Hindawi
Journal of Electrical and Computer Engineering
Volume 2019, Article ID 4792461, 18 pages
https://doi.org/10.1155/2019/4792461

Review Article

Technology and Modeling of Nonclassical Transistor Devices

George V. Angelov,[1] **Dimitar N. Nikolov**,[2] **and Marin H. Hristov**[1]

[1]*Department of Microelectronics, FETT, Technical University of Sofia, 1756 Sofia, Bulgaria*
[2]*Department of Electronics, FETT, Technical University of Sofia, 1756 Sofia, Bulgaria*

Correspondence should be addressed to George V. Angelov; gva@ecad.tu-sofia.bg

Received 23 June 2019; Accepted 8 October 2019; Published 3 November 2019

Academic Editor: Herve Morel

This paper presents a comprehensive outlook for the current technology status and the prospective upcoming advancements. VLSI scaling trends and technology advancements in the context of sub-10-nm technologies are reviewed as well as the associated device modeling approaches and compact models of transistor structures are considered. As technology goes into the nanometer regime, semiconductor devices are confronting numerous short-channel effects. Bulk CMOS technology is developing and innovating to overcome these constraints by introduction of (i) new technologies and new materials and (ii) new transistor architectures. Technology boosters such as high-*k*/metal-gate technologies, ultra-thin-body SOI, Ge-on-insulator (GOI), A^{III}–B^{V} semiconductors, and band-engineered transistor (SiGe or Strained Si-channel) with high-carrier-mobility channels are examined. Nonclassical device structures such as novel multiple-gate transistor structures including multiple-gate field-effect transistors, FD-SOI MOSFETs, CNTFETs, and SETs are examined as possible successors of conventional CMOS devices and FinFETs. Special attention is devoted to gate-all-around FETs and, respectively, nanowire and nanosheet FETs as forthcoming mainstream replacements of FinFET. In view of that, compact modeling of bulk CMOS transistors and multiple-gate transistors are considered as well as BSIM and PSP multiple-gate models, FD-SOI MOSFETs, CNTFET, and SET modeling are reviewed.

1. Introduction

The progress of VLSI CMOS technology continuously follows Moore's Law (1965). Currently, the dimensions of the conventional metal-oxide-semiconductor (MOS) transistors, respectively, the channel sizes, are scaling into deep nanoscale under 10 nanometers. The hi-tech industry has been foreseeing these developments with the notable International Technology Roadmap for Semiconductors (ITRS) [1]. This roadmap is crucial for the hi-technologies as it takes 10–15 years of research and development efforts for pioneering a new VLSI technology to commercialize and propagate in the market.

Technology advances manifest in a phenomenal increase in device density and performance, which serve as the basis for numerous data-intensive applications. The current technology drivers include autonomous driving, 5G communications, cognitive computing, artificial intelligence, cloud computing, the Internet of Things (IoT), blockchain, and cyber security.

On the other hand, Moore's Law technology developments are accompanied by lots of unwanted *short-channel effects* (SCEs) that are leading to serious technical hitches such as increased leakage currents, larger parameter variations, higher manufacturing costs, etc. [1]. In order to extend Moore's Law, certain strategies to circumvent these constraints need to be implemented [2]. Numerous innovations in CMOS technologies involving advanced materials and novel device structures have been proposed and implemented [3]. Advanced structures allowing more aggressive downscaling compared to classical bulk Si transistors such as ultra-thin-body (UTB) silicon-on-insulator (SOI) single-gate transistors and multiple-gate field-effect transistors (FETs) including the already implemented FinFETs, are developed as extensions to conventional MOSFETs. Aforesaid structures are referred to as *nonclassical devices*. There are more than a few nonclassical nanoelectronic devices that could complement and/or extend state-of the-art CMOS devices, e.g., graphene transistors (GFETs), nanowire or nanosheet transistors, etc. [4, 5].

Device modeling represents the cardinal tie of semiconductor technologies to integrated circuit design. Overall, compact device circuit models confront two contradictory requirements: they have to be both unsophisticated and highly accurate. In order to ensure the design quality, compact circuit models should be scalable to match the technology scaling while staying accurate across a broad range of operating and process conditions. Therefore, in this paper, we are reviewing modern compact device models and concepts used to predict the behavior of devices with nonclassic geometries that extend Moore's Law. For that purpose, the challenges and respective modeling approaches are highlighted and discussed.

2. Scaling and Limitations

2.1. Scaling Concept. Moore's Law represents the downscaling concept and the device functionality growth while reducing manufacturing cost per function [6, 7]. In 1965, it was an observational prediction on the rate of development of semiconductor technology by forecasting the amount of transistors that can be integrated into a microchip for the decade between 1965 and 1975. But the tendency remained just slightly changed over the following 5 decades. The fundamental key to this progress has been the push to smaller sizes by implementing the scaling rules initially formulated in the 1970s [8].

Scaling lets hundreds of millions to tens of billions of transistor devices to be fitted in a single silicon die. Device dimensions' downscaling is related to the following 4 aspects: density, performance, economy, and power. Device density (the quantity of transistors per unit area) growth and performance gain have been forecasted 50 years ago to double every year; nowadays, this progress has slowed down to doubling every three years. Economy (the cost per single transistor) has been initially predicted by Moore's Law to halve the costs every year, and it is still keeping this comparative pace in course of the last 50 years. Power consumption has increased its significance with the wide-spread of mobile devices.

The empirical observation made by Gordon Moore in 1965 was that the amount of devices in a semiconductor chip doubles by introducing each new technology generation (node). Each technology node, e.g., 0.18 μm, 0.13 μm, 90 m, 65 nm, 45 nm, 22 nm, 14 nm, 10 nm, 7 nm, and 5 nm (the number refers to the minimum channel length) has its distinct characteristics and each successive technology node reduces the sizes to 70% of the previous technology node—this is the technology *miniaturization* or *scaling*. In result, each new technology node gives twice bigger device density compared to the previous node or 50% smaller chip area (70% reduction of the prior linear size is giving ≈50% reduction in area: $0.7 \times 0.7 = 0.49$). For reference, a 7 nm chip contains c.a. 20 billion transistors per chip, while a 5 nm chip contains up to 30 billion transistors. The economic dimension of scaling is that the cost per circuit with each new node reduces considerably.

Classical *metal-oxide-semiconductor field-effect transistor* (MOSFET) is an unipolar 3-terminal device that has been effectively downscaled to 32 nm channel length [9]. MOSFET operates on the principle of modulating the electric field in the transistor's substrate under the gate, and in result, the flow of current between the source and drain electrodes is controlled by the gate voltage. MOSFET is intrinsically a 2D device having the current flow in a direction that is perpendicular to the gate field. As such, for decades, MOSFET has provided suitable platforms for applications such as processors, sensors, memory cells, solar cells, etc.

The concept of *scaling* implies that performance is improved by reducing the transistors sizes in the integrated circuits (gate length, L, and gate width, W, and the thickness, t_{ox}, of the oxide insulation layer between the gate and the substrate) along with reducing the supply voltage and increasing the impurity doping concentration so that the electric field in the device is kept constant (i.e., "*constant field scaling*"). Device scaling should go on in a balanced way between the lateral and transversal dimensions. If scaling is done by a factor of κ (κ—unitless scaling constant; $\kappa \approx 1.43 \approx 1/0.7$), in order to keep the same electric-field patterns as it was the original FET, the applied voltages should be reduced by κ and, at the same time, the impurity doping concentration as well as the charge densities should be also increased by the same factor of κ. Along with the decreased applied voltage, the volume of the depletion regions under the transistor electrodes (source, gate, and drain) shrinks as well. Generally, depletion regions should be set aside separated in order to control the transistor *on* and *off* state by the gate voltage [10].

Introduced by Dennard et al. in [8, 11], the *constant field scaling concept* is based on simple electrostatics. It leads to the following benefits: (i) the circuit density increases as κ^2, (ii) the circuit speed rises proportionally to κ because it is related to g_m/C (the capacitance of tinier devices is reduced by κ whereas the transconductance g_m remains the same), (iii) power dissipation decreases by κ^2, etc. [12] (see Table 1).

The constant field scaling concept has been valid until device dimensions went down into the submicron scale where the electric field in the channel starts to change significantly across the channel. The widely adopted *gradual channel approximation* (GCA), firstly used by Shockley [13], presumes that the electrical field below the gate of the MOSFET changes more gradually in the channel direction than in the gate direction (the fact that the electrical field in the MOSFET changes more gradually in the channel direction than in the gate direction means that $\partial Ey/\partial y << \partial Ex/\partial x$ (x—the gate direction, y—the channel direction; E—electrical field); this allows to simplify); it remains valid no matter how small the device sizes are with the only exception of saturation region where it fails [14, 15]. Besides, downscaling the threshold voltage together with the applied power-supply voltage lifts up the standby leakage current; the latter restricts the scale-down of the power-supply voltage.

A crucial part in obtaining high performance is the small *subthreshold slope* (or *subthreshold swing*) *SS*. The subthreshold slope depicts how sharply the transistor current turns *off* below the threshold. Another issue is the *saturation* of the carrier velocities because of the nonscaling nature of

Table 1: Classical device scaling.

Device/circuit parameter	Scaling factor constant electric field scaling	Scaling factor generalized scaling (E > 1)
Linear dimensions (L, W, t_{ox})	$1/\kappa$	$1/\kappa$
Electric field intensity	1	ε
Supply voltage (V, V_{DD})	$1/\kappa$	ε/κ
Current (I)	$1/\kappa$	ε/κ
Capacitance (C)	$1/\kappa$	$1/\kappa$
Inversion charge density (Q_I)	1	ε
Speed ($\sim g_m/C$)	κ	κ
Chip area (A)	$1/\kappa^2$	$1/\kappa^2$
Power dissipation ($P = VI$)	$1/\kappa^2$	ε^2/κ^2
Power density ($\sim P/A$)	1	ε^2
Doping concentration (N_A, N_D)	κ	$\varepsilon\kappa$

the subthreshold slope SS and the *off*-current I_{off}; velocity saturation limits the increase of speed by increasing the current. The "ideal" value of the subthreshold slope is $SS \approx 62$ mV/decade. The "ideal" slope SS lets the device turn *off* quickly, which highly reduces the leakage current while maintaining the drive current. Then again, a steep SS might be used to lower the threshold voltage V_T and thus to attain a larger drive current I_{on} while maintaining the identical *off*-current I_{off}. Besides, an increased series resistance of graded junctions is observed—this is needed due to reliability considerations for hot carriers at the higher voltage levels.

These effects imply that not all parameters should be scaled by one factor. Brews et al. [16] and Baccarani et al. [17] have addressed this by formulating the *generalized scaling concept*. The basic idea is to let the intensity of electric field change by a factor of ε ($\varepsilon > 1$) that is not equal to κ [17]. Hence, the power-supply voltage is scaled less quickly (less quickly than the electric field intensity) by a factor of ε/κ—see Table 1. In addition, the supply voltage V_{DD} cannot be scaled as quick as the device sizes due to the nonscaling nature of the threshold voltage V_T and the subthreshold slope (nonscaling because the silicon's band structure is fixed). The electric field patterns in the MOSFET are kept the same by increasing the impurity doping concentration by ε, which upholds the depletion regions separated. In case that voltages cannot scale in any way, we have the *fixed voltage scaling* scenario ($V = 1$); in this situation, power does not scale whatsoever; nonetheless, power density rises with κ^2 because each device continues to get smaller.

The utmost integration density (i.e., the number of transistors per unit area) is controlled either by the density of power dissipation or by the area dominated by transistors, passive devices, and interconnections. According to the generalized scaling concept, from performance point of view, the active power density will be noticeably higher for the scaled devices because of the raised electric field. At the same time, as the power-supply voltage is decreased, the performance significantly worsens at larger threshold voltages V_T and also becomes more susceptible to tolerances in V_T. Consequently, from the performance viewpoint, with power supply being downscaled, the threshold voltage should be reduced and also V_T control should be improved (i.e., the V_T variation caused by process tolerances should be reduced) [12].

