American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-03, Issue-01, pp-222-232
www.ajer.org

Research Paper Open Access

An Evolutionary Transition of conventional n MOS VLSI to CMOS considering Scaling, Low Power and Higher Mobility

Md Mobarok Hossain Rubel[1], Muhammad Minhazul Haque Bhuiyan[2]
[1]*Electrical and Electronic Engineering, Leading University, Sylhet, Bangladesh)*
[2]*(Computer Science and Engineering, Leading University, Sylhet, Bangladesh)*

***Abstract:* -** This paper emphasizes on the gradual revolution of CMOS scaling by delivering the modern concepts of newly explored device structures and new materials. After analyzing the improvements in sources, performance of CMOS technology regarding conventional semiconductor devices has been thoroughly discussed. This has been done by considering the significant semiconductor evolution devices like metal gate electrode, double gate FET, FinFET, high dielectric constant (high k) and strained silicon FET. Considering the power level while scaling, the paper showed how nMOS VLSI chips have been gradually replaced by CMOS aiming for the reduction in the growing power of VLSI systems.

***Keywords:* -** *Dielectric, Electrostatic, Low power Design, Moore's Law, Microelectronics, Nanowire, Nanotechnology, VLSI, Work function.*

I. INTRODUCTION

Early 1970 was an era when various methods of scaling MOS devices were explored and it was found that if the voltages with lithoographic dimensions were scaled, benefits of scaling like faster, low energy consumption and cheaper gates would be made easily. Semiconductor industry has been so successful that Semiconductor Industry Association (SIA) has published roadmaps [1] for semiconductor technology since 1992. The only objective of the roadmap incorporating the industries in many developed nations was to pursue with Moore's law [2], which is generally known as the doubling of transistors performance and quadrupling of the number of devices on a cheap every three years. As the MOSFET's power performance was improved, it literally followed the evolution of CMOS technology which was introduced in the late 1970. Power FET technologies use depreciated CMOS basics, with the leading edge with a time delay in the order of feature size as $1\mu m$, $0.8\mu m$, $0.5\mu m$, $0.35\mu m$, $0.25\mu m$, $0.18\mu m$ etc. The outstanding progress signified by Moore's law leaded VLSI circuits to be used in electronic applications like computing, portable electronics and telecommunications [3].

But it is a matter of disgrace that no hypothesis can last forever and recently scaling has been diverged from its ideal characteristics that were assumed before. The problem was found critical when it was seen that all device voltages can not scale; since kT/q does not scale and leakage currents are set by the transistor's threshold voltage, certainly there was a limit to how transistor's V_{th} can be made. Fixing V_{th}, changing V_{dd} simply trades off energy and performance.

Shrinking the conventional MOSFET beyond 50-nm-technology node requires innovations to circumvent barriers due to the fundamental physics that constraints the conventional MOSFET.

Unreliable power scaling, combined with previously applied aggressive performance scaling strategy has made power the most vital problem in modern chip design. Manufacturers can no longer focus on creating the highest performance chips just because of uncertainty whether the chips will dissipate more power. The limitations must be included with quantum mechanical tunneling of carriers through the thin gate oxide, quantum mechanical tunneling of carriers from source to drain and drain to the body of the MOSFET, control of the density and location of the dopant atoms in the MOSFET channel and source/drain region to provide a high on-off current ratio and finally the finite subthreshold slope. These predominant limitations led semiconductor industries to pessimistic predictions of the significant end of technological progress [1].

The organization of this paper is first to address opportunities for the silicon MOSTFET that usually deviate from conventional scaling techniques like doping profile control and thin silicon dioxide gate dielectrics. Later discussions include high dielectric constant gate dielectric, metal gate electrode, double gate FET and strained silicon FET. Following the fact, the paper also shows the difference between conventional microelectronics technology and the more predefined nanotechnology.

II. EARLY MOSFETs AND THE DEVELOPMENT IN THE FIELD

The first generation of macrocell power MOSFET transistors were double diffused MOSFET (DMOS) which was introduced by International Rectifier into the market. This was simply known as planer power MOSFET. The second generation of macrocell technology TrenchFET introduced by Siliconix became popular in the 1990. This actually offered improved switch resistance. This technology was more advantageous than the previous one as it was designed for a drain voltage capability lower than 100V. However, soon the switching loss that was assumed to be very important in switch mode power supply (SMPS), remain the main hindrance. Transient response has become the burning question to be improved as well as the converter’s switching frequency. Macrocell power MOSFET recently introduced by Texas Instruments, NexFETTM technology offers a specific R_{DSON} competitive to the TrenchFET which is in the order to reduce the input and Miller capacitances significantly. This new generation MOSFET reduces switching losses in SMPS applications and enables operation at high switching frequencies. It has been proved to be promising at 30V and below which is desirable for distributed bus architecture prevalent in today’s end systems.

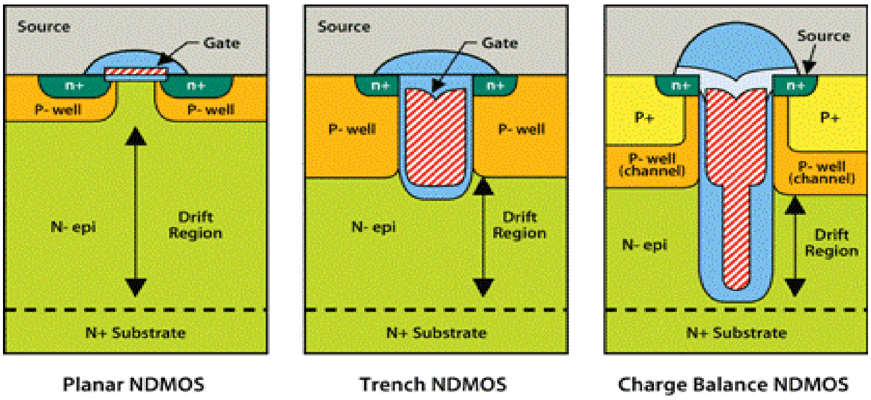

Fig. 1: Comparison of planar DMOS and TrenchFET device structures

III. SILICON GROWTH AND INSTABILITY AND MOSFET

The main problem of electron transport in SiO_2 was high field electron transport in polar insulators which was demonstrated by Karel Throbner in 1970 when he was pursuing his PhD thesis with Richard Feynman. Experimental observations do not show predicted run-away at $2-3MV/cm$ and as a result, Umklapp scattering with acoustic phonons keeps electron energy under control.

