Cmos Scaling: Present, Past and Future

Tanmay Sinha Roy*, Sk Babul Akhtar, Shreya Adhikary, Tomal Suvro Sannyashi, Debasis Mondal
Department of Electronics and Communication Engineering

Swami Vivekananda University, Barrackpore, West Bengal, India

**Corresponding Author*

ABSTRACT

The evolution of Complementary Metal Oxide Semiconductor (CMOS) technology plays a crucial role in modern advancements. As CMOS scales down beyond 22nm towards 7nm, it encounters various challenges and opportunities in design. This miniaturization is examined through scaling theory, with a focus on issues such as performance, power consumption, cost-effectiveness, technological constraints, and reliability. The anticipated breakthrough in overcoming the 5nm physical gate length barrier is expected by 2026, thanks to the use of High-k materials, which also help mitigate current leakage issues. Additionally, lithography technology is a key process driving transistor downsizing, with several concerns being addressed regarding performance, power consumption, materials, cost, and technological limitations.

***Key words*:** 7nm Gate Length, CMOS, Downsizing, High-k, VLSI.

Introduction

The advancement of CMOS technology into the nanometer scale has become a critical issue in modern integrated circuits (ICs) (Akter et al. 2008a, b; Reaz et al. 2007a, b; Marufuzzaman et al. 2010; Reaz et al. 2003; Reaz et al. 2005; Iwai, 2012). Today's cutting-edge communication and engineering technologies would be inconceivable without the significant progress in integrated circuits (Iwai, 2003; Reaz et al. 2006; Reaz and Wei 2004; Mohd-Yasin et al. 2004; Mogaki et al. 2007). Moreover, everyday activities, manufacturing, commerce, transportation, medical care, and education all depend on CMOS technology (Iwai, 2008). Consequently, the evolution of CMOS technology is crucial for both the semiconductor industry and the global economy. Figure 1 illustrates the evolution of electronic circuits in relation to component dimensions (Iwai and Ohmi, 2002).

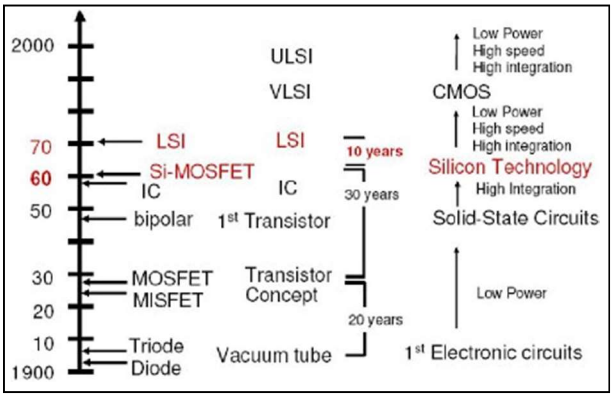

Fig. 1: The downsizing of component dimension.

CMOS downsizing offers several advantages, including improved performance, reduced power consumption, increased density, and lower costs (Kin and Park, 2011). New LSI products quickly replace previous-generation products in the market. This paper reviews the evolution of CMOS downsizing technology, examining scaling theory and addressing limitations related to performance, power usage, economic factors, technological challenges, and reliability.

Background:

For over 40 years, the semiconductor industry has been characterized by rapid advancements in its products. In 1965, Gordon E. Moore, co-founder of Intel, predicted that the number of transistors on cutting-edge integrated circuits would approximately double every two years without a corresponding increase in chip costs (Moore, 1965). Figure 2 illustrates Moore's Law, which tracks the growth in transistor numbers in Intel's microprocessors. This prediction has largely held true, with the number of transistors increasing from 2,250 in Intel's 4004 (1971) to 731 million in the Intel Core i7 (2008).