The relative device density (the number of devices per chip) is reciprocal to the square of the relative lithography dimensions for each technology generation. These lithography rules determine the minimum possible dimensions that could be realized on the die. On the other hand, the minimum gate length at each technology generation is determined by analyzing the 2D field effects inside the MOSFET and in particular by predicting the functional form of the potential variation along the transistor channel.

2.2. Physical Limitations. The above considerations are typical for the classical MOSFET scaling, which ended at the 0.13 μm technology node in 2001. Beneath the 100 nm boundary, CMOS design options are significantly limited by quantum mechanical effects of tunneling currents (these include gate oxide tunneling, band-to-band tunneling, and source to drain tunneling) and carrier confinement, and voltage nonscaling. When semiconductor structures are thin enough, current could pass ("*tunnel*") through the various potential barriers in the MOSFET and thus to degrade the typical device's behavior.

Current tunneling through the gate oxide insulator is one of the most pronounced constraints to scaling. It causes the unwanted leakage current that prevents complete *on*/*off* device control by the gate voltage. By rule, lower power devices require thicker oxide thicknesses. Today's thickness of the SiO_2 gate insulator is in the interval of 1.0–1.5 nm, and in this interval, the density of leakage current becomes higher than 1 A/cm^2 [18]. This contributes tens of milliwatts to the total chip dissipation, which is challenging for low-power applications. Accordingly, the minimum thickness of SiO_2, in order to act as an insulator, is at least 0.7–0.8 nm.

There are three options to tackle the direct *gate oxide tunneling* limitation. One is to discontinue scaling the oxide and to scale the rest of the FET in such a way that the thicker oxide could be compensated. The second is to modify the transistor's structure in such a manner as to enable further MOSFET scaling, even with a fairly thicker oxide—dual- and multiple-gate FETs are an illustration of this approach. And the third is to alter the oxide of gate insulator to another material with respective capacitive properties such that the *effective oxide thickness* can be reduced without increasing the tunneling current—high-k gate dielectrics are an example of this approach [19].

Band-to-band tunneling (BTBT) effect typically takes place amid the drain and body of a MOS transistor. The high doping of the channel (that goes along with the scaling) results in a high electric field across the depletion layer at the reverse-biased drain junction. The high electric field (~10^6 V/cm) encourages a parasitic leakage current that is associated mostly with the tunneling of electrons between the valence band in the channel to the conduction band in the drain [20]. The nearness between the source and drain junctions may result in quantum mechanical tunneling that will enlarge the total leakage current of the transistor [21];

source-to-drain tunneling is another potential cause of tunneling current that can disturb the operation of short-channel MOSFETs.

Voltage scaling is limited in numerous aspects. The voltages of built-in junction are preset by the native bandgap potential of silicon, which cannot be scaled down. Accordingly, the internal fields do not necessarily scale as desired as the applied voltages are scaled down toward 1 V. An analogous problem occurs when trying to scale down V_T, which is related to the nonscaling behavior of the subthreshold slope and its impact on the *off* current [19].

As device dimensions get smaller, the influence of parasitic resistances and parasitic capacitances of the drain and source become substantial. In particular, parasitic S/D resistances, which elucidate the voltage drop between S/D contacts and the channel, increasingly affect transistor *on*-current. Hence, in order to enable performance growth, suitable control of parasitic S/D resistance is required. Forming shallower junctions allows effective control of parasitic resistances, but it is difficult to be implemented; besides, it increases the BTBT leakage current. For digital applications, lower capacitance is a key for high-performance CMOS circuits. Dimension scaling, however, increases parasitic capacitances, which requires techniques for their control [22].

2.3. Short Channel Effects. The *short channel effects* (SCEs) signify the disadvantages originating from technology downscaling. Many device nonidealities make the device characteristics very different from the desired ones. The horizontal and vertical electric fields escalate massively and affect numerous MOSFET parameters. The electric field rises because voltages are not scaled down along with the other parameters in order to maintain the speed of the MOSFET. With downscaling transistor gate length L_G, the transfer characteristic (I_D vs. V_G) worsens because (i) the subthreshold slope SS flattens and the threshold voltage V_T decreases; hence, the device cannot be switched *off* by pulling down the gate voltage V_G, and (ii) the slope SS and voltage V_T get more sensitive to variations of L_G that are inherent to the manufacturing processes.

In the long-channel case, a transistor is turned *on* or *off* by raising or lowering V_G, respectively, which in its turn controls the channel potential via the gate-to-channel capacitance C_G. Hence, the electrostatic control of the channel potential is achieved by V_G and C_G only. In the short channel case, the channel potential is affected also by the drain voltage V_D via the drain-to-channel capacitance C_D because the value of C_D is no longer negligible compared to the value of C_G; in the very short channel length case, it even can occur that the transistor is turned on by V_D only (cf. the DIBL effect below). In result, the gate itself cannot effectively control the channel. One solution to this problem is to reduce the *gate oxide thickness*. This is a crucial parameter, which has been severely scaled to attain enough drive current and to control short channel effects. Another solution is to increase the *doping* of the channel—it leads, however, to other disadvantages such as decreased mobility, hindered device performance, etc.

Some of the common SCEs include the following. *Channel length modulation* (CLM) is the effect which leads to a linear increase of the drain current in saturation region that is proportional to the drain voltage in the saturation region. *Carrier field mobility degradation* is an effect of degradation of carrier mobility due to the scattering mechanisms associated with the influence of the horizontal and vertical electrical fields (horizontal-lateral; vertical-transversal) present in the transistor channel and the increased temperature. *Hot carrier effects* (HCEs) are due to the presence of high electric fields in a device—at the high electric fields, the carriers in the channel may get high kinetic energy and attain sufficient energy to overcome potential barriers, in result, these carriers might drift into an undesirable area such as the gate dielectric, the gate, or the substrate, and to cause impact ionization increase of the subthreshold current or carrier injection into the gate dielectric. This shifts the threshold voltage and compromises the transistor behavior. *Drain induced barrier lowering* (DIBL) refers to the effect of lowering the threshold voltage due to the depletion in the channel induced by the drain potential: at submicron and sub-100-nm dimensions, the increased depletion implies inversion at lower threshold voltage—this leads to increased leakage and static power dissipation. As discussed above, *gate oxide tunneling* due to thin oxide thickness is the reason for enlarged power dissipation and deficit of *on*-current density. The *polysilicon depletion* effect refers to the potential across the polysilicon gate that depletes the carriers in the gate; in result, the effective voltage decreases, which lessens the inversion and therefore the drain current. At sub-100-nm scales, aggravate *statistical variability* effects such as statistical variation of the dopants (random dopants) in the substrate, present manufacturing methods, do not control the exact positioning of dopant atoms, and therefore, large statistical variations become likely which limit device scaling. Finally, *ballistic transport* effects occur in sub-100-nm channel devices where carriers do not scatter off of semiconductor lattice ions since the channel length is shorter compared to the average free path of carriers.

3. Technology Evolution

3.1. Bulk CMOS. To overcome physical limits of device miniaturization, new device structures and new materials are being developed and implemented. These innovations could be divided in two trends: (i) modification of CMOS technologies to extend Moore's Law—the so called "More Moore" and (ii) implementation of novel device structures—the so called "More than Moore" or "Beyond Moore." The "More Moore" direction includes measures for strengthening the electrostatic control of the channel by means of continued scaling of effective oxide thickness with high-k/metal gate stack and multiple-gate structures for higher drive current at constant overdrive voltage (V_{DD}-V_T), improving mobility of carriers by implementing high-mobility materials and techniques such as strain engineering for the channel, semiconductor band-gap engineering, reducing parasitics, etc. [23].

The "More than Moore" trend is about diversification of semiconductor device functionalities, and as such, this trend does not represent an alternative to Moore's Law. Analog, digital, and mixed-signal processing are extended by incorporating nondigital functionalities such as micro-electromechanical devices, microfluidic devices, sensors, and actuators. These add-ons along with the associated power supplies constitute the interface to the outside world through sensors, actuators, etc. In result, an overall system miniaturization is achieved despite that the added nondigital parts do not scale at the same rate as the semiconductor digital processing parts.

Planar bulk CMOS (complementary metal-oxide-semiconductor) technology has been the mainstream VLSI technology for the past decades. CMOS technology is based on the exploitation of complementary transistors, labelled by *n*- and *p*-types, in which carriers are electrons and holes, respectively. The use of complementary devices facilitates the decrease of power consumption and the increase of gain and is suitable for realization of logic circuits. The key areas of bulk CMOS technology include the gate dielectric and gate stack, source/drain resistances, channel doping profile, isolation, lithography, and etching means and interconnect matters for high-performance applications [12].

A portion of the above technology areas are performance-driven— in order to improve it, more abrupt profiles of the source/drain (S/D) are needed to lower series resistance, slimmer gate dielectrics are necessary to enlarge drive current, and silicided gates are implemented to decrease the resistive/capacitive delay for devices with greater width. Other technology areas are density-driven—*shallow trench isolation* (STI) or silicided S/D junctions. STI is a technique that enables denser circuits by providing better planar isolation limited only by the constraints of lithography. In addition, STI increases device performance in consequence of the minimized perimeter junction capacitance; the latter is achieved by shifting the doping profile's field far from the junction edge. The silicided source/drain junctions contribute to a better density by means of cutting down the number of contacts to the source/drain area and thus leaving free wiring tracks [12].

In a scaled bulk CMOS device, as mentioned above, the channel is getting shorter and the electric field higher; thus, the gate cannot control over the carriers' inversion in the channel. By means of increasing the channel doping, gate control is improved, but in a heavily doped channel, carrier mobility is significantly degraded due to impurity scattering and an increased vertical electric field. Today's CMOS technology features transistor gate oxide thickness between 1.0 and 1.5 nm and channel's doping of the order of $1 \times 10^{18}\,\mathrm{cm}^{-3}$. Heavy channel doping (i) enhances the body-to-drain leakage current due to band-to-band tunneling (BTBT), (ii) reduces carrier mobility, (iii) increases the junction capacitance, and (iv) degrades the subthreshold slope. In addition, thin gate oxides have shorter width of the potential barrier that divides the channel from the gate thus increasing the leakage current due to direct gate oxide tunneling. All these phenomena could cause incorrect transistor operation and higher power dissipation. At the same time, considerable depletion charge emerges in the transistor channel, which augments the vertical electric field in channel. This leads to an enlarged depletion capacitance (and therefore to an enlarged subthreshold slope) and raised phonon scattering effects in the channel. Hence, V_T should be elevated to keep down the leakage *off*-current, but this reduces the drive *on*-current.

Numerous solutions and innovations have been required to elude the constrains due to fundamental physics under the 100 nm linewidth. In summary, the latter implies the need to (1) reduce the effects of quantum-mechanical tunneling of carriers (1.1) through the thin gate oxide, (1.2) from source to drain, and (1.3) from drain to transistor body; (2) to keep a high *on-off* current ratio, which necessitates to control the dopant atoms density and location in (2.1) the transistor channel and in (2.2) the source/drain regions; and (3) the finite subthreshold slope [3, 19].

An essential innovation enabler for sub-100-nm VLSI technology nodes and those under 10 nanometers is the high-*k* gate oxide transistors. In sub-100-nm transistors, the thickness of gate oxide dielectrics is ~1.0–1.5 nm and gate tunneling leakage increases exponentially because of the quantum mechanical tunneling. This compromises device operation and increases power dissipation. A workaround is to replace the SiO_2 as a gate insulator with a high-*k* (high dielectric permittivity) gate insulator that allows a thicker dielectric layer to behave as a thinner SiO_2 layer—high-*k* dielectrics have thinner *effective oxide thickness* ("effective" in reference to the thickness of SiO_2) [24].

