Hindawi Publishing Corporation
VLSI Design
Volume 2011, Article ID 178516, 19 pages
doi:10.1155/2011/178516

Review Article

Advancement in Nanoscale CMOS Device Design En Route to Ultra-Low-Power Applications

Subhra Dhar,[1] Manisha Pattanaik,[1] and Poolla Rajaram[2]

[1] *Department of Information Technology, VLSI Design Laboratory, ABV-Indian Institute of Information Technology and Management, Madhya Pradesh, Gwalior 474010, India*
[2] *School of Studies in Physics, Jiwaji University, Madhya Pradesh, Gwalior 474011, India*

Correspondence should be addressed to Subhra Dhar, sdharmos@gmail.com

Received 9 September 2010; Revised 11 January 2011; Accepted 11 March 2011

Academic Editor: A. G. M. Strollo

Copyright © 2011 Subhra Dhar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, the demand for power sensitive designs has grown significantly due to the fast growth of battery-operated portable applications. As the technology scaling continues unabated, subthreshold device design has gained a lot of attention due to the low-power and ultra-low-power consumption in various applications. Design of low-power high-performance submicron and deep submicron CMOS devices and circuits is a big challenge. Short-channel effect is a major challenge for scaling the gate length down and below 0.1 μm. Detailed review and potential solutions for prolonging CMOS as the leading information technology proposed by various researchers in the past two decades are presented in this paper. This paper attempts to categorize the challenges and solutions for low-power and low-voltage application and thus provides a roadmap for device designers working in the submicron and deep submicron region of CMOS devices separately.

1. Introduction

Metal-Oxide-Semiconductor Field-Effect-Transistor (MOSFET) has been the major device for integrated circuits over the past two decades. With technology advancement and the high scalability of the device structure, silicon MOSFET-based VLSI circuits have continually delivered performance gain and/or cost reduction to semiconductor chips for data processing and memory functions. A lot of research has gone into device design over the last thirty years, but the evolution of process technologies brings new obstacles as well as new opportunities to device designers. Continuous CMOS scaling has been the main driving factor of silicon technology advancement to improve the performance. Design of low-power high-performance CMOS devices and circuits is a big challenge. The process parameters in low power design are channel length, oxide thickness, threshold voltage, and doping concentration in the channel. As the technology is scaled down, process parameter variations have become severe problem for low-power design. The low-power design technique should be such that it is less sensitive to the process parameter variations. As technology scales down, the variations of these process parameters are expected to be significant in future generations. As a result, the yield of the circuit will be less. The variation of leakage power and delay in the transistors on a given die are different for different low-power design techniques. The role of threshold voltage (V_{th}) and subthreshold swing (S) has become increasingly important with VLSI applications emphasizing on low-voltage, low-power, and high-speed design. Short-channel effect is a major challenge for scaling the gate length down and below 0.1 μm. The dependence of V_{th} on channel length is stronger as compared to other factors that also cause V_{th} fluctuation at small device dimension such as random dopant distribution [1]. The gate oxide thickness in recent process technologies has approached the limit when direct tunneling starts to play a significant role in both off-state and on-state MOSFET transistor operation modes. This phenomenon, in addition to subthreshold leakage, results in a dramatic total standby leakage power dissipation. Thus, better design strategies to control the total leakage power are necessary. It is well known that gate tunneling currents are highly sensitive to the voltage variation across the gate oxide. Supply voltage attenuation can give significant reduction in

gate leakage power consumption. Circuit design techniques to mitigate the impact of gate leakage would be much less efficient than the use of high-k material since gate leakage is a stronger function of process-induced oxide thickness fluctuation as compared to change in V_{dd} and threshold voltage. In addition to the gate-oxide scaling issue, higher doping concentrations would degrade subthreshold swing (S). Furthermore, V_{dd} scaling necessitates threshold voltage (V_{th}) reduction, which exponentially increases I_{OFF}. I_{OFF} reduction is critical where chips are often in standby mode of operation, and even during active operation, acceptable I_{OFF} is required since leakage power consumption is rapidly increasing at a much faster rate compared to dynamic power [2]. Few front ends of line challenges faced by transistors in the deep submicron and ultradeep submicron region include diminishing $V_{gs} - V_{th}$, larger V_{th}/V_{dd}, thin junctions drive dopant levels to saturation, dopant loss and statistical dopant fluctuation on small geometry devices increase device variability, and increase in channel doping concentration to control DIBL reduces carrier mobility while increasing body effect, gate oxide scaling slowing as it approaches the monolayer thickness of SiO_2[3]. This paper is divided into three main sections. Section 2 discusses the CMOS scaling trends facing the challenges due to the feature size, supply voltage, threshold voltage, and oxide thickness. Section 3 focuses on the limitations of technology scaling, and therefore the potential solutions to be sought to face the above challenges are dealt with in Section 4. Section 5 gives the complete roadmap to device designers working below 0.1 micron region in tabular form and thus concludes the paper.

2. CMOS Scaling Trends

Device scaling is based upon simple principles; by reducing the sizes of devices and interconnects, density, and power, the speed and performance of transistors can be improved. This paper mainly focuses on the

(i) scaling of threshold voltage with feature size,

(ii) scaling of gate oxide thickness with feature size,

(iii) scaling of supply voltage with feature size.

With technology scaling, the MOS device channel length is reduced. When the dimensions of a MOSFET are scaled down, both the voltage level and the gate-oxide thickness are also reduced. The supply voltage V_{dd} has been scaled down in order to keep the power consumption under control. The transistor V_{th} had to be scaled to maintain a high drive current and achieve performance improvement. In a given technology generation, since the source-body and drain-body depletion widths are predefined based on the dopings, the rate at which the barrier height increases as a function of distance from the source into the channel is constant. Therefore, when the channel length is reduced, the barrier for the majority carriers to enter the channel also is reduced as indicated in Figure 1. This results in reduced threshold voltage. In short channel transistor, the barrier height and therefore the threshold voltage are a strong function of the

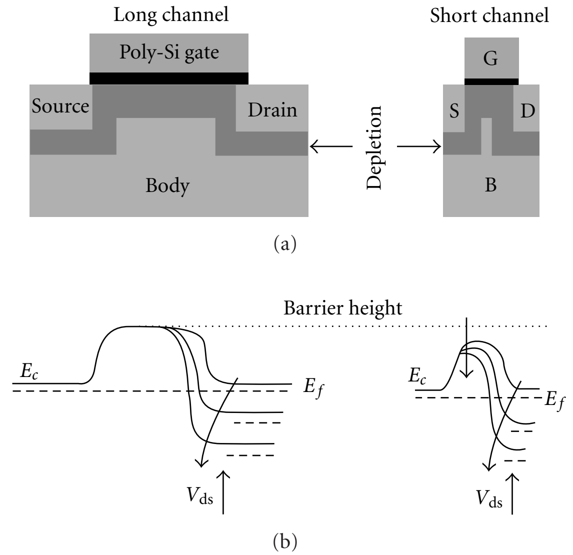


Figure 1: Channel length reduces the barrier for the majority of carriers to enter the channel [4].

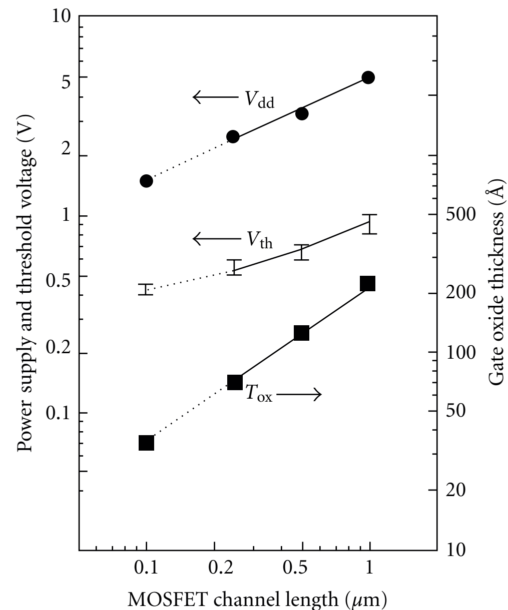


Figure 2: Scaling trend of power supply voltage (V_{dd}), threshold voltage (V_{th}), and gate oxide thickness (T_{ox}) as a function of CMOS channel length [5].

drain voltage [4]. As Figure 1 indicates, the barrier reduces as the drain voltage is increased.

The gate material has long been polysilicon with silicon dioxide as the insulator between the gate and the channel; aggressive scaling of CMOS technology in recent years has reduced the silicon dioxide (SiO_2) gate dielectric thickness below 20 Å (see Figure 3) [6]. In 90 nm, the gate oxide consists of about 5 atomic layers equivalent to 1.2 nm in thickness. The thinner the gate oxide, the higher the transistor current and consequently the switching speed.

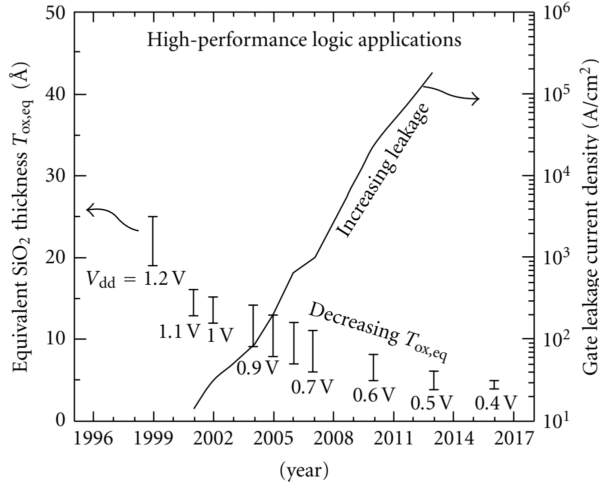


Figure 3: Scaling of CMOS technology versus gate dielectric thickness below 20 Å [6].

The sustained scaling process in the subnanometer regime (<25 nm) requires ultimately scaling down of the gate oxide thickness and increasing channel/body doping densities to overcome severe short-channel effects [2].

The scaling trend of power supply voltage (V_{dd}), threshold voltage (V_{th}), and gate-oxide thickness (T_{ox}) as a function of CMOS channel length [5] is shown in Figure 2. When V_{dd} is reduced towards shorter channel lengths, it becomes increasingly difficult to satisfy both the performance and the off current requirements. Tradeoff between leakage current and circuit speed stems due to subthreshold nonscalability. For this reason and for compatibility with the standardized power supply voltage of earlier generation systems, the general trend is that V_{dd} has not been scaled down in proportion to L and V_{th} has not been scaled down in proportion to V_{dd} as is shown in Figure 2. Aggressive scaling of CMOS technology in recent years has reduced the silicon dioxide gate dielectric thickness below 20 Å.

The effect of important factors like the gate oxide thickness and the gate leakage on CMOS scaling is quantified and is shown in Figure 3.

In Table 1, scaling relations are shown along with the scaling behavior of some of the other important physical parameters.

Scaling trends in [7] highlight that new device structures need to be judged on parasitic resistance and capacitance more than on channel transport properties. The sudden rise in parasitics at the end of the roadmap can be qualitatively understood as resulting from the space between neighboring devices decreasing to tens of nanometers since the source/drain and contact size need to be aggressively scaled to support the increased density in the absence of gate-length scaling. Figure 4 depicts the typical design rules of planar Si transistors for the 32-nm technology node. It can be seen that the source/drain contact and the gate are only tens of nanometers apart, which is undesirable in terms of parasitic resistance. Such small contact size leads to higher contact resistance and contact-to-gate capacitance.

Table 1: Technology scaling rules for three cases [8].

Physical parameters	Constant electric field scaling factor	Generalized scaling factor	Generalized selective scaling factor
Channel length, insulator thickness	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Wiring width, channel width	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Electric field in device	1	ε	ε
Voltage	$1/\alpha$	ε/α	ε/α_d
On-current per device	$1/\alpha$	ε/α	ε/α_d
Doping	α	$\varepsilon\alpha$	$\varepsilon\alpha_d$
Area	$1/\alpha^2$	$1/\alpha^2$	$1/\alpha_w^2$
Capacitance	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Gate delay	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Power dissipation	$1/\alpha^2$	ε^2/α^2	$\varepsilon^2/\alpha_w\alpha_d$
Power density	1	ε^2	$\varepsilon^2\alpha_w/\alpha_d$

The parasitic resistance and capacitance are now becoming comparable and are on course to become even larger than the intrinsic device resistance and capacitance.

3. Challenges in the Submicron and Deep Submicron Region

To understand the behavior of sub-100-nm devices, numerous of studies have been conducted in nanoscale MOSFETs. The main challenges encountered by bulk MOSFETS in the submicron and deep submicron region for low-power, low-voltage applications are the short-channel effects, gate-induced drain leakage, threshold voltage roll-off, DIBL, hot carrier effects, polydepletion effects, BTBT, and so forth. These are discussed in detail in next subsections.

