Innovative wafers for energy-efficient CMOS technology

ABSTRACT

For continued attractiveness and competitiveness of advanced electronic appliances such as smartphones, TVs, notebooks or tablets, the semiconductor industry is moving to "fully depleted" transistor technology to build integrated circuits. This technology comes in two flavors: planar and tri-dimensional (FinFET), each with its own advantages and challenges. This White Paper explains how innovative wafers, which are the foundations of silicon chips, will play a role to enable or facilitate the introduction of the planar and non-planar approaches to fully depleted technology, starting at the 28nm node. It also outlines the benefits that users can expect.

August, 2012

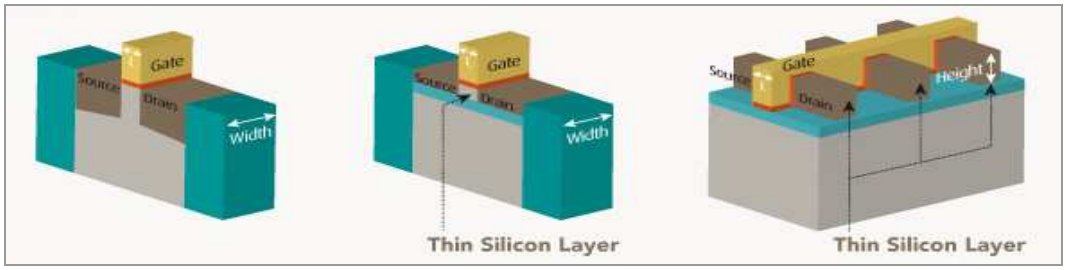
Rel.0.1.b

Semiconductor technology will evolve to serve consumer needs

To stay competitive in their market segments, computers, TVs, smartphones, tablets and other electronic devices must provide increasingly better value for money and an improved user experience – with, for example, new and compelling uses, powerful yet intuitive interfaces, faster response times, longer battery autonomy, and so on. To achieve this, the electronics industry has relied for decades on the continued miniaturization of transistors, which are the basic building elements of the integrated circuits ("chips"), at the heart of those products.
During these years of "happy scaling" of transistor technology, it was possible to add more and more functionality on a single chip while simultaneously increasing its performance (the operating speed of, e.g., the processor) and keeping its power consumption under control. However, those days are over. At the extremely small dimensions that transistors are reaching today (28nm and below), several effects that were secondary in the past are becoming predominant and very deleterious. The net result is the inability, at upcoming technology generations, to scale conventional semiconductor technology and obtain a satisfactory combination of higher transistor density, significant performance gains and low power consumption – despite soaring development costs.

To overcome these hurdles, there is consensus in the industry that the way to go is "ultra-thin body" or "fully depleted" devices. This is reflected in the ITRS (International Technology Roadmap for Semiconductor) *[1]*. The "ultra-thin body" qualifier refers to the geometry of the transistor, which gives it an advantageous behavior. The "fully depleted" qualifier, used somewhat equivalently, refers to an electrical property of such transistors – a full explanation requires serious microelectronics background, but is of marginal importance for this discussion. In a nutshell, it refers to the absence – or "depletion" – of some type of electrical charges (the majority carriers) in the transistors, which when present contribute to creating leakage currents.

A fully depleted (FD) transistor may be planar or tri-dimensional. In both cases, in contrast with traditional technology, the current between source and drain is only allowed to flow through a thin silicon region, defined by the geometry of the transistor (Figure 1). In addition, such transistors can eliminate the need for implanting "dopant" atoms into their channel. As a result, the behavior of the transistor is governed by different physical rules compared to conventional technology and is greatly improved. With those improvements, it becomes possible to go back to a much more satisfactory situation where further shrinking of the transistor dimensions enables creating more complex chips with better performance and, most importantly, with power consumption kept under tight control.



