

Nanoscale CMOS Scaling Trends, Challenges and Their Solutions

Misha Urooj Khan^{*1}, Aqsa Arshad³, Asad-ur-Rehman², Syeda Ume Rubab Bukhari², Aqsa Samer¹, Faizan Tariq²

¹Department of Electronics Engineering, University of Engineering and Technology (UET), Taxila, Pakistan

²Department of Electrical Engineering, University of Engineering and Technology (UET), Taxila, Pakistan

³Department of Computer Engineering, University of Engineering and Technology (UET), Taxila, Pakistan

*misha.urooj@students.uettaxila.edu.pk

Abstract

The market for power-sensitive designs has increased dramatically in recent years, owing to the growing production of devices that are powered by portable batteries. Subthreshold computer architecture has grown in importance as technology scaling continues unabated. The short-channel effect is a significant problem when reducing the gate length to less than 0.1 μm . This paper provides a systematic analysis and possible strategies for the emergence of CMOS, which is the leading technology today, as suggested by numerous researchers. In addition, we identify and explain the problems and remedies for low voltage and power applications, providing a foundation for computer developers operating independently in the submicron (SM) and deep submicron (DSM) regions of CMOS systems.

Keywords

CMOS, deep submicron, low power, low voltage and submicron

1. Introduction

Metal oxide semiconductor field-effect transistor (MOSFET) is the primary component used in integrated circuits. It has been out of date for the last two decades. With technical innovation, silicon MOSFET-based electronic devices and circuits have consistently offered efficiency enhancements and cost savings, as well as system design robustness [1][2][3]. CMOS technology is one of the most promising innovations in the microchip production environment, and it is commonly used in multiple and varied application areas to build CPUs, and electronic devices justifiably making the most of these new technologies because of a handful of significant benefits in integrated circuit design [9]. In today's digital memories, Both P channel and N channel semiconductor systems are used for this application [9]. The CMOS system is one of the most common MOSFET technologies accessible today [3]. This is the prevailing semiconductor technology for microcontrollers, microprocessor modules, memories, and integrated circuits which are unique to use [4-5]. **Figure 1** shows the scaling trends from 2005 till now [15].

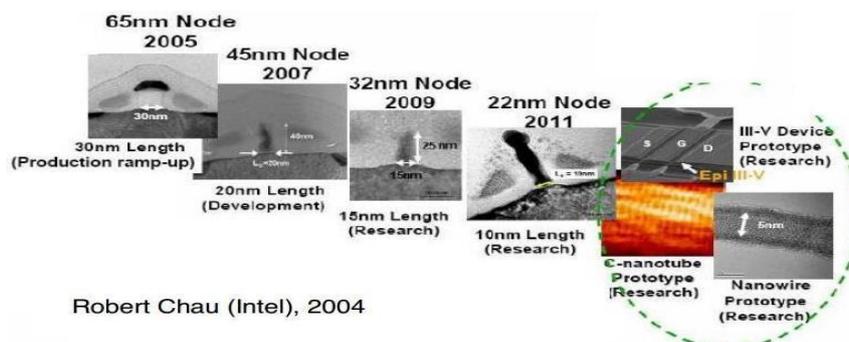


Figure 1. CMOS scaling trends by Robert Chau Intel (2004) [15].

A. Low Power Design Parameters

The low power design parameters which affect the nanoscale CMOS trends are threshold voltage, length of the channel, thickness, of the oxide, and doping concentration of the channel [3]. Differences in process parameters are now a big problem for low-power design, as the equipment is downscaled [4] [7] [9-10]. The low power system design needs to be such that variations in design parameters are less durable. The variation in these device parameters is found to be substantial in the coming generations as technology falls monotonically [3-7]. As a result, the circuit efficiency would be lower. Variability of the leakage capacity and delay on a given die in the transistors is typical of numerous low-power design techniques [1-10]. Direct tunneling is important in the off and on MOSFET transistor modes. The thickness of gate oxide has also been surpassed in modern manufacturing techniques. This process also results in a dramatic dispersion of the cumulative standby leakage capacity, in comparison to the subthreshold leakage [9-11]. **Figure 2** shows the block diagram of CMOS (Complementary Metal Oxide Semiconductor).