3.1. Gate-Leakage Current. The increased gate-leakage current is a well-recognized challenge to continued MOSFET scaling. It is the current flowing into the gate of the transistor also called the tunneling current. With current process technology parameters, gate leakage has increased to more than double the subthreshold current and will continue to increase at a much higher rate mandating the use of high-k materials other than silicon dioxide to enable the use of thicker oxide thicknesses [10]. Control of gate-leakage current is paramount in low-power CMOS circuit design. The evaluation of gate current due to tunneling through the thin-gate oxide between polysilicon gate and substrate has attracted special attention, since it represents one of the

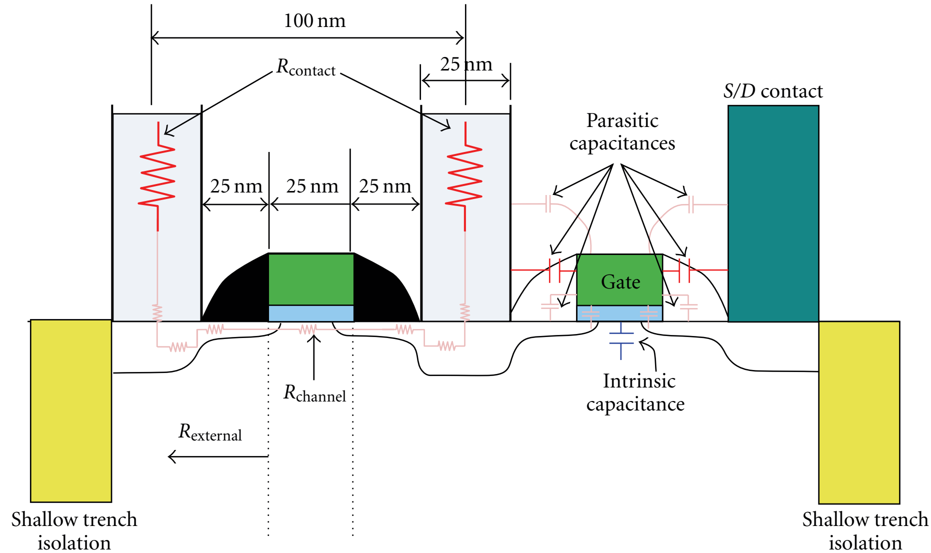


Figure 4: Design rules of planar Si transistors for the 32 nm technology node [7].

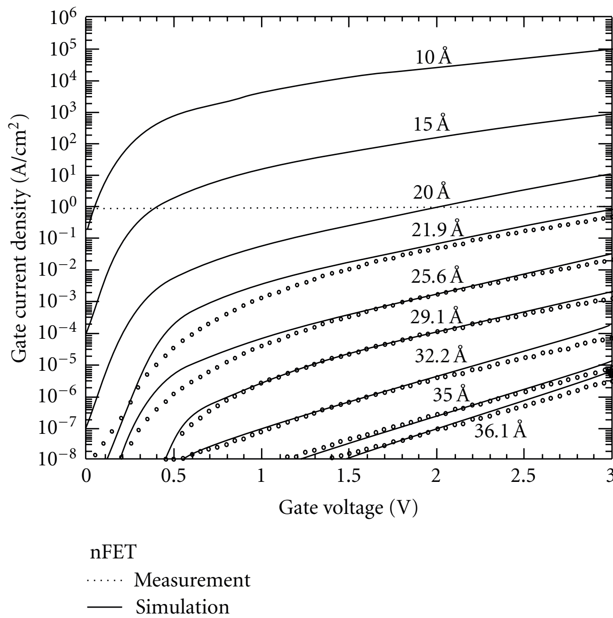


Figure 5: Tunneling currents for oxide thicknesses ranging from 3.6 to 1.0 nm versus gate voltage [8].

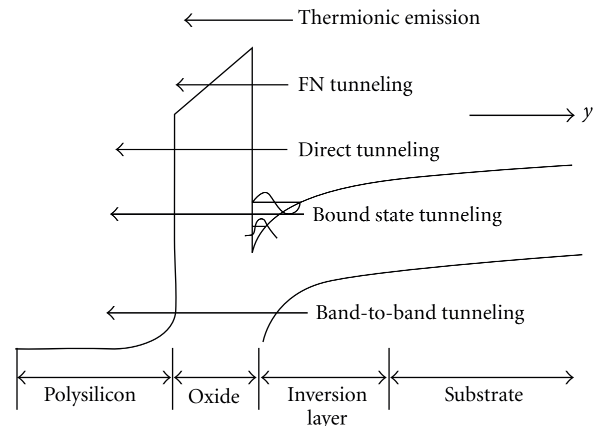


Figure 6: Gate currents components between polysilicon gate and semiconductor substrate in nanoscale MOSFET [9].

major leakage current components in nanoscale MOSFETs. The gate leakage current increases exponentially as the oxide thickness is scaled down. Gate currents components between polysilicon gate and semiconductor substrate in nanoscale MOSFETS are shown in Figure 6. Tunneling currents for oxide thicknesses ranging from 3.6 to 1.0 nm are plotted in Figure 5.

3.2. Gate Oxide Thickness. For CMOS devices with channel lengths of 100 nm or less, an oxide thickness of ∼3 nm is needed. This thickness comprises only a few layers of atoms and is approaching fundamental limits. Scaling which reduces gate length also reduces the dielectric thickness, but the continued thickness reduction of conventional oxides results in reliability degradation and unacceptable current leakage. In a nanoscale device where SCE is extremely severe, an increase in the oxide thickness will increase the subthreshold leakage. To keep adverse 2D electrostatic effects on threshold voltage (i.e., short-channel effects) under control, gate-oxide thickness is reduced nearly in proportion to channel length. Researchers in [5] have shown the gate oxide scaling to a thickness of only a few atomic layers, where quantum mechanical tunneling gives rise to a sharp increase in gate leakage currents. Major causes for concern in further reduction of the SiO_2 thickness include increased polysilicon gate depletion, gate dopant penetration into the channel region,

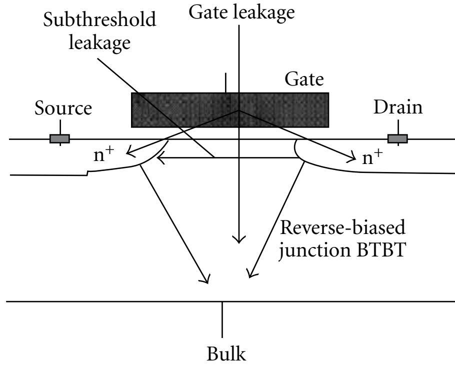


Figure 7: Static CMOS leakage sources [12].

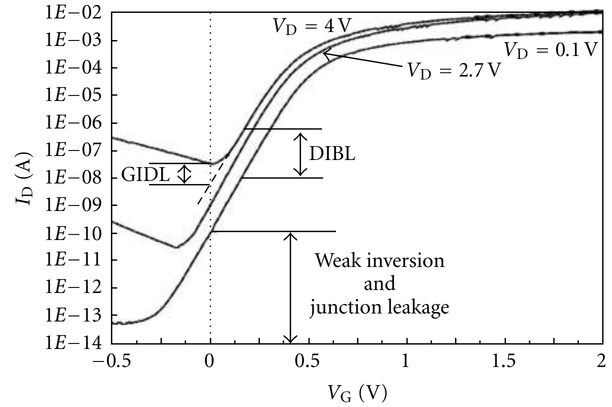


Figure 8: n-channel I_D versus V_G showing DIBL, GIDL, weak inversion, and p-n junction reverse-bias leakage components [13].

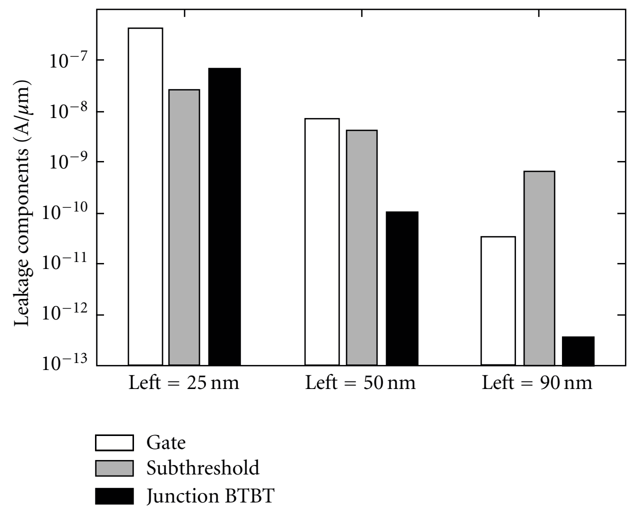


Figure 9: Contribution of different leakage components in NMOS devices at different technology generations [14].

and high direct-tunneling gate-leakage current, which leads to questions regarding dielectric integrity, reliability, and stand-by power consumption.

Devices with thick gate oxides have total gate capacitance approximately equal to the oxide capacitance. The situation changes for MOSFETs with very thin gate oxides (less than 10 nm thick). Since the device scaling rules demand thinner gate oxides, the degradation of the total gate capacitance, which defines MOSFET's transconductance, is expected to have very important consequences on their performance. The degradation of the total gate capacitance due to finite inversion layer capacitance is significant in scaled silicon MOSFET's [5]. C-V data of 15 Å oxides show about 40% less capacitance at inversion than that of the physical oxide; the currents are degraded by only 10–20%.

3.3. Gate-Induced Drain Leakage (GIDL). GIDL at the NMOS gate drain edge is important at low current levels and high applied voltages. The electron thermal voltage, kT/q, is a constant for room temperature electronics; the ratio between the operating voltage and the thermal voltage shrinks. This leads to higher source-to-drain leakage currents stemming from the thermal diffusion of electrons. Thinner oxide thickness and higher V_{dd} enhance the electric field increasing GIDL [11]. At low drain doping values, the electric field not being high enough causes tunneling, whereas for very high drain doping, the depletion width and tunneling will be limited causing less GIDL.

3.4. Off-State Leakage Current. Static CMOS Leakage sources are shown in Figure 7. Other sources of leakage currents including DIBL, GIDL, weak inversion, and p-n junction reverse bias leakage components [8] are shown in Figure 8. The largest contributor of leakage currents in the submicron region is the off-state drain to source leakage (I_{OFF}).

Contribution of different leakage components in NMOS devices at different technology generations is shown in Figure 9. Magnitude of the leakage components and their relative dominance on each other depends strongly on device geometry and doping profile [15].

It can be observed that, for the 90-nm device, the major leakage component is the subthreshold leakage, but in the scaled devices, contributions of the junction leakage and the gate leakage have significantly increased. Variation of different leakage components with technology generation, oxide thickness, and doping profile are shown in [14].

The off-state current I_{OFF} is also increased by drain-induced barrier lowering (DIBL). As the channel length becomes shorter, both channel length and drain voltage reduce this barrier height. This two-dimensional effect makes the barrier height be modulated by channel length variation resulting in threshold voltage variation as shown in Figure 10.

Transistor I_{OFF} is related to V_{th} which in turn is modulated by DIBL. Drain-induced barrier lowering (DIBL) reduces threshold voltage for short channel devices and increases threshold voltage roll-off. For short channel devices, channel length variation (ΔL) translates to threshold voltage variation (ΔV_{th}) [4]. DIBL occurs when the depletion region of the drain interacts with the source near the channel

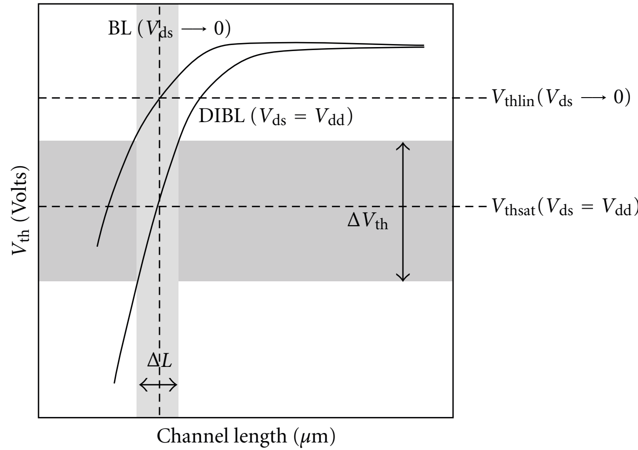


Figure 10: Barrier lowering (BL) resulting in threshold voltage roll-off with channel length reduction [4].

surface to lower the source potential barrier. It happens when a high drain voltage is applied to a short-channel device, lowering the barrier height and resulting in further decrease of the threshold voltage. The source then injects carriers into the channel surface without the gate playing a role. DIBL is enhanced at a higher drain voltage and shorter L_{eff}. Surface DIBL typically happens before deep bulk punch-through. Ideally, DIBL does not change the slope S, but it does lower V_{th}. For a device with zero gate voltage and drain potential V_{dd}, significant band bending in the drain allows electron-hole pair generation. A deep depletion condition is created since the holes are rapidly swept out. This leakage mechanism is exacerbated by high source or drain-to-body voltages as well as high drain to gate voltages.