MOSFETs use polysilicon (polycrystalline silicon) for gate electrode for more than three decades. In nanometer transistors, however, the gate becomes more heavily doped which leads to increased resistance and hence substantial resistive/capacitive delay. In addition to the enlarged doping concentration, polysilicon cannot resist the depletion that occurs at the SiO_2/poly-Si interface during inversion—the *polysilicon depletion effect*. A solution to these problems is the use of a metal for gate material with the metal being compatible to the novel high-*k* dielectrics.

Carrier mobility loss which is a result from higher levels of channel doping and downscaled gate dielectrics have to be compensated in order to keep the rate of increase in device performance. Solutions to this problem are SOI technology and mobility-enhancement technology.

3.2. Silicon-on-Insulator. Performance enhancements can be accomplished by realizing shrunk silicon CMOS architecture over an insulator layer—*silicon-on-insulator* (SOI). SOI wafer has a very thin Si-layer (of the order of tens of nanometers) placed on top of a SiO_2 insulating layer called *buried oxide* (BOX) (its thickness is of the order of 100–200 nm). It is located above the Si-substrate. The performance improvement of SOI over bulk CMOS is for the most part because of the decreased parasitic capacitances (owing to the reduced source/drain junction capacitances) and body effect due to its buried oxide layer. It isolates the channel from the substrate and blocks the increase of the transistor threshold voltage. SOI transistors might be "*partially*

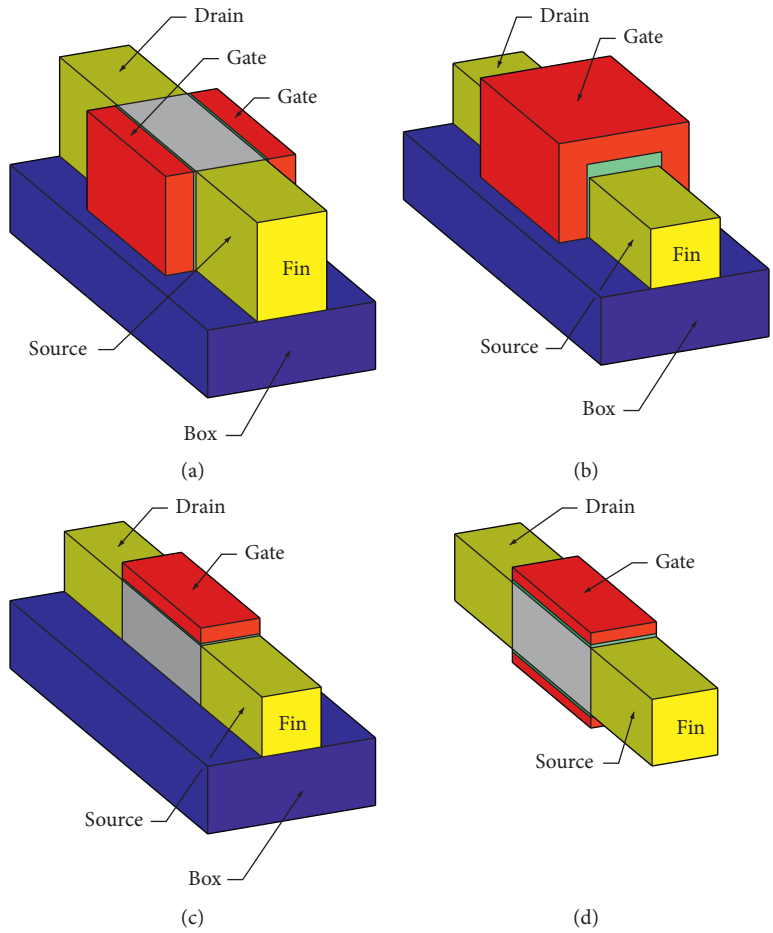

Figure 1: Different types of SOI devices: (a) independent double-gate FinFET on SOI, (b) double-gate nonplanar FinFET (common gate FinFET), (c) single-gate SOI transistor, and (d) double-gate planar SOI transistor.

depleted" (PD-SOI) or "*fully depleted*" (FD-SOI). In PD-SOI transistors, the Si-film (normally larger than 100 nm) over the insulator is thicker than the depletion thickness under the gate oxide. In FD-SOI transistors, the silicon film is so thin (usually less than 50 nm) that the body is fully depleted; full depletion can occur also when the doping concentration of the body is low enough to let its full depletion.

Ultrathin body (UTB) SOI device is a 3D structure as it has a very thin vertical Si-body (referred to as "*fin*"), which forms the channel of the MOS transistor (Figure 1). The "fin" transistors are used in the latest 14 nm, 10 nm, and 7 nm technology nodes. The 3D transistors with the fin surrounded by the gate allows for enhanced power and efficiency compared to the classical 2D planar transistors.

Single- or double-gate MOSFETs could be realized through the UTB concept. The *single-gate* (SG) layout (Figure 1(c)) has the source, channel, and drain placed over the insulator and the gate on top the channel. The *double-gate* (DG) layout features two gates at the opposite sides of the thin body. There are two modifications of the DG transistor: a planar DG SOI transistor (called also *UTB SOI transistor*) (Figure 1(d)) and DG nonplanar transistor called *FinFET*. The DG SOI has two gates that can be either biased independently (an independent DG FinFET) (Figure 1(a)) or they could be connected (a common gate DG FinFET) (Figure 1(b)). The firm coupling between two gates heads to an "ideal" subthreshold slope of the DG transistor. If the body of DG MOS transistors is undoped, the source and drain junction capacitances are significantly reduced; *undoped* means that the silicon layer is assumed to have intrinsic doping concentration. Also, the reduction of junction capacitance leads to higher switching speed of DG transistors.

A highly doped region in the silicon under the BOX layer works as a back gate. The front gate forms the inversion in the channel while the back gate is principally used for adjusting the threshold voltage V_T; typically, the buried oxide is thick enough that the back-gate does not produce an inversion layer at the back surface. This means that from electrostatic point of view, the DG transistor is better than the SG one because the two gates control the channel from both sides and any paths for hypothetical leakage currents will be close to one or the other gate; hence, these paths will be cut off by the electrical fields of the gates. The second gate also eliminates the drain electric field penetration through the BOX layer, which additionally improves the gate control of the channel and mitigates SCEs.

The major benefit of SOI technology is that it serves as a fundament for novel structures such as *multiple-gate (MuG)*

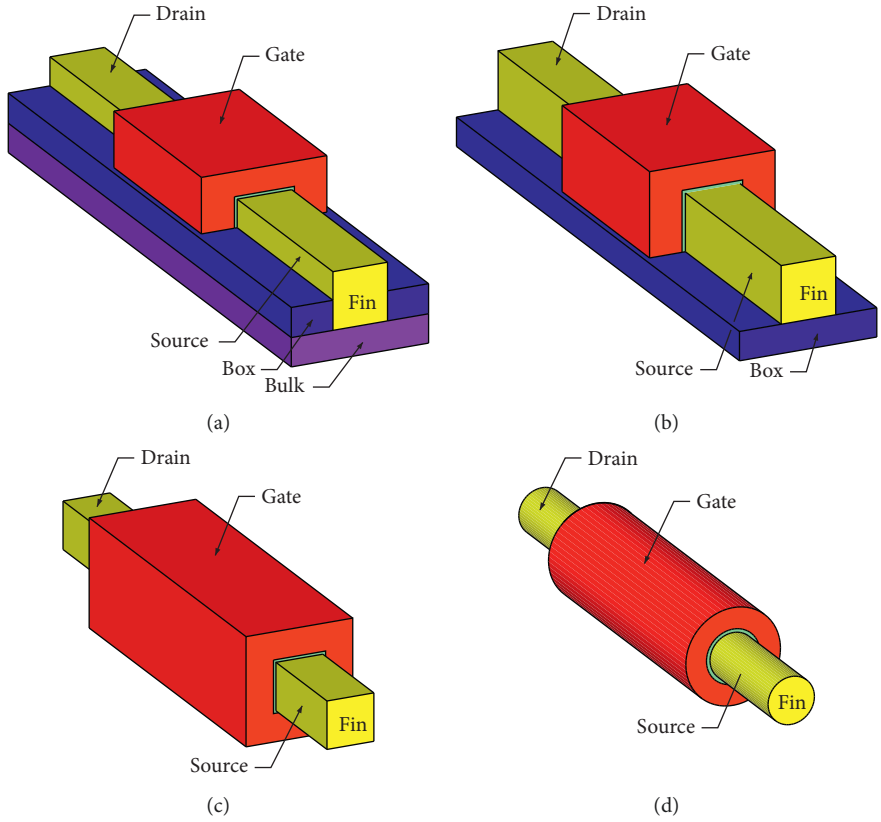

Figure 2: Different types of SOI devices: (a) triple-gate FinFET on Bulk Si, (b) triple-gate FinFET on SOI, (c) quadruple-gate-all-around FET, and (d) surrounding-gate-all-around FET (nanowire FET).

FETs. The above-described DG transistor is the first representative of MuG FET structures. Along with the double-gate FETs, there are triple-gate transistors and gate-all-around transistors (Figures 2(a) and 2(b)). The manufactural-friendly version of the double- and triple-gate FETs is the FinFET structure. The *gate-all-around* (GAA) transistors have three modifications—quadruple-gate-all-around FETs (Figure 2(c)), surrounding-gate-all-around FETs, such as nanowire-channel transistor devices (Figure 2(d)). Depending on the manufacturing process, the nanowire channel may be horizontally or vertically oriented. These structures improve the electrostatic control of the channel and the drive current even better than the DG; the nanowire FET is most effective, though still not manufactural-friendly [24].

For ultrathin body MOSFETs, short-channel effects are sharply dependent on the body thickness and can be assessed by the proportion of the gate length to the body thickness [23]. The advantages of ultrathin body include reduced junction capacitance, no need for high channel doping to withhold short-channel effects (light doping lessens the vertical electric field in the channel), improved electrical isolation, and reduced tunneling leakage through the gate oxide. UTB and especially DG devices feature a performance speedup (~5–10%) because the capacitive load is lessened by the removal of both junction and depletion capacitances. In addition, dopant fluctuation is eliminated that reduces the overall device variation [25].

3.3. High Carrier Mobility Devices. Carrier transport in Si MOS transistors is key for device performances. In long-channel devices, the stationary (diffusion) transport dominates the carrier transport and carrier scattering events that happen inside the channel. As the channel shortens, the nonstationary transport and scattering events occur more rarely in the channel—quasiballistic transport. Going to channel lengths under 10 nm, almost no scattering events occur and ballistic transport dominates the carrier transport inside the channel [26].

Carrier transport mechanisms determine the drive current in the channel (i.e., the *on*-current). The increase of the drive current in the channel results in a reduced supply voltage and therefore lower power consumption keeping the performance of CMOS circuits high. The effective mass decrease is rather effective in increasing *on*-current because it directly lifts up the velocity of carriers.