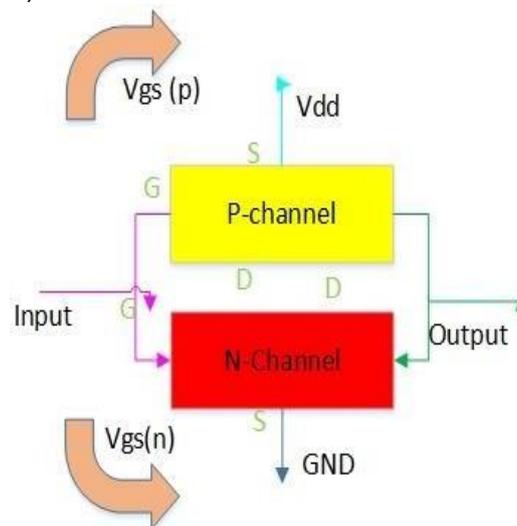


Figure 2. Block diagram of CMOS.

B. Front End Line Difficulties (FELD)

Some FELD posed by transistors in the DSM areas involve,

- i. Interface heterogeneity.
- ii. Increased V_{th} or V_{dd} .
- iii. Loss of doping.
- iv. Thin interconnections leading to saturation doping.
- v. Statistical doping variations on small geometry systems.
- vi. Increasing channel doping concentrations.
- vii. Decreasing carrier mobility.

2. Scaling Trends of Nanoscale CMOS

Scaling of the nanoscale CMOS devices is based on the most basic principles found such as the size, density, power of equipment and connections, performance, and reliability [3]. The key emphasis of this paper is the,

- i. Threshold voltage scaling with the size of the function,
- ii. Characteristic size,
- iii. Supply voltage.

By hardware scaling the length of the MOS channel is shortened. The gate-oxide voltage and thickness are therefore reduced as the dimensions of the MOSFET are lessened. V_{dd}'s supply voltage is also minimized to limit energy consumption. Since the diameter of the source and drain body degradation is set depending on the material, the rate by which the height of the barrier decreases as a function of distance around the pipe from the source is constant [1]. Polysilicon has traditionally been used as the gate material, with silicon dioxide serving as the insulating material. The rapid growth of CMOS in recent years has resulted in a decrease in silicon dioxide gate dielectric thickness, which is now less than 20° A. The scaling trends in [4] indicate that new device architectures should be tested not only for channel electrical conductivity but also for resistance and parasitic capacitance resistance. As the distance between neighbouring devices is reduced to tens of nanometres, the gradual increase in parasites can be viewed qualitatively at the end of the road map, as the source/drain and contact size must be significantly reduced to allow increased density if the gate length is not scaled. The effect of critical variables such as oxide thickness and CMOS scale leakage is measured. Because of the small sized contact scale, contact resistance and capacitance are increased. **Figure 3** shows the trends of nanoscale CMOS circuits over the past many generations of Intel devices [4].