(a) bulk silicon transistor *(b) planar FD transistor* *(c) FinFET transistor*

Fig. 1: Contrary to conventional CMOS (a), with planar FD (b) or tri-dimensional FD a.k.a FinFET (c) the current is confined within a thin layer of silicon *[notional diagrams – not to scale]*

In the planar flavor of fully depleted technology (Figure 1.b), transistors are built flat on the silicon, as has been the case for decades. In contrast, the tri-dimensional flavor (Figure 1.c) requires fabrication of thin vertical "fins" of silicon in which current will flow from source to drain. There may be one or several fins per such FinFET transistor, around which the gate wraps. There are specific benefits and challenges for each of these two approaches. The semiconductor industry is introducing planar FD starting at the 28nm node, with first IC silicon expected in late 2012 and production in the following year. Tri-dimensional FD or FinFET, on the other hand, is expected below 20nm (and marketed as 16nm or 14nm) with, for the most advanced players, production in the 2015 time-frame (with the exception of Intel's "TriGate", launched in production at the 22nm node in 2012).

Wafers – the foundations of integrated circuits

The semiconductor industry fabricates integrated circuits side-by-side on wafers of silicon (Figure 2), their diameter is 300 mm (12'') for modern technologies (transition to 450 mm is envisioned for the end of the decade). The wafers are sawn into individual dice at the end of the fabrication process. Each die is then molded into a plastic package with metal pins or balls that are soldered onto the final application board, connecting the silicon die to the external world.

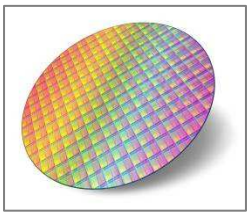


Fig. 2: Chips on a wafer

With conventional CMOS transistor technology, the blank wafers (on which the transistors that make up the chips are fabricated) consist of bulk silicon.
With FD technology, either planar or tri-dimensional, the transistors are either necessarily or advantageously fabricated on innovative silicon-on-insulator wafers. These consist of a thin layer of high-quality crystalline silicon, separated from a silicon base by a high-quality oxide, with very high bonding strength between those layers (Figure 3). The process to produce these wafers is the Smart Cut™ technology that Soitec employs and licenses to third-parties, as outlined in Figure 4.

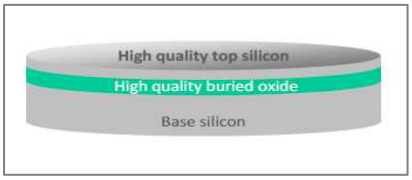


Fig. 3: general structure of a silicon-on-insulator wafer

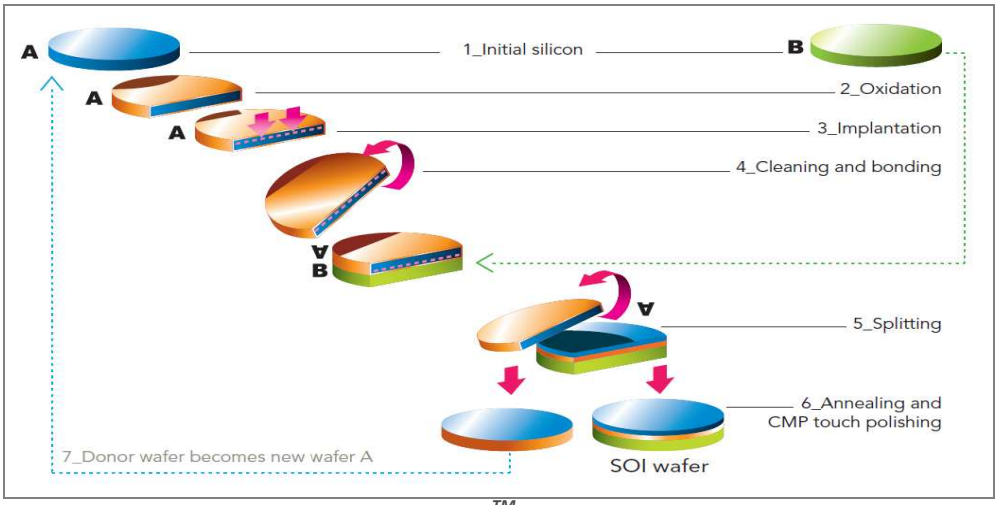


Fig. 4: Smart Cut™ process outline

Two different wafer product lines are available to serve the needs of both those customers who are keen to access FD technology as soon as the 28nm node with a planar approach and those who will industrialize FinFET technology.