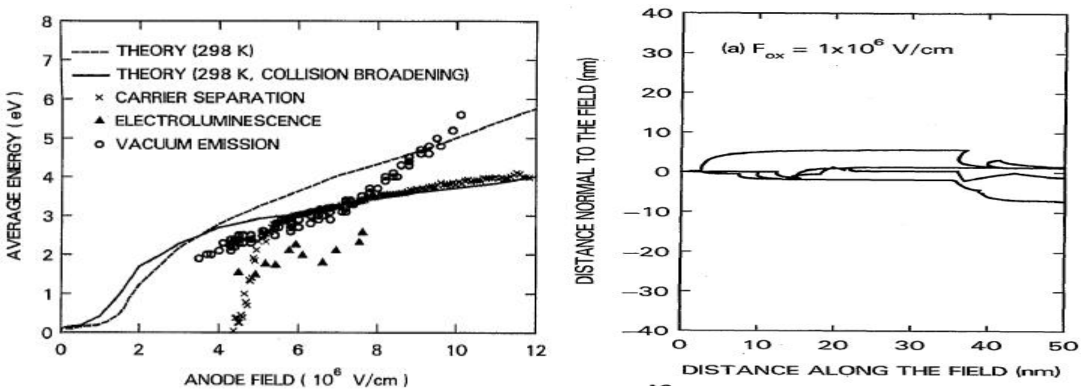

Fig. 2: LO-phonon scattering run-away connected to dielectric breakdown and Small polaron of time-of-flight experiments

Although there few drawbacks for which the consequences of injecting hot electrons in constant voltage scaled MOSFET were in the highlights. The two main problems understood the origin/spectrum of hot carrier and the nature/process of damage generation. Practical problems were also revealed pointing the unnecessary and expensive burn in and Wall Street big glitch in 1994. For some digital circuits, a figure of merit for MOSFET's for unloaded circuit is CV/I, where C the gate capacitance is, V is the voltage swing and I is the current drive of the MOSFET. For the loaded circuits, the current drive of the MOSFET is of paramount importance. Historical data indicate the scaling the MOSFET channel length improves circuit speed as suggested by scaling theory. Figure 1 shows how the injection of electrons affects the scattering runaway to dielectric breakdown. The off-current specification for CMOS has been rising rapidly to keep the speed performance high. While $1nA/\mu m$ was the maximum off-current allowed in the late 1990's, off currents in excess of $100nA/\mu m$ are proposed today.

Keeping in mind both CV/I metric and the benefits of a large current drive, we note that device performance maybe improved by 1) inducing a larger charge density for a given gate voltage drive; 2) enhancing the carrier transport by improving the mobility, saturation velocity o ballistic transport; 3) ensuring device scalability to achieve a shorter channel length and 4) reducing parasitic capacitances and parasitic resistances.

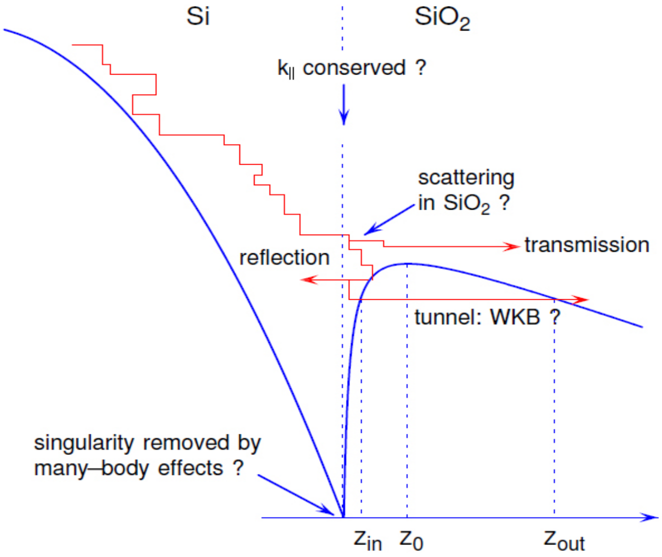

Fig. 3: Electron injection in SiO_2

IV. MOSFET GATE STACK

The reduction in the gate dielectric thickness is required for continuous device scaling. This has actually two different considerations: controlling the short channel effect and achieving a high current drive by keeping the amount of charge induced in the channel large as the power supply voltage decrease. It is the electrical thickness that is significant. The electrical inversion is determined by the series combination of three capacitances in the gate stack: the depletion capacitance of the gate electrode, the capacitance of the gate dielectric and the capacitance of the inversion layer in the silicon.

In the contrast, the direct tunneling current through the gate dielectric grows exponentially with decreasing physical thickness of the gate dielectric [7]. The tunneling current has a direct impact on the standby power of the chip and puts a lower limit on unabated reduction of the physical thickness of the gate dielectric. It is likely that tunneling currents arising from silicon dioxides (SiO_2) thinner than $0.8nm$ cannot be tolerated, even for high performance systems [8]. High dielectric constant gate dielectrics and metal gate electrodes were explored through the introduction of new materials. Figure 4 shows the depletion capacitance of the electrode, the capacitance of the gate dielectric, and the capacitance of the inversion layer in the silicon.

