Fully Depletion of Advanced Silicon on Insulator MOSFETs

M.K. Md Arshad, N. Othman & U. Hashim

Institute of Nano Electronic Engineering, Universiti Malaysia Perlis, Kangar, Perlis, Malaysia

Published online: 18 Feb 2015.

To cite this article: M.K. Md Arshad, N. Othman & U. Hashim (2015): Fully Depletion of Advanced Silicon on Insulator MOSFETs, Critical Reviews in Solid State and Materials Sciences, DOI: 10.1080/10408436.2014.978447

To link to this article: http://dx.doi.org/10.1080/10408436.2014.978447

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Fully Depletion of Advanced Silicon on Insulator MOSFETs

M.K. Md Arshad*, N. Othman, and U. Hashim

Institute of Nano Electronic Engineering, Universiti Malaysia Perlis, Kangar, Perlis, Malaysia

Scaling of the transistor has been tremendous successful in the beginning with reduction of the gate oxide thickness and increase of doping concentration. Moving into smaller dimension, those are not enough to overcome the short channel effect. Starting with changing in materials and followed by device architecture is needed which require fully depletion operation. This article reviews the fully-depletion operation of thin body of silicon on insulator of advanced MOSFETs.

Keywords: fully-depletion, SOI MOSFETs, thin-body and thin-buried oxide (UTBB)

Table of Contents

1. INTRODUCTION ............................................................................................................................... 1

2. SILICON ON INSULATOR (SOI) ........................................................................................................ 3
   2.1. PD and FD-SOI MOSFETs ......................................................................................................... 3
   2.2. Short Channel Effects in SOI ................................................................................................... 5

3. FD-SOI MOSFETs ............................................................................................................................... 5
   3.1. MASTAR Equations ............................................................................................................ 5
   3.2. UTBB Electrostatic Features .................................................................................................. 7
   3.3. Figures-of-Merit ........................................................................................................... 8

4. CONCLUSION ............................................................................................................................... 9

REFERENCES ....................................................................................................................................... 10

1. INTRODUCTION

In 1965, Gordon Moore, a co-founder of both Fairchild and Intel, described in his paper, based on five data points available from Fairchild at that time: “with unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip.”\(^1\) Ten years later, in 1975, he revisited the prediction and surprisingly the actual data scatter pretty well along the prediction line, which evidenced that the number of devices per integrated circuit increase exponentially with increasing time. This astonishing prediction was later famously named as Moore’s Law and has brought significant impact to the industry. Moreover, it later became a basis of industry forecasts such as those of the ITRS\(^2\) for transistor scaling.

The successful of this classical scaling is initially supported with the scaling concept published by Dennard in 1974,\(^3\) as illustrated in Table 1 (constant-field scaling). The simple concept of scaling is to reduce all of the physical dimensions by the same amount of \(\alpha\). To keep the electric field constant, the channel doping needs to be increased and the applied voltage to be reduced. This will cause the depletion regions within the devices to scale as much as the other dimensions. A first important result of scaling is increased circuit density since the area is divided by \(\alpha^2\). This was seen as a key to reducing

\(^1\)Critical Reviews in Solid State and Materials Sciences, 0:1–15, 2014

Copyright © Taylor & Francis Group, LLC
ISSN: 1040-8436 print / 1547-6561 online
DOI: 10.1080/10408436.2014.978447

\(^*\)E-mail: mohd.khairuddin@unimap.edu.my

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/bsms.
manufacturing costs. A second important result that underlies the speed and power benefits by a factor of $2^\alpha$ and $2^{\alpha^2}$, respectively, is the reduction of capacitance per circuit. Furthermore, as a constraint of the voltage that is not usually scaled as fast as the linear dimension due to subthreshold leakage constraint, additional scaling factor $\varepsilon$ for the electric field is introduced to account for the increased of electric field ($\varepsilon$ is greater than one). This is summarized under “generalized scaling factor.”

The continuous and systematic increase in transistor density and performance, as described in Moore’s Law and guided by Dennard’s scaling theory, has been remarkably successful for the development of silicon MOSFET technology for the past 50 years. This can be seen with many generations of smaller devices and significant increase of transistors count per chip for faster, power efficient and high performance microprocessors and more importantly at reduced cost. This classical MOSFET scaling technique was followed successfully until 90 nm transistor generation (130 nm was the last CMOS MOSFET scaling technique was followed successfully until 90 and 65 nm generations when the classical MOSFET scaling methods are no longer effective.