3.5. Band-to-Band Tunneling Currents. Transistor scaling has led to increasingly steep halo (or pocket) implants, where the substrate doping at the junction interfaces is increased while the channel doping is low. This allows punch-through and DIBL control with less impact on channel mobility. The resulting steep doping profile at the drain edge increases band-to-band tunneling currents there, particularly as V_{tb} is increased. Diode area leakage components from both the source drain diodes and the well diodes are generally negligible with respect to I_{OFF} junction band to band tunneling and GIDL components.

3.6. Threshold Voltage. Threshold Voltage is one of the most important parameters in technology and circuit design. Scaled transistors reduce the threshold voltage V_{th} to maintain performance by maintaining $V_{gs} - V_{th}$, gate overdrive, as V_{dd} is lowered. This increases the I_{OFF} exponentially as the subthreshold slope S is essentially fixed. Unlike the strong inversion region in which the drift current dominates, subthreshold conduction is dominated by the diffusion current. The carriers move by diffusion along the surface similar to charge transport across the base of bipolar transistors. The exponential relation between the driving voltage on the gate and the drain current is a straight line in a semilog plot.

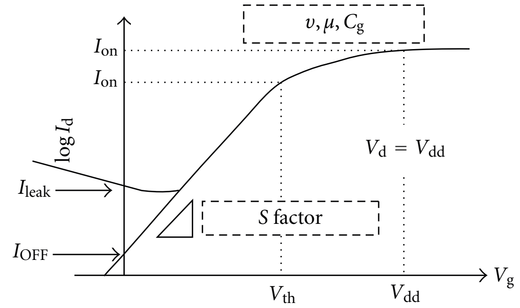


Figure 11: Schematic $\log I_{ON}$–V_g characteristics to show the factor affecting power consumption [4].

Weak inversion typically dominates modern device off-state leakage due to the low V_{th} used.

The subthreshold slope indicates how effectively the flow of the drain current of a device can be stopped when V_{gs} is decreased below V_{th}. Figure 11 shows schematic $\log I_{ON}$–V_g characteristics to show the factor affecting power consumption. As the device dimensions and the supply voltage are being scaled down to enhance performance, power efficiency, and reliability, this characteristic becomes a limitation on how limited a power supply can be used. The parameter S is measured in millivolts per decade. For the limiting case of $T_{ox} \sim 0$ and at room temperature, $S \sim 60$ mV/decade. Typical S values for a bulk CMOS process can range from 80 mV/decade to 120 mV/decade or more. A low value for subthreshold slope is most desirable. As noted in Section 3, threshold voltage is one of the most important parameters in technology and circuit design. Scaling of transistor threshold voltage is associated with exponential increase in subthreshold leakage current, along with other negative impacts such as increased DIBL, V_{th} roll-off, reduced on-off current ratio, and increased source drain résistance. Figures 12 and 13 show the relation between scaling and switching power, threshold voltage, and off current.

Figures 14(a) and 14(b) show the experimental threshold voltage measured on real bulk devices that were fabricated using technological dimensions issued from CMOS 90- and 65-nm technologies, respectively. The main features are $T_{ox} = 1.6/1.2$ nm, $L_{nominal} = 45/70$ nm, and $V_d = 1/1$ V for the devices based on the 90-nm/65-nm technology, respectively [15]. Figure 14(c) shows the on and off-state currents for the 90 nm and 65-nm technology.

3.7. System Performance and Power Considerations. The delay of a CMOS gate may be expressed as

$$T_p = \frac{k_d C_L V_{dd}}{I_{ON}}, \quad (1)$$

where C_L is the load capacitance and k_d is a fitting parameter. The delay expression is clearly dominated by an exponential dependence on V_{dd} and V_{th}. The simulated delay of a CMOS inverter is shown in Figure 15 at nominal V_{dd} and at 250 mV. The delay at nominal V_{dd} improves with L_{poly}, through a rate that is slower than the target of 30% per generation under generalized scaling. With the exception of the 32-nm

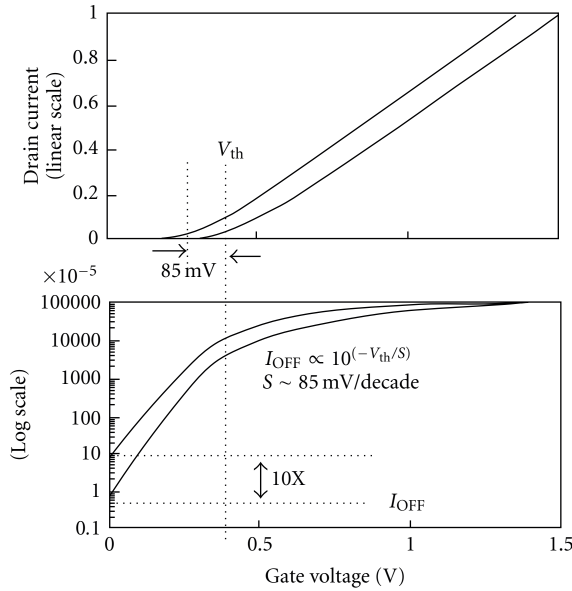


Figure 12: Relationship between threshold voltage (V_{th}) and subthreshold leakage current (I_{OFF}) [4].

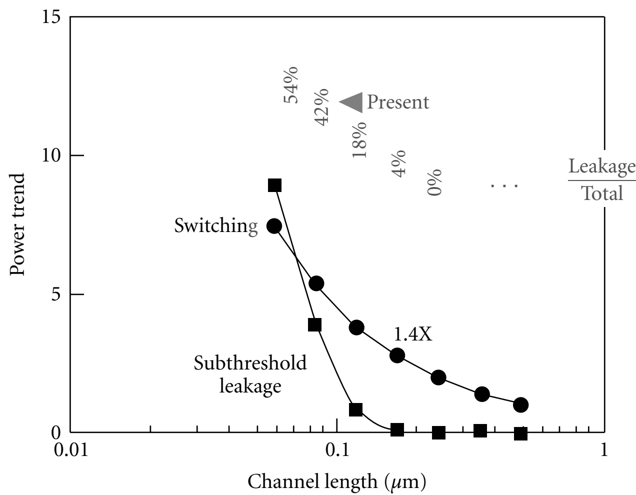
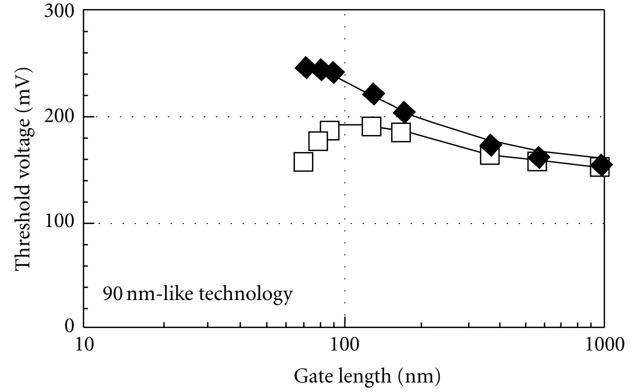
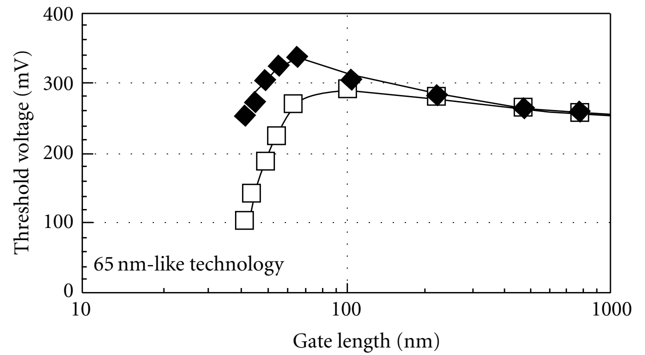


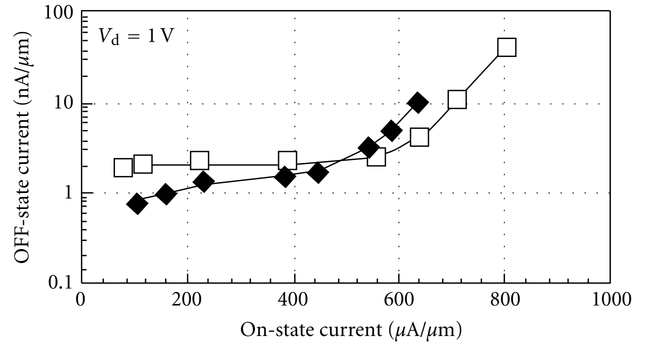
Figure 13: Trend in subthreshold leakage and switching power with technology scaling [4].



(a)



(b)



(c)

Figure 14: (a)–(c) Threshold voltage measured on real bulk devices CMOS 90- and 65-nm technologies, I_{ON}–I_{OFF}, respectively [15].

device, the delay actually increases with device scaling at $V_{dd} = 250$ mV due to strict leakage constraints during device optimization as well as degraded S [16]. CMOS delay degradation due to polydepletion is studied in [17].

Extrapolating current power trends, the off-state power consumption would be about equal to active power around the 10-nm node. As reported in [18], off-state power needs to be less than 10% to 20% of total power consumed, and current 90-nm technology is already approaching this limit. For low-power applications, meeting challenges in the subnanometer range is to meet performance and leakage requirements for highly scaled MOSFETS. For the 10-nm node, for low-standby-power (LSTP) applications at 25°C = 10 pA/μm, nominal saturation current drive = 700 μA/μM is predicted according to [19]. For low-operating-power (LOP) applications, I_{leak} = 10000 pA/μm and I_{dd} = 900 μA/μM; for high-performance (HP) applications, I_{leak} = 10 μA/μM and I_{dd} = 1500 μA/μM. For low-power transistors, the gate length lags behind the high-performance transistor gate length by

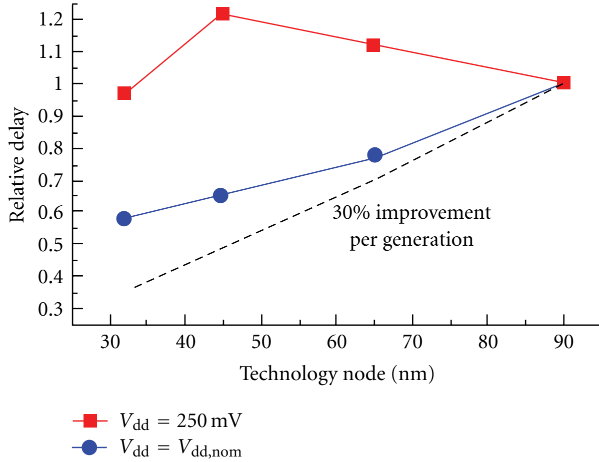


Figure 15: Simulated delay of a CMOS inverter is shown at nominal V_{dd} and at 250 mV [16].

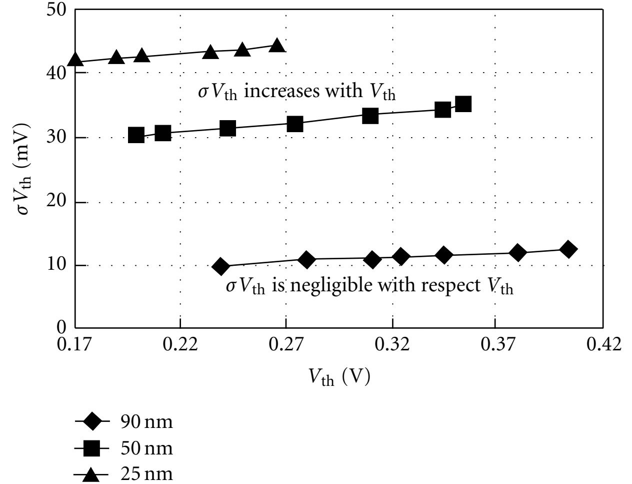


Figure 16: σV_{th} due to random dopant fluctutation versus V_{th} [20].

two years reflecting historical trends and the need for very low leakage current in mobile applications.