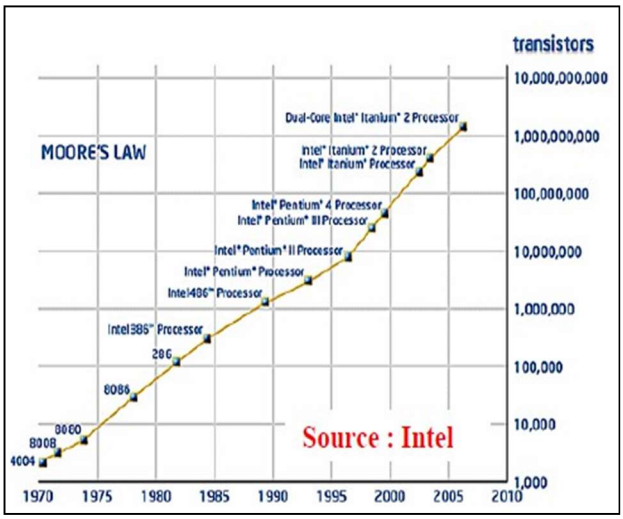

Fig. 2: Moore's Law in Microprocessors.

The remarkable advancement in microprocessors has been largely driven by the continuous downsizing of metal oxide semiconductor (MOS) field effect transistors. As these transistors shrink, they become less expensive, consume less power, operate faster, and enable more functions per unit area of silicon. This miniaturization of silicon ICs enhances performance while reducing the cost per function. In theory, Dennard's ideal scaling approach improves performance and integration significantly without a notable increase in power consumption, provided the chip area remains constant (Dennard et al. 1972). However, the actual scaling progress over the past 30 years since 1970 has been more aggressive. Table 1 illustrates the advancements in ICs with examples of each trend.

Table 1: Improvement trends for ICs enabled by feature scaling.

Trend	Example
Integration Level	Components/ Chips, Moore's law
Cost	Cost Per Function
Speed	Microprocessor throughput
Power	Laptop or cellphone battery life
Compactness	Small and light weight products
Functionality	Non-volatile memory, imager

The core technology driving this evolution is CMOS downsizing. MOS transistors, which are fundamental to information infrastructure, have enabled increasingly dense and faster integration through their scaling. Figure 3 illustrates the fundamental structure of

CMOS technology.

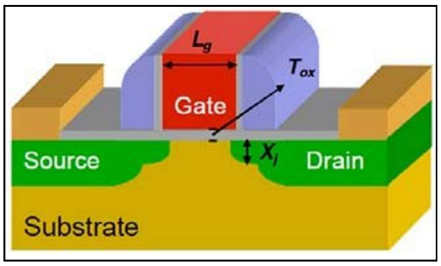

Fig. 3: The structure of CMOS.

The primary concept behind downsizing is to reduce all transistor dimensions by a specific factor, . Along with size reduction, certain parameters must be adjusted accordingly. As the channel length decreases, performance enhances, power consumption per switch decreases, and density increases. Figure 4 illustrates CMOS downsizing, and Table 2 provides a summary of the CMOS parameters involved in this process.

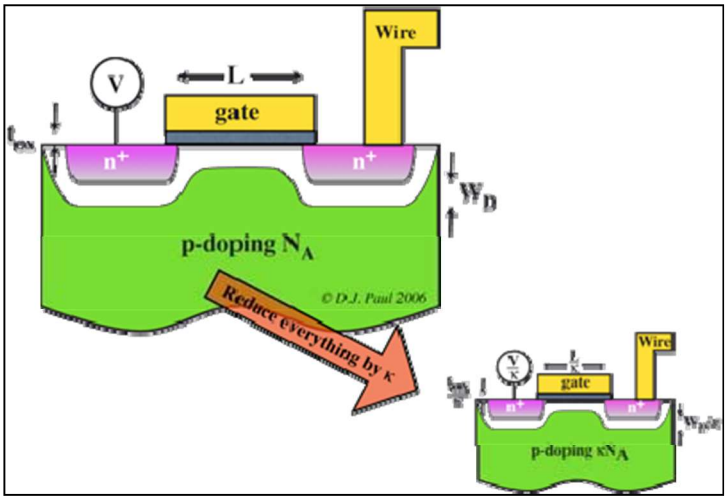

Fig. 4: The illustration of CMOS downsizing.

Table 2: The CMOS parameters involved in downsizing.

