International Journal of Advance Research in Engineering, Science & Technology(IJAREST),
ISSN(O):2393-9877, ISSN(P): 2394-2444,
Volume 2, Issue 6, June- 2015, Impact Factor: 2.125

SCALING CHALLENGES IN CMOS TECHNOLOGY

Adarsh H C, Nithin M

[1]*PG Scholar, Dept. of E&C, RVCE, Bengaluru, India*, [2]Asst. Prof., Dept. of E&C, RVCE, Bengaluru, India
[1]adarsh.heggere@gmail.com, [2]nithinm@rvce.edu.in

Abstract

This paper explores scaling challenges in CMOS transistor. Conventional devices have been scaled by thinning gate dielectrics, forming shallower extensions, increasing channel doping, and lowering power supply voltages. Many of these key scaling methods are reaching fundamental limitations. New thin body device architectures such as FinFETs are emerging which do not rely on the conventional scaling approach. The short channel effects for these new device options improve as the channel thickness is reduced. A brief overview of transistor architectures such as extremely thin silicon-on-insulator and MUGFET (FinFET), as well as nanowire device architectures are discussed here.

Keywords—*Complementary metal-oxide semiconductor (CMOS), FinFET, mobility, nanowire, Silicon on Insulator (SOI), strain engineering.*

1. INTRODUCTION

For the past 40 years, relentless focus on Moore's Law transistor scaling has provided ever-increasing transistor performance and density. The motivation factors for scaling is to achieve high density of circuits (i.e. same functionality in a smaller area or more functionality in the same area), cost reduction and for faster switching of transistors (i.e. to decrease delay). For further scaling, both the familiar challenges of historical scaling and the new challenges associated with length scales on the order of atomic dimensions needs to be addressed.

In these advanced devices, the traditional issues of channel mobility, short-channel control, and parasitic resistance and capacitance are still critically important. However, in addition to these traditional issues, there are new issues of atomic

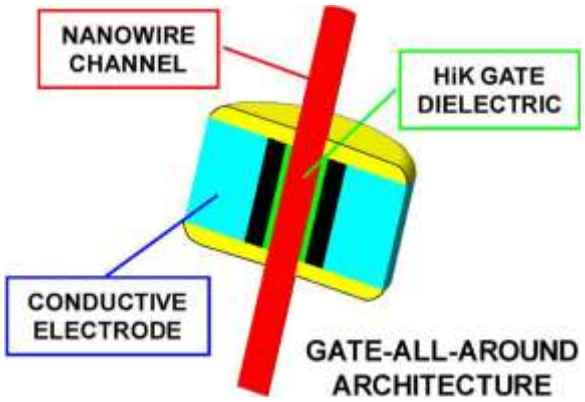

Fig.1. Basic components of the ultimate CMOS device [1]

spacing limiting critical dimensions, interface and support layers dominating the physical structures, and quantum confinement and scattering effects.

Consider a device as shown in fig.1 with a nanowire channel, a gate-all-around (GAA) architecture, a high-*k* gate dielectric, and a conductive gate electrode stack. The minimum channel dimensions will be determined by quantum confinement effects and scattering at atomic dimensions. The nanowire architecture is determined by electrostatic requirements to achieve the best possible short-channel control. Each of the various gate layers (interface layer (IL), high-*k* layer, threshold voltage (V_T) control layer, primary work function layer, conduction layer, and so on) is limited by material properties at atomic dimensions.

Parasitic capacitance impacts circuit performance by increasing the capacitive load and thus increasing delay. In addition, the active power in a circuit is proportional to $C_{dyn}V^2f$ (where C_{dyn} is the total dynamic capacitance, V is the operating voltage, and f is the frequency); thus, reducing C_{dyn} improves active power.

