How Many Gates Do We Need in A Transistor: One, Two, Three or Four?

S. CRISTOLOVEANU

Institute of Microelectronics, Electromagnetism and Photonics
(UMR 5531, CNRS-INPG-UJF)
INP Grenoble – Minatec, 3 Parvis Louis Néel, BP 257,
38016 Grenoble Cedex 1, France
E-mail: sorin@enserg.fr

Abstract. The migration from single-gate to multiple-gate SOI transistors is inexorable but still includes a number of hesitations and dilemmas. In this context, we review the main motivations, technological achievements, physics-related implications and future challenges. Specific aspects governing the operation of multiple-gate transistors are discussed with the aim of clarifying the strategic trends. The physics mechanisms are briefly evoked. The road is difficult with many turns; hopefully, it looks long enough. Generic milestones are: nano-size, multiple-gates, innovative ideas and multi-facet SOI technology.

1. Context

Gradually increasing the channel doping was a very safe, yet extremely successful strategy so far for solving CMOS scaling problems. The summit of this sophisticated art consists in decorating the transistor body with heavy concentrations of dopant (halos, pockets, etc.), engineering very thin source-drain regions with various compositions, and adding strain [1].

Then SOI came into the picture [2–7]. Not only because the transistor body is isolated by the buried oxide (BOX), but mainly because the BOX being available it can sustain a silicon film with adjustable thickness. Thinning down the silicon film appears to be an easier and more attractive solution for scaling than increasing the body doping beyond unreasonable levels [1–14]. In ultra-thin SOI MOSFETs (~ 10 nm), the doping is irrelevant and can be safely omitted [10–11].

The plain control of the transistor electrostatics is the guarantee for acceptable short-channel effects. In SOI, it requires a film thickness just one fourth of the channel...
length. Instead of having to place and discipline one dopant atom every 10 silicon atoms, we may need a Si film thin enough to contain 10 atoms or less. Of course, this is another sort of challenge.

In order to relax the film thickness constraint, it is necessary to reinforce the electrostatic control. The obvious solution is to border the transistor body with additional gates (Fig. 1) [1–10]. For a transistor featuring $N_G$ gates and a channel length $L$, the smallest dimension of the body (i.e., film thickness for planar MOSFETs or fin width for FinFETs and vertical MOSFETs) is relaxed from $L/4$ to $N_G L/4$. For example, a double-gate (DG) device affords to be twice as thick as a single-gate SOI transistor. This is extremely important and more realistic in terms of current technology: it also alleviates the pressure on SOI material engineers. A 3–5 nm thick Si film with 5–10% uniformity on large wafers (30 cm diameter) is a nightmare.

![Fig. 1. Schematic configurations of multiple-gate SOI transistors: (a) double-gate mode emulated in a FD MOSFET, (b) planar double-gate MOSFET, (c) Gate-All-Around MOSFET, (d) Four-Gate transistor, (e) FinFET variants including the original DG FinFET, the MuGFET, the Triple-Gate FET, the Pi-FET, and the Omega-FET.](image)

As a bonus, the drain current and trans-conductance being increased by roughly a factor of two, the circuit designers are delighted. They can even play with the two gates, by keeping them interconnected or biasing them independently. The DG
FinFET with two independent gates (like MuGFET, Fig. 1e) is a 4 terminal device where the two separate gates can be taken advantage of. The result is beneficial for new devices and reconfigurable circuits with enhanced functionality. Expressions like ‘paradigm shift’ or ‘revolutionary’ design summarize a quantitative approach (reduced transistor count for a given circuit or logic function) or a qualitative breakthrough (novel types of memories, logic functions, etc).

Triple-Gate FinFETs [15–21] and Gate-All-Around (GAA) MOSFETs [3] show even better electrostatic behavior while requiring less dramatic thinning: the body size and channel length become comparable. However, these devices are not really multiple-gate transistors because one single gate covers most of the exposed sections of the body (Fig. 1e). They feature multiple interfaces and corresponding channels rather than multiple gates.

A genuine 4-gate transistor has been demonstrated [22]: it has two MOS gates and two junction (JFET) gates, all of them independent of each other (Fig. 1d). It is difficult to imagine a transistor with more than 4 gates, unless the channel is partitioned in sub-channels with individual gates. Such a partition excludes applications for short channels which was our starting reflection point.

In the next sections, a number of key issues faced by these various SOI approaches will be discussed. Recent and important technological achievements include strained Si, SiGe, Ge, dual crystal orientations, etc. These materials are not competing with SOI because they must be ‘On Insulator’. Otherwise, the electrostatic behavior of the transistor, which is qualitatively independent on the material nature, will erase all benefits.