V. HIGH k GATE DIELECTRIC

A gate dielectric with a dielectric constant k substantially higher than that of SiO_2 (k_{ox}) will achieve a smaller equivalent electrical thickness (t_{eq}) than the SiO_2, even with a physical thickness (t_{phys}) larger than that of the SiO_2 (t_{ox}): $t_{eq} = (\frac{k_{ox}}{k})t_{phys}$

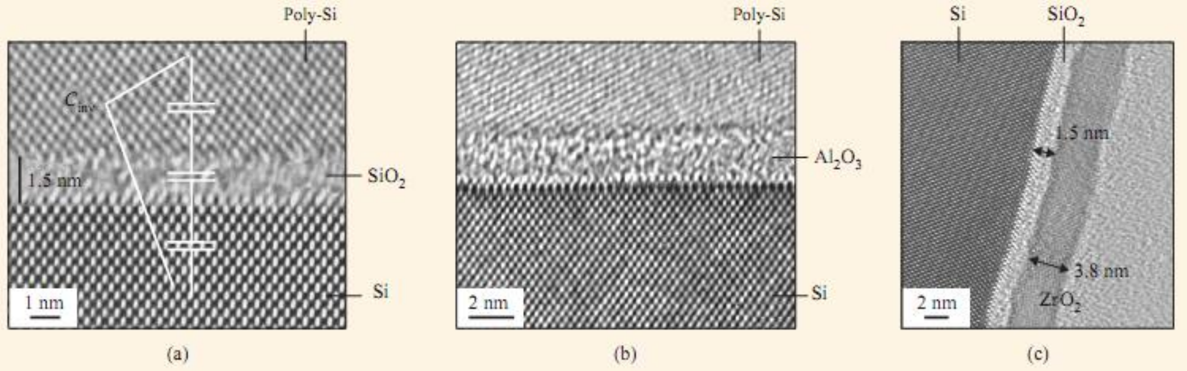

Fig 4. a) Transmission electron micrograph (TEM) of a conventional silicon dioxide (oxynitride) with a physical thickness of $1.5nm$. b) TEM of a $2.2nm$ Al_2O_3 with an equivalent electrical thickness of $1nm$. C) TEM of a $3.8nm$ ZrO_2 on an $1.5nm$ interfacial silicon dioxide. Adapted with permission from Gusev et al. 2001 IEEE.

It is not that simple to replace SiO_2 with a material having the same dielectric constant. Thermal stability with respect to silicon is more important consideration, since high temperature anneals are generally employed to activate dopants in the source/drain as well as the polysilicon gate. Although many binary and ternary oxides are predicted to be thermally stable with respect to silicon [9], recent research on high dielectric constant gate insulators have focused primarily on metal oxides such as Ta_2O_5, Al_2O_3, La_2O_3, HfO_2 and GdO_3 and their silicates [10]. Large silicon to insulator energy barrier height is exponentially dependent on the square root of the barrier height [11]. Hot carrier emission into the gate insulator is also related to the same barrier height [12]. The high k material should therefore not only have a large bandgap but also have a band alignment which results in a large barrier height.

VI. METAL GATE ELECTRODE

Metal gate electrode has numbers of advantages compared to the doped polysilicon gate used almost exclusive today. Due to the depletion of the doped polysilicon gate capacitance degrades for $0.4-0.5nm$ of the equivalent oxide thickness of the total gate capacitance at inversion. Considering the gate equivalent oxide of less than $1.5nm$ at inversion, substantial amount like sub $50nm$ CMOS is required. Thermal instability may require the use of a low thermal budget process after the gate dielectric deposition. From a device design point of view, the most important consideration for the gate electrode is the work function of the material. When the polysilicon gate technology has somehow got locked in the gate work functions to values close to the conduction band and the valence band of silicon, the use of the metal gate material opens up the opportuinity to choose the work function of the gate and the redesign the device to achieve the best combination of work function and channel doping. A mid gap work functions results in either a threshold voltage that is too high for high performance applications or compromised short channel effects since the channel must be counterdroped to bring the threshold voltage down. For double gate FET's where the short channel effects are controlled by the device geometry, the threshold voltage is determined mainly by the gate work function [13-15]. Therefore, for double gate FET, the choice of the gate electrode is particularly important.

The requirements of a low gate dielectric/silicon interface state density and low gate dielectric fixed charges imply that a damage free metal deposition process like CVD instead of sputtering is required. The deposition process must not introduce impurities like traces of the CVD precursor materials into the gate stack. The thermal stability of the metal electrode must withstand the thermal anneals required to passivate at the silicon/gate dielectric interface after the metal deposition as well as the thermal processing of the back end metallization processes. Moreover, it is likely to be expected to have a low resistivity at least similar to conventional silicides such as $CoSi_2$, although this requirement may be relaxed by strapping the gate electrode of the proper work function with a lower resistivity material on top.

In the replacement gate technology [16], a dummy gate material is used for the formation of the self aligned gate to source/drain structure. As a result, the dummy gate material is removed and replaced with the desired gate dielectric and electrode [16]. In the other hand, the metal gate electrode may be etched in a way similar to the polysilicon gate technology. In addition, thermal stability issues from the source/drain dopant activation anneal must be addressed. In both of the cases, if metals with two different work functions are employed for n-FET and p-FET, respectively the integration of n-FET and p-FET in a CMOS process remains a challenge. Since 1) the deposition of the metals for n-FET and p-FET must be done separately and 2) one must find a way to strap the two different metals in a compact way to connect the n-FET and p-FET gates.

VII. DOUBLE GATE FET (DGFET) AND ELECTROSTATIC

In the early 1980's, double gate FET was introduced for the first time. Many groups explored the concept both experimentally and theoretically [18]. The Monte Carlo and drift diffusion modeling work by Fiegna at al. [17] and Frank at al. [19] clearly showed that a DGFET can be scaled to a very short channel thickness about 15nm while achieving the expected performance derived from scaling. Although the initial work focused on the better scalability of DGFET, current researches suggest that the scalability advantage may not be as large as previously envisioned [20, 21], but the carrier transport benefits may be substantial.