II. SCALING IN PLANAR CMOS

CMOS technology scaling continues to follow Moore's law well into the nano-scale regime. The number of devices per chip and the system performance has been improving exponentially over the last two decades. As the channel length is reduced, the performance improves, the power per switching event decreases, and the density improves. But the power density, total circuits per chip, and the total chip power consumption has been increasing along with following drawbacks: 1) higher sub-threshold conduction, 2) increased gate oxide leakage, 3) increased junction leakage, 4) DIBL (Drain Induced Barrier Loading) and V_T roll off, 5) lower output resistance, 6) lower transconductance, 7) interconnect capacitance, 8) heat production, 9) process variations and 10) modeling challenges.

With the aggressive reduction of MOSFET dimensions into the deep sub micrometer regime, hot-carrier degradation (HCD) is becoming an important reliability issue. HCD results from heating and subsequent injection of carriers into the gate oxide, which results in a localized and non-uniform buildup of interface states and oxide charges near the drain junction of the transistor. The generated defects produce threshold voltage shift, transconductance degradation, drain current reduction, etc., and eventually lead to device failure[2].

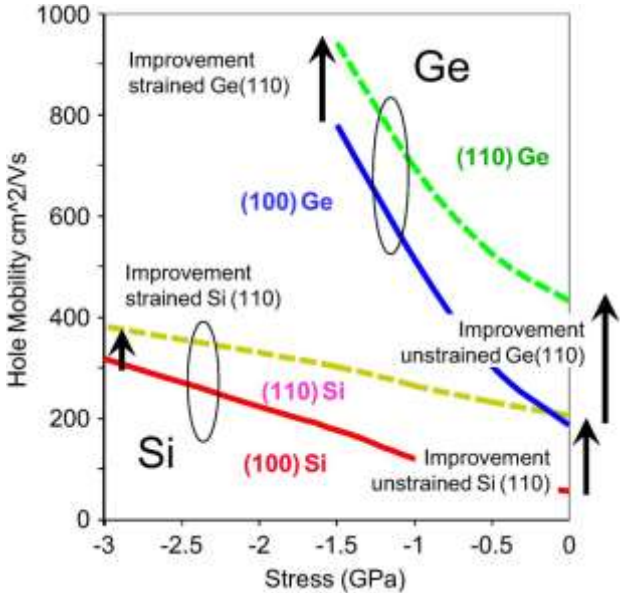

Fig.2. Mobility and strain in Si and Ge as a function of stress and wafer orientation [1]

The need for more performance and integration has accelerated the scaling trends in almost every device parameter, such as lithography, effective channel length, gate dielectric thickness, supply voltage, device leakage, etc. Some of these parameters are approaching fundamental limits, and alternatives to the existing material and structures may need to be identified in order to continue scaling.

III. SCALING SOLUTIONS

Several innovations were introduced to extend mainstream CMOS chip scaling like channel mobility enhancement by process-induced strain, capacitance-equivalent oxide thickness (CET) scaling and gate tunneling reduction by high-K/metal gate and better electrostatic control by FinFET/Multi-Gate FET (MUGFET) 3D structure. By providing better gate control, FinFET/MUGFET enables fully depleted channel and pushes the sub-threshold swing very close to the thermal limit of 60mV/dec at room temperature. This allows further V_{DD} and V_T scaling for power and leakage reductions.[3]

A. Strain engineering

In 1992 it was first demonstrated that n-channel MOSFETs on a strained Si substrate, exhibit a 70% higher effective mobility (μ_{eff}) than those on unstrained substrates as shown in fig.2 and thus increases the current I_D. The two approaches to introduce strain in Si channel of MOSFETs are, a global one, where stress is introduced across the entire substrate, and a local approach, where stress is engineered into the device by means of shallow-trench-isolation, epitaxial layers and/or highly stressed nitride capping layers. For a PMOS, a compressive linear stress is applied whereas for a NMOS, a tensile linear stress is applied to increase I_D in it.