Electron and hole mobility engineering can also be achieved by reducing device self-heating [23]; this implies replacing the SiO$_2$ buried oxide (BOX) by a different dielectric with improved thermal conductivity [24–26].

From a fundamental point of view, these various approaches are just ‘embellishments’ of SOI. Changing the meaning of SOI, from silicon-on-insulator to semiconductor–on-insulator, clarifies the ambiguity and provides a more general view.

### 2. Zero-Gate transistor

The Zero-Gate transistor, actually called Pseudo-MOSFET or Ψ-MOSFET, serves for SOI material characterization. Although there is no apparent top gate like in a normal MOS transistor, the Si substrate underneath the BOX can be biased. A conduction channel (inversion or accumulation) is formed at the film-BOX interface. The BOX plays the role of a gate oxide and the Si film represents the transistor body. Low-pressure probes are placed on the film to form source and drain point contacts [27, 28].

The current-voltage characteristics contain rich information about the material quality. Standard MOS methods allow extracting key parameters (defect density, carrier lifetime and mobility of both electrons and holes, etc) for SOI wafer process monitoring and optimization. In ultra-thin structures, the quality of the top free surface and the properties of whatever is deposited on it can also be sensed.
The pseudo-MOSFET still works if the pressure probes are replaced by mercury probes or by evaporated circular contacts (for magnetoresistance measurements in Corbino geometry) [29]. A more astonishing variant is the current-less transistor, where the current is replaced by an optical signal (second harmonic generation, SHG) still function of the substrate bias [30].

Recent interest is focused on ultra-thin structures with film and BOX in the sub-10 nm range, strained SOI, Ge on insulator, flexible substrates, alternative dielectrics, etc.

3. Single-Gate transistors

Conventional SOI transistors are categorized as partially depleted (PD) or fully depleted (FD). PD circuits are currently making the success of the SOI microprocessor industry, whereas FD devices have far more potential for future scaling. The distinction between PD and FD becomes a difficult exercise in short-channel transistors which tend to mix operation mechanisms borrowed from both families [31–33].

The major short-channel effect in SOI is due to the penetration of the electric field from the drain into the BOX and substrate (inset of Fig. 2). The fringing field rises the surface potential at the film-BOX interface: drain-induced virtual substrate biasing (DIVSB) [10–11]. Since the front and back interfaces are naturally coupled, the front-channel threshold voltage is lowered with increasing drain bias.

![Fig. 2. Threshold voltage lowering $\Delta V_T$ induced by DIBL and DIVSB effects versus film thickness and channel doping ($V_D = 1.2$ V, $L = 100$ nm). The doping effect is canceled for SOI films thinner than 15 nm (adapted from [5, 10]).](image)

Figure 2 illustrates the scaling strategy in SOI, based on film thinning. A difference in body doping of 3 orders of magnitude does not have any effect on the threshold voltage lowering if the film thickness is correctly scaled down [11]. Additional recipes include an extremely thin BOX [34] and a ground plane underneath [10].
In sub-10 nm thick films, interesting effects take place which refresh the transistor physics. Vertical quantum confinement [35] and sub-band splitting become significant, affecting the threshold voltage and the carrier mobility. The coupling of the channel to the BOX defects and substrate biasing is amplified [36–42]. The series resistance increases drastically and may offset the excellent transport properties in the transistor body.

Additional dimensional effects occur as a result of the simultaneous shrinking of the other dimensions: narrow channel effects [43–45], gate-induced floating-body effects due to tunnelling through nano-meter thick gate dielectrics [46-48], etc. Strain is also size dependent, either if it is embedded in the original SOI wafer or generated by the CMOS process.

4. Double-Gate transistors

An ideal DG MOSFET is viewed as a perfectly symmetrical device (Fig. 1b). The two channels facing each-other are activated simultaneously and feature identical charge and mobility. The subthreshold slope matches the minimum theoretical value \((2.3 \frac{kT}{q})\). As the body thickness decreases, the centroid of the inversion charge moves from the interfaces into the volume. This effect is called volume inversion [49]. Combined with a zero electric field in the middle of the body, volume inversion enables the mobility enhancement [10, 50–54]. Monte-Carlo simulations predict a maximum value of the electron mobility in 3–5 nm thick films as a result of competing effects: phonon confinement, subband splitting which lowers the effective mass, and reduced inter-band scattering [50, 53].