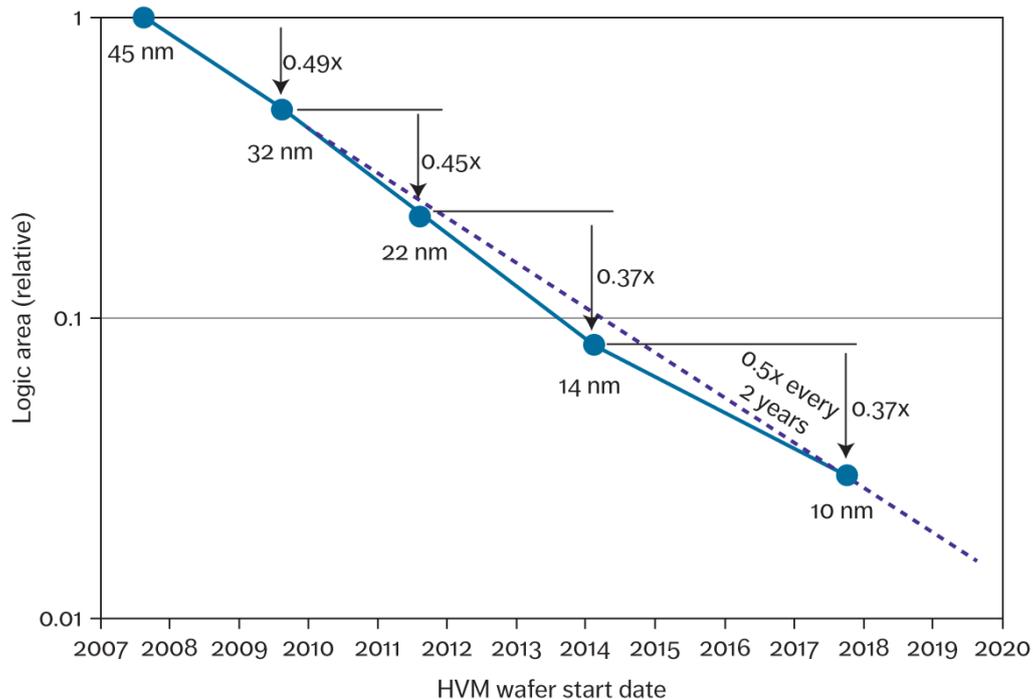


Figure 3. Nanoscale CMOS trends [4].

3. Challenges Faced During the Submicron & Deep Submicron CMOS Region's

Many studies have been conducted on nanoscale CMOS and MOSFET to fully comprehend the behaviour patterns of sub-100 nm devices. The major problems are bulk MOSFETS in the SM and DSM areas, they are explored in detail in the below sections.

A. Gate Oxide Thickness

An oxide thickness of 3 nm is needed for CMOS devices with lengths of 100 nm or less. This thickness is composed of just a few atomic particles and adheres to simple standards. Scaling that increases gate duration which often lowers dielectric thickness, thus raising the thickness of standard oxides which in turn continuously contributes to loss in efficiency and undesirable current leakage. The thickness of the oxide raises the leakage sub-threshold in a nanoscale application where the short channel effect (SCE) is highly severe [6]. The thickness of the

gate-oxide is limited almost concerning the duration of the channel so that it would keep under control the undesirable electrostatic impact on the threshold stress (i.e., small channel impact).

B. Leakage current

The operation of gate-leakage current is critical in low power CMOS circuit development. Specific emphasis is also being given to the evaluation of gate current via a thin-gap oxide tunnelling between both the polysilicon gap and the substrate as it is a significant leakage feature in nanoscale MOSFET and CMOS. When the oxide grows steadily, the scaling down the thickness of the gate is also being increased. **Figure 4** indicates that the key leakage factor for 45-nm CMOS technology is the sub-threshold leakage, but leakage junctions and gate leakage contributions have greatly improved in scaled systems [8]. In terms of leakage, oxide thickness, and doping profiles, various components of the nanotechnology behave differently [4]. The elimination of drain-inducing barriers increases the off-state I_{OFF} as well Drain-induced barrier lowering (DIBL). The barrier height limits both channel length and drains voltage due to the shorter channel length. Disparity in channel length modulates the height of the firewall, resulting in a decrease in threshold height [6].

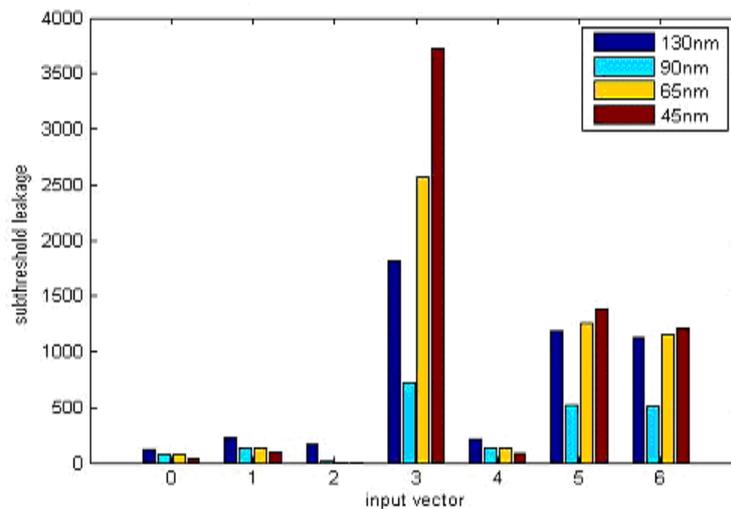


Figure 4. Comparison of leakage currents of different nanoscale CMOS technologies [8].