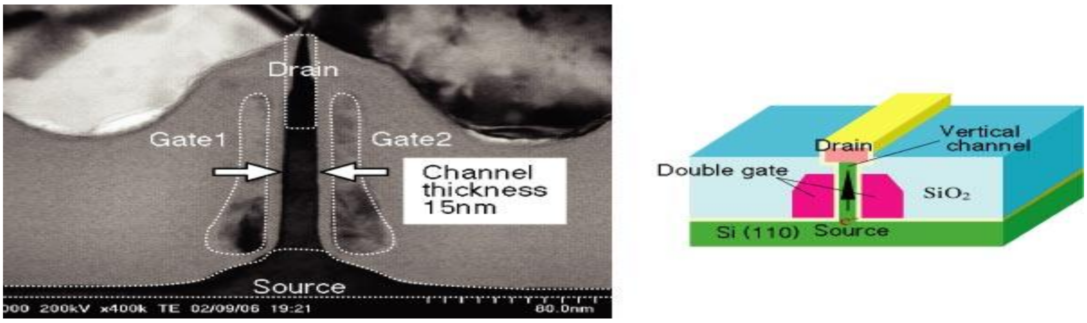

Fig. 5: Scaling- Electrostatic integrity: Double gate FET

DGFET has the unique features like [24] 1) short channel control effects by device geometry,a s compared to bulk FET where short channel effects are controlled by doping and 2) a thin silicon channel leading to tight coupling of the gate potential with the channel potential. These features provide potential DGFET advantages like reduced 2D short channel effects leading to a shorter allowable channel length to bulk FET and a sharper subthreshold slope like $60mV/dec$ compared to > $80mV/dec$ for bulk FET which allows for a greater gate override for the same power supply and the same off current and better carrier transport as the channel doping is reduced. When the channel doping is reduced, it relieves a significant scaling limitation due to the drain to body band to band tunneling leakage current. Hence, there is more current drive per device area, and this density improvement critically depends on the specific fabrication methods employed and is not intrinsic to the device structure. DGFET can be switched with its two gates simultaneously. The one gate can be switched only and another one is used to apply bias to dynamically alter the threshold voltage of the FET [22, 23]. A thin gate dielectric at the nonswitching gate reduces the voltage required to specify the threshold voltage and preserves the drain field shielding advantage of the double gate device structure. Moreover, a thinner gate dielectric also means extra capacitance that does not contribute to channel charge for switching. To evaluate the scalability of FET's, the concept of the "scale length" for a MOSFET is useful [24, 25, 26]. The electrostatic potential of the MOSFET channel can be approximated by analytically solving the 2D Laplace equation using the superposition principle and the short channel behavior can be described by a characteristic "scale length." [27]. By the amount of 2D short channel effects, the minimum gate length can be determined. From the figure 7, it can be seen that the trend of these 2D effects as the channel length is decreased with respect to the scale length of the MOSFET. With the same scaling formation, figure 6 shows the electrostatic integrity of Si nanowire transistors where 10nm Al_2O_3 blocking layer has been injected. SiO_2 layer is still present for the predetermined presence of nanowires expressing the 2D electrostatic behavior. With typical tolerance of 20-30% gate length variation, an L/λ of 1.5 is required. Conventional short channel effect theory [28] correlates the junction depth to the shorter channel effects. For DGFET, consideration of junction depth is moot, since the 2D electrostatic behavior is controlled by the thickness of the silicon channel instead of the junction depth.

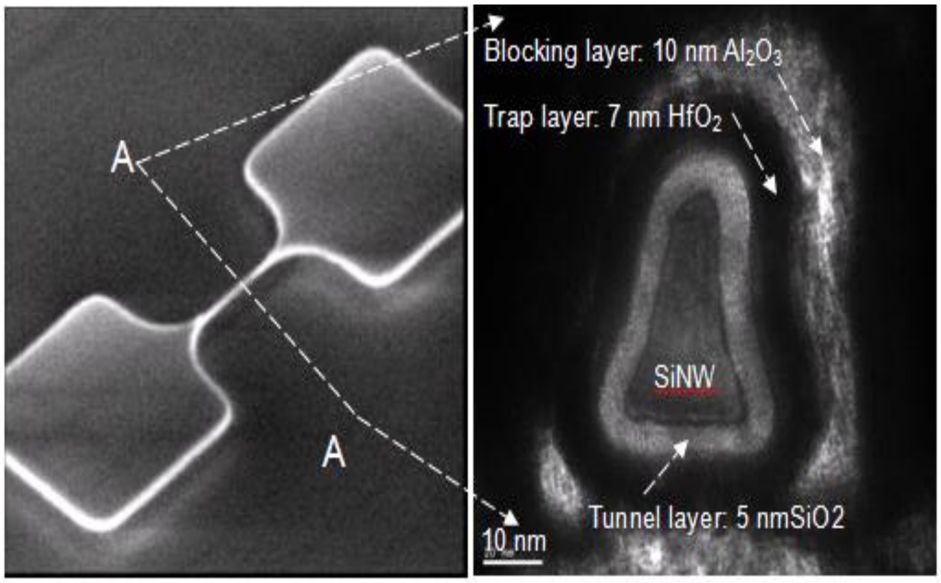

Fig. 6: Scaling- Electrostatic integrity: Si Nanowire Transistors

But the steepness of the source/drain junction is still an important consideration as in the case of bulk FETs [21]. Figure 7 illustrates the threshold voltage roll-off characteristics of the DGFET with lateral junction profile gradients of 2, 4 and 6 nm which is known as Gausian analytical profile.

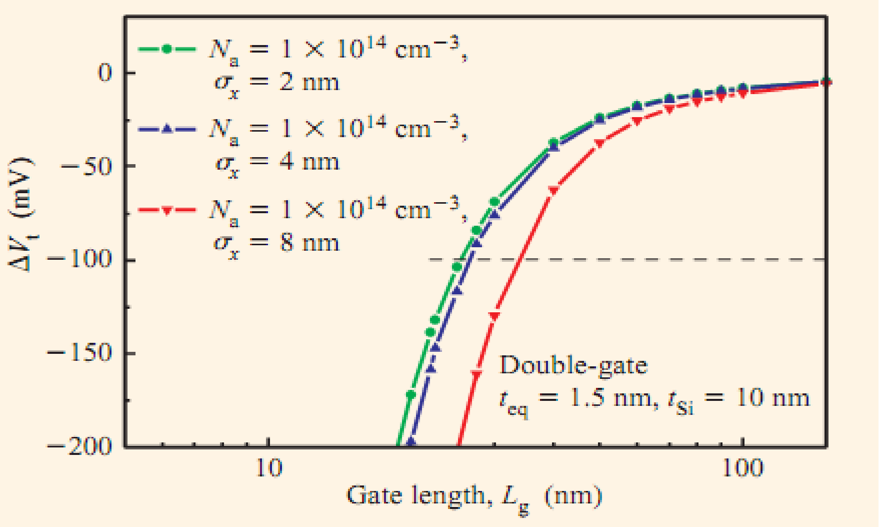

Fig. 7: Threshold voltage roll-off characteristics of double gate FET with different junction gradients, illustrating the importance of maintaining a sharp doping profile for DGFET even though the junction depth is no longer important for DGFET. The silicon channel thickness t_{Si} is 10nm and the equivalent gate dielectric thickness t_{eq} is 1.5 nm.