‘Genetically’ modified DG MOSFETs may have dual gate stacks, where the difference in work functions is useful for threshold voltage control. They also may feature a longer back gate, partially over-lapping the source and drain. Such a non-ideal DG transistor may outperform the ‘ideal’ structure because the bottom gate tends to drive the source and drain regions into strong accumulation, so that the series resistance is dynamically lowered [55–56].

The problem is that nobody knows how to fabricate realistic planar DG MOSFETs. Many processing ideas have been tested. They were successful in generating papers, but not enough convincing for mass production.

It is interesting to note that a standard planar FD MOSFET stands as the simpler form of asymmetric DG transistor. Indeed, the Si substrate underneath the BOX can be used as a second gate. The trend in FD devices is to use a very thin BOX, with a ground plane underneath [10], which renders the back gate action very efficient. For balanced DG operation, the inversion charges in the front and back channels must be equated. This is achieved by selecting the threshold voltages and gate biases such as to account for the device asymmetry [57].

On the other hand, when one channel is inverted and the opposite channel is forced into accumulation, interesting effects applicable to innovative capacitor-less DRAMs are observed [58]. The basic idea is that the supply of majority carriers (by
impact ionization, band-to-band tunneling, etc) is a relatively slow process, leading to hysteresis and related memory effects.

However, in films thinner than 5–10 nm, it becomes difficult to maintain one channel in inversion while driving the opposite channel into accumulation. The front and back channels actually tend together to reach inversion or accumulation, depending on the dominating gate voltage. The interface coupling effects are exacerbated leading to super-coupling. The film behaves as a rigid quasi-rectangular well: if the front gate voltage is swept, the potential of the entire film follows [42]. In this case, the notion of front and back channels becomes obsolete. An important consequence is that the carrier mobility must be viewed as an in-depth integral.

FinFETs are pragmatic DG devices, easy to fabricate, more difficult to optimize [15, 16]. Since the gates are on the lateral sides, the edges of the transistor need to be smooth and the inter-gate distance should be in the 10 nm range. The basic mechanisms in planar DG and FinFETs are more or less the same, simply transformed by a rotation from horizontal to vertical direction. Separating the two gates (MugFET) is very useful for signal mixing and threshold voltage adjustment. But, if the two gates are meant to stay interconnected, a more efficient solution is to fabricate Triple-Gate FinFETs.

5. Triple-Gates and more

A FinFET with an active top gate is named Triple-Gate MOSFET. One single gate controls three different sections of the channel: two vertical and one horizontal (Fig. 1e). The discrimination of the different channels is possible by using the backgate action and the variable fin width [59]. The advantages of Triple-Gate FinFETs in terms of integration density and performance are becoming well understood. Several aspects have attracted recent attention:

- In a very tiny transistor, all conductive regions collapse in a unique central channel. This is another form of volume inversion.
- The ‘corner effect’ is due to the coupling between the adjacent gates which induces a local lowering of the threshold voltage [60]. This parasitic channel may cause an increase of the off-state leakage current. The corner effect is negligible in undoped transistors whereas, in heavily doped bodies, it can be suppressed by corner rounding.
- The ‘local’ threshold voltage varies along the vertical channel, from the BOX to the top gate. This non-uniformity increases with the doping level but remains lower than 25 mV [17].
- The top and lateral channels have a different crystal orientation and interface quality. The carrier mobility can be lower on the sidewalls. The challenge is to achieve perfectly vertical and defect-free sidewalls.
- In a very small body with sub-10 nm thickness, width and length, the number of Si atoms is $10^4$ or even less. What is the minimum silicon cluster size below
which the concepts of effective mass and energy bands should be revisited?

What are the meaning and impact of one single impurity in the body?

The optimization of Triple-Gate FinFETs is a matter of trade-off between drive current (wide top channel) and electrostatic control (narrow fin). The coupling effects are truly 3D, with three components [17]: (i) lateral coupling between the side gates, (ii) vertical coupling between the top gate and the bottom gate (substrate, Fig. 3), and (iii) longitudinal coupling between the drain and the body via the fringing fields penetrating in the BOX and substrate.

![Fig. 3. Threshold voltage as a function of substrate bias in triple-gate FinFETs with wide, square, and tall fin configurations (aspect ratios in nm: \(\text{t}_{\text{si}}/W = 20/80, 20/20, 80/20\)).

The substrate effect disappears in transistors that are tall and very narrow, leading to radiation-hard capability and reduced short-channel effects (after Ritzenthaller et al. [17]).

Figure 3 shows that wide fins behave as FD MOSFETs with strong vertical coupling and little impact of the lateral gates. In tall and narrow fins, the electrostatics is governed by the coupling of the lateral gates, which inherently tends to reduce or suppress the vertical coupling to the bottom gate. The surface potential at the fin-BOX interface can be totally controlled by the fringing field between the lateral gates [17].

