Fully depleted silicon technology to underlie energy-efficient designs at 28 nm and beyond

Xavier Cauchy - February 20, 2013

Industry Need for Continued Scaling
Technological advances in transistor scaling have had a dramatic effect on consumer electronics and their corresponding use cases. In 1973, Motorola developed the first mobile phone, which weighed 2.5 pounds, was 9 inches long, had limited battery life and only allowed users to make and receive calls. Fast forward to today's mobile devices that fit in the palm of your hand, with batteries that last all day and more computing power than ever thought possible.

While it has taken 40 years to come this far, innovation has been exceptionally rapid over the course of the past 10 years, and consumer expectations have accelerated at a similar pace. What sort of features and computing capabilities will we expect of our mobile devices five years from now? How about in 10 years? Future improvements largely hinge on the industry's ability to continue on the path of Moore's Law by producing ever-smaller transistors with ever-greater performance. Satisfactory scaling fulfills two core requirements: the need for smaller transistors that reduce costs and a parallel need for improved performance and lower power consumption.

To date, transistor scaling has continued in accordance with Moore's Law down to 32 nm. Engineering challenges, however, are forcing chipmakers to compromise performance and power efficiency in order to reach smaller nodes - unless they switch to new technologies that help better solve these challenges. Today, the semiconductor industry is starting to deploy such new technologies, largely relying on "fully-depleted" transistors for continued scaling and performance gains.

Fully Depleted Silicon Technology
A fully depleted (FD) transistor can be planar or tri-dimensional. In each case, in direct contrast with other technologies commonly used today, the current between source and drain is allowed to flow only through a thin silicon region, defined by the physical parameters of the transistor.

In the planar design of fully depleted technology, transistors are built flat on the silicon. For the three-dimensional alternative, manufacturers fabricate thin vertical "fins" of silicon in which current will flow from source to drain. Additionally, FD transistors can eliminate the need for implanting "dopant" atoms into the channel. These improvements help chipmakers secure gains in both energy efficiency and performance that are required from scaling silicon technology.
In Figure 1a representing a conventional CMOS transistor, the extension of the depletion zone is a function of doping and is variable (modulated by the drain voltage). Its extent correlates well with the notorious "short channel effects" which severely affect the behavior of the transistor.

In contrast, FD transistors, illustrated in figures 1b and 1c, differ from traditional transistor architectures by having a channel that is not defined by its doping level, but rather by physical boundaries - in which case the depletion zone fills the full body of the transistor. This design improves gate control over the channel, which enhances performance and cuts leakage.

Three-dimensional FinFET architectures are in the short-term planning phases of several world-leading foundries to scale CMOS technology to 16 nm and beyond, with the most aggressive schedules aiming for high-volume production no later than 2015, despite the challenges of manufacturing this technology.
For planar architecture, FD-SOI is currently being implemented at the 28-nm node by industry-leading companies such as ST Microelectronics and its partner ST-Ericsson. Allowing for a smooth evolution from conventional planar CMOS technology, FD-SOI is a comparatively simple technology, with its proponents arguing that both power consumption and performance figures are comparable to those of FinFET.

One aspect to keep in mind when considering scaling options is lithography. The need to print smaller dimensions in order to fabricate advanced chips is dramatically increasing the cost of lithography and its share in the overall process cost. To make things worse, beyond 28nm, sophisticated and costly multi-pass patterning, such as double patterning, is required to continue using existing lithography equipment based on 193nm wavelength light sources.

Upcoming lithography based on Extreme Ultra-Violet (EUV) wavelength light sources will have the ability to print tiny dimensions in a single pass, but volume production using EUV technology is not expected for several years. These soaring costs are prompting some companies to think twice before moving their chips to smaller nodes and to seriously evaluate alternatives that offer immediate, competitive results for lower R&D and manufacturing costs. In this context, 28nm planar FD-SOI technology can be a smart choice to obtain next-generation performance and energy efficiency without rushing to 20nm or 14nm.