VIII. SCALING kT/q AND THE PROBLEM:

It took the first power crisis in the 1980 while CMOS technology was invented, caused VLSI chips to switch from nMOS which during the late 1970s was the dominant VLSI technology. During the period V_{dd} was fixed to 5V, and was not scaling with technology to maintain system compatibility. The depletion thresholds for the nMOS loads did not scale rapidly, so the current per minimum gate scaled only slowly. The power of the chips started to grow with the complexity and chips rapidly went from a watt to multiple watts with the final nMOS VLSI chips dissipating over 10W [29]. While the peak currents in CMOS were as large as nMOS, since they were transients that lasted roughly 1/20 of a clock cycle, a CMOS processor ran at roughly 10x lower power than a similar nMOS chip. Figure 8 uses microprocessor data to track CMOS technology scaling since the mid 1980 to today. Through four generations of technology, from the $2\mu m$ generation in the early 1980s to the $0.5\mu m$ generation in the mid 1990s, the power savings from switching to CMOS was large enough that V_{dd} did not need to scale and was kept constant at 5V.

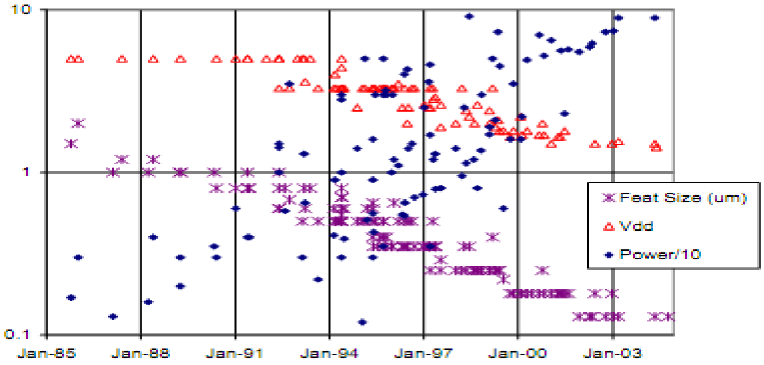

Fig. 8: Microprocessor V_{dd} , Power/10, and feature size versus year from 1994 to today V_{dd} has roughly tracked feature size

Power continued to increase during this time. Part of this increase in power was due to increase in area but power density increased by 30x during this period as well. This was due to the performance optimizations such as improved circuit design, better sizing optimization and deeper pipelines.

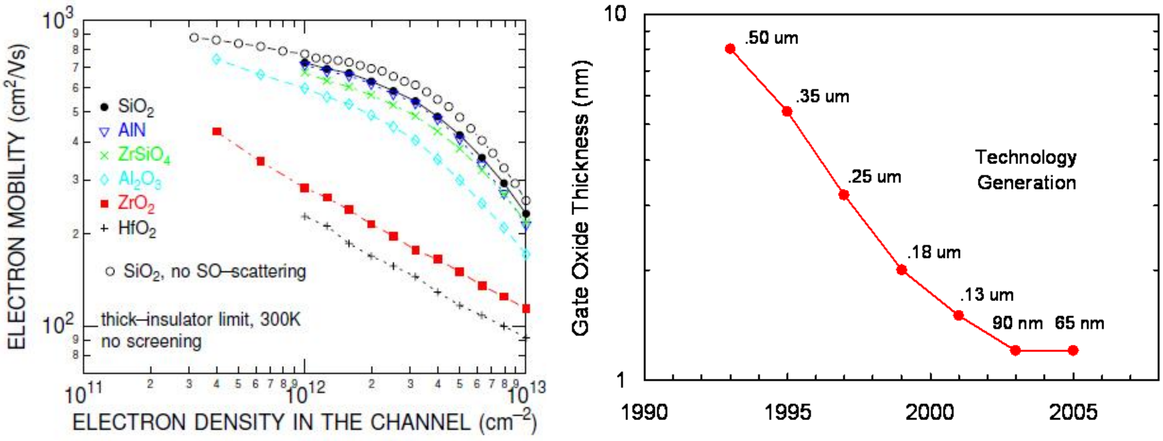

Fig. 9: a) Scaling- reduce leakage, low mobility in high k MOS systems scattering with interfacial optical phonons, b) Scaling- reduce leakage: gate oxide scaling at Intel

Figure 9 (a) illustrates the accepted value of off-leakage increasing for $I_{off} / I_{on} \approx 10^{-4}$ for the 32nm mode and the electrostatic integrity stands for junction leakage and gate leakage. For the gate leakage high k insulators such as HfO_2, ZrO_2, Al_2O_3 etc, electron mobility decreases as the electron density increases. Figure 9 (b) shows the variation of gate oxide thickness as the generations meet up their new challenges through years.

IX. INCREASED IMPROVMNETS IN SCALING AND HIGHER MOBILITY

The double gate FET carrier transport pointed out its importance of a low doped channel for carrier transport in DGFET. A higher carrier mobility and saturation velocity can be found through the choice of material for the FET channel. Fischetti and Laux [89] compared the performance of several semiconductors that have high carrier mobilities and saturation velocities including Ge, InP, InGaAs, GaAs etc. These materials provide a significantly higher carrier mobility which give only a moderate performance advantage over a lower mobility material such as silicon. The band structure which determines the density of states like the inversion capacitance [30] and the carrier scattering rates at high carrier energies are just as important as the carrier mobility.