As compared to the pure Triple-Gate FinFET (where \(N_G = 3\), Fig. 1e), the control of back-surface potential is reinforced by letting the lateral gates to extend vertically into the BOX (\(\pi\)-gate where \(N_G \approx 3^+\)) and laterally underneath the film (\(\Omega\)-gate where \(N_G \approx 4^-\), Fig. 1e). As a result, the defects generated in the BOX, by processing, radiations or hot carrier injection, have no impact on the device performance [61]. In addition, the longitudinal drain-to-body coupling is screened, reducing the short-channel effects [17]. Narrow Triple-Gate FinFETs are expected to have a considerable impact on the future market place because they are intrinsically radiation-hard [61] and tolerant to short-channel effects.

The Gate-All-Around (GAA) MOSFET features 4 channel sections (or apparent gates, \(N_G = 4\)), unless the body is cylindrical. A GAA combines channel coupling...
and quantum effects [3]. It is the ideal device in terms of electrostatic control as it perfects the qualities of Ω–FETs. Recent achievements include GAA nanowires which can even be superposed for more efficient integration.

The on/off current levels, transconductance and short-channel effects are very attractive in GAA MOSFETs and Triple-Gate FinFETs. Still these elegant and performing devices are conceptually poorer than a DG MOSFET with two independent gates. In this sense, their flexibility is even lower compared to a planar fully depleted single-gate (SG) transistor, where a back gate can be activated by biasing the substrate (Fig. 1a). It is a lot easier to apply two different signals to a SG MOSFET, by promoting the substrate as a back gate, than to apply 3 simultaneous signals to a unique gate even if this gate is named Triple-Gate. Ironically, in GAA and well optimized Triple-Gate MOSFETs, the back-gate effect is no longer an available option because the lateral gates control the potential everywhere and screen the substrate effect.

6. Four-Gate transistor

The G\textsuperscript{4}–FET is derived from a basic SOI MOSFET provided with two lateral body contacts [22]. The current is driven by majority carriers flowing between the body contacts, which are promoted as source and drain (Fig. 1d). There are 4 independent gates: the usual front and back MOS gates govern the surface accumulation or vertical depletion regions, whereas the two lateral junctions control the effective width of the body through the extension of the horizontal depletion regions.

The conduction path is modulated by mixed MOS–JFET effects: from wire-like volume conduction to a strongly accumulated front and/or back interface channels. Analytical models are available for each mode [62].

The volume conduction mode is very efficient to suppress trapping noise and radiation damage [63, 64]. This is because the channel is double shielded from the interfaces by the depletion region and by the inversion layers. Changing the biasing mode made it possible to elucidate the origin of 1/f noise: mobility fluctuations in volume mode, where trapping is absent, and carrier number fluctuations via trapping in surface conduction mode [63].

On the practical side, the fact that each gate has the capability of switching the transistor on and off is an asset for novel applications. For example, the characteristics of an inverter can be translated or adjusted by appropriately biasing the lateral gates. A simple inverter can thus be reconfigured in a NOR, NAND or majority voting gate without any hardware. Other preliminary examples include mixed-signal circuits, compact logic functions, single-device memory, etc.

7. Conclusions

The nano-size multiple-gate SOI MOS transistor is the ideal candidate for the future landscape of the microelectronics, including the transition from micro- to nano-devices and systems. The MOSFET scaling is intrinsically easier in SOI than in bulk
Si, where it is becoming a desperate issue. The key condition is to utilize ultra-thin bodies and to open the SOI family to any kind of semiconductor, strained or not, on any type of dielectric.

Two, three or four gates can collaborate for enabling enhanced scaling, performance, functionality, and flexibility. Since the device operation is governed by 3-D effects, there are complex coupling mechanisms along the longitudinal, lateral and vertical directions. A given size effect (length, width, or thickness) is modulated by the other dimensions. All device architectures (planar or vertical double gates, \( \Omega \) or \( \pi \) topologies, 4-gates and gate-all-around transistors) have their merits and drawbacks. The winners will be selected with processing feasibility criteria.

An amazing horizon is opened to circuit designers looking in the direction of multiple addressable gates. The possibility to apply 2 or more independent signals concomitantly on various gates is the exciting feature for conceiving new memory effects, standard logic functions with economic transistor count or, even better, new functionalities.

Acknowledgements. I warmly thank all my colleagues who tried hard and eventually succeeded to keep my eyes open. The list is long, most of my friends are in.

References

How Many Gates Do We Need in A Transistor?