C. Drain Leakage due to Gate induction

At low current and high applied voltages, the NMOS gate drainage current is crucial. In-room temperature, the thermal voltage (kT/q) of the electron is stable; the working voltage-to-thermal voltage ratio decreases. This improves electron leakage from the heat diffusion source to the drain. Gate Induced Drain Leakage (GIDL) increases the electric field [11] by having a thinner oxide layer and a higher V_{DD} . The electric fields are insufficiently high in tunnels with weak drainage values, and the depletion diameter and tunnelling are reduced at extremely high doping speeds, resulting in a lower GIDL [12].

D. Tunnelling Currents

The scaling of the transistors resulted in steeper halo implants whereby substrate doping is enhanced while the channel doping remains small. It offers DIBL regulation with lower channel versatility results. The subsequent high doping profile at the drainage edge raises the tunneling current from band to band. The leakage of the diode field in the source and diodes is typically negligible for the I_{OFF} belt [1].

E. Threshold Voltage

When the sub-threshold slope is essentially fixed, it increases I_{OFF} exponentially. The diffusion current is governed by sub-threshold conduction in the strong inversion region where the drift

current operates. Carriers migrate around the surface through diffusion, which is analogous to the charging of transportation around the base of bipolar transistors [3] [5].

4. Solution of Challenges

Sub-thresholds are an efficient means of obtaining ultra-low power for applications that do not need high efficiency. Lowering the voltage level is not only helpful for mitigating active power but also for reducing leakage. When the gate overdrives $V_{gs} - V_{th}$ reduces, Scaling V_{dd} slows down the chain. In this context, [12] recommends the implementation of dynamic voltage scaling to the lower V_{dd} systems that needed to fulfil the output requirements. To eliminate toxins or a lower doping standard of the substrate which results in a higher W_{dm} , the slope of the substrate S may be decreased by utilizing a thinner layer of oxide. Increase in working conditions, i.e., lower temperatures or bias in the substrate, often rise the value of S [13]. To minimize short-channel effect, oxide thickness and higher & non-uniform doping will be introduced when the devices are positioned in a nanometer structure.

A. Use of High k Dielectrics

The primary explanation for replacing SiO_2 with a dielectric substitute gate is the removal of leakage. As indicated in [8], a solution may have a higher quality of k value than silica. This would increase the real thickness of the dielectric gate to preserve the same electromagnetic current in the channel. By 2016, Si_3N_4 or SiO_xN_y should be able to be used in elevated performance, reduced power applications [2]. Transistors with a 30 nm length, a dielectric with $7A^\circ$, and a SiO_xN_y gate have shown to have outstanding electrical characteristics, indicating that this gate is suitable for high-performance applications at least up to $7A^\circ$.

B. Strained Si

The process by which the silicone grid is dynamically stressed is known as strained Si [3]. As compared to other high-mobility III–V semiconductors, the incorporation of a lattice strain Si channel changes the framework of the belt and solves the Si transport defect. The strain is often used for both low channel conductivity weight and high state power, resulting in an extremely large conductivity weight perpendicular to channel orientation at the transistor level [7]. The strain increases transistor efficiency by increasing channel mobility by decreasing the productive NMOS weight and interval dispersion rate of the electron and growing the effective PMOS weight and band dispersion rate of the hole [6]. **Figure 5** shows a simple silicon vs strained silicon [4].