Nevertheless, there was a need to reduce the gate-oxide leakage with high-$k$ technology for reduced leakage and improved performances. A hafnium-based dielectric (high-$k$ material) was introduced in 45 nm generation node as replacement to SiO2. High-$k$ material provides a gate oxide that is physically thicker, which reduces leakage, but has thinner electrical equivalence, which improves transistor performance. The polysilicon gate electrode was replaced by two different metal-gate materials (for NMOS and PMOS) to eliminate the poly-depletion effect. The thin gate oxide provided by high-$k$ and the metal gate also helped to reduce transistor $V_{th}$ variability. In addition, the 45 nm generation introduces different process flow known as gate-last. The gate-last flow allows the metal gate materials to be deposited after high temperature source and drain formation steps are completed, thus offering a wider set of materials options for tuning NMOS and PMOS gate-work functions. As a result, the gate-last flow has additional benefit of enhancing channel strain during the removal of the sacrificial polysilicon gate which further increase the transistor performances.

The 32 nm technology node utilized second generation of high-$k$ and metal-gate transistors along with previous benefits obtained in fourth generation of strained silicon.

For the latest 22 nm generation node, Trigate 3D-MOSFET architecture has been opted to extend the scalability with improvement in performances compared to classical planar transistor. Trigate provides fully depleted operation due to narrow fins, as the fins are wrapped by the gate electrode around three sides, providing improved electrostatic control of the channel. The improved channel control provides steeper sub-threshold slope, reduced leakage current, higher performances, etc. Although there is a significant improvement in term of electrostatic behavior, but this alone is not sufficient to deliver the expected performance and benefits. The strained Si, gate-last process, and high-$k$ and metal gate still need to be applied as in previous nodes.

Figure 1, depicts the chronological advancement in materials and structures for 90 nm down to 22 nm technology nodes as previously discussed.

Now, one can see that, as the transistor moves into extreme smaller dimensions, it needs to incorporate innovations such as strained-Si, high-$k$ + metal gate, and 3D-structures. It is no longer limited to shrinking the geometries to make more space available for more transistors at reduced cost per transistor, the transistors to provide tensile channel strain to improve electrons mobility. The 65 nm generation introduced two years later further improved these strain techniques and subsequently increase the transistor performances.

### Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Constant-field scaling</th>
<th>Generalized scaling factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical dimensions ($L, W, T_{ox}, X_j$)</td>
<td>$1/\alpha$</td>
<td>$1/\alpha$</td>
</tr>
<tr>
<td>Electric field ($\varepsilon$)</td>
<td>$1/\alpha$</td>
<td>$\varepsilon/\alpha$</td>
</tr>
<tr>
<td>Body doping concentration ($N_a$)</td>
<td>$\alpha$</td>
<td>$\varepsilon/\alpha$</td>
</tr>
<tr>
<td>Supply voltage ($V_{dd}$)</td>
<td>$1/\alpha$</td>
<td>$\varepsilon/\alpha$</td>
</tr>
<tr>
<td>Transistor current ($I$)</td>
<td>$1/\alpha$</td>
<td>$\varepsilon/\alpha$</td>
</tr>
<tr>
<td>Capacitance ($C = \varepsilon_{ox} A / T_{ox}$)</td>
<td>$1/\alpha^2$</td>
<td>$1/\alpha^2$</td>
</tr>
<tr>
<td>Area ($A$)</td>
<td>$1/\alpha^2$</td>
<td>$\varepsilon^2/\alpha^2$</td>
</tr>
<tr>
<td>Power dissipation ($P \sim IV_{dd}$)</td>
<td>$1/\alpha^2$</td>
<td>$\varepsilon^2/\alpha^2$</td>
</tr>
<tr>
<td>Power density ($P/A$)</td>
<td>$1/\alpha^2$</td>
<td>$\varepsilon^2/\alpha^2$</td>
</tr>
</tbody>
</table>
but it involves materials advancement, new structures as well as advances in processes which need to be integrated together for achieving the same goals. In the next generation, continuing the shrinking is even more challenging. In response to that, several approaches have been identified with some of them addressed and implemented while some are still at the research levels. Figure 2 illustrates the possible solutions:\textsuperscript{20,21}

- **Improve the carrier transport properties by introducing new materials.** The strained-Si technology has been widely used to achieve high drive currents. As shown in Figure 2, silicon only as a channel is not necessarily the best material anymore. Changing the channel material or incorporating silicon channel with higher mobility materials (SiGe, Ge, III-V semiconductors, or carbon nanotubes) will lead to further performance improvement to extend the CMOS generation.

- **Improve the electrostatics of MOSFETs by introducing new structures.** Different transistor structures (Figure 3) should be considered to improve the electrostatics and control of short-channel effect especially when entering sub-nanometer regime. A double-gate transistor has much better short channel effect immunity than a single-gate transistor because of better control of the channel by the gates. The ideal device structure in terms of the electrostatic control is a gate-all-around which leads to the nanowire transistor structure.