FD-2D – The early and evolutionary path to fully depleted technology

The benefits of planar FD transistors have been extensively studied for over 15 years. However, this technology puts tight requirements upon starting wafers to live up to its promises. For advanced technology nodes, the top silicon layer must be ultra-thin – just a few nanometers thick (1nm = 10^{-9}m) – and extremely uniform, with maximum variation in the range of a few Angstroms (1Å = 0.1 nm). This is because, with FD, variations in silicon thickness translate into variations in the electrical behavior of transistors. Until recently many thought it was not realistic to expect cost-effective wafers meeting so stringent specifications.

However, today, manufacturers such as Soitec with its FD-2D product line provide starting wafers that meet these requirements and make planar FD technology a reality, with high volume availability to respond to market demand. The Smart Cut™ technology is licensed to ensure multi-sourcing options. This allows chipmakers to take advantage of the benefits of fully depleted silicon technology – specifically, energy efficiency and performance without compromising on transistor miniaturization – before FinFET (the tri-dimensional flavor of FD technology) is industrialized in foundries.

Figure 5 outlines the structure of a planar fully depleted transistor fabricated from an FD-2D wafer. Initially, for the 28nm technology node, the top silicon is generally 12nm thick and the buried oxide is 25nm thick. This enables fabrication of transistors with, depending on technology implementation choices, 5nm to 8nm silicon under the gate. Starting silicon thickness is always slightly greater than final thickness to allow for some silicon consumption inherent to the CMOS transistor fabrication process. Future generations can leverage even thinner buried oxide layers down to 10nm thick, providing a path for scalability of planar FD transistor technology down to the 14nm node.

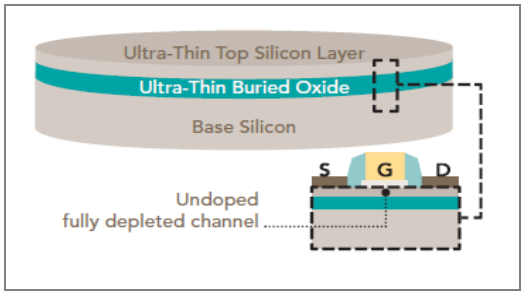


Fig. 5: Cross-section of a planar FD transistor fabricated on an FD-2D wafer *[notional]*

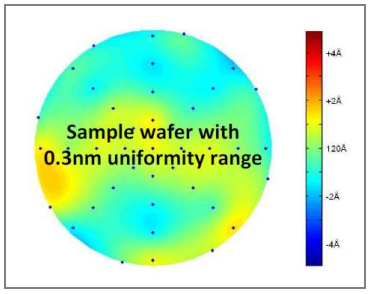


Fig. 6: Top silicon thickness map of a 300mm diameter wafer, with measured min-max range at 0.3nm

Soitec's FD-2D wafers have a top silicon thickness that is controlled to within just a few atomic layers, using the Smart Cut™ process as an "atomic scalpel". Figure 6 is an example of a wafer map that shows a silicon uniformity as good as 0.3nm, as measured by 41 point ellipsometry –

that's approximately equivalent to controlling to within 1mm the thickness of a layer that would span a range of 1000km, or 0.07 inches over a range of 1000 miles.