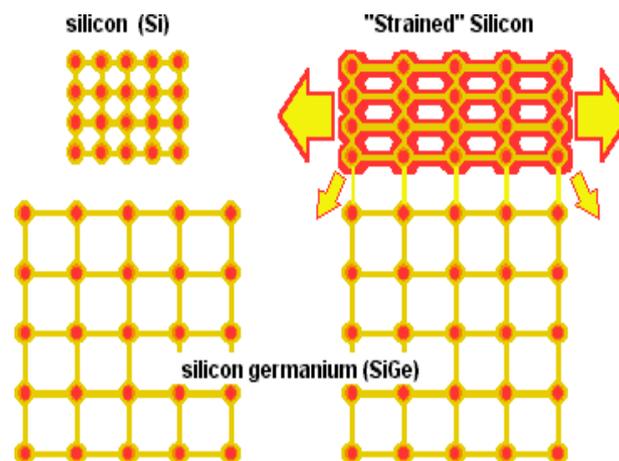


Figure 5. Simple silicon vs strained silicon [4].

C. Gate Stack Structures with High-k values

To minimize gate current in nanoscale CMOS, silicon dioxide sheets have been replaced by high-k gate stack structures. Nitrogen build-up in the dielectric prevents doping from the polysilicon gate from diffusing into the conduit, resulting in threshold voltage increases. Nitrogen absorption in hafnium silicate, hafnium aluminate, and HfO_2 improves dielectric

silicate stiffness, inhibits doping diffusion in high-temperature polysilicon channel cooling, and increases the crypto temperature of high-k stacks [9].

D. Metal Gate Technology High-k values

The high-k metal-gate method reduces intrinsic volatility. A new type of modification is created by the usage of metal as a gating substrate owing to the mechanical dependency of metal as the direction of the metal grain. Influences of variations in the intrinsic parameters and variations in the operation of the metal gate, method vector influence can also be decreased significantly. The 16-nm CMOS with dc, timing details, power information, and high-frequency circuits are recorded in [4]. The detailed research analyses variations in the nature and efficiency of CMOS technology and durability in the sub-22 nm period [6]. **Figure 6** shows the usual silicon transistor vs a high k-based transistor [5].

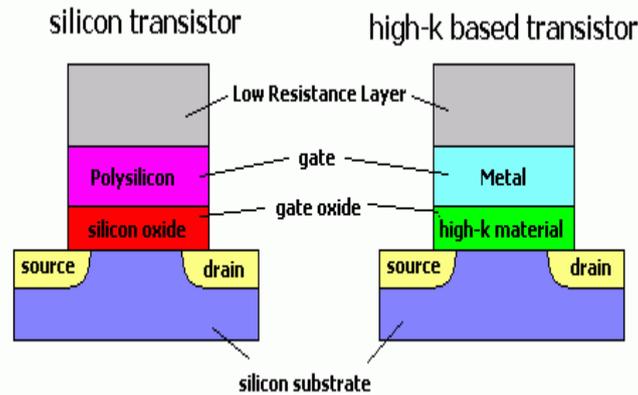


Figure 6. Metal Gate Technology and High-k based transistor [5].

E. Insulated Shallow Extension Structure

Excellent power over SCE is accomplished using shallow junction technologies with gate lengths as small as 20 nm [5]. For optimized short-circuit behaviour, it is important to control the threshold voltage roll-off [4].

F. Lateral Channel Engineering

Halo profiles work to render higher for shorter FETs Average doping in the system as shown by a longer FET path, thereby appearing to boost V_{th} as compared to the short-channel results that lower it. These halos are used for the construction of a 25-nm CMOS [6]. This strategy adds a layer of independence to efforts to mitigate the negative effects of SCE [3].

G. Highly Doped Gates

The results of poly depletion would intensify with the start of MOSFET Scaling due to Corner and Edge's impact. The most promising solution to resolving such problems would be to achieve heavily doped gates with a lower abomination gradient in the Polygate.

H. Lowering the Supply Voltage

Bulk sizing Halo implants are used in Ge PMOS technologies to ensure that the transistor channel is properly controlled and similar to silicone. The halos increase the electrical field on the drain line, resulting in a larger drain leak. Since the BTBT system is highly electrically dependent, lowering the supply voltage will result in a lower voltage of the device [1][3].

5. CMOS Roadmap in Tabular Form

High-performance (HP) systems have the most extreme scaling because very small threshold voltages are typically used, requiring high SCE and DIBL voltages. Based on the three broad component families as high performance (HP), low operating power (LOP), and low standby power (LSTP), a guide for SM and DSM systems is set out in Tables 1 and 2.