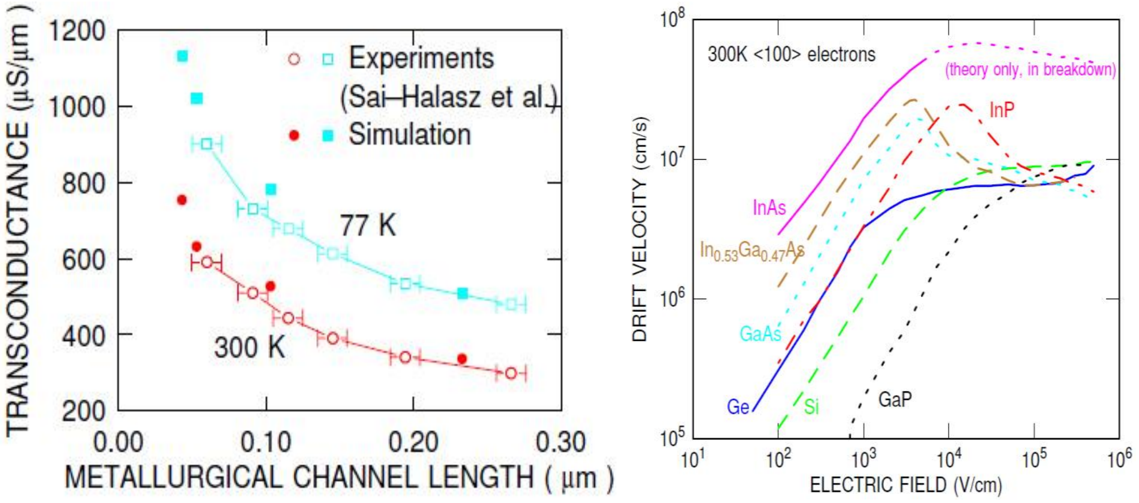

Fig. 10 a) Scaling improve performance – transconductance varies with the temperature with channel length, b) high velocity and low effective mass semiconductors

The carrier mobility in silicon under biaxile tensile strain is enhanced. [31-36]. The most commonly cited reason for electron mobility enhancement in strained silicon is that under the biaxile tensile strain, the sixfold degeneracy of the conduction band of silicon is lifted, raising the higher effective mass fourfold degenerate ellipsoids. The use of strained silicon provides a trustworthy trade off between moderate levels of performance enhancement over silicon an ease of fabrication and integration with silicon as compared to other higher mobility materials such as Ge, InGaAs, InAs, GaAs and InP that has been shown in figure 10 (b). Recent work provided promising experimental evidence that introducing the biaxial tensile strained silicon through a layer of relaxed SiGe may provide adequate performance gains for incorporation into conventional CMOS technologies.

Another improvement can be made by stretching the silicon atoms beyond their normal interatomic distance. This can be done by putting the layer of silicon over a substrate of silicon germanium (SiGe). In the silicon layer atoms align with atoms underlying silicon germanium layer, so the links between the silicon atoms become stretched thereby approach to the formation of strained silicon. Figure 11 shows the scaling improves performance with strained silicon that has been performed to make IBM 32nm strained silicon nFET on silicon germanium virtual substrate and Intel 45nm strained silicon pFET.

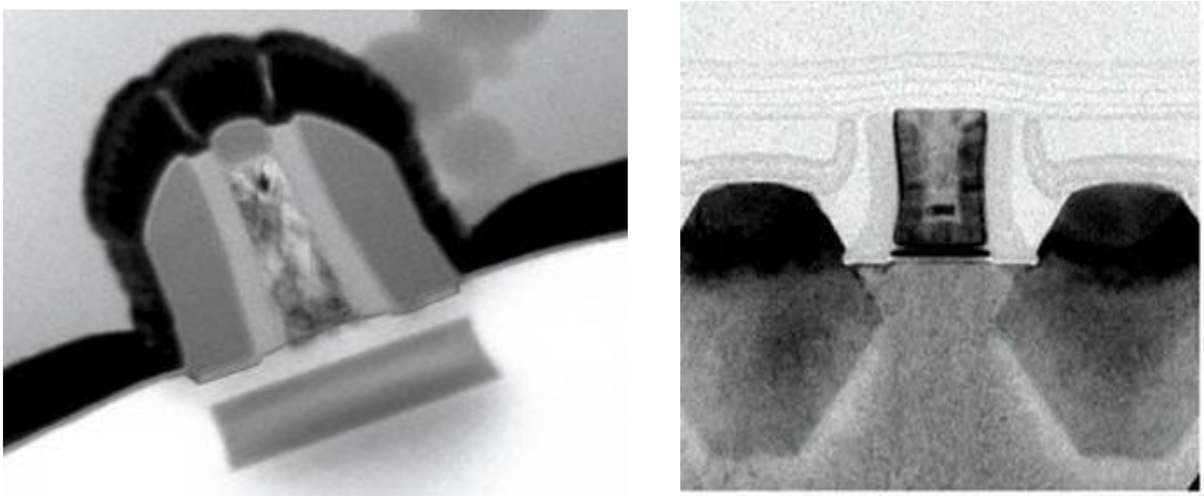

Fig. 11: Scaling- Improve performance: Strained Si a) IBM 32 nm strained (tensile) Si nFET on SiGe virtual substrate, b) Intel 45 nm strained (compressive) Si pFET with regrown SiGe S/D

X. OPTIMIZATION PERSPECTIVE

Let us assume that there is an attempt to try all the different ways to build a unit using all possible transistor sizes, circuit methods and supply and threshold voltages. The optimal design point depends on the application constraints like maximum power or minimum performance requirements, but will always lie on the lower right edge of the feasible set that forms the Pareto optimal points.

Figure 12 (a) shows the result of plotting all of the solutions on a graph with performance on one axis and the energy consumed for a single operation on the other. Figure 12 (b) estimates the energy performance trade offs using published microprocessor data. While a complete optimizer does not exist, tools that optimize a subset of the parameters exist. The result of a tool is a sized circuit, and the optimal values of V_{dd} and V_{th} to use for the circuit. Table 1 shows the optimal result of the voltages with respective sensitivity.

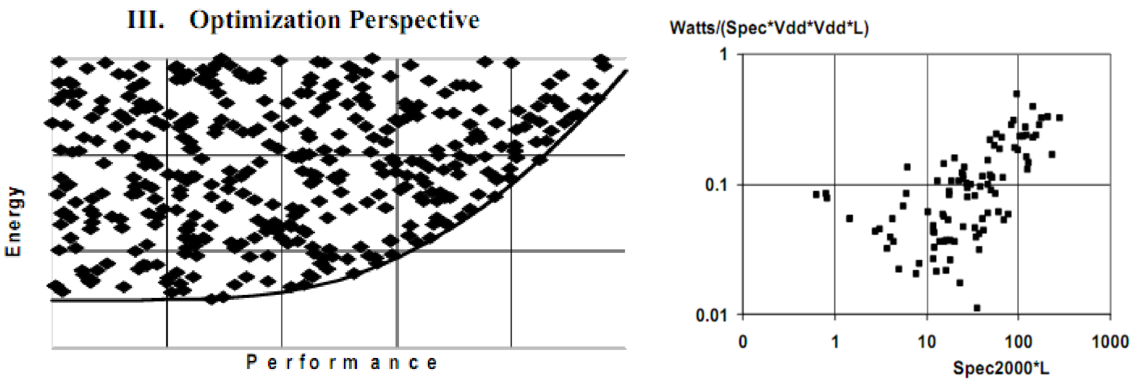

Fig. 12 a) The optimal curve is the boundary of the space of all possible solutions in the Energy Performance plane, b) Energy consumed per operation for CMOS processors built during the past 20 years.