Parameter	Constant F downsizing	Generalized Downsizing
Device Dimension (L_g, W, t_{ox})	$1/\kappa$	$1/\kappa$
Doping Concentration	κ	$\alpha\kappa$
Voltage (V)	$1/\kappa$	α/κ
Electric Field ($\vec{F}$)	1	α

Carrier velocity (v)	1	1
Depletion Layer Width (W_D)	$1/\kappa$	$1/\kappa$
Capacitance, ($C = \varepsilon A/t_{ox}$)	$1/\kappa$	$1/\kappa$
Current (I)	$1/\kappa$	α/κ α/κ
Circuit Delay Time ($\tau \sim CV/I$)	$1/\kappa$	$1/\kappa^2$
Power Dissipation per Circuit ($P \sim IV$)	$1/\kappa^2$	α/κ
Power-delay product per circuit ($P\tau$)	$1/\kappa^2$	
Circuit Density ($\propto 1/A$)	$1/\kappa^2$ κ^2	
Power Density (P/A)	1	α^2

Discussion:

Evolution In Cmos Downsizing:

The evolution of CMOS downsizing has involved progressively reducing the size of components such as MOSFETs. Since the early 1970s, various limits have been proposed for downsizing. In the mid-1980s, a 1 µm limit was considered feasible due to anticipated issues with short-channel effects and optical lithography (Iwai, 2004). By the late 1990s, the limit was thought to be 0.25 µm due to increasing source/drain resistance, direct-tunneling leakage in gate oxides, and dopant fluctuations in the channel (Iwai, 2009). In the early 2000s, 100 nm was expected to be the limit due to challenges in further reducing MOSFET physical dimensions. However, these predicted limits have been exceeded, as evidenced by the successful development of commercial products with smaller MOSFET sizes.

Currently, the 22 nm node is the CMOS technology phase succeeding the 32 nm node. As illustrated in Figure 5, transistor performance decreased with the shrinking CMOS generations from 90 nm to 22 nm. Nonetheless, enhancements such as strain polysilicon in the 90 nm and 65 nm nodes, and strain plus high-k metal gates in the 45 nm and 32 nm nodes, have been implemented to continue advancing the transistor roadmap.

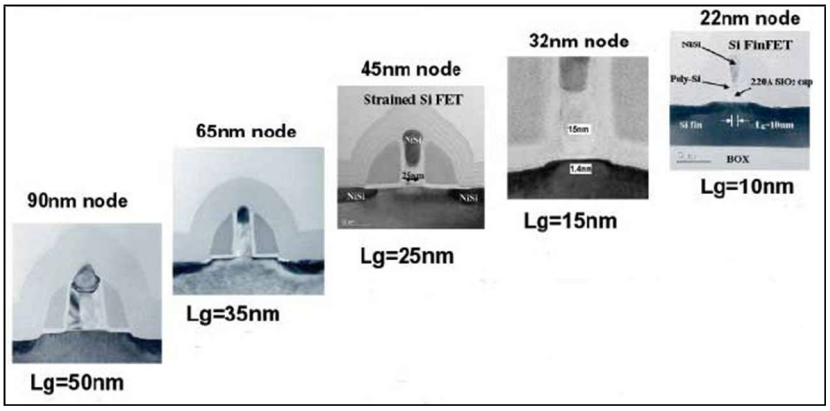

Fig. 5: Transistor Continuous Scaling.

The next milestone following the 22 nm gate length is expected to be a 7 nm gate length, as predicted by the International Technology Roadmap for Semiconductors (ITRS) for the decade up to 2018 (Iwai, 2009). The gate oxide thickness needs to be about two orders of magnitude smaller than the gate length, ideally around 0.7 nm. Currently, a 1.2 nm thick oxynitride film is used in manufacturing, with projections suggesting that the silicon dioxide equivalent thickness will decrease by 0.5 nm over the next decade (Iwai, 2004). While MOS transistors with a 0.8 nm oxynitride gate insulator have been demonstrated to function, there are anticipated challenges in developing LSIs with such thin gate insulators. Predicting the limits of further downsizing is difficult, and several potential limits are proposed as outlined in Table 3.

Table 3: Predicted limitations for downsizing.