B. ETSOI (Extremely Thin SOI)

A silicon-on-insulator (SOI) device has the potential to improve planar short-channel properties by forming a channel in a silicon film whose thickness (T_{si}) is thinner than the channel depletion depth. Such a device is called an extremely thin SOI (ETSOI) device [alternatively referred to as an ultrathin-body SOI device or a fully depleted SOI device (FDSOI)].

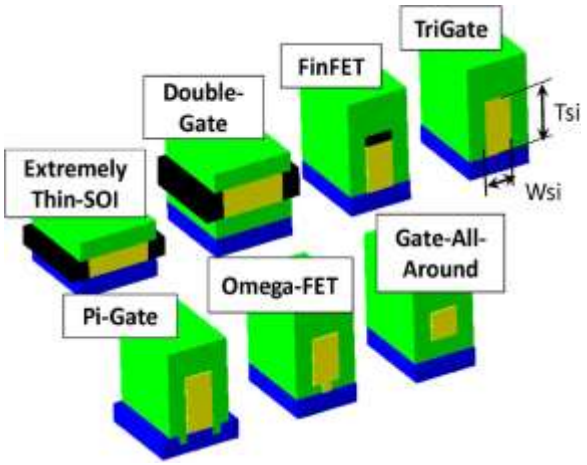

Fig.3. Architectures which reduce source–drain interaction [1]

ETSOI devices possess improved short-channel properties and lower channel doping (with associated benefits in random-dopant fluctuations and mobility) and offer the possibility for body-bias [using thin buried oxide (BOX)].

ETSOI shows significant body-bias effect as well as limited sensitivity of the body effect to L_{eff}. As a disadvantage, body-bias with thin BOX can degrade sub-threshold slope (if $V_B > 0$) by affecting the potential barrier at the back-gate. In addition, all body-bias schemes (particularly those which selectively alter the V_T of individual devices) must take into account the density degradation due to the additional routing and taps required to access the body.

The major challenge of ETSOI devices centers on the stringent (< 10 nm) thickness requirements for T_{Si}. These small dimensions create several significant challenges, including the following: 1) thickness targeting and variation in ETSOI source material; 2) performance issues, including high parasitic S/D resistance and strain; and 3) quantum confinement and scattering effects. Overall, in spite of advances in resistance and strain engineering, ETSOI devices continue to perform with lower drive currents than comparable bulk planar or fully depleted devices created using multiple gate techniques.

C. MUGFET (Multi Gate FET)

An alternative approach for short-channel control is to surround the channel with two (or more) opposing gates. Each additional gate improves the short-channel control. These gates can be oriented horizontally (a double-gate device) or vertically (a FinFET device). SOI is not required in these multiple-gate devices. A simple but effective way to portray the improvement in fully depleted multiple gate devices is to use the "natural channel length" parameter λ_N. This parameter represents the extension of the electric field lines from the S/D

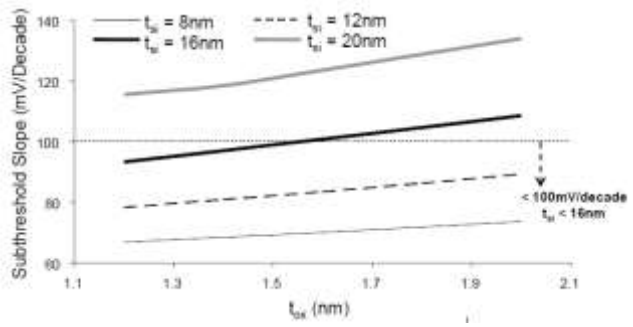

Fig.4. The sub-threshold slope of an N-type FinFET with the different fin and gate oxide thicknesses. T = 27^0C. VDD = 0.8V [5]