3.8. Statistical Variations. Controlling systematic and random variation in device parameters during fabrication is becoming a great challenge for scaled technologies. The delay and leakage currents in a device depend on the transistor geometry (gate length, oxide thickness, width, the doping profile, "halo" doping concentration, etc.), the flat-band voltage, and the supply voltage. Any statistical variation in each of these parameters results in a large variation in different leakage components and significant spread in delay. Among the statistical variations, the random placement of dopants is of great concern because it is independent of transistor spatial location and causes threshold voltage mismatch between transistors even though they may be close to each other (intradie variation) resulting in significant leakage and delay variation of logic gates and circuits. Hence, any low leakage design needs to consider the spread of leakage and delay, both at circuit and device design phase, to minimize overall leakage, while maintaining yield with respect to a target delay under process variation. Figure 16 shows that σV_{th} is negligible with respect to nominal V_{th} in 90-nm devices but becomes significant in 50-nm and 25-nm devices resulting in considerable spread in drive current and leakage power.

3.9. Bulk Impurity Concentration. As MOSFET's scale to shorter channel lengths, channel doping densities increase in order to suppress undesirable short-channel effects such as punchthrough and drain-induced barrier lowering. The doping concentration is raised to values of 1×10^{18} cm^{-3} as the channel length is scaled to 0.1 m and below, such high bulk-impurity concentrations lead to an increase of threshold voltage and to an electron mobility decrease, which strongly degrades the electric properties of the device [8].

The above section concludes that aggressive scaling of the devices not only leads to increased subthreshold leakage but also has other negative impacts such as increased drain-induced barrier lowering (DIBL), V_{th} roll-off, reduced ON-current to off current ratio, and increased source drain resistance. Limitations of scaling also include electronic thermal energy, quantum mechanical tunneling, standby power-active power increases, and slow performance gain. Thus, summarizing the challenges faced by submicron and deep submicron devices, it is clear that the three key factors limiting continued scaling in CMOS are

(i) minimum dimensions that can be fabricated,

(ii) diminishing returns in switching performance,

(iii) off-state leakage.

4. Detailed Solutions for Achieving Low-Power and Low-Voltage Applications in Submicron and Deep Submicron Region

Subthreshold operation is an effective way to achieve ultralow power for systems which do not have high-performance requirement. Lowering operating voltage is useful to lower not only active power, but also leakage. Scaling V_{dd} slows a circuit down since the gate overdrive $V_{gs} - V_{th}$ is reduced. To deal with this, [22] suggests using dynamic voltage scaling for systems to allow the lowest V_{dd} necessary to meet performance requirements. The subthreshold slope S can be made smaller by using a thinner oxide (insulator) layer to reduce T_{ox} or a lower substrate doping concentration (resulting in larger W_{dm}). Changes in operating conditions, namely, lower temperature or a substrate bias, also cause S to decrease [13]. To avoid the short-channel effects, oxide thickness scaling and higher and nonuniform doping ("halo" and "retrograde well") need to be incorporated as the devices are scaled to the nanometer regime. Solutions suggested by various researchers include optimizing doping profile, pushing Si depletion width to tunneling limit, and scaling till 20-nm gate length with non-scaled gate oxides and voltage levels; cooling to low temperature can provide additional design space needed to extend CMOS devices to 10 nm for

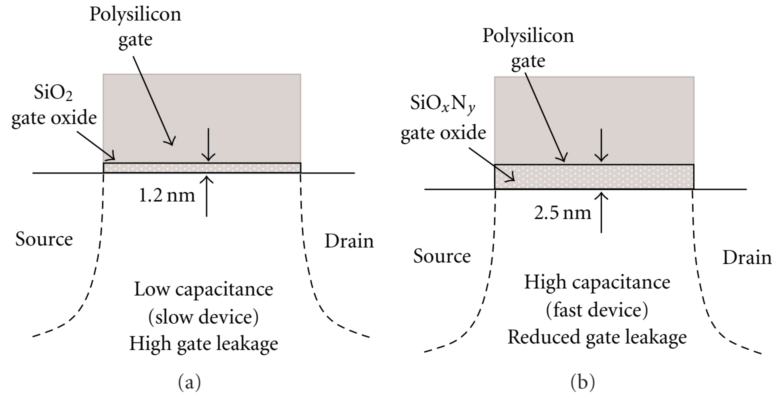


Figure 17: SiO_xN_y gate oxide of high dielectric constant proves to be of much better use [21].

server applications. The following subsections are divided into solutions feasible for few important aspects of low-power operation in nanoscaled technologies. Most of the solutions presented here are based on numerical simulation.

4.1. High-k Dielectrics to Combat Gate Tunneling Currents. Gate-leakage reduction is the key motivation for the replacement of SiO_2 with alternative gate dielectrics. Higher k value material than the silica (SiO_2) may be a solution as proposed in [18]. This would allow the actual thickness of the gate dielectric to be increased while still maintaining the same electric field in the channel.

For high-performance and low-operating-power logic applications, it is predicted that Si_3N_4 or SiO_xN_y will be usable through 2016. The need for high-k gate dielectrics such as HfO_2 is more urgent for low-standby-power applications compared to high-performance logic applications. As the dielectric thickness indirectly controls the gate length, the effective gate length needs to be 40 times the dielectric thickness to properly control short channel effects (SCE). Thus, new materials and material modifications including low-resistivity conductors, strained Si, and low-k dielectrics are suggested to improve performance and overcome short channel effects. Promising results have been reported on the development of low-standby-power CMOS technology using HfO_2 gate dielectric. Transistors with gate length down to 55 nm, HfO_2 gate dielectric with electrical oxide thickness down to 15 Å, and off-state leakage current of 25 pA/μm were demonstrated in [10]. As a result, there is immense interest in alternative gate dielectrics with higher relative permittivities. For a given equivalent SiO_2 thickness, using a high-k gate dielectric factor compared with SiO_2 achieves significant suppression of the direct-tunneling gate current and therefore results in considerable reduction in power consumption. Gate dielectrics such as silicon nitride (Si_3N_4) or oxynitride (SiO_xN_y), zirconium oxide (ZrO_2), hafnium oxide (HfO_2), aluminum oxide (Al_2O_3), and lanthanum oxide (La_2O_3) have been thoroughly investigated. For physical gate length of 25 nm, supply voltage is 0.7 V, and the range of $T_{ox,eq}$ is 6–11 Å. At this technology node, the maximum tolerable gate dielectric leakage per unit device width for high-performance logic applications is 1 μA/μm. Transistors with 30-nm gate length and a 7 Å SiO_xN_y gate dielectric have been demonstrated with excellent electrical characteristics, suggesting that an SiO_xN_y gate dielectric can be used at least down to 7 Å for high-performance applications. Figure 17 compares the conventional SiO_2 gate oxide with SiO_xN_y gate oxide for enhanced performance. If higher leakage-current density can be tolerated, the scaling limit of an SiO_xN_y gate dielectric can be pushed well below 7 Å.

4.2. Strained Si Helps to Improve Electron and Hole Mobility. Strained Si is the process of introducing physical strain on the silicon lattice. Introducing lattice strain into the Si channel alters the band structure and addresses Si transport deficiencies compared with other high-mobility III–V semiconductors. Strain is used to create both a low conductivity mass in the channel direction and a high density of states by creating a very large conductivity mass in the plane of the transistor perpendicular to the channel direction [7]. Strain improves transistor performance by enhancing the channel mobility through reduced electron effective mass and intervalley scattering rate for NMOS and reduced hole effective mass and interband scattering rate for PMOS. To continue transistor I_{dsat} improvement, efforts are being poured into strained Si channel transistors; 10 to 20% improvements have been reported using SiGe strained silicon [3]. Results from [23] show that the PMOS mobility gain induced by the biaxial compressive strain in SiGe increases with increasing Ge% in the SiGe layer and that the mobility gain does not reduce at high transverse channel electric fields as in the case of the biaxial tensile strain in Si.

4.3. High-k Gate Stack Structures to Reduce Gate Current. Replacing silicon dioxide layer with high-k gate stack structures for nanoscale MOSFETs has been used to reduce gate current. Incorporation of nitrogen at the dielectric interface suppresses dopant diffusion from gate polysilicon into the channel, which can cause a shift of threshold voltage. Incorporation of nitrogen into hafnium silicate (Hf/SiO), hafnium aluminate (Hf/AlO), and HfO_2 greatly enhances

the dielectric constant of silicates, suppresses dopant diffusion from gate polysilicon into the channel during high-temperature annealing process, and increases crystallization temperature of the high-k stacks [9].

4.4. High-k/Metal Gate Technology to Reduce the Intrinsic Parameter Fluctuations. The high-k/metal-gate technology is the key to reduce the intrinsic-parameter fluctuations. The use of metal as a gate material introduces a new source of random variation due to the dependence of work function on the orientation of metal grains. Influences of the intrinsic parameter fluctuations and the metal gate work-function fluctuation (WKF), process-variation effect (PVE), and random dopant fluctuation (RDF) on 16-nm-gate bulk MOSFETs' *dc*/*ac* and circuits' timing/power/high frequency characteristics are reported in [24]. The vast study assesses the fluctuations on digital circuit performance and reliability, which may benefit CMOS design and technology in the sub-22-nm era. Intrinsic parameter fluctuations were generated in 16-nm-gate bulk MOSFETs with amorphous-based Tin/HfSiON gate stacks with an EOT of 1.2-nm pMOSFETs with nMOSFETs and pMOSFETs randomly selected and used to study the circuit characteristic fluctuations. The WKF has shown its increasing importance in nanoscale transistors, particularly for pMOSFETs' characteristics. This work shows that the WKF and RDF dominate the device threshold-voltage fluctuation and therefore rule the delay time of the explored digital inverter circuits.

4.5. Optimizing Doping Profile to Improve Short-Channel Behavior. Aggressive scaling of CMOS technology below 0.1 μm gate pitch has stringent requirements on standby power consumption and parametric yields. To control the short channel effects for higher parametric yield, substrate solutions such as the super halo doping and SOI structures become strong options in device design.

4.5.1. Stepped Doping Profile. The use of non uniform doping concentrations has been methodically investigated to increase bulk impurity concentration through boron implant, indium channel implant, silicon epitaxial growth on heavily doped substrates, and delta-doped MOSFET's. A doping profile, which provides low channel doping and a high doping concentration just below the inversion zone, improves short-channel behavior of the device while decreasing the threshold voltage and raising carrier mobility. The use of a low-doped ultrathin epitaxial layer over a heavily doped substrate can also achieve similar result. With this stepped doping profile, the low-doped zone would provide a reduction in threshold voltage, which is essential in order to realize low-power CMOS. A stepped doping profile, with the appropriate choice of the low-doped zone thickness, enhances the electron mobility and reduces the threshold voltage while maintaining good electric behavior versus short-channel effects. The higher mobility and the lower threshold voltage provided by the low-doped zone near the interface greatly improve the electric behavior of submicrometer devices. By adding a low-doped zone of

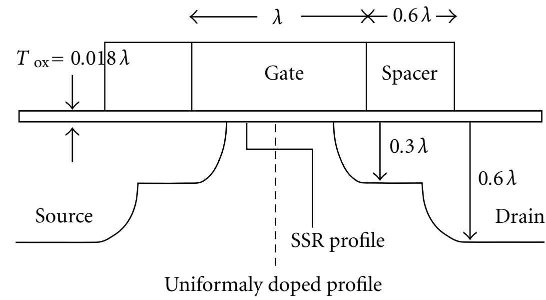


Figure 18: SSR profile with a lightly doped layer of thickness X_j where X_j is the extension depth [26].

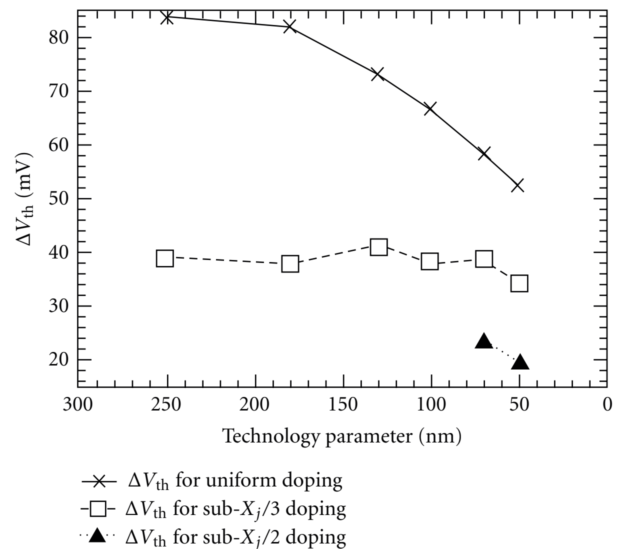


Figure 19: V_{th} roll-off quantified by change in V_{th} between L_- and L_+ devices with drain biased at V_{dd} [26].

convenient thickness next to the interface over a high-doping substrate, both the electron mobility and the threshold voltage of the device are improved while avoiding short channel effects [25].