Year	Expected Limit Size
1971	10 µm
1975	3 µm
1982	1.5 µm
1985	1 µm
1989	800 nm
1994	600 nm
1995	350 nm
1998	250 nm
1999	180 nm
2000	130 nm
2002	90 nm

2006	65 nm
2008	45 nm
2010	32 nm
2011	22 nm
2013	16 nm
2015	11 nm
2018	9 nm
2024	7 nm
Future (approximate 2026)	5 nm

Limitations in CMOS Downsizing:

ITRS has projected future scaling trends up to 2018, forecasting that the physical gate length will reach 7nm (Iwai and Wong, 2006). CMOS downsizing presents several integration challenges, including performance issues, power consumption concerns, economic challenges, and technological obstacles.

Performance Limitations:

The primary issue causing performance degradation in ultra-large scale circuits is interconnect delay, which arises from increased resistance and capacitance in narrow and densely packed metal lines (Kuhn, 2009). For example, when the width of copper wires is reduced to less than 100nm, the resistivity of the conductor increases due to surface scattering effects, as illustrated in Figure 6 (Iwai, 2007). According to scaling theory, the drain current per unit gate width should remain stable. However, a significant reduction in drain current per unit gate width has been observed in sub-100nm gate length MOSFETs (Iwai, 2007). This decline is attributed to the suboptimal MOSFET structures and processes currently in use.

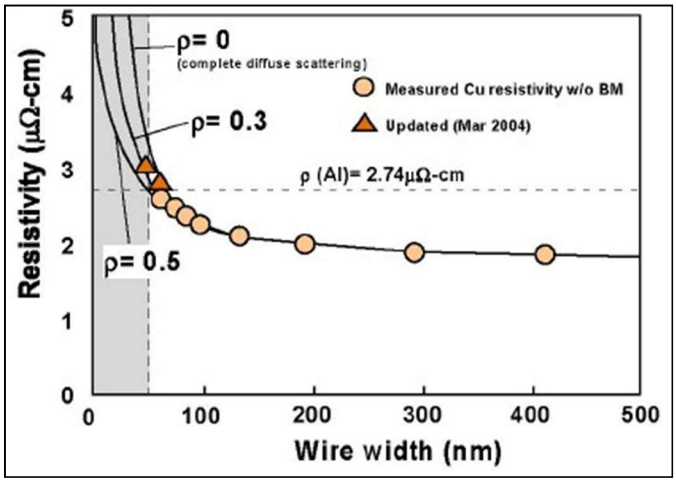

Fig. 6: Increase of resistivity because of surface scattering effect.

Currently, insufficient advancements in MOSFET structures and processes are seen as factors limiting progress and driving the need for new technologies. Innovations such as raised source-drain structures, plasma doping combined with flash or laser annealing, NiSisalicide, mobility enhancement through strained Si channels, silicon-on-insulator (SOI), three-dimensional structures, high-k gate insulators, and low-k interlayer dielectrics for interconnects are expected to address these issues and improve performance (Kuhn, 2011). Low-k materials are used in back-end processes but face challenges due to high mechanical and thermal stress during packaging. High-k materials could reduce current leakage by allowing for narrower dielectrics to maintain physical scaling (Kuhn, 2011). However, these materials' tendency to change properties at high temperatures remains a challenge that requires additional industrial processes.

In addition, advancements in circuit and system design architectures are also enhancing integrated circuit performance. Examples include parallel processing and optimized interconnects supported by steering devices. The system-on-chip (SoC) approach, which integrates DRAM with logic units, improves data transfer speeds between logic and memory (Haron and Hamdioui, 2010). Overall, it is expected that electronic system performance will improve, potentially down to 20nm or 10nm production scales, due to better tool technology and new system architectures. For sub-10nm gate length transistors, improvements in drain current levels are anticipated.

Material Limitations:

For the continued scaling of CMOS transistors, new materials are being proposed, as illustrated in Figure 7 (Haron and Hamdioui, 2010). Traditional materials like Silicon (Si), Silicon dioxide (SiO2), Aluminum (Al), Copper (Cu), and Salicides are limited by their physical properties, such as relative dielectric constant, carrier mobility, carrier saturation velocity, breakdown field strength, and conductivity. As these materials reach their physical limits, maintaining optimal device performance becomes challenging. Although Copper is less prone to electromigration compared to Aluminum, it is more vulnerable to open defects when used as interconnect wiring. Ultra-thin SiO2 gates have shown greater reliability due to better uniformity in thin films and fewer trapped charges from tunneling. To address these limitations, new materials such as high-k gate dielectrics and novel structures like three-dimensional MOSFETs must be carefully introduced to prevent potential issues with reliability and yield during integration (Martin, 2011).