- **Suppress the variability.** In order to thoroughly eliminate Random Dopant Fluctuation (RDF) problem, a transistor with an intrinsic channel is required, which can be achieved in fully-depleted silicon-on-insulator (FD-SOI) structure.

The next generation of transistors should be able to solve all the challenges mentioned, with combinations of solutions preferable to achieve the improvement in performances. Of course to discuss everything is single article will be very lengthy, thus we focus on the thin body silicon-on-insulator.

2. **Silicon on Insulator (SOI)**

SOI technology provides a good insulation from substrate injection noise (crosstalk) due to the oxide insulation between devices especially between the analog and digital devices. It provides an additional degree of freedom when selecting the resistivity of the substrate to reduce the capacitive coupling and the losses.\textsuperscript{22–24} In addition, the shallow junction is easy to achieve (in the case of thin silicon film) thanks to its dependence on silicon thickness.

SOI also features some drawbacks such as self-heating. In SOI devices the heat dissipation is not as efficient as in bulk Si due to presence of the insulator which has lower thermal conductivity than silicon. This can lead to performance device degradation or sometime failure over the operating period. However, self-heating is expected to be improved with implementation of thin BOX.\textsuperscript{25–27}

For the case with thicker silicon film, partially depleted devices are more prone to undesirable results due to floating-body effects. For fully depleted SOI (thinner Si film), one needs stringent manufacturability process control on the uniformity of the silicon thickness and to prevent the surface states at the BOX/substrate interface. Similarly if the thin BOX is employed, the quality of thin BOX and the interface requires stringent control.\textsuperscript{28,29}

2.1. **PD and FD-SOI MOSFETs**

In classical SOI transistor, the current flows between source and drain are controlled primarily by the gate and secondarily by the substrate bias that acts a back gate. As a result, two inversion channels may be activated, i.e., at the front Si-SiO\textsubscript{2} interface (front interface) and at the Si-BOX interface (back interface). In PD-SOI the two space-charge regions are independent and for FD they are coupled. This behavior is very much related to the thickness and doping of the silicon films and can be categorized physically depending on the maximum depletion width ($X_{dmax}$) given by (1):

$$X_{dmax} = \sqrt{\frac{4\varepsilon_S \Phi_f}{qN_A}}$$  \hspace{1cm} (1)
Partially depleted SOI (Figure 4a) can be defined as a device for which the silicon film is larger than twice the value of $X_{dmax}$. This commonly occurs in thick film SOI devices or depending on doping concentration in the Si films. The silicon and buried oxide thickness of 50–90 nm and 50–145 nm, respectively, can be considered as PD-SOI devices. PD-SOI features similar characteristics as in bulk as the body is never completely depleted. Two independent channels may be formed close to the interfaces, with presence of a neutral region in between. This neutral region is known as a floating body effect (FBE) which leads to undesirable device performance such as kink-effect, parasitic BJT effects, anomalous subthreshold slope, reduced drain breakdown voltage, and other effects. The problems are diminished in body-tied or dynamic-threshold MOSFET (DTMOS) structures.

Fully depleted (Figures 4b,c) SOI can be obtained when the silicon film thickness is smaller than $X_{dmax}$. Ultra-thin silicon with standard buried oxide (UTB; Figure 4b) and ultra-thin silicon with ultra-thin buried oxide (UTBB; Figure 4c) can be considered as fully-depleted devices. As a guide, silicon and BOX thicknesses of less than 30 nm and 150 nm is considered as UTB, while for UTBB, the BOX is much thinner, notably below 50 nm.

In FD-SOI, the depletion region covers the whole transistor body irrespective of the bias applied to the back gate. As a result, the front- and back-channel are coupled, therefore the electrical characteristics of one channel depends on the bias applied to the opposite gate. This excellent coupling offers improved drive current and near ideal subthreshold slope due to smaller body factor. Body factor (normally represented by the letter $n$), is an image of the coupling efficiency between the front-gate voltage and the channel. The closer $n$ is to unity, the sharper the transition between the off and on-states of the transistor. More detailed explanation about the body factor can be found here. Since the body film is fully depleted, the device characteristics become more complex, i.e., the front-gate characteristics include contributions from the BOX/Si interface and highly depend on the back-gate bias. The FD-SOI devices can be considered free from FBE effects. For FD-SOI with ultra-thin body (typically $\sim 10$ nm), due to super-coupling effect accumulation regime at the back interface cannot be achieved with simultaneous inversion at front interfaces or visa-versa. In this case, the potential distribution remains constant along the front- to the back-gate of the channel. Super-coupling effect is not only dependent on silicon body thickness, but also ultra-thin BOX with combination of applied voltage between the front and back-gates.

The above definitions can be used to distinguish between PD and FD-SOI, typically for thick and thin Si body, respectively. However, there is a region with medium thickness of silicon body in which conventional depletion or floating body effects coexist. For devices with this intermediate body thickness, the electrical characteristics may be more complex than those in pure PD or pure FD devices.