4.5.2. Super Steep Retrograde (SSR) Profile to Obtain Higher Threshold Voltage. Super steep retrograde doping is just a case of retrograde channel profile where the transition from the lightly doped surface to the heavily doped substrate is sharp. A retrograde doping profile which reduces the channel depletion width may give a more optimal tradeoff between short-channel characteristics and body effect. L_+ and L_- refer to devices with gate lengths 0.2 λ greater and 0.2 λ less than the nominal length, where λ is the technology parameter represented by the nominal length of the polygate. Figure 18 shows SSR profile with a lightly doped layer of thickness X_j where X_j is the extension depth; the technology parameter is the nominal gate length λ.

The V_{th} roll off quantified by change in V_{th} between L_- and L_+ devices with drain biased at V_{dd} is shown in Figure 19; the results are those arrived due to numerical simulation. The main effect of SSR MOSFETs is to achieve low V_{th} without

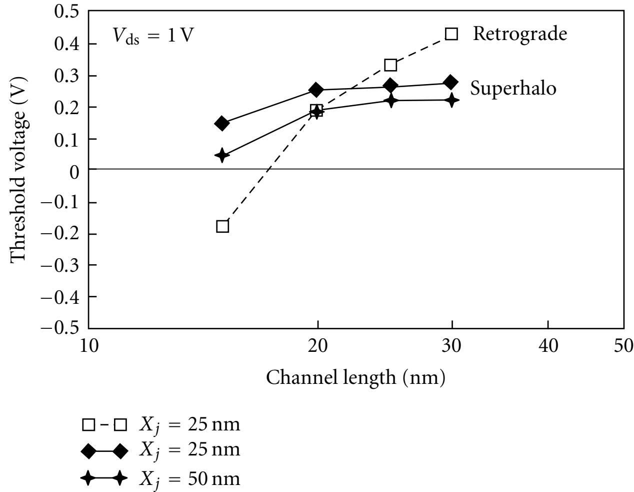


Figure 20: The superior short-channel effect obtained with the superhalo compared with a nonhalo retrograde profile [8].

worsening V_{th} roll-off and hence better short channel characteristics than uniformly doped MOSFETs [26].

4.5.3. Superhalo Doping. With the nonscaled gate oxide and supply voltage, an optimized vertically and laterally non uniform doping profile called the superhalo is suggested for controlling short-channel effects. Such a profile can be realized by ion implantation self-aligned to the gate edges with very restricted amount of diffusion. One of the possible ways suggested is to create sub-50-nm bulk CMOS devices through the use of super halo doping. The highly non uniform profile sets up a higher effective doping concentration toward shorter devices, which counteracts short-channel effects. This results in off currents insensitive to channel length variations and allows CMOS scaling to the shortest channel length possible [8].

The superior short-channel effect obtained with the superhalo compared with a nonhalo retrograde profile is shown in Figure 20. The results obtained are due to device simulation. I_{OFF} is nearly independent of channel length variations between 20 and 30 nm. Because of the nearly flat dependence on channel length, superhalo allows a nominal device to operate at a lower threshold voltage, thereby gaining significant performance benefit of 30–40% over nonhalo devices for 25-nm CMOS at 1.0 V. Higher surface and channel doping and shallow source/drain junction depths reduce the DIBL effect on the subthreshold leakage current. While high DIBL effects override body bias at high V_{ds}, at low V_{ds} it can still be useful. It should be pointed out that DIBL, which is still present in superhalo devices, has only a minor effect on the delay performance for a given high-drain bias.

Since the channel doping mainly originates from tilted implantation through self-aligned source/drain extension, an additional mask can create selective source or drain doping. Doping implantation needs to be optimized to adjust the channel implant profiles to minimize the band to band

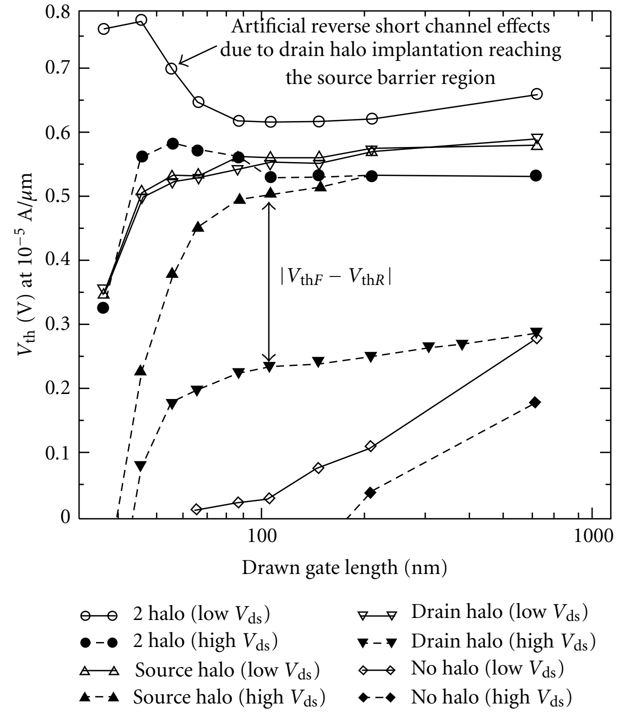


Figure 21: SCEs in the super halo bulk CMOS devices [27].

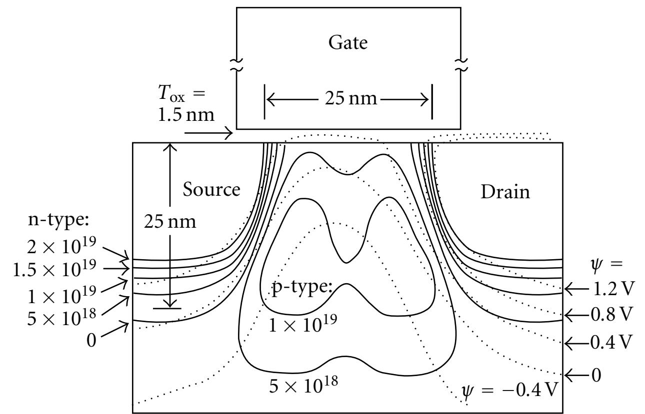


Figure 22: Source, drain, and superhalo doping contours in a 25-nm nMOSFET design [8].

tunneling while maintaining tight control on SCE. Figure 21 shows simulated results of SCEs in the super halo bulk CMOS devices. One-sided super halo design can achieve acceptable SCE behavior below 70-nm drawn gate length [27]. Figure 22 shows a doping profile for 25-nm MOSFET. This profile is realized by ion implantation self aligned to the gate edges with very restricted amount of diffusion.

4.5.4. Abrupt Doping. In the design of 25-nm CMOS devices, [28] predicted good short channel effects (SCE) for gate length $L \leq 25$ nm when the net doping concentration decreases laterally with a slope of 4-5 nm/decade. For nMOSFET, [17] showed that superhalo doping could provide the required

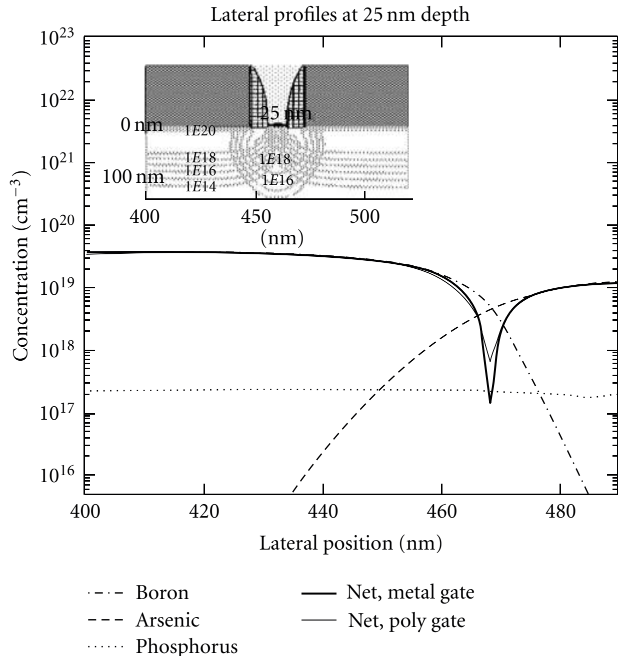


Figure 23: Lateral concentration distribution of boron, arsenic, and phosphorus and lateral net doping profile of a metal gate SALVO PMOS [29].

lateral slope. For an optimized device design, in the lateral direction, the *S/D* doping which is decreasing toward the gate edge can be balanced by the local channel doping which is increasing, such that the net doping profile is more abrupt than either the *S/D* or the local-channel doping profiles. With the process flow optimized for pMOSFET, 25-nm nMOSFET can also be designed with good characteristics, since the n^+ *S/D* doping can be made quite abrupt with low-energy arsenic implant.

4.5.5. Self-Aligned Local Channel V-Gate by Optical Lithography (SALVO). SALVO uses optimized local channel doping to sharpen the lateral junctions in order to minimize the short channel effect for gate lengths down to 25 nm. This novel design flow shows that SALVO is capable of producing pMOSFET devices with lateral net doping slope as abrupt as 6 nm/dec for a metal gate and 8.5 nm/decade for a polysilicon gate, ensuring acceptable SCE down to 25-nm gate length [29]. Experimental results showing lateral concentration distribution of boron, arsenic, and phosphorus and lateral net doping profile of a metal gate SALVO pMOSFET and of the optimized polysilicon-gate PMOS with $T_{\text{poly}} = 95°$C are shown in Figure 23.

4.6. Lateral Channel Engineering to Limit Short Channel Effects. For shorter FETs, halo profiles work to create a higher average doping in the channel than that seen by a longer channel FET, thus tending to raise the V_{th} in opposition to short-channel effects that are lowering it. Such halos are used to achieve the 25-nm bulk CMOS design. The primary advantages of these alternate device structures are a better ideality factor, near unity, and the possibility of thinner Si channels than would be possible in bulk devices except at very low temperature [8]. This technique gives an extra degree of freedom in efforts to limit adverse SCE effects. Locally high doping concentrations in channels near source/drain junctions have been employed via lateral channel engineering, for example, Halo, pocket, tilted channel implantation, and tilted implantation punch-through stopper.

4.6.1. In-Halo Structures to Obtain High Performance. Indium has been successfully used in fabricating abrupt n-MOSFET profiles because of its low diffusion constant which leads to shallower deep submicron nMOSFET source/drain extension halo profiles. In-halo structures allow reduced $V_{\text{th}}S$ while increasing device resistance to SCEs. In halo nMOSFET, the junction leakage and capacitance are affected by indium dosages levels and off-state leakage increased due to an increase in the n^+ junction break down field [30]. In-halo structures show better SCE margin for high-speed device design, and smaller poly gate overlapped channel diffusion areas can be obtained due to shorter poly-Si gate lengths.

4.6.2. Pocket Implant. This is a popular technique for improving short channel effect [1]. Compared with substrate engineering, pocket implant can place the implanted ions near the location where they are needed the most around the drain (and the source). Compared with epitaxial channel, pocket implant using ion implantation is rather of low cost.

4.6.3. Delta Doping. For applications whose I_{OFF} is not a major concern but V_{th}, roll-off control is vital; delta-doped MOSFET is the choice when one wants to reduce oxide stress and oxide leakage. MOSFET with uniform channel doping has lower I_{OFF} than delta-doped MOSFET required for some applications. Uniformly doped MOSFET of 0.09 nm minimum channel length $L_{\min}$ can be achieved with $T_{\text{ox}} = 30\,\text{Å}$, $N_{\text{sub}} = 10^{18}\,\text{cm}^{-3}$, $X_j \sim 200\,\text{Å}$, and $V_{\text{th}} \sim 0.35\,\text{V}$. Ideal delta doping profile can improve $L_{\min}$ by 20% and $L_{\min}$ of 0.07 μm can be achieved. This MOSFET is suitable for applications where larger current drive good digital design noise margin is needed but the stand-by power consumption can be relaxed, such as in high-end microprocessors. For applications where I_{OFF} is important such as devices in memory array, uniformly doped MOSFET is better than delta-doped MOSFET in terms of $I_{\text{dsat}}/I_{\text{OFF}}$ [1].

The schematic device cross sections of MOSFET devices and the definitions of their back gate thicknesses X_{bg} are shown in Figure 24, (a) Uniformly-doped MOSFET is used to approximate conventional MOSFET X_{bg} is depletion width. (b) Delta-doped MOSFET X_{bg} is lightly-doped layer width (X_{ld}). (c) Pocket-Implanted MOSFET, X_{bg} is depletion width (d) Fully-depleted SO1 MOSFET's which has large X_{bg}.