V_{dd}	nMOS V_{th}	Sensitivity $\frac{\partial E / \partial V_{dd}}{\partial Perf. / \partial V_{dd}}$
550 mV	321 mV	0.031
700 mV	189 mV	0.194
850 mV	183 mV	0.7633
1 V	182 mV	1.8835

Table 1: Optimal V_{dd} , V_{th} and sensitivity for a 90 nm inverter at $80^0 C$ with 20% activity factor driving a fixed capacitive load.

XI. LOW POWER CIRCUITS AND ARCHITECTURE

A technique with moderate performance cost might be well suited for a low speed machine with a large marginal delay cost per unit energy, but would actually make the power higher if it was applied to a fast machine with a small marginal delay cost for lower energy consumption. The energy reduction technique generally involves problem reformulation or algorithmic changes that allow the desired task to be accomplished with less computation than before. These techniques can change the power required for a task by orders of magnitude [37], more than any other method. Before power became a critical problem, designers were rarely concerned whether a unit was doing useful work; they were only concerned about functionality and performance. The larger output reductions come from applying this idea at the system level. Subsystems often support different execution states, from powered off, to ready-to-run. Modern PCs use an interface called ACPI to allow the software to deactivate unused units so that they don’t dissipate power [38]. The reducing of energy with no performance cost are techniques that improve performance with no energy cost. For applications with data parallelism, it is possible to use two functional units each running at half rate rather than using a single unit running a full rate. As the energy per operation is lower as one decreases performance, this parallel solution will dissipate less power than the original solution. Most of the remaining low power techniques are really methods of dealing with application, environmental or fabrication uncertainty, so the energy cost of variability should be considered.

XII. CONCLUSION

In the scaling, power has always been a concern. Rise in the power levels of nMOS VLSI chips in the 1980s caused the industry to switch to CMOS. In the early 1990, power became the issue of talking in designing of CMOS; many approaches were there to reduce the growing power of VLSI systems. Energy efficiency of technology scaling, and system level optimization were the most successful approaches for reduction in the reduced computation. One thing should be kept in mind that power and performance are integrally connected for reducing chip power. By reducing the performance, power can be lowered but the technique is to lessen the energy without affecting the circuit's performance. Power growth must be addressed by application specific system level optimization. Unless they become impractical, conventional devices and materials will continue to be used. In this paper, we review the approaches to circumvent or surmount the barriers to device scaling. Discussing the new materials and new device structures, we showed innovations of materials for the gate stack and transistor channel. Double gate FET structure has also been shown. Gradually approaching the facts of dopant profile control and contact formation, unconventional to conventional technologies have been employed. In the silicon microelectronics technology, as nanotechnology may be seen successful, it is proven that it will be many years before nanotechnology can reach the level of maturity of the current silicon technology. As been seen, in the near future, there will be a gigantic shift of microelectronics to nanotechnology, hence, at present; it is somehow exposed by recent researches to make plenty of applications for the continuous technological progress.

REFERENCES

[1] *International Technology Roadmap for Semiconductors, 1999 Edition, Semiconductor Industry Association (SIA),* Austin, Texas: SEMATECH, USA, 2706 Montopolis Drive, Austin, Texas 78741; http://www.itrs.net/ntrs/publntrs.nsf.

[2] G. Moore, "Progress in Digital Integrated Electronics," *IEDM Tech. Digest*, pp. 11–13 (1975)

[3] R. Dennard, F. H. Gaensslen, H. N. Yu, L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *IEEE, J. Solid State Circuits* SC-9, 256 (1974).

[4] B. Hoeneisen and C. A. Mead, "Fundamental Limitations in Microelectronics—I MOS Technology," *Solid State Electron*. 15, 819 (1972).

[5] S. Asai and Y. Wada, "Technology Challenges for Integration Near and Below 0.1 nm," *Proc. IEEE* 85, 505 (1997).

[6] H.-S. P. Wong, D. Frank, P. M. Solomon, H.-J. Wann, and J. Welser, "Nanoscale CMOS," *Proc. IEEE* 87, 537 (1999).

[7] S.-H. Lo, D. Buchanan, Y. Taur, and W. Wang, "Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFETs," *IEEE Electron Device Lett*. 18, 209–211 (1997)

[8] D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H.-S. Wong, "Device Scaling Limits of Si MOSFETs and Their Application Dependencies," *Proc. IEEE 89*, 259 –288 (2001).

[9] HR. W. Keyes, "Fundamental Limits of Silicon Technology," *Proc. IEEE 89*, 227–239 (2001).

[10] K. Hubbard and D. Schlom, "Thermodynamic Stability of Binary Oxides in Contact with Silicon," *J. Mater. Res. 11*, 2757 (1996).

[11] E. Gusev, D. Buchanan, E. Cartier, A. Kumar, D. DiMaria, S. Guha, A. Callegari, S. Zafar, P. Jamison, D. Neumayer, M. Copel, M. Gribelyuk, H. Okorn-Schmidt, C. D'Emic, P. Kozlowski, K. Chan, N. Bojarczuk, L.-A. Rannarsson, P. Ronsheim, K. Rim, R. Fleming, A. Mocuta, and A. Ajmera, "Ultrathin High-k Gate Stacks for Advanced CMOS Devices," *IEDM Tech. Digest*, pp. 451– 454 (2001).

[12] J. Robertson, "Band Offsets of Wide-Band-Gap Oxides and Implications for Future Electronic Devices," *J. Vac. Sci. Technol*. B 18, 1785–1791 (2000).