![Basic architecture of (a) partially depleted (PD), (b) ultra-thin body with standard buried oxide thickness (UTB) FD, and (c) ultra-thin body and ultra-thin buried oxide (UTBB) FD-SOI MOSFETs.](image-url)
SOI body where \( X_{\text{dmax}} < t_{\text{Si}} < 2X_{\text{dmax}} \) which can be said as transition between partial to full depletion. In this case, the device operation can be changed from PD to FD due to contribution of source and drain lateral depletion regions. It is dependent on gate length, channel doping and back-gate bias.\(^{22,29,35}\) This criterion is only valid for classical SOI MOSFETs and it does not occur for advanced SOI MOSFETs featuring thin Si body and thin buried oxide.

### 2.2. Short Channel Effects in SOI

Short channel effects (SCE) are the phenomena by which the threshold voltages are seen to decrease and off-state current seen to increase as the gate length is reduced. This is a consequence of the fact that as gate length is decreased, the depletion regions associated with the source and drain regions become closer and start to interact with each other. These depletion regions are the region of high electric fields which propagate through the depletion regions associated with the junctions. They facilitate carrier transport directly between the source and drain regions, which gives rise to the observed phenomena of higher off-state currents, reduced threshold voltages, and increased statistical fluctuations, all of which pose severe drawbacks such as degraded mobility, increased capacitances, and parasitic resistances, and large interconnect delays, etc.

Figure 5a illustrates the electric field lines in bulk Si devices. Classical bulk Si scaling approaches have dealt with this problem by requiring an increase of body doping (region I and II in Figure 5a), thereby decreasing the depletion widths and electric field propagations associated with the source and drain, so that these two junctions are kept separated to the larger extent possible. However, increasing the body doping is accompanied by severe drawbacks such as degraded mobility, increased capacitances, and increased statistical fluctuations, all of which pose serious challenges to scaling. Other scaling challenges include those such as high gate currents, high parasitic resistances and capacitances, large interconnect delays, etc.

Figure 5b illustrates the electric field in SOI MOSFETs device. In SOI, most of the electric field propagates through the BOX before reaching the channel region. Short channel effects in SOI devices can be controlled in different way as derived from (2):\(^{36,37}\)

\[
\lambda = \sqrt{\frac{e_{\text{Si}}}{e_{\text{ox}}} T_{\text{Si}} T_{\text{ox}}},
\]

where \( \lambda \) is the natural length which represents the spread of the electric potential in the x-direction along the source and drain. It shows dependence on the gate oxide and silicon oxide thicknesses. The thinner the gate oxide and/or the silicon film, the smaller the natural length, and hence it allows to reduce the influence of the lateral electric field in the channel region. Unlike the approach used in bulk Si to reduce SCE, the increase of channel doping in SOI devices is less efficient compared with reducing the natural length directly.\(^{36}\) Incorporate SOI devices with ground plane and thin buried oxide allows further improvement since most of the electric fields from the source and drain terminate on the ground plane instead of channel region.

### 3. FD-SOI MOSFETs

#### 3.1. MASTAR Equations

MASTAR stands for Model for Assessment of CMOS Technologies And Roadmaps.\(^{38}\) MASTAR equations are derived based on empirical data and have been established and confirmed throughout a number of past CMOS generations.\(^{39-42}\) Moreover, this model has been implemented in the predicting tools used by ITRS\(^{2}\) for Process Integration, Devices, and Structures to calculate the impact of transistor scaling on electrical characteristics.

For bulk,

\[
\text{EI} = \left(1 + \frac{X_{\text{j}}^2}{L_{\text{el}}^2}\right) \left(\frac{T_{\text{oxel}}}{L_{\text{el}}} \frac{T_{\text{dep}}}{L_{\text{el}}}\right)
\]

(3)

\[
\text{EIS} = \left(\frac{T_{\text{oxel}}}{L_{\text{el}}}, X_{\text{j}}\right) \left(1 + \frac{3}{4} \frac{T_{\text{dep}}}{L_{\text{el}}}\right)
\]

(4)

\[
\text{SCE} = 0.64 \left(\frac{T_{\text{Si}}}{T_{\text{ox}}}\right) . \text{EI} . \Phi_{\text{d}}
\]

(5)

\[
\text{DIBL} = 0.80 \frac{e_{\text{Si}}}{e_{\text{ox}}} (\text{EI}) V_{\text{ds}}
\]

(6)

where \( X_{\text{j}} \) is junction depth, \( L_{\text{el}} \) is electrical gate length, \( T_{\text{oxel}} \) is electrical oxide thickness, \( T_{\text{dep}} \) is depletion thickness, \( T_{\text{Si}} \) is silicon thickness, \( T_{\text{ox}} \) is oxide thickness, \( V_{\text{ds}} \) is drain potential, \( C_{\text{dep}} \) is depletion capacitance, and \( C_{\text{ox}} \) is oxide capacitance.