4.6.4. Dual-V_{th} Design to Reduce Subthreshold Leakage. Dual-V_{th} design technique has proven to be extremely effective in reducing subthreshold leakage in both active and

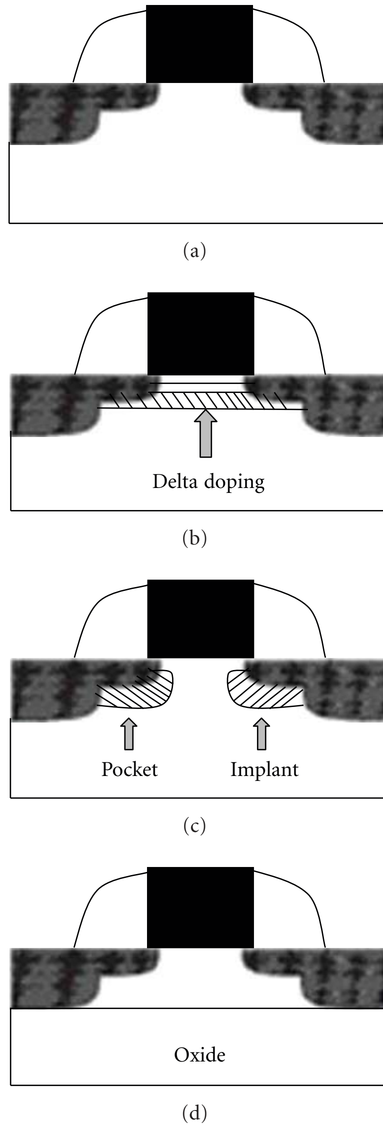


Figure 24: schematic device cross-sections of MOSFET devices shown. (a) Uniformly-doped (UD) MOSFET; (b) delta-doped (DD) MOSFET; (c) pocket-Implanted (PI) MOSFET; (d) Fully-depleted SO1 (FDSOI) MOSFET's [1].

standby mode of operation of a circuit in submicrometer technologies. In nanoscaled bulk silicon technologies, high-V_{th} devices are obtained by changing the peak halo density and its location. An increase in the strength of the "halo" reduces subthreshold leakage and improves short channel effects. It also increases the variability due to random dopant fluctuation and the junction capacitance. Metal gates are being explored not only to have proper control on realizing devices having high V_{th}, but also to achieve high performance while maintaining short channel effect [20].

4.6.5.Highly Doped Drain (HDD) Implants to Ensure Good Short Channel Behaviour. In bulk Ge or on thick germanium-on-silicon substrates, fabrication of short-channel transistors requires a combination of well, extension, halo, and highly doped drain (HDD) implants to ensure good short-channel behavior. For decananometer CMOS p^+-n junctions, it was found that the area component of the leakage depends heavily on the n-type doping level at the metallurgical junction, originating from the halo and well implants. As CMOS technologies are scaled, stronger halos are required to provide sufficient short-channel control. An important observation from [31] is that the halo-implant conditions with good short-channel control, evidenced by a low drain-induced barrier lowering (DIBL), show a high extension leakage J_E and vice versa. All halo conditions with a short-channel DIBL below 200 mV/V have a J_E above 4.3×10^{-7} A/μm or higher.

4.6.6. Lowering the Supply Voltage of the Technology to Obtain Lower Extension Leakage J_E. Scaling a bulk Ge PMOS technology requires halo implants to provide sufficient gate control over the transistor channel, similar to that in silicon technology. These halos increase the electric field at the drain side, leading to enhanced drain leakage. As the BTBT mechanism depends strongly on the electric field, decreasing the supply voltage of the technology is one way to obtain lower extension leakage $J_{E.}$. It was experimentally confirmed that J_E is the dominant junction-leakage component in short-channel Ge pFETs. A tradeoff was found between short-channel control on the one hand and drain leakage on the other. Changing the halo or extension implant will provide better short-channel control only at the expense of increased leakage or vice versa. As a consequence, in sub-70-nm gate-length technologies, the drain leakage is likely 4×10^{-7} A/μm or higher at a supply voltage of 1 V, at room temperature. Through device simulations, it was found that the highest electric fields in off-state condition occur directly under the gate and under the extension region and reach up to 1.1 MV/cm for the centerpoint halo. As BTBT is generated in the presence of high electric fields, one option to reduce J_E is to lower the supply voltage V_{dd}. For supply voltages below 0.7 V, it is shown that it is possible to obtain J_E values below 1×10^{-7} A/μm, combined with DIBL values below 200 mV/V for 70 nm gate-length Ge transistors. Another suggested parameter that may influence J_E is the spacer width; as a large portion of J_E is generated under the spacers, it is possible that narrower spacers lead to lower leakage, at the expense of a reduced channel control [31].

4.6.7. Asymmetric Halo Doping to Reduce Static Leakage. Asymmetric halo bulk MOSFETs have been advocated over conventional symmetric halo MOSFETs for sub-50-nm gate lengths. They have negligible band to band tunneling leakage in the reverse-biased drain substrate junctions and comparatively lower subthreshold leakage, therefore dissipating less static power in the circuits [32].

For the same I_{ON}, there is a slight improvement in the subthreshold swing in AH transistor resulting in reduced I_{OFF}. High halo doping on the source side in this nMOSFET reduces-drain induced barrier lowering (DIBL). AH CMOSFETs are more effective in saving power dissipation when the clock frequency is low and when static power is dominant [32].

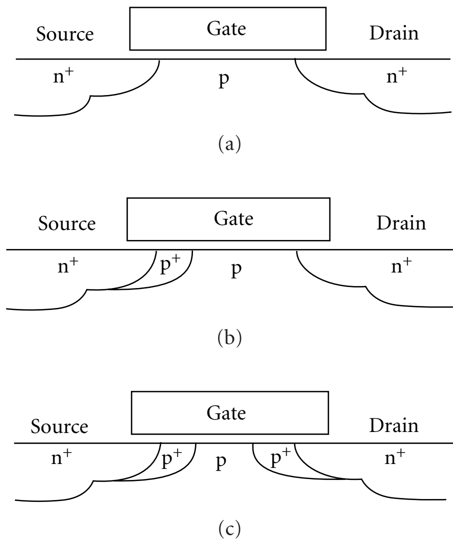


FIGURE 25: Structures of different n-channel MOSFETs. (a) Conventional structure, (b) LAC structure showing a p^+-pocket near the source end, and (c) DH structure showing a p^+-pocket near both the source and drain ends [33].

4.6.8. Single Halo (SH), Lateral Asymmetric Channel (LAC), and Double Halo (DH) to Improve Device Performance. In any halo implant, it is important to optimize the tilt angle of implant. The effects of halo implant on the subthreshold performance of 100-nm CMOS devices and circuits for ultralow power analog/mixed-signal applications have been systematically investigated in [33]. Improvement in the device performance parameters like g_m/I_d, V_A, intrinsic gain, and so forth, for such applications has been observed in the halo-implanted devices while the improvement is significant in LAC devices. Lower tilt angle has been found to produce better results in both LAC and DH devices. A more than 100% improvement in voltage gain has been observed in a current source CMOS amplifier with the use of LAC doping on both the p- and n-channel devices of such amplifiers in the subthreshold regime of operation.

Structures of different n-channel MOSFETs; (a) conventional structure, (b) LAC structure showing a p^+-pocket near the source end, and (c) DH structure showing a p^+-pocket near both the source and drain ends are shown in Figure 25 [33].

So, for the planar forms of the device, it is best to implant halo doping profiles into the channel, but then this would be subject to more fluctuations than for bulk devices because the volume available for such a doping is smaller due to the thikness of the channel. Lateral variations in the gate work function might also be promising. Interface chemistry might necessitate the use of metallic gate electrodes in which case metals must be found with work functions near those of n and p-poly-Si to achieve low $V_{th}S$ [8].

4.7. Highly Doped Gates to Deal with Polydepletion Effects. Polydepletion effects will become more severe with continued scaling of MOSFETs due to the significance of corner and edge effects. Achieving highly doped gates with less dopant gradient in the polygate can be the most appropriate solution to overcome these problems.

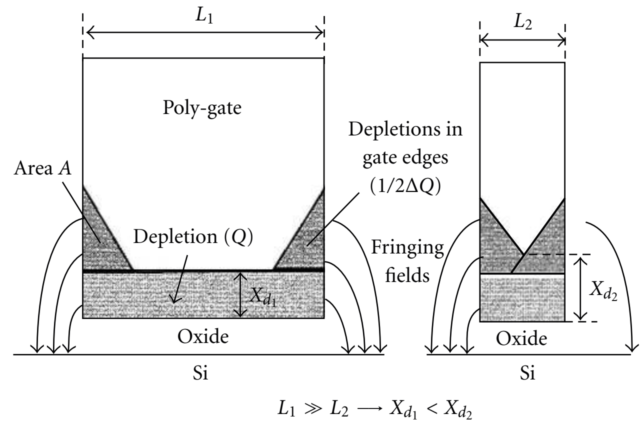


FIGURE 26: Gate-length effect on polydepletion [34].

Gate-length effect on polydepletion in Figure 26 shows that an effective depletion width is wider for the shorter gate length, leading to an additional potential drop. Impact of non uniform dopant distributions and gate geometry to polydepletion effects is reported based on a study using device simulation. Vertically nonuniform and steep dopant profiles in poly-gate result in built-in electric field effects and potential drops in the gate region. Laterally uniform and convex dopant profiles in the poly-gate cause substantial edge potential drops for short gate lengths, mainly due to fringing fields [34].

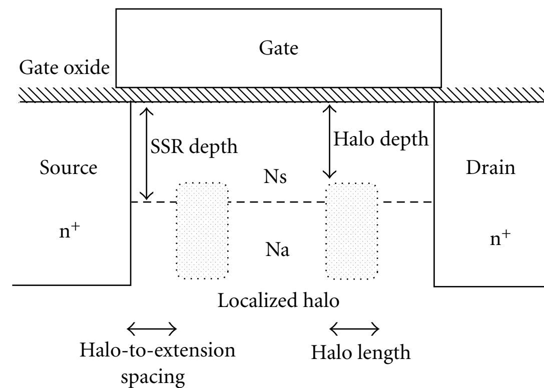


FIGURE 27: Highly localized uniform halo imposed on a conventional SSR channel [35].

4.8. Insulated Shallow Extension (ISE) Structure to Suppress Short Channel Effects. An excellent control over SCE down to 20-nm gate length is realized with shallow junction technology [35]. For an optimized short channel behavior, it is indispensable to control the roll-off of the threshold voltage while keeping both the drain leakage current and threshold voltage at an acceptably low level. Figure 27 is the structure used to examine the effects of localized halo

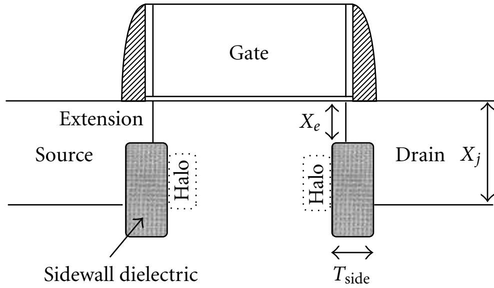


Figure 28: Insulated shallow extension (ISE) structure [35].

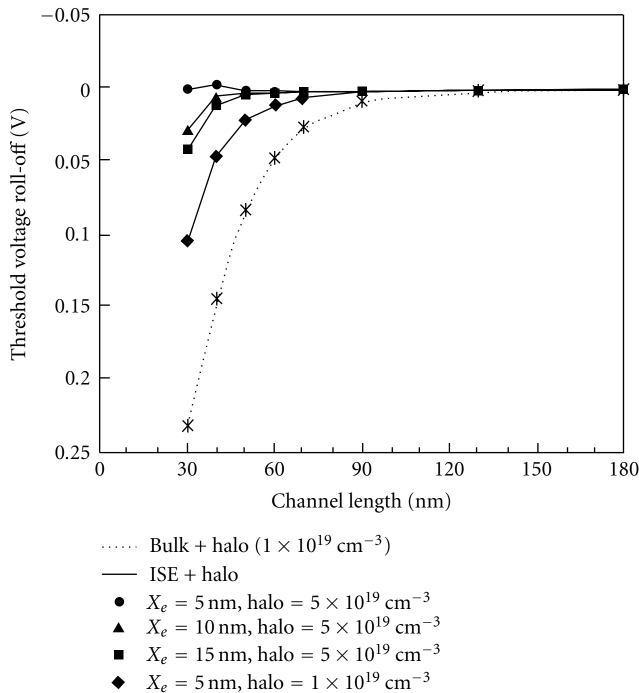


Figure 29: Excellent roll-off behavior achieved for sub-50-nm bulk devices with localized halo ISE structures [35].

structures to minimize the roll-off of short channel threshold voltage with suitably low drain leakage current and threshold voltage level. To reduce the impact of heavily doped halo on the threshold voltage, it is essential to have a deep and short enough halo doping profile. The halo-to-extension spacing is most effective in the suppression of the short channel roll off and the drain leakage, while the use of large halo depth is required to achieve suitably low threshold voltage.

