease. DT-MOSFETs achieve perfect gate-charge coupling, maximum subthreshold slope, and enhanced current, which are attractive features for sub-0.6 V CMOS circuits. A simple model compares the performance of DT and body-grounded (BG) MOSFETs using an analytically defined enhancement factor $k$: $V_{DT} = V_{BG}/k$, $S_{DT} = S_{BG}/k$, and $\mu_{DT} = \mu_{BG} \times k$ [22].

5.3.2. Ground-plane MOSFET

Initially, the ground plane (GP) for SOI was imagined as a highly conducting layer, made by pulse doping, at the interface, film–BOX (Fig. 4b) [23]. Such a GP cancels the fringing field, by providing an equipotential reference at the back interface which terminates the potential contours. A different approach consists in

---

Fig. 4. Schematic structures envisioned as ultimate SOI MOSFETs: (a) fully depleted, (b) pulsed doping, (c) field plate, (d) ground plane, (e) dynamic threshold, and (f) double gate.
locating a field plate underneath a very thin BOX (<6 nm, Fig. 4c) [24]. The field plate can be biased (i.e. second gate) to adjust the back channel potential and control the front channel threshold.

An alternative GP, more compatible with the present technology, can easily be fabricated, by ion implantation or bonding, underneath a conventional buried oxide (Fig. 4d). An optimum MOSFET architecture can be achieved by combining such a GP, with a 50–100 nm thick BOX, a low-film doping and a mid-gap gate [18]. Additional improvements are expected from the use of buried insulators with lower dielectric constant, including buried air gap structures [25].

5.3.3. Double-gate MOSFETs

The operation of double-gate (DG) SOI MOSFETs is based on the concept of volume inversion [26]: the formation of front and back inversion channels causes, by continuity, the spreading of minority carriers in the volume of a thin SOI film [26]. As many minority carriers flow in the middle of the film, surface scattering is reduced, enabling a higher transconductance and a reduced 1/f noise.

It is admitted that the DG-MOSFET will represent the final metamorphosis of an MOS transistor [27]. This is so because the scalability is improved by the double gate potential control which lowers both CSE and DIBL [23,27]. DG-MOSFETs with symmetrical configuration have been fabricated with GAA [12], Delta [28], and lateral epitaxial growth [29] technologies.

We now discuss the case of ultimately thin (down to 1 nm!) MOSFETs, fabricated by sacrificial oxidation. Their characteristics, in single-gate or double-gate modes, are very well behaved. It is fascinating that just a few monolayers of silicon are still able to maintain an MOS-like functionality.

When biased in volume inversion, with \( V_{G1} = \frac{t_{ox2}}{t_{ox1}} V_{G1} \), the 3-nm thick transistors reveal an outstanding increase in transconductance as compared to single-gate (SG) operation (Fig. 5a) [30]. This difference is explained using the Poisson and Schrödinger equations, solved self-consistently.

As the film (or quantum well) becomes thinner, the energy levels and their separation increase [31]. The difficulty to populate the upper levels explains the ‘abnormal’ increase in threshold voltage observed in sub-10 nm thick SOI MOSFETs [32].

In DG-MOSFETs, the in-depth potential profile is symmetrical, and the vertical field cancels in the middle of the film. The quantization effects are essentially thickness-defined, hence lesser than in SG-MOSFETs. In DG mode, \( V_T \) is lower and several subbands are populated, whereas in SG mode, only the ground level is involved in charge transport.

The spatial distribution of charge is symmetrical in DG-MOSFETs and most of the carriers flow in the middle of the film (Fig. 5b). In other words, quantum calculations reinforce the volume inversion concept as compared to the classical viewpoint: the Poisson-defined distribution showed more carriers at the film interfaces. The subthreshold swing in DG mode is ideal (60 mV/decade), better than in SG mode. In strong inversion, the total charge in DG mode is marginally higher than twice the inversion charge in SG mode. This implies that the difference in transconductance between SG and DG modes is related to the carrier mobility rather than to a charge effect.

A first-order model predicts the impact of carrier and field distributions on the mobility [30]. The electric field is negligible in the middle of the DG-MOSFET where most of the inversion charge is located. As electron-phonon scattering strongly depends on the field, the mobility is presumably enhanced in the center of the film. It is also assumed that surface roughness degrades the mobility near each interface (gray areas in Fig. 5b). The calculations show that the average field-effect mobility is far higher in DG mode than twice the value in SG mode, where the vertical field is stronger and many carriers flow in the ‘rough’ region near the front interface (Fig. 5b).

6. Conclusions

SOI technology offers the opportunity to integrate high performance and/or innovative devices which can push away the present frontiers of the CMOS downscaling. SOI actually appears as the final relay of the Si-based microelectronics. The short-term prospects of SOI depend on the penetration rate of SOI circuits into the market. Not only does SOI offer enhanced performance, but also most of SOI disadvantages (self-heating, hot-carriers, early breakdown, etc.) disappear for operation at low voltage. But the key challenge to SOI designers, process engineers, and managers is still to overcome the bulk-Si monopoly.
Acknowledgements

The prospective part of this work has been performed at the Center for Projects in Advanced Microelectronics (CPMA), Grenoble, France. The CPMA is a multiproject institute operated by the Centre National de la Recherche Scientifique (CNRS), the Laboratoire d’Electronique, de Technologie et d’Instrumentation (LETI, Grenoble), the Institut National Polytechnique de Grenoble (INPG), and the Institut National des Sciences Appliquées (INSA, Lyon). Special thanks are due to my colleagues T. Ernst, D. Munteanu, N. Hefyene, G. Ghibaudo, M. Gri and T. Ouisse.

References

[29] Wong HS, Chan KK, Taur Y. Self-aligned (top and bottom) double-gate MOSFET with a 25 nm thick silicon channel. IEDM Techn Dig 1997;427.