By enabling planar implementation of fully depleted technology, these wafers offer the opportunity to **access the benefits of FD technology today** – no need to wait for the 16nm/14nm technology node and the advent of FinFET technology. These benefits include:

- **Performance** *(GHz),* **power efficiency**, **operation at very low Vdd** (logic and SRAM),
- **Enhanced efficiency of some design techniques** (e.g., DVFS – Dynamic Voltage and Frequency Scaling),
- Plus, **unique back-bias efficiency** which can optionally be used to temporarily boost performance or cut leakage.

Early adopters of the technology are announcing very substantial performance and leakage gains as well as much improved energy efficiency *[2].* In addition, exceptional performance is maintained at very low power supply (e.g. 0.5V-0.7V), enabling ultra-low-power consumption in many use cases. At final product level, the early adopters forecast that a smartphone relying on this technology can enjoy an additional day of use between battery charges, or several additional hours of non-stop intensive usage such as high-definition video recording or playback, continuous high-speed Web browsing, etc *[3].*

In addition, with planar FD transistor technology, designers retain the methodologies, design flows and electronic design automation (EDA) tools they use with today's conventional technology, and can port legacy IP blocks and circuits in a straightforward way *[4]*. Furthermore, chip manufacturers use the same fab tools and production lines as well as extremely similar process steps. For example, at the 28nm technology node, the back-end of line (BEOL) of the CMOS process is 100% identical to a conventional low-power process on bulk silicon and the front-end of line (FEOL) has 80% in common, with fewer process steps overall *[2]* (about 10% fewer, including the elimination of all the steps related to channel doping in conventional CMOS).

Finally, different studies *[2, 5]* indicate that the cost of ownership of planar FD technology is extremely competitive compared to alternatives. For example, already at the 28nm node, the cost of finished FD wafers (i.e., at the end of CMOS processing) is on par with that of conventional low-power technology on bulk silicon while surpassing the overall performance of the more complex and more expensive high-performance (sometimes called "G-type") bulk technology. The final cost at die level could be even more favorable owing to different factors, including the fact that the improved electrical characteristics of FD will enable better parametric yield.

FinFET – Transition facilitated by innovative wafers

As outlined in Figure 1.c, a FinFET transistor consists of one or several fins of silicon around which the gate wraps. The fins must be electrically isolated from the underlying substrate.

One solution (Figure 7.a) consists of starting from a traditional bulk silicon wafer and completely handling the fin creation and electrical isolation through the CMOS process. The alternative

(Figure 7.b) is to start from a wafer that pre-defines some of the fin characteristics and embeds the electrical isolation, thus simplifying the CMOS process.

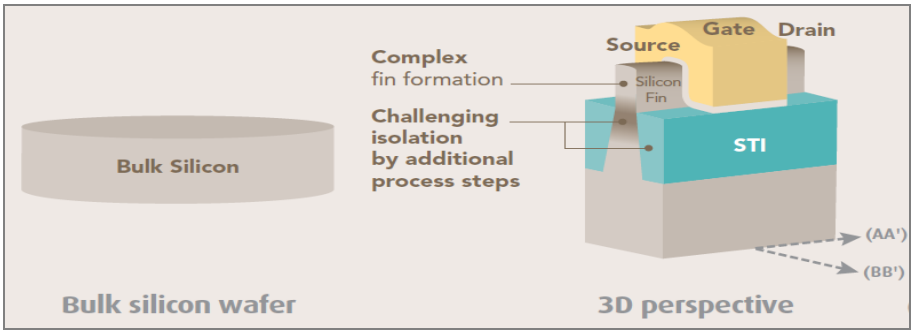


Fig. 7.a: FinFET on bulk silicon wafer (one fin shown) *[notional – not to scale]*