An essential feature in Si and Ge conduction bands is the anisotropy of effective mass. Therefore, subband structures are considerably dependent on surface orientation of the crystal. This anisotropy enables to design an optimum effective mass by choosing appropriate surface orientation Si and Ge crystal layers. Correctly applied strain to the channel of MOSFET can considerably enhance its performance by changing energy bands and increasing carriers mobility (for reference, we give the electron/hole mobility of Si, Ge, and several AIII–BV compound semiconductors: Si—1600/430 cm^2/Vs, Ge—3900/

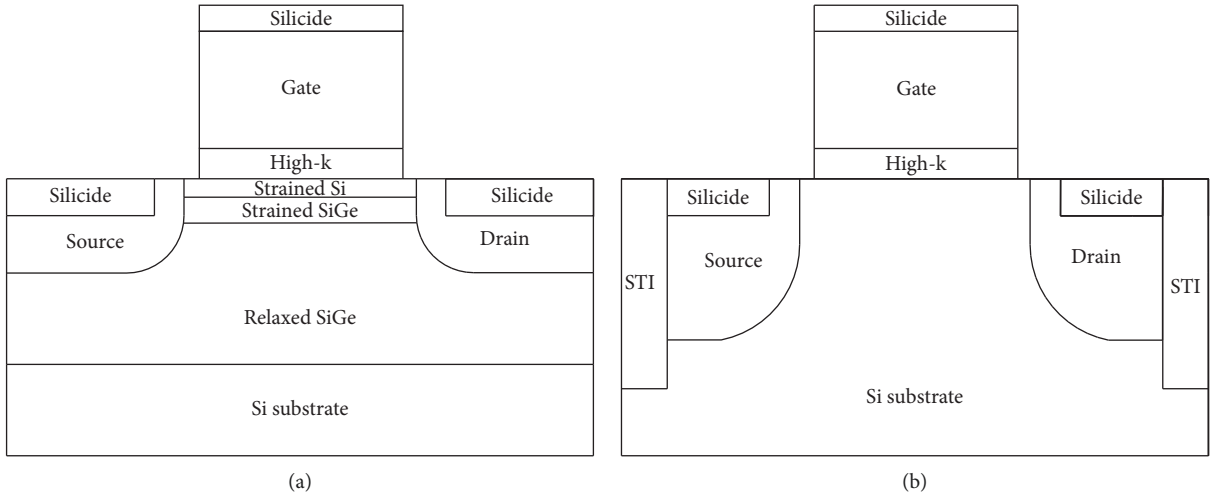

Figure 3: MOS transistor (a) with a strained Si-channel and (b) without a strained Si-channel.

1900 cm^2/Vs, GaAs—9200/400 cm^2/Vs, InP—5400/200 cm^2/Vs, InAs—40000/500 cm^2/Vs, and InSb—77000/850 cm^2/Vs [28]).

Strain engineering technology (or *mobility-enhancement technology*) implements channel materials (such as Ge and GaAs) with high carrier mobility (low effective mass) and/or strained channel materials by means of strain engineering for mobility enhancement. Welser et al. from the Stanford group demonstrated strained-silicon MOSFETs in 1992 [27] for the first time. Afterwards, strained-Si technologies have been developed and implemented in the sub-100-nm VLSI CMOS technologies. Strain engineering technologies are used even in low-power ICs, because of the performance improvement without any scaling and because of their compatibility to conventional CMOS technologies. At present, these technologies rely on manufacturing processes utilizing materials for the transistor channel: strained-silicon (SiGe), germanium-on-insulator (Ge-OI), and compounds of A^{III}-B^{V} semiconductors. A^{III}-B^{V} compound semiconductors are prospective candidates for channel materials of sub-10-nm technology nodes. In combination with multiple-gate structures, excellent solutions could be reached. The main challenge in realizing such devices are the integration technology and compatibility to CMOS technology [24].

Global and local strain are the two categories of strain engineering. Global strain technology for strained-Si MOS transistors is related to depositing thin strained layers of silicon on entire bulk relaxed SiGe. Such devices are bulk strained-Si/bulk relaxed-SiGe, strained-Si/SiGe-on-insulator (SiGe-OI), and single strained silicon-on-insulator (SOI) substrates [28]. Local strain can be introduced by using the thermal mismatch of silicon and the insulation materials, for example, shallow trench isolation (STI), silicidation at the source/drain area, and nitride contact etch stop liners (CESL). Another method to introduce tensile strain is by growing epitaxial silicon layer on material with different lattice constant such as relaxed-SiGe [3]. A cross section of strained Si-channel transistors created by both global and local strains is depicted in Figure 3.

The ternary compound of InGaAs requires special accent, as it has remarkable electron transport characteristics. Depending on strain and composition, electron mobility of InGaAs varies between 6,000 and 30,000 cm^2/Vs at room temperature [29]. This makes InGaAs outstanding for high-speed transistor applications (signal amplification at frequencies of ~500 GHz up to THz-es) in the so called High-Electron Mobility Transistors (HEMTs)—the HEMT was firstly demonstrated by Mimura et al. from Fujitsu in 1980 [30]; another type of HEMTs are based on the binary compound of GaN. HEMTs are heterostructure field-effect transistors. Scaling of HEMTs reached its limit because of the gate leakage current that originates from the relatively small bandgap of the barrier material. A solution to this issue is to use barrier with wider bandgap. Another approach is the so called Quantum-Well InGaAs MOSFET, which removes the wide bandgap barrier and gives reduction of parasitic resistance [31].

3.4. 3D Integration. Classical 2D CMOS technology might be elaborated by using also the vertical dimension in order to meet the demands of technology device scaling. The third dimension could be used in two ways: (i) by stacking the integrated circuit chips one on top of the other and then interconnecting them and (ii) by stacking complementary (*n*- and *p*-type) FETs layers one on top of the other in a single chip. The first option could be considered as a 3D packaging technique along with package-on-package (PoP) and side-by-side mounting. The second option is representing the so called 3D monolithic integration (also referred to as sequential-3D integration). This concept aligns not only to the constantly downscaling dimensions by integrating more functionality in a smaller footprint, but it also overcomes lots of challenges such as control of the strain in transistor channels and hence control of carrier mobility and control of threshold voltage.

3D integration technology starts with fabrication of MOSFETs in a conventional bulk or SOI substrate for the

base layer onto which upper semiconductor substrate layer(s) are added by processes such as molecular bonding, polycrystalline deposition and recrystallization or selective epitaxial overgrowth. Typically, the base layer consists of *n*-type MOSFETs and the upper layer—of *p*-type MOSFETs. Insulation between layers (interlayer dielectrics with thickness ~100 nm) is realized via silicidation process. The interconnects between the vertical layers are done by 3D contacts (vias) using a single lithography. It is important to achieve low temperature (referred to as thermal budget) 3D integration processes in order to safe-keep the lower layer devices, interconnects, and bonding interfaces from degradation during formation of upper layer(s). In addition to thermal issues, there are technology challenges related to interconnect metals, contacts, and gate stacks. Main applications of 3D integration technology are in memory ICs, in particular, in embedded SRAMs [32, 33].

3.5. Carbon Nanotube Transistors. Carbon nanotube field-effect transistors (CNTFETs) are similar to the MOSFETs as they also have three terminals (source, drain, and gate), and the gate controls the current between the source and drain through a channel. The difference of CNTFETs compared to MOSFETs is that the channel in CNTFETs is a carbon nanotube, while in MOSFETs, the channel is made of heavily doped silicon. Matching up nanowire FETs against CNTFETs, the latter have high carrier mobility and smooth channel shell due to the carbon-nanotube nature of the channel [34].

Carbon nanotubes (CNTs) are cylinders of rolled-up carbon atoms sheets with nanoscale dimensions (diameter ~ 1–2 nm) and high aspect ratio (length-to-diameter ratio). Their electrical, thermal, and mechanical properties predetermine their applications in electronics, e.g., CNT transistors, electrical interconnects, electrochemical energy storage, and mechanical additives to diverse structural materials. Carbon nanotubes exhibit semiconducting or conducting properties according to their chirality (i.e., the angle and the diameter of folding the carbon sheets in cylinders). For transistor applications, the bandgap of semiconducting CNTs is important as it determines the threshold voltage—the bandgap is inversely related to the nanotube diameter [3].

In CNTFETs, CNTs with specific chirality that determines their semiconducting behavior are implemented to form the transistor channel. CNTFETs have several technology advantages such as significant performance improvement (low power and high speed), easy incorporation of high-*k* dielectrics, and the almost identical *IV* characteristics of *p*- and *n*-type transistors, which makes them just right for CMOS circuit design. The downsides include the lack of a precise fabrication process for synthesizing nanotubes with identical chiralities and parameter, the difficulties to control the precise placement of nanotubes, etc. [24].

With their properties, CNTFETs emerged as capable runners for next generation semiconductor devices below the 10 nm linewidth. They are a typical representative of sub-10-nm gate-all-around (GAA) transistors. There are several types of CNTFETs: MOSFET-like CNTFETs, Schottky-barrier CNTFETs, and tunneling CNTFETs [35]. Compared to conventional CMOS technology, the CNTFETs show better device performance—sub-10-nm CNT transistors outperform MOS transistors with more than 4× the current density [36].

3.6. Transistors with 2D Semiconductor Materials. The implementation of two-dimensional semiconductor materials in the sub-20-nm channel MOSFETs is a consequence of the quest to diminish the undesirable short channel effects. The main manifestation of short channel effects in bulk (3D) MOSFETs is the escalation of static power due to the occurrence of off-state current and the consequent escalation of leakage current. The latter is related to the heat dissipation accumulated by the raised static power and it appears from "bad" electrostatics of carriers in the channel and the electric field aroused by the gate voltage. In 3D ultra-thin body FETs, the performance is deteriorated (there is carrier scattering) due to the presence of dangling bonds, unwanted coupling with phonons, and the formation of interface states. If the transistor channel is made from 2D material, the transistor body is ultra-thin (atomic thickness) by nature. All the carriers are confined in the atomic-thick channel, and all of them are uniformly controlled by the electric field of the gate. Therefore, the leakage current is suppressed, and there is no carrier scattering as there is no dangling bonds in the channel. These properties along with their low dielectric constant, high on-state to off-state current ratio, and wide-bandgap enable the application of 2D materials for low-power subthreshold semiconductor electronics.

Typical 2D materials that are applicable to FETs are graphene, the transition metal dichalcogenides (TMDs, denoted also with their general formula MX_2), silicone, phosphorene, and boron nitride. The first 2D material, the graphene, is not appropriate for MOSFET applications as it is a zero bandgap semiconductor; it has no energy bandgap, which is needed to ensure the MOSFET switching properties, i.e., high I_{on} and low I_{off} currents.

Most prospective for industrial implementation are the TMDs. The 2D layers of TMDs are weakly coupled by van der Waals forces which allow their mechanical exfoliation. Useful property for transistor applications is the dependence of the energy bandgap on the layer thickness. In general, TMDs can be insulators, semiconductors, or metals: HfS2 is an insulator; MoS_2, WS_2, and $MoSe_2$ are semiconductors; WTe_2, $TiSe_2$, NbS_2, and VSe_2 are metals. Currently, primarily MoS_2 proves potential for practical sub-20-nm channel transistors applications—it is available as natural substance, it could be easily obtained in 2D crystal form by exfoliation techniques, and it has low dielectric constant ($\varepsilon = 4$–7), thin body (~0.7 nm per layer), quite high energy bandgap (1.85 eV per layer), and high effective mass (ensuring a high I_{on}/I_{off} ratio) with almost symmetrical values for electrons and holes. There are, however, technology challenges for realizing MoS_2 short channel FETs. These include structural

defects during exfoliation (sulfur vacancies and impurities that restrict transistor performance), doping (usually MoS_2 MOSFETs demonstrate *n*-type behavior without intentional doping), and lithography (EB lithography can deliver 10 nm resolution, but with limited productivity and difficult dimensions' control) [37, 38].

So far, extensive research efforts are focused on studying at least three key subjects in order to realize a high performance MoS_2 MOSFET: (i) dielectric deposition onto the 2D MoS_2 crystal, (ii) manufacturing of low-resistivity semiconductor-to-metal junction for contact pads, and (iii) short channel effects. There are successful high-*k* dielectric deposition techniques demonstrated. For achieving MoS_2/metal contact resistance that is comparable to the contact resistances of other low-dimensional systems such as graphene or carbon nanotube, the research is focused on finding metals/alloys with work function positioned on or close to the edge of the conduction band (for *n*-type transistors) or on the valence band edge (for *p*-type transistors). The short-channels subject should be also addressed in the context of the overall transistor scaling and increasing integration density on single chip rather as "short-channel properties" of the MoS_2 are already widely investigated [39].