[13] J. Hauser, "Gate Dielectrics for Sub-100 nm CMOS," in IEDM Short Course: Sub-100 nm CMOS, M. Bohr, Ed., *IEDM Tech. Digest* (1999)

[14] Y.-S. Suh, G. Heuss, H. Zhong, S.-N. Hong, and V. Misra, "Electrical Characteristics of TaSi, Ny Gate Electrodes for Dual Gate Si-CMOS Devices*," Symposium on VLSI Technology, Digest of Technical Papers*, 2001, pp. 47– 48.

[15] D.-G. Park, K.-Y. Lim, H.-J. Cho, T.-H. Cha, J.-J. Kim, J.-K. Ko, I.-S. Yeo, and J.-W. Park, "Novel Damage-Free Direct Metal Gate Process Using Atomic Layer Deposition," *Symposium on VLSI Technology, Digest of Technical Papers*, 2001, pp. 65– 66.

[16] A. Chatterjee, R. Chapman, G. Dixit, J. Kuehne, S. Hattangady, H. Yang, G. Brown, R. Aggarwal, U. Erdogan, Q. He, M. Hanratty, D. Rogers, S. Murtaza, S. Fang, R. Kraft, A. Rotondaro, J. Hu, M. Terry, W. Lee, C. Fernando, A. Konecni, G. Wells, D. Frystak, C. Bowen, M. Rodder, and I.-C. Chen, "Sub-100 nm Gate Length Metal Gate NMOS Transistors Fabricated by a Replacement Gate Process*," IEDM Tech. Digest*, pp. 821– 824 (1997).

[17] C. Fiegna, H. Iwai, T. Wada, T. Saito, E. Sangiorgi, and B. Ricco, "A New Scaling Methodology for the 0.1– 0.025nm MOSFET," *Symposium on VLSI Technology, Digest of Technical Papers*, 1992, p. 33.

[18] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-Gate Silicon-on-Insulator Transistor with Volume Inversion: A New Device with GreatlyEnhanced Performance," *IEEE Electron Device Lett*. 8,410 (1987)

[19] D. Frank, S. Laux, and M. Fischetti, "Monte Carlo Simulation of a 30nm Dual-Gate MOSFET: How Far Can Si Go?," *IEDM Tech. Digest*, p. 553 (1992).

[20] D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H.-S. Wong, "Device Scaling Limits of Si MOSFETs and Their Application Dependencies," *Proc. IEEE* 89, 259 –288 (2001)

[21] Y. Taur, C. Wann, and D. J. Frank, "25 nm CMOS Design Considerations," *IEDM Tech. Digest*, pp. 789 –792 (1998).

[22] I. Yang, C. Vieri, A. Chandrakasan, and D. Antoniadis, "Back-Gated CMOS on SOIAS for Dynamic Threshold Voltage Control," *IEEE Trans. Electron Devices* 44, 822 (1997).

[23] H.-S. Wong, D. Frank, and P. Solomon, "Device Design Considerations for Double-Gate, Ground-Plane, and Single-Gated Ultra-Thin SOI MOSFET's at the 25 nm Channel Length Generation," *IEDM Tech. Digest*,p. 407 (1998).

[24] R. Yan, A. Ourmazd, and K. Lee, "Scaling the Si MOSFET: From Bulk to SOI to Bulk*," IEEE Trans. Electron Devices* 39, 1704 (1992)

[25] K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, and Y. Arimoto, "Scaling Theory for Double-Gate SOI MOSFET's," *IEEE Trans. Electron Devices* 40, 2326 (1993).

[26] D. Frank, Y. Taur, and H.-S. P. Wong, "Generalized Scale Length for Two-Dimensional Effects in MOSFET's," *IEEE Electron Device Lett*. 19, 385 (1998).

[27] C. Y. Chang and S. M. Sze, Eds., ULSI Devices, Wiley, New York, 2000, Ch. 3

[28] S. M. Sze, Physics of Semiconductor Devices, Wiley, New York, 1981

[29] M. Forsyth, W.S. Jaffe, D. Tanksalvala, J. Wheeler, and J. Yetter, "A 32-bit VLSI CPU with 15-MIPS Peak Performance," *IEEE Journal of Solid-State Circuits*, Oct. 1987.

[30] S. Takagi, M. Takaytanagi-Takagi, and A. Toriumi, "Accurate Characterization of Electron and Hole Inversion-Layer Capacitance and Its Impact on Low Voltage Operation of Scaled MOSFETs," *IEDM Tech. Digest*, pp. 619 – 622 (1998)

[31] T. Vogelsang and H. R. Hofmann, "Electron Transport in Strained Silicon Layers on Six Gex Substrates," *Appl. Phys. Lett*. 63, 186 (1993)32.

[32] D. Nayak, J. Woo, J. Park, K. Wang, and K. MacWilliams, "High-Mobility p-Channel Metal-Oxide-Semiconductor Field-Effect Transistors on Strained Si," *Appl. Phys. Lett.* 62, 2853–2855 (1993)

[33] J. Welser, J. Hoyt, S. Takagi, and J. Gibbons, "Strain Dependence of the Performance Enhancement in Strained-Si n-MOSFETs," *IEDM Tech. Digest, pp*. 373– 376 (1994).

[34] M. Fischetti and S. Laux, "Band Structure, Deformation Potentials, and Carrier Mobility in Strained Si, Ge, and SiGe Alloys," *J. Appl. Phys*. 80, 2234 (1996).

[35] S. Tiwari, M. Fischetti, P. Mooney, and J. Welser, "Hole Mobility Improvement in Silicon-on-Insulator and Bulk Silicon Transistors Using Local Strain," *IEDM Tech. Digest*, pp. 939 –941 (1997)

[36] K. Rim, J. Hoyt, and J. Gibbons, "Transconductance Enhancement in Deep Submicron Strained-Si n-MOSFETs," *IEDM Tech. Digest*, p. 707 (1998)

[37] N. Zhang and R. Brodersen, "The cost of flexibility in systems on a chip design for signal processing applications," http://bwrc.eecs.berkeley.edu/Classes/EE225C/Papers/arch_design. doc, 2002.

[38] "*Advanced Configuration and Power Interface Specification*," Hewlett-Packard Corp., Intel Corp., Microsoft Corp., Phoenix Technologies Ltd., and Toshiba Corp., http://www.acpi.info/ DOWNLOADS/ACPIspec30.pdf, Sept. 2004.