Electrostatic Integrity (EI) in (3) has relation to SCE and DIBL, while EIS (Electrostatic Integrity specifically for subthreshold slope) in (4) is related to subthreshold slope. EI and EIS equations will be used to explain the behaviors of FD-SOI.
devices as the gate length is scaled and for better understanding, the comparison is made with planar bulk Si devices. 

The ratios of $T_{ox}/L_{el}$, $X_j/L_{el}$, and $T_{dep}/L_{el}$ that appear in EI and EIS are called “good technology rules” and these parameters are depicted in Figure 5. For (a) bulk Si and (b) SOI MOSFETs. These rules describe the quality of the transistor in the subthreshold regime by ensuring good electrostatic integrity of the device. In other words, keeping these ratios constant from one generation to another preserves the value of SCE, DIBL and S throughout the scaling devices. Typically, the three ratios used to reach 1/27, 1/2, and 1/2, respectively.39

In the first parameter, to avoid high off-state current and poly-depletion effect, the scaling of $T_{ox}/L_{el}$ ratio needs to be sustained. With scaling of SiO$_2$ approaching the limit, high-$\kappa$ materials are required. For bulk CMOS, high-$\kappa$ materials co-integration with dual metal schemes significantly increase the process complexity compared to a poly-silicon-gate technology.39 The big advantage of polysilicon gates is the modulation of its work function by implantation. This property is not obtainable with metallic gates. Implementation of single-gate (mid-gap gate function, e.g., TiN and W) metals is inadequate for advanced bulk silicon as follows.42,43

- Resulting threshold voltages are too large for low-voltage operation. They are increased or decreased by half of the silicon gap for NMOS and PMOS, respectively.
- Short channel characteristics are severely degraded due to low substrate doping density.

However, there are several approaches to overcome the problem with co-integration of dual metallic gates into the CMOS gate stack,39,44 i.e., traditional etching,42 Total Sili-cidation (TOSI) or Fully Sili-cidation (FUSI),42,43 and gate-last damascene,42,43 with the latter method being most preferable.44 While in FD-SOI ultra-thin body devices, the implementation of a single gate (mid-gap) is straightforward39,42,45,46 since the difference between NMOS and PMOS gate work function is relatively small for undoped channel. Undoped channel is the channel that does not require additional intrinsic carriers, i.e., the channel doping can be kept to low about $\sim 10^{15}$ cm$^{-3}$, which is impossible for bulk Si due to SCE. In addition, UTB offers lower threshold voltage than their bulk counterparts thanks to negligible depletion charge in the channel, without increasing the off current.39

The second parameter which controls the scaling of the device is a reduction in junction depth following the reduction of gate length, i.e., $X_j/L_{el}$. This term faces difficulties in bulk Si due to uncertain technological feasibility,47 but also the trade-off between depth and resistivity is an issue. If the junction depth is shallow and smaller than the inversion layer, the source, drain, and channel regions will be partially disconnected, or the resistance between the regions will be very high, resulting in large degradation of the drive current.3,48 Another foreseen issue is that it requires tight control of manufacturability in terms of narrow doping profiles. Therefore, UTB can offer great advantage with respect to the shallow-junction problem since the junction depth is then limited to the thickness of the channel. Thanks to that, the depth of the junction is also decoupled from its doping level, thus diminishing the junction resistivity problem.39,40

The third parameter is the depletion depth ($T_{dep}$), similar to junction depth; the depletion depth can also be limited by geometry in UTB devices. In bulk devices, the limitation in the channel hinders the $T_{dep}$ scaling (regions I and II). Several limitations prevent high channel doping i.e. significant junction leakage, increased statistical fluctuation due to random dopant fluctuation in the channel, strong degradation in mobility due to impurity scattering and an increased transverse electric field. In extreme case, the mobility degradation may dominate the beneficial effect of channel shortening. This eventually leads to a current decrease instead of increase in shorter devices.39,41,42,49 In UTB devices, the effective depletion depth is geometrically limited to the silicon, BOX thickness, and sometimes small effect of substrate depletion depth ($T_{Si}$).