Other promising 2D material is the phosphorene (black phosphorus), first synthesized in 1914. It has anisotropic transport behavior. Its atoms join together to form a 2D crumpled sheet. Phosphorene is a direct energy bandgap material that is adjustable by the applied strain. Similarly to graphite, phosphorene is composed of many layers, and it has thickness dependent bandgap (i.e., the bandgap depends on the number of layers)—the higher the number of layers, the narrower the bandgap. For comparison, MoS_2 materials distinctly change from indirect bandgap in multilayer MoS_2 to direct bandgap in single-layer MoS_2. For these reasons, FETs fabricated using phosphorene have better performance than FETs based on TMDs. However, there are two major challenges that restrict phosphorene industrial applications: (i) there is still no technique for large-scale synthesis of phosphorene and (ii) phosphorene is unstable and prone to degradation [38, 40].

3.7. Single Electron Transistors. Single electron transistors (SETs) are also 3-terminal devices analogous to conventional MOSFET, with an optional second gate. Between the source and drain the current is "composed" of a single electron, i.e., the device operates completely on the quantum mechanical effect of tunneling. The SET is comprised of one small conductive "island" (a quantum dot) that is stacked between two tunnel junctions with high resistivity, and the island is coupled with one or two gates. Two conducting layers separated by a thin layer form the barrier. The "island" (i.e., the quantum dot) is stacked and has a fixed number of elementary charges. This is due to the Coulomb blockade effect (the effect of Coulomb blockade is the electrostatic repulsion felt by an electron that is coming close to a miniature negatively charged area), which controls the number of negative carriers ("electrons") in the island. In result, for definite values of gate and drain voltages, only an interval of charges could undergo quantum tunneling through the channel; the "channel" is induced by a tunneling between the source and drain through tunnel capacitors. The gate voltage attracts carriers to the island through the barrier only. The barrier of the tunnel controls every electron passing, where comes the name from—single electron transistor [41].

Having current of single electrons greatly reduces power consumption; devices with 2 nm sizes could be achieved. But the SETs' shortcomings are high impedance and low gain and they are prone to noise. In order to operate optimally (at ~1 GHz speed), SETs need temperatures far less than the room temperature [42]. There is a large threshold voltage variation as well. Due to these problems, SETs are very unlikely to become the replacement of MOSFETs.

3.8. FinFETs. FinFET is the pioneer of GAA-like devices in mainstream sub-20-nm technologies. Classical MOS transistors are planar (2D), i.e., they form an inversion conducting channel in the silicon region under the gate when the gate voltage is above the threshold voltage. FinFET transistors are 3D structures, i.e., they form the conducting channels on three sides of a vertical fin structure, running fully depleted operation; FinFETs can have multiple fins connected together to increase total drive current for a better performance. FinFET is today's state-of-the-art transistor—implemented in commercial production in 2011 by Intel (referred to as "Tri-Gate transistor"); it underpins the last 22 nm, 14 nm, 10 nm, and 7 nm technology nodes. As noted above, being surrounded from three sides by the gate, the fin hosts the flow of carriers (i.e., the channel) when transistor is switched *on* and prevents the carriers from leaking out when the transistor is switched *off*. FinFETs can be built on bulk (Figure 2(a)) or SOI substrates (Figure 2(b)); therefore, there are two modifications: bulk FinFETs or the SOI FinFETs. SOI FinFETs can be (i) with two gates (DG FinFET)—when there is an oxide isolation on top of the fin below the gate (Figure 1(b))—and (ii) with three gates (tri-gate FinFET)—when the top of the fin as well as its sides all conduct current (Figures 2(a) and 2(b)).

3.9. Nanowire and Nanosheet Transistors. The GAA FET structures comprised of stacked nanowires (NWs) give most promising results to become the successors of the FinFET [43, 44]. For the upcoming technology nodes under 7 nanometers, a device better than FinFET is needed to empower the ultimate gate-length scaling and to achieve good layout area efficiency [1].

Nanowire field-effect transistors (NWFETs) could be classified as a type of *gate-all-around* FETs where the channel is formed by a thin nanowire. The channel being surrounded by the gate from all sides makes the nanowire FET an iteration of the FinFET in technology development as the FinFET has the channel surrounded by the gate from three sides only. This is accompanied by a considerably higher fabrication complexity. NWFETs are considered as a prospective candidate for continuation of transistor scaling because their geometry ensures better electrostatic channel

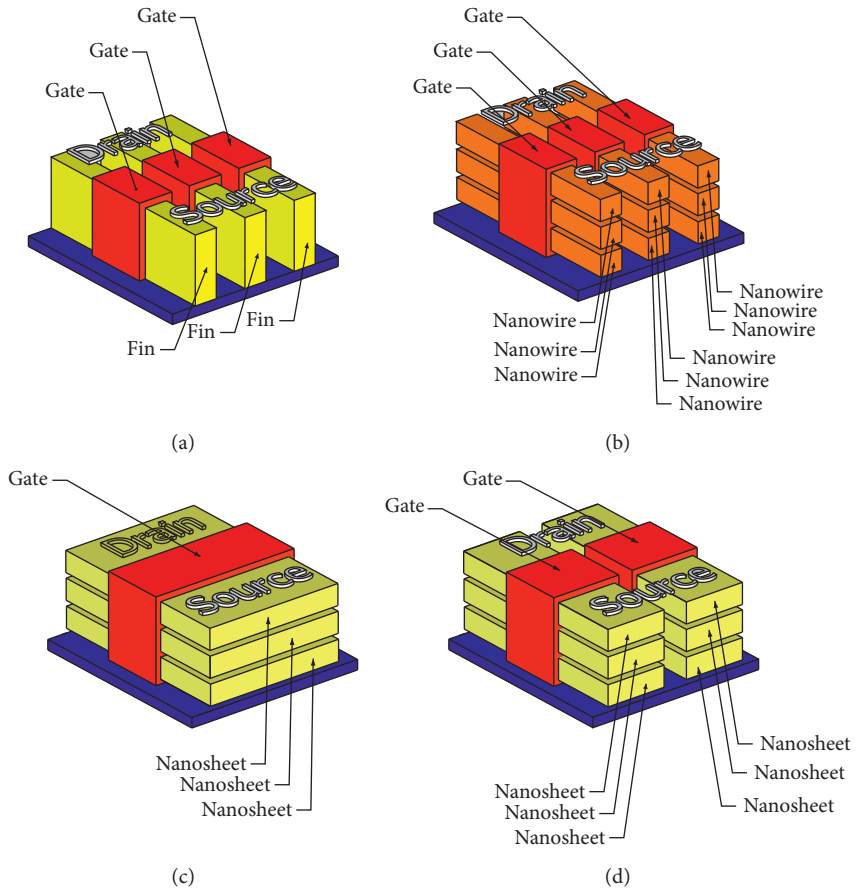

Figure 4: (a) State-of-the-art FinFET, (b) nanowire FET, (c) single-stack nanosheet FET, and (d) double stack nanosheet FET.

control compared to the planar MOSFETs. The better scalability originates from technology reasons as the nanowire diameter could be well controlled to sizes below 10 nm [45].

The particular technological challenges of the NWFETs include the reduced channel width per layout footprint (W_{eff}/layout footprint), the GAA parasitic capacitances, and the complexity of integration of the gate stack [44]. Besides, carriers in the nanowire channel exhibit ballistic transport properties. The diameters of the nanowires vary, which is related to the fabrication technology; these diameter variations might result in variations in the transistor threshold voltage. Finally, there are quantum confinement effects that cause the inversion charge to occur not at the surface but in the bulk; these quantum confinement effects also make it difficult to model the device behavior.

The group of *C. Dupre* from LETI (Grenoble, France) fabricated and analyzed the so-called ΦFET (or 3D-NWFET) in 2008. The ΦFET is a FinFET-like transistor with stacked 15 nm diameter nanowires with an independent gate nanowire structure. The nanowires (which host the transistor channels), that are surrounded by gate, proved to be a structure for effective reduction of the *off* current while yielding very high *on* current levels. In addition, independent gates architecture enables the possibility for decrease of power consumption [43].

In 2013, Hur et al. from Samsung fabricated a GAA nanowire-on-insulator (NOI) structure for devices under 10 nm called "practical" silicon NWFET. The transistor features high-k/metal gate stacks. It exhibits very good short-channel characteristics: subthreshold slope $SS \approx 74$ mV/dec, DIBL ≈ 50 mV/V [46].

In 2015 Zheng et al. from UC Berkley proposed "iFinFET," a novel stacked transistor structure for scaling under the 10-nanometer linewidth. iFinFET stands for "inserted-oxide FinFET," and this is an evolutionary multiple-gate transistor architecture giving improved gate control without additionally complicating the fabrication process. The analyses showed better electrostatic integrity compared to the FinFET (subthreshold slope $SS = 81$ mV/dec) and less gate capacitance penalty compared to the GAA FET [47, 48].

The team of Mertens et al. from IMEC (Leuven, Belgium) have proposed a *nanowire* (NW) gate-all-around transistor structure composed of dense vertical stacks of 8-nm-diameter silicon nanowires; the nanowires are horizontal, and they are built onto conventional bulk silicon substrates (Figure 4(b)). The fabrication relies on *replacement metal gate* process. The IMEC team reported superb short-channel characteristics (subthreshold slope $SS = 65$ mV/decade, DIBL $= 42$ mV/V) comparable to respective FinFET devices. By means of ground plane engineering, the parasitic current channels under the silicon nanowires are efficiently blocked [49].

Nanosheet (NS) transistor is a prospective candidate for the upcoming 5 nm and 3 nm technology nodes because after the 7 nm node for the state-of-the-art, FinFETs will become difficult to switch *off* as the carriers would leak out despite that the channel is surrounded from three sides by the gate. The nanosheet transistor is a variation of the nanowire transistor with layering silicon nanosheets rather than nanowires with the gate completely surrounding the transistor's channel. If the FinFET could be perceived as a vertical structure, the nanosheet FET is a horizontal structure. Nanosheets can bring back the benefits of pre-FinFET planar designs. Horizontally stacked gate-all-around transistors might be fabricated with slight variation of state-of-the-art FinFET technology.

The IBM team with partners from Samsung and Global Foundries fabricated a nanosheet gate-all-around transistor composed of 3 horizontally stacked Si-sheets, each only 5 nanometers thick; prototyped are structures with one or two stacks of nanosheets (Figures 4(c) and 4(d)). The transistors obtained have excellent short-channel characteristics: subthreshold slope $SS = 75$ mV/decade, DIBL = 32 mV/V. Nanosheets' width could vary between 8 and 50 nm. Wider nanosheets provide better performance while narrower nanosheets provide lower power consumption. This would allow for circuit designers to choose the right transistors depending on the performance/power priority of application they do design [50].

4. Compact Modeling of Bulk CMOS Transistors

Design of silicon monolithic integrated circuits for digital and analog applications containing millions to hundreds of millions of transistors is deeply dependent on computer-aided design (CAD) tools and in particular on device modeling and circuit simulation. Needs for circuit simulation are growing as a result of device downscaling under the 10 nm linewidth and the integration of systems with numerous functions on a single integrated circuit. The objective of circuit models is to formulate simple, accurate, and quick analytical equations for the electrical characteristics at the terminals of semiconductor devices that produce numerically efficient results out of circuit simulators; such models are called *compact models*. Hence, compact models have to meet the conflicting requirements of simplicity versus accuracy [25].

In mass production usually different types of MOSFETs are fabricated, so compact models should cover as many transistor types as possible. In addition, precise compact models are needed for the sake of designing and developing analog, digital, and mixed-signal circuits. It is greatly desired that compact device models are physics-based since equations and parameters that are based on device physics are related to real device properties and effects.

The basic equations for expressing device characteristics originate from classical physics: Poisson equation (the relation between the electric field and the gradient of a potential), current-density equations, and continuity equations. Usually, inversion charge or density of inversion charge in the channel is calculated first, and next, the current is determined taking the time derivative of this charge. The majority of device models are built upon a "core model" whilst advanced physical effects such as short channel effects and quantum mechanical effects are implemented as add-ons to core models. Device simulators computationally solve all equations to find device characteristics [51].