Power Utilization limitations:

Power consumption is a key limiting factor for high-performance logic CMOS integrated circuits. Reducing the supply voltage is the most effective method for decreasing dynamic power consumption (Iwai, 2009). If current trends in chip frequency and the number of transistors continue, the power consumption of high-performance microprocessors could reach 10 kW within a few years, and heat generation on the silicon chip surface could reach 1000 W/cm², comparable to the surface temperature of a rocket nozzle (Iwai, 2007). This significant increase in power density results from inadequate supply voltage reduction and the exponential growth in transistor density. Although challenging, low-voltage technology combined with appropriate control of chip density and dimensions, along with new cooling technologies, could partially address this issue. Additionally, innovative system power management techniques, such as variable clock frequency and variable voltage supply, will also help mitigate this problem. Furthermore, gate leakage current can be reduced by using thicker high-K dielectrics, and sub-threshold leakage current can be minimized by employing a three-dimensional (3D) structure like finFET (Iwai and Wong, 2006), as shown in Figure 8. These approaches are proposed for low standby power devices, potentially even before high-performance logic units are fully realized and manufacturing costs become feasible (Iwai and Wong, 2006).

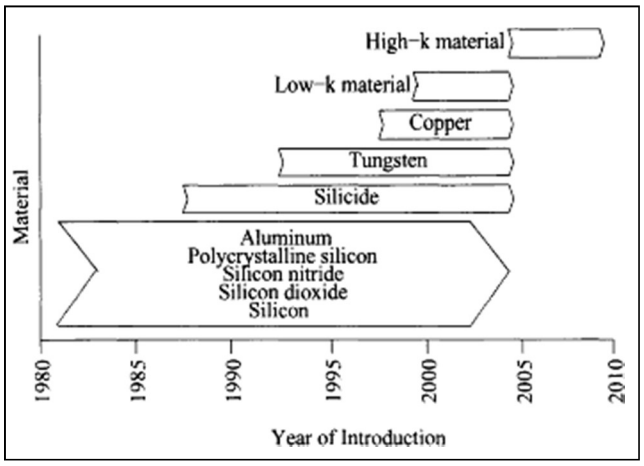

Fig. 7: Introduction of new material.

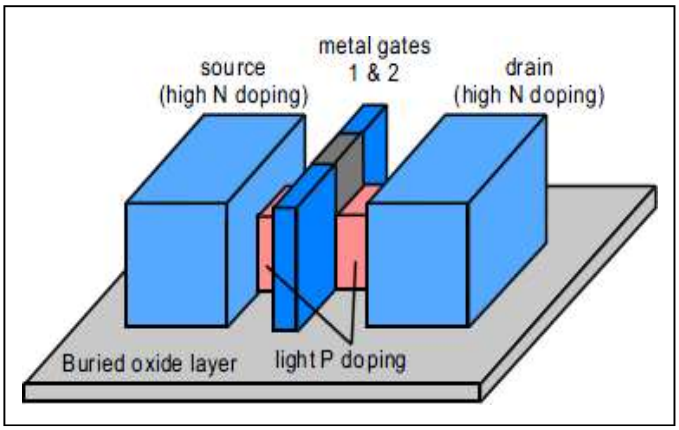

Fig. 8: Multiple gate or Fin-FET structure should be used to have a better control of the short-channel effects.

Economical Limitations:

The rising costs in the semiconductor industry are primarily driven by production and testing expenses, which have been growing exponentially as CMOS sizes decrease. A new wafer foundry, as illustrated in Figure 9, currently costs around $25 million, and the National Institute of Standards and Technology (NIST) projected that this cost would double by 2010. The increase in costs is largely attributed to the expense of equipment, clean room facilities, and the complexities of the lithography process (Wider and Neppi, 1992). The reduced size of circuits makes them more susceptible to hard and soft faults, necessitating thorough verification for quality assurance. Additionally, the complexity of testing techniques requires more steps and time, further driving up testing costs.