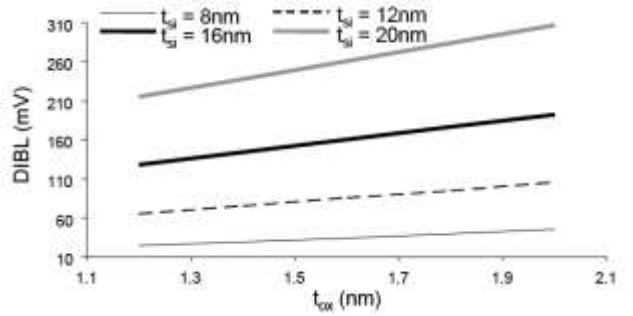

Fig.5. The drain-induced-barrier-lowering of an N-type FinFET with the different fin and gate oxide thicknesses. T = 27^0C [5]

regions into the channel region. A device will have minimal short-channel effects if L_{eff} is approximately 6× longer than λ_N. Note that the effective length λ_N can be improved by increasing the number of gates, decreasing the gate dielectric thickness (T_{ox}), decreasing the channel thickness (T_{si}), or decreasing the permittivity of the channel (ε_{Si}).

In double-gate devices, the improved short-channel effects resulting from more than two gates led to several important modifications including the TriGate architecture (gates on two sides and the top), Pi-Gates (the side gates extend below the channel), and Omega-FETs (the gate not only wraps around two sides and the top but under laps part of the fourth).

The attractiveness of FINFET consists in the realization of self-aligned double-gate devices with a conventional CMOS process. This allows extending the gate scaling beyond the planar transistor limits, maintaining a steep sub-threshold slope, better performance with bias voltage scaling and good matching due to low doping concentration in the channel [4].

These multiple-gate devices have electrostatic advantages over conventional planar devices. In addition, the increased electrostatic confinement provided by multiple gates relaxes the manufacturing constraints in comparison to ETSOI. Multiple gate devices have an additional critical benefit over FDSOI devices in that the total electrical area may be significantly larger than the total footprint area. In addition to providing a potential layout density benefit, modern products will also see a performance benefit from the increased drive current (in spite of the increased effective channel width) as products are typically more heavily loaded by parasitic capacitances (70%) than gate-originated capacitances (30%). Since fully depleted devices [ETSOI and multiple gate field effect transistors (MuGFETs)] control I_{off} through architecture rather than doping profiles, the channel can remain undoped (V_T targeting can be done through altering the work function of the gate). Undoped channels have the potential for low random variation due to minimization of random dopant fluctuations. Undoped devices display the lowest measured random V_T variation values in the literature. Manufacturing and design complexity continue to be the most significant challenges for future MuGFET devices. Horizontally oriented MuGFET devices (double-gate devices) face the difficult challenges of a release etch to access the lower gate, as well as the requirements for highly conformal atomic layer deposition (ALD) gate dielectric and metal electrode processes. Vertically oriented MuGFET devices (FinFET/TriGate devices) face significant fin and gate patterning challenges associated with the non-planar architecture. The high aspect ratios at tight fin pitches introduce new challenges for S/D and extension doping, likely requiring creation of new doping and annealing techniques. The granularity of the FinFET/TriGate architectures (a transistor can only have an integral number of fins) introduces significant new complexity into the circuit design process, particularly for low power geometries where single-fin devices may be common. Register files and memory circuits also face significant challenges due to the quantization of fins and the limited flexibility to tune single-fin solutions for optimal circuit stability. Overall, in spite of the challenges of the non-planar architecture, recent MuGFET devices have achieved higher drive currents than comparable generation FDSOI devices while retaining equivalent short-channel control. In spite of the significant manufacturing issues, TriGate devices have been implemented successfully into manufacturing on the 22-nm node.

D. GAA (Gate All Around) Architecture

Gate-all-around (GAA) and nano-wire (NW) FET structure pushes multi-gate architecture to the extreme, which can further boost performance by enabling even more aggressive gate length scaling. GAA/NW FETs have been built in a vertical channel configuration or in a horizontal hanging structure with multiple wires stacked vertically. Both approaches have further benefit of increasing transistor density to boost system throughput and performance [3].