Circuit compact modeling of bulk MOS transistors has been pursuing technology developments ever since the beginning of the integrated circuits' fabrication. This progress started with simple piecewise models and evolved to complex continuous compact models. The first SPICE models of Level 1, Level 2, and Level 3 as well as the subsequent BSIM (Berkeley Short-channel IGFET Model by University of California Berkeley) and BSIM2 are all piecewise models that describe transistor *IV* characteristics separately in different regions of operation. The discontinuities of 1st order derivatives of current and capacitances lead to inaccurate and nonphysical description of the transition region (between linear mode and saturation mode of operation). Derivatives' discontinuity triggers convergence difficulties in circuit simulations. The later models such as BSIM3, BSIM4, BSIM5, and PSP employ smoothing functions to connect the operation regions of the MOS transistor. In this way, the discontinuities are overcome and convergence in circuit simulation is achieved. The transition region is better described where analog circuits are usually biased in.

The conventional long-channel MOSFET could be considered as a 2D device as its input gate voltage in the x-direction that is normal to the semiconductor surface in order to control the channel current which passes in the y-direction under the semiconductor surface when a voltage is applied between the source and drain.

Analytical models are created by decomposing the 2D problem into two 1D problems. The equation in x-direction relates the applied gate voltage to the electric conditions of the semiconductor surface; it is called input voltage equation. Its solution is an electrostatic solution of Poisson's equation. The equation in y-direction relates the current flowing in the channel to the voltage between the drain and source and to the x-solution; it is called output current equation. Its solution includes the drift and diffusion currents. This split-up of the 2D problem in two 1D problems is known as the *gradual channel approximation*. It assumes that the potential between the source and drain (i.e., along the channel) alters gradually so that the electrostatic 1D solution is acceptable [52]. Pao and Sah solved the 2D problem in the form of *double integral* over the channel length and inversion layer thickness [53]. It is accurate and valid in all operating regions. This is a physics-based solution, but it was too complicated to be directly used for compact model circuit simulation at that time.

To simplify the computationally cumbersome physical equations of Pao-Sah, *charge-sheet approximation* is developed. It presumes that the inversion charge layer has infinitesimal thickness so that the potential does not change across the inversion region [54]. This assumption permits a huge simplification of the *double integral*.

The double integral cannot be calculated analytically, and the charge-sheet approximation is needed because both mobile and depletion charges are present in this double integral. In result, equations for the surface potential in terms of applied voltages are obtained. In order to get explicit solution for the density of inversion charge, two separate equations are formulated corresponding to the two operating regions (linear and saturation) of the MOSFET—the so-called *threshold voltage based* formulation. Below threshold voltage (V_T) the surface potential is a linear function of the input gate voltage (V_G) and above V_T—it is constant for V_G. The solutions obtained are separate for the two regions; consequently, *smoothing functions* (they are inferred from experiment) are required to connect them [55]. BSIM3 and BSIM4 and MM9 (MOS Model 9 by Philips) are examples of models that are based on threshold voltage formulation [51].

Transistor *IV* characteristics might be determined proceeding from the surface potential or from the density of inversion charge in the channel and then taking the time derivative to obtain the current. These approaches are called *surface potential based* and *charge based* formulation, respectively. Such models are MM11 (MOS Model 11 by Philips), EKV (Enz-Krummenacher-Vittoz by EPFL-Lausanne), BSIM5 (by UC Berkeley), PSP model (a surface-potential based model developed by Pennsylvania State University and NXP Semiconductors, formerly Philips Semiconductors) [56], and HiSIM model [57]. Surface potential-based models surmount the drawback of early BSIM models. These models solve a single equation of the surface potential (ϕ_s) that is valid in all the operating rather than solving two separate equations depending on the threshold voltage. Hence, the ϕ_s-based models are continuous in all operating regions.

5. Modeling of Multiple-Gate Transistors

As discussed above, the principal advantages of multiple-gate transistors do the better handling of short channel effects and the better *on*-state drive current, which entails quicker circuit speeds. There are many types of multiple-gate transistors such as planar double-gate and triple-gate transistors on bulk-Si substrate or on SOI substrate (e.g., FinFETs), gate-all-around transistors (e.g., nanowire MOSFETs), etc. Multiple-gate MOSFETs can be classified in general as common multiple-gate (CMG) and independent multiple-gate (IMG). CMGs are a special case of IMG where the gates are electrically connected, and hence, the same electrical gate voltage is applied to them—gates have the same material, gate oxide thickness, bias, and work function (the regular FinFET is a CMG device) [58]. On the other hand, "independent" denotes that gates are not electrically connected to each other; hence, different gate voltages might be applied to them.

5.1. Double-Gate Transistor Models. Core compact models of DG SOI devices could be either surface potential-based or charge-based; the threshold voltage-based formulation is not relevant because of the undoped Si-body in these devices. The double-gate (DG) MOS device features two gates that are generally asymmetric (or independent) and have different oxide thicknesses. The bottom gate permits for full depletion of the channel. When the two gates are connected and the oxide thicknesses are equal, we have a symmetric device—a common-gate DG MOSFET.

Symmetric undoped body DG SOI MOSFET represents the simplest case as the gates are of the same substance with equal thickness and have same work function and applied bias. As mentioned above, Pao-Sah's double integral of the drain current includes both depletion and mobile charges. Having an undoped (or lightly doped) body means that depletion charges are negligible. Hence, in Poisson's equation remains the mobile charge term only. In result, exact solutions to Poisson's equation and current continuity equation are obtained proceeding from just the gradual channel approximation, i.e., without the use of charge-sheet approximation [59].

For long-channel symmetric DG SOI devices, two ϕ_s-based models stand out. Lu et al. proposed exact solution to Pao-Sah's double integral as a closed-from function without applying the charge-sheet approximation [59, 60]. The other model is developed by Ortiz-Conde et al. [61], it has a physics-based single analytic equation for the current that includes both drift and diffusion components and gives a continuous expression valid for all operating regions.

Two ϕ_s-based models will be mentioned for symmetric DG SOI devices: by Taur et al. [60] and Ortiz-Conde et al. [61]. Three *charge*-based models for symmetric undoped DG MOSFETs are available: by Sallese and the EPFL group (the DG EKV model) [62] and Fossum et al.—the UFDG model (University of Florida DG model) that implements a Poisson–Schrödinger solver for standard undoped DG MOSFETs [63, 64]. The third charge-based model for undoped symmetric DG MOSFETs does not use charge-sheet approximation, and its mathematical formulation proceeds from the exact solution of Poisson's equation; it is given by He et al. [65]. For undoped MOS devices, ϕ_s-based models might be converted into charge-based models [66].

Real devices are usually asymmetric and lightly doped (residual impurities in concentrations of ca. 10^{15} cm^{-3}), so the above assumptions for symmetric and especially undoped devices are hardly applicable. *Asymmetric undoped* DG SOI MOSFETs devices have different work functions of the two gate oxides. The asymmetry is due to unequal oxide thicknesses, unequal flatband voltages of the two gates, and unequal gate voltages. Three models for the asymmetric devices are available: by Lu and Taur [59], by Ortiz-Conde et al. [67], and by Roy et al. [68]. A couple of models consider the case of *asymmetric doped* double-gate MOSFETs numerically. These include Ortiz-Conde et al. [69] who present a numerical solution for calculating the potentials and charges, Jin et al. who numerically calculate the surface potential and drain current [70], and the EPFL group which accounts for the silicon layer doping relying on the hypothesis of equivalent thickness of the Si-layer [71].

5.2. BSIM and PSP Multiple-Gate Models. The BSIM group developed two compact models for the two types of multiple-gate transistors, symmetric and asymmetric: common multigate (BSIM-CMG)—for transistors with symmetric gates—and independent multigate (BSIM-IMG)—for asymmetric/independent gates transistor [72]. As multiple-gate transistors could be built on bulk silicon or SOI—accordingly, BSIM-CMG offers two modes: "SOI mode" or "bulk SOI mode" [73].

BSIM-CMG and BSIM-IMG are models based on surface potential-formulation. As such type of models, they feature continuous and smooth descriptions for *IV* and *CV* characteristics over the transistor's operating regions. All electrical quantities the terminal charges or the terminal currents are either an explicit or implicit function of ϕ_s. Models' core proceeds from the long-channel concept. The calculation of the surface potential includes doped body effect, poly-depletion, short-channel effect, and quantum confinement which are all introduced in the core model. To correctly model MG transistor behavior and effects, the BSIM-CMG has over 150 parameters, split in device and technology parameters, design and physical parameters, fitting or smoothing parameters, etc. The BSIM-CMG model utilizes the notion of quantities depending on the length for velocity saturation and smoothing parameters. By means of 3D modeling of SCEs, the double-gate model core of BSIM-CMG is broadened to model also triple- and quadruple-gate FinFETs [74, 75].

Unlike the BSIM-CMG, which allows for modeling both doped and undoped transistor bodies, for simplicity in solving the Poisson equation, BSIM-IMG presumes a lightly doped transistor body. BSIM-IMG borrows from the BSIM-CMG model many physical effects with the respective modifications to account for the independent (asymmetric) gate operation [76].

The group of Gildenblat at Arizona State University (formerly at Pennsylvania State University) and the NXP group (formerly part of Philips) proposed a compact circuit model for symmetric three-terminal FinFETs with undoped or lightly doped thin body that is based on PSP model. It is comprised of simple analytical expressions for currents and charges. PSP-DGFET is a ϕ_s-based model that keeps the basic structure and formulation of the PSP model for the conventional planar transistor. It consists of a global level, that models geometry scaling effects, and a local level, with parameters that affect the electrical characteristic of devices with specific dimensions. The PSP-DGFET model assumes an undoped or lightly doped silicon channel and ignores the edge effects in the fin corners [77, 78].

5.3. FD-SOI MOSFETs. UTBB technology on thin buried oxide (BOX) has two benefits of a manufacturing process and optimized power consumption [79]. The CEA-LEI group developed the Leti-UTSOI compact model for independent double gate (IDG) devices. It describes the intrinsic charges and currents, including all physical effects of decananometer transistors. The model is valid and accurate for all independent double-gate structures and also features outstanding predictability over technological parameters [80].

6. CNTFET and 2D Transistor Modeling

The core of CNTFETs compact models needs to express the surface potential in terms of the applied gate bias in such a way that the full-band carrier density can be obtained including the unique quantum capacitance (CQ) characteristics of the CNT. Akinwande et al. derived an analytical expression for the *IV* and *CV* of ballistic transport CNFET [81].

An universal CNTFET compact model was developed by the Stanford group in 2007. It describes device nonidealities such as quantum confinement on both axial and circumferential directions, the phonon scattering (acoustical/optical) in the channel region, and the screening effect by the parallel CNTs for CNTFET with multiple CNTs [82]. This model might be stretched to include modeling of elastic scattering in the channel, the parasitic gate capacitances, the resistive source/drain, and the Schottky-barrier resistance [83].

The Stanford group released a sub-10-nm CNTFETs compact model in 2015, which is based on the virtual-source (VS) approach. It describes the carrier VS velocity extracted from experimental devices with gate lengths down to 15 nm and SCEs such as inverse drain-induced barrier lowering and subthreshold slope degradation depending on the device dimensions, etc. [84, 85].

MOSFETs with channels made from 2D materials and in particular MoS_2 transistors could be modeled using the *MIT virtual source* (MVS) compact model [86]. The MVS model has been validated also for short-channel A^{III}-B^{V} compound semiconductors and graphene. With proper extraction of parameters such as gate capacitance, DIBL, SS (subthreshold swing), injection velocity, carrier mobility, and junction resistance, the *I–V* characteristics of MoS_2 FETs could be successfully simulated [37].