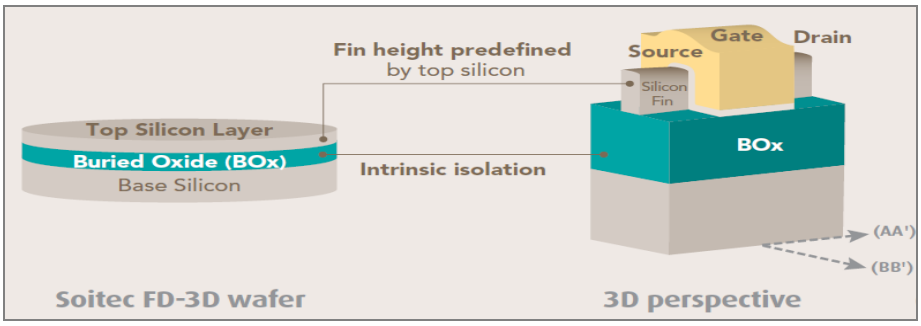


Fig. 7.b: FinFET on FD-3D wafer (one fin shown) *[notional – not to scale]*

Challenges of FinFET on traditional wafers

In the first case, fin formation typically involves creating trenches and filling them with an oxide (involving "shallow trench isolation", STI, process steps), followed by planarization and recessing the oxide by time etch (Figure 8) – that is, some of the oxide is etched away to uncover the silicon fins, with the amount of oxide removed (hence the height of the fins) dependent on the duration of the etch process.

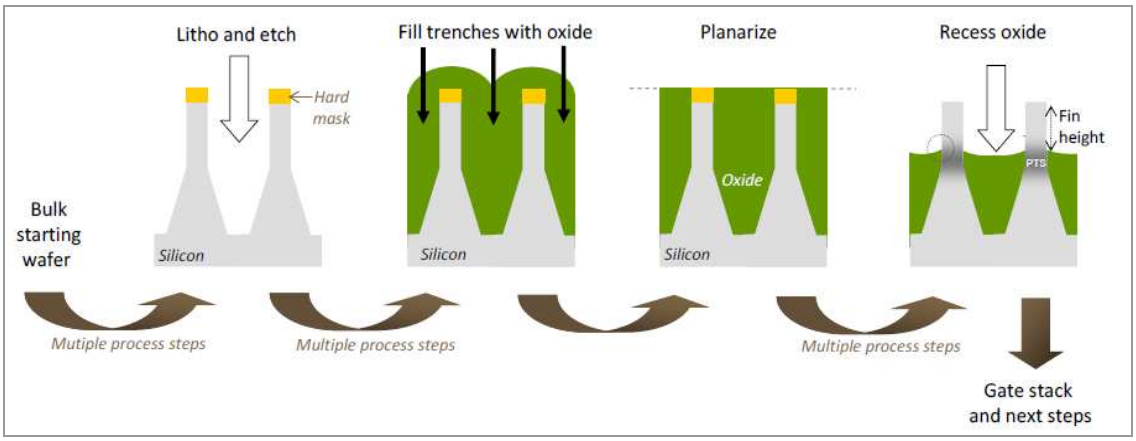


Fig. 8: Defining the fin on a bulk silicon wafer (showing BB' fin cross-section) – Outline of typical steps

Clearly, with this approach, controlling and repeating the same height and profile for a multitude of fins (also from wafer to wafer) is complicated and is likely to be achieved to a limited extent only or with trade-offs. And because of the physics of FinFET, any variability or trade-off in fin geometry will translates as variability or trade-off in electrical characteristics.

The oxide that is left between the fins ensures electrical isolation from one fin to the next. In addition, there is also a need to break any electrical path under the fin, to avoid "punch-through" current leaking from source to drain. A typical solution relies on inserting, by some adequate steps in the CMOS process, a highly doped layer (PTS, punch-through stopper) under the fin to create a junction and provide that isolation. However, optimizing this layer and properly implementing the associated process steps is challenging. Its exact location and its dopant concentration (not too high, not too low) are critical to achieve acceptable leakage *[6]*. Finally, creating a highly doped layer under the fin without unintentionally doping the fin is extremely challenging too. If the fin is doped, then variability creeps in because, given the tiny dimensions of transistors in advanced technology, the electrical behavior of the fin will depend on the exact number and location of the dopant atoms – which are uncontrollable. Excessive variability contributes to increasing total leakage at the circuit level and limits the ability to operate at reduced power supply and low energy consumption. Therefore, overall, the fabrication process is complex, likely to pose serious challenges in terms of controllability in a repeatable and industrial way, and raises the risk of disappointing electrical results – especially as dimensions continue to shrink beyond the 16nm node.