A. Comparison of SM CMOS technology

Following tables shows the comparison of SM CMOS technology, their performance factors, the challenges faced by them, solution to their challenges, their enhanced performances and their applications.

Table 1 - Comparison of SM CMOS devices and technology

Operation Range	Performance factors	Challenges	Solution	Applications
Sub-100nm	Increased off-state leakage because of lowering V_{th} under tech scalable sizing.	Reduction of leakage and gate currents.	Temperature reduction. Solutions for insulation.	LOP [3]
Sub-95 nm	As devices downscale, their output degrades due to poly depletion impacts.	Poly-depletion effects.	Using less dopant Gradient. Highly doped gates	LSTP [4]
Sub-90 nm	V_{th} decreases in the short channel area.	Subthreshold swing. Electric field reduction at drain.	Increasing junction depths and channel doping MOSFET with grooved gate	LOP [7]
Sub-80 nm	Controlling gate-leakage current is critical in the design of low-power CMOS circuits.	Reduced gate currents.	There is an interfacial oxide layer between the high-k layer and silicone substratum and the intermediate layer between the high-k dielectric and silicone substrate is accessible. Pile of high-k gate	LOP [8]
Sub-70 nm	Adverse V_{th} roll off	V_{th} roll off control	Ion implantation Pocket implant	HP-LOP [9]
Sub-65 nm	A high concentration of bulk impurities causes an increase in V_{th} .	Reduce V_{th}	Low-doped region adjacent to high-doped surface Doping performance in stages	HP [10]
Sub-45 nm	Barrier reduction in both the inversion channel and the body depletion field.	V_{th} adjustment Body punch-through	Implantation with a slanted angle. A super halo and a retrograde channel	HP-LOP [11]

B. Comparison of deep submicron CMOS technology

Following tables shows the comparison of SM CMOS technology, their performance factors, the challenges faced by them, solution to their challenges, their enhanced performances and their applications.

MOSFET transistor researchers are exploring system layout and channel material improvements to enable more generations of MOSFET scaling [1]. The MOSFET implemented with multiple horizontal nanowire stacks is one alternative that, due to its superior electrostatics, may require more gate-length scaling beyond [4] what F in FET can achieve. MOSFETs with III-V semiconductor channel materials are a good choice for a higher-mobility channel [9] than silicon.

Table 2 - Comparison of deep SM CMOS devices and technology

Operation range	Performance factors	Challenges	Solution	Applications
Sub-55nm	SCE improvement and push existing loss	Voltage scaling is limited. V _{th} scalability	Doping profile that is not standardised. Engineering of lateral channels.	HP [2]
Sub-50 nm	Revere-biased diode junc. BTBT current	Reduce mobility and band to band tunnelling current.	Reduce the concentration of peak halo doping. Doping with asymmetric halo (AH)	LOP-LSTP [3]
Sub-45 nm	Because of changes in system parameters, there is a wide range of variance in various leakage components.	Variability in V _{th} caused by spontaneous dopant volatility and junction capacitance.	Strengthening of the halo. AH halo that has been changed.	LOP [6]
Sub-48 nm	V _{th} roll-off detrimental	Drain-induced barrier lowering (DIBL)	On the root line, there is a significant amount of halo doping. Non-uniform doping	LOP [8]
Sub-40 nm	Quick channel V _{th} roll off and gate/drain leaking.	Band to band tunnelling control. V _{th} roll-off.	Halo for distance enlargement. Localized halo under the top of the channel.	LOP [9]
Sub-35 nm	It is challenging to employ superhalo for PMOS horizontally with a slope of 4-5 nm/dec net doping concentrations	Short channel results that are appropriate.	SALVO's lateral net doping profile	LSTP [10]
Sub-30 nm	The job mechanism of metal gates influences the V _{th} of the system as well as the tuning and power of digital circuits.	Fluctuations of intrinsic parameters.	High-k dielectrics used in metal gate technology.	HP-LOP [11]
Sub-18 nm	V _{th} of unit is affected by process heterogeneity and spontaneous dopant fluctuations.	Variation of inherent values.	Work function varies at random. The use of metal gates	HP-LSTP [12]