7. SET Modeling

Several analytical models of single electron transistors (SETs) are available, each of them proceeding from the single electron tunneling theory and the master equation method of steady-state. The three leading SET models for circuit simulation include master equation model, Monte Carlo model, and macro model.

Uchida et al. developed a compact analytical physics-based SET model. It describes the characteristics of resistively symmetric SETs over a wide source-to-drain voltage range of voltages [87]. This model is applicable for modeling of digital circuits. Inokawa and Takahashi [88] have broadened Uchidas' model to account for asymmetric devices plus the background charge effect which is significant for SET operation. Le Royer et al. [89] and Lientschnig et al. [90] have released models that are based on master equation method. Another compact analytical model is the MIB model (MIB stands for the initials of Mahapatra, Ionescu, and Banerjee) [91]. MIB model also uses the master equation

method and it is verified with Monte Carlo simulations for a wide range of voltages. It is applicable for digital and analog circuits design. The MIB model variation called hybrid MIB model describes a generalized asymmetric device that is crucial for circuit design with SETs [92].

8. Conclusions

This paper has reviewed CMOS technology scaling and device evolution from classical to nonclassical structures plus the corresponding compact models used in circuit design. Physical limitations and short-channel effects have been considered. Currently, at device structure level, the R&D is focused on improving transistor's gate control of the channel current. Nonclassical device architectures as well as advanced materials and technology innovations for improved short-channel control has been examined. Silicon-on-insulator and high-mobility device technologies have been surveyed. In particular, FinFETs and multiple-gate transistors, carbon nanotube transistors, transistors based on A^{III}-B^{V} materials, transistors with channels made of layered 2D semiconducting crystals, single electron transistors, and the prospective nanowire and nanosheet transistors have been scrutinized.

The importance of compact circuit models as a crucial link amid technology and design has been articulated. Circuit design in nanoscale dimensions increasingly relies on compact modeling of devices used. This is due to the progressively complicating technologies for fabrication. The nonclassical device modeling approaches has been enumerated with their principle characteristics including compact models of double-gate transistor, BSIM and PSP for multiple-gate transistors, FD-SOI transistors, CNTFET models, and SET models.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors thank the financial support by the project "Modeling, Characterization, Testing and Development of PCBs for Industrial Electronics Applications" (contract no. Д-057-2018/20.06.2018). The authors also acknowledge the support of Research & Development & Innovations Consortium at Sofia Tech Park.

References

[1] International Technology Roadmap for Semiconductors 2.0, 2015, http://www.itrs2.net/itrs-reports.html.

[2] G. Dan Hutcheson, "Moore's law, lithography, and how optics drive the semiconductor industry," in *Proceedings of SPIE Advanced Lithography, Volume 10583, Extreme Ultraviolet (EUV) Lithography IX*, San Jose, CA, USA, March 2018.

[3] H.-S. P. Wong, "Beyond the conventional transistor," *IBM Journal of Research and Development*, vol. 46, no. 2-3, pp. 133–168, 2002.

[4] J. B. Chang, M. Guillorn, P. M. Solomon et al., "Scaling of SOI FinFETs down to fin width of 4 nm for the 10 nm technology node," in *Proceedings of the Symposium on VLSI Technology—Digest of Technical Papers*, pp. 12-13, Honolulu, HI, USA, January 2011.

[5] K. J. Kuhn, "Considerations for ultimate CMOS scaling," *IEEE Transactions on Electron Devices*, vol. 59, no. 7, pp. 1813–1828, 2012.

[6] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 86, no. 1, pp. 82–85, 1998.

[7] G. Moore, "Progress in digital integrated electronics," in *Proceedings of the Technical Digest—International Electron Devices Meeting*, pp. 11–13, Washington, DC, USA, December 1975.

[8] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, 1974.

[9] K. Mistry, "Tri-gate transistors: enabling Moore's law at 22 nm and beyond," in *Proceedings of the SEMICON West*, San Francisco, CA, USA, July 2012.

[10] B. Hoeneisen and C. A. Mead, "Fundamental limitations in microelectronics-I. MOS technology," *Solid-State Electronics*, vol. 15, no. 7, pp. 819–829, 1972.

[11] R. H. Dennard, F. H. Gaensslen, L. Khun, and H. N. Yu, "Design of micron MOS switching devices," in *Proceedings of the International Electron Devices Meeting (IEDM)*, Washington, DC, USA, December 1972.

[12] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS scaling for high performance and low power-the next ten years," *Proceedings of the IEEE*, vol. 83, no. 4, pp. 595–606, 1995.

[13] W. Shockley, "A unipolar "field-effect" transistor," *Proceedings of the IRE*, vol. 40, no. 11, pp. 1365–1376, 1952.

[14] Y. Taur and H.-H. Lin, "Modeling of DG MOSFET," *IEEE Transactions on Electron Devices*, vol. 65, no. 5, pp. 1714–1720, 2018.

[15] Y. Taur, W. Choi, J. Zhang, and M. Su, "A non-GCA DG MOSFET model continuous into the velocity saturation region," *IEEE Transactions on Electron Devices*, vol. 66, no. 3, pp. 1160–1166, 2019.

[16] J. R. Brews, W. Fichtner, E. H. Nicollian, and S. M. Sze, "Generalized guide for MOSFET miniaturization," in *Proceedings of the 1979 International Electron Devices Meeting*, pp. 2–4, Washington, DC, USA, December 1979.

[17] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a 1/4 micron MOSFET design," *IEEE Transactions on Electron Devices*, vol. 31, no. 4, pp. 452–462, 1984.

[18] Y. Taur, "CMOS scaling beyond 0.1 μm: how far can it go," in *Proceedings of the Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)*, pp. 6–9, Taipei, Taiwan, June 1999.

[19] D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H.-S. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proceedings of the IEEE*, vol. 89, pp. 259–288, 2001.

[20] Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.

[21] J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?," in *Proceedings of the Technical Digest of International Electron Devices Meeting (IEDM)*, pp. 707–710, San Francisco, CA, USA, December 2002.

[22] M. C. Chang, C. S. Chang, C. P. Chao et al., "Transistor- and circuit-design optimization for low-power CMOS," *IEEE*

Transactions on Electron Devices, vol. 55, no. 1, pp. 84–95, 2008.

[23] L. Chang, Y.-K. Choi, D. Ha et al., "Extremely scaled silicon nano-CMOS devices," *Proceedings of the IEEE*, vol. 91, no. 11, pp. 1860–1873, 2003.

[24] Y.-B. Kim, "Challenges for nanoscale MOSFETs and emerging nanoelectronics," *Transactions on Electrical and Electronic Materials*, vol. 11, no. 3, pp. 93–105, 2010.

[25] Y. Chauhan, D. D. Lu, V. Sriramkumar et al., *FinFET Modeling for IC Simulation and Design. Using the BSIM-CMG Standard*, Academic Press, Elsevier, Amsterdam, Netherlands, 2015.

[26] K. Natori, "Scaling limit of the MOS transistor—a ballistic MOSFET," *IEICE Transactions on Electronics*, vol. E84C, no. 8, pp. 1029–1036, 2001.

[27] J. Welser, J. L. Hoyt, and J. F. Gibbons, "NMOS and PMOS transistors fabricated in strained-silicon/relaxed silicon-germanium structures," in *Proceedings of the International Technical Digest on Electron Devices Meeting*, pp. 1000–1002, San Francisco, CA, USA, 1992.

[28] S. Takagi, T. Irisawa, T. Tezuka et al., "Carrier-transport-enhanced channel CMOS for improved power consumption and performance," *IEEE Transactions on Electron Devices*, vol. 55, no. 1, pp. 21–39, 2008.

[29] J. A. del Alamo, "Nanometer-scale InGaAs field-effect transistors for THz and CMOS technologies," in *Proceedings of the ESSCIRC*, Bucharest, Romania, September 2013.

[30] T. Mimura, S. Hiyamizu, T. Fujii, and K. Nanbu, "A new field-effect transistor with selectively doped GaAs/n-Al_xGa_{1-Ax} as heterojunctions," *Japanese Journal of Applied Physics*, vol. 19, no. 5, Article ID L225, 1980.

[31] D.-H. Kim, J. A. del Alamo, D. A. Antoniadis et al., "$L_g = 60$ nm recessed $In_{0.7}Ga_{0.3}As$ metal-oxide-semiconductor field-effect transistors with Al_2O_3 insulator," *Applied Physics Letters*, vol. 101, no. 22, Article ID 223507, 2012.

[32] M. Vinet, P. Batude, C. Tabone et al., "3D monolithic integration: technological challenges and electrical results," *Microelectronic Engineering*, vol. 88, no. 4, pp. 331–335, 2011.

[33] A. Mallik, A. Vandooren, L. Witters et al., "The impact of sequential-3D integration on semiconductor scaling Roadmap," in *Proceedings of the 2017 IEEE International Electron Devices Meeting (IEDM)*, pp. 32.1.1–31.1.4, San Francisco, CA, USA, December 2017.

[34] K. Natori, Y. Kimura, and T. Shimizu, "Characteristics of a carbon nanotube field-effect transistor analyzed as a ballistic nanowire field-effect transistor," *Journal of Applied Physics*, vol. 97, no. 3, Article ID 034306, 2005.

[35] M. H. Moaiyeri and F. Razi, "Performance analysis and enhancement of 10-nm GAA CNTFET-based circuits in the presence of CNT-metal contact resistance," *Journal of Computational Electronics*, vol. 16, no. 2, pp. 240–252, 2017.

[36] A. D. Franklin, S.-J. Han, G. S. Tulevski et al., "Sub-10-nm carbon nanotube transistor," *Nano Letters*, vol. 12, no. 2, pp. 758–762, 2012.

[37] A. Nourbakhsh, A. Zubair, R. N. Sajjad et al., "MoS_2 field-effect transistor with sub-10 nm channel length," *Nano Letters*, vol. 16, no. 12, pp. 7798–7806, 2016.

[38] M. Chhowalla, D. Jena, and H. Zhang, "Two-dimensional semiconductors for transistors," *Nature Reviews Materials*, vol. 1, Article ID 16052, pp. 1–15, 2016.

[39] H. Liu, A. T. Neal, and P. D. Ye, "Channel length scaling of MoS_2 MOSFETs," *ACS Nano*, vol. 6, no. 10, pp. 8563–8569, 2012.

[40] A. Khandelwal, K. Mani, M. H. Karigerasi, and I. Lahiri, "Phosphorene—the two-dimensional black phosphorous: properties, synthesis and applications," *Materials Science and Engineering:B*, vol. 221, pp. 17–34, 2017.

[41] Y. Takahashi, A. Fujiwara, Y. Ono, and K. Murase, "Silicon single-electron devices and their applications," in *Proceedings of 30th IEEE International Symposium on Multiple-Valued Logic (ISMVL'2000)*, pp. 411–420, Portland, OR, USA, Portland, OR, USA, May 2000.

[42] M. Saitoh, H. Harata, and T. Hiramoto, "Room-temperature demonstration of integrated silicon single-electron transistor circuits for current switching and analog pattern matching," in *Proceedings of the IEEE Electron Devices Meeting (IEDM)*, pp. 187–190, San Francisco, CA, USA, December 2004.

[43] C. Dupré, A. Hubert, S. Bécu et al., "15 nm-diameter 3D stacked nanowires with independent gates operation: FFET," in *Proceedings of the IEDM Technical Digest*, pp. 1–4, Washington, DC, USA, December 2008.

[44] S.-D. Kim, M. Guillorn, I. Lauer, P. Oldiges, T. Hook, and M.-H. Na, "Performance trade-offs in FinFET and gate-all-around device architectures for 7 nm-node and beyond," in *Proceedings of the IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, pp. 1–3, Rohnert Park, CA, USA, January 2015.