FD-3D wafers simplify FinFET technology

An alternative solution consists of starting from a wafer such as Soitec's FD-3D (Figure 9): it pre-integrates critical structures of the transistor, in the sense that fin height is directly driven by the thickness of the top silicon, and fin isolation is natively present under the form of the buried oxide. This means fewer steps are necessary in the CMOS process to define the fins and perform fin isolation, as outlined in Figure 10.

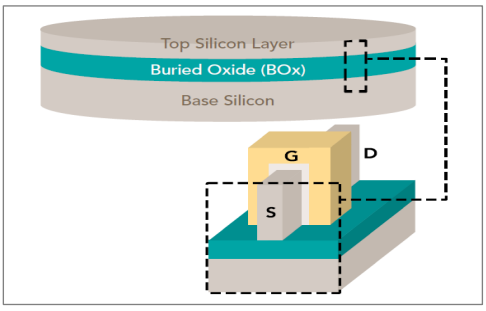


Fig. 9: Perspective view of a FinFET transistor fabricated on an FD-3D wafer

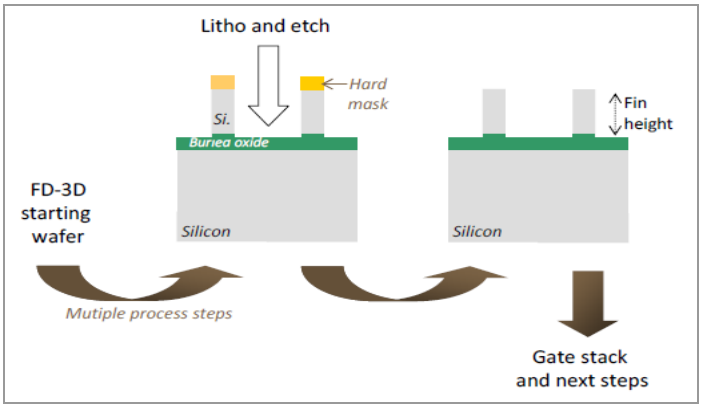


Fig. 10: Simplified fin formation and isolation with FD-3D wafer – outline flow

As a result, the FinFET fabrication process on FD-3D wafers is significantly simpler than on bulk silicon *[7]*. Specifically:

- Creating fins is facilitated by the fact that the buried oxide acts as an etch stop. Therefore the height of fins can be determined from the thickness of the top silicon on the starting FD-3D wafer, rather than depending on the accuracy of an oxide recess and on the shape of the oxide at the foot of the fin. Consequently each fin is more clearly defined; its height and profile are more easily repeatable.
- The buried oxide embedded in the FD-3D wafer provides both the isolation between fins and optimum isolation of each fin from the underlying substrate thus cutting leakage. Therefore, there is no need for STI nor PTS and their complex optimization.
- The gate, source, drain and channel all stop abruptly on the buried oxide, in a reproducible way. This helps to obtain a good performance/leakage ratio as well as low variability.

Typically, four lithography steps and over 55 process steps – some of them quite challenging – can be saved. That translates as lower capital expenditure, lower operating expenses, higher production throughput and lower overall cost of ownership. That also means fewer industrialization challenges, which presents benefits in terms of technology development and time to market for this technology.