Based on those merits, one can see that UTB FD-SOI is superior with respect to EI. The main key is that, both the $X_j$ and $T_{dep}$ are no longer results of doping and diffusion. Instead, they are set by the thickness of the silicon film and therefore, can be much smaller than $X_j$ and $T_{dep}$ in bulk devices. Moreover, based on those explanations, UTB devices have potential to extend MOSFET scalability. However, when discussing about scalability of FD-SOI with thin body, it is more appropriate to include thin buried oxide as well (UTBB). This is due to the fact that only thinning silicon body is not sufficient for very good scalability of SOI MOSFETs but also the thickness of underlying BOX. This can be explained with significant reduction of electrostatic field penetration, emanating laterally from the drain into the BOX towards the source and channel inversion39,50–52 and thus degrading the sub-threshold characteristics. For that reason, when hypothesizing about scalability in FD-SOI MOSFETs, it is more appropriate to focus on UTBB devices. However, thin BOX triggers stronger coupling between the gate and substrate, and if substrate is lightly doped (particularly underneath the BOX/substrate interface), it will be depleted by the gate and drain fields. This depleted layer will then behave as dielectric and effectively add to the physical thickness of the BOX53 and thus loses it advantages. Ground plane implementation can be used to suppress this unwanted effect.

As a final point, one can see, for scalability of UTB and UTBB SOI MOSFETs, the terms $X_j$ and $T_{dep}$ in (3) and (4) can be replaced with $T_{Si}$ and $T_{Si} + \gamma T_{BOX}$,38,39 respectively with $\gamma$ to take into account the substrate depletion depth.
3.2. UTBB Electrostatic Features

As previously discussed, the scaling principles of bulk silicon MOSFETs device require a reduction of junction depth and an increase of doping level, which adversely affect the junction capacitance and carrier mobility. Similar difficulty arises for PD-SOI with regards to channel and junction doping since PD-SOI characteristics are almost similar to bulk silicon (due to thicker silicon body). Fortunately, FD-SOI has emerged as promising technology as those two bottleneck parameters are diminished in FD-SOI. In addition, the scaling rules and design windows are more relaxed because additional tunable parameters (BOX thickness, channel free from doping, underlap channel, substrate doping, and backgate biasing) are available for device optimization. In this section, we will look into the electrostatic integrity features of FD-SOI MOSFETs (particularly UTBB) to extend the scalability of CMOS and to continue enjoying increased performances.

First, the basic concept of using ultra-thin silicon body is to reduce the SCE as explained by natural length expression (2) and MASTAR equation. Second, by reducing the silicon body thickness, the leakage current which flows along the bottom between source and drain, which is less effectively controlled by the front gate, can be reduced.\(^{50-56}\) This allows the channel to be undoped. With undoped channel, higher carrier mobility can be reached that further improves the drive current, thanks to lower transverse electrostatic field and negligible impurity scattering. It also features negligible depletion charge and low junction capacitances which yield steep sub-threshold slope.\(^{49}\) Moreover, significantly improved variability control can be achieved due to elimination of random dopant fluctuation.\(^{51,57-62}\)

However, thin silicon body induces high parasitic source and drain resistances. There are several approaches reported in the literature to overcome this effect. One is to use thicker raised source and drain (RSD) to minimize the parasitic series resistances and achieve high drain current.\(^{49,51}\) However, there is a trade-off with this approach as one can gain benefit from minimizing the effect of parasitic resistances but degrade the circuit performances due to increased parasitic capacitances between the gate and source and drain electrodes, i.e., fringing capacitances. Therefore, in this approach the spacer width between gate and RSD should be optimized. The second approach to reduce the parasitic resistances is to use metallic or siliconized source and drain,\(^{51,54,55}\) with very low Schottky barrier in order to not degrade the drive current. A last method is known as faceted raised source and drain (faceted RSD). The faceted RSD do not only reduce the series resistance but also minimize the parasitic capacitances.\(^{63-66}\) The method has been implemented in SRAM cell with reduction delay time, thus affirming the benefit of capacitance reduction.\(^{67}\)

Second, in FD-SOI, the channel can be kept undoped due to above reason, such that the implementation of underlap is more feasible in FD-SOI\(^{68}\) compared to bulk.\(^{69}\) With underlap, the effective channel length increases thus decreasing the drive current due to an increase of series source and drain resistances. Moreover, the off-state current is reduced due to the suppression of DIBL and subthreshold slope is improved. Hence, there is a trade-off between improved short channel behavior and reduced drive current in devices with longer extension lengths.\(^{68,70,71}\) Nevertheless, underlap is reported to improve the analog and RF figures-of-merit at low current level thanks to minimal impact of series resistance at this current level and reduction of fringing capacitances,\(^{72,73}\) thanks to longer effective channel length.