[45] M. G. Bardon, Y. Sherazi, D. Jang et al., "Power-performance trade-offs for lateral NanoSheets on ultra-scaled standard cells," in *Proceedings of the IEEE Symposium on VLSI Technology*, pp. 143-144, Honolulu, HI, USA, June 2018.

[46] S.-G. Hur, J.-G. Yang, S.-S. Kim et al., "A practical Si nanowire technology with nanowire-on-insulator structure for beyond 10 nm logic technologies," in *Proceedings of the IEEE International Electron Devices Meeting (IEDM)*, pp. 26.35.1–26.5.4, Washington, DC, USA, December 2013.

[47] P. Zheng, D. Connelly, F. Ding, and T.-J. K. Liu, "Simulation-based study of the inserted-oxide FinFET for future low-power system-on-chip applications," *IEEE Electron Device Letters*, vol. 36, no. 8, pp. 742–744, 2015.

[48] P. Zheng, D. Connelly, F. Ding, and T.-J. K. Liu, "Inserted-oxide FinFET (iFinFET) design to extend CMOS scaling," in *Proceedings of the Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)*, Hsinchu, Taiwan, April 2015.

[49] H. Mertens, R. Ritzenthaler, A. Hikavyy et al., "Gate-all-around MOSFETs based on vertically stacked horizontal Si nanowires in a replacement metal gate process on bulk Si substrates," in *Proceedings of the IEEE Symposium on VLSI Technology*, pp. 1-2, Honolulu, HI, USA, June 2016.

[50] N. Loubet, T. Hook, P. Montanini et al., "Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET," in *Proceedings of the IEEE Symposium on VLSI Technology*, pp. T230–T231, Kyoto, Japan, June 2017.

[51] G. Angelov, T. Takov, and S. Ristiç, "MOSFET models at the edge of 100-nm sizes," in *Proceedings of the 24th International Conference on Microelectronics (MIEL 2004)*, vol. 1, pp. 295–298, Niš, Serbia, May 2004.

[52] C.-T. Sah, *Fundamentals of Solid State Electronics*, World Scientific Publishing Co, Singapore, 1991.

[53] H. C. Pao and C. T. Sah, "Effects of diffusion currents on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid State Electronics*, vol. 9, no. 10, pp. 927–937, 1966.

[54] J. R. Brews, "A charge-sheet model of the MOSFET," *Solid State Electronics*, vol. 21, no. 2, pp. 345–355, 1978.

[55] Y. Cheng, K. Chen, K. Imai, and C. Hu, "A unified MOSFET channel charge model for device modeling in circuit

simulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, no. 8, pp. 641–644, 1998.

[56] G. Gildenblat, X. Li, W. Wu et al., "PSP: an advanced surface-potential-based MOSFET model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1979–1993, 2006.

[57] M. Miura-Mattausch, N. Sadachika, D. Navarro et al., "HiSIM2: advanced MOSFET model valid for RF circuit simulation," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1994–2007, 2006.

[58] D. Lu, C.-H. Lin, A. Niknejad, and C. Hu, "Multi-gate MOSFET compact model BSIM-MG," in *Chapter 13: Compact Modeling. Principles, Techniques and Applications*, G. Gildenblat, Ed., Springer, New-York, NY, USA, 2010.

[59] H. Lu and Y. Taur, "An analytic potential model for symmetric and asymmetric DG MOSFETs," *IEEE Transactions on Electron Devices*, vol. 53, no. 5, pp. 1161–1168, 2006.

[60] Y. Taur, X. Liang, W. Wang, and H. Lu, "A continuous, analytic drain-current model for DG MOSFETs," *IEEE Electron Device Letters*, vol. 25, no. 2, pp. 107–109, 2004.

[61] A. Ortiz-Conde, F. J. García-Sánchez, and J. Muci, "Rigorous analytic solution for the drain-current of undoped symmetric dual-gate MOSFETs," *Solid State Electronics*, vol. 49, no. 4, pp. 640–647, 2005.

[62] J. M. Sallese, F. Krummenacher, F. Prégaldiny, C. Lallement, A. Roy, and C. Enz, "A design oriented charge-based current model for symmetric DG MOSFET and its correlation with the EKV formalism," *Solid State Electronics*, vol. 49, no. 3, pp. 485–489, 2005.

[63] J. G. Fossum, L. Ge, M.-H. Chiang et al., "A process/physics-based compact model for nonclassical CMOS device and circuit design," *Solid State Electronics*, vol. 48, pp. 919–926, 2004.

[64] http://www.soi.tec.ufl.edu/ufdg.html, 2011.

[65] J. He, X. Xi, C. H. Lin, M. Chan, A. Niknejad, and C. Hu, "A non-charge-sheet analytic theory for undoped symmetric double-gate MOSFET from the exact solution of Poisson's equation using SSP approach," in *Proceedings of the Workshop on Compact Modeling (WCM)*, pp. 124–127, Boston, MA, USA, March 2004.

[66] A. Ortiz-Conde, F. J. García-Sánchez, J. Muci, S. Malobabic, and J. J. Liou, "A review of core compact models for undoped double-gate SOI MOSFETs," *IEEE Transactions on Electron Devices*, vol. 54, no. 1, pp. 131–140, 2007.

[67] A. Ortiz-Conde, F. J. García-Sánchez, S. Malobabic, J. Muci, and R. Salazar, "Drain–current and transconductance model for the undoped body asymmetric double-gate MOSFET," in *Proceedings of the 8th IEEE International Conference on Solid-State and Integrated Circuit Technology*, pp. 1239–1242, Shanghai, China, October 2006.

[68] A. S. Roy, J. M. Sallese, and C. C. Enz, "A closed-form charge-based expression for drain–current in symmetric and asymmetric double gate MOSFET," *Solid State Electronics*, vol. 50, no. 4, pp. 687–693, 2006.

[69] A. Ortiz-Conde and F. J. García-Sánchez, "A rigorous classical solution for the drain current of doped symmetric double-gate MOSFETs," *IEEE Transactions on Electron Devices*, vol. 59, no. 9, pp. 2390–2395, 2012.

[70] X. Jin, X. Liu, J.-H. Lee, and J.-H. Lee, "A continuous current model of fully-depleted symmetric double-gate MOSFETs considering a wide range of body doping concentrations," *Semiconductor Science and Technology*, vol. 25, no. 5, Article ID 055018, p. 4, 2010.

[71] J. M. Sallese, N. Chevillon, F. Pregaldiny, C. Lallement, and B. Iñiguez, "The equivalent-thickness concept for doped symmetric DG MOSFETs," *IEEE Transactions on Electron Devices*, vol. 57, no. 11, pp. 2917–2924, 2010.

[72] http://bsim.berkeley.edu/.

[73] J. P. Colinge, *FinFETs and Other Multi-Gate Transistors*, Springer, Berlin, Germany, 2008.

[74] M. V. Dunga, C.-H. Lin, D. D. Lu et al., "BSIM-mg: a versatile multi-gate FET model for mixed-signal design," in *Proceedings of the IEEE Symposium on VLSI Technology*, pp. 60-61, Kyoto, Japan, June 2007.

[75] M. V. Dunga, *Nanoscale CMOS Modeling*, Ph.D. thesis, EECS Department, University of California, Berkeley, CA, USA, 2008.

[76] J. P. Duarte, S. Khandelwal, A. Medury et al., "BSIM-CMG: standard FinFET compact model for advanced circuit design," in *Proceedings of the IEEE European Solid-State Circuit Conference (ESSCIRC)*, pp. 196–201, Graz, Austria, 2015.

[77] G. D. J. Smit, A. J. Scholten, G. Curatola, R. van Langevelde, G. Gildenblat, and D. B. M. Klaassen, "PSP-based scalable compact FinFET model," *Nanotech*, vol. 3, pp. 520–525, 2007.

[78] G. Dessai, A. Dey, G. Gildenblat, and G. D. J. Smit, "Symmetric linearization method for double-gate and surrounding-gate MOSFET models," *Solid State Electronics*, vol. 53, no. 5, pp. 548–556, 2009.

[79] P. Magarshack, P. F. Magarshack, and G. Cesana, "UTBB FD-SOI: a process/design symbiosis for breakthrough energy-efficiency," in *Proceedings of the Conference on Design, Automation and Test in Europe (DATE)*, pp. 952–957, Grenoble, France, March 2013.

[80] T. Poiroux, O. Rozeau, P. Scheer et al., "Leti-UTSOI2.1: a compact model for UTBB-FDSOI technologies—part II: DC and AC model description," *IEEE Transactions on Electron Devices*, vol. 62, no. 9, pp. 2760–2768, 2015.

[81] D. Akinwande, J. Liang, S. Chang, Y. Nishi, and H.-S. P. Wong, "Analytical ballistic theory of carbon nanotube transistors: experimental validation, device physics, parameter extraction, and performance projection," *Journal of Applied Physics*, vol. 104, no. 12, Article ID 124514, 2008.

[82] J. Deng and H.-S. P Wong, "A compact SPICE model for carbon-nanotube field-effect transistors including non-idealities and its application—part I: model of the intrinsic channel region," *IEEE Transactions on Electron Devices*, vol. 54, no. 12, pp. 3186–3194, 2007.

[83] J. Deng and H.-S. P. Wong, "A compact SPICE model for carbon-nanotube field-effect transistors including non-idealities and its application—part II: full device model and circuit performance benchmarking," *IEEE Transactions on Electron Devices*, vol. 54, no. 12, pp. 3195–3204, 2007.

[84] C.-S. Lee, E. Pop, A. D. Franklin, W. Haensch, and H.-S. P. Wong, "A compact virtual-source model for carbon nanotube FETs in the sub-10-nm regime—part I: intrinsic elements," *IEEE Transactions on Electron Devices*, vol. 62, no. 9, pp. 3061–3069, 2015.

[85] C.-S. Lee, E. Pop, A. D. Franklin, W. Haensch, and H.-S. P. Wong, "A compact virtual-source model for carbon nanotube field-effect transistors in the sub-10-nm regime—part II: extrinsic elements, performance assessment, and design optimization," *IEEE Transactions on Electron Devices*, vol. 62, no. 9, pp. 3070–3078, 2015.

[86] A. Khakifirooz, O. M. Nayfeh, and D. Antoniadis, "A simple semiempirical short-channel MOSFET current–voltage model continuous across all regions of operation and employing only

physical parameters," *IEEE Transactions on Electron Devices*, vol. 56, no. 8, pp. 1674–1680, 2009.

[87] K. Uchida, K. Matsuzawa, J. Koga, R. Ohba, S.-I. Takagi, and A. Toriumi, "Analytical single-electron transistor (SET) model for design and analysis of realistic SET circuits," *Japanese Journal of Applied Physics*, vol. 39, no. 4B, pp. 2321–2324, 2000.

[88] H. Inokawa and Y. Takahashi, "A compact analytical model for asymmetric single-electron tunneling transistors," *IEEE Transactions on Electron Devices*, vol. 50, no. 2, pp. 455–461, 2003.

[89] C. Le Royer, G. Le Carval, D. Fraboulet, and M. Sanquer, "Accurate modelling of quantum-dot based multi-tunnel junction memory," in *Proceedings of the ESSDERC*, pp. 403–406, Firenze, Italy, September 2002.

[90] G. Lientschnig, I. Weymann, and P. Hadley, "Simulating hybrid circuits of single-electron transistors and field-effect transistors," *Japanese Journal of Applied Physics*, vol. 42, no. 10, pp. 6467–6472, 2003.

[91] S. Mahapatra, V. Vaish, C. Wasshuber, K. Banerjee, and A. M. Ionescu, "Analytical modeling of single electron transistor for hybrid CMOS-SET analog IC design," *IEEE Transactions on Electron Devices*, vol. 51, no. 11, pp. 1772–1782, 2004.

[92] S. Mahapatra and A. M. Ionescu, *Hybrid CMOS Single-Electron-Transistor Device and Circuit Design*, Artech House, Norwood, MA, USA, 2006.