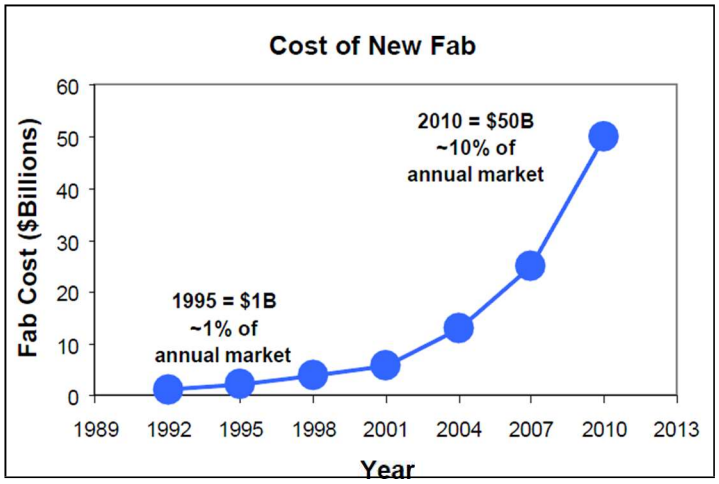

Fig. 9: Wafer foundry cost.

The minimum cost trends have shifted downward year by year, as depicted in Figure 10. Meanwhile, ongoing technological advancements have enabled an increase in the number of components on an IC chip (Schwierz et al., 2010). This study is fundamentally important because cost is a critical issue in the semiconductor industry. As long as the cost per component can be realistically reduced while increasing the complexity of ICs, the number of devices per chip will continue to grow from one generation of production to the next.

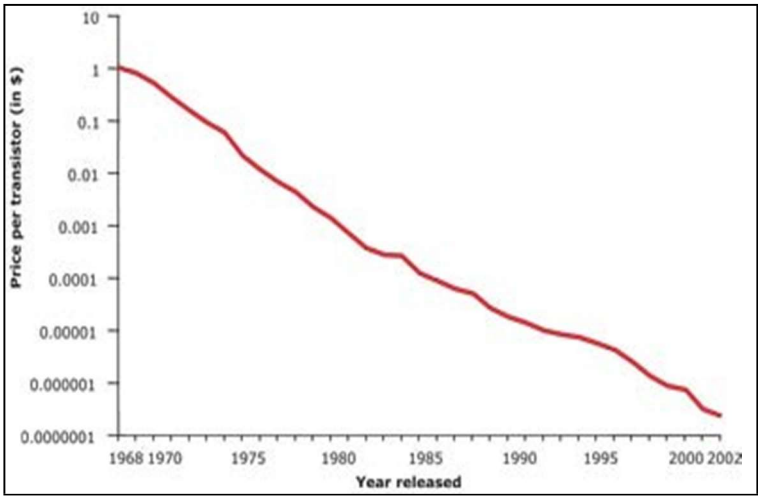

Fig. 10: Average transistor prices by year.

Technological Limitations:

CMOS transistors are created by patterning wafers using lithography and masks. Lithography technology is a key driver of transistor miniaturization. However, it struggles

to keep up with the shrinking dimensions of CMOS transistors. Techniques like proximity X-ray steppers and ion beams face challenges due to the difficulties in managing the mask-wafer distance and ensuring uniform exposure of photoresists on the wafer. Another issue is the inability of the polishing process to maintain consistent wafer thickness and reliable mask quality, as noted by Gupta et al. (2003). According to Skotnicki et al. (2005), patterning features smaller than the wavelength of light requires a trade-off between complexity, expensive masks, and potential design limitations. Figure 11 illustrates the evolution of masks from the 180 nm technology to the most recent advancements.