Third, as previously mentioned in natural length expression (2) and MASTAR equation, thin BOX further improves the SCE. Employment of thin BOX of 50 nm or below\(^{53,74}\) allows suppression of fringing electric fields through the BOX thus improving front-gate-to-channel controllability and reducing DIBL. Despite degraded subthreshold slope in long devices with thin BOX due to capacitance increases induced by the use of thin BOX, this degradation becomes smaller in shorter devices,\(^{75-77}\) thanks to better electrostatic coupling and control of SCE. This is translated into thinner depletion thickness in shorter compared to long channel.\(^{78}\) In addition, thin BOX is also suitable for implementing back-gate biasing schemes used for tuning device characteristics.\(^{79-81}\) However, the drawback is that it enhances the channel coupling through substrate and if the substrate is depleted, the thin BOX does not gain its advantages. Ground plane implementation with heavily doped substrate\(^{53,82,83}\) or localized doping,\(^{84-89}\) have been discussed to reduce the depletion effects in the substrate. The methods in Makiyama et al.\(^{87}\) and Yan et al.\(^{88,89}\) claimed through numerical simulation, able to provide better control of short channel effects and thus better DIBL. Eventually, incorporating the GP substrate offers additional attractive package for device scaling solution. It provides as well an easy way to modulate the threshold voltage with Fenouillet-Beranger et al.\(^{77,90}\) and Thomas et al.\(^{86}\) combining back-gate biasing, constant or connected to drain bias.\(^{84-86}\) In such ways, multi-\(V_{\text{th}}\) option can be achieved based on single metal gate, without complexity in process and channel doping adjustment. With back-gate biasing, the modulation of threshold voltage can even be achieved for device with standard resistivity substrate.\(^{77}\) Another interesting method to achieve multi-\(V_{\text{th}}\) option is using gate materials\(^{58,91,92}\) with different work-functions. In this approach, it requires integration of dual-metal gates scheme and two different ground planes doping (p- or n-GP). Four differences \(V_{\text{th}}\) can be achieved ranging from 0.32 V up to 0.6 V for both n-MOS and p-MOS devices.\(^{91}\)

Finally, exploiting the physical nature of SOI, where the silicon body is located between the gate and buried oxides, basically allows for emulating a double-gate mode (so-called quasi-double gate, QDG, with \((V_{\text{sub}} - V_{\text{Th2}}) = T_{\text{BOX}} / T_{\text{gox}} (V_{G} - V_{\text{Th1}}).\) The technique was originally proposed by Balestra et al.\(^{93}\) in 1987 for 200 nm and 380 nm of silicon body and BOX, respectively. It is more attractive in the
case of thin body Si where the front and back channels are no longer independent, and as a result of volume inversion concept, controlling the channels from both sides at the same time forces most of the carriers to flow through the middle of the film. The electrons are then localized further away from the oxide interfaces and hence suffer less surface roughness scattering which later significantly increases the carrier mobility compared to single gate device with the same parameters. Significant improvements of subthreshold slope below the theoretical limit of 60 mV/dec have been reported experimentally. In addition, it is shown that the transconductance in QDG mode exceeds twice  or four times the value observed in SG-mode depending on the back-gate biasing even to enhance mobility. However, full QDG mode realizations in such devices may require an over-voltage to be applied to the back gate, which is very complex for CMOS circuits. Thus, in UTBB, as a result of the ultra-thin BOX and the availability of GP, a similar approach can be realized by simply connecting the substrate (or GP contact) to the front gate, i.e., $V_{\text{Sub}} = V_g$. Such a connection is known as asymmetric double-gate (ADG) regime and is expected to give a weaker improvement than a pure QDG, but is more practical and easier to employ in circuits. Such operation applied to UTBB devices without a GP was shown to enhance drive current ($I_{\text{on}}$) and with GP, improve $I_{\text{on}}$ and lower DIBL as well (see Figure 6).

### 3.3. Figures-of-Merit

Figure 7 shows the DIBL comparison between bulk, FD-SOI (referring to UTBB) and FinFET devices. Due to better electrostatic integrity, FD-SOI devices are known to be suitable to extend the scalability of MOSFETs. Only FinFET with very aggressive scaling fin (13 nm for $L = 25$ nm) can reduce DIBL further but FinFET poses manufacturing challenges due to 3D-structure devices.

Figure 8 reports the benchmarking of the variability between planar FD-SOI and bulk silicon. Thanks to the use of an undoped channel, which suppresses the random dopant fluctuations, the advantages of FD-SOI are clearly seen with excellent control of threshold voltage variability with lower matching factor ($A_{\text{vth}}$) reached for both planar FD-SOI or SOI-FinFETs.