With the ISE structure, a localized heavily doped halo up to 5×10^{19} cm^{-3} can be used to suppress the SCEs without limitation imposed by the drain leakage current. Figure 28 shows the sidewall oxide of the insulated shallow extension (ISE) structure to suppress the short channel effects of bulk MOSFETS.

An excellent roll-off behavior achieved for sub-50-nm bulk devices with localized halo ISE structures is shown in Figure 29. The results of drain current for conventional bulk devices and localized halo ISE devices are shown in Figure 30.

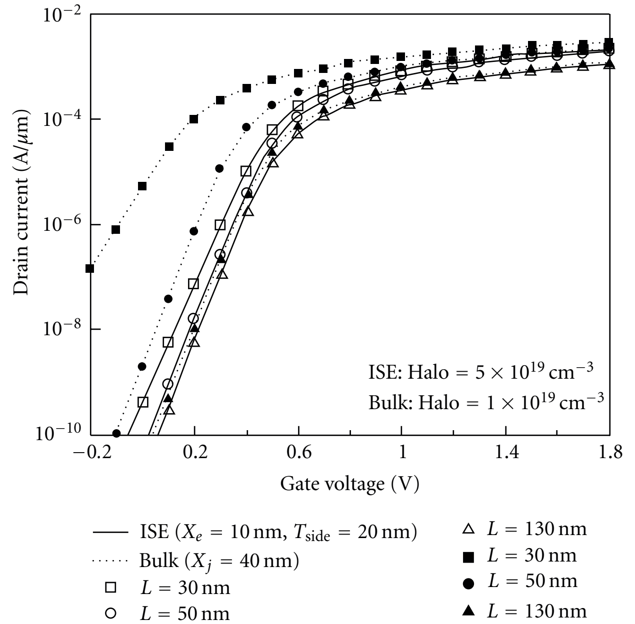


Figure 30: Drain current for conventional bulk devices and localized halo ISE devices [35].

Very good subthreshold behavior is observed for all localized halo ISE devices. The use of localized halo profile is very important for the design of sub-50-nm bulk devices. To reduce the impact of heavily doped halo on the threshold voltage, it is essential to have a deep and short enough halo doping profile. The halo-to-extension spacing is the most effective means to control the band to band tunneling leakage current and the threshold voltage roll off. Implanting the localized halo beneath the channel surface and away from the extension junction edges using ISE structure can realize sub-50-nm bulk MOSFET devices with reduced SCEs [35].

4.8.1. Limitations Imposed by Halo Doping. As CMOS devices are scaled down, the number of dopant atoms in the depletion region of a minimum geometry device decreases. Due to the discreteness of atoms, there is a statistical random fluctuation of the number of dopants within a given volume around its average value. Although the average concentration of doping is quite well controlled by ion implantation and annealing processes, these processes lead to randomness at the atomic scale in the form of spatial fluctuations in the local doping concentration, which in turn cause device-to-device variation in MOSFET threshold voltages. The threshold uncertainty increases with increasing junction depth because of the increasing body doping needed to maintain a fixed off current of 1 nA/m [8]. By separately simulating the effects of discrete donors or discrete acceptors, the simulations also show that the effect of discreteness in the source drain is usually negligible. Stochastic simulations also confirm the analytically predicted result that highly retrograde channel doping profiles can yield significantly lower σV_{th}s than uniformly doped channels. This is because the doping fluctuations are moved further away from the

Table 2: Roadmap for submicron CMOS devices.

Range of operation	Performance factors	Challenges	Technique/Solution	Enhanced performance	Applications
Sub-100 nm	Elevated off-state leakage due to lowering of V_{th} under tech. scaling	Reduce leakage currents	Lowering of temperature Cooling solutions	Total leakage current may be reduced by more than 200x by cooling from 127°C to 0°C	LOP [36]
Sub-100 nm	Device performance deteriorates due to polydepletion effects as devices scale down	Polydepletion effects	Using less dopant gradient Highly doped gates	Vertically nonuniform and steep dopant profile in polygate result in built electric field effect and potential drops in the gate region. Laterally nonuniform and convex dopant profiles in the polygate cause substantial edge pot. drops for short L_gs	LSTP [34]
Sub-100 nm	V_{th} lowering in short channel region	Subthreshold swing, reduced electric field at the drain	Optimizing junction depths and channel doping Grooved gate MOSFET	V_{th} lowering not observed even beyond 0.1 m due to corner effect, V_{th} increases for shorter L for grooved gate MOSFETs	LOP [37]
Sub-100 nm	Control of gate-leakage current is paramount inlow-power CMOS circuit design	Reduce gate currents	Between the high-k layer and silicon substrate an interfacial oxide layer is present, and a transition layer between high-k dielectric and silicon substrate may exist High-k gate stack	The gate current is reduced by the addition of a transition layer and with increasing thickness of the transition layer for the same EOT	LOP [9]
Sub-100 nm	Adverse V_{th} roll off	V_{th} roll off control	Ion implantation Pocket implant	Pocket implant device through optimization of k parameter pushes L_{min} to 55–60% of that of UD channel device, V_{th} overshoots 100 mV	HP-LOP [38]
Sub-100 nm	High bulk impurity conc. leads to an increase in V_{th}	Reduce V_{th}	Low-doped zone next to high-doped substrate Stepped doping profile	Higher mobility and lower threshold voltage, $\Delta V_{th} = 0.5$ V	HP [25]
Sub-70 nm	Pot. Barrier lowering both in the inversion channel and in the body depletion region	V_{th} adjustment, body punch-through	Angle tilted implantation Superhalo and a retrograde channel	Good I_{drive}/I_{OFF} performance achieved for both n-p channel devices, excellent control of V_{th} roll-off down to 40 nm L_g, 935 μA/μm, 395 μA/μm on-state drive currents with V_{dd} of 1.5 V	HP-LOP [37]

channel and closer to the body and so have less effect since they are screened by the free carriers in the body. Thus, it is widely understood that V_{th} should be set differently for different applications to control the subthreshold leakage dissipation.

4.9. Cooling Solutions to Reduce Leakage Currents. The decrease of thermal sensitivity of threshold voltage under technology scaling is one of the major reasons of effectiveness decrease of weak inversion current reduction by cooling. Punch-through leakage current is substantially reduced by the temperature lowering.

The simulated results of technology scaling and temperature reduction on the different leakage mechanisms of subquarter micron MOSFETs are investigated in Figures 31(a) and 31(b).

It shows GIDL and bulk BTBT currents as a function of technology scaling and temperature. The absolute value of the subthreshold leakage current is significantly increased with technology scaling. This component is reduced with reduction in temperature. However, the rate of leakage reduction with temperature is diminished with technology scaling. Punch-through leakage current is increased with technology scaling; however, it can be significantly reduced

Table 3: Roadmap for deep submicron CMOS devices.

Range of operation	Performance factors	Challenges	Technique/Solution	Enhanced performance	Applications
Sub-50 nm	Loss of drive current and enhancement in SCE	Voltage scaling limitation, V_{th} scaling	Nonuniform doping profile (SSR) Lateral channel engineering	High linear current 50 mV drain bias, exhibit smaller roll-off, currents less than UD devices	HP [26]
Sub-50 nm	Revere-biased diode junc. BTBT current	Reduce BTBT	Reduce peak halo doping conc. Asymmetric halo (AH) doping	Dissipates less static power in circuits, improvement in performance	LOP-LSTP [32]
Sub-50 nm	Large variation in different leakage components due to variation in device parameters	V_{th} variability, due to random dopant fluctuation and junction capacitance	Increase in strength of halo Modified AH halo	Min jn. BTBT and highest performance for a given subthreshold leakage, reduces subthreshold leakage, improves SCE	LOP [26]
Sub-50 nm	Adverse V_{th} roll-off	DIBL	High halo doping on source side Non uniform doping	V_{th} drop as a result of DIBL effect is reduced	LOP [32]
Sub-50 nm	Roll off of short channel V_{th} and gate/drain leakage	BTBT control, V_{th} roll-off	Halo to extension spacing Implant localized halo beneath the channel surface	Sub-50-nm bulk MOSFET devices can be achieved with small V_{th} roll-off, low DIBL, suitable V_{th}	LOP [35]
Sub-50 nm	Net doping conc. decreases laterally with a slope of 4-5 nm/dec, difficult to use superhalo tech. for PMOS	Acceptable short channel effects	Lat. conc. dis. of B, Ar, p^+ optimized poly-Si gate Lateral net doping profile in SALVO design	Produces PMOS devices with lateral net doping slope as abrupt as 6 nm/dec for a metal gate, 8.5 nm/dec for a poly Si gate down to 25 nm L_g	LSTP [29]
Sub-22 nm	Metal gate work function affects V_{th} of device and tuning and power of digital circuits	Intrinsic parameter fluctuations	Metal gate technology High-k dielectrics	Intrinsic parameter fluctuations controlled	HP-LOP [24]
Sub-22 nm	Process variation effects, random dopant fluctuation affects V_{th} of device	Intrinsic parameter fluctuation	Random variation of work function Metal gate technology	Metal as a gate material introduces a new source of random variation due to the dependence of work functions on the orientation of metal grains	HP-LSTP [24]

by cooling and by anti-punch-through implantation. Leakage current due to DIBL effect is increased with scaling and is almost insensitive to the temperature. Impact ionization leakage current is increased with technology scaling, and it is almost temperature independent in temperature interval from 300 to 77 K. GIDL and bulk BTBT currents are significantly increased with technology scaling and are almost temperature independent [36].

4.10. Grooved Gate MOSFETs. According to nonplanar device simulated results obtained by [37], grooved gate MOSFETS are highly resistant to short channel effects, apart from being highly well behaved in the sub micron region; these devices also offer a high packing density when compared to conventional planar MOSFETS.

The above analysis is summarized in the form of a roadmap for quick reference for the method to be applied

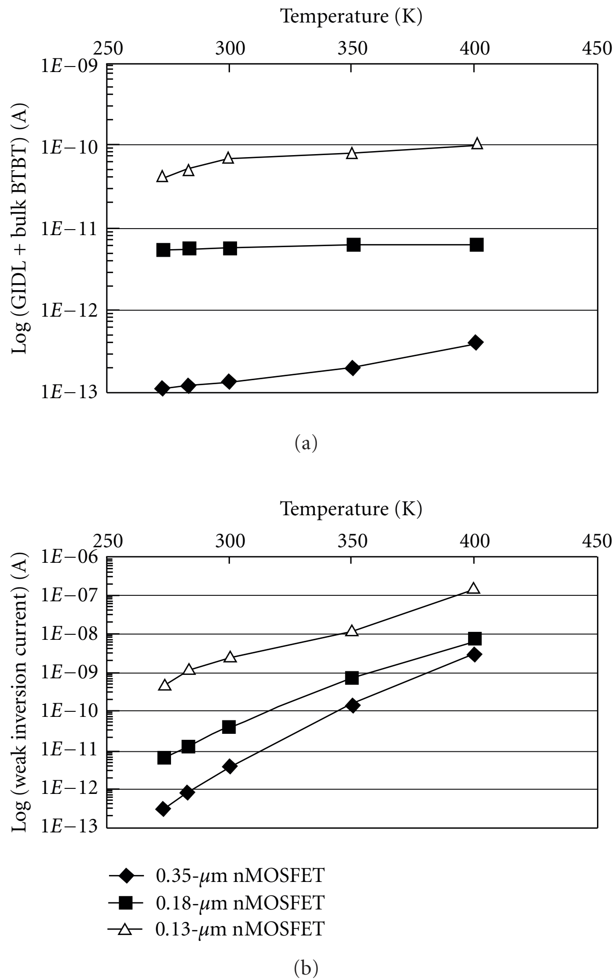


Figure 31: (a) and (b) GIDL and bulk BTBT currents as a function of technology scaling and temperature [36].

for a particular range of operation for optimum results in various applications.

5. Roadmap in a Tabular Form

The high-performance devices (HP) employ the most aggressive scaling as very low threshold voltages are traditionally used, authorizing large SCE and DIBL. Low power in standby mode (LSTP) technologies nominal threshold voltage is large and thus their SCE, DIBL and S are required to be much better than those of HP technologies. The low-operating-power technology (LOP) requires speed in the active mode and low power in the standby mode [39]. Considering the three big families of products as high performance (HP), low operational power (LOP), and low standby power (LSTP), a roadmap is created as a ready reckoner for sub micron and deep submicron devices in Tables 2 and 3.