6. Conclusions

Sub-threshold architecture is an obvious choice for achieving low power consumption on the semiconductor roadmap. Device, circuit, and architecture level changes unique to the sub-threshold step must be applied to achieve optimal output. Since substrate circuits are extremely vulnerable to process irregularities, it is important to use advanced design strategies to improve circuit effectiveness. Improved channel stability due to channel pressure is a key factor in achieving MOSFET efficiency specifications. Numerous significant technologies and material developments, such as high-k dielectric, metal gate electrodes, elevated source/drain, and doping techniques, must be applied to effectively scale ICs to achieve efficiency, leakage current, and other specifications. Disruptions, statistical process instability, the influence of quantum effects, line edge friction, and variation in ultra-thin body width must all be understood for improved DSM performance, especially in low-power applications.

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-Meim and, "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.

- [17] M. U. Khan, S. Z.-Z. Ali, A. Ishtiaq, K. Habib, T. Gul, and A. Samer, "Classification of Multi-Class Cardiovascular Disorders using Ensemble Classifier and Impulsive Domain Analysis," in *Mohammad Ali Jinnah university International Conference on Computing 2021 (MAJICC'21)*, 2021, pp. 1–8.
- [18] M. U. Khan, A. Farman, A. ur Rehman, N. Israr, M. Z. H. Ali, and Z. A. Gulshan, "Automated System Design for Classification of Chronic Lung Viruses using Non-Linear Dynamic System Features and K-Nearest Neighbour," in *Mohammad Ali Jinnah university International Conference on Computing 2021 (MAJICC'21)*, 2021, pp. 1–8.
- [19] M. U. Khan, D. S. A. A. Zaidi, A. Ishtiaq, S. U. R. Bukhari, S. Samer, and A. Farman, "A Comparative Survey of LiDAR-SLAM and LiDAR based Sensor Technologies," in *Mohammad Ali Jinnah university International Conference on Computing 2021 (MAJICC'21)*, 2021, pp. 1–8.
- [20] M. U. Khan, A. Mobeen, S. Samer, and A. Samer, "Embedded System Design for Real-time Detection of Asthmatic Diseases Using Lung Sounds in Cepstral Domain," in *6th International Electrical Engineering Conference (IEEC 2021) April, 2021 at NEDUET, Karachi, Pakistan, 2021*, pp. 1–5.
- [22] A. Zainab, M. U. Khan, H. Khan, and S. U. R. Bukhari, "Intelligent System Design for Early Diagnosis of Faults in Machine Bearings," in *6th International Electrical Engineering Conference (IEEC 2021) April, 2021 at NEDUET, Karachi, Pakistan, 2021*, pp. 1–5.
- [23] S. Z. H. Naqvi, M. Arooj, S. Aziz, M. U. Khan, M. A. Choudhary, and M. N. Ul Hassan, "Spectral Analysis of Lungs sounds for Classification of Asthma and Pneumonia Wheezing," in *2nd International Conference on Electrical, Communication and Computer Engineering, ICECCE 2020, Jun. 2020*, pp. 1–6, doi: 10.1109/ICECCE49384.2020.9179417.
- [24] A. Ishtiaq, M. U. Khan, S. Z. Ali, K. Habib, S. Samer, and Enass Hafeez, "A Review of System on Chip (SOC) Applications in Internet of Things (IOT) and Medical," in *ICAME21, International Conference on Advances in Mechanical Engineering, Pakistan, 2021, no. August*, pp. 1–10.
- [25] M. U. Khan, S. Z. Ali, K. Habib, H. Khan, F. Tariq, and S. Bibi, "A Novel Intelligent Model For COVID-19 Detection Using Cough Auscultations and Hjorth Descriptors," in *ICAME21, International Conference on Advances in Mechanical Engineering, Pakistan, 2021, no. August*, pp. 1–10.
- [26] M. U. Khan, A. Mobeen, S. Samer, and A. Samer, "Embedded System Design for Real-time Detection of Asthmatic Diseases Using Lung Sounds in Cepstral Domain," in *6th International Electrical Engineering Conference (IEEC 2021)*, 2021, p. 6, Accessed: Aug. 02, 2021.