The device scaling for RF is more relaxed, since it is not aggressive as in digital devices. Most available information in literature is dominated by either PD-SOI or bulk Si. Although the scaling is less crucial, nonetheless the benefit of scaling gained in digital can be translated into benefit of RF FoM with condition that the parasitic resistances (gate, source, and drain) and parasitic capacitances are not amplified through fabrication process or architectural technologies. For PD-SOI (mostly reported by IBM), among the highest $f_T$ and $f_{\text{max}}$ ever reported
are 330 and 450 GHz, respectively, for L = 27 nm. The device features Ni silicidation to minimize gate resistances, stress channel engineering for performances enhancement, and optimized layout to minimize the parasitic capacitances. While for bulk Si (mostly reported by Intel), the highest fT and fmax are 330 and 420 GHz, respectively, for L = 29 nm. The device employing uni-axial strained silicon, Ni silicide, and excellent control of junction engineering to manage the parasitic resistances. Recently reasonably high fT of 300 GHz was reported for FD-SOI (UTB with L = 25 nm, TSi = 6 nm, and TBOX = 145 nm, by IBM) devices with faceted RSD and undoped channel. As previously mentioned, the faceted RSD does not only reduce the series resistance but also minimizes the parasitic capacitances. In addition, the faceted RSD increases the strain coupling from stress liners to the channel and hence the silicide contact area.

Figure 9 and 10 show the benchmarking of g_m and fT and, respectively, for UTBB with other state-of-the art technologies, i.e., PD-SOI, FinFET, Bulk, UTB, or BULK and FD-SOI (ETSOI and UTBB).

Here, one can see the g_m and fT of state-of-the art technologies can be grouped differently i.e., with and without strain/stress technologies. The devices with strain/stress technologies result in highest g_m which translated into highest fT. The performance of UTBB device is comparable with FinFET, PD-SOI, UTB and BULK for similar gate length. As the gate length scaled down to 30 nm, extrinsic fT as high as ~160 GHz and extrinsic g_m of ~1000 mS/mm can be achieved. First, the increase of fT is attenuated in comparison to expected increase at shorter gate length. This is due to the transistor suffers from the relative increase of the parasitic parameters outside the transistor channel such as the series resistances, overlap and fringing capacitances which do not properly scale down with gate length. In addition to that, for short channel length, the carriers can reach the velocity saturation at high drain voltage and thus g_m is no longer proportionate to 1/L. Velocity saturation is the phenomena where the carriers reach their peak velocities under the influence of the lateral drain fields. As a result fT improvement is less significant in short channel compared to long channel devices. Nevertheless, one still see fT is higher in shorter devices compared to long one due to improvement in g_m and lower parasitic capacitances. Second, the value of extrinsic fT is still below ITRS roadmap. Third, fT and g_m have direct correlation, projected in Figures 9 and 10 where higher fT is achieved as a result of higher g_m. (i.e. for stress/strain engineering devices). Finally, ETSOI devices (also with strained silicon engineering) present fT almost at the same level compared to other state-of-the arts strain/stress technologies.

In Md Arshad et al., the estimation of improvement gained in UTBB devices with other state-of-the art technologies have been made based on intrinsic g_m and fT. One can see that intrinsic g_m is significantly high, with a value, which is comparable to extrinsic g_m obtained from stress/strain engineering devices. However, such improvement is seemed to be not directly translated into intrinsic fT (w/o Rsd) (Figure 10). This gives indication that, the parasitic capacitance is still high in UTBB which affects the intrinsic fT (w/o Rsd). Next, by removing C_gge one can see that the maximum improvement can be obtained in UTBB devices as demonstrated by intrinsic fT (w/o Rsd and C_gge), shown in Figure. 10. This, however, provides an overestimated value of fT. Next, they consider C_gge = 300 Ω.μm and C_gge = 0.7 fF/μm as in ITRS requirement for FD-SOI technology. The result show that the g_m (w/ Rsd ITRS) and fT (w/ Rsd and C_gge ITRS) as high as ~1350 mS/mm (Figure 9) and ~310 GHz (Figure 10) can be achieved, respectively. Such values are comparable to strained silicon technology and ETSOI (also with strained silicon technology).
4. CONCLUSION

In this article, we start with comparison between bulk and silicon-on-insulator, then we highlight the main parameter in thin body SOI controlling the short channel effects. After that, we present the digital and RF figures of merit in thin body SOI MOSFETs. The main advantages of thin body MOSFETs are the back-gate biasing either from the substrate or asymmetrical double gate. These configurations are not achievable in other technologies such as FinFET. Although the thin SOI device is better since its drawbacks are largely overshadowed by its significant advantages, the choice is still surprising since SOI starting material is more expensive than bulk Si as it requires extra processing steps to manufacture. This poses an obstacle to the selection by industry although SOI brings more benefits. However, as the device gate length shrinks down, it is believed the thin body SOI device offers better choice in the future.

FUNDING

The authors are grateful to the Department of Higher Education, Ministry of Higher Education, (KPT) for funding this research through the Fundamental Research Grant Scheme (FRGS) with the grant number 9003-00380.

REFERENCES

5. K. J. Kuhn, Moore’ s law past 32nm: Future challenges in device scaling, in Inter. Workshop on Computational Electronics 1–6 (2009).


103. B.-Y. Nguyen and C. Mazur.