Benefits for the final products

The ability to build fully isolated FinFETs on FD-3D wafers makes it possible, if wished, to have truly undoped fins – be it in selected macros (e.g. SRAMs) or in a more generalized way – and to avoid variability from random dopant fluctuations. That, plus superior fin geometry control, opens the door to **better VDDmin** (the minimum voltage at which SRAM and logic can be properly operated) and therefore **lower power consumption**, greater opportunities **for better parametric yield**, and better global SOC leakage power (as less variability means fewer transistors with leakage much higher than nominal) *[7]*. The built-in isolation (BOX layer) helps obtaining improved transistor leakage which completes the **reduction of SOC leakage power**. Finally, using a significantly simpler fabrication process with fewer lithography and processing steps, on top of contributing to improving the processed wafer cost, opens up opportunities to achieve **better defectivity** and better functional yield.

Longer term

The next step for the semiconductor industry, after the transition to planar and tri-dimensional fully depleted technologies – still essentially silicon-based – is bound to be the introduction of new materials in the channel. Solutions based on germanium and III-V compounds such as InP, GaAs and others are being actively researched. In an industrial solution, it is likely that silicon or in any case a low-cost material will still be used as mechanical support for the starting wafers. In parallel, the transition to 450mm diameter wafers is expected to occur for advanced nodes entering production at the end of this decade.

In this context, the Smart Cut™ layer transfer technology may, again, prove extremely valuable to facilitate reaching a satisfactory solution, by enabling independent control over various optimization knobs such as electrical properties, mechanical support, crystalline quality or

thermal conduction. For example, transferring a thin layer of high-quality III-V material onto a low-cost handle wafer (silicon or other), with an optimized interfacing layer, could be an interesting option.
Besides the introduction of new materials, one may also see the advent of new transistor architectures such as gate-all-around or nanowires. Again, substrate engineering will have a role to play to provide the industry with the most suitable starting wafers.

Conclusion

Fully depleted silicon technology is coming. The question is how fast and how easily this transition can be accomplished. Innovative wafers provide part of the answer, both for planar and tri-dimensional implementations. FD-2D wafers from Soitec and its licensees enable a planar implementation of fully depleted technology, providing the semiconductor ecosystem with an early and low-risk path to reap the benefits of FD-based circuits as soon as the 28nm node – thereby offering optimal performance and power efficiency across all use cases, today. For those chip makers whose products will afford post-20nm nodes, FinFET is the most favored option at such geometries, but its manufacturability and cost competitiveness are extremely challenging, especially as dimensions continue shrinking. Again, innovative wafers such as Soitec's FD-3D can play a role in addressing these challenges in a timely manner and making the most of FinFET technology. Looking further ahead, the Smart Cut™ technology will enable independent optimization of various key aspects of the wafers, continuing to simplify the implementation of the next silicon technology breakthroughs.

About the author
Xavier Cauchy is in charge of Digital Applications and Strategic Marketing at Soitec.

References

[1] International Technology Roadmap for Semiconductors (ITRS), *www.itrs.net*
[2] White Paper, "Planar fully depleted silicon technology to design competitive SOC at 28nm and beyond", STMicroelectronics - *http://www.soiconsortium.org/about-soi/white-papers.php*
[3] ST-Ericsson data – Disclosed at MWC ST-Ericsson Media & Analyst Briefing at Mobile World Congress, February, 2012 - *http://www.media-server.com/m/p/7k42mcqs*
[4] White Paper, "Considerations for bulk CMOS to FD-SOI design porting", SOI Consortium
[5] Cost study by IC Knowledge, - *http://www.icknowledge.com/news/presentations.html*
[6] "SOI versus bulk-silicon nanoscale FinFETs", *Jerry G.Fossum et al., Solid-State Electronics, Volume 54, Issue 2, February 2010.*
[7] « SOI Value in IBM Silicon Technology », W. Clark – E. Nowak, IBM, Oct.2011 - *http://www.gsaglobal.org/3dic/docs/20111019_IBM_SOI_Value_GSA.pdf*