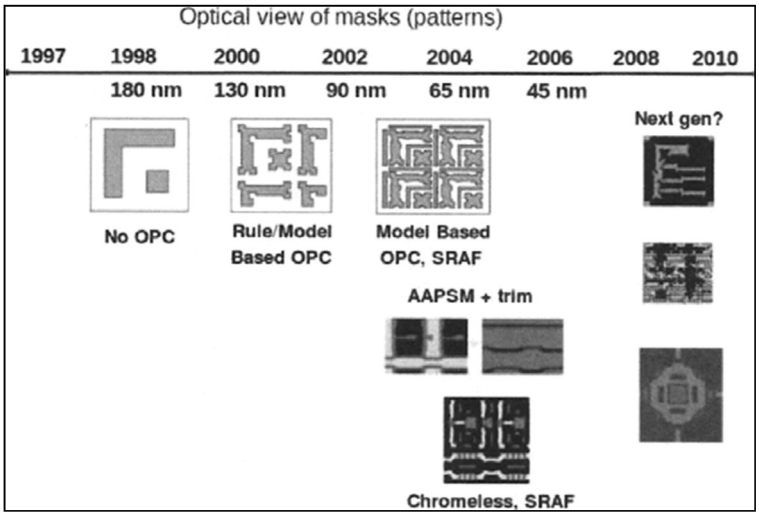

Fig. 11: Evolution of CMOS mask pattern.

Conclusion:

The advancement of CMOS technology as it scales down from 22nm to 7nm has presented numerous design challenges and opportunities. The limitations associated with CMOS miniaturization are examined in terms of performance, materials, power consumption, cost, and technological constraints. The development of new materials and technologies is expected to enhance CMOS performance. High-k dielectric materials, in particular, are being investigated for their potential to reduce current leakage and decrease power consumption, aided by adjustable clock frequencies and voltage supplies. However, while lithography remains a crucial process in transistor downsizing, it faces significant difficulties and high costs that hinder its ability to keep pace with CMOS scaling. The focus of technological advancements will be on manufacturing processes aimed at producing affordable chips. As a result, manufacturing progress is anticipated,

continuing the technological evolution in clean rooms, wafers, and related equipment.

References

Akter, M., M.B.I. Reaz, F. Mohd-Yasin and F. Choong, 2008. Hardware implementations of an image compressor for mobile communications. Journal of Communications Technology and Electronics, 53(8): 899-910.

Akter, M., M.B.I. Reaz, Mohd-Yasin, F. and F. Choong, 2008. A modified-set partitioning in hierarchical trees algorithm for real-time image compression. Journal of Communications Technology and Electronics, 53
(6): 642-650.

Dennard, R.H., F.H. Gaensslen, L. Kuhn and H.N. Yu, 1972. Design of Micron MOS Switching Devices. In International Electron Devices Meeting, pp: 168-170.

Gupta, P. and A.B. Kahng, 2003. Manufacturing-aware physical design. In Proceeding of IEEE IACM International Conference on Computer-Aided Design, pp: 681-687.

Haron, N.Z. and S. Hamdioui, 2010. Why CMOS Scaling Coming to an END. In the Proceedings of 3rd International Design and Test Workshop (IDT), pp: 98-103.

Iwai, H. and H. Wong, 2006. Nano-CMOS Technology for Next Fifteen Years. In IEEE Region 10 Conference TENCON, pp:1-4.

Iwai, H. and S. Ohmi, 2002. Silicon Integrated Circuit Technology from Past to Future. Microelectronics Reliability, 42(4-5): 465-491.

Iwai, H., 2003. CMOS Scaling toward sub 10nm regime. In the Proceedings of 11th IEEE International Symposium on Electron Devices for Microwave and Optoelectronic Applications, pp: 30-34.

Iwai, H., 2007. Future of Nano CMOS Technology. In the Proceeding of 2007 International Workshop on Electron Devices and Semiconductor Technology (EDST), pp: 57-61..

Iwai, H., 2007. Future of Silicon Integrated Circuit Technology. In the Proceeding of 2nd International Conference on Industrial and Information Systems (ICIIS 2007), pp: 571-576..

Iwai, H., 2008. CMOS technology after reaching the scale limit.In the proceedings of the IEEE-EDS 8th International Workshop on Juction Technology, pp: 1-2.

Iwai, H., 2009. Roadmap for 22nm and beyond. Microelectronic Engineering, 86(7-9): 1520-1528.