6. Conclusion

Subthreshold design is an inevitable choice in the semiconductor roadmap for achieving ultra-low-power consumption. In order to achieve optimal performance, device, circuit, and architectural level optimizations specific to the subthreshold operation need to be applied. Due to the high sensitivity of the subthreshold circuits to process variations, it is imperative to use innovative design techniques to improve circuit robustness. Enhanced channel mobility due to applied strain to the channel is a major contributor to meeting the MOSFET performance requirements. In order to successfully scale ICs to meet performance, leakage current, and other requirements, numerous major processes and material innovations, such as high-k dielectric, metal gate electrodes, elevated source/drain, and doping techniques, need to be implemented. Dealing with fluctuations, statistical process variations, impact of quantum effects, line edge roughness, and variation in the ultra thin body width needs to be understood for better deep submicron performance especially for low-power applications.

Acknowledgment

The authors would like to thank Dr. Joydip Dhar, Associate Professor, ABV-IIITM, Gwalior for the help and support extended throughout this work to them. The authors also gratefully acknowledge the Director, ABV-IIITM, Gwalior, for the support and encouragement for this research work.

References

[1] C. H. Wann, K. Noda, T. Tanaka, M. Yoshida, and C. Hu, “A comparative study of advanced MOSFET concepts,” *IEEE Transactions on Electron Devices*, vol. 43, no. 10, pp. 1742–1753, 1996.

[2] K. Kim, K. K. Das, R. V. Joshi, and C. T. Chuang, “Nanoscale CMOS circuit leakage power reduction By Double-gate device,” in *Proceedings of the International Symposium on Lower Power Electronics and Design (ISLPED ’04)*, pp. 102–107, August 2004.

[3] B. P. Wong et al., “Nano CMOS scaling problems and implications,” in *Nano-CMOS Circuit and Physical Design*, John Wiley and Sons, 2005.

[4] S. G. Narendra, “Challenges and design choices in nanoscale CMOS,” *ACM Journal on Emerging Technologies in Computing Systems*, vol. 1, no. 1, 2005.

[5] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, 1998.

[6] Y. C. Yeo, T. J. King, and C. Hu, “MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations,” *IEEE Transactions on Electron Devices*, vol. 50, no. 4, pp. 1027–1035, 2003.

[7] S. E. Thompson and S. Parthasarathy, “Moore’s law: the future of Si microelectronics,” *Materials Today*, vol. 9, no. 6, pp. 20–25, 2006.

[8] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H. S. P. Wong, “Device scaling limits of Si MOSFETs and their application dependencies,” *Proceedings of the IEEE*, vol. 89, no. 3, pp. 259–287, 2001.

[9] J. P. Sun, W. Wang, T. Toyabe, N. Gu, and P. Mazumder, “Modeling of gate current and capacitance in nanoscale-MOS structures,” *IEEE Transactions on Electron Devices*, vol. 53, no. 12, pp. 2950–2956, 2006.

[10] W. M. Elgharbawy and M. A. Bayoumi, "Leakage sources and possible solutions in nanometer CMOS Technologies," *IEEE Circuits and Systems Magazine*, vol. 5, no. 4, pp. 6–16, 2005.

[11] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current in deep-submicron CMOS circuits," *Journal of Circuits, Systems and Computers*, vol. 11, no. 6, pp. 575–600, 2002.

[12] N. Ekekwe and R. Etienne-Cummings, "Power dissipation sources and possible control techniques in ultra deep submicron CMOS technologies," *Microelectronics Journal*, vol. 37, no. 9, pp. 851–860, 2006.

[13] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 305–327, 2003.

[14] S. Mukhopadhyay, A. Raychowdhury, and K. Roy, "Accurate estimation of total leakage in nanometer-scale bulk CMOS circuits based on device geometry and doping profile," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 3, pp. 363–381, 2005.

[15] A. Pouydebasque, C. Charbuillet, R. Gwoziecki, and T. Skotnicki, "Refinement of the subthreshold slope modeling for advanced bulk CMOS devices," *IEEE Transactions on Electron Devices*, vol. 54, no. 10, pp. 2723–2729, 2007.

[16] S. Hanson, M. Seok, D. Sylvester, and D. Blaauw, "Nanometer device scaling in subthreshold circuits," in *Proceedings of the 44th ACM/IEEE Design Automation Conference (DAC '07)*, pp. 700–705, June 2007.

[17] Y. Taut and E. J. Nowak, "CMOS devices below 0.1 ¯m: how high will performance go?" in *Proceedings of the International Electron Devices Meeting (IEDM '97)*, pp. 215–218, December 1997.

[18] D. Rairigh, "Limits of CMOS technology scaling and technologies beyond-CMOS," *IEEE*, 2006.

[19] ITRS, 2001.

[20] A. Agarwal, K. Kang, S. Bhunia, J. D. Gallagher, and K. Roy, "Device-aware yield-centric dual-V design under parameter variations in nanoscale technologies," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 15, no. 6, pp. 660–671, 2007.

[21] *Introducing 65nm Technology in Micro Wind 3*, 2006.

[22] L. T. Clark, R. Patel, and T. S. Beatty, "Managing standby and active mode leakage power in deep sub-micron design," in *Proceedings of the International Symposium on Lower Power Electronics and Design (ISLPED '04)*, pp. 274–279, August 2004.

[23] R. Chau, M. Doczy, B. Doyle et al., "Advanced CMOS transistors in the nanotechnology era for high-performance, low-power logic applications," in *Proceedings of the International Conference on Solid-State and Integrated Circuits Technology Proceedings (ICSICT '04)*, vol. 1, pp. 26–30, 2004.

[24] Y. Li, C. H. Hwang, T. Y. Li, and M. H. Han, "Process-variation effect, metal-gate work-function fluctuation, and random-dopant fluctuation in emerging CMOS technologies," *IEEE Transactions on Electron Devices*, vol. 57, no. 2, pp. 437–447, 2010.

[25] J. A. López-Villanueva, F. Gámiz, J. B. Roldán, Y. Ghailan, J. E. Carceller, and P. Cartujo, "Study of the effects of a stepped doping profile in short-channel mosfet's," *IEEE Transactions on Electron Devices*, vol. 44, no. 9, pp. 1425–1431, 1997.

[26] I. De and C. M. Osburn, "Impact of super-steep-retrograde channel doping profiles on the performance of scaled devices," *IEEE Transactions on Electron Devices*, vol. 46, no. 8, pp. 1711–1717, 1999.

[27] E. C. Kan, V. Narayan, and G. Pei, "Band to band tunneling by Monte Carlo simulation for prediction of MOSFET's," *IEEE Transactions On Electron Devices*, vol. 44, no. 9, 1997.

[28] Y. Taur, C. H. Wann, and D. J. Frank, "25 nm CMOS design considerations," in *Proceedings of the IEEE International Electron Devices Meeting (IEDM '98)*, pp. 789–792, December 1998.

[29] H. H. Vuong, C. P. Chang, and C. S. Pai, "Design of 25-nm SALVO PMOS devices," *IEEE Electron Device Letters*, vol. 21, no. 5, pp. 248–250, 2000.

[30] W. K. Yeh and J. W. Chou, "Optimum halo structure for sub-0.1 μm CMOSFETs," *IEEE Transactions on Electron Devices*, vol. 48, no. 10, pp. 2357–2362, 2001.

[31] G. Eneman, B. De Jaeger, E. Simoen et al., "Quantification of drain extension leakage in a scaled bulk Germanium PMOS technology," *IEEE Transactions on Electron Devices*, vol. 56, no. 12, 2009.

[32] A. Bansal and K. Roy, "Asymmetric halo CMOSFET to reduce static power dissipation with improved performance," *IEEE Transactions on Electron Devices*, vol. 52, no. 3, pp. 397–405, 2005.

[33] S. Chakraborty, A. Mallik, C. K. Sarkar, and V. R. Rao, "Impact of halo doping on the subthreshold performance of deep-submicrometer CMOS devices and circuits for ultralow power analog/ mixed-signal applications," *IEEE Transactions on Electron Devices*, vol. 54, no. 2, pp. 241–248, 2007.

[34] C. H. Choi, P. R. Chidambaram, R. Khamankar, C. F. Machala, Z. Yu, and R. W. Dutton, "Dopant profile and gate geometric effects on polysilicon gate depletion in scaled MOS," *IEEE Transactions on Electron Devices*, vol. 49, no. 7, pp. 1227–1231, 2002.

[35] C. H. Shih, Y. M. Chen, and C. Lien, "Design strategy of localized halo profile for achieving sub-50 nm bulk MOSFET," *Microelectronics Reliability*, vol. 44, no. 7, pp. 1069–1075, 2004.

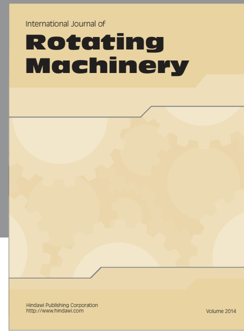
[36] O. Semenov, A. Vassighi, and M. Sachdev, "Impact of technology scaling on thermal behavior of leakage current in sub-quarter micron MOSFETs: perspective of low temperature current testing," *Microelectronics Journal*, vol. 33, no. 11, pp. 985–994, 2002.

[37] J. Tanaka, "Simulation of sub 100nm MOSFETS with completely suppressed short channel effects," *IEEE Electron Device Letters*, vol. 14, no. 8, 1993.

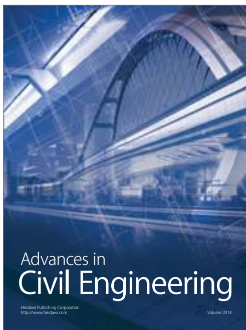
[38] B. Yu, C. H. J. Wann, E. D. Nowak, K. Noda, and C. Hu, "Short-channel effect improved by lateral channel-engineering in deep-submicronmeter MOSFET's," *IEEE Transactions on Electron Devices*, vol. 44, no. 4, pp. 627–634, 1997.

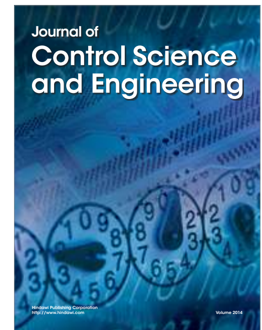
[39] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon et al., "Innovative materials, devices, and CMOS technologies for low-power mobile multimedia," *IEEE Transactions on Electron Devices*, vol. 55, no. 1, pp. 96–130, 2008.

Journal of
Engineering

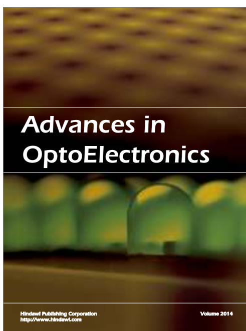
The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

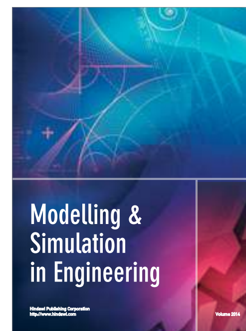
International Journal of
Distributed
Sensor Networks

Advances in
Civil Engineering

Hindawi
Submit your manuscripts at
http://www.hindawi.com

Journal of
Control Science
and Engineering

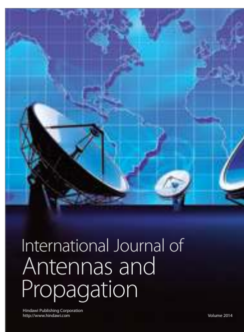
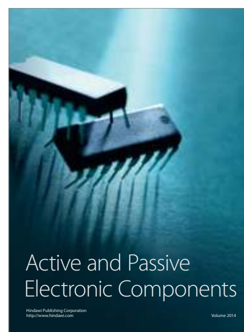
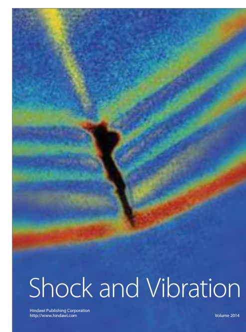
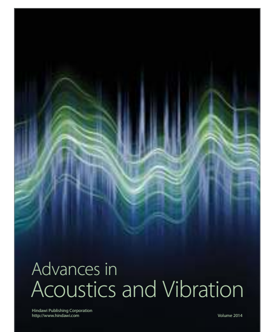
Journal of
Robotics

Journal of
Electrical and Computer
Engineering

Advances in
OptoElectronics

VLSI Design

International Journal of
Navigation and
Observation

Modelling &
Simulation
in Engineering

International Journal of
Aerospace
Engineering

International Journal of
Chemical Engineering

International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components

Shock and Vibration

Advances in
Acoustics and Vibration