The planar FD-SOI approach

While fully-depleted silicon technology has been studied for more than a decade, strict manufacturing standards are required from the blank starting wafers on which transistors are built in order to exploit the technology's full potential. For advanced technology nodes, the top silicon layer, which drives channel thickness under the gate, must be just a few nanometers thick (1nm = 10^{-9}m) and with a maximum variation in the range of just a few Angstroms (1Å = 0.1 nm). This is because, with planar FD-SOI, variations in silicon thickness translate into variations in the electrical behavior of transistors.

Today, manufacturers such as Soitec, with its FD-2D product line, provide starting wafers that meet these requirements and make planar FD-SOI technology a cost-effective option for chip designers, giving them the opportunity to obtain rapid access to performance and efficiency gains.

Adopters of the technology are experiencing a wide range of improvements over traditional technologies - including better performance and energy efficiency across the full range of power supply, exceptional performance at very low Vdd (e.g., 0.6V-0.7V), enhanced efficiency of DVFS (Dynamic Voltage and Frequency Scaling) and significant boosts in performance and leakage control.
through the optional use of a back-bias.

For example, ST and ST-Ericsson have fabricated a smartphone chip based on 28nm FD-SOI in which the ARM dual Cortex-A9 CPU can reach 800MHz at a mere 0.6V and over 1.5GHz at just 0.85V [1]. In comparison, a low-power 28LP technology delivers only a few hundred MHz of processing power, while a high-performance 28HP technology would consume much more leakage power and dynamic power to deliver comparable performance [2].

FD-SOI also provides best-in-class peak performance: the same physical Cortex-A9 cores reach an impressive 2.5GHz operating frequency when the supply voltage is raised, with substrate biasing usable as a powerful knob to get very high performance at optimized power consumption. The end result of these improvements is mobile devices that offer an additional day of medium use between battery charges, or several additional hours of intensive applications, viewing of HD multimedia content or high-speed web browsing.

In addition, because FD-SOI is fully compatible with traditional planar technology, it does not disrupt the design methodology, meaning designers keep the same flows and tools as what they would use with conventional CMOS design.

The three-dimensional (FinFET) approach A FinFET transistor consists of one or several fins of silicon around which the gate wraps. All fins have the same physical width and height, and the driving strength of the transistor is determined by its number of fins and is therefore quantized - contrary to a planar transistor whose driving strength can be selected from a continuum of values by adjusting the transistor width.

Controlling the exact shape and dimensions of each fin in each chip, especially their height, is important to reap the benefits of FinFET technology. Additionally, to avoid “punch-through” current leaking from source to drain at the bottom of the fin, efficient electrical isolation beneath each fin also is required.

FinFET architectures can be produced either on bulk silicon or on silicon-on-insulator starting wafers. Building FinFET on the latter presents a number of advantages, especially as dimensions continue to shrink beyond the 16nm node.

The advantages of building FinFETs on silicon-on-insulator wafers stem mainly from easier fabrication of the fins and better control over their physical parameters [3]. To accomplish this, these wafers pre-integrate critical parameters of the transistors, that is, fin height is directly driven by the thickness of the top silicon on the wafers and fin isolation is natively present and total, under the form of a buried oxide.

Conversely, when starting from bulk silicon, the fins and their isolation have to be formed as part of the CMOS process, which is more complex and more subject to variability. Fin isolation in that case typically involves forming a highly doped layer called PTS (punch-through stopper) under the fin - which is complex to optimize.

Moreover, the ability with silicon-on-insulator wafers to build fully isolated FinFETs makes it possible, if wished, to have undoped fins for use in selected macros (like SRAMS) or in a more generalized way, thus avoiding the variability associated with random dopant fluctuations (RDF). The combination of low RDF and superior fin geometry control opens the possibilities of better VDDmin (the minimum voltage at which SRAM and logic can be properly operated) and lower power consumption and allows for better parametric yield and improved system-on-chip leakage power.
References


About the author
Xavier Cauchy is currently Digital Applications & Strategic Marketing manager at Soitec. He has held various technical, business and management positions over his 18+ years in the high-tech industry, many of them in the field of System-on-Chip development. He holds a post-graduate degree in electronics from ISEN (France) and an MSc in digital communications from King's College, University of London.