GAA devices differ from Omega-FETs in that the gate wraps entirely around the device. Note that both lateral and vertical devices are possible with GAA architecture. Both types provide optimal electrostatic confinement with the associated short-channel effect benefits.

Nanowires are an extreme case of GAA devices, having height and width dimensions roughly the same (or even cylindrical) and atomically small (< 10 nm) dimensions. These devices operate in a size and field regime where carrier conduction moves from the surface of the device (as in conventional planar and finned devices) to the center of the device. Nanowires represent the extreme limit of MuGFET

scaling as they operate in the regime of fully depleted and quantum confined with associated changes in the transport physics. Theoretical low field mobility studies for NMOS nanowires suggest flat or improved mobility down to a certain size (6–8 nm) and then rapid degradation in mobility at smaller sizes due to phonon and surface

roughness scattering. Theoretical mobility studies for PMOS are more complex (due to strain and band non-parabolicity) but also more optimistic, suggesting that _110_ channel direction hole mobility down to 5-nm wire sizes remains competitive to planar mobility for high field and stress.

IV. CONCLUSION

Table 1: The analytical expressions for the natural length of multigate architectures [6]

Gate architecture	Natural length
Single gate	$\lambda_1 = \sqrt{\frac{\varepsilon_{si}}{\varepsilon_{ox}} t_{si} t_{ox}}$
Double gate	$\lambda_2 = \sqrt{\frac{\varepsilon_{si}}{2\varepsilon_{ox}}\left(1+\frac{\varepsilon_{ox}}{4\varepsilon_{si}}\frac{t_{si}}{t_{ox}}\right) t_{si} t_{ox}}$
Triple gate, square section	$\lambda_3 = \sqrt{\frac{\varepsilon_{si}}{3\varepsilon_{ox}}\left(1+\frac{\varepsilon_{ox}}{4\varepsilon_{si}}\frac{t_{si}}{t_{ox}}\right) t_{si} t_{ox}}$
Quadruple gate, square section	$\lambda_4 = \sqrt{\frac{\varepsilon_{si}}{4\varepsilon_{ox}}\left(1+\frac{\varepsilon_{ox}}{4\varepsilon_{si}}\frac{t_{si}}{t_{ox}}\right) t_{si} t_{ox}}$
Cylindrical GAA	$\lambda_{GAA} = \sqrt{\frac{2\varepsilon_{si} R^2 \ln(1+\frac{t_{ox}}{R}) + \varepsilon_{ox} R^2}{4\varepsilon_{ox}}}$

The scaling challenges in CMOS transistor are reviewed. The solutions offered to overcome these challenges are discussed i.e. by introducing strain or by moving from planar to non-planar architectures like MUGFET and GAA. The multigate transistor structure achieves improved control of short-channel effects as depicted in table 1. By increasing the number of gates, λ decreases and improves short channel effects.

REFERENCES

[1] Kelin J. Kuhn, "Considerations for Ultimate CMOS Scaling" IEEE Transactions on Electron Devices, Vol. 59, No. 7, July 2012, p.1813-1828

[2] S. Mahapatra, Chetan D. Parikh, V. Ramgopal Rao, Chand R. Viswanathan and Juzer Vasi, "Device Scaling Effects on Hot-Carrier Induced Interface and Oxide-Trapped Charge Distributions in MOSFET's" IEEE Transactions on Electron Devices, Vol. 47, No. 4, April 2000, p.789-796

[3] Jack Y.-C. Sun, "System Scaling and Collaborative Open Innovation"

[4] M.Jurczak, N.Collaert, A.Veloso, T.Hoffmann, S.Biesemans, "Review of FINFET technology", IEEE 2009

[5] Sherif A. Tawfik, Volkan Kursun, "FinFET Technology Development Guidelines for Higher Performance, Lower Power, and Stronger Resilience to Parameter Variations" IEEE 2009

J.-P. Colinge, "Multigate Transistors-Pushing Moore's law to limit" IEEE 2014