Iwai, H., 2009. Technology Roadmap for 22nm and Beyond. In the proceedings of 2nd International Workshop on Electron Devives and Semiconductor Technology (IEDST'09), pp:1-4.

Iwai, H., 2012. Materials and structures for future nano CMOS. In the IEEE Nanotechnology Materials and Devices Conference (NMDC), pp: 14-18.

Kin, K. and G.S. Park, 2011. Landscape for Semiconductor Analysis: Issues and Challenges. In proceeding of the 18th IEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA), pp: 1-9.

Kuhn, K.J., 2009. Moore's Law past 32nm: Future Challenges in Device Scaling. In the 13th International Workshop on Computational Electronics (IWCE'09), pp: 1-6.

Kuhn, K.J., 2011. CMOS Scaling for the 22nm Node and Beyond: Device Physics and Technology. In the Proceeding of International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA), pp: 1-2.

Martin, M.F., 2011. High-k/Metal Gate Innovations Enabling Continued CMOS Scaling. In the Proceedings of ESCCIRC 2011, pp: 50-58.

Marufuzzaman, Mohd., M.B.I. Reaz, M.S. Rahman and M.A. Mohd. Ali, 2010. Hardware prototyping of an intelligent current dq PI controller for FOC PMSM drive. In Proceedings of 6th International Conference on Electrical and Computer Engineering, pp: 86-88.

Mogaki, S., M. Kamada, T. Yonekura, S. Okamoto, Y. Ohtaki and M.B.I. Reaz, 2007. Time-stamp service makes real-time gaming cheat-free. Proceedings of the 6th ACM SIGCOMM Workshop on Network and System Support for Games, NetGames '07, pp: 135-138.

Mohd-Yasin, F., A.L. Tan and M.B.I. Reaz, 2004. The FPGA prototyping of Iris recognition for biometric identification employing neural network. Proceedings of the International Conference on Microelectronics, ICM., pp: 458-461.

Moore, G.E., 1965. Cramming More Components onto Integrate Circuits. Electronics Magazine, 38(8).

Reaz, M.B.I. and L.S. Wei, 2004. Adaptive Linear Neural Network Filter for Fetal ECG Extraction. Proceedings of the International Conference on Intelligent Sensing and Information Processing, pp: 321-324.

Reaz, M.B.I., F. Mohd-Yasin, S.L. Tan, H.Y. Tan and M.I. Ibrahimy, 2005. Partial encryption of compressed images employing FPGA. In Proceedings of IEEE International Symposium on Circuits and Systems, pp: 2385-2388.

Reaz, M.B.I., F. Choong and F. Mohd-Yasin, 2006. VHDL modeling for classification of power quality disturbance employing wavelet transform, artificial neural network and fuzzy logic. Simulation, 82(12): 867-881.

Reaz, M.B.I., F. Choong, M.S. Sulaiman and F. Mohd-Yasin, 2007. Prototyping of wavelet transform, artificial neural network and fuzzy logic for power quality disturbance classifier. Electric Power Components and Systems, 35 (1): 1-17.

Reaz, M.B.I., M.I. Ibrahimy, F. Mohd-Yasin, C.S. Wei and M. Kamada, 2007. Single core hardware module to implement encryption in TECB mode.Informacije MIDEM, 37(3): 165-171.

Reaz, M.B.I., M.T. Islam, M.S. Sulaiman, M.A.M. Ali, H. Sarwar and S. Rafique, 2003. FPGA realization of multipurpose FIR filter. In Proceedings of Parallel and Distributed Computing, Applications and Technologies, pp: 912-915.

Schwierz, F., H. Wong and J.J. Liou, 2010. Nanometer CMOS. Pan Standford.

Skotnicki, T., Hutchby, J.A., T. King, H.-S.P. Wong, F. Boeuf, 2005. The End of CMOS Scaling: Towards the Introduction of New Materials and Structural Changes to Improve MOSFET Performance. IEEE Circuits and Devices Magazine, 21(1):16-26.

Wider, A.W. and F. Neppi, 1992. CMOS Technology Trends and Economics. IEEE Micro, 